

## Seasonal soil moisture forecasting using the AWRA landscape water balance model

**E. Vogel**<sup>a,b</sup>, C. Holgate<sup>c,b</sup>, J. Cairns<sup>d</sup>, A. Ramchurn<sup>e</sup>, A.J. Frost<sup>e</sup>, A.I.J.M. van Dijk<sup>e</sup>

<sup>a</sup> Australian-German Climate & Energy College, The University of Melbourne, Melbourne, Australia

<sup>b</sup> ARC Centre of Excellence for Climate System Science, Australia

<sup>c</sup> Fenner School of Environment & Society, The Australian National University, Canberra, Australia

<sup>d</sup> Department of Geography, The University of Otago, Dunedin, New Zealand

<sup>e</sup> Bureau of Meteorology, Melbourne, Australia

Email: [elisabeth.vogel@climate-energy-college.org](mailto:elisabeth.vogel@climate-energy-college.org)

**Abstract:** The Australian Bureau of Meteorology (BOM) publishes and presents to stakeholders seasonal outlooks of key environmental variables including precipitation, temperature and streamflow. A similar forecast system for soil moisture (SM), if skilful, could benefit many stakeholders and sectors. Potential applications include agricultural planning (e.g. irrigation and sowing scheduling, planning of tilling, sowing and harvesting), livestock management, economic forecasts (e.g. improved crop yield predictions) and emergency management (e.g. bushfire and flood risk assessment). The aim of this study was to develop a proof-of-concept for a seasonal SM forecasting system for Australia and to test its performance. Specifically, we aimed to answer the research question: Can SM over the coming three months be forecast with better skill than the climatological expectation?

This paper presents a SM forecasting methodology that combines the POAMA-2 seasonal climate prediction system with the AWRA landscape water balance model (AWRA-L). POAMA-2 is a global coupled ocean-atmosphere model that currently underpins the BOM seasonal climate outlooks. POAMA-2 hindcasts of precipitation, air temperature and solar radiation, available for 1981-2014 on a 2.5° x 2.5° grid, were converted into a suitable format to force AWRA-L. The AWRA-L gridded water balance model simulates daily land surface water fluxes and stores, including SM for three soil layers (0-10 cm, 10-100 cm, 100-600 cm), on a national 0.05° grid. Combining these two modelling systems, we produced hindcasts of monthly SM stored in the AWRA shallow soil layer (10-100 cm soil depth: corresponding to the root depth of many agricultural crops) for the period 1981-2013, with 12 hindcasts per year starting on the 1<sup>st</sup> of each month and 22 ensemble members per hindcast. Further, we calculated probabilistic hindcasts of the occurrence of a potentially harmful event (PHE). The PHE aims to capture critically dry soil conditions with potentially adverse effects on, for example, agriculture. It is defined as a month in which the 20<sup>th</sup> percentile of SM in this month is drier than a 1-in-5-year event. Each hindcast had a length of three months, and all individual forecasts at one-, two- and three-month lead time were subsequently combined into continuous monthly time series with 22 ensemble members per time series.

We assessed the skill of the hindcasts of soil moisture against i) historical SM simulations using AWRA-L forced with daily gridded climate analysis data produced by BOM and ii) ground-based *in situ* measurements of SM (0-90cm) from the OzNet network in the Murrumbidgee catchment. The predictive capacity of the hindcasts at one-month lead time appear relatively high and are mostly related to the state initialisation of AWRA-L, highlighting the importance of soil memory and SM initialisation at shorter time scales. However, at two- and three-month lead times, our results show relatively low predictive skill across Australia, with a few exceptions in wet regions in the northern, southern and south-western parts of Australia. Most of the skill stems from the ability of the forecasting system to reproduce the mean spatial and seasonal pattern of SM, however, increasing the skill to predict deviations from the mean climatology requires further work. This study presents baseline estimates of predictive skill at different time scales, which may be used as reference for future developments of SM forecasting systems, and provides recommendations for future research.

**Keywords:** Soil moisture, seasonal forecasting, soil moisture hindcasts, climate services

## 1. INTRODUCTION

Different sectors, both in Australia and internationally, are moving toward utilisation of seasonal forecasts of soil moisture (SM). The agricultural sector can utilise forecasts to improve irrigation scheduling, planning of tilling, sowing and harvest, and livestock management. Emergency service authorities may utilise routine seasonal SM forecasts to improve the reliability of flood risk and bushfire hazard assessments. Moreover, foreknowledge of SM has the potential to improve seasonal forecasts of temperature and rainfall through feedbacks between the land surface and the atmosphere (Hirsch et al., 2014; Koster et al., 2003), although the strength of feedback across Australia, and the improvement to seasonal climate and hydrological prediction has yet to be established. The Australian Bureau of Meteorology (BOM) publishes seasonal climate outlooks ([www.bom.gov.au/climate/ahead](http://www.bom.gov.au/climate/ahead)) of rainfall, streamflow and temperature several months ahead; however, a forecast of SM conditions at similar time scales does not yet exist.

The aim of this study was therefore to develop a prototype seasonal SM forecast system for Australia to answer the question: Can SM over the coming three months be forecast with better skill than the climatological expectation? This aim was pursued by combining hindcasts of a global coupled atmosphere-ocean model that currently underpins the BOM seasonal climate outlooks (POAMA-2) and a process-based landscape water balance model (AWRA-L), which is currently used operationally to simulate landscape water fluxes and stores across Australia. The specific objectives of the study included:

1. Develop a methodology to process the POAMA-2 climate forecast output and use it to force the AWRA-L model, producing hindcasts of SM stored in the shallow soil layer (10-100 cm) across Australia at one-, two- and three-month lead times;
2. Assess the *theoretical skill* (defined in Methodology) of the hindcasts against historical SM, simulated by AWRA-L with forcing from an observational climate data set, and
3. Assess the *actual skill* of the seasonal forecasts against ground-based measurements of *in situ* SM.

The research presented in this paper was conducted as a short-term graduate student project within the 2016 Australian Climate and Water Summer Institute. As part of the programme, participants conducted research projects with partner institutions (including BOM, CSIRO and Geoscience Australia) in a four-week time frame aiming to fill research gaps related to climate and water services in Australia. Due to the limited time frame, the development of a fully operational forecasting system was beyond the scope of this study. Instead, this paper presents a simple forecasting methodology that was developed to produce a comprehensive set of SM hindcasts. It provides baseline estimates of predictive skill at different time scales and outlines recommendations for future work.

## 2. METHODOLOGY

### 2.1 Calculation of seasonal hindcasts of SM

#### *Models*

Seasonal hindcasts of SM were created by combining the POAMA-2 (Predictive Ocean Atmosphere Model for Australia) coupled ocean-atmosphere model (Hudson *et al.*, 2013) with the AWRA-L (Australian Water Resource Assessment – Landscape) water balance model version 5 (Frost *et al.*, 2016a; Viney *et al.*, 2015; original model version described in Van Dijk, 2010).

**POAMA-2** is the current operational seasonal prediction system used by BOM. POAMA-2 forecasts consist of 33 ensemble members; 11 members from three different model configurations. POAMA-2 hindcasts are available for 1981-2014, with 6 starting dates each month and a length of 270 days per hindcast, at 2.5° x 2.5° resolution.

**AWRA-L** is a gridded water balance model that simulates land surface water fluxes and stores, including runoff, evapotranspiration, deep drainage and SM for three soil layers (top: 0-10 cm, shallow: 10-100 cm, deep: 100-600 cm), on a 0.05° x 0.05° grid. Vegetation is modelled as two hydrological response units per grid cell, representing deep rooted vegetation (trees) and shallow rooted vegetation (grass). The model was developed by CSIRO and BOM and is run operationally by BOM from 1911 until yesterday ([www.bom.gov.au/water/landscape](http://www.bom.gov.au/water/landscape)) to inform the National Water Accounts, Water resource assessment, monitoring of soil moisture conditions in Australia and a number of registered users. The pre-calibrated AWRA-L model used in this study, and the *in situ* data used for evaluation of AWRA, are supplied as part of the AWRA Community modelling system (AWRA CMS: see [https://github.com/awracms/awra\\_cms](https://github.com/awracms/awra_cms)).

### Calculation of SM hindcasts

By forcing the AWRA-L model with POAMA-2 climate hindcasts, we produced hindcasts of daily SM (10–100 cm soil depth) for the period 1981–2013, starting on the 1<sup>st</sup> of each month, with each hindcast having a length of three months and composed of 22 ensemble members. POAMA-2 output was initially pre-processed to suit AWRA-L input, including: 1) extraction of required variables (precipitation, minimum and maximum air temperature, solar radiation), 2) regriding to AWRA-L resolution by nearest neighbour interpolation, 3) cropping to the AWRA-L extent, and 4) unit conversions. At the start of each hindcast, state variables were initialised using output from a historical AWRA-L run using the Australian Water Availability Project data (described in the following section).

The daily output of AWRA-L soil moisture was aggregated into monthly time series and monthly summary statistics were calculated (monthly mean, minimum, maximum, 20<sup>th</sup> and 80<sup>th</sup> percentiles). In addition, probabilistic hindcasts of the occurrence of a potentially harmful event (PHE) were calculated. The PHE aims to capture critically dry soil conditions with potentially adverse effects on, for example, agriculture or bushfire risk. It is defined here as a month in which the 20<sup>th</sup> percentile of daily mean SM (referred to as SM\_20pctl) is drier than a 1-in-5-year event. In this study, the 20<sup>th</sup> percentile of daily SM was chosen as an indicator for dry conditions as it describes approximately the driest six days of the month, unlike the total monthly minimum, which only contains information of the driest day and is capped at zero.

### Calculation of historical SM reference data

A reference dataset was separately calculated for 1911–2013 by forcing AWRA-L with historical climate data from the Australian Water Availability Project (AWAP) (Jones et al., 2009; <http://www.csiro.au/awap/>). The AWAP dataset provides gridded climate data of site-based surface observations on a 0.05° x 0.05° grid. Since solar radiation was only available from 1990 onwards, pre-1990 solar radiation data was padded with its daily climatology (using the default gap-filling method in AWRA-L) and combined with daily precipitation, minimum and maximum air temperature data covering the period 1911–2013. Daily outputs of SM were aggregated to monthly values and statistics calculated. The output from this historical SM simulation was used for a) the initialisation of hindcasts, b) as a historical reference for threshold-based SM indicators and c) for the evaluation of the predictive skill of the SM hindcasts.

## 2.2 Tests of forecast skill

### Deterministic forecasts of mean monthly SM

We tested the *theoretical* and *actual* skill of the deterministic SM forecasts at one-, two- and three-month lead times using the hindcast ensemble mean of monthly mean SM. The *theoretical* or potential skill of the hindcasts was assessed by comparing the SM hindcasts with the historical reference (AWRA-L estimates forced by the AWAP data, described in the previous section), for every grid cell separately. This comparison assumes that the historical reference time series represents the “true” historical SM conditions across the country. It captures SM prediction errors due to uncertainties in the climate data (model biases in POAMA-2 and initialisation uncertainties of the climate data) and their propagation through to errors in SM predictions, but it does not capture systematic model biases in AWRA-L and errors related to the SM initialisation. The *actual* skill captures uncertainties related to climate inputs, model biases and uncertainties in the initialisation of both the climate and SM hindcasts.

The *actual* skill was calculated by comparing the SM hindcasts with point-based SM *in situ* measurements within the OzNet network (Smith et al., 2012), with observations covering 0–90cm depth. These data were collated (with a range of other soil moisture, evapotranspiration, streamflow and recharge datasets) and formatted for AWRA-L evaluation and comparison purposes (Frost et al., 2016b). We compared the hindcasts with the ground-based measurements from 38 sites covering 2001–2013 across four sampling areas

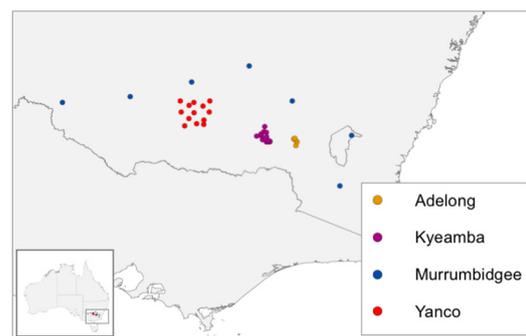


Figure 1. Locations of *in situ* soil moisture measurements within OzNet network

within the Murrumbidgee catchment in South Eastern Australia (see **Figure 1**: Murrumbidgee, Adelong, Kyeamba and Yanco).

For both the theoretical and actual skill assessment, the coefficient of determination ( $R^2$ ) from a regression of predicted SM against reference and observed SM was calculated. For the theoretical skill, we additionally calculated the  $R^2$  from anomalies of SM after subtraction of the mean climatology to assess whether the model is, for example, skilful for SM anomalies, but not SM itself due to biases introduced by POAMA-2.

### ***Probabilistic forecasts of a potentially harmful event***

The probabilistic forecast of a PHE – i.e. the likelihood of such an event occurring – was assessed against the historical reference (assessment of theoretical skill), but not against OzNet measurements, as their time range was insufficiently long to determine the relative thresholds required for the PHE calculation. We assessed the skill by calculating the Brier score and Brier skill score.

The Brier score captures the magnitude of the probability forecast errors (Brier, 1950; Wilks, 2011) relative to a reference, and is calculated as follows:

$BS = \frac{1}{N} \sum_{i=1}^N (p_i - o_i)^2$ ; with  $p_i$  being forecast probability of an event in observation  $i$ , and  $o_i$  the actual occurrence of event in observation  $i$  (0 or 1).

The Brier score is analogous to the mean squared error (MSE), but for discrete or binary target variables. It ranges between zero and one, with zero being the perfect score (Wilks, 2011). The Brier skill score evaluates the relative skill of a forecast over that of the long-term climatology; it is calculated as follows:

$BSS = 1 - \frac{BS}{BS_{reference}}$ ; with  $BS$  being the Brier score of the forecast, and  $BS_{reference}$  the Brier score of the climatology or a “naïve” reference forecast.

Here, we used a 20% probability of a PHE occurring in every grid cell and year as reference forecast, related to the percentile-based definition of the PHE as an event occurring – on average – in one out of five years. A Brier skill score of one indicates a perfect forecast; zero indicates that the forecast and the reference have equal skill, and a negative value indicates that the forecast is worse than the climatological expectation (Wilks, 2011).

## **3. RESULTS AND DISCUSSION**

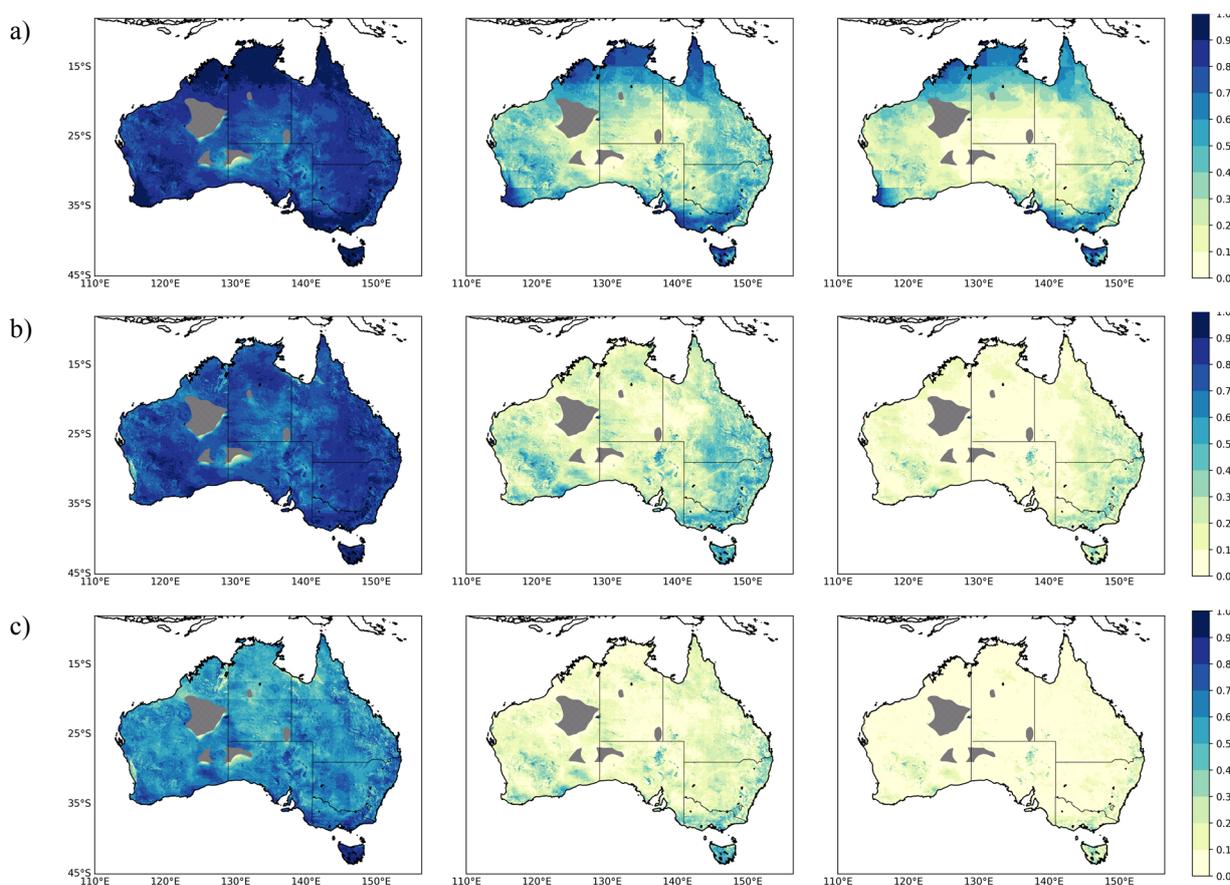
Hindcasts of monthly SM (10-100 cm soil depth) were produced for the period 1981-2013, comprising 12 hindcasts per year starting on the 1st of every month. Each hindcast had a length of three months and consisted of 22 ensemble members each comprising the following variables: i) monthly mean, minimum and maximum SM, ii) the 20<sup>th</sup> and 80<sup>th</sup> percentile of daily SM and iii) the probability of a PHE.

### **3.1 Theoretical skill – comparison with historical AWRA-L run**

We compared the hindcast ensemble mean with the historical reference time series to assess the theoretical skill of the seasonal forecasts of mean monthly SM. The coefficient of determination of SM is very high throughout the country at one-month lead time (mean  $R^2$  of 0.8) and decreases with increasing lead times (mean  $R^2$  values of 0.4 and 0.3 at two- and three-month lead time) (**Figure 2a**). The spatial patterns look similar for different lead times. The explained variance of SM anomalies is highest in regions in equatorial and tropical regions of Western Australia, Northern Territory and Queensland as well as in temperate regions, particularly South-West Australia, Tasmania, large parts of Victoria and parts of New South Wales and South Australia. Some regions in inland Australia, particularly the Gibson Desert in Western Australia, exhibit very low predictive skill due to a lack of proximal rain gauges in these regions and therefore poor or missing rainfall estimates in the AWAP dataset (regions with less than ten years of rainfall data were masked).

In terms of SM anomalies (deviations from the mean climatology of SM), the explained variance is particularly high for the forecasts at one-month lead time, however,  $R^2$  values decrease sharply at two-month and three-month lead times (mean  $R^2$  at one-, two- and three-month lead time are 0.75, 0.24 and 0.1, respectively) and are considerably lower than for the explained variance of actual SM (**Figure 2b**). The difference in skill between actual SM and SM anomalies (panel a and b) suggests that most of the predictive capacity at two- and three-month lead time is associated with the capacity of the forecasting system to correctly reproduce the mean climatology of SM.

We further assessed the contribution of SM memory to predictive skill by comparing the explained variance of the hindcasts with hindcasts that were initialised using the climatology of SM, as default states provided in AWRA-L. The comparison shows that most of the explained variance at one-month lead time is related to the SM initialisation, highlighting the importance of memory for the prediction of SM at shorter time scales (**Figure 2c**). However, at two- and three-month lead time, the initialisation of SM plays a minor role, as the difference in forecast skill is negligible in most parts of the country. This is likely related to the fact that most of the explained variance at these lead times was found to be associated with the ability of the forecasting system to reproduce the mean climatology of SM; and therefore, initialising the hindcasts with default SM states in this case does not reduce the explained variance, compared to using the correct initialisation. The strongest contribution of initial conditions to forecasting skill (i.e. a large difference in  $R^2$  between default and correct initialisation) can be found in southern Victoria and Tasmania, which may indicate higher SM memory in these regions than in other parts of Australia. The difference between panel b and c in **Figure 2** is equal to the skill attributable to the POAMA-2 hindcasts, indicating particularly high contribution of predictive capacity of the seasonal forecast system in parts of Victoria, New South Wales and the South-West of Western Australia.

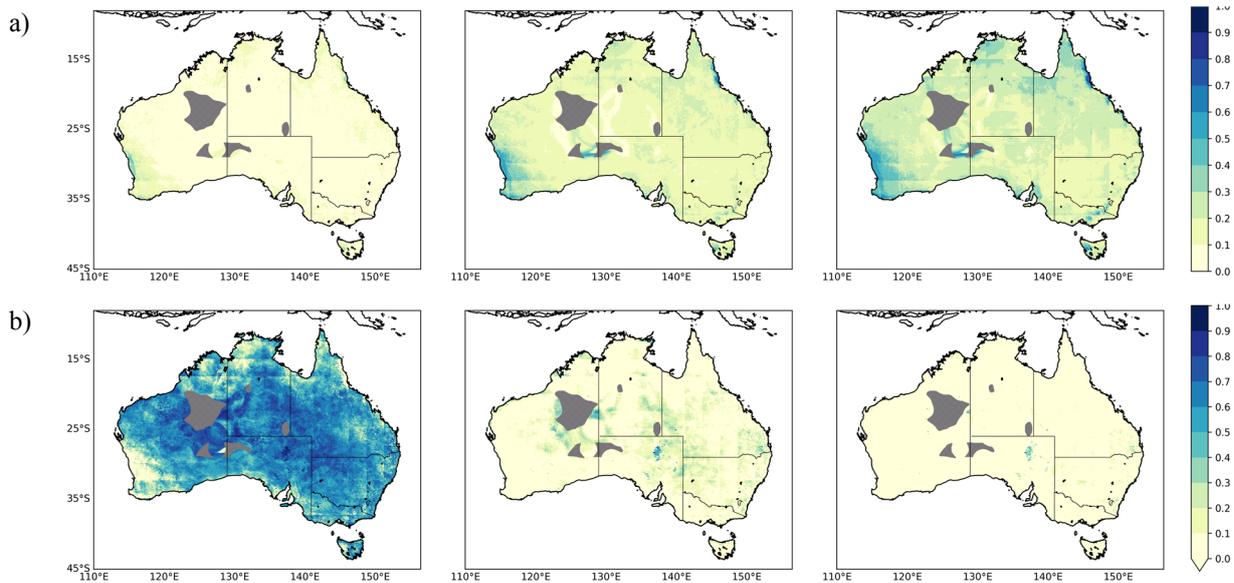


**Figure 2.** Theoretical hindcast skill for mean monthly SM at one-, two- and three-month lead times (columns 1, 2 and 3, respectively). Skill is expressed as the fraction of explained variance ( $R^2$ ) in reference model simulations using historical forcing. a) Skill of actual SM; b) skill of SM anomalies relative to climatology; c) contribution of SM initial state to skill improvement (difference in  $R^2$  of hindcasts using actual initial states and climatology initial states). Regions with fewer than 10 daily rainfall data points in the AWAP dataset were masked.

**Probabilistic forecasts of potentially harmful event**

The Brier score of the probabilistic forecast of the PHE is relatively homogenous throughout most of the country (**Figure 3a**) and increases with lead time, meaning the skill decreases with increasing lead time. The mean Brier scores across Australia are 0.07, 0.17 and 0.23, at one-, two- and three-month lead times respectively. The maps of Brier skill score indicate that the seasonal forecast system of the binary event increases the skill considerably at one-month lead time, to some degree in certain regions at two-month lead

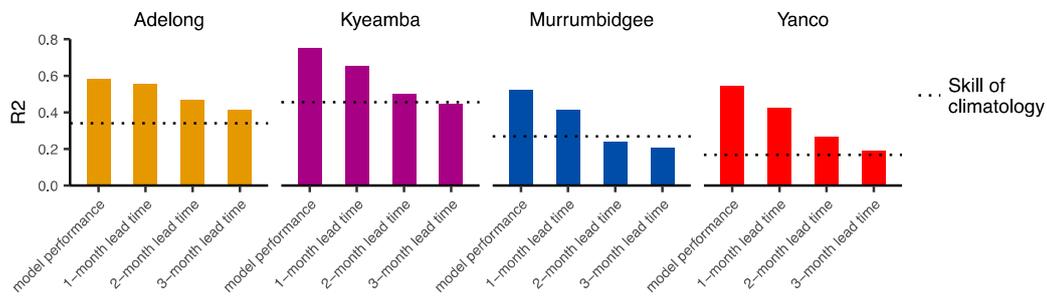
time and no skill is added at three-month lead time (**Figure 3b**). At one-month lead time, the mean Brier skill score is 0.55, i.e. on average the seasonal forecast reduces the error by 55% compared to the naïve forecast.



**Figure 3.** Skill assessment for probabilistic forecast of a PHE at one-, two- and three-month lead time (columns 1, 2 and 3 respectively). The probabilistic PHE forecast describes the likelihood of the six driest days being at or below a 1-in-5-year event threshold. a) Brier score. b) Brier skill score (BSS) as comparison against reference forecast, predicting a 20% chance of the PHE in every year and grid cell. Negative BSS values (indicating worse skill than the reference) were set to zero. Regions with fewer than 10 daily rainfall data points in the AWAP dataset were masked.

### 3.2 Actual skill – comparison with OzNet *in situ* measurements

The coefficient of determination based on the AWRA-L historical run represents the model performance, i.e., the maximally attainable skill, and ranges between approximately 0.45 and 0.75 (**Figure 4**, top panel). The findings of the explained variance at the very local scale are comparable to the results at grid cell scale. As can be expected, the SM hindcasts have decreasing skill with increasing lead times. The hindcasts at one- and two-month lead time have higher skill than the SM climatology across all locations. However, at all measurement sites except Adelong, the three-month hindcasts have less or similar skill as the mean climatology. Further, we find that the actual skill is equal to the product of theoretical skill and the historical model performance of AWRA-L – expressed as  $R^2$  – for each station and lead time (values not shown here), which is similar to results found in previous research on stream flow forecasts (Van Dijk et al., 2013). These findings suggest that the actual skill at a new site can be estimated by combining the theoretical skill maps with data on the historical model performance of AWRA-L. This allows to estimate the skill at new locations as well as to separate sources of error into biases in the water balance model or the seasonal climate forecast.



**Figure 4.** Mean coefficient of determination ( $R^2$ ) from a regression of monthly mean SM (historical simulation and hindcasts at three lead times) against OzNet *in situ* SM measurements, averaged over all sites at one location.

#### 4. CONCLUSIONS

This study presented a proof-of-concept of a seasonal forecasting system of SM for Australia based on POAMA-2 and AWRA-L and provides baseline estimates of forecast skill. We developed seasonal hindcasts of SM in the shallow soil layer for 1982-2013 and assessed the predictive skill using a range of skill measures that capture the capacity of the forecasts to reproduce mean monthly SM temporal dynamics, relative anomalies of SM as well as the likelihood of a potentially harmful event. We find that using the current modelling set-up, the predictive skill is good at one-month lead time, but decreases considerably at two- and three-month lead time. This study has a number of limitations which may be addressed in future studies; steps to develop the seasonal forecasting system further include:

- Improving the spatial resolution of the climate forecasts by applying an appropriate downscaling method to the POAMA-2 hindcasts (e.g. a bias correction, quantile-quantile regression, analogue downscaling);
- Broadening the skill assessment, e.g. calculating ROC curves and reliability diagrams (Wilks, 2011); comparing with additional *in situ* SM data from elsewhere in Australia; testing skill separated by month and season and at different time scales (daily vs. monthly), and adding skill for other prediction targets;
- Investigating skill improvements achievable by using seasonal climate forecasts of precipitation, air temperature, solar radiation, from the Bureau's next generation seasonal forecast climate system (ACCESS-S) once available, and comparing the resulting skill with that of SM estimates internal to the ACCESS-S model.

#### ACKNOWLEDGEMENTS

This study was conducted as part of the 2016 Australian Climate and Water Summer Institute and would not have been possible without the support, organisation and funding by the OzEWEX community and partner institutions. We would particularly like to express our gratitude to Amgad Elmahdi, Julien Lerat and Francis Chiew for sharing their knowledge and expertise with us. Further, we would like to thank Rowena Smith and Zac Hatfield-Dodds for their organisational and technical support during the Summer Institute. We would like to acknowledge all lecturers, presenters and fellow participants at the Summer Institute for their insights and valuable feedback. Authors E. Vogel and C.M. Holgate acknowledge the support of the ARC Centre of Excellence for Climate System Science. E. Vogel and C.M. Holgate are supported by an Australian Government Research Training Award.

#### REFERENCES

- Brier, G.W. (1950). Verification of forecasts expressed in terms of probability. *Mon. Weather Rev.* **78**, 1–3.
- Frost, A.J., Ramchurn, A., and Smith, A. (2016a). The Bureau's Operational AWRA-L Model: Technical description of AWRA-L version 5. Bureau of Meteorology Technical Report.
- Frost, A.J., Ramchurn, A., and Hafeez, M. (2016b). Evaluation of the Bureau's Operational AWRA-L Model. Bureau of Meteorology Technical Report.
- Hirsch, A.L., Pitman, A.J., and Kala, J. (2014). The role of land cover change in modulating the soil moisture-temperature land-atmosphere coupling strength over Australia. *Geophys. Res. Lett.* **41**, 5883–5890.
- Hudson, D., Marshall, A.G., Yin, Y., Alves, O., and Hendon, H.H. (2013). Improving intraseasonal prediction with a new ensemble generation strategy. *Mon. Weather Rev.* **141**, 4429–4449.
- Jones, D.A., Wang, W., and Fawcett, R. (2009). High-quality spatial climate data-sets for Australia. *Aust. Meteorol. Oceanogr. J.* **58**, 233.
- Koster, R.D., Suarez, M.J., Higgins, R.W., and Van den Dool, H.M. (2003). Observational evidence that soil moisture variations affect precipitation. *Geophys. Res. Lett.* **30**, 1241.
- 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 Resour. Res.* **48**, W07701.
- Van Dijk, A. (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 Dijk, A.I.J.M., Peña-Arancibia, J.L., Wood, E.F., Sheffield, J., and Beck, H.E. (2013). Global analysis of seasonal streamflow predictability using an ensemble prediction system and observations from 6192 small catchments worldwide. *Water Resour. Res.* **49**, 2729–2746.
- Viney, N., Vaze, J., Crossbie, R., Wang, B., Dawes, W., and Frost, A. (2015). AWRA-L v5.0: Technical description of model algorithms and inputs (CSIRO Land and Water).
- Wilks, D.S. (2011). *Statistical methods in the atmospheric sciences* (Academic press).