Assessing spatio-temporal hydrologic variability:
a case-study in western Victoria

by

Stuart Cameron Brown, BEnvSci (Hons)

Submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

Deakin University

September, 2015
I am the author of the thesis entitled

‘Assessing spatio-temporal hydrologic variability: a case-study in western Victoria’

submitted for the degree of Doctor of Philosophy

This thesis may be made available for consultation, loan and limited copying in accordance with the Copyright Act 1968.

'I certify that I am the student named below and that the information provided in the form is correct'

Full Name:  Stuart Cameron Brown

Signed: [Signature Redacted by Library]

Date:
I certify the following about the thesis entitled (10 word maximum)

‘Assessing spatio-temporal hydrologic variability: a case-study in western Victoria’

submitted for the degree of Doctor of Philosophy

a. I am the creator of all or part of the whole work(s) (including content and layout) and that where reference is made to the work of others, due acknowledgment is given.

b. The work(s) are not in any way a violation or infringement of any copyright, trademark, patent, or other rights whatsoever of any person.

c. That if the work(s) have been commissioned, sponsored or supported by any organisation, I have fulfilled all of the obligations required by such contract or agreement.

I also certify that any material in the thesis which has been accepted for a degree or diploma by any university or institution is identified in the text.

'I certify that I am the student named below and that the information provided in the form is correct'

Full Name: Stuart Cameron Brown

Signed: [Signature Redacted by Library]

Date:
AUTHORSHIP STATEMENT

1. Details of publication and executive author

<table>
<thead>
<tr>
<th>Title of Publication</th>
<th>Publication details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessing the impact of drought and forestry on streamflows in south-eastern Australia using a physically-based hydrological model</td>
<td>Submitted to ‘Environmental Earth Sciences’ with minor revisions</td>
</tr>
<tr>
<td>Name of executive author</td>
<td>School/Institute/Division if based at Deakin; Organisation and address if non-Deakin</td>
</tr>
<tr>
<td>Stuart C Brown</td>
<td>School of Life &amp; Environmental Sciences</td>
</tr>
</tbody>
</table>

2. Inclusion of publication in a thesis

| Is it intended to include this publication in a higher degree by research (HDR) thesis? | Yes | If Yes, please complete Section 3 If No, go straight to Section 4. |

3. HDR thesis author’s declaration

| Name of HDR thesis author if different from above. (If the same, write “as above”) | School/Institute/Division if based at Deakin | Thesis title |
| As above | School of Life & Environmental Sciences | Assessment of Spatio-temporal Hydrologic Variability: a case-study in Western Victoria |

If there are multiple authors, give a full description of HDR thesis author’s contribution to the publication (for example, how much did you contribute to the conception of the project, the design of methodology or experimental protocol, data collection, analysis, drafting the manuscript, revising it critically for important intellectual content, etc.)

I declare that the above is an accurate description of my contribution to this paper, and the contributions of other authors are as described below.

Signature and date

4. Description of all author contributions

<table>
<thead>
<tr>
<th>Name and affiliation of author</th>
<th>Contribution(s) (for example, conception of the project, design of methodology or experimental protocol, data collection, analysis, drafting the manuscript, revising it critically for important intellectual content, etc.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Versace, V.L. (Federation University)</td>
<td>Drafting of manuscript</td>
</tr>
<tr>
<td>Lester, R.E</td>
<td>Drafting of manuscript</td>
</tr>
<tr>
<td>Walter, M.T. (Cornell University)</td>
<td>Drafting of manuscript</td>
</tr>
</tbody>
</table>
5. Author Declarations

I agree to be named as one of the authors of this work, and confirm:

i. that I have met the authorship criteria set out in the Deakin University Research Conduct Policy,

ii. that there are no other authors according to these criteria,

iii. that the description in Section 4 of my contribution(s) to this publication is accurate,

iv. that the data on which these findings are based are stored as set out in Section 7 below.

If this work is to form part of an HDR thesis as described in Sections 2 and 3, I further

v. consent to the incorporation of the publication into the candidate’s HDR thesis submitted to Deakin University and, if the higher degree is awarded, the subsequent publication of the thesis by the university (subject to relevant Copyright provisions).

<table>
<thead>
<tr>
<th>Name of author</th>
<th>Signature*</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Versace, V.L.</td>
<td>Signature Redacted by Library</td>
<td></td>
</tr>
<tr>
<td>Lester, R.E</td>
<td>Signature Redacted by Library</td>
<td></td>
</tr>
<tr>
<td>Walter, M.T.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6. Other contributor declarations

I agree to be named as a non-author contributor to this work.

<table>
<thead>
<tr>
<th>Name and affiliation of contributor</th>
<th>Contribution</th>
<th>Signature* and date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assoc. Prof. Giovanni Turchini, Assoc. HoS, School of Life &amp; Env. Sci.</td>
<td>signing on behalf of M. T. Walter</td>
<td>Signature Redacted by Library</td>
</tr>
</tbody>
</table>

* If an author or contributor is unavailable or otherwise unable to sign the statement of authorship, the Head of Academic Unit may sign on their behalf, noting the reason for their unavailability, provided there is no evidence to suggest that the person would object to being named as author.

7. Data storage

The original data for this project are stored in the following locations. (The locations must be within an appropriate institutional setting. If the executive author is a Deakin staff member and data are stored outside Deakin University, permission for this must be given by the Head of Academic Unit within which the executive author is based.)

<table>
<thead>
<tr>
<th>Data format</th>
<th>Storage Location</th>
<th>Date lodged</th>
<th>Name of custodian if other than the executive author</th>
</tr>
</thead>
</table>

This form must be retained by the executive author, within the school or institute in which they are based.

If the publication is to be included as part of an HDR thesis, a copy of this form must be included in the thesis with the publication.
## AUTHORSHIP STATEMENT

### 1. Details of publication and executive author

<table>
<thead>
<tr>
<th>Title of Publication</th>
<th>Publication details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessment of spatio-temporal varying relationships between rainfall, land cover and surface water area using geographically weighted regression</td>
<td>DOI: 10.1007/s10666-011-9289-8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name of executive author</th>
<th>School/Institute/Division if based at Deakin; Organisation and address if non-Deakin</th>
<th>Email or phone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stuart C Brown</td>
<td>School of Life &amp; Environmental Sciences</td>
<td><a href="mailto:stuart.brown@research.deakin.edu.au">stuart.brown@research.deakin.edu.au</a></td>
</tr>
</tbody>
</table>

### 2. Inclusion of publication in a thesis

<table>
<thead>
<tr>
<th>Is it intended to include this publication in a higher degree by research (HDR) thesis?</th>
<th>Yes</th>
<th>If Yes, please complete Section 3 If No, go straight to Section 4.</th>
</tr>
</thead>
</table>

### 3. HDR thesis author’s declaration

<table>
<thead>
<tr>
<th>Name of HDR thesis author if different from above. (If the same, write “as above”)</th>
<th>School/Institute/Division if based at Deakin</th>
<th>Thesis title</th>
</tr>
</thead>
<tbody>
<tr>
<td>As above</td>
<td>School of Life &amp; Environmental Sciences</td>
<td>Assessment of Spatio-temporal Hydrologic Variability: a case-study in Western Victoria</td>
</tr>
</tbody>
</table>

If there are multiple authors, give a full description of HDR thesis author’s contribution to the publication (for example, how much did you contribute to the conception of the project, the design of methodology or experimental protocol, data collection, analysis, drafting the manuscript, revising it critically for important intellectual content, etc.)

Conceptualised the method, performed and interpreted the analyses, drafted and proofed the manuscript.

I declare that the above is an accurate description of my contribution to this paper, and the contributions of other authors as described below.

**Signature and date**

[Signature Redacted by Library]

### 4. Description of all author contributions

<table>
<thead>
<tr>
<th>Name and affiliation of author</th>
<th>Contribution(s) (for example, conception of the project, design of methodology or experimental protocol, data collection, analysis, drafting the manuscript, revising it critically for important intellectual content, etc.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Versace, V.L.</td>
<td>Interpretation of analysis, drafting the manuscript</td>
</tr>
<tr>
<td>Laurensen, L.</td>
<td>Drafting the manuscript</td>
</tr>
<tr>
<td>Ierodiaconou, D.</td>
<td>Drafting the manuscript</td>
</tr>
<tr>
<td>Fawcett, J.</td>
<td>Interpretation of analysis, drafting the manuscript</td>
</tr>
<tr>
<td>Salzman, S</td>
<td>Drafting the manuscript</td>
</tr>
</tbody>
</table>
5. Author Declarations

I agree to be named as one of the authors of this work, and confirm:

i. that I have met the authorship criteria set out in the Deakin University Research Conduct Policy,

ii. that there are no other authors according to these criteria,

iii. that the description in Section 4 of my contribution(s) to this publication is accurate,

iv. that the data on which these findings are based are stored as set out in Section 7 below.

If this work is to form part of an HDR thesis as described in Sections 2 and 3, I further

v. consent to the incorporation of the publication into the candidate’s HDR thesis submitted to Deakin University and, if the higher degree is awarded, the subsequent publication of the thesis by the university (subject to relevant Copyright provisions).

<table>
<thead>
<tr>
<th>Name of author</th>
<th>Signature*</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Versace, V.L.</td>
<td>Signature Redacted by Library</td>
<td></td>
</tr>
<tr>
<td>Laurenson, L.</td>
<td>Signature Redacted by Library</td>
<td>22 May 2015</td>
</tr>
<tr>
<td>Ierodiaconou, D.</td>
<td>Signature Redacted by Library</td>
<td></td>
</tr>
<tr>
<td>Fawcett, J.</td>
<td>Signature Redacted by Library</td>
<td>22 May 2015</td>
</tr>
<tr>
<td>Salzman, S</td>
<td>Signature Redacted by Library</td>
<td></td>
</tr>
</tbody>
</table>

6. Other contributor declarations

I agree to be named as a non-author contributor to this work.

<table>
<thead>
<tr>
<th>Name and affiliation of contributor</th>
<th>Contribution</th>
<th>Signature* and date</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* If an author or contributor is unavailable or otherwise unable to sign the statement of authorship, the Head of Academic Unit may sign on their behalf, noting the reason for their unavailability, provided there is no evidence to suggest that the person would object to being named as author.

7. Data storage

The original data for this project are stored in the following locations. (The locations must be within an appropriate institutional setting. If the executive author is a Deakin staff member and data are stored outside Deakin University, permission for this must be given by the Head of Academic Unit within which the executive author is based.)

<table>
<thead>
<tr>
<th>Data format</th>
<th>Storage Location</th>
<th>Date lodged</th>
<th>Name of custodian if other than the executive author</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This form must be retained by the executive author, within the school or institute in which they are based.

If the publication is to be included as part of an HDR thesis, a copy of this form must be included in the thesis with the publication.
# AUTHORSHIP STATEMENT

1. **Details of publication and executive author**

<table>
<thead>
<tr>
<th>Title of Publication</th>
<th>Publication details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hydrologic Landscape Regionalisation Using Deductive Classification and Random Forests.</td>
<td>DOI: 10.1371/journal.pone.0112856</td>
</tr>
<tr>
<td>Name of executive author</td>
<td>School/Institute/Division if based at Deakin; Organisation and address if non-Deakin</td>
</tr>
<tr>
<td>Stuart C Brown</td>
<td>School of Life &amp; Environmental Sciences</td>
</tr>
<tr>
<td></td>
<td><a href="mailto:stuart.brown@research.deakin.edu.au">stuart.brown@research.deakin.edu.au</a></td>
</tr>
</tbody>
</table>

2. **Inclusion of publication in a thesis**

| Is it intended to include this publication in a higher degree by research (HDR) thesis? | Yes | If Yes, please complete Section 3 If No, go straight to Section 4. |

3. **HDR thesis author’s declaration**

<table>
<thead>
<tr>
<th>Name of HDR thesis author if different from above. (If the same, write “as above”)</th>
<th>School/Institute/Division if based at Deakin</th>
<th>Thesis title</th>
</tr>
</thead>
<tbody>
<tr>
<td>As above</td>
<td>School of Life &amp; Environmental Sciences</td>
<td>Assessment of Spatio-temporal Hydrologic Variability: a case-study in Western Victoria</td>
</tr>
</tbody>
</table>

If there are multiple authors, give a full description of HDR thesis author’s contribution to the publication (for example, how much did you contribute to the conception of the project, the design of methodology or experimental protocol, data collection, analysis, drafting the manuscript, revising it critically for important intellectual content, etc.)

I declare that the above is an accurate description of my contribution to this paper, and the contributions of other authors are as described below.

<table>
<thead>
<tr>
<th>Signature and date</th>
<th>Signature Redacted by Library</th>
</tr>
</thead>
</table>

4. **Description of all author contributions**

<table>
<thead>
<tr>
<th>Name and affiliation of author</th>
<th>Contribution(s) (for example, conception of the project, design of methodology or experimental protocol, data collection, analysis, drafting the manuscript, revising it critically for important intellectual content, etc.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lester, R.E</td>
<td>Assistance with experimental design and analysis, drafting of manuscript</td>
</tr>
<tr>
<td>Versace, V.L.</td>
<td>Drafting of manuscript</td>
</tr>
<tr>
<td>Fawcett, J.</td>
<td>Assistance with experimental design and analysis, drafting of manuscript</td>
</tr>
<tr>
<td>Laurenson, L.</td>
<td>Drafting of manuscript</td>
</tr>
</tbody>
</table>
5. Author Declarations

I agree to be named as one of the authors of this work, and confirm:

i. that I have met the authorship criteria set out in the Deakin University Research Conduct Policy,
ii. that there are no other authors according to these criteria,
iii. that the description in Section 4 of my contribution(s) to this publication is accurate,
iv. that the data on which these findings are based are stored as set out in Section 7 below.

If this work is to form part of an HDR thesis as described in Sections 2 and 3, I further
v. consent to the incorporation of the publication into the candidate’s HDR thesis submitted to Deakin
University and, if the higher degree is awarded, the subsequent publication of the thesis by the
university (subject to relevant Copyright provisions).

<table>
<thead>
<tr>
<th>Name of author</th>
<th>Signature*</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lester, R.E</td>
<td>Signature Redacted by Library</td>
<td>22/5/2015</td>
</tr>
<tr>
<td>Versace, V.L.</td>
<td>Signature Redacted by Library</td>
<td></td>
</tr>
<tr>
<td>Fawcett, J.</td>
<td>Signature Redacted by Library</td>
<td>22/05/2015</td>
</tr>
<tr>
<td>Laurenson, L.</td>
<td>Signature Redacted by Library</td>
<td>22 May 2015</td>
</tr>
</tbody>
</table>

6. Other contributor declarations

I agree to be named as a non-author contributor to this work.

<table>
<thead>
<tr>
<th>Name and affiliation of contributor</th>
<th>Contribution</th>
<th>Signature* and date</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* If an author or contributor is unavailable or otherwise unable to sign the statement of authorship, the Head of
  Academic Unit may sign on their behalf, noting the reason for their unavailability, provided there is no evidence to
  suggest that the person would object to being named as author

7. Data storage

The original data for this project are stored in the following locations. (The locations must be within an appropriate
institutional setting. If the executive author is a Deakin staff member and data are stored outside Deakin
University, permission for this must be given by the Head of Academic Unit within which the executive author is
based.)

<table>
<thead>
<tr>
<th>Data format</th>
<th>Storage Location</th>
<th>Date lodged</th>
<th>Name of custodian if other than the executive author</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This form must be retained by the executive author, within the school or institute in which they are based.

If the publication is to be included as part of an HDR thesis, a copy of this form must be included in the thesis
with the publication.
'If I can’t picture it, I can’t understand it.’ – Albert Einstein

‘Essentially, all models are wrong, but some are useful.’ – George Box
Acknowledgments

This thesis would not have been possible without the support, guidance, input and advice from many individuals over the years. First and foremost, I would like to thank my supervisors who imparted their wisdom during my candidature. Firstly, my primary co-supervisors, Associate Professor Laurie Laurenson and Dr Rebecca Lester. Laurie, the freedom you granted me during my candidature was very much appreciated, as was the direction and support I knew that I could count on you for when I needed it. Rebecca, although you only became a supervisor towards the end of my project, your encouragement, support and patience with me have been very, very much appreciated. Your advice on different aspects of my research was tremendous, and I thank you greatly for that. My writing skills have (hopefully) improved dramatically with your guidance and you have proven invaluable in helping me to get published and complete my studies. To my other supervisors (official and otherwise); Dr Vincent Versace, you were a continual source of encouragement and always available for discussions. Your enthusiasm and fervour towards my project definitely helped to keep me on the path to completion. I cannot thank you enough for your guidance over the years, and in particular for encouraging me to present my work in Europe. I am also indebted to you for organising my trip to Cornell University with Professor Todd Walter. Dr Jonathon Fawcett, early on in my candidature we were fortunate enough to meet where we “discussed” some of my research ideas. If it was not for that meeting (and a few others over dinner and red wine), I doubt that a significant body of work in this thesis would have evolved into what it did. Finally I would like to thank Dr Daniel Ierodiaconou and Dr Scott Salzman for
discussions and guidance early in my candidature that helped to set the direction of my work.

I would also like to thank the many friends and housemates that have been involved in my life during my candidature. If it wasn’t for you, I probably would have finished this thesis on time. Last, but certainly not least, I would like to acknowledge the support of my immediate family. To my mum, Marie; thank you for your support and belief in me over the years. Hopefully you will read this thesis one day and realise what it was I was doing all this time! To my partner, Becc; your continual encouragement, support, and patience over the years has been incredible to say the least. The sense of perspective and self-belief you have instilled in me have been tremendous and I cannot thank you enough for your positive attitude and understanding.

Many of the opportunities I was fortunate enough to enjoy during my candidature would not have been possible without assistance from numerous funding bodies. As such, I would also like to acknowledge the financial assistance I received as part of my Australian Postgraduate Award, the Victorian Department of Environment and Primary Industries, and the Glenelg Hopkins Catchment Management Authority.
List of peer-reviewed publications

Chapter 2 has been published in Environmental Earth Sciences:


Chapter 3 has been published in Environmental Modeling and Assessment:


Chapter 5 has been published in *PLoS One*:

Abstract

This thesis investigates three methods of identifying, quantifying and predicting spatio-temporal variability in relationships among environmental and hydrological variables. Spatio-temporal variability, particularly of model coefficients and residuals, is an inherent issue in environmental studies that is typically overlooked. Overlooking such variability can create difficulties when making inferences from modelling, particularly to inform management, as variability can lead to unexpected outcomes where responses to management actions differ among sites and regions. As such, detailed methods of investigating and quantifying spatio-temporal variability in environmental systems are required. The overarching aim of the thesis is, therefore, to present a number of different methods that can be used to model and examine spatially-varying environmental relationships. The methodologies are applied to examine the spatio-temporal variability in the relationships among surface water systems and a host of environmental, climatological and landscape variables of western Victoria, Australia. The Glenelg-Hopkins region of western Victoria is used as a case study to further model spatio-temporal variability in, and the impact of land-cover transitions on, regional hydrology which have been documented in previous research.

The first method assessed involved the use of the Soil and Water Assessment Tool (SWAT) to model the effect of the introduction of *Eucalyptus* (*Eucalyptus globulus*) plantation forestry in the region. The introduction of *Eucalyptus* coincided with a severe drought and anecdotal evidence suggested that the introduction of *Eucalyptus* into predominantly-agricultural catchments had caused a substantial reduction in streamflows within those catchments. The modelled introduction of *Eucalyptus*
plantation forestry did not significantly change streamflows compared with a scenario which did not include the land-cover change but did include the concomitant drought conditions, suggesting that the modelled realistic rate of land-cover change was not sufficiently extensive to have an impact on streamflows. The SWAT models developed will be invaluable as a basis for future use in regional climate-change studies and for the assessment of land management and land-cover change impact on streamflows.

Inherently linked to the streamflows of the region is the extent of permanent wetlands. The second method explored within this thesis therefore examines the spatio-temporal relationship between climate, land cover and wetland extent. To do this, geographically weighted regression (GWR) was used to examine this relationship in 149 sub-catchments covering the Glenelg-Hopkins region. The application of GWR revealed that the relationships among land cover, rainfall and surface water extent exhibit significant spatial non-stationarity. The models suggested that the amount of surface water area in the landscape was likely related to anthropogenic drainage practices enhancing runoff to facilitate intensive agriculture and increased plantation forestry. However, with some key soil variables not present in our analysis, the strength of this relationship could not be qualified. A temporal assessment of the relationship suggested that the strength of the relationship varied independently of the spatial heterogeneity. The results of this approach support previous research that suggested a much more variable relationship observed in the Glenelg catchment – the western half of the Glenelg-Hopkins region, than in the Hopkins catchment.

Building on the relationships developed using GWR, seven key hydrologic soil variables, which became available after the initial models were developed, were included
to strengthen and re-examine the previous analyses. For two of the three modelled years, soil variability helped to explain the spatial variability in wetland extent. An assessment of model coefficients for the soil variables suggested that there was temporal variation in the ability of soil properties to control surface water extent. Unfortunately, a number of counter-intuitive relationships between soil variables and wetland extents made it difficult to quantify the links between surface water, land cover, soils and climate. Furthermore, a number of contradictory results between the results of GWR models here and those created earlier, potentially due to local correlations among model parameters, complicated model interpretation. Nonetheless, while no clear links between environmental variables and wetland extents were discernible from the modelled relationships, the research has contributed to the growing body of evidence of significant spatio-temporal variability in the Glenelg-Hopkins region, and further highlighted the utility of methods such as GWR in the analysis of spatial relationships.

The final method explored extends beyond the Glenelg-Hopkins case study region and presents a methodology for classifying hydrologic landscapes by employing non-parametric statistics and hybrid image classification. This approach differs from previous classifications which have required the use of an \textit{a priori} spatial unit (e.g. a catchment), necessarily resulting in the loss of variability that is known to exist within those units. A simple statistical approach to identify an appropriate number of classes eliminated the need for the selection and justification of an arbitrary number of groups, or post-hoc testing with different numbers. The use of a hybrid classification employing random forests extended this statistical clustering to an area of approximately 228,000 km$^2$ of south-eastern Australia. This extension resulted in a highly-accurate
regionalisation at both 30-m and 2.5-km resolutions, and a less-accurate 10-km classification that would be more appropriate for use at a continental scale. An assessment of the Glenelg-Hopkins region demonstrated that the method preserved the documented intra- and inter-catchment variability in local hydrology. The methodology presented simplifies current classification frameworks that are becoming more popular in ecohydrology, while better retaining small-scale variability in hydrology, enabling future attempts to explain and visualise broad-scale hydrologic trends at the scale of catchments and continents. An initial assessment of the method concluded that, in the absence of comprehensive stream gauge data which have traditionally been used to define catchments for hydrological classifications, the method is suitable for categorising regions that should behave similarly hydrologically.

The new regionalisation classification framework was then further validated using random forests (RF) and classification and regression trees (CART). The preliminary validation established that there was enough variation in the streamflow indices to suggest that all classification groups were significantly different from one another. However, when using the streamflow indices to predict regionalisation group membership, a constrained discriminant analysis was able to achieve overall classification accuracies of 48%. Classification success using CART and RF was higher than previous estimates from the preliminary validation and ranged from 51 to 62%, further validating the regionalisation framework developed. Also, a local analysis of spatial auto-correlation of model residuals suggested that there was very limited auto-correlation present in the Glenelg-Hopkins region, and indeed over the rest of the 228,000 km² region – suggesting the spatial variability in the regionalisation had been accounted for.
in the classification models. A comparison to a new flow-based classification, demonstrated that regionalisation and flow-based approaches emphasised different aspects of hydrology and that the most suitable approach is dependent on the study region and the objectives of the research. Generally, however, in the absence of comprehensive stream gauge data, regionalisation studies can be used to infer streamflow behaviour, but when possible, flow-based classifications are likely to be more suitable.

Given the variability that has been identified in the Glenelg-Hopkins region, the methods presented in this thesis are likely to be suitable for the quantification of spatio-temporal variability in environmental relationships in other, heavily-modified catchments. Long-term water availability is of particular concern to regional management authorities and, in the face of climate change, further land-cover changes and extended drought conditions, a comprehensive modelling and assessment of regional hydrologic dynamics is critical. The research highlights the effects of land cover and climate controls on streamflows and surface water extents and recommends a cautious, variable approach to management of regional water resources as management actions could have different results in geographically-close regions as a result of small-scale variability. The methods presented in this thesis vary in their assumptions and thus should be applied with careful thought as to when each is most appropriate but, as a suite, should be applicable to a wide range of environmental research questions – particularly when relationships among variables are thought to vary spatially and temporally
# Table of contents

Acknowledgments ..................................................................................................... I
List of peer-reviewed publications ......................................................................... III
Abstract ..................................................................................................................... I
Table of contents .................................................................................................... VI

**General introduction** ......................................................................................... 1
1.1 Impacts of a changing global climate on hydrology ........................................ 1
1.2 Impacts of climate and land cover on water quality ........................................ 6
1.3 The impact of climate and land cover on water quantity .................................. 7
1.4 Catchment processes and heterogeneity of landscapes ..................................... 9
1.5 Hydrological modelling .................................................................................... 11
1.6 Advances in hydrological modelling ................................................................ 14
  1.6.1 From empirical to physical hydrological models .................................... 15
  1.6.2 Statistical and machine learning methods in hydrology ....................... 17
1.7 Regionalisation and hydrological classification approaches to hydrology .......... 20
1.8 Study site description ..................................................................................... 22
  1.8.1 Location, landforms and climate of Victoria ........................................ 22
  1.8.2 Location of the Glenelg-Hopkins region ............................................. 25
  1.8.3 Climate of the Glenelg-Hopkins region .............................................. 27
  1.8.4 Geomorphology in the Glenelg-Hopkins region ................................... 32
  1.8.5 Land cover in the Glenelg-Hopkins region .......................................... 32
  1.8.6 Regional hydrology in the Glenelg-Hopkins region ............................. 35
1.9 Rationale for selecting the Glenelg-Hopkins region as a case study .................. 36
1.10 Objectives of the thesis ............................................................................... 36
1.11 Thesis orientation ......................................................................................... 37

**Assessing the impact of drought and forestry on streamflows in south-eastern Australia using a physically-based hydrological model** ................................................. 41
2.1 Introduction ..................................................................................................... 41
  2.1.1 Land-use and land-cover changes ......................................................... 41
  2.1.2 LULCC impacts on the hydrological cycle .......................................... 42
  2.1.3 SWAT model ..................................................................................... 43
  2.1.4 Aim of study .................................................................................... 46
2.2 Materials and methods .................................................................................. 47
  2.2.1 Australian LULCC ............................................................................ 47
  2.2.2 Australian hydrological conditions ..................................................... 47
  2.2.3 Study site ......................................................................................... 48
  2.2.4 SWAT setup, calibration, sensitivity analysis, validation and LULCC assessment .................................................. 51
2.3 Results ................................................................................................................ 60
  2.3.1 Model calibration, sensitivity analysis and validation ................................. 60
  2.3.2 Monthly, annual and seasonal flows .......................................................... 63
  2.3.3 Flow duration curves ................................................................................ 67
  2.3.4 Mean daily streamflows, assessment of LULCC and monthly ET ................. 70

2.4 Discussion .......................................................................................................... 73
  2.4.1 Applicability of SWAT to the Hopkins and Mount Emu Creek sub-catchments 73
  2.4.2 Monthly, annual, seasonal and drought period flow assessment .................... 76
  2.4.3 Assessment of increased plantation forestry on streamflow ............................ 78

2.5 Conclusions ....................................................................................................... 80

Assessment of spatio-temporally varying relationships between rainfall, land cover and surface water area using geographically weighted regression ....................... 82

3.1 Introduction ....................................................................................................... 82
  3.1.1 Climate change and wetland extents .......................................................... 82
  3.1.2 Factors affecting surface water extents .................................................... 83
  3.1.3 Large-scale data acquisition ...................................................................... 84
  3.1.4 Regression modelling for spatio-temporal assessments of change ................. 84
  3.1.5 Aim of study ............................................................................................ 86

3.2 Data collection and methods .............................................................................. 86
  3.2.1 Study site ................................................................................................. 86
  3.2.2 Site description ........................................................................................ 89
  3.2.3 Sampling sites and land-cover data ........................................................... 91
  3.2.4 Regional climate data ............................................................................... 92
  3.2.5 Modelling methods ................................................................................... 93
  3.2.6 Modelling background .............................................................................. 95
  3.2.7 Comparisons between OLS and GWR model results ................................. 97
  3.2.8 Residual analysis and tests for spatial autocorrelation and variance .............. 98

3.3 Results ................................................................................................................ 98
  3.3.1 Comparisons between OLS and GWR modelling approaches ....................... 98
  3.3.2 Results of GWR modelling ...................................................................... 101

3.4 Discussion ........................................................................................................ 107
  3.4.1 Model performance and interpretation .................................................... 107
  3.4.2 Regional changes in land cover and drainage ............................................ 111

3.5 Conclusions ..................................................................................................... 113

Getting down and dirty – can soil attributes help to quantify a spatially-varying relationship between rainfall, land cover and wetland extents? ............................ 115

4.1 Introduction ....................................................................................................... 115

4.2 Methods .......................................................................................................... 116
  4.2.1 Model design ......................................................................................... 116
4.2.2 Model comparisons ................................................................. 117
4.3 Results ....................................................................................... 118
4.4 Discussion ............................................................................... 122
4.5 Conclusions ............................................................................ 124

Hydrologic landscape regionalisation using deductive classification and random forests ............................................. 126

5.1 Introduction ................................................................................ 126
5.1.1 Flow variability and ecological controls .................................. 126
5.1.2 Landscape and hydrologic units .............................................. 127
5.1.3 Deductive and inductive landscape classification .................... 128
5.1.4 Statistical clustering and multivariate analyses ....................... 131
5.1.5 Supervised classification of landscapes ................................. 132
5.1.6 Aim of study ......................................................................... 133

5.2 Materials and methods ............................................................... 134
5.2.1 Site description ..................................................................... 136
5.2.2 Site description of case-study area in western Victoria .......... 138
5.2.3 Variable selection and processing ......................................... 138
5.2.4 Development of classification groups ................................. 140
5.2.5 Hybrid classification with random forests ......................... 146
5.2.6 Accuracy assessment ............................................................. 147
5.2.7 Relationships between the regionalisation and hydrologic indices ... 150

5.3 Results ................................................................. 154
5.3.1 Spatial distribution of sample points ..................................... 154
5.3.2 Clustering and ordination ....................................................... 155
5.3.3 Random forests ................................................................. 160
5.3.4 Comparisons between original, PCA and resampled classifications .... 164
5.3.5 Spatial variability in western Victoria ................................. 169
5.3.6 Relationships between the regionalisation and hydrologic indices .... 173

5.4 Discussion ................................................................................ 174
5.4.1 Differences to previous regionalisation studies ..................... 175
5.4.2 Statistical evaluation, clustering, and PCA ......................... 176
5.4.3 Classification by random forests ......................................... 178
5.4.4 Case study on spatial variability in western Victoria ............ 180
5.4.5 Relationships between the regionalisation and hydrologic indices .... 181

5.5 Conclusions ............................................................................ 182
Classifying stream variability in ungauged basins from a pixel-based hydrological regionalisation ................................................................. 184

6.1 Introduction .............................................................................................................. 184
  6.1.1 Prediction in ungauged basins ......................................................................... 184
  6.1.2 Regression and machine-learning methods for ungauged flow prediction .... 185
  6.1.3 Identifying associations between catchment form and function .................. 187
  6.1.4 Aim of study ................................................................................................. 188

6.2 Methods ..................................................................................................................... 189
  6.2.1 Site description ................................................................................................. 189
  6.2.2 Development of the hydrological regionalisation ............................................ 191
  6.2.3 Streamflow indices ......................................................................................... 191
  6.2.4 Flow classification ......................................................................................... 192
  6.2.5 Variable contributions to the classifications ................................................. 193
  6.2.6 Modelling approaches ................................................................................... 193
  6.2.7 Model tuning and evaluation ......................................................................... 196
  6.2.8 Spatial autocorrelation analysis .................................................................. 198

6.3 Results ....................................................................................................................... 199
  6.3.1 Flow classification ......................................................................................... 199
  6.3.2 Variable contributions to the classifications ................................................. 202
  6.3.3 Modelling approaches ................................................................................... 207
  6.3.4 Comparisons between the regionalisation and flow classifications ............ 219
  6.3.5 Spatial autocorrelation analysis .................................................................. 222

6.4 Discussion ................................................................................................................. 222
  6.4.1 Ability of streamflow indices to predict hydrologic regionalisation groups ... 223
  6.4.2 Ability of environmental variables to predict flow classification groups ...... 225
  6.4.3 Flow indices and environmental variables controlling the classifications .... 229
  6.4.4 Potential methods to improve the ability of the regionalisation to predict flow regimes ................................................................. 231

6.5 Conclusions .............................................................................................................. 235

6.6 General discussion .................................................................................................... 237

7.1 Research highlights ................................................................................................. 238

7.2 A conceptual framework for investigating spatio-temporal hydrologic variation ......................................................................................... 243

7.3 Methodological implications for hydrological and environmental science .... 247

7.4 Uncovering spatial variability in the Glenelg-Hopkins catchment .................... 254

7.5 Future work ............................................................................................................. 259

7.6 Conclusions .............................................................................................................. 261

References ..................................................................................................................... 263
### Appendix I

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.1</td>
<td>Supporting tables</td>
<td>286</td>
</tr>
<tr>
<td>9.2</td>
<td>R code for permutation test</td>
<td>293</td>
</tr>
<tr>
<td>9.3</td>
<td>Supporting figures</td>
<td>295</td>
</tr>
</tbody>
</table>

### Appendix II

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.1</td>
<td>Supporting tables</td>
<td>305</td>
</tr>
<tr>
<td>10.2</td>
<td>Supporting figures</td>
<td>308</td>
</tr>
</tbody>
</table>
**General introduction**

Climate change is widely accepted by the scientific community, with current research suggesting that temperature and rainfall patterns are likely to change dramatically over the next 50 years (IPCC, 2007b, 2014). Extended periods of drought and associated reductions in precipitation, runoff and soil moisture are forecast, and a range of hydrological and ecohydrological models suggest that groundwater and surface water systems will be under increased stress (Lake, 2003). While landscape characteristics affecting the quantity, quality and movement of water are extremely complex (Winter, 2001), landscapes and catchments that appear unique and diverse often have common attributes governing water movement. Thus, there are generally patterns across spatial and temporal scales which could be exploited to improve hydrologic predictability when, and if, those patterns are appropriately identified and explained (Troch et al., 2009). In an attempt to explore, explain and predict these patterns among catchments, regions and continents, research is now shifting away from simple regression and statistical approaches towards physically-based streamflow models, more advanced statistical techniques, data-mining methods and hydrologic regionalisation studies.

1.1 Impacts of a changing global climate on hydrology

Evidence that the Earth’s climate continues to warm is unequivocal, with the majority of evidence indicating that anthropogenic greenhouse gas emissions are the dominant cause of recent warming (IPCC, 2007b, 2014). While climate change is typically associated with increasing temperatures, this is not the only climatic variable affected. Rainfall, evapotranspiration (ET) and/or wind-speed are also predicted to
change in some regions and may prove to be more influential in impacting local climate than changes in temperature (Hulme, 2005).

Climate scenarios predict an increase in rainfall in the tropics (i.e. increases in monsoon rainfall) and at high latitudes, with general decreases in rainfall predicted in the sub-tropics. Global mean annual rainfall shows a slightly increasing, albeit very variable, trend (Figure 1.1), and climate projections currently include increases in globally-averaged mean water vapour, evaporation and rainfall (IPCC, 2008, 2014). The increase in global mean annual temperatures is clear (Figure 1.2), having risen by 0.85 ± 0.2 °C between 1880 and 2012, with the 20 warmest years on record having occurred after 1990 (Figure 1.2; Bureau of Meteorology, 2014). The Fifth Assessment Report by IPCC (2014, p. 10) suggests that it is “virtually certain” that there will be more frequent hot and fewer cold temperature extremes over most land areas on daily and seasonal timescales as global mean surface temperatures increase. The IPCC (2014, p. 10) also report that it is “very likely” that heat waves will occur with a higher frequency and longer duration.

Increases in the frequency and/or intensity of ecosystem disturbances such as droughts have been observed in different parts of the world and, in some cases, can be attributed to changes in climate (IPCC, 2014). It is also likely that rainfall variability associated with the El Niño Southern Oscillation (ENSO) will intensify on regional scales as a result of climate change (IPCC, 2014). While natural droughts are not a well-understood phenomenon, it is theorised that ENSO is the major climatic phenomenon responsible for weather extremes, including droughts, in Australia (Cleugh et al., 2011).
Figure 1.1: Global annual rainfall anomalies (mm) between 1900 and 2014. The black trend line is the 10-year moving average. Global annual rainfall is highly variable and, while a small increase in annual rainfall is shown by the trend line, inter-annual variability is also large. Data extracted from bom.gov.au/climate/change/.
Figure 1.2: Global annual mean temperature anomalies (°C) between 1850 and 2014. The black trend line is the 10-year moving average. A significant increase in global mean temperatures can be observed from the 1980s, while the 20 hottest years on record can be seen after 1990. Data extracted from bom.gov.au/climate/change/.
In addition, there are complex feedback loops between regional climatic conditions and hydrology (IPCC, 2008) and other influences, both naturally occurring and anthropogenic, that can exacerbate the effects of drought. Surface water balances broadly reflect the availability of both water and energy in the hydrologic cycle. In regions where plant available water is high, ET is controlled by prevailing meteorological conditions and land cover (Zhang et al., 2001). Changes in the surface water balance, for example as a result of afforestation or deforestation, can feed back into the climate system by recycling water into the atmosphere (instead of allowing it to become surface runoff or to penetrate into deeper soil layers) (Seneviratne et al., 2010). Declines in soil moisture have been linked to vegetation die-off, reducing surface shading which further increased soil moisture deficiencies – effectively prohibiting the re-establishment of tree cover which had been shown to maintain a higher soil water content than adjacent bare/scrub patches (D’Odorico et al., 2007). These changes can then feed back into the hydrologic cycle as the triggering of rainfall may itself be the result of enhanced atmospheric instability induced by wet soil conditions (Seneviratne et al., 2010). It has also been shown that changes in regional climates (at high latitudes) can be influenced by land-cover changes in distant (tropical) systems as a result of changes in atmospheric convection (Chase et al., 2000). Depending on the local environment, however, the sign and magnitude of these effects are highly variable. While often these feedback loops may be relatively small at global scales, they can become exceptionally important at smaller spatial and temporal scales – a process which can lead to regional and/or local changes in climatic variability and extremes (IPCC, 2008).
The ‘snow-ball’ effect of negative feedback loops has already been identified as contributing to changes in water balances (D’Odorico et al., 2007; Seneviratne et al., 2010), and climate change is expected to exacerbate these disturbances by affecting global hydrology (IPCC, 2008, 2014). As temperatures rise and rainfall variability increases, combined with an expected increase in ENSO-related weather extremes (IPCC, 2014), there are likely to be significant changes in the spatial and temporal dynamics of hydrologic systems. Through the mutual effects of an increase in global water use (as human populations increase), the cyclic nature of droughts and an increase in climatic variability, it is now becoming generally accepted that climate changes further threaten the quality and availability of freshwater.

1.2 Impacts of climate and land cover on water quality

Approximately one in six people lack adequate access to clean drinking water, and more than double that number lack water needed for basic sanitation (Riley et al., 2011). Depending on the intended use, there are various different requirements for water quality to ensure that there is minimal risk to end users and to minimise environmental damage (ANZECC and ARMCANZ, 2000). For example, human infection is possible following irrigation of crops with polluted water (Hamilton et al., 2006), while land-based disposal of sewage effluent as irrigation has been shown to retard plant growth due to the presence of heavy metals (Smith et al., 1996). It is obvious that climate impacts, such as droughts, may lead to a reduction in quality of surface or groundwater – for example, by increasing the concentration of solutes such as salts and nutrients. In spite of this, until recently, there has been limited peer-reviewed research on the impacts of climate change on water quality (Delpla et al., 2009). Water quality parameters that are
susceptible to climate changes, (e.g. as a result of increasing air, soil and water
temperatures and an increase in heavy rainfalls in temperate regions) include: dissolved
organic matter (Evans et al., 2005); organic and inorganic micro-pollutants (e.g.
pesticides and heavy metals, Bloomfield et al., 2006; van Vliet and Zwolsman, 2008);
pathogens (e.g. coliform bacteria such as Escherichia coli, Pednekar et al., 2005); and
cyanobacteria and associated cyanotoxins (Jöhnk et al., 2008).

The extent of wetlands and riparian vegetation is known to influence the water
quality in streams, with wetlands being used globally to reduce nutrient concentrations
in aquatic systems. Measurements from different regions around the globe suggest that,
when wetlands comprise 2–7 % of the land cover within a catchment, significant
improvements in water quality can be observed (Verhoeven et al., 2006). There is also
ample evidence linking the amount of vegetation within catchments to water quality
(Vinten and Dunn, 2001; Tong and Chen, 2002; Versace et al., 2008b; Tu, 2009).
While wetlands are typically used for water treatment and to reduce nutrient loads in
terrestrial and aquatic ecosystems, there is also evidence that the resulting nutrient
enrichment of wetlands can often have significant effects on primary productivity,
nutrient cycling and leaching, and shifts in the species composition within those
wetlands – particularly in agricultural catchments (Verhoeven et al., 2006).

1.3 The impact of climate and land cover on water quantity

It has recently been shown that water resources within developed nations are
more sensitive to climatic conditions, in contrast to resources in developing countries
where stress is far more likely to be a result of socioeconomic growth (Schlosser et al.,
2014). With the global population rising and an increase in severe drought conditions
predicted under a number of climate change scenarios (IPCC, 2008, 2014), it has been estimated that an additional 1.3 billion of the world’s projected 2050 population will be living in regions where water demand will consistently exceed surface water supplies (Schlosser et al., 2014). Floods, droughts and prevailing meteorological conditions are the leading drivers of water availability, particularly of surface flow to streams (Li et al., 2009; Teng et al., 2012). Climate change is expected to increase inter-annual hydrologic variability, with one study from Britain, for example, indicating a minor increase in mean monthly flow but a significant reduction (40 %) in low-flow volumes by 2080 (Arnell, 2003). As a result of increased evaporation and reduced rainfall, combined with an increased demand for water, Australian water availability is expected to decline with median runoff reductions of 9 % in the north of the Murray-Darling Basin and 13 % reductions in the south by 2030 (Cleugh et al., 2011).

Spatially, land-cover impacts on hydrology are particularly obvious at small scales such as hillslopes or fields (Tollan, 2002). However, it has been shown that alterations in land cover can drastically affect runoff rates in small (<10 km²) (Scott and Lesch, 1997; Rodriguez Suarez et al., 2014) through to macro-scale (>10,000 km²) catchments (Costa et al., 2003). The alteration in surface water availability has been linked primarily to differences in ET as a result of processes such as afforestation (Zhang et al., 2001; Peel et al., 2010). The compensatory effects of the hydrological cycle at large scales typically result in few changes being observed at the catchment level (Fohrer et al., 2001; Peel et al., 2010) and, indeed, the effect of vegetation on ET has been shown to diminish as the area of a catchment increases (Peel et al., 2010). Brown et al. (2005)
indicated that for any impact of vegetation change to be detected in streamflows, at least 20% of the catchment needed to have undergone a transition in land cover.

Global wetland extents have also decreased to approximately half of their original extents (OECD, 1996). The drainage of wetlands to facilitate intensive agriculture is likely to be the largest threat to wetland extents and, in 2005, wetlands were estimated to cover an area of between 5.3 and 12.8 million km² globally (Zedler and Kercher, 2005). While there are differences in the spatial and temporal resolution of many wetland inventories and the spatial extent of wetland coverage varies greatly among countries, the overall trend indicates an indisputable reduction in surface area, condition and associated biodiversity of global wetlands (Brinson and Malvarez, 2002; Zedler and Kercher, 2005; Verhoeven et al., 2006).

Despite the implicit links between climate, land cover and catchment processes, many regions that have undergone extensive land-cover change do not have reliable information on the relationships between land cover and water quantity. These changes are in addition to changes associated with impoundment and diversion of water for agricultural and urban uses.

1.4 Catchment processes and heterogeneity of landscapes

Catchments are topographically-defined units that provide spatial and temporal continuity of hydrological processes across landscapes and regions. The role of catchments in controlling fluxes of energy, water and carbon (which are driven by temperature, rainfall and chemical gradients) has long been recognised by hydrologists, ecologists and soil scientists as being moderated by catchment features such as vegetation, soil structure and geology (Troch et al., 2009). Traditionally, hydrological
science took the view that catchment behaviour could be inferred from processes occurring on a much smaller scale (e.g. hillslope), and it has long been assumed that landscapes that appear unique and diverse often have common sets of attributes governing the movement of water (Beven et al., 1988; Winter, 2001).

There has, however, been a recent paradigm shift away from this assumption to a more ‘holistic’ view of catchment hydrology which explicitly accounts for landscape heterogeneity (Troch et al., 2009) and the general consensus among hydrologists is that the relative contribution of small-scale processes decreases with increasing catchment size (Li and Sivapalan, 2011; Hrachowitz et al., 2013). The inherent heterogeneity of catchments results in significant spatio-temporal variability in hydrologic states, particularly for scale-dependent flow and transport of nutrients and energy (Olden and Poff, 2003; Sawicz et al., 2011), which then leads to an incomplete understanding of hydrological processes if that heterogeneity is not considered (Troch et al., 2009; Hrachowitz et al., 2013).

An important aspect of hydrological modelling, and in particular eco-hydrological modelling, is being able to identify, conserve and quantify spatio-temporal variability in the relationships being explored. For example, the natural flow paradigm (Poff et al., 1997) emphasises that the form and function of aquatic ecosystems (e.g. rivers, wetlands) can be maintained, provided that a regulated flow regime mimics the natural flow regime. However, the flow regime of a given system is representative of long-term flow behaviour and inherently incorporates influences that operate over a large range of temporal and spatial scales (Arthington and Pusey, 2003; Thoms and Parsons, 2003). The importance of considering spatial and temporal distributions of
rainfall in hydrological modelling is well known (Tetzlaff and Uhlenbrook, 2005) and, although it has been demonstrated that temporal variability is just as important as spatial variability in hydrological modelling (Bürger and Chen, 2005), there is a disparity between the perceived importance of variability and representation in current empirical models (Bürger and Chen, 2005). Hydrological variation, both spatial and temporal, can lead to potentially complex, non-linear patterns of hydrological characteristics. Furthermore, this variation can affect ecosystem services which are often provided by complex interactions between terrestrial and aquatic ecosystems (D’Odorico et al., 2010). However, many environmental flow and hydrological assessments, which can be used to inform management for flow regulation or aquatic conservation purposes (Poff et al., 2010) fail to recognise, preserve and quantify the spatio-temporal dynamics of a given hydrological system (Thoms and Parsons, 2003). Provided that the complexities and variability of a river or wetland system can be identified and modelled adequately, environmental flow targets or aquatic conservation efforts based on methods that incorporate spatial and temporal variability may be more practical because they should more accurately reflect the natural flow regime and are less likely to result in unintended consequences as a result of system heterogeneity. Catchment hydrology is now moving towards being an interdisciplinary science in an attempt to develop new hydrological theories and models that can account for such heterogeneity (Troch et al., 2009; Hrachowitz et al., 2013).

1.5 Hydrological modelling

All hydrological models are simplified representations of real world processes. There are three commonly-used frameworks for investigative hydrological modelling:
experimental paired-catchment studies; statistical time-series analysis; and computer-based hydrological modelling. Each of these methods provides its own unique framework for the conceptualisation and investigation of hydrological relationships. For example, paired-catchment studies are typically employed when determining the magnitude of water yield changes resulting from alterations in land-cover or management practices (Brown et al., 2005). Time-series analysis of streamflow records can be used for determining flow probabilities or the recurrence intervals of floods or low flows, while more advanced time-series analysis can be used for examining water fluxes at the river–groundwater interface (Keery et al., 2007). Finally, computer-based hydrological models are useful for modelling interactions between soil, water, climate and land-cover, and the associated hydrologic processes such as precipitation, evapotranspiration, infiltration, surface runoff and groundwater flow (Arnold et al., 1998; Gassman et al., 2007; Todini, 2007).

Historically, paired-catchment approaches have largely been applied at small scales (~10s km²) and have been used to estimate effects of land-cover on hydrology (Brown et al., 2005). Recently, however, there has been a shift in water resources management and hydrological research to meso- and macro-scales as these scales are more appropriate to the provision of information for economical, societal and environmental management and planning (Lahmer et al., 2001). Assessments at larger scales (e.g., Costa et al., 2003; Xu and Singh, 2004), either using time-series analyses or computer-based hydrological modelling, have been made possible through the rapid development of geographic information system (GIS) technologies and the integration of physically-based, (semi-)distributed hydrological models into GIS suites. Typically,
the required inputs for these models (e.g. rainfall, land cover, soil properties and topography) are available as spatial databases (either raster- or vector-based) which streamlines their input as model parameters (Xu and Singh, 2004; Gassman et al., 2007) thereby simplifying the modelling process.

While there are many hydrological models available, they differ in spatial and temporal detail (of both input and output), but all provide users with the ability to generate predictions of catchment behaviour such as streamflow and sediment and nutrient exports (Martina and Todini, 2009; Moradkhani and Sorooshian, 2009). The quality of the model output however, is dependent on several factors. The first consideration is data availability. Model outputs will only be as good as the data used during calibration, since uncertainty in the input data is frequently amplified in the output (Todini, 2007). Uncertainty in predictions can also be influenced by calibration accuracy, parameter sensitivity (Moriasi et al., 2007), and parameter uncertainty (Abbaspour et al., 2007; Todini, 2007). The choice of model complexity also needs to be tailored to data availability. Traditionally, the data needed to be able to reliably design and apply mechanistic or physically-based hydrological models were expensive and time-consuming to collect over regional catchment scales. However, the development of high-resolution international and continental datasets which are largely cost-free means that data limitations are becoming less frequent, allowing an increased understanding of the complex nature of the hydrological cycle at the catchment scale. Examples of such datasets include topography (ASTER GDEM Validation Team, 2011), land cover (Global Land Cover 2000 database, 2003), soil (FAO, 2007), climatic attributes (Fuka et al., 2014) and satellite-based remotely-sensed data (e.g. LANDSAT, MODIS). Second,
there is the question of whether the processes being modelled are actually occurring in situ. That is, is the conceptual basis underpinning the model relevant to the catchment of interest? For example, is the application of a model designed to simulate flows in snow-melt dominated streams, appropriate for use in semi-arid regions? As such, careful consideration needs to be taken as to the choice, design and application of hydrological models to a catchment of interest (Abbott and Refsgaard, 1996; Martina and Todini, 2009; Moradkhani and Sorooshian, 2009).

1.6 Advances in hydrological modelling

Deterministic (as opposed to stochastic, which are not covered here) hydrological models can be classified as either lumped or distributed depending on the spatial scheme used during the modelling process. Given the same set of hydrological inputs (e.g. rainfall), deterministic models will always produce the same output values (e.g. streamflow); stochastic models, by definition will not as they include elements of randomness. Spatial variability is ignored in lumped models, while the opposite is true for distributed models; that is, the spatial variability in parameters such as topography, vegetation, soil and climate is taken into consideration. The spatial aggregation for distributed models can take on one of three broad discretisation forms (i.e. the “grid-cell”): Orthogonal grid-based, irregular grid-based and hydrological response unit (HRU)-based (Kite and Pietroniro, 1996).

Orthogonal grid-based discretisation involves dividing the catchment into a series of rectangular grids. An example of the orthogonal grid-based system that is used by a number of hydrologic models (e.g. SHE; Abbott et al., 1986a, 1986b) is the Representative Elemental Areas (REA) discretisation scheme. Here, areal units are
identified within a catchment where the hydrological properties are definable and would not be significantly different if a smaller scale of discretisation were used. The REA approach influences the hydrological response depending on the elements’ location within the catchment (Kite and Pietroniro, 1996). Irregular grid-based discretisation divides a catchment into irregular elements depending on catchment topography and terrain features. Hydrologic modelling elements are generally defined by triangulated irregular networks (TIN), but can contain information on streamlines or catchment boundaries (Kite and Pietroniro, 1996; Vivoni et al., 2004). The primary benefits of the irregular grid-based method is the variable resolution offered by the TIN – essentially regions of high variability can be modelled more precisely than regions of low variability (Vivoni et al., 2004). Lastly, the HRU-based discretisation first divides the catchment into a number of sub-catchments, and then into smaller areal units where each unit consists of a hydrologically-homogeneous set of landscape characteristics such as land cover, slope and soil attributes. HRUs each generate a distinct hydrological response, but unlike the REA approach to discretisation, their location is generally only important for routing water through each component of the model (Kite and Pietroniro, 1996). This method of spatial discretisation is very popular for physically-based models such as SWAT (Arnold et al., 1998).

1.6.1 From empirical to physical hydrological models

Models can further be classified as empirical, conceptual or physically-based (Moradkhani and Sorooshian, 2009). Empirical models typical involve mathematical equations that have been determined by analysis of input and output time series data and not from physical processes occurring in the catchment (Abbott and Refsgaard,
Examples of empirical models include the Rational Method (Mulvaney, 1850) which relates the time of concentration of small catchments to peak runoff rates, and the unit hydrograph (Sherman, 1932) which provides an estimate of streamflow given a specific rainfall amount for a given catchment.

Conceptual models typically involve the configuration of “stores” of water and use mathematical functions to transfer the water between the stores or into the stream. Conceptual runoff models are frequently used to compensate for a lack of measurements, for example to extend time series flow data, or to predict nutrient fluxes or the effects of climate and land cover changes. As a consequence, calibration of the model against the variable of interest (e.g. flow, nutrient concentrations) is required (Ye et al., 1997). Popular conceptual models include TOPMODEL (Beven and Kirkby, 1979) which is a (semi-)distributed rainfall-runoff model that utilises topographic information (i.e. specific catchment area and wetness index) to generate runoff estimates. Beven et al. (1995) consider TOPMODEL to be a set of conceptual tools that can be used to simulate the dynamics of surface or subsurface contributing areas to hydrological processes rather than a generally applicable single model structure.

Likewise, the U.S. National Weather Service River Forecast System is a model based on a generalised hydrologic model (Burnash et al., 1973) that is designed to conceptually model hydrologic processes of catchment headwaters.

All physically-based distributed hydrologic models tend to represent real world processes, although different models are likely to have different assumptions and structures (Martina and Todini, 2009). The primary components of the land phase of the hydrologic cycle are typically taken into consideration with physically-based models.
These include interception, snowmelt, evapotranspiration, surface and subsurface runoff, groundwater flow and channel routing (Abbott and Refsgaard, 1996; Arnold et al., 1998). In physically-based hydrologic models, hydrological processes are modelled by partial differential equations (such as Richards’ equation for unsaturated flow, or Boussinesq’s equation for groundwater flow), or alternatively by empirical equations (Abbott et al., 1986b). The governing equations are usually ‘point-scale’, while the models themselves typically refer to a finite scale such as the ‘grid cell’. How the point-scale equations are integrated into the grid cell strongly affects the sensitivity of the model to different scales of space and time and, more specifically, the model’s capability to preserve physical representativeness at different scales (Martina and Todini, 2009).

### 1.6.2 Statistical and machine learning methods in hydrology

All statistical techniques are subject to sets of assumptions that are typically violated in hydrology (Holder, 1985). For example, the presence of (multi-)collinearity among predictor variables, heteroscedasticity between predictor and response variables, or the existence of a non-random distribution of model residuals can invalidate simple regression methods such as OLS (Quinn and Keough, 2002). Typically, violations of statistical assumptions will result in model misspecification which can lead to illogical model coefficients (Wheeler, 2009), spatial-autocorrelation of residuals (Zhang et al., 2005) or non-stationarity between response and predictor variables (Fotheringham et al., 2002). Nevertheless, when model assumptions are correctly satisfied, problems facing hydrologists that can typically be addressed using simple regression and statistical methods are: 1) the development of streamflow estimates for sites where no records exist; and 2) the extension of short-term records. However, to estimate these quantities
requires data from nearby hydrologically-homogenous catchments (Gordon et al., 2004). Recently, simple and multiple linear regressions have been used to model water quality in streams and lakes (Versace et al., 2008b; Sheela et al., 2011), relate baseflow to catchment properties (Mazvimavi et al., 2005), explore relationships between topography and precipitation (Um et al., 2011) and, finally, make predictions of the likely impact of climate change on aquatic ecosystems across large spatial and temporal scales (Lester et al., 2014). The relative simplicity of regression approaches has been shown to be of value when examining first-order (e.g. topographical) control mechanisms on catchment function in data-scarce regions (Hrachowitz et al., 2013).

Recent improvements and developments in regression models have had some applications in hydrology. Quantile regressions have been used in regional flood-frequency analysis (Haddad and Rahman, 2012) and in measuring the predictive uncertainty of rainfall-runoff forecasts (Weerts et al., 2011). Geographically weighed regression (GWR) – an improvement on least-squares regression that accounts for spatial variation in model parameters – has also been utilised recently (Fotheringham et al., 2002). The method has been used to model groundwater and surface water quality as a function of differences in land cover (Tu and Xia, 2008; Tu, 2013; Javi et al., 2014); while Chang and Psaris (2013) found that maximum stream temperature in the Columbia River catchment, USA, could be predicted using GWR based on baseflow index, percent forest cover, and stream order. GWR has been shown to outperform a number of other modelling methods, specifically with regard to spatial autocorrelation of model residuals (Zhang et al., 2005), and recent improvements to the method can also account for localised collinearity among variables (which may not be readily
obvious) that can influence model interpretation (Wheeler, 2009). It has also been shown that, in some instances, multiple regression analyses can outperform more advanced data mining/machine learning techniques, but only where the relationship between the variables was linear (Mazvimavi et al., 2005).

Conceptual or physically-based models are of importance in the understanding of hydrological processes. It is often desirable however, to implement simple “black-box” models that can provide detailed outputs (e.g. streamflow) without direct consideration of the physical processes between the predictor and response variables (Dibike and Solomatine, 2001). While simple relationships and multiple regressions have proven useful for hydrological modelling, recent work has seen a shift away from classic statistical techniques to more advanced machine-learning methods such as neural networks (Dibike and Solomatine, 2001; Mazvimavi et al., 2005; Bhattacharya et al., 2007; Hong, 2008), classification and regression trees (CART) (Bhattacharya et al., 2007; Kennard et al., 2010b; Poor and Ullman, 2010; McManamay et al., 2012), random forests (RF) (Carlisle et al., 2010; Rodriguez-Galiano et al., 2014) and support vector machines (Hong, 2008; Rasouli et al., 2012; Raghavendra and Deka, 2014). Machine-learning techniques provide flexible mathematical structures and frameworks which are capable of identifying complex non-linear relationships between the input and output data without attempting to understand the nature of the phenomena. These properties make machine-learning methods more appropriate than classical statistical models when the emphasis is on prediction rather than explanation. This is particularly true when the predictor variables are correlated and there are many complex, possibly non-linear,
interactions which typically violate the assumptions of methods such as multiple regression (Holder, 1985; Dibike and Solomatine, 2001).

1.7 Regionalisation and hydrological classification approaches to hydrology

Streamflow prediction in ungauged basins (PUB) is a significant challenge for hydrologists. Beginning in 2003, the International Association of Hydrological Sciences began a shift in hydrology from a focus on model form towards a focus on improved scientific understanding of hydrological processes. This initiative was concluded in 2012 and became known as the PUB decade (Hrachowitz et al., 2013). The purpose of the initiative was to facilitate the inference of hydrological functions and controls in data-scarce regions from metrics related to catchment form – for example, the combined effects of climate, topography, geology, soil type and land cover – independent of existing streamflow data. However, catchment heterogeneity has hindered attempts at classification and has made it difficult to understand and explain the spatio-temporal variation and control of hydrological processes at the catchment scale (Hrachowitz et al., 2013). Wagener et al. (2007) suggested that if a framework could be developed based on the concept of catchment function, including an explicit link between climate, landscape attributes and streamflow indices and the associated uncertainty and heterogeneity at multiple temporal and spatial scales, then it could serve as a potentially predictive model of catchment function in ungauged basins.

While hydrologic classification can refer to an assortment of methods, a review by Olden et al. (2012) recognised two broad approaches to hydrologic classification; deductive and inductive approaches. The inductive approach uses the emergent
properties of discharge time series data to generate classes – i.e. flow-based classification.

In contrast, the deductive approach to classification is used when attempting to describe broad spatial patterns in flow regime variability where there is a lack of gauged or modelled streamflow data available – i.e. hydrological regionalisation. Deductive methods of environmental classification are commonly used when the objective is to quantify and describe spatial variation in flow regime attributes. This approach to classification identifies groups on the basis of physical and climatic attributes that, over broad scales, produce similar hydrologic responses in stream systems (Olden et al., 2012).

The increased availability of high-quality, hydrologically-relevant spatial datasets (e.g. climate, topography, land cover) makes deductive reasoning an appealing method when attempting to define spatial similarities or dissimilarities in hydrological characteristics (Olden et al., 2012).

Wolock et al. (2004) used the concept of hydrologic landscapes introduced by Winter (2001) to create a regionalisation of nearly 44,000 catchments (~200 km² in area each) using a combination of multivariate ordination and cluster analyses. Kennard et al. (2010b) presented a method that utilised statistical clustering of a range of environmental, geological, topographical and meteorological data to create a hydrological regionalisation. They then attempted to relate their regionalisation to regional streamflow indices with mixed success. Sawicz et al. (2011) employed the use of precipitation-temperature-streamflow signatures and Bayesian clustering to characterise 280 non-contiguous catchments located in eastern USA so as to understand similarities in climatic and landscape attributes across the region. They found that signatures which varied along climatic gradients exerted a stronger influence on cluster separation than
those signatures which varied as a result of geology or land cover. McManamay et al. (2012) found that a hydrological regionalisation developed by Wolock et al. (2004) explained only 7% to 39% of the variation in a number of flow indices, but that a flow-based classification approach was able to explain 9% to 87% of the variation in the same indices. However, it has been demonstrated that a hydrologic regionalisation, when combined with stream classification, can help in the prediction of streamflow metrics (Snelder et al., 2005; Santhi et al., 2008). Hydrologic regionalisations have also been shown to improve predictive streamflow models when those models are stratified by hydrologic regions (Carlisle et al., 2010). The creation of these regionalisations allows inferences to be made regarding hydrologic processes even in the absence of rigorous mechanistic models or detailed hydrologic data (Hrachowitz et al., 2013).

1.8 Study site description

The area used as a case study for the research presented in this thesis is the Glenelg-Hopkins region of western Victoria, Australia. Given below is a brief description of the state of Victoria, followed by a more in depth description of the Glenelg-Hopkins region. Due to a number of chapters from this thesis having already been published, there is overlap between this section (Section 1.8) and the research chapters.

1.8.1 Location, landforms and climate of Victoria

Victoria comprises an area of 227,594 km² and is the southernmost state of mainland Australia. Victoria contains a diverse range of topographic, geologic and climatic conditions, with relatively flat, semi-arid plains in the west and northwest, to alpine areas in the east and northeast (Figure 1.3). Rainfall across the state is variable,
with median annual rainfall ranging from more than 2,500 mm in mountainous regions
to less than 300 mm in the west and northwest (Department of Environment and
Primary Industries, 2013a). The majority of the state experiences a Mediterranean
climate consisting of hot, dry summers and cool, wet winters, with some snowfall
occurring in the alpine regions (Department of Environment and Primary Industries,
2013b). Victoria has an extensive river and wetland network, with the largest river being
the Murray River system in the north, and nearly 17,000 wetlands larger than 0.01 km²
in surface area (Corrick, 1992).
Figure 1.3: Location of the state of Victoria in south-eastern Australia. Dark blue lines represent perennial rivers, while the colour gradient represents elevations, with darker browns indicating higher elevations. ‘m.a.s.l’ is metres above mean sea level.
1.8.2 Location of the Glenelg-Hopkins region

The Glenelg-Hopkins region of western Victoria covers approximately 27,000 km², and supports a permanent population of 130,000 with population densities highest around Ballarat, Warrnambool, Hamilton, Port Fairy, Portland and Ararat (Glenelg Hopkins CMA, 2013). The region is situated south-west of the Great Dividing Range and is comprised of three major catchments: Glenelg; Hopkins; and Portland Coast. The region extends from close to Ballarat in the Victorian Central Highlands, west to the South Australian border, and south towards the coast (Figure 1.4).
Figure 1.4: Location of the Glenelg-Hopkins region in western Victoria. Dark blue lines represent perennial rivers, dark blue areas represent wetland habitat, while the colour gradient represents elevations, with darker browns indicating higher elevations. The three major drainage basins are outlined in red (Glenelg), black (Hopkins) and orange (Portland Coast). ‘m.a.s.l.’ is metres above mean sea level.
1.8.3 Climate of the Glenelg-Hopkins region

Like much of the rest of Victoria, the Glenelg-Hopkins catchment experiences a Mediterranean climate, with hot, dry summers and cool, wet winters. Mean maximum summer temperatures range from 22 to 24 °C in coastal and elevated areas but increase further inland to 25 to 27 °C. In winter, mean maximum temperatures range from 12 to 14 °C with frosts commonly occurring inland (Department of Sustainability and Environment, 2008a). The mean annual rainfall in the area ranges from 500 mm at Lake Bolac in the north east of the region to 900 mm near Heywood and west of the Grampians Ranges in the Cobboobonee Forest (Department of Sustainability and Environment, 2008a). Rainfall is consistent across the region, with at least 1 mm of rainfall occurring on 129 days each year, on average (Department of Sustainability and Environment, 2008a). The region has suffered prolonged periods of below average rainfall in the past (Raleigh and Dixon, 2005) and, between 1998 and 2007, the region’s average annual rainfall was 10 % below the 1961 to 1990 average (Department of Sustainability and Environment, 2008a). This cycle of prolonged drought has been evident across much of Victoria since 1996 (Turner et al., 2004) and is reflected in long-term hydrographs (Figures 1.5 – 1.10) which show significant declines in discharge between 1997 and 2010. Annual stream flows fell by 40, 56 and 65 % respectively in the Hopkins, Portland Coast and Glenelg basins (Glenelg Hopkins CMA, 2013) during the so-called ‘Millennium Drought’ of 1997 to 2009 (CSIRO, 2012) (see Figures 1.5 – 1.10 for changes in streamflow for six of the major rivers in the region). Streamflow across the region increased drastically after the Millennium Drought during extended periods of above average rainfall experienced during 2010 and 2011 (CSIRO, 2012).
Declines in runoff are predicted to continue under climate change and it has been forecast that reductions of 5 and 30 % can be expected by 2030 in the Hopkins and Glenelg Rivers, respectively (Department of Sustainability and Environment, 2008a).
Figure 1.5: Mean daily flow (ML/day) for Glenelg River at Dartmoor (Gauge 238206)

Figure 1.6: Mean daily flow (ML/day) for Glenelg River at Big Cord (Gauge 238228)
Figure 1.7: Mean daily flow (ML/day) for Mount Emu Creek at Skipton (Gauge 236203)

Figure 1.8: Mean daily flow (ML/day) for Hopkins River at Hopkins Falls (Gauge 236209)
Figure 1.9: Mean daily flow (ML/day) for Moyne River at Toolong (Gauge 237200)

Figure 1.10: Mean daily flow (ML/day) for Eumerella River at Codrington (Gauge 236206)
1.8.4 Geomorphology in the Glenelg-Hopkins region

Flat volcanic plains dominate the southern and eastern areas of the Glenelg-Hopkins region, while older landforms such as the Grampians Mountains, the Dundas Tablelands and the central highlands dominate the north (Glenelg Hopkins CMA, 2002) (Figure 1.4). Soil distribution in the region is complex, with 49 types identified primarily consisting of red, yellow, brown and grey duplex soils, grey cracking clays and lateritic/stony profiles (Glenelg Hopkins CMA, 2002). The large variations in soil texture, structure, fertility, and drainage characteristics have been attributed to different ages and parent material geology, as well as the climatic conditions under which soil formation occurred across the region (Glenelg Hopkins CMA, 2002).

1.8.5 Land cover in the Glenelg-Hopkins region

Since European settlement in 1827, there has been extensive land-cover and land-use change in the Glenelg-Hopkins region (Dixon, 2000). Agricultural land cover dominates with dryland pasture and cropping covering approximately 70 % of the region (Figure 1.11). Pine (*Pinus radiata*) and *Eucalyptus* (*Eucalyptus globulus*) plantations are also a dominant feature of the landscape covering about 3 % and 5 % of the region, respectively, in 2002 (Ierodiaconou *et al.*, 2005). Contemporary land-cover changes have been well studied in the Glenelg-Hopkins region (Ierodiaconou *et al.*, 2005; Versace *et al.*, 2008a; Versace *et al.*, 2008b), while the long term impacts of an increase in plantation forestry (in particular *E. globulus*) has raised questions of impacts on streamflow and groundwater resources (Raleigh and Dixon, 2005; Benyon *et al.*, 2008). Some impacts of land-cover changes, particularly the replacement of grazing land by wheat and canola crops, have been identified in streamflow and groundwater records.
(Yihdego and Webb, 2011; Yihdego and Webb, 2013). Land-cover changes have also been implicated in dryland salinisation processes in the region (Versace, 2007). Lastly, the region contains approximately 44% of Victoria’s wetlands and it is thought that, since European settlement, over 75% of the region’s wetlands have been modified or degraded due to agricultural drainage (Corrick, 1992).
Figure 1.11: Aggregated land-cover in the Glenelg-Hopkins region in 2002. From Ierodiaconou et al. (2005).
1.8.6 Regional hydrology in the Glenelg-Hopkins region

Anthropogenic changes in the region have led to decreases in surface water quantity and quality. Declines in groundwater levels have been observed since the mid-to late-1980s in the majority of key bores in the region (Raleigh and Dixon, 2005). Yihdego and Webb (2011) suggested that changes in groundwater and lake water levels that had previously been attributed to changes in land cover were actually outweighed by climate variability. However, it has also been shown that an expansion in the number of farm dams for livestock watering reduced streamflows in the 1970s and 1980s and, furthermore, that decreases in streamflows in the 1990s were a result of the widespread replacement of grazing land by wheat and canola crops (Yihdego and Webb, 2013).

Water quality and associated trends vary across the region. Smith and Nathan (1998) demonstrated that gauging stations in the north-west of the region primarily showed a decline of in-stream salinity, while gauges in the lowland areas of the region displayed moderate increases in salinity over a 10-year period. In contrast, however, Dahlhaus et al. (2002) suggested that both land and water salinity were more prevalent prior to extensive land-cover change (i.e. post-European settlement) than previously thought. Relationships between the proportion of native vegetation and in-stream salinity were strong for the Glenelg and Portland Coast catchments but less evidence for the same relationships was identified in the Hopkins catchment (Versace et al., 2008b). This difference was suggested to be a result of the degraded nature of the Hopkins catchment, relative to the other two (Versace et al., 2008b).
1.9 Rationale for selecting the Glenelg-Hopkins region as a case study

The Glenelg-Hopkins region has undergone considerable land-cover change since European settlement. These changes are thought to have contributed to the present-day condition of both land and water (Ierodiaconou et al., 2005; Versace et al., 2008a; Versace et al., 2008b; Yihdego and Webb, 2011; Yihdego and Webb, 2013). A common theme amongst previous studies conducted in the region was that a lack of data availability and time precluded the use of complex hydrological models to examine impacts of climate and land cover on regional water resources. Likewise, previous studies have also suggested that there was significant spatial variability in the models they created. Therefore, with the recent development of appropriate datasets and frameworks linking hydrological models to spatial databases, the Glenelg-Hopkins region provides a suitable setting to develop a range of hydrological models to examine impacts of land-cover and climate change on streamflows, and to develop and refine methodologies to examine spatially-variable relationships between environmental variables and hydrology.

1.10 Objectives of the thesis

The overall objective of this thesis is to assess the ability of three different methods to examine and explain spatial and temporal variation in the relationships between environmental variables and hydrological responses. Specifically, this will be done by:

1) Creating a complex, semi-distributed, physically-based hydrological model. This objective allowed an assessment of the impact of land-cover changes that occurred in unison with an extended drought period. The creation of two independent models,
on two sub-catchments that have differing land covers and climates, allowed for a rigorous, mechanistic assessment of spatial differences in hydrology that could be exacerbated by anthropogenic impacts.

2) Examining spatio-temporal trends in the relationship between wetland extents and land cover. This objective utilised geographically weighted regression, and combined land cover maps from 1980, 1995 and 2002 with an appropriate climate indicator. This objective gave an indication of the strength of spatially-varying relationships in regional hydrology that have been commonly identified in (or hypothesised from) previous research.

3) Developing a hydrologic landscape regionalisation. A range of spatial databases on land cover, climate, geology and hydrology, were combined with non-parametric statistics and supervised image classification. The creation of this regionalisation allowed a spatially-explicit assessment of differences in regional hydrology, and defined a new framework that inherently preserves inter-catchment variability that is frequently lost using other hydrologic regionalisation frameworks.

4) Examining the relationship between the regionalisation classes and hydrology using the regionalisation created for Objective 3. By linking the hydrology of catchments back to the regionalisation, it was possible to draw conclusions about hydrology directly from the regionalisation – even in the absence of comprehensive streamflow data.

1.11 Thesis orientation

This thesis examines the spatial variability in the relationships between land cover and water quantity in a highly-modified, mixed-use regional catchment. A
multidisciplinary approach incorporating physical and statistical modelling, land-cover maps, historical streamflow data, and spatial databases of soil, geology and climate were used to assess the spatial variability in the relationships, and the strength of the relationship, among the variables.

Chapter 2 (‘Assessing the impact of drought and forestry on streamflows in south-eastern Australia using a physically-based hydrological model’) covers the development of a physically-based, semi-distributed streamflow model that links changes in land cover to streamflows during an extended drought period. The methodology developed in this chapter addressed the concerns of previous research that there is a lack of suitable data which has historically precluded the use of such methods. The development of this model was useful for assessing spatial differences in anthropogenic effects on streamflow, where differences existed. This research chapter is aligned with Objective 1 (‘1.10 Objectives of the thesis’) and has been accepted for publication in Environmental Earth Sciences.

Chapter 3 (‘Assessment of spatio-temporal varying relationships between rainfall, land cover and surface water area using geographically weighted regression’) applied geographically weighted regression (GWR), considered to be a recent improvement to ordinary least-squares regression. Though there has been some mention of spatial variability in the relationship between hydrology and land cover in previous research, there has been no work to identify and quantify the strength of this variability. This is the first comprehensive attempt to explicitly map spatial variability in the region. This research chapter is aligned with Objective 2 (‘1.10 Objectives of the thesis’) and has been published in a slightly modified form in Environmental Modeling and Assessment (Brown
et al., 2012). Chapter 2 was also presented at the “2nd EGU Leonardo Conference on the Hydrological Cycle – Looking at catchment in colors”. Luxembourg City, Luxembourg, 2010.

Chapter 4 (‘Getting down and dirty – can soil attributes help to quantify a spatially-varying relationship between rainfall, land cover and wetland extents?’) attempts to build on the results, and address some shortcomings identified with the method presented in Chapter 3. The chapter is written in the style of a short communication, and provides an assessment of modelling that incorporates a range of soil parameters, which were not readily available at the time of model development for Chapter 3. This research chapter is aligned with Objective 2 (‘1.10 Objectives of the thesis’).

Chapter 5 (‘Hydrologic landscape regionalisation using deductive classification and random forests’) developed a new framework for the creation of hydrologic landscape regionalisations. This framework was designed to preserve intra-catchment hydrologic variability that is often lost using other approaches. An initial assessment of the applicability of the method that links the regionalisation to regional hydrology is also presented. This research chapter is aligned with Objective 3 (‘1.10 Objectives of the thesis’) and has been published in PloS One (Brown et al., 2014).

Chapter 6 (‘Classifying stream variability in ungauged basins from a pixel-based hydrological regionalisation’) further validated the regionalisation created in Chapter 5 by linking a number of streamflow indices that are commonly used to describe stream behaviour directly to the regionalisation classes. Classification by two machine-learning methods was used to predict classes from the previously created regionalisation and from a new flow-based classification. A comparison between the regionalisation and the flow-
based classification was also conducted to test the applicability and merits of both approaches to regions with and without, extensive stream gauge networks. This research chapter is aligned with Objective 4 (‘1.10 Objectives of the thesis’) and is currently being prepared for submission to *Journal of Hydrology*.

The final chapter of the thesis (‘General discussion’) synthesises the key findings of each research chapter. This chapter is divided into five sections that: 1) discuss the broad outcomes of each chapter; 2) presents a simple, conceptual framework linking the methods presented in each chapter to spatio-temporal assessment of hydrologic variability; 3) considers the methodological implications to hydrological science and provides general recommendations for the examination of spatially-varying relationships in hydrologic and environmental studies; 4) reflects on what the research has contributed to the management of the case study region; and 5) identifies the next steps for future research.
Assessing the impact of drought and forestry on streamflows in south-eastern Australia using a physically-based hydrological model

2.1 Introduction

2.1.1 Land-use and land-cover changes

Land-use and land-cover changes (LULCC) have the ability to alter biotic diversity, affect local and regional climate change (Chase et al., 2000) and affect the capacity of biological systems to support human needs via ecosystem services (Lambin et al., 2001). Globally, LULCC is believed to be occurring at an unprecedented magnitude and rate (Lambin et al., 2001). Contemporary LULCC is predominately anthropogenic, and is usually directed at altering terrestrial systems for individual, societal or economic wants or needs. Agriculture has been the greatest force behind LULCC. Nearly one third of the Earth’s land surface is used for cropping or grazing, at the expense of remnant forests, grasslands, and wetlands (Ramankutty et al., 2006). For example, forest area has decreased globally from an estimated 53 million km² in 1700 to approximately 43.5 million km² in 2006, while savannas and grasslands have reduced from 30-32 to 12-23 million km² over the same timeframe (Ramankutty et al., 2006). More forests were cleared between 1950 and 1980 than in the early 18th and 19th Centuries combined (Ramankutty et al., 2006). Global wetland extents have also decreased by about half (OECD, 1996) and were estimated to be between 5.3 and 12.8 million km² in 2005 (Zedler and Kercher, 2005). The rapid global LULCC that has occurred is neither randomly nor uniformly distributed, but clustered in particular locations such as at forest edges and along road and transportation networks (Ramankutty et al., 2006).
2.1.2 LULCC impacts on the hydrological cycle

LULCC is one of the two dominant factors affecting the hydrological cycle, with the other being climate variability (McMahon and Finlayson, 2003; Li et al., 2009). Surface flow to streams is largely controlled by prevailing meteorological conditions and is generally greatest during and immediately after storm or rain events. Baseflow to streams, comprised of groundwater flow and interflow (lateral flow of water through the soil profile), is the dominant source of flow during dry periods. LULCC has been shown to alter baseflows and stream discharges (Costa et al., 2003) while also affecting flood frequency and severity (Brath et al., 2006). Changes in hydrology as a result of LULCC have been linked to vegetation primarily by changing evapotranspiration (ET) (Peel et al., 2010). However, the effect of vegetation on ET diminishes as the area of a catchment increases (Peel et al., 2010).

The impacts of LULCC on hydrology are variable in both time and space. Temporally, LULCC impacts have been observed affecting peak run-off rates of hydrographs, while long-term temporal changes have been observed for average annual runoffs (Costa et al., 2003; Brath et al., 2006). Spatially, LULCC impacts are most obvious at the slope, patch or field scale (Tollan, 2002), while the impact on annual water balance at the catchment scale is generally relatively small due to the complex nature and compensatory effects of the hydrological cycle at large scales (Fohrer et al., 2001; Peel et al., 2010). However, some evidence exists that LULCC and related anthropogenic activities can drastically alter the hydrological dynamics of large catchments more than climate variability with, for example, fluctuations in annual mean...
stream discharges following LULCC irrespective of prevailing weather conditions (Costa et al., 2003; Fan et al., 2010).

Generally, the removal of remnant vegetation and its replacement with shallow-rooted pasture or cropping species results in increased water yields, while afforestation of grasslands, pastures or crops results in a decrease in water yields. Studies investigating changes in small-scale catchments (approximately 10 km² or less) generally exhibit an increase in annual mean discharges as a result of deforestation, but few studies at meso- and macro-scales (>10,000 km²) are able to identify the same relationships (Costa et al., 2003). Therefore, it is essential to understand the impact that LULCC has on hydrological conditions, particularly baseflow, at a variety of temporal and spatial scales in low-flow systems such as those typically found in arid, semi-arid and Mediterranean climates.

2.1.3 SWAT model

There are currently three types of investigative hydrological modelling methods, each providing its own unique framework to help conceptualise and investigate hydrological relationships: experimental paired-catchment studies; statistical time-series analysis; and computer-based hydrological modelling. Statistical time-series analysis is generally simple and is useful for calculating flow probabilities or average recurrence intervals of floods and low flows; most computer-based hydrological models are however, complicated and require extensive data inputs, although the information they are required to output is often simple. Most of the previous research of LULCC impacts on the hydrological cycle has been conducted in small-scale catchments (~10s km²) using a paired-catchment approach. Research is now shifting towards regional
water management at meso- and macro-scales, as these studies are more relevant to provide information for economical, societal and environmental management and planning (Lahmer *et al.*, 2001). The rapid development of GIS technologies is also assisting in the ability to comprehensively model hydrological systems of catchments larger than 1 million km² (Xu and Singh, 2004). Assessments at large scales have been made easier with the development of physically-based, distributed hydrological models as they relate observable land and environmental characteristics (e.g. rainfall, land cover, soil properties and topography) directly to model parameters.

One of the most widely-used examples is the Soil and Water Assessment Tool (SWAT, http://swatmodel.tamu.edu; Arnold *et al.*, 1998; Gassman *et al.*, 2007). SWAT operates on a daily time step at the catchment scale and is designed to assess the impact of climate variability and land management practices on hydrology. SWAT was developed to predict the hydrological impact of land management practices over long periods of time in large, complex catchments with heterogeneous soils, topography, land cover and management practices (Gassman *et al.*, 2007). To achieve this, the model is semi-distributed and physically-based, requiring specific information on weather, soil properties, topography, land cover and land management practices within the catchment (Neitsch *et al.*, 2011). Catchment delineation and the creation of multiple homogenous hydrological response units (HRUs) are achieved through a two-step method. Firstly, a digital elevation model (DEM) is processed to generate sub-catchments, hill slopes, stream networks and channel lengths. The second step involves the overlay of land cover, soil classes and slope classes to generate multiple HRUs within each sub-catchment. The benefits of using a physically-based model include that ungauged catchments can be
modelled and the relative impact of climate, vegetation and land management on catchment hydrology can be estimated (Neitsch et al., 2011).

For SWAT to accurately predict the transportation of water, nutrients, sediments and pesticides through the catchment, the simulated hydrological cycle of the model must approximate reality. To do this, the hydrological cycle of the simulation is separated into two major divisions; the land phase and the routing phase. The driving force behind the land phase of SWAT is the water balance equation:

\[
SW_t = SW_0 + \sum_{i=1}^{t} \left( R_{i}^{\text{day}} - Q_{i}^{\text{surf}} - E_{i}^{a} - w_{i}^{\text{vap}} - Q_{i}^{\text{gw}} \right)
\]

where \(SW_t\) is the final soil water content and \(SW_0\) is the initial soil water content, \(R_{i}^{\text{day}}\) is the amount of precipitation, \(Q_{i}^{\text{surf}}\) is the amount of surface runoff, \(E_{i}^{a}\) is evaporation, \(w_{i}^{\text{vap}}\) is the amount of water entering the vadose zone from the soil profile and \(Q_{i}^{\text{gw}}\) is the amount of return baseflow, all on day \(i\) (Neitsch et al., 2011). All parameters are measured in mm.

Once the land phase of the simulated hydrological cycle has been modelled, SWAT then models the routing phase of the hydrological cycle. The routing phase is where water loadings, sediment, nutrients and pesticides are routed through the stream network. Routing through the main channel is separated into four components; flood routing, sediment routing, nutrient routing and channel pesticide routing. For a comprehensive explanation of the simulation of the water balance equation, the modelling of the land phase and the routing modelling used by SWAT, refer to Neitsch et al. (2011) and Arnold et al. (1998). SWAT has had limited application in Australia, but in recent years the model is becoming more popular, with modelled streamflows
closely resembling observed streamflows suggesting that the SWAT model is suitable for simulating catchment scale hydrology in Australia (Saha et al., 2014).

2.1.4 Aim of study

There has been anecdotal evidence that the introduction of plantation forestry into predominantly-agriculturally catchments has caused substantial reductions in streamflows within those catchments, leading to concern amongst environmental managers. One such example has occurred with the introduction of \textit{Eucalyptus globulus} plantations into the Hopkins River catchment, Australia. Earlier work in the region, and nearby, has indicated that it is possible that LULCC may be influencing wetland extents (Chapter 3), groundwater elevations (Yihdego and Webb, 2011) and streamflows (Yihdego and Webb, 2013). The introduction of the \textit{Eucalyptus}, however, coincided with a severe, long-term drought (“Millennium Drought”; 1997 – 2009, CSIRO, 2012). Therefore, the aim of this study was to use SWAT to create a hydrological model to examine differences in streamflows so as to assess the impacts of the introduction of plantation forestry independent of drought, using two sub-catchments of the Hopkins River catchment as a case study. The daily streamflow models (i.e. plantation and no-plantation model) for each sub-catchment (Hopkins and Mount Emu Creek) were then aggregated to monthly, annual and seasonal mean daily flows to allow assessments at scales more appropriate for management.
2.2 Materials and methods

2.2.1 Australian LULCC

In Australia, LULCC was a feature of landscape transition prior to European settlement through indigenous land management practices (Ramankutty et al., 2006). However, extensive and rapid land-cover modification occurred after European settlement (from 1788) with land clearing for agricultural pursuits beginning in the late 19th Century (Ramankutty et al., 2006). The expansion in agriculture is estimated to have degraded, removed or modified 69% of the remnant native forest in the state of Victoria and 50% of the remnant native forest of New South Wales (Ramankutty et al., 2006). Relative to size, estimates suggest that Victoria has undergone the greatest level of land clearing of all the Australian states and territories (Graetz et al., 1995). While rates of land clearing between 1980 to 1990 were spatially varied, clearing was always at the expense of remnant native forest, woodlands or scrublands, while the primary drivers for the change were wool, beef, milk and wheat production and, to a lesser extent, urban and industrial expansion (Australian Greenhouse Office, 2000).

2.2.2 Australian hydrological conditions

Australia is the world’s driest inhabited continent, with 66 % of the continent receiving <500 mm of rainfall per annum (Arthington and Pusey, 2003). Average rainfall across the entire continent is only 455 mm per annum. Evapotranspirative losses are extremely high, resulting in a mean annual runoff of only 12 % of rainfall, with 75 % of the Australian continent receiving <12.5 mm of annual runoff on average. In contrast, North America has a mean annual percentage rainfall as runoff of 33 % of rainfall (Arthington and Pusey, 2003). Furthermore, previous modelling has shown that, during
typical drought years, it is not uncommon for ET to greatly exceed rainfall, which can result in significant losses of soil water (Saha et al., 2014). Due to these low runoff ratios and high evaporation and ET losses, Australia’s perennial rivers are dominated by baseflow (McMahon and Finlayson, 2003). These baseflows are necessary to maintain rivers in the arid, semi-arid and Mediterranean climate zones and during extended dry periods typical of much of the Australian continent.

2.2.3 Study site

The Hopkins River catchment forms part of the Glenelg-Hopkins region of western Victoria, Australia. The Hopkins catchment covers 9,832 km² of the eastern side of the Glenelg-Hopkins region. The catchment is dominated by agriculture, with 90 % being used for pasture and dryland grain crops, and 8 % supporting remnant native vegetation and wetlands. Following its introduction to the region, Eucalyptus plantation forestry accounts for <2 % of the catchment. The location and land cover of the two sub-catchments (Hopkins [South], Mount Emu Creek [North]) used in this study are presented in Figure 2.1, with tabulated land cover and sub-catchment statistics provided in Table 2.1.
Figure 2.1: Location and land cover of the two sub-catchments within the Hopkins catchment of south-eastern Australia. Land-cover data are from Ierodiaconou et al. (2005). A number of catchment statistics on climate, topography and land cover are provided in Table 2.1.
Table 2.1: Climate, topography, and land cover statistics for the two case study sub-catchments, Hopkins and Mt Emu Creek. ‘m.a.s.l’ is metres above mean sea level. Land cover acronyms in parentheses represent the SWAT land cover database code used for modelling. * indicates the custom evergreen forest class that was created to simulate *Eucalyptus*.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Hopkins</th>
<th>Mt Emu Creek</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Topography</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area (km²) (not including upstream sub-catchments)</td>
<td>407</td>
<td>570</td>
</tr>
<tr>
<td>Maximum Elevation (m.a.s.l)</td>
<td>259</td>
<td>626</td>
</tr>
<tr>
<td>Minimum Elevation (m.a.s.l)</td>
<td>24</td>
<td>276</td>
</tr>
<tr>
<td>Mean Slope (degrees)</td>
<td>1.8</td>
<td>2.3</td>
</tr>
<tr>
<td>Maximum Slope (degrees)</td>
<td>42.3</td>
<td>38.8</td>
</tr>
<tr>
<td><strong>Climate (1961-1990)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Annual Precipitation (mm)</td>
<td>806</td>
<td>686</td>
</tr>
<tr>
<td>Mean Annual Temperature (°C)</td>
<td>14</td>
<td>12</td>
</tr>
<tr>
<td>Mean Annual Evapotranspiration (mm)</td>
<td>611</td>
<td>605</td>
</tr>
<tr>
<td>Mean Annual Discharge (mm)</td>
<td>24</td>
<td>30</td>
</tr>
<tr>
<td><strong>Land cover</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pasture (PAST) (%)</td>
<td>90.75</td>
<td>76.95</td>
</tr>
<tr>
<td>Crops (AGRR) (%)</td>
<td>0.15</td>
<td>4.34</td>
</tr>
<tr>
<td>Irrigated Pasture (PAST) (%)</td>
<td>1.18</td>
<td>0.00</td>
</tr>
<tr>
<td>Pine Plantation (PINE) (%)</td>
<td>0.00</td>
<td>0.95</td>
</tr>
<tr>
<td><em>Eucalyptus</em> Planation (EUCL*) (%)</td>
<td>1.69</td>
<td>3.42</td>
</tr>
<tr>
<td>Remnant Vegetation (FRSE) (%)</td>
<td>4.79</td>
<td>6.84</td>
</tr>
<tr>
<td>Urban (URBN) (%)</td>
<td>0.54</td>
<td>0.28</td>
</tr>
<tr>
<td>Wetlands (WETL) (%)</td>
<td>0.90</td>
<td>7.22</td>
</tr>
</tbody>
</table>
2.2.4 SWAT setup, calibration, sensitivity analysis, validation and LULCC assessment

The elevation data used in the study consisted of a smoothed 90-m resolution DEM (Geoscience Australia, 2011). This DEM was used by SWAT to generate sub-catchment boundaries and the stream network. A pre-existing stream network was also used to ‘burn in’ DEM cells to assist in the delineation, as the study area is quite flat. Land-cover data for the region was obtained from Ierodiaconou et al. (2005). Soil data were derived from the FAO Digital Soil Map of the World database (FAO, 2007). The database consists of a raster and a linked database of soil properties, for two soil horizons, at a 10-km resolution. Gridded datasets (5 km resolution) (Bureau of Meteorology, 2012) were used to calculate average daily rainfall, and minimum and maximum temperatures in the sub-catchments, while two long-term weather monitoring stations near the catchments were used to generate climate statistics. Daily streamflows used for calibration and validation were extracted from the Water Measurement Information System (http://data.water.vic.gov.au/monitoring.htm).

A SWAT 2009 model (revision 510) was created for each sub-catchment using the ArcSWAT interface (version 2009.93.7b), with 26 and 15 HRUs defined for the Hopkins and Mount Emu Creek sub-catchments, respectively. Hargreaves Potential ET (PET) method was used to estimate PET as it only requires air temperature as input data (Neitsch et al., 2011). Surface runoff was calculated using the Soil Conservation Service (SCS) curve number method, while flow routing through the streams was calculated with the variable storage method (Neitsch et al., 2011).

Minor manual model calibration was conducted prior to automatic calibration. Manual calibration included adjustment of a number of model parameters (Neitsch et
al., 2011): the baseflow recession constant (ALPHA_BF; Arnold and Allen, 1999), the maximum canopy storage for forests, pastures and crops (CANMX; Breuer et al., 2003; Ladson, 2011), the plant uptake compensation factor (EPCO; Neitsch et al., 2011) and the SCS moisture condition II curve number (CN2; Neitsch et al., 2011) parameters.

Manual calibration prior to automatic calibration has been shown to improve automatic calibration results in other Australian SWAT models (Saha et al., 2014) and is generally recommended by the SWAT development team (R. Srinivasan, pers. comm.). Automatic model calibration was conducted using the SUFI-2 algorithm (Abbaspour et al., 2007) and SWAT-CUP (version 4.3.7) (Abbaspour, 2011) on daily flow data between January 1980 and December 1991. These dates were deemed suitable as they represented a long calibration period where there were representative periods of high, average and low flow conditions (Figure 2.2). Statistical evaluation of the performance of the calibration was assessed using the Nash-Sutcliffe Efficiency ($E_{NS}$)

$$E_{NS} = 1 - \frac{\sum_{i=1}^{n} \left( Y_{i}^{obs} - Y_{i}^{sim} \right)^2}{\sum_{i=1}^{n} \left( Y_{i}^{obs} - Y_{mean} \right)^2}$$

in SWAT-CUP, where $Y_{i}^{obs}$ is the observed flow, $Y_{i}^{sim}$ is the simulated flow, both on day $i$, and $Y_{mean}$ is the mean observed flow for the period of record. $E_{NS}$ ranges between $-\infty$ and 1, with $E_{NS} = 1$ indicating a perfect match between simulated and observed data. Values <0 imply that the mean of the observed data is a better predictor than the simulated values and indicate unacceptable performance (Moriasi et al., 2007).
Further evaluation of the calibration was conducted with the R (R Core Team, 2014) hydroGOF package (Zambrano-Bigiarini, 2014a) using: Percentage Bias (PBIAS)

\[
PBIAS = 100 \times \left[ \frac{\sum_{i=1}^{n} (Y_{o,i} - Y_{s,i})^2}{\sum_{i=1}^{n} Y_{o,i}^2} \right]
\]

where terms are defined as above. Positive values of PBIAS indicate model overestimation bias, while negative values indicate underestimation (Zambrano-Bigiarini, 2014a); and the Root Mean Square Error (RMSE) observations standard deviation ratio (RSR)

\[
RSR = \frac{RMSE}{STDEV_{o,i}} = \frac{\sqrt{\sum_{i=1}^{n} (Y_{o,i} - Y_{s,i})^2}}{\sqrt{\sum_{i=1}^{n} (Y_{o,i} - Y_{mean})^2}}
\]

where terms are defined as above. Lower values for RSR indicate a lower RMSE, and therefore better model performance.

Criteria provided by Moriasi et al. (2007) explain that streamflow statistics with ENS > 0.50, -25% ≤ PBIAS ≥ 25% and RSR < 0.70 are considered good fits on models created at monthly time steps and can be relaxed to account for finer (e.g. daily as used here) time steps (Moriasi et al., 2007). Models were validated using the same statistics with streamflows from January 1992 to December 2009, using the best calibrated parameter ranges. There are numerous studies demonstrating the suitability of the SUFI-2 algorithm to model calibration for large and small catchments (Abbaspour et al., 2007). One major benefit SUFI-2 is its ability to calculate uncertainty ranges for the conceptual model, input uncertainty and parameter uncertainty. These uncertainties are measured by the P factor, which is the percentage of observed data that falls within a 95
% prediction uncertainty (95PPU) band calculated by Latin Hypercube (LH) sampling of the simulated data (Abbaspour et al., 2007). A higher P factor indicates that more of the observed data are contained within the uncertainty bands of the simulated data, suggesting that the conceptual model and the chosen parameter ranges are suitable. An R factor is also calculated, which is the average range between the upper and lower bounds of the 95PPU divided by the standard deviation of the observed data (Abbaspour et al., 2007). SUFI-2 attempts to maximise the P factor, while minimising the R factor (Abbaspour et al., 2007). Calibration parameters, final parameter ranges and the results of global sensitivity analysis are presented in Table 2.2.
Figure 2.2: Calibration and validation hydrographs for the two sub-catchments. Automatic model calibration resulted in very-good fits to observed streamflows using a number of statistics during both calibration and validation periods.

- a) Hopkins River calibration
  - Observed flow
  - Simulated flow
  - 95 PPU
  - p-factor = 0.65
  - r-factor = 0.18

- b) Hopkins River validation
  - Observed flow
  - Simulated flow

- c) Mount Emu Creek calibration
  - Observed flow
  - Simulated flow
  - 95 PPU
  - p-factor = 0.47
  - r-factor = 0.25

- d) Mount Emu Creek validation
  - Observed flow
  - Simulated flow
In the context of this study, global sensitivity analysis is a procedure used to determine the rate of change of modelled streamflows in response to variations in different calibration parameters (Moriasi et al., 2007). This process can identify the parameters to which the calibration procedure is most sensitive. As for the calculation of the 95PPU, SWAT-CUP performs global sensitivity analysis using LH sampling. In an LH loop, parameter values are altered and assessed against the average change in $E_{NS}$ as a result from changes in each parameter, while all other parameters are changing. A $t$-test is then used to identify the relative significance of each parameter to calibration (Abbaspour, 2011). This process provides relative sensitivities based on linear approximations between the model and $E_{NS}$ and, as such, only provides partial information about the sensitivity of the calibration procedure to the calibration parameters (Abbaspour, 2011). Conversely, One-At-a-Time (OAT) sensitivity analysis calculates the sensitivity of a variable to the changes in $E_{NS}$ if all other calibration parameters are kept at some constant value. OAT sensitivity analysis was not conducted for this study as it is particularly time-consuming for large numbers of variables, and there was no a priori method to determine what the value of the other parameters (held constant) should be. This was an important consideration in the choice of approach as the sensitivity of one parameter can depend on the values of other parameters (Abbaspour, 2011).

Annual flow duration curves (FDC) were produced to help visualise differences in streamflows, while hydroTSM (Zambrano-Bigiarini, 2014b) and hydroGOF were used for the aggregation of monthly, seasonal and annual time-series from the daily streamflow data and visual assessment of seasonal differences between the models.
Baseflows were separated from the observed and simulated streamflows using the method described in Arnold and Allen (1999) to enable assessment of water balance partitioning by SWAT.
Table 2.2: Parameters used in flow calibration, final parameter ranges and results of sensitivity analysis. See Neitsch et al. (2011) for detailed parameter explanations. The input adjustment ranges were the values used in SWAT-CUP for automatic calibration while the final parameter ranges represent the final adjusted ranges applied by the models. * = parameter was adjusted depending on land cover and/or soil type/layer. ** = baseflow recession is equal to (1 / number of days for stormflow to recede).

Adjustment Types: a = a random value from the adjustment range is added to the existing parameter value, r = the existing parameter value is multiplied by (1+ a random value from the adjustment range), v = the existing parameter value is to be replaced by a random value from the adjustment range. See Abbaspour (2011) for more details.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Unit</th>
<th>Adjustment</th>
<th>Description</th>
<th>Adjustment Ranges</th>
<th>Calibrated Ranges</th>
<th>Sensitivity Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alpha BF[^a]</td>
<td>1/day[^*]</td>
<td>a</td>
<td>Baseflow recession constant</td>
<td>0.0 – 0.5</td>
<td>0.61 – 0.93</td>
<td>3.36 0.01 20.88 0.01</td>
</tr>
<tr>
<td>Canmx[^b]</td>
<td>mm H<del>2</del>O</td>
<td>v</td>
<td>Maximum amount of water that can be trapped in the fully developed canopy</td>
<td>0 – 3</td>
<td>0 – 2</td>
<td>4.40 0.01 4.36 0.01</td>
</tr>
<tr>
<td>CH_K2[^c]</td>
<td>mm H<del>2</del>O/hour</td>
<td>v</td>
<td>Effective hydraulic conductivity of main channel alluvium</td>
<td>0 – 150</td>
<td>1.00 – 27.12</td>
<td>0.92 0.36 -1.01 0.31</td>
</tr>
<tr>
<td>CN2[^d]</td>
<td></td>
<td>r</td>
<td>Moisture condition II curve number</td>
<td>-0.25 – 0.25</td>
<td>63.2 – 79.6</td>
<td>-14.75 0.01 9.02 0.01</td>
</tr>
<tr>
<td>EPCO[^e]</td>
<td></td>
<td>v</td>
<td>Plant soil-water uptake compensation factor</td>
<td>0.01 – 1.00</td>
<td>0.42 – 0.91</td>
<td>1.46 0.14 -2.77 0.01</td>
</tr>
<tr>
<td>ESCO[^f]</td>
<td></td>
<td>v</td>
<td>Soil evaporation compensation coefficient</td>
<td>0.01 – 1.00</td>
<td>0.73 – 0.96</td>
<td>-2.14 0.03 0.00 1.00</td>
</tr>
<tr>
<td>GW_REVAP[^g]</td>
<td>mm H<del>2</del>O</td>
<td>v</td>
<td>Revaporation coefficient</td>
<td>0.02 – 0.20</td>
<td>10.79 – 221.87</td>
<td>11.00 0.01 -2.89 0.01</td>
</tr>
<tr>
<td>GWDELAY[^h]</td>
<td>days</td>
<td>v</td>
<td>Delay time for aquifer recharge</td>
<td>1 – 20</td>
<td>0.4 – 1.6</td>
<td>6.26 0.01 4.77 0.01</td>
</tr>
<tr>
<td>GWQMN[^i]</td>
<td>mm H<del>2</del>O</td>
<td>v</td>
<td>Threshold depth of water in shallow aquifer for baseflow to occur</td>
<td>0 – 300</td>
<td>9.28 – 120.58</td>
<td>23.34 0.01 1.65 0.10</td>
</tr>
<tr>
<td>RCHRG_DP[^j]</td>
<td></td>
<td></td>
<td>Aquifer percolation coefficient</td>
<td>0.0 – 0.5</td>
<td>0.09 – 0.25</td>
<td>6.69 0.01 6.24 0.01</td>
</tr>
<tr>
<td>REVAPMN[^k]</td>
<td>mm H<del>2</del>O</td>
<td>v</td>
<td>Threshold depth of water in shallow aquifer for revaporation to occur</td>
<td>100 – 1500</td>
<td>350 – 1240</td>
<td>-12.23 0.01 -2.89 0.01</td>
</tr>
<tr>
<td>SOL_AWC[^l]</td>
<td>mm H<del>2</del>O/mm soil</td>
<td>r</td>
<td>Soil available water capacity</td>
<td>-0.25 – 0.25</td>
<td>0.134 – 0.175</td>
<td>1.86 0.05 -2.41 0.02</td>
</tr>
<tr>
<td>SOL_Z[^m]</td>
<td>mm</td>
<td>r</td>
<td>Soil depth</td>
<td>-0.25 – 0.25</td>
<td>350 – 1000</td>
<td>0.68 0.50 0.14 0.89</td>
</tr>
<tr>
<td>SURLAG[^n]</td>
<td>days</td>
<td>a</td>
<td>Surface runoff lag coefficient</td>
<td>1 – 6</td>
<td>0.5 – 1.2</td>
<td>-36.98 0.01 -0.50 0.62</td>
</tr>
</tbody>
</table>
To quantify the impact of the introduction of *Eucalyptus* forestry on the streamflows within the sub-catchments independent of the drought-induced changes in streamflow, the SWAT Land Use Update module was used. This module allows land cover within the model to be altered during simulation. The existing evergreen forest class (FRSE) was altered using the parameters presented in Watson (2006, Appendix B.3) to enable representation of *Eucalyptus* which is currently not present in the database. This was done to enable SWAT to more accurately model *Eucalyptus* than would have been possible had the default evergreen forest class been used for simulation. The Land Use Update model currently only allows for integer conversions of land-cover classes so an additional pasture class was also created to enable conversion of pasture to *Eucalyptus* at the appropriate (non-integer) percentages.

For this study, models were calibrated without the presence of *Eucalyptus*, reflecting land cover in the region prior to the introduction of *Eucalyptus*. The *Eucalyptus* plantations were introduced (plantation scenario) by converting the custom pasture class to *Eucalyptus* in 2001, nine years after the beginning of the validation period to coincide with the majority of *Eucalyptus* planting in the two sub-catchments in 2001 (Ierodiaconou *et al.*, 2005). A second simulation was also run for the same validation period, where the custom pasture class was not converted to *Eucalyptus* (no-plantation scenario). By altering only the land-use component, SWAT could reliably simulate changes in streamflows as a result of climate relative to changes as a result of land-cover change, allowing an independent assessment of the impacts of LULCC and drought on streamflows. The streamflows from the plantation and no-plantation scenarios for both sub-catchments were then assessed with a Mann-Whitney U test (R Core Team, 2014) to
test for significant changes as a result of an increase in plantation forestry. Modelled ET for the two scenarios were assessed with a paired-sample Wilcoxon test (R Core Team, 2014) to test for differences as a result of an increase in plantation forestry. The modelled ET from the plantation scenario was also compared against long-term mean observed monthly ET as an assessment of the models ability to realistically simulate ET.

2.3 Results

2.3.1 Model calibration, sensitivity analysis and validation

The calibration and validation of each model was determined to be very good (Figure 2.2), with high ENS values, indicating high efficiency, and low PBIAS and RSR values, indicating low unexplained variance and bias in the simulated data (Table 2.3). The model for the Hopkins sub-catchment was able to capture 96 % of the flow variability and magnitude during the calibration and validation stages, while the Mount Emu Creek sub-catchment model was able to capture 77 and 80% of the flow variability and magnitude during the calibration and validation stages, respectively. A relatively high P factor and low R factor were observed during the calibration stage of developing both models, indicating that conceptually the models were acceptable, capturing an acceptable level of parameter and prediction uncertainty (Figure 2.2). This was defined as the majority of observed streamflows being contained by the uncertainty bounds (95PPU) of the simulated streamflow, while keeping the width of the uncertainty bounds as small as possible.

Visual and statistical comparison of daily simulated and observed baseflows suggested that baseflow timing and magnitude estimation were very good for both models (Table 2.3). The Hopkins sub-catchment model tended to over-estimate
baseflows during the calibration period and then consistently under-estimated baseflows during the validation period. The Mount Emu Creek sub-catchment model consistently over-estimated baseflows during both the calibration and validation periods.

When considering the period of interest, the Millennium Drought, the Hopkins sub-catchment model produced a very good fit for the observed streamflow during that period ($E_{NS} = 0.87$, $RSR = 0.27$) and, although PBIAS (-24.7 %) suggested flows were under-estimated, they were still considered satisfactory according to the criteria from Moriasi et al. (2007). The Mount Emu Creek sub-catchment model produced satisfactory fits for the observed data during the same period based on $E_{NS}$ (0.58) and $RSR$ (0.65) values, however, flows were unsatisfactorily over-estimated according to the PBIAS statistic (86 %). As flows were underestimated in the Hopkins and overestimated in the Mount Emu Creek sub-catchment during the Millennium Drought, an analysis of daily baseflows was not considered. Likewise, analysis of flows at monthly and annual time scales was also not considered during the Millennium Drought.
Table 2.3: SWAT model evaluation statistics. The monthly and annual parameters were calculated after aggregating the daily simulated and observed streamflows to the appropriate time scales. Streamflow = surface flow + baseflow, Baseflow = baseflow contribution. The calibration period was between January 1980 and December 1991, while model validation was January 1992 to December 2009. The Millennium Drought occurred between 1997 and 2009. ENS = Nash-Sutcliffe Efficiency, PBIAS = Percentage Bias, RSR = Root Mean Square Error observations standard deviation ratio. Guidelines for the statistics are summarised in Section 2.2.4 and additional detail can be found in Moriasi et al. (2007) and Zambrano-Bigiarini (2014a).

<table>
<thead>
<tr>
<th></th>
<th>Daily</th>
<th>Monthly</th>
<th>Annual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Streamflow</td>
<td>Baseflow</td>
<td>Streamflow</td>
</tr>
<tr>
<td></td>
<td>ENS</td>
<td>PBIAS</td>
<td>RSR</td>
</tr>
<tr>
<td><strong>Hopkins River</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calibration</td>
<td>0.96</td>
<td>8.0</td>
<td>0.20</td>
</tr>
<tr>
<td>Validation</td>
<td>0.96</td>
<td>-11.6</td>
<td>0.19</td>
</tr>
<tr>
<td>Millennium Drought</td>
<td>0.87</td>
<td>-24.7</td>
<td>0.27</td>
</tr>
<tr>
<td><strong>Mount Emu Creek</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calibration</td>
<td>0.77</td>
<td>0.1</td>
<td>0.48</td>
</tr>
<tr>
<td>Validation</td>
<td>0.80</td>
<td>4.0</td>
<td>0.44</td>
</tr>
<tr>
<td>Millennium Drought</td>
<td>0.58</td>
<td>86.0</td>
<td>0.65</td>
</tr>
</tbody>
</table>
The sensitivity analysis in SWAT-CUP showed that the majority of the parameters significantly affected the calibration ($P \leq 0.05$). The baseflow recession constant (ALPHA_BF), maximum canopy storage (CANMX), soil moisture condition II curve number (CN2), the groundwater delay time (GW_DELAY), groundwater revaporation coefficient (GW_REVAP), aquifer percolation coefficient (RCHRG_DP), and the soil available water capacity (SOL_AWC) were all shown to significantly affect auto-calibration in both sub-catchments (Table 2.2). The surface lag runoff coefficient (SURLAG) and the baseflow recession constant (ALPHA_BF) were found to be the single most sensitive calibration parameters for the Hopkins and Mount Emu Creek models, respectively. The plant uptake compensation factor (EPCO) was significant for the Mount Emu Creek model, while the soil evaporation compensation factor (ESCO) was significant for the Hopkins model. Only two of the parameters modified during calibration were not found to be sensitive for either sub-catchment: soil depth (SOL_Z); and the effective hydraulic conductivity of the main channel alluvium (CH_K2).

2.3.2 Monthly, annual and seasonal flows

Analysis of monthly and annual daily average flows indicated non-significant increases in model fit over those observed for daily simulated flows, with very similar ENS, PBIAS and RSR values (Table 2.3, Figure 2.3).
Figure 2.3: Hydrographs of average daily flows aggregated to monthly and annual time steps. The vertical blue line represents the beginning of the validation period. Very good model fits were observed when modelled daily flows were aggregated to monthly and annual time steps, but the improvement was not significant.
Seasonal streamflow analysis demonstrated mixed results (Table 2.4, Figure 2.4). Simulated flows for winter and spring were very good for both models; while summer and autumn predictions were very good regarding $E_{NS}$ and RSR for the Hopkins model, and either satisfactory or good for the Mount Emu Creek model; both models performed poorly with regards to PBIAS. The negative PBIAS values for the Hopkins during summer and autumn demonstrate that the model underestimated streamflows, while positive values for Mount Emu Creek demonstrate overestimation (Table 2.4).

Similarly seasonal baseflow analysis results were also mixed (Table 2.4). Hopkins modelled baseflows were very good regarding all metrics during winter and spring ($E_{NS} > 0.9$, -10% $< $ PBIAS $> $ 10%, RSR $< $ 0.25). Autumn baseflows for the Hopkins model were good with regard to $E_{NS}$ and RSR; PBIAS however was considered unsatisfactory and demonstrated a large underestimation of baseflows (PBIAS = -25.9%).

The Mount Emu Creek model also demonstrated similar results. Winter and spring simulated baseflows were very good ($E_{NS} = 0.9$, -10% $< $ PBIAS $> $ 10%, RSR $< $ 0.25) with the exception of winter PBIAS which was rated as good (PBIAS = 15.6%). Summer baseflows were good with regard to $E_{NS}$ and RSR, while PBIAS was unsatisfactory and indicated significant overestimation (PBIAS = 51.2%). Autumn baseflows in the Mount Emu Creek model were considered unsatisfactory for all metrics and were severely overestimated.
Figure 2.4: Hydrographs of average daily flows aggregated to seasons. Analysis of seasonal flows provided a mix of results depending on season and sub-catchment.
2.3.3 Flow duration curves

Examination of the FDC for both sub-catchments (Figure 2.5) suggested that simulated baseflow contributions to the streams differed significantly from the observed data. The Hopkins model did not produce flow rates (or discharge; Q) with a probability of exceedance greater than 80 % \((Q_{80})\) and ~50 % \((Q_{30} > Q < Q_{80})\) of simulated streamflows were lower than those observed, again suggesting relatively poor baseflow simulation (Figure 2.5). The Mount Emu Creek model failed to produce flows greater than \(Q_{95}\) of the observed Mount Emu Creek flow. The mid-range flows of the Mount Emu Creek sub-catchment \((Q_{25} > Q < Q_{75})\) were higher than observed flows suggesting an over-contribution of baseflows to the model. These results support the analyses of baseflows above (Section 2.3 and Section 2.3.2).

The high-flow ends of both FDC suggests that the model is simulating storm flows and surface runoff accurately with no observable difference in the \(Q_{25}\) of the Hopkins model (top row, Figure 2.5) nor in the \(Q_{90}\) flows for Mount Emu Creek (bottom row, Figure 2.5). Visual examination of the Hopkins model FDC shows no difference between the plantation and no-plantation scenarios, suggesting that LULCC did not affect streamflows in this sub-catchment, while minor differences were observed for the Mount Emu Creek scenarios \((\geq Q_{80})\) implying slightly higher flows during the no-plantation scenario (Figure 2.5).
Table 2.4: Evaluation statistics for seasonal flows for both sub-catchments. Winter and spring flows consistently outperformed summer and autumn flows. Streamflow = surface flow + baseflow, Baseflow = baseflow contribution. ENS = Nash-Sutcliffe Efficiency, PBIAS = Percentage Bias, RSR = Root Mean Square Error-observations standard deviation ratio. Guidelines for the statistics are summarised in Section 2.2.4 and additional detail can be found in Moriasi et al. (2007) and Zambrano-Bigiarini (2014a).

<table>
<thead>
<tr>
<th></th>
<th>Streamflow</th>
<th>Baseflow</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Summer</td>
<td>Autumn</td>
</tr>
<tr>
<td>Hopkins</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENS</td>
<td>0.97</td>
<td>0.86</td>
</tr>
<tr>
<td>PBIAS</td>
<td>-17.9</td>
<td>-28.0</td>
</tr>
<tr>
<td>RSR</td>
<td>0.26</td>
<td>0.37</td>
</tr>
<tr>
<td>Mount Emu Creek</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENS</td>
<td>0.62</td>
<td>0.50</td>
</tr>
<tr>
<td>PBIAS</td>
<td>45.3</td>
<td>81.5</td>
</tr>
<tr>
<td>RSR</td>
<td>0.61</td>
<td>0.70</td>
</tr>
</tbody>
</table>
Figure 2.5: Flow duration curves (log$_{10}$ m$^3$/s) for both sub-catchments. Visual analysis of the flow duration curves indicated that both models were good at simulating storm flows and surface runoff; however, baseflows were poorly simulated.
2.3.4 Mean daily streamflows, assessment of LULCC and monthly ET

Significant differences were observed in both sub-catchments between the mean daily flows for the pre-drought compared with the drought periods of the observed flow data (Mean daily flow for Hopkins: 9.3 m$^3$/s pre-drought and 2.1 m$^3$/s during the drought, respectively and Mount Emu Creek: 1.8 m$^3$/s pre-drought and 0.2 m$^3$/s drought, respectively). Significant differences also existed between the mean daily flows of the simulated data for the pre-drought and drought periods (Mean daily flow for Hopkins: 9.6 m$^3$/s pre-drought and 1.6 m$^3$/s during the drought respectively and Mount Emu Creek: 1.8 m$^3$/s pre-drought and 0.3 m$^3$/s drought, respectively).

There were no significant differences detected between the daily simulated flows when examining the plantation scenarios post-planting date, i.e. the presence or absence of *Eucalyptus* did not affect mean daily simulated flows between January 2001 and December 2009 (Mean daily flow for Hopkins: 1.6 m$^3$/s for the plantation and 1.7 m$^3$/s for the no-plantation scenarios, respectively [W = 5395321, P = 0.93]; Mount Emu Creek: 0.3 m$^3$/s for each of the plantation and no-plantation scenarios [W = 5317860, P = 0.27]).

There were no significant differences in simulated ET between the plantation and no-plantation scenarios in either sub-catchment (Hopkins: V = 45, P = 0.31; Mount Emu Creek: V = 25, P = 0.29). Significant differences were identified between observed long-term and simulated average monthly ET in both sub-catchments (Hopkins: V = 74, P < 0.001; Mt Emu Creek: V = 78, P < 0.001). However, visual examination of monthly ET values shows that the simulation follows a very similar pattern to the observed data (Figure 2.6). Total annual average ET values for the Hopkins sub-
catchment were ~100 mm less per year than observed and ~130 mm less per year in the Mt Emu Creek sub-catchment.
Figure 2.6: Average monthly ET (mm) values for the two sub-catchments. While ET was underestimated by both models, modelled monthly ET values closely resembled long-term observed ET.
2.4 Discussion

The objective of this study was to create a hydrological model for two meso-scale catchments in south-eastern Australia and to assess the impacts of a prolonged drought compared with those of a small increase in *Eucalyptus* plantation forestry within those catchments. Anecdotal evidence suggested that the increase in plantation forestry had caused a decrease in streamflows but south-eastern Australia also experienced a severe, long-term drought between 1997 and 2010 (CSIRO, 2012) which made quantification of the impact of increased plantation forestry difficult. The SWAT models developed demonstrated very good calibration and validation results at a daily time step. These models allowed for an assessment of the impacts of forestry and of drought on streamflows independently.

2.4.1 Applicability of SWAT to the Hopkins and Mount Emu Creek sub-catchments

The two SWAT models created for this study were classified as very good using the guidelines provided by Moriasi *et al.* (2007). Except for several high-flow events during the calibration stage of the Mount Emu model, streamflows and baseflows were generally matched very well to the observed data. As the SWAT model was not created to model single-event, high-flow conditions (Arnold *et al.*, 1998), this is an acceptable result that is in line with recent work in Australia (Saha *et al.*, 2014). Saha *et al.* (2014) found that a SWAT model created for the Yass River, Australia overestimated baseflow contributions to streams during both the calibration and validation stages by 6 % and 3 %, respectively. The models generated in this study showed a mixture of over- and underestimation of baseflows suggesting an issue with water balances. The model for the
Hopkins sub-catchment was found to overestimate baseflows by ~10% during calibration and underestimate by ~13% during validation. The Mount Emu Creek model overestimated baseflows during both calibration and validation by ~7% and ~18%, respectively. Nonetheless, these values still demonstrate a good model fit based on the criteria used.

Predictably, as the flows in both sub-catchments are dominated by baseflow, the majority of parameters identified as sensitive during calibration affected either soil moisture or groundwater flow. Parameters such as the baseflow recession coefficient and the groundwater delay time, which control the rate and timing of baseflow releases and recharge of the shallow aquifer, were shown to be very sensitive to calibration and suggest that the region has a quick response to groundwater recharge (Smedema and Rycroft, 1983). Parameters regarding soil moisture content were also predictably identified as very sensitive in both models. Increases in plant available soil water allow higher ET rates as more water becomes available for transpiration, while decreases in EPCO and ESCO allow more plant water and evaporative water demand to be met from deeper in the soil column – which both ultimately result in lower total water yields as soil moisture decreases (Neitsch et al., 2011). The soil moisture condition II curve number was also very sensitive to calibration, as expected as it largely controls partitioning of rainfall into surface and subsurface flows and is a function of the soil’s antecedent soil water conditions, permeability and land cover (Neitsch et al., 2011).

Surface lag time which controls the fraction of the surface runoff that can enter the stream on any given day was also identified as sensitive in the Hopkins model. This parameter is important for controlling multiday storm-flows in catchments – essentially
lagging the surface runoff by holding it in storage temporarily. Saha et al. (2014), identified CN2, ESCO, GWQMN, SOL_AWC and ALPHA_BF as the five most sensitive parameters in their model at a monthly time-step, while the same parameters were also among the most sensitive parameters at a daily time step. A review of SWAT model applications also consistently identified parameters linked to baseflow and soil moisture as being highly sensitive to calibration (Gassman et al., 2007), supporting our findings and indicating that they should be the focus of calibration to ensure that SWAT models are conceptually consistent with regional hydrological conditions.

SWAT models can be sensitive to the resolution of soil input data due to the spatial aggregation needed to create HRUs, which can ultimately affect sediment and nutrient yields (Romanowicz et al., 2005). While differences in streamflow have been noted in smaller (~125 km²) sub-catchments as a result, the models were still considered satisfactory (Geza and McCray, 2008). There have also been studies that have observed no significant differences in streamflow when using coarse (e.g. 1:250,000) and fine (e.g. 1:12,000 to 1:63,360) soil datasets in catchments on a similar scale to ours (Moriasi and Starks, 2010). In spite of the lack of consensus on the effects of soil resolution on simulation accuracy, there is a general agreement that the use of optimal soil (and other) datasets is important for hydrological modelling. The FAO (2007) soil dataset used here was the only suitable dataset available at the time of model development. New finer-scale data will become available in time but more effort is required to prepare and calibrate such models and therefore the benefits need to be carefully considered (Geza and McCray, 2008).
The creation of independent models for the two catchments allowed an explicit assessment of spatio-temporal hydrologic variability which has been consistently found in the region (Chapters 3, 4 and 5). Previous approaches to regional hydrology (Versace et al., 2008b) have relied on linear regression methods and may have been confounded by other factors such as local geomorphology and climate (Yihdego and Webb, 2013); the SWAT models used here are more complex than the monthly empirical water-balance models than have been used recently to assess LULCC impacts on streamflows (Yihdego and Webb, 2013) and are the next logical step in improving the understanding of regional hydrological conditions. Further work to extend the models to cover the entire Glenelg-Hopkins region would allow for an assessment of management influences and potential effects of LULCC and climate change at a much larger scale.

2.4.2 Monthly, annual, seasonal and drought period flow assessment

The models were able to predict average daily flows at monthly and annual time steps very accurately. Average daily baseflow predictions were also very good at the monthly and annual time steps (Table 2.3) and are therefore suitable for prediction at those time steps. The models had mixed success in predicting seasonal average daily streamflows (Table 2.4, Figure 2.4). Winter and spring streamflows and baseflows were very accurately reproduced, likely due to increased surface runoff to streamflow ratio (i.e. less baseflow relative to streamflow). Summer and autumn flows had a higher baseflow contribution and the models were not able to accurately capture this. The models here were able to accurately model streamflows during drought although average daily flows were underestimated in the Hopkins model and overestimated in the Mount Emu Creek model. These model limitations are unfortunate, given that, in the face of climate
change, streams in the region are likely to become more dependent on baseflows to maintain conditions during extended periods of drought such as the Millennium Drought (CSIRO, 2012). If regional managers are to use similar physically-based models to plan for future streamflows under climate change, a balance between the accuracy of total modelled streamflows and baseflow is needed, depending on the desired management objectives and the hydrology of the relevant systems.

Climate mechanisms (such as the *El Niño* Southern Oscillation) are believed to be the primary driver behind drought in Australia, and it is not uncommon for streams to have consecutive years below long-term average flows (McMahon and Finlayson, 2003). The Millennium Drought was unusual in terms of its severity, duration, geographical location and absence of any intervening wet years (CSIRO, 2012). Additionally, the decline in rainfall occurred mainly in autumn and early winter and resulted in drier soil conditions and less runoff during the historical wetter season (CSIRO, 2012). Previous modelling has shown that, during drought in south-eastern Australia, a 10 % reduction in annual rainfall can result in a 20-30 % decrease in mean annual runoff, while increases in potential ET can cause equivalent reductions in runoff (Teng *et al.*, 2012). The Glenelg-Hopkins region has historically experienced prolonged periods of below-average rainfall and, between 1998 and 2007, the region’s average annual rainfall was 10 % below the long-term average (Department of Sustainability and Environment, 2008a). Streamflows across the region increased subsequently during extended periods of above-average rainfall (2010-2011; CSIRO, 2012). Declines in runoff are predicted to continue under climate change and reductions of 5 % are forecast by 2030 in the Hopkins River (Department of Sustainability and Environment,
A reduction in flows could lead to alteration in current river regulation practices, particularly as river and wetlands are recognised as legitimate ‘users’ of water in Australia (Arthington and Pusey, 2003). However, flow regulation is also a major cause of ecological deterioration in many Australian rivers and as such, any changes in regulation are likely to affect ecological processes and biodiversity (see Arthington and Pusey, 2003, for an extensive set of examples).

The high inter-annual and inter-decadal variability in Australian rainfall streamflow presents particular challenges for management of water resources (CSIRO, 2012). The SWAT models created for this study could be useful for management in the prediction of future streamflow conditions provided accurate datasets of predicted climate variables are available for the period of interest. The availability of downscaled climate model data is thus essential to inform adaptation planning for water resource management (Girvetz et al., 2013), particularly when physically-based models are being employed for the prediction and assessment of streamflows and landscape water balances (Saha et al., 2014).

2.4.3 Assessment of increased plantation forestry on streamflow

Contemporary LULCC has been well studied in the Glenelg-Hopkins region (Ierodiaconou et al., 2005; Versace et al., 2008a; 2008b), but the introduction of Eucalyptus forestry coinciding with a severe long-term drought made an assessment of hydrologic effects difficult. The SWAT models created for this study suggested that streamflows would not have been significantly different in either sub-catchment had the expansion of Eucalyptus forestry not occurred. Previous work in and surrounding the Glenelg-Hopkins region has suggested changes in groundwater and lake water levels.
through LULCC are outweighed by climate variability (Yihdego and Webb, 2011), although impacts have been detected in streamflow records in the 1970s and 1980s as a result of an expansion in the number of farm dams for livestock watering (Yihdego and Webb, 2013). Small decreases in streamflow during the 1990s have also been related to the widespread replacement of grazing land by wheat and canola crops (Yihdego and Webb, 2013). However, in those studies, it was not clear what, if any, influence the expansion of *Eucalyptus* had on contemporary regional streamflows as the changes occurred in unison with an increase in cropping land covers (Yihdego and Webb, 2013).

The lack of difference in streamflow associated with increased forestry may have been a result of the models under-predicting forest growth, given that simulated ET rates were lower than long-term observed averages. Benyon *et al.* (2008) suggests that *Eucalyptus* water use is low in the first two years of a plantation rotation, that surface runoff or groundwater recharge in a new plantation can be up to twice that of the pasture the plantation replaced, and that it can take up to four years for the ET in a well-managed plantation to account for all of the rainfall (i.e. for ET to be greater than or equal to rainfall). Such an effect does not explain the observations in this study, given that our models however would have accounted for this ‘lag’ in water use as they simulated nine full years of rotation (January 2001 – December 2009).

Another possible explanation for the simulated patterns, perhaps more likely here, is that the extent of *Eucalyptus* was simply too small to influence streamflows. Previous work, both locally (Benyon *et al.*, 2008; Sinclair-Knight-Merz, 2008; Yihdego and Webb, 2013) and internationally (Scott and Lesch, 1997; Costa *et al.*, 2003; Fan *et al.*, 2010; Rodriguez Suarez *et al.*, 2014), suggests that LULCC needs to occur on a
much larger scale (relative to catchment size) for the impacts to be readily observed in streamflows and to be distinguishable from climatic variability.

A 22 % reduction in modelled streamflows was identified during the first ten years of the plantation rotation for a small (4 km²) catchment in Spain, where 3.2 km² of grassland was converted to *Eucalyptus*, with differences between modelled streamflows with and without *Eucalyptus* being up to 30 % (Rodriguez Suarez et al., 2014). Likewise, the complete afforestation of a small catchment (< 1 km²) in South Africa with *Eucalyptus* resulted in a statistically-significant decrease in streamflow three years after planting, with the stream drying completely in the ninth year of the rotation (Scott and Lesch, 1997). After clear-felling removal of the eucalypts, it was a further five years before flows returned to the stream and then only as an ephemeral response to the largest storms (Scott and Lesch, 1997). Twenty-five percent of the Crawford River catchment, Australia, which is to the west of our sub-catchments, was converted from grazing land to plantation forestry (Benyon et al., 2008) and a 35 % reduction in the combined long-term mean annual streamflow and groundwater recharge is expected at full canopy cover (Sinclair-Knight-Merz, 2008). In contrast, our modelled plantations were able to reach full canopy maturity with no significant decreases in streamflows further supporting our assertion that LULCC was not extensive enough to influence streamflows.

### 2.5 Conclusions

This study created and applied SWAT models to two sub-catchments in south eastern Australian. Auto-calibration after some initial minor manual calibration resulted in very good model fits to observed data using a range of statistics. Analysis of model outputs under a variety of scenarios, including drought and land-use change,
demonstrated that the models could satisfactorily predict streamflows and baseflows.

Seasonal analysis, however, displayed mixed results with summer and autumn streamflows and baseflows being unsatisfactorily represented in both models. The evaluation of land-cover changes in the region suggested that the introduction of *Eucalyptus* plantation forestry was not extensive enough to influence streamflows in the sub-catchments and that the major reductions that had been observed were likely to be climate related. Further model refinement (e.g. to incorporate nutrient and sediment modelling) could extend the applicability of the models described here to allow examination of further land-cover change scenarios, climate change scenarios, and water quality modelling for nutrients and sediment.
Assessment of spatio-temporally varying relationships between rainfall, land cover and surface water area using geographically weighted regression

3.1 Introduction

3.1.1 Climate change and wetland extents

Climate change is now widely accepted by the scientific community and current research suggests that temperature and rainfall patterns are likely to change dramatically over the next 50 years (IPCC, 2007a, 2014). Extended periods of drought and associated reductions in precipitation, run-off and soil moisture are forecast. Hydrological models suggest groundwater levels and streamflows will be under increasing stress (Lake, 2003). With perennial surface-water habitat being largely dependent on rainfall, an increased understanding of the effects of climate on aquatic habitat availability at regional scales is necessary for the development and improvement of management plans designed to enhance and conserve these habitats.

Global wetland extent is currently estimated at between 5.3 and 12.8 million km$^2$ with about half of the original extent having been lost (O’Connell, 2003; Zedler and Kercher, 2005). These extents however are not reliable as many countries lack comprehensive wetland inventories (Finlayson et al., 1999). Drainage of wetlands for agriculture is perceived to have caused the biggest loss of wetlands to date, with an estimated 26% of global wetland area being drained for intensive agriculture (Zedler and Kercher, 2005). In Asia alone, annual decreases of about 5000 km$^2$ are lost primarily to agriculture and through dam construction (Zedler and Kercher, 2005). Although the available information varies in resolution and spatial extent, the overall trend indicates an indisputable reduction in surface area, condition and associated biodiversity of global
wetlands (Brinson and Malvarez, 2002; Zedler and Kercher, 2005). Perennial wetlands in particular are essential for ecosystem function as biodiversity hotspots (Williams et al., 2003; Dudgeon et al., 2006; Davies et al., 2008), in the biogeochemical cycling of nutrients (Mitsch, 1995; Verhoeven et al., 2006), as refuges across a variety of spatial scales (Magoulick and Kobza, 2003; Kobza et al., 2004; Canepuccia et al., 2007), as well as providing ecosystem services widely benefiting human populations. Unfortunately however, freshwaters ecosystems are amongst the most threatened in the world (Dudgeon et al., 2006; Moilanen et al., 2008). Jensen (1999) wrote that even though Australia is the driest inhabited continent, water is not valued enough to ensure wetlands, rivers, catchments and water supply resources are effectively protected. Brinson and Malvarez (2002) agree and reported that due to the Australian climate, wetlands and anthropogenic activities are in direct competition for water and that, without reductions in human water usage, there will be few opportunities to improve the status of Australian wetlands over the next few decades.

3.1.2 Factors affecting surface water extents

There have been a number of studies on the separate impacts of climate and land-cover changes on hydrology and water quality (Verhoeven et al., 2006; Delpla et al., 2009). However, few studies examine the impacts of climate and land-cover changes on the area of surface-water habitats. There are several factors affecting surface hydrology and surface water availability; rainfall variability (Peel et al., 2004), soil type and infiltration capacity (D’Odorico et al., 2007; Ranatunga et al., 2008; Seneviratne et al., 2010; Li and Sivapalan, 2011), topography (Dirnböck et al., 2002), and vegetation type (Brown et al., 2005; Peel et al., 2010). Vegetation has been shown to affect surface
hydrology, primarily by changing evapotranspiration (Li et al., 2009; Peel et al., 2010),
with temperate forested catchments displaying statistically significantly higher
evapotranspiration than non-forested catchments (Peel et al., 2010). However, the effect
of vegetation on evapotranspiration has been shown to diminish as the area of a
catchment increases (Peel et al., 2010).

3.1.3 Large-scale data acquisition

Advances in data collection using a variety of space-based remote sensors have
permitted the systematic acquisition of land-cover data over large spatial and temporal
scales. The LANDSAT sensor is the longest running of the land-cover sensors and has
been used extensively in the production of land-cover maps, in landscape ecology
research and in detecting land-cover changes (Williams et al., 2006). The sensor provides
cost-effective, high-resolution images (30 m in the visible bands) that are particularly
useful for environmental monitoring due to the 16-day temporal resolution. These
remotely-sensed images coupled with statistical modelling techniques have allowed
scientists and natural resource managers to begin to link pattern to process (Gillanders et
al., 2008). A common methodological framework to achieve this has been the use of
regression techniques in conjunction with classified land-use maps (Versace et al.,
2008b).

3.1.4 Regression modelling for spatio-temporal assessments of change

Traditionally, regression techniques have been limited to methods such as
ordinary least squares (OLS) regression which is subject to some restrictive assumptions
related to normality and variance distributions (Quinn and Keough, 2002). These
assumptions include that residuals are not correlated (i.e. there is no autocorrelation), that they are normally distributed and display homogeneity of variance (i.e. homoscedasticity). According to Tu and Xia (2008), studies concerned with the aquatic environment often violate these assumptions resulting in additional problems such as spatial autocorrelation and spatial non-stationarity. Spatial autocorrelation is a phenomenon where the values for a given variable at location \( x \) are related to the values for the same variable at locations nearby. It is possible that different degrees of spatial autocorrelation can be present within the same dataset and therefore global models to test for spatial autocorrelation would fail to detect it (Fotheringham et al., 2002). Spatial non-stationarity occurs when the relationships between the response and predictor variables are not constant over space (Fotheringham et al., 2002). These issues of spatial autocorrelation and non-stationarity can be a result of model misspecification. For example, using OLS to identify patterns that are known or thought to vary over space, will likely result in significant spatial autocorrelation between residuals as the model will not be able to effectively explain local variations in the relationship.

Geographically weighted regression (GWR) has been developed in an attempt to explore and explain spatially varying relationships, by essentially allowing model parameters to vary over space and thus attempt to overcome some of the restrictive assumptions of OLS regression (Fotheringham et al., 2002). Fotheringham et al. (2002) gives a detailed explanation of the theoretical background behind GWR and explains the applicability of the method to explore spatially varying relationships. The technique has broad applications across a number of fields including health (Nakaya et al., 2005), forestry (Lazarus et al., 2006), aquatic science (Tu and Xia, 2008; Chang and Psaris,
GWR has also been shown to provide better localised prediction results than other techniques (Zhang et al., 2005). The strengths of GWR make it an ideal technique to explore the spatio-temporally varying relationships among rainfall, land cover and area of surface water habitat.

### 3.1.5 Aim of study

The primary purpose of this chapter is to evaluate the suitability of two regression methods (GWR and OLS) to explain spatially-varying data in a predominately dryland catchment in southwest Victoria, Australia. In addition to comparing the two modelling approaches, the results generated will allow an assessment of the influence of rainfall and land-cover changes on surface water extent at the sub-catchment and regional scale. Therefore, the secondary purpose of the chapter is to provide regional natural resource managers with further information that will assist long-term strategic catchment planning. A similar approach was adopted by Tu and Xia (2008), who used their paper not to determine whether any relationships existed between land cover and water quality but to examine whether any interesting spatial variations existed in the relationship among the variables.

### 3.2 Data collection and methods

#### 3.2.1 Study site

The Glenelg-Hopkins region is situated south-west of the Great Dividing Range of eastern Australia and is located in the state of Victoria (Figure 3.1). It covers approximately 27,000 km² and the cities of Warrnambool, Ararat, Hamilton, Portland...
and the western fringes of Ballarat are within its boundary. The region contains the
Grampians Ranges in the north, but is generally a low-lying series of catchments. The
three major drainage basins within the region are the Glenelg, Hopkins and Portland
Coast; drainage across the basins is generally poor, resulting in the formation of many
lakes and wetlands (Glenelg Hopkins CMA, 2006b). The region experiences a
Mediterranean climate characterised by hot dry summers and cool, wet winters. Average
annual rainfall ranges from 500 mm/yr around Lake Bolac to >900 mm/yr in the far
south of the region and the upper reaches within the Grampians Ranges in northern
headwaters of the Glenelg catchment.
Figure 3.1: Location of the Glenelg-Hopkins region in western Victoria. Dark blue lines represent perennial rivers, dark blue areas represent wetland habitat, while the colour gradient represents elevations, with darker browns indicating higher elevations. The three major drainage basins are outlined in red (Glenelg), black (Hopkins) and orange (Portland Coast).
3.2.2 Site description

Agricultural land cover dominates the Glenelg-Hopkins region, with dryland pasture and crops covering approximately 70% of the region. Pine (*Pinus radiata*) and *Eucalyptus* (*Eucalyptus globulus*) plantations are also a dominant feature of the landscape covering about 3% and 5% of the region, respectively, in 2002 (Ierodiaconou *et al.*, 2005). Remnant native vegetation covers approximately 16% of the catchment and is largely accounted for by the Grampians National Park and the Lower Glenelg National Park. Figure 3.2 shows aggregated regional land cover in 2002 based upon the classifications of Ierodiaconou *et al.* (2005). Since European settlement, land cover in the region has undergone dramatic changes (Dixon, 2000; Ierodiaconou *et al.*, 2005; Versace *et al.*, 2008a). Between 1980 and 2002, 16% of the region underwent land-cover transition (Ierodiaconou *et al.*, 2005). The region contains approximately 44% of Victoria’s wetlands and it is thought that, since European settlement, over 75% of the region’s wetlands have been modified by agricultural drainage (Corrick, 1992).
Figure 3.2: Aggregated regional land cover in 2002 (Ierodiaconou et al., 2005). Agricultural land cover dominates the region, while remnant native vegetation and plantation forestry are also widespread.
Climate change scenarios forecast increasing stress on Australia’s scarce freshwater resources (Watson et al., 1997; Pittock et al., 2001). With Australia under conditions of major to severe drought 50% of the time since records began (McKernan, 2005), future climate scenarios indicate a continuing trend of rainfall deficiency. Although the entire study region experiences low inter-annual rainfall variability (Department of Sustainability and Environment, 2008a), reductions in rainfall are expected for the region, with a decrease of 4% by 2030 and 10% by 2070 (Department of Sustainability and Environment, 2008a). Evaporation is expected to increase by 2% by 2030 and by 6% by 2070 (Department of Sustainability and Environment, 2008a). Declines in run-off are predicted to continue under climate change and it has been forecast that reductions of 5 and 30% can be expected by 2030 in the Hopkins and Glenelg Rivers, respectively (Department of Sustainability and Environment, 2008a), with reductions greater than 50% by 2070 (Jones and Durack, 2005).

The region has been identified as a potential ‘food bowl’ and, in the face of climate change, an intensification of agricultural pursuits in the region can be expected. With a strategically changing land cover and serious reductions in runoff predicted as a result of climate change, water and land managers in this region need a greater understanding of where surface water habitats are likely to be affected by future climate and land cover changes.

3.2.3 Sampling sites and land-cover data

Polygons for the catchments were obtained from the Glenelg-Hopkins Catchment Management Authority (GHCMA) and represent salinity management units used within the region and are based on existing sub-catchments within the region.
(Anderson, 2005). Polygons that were smaller than 2.5 km$^2$ were removed from the dataset as they were determined to be artefacts of the digitisation process and were not included in the analyses. Removal of these fragments resulted in 149 sub-catchments being used in the study ranging in size from 2.5 – 716 km$^2$.

Land-cover data for the region between 1980 and 2002 was obtained from Ierodiaconou et al. (2005). Land-cover data were aggregated from 11 Level 1 classes for 1980 and 1995 and 12 Level 1 classes for 2002 to four classes for the analyses. For the analyses in this chapter, the land-cover data were aggregated to minimise issues of collinearity among land cover classes. The four classes were perennial water, plantation forestry (pine and *Eucalyptus*), agriculture (dryland cropping, dryland pasture, irrigated agriculture and irrigated pasture) and remnant native vegetation. Other classes that were excluded from the study included area subject to inundation (for issues of auto-correlation with area of perennial water), and sand and urban areas (due to being a very small component of the catchment). The area of each of the land classes was determined for each sub-catchment for analysis. Figure 3.3 displays regional land cover within the salinity sub-catchments at each observation (1980, 1995 and 2002).

3.2.4 Regional climate data

Monthly rainfall data with a 5-km resolution was obtained for Australia between 1979 and 2002 from the Bureau of Meteorology (Bureau of Meteorology, 2012). Long-term average annual rainfall data between 1961 and 1990 were also obtained (Bureau of Meteorology, 2012). Temperature data were not used as temperatures across the region display very little variation both temporally and spatially. Total rainfall for the 12 months’ prior to each land-cover map was determined by summing the monthly
rainfall rasters for the region. Mean total annual rainfall (for the year prior to each land-
cover map) and long-term (1961 – 1990) total annual average rainfall for each catchment
was then calculated.

To identify patterns of spatial variability in precipitation, a rainfall residual was
calculated. This was done by subtracting the average total annual rainfall from the long-
term average annual rainfall for each polygon. The rainfall residual for each polygon was
then used in the analyses as the climatic variable.

3.2.5 Modelling methods

The dependent variable (area of perennial water) was assessed for normality,
using histograms and skewness and kurtosis coefficients. The raw data for all years
displayed a distinct positive skew and was log(x+1) transformed before the analyses. The
independent variables were not transformed. The GWR analyses were performed using
GWR 3.0 software (Charlton et al., 2003) with outputs then imported to ArcGIS for
further analyses. OLS analyses were performed within ArcGIS 9.3 using the ordinary
least squares tool. Six different models were run for each observation (i.e. 1980, 1995 and
2002), using both OLS and GWR for a total of 36 models. Log(x+1) area of perennial
water was used in all models as the dependent variable.
Figure 3.3: Regional land cover at each observation from Ierodiaconou et al. (2005). The expansion of plantation forestry in the region is highly evident between 1980 and 2002.
The models used were: Rainfall, where rainfall residual was used as the independent variable; Mixed Land Cover, where areas of the three different aggregated land covers were used as the independent variables; TotalAg, RemnVeg, and Plantation, where the area of each respective land-cover class was used as the independent variable and the three land-cover classes were run separately in order to avoid issues of multicollinearity among land-cover variables. This is a similar approach to that presented by Tu and Xia (2008). Finally, an All Variables model was also run, which was a combination of the Rainfall and Mixed Land Cover models.

3.2.6 Modelling background

Traditional regression modelling techniques such as OLS assume that patterns in the data are spatially constant and therefore parameter estimates are the same for the entire study area. The parameter estimates of an OLS model can therefore be considered global statistics and can hide important variations in the spatial distribution and relationship between the independent and dependent variables. OLS is therefore considered a ‘global’ modelling approach. An OLS model can be expressed as:

\[
y = \beta_0 + \sum_{i=1}^{p} \beta_i x_i + e_j
\]

where \(y\) is the dependent variable, \(\beta_0\) is the intercept, \(\beta_i\) is the global parameter estimate (coefficient) for the independent variable \(x_i\), \(p\) is the number of independent variables and \(e\) is the error term.

GWR is an extension to traditional OLS regression techniques that allows local rather than global statistics to be estimated and explored. By calculating local statistics, spatial relationships between the variables in the model can be easily examined and
patterns identified (Fotheringham et al., 2002). As GWR estimates local statistics, it is considered to be a ‘local’ model and is more appropriate to use when relationships are thought to or are known to vary spatially. GWR is an improvement on OLS modelling and is expressed as:

\[ y = \beta_0(u_j, v_j) + \sum_{i=1}^{n} \beta_i(u_j, v_j)x_i + \varepsilon_j \]

where \( y \) is the dependent variable, \( u_j \) and \( v_j \) are the co-ordinates for the observation \( j \), \( \beta_0(u_j, v_j) \) is the intercept for location \( j \), \( \beta_i(u_j, v_j) \) is the local parameter estimate for the independent variable \( x_i \) at location \( j \) and \( \varepsilon \) is the error term (Fotheringham et al., 2002).

GWR employs a weighted distance decay function for model calibration. This assumes that observations closer together will have more impact on each other than on observations further apart. The weighting function for including related samples can be calculated using the exponential distance decay function:

\[ \omega_{ij} = \exp\left(\frac{-d_{ij}^2}{b^2}\right) \]

where \( \omega_{ij} \) is the weight of observation \( j \) for observation \( i \), \( d_{ij} \) is the distance between observation \( i \) and \( j \), and \( b \) is the kernel bandwidth. When the distance between observations is greater than the kernel bandwidth, the weight rapidly approaches zero (Fotheringham et al., 2002). With the GWR software, both fixed and adaptive bandwidths can be chosen. Fixed bandwidth kernel calculates a bandwidth that is held constant over space, whereas the adaptive bandwidth kernel can adapt bandwidth distance in relation to variable density; bandwidths are smaller where data are dense and larger when data are sparse (Fotheringham et al., 2002; Tu and Xia, 2008).
In this study, all GWR models used the adaptive bi-square kernel bandwidth as sample densities varied spatially. The optimal bandwidth distance was determined automatically in GWR 3.0 using the corrected Akaike Information Criterion (AICc). Additionally, a Monte-Carlo significance test was also conducted to test for significant spatial variability in model coefficients.

3.2.7 Comparisons between OLS and GWR model results

A number of tests were conducted to compare the performance of GWR and OLS models. The comparison was performed by comparing R^2 and AICc values among models. In addition to using AICc to calculate an optimal bandwidth distance, the GWR 3.0 software calculates another AICc value which is used for comparisons among different models. Higher R^2 values indicate the model’s ability to explain more variance in the dependent variable as a function of the independent variables. The AICc was used in this instance as a test among models, with smaller values indicating better, more parsimonious results. The AICc is an indicator of model accuracy and complexity where decreases in the AICc value indicate a closer approximation of the model to reality (Quinn and Keough, 2002). Statistically significant model improvements between GWR and OLS models were identified using an approximate likelihood ratio (ALR) test, which is based on the F-test (see Fotheringham et al., 2002, pg. 94). If the results of the ALR test are significant (P ≤ 0.05), then the GWR model is considered a statistically-significant improvement over the OLS model.
3.2.8 Residual analysis and tests for spatial autocorrelation and variance

The standardised residuals of the GWR and OLS models were checked for normality through visual histogram interpretation. Standardised residuals of the OLS and GWR models were also analysed for spatial autocorrelation using Global Moran’s I and local indicators of spatial association (LISA) analysis (Anselin, 1995). Global Moran’s I values can range from -1 to 1. A value of 1 indicates perfect spatial autocorrelation where high values, or low values, cluster together. A value of -1 indicates perfect negative spatial autocorrelation with values representing a checkerboard (Tu and Xia, 2008). A value of 0 indicates perfect random spatial variability. LISA measures the degree of local spatial autocorrelation at each sampling point by using a localised Moran’s I. Global Moran’s I and LISA were calculated using GeoDa 0.9.5-i (Beta) analysis software (Anselin et al., 2006). Results of the Monte-Carlo significance test were included in the GWR output to identify statistically-significant spatial variation in the model variables. This was used in conjunction with the results of the ALR test as an indicator of the applicability of GWR to improve parameter estimates over OLS. As one of the assumptions of OLS is that parameters are constant over space, a deviation from this condition would suggest that OLS will not be a good predictor under these circumstances.

3.3 Results

3.3.1 Comparisons between OLS and GWR modelling approaches

Improvements in both $R^2$ and AIC$_C$ were observed for GWR models over OLS counterparts for all models used in the study (Table 3.1). All OLS models displayed non-normal residuals, while six of 18 GWR models displayed non-normal residuals. All
GWR models displayed higher R² values than the analogous OLS models. Additionally, all GWR models, with the exception of the 1980 rainfall model, had smaller AICc values by at least 3. All of the OLS models displayed significant (P ≤ 0.05) global spatial autocorrelation, while only five of 18 GWR models displayed significant global spatial autocorrelation (Table 3.1). Coefficients between models and model types displayed substantial variation (Table 3.2). In some cases, an order of magnitude difference was found between the OLS model coefficients and the median GWR coefficients for the same model. This was also observed for model R² values (Table 3.2). Mixes of negative and positive coefficients were also seen across models, model types and years. The results of the Monte-Carlo significance test showed that, for the 1980 models, all independent variables, with the exception of area of plantation in the Plantation, Mixed and All Variable models, displayed non-significant spatial variability. The results for the 1995 models showed that all variables in the Mixed and All Variables models, area of plantation in the Plantation model and rainfall residual in the Rainfall model all displayed significant (P ≤ 0.05) spatial variation. The results of the 2002 models showed that area of remnant native vegetation in the RemnVeg, Mixed and All Variable models, area of plantation in the Plantation, Mixed and All Variable models, and the area of agriculture in the All Variable model all displayed significant (P ≤ 0.05) spatial variation. Statistically-significant improvements of parameter estimates by GWR modelling was supported by the results of the ALR test which demonstrated that all GWR models, with the exception of the 1980 Rainfall model, were statistically-significant (P ≤ 0.05) improvements over comparative OLS models (Table 3.1).
Table 3.1: Results of OLS and GWR analyses. Analysis of model performance, indicated that with the exception of one model (1980, Rainfall) all GWR models outperformed the analogous OLS model. Adjusted R² values for GWR models are mean values, * indicates statistically-significant ($P \leq 0.05$) residual global spatial autocorrelation, # indicates non-significant ($P > 0.05$) model improvement over OLS.

<table>
<thead>
<tr>
<th>Model</th>
<th>Year</th>
<th>Type</th>
<th>Model</th>
<th>Adj. R²</th>
<th>Moran’s I</th>
<th>ALR Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Agriculture</td>
<td>1980</td>
<td>OLS</td>
<td>621.43</td>
<td>0.15</td>
<td>*0.23</td>
<td>3.82</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GWR</td>
<td>599.37</td>
<td>0.43</td>
<td>-0.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1995</td>
<td>OLS</td>
<td>638.83</td>
<td>0.13</td>
<td>*0.24</td>
<td>3.85</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GWR</td>
<td>616.21</td>
<td>0.41</td>
<td>-0.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2002</td>
<td>OLS</td>
<td>566.55</td>
<td>0.27</td>
<td>*0.15</td>
<td>3.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GWR</td>
<td>552.06</td>
<td>0.49</td>
<td>-0.06</td>
<td></td>
</tr>
<tr>
<td>Remnant Native Vegetation</td>
<td>1980</td>
<td>OLS</td>
<td>636.27</td>
<td>0.05</td>
<td>*0.24</td>
<td>6.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GWR</td>
<td>615.30</td>
<td>0.22</td>
<td>*0.09</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1995</td>
<td>OLS</td>
<td>650.74</td>
<td>0.05</td>
<td>*0.24</td>
<td>7.38</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GWR</td>
<td>523.75</td>
<td>0.25</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2002</td>
<td>OLS</td>
<td>610.63</td>
<td>0.02</td>
<td>*0.24</td>
<td>4.89</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GWR</td>
<td>588.24</td>
<td>0.22</td>
<td>*0.08</td>
<td></td>
</tr>
<tr>
<td>Plantation</td>
<td>1980</td>
<td>OLS</td>
<td>646.58</td>
<td>-0.01</td>
<td>*0.20</td>
<td>6.57</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GWR</td>
<td>628.53</td>
<td>0.14</td>
<td>*0.11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1995</td>
<td>OLS</td>
<td>661.07</td>
<td>-0.01</td>
<td>*0.19</td>
<td>5.91</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GWR</td>
<td>640.77</td>
<td>0.15</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2002</td>
<td>OLS</td>
<td>614.80</td>
<td>0.00</td>
<td>*0.19</td>
<td>4.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GWR</td>
<td>591.26</td>
<td>0.30</td>
<td>-0.03</td>
<td></td>
</tr>
<tr>
<td>Mixed Land Cover</td>
<td>1980</td>
<td>OLS</td>
<td>614.64</td>
<td>0.20</td>
<td>*0.21</td>
<td>5.72</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GWR</td>
<td>594.37</td>
<td>0.34</td>
<td>*0.11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1995</td>
<td>OLS</td>
<td>631.47</td>
<td>0.18</td>
<td>*0.23</td>
<td>3.46</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GWR</td>
<td>618.18</td>
<td>0.56</td>
<td>-0.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2002</td>
<td>OLS</td>
<td>561.25</td>
<td>0.31</td>
<td>*0.16</td>
<td>3.43</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GWR</td>
<td>546.06</td>
<td>0.54</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Rainfall</td>
<td>1980</td>
<td>OLS</td>
<td>638.92</td>
<td>0.04</td>
<td>*0.13</td>
<td>2.43</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GWR</td>
<td>637.83</td>
<td>0.26</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1995</td>
<td>OLS</td>
<td>660.96</td>
<td>-0.01</td>
<td>*0.19</td>
<td>3.21</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GWR</td>
<td>649.77</td>
<td>0.40</td>
<td>-0.08</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2002</td>
<td>OLS</td>
<td>615.62</td>
<td>-0.01</td>
<td>0.17</td>
<td>4.77</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GWR</td>
<td>599.09</td>
<td>0.15</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>All Variables</td>
<td>1980</td>
<td>OLS</td>
<td>607.90</td>
<td>0.24</td>
<td>*0.15</td>
<td>3.47</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GWR</td>
<td>600.37</td>
<td>0.31</td>
<td>*0.12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1995</td>
<td>OLS</td>
<td>633.59</td>
<td>0.18</td>
<td>*0.23</td>
<td>3.41</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GWR</td>
<td>621.39</td>
<td>0.55</td>
<td>-0.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2002</td>
<td>OLS</td>
<td>562.37</td>
<td>0.31</td>
<td>*0.17</td>
<td>3.45</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GWR</td>
<td>555.86</td>
<td>0.65</td>
<td>-0.02</td>
<td></td>
</tr>
</tbody>
</table>
3.3.2 Results of GWR modelling

The single land-cover models (TotalAg, RemnVeg and Plantation) all appear to effectively explain spatial differences in the relationships between the dependent and independent variables, with higher $R^2$ values observed where each independent land cover dominates the landscape (Figure 3.4). The ability of the respective land-cover models to explain the relationship drops off in areas dominated by other land covers (Figure 3.4). Rainfall residual was not as effective at explaining the spatial variation in surface water area, with relatively poor $R^2$ values when compared to the other models (Table 3.1, Figure 3.4). The results of the Mixed and All Variable models were very similar with regard to the strength of the relationship and the spatial distribution of the $R^2$ values (Table 3.1, Figure 3.4).

Area of Total Agriculture displayed a mix of both negative and positive parameter coefficients with negative coefficients being observed in five of the nine models in which it was a variable, suggesting that, depending on the catchment, an increase (or decrease) in agriculture can result in smaller (or larger) wetland extents. Remnant Native Vegetation displayed positive coefficients for all three years of the RemnVeg model. This suggests that, as the area of remnant vegetation within a catchment increases, so will the extent of surface water. However, area of remnant vegetation displayed negative coefficients in the 1995 and 2002 Mixed Land Cover and All Variables models (Table 3.2, Figure 3.5), contrasting the other single variable RemnVeg models, and suggesting that an increase in remnant vegetation will result in a reduced surface water extent in the presence of other land covers. Area of Plantation was shown to exhibit negative coefficients across all years of all models, with the single
exception of the 2002 Plantation model, while rainfall residual displayed negative coefficients across all models and all years (Table 3.2).

Results of LISA analysis of the residuals for the GWR models demonstrate very little local spatial autocorrelation (Figure 3.6). Temporally and spatially, the results for all models show that GWR is much better at accounting for spatial non-stationarity of variables than OLS. This is displayed through very minimal statistically-significant clustering (i.e. residuals are not clustered with other comparative residuals).
Table 3.2: Coefficient estimates for the independent variables in each of the models. The Min, Med, and Max coefficients are for the GWR models. The GWR models displayed a mix of positive and negative model coefficients indicating that there was a spatially-varying relationship between the independent and dependent variables. Rainfall res. = rainfall residual; Remn. Veg. = area of remnant vegetation; Total Ag. = area of agriculture; Total plant = area of plantation.

<table>
<thead>
<tr>
<th>Model</th>
<th>Year</th>
<th>Variable</th>
<th>OLS</th>
<th>Min.</th>
<th>Med.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Agriculture</td>
<td>1980</td>
<td>Total Ag.</td>
<td>0.00008</td>
<td>-0.00002</td>
<td>0.00007</td>
<td>0.00019</td>
</tr>
<tr>
<td></td>
<td>1995</td>
<td>Total Ag.</td>
<td>0.00007</td>
<td>-0.00004</td>
<td>0.00006</td>
<td>0.00019</td>
</tr>
<tr>
<td></td>
<td>2002</td>
<td>Total Ag.</td>
<td>0.00009</td>
<td>0.00002</td>
<td>0.00010</td>
<td>0.00023</td>
</tr>
<tr>
<td>Remnant Native</td>
<td>1980</td>
<td>Remn Veg.</td>
<td>0.00009</td>
<td>0.00005</td>
<td>0.00012</td>
<td>0.00021</td>
</tr>
<tr>
<td>Vegetation</td>
<td>1995</td>
<td>Remn Veg.</td>
<td>0.00010</td>
<td>0.00006</td>
<td>0.00013</td>
<td>0.00022</td>
</tr>
<tr>
<td></td>
<td>2002</td>
<td>Remn Veg.</td>
<td>0.00007</td>
<td>0.00004</td>
<td>0.00011</td>
<td>0.000133</td>
</tr>
<tr>
<td>Plantation</td>
<td>1980</td>
<td>Total plant.</td>
<td>0.00009</td>
<td>-0.00444</td>
<td>0.00057</td>
<td>0.01458</td>
</tr>
<tr>
<td></td>
<td>1995</td>
<td>Total plant.</td>
<td>0.00007</td>
<td>-0.00339</td>
<td>0.00040</td>
<td>0.04794</td>
</tr>
<tr>
<td></td>
<td>2002</td>
<td>Total plant.</td>
<td>0.00008</td>
<td>0.00011</td>
<td>0.00036</td>
<td>0.00556</td>
</tr>
<tr>
<td>Mixed Land Cover</td>
<td>1980</td>
<td>Total Ag.</td>
<td>0.00007</td>
<td>0.00004</td>
<td>0.00013</td>
<td>0.00019</td>
</tr>
<tr>
<td></td>
<td>1995</td>
<td>Remn Veg.</td>
<td>0.00010</td>
<td>0.00002</td>
<td>0.00006</td>
<td>0.00010</td>
</tr>
<tr>
<td></td>
<td>2002</td>
<td>Total plant.</td>
<td>-0.0023</td>
<td>-0.00322</td>
<td>-0.0010</td>
<td>0.00511</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total Ag.</td>
<td>0.00007</td>
<td>-0.00005</td>
<td>0.00007</td>
<td>0.00021</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Remn Veg.</td>
<td>0.00011</td>
<td>-0.00888</td>
<td>0.00006</td>
<td>0.00084</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total plant.</td>
<td>-0.0021</td>
<td>-2.42545</td>
<td>-0.0092</td>
<td>0.17143</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total Ag.</td>
<td>0.00010</td>
<td>0.00002</td>
<td>0.00009</td>
<td>0.00017</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Remn Veg.</td>
<td>0.00008</td>
<td>-0.00209</td>
<td>0.00004</td>
<td>0.00021</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total plant.</td>
<td>-0.0008</td>
<td>-0.00061</td>
<td>0.00014</td>
<td>0.00208</td>
</tr>
<tr>
<td>Rainfall</td>
<td>1980</td>
<td>Rainfall res.</td>
<td>-0.01126</td>
<td>-0.04331</td>
<td>0.00382</td>
<td>0.03214</td>
</tr>
<tr>
<td></td>
<td>1995</td>
<td>Rainfall res.</td>
<td>-0.00174</td>
<td>-0.23333</td>
<td>0.00478</td>
<td>0.14215</td>
</tr>
<tr>
<td></td>
<td>2002</td>
<td>Rainfall res.</td>
<td>0.00248</td>
<td>-0.00818</td>
<td>0.00808</td>
<td>0.01955</td>
</tr>
<tr>
<td>All variables</td>
<td>1980</td>
<td>Total Ag.</td>
<td>0.00007</td>
<td>0.00004</td>
<td>0.00006</td>
<td>0.00010</td>
</tr>
<tr>
<td></td>
<td>1995</td>
<td>Remn Veg.</td>
<td>0.00011</td>
<td>0.00003</td>
<td>0.00013</td>
<td>0.00018</td>
</tr>
<tr>
<td></td>
<td>2002</td>
<td>Total plant.</td>
<td>-0.00005</td>
<td>-0.00266</td>
<td>-0.00007</td>
<td>0.00531</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rainfall res.</td>
<td>-0.01139</td>
<td>-0.01190</td>
<td>-0.00686</td>
<td>0.00360</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total Ag.</td>
<td>0.00007</td>
<td>-0.00006</td>
<td>0.00007</td>
<td>0.00020</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Remn Veg.</td>
<td>0.00011</td>
<td>-0.00308</td>
<td>0.00008</td>
<td>0.00045</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total plant.</td>
<td>-0.00023</td>
<td>-1.68390</td>
<td>-0.00142</td>
<td>0.06871</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rainfall res.</td>
<td>0.00086</td>
<td>-0.07777</td>
<td>0.00721</td>
<td>0.03551</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total Ag.</td>
<td>0.00010</td>
<td>-0.00001</td>
<td>0.00009</td>
<td>0.00020</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Remn Veg.</td>
<td>0.00009</td>
<td>-0.00306</td>
<td>0.00006</td>
<td>0.00058</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total plant.</td>
<td>-0.00010</td>
<td>-0.00143</td>
<td>0.00019</td>
<td>0.00370</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rainfall res.</td>
<td>0.00331</td>
<td>-0.02789</td>
<td>0.00796</td>
<td>0.04365</td>
</tr>
</tbody>
</table>
Figure 3.4: Results of the GWR models showing local $R^2$ values for each of the models. Significant variability in the explanatory power of the relationship is obvious from local $R^2$ results. Rows in the image represent years, while columns show the individual models.
Figure 3.5: Spatial distribution of regression coefficients for the 2002 All Variables model. Significant spatial differences in coefficients were observed across the region, suggesting that the relationship between variables in the model had significant spatial variability which would have been erroneously modelled with OLS regression.
Figure 3.6: Results of the LISA analysis showing localised spatial autocorrelation of GWR model residuals. Minimal spatial clustering of residuals indicated that the GWR models were able to effectively explain the spatial variance in the relationship between the independent and dependent variables. Rows in the image represent years, while columns show the individual models. The blank catchments are representative of no significant spatial autocorrelation.
3.4 Discussion

3.4.1 Model performance and interpretation

The higher $R^2$ and lower AICC values associated with the GWR models support previous research suggesting that GWR is better at explaining spatially-varying relationships than OLS (Fotheringham et al., 2002; Zhang et al., 2005; Tu and Xia, 2008). The distribution of the $R^2$ values (Figure 3.4) confirms our assertion that there is a spatially-varying relationship between land cover and the area of surface water within the region. The increased explanatory power of GWR was confirmed by the results of the ALR test, which showed statistically-significant ($P \leq 0.05$) improvements over analogous OLS models. Interpretation of residual histograms also supported the ability of GWR to better model spatially-varying data with six of 18 models displaying non-normal residuals compared to all OLS models displaying non-normal residuals. Global and local residual analysis also confirmed that GWR is a better predictor of spatially-varying relationships, with five of 18 GWR models displaying significant global spatial autocorrelation compared to 17 of the 18 OLS models (Table 3.1). The results of LISA analysis demonstrated the ability of GWR to better model spatially-varying data with very minimal clustering of residuals (Figure 3.6) indicating that the GWR models are not over- or underestimating the magnitude of the dependent variable in a spatially-correlated fashion (Zhang et al., 2005). Furthermore, the assumptions of OLS regression were not met and therefore the validity of the models is questionable. When OLS assumptions are violated, regression efficiency is reduced and model results can be misinterpreted (Quinn and Keough, 2002). In particular, the spatial autocorrelation issues of the OLS analyses (Table 3.1) severely limit the inferences that can be drawn.
from the analyses. OLS assumes the relationship to be stationary, but resulted in spatially-correlated residuals, indicating that the relationship is not stationary. In contrast, the application of GWR reduced global spatial autocorrelation of residuals and displayed minimal local spatial autocorrelation and therefore represents a statistically more reliable model. The comparatively poor performance of the OLS models (non-normal residuals, relatively low $R^2$, spatial autocorrelation of residuals) compared to the GWR results further reinforces the utilisation of new regression techniques such as GWR when investigating relationships that are believed or known to vary spatially.

The processes affecting the distribution of surface water within a landscape were a combination of soil and water interactions (Castillo et al., 2003; D’Odorico et al., 2007; Li and Sivapalan, 2011), vegetation (Brown et al., 2005; Li et al., 2009; Peel et al., 2010), topography (Dirnböck et al., 2002) and climate (Bronstert et al., 2002). Within our study area, we found there was a spatio-temporally varying relationship between land cover and the extent of surface water within the region. The performance of the Mixed Land Cover model and the All Variables model were very similar in terms of the spatial and temporal distribution of $R^2$ values (Figure 3.4). This suggests that, in this instance, land cover is capable of explaining the majority of the relationship between area of surface water and land-cover type, independent of rainfall. While this seems counter-intuitive, as rainfall is the primary driver of the hydrological cycle, the history of agricultural development in the study area can offer an explanation of this result. Specifically, the results of the Mixed Land Cover models suggest that the surface area of water in the landscape is likely related to anthropogenic drainage practices which enhance runoff to facilitate intensive agriculture and the decreased runoff associated
with increased plantation forestry (Turner et al., 2004; Van Dijk et al., 2004; Brown et al., 2005; Peel et al., 2010). These factors (land-cover changes and associated changes in drainage and runoff) have potentially contributed to the available rainfall being less likely to be able to maintain existing surface-water habitat. Furthermore, the model also suggests that rainfall variability has had less influence in the reduction of the area of surface water within the region than anthropogenic drainage has impacted these areas. However, as rainfall is an essential part of the hydrological cycle, the results of the All Variables model cannot be ignored. The coefficients for the 2002 All Variables model (Figure 3.5) demonstrate that there was significant variation in model coefficients depending on the variable of interest and the spatial location of the sub-catchment. The All Variables model coefficients displayed similar distributions to the individual land cover (TotalAl, RemnVeg and Plantation) model $R^2$ with higher coefficients observed where a particular land cover dominates (Figures 3.4 and 3.5).

Agricultural land cover dominates throughout the Hopkins Basin and therefore the highest coefficients are observed throughout that basin for area of total agriculture. Similarly, as remnant native vegetation dominates the Grampians National Park in the north-central area of the region, the highest coefficients are seen in this area for area of remnant vegetation. It is noteworthy that high coefficients were generated in the Hopkins Basin for area of plantation. This is interesting because, whilst *Eucalyptus* forestry expansion was a region-wide land-use change (Ierodiaconou et al., 2005; Versace et al., 2008a), the ability of area of plantation to explain surface water conditions is greatest in the Hopkins Basin where the introduction of *Eucalyptus* plantations was at a smaller scale than that observed in the other regional basins (Ierodiaconou et al., 2005),
and is particularly interesting as the area of pine plantations within the area of the high coefficients is limited. However, the size of the coefficients output by the analyses (Table 3.2, Figure 3.5) suggests that there were other factors controlling the distribution and area of surface water within the region. These are likely, soil type/infiltration capacity and topographical variables such as slope or topographic wetness index.

The presence of negative coefficients within the All Variables model for both area of remnant native vegetation and area of plantation is expected as previous research has shown that both of these land cover types reduce run-off compared to agricultural land (Zhang et al., 2001; Van Dijk et al., 2004; Brown et al., 2005; Peel et al., 2010) and hence, negatively affect surface water accumulation. What was not expected however, was the presence of negative model coefficients for rainfall residual.

Rainfall was shown to have a number of negative coefficients across models, years and model types (Table 3.2, Figure 3.5). This is counter-intuitive as it suggests that in areas where there is less rainfall (i.e. a greater rainfall residual) there will be greater surface water extents. We suspect this may be a result of using either rainfall residual as the rainfall variable or timing of the LANDSAT image capture. The use of total annual rainfall, as opposed to rainfall residual, should be investigated in future research. The timing of the LANDSAT image capture could have affected model results with the use of a rainfall residual. If there had been a year of below-average annual rainfall but the months leading up to the image capture were relatively wet, there would be a lot of surface water across the catchment. Instances like this would limit the ability of a rainfall residual to accurately explain surface water differences between catchments.
3.4.2 Regional changes in land cover and drainage

Since European settlement, large-scale changes in the landscape have occurred within the region (Ierodiaconou et al., 2005) and it is estimated that 75% of the region’s wetlands have been severely modified by agricultural drainage (Corrick, 1992). Contemporary changes (1995 - 2002) in land cover have largely been associated with the widespread adoption of dryland cropping (~6.5% increase) and the development of Eucalyptus plantation forestry (~5% increase; Ierodiaconou et al., 2005). Transition analysis by Versace et al. (2008a) reported that expansions in cropping and forestry land cover did not affect the distribution or percentage cover of remnant native vegetation within the region, and that systematic gains in dryland cropping and Eucalyptus forestry occurred at the expense of dryland pasture.

A partial explanation of the expansion of dryland crops in the region was the decline in profitability of traditional dryland pastures and sheep grazing and the development of raised-bed cropping technologies which allowed land owners to crop in previously waterlogged soils (Versace et al., 2008a). Consequently, the expansion of dryland cropping within the region may have contributed to altering the hydrological dynamics of the region by affecting runoff and decreasing the amount of waterlogged soils. The pastures that dryland cropping has replaced were largely poorly drained and, during times of high rainfall, existed on waterlogged soils. As a result there was likely more surface water habitat, albeit ephemeral, found before the widespread introduction of raised bed crops. The introduction of Eucalyptus plantation forestry in the region between 1995 and 2002, at the expense of dryland pasture, was a response to economic and environmental conditions by both farmers and timber companies (Ierodiaconou et
Runoff across the region has decreased and may be partially attributed to increased water consumption by deep-rooted woody vegetation as a result of greater rain interception and deeper root systems than perennial pastures and non-irrigated agricultural crops (Benyon et al., 2006; Benyon et al., 2008; Benyon and Doody, 2014). Turner et al. (2004) compared mean annual runoff between grassland and native *Eucalyptus* forests from two rainfall zones of 800 mm and 1200 mm. The comparisons showed that grassland had a mean annual runoff of 210 mm and 493 mm, while the *Eucalyptus* forests had mean annual runoff of 45 mm and 228 mm from the respective rainfall zones. As much of the study region has annual rainfall below 800 mm, it is not inconceivable that the expansion of *Eucalyptus* forestry within the region has affected runoff (but see Chapter 2). Afforestation has also been shown to exhibit sometimes severe impacts on water resources (Zhang et al., 2001; Van Dijk et al., 2004). Van Dijk et al. (2004) reported that for an 800 mm rainfall zone, land-cover transition from perennial pastures to forestry (either plantation or revegetation) resulted in an average water yield reduction of about 1.5 ML for each hectare planted.

Many long-term land-use management decisions can be very sensitive to changes in physical climate conditions and there is an awareness that many decisions already occurring need to take long-term climate change into account (Hallegatte, 2009). From an integrated water resources management perspective, those in charge of maintaining resources need to take into account the effects of land-cover changes that can drastically alter catchment hydrology and prepare for consequences that may not be observed for some time. However, the influence of rainfall and temperature variability should be assessed independently when quantifying the hydrological effect of land-cover
changes (Li et al., 2009). Meinke and Stone (2005) propose that changes in agricultural industries (whether crops or pastures) occur on inter-decadal scales (10 – 20 years) with broader land-use changes (e.g. agricultural or natural systems) occurring on multi-decadal scales (> 20 years). These decisions, particularly those related to water infrastructure and land-use planning have consequences over 50 – 200 years (Hallegatte, 2009). Whilst Eucalyptus forestry is not likely to expand in the future due to prevailing economic factors; quantifying its expansion from the 2002 extent could provide further evidence of increased plantation forestry severely impacting regional hydrological dynamics and further limiting the ability of rainfall to maintain the ever-decreasing surface water habitat within the region. There is anecdotal evidence that land-cover changes in the region post-2002 have seen the further southward expansion of dryland cropping as a result of observed and expected rainfall changes within the region, coupled with economic drivers beyond the region. With wetland systems continuing to recede and with small changes in climate expected within the region (Department of Sustainability and Environment, 2008a), empirical evidence gained from this study suggests that the declining state of regional wetlands is linked to the dynamic nature of regional land cover and the associated hydrological changes.

3.5 Conclusions

Over the next 50 years, the climate of the world is expected to change quite drastically as a whole. In some regions however, climate changes are not expected to be as severe and other factors may pose a more immediate, but arguably manageable, threat to aquatic ecosystems. This study has demonstrated the superior ability of GWR to model spatially-varying relationships over OLS regression. Studies concerned with any
form of spatial analyses need to take the limitations of OLS and other similar linear regression methods into consideration and investigate newer, more suitable methods when attempting to explain spatial relationships. We also demonstrated, through the application of GWR to historical land-cover and rainfall data, that land-cover change can influence surface water area within a catchment, however the effects of land-cover changes could not be quantified as some key variables on soil type/infiltration capacity and topographical variables such as slope were absent from our analyses. Future work will include these variables as they become more readily available and accurate.

Management agencies have a responsibility to ensure that they are aware of the impacts of these changes on the resources for which they are responsible and methods like those presented here may do a great deal in increasing that understanding. While planning for the future should no doubt include the possibilities of a changing climate seriously affecting water resources, changes in land cover cannot be underestimated in their ability to alter topography, runoff and drainage at catchment scales.
Getting down and dirty – can soil attributes help to quantify a spatially-varying relationship between rainfall, land cover and wetland extents?

4.1 Introduction

Previous modelling has found that there was significant spatial variability in the relationship between wetland extent, land cover and climate in the Glenelg-Hopkins region (Chapter 3). The identified spatial non-stationarity of model coefficients violated the assumptions of traditional OLS regression and a newer regression technique, geographically weighted regression (GWR; Fotheringham et al., 2002), was shown to greatly improve model performance compared to OLS in terms of higher R^2 and lower corrected Akaike Information Criterion (AIC_c). The GWR models suggested that the area of surface water in the landscape was related to the amount of agriculture and plantation forestry within a catchment, which could be interpreted as anthropogenic drainage practices enhancing runoff to facilitate intensive agriculture and increased plantation forestry. However, with some key soil variables not able to be included in the previous analysis, the validity of this conclusion could not be assessed.

Spatial differences in soils, and in particular soil moisture, have been shown to affect runoff responses in semi-arid catchments (Castillo et al., 2003) and affect the distribution of vegetation (D’Odorico et al., 2007), while saturated soil hydraulic conductivity has been linked to spatial heterogeneity in runoff processes at a range of scales (Li and Sivapalan, 2011). The development of fine-scale spatial databases of soil parameters is continuing to improve hydrologic modelling efforts and some models have been shown to be incredibly sensitive to the quality of the soil data used in their parameterisation (Romanowicz et al., 2005).
The purpose of this short chapter is therefore to build on previous analyses to investigate the ability of a range of hydrologic soil properties (Western and McKenzie, 2006) that were not able to be included in the previous analysis (Chapter 3), and to assess previous conclusions regarding the observed spatial relationships.

4.2 Methods

4.2.1 Model design

Following Chapter 3, there were 149 sub-catchments used in this study. Seven new variables related to soil hydrology were included in this analysis, in addition to the variables included from the previous chapter (see below). The mean saturated conductivity of the A and B horizons, the plant available water capacity of the A and B horizons, the thickness of the A and B horizons and the total depth of the soil profile were determined for each of the 149 sub-catchments from a 1-km resolution spatial database (Western and McKenzie, 2006). The GWR analyses were performed using GWR 4.0 (Nakaya, 2014a) which has improvements over GWR 3.0 which was used previously. These improvements include corrections to the calculation of local diagnostic statistics, including local sigma and local $R^2$, and a method for automatically modelling variables as either locally variable or globally constant (Nakaya, 2014b). Models that were previously created in GWR 3.0 (i.e. those in Chapter 3) were rebuilt for this study using GWR 4.0 to permit direct comparisons between model $AIC_c$ and localised $R^2$ values. All models used $\log(x+1)$ area of perennial water as the dependent variable. In addition to the soil parameters, the independent variables were the area of remnant vegetation, the area of plantation forestry, the area of agriculture, and the mean rainfall residual for each catchment (Chapter 3).
Six semi-parametric GWR models (one with and one without the soil parameters for each of 1980, 1995 and 2002) were created to analyse the impact of soil variables on the previously-modelled relationships. For consistency with Chapter 3, models without soil parameters are referred to as ‘All variables’ and models with soil parameters are referred to as ‘All variables + soil’. No stepwise variable selection was conducted and all variables were included in the respective models. Semi-parametric GWR models are similar to partial linear models or mixed models in that some parameters were allowed to vary spatially, while others were modelled with global coefficients (Nakaya et al., 2009). A geographic variability test was used to assess whether independent variable coefficients varied spatially (Nakaya, 2014b). If variables were considered to not show significant spatial variability (relative to the full GWR model where all variables are considered spatially varying), they were automatically modelled with global (as opposed to local) coefficients. GWR models employ a weighted-distance decay function that assumes that observations closer together will have more impact on each other than on observations further apart (Fotheringham et al., 2002). The optimum bandwidth for this decay function was chosen automatically in GWR 4.0, with all models utilising the adaptive bi-square kernel bandwidth as sample densities varied spatially (Nakaya, 2014b).

4.2.2 Model comparisons

GWR 4.0 calculates an $\text{AIC}_c$ value which is used for comparisons among different models. $\text{AIC}_c$ is an indicator of model accuracy and complexity where decreases in $\text{AIC}_c$ of 3 or more indicate a better model (Quinn and Keough, 2002). As previous modelling (Chapter 3) identified that all the GWR models were significant
improvements over analogous OLS models, no comparison between OLS and GWR regression was attempted in this study. Comparisons between models with and without the soil parameters were based on the observed decrease in $AIC_c$ value and Wilcoxon signed-rank tests between the local $R^2$ coefficients.

4.3 Results

The addition of the soil parameters to the relationship resulted in improved models for 1980 and 2002 where decreases of 3 and 8 in $AIC_c$ were observed respectively (Table 4.1). No change in the $AIC_c$ for the 1995 model suggested that the addition of the soil parameters contributed little to the relationship. This was a rather unexpected result and suggests that spatial variability in soil parameters did not have a strong control on the distribution and extent of surface water habitat across the region, at least for the 1995 time point. However, all models that included soil parameters demonstrated significant ($P \leq 0.001$) improvements in local $R^2$ (Figure 4.1) according to the results of the Wilcoxon signed-rank tests. Unlike the decrease in $AIC_c$, this was to be expected as more independent variables typically increase the ability of regression models to explain variance in a dependent variable and does not necessarily indicate a significant control of surface water habitat by soil parameters in the region. The pattern of improved models across two of the three modelled years supports the notion that there is a significant control of surface water habitat by soil parameters.
Table 4.1: The results of GWR modelling and model comparisons for the new models which contained information on hydrologically-relevant soil parameters. According to AICc, the addition of soil parameters improved the model fit of the 1980 and 2002 models. The global variables did not display significant local variation and therefore not modelled with locally-varying coefficients. The final column is the median coefficient for each variable that was modelled as locally varying for each of the GWR models. Abbreviations of variables included are as follows: KSAT = hydraulic conductivity of the respective soil horizon (i.e. A or B); PAWC = plant available water of the respective soil horizon; Rainfall Res. = rainfall residual; Remn. Veg. = area of remnant vegetation; Soil Depth = total depth of soil profile; Thick = thickness of the respective soil horizon; Total Ag. = area of agriculture; Total Plant. = area of plantation.

<table>
<thead>
<tr>
<th>Model</th>
<th>Year</th>
<th>AICc</th>
<th>Mean Adj.R²</th>
<th>Global variables</th>
<th>t-stat</th>
<th>Local variables</th>
<th>Median local coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>All variables</td>
<td>1980</td>
<td>587</td>
<td>0.41</td>
<td>Total Plant.</td>
<td>1.04</td>
<td>Remn. Veg</td>
<td>0.00007</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Rainfall Res.</td>
<td>1.63</td>
<td>Total Ag.</td>
<td>0.00006</td>
</tr>
<tr>
<td>All variables + soil</td>
<td>1980</td>
<td>584</td>
<td>0.57</td>
<td>Total Plant.</td>
<td>0.69</td>
<td>Remn. Veg</td>
<td>0.00007</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Rainfall Res.</td>
<td>0.79</td>
<td>Total Ag.</td>
<td>0.00007</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>KSAT [A]</td>
<td>0.44</td>
<td>PAWC [A]</td>
<td>0.03139</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Soil Depth</td>
<td>0.04</td>
<td>Thick [A]</td>
<td>-12.80444</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>KSAT [B]</td>
<td>0.00118</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PAWC [B]</td>
<td>0.00584</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Thick [B]</td>
<td>-5.49447</td>
</tr>
<tr>
<td>All variables</td>
<td>1995</td>
<td>596</td>
<td>0.44</td>
<td>Total Plant.</td>
<td>2.11</td>
<td>Remn. Veg</td>
<td>0.00009</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Rainfall Res.</td>
<td>0.38</td>
<td>Total Ag.</td>
<td>0.00006</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>KSAT [A]</td>
<td>0.41</td>
<td>PAWC [A]</td>
<td>0.10743</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Soil Depth</td>
<td>5.75</td>
<td>Thick [A]</td>
<td>-32.64193</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>KSAT [B]</td>
<td>0.02000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PAWC [B]</td>
<td>0.01565</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Thick [B]</td>
<td>-12.71624</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Soil Depth</td>
<td>5.75811</td>
</tr>
<tr>
<td>All variables</td>
<td>2002</td>
<td>532</td>
<td>0.48</td>
<td>Total Plant.</td>
<td>1.30</td>
<td>Remn. Veg</td>
<td>0.00007</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Rainfall Res.</td>
<td>3.82</td>
<td>Total Ag.</td>
<td>0.00007</td>
</tr>
<tr>
<td>All variables + soil</td>
<td>2002</td>
<td>524</td>
<td>0.54</td>
<td>Total Plant.</td>
<td>-0.01</td>
<td>Remn. Veg</td>
<td>0.00011</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Thick [A]</td>
<td>-0.02</td>
<td>Total Ag.</td>
<td>0.00008</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Rainfall Res.</td>
<td></td>
<td>Rainfall Res.</td>
<td>0.02222</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>KSAT [A]</td>
<td></td>
<td>KSAT [A]</td>
<td>0.02871</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PAWC [A]</td>
<td></td>
<td>PAWC [A]</td>
<td>0.48828</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>KSAT [B]</td>
<td></td>
<td>KSAT [B]</td>
<td>0.02106</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PAWC [B]</td>
<td></td>
<td>PAWC [B]</td>
<td>-0.03563</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Thick [B]</td>
<td>-34.34656</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Soil Depth</td>
<td>27.45593</td>
</tr>
</tbody>
</table>
Figure 4.1: Local $R^2$ from GWR models. All models displayed significant spatial and temporal variability, which has also been observed in previous studies (Chapter 3). With the exception of the 1995 ‘All variables + soil’ (panel d), the increased explanatory power of the models containing the seven hydrologically-relevant soil parameters was considered significant according to comparisons of corrected Akaike Information Criterion values.
The area of plantation in a catchment was modelled as a global coefficient across all models suggesting that there is no relationship between the extent of surface water, the position of a given catchment in the region and the area of plantation in that catchment, regardless of the long-term rainfall residual or as a function of soil conditions. With the exception of the 2002 ‘All variables + soil’ model, area of plantation had a positive global relationship with the area of surface water habitat (i.e. a positive model coefficient) suggesting that, as plantation area increased, an increase in surface water extent also occurred. Similarly, the positive model coefficients for the area of remnant vegetation of a catchment suggested that increases in extents of remnant vegetation result in an increase in the area of surface water (Table 4.1).

Saturated hydraulic conductivity of the A horizon (KSAT [A]) was set as a global coefficient in the 1980 and 1995 model, while thickness of the A horizon (Thick [A]) was set as global in the 2002 model (Table 4.1). As the actual values of the soil parameters did not change among modelled years, this suggests that there was temporal variation in the relationship among the parameters. Furthermore, the negative coefficients for Thick [A] and Thick [B] in all models suggest that, as the thickness of the respective soil layers increased, there was a decrease in surface water extent. However, the total soil profile depth ([Soil Depth] which was not equal to the sum of Thick [A] and Thick [B] as it is a weighted average depth; Western and McKenzie, 2006) showed positive coefficients across all years, suggesting that where there was a deeper soil profile there would be more surface water (Table 4.1). Somewhat unexpectedly, with the exception of the 1980 ‘All variables + soil’ model, the saturated hydraulic conductivity of both layers was positively associated with surface water extent. This suggests that, as the
hydraulic conductivity of soils increases, there is an increase in surface water extent. A very similar relationship was observed with the plant available water capacity of the A and B soil horizon (PAWC [A] and PAWC [B]) (Table 4.1).

4.4 Discussion

Previous research has illustrated the value of applying GWR in modelling the relationship between climate, land cover and wetland extents in the region (Chapter 3). The results presented here support previous modelling efforts, in that we have again identified a spatio-temporally-variable relationship among climate, land cover and wetland extent (Table 4.1, Figure 4.1); however some of the conclusions are different. For example, Chapter 3 (Table 3.2) suggested that, in areas where there was a greater proportion of plantation forestry and remnant vegetation, there would be less surface water extent. Likewise, the models also suggested that anthropogenic drainage practices linked to agriculture would limit the extent of surface water. The GWR models presented here do not support this previous assessment and instead suggest that, as the area of a catchment used for agriculture, plantation forestry or remnant vegetation increases, so will the extent of surface water. This was observed even in models that did not contain any of the new soil variables and could be related to changes in the software between version 3.0 and 4.0 (Nakaya, 2014b), specifically those related to the calculation of local model coefficients (although the mechanism for such an impact is unknown, given that the nature of those changes are not made explicit in software documentation).

As the distribution of water within a landscape is a combination of processes related to vegetation (Brown et al., 2005), climate (Bronstert et al., 2002) and soil (Castillo et al., 2003; D’Odorico et al., 2007; Li and Sivapalan, 2011), the next logical
step to improve on the previously-modelled relationships (Chapter 3) was to include variables relevant to soil hydrology (Western and McKenzie, 2006). Unfortunately, the results of the modelling have not clarified the relationship the between soil parameters, land cover, climate and surface water extents. There were a number of results that were counter-intuitive, such as the suggestion that areas with higher saturated hydraulic conductivity would have greater wetland extents. As hydraulic conductivity increases, by definition, the ability of soils to effectively drain water also increases. It follows that there should be less surface water in areas where soils have an increased ability to drain but such a relationship was not suggested by the results of this modelling. Similarly, an increase in plant available water capacity was also associated with an increased wetland extant; an unlikely relationship given that, as there is more water available from the soil for uptake by plants, vegetation densities (D’Odorico et al., 2007) and evapotranspiration (Brown et al., 2005) should be higher, ultimately leading to a decrease in surface water extents.

While the models created here were able to demonstrate that there is a spatially-varying relationship among the variables, the contradictory results between this research and previous modelling (Chapter 3), makes it difficult to quantify the links between surface water, land cover, soils and climate in the region. In particular, the counter-intuitive relationships between soil variables and wetland extents in the region was particularly interesting as soil variability is known to affect runoff processes (Castillo et al., 2003), and soil moisture (and by extension, wetland extent) is a synthesis of the dynamic effects of climate, soil and vegetation in dryland ecosystems (D’Odorico et al., 2007). The coefficients for the predictor variables could be affected by localised
collinearity in model predictors, which has been shown to influence model coefficients and consequently interpretation (Wheeler, 2009). Penalisation methods such as such as geographically weighted lasso regression (GWR-L) have been developed to reduce issues of localised correlation between predictor variables, however the method lowers prediction and estimation error of the response variable (Wheeler, 2009). Future work could investigate these methods, particularly in light of the counter-intuitive regression coefficients observed here. As such, there are no clear recommendations that can be taken from this research for management of the water resources in the Glenelg-Hopkins region. However, this research has contributed to the growing body of evidence of significant spatio-temporal variability in relationship between water resources and environmental conditions in the region. The spatial fluctuations in GWR model coefficients suggest that land-cover and water resources management policies should be tailored to specific areas across the region, as an approach that works in one of the 149 sub-catchments modelled may not work in other, nearby catchments. Further model improvements, such as the use of hydrological response units ([HRU], where the unit of analyses are hydrologically-homogeneous set of landscape characteristics such as slope and soil attributes; Kite and Pietroniro, 1996), as opposed to catchments, for example, or the utilisation of GWR-L, may help clarify the relationships revealed here and lead to more robust management recommendations.

4.5 Conclusions

This short study has once again confirmed the capability of GWR to model systems where there is a significant amount of spatial variation in the relationship among dependent and independent variables. The results of the models presented here suggest
that there is a positive link between the areas of agriculture, plantation forestry, remnant
native vegetation and surface water extents in the region. Unfortunately, results of the
models created here do not agree with previous modelling efforts in the region, with
some of these inconsistencies potentially related to changes in the model software itself.
With the exception of one model, the addition of a number of hydrologically-relevant
soil parameters did improve the performance of the models according to AICc. More
research is therefore recommended to increase the ability of this approach to elucidate
spatio-temporal relationships between climate, land cover, soils and water in the region –
perhaps by using proportional areas rather than total areas of different land cover within
a catchment, or the use of HRUs instead of catchments as the unit of analyses. Land and
water managers need to be aware that spatial differences in environmental relationships
could affect management measures and that simple methods such as GWR can help
identify the extent and magnitude of spatial variability that exists within a given system.
Hydrologic landscape regionalisation using deductive classification and random forests

5.1 Introduction

5.1.1 Flow variability and ecological controls

Long-term trends in flow variability in streams have the ability to create and maintain ecosystem dynamics for a range of ecologically-important conditions (Bunn and Arthington, 2002; Poff et al., 2010) and can therefore influence biotic communities and abiotic conditions at local to regional scales, both temporally and spatially (Poff et al., 1997; Bunn and Arthington, 2002). These long-term trends are controlled by the same factors influencing the hydrologic cycle in a landscape and ultimately influence physical habitat and refuge availability, food distribution and abundance, and opportunities for migration, reproduction and recruitment (Naiman et al., 2008). Given this ability for hydrologic variability to control the ecological and biophysical attributes of in-stream and riparian systems, landscapes that have similar hydrologic properties should have similar biological and ecological assemblages (Poff and Allan, 1995). Furthermore, if the same or similar hydrologic landscapes can exist in multiple spatial locations within bioregions, it stands to reason that the ecology of these systems should also be similar, regardless of spatial location. The ability to identify, classify, and validate spatial patterns in hydrologic landscapes is an important step in creating a solid foundation to assess the impact of natural flow variability, associated ecological conditions and management of water resources across a range of spatial scales. As such, hydrologic classification has been identified as a critical step in providing a spatially-explicit understanding of the magnitude and timing of flow regime variation within and between rivers and regions (Kennard et al., 2010b; Poff et al., 2010).
5.1.2 Landscape and hydrologic units

Landscape characteristics affecting the quality, quantity, and movement of water are extremely complex (Winter, 2001). The earth is made up of a number of different landforms, geological settings and climatic conditions, and the idea of a simple, unifying conceptual hydrologic framework may seem impossible to achieve (Winter, 2001). However, landscapes that appear unique and diverse often actually have a common set of attributes (e.g. governing the movement of water). Winter (2001) introduced the concept of hydrologic landscape units, which suggests that the complete hydrologic system (i.e. incorporating surface runoff, groundwater flow and atmospheric water) interacts with simple physiographic features, and that these features then become the building blocks of all hydrologic landscapes. Therefore, by this rationale, the movement, storage and release of surface and subsurface water are controlled by a common set of physical principles regardless of the geographic location of the landscape (Wolock et al., 2004). Winter (2001) termed these ‘fundamental hydrologic landscape units’ (FHLU), and defined the conceptual unit as a land surface form which includes an upland, an adjacent lowland and the valley side that separates them. The hydrologic system of an FHLU consists of: 1) the movement of surface water (controlled by the slopes and permeability of the landscape); 2) the movement, storage and release of groundwater (a function of the geologic setting); and 3) atmospheric water exchange (controlled by climate) (Winter, 2001). Much peer-reviewed research supports the idea that all hydrologic landscapes can be considered to be variations and multiples of FHLUs, and that these can then be used to describe major, spatially-contiguous and discrete landscape types that should have similar hydrologic conditions (e.g. Wolock et
al., 2004; Kennard et al., 2010b; Olden et al., 2012). Since the concept was first introduced, further research has been conducted to delineate hydrologic landscape regions based on a number of different approaches and across a variety of scales (see Olden et al. (2012) and Kennard et al. (2010b) for an extensive list of examples).

5.1.3 Deductive and inductive landscape classification

Classification is the process of systematically placing objects into classes that are similar with respect to a set of variables or characteristics. Hydrologic classification is therefore the process of systematically arranging streams, rivers or catchments into classes that are similar with respect to their flow regime (Kennard et al., 2010b; Olden et al., 2012). While hydrologic classification can refer to a broad assortment of methods, a review by Olden et al. (2012) recognises two broad approaches to hydrologic classification; deductive and inductive approaches (not to be confused with top-down and bottom-up logic; see below). The inductive approach uses the emergent properties of discharge time series data to generate classes – i.e. flow-based classification. In contrast, the deductive approach to classification is used when attempting to describe broad spatial patterns in flow regime variability where there is a lack of gauged or modelled streamflow data available – i.e. hydrological regionalisation. Deductive methods of environmental classification are commonly used when the objective is to quantify and describe spatial variation in flow regime attributes. This approach to classification identifies groups on the basis of physical and climatic attributes that, over broad scales, produce similar hydrologic responses in stream systems (Olden et al., 2012). The increased availability of high-quality, hydrologically-relevant spatial datasets (e.g. climate, topography, land cover) makes deductive reasoning an appealing method when
attempting to define spatial similarities or dissimilarities in hydrological characteristics (Olden et al., 2012). It has been demonstrated that the deductive approach to hydrological classification can help in the prediction of streamflow metrics (Snelder et al., 2005; Santhi et al., 2008), and that it improves predictive streamflow models when those models are stratified by hydrologic regions (Carlisle et al., 2010). However, some facets of flow regimes (e.g. low flow magnitude and duration) are difficult to accurately characterise and quantify with this approach due to limitations in data quality and conceptual knowledge of the systems, and spatial variability of hydrological processes in many regions (Kennard et al., 2010b; McManamay et al., 2012; Olden et al., 2012).

Wolock et al. (2004) used the concept of hydrologic landscapes introduced by Winter (2001) to classify nearly 44,000 catchments (~200 km² in area each) using a combination of multivariate ordination and cluster analyses. Kennard et al. (2010b) presented a method combining non-hierarchical clustering of climate, topography, soils and geology, vegetation and flow data to group Australian streams at a continental scale with mixed success. Sawicz et al. (2011) employed the use of precipitation-temperature-streamflow signatures and Bayesian clustering to characterise 280 non-contiguous catchments located in eastern USA so as to understand similarities in climatic and landscape attributes across the region. Their work found that signatures which vary along climatic gradients exerted a stronger influence on cluster separation than those signatures which may vary as a result of geology or land cover. It has also been shown by McManamay et al. (2012) that hydrological regionalisations (Poff, 1996) can be severely lacking in their ability to explain variation in a number of streamflow metrics.
The approaches by Wolock et al. (2004), Kennard et al. (2010b), Sawicz et al. (2011) and others all require the use of catchments or some choice of arbitrary spatial unit (e.g. eco/bio-region) to delineate and display the results of the clustering. However, there is evidence of significant flow variability within river catchments (Poff et al., 2006; Kennard et al., 2010b) and significant spatial variability in climate and land cover within sub-catchments that affect wetland extent (Chapter 3; Chapter 4). The approach of delineating spatial units \textit{a priori} leads to a loss of spatial variability, particularly as the catchment or spatial units become larger. Olden et al. (2012) state that while deductive classification is common in the literature, hydrologic landscape regions and other similar concepts that are founded on physical principles have rarely been tested with this approach. The \textit{a priori} (or ‘top-down’) specification of boundaries between classes has been criticised, while alternative ‘bottom-up’ approaches, where groups are developed as an emergent property of the data (Mackey et al., 2008) (not to be confused with inductive reasoning which relies on time series hydrologic data) have been considered to be in keeping with physical ecohydrological principles (Olden et al., 2012). Using a bottom-up approach, spatial and group clustering patterns are generated based on the analysis of a large number of units, such as pixels or micro-catchments. These units are then allocated into clusters based on their multivariate similarity (Mackey et al., 2008). However, a number of subjective choices as to which datasets to include, classification strategies and the number of groups in the classification process still need to be made. Such decisions could affect the quality and repeatability of the classification process when applied to different regions and datasets (Mackey et al., 2008; Stein et al., 2009; Sawicz et al., 2011; Olden et al., 2012). Despite the potential limitations, the routine
availability of these datasets and the application of statistical clustering and analyses have allowed scientists to begin to link spatial patterns to ecohydrological processes.

5.1.4 Statistical clustering and multivariate analyses

Statistical clustering and multivariate analyses are important and powerful tools in the identification of spatial and temporal gradients. There is a multitude of variations on the theme of statistical clustering (Cormack, 1971; Clifford and Stephenson, 1975; Everitt, 1980), but the most commonly used are hierarchical agglomerative methods (Clarke and Warwick, 2001) which fuse individual samples into like groups, gradually increasing the similarity within groups while lowering the similarity level between groups; i.e. each sample starts as its own group and pairs of groups are merged moving up a hierarchy. The process is considered complete when all samples are contained within a single group or cluster. Unlike hierarchical clustering, non-hierarchical clustering places samples into groups that are not related hierarchically, but differ from each other significantly in multivariate space. Described simply, non-hierarchical clustering tends to work by assigning each sample \((n)\) into a pre-defined number of clusters \((k)\) and then cluster membership of the samples is iteratively reassessed, usually with the criterion of maximising between-cluster variance while simultaneously minimising within-cluster variance. The most common example of non-hierarchical clustering is the \(k\)-means algorithm (Legendre and Legendre, 1998). In some instances, the groups extracted by hierarchical and non-hierarchical algorithms do not differ significantly (Belbin and McDonald, 1993), but non-hierarchical methods can be much more efficient at extracting groups from large datasets (Belbin, 1987).
Once statistical clustering has occurred, an analysis of the performance of the clustering can be conducted through the use of multivariate statistics and, in particular, ordination plots. There are a number of ordination techniques (e.g. principal components analysis [PCA], and multidimensional scaling [MDS]) that, broadly speaking, reduce multidimensional space so that objects can be compared graphically in two or three dimensions, without a significant loss of explanatory information. It is also possible to use the clustering information to train predictive models to classify samples not included in the original classification. This is where supervised classification algorithms, such as random forests (RF; Breiman, 2001; Held et al., 2012), coupled with geographic information systems (GIS) and image processing software can be applied to extend the applicability of deductive landscape classification approaches to regionalisation studies.

5.1.5 Supervised classification of landscapes

One of the most common applications of remotely-sensed images and data is the creation of maps of vegetation type, soil properties or other discrete classes. In supervised classification, the location of known classes on those maps (i.e. training sites) is used by the software to determine the spectral signature of the pixels belonging to each of those classes. Each pixel in the image (i.e. outside the training sites) is then assigned, based on its spectral signature, to the class it most closely matches. Supervised classification can be applied at the individual pixel level or to groups of adjacent, similar pixels for the creation of contiguous regions. However, for the classification to work effectively, *a priori* knowledge of where the classes of interest (e.g. land-cover types) are located is required. When supervised classification is combined with, for example, an
unsupervised statistical classification, the process is referred to as hybrid (or semi-supervised) classification (Lillesand et al., 2008). A major benefit of hybrid classifications for landscape regionalisation is that they permit the bottom-up approach to deductive classification as recommended by Mackey et al. (2008) and Olden et al. (2012). This eliminates the need for survey approaches to develop a priori knowledge of the location of classes of interest which require expert opinion and substantial amounts of qualitative evidence which is not always available or suitable. The hybrid approach also eliminates the need to define a spatial unit a priori (e.g. a catchment) and allows small-scale (e.g. intra-catchment) variability to be identified and preserved where it may otherwise be lost.

5.1.6 Aim of study

The aim of this study was to create a hydrologic landscape regionalisation using deductive reasoning and a bottom-up approach to statistical clustering combined with a hybrid classification. The regionalisation was then assessed based on its ability to discriminate between groups (regions) based on a number of streamflow indices. In this research, we used unsupervised classification (i.e. the statistical clustering) to first determine class membership based on multivariate space and then used supervised image classification to classify the remaining pixels from a number of ecohydrologically-important layers into the classes of interest as defined by the statistical clustering. This approach will permit the regionalisation of spatially non-contiguous regions, while maintaining small-scale intra-catchment variability that would be lost using catchments as the unit of classification as has often been done in the past. The assessment of the ability of the regions to differentiate among streams based on a number of flow indices
provides insight into the utility of the method in predicting streamflow characteristics in ungauged catchments.

5.2 Materials and methods

To clarify the process used in the creation of the hydrological regionalisation and the validation and training methods for the RF models, a graphical overview of the methods is presented in Figure 5.1.
Figure 5.1: Graphical overview of the methods employed in the creation of the hydrological regionalisation; the training and validation of the RF models used to extend the statistical clustering to the state of Victoria; and of the validation of the classification with hydrological data. Each colour coded section of the figure corresponds to a section in the methods. Green: Variable Selection and Processing; Orange: Development of Classification Groups; Beige: Hybrid Classification with Random Forests; Light Blue: Accuracy Assessment; Grey: Relationship between the Regionalisation and Hydrologic Indices. The process for the ALOC 20100 % models did not involve splitting the ALOC classified random sample points into training and validation subsets and model accuracy was only assessed with OOB accuracy from EnMap-Box.
5.2.1 Site description

Victoria is the southernmost state of mainland Australia, comprising an area of 227,594 km², and bordered by the southern bank of the Murray River to the north, South Australia to the west and separated from Tasmania by Bass Strait to the south. Topographically, geologically, and climatically, Victoria is diverse, varying from wet temperate climates in the southeast to alpine areas rising to ~2000 m altitude in the northeast (Figure 5.2). Median annual rainfall in Victoria exceeds 2,500 mm in some parts of the mountainous northeast but is less than 300 mm in a large part of the west and northwest (Department of Environment and Primary Industries, 2013a). To the west and northwest are extensive, flat areas of semi-arid plains, while most of the rest of the state experiences a Mediterranean climate consisting of hot, dry summer and cool, wet winters (Department of Environment and Primary Industries, 2013b). Generally, snowfall is only observed in the mountains and hills to the east and centre of the state. Victoria has an extensive wetland system, with nearly 17,000 wetlands larger than 0.01 km² in surface area (Corrick, 1992), and a large river network, with the largest being the Murray River system.
Figure 5.2: Location of the study area in south-eastern Australia. Dark blue lines represent perennial rivers, while the colour gradient represents elevations, with darker browns indicating higher elevations.
5.2.2 Site description of case-study area in western Victoria

The Glenelg-Hopkins region of western Victoria covers approximately 27,000 km² and the regional cities of Warrnambool, Ararat, Hamilton, Portland and the western fringes of Ballarat are within its boundary. The region contains the Grampians Ranges in the north but is generally a low-lying series of catchments across three major catchments – Glenelg, Hopkins and Portland. The region has been previously studied with respect to land-use and land-cover changes (Versace et al., 2008a) and the associated impacts on nutrient exports, in-stream salinity and dryland salinity (Ierodiaconou et al., 2005; Versace et al., 2008b), while recent work has examined the spatio-temporal variability between land cover, climate and wetland extent (Chapter 3; Chapter 4), the impact of land-cover changes on groundwater levels (Yihdego and Webb, 2011) and empirically modelled streamflow response to land-cover change (Yihdego and Webb, 2013).

5.2.3 Variable selection and processing

The first phase of the classification involved selecting suitable variables upon which to base our classification. Steps associated with variable selection and processing, described in this section, are outlined in green in Figure 5.1. Based on the concept of FHLUs, 25 variables were chosen that could explain the storage, movement, and quality of surface water, groundwater, and atmospheric water. A full list and brief description of each of the variables are presented in Appendix I (Table 9.1). All raster calculations and raster analysis for the processing of variables was conducted in ArcGIS 10.1 (ESRI, 2012).

The raster datasets employed in the study covered a wide range of resolutions (30 m – 10 km). Typically, with GIS, analyses are only considered to be suitable if all
rasters are resampled to the coarsest resolution. However, this can result in the loss of a substantial amount of detail and information and can affect the ability of supervised classification methods to successfully classify pixels (See Figure 5.3 for a comparison between the 30-m and 10-km Landscape Development Index [LDI]). Therefore, in this study, two approaches were used to standardise the scale of our raster data. The first approach was to resample all datasets to the finest resolution (30 m); and the second involved re-sampling all the raster datasets to the coarsest resolution found in our datasets (10 km). All rasters were continuous in their spatial coverage with the exception of the soil hydrological properties (KSAT, PAWC and soil horizon thickness) which had significant gaps where large lakes and wetlands were found. There was also a significant gap in coverage on the eastern headland of Port Phillip Bay. To ensure that all datasets aligned correctly and had the same degree of spatial continuity, the digital elevation model (DEM) was used as a snap raster for the resampling. Once the resampling had been completed using a nearest neighbour algorithm, the now 30-m soil properties were used as a mask to extract all other raster values. The result of this was that all of the datasets used in the analysis had a 30-m spatial resolution and all had corresponding areas of missing data that would be excluded from any analysis.

For the second approach, all of the original datasets were resampled to 10 km using a nearest neighbour algorithm and the mean annual evapotranspiration raster as the snap raster. Two different datasets were used as snap rasters so that the pixels of the resampled rasters (at either 30 m or 10 km) would be aligned correctly at the respective resolution. If this study had been conducted at a continental scale, then a coarser resolution would be more suitable, however as it was conducted on a relatively small
scale, we considered that resampling to a finer resolution was both suitable and justifiable. Furthermore, limiting our analysis to a coarser resolution (on the basis of a single coarse-scaled dataset; mean annual evapotranspiration), would have significantly affected the applicability and usefulness of the method presented here, specifically, the ability of the RF model to accurately recover and reproduce allocated class information. Previous research has shown that the accuracy of supervised classifications, at both an overall and per class level, can be affected by the spatial resolution of the input images (Mumby and Edwards, 2002; Sprintsin et al., 2007). As such, we assessed this issue through the accuracy of the RF model and the ability of the hybrid classifications to accurately recover the class information using the coarser dataset (see section on accuracy assessment). A layer stack of both sets of variables (30-m and 10-km resolution) was produced in ENVI 4.8 (Exelis Visual Information Solutions, 2010) for later use with the RF model.

Two random distributions of sampling points ($n = 10,000$ and $n = 410$; with minimum distances between points of 30 m and 10 km respectively) were then created. Ripley’s $K$ function (Dixon, 2006), which determines whether features are significantly clustered or dispersed over a range of distances, was then used to assess whether both sets of points were distributed across our sampling area. Using the random-sampling points, raster values were extracted from each of the raster layers for later use in clustering, and then the training and validation of the RF model.

5.2.4 Development of classification groups

The selected variables were then statistically analysed to develop the groups (known as ‘regions’) to be used as the basis for classification. The steps involved are
outlined in orange in Figure 5.1. All sampling points that were found to contain missing
data were removed prior to analysis, resulting in $n = 9,958$ (30-m data) and $n = 406$ (10-km data) for the two sets of data points. For the 30-m data, three 1000-point subsamples were taken for initial statistical analysis, while all data points were used for the 10-km dataset. In order to develop a bottom-up approach to classification, where groups are an emergent factor of the data rather than defined a priori, we cluster-analysed each of the subsamples individually. Due to the large ranges and different scales used across our variable set, Gower similarity matrices were constructed for each of the initial analysis datasets using PRIMER 6 (Clarke and Gorley, 2006).
Figure 5.3: An example of the differences in resolution identified when working in relatively small study areas. The left hand image is the Landscape Development Intensity index (LDI) at a 30-m resolution while the image on the right is the LDI at a 10-km resolution. The accuracy of supervised classifications can be affected by the spatial resolution of the input images and as such we developed models at both resolutions.
The CLUSTER (Clarke, 1993) function was then used with a SIMPROF (Clarke et al., 2008) test, to identify the number of statistically-significant ($\alpha = 0.05$) groups within the datasets. Essentially, SIMPROF determines the number of significant groups with the assumption of no \textit{a priori} groups by calculating similarities between every pair of samples using the chosen resemblance matrix and a hierarchical cluster dendrogram. Beginning at the top of an already-defined hierarchy (i.e. by the CLUSTER function), progression down the divisions or branches of the dendrogram is only permitted if the current set of samples is deemed to still have statistically-significant dissimilarity. Upon encountering a non-significant result (i.e. the samples are similar), no further tests are performed down that branch of the dendrogram and all samples below are considered part of the same group (Clarke et al., 2008). A limitation of the SIMPROF test is that groups identified by the test may be at too fine a level of detail for practical purposes. However, if the resulting clusters are super-sets of the SIMPROF-defined groups, it is appropriate to define coarser groupings based on an arbitrary slice at some chosen level of similarity (Clarke et al., 2008). As the 30-m SIMPROF tests were conducted on three 1000-point subsamples, the subsample that produced the largest number of groups was used to determine the number of groups for the 30-m data. Even though SIMPROF uses a hierarchical relationship between sampling points to determine the number of clusters present in the data, we believe this approach is suitable for estimating an appropriate number of non-hierarchical groups as opposed to choosing an arbitrary $k$ number of groups.

We then classified the full datasets ($n = 9,958$ and $n = 406$) into the number of classes suggested by the SIMPROF tests using the non-hierarchical clustering algorithm
ALOC (Belbin, 1987) and the Gower metric in PATN v3.1.2 (Blatant Fabrications Pty. Ltd., 2009). Group allocations were then exported from PATN and joined to the original datasets as factors for further analysis. Once group membership information was in PRIMER, the ANalysis Of SIMilarities (ANOSIM; Clarke, 1993) routine was used to test for statistically-significant differences among sample groups. Based on the $R$ statistic, which is scaled to be between -1 and +1, global $R$ values > 0 indicate greater dissimilarity between groups than within groups (Clarke and Warwick, 2001). Group averages were calculated using the AVERAGE (Clarke and Gorley, 2006) tool in PRIMER to visually analyse group separation using MDS. MDS is useful in providing a visual representation of the pattern of similarities between objects or groups while reducing the multidimensional space to be more readily interpretable (i.e. reducing data to two or three dimensions). The ability of MDS to reduce the degree of multidimensional space is measured with a stress value. Essentially, stress is the mismatch between distances between all samples in the plot in multidimensional space and the calculated estimate of their respective locations in two or three dimensions, with lower values indicating better representation (Clarke and Warwick, 2001). The CLUSTER and SIMPROF routines were then used to hierarchically cluster the ALOC generated group averages into ‘meta-groups’. By definition, non-hierarchical groups are not linked based on their hierarchical multivariate relationship to each other, but rather are defined by their multivariate dissimilarity. As such, group $x$ may not be closely related to group $y$ but could be more closely related to group $z$. By hierarchically clustering our ALOC generated groups, were we able to determine which ALOC groups were more closely related to each other based on their multivariate means. A standardised Euclidean distance similarity matrix was
then created and the SIMilarity PERcentage (SIMPER; Clarke, 1993) routine was then used to analyse variable contribution to each of the meta-groups and to examine between meta-group similarity, while the Kruskal-Wallis (R Core Team, 2014) statistic was used to assess the ability of each of the variables to differentiate between clusters.

Previous studies (e.g. Wolock et al., 2004) have employed PCA (Clarke, 1993) to reduce dimensionality and reduce multi-collinearity among variables. Our method relied firstly on using ‘raw’ data (i.e. the data was not transformed in any way) to extract spectral information for the statistical clustering and then classification. The results of this method were then compared against a classification based on PCA-transformed data. PCA is a procedure where possibly correlated variables are orthogonally transformed into a new set of linear, uncorrelated variables known as principal components (Clarke and Warwick, 2001). The transformation results in the first component explaining most of the data variance (i.e. the first component explains as much of the multivariate data as possible) while each additional component in turn is then created to explain the remaining variance. The number of components is less than or equal to the number of original variables and each additional component is created under the condition that it is uncorrelated with all of the preceding components (Clarke and Warwick, 2001). Here, a PCA was conducted in PRIMER on the same standardised Euclidean distance matrix used for the SIMPER analysis. The eigenvectors, for the first five principal components (eigenvalues ≥ 1) were used in ArcMap to generate PCA bands and new PCA raster stacks were created in ENVI.
5.2.5 Hybrid classification with random forests

The regions that were developed in the previous step were then used to classify the raster stack of the variables for the entire study area. Steps in this section are outlined in beige in Figure 5.1. RF is an ensemble machine-learning method used in classification and regression (Breiman, 2001). The RF method is relatively unknown in land remote sensing and has not been thoroughly evaluated by the remote-sensing community, although it has been shown to be more accurate than single decision-tree classifiers (Rodriguez-Galiano et al., 2012). RF requires two parameters for generating a predictive model: the number of trees \( k \) and the number of variables used for growing the trees \( mtry \). Therefore, a dataset can be classified by defining a constant number of \( mtry \) variables, while each of the training samples is classified by \( k \) trees.

Classification is determined by using the mode of the classes output by individual trees for each training site \((x)\) using the equation,

\[
\hat{C}_b(x) = \text{majority vote} \left\{ \hat{C}_i(x) \right\}_{1}^{B}
\]

where \( \hat{C}_b(x) \) is the class prediction of the \( b \)th RF tree from a possible \( B \) classes (Rodriguez-Galiano et al., 2012). RF increases the diversity of the constituent trees by making them grow from different training data through bootstrap aggregation which involves random re-sampling (without deletion) of the original training dataset (Breiman, 2001). Therefore some data may be used more than once in the training of the model, while some may not be used at all (Rodriguez-Galiano et al., 2012). Being an ensemble method, multiple models (trees) are used allowing the algorithm to obtain better predictive performance than that which would be obtained by using any of the
constituent models individually. RF is becoming increasingly popular in data mining, remote sensing and landscape ecology as it is non-parametric, can generate internal, unbiased error estimates and variable importance, is robust to training data reduction and noise, and is highly accurate (Breiman, 2001; Rodriguez-Galiano et al., 2012).

In order to develop the RF models, the sampling points were randomly split into independent training (80 %) and validation (20 %) datasets and then stratified using the cluster-membership allocations from PATN. Due to the small sample size of the 10-km dataset, an additional RF model was created using 100 % of the sample for RF training. Using ENVI, layer stacks of the raster data (30 m, 10 km, and PCA at both resolutions) were constructed and masked. Layer stacks, masks and training regions of interest were imported in to EN-Map Box 1.4 (Held et al., 2012) to permit the building of RF models and the classification of the image stacks. EN-Map Box was set to use 200 trees ($k$) per sample, and the square root of the number of input variables on the non-PCA transformed data ($mtry = \sqrt{25} = 5$), or $m-1$ variables ($mtry = 4$) for the PCA models. The Gini coefficient (Breiman, 1993) was used to calculate impurity, which is one method used to evaluate the best split decision for each tree.

5.2.6 Accuracy assessment

The ability of the classification to accurately represent the information across a range of resolutions was then tested, with the relevant steps outlined in light blue in Figure 5.1. The accuracy of the RF models in recovering and classifying the image stacks into the ALOC classes was assessed with out-of-bag (OOB) error rates generated in EN-MAP box, and using the independent 20 % validation dataset to calculate percent agreement between classified and validation data, user and producer accuracies, and
Kappa (κ) coefficients (Cohen, 1960) in ENVI and R (R Core Team, 2014) with the
psych (Revelle, 2013) and irr packages (Gamer, 2012). User accuracy refers to the
probability that a pixel classified into a certain class really belongs to that class, while
producer accuracy refers to the probability that a certain class is classified correctly. The
locations of the validation pixels were used to extract class information from the original
and PCA classifications and percent agreement and κ coefficients between model types
were examined. High levels of agreement between the original and PCA classifications
would indicate an insignificant amount of multivariate information loss by PCA and
further support the use and application of methods that reduce data dimensionality and
multi-collinearity between variables in regionalisation studies. The 10-km sample that
used 100 % of the data for RF training could not be assessed for accuracy independently
and was therefore only assessed with OOB error and class distributions.

We also decided that, due to the resampling of the original data (to 30-m from
a range of resolutions), it was worth investigating the effect of resampling the 30-m
classifications to a coarser resolution (using the majority filter) to help remove some of
the finer-scale variability in the data. A resolution of 2.5 km was chosen as a suitable
pixel resolution, to ensure that our resampled assessment points were further apart than
the mean distance of the original samples (see below), this meant our resampled
classifications were equal to or larger than the resolution of the majority of the datasets
while still being finer than the soil erosivity index. To assess agreement between the 30-
m and 2.5-km resampled classifications (i.e. original and PCA both at 30-m and
resampled 2.5-km), 300 random validation points were selected from the 20 %
validation datasets, with minimum distances between points of 2.5 km, and used to
extract class information from the 30-m and 2.5-km resampled classifications. Agreement between the validation data and the 30-m and 2.5-km resampled classifications was assessed with percent agreement and $\kappa$ coefficients between model resolutions. Visual inspection of the class distributions was also conducted; for this the 300 random validation points were used for the 30-m and 2.5-km classifications, while all 81 validation points for the 10-km classifications were used. McNemar’s chi-square test (McNemar, 1947; Fay, 2010) was used to formally test for statistically-significant differences between model types (e.g. 30-m original and 30-m PCA) and resolutions (e.g. 30-m original and resampled 2.5-km original) and the validation dataset using the same 300 random validation points; however we recognised $a priori$ that the relatively large number of random validation points would likely make any difference statistically significant and that the resolution tests would not be independent of one another. As such, a permutation-based method ($n = 9999$; Appendix I, Section 9.2) was also used to assess for statistically-significant differences in the $\kappa$ values of the classifications (McKenzie et al., 1996; Foody, 2004).
The permutation test worked as follows:

- If \( V = \) validation data; \( A = \) classification 1; \( B = \) classification 2
- Let \( \text{test}(x, y) \), be a function that calculates the test statistic (\( \kappa \)) for the classifications,
- \( H_0 \) if \( A \) and \( B \) are approximately equal in their classification accuracy, observations in \( A \) and \( B \) can be exchanged without affecting \( \kappa \), given by \( \text{test}(V, A) \) and \( \text{test}(V, B) \),
- Randomly exchange data between \( A \) and \( B \) \( n \) times, and observe how these changes affect the \( \kappa \) of \( A \) and \( B \), i.e. \( \text{test}(V, A) - \text{test}(V, B) \),
- \( n \) permutations would result in \( N \) data points. Rank the \( N \) data points and observe where the \( \kappa \) from the original test (i.e. not the permutated data) is located among the \( \kappa \) values from the permutated data points. If the original \( \kappa \) is outside the 0.975 percentile (or below the 0.025 percentile) then we can claim that the two classifications are different at \( \alpha = 0.05 \).

If a high level of agreement was found, as indicated by high percent agreement and high \( \kappa \), then resampling to 2.5-km \textit{a posteriori} could be considered a suitable and justifiable method for smoothing the finer-scale variability.

5.2.7 Relationships between the regionalisation and hydrologic indices

Finally, we tested the results of our classification against a traditional classification based on hydrologic indices. Steps in this section are outlined in grey in Figure 5.1. As a preliminary assessment of the ability of the regionalisation to differentiate among streams with differing hydrology, we calculated a range of streamflow indices based on the recommendations of Olden and Poff (2003) and then explored the relationships between the regions and streams with a permutation-based ANOVA (PERMANOVA; Anderson, 2001) and a constrained discriminant ordination (Canonical Analysis of Principal Coordinates [CAP]; Anderson and Robinson, 2003; Anderson and Willis, 2003).
Streamflow gauge locations were downloaded from the Water Measurement Information System (data.water.vic.gov.au/monitoring.htm). These locations were then overlayed on the regionalisation (30-m ALOC 23 meta-group classification, see Section 5.3.3) and had group (region) information appended to them. For example, a gauge that was located in regionalisation group E, was assigned to group E regardless of the upstream contributions of the regionalisation classes. In regions where there were more than 50 stream-gauges present, 50 were chosen at random to be included in the analysis. Daily streamflow data between 1980 and 2010 were then downloaded for 564 gauges throughout Victoria. A minimum record length of 15 years within the 30-year temporal window was required for a gauge to be included in the analysis (Kennard et al., 2010a). Stream gauges that were potentially subject to modification by weirs, dams or water extractions were not specifically excluded from the analysis. Where there were missing periods of flow information (to a maximum of 20 days in any single event) the record was in-filled using linear interpolation (Kennard et al., 2009) with the Time Series Manager module of the River Analysis Package (RAP; Marsh et al., 2003). Gauges that had a single period of missing data greater than 20 days were excluded from the analysis. Thirty-two indices characterising different aspects of the flow regime (Olden and Poff, 2003, Table III, All Streams) for each stream were calculated using the Time Series Analysis module of RAP. Indices that were related to discharge (i.e. those divided by catchment area) were not included in the analysis.

To test for differences among our groups, the PERMANOVA+ add-on (Anderson et al., 2008) for PRIMER was used. One-way PERMANOVAs (999 permutations), using Group as a fixed factor, were run for the dataset of flow indices
based on the original gauges ($n = 201$) and for an additional dataset consisting of a bootstrapped sample of those flow indices ($n = 383$), based on normalised Euclidean distance matrices. Analyses tested both for main effects and pairwise differences among groups. Traditional Analysis of Variance (ANOVA) is powerful for univariate data however the traditional multivariate analogues (e.g. MANOVA), are too stringent in their assumptions; in particular, that of multivariate normality which is frequently untrue in ecological data (Belbin and McDonald, 1993; Anderson, 2001), for use in ecology. As such, permutation-based non-parametric methods are preferred (Anderson, 2001). PERMANOVA uses permutation methods to test the simultaneous response of one or more variables to one or more factors in an analysis of variance. The use of permutations in PERMANOVA removes the assumption of normal distributions which are required for traditional ANOVA/MANOVA testing and, as such, the only assumption of the test other than independence is that the observations can be exchanged under a true null hypothesis (Anderson, 2001). Another benefit of using a permutation approach is that the permuted $P$-values provide an exact test of each individual null hypothesis of interest, and as such ad-hoc pairwise corrections (e.g. Bonferroni) are not strictly necessary (Anderson et al., 2008).

When data are classified into a priori groups, unconstrained ordinations (PCA, MDS) are extremely useful for visualising patterns from a multivariate space. However, the overall dispersion of points (when reduced to two or three dimensions) can often hide the true multivariate differences among those groups and it may be very possible to discriminate among groups through another direction or dimension of the multivariate data (Anderson et al., 2008). Unlike unconstrained ordinations, constrained ordinations
have an *a priori* hypothesis which controls the way the multivariate data can be interpreted in an attempt to relate predictor variables (streamflows indices) to response variables (groups) (Anderson and Willis, 2003). In a discriminant analysis, the ordination axes are interpreted in such a way as to maximise the differences among *a priori* groups, while in a canonical correlation, the axes are interpreted to maximise correlations among variables. CAP first calculates the principal coordinate axes (PCO) among $N$ samples, and then chooses an appropriate number of PCO axes ($m$) for interpretation based on a number of criteria (see Anderson *et al.*, 2008, for details), including a leave-one-out cross validation procedure which attempts to maximise classification success. The benefits of CAP over other constrained ordination methods are its ability to use any distance or dissimilarity measure, conduct permutation tests for significance of relationships among variables and predict group membership of new samples (Anderson and Willis, 2003). To assess the ability of our regionalisation to discriminate groups based on streamflow indices, three CAPs were conducted using the PERMANOVA+ add-on for PRIMER. Two of the analyses were performed against the group information extracted for each stream gauge using the same normalised Euclidean distance matrices that were used in the PERMANOVA tests (i.e. the original dataset and the bootstrapped dataset of hydrologic indices). The number of axes ($m$) was not specified and a permutation test ($n = 999$) was conducted to test the strength of the relationship. In addition, a third CAP was conducted using the bootstrap dataset where a stratified random sampling approach was used to remove 20% of the group information as a validation sample. The CAP was conducted as before, with the exception being that the validation cases were allocated groups based on the results of
the CAP. Allocation accuracy by CAP was assessed by calculating percent agreement between the CAP allocated group and the original group information, and κ coefficients with the irr and psych packages in R. Pearson’s r (R Core Team, 2014) was calculated between the number of gauges in each group and the number of gauges classified correctly from each group to assess for thresholds at which a specified level of accuracy could be achieved.

5.3 Results

5.3.1 Spatial distribution of sample points

Analysis of spatial distribution of the sampling points concluded that the mean, minimum and maximum distances to the nearest neighbour among the 30-m points were 2.4 km, 202 m and 8.8 km, respectively. The corresponding values for the 10-km points were 15 km, 10 km, and 47.4 km, respectively. Ripley’s K, based on 999 permutations, indicated significant over-dispersion of the 30-m sample points to a distance of ~250 m and a significant clustering at distances greater than ~650 m, while the 10-km sample points were significantly over-dispersed at distances less than ~12 km, but were significantly clustered at distances greater than ~18 km. Based on these values, the spatial distribution of our sampling points was considered suitable for the analyses as no significant clustering was displayed by the sample points at the resolution of the datasets they would be sampling (e.g. there was no significant over-dispersion or clustering at the mean distance of 15 km for the 10-km points).
5.3.2 Clustering and ordination

Twenty-three groups were identified within the 30-m data points and 20 groups for the 10-km data. As a result, the data were allocated into 23 (ALOC 23 [30-m data]) and 20 (ALOC 20 [10-km data]) non-hierarchical groups. There were wide variations in the ability of the model variables to differentiate among clusters (Appendix I, Tables 9.2 and 9.3) and the group populations had differing multivariate distributions. The allocated groups were well separated (Appendix I, Figure 9.1) with global-$R$ values of 0.852 ($P = 0.001$) for the ALOC 23 clustering and 0.908 ($P = 0.001$) for the ALOC 20 clustering, indicating that cluster membership was highly unlikely to be a result of chance alone. This was supported by the fact that neither ANOMSIM resulted in any permutations that had $R$-statistics greater than or equal to the global-$R$ value.

The allocated groups were then further clustered into hierarchical meta-groups. Using the group averages for ALOC 23 and ALOC 20, 11 and ten meta-groups were generated (Appendix I, Figures 9.1 and 9.2). ANOSIM analysis again indicated that cluster membership was highly unlikely to be a result of chance alone and suggested that the meta-groups for both the ALOC 23 and ALOC 20 models were well separated with global-$R$ values of 0.668 and 0.762 ($P = 0.001$). Again no permutations had $R$-statistics that were greater than or equal to the global-$R$ value. Some variables exhibited no relationship between the observed values and the meta-groups, while other variables show very clear relationships to the meta-groups (Appendix I, Figures 9.3 – 9.7). For example, the values for the aridity index (low aridity index values represent drier regions) and rainfall decreased from groups A to K. The opposite relationship was observed for
maximum and minimum temperatures, again suggesting that, as we move through regions A to K, the environment becomes drier and warmer. The BioClim variables 8, 9, 15 and 16 also supported this relationship, with increases in BIO08 and BIO09 (mean temperature of the wettest and driest quarter, respectively), and decreases in BIO16 and BIO17 (precipitation of the wettest and driest quarters, respectively). Variable percentage contributions to the meta-groups for both the ALOC 23 and ALOC 20 classifications differed markedly (Appendix I, Tables 9.2 and 9.3, Figure 9.8) with, for example, saturated hydraulic conductivity of the B soil horizon (B_KSAT) contributing 0 % to a number of ALOC 23 meta-groups, while contributing 46 % to ALOC 23 meta-group F. The within-group variation was highly variable among groups, with average squared distances of groups ranging from 4.18 to 11.77 for the ALOC 23 meta-groups, and 4.44 to 13.45 for the ALOC 20 meta-groups (Appendix I, Tables 9.2 and 9.3).

Ordination by PCA resulted in the creation of five principal components (PC) for both ALOC 23 and ALOC 20, using a minimum eigenvalue of one. The first PC for each of the ALOC 23 and ALOC 20 clustering explained 45 % and 53 % of the data variance, with eigenvalues of 11.2 and 13.2, respectively (Tables 5.1 and 5.2). The five PCs combined explained 78 % and 87 % of the variance, with the final PC having eigenvalues of 1.39 and 1.25, respectively. PC 1 can be interpreted to represent those areas that are wet, cool, heavily vegetated, high elevation environments with steep slopes and low erodibility soils (Tables 5.1 and 5.2, $0.2 < r < -0.2$). With the exception of PC 1, none of the PCs for either of the datasets can be interpreted to represent similar environments across the two classifications. For example, PC 2 of the ALOC 23 dataset can be interpreted to represent areas with thick, weathered, A-horizon soils with high
levels of plant available water, with relatively small variations in temperature seasonality and poorly developed B-horizon soils. For the ALOC 20 dataset, on the other hand, PC 2 suggests areas of poorly developed A-horizon soils, but with thick B horizon soils, and cooler, wetter summers.
Table 5.1: Results of principal components analysis for the ALOC 23 classification. Eigenvalues are presented in the top row while the numbers in brackets represents the cumulative percentage variance explained by the PCA. Bold numbers indicate variables with a correlation ≥ 0.2 or ≤ -0.2 which was used for PC interpretation.

<table>
<thead>
<tr>
<th>Variable</th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>PC4</th>
<th>PC5</th>
</tr>
</thead>
<tbody>
<tr>
<td>A_KSAT</td>
<td>0.14</td>
<td>0.14</td>
<td>-0.42</td>
<td>-0.30</td>
<td>0.18</td>
</tr>
<tr>
<td>A_PAWC</td>
<td>0.01</td>
<td>0.48</td>
<td>-0.18</td>
<td>-0.02</td>
<td>-0.12</td>
</tr>
<tr>
<td>A_THICK</td>
<td>0.10</td>
<td>0.45</td>
<td>-0.21</td>
<td>-0.12</td>
<td>-0.02</td>
</tr>
<tr>
<td>ARIDITY_INDEX</td>
<td>0.29</td>
<td>0.01</td>
<td>0.04</td>
<td>0.02</td>
<td>0.03</td>
</tr>
<tr>
<td>B_KSAT</td>
<td>0.15</td>
<td>0.05</td>
<td>-0.40</td>
<td>-0.31</td>
<td>0.20</td>
</tr>
<tr>
<td>B_PAWC</td>
<td>0.12</td>
<td>-0.34</td>
<td>-0.09</td>
<td>-0.41</td>
<td>0.26</td>
</tr>
<tr>
<td>B_THICK</td>
<td>0.06</td>
<td>-0.38</td>
<td>0.14</td>
<td>-0.38</td>
<td>0.16</td>
</tr>
<tr>
<td>BIO04</td>
<td>-0.08</td>
<td>-0.34</td>
<td>-0.29</td>
<td>0.08</td>
<td>-0.39</td>
</tr>
<tr>
<td>BIO08</td>
<td>-0.22</td>
<td>0.05</td>
<td>-0.10</td>
<td>0.20</td>
<td>0.44</td>
</tr>
<tr>
<td>BIO09</td>
<td>-0.24</td>
<td>-0.10</td>
<td>-0.17</td>
<td>-0.04</td>
<td>-0.19</td>
</tr>
<tr>
<td>BIO15</td>
<td>0.08</td>
<td>0.14</td>
<td>0.16</td>
<td>-0.43</td>
<td>-0.53</td>
</tr>
<tr>
<td>BIO16</td>
<td>0.28</td>
<td>0.03</td>
<td>0.07</td>
<td>-0.05</td>
<td>-0.05</td>
</tr>
<tr>
<td>BIO17</td>
<td>0.27</td>
<td>-0.02</td>
<td>0.03</td>
<td>0.18</td>
<td>0.23</td>
</tr>
<tr>
<td>ELEVATION</td>
<td>0.25</td>
<td>-0.14</td>
<td>-0.08</td>
<td>0.04</td>
<td>-0.16</td>
</tr>
<tr>
<td>ET_ANNUAL</td>
<td>0.24</td>
<td>0.14</td>
<td>0.28</td>
<td>0.06</td>
<td>0.10</td>
</tr>
<tr>
<td>GW_SWL</td>
<td>0.17</td>
<td>-0.08</td>
<td>-0.29</td>
<td>0.09</td>
<td>-0.14</td>
</tr>
<tr>
<td>GW_TDS</td>
<td>-0.17</td>
<td>-0.10</td>
<td>-0.32</td>
<td>0.08</td>
<td>0.02</td>
</tr>
<tr>
<td>LDI</td>
<td>-0.11</td>
<td>-0.11</td>
<td>0.16</td>
<td>-0.07</td>
<td>0.05</td>
</tr>
<tr>
<td>MAX_TEMP</td>
<td>-0.27</td>
<td>-0.05</td>
<td>-0.18</td>
<td>0.00</td>
<td>-0.07</td>
</tr>
<tr>
<td>MIN_TEMP</td>
<td>-0.25</td>
<td>0.08</td>
<td>-0.05</td>
<td>0.00</td>
<td>0.15</td>
</tr>
<tr>
<td>RAIN_ANNUAL</td>
<td>0.29</td>
<td>0.00</td>
<td>0.04</td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>SLOPE_RAD</td>
<td>0.23</td>
<td>-0.10</td>
<td>-0.20</td>
<td>0.22</td>
<td>-0.03</td>
</tr>
<tr>
<td>SOIL_EROS</td>
<td>0.26</td>
<td>-0.07</td>
<td>-0.05</td>
<td>0.09</td>
<td>0.03</td>
</tr>
<tr>
<td>TWI</td>
<td>-0.18</td>
<td>-0.01</td>
<td>0.16</td>
<td>-0.27</td>
<td>0.03</td>
</tr>
<tr>
<td>WEATH_IND</td>
<td>-0.16</td>
<td>0.23</td>
<td>0.10</td>
<td>-0.26</td>
<td>0.14</td>
</tr>
</tbody>
</table>

Eigenvalue: ALOC 23

<table>
<thead>
<tr>
<th>Eigenvalue</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>11.2 (45)</td>
<td>3.26 (58)</td>
<td>2.07 (66.3)</td>
<td>1.58 (72.6)</td>
<td>1.39 (78.2)</td>
</tr>
</tbody>
</table>
Table 5.2: Results of principal components analysis for the ALOC 20 classification. Eigenvalues are presented in the top row while the numbers in brackets represents the cumulative percentage variance explained by the PCA. Bold numbers indicate variables with a correlation ≥ 0.2 or ≤ -0.2 which was used for PC interpretation.

<table>
<thead>
<tr>
<th>Variable</th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>PC4</th>
<th>PC5</th>
</tr>
</thead>
<tbody>
<tr>
<td>A_KSAT</td>
<td>0.13</td>
<td>-0.22</td>
<td>0.15</td>
<td>0.53</td>
<td>0.03</td>
</tr>
<tr>
<td>A_PAWC</td>
<td>0.05</td>
<td>-0.45</td>
<td>0.12</td>
<td>0.10</td>
<td>-0.07</td>
</tr>
<tr>
<td>A_THICK</td>
<td>-0.04</td>
<td>-0.45</td>
<td>0.10</td>
<td>-0.04</td>
<td>-0.12</td>
</tr>
<tr>
<td>ARIDITY_INDEX</td>
<td>0.26</td>
<td>-0.09</td>
<td>0.00</td>
<td>-0.03</td>
<td>-0.02</td>
</tr>
<tr>
<td>B_KSAT</td>
<td>0.16</td>
<td>-0.22</td>
<td>0.26</td>
<td>0.29</td>
<td>-0.08</td>
</tr>
<tr>
<td>B_PAWC</td>
<td>0.17</td>
<td>0.18</td>
<td>0.01</td>
<td>0.47</td>
<td>0.29</td>
</tr>
<tr>
<td>B_THICK</td>
<td>0.10</td>
<td>0.30</td>
<td>-0.10</td>
<td>0.35</td>
<td>0.42</td>
</tr>
<tr>
<td>BIO04</td>
<td>-0.21</td>
<td>-0.01</td>
<td>-0.26</td>
<td>0.26</td>
<td>-0.36</td>
</tr>
<tr>
<td>BIO08</td>
<td>-0.25</td>
<td>-0.01</td>
<td>0.19</td>
<td>0.09</td>
<td>0.07</td>
</tr>
<tr>
<td>BIO09</td>
<td>0.02</td>
<td>-0.26</td>
<td>0.35</td>
<td>-0.26</td>
<td>0.52</td>
</tr>
<tr>
<td>BIO15</td>
<td>0.26</td>
<td>-0.09</td>
<td>0.00</td>
<td>-0.11</td>
<td>0.10</td>
</tr>
<tr>
<td>BIO16</td>
<td>0.26</td>
<td>0.01</td>
<td>-0.19</td>
<td>0.03</td>
<td>-0.13</td>
</tr>
<tr>
<td>BIO17</td>
<td>-0.04</td>
<td>0.38</td>
<td>0.38</td>
<td>-0.08</td>
<td>-0.01</td>
</tr>
<tr>
<td>ELEVATION</td>
<td>0.24</td>
<td>0.18</td>
<td>0.11</td>
<td>-0.11</td>
<td>-0.01</td>
</tr>
<tr>
<td>ET_ANNUAL</td>
<td>0.21</td>
<td>-0.10</td>
<td>-0.41</td>
<td>-0.09</td>
<td>0.10</td>
</tr>
<tr>
<td>GW_SWL</td>
<td>0.23</td>
<td>0.11</td>
<td>0.20</td>
<td>0.05</td>
<td>-0.24</td>
</tr>
<tr>
<td>GW_TDS</td>
<td>-0.16</td>
<td>0.11</td>
<td>0.41</td>
<td>0.13</td>
<td>-0.25</td>
</tr>
<tr>
<td>LDI</td>
<td>-0.15</td>
<td>-0.08</td>
<td>-0.26</td>
<td>0.18</td>
<td>0.08</td>
</tr>
<tr>
<td>MAX_TEMP</td>
<td>0.26</td>
<td>-0.09</td>
<td>0.01</td>
<td>-0.03</td>
<td>0.00</td>
</tr>
<tr>
<td>MIN_TEMP</td>
<td>0.25</td>
<td>0.02</td>
<td>0.07</td>
<td>0.08</td>
<td>-0.21</td>
</tr>
<tr>
<td>RAIN_ANNUAL</td>
<td>0.27</td>
<td>-0.02</td>
<td>0.04</td>
<td>0.01</td>
<td>-0.07</td>
</tr>
<tr>
<td>SLOPE_RAD</td>
<td>-0.26</td>
<td>0.09</td>
<td>0.05</td>
<td>0.10</td>
<td>-0.04</td>
</tr>
<tr>
<td>SOIL_EROS</td>
<td>-0.24</td>
<td>-0.16</td>
<td>-0.07</td>
<td>0.13</td>
<td>0.01</td>
</tr>
<tr>
<td>TWI</td>
<td>-0.23</td>
<td>0.00</td>
<td>0.03</td>
<td>-0.06</td>
<td>0.10</td>
</tr>
<tr>
<td>WEATH_IND</td>
<td>-0.19</td>
<td>-0.17</td>
<td>-0.03</td>
<td>0.11</td>
<td>0.27</td>
</tr>
</tbody>
</table>
5.3.3 Random forests

Two hundred trees were sufficient for the RF models to achieve acceptable accuracies. Exploratory analysis, using models with 5000 trees (not presented) showed little improvement in OOB error (<1%). OOB error rates were low for the ALOC 23 and ALOC 23 PCA RF models, with estimated maximum accuracies of 95% and 92%, respectively. The OOB error rate of the ALOC 20, ALOC 20 PCA and the 100% ALOC 20 and 100% ALOC 20 PCA models (the latter two were created due to the small sample size for the 10-km dataset as described above) was significantly worse, with estimated accuracies of 59%, 56%, 56% and 54%, respectively (Figure 5.4).

Classification accuracy was 95% (κ = 0.94) for the ALOC 23 classification and 92% (κ = 0.92) for the ALOC 23 PCA classification (Appendix I, Table 9.4) when tested against the validation dataset. The accuracy of the ALOC 20 and ALOC 20 PCA classifications decreased relative to those estimated by the RF OOB error, with accuracies of 46% (κ = 0.42) and 47% (κ = 0.44) (Appendix I, Table 9.4). The producer accuracies differed significantly for each of the classifications (Appendix I, Table 9.4), with observed minimum producer classification accuracies of 81% for the ALOC 23 classification and 59% for the ALOC 23 PCA classification. Likewise, the ALOC 20 and ALOC 20 PCA classification also exhibited low producer accuracies with minima of 0% observed for a number of classes in each classification. Visual inspection of the resulting classifications showed few obvious differences among the various ALOC 23 classifications (Figure 5.5), but more differences were apparent among the ALOC 20 classifications (Figure 5.6).
Figure 5.4: Out-Of-Bag (OOB) percent accuracies for the ALOC clusterings as classified by random forests. The 30-m ALOC 23 and ALOC 23 PCA models were significantly more accurate than the 10-km ALOC 20 classifications.
Figure 5.5: Results of the ALOC 23 (30 m) classifications. Top row - ALOC 23 and ALOC 23 PCA; Bottom row - ALOC 23 and ALOC 23 PCA resampled to 2.5 km. Colours represent each of the ALOC non-hierarchical clusters. Similar colours and cluster numbers do not necessarily represent related groups.
Figure 5.6: Results of the ALOC 20 (10 km) classifications. Top row - ALOC 20 and ALOC 20 PCA; Bottom row - ALOC 20 (100 %) and ALOC 20 PCA (100 %). Note that not all ALOC clusters are present in the final classifications. Colours represent each of the ALOC non-hierarchical clusters. Similar colours and cluster numbers do not necessarily represent related groups.
5.3.4 Comparisons between original, PCA and resampled classifications

The class distributions of the RF and resampled classifications illustrate that the RF classifications at 30-m and resampled to 2.5-km (bottom row, Figure 5.5) were quite successful in maintaining the distribution of the validation dataset (Figure 5.7). All 10-km models were missing classes 7, 8, 12, 18, 19 and 20 from the validation dataset (Figure 5.7), meaning that accuracy assessment of these classes was not possible, although classes 7, 8, 12, and 18 were present in the final classification (Figure 5.6). Class 1 tended to be over-classified by the ALOC 20 RF models, as evidenced by the large number of validation points classified as such (Figure 5.7). The ALOC 20 100 % classification was also the only classification to have classes 6 and 15 represented at the validation locations.

Agreement between the ALOC 23 and ALOC 23 PCA classifications was high at 93 % ($\kappa = 0.93$). There was no statistically-significant difference between the two classifications ($\chi^2 = 2.72, P = 0.1$). Agreements for the ALOC 20 and ALOC 20 PCA classifications was lower than that observed for the ALOC 23 classifications, at 72 % ($\kappa = 0.67$), which was also non-significant ($\chi^2 = 0.24, P = 0.63$). This indicates that the classification based on the five principal components captured the vast majority of the variability among points at the 30-m scale, and most of the variability at the 10-km scale.
Figure 5.7: Class distributions from the final classified images. The 20 % validation points were used to extract class information from the classified images. The resulting figure shows that the ALOC 23 classifications were much better at maintaining the class distribution of the validation dataset (shown above in red) than the ALOC 20 classifications.
Visual inspection of the resampled classifications (bottom row, Figure 5.5) showed very similar results to the original 30-m classifications. Agreement between the resampled ALOC 23 classification and the original 30-m classification was high (92 %, $\kappa = 0.92$) and an accuracy of 91 % ($\kappa = 0.9$) was observed with the 300 randomly chosen points from the validation dataset. This was a 4 % difference in accuracy to the original 30-m ALOC 23 classification. However, McNemar’s test suggested that the difference in classification accuracy was statistically significant ($\chi^2 = 12.56, P < 0.001$). As the two classifications are not independent, the permutation test was also used to compare the two. This supported the results of the McNemar’s test and a permutated $\kappa$ difference of 0.062 was deemed significant ($P < 0.001$). The ALOC 23 PCA resampling results were similar, with a relatively high agreement with the original 30-m PCA classification (88 %, $\kappa = 0.87$), and 87 % ($\kappa = 0.86$) agreement with the validation dataset, a 5 % difference compared to the original ALOC 23 PCA classification. These differences were also statistically significant (McNemar’s test, $\chi^2 = 11.76, P < 0.001$; permutation test, $\kappa$ difference = 0.073, $P < 0.001$). Agreement was quite high between the 2.5-km resampled ALOC 23 and 2.5-km resampled ALOC 23 PCA classifications (93 %, $\kappa = 0.92$), however unlike the 30-m classification results, when comparing the agreement between the two resampled classifications, McNemar’s test suggested the results were significantly different ($\chi^2 = 5.88, P = 0.015$). As these two classifications were independent of one another the permutation test was not required.
Once the classification assessment was finalised, the meta-group allocations were appended to the classifications. The ALOC 23 meta-group allocations were examined visually (Figure 5.8) and were deemed suitable given that regions having similar groups (e.g. A and B, represented by distinct colours in Figure 5.8) exist in similar areas among the two classifications and show quite obvious spatial relationships. While the performance of the ALOC 20 models was weaker, the meta-group assignment results were similar to those observed for the ALOC 23 classification in that similar meta-groups existed closer spatially (not presented).
Figure 5.8: Results of the ALOC 23 (30 m) meta-group allocations. Top row - ALOC 23 and ALOC 23 PCA; Bottom row - ALOC 23 and ALOC 23 PCA resampled to 2.5 km. Colours represent each of the hierarchical meta-groups as defined by SIMPROF. Similar colours and group letters indicate a closer relationship than those further apart.
5.3.5 Spatial variability in western Victoria

The variability of model results was not uniform across the entire Glenelg-Hopkins case-study region (Figure 5.9). While the eastern and central parts of the region appeared to be relatively spatially uniform (i.e. they do not show a significant amount of variation in ALOC groupings); the western and north-western parts of the region are comprised of a number of ALOC classes. This suggests that the hydrological system varies spatially across the region, with the most variability likely occurring in the western half of the catchment.

The meta-group classifications support the idea that there was a difference in the hydrological systems of the three major catchments of the Glenelg-Hopkins region (Figure 5.10). Interestingly, the amount of spatial variation did not decrease when the meta-group assignments of the original ALOC classes were examined, suggesting that hydrologic responses could be very different in areas that are quite close together. In the Glenelg River catchment (Figure 5.10, shown in red), there is obvious spatial variation in the assigned hydrological classes. Of particular interest are the two catchments in the north-east of the catchment that contain Rocklands Reservoir and the majority of the Grampians ranges, as they each consist of five different hydrological meta-groups (E, F, H, I & J). Even though the meta-groups represented in those particular catchments occupy a small area, they are still all present in the resampled classification (Figure 5.10, bottom row). While the differences were less pronounced in the Hopkins River catchment (Figure 5.10, outlined in black), there was some spatial variability in classes in the north (dominated by meta-group I, with some small patches comprised of meta-group E), while most of the catchment belongs to meta-group I and the two
southernmost catchments belong to meta-group G. There was even less variation in the classification of the Portland catchment (Figure 5.10, outlined in orange) with meta-group I dominating that catchment. Nonetheless, there were small areas of meta-groups F and J in the south-west catchments of the catchment. Visual examination of the PCA and the resampled classifications showed little difference to that observed in the original 30-m classifications. The most obvious change was the small area in the southern catchment of the Hopkins River catchment (outlined in black), and the easternmost parts of the Portland catchment (outlined in orange), that was classified as meta-group J in the PCA classifications.
Figure 5.9: Results of the ALOC 23 (30 m) classifications for the Glenelg-Hopkins region. Top row - ALOC 23 and ALOC 23 PCA; Bottom row - ALOC 23 and ALOC 23 PCA resampled to 2.5 km. Colours represent each of the ALOC non-hierarchical clusters. Similar colours and cluster numbers do not necessarily represent related groups. Higher variability is obvious in the NW side of the region, compared to the east and south. Red: Glenelg Catchment; Black: Hopkins Catchment, Orange: Portland Catchment.
Figure 5.10: Results of the ALOC 23 (30 m) meta-group allocations for the Glenelg-Hopkins region. Top row - ALOC 23 and ALOC 23 PCA; Bottom row - ALOC 23 and ALOC 23 PCA resampled to 2.5 km. Colours represent each of the hierarchical meta-groups. Similar colours and group letters indicate a closer relationship than those further apart. Red: Glenelg Catchment; Black: Hopkins Catchment, Orange: Portland Catchment.
5.3.6 Relationships between the regionalisation and hydrologic indices

A total of 201 gauges were deemed suitable for inclusion in the analysis based on the criteria specified above (Section 5.2.7). With the exception of groups A, K and H, all regions had multiple suitable gauges. There were significant differences among groups based on their hydrologic indices for both the original (pseudo-$F_{1,7} = 3.655$, $P = 0.001$) and the bootstrapped datasets (pseudo-$F_{1,7} = 9.304$, $P = 0.001$). *A posteriori* pairwise comparisons of groups using the original dataset indicated that the differences were significant ($P \leq 0.05$) between all groups with the exception of pairs B:E, C:D, C:E and F:I. The pairwise comparisons on the bootstrapped dataset were all significant ($P < 0.05$).

The ability of CAP to correctly classify cases within the original dataset based on their hydrologic indices was relatively poor, with only 92 samples correctly classified, but the model was statistically significant (48 %, $m = 30$, $P = 0.001$). Stream gauges from meta-group C had the lowest classification accuracy with only 10 % of gauges being successfully allocated. The highest classification accuracy was observed for both meta-groups B and G, with 67 % of gauges correctly allocated to each. The bootstrap dataset performed better with 253 samples being correctly classified (66 %, $m = 28$, $P = 0.001$). The lowest classification accuracy was observed for meta-group E with only 47 % of gauges being correctly allocated. The highest classification accuracy was meta-group I with 74 % of gauges correctly allocated.

When 20 % of cases were used as a validation sample within the bootstrapped dataset, CAP performed reasonably, with 208 of the gauges correctly classified (67 %, $m = 32$, $P = 0.001$). The lowest classification accuracy was observed for meta-group J with
58% of gauges being correctly allocated. The highest classification accuracy was meta-group F with 77% of gauges correctly allocated. Agreement between the samples that were allocated to new groups and their original group was relatively high at 60% ($\kappa = 0.53$). The lowest successful group allocation was observed for meta-group E where only 20% of gauges were correctly classified. The highest allocation accuracy was observed for meta-group D where 100% of gauges were correctly classified.

Pearson’s correlation indicated there was a statistically-significant positive relationship between the number of gauges in each group and the number of gauges classified correctly from each group (Appendix I, Figure 9.9) for the original ($r = 0.91, P = 0.002, n = 8$), the bootstrapped dataset ($r = 0.93, P = 0.001, n = 8$), and the bootstrapped dataset with the validation samples ($r = 0.96, P < 0.001, n = 8$). This indicated that the model was most likely to correctly classify groups that were common in the dataset, with uncommon groups being correctly classified less often. All classifications performed highly favourably compared to a random allocation of cases among 11 groups which, assuming equal sample sizes, results in 9% of cases being correctly classified.

5.4 Discussion

By incorporating a number of environmental variables likely to influence regional hydrology and a range of non-parametric statistical and classification methods, this study aimed to generate a hydrologic landscape classification that did not require the use of an a priori selection of a spatial unit such as a catchment. The main objective of the approach was to see whether it was possible to create a classification that could preserve the environmental and hydrological variability that are known to influence
streamflows within and among catchments that has typically been lost in previous regionalisation studies. An analysis of the ability of the classification to differentiate between streams from each group based on a number of streamflow indices was also undertaken.

5.4.1 Differences to previous regionalisation studies

Inductive methods of hydrologic regionalisation have been popular in the past (see Table 2 in Olden et al., 2012) and, while there have been a number of studies that have focused on deductive methods (see Table 1 in Olden et al., 2012), the choice of variables, their resolution (temporal and spatial), the classification method, the spatial scale of the classification and the number of groups are all known to influence deductive classifications (Sawicz et al., 2011; Olden et al., 2012). While the final product of deductive methods is a spatial mosaic of independent hydrologic types, the final classifications do not always only identify hydrologic variation (Snelder et al., 2005; Carlisle et al., 2010). Inter-catchment variability can limit the applicability of hydrologic regionalisations to generalise and predict catchment behaviour as a function of climatic and environmental gradients (Sawicz et al., 2011). As previous studies have relied on catchments, landscape units or stream sections (Wolock et al., 2004; Stein et al., 2009; Sawicz et al., 2011; Olden et al., 2012), an issue that is more apparent in deductive regionalisation studies is the loss of small-scale spatial hydrologic variability (Poff et al., 2006; Kennard et al., 2010b) as the unit of analysis gets larger. Our method relied on using an accurate supervised image classification method to extend our statistical clustering to an area covering ~228,000 km² without the need to rely on catchments, landscape units or stream sections. We opted for this approach as it is well known in the
literature that there is significant flow variability within and among catchments and that the variables governing flow variability are dependent on scale (McManamay et al., 2012). While other regionalisation studies have identified that the primary drivers of catchment function are largely related to climatic gradients (Sawicz et al., 2011), our results suggest that a mixture of climatic, geological and environmental functions are driving catchment, and thus hydrologic, variability (Appendix I, Tables 9.2 and 9.3, Figure 9.8) in our regionalisation. It would be expected in a traditional approach to hydrologic regionalisations that some of this variability would be lost – which could explain why other studies have largely identified climatic gradients that vary slowly with space to be the primary drivers of catchment function. Our approach allows for different classes to be represented within a single catchment, thus preserving intra-catchment variability.

5.4.2 Statistical evaluation, clustering, and PCA

Traditional parametric statistical analyses and clustering algorithms such as $k$-means tend to have restrictive assumptions regarding independence of samples, multivariate normality and collinearity. The assumption of samples being distributed normally through multivariate space, for example, is unlikely to be true in most ecological and environmental datasets (Belbin and McDonald, 1993). The approach that we employed relied on the use of non-parametric and permutation-based statistical methods in conjunction with the RF classification algorithm. This approach had far fewer assumptions relating to data normality and collinearity (Clarke, 1993; Breiman, 2001; Clarke et al., 2008). The approach also allowed the decision regarding the number of groups used in the analysis to be statistically-justified, when this decision is typically
arbitrary. Our method was supported by the application of both ANOSIM and MDS to assess group separation, with each suggesting that the groups were distinct and that we had chosen an appropriate number of groups for our dataset. While not perfect (SIMPROF, by design, tests for hierarchically-related groups and we were after non-hierarchical groups), we believe this approach to be simpler, more statistically sound and more efficient than methods employed in the past which require large amount of *a posteriori* or *post-hoc* statistical testing (Snelder et al., 2005; Stein et al., 2009).

The use of PCA-transformed data was shown not to significantly affect the classification accuracy of the model, even though the PCA was only able to explain <80% of the data variability (Table 5.1). This suggests that future classifications could be conducted on PCA-transformed datasets while still producing accurate classification schemes. Our method has essentially shown that it is possible to extract the same number of groups from PCA-transformed data as it was from the non-PCA transformed data. However, using PCA from the beginning has the potential to influence the number of groups identifiable by SIMPROF (as there are fewer data and less variance in the remaining data) and therefore influence the overall classification process. Using PCA could, however, make the process more efficient in that having a reduced number of groups to begin with could remove the need to first use non-hierarchical classification before hierarchically classifying the groups. The major benefit of using PCA-transformed data in this study was that the time to parameterise and classify the raster data with RF was substantially reduced.
5.4.3 Classification by random forests

The non-parametric and highly accurate RF classifier (Breiman, 2001; Rodriguez-Galiano et al., 2012) was very successful in recovering and classifying the ALOC class information of the remaining pixels in the raster datasets. While the ALOC 23 RF models had very high classification accuracies (95 % [κ = 0.94] and 92 % [κ = 0.92]), our hypothesis that the classification of the coarser datasets would be inferior was confirmed by the low accuracies of the ALOC 20 classifications (46 % [κ = 0.42] and 47 % [κ = 0.44]). As our sampling density was severely limited by pixel size in the 10-km models, this further supports previous research showing that the overall and per-class accuracy of supervised classifications can be limited by pixel size (Mumby and Edwards, 2002; Sprintsin et al., 2007). While the RF classifier has been shown to be robust against statistical noise and training data reduction (Rodriguez-Galiano et al., 2012), it is possible that, in this case, there were simply not enough training sites to allow for the creation of an appropriate model at the 10-km scale. This analysis used only 325 training sites (80 % of \( n = 406 \) 10-km sample points) to try to produce a classifier for 20 classes and another 33 % of the training site data was excluded for OOB accuracy assessment (Rodriguez-Galiano et al., 2012). This left the RF algorithm with only 215 points to generate the required classification trees. The two ALOC 20 100 % models used 268 points (66 % of \( n = 406 \) 10-km sample points) to create an RF model but still performed poorly with OOB error rates limiting accuracies to 56 % and 54 %. As all classes were included in training data for the ALOC 20 100 % models, their omission from the final classification (Figures 5.6 and 5.7) suggests that they, by chance, happened to be excluded from the training data selected by the bootstrapping step used
to calculate OOB error. This could partly explain the high OOB error rates observed for those models. The missing classes from the ALOC 20 classifications (Figure 5.7; Appendix 1, Table 9.4) could be similarly explained, although it is also possible that they were excluded randomly from the 80% training data at the previous step. As expected, the exclusion of data, either manually for validation purposes or automatically by RF to enable an OOB estimate, appeared to severely limit the classification accuracy when using small amounts of training data (Rodriguez-Galiano et al., 2012). This further supports our hypothesis that limiting the study to a 10-km resolution based on a single, coarse dataset would influence the results presented here, particularly as this study was conducted over a relatively small area. If the study had been conducted at a continental scale, for example, it would be possible to generate more than 15,500 random points at a minimum distance of 10 km and, therefore, resampling the same datasets to a finer resolution (to avoid sub-sampling of pixels) would not be necessary.

The resampling of the ALOC 23 models from a 30-m to 2.5-km resolution appeared to remove some of the finer-scale spatial variability in the classifications (Figures 5.5 and 5.8). While not appearing to constitute a significant change between the original and resampled classifications (percent agreements between the resampled and original classifications were 92% [κ = 0.92] for the ALOC 23 model and 88% [κ = 0.87] for the ALOC 23 PCA model), the results of both McNemar's test and the permutation test indicated that the resampled classifications were significantly different from their equivalent 30-m classifications. This may seem like a serious drawback to the method, however, when compared to the validation dataset the accuracy of the 2.5-km ALOC 23 classification was only 4% less than that observed for the 30-m model, while
the accuracy of the 2.5-km ALOC 23 PCA classification was only 5% less than that of the 30-m PCA model. While statistically the difference may be deemed significant, we contend that in reality a difference of ≤ 5% would likely not be ecologically or environmentally important and thus, would not affect the ability of the method to create a hydrological landscape classification that could be used to explain spatial differences in streamflow metrics. Additionally, the resampling step was performed \textit{a posteriori} and therefore may not actually be necessary in all cases.

5.4.4 Case study on spatial variability in western Victoria

The spatial variability in the classification of the Glenelg-Hopkins region was most evident in the Glenelg catchment, and less so in the Hopkins and Portland catchments (Figures 5.9 and 5.10). Spatial hydrologic variability has been observed in a number of studies in the past and the strongest and most variable relationships between environmental factors and water quality and quantity have consistently been found in the Glenelg catchment. Modelling results from Chapter 3 found that the Glenelg catchment exhibited the most variability in the relationships between climate, land cover and wetland extent, which may help to explain some of the variability observed here. Relationships explored in the past relating the proportion of native vegetation to in-stream salinity (Versace \textit{et al.}, 2008b) showed strong relationships in the Glenelg and Portland catchments, but less evidence for the same relationships in the Hopkins catchment and it was suggested that this may have been due to the degraded nature of the catchment. The degraded nature of the Hopkins catchment could also explain the lack of variability observed in that catchment in this study. A spatially-varying relationship between nutrient exports and land cover has also been observed in the...
Glenelg-Hopkins region (Ierodiaconou et al., 2005) although, due to the lack of suitable stream-gauge data, it is not clear whether this relationship holds for the streams and rivers of the region. However, the relationships in the region that have been described in the past (Chapters 3 and 4; Ierodiaconou et al., 2005; Versace et al., 2008b) are likely to be complicated due to variations in geomorphology, groundwater levels and salt concentrations (Yihdego and Webb, 2011); conditions that we have attempted to account for in this study. Water resource managers in the region need to take into account possible differences in intra- and inter-catchment hydrology that could drastically affect river management and restoration plans and regionalisation studies such as the one presented here could assist in identifying that variability.

5.4.5 Relationships between the regionalisation and hydrologic indices

Understanding, and being able to accurately predict, streamflow characteristics in ungauged locations is crucial for ecohydrological and other studies (Kennard et al., 2010b; Poff et al., 2010). Our method set out to test a new approach to hydrological regionalisation that removed the need for catchments as a spatial unit of analysis for our statistical clustering (Wolock et al., 2004; Kennard et al., 2010b; Sawicz et al., 2011). However, the ability to link the results of the regionalisation to streamflow indices could have presented an issue given that we did not rely on catchments as a spatial unit. The results supported our hypothesis that our method would be able to identify and preserve inter- and intra-catchment variability. Pairwise comparisons suggested that, even in the original dataset \( n = 201 \), there was enough variability in our 32 streamflow indices to separate all but four pairs (of a total of 28 pairs). A simple bootstrap (with replacement) to \( n = 383 \) gauges was sufficient for all groups to be easily identified as distinct
suggesting there is a minimum number of gauges for the approach to be reliable (likely to relate to the number of each individual class in the dataset). In addition, CAP demonstrated that the regionalisation was able to discriminate among streams from different groups. While the classification was not perfect, the analysis conducted on the bootstrap datasets indicated that the stream gauges could be classified correctly significantly better than chance alone, and that gauges from one class could be correctly classified 100% of the time. In general, common classes were correctly classified more frequently than uncommon classes. While the correlation between the number of gauges in each class and the number of correctly-allocated samples was significant, there was no clear threshold (i.e. the relationship was linear), so it is difficult to identify a single minimum number of gauges that could be implemented to ensure that results met pre-defined criteria for reliability. Therefore, future users should interpret results for rare groups with caution. However, we believe that, based on this preliminary assessment of our method, the results illustrate that there is promise in the method for categorising regions, particularly in the absence of comprehensive streamflow data as is the case in many regions in Australia and elsewhere. Further investigation of the validation (i.e. linking the regionalisation to streamflow indices) using more in-depth data mining approaches (e.g. decision trees, Breiman, 1993) is likely to produce even greater classification success.

5.5 Conclusions

Hydrologic classifications are increasingly being employed in the management and research of aquatic resources. Our approach differed from inductive hydrological regionalisation where membership is defined quantitatively based on indices of stream
flow, and traditional deductive regionalisation which requires the use of catchments or other appropriate spatial units. Instead, membership of pixels was defined qualitatively with the random forest classifier based on a statistical classification of a number of environmental variables that we believe could have a direct influence on the hydrologic cycle. In essence, we present a method that allowed the creation of spatially-independent hydrological regions; these regions represent a series of fundamental hydrologic landscape units that exist in multiple locations depending on environmental similarity rather than a combination of environment and streamflow-metric response similarity. To our knowledge, the application of deductive reasoning and hybrid classification is a novel approach in hydrological regionalisation. This method has removed the need to rely on a spatial unit specified \textit{a priori}, such as a catchment or ecoregion, and has allowed the preservation of intra-catchment variability. Thus, it should be useful in the spatial explanation and prediction of streamflow responses that are known, or suspected, to vary within catchments. The ability of our regionalisation to discriminate among streams from different groups based on their range of flow indices highlights the value of this approach, particularly in regions where streamflow data are lacking.
6.1 Introduction

6.1.1 Prediction in ungauged basins

Streamflow prediction in ungauged basins (PUB) is a significant challenge for hydrologists. Beginning in 2003, the International Association of Hydrological Sciences began a shift from a focus on model form towards improved scientific understanding of hydrological processes. This initiative was concluded in 2012 and became known as the PUB decade (Hrachowitz et al., 2013). The purpose of the initiative was to recognise that, in data-scarce regions, it would be beneficial to be able to infer hydrological functions and controls from metrics related to catchment form – for example, the combined effects of climate, topography, geology, soil type and land cover – independent of existing streamflow data. Previously, much of scientific hydrology had been driven by an attempt to understand whether physically-based, index-based or conceptual models would be better suited for reproducing hydrological processes across a comprehensive variety of gauged and ungauged catchments (Hrachowitz et al., 2013). The ability to accurately predict streamflow regimes in ungauged sites is not only important for water resources management to estimate yields, storage and the behaviour of streams during and after extreme climatic events (Sanborn and Bledsoe, 2006), but is increasingly important in ecohydrology (Poff et al., 1997; Bunn and Arthington, 2002; Kennard et al., 2010a). A streamflow regime consists of components relating to the magnitude, timing, duration and frequency of average, high- and low-flow conditions, the rate of change of streamflow, and inter- and intra-annual variation (Poff et al., 1997). Biotic communities and abiotic conditions can be influenced by long-term trends in...
streamflow regime, which can create and maintain ecosystem dynamics for a range of
ecologically-important conditions both temporally and spatially (Poff et al., 1997; Bunn
and Arthington, 2002; Lester et al., 2014).

6.1.2 Regression and machine-learning methods for ungauged flow prediction

In many instances, gauged streamflow records are limited or not available at
sites of interest, particularly on small streams and, as such, typical problems facing
ecohydrologists are: 1) the development of streamflow estimates for sites where no
records exist; and 2) the extension of short-term records. Both problems can be
addressed by simple regression methods; however these estimates rely on having gauged
streamflow data from nearby hydrologically-similar catchments (Gordon et al., 2004).
Some methods allow for the grouping of large and small catchments but, as a rule,
catchments of different sizes typically behave very differently hydrologically (Gordon et
al., 2004). For example, mean annual flow for an ungauged site can be estimated from
nearby gauged sites by adjusting for differences in area:

\[
\bar{x}_i = \bar{x}_2 \left( \frac{A_1}{A_2} \right)^a
\]

where \(\bar{x}_i\) = mean annual flow (volume units) for the ungauged site, \(\bar{x}_2\) = mean annual
flow (volume units) for the gauged sites, \(A_1\) and \(A_2\) are the areas of the ungauged and
gauged catchments, and \(\alpha\) is a calibration constant that is generally less than 1.0
(McMahon, 1976).
Likewise, the coefficient of variation in streamflows can be estimated by adjusting for differences in mean annual runoff:

\[ C_{V_i} = C_{V_g} \left( \frac{\bar{X}_i / A_i}{\bar{X}_g / A_g} \right)^b \]

where \( C_{V_i} \) and \( C_{V_g} \) are the coefficient of variation in streamflow for the ungauged and gauged sites, respectively, and \( b \) is a calibration constant that is generally less than 0 (McMahon, 1976). Definitions for the other symbols are consistent with the previous formula.

Historically, multiple regression methods have also been used extensively in hydrology (Holder, 1985). Recently, they have been used to model water quality in streams and lakes (Sheela et al., 2011), assess relationships among land cover, climate, soils and wetland extents (Chapters 3 and 4), relate baseflow to catchment properties (Mazvimavi et al., 2005), predict average annual and monthly runoffs (Mazvimavi et al., 2005), and to explore relationships between topography and precipitation (Um et al., 2011). The relative simplicity of regression approaches can be advantageous when examining first-order (e.g. topographical) control mechanisms on catchment function in data-scarce regions (Hrachowitz et al., 2013). The PUB decade, however, brought about an increased understanding of the importance of developing more complex models to better understand the patterns and dynamics of the underlying processes in both gauged and ungauged catchments (Hrachowitz et al., 2013).

While simple relationships and multiple regressions have proven useful in the past for PUB, recent work has seen a shift away from classic statistical techniques to more advanced machine-learning methods such as neural networks (Mazvimavi et al.,
2005), classification and regression trees (CART) (Kennard et al., 2010b; Poor and Ullman, 2010; McManamay et al., 2012), random forests (RF) (Carlisle et al., 2010; Rodriguez-Galiano et al., 2014) and support vector machines (Raghavendra and Deka, 2014). Machine-learning methods can be more appropriate than classical statistical models (e.g. regression analysis) when the emphasis is on prediction rather than explanation, particularly when the predictor variables are correlated and there are many complex, possibly non-linear interactions which typically violate the assumptions of methods such as multiple regression (Mazvimavi et al., 2005).

6.1.3 Identifying associations between catchment form and function

One of the key concepts arising from the PUB decade was that an improved understanding of the associations between catchment form and function, based on inductive or emergent properties (i.e. streamflow characteristics or catchment signatures), could be used as first steps towards functional catchment classification schemes (Hrachowitz et al., 2013). However, catchment heterogeneity has hindered attempts at classification and has made it difficult to understand and explain the spatio-temporal variation and control of hydrological processes at the catchment scale (Hrachowitz et al., 2013). Wagener et al. (2007) suggested that a descriptive and potentially predictive framework could be developed based on the concept of catchment function, including an explicit link among climate, landscape attributes and streamflow indices. Recent attempts to apply this approach revolve around the use of hydrologic regionalisations, which attempt to group streams or landscapes based on a suite of physical and climatic attributes that ideally produce similar hydrologic responses in stream systems.
Chapter 5 and Kennard *et al.* (2010b) both presented methods that utilised statistical clustering of a range of environmental, geological, topographical and meteorological data, to create a hydrological regionalisation, and attempted to relate their regionalisations to regional streamflow indices – both with mixed success. McManamay *et al.* (2012) found that a hydrological regionalisation by Poff (1996) explained only 7% to 39% of the variation in a number of streamflow indices, but that a flow classification approach was able to explain 9% to 87% of the variation in the same indices. However, it has been demonstrated that a hydrologic regionalisation when combined with stream classification can help in the prediction of streamflow metrics (Snelder *et al.*, 2005; Santhi *et al.*, 2008). Hydrologic regionalisations have also been shown to improve predictive streamflow models when those models are stratified by hydrologic regions (Carlisle *et al.*, 2010). The creation of these regionalisations allows some inferences to be made regarding hydrologic process relationships even in the absence of rigorous mechanistic models (Hrachowitz *et al.*, 2013).

### 6.1.4 Aim of study

This chapter investigates the suitability of using CART and random forests to improve the validation of a previously-developed hydrologic regionalisation (Chapter 5). An initial validation established that there was enough variation in the streamflow indices to suggest that all classification groups were significantly different from one another and, when using the indices to predict group membership, a constrained discriminant analysis was able to achieve overall classification accuracies of 67%. However, this estimate was likely over-optimistic and a more conservative estimate of classification success was 48% (Chapter 5, Section 5.4.5). The validation conducted
here uses more advanced machine-learning techniques to improve upon the previous validation and provide more evidence of the utility of the hydrologic regionalisation framework created in Chapter 5. Additionally, a new classification and validation was conducted that assesses the ability of the regionalisation variables (i.e. the variables that were used to create the original regionalisation) to discriminate among classes in a classification based on indices of long-term streamflow. This approach allows an assessment of the relative merits of both approaches (i.e. hydrological regionalisation and flow classification), and identifies any potential shortcomings of the regionalisation approach utilised in Chapter 5. A comparison of the predictive power of CART and RF is also conducted for the respective classification (i.e. regionalisation and flow-based) models.

6.2 Methods

6.2.1 Site description

Victoria is the southernmost state of mainland Australia, comprising an area of approximately 228,000 km² (Figure 6.1). Topographically, geologically and climatically, Victoria is diverse, varying from flat semi-arid plains to the west and northwest, wet temperate climates in the southeast, to alpine areas rising to ~2000 m altitude in the northeast. Median annual rainfall is less than 300 mm in a large part of the west and northwest but exceeds 2,500 mm in some parts of the mountainous northeast (Department of Environment and Primary Industries, 2013a). Most of the state experiences a Mediterranean climate consisting of hot, dry summer and cool, wet winters (Department of Environment and Primary Industries, 2013b).
Figure 6.1: Location of the study site in south-eastern Australia. Dark blue lines represent perennial rivers, while the colour gradient represents elevations, with darker browns indicating higher elevations. Catchments used in the analyses are outlined in black, with gauge (n = 201) locations marked in red.
6.2.2 Development of the hydrological regionalisation

Chapter 5 created a hydrological regionalisation using 25 variables that could explain the storage, movement, and quality of surface water, groundwater, and atmospheric water. The method worked by classifying individual pixels into a number of independent groups based on multivariate similarity and, as such, removed the need to rely on a spatial unit specified a priori, such as a catchment or ecoregion, which allowed the preservation of intra-catchment variability which is typically lost in regionalisation studies. The regions were first defined statistically for individual pixels using a non-hierarchical classification scheme based on 25 variables related to climate, geology and topography, which would be expected to influence regional hydrology (Chapter 5 and Section 6.2.6). The use of a hybrid classification (employing RF) extended this pixel-based statistical clustering to the rest of the state without the need to rely on catchments, landscape units or stream sections. The ability of RF to extend the statistical clustering was excellent, and validation of classification groups using a constrained discriminant analysis allowed for a rapid assessment of the discrimination of groups based on 32 streamflow indices. Based on this initial assessment of the method, results from Chapter 5 suggested that, in the absence of comprehensive stream gauge data which has traditionally been used to define catchments for hydrological classifications (Poff, 1996; Kennard et al., 2010b), the method was suitable for categorising regions that should behave similarly hydrologically.

6.2.3 Streamflow indices

In order to characterise different aspects of the flow regimes, 32 indices (Appendix II, Table 10.1; Table III, Olden and Poff, 2003) were created for 201 stream
gauges in Victoria using the Time Series Analysis module of the River Analysis Package (Marsh et al., 2003). Indices that were related to discharge (i.e. those divided by catchment area) were not included in the analysis conducted in Chapter 5. Details of the method used for site selection, record cleaning and metric calculation can be found in Chapter 5 (Section 5.2.7).

6.2.4 Flow classification

A new classification was undertaken to compare with the regionalisation approach utilised previously. Based on the 33 streamflow indices and the catchment area of each gauge (km²), 201 streamflow gauges were classified into eight non-hierarchical groups using the ALOC algorithm and the Gower metric in PATN v3.1.2 (Blatant Fabrications Pty. Ltd., 2009). These groups were then hierarchically classified based on their multivariate group averages. This step was necessary to identify classes that were more closely related to each other than those that were not. An ANalysis Of SIMilarities (ANOSIM; Clarke, 1993) test, based on a standardised Euclidean distance matrix, was used to test for significant differences among the hydrological classification groups. The test uses the $R$ statistic, which is scaled to be between -1 and +1; global $R$ values > 0 indicate greater dissimilarity among groups than within groups. Additionally, to further test for significant differences among the flow classification groups, a permutation-based multivariate ANOVA (PERMANOVA; Anderson, 2001) was conducted using flow classification group as a fixed factor. Analyses tested both for main effects and pairwise differences among groups.
6.2.5 Variable contributions to the classifications

To examine the relative contribution of each of the predictor variables (see Section 6.2.6 below) to both the regionalisation and flow classification groups, the SIMilarity PERcentage (SIMPER; Clarke, 1993) routine in PRIMER 6 was conducted on standardised Euclidean-distance similarity matrices. A Kruskal-Wallis test was used to evaluate the ability of each of the indices to differentiate among classification groups.

6.2.6 Modelling approaches

Two datasets were used during the modelling used to validate the regionalisation – a flow indices dataset which was used to predict the regionalisation groups, consisting of the mean annual runoff of each catchment (m$^3$ s$^{-1}$ km$^{-2}$) and the 32 indices described above (in a similar fashion to that Chapter 5) and an environmental dataset that was used to predict the newly-created flow classification. The variables for the environmental dataset were related to catchment properties which would be expected to influence regional hydrology. Variables included: multiple soil properties such as horizon thickness and saturated hydraulic conductivity; climate variables such as mean temperatures, rainfall and bioclimatic variables; and landscape variables such as topographic wetness index and elevation. Details on the 25 variables in the environmental dataset can be found in Appendix I (Table 9.1). The mean and standard deviation for each of the environmental variables was calculated for the catchments upstream of each gauge so as to allow intra-catchment variability to be modelled and assessed.

Three different modelling methods were applied to test the ability of the variables in each dataset to correctly classify gauges based on the hydrological
regionalisation and the flow classification. The first method was Canonical Analysis of Principal Coordinates (CAP; Anderson and Willis, 2003) which was employed in Chapter 5 for the preliminary validation of the hydrological regionalisation. The other two methods were more advanced machine-learning approaches, namely classification trees (CART; Breiman, 1993) and random forests (RF; Breiman, 2001). Although CART and RF were used on both classifications, CAP was not used to reclassify the regionalisation groups using the flow indices as that analysis had previously been conducted (Chapter 5, Section 5.2.7). For clarity, in the results and discussion, classes arising from the hydrological regionalisation are named using capital letters, while classes arising from the flow classification are named using Roman numerals.

6.2.6.1 CAP

Unlike unconstrained ordinations (e.g. PCA, MDS), constrained ordinations such as CAP have an a priori hypothesis which controls the interpretation of the multivariate data used to relate predictor to response variables (Anderson and Willis, 2003). For classification applications, the ordination axes are interpreted in such a way as to maximise the differences among groups identified a priori. CAP analyses were conducted using the PERMANOVA+ add-on for PRIMER (Anderson et al., 2008), the number of axes (m) was not specified and a permutation test (n = 999) was conducted to test the strength of the relationship.

6.2.6.2 CART

Tree-based models are fundamentally comprised of one or more nested if-then statements (splits) between the predictors in a given dataset. There are a number of
methods to control the depth and complexity of a model and one of the most common is the cost complexity parameter (cp), which is a penalising factor for growing too large a tree (Breiman, 1993). The idea behind tuning for cp is, for a specific value, to find the smallest “pruned tree” that gives the lowest penalised error (Kuhn and Johnson, 2013); i.e. it prevents a tree from growing too large and over-fitting. CART models were tuned independently for cp, which was limited to be between 0.001 and 0.1. Due to imbalances in the number of cases belonging to each group, the CART classification models included a misclassification cost, whereby the cost increased the further the classified group was from the true group where groups were assigned letters in order of their relative similarity (e.g. Classified as B, true group B, cost = 0, classified as C, true group B, cost = 1, etc.). CART trees were limited to a maximum depth of 30 branches, the minimum number of observations for a split was four and a minimum leaf node size of two was chosen. CART was used to predict both the gauge class from the regionalisation (using the flow indices as predictors), and the flow classification class (using the environmental variables as predictors) in two independent analyses.

6.2.6.3 Random forests

RF is an ensemble machine-learning method used in classification and regression (Breiman, 2001). As an ensemble method, multiple models (trees) are grown and used for prediction allowing the algorithm to obtain better performance than single decision-tree classifiers (Rodriguez-Galiano et al., 2012). Recently, RF has become more popular in both hydrology and ecohydrology (Peters et al., 2007; Rodriguez-Galiano et al., 2014). However, unlike CART models, models produced by RF are not interpretable – they are considered ‘black-box’. RF requires two parameters for generating a predictive
model: the number of trees \((k)\); and the number of random predictor variables used at each split during the growing of the trees \((mtry)\). Therefore, a dataset can be classified by defining a constant number of \(mtry\) variables, while each of the training samples is classified by \(k\) trees. Classification is determined by using the mode of the classes output by individual trees for each training site \((x)\) using the equation:

\[
\hat{C}_r^b = \text{majority vote}\left\{ \hat{C}_b(x) \right\}_{1}^{B}
\]

where \(\hat{C}_b(x)\) is the class prediction of the \(b\)th RF tree from a possible \(B\) classes (Rodriguez-Galiano et al., 2012). All RF models consisted of 1000 trees \((k = 1000)\) and were tuned independently for \(mtry\), which was limited to between 2 and \((x - 1)\), where \(x\) was the number of independent variables. The internal bootstrapping procedure of the RF classification models was stratified by group to account for imbalances in the number of observations per group (Breiman, 2001; Liaw and Wiener, 2002). RF was employed in the same manner as CART, i.e. to predict classes from both the regionalisation and flow classifications.

### 6.2.7 Model tuning and evaluation

The caret package (Kuhn, 2014) for R 3.1.2 was used as an interface to the rpart (Therneau et al., 2014) and randomForest (Liaw and Wiener, 2002) packages for the creation, parameter tuning and cross-validation of performance for each of the models used in this study. The most suitable variable for each split in the classification CART and RF models was chosen based on the Gini index (Breiman, 1993), which is a measure of the probability that a randomly-chosen sample would be incorrectly classified if it
were randomly classified according to the group distributions in the dataset. In this study, Kappa (κ) coefficients and percent agreement were used to assess the performance of CART and RF for the assessment of each of the regionalisation and the flow classifications. The best-performing model for each classification was chosen using the one-standard error rule (Breiman, 1993). Variable importance to the classification models was assessed based on the mean decrease of the Gini index that would occur if a variable was removed from the model – a larger mean decrease in Gini values indicates that a particular predictor variable plays a greater role in partitioning the data into the defined classes (Breiman, 1993).

Resampling methods, such as the bootstrap, can be used to produce appropriate, unbiased estimates of model performance using the training set when there are a limited number of observations (Kuhn and Johnson, 2013). Essentially, resampling methods create a model on a subset of the training samples and then assess performance using the remaining samples. Bootstrap resampling works by creating a new dataset equal in size to the original dataset; however the samples are generated with replacement. The samples not included in the new bootstrapped dataset are referred to as ‘out-of-bag’ samples. The model is fit using the bootstrapped dataset and is used to predict the out-of-bag samples (Kuhn and Johnson, 2013). The “0.632 bootstrap” (Efron, 1983) addresses the pessimistic bias issue of the simple bootstrap by creating an error estimate ($\hat{\text{Err}}^{(0.632)}$) that is a combination of the pessimistic simple bootstrap error (from the out-of-bag samples) and the (optimistic) error from re-predicting the training samples:

$$\hat{\text{Err}}^{(0.632)} = 0.368 \times \text{err} + 0.632 \times \hat{\text{Err}}^{(1)}$$
where $\overline{err}$ is the error estimate from re-predicting the training dataset and $\hat{Err}^{(1)}$ is the error estimate from the simple bootstrap. The 0.632 bootstrap reduces the bias of the original bootstrap, but can also be unstable with small sample sizes (Efron and Tibshirani, 1997). In this study, the 0.632 bootstrap was used as a resampling method for model tuning. One thousand bootstrap datasets were created and the best tuning parameter results ($cp$ or $mtry$) across the bootstrap datasets were then applied to the original (non-bootstrapped) dataset to create a final model (Kuhn, 2014). Differences between the models (CART and RF) were assessed with Bonferroni-corrected paired $t$-tests (Kuhn and Johnson, 2013; Kuhn, 2014) to evaluate whether the models had equivalent accuracies across the bootstrapped datasets. Wilcoxon signed-rank tests were also conducted to test for pairwise differences in group error rates between analogous CART and RF models. To test which classification (regionalisation or flow classification) had superior predictive ability, Welch’s $t$-tests were conducted between like models (CART or RF) to evaluate whether they had equivalent accuracies across the bootstrapped datasets. Finally, to investigate variance between the two classifications, an assessment was conducted using a confusion matrix to compare the distribution of cases among the regionalisation groups with the distribution of those cases among the flow classification groups.

6.2.8 Spatial autocorrelation analysis

Group predictions for each sample were extracted from each of the final classification models. Residuals of the classification models for each gauging station were determined by counting the relative distance (i.e. number of groups) between the observed and predicted gauge groups for each of the regionalisation and flow
classifications (calculated in a manner consistent with the misclassification cost). The residuals were analysed using local indicators of spatial association (LISA) analysis which measures the degree of local spatial autocorrelation at each sampling point by using a localised version of Moran’s $I$ statistic (Anselin, 1995). LISA allows for the identification and visualisation of high-high (relative to surrounding residuals) and low-low residual locations (i.e. spatial clusters which show positive local spatial autocorrelation) and high-low and low-high locations (i.e. spatial outliers which show negative local spatial autocorrelation). While CART and RF make no assumptions about spatial autocorrelation, an assessment of the predictive errors of the resultant models could be useful in identifying regions where the models perform poorly or have not been able to account for spatial variability in the model. Inverse distance weighting with a threshold distance of 13,965 m (the mean distance between all pairs of gauges) and a minimum of three nearest neighbours was used to define neighbours for residual analysis of the regionalisation classification models, as they only relied on gauge information. The flow classification models all used catchment contiguity where catchments that shared a boundary were considered neighbours – in the case of “island” catchments where they were no boundary neighbours, the three closest catchments were considered.

6.3 Results

6.3.1 Flow classification

As a result of there being suitable gauge data for only eight of the 11 regionalisation classes identified in Chapter 5 (Section 5.3.6), gauges were allocated into eight non-hierarchical groups using the ALOC algorithm. ANOSIM suggested that all groups were well separated, with global $R$ values of 0.555 ($P = 0.001$), indicating that
group membership was highly unlikely to be a result of chance alone. This was supported by the fact that none of the pairwise ANOSIM permutations resulted in $R$ statistics greater than or equal to the global $R$ value. Hierarchical group-average clustering permitted the arrangement of the non-hierarchical groups into a suitable hierarchy, where groups that were related to one another were closer to each other in multivariate space than those that were not (Figure 6.2). PERMANOVA supported the results of ANOSIM and suggested that there were significant differences among all groups (pseudo-$F_{7,193} = 22.05$, $P = 0.001$), while post-hoc pairwise comparisons of groups indicated that the differences were significant ($P \leq 0.05$) between all hierarchical groups with the exception of pairs I:II and I:III.
Figure 6.2: Spatial arrangement of the flow classification groups. Colours represent each of the groups as defined by group-average hierarchical clustering of the ALOC-defined non-hierarchical classes. Similar colours and group letters indicate a closer relationship than those further apart.
6.3.2 Variable contributions to the classifications

6.3.2.1 Relationship between flow indices and hydrological regionalisation

The within-group variation of flow indices within the hydrological regionalisation was quite high, with average squared distances ranging from 8.6 to 45.1 for the eight regionalisation groups (Table 6.1), suggesting large multivariate dispersion and considerable variability in flow indices within groups. The majority of the flow indices were able to differentiate among groups, with all indices bar four (high flood pulse count [FH3], flood frequency [FH6], mean maximum August flows [MH8] and Colwell constancy [TA1]) showing significant differences in distributions across the eight groups according to the Kruskal-Wallis test (Table 6.1). This indicated that these were the only four variables whose distributions did not differ among the regionalisation groups.

Variable contribution to each of the regionalisation groups differed markedly, according to SIMPER analysis, with skewness in daily flows [MA5] contributing 0 % to six of the eight groups, while contributing 2 and 12 % to the remaining two regionalisation groups, for example (Table 6.1). Similarly, variability in Julian date of annual minimum flows [TL2] showed large variation among groups, contributing 8 % of the variation to group F, while contributing ≤ 4 % to the remaining seven groups. The minimum number of flow indices cumulatively contributing > 50 % of the variation to each group differed among groups, with as few as 21 (Group E) to as many as 27 (Group C) indices required (Table 6.1).
Table 6.1: Flow index contribution to each of the regionalisation groups as calculated using SIMPER. Numbers represent the percentage contribution of each of the flow indices to the regionalisation groups using a standardised Euclidean distance similarity matrix. Bold numbers denote the indices with the greatest and least contribution to each group, while ‘-’ denotes indices that had zero percentage contribution to the regionalisation groups. KW = Kruskal-Wallis statistic, with higher values indicating a better ability of that flow index to discriminate among groups. # indicates non-significant ($P > 0.05$) differences in variable distributions across regionalisation groups according to KW test statistic.

<table>
<thead>
<tr>
<th>Flow Index</th>
<th>Average Squared Distance</th>
<th>Regionalisation Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>B</td>
</tr>
<tr>
<td>MA5</td>
<td>49.6</td>
<td>-</td>
</tr>
<tr>
<td>DL18</td>
<td>47.2</td>
<td>0.2</td>
</tr>
<tr>
<td>ML17</td>
<td>44.4</td>
<td>4.5</td>
</tr>
<tr>
<td>TL1</td>
<td>44.2</td>
<td>2.8</td>
</tr>
<tr>
<td>MA11</td>
<td>43.9</td>
<td>1.9</td>
</tr>
<tr>
<td>ML18</td>
<td>43.7</td>
<td>1.7</td>
</tr>
<tr>
<td>MA3</td>
<td>42.7</td>
<td>1.0</td>
</tr>
<tr>
<td>DL13</td>
<td>40.4</td>
<td>3.8</td>
</tr>
<tr>
<td>MH16</td>
<td>40.2</td>
<td>4.1</td>
</tr>
<tr>
<td>RA9</td>
<td>39.1</td>
<td>1.3</td>
</tr>
<tr>
<td>ML21</td>
<td>38.7</td>
<td>2.2</td>
</tr>
<tr>
<td>ML4</td>
<td>37.1</td>
<td>0.5</td>
</tr>
<tr>
<td>FH2</td>
<td>35.6</td>
<td>1.7</td>
</tr>
<tr>
<td>DH13</td>
<td>35.5</td>
<td>1.9</td>
</tr>
<tr>
<td>FL3</td>
<td>33.2</td>
<td>7.6</td>
</tr>
<tr>
<td>TL2</td>
<td>33.0</td>
<td>0.6</td>
</tr>
<tr>
<td>MH14</td>
<td>30.4</td>
<td>3.0</td>
</tr>
<tr>
<td>FH7</td>
<td>30.3</td>
<td>5.7</td>
</tr>
<tr>
<td>MA41</td>
<td>27.6</td>
<td>3.3</td>
</tr>
<tr>
<td>RA8</td>
<td>24.7</td>
<td>4.4</td>
</tr>
<tr>
<td>RA5</td>
<td>24.7</td>
<td>4.4</td>
</tr>
<tr>
<td>DH16</td>
<td>21.6</td>
<td>1.8</td>
</tr>
<tr>
<td>DH20</td>
<td>19.7</td>
<td>6.8</td>
</tr>
<tr>
<td>DH15</td>
<td>18.9</td>
<td>6.3</td>
</tr>
<tr>
<td>FL2</td>
<td>17.2</td>
<td>1.7</td>
</tr>
<tr>
<td>DL17</td>
<td>17.1</td>
<td>2.4</td>
</tr>
<tr>
<td>MH10</td>
<td>16.0</td>
<td>1.8</td>
</tr>
<tr>
<td>FL1</td>
<td>14.7</td>
<td>1.8</td>
</tr>
<tr>
<td>DL16</td>
<td>14.3</td>
<td>1.0</td>
</tr>
<tr>
<td>TAI1#</td>
<td>13.0</td>
<td>7.5</td>
</tr>
<tr>
<td>FH30</td>
<td>11.8</td>
<td>2.5</td>
</tr>
<tr>
<td>MH8#</td>
<td>11.8</td>
<td>3.8</td>
</tr>
<tr>
<td>FH6#</td>
<td>7.9</td>
<td>6.0</td>
</tr>
</tbody>
</table>
6.3.2.2 Relationship between environmental variables and flow classification

Similar to the relationship between flow indices and the hydrological regionalisation, the within-group variation in environmental variables was high, with average squared distances ranging from 9.6 to 45.3 for the eight flow classification groups, again suggesting large multivariate dispersion within groups. The ability of the environmental variables to differentiate among groups was mixed, with 12 of the 25 variables showing non-significant differences in distributions across the eight flow classification groups according to the Kruskal-Wallis test (Table 6.2), indicating that there were only 13 variables whose distributions differed significantly among the flow classification groups.

Variable contribution to each of the flow classification groups differed markedly according to SIMPER analysis, with landscape development intensity, for example, contributing ≤ 10 % to seven of the eight groups, while contributing 40 % to the remaining group (Table 6.2). Interestingly, the aridity index (rainfall/potential ET) displayed similar levels of contribution of variance to all the groups (1 – 5 %) yet had the largest KW value of all the variables suggesting that the aridity index was the variable that could most easily discriminate between groups (Table 6.2). Similar patterns were observed when comparing the flow indices to the regionalisation groups, in that there was significant variation in the number of environmental variables cumulatively contributing > 50 % of the variation to each group, with a minimum of 15 (Groups V and VI) and a maximum of 22 (Groups I and III) (Table 6.2).
Table 6.2: Environmental variable contribution to each of the flow classification groups as calculated using SIMPER. Numbers represent the percentage contribution of each of the variables to the classification groups using a standardised Euclidean distance matrix. Bold numbers denote the variables with the greatest and least contribution to each group, while ‘-‘ denotes indices that had zero percentage contribution to the regionalisation groups. KW = Kruskal-Wallis statistic, with higher values indicating a better ability of that variable to discriminate among groups. # indicates non-significant ($P > 0.05$) differences in variable distributions across regionalisation groups according to KW test statistic.

<table>
<thead>
<tr>
<th>Environmental variable</th>
<th>Average Squared Distance</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
<th>VI</th>
<th>VII</th>
<th>VIII</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aridity Index</td>
<td>56.3</td>
<td>3.5</td>
<td>2.4</td>
<td>1.3</td>
<td>1.6</td>
<td>5.1</td>
<td>3.0</td>
<td>2.2</td>
<td>2.0</td>
</tr>
<tr>
<td>Mean Annual Rainfall</td>
<td>52.6</td>
<td>4.6</td>
<td>1.4</td>
<td>1.0</td>
<td>1.3</td>
<td>5.0</td>
<td>3.2</td>
<td>2.3</td>
<td>2.3</td>
</tr>
<tr>
<td>Rainfall Of Wettest Quarter</td>
<td>48.6</td>
<td>6.8</td>
<td>1.0</td>
<td>1.2</td>
<td>1.9</td>
<td>5.1</td>
<td>3.3</td>
<td>2.3</td>
<td>3.0</td>
</tr>
<tr>
<td>Soil Erosivity Index</td>
<td>45.5</td>
<td>7.7</td>
<td>1.2</td>
<td>-</td>
<td>0.3</td>
<td>5.2</td>
<td>2.9</td>
<td>2.4</td>
<td>1.7</td>
</tr>
<tr>
<td>Rainfall Of Driest Quarter</td>
<td>45.5</td>
<td>2.8</td>
<td>1.0</td>
<td>0.3</td>
<td>0.4</td>
<td>3.3</td>
<td>3.7</td>
<td>4.4</td>
<td>5.3</td>
</tr>
<tr>
<td>Groundwater Total Dissolved Solids</td>
<td>42.0</td>
<td>4.5</td>
<td>-</td>
<td>15.1</td>
<td>9.6</td>
<td>5.8</td>
<td>2.1</td>
<td>4.1</td>
<td>3.0</td>
</tr>
<tr>
<td>Mean Annual Evapotranspiration</td>
<td>27.6</td>
<td>2.6</td>
<td>0.4</td>
<td>6.6</td>
<td>2.4</td>
<td>3.9</td>
<td>3.2</td>
<td>5.3</td>
<td>6.8</td>
</tr>
<tr>
<td>Weathering Intensity Index</td>
<td>27.6</td>
<td>6.9</td>
<td>1.3</td>
<td>0.7</td>
<td>1.2</td>
<td>2.7</td>
<td>5.0</td>
<td>5.9</td>
<td>5.6</td>
</tr>
<tr>
<td>Temperature Seasonality</td>
<td>23.7</td>
<td>1.5</td>
<td>2.3</td>
<td>3.3</td>
<td>1.2</td>
<td>3.0</td>
<td>5.8</td>
<td>4.7</td>
<td>4.1</td>
</tr>
<tr>
<td>Saturated Conductivity – A Horizon</td>
<td>19.4</td>
<td>0.4</td>
<td>6.2</td>
<td>2.3</td>
<td>5.1</td>
<td>3.6</td>
<td>4.5</td>
<td>2.7</td>
<td>4.8</td>
</tr>
<tr>
<td>Mean Temperature Of Driest Quarter</td>
<td>15.7</td>
<td>0.1</td>
<td>9.8</td>
<td>3.2</td>
<td>3.3</td>
<td>3.6</td>
<td>3.8</td>
<td>5.8</td>
<td>7.1</td>
</tr>
<tr>
<td>Mean Maximum Temperature</td>
<td>14.3</td>
<td>-</td>
<td>14.8</td>
<td>2.6</td>
<td>3.4</td>
<td>4.8</td>
<td>4.0</td>
<td>3.0</td>
<td>5.0</td>
</tr>
<tr>
<td>Mean Temperature Of Wettest Quarter$^e$</td>
<td>13.9</td>
<td>1.8</td>
<td>1.5</td>
<td>2.1</td>
<td>2.1</td>
<td>5.5</td>
<td>2.7</td>
<td>4.6</td>
<td>5.4</td>
</tr>
<tr>
<td>Elevation$^e$</td>
<td>13.9</td>
<td>1.4</td>
<td>8.3</td>
<td>0.4</td>
<td>2.2</td>
<td>4.4</td>
<td>5.0</td>
<td>3.1</td>
<td>2.9</td>
</tr>
<tr>
<td>Plant AWC – A Horizon$^e$</td>
<td>13.3</td>
<td>0.6</td>
<td>-</td>
<td>22.2</td>
<td>16.0</td>
<td>2.7</td>
<td>3.8</td>
<td>5.1</td>
<td>3.0</td>
</tr>
<tr>
<td>Slope</td>
<td>13.3</td>
<td>3.2</td>
<td>0.2</td>
<td>0.4</td>
<td>2.2</td>
<td>3.3</td>
<td>3.9</td>
<td>6.1</td>
<td>3.8</td>
</tr>
<tr>
<td>Landscape Development Intensity$^e$</td>
<td>12.1</td>
<td>40.6</td>
<td>2.2</td>
<td>10.2</td>
<td>6.6</td>
<td>5.5</td>
<td>2.5</td>
<td>2.2</td>
<td>7.1</td>
</tr>
<tr>
<td>Mean Minimum Temperature$^e$</td>
<td>9.9</td>
<td>0.8</td>
<td>12.4</td>
<td>1.6</td>
<td>2.2</td>
<td>4.2</td>
<td>5.1</td>
<td>4.2</td>
<td>2.4</td>
</tr>
<tr>
<td>Saturated Conductivity – B Horizon$^e$</td>
<td>9.8</td>
<td>-</td>
<td>9.5</td>
<td>1.3</td>
<td>3.8</td>
<td>3.9</td>
<td>4.7</td>
<td>2.6</td>
<td>3.4</td>
</tr>
<tr>
<td>Topographic Wetness Index$^e$</td>
<td>8.2</td>
<td>0.7</td>
<td>0.5</td>
<td>1.1</td>
<td>7.3</td>
<td>3.4</td>
<td>3.2</td>
<td>5.3</td>
<td>6.0</td>
</tr>
<tr>
<td>Layer Thickness – B Horizon$^e$</td>
<td>6.7</td>
<td>0.1</td>
<td>7.1</td>
<td>4.8</td>
<td>6.7</td>
<td>3.3</td>
<td>5.6</td>
<td>3.0</td>
<td>2.1</td>
</tr>
</tbody>
</table>

205
<table>
<thead>
<tr>
<th>Environmental variable</th>
<th>Average Squared Distance</th>
<th>Flow Classification Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>KW</td>
<td>I</td>
</tr>
<tr>
<td>Plant AWC – B Horizon⁶</td>
<td>6.4 0.2 9.0 1.0 0.9 3.1 6.0 2.2 2.8</td>
<td></td>
</tr>
<tr>
<td>Layer Thickness – A Horizon⁶</td>
<td>6.0 - 2.9 14.8 12.2 2.3 4.9 4.7 2.8</td>
<td></td>
</tr>
<tr>
<td>Rainfall Seasonality⁷</td>
<td>4.3 4.6 4.2 2.2 3.6 4.1 4.2 6.5 4.6</td>
<td></td>
</tr>
<tr>
<td>Groundwater Static Water Level⁷</td>
<td>3.7 4.5 0.4 0.2 2.5 3.5 3.9 5.6 2.8</td>
<td></td>
</tr>
</tbody>
</table>
6.3.3 Modelling approaches

6.3.3.1 CAP

*Distinguishing among flow classification groups using environmental variables*

The ability of CAP to correctly classify gauges into the correct flow classification group based on the environmental variables was marginal, with 96 of 201 samples correctly classified, but the model was statistically significant (48 %, $m = 42$, $P = 0.001$). This result is substantially better than the level of correct classification expected by chance alone which, assuming eight groups and equal sample sizes, results in 12.5 % of cases being correctly classified. Groups I and III in the flow classification had the lowest success rate with zero samples correctly allocated, followed by Group II which had a 25 % correct allocation rate. The highest correct classification rate was for Group VIII, in which 78 % of samples were correctly classified. Due to the low number of gauges in Groups I, II and III (2, 4 and 3, respectively) and the results of the PERMANOVA suggesting no significant differences between pairs I:II, and I:III, I removed those gauges from any further analyses on the flow classification groups. The capacity of CAP to correctly classify gauges increased with the removal of the nine gauges belonging to classes I, II and III, with 105 of the remaining 192 gauges classified correctly (55 %, $m = 42$, $P = 0.001$). Again, Class VIII had the highest correctly classified rate, still with 78 % of samples correctly classified, while all other classes were correctly allocated ~ 50 % of the time. Allowing for a one class either side allocation, (e.g. True Class VII, allocated to class VI, VII or VIII), the success rate of the CAP increased drastically with classes V, VI, VII and VIII having percentage success rates of >80 %,
while class IV increased from 50 % to 67 %, suggesting that there was some overlap in
the environmental characteristics of adjacent classes.

6.3.3.2 CART

Identifying regionalisation classification groups using flow indices

Parameter tuning for the CART model of regionalisation group classification
(Figure 6.3) resulted in a $cp$ value of 0.003 using the flow indices dataset. The final
tuned $k$ value ($\pm$ s.d.) using the best tuned $cp$ value for the regionalisation classification
model was 0.42 ($\pm$ 0.06), while classification accuracy was 51 ($\pm$ 5 %) (Table 6.3). These
results are similar to those observed during the previous validation conducted in Chapter
5 (Section 5.3.6) and signify that a large amount of variation in regional hydrology was
accounted for under that regionalisation framework and classification. Bootstrap-
estimated misclassification rates were high (> 50 %) for all of the regionalisation classes
on the flow indices dataset (Table 6.4). However, as before, allowing for a classification
tolerance of one class either side of the true class, classification accuracy increased
markedly with, for example, Group I classification error decreasing from 60 % to 31 %.
Figure 6.3: Regionalisation group classification tree built using 33 flow indices. Final classification accuracy was 51 %, while analysis indicated that 23 of the 33 predictor variables were actually used in producing and tuning the classification tree. The top line in each node represents the majority class of the node; the second line is the probability of the majority node being correct; the bottom line represents the splitting variable. Split values are placed above the subsequent node. The pie charts show the distribution of all classes present in each node. For details of the split variables please refer to Appendix II (Table 10.1).
Table 6.3: Bootstrap assessment of the performance of the classification models – kappa and percentage accuracy of the CART (left) and RF (right) models used to validate the hydrological regionalisation following from Chapter 5, and the new flow classification conducted in this paper. The Min., Mean and Max. columns represent model performance across the \((n = 1000)\) 0.632 bootstrap datasets for the final chosen \(cp\) or \(mtry\) value. The final column represents the performance of the final tuned model on the original (non-bootstrapped) datasets.

<table>
<thead>
<tr>
<th>Classification Models</th>
<th>CART</th>
<th>RF</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Kappa</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regionalisation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(flow indices dataset)</td>
<td>Min.</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>Max.</td>
<td>0.42</td>
</tr>
<tr>
<td></td>
<td>Final</td>
<td></td>
</tr>
<tr>
<td>Flow classification</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(environmental variables dataset)</td>
<td>Min.</td>
<td>17%</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>27%</td>
</tr>
<tr>
<td></td>
<td>Max.</td>
<td>61%</td>
</tr>
<tr>
<td></td>
<td>Final</td>
<td></td>
</tr>
</tbody>
</table>

| **Accuracy**          |      |    |
| Regionalisation       |      |    |
| (flow indices dataset)| Min. | 17%  | 19%  | 51%  | 51%  | 26%  | 43%  | 62%  |
|                       | Mean | 27%  | 44%  | 59%  | 61%  | 38%  | 52%  | 67%  |
|                       | Max. | 61%  | 61%  | 61%  | 59%  | 48%  | 67%  | 69%  |
|                       | Final|      |      |      |      |      |      |      |
Table 6.4: Bootstrap \((n = 1000)\) confusion matrices for CART for the regionalisation and flow classifications. Entries are percentages of table totals across all rows and columns. The diagonal (italics) represent the percentage of cases that were correctly classified by CART while values off the diagonal represent the number of misclassified cases for each combination of predicted and actual groups. Numbers in brackets are the actual group counts from the original (non-bootstrapped) dataset. Flow classification groups A, B and C were excluded from the CART analysis (see Section 6.2.6.2).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>1.50</td>
<td>0.10</td>
<td>0.70</td>
<td>1.80</td>
<td>0.30</td>
<td>0.60</td>
<td>0.40</td>
<td>0.60</td>
</tr>
<tr>
<td>C</td>
<td>0.20</td>
<td>0.40</td>
<td>0.80</td>
<td>0.70</td>
<td>0.30</td>
<td>1.30</td>
<td>0.40</td>
<td>0.40</td>
</tr>
<tr>
<td>D</td>
<td>0.70</td>
<td>1.30</td>
<td>3.60</td>
<td>1.90</td>
<td>0.50</td>
<td>2.30</td>
<td>0.70</td>
<td>0.90</td>
</tr>
<tr>
<td>E</td>
<td>2.00</td>
<td>0.90</td>
<td>1.90</td>
<td>8.00</td>
<td>1.10</td>
<td>3.50</td>
<td>3.30</td>
<td>1.90</td>
</tr>
<tr>
<td>F</td>
<td>0.30</td>
<td>0.30</td>
<td>0.40</td>
<td>1.10</td>
<td>0.50</td>
<td>1.30</td>
<td>0.90</td>
<td>0.80</td>
</tr>
<tr>
<td>G</td>
<td>0.50</td>
<td>1.40</td>
<td>2.20</td>
<td>3.10</td>
<td>1.40</td>
<td>5.90</td>
<td>3.30</td>
<td>1.90</td>
</tr>
<tr>
<td>I</td>
<td>0.50</td>
<td>0.30</td>
<td>0.70</td>
<td>3.10</td>
<td>0.80</td>
<td>3.00</td>
<td>7.40</td>
<td>2.20</td>
</tr>
<tr>
<td>J</td>
<td>0.40</td>
<td>0.30</td>
<td>0.60</td>
<td>1.50</td>
<td>1.10</td>
<td>1.60</td>
<td>2.00</td>
<td>4.20</td>
</tr>
<tr>
<td>Classification Error</td>
<td>75%</td>
<td>92%</td>
<td>67%</td>
<td>62%</td>
<td>92%</td>
<td>70%</td>
<td>60%</td>
<td>67%</td>
</tr>
<tr>
<td>Classification Error (Tolerance = 1)</td>
<td>72%</td>
<td>64%</td>
<td>42%</td>
<td>48%</td>
<td>50%</td>
<td>48%</td>
<td>31%</td>
<td>50%</td>
</tr>
</tbody>
</table>

Flow classification (environmental variables dataset)

<table>
<thead>
<tr>
<th>Predicted Group</th>
<th>I (2)</th>
<th>II (4)</th>
<th>III (3)</th>
<th>IV (6)</th>
<th>V (53)</th>
<th>VI (67)</th>
<th>VII (39)</th>
<th>VIII (27)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-0.40</td>
<td>0.80</td>
<td>0.90</td>
<td>0.30</td>
<td>0.30</td>
</tr>
<tr>
<td>II</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>III</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>IV</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.10</td>
<td>12.50</td>
<td>7.90</td>
<td>3.90</td>
<td>1.50</td>
</tr>
<tr>
<td>V</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>8.90</td>
<td>17.60</td>
<td>5.50</td>
<td>3.00</td>
</tr>
<tr>
<td>VI</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.40</td>
<td>4.40</td>
<td>5.90</td>
<td>7.20</td>
<td>3.00</td>
</tr>
<tr>
<td>VII</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.30</td>
<td>1.10</td>
<td>2.70</td>
<td>3.10</td>
<td>6.30</td>
</tr>
<tr>
<td>VIII</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Classification Error</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>88%</td>
<td>55%</td>
<td>50%</td>
<td>64%</td>
<td>55%</td>
</tr>
<tr>
<td>Classification Error (Tolerance = 1)</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>53%</td>
<td>20%</td>
<td>10%</td>
<td>21%</td>
<td>34%</td>
</tr>
</tbody>
</table>
Analyses of the final classification tree indicated that 23 of the 33 predictor variables were actually used in generating the classification tree. Of those 23, analysis of variable importance indicated that the four most important predictors for the regionalisation CART model built on the flow indices dataset were: variability in high-flood pulse count [FH2]; variability in high-flow pulse duration [DH16]; mean annual runoff [MA41]; and high-flood pulse count [FH3]. These variables suggest that differences in high-flow hydrology were more important in discriminating between regionalisation groups than other variables (e.g. those concerning low flows or baseflows). Conversely, variable importance based on the KW statistic (Table 6.1), suggested that the variables that could best discriminate groups were skewness in daily flows [MA5], mean number of zero flow days [DL18], baseflow index [ML17] and Julian date of annual minimum flows [TL1] – all statistics related to low flow hydrology. Intriguingly, FH3 was deemed to not have significant between-group variation according to the KW statistic (Table 6.1), however CART modelling revealed it to be a very important variable in discriminating between groups. This result is interesting and suggests that, depending on the classification or analysis method used, there can be very significant differences in the ability of individual variables to discriminate among groups, potentially related to the type of relationships that exist between variables, or the assumptions of the test in question.

Identifying flow classification groups using environmental variables

Parameter tuning for the CART model of flow classification groups (Figure 6.4) resulted in a $cp$ value of 0.012 using the environmental variables dataset.
Figure 6.4: Flow group classification tree built using 50 environmental variables related to topography, climate, geology and land cover. Final classification accuracy was 61\%, while analysis indicated that 21 of the 50 predictor variables were actually used in producing and tuning the classification tree. The top line in each node represents the majority class of the node; the second line is the probability of the majority node being correct; the bottom line represents the splitting variable. Split values are placed above the subsequent node. The pie charts show the distribution of all classes present in each node. For details of the split variables please refer to Appendix I (Table 9.1).
The final tuned $\kappa$ value (± s.d.) using the best tuned $cp$ value for the flow classification model was 0.48 (± 0.08), while classification accuracy was 61 (± 6 %) (Table 6.3). These higher classification accuracies suggest that the environmental variables may be more useful in explaining variability in flow classification groups than hydrologic indices were when using a regionalisation approach. Bootstrap-estimated misclassification rates were high (>50 %) for the five flow classes included in the flow classification model (Table 6.4). Allowing for a classification tolerance of one class either side of the true class however, classification accuracy increases were greater than those observed for the regionalisation model with, for example, Class F showing an increase of 80 % accuracy (classification error decreased from 50 % to 10 %).

Analyses of the final classification tree indicated that 21 of the 50 predictor variables (the mean and standard deviation of each of the environmental variables) were actually used in generating the classification tree. Analysis of variable importance indicated that the four most important predictors for the flow classification CART model built on the environmental variables dataset were: mean annual rainfall; mean aridity index; mean precipitation of the wettest quarter; and mean soil erosivity index of each catchment. These variables suggest that climatic conditions, particularly rainfall, are largely responsible for differences in the flow classification groups. Unlike the regionalisation CART model, the four variables identified as the most important for the flow classification model were also the variables with the highest KW scores (Table 6.2).
6.3.3.3 Random forests

*Identifying regionalisation classification groups using flow indices*

Parameter tuning for \( mtry \) resulted in the final regionalisation random forest model using two of the streamflow indices per split for tree growing. One thousand trees were sufficient to allow the models to reach a stable solution with regard to out-of-bag (OOB) error, and there was little improvement in model accuracies after approximately 500 trees (Appendix II, Figure 10.1). The final tuned model had an accuracy of 62 % (\( \kappa = 0.53 \)) when using the flow indices to predict the regionalisation groups (Table 6.3). Pairwise \( t \)-tests indicated that both \( \kappa \) (\( t_{(999)} = 43.2, P < 0.001 \)) and percent agreement (\( t_{(999)} = 46.3, P < 0.001 \)) were significantly higher for the RF model when compared against bootstrap values from the CART model. However, Wilcoxon signed-rank tests suggested that there were no significant differences between pairwise group error rates (\( V = 6, P = 0.18 \)).

Kappa and model accuracy varied across a very small range of values for all values of \( mtry \) on the original (non-bootstrapped) dataset with \( \kappa \) ranging from a maximum of 0.54 for \( mtry = 2 \) to a minimum of 0.52 for \( mtry = 30 \). Percent accuracy followed the same pattern and varied between 60 and 62 % for the same values of \( mtry \). Mean bootstrapped \( \kappa \) values for \( mtry = 2 \) was 0.31 (± 0.01), while the mean accuracy value was 43 (± 1 %) (Table 6.3). Bootstrap estimated misclassification rates were high (> 50 %) with the exception of groups D, E, G and I (Table 6.5). As before, allowing a tolerance of one class, misclassification rates dropped substantially with all regionalisation classes except B and C having error rates ±50 % (Table 6.5).
The four most important predictors for the RF model used to classify the regionalisation groups were: variability in high-flood pulse count [FH2], coefficient of variation of the baseflow index [ML18], spread in daily flows [MAI1] and Julian date of annual minimum [TL1]. These variables suggest that different aspects of the flow regime were important in distinguishing between regionalisation classes as indices related to both high and low flows were considered important for model accuracy. These variables were also considered relatively important according to the KW statistic with TL1, MAI1 and ML18 being amongst the six variables with the highest KW values (Table 6.1).

Additional analyses, not presented here, demonstrated that when the percent area of each class in the regionalisation was included as additional predictor variables (to estimate inter-catchment variability), accuracy of the tuned CART models increased to 77 %, while the tuned RF model accuracy increased to 83 %. This was possibly due to the fair agreement between the gauge class and the percent area of the majority class of each catchment, suggesting that the lower accuracies observed for the hydrological regionalisation are likely to be, at least in part, due to inter-catchment variability.
Table 6.5: Bootstrap ($n = 1000$) confusion matrices for RF for the regionalisation and flow classifications. Entries are percentages of table totals across all rows and columns. The diagonal (italics) represent the percentage of cases that were correctly classified by RF while values off the diagonal represent the number of misclassified cases for each combination of predicted and actual groups. Numbers in brackets are the actual group counts from the original (non-bootstrapped) dataset. Flow classification groups A, B and C were excluded from the RF analysis (see Section 6.2.6.3).

### Regionalisation classification (hydrological indices dataset)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>0.80</td>
<td>0.00</td>
<td>0.30</td>
<td>0.70</td>
<td>0.10</td>
<td>0.00</td>
<td>0.00</td>
<td>0.30</td>
</tr>
<tr>
<td>C</td>
<td>0.00</td>
<td>0.20</td>
<td>0.50</td>
<td>0.20</td>
<td>0.10</td>
<td>0.10</td>
<td>0.00</td>
<td>0.10</td>
</tr>
<tr>
<td>D</td>
<td>0.90</td>
<td>2.00</td>
<td>5.40</td>
<td>0.60</td>
<td>0.40</td>
<td>1.70</td>
<td>0.20</td>
<td>0.60</td>
</tr>
<tr>
<td>E</td>
<td>3.10</td>
<td>0.80</td>
<td>2.10</td>
<td>11.80</td>
<td>1.40</td>
<td>2.80</td>
<td>2.40</td>
<td>2.40</td>
</tr>
<tr>
<td>F</td>
<td>0.00</td>
<td>0.10</td>
<td>0.00</td>
<td>0.00</td>
<td>0.50</td>
<td>0.20</td>
<td>0.20</td>
<td>0.10</td>
</tr>
<tr>
<td>G</td>
<td>0.40</td>
<td>1.90</td>
<td>1.80</td>
<td>3.10</td>
<td>2.00</td>
<td>10.10</td>
<td>3.80</td>
<td>2.00</td>
</tr>
<tr>
<td>I</td>
<td>0.60</td>
<td>0.10</td>
<td>0.80</td>
<td>4.40</td>
<td>0.90</td>
<td>3.80</td>
<td>10.50</td>
<td>3.70</td>
</tr>
<tr>
<td>J</td>
<td>0.10</td>
<td>0.00</td>
<td>0.00</td>
<td>0.30</td>
<td>0.70</td>
<td>0.30</td>
<td>1.30</td>
<td>3.00</td>
</tr>
</tbody>
</table>

Classification Error
- 86% 96% 50% 44% 92% 48% 43% 75%

Classification Error (Tolerance = 1)
- 86% 57% 27% 41% 35% 28% 15% 45%

### Flow classification (environmental variables dataset)

<table>
<thead>
<tr>
<th>Predicted Group</th>
<th>I (2)</th>
<th>II (4)</th>
<th>III (3)</th>
<th>IV (6)</th>
<th>V (53)</th>
<th>VI (67)</th>
<th>VII (39)</th>
<th>VIII (27)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>II</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>III</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>IV</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.10</td>
<td>0.80</td>
<td>0.80</td>
<td>0.30</td>
<td>0.00</td>
</tr>
<tr>
<td>V</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.10</td>
<td>13.10</td>
<td>7.60</td>
<td>4.20</td>
<td>0.80</td>
</tr>
<tr>
<td>VI</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.60</td>
<td>10.30</td>
<td>21.00</td>
<td>4.40</td>
<td>3.00</td>
</tr>
<tr>
<td>VII</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.30</td>
<td>3.10</td>
<td>4.40</td>
<td>9.50</td>
<td>2.70</td>
</tr>
<tr>
<td>VIII</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.00</td>
<td>0.30</td>
<td>1.10</td>
<td>1.70</td>
<td>7.60</td>
</tr>
</tbody>
</table>

Classification Error
- NA NA NA 97% 53% 40% 53% 46%

Classification Error (Tolerance = 1)
- NA NA NA 61% 12% 5% 22% 27%
Parameter tuning for \textit{mtry} resulted in the final RF flow classification model using four of the environmental variables per split for tree growing. As for the RF regionalisation model, 1000 trees was shown to be large enough to allow the flow classification models to reach a stable solution with regard to out-of-bag (OOB) error, and there was little improvement in model accuracies after approximately 500 trees (Appendix II, Figure 10.1). The final tuned model had an accuracy of 69 \% (\(\kappa = 0.58\)) when using the environmental variables to predict the flow classification groups (Table 6.3). Again, pairwise \(t\)-tests indicated that both \(\kappa (t_{999}) = 29.2, P < 0.001\) and percent agreement (\(t_{999} = 31.0, P < 0.001\)) were significantly higher for the RF model when compared against bootstrap values from the CART model. Wilcoxon signed-rank tests again suggested that there were no significant differences between pairwise group error rates \((V = 2, P = 0.14)\).

Kappa and model accuracy showed very small variability for all values of \textit{mtry} on the original (non-bootstrapped) dataset with \(\kappa\) ranging from a maximum of 0.58 for \(mtry < 3\) to 0.57 for \(mtry > 20\). Percent accuracy was approximately 69 \% for all values of \textit{mtry}. Mean bootstrapped \(\kappa\) values for \textit{mtry} = 4 was 0.33 (± 0.001), while the mean accuracy value was 52 \% (± 0.2 \%) (Table 6.3). Bootstrap estimated misclassification rates were relatively high (> 50 \%) with the exception of groups VI and VIII (Table 6.5). As before, allowing a tolerance of one class, misclassification rates dropped substantially, with all flow classes except IV having classification accuracies ≥ 73 \% (Table 6.5). The four most important predictors for the RF model built on the original dataset were: mean aridity index; mean precipitation of the wettest quarter; mean annual rainfall; and
mean value of the landscape development intensity index (LDI). These variables suggest that characteristics related to climate and land-cover conditions were important for defining hydrologic patterns across landscapes. Unlike the flow classification CART model, only three of the four variables identified as the most important for the RF flow classification model were also the variables with the highest KW scores. The LDI had a relatively low KW score and was shown to not exhibit significant differences between flow classification groups (Table 6.2).

6.3.4 Comparisons between the regionalisation and flow classifications

The ability of CART and RF to predict classification groups varied with classification type (regionalisation or flow) (Table 6.3). Welch’s t-tests indicated that CART and RF were significantly better at predicting membership of samples for the flow classification than they were at predicting regionalisation group membership in terms of both κ (CART: $t_{(1929.7)} = 17.6, P < 0.001$; RF: $t_{(1989.6)} = 8.5, P < 0.001$) and percent accuracy (CART: $t_{(1993.8)} = 50.7, P < 0.001$; RF: $t_{(1974.4)} = 36.7, P < 0.001$).

Analysis of the distribution of the regionalisation groups (B, C, D etc.) across the flow classification classes (I, II, III etc.) showed very low agreement if the classes were treated as being ordinal based on the number of samples assigned to each class (i.e. regionalisation group B was equivalent to flow class I, E = V, I = VIII, etc.). For example, regionalisation classes B and C (there were no samples from classes A, H or K; Chapter 5, Section 5.3.6) had zero samples allocated to the first two flow classes (I, II); while regionalisation groups E and G had the highest agreement with the appropriate flow class (V and VI) (Table 6.6). Overall percentage agreement (assuming ‘ordinal’ classes) was only 27%.
The relationship among the classes in the respective classifications suggests that a smaller number of flow classes were able to explain the same variation that the regionalisation classes were designed to explain. For example, Table 6.6 reveals that five flow classes can explain 96% of the variation in the regionalisation classes with 192 cases assigned to only five flow classes (IV – VIII). This may, in part, reflect biases in the location of the available gauges.
Table 6.6: Distribution of regionalisation groups across the newly-created flow classes. The low agreement and specificity of the regionalisation groups to the flow classes, suggests that there is considerable hydrologic variability in the regionalisation groups. The diagonal (italics) represent the number of cases that were in the same ordinal group (e.g. regionalisation class B, flow class I). Agreement = the percentage agreement along the diagonal for each of the regionalisation groups and flow classes. Precision = number of correct samples / number of total samples in each flow class, assuming ordinal groups.

<table>
<thead>
<tr>
<th>Flow Class</th>
<th>B</th>
<th>C</th>
<th>F</th>
<th>J</th>
<th>E</th>
<th>G</th>
<th>D</th>
<th>I</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0%</td>
</tr>
<tr>
<td>II</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>III</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>33%</td>
</tr>
<tr>
<td>IV</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>50%</td>
</tr>
<tr>
<td>V</td>
<td>7</td>
<td>5</td>
<td>1</td>
<td>7</td>
<td>17</td>
<td>6</td>
<td>9</td>
<td>1</td>
<td>32%</td>
</tr>
<tr>
<td>VI</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>14</td>
<td>16</td>
<td>8</td>
<td>11</td>
<td>24%</td>
</tr>
<tr>
<td>VII</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>13</td>
<td>5</td>
<td>10</td>
<td>13%</td>
</tr>
<tr>
<td>VIII</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>13</td>
<td>48%</td>
</tr>
</tbody>
</table>

| Agreement  | 0% | 0% | 8% | 12%| 40%| 41%| 23%| 35%|           |
6.3.5 Spatial autocorrelation analysis

LISA analysis of residuals demonstrated very little spatial autocorrelation for all of the regionalisation and flow classification models. RF consistently outperformed the CART classification models in terms of having a smaller number of spatial clusters (i.e. illustrative of positive spatial autocorrelation) and/or outliers (i.e. illustrative of negative spatial autocorrelation) on both datasets (Appendix II, Figure 10.2). The low levels of spatial autocorrelation exhibited by all the models suggests that the majority of the spatial variation in the relationships is being accounted for by both methods and is captured in the regionalisation created in Chapter 5.

6.4 Discussion

This paper investigated the ability of CART and RF to improve the validation of a hydrologic regionalisation created for Chapter 5 which was developed so as to not require use of a spatial unit defined *a priori* (e.g. catchment) nor rely on hydrologic gauge data to create the classes. Here, two different predictive models were created to analyse the ability of 33 streamflow indices to differentiate among the different groups in the regionalisation, which were then compared with another three predictive models which were created based on the more common method of identifying flow classification groups based on environmental predictor variables. All models were tuned for parameters controlling the complexity and sensitivity of the final model and, in the case of CART and RF, accuracy was assessed with 0.632 bootstrap resampling. Also, an analysis of the amount of local spatial autocorrelation was undertaken to examine visually where the methods that had been utilised may not have been able to account for spatial variability.
6.4.1 Ability of streamflow indices to predict hydrologic regionalisation groups

Using a constrained discriminant analysis for validation of the hydrological regionalisation, Chapter 5 reported a classification accuracy of 48 %, with 92 of 201 stream gauges correctly classified. The CART and RF models developed here achieved accuracies of 51 and 62 %, respectively for the same gauges. However, resampled performance estimates using 0.632 bootstrapping implied upper and lower accuracies of 62 and 17 %, respectively, suggesting that, even if more streamflow gauges were available, classification accuracies may not be any greater than those observed here, assuming a similar distribution of gauges among classes. As previously observed (Chapter 5), common classes were correctly classified more frequently than uncommon classes and previous recommendations of interpreting the results of the less common classes with caution are still appropriate.

It has been suggested that catchments are the appropriate scale to relate landscape and environmental characteristics to flow variability, and that the spatial scale of regionalisations will influence their ability to accurately predict a response (McManamay et al., 2012). I have shown that, while an analysis of the ability of a regionalisation to predict hydrological response is dependent on the use of catchments as a spatial scale, the regionalisation itself can be created independently of those catchments and can still accurately predict hydrology. Furthermore, Poff et al. (2006) suggested that, within hydro-climatically similar regions, incorporating a finer spatial scale analysis of flow regime type could improve hydrologic stratification based solely on regionally-relevant components of flow variability. Geomorphic stratification could also be applied
to identify streams with similar hydro-geomorphic properties which could impact streamflow regimes and variability (Poff et al., 2006).

It has also been hypothesised that catchments become less distinct from one another and streamflow regimes become less spatially diverse as catchment scale increases (Gustard, 1992). As a result, the applicability of hydrologic regionalisations to hydrological prediction has been questioned when the relative importance of the variables that create them is not allowed to change with scale (Buttle, 2006; McManamay et al., 2012). McManamay et al. (2012) proposed that a major issue with hydrological regionalisations is that catchments are assigned to a single, dominant regionalisation class. Catchments, however, are likely composed of a number of different classes and a combination of localised and catchment factors need to be taken into account when attempting to predict flow variability (Snelder et al., 2005; Carlisle et al., 2010; McManamay et al., 2012).

The ‘error’ in the classification accuracies observed here could, in fact, possibly be explained by the regionalisation preserving a large amount of the intra-catchment variability that is known to exist in general (Poff, 1996; Poff et al., 2006; Kennard et al., 2010b), and across the study site (Chapters 3, 4 and 5). Although many of the regionalisation classes were predicted with moderate accuracy using a range of flow indices, particularly when a classification tolerance of one class was permitted, the ability of the streamflow indices to predict regionalisation groups was generally mixed. While the regionalisation created for Chapter 5 was specifically designed to account for intra-catchment variability, in this analysis it was only possible to assign a single class to the location of each stream gauge. This approach effectively limited each catchment to a
single class for the purposes of the validation exercise. It has been observed previously that hydrologic regions are typically unable to explain much of the variation in individual hydrologic parameters (McManamay et al., 2012) and this may have also contributed to the error in the classifications presented here. So, our approach allows some inferences to be made regarding the hydrology of those classes, particularly of the more common classes, but an explicit method of accounting for the variability within each class would help increase prediction accuracy which would lead to an even greater understanding of the hydrology of each class.

Further work is needed to investigate different methods of incorporating the catchment variability that was preserved in the regionalisation assessed in this study to help improve in the prediction of streamflow indices. If this variability can be adequately incorporated, these methods (i.e. the hydrologic regionalisation built using the framework of Chapter 5) and validation using machine learning methods such as RF should be particularly useful for predicting streamflow in ungauged basins. These approaches will permit inferences to be made regarding hydrologic process relationships (Hrachowitz et al., 2013) and represent a substantive step towards creating a framework that explicitly links climate, landscape and streamflow (Wagener et al., 2007).

6.4.2 Ability of environmental variables to predict flow classification groups

In the past, environmental variables have been used to predict a single or limited number of flow indices. For example, Santhi et al. (2008) used a range of geologically-relevant environmental variables (that had previously been used in delineating hydrologic landscape regions) to predict baseflow with excellent result for the conterminous United States. Mazvimavi et al. (2005) utilised neural networks and
multiple regression to predict average annual and monthly runoff and base flow for 52 catchment in Zimbabwe using catchment characteristics such as precipitation, lithology and slope as predictor variables. Water quality has also been estimated for both surface (Poor and Ullman, 2010) and groundwater (Rodriguez-Galiano et al., 2014) utilising a range of environmental and catchment variables with excellent results. In addition to such approaches, the one utilised in this study, to predict flow classification groups (based on a range of streamflow indices) using the environmental variables that were used to create the hydrologic regionalisation for Chapter 5, is also common in the literature (see Table II in Olden et al., 2012). Indeed, it has been suggested that this method (classifying a priori flow classes using environmental variables) may be more appropriate than regionalisation approaches (McManamay et al., 2012) because of the limited ability of hydrologic regions to explain variance in a series of individual flow indices.

Canonical analysis of our flow classification gave an initial indication of the ability of environmental characteristics to predict flow classes; while classification accuracies were modest, they were substantially better than those expected by chance alone. For example, a sample of 201 gauges randomly allocated into eight classes would result in class accuracies of approximately 12.5 %, purely by chance. Our results were also similar to those observed in the past when trying to relate flow indices to regionalisation classes (Chapter 5, Section 5.3.6). The CART and RF models developed to predict flow classification membership did so with accuracies of 61 and 69 %, respectively. These results are comparable to previous research in Australia where a number of variables describing catchment topography, geology, vegetation cover and
climate variables resulted in a model that could accurately predict an *a priori* flow class 62% of the time for 830 stream gauges (Kennard *et al.*, 2010b). Furthermore, with the exception of flow class IV, a classification tolerance of one class either side of the true class resulted in excellent classification results, with 66 to 95% of the samples correctly allocated to their *a priori* flow class (Table 6.4, Table 6.5). This is a significant improvement on the accuracies observed for the regionalisation approach, and may in fact be more useful as it could provide ‘ranges’ of the individual flow indices for each class. For example, classifying an ungauged basin to a particular class based on its environmental characteristics could include a confidence interval of sorts, calculated for potential streamflow behaviour by incorporating the streamflow information based on the classes either side.

The low levels of group agreement and precision between the flow and the regionalisation classes (Table 6.6) suggests that there is significant variability in the flow regimes of the streams within each of the regionalisation groups. For example, flow classes V, VI and VII all had at least one gauge from each of the regionalisation groups assigned to them. The variability within these classes is significant and suggests that gauges from mountainous upland areas (regionalisation group B) are behaving similarly to gauges from lowland, floodplain areas (regionalisation group J) (refer to Figure 5.8 for the spatial distribution of the regionalisation groups). Essentially, the regionalisation and flow classification groups used here are capturing different aspects of variability – one based on environmental indicators of hydrology, the other on streamflow. As a result, the most fitting approach will depend on the objectives of any given study. In general however, I believe that classifications based on flow regime may be more appropriate.
than regionalisation-based approaches when comprehensive stream gauge data are available (Kennard *et al.*, 2010b) unless the regionalisation can specifically account for spatial variability in a river network (Snelder *et al.*, 2005).

The lack of congruence across the two sets of classes may also be caused by the inclusion of stream gauges that were heavily regulated, such as those where significant extraction of water or damming had occurred. Regulating a stream and thus altering its flow regime would influence the flow class to which that stream was assigned. However, as the regionalisation was based on environmental variables, the impacts of regulation would not necessarily be apparent in the regionalisation classes. This could lead to regulated streams appearing in different flow classes than may have occurred under a regionalisation approach. While previous research has focused on modelling natural flow regimes and, therefore, explicitly excluded modified streams (Kennard *et al.*, 2010b), I believe that the inclusion of modified streams is important to identify impacts of that modification on streamflow behaviour. I believe that including regulated streams in flow-based classifications is likely to be more useful in many cases as it can be used to explain the variability in hydrologic response even in heavily-disturbed and hydrologically-modified catchments independent of scale effects on flow. That is, such an approach could demonstrate that upland streams can be modified to the extent that they represent lowland streams with regard to their streamflow characteristics, regardless of catchment size, which can have concomitant effects that influence instream and riparian ecology (Poff *et al.*, 1997).
6.4.3 Flow indices and environmental variables controlling the classifications

Being able to identify variables that can most easily discriminate between classes in regionalisation and flow classification studies can be useful in helping to develop conceptual models of streamflow regime (Kennard et al., 2010b; McManamay et al., 2012). At large scales (regions or continents), climatic gradients have been linked to catchment function while soil parameters and combinations of climatic, geologic and environmental variables have been identified as potential drivers of hydrologic variability at a variety of other scales (Kennard et al., 2010b; Sawicz et al., 2011; McManamay et al., 2012). For the regionalisation, with the exception of high flow variability (FH2), the CART model suggested that high-flow variables were best for discriminating between classification groups, while the RF model indicated that low-flow variables were best. The lack of consensus between the two methods, and between those methods and the KW statistics (Table 6.1), on the relative importance of the variables is fascinating and suggests that, depending on the method employed, different variables can be used to discriminate between groups. Both the RF model and KW statistics suggested that differences in the low-flow indices were more important in discriminating among groups, given that three of the four most important variables identified using each method was concerned with low-flow hydrology. While I was not able to find other studies assessing the strength of a regionalisation using flow indices in the manner I have presented here (i.e. using variable importance measures), McManamay et al. (2012) suggested that mean annual runoff, daily flow variability, predictability of flow and seasonal predictability of moderate to high flows (Colwell, 1974) were the four most important indices for discriminating between the regionalisation classes of Poff (1996).
The only two variables in common with our study were daily flow variability (MA3) and mean annual runoff (MA41) – two variables that were not considered important for either the CART or RF models, making comparisons to other regionalisation studies difficult.

For the flow classification, with the exception of landscape development intensity and soil erosivity, both CART and RF identified mean annual rainfall, rainfall in the wettest quarter and the mean aridity index as best able to discriminate among classes. These were also the three most important variables according to the KW statistic (Table 6.2). It is possible that these variables are important to the classification as they vary slowly in space and there were large amounts of spatial connectivity among the classes in our flow classification (Sawicz et al., 2011) (Figure 6.2). Another possible explanation is that 19 of the 33 streamflow indices (which were used to create our flow-classes) were concerned with annual streamflow regimes (Appendix II, Table 10.1) and, as such, variables that were important are more relevant at larger temporal scales. For example, variability in mean annual rainfall is likely to affect mean annual runoff more than daily flow variability. Therefore environmental variables that are relevant at annual time scales (such as mean annual rainfall) are more likely to have been identified as important.

Kennard et al. (2010b) found that mean rainfall in the coldest and driest quarters, mean August rainfall and mean March areal ET were the four most important predictor variables in their best-performing model. These results, along with mine, suggest that climate and not landscape variables control streamflow behaviour. However, Carlisle et al. (2010) found substantial variation in environmental drivers of streamflow
within what were believed to be relatively homogeneous hydrological regions of the United States. McManamay et al. (2012) suggested that the variables that create regionalisations may be more important than the regionalisation itself when attempting to predict flow classes or explain hydrologic variability. Given the higher classification accuracies that I achieved here using environmental variables to predict an a priori flow class relative to those observed for the regionalisation, my results support this previous work although I did not explicitly test the ability of my regionalisation to predict flow classes.

Contrasting results that have been found in the literature make it difficult to provide recommendations on which approach (regionalisation or flow-based classification) is more appropriate for predictions of hydrological behaviour in ungauged basins. Olden et al. (2012) suggested that regionalisation studies are appropriate for extending understanding of well-gauged regions to ungauged or sparsely-gauged regions, where similarities in environmental characteristics should influence hydrologic behaviour. In the absence of comprehensive stream gauge data, regionalisation studies can be used to infer streamflow behaviour, but when possible, flow-based classifications are likely more suitable.

6.4.4 Potential methods to improve the ability of the regionalisation to predict flow regimes

The classifications presented here have demonstrated that there is potential for the regionalisation framework from Chapter 5 to be useful for predicting hydrology in ungauged basins. The regionalisation identified significant amounts of variation in landscapes, even within catchments (Chapter 5, Section 5.3.5) and the CART and RF
models were both able to model this variability, as is evidenced by the low levels of spatial autocorrelation in model residuals (Appendix II, Figure 10.2). The CART and RF models were both designed to account for class imbalances in the sample datasets for the respective classifications and penalise misclassifications.

While applying a specific cost of misclassification in the CART models did not seem to affect final classification accuracies (i.e. there were no differences in pairwise group error rates between CART and RF – see Section 6.3.3), this could be a function of differences in sample sizes across groups. It may have been more appropriate to apply a higher cost of misclassification for classes with few cases or to increase the prior probability of their occurrence compared with the more common classes, to actively bias the classification towards them and therefore improve their classification rate. Previous research indicates that assigning higher prior probabilities for smaller classes leads to models that tend to predict these samples more effectively than if uniform misclassification rates are used (Breiman, 1993). Similarly for the RF models, increasing the prior probabilities of the under-represented classes may have increased the prediction accuracy of classes with few cases as a result of the internal bootstrapping procedure (Breiman, 2001; Liaw and Wiener, 2002). While the CART models would still have likely been outperformed by the RF models, an increase in their predictive accuracy could make extrapolating to other, ungauged regions, easier as CART models are very easily interpretable. However, I did not attempt that approach here, as there are assumptions about the representativeness of cases which may not be appropriate and are not able to be tested with the data available at present.
In contrast to the regionalisation, in the flow-based classification all classes bar one had at least 25 cases, meaning that low sample sizes were unlikely to have contributed to the misclassifications. The misclassifications are therefore likely due to the groups existing closely in multivariate space, which is evidenced by the much higher classification accuracies when a misclassification tolerance of one class was permitted (Table 6.4, Table 6.5). While there was not as much spatial variability in the flow classification (Figure 6.2) when compared to the regionalisation, the low levels of autocorrelation of model residuals (Appendix II, Figure 10.2) also support the use of the methods presented here (i.e. flow based classification) when modelling variability in hydrologic systems.

Within-group variance, particularly for the less common classes, may also have contributed to the low classification accuracies. For example, Group F (12 cases), had a small sample size and the largest average squared within-group distance (Table 6.1) – a combination which likely contributed to the poor classification accuracy for the group. However, Groups B (12) and C (10) had relatively small average squared distances in comparison. Another possibility is that the variability within the catchments may not have been adequately characterised by the relatively small number of flow indices that were used. Those indices were selected to be generally applicable across all of the catchment types found in the study region, as opposed to those that may have been better able to highlight different stream types (e.g. harsh intermittent or stable groundwater dominated perennial rivers; Olden and Poff, 2003). As there are no more stream gauges that fit the criteria to be included in this analysis, any future analyses could employ a greater range of flow indices (Olden and Poff, 2003; Kennard et al., 2003; Olden and Poff, 2003; Kennard et al., 2003).
(Chapter 5). Alternatively, Snelder et al. (2005) assessed a framework that explicitly subdivided larger catchments into smaller catchments by delineating sections of streams (and their respective catchments) into a number of groups based on climate and source-of-flow variables and then assessed the classification strength using 13 flow indices. Their results indicated that this “River-Environment-Classification” (REC) was stronger than two existing hydrological regionalisations and a stream classification based on climate than did not account for the river network. This increased classification strength was attributed to the explicit consideration of causes of spatial variation and the spatial structure of river networks within the REC.

The large amount of spatial connectivity among the flow classes (Figure 6.2), suggests that catchments that are near each other are likely to display similar levels of hydrological homogeneity. Spatial connectively and proximity among classes has been proposed as a good indicator of similarity when examining and attempting to predict flow classes (Sawicz et al., 2011). In a similar study, Kennard et al. (2010b) explored the inclusion of location data in their models, resulting in a model that was able to correctly classify 48 % of samples using only those location data. While I did not include gauge location, the spatial cohesion of my flow classes (Figure 6.2) suggests that including this information could improve model accuracy. However there is evidence that independent sections of river, with dissimilar catchment characteristics and thus hydrologic regimes, can be spatially close (Snelder et al., 2005), and conversely, there is evidence showing non-contiguity of certain streamflow regime classes where hydrology is similar (Kennard et al., 2010b). As a result, Kennard et al. (2010b) advise that restraint should be used
when extrapolating streamflow regime characteristics to ungauged areas, even when those areas are relatively close to gauged and classified catchments. Furthermore, the utilisation of fuzzy partitional clustering (e.g. Bayesian mixture modelling; Kennard et al., 2010b) with multi-label classification (Grigorios and Ioannis, 2007) where streams can be assigned to more than one class based on a probability measure may be an option to improve the results of the flow classification.

6.5 Conclusions

Different methods of hydrologic classification are becoming more popular in hydrology and ecohydrology. A previous hydrologic regionalisation developed a number of spatially-independent regions that removed the need to rely on defining an *a priori* spatial unit, such as a catchment, thus preserving variability within that spatial unit. An initial validation of the regionalisation (conducted in Chapter 5), indicated that there was some merit to the method and that there were links between streamflow behaviour and regionalisation classes. In this study, I built upon that previous validation using more advanced machine-learning methods that can easily handle complex, non-linear relationships between variables. The new validation supported previous assertions that the regionalisation had indeed captured significant amounts of landscape and hydrologic variability and suggested that it is possible to create a hydrologic regionalisation that does not require an *a priori* spatial unit. The regionalisation classification was also compared to a new classification of streams based on flow indices, an approach that is commonly used to examine links between streamflows and environmental characteristics (i.e. the variables that initially created the regionalisation). The results of this chapter suggest that, in the absence of a comprehensive stream gauge network, regionalisations
can be created to predict regional hydrology but, where gauge data are available, a flow-based classification, where environmental variables are summarised at the catchment scale, are likely more suitable for creating a classification that can be used for streamflow prediction.
**General discussion**

This thesis examined three different methods to quantify the spatial and temporal variability in water resources that has consistently been identified in environmental and hydrological research in the Glenelg-Hopkins catchment in Victoria, Australia. While the methods presented in this thesis were applied to hydrological systems, the principles behind them (e.g. assessments of spatial autocorrelation, spatial non-stationarity, and scaling issues) can be applied to the broader field of environmental science where similar variability is also relevant. For example, issues of spatial-autocorrelation, spatial heterogeneity or non-stationarity can be identified using methods such as GWR or spatially explicit structural equation models (SE-SEM). Such methods are beneficial when spatial phenomena are not adequately described by global modelling methods (e.g. OLS regression), and are particularly useful for identifying regions or locations where a suitable localised model could provide a better description. Subsequent analyses can then be spatially explicit or not, where the former can incorporate stationary or non-stationary decisions based on results of GWR or SE-SEM modelling.

The research attempted to address a number of knowledge gaps, using the Glenelg-Hopkins region as a case study; specifically: 1) did the introduction of *Eucalyptus* plantation forestry negatively affect streamflows during an extended drought period?; 2) how strong is the link, if any, between climate, land cover, soil variability and wetland extents in the region?; and 3) is it possible to use climatic and landscape variables to predict regional hydrology?
7.1 Research highlights

The outcomes of this research have contributed to the growing body of evidence that significant spatial and temporal variability exist across landscapes in the region (e.g. Ierodiaconou et al., 2005; Versace et al., 2008b), as is the case in many other regions worldwide. The outcomes have also contributed to a better understanding of water resources research and management in the region with implications for other similarly-variable regions elsewhere. The employment of a physically-based, semi-distributed hydrological model identified that climate extremes (i.e. drought) were far more influential in reducing streamflows than an increase in Eucalyptus plantation forestry; the utilisation of geographically weighted regression (GWR) to model 149 sub-catchments in the region identified and, for the first time, quantified the spatio-temporal variability that exists in relationships between climate, land cover, soils and surface water extents; while the design and application of a novel approach to hydrological regionalisation provided evidence of significant hydrological variation that could be identified independently of a comprehensive stream gauge network and without the need to use spatial units such as a catchment that needed to be defined a priori.

The application of the SWAT model (Chapter 2) to the two sub-catchments in the Glenelg-Hopkins region identified that the modelled sub-catchments were hydrologically heterogeneous, as a result of differing land covers, topography and climates (Table 2.1), yet demonstrated a similar lack of response when realistic land-cover changes were modelled. Analysis of model outputs under a variety of scenarios, including drought and land-cover change, demonstrated that the models could
satisfactorily predict streamflows and baseflows (Moriasi et al., 2007; Saha et al., 2014) – the first time such models have been created for the region. A major benefit of using SWAT over alternative empirical methods (Versace et al., 2008b; Yihdego and Webb, 2013) is that potential future land-cover changes, climate changes, and impacts of catchment management plans (e.g. the establishment of riparian buffer zones or changes in fertiliser and pesticide applications) can be easily incorporated into the model and assessed quantitatively or predictively (Gassman et al., 2007). Furthermore, once a SWAT model for a region has been successfully calibrated and validated, SWAT can be used to predict the impact of management on water, sediment, and agricultural chemical yields in nearby, ungauged catchments (Gassman et al., 2007). SWAT has had limited application in Australian hydrological research and, while the results here were not entirely unexpected, the design and application of the model to Australian hydrological conditions is important as it supports the use of physically-based semi-distributed hydrological modelling tools in Australia, and contributes to their application in low-flow systems such as those modelled here.

The next two research chapters (Chapters 3 and 4) employed the use of GWR, a relatively new regression method that can be used in the assessment of relationships that are thought or known to vary spatially. The models created for Chapter 3 examined spatio-temporal variability in the relationship between wetland extents, land cover and rainfall. Models in Chapter 4 built on this research and included an additional seven key soil properties that had not previously been available for inclusion in Chapter 3 but that can influence hydrology to examine the same relationships. Unfortunately, the models that were re-created in Chapter 4 produced contradictory results compared to Chapter
3, despite utilising the same variables, making it difficult to quantify the links between surface water, land cover, soils and climate in the region. This change appeared to be as a result of updates to the software applying the GWR, perhaps as a result of location correlations among predictor variables. Furthermore, when using a localised regression method such as GWR, issues of localised non-linearity and (multi-)collinearity could result in model misspecification which could influence the model coefficients and consequently interpretation (Wheeler, 2009). Thus, the results suggest that future environmental or hydrological research in the region needs to be based on careful consideration of the most appropriate modelling approach and, where possible, the impact of the modelling assumptions on model outputs need to be assessed (Zhang et al., 2005).

The finding that the inclusion of variables describing soil variability did not improve the relationship between climate, land cover and wetland extent in the region (according to model AICc) for one of the modelled time points (1995) was unexpected. This is because soil variability is known to affect runoff processes, primarily through differences in infiltration rates caused by antecedent soil moisture conditions and dominant runoff generation mechanisms (Castillo et al., 2003; Li and Sivapalan, 2011). The research did, however, demonstrate the superior ability of GWR to model spatially-varying relationships over ordinary least squares (OLS) regression, which is consistent with other hydrological studies that have utilised GWR (Chang and Psaris, 2013; Tu, 2013; Javi et al., 2014). The results of these two chapters add to previous research that suggests that studies concerned with any form of spatial analyses need to take the limitations of OLS and other similar linear regression methods into consideration and
investigate newer, more suitable methods when attempting to explain, or quantify spatial relationships and variability (Zhang et al., 2005).

The final two research chapters (Chapters 5 and 6) developed and validated a novel methodology to explain differences in regional hydrology using classification based on landscape and climate characteristics. The method describes the creation of a series of spatially-independent hydrological regions; regions that can exist in multiple disparate locations depending on multivariate similarity (Chapter 5). The benefit of this new method is that it reduced the number of a priori decisions that are required and enables classification of streams even in the absence of flow data. Unlike traditional methods that are based on a combination of environment and streamflow-response similarity (Olden et al., 2012), and require the use of catchments or some other spatial unit to be selected a priori (McManamay et al., 2012), the new framework does not require a priori unit selection, but is instead based on pixel resolution. As such, the method preserves intra- and inter-catchment variability that is typically lost in other hydrologic regionalisation studies (Poff, 1996; Kennard et al., 2010b; McManamay et al., 2012). Through creating a framework that retains this variability, it should be possible to use hydrologic regionalisations to make informed predictions about the hydrology of an ungauged, heterogeneous catchment (Snelder et al., 2005; Wagener et al., 2007; Hrachowitz et al., 2013).

Validation of the methodology (Chapter 6) using gauged streamflow data using more advanced machine-learning methods designed to handle complex, non-linear relationships among variables built upon a previous preliminary validation (Chapter 5). The new validation supported the use of the framework to create the regionalisation and
suggested that such a hydrologic regionalisation in the absence of an *a priori* spatial unit can be used to estimate relevant hydrologic properties within the landscape. Analysis suggested the regionalisation had captured significant amounts of spatial variability with 62% (125 of 201) of stream gauges included in the analyses being correctly allocated to their regionalisation class based on 33 flow indices. An optimistic method of including the preserved catchment variability from the regionalisation, whereby the proportional area of each class within each catchment was used as a predictor variable, suggested that classification accuracies could be as high as 83%. However, this was not ideal as there was a modest agreement between the highest proportional area of the classes and the gauge class which confounded the classification, and consequently these results were not presented (Chapter 6, Section 6.3.3.3). As such, no suitable method of including the preserved variability from the regionalisation has thus far been identified.

The final two chapters add to the existing research on hydrological prediction in ungauged basins by providing a new framework that, by design, specifically retains intra-catchment variability. Catchment heterogeneity has long hindered attempts to deepen the understanding of links between hydrological processes and landscapes, and hydrological regionalisation (or catchment classification) frameworks are perceived as a way of synthesising process understanding that is valid across multiple spatial and temporal scales (Hrachowitz *et al.*, 2013). Provided that the links between regional hydrology and variables describing climate and landscape can be elucidated, the frameworks presented in this thesis (i.e. for regionalisation and validation) will permit inferences to be made regarding hydrologic process relationships and represent a
substantive step towards creating a framework that explicitly links climate, landscape and streamflow.

7.2 A conceptual framework for investigating spatio-temporal hydrologic variation

The methods explored to address the knowledge gaps presented in this thesis ranged from a complex, physically-based, semi-distributed hydrological model (SWAT), to geographically weighted regression (GWR), and finally an investigation of hydrologic landscape regionalisations. While in this thesis, the methods were presented in order of increasing scale (SWAT \(\rightarrow\) GWR \(\rightarrow\) Hydrologic Landscape Regionalisation), a possible conceptual framework (Figure 7.1) that allows each of the methods to build on the results of the previous investigation, would see the methods used in the reverse order. Using the presented conceptual framework (Figure 7.1), preliminary investigations for determining whether a stream gauge network is suitable for such an approach would need to incorporate guidelines on both spatial and temporal coverage of flow data. Kennard et al., (2010a) provide guidelines on suitable temporal coverage, however no such guidelines have been identified regarding spatial coverage. If a gauge network is considered ‘sparse’ (in regards to spatial coverage and temporal continuity/availability of flow data), then the regionalisation approach presented in Chapter 5 may well be suitable, however some flow data will still be required to validate the hydrologic landscape regionalisation. If however the network can be considered ‘comprehensive’ then the flow based regionalisation presented in Chapter 6 may be more appropriate. This approach however, requires the availability of suitable, spatially explicit environmental datasets (e.g. climate, soils, topography) for validation.
Figure 7.1: Graphical design of the conceptual framework linking the methods that have been applied in each of the research chapters. ‘Sparse’ and ‘Comprehensive’ stream gauge networks refer to the spatial coverage of stream gauges, and the temporal continuity/availability of measured streamflow data. The solid lines show suggested inputs to each of the modelling stages, while the dashed lines illustrate the connection between the methods and a detailed assessment of spatio-temporal hydrologic variation.
Following the creation and validation of a suitable hydrologic regionalisation, assessments between land cover and regional hydrology can then be explored with methods such as GWR or SE-SEM. These assessments could utilise the previously created hydrologic regions as a form of stratification, as it can be reasonably expected that hydrological functioning will be different within each of the hydrologic regions (McManamay et al., 2012). In the presence of a sparse stream gauge network, methods such as those presented in Chapters 3 and 4 which do not require streamflow measurements, could be conducted – these methods are suitable for an assessment of inter-relationships among land cover, topography, and climate. While in the presence of a comprehensive stream gauge network it should be possible to also incorporate measures of streamflow, for example through the use of streamflow indices (e.g. Appendix II, Table 10.1) as covariates, which could further elucidate these relationships. The ability of these methods to accurately model the relationships between land cover, topography, climate and soils could be improved through the application of geographically weighted ridge or lasso regression (Wheeler, 2009) – particularly in the presence of local collinearity between explanatory variables.

Subsequent to the analysis of the relationships between land cover, topography and climate; and to the creation of hydrologic regions, the application of process based hydrologic model such as SWAT can then be conducted. There are a number of difficulties in the application of physically based models such as SWAT, and these are mainly concerned with data availability and suitability. Typically, these models are data intensive, and require long-term daily streamflow measurements for calibration and validation. As such, these methods are only suitable when a stream gauge network can be
considered comprehensive (specifically, with regards to temporal continuity/availability of flow data). By incorporating process based hydrologic models as the last stage of the conceptual framework, they should ideally, be able to model streamflows more accurately and more influential relationships between hydrology and environmental conditions should be able to be explored. For example, Carlisle et al., (2010) demonstrated increases in the accuracy of a predictive streamflow model when the model was stratified by hydrologic regions. This is a result of streamflow (measured at any point – e.g. a stream gauge) being a cumulative function of upstream processes; therefore stratifying by hydrologic regions provides a mechanism to improve the performance of SWAT models. Examination of GWR results could also be used during the application of a SWAT model by identifying areas that are spatially auto-correlated – and as such should behave similarly in regards to hydrology. Furthermore, these simple methods (hydrologic landscape regionalisations and GWR) may provide some insight into relationships between land cover and regional hydrology. If for example, a particular land cover is not important in predicting the presence/amount of wetland habitat, it may not contribute much to regional hydrological functioning, and may therefore not influence the output of a SWAT model.

The framework presented in Figure 7.1 provides a simplistic, conceptual connection between the methods presented in this thesis. Through a detailed analysis of model results (e.g. the location and occurrence of hydrologic regions, spatial assessment of model residuals and coefficients, and predicted streamflows) it is possible to combine all the methods to permit an assessment of spatio-temporal variability in hydrology such
has been presented here. While the results presented here are in order of increasing spatial scale, the methods were conducted following such a framework.

7.3 Methodological implications for hydrological and environmental science

The research in this thesis presented a number of methods that could be used to investigate spatially (and temporally) varying relationships. Applying and critically evaluating the methods employed here highlighted a number of methodological issues that are relevant for the application of these approaches in other regions. If these issues can be addressed, these methods are likely very suitable for utilisation in a range of environmental studies in highly-modified, mixed-used catchments such as the Glenelg-Hopkins.

In the past, simpler empirical approaches for examining land-cover impacts on hydrology in the region have been used, while the application of physically-based models such as SWAT has been critiqued as being time consuming and data intensive (Yihdego and Webb, 2013). However, while the SWAT models created here are more complex, they are also better suited to predicting future hydrologic conditions (under a range of land-cover scenarios that could occur in tandem with extreme climatic conditions such as extended droughts) than the monthly empirical water balance models than have been used recently to assess LULCC impacts on streamflows (Yihdego and Webb, 2013). As such, they are the next logical step in improving the understanding of regional hydrological conditions.

The availability of high-resolution digital elevation models (Geoscience Australia, 2011), land cover (Ierodiaconou et al., 2005) and climatic datasets (Bureau of...
Meteorology, 2012), for the region made this approach possible, and the unavailability of similar data previously has likely precluded the use of such methods in the past (Versace, 2007). Unfortunately, the only suitable soil dataset available at the time of model creation was coarse in comparison to other datasets (FAO, 2007). The resolution of soil datasets in SWAT models has been shown to affect modelled sediment and nutrient yields (Romanowicz et al., 2005), but the impact on modelled streamflows changes inversely relative to catchment scale (Geza and McCray, 2008; Moriasi and Starks, 2010; Li et al., 2012). Future research using the models created in the study, particularly for any nutrient or sediment assessments, will need to cautiously interpret model outputs unless suitable, finer-scale soil databases become readily available, in which case the models will need to be re-conceptualised.

Previous research in the region examining links between land cover and instream salinity (Versace et al., 2008b) has likely been confounded by localised variations in geomorphology and climate (Yihdego and Webb, 2013) and a major advantage of employing the SWAT model was that this variability could be accounted for explicitly. Although this is largely an advantage of utilising a complex, physically-based model, any reduction in confounding effects can only improve model validity and interpretation. As outlined in Section 7.1, a major benefit of the use of models such as SWAT is that, once they are created, they can be used for the assessment of a range of catchment management goals (Glenelg Hopkins CMA, 2006a, 2013), from the estimation of pollutant loads from non-point sources, crop yields, vegetation biomass and climate change impacts through to simple hydrologic assessments (Gassman et al., 2007).
Previous criticisms of models such as SWAT (Yihdego and Webb, 2013) are somewhat justified as the models produced for this study were very time consuming to create. However, their effectiveness as a tool to assist in future management and investigation into regional hydrology are likely to outweigh the negatives associated with the time to produce the models, particularly when assessing the effects of multiple concurrent disturbances, or when wanting to predict the outcomes of a range of management plans on hydrology. Outputs required from hydrological models are typically simple (Arnold et al., 1998; Todini, 2007), and this research was no exception, with only modelled streamflow and evapotranspiration required to answer the research question for this part of the thesis. However, the task of assessing the relative impacts of land-cover change and climate in large reference catchments that undergo gradual rates of systematic change (Versace et al., 2008a) necessitated the use of hydrological models (Yihdego and Webb, 2013), rather than following most other land-cover change assessments which have been conducted in smaller, paired catchments with typically instantaneous land-cover changes (Brown et al., 2005).

In contrast to the time-consuming, data-intensive SWAT models, the GWR modelling of the relationship between land cover, climate, rainfall, and wetland extents was simple, yet statistically sound. There are many examples of the employment of GWR in the literature and the method has broad applications across a number of fields from health (Nakaya et al., 2005) to aquatic science (Tu and Xia, 2008; Chang and Psaris, 2013; Javi et al., 2014). Due to the prevalence of the method in aquatic (and environmental) science (Chapter 1, Section 1.6.2; Chapter 3, Section 3.1.4) it was determined that GWR would present a suitable method for modelling relationships in
environmental variables in the region. The aggregation of the land classes from the 11 (for 1980 and 1995) and 12 (for 2002) (as defined by Ierodiaconou et al., 2005) to only four classes that were included in the model, describing similar land covers (e.g. the agricultural land class was the aggregation of dryland cropping, dryland pasture, irrigated agriculture and irrigated pasture; see Chapter 3, Section 3.2.3 for details of the other aggregated land classes) minimised issues of multi-collinearity among predictor variables (Versace, 2007), while the log transformation of the dependent variable (area of wetland extent [km²]) helped with assumptions of normality and variance of model residuals (Quinn and Keough, 2002). However, it has been shown that localised correlation in predictor variables (which may not be as readily obvious as global correlations) will result in strongly-correlated GWR model coefficients and, hence, inferences between variables can be problematic (Wheeler, 2009).

Current modelling practices largely revolve around the assessment of model performance in regards to overall model accuracy, while relatively little attention is given to assessment of the spatial heterogeneity of model error (Zhang et al., 2005). The methods used in this thesis (Chapter 3, Sections 3.2.7 and 3.2.8), however, specifically investigated the global versus local spatial autocorrelation of model residuals and found that GWR was more appropriate than OLS for investigating the relationships explored. Specifically, all the GWR models produced higher $R^2$ and were less likely to exhibit global spatial autocorrelation of residuals than the analogous OLS models. Furthermore, the GWR models also demonstrated very low levels of localised spatial autocorrelation of model residuals, indicative of a suitable spatially varying model (Figure 3.6). GWR has also been shown to provide better localised prediction results than other techniques.
including non-linear neural networks, which generate spatial patterns of model residuals similar to OLS (Zhang et al., 2005). Whilst other methods may produce more accurate models (Zhang et al., 2005), my study suggests that methods such as GWR, which allow for model coefficients to vary spatially (Fotheringham et al., 2002), are likely more appropriate for spatially-explicit data, as they can indicate where spatial non-stationarity is occurring in the model. Such methods are likely to be appropriate given that assessments with GWR are typically associated with measuring and then interpreting statistically-significant variation in regression coefficients, as opposed to fitting a curve to a response variable for prediction (Fotheringham et al., 2002; Wheeler, 2009) and so can indicate where locally-weighted regression coefficients deviate from their global ‘fixed’ values.

Due to the restrictive assumptions of OLS and similar regression methods, which are generally violated in environmental science (Zhang et al., 2005; Tu and Xia, 2008), the emergence of additional problems such as spatial autocorrelation (typically indicative that a relationship is spatially non-stationary) are predictable. Recent improvements to GWR, which now allow some variables to be modelled globally (i.e. fixed), while others are allowed to vary spatially (i.e. local), essentially make GWR similar to a partial linear or mixed effects model (semi-parametric GWR), while still permitting observations closer together to have more impact on each other than on observations further apart (Fotheringham et al., 2002; Nakaya et al., 2009). In addition, penalisation methods such as Geographically Weighted Lasso Regression (GWR-L) have been shown to lower prediction and estimation error of the response variable, and lessen issues of localised correlation of predictor variables (Wheeler, 2009). Future work
investigating the relationships explored here could investigate these model improvements (GWR-L), particularly in light of the counter-intuitive regression coefficients observed in Chapter 4 which were developed using semi-parametric GWR (Section 4.3, Table 4.1).

The application of GWR in modelling the spatial variability in the relationships among land cover, rainfall, climate, soil attributes and wetland extents in this case study region (Chapters 3 and 4) is, to my knowledge, the first attempt to quantitatively model the spatial variability in hydrology. While the models demonstrated that GWR greatly outperformed analogous OLS models, with higher $R^2$ and very minimal global and local spatial autocorrelation (Chapter 3), there were contrasting results between the chapters (Chapters 3 and 4) making quantitative assessments of the relationship difficult. Nonetheless, the research still contributes to the growing body of evidence that methods such as GWR that allow model coefficients to vary based on nearby observations are likely to be suitable for use in environmental and hydrological research. Studies concerned with any form of spatial analyses need to take the limitations of OLS and other more advanced methods (e.g. neural networks; Zhang et al., 2005) into consideration and investigate newer, more suitable methods when attempting to explain spatial relationships – particularly with regards to model residuals.

Building on the evidence of spatial variability in regional hydrology identified by the GWR methods, the research in Chapters 5 and 6 conceptualised, designed, and validated a new framework for hydrological regionalisation. Recent initiatives in hydrology have revolved around prediction in ungauged basins (PUB; Hrachowitz et al., 2013) and have recognised that, in data-scarce regions as opposed to regions with
extensive stream gauge networks, the need to infer hydrological functions and controls from metrics related to catchment form would be beneficial (Hrachowitz et al., 2013). The ability to accurately predict streamflow regimes at ungauged sites is not only important for water resources management (Sanborn and Bledsoe, 2006), but is increasingly important in ecohydrology (Poff et al., 1997; Bunn and Arthington, 2002; Kennard et al., 2010b). The new framework differed from traditional regionalisation studies which require the a priori use of catchments or other appropriate spatial units (e.g. Poff, 1996). Instead, membership of pixels was defined qualitatively with the random forest classifier based on a statistical classification of a number of environmental variables that could have a direct influence on the hydrologic cycle. A thorough literature review did not reveal any previous work employing this proposed approach and therefore the application of deductive reasoning and hybrid classification is considered to be a novel approach to hydrological regionalisation.

The validation process (Chapter 6), however demonstrated that there was not complete congruence between this approach and a flow-based classification and, to date, a suitable method to incorporate intra-catchment variability into the validation has not be identified. The inherit connectivity and spatial structure of riverine networks and catchments makes it difficult to produce a regionalisation that can effectively retain this variability (Snelder et al., 2005). Considering this, it has been suggested that the classification of streams (or catchments) may be more appropriate than regionalisation approaches because of the limited ability of hydrologic regions to explain variance in a series of flow indices (McManamay et al., 2012). However, the most fitting approach will depend on the objectives of a given study. When comprehensive stream gauge data
are available classifications based on flow regime may be more appropriate than regionalisation-based approaches, unless the regionalisation can specifically account for spatial variability in the river network (Snelder et al., 2005), or the variability can be maintained and incorporated into the validation of the regionalisation. Hydrological frameworks such as the one created and validated here, work under the assumption that there are broad, landscape-scale patterns that will be represented in a physical system (i.e. hydrological response; Sawicz et al., 2011) or a biotic system (i.e. influence instream biotic communities; Bunn and Arthington, 2002; Naiman et al., 2008). While the work in this thesis identified that such landscape scale patterns do exist in regional hydrology, managers and researchers will need to exercise caution when selecting variables for future frameworks that may be used in a biological assessment of river classification (McManamay et al., 2012) as different aspects of the flow regime, particularly variability, affect aquatic biodiversity and resilience (Poff et al., 1997; Bunn and Arthington, 2002; Naiman et al., 2008). As such, any future frameworks that are designed to specifically assess riverine biodiversity will need to ensure that streamflow variability is inherently preserved and modelled accurately in such a framework.

7.4 Uncovering spatial variability in the Glenelg-Hopkins catchment

A common theme amongst previous studies conducted in the Glenelg-Hopkins region was that there was significant spatio-temporal variability in the regions hydrological and environmental relationships. For example, Ierodiaconou et al. (2005) demonstrated that rapid land-cover change between 1980 and 2002 resulted in an increase in modelled nitrogen and phosphorous loadings, and that there was considerable regional variation in nutrient yields; while Versace et al. (2008b) noted an
inverse relationship between in-stream salinity and a higher proportion of native vegetation cover across the same time period. Furthermore, a lack of historical and contemporary spatial datasets related to land-cover, soils and climate had precluded the use of complex hydrological models to examine impacts of climate and land cover on regional water resources (Versace, 2007). Land cover in the region has historically undergone a number of random and systematic transitions (Versace et al., 2008a) and, as such, water and land managers in this region need a greater understanding of where regional river and surface water habitats are likely to be affected by future climate and land cover changes.

The results of this thesis provide strong empirical evidence of significant spatial variability in the relationships explored. Chapter 2 (‘1.10 Objectives of the thesis – Objective 1’) identified a very similar response to a modelled land-cover change in two hydrologically-heterogeneous catchments – suggesting a distinct lack of spatial variability in hydrology. The variability in the region did, however, become apparent during the interpretation of the analysis conducted in Chapters 3 (‘1.10 Objectives of the thesis – Objective 2’), and 4 (‘1.10 Objectives of the thesis – Objective 2’). Further evidence of the extent of variability in regional hydrology was apparent with the production of the hydrological regionalisation in Chapter 5 (‘1.10 Objectives of the thesis – Objective 3’). While no direct assessment of spatial variability in the regional hydrology of the Glenelg-Hopkins region was conducted in Chapter 6 (‘1.10 Objectives of the thesis – Objective 4’), the validation analyses of the regionalisation indicated that the framework of Chapter 5 was effective in predicting hydrology in the absence of a comprehensive stream gauge network.
Anecdotal evidence had suggested that an observed reduction in streamflows in the region could be linked to the introduction of *Eucalyptus* forestry. Despite this, a quantification of the reduction in streamflows due to the independent effects of *Eucalyptus* and drought in the region had not been conducted, though assessments have been conducted for other historical land-cover changes (Yihdego and Webb, 2013). The models were able to accurately simulate streamflows in both sub-catchments over a long time period (1980-2009), and a model that did not include the introduction of *Eucalyptus* suggested that the introduction of the plantations was not extensive enough to alter streamflows significantly (Chapter 2, Section 2.3.4). This lack of response to the introduction of *Eucalyptus* suggests that the major reductions in observed streamflows within the region were due to the concurrent extended drought conditions. This is a significant finding for the region, particularly in the management of water resources in the face of climate change, and possible future land-cover changes (Glenelg Hopkins CMA, 2006a). More specifically, the models provide a basis for making estimates of the water yield impacts of any future broader-scale afforestation or revegetation in the region, both of which are embedded in current management policy (Glenelg Hopkins CMA, 2006a, 2013). However, in 2002, *Eucalyptus* plantations covered 5% of the Glenelg-Hopkins region and their introduction was systematic (at the expense of dryland pastures; Versace *et al.*, 2008a). This systematic expansion has resulted in some sub-catchments undergoing significant greater amounts of land-cover change relative to others in the region, and the response that we observed here will not necessarily hold true for all sub-catchments in the region (Benyon *et al.*, 2008).
Although spatial environmental variability has been observed in a number of studies in the past (Ierodiaconou et al., 2005; Versace et al., 2008b), the strongest and most variable relationships between environmental factors and hydrological conditions have consistently been found in the Glenelg catchment. Modelling results from Chapters 3, 4 and 5 found that the Glenelg catchment exhibited the most variability in the relationships explored, relative to those observed in the Hopkins catchment. The lack of variability in the Hopkins catchment has been attributed to the degraded nature of the catchment (Versace et al., 2008b); while coefficients from the GWR models in Chapters 3 (Figure 3.5) and 4 (Table 4.1) suggest that there is significant spatial, and temporal, variability between land cover, soils and climate, that affects the extent of wetlands throughout the region (i.e. including the Hopkins catchment). Further evidence of the spatial variability in the region is apparent from Chapter 5 which used the hydrological regionalisation, developed at a much larger scale, to examine spatial variability in the Glenelg-Hopkins region. Previously modelled relationships (e.g. Versace et al., 2008b) were likely confounded by variations in geomorphology, groundwater levels and other catchment or landscape variables (Yihdego and Webb, 2013) – factors that were accounted for in the regionalisation. This suggests that, with regard to hydrology, the variability apparent in the regionalisation, and the GWR models, is likely a good indicator of the state of intra- and inter-catchment variability in the region.

Water resource managers need to account for such variability in intra- and inter-catchment hydrology because of its potential effect on river management and restoration plans (Kennard et al., 2010b; McManamay et al., 2012), and frameworks
such as the one developed in Chapter 5 ought to allow such assessments to be made. Interestingly, the hybrid classification process (Chapter 5) identified varying amount of
spatial variability in the case study region. A smaller number of regionalisation classes
was observed in the eastern half (Hopkins catchment) relative to that observed in the
west (Glenelg) which concurs with previous research (Chapters 3 and 4; Versace et al.,
2008b) that suggests there is more spatial variability in relationships between
environmental variables in the Glenelg catchment.

While a specific assessment of the variability in the flow classification approach
was not conducted for the Glenelg-Hopkins region, as opposed to the Glenelg-Hopkins-
specific assessment of the regionalisation in Chapter 5 (Section 5.3.5), the spatial
arrangement of flow classes in Figure 6.2 suggests that there was very little variability in
the flow regimes of the streams in the Glenelg-Hopkins region. This could explain how
the sub-catchments that were modelled with SWAT, despite being hydrologically
heterogeneous, both displayed a similar lack of response to the modelled land-cover
changes (Chapter 2). Furthermore, the flow classification suggested that, for there to be
such minimal variability in the Glenelg-Hopkins catchment in spite of the wide range of
hydrological regions (Chapter 5), regulation or management of flows is likely affecting
regional hydrology (Chapter 6, Section 6.4.2) (Arthington and Pusey, 2003). This
accords with the previous view that degradation may be influencing the amount of
variability observed in some sub-catchments (Versace et al., 2008b).

The Glenelg-Hopkins Catchment Management Authority’s Native Vegetation
Plan (Glenelg Hopkins CMA, 2006a) has broad aims to cover 30 % of the region with
native vegetation by 2030 and it is not unreasonable to expect this change to exert more
pressure on the already-stressed water resources of the region. To protect the water resources of the region it is essential to implement management strategies that ensure that environmental degradation is minimised; therefore, it is critical that catchment managers have access to tools and models that will enable them to predict the impacts of future land cover and climate changes (Glenelg Hopkins CMA, 2006a; Department of Sustainability and Environment, 2008a; Glenelg Hopkins CMA, 2013). Moreover, water resource managers in the region need to take into account possible differences in intra- and inter-catchment hydrology that could drastically affect river management and restoration plans and studies such as those presented in this thesis could assist in identifying that variability (Poff et al., 1997; Kennard et al., 2010b).

7.5 Future work

The research presented in this thesis has built upon previous work in the Glenelg-Hopkins region, and has demonstrated a number of methods that can be utilised to model spatio-temporal variability in hydrologic systems. In order to facilitate that utilisation, there are a number of recommendations that should be considered to improve future, similar studies in the region. Specifically:

1) The framework presented for the regionalisation could be improved by utilising fuzzy (i.e. probability based; Kennard et al., 2010b) and multi-label classification (i.e. where each sample can be assigned more than one class; Grigorios and Ioannis, 2007) approaches for validation. The dramatic increases in accuracy rates when samples in the analyses presented here were allowed a tolerance of one class, suggests that an approach utilising this method may be more useful in terms of increased model accuracy and therefore, applicability in predicting flows in ungauged basins. The flow-
based classification could also be improved by utilising the methods presented in Chapter 5, specifically those related to selecting the number of classes for clustering. Further improvements could also incorporate categorical variables with, for example, information such as whether or not a catchment is heavily regulated (e.g. presence of large impoundments within the catchment) also included in the clustering process.

2) The GWR models may be able to be improved by utilising a modified hydrological response units (HRU) instead of catchments as the unit of analysis. As a HRU is representative of an area with unique land cover, topographical and soil attributes, an analysis conducted at this scale could be more appropriate than at the sub-catchment level employed here. To improve on our analysis however, the HRUs would need to be created using only soil and topographical information, as land cover information would be required as predictor variables. Conceptualising hydrological models using HRUs has been found previously to be a suitable method for regional hydrological modelling (Flügel, 1995) and the relationships explored here could also benefit from this approach, particularly if it is able to overcome potential issues of localised auto-correlation of predictor variables (Wheeler, 2009).

3) The SWAT models can, and should, be extended to cover more of the Glenelg-Hopkins region. This however, will only be possible if a new land-cover map is produced, and if detailed information on agricultural management practices (particularly on crop-rotations, planting and harvest data, irrigation and agro-chemical application rates and timing) in the region are available. The extent of inter-catchment transfers of water (if any), and the timing and magnitude of controlled releases from impoundments would also be required. Furthermore, the development and availability of suitable soil
data at a finer scale than that employed here, would allow more accurate assessments of sediment loadings in regional streams to be conducted. The appropriate scale for the model is limited, however, by the distribution and availability of long-term stream-gauge records necessary for calibration and validation.

4) The previous land-cover map (Ierodiaconou et al., 2005), which has been vital to research and management in the region, is now 13 years old and, therefore, a high priority should be the generation of a new land-cover map for the Glenelg-Hopkins region. While this recommendation is not new (Versace, 2007), there has, as far as I am aware, not been a coordinated effort to produce a newer, post-2002 version. An updated land-cover map is vital for quantifying changes in the region post-2002 (e.g. Versace et al., 2008a), and would also be useful for relating land-cover changes to changes in streamflows and regional wetlands (Chapters 2, 3 and 4; Yihdego and Webb, 2013).

7.6 Conclusions

Overall, this thesis has demonstrated three different methods of identifying and quantifying hydrologic variability. The application of the SWAT model to two hydrologically-dissimilar catchments demonstrated that climate has been far more influential in altering streamflows than the introduction of Eucalyptus plantation forestry. GWR has, for the first time, quantified the breadth and magnitude of variability in relationships between environmental variables that has been suggested by other research in the region. Finally, the design and validation of a new framework for creating a hydrologic landscape regionalisation has further highlighted the variability in environmental relationships in the study region, but also provided a method for use in the absence of detailed hydrologic data, or when the use of a priori spatial units is
undesirable. While, for this thesis, hydrologic variability was assessed, the principles used in this thesis are applicable to many areas of environmental research. Provided that the model is specified correctly and all assumptions are met (including those of no or minimal localised correlation among predictor variables), simple methods such as GWR can be used as a first approach to investigate a relationship that is believed to vary spatially; regionalisation studies can be used to further examine spatially-varying relationships and, provided that suitable data exist for some part of the study region for validation, such regionalisations can then be used for making predictions in other parts of the study region where no suitable data exist; and, finally, deterministic or mechanistic models can be applied when specific, quantifiable results are necessary provided that the (typically) extensive data requirements for model design and calibration are met.
References


CSIRO (2012) *Climate and Water Availability in South-Eastern Australia: A Synthesis of Findings from Phase 2 of the South Eastern Australian Climate Initiative (SEACI)*. Australia: CSIRO.


Department of Sustainability and Environment (2008a) *Climate Change in the Glenelg Hopkins Region*; Department of Sustainability and Environment, editor. Melbourne: Department of Sustainability and Environment.

Department of Sustainability and Environment (2012a) *Victorian Statewide Watertable Mapping*. Melbourne: Department of Sustainability and Environment.

Department of Sustainability and Environment (2012b) *Victorian Statewide Watertable Salinity Mapping*. Melbourne: Department of Sustainability and Environment.


Appendix I

9.1 Supporting tables
Table 9.1: The variables used in the creation of the hydrological regionalisation. A number of variables describing the storage, transport and release of surface water, groundwater and atmospheric water were included in the analysis. *The DTM data was resampled to 30 m to enable geo-TIFF compatibility with ENVI 4.8.

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Abbreviation</th>
<th>Description</th>
<th>Source</th>
<th>Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plant AWC – A Horizon</td>
<td>A_PAWC</td>
<td>Plant available water capacity A horizon (mm)</td>
<td>Western and McKenzie (2006)</td>
<td>1 km</td>
</tr>
<tr>
<td>Plant AWC – B Horizon</td>
<td>B_PAWC</td>
<td>Plant available water capacity B horizon (mm)</td>
<td>Western and McKenzie (2006)</td>
<td>1 km</td>
</tr>
<tr>
<td>Layer Thickness – A Horizon</td>
<td>A_THICK</td>
<td>The weighted average A horizon thickness (m)</td>
<td>Western and McKenzie (2006)</td>
<td>1 km</td>
</tr>
<tr>
<td>Layer Thickness – B Horizon</td>
<td>B_THICK</td>
<td>The weighted average B horizon thickness (m)</td>
<td>Western and McKenzie (2006)</td>
<td>1 km</td>
</tr>
<tr>
<td>Saturated Conductivity – A Horizon</td>
<td>A_KSAT</td>
<td>The weighted average of median A horizon saturated hydraulic conductivity (mm/hour)</td>
<td>Western and McKenzie (2006)</td>
<td>1 km</td>
</tr>
<tr>
<td>Saturated Conductivity – B Horizon</td>
<td>B_KSAT</td>
<td>The weighted average of median B horizon saturated hydraulic conductivity (mm/hour)</td>
<td>Western and McKenzie (2006)</td>
<td>1 km</td>
</tr>
<tr>
<td>Soil Erosivity Index</td>
<td>SOIL_EROS</td>
<td>Rainfall erosivity (MJ.mm/ha.hour.year) for soils.</td>
<td>Lu and Yu (2002)</td>
<td>5 km</td>
</tr>
<tr>
<td>Groundwater Static Water Level</td>
<td>GW_SWL</td>
<td>Static water level of groundwater aquifer (MASL)</td>
<td>Department of Sustainability and Environment (2012a)</td>
<td>100 m</td>
</tr>
<tr>
<td>Groundwater Total Dissolved Solids</td>
<td>GW_TDS</td>
<td>Groundwater aquifer salinity (TDS)</td>
<td>Department of Sustainability and Environment (2012b)</td>
<td>100 m</td>
</tr>
<tr>
<td>Mean Maximum Temperature</td>
<td>MAX_TEMP</td>
<td>Mean maximum annual temperature (°C)</td>
<td>Bureau of Meteorology (2012)</td>
<td>2.5 km</td>
</tr>
<tr>
<td>Mean Minimum Temperature</td>
<td>MIN_TEMP</td>
<td>Mean minimum annual temperature (°C)</td>
<td>Bureau of Meteorology (2012)</td>
<td>2.5 km</td>
</tr>
<tr>
<td>Temperature Seasonality</td>
<td>BIO04</td>
<td>BIOCLIM 4</td>
<td>Hijmans et al. (2005)</td>
<td>1 km</td>
</tr>
<tr>
<td>Mean Temperature of Wettest Quarter</td>
<td>BIO08</td>
<td>BIOCLIM 8</td>
<td>Hijmans et al. (2005)</td>
<td>1 km</td>
</tr>
<tr>
<td>Mean Annual Rainfall</td>
<td>RAIN_ANNUAL</td>
<td>Mean annual rainfall (mm)</td>
<td>Bureau of Meteorology (2012)</td>
<td>2.5 km</td>
</tr>
<tr>
<td>Precipitation Seasonality</td>
<td>BIO15</td>
<td>BIOCLIM 15</td>
<td>Hijmans et al. (2005)</td>
<td>1 km</td>
</tr>
<tr>
<td>Precipitation of Wettest Quarter</td>
<td>BIO16</td>
<td>BIOCLIM 16</td>
<td>Hijmans et al. (2005)</td>
<td>1 km</td>
</tr>
<tr>
<td>Precipitation of Driest Quarter</td>
<td>BIO17</td>
<td>BIOCLIM 17</td>
<td>Hijmans et al. (2005)</td>
<td>1 km</td>
</tr>
<tr>
<td>Mean Annual Evapotranspiration</td>
<td>ET_ANNUAL</td>
<td>Mean annual evapotranspiration (mm)</td>
<td>Bureau of Meteorology (2012)</td>
<td>10 km</td>
</tr>
<tr>
<td>Variable Name</td>
<td>Abbreviation</td>
<td>Description</td>
<td>Source</td>
<td>Resolution</td>
</tr>
<tr>
<td>----------------------------</td>
<td>--------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>Aridity Index</td>
<td>ARIDITY_INDEX</td>
<td>Mean annual precipitation / Potential evapotranspiration (mm/mm)</td>
<td>Produced in ArcGIS 10.1</td>
<td>2.5 km / 10 km</td>
</tr>
<tr>
<td>Landscape Development Intensity</td>
<td>LDI</td>
<td>An index for measuring anthropogenic loadings on landscapes. Values range from 1-10 on a normalised natural log scale, with higher values indicating more intensive landscape development/use.</td>
<td>Produced in ArcGIS 10.1 Brown and Vivas (2005); Bureau of Rural Sciences (2010)</td>
<td>50 m</td>
</tr>
<tr>
<td>Elevation</td>
<td>ELEV</td>
<td>Elevation (metres above mean sea level [MASL])</td>
<td>Department of Sustainability and Environment (2008b)</td>
<td>30 m*</td>
</tr>
<tr>
<td>Slope</td>
<td>SLOPE_RAD</td>
<td>Slope steepness from elevation dataset (radians)</td>
<td>Produced in ArcGIS 10.1</td>
<td>30 m</td>
</tr>
<tr>
<td>Topographic Wetness Index</td>
<td>TWI</td>
<td>High TWI values represent drainage depressions; lower values represent crests and ridges. $TWI = \ln(\alpha/\tan \beta)$, where $\alpha$ = upstream contributing area, $\beta$ = slope in radians</td>
<td>Produced in ArcGIS 10.1</td>
<td>30 m</td>
</tr>
<tr>
<td>Weathering Intensity Index</td>
<td>WEATH_IND</td>
<td>Ranges in value from 1 – 6. Low values indicate unweathered bedrock; high values indicate heavily weathered rock</td>
<td>Wilford (2012)</td>
<td>100 m</td>
</tr>
</tbody>
</table>
Table 9.2: SIMPER results from PRIMER 6. Numbers represent percent contribution of each of the variables to the ALOC 23 meta-groups on a standardised Euclidean distance matrix. Meta-groups C, H and J only contained 1 ALOC cluster and therefore percent contribution could not be calculated using SIMPER, so these were omitted from the table. KW = Kruskal-Wallis statistic, with higher values indicating a better ability of that variable to discriminate between clusters. All KW values were significant at $P < 0.001$.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Avg. Squared Dist.</th>
<th>KW</th>
</tr>
</thead>
<tbody>
<tr>
<td>A_KSAT</td>
<td>3885.0</td>
<td>0.0</td>
</tr>
<tr>
<td>A_PAWC</td>
<td>2770.2</td>
<td>0.0</td>
</tr>
<tr>
<td>A_THICK</td>
<td>3269.4</td>
<td>0.0</td>
</tr>
<tr>
<td>ARIDITY_INDEX</td>
<td>4486.7</td>
<td>10.1</td>
</tr>
<tr>
<td>B_KSAT</td>
<td>2746.8</td>
<td>0.1</td>
</tr>
<tr>
<td>B_PAWC</td>
<td>3277.6</td>
<td>0.3</td>
</tr>
<tr>
<td>B_THICK</td>
<td>3639.9</td>
<td>0.1</td>
</tr>
<tr>
<td>BIO04</td>
<td>3235.3</td>
<td>0.7</td>
</tr>
<tr>
<td>BIO08</td>
<td>3268.8</td>
<td>5.0</td>
</tr>
<tr>
<td>BIO09</td>
<td>3907.6</td>
<td>6.8</td>
</tr>
<tr>
<td>BIO15</td>
<td>1713.7</td>
<td>7.6</td>
</tr>
<tr>
<td>BIO16</td>
<td>4383.7</td>
<td>0.3</td>
</tr>
<tr>
<td>BIO17</td>
<td>4226.6</td>
<td>0.7</td>
</tr>
<tr>
<td>ELEVATION</td>
<td>3687.4</td>
<td>12.9</td>
</tr>
<tr>
<td>ET_ANNUAL</td>
<td>4204.7</td>
<td>0.1</td>
</tr>
<tr>
<td>GW_SWL</td>
<td>1933.6</td>
<td>0.0</td>
</tr>
<tr>
<td>GW_TDS</td>
<td>3644.2</td>
<td>0.0</td>
</tr>
<tr>
<td>LDI</td>
<td>1240.2</td>
<td>1.8</td>
</tr>
<tr>
<td>MAX_TEMP</td>
<td>4014.4</td>
<td>13.7</td>
</tr>
<tr>
<td>MIN_TEMP</td>
<td>3640.2</td>
<td>16.1</td>
</tr>
<tr>
<td>RAIN_ANNUAL</td>
<td>4486.2</td>
<td>9.3</td>
</tr>
<tr>
<td>SLOPE_RAD</td>
<td>2870.5</td>
<td>1.6</td>
</tr>
<tr>
<td>SOIL_EROS</td>
<td>4252.9</td>
<td>15.2</td>
</tr>
<tr>
<td>TWI</td>
<td>1990.9</td>
<td>1.5</td>
</tr>
<tr>
<td>WEATH_IND</td>
<td>2459.4</td>
<td>0.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ALOC 23 meta-group</th>
<th>A</th>
<th>B</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>I</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4.70</td>
<td>6.03</td>
<td>4.18</td>
<td>3.19</td>
<td>9.58</td>
<td>6.19</td>
<td>5.36</td>
<td>11.77</td>
</tr>
<tr>
<td>A</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
<td>0.1</td>
<td>0.0</td>
<td>11.2</td>
</tr>
<tr>
<td>B</td>
<td>0.0</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>4.8</td>
<td>0.1</td>
<td>12.6</td>
<td>10.2</td>
</tr>
<tr>
<td>D</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.1</td>
<td>4.5</td>
<td>0.4</td>
<td>0.0</td>
<td>8.0</td>
</tr>
<tr>
<td>E</td>
<td>0.1</td>
<td>0.0</td>
<td>0.1</td>
<td>0.0</td>
<td>46.0</td>
<td>0.5</td>
<td>0.0</td>
<td>4.8</td>
</tr>
<tr>
<td>F</td>
<td>10.1</td>
<td>8.3</td>
<td>3.0</td>
<td>8.5</td>
<td>0.3</td>
<td>3.0</td>
<td>0.4</td>
<td>13.3</td>
</tr>
<tr>
<td>G</td>
<td>19.1</td>
<td>1.1</td>
<td>0.5</td>
<td>0.1</td>
<td>1.1</td>
<td>0.5</td>
<td>0.0</td>
<td>28.7</td>
</tr>
<tr>
<td>I</td>
<td>17.7</td>
<td>1.9</td>
<td>1.7</td>
<td>1.9</td>
<td>0.4</td>
<td>1.7</td>
<td>0.6</td>
<td>0.3</td>
</tr>
<tr>
<td>K</td>
<td>6.5</td>
<td>0.1</td>
<td>0.0</td>
<td>28.7</td>
<td>0.3</td>
<td>0.0</td>
<td>0.0</td>
<td>1.7</td>
</tr>
</tbody>
</table>

289
Table 9.3: SIMPER results from PRIMER 6. Numbers represent percent contribution of each of the variables to the ALOC 20 meta-groups on a standardised Euclidean distance matrix. Meta-groups A, C, F, H and I only contained 1 ALOC cluster and therefore percent contribution could not be calculated using SIMPER, so these were omitted from the table. KW = Kruskal-Wallis statistic, with higher values indicating a better ability of that variable to discriminate between clusters. All KW values were significant at $P < 0.001$.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Avg. Squared Dist.</th>
<th>KW</th>
<th>ALOC 20 meta-group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>B</td>
</tr>
<tr>
<td>A_KSAT</td>
<td>227.1</td>
<td>17.6</td>
<td>1.5</td>
</tr>
<tr>
<td>A_PAWC</td>
<td>198.7</td>
<td>40.9</td>
<td>3.9</td>
</tr>
<tr>
<td>A_THICK</td>
<td>157.0</td>
<td>35.4</td>
<td>8.3</td>
</tr>
<tr>
<td>ARIDITY_INDEX</td>
<td>351.9</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td>B_KSAT</td>
<td>176.2</td>
<td>1.8</td>
<td>0.1</td>
</tr>
<tr>
<td>B_PAWC</td>
<td>152.3</td>
<td>2.0</td>
<td>15.7</td>
</tr>
<tr>
<td>B_THICK</td>
<td>226.5</td>
<td>0.0</td>
<td>16.5</td>
</tr>
<tr>
<td>BIO04</td>
<td>209.6</td>
<td>0.1</td>
<td>14.9</td>
</tr>
<tr>
<td>BIO08</td>
<td>169.7</td>
<td>0.5</td>
<td>1.9</td>
</tr>
<tr>
<td>BIO09</td>
<td>301.1</td>
<td>0.1</td>
<td>5.0</td>
</tr>
<tr>
<td>BIO15</td>
<td>100.5</td>
<td>0.1</td>
<td>3.3</td>
</tr>
<tr>
<td>BIO16</td>
<td>335.3</td>
<td>0.0</td>
<td>1.3</td>
</tr>
<tr>
<td>BIO17</td>
<td>344.4</td>
<td>0.1</td>
<td>0.9</td>
</tr>
<tr>
<td>ELEVATION</td>
<td>192.5</td>
<td>0.1</td>
<td>1.9</td>
</tr>
<tr>
<td>ET_ANNUAL</td>
<td>318.3</td>
<td>0.2</td>
<td>3.6</td>
</tr>
<tr>
<td>GW_SWL</td>
<td>151.6</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>GW_TDS</td>
<td>260.7</td>
<td>0.0</td>
<td>0.9</td>
</tr>
<tr>
<td>LDI</td>
<td>86.0</td>
<td>0.3</td>
<td>0.5</td>
</tr>
<tr>
<td>MAX_TEMP</td>
<td>306.2</td>
<td>0.0</td>
<td>0.9</td>
</tr>
<tr>
<td>MIN_TEMP</td>
<td>206.0</td>
<td>0.0</td>
<td>0.2</td>
</tr>
<tr>
<td>RAIN_ANNUAL</td>
<td>349.9</td>
<td>0.0</td>
<td>0.5</td>
</tr>
<tr>
<td>SLOPE_RAD</td>
<td>215.2</td>
<td>0.1</td>
<td>3.9</td>
</tr>
<tr>
<td>SOIL_EROS</td>
<td>342.7</td>
<td>0.1</td>
<td>2.5</td>
</tr>
<tr>
<td>TWI</td>
<td>148.8</td>
<td>0.0</td>
<td>4.8</td>
</tr>
<tr>
<td>WEATH_IND</td>
<td>174.5</td>
<td>0.5</td>
<td>6.0</td>
</tr>
</tbody>
</table>
Table 9.4: Total accuracies and kappa (\(\kappa\)) statistics for the 4 RF-classified models, and producer and user accuracies for each of the classes defined by the ALOC algorithm as classified by RF. N/A indicates groups that were missing from the classified dataset as a result of exclusion from the samples used to train the RF model. In some cases, groups were absent from the 80% training data, while others were excluded by the bootstrap aggregation step used to train the RF models.

<table>
<thead>
<tr>
<th>Model</th>
<th>ALOC 23 (94.9%, (\kappa = 0.94))</th>
<th>ALOC 23 PCA (92.1%, (\kappa = 0.92))</th>
<th>ALOC 20 (46.1%, (\kappa = 0.42))</th>
<th>ALOC 20 PCA (47.4%, (\kappa = 0.44))</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALOC Class</td>
<td>Producer</td>
<td>User</td>
<td>Producer</td>
<td>User</td>
</tr>
<tr>
<td>ALOC 01</td>
<td>99.3</td>
<td>95.2</td>
<td>99.3</td>
<td>91.2</td>
</tr>
<tr>
<td>ALOC 02</td>
<td>96.2</td>
<td>92.1</td>
<td>93.6</td>
<td>92.5</td>
</tr>
<tr>
<td>ALOC 03</td>
<td>96.6</td>
<td>83.6</td>
<td>89.7</td>
<td>83.9</td>
</tr>
<tr>
<td>ALOC 04</td>
<td>95.9</td>
<td>98.6</td>
<td>95.9</td>
<td>100.0</td>
</tr>
<tr>
<td>ALOC 05</td>
<td>92.7</td>
<td>92.7</td>
<td>98.9</td>
<td>90.9</td>
</tr>
<tr>
<td>ALOC 06</td>
<td>99.2</td>
<td>99.6</td>
<td>99.6</td>
<td>100.0</td>
</tr>
<tr>
<td>ALOC 07</td>
<td>100.0</td>
<td>100.0</td>
<td>91.1</td>
<td>100.0</td>
</tr>
<tr>
<td>ALOC 08</td>
<td>88.2</td>
<td>93.8</td>
<td>58.8</td>
<td>83.3</td>
</tr>
<tr>
<td>ALOC 09</td>
<td>83.3</td>
<td>96.8</td>
<td>72.2</td>
<td>89.7</td>
</tr>
<tr>
<td>ALOC 10</td>
<td>94.1</td>
<td>100.0</td>
<td>97.7</td>
<td>93.0</td>
</tr>
<tr>
<td>ALOC 11</td>
<td>91.7</td>
<td>94.8</td>
<td>90.0</td>
<td>88.5</td>
</tr>
<tr>
<td>ALOC 12</td>
<td>100.0</td>
<td>96.1</td>
<td>100.0</td>
<td>94.2</td>
</tr>
<tr>
<td>ALOC 13</td>
<td>96.9</td>
<td>98.9</td>
<td>96.6</td>
<td>93.6</td>
</tr>
<tr>
<td>ALOC 14</td>
<td>81.1</td>
<td>89.6</td>
<td>91.9</td>
<td>86.1</td>
</tr>
<tr>
<td>ALOC 15</td>
<td>93.7</td>
<td>92.2</td>
<td>98.5</td>
<td>85.1</td>
</tr>
<tr>
<td>ALOC 16</td>
<td>91.0</td>
<td>93.6</td>
<td>87.6</td>
<td>90.1</td>
</tr>
<tr>
<td>ALOC 17</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>ALOC 18</td>
<td>95.2</td>
<td>95.2</td>
<td>85.7</td>
<td>88.5</td>
</tr>
<tr>
<td>ALOC 19</td>
<td>85.3</td>
<td>96.7</td>
<td>73.5</td>
<td>92.6</td>
</tr>
<tr>
<td>ALOC 20</td>
<td>87.5</td>
<td>89.5</td>
<td>81.6</td>
<td>89.5</td>
</tr>
<tr>
<td>Model</td>
<td>ALOC 23 (94.9%, κ = 0.94)</td>
<td>ALOC 23 PCA (92.1%, κ = 0.92)</td>
<td>ALOC 20 (46.1%, κ = 0.42)</td>
<td>ALOC 20 PCA (47.4%, κ = 0.44)</td>
</tr>
<tr>
<td>-------</td>
<td>--------------------------</td>
<td>-----------------------------</td>
<td>-----------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>ALOC 21</td>
<td>98.8</td>
<td>96.6</td>
<td>95.4</td>
<td>95.4</td>
</tr>
<tr>
<td>ALOC 22</td>
<td>96.3</td>
<td>94.1</td>
<td>89.0</td>
<td>90.1</td>
</tr>
<tr>
<td>ALOC 23</td>
<td>100.0</td>
<td>96.0</td>
<td>91.7</td>
<td>81.5</td>
</tr>
</tbody>
</table>
9.2 R code for permutation test

```r
# load library and set random seed
if(!is.element('psych', installed.packages()[1])) {
  install.packages('psych')
} else {print("psych" library already installed", quote=FALSE)}
library(psych)
set.seed(1)

# KAPPA STATISTICS PERMUTATION TEST
## test statistics = difference between two kappa values

# This function uses a permutation test to assess if two kappa test results are
# significantly different.
perm_test <- function(input_data_frame, validation, classifier1, classifier2, iterations) {

  ## format input_data
  scores <- input_data_frame[, c(validation, classifier1, classifier2)]

  ## Classification 1 kappa
  classifier1.kappa <- cohen.kappa(scores[, c(validation, classifier1)], alpha=.05)
  classifier1.kappa

  ## Classification 2 kappa
  classifier2.kappa <- cohen.kappa(scores[, c(validation, classifier2)], alpha=.05)
  classifier2.kappa

  observed.kappa.difference <- round(classifier1.kappa$kappa, 3) -
                              round(classifier2.kappa$kappa, 3)
  observed.kappa.difference

  ## Permutation test
  N.perm <- iterations

  perm.kappa.differences <- c(rep(NA, length = N.perm), observed.kappa.difference)

  ## Under H0 that the resampling of the 30m classification to 2.5km is identical
to the original 30m classification, we can exchange classes between the two.

  for (i in 1:N.perm) {
    perm.scores <- scores
    perm.scores[, c(classifier1, classifier2)] <-
    t(apply(perm.scores[, c(classifier1, classifier2)], 1,
        function(x) x[sample(1:2)]) )
    perm.kappa.differences[i] <-
    round(cohen.kappa(perm.scores[, c(validation, classifier1)],
                      alpha=.05)$kappa, 3) -
    round(cohen.kappa(perm.scores[, c(validation, classifier2)],
                      alpha=.05)$kappa, 3)
  }

  p.value <-
  mean(perm.kappa.differences <= (-1.0 * abs(observed.kappa.difference)) |
       perm.kappa.differences >= abs(observed.kappa.difference))

  results <- p.value
  attr(results, 'perm.kappa.differences') <- perm.kappa.differences
  attr(results, 'observed.kappa.difference') <- observed.kappa.difference

  return(results)
}
```

293
# Load data file
## data columns must be set-out: PointID, Validation, Classifier1, Classifier2.
data <- read.csv(file.choose(), header=TRUE)
names <- attr(data,'names')

# ALWAYS assume the following
## names[1] is the point ID column name
## names[2] is the validation column name
## names[3:length(names)] is the classifiers column names
validation <- names[2]
classifiers <- names[3:length(names)]

# set the number of iterations for permutation. Recommend 9999.
N.perm <- 9999
print(paste("Number of iterations: ",N.perm+1), quote=FALSE)

# create empty matrices for test statistics
p.values <- matrix(rep(1,(length(classifiers)^2)), nrow=length(classifiers), ncol=length(classifiers))
observed.kappa.differences <- matrix(rep(0,(length(classifiers)^2)), nrow=length(classifiers), ncol=length(classifiers))
perm.kappa.differences <- array(rep(0,(length(classifiers)^2)*length(data[,1])), c(length(classifiers), length(classifiers),(N.perm+1)))

# loop for the perm test
row_count <- 1
for (classifier1 in classifiers[1:(length(classifiers)-1)]) {
  col_count <- row_count+1
  for (classifier2 in classifiers[(row_count+1):length(classifiers)]) {
    ## perform test
    results <- perm_test(data,validation,classifier1,classifier2,N.perm)
    ## record p-values
    p.values[row_count,col_count] <- results
    ## record observed kappa differences
    attr(results,'observed.kappa.difference')
    observed.kappa.differences[row_count,col_count] <- attr(results,'observed.kappa.difference')
    ## record perm kappa differences
    perm.kappa.differences[row_count,col_count,] <- attr(results,'perm.kappa.differences')
    perm.kappa.differences[col_count,row_count,] <- attr(results,'perm.kappa.differences')
    col_count <- col_count + 1
  }
  row_count <- row_count+1
}
print(paste("Observed Kappa Difference = ",round( observed.kappa.differences[1,2],3),"(after",N.perm+1,"permutations) "), quote=FALSE)

if (p.values[1,2] <= 0.05)
  print("Observed Kappa difference is significant at .05", quote = FALSE)
  print(paste("p ", p.values[1,2]), quote=FALSE)
if (p.values[1,2] > 0.05)
  print("Observed Kappa difference is insignificant at .05", quote = FALSE)
  print(paste("p ", p.values[1,2]), quote=FALSE)
9.3 Supporting figures
Figure 9.1: MDS analysis plots for the ALOC 23 and ALOC 20 models. Top row shows the ALOC 20 and ALOC 23 groups, while the bottom row shows the ALOC 20 and ALOC 23 meta-group plots.
Figure 9.2: ALOC 23 and ALOC 20 dendrogram demonstrating the hierarchical relationships between the non-hierarchical groups as defined from the ALOC group averages using SIMPROF. Red dotted lines indicate no significant differences among groups, while solid black lines indicate statistically-significant differences. The vertical axes represent the percentage similarity between groups. Letters in green represent the meta-groups each combination of non-hierarchical groups belongs to.
Figure 9.3: BioClim variable distributions across each of the ALOC 23 meta-groups.
Figure 9.4: Groundwater variable distributions across each of the ALOC 23 meta-groups. Note that observations >30,000 have been removed from GW_TDS for plotting purposes.
Figure 9.5: Landscape variable distributions across each of the ALOC 23 meta-groups.
Figure 9.6: Soil variable distributions across each of the ALOC 23 meta-groups.
Figure 9.7: Climate variable distributions across each of the ALOC 23 meta-groups.
Figure 9.8: Variable contribution to each of the hierarchical meta-groups calculated using SIMPER on a standardised Euclidean distance matrix. Any variables contributing <5% to the variance were pooled together and are represented by ALL_OTHER_VARS. Missing groups contained only one ALOC cluster and therefore percent variable contribution could not be calculated with SIMPER.
Figure 9.9: The correlation between the number of samples from each class and the number of samples correctly classified by CAP was highly significant, with the linear relationship among the variables for each sample illustrated in the figure. However, as the relationship was linear there was no clear threshold suggesting a minimum number of gauges needed to guarantee an acceptable level of accuracy. The CAP on the original dataset \((n = 201)\) was quite poor (classification accuracy = 48\%, \(m = 30, P = 0.001\)), while the bootstrapped dataset \((n = 383)\) was a significant improvement (classification accuracy = 66\%, \(m = 28, P = 0.001\)). CAP on the bootstrap dataset with a 20\% validation sample also performed reasonably (classification accuracy = 67\%, \(m = 32, P = 0.001\)).
Appendix II

10.1 Supporting tables
Table 10.1: Description of the 33 streamflow indices used to discriminate between regionalisation groups, and for generating clusters in the flow-based classification. For further details see Olden and Poff (2003). ‘-’ indicates dimensionless variables. Note that some flow variables are divided by mean flows, giving a dimensionless variable (e.g. DH13).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Unit</th>
<th>Time Step</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>DH13</td>
<td>Mean 30-day Maximum Discharge</td>
<td>-</td>
<td>Daily</td>
<td>Mean 30 day maximum flow divided by median monthly flow</td>
</tr>
<tr>
<td>DH15</td>
<td>High Flow Pulse Duration</td>
<td>days</td>
<td>Annual</td>
<td>Mean duration of high flow pulses exceeding 75th percentile</td>
</tr>
<tr>
<td>DH16</td>
<td>Variability in High Flow Pulse Duration</td>
<td>-</td>
<td>Annual</td>
<td>Coefficient of variation in DH15</td>
</tr>
<tr>
<td>DH20</td>
<td>High Flow Duration</td>
<td>days</td>
<td>Annual</td>
<td>Mean duration of flows exceeding 25th percentile</td>
</tr>
<tr>
<td>DL13</td>
<td>Mean 30-day Minimum Discharge</td>
<td>-</td>
<td>Daily</td>
<td>Mean 30 day minimum flow divided by median monthly flow</td>
</tr>
<tr>
<td>DL16</td>
<td>Low Flow Pulse Duration</td>
<td>days</td>
<td>Annual</td>
<td>Mean duration of low flow pulses below 25th percentile</td>
</tr>
<tr>
<td>DL17</td>
<td>Variability in Low Flow Pulse Duration</td>
<td>-</td>
<td>Annual</td>
<td>Coefficient of variation in DL16</td>
</tr>
<tr>
<td>DL18</td>
<td>Number of Zero Flow days</td>
<td>year⁻¹</td>
<td>Annual</td>
<td>Mean number of days where streams cease to flow</td>
</tr>
<tr>
<td>FH2</td>
<td>Variability in High Flood Pulse Count</td>
<td>-</td>
<td>Annual</td>
<td>Coefficient of variation in mean number of pulses exceeding 75th percentile</td>
</tr>
<tr>
<td>FH3</td>
<td>High Flood Pulse Count 2</td>
<td>year⁻¹</td>
<td>Annual</td>
<td>Number of high flow pulses, where a pulse is equal to 3x mean daily flow</td>
</tr>
<tr>
<td>FH6</td>
<td>Flood Frequency 1 (3xMDF)</td>
<td>year⁻¹</td>
<td>Annual</td>
<td>Mean number of high flow events using a threshold of 3x mean daily flow</td>
</tr>
<tr>
<td>FH7</td>
<td>Flood Frequency 2 (7xMDF)</td>
<td>year⁻¹</td>
<td>Annual</td>
<td>Mean number of high flow events using a threshold of 7x mean daily flow</td>
</tr>
<tr>
<td>FL1</td>
<td>Low Flood Pulse Count</td>
<td>year⁻¹</td>
<td>Annual</td>
<td>Number of annual occurrences where flows drop below 25th percentile</td>
</tr>
<tr>
<td>FL2</td>
<td>Variability in Low Flood Pulse Count</td>
<td>-</td>
<td>Annual</td>
<td>Coefficient of variation in FL1</td>
</tr>
<tr>
<td>FL3</td>
<td>Frequency of Low Flow Spells</td>
<td>year⁻¹</td>
<td>Annual</td>
<td>Total number of low flow spells (threshold equal to 5% of mean daily flow) divided by the record length in years</td>
</tr>
<tr>
<td>MA11</td>
<td>Spread in Daily Flows</td>
<td>1/m³·s⁻¹</td>
<td>Daily</td>
<td>Ratio of 25th/75th percentile divided by median daily flow</td>
</tr>
<tr>
<td>MA3</td>
<td>Variability in Daily Flows</td>
<td>-</td>
<td>Daily</td>
<td>Coefficient of variation in daily flows</td>
</tr>
<tr>
<td>MA41</td>
<td>Mean Annual Runoff</td>
<td>m³·s⁻¹·km⁻²</td>
<td>Annual</td>
<td>Mean annual flow divided by catchment area</td>
</tr>
<tr>
<td>MA5</td>
<td>Skewness in Daily Flows</td>
<td>-</td>
<td>Daily</td>
<td>Mean daily flows divided by median daily flows</td>
</tr>
<tr>
<td>Variable</td>
<td>Description</td>
<td>Unit</td>
<td>Time Step</td>
<td>Definition</td>
</tr>
<tr>
<td>------------</td>
<td>--------------------------------------</td>
<td>------</td>
<td>-----------</td>
<td>----------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>MH10</td>
<td>Mean Maximum Monthly Flow (October)</td>
<td>m$^3$/s</td>
<td>Monthly</td>
<td>Mean of the maximum monthly flows for October across all years of record</td>
</tr>
<tr>
<td>MH14</td>
<td>Median of Annual Maximum Flows</td>
<td>-</td>
<td>Annual</td>
<td>Median of the highest annual daily flow divided by the median annual daily flow averaged across all years</td>
</tr>
<tr>
<td>MH16</td>
<td>High Flow Discharge</td>
<td>-</td>
<td>Annual</td>
<td>Mean of the 10th percentile from the flow duration curve divided by median daily flow averaged across all years</td>
</tr>
<tr>
<td>MH8</td>
<td>Mean Maximum Monthly Flow (August)</td>
<td>m$^3$/s</td>
<td>Monthly</td>
<td>Mean of the maximum monthly flows for August across all years of record</td>
</tr>
<tr>
<td>ML17</td>
<td>Baseflow Index</td>
<td>-</td>
<td>Annual</td>
<td>Seven-day minimum flow divided by mean annual daily flows averaged across all years</td>
</tr>
<tr>
<td>ML18</td>
<td>CV Baseflow Index</td>
<td>-</td>
<td>Annual</td>
<td>Coefficient of variation in ML17</td>
</tr>
<tr>
<td>ML21</td>
<td>Variability across Annual Minimum Flows</td>
<td>-</td>
<td>Annual</td>
<td>Coefficient of variation in annual minimum flows averaged across all years</td>
</tr>
<tr>
<td>ML4</td>
<td>Mean Minimum Monthly Flow (April)</td>
<td>m$^3$/s</td>
<td>Monthly</td>
<td>Mean of the minimum monthly flows for April across all years of record</td>
</tr>
<tr>
<td>RA5</td>
<td>Number of Rises</td>
<td>-</td>
<td>Daily</td>
<td>Ratio of days where the flow is higher than the previous day</td>
</tr>
<tr>
<td>RA8</td>
<td>Number of Reversals</td>
<td>-</td>
<td>Daily</td>
<td>Number of negative and positive changes flows from one day to the next</td>
</tr>
<tr>
<td>RA9</td>
<td>Variability in Number of Reversals</td>
<td>-</td>
<td>Daily</td>
<td>Coefficient of variation in RA8</td>
</tr>
<tr>
<td>TA1</td>
<td>Constancy</td>
<td>-</td>
<td>Daily</td>
<td>Colwell's (1974) constancy of mean daily flow</td>
</tr>
<tr>
<td>TL1</td>
<td>Julian Date of Annual Minimum</td>
<td>-</td>
<td>Daily</td>
<td>The mean Julian date of the 1-day annual minimum flow over all years of record</td>
</tr>
<tr>
<td>TL2</td>
<td>Variability in Julian Date of Annual Minimum</td>
<td>-</td>
<td>Daily</td>
<td>Coefficient of variation in TL1</td>
</tr>
</tbody>
</table>
10.2 Supporting figures
Figure 10.1: Mean random forest (RF) model accuracy across $n = 1000$ bootstrap datasets. Model performance did not increase much after 500 trees. The RF model that classified the flow classification groups using only environmental variables consistently outperformed the regionalisation RF model that attempted to classify regionalisation groups based on streamflow indices.
Figure 10.2: Results of LISA residual analysis showing the arrangement of spatial clusters (positive spatial autocorrelation) and spatial outliers (negative spatial autocorrelation). The very limited amount of local spatial autocorrelation (positive and negative) suggests that the methods utilised for the analysis and classification have been able to account for spatial variability. HH = High-High residual spatial clustering; LL = Low-Low residual spatial clustering; HL = High-Low residual spatial outliers; LH = Low-High residual spatial outliers.