The Australian Water Resources Assessment System
Technical Report 4. Landscape Model (version 0.5) Evaluation Against Observations
AIJM Van Dijk and G Warren

17 June 2010

A water information R & D alliance between the Bureau of Meteorology
and CSIRO’s Water for a Healthy Country Flagship
Australia is founding its future on science and innovation. Its national science agency, CSIRO, is a powerhouse of ideas, technologies and skills. CSIRO initiated the National Research Flagships to address Australia’s major research challenges and opportunities. They apply large scale, long term, multidisciplinary science and aim for widespread adoption of solutions. The Flagship Collaboration Fund supports the best and brightest researchers to address these complex challenges through partnerships between CSIRO, universities, research agencies and industry.

The Water for a Healthy Country Flagship aims to provide Australia with solutions for water resource management, creating economic gains of $3 billion per annum by 2030, while protecting or restoring our major water ecosystems.

The work contained in this report is collaboration between CSIRO’s Water for a Healthy Country Flagship and the Bureau of Meteorology’s Water Division under the Water Information Research and Development Alliance (WIRADA).

WIRADA is a five-year, $50 million partnership that will use CSIRO’s R&D expertise to help the Bureau of Meteorology report on availability, condition and use of Australia's water resources.

For more information about the partnership visit www.csiro.au/partnerships/WIRADA.html
For more information about Water for a Healthy Country Flagship or the National Research Flagship Initiative visit www.csiro.au/org/HealthyCountry.html


Copyright and Disclaimer
© 2010 Bureau of Meteorology and CSIRO To the extent permitted by law, all rights are reserved and no part of this publication covered by copyright may be reproduced or copied in any form or by any means except with the written permission of Bureau of Meteorology and CSIRO.

Important Disclaimer:
Bureau of Meteorology and CSIRO advise that the information contained in this publication comprises general statements based on scientific research. The reader is advised and needs to be aware that such information may be incomplete or unable to be used in any specific situation. No reliance or actions must therefore be made on that information without seeking prior expert professional, scientific and technical advice. To the extent permitted by law, Bureau of Meteorology and CSIRO (including its employees and consultants) exclude all liability to any person for any consequences, including but not limited to all losses, damages, costs, expenses and any other compensation, arising directly or indirectly from using this publication (in part or in whole) and any information or material contained in it.

Cover Photo: © 1989 CSIRO. Photographer: Willem van Aken. Description: Flooded Murray River at Mildura, VIC. From Science Imagine Online (www.scienceimage.csiro.au), file BU5881.jpg
ABOUT THIS REPORT

Through the Water Information Research and Development Alliance, the Bureau of Meteorology and CSIRO are developing the Australian water resources assessment system (AWRA). The AWRA system is integrated software that combines hydrological models with a variety of observations and observation-derived products.

The purpose of the AWRA system is to operationally provide up to date, credible, comprehensive, accurate and relevant information about the history, present state and future trajectory of the water balance, with sufficient detail to inform water resources management. Intended dissemination of the information is through the Bureau’s water information services, in particular, occasional and scheduled water resources assessments and the annual National Water Account.

The technical details and evaluation of the AWRA system are documented in a series of reports that are updated as new system or component versions are developed. Reports in this series are (the current report is highlighted):

- 1. System Conceptual Design
- 2. Implementation Document
- 3. Landscape Model (version 0.5) Technical Description
- 4. Landscape Model (version 0.5) Evaluation Against Observations
EXECUTIVE SUMMARY

This report is one of a series on the Australian Water Resources Assessment system (AWRA), and evaluates simulations by the current model version 0.5 against various observation types. This was done to: (1) set a benchmark against which future modifications can be tested; (2) understand how and against which observations model estimates are most usefully compared; (3) identify processes or quantities that are not described well by the model; (4) inform the development of model-data assimilation techniques; and (5) allow the model results to be used with appropriate caveats in water balance assessment and water accounting.

The same model parameterisation was used in all cases. An important caveat is that the current model version uses the assumption that local precipitation is the only source of water. Therefore simulations are likely to be of poor quality for areas receiving important lateral influxes of river and groundwater; for example, floodplains and irrigation areas. A second caveat is that the model received very minimal parameter calibration and was used without data assimilation or output bias correction. Model-data fusion techniques will without a doubt much improve the agreement between model estimates and observations, but would have pre-empted a meaningful assessment of model performance.

Four types of observations were used in the current evaluation. Additional on-ground and satellite observation data sets were identified and will be considered for future updates of this document. Conclusions with regards to the four observation types are summarised below.

**Catchment streamflow** records from 362 small catchments minimally affected by regulation were reproduced well, despite the lack of parameter calibration. Calculated performance statistics were equivalent or better than those obtained with other rainfall-runoff models in comparable experiments. Important features were the lack of bias and the strong improvement in performance as data was aggregated for multiple catchments and over longer time scales. The model structure appeared robust and development of operational model-data fusion techniques are recommended.

**Flux tower ET observations** at four sites across Australia were made available by the principal investigators. The model reproduced observations well, and the relative and absolute agreement increased as daily data were aggregated to monthly values. The quality of model rainfall forcing was the main source of error, whereas rainfall interception losses were an important source of uncertainty in the observations. Aside from these two issues, there appears to be only modest scope to improve model performance.

**Satellite-derived top soil water content** (SWC) estimates from the ASAR GM radar satellite instrument were made available by the Technical University of Vienna with support from the European Space Agency. The modelled SWC estimates agreed well with satellite top soil moisture content, although known sources of error in observations meant that in regions with dense vegetation and strong relief estimates could not be evaluated.

**Satellite-derived vegetation properties** describing vegetation canopy cover fraction, density and chlorophyll content were derived from the AVHRR and MODIS optical satellite instruments. Simulation of these vegetation properties is not an model objective by itself, but accurate representation of vegetation dynamics can be a prerequisite to accurate estimation of ET fluxes. The AWRA model reproduced the response of grasses, herbs and crops to temporal patterns in water availability well. The effects of temperature limitation at high elevations and the generally small variations in forest canopy properties were not reproduced. Where the model did not perform well, the implications for water balance estimation may still be modest. A possible exception is the estimation of rainfall interception losses in high elevation areas. It is recommended to investigate opportunities to improve rainfall estimates and/or quantify lateral water influxes by assimilation of satellite derived vegetation properties into the model.

Overall model performance was considered to be very satisfactory when compared to previously published model evaluations. Despite opportunities for improvement, the model is recommended as suitable for the production of regular water resource assessments and water accounts.
# CONTENTS

About this report ..........................................................................................................iii

Executive Summary .....................................................................................................iv

Contents.........................................................................................................................v

Acknowledgements ...................................................................................................viii

List of Acronyms and Symbols ..................................................................................ix

1. Introduction ..........................................................................................................1
   1.1 Terms used and avoided ..................................................................................1
       1.1.1 What is being evaluated? .................................................................1
       1.1.2 Model estimates versus predictions or forecasts .........................2
       1.1.3 Difference versus accuracy, precision, and error .......................2
       1.1.4 Evaluation versus validation ........................................................3
       1.1.5 Evaluation versus uncertainty assessment ................................3
   1.2 Level of independence of this evaluation ......................................................3

2. Evaluation metrics ...............................................................................................5
   2.1 Statistics of the distribution ..........................................................................5
   2.2 Absolute agreement .......................................................................................5
   2.3 Relative agreement .........................................................................................6

3. Data used ..............................................................................................................8
   3.1 Modelled data ...............................................................................................8
       3.1.1 Runs produced .................................................................................8
       3.1.2 Parameterisation ..............................................................................8
   3.2 Observations .................................................................................................10

4. Gauged streamflow ...........................................................................................11
   Summary 11
   4.1 Introduction .................................................................................................11
       4.1.1 Justification .....................................................................................11
       4.1.2 Observations ..................................................................................12
       4.1.3 Caveats ..........................................................................................12
   4.2 Method ..........................................................................................................13
       4.2.1 Model run .......................................................................................13
       4.2.2 Data aggregation ............................................................................14
       4.2.3 Calculated metrics and interpretation ..........................................14
       4.2.4 Comparison with performance of Zhang model .......................15
   4.3 Results ............................................................................................................15
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.3.1</td>
<td>Mean streamflow</td>
<td>15</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Daily streamflow patterns and flow duration curves</td>
<td>18</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Effect of temporal aggregation</td>
<td>20</td>
</tr>
<tr>
<td>4.3.4</td>
<td>Effect of spatial aggregation</td>
<td>22</td>
</tr>
<tr>
<td>4.4</td>
<td>Discussion</td>
<td>26</td>
</tr>
<tr>
<td>4.4.1</td>
<td>Mean streamflow</td>
<td>26</td>
</tr>
<tr>
<td>4.4.2</td>
<td>Daily streamflow patterns and flow duration curves</td>
<td>26</td>
</tr>
<tr>
<td>4.4.3</td>
<td>Effect of temporal aggregation</td>
<td>28</td>
</tr>
<tr>
<td>4.4.4</td>
<td>Effect of spatial aggregation</td>
<td>29</td>
</tr>
<tr>
<td>4.5</td>
<td>Conclusions</td>
<td>29</td>
</tr>
<tr>
<td>4.6</td>
<td>Recommendations</td>
<td>30</td>
</tr>
<tr>
<td>5.1</td>
<td>Introduction</td>
<td>31</td>
</tr>
<tr>
<td>5.1.1</td>
<td>Justification</td>
<td>31</td>
</tr>
<tr>
<td>5.1.2</td>
<td>Caveats</td>
<td>32</td>
</tr>
<tr>
<td>5.1.3</td>
<td>Literature review</td>
<td>33</td>
</tr>
<tr>
<td>5.2</td>
<td>Descriptions of sites</td>
<td>34</td>
</tr>
<tr>
<td>5.3</td>
<td>Method</td>
<td>37</td>
</tr>
<tr>
<td>5.3.1</td>
<td>Processing of observed data</td>
<td>37</td>
</tr>
<tr>
<td>5.3.2</td>
<td>Model run</td>
<td>38</td>
</tr>
<tr>
<td>5.3.3</td>
<td>Comparison</td>
<td>38</td>
</tr>
<tr>
<td>5.4</td>
<td>Results</td>
<td>39</td>
</tr>
<tr>
<td>5.4.1</td>
<td>Components of average ET</td>
<td>39</td>
</tr>
<tr>
<td>5.4.2</td>
<td>Daily flux patterns</td>
<td>40</td>
</tr>
<tr>
<td>5.4.3</td>
<td>Monthly flux patterns</td>
<td>42</td>
</tr>
<tr>
<td>5.4.4</td>
<td>Precipitation</td>
<td>44</td>
</tr>
<tr>
<td>5.4.5</td>
<td>Enhanced Vegetation Index</td>
<td>45</td>
</tr>
<tr>
<td>5.5</td>
<td>Discussion</td>
<td>46</td>
</tr>
<tr>
<td>5.5.1</td>
<td>Overall agreement</td>
<td>46</td>
</tr>
<tr>
<td>5.5.2</td>
<td>Sources of error</td>
<td>46</td>
</tr>
<tr>
<td>5.5.3</td>
<td>Influence of night-time ET</td>
<td>47</td>
</tr>
<tr>
<td>5.5.4</td>
<td>Influence of rainfall interception</td>
<td>47</td>
</tr>
<tr>
<td>5.6</td>
<td>Conclusions</td>
<td>47</td>
</tr>
<tr>
<td>5.7</td>
<td>Recommendations</td>
<td>48</td>
</tr>
<tr>
<td>6.1</td>
<td>Introduction</td>
<td>49</td>
</tr>
<tr>
<td>6.1.1</td>
<td>Justification</td>
<td>49</td>
</tr>
<tr>
<td>6.1.2</td>
<td>Caveats</td>
<td>50</td>
</tr>
<tr>
<td>6.1.1</td>
<td>Literature review</td>
<td>51</td>
</tr>
<tr>
<td>6.2</td>
<td>Methods</td>
<td>53</td>
</tr>
<tr>
<td>6.2.1</td>
<td>Observations used</td>
<td>53</td>
</tr>
</tbody>
</table>
6.2.2 Model data used ................................................................. 53
6.2.3 Comparison............................................................................ 54
6.3 Results............................................................................................ 54
6.3.1 Descriptive statistics .............................................................. 54
6.3.2 Bias ............................................................................................. 55
6.3.3 Relative agreement ................................................................. 56
6.3.4 Absolute agreement ................................................................. 57
6.4 Discussion ..................................................................................... 58
6.4.1 Overall agreement ................................................................. 58
6.4.2 Causes for differences .............................................................. 59
6.5 Conclusions ................................................................................... 60
6.6 Recommendations........................................................................ 60
7. AVHRR/MODIS vegetation properties ............................................. 62
   7.1 Introduction ............................................................................... 62
      7.1.1 Justification........................................................................... 62
      7.1.2 Caveats .................................................................................. 63
      7.1.3 Data sets used and caveats ....................................................... 65
      7.1.4 Summary................................................................................ 68
    7.2 Methods .................................................................................... 68
      7.2.1 Observations used ................................................................. 68
      7.2.2 Model data used ................................................................. 69
      7.2.3 Comparison........................................................................... 69
    7.3 Results............................................................................................ 71
      7.3.1 AVHRR FPAR................................................................. 71
      7.3.2 MODIS FPAR................................................................. 74
      7.3.3 MODIS fraction photosynthetic vegetation ................................ 77
      7.3.4 LAI ....................................................................................... 80
      7.3.5 EVI ....................................................................................... 83
  7.4 Discussion ..................................................................................... 86
      7.4.1 Overall comparison .............................................................. 86
      7.4.2 Regional comparison ............................................................. 86
      7.4.3 Implications for water balance estimation ................................ 88
    7.5 Conclusions ................................................................................ 89
    7.6 Recommendations........................................................................ 89

References .......................................................................................... 91

Appendix A. Model code ...................................................................... 96
The following individuals have contributed data or otherwise provided valuable assistance for the production of this report:

- Juan Pablo Guerschman, Randall Donohue, Tim MeVicar (CSIRO Land and Water, Canberra)
- Steve Marvanek (CSIRO Land and Water, Adelaide)
- Yi Liu (University of New South Wales, Sydney)
- Jorge Peña Arancibia (King’s College, London, UK)
- Ray Leuning, Eva Van Gorsel (CSIRO Marine and Atmospheric Research, Canberra)
- Tom Van Niel (CSIRO Land and Water, Perth)
- Marcela Doubkova, Wolfgang Wagner (Technical University Vienna, Austria)
- Jeff Walker, Robert Pipunic (University of Melbourne)
- Jason Beringer (Monash University, Melbourne)
- Lindsay Hutley (Charles Darwin University, Darwin)

This report has benefited greatly from discussions with and/or reviews from Andrew Frost and Sri Srikanthan (all Bureau of Meteorology, Water Division); Warrick Dawes, Francis Chiew and Juan Pablo Guerschman (all CSIRO Land and Water); and Ray Leuning (CSIRO Marine and Atmospheric Research).

These contributions are gratefully acknowledged.
## LIST OF ACRONYMS AND SYMBOLS

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AND</td>
<td>absolute normalised difference</td>
</tr>
<tr>
<td>ASAR GM</td>
<td>advanced synthetic aperture radar, operated in global monitoring mode</td>
</tr>
<tr>
<td>AVHRR</td>
<td>advanced very high resolution radiometer</td>
</tr>
<tr>
<td>AWRA</td>
<td>Australian Water Resources Assessment System</td>
</tr>
<tr>
<td>AWRC</td>
<td>Australian Water Resources Council</td>
</tr>
<tr>
<td>BAWAP</td>
<td>Bureau of Meteorology component of Australian Water Availability Project</td>
</tr>
<tr>
<td>CRV</td>
<td>correlated residual variance</td>
</tr>
<tr>
<td>CSIRO</td>
<td>Commonwealth Scientific and Industrial Research Organisation</td>
</tr>
<tr>
<td>ERS</td>
<td>European Remote Sensing satellite</td>
</tr>
<tr>
<td>ET</td>
<td>evapotranspiration</td>
</tr>
<tr>
<td>EVI</td>
<td>enhanced vegetation index</td>
</tr>
<tr>
<td>FDC</td>
<td>flow-duration curve</td>
</tr>
<tr>
<td>FPAR</td>
<td>absorbed fraction photosynthetically active radiation</td>
</tr>
<tr>
<td>FPV</td>
<td>fraction photosynthetic vegetation</td>
</tr>
<tr>
<td>HRU</td>
<td>hydrological response unit</td>
</tr>
<tr>
<td>LAI</td>
<td>leaf area index</td>
</tr>
<tr>
<td>MDA</td>
<td>model-data assimilation</td>
</tr>
<tr>
<td>MODIS</td>
<td>moderate resolution imaging spectrometer</td>
</tr>
<tr>
<td>MRE</td>
<td>mean relative difference</td>
</tr>
<tr>
<td>NASA</td>
<td>National Aeronautics and Space Administration (USA)</td>
</tr>
<tr>
<td>NIR</td>
<td>near infrared (radiation)</td>
</tr>
<tr>
<td>NDVI</td>
<td>normalised difference vegetation index</td>
</tr>
<tr>
<td>NOAA</td>
<td>National Oceanic and Atmospheric Administration (USA)</td>
</tr>
<tr>
<td>NSME</td>
<td>Nash-Sutcliffe model efficiency</td>
</tr>
<tr>
<td>NSW</td>
<td>New South Wales</td>
</tr>
<tr>
<td>NT</td>
<td>Northern Territory</td>
</tr>
<tr>
<td>OER</td>
<td>overestimation rate</td>
</tr>
<tr>
<td>PCI</td>
<td>photosynthetic capacity index</td>
</tr>
<tr>
<td>PET</td>
<td>potential evapotranspiration</td>
</tr>
<tr>
<td>SD</td>
<td>standard difference</td>
</tr>
<tr>
<td>SE</td>
<td>south-east</td>
</tr>
<tr>
<td>SNR</td>
<td>signal-to-noise ratio</td>
</tr>
<tr>
<td>SWC</td>
<td>soil water content</td>
</tr>
<tr>
<td>R</td>
<td>correlation coefficient</td>
</tr>
<tr>
<td>R^2</td>
<td>coefficient of determination</td>
</tr>
<tr>
<td>RR^2</td>
<td>ranked or non-parametric coefficient of determination</td>
</tr>
<tr>
<td>VI</td>
<td>vegetation index</td>
</tr>
<tr>
<td>WA</td>
<td>Western Australia</td>
</tr>
<tr>
<td>μ</td>
<td>mean</td>
</tr>
<tr>
<td>σ</td>
<td>standard deviation</td>
</tr>
</tbody>
</table>
1. INTRODUCTION

The aim of this report is to compare AWRA-L model estimates to observations, and to quantify and discuss the likely causes of the differences encountered. There are several reasons why such an exercise is important:

- To set a performance benchmark against which future modifications can be tested.
- To understand observations and comparison techniques that are or are not suitable for model evaluation.
- To identify processes or quantities that are not well described by the model, and the conditions under which this occurs.
- To inform future model-data assimilation experiments; a prerequisite for which is good understanding of differences between model and the assimilated observations.
- To allow the model results to be used with appropriate caveats and understanding of uncertainty.
- To support a more comprehensive uncertainty assessment.

It is important to emphasise up front that this report does not offer a formal validation or a reliable characterisation of all errors in the model. It is also important to emphasise that there has so far been limited effort in estimating model parameters values or fields, which means that further reduction in differences are feasible even without changing model structure.

Instead, this report offers an assessment of the degree of uncertainty in model estimates, as background information to support its use and further development. This is further explained below.

1.1 Terms used and avoided

Model evaluation and uncertainty assessment is an area of science that has many semantic pitfalls and therefore it is attempted below to provide clear working definitions. The main points of this section can be summarised as follows:

- Terms used in this report are: model estimate, difference, agreement, and evaluation.
- Terms generally avoided are: model prediction, model forecast, precision, accuracy, error, validation, and uncertainty assessment.
- The term ‘model’ here only refers to the modelled data produced by the specific AWRA model run used for this report.

The rationale for the use of these terms is given below.

1.1.1 What is being evaluated?

A model is a representation of reality. Models used in environmental prediction include conceptual, statistical, mathematical, physical and computer models, as well as, in an information technology context, data models.

The term model in this report refers to the AWRA computer model and its components, and the evaluation is limited to the estimates that it produces. The theories, concepts and data model...
embedded within it are not evaluated in this report, except to the extent that inferences can be made by comparing model estimates to observations.

Model verification usually refers to analyses intended to ascertain that the computer model does not contain numerical errors, such as code bugs, loss of mass balance, mislabelling of files or outputs, and other consistency tests. The AWRA model has been verified in this manner, but those tests are not described in this report.

The evaluation in this report would change if different parameter values or inputs were used. Therefore this evaluation is limited to the specific model run that was produced for the purposes of this report. That model run is described in Section 2.

1.1.2 Model estimates versus predictions or forecasts

The term model prediction and forecast can be used in a way that is synonymous to estimation and simulation. Applied retrospectively, with observed or estimated historic model boundary conditions such as climate records, model estimates are sometimes referred to as predictions or ‘hind-casts’. In this report, the terms prediction and forecast are avoided for two reasons: to many it might imply a prediction about the future rather than the past; to others, a prediction of historical conditions to formally validate a model. Neither has been attempted in this report.

1.1.3 Difference versus accuracy, precision, and error

In observational science and engineering, precision is usually defined as the reproducibility of the measurement – for example, the standard deviation in subsequent readings of the same quantity – whereas accuracy is the difference between the average reading and the real quantity. Defined this way, in modelling perfect precision is always achieved unless there is a random process in the calculations, which is not the case in AWRA. Therefore, the term precision is not used in this report.

The terms accuracy and error are avoided where a difference is found between model estimates and observations, because these terms imply that the reason for the difference is known. In principle, the precision and accuracy in an instrumental record may be known. If the difference with model estimates is several times greater than both these, most of that difference may confidently be attributed to model error. Ideally, the observations used are as much as possible the direct observations; for example, water level rather than streamflow; remotely sensed radiances rather than derived products. In that case an ‘observational model’ may need to be attached to the biophysical model to translate states and fluxes to the observed quantities. If the precision and accuracy of the instrument used are specified, this approach ensures that errors can be attributed to the model – including its observation model component – rather than to the observations.

However, this approach may not always be useful, and can be impossible or impractical. It may not be useful because the source of the error remains unknown (that is, the observational model or the biophysical model). It may be impossible or impractical because the original instrument readings no longer exist or are unavailable, or because the details of the method by which they were converted to biophysical quantities no longer exists or is available. The ‘observations’ used in this report are likely to have many errors of their own, for example because of instrument malfunctions or unremoved gap-filled data; because of errors in corrections, rating curves or retrieval models; or because the measurement is wrongly assumed to be representative at the resolution of modelling.
If follows that the attribution of differences between model and observations has a subjective element and hence the term ‘error’ is not generally used in this report, except where its source is unambiguous.

### 1.1.4 Evaluation versus validation

For instruments, validation implies that the instrument measures a quantity with or within estimated accuracy and precision. In science, model validation is achieved if the model accurately predicts observed phenomena. It has been argued that natural systems are too complex and uncertain, and the opportunities for controlled experiments too few, for a model ever to be valid in a strict sense (Oreskes et al. 1994).

In a more pragmatic interpretation the term validation is still widely used in model assessment however. Usually it describes whether a model sufficiently accurately describes observations to support a certain theory, interpretation or practical use. *Independent* validation does require that the observations against which the model is validated have in no way been used to formulate the model in the first place. Unfortunately, model applications without any use of observations in their formulation commonly lead to very poor predictions indeed and therefore fully independent validation tends to be an unsatisfactory exercise. To avoid this, very often some of the observations are withheld from the model, and the remainder used to constrain model behaviour, for example by using mathematical error minimisation (or optimisation) techniques to estimate (or calibrate) model parameter values. These approaches are usually described ‘calibration-validation’, ‘split sample’ or ‘cross validation’ techniques. While pragmatic solutions, they remove the process further away from a formal and independent validation.

The model predictions being evaluated in this report are also informed, to varying degrees, by many of the observation types used, as explained further below. Therefore the term ‘validation’ is avoided in this report in favour of ‘evaluation’.

The interpretation of the evaluation allows a judgement to be formed of whether model performance is acceptable or satisfactory to be used for a given purpose, rather than ‘valid’ per se. Such a judgement is likely to be highly purpose specific and will inevitably be subjective.

### 1.1.5 Evaluation versus uncertainty assessment

Uncertainty assessment is a vital exercise when the intended use of model results is associated with considerable risks. To be useful, the scope and emphasis of an uncertainty assessment requires the specific purposes of use and the associated risks to be identified (e.g. Van Dijk et al. 2008). Sources of uncertainty may include both known and, paradoxically, unknown errors in the model concepts, structure, parameter estimates, assumptions, and input data. This is well beyond the scope of the current evaluation.

Since the current model evaluation does not consider any specific application of the results, it should not be considered as an uncertainty assessment and hence the term is avoided. The results of this evaluation can provide valuable information for an uncertainty analysis however.

### 1.2 Level of independence of this evaluation

For reasons discussed above, a fully independent model evaluation tends to be impossible, unsatisfactory or undesirable. Most of the observations used for evaluation in this report where somehow used in model structure development or parameter estimation. This was not done in a way that makes their use for model evaluation trivial or useless, but it is important to
appropriately recognise in which respects they are and are not independent from the model simulations. Further detail is provided in the various sections, but a summary is as follows:

- Daily streamflow data was used to infer model structures and parameter estimation equations (Van Dijk 2009a; Van Dijk 2010). This produced a global calibration. Direct model parameter calibration against some or all catchment streamflow data was not performed however. Given the large number of catchments used in the global calibration (183 to 326 depending on the parameter) individual catchment records can to some degree be considered independent from the model estimates.

- Daily and monthly flux tower ET and MODIS EVI patterns for four sites were used to develop the structure of the vegetation hydrology and phenology component. Observations were not directly used for parameter calibration in the model run reported here, but prior estimates of some of the parameters have been informed by comparison in a more indirect manner in earlier publications (e.g. Van Dijk and Renzullo 2009).

- Continental MODIS albedo, EVI, and FPAR products as well as ASAR GM soil moisture product for four selected dates used to develop albedo equation and estimate conversion between canopy cover and vegetation photosynthetic index (Van Dijk and Warren 2010; Van Dijk et al. 2010).

- The AVHRR FPAR data (Donohue et al. 2007) were used to parameterise the fraction of deep-rooted vegetation, estimated as the minimum FPAR in the time series (Figure 1).

Further details on the way these observations were used are provided in the AWRA Technical Report 3. Their implications are considered in the discussion on differences for each section.

![Fraction](image.png)

Figure 1. Fraction of deep-rooted vegetation, estimated as minimum AVHRR FPAR for the period 1980-2006.
2. EVALUATION METRICS

A fixed set of statistical metrics was used in this evaluation and is defined below. Metrics chosen represent the statistics of the respective distributions; absolute differences between model and observations; and relative agreement in spatial or temporal patterns or both. Each of these aspects contains information that can be useful for the interpretation of differences encountered, but not all statistics were considered equally useful in all cases.

2.1 Statistics of the distribution

The statistics of the distribution of observed and estimated variables include the mean, standard deviation and median. These are defined as follows for model estimates (and analogous for the observations). For the sake of completeness, the **mean** ($\mu_M$) and **standard deviation** ($\sigma_M$) are calculated as:

$$\mu_M = \frac{1}{n} \sum M$$  \hspace{1cm} [2-1]

and

$$\sigma_M = \sqrt{\text{var}(M)} = \sqrt{\frac{1}{n} \sum (M - \bar{M})^2}$$  \hspace{1cm} [2-2]

where $O$ is the observations matrix, $M$ the model estimate matrix, $n$ the number of elements, and $\text{var}(M)$ the variance of $M$.

The $x^{th}$ **percentile** is calculated by finding the value of $M$ that is exceeded by $x$ percent of all values in the matrix. The **median** is the 50% percentile. The same statistics listed can all also be calculated after aggregating the data to a greater time step or area, whether through averaging or summation. For example, standard deviation can be constructed for daily, monthly or annual flow.

A graphic use of percentiles is the **exceedence curve** which shows the full range of percentiles (0 to 100%) on the horizontal axis, and the corresponding percentile values on the y-axis. An instance of this is the **flow duration curve** (FDC) which, in the same way, shows the non-exceedence curve for a stream flow time series for a given location (see Section 3).

2.2 Absolute agreement

The **standard difference** (SD) is defined as the square root of the residual variance $\text{res}(O,M)$ or:

$$SD = \sqrt{\text{res}(O,M)} = \sqrt{\frac{1}{n} \sum (O - M)^2}$$  \hspace{1cm} [2-3]

The **absolute and relative bias** are calculated as the difference in means:

$$\text{AbsBias} = \mu_M - \mu_O$$  \hspace{1cm} [2-4]

$$\text{RelBias} = \frac{\mu_M - \mu_O}{\mu_O}$$  \hspace{1cm} [2-5]
The overestimation rate (OER) is expressed as the difference between the frequency with which the model produced higher estimates than the observations, compared to the frequency expected for an unbiased estimate.

\[
OER = n(M > O) - \frac{1}{2}n
\]  

[2-6]

It provides an estimate of bias that is not affected by the distribution of values\(^1\).

The Nash-Sutclifffe model efficiency (NSME) expresses the fraction of total variance in the observations (\(\text{var}(O)\)) that is reproduced by the model. It is defined as:

\[
NSME = 1 - \frac{\text{res}(O,M)^2}{\text{var}(O)^2} = 1 - \frac{\sum(O - M)^2}{\sum(O - \bar{O})^2}
\]  

[2-7]

An NSME value of unity indicates perfect agreement, whereas a value less than zero indicates that the average observed value is a better predictor than the model (that is, the residual variance is greater than the observed variance).

### 2.3 Relative agreement

The median relative difference (MRD) is defined as the median of the ‘absolute normalised difference’, defined as (AND).

\[
AND = \left| \frac{M - O}{O} \right|
\]  

[2-8]

The median is chosen because some data sets are strongly skewed, include zero values – leading to infinitely high AND – or have estimates and observations with different sign. The MRE provides an estimate of the relative agreement between model and observations that can be expected to be achieved for half of the data.

The coefficient of determination (\(R^2\)) is:

\[
R^2 = \left[ \frac{\text{cov}(M,O)}{\sigma_M \sigma_O} \right]^2
\]  

[2-9]

The ranked or non-parametric coefficient of determination (\(RR^2\)) indicates the degree to which the model and observations show the same relative patterns. It is calculated by first replacing the numerical values in both the observed and estimated matrix by their rank, and subsequently calculating the coefficient of determination for the result. Where \(RR^2\) is higher than \(R^2\), there is a possibility that part of the difference is structure and can be removed by a non-linear bias correction.

The correlated residual variation is calculated as the difference between \(R^2\) and NSME and indicates the fraction of the observed variance that is correlated to the model estimates but not explained by them. It describes systematic differences between observations and estimates, such as constant bias and proportional differences, as opposed to apparently random differences, which can be calculated from \(1-R^2\). CRV can be very useful in identifying opportunities to

---

\(^1\) This ignores the possibility that model estimates are exactly equal to observations. Given the number of decimals in the observations and model estimates used here there will be no difference.
improve model performance through parameter estimation or bias correction (e.g. see Van Dijk et al. 2008 for application to river models).
3. DATA USED

3.1 Modelled data

3.1.1 Runs produced

Model runs were alternatively performed with the grid-based continental AWRA system and a more flexible one-dimensional implementation of the model that allows it to be used with pre-processed climate time series. Both implementations use the same structure, parameters and forcing. The model structure and assumptions are described in AWRA Technical Report 3, while the implementations are described in AWRA Report No. 2.

The continental simulations were done for the period 1980 to 2009, using the BAWAP climate inputs described in AWRA Report No. 2. The time period modelled for the flux tower sites and catchments are described in the respective sections. Generally, the same BAWAP input data were used but some of the data was replaced with SILO gridded climate data, either because preformatted BAWAP data averaged by catchment were not available at the time, or in case of the BAWAP radiation product, because there were missing data values.

3.1.2 Parameterisation

All model runs performed for this report were done using the same set of parameter estimates. The two catchment parameters ($K_{gw}$ and $K_{rout}$) and three HRU specific parameters ($F_{drainFC}$, $S_{gref}$ and $F_{hru}$) are estimated from climate and land cover input data as part of model initialisation in the manner described in AWRA Technical Report 3. Default values were used for the remaining HRU parameters with fixed values, with the exception of three parameters that were iteratively estimated (Table 1):

- Reference daily rainfall ($P_{refR}$), describing the relationship between net rainfall and infiltration-excess surface runoff. The default parameter (1013 mm) produced severe underestimates of storm runoff response for reasons that are yet to be investigated. Limited manual iteration was performed; testing values of 50, 100, 150, 250 and 500 mm. A value of 150 mm appeared to reproduce the overall flow duration curve for all daily stream flow (in mm/day) pooled across the 326 catchments reasonably well and was adopted.

- Shallow root zone water storage at field capacity ($S_{sFC}$). No default estimate was available. Values of 50, 100, 150, 200, 250 and 300 mm were tested, and a value of 200 mm appeared to reproduce greenness patterns for two flux tower sites in grassland and open woodland (Kyeamba and Virginia Parks; Section 5) reasonably well.

- Deep root zone water storage at field capacity ($S_{sFC}$). No default estimate was available. Values of 200, 600 800, 1000 and 2000 mm were tested, and a value of 1000 mm appeared to reproduce greenness patterns for two flux tower sites in forest and savannah woodland (Tumbarumba and Howard Springs, cf. Section 5) reasonably well.

---

2 Words printed in Courier font refer to model code; see AWRA Report 3.0.5 for equivalent mathematical symbols.
Table 1. Parameter settings used in all runs performed for the current evaluation of AWRA version 0.5. All values are default values except those shaded and printed bold (see text). Listed are the ‘model code’ notation used in the MATLAB code for AWRA-L version 0.5, and the ‘symbol’ notation used in the textual model description. Both can be found in AWRA Technical Report 3.

<table>
<thead>
<tr>
<th>Code</th>
<th>Symbol</th>
<th>HRU1</th>
<th>HRU2</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>alb_dry</td>
<td>$a_{dry}$</td>
<td>0.26</td>
<td>0.26</td>
<td>dry soil albedo (-)</td>
</tr>
<tr>
<td>alb_wet</td>
<td>$a_{wet}$</td>
<td>0.16</td>
<td>0.16</td>
<td>wet soil albedo (-)</td>
</tr>
<tr>
<td>beta</td>
<td>$\beta$</td>
<td>4.5</td>
<td>4.5</td>
<td>coefficient describing rate of hydraulic conductivity increase with water content (-)</td>
</tr>
<tr>
<td>cGsmax</td>
<td>$n/a$</td>
<td>0.03</td>
<td>0.03</td>
<td>multiplier to estimate Gsmax from PCI (m s$^{-1}$)</td>
</tr>
<tr>
<td>ER_frac_ref</td>
<td>$F_{ER0}$</td>
<td>0.20</td>
<td>0.05</td>
<td>average ratio of wet canopy evaporation rate and rainfall rate for full canopy cover (-)</td>
</tr>
<tr>
<td>Fgw_conn</td>
<td>$F_{dg}$</td>
<td>1</td>
<td>1</td>
<td>factor describing soil-groundwater connectivity (-)</td>
</tr>
<tr>
<td>FsoilEmax</td>
<td>$f_{sEmax}$</td>
<td>0.2</td>
<td>0.5</td>
<td>maximum soil evaporation fraction (-)</td>
</tr>
<tr>
<td>fvegref_G</td>
<td>$F_{S,ref}$</td>
<td>0.15</td>
<td>0.15</td>
<td>reference soil cover fraction that determines the rate of decline in energy loss with increasing canopy cover (-)</td>
</tr>
<tr>
<td>FwaterE</td>
<td>$F_{OW}$</td>
<td>0.7</td>
<td>0.7</td>
<td>open water evaporation scaling factor (-)</td>
</tr>
<tr>
<td>Fmax</td>
<td>$F_{loss,max}$</td>
<td>0.3</td>
<td>0.3</td>
<td>maximum fraction of daytime net radiation 'lost' to heat storage when there is no vegetation (-)</td>
</tr>
<tr>
<td>hveg</td>
<td>$h$</td>
<td>10</td>
<td>0.5</td>
<td>vegetation canopy height (m)</td>
</tr>
<tr>
<td>InitLoss</td>
<td>$I_0$</td>
<td>5</td>
<td>5</td>
<td>initial retention capacity (mm)</td>
</tr>
<tr>
<td>LAI_max</td>
<td>$\Lambda_{max}$</td>
<td>8</td>
<td>8</td>
<td>maximum achievable LAI (-)</td>
</tr>
<tr>
<td>LAI_ref</td>
<td>$\Lambda_{ref}$</td>
<td>2.5</td>
<td>1.4</td>
<td>reference LAI determining canopy cover (-)</td>
</tr>
<tr>
<td>PrefR</td>
<td>$P_{ref}$</td>
<td>150</td>
<td>150</td>
<td>reference event precipitation for runoff generation (mm d$^{-1}$)</td>
</tr>
<tr>
<td>S_sls</td>
<td>$s_V$</td>
<td>0.1</td>
<td>0.1</td>
<td>canopy storage capacity per unit leaf area (mm)</td>
</tr>
<tr>
<td>S0FC</td>
<td>$S_0FC$</td>
<td>30</td>
<td>30</td>
<td>accessible top soil water storage at field capacity (mm)</td>
</tr>
<tr>
<td>SdFC</td>
<td>$S_{dFC}$</td>
<td>1000</td>
<td>1000</td>
<td>accessible deep soil water storage at field capacity (mm)</td>
</tr>
<tr>
<td>SsFC</td>
<td>$S_{sFC}$</td>
<td>200</td>
<td>200</td>
<td>accessible shallow soil water storage at field capacity (mm)</td>
</tr>
<tr>
<td>SLA</td>
<td>$C_{SLA}$</td>
<td>3</td>
<td>10</td>
<td>specific leaf area per unit dry leaf biomass (m$^2$ kg$^{-1}$)</td>
</tr>
<tr>
<td>Tgrow</td>
<td>$t_{grow}$</td>
<td>1000</td>
<td>150</td>
<td>time constant determining rate of canopy increase (d)</td>
</tr>
<tr>
<td>Tsenc</td>
<td>$t_{senesce}$</td>
<td>60</td>
<td>10</td>
<td>time constant determining rate of canopy decrease (d)</td>
</tr>
<tr>
<td>U0d</td>
<td>$U_{0d}$</td>
<td>4</td>
<td>0</td>
<td>maximum root water uptake from deep soil (mm d$^{-1}$)</td>
</tr>
<tr>
<td>U0s</td>
<td>$U_{0s}$</td>
<td>6</td>
<td>6</td>
<td>maximum root water uptake from shallow soil (mm d$^{-1}$)</td>
</tr>
<tr>
<td>Vc</td>
<td>PCI</td>
<td>0.35</td>
<td>0.65</td>
<td>photosynthetic capacity index</td>
</tr>
<tr>
<td>w0limE</td>
<td>$w_{0lim}$</td>
<td>0.85</td>
<td>0.85</td>
<td>relative top soil water content at which evaporation is reduced (-)</td>
</tr>
<tr>
<td>w0ref_alb</td>
<td>$w_{0,ref}$</td>
<td>0.3</td>
<td>0.3</td>
<td>reference value of $w_0$ describing the relationship between albedo and top soil wetness (-)</td>
</tr>
<tr>
<td>wdlimU</td>
<td>$w_{dlim}$</td>
<td>0.3</td>
<td>0.3</td>
<td>relative deep soil water content at which root uptake is reduced (-)</td>
</tr>
<tr>
<td>wslimU</td>
<td>$w_{slim}$</td>
<td>0.3</td>
<td>0.3</td>
<td>relative shallow soil water content at which root uptake is reduced (-)</td>
</tr>
</tbody>
</table>

The notations used in the model code (reproduced in Appendix A of this report) vary from those used in the model description in AWRA Technical Report 3, due to revisions in the latter after
model development and evaluation. This will be addressed in future versions. To assist interpretation both are listed in Table 1.

### 3.2 Observations

A wide variety of on-ground and satellite observations is available and could potentially be used to evaluate model performance. A key requirement is that the observations are equivalent to a model-simulated quantity, or at least can usefully be compared to it or estimated from it. In this first evaluation, the emphasis was on observations that either have traditionally been considered to provide a reliable indication of the quality of water balance observations; or were readily available; or both:

- gauged streamflow records for 362 small, unimpeded catchments;
- eddy covariance flux tower evapotranspiration data at four sites;
- radar remote sensing (ASAR GM) derived top soil water content estimates;
- optical remote sensing (AVHRR, MODIS) derived vegetation fractional cover, leaf area index and greenness;

A number of data sets has been identified and is intended for use in future evaluation. These include the following on-ground measurements:

- streamflow data for a larger number of catchments representing a wider range of environments;
- long-term average deep drainage estimated by a range of field techniques (R. Crosbie, pers. comm.);
- *in situ* soil water content measurements available from a series of field campaigns and ongoing measurements by University of Melbourne researchers (e.g. Merlin et al. 2008).

Additional satellite data sets that will be considered include:

- optical remote sensing (MODIS) derived surface albedo;
- passive microwave remote sensing derived top soil water content estimates (Liu et al. 2009);
- estimates of total terrestrial water storage derived from observations by the Gravity Recovery and Climate Experiment (GRACE).

It is intended that these will be considered for inclusion in future updates of this evaluation document. Additional data may have become available by then. One of those data sets is *in situ* soil water content measurements made by eleven cosmic ray soil moisture sensors (Zreda et al. 2008) that were recently purchased by CSIRO and will be deployed at sites across Australia.
4. GAUGED STREAMFLOW

Summary

Catchment streamflow records from 362 small catchments minimally affected by regulation were reproduced well, despite the lack of parameter calibration. Calculated performance statistics were equivalent or better than those obtained with other rainfall-runoff models in comparable experiments. Important features were the lack of bias and the strong improvement in performance as data was aggregated for multiple catchments and over longer time scales. The model structure appeared robust and development of operational model-data fusion techniques are recommended.

4.1 Introduction

4.1.1 Justification

One of the key variables of interest to be produced by the AWRA system will be estimates of streamflow. These estimates are of use in the production of annual water accounts as well as annual, seasonal or perhaps more frequent or occasional water resources assessments. Water accounts will be generated for the continent, for drainage divisions, river basins and perhaps smaller spatial units depending on their relevance (D. Barratt, Bureau of Meteorology, pers. comm.). Water assessments may be produced at similar spatial scales, and report on streamflow over the past week, month, season or year.

Most commonly, evaluation of streamflow estimation methods is done by using selected streamflow records. These records are usually produced using an automatic water level recorder or logger that measures the water level at a defined river cross section on a regular basis every few minutes. These are combined with a stream flow rating curve, which is based on irregular hand measurements of stream flow rate averaged across the cross section at known water levels. This procedure, as well as interruptions in the water level measurements, means that the quality and completeness of streamflow records can vary.

To evaluate models that describe the generation of streamflow in small catchments (as opposed to routing and other hydrological processes within longer rivers), usually records are selected for those catchments where it is expected that important assumptions in the model are met. This excludes, for example, catchments with extractions, strong riparian losses, water bodies, inter-basin transfers, etc. In Australia, the most common approach has been to use the data set collated initially for the National Land and Water Resources Audit (Peel et al. 2000) and reassessed and extended for subsequent projects (e.g. CSIRO’s ‘Sustainable Yields’ projects, South East Australia Climate Initiative, CRC eWater product development).

These data were also used here and are described below. Their advantage is that they are readily available and relatively well quality controlled. Some disadvantages are that they are generally small and that the sample is biased, for example, towards southeast Australia, wetter catchments, and higher relief catchments. Their small size also means that they will show a fast response between rainfall and streamflow with steeper peaks (that is, more high frequency variation) and are more likely to be influenced by the effects of smaller scale variation in precipitation, terrain and geology. Therefore, it is worthwhile to consider the effect that
averaging over longer time steps and larger areas has on the model evaluation. A method to do so is described below.

4.1.2 Observations

Daily streamflow data (in ML d⁻¹) were collated for 362 catchments across Australia as part of previous studies (Peel et al. 2000; Guerschman et al. 2008; 2009b). Streamflow data for these selected catchments were considered of satisfactory quality and any influence of river regulation, water extraction, urban development, or other processes upstream considered unimportant. Large lakes or wetlands do not occur in any of the catchments, but smaller impoundments can occur. The data set includes catchments under native forest, catchment fully cleared for grazing, and catchments with a varying combination of cropping, grazing, plantation forestry and native vegetation. From the data set, those records were selected that had good quality observations for at least five years during the period 1990–2006.

The selected 278 stations are mostly located in southwest West Australia, Tasmania, and the coastal regions and Great Divide of the eastern states. The contributing catchments of all gauges were delineated through digital elevation model analysis and visual quality control. Catchment areas varied between 23–1937 (median 278) km². Daily streamflow volumes were converted to streamflow depths (Q, mm d⁻¹) and varied from 4 to 1937 (median 114) mm y⁻¹.

4.1.3 Caveats

For reasons explained in Section 1 differences between model estimates and observations are often not easily and confidently attributed to lacking model performance alone. Based on historic experience with daily streamflow observations, a number of common sources of difference between model estimates and records can be identified:

Model input: in particular, catchment-average rainfall data are often based on a small number of rain gauges. For the 278 catchments considered here, on average three (range 0–22) gauges were inside or within 5 km of each catchment were used in the interpolation product. There is much opportunity for bias: precipitation gauges tend to be located where they are of interest and easily read (such as on cleared land, in settlements, along roads). Biases also tend to arise due to interpolation of sparse data (Li and Shao accepted).

Model structure: inevitably the model makes simplifying assumptions about the way in which streamflow is generated, and therefore differences arising from these simplifications would be fully expected.

Model parameterisation: even if the model structure were perfect and all parameters had a well-defined physical meaning, the lack of knowledge and observations to estimate these parameter values for each individual catchment will be imperfect and therefore will lead to differences.

Catchment delineation: this affects the model through the selection of gridded input data, but particularly affects the observations, as catchment area is used to convert streamflow volume to depth. Errors can occur because of the quality or use of the digital elevation model used, or because of the groundwater from outside the surface catchment surfaces enters the stream. Previous assessment showed that errors of 10-20% easily occur.

Gauging errors: water level records can also be in error, for example due to unidentified blocking of the connecting pipe, sedimentation of the stilling well, instrument malfunction, etc. Another common problem is the occurrence of ungauged bypass flows. These can occur as subterranean transfers through the channel substrate or as deeper groundwater flows, which will
result in an apparent overestimation of low flows by the model. Bypass flows are also quite common at very high water levels, where streamflow tops the gauged channel and radically changes the shape of the rating curve in a way that is not taken into account.

**Rating curves:** water level-discharge rating curves are used to estimate daily stream flow from water level readings. This assumes that there is a monotonic unique relationship between water level and discharge, which is often not the case due to, for example, slow or fast channel morphology changes, backwater effects, tidal influences, and water temperature variations. The extent to which the streamflow regime is covered is also variable; often measurements for the rating curve are not available or questionable at high flow rates, and therefore these estimates rely on extrapolation of the rating curve. The magnitude of likely errors varies as a function of the stability of the river cross-section, the accuracy of rating curve measurements and the range of experienced flows that are rated.

**Streamflow data processing errors:** Most of the streamflow data used in this study included data quality flags indicating were measurements were questionable or missing. However the accuracy of quality flags is known to be lacking sometimes (R. Sparks, CSIRO, pers. comm.) and sometimes erroneous or gap-filled data will have remained in the data used. Guidance on the errors in streamflow estimates is difficult to provide, because errors will vary from station to station, and because it is not always clear what part of the errors may be systematic (‘red noise’) and what part random (‘white noise’) and therefore lost in aggregation. Published error assessments suggest that the most important source of error in accurate discharge estimation is normally in the rating curve (Di Baldassarre et al. 2006; Harmel et al. 2006). The latter authors estimated errors to increase from 10% at low flows to ca. 25% at the highest rated flows and up to 40% for peak flows, which are often unrated. Numerical experiments for this scenario and using the same set of streamflow used in this evaluation (author, unpublished) suggest that even if all errors were random and the model perfect, these errors would degrade the best achievable NSME to ca. 0.75 on average and create an apparent bias of up to 5% for catchments with few flow events. Conversely, a systematic bias of <10% would not reduce the achievable NSME much but would add directly to the apparent bias. An unavoidable bias may occur due to bias in streamflow records – due to, for example, wrong estimates of the catchment area, bypass flows, or erroneous extrapolation of the rating curve –or the model inputs; rainfall in particular. As a conservative estimate it may be assumed that agreement between average flows should not be expected to be better than 10–15%. Where the same record is used for both calibration and validation, much better performance could in fact indicate ‘over-fitting’ (that is, calibration reproduces errors in the record). Such over-fitting is unlikely to have occurred in the current case, where a single parameter set was used for all catchments, unless it were a bias common to all catchments; for example, due to a bias intrinsic to the gauging technique or rainfall estimation. Similarly, averaging data from more than one gauge record would be expected to reduce much of the bias in the observations unless they were intrinsic. This is not necessarily true for temporal aggregation for an individual record. These caveats should be considered when interpreting the evaluation results.

### 4.2 Method

#### 4.2.1 Model run

Daily precipitation and atmospheric variables were calculated from the SILO gridded 0.05° climate products derived by interpolation of station data (Jeffrey et al. 2001). Fraction deep-rooted vegetation was estimated from the AGO Landsat-based tree cover mapping product
(Furby 2002). Using the GIS catchment coverage, catchment-average time series of the SILO data and of tree cover were calculated.

The range of average annual rainfall for the catchment sample was 409–2983 (median 852) mm y\(^{-1}\); precipitation other than rainfall was insignificant. Priestley-Taylor potential evapotranspiration (Priestley and Taylor 1972) \((E_0)\) was 651–2120 (1258) mm y\(^{-1}\) and catchment humidity \((H,\) the ratio of average rainfall over average \(E_0)\) was 0.2–3.5 (0.7). Tree cover varied between zero and 90%.

The one-dimensional AWRA version was run with these data inputs and parameters as described in Section 3.1, for the periods 1990-2006. The years 1985-1990 were spin-up years and are not considered in the evaluation.

### 4.2.2 Data aggregation

The degree to which the daily streamflow values and patterns in small catchments are reproduced is arguably one of the more stringent tests of model performance. In practical applications, however, monthly, seasonal or even annual streamflow totals may be of greater interest, equally streamflow may be of greater interest at the scale of large catchments; for example, the combined catchments contributing to a storage reservoir. Prior expectation would be that the difference between estimates and observations is reduced as larger scales are considered, but any systematic differences occurring for all catchments would be expected to remain. To analyse this, the observed and modelled data were transformed to larger temporal and spatial scales.

Daily totals were aggregated to weekly, monthly, seasonal (January-March, and so on) and annual (January-December) means. Time step means were only calculated if more than 70% of daily records were available for that time step.

To investigate the effect of spatial averaging, catchment streamflow was aggregated for all catchments within larger AWRC river basins and drainage divisions based on gauge code (AWRC 1975) weighted by its area to create aggregated area averages. Aggregated data were analysed if they included at least 3 catchments. The number of catchments included in the calculations may sometimes vary over time depending on data availability. To avoid too many variations on the same theme, this experiment was only done for monthly streamflow.

### 4.2.3 Calculated metrics and interpretation

All metrics listed in Section 2 were calculated, for all levels of spatial and temporal aggregation. To allow appropriate comparison of the observed and estimated values, missing values in the observations were also ignored in the model estimates when aggregating. This lack of gap-filling means that period average values of one variable are not always necessarily consistent with period averages of other variables (such as rainfall) nor necessarily provide the best estimate of period-average streamflow.

Also produced were time series plots and flow duration curves (FDCs) for the catchments corresponding to the NSME quartiles; that is, 25% interval percentiles, which includes the catchment corresponding to the worst, best and median performance.

To support interpretation, it was investigated whether some of the variations in any of the performance metrics could be explained to catchment attributes such as size, humidity, number of catchments aggregated, and so on. This was done through assessment of scatter plots.
4.2.4 Comparison with performance of Zhang model

Simple models exist to estimate long-term average streamflow based on catchment climate alone and have been fairly widely used in Australia as well. This provides an opportunity to compare AWRA performance with these other models.

Several equations have been proposed to describe the relationship between longer term average precipitation and potential evapotranspiration (PET) on one hand, and the partitioning of rainfall into actual ET and catchment streamflow on the other, by considering the demand (PET) and supply (precipitation) limits on actual ET (Oudin et al. 2008). These equations are sometimes referred to as Budyko-type models, and typically have a single parameter that describes how efficiently ET can occur; that is, how closely actual ET approaches potential ET when demand approaches supply. Despite their simplicity, Budyko-type models typically provide a robust performance for long-term average fluxes and as such set a useful performance benchmark. It was shown before that most of the Budyko-type models produce near similar performance and therefore only one of them was selected here for comparison, namely, the formulation of Zhang and colleagues (Zhang et al. 2001; Zhang et al. 2004). The ‘Zhang equation’ is given by:

$$
\mu_O = \left[1 + \frac{1}{H} + \theta \left(\frac{1}{H}\right)^2\right]^{-\frac{1}{\theta}} \mu_P \quad [4-1]
$$

with

$$
H = \frac{\mu_P}{\mu_{PET}} \quad [4-2]
$$

where $\mu_O$ is long-term average streamflow (in mm/year), $H$ catchment humidity, $\theta$ the model parameter, and $\mu_P$ and $\mu_{PET}$ mean rainfall and potential evapotranspiration (PET) respectively, which were calculated from gridded SILO data; the latter using the Priestley-Taylor formulation.

Zhang et al. allow for a different parameter to be used for the tree and non-tree covered fraction of a catchment based on measurements in experimental catchments. However, subsequent research has shown that no evidence can be found for these differences in ‘real-world’ catchments (van Dijk et al. 2007; Oudin et al. 2008) and therefore only a single parameter was fitted here through least squares optimisation, producing a value of $\theta=1.95$.

4.3 Results

All calculated metrics are available in a Microsoft Excel spreadsheet that accompanies this report (“AWRA_405_Supplement.xls”). Below summaries of these data are presented

4.3.1 Mean streamflow

The mean recorded ($\mu_O$) and model estimated ($\mu_M$) average streamflow are shown on a linear scale in Figure 2a; and on a double logarithmic scale in Figure 2b to better show results for drier catchments. Several metrics are listed in Table 2. The following observations are made:
Both the graphical comparison (Figure 2) and the performance metrics suggest there is no bias or other systematic difference between the estimates and observations: bias is minimal, NSME equals $R^2$, and OER is low.

The difference between the estimates and observations increases as a function of estimated or observed streamflow (that is, the distribution of differences is heteroscedastic; Figure 2a) but appears to decreases somewhat expressed relative to the estimate (Figure 2b) – that is, the residual differences are neither constant nor proportional, but somewhere in between (Figure 2a and b).

Model performance is very similar if marginally better than that of the Zhang equation, explaining 2% more of the overall variance; equivalent to 7% of the residual variance.

Table 2. Performance metrics for period average streamflow for the stations in the analysis (1990-2006, N=278). Acronyms are explained in Section 2. Left AWRA model estimates, right estimates derived with the Zhang model with the parameter fitted to 1.95. Means were converted from mm/day to mm/year by multiplication with 365.25

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AWRA</td>
<td></td>
<td>Zhang model</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>182 242</td>
<td>Observations</td>
<td>182 242</td>
</tr>
<tr>
<td>Model estimates</td>
<td>184 224</td>
<td>Model estimates</td>
<td>187 209</td>
</tr>
<tr>
<td>SD (mm/year)</td>
<td>78</td>
<td>SD (mm/year)</td>
<td>84</td>
</tr>
<tr>
<td>bias (mm/year)</td>
<td>-1</td>
<td>bias (mm/year)</td>
<td>-4</td>
</tr>
<tr>
<td>OER</td>
<td>4%</td>
<td>OER</td>
<td>11%</td>
</tr>
<tr>
<td>AND</td>
<td>28%</td>
<td>AND</td>
<td>28%</td>
</tr>
<tr>
<td>NSME</td>
<td>0.896</td>
<td>NSME</td>
<td>0.879</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.896</td>
<td>$R^2$</td>
<td>0.885</td>
</tr>
<tr>
<td>$RR^2$</td>
<td>0.820</td>
<td>$RR^2$</td>
<td>0.819</td>
</tr>
</tbody>
</table>

Figure 2. Comparison of average AWRA-L estimated and recorded streamflow, expressed on a per annum basis, for days with streamflow records available, for each of the 278 catchments with sufficient data. Solid line is the linear regression equation; dotted line the 1:1 line.
4.3.2 Daily streamflow patterns and flow duration curves

Some characteristics of the distribution of performance for daily streamflow patterns for the 278 catchments are listed in Table 3.

Table 3. Performance metrics for daily streamflow patterns for the stations in the analysis (1990-2006, N=278) (acronyms are explained in Section 2).

<table>
<thead>
<tr>
<th></th>
<th>σₒ</th>
<th>σₘ</th>
<th>SD</th>
<th>bias</th>
<th>OER</th>
<th>AND</th>
<th>NSME</th>
<th>R²</th>
<th>RR²</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>1.74</td>
<td>1.18</td>
<td>1.35</td>
<td>-0.01</td>
<td>0.06</td>
<td>1.19</td>
<td>-6.01</td>
<td>0.40</td>
<td>0.90</td>
</tr>
<tr>
<td>st.dev.</td>
<td>1.79</td>
<td>1.07</td>
<td>1.19</td>
<td>0.22</td>
<td>0.22</td>
<td>1.13</td>
<td>97.09</td>
<td>0.21</td>
<td>0.08</td>
</tr>
<tr>
<td>minimum</td>
<td>0.02</td>
<td>0.20</td>
<td>0.20</td>
<td>-0.80</td>
<td>-0.43</td>
<td>0.34</td>
<td>-1583</td>
<td>0.00</td>
<td>0.62</td>
</tr>
<tr>
<td>25% percentile</td>
<td>0.65</td>
<td>0.61</td>
<td>0.67</td>
<td>-0.09</td>
<td>-0.11</td>
<td>0.75</td>
<td>-0.01</td>
<td>0.24</td>
<td>0.86</td>
</tr>
<tr>
<td>median</td>
<td>1.11</td>
<td>0.88</td>
<td>0.91</td>
<td>-0.01</td>
<td>0.07</td>
<td>0.94</td>
<td>0.30</td>
<td>0.41</td>
<td>0.92</td>
</tr>
<tr>
<td>75% percentile</td>
<td>2.15</td>
<td>1.43</td>
<td>1.64</td>
<td>0.07</td>
<td>0.24</td>
<td>1.00</td>
<td>0.47</td>
<td>0.57</td>
<td>0.95</td>
</tr>
<tr>
<td>maximum</td>
<td>11.09</td>
<td>8.99</td>
<td>6.84</td>
<td>1.01</td>
<td>0.50</td>
<td>11.65</td>
<td>0.75</td>
<td>0.85</td>
<td>1.00</td>
</tr>
</tbody>
</table>

The following observations are made:
- The σₘ is smaller than σₒ (Table 3). This may be attributed to the frequent failure of the model to reproduce the largest peak flows, which have the greatest variance.
- The SD is large (median SD=0.91 mm/day) and in the same order of magnitude as the standard deviation (median σₒ=1.11 mm/day).
- Necessarily, the median bias found is the same as that found for long-term average streamflow values.
- There is no obvious tendency for the model to consistently over- or underestimate flows (median OER is 0.07).
- The relative error in daily flows tends to be large (median AND 0.94). This value is mainly due to overestimates of very low flows, which easily leads to very high AND values (relative differences are much smaller for higher flows).
- The median NSME is 0.30 and 26% of the catchments had a NSME that was less than zero.
- The relative agreement as expressed in R² was somewhat better, with a median value of R²=0.41.
- The ranking of flows is generally reproduced very well, with a median RR² of 0.92.

Plots of daily streamflow are not very informative due to the strong skewness and high temporal variability in the time series, but monthly averages as well as the daily flow duration curves (FDCs) are shown in Figure 3 for the six catchments corresponding to the quartiles of the NSME distribution.
Figure 3. Five examples of model performance corresponding with quartiles of the daily flow NSME distribution, from (top) worst to (bottom) best. Observations are shown in blue and model estimates in red, presented as 30-day running averages (left panel) and daily FDCs (right panel). Note that vertical scales do vary.
The following observations are made about these figures:

- Figure 3a shows the catchment with the poorest performance (Dart Brook at Aberdeen) and suggests a vast overestimation of recorded rainfall runoff response. The observations appear to suggest that there is no baseflow, whereas the estimates do. This was observed for some other records as well.

- The poorest model performance (in the figures shown in terms of NSME) is found for catchments with low average runoff rates, whereas the best performance is typically found for catchments with high average runoff (note that the vertical scales vary in Figure 3).

- Both over- and underestimations of high flows occur, as evidenced by considering the peak monthly flows and the high flow end of the flow duration curve.

- Agreement in the flow duration curves tends to be poor for the lowest flows (<0.1 mm/day).

- Figure 3b appears to indicate a shift change in recorded rainfall runoff response occurring around 1997 for this catchment (Avoca River at Amphitheatre) that is not reproduced by the model. Similar temporal non-stationarity in the record-model differences was identified in some other catchments, often leading to low performance scores.

### 4.3.3 Effect of temporal aggregation

In Figure 4 the effect of temporal aggregation on some of the calculated metrics is shown. The following observations are made:

- The standard deviation in the observations (as well as in the estimates; data not shown) rapidly decreases through aggregation, from 1.11 mm/day on a daily time step to 0.22 mm/day (80 mm/year) on an annual time step.

- The standard difference (SD) rapidly decreases when aggregating to weekly and monthly averages, from a median 0.91 mm/day for daily totals to 0.36 mm/day (131 mm/year) for annual time step (and 78 mm/year for long term averages; cf. Section ‘Mean Streamflow’).

- The median value of the relative difference (AND) reduces from 94% for daily to monthly time steps, to 42% for annual values (and 28% for long-term averages, cf. Section 4.3.1).

- The median NSME for streamflow values increases somewhat through temporal aggregation, from a median value of 0.30 for daily values to 0.40–0.41 for seasonal and annual streamflow totals.

- By comparison, $R^2$ values increase steadily through aggregation; from 0.41 for daily data to 0.72 for annual totals. The difference between $R^2$ and NSME values is equivalent to the fraction of systematic differences. It follows that over longer time steps the streamflow patterns are increasingly well reproduced, but remaining bias means that NSME values remain lower than $R^2$.

- RR$^2$ values are consistently high (0.90 vs. 0.95), implying that the ranking of flows is generally reproduced very well. This suggests that applying a bias correction function would allow a large part of the variance to be explained. RR$^2$ changes little through temporal aggregation.

- Similar as for differences in mean streamflow, SD and AND for daily streamflow are highly heteroscedastic: the absolute errors are greater and relative errors smaller for high flows than for low flows (data not shown).
Figure 4. Distribution of metrics of model performance and the effect of temporal aggregation. Shown are the median value of the metric (solid line and markers), the minimum and maximum values (dotted line without markers), and the 25% and 75% quartiles (dashes line with open markers). Acronyms as explained in Section 2. (Note that the minimum and maximum sometimes are beyond the range of the vertical axis).
4.3.4 Effect of spatial aggregation

Spatial averages were calculated for 39 river basins and 6 drainage divisions. The river basins varied from size and number of catchments from a river basin with 3 catchments together covering 321 km² to one with 23 catchments covering 10,202 km². The 6 drainage divisions with more than 2 catchments included the North East coast (20 catchments), South East coast (148), South West coast (10), Tasmania (12), the South East Gulf (9) and the Murray-Darling Basin (123). It is noted that the aggregated catchments still only represent a small fraction of total area (for example, 53,000 km² out of 1.06 million km² in the case of the Murray-Darling Basin, or 5% of total area).

As would be expected, on average there was an approximately proportional relationship between the number of catchments and the total area (Figure 5a). It appeared however, that the number of catchments that was aggregated were consistently a better predictor of the performance metrics than the total area involved. This could suggest that measurement errors or variability at the scale of individual catchments play an important role in causing differences between model estimates and records.

In Figure 5b-h, the effect of spatial aggregation on several indicators of model performance is shown. The following observations are made:

- There is a strong relationship between the number of catchments that are combined and the resulting $\sigma_0$ values. Because most of the standard deviation is caused by flow peaks associated with rainfall events, it would seem logical to conclude that averaging across more catchments reduces the local effects associated with spatial rainfall variability.

- All indicators of model performance improve as the data is aggregated across more catchments.

- The bias between model estimates and records reduces very rapidly as data from several catchments is aggregated. The absolute bias (bias) reduces from a median of ±0.09 mm/day for individual catchments, to about four orders of magnitude less (ca. ±0.0001 mm/day) for large numbers of catchments. This further confirms that the model estimates almost free of bias.

- There is a reduction in SD that is equally strong as in $\sigma_0$; SD reduces almost three orders of magnitude going from a single catchment to data averaged over more than 100 catchments.

- There is some suggestion that relative error (AND) reduces as data is spatially aggregated, but the relationship is weak.

- The absolute agreement (NSME) appears to improve as data is aggregated, from a median NSME<0.4 for individual catchments, to ca. 0.70 for large numbers of catchments. The relationship is not very strong however.

- The relative agreement ($R^2$) appears to improve through spatial aggregation, but resulting values are generally very similar to NSME values. This implies that spatial aggregation appears to remove much of the bias.

- Spatial aggregation appears to have little influence on ranking performance ($RR^2$).
Figure 5. Effect of spatial aggregation on selected metrics of model performance. Shown in (a) is the relationship between number of catchments and aggregate area. Open dots indicate the median value for the individual catchments (cf. previous section). Dashed lines are linear or power regression equations added to assist visual interpretation of trends. Acronyms as explained in Section 2.
In Figure 6, monthly streamflow time series and monthly FDCs are shown for those catchments aggregates corresponding to the quartiles of the NSME for monthly streamflow. The following observations are made:

Compared to monthly streamflow patterns in individual catchments, the aggregated data shown much stronger agreement between model estimates and records.

Systematic over- and underestimations still appear to occur, both for high flows and low flows, and sometimes both (e.g. Figure 6a and b).

The flow duration curves generally agree quite well, but some apparently systematic over- and underestimations in different flow ranges are still evident.
Figure 6. Five examples of model performance corresponding with quartiles of the weekly flow NSME distribution for flows aggregated by river basin or drainage division, from (top) worst to (bottom) best. Observations are shown in blue and model estimates in red, presented as 7-day running averages in both panels.
4.4 Discussion

4.4.1 Mean streamflow

Mean estimated streamflow in the 278 catchments generally agreed well with observations, and produced slightly better results than a calibrated Budyko-type (or ‘Zhang’) model. A large bias was not to be expected, since three important model parameters were iteratively changed (Section 3) to produce mean streamflow totals that were in approximate agreement with the observations. However, the virtual absence of any bias in the results cannot be fully attributed to the limited amount of iterative parameter estimation. It may be simply be fortuitous. In this light, the fact that the estimates are already better than those produced by the Zhang equation are encouraging, considering that the model parameter of that equation was fully optimised to minimise the SD.

The non-homogeneous difference distribution means that it is not straightforward to compare the metrics found here with values reported elsewhere. For example, ignoring 11 catchments with rainfall in excess of 1500 mm/year rainfall, mainly located in North Queensland and Tasmania, reduced SD from 78 to 65 mm/year. Similarly, only using catchments located in the Murray-Darling Basin \( n = 104 \) reduced SD to 49 mm/year. Most of this could be attributed to the removal of catchments with higher streamflow and higher absolute (but smaller relative) differences.

4.4.2 Daily streamflow patterns and flow duration curves

The estimates performed less well in reproducing daily streamflow observations. The low NSME scores and high SD can be attributed to the general underestimation of the highest peak flows. The flow duration curves and the OER scores demonstrate that conversely low flows tend to be overestimated.

Estimation of high and low flows

Many catchments showed an underestimation of peak flows. It is emphasised that the model was used with a global parameter set that received very little calibration, and it therefore seems likely that better estimates might be achieved by optimising the single estimate of \( P_{\text{ref}} \) (and perhaps \( S_{\text{gref}} \)) across all catchments.

Limited testing for individual catchments also showed that considerably better SD, NSME and \( R^2 \) metrics can be achieved by adjusting the two parameters involved in simulating storm flow \( (P_{\text{ref}} \text{ and } S_{\text{gref}}; \text{ see AWRA Technical Report 3}) \). These metrics are calculated after squaring differences (Section 2), thereby putting much more emphasis on days with high flows. (Partly for this reason, the suitability NSME and \( R^2 \) as general purpose goodness-of-fit measures has been questioned; Legates and McCabe 1999; Criss and Winston 2008). Calibrating these two parameters alone is already likely to lead to improved performance metrics. Whether calibration also helps improve performance in ungauged catchments depends on the spatial coherence in parameter values. Recent work with a very similar set of catchments and model structure suggested that spatial correlation in these two parameters exists but is comparatively weak (Van Dijk 2010).

There will likely be a limit to the accuracy with which observed high flows can be reproduced, for two main reasons. Firstly, the accuracy of high flow data in the records can be questionable because of gauging and rating problems (see Caveats). Secondly, the role of rainfall infiltration excess runoff means that runoff response will also be a function of the spatial and temporal distribution of peak rainfall intensities in the catchment. It is common for (calibrated) rainfall-
runoff models to underestimate the highest flows, and this may be one reason for it, because calibration of model parameters to the largest runoff events will lead to overestimation of runoff in other, smaller events, see for example Figure 3d. Currently no spatial data on rainfall intensity are available, although methods to generate this are under development in WIRADA.

**Overestimation of low flows**

The flow duration curves suggested that low flows are sometimes overestimated. This might well change if the surface runoff model parameters were changed, since greater surface runoff would lead to lesser soil water infiltration and groundwater recharge in the model, and therefore reduced low flows. Some of the catchments showed no or negligible baseflow where the model estimates suggest there was. This may imply that there really is no slow flow component in these catchments, or it may be that groundwater is lost from the catchment without flowing through the gauged channel (see Caveats). It is worth noting that it has been found previously that the baseflow recession coefficient, an important determinant of low flow behaviour, shows a fair degree of spatial correlation (Van Dijk 2009b). This spatial correlation was not exploited in parameterising the model for the current evaluation, but techniques to do so are currently being implemented in further model application.

**Comparison with published modelling studies**

No studies have been published that use the exact same data set and metrics and therefore a direct comparison cannot be made. Those studies that did use a similar data set involved much more intensive parameter calibration (e.g. Peel et al. 2000; Viney et al. 2008; Zhang and Chiew submitted). It is all but certain that a better performance would have been achieved had the AWRA model been calibrated following these same approaches. Nonetheless, it may be of interest to assess the difference between the performance that is currently achieved and the performance achieved in those studies, as an indicator of the improvement that may be possible through more effort in global and local parameter estimation. Optimal parameter calibration and parameter estimation methods and model-data fusion techniques for application in AWRA are currently being developed in or implemented through WIRADA research.

Peel and colleagues (2000) calibrated the SIMHYD model to a data set of 331 catchments, of which many catchments are also in the data set used here. For each, they calculated the AND for average streamflow and NSME for daily streamflow for one part of the record, after calibrating several model parameters to the other part of the record (that is, a split-sample cross-validation). A median NSME of approximately 0.75 was achieved, with NSME>0.60 for about 76% of the catchments. By comparison, these numbers are ca. 0.30 and zero, respectively in the current evaluation. Furthermore, “for 95% of catchments the cross validation estimate of total streamflow was within 15% of the recorded total streamflow, for 87% of catchments within 10% and for 68% of catchments within 5%” (Peel et al., 2000). These numbers are 29%, 22% and 10% of the catchments, respectively, for the estimates evaluated here. Given that the ‘validation’ data were far from independent of the calibration, the numbers quoted in Peel et al. (2000) are probably an optimistic estimate of the performance that can realistically be achieved. For example, the approach will have ensured that any systematic bias in rainfall estimates will have been ‘corrected’ by the calibrated model parameters. Indeed, as streamflow gauging is estimated to be in error by 15% on average (much of which may be systematic) the fact that 95% of all gauges produces total flows within this 15% may well indicate over-fitting, that is, the parameters to some degree may reproduce observational errors rather than reality.
Separate, more independent calibration/verification experiments were reported by Viney et al. (2008) and Zhang and Chiew (2009) using data that are in part the same as those used here. Using calibrated parameters from a single donor catchment Viney et al. achieved NSME values of 0.21-0.61 and 0.29-0.50 for respectively the Sacramento and Simhyd rainfall-runoff models. They obtained slightly better results again (median NSME of around 0.65) by using the average of ten to twelve estimates obtained with the two models and parameter sets for the five to six closest catchments. Following the same approach, they found a median difference in average streamflow of -15% to +10% for the two respective models, while estimates were within ca. ±20% of recorded values for about half of all catchments. Zhang and Chiew obtained a somewhat lower NSME values for daily flows of 0.30-0.50 using an eight-member multi-donor technique.

Similar statistics calculated for the currently evaluated model estimates are median NSME for daily flows of 0.30, median relative bias of +4% and half of estimates within ±27% of observed average flows. A direct comparison is impossible however, since none of the reviewed studies assessed the performance that could be achieved with one single, a priori estimated parameter set for all catchments, as was done for the current evaluation. In this respect, the results obtained here are encouraging. However, given the many differences between the studies (in parameter estimation, in the streamflow records used) any conclusions are necessarily tentative. A formal benchmarking study is currently underway and should better allow conclusions about the relative merits of alternative model structures to be drawn. Even so, the results obtained by these earlier studies suggest that an NSME higher than ca. 0.60-0.65 should not be expected even with the best parameter estimation technique.

### 4.4.3 Effect of temporal aggregation

Temporal aggregation had a positive effect on most performance metrics, with the exception of bias and the ranked correlation coefficient. The following processes lead to improved performance:

- **The effect of timing differences in runoff peaks are reduced.** Peaks have considerable influence on squared-difference based performance metrics for daily flows, and can occur because of the mismatch in rainfall days (starting 9 am) and streamflow days (unclear, but presumably starting midnight); potential confusion on whether the date label refers to start or end date of the time step; and the delaying of runoff peaks due to routing. Aggregating to greater time steps reduces these mismatching differences.

- **Averaging also reduces also differences due to timing of the release of catchment streamflow, for example storm flow versus baseflow.** Thus, the description of storm flow recession will have negligible impacts at timescales of months or longer, whereas the description of baseflow recession may have little influence at seasonal or annual time scale.

- **Averaging reduces the variance in the observations as well as the residual variance.** Because NSME and $R^2$ increase, however, it must be concluded that the residual variance between model and observations is reduced more than the observed variance.

- **The relative differences calculated are greatest for low flows, when a small absolute difference can mean a large relative difference.** Through aggregation the lowest flow values tend to be removed and as a consequence the AND values calculated are smaller.

- **With annual water accounting in mind, it may be of interest to consider streamflow estimates at annual time step.** For the small catchments analysed here, it appears that the median error that can be expected in the order of 131 mm/year or on average 42% of observed values. However the non-homogeneous difference distribution means that these metrics are best not applied generally; for example, average performance in catchments or
months with high streamflow will be worse in absolute terms but better in relative terms than implied by these numbers.

- Finally, the difference between the $R^2$ and NSME scores increases as time scales are increased. This implies that as data is aggregated to longer intervals, an increasing component of the unexplained variance is systematic. This is promising, since such systematic differences are likely to be easier to address through parameter calibration and model-data fusion than are random errors.

### 4.4.4 Effect of spatial aggregation

Spatial aggregation had a strongly positive effect on most performance metrics. Particularly encouraging was that the bias and SD very rapidly decreased as data from more than one catchment was aggregated. It is concluded that the following processes are likely to have lead to this improvement:

- The effect of systematic errors in rainfall estimates and streamflow records for individual catchments is reduced as data from more catchments are combined.
- The influence of local scale variability in substrate and characteristics of individual catchment is reduced.
- The effect of local variability in unmeasured rainfall properties, in particular the temporal and local scale spatial distribution of rainfall intensity, is averaged out.
- The relative importance of these processes in explaining the effects of spatial aggregation is unclear, however.

### 4.5 Conclusions

Overall it is concluded that even without local catchment calibration the AWRA model provides useful estimates of catchment streamflow. Important characteristics are the apparent lack of bias and the strongly improved performance at longer time scales and larger spatial scales. On the basis of these findings, it is recommended that the results can already be used for water accounting and assessment purposes as is.

Comparison with published studies demonstrates that there is scope for improvement through more effort in parameter estimation, however, particularly in the estimation of local and short-term flow patterns. Although the uncertain quality of rainfall estimates and streamflow records should lead to some caution in putting too much effort in parameter calibration (lest the model reproduces estimation and observation errors rather than reality), previous regionalisation studies have shown that better performance at these shorter and smaller scales can be achieved (e.g. Viney et al. 2008). Based on these studies, an improvement of NSME from 0.30 to around 0.60 would appear feasible.

The fact that the AWRA model showed quite satisfactory performance without any concerted effort in parameter optimisation or estimation suggests that the model structure is robust. This would also be expected on the basis of the analysis that led to the selection of the storm flow and baseflow representations within the model (Van Dijk, 2009, 2010). However, the current evaluation does not allow any conclusive statement to be made as to the relative performance of the AWRA model in reproducing streamflow observations when compared to other possible model formulations, such as SIMHYD, Sacramento, and so forth. Benchmarking experiments are currently underway through a combined effort of WIRADA and the CRC eWater that will allow alternative model structures to be assessed in an objective manner.
The current parameter set was not optimised in any formal way and this is the most obvious way by which to improve model streamflow estimates. Furthermore, earlier analysis (Van Dijk, 2009; 2010) demonstrated that some of the model parameters show spatial correlation that cannot be related to any measurable and continentally available catchment attribute, but which can still be exploited through data interpolation approaches such as kriging. This is particularly true for the groundwater recession coefficient ($k_G$).

In addition to improved parameter estimation, in retrospective applications such as water accounts and assessments, recorded streamflow data can also be used to adjust model estimates further after simulations using spatial model-data fusion techniques. Such a method is currently being trialled in WIRADA and will be assessed in future.

### 4.6 Recommendations

The following recommendations are made:

- The streamflow estimates are sufficiently accurate and bias-free to form the basis of annual water assessments and accounts for catchments, basins and drainage divisions.

- Benchmarking studies should be continued to assess whether with local calibration, the AWRA model structure is suitable to provide estimates of short-term streamflow (days and weeks) or whether better model structures can be defined.

- Automated model parameter fitting should be pursued to derive best estimates for key parameters for individual catchments. Where spatial correlation in parameter values can be established and occurs over sufficiently large lengths to be useable, kriging should be undertaken to exploit this correlation.

- For the purpose of water balance assessment and accounting, a simple model-data fusion approach (e.g. optimal interpolation) should be developed to allow model estimates to be adjusted on the basis of observations. Provided this is justified by appropriate assumptions about model and observation error, this will bring recorded and modelled streamflow into the best possible agreement.
5. FLUX TOWER EVAPOTRANSPIRATION

Summary

Flux tower ET observations at four sites across Australia were made available by the principal investigators. The model reproduced observations well, and the relative and absolute agreement increased as daily data were aggregated to monthly values. The quality of model rainfall forcing was the main source of error, whereas rainfall interception losses were an important source of uncertainty in the observations. Aside from these two issues, there appears to be only modest scope to improve model performance.

5.1 Introduction

5.1.1 Justification

Next to rainfall, evapotranspiration (ET) is usually the largest term in the water budget. In a water balance model, errors in ET estimation are likely to be amplified into the other water balance terms. Several techniques have been developed to measure ET, which can broadly be classified as water budget and micrometeorological techniques.

Water budget methods infer ET as a residual by measurement or estimation of all other terms (inputs, runoff, storage changes) and can be done at large scales, for example in catchment or irrigation area studies, or at small scale, for example in lysimeter or soil moisture studies. The accuracy depends on the accuracy and comprehensiveness with which these terms can be measured. There is a trade-off between precision and representativeness. For example, weighing lysimeters arguably provide the most accurate estimates of ET but their high construction costs combined with the effects of soil disturbance and edge effects mean that they are often not representative of their surrounds (Allen et al. 1991). On the other extreme, catchments can only provide accurate estimates over longer periods (because storage changes need to be ignored or estimated) and can suffer from uncertainty in estimated precipitation (see Section 5).

Micrometeorological methods allow the measurement of ET fluxes by simultaneously measuring components of the radiation energy balance and high frequency variations in vertical wind speed, temperature and humidity. Instruments include eddy covariance instruments, scintillometers, and (now largely disused) thermocouples. In the context of these measurements, the ET flux is usually referred to as the latent heat flux, which can be expressed as $\lambda E$ (in W m$^{-2}$) or $E$ (in mm per time unit)$^3$. Micrometeorological methods are prone to uncertainties and issues of representativeness of their own (see further on) but have the advantage of providing ET information with high temporal resolution, along with comprehensive data on meteorological conditions. This offers better opportunities to identify likely causes for differences between model and observations.

---

$^3$ Conversion requires division with the (slightly temperature dependent) latent heat of vaporisation ($\lambda$) and conversion from seconds to the time unit of interest.
It is noted that other techniques exist to estimate individual components of total ET, for example, sap flow to measure transpiration of individual plants, and canopy precipitation budget measurements to estimate rainfall interception losses. These were not considered in this evaluation but may be considered in future studies.

5.1.2 Caveats

Differences between model estimates and observations are often not easily and confidently attributed to lack of model performance alone (see Section 1). Based on published experiences in comparing modelled ET with flux tower observations, a number of common sources of difference between model estimates and records can be identified:

Model input: in particular, local rainfall data may be based on a small number of nearby rain gauges or represent an average for a larger area.

Model structure: inevitably the model makes simplifying assumptions about the way in which ET is generated, and therefore differences arising from these simplifications would be fully expected. For example, night-time ET is not explicitly estimated by the AWRA model but rather, implicitly included in total ET.

Model parameterisation: even if the model structure were perfect and all parameters had a well-defined physical meaning, the lack of knowledge and observations to estimate these parameter values for each individual flux tower site will be imperfect and therefore will lead to differences.

Errors of representation. The footprint area of flux tower flux measurements varies as a function of measurement height, local vegetation and topography, aerodynamic stability, and wind speed and direction. Fluxes are usually assumed to originate from a more or less Gaussian elliptical footprint upwind from the flux tower, with a maximum diameter of perhaps up to 1 km (Baldocchi 1997) but usually less. This is a common issue in comparing modelled fluxes, which are normally produced using inputs at coarser resolution (for example, ca. 5 km in the case of AWRA). In running the AWRA model, this was partly taken into account by assuming an approximate tree cover value that appeared representative for the surroundings of the flux tower rather than using the cell average values (see further on). However, the varying size and location of the source area inevitably means that the modelled and observed values to some extent will apply to different areas.

Flux measurement errors. There are various sources of both random and systematic errors in $\lambda E$ estimated from eddy covariance measurements. Errors in flux tower $\lambda E$ measurement include errors introduced by the various processing steps and corrections required to calculate latent heat fluxes from the original measurements, and problems of measuring under conditions with rainfall, high humidity, or very low wind speed. It has been suggested that alternative data processing techniques can introduce differences of 15% in instantaneous fluxes (Mauder et al. 2007). A popular test to diagnose possible measurements errors is to consider energy balance closure, that is, whether energy appears to be conserved based on estimates of the component fluxes; the sum of latent and sensible heat fluxes needs to equal net available energy, calculated as net radiation minus ground heat flux and heat storage. Wilson et al. (2002) reviewed flux observations and estimated that on average the energy balance was not closed within 5–30% of latent heat flux on an half-hourly basis. Underestimation of $\lambda E$ appears more common than overestimation. Additional uncertainty is introduced by the need for gap-filling. Falge et al. (2001) analysed the uncertainty introduced by alternative gap-filling strategies and estimated this at approximately 10% of the monthly average for a typical missing data percentage of 30%.
5.1.3 Literature review

There have been a number of studies comparing ET estimated by land surface models (sometimes parameterised with MODIS observations) with flux tower based estimates (see e.g. references in Baldocchi 2008). Some quantitative information based on a non-exhaustive literature review is provided below.

Published model studies cannot be compared directly as the observations and environments pose varying challenges. For example, Morales et al. (2005) compared ET modelled by several land surface scheme to measurements at 15 flux towers across Europe, of which one in a water limited Mediterranean environments (Castelporziano, Italy). They found SD values for monthly average ET of 0.4–0.6 mm/day and typical $R^2$ values around 0.80. However, the drier site showed markedly lower model performance: $R^2$ was ca. 0.30 and SD 0.7 mm/d. Stöckli et al. (2008a) compared the performance of a land surface model with flux tower ET measurements at 15 sites across the globe, among which three sites had a Mediterranean climate. They too found that model performance was consistently poorer at the water-limited sites, with $R^2$ values of around 0.30, compared to ca. 0.50-0.80 for the other, wetter sites. SD values were similar however, at ca. 1–1.5 mm/d, again for monthly totals.

Several studies calibrated and the evaluated MODIS-based ET models against flux tower observations using split sample techniques. Mu et al. (2007) compared MODIS ET estimates with measurements at 19 North American sites and calculated a SD of 0.9 mm/d for 8-day average fluxes. Cleugh et al. (2007) achieved the same results for two of the sites used in the current study (Tumbarumba and Virginia Park). Leuning et al. (2008) evaluated an algorithm that uses the MODIS LAI product for 15 flux tower sites in Europe, USA and Australia, both with explicit site calibration and with a generalised set of parameters. The model variant with six parameters calibrated against the observations produced an average $R^2=0.73$ for 8-day ET averages, varying from 0.43–0.87 between sites; humid sites showed better performance than drier sites. The generalised algorithm produced $R^2=0.68$. Zhang et al. (2009) compared daily ET estimates with measurements at sites in the American Pan-Arctic region, after calibrating the algorithm to 6 other sites in the same region. They calculated an $R^2$ for daily data of 0.66–0.71 (varying between sites), with an SD equivalent to ca. 0.6–1.3 mm/day and a bias of 0.1–0.5 mm/day. It should be noted however that annual average ET rates at these cold sites were only 0.16–0.77 mm/day. Aggregating to monthly average ET rates improved $R^2$ to 0.79–0.85, SD to 0.16–0.2 mm/day; curiously but presumably coincidentally, the results in calibration were worse than those in validation. Using cross-validation Guerschman et al. (2008; 2009b) compared monthly MODIS-based ET estimates against, among others, the four sites used here. They found $R^2$ values of 0.54–0.92 and SD of 0.31–1.03 mm/day. Interestingly, they found better model performance for drier sites than for wetter sites, which was explained by the predictive value of MODIS greenness.

An alternative estimate of the level of agreement that may be expected is provided by published comparisons of alternative field ET measurement techniques. Wilson et al. (2001) and Scott (2009) compared eddy covariance ET measurements to ET estimates obtained with catchment or site water balance measurements. They found annual totals to agree within ±31-50 mm/year or ca. 10% of annual ET. For each estimation approach, different sources of estimation error were identified and appeared to lead to a similar magnitude of uncertainty.

The information in this and the previous section is interpreted as follows:

---

4 published statistics are sometimes in W/m². These numbers are converted to approximate mm d⁻¹ by multiplying with 0.037.
• An error of 10–30% in daily ET estimates should not surprise, due to both measurement and model errors. Aggregating data from daily to monthly time steps markedly improves the agreement between model and observations.

• A high bias between model estimates and eddy covariance measurements would be more expected than a low bias, given that flux tower technology appears more prone to underestimation than to overestimation of real ET. An agreement of better than 10% in annual and long-term averages should not be expected.

• Expressed in terms of $R^2$ or similar statistics, model performance may be expected to be best for at humid sites with high PET seasonality, and worst for water-limited sites with little seasonality in PET.

• Model performance at water limited sites with seasonally green vegetation may well depend primarily on the ability to reproduce patterns of water availability and vegetation vigour.

5.2 Descriptions of sites

The locations and some attributes of the four sites are described in Table 4, whereas the sites and their surroundings are shown in Figure 8. Further details and data analysis are published in the literature for Tumbarumba and Virginia Park (Leuning et al. 2005) and for Howard Springs (Beringer et al. 2003; Beringer et al. 2007). Some further details for the Kyeamba site are provided on the data web site\(^5\). A brief description of each site follows.

The Howard Springs site (HoSp) is near Darwin (NT) and experiences a monsoonal tropical savannah climate. The vegetation is classified as an open forest savannah. Controlled burning of the understory has been undertaken during the dry period in some years, and also occurs in surrounding areas.

The Kyeamba site (Kyem) is in the middle Murrumbidgee catchment (NSW) and experiences a temperate climate with no dry season but a hot summer. The surroundings are mainly grazing land with some trees along fence lines and a nearby creek. The pasture senesces during hot summer periods.

The Tumbarumba site (Tumb) is in the upper Murray catchment (NSW) and experiences a temperate climate with a mild summer and no dry season. The vegetation is classified as wet open sclerophyll forest. Although rainfall exceeds PET in 44% of months, average annual rainfall exceeds PET and therefore the deep rooted vegetation likely has sufficient access to water for most of the time.

The Virginia Park site (ViPa) is located in northern Queensland and experiences a subtropical climate with a dry winter. The vegetation is classified as open woodland savannah. This location receives annual rainfall that is similar to the Kyeamba site but has a higher PET (Table 4).

Table 4: Description of the four flux tower sites (source: Guerschman et al. 2008)

<table>
<thead>
<tr>
<th>Site Name</th>
<th>Howard Springs</th>
<th>Kyeamba</th>
<th>Tumbarumba</th>
<th>Virginia Park</th>
</tr>
</thead>
<tbody>
<tr>
<td>State</td>
<td>NT</td>
<td>NSW</td>
<td>NSW</td>
<td>Qld</td>
</tr>
<tr>
<td>Code</td>
<td>HoSp</td>
<td>Kyem</td>
<td>Tumb</td>
<td>ViPa</td>
</tr>
<tr>
<td>Latitude</td>
<td>-12.4952</td>
<td>-35.324</td>
<td>-35.6557</td>
<td>-19.8833</td>
</tr>
<tr>
<td>Longitude</td>
<td>131.1501</td>
<td>147.5348</td>
<td>148.1521</td>
<td>146.5539</td>
</tr>
<tr>
<td>Annual rainfall</td>
<td>1764</td>
<td>502</td>
<td>1027</td>
<td>524</td>
</tr>
<tr>
<td>(mm/y)</td>
<td>2076</td>
<td>1510</td>
<td>1198</td>
<td>1884</td>
</tr>
<tr>
<td>Annual PET</td>
<td>31%</td>
<td>23%</td>
<td>44%</td>
<td>7%</td>
</tr>
<tr>
<td>(mm/y)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 7. Range of monthly precipitation (blue and) and Priestley-Taylor PET (orange band) for the four sites between 1999 and 2008 (data from SILO).
Figure 8. Imagery and photos illustrating the four sites. The satellite images are Landsat TM false colour composites. Green indicates areas with high chlorophyll, purple areas low chlorophyll, dark blue and black areas indicate water; a + marks the spot.
5.3 Method

5.3.1 Processing of observed data

Eddy covariance $\lambda$E estimates as well as rainfall, air temperature, incoming shortwave radiation and net radiation were made available for the four sites by the principal investigators. Data were provided as half-hourly or hourly flux estimates. Some characteristics describing data quantity and quality are listed in Table 5. The data for Tumbarumba were almost entirely gap-filled, whereas data for the other sites was not gap-filled.

The $\lambda$E estimates were not directly converted to daily totals and compared to model estimated daily ET. There were several reasons for this:

- The measurement time series contained missing data that needed to be interpolated to achieve estimates of daily total ET.
- The sub-daily measurement interval and the simultaneous measurement of rainfall rates means that ET rates could be estimated separately for dry intervals and intervals with rainfall. This can help to separately compare ‘dry’ ET and ET during rainfall and identify some of the causes for differences found.
- The sub-daily measurement interval and the simultaneous measurement of incoming radiation means that daytime and night-time ET rates could be estimated separately. This can help assess the magnitude of night-time ET and its influence on the comparison.

All $\lambda$E estimates were first converted to E estimates (in mm) by division with the $\lambda$ calculated from simultaneous air temperature measurements (for equation see Appendix in AWRA Technical Report 3) and conversion from seconds to hour or half-hour. Daytime and night-time intervals were identified on the basis of net radiation, using a threshold of 5 W m$^{-2}$ for all sites except Tumbarumba, where a value of 10 W m$^{-2}$ produced a visually better separation.

To estimate daytime total ET for days with missing data, the total of measured E was calculated, as well as the summed net radiation for these intervals. Rainfall and night-time intervals were not included. If this radiation sum amounted to less than 60% of the net radiation, all data for this day was ignored and no estimate made. If the sum of ‘represented’ net radiation exceeded 60% of total net radiation, total daytime ET was estimated using the average evaporative fraction for the measured intervals.

To obtain an approximate estimate of the importance of night-time ET, the average (not gap-filled) night-time and (gap-filled) day-time ET rates were calculated. To obtain an estimate of the magnitude of wet canopy evaporation rates, the average ET rate for all rainfall intervals was calculated. To estimate rainfall interception losses during rainfall, the ratio of storm ET and average rainfall rate ($R$) was calculated.

6 The provision of data by Drs. R. Leuning (CSIRO, Canberra), J. Beringer (Monash University Melbourne), L. Hutley (Charles Darwin University, Darwin) and R. Pipunic (Melbourne University) is gratefully acknowledged.
Table 5. Summary of the quantity and quality of available flux tower data. Completeness was calculated for all days with (some) data. Energy balance closure estimates based on publication or communication with principal investigator.

<table>
<thead>
<tr>
<th>Site</th>
<th>HoSp</th>
<th>Kyem</th>
<th>Tumb</th>
<th>ViPa</th>
</tr>
</thead>
<tbody>
<tr>
<td>completeness</td>
<td>0.89</td>
<td>0.45</td>
<td>n/a</td>
<td>0.92</td>
</tr>
<tr>
<td>energy balance closure</td>
<td>89%</td>
<td>80%</td>
<td>&gt;90%</td>
<td>&gt;90%</td>
</tr>
<tr>
<td>days with (some) data</td>
<td>1285</td>
<td>323</td>
<td>2148</td>
<td>176</td>
</tr>
<tr>
<td>days accepted</td>
<td>1223</td>
<td>184</td>
<td>2148</td>
<td>176</td>
</tr>
<tr>
<td>months with &gt;10 days</td>
<td>45</td>
<td>9</td>
<td>49</td>
<td>20</td>
</tr>
</tbody>
</table>

5.3.2 Model run

Daily precipitation and atmospheric variables were calculated from the SILO gridded 0.05º climate products derived by interpolation of station data (Jeffrey et al. 2001).

The fraction area with deep-rooted vegetation was estimated from publications and satellite and on-ground images for each of the sites. Estimated deep-rooted cover fractions were 50% (HoSp), 20% (Kyem), 95% (Tumb) and 40% (ViPa). It is noted that these numbers should not be conceptualised as projected canopy cover, but as the lateral fraction of soil explored by deep-rooted vegetation.

The one-dimensional AWRA version was run with these data inputs and parameters as described in Section 5.2, for the periods 1985-2006. The years 1985-2000 were model warm up years and not used in the evaluation. Daily totals of both total ET and ‘dry’ ET (excluding rainfall interception losses) were used.

5.3.3 Comparison

Measured and modelled ET for all available days were compared using the metrics listed in Section 5.2. In evaluation, allowance needs to be made for the conceptual difference between storm ET (removed from the estimated flux tower ET) and modelled rainfall interception losses. While ET during storm will mostly represent evaporation of water from the wet canopy, there may still be a transpiration component. Conversely, and perhaps more importantly, after rainfall ceases there can still be considerable evaporation from the wet canopy for some hours after the storm. Common estimates of canopy storage capacity are in the order of a few millimetres (see AWRA Technical Report 3) and therefore this component can be a large component of total ET on days with rainfall, particularly if rain falls intermittently and during daytime. As a consequence, measured ET would be expected to exceed modelled ‘dry’ ET on such days.

Averages of measured daily ET were calculated for each month that had 10 days or more of data available. Using data for the same days, average modelled ET was also calculated. These average values are not necessarily good estimates of ‘real’ monthly averages, but the estimated and modelled quantities compared are mutually consistent.

To help in interpretation, an indicative ET budget was calculated at each of the sites. This was done by averaging the available monthly averaged measured and modelled ET estimates; again, the two quantities compared are for the same days, but neither should be interpreted as a best-estimate long-term ET budget. As part of this, measured night-time ET was estimated from the ratio of night-time and daytime ET rates, whereas storm ET was estimated by applying the average storm $E/R$ ratio to average daily rainfall.
The available site rainfall measurements offer an opportunity to evaluate the quality of gridded rainfall estimates for each site, and help to understand the role of rainfall estimation errors on model ET estimation. In addition, in the interpretation of ET differences it may be helpful to compare model simulated and satellite observed indicators of vegetation vigour. For that reason, Enhanced Vegetation Index (EVI) time series were calculated from MODIS reflectance measurements. Reflectance data were extracted for the 0.05° cells containing the flux tower location following the methods described in Guerschman et al. (2008).

5.4 Results

5.4.1 Components of average ET

Statistics of ET during rainfall and night-time conditions are listed in Table 6, as well as estimated ET components and corresponding modelled fluxes.

Table 6. Mean measured and estimated daily ET fluxes and indicators of the magnitude of ET during rainfall conditions and night-time.

<table>
<thead>
<tr>
<th>Site</th>
<th>HoSp</th>
<th>Kyem</th>
<th>Tumb</th>
<th>ViPa</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Storm ET</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N intervals</td>
<td>782</td>
<td>0</td>
<td>4319</td>
<td>150</td>
</tr>
<tr>
<td>mean $E$ rate (mm/h)</td>
<td>0.21</td>
<td>-</td>
<td>0.033</td>
<td>0.13</td>
</tr>
<tr>
<td>mean rainfall rate ($R$, mm/h)</td>
<td>4.18</td>
<td>-</td>
<td>1.25</td>
<td>2.36</td>
</tr>
<tr>
<td>$E/R$</td>
<td>5.0%</td>
<td>-</td>
<td>2.6%</td>
<td>5.7%</td>
</tr>
<tr>
<td><strong>Night-time $E$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$E$ (mm/h)</td>
<td>0.0014</td>
<td>0.0001</td>
<td>0.012</td>
<td>0.007</td>
</tr>
<tr>
<td>ratio night/daytime ET</td>
<td>5.0%</td>
<td>0.6%</td>
<td>5.8%</td>
<td>7.0%</td>
</tr>
<tr>
<td><strong>Estimated ET (mm/d)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>daytime dry ET</td>
<td>2.87</td>
<td>1.76</td>
<td>2.03</td>
<td>1.13</td>
</tr>
<tr>
<td>night-time dry ET</td>
<td>0.14</td>
<td>0.09#</td>
<td>0.12</td>
<td>0.08</td>
</tr>
<tr>
<td>storm ET</td>
<td>0.25</td>
<td>0.08#</td>
<td>0.06</td>
<td>0.05</td>
</tr>
<tr>
<td>total ET</td>
<td>3.27</td>
<td>1.92</td>
<td>2.21</td>
<td>1.26</td>
</tr>
<tr>
<td><strong>Modelled ET (mm/d)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dry canopy ET</td>
<td>2.49</td>
<td>1.55</td>
<td>1.82</td>
<td>1.22</td>
</tr>
<tr>
<td>Total ET</td>
<td>2.87</td>
<td>1.67</td>
<td>2.35</td>
<td>1.26</td>
</tr>
</tbody>
</table>

# could not be calculated from the data, 5% assumed based on the other sites.
The following observations are made:

- Night-time ET is estimated to amount to 5-7% of daytime dry-period ET.
- Rainfall period ET is estimated to evaporate 2.6–5.7% of rainfall; this includes both daytime and night-time intervals. Extrapolating this estimate suggests that average ET during rainfall periods is equivalent to 2.9–8.8% of the daytime ET rate that occurs under dry conditions.
- The uncertainty in estimated and modelled interception losses appears small for the two drier sites (Kyeamba, Virginia Park), but may be considerable for the two wetter sites. Estimated ET during rainfall and modelled interception losses are of similar magnitude for Howard Springs, but estimated values are lower than modelled values for Tumbarumba (it is reiterated that the two components are not conceptually equal, however.
- Modelled dry canopy ET is generally similar or lower than daytime dry ET. Modelled dry canopy ET is 15–17% lower than the sum of dry daytime and night-time ET for three sites, but almost equal (+1%) for Virginia Parks
- Modelled total ET is 12% lower than total estimated ET for Howard Springs, 13% lower for Kyeamba, 7% higher for Tumbarumba and equal for Virginia Parks.

### 5.4.2 Daily flux patterns

Estimates of total daily ET were not compared due to the uncertainty in estimating night-time and storm ET fluxes on a daily basis. Statistics comparing daily measured dry period ET and modelled dry canopy ET are listed in Table 7. Scatter plots of the daily data are shown in Figure 10.
Table 7. Results of comparison of measured and modelled “dry” ET.

<table>
<thead>
<tr>
<th>Site</th>
<th>HoSp</th>
<th>Kyem</th>
<th>Tumb</th>
<th>ViPa</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_o$ (mm/d)</td>
<td>2.87</td>
<td>1.76</td>
<td>2.03</td>
<td>1.13</td>
</tr>
<tr>
<td>$\mu_m$ (mm/d)</td>
<td>2.49</td>
<td>1.55</td>
<td>1.82</td>
<td>1.22</td>
</tr>
<tr>
<td>$\sigma_o$ (mm/d)</td>
<td>1.25</td>
<td>1.49</td>
<td>1.36</td>
<td>0.86</td>
</tr>
<tr>
<td>$\sigma_m$ (mm/d)</td>
<td>0.90</td>
<td>1.24</td>
<td>1.06</td>
<td>1.08</td>
</tr>
<tr>
<td>SD (mm/d)</td>
<td>0.90</td>
<td>0.58</td>
<td>0.72</td>
<td>0.79</td>
</tr>
<tr>
<td>absolute bias (mm/d)</td>
<td>-0.38</td>
<td>-0.20</td>
<td>-0.21</td>
<td>0.09</td>
</tr>
<tr>
<td>relative bias</td>
<td>-13%</td>
<td>-12%</td>
<td>-10%</td>
<td>8%</td>
</tr>
<tr>
<td>MRD</td>
<td>21%</td>
<td>23%</td>
<td>22%</td>
<td>43%</td>
</tr>
<tr>
<td>NSME</td>
<td>0.48</td>
<td>0.85</td>
<td>0.72</td>
<td>0.15</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.57</td>
<td>0.88</td>
<td>0.75</td>
<td>0.48</td>
</tr>
</tbody>
</table>

Figure 10. Comparison of daytime dry period ET derived from measurements and model estimated dry canopy ET. Blue dots represent days with more than 0.5 mm precipitation (as measured at the flux tower); orange dots days without precipitation. Regression lines shown are for all data combined.
The following observations are made:

- Differences between measured and modelled dry ET are noticeably larger at Virginia Park than for the other three sites. Also, unlike for the other sites the standard deviation in model estimates is higher than that in the measurements. The median relative difference (MRD) is 43% at Virginia Park and 21-23% at the other three sites.

- There is a positive bias in model estimates only at Virginia Park (+8%), model estimates are 10–13% lower than measured values at the other three sites.

- Figure 10 suggests that wet canopy evaporation after rainfall has ended explains some of the negative bias for Howard Springs, Kyeamba and Virginia Parks, but probably not for Tumbarumba; the ‘blue dots’ tend to be located above the 1:1 line.

- The $R^2$ is 0.48 for Virginia Park but higher (0.57–0.88) for the other three sites. SD is 0.58–0.90 mm/d; Virginia Park falls within this range.

- The regression equation for Virginia Park (Figure 10) suggests that the model may systematically underestimate ET under very dry conditions.

### 5.4.3 Monthly flux patterns

Statistics of a comparison between monthly average values of measured dry period ET are shown together with modelled dry canopy and total ET in Figure 11. Statistics of the comparison of measured dry period ET with modelled dry canopy ET are listed in Table 8 and for modelled total ET in Table 9. Results of a similar comparison using MODIS-based ET estimates reported by Guerschman et al. (2008) are listed in Table 10.

Figure 11. Comparison of monthly average “dry” ET derived from measurements and model results, respectively. The lower and upper limit of the orange band represents the modelled dry canopy ET and total ET (including wet canopy evaporation), respectively.
The following observations are made:

- The uncertainty associated with interception losses is small for Kyeamba and Virginia Park, as evident from the narrow range in model estimates in Table 10.

- For Howard Springs, measured monthly dry period ET is generally within the modelled range, but wet season measurements are frequently closer to total ET than to dry canopy ET (Figure 11).

- For Tumbarumba, measurements are generally within the modelled range, but are generally much closer to modelled dry canopy ET during wet periods, and higher than total ET for dry periods, as evident from the narrow range in modelled ET (Figure 11).

- For Kyeamba, there is very good agreement between model and measurements, although data was only available for nine months (Figure 11). For Virginia Park there is generally reasonably good agreement between modelled and measured ET, but measured ET appears generally higher than modelled values during dry periods (as evident from low ET) and there is a large mismatch for the last three months of measurement (Figure 11).

- The SD between monthly average measured dry ET and dry canopy ET is 0.35–0.63 mm/d. NSME is 0.50–0.93 and $R^2$ 0.70–0.96 (Table 8). Performance is generally poorest for Virginia Park.

- Comparison of Tables 8 and 10 shows that the performance for monthly average ET is better than values reported by Guerschman et al. (2008), with the exception of Virginia Park. The latter is explained by the poor agreement of the AWRA model estimates for the last three months of measurement.

---

7 Guerschman et al. (2008) used somewhat different interpolation and aggregation methods and used a differently processed earlier data set for Tumbarumba, which may have an influence.
5.4.4 Precipitation

Monthly accumulated rainfall as measured at the flux tower and as estimated from SILO and used in modelling are compared in Figure 12, whereas period average values are compared in Table 11. Data aggregation was performed the same way as for ET. This comparison is provided to assist in interpretation of the results.

Table 11. Average rainfall for the four sites based on monthly aggregate estimates, based on measurements at the tower and as derived from SILO respectively.

<table>
<thead>
<tr>
<th>Site</th>
<th>HoSp</th>
<th>Kyem</th>
<th>Tumb</th>
<th>ViPa</th>
</tr>
</thead>
<tbody>
<tr>
<td>tower (mm/d)</td>
<td>5.07</td>
<td>1.63</td>
<td>2.30</td>
<td>0.93</td>
</tr>
<tr>
<td>SILO (mm/d)</td>
<td>4.43</td>
<td>1.74</td>
<td>2.60</td>
<td>1.36</td>
</tr>
<tr>
<td>bias (mm/d)</td>
<td>-0.64</td>
<td>0.11</td>
<td>0.30</td>
<td>0.42</td>
</tr>
<tr>
<td>difference</td>
<td>-13%</td>
<td>7%</td>
<td>13%</td>
<td>46%</td>
</tr>
</tbody>
</table>

The following observations are made:

- There is considerable bias between tower rainfall measurements and SILO precipitation estimates. For Virginia Park, SILO average rainfall is 46% greater than rainfall measured at the tower, mainly due to a large overestimation of rainfall in December 2002 and January 2003. Biases are ±7–13% for the other three sites.

- This bias would be equivalent to ±42–235 mm per year if average bias in daily precipitation is multiplied by the number of days in an average year, and ±100–590 mm if the relative bias is applied to long-term annual average SILO rainfall (Table 4). These numbers are only provided for illustration: it is likely that this bias includes an effect of random differences, and therefore part of the bias would be reduced over a longer time integration period.
5.4.5 Enhanced Vegetation Index

Eight-day composite MODIS EVI patterns are compared with simulated EVI in Figure 13 to assist in interpretation of differences between modelled and measured ET.

![Graphs showing EVI comparison](image)

Figure 13. Comparison of the Enhanced Vegetation Index (EVI) as derived from the 8-day MODIS EVI product and as predicted by the model, respectively. Coefficient of correlation ($R$) is 0.52 for HoSp, 0.81 for Kyem, 0.18 for Tumb and 0.70 for ViPa.

The following observations are made:

- The approximate magnitude and temporal patterns in modelled and observed EVI appear to agree reasonably well.
- For Howard Springs, the onset of both greening and senescence is earlier than modelled. The observations also show the occurrence of burning in the season of most years, evident from a sudden drop in EVI. Some data gaps exist due to the near continuous cloud cover during the monsoon season.
- For Kyeamba, modelled and observed EVI agree rather well, although modelled values are generally somewhat lower than observed values.
- For Tumbarumba, the small seasonal variation in EVI is not captured at all by the model and indeed a small seasonal variation with opposite pattern is simulated.
- For Virginia Park, patterns generally agree reasonably well, but modelled dry season EVI is somewhat lower than observed values. Also note that the strong overestimation of EVI in early 2003.
5.5 Discussion

5.5.1 Overall agreement

The evaluation suggests that the model without extensive parameter calibration already provides realistic estimates of average ET and its component fluxes. Average total modelled ET and ET estimated from the measurements agreed within 0–13%. This is probably as good as could be expected given the sources of error both in the measurements and in the modelling (Section 5.1.2).

The performance in terms of SD and $R^2$ for monthly averages is as good as was expected in advance based on literature review (Section 5.1.3) and generally represent an improvement on the MODIS ET estimates produced by Guerschman et al. (2008). An exception was found for Virginia Park, where the model performed less well due to errors in the SILO rainfall input. This reiterates two important distinctions between the MODIS ET algorithm and the model: the model relies on the assumption that local precipitation is the only source of water for ET; and propagates any errors in precipitation forcing. The MODIS ET algorithm is much less affected by errors in rainfall estimates during wet periods, which are used to estimate the rainfall interception component, and not affected at all during dry periods. Conversely, the MODIS ET estimates are also not constrained by rainfall through a running water balance.

It is concluded that model ET estimates are likely to be more accurate where rainfall information is relatively accurate and precipitation the only source of water evaporated. The MODIS ET estimates may be more accurate where these conditions are not met.

5.5.2 Sources of error

The excellent agreement in mean total ET (0% difference) for Virginia Park is fortuitous: a negative bias for model estimates during dry periods was compensated by a positive bias towards the end of the time series that can be attributed to overestimates in the SILO rainfall forcing for two months. Conversely, the 12% difference for Kyeamba does not reflect the very good agreement in daily and monthly ET patterns. The negative bias in the model is almost entirely attributable to an overestimation of rainfall for November 2005 for this site (cf. Figure 11 and Figure 12).

The differences in estimated total ET for the other two sites may also be at least partially attributable to rainfall estimation errors: for Howard Springs, the bias in model ET of $-0.40$ mm d$^{-1}$ corresponded with a similar bias of $-0.64$ mm d$^{-1}$ in SILO rainfall; for Tumbarumba an ET bias of $+0.14$ mm d$^{-1}$ corresponded with a rainfall bias of $+0.30$ mm d$^{-1}$.

---

8 This may be a pessimistic assessment: another SILO interpolated rainfall data set appears to agree much better with precipitation recorded at Virginia Park (A. Frost, Bureau of Meteorology, pers. comm.). The reason for this discrepancy remains to be investigated.
5.5.3 Influence of night-time ET

The measurements suggested that night-time ET represented around 5% of daytime ET. This process is not explicitly represented in the AWRA model. Given the uncertainty in estimating effective daytime meteorological variables such as radiation, ground heat flux, humidity and wind speed (cf. AWRA Technical Report 3), this component can reasonably be assumed to be implicitly lumped into daytime ET within the model.

5.5.4 Influence of rainfall interception

Considering the components of ET, the greatest uncertainty was associated with rainfall interception at the wetter sites. For Howard Springs, storm ET and modelled wet canopy ET were of comparable magnitude, particularly when considering that part of the interception losses were included in dry period ET for the measurements (Figure 10). While this conceptual difference also exists for Tumbarumba, this does not appear to explain the large difference between measured and modelled ET. Model estimates of total ET for periods with rainfall were considerably higher than observations, and this difference appears mainly attributable to differences in estimated interception losses (cf. Figure 10 and Figure 11). Because total ET was underestimated during dry periods, overall average ET was still reproduced relatively well.

It is not clear whether the interception evaporation estimated by the model or that estimated from the observations is closer to the truth. It has been argued that eddy covariance technology is not suited to measure ET rates under rainfall conditions, although Czkowsky and Fitzjarrald (2009) appeared to be able to derive feasible estimates of interception loss for Amazonian rain forest. Storm ET rates were 0.13–0.21 mm h\(^{-1}\) for the Howard Springs and Virginia Park sites and ratios of eddy covariance storm ET over rainfall rate (\(E/R\)) 0.050–0.057. These values would appear reasonable when compared to values published in literature (see review in AWRA Technical Report 3). However for Tumbarumba a very low storm ET rate of 0.033 mm h\(^{-1}\) and E/R ratio of 0.026 were calculated. This coincides with a lower average rainfall intensity (1.25 mm h\(^{-1}\)); less available evaporative energy, particularly during the cool season; and perhaps more exposed and windier conditions. High interception losses that are not readily explained by Penman-Monteith theory were also reported in many other studies under similar conditions (see e.g. Roberts 1999). The mechanisms by which this occurs are still a matter of speculation, however (see AWRA Report 3 for some discussion).

5.6 Conclusions

It is concluded that measured and modelled ET generally agree rather well, with the relative and absolute agreement increasing as daily data are aggregated to monthly averages. The main source of error was found to be associated with the difference between tower based precipitation measurements and SILO grid-based estimates; it seems reasonable to assume that the tower based measurements are the more accurate ones where differences are large. An important source of uncertainty in the interpretation was associated with rainfall interception losses estimated from the measurements.

Overall, it would appear that given uncertainty in the model forcing and in aspects of the measurements, scope to improve performance with regards to flux tower ET by changes to the
model structure may be modest. A formal parameter sensitivity studies would be required to assess to what extent better parameter estimation can improve model performance. Instead, recommended research priorities are to explore approaches to improve the quality of rainfall information, and to obtain better quantitative understanding of the processes leading to rainfall interception loss.

An important caveat to this is that the model was not evaluated for sites that are known to rely on water inflows from other areas, such as water bodies, wetlands and irrigation areas. Realistic estimation of ET at such locations will require knowledge of the additional water inputs, or inference from satellite observations of for example greenness and land surface temperature using model-data assimilation approaches.

5.7 Recommendations

In summary, the following recommendations are made:

- The existing AWRA model and parameterisation appears to produce useful estimates of ET, provided that precipitation is the main source of evaporated water and provided that estimates of precipitation are of reasonable quality. Estimates are of useful accuracy at daily time step as well as aggregated over longer time scales.

- It is suggested that the model performance against ET measurements at the four flux towers documented here can serve as a benchmark for future model and/or parameterisation improvements as well as for alternative estimation approaches.

- Improving rainfall estimates is a likely precondition to improving ET estimates in areas where the current quality is modest (that is, more sparsely gauged regions). This may be achieved by blending gauge data with additional observations, for example satellite rainfall products, rainfall radar, and numerical weather predictions. Methods to do so are being developed through WIRADA.

- The mechanisms and sources of energy of rainfall interception are insufficiently understood to allow more definite interpretation of the modelled interception estimates. Process research is needed to clarify the causes for greater forest interception losses than would be expected on the basis of energy balance considerations alone. This research is not currently planned in WIRADA.

- Assimilation of satellite observed vegetation indicators (such as EVI) and perhaps land surface temperature are likely to improve ET estimation accuracy, particularly where rainfall estimates are of lesser quality, or where water inputs in addition to rainfall occur. This will require development of new model-data fusion approaches that address the current reliance on prior rainfall estimates in water balance modelling. Methods to do so are being developed through WIRADA.
6. **ASAR GM SOIL MOISTURE**

**Summary**

Satellite-derived top soil water content (SWC) estimates from the ASAR GM radar satellite instrument were made available by the Technical University of Vienna with support from the European Space Agency. The modelled SWC estimates agreed well with satellite top soil moisture content, although known sources of error in observations meant that in regions with dense vegetation and strong relief estimates could not be evaluated.

**6.1 Introduction**

**6.1.1 Justification**

Accurate estimation of soil water content (SWC) is of interest for several reasons. Firstly, accurate estimates (and forecasts) of relative soil water availability can be of interest for dry land farmers and support drought related policies. For example, the Australian Water Availability Project (2004-2007) produced a prototype soil moisture monitoring system to support the Exceptional Circumstances program\(^9\). Secondly, soil water storage is a term in the water accounts, and therefore needs to be estimated. Thirdly, the accurate simulation of soil wetness may potentially improve short-term streamflow forecasting. Finally, good agreement in the simulation of soil moisture provides another line of evidence to build trust in the estimation of other water balance terms.

Conventionally there have been two approaches to measure SWC, in the field using soil moisture probes and using satellite remote sensing. Several in-field techniques exist to measure the water content of a very small volume (normally <1 dm\(^3\)) of soil, with instruments that rely on the influence of SWC on electric resistivity, permittivity or the attenuation of neutrons (Robinson \textit{et al.} 2008; Vereecken \textit{et al.} 2008). Because of the local nature of the measurement and the potential for errors associated with poor soil contact and soil disturbance (all are invasive techniques) there is considerable uncertainty in scaling these measurements up to larger areas. The very recent development of non-invasive cosmic ray sensor technology (Zreda \textit{et al.} 2008), which in addition provides integrated estimates of SWC over larger areas (>200m laterally, a few dm vertically) is promising but observation records are not yet available for Australia\(^10\).

The SWC of the very top layer of soil (<10 cm) can also be inferred from satellite remote sensing with useful accuracy, at least in a relative sense. Evaluations of satellite-derived SWC estimates against \textit{in situ} measurements have been published (e.g. Reichle \textit{et al.} 2004; Wagner \textit{et al.} 2007; Draper \textit{et al.} 2009; Rüdiger \textit{et al.} 2009) and show that passive and active microwave measurements can both provide useful SWC estimates depending on land cover and terrain characteristics. Passive microwave remote sensing (radiometry) relies on the influence of soil moisture on microwave radiation emitted by the Earth’s surface at different frequencies and polarisations. Active microwave methods involve the use of scatterometers, which measure the


\(^{10}\) CSIRO has recently purchased 11 such sensors that will be deployed to sites across Australia to provide these measurements in future.
return of directional radar emissions. Both approaches require the use of so-called retrieval
models to account for influences other than soil moisture content. These can include vegetation
cover overhead, surface roughness, the angle between the direction of observation and the
earth’s surface, soil temperature, and atmospheric interferences. The importance of these factors
varies between active and passive methods, but is understood reasonably well and can partially
be accounted for.

The potentially beneficial use of remotely sensed soil moisture for land surface modelling has
been recognised, and experiments have been carried out to assimilate these observations in large
scale hydrological and weather forecasting models (e.g. Reichle and Koster 2005; Drusch
2007). Several studies have shown reasonable consistency between observed and model-
estimated soil moisture estimates, but the two data types often have different absolute ranges.

Most passive and active microwave instruments make frequent measurements over Australia at
course resolution (every 2 to 3 days, >25 km), and the earliest measurements were made in 1978
(Liu et al. 2009). The resolution of passive microwave methods is limited by the sensitivity and
signal-to-noise ratio (SNR) of the observing instruments. For radar methods the factor limiting
resolution is the ability to make repeated observations. An unusually high resolution active
microwave product available for some parts of the world, including Australia, is developed from
observations by the advanced synthetic aperture radar (ASAR) instrument on board the Envisat
satellite. This instrument can measure in three modes that represent a trade-off between spatial
resolution, temporal measurement frequency, and energy use. For practical reasons related to
operational mission design, the instrument measures most of Australia approximately every
three days at ~1 km resolution in its Global Mode (ASAR GM).

The Technical University of Vienna has developed a prototype ASAR GM soil moisture
product for Australia using the same method that was used to produce the more extensively
validated, coarser scale ERS soil moisture product (Wagner et al. 1999). While the ASAR
product is relatively new, it has already received some evaluation against field observations
(Pathe et al. 2009) and studies to date confirm it has similar accuracy and sources of error as the
courser scale ERS radar product, but with the added benefit of higher spatial resolution. The
ASAR GM product was made available to CSIRO by the Technical University of Vienna as part
of a collaborative research project, and was used for evaluation of AWRA simulated relative top
soil SWC.

6.1.2 Caveats

The absolute precision of satellite observed SWC is not well quantified and can vary in space
and time. Therefore, perhaps more so than with some comparatively more direct measurements
(e.g. streamflow) it is difficult to attribute differences between model estimates and observations
(see Section 1). Based on published experiences in comparing satellite-derived soil moisture
(SSM) with in situ and modelled values, the following common sources of difference can be
identified:

Model input: in particular, at continental scale large areas of Australia have rather poor spatial
rainfall data, and this has been shown to have a direct effect on the agreement in satellite and
model SWC, particular over daily time scales (see McCabe et al. 2008; Draper et al. 2009).

Model structure: inevitably the model makes simplifying assumptions about the processes and
properties governing soil SWC dynamics in the top soil. For example, in the current AWRA
model version soil evaporation draws on top soil SWC, but transpiration does not.

Model parameterisation: even if the model structure were perfect and all parameters had a
well-defined physical meaning, the knowledge and observations to estimate the parameters
across large areas is imperfect at best and this will lead to errors in model estimation. Of
particular importance in this context is the available water storage in the top soil, which for lack of better spatial information was estimated at 20 mm across the continent. Other parameters that are poorly known are those related to soil water evaporation rate and top soil drainage characteristics (see AWRA Technical Report 3).

Errors of representation. Published studies comparing modelled and satellite observed SWC studies demonstrate uncertainty due to the different depth intervals for which modelled and satellite-observed SWC are representative. It is commonly assumed that active and passive microwave techniques reflect conditions in the few top centimetres of soil (varying between instruments and wavelengths used), but a predictive method to accurately estimate the source depth does not exist. Due to the hydraulic coupling between top soil and deeper soil layers these shallow observations may still correlate well to deeper SWC (Van Dijk and De Jeu 2008). By comparison, the top soil layer in AWRA has a storage capacity of 20 mm water between field capacity and the point at which evaporation ceases. The corresponding soil depth depends on the pore size distribution, but would be expected to be in the order of perhaps 5 to 10 cm (see AWRA Technical Report 3).

Measurement precision. Because of the energy requirements of radar technology, there is a trade-off between spatial and temporal resolution and SNR, which can be improved by repeated measurement at the cost of greater energy use. ASAR GM reflects a choice for higher spatial and temporal resolution and lower SNR. The SNR can be improved by averaging over larger areas, over more than one overpass, or both.

Interpretation error. There are various sources of random and systematic error in ASAR GM derived SWC estimates, which are summarised well in Pathé et al. (2009). Importantly, ASAR GM and other synthetic aperture radar (SAR) measurement techniques are very sensitive to structure in the geometrical arrangement of radar scattering elements, including the roughness of the soil and the vegetation. Because this effect is stronger than that of SWC, radar cannot be used reliably to infer absolute SWC. However, change detection techniques can be used to estimate the relative wetness, on the assumption that surface roughness and vegetation change little or slowly. Furthermore, dense vegetation will intercept much of the energy emitted by the instrument, and therefore detected changes may not reflect changes in SWC. These sources of error are discussed further below.

6.1.1 Literature review

There have been several studies comparing radar derived SWC with in-situ and modelled estimates. In comparing ASAR GM and ERS data to in-situ observations for the same dates from the Oklahoma Mesonet, Pathé et al. (2009) found correlations (\(R\)) between ca. 0.15 and 0.85, with median values of around 0.65 for ASAR GM averaged to 50 km and ERS, and ca. 0.55 for ASAR GM averaged to 3 km. The correlation between 50-km ASAR GM and ERS was high (ca. 0.9) suggesting that the poorer agreement at higher resolution was mainly due to the reduced SNR.

Vischel et al. (2008) compared average ERS soil moisture for a catchment in South Africa with whole-of-soil SWC modelled with the TOPKAPI (for ‘TOPographic Kinematic Approximation and Integration’) model. It is noted that an exponential filter was applied to the ERS data, which will have had the effect of smoothing the signal and increasing the SNR, as well as empirically increasing the correlation with deeper soil moisture (Van Dijk and De Jeu 2008). Obtained \(R^2\) values were 0.76-0.92 within individual seasons for the entire 4625 km² catchment (in which case values from 3 footprints were averaged), and 0.68-0.88 for individual footprints.

Both studies also make it clear that (a) strong seasonal patterns in soil SWC enhance correlation; (b) temporal aggregation and smoothing emphasise these low frequency trends and
suppress high frequency variations associated with rainfall events, model error and measurement noise. In addition spatial averaging appears to increase the accuracy of the satellite estimates. Hence it is difficult to provide precise guidance on what might constitute an acceptable agreement.

The effect of vegetation and surface roughness varies as a function of incidence angle (the angle between the direction of view of the instrument and the earth surface). Based on this principle Pathe et al. (2009) developed a method to estimates the relative error in SWC. The relative error in ASAR GM SWC estimates over Australia estimated using this method is shown in Figure 14 (Doubkova, unpublished). Comparison with Figure 15 shows that the satellite estimates are expected to be of best quality over herbaceous or shrub vegetation on regolith (likely to generally coincide with low relief alluvial soils), and worst over dense forests and non-regolith materials (likely to generally coincide with rock outcrops).

Figure 14. Relative error in ASAR GM soil moisture estimates over Australia estimated using the method of Pathe et al. (2009) (source: Doubkova, unpublished).

Figure 15. (left) simplified geological map for mainland Australia (source: Doubkova, unpublished); (right) growth forms (source: Desert CRC).
The above can be summarised to provide the following prior expectations about the agreement between modelled and ASAR GM observed SWC:

- Only the relative agreement between modelled and observed SWC can be expected to be good, and it is less useful to compare absolute numbers. The relation between absolute numbers may contain information on surface roughness and vegetation.
- For comparison, the ca. 1 km resolution and 3-daily ASAR GM observations are best aggregated to a longer time step and/or coarser resolution increase the SNR of the satellite estimates.
- The agreement would be expected to be best in comparatively low relief, alluvial areas with sparse vegetation and worst in dense vegetation and higher relief terrain.
- The agreement would be expected to be best for areas with a clear seasonality in SWC, and least in areas that do not have pronounced seasonality, that is, areas that have very erratic or consistently high or low rainfall.

6.2 Methods

6.2.1 Observations used

The ASAR GM data used in this evaluation where provided by TU Vienna as part of a joint research project supported by the European Space Agency. The ASAR instrument is onboard the European Environmental Satellite Envisat, which was launched on 1 March 2002. It has a sun-synchronous orbit and a nominal repeat rate of 35 days, crossing the equator at 10:00 a.m. in descending mode. Further mission details are described in Pathe *et al.* (2009).

All ASAR modes except GM are acquired only on user request: GM data are obtained within the so-called background mission, which is active whenever no other data request has been placed by the ground control centre. As a consequence a small amount of data may be missing for dates where the instrument was used to collect higher resolution data, for example during flood events.

The SWC estimates used were produced using the change detection technique described in Pathe *et al.* (2009) to estimate relative SWC, then reprojected to regular grids. The data used here were acquired between 4 December 2004 and 25 January 2009.

It is noted that data are not available over Tasmania due to the ASAR mission design. Some additional areas with data that are considered consistently anomalous or representing surface water bodies are also masked out.

6.2.2 Model data used

The AWRA version 0.5 model was run with default parameters as described in Section 3. To provide a model input that was most similar to the satellite observed data, a model estimated relative ‘surface wetness’ or topsoil SWC \( (SWC_m) \) was calculated from simulated top soil water storage \( (S_0) \) in mm) and saturated fraction \( (f_{sat}) \):

\[
SWC_m = f_{sat} + (1 - f_{sat}) \frac{S_0}{S_{0\text{FC}}}
\]  

[6.1]
where \( S_{0FC} \) is top soil water storage at field capacity. The resulting value is scaled between the point of zero evaporation (\( SWC_m=0 \)) and field capacity (\( SWC_m=1 \)) and therefore not directly equivalent to volumetric water content, although there is a linear relationship between the two.

### 6.2.3 Comparison

Prior to the statistical analysis the observations and model estimates were spatially re-sampled to the smallest common denominator of 0.05° and a common spatial extent was applied to all data. Both data sets were temporally aggregated to MODIS-compliant 8-day median composites to increase the SNR. A median rather than mean filter was used to reduce the influence of incidental outliers in the data. Maps of all statistics listed in Section 2 were calculated for each grid cell (G. Warren, unpublished). Only those that appeared informative are reproduced here however.

### 6.3 Results

#### 6.3.1 Descriptive statistics

Maps showing the median soil moisture content as calculated from the observed and the modelled SWC are shown in Figure 16, whereas the standard deviation of the two data sets us shown in Figure 17.

![Figure 16](image-url)
The following conclusions are drawn:

- There is some qualitative spatial agreement in patterns; areas with higher median values or standard deviations in the observations also show higher values in the modelled SWC. Absolute values are quite different, however.

- The spatial range in median values is less for the ASAR observations (ranging from ~0.25 in dry regions to ~0.50 in humid regions) compared to modelled values (range ~0.10 to ~0.75).

- Spatial patterns in standard deviation show some differences. In particular, there appears to be a sampling artefact along a latitudinal band in central NSW; and higher standard deviations are found in some areas that appear to correspond to areas with greater estimated error (cf. Figure 14).

### 6.3.2 Bias

Maps showing the absolute and relative bias in SWC are shown in Figure 18. The following observations are made:

- The modelled values are up to ~0.4 higher for humid areas, and ~0.2 lower for arid areas. This corresponds with the greater spatial range in median modelled SWC (Figure 16).

- The relative bias shows a similar pattern, with modelled median values being almost twice observed values in humid areas and half in the arid interior.
6.3.3 Relative agreement

The coefficient of correlation $R$ is a measure of the degree of relative agreement between observed and modelled SWC. In this case $R$ was considered a more useful measure than the coefficient of determination $R^2$, as in principle a negative $R$ could still create a high $R^2$ value, even though patterns would not agree at all in such a case (that is, they would appear to be opposite). A map of $R$ values is shown in Figure 19. The slope and intercept of the corresponding linear regression of observed on modelled values is shown in Figure 20.
Figure 20. (left) slope and (right) intercept of the linear regression of ASAR GM on AWRA surface wetness. Missing or invalid data is shown in black.

The following observations are made:

- For most parts of Australia, $R$ values are greater than 0.50. Very low values ($R<0.10$) are found for salt lakes and (ephemeral) water bodies. Low values (0.10 to 0.40) are found for parts of the arid interior that appear to correspond with rock outcrops (cf. Figure 15).
- High $R$ values (>0.75) are found for drier and seasonally wet cropping and grazing areas, and medium $R$ values (0.50 to 0.75) are generally found for the regions that have tall vegetation and/or do not have strong rainfall seasonality.
- The regression relationship shows a near-zero or even negative slope for some of the areas with low $R$, as would be expected. For areas with comparatively the highest $R$, the slope of the regression appears to be around 0.50, with lower values for the remaining areas.
- The intercept of the regression equation is between zero and 0.20 for most of Australia, with higher values for regions with very low $R$.

### 6.3.4 Absolute agreement

The standard difference (SD) and NSME are measures of absolute agreement between observed and modelled SWC and maps of both quantities are shown in Figure 21.
The following observations are made:

- Both SD and NSME suggest overall poor absolute agreement. This is not surprising given the different mean and standard deviation of the two data sets.
- SD values are between 0.25 and almost 1. Lowest values (SD<0.25) are found in arid areas, and highest values in humid areas and areas with rock outcrops (cf. Figure 15; e.g. the Pilbara).
- NSME values are almost uniformly less than zero, which is attributed to the different distributions.

6.4 Discussion

6.4.1 Overall agreement

The comparison showed generally reasonable to good relative agreement in SWC patterns and poor agreement in absolute values. The scaling of the ASAR SWC is different from that of the modelled SWC, and therefore agreement in absolute values was not to be expected. There does appear to be some consistency in the slope and intercept of the regression relationship, or at least for areas where there is meaningful correlation. The regression equation for those areas with relatively high $R$ suggested that the dynamic range in the observed SWC is about half that of the modelled values, whereas the intercept of the equation suggests that under very dry conditions, when modelled SWC is reduced to zero, the observed SWC is on average 0.20.

The relative agreement between the two data sets, as expressed in $R$, was satisfactory for many regions. The literature review suggested that it is difficult to make a comparison between evaluations that aggregate the observations to different spatial and temporal time steps; particularly because random noise in the observations is significant and is reduced by aggregation. The results also confirm that higher $R$ values are to be expected in seasonally wet areas, and lower $R$ values in dry and humid areas that do not have a clear seasonality.
6.4.2 Causes for differences

Spatial differences in the agreement between observed and modelled SWC strongly support the importance of the observation error sources that were anticipated. Agreement was comparatively better for regions with alluvial soils with low vegetation density, and poorer for regions with rock outcrops or dense vegetation. Superimposed onto this is the effect of a seasonal soil moisture cycle in strengthening the correlation. The effect of observational errors can be estimated from the residual variance of the regression equation. This was done in a separate study (Doubkova, unpublished) but using daily values and at higher spatial resolution. The result of this is shown in Figure 22, along with the errors that were estimated a priori (same as Figure 14).

Figure 22. (left) Calculated root mean square “error” (RMSE) between daily ASAR and AWRA SWC values; (right) prior estimates of the error in ASAR estimates based as derived directly from the observations.

The agreement in spatial patterns between the two maps is encouraging and reinforces that most of the residual difference between bias-corrected observed and modelled SWC can be attributed to observational errors. Some differences can be identified however; for example calculated differences for wheat growing areas in Western and South Australia are greater than errors estimated a priori. It is speculated that this may be related to the effects of soil tillage on surface roughness.

Given that normalised errors in the observed ASAR appear larger than those in modelled values, this evaluation does not provide a good test of the accuracy of model SWC. Further assessment is required, comparing both modelled and observation-based SWC to in-situ measurements and other observation-based estimates, in particular from passive microwave remote sensing.

There have been several studies to assess the feasibility of assimilating satellite soil moisture observations to improve soil water balance estimates. Potentially, rainfall estimates can be improved at the same time (Crow et al. 2009), particular where gauging is sparse. The spatially varying error in ASR estimates poses a challenge in model-data assimilation, and therefore the demonstrated effectiveness of the method of Pathe et al. (2009) to predict the relative magnitude of errors is an important step. Another promising method to provide useful estimates of the error structure in ASAR and modelled SWC is the triple collocation method, which requires a third data set with entirely uncorrelated errors. Scipal et al. (2009) applied this method in a very similar context, using SWC estimates derived from passive microwave observations as the third data set. While it is unlikely that errors between radar and passive microwave methods are entirely uncorrelated (for example, both suffer from vegetation density effects) the importance
of difference error sources does seem to vary between the two approaches and therefore it may be valid by good approximation. The results obtained by Scipal et al. at least suggest that the approach has some promise.

A promising pathway to not only quantify, but also reduce SWC observational error is by blending active and passive microwave observations. This is an important objective of the planned NASA SMAP mission\(^\text{11}\), but it should also be possible to achieve this using independent active and passive microwave platforms. It will require a method to reconcile the achievable ASAR product resolution with the much coarser spatial resolution (>25 km) of passive microwave SWC products.

### 6.5 Conclusions

It is concluded that measured and modelled top soil SWC appear to as well as could be expected given the observational errors and the different scaling of the two estimates. The main source of difference was found to be the errors in the observations associated with their sensitivity to surface roughness and vegetation density.

The comparison suggests that the existing AWRA model and parameterisation provide useful estimates of top soil SWC. The analysis was only performed over 8-day periods and 0.05° resolution. However, a separate analysis (Doubkova, unpublished) suggests that these findings also apply to daily observations and higher spatial resolution, although noise in the observations increases.

Neither the observations nor the modelled values are expressed in volumetric water content and therefore no comment can be made about the accuracy achievable in estimating this quantity. Most likely, this will primarily depend on the ability to estimate top soil water retention characteristics over larger areas. With the possible exception of certain locations or small regions that have undergone more intensive data collections, such estimates are not likely to be particularly accurate. Perhaps more important however is whether such volumetric estimates actually would be more useful. It is not obvious what purpose would require SWC to be accurate in volumetric terms rather than in absolute (mm) or relative terms.

This comparison does not provide clear guidance as to how the model could be improved. Given the shallow depth of the layer considered, it might be expected that the greatest model errors would be associated with errors in rainfall forcing, at least where gauging is sparse.

### 6.6 Recommendations

The following recommendations are made:

- The modelled SWC estimates appear to agree with ASAR GM derived values as well as can be expected given the nature of the observations. The quality of model estimates under dense vegetation and in areas with relief could not be evaluated.

- It is recommended to further assess the performance of SWC estimation by the model by further comparison to in-situ measurements and other observation-based estimates, in particular from passive microwave remote sensing.

- The quality of rainfall estimates may well be the main constraint on accurate top soil moisture estimation. It is recommended to develop methods to quantity and reduce the

\(^\text{11}\) http://smap.jpl.nasa.gov/
error in satellite soil moisture estimates, for example, through data fusion using both passive and active remote sensing methods, and to develop methods to assimilate these observations in modelling. Methods to do so are being developed through WIRADA.
7. AVHRR/MODIS VEGETATION PROPERTIES

Summary

Satellite-derived vegetation properties describing vegetation canopy cover fraction, density and chlorophyll content were derived from the AVHRR and MODIS optical satellite instruments. Simulation of these vegetation properties is not an model objective by itself, but accurate representation of vegetation dynamics can be a prerequisite to accurate estimation of ET fluxes. The AWRA model reproduced the response of grasses, herbs and crops to temporal patterns in water availability well. The effects of temperature limitation at high elevations and the generally small variations in forest canopy properties were not reproduced. Where the model did not perform well, the implications for water balance estimation may still be modest. A possible exception is the estimation of rainfall interception losses in high elevation areas. It is recommended to investigate opportunities to improve rainfall estimates and/or quantify lateral water influxes by assimilation of satellite derived vegetation properties into the model.

7.1 Introduction

7.1.1 Justification

The AWRA system dynamically simulates several vegetation properties. Deriving information on these vegetation properties is not an objective by itself, but accurate representation of vegetation dynamics can be a prerequisite to accurate estimation of ET fluxes. The relationship between vegetation properties and ET is particularly strong in water limited environments, and this is the basis for a category of ET estimation algorithms that relies on satellite-observed vegetation properties (e.g. Cleugh et al. 2007; Mu et al. 2007; Guerschman et al. 2009b).

Long-term average vegetation properties and regular seasonal cycles can be calculated directly from remote sensing products and in principle do not need to be simulated. However water availability often varies from one year to the next, and will not be captured well in the average behaviour. Therefore, a predictive vegetation phenology model is likely to be advantageous for periods when satellite observations are not available. The vegetation phenology model structure is described in AWRA Technical Report 3.

In conventional soil vegetation atmosphere transfer models, the variable leaf area index (LAI) is used to describe the dynamic influence of vegetation density on the surface water and energy balance, with linear or hyperbolic relationships to estimate associated properties such as canopy cover and surface conductance (e.g. Zhang and Dawes 1998). LAI is commonly defined as the one-sided area of living leaves per unit ground area, although this definition is not without its semantic challenges (Barclay 1998; Breda 2003). The emphasis on LAI as the primary descriptor is probably because it was comparatively the more easily determined before the advent of optical field and remote sensing methods (Glenn et al. 2008). Many spatial land surface models still use LAI as a driving variable, and infer parameter values from remote sensing. However LAI is not directly observed through remote sensing, but rather needs to be estimated from calculated vegetation indices (VIs), introducing various assumptions, errors and uncertainties along the way (Glenn et al. 2008).
Several studies have found that in fact many of the properties and fluxes of interest are better related to remotely sensed VIs than they are to LAI (Glenn et al. 2008). For example, light absorption is better estimated from remotely sensed albedo, whereas canopy cover, the fraction of absorbed photosynthetically active $^{12}$ radiation (FPAR), and leaf assimilation capacity can be estimated rather well from the Normalised Difference Vegetation Index (NDVI) and the Enhanced Vegetation Index (EVI).

In this section, model estimates of FPAR, EVI, fraction canopy cover and LAI are compared to values calculated or estimated from observations by the Advanced Very High Resolution Radiometer (AVHRR) and the Moderate Resolution Imaging Spectroradiometer (MODIS) satellite instruments. The AVHRR instrument has been operating from a series of NOAA polar satellites providing a near-continuous record since June 1981. The MODIS instrument has been operating from a pair of scientific NASA satellites (Terra and Aqua) since February 2000. Although strictly speaking a research mission, observations by MODIS are used in several operational applications including, for example, Australia’s Sentinel Hotspots national bushfire monitoring system.

The use of satellite vegetation observation in developing, calibrating and evaluating models of vegetation phenology is a recent development and there are few published studies. Some likely reasons include the relatively recent appreciation of vegetation phenology in governing carbon fluxes and its potential to provide feedbacks in global warming scenarios; the limited accessibility of remotely sensed vegetation products; and the challenge in interpreting differences between model and observations. The few studies that are published have primarily used satellite vegetation indices such as NDVI and FPAR to evaluate winter/summer phenology of higher latitude ecosystems. Bondeau et al. (1999) compared seasonal FPAR patterns simulated by seven prognostic large-scale vegetation models to AVHRR observations, and also included an African transect. They conclude that the models tend to reproduce rain green phenology poorly and overestimate the length of the growing season in savannah ecosystems. Stöckli et al. (2008b) review the representation of phenology in several state-of-the-art vegetation models and compare model LAI to MODIS LAI (collection 4). They conclude that the models have little prognostic value in general and even less so for water availability driven phenology. They overcame this to some extent by assimilating MODIS FPAR and LAI products into their model. Model-data assimilation (MDA) experiments were also done using MODIS EVI and an earlier AWRA-L version (Van Dijk and Renzullo 2009). This suggested that AWRA-L reproduced seasonal patterns reasonably well for shallow-rooted vegetation. It also demonstrated that MDA slightly improved water balance agreement with passive microwave observations, but had little effect on the agreement with flux tower ET observations.

### 7.1.2 Caveats

The data sets used in this evaluation are described in some more detail in Section 7.1.3, along with published information on their accuracy against field measurements. Some additional sources of difference between model estimated and satellite derived vegetation properties can be identified as follows:

**Observation or retrieval model.** Probably the most important source of uncertainty is associated with the indirect relationship between satellite observations and the vegetation property of interest. AWRA-L model simulates vegetation properties that need to be converted

---

12 usually equated to radiation in 0.4-0.7 μm wavelengths
13 USA National Oceanic and Atmospheric Administration (http://noaasis.noaa.gov/NOAASIS/ml/avhrr.html)
14 USA National Aeronautics and Space Administration (http://modis.gsfc.nasa.gov/)
to satellite VIs using so-called observational models. Conversely, several satellite products representing physiological vegetation properties, such as LAI and fraction canopy cover, are derived using a so-called satellite retrieval model. The relationship with model estimated quantities is more direct for these products, but the uncertainties are not necessarily less – effectively the retrieval model is equivalent to an inverted form of the observation model, and therefore suffers from similar errors and uncertainties. Both observation and retrieval models require simplifying assumptions and parameter estimation, and this introduces additional errors that can potentially make it harder to determine whether differences between the ‘observations’ (that is, the satellite product) and the modelled quantities should be primarily be attributed to the measurements, to the observation or retrieval model, or to the biophysical model itself. This is discussed in some more detail further on.

**Model input.** Over large parts of Australia the vegetation responds primarily to water availability. At the same time spatial rainfall data for many parts of Australia is rather poor, and this can lead to errors in estimated vegetation patterns, particularly for ephemeral vegetation that relies on isolated storm periods (for example, in arid areas).

**Model structure.** The AWRA model makes simplifying assumptions about the way that vegetation responds to environmental factors. Importantly, the current model version only considers the influence of water availability on vegetation density. This may be a reasonable assumption for much of Australia, but temperature, radiation and nutrient availability can be more important controls in some areas, particularly at higher elevations and latitudes.

**Model parameterisation.** Even if the model structure were perfect and all parameters had a well-defined physical meaning, the lack of knowledge and observations to estimate these parameter values would be imperfect and therefore will lead to differences. For example, only two vegetation types are recognised (shallow- and deep-rooted vegetation) and each has a spatially constant set of parameters. Spatial constants are also used to convert between leaf canopy cover and remotely sensed vegetation indices, without considering the spatially and temporally varying effect of, for example, soil background colour.

A further caveat to consider is that one of the data sets (AVHRR FPAR) was in fact used to estimate the fraction deep-rooted vegetation for each 0.05° grid cell. This was done by assuming this fraction to be equal to the minimum FPAR in the 1980–2006 time series (Figure 1). This has two consequences. Firstly, a degree of spatial agreement in modelled and observed FPAR patterns is to be expected given the observations have already been used in parameterisation and therefore should not be considered an indicator of good model performance.

Secondly, the assumption of a direct relationship between minimum FPAR and fraction deep-rooted vegetation will introduce errors. The assumption was made because other continuous land cover fraction data sets for Australia were not yet available or assessed at the time of the model run. The assumption made is expected to over-estimate the fraction of deep-rooted vegetation in consistently wet regions, which may be dominated by shallow-rooted vegetation but still show a high minimum FPAR (for example, wetlands, coastal plains). It is further expected that the approach followed may underestimate the vegetation density and cover of deep-rooted vegetation in areas that are resource limited. For example, trees in open woodland may fully explore the soil, but find too little water to support a continuously closed canopy. By way of calculation example, FPAR may be 0.50 in this case, leading to the model assumption that 50% of the grid cell is covered by deep-rooted vegetation. If the modelled FPAR of the two vegetation components were otherwise estimated accurately, this would produce an underestimation of FPAR for the grid cell when all shallow-rooted vegetation has senesced: FPAR for the shallow- and deep-rooted fractions would be accurately estimated at 0 and 0.50, respectively, but the area-weighted FPAR consequently inaccurately estimated as 0.25. In summary, overestimation of the mean and variability of vegetation density is to be expected for areas where deep-rooted vegetation do not always achieve a closed canopy.
7.1.3 Data sets used and caveats

The data sets used are described below, along with some known issues that are relevant to the interpretation of the model evaluation.

**AVHRR FPAR**

Donohue et al. (2008) developed a monthly FPAR time series product for Australia for the entire AVHRR record from July 1981 onwards, at a resolution that varies from 0.081° to 0.011° over time. Observations were spectrally calibrated using a so-called ‘invariant-cover-triangle’ method (Donohue et al. 2007), the main assumption of which is that (a) changes in the relationship between red and near infrared (NIR) reflectance are due to vegetation change, and (b) the influence of soil colour does not vary over time. From the calibrated reflectance NDVI can be calculated as (Rouse et al. 1974):

\[
NDVI = \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED}}
\]

where \( \rho_{NIR} \) and \( \rho_{RED} \) are the reflectance in the NIR and red bands, respectively. To calculate FPAR, equivalence with NDVI was assumed for NDVI in the range between 0 and 0.95, and set to this minimum and maximum value, respectively, for NDVI values outside the range. There has not been any field validation of the FPAR values produced by this method and therefore its accuracy is unknown. Errors can originate both from the precision and accuracy of reflectance measurement and from the assumed relationship between NDVI and FPAR.

Known issues with regards to the accuracy of AVHRR reflectance measurements include sensor calibration degradation (the instrument does not have on-board calibration), the effect of water vapour on the comparatively sensitive NIR measurement, and some other atmospheric effects. Analysis of the results for overlapping AVHRR periods suggest that the calibrated data have a SD of ±0.027 FPAR units (Donohue et al. 2007). Some general conclusions about the accuracy of other similar AVHRR NDVI products can be derived from the literature. For example, a comparison of long-term trends calculated from differently processed AVHRR NDVI data sets over the Iberian peninsula showed considerable differences, even between the two newest data sets that were considered to have had the highest quality processing (Alcaraz-Segura et al. 2010). A review of published analyses of the agreement between AVHRR and MODIS NDVI (Ji et al. 2008) suggests that correlations are generally reasonably high \((R^2>0.7)\) with average differences mostly in the order of 0.05-0.10 NDVI units, but considerable spatial and temporal differences can occur. AVHRR NDVI values are usually found to have a smaller dynamic range and be somewhat lower than MODIS NDVI.

Additional errors are introduced when estimating FPAR from NDVI. In particular, while there generally appears to be a strong (near) linear relationship between NDVI and FPAR for a given vegetation type, the constant of proportionality can vary considerably (Glenn et al. 2008). NDVI is primarily a measure of the ‘greenness’ or chlorophyll concentration of the canopy. As a consequence, it does not account for the absorption of PAR by other components of the canopy, such as branches and dead leaves. This conceptual difference with ‘real’ PAR is not a problem for evaluation, as long as FPAR is compared with modelled fractional cover of living leaves. However NDVI is also affected by the actual concentration of chlorophyll in living leaves and by the reflective properties of the soil surface, which leads to some conceptual inconsistencies with model estimated leaf canopy cover.
MODIS LAI/FPAR

The MODIS global LAI/FPAR product (short name MOD15A2) is composited every 8 days at 1-kilometer resolution. The method considers spectral information as a function of sensor geometry\textsuperscript{16}. The algorithm uses observations obtained under up to seven different geometries and produces the most probable values for pixel LAI and FPAR as well as information on uncertainty (Knyazikhin et al. 1999; Myneni et al. 2002). A look-up-table is used to invert the three-dimensional observation model. When the method fails to produce a solution, a back-up method is used that estimates LAI and FPAR directly from NDVI, with relationships that vary between different biomes represented in the MODIS land cover product. The NDVI is therefore only directly used to estimate FPAR and LAI when the primary method fails. Correlation is reasonably strong, however: comparison between NDVI and retrieved FPAR shows an approximately linear relationship but with an intercept, such that FPAR is zero at NDVI around 0.10. The relationship between LAI and NDVI conforms to an asymptotic hyperbole, with a change point at an LAI of around 2 (Myneni et al. 2002). The MODIS LAI/FPAR product receives regular modifications; in the current study the collection 5 (c5) data were used. Some differences with previous versions include algorithm refinements to improve data quality for woody vegetation, and use of an improved MODIS land cover product that distinguishes eight instead of six biomes.

Atmospheric influences on retrieved FPAR and LAI should be better accounted for in MODIS than in AVHRR based products. However, in estimating FPAR assumptions about the equivalence between NDVI and FPAR are replaced by assumptions about leaf reflective properties for each of the biomes considered. The accuracy of the FPAR and LAI product has been assessed at various locations via ground-truth and validation efforts as well as several volunteered assessments published in the literature. Most published comparisons are for the collection 4 (c4) product or earlier versions. This includes a comparison against collated field LAI measurements in Australia that found that the MODIS c4 product appeared to provide reasonable estimates for most land cover types, but that strong overestimation occurred for some open forests and woodlands in eastern Australia (Hill et al. 2006). Fuentes et al. (2008) compared MODIS c4 LAI with values estimated from hemispherical photographs for several West Australian sites and found that MODIS LAI overestimated field LAI reasonably consistently by ca. 0.3 LAI units. Comparison of MODIS FPAR and LAI for water limited ecosystems in Senegal and Zambia also indicated an overestimation of field values (Fensholt et al. 2004; Huemmrich et al. 2005). FPAR was most severely overestimated at very low values (by as much as 0.3–0.4 units) but converged to within ~0.1 at FPAR greater than 0.4. MODIS LAI appeared to show a consistent offset of 0.2–0.5 units. Documentation associated with the c5 LAI product indicates that recent modifications have reduced the overestimation at the Zambian site\textsuperscript{17}, although it is unclear to what extent this may also be true for Australian woodlands.

The comparison of satellite derived FPAR and LAI with field-measured values does have some conceptual problems. For short vegetation FPAR can accurately be measured as the difference between incident and reflected PAR, but this is often impractical for tall vegetation. In those circumstances, FPAR is commonly assumed to be equivalent to intercepted PAR (the difference between incoming and transmitted PAR measured below the canopy). This generally appears a reasonable assumption, although the two quantities may vary by 0.05 units or so (Huemmrich et al. 2005). Destructive methods to measure LAI are arguably the most accurate but also the most laborious and impractical for tall vegetation types. Instead LAI estimates are commonly inferred

\textsuperscript{16} the arrangement of the earth surface, direction of sun light, and the instruments viewing direction

\textsuperscript{17} https://lpdaac.usgs.gov/lpdaac/products/modis_products_table/leaf_area_index_fraction_of_photosynthetically_active_radiation/8_day_l4_global_1km/v5/terra
from sunlight interception, for example using hemispherical photos or light interception measurements. This introduces uncertainties due to the effects of leaf clumping, dependency on viewing angle, and light interception by non-leaf elements such as branches and stems.

**MODIS fraction photosynthetic vegetation**

Using processed MODIS reflectance data, Guerschman et al. (2009a) estimated fractions of bare soil (FBS), non-photosynthetic vegetation (FNPV; for example, litter, branches) and fraction photosynthetic vegetation (FPV) across Australia. The approach involved calculation of NDVI as well as a Cellulose Absorption Index. End members (that is, values for pixels with 100% cover of the respective classes) were empirically estimated for each of the three cover types and these were used to estimate the respective fractions.

Because the AWRA version 0.5 model does not estimate FBS and FNPV separately these fractions were not considered in evaluation, however the FPV estimates can be compared. Conceptually, FPV would seem an improvement on FPAR for the purpose of evaluation, since it explicitly aims to estimate the spatial fraction of living leaves. In practice, however, the manner by which it is calculated is very similar to the formulation of FPAR and therefore the same caveats apply. The main difference with a simple estimation from NDVI is that the method makes some allowance for soil background: for bare soil conditions, NDVI is scaled between 0.100 and 0.800 to estimate FPV; whereas for conditions with fully (litter) covered soil, NDVI is scaled between 0.175 and 0.800. It has not yet been tested whether these changes lead to an improvement in FPV estimates. Comparison with FPV measurements at ten grassland sites across Australia suggested a standard error of 0.05–0.54 units, although the higher errors were more likely due to problems in scaling from site to pixel resolution (Guerschman et al. 2009a).

**MODIS EVI**

The MODIS Vegetation Indices product (short name MOD13Q1) includes two vegetation indices: NDVI and the Enhanced Vegetation Index (EVI; Huete et al. 2002). It is calculated from NIR, red and blue band reflectance ($\rho_{NIR}$, $\rho_{RED}$ and $\rho_{BLUE}$) as:

$$EVI = 2.5 \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + 6\rho_{RED} - 7.5\rho_{BLUE} + 1}$$

EVI is an improvement on NDVI in that it reduces soil background effects while increasing sensitivity over dense vegetation conditions: the relationship between NDVI appears to become insensitive to canopy density greater than an LAI of around 2, whereas EVI is sensitive up to an LAI of perhaps 4 (Huete et al. 2002). The EVI also uses the blue band to remove residual atmosphere contamination caused by smoke and sub-pixel thin cloud. MOD13Q1 data are provided every 16 days at 250-meter spatial resolution. Because blue reflectance measurements are only available at 500m resolution, these are used to correct for residual atmospheric effects, with apparently negligible negative impact\(^{18}\). The collection 5 EVI product was used here, but modifications from earlier products appear to be marginal.

Since EVI is a radiometric index, the only sources of error are instrument precision and the correction for atmospheric effects and viewing angle. Agreement with field measurements has generally been very good, with MODIS and field EVI agreeing with 0.05 units (Huete et al. 2002).

\(^{18}\) https://lpdaac.usgs.gov/lpdaac/products/modis_products_table/vegetation_indices/16_day_l3_global_250m/v5/terra
7.1.4 Summary

The previous sections are summarised as follows:

- Few studies have evaluated model estimated vegetation properties in space and in time against satellite-derived vegetation properties. Those that have typically find little predictive skill for rain deciduous vegetation.

- Due to the approach to parameter estimation, it is to be expected that the model will overestimate the mean and variability of vegetation density in areas where deep-rooted vegetation (for example, trees) cannot achieve a closed canopy.

- Apart from the ability of the biophysical model to accurately estimate biophysical measures (for example, LAI and canopy cover fraction) additional errors will be introduced by the need to predict their relationship to quantities observed by the satellite (e.g. NDVI and EVI).

- Published field validation studies suggest that there is usually a reasonably strong and near-linear relationship between NDVI, APAR and intercepted PAR. In turn, intercepted PAR is closely related to projected canopy cover. The accuracy of the MODIS FPV product is not well known, but a working assumption may be that it is likely to have similar characteristics as the MODIS FPAR product.

- Additional uncertainty may result from inconsistencies between the formal definition of FPAR and LAI on one hand, and other influences that may affect estimated values, in ways that depend on the measurement or model. Such influences include non-photosynthetic canopy elements and varying leaf chlorophyll concentration.

- As a result, while there are often reasonably strong to strong relationships between NDVI, FPAR and LAI, the parameters of this relationship can vary between vegetation types. Under such circumstances, a good relative agreement between modelled and satellite-derived properties might still occur, but good absolute agreement would not necessarily be expected.

7.2 Methods

7.2.1 Observations used

The products used were described in Section 7.1.3; some further details of the actual data files used in this evaluation are provided in Table 12. The MODIS data were downloaded from NASA’s Land Processed Distributed Active Archive Centre (LPDAAC) and mosaiced and remapped for Australia as described by Paget and King (2008). All data sets are also publicly available from the CSIRO web site19

<table>
<thead>
<tr>
<th>sensor variable</th>
<th>AVHRR</th>
<th>MODIS</th>
<th>MODIS</th>
<th>MODIS</th>
<th>MODIS</th>
<th>MODIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>data product</td>
<td>PAL*</td>
<td>MCD43A4*</td>
<td>MOD15A2, c5</td>
<td>MOD15A2, c5</td>
<td>MOD13Q1, c5</td>
<td></td>
</tr>
</tbody>
</table>

19 http://wron.net.au/data/
7.2.2 Model data used

The AWRA version 0.5 model was run with default parameters as described in Section 3. The model provides estimates of LAI and leaf canopy cover fraction using methods described in AWRA Technical Report 3. In this comparison, the following observation models are used to estimate grid cell average vegetation properties from the model results (subscript $M$):

$$\text{LAI}_M = \sum f_i \text{LAI}_i \tag{7.3}$$

$$\text{FPAR}_M = \sum f_i f_{v,i} \tag{7.4}$$

$$\text{FPV}_M = \text{FPAR}_M \tag{7.5}$$

$$\text{EVI}_M = \sum f_i (f_{v,i} \text{PCI}_{v,i} + \text{EVI}_0) \tag{7.6}$$

where $f_i$ is the fraction of a pixel covered by the two respective hydrological response units (that is, tall deep-rooted and short shallow-rooted vegetation); $\text{LAI}_i$ the LAI, $f_{v,i}$ the vegetation canopy cover fraction and $\text{PCI}_{v,i}$ the specific photosynthetic capacity index for each HRU; and $\text{EVI}_0$ is a background EVI for which a uniform estimate of 0.07 was used.

7.2.3 Comparison

Prior to the statistical analysis each pair of satellite observations and the model estimates was spatially and temporally resampled to the smallest common denominator, and a common spatial extent was applied. All AVHRR and MODIS data were aggregated to the model resolution (0.05°) and the model results were temporally averaged to the time step of the satellite data sets (8 days to a month). For each grid cell, all statistics described in Section 2 were calculated, producing maps of the statistics (G. Warren, unpublished). Only those that appeared informative are reproduced here. To assist in interpretation, a map of major vegetation types across Australia is provided in Figure 23 and a map of soils in Figure 24.
Figure 23. Major vegetation types in Australia in 1988 (source: Williams et al. 2001)

Figure 24. Soils map (following the Australian Soil Classification; source: McKenzie et al. 2004)
7.3 Results

7.3.1 AVHRR FPAR

Maps showing statistics of the distribution of observed and modelled values are shown in Figure 25, as well as the absolute and relative bias. The following observations are made:

- The mean observed and modelled FPAR show similar spatial patterns, varying from <0.25 units in arid regions to >0.75 in wet regions. Spatial gradients appear stronger in the modelled FPAR than in the AVHRR FPAR however.
- The spatial pattern in standard deviation in modelled and AVHRR FPAR are similar and between 0.05 and 0.40 for most of the continent. Model standard deviation is greater in the open woodland and scrubland of Northern and West Australia.
- Compared to AVHRR FPAR, modelled FPAR shows the greatest bias for parts of the highlands of Tasmania (more than +0.25 units). Modelled values are greater by 0.10–0.25 units for higher altitude regions in SE Australia and some of the woodlands and open woodlands of Northern Australia.
- Mean modelled FPAR is lower by 0.10–0.25 units for the woodland region of south central WA, woodlands of SE Queensland, and the floodplain and irrigation areas of the Murray-Darling Basin.
- Mean modelled FPAR is lower by up to ~0.10 for most parts of arid inland Australia. Because of the low mean FPAR this leads to large negative bias in the modelled FPAR (less than ~50%).
- The greatest relative positive bias (more than +50%) occurs on the drier open woodland regions of Northern Australia, for lakes and salt lakes, and for the high elevations parts of Tasmania.

Maps showing statistics of the agreement in temporal patterns are shown in Figure 26. The following observations are made:

- The standard error (SD) is generally around 0.15, with lower values (~0.05) in the arid interior and >0.40 in Northern Australia and the higher altitude regions and floodplains of Eastern Australia and Tasmania.
- The coefficient of correlation is good ($R>0.75$) for all pasture and cropping regions and most of grasslands of Queensland and NT, with the apparent exception of the ranges (for example McDonnell Ranges; cf. Figure 15).
- Correlation is very poor ($R<0$) for all forests and most woodlands in humid regions. Very poor correlation was also found for some of the arid inlands, without an obvious relation to vegetation types.
- The regression equation confirmed that the model has no predictive value for forested areas (slope is close to zero and the intercept will be the mean). Negative slopes are found for higher altitude areas in Tasmania and SE Australia; water availability is unlikely to be the main determinant of FPAR in these regions. Slopes of 0.50 to 1.0 and a small intercept are found for the areas that show good correlation between modelled and observed FPAR.
- Reasonable NSME values are found for some of the areas with high correlation, whereas areas with high CRV indicate those areas where NSME would be appreciable after bias correction.
Figure 25. Mean, standard deviation and bias (model vs. observations) between modelled FPAR and AVHRR FPAR.
Figure 26. Standard difference (SD), coefficient of correlation (R), intercept and slope of a linear regression of AVHRR FPAR on modelled FPAR, Nash-Sutcliffe Model Efficiency (NSME), and correlated residual variance (CRV).
7.3.2 MODIS FPAR

Maps showing statistics of the distribution of observed and modelled values are shown in Figure 27, as well as the absolute and relative bias. The following observations are made:

- Areas without MODIS data (shown in black) mostly coincide with water bodies, salt lakes and floodplains, and areas with high albedo.

- The mean observed and modelled FPAR show similar spatial patterns, varying from <0.25 units in arid regions to >0.75 in wet regions. However, the modelled FPAR appears to have stronger spatial gradients than the MODIS FPAR, as well as lower minimum values in arid Australia and higher maximum values in humid regions.

- The spatial pattern in standard deviation in modelled and MODIS FPAR are rather different. MODIS FPAR shows greater standard deviation for the pasture and cropping regions and for dense forests (0.50–0.75) than does the modelled FPAR (0.15–0.30).

- Compared to MODIS FPAR, modelled FPAR shows the greatest bias for the high altitude parts of Tasmania (more than +0.25 units). Modelled values are greater by 0.10–0.25 units for higher altitude forests in SE Australia and woodlands of Northern Australia.

- Mean modelled FPAR is lower by 0.10–0.20 units for woodland regions in south central WA and SE Queensland, and the floodplain and irrigation areas of the Murray-Darling Basin.

- Mean modelled FPAR is lower by up to ~0.10 for most parts of arid inland Australia. Because of the low mean FPAR this leads to large negative bias in the modelled FPAR (less than –50%)

- The greatest relative positive bias (more than +50%) occurs for mountainous parts of Tasmania. Somewhat lesser positive bias (30-50%) is calculated for Northern Australia.

Maps showing statistics of the agreement in temporal patterns are shown in Figure 28. The following observations are made:

- The standard error (SD) is generally 0.15–0.20, with lower values (~0.10) in the most arid regions and SD>0.40 in high altitude areas of Tasmania.

- The coefficient of correlation is good ($R>0.75$) for the pasture and cropping regions of WA and SE Australia, and some of the grasslands of Queensland and NT (cf. Figure 15).

- Correlation is very poor ($R<0$) for forests. Very poor correlation was also found for some of the arid inlands.

- The regression equation confirmed that the model has no predictive value for forested areas (slope is close to zero and the intercept will be the mean). Negative slopes are found for Tasmania and higher altitude forests in SE Australia, indicating that water availability is unlikely to be the main determinant of FPAR in these regions. Slopes of 0.50 to 1.0 and a small intercept are found for the areas that show good correlation between modelled and observed FPAR, as well as for the open scrublands of the Nullarbor.

- NSME values are low throughout, indicating a bias exists even where $R$ is comparatively high.
Figure 27. Mean, standard deviation and bias (model vs. observations) between modelled FPAR and MODIS FPAR. Areas masked out in the product are shown in black.
Figure 28. Standard difference (SD), coefficient of correlation (R), intercept and slope of a linear regression of MODIS FPAR on modelled FPAR, Nash-Sutcliffe Model Efficiency (NSME), and correlated residual variance (CRV).
7.3.3 MODIS fraction photosynthetic vegetation

Maps showing statistics of the distribution of observed and modelled values are shown in Figure 29, as well as the absolute and relative bias. The following observations are made:

- The mean observed and modelled FPV show rather similar spatial patterns, varying from <0.25 units in arid regions to >0.75 in wet regions. The modelled FPV appears to have stronger spatial gradients than the MODIS FPV, particularly in N Australia, whereas modelled FPV for forests (>0.80) is higher than MODIS FPV (<0.75).

- The spatial pattern in standard deviation in modelled and MODIS FPV are rather different. MODIS FPV shows greater standard deviation in forested (>0.75) than does the modelled FPV (<0.25), whereas standard deviation in modelled values is greater for open woodland regions.

- Compared to MODIS FPV, modelled FPV shows the greatest bias for the higher altitudes of Tasmania and SE Australia, as well as for the Kimberley and other ranges of NT (0.2–0.4).

- Mean modelled FPV is lower by 0.20–0.40 units for woodland regions in south central WA and SE Queensland and the floodplain and irrigation areas of the Murray-Darling Basin.

- Mean modelled FPV is higher by up to ~0.15 for most parts of arid inland Australia. Because of the low mean FPV this leads to a large positive relative bias (more than 50%).

- The greatest relative negative bias (less than −50%) occurs for inland S WA, western NSW and SE Queensland.

Maps showing statistics of the agreement in temporal patterns are shown in Figure 30. The following observations are made:

- The standard error (SD) is generally 0.15–0.20, with lower values (~0.10) in the most arid regions and very high values SD>0.75 in high altitude areas of Tasmania and coastal far north Queensland.

- The coefficient of correlation is reasonable \( R > 0.50 \) for most of Australia with the exception of the more humid forests along the coast \( R < 0 \). Poor correlation is also found for the arid regions of S Australia. Best correlations \( R > 0.75 \) are found for pasture and croplands areas and grasslands in Queensland and NT (cf. Figure 15).

- The regression equation confirmed that the model has no predictive value for forested areas (slope is close to zero and the intercept will be the mean). Negative slopes are found for higher altitude regions in Tasmania and SE Australia, indicating that water availability is unlikely to be the main determinant of FPAR in these regions. Slopes of 0.50 to 1.0 and a small intercept are found for the areas that show good correlation between modelled and observed FPAR.

- NSME values are low throughout except for parts of the pasture and cropping regions, indicating a bias exists even where \( R \) is comparatively high. The CRV shows areas where bias correction can improve NSME.
Figure 29. Mean, standard deviation and bias (model vs. observations) between modelled and MODIS-derived fraction photosynthetic vegetation (FPV).
Figure 30. Standard difference (SD), coefficient of correlation (R), intercept and slope of a linear regression of MODIS-derived fraction photosynthetic vegetation on modelled values, Nash-Sutcliffe Model Efficiency (NSME), and correlated residual variance (CRV).
7.3.4 LAI

Maps showing statistics of the distribution of observed and modelled values are shown in Figure 31, as well as the absolute and relative bias. The following observations are made:

- The mean observed and modelled LAI show very similar spatial patterns, varying from LAI<1 in arid regions to >3 in wet regions. The modelled LAI appears to be consistently higher in the woodlands of N Australia, whereas modelled LAI for higher altitude regions in Tasmania (>4) is much higher than MODIS LAI (1-3).

- The spatial pattern in standard deviation in modelled and MODIS LAI are very different however. MODIS LAI shows lesser standard deviation (by up to ca. 1.5 units) than modelled LAI, in particular in Northern Australia.

- Mean modelled LAI is within ±0.5 units of MODIS LAI for most of Australia. Compared to MODIS LAI, modelled LAI shows the greatest bias for higher altitude regions of Tasmania and SE Australia (>1 units) and to a lesser extent the more dense woodlands of Northern Australia, for example Kimberly, top end and Cape York Peninsula (>0.75 units).

- Because of the low mean MODIS LAI in arid parts of Australia there was a strong negative relative bias (− 50–100%). There was a strong positive relative bias for higher altitude forests in Tasmania, SE Australia and SW WA, as well as for Northern Australia (+50 to +100%)

Maps showing statistics of the agreement in temporal patterns are shown in Figure 32. The following observations are made:

- The standard error (SD) between MODIS and modelled LAI is around 0.5–1.0 for most of Australia. Very high values (>2) are found in high altitude areas of Tasmania and along the east coast.

- The coefficient of correlation is reasonable ($R>0.50$) for most of Australia. Very poor correlation ($R<0$) is calculated for the forested areas along the coasts, as well as parts of arid inland Australia.

- For much of Australia, the regression equation acted to reduce the variance in LAI and increase the minimum LAI (slope <1 and intercept>0). A positive intercept was still found for those areas with high $R$, suggesting that here too, minimum LAI was estimated lower by the model. Negative slopes are higher altitude regions of Tasmania and SE Australia.

- NSME values are low throughout except for parts of the wheat belt and the SE Australian cropping regions.
Figure 31. Mean, standard deviation and bias (model vs. observations) between modelled and MODIS LAI.
Figure 32. Standard difference (SD), coefficient of correlation (R), intercept and slope of a linear regression of MODIS LAI on modelled LAI, Nash-Sutcliffe Model Efficiency (NSME), and correlated residual variance (CRV).
7.3.5 EVI

Maps showing statistics of the distribution of observed and modelled values are shown in Figure 33, as well as the absolute and relative bias. The following observations are made:

- The mean observed and modelled EVI show very similar spatial patterns, varying from EVI < 0.25 in arid regions to > 0.50 in wet regions.
- The spatial patterns in σ of EVI shows higher modelled than MODIS values across most of the continent, and particularly for northern Australia, where model σ is ~ 0.25 and MODIS σ ~ 0.10.
- Mean modelled EVI is within ± 0.1 units of MODIS EVI for most of Australia. Compared to MODIS EVI, modelled EVI shows the greatest bias for the higher altitudes of Tasmania and SE Australia (> 0.15 units) and to a lesser extent the more dense woodlands of Northern Australia (Kimberly, top end and Cape York Peninsula; 0.05–0.10 units).
- Modelled EVI is considerably lower for the floodplains and irrigation areas of the MDB and some of the humid regions of Queensland.
- In relative terms, modelled and observed EVI are in good agreement for most of Australia. The strongest positive EVI bias (+ 50–100%) was found for Tasmania, high elevations of SE Australia, Kimberly and the lakes of inland Australia. There was a reasonably strong negative bias for the scrublands of the Nullarbor and Mallee and woodland areas in SE Queensland, as well as for the floodplain and irrigation areas of the Murray-Darling and Lake Eyre basin (~ 30–50%).

Maps showing statistics of the agreement in temporal patterns are shown in Figure 34. The following observations are made:

- The standard error (SD) between MODIS and modelled EVI is < 0.15 units for most of Australia. Very high SD (> 0.30) are found in high altitude areas of Tasmania, whereas SD of 0.15–0.30 was found for most of Northern Australia.
- Linear edges are visible in the R map. This is associated with 11 MODIS images that only cover part of Australia. The cause for these missing data and its implications have not been investigated.
- The coefficient of correlation is very good (R > 0.80) for the pastures and croplands of WA and SE Australia and some of the open grasslands of Queensland and NT. Correlation is very poor (R < 0) for all closed forests, as well as parts of arid inland Australia.
- For much of Australia, the regression equation acted to reduce the variance in LAI and increase the minimum LAI (slope < 1 and intercept > 0). Negative slopes are found for higher altitude regions in Tasmania and SE Australia.
- NSME values are low throughout except for parts of the pasture and croplands of WA and SE Australia.
Figure 33. Mean, standard deviation and bias (model vs. observations) between modelled EVI and MODIS EVI.
Figure 34. Standard difference (SD), coefficient of correlation (R), intercept and slope of a linear regression of MODIS EVI on modelled EVI, Nash-Sutcliffe Model Efficiency (NSME), and correlated residual variance (CRV).
7.4 Discussion

7.4.1 Overall comparison

Some overall results are summarised as follows:

- The AVHRR FPAR, MODIS FPAR and MODIS FPV were assumed conceptually identical for the purposes of this evaluation, but showed some systematic differences. Further research may be required to better understand the relative accuracies of these products, and to develop a best estimate dynamic canopy fraction data set from these.

- Large-scale patterns in mean vegetation density were reproduced reasonably well by the model. For example, mean model LAI was generally within 0.5 units of MODIS LAI, and EVI generally within 0.1 units.

- At large scale, the main areas of difference were high altitude vegetation and transitional regions, where modelled FPAR/FPV gradients appear stronger than AVHRR and MODIS FPAR and FPV. Model EVI patterns appear similar to MODIS EVI however.

- The standard deviation (σ) in modelled FPAR is of similar magnitude and pattern as AVHRR FPAR, but only about half of the σ in MODIS FPAR (particularly for cropping regions and forests). Conversely, standard deviation in modelled FPV is greater than MODIS FPV for subhumid regions, but considerably less for densely vegetation areas (for example forests). Standard deviation in modelled LAI and EVI were also considerably greater than for MODIS LAI and EVI, by ca. 1.5 and 0.15 units, respectively. The reasons for these apparently inconsistent findings are unclear.

7.4.2 Regional comparison

The results are summarised and discussed below for individual regions that showed some consistency.

Pasture and cropping areas. For all vegetation data sets evaluated, the model showed the best performance in reproducing averages and temporal patterns the pasture and cropping areas, including WA wheat belt and the humid and subhumid regions of SA, Victoria and NSW. The mean and variation or standard deviation in modelled FPAR was similar to that of AVHRR FPAR and MODIS EVI. However, the mean was slightly less and the variation almost half that of MODIS FPAR, whereas conversely, mean modelled FPV and LAI were generally somewhat higher and the variation greater than the MODIS equivalents. R values were reasonable to good (>0.75 for AVHRR and MODIS FPAR/FPV, and >0.80 for MODIS EVI), and even the absolute agreement was satisfactory (NSME>0.50) for AVHRR FPAR, and in smaller areas for the other products.

Modelled and MODIS EVI time series are shown in Figure 13 for the Kyeamba (NSW) flux tower site, which falls within this vegetation category. The magnitude and strong variability in EVI is reproduced well by the model and explain the high correlation (R=0.81 in this case). The same figure also shows that there appears to be some offset between the modelled and observe data, which will affect statistics of the absolute agreement such as NSME. Overall, vegetation dynamics in this biome indeed appear mostly controlled by soil water availability and hence can be reproduced by the model.
High altitude areas of Tasmania and southeast Australia. The model overestimated all measures of vegetation density and greenness for these areas. This was particularly true for the higher altitude areas of Tasmania: mean modelled FPAR was more than 0.25 greater than AVHRR/MODIS FPAR and FPV, and modelled LAI and EVI was greater than the MODIS equivalents by 1–2 and 0.15 units, respectively. A similar but mostly slightly less overestimation occurred for the higher altitude regions of SE Australia. In addition, the seasonality was in fact opposite from that which was modelled, leading to negative but low correlations.

The Tumbarumba flux tower site is located in this vegetation type, if towards the warmer end, occasionally receiving snow. Figure 13 illustrates some of the reasons for the poor model performance. The modelled EVI also shows a low variability, but decreases somewhat in summer, due to the modelled effect of a modest moisture limitation. The observed values show no such decrease. Some low frequency variation in EVI is visible with (on average) the highest EVI values in May and the lowest in October, indicating a temperature limitation. This summer/winter phenology will be stronger for higher altitude areas and is not reproduced at all by the current model version. The figure also shows considerable noise in the EVI values, which has been attributed to the potential effects of wind and incidence angle on the reflective behaviour of eucalypt canopies (Hill et al. 2006). Snow background effects may cause further errors in the retrieved vegetation variables.

Considering the bias and low correlation, it may well be that the mean seasonal pattern in vegetation properties (a so-called ‘climatology’) provides a better estimate of vegetation behaviour than the vegetation model.

Forests. Seasonal patterns in vegetation density closed and open forests along the Australian coast were reproduced very poorly ($R<0$ for all measures). For many of the forests this is because there is no strong seasonal pattern that can be reproduced, as mentioned above. The mentioned problems of measurement over eucalypt forests may further increase measurement error and hence further reduce the agreement with modelled values.

Floodplains and irrigation areas. Modelled vegetation density was lower for several floodplain and irrigation areas in the Murray-Darling Basin and the channel country (Lake Eyre basin): modelled FPAR was lower than AVHRR and MODIS FPAR (by 0.10–0.25 units), MODIS FPV (0.20–0.40), MODIS LAI (0.50) and MODIS EVI (0.10–0.20). The poor performance of the model in these environments is not surprising and can be attributed to the model assumption that all available water is derived from local rainfall. This is a known limitation of the current model version.

Northern grasslands. Model performance for the grasslands of Queensland and Northern Australia was similarly good as for pasture and cropping areas in the southern parts of the continent. Much of this vegetation consists of tussock grassland on cracking clay soils or Vertosols (cf. Figure 23 and Figure 24). The distribution statistics and relative and absolute agreement were similar to those for pasture and cropping areas. As for that category, differences in the mean and standard deviation varied between the different products, but were generally modest, while $R$ values were also similarly good. The good performance is attributed to the accurately reproduced response of these grasses to water availability.

Northern savannah. Model estimated vegetation density for the Northern Australian savannah woodlands was generally greater than for the satellite products, particularly for the denser woodlands at the wetter, northern side (for example in Kimberly, top end and northern part of Cape York Peninsula). Mean modelled metrics were 0.10–0.25 units greater than both AVHRR and MODIS FPAR; 0.20–0.40 units greater than MODIS FPV; 0.75 units greater than MODIS LAI; and 0.05–0.10 units greater than MODIS EVI. The modelled variation in LAI and EVI was also greater than for the MODIS products.
The correlation in temporal patterns was modest to reasonable over the drier savannah areas (open woodland), with $R$ of 0.5–0.7 for MODIS FPAR, LAI and EVI. Correlation was low (0.25–0.50) for the more dense woodlands to the north. This contrast was even stronger in the AVHRR FPAR and MODIS FPV, with sometimes higher $R$ values over the southern savannahs, and very low correlation ($R<0.25$) over the northern woodlands.

Modelled and MODIS EVI are shown for the Howard Springs (NT) savannah flux tower site in Figure 13, which is in the wetter north. The range in EVI appears reasonably similar. However there is a clear timing difference for both the growth and senescence stages, which goes some way to explain the low correlation coefficient. Due to the model logic the vegetation does not increase until soil water stores are replenished by the first rains. In reality the phenology of the deep-rooted vegetation component appears to anticipate the wet season and leaf flush occurs before the onset of the rains (possibly in response to air humidity; L. Hutley, pers. comm.).

**Sub-humid eucalypt woodlands.** Remnant eucalypt woodlands in south-central WA, the Murray-Darling Basin Mallee region and SE Queensland all show similar features in the comparison. The modelled mean vegetation density was somewhat lower than those in the satellite products (by ca. 0.1–0.2 AVHRR/MODIS FPAR, 0.2–0.4 MODIS FPV, 0.5 LAI, and 0.1 EVI units respectively). The relative agreement in temporal patterns was poor in all cases ($R<0.25$), and the temporal variation in modelled properties greater than that in the satellite products. The expected underestimation of the fraction deep-rooted vegetation provides a feasible explanation for this: the minimum FPAR was around 0.25–0.35 for this vegetation (Figure 1) and hence this was also the estimated fraction deep-rooted vegetation. It seems likely that the actual fraction of areas explored by the plant roots is much higher than this.

**Arid regions.** Modelled FPAR was 0.10 lower than AVHRR and MODIS FPAR for much of arid inland Australia, but up to 0.15 higher than MODIS FPV. EVI generally agreed reasonably well, but modelled values were noticeably lower than MODIS values for the Nullarbor shrublands. The degree to which modelled and observed temporal patterns in vegetation density were correlated varied, with very poor correlations ($R<0$) consistently found for some regions (for example large parts of South Australia), modest to reasonable correlation ($R>0.50$) for some other regions (for example Pilbara, central Australia), and inconsistent results for some other areas (for example central WA). The MODIS FPAR and LAI products were not available for significant parts of the arid interior. This appears associated with the high reflectivity of these areas, which may have been masked out by the quality control algorithm. It is also notable that $R$ values were generally somewhat better for the MODIS FPV product than for the other products. This suggests that the effect of soil litter on FPAR and other parameters may play a role. There are several other reasons why a good agreement between modelled and observed vegetation properties would not necessarily be expected in many of these regions: rainfall is erratic, vegetation greening short-lived, and rainfall forcing data quality low in many of the areas.

### 7.4.3 Implications for water balance estimation

Inadequate simulation of vegetation dynamics can affect water balance estimation mainly by leading to errors in the estimation of actual ET. The likelihood of this occurring varies from between vegetation types.

The largest difference between modelled and observed vegetation dynamics was found for higher altitude regions in Tasmania and SE Australia. In these regions precipitation will be considerably greater than potential ET and therefore the effect of poor vegetation simulation on ET may be limited. The greatest effect may be that associated with the estimated rainfall interception losses, which are not constrained by potential ET and therefore can be
overestimated. This adds to the overall uncertainty associated with the estimation of rainfall interception losses (Section 5).

For other vegetation types, the mean and temporal patterns in vegetation metrics were reproduced to varying degrees, partly depending also on errors and variability (respectively) in the observation-based products. An important source of difference between model and observations in dry regions may be the often poor quality of rainfall estimation (see Section 5), which also affects water balance estimation. Differences may not always have important consequences for water balance estimates. A case in point is the Howard Springs savannah woodland site. While correlation between modelled and observed EVI was modest ($R=0.52$), this did not appear to be an important source of differences between modelled and flux tower ET ($R^2=0.75$; Table 7)

Large differences also exist for floodplains, irrigation areas and other areas that receive additional surface or groundwater inflows from elsewhere. This uncertainty in the model results is important and needs to be addressed. There appears to be scope to assimilate satellite observations. Without updating the soil moisture store, this would improve the agreement with observed vegetation density were soil water is not limiting, and could reproduce for example disturbances, temperature driven changes, and to some extent also the phenology of water limited (deep-rooted) vegetation (Stöckli et al. 2008b; Van Dijk and Renzullo 2009).

Where the connection is strong enough and model and observation errors sufficiently well understood, data assimilation may even help to improve the estimate of soil water storage, which in turn may make it possible to improve precipitation estimates or estimate the quantity of water added by surface or groundwater inflows (cf. Section 5). This remains to be tested, however.

7.5 Conclusions

The discussion is summarised and main conclusions drawn as follows:

- The AWRA version 0.5 model simulates the response of shallow-rooted vegetation (grasses, herbs and crops) to water availability rather well.
- The agreement in seasonal dynamics of semi-arid and arid regions varied. Some of the lesser performance may be attributable to errors in rainfall forcing and erroneous estimation of the fraction of soil explored by deep-rooted vegetation.
- Seasonal patterns in evergreen forests were reproduced poorly, which can partly be attributed to the low temporal variability in vegetation density and errors in remote sensing over (eucalypt) forests. Furthermore, the effect of cool temperatures on the vegetation is not simulated by the model and hence vegetation dynamics at higher elevations in Tasmania and SE Australia were poorly simulated, and vegetation density overestimated. This might have an effect on water balance estimates through some overestimation of rainfall interception losses.
- The density of northern savannah woodlands appeared to be overestimated and there was a mismatch in the timing of greening and senescing. This does not necessarily lead to large errors in water balance estimation however.

7.6 Recommendations

Based on this analysis, it is recommended that:
Further research be carried out to better understand the relative accuracies of the AVHRR FPAR, MODIS FPAR and MODIS FPV products, and to develop a method to produce dynamic, best estimate canopy fraction data;

Alternative dynamic vegetation model approaches be developed for vegetation types and regions where temperature rather than water availability is the growth limiting factor. This could be done through use of a vegetation ‘climatology’ or alternatively, by developing a simplified temperature driven phenology model (cf. van Dijk et al. 2005; Stöckli et al. 2008b).

Experiments be done to assess whether water balance estimates for water-limited vegetation types can be improved by assimilating satellite vegetation observations. This could also open opportunities to improve precipitation forcing and estimate the influx of surface and groundwater.
REFERENCES


Donohue RJ, Roderick ML, McVicar TR (2007) 'Correcting Long-Term AVHRR Reflectance Data Using the Vegetation Cover Triangle.' CSIRO, Canberra.


The Australian Water Resources Assessment system


Li M, Shao Q (accepted) An improved statistical approach to merge satellite rainfall estimates and rain gauge data. *Journal of Hydrology*. 

92


APPENDIX A. MODEL CODE

The AWRA-L implementation used to produce simulations for this report was written in MATLAB™. The dynamic part of the model (that is, the code executed at each time step to evolve the model) is reproduced below. Model parameterisation and pre-processing of the input data and the overall workflow execution are done by other scripts that available upon request.

Note that parameter and variable references may occasionally vary from those used in this report. Further note that the comments below (in green) can vary somewhat from the comments in the original code, for example due to changes in the section numbering after revision of the reports. The section numbering below applies to AWRA Technical Report 3.

```
function [state,out]=timestep_05(in,state,par)

% Australian Water Resources Assessment Landscape (AWRA-L) model
% This script contains the AWRA-L time step model
% Full documentation is found in the technical reference:
% Technical Report 3. Landscape Model (version 0.5) Technical Description
% The section references below refer to the sections in the above report.

% ASSIGN STATE VARIABLES
S0          =   state.S0;
Ss          =   state.Ss;
Sd          =   state.Sd;
Sg          =   state.Sg;
Sr          =   state.Sr;
Mleaf       =   state.Mleaf;

% ASSIGN INPUT VARIABLES
Pg          =   in.Pg;
Rg          =   in.Rg;
Ta          =   in.Ta;
pe          =   in.pe;
pair        =   in.pair;
u2          =   in.u2;

% ASSIGN PARAMETERS
Nhru        =   par.Nhru;
Fhru        =   par.Fhru;
SLA         =   par.SLA;
LAIref      =   par.LAIref;
Sgref       =   par.Sgref;
S0FC        =   par.S0FC;
SsFC        =   par.SsFC;
SdFC        =   par.SdFC;
fday        =   par.fday;
Vc          =   par.Vc;
alb_dry     =   par.alb_dry;
alb_wet     =   par.alb_wet;
w0ref_alb   =   par.w0ref_alb;
Gfrac_max   =   par.Gfrac_max;
fvegref_G   =   par.fvegref_G;
hveg        =   par.hveg;
Us0         =   par.Us0;
Ud0         =   par.Ud0;
wslimU      =   par.wslimU;
wdlimU      =   par.wdlimU;
cGsmax      =   par.cGsmax;
FsoilEmax   =   par.FsoilEmax;
w0limE      =   par.w0limE;
FwaterE     =   par.FwaterE;
s_sls       =   par.s_sls;
```

ER_frac_ref = par.ER_frac_ref;
InitLoss = par.InitLoss;
PrefR = par.PrefR;
FdrainFC = par.FdrainFC;
beta = par.beta;
Fgw_conn = par.Fgw_conn;
K_gw = par.K_gw;
K_rout = par.K_rout;
LAImax = par.LAImax;
Tgrow = par.Tgrow;
Tsenc = par.Tsenc;

% diagnostic equations
LAI = SLA.*Mleaf;    % (5.3)
fveg = 1 - exp(-LAI./LAIref);     % (5.3)
fsoil = 1 - fveg;
w0 = S0./S0FC;     % (2.2)
ws = Ss./SsFC;     % (2.2)
wd = Sd./SdFC;     % (2.2)

% Spatialise catchment fractions (3.8)
for i=1:par.Nhru
    fwater = [fwater; min(0.005,0.007.*Sr.^0.75)];
    fsat = [fsat; min(1,max(min(0.005,0.007.*Sr.^0.75),Sg./Sgref))];
    Sghru = [Sghru; Sg];
end

% CALCULATION OF PET
% Conversions and coefficients (3.2)
pes = 610.8.*exp(17.27.*Ta./((237.3+Ta)));
fRH = pe./pes;
cRE = 0.03449+4.27e-5.*Ta;
Caero = fday.*0.176.*(1+Ta./209.1).*(pair-0.417.*pe).*(1-fRH);
keps = 1.4e-3.*((Ta./187).^2+Ta./107+1).*6.36.*pair./pes;
Rgeff = Rg./fday;
% shortwave radiation balance (3.4)
alb_veg = 0.452.*Vc;
alb_soil = alb_wet+(alb_dry-alb_wet).*exp(-w0./w0ref_alb);
alb = fveg.*alb_veg+fsoil.*alb_soil;
RSn = (1-alb).*Rgeff;
% longwave radiation balance
StefBolz = 5.67e-8;
Tkelv = Ta+273.16;
RLin = (0.65.*(pe./Tkelv).^0.14).*StefBolz.*Tkelv.^4;     % (3.5)
RLout = 1.*StefBolz.*Tkelv.^4;      % (3.6)
RLn = RLin-RLout;
fGR = Gfrac_max.*(1-exp(-fsoil./fvegref_G));     % (3.7)
Rneff = (RSn+RLn).*(1-fGR);
% Aerodynamic conductance (3.9)
fh = log(813./hveg-5.45);
ku2 = 0.305./(fh.*((fh+2.3));
ga = ku2.*u2;
% Potential evaporation (3.8)
kalpha = 1+Caero.*ga./Rneff;
E0 = cRE.*(1./(1+keps)).*kalpha.*Rneff.*fday;

% CALCULATION OF ET FLUXES AND ROOT WATER UPTAKE
% Root water uptake constraint (4.4)
Usmax = Us0.*min(1,ws./wslimU);
Udmax = Ud0.*min(1,wd./wdlimU);
U0 = max(Usmax,Udmax);
% Maximum transpiration (4.3)
Gsmax = cGsmax.*Vc;
gs = fveg.*Gsmax;
ft = 1./(1+keps./((1+keps)).*ga./gs);
Etmax = ft.*E0;
% Actual transpiration (4.1)
Et = min(U0, Etmax);
% Root water uptake distribution (2.4)
Us = min( (Usmax/(Usmax+Udmax)).*Et, Ss-1e-2 );
Ud = min( (Udmax/(Usmax+Udmax)).*Et, Sd-1e-2 );
Et = Us + Ud;   % to ensure mass balance
% Soil evaporation (4.5)
fsol1E = Fsoil1Emax.*(1,ws./wslimE);
Es = (1-fsat).*fsol1E.*(E0-Et);
% Groundwater evaporation (4.6)
Eg = (fsat-fwater).*FsoilEmax.*(E0-Et);
The Australian Water Resources Assessment system

% Open water evaporation (4.7);
Er = fwater.*FwaterE.*( E0-Et ) ;
% Rainfall interception evaporation (4.2)
Sveg = S_sls.*LAI;
ER = ER_frac_ref.*fveg;
Pwet = -log(1-ER./fveg).*Sveg./ER;
Ei = (Pg<Pwet).*fveg.*Pg+(Pg>Pwet).*(fveg.*Pwet+ER.*(Pg-Pwet));

% CALCULATION OF WATER BALANCES
% soil surface water fluxes (2.3)
Pn = max(0, Pg - Ei - InitLoss) ;
Rhof = (1-fsat).*( Pn/((Pn+PrefR) )*Pn ;
Rsof = fsat.*Pn ;
QR = Rhof + Rsof ;
I = Pg - Ei - QR ;
% SOIL WATER BALANCES (2.1 inc drainage 2.5)
% Topsoil (S0)
S0 = S0 + I - Es ;
SzFC = S0FC;
Sz = S0;
wz = max(1e-2,Sz)./SzFC;
fD = (wz>1).*max(FdrainFC,1.-1./wz) + (wz<=1).*FdrainFC.*exp(beta.*(wz-1) );
Dz = min(fD.*Sz,Sz-1e-2);
D0 = Dz;
S0 = S0 - D0 ;
% Shallow soil(Ss)
Ss = Ss + D0 - Us;
SzFC = SsFC;
Sz = Ss;
wz = max(1e-2,Sz)./SzFC;
fD = (wz>1).*max(FdrainFC,1.-1./wz) + (wz<=1).*FdrainFC.*exp(beta.*(wz-1) );
Dz = min(fD.*Sz,Sz-1e-2);
Ds = Dz;
Ss = Ss - Ds ;
% Deep soil (Sd) (inc capillary rise 2.7)
Sd = Sd + Ds - Ud;
SzFC = SdFC;
Sz = Sd;
wz = max(1e-2,Sz)./SzFC;
fD = (wz>1).*max(FdrainFC,1.-1./wz) + (wz<=1).*FdrainFC.*exp(beta.*(wz-1) );
Dz = min(fD.*Sz,Sz-1e-2);
Dd = Dz;
Sd = Sd - Dd;
Y = min(Fgw_conn.*max(0,wdlimU.*SdFC-Sd),Sghru);
Sd = Sd + Y;
% CATCHMENT WATER BALANCE
% Groundwater store water balance (Sg) (2.6)
NetGf = sum(Fhru.*(Dd - Eg - Y));
Sg = Sg + NetGf;
Qg = min(Sg, (1-exp(-K_gw)).*Sg) ;
Sg = Sg - Qg;
% Surface water store water balance (Sr) (2.8)
Sr = Sr + sum(Fhru.*(QR - Er) ) + Qg ;
Qtot = min(Sr, (1-exp(-K_rout)).*Sr) ;
Sr = Sr - Qtot;
% VEGETATION ADJUSTMENT (5)
fvmax = 1-exp(-LAImax./LAIref);
fveq = (1./max(E0.U0)-1,1e-3)).*(keps/(1+keps)).*(ga./Gsmax);
dMleaf = -log(1-fveq).*LAIref./SLA-Mleaf ;
Mleafnet = (dMleaf>0).*min(dMleaf/Tgrow,MaxGrow) +(dMleaf<0).*dMleaf./Tsenc;
Mleaf = Mleaf + Mleafnet;
% Updating diagnostics
LAI = SLA.*Mleaf;  % (5.3)
fveg = 1 - exp(-LAI./LAIref) ;  % (5.3)
fsoil = 1 - fveg;
w0 = S0./S0FC;  % (2.1)
ws = Ss./SsFC;  % (2.1)
w1 = Sd./SdFC;  % (2.1)
% ASSIGN OUTPUT VARIABLES
% fluxes
out.E0 = sum(Fhru.*E0);
out.Ee = sum(Fhru.*(Es + Eg + Er + Ei));
out.Et = sum(Fhru.*Et);
Landscape Model (version 0.5) Evaluation Against Observations

out.Ei = sum(Fhru.*Ei);
out.Etot = out.Et + out.Ee;
out.Qtot = Qtot;
out.gwflux = NetGf;

% states
out.S0 = sum(Fhru .* S0);
out.Ss = sum(Fhru .* Ss);
out.Sd = sum(Fhru .* Sd);
out.Sg = Sg;
out.Stot = out.S0 + out.Ss + out.Sd + Sg + Sr + sum(Fhru .* Mleaf.*(0.8/0.2));
% NOTE: 0.8 because wet leaf biomass is assumed to consist of 80% water
out.Mleaf = sum(Fhru .* Mleaf);
out.LAI = sum(Fhru .* LAI);
out.fveg = sum(Fhru .* fveg);
out.fsat = sum(Fhru .* fsat);
out.wunsat = sum(Fhru .* w0);

% synthetic satellite products
out.albedo = sum(Fhru .* alb);
out.EVI = sum(Fhru .* (Vc.*fveg+0.07));
% NOTE 0.07 is assumed EVI for bare soil
out.fsat = sum(Fhru .* fsat);
out.wunsat = sum(Fhru .* w0);

% ASSIGN STATE VARIABLES
state.S0 = S0;
state.Ss = Ss;
state.Sd = Sd;
state.Sg = Sg;
state.Sr = Sr;
state.Mleaf = Mleaf;

%=========EOF=========