Australian Government
Bureau of Meteorology

Evaluation of the Australian Landscape Water Balance model: AWRA-L v6

A comparison of AWRA-L v6 against Observed Hydrological Data and Peer Models

Citation: Frost, A. J. and Wright, D. P., (2018). Evaluation of the Australian Landscape Water Balance model: AWRA-L v6. Bureau of Meteorology Technical Report.

Version number/type	Date of issue
3	28/11/2018

© Commonwealth of Australia 2018

This work is copyright. Apart from any use as permitted under the Copyright Act 1968, no part may be reproduced without prior written permission from the Bureau of Meteorology. Requests and inquiries concerning reproduction and rights should be addressed to the Production Manager, Communication Section, Bureau of Meteorology, GPO Box 1289, Melbourne 3001. Information regarding requests for reproduction of material from the Bureau website can be found at www.bom.gov.au/other/copyright.shtml

Published by the Bureau of Meteorology

Contact details

Dr. Andrew Frost

Senior Hydrologist, Water Resources Modelling Unit

Bureau of Meteorology
PO Box 413, Darlinghurst NSW 1300

Tel: +61 2 9296 1517
Email: Andrew.Frost@bom.gov.au

Dr. Chantal Donnelly

Unit Head, Water Resources Modelling Unit

Bureau of Meteorology
GPO Box 413 Brisbane QLD 4001

Tel: +61 7 3239 8767
Email: Chantal.Donnelly@bom.gov.au

AWRA Modelling Team:

Water Resources Modelling Unit, Water Program

Email: awrams@bom.gov.au

Summary

This technical report details the scientific evaluation of the Bureau of Meteorology (BoM: hereafter called the Bureau) operational Australian Water Resources Assessment Landscape (AWRA-L version 6) modelling system. The evaluation used a range of the best measurements/estimates of hydrological variables available nationally, including streamflow, soil moisture, actual evapotranspiration (ET) and groundwater recharge. In addition, the performance of the operational AWRA-L version 6 model (hereafter called AWRA-L v6.0) is compared to the previous AWRA-L v5.0, and two other national, gridded, land-surface models i.e. CABLE-SLI and WaterDyn. Runoff simulated by AWRA-L v6 is also compared with simulated streamflow from individual conceptual rainfall runoff models using at-site calibration and nearest neighbour regionalisation for predictions in ungauged basins.

AWRA-L and the peer models are assessed and compared according to various performance statistics for each set of evaluation data. Select key indicators of AWRA-L model performance are provided. These *benchmark* statistics provide a baseline against which future model improvements can be compared, using the same comparison data. Aspirational targets for overall performance are also provided. The AWRA model is optimised and selected for performance across the water balance, nationally.

The results show that AWRA-L v6.0 performs relatively well according to streamflow nationally (295 unimpaired catchments used in calibration and 291 separate catchments used in validation) in comparison to WaterDyn and CABLE, reflecting that AWRA-L is calibrated to streamflow. AWRA-L v6 also performs well according to probe-based point measurements of root zone (0-90cm) soil moisture from 51 sites in South Eastern Australia. AWRA-L v6.0 performs relatively poorly in comparison to WaterDyn and CABLE for ET (25 flux tower measurements nationally), although median monthly correlation is now equal to that in CABLE, improving over previous performance of AWRA-L v5.0. Comparison of AWRA-L model deep drainage output (against a long term average and annual time-series recharge dataset) showed that observed drainage biases are driven predominantly by the saturated hydraulic conductivity rather than rainfall variability, noting high uncertainty in these recharge estimates. AWRA-L v6.0 improves over AWRA-L v5.0 for the key variables tested; noting performance for 0-5/0-8 cm soil moisture correlation decreases in AWRA-L v6.0.

The improved performance of the AWRA-L v6.0 model for streamflow is a result of improvements to the nationwide calibration target and the conceptual hydrological structure. When comparing to peer models, CABLE is equivalent to AWRA-L v6.0 in terms of soil moisture, and marginally better regarding ET as expected from its purpose as a model for land/atmosphere exchange, along with calibration to flux tower and derived catchment ET. WaterDyn performs well for ET, but both these peer models perform worse for streamflow and root zone soil moisture. Comparison to locally calibrated, nearest neighbour regionalised rainfall-runoff models show that while AWRA-L calibrated nationally does not perform as well in calibration (as there is one set of parameters

applying nationally), performance for ungauged basins approaches that of the locally calibrated models; giving confidence in the use of AWRA-L nationally for runoff prediction; along with other components of the water balance.

Each of the models have differing strengths and weaknesses. Overall, given runoff/streamflow is the dominant hydrological variable used in surface water resource assessment, and that AWRA-L simultaneously performs well for root zone soil moisture (a key agricultural variable), AWRA-L is considered the most fit-for-purpose national hydrological model estimator for water resource and agricultural applications.

In addition to the scientific evaluation against observed hydrological datasets and other peer models, annual national maps and monthly catchment time-series of the outputs from AWRA-L are presented for scientific understanding of hydrological processes. Overall, the spatial plots and the time-series give confidence in the use of AWRA-L for water resources assessment; as they broadly follow the expected catchment responses and spatial and temporal trends expected across Australia.

Table of Contents

List of Tables

List of Figures

List of Acronyms

ACCESS: Australian Community Climate and Earth System Simulator climate model

AMSR-E: Advanced Microwave Scanning Radiometer for the Earth Observing System

ASCAT: Advanced Scatterometer aboard the MetOp-A satellite

ASRIS: Australian Soil Resource Information System

AWAP: Australian Water Availability Project

AWRA-L: Australian Water Resources Assessment Landscape Model

AWRA-R: Australian Water Resources Assessment River Model

AWRAMS: Australian Water Resource Assessment modelling system

BoM: Bureau of Meteorology

CABLE: Community Atmosphere Biosphere Land Exchange model

CMRSET: CSIRO MODIS reflectance-based Scaling ET

CSIRO: Commonwealth Scientific and Industrial Research Organisation

DINGO: Dynamic INtegrated Gap filling and partitioning for OzFlux

ET: Evapotranspiration

LAI: Leaf Area Index

NWA: National Water Account

MODIS: Moderate Resolution Imaging Spectroradiometer

RWI: Regional Water Information

SASMAS: Scaling and Assimilation of Soil Moisture and Streamflow

SLI: Soil-Litter-Iso

WIA: Water in Australia

WIRADA: Water Information Research and Development Alliance

1 Introduction

Prolonged extreme drought and resulting water shortages within Australia during the 'Millennium drought', over the period 1997 to 2009, triggered the implementation of the federally mandated Water Act (2007) towards better monitoring of water availability and water use nationwide. As a result, the Australian Bureau of Meteorology (the Bureau) was given responsibilities including collating water data from jurisdictional agencies and analysing and reporting on water status, in addition to its existing weather and flood forecasting responsibilities.

The Australian Water Resources Assessment (AWRA) Modelling System underpins the Bureau's water information services for national water resource assessment reporting, water use accounting and situation monitoring. The modelling system has been developed by the Bureau and CSIRO over the last decade and is run operationally within the Bureau to provide both situational awareness and national retrospective water resource assessment.

The AWRA-L (landscape) model runs on a daily timestep and 0.05° grid (approximately 5 km) simulating the landscape water balance for Australia from 1911 to yesterday. Key outputs from the AWRA-L model include surface runoff, soil moisture, evapotranspiration and deep drainage. Outputs from the model are available through the website interface (http://www.bom.gov.au/water/landscape); or by request as a registered user through other means.

AWRA-L is optimised to the whole water balance using a national streamflow dataset along with satellite derived soil moisture and evapotranspiration estimates. The model is validated against a wide range of observational datasets including point scale soil moisture probe data, flux tower estimates and groundwater recharge estimates. The modelling system has recently been released as a community modelling system (https://github.com/awracms/awra_cms), enabling application and development by the wider research community.

Since the operational AWRA-L version 5 modelled outputs have been made publicly available in November 2015, the modelled fluxes have been used internally and externally for various climatological, flood, water and agriculture applications across Australia. This report documents the testing of AWRA-L v6 (released operationally in late 2018). The model description of v6 is found in Frost et al (2018). This technical report updates previous evaluation undertaken for AWRA-L v5 within Frost et al (2016a) to include evaluation of AWRA-L v6.

This report evaluates and compares the hydrologic performance of the AWRA-L v5 and v6 models, a national water balance model (WaterDyn) and a global biogeochemical land surface scheme (CABLE), applied regionally. These models were compared against catchment streamflow, point estimates of flux tower derived evapotranspiration across Australia, and point estimates of 0-90cm profile soil moisture over the

Murrumbidgee and Upper Hunter Catchments. Satellite derived estimates of evapotranspiration (CMRSET, SLST) and soil moisture (AMSR-E, ASCAT) are also compared to the point based observations.

The three models were further compared to a collated national long term average recharge dataset and a set of annual recharge time-series data within South Australia. Finally, locally calibrated nearest neighbour regionalised conceptual rainfall runoff models are also compared to AWRA-L, WaterDyn and CABLE to see how they perform relatively for streamflow (given a range of outputs are provided by the national models).

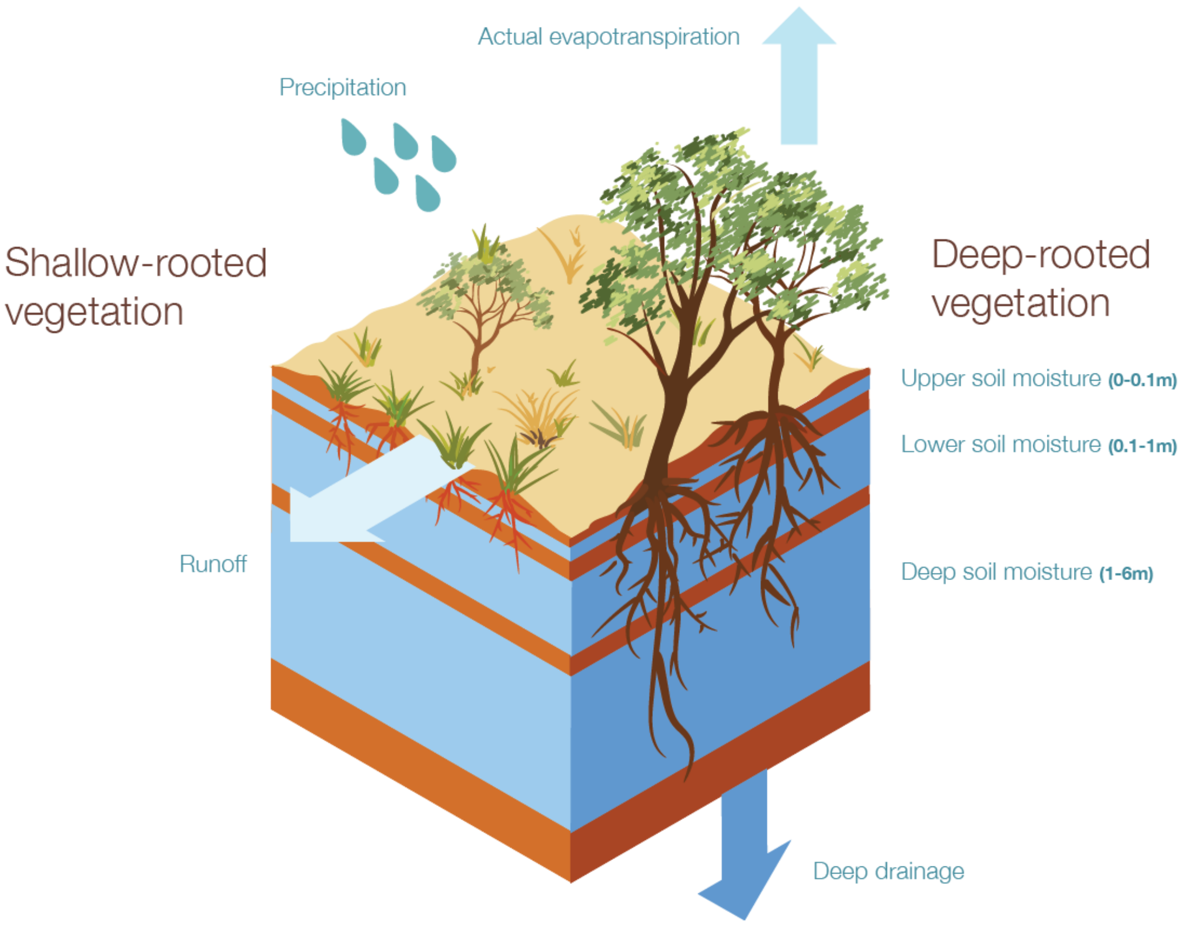

Figure 1. Conceptual AWRA-L grid cell with key water stores and fluxes shown

This report is structured as follows:

- Chapter 2: Forcing and evaluation data
- Chapter 3: Model descriptions
- Chapter 4: Evaluation approach
- Chapter 5: Evaluation according to data
- Chapter 6: Evaluation for reporting purposes
- Chapter 7: Conclusions

2 Data

Various organisations provide good monitoring of rainfall and streamflow across the nation; albeit coverage is sparse in some areas (e.g. arid interior of Western Australia). Monitoring of other hydrological fluxes and stores (e.g. soil moisture, ET, deep drainage) is less prevalent. However, since 2000 many universities and research groups have established an excellent ground-based network for physical measurement of hydrological fluxes in select catchments and locations around Australia. The Bureau has utilised these hydrological catchment-based and point-based datasets for scientific evaluation of AWRA-L model performance in this report.

2.1 Climate forcing data

All models were forced using the daily gridded Australian Water Availability Project (AWAP) climate data set that consists of air temperature (daily minimum and maximum) and daily precipitation from 1st January 1911 to yesterday (Jones et al., 2009). The climate data is interpolated from station records and provided on a 0.05° (approximately 5 km) grid across Australia. Additionally, daily solar exposure (downward shortwave radiation) is produced from geostationary satellites (Grant et al., 2008) and aggregated to the same 0.05° AWAP grid. The solar radiation record is from 1990 to yesterday, with the Himawari-8 satellite used since 23rd March 2016. Prior to that date the GMS-4, GMS-5, GOES-9 and MTSAT-1R satellites were used. All model simulations cover at least the period of 1950 until 2013.

2.2 Evaluation data

2.2.1 Streamflow

A set of 782 unimpaired catchments with gauged flow records of reasonable length across Australia were collated by Zhang et al. (2013), according to the following criteria: (a) catchment area is greater than 50 km^2, (b) the stream is unregulated (no dams or reservoirs), (c) no major impacts of irrigation or land use change, (d) observed record has at least 10 years of data between 1975 and 2011. The catchments (delineated using the BoM's national catchment Geofabric product: www.bom.gov.au/water/geofabric) were collated for use in calibration and evaluation of AWRA-L. The spatial distribution of catchments reserved for calibration and validation of AWRA-L is shown in Figure 2; with regional divisions showing areas of similar climate. Data from 295 catchments covering the period 1/1/1981-30/12/2011 were used in calibration of AWRA-L while 291 catchments not used in calibration are used for validation. The remaining catchments were excluded (a) if greater than 5000 km^2 as there are current no streamflow routing processes in AWRA-L and (b) to exclude nested catchments to ensure independence of records (see Zhang et al., 2013 section 5.3).

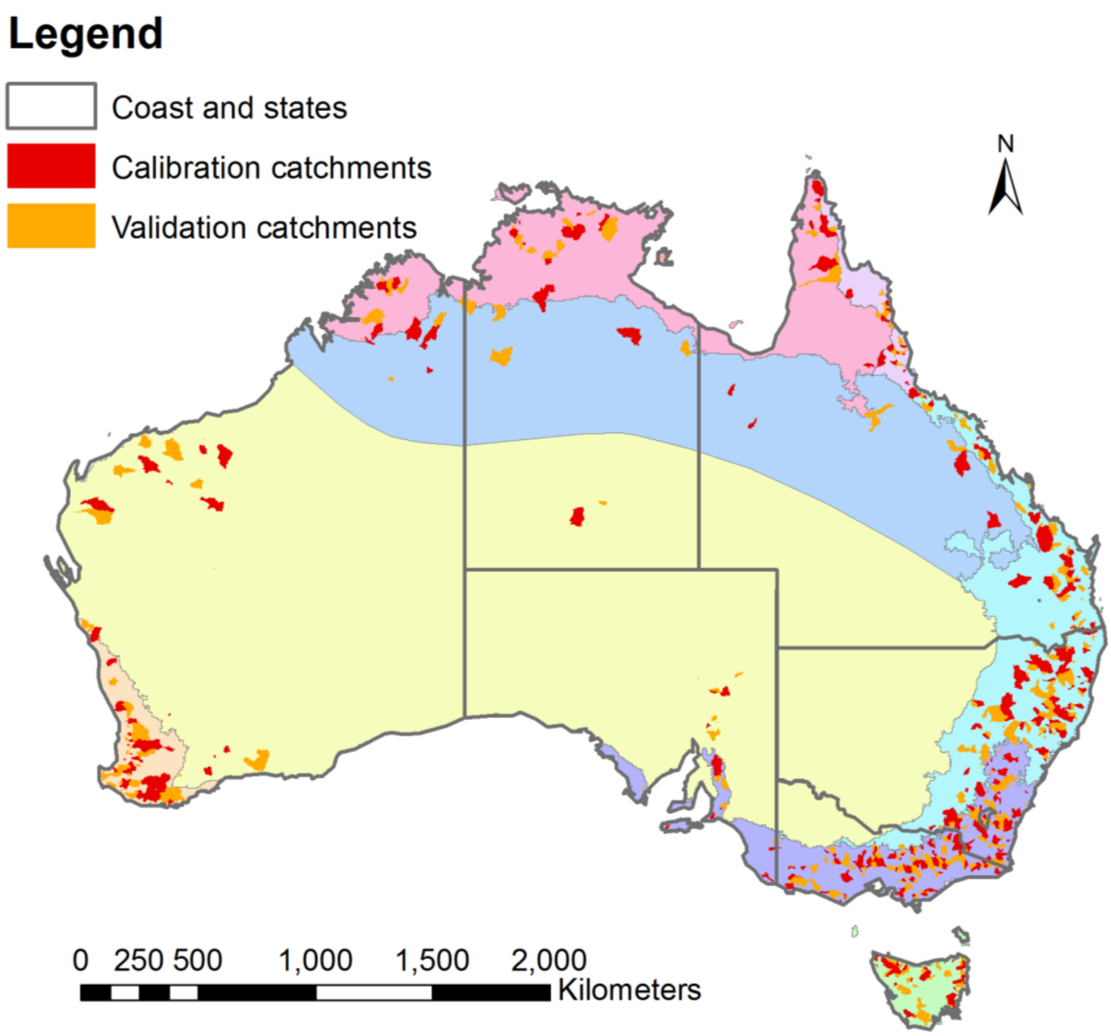

Figure 2. Location of unimpaired catchments used for model evaluation

2.2.2 Soil moisture

The following soil moisture data sets have been used for evaluation of the modelled outputs.

OzNet network: Time-series of volumetric soil moisture at various depths within the soil profile (0-5cm/8cm, 0-30cm, 30-60cm, 60-90cm) for 38 sites across the Murrumbidgee catchment, NSW (see Figure 3(a)) were used in evaluation (and not calibration) of the models. These time series were derived from reflectometer measurements from the OzNet network, setup and maintained by the University of Melbourne and Monash University (Smith et al., 2012). The reflectometers were calibrated according to independent measurements (Rüdiger et al., 2010), and had a median of 67% of monthly data available over the 2001 – 2013 period considered. For soil moisture evaluations, model soil layers were weighted according to the fraction of overlap they have with the observations they are being compared with (0-90cm for profile).

Scaling and Assimilation of Soil Moisture and Streamflow (SASMAS) Network: Time series of water content reflectometer measurements of soil moisture at various depths within the profile (0-5cm, 0-30cm, 30-60cm, 60-90cm) within the Upper Hunter River, NSW (Rüdiger et al., 2007) - see Figure 3(b) - were used in evaluation (and not calibration) of the models. These time series were collated as part of the SASMAS project monitoring sites (managed by the University of Newcastle). There were 13 sites with profile (0-90cm) data available, with a median 75% of monthly data available over the period 2003-2011.

ASCAT and AMSR-E satellite based gridded estimates: of soil moisture (along with modelled estimates) were compared to the point probe based estimates where available (surface soil moisture) to determine their value for evaluation, AWRA-L calibration and as a potential rival modelled product purposes or for eventual assimilation into AWRA-L. ASCAT is a Technische Universitat Wien (TUW) product (Bartalis et al., 2007), active Advanced Scatterometer aboard the MetOp-A satellite covering 1/07/2007-31/12/2011. The Vrije Universiteit Amsterdam (VUA)-NASA AMSR-E product (Owe et al., 2008) is derived from passive Advanced Microwave Scanning Radiometer for the Earth Observing System aboard the Aqua polar orbiting satellite, and covers 2002-2011. The methods used to derive satellite data here are further discussed in Renzullo et al (2014). It is noted that catchment averages of AMSR-E soil moisture has been used for calibration of the AWRA-L model covering the same time period – see section 3.1.

(a)

(b)

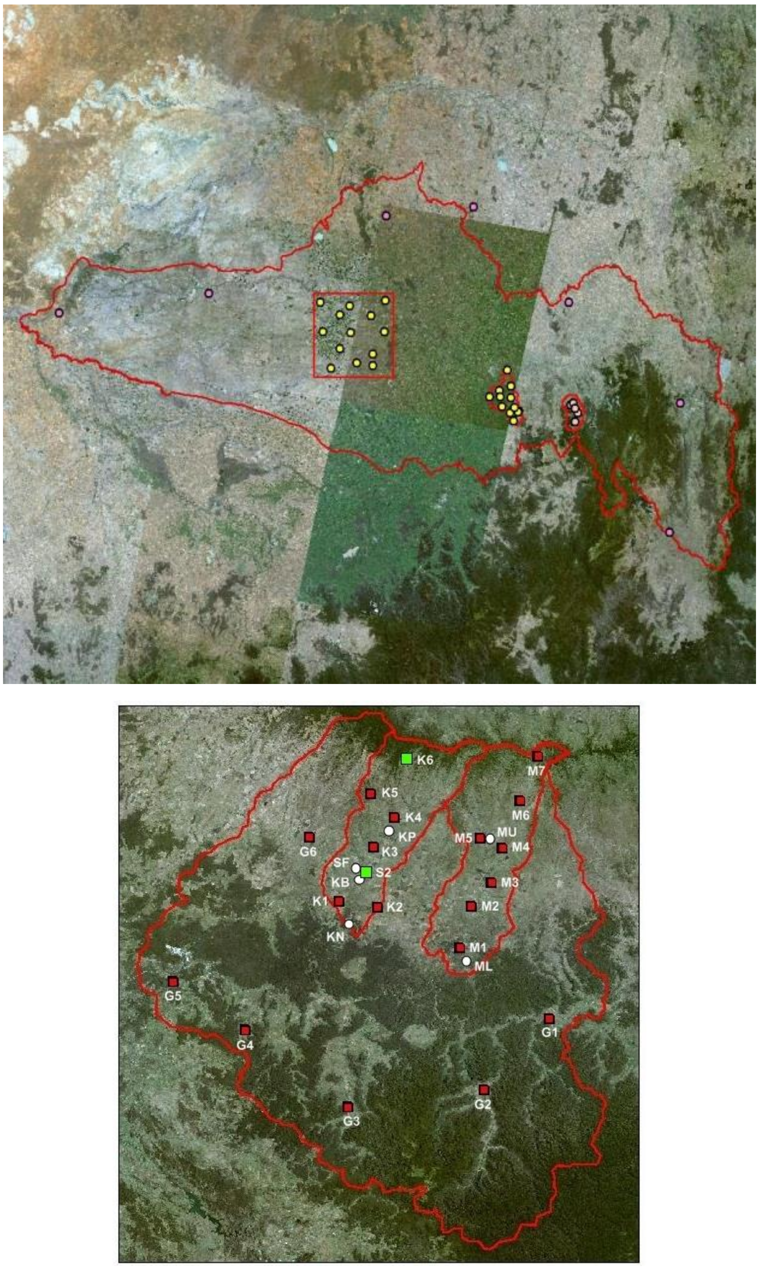

Figure 3. (a) OzNet Murrumbidgee soil moisture (from www.oznet.org.au/murrumbidgeesm.html), (b) SASMAS Goulburn soil moisture (from www.eng.newcastle.edu.au/sasmas/SASMAS/sasdata.html)

2.2.3 Actual Evapotranspiration

The following data sets were used for evaluation of the modelled outputs:

OzFlux Network: Daily evapotranspiration estimates were derived from flux stations from the OzFlux network (Beringer, Hutley et al., 2016a; www.ozflux.org.au; see Figure 4 for locations (see Appendix A: ET monitoring site details) with average annual rainfall overlain to give an indication of the variety of climate areas sampled. Latent heat was obtained using the DINGO (Dynamic INtegrated Gap filling and partitioning for OzFlux) methodology for processing raw flux tower data (Beringer, McHugh et al., 2016). Eddy covariance datasets were quality assured and quality controlled (QA/QC) using the OzFlux standard processing protocol OzFluxQCv2.8.5. The QA/QC processes and corrections involved in the OzFluxQC protocol are described in Eamus et al., (2013).

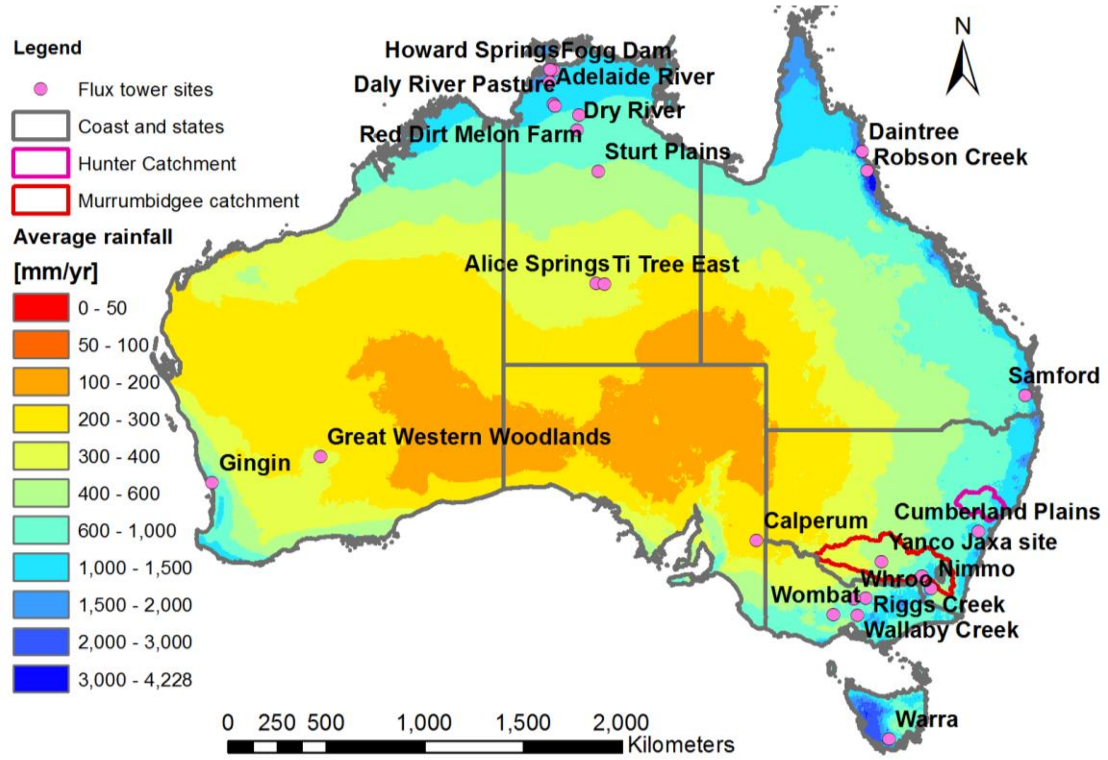

Figure 4. ET flux towers locations and soil moisture monitoring catchment locations also shown. Average annual rainfall is shown to give an indication of the range of climate conditions sampled within Australia.

The period 2001-2013 was used for scientific evaluation, being the intersection of years available for all models available, and had a median of 30% months available for the

25 sites tested (after infilling using the DINGO). This data was not used in calibration of AWRA-L, but some flux tower data was used in calibration of CABLE (see section 3.3).

Satellite retrieval based gridded estimates: CSIRO MODIS reflectance-based Scaling ET (CMRSET; Guerschman et al., 2009) satellite ET covering 2001-2010 and the CSIRO developed Simplified Land Surface Temperature (SLST) algorithm (Van Niel et al., 2012), were compared to the observed point estimates. CMRSET was used in AWRA-L calibration, and also for evaluation purposes. CMRSET is run operationally within the Bureau and produces 250m gridded 8-day cycle national maps of Actual ET based on MODIS satellite data and AWAP climate data, see example AET map for Australia in Figure 5.

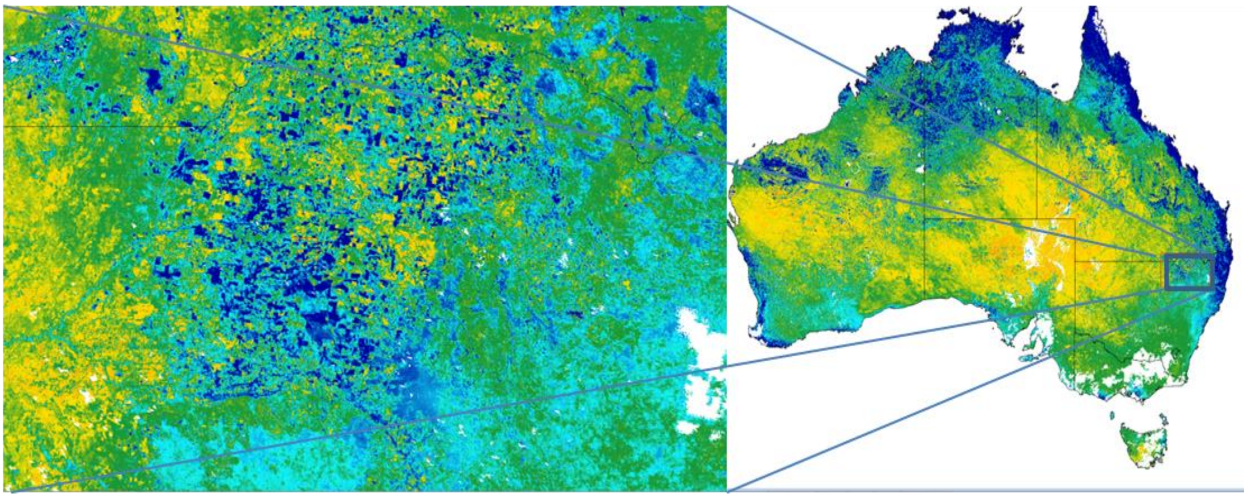

Figure 5. CMRSET derived map of 8 day Actual Evapotranspiration for 04/07/2014 (noting white area shows no data, most likely affected by clouds). Courtesy Juan Pablo-Guerschman CSIRO.

2.2.4 Groundwater Deep Drainage

Shi et al. (2015) collated various datasets which could be used for evaluating AWRA-L modelled deep drainage across Australia:

1. **Long term average:** A long term average recharge dataset has been processed from 6343 individual field estimates of estimates collated by Crosbie et al (2010a; 2010b) with some additional points added that were generated from the Bioregional Assessment Programme (www.bioregionalassessments.gov.au). It

was filtered to remove any data points that had recharge equal to zero or any points that had recharge more than two thirds of the mean annual rainfall. The remaining points were averaged for 2282 grid cells (0.05˚) that are coincident with the AWRA-L model by taking the geometric mean – see Figure 6. The majority of the recharge estimates are based on chloride mass balance estimates, which represent long-term mean annual recharge at the point.

2. **Annual recharge time series:** dataset was created using the water table fluctuation (WTF) method and data for the period 1970-2012 at 438 boreholes in the southeast of South Australia and southwest of Victoria – see Figure 7.
3. **Monthly time-series:** A further monthly time series dataset covering 6 sites over August 2000-Decmber 2002 in the Tomago sand beds in NSW is available.

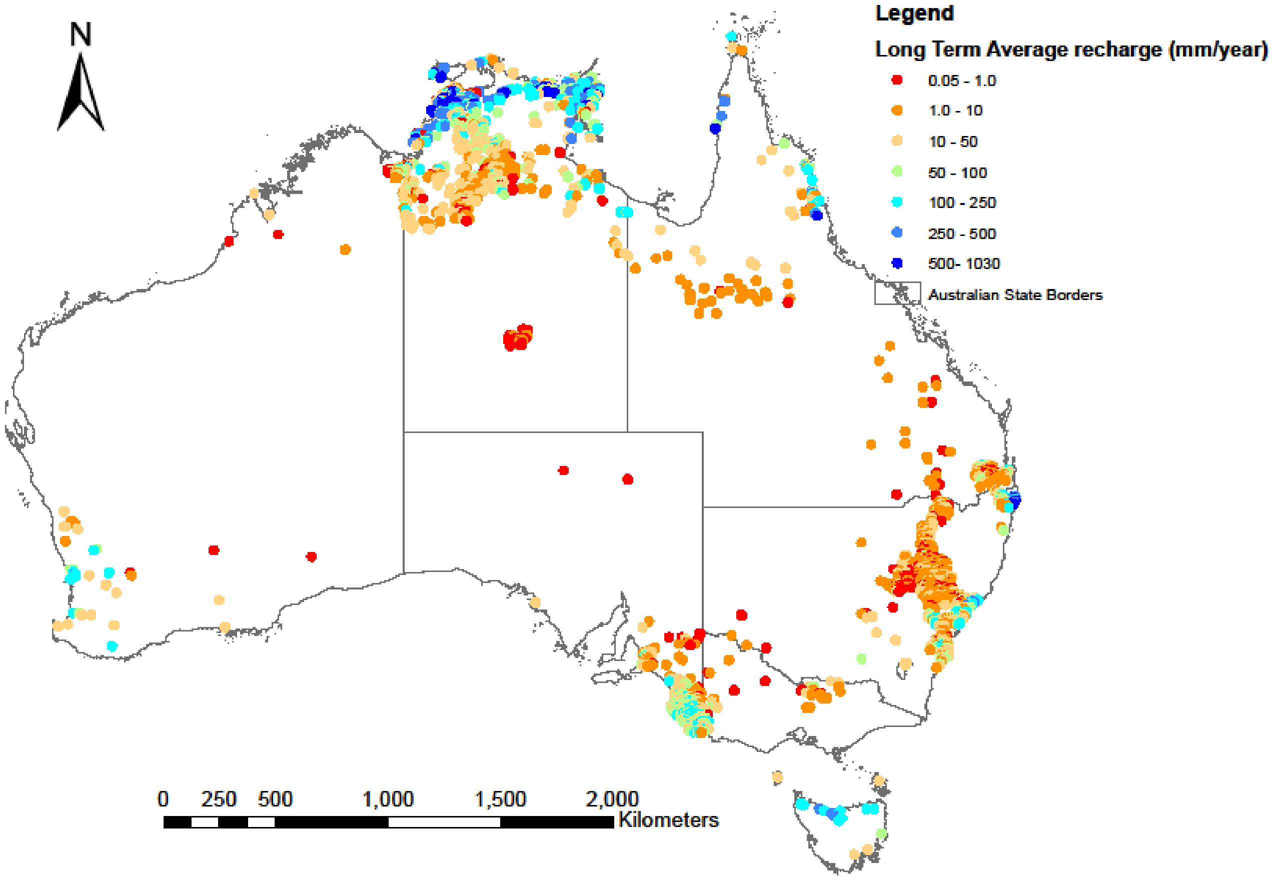

Figure 6. Shi et al (2015) long term average recharge estimates

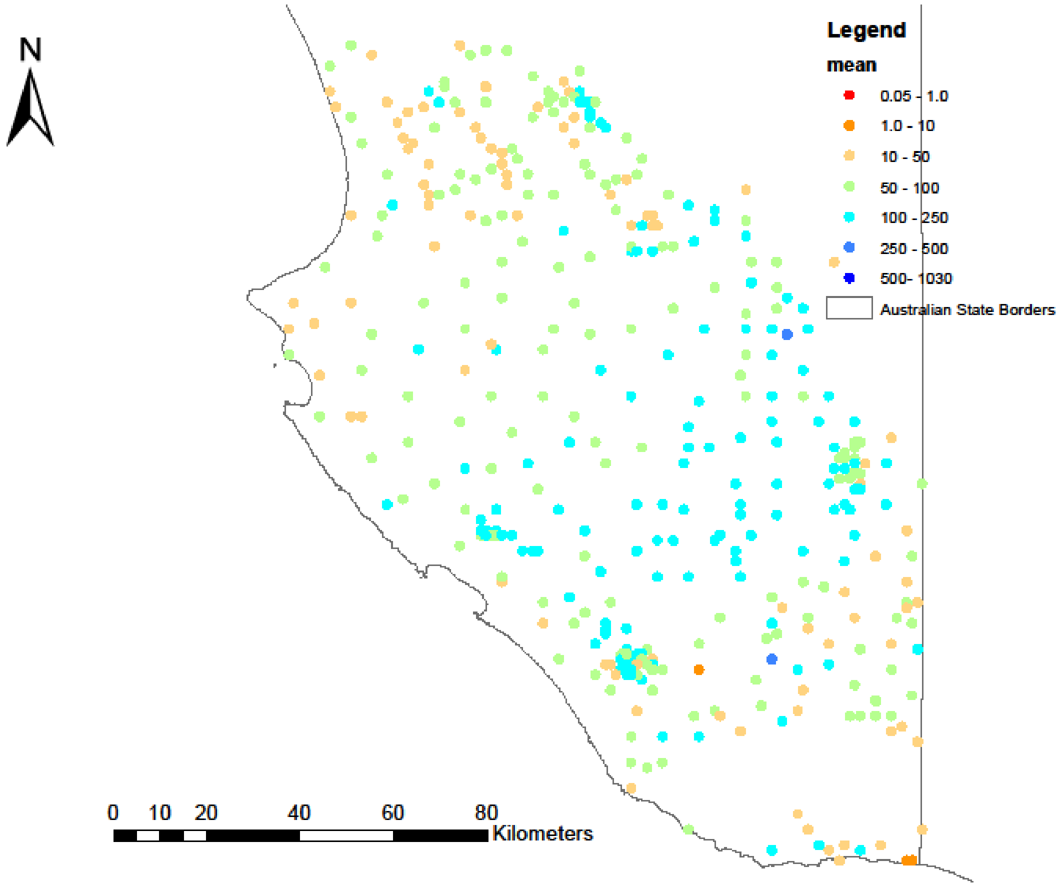

Figure 7. Shi et al (2015) location of annual estimates in South Australia/Victorian state border.

Considering huge variability of deep drainage at any point, and uncertainties associated with derivation of evaluation datasets, it is very hard to have absolute validation of modelled deep drainage of a 25 km^2 grid with the field data. Nevertheless it is of interest to be aware of how AWRA deep drainage estimates compare to other recharge estimates.

3 Models

3.1 AWRA-L

AWRA-L (Van Dijk, 2010; Viney et al., 2014; Viney et al., 2015; Frost et al., 2016b; Frost et al., 2018) is a one dimensional, 0.05° grid based water balance model over the continent that has semi-distributed representation of the soil, groundwater and surface water stores. AWRA-L is a three soil layer (top: 0-10cm, shallow: 10cm-100cm, deep: 100cm-600cm), two hydrological response unit (shallow rooted versus deep rooted) model (Figure 8).

AWRA-L models hydrological processes for:

- Saturation excess overland flow (depending on groundwater store saturation level)
- Infiltration and Hortonian (infiltration excess) overland flow
- Saturation, interflow, drainage and evapotranspiration from soil layers
- Baseflow, evaporation and capillary rise from the groundwater store

With the soil layers modelled separately for 2 (shallow and deep rooted) hydrological response units.

Various spatial datasets are also used to parameterise AWRA-L spatially including:

- Vegetation properties: Estimates of satellite observation derived forest height (1km lidar based estimated derived by Simard et al., 2011), maximum Leaf Area Index (LAI: from analysis of time series of MODIS LAI images) and importantly the proportion deep/shallow rooted (based on estimate of fraction persistent and recurrent vegetation as derived by Donohue et al., 2008).
- Slope and hydraulic conductivity affecting infiltration capacity
- Soil drainage/storage parameters:
- soil hydraulic conductivity and fractional water storage capacity from pedotransfer function applied to clay content from the Soil and Landscape Grid of Australia (www.clw.csiro.au/aclep/soilandlandscapegrid)
- Topology and effective porosity effecting baseflow/saturation (Peeters et al., 2011)

The bottom levels of the top, middle and deep soil layers within AWRA-L are chosen to be 0.1m, 1m and 6m respectively. For further details of the AWRA-L v6.0 algorithms and input data see Frost et al. (2018).

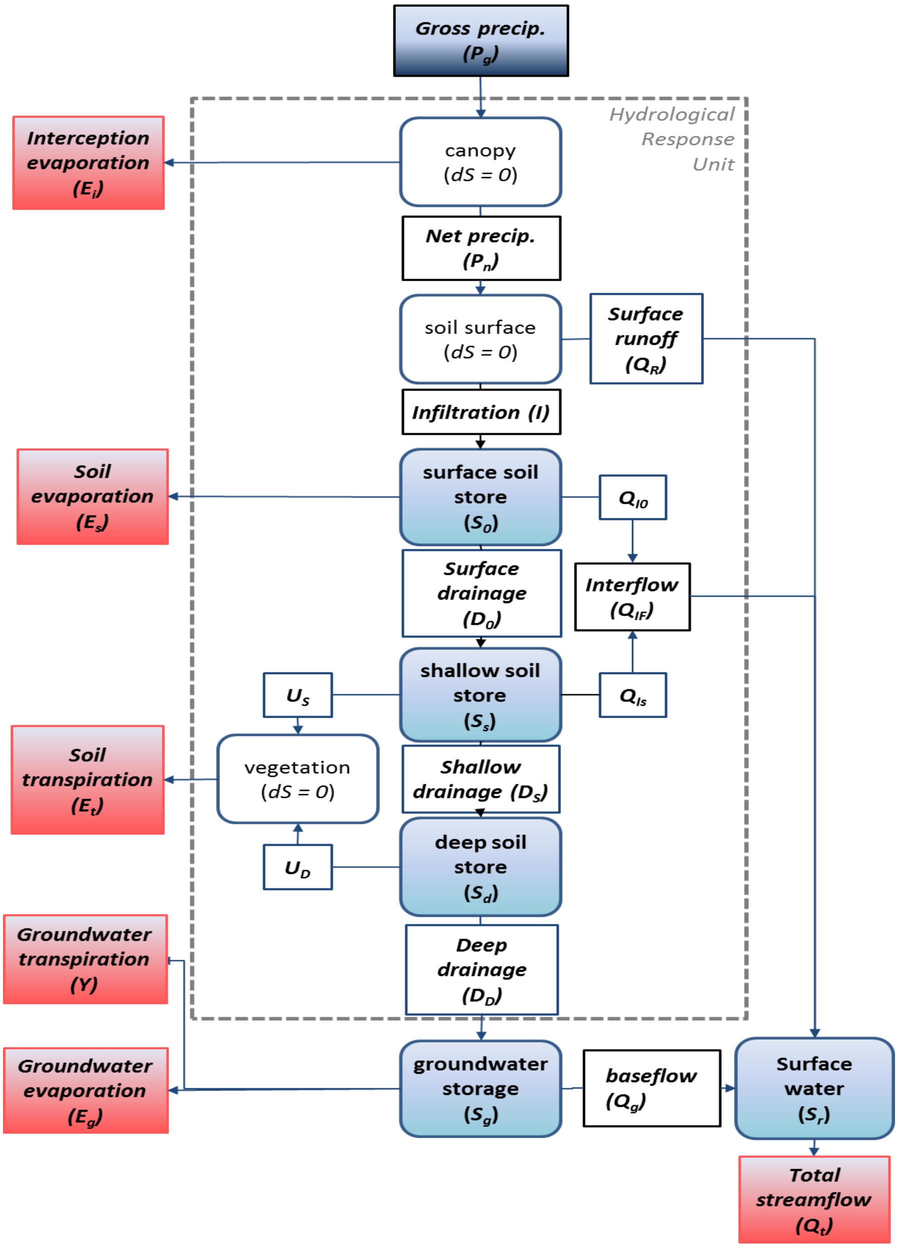

Figure 8. AWRA-L model conceptual diagram showing different hydrological processes

The AWRA-L model has been calibrated to streamflow and catchment averaged soil moisture and ET across Australia. AWRA-L model parameters are currently calibrated nationally over 295 unimpaired catchments as identified within Zhang et al. (2013) – see Figure 2. The remaining 291 separate catchments are kept for scientific validation purposes. Three different datasets are used in calibration over these catchments across Australia including:

- **Catchment streamflow**: covering the period of 1981-2011
- **Catchment evapotranspiration:** CMRSET satellite ET - Satellite retrieval based grid estimates of evapotranspiration covering 2001-2010.
- **Catchment soil moisture:** AMSR-E product (Owe et al., 2008) - Satellite retrieval based grid estimates of soil moisture, covering the period of 2002-2011 have been used.

AWRA-L parameters (i.e. 21 parameters chosen to be free, rather than fixed) are optimised across the continent to maximise a composite function combining the performance according to streamflow, ET and soil moisture at all sites. The following streamflow objective function is evaluated for each catchment simulation (as derived by Viney et al., 2009):

$$F_s = NSE_d - 5\,|\ln(1 + B)|^{2.5} \qquad (1)$$

where NSE_d is the daily Nash-Sutcliffe Efficiency (Eq. 6) and B is relative bias (B) (Eq. 4 - see section 4.1). It is noted that this differs from the objective used for AWRA-L v5 which also used NSE_m the daily Nash-Sutcliffe Efficiency:

$$F_s = (NSE_d + NSE_m)/2 - 5\,|\ln(1 + B)|^{2.5} \qquad (2)$$

Daily soil moisture correlation (R_{SM}) and monthly evapotranspiration (R_{ET}) (defined in Eq. 7) are also used for each catchment according to the weighted function:

$$F = 0.7 * F_s + 0.15 * R_{SM} + 0.15 * R_{ET} \qquad (3)$$

Finally, the national calibration of AWRA-L maximises the grand objective function:

$$grandF = mean(F25\%, F50\%, F75\%, F100\%) \qquad (4)$$

where FX% being the Xth ranked site percentile F value. This objective function aims to get an adequate fit over a wide range of sites, but also to exclude very poor fitting areas (i.e. those below the 25%).

3.2 WaterDyn

The WaterDyn model, developed by CSIRO Marine and Atmospheric Research (Raupach et al., 2009), as part of the AWAP, is another daily national 0.05° grid-based biophysical model of the water balance between the atmosphere and soil which runs at a daily timestep, with monthly and weekly outputs published operationally by CSIRO.

Fluxes contributing to streamflow consist of two components: surface runoff and deep drainage. Surface runoff occurs only when the upper soil layer is completely saturated and is then equal to the rate of precipitation. Deep drainage is a function of the relative soil moisture and the saturated hydraulic conductivity of the soil layer.

WaterDyn, like AWRA-L, also uses daily input gridded data (0.05°) from AWAP although WaterDyn uses the recalibrated daily rainfall surfaces (monthly interpolated surfaces disaggregated daily according to the daily rainfall interpolations), as opposed to the standard daily rainfall surfaces as used by AWRA-L across Australia.

WaterDyn model has two soil layers (and no groundwater store) and is run using various spatial datasets including thickness of soil and saturated volumetric water content of upper/lower soil layers, while constant saturated hydraulic conductivity values were used nationally.

WaterDyn was parameterised using calibration, and investigation of parameter uncertainty, to streamflow from six unimpaired catchments within the Murrumbidgee (see Raupach et al., 2009 for more details). Monthly simulation values were available for evaluation covering January 1900 to February 2014, according to run 26j (www.csiro.au/awap/doc/AWAP_readme_v9.txt).

3.3 CABLE

The CSIRO Atmosphere Biosphere Land Exchange (CABLE) model, is a community global land-surface model developed by CSIRO, the Bureau and partner universities (Kowalczyk et al., 2006; Wang et al., 2011). The CABLE model is being developed with the intention of use within the Australian Community Climate and Earth System Simulator climate model (ACCESS). CABLE is a land surface model, used to calculate the fluxes of momentum, energy, water and carbon between the land surface and the atmosphere and to model the major biogeochemical cycles of the land ecosystem.

Fluxes contributing to streamflow consist of two components: surface runoff and deep drainage. Drainage, compared to WaterDyn and AWRA-L, is modelled as gravitational drainage from the lowest soil layer (with draining in the soil layers modelled according to Richard's equation solution assuming a relationship between hydraulic conductivity and soil moisture content).

CABLE uses daily input climate gridded data (0.05°) from the Bureau operational AWAP service. It is noted that CABLE (like WaterDyn) uses the recalibrated daily rainfall surfaces (monthly interpolated surfaces disaggregated daily according to the daily rainfall interpolations), as opposed to the standard daily rainfall surfaces as used by AWRA-L model. Data are downscaled from daily to hourly time steps (on the half-hourly) using a weather generator (Haverd et al., 2013).

10 soil layers are included in this implementation of CABLE (0.022, 0.058, 0.07, 0.15, 0.30, 0.30, 0.30, 1.20, 3.0, and 4.5m thicknesses depth from topmost to bottommost layer). Secondly, the default CABLE v1.4 soil and carbon modules were replaced respectively by the Soil-Litter-Iso (SLI) soil model (Haverd and Cuntz, 2010) and the CASA-CNP biogeochemical model (Wang et al., 2010) – see Haverd et al (2013). Spatially varying soil properties used by BIOS2 are bulk density, clay and silt fractions, saturated hydraulic conductivity, suction at saturation, field capacity, wilting point, and saturated volumetric water content – see Haverd et al (2013) for further details.

CABLE parameters were calibrated/constrained according to:

- 50 unimpaired catchment streamflow records spread across Australia (10 from each bioclimatic region except desert) used to compare to long term streamflow (precipitation-ET) from the model. ie. does not attempt to model short term temporal dynamics of streamflow.
- 6 OzFlux sites Evapotranspiration and gross primary production of Carbon (Howard Springs, Daly River Savanna, Daly River Pasture, Sturt Plains, Tumbarumba, Virginia Park)

Monthly simulation values were available for evaluation covering January 1900 to December 2013.

3.4 Summary of model characteristics

The salient features of AWRA-L and peer models (WaterDyn and CABLE) are summarised in Table 1.

Table 1. Summary of AWRA-L, WaterDyn, and CABLE model characteristics

	WaterDyn	***CABLE***	***AWRA-L (v5.0 and v6.0)***
Reference	Raupach et al (2009)	Wang et al (2011); Haverd et al (2013)	Viney et al (2015); Frost et al (2016); Frost et al (2018).
Developer	CSIRO/BoM/ABARES	CSIRO/BoM + universities	CSIRO/BoM
Purpose	Monitoring terrestrial water balance	Land surface scheme for the Australian Community Climate and Earth-System Simulator (ACCESS)	Water resources reporting, assessment and monitoring
Soil layers (spatially varying properties)	2 (depth, saturated volumetric water content)	10 (saturated hydraulic conductivity, field capacity, etc)	3 (saturated hydraulic conductivity, % available water holding capacity)
Calibration	Parameter calibration and sensitivity analysis to 6 catchments in Murrumbidgee	Calibration to derived ET (50 catchments across 10 climate zones within Australia) and flux tower data	Streamflow over ~300 catchments and satellite soil moisture and ET

3.5 Lumped-rainfall runoff models

Two lumped catchment conceptual rainfall-runoff models are used for streamflow comparison purposes against AWRA-L model:

- GR4J (Perrin et al., 2003): a 4-parameter model derived from empirical analysis over many catchments towards finding the most efficient/parsimonious model structure.

- Sacramento (Burnash, 1995): The Sacramento model is a conceptual catchment water balance model developed for the U.S. National Weather Service that models the rainfall–runoff process at daily time-steps. A 13-parameter implementation was used here.

These models are calibrated in a different way to AWRA-L, in that they are calibrated for individual catchments, rather than finding a single parameter set to cover the entire model domain. Once the parameters are found for the calibration catchments, they are transferred by nearest-neighbour regionalisation to the closest validation catchments nearby. Nearest-neighbour regionalisation is a method used as a practical approach to regionalisation/predictions in ungauged basins, to produce the best performance possible where calibration is possible, but to also allow prediction in areas where the model cannot be calibrated. It is noted that the calibration process only uses streamflow (rather than also using satellite derived soil moisture and evapotranspiration as now used in AWRA-L). For further details of the methods applied for the conceptual rainfall runoff modelling approach used here see Ramchurn and Frost (2014).

4 Evaluation approach

4.1 Statistics used in evaluation

Various statistics are calculated for each catchment/point to assess the models including:

Relative bias (B)

$$B_i = \sum_{t=1}^{T} \frac{Q_{mit} - Q_{oit}}{\bar{Q}_{oi}} \tag{5}$$

Monthly and daily Nash-Sutcliffe Efficiency (NSE), and

$$NSE_i = 1 - \sum_{t=1}^{T} \frac{(Q_{mit} - Q_{oit})^2}{(Q_{oit} - \bar{Q}_{oi})^2} \tag{6}$$

Pearson's correlation coefficient (r)

$$r_i = \frac{\sum_{t=1}^{T}(Q_{oit} - \bar{Q}_{oi})(Q_{mit} - \bar{Q}_{mit})}{\sqrt{\sum_{t=1}^{T}(Q_{oit} - \bar{Q}_{oi})^2}\sqrt{\sum_{t=1}^{T}(Q_{mit} - \bar{Q}_{mi})^2}} \tag{7}$$

where Q_{mit} and Q_{oit} are the modelled and observed values respectively for site *i* and time step *t*, for a total of T_i available observations, and $\bar{Q}_{oi}$ is the mean observed and $\bar{Q}_{mi}$ the mean modelled data for site *i*.

The bias and monthly NSE statistics in particular are seen as good metrics for judging the models performance for the purposes of AWRA-L streamflow. Pearson's correlation coefficient is a good indicator for variables where the bias (and absolute value) of the variable is not as important as matching the variability (e.g. soil moisture and actual ET). Finally, the Kling-Gupta Efficiency (KGE; Gupta et al., 2009) is also used for decomposing streamflow performance across the country according to (a) correlation, (b) alpha (ratio of standard deviation of model over observed standard deviation, a measure of variability), and (c) beta (ratio of mean of model over observed mean, a measure of bias).

4.2 Evaluation criteria

The AWRA-L model was primarily developed for water resource applications across Australia. Therefore, the evaluation criteria are primarily based on the available observed hydrological data across Australia. In general, improvements in model performance should be judged on data reserved for validation (i.e. separate to calibration data) – so that performance is more assured for predictions in ungauged basins – following the principles outlined in Refsgaard and Henriksen (2004). It is to be noted that all observed datasets have uncertainty associated with them, and are essentially a model also. Future improvement of the AWRA-L (and other) models can be judged according to the performance of AWRA-L v6 according to these metrics.

Primary metric – Assessment of AWRA-L against observed streamflow

- NSE for daily (NSE_d) and monthly (NSE_m) runoff
- Relative Bias (B) in long-term averages

Secondary metric – Assessment of AWRA-L against derived data

- Soil moisture: Daily and monthly correlation of probe-based point soil moisture sampled for the profile (0-90cm) with AWRA-L soil moisture.
- Actual ET: Daily and monthly correlation of flux tower ET with AWRA-L ET.
- Deep drainage: Correlation between long-term reliable point measurements of recharge with AWRA-L deep drainage.

Tertiary metric: Behaviour

- Checking AWRA-L simulations of internal fluxes and checking sensibility of national AWRA simulations for reporting purposes (e.g., no major irregular spatial patterns due to regionalisation, time-series plots for select locations).

The assessment criteria according to the observed data (the Primary and Secondary metrics above) are provided in Table 2. Aspirational targets are provided based on how the Bureau would like the AWRA model to perform, based on what we consider to be reasonable baseline performance characteristics and past experience of performance with peer models. For example: For the majority of catchments to perform better than the average/climatology for streamflow – therefore, we want to have less than 5% at zero NSE (equivalent to climatology) – and have at least half of the catchments above 0.5 NSE (considered good performance for how the model is intended to be used). AWRA-L is assessed against these criteria in Chapter 5.

Chapter 6 presents a brief examination of AWRA-L outputs according to the Tertiary metric based on behaviour of the model for reporting purposes.

Table 2. AWRA-L assessment criteria

Variable	Assessed against	Assessment criteria	Comparison with simulations from alternative models	Aspirational target
Streamflow	Gauged streamflow (calibration and validation sites)	Daily NSE Monthly NSE Bias	CABLE WaterDyn Rainfall-runoff model (Sacramento and/or GR4J) – local calibration/ nearest neighbour regionalisation	**Daily NSE**: Less than 5% catchments with NSE<0 greater than 50% catchments with NSE>0.5 **Bias:** (a) 50 % of catchments with -30%<bias<30%, (b) 90% with -50%<Bias<100%, and (c) No systematic spatial pattern of under- or over-estimation (i.e. low Bias when aggregated, mean and median bias close to 0)
Soil moisture	Profile soil moisture from dedicated field observations	Daily and monthly correlation	CABLE WaterDyn	50% with daily correlation > 0.75 50% with monthly correlation >0.75
Actual ET	Flux ET	Monthly correlation	CABLE WaterDyn	Monthly correlation – 95% sites/cells with R>0.5, >50% sites/cells with R>0.8
Deep drainage	National Long term average dataset Annual time series dataset	Bias Annual correlation	CABLE WaterDyn	25% bias value below zero 75% bias value above zero Median annual correlation above 0.5

5 Evaluation according to observed data

Various statistics for calibration and validation catchments are now presented for each model to assess their performance against observed hydrological data sets including streamflow, soil moisture, ET and recharge. Calibration/validation catchment statistics are presented using boxplots, showing the cumulative distribution of the statistics across all sites, with the box indicating the 25% percentile, median and 75% percentile (e.g. 25% percentile for the 295 calibration sites means that 74 sites have lower values). Tables of statistics are presented in the case where there are insufficient sites for representation as a cumulative distribution (e.g. flux tower ET), or where alternative presentation of the statistics was meaningful (e.g. recharge).

5.1 Streamflow

The AWRA-L model performance has been assessed against other national models (WaterDyn and CABLE) as well as typical rainfall-runoff catchment scale models (GR4J and Sacramento) across Australia. Calibration and validation daily NSE/KGE and monthly NSE and relative bias are plotted in Figure 9. AWRA-L v6 improves over AWRA-L v5 for daily NSE and KGE (Fig 10abcd), with a slight degradation according to monthly NSE (Fig 10(ef)), reflecting the updated objective function removing the monthly NSE component.

For the national landscape/landsurface models, the results show that AWRA-L model performs better for streamflow than WaterDyn and CABLE according to monthly NSE and bias (Figure 9(efgh)) over the AWRA-L calibration and validation catchments across Australia. This result is expected due to a) AWRA-L being designed to represent runoff characteristics more accurately; and b) AWRA-L is calibrated directly to streamflow characteristics nationally.

For the locally calibrated nearest neighbour regionalised rainfall runoff models, AWRA-L performs worse in the calibration catchments than the locally calibrated models, due to the differing calibration approach used (AWRA sacrifice local performance for as good performance as possible across multiple sites and variables). In particular, bias is near zero for the locally calibrated models due to each of the models having terms that can effectively match the average flow at a particular site where calibrated, while AWRA-L tries to minimise the bias over a set of sites. However, over the validation catchments AWRA-L bias has less spread about zero, providing confidence in the spatial predictive qualities of AWRA-L which has significance for predictions in ungauged basins. *Significantly, AWRA-L v6 monthly performance for the validation catchments (Figure 9(f)) is approaching the performance of nearest neighbour regionalised rainfall-runoff models (GR4J and Sacramento), even though AWRA-L is not calibrated purely to streamflow like the lumped rainfall-runoff models*. AWRA-L performs approximately 0.07 worse for daily NSE (Figure 9(f)) than the locally calibrated models.

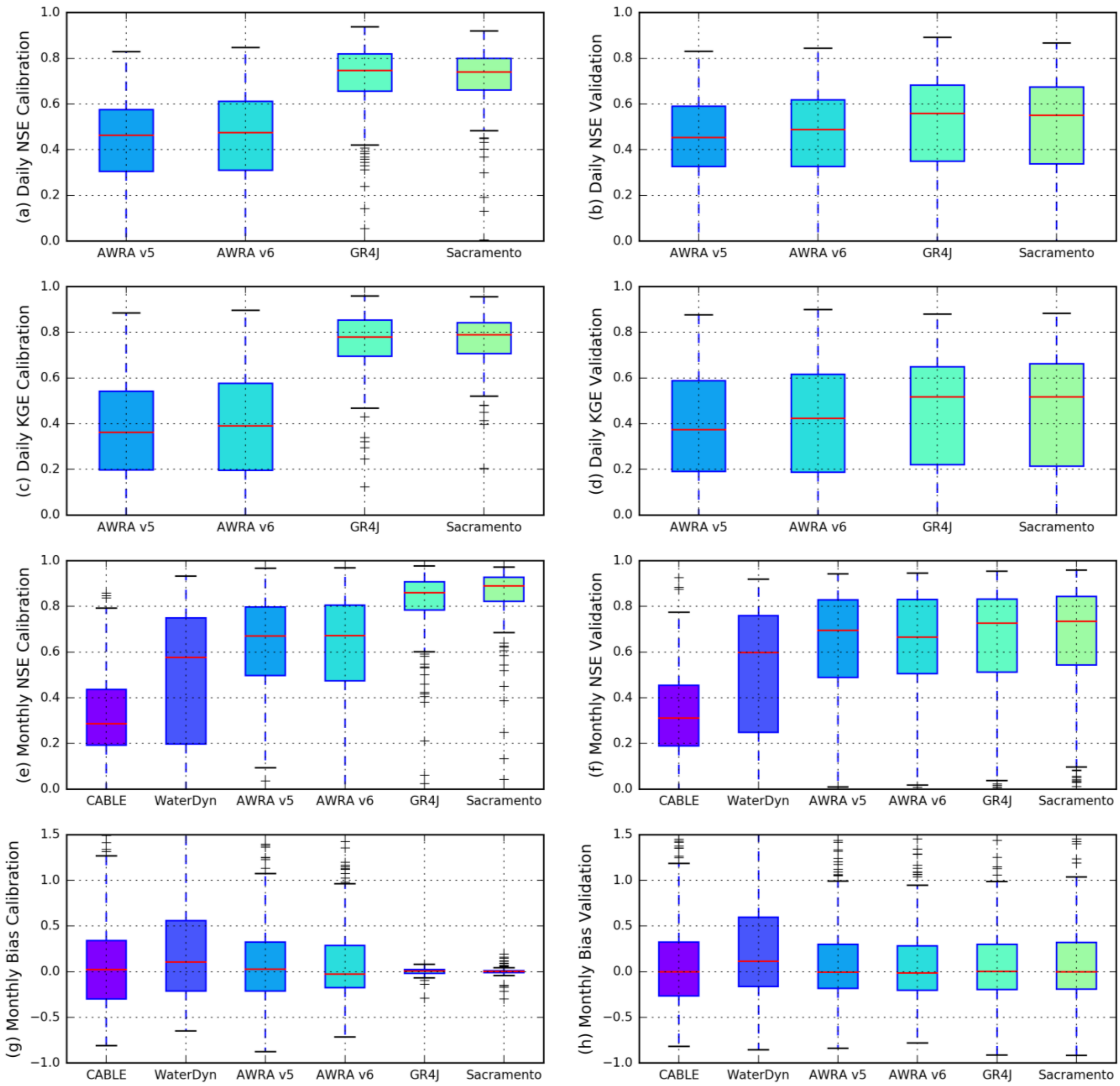

Figure 9. Streamflow statistics for calibration (left) and validation (right) catchments; including Daily NSE (a,b) and KGE (c,d) and Monthly NSE (e,f) and Bias (g,h)

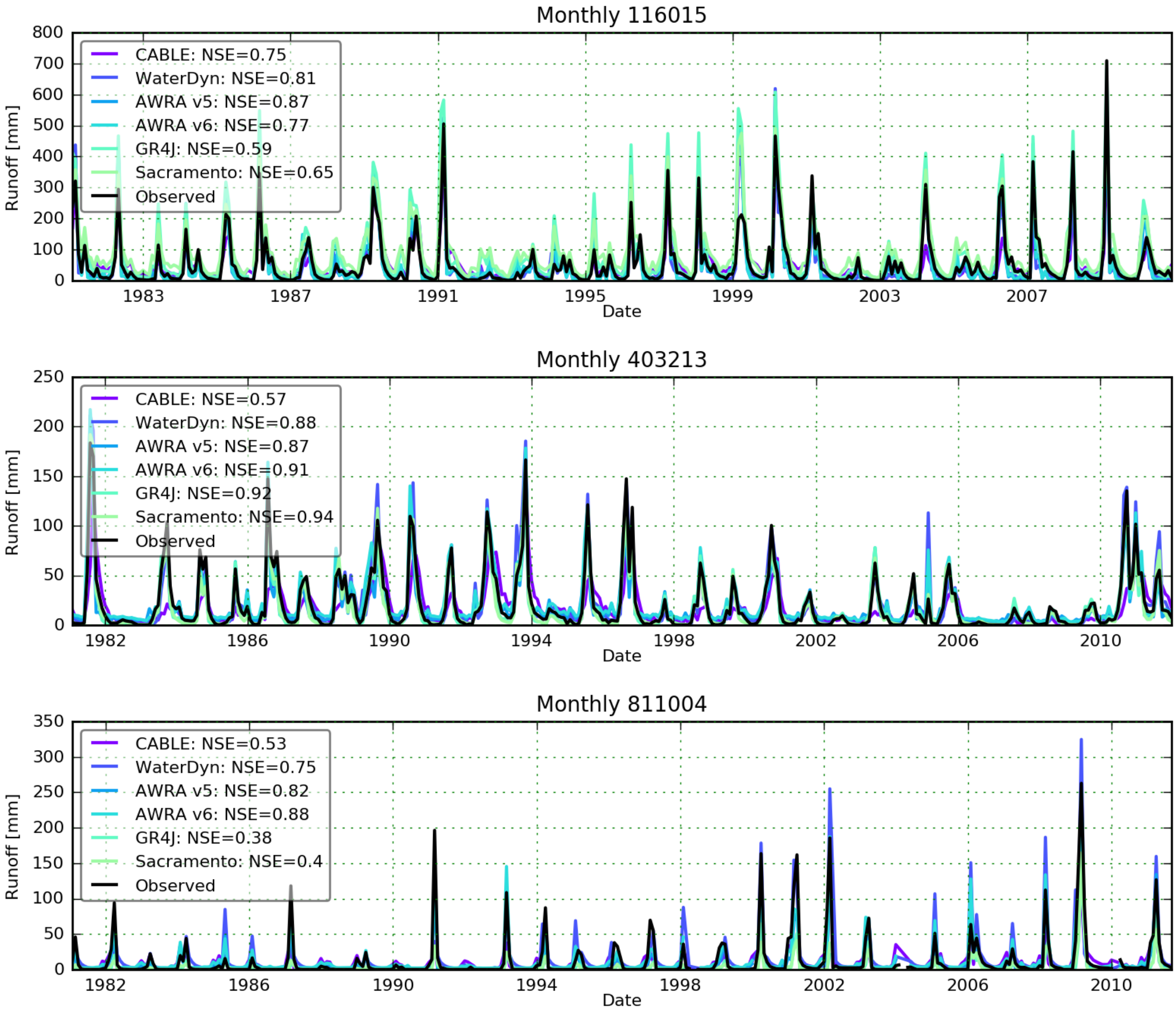

Figure 10. Select monthly catchment runoff time series comparing modelled and observed flow.

Selected catchment monthly time series for AWRA-L (and WaterDyn and CABLE) are compared to those from the lumped rainfall models obtained when using a nearest neighbour method (which uses the parameters obtained from the closest locally calibrated catchment) in Figure 10. Catchment details and AWRA-L water balance time series are provided in Section 6. This comparison indicates that AWRA-L provides more reliable estimates than a simple method of estimating flow in ungauged catchments using locally calibrated models in these locations. Furthermore, AWRA-L provides a range of water balance outputs (ET, soil moisture and deep drainage) – where the lumped conceptual model does not. *Overall, the result that AWRA-L performs well in validation provides confidence in its use for spatial prediction – across the country – for water resource assessment and scenario analysis purposes.*

The performance of AWRA-L according to daily NSE, monthly NSE and relative bias are presented in Table 3 to Table 5. Evaluation criteria listed in Table 2 are bold in the tables for model benchmarking purposes and comparison to the aspirational targets.

Table 3. Daily NSE percentiles for each model

Calibration	0%	5%	25%	50%	75%	95%	100%	Validation	0%	5%	25%	50%	75%	95%	100%
CABLE*								**CABLE***							
WaterDyn*								**WaterDyn***							
AWRA-Lv5	-15.42	-1.16	0.30	0.46	0.58	0.71	0.83	**AWRA-Lv5**	-41.79	-0.30	0.33	0.45	0.59	0.73	0.83
AWRA-Lv6	-14.70	**-0.65**	0.31	**0.47**	0.61	0.73	0.85	**AWRA-Lv6**	-11.57	**-0.06**	0.33	**0.49**	0.62	0.74	0.84
GR4J	0.00	0.47	0.66	0.75	0.82	0.88	0.94	**GR4J**	-11.12	-0.76	0.35	0.56	0.68	0.79	0.89
Sacramento	-1.95	0.50	0.66	0.74	0.80	0.86	0.92	**Sacramento**	-8184.58	-1.73	0.34	0.55	0.67	0.80	0.87
Benchmark		**-0.65**		**0.47**				**Benchmark**		**-0.06**		**0.49**			

* Daily results are not available for the comparison

Table 4. Monthly NSE percentiles for each model

Calibration	0%	5%	25%	50%	75%	95%	100%	Validation	0%	5%	25%	50%	75%	95%	100%
CABLE	-286.86	-0.72	0.19	0.29	0.44	0.71	0.86	**CABLE**	-23.80	-0.38	0.19	0.31	0.45	0.73	0.93
WaterDyn	-515.30	-3.25	0.20	0.58	0.75	0.86	0.93	**WaterDyn**	-43.96	-1.76	0.25	0.60	0.76	0.88	0.92
AWRA-L v5	-22.55	-0.51	0.50	0.67	0.80	0.89	0.97	**AWRA-L v5**	-43.77	-0.23	0.49	**0.69**	0.83	0.91	0.94
AWRA-L v6	-22.76	**-0.42**	0.47	**0.67**	0.81	0.91	0.97	**AWRA-L v6**	-13.56	**-0.19**	0.50	0.67	0.83	0.91	0.95
GR4J	-0.02	0.60	0.78	0.86	0.91	0.94	0.98	**GR4J**	-16.81	-0.24	0.51	0.73	0.83	0.91	0.95
Sacramento	-4.95	0.67	0.82	0.89	0.93	0.96	0.97	**Sacramento**	-1943.31	-0.25	0.54	0.74	0.84	0.93	0.96
Benchmark		**-0.42**		**0.67**				**Benchmark**		**-0.19**		**0.69**			

Table 5. Relative bias percentiles for each model

Calibration	0%	5%	25%	50%	75%	95%	100%	Validation	0%	5%	25%	50%	75%	95%	100%
CABLE	-0.81	-0.54	-0.30	0.03	0.34	1.36	130.68	**CABLE**	-0.82	-0.56	-0.27	0.00	0.32	1.44	10.84
WaterDyn	-0.65	-0.48	-0.21	0.11	0.56	1.83	113.41	**WaterDyn**	-0.85	-0.46	-0.16	0.11	0.60	2.38	14.98
AWRA-L v5	-0.87	-0.49	-0.21	0.02	0.32	**1.36**	21.24	**AWRA-L v5**	-0.84	-0.46	**-0.18**	-0.01	0.29	**1.28**	8.69
AWRA-L v6	-0.71	**-0.43**	**-0.18**	-0.03	**0.29**	1.38	31.20	**AWRA-L v6**	-0.78	**-0.43**	-0.20	-0.01	**0.28**	1.40	5.44
GR4J	-0.29	-0.04	-0.02	0.01	0.02	0.05	0.09	**GR4J**	-0.91	-0.60	-0.20	0.00	0.30	1.34	7.41
Sacramento	-0.30	-0.02	-0.01	0.00	0.01	0.05	1.78	**Sacramento**	-0.92	-0.54	-0.19	0.00	0.32	1.41	7.36
Benchmark		**-0.43**	**-0.18**		**0.29**	**1.36**				**-0.43**	**-0.18**		**0.28**	**1.28**	

AWRA-L v6 currently does not meet the aspirational daily NSE criteria (50% NSE of 0.49 rather than 0.5, 5% NSE at -0.1 rather than above zero); although it is not far off. In terms of bias AWRA-L does meet the criteria in validation for 50% of sites (25% to 75%) to be within -0.3 and 0.3. It does not meet the criteria for the 90% of sites (5% to 95%) being between -0.5 and 1; with the 95% value showing a bias of 140% in validation.

AWRA-L (a) daily NSE and (b) monthly relative bias are plotted in Figure 12 to evaluate spatial performance. AWRA-L v6 performs well (above 0.5 daily NSE) in Coastal NSW and Victoria, the majority of Queensland, the majority of Tasmania, South Western West Australia and coastal catchments in the Northern Territory. AWRA-L has lower performance for catchments along the Great Dividing Range (from Victoria to NSW/Queensland border) and also in Western Australia along the Darling Scarp. This appears to be partly due to positive bias in these areas. Possible reasons for this bias include (a) deep soil store rooting depth being insufficient (e.g. Jarrah forests of Darling Scarp having roots to 20 metres rather than 6m currently) causing underestimated ET, (b) losses to groundwater systems/transfer that are currently unaccounted for (ie. losses cannot be included in the system currently) and (c) losses due to inadequate routing procedure, amongst other possibilities.

AWRA-L daily (a) KGE, (b) correlation, (c) KGE alpha (std. dev. modelled/ std. dev. observed) and (d) KGE beta (mean modelled/mean observed) are plotted spatially in Figure 12 to further investigate spatial performance. Underestimation of variability (alpha) appears to be an aspect of poor performance. Further investigation is required to determine the reasons for this underestimation.

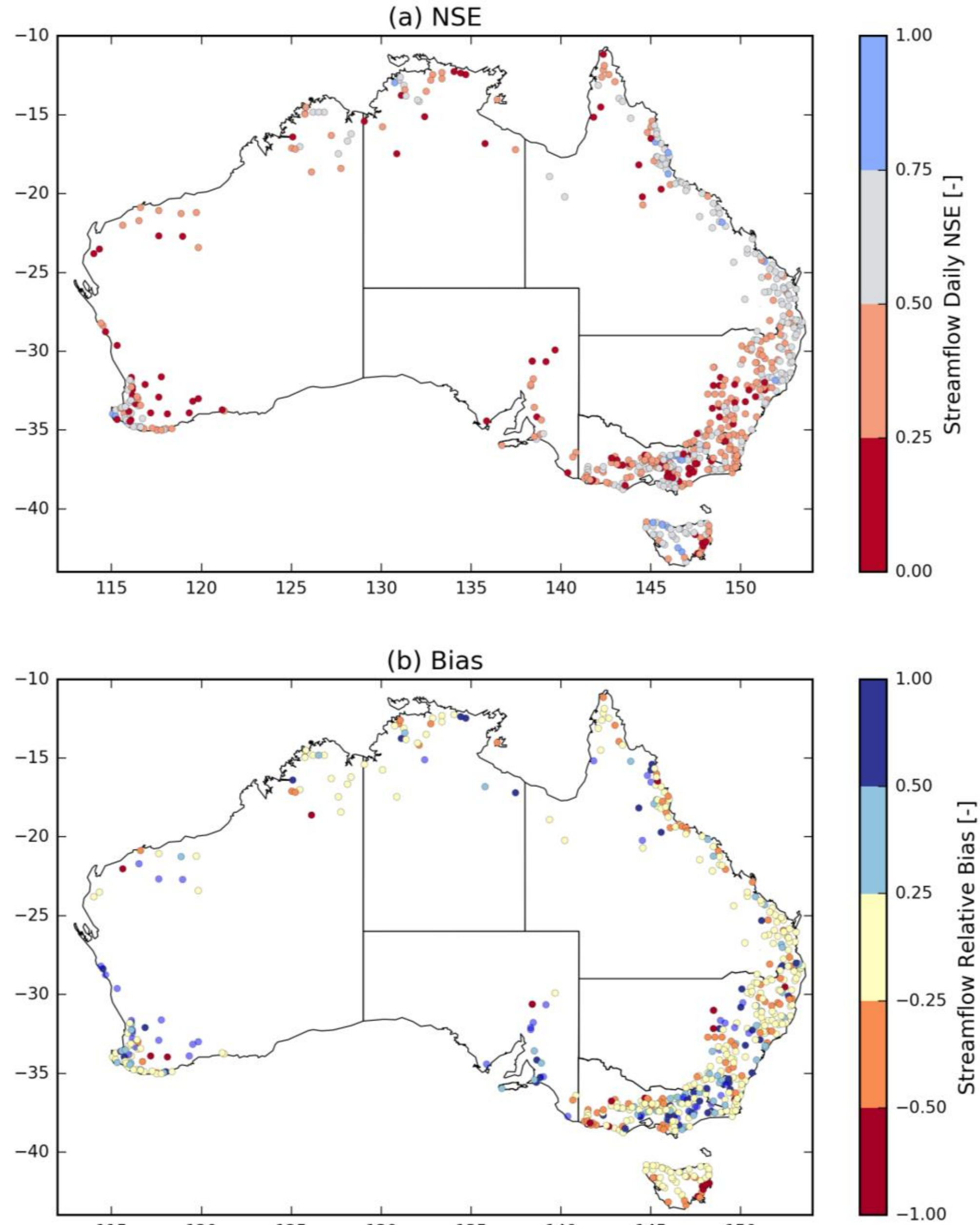

Figure 11. Map of AWRA-L v6 runoff (a) daily NSE and (b) monthly relative bias compared to streamflow. Calibration and validation sites shown.

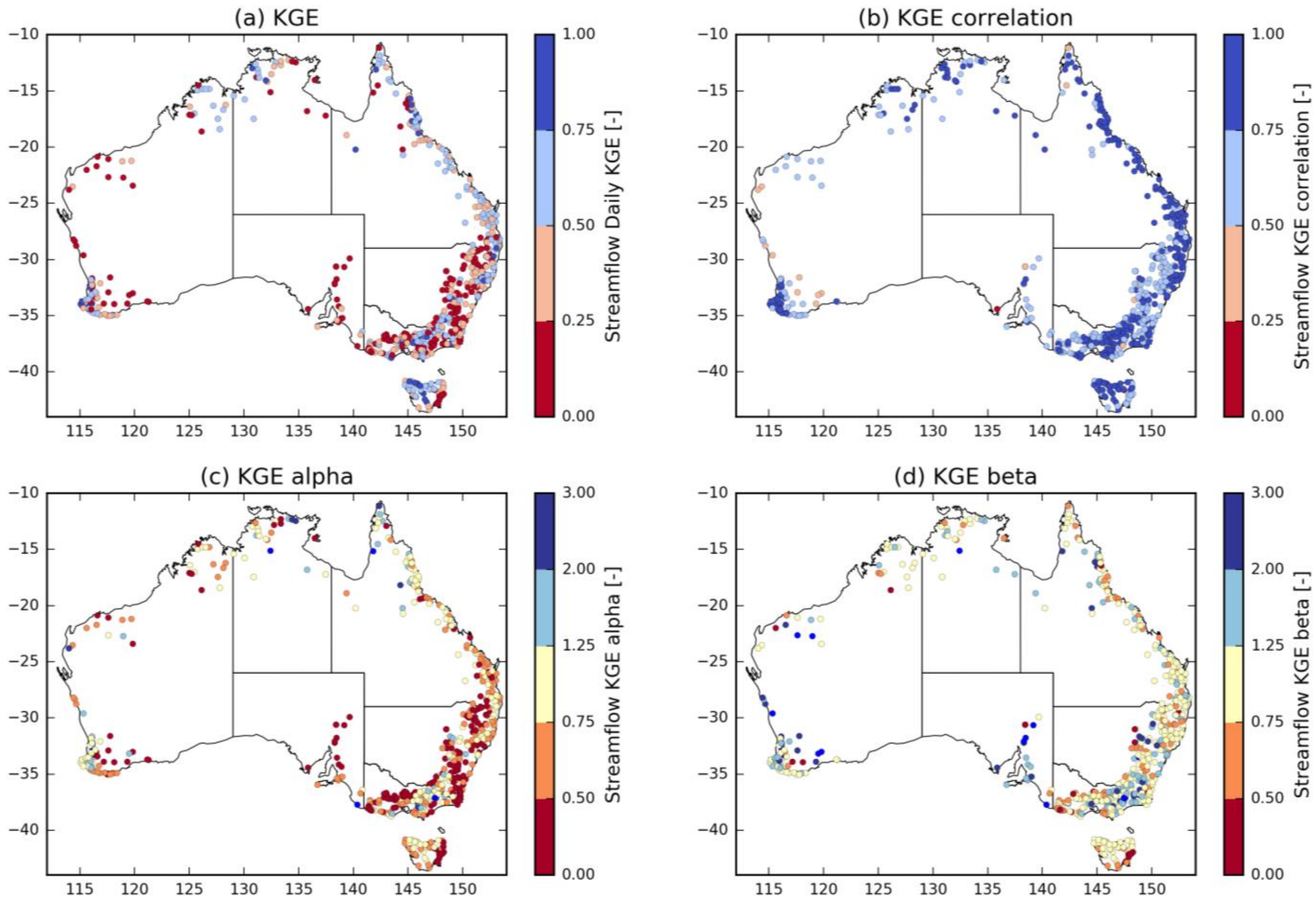

Figure 12. Map of AWRA-L v6 runoff (a) KGE, (b) KGE correlation, (c) KGE alpha and (d) KGE beta compared to streamflow. Calibration and validation sites shown.

5.2 Soil moisture

The point based profile (0-90cm) estimates of soil moisture are compared to the layers of each model (weighted according to degree of overlap) according to monthly correlation for the OzNet Murrumbidgee data (Figure 13) and Upper Hunter SASMAS data (Figure 14).

This comparison uses the entire record that is available covering the model simulations (i.e. up until 2013). AWRA-L v6 improves over AWRA-L v5 for median correlation, while performance drops for the 25th and 75th percentile for the SASMAS set. AWRA-L and

CABLE perform similarly for profile soil moisture, with WaterDyn worse for the SASMAS Upper Hunter evaluation (noting WaterDyn was parameterised based on testing in 6 catchments in the Murrumbidgee).

The overall result of this comparison is therefore that AWRA-L represents profile (0-90cm) soil moisture temporal dynamics as well as CABLE, and better than WaterDyn (particularly for the Upper Hunter SASMAS data).

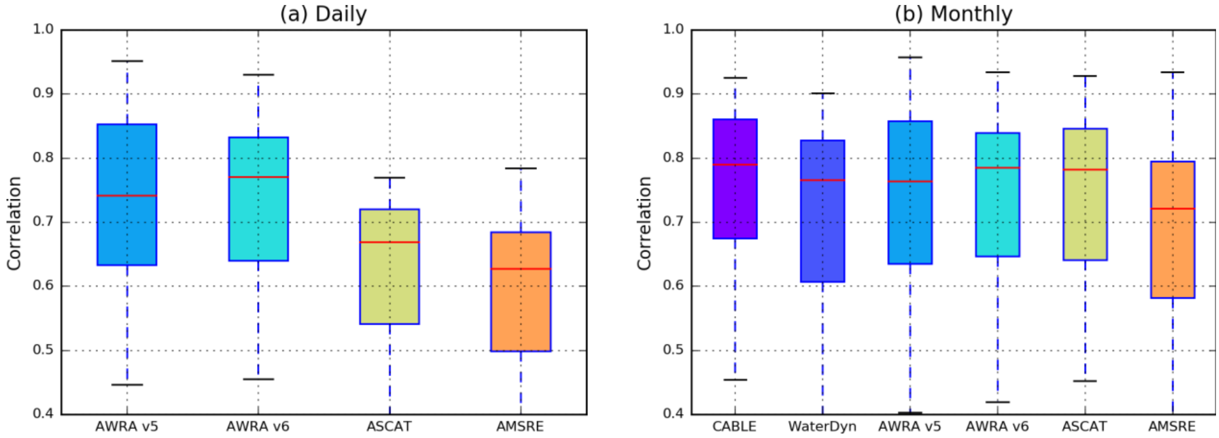

Figure 13. (a) Daily and (b) Monthly correlation of models against Murrumbidgee OzNet data 2001-2013 profile (0-90cm) soil moisture. Satellite data is evaluated over a shorter period (AMSR-E: 2002-2011, ASCAT: 2007-2011) – and relates only to the top few cm.

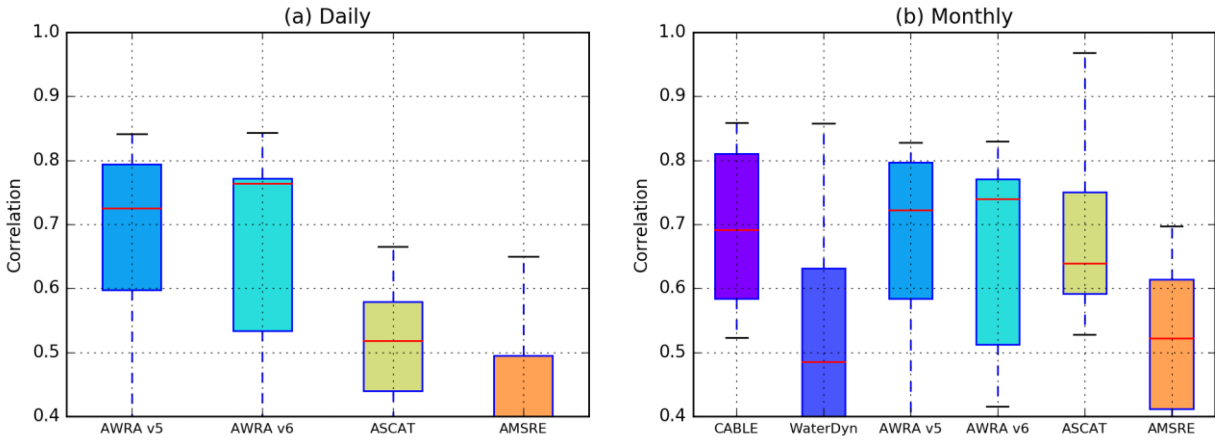

Figure 14. (a) Daily and (b) Monthly correlation of models against Upper Hunter SASMAS 2003-2011 profile (0-90cm) soil moisture. Satellite data is evaluated over a shorter period (AMSR-E: 2002-2011, ASCAT: 2007-2011) – and relates only to the top few cm.

Table 6 presents the daily and monthly profile (0-90cm) correlation statistics, for evaluation against the Evaluation criteria listed in Table 2. AWRA-L v6 performs best and is above 0.75 daily correlation for the SASMAS and OzNet datasets, with only monthly SASMAS slightly below at 0.74.

Table 6. Ranked correlation of profile (0-90cm) daily and monthly soil moisture AWRA-L and satellite estimates against OzNet (2001-2013) and SASMAS (2003-2011) data. Noting satellite data is evaluated over a shorter period (AMSR-E: 2002-2011, ASCAT: 2007-2011) – and relates only to the top few cm.

OzNet	0%	5%	25%	50%	75%	95%	100%	SASMAS	0%	5%	25%	50%	75%	95%	100%
								Daily Soil Moisture							
AWRA-L v5	0.45	0.51	0.64	0.74	0.83	0.93	0.95	**AWRA-L v5**	0.27	0.34	0.60	0.73	0.79	0.84	0.84
AWRA-L v6	0.46	0.57	0.64	**0.77**	0.83	0.91	0.93	**AWRA-L v6**	0.23	0.37	0.53	**0.76**	0.77	0.83	0.84
ASCAT	0.40	0.43	0.54	0.67	0.72	0.75	0.77	**ASCAT**	0.06	0.18	0.44	0.52	0.58	0.65	0.66
AMSRE	0.30	0.39	0.51	0.61	0.68	0.77	0.78	**AMSRE**	0.08	0.13	0.37	0.39	0.49	0.58	0.65
Benchmark				**0.77**				**Benchmark**				**0.76**			
								Monthly Soil Moisture							
CABLE	0.38	0.52	0.67	0.79	0.86	0.90	0.93	**CABLE**	0.11	0.18	0.58	0.69	0.81	0.83	0.86
WaterDyn	0.29	0.39	0.61	0.75	0.83	0.85	0.90	**WaterDyn**	0.17	0.25	0.36	0.49	0.63	0.75	0.86
AWRA-L v5	0.40	0.52	0.64	0.76	0.85	0.93	0.96	**AWRA-L v5**	0.14	0.23	0.58	0.72	0.80	0.82	0.83
AWRA-L v6	0.42	0.56	0.65	**0.79**	0.84	0.92	0.93	**AWRA-L v6**	0.10	0.29	0.51	**0.74**	0.77	0.82	0.83
ASCAT	0.45	0.48	0.64	0.78	0.85	0.89	0.93	**ASCAT**	0.53	0.55	0.59	0.64	0.75	0.91	0.97
AMSRE	0.39	0.44	0.58	0.72	0.79	0.92	0.93	**AMSRE**	0.00	0.12	0.41	0.52	0.61	0.69	0.70
Benchmark				**0.79**				**Benchmark**				**0.74**			

The profile layer monthly correlation values are plotted for the OzNet and SASMAS sites (Figure 15) to give an indication of how AWRA-L performs spatially following the analysis presented in Frost et al. (2015). Further, daily time series of the profile soil moisture for AWRA v5 and v6 against the probe data is presented, to provide examples of how well AWRA-L produces drying and wetting of the soil as experienced during the Millennium drought, particularly the years 2006 and 2007 (see Potter et al, 2010). The observed data is plotted as percentage water per volume, while the modelled values are transformed from mm to percentage water per volume according to mapped soil field capacity and wilting point values – see Frost et al. (2015) for further details.

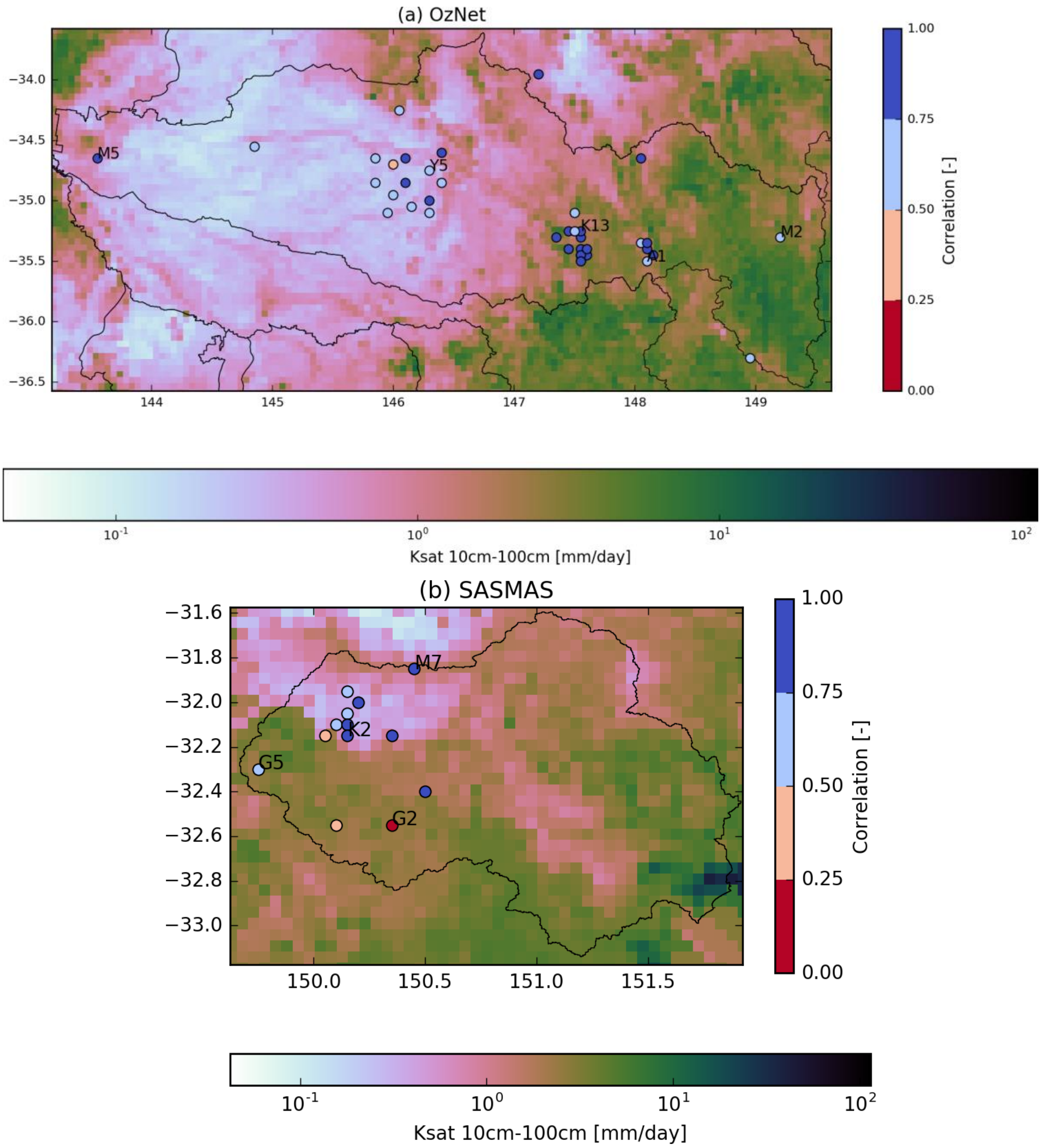

Figure 15. AWRA-L Monthly correlation for profile soil moisture of a) Murrumbidgee (OzNet) and b) Upper Hunter (SASMAS) data. AWRA-L saturated conductivity (Ksat) for shallow layer (10cm-100cm) underlain. Select site labels are shown.

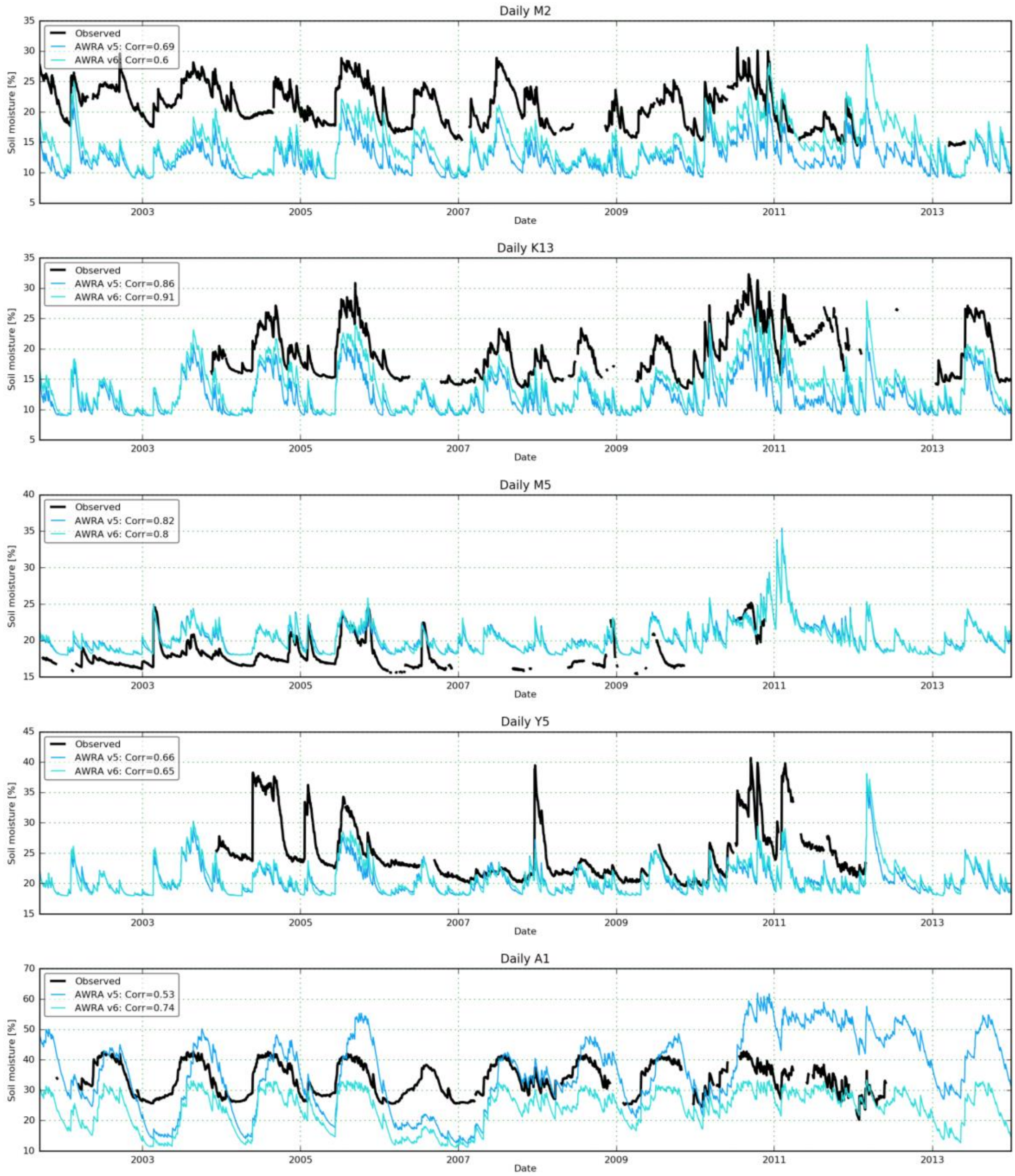

Figure 16. Five Murrumbidgee OzNet sites daily profile (0-90cm) soil moisture and model/satellite estimates.

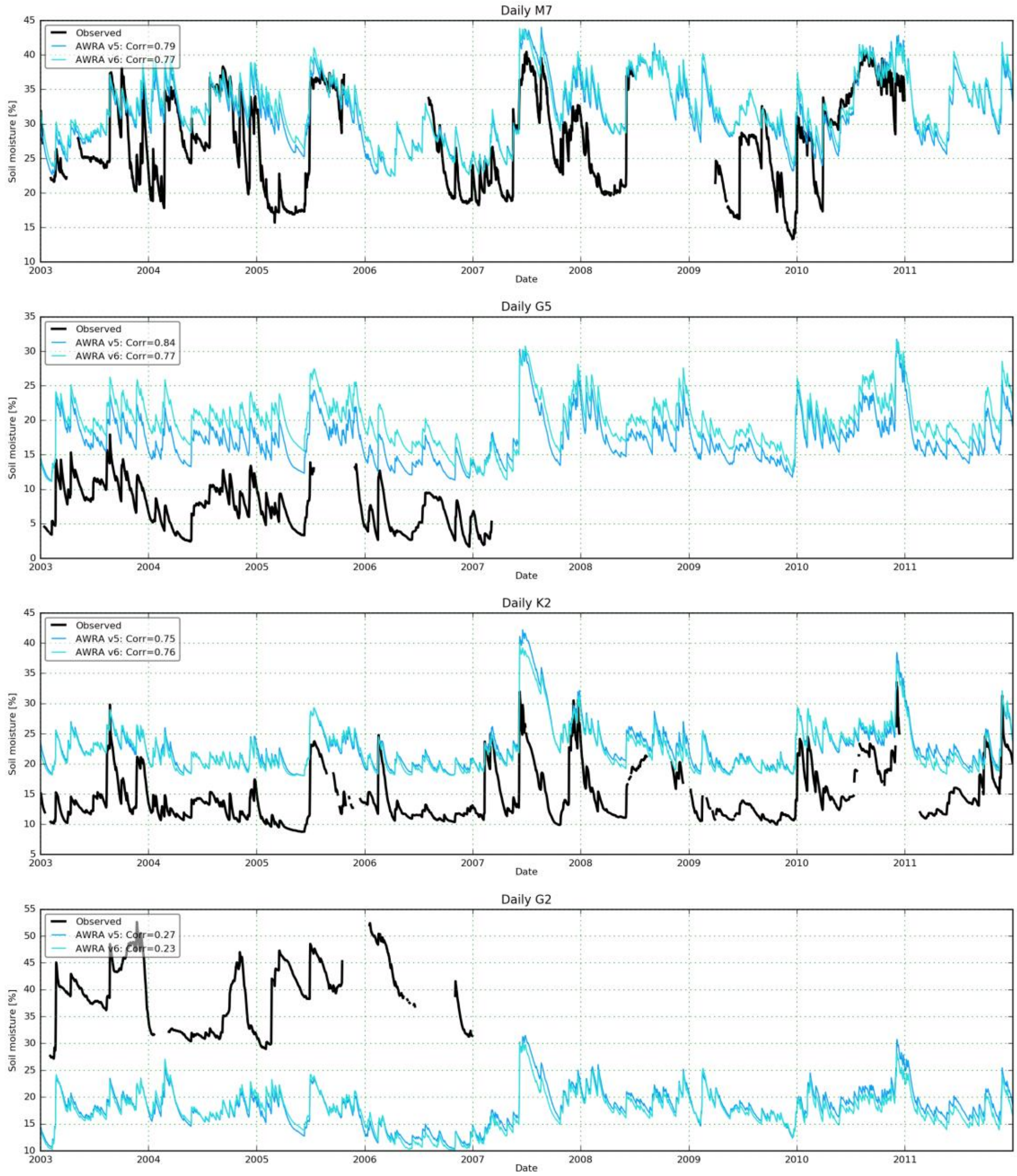

Figure 17. Four Upper Hunter SASMAS sites daily profile (0-90cm) soil moisture and model/satellite estimates

A brief evaluation of the performance of the models at reproducing the top layer soil moisture is presented in Appendix C. Following the results presented in Frost et al. (2015), AWRA-L performs relatively worse compared to WaterDyn and CABLE when evaluated against the probe 0cm-5cm/8cm and satellite based data.

It is noted interpretation of the results presented on soil moisture should consider:

- The difference in point scale observations compared to large grid scale (~25 km by 25km for the models, larger for satellite data) outputs – with the point not reflecting the sampling area of the models evaluated.
- Uncertainties in probe calibrations: with some sites being better calibrated than others.
- Inaccuracies of transfer and quality control: with some sites likely to have timing errors and/or the wrong data.
- Inaccuracies in satellite soil moisture product derivations.

It is expected these datasets will improve over time, with further calibration and quality control.

5.3 Actual Evapotranspiration

The point based estimates of actual ET derived from infilled flux tower data (DINGO) at 25 sites was compared to the CABLE, WaterDyn, AWRA-L, CMRSET and SLST gridded outputs over the entire simulation period (2001-2013; noting that the CMRSET and SLST do not cover this entire period) available according to correlation (Figure 18) and relative bias (Figure 19). CABLE and WaterDyn are roughly equal in terms of monthly correlation and better than AWRA-L v5. AWRA-L v6 improves over AWRA v5 in terms of correlation and bias; although is below CABLE and WaterDyn in terms of monthly correlation and bias performance. CABLE is expected to perform best here, as: (a) it is calibrated to the Tumbarumba, Howard Springs and Virginia Park ET (albeit over a different time period), while the other models are not, and (b) it contains a more complete formulation of land-surface energy and water related dynamics.

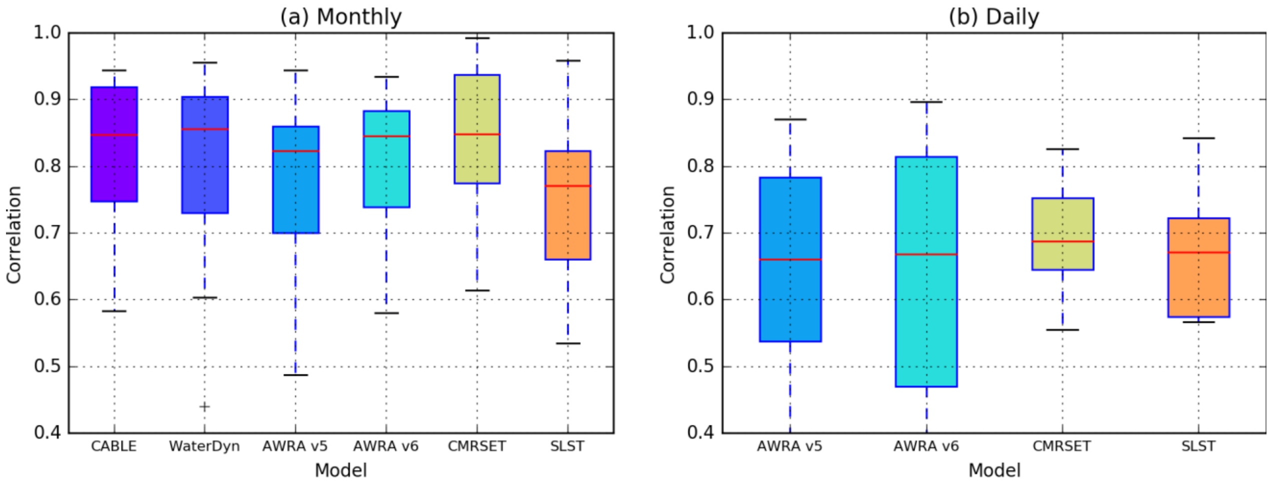

Figure 18. Correlation over 2001-2013 of DINGO derived actual ET compared to modelled (a) Monthly and (b) Daily data

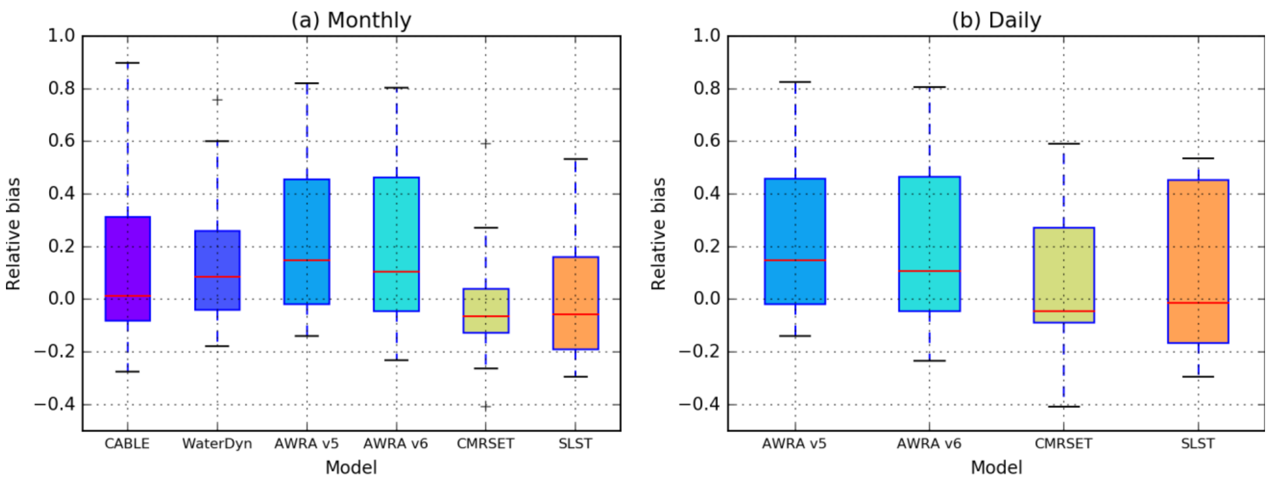

Figure 19. Relative bias over 2001-2013 of DINGO derived actual ET compared to modelled (a) Monthly and (b) Daily data

Table 7 and Table 8 present the monthly and daily correlation and bias statistics, for evaluation against the criteria listed in Table 2. AWRA-L monthly correlation is close to the target criteria (monthly correlation for the 5th percentile catchment is greater than 0.5 and 50th percentile is greater than 0.9) – but below CABLE and WaterDyn particularly at the 5th percentile. AWRA-L v6 improves over AWRA-L v5, and performs equivalently to CABLE at the 50th percentile, although lower at the 5th percentile. CABLE and WaterDyn provide benchmarks for future performance testing.

Table 7. Monthly (a) correlation and (b) relative bias of modelled estimates compared to DINGO data 2001-2013. Noting satellite based estimates CMRSET and SLST do not cover the same period as models.

Correlation	0%	5%	25%	50%	75%	95%	100%	Relative bias	0%	5%	25%	50%	75%	95%	100%
CABLE	-0.01	0.32	0.75	0.85	0.92	0.94	0.94	**CABLE**	-0.28	-0.17	-0.08	0.01	0.31	0.71	0.90
WaterDyn	0.44	0.61	0.73	0.86	0.90	0.95	0.96	**WaterDyn**	-0.18	-0.14	-0.04	0.08	0.26	0.60	0.76
AWRAL v5	0.27	**0.49**	0.70	0.82	0.86	0.93	0.94	**AWRA-L v5**	-0.14	-0.11	-0.02	0.15	0.46	0.70	0.82
AWRAL v6	0.12	0.41	0.74	**0.85**	0.88	0.92	0.93	**AWRA-L v6**	-0.23	-0.17	-0.05	0.11	0.46	0.79	1.80
CMRSET	0.61	0.70	0.77	0.85	0.94	0.98	0.99	**CMRSET**	-0.41	-0.31	-0.13	-0.07	0.04	0.38	0.59
SLST	0.17	0.41	0.66	0.77	0.82	0.92	0.96	**SLST**	-0.29	-0.27	-0.19	-0.06	0.16	0.52	0.53
Benchmark		**0.49**		**0.85**											

Table 8. Daily correlation of AWRA-L compared to DINGO data 2001-2013.

Correlation	0%	5%	25%	50%	75%	95%	100%
AWRAL v5	0.29	0.37	0.54	0.66	0.78	0.86	0.87
AWRAL v6	0.15	0.19	0.47	0.67	0.81	0.85	0.90
CMRSET	0.14	0.28	0.64	0.69	0.75	0.82	0.83
SLST	0.19	0.21	0.57	0.67	0.72	0.84	0.84

A second comparison (Figure 20) was undertaken using the time period that the satellite ET data was available (2001- 2010). This gives an indication of how well the satellite data represents ET, compared to the three models. WaterDyn, CABLE and AWRA-L are superior in terms of median monthly correlation, although CMRSET produces some correlations at the high deciles that are higher than the models, while

SLST performs relatively poorly. This suggests that this version of CMRSET provides some value in terms of a dataset that can be used for evaluation, calibration and assimilation into AWRA-L. However, given that some correlations in the lowest performing sites (25th percentile) are below that of AWRA-L, the use of the data in calibration may be detracting ET performance in some cases. Further work is required to ensure that the actual ET dataset adds value to AWRA-L calibration.

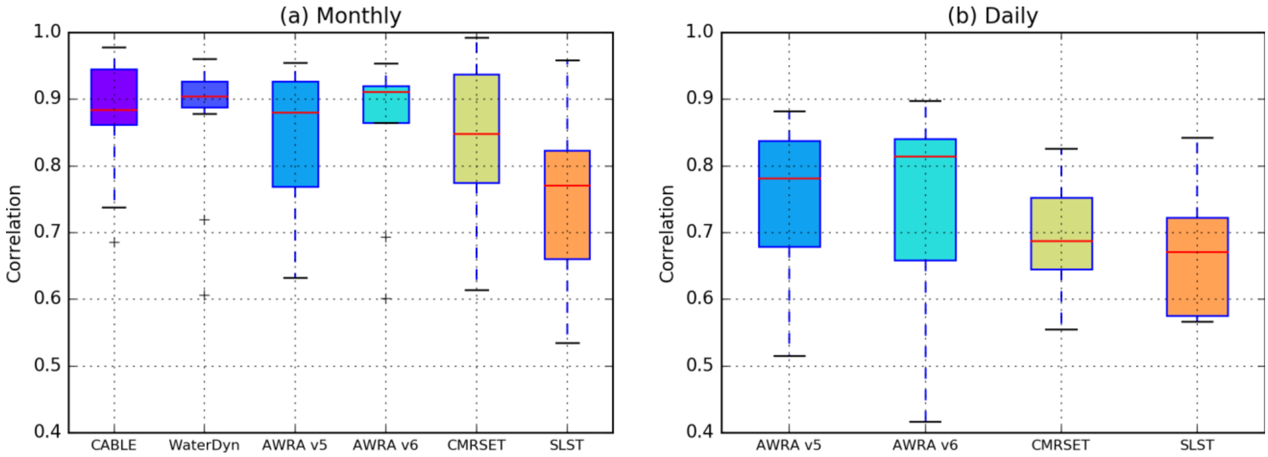

Figure 20. Correlation over 2001-2010 of DINGO derived actual ET compared to modelled (a) Monthly and (b) Daily data

Figure 21 shows the spatial plots of the AWRA-L (a) correlation and (b) relative bias compared to the DINGO ET data. Spatially we see that AWRA-L overestimates in some areas when compared to DINGO ET (several sites in central Australia and eastern Australia), however the remaining sites compare favourably.

Finally it is noted that there is significant uncertainty associated in closing energy balance from flux tower data. Wilson et al. (2002) carried out a comprehensive evaluation of energy balance closure across 22 sites using eddy covariance flux towers ranging from Mediterranean to temperate and arctic climate. Results indicated a general lack of energy balance closure at most sites, with a mean imbalance in the order of 20%. Further, the infilling procedure used here for infilling also has uncertainties. In particular, the method used to infill data up until the start of the calendar year (before the flux tower observations start) shows significant uncertainty (e.g. Cumberland in early 2011 before start in September 2011 – see Figure 40 in Appendix A: ET monitoring site details and time series). In general, evapotranspiration is difficult to definitively measure and all comparisons are therefore indicative only.

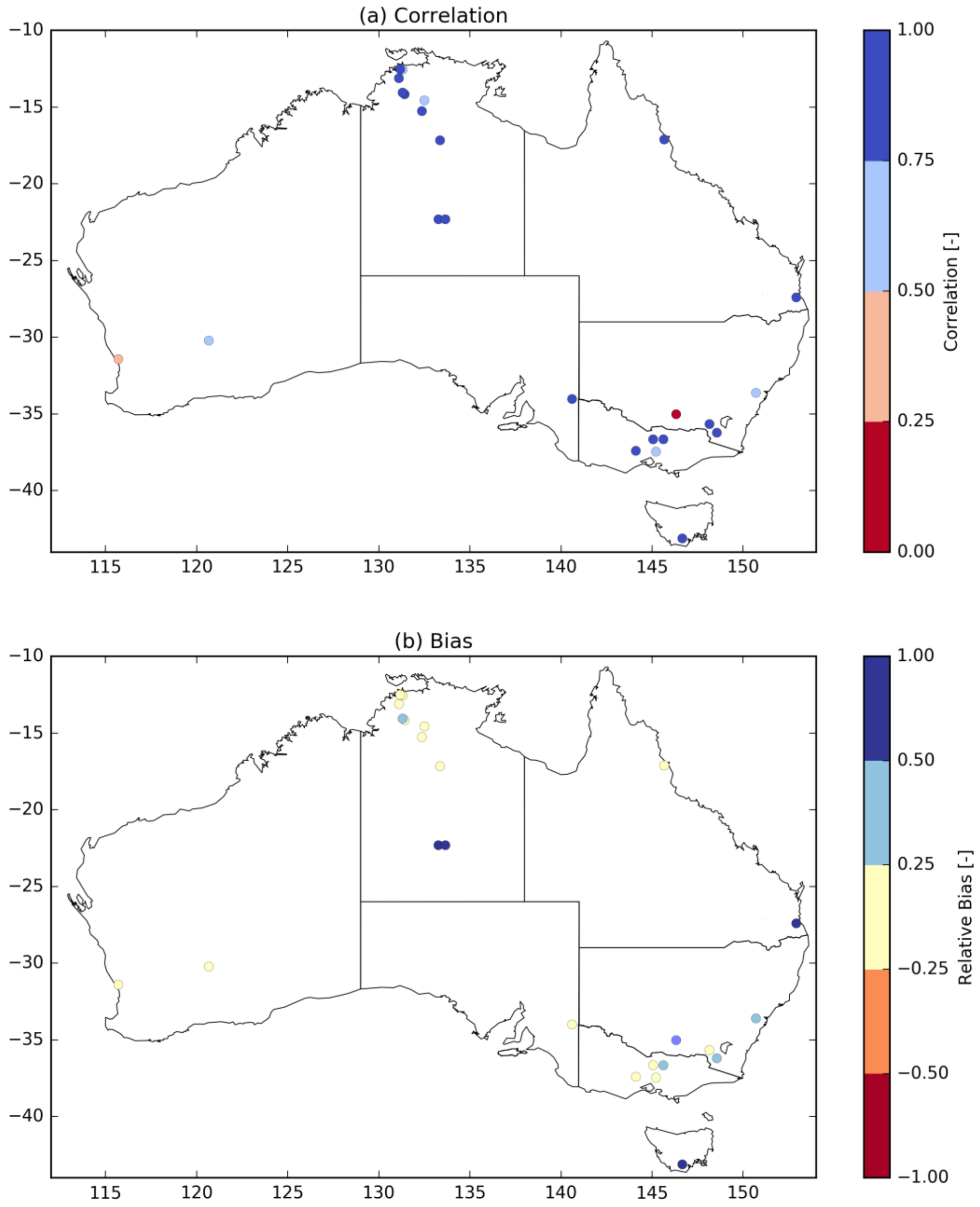

Figure 21. AWRA-L ET monthly (a) correlation and (b) bias compared with DINGO data

5.4 Groundwater deep drainage

Modelled deep drainage was compared against the Long Term Average national collated recharge dataset covering 2282 grid cells – with relative bias calculated (Figure 22). Modelled deep drainage was also compared to the annual time series recharge dataset spanning 1970-2012 covering 438 sites using the water table fluctuation method (Figure 23); with annual correlation and relative bias presented.

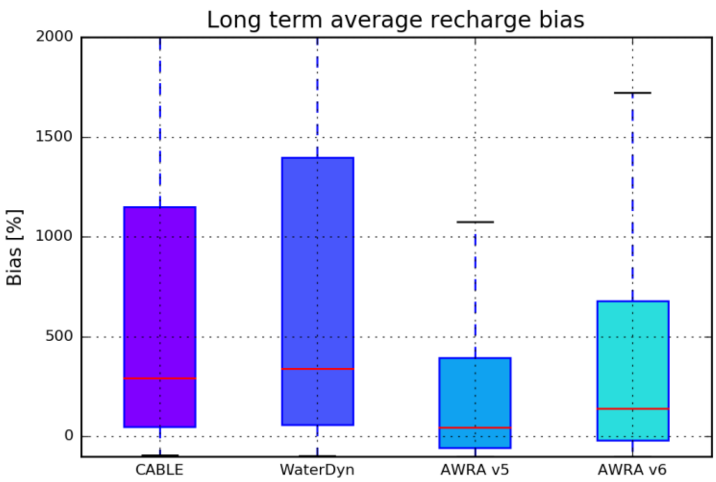

Figure 22. Modelled outputs versus Long Term Average recharge dataset (2282 grid cells across Australia) relative bias

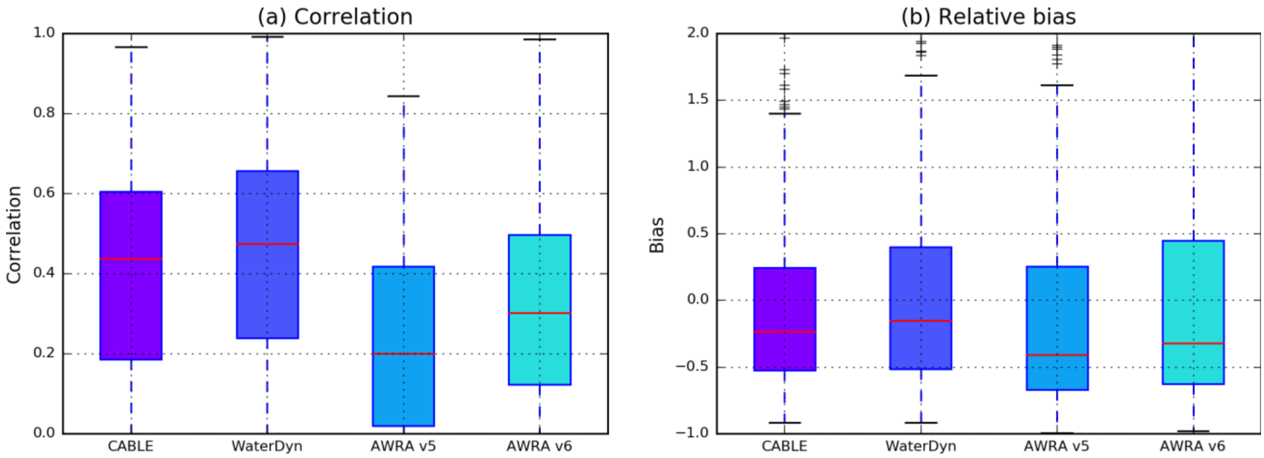

Figure 23. Modelled outputs versus annual recharge dataset (438 sites in South Australia) (a) correlation and (b) relative bias

AWRA-L performs well comparatively against the national long term average recharge dataset, with a median bias just above zero. However, for the annual time-series, AWRA-L is most biased (-40%), and has the lowest median correlation of the 3 models. This difference in results between the two datasets is attributed to variability in local performance in differing areas by AWRA-L. AWRA-L v6.0 improves in terms of performance according to annual correlations compared to AWRA-L v5.0.

The three models are now compared against the specified evaluation criteria:

- Bias for at least 25% of the Long term average sites to be below zero and bias for at least 25% to be greater than zero.
- Annual correlation (for annual data) of at least 50% of sites to be greater than 0.5.

AWRA-L accords with the bias constraint, where the other models do not. All models do not achieve the aspirational target of 0.5 median correlation.

Table 9. Deep drainage evaluation criteria

	0%	5%	25%	50%	75%	95%	100%
Model	**National Long Term Average Dataset relative bias**						
CABLE	-0.94	-0.50	0.48	2.93	11.50	85.40	1827.24
WaterDyn	-0.97	-0.50	0.60	3.40	13.98	111.69	1514.61
AWRA-L v5	-1.00	-0.93	-0.58	0.45	**3.95**	41.53	1199.26
AWRA-L v6	-0.99	-0.77	**-0.20**	1.42	6.77	58.17	2150.01
Benchmark			**-0.20**		**3.95**		
Model	**Annual time series correlation**						
CABLE	-0.76	-0.29	0.18	0.44	0.60	0.78	0.97
WaterDyn	-0.61	-0.10	0.24	0.47	0.66	0.84	0.99
AWRA-L v5	-0.99	-0.35	0.02	0.20	0.42	0.68	0.84
AWRA-L v6	-0.89	-0.14	0.12	**0.30**	0.50	0.71	0.99
Benchmark				**0.30**			

Example annual time series for two sites are presented for the 3 models in Figure 24. This plot gives an indication of the variability between models and data.

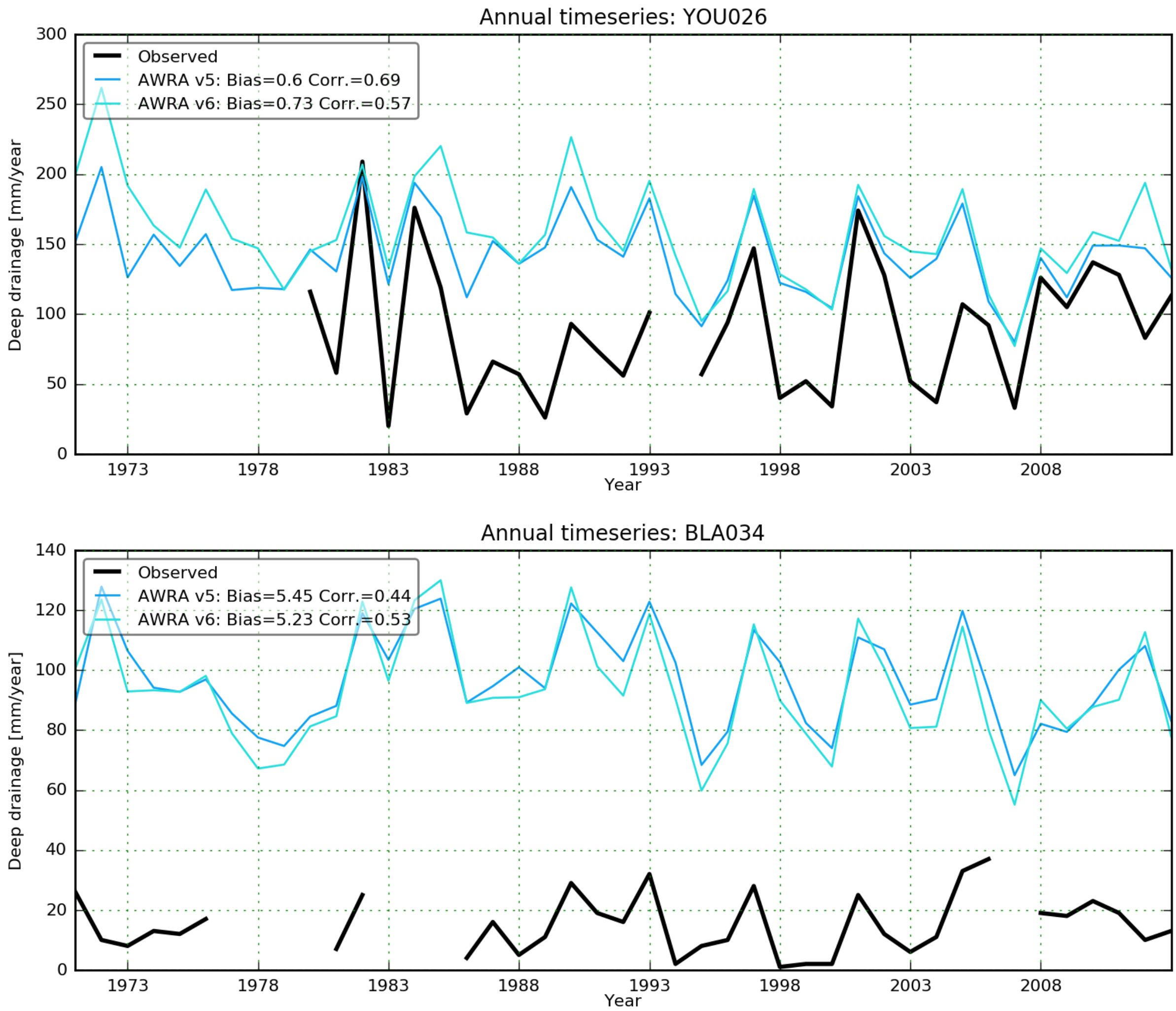

Figure 24. Example annual deep drainage time-series for two sites.

Figure 25 shows the relative bias value of the AWRA-L model compared to the Long Term Average data Australia wide; with the AWRA-L v6 shallow layer saturated soil conductivity underlain. Figure 26 shows the AWRA-L performance according to the annual data spatially – with the AWRA-L map of the maximum shallow layer soil storage and saturated conductivity is also underlain. The recharge values tend to be biased positively in areas where the saturated degree of conductivity is high, and negatively biased where saturated conductivity is low.

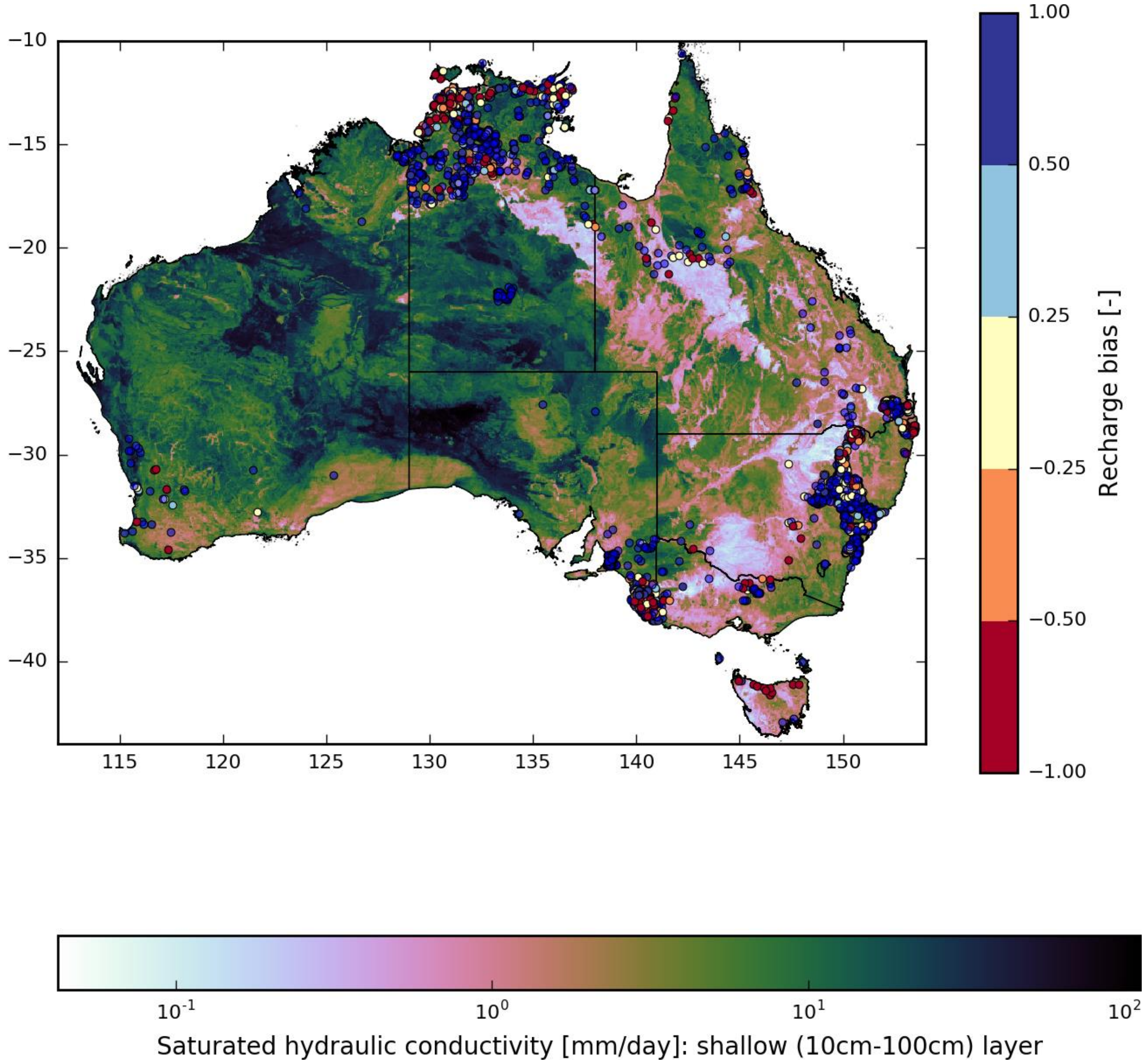

Figure 25. AWRA-L v6 relative bias of deep drainage compared to Long Term Average estimates over Australia. AWRA-L v6 shallow layer soil saturated hydraulic conductivity (Ksat) is also mapped.

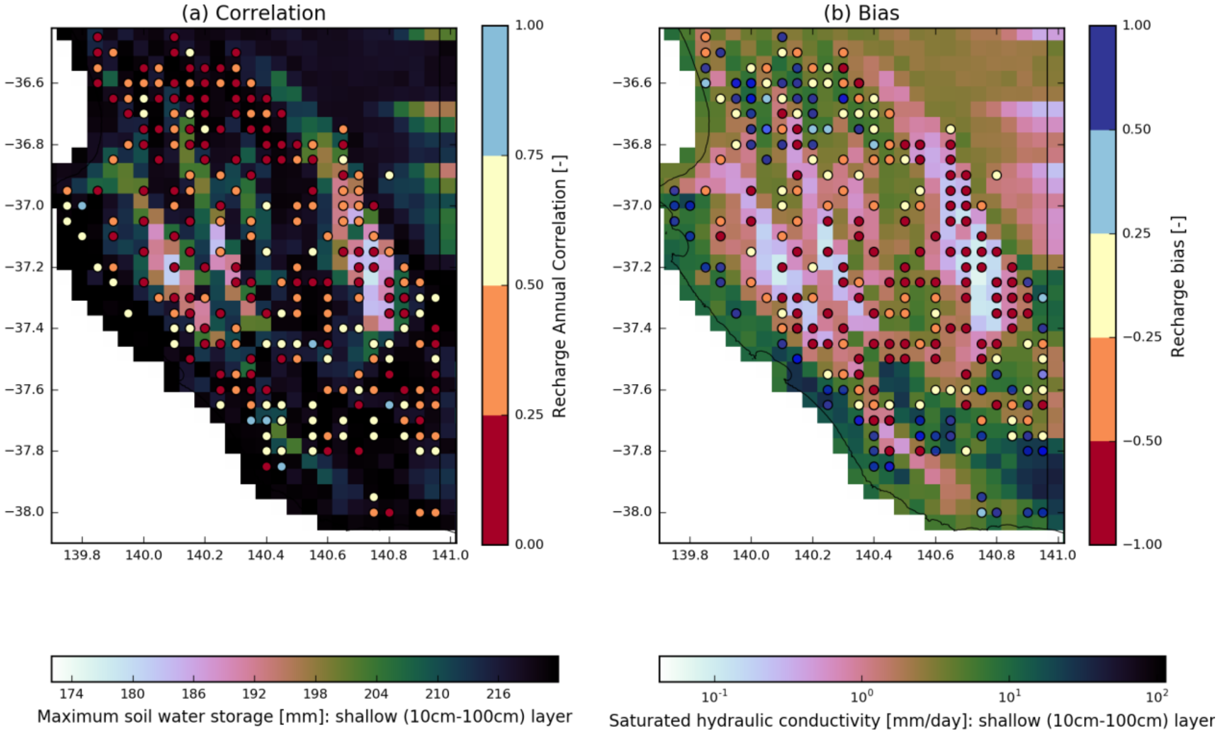

Figure 26. AWRA-L correlation and relative bias of deep drainage compared to annual time series estimates over South Eastern South Australia. AWRA-L v6 shallow layer maximum soil storage (Ssmax) and saturated hydraulic conductivity (Ksat) is also mapped.

These results appear to indicate that the saturated conductivity layer is controlling drainage too strongly; and other factors such as rainfall gradient and variability are not affecting drainage enough. This repeats similar results previously found for AWRA-L v5, although the degree of bias is lower.

It is noted that this timespan used here in evaluation is much smaller than that estimated according to the long term average through chloride mass balance, and also that land use changes may mean that the long term averages are not representative for the period compared. The simulation period considered was 1970-2012, consistent with the evaluation against recharge annual time series span. However, the pattern of bias against the long-term recharge dataset is consistent nationally.

5.5 Summary according to benchmark statistics

Table 10 shows a summary of the performance of all the models considered against the all the available observation data sets across the water balance. This table provides a quick overview of the trade-offs in performance between different models across the water balance and the opportunity to compare all models overall performance across the water balance. Based on these results, AWRA-L v6 was deemed to improve on AWRA-L v5 due to the improvements in soil moisture, recharge and runoff performance. Soil moisture and runoff were considered the focus variables as most use of the product is soil moisture and runoff data.

Table 10. Performance according to benchmark validation statistics. Percentile indicates the ranked site value for a given statistic. The red, white and blue colouring indicates the rank of the model according to the statistic. Dark blue indicates highest rank, white middle rank, dark red lowest rank. *Note: satellite based estimates (ASCAT, AMSRE, CMRSET, SLST) do not cover full time range of observed data.

Data and statistic	Percentile	Target	Best	CABLE	WaterDyn	v5	v6	GRJ	Sacramento	ASCAT*	AMSRE*	CMRSET*	SLST*
Streamflow													
Daily NSE	0.05	0	-0.06			-0.3	-0.06	-0.76	-1.73				
	0.5	0.5	0.56			0.45	0.49	0.56	0.55				
Monthy NSE	0.05	0	-0.19	-0.38	-1.76	-0.23	-0.19	-0.24	-0.24				
	0.5	0.5	0.74	0.31	0.60	0.70	0.67	0.73	0.74				
Monthly Relative Bias	0.05	>-0.5		-0.56	-0.46	-0.46	-0.43	-0.60	-0.54				
	0.25	>-0.3		-0.27	-0.16	-0.18	-0.20	-0.20	-0.19				
	0.75	<0.3		0.32	0.60	0.29	0.28	0.30	0.32				
	0.95	<1.5		1.44	2.38	1.28	1.40	1.34	1.41				
Soil moisture													
SASMAS 0-90cm Daily correlation	0.5	0.75	0.76			0.73	0.76			0.52	0.39		
SASMAS 0-90cm Monthly correlation	0.5	0.75	0.74	0.69	0.49	0.72	0.74			0.64	0.52		
OzNet 0-90cm Daily correlation	0.5	0.75	0.77			0.74	0.77			0.67	0.63		
OzNet 0-90cm Monthly correlation	0.5	0.75	0.79	0.79	0.77	0.76	0.79			0.78	0.72		
Actual Evapotranspiration													
Monthly correlation	0.05	0.5	0.70	0.32	0.61	0.49	0.41					0.70	0.41
	0.5	0.8	0.86	0.85	0.86	0.82	0.85					0.85	0.77
Daily correlation	0.05		0.37			0.37	0.19					0.28	0.21
	0.5		0.69			0.66	0.67					0.69	0.67
Recharge													
Long Term Average Relative bias	0.25	<0		48	60	-58	-20						
	0.75	>0		1150	1398	395	677						
Annual correlation	0.5	0.5	0.47	0.44	0.47	0.20	0.30						

6 Evaluation of AWRA-L for reporting purposes

Gridded outputs and select catchment based time series from AWRA-L v5 and v6 are presented here to give an understanding of the AWRA-L model spatial and temporal dynamics and the extent of changes presented by the updated version.

National annual spatial plots: Annual totals and deciles for the years covering 2008-2017 are plotted in Figure 27 and Figure 28. These plots show year to year variability, in particular the ending of the Millennium drought. These plots show the spatial variability expected across Australia according to climate and catchment conditions. It is noted there is an area in arid Western Australia where rainfall is sparsely monitored, leaving a hole in the interpolated rainfall in the region. This further translates into the water balance components having no flows in this area.

Catchment Time series: Ten catchments were selected for evaluation of states/flux time-series as shown in Figure 29; with site plots shown in Figure 30 to Figure 39. Key features of these sites are presented in Table 11.Time series of the following variables are plotted for each of these catchments: potential and actual evapotranspiration, runoff, deep drainage to the groundwater store, top 0-10cm soil moisture, shallow 10-100cm soil moisture, deep 100-600cm soil moisture. Observed streamflow, along with satellite based ET and soil moisture is also plotted for comparison purposes. These plots give an indication of the seasonal and inter-annual variability present at each of these locations for the key water variables output by AWRA-L.

Key observations from these site plots include:

- the deep storage and deep drainage shows drawdown over the Millennium drought period – as expected in this area.
- Upper soil moisture (0-10cm) is low compared to satellite data for both AWRA v5 and v6. AWRA v6 is lower than v5.
- Lower soil moisture (10-100cm) is similar for v5 and v6.
- Streamflow performance varies according to location, with some sites showing good performance (eg. Site 116013), and others overestimating (eg. Site 226222) or underestimating (e.g. site 410048) variance.
- Deep drainage and deep soil moisture increases for most sites for v6 compared to v5.
- AWRA v6 PET is higher than v5.

Overall the spatial plots and time series give confidence in the use of AWRA-L for water resources assessment; as they broadly follow the expected catchment responses and spatial and temporal trends expected across Australia. The comparisons between the outputs of AWRA v5 and AWRAv6 emphasise that Landscape Water Balance output users will need to be aware of the model updates and in some cases review any relationships they have created between their models and AWRA data.

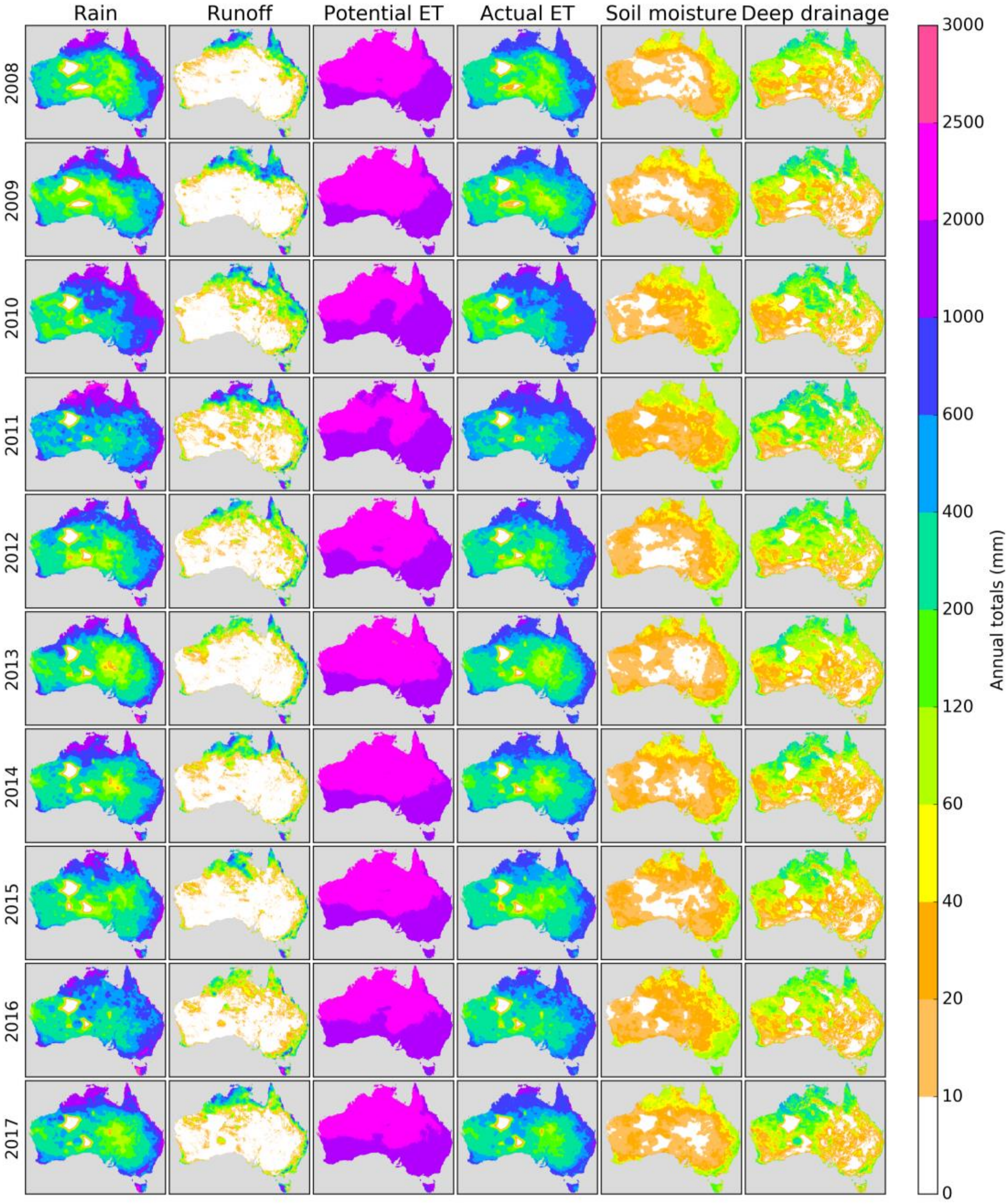

Figure 27. AWRA-L v6 mean annual rainfall, runoff, PET, AET, soil moisture and deep drainage 2008-2017. Units=mm.

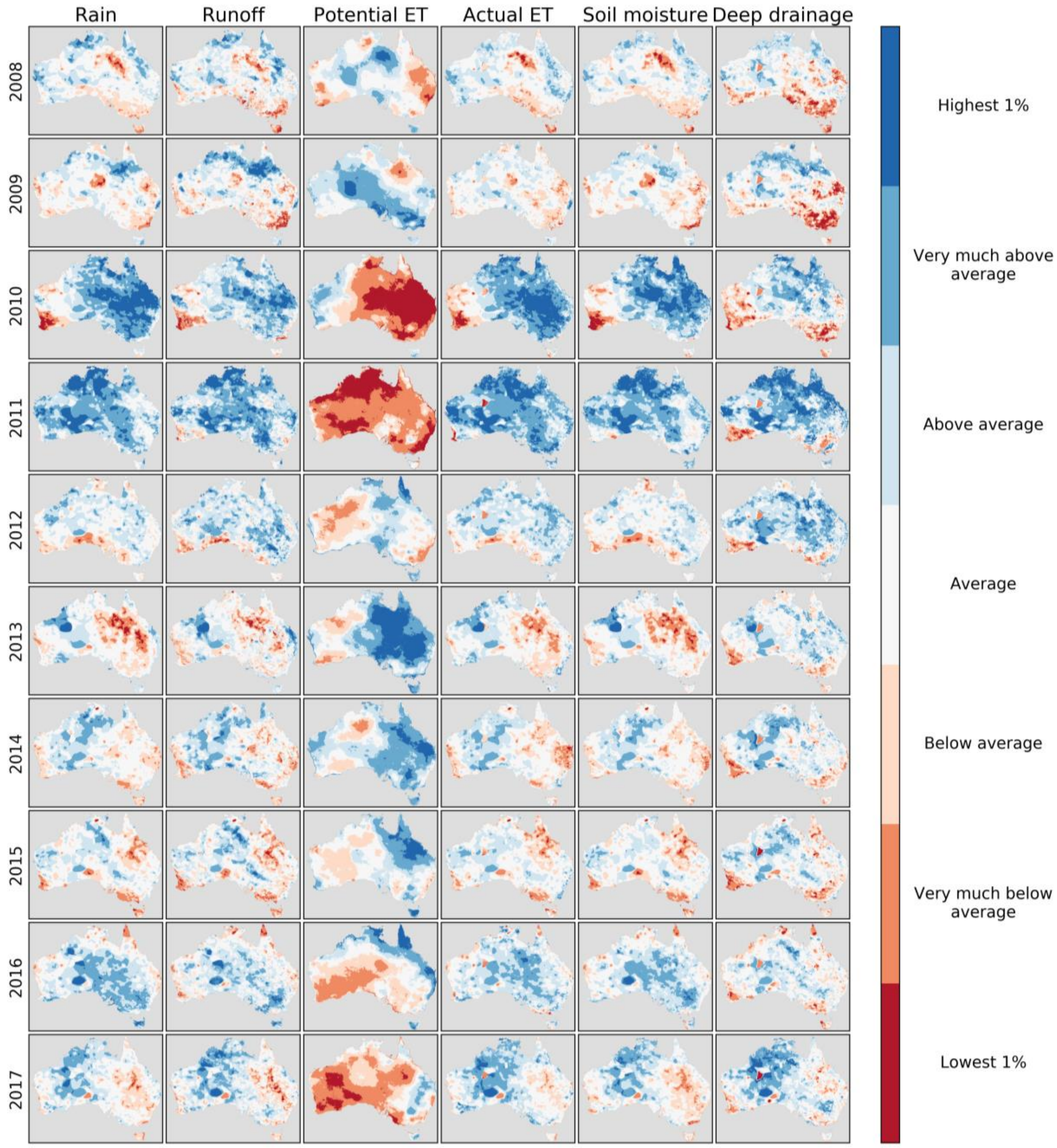

Figure 28. Annual rain, runoff, ET, soil moisture and deep drainage deciles for 2008-2017

Table 11. Selected catchments for detailed evaluation

ID	Name	River	State	Lat. (°)	Lon. (°)	Area (km²)	Elev. (m)	Slope (%)	Ave. Precip. (mm)	Ave.PET (mm)	Forest (-)
145105	Beaudesert Pump Stn	Albert	QLD	-28.02	153.06	266	326	8	1209	1443	0.6
116013	Archer Ck	Millstream	QLD	-17.65	145.34	315	911	4	1589	1714	0.5
226222	near Noojee (U/S Ada R)	Latrobe	VIC	-37.88	145.89	65	480	8	1352	1103	0.9
403213	Greta South	Fifteen Mile Ck	VIC	-36.62	146.24	231	549	7	1032	1214	0.6
410048	Ladysmith	Kyeamba Ck	NSW	-35.20	147.53	548	321	3	641	1217	0.3
501503	U/S Victor Harbour Stw	Inman	SA	-35.54	138.58	165	168	4	701	1190	0.4
614044	Yarragil Formation	Yarragil Brook	WA	-32.81	116.15	71	288	2	904	1489	0.7
607155	Malimup Track	Dombakup Brook	WA	-34.58	115.97	116	87	1	1129	1288	0.7
814011	Manbulloo Boundary	Dry	NT	-15.08	132.41	4786	204	0	896	2091	0.2
811004	Victoria HWY	East Baines	NT	-15.77	130.03	2443	195	2	833	1988	0.2

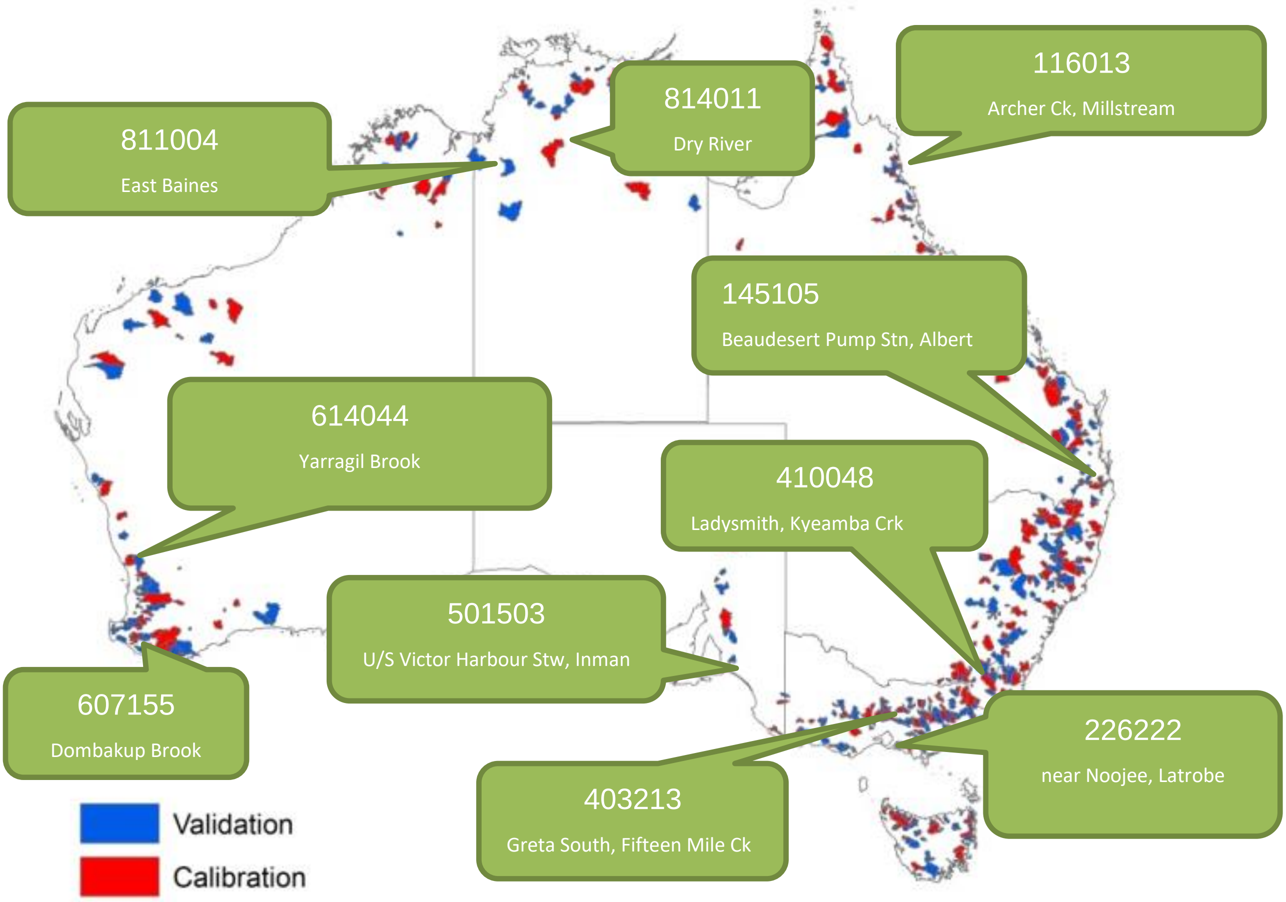

Figure 29. Locations of selected catchments for detailed evaluation

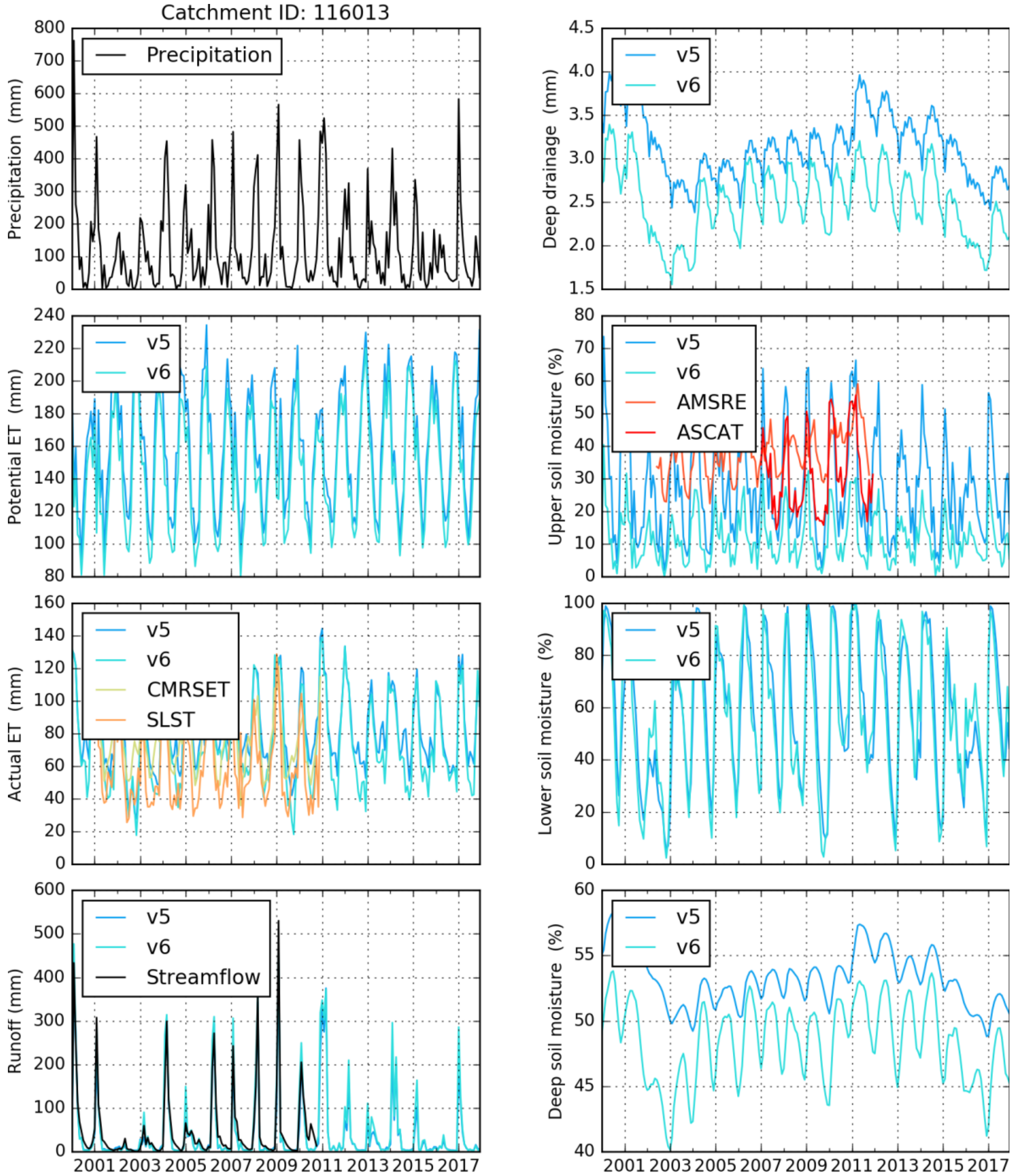

Figure 30. 116013 Archer Creek @ Millstream QLD AWRA-L monthly simulations.

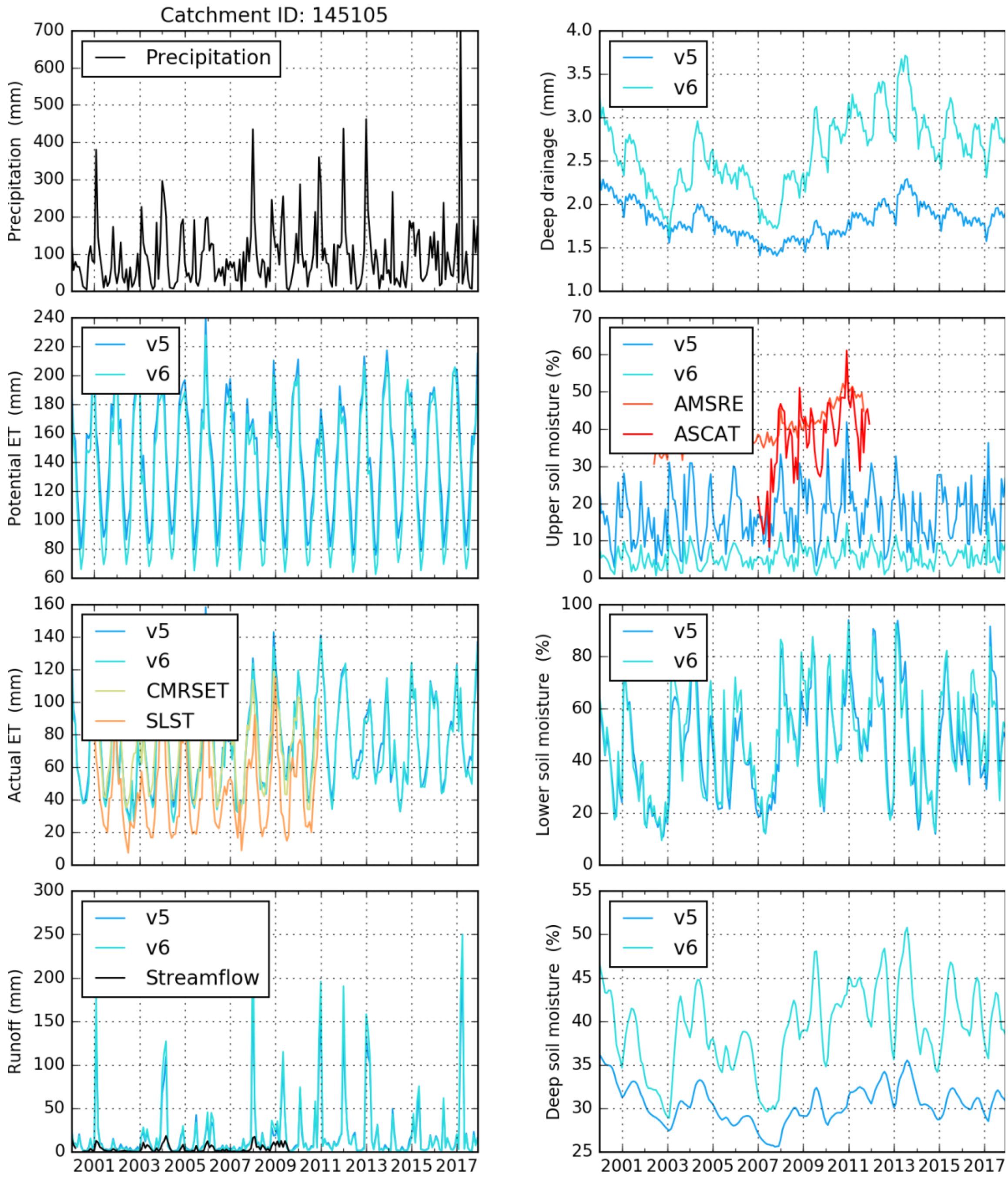

Figure 31. 145105 Beaudesert, Albert River QLD AWRA-L monthly simulations.

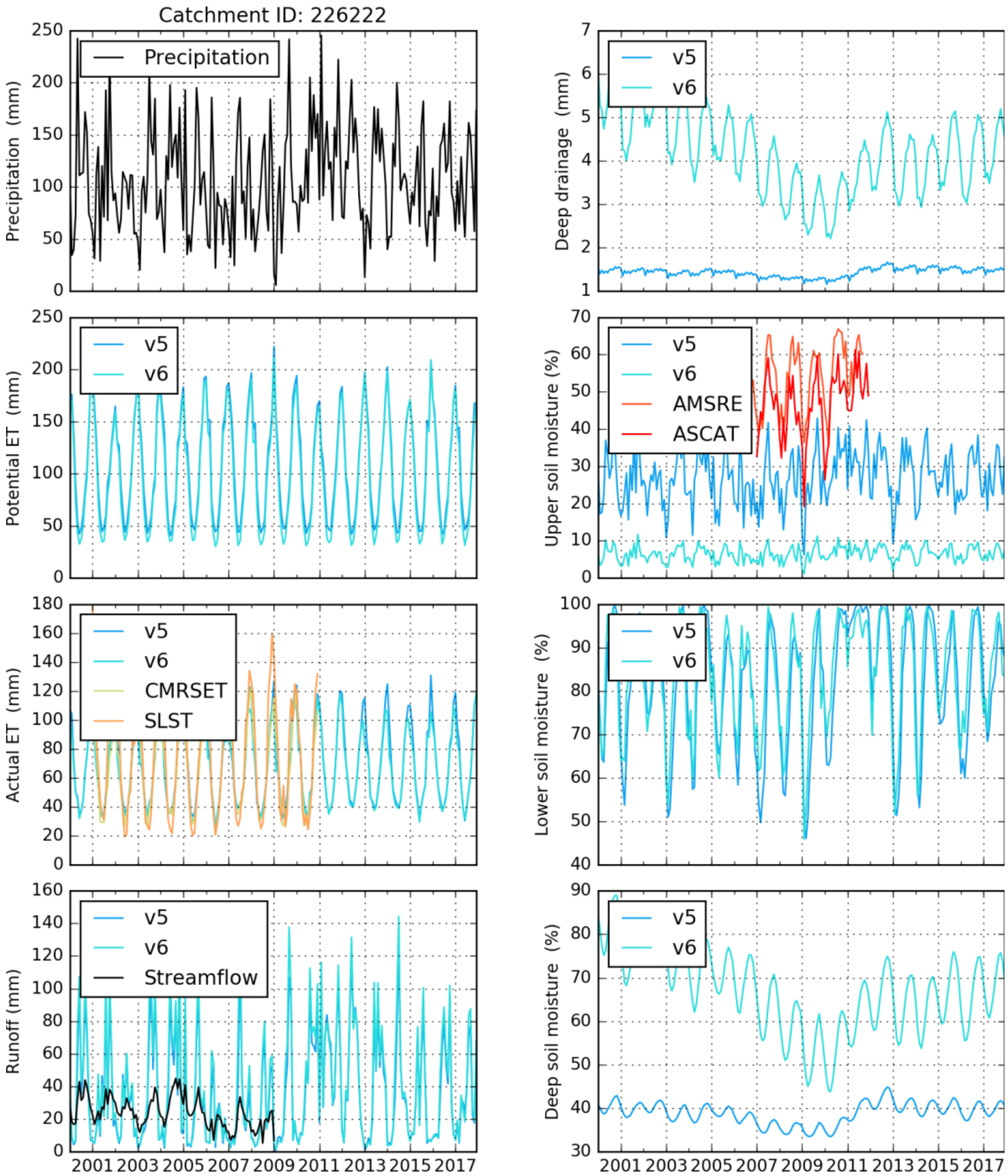

Figure 32. 226222 near Noojee VIC AWRA-L monthly simulations.

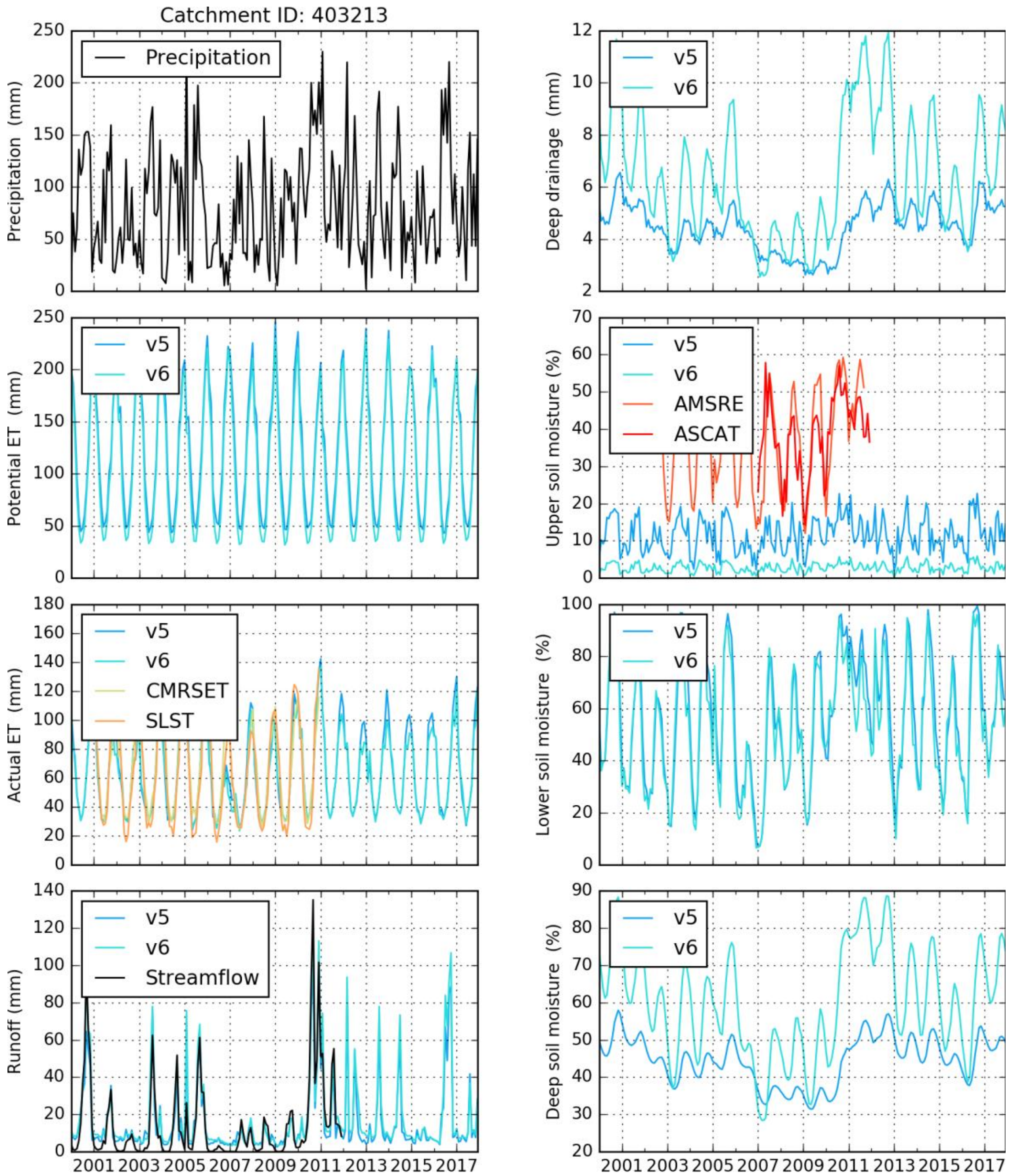

Figure 33. 403213 FIFTEEN MILE CREEK, GRETA SOUTH AWRA-L monthly simulations.

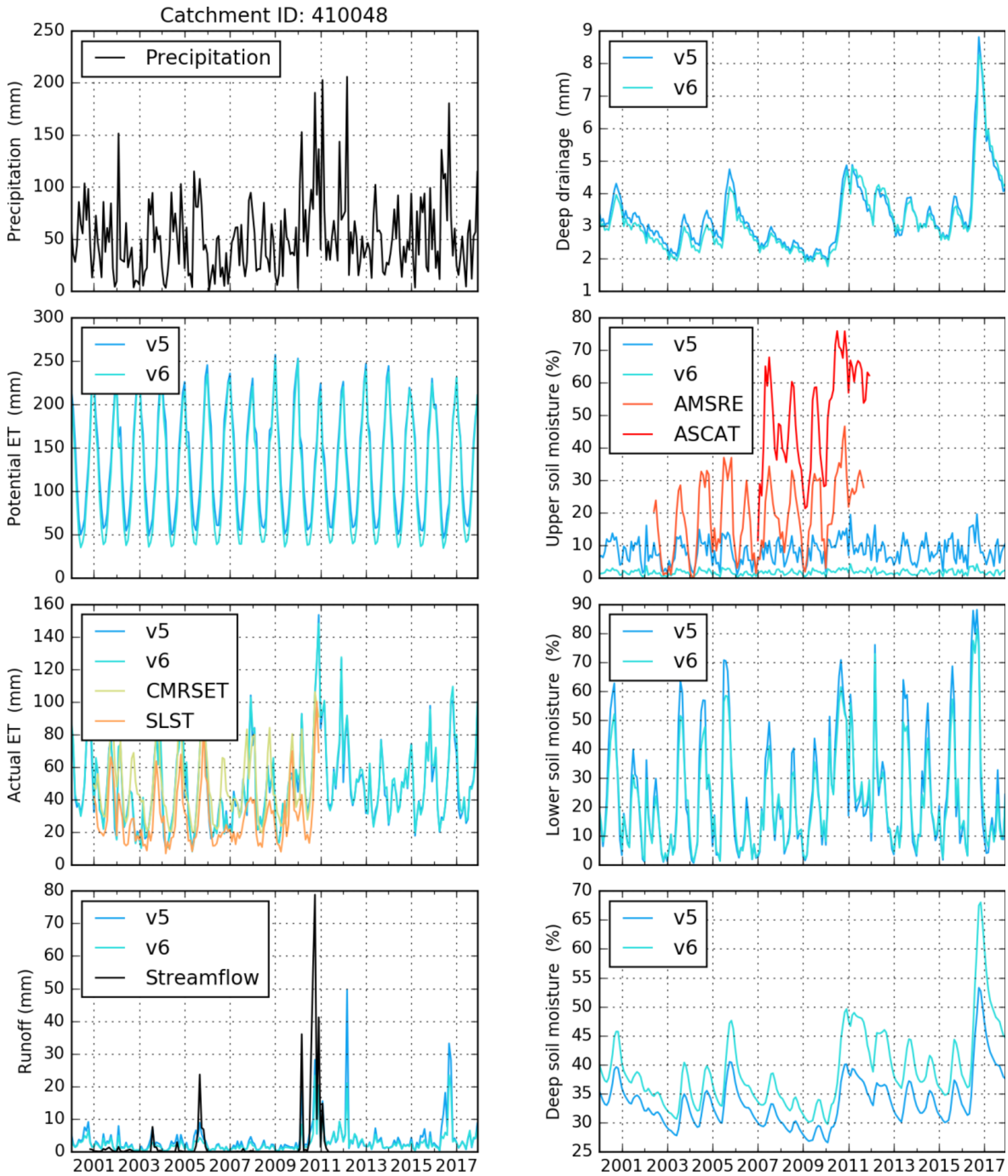

Figure 34. 410048 Ladysmith, Kyeamba Creek NSW AWRA-L Monthly simulations.

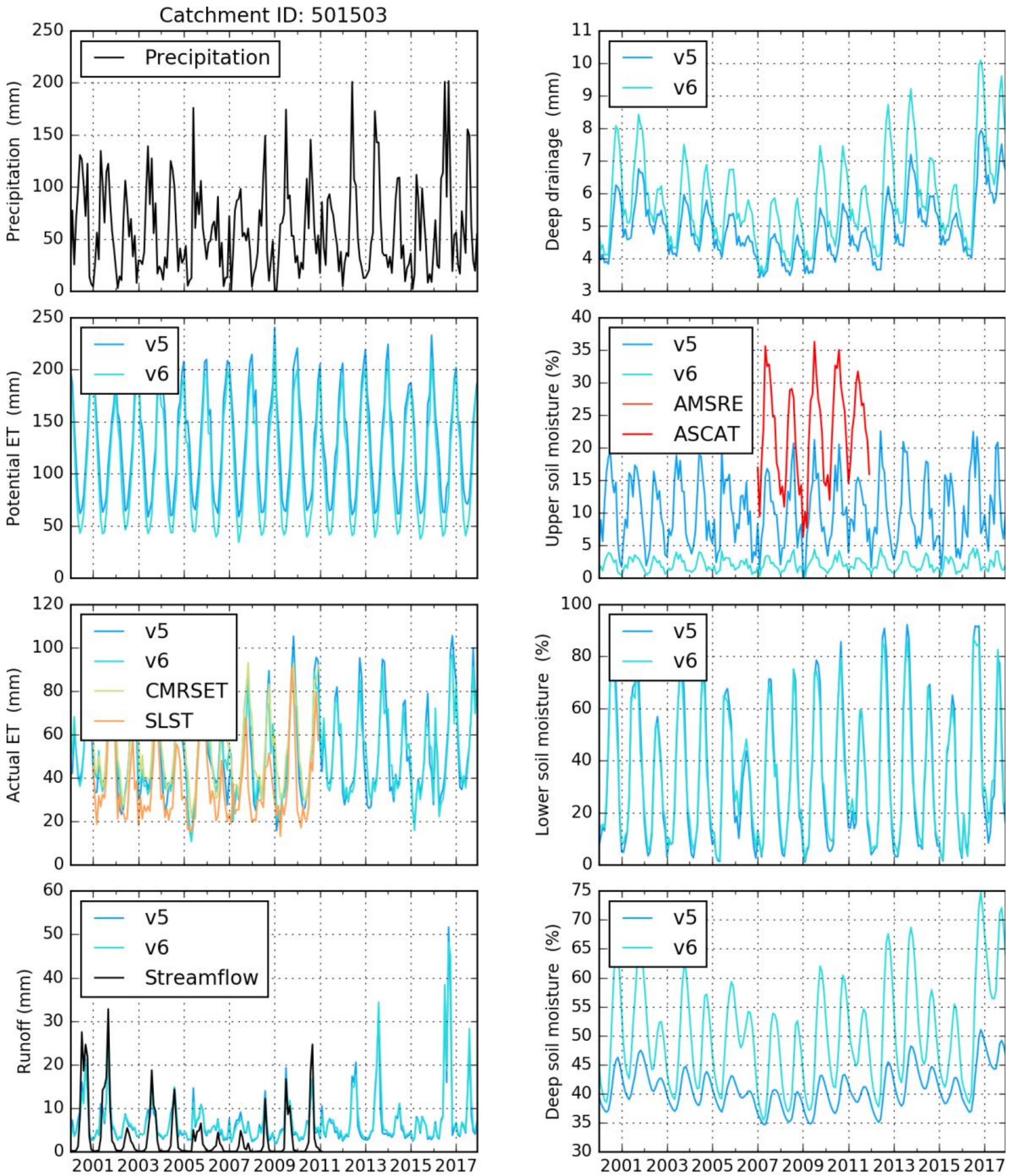

Figure 35. 501503 US Victor harbour, Inman River SA AWRA-L monthly simulations.

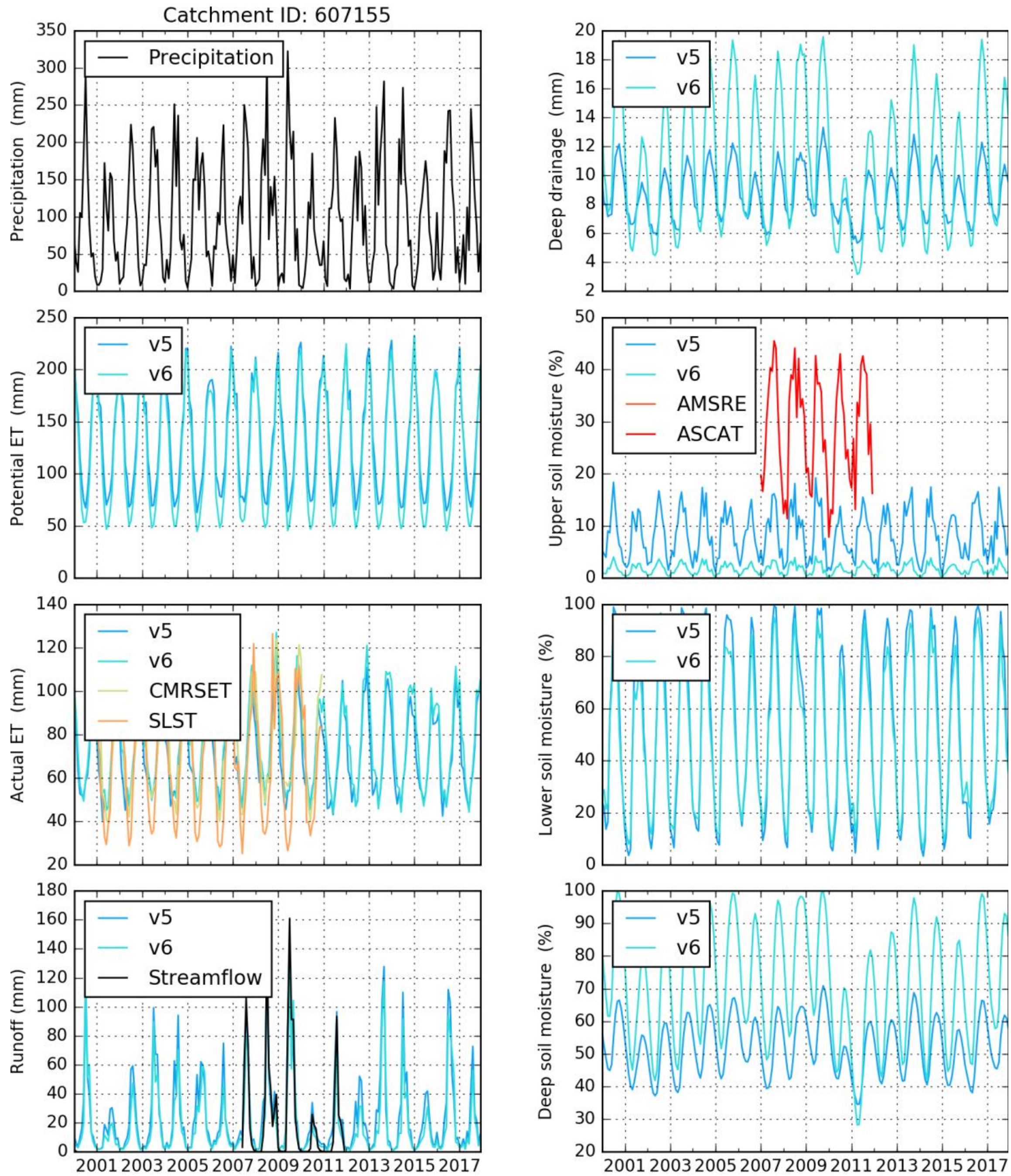

Figure 36. 607155 Malimup Track, Dombakup Brook WA AWRA-L Monthly simulations.

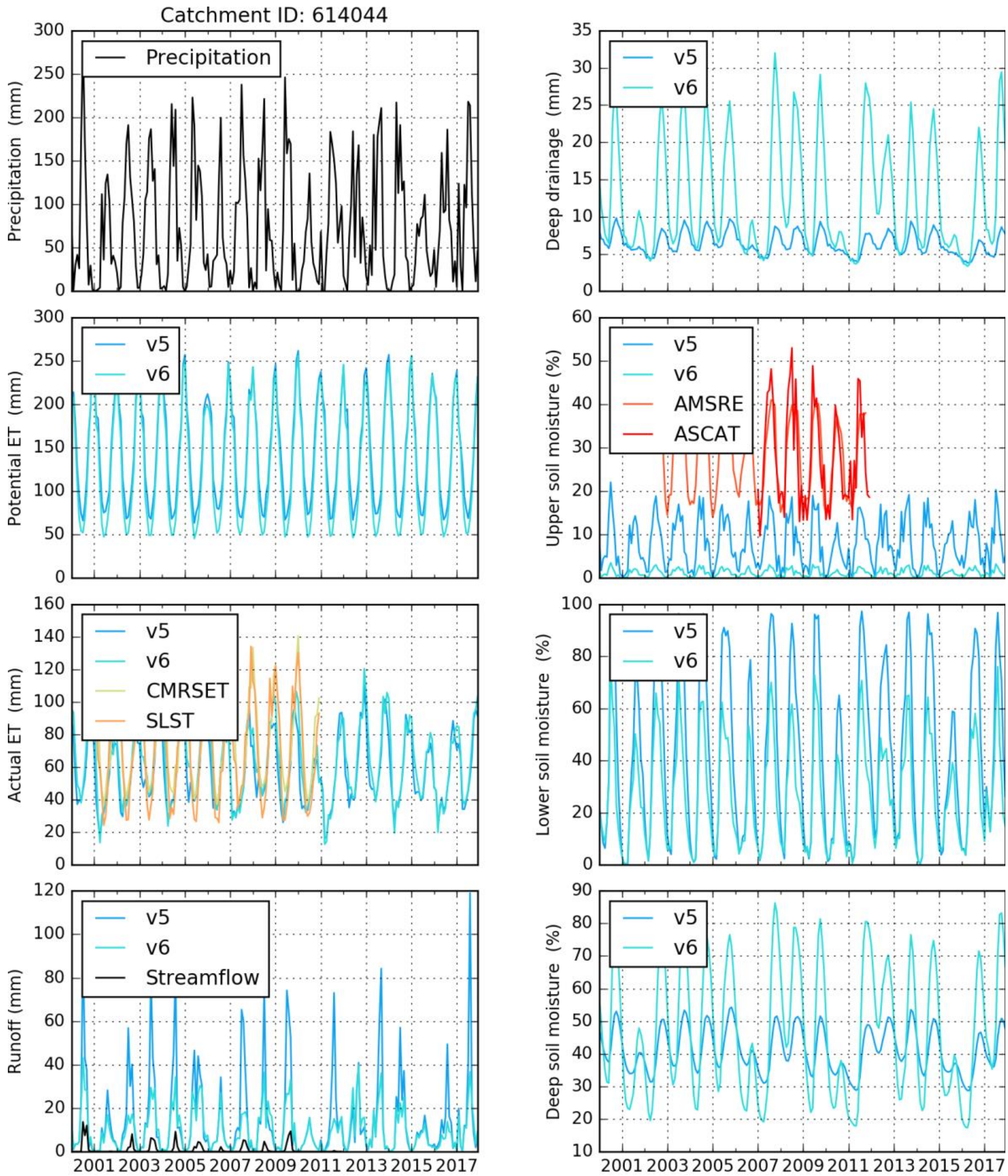

Figure 37. 614044 Yarragil Formation, Yarragil Brook WA AWRA-L monthly simulations.

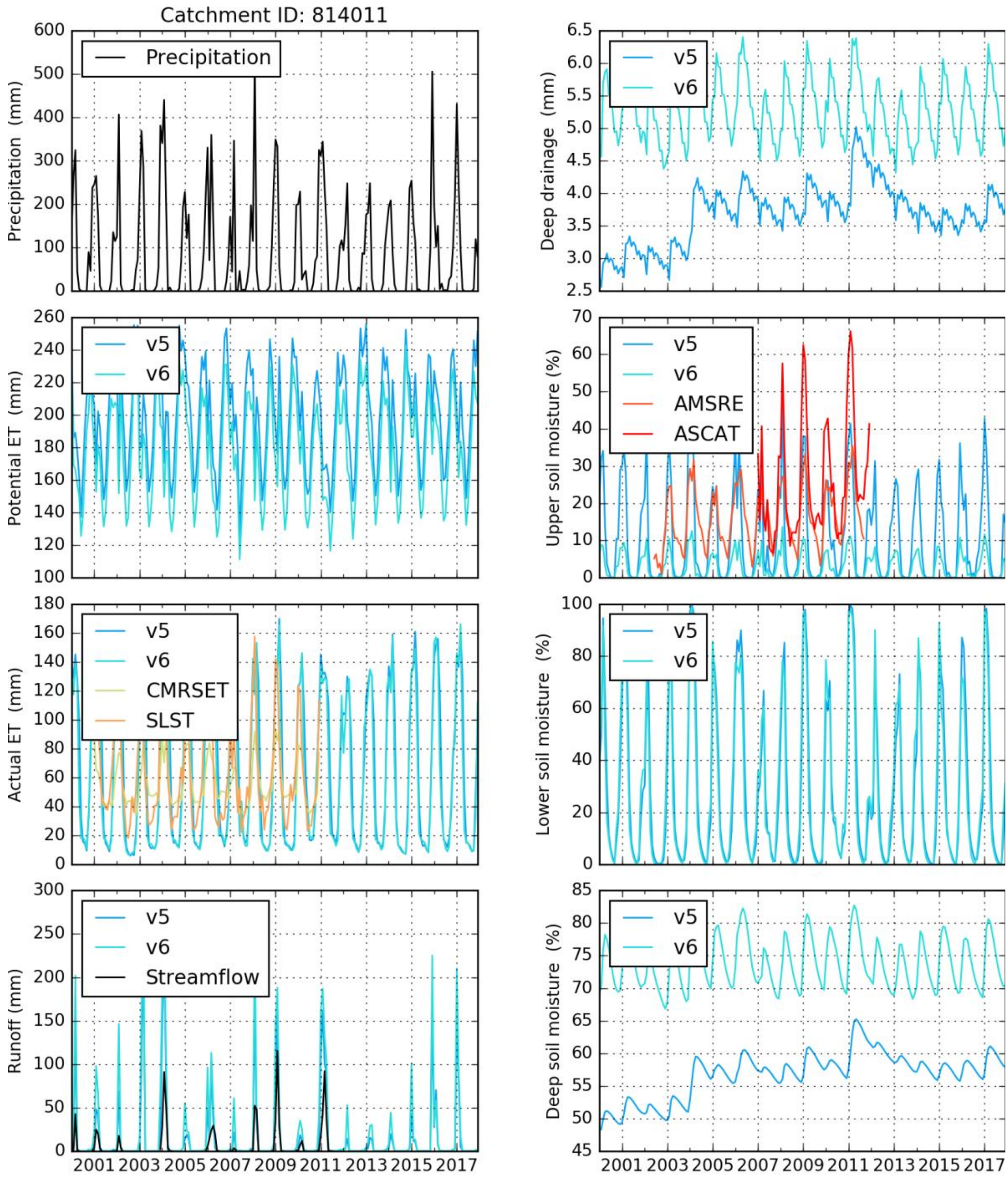

Figure 38. 814011 Dry River NT AWRA-L monthly simulations.

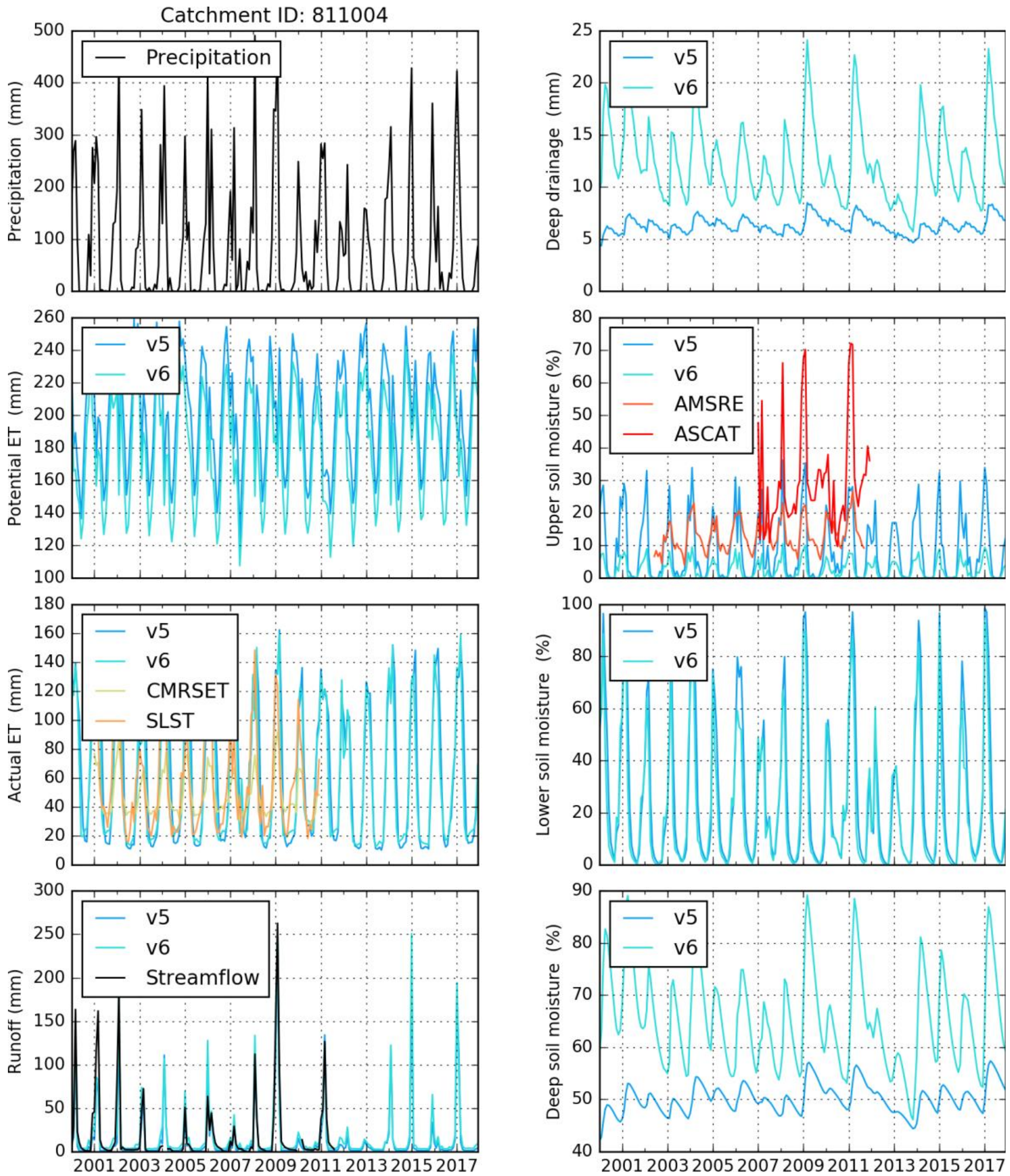

Figure 39. 811004 Victoria HWY East Baines NT AWRA-L monthly simulations.

7 Conclusions

AWRA-L v5 and v6 performance was evaluated using available Australian streamflow, soil moisture, evapotranspiration and groundwater recharge hydrological data sets and compared to two peer national models (WaterDyn and CABLE) as well as two locally calibrated nearest-neighbour regionalised rainfall-runoff models. Performance against key evaluation criteria was undertaken, and results presented in

Table 10 provide a benchmark from which future versions of AWRA-L and other models can be compared. Aspirational targets for model performance are set based on past experience. AWRA-L performs well across the range of variables tested. In certain cases AWRA-L does not reach the aspirational targets set – leaving room for future improvement.

AWRA-L reproduces streamflow relatively well over the 291 catchments reserved for validation. It performs particularly well considering it is approaching the performance of locally calibrated-nearest neighbour regionalised rainfall runoff models (and is superior in terms of bias). AWRA-L v6 improves at the daily timescale according to NSE over AWRA-L v5. AWRA-L v6 performance according to daily Nash-Sutcliffe Efficiency at the 5% / 50% is -0.1 / 0.49 in validation, with aspirational performance set at 0.0 / 0.5 respectively.

AWRA-L and CABLE perform similarly for root-zone (profile 0-90cm) soil moisture, with WaterDyn worse. Current AWRA-L performance according to daily and monthly correlation at the 50% is 0.74-0.79 (for the Murrumbidgee and Hunter sites), with aspirational performance set at 0.75. AWRA-L v6 improves over AWRA-L v5.

For actual ET, CABLE and WaterDyn are better overall than AWRA-L model, although AWRA-L v6 median monthly correlation is equivalent to CABLE. AWRA-L v6 performance according to monthly correlation to DINGO flux tower data at 5% / 50% is 0.41 / 0.85 with aspirational performance set at 0.5 / 0.8 respectively. AWRA-L v6 is an improvement over v5 for median correlation.

AWRA-L appears to not match the spatial patterns of the national recharge dataset, due to drainage currently being overly dependent on saturated conductivity, and not enough on rainfall variability. However, it is noted that there is high uncertainty in this evaluation data. Current AWRA-L v6 performance according to relative bias against the national long term average recharge dataset at 25% / 75% is -20% / 677% with aspirational performance set at being less than zero / greater than zero respectively. Secondly AWRA-L v6 has a median annual correlation against the South Australian annual time-series dataset of 0.3, below the aspirational target of 0.5. AWRA-L v6 improves in terms of correlation performance compared to v5.

Each of the models have differing strengths and weaknesses. Overall, given runoff/streamflow is the dominant hydrological variable used in surface water resource assessment, and that AWRA-L performs well for root zone soil moisture (a key agricultural variable), AWRA-L v6 is considered most fit for purpose water balance estimation purposes.

Various maps and time series are presented to give an understanding of the spatial and temporal dynamics of the model and key output variables according to the AWRA-L v5 and v6. Overall from the spatial plots and the time series give confidence in the use of AWRA-L for water resources assessment; as they broadly follow the expected catchment responses and spatial and temporal trends expected across Australia.

It is noted that there are various improvements to AWRA-L underway including:

- Improving spatial resolution of AWRA-L model, from the current ~5km grid scale down to ~1km
- Implementation of grid flow routing
- Increasing the number of Hydrological Response Units from two (shallow and deep rooted vegetation) to five to include:
  - impervious areas (e.g. Urban)
  - irrigated areas, and
  - permanently inundated areas/lakes.
- Trial of use of mapped groundwater elevation surfaces in calibration to better constrain the overall water balance, and produce better recharge estimates
- Regional calibration of the model: towards better matching local variability
- Assimilation of satellite observations (e.g. surface soil moisture and vegetation) and ground based observations (e.g. streamflow).

Once these additions have been parameterised and shown to improve the model, they will be incorporated into the operational system. Finally, the AWRA-L model is being released as a community model to enable use and development of the system by a wide range of stakeholders.

References

Bartalis, Z., Wagner, W., Naeimi, V., Hasenauer, S., Scipal, K., Bonekamp, H., Figa, J. and Anderson, C., 2007. Initial soil moisture retrievals from the METOP-A Advanced Scatterometer (ASCAT). Geophysical Research Letters, 34(20): L20401.

Beringer, J., Hutley, L. B., McHugh, I., Arndt, S. K., Campbell, D., Cleugh, H. A., Cleverly, J., Resco de Dios, V., Eamus, D., Evans, B., Ewenz, C., Grace, P., Griebel, A., Haverd, V., Hinko-Najera, N., Huete, A., Isaac, P., Kanniah, K., Leuning, R., Liddell, M. J., Macfarlane, C., Meyer, W., Moore, C., Pendall, E., Phillips, A., Phillips, R. L., Prober, S. M., Restrepo-Coupe, N., Rutledge, S., Schroder, I., Silberstein, R., Southall, P., Yee, M. S., Tapper, N. J., van Gorsel, E., Vote, C., Walker, J., and Wardlaw, T, 2016. An introduction to the Australian and New Zealand flux tower network – OzFlux, Biogeosciences, 13, 5895-5916.

Beringer, J., McHugh, I., Hutley, L.B., Isaac, P. and Kljun, N., 2016. Dynamic INtegrated Gap-filling and partitioning for OzFlux (DINGO). Biogeosciences Discuss., 2016: 1-36.

Burnash, R.J., 1995. The NWS river forecast system–catchment modelling. In: V.P. Singh (Editor), Computer Models of Watershed Hydrology. Water Resources Publications, Littleton, Colorado, pp. 311-366.

Crosbie, R., Jolly, I.D., Leaney, F.W., Petheram, C. and Wohling, D., 2010a. Review of Australian groundwater recharge studies, CSIRO: Water for a Healthy Country National Research Flagship.

Crosbie, R.S., Jolly, I.D., Leaney, F.W. and Petheram, C., 2010b. Can the dataset of field based recharge estimates in Australia be used to predict recharge in data-poor areas? Hydrol. Earth Syst. Sci., 14(10): 2023-2038.

Donohue, R.J., Roderick, M.L. and McVicar, T.R., 2008. Deriving consistent long-term vegetation information from AVHRR reflectance data using a cover-triangle-based framework. Remote Sensing of Environment, 112(6): 2938-2949.

Draper, C.S., Walker, J.P., Steinle, P.J., de Jeu, R.A.M. and Holmes, T.R.H., 2009. An evaluation of AMSR-E derived soil moisture over Australia. Remote Sensing of Environment, 113(4): 703-710.

Eamus, D., Cleverly, J., Boulain, N., Grant, N., Faux, R. and Villalobos-Vega, R., 2013. Carbon and water fluxes in an arid-zone Acacia savanna woodland: An analyses of seasonal patterns and responses to rainfall events. Agricultural and Forest Meteorology, 182–183: 225-238.

Frost, A.J., Ramchurn, A. and Hafeez, M., 2015. Evaluation of AWRA-L for national drought and soil moisture monitoring, 36th Hydrology and Water Resources Symposium: The art and science of water. Barton, ACT: Engineers Australia, Hobart, pp. 1496-1504.

Frost, A. J., Ramchurn, A., Hafeez, M. 2016a. Evaluation of the Bureau's Operational AWRA-L Model. Bureau of Meteorology Technical Report. See http://www.bom.gov.au/water/landscape/assets/static/publications/Frost_Evaluation_Report.pdf

Frost, A. J., Ramchurn, A., and Smith, A. 2016b. The Bureau's Operational AWRA Landscape (AWRA-L) Model. Bureau of Meteorology Technical Report. See http://www.bom.gov.au/water/landscape/assets/static/publications/Frost__Model_Description_Report.pdf

Frost, A. J., Ramchurn, A., and Smith, A., 2018. The Australian Landscape Water Balance model (AWRA-L v6). Technical Description of the Australian Water Resources Assessment Landscape model version 6. Bureau of Meteorology Technical Report.

Grant, I., Jones, D., Wang, W., Fawcett, R., and Barratt, D., 2008. Meteorological and remotely sensed datasets for hydrological modelling: a contribution to the Australian Water Availability Project. Proceedings of the Catchment-scale Hydrological Modelling & Data Assimilation (CAHMDA-3) International Workshop on Hydrological Prediction: Modelling, Observation and Data Assimilation, January 9 2008–January 11 2008.

Gupta, H. V, Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. J. Hydrol. 377, 80–91. https://doi.org/https://doi.org/10.1016/j.jhydrol.2009.08.003

Guerschman, J.P., Van Dijk, A.I.J.M., Mattersdorf, G., Beringer, J., Hutley, L.B., Leuning, R., Pipunic, R.C. and Sherman, B.S., 2009. Scaling of potential evapotranspiration with MODIS data reproduces flux observations and catchment water balance observations across Australia. Journal of Hydrology, 369(1-2): 107-119.

Haverd, V. and Cuntz, M., 2010. Soil–Litter–Iso: A one-dimensional model for coupled transport of heat, water and stable isotopes in soil with a litter layer and root extraction. Journal of Hydrology, 388(3–4): 438-455.

Haverd, V., Raupach, M.R., Briggs, P.R., Canadell, J.G., Isaac, P., Pickett-Heaps, C., Roxburgh, S.H., van Gorsel, E., Viscarra Rossel, R.A. and Wang, Z., 2013. Multiple observation types reduce uncertainty in Australia's terrestrial carbon and water cycles. Biogeosciences, 10(3): 2011-2040.

Kowalczyk, E.A., Wang, Y.P., Law, R.M., Davies, H.L., McGregor, J.L. and Abramowitz, G., 2006. The CSIRO Atmosphere Biosphere Land Exchange model for use in climate models and as an offline model, CSIRO.

Lacava, T., Brocca, L., Faruolo, M., Matgen, P., Moramarco, T., Pergola, N. and Tramutoli, V., 2012. A multi-sensor (SMOS, AMSR-E and ASCAT) satellite-based soil moisture products inter-comparison, 2012 IEEE International Geoscience and Remote Sensing Symposium, pp. 1135-1138.

Leroux, D., Kerr, Y.H., Albitar, A., Bindlish, R., Jackson, T.J., Berthelot, B. and Portet, G., 2013. Comparison between SMOS, VUA, ASCAT, and ECMWF soil moisture products over four watersheds in U.S. IEEE Transactions on Geoscience and Remote Sensing: 1-20.

Owe, M., de Jeu, R. and Holmes, T., 2008. Multisensor historical climatology of satellite-derived global land surface moisture. Journal of Geophysical Research: Earth Surface, 113(F1): F01002.

Peeters, L.J.M., Doble, R.C., Crosbie, R.S. and van Dijk, A.I.J.M., 2011. Incorporating topography-dependent groundwater storage in AWRA-L improves groundwater flux estimation, 19th International Congress on Modelling and Simulation, Perth, Australia, pp. 4064-4070.

Perrin, C., Michel, C. and AndrÃeassian, V., 2003. Improvement of a parsimonious model for streamflow simulation. Journal of Hydrology, 279(1-4): 275-289.

Potter, N. J., F. H. S. Chiew and A. J. Frost (2010). *An assessment of the severity of recent reductions in rainfall and runoff in the Murray-Darling Basin*, Journal of Hydrology 381(3-4): 52-64

Raupach, M.R., Briggs, P.R., Haverd, V., King, E.A., Paget, M. and Trudinger, C.M., 2009. Australian Water Availability Project (AWAP) - CSIRO Marine and Atmospheric Research Component: Final Report for Phase 3, CSIRO Marine and Atmospheric Research, Canberra, Australia.

Ramchurn, A. and Frost, A.J., 2013. Potential improvements to the Australian Water Resources Assessment system landscape (AWRA-L) model, 20th International Congress on Modelling and Simulation, Adelaide, Australia, pp. 3008-3014.

Refsgaard, J.C. and Henriksen, H.J., 2004. Modelling guidelines--terminology and guiding principles. Advances in Water Resources, 27(1): 71-82.

Renzullo, L.J., van Dijk, A.I.J.M., Perraud, J.M., Collins, D., Henderson, B., Jin, H., Smith, A.B. and McJannet, D.L., 2014. Continental satellite soil moisture data assimilation improves root-zone moisture analysis for water resources assessment. Journal of Hydrology, 519, Part D(0): 2747-2762.

Rüdiger, C., Hancock, G., Hemakumara, H.M., Jacobs, B., Kalma, J.D., Martinez, C., Thyer, M., Walker, J.P., Wells, T. and Willgoose, G.R., 2007. Goulburn River experimental catchment data set. Water Resources Research, 43(10): W10403.

Rüdiger, C., Western, A.W., Walker, J.P., Smith, A.B., Kalma, J.D. and Willgoose, G.R., 2010. Towards a general equation for frequency domain reflectometers. Journal of Hydrology, 383(3–4): 319-329.

Shi, X., R.S., C. and Vaze, J., 2015. Evaluation of AWRA recharge: Comparison to field estimates of recharge across Australia. , CSIRO.

Simard, M., Pinto, N., Fisher, J.B. and Baccini, A., 2011. Mapping forest canopy height globally with spaceborne lidar. Journal of Geophysical Research: Biogeosciences, 116(G4): G04021.

Smith, A.B., Walker, J.P., Western, A.W., Young, R.I., Ellett, K.M., Pipunic, R.C., Grayson, R.B., Siriwardena, L., Chiew, F.H.S. and Richter, H., 2012. The Murrumbidgee soil moisture monitoring network data set. Water Resources Research, 48(7): W07701.

Van Dijk, A.I.J.M., 2010. The Australian Water Resources Assessment System. Technical Report 3. Landscape Model (version 0.5) Technical Description CSIRO: Water for a Healthy Country National Research Flagship.

Van Niel, T.G., McVicar, T.R., Roderick, M.L., van Dijk, A.I.J.M., Beringer, J., Hutley, L.B. and van Gorsel, E., 2012. Upscaling latent heat flux for thermal remote sensing studies: Comparison of alternative approaches and correction of bias. Journal of Hydrology, 468–469(0): 35-46.

Vaze, J., Viney, N., Stenson, M., Renzullo, L., Van Dijk, A., Dutta, D., Crosbie, R., Lerat, J., Penton, D., Vleeshouwer, J., Peeters, L., Teng, J., Kim, S., Hughes, J., Dawes, W., Zhang, Y., Leighton, B., Perraud, J., Joehnk, K., Yang, A., Wang, B., Frost, A., Elmahdi, A., Smith, A. and Daamen, C., 2013. The Australian Water Resource Assessment System (AWRA), Proceedings of the 20th International Congress on Modelling and Simulation, Adelaide, Australia.

Viney, N., Vaze, J., Crosbie, R., Wang, B., Dawes, W. and Frost, A., 2014. AWRA-L v4.5: technical description of model algorithms and inputs, CSIRO, Australia.

Viney, N., Vaze, J., Crosbie, R., Wang, B., Dawes, W. and Frost, A., 2015. AWRA-L v5.0: technical description of model algorithms and inputs, CSIRO, Australia.

Viney, N.R., Perraud, J., Vaze, J., Chiew, F.H.S., Post, D.A. and Yang, A., 2009. The usefulness of bias constraints in model calibration for regionalisation to ungauged catchments. In: R.S. Anderssen, R.D. Braddock and L.T.H. Newham (Editors), 18th World IMACS / MODSIM Congress. Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in

Simulation, Cairns, Australia, pp. 3421-3427, http://www.mssanz.org.au/modsim09/I7/viney_I7a.pdf.

Wagner, W., Brocca, L., Naeimi, V., Reichle, R., Draper, C., Jeu, R.d., Ryu, D., Su, C.H., Western, A., Calvet, J.C., Kerr, Y.H., Leroux, D.J., Drusch, M., Jackson, T.J., Hahn, S., Dorigo, W. and Paulik, C., 2014. Clarifications on the "Comparison Between SMOS, VUA, ASCAT, and ECMWF Soil Moisture Products Over Four Watersheds in U.S.". IEEE Transactions on Geoscience and Remote Sensing, 52(3): 1901-1906.

Wang, Y.P., Kowalczyk, E., Leuning, R., Abramowitz, G., Raupach, M.R., Pak, B., van Gorsel, E. and Luhar, A., 2011. Diagnosing errors in a land surface model (CABLE) in the time and frequency domains. Journal of Geophysical Research: Biogeosciences, 116(G1): G01034.

Wang, Y.P., Law, R.M. and Pak, B., 2010. A global model of carbon, nitrogen and phosphorus cycles for the terrestrial biosphere. Biogeosciences, 7(7): 2261-2282.

Wilson, K., Goldstein, A., Falge, E., Aubinet, M., Baldocchi, D., Berbigier, P., Bernhofer, C., Ceulemans, R., Dolman, H., Field, C., Grelle, A., Ibrom, A., Law, B.E., Kowalski, A., Meyers, T., Moncrieff, J., Monson, R., Oechel, W., Tenhunen, J., Valentini, R. and Verma, S., 2002. Energy balance closure at FLUXNET sites. Agricultural and Forest Meteorology, 113(1–4): 223-243.

Zhang, Y.Q., Viney, N., Frost, A., Oke, A., Brooks, M., Chen, Y. and Campbell, N., 2013. Collation of streamflow and catchment attribute dataset for 780 unregulated Australian catchments, CSIRO: Water for a Healthy Country National Research Flagship.

Appendices

Appendix A: ET monitoring site details and time-series

Table 12. Flux tower site details (data source noted)

Site Name	Citation	Temporal coverage
Adelaide Riv.	*Jason Beringer (2013) Adelaide River OzFlux tower site OzFlux: Australian and New Zealand Flux Research and Monitoring hdl: 102.100.100/14228*	2007-01 - 2009-05
Alice Springs	*James Cleverly (2011) Alice Springs Mulga OzFlux site OzFlux: Australian and New Zealand Flux Research and Monitoring hdl: 102.100.100/14217*	2010-09 - 2013-12
Calperum	*Calperum Tech (2013) Calperum Chowilla OzFlux tower site OzFlux: Australian and New Zealand Flux Research and Monitoring hdl: 102.100.100/14236*	2010-01 - 2013-12
Cumberland	*Elise Pendall (2015) Cumberland Plain OzFlux Tower Site OzFlux: Australian and New Zealand Flux Research and Monitoring hdl: 102.100.100/25164*	2012-01 - 2013-12
Daintree	*Mike Liddell (2013) Daintree Ozflux tower site OzFlux: Australian and New Zealand Flux Research and Monitoring hdl: 102.100.100/14244*	2011-01- 2013-12
Daly Pasture	*Jason Beringer (2013) Daly Pasture OzFlux tower site OzFlux: Australian and New Zealand Flux Research and Monitoring hdl: 102.100.100/14238*	2007-01- 2010-05
Daly Uncleared	*Jason Beringer (2013) Daly Uncleared OzFlux tower site OzFlux: Australian and New Zealand Flux Research and Monitoring hdl: 102.100.100/14239*	2007-01- 2013-12
Dry River	*Jason Beringer (2013) Dry River OzFlux tower site OzFlux: Australian and New Zealand Flux Research and Monitoring hdl: 102.100.100/14229*	2008-01- 2013-12
Fogg Dam	*Jason Beringer (2013) Fogg Dam OzFlux tower site OzFlux: Australian and New Zealand Flux Research and Monitoring hdl: 102.100.100/14233*	2006-02- 2008-12
Gingin	*Craig Macfarlane (2012) Gingin OzFlux: Australian and New Zealand Flux Research and Monitoring hdl: 102.100.100/14223*	2011-01- 2013-11
GWW	*Craig Macfarlane (2013) Great Western Woodlands OzFlux: Australian and New Zealand Flux Research and Monitoring hdl: 102.100.100/14226*	2013-01- 2013-12
Howard Spr	*Jason Beringer (2013) Howard Springs OzFlux tower site OzFlux: Australian and New Zealand Flux Research and Monitoring hdl: 102.100.100/14234*	2001-01- 2013-12
Nimmo	*Robert Simpson (2012) Nimmo High Plains OzFlux Tower Site OzFlux: Australian and New Zealand Flux Research and Monitoring hdl:*	2007-01- 2013-12
RDMF	*Jason Beringer (2014) Red Dirt Melon Farm OzFlux tower site OzFlux: Australian and New Zealand Flux Research and Monitoring hdl:*	2011-09-2013-07
Riggs Creek	*Jason Beringer (2014) Riggs Creek OzFlux tower site OzFlux: Australian and New Zealand Flux Research and Monitoring hdl: 102.100.100/14246*	2011-01- 2013-12
Robson Ck	*Mike Liddell (2013) Robson Creek Ozflux tower site OzFlux: Australian and New Zealand Flux Research and Monitoring hdl: 102.100.100/14243*	2013-01- 2013-12
Samford	*David Rowlings (2011) Samford Ecological Research Facility OzFlux tower site OzFlux: Australian and New Zealand Flux Research and Monitoring hdl:*	2010-01- 2013-12
Sturt Plains	*Jason Beringer (2013) Sturt Plains OzFlux tower site OzFlux: Australian and New Zealand Flux Research and Monitoring hdl: 102.100.100/14230*	2008-01- 2013-12
Ti Tree East	*James Cleverly (2013) Ti Tree East OzFlux Site OzFlux: Australian and New Zealand Flux Research and Monitoring hdl: 102.100.100/14225*	2012-08- 2013-12
Tumbarumba	*Eva vanGorsel (2013) Tumbarumba OzFlux tower site OzFlux: Australian and New Zealand Flux Research and Monitoring hdl: 102.100.100/14241*	2001-01 - 2013-12
Wallaby Ck	*Jason Beringer (2013) Wallaby Creek OzFlux tower site OzFlux: Australian and New Zealand Flux Research and Monitoring hdl: 102.100.100/14231*	2005-01- 2012-12
Warra	Emma White (2014) Warra OzFlux tower site OzFlux: Australian and New Zealand Flux Research and Monitoring hdl: 102.100.100/16188	2013-03 - 2013-12
Whroo	*Jason Beringer (2013) Whroo OzFlux tower site OzFlux: Australian and New Zealand Flux Research and Monitoring hdl: 102.100.100/14232*	2011-12- 2013-12
Wombat	*Stefan Arndt (2013) Wombat State Forest OzFlux-tower site OzFlux: Australian and New Zealand Flux Research and Monitoring hdl: 102.100.100/14237*	2010-01- 2013-12
Yanco	*Jason Beringer (2013) Yanco JAXA OzFlux tower site OzFlux: Australian and New Zealand Flux Research and Monitoring hdl: 102.100.100/14235*	2012-01- 2013-12

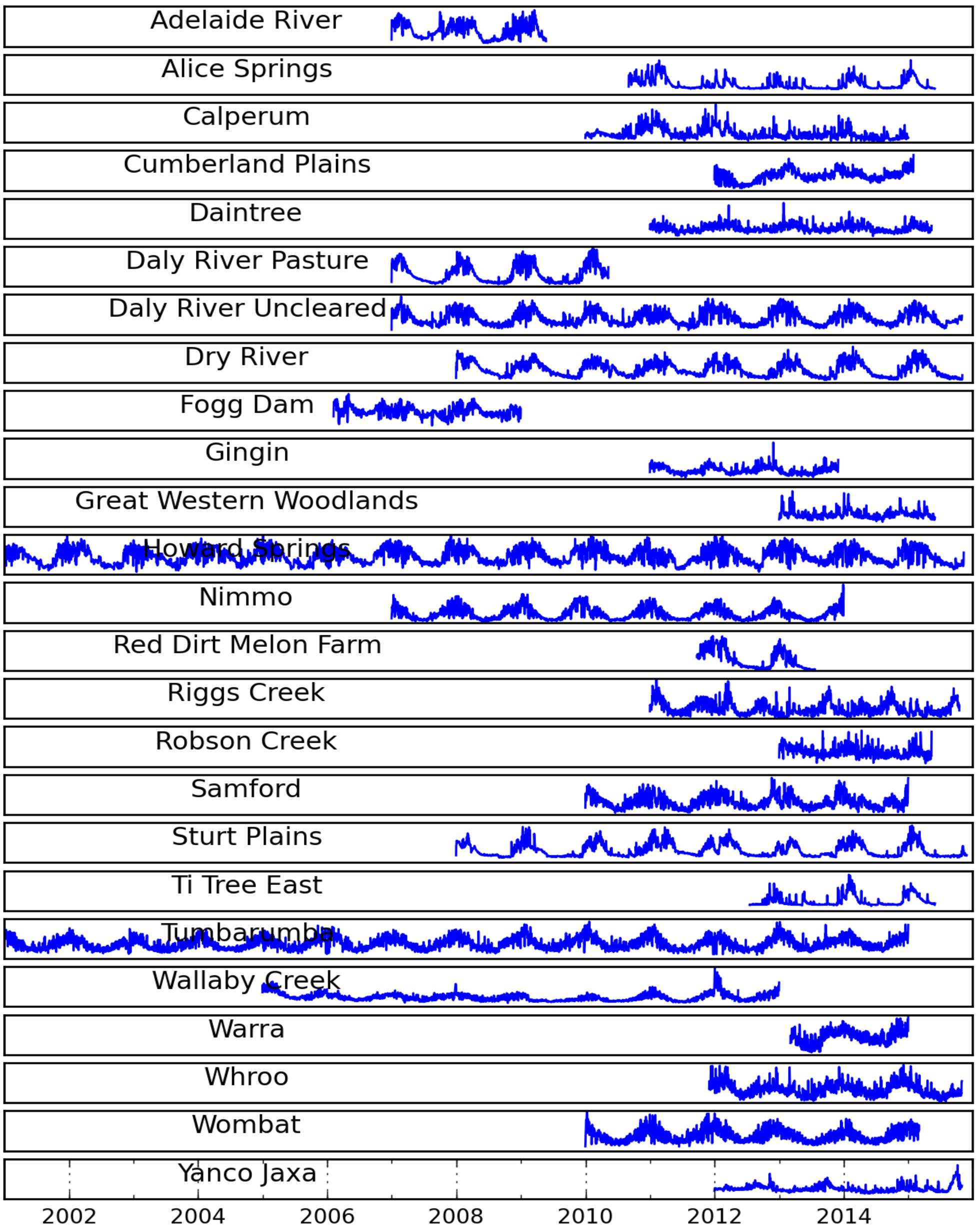

Figure 40. Indicative site time-series of DINGO Evapotranspiration (mm). Axis scale omitted for space purposes.

Appendix B: Soil moisture monitoring site details and time-series

Table 13. OzNet site details

OzNet Site	Start Date	End Date	Daily avail.	Monthly avail.
A1	1/12/2001	31/05/2012	74%	80%
A2	1/12/2001	30/05/2011	37%	41%
A3	1/12/2001	30/11/2012	66%	72%
A4	1/12/2001	31/08/2011	34%	36%
A5	25/11/2001	5/02/2010	58%	64%
K1	15/11/2001	27/09/2012	73%	76%
K10	6/12/2003	31/05/2011	49%	56%
K11	6/11/2003	28/08/2009	46%	47%
K12	5/11/2003	31/05/2011	51%	55%
K13	16/11/2003	31/12/2013	65%	71%
K14	6/11/2003	31/05/2011	56%	59%
K2	16/11/2001	3/09/2010	69%	72%
K3	16/11/2001	24/08/2012	71%	82%
K4	15/11/2001	26/07/2012	80%	84%
K5	14/11/2001	25/06/2012	66%	70%
K6	5/11/2003	16/04/2013	62%	70%
K7	5/11/2003	31/05/2011	59%	61%
K8	5/11/2003	16/04/2013	52%	60%
M1	13/09/2001	1/02/2012	73%	76%
M2	13/09/2001	31/05/2013	79%	84%
M3	15/11/2001	31/05/2013	24%	25%
M4	15/09/2001	31/05/2011	75%	79%
M5	27/09/2001	15/12/2010	49%	61%
M6	27/09/2001	31/05/2011	71%	77%
M7	28/09/2001	1/02/2012	82%	85%
Y1	27/12/2003	31/12/2013	59%	67%
Y10	9/01/2004	31/12/2013	70%	76%
Y11	8/01/2004	31/12/2013	59%	64%
Y12	11/12/2003	31/12/2013	62%	68%
Y13	11/12/2003	31/12/2013	65%	72%
Y2	16/01/2004	31/12/2013	55%	65%
Y3	28/09/2001	17/04/2002	4%	5%
Y4	21/12/2003	23/06/2013	58%	66%
Y5	9/12/2003	28/02/2012	60%	65%
Y6	21/12/2003	20/10/2013	54%	64%
Y7	17/12/2003	31/12/2013	63%	66%
Y8	11/12/2003	31/12/2013	56%	61%
Y9	17/12/2003	25/12/2013	65%	72%

Table 14. SASMAS site details

Site	Start Date	End Date	Daily avail.	Monthly avail.
G1	3/02/2003	16/10/2008	55%	59%
G2	3/02/2003	31/12/2006	34%	38%
G3	1/01/2003	31/12/2006	34%	36%
G4	NA	NA	0%	0%
G5	14/01/2003	6/03/2007	42%	44%
G6	NA	NA	0%	0%
K1	1/01/2003	31/12/2011	89%	94%
K2	1/01/2003	31/12/2011	90%	97%
K3	1/01/2003	31/12/2009	72%	75%
K4	1/01/2003	31/12/2010	74%	76%
K5	1/01/2003	31/12/2011	90%	93%
K6	NA	NA	0%	0%
M1	NA	NA	0%	0%
M2	1/01/2003	11/07/2007	49%	51%
M3	NA	NA	0%	0%
M4	NA	NA	0%	0%
M5	NA	NA	0%	0%
M6	NA	NA	0%	0%
M7	1/01/2003	31/12/2010	69%	72%
S1	4/02/2003	31/12/2010	82%	84%
S2	NA	NA	0%	0%
S3	NA	NA	0%	0%
S4	NA	NA	0%	0%
S5	4/02/2003	31/12/2011	88%	94%
S6	NA	NA	0%	0%
S7	NA	NA	0%	0%

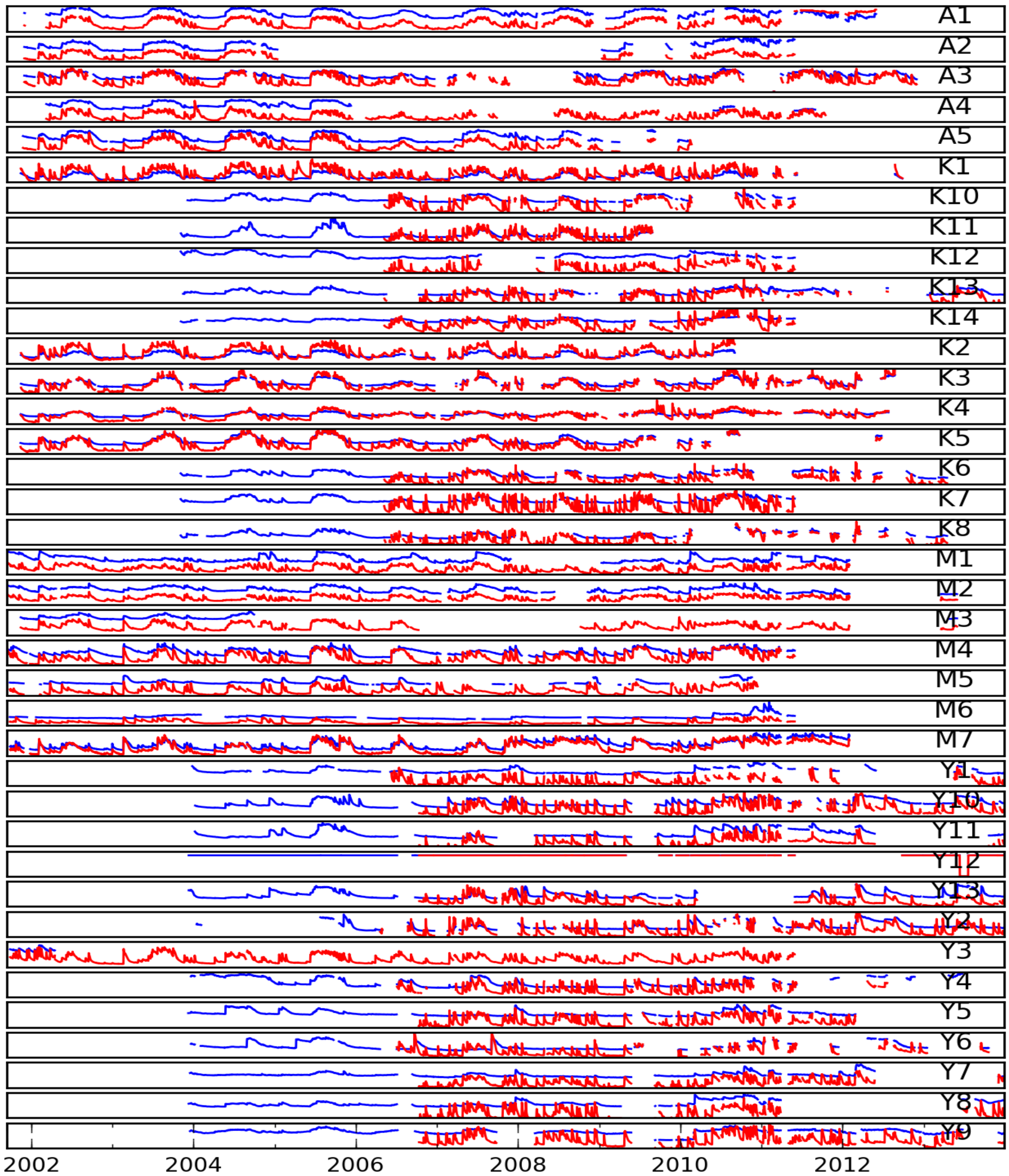

Figure 41. Indicative site daily time-series of OzNet top layer (red: 0-5/8cm) and profile (blue: 0-90cm) volumetric soil moisture. Site numbers are listed far right. Axis scale omitted for space purposes.

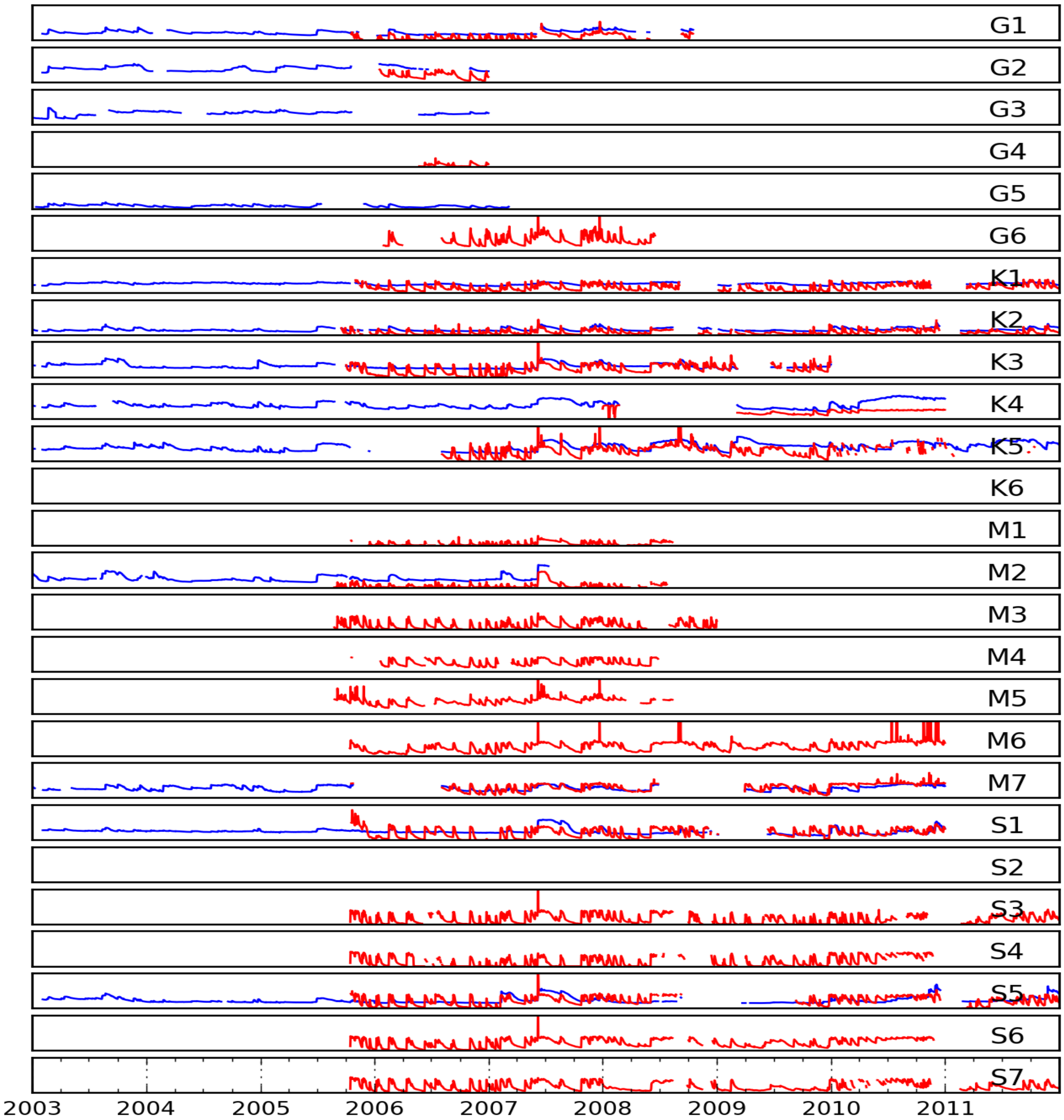

Figure 42. Indicative site daily time-series of SASMAS top layer (red: 0-5/8cm) and profile (blue: 0-90cm) volumetric soil moisture. Site numbers are listed far right. Axis scale omitted for space purposes.

Appendix C: Evaluation against top layer soil moisture

A comparison was undertaken using the time period that the satellite data was available (Jan 2007-Sept 2011) for the OzNet Murrumbidgee and SASMAS Hunter data (Figure 43. Correlation of (a) daily and (b) monthly top layer (0-5/8cm) soil moisture of models against Murrumbidgee OzNet for Jan 2007-Sept 2011). This gives an indication of how well the satellite data represents surface and profile soil moisture, compared to AWRA-L. CABLE and WaterDyn perform better than AWRA-L. AWRA-L v6 performs worse that AWRA v5. ASCAT appears to perform slightly better than AMSRE in general. AWRA-L performs better than AMSRE for surface soil moisture for the SASMAS Hunter sites, but not as well for the Murrumbidgee (for daily data also).

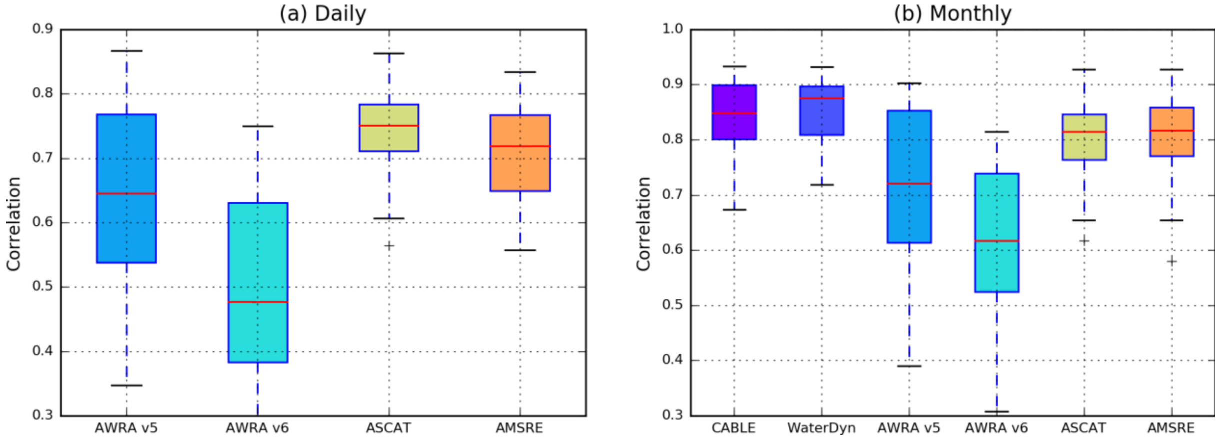

Figure 43. Correlation of (a) daily and (b) monthly top layer (0-5/8cm) soil moisture of models against Murrumbidgee OzNet for Jan 2007-Sept 2011

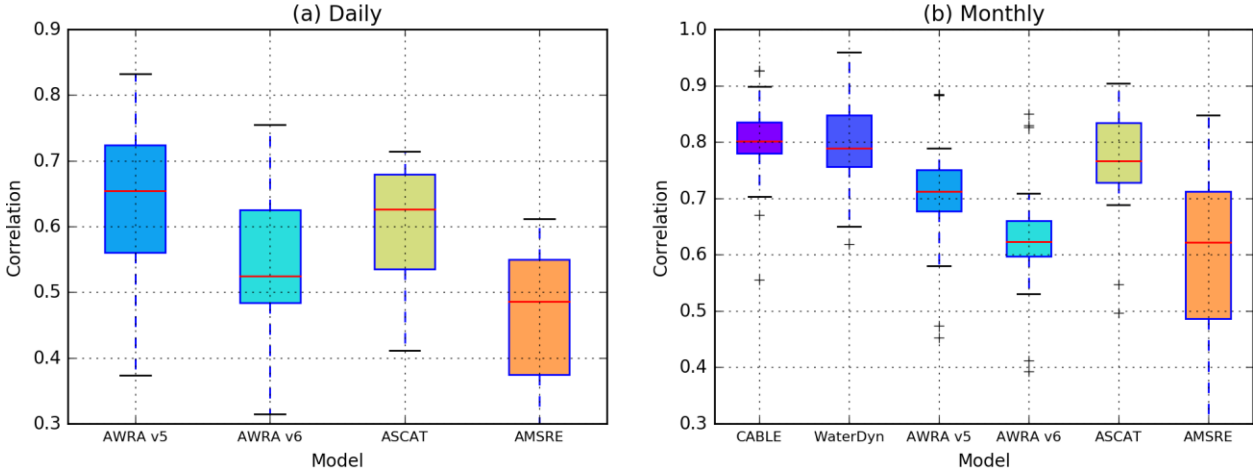

Figure 44. Correlation of (a) daily and (b) monthly top layer (0-5/8cm) soil moisture of models against Upper Hunter SASMAS data for Jan 2007-Sept 2011

Satellite data provides relatively accurate estimate of monthly and daily (not shown) surface soil moisture compared to AWRA-L in some areas (OzNet Murrumbidgee), while the model based estimates are superior in other areas (SASMAS). The difference in AMSR-E performance over the two areas follows that found (in comparing AMSR-E performance in these two areas) by Draper et al., (2009), noting that AMSR-E is no longer operational. There are multiple candidate satellite derived products available for evaluation/assimilation/calibration of water balance/land-surface models – and some debate over which satellite is best to use. See Lacava et al. (2012) for comparison of SMOS, AMSR-E and ASCAT, Leroux et al. (2013) for a comparison of SMOS, VUA (AMSR-E), ASCAT satellite based and ECMWF model forecast for surface soil moisture, and the subsequent clarification paper by Wagner et al., (2014) presenting differing results depending on the version of satellite data used and analysis method. However, all products do show use in terms of correlation to surface soil moisture. These datasets therefore serve as valuable tools for verification and calibration of AWRA-L, and more recent products will be evaluated for this purpose in future.