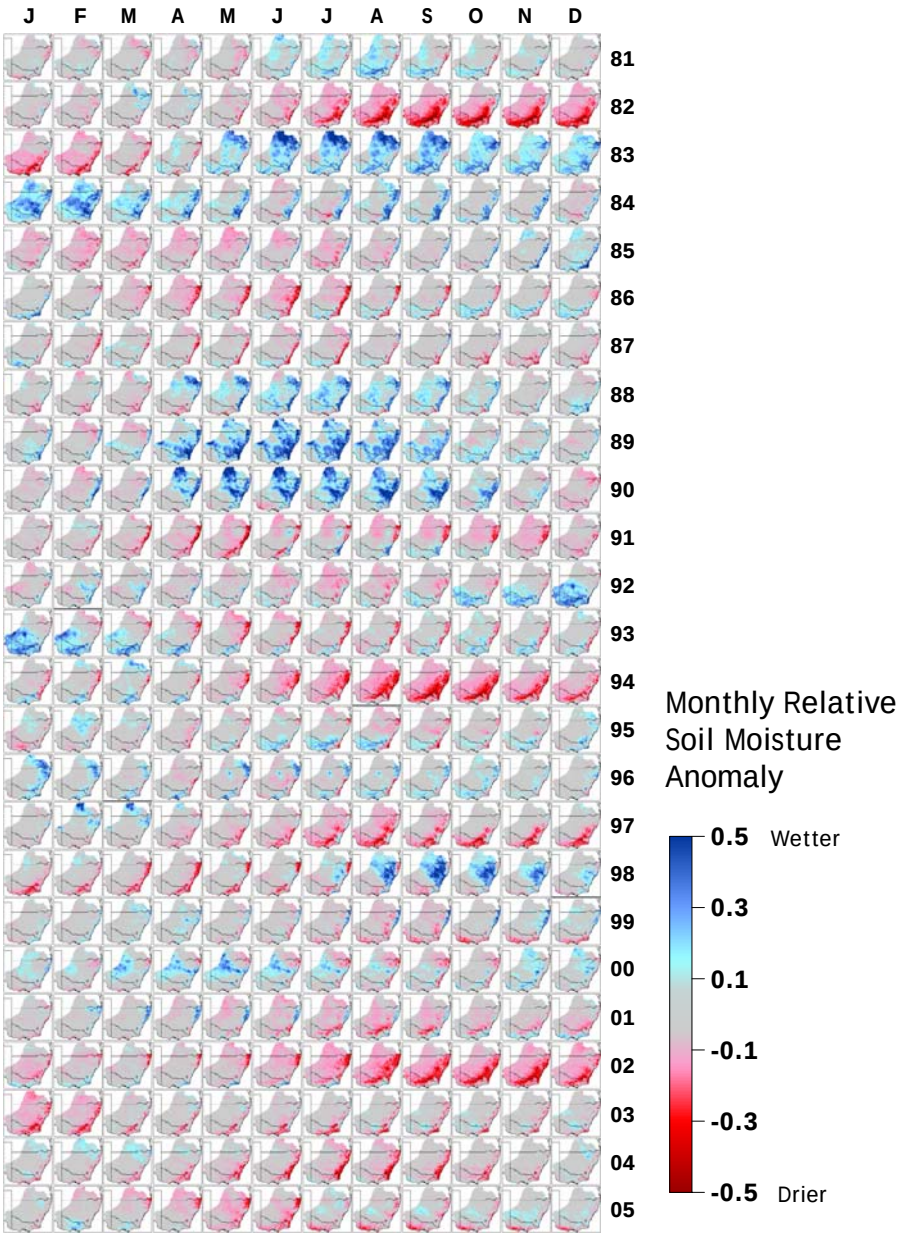
Australian Water Availability Project

CSIRO Marine and Atmospheric Research (Earth Observation) Component

Final Report

M.R. Raupach, C.M. Trudinger, P.R. Briggs and E.A. King

30 June 2006



CSIRO

Table of Contents

Summary

This is the Final Report of the CSIRO Marine and Atmospheric Research (former Earth Observation Centre) component of the Australian Water Availability Project, Phase 1 (1 January 2005 to 30 June 2006). The objectives of this component (in summary) are:

- To develop a "Hydrological and Terrestrial Biosphere Data Assimilation System" (HTBDAS) for the near-real-time estimation and prediction of water availability (relative soil moisture and water balance fluxes including transpiration, soil evaporation, runoff and drainage) with daily time resolution and 5 km spatial resolution;
- To evaluate the system for 5 selected catchments in the Murray-Darling Basin;
- To apply the system over the Murray-Darling Basin and adjacent coastal region in southeast Australia, obtaining climate histories of water availability over the last two decades.

The project has met its primary objectives. A working model has been developed, tested and applied to obtain histories of relative soil moisture and water balance fluxes (transpiration, soil evaporation, runoff and drainage) over southeast Australia for the 25 years from 1981 to 2005.

Specific achievements fall into three categories (model, data, assimilation) representing the essential generic components of a data assimilation system.

Model:

- Dynamic and observation models have been developed and tested in point mode;
- Forward model predictions have been extended successfully to spatial mode (southeast Australia, 25 years from 1981 to 2005);
- Model results are encouraging, even with prior-knowledge parameter values;
- Tests have been carried out for over 150 unimpaired gauged catchments, by comparing predicted and measured runoff.

Data:

- Remotely sensed (vegetation greenness and surface temperature) data have been assembled;
- Discharge data have been assembled for over 200 unimpaired catchments;
- Both datasets have been subjected to extensive quality assessment and evaluation.

Assimilation:

- The Ensemble Kalman filter (EnKF) has been developed and tested for sequential parameter estimation of catchment discharge at the scale of individual unimpaired catchments and collections of catchments taken together;
- Traditional (down-gradient) parameter estimation is being used as a check on the EnKF.

There are several remaining challenges and issues for future work. These include:

- Further improvements to the dynamic and observation models;
- Finalisation of model parameter estimation in spatial mode with both down-gradient and ensemble Kalman filter methods;
- Assimilation of more data streams (especially land surface temperature);
- Extension of the spatial domain to the whole of the Australian continent.

Preface

This is the Final Report for the CSIRO Marine and Atmospheric Research (former Earth Observation Centre) component of the Australian Water Availability Project, Phase 1 (1 January 2005 to 30 June 2006).

The report is structured as follows: In the introductory Section 1, the aims of the project are defined. Section 2 provides an overview of data assimilation in the context of water balance estimation. Section 3 describes the input observations, including both meteorological data used to drive the model and the remote sensing and runoff data which are assimilated as the model is running to improve the predictions. Section 4 describes the model development and presents the main results from forward runs for southeast Australia. Section 5 describes the data assimilation (model-data fusion) methodology, and presents results. Section 6 describes future developments, including (1) challenges remaining from the present work, and (2) the path to prediction, by linking water availability products to seasonal forecasts and climatological scenarios.

1 Introduction

The Australian Water Availability Project (AWAP) is a partnership between the Bureau of Rural Science, CSIRO (EOC and Land and Water), and the Bureau of Meteorology. The overall objective and outcome statement for AWAP are (AWAP WorkPlan 2004-2006, Version 6, 27 September 2004):

Objective: *To develop an operational prototype of a new and integrated approach to monitoring and predicting soil moisture and other components of the water balance, for space scales ranging from 1km to all Australia, and for time scales ranging from fortnightly to decades.*

Outcome statement: *This project will demonstrate a new and integrated approach to monitor and predict seasonal and multi-year water balance status from paddock to national scales. It will be focused on the needs of government, industry sectors and individuals to manage the landscape more productively and sustainably. Importantly, the products will be scalable from paddock to catchment to continent. Outputs will include maps and electronic information on past, present and future levels of the components of the water balance, from rainfall to soil moisture, and will be relevant to a broad range of government and non-government programs.*

The CSIRO Marine and Atmospheric Research (former Earth Observation Centre) component of AWAP, the subject of this report, has the following more specific objectives (summarised from the AWAP WorkPlan 2004-2006, Version 6, 27 September 2004):

- *To develop a "Hydrological and Terrestrial Biosphere Data Assimilation System" (HTBDAS) for the near-real-time estimation and prediction of water availability (relative soil moisture and water balance fluxes including transpiration, soil evaporation, runoff and drainage) with daily time resolution and 5 km spatial resolution;*
- *To evaluate the system for 5 selected catchments in the Murray-Darling Basin;*
- *To apply the system over the Murray-Darling Basin and adjacent coastal region in southeast Australia, obtaining climate histories of water availability over the last two decades.*

Appendix A gives the milestone schedule for the CSIRO components of the project, including contributions from both CSIRO Land and Water (CLW) and the Earth Observation Centre (EOC), now part of CSIRO Marine and Atmospheric Research (CMAR). Three EOC/CMAR milestone reports[1] have described progress towards these milestones.

[1] Raupach, M.R., Briggs, P.R. and Trudinger, C. (2005) Australian Water Availability Project (CSIRO Earth Observation Centre Component) Milestone Report 1: Proposed methods for generating an operational hydrological and terrestrial-biosphere data assimilation system (HTBDAS). 7 May 2005.
Raupach, M.R., Briggs, P.R. and Trudinger, C. (2005) Australian Water Availability Project (CSIRO Earth Observation Centre Component) Milestone Report 2. 18 November 2005.
Raupach, M.R., Briggs, P.R. and Trudinger, C. (2005) Australian Water Availability Project (CSIRO Earth Observation Centre Component) Milestone Report 3. 17 March 2006.

2 Overview of Data Assimilation for Water Availability Estimation

Like a financial balance sheet, the terrestrial water balance includes "revenue" (the inflows of water from precipitation and possibly irrigation) and "costs" (the outflows of water through plant transpiration, soil evaporation, runoff and drainage to groundwater). The "net cash result" or "cash reserve" of the water balance is soil moisture, the amount of water stored in the soil profile for use in plant growth. Soil water reserves are the most important indicators of drought status, but unlike the financial analogue, nature does not issue a regular balance sheet. Instead, indirect means must be used to map the soil moisture across large regions. Traditionally there are two main options, measurements and modelling.

Measurement: This involves measuring a surrogate (something related to soil moisture), and inferring soil moisture indirectly from this quantity. A common surrogate is vegetation greenness, measurable by satellite through numerous "vegetation indices". These all rely on measuring the strength of spectral features such as the "red edge", whereby green, growing vegetation absorbs strongly in the red (around 0.7 μm) and reflects strongly in the near infrared (NIR) (around 0.9 μm). One common and longstanding vegetation index is NDVI (Normalised Difference Vegetation Index). A different possibility is Land Surface Temperature (LST), also measurable from satellites, which (together with *in-situ* meteorological information) can provide information about the surface energy balance and thence the total evaporation rate, which in turn is related to soil moisture.

Modelling: This involves the calculation of soil moisture using a water balance model, by (1) measuring the water inflows (rainfall, irrigation), (2) modelling the outflows (transpiration, soil evaporation, runoff and drainage, which all depend on the soil moisture itself, together with weather, vegetation and soil properties), and (3) combining all this information using the water balance equation

$$\begin{aligned}[\text{change in soil water}] &= [\text{inflows}] - [\text{outflows}] \\ &= [\text{rainfall}] + [\text{irrigation}] \\ &\quad - [\text{transpiration}] - [\text{soil evaporation}] - [\text{runoff}] - [\text{drainage}]\end{aligned}$$

where input and output fluxes are identified by colour.

Each approach has its own has generic problems. Measurements are always indirect and depend on knowing complex relationships between the surrogate and soil moisture itself, as well as being subject to errors in the measurements themselves. Models are subject to a range of errors and uncertainties in model formulation and parameterisation, as well as uncertainties in measured input and ancillary data used by the model.

The intent of this project is to map and predict soil moisture and related quantities by using model-data fusion methods to combine both measurements and modelling. This approach (1) overcomes the individual weaknesses of measurement and modelling approaches on their own; and (2) provides information not only about soil moisture but also about other important quantities such as water flows through transpiration, evaporation, runoff and drainage, and vegetation growth itself.

The suite of techniques for combining measurements and models is collectively called "model-data fusion" or "model-data synthesis", a term which covers two broad approaches (each with numerous variants): “parameter estimation” and “data assimilation”. All approaches involve minimising some measure of the disagreement between the model and the measurements, such as a search for “least-squared-error”. *Parameter estimation* involves finding values of parameters (poorly known but supposedly constant quantities) that appear in all biophysically based models of landscape processes. It is almost always necessary to choose these parameters so that the model best fits some set of test data which it is attempting to predict. Techniques for finding the best (“optimum”) parameters range from simple graphical methods to advanced search procedures for finding multiple parameters simultaneously. *Data assimilation* is a more subtle way of combining models with measurements, in which either or both of model state variables (for example, soil moisture in the present work) and parameters are refined step by step in time as the model is run forward and as new data become available to correct the model. Data assimilation has been developed over the last two decades in meteorology and oceanography, and has contributed to substantial improvements in the accuracy of weather forecasts. The extent of this improvement is evident in Figure 1, showing the increase in weather forecast skill over the period 1981-2001 in both the northern and southern hemispheres.

This project uses model-data fusion (with a combination of data assimilation and parameter estimation) to provide maps and predictions of landscape processes (particularly soil moisture). The benefit for landscape management and productivity is potentially large. In the long term, the outcomes of such a development will be (a) integrated monitoring of the state and metabolism of landscape systems; thence (b) monitoring of system responses to resource management and climate variability and change; and thence (c) guidance for adaptive, system-wide management through continual feedback via measurement and monitoring.

The vehicle for implanting this vision in the context of the AWAP is a "Hydrological and Terrestrial-Biosphere Data Assimilation System" (HTBDAS). Figure 2 shows its six main components: (a) observations (covered here in Section 3), (b) a model (Section 4), (c) prior information (Section 4), (d) a model-data fusion process (Section 5), (e) a product interface, and (f) mechanisms for product utilisation (covered by other AWAP components). The first three of these are components for information supply, the fourth is a component for information integration, and the last two carry out information delivery and uptake. The major innovation is in the integration or model-data fusion component, (4).

3 Observations

In this project, two kinds of observation[2] are *assimilated* into the model: measures of vegetation greenness from remote sensing, and *in-situ* data on river flows, where available. A third class of observations, weather data, is used to *drive* the model. The distinction between driving data and

[2] We had intended to assimilate land surface temperature (LST) as an additional remotely sensed data stream, but delays in the availability of LST from elsewhere mean that this is no longer possible. Therefore it has been agreed that LST data will not be used in the HTBDAS in the current project. This will not affect the ability of this project to meet its prototype demonstration goal.

assimilated data is that the driving data are applied as external inputs to the model, while the assimilated data are compared with model outputs and the resulting discrepancy is minimised to produce best model performance.

The data are described in the following subsections, in each case with separately headed outlines of (1) data sources, (2) interpretation methods, and (3) analysis carried out in this project to assess the character and quality of each data set.

3.1 Remotely Sensed Measures of Vegetation Greenness

Data sources: The main remote sensing data used in this project are from the long time series available from the NOAA-AVHRR sensor aboard the operational series of NOAA polar orbiting satellites, processed and archived for the Australian continent in the CSIRO AVHRR Time Series (CATS) (King 2003). These data are supplemented with data from several other sensors (MODIS, SeaWiFS, SPOT-VGT) and a compilation project (GlobCarbon, European Space Agency) for comparison. Details are given in Appendix B. All these data sources provide measures of vegetation condition covering the entire Australian continent, at spatial resolutions from less than 1 km to 25 km, and over various multi-year periods (see below).

Data Interpretation: Our primary measure of land condition is the Normalised Difference Vegetation Index (NDVI), based on the difference in reflectance of the surface to light in visible (red, wavelength approximately 0.65 μm) and near-infrared (NIR, wavelength approximately 0.7 to 1 μm) spectral bands. The NDVI is defined as NDVI = (NIR−Red)/(NIR+Red). This quantity has been shown to be a useful index of the "vegetation greenness". More formally, NDVI is a (far from perfect) measure of two biophysical vegetation properties: (1) the green-vegetation cover fraction (v), the fraction of the surface covered by green or actively growing vegetation at nadir view, and (2) the green leaf area index (Λ) (Tucker 1979; Sellers 1985; Sellers *et al.* 1992; Lu *et al.* 2003). The relationship between these two biophysical quantities is complex, depending on factors such as leaf clumping and leaf angle distribution. However, a common simplification is to treat the canopy as an absorbing medium obeying Beer's law of absorption, which implies that

$$v = 1 - \exp\left(-c_{Ext}\Lambda\right)$$

where c_{Ext} is an extinction coefficient. This assumption is used here with $c_{Ext} = 0.5$, the theoretical value for isotropically oriented vegetation elements.

The relationship between the biophysical vegetation properties (either v or Λ) and remotely sensed variables such as NDVI is not unique and depends on a range of non-vegetation factors including sun and view angles and soil colour, in addition to sensor characteristics such as precise spectral wavebands of the Red and NIR channels. Roughly, the NDVI observed on a surface with no green vegetation ($v = 0$) is around 0.1 to 0.3 (with significant dependence on factors such as background soil colour), while the NDVI on a surface with 100% green vegetation cover ($v_c = 1$) is around 0.7 to 0.8 (Lu *et al.* 2003). NDVI values for the same land-surface location and time from AVHRR and other sensors (such as MODIS) are not identical, even with identical pixel sizes and sun and view angles, because of differing Red and NIR spectral wavebands.

Data Assessment: The overall spatial behaviour of vegetation greenness across the Australian continent is shown in Figure 3, as the mean and standard deviation for the period 1981-2002 of NDVI (from BPAL AVHRR; see Appendix B). The mean NDVI follows the continental rainfall distribution quite closely because vegetation in Australia is predominantly water limited. The standard deviation is largest in the cropping regions because of the annual forcing of land cover by management (so that standard deviation of vegetation greenness is a remotely-sensed index of cropping as a land use).

We have assessed the overall integrity of the NDVI time series from AVHRR (our primary source) by comparing long-term trends in large-area spatial averages of vegetation greenness, as measured by all the satellite sources summarised in Appendix B. Figure 4 shows such a comparison where the averaging area is the entire Australian continent (similar comparisons have been done for 17 smaller regions). Three groups of data are evident in this figure:

- AVHRR-NDVI is represented by the "BPAL" series (1981 to 2003 at 8 km resolution, resampled to 5 km for consistency with other spatial data), and by the "CATS2a" series (1992 to 2005 at 1 km resolution, with BRDF correction and cloud clearing; see Appendix B). In principle, the CATS2a series represents best currently available processing practice for AVHRR data.
- The data from MODIS, SeaWiFS and SPOT-VGT represent NDVI results from "modern" (1990s) environmental satellites with on-board calibration and active position control. (The NOAA satellites which carry the AVHRR sensors have neither of these attributes).
- The vegetation cover fraction series (from GlobCarbon; see Appendix B) is an estimated biophysical product, as opposed to a remotely sensed index, though it has been derived solely from remotely sensed data (see http://geofront.vgt.vito.be/geosuccess/ for algorithm details).

The following conclusions emerge from Figure 4.

- All time series show significant declining trends in vegetation greenness, especially over the period 2000 to 2003.
- More spatially explicit investigation (not shown here) suggests that the observed declining trend in vegetation greenness is most pronounced in arid and semi-arid regions (annual rainfall less than around 500 mm).
- The AVHRR data set suggests that this decline is continuing. This may be caused in part by calibration drifts in the afternoon-overpass AVHRR satellite active through this period (NOAA-16), but the question is under continuing investigation.
- The data from "modern" environmental satellites (MODIS, SeaWiFS SPOT-VGT) are closely grouped, except in some details such as discrepancies in SeaWiFS and SPOT-VGT before 2000.
- Differences in overall NDVI levels between this group of modern sensors and AVHRR are seen, and are expected because of differing Red and NIR spectral wavebands (AVHRR spectral filters are coarser than those in modern sensors). Despite this, there is a high degree of similarity in the patterns of temporal variability seen between AVHRR and the modern sensors.

3.2 River Discharge Data

Data Sources: River discharge data have been obtained from Dr Francis Chiew (e-Water CRC and CSIRO Land and Water). These data are presently available in archive form only, for model development and testing. The data represent monthly (sometimes daily) measurements of catchment discharge at outlet stream gauging stations, for around 300 "unimpaired" catchments for which water extraction and diversion within the catchment is minimal. The durations of the records are variable, ranging from a few years to nearly 100 years. About 200 of these gauging stations are providing ongoing data (as of December 2004), while for the others, historical but not contemporary records are available.

Data Interpretation: Because these catchment discharge data are restricted to unimpaired catchments as defined above, we interpret the measured catchment discharge as being equal to the sum of surface runoff and subsurface drainage into streams, spatially integrated across the catchment. This approximation neglects in-stream losses of water (such as evaporation and drainage from the stream itself) and assumes that the catchment outlet is the only route by which water leaves the catchment through horizontal transport.

Data Assessment: Results presented later (Section 4.2) will show the nature of these data and offer a comparison with model predictions.

3.3 Weather Data

Data Sources: The required weather data to drive the HTBDAS include daily rainfall, solar radiation, maximum and minimum temperature, gridded to a spatial resolution of 5 km. Two sources are available:

- Gridded data on daily rainfall, solar radiation, maximum and minimum temperature (5 km resolution, continental domain, time period 1980 to present) are being produced by the Bureau of Meteorology (BoM) as a component of the Australian Water Availability Project. At the time of this report, all variables are available except for solar radiation.
- Use is also being made of the SILO historic gridded weather dataset for Australia from the Queensland Department of Natural Resources, Mining and Environment. This covers all required variables (rainfall, solar radiation, maximum and minimum temperature, at 5 km resolution, continental domain, time period from pre-1957 to present, with decreasing accuracy before 1957).

Data Interpretation: For model runs to date, we have used the SILO archive because of the availability of solar radiation. Future runs will switch to the BoM dataset.

Data Assessment: Figure 5 compares the continentally and annually averaged rainfall, daily maximum and daily minimum temperature series from the BoM and SILO gridded weather archives. In addition to the gridded archives, the figure shows nongridded series from Torok and Nicholls 1996 for temperature, Lavery *et al.* 1997 for precipitation, and from the National Climate Centre (BoM) at http://www.bom.gov.au/cgi-bin/silo/reg/cli_chg/timeseries.cgi for both rainfall and temperature.

There is good agreement between all series at this coarse aggregation level, with excellent agreement between the two gridded data sets for all plotted variables. There is a small systematic difference of about 0.3 degC between the gridded and nongridded series for the annual mean of daily minimum temperature, with the nongridded series being higher.

4 Models

4.1 Dynamic and Observation Models

Two kinds of model are needed: a *dynamic model* and a set of *observation models*. The dynamic model predicts the landscape state, quantified by a set of state variables including soil moisture and other variables such as carbon stores. This is done using the water balance equation (Section 2) with specifications for all soil-water-dependent fluxes, together with similar equations for other state variables. The observation models predict what the assimilated observables (such as NDVI, LST, discharge) should be, from the landscape state predicted by the dynamic model. The difference between the predicted and measured observables is then used in the model-data fusion step to correct the model.

- *Dynamic model:* This is a simplification of the Bios and SimBios models, developed at CSIRO Land and Water and the CSIRO Earth Observation Centre (Raupach *et al.* 2001a; Raupach *et al.* 2001b). The dynamic model has two state variables: the soil water in a single store (W), and green leaf carbon (C_L). The model equations are summarised in Appendix C. They consist of two mass-balance equations for W and C_L respectively, together with phenomenological equations relating the fluxes of water and carbon in these mass-balance equations to the state variables (W, C_L), driving meteorological variables (daily rainfall, solar radiation, maximum and minimum temperature) and model parameters. The reasons for including green leaf carbon in the model are (1) a leaf vegetation estimate is needed to assimilate both NDVI and LST, and (2) vegetation greenness is a diagnostic of plant growth, a crucial quantity in the dynamics of the terrestrial biosphere and in particular for the phenomenological equation for transpiration.

- *Observation models:* These express the observables in terms of predicted quantities available from the dynamic model (the state variables W and C_L, and the water and carbon fluxes which depend on them). The observables for which observation models have been written are NDVI, land surface temperature (LST)[3], and catchment discharge. The present equations for the observation models are summarised in Appendix D.

The present formulation of both the dynamic and observation models is deliberately minimal (only essentials have been included). The basic reason for this approach is to allow the eventual level of complexity in both the dynamic and observation models to be an emergent property of the model-data fusion process, rather than being imposed at the beginning. The tension between simplicity and complexity in terrestrial biosphere modelling is discussed by (Raupach *et al.*

[3] LST is not being assimilated in the project, but the observation model has been written in WaterDyn.

2005a). In this methodology, "model failure" is demonstrated by irreconcilable discrepancies between model and data which cannot be removed by the model-data fusion process. Model failure in this sense is a good thing, as it precisely identifies areas where the dynamic and/or observation models are oversimplified and need to be increased in sophistication. More sophisticated algorithms can then be developed in these areas. Hence we are developing models of the right level of sophistication for this application by starting simple and working up, rather than starting from a complicated initial formulation and attempting to work down without empirical guidance as to where the model is over-complicated.

4.2 Prior information

Prior information includes all prior knowledge (expressed as values and uncertainties) brought to bear on the problem. It is of two kinds.

- *Prior estimates and constraints on process parameters:* These have been derived from the literature and previous experience with antecedents to the WaterDyn model (Raupach *et al.* 2001a; Raupach *et al.* 2001b). These estimates are not definitive, but when used with appropriate (often wide) uncertainty bounds, they provide valuable initial constraints on model process parameters such as water and light use efficiencies, or lower-level parameters that determine these aggregated model properties.

- *Information on soils and vegetation types:* Several soil and vegetation properties are required, such as the depth of available soil water and the decay time constant for water extraction. Initial estimates are derived from soil and vegetation type maps gathered under the National Land and Water Resources Audit (NLWRA 2001a; NLWRA 2001b). The original source of the soil data is the Atlas of Australian Soils (McKenzie and Hook 1992; McKenzie *et al.* 2000).

4.3 Model Setup

The dynamic and observation models are both embodied in an overall code called WaterDyn (for Water Dynamics), written in Fortran 95. The model time step is daily and the present spatial resolution is 0.05 degrees of latitude and longitude (approximately 5 km). The results presented here cover a spatial domain consisting of the entire southeast Australian region[4], defined as the Murray-Darling Basin and the adjacent coastal regions covered by the Southeast Coast Drainage Division (NLWRA 2001a; NLWRA 2001b). Figure 6 shows the spatial domain, identifying drainage basin boundaries, rivers and main towns. Figure 7 shows the gauged unimpaired catchments (see Section 3.2) within the spatial domain.

The model is run in one of two modes, "forward mode" and "assimilation mode". In forward mode, the model makes predictions of the observed quantities to be used for data assimilation but

[4] Our spatial domain covers (and extends beyond) a rectangular area in southeast Australia defined as the spatial domain for AWAP. We have used a domain based on drainage boundaries rather than rectangles because much of our model testing relies on catchment discharge data for which catchments provide the natural spatial boundaries.

these predictions are not actually assimilated. In assimilation mode, differences between predicted and actual observations are used to correct the model, as described in Sections 2 and 5. The results presented in this section are all from the model in forward mode; results in assimilation mode are given in Section 5.

For the forward runs, prior values of all model parameters were selected to give credible model performance for one particular catchment (Adelong, near the Tumbarumba flux tower site). These parameters were then applied unchanged across the spatial domain. The initial conditions for the state variables (W and C_L) were set to uniform values across the entire spatial domain.

To illustrate the typical model behaviour, Figure 8 shows time series for the meteorological driving data (top), the predicted model state variables W and C_L (bottom) and the water fluxes in the water balance (middle), for a 25-year run (1 Jan 1981 to 31 Dec 2005). All quantities in Figure 8 are averaged spatially across the Murrumbidgee Basin (the stippled area in Figure 7). The time series are shown at monthly-averaged time intervals for clarity, though the model runs at a daily time step and produces all outputs at daily time resolution.

In Figure 8, the first few months of the run stand out as anomalous because the model takes a few months to "forget" its assumed initial conditions for W and C_L and settle into a realistic pattern of interannual variability in W and C_L determined by the forcing meteorology (mainly rainfall). The strong trends near the beginning of the prediction period are a result of this settling down. To avoid this problem the model is normally "spun", that is, run twice through the 25-year period (1 Jan 1981 to 31 Dec 2005). The end values of W and C_L on 31 Dec 2005 from the first 25-year spin are used as realistic estimates of the initial conditions for the second spin, which provides the results actually used. (In practice, archived results for W and C_L at the end of the first spin are usually used to start the second spin, rather than running the model twice on every occasion).

4.4 Model Results in Forward Mode

Here we present the basic model predictions in forward mode across the entire spatial domain, for the 25-year period January 1981 to December 2005. Figure 9 shows a monthly map sequence of rainfall (the most variable meteorological forcing quantity) across this space and time domain. The temporal variability is made clearer by plotting the corresponding map sequence for rainfall anomaly, in Figure 10. (The anomaly in defined in the next paragraph). Figures 9 and 10 together show the high variability of Australian rainfall, and that much of this variability at the wet end arises from individual very wet months.

In general, we define the anomaly for a quantity as [current value of quantity] − [monthly mean value of quantity]. Hence, the anomaly is the quantity with the regular seasonal cycle subtracted out. In figure 10 and subsequent anomaly plots, grey denotes a zero anomaly (indicating that the plotted quantity at that location and time is average), red denotes a negative anomaly (indicating that the quantity is below average for that month), and blue denotes a positive anomaly (indicating an above-average value for the plotted quantity in that month). The anomaly plots are not normalised, so they have the same units as the quantity itself.

Figures 11 shows the relative soil moisture over southeast Australia as predicted by a 25-year forward run of WaterDyn, and Figure 12 shows the soil moisture anomaly (as defined above). The predicted relative soil moisture scales from 0 to 1 through the range from wilting point to

field capacity. Figures 11 and 12 show that there is strong interannual variability in soil moisture, superimposed on a regular annual pattern in which the relative soil moisture is high in late winter and spring, especially in the southeast highlands and southern Victoria. The soil moisture anomaly (Figure 12) varies much more smoothly than the rainfall anomaly (Figure 10) because of the lag imparted by soil moisture storage, which is of the order of months. A positive soil moisture anomaly is correlated with an antecedent positive rainfall anomaly over this time scale, rather than the current rainfall anomaly.

Dominant features of the interannual variability in soil moisture over the period 1981-2005 are the major droughts in 1982, 1994 and 2002 (in which relative soil moisture throughout the basin remains less than around 0.3 for the whole year except in the wet southeast corner), and the very wet years of 1983-84 and 1988-89. There are also the milder low-moisture anomalies in 1986, 1991 and 1997 and the whole period following the major 2002 drought to the end of 2005. These features are correlated with the ENSO (el Nino-Southern Oscillation) cycle in the equatorial Pacific Ocean. Three indices of ENSO are plotted in Figure 13: the Southern Oscillation Index (SOI) based on the Tahiti-Darwin air pressure difference, the Nino4 index based on central Pacific sea surface temperature, and the Multivariate ENSO Index (MEI) based on a combination of six atmosphere-ocean variables. All three indices agree well. Overall, dry soil moisture anomalies in southeast Australia align with strong el Nino events in the Pacific Ocean (positive values of the indices in Figure 13), and wet anomalies with strong la Nina events (negative indices in Figure 13). However, the agreement is not complete: for example, 2002 was drought with a near-record soil moisture negative anomaly (Figure 12) but was only a mild ENSO event, whereas 1997 was a near-record ENSO event but only a mild drought.

Figure 14 shows the total evaporation (plant transpiration plus soil evaporation), and Figure 15 its anomaly. The energy-driven annual cycle in total evaporation is clearly evident, but with wet and dry anomalies consistent with those for soil moisture. The evaporation anomalies in winter are substantially less than in other seasons because the evaporation itself is very much lower (recalling that the anomalies plotted here are absolute, not relative).

Figures 16 and 17 show the values and anomalies (respectively) of net primary production of plant carbon (NPP). The NPP is very closely coupled to total evaporation, as is expected in a water-limited environment. The closeness of this coupling is real but is perhaps a little overstated by the current version of the WaterDyn model, which assumes a highly simplified relationship between transpiration and NPP based on water use efficiency.

To give an idea of the relationship between the variations in the four quantities shown above (rainfall, soil moisture, evaporation and NPP), Figure 18 shows the annual mean of each quantity over each year from 1981 to 2005. The trends discussed above are apparent, though less dramatically in a comparison of mean annual quantities.

4.5 Model Tests in Forward Mode

The results presented in the last section are untested model predictions. To test them, we have compared the results of the discharge observation model in WaterDyn with actual observations of discharge from unimpaired catchments.

We focus first on the Murrumbidgee Basin (the stippled area in Figure 7) which contains 11 unimpaired catchments with sufficient discharge record durations in the period 1981 to 2005 for testing to be possible. Figure 19 shows a plot of predicted against measured catchment discharge for these catchments. The points represent average discharge over the whole of the available record for each catchment. Details of the catchments are given in the table accompanying Figure 19. For these 11 catchments, Figure 20 shows month-by-month comparisons of predicted and measured discharge. The scatter here is much higher, for at least two reasons: (1) the model does not include physics to enable an event-based prediction of storm hydrographs, and (2) the spatial rainfall pattern across these small catchments is spatially undersampled by the daily rainfall data at 5 km resolution, a problem which is more severe at short than long time scales.

Turning to the entire spatial domain, Figure 21 shows a comparison of predicted against observed average discharge for 153 unimpaired catchments for which discharge data are available. These catchments are mainly in the hilly southeast of the spatial domain where rainfall is medium to high (see Figure 7). Figure 21 reveals a systematic underestimate of discharge by the forward model with prior parameters, amounting to a low bias of order 100 mm/y in the estimate of discharge for catchments with discharge significantly above this value.

Figure 21 provides evidence of model failure (see discussion in Section 4.1). A possible cause of the problem is the use of a single water store, which means that surface runoff cannot occur until the whole of the single store has reached maximum water-holding capacity. To test this hypothesis, work in the immediate future will include the incorporation of multiple (initially two) water stores in the dynamic model, following earlier models (Raupach *et al.* 2001a; Raupach *et al.* 2001b) from which the present model was simplified.

Figure 21 also shows that the model-measurement disagreement has two components, one due to bias (discussed above) and the other due to scatter. The scatter component is comparable with or smaller than the bias.

5 Model-Data Fusion

5.1 Description

A central component and major innovation in this project is the model-data fusion or integration process. A basic description of this process has been given in Section 2. The main model-data fusion approach being used here is "sequential parameter estimation with the Ensemble Kalman Filter (EnKF)". This terminology can be unpacked to describe the model-data fusion approach in a little more detail.

- *Kalman Filter (KF):* The Kalman Filter is a group of algorithms for the sequential (step-by-step through time) combination of information from a dynamic model and a set of observations. Each step forward in time consists of two substeps: a prediction substep, in which the dynamic model is used to obtain a "prior state estimate" at the next time step, and a correction or analysis substep in which this prior estimate is corrected by comparison with observations to yield the final "posterior state estimate". The analysis step takes the form

$$\left[\begin{array}{c}\text{posterior}\\ \text{state estimate}\end{array}\right] = \left[\begin{array}{c}\text{prior state}\\ \text{estimate}\end{array}\right] + \left[\text{Kalman gain}\right]\bullet\left[\text{innovation}\right]$$

$$\left[\text{innovation}\right] = \left[\text{observation}\right] - \left[\begin{array}{c}\text{prior estimate}\\ \text{of observation}\end{array}\right]$$

The innovation is the difference between the actual observation and the observation as predicted from the prior state by using the observation model. The update to the prior state estimate is obtained by (matrix) multiplication of the innovation by the "Kalman gain matrix" **K**, a function of the prior state uncertainty, **P** (which depends on the error **Q** in the dynamic model) and the observation or measurement error, **R**. The formal functional dependence is **K** = **K**(**P**(**Q**), **R**). The uncertainties **P**, **Q** and **R** are all covariance matrices. The KF is now widely used (navigation, weather forecasting, earth system science, hydrology, finance, …) and has been called "the most useful piece of mathematics of the twentieth century" (Casti 2000). Outlines of the mathematical details are given by Grewal and Andrews 1993 and Enting 2002, and for the present context by Raupach *et al.* 2005b.

- *Ensemble Kalman Filter (EnKF):* The ensemble Kalman filter (EnKF) (Evensen 1994) is a KF implementation appropriate for nonlinear problems which are high-dimensional (involving many state variables). The present problem is both nonlinear (because of the inherent nonlinearity of the dynamic model; see Appendix C) and high-dimensional (because estimation of a common set of parameters across multiple land surface elements). The EnKF involves simultaneous solution of the dynamic model for an ensemble of typically 100 members which differ randomly because of the combined effects of initial state uncertainty ($\mathbf{P}_0$), model error (**Q**) and errors in past measurements (**R**). The current prior state covariance matrix (**P**) is obtained from the covariance of the ensemble.

- *Sequential parameter estimation:* In this implementation the "target variables" (the model variables being adjusted to give best fit between observations and model) are model parameters which are nominally time-independent. They are estimated sequentially, by refining current parameter estimates at each time step of the model run on the basis of model-measurement discrepancies. This is done by treating the parameters as extra elements of the state vector, so that the state being estimated in the above KF equation is an "augmented state vector" consisting of the union of the dynamic state vector and the parameters. Sequential parameter estimation has two advantages: (1) it maximises the predictive power of the model, since all quantities represented in the model can be predicted into the future using current best estimates of parameters; and (2) it allows for detection and estimation of changes in the nominally constant parameters through time (for instance due to a change in the system not accommodated in the model, such as land use change), as these changes are detected through model-measurement discrepancies.

In addition to the EnKF, we are using a more traditional down-gradient method of parameter estimation as a check on the EnKF. In this method, fixed parameter estimates are used for a complete run of the model, yielding a set of predicted observables which are then compared with the actual observations at all times simultaneously. Parameter optimisation then involves adjusting the parameters to minimise the sum of the squares of (observation − prior prediction), over all observations. This is a "batch" or "non-sequential" method of parameter estimation, in

contrast with the sequential parameter estimation with the EnKF described above. Differences between sequential and batch parameter estimation methods are outlined by Raupach *et al.* 2005b. To implement the down-gradient (batch) method here, we use the standard PEST (Parameter ESTimation) package.

5.2 Results

To date, our applications of model-data fusion to estimate parameters have focused on use of observations of monthly discharge from unimpaired catchments (Figure 7). We have not yet used other potential observations for assimilation (NDVI and LST) in the model-data fusion process. Therefore, in this section we report the results of assimilation of monthly discharge data only.

Five different parameters in the model were estimated, with all others being held at constant prior values. The parameters to be estimated were selected as those to which the model predictions are most sensitive, on the basis of past experience. The five estimated parameters are:

- RateEW: a specifier for the decay rate of water extraction by roots from a drying soil under water-limited transpiration (RateEW = k_E in Appendix C)

- RateLch: a specifier for the depletion rate of soil water by deep drainage (RateLch = k_D in Appendix C)

- PwrFWSoil: an exponent specifying the rate of decrease of soil evaporation with decreasing relative soil water (PwrFWSoil = β in Appendix C)

- PrwFWLch: an exponent specifying the rate of decrease of deep drainage with decreasing relative soil water (PwrFWLch = γ in Appendix C)

- ZSoilMult: a multiplier applied to the soil depth given by the maps of soil types and pedotransfer functions from the Atlas of Australian Soils (McKenzie and Hook 1992; McKenzie *et al.* 2000). This recognises that the model parameter for soil depth (the depth of extractable soil water, Z_W in Appendix C) is not the same as the soil depth in the Atlas, our source of information about spatially distributed soil properties. However, the Atlas soil depth is expected to provide some prior information about the spatial distribution of Z_W. We therefore assume that Z_W = ZSoilMult*(Atlas soil depth), where ZSoilMult is a spatially uniform multiplier. The Atlas provides depths of two soil layers (upper and lower), which are summed to obtain the Atlas soil depth.

The five parameters were estimated using the EnKF for sequential parameter estimation, with an ensemble of 100 members and with the WaterDyn model run in assimilation mode. The assimilated observable was measured monthly discharge in one or more unimpaired catchments. Parameters and model state variables were assumed to have log-normal distributions, meaning that log(quantity) has a Gaussian distribution. This assumption ensures that the EnKF cannot produce negative parameters or state variables in the model ensemble, which for most parameters and state variable would be physically unrealistic and likely to cause unpredictable model behaviour.

The results are presented in two stages: first we show parameter estimation results for one particular catchment (Adelong). Second, we estimate a uniform set of parameters for 9 different unimpaired catchments in the Murrumbidgee Basin, using discharge data from all 9 catchments together.

1. Parameters estimates for the Adelong catchment: Evolving estimates of the parameters themselves in the Adelong catchment are shown in Figure 22, with the heavy solid lines showing the optimal estimate of each parameter and the pair of dashed lines the parameter uncertainty (±1 standard deviation of log(parameter), transformed back to determine an uncertainty for the parameter itself). The parameter estimates depart rapidly from their initial (prior) values in the first year or two of the run, under the influence of information from the measurements which requires different optimum parameter values from the prior assumed values for the parameters. The parameter estimates then settle into a slow evolution, with the optimum values changing little but the uncertainties slowly decreasing as the information content from the measurements increases. The estimates change most rapidly when the EnKF receives significant new information from the observations, for example in major rainfall events; this causes the evolution of the estimates to have a stepwise character. The uncertainties are not symmetric on either side of the optimum estimate because of the use of log-normal distributions (see above).

Figure 23 compares the predicted and observed discharge for the Adelong catchment with the predictions being computed using the evolving parameter estimates shown in Figure 22. The agreement between predictions and observations is much improved by sequential parameter estimation, compared with the level of agreement obtained with prior parameter estimates (that is, without parameter estimation) shown in Figure 20.

The innovation (the difference between observations and predictions) is an indicator of the amount of new information being brought to the sequential parameter estimation by each successive observation. The innovation for this parameter estimation in the Adelong catchment is shown in Figure 24. It decreases with time through the 25-year run, for two reasons which are each responsible for about half of the decrease: (1) the parameter estimates progressively improve as data is assimilated, so that model performance becomes progressively better; (2) by chance in this 25-year record, the overall magnitude and variability in the observations decreases with time.

2. Uniform parameter estimates for 9 different unimpaired catchments, treated together: Having obtained satisfactory performance of the parameter estimation process for one catchment, the next step is to apply the process over a wider spatial domain. For this purpose, we used a domain consisting of the union of 9 unimpaired catchments in the Murrumdidgee Basin. (Of the 11 catchments shown in Figure 19 and 20, two were not used here because observation record lengths were not long enough).

Parameter values, held spatially uniform across all 9 catchments, were estimated by applying the EnKF with 100 ensemble members to assimilate observed monthly discharge from each catchment. The spatial domain for the WaterDyn model consisted of the union of all 9 catchments.

A problem encountered in this exercise was that occasionally one ensemble member would stray into physically unrealistic parts of state and parameter space. The most common example of this was that negative soil moisture could occur for certain parameter combinations, particularly at dry locations and times. The EnKF uses the logarithm of state variables, so is not able to handle negative soil moisture. The cure for this problem is to make the dynamic model for application in assimilation mode numerically robust, to standards that are much higher than for the model in forward mode (because forward-mode predictions only require the model to be robust in physically realistic parts of state and parameter space). The process of making the dynamic model meet these high standards of robustness is still going on at the time of writing. Therefore, in this report we present indicative results over a shortened run (1990 to 1997) for which all ensemble members were well behaved.

Figure 25 shows the parameter estimates obtained in this way. In comparison with the parameter estimates for the Adelong catchment (Figure 22), the uncertainties are significantly higher and the rate of convergence of parameter estimates is lower. This is a result of three factors: the shortened run duration (7 years as opposed to 25), the fact that we are applying spatially uniform parameters across catchments which differ physically and climatically in ways which lead to differences among optimum parameters estimated for each catchment separately, and the use of large assumed data uncertainties in this run to promote stability in the EnKF and minimise the problem caused by occasional stray ensemble members (see above).

Figure 26 compares the observed and predicted discharge for each of the 9 catchments. As in Figure 25, identical parameter values have been applied across all catchments. The agreement is fair but not as good as with parameters estimated and applied for a single catchment, as for the Adelong catchment in Figure 23. This shows that spatial smoothing of parameters leads to a loss of predictive accuracy.

6 Conclusions

The CMAR component of the AWAP project has met its primary objectives for Phase 1 of the Australian Water Availability Project (January 2005 to June 2006). Considering the model, data and assimilation components of the project in turn, specific achievements are as follows.

- *Model:* A water balance model has been developed, tested and applied to obtain histories of relative soil moisture and water balance fluxes (transpiration, soil evaporation, runoff and drainage) over southeast Australia for the 25 years from 1981 to 1995. Results have been expressed both as values of the soil water store and fluxes, and also as monthly anomalies of these quantities over the period 1981-2005.

- *Data:* Remotely sensed data for vegetation greenness have been assembled from several sources and intercompared, showing broadly consistent trends but some differences among different measures of essentially the same quantity. Discharge data have been assembled for over 200 unimpaired catchments. Gridded meteorological data (rainfall, solar radiation, daily maximum and minimum temperature at 5 km spatial and daily time resolution) have been gathered from two sources (the SILO archive from QDNRM and from the Bureau of

Meteorology as part of the AWAP project itself): intercomparison of these datasets shows excellent agreement at large spatial scale.

- *Assimilation:* The Ensemble Kalman filter (EnKF) has been developed and tested successfully for sequential parameter estimation of catchment discharge at the scale of individual unimpaired catchments and collections of catchments taken together. In addition, traditional (down-gradient) parameter estimation is being used as a check on the EnKF.

There are several remaining challenges and issues for future work. These are summarised here in the context of the extension of AWAP to Phase 2 for the period July 2006 to June 2007, with the central goal of implementing a prototype operational system for the determination of water availability in near-real-time across the whole of the Australian continent. To build on the achievements of AWAP Phase 1 and achieve the Phase 2 goal, there are challenges in each major area of the project, as follows.

- *Model:* Further improvements in the dynamic model will be in two main areas: (1) to increase numerical robustness for application in assimilation mode (see Section 5.2), and to explore the use of multiple water stores (see Section 4.4). A third possible area of improvement is the use of multiple carbon stores following Raupach *et al.* 2001a and Raupach *et al.* 2001b. For the observation models, improvements are necessary in the skill of the observation model for vegetation greenness. The observation model for land surface temperature (LST) requires testing.

- *Data:* The main data requirements for Phase 2 are (1) to test multiple sources of LST data, as already done in Phase 1 for vegetation greenness; and (2) to obtain all necessary data streams in near-real-time. The latter requires close cooperation among all AWAP participants.

- *Assimilation:* The challenges are to overcome some technical issues which still remain with assimilation of discharge (See Section 5.2), and to extend the assimilation process to continental scale with assimilation of remotely sensed data, both for vegetation greenness and for LST.

Acknowledgments

The work outlined here, while primarily the direct effort of the authors of this report, has benefited enormously from cognate projects by close colleagues. In particular, we acknowledge the work by Matt Paget, Michael Schmidt and Tim McVicar in developing the CATS series. We are grateful for the generous access to the CRCCH catchment discharge data provided by Dr Francis Chiew (now of CSIRO Land and Water). As ever, discussions with and encouragement from close colleagues – particularly Helen Cleugh, Pep Canadell and Damian Barrett – has been important. We are also grateful to our AWAP colleagues in BoM (David Jones and Neil Plummer) and BRS (David Barratt, John Sims) for their continued engagement, skill and good humour.

Appendix A: Australian Water Availability Project Payment Schedule

The table shows the milestone schedule for the CSIRO components of AWAP, including contributions from both CSIRO Land and Water (CLW) and the Earth Observation Centre (EOC), now part of CSIRO Marine and Atmospheric Research (CMAR). The present report describes only the EOC-CMAR contribution.

Milestone	Date	Unit	Payment
Project commencement Phase	on signing of MOU	EOC, CLW	$50,000
AWAP 01 Report on the proposed methods for generating remotely sensed moisture availability spatial products	30 April 2005	CLW	$50,000
AWAP 02 Report on the specifications and outputs of a full terrestrial-biosphere data assimilation system for water balance	30 April 2005	EOC	$50,000
AWAP 02.1 Report on observation models relating the variables represented in the dynamic model to observed quantities, and a model-data synthesis scheme based on the Ensemble Kalman Filter or similar method	30 September 2005	EOC	$25,000
AWAP 01.1 Report on improved methods to downscale meteorological data from daily observations to time-of-overpass	28 February 2006	CLW	$25,000
AWAP 02.2 Report on evaluation of the complete HTBDAS for approximately 5 selected sub-catchments (ANRA Basins) from within the study area, including a simple demonstration of the forecast capabilities of the HTBDAS using "retrospective" seasonal forecasts	31 March 2006	EOC	$25,000
AWAP 03 Draft final report of a prototype near-real-time operational system for direct interpretation of soil moisture from remotely sensed data	31 March 2006	CLW	$25,000
AWAP 04 Final comprehensive project report including sections on methodologies for linking available soil water and water balance spatial products to seasonal forecasting models, and possible future water availability scenarios	30 June 2006	EOC, CLW	$50,000

Appendix B: Details of Remote Sensing Data Sources

This Appendix summarises information on several remote sensing data sources used here. These include data from the following sensors: (1) NOAA-AVHRR; (2) MODIS, (3) SeaWiFS, (4) SPOT-VGT. Of these, the NOAA time series is the only one operationally supported for the next decade. In addition we have used data from (5) the GlobCarbon project, a compilation by the European Space Agency of satellite-based vegetation data from several sources. Details are summarised in Table B1.

NOAA-AVHRR: The NOAA polar-orbiting satellites have carried AVHRR instruments operationally since the launch of NOAA-6 in 1979. The AVHRR sensor records data in 5 spectral bands of the electromagnetic spectrum: (1) red (580-680 nm); (2) Near Infrared (NIR) (725-1100 nm); (3) 3.55-3.93 μm; (4) 10.5-11.3 μm; and (5) 11.5-12.5 μm. The spacecraft are in sun-synchronous polar orbits of approximately 100 min duration at an altitude of about 700km. Each orbit comprises an ascending and a descending component corresponding to whether the spacecraft is travelling northwards or southwards respectively. The overpass time of the ascending node is nominally around 1330 local solar time but changes slowly with orbital drift (typically at 0.25 to 0.5 h/y). The spatial resolution of AVHRR data is 1.1 km at nadir (the point directly beneath the satellite), increasing to 5.4 km at the edge of the swath where the scan angle is 55°. For details see Cracknell 1997.

The CSIRO AVHRR archive is maintained by CSIRO Marine and Atmospheric Research. Between 1981 and 1986 the basic data are coarse-resolution (about 8 km at nadir) Global Area Coverage (GAC) data Cracknell 1997 provided by NOAA. These data are also used to supplement the limited full resolution data from 1986 until 1992 when the data began to be comprehensively archived from direct broadcasts from the NOAA satellites received in Australia. Since 1992, the data from a number of Australian reception stations have been combined by stitching the different segments from each station to eliminate redundancy and produce a single best-quality scene for each overpass (King 2000; Lovell *et al.* 2003; King 2003). The daily coverage in this archive is an area of approximately 50 million km^2, including the entire Australian land surface and surrounding regions to at least 2000km from the Australian coast. Since 1992, coverage of this area has been obtained four times daily.

The Australian AVHRR archive is available in two forms, both used in this work.

1. The "BPAL AVHRR" archive has been compiled from various sources and processed to produce complete, cloud-free, calibrated, geolocated, continental coverage of all AVHRR channels as seen by the afternoon overpass, at 0.05 degree (about 5 km) spatial resolution, covering land only, for 1981 to present. Compositing (maximum-NDVI), to approximately 10-day time resolution, was used as a first-order means of removing cloud effects. Additional "BISE" (Best Index Slope Extraction) filtering was used to further reduce cloud contamination and the effects of variations in view and sun angles (Lovell and Graetz 2001). The "BPAL" terminology arises because this archive extends a series available from NASA for the period 1981-1994 called the PAL (Pathfinder AVHRR Land) dataset, using BISE filtering. The BPAL AVHRR archive is used here for analysis of land condition.

2. The "CATS" (CSIRO AVHRR Time Series) archive is currently available from 1992 to present. This archive includes all AVHRR channels at a nominal spatial resolution of 1.1 km at nadir and temporal resolution of up to four overpasses per day. The data are calibrated and geolocated, with other processing in several versions described in Table B2.

MODIS: The Moderate Resolution Imaging Spectroradiometer (MODIS) is a key instrument aboard the Terra and Aqua satellites, launched in December 1999 and March 2002, respectively. The orbit of Terra around the Earth is timed so that it passes from north to south in the morning (about 1030 local solar time), while Aqua passes south to north in the afternoon (about 1330 local solar time). Terra MODIS and Aqua MODIS each view the entire Earth surface every day, acquiring data in 36 spectral bands ranging from 405 nm to 14.4 μm. The spatial resolution of MODIS data is 250 m for one spectral band in the visible and one in the NIR, 500 m in 5 visible to mid-infrared bands and 1000 m in all other bands. The data used in this study are standard global products obtained from NASA by DLT tapes (due to the enormous amount of data) and subsequently uploaded to the CSIRO MODIS Data Storage Cluster. These standard products may also be downloaded by FTP from the NASA website.

SeaWiFS: The Sea-viewing Wide Field-of-view Sensor (SeaWiFS) is carried on the SeaStar satellite launched on August 1, 1997 as part of NASA's "Mission to Planet Earth". The SeaStar maintains a sun-synchronous 705 km altitude orbit, with a north-to-south equatorial crossing at 12:20 local solar time, covering the Earth's surface once a day. The SeaWiFS sensor has 8 bands in the 402 nm (violet) to 885 nm (NIR) range. It differs from the AVHRR sensor in that it can tilt to avoid sunglint on the sea. SeaWiFS transmits local area coverage (LAC) data in real-time at a spatial resolution of 1.1km, with global area coverage (GAC) data archived and transmitted at 4.5 km resolution. The data used in this project are the SeaWiFS level 3 monthly NDVI product at 9 km resolution. Information about the SeaWiFS project can be obtained from http://oceancolor.gsfc.nasa.gov/SeaWiFS/

SPOT-VGT: The "Vegetation" (VGT) instrument is a wide-field sensor carried as part of the SPOT 4 and 5 satellite payloads launched on March 24, 1998 and May 4, 2002. The SPOT 4/5 satellites maintain a sun-synchronous polar orbit at ~830 km altitude. The Vegetation instrument has 4 non-contiguous bands in the visible, NIR, and MIR range (430-1750 nm), with a swath width of 2250 km at 1.165 km spatial resolution, allowing 90% global coverage in one day. Several products are available, including daily and ten-day synthesis products (S10) at full resolution as well as 4 km and 8 km reduced resolutions. The VGT images are processed and archived by the Belgian research institute VITO. The data used in this project are the SPOT VGT-S10 NDVI series at 1 km resolution.

GlobCarbon: The GlobCarbon project is part of the Global Terrestrial Observing System coordinated by the FAO. GlobCarbon uses data supplied by the European Space Agency to produce a range of fully calibrated satellite estimates of global land surface properties (fire, albedo, fAPAR, LAI, vegetation growth cycle) which are nearly independent of the original data source. The focus of the project is on the seven years 1997 to 2003, a period of overlap between various satellite measurements. GlobCarbon products and services are managed by VITO and various other European agencies, and distributed by VITO at http://geofront.vgt.vito.be/geosuccess/. The GlobCarbon data available for this project were monthly LAI from 1999 to 2002, generated from SPOT VGT and ATSR data.

Sensor	Agency (Country)	Platform (Launch)	Swath (km)	Revisit time (overpass time)	Spectral Bands	Nadir spatial resolution (m)
AVHRR		NOAA (numerous satellites)	2500	1 day (nominal: 0900, 1400)	5 bands (visible, NIR, thermal)	1100
MODIS	NASA (USA)	Terra (Dec 1999) Aqua (May 2002)	2330	1-2 days (Terra: 1030) (Aqua: 1330)	36 bands (visible, NIR, MIR, thermal)	250 to 1000
SeaWiFS	NASA (USA)	SeaStar (Aug 1997)	2801 (LAC), 1502 (GAC)	1 day (1220)	8 bands (visible, NIR)	1100 (LAC) 4500 (GAC)
VGT	CNES (France)	SPOT 4 / 5 (Mar 1998 / May 2002)	2250	1-2 days	4 bands (visible, NIR, MIR)	1165

Table B1: Details of satellite sensors providing data utilised in this work

CATS Version	Orbit	Time Res	Calibration	Cloud Corr	BRDF Corr	Compositing	Spatial Res	Time Span
BPAL	pm	8-11 day	hybrid	BISE		Max NDVI	8km	81/07 - 04/12
1	pm	mthly	Calwatch		Polder	Max NDVI	1km	92/04 - 04/12
2a	pm	mthly	Calwatch	CLAVR	Polder	Max NDVI	1km	92/04 - 04/12
2b	pm	mthly	Calwatch	CLAVR		Max NDVI	1km	92/04 - 04/12
4	am	mthly	Calwatch	CLAVR	Polder	Max NDVI	1km	92/04 - 04/12
5	am	mthly	Calwatch	CLAVR		Max NDVI	1km	92/04 - 04/12

Table B2: Versions of the CSIRO AVHRR Time Series.

Appendix C: Summary of Water and Leaf Carbon Model

This Appendix summarises the equations currently used in the WaterDyn dynamic model for soil water and green-leaf carbon (October 2005, WaterDyn06M).

State Variables and Balance Equations

The model has two state variables, a soil water store W [mol-water m^{-2}] and a green-leaf carbon store C_L [molC m^{-2}]. Corresponding dimensionless variables are the relative soil water w and vegetation cover fraction or green-leaf cover v (both between 0 and 1), respectively related to W and C_L by

$$w = W/(\rho_W Z_W) \tag{1}$$

$$v = 1 - \exp(-c_{Ext}\Lambda) = 1 - \exp(-c_{Ext}C_L/C_{L0}) \tag{2}$$

In Equation (1), ρ_W is the density of liquid water [mol-water m^{-3}] and Z_W the depth of the extractable soil water [m-liquid-water]. In Equation (2), c_{Ext} is the exponential light extinction coefficient in the canopy, Λ the leaf area index and C_{L0} the green-leaf carbon store at $\Lambda = 1$, given by $C_{L0} = d_L\rho_{CL}$, where d_L is the leaf thickness and ρ_{CL} the density of carbon in green leaf [molC m^{-3}].

The water and carbon balance equations governing W and C_L are

$$\frac{dW}{dt} = \rho_W Z_W \frac{dw}{dt} = \underbrace{F_{WP}}_{\text{Precipitation}} - \underbrace{F_{WT}}_{\text{Transpiration}} - \underbrace{F_{WS}}_{\text{Soil Evaporation}} - \underbrace{F_{WR}}_{\text{Runoff}} - \underbrace{F_{WD}}_{\text{Deep Drainage}} \tag{3}$$

$$\frac{dC_L}{dt} = a_L \underbrace{F_{CP}}_{\text{NPP}} - \underbrace{k_L C_L}_{\text{Leaf decay}} \tag{4}$$

In Equation (3), all water fluxes (F_W) are in metres of water per day [mWater day^{-1}]. In Equation (4), F_{CP} is the plant carbon production flux or net primary productivity [molC m^{-2} day^{-1}], a_L the allocation coefficient for growth carbon to leaf, and k_L the decay rate for leaf carbon [day^{-1}].

The model uses standard MKS units, but with time in days rather than seconds.

Phenomenological Equations for Water Fluxes

The phenomenological equations governing the water fluxes in Equation (3) are as follows.

(1) *Precipitation* (F_{WP}) is an external input.

(2) *Transpiration* (F_{WT}) is the lesser of an energy-limited and a water-limited transpiration rate:

$$\begin{aligned} F_{WT} &= \min\left(F_{WT}\left(\text{energy-lim}\right), F_{WT}\left(\text{water-lim}\right)\right) \\ F_{WT}\left(\text{energy-lim}\right) &= \nu F_W\left(\text{PT}\right) \\ F_{WT}\left(\text{water-lim}\right) &= \nu k_E Z_W w = \nu k_E \frac{W}{\rho_W} \end{aligned} \tag{5}$$

Here F_W(PT) is the Priestley-Taylor evaporation rate [mWater day^{-1}], and k_E is a rate [day^{-1}] for the decay of water extraction by roots from a drying soil under water-limited transpiration.

Priestley-Taylor evaporation is a thermodynamic estimate of the energy-limited evaporation rate for the whole surface (vegetation plus soil). From Raupach 2000 and Raupach 2001, it is

$$F_W\left(\text{PT}\right) = \frac{c_{PT}\Phi_{Eq}}{\rho_W \lambda_W} \tag{6}$$

where ρ_W is the density of liquid water [molWater m^{-3}], λ_W is the latent heat of vaporisation of water [J mol-water^{-1}], Φ_{Eq} is the thermodynamic equilibrium latent heat flux [J m^{-2} day^{-1}], and c_{PT} is the Priestley-Taylor coefficient, a number which is well constrained at about 1.26 Priestley and Taylor 1972; Raupach 2001. The equilibrium latent heat flux is given by

$$\Phi_{Eq} = \frac{p\varepsilon\Phi_A^*}{p\varepsilon + 1} \tag{7}$$

where Φ_A^* is the isothermal available energy flux [J m^{-2} day^{-1}], ε is the ratio of latent to sensible heat content of saturated air (2.2 at 20 deg C, roughly doubling with each 13 deg C temperature increase) and p is a number slightly less than 1 accounting for radiative coupling, defined in the next equation. The isothermal available energy flux Φ_A^* is given by

$$\Phi_A^* = \left(1-a\right)\Phi_{S\downarrow} + e\left(\Phi_{L\downarrow} - \sigma T_a^4\right); \qquad p = \frac{G_a}{G_a + G_r} \tag{8}$$

$\Phi_{S\downarrow}$ and $\Phi_{L\downarrow}$ are the downward shortwave and longwave irradiances; a and e are whole-surface albedo and emissivity, respectively; σ is the Stefan-Boltzmann constant; T_a [degK] is the air temperature at a reference height; G_a is the aerodynamic conductance; $G_r = 4e\sigma T_a^3/(\rho_A c_{PA})$ is the radiative conductance; ρ_A is the density of air c_{PA} is the specific heat of air at constant pressure.

In actual model coding, the energy fluxes (denoted Φ above) and mass fluxes (denoted F) are calculated separately for the day (sunlit) and night parts of each 24-hour day, and then summed.

(3) *Soil evaporation* (F_{WS}) is given by

$$F_{WS} = (1-v)\, w^{\beta} F_W (\mathrm{PT}) \tag{9}$$

where β is an exponent specifying the rate of decrease of soil evaporation with decreasing relative soil water w.

(4) *Runoff* (F_{WR}) is given by

$$F_{WR} = F_{WP}\, \mathrm{Step}(w-1) \tag{10}$$

so that $F_{WR} = F_{WP}$ (all precipitation runs off) when the soil is saturated ($w = 1$), and there is no runoff otherwise.

(5) *Deep drainage* (F_{WD}) is given by

$$F_{WD} = k_D Z_W w^{\gamma} \tag{11}$$

where γ is an exponent specifying the rate of decrease of deep drainage with decreasing relative soil water w, and k_D is a rate [day^{-1}] for the depletion of soil water by deep drainage.

Phenomenological Equations for Carbon Fluxes

In Equation (4), the phenomenological equation for the plant carbon production flux or net primary productivity (F_{CP}) is

$$F_{CP} = \left[\left(\alpha_Q v F_Q\right)^{-1} + \left(\alpha_W \rho_W F_{WT}\right)^{-1} \right]^{-1} \tag{12}$$

where F_Q is the incident quantum flux of photosynthetically active radiation (PAR) on the surface [mol-quanta m^{-2} day^{-1}], and α_Q and α_W are respectively a PAR use efficiency [molC mol-quanta^{-1}] and a transpired-water use efficiency [molC mol-water^{-1}]. The leaf allocation coefficient responds to soil water through

$$a_L = \frac{\sqrt{w}}{\sqrt{w} + \sqrt{w_0}} \tag{13}$$

where w_0 is the relative soil water at which $a_L = 0.5$. Equations (12) and (13) are taken from a new analysis of carbon allocation using ecological optimality principles (Raupach 2005).

Appendix D: Initial Observation Model

This Appendix summarises the equations currently used in the WaterDyn observation model (March 2006, WaterrDyn07M).

(1) *Normalised Difference Vegetation Index (NDVI):* The NDVI (N) is assumed to be linearly related to the green-leaf cover fraction (v), through

$$N = N_0 + v\frac{N - N_0}{N_1 - N_0} \tag{14}$$

where N_0 and N_1 are NDVI values for bare soil and full canopy cover, respectively.

(2) *Land Surface Temperature (LST):* The observation model for LST (T_s) needs to take into account the difference between conditions at time of satellite overpass (t_p) and the average diurnal conditions described by the dynamic model (Appendix B). We first relate T_s at time t_p to the air temperature and sensible heat flux (Φ_H) at that time:

$$T_s\left(t_p\right) = T_a\left(t_p\right) + \frac{\Phi_H\left(t_p\right)}{\rho_a c_p G_a\left(t_p\right)} \tag{15}$$

The instantaneous (t_p) values of T_a and Φ_H can be related to model (diurnally averaged values through the following simple, empirical assumptions:

$$T_a\left(t_p\right) = 0.5\left(T_{a\max} + T_{a\min}\right) + 0.5\left(T_{a\max} - T_{a\min}\right)\cos\left(\frac{2\pi\left(t_p - t_{Ta\max}\right)}{t_{day}}\right) \tag{16}$$

$$\Phi_H\left(t_p\right) = 2\cos\left(\frac{\pi\left(t_p - t_{noon}\right)}{f_{daylight}T_{day}}\right)\Phi_H\left(\text{diurnal average}\right) \tag{17}$$

where $t_{a\max}$ is the time of maximum temperature in the day, T_{day} the 24-hour day duration and $f_{Daylight}$ the daylight fraction of the 24-hour day. (Note: these simple expressions represent starting assumptions only, and may be improved and made more complicated if necessary).

Finally, the (diurnally averaged) sensible heat flux is related to the other energy fluxes in the surface energy balance, as

$$\Phi_A^* = \Phi_E + \frac{\Phi_H}{p} \tag{18}$$

where notation follows Appendix B. Thus $T_a(t_p)$ is expressible in terms of the latent heat flux, and thence the total evaporation, since

$$\Phi_E = \rho_W \lambda_W \left(F_{WT} + F_{WS} \right) \tag{19}$$

The end result is an expression for $T_s(t_p)$ in terms of meteorological forcing variables, diurnally averaged fluxes and state variables available to the dynamic model.

(2) *Catchment outflow at monthly scale:* Initially, this is described as the sum over a month and a catchment area of the surface runoff and drainage from each land element in the catchment:

$$\text{Catchment Outflow} = \int_{\text{month}} dt \sum_{\text{catchment}} \left(F_{WR} + F_{WL} \right) \tag{20}$$

To avoid difficulties with water extractions from rivers (farm dams, irrigation, offtakes etc) we use only data from nominally unimpaired catchments as identified by Francis Chiew (e-Water CRC). These are small catchments for which lags between elemental-area water fluxes and catchment outflow can initially be assumed to be not significant at monthly time scale.

Some of the assumptions in this simple model may be violated in some catchments. Where significant parameterisation problems appear to be attributable to this cause, a more sophisticated observation model for catchment outflow will be sought through interactions with colleagues in CSIRO Land and Water and elsewhere.

Figures

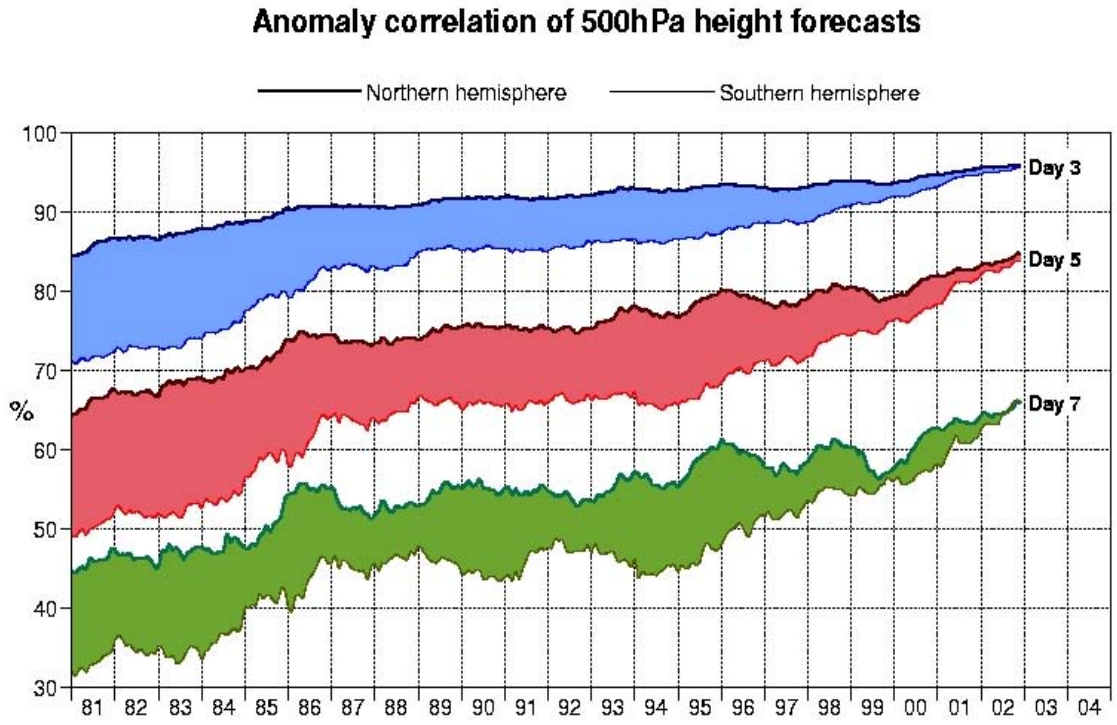


Figure 1: Evolution of 3-day, 5-day and 7-day weather forecast skills for northern (NH) and southern (SH) hemispheres, 1981-2002. The upper and lower lines bounding each coloured region are the skills for the NH and SH, respectively. The overall improvement over the period 1981 to 2001 is due to (1) better weather models; (2) better data, largely from remote sensing; and (3) continuous assimilation of data into models. The importance of remotely sensed data is made clear by the convergence of the NH and SH skills by 2001, since remote sensing adds relatively more information in the SH than in the NH because of the sparseness of SH *in-situ* observations. (Courtesy A. Hollingsworth, ECMWF; from a presentation "Goals of GEMS" at the Integrated Global Carbon Observation (IGCO) Implementation Workshop, Frascati, Italy, 3-5 November 2004).

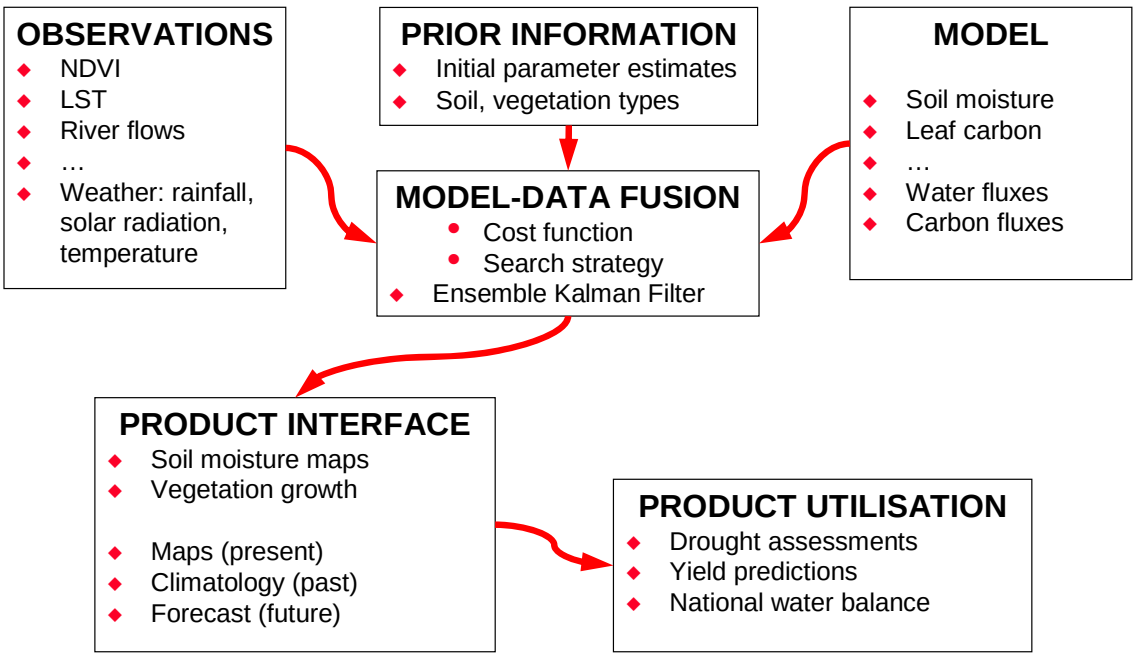


Figure 2: Schematic representation of components of a hydrological and terrestrial-biosphere data assimilation system (HTBDAS).

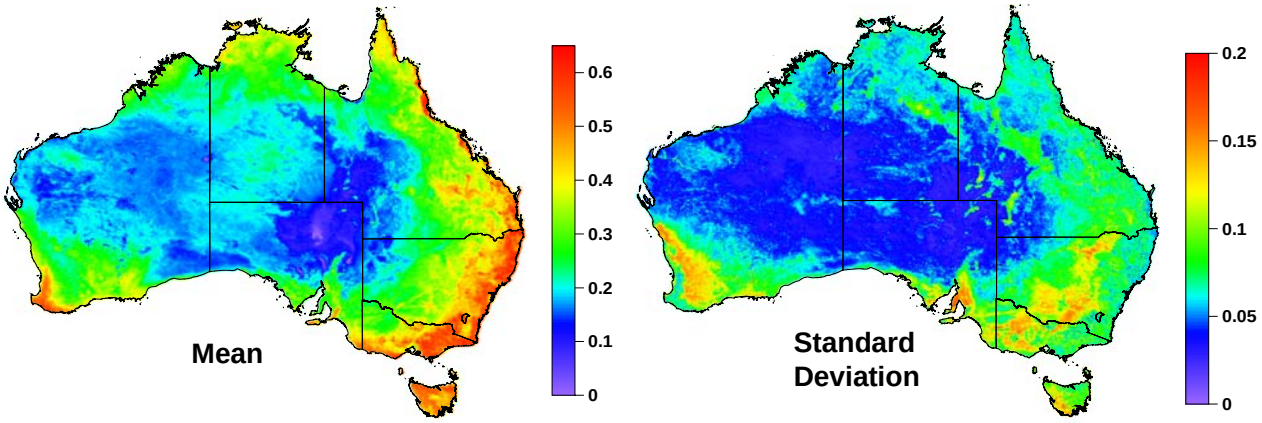


Figure 3: Mean and standard deviation of NDVI from BPAL AVHRR (1982-2002 excluding 1994). See Appendix B for details of BPAL AVHRR series.

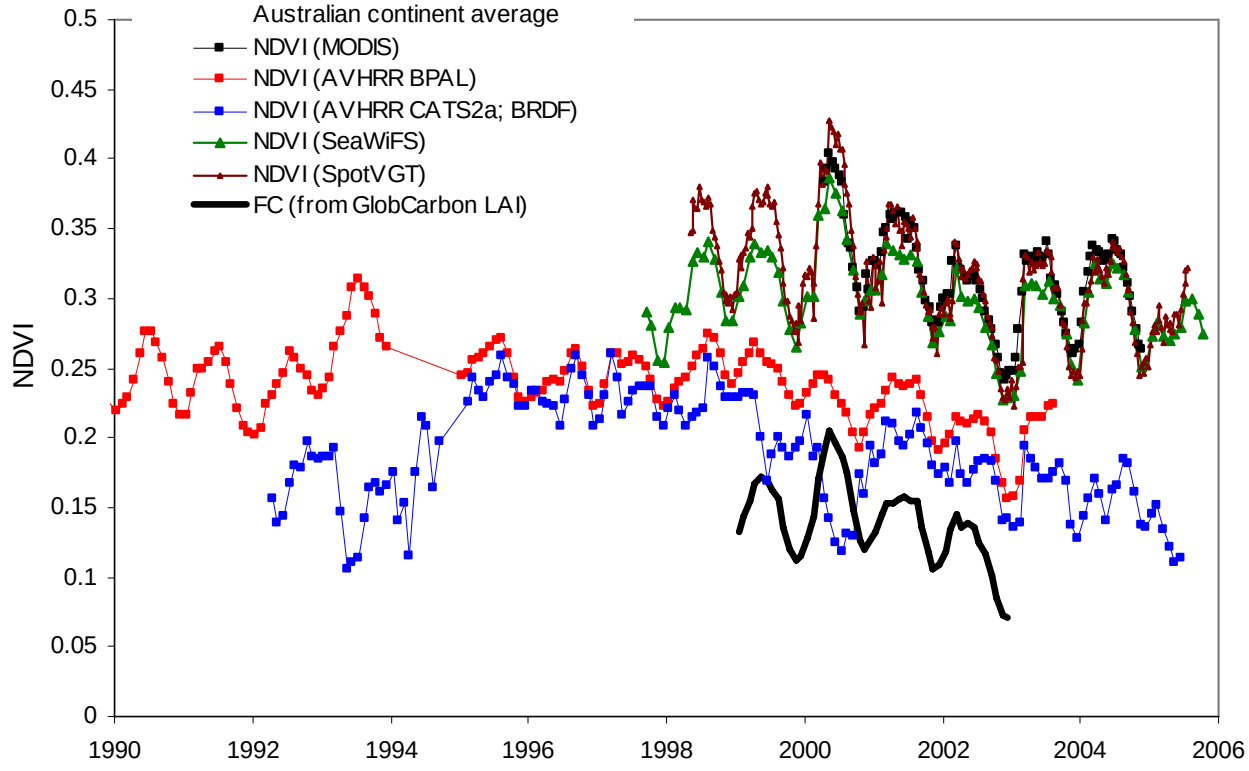


Figure 4: Time series of continentally averaged measures of vegetation greenness (NDVI vegetation fraction cover) from several data sources described in Appendix B.

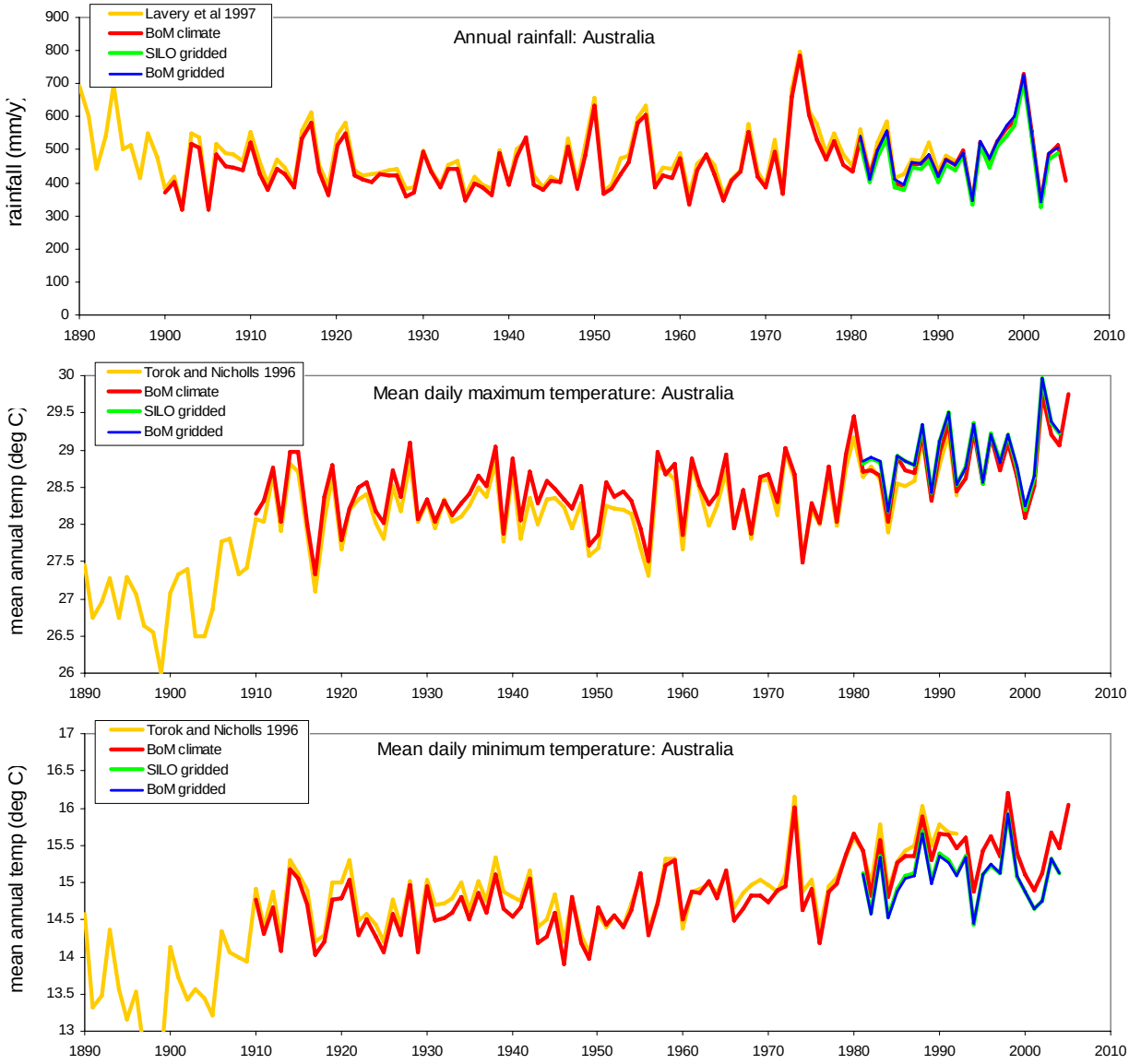


Figure 5: Comparison of several datasets for continentally and annually averaged Australian rainfall (top), daily maximum temperature (middle) and daily minimum temperature (bottom). The "SILO gridded" and "BoM gridded" series are from the SILO and BoM gridded (daily, 5 km) weather archives. The "BoM climate" series are from the continentally averaged data from the National Climate Centre (http://www.bom.gov.au/cgi-bin/silo/reg/cli_chg/timeseries.cgi) at the Bureau of Meteorology. Also shown are published continentally and annually averaged series for temperature Torok and Nicholls 1996 and rainfall Lavery *et al.* 1997

Figure 6: The spatial domain for this work, consisting of the Murray-Darling Basin and South-East Coast Drainage Division in South Eastern Australia. Major rivers and towns are shown. Olive borders and numbers identify NLWRA Drainage Basins (NLWRA 2001a; NLWRA 2001b).

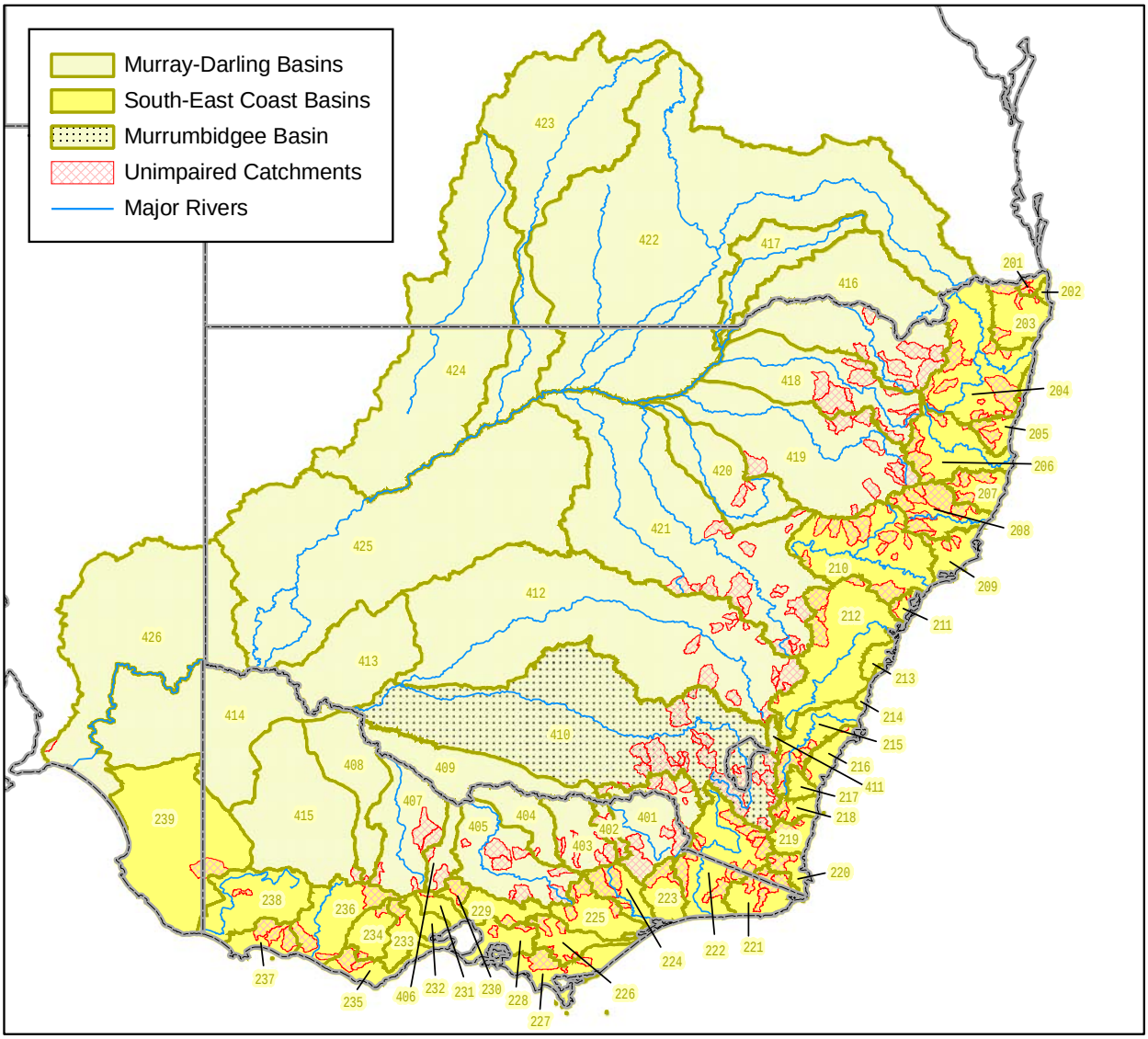


Figure 7: Location of unimpaired catchments in South Eastern Australia (hatched red areas), superimposed on the spatial domain map (Figure 6). The Murrumbidgee basin (410) is stippled in black.

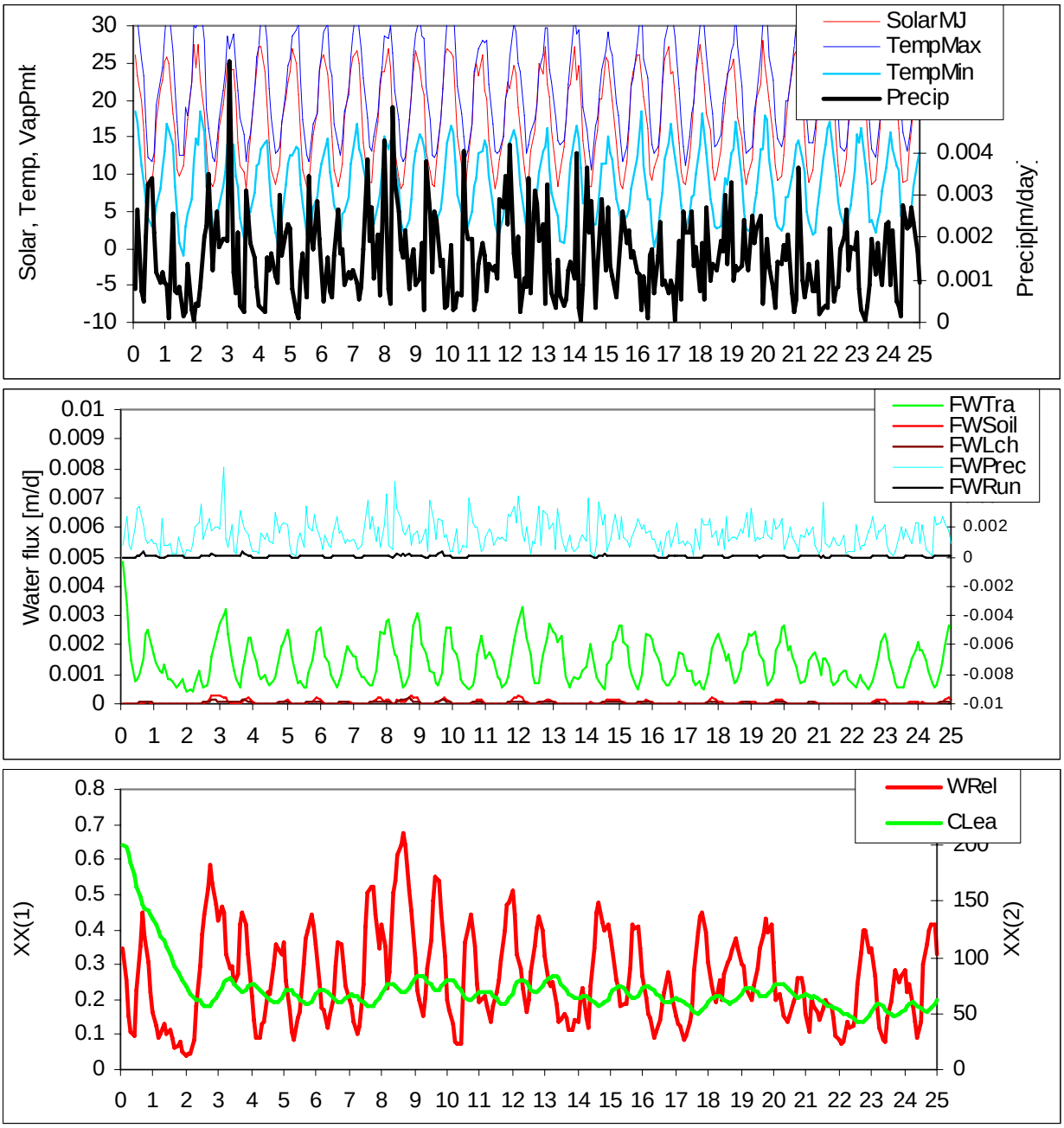


Figure 8: Spatially averaged (Murrumbidgee Basin) monthly time series for a 25-year run of the forward model: meteorological forcing (top panel); model output of water balance terms (middle panel); model output of state variables (XX(1) = relative soil moisture w, XX(2) = green-leaf carbon store C_L (molC m^{-2}) (bottom panel). Model version: WaterDyn09Ma.

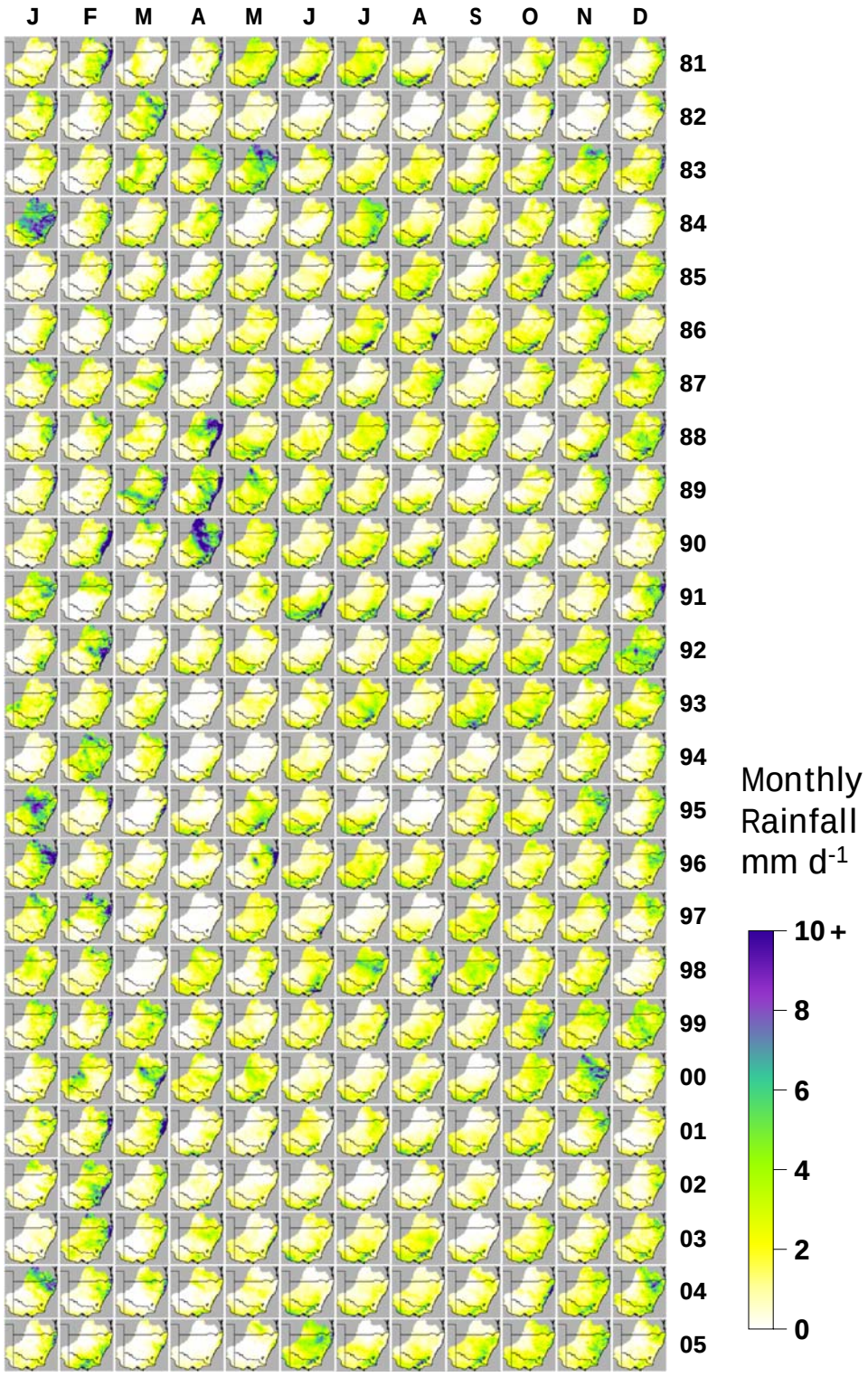


Figure 9: 25-year time series of monthly rainfall over Southeast Australia.

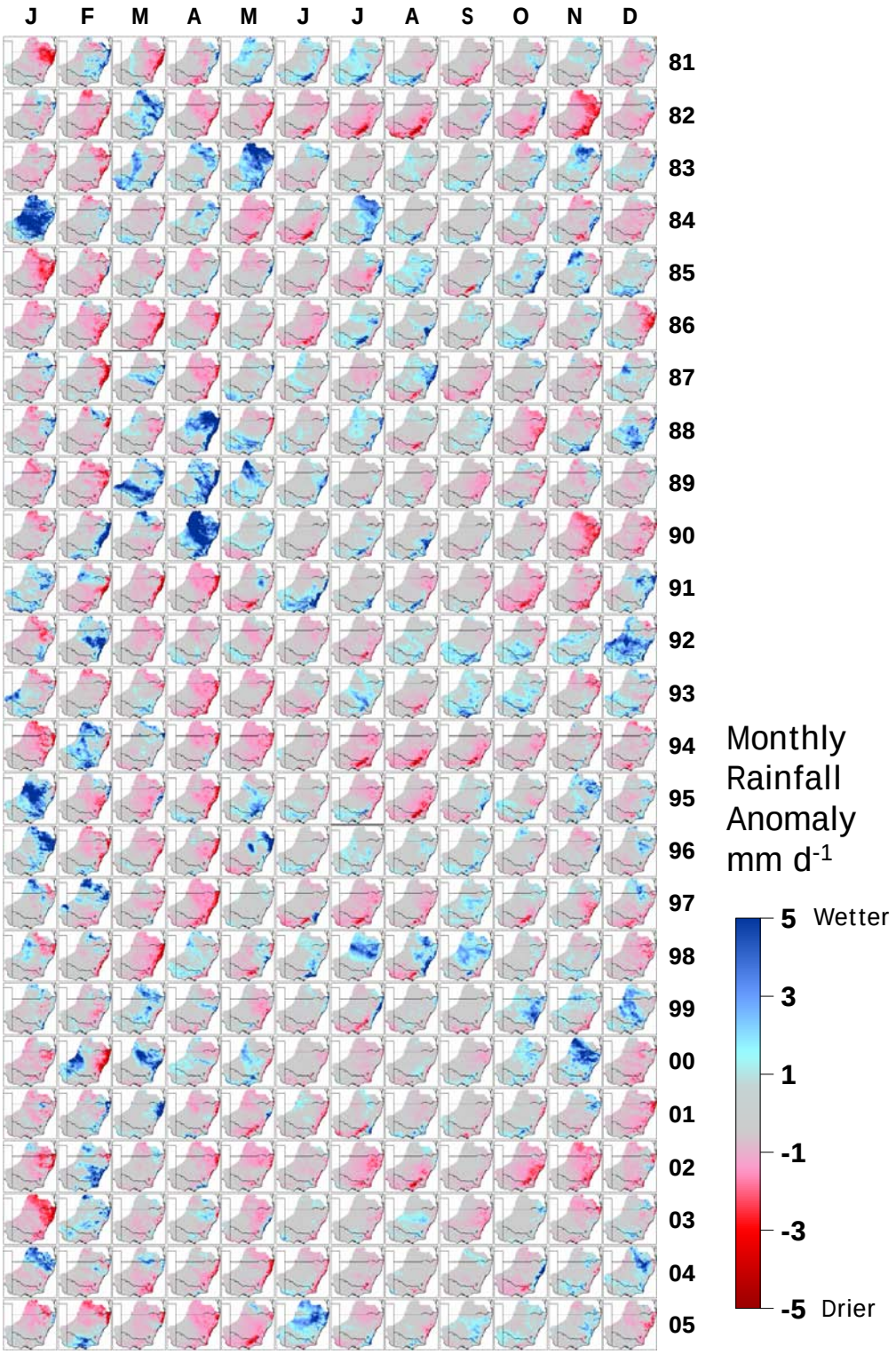


Figure 10: 25-year time series of monthly rainfall anomaly over Southeast Australia.

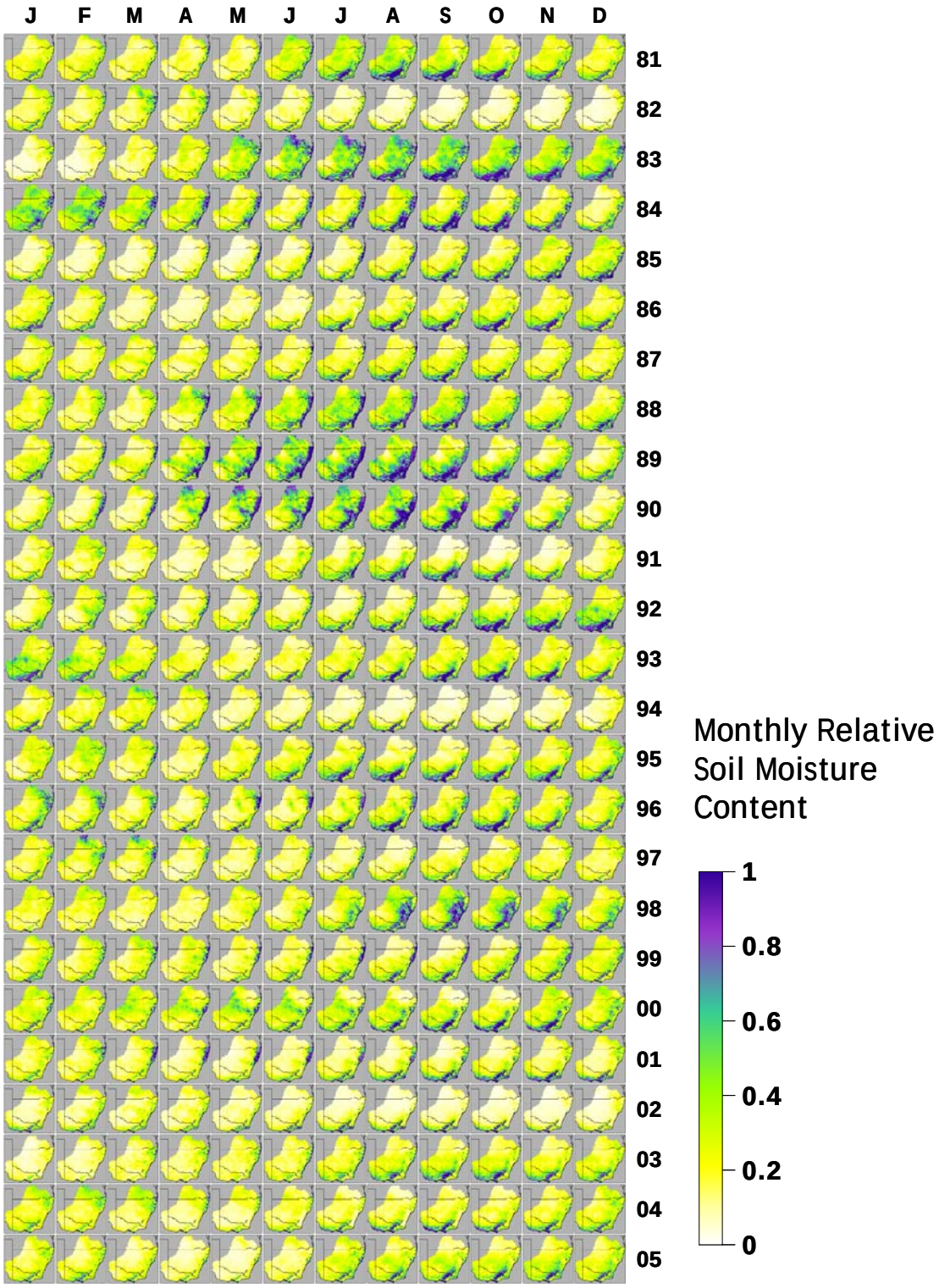


Figure 11: 25-year time series of monthly relative soil moisture content over Southeast Australia. Model version: WaterDyn10Mb.

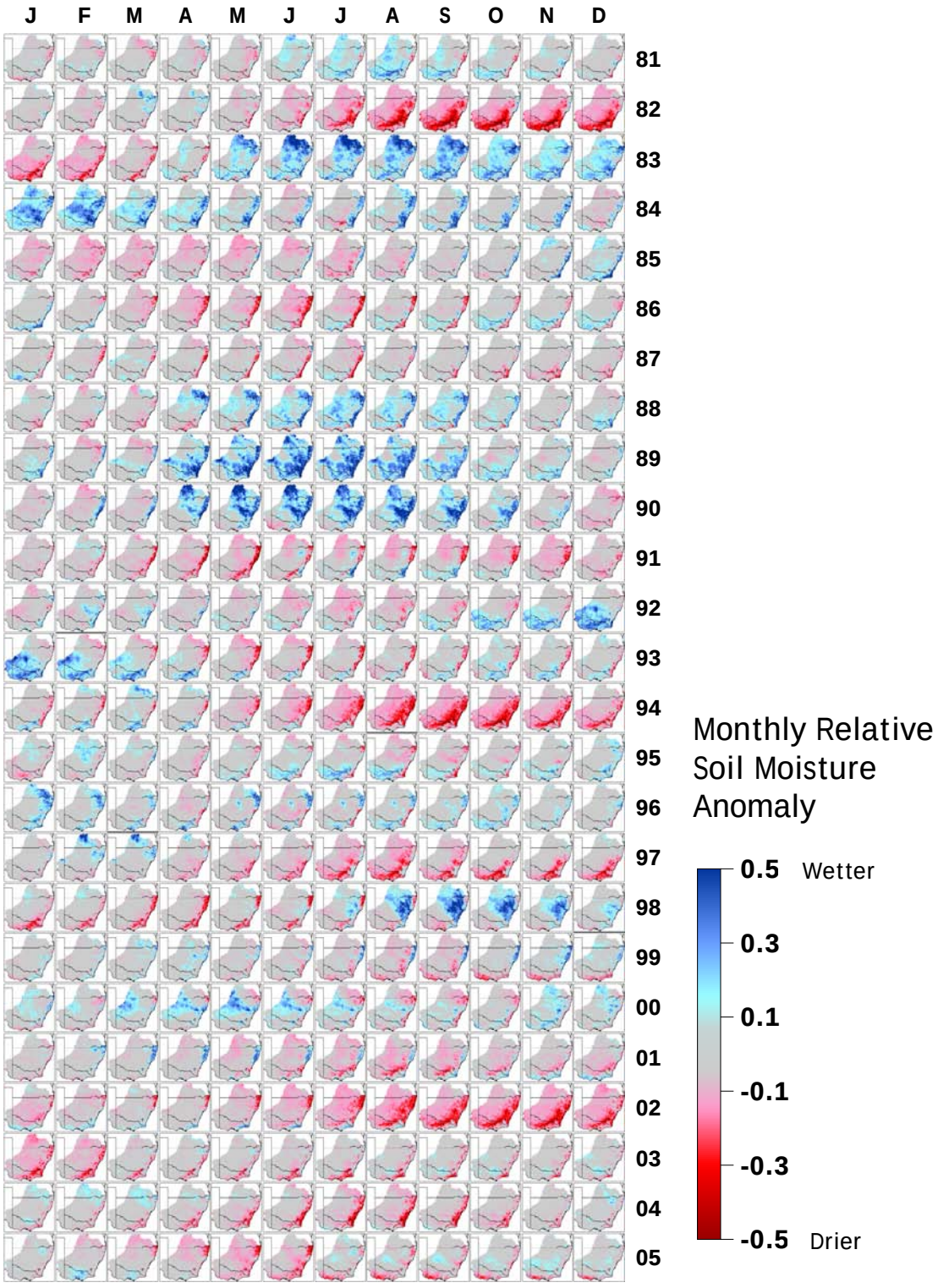


Figure 12: 25-year time series of monthly relative soil moisture anomaly over Southeast Australia. Model version: WaterDyn10Mb.

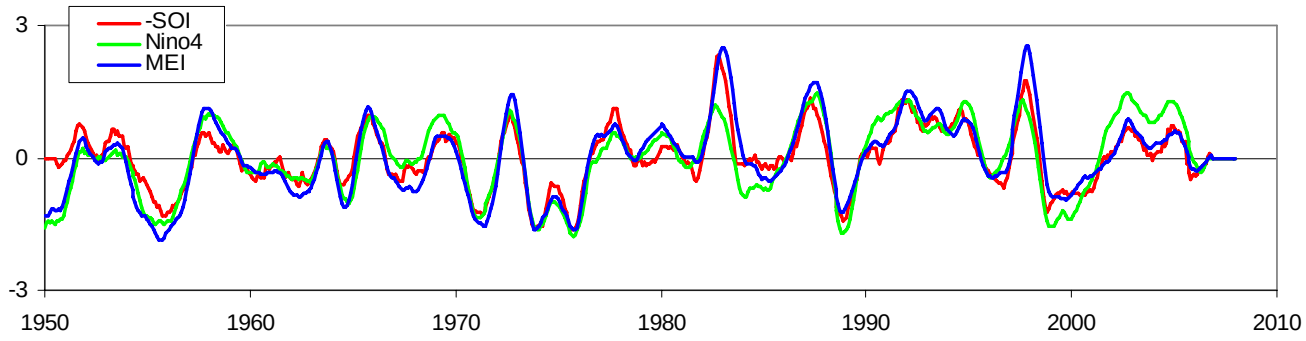


Figure 13: Time series for 1950 to 2006 of three indicators of the ENSO (El Nino-Southern Oscillation) climate phenomenon in the equatorial Pacific Ocean. SOI (Southern Oscillation Index) is the difference between standardised pressure series at Tahiti and Darwin, where standardisation implies transformation to a series with mean zero and variance one. Nino4 is the standardised sea surface temperature (SST) in a rectangular region in the Central Tropical Pacific (5N to 5S, 160E to 150W). MEI is the Multivariate ENSO Index (Wolter and Timlin 1993; Wolter and Timlin 1998), based on a combination of six observed variables over the tropical Pacific (sea-level pressure, zonal and meridional components of the surface wind, sea surface temperature, surface air temperature, and total cloudiness fraction of the sky). The MEI is the first unrotated Principal Component (PC) of all six observed fields combined. All series have been smoothed with an 11-month running mean. Positive values of all indices are correlated with drought in eastern Australia. Note that SOI defined here is $-0.1\ \mathrm{SOI}_{\mathrm{BoM}}$, where $\mathrm{SOI}_{\mathrm{BoM}}$ is the SOI used by the Australian Bureau of Meteorology, which is correlates positively with increased rainfall and takes values of ±10 at ±1 standard deviation. Data source: http://www.cdc.noaa.gov/ClimateIndices/List/.

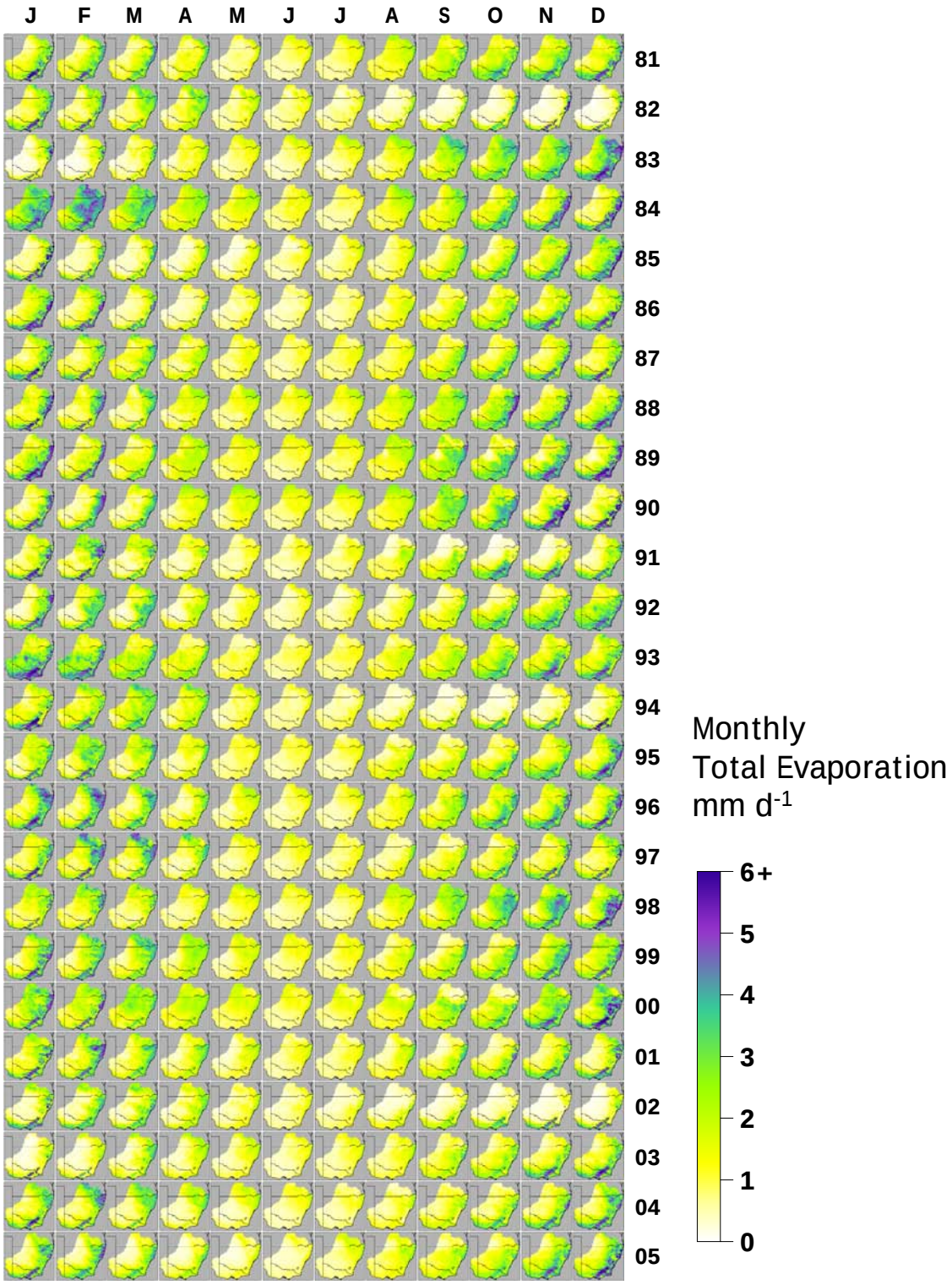


Figure 14: 25-year time series of monthly total evaporation (plant+soil) over Southeast Australia. Model version: WaterDyn10Mb.

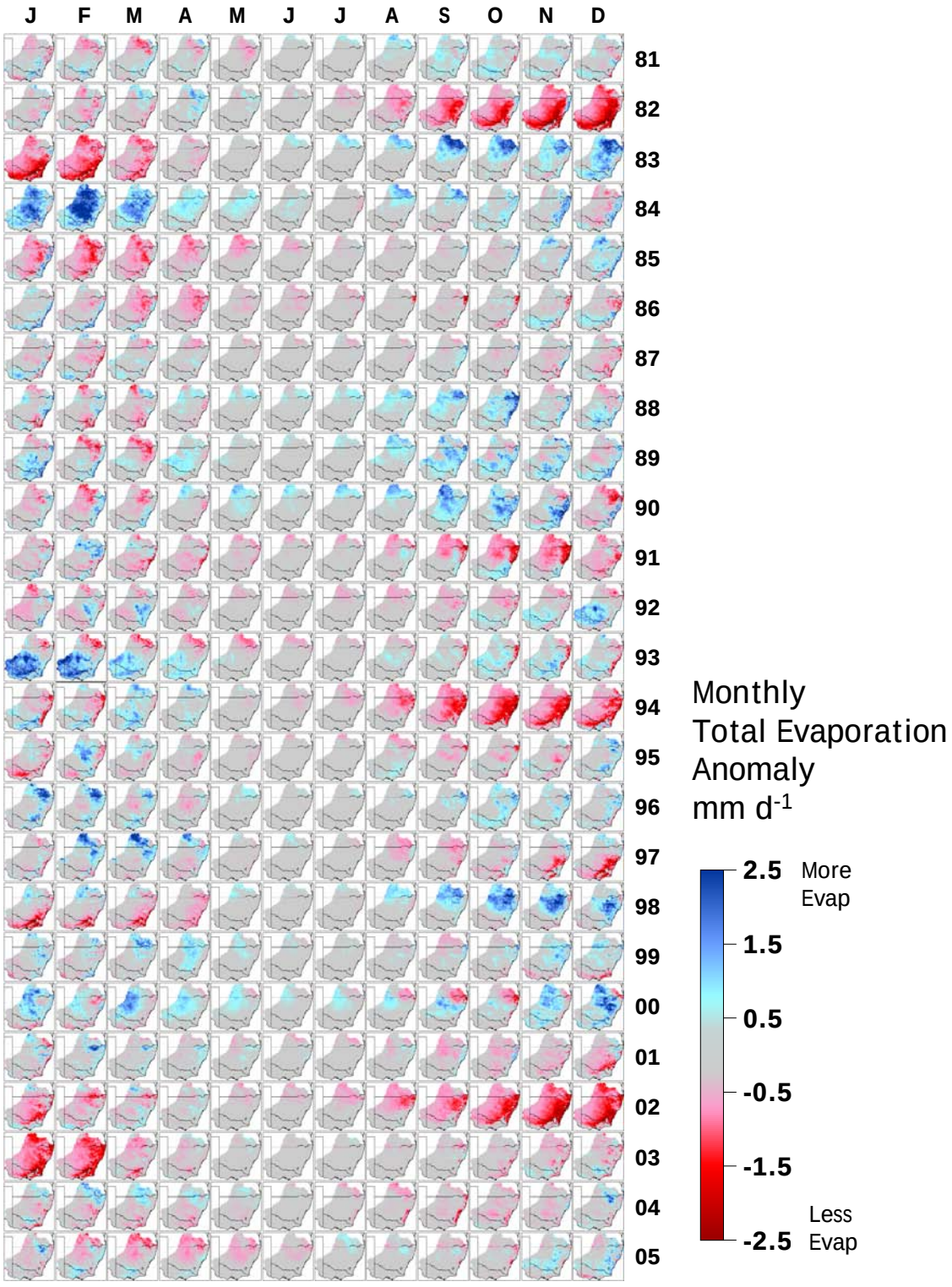


Figure 15: 25-year time series of anomaly in monthly total evaporation (plant+soil) over Southeast Australia. Model version: WaterDyn10Mb.

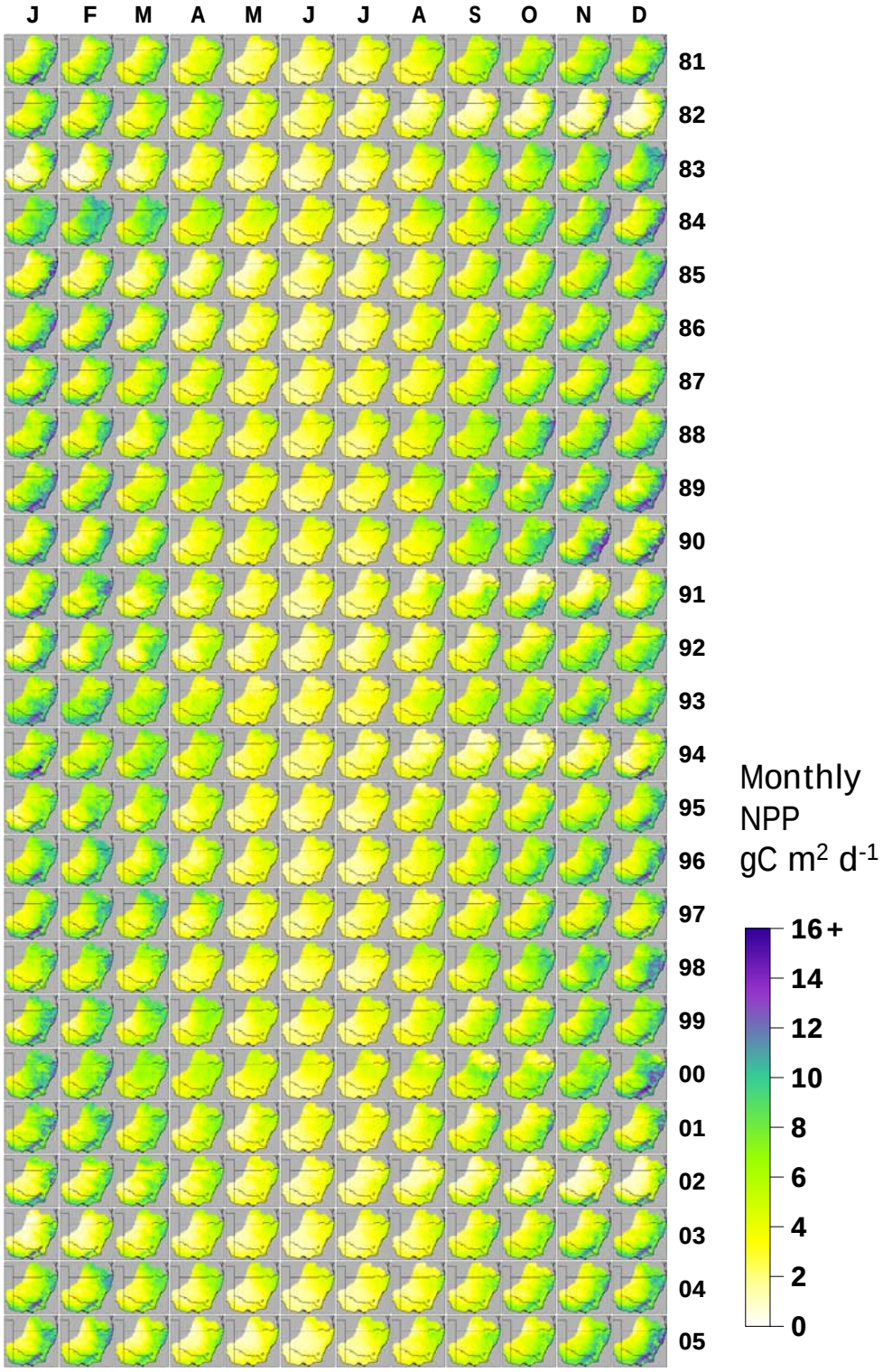


Figure 16: 25-year time series of monthly net primary production (NPP, in gC m^{-2} d^{-1}) over Southeast Australia. Model version: WaterDyn10Mb.

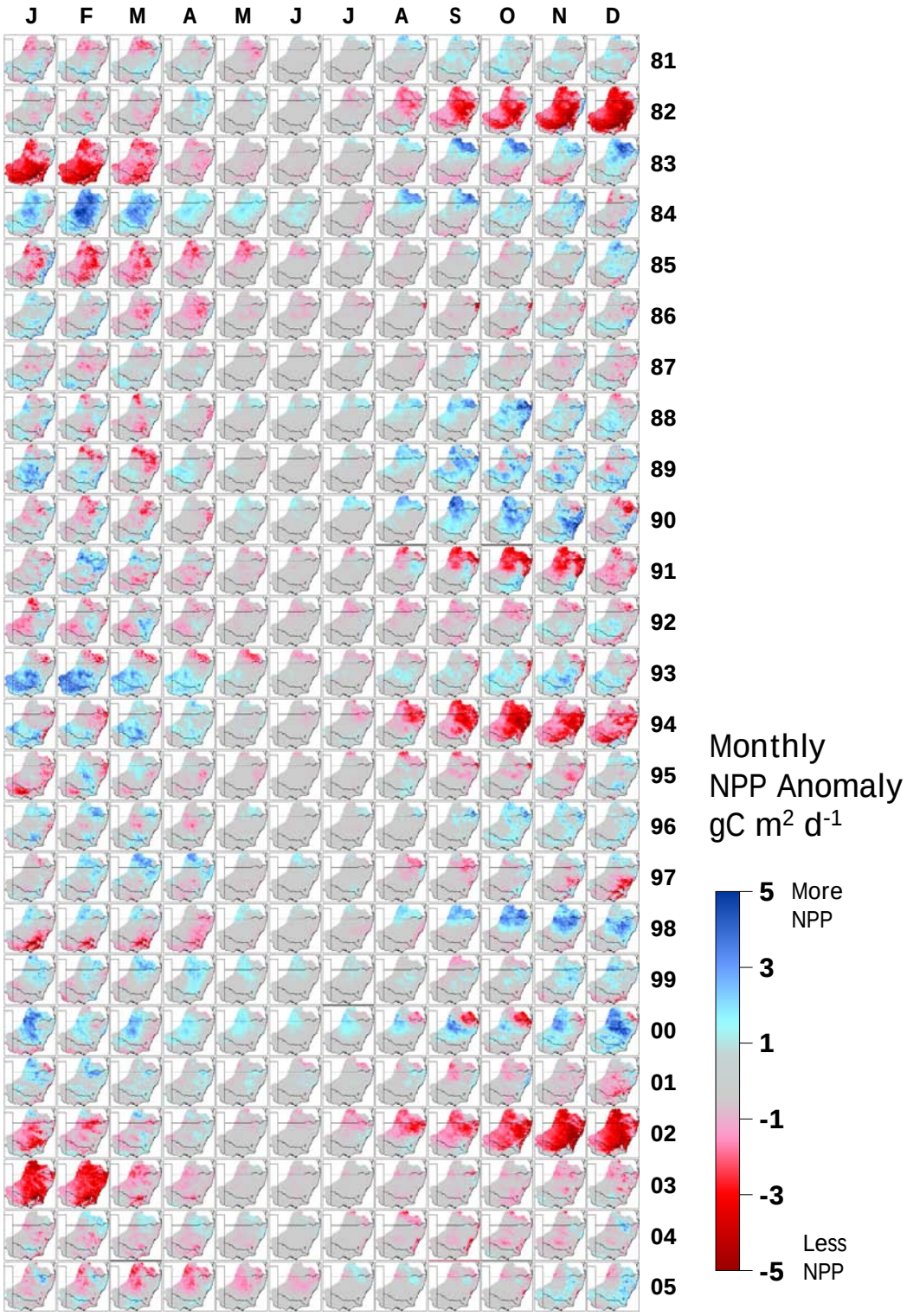


Figure 17: 25-year time series of anomaly in monthly net primary production (NPP, in gC m^{-2} d^{-1}) over Southeast Australia. Model version: WaterDyn10Mb.

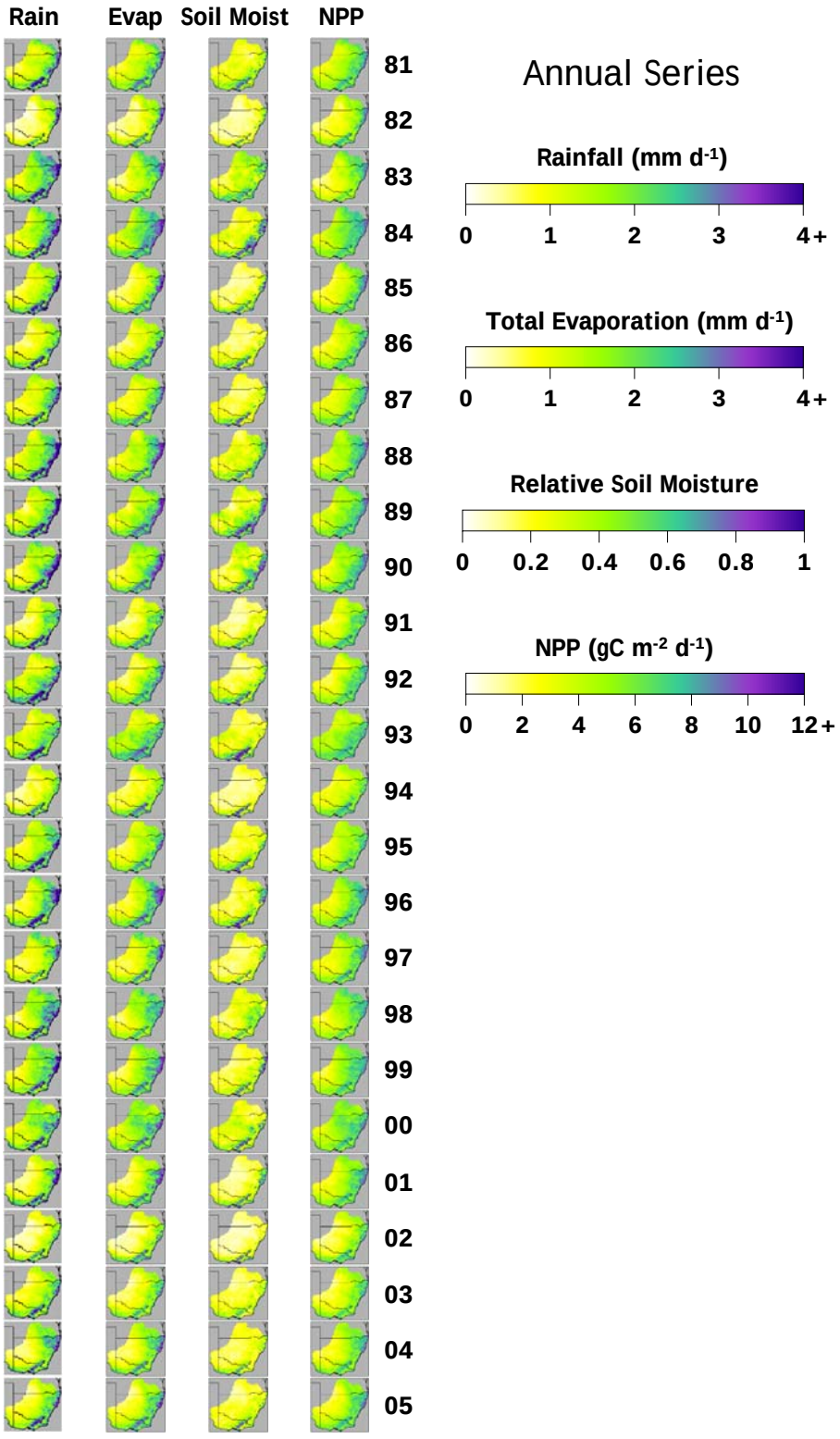
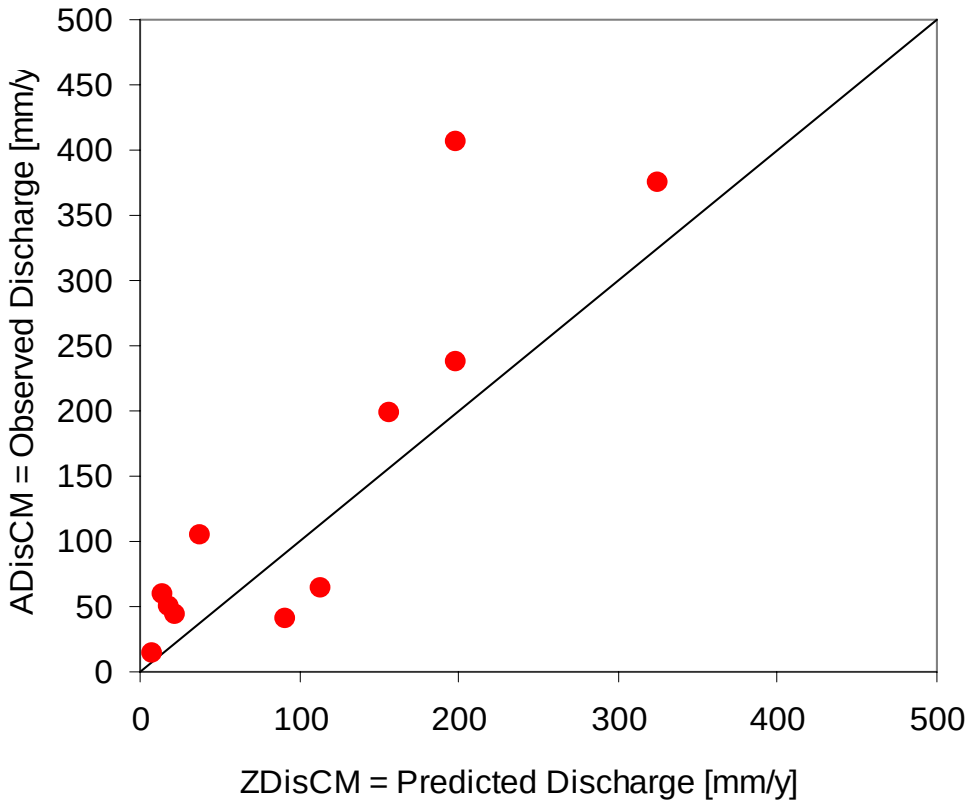


Figure 18: Comparison of annual means from 1981 to 2005 of rainfall, relative soil moisture, total evaporation and NPP. Model version: WaterDyn10Mb.



Catchment	Name	Area(km^2)	Rainfall(mm/y)	Runoff(mm/y)
410044	MuttamaCreek@Coolac	1025	633.3	51.2
410038	AdjungbillyCreek@Darbalara	411	1172.3	225.3
410047	TarcuttaCreek@OldBorambola	1660	810.1	98.1
410048	KyeambaCreek@Ladysmith	530	661.5	68.6
410057	GoobarragandraRiver@Lacmalac	673	1339.4	462.3
410061	AdelongCreek@BatlowRoad	155	1141.2	274.1
410059	GilmoreCreek@Gilmore	233	1217.6	388.1
410097	BillabongCreek@Aberfeldy	331	730.8	77.7
410033	MurrumbidgeeRiver@MittagangCrossing	1891	882	154.4
410141	MicaligoCreek@Michelago	190	745.4	50.4
222007	WullwyeRiver@Woolway	520	585.8	51.8

Figure 19: Comparison of observed versus predicted discharge at the outflow points of 11 unimpaired catchments within the Murrumbidgee basin. Mean rainfall values in the table are catchment area averages from the SILO weather archive. Mean runoff values are observed. Model version: WaterDyn09Ma.

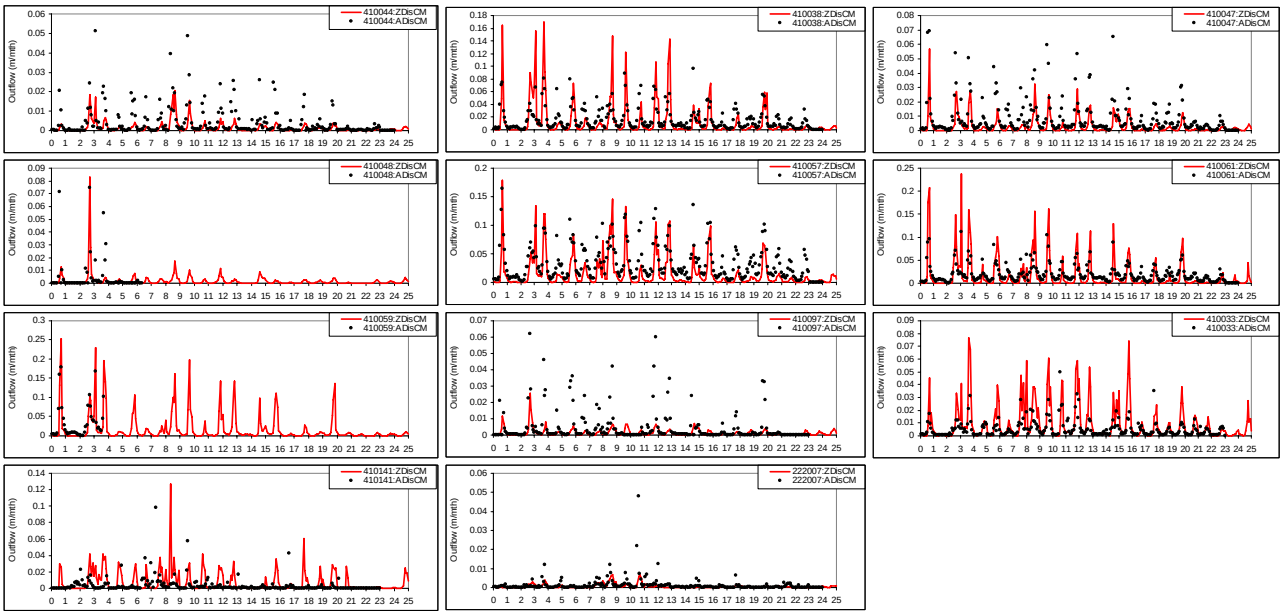


Figure 20: Predicted and observed discharge for 11 unimpaired catchments in the Murrumbidgee basin for a 25-year time series: Jan 1981 to December 2005. Model version: WaterDyn09Ma.

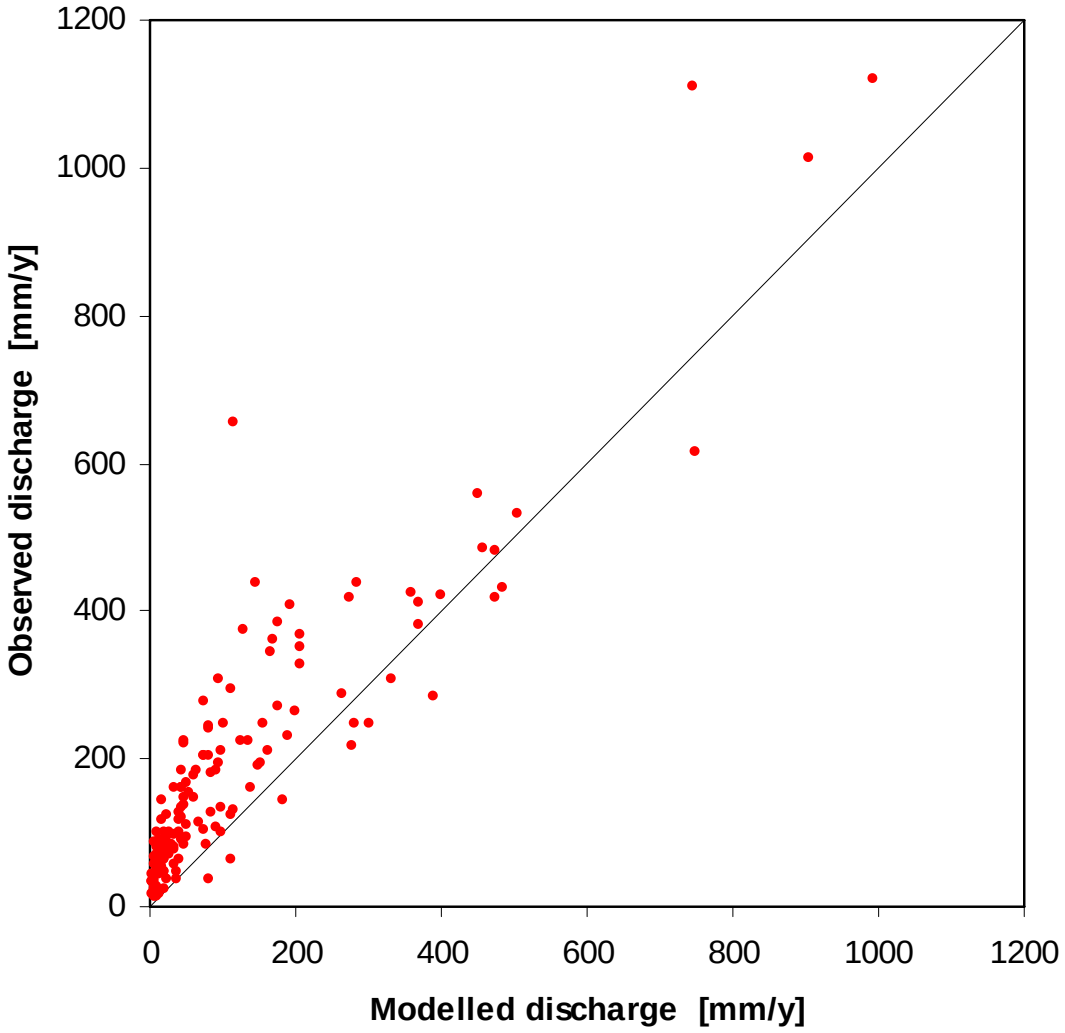


Figure 21: Comparison of observed versus predicted discharge at the outflow points of 153 unimpaired catchments in the spatial domain (see Figure 7). Model version: WaterDyn10Mb.

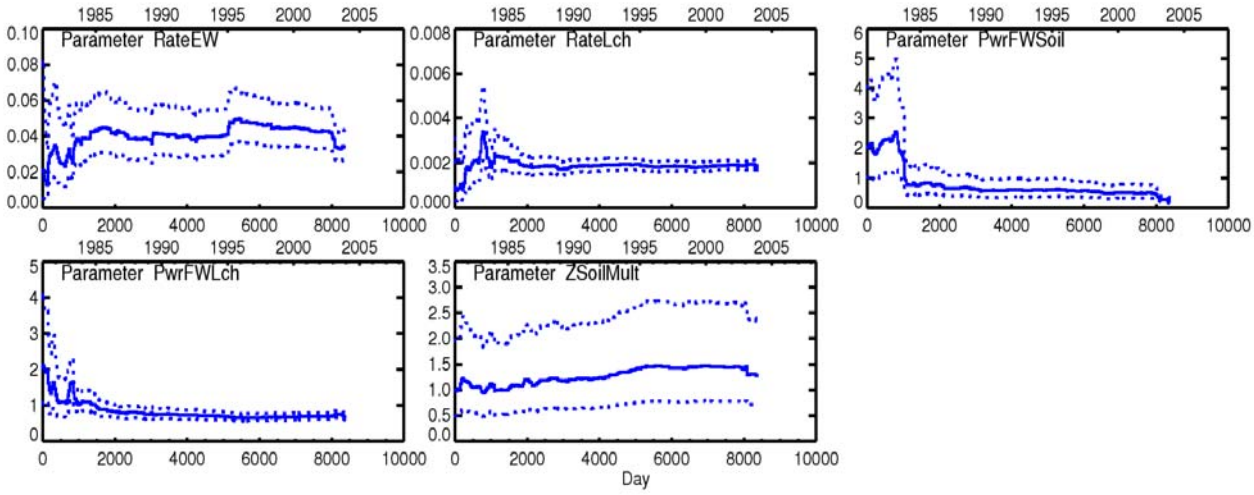


Figure 22: Sequential estimates of five parameters for the Adelong catchment, by assimilation of observed monthly discharge using the EnKF, with the WaterDyn model in assimilation mode and an ensemble of 100 members.

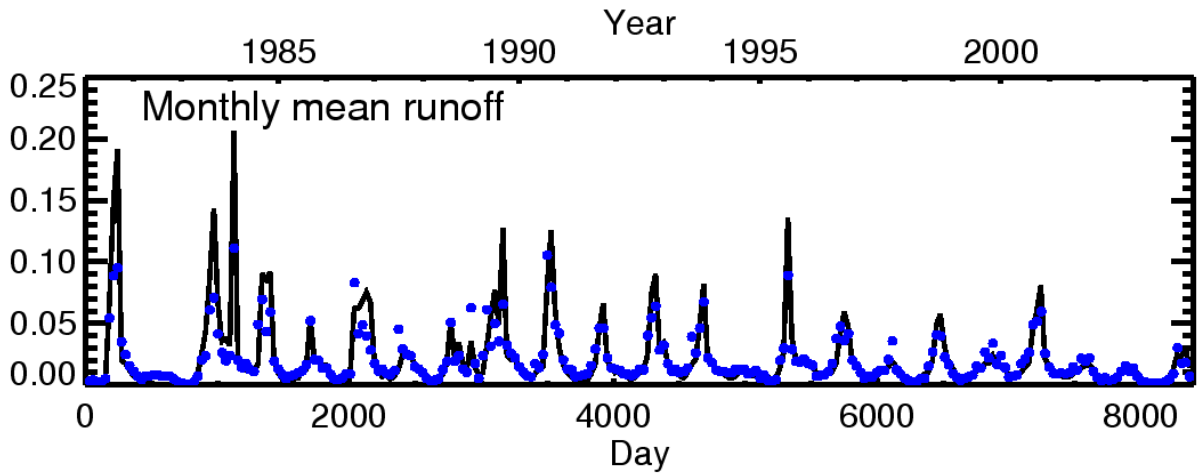


Figure 23: Comparison of observed (blue points) and predicted (black line) discharge (m/month) for the Adelong catchment, for the same 25-year run as in Figure 22.

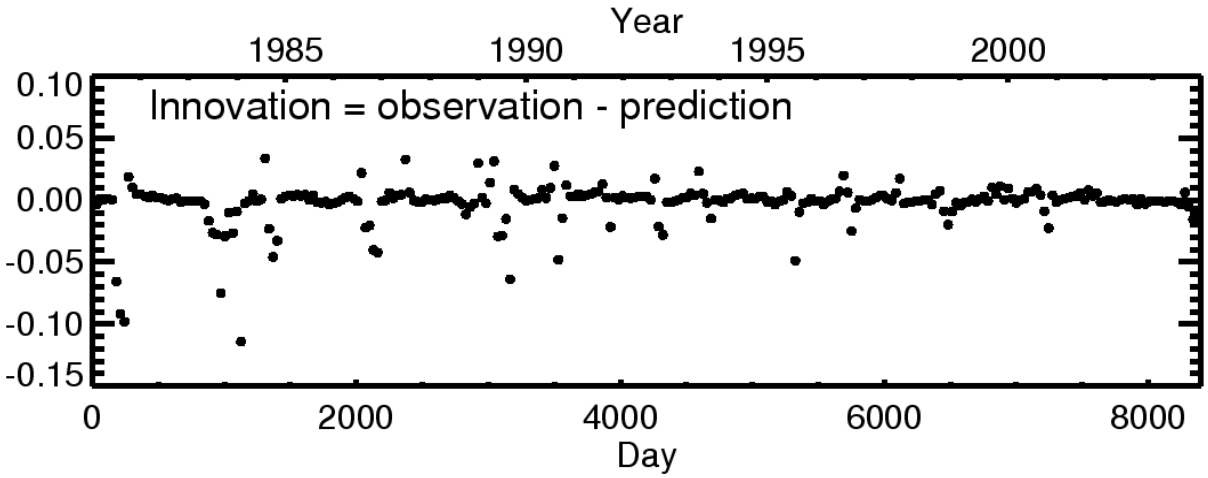


Figure 24: Innovation (= observation − prediction) in monthly mean discharge (m/month) for the Adelong catchment, for the 25-year run shown in Figures 22 and 23.

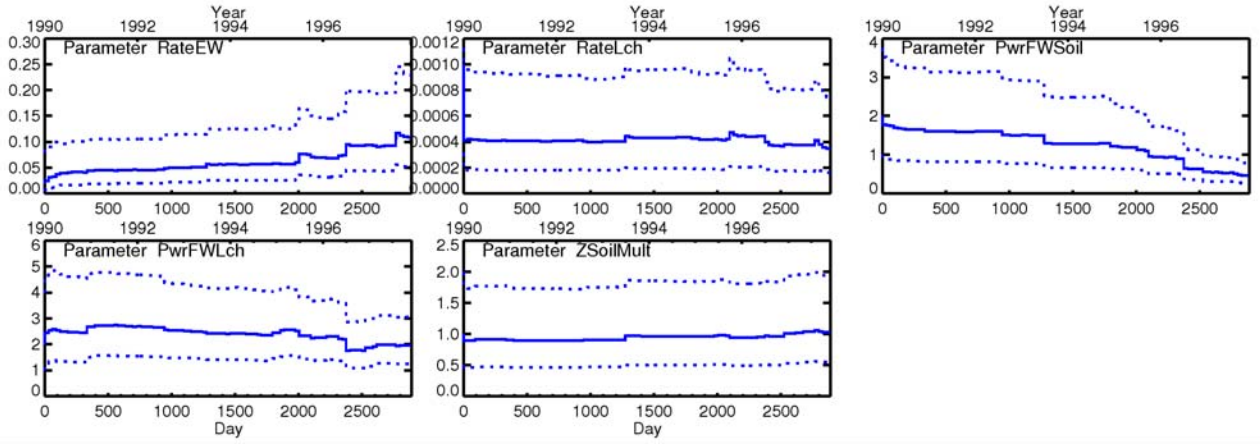


Figure 25: Sequential estimates of five parameters for a spatial region consisting of 9 catchments in the Murrumbidgee Basin, where identical parameter values are applied across all catchments. The parameters are estimated by applying the EnKF with 100 ensemble members and the WaterDyn model in assimilation mode to assimilate observed monthly discharge from each catchment. The spatial domain consists of the union of all 9 catchments, and the time domain runs from 1990 to 1997.

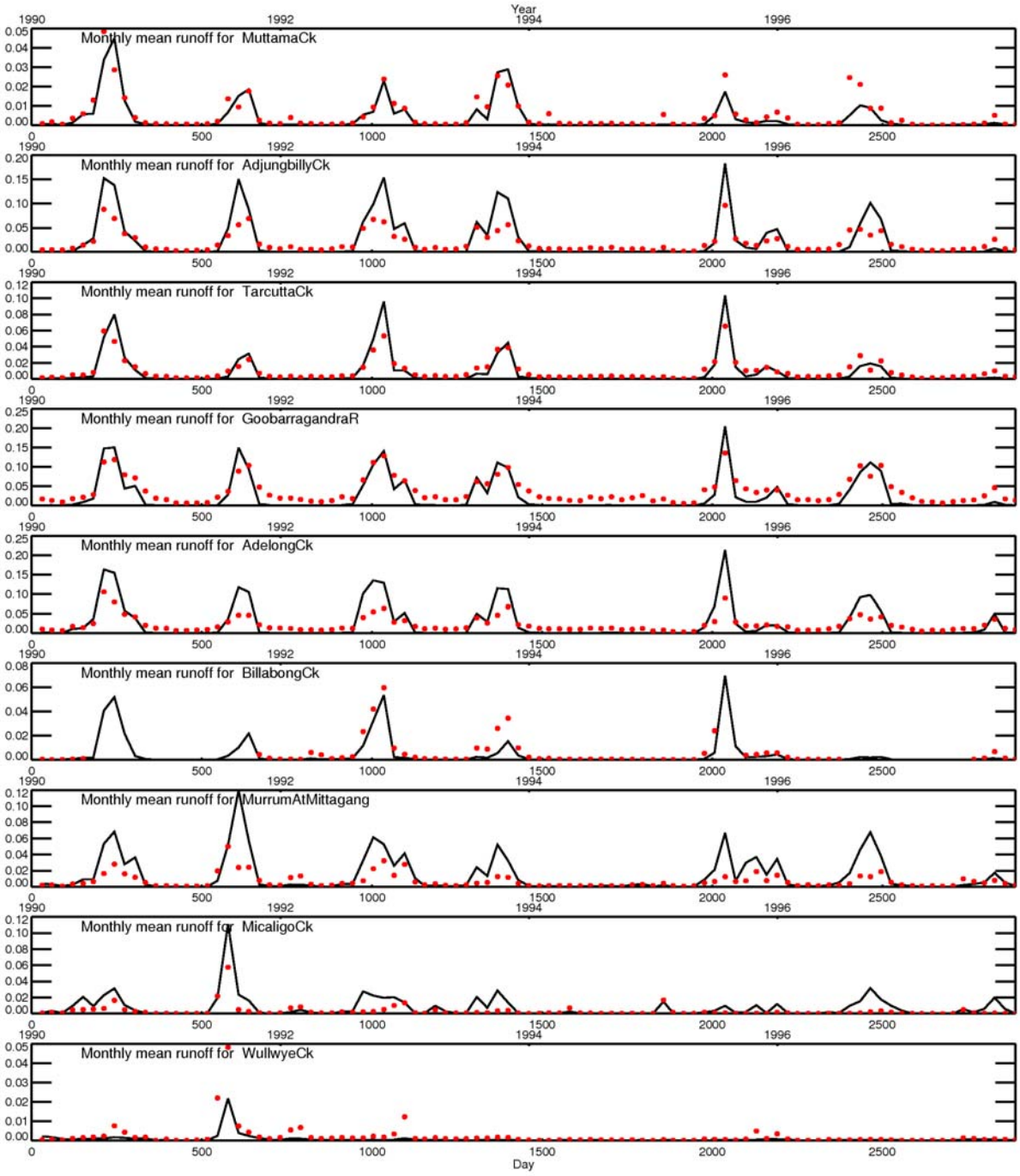


Figure 26: Comparison of observed (points) and predicted (line) discharge (m/month) for 9 catchments in the Murrumbidgee Basin, where identical parameter values are applied across all catchments. The parameters are estimated by applying the EnKF with 100 ensemble members and the WaterDyn model in assimilation mode to assimilate observed monthly discharge from each catchment. The spatial domain consists of the union of all 9 catchments, and the time domain runs from 1990 to 1997.

References

Casti JL (2000) *Five more golden rules*. John Wiley and Sons, Inc., New York, 1

Cracknell AP (1997) *The advanced very high resolution radiometer (AVHRR).* Taylor and Francis, London and Bristol, PA,

Enting IG (2002) *Inverse problems in atmospheric constituent transport*. Cambridge University Press, Cambridge, 392 pp

Evensen G (1994) Sequential Data Assimilation with A Nonlinear Quasi-Geostrophic Model Using Monte-Carlo Methods to Forecast Error Statistics. *Journal of Geophysical Research-Oceans* **99**:10143-10162

Grewal MS, Andrews AP (1993) *Kalman Filtering: Theory and Practice*. Prentice-Hall, Englewood Cliffs, NJ, 381 pp

King EA (2000) Stitching the Australian 1-km archive. Proceedings of the Land EnvSat Workshop, 10th Australasian Remote Sensing and Photogrammetry Conference, Adelaide, Australia, 2000, CSIRO, pp 41-50

King EA (2003) The Australian AVHRR data set at CSIRO/EOC: origins, processes, holdings and prospects. EOC Report 2003/04, CSIRO Earth Observation Centre, Canberra, ACT

Lavery B, Joung G, Nicholls N (1997) An extended high-quality historical rainfall dataset for Australia. *Aust. Met. Mag.* **46**:27-38

Lovell JL, Graetz RD (2001) Filtering pathfinder AVHRR land NDVI data for Australia. *Int. J. Remote Sens.* **22**:2649-2654

Lovell JL, Graetz RD, King EA (2003) Compositing AVHRR data for the Australian continent: seeking best practice. *Canadian Journal of Remote Sensing* **29**:770-782

Lu H, Raupach MR, McVicar TR, Barrett DJ (2003) Decomposition of vegetation cover into woody and herbaceous components using AVHRR NDVI time series. *Remote Sensing of Environment* **86**:1-18

McKenzie NJ, Hook J (1992) Interpretations of the Atlas of Australian Soils. Technical Report 94/1992, CSIRO Division of Soils, Canberra

McKenzie NJ, Jacquier DW, Ashton LJ, Cresswell HP (2000) Estimation of soil properties using the Atlas of Australian Soils. CSIRO Land and Water Technical Report 11/00, CSIRO Land and Water, Canberra, Australia, 24 pp

NLWRA (2001a) Australian Agriculture Assessment 2001. National Land and Water Resources Audit, Commonwealth of Australia, Canberra, Australia, 480 pp

NLWRA (2001b) Australian Water Resources Assessment 2000. National Land and Water Resources Audit, Commonwealth of Australia, Canberra, Australia, 160 pp

Priestley CHB, Taylor RJ (1972) On the assessment of surface heat flux and evaporation using large-scale parameters. *Mon. Weather Rev.* **100**:81-92

Raupach MR (2000) Equilibrium evaporation and the convective boundary layer. *Boundary-Layer Meteorol.* **96**:107-141

Raupach MR (2001) Combination theory and equilibrium evaporation. *Quart. J. Roy. Meteorol. Soc.* **127**:1149-1181

Raupach MR (2005) Dynamics and optimality in coupled terrestrial energy, water, carbon and nutrient cycles. In: Franks SW, Sivapalan M, Takeuchi K, Tachikawa Y (eds) *Prediction in Ungauged Basins: International Perspectives on the State-of-the-Art and Pathways Forward (IAHS Publication No. 301).* IAHS Press, Wallingford

Raupach MR, Barrett DJ, Briggs PR, Kirby JM (2005a) Simplicity, complexity and scale in terrestrial biosphere modelling. In: Franks SW, Sivapalan M, Takeuchi K, Tachikawa Y (eds) *Prediction in Ungauged Basins: International Perspectives on the State-of-the-Art and Pathways Forward (IAHS Publication No. 301).* IAHS Press, Wallingford

Raupach MR, Kirby JM, Barrett DJ, Briggs PR (2001a) Australian Landscape Balances of Water, Carbon, Nitrogen and Phosphorus: (1) Project Description and Results. CSIRO Land and Water Technical Report 40/01, CSIRO Land and Water, Canberra, Australia

Raupach MR, Kirby JM, Barrett DJ, Briggs PR, Lu H, Zhang L (2001b) Australian Landscape Balances of Water, Carbon, Nitrogen and Phosphorus: (2) Model Formulation and Testing. CSIRO Land and Water Technical Report 41/01, CSIRO Land and Water, Canberra, Australia

Raupach MR, Rayner PJ, Barrett DJ, DeFries RS, Heimann M, Ojima DS, Quegan S, Schmullius CC (2005b) Model-data synthesis in terrestrial carbon observation: methods, data requirements and data uncertainty specifications. *Global Change Biol.* **11**:10.1111/j.1365-2486.2005.00917.x

Sellers PJ (1985) Canopy reflectance, photosynthesis and transpiration. *Int. J. Remote Sens.* **6**:1335-1372

Sellers PJ, Berry JA, Collatz GJ, Field CB, Hall FG (1992) Canopy reflectance, photosynthesis, and transpiration 3: A reanalysis using improved leaf models and a new canopy integration scheme. *Remote Sens. Environ.* **42**:187-216

Torok SJ, Nicholls N (1996) A historical annual temperature dataset for Australia. *Aust. Met. Mag.* **45**:251-260

Tucker CJ (1979) Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment* **8**:127-150

Wolter K, Timlin MS (1993) Monitoring ENSO in COADS with a seasonally adjusted principal component index. Proc. of the 17th Climate Diagnostics Workshop, Norman, OK, NOAA/N MC/CAC, NSSL, Oklahoma Clim. Survey, CIMMS and the School of Meteor., Univ. of Oklahoma, 52-57. 1993, University of Oklahoma

Wolter K, Timlin MS (1998) Measuring the strength of ENSO - how does 1997/98 rank? *Weather* **53**:315-324