Performance of ACCESS-S1 for key horticultural regions

Hudson, D., Shi, L., Alves, O., Zhao, M., Hendon, H.H. and Young, G.

March 2017
Performance of ACCESS-S1 for key horticultural regions

Hudson, D., Shi, L., Alves, O., Zhao, M., Hendon, H.H., Young, G.

Bureau Research Report No. 020

March 2017

National Library of Australia Cataloguing-in-Publication entry

Author: Hudson, D., Shi, L., Alves, O., Zhao, M., Hendon, H.H., Young, G.

Title: Performance of ACCESS-S1 for key horticultural regions

ISBN: 978-0-642-70688-1

Series: Bureau Research Report – BRR020
Enquiries should be addressed to:

Contact Name: Debra Hudson

Bureau of Meteorology
GPO Box 1289, Melbourne
Victoria 3001, Australia

Contact Email: D.Hudson@bom.gov.au

Copyright and Disclaimer

© 2017 Bureau of Meteorology. To the extent permitted by law, all rights are reserved and no part of this publication covered by copyright may be reproduced or copied in any form or by any means except with the written permission of the Bureau of Meteorology.

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

Abstract ........................................................................................................................................... 1

1. Introduction ................................................................................................................................. 2

2. The ACCESS-S1 forecast system .............................................................................................. 3
   2.1 The coupled model, data assimilation and ensemble generation ........................................ 4
   2.2 Hindcast configuration ............................................................................................................ 5

3. Skill evaluation .......................................................................................................................... 6
   3.1 Seasonal forecasts ............................................................................................................... 6
       3.1.1 Rainfall ......................................................................................................................... 7
       3.1.2 Maximum temperature ............................................................................................... 8
       3.1.3 Minimum temperature ............................................................................................... 10
   3.2 Multi-week forecasts .......................................................................................................... 11
   3.3 Performance of three-category seasonal forecasts ............................................................ 15

4. Discussion and Conclusions ...................................................................................................... 21

Acknowledgements ..................................................................................................................... 23

References ....................................................................................................................................... 23

Appendix A: Skill maps for Australia ............................................................................................ 27

Appendix B: Land surface initial conditions experiment ............................................................. 31
List of Tables

Table 1: The ACCESS-S1 forecast system.................................................................................. 3

List of Figures

Figure 1: The nine horticulture regions analysed in the skill evaluation................................. 6

Figure 2: Correlation skill for 0-month lead (top) forecasts of seasonal mean rainfall anomalies for the FMA (red), MJJ (blue), ASO (green) and NDJ (orange) seasons; and for 1-month lead (bottom) forecasts of the MAM (red), JJA (blue), SON (green) and DJF (orange) seasons. Skill is shown for the horticulture regions (Figure 1), indicated from left to right: Bowen, Bundaberg, Sydney Basin, Werribee/Cranbourne/Gippsland; Mildura; Devonport; Adelaide Plains; south western WA and Carnarvon. Statistically significant correlations exceed 0.4 (at the 95% confidence level, n=23). .............................................................................. 8

Figure 3: As for Figure 2, but for Tmax. .................................................................................... 9

Figure 4: As for Figure 2, but for Tmin. .................................................................................... 11

Figure 5: Correlation skill for multi-week rainfall anomalies for the summer half of the year (i.e., forecasts initialised on the 25 Oct, 1 Nov, 25 Jan, 1 Feb) (top) and the winter half of the year (i.e., forecasts initialised on the 25 Apr, 1 May, 25 Jul, 1 Aug) (bottom). Skill is shown for fortnight 1 (weeks 1+2; Red), fortnight 2 (weeks 2+3; Blue), and fortnight 3 (weeks 3+4; Green) of the forecast for each of the horticulture regions (Figure 1), indicated from left to right: Bowen, Bundaberg, Sydney Basin, Werribee/Cranbourne/Gippsland; Mildura; Devonport; Adelaide Plains; Southern Western Australia and Carnarvon. Statistically significant correlations exceed 0.2 (at the 95% confidence level, n=92) ................................... 13

Figure 6: As for Figure 5, but for Tmax. .................................................................................... 14

Figure 7: As for Figure 5, but for Tmin. .................................................................................... 15

Figure 8: Schematics to describe the tercile scoring used in Figures 9-12. Panel a) indicates a hypothetical time series of tercile forecasts and observations for rainfall for a given region (it applies equally to temperature, except the y-axis would read "colder than normal" and "hotter than normal") (after Abawi et al. 2008). "O" denotes the tercile with the observed value of rainfall e.g., in 1990 the rainfall was wetter than normal and in 1992 it was drier than normal. "P" denotes the tercile with the highest forecast probability e.g., the model predicted that it would most likely be wetter than normal in 1990 and most likely be normal in 1992. The year is shaded in blue if the observed rainfall was in the same tercile as the tercile with the highest probability (consistent forecast). Red indicates that the observed rainfall was different by two categories from the most probable forecast category (inconsistent forecast) and green indicates that the observed was in the neighbouring category (near consistent forecast). This colour scoring is summarised in the table in b). .................................................................................. 16

Figure 9: Time series of consistent (blue), near consistent (green) and inconsistent (red) (as defined in Figure 8) forecasts for each of the horticulture regions for forecasts of the early spring season (ASO) rainfall at 0-month lead time for 1990-2012. The numbers in the right hand panel indicate the number of forecasts (out of the 23...
years) that were either consistent, near consistent or inconsistent. The horticulture regions are indicated on the y-axis and are, from bottom to top: Bowen, Bundaberg, Sydney Basin, Werribee/Cranbourne/Gippsland; Mildura; Devonport; Adelaide Plains; Southern Western Australia and Carnarvon. ................................

Figure 10: The percentage of the 23 years that have consistent (blue), near consistent (green) and inconsistent (red) forecasts for 0-month lead forecasts of rainfall for the seasons indicated on the panels. The horticulture regions are indicated from left to right: Bowen, Bundaberg, Sydney Basin, Werribee/Cranbourne/Gippsland; Mildura; Devonport; Adelaide Plains; south western WA and Carnarvon. ................

Figure 11: As for Figure 10, but for Tmax. .................................................................................

Figure 12: As for Figure 10, but for Tmin. ..................................................................................

List of Figures in Appendix A

Figure A 1: Correlation skill for 0 month-lead forecasts of seasonal mean rainfall anomalies for the FMA, MJJ, ASO and NDJ seasons. The dashed line indicates the correlation value above which the correlations are statistically significant (i.e. r>0.4 is significant, 95% confidence level, n=23). .............................................................

Figure A 2: Correlation skill for 1 month-lead forecasts of seasonal mean rainfall anomalies for the MAM, JJA, SON and DJF seasons. The dashed line indicates the correlation value above which the correlations are statistically significant (i.e. r>0.4 is significant, 95% confidence level, n=23). .............................................................

Figure A 3: As for Figure A1, but for Tmax.................................................................................

Figure A 4: As for Figure A2, but for Tmax ................................................................................

Figure A 5: As for Figure A1, but for Tmin .................................................................................

Figure A 6: As for Figure A2, but for Tmin ................................................................................

Figure A 7: Correlation skill for the fortnights comprising weeks 1 and 2 (left column), weeks 2 and 3 (middle column) and weeks 3 and 4 (right column) for rainfall (top row), Tmax (middle row) and Tmin (bottom row) using all of the 8 forecast start dates noted in section 2.2. The dashed line indicates the correlation value above which the correlations are statistically significant (i.e. r>0.14 is significant, 95% confidence level, n=184). If there is no dashed line shown on a map, then all the correlations are significant......................................................

List of Figures in Appendix B

Figure B 1: Correlation skill for Tmax for 0-month lead forecasts of the month of May (1990-2012) from (a) the sensitivity experiment where the forecasts are initialised with realistic soil moisture conditions, and (b) ACCESS-S1, where the forecasts are initialised using climatological soil moisture..................................................
ABSTRACT

ACCESS-S1 will be the next version of the Bureau of Meteorology's seasonal prediction system, due to become operational in 2017. Here we evaluate the performance of ACCESS-S1 for forecasts of Australian climate on seasonal and multi-week timescales, focusing on 9 key regions of importance to the horticultural industry, specifically vegetable growers.

We have examined the skill of forecasts of rainfall, maximum temperature (Tmax) and minimum temperature (Tmin). The skill of the forecasts varies with region, time of year, variable and forecast lead time (i.e. how much advance warning). For example, the spring season tends to be the most skilful, particularly for the eastern and south-eastern regions for rainfall and Tmax. In general, forecasts for temperature are more skilful than those for rainfall. Early summer Tmax forecasts exhibit good skill across all the horticulture regions. For a given forecast period, skill decreases as the lead time increases (e.g. fortnight 1 of the forecast is more skilful than fortnight 2). Our analysis has shown that there is scope for improved performance in ACCESS-S2 (the next version of the system). ACCESS-S1 is reliant on the UK Met Office's (UKMO) initial conditions of the atmosphere, ocean and land. The UKMO initialise the land surface using climatological conditions, rather than time-varying realistic soil moisture. Our experiments have shown that this impacts negatively on the skill of the multi-week (from week 4) and seasonal forecasts, particularly for Tmax over the eastern horticulture regions and to a lesser degree on Tmin over the south-eastern regions for forecasts in late autumn and winter. In ACCESS-S2 we will not be reliant on the UKMO initialisation strategy; instead the Bureau's data assimilation will be used, whereby the land surface will be initialised with realistic initial conditions.

The evaluation of forecast performance has focused on key regions of importance to vegetable growers. Going forwards, to get optimal use out of climate forecasts, there is a need to partner with the industry to determine both the level of skill that is useful and what management decisions could be made on the basis of the forecasts at their given level of skill. For each region and crop there may be different climate related risks and management decisions at different times of the year. The key question is whether the forecasts lead to better decision-making.
1. INTRODUCTION

Climate variability is the largest driver of annual agricultural income and production variation in Australia (Love 2005; Howden et al. 2007; Stokes and Howden 2010) and the horticultural industry is no exception. Horticulture Innovation Australia is funding a project to evaluate the performance of multi-week and seasonal forecasts from ACCESS-S1 for regions relevant to the Australian vegetable industry. Climate forecasts provide agricultural users with information about the expected temperature and rainfall conditions in the coming weeks to seasons, providing essential information for planning and investment decisions.

ACCESS-S1 (the seasonal prediction version of the Australian Community Climate and Earth-System Simulator) will be the Bureau of Meteorology’s next generation seasonal prediction system, replacing the current operational system, POAMA, in 2017. ACCESS-S1 has considerable enhancements compared to POAMA, including higher resolution of the component models and state-of-the-art physics parameterisation schemes. ACCESS-S1 uses the seasonal prediction global coupled model from the UK Met Office (UKMO), as well as their initial conditions, but the system is enhanced with our own perturbation scheme for the generation of the forecast ensemble (to make it appropriate for multi-week forecasts), a larger forecast ensemble and a longer hindcast period. The first version of ACCESS-S is essentially a "fast-track" option, as it includes a number of compromises (e.g., it will not include the locally developed coupled assimilation strategy), but the benefit is earlier delivery of higher resolution forecasts for Australia.

A previous report evaluated the forecast performance for the horticulture regions of a preliminary set of hindcasts and compared the skill to that of POAMA (Shi et al. 2016). The results showed that the forecasts have much more regional detail than the POAMA forecasts and a considerably better simulation of the mean state for rainfall, maximum temperature (Tmax) and minimum temperature (Tmin) over Australia. Overall the results were positive, with the forecasts being generally more skilful than POAMA. For the nine horticulture regions considered together, there was generally more improvement in skill compared to POAMA than there was degradation, in all except three cases. These exceptions were for forecasts of seasonal mean Tmax for late autumn and winter, and forecasts of fortnightly mean Tmax for fortnight 3 (weeks 3+4 of the forecast) in autumn months, where POAMA was better. The multi-week forecasts were shown to be clearly more skilful than POAMA for the first two fortnights of the forecast (i.e. weeks 1+2, weeks 2+3) for all three variables and horticulture regions. This was a preliminary study since confidence in the results was limited by the relatively small ensemble size (3-members per start date) and the short hindcast period (14 years).

This report builds on previous work, focusing on the performance of the final configuration of the ACCESS-S1 system, which has a larger hindcast period and ensemble size than analysed previously. Here we report on the skill of seasonal and multi-week forecasts of rainfall, Tmax and Tmin for nine key horticultural regions.
2. THE ACCESS-S1 FORECAST SYSTEM

The ACCESS-S1 forecast system is summarised in Table 1 and described below.

Table 1: The ACCESS-S1 forecast system

<table>
<thead>
<tr>
<th>Component</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Atmospheric model</strong></td>
<td>Global Atmosphere 6.0: The Unified Model (UM; Walters et al. 2014), which is part of the UKMO coupled model GC2 (Williams et al. 2015).</td>
</tr>
<tr>
<td></td>
<td>Horizontal resolution: N216 (~60 km in the mid-latitudes)</td>
</tr>
<tr>
<td></td>
<td>Vertical resolution: 85 levels (extending into the stratosphere)</td>
</tr>
<tr>
<td><strong>Land surface model</strong></td>
<td>Global Land 6.0: Joint UK Land Environment Simulator (JULES; Best et al. 2011; Walters et al. 2015) with 4 soil levels, which is part of the UKMO coupled model GC2.</td>
</tr>
<tr>
<td><strong>Ocean model</strong></td>
<td>Global Ocean 5.0: Nucleus for European Modelling of the Ocean (NEMO; Madec 2008, Megann et al. 2014) which is part of the UKMO coupled model GC2.</td>
</tr>
<tr>
<td></td>
<td>Horizontal resolution: 0.25°</td>
</tr>
<tr>
<td></td>
<td>Vertical resolution: 75 levels. Level thicknesses range from 1 m near the surface to about 200 m near the bottom (6000 m depth)</td>
</tr>
<tr>
<td><strong>Sea ice model</strong></td>
<td>Global Sea Ice 6.0: Los Alamos sea ice model (CICE; Rae et al. 2015), which is part of the UKMO coupled model GC2.</td>
</tr>
<tr>
<td><strong>Atmosphere initial conditions</strong></td>
<td>ERA-Interim (Dee et al. 2011) for the hindcasts</td>
</tr>
<tr>
<td></td>
<td>Global NWP analysis for the real-time forecasts</td>
</tr>
<tr>
<td><strong>Ocean and ice initial conditions</strong></td>
<td>UKMO NEMO 3-dimensional variational ocean data assimilation (NEMOVAR; Mogensen et al. 2009, 2012)</td>
</tr>
<tr>
<td><strong>Land surface initial conditions</strong></td>
<td>Climatological soil moisture (MacLachlan et al. 2014)</td>
</tr>
<tr>
<td><strong>Ensemble generation</strong></td>
<td>Initial condition uncertainty: BoM perturbation scheme and lagged initial conditions</td>
</tr>
<tr>
<td></td>
<td>Model uncertainty/unresolved processes: UKMO stochastic physics scheme (SKEB2; Bowler et al. 2009).</td>
</tr>
</tbody>
</table>
2.1 The coupled model, data assimilation and ensemble generation

The coupled model and the data assimilation for ACCESS-S1 are the same as that used for seasonal forecasting at the UKMO i.e., the GloSea5-GC2 system. The key differences between ACCESS-S1 and GloSea5-GC2 are in the ensemble generation and the hindcast configuration.

The global coupled model is the UKMO's Global Coupled configuration 2 (GC2) and is comprised of: Global Atmosphere 6.0 (GA6.0), Global Ocean 5.0 (GO5.0), Global Land 6.0 (GL6.0) and Global Sea Ice 6.0 (GSI6.0) (Williams et al. 2015). The atmosphere model is run at a ~60 km horizontal resolution (N216 resolution) with 85 levels in the vertical (Walters et al. 2014). The land surface model is the JULES (Joint UK Land Environment Simulator) model (Best et al. 2011; Walters et al. 2015) and is tightly coupled to the atmosphere model on the same horizontal grid. It has 4 soil levels and a sophisticated representation of soil and surface hydrology and of fluxes of heat and moisture within the soil and to the atmosphere. Land cover heterogeneity is represented. The ocean model is based on the NEMO3.4 (Nucleus for European Modelling of the Ocean) model and has a 0.25° horizontal resolution with 75 levels in the vertical (Madec 2008, Megann et al. 2014). The sea ice model is the CICE4.1 model (Rae et al. 2015) and has 5 thickness categories. It is tightly coupled to the ocean model on the same horizontal grid. The atmosphere/land and the ocean/sea-ice are coupled every 3 hours using the Ocean Atmosphere Sea Ice Soil coupler (OASIS3, Valcke, 2013).

Climate forcing for greenhouse gases (e.g. CO₂ and methane) are set to observed values up to the year 2005 and after this the emissions follow the Intergovernmental Panel on Climate Change (IPCC) RCP4.5 scenario (MacLachlan et al. 2014). For ozone, the observational seasonally-varying climatology is used.

The atmospheric initial conditions for both ACCESS-S1 and GloSea5-GC2 real-time forecasts are taken from the UKMO numerical weather prediction (NWP) analysis (which has the same grid as the GC2 atmospheric model) and the hindcasts use the ERA-Interim reanalysis (Dee et al. 2011) that are interpolated onto the GC2 grid. The ocean and sea ice are initialised using the FOAM (Forecast Ocean Assimilation Model) analyses (Blockley et al. 2014) in both the real-time forecasts and the hindcasts. FOAM uses the same ocean and sea-ice models as GC2 and uses the NEMO 3-dimensional variational ocean data assimilation (NEMOVAR; Mogensen et al. 2009, 2012). Soil moisture is initialised with a monthly climatology of a land surface reanalysis using the JULES model (run in stand-alone mode) forced with the Integrated Project Water and Global Change Forcing Data methodology applied to ERA-Interim data (Weedon et al. 2011, MacLachlan et al. 2014). Soil temperatures are initialised using time-varying ERA-Interim data.

Model uncertainty in ACCESS-S1 (and GloSea5-GC2) is simulated using the Stochastic Kinetic Energy Backscatter (SKEB2) scheme (Bowler et al. 2009). The stochastic physics scheme (SKEB2) is designed to represent unresolved processes and to provide grid-scale perturbations during the model integration (MacLachlan et al. 2014). Each ensemble member evolves differently due to these grid-scale perturbations.

The representation of initial condition uncertainty for ACCESS-S1 differs from that of the UKMO GloSea5-GC2 system. The UKMO GC2 system simulates initial condition uncertainty using a lagged ensemble (combining forecasts from earlier start times). There are no perturbed initial conditions i.e., ensemble members that are initialised on the same date use the same initial conditions and will diverge from each other during the forecast as a result of the stochastic physics. This approach is appropriate for seasonal prediction, which is the focus of
the UKMO GloSea5-GC2 system, but is not optimal for multi-week forecasting since it does
not create enough ensemble spread in the first month of the forecasts. Furthermore, the design
of the hindcast set as produced at the UKMO, with 4 forecast starts per month and 3 ensemble
members per start (Shi et al. 2016), means that a viable lagged ensemble cannot be generated or
assessed for multi-week predictions, because the time between forecast starts is too great and it
would impact negatively on the skill. Ultimately, for the next version of the system, ACCESS-
S2, perturbed ensembles will be produced using the Bureau's local assimilation scheme (Hudson
et al. 2013; Yin et al. 2011). However, in the meantime, a viable yet practical approach has been
developed to produce ensemble forecasts for ACCESS-S1 that are suitable for multi-week
prediction in real-time and in the hindcasts.

The new ensemble perturbation scheme developed for ACCESS-S1 generates 10 additional
ensemble members by perturbing the atmosphere. A larger (lagged) ensemble can be generated
by optimal combination of forecasts from succeeding previous initialisation day(s). The method
of creating these perturbations is described in detail in Lim et al. (2016). The 11 member
ensemble for each forecast start date will diverge as a result of the initial condition
perturbations, as well as the stochastic physics.

### 2.2 Hindcast configuration

The start dates of the hindcasts are restricted by the availability of initial conditions from the
UKMO, which is currently for the 23 year period 1990-2012. Initial conditions for the
hindcasts are only available on the 1<sup>st</sup>, 9<sup>th</sup>, 17<sup>th</sup>, and 25<sup>th</sup> of the month. Hence, the hindcast set
for ACCESS-S1 will consist of an 11 member ensemble from those dates each month for the
period 1990-2012. The full hindcast set is still in the process of being generated.

The hindcasts generated on the following 8 start-dates are used in this paper to estimate the
performance of ACCESS-S1 over the 23 year period: 25<sup>th</sup> of the months of January, April, July,
October; and 1<sup>st</sup> of the months of February, May, August and November. For the verification of
3-month mean seasonal forecasts a time-lagged 22 member ensemble has been created by
combining the 1<sup>st</sup> of the month forecast with the 25<sup>th</sup> of the month prior (e.g. 11 members from
25<sup>th</sup> Jan combined with 11 members from 1<sup>st</sup> Feb). These dates allow analysis of the
performance of forecasts of the 4 standard seasons at a 1 month lead time (i.e. DJF, MAM, JJA,
SON). A time-lagged ensemble is not used for verification of the multi-week forecasts (i.e. the
11 member ensemble from the 1<sup>st</sup> and 25<sup>th</sup> of the months are used as independent forecasts). For
forecasts on the multi-week timescale the skill of forecasting fortnight 1 (weeks 1+2), fortnight
2 (weeks 2+3) and fortnight 3 (weeks 3+4) of the forecast is assessed.

Although a 23-year hindcast period is significantly longer than the UKMO's 14-year hindcast
period, it is still regarded as too short for obtaining a robust estimate of the skill for forecasts on
the seasonal timescale. A long enough hindcast set is necessary for including sufficient cases of
the low frequency influences (climate drivers) on Australian climate, like the El Niño Southern
Oscillation and for knowing how that skill may vary based on the state of these climate drivers.
In addition, a sample size of 23 is considered small for obtaining statistically robust results. For
the next version of ACCESS-S (version 2) the hindcast period will be increased to 30+ years
(this is possible because the Bureau will be implementing its own data assimilation system in
version 2 and will not be reliant on the UKMO initial conditions, as it is for version 1). In this
report, sampling issues due to the short hindcast period analysed are addressed by indicating
which correlations are statistically significant given the sample size.
3. SKILL EVALUATION

To assess the performance of the forecasts, correlation of the ensemble mean forecast anomaly with the observed anomaly is used. It is common-practice in seasonal prediction to bias-correct the forecasts and assess anomalies (e.g. Stockdale, 1997). Forecasts of rainfall, Tmax and Tmin anomalies over Australia are evaluated. These anomalies are created by producing a climatology from the ensemble mean hindcasts that is a function of both start date, lead time and grid box. The climatology is then subtracted from a given ensemble mean forecast to produce the forecast anomalies, and in so doing a first-order linear correction for model bias or drift is made. The forecasts are compared with the observations, as provided by the 0.25° resolution Australian Water Availability Project (AWAP) gridded datasets (Jones et al. 2009), and observed anomalies are calculated using the AWAP climatology for the 23 year period assessed (1990-2012).

Performance of the forecasts are assessed for the whole of Australia (Figures shown in Appendix A), but the focus of this report is the evaluation over areas of interest to vegetable growers. Nine regions have been selected in consultation with the horticultural industry (specifically Will Gordon, Industry Services Manager, Horticulture Innovation Australia) (Figure 1). These regions have been defined on the 0.25° (~25 km) AWAP observations grid (Jones et al. 2009). Given that ACCESS-S1 has grid boxes larger than 25 x 25 km, the model grid boxes are weighted depending on their contribution to a given region. These weights are used when creating the average skill for a given region.

The performance of tercile forecasts (i.e., 3-category forecasts: drier/colder, neutral or wetter/hotter conditions than normal) for the horticulture regions are also assessed. The details of this assessment are provided in section 3.3.

![Figure 1: The nine horticulture regions analysed in the skill evaluation.](image)

3.1 Seasonal forecasts

Performance of the seasonal forecasts is evaluated for two lead times: 0-month and 1-month leads. However, given the start dates available for analysis, the seasons analysed are slightly different. The 0-month lead seasonal forecasts are evaluated for the seasons Feb-Mar-Apr
This means that the 22 member forecast ensemble is initialised at the start of the season (i.e., the FMA forecast is initialised using 11 members from the 25th January and 11 members from the 1st February). The 1-month lead forecasts are evaluated for the seasons Mar-Apr-May (MAM), Jun-Jul-Aug (JJA), Sep-Oct-Nov (SON) and Dec-Jan-Feb (DJF) and are initialised one month prior to the verification season (i.e., the MAM forecast ensemble is initialised on the 25th January and the 1st February).

### 3.1.1 Rainfall

The correlation skill over Australia for the 0-month and 1-month lead seasonal forecasts is shown in Figures A1 and A2 respectively. Given the relatively small sample size (n=23) only correlations greater than 0.4 can be considered statistically significant (at the 95% confidence level). Overall, the spring time of year (ASO and SON) has the most skill and summer (NDJ and DJF) the least, for both lead times. This is, however, not the case for all regions, which will become apparent when focussing on the horticulture regions. The El Niño Southern Oscillation (ENSO) and the Indian Ocean Dipole (IOD) are key climate drivers that provide predictability over eastern and southern Australia, particularly in winter and spring (e.g., Risbey et al. 2009; Langford et al., 2011). ACCESS-S1 produces skilful forecasts of the ENSO and IOD indices from August (correlations exceed 0.9 for the NINO3 index and 0.6 for the IOD over the first 4 months of the forecast), and these indices are more skilfully predicted at this time of year than any other (Lim et al., 2016). However, skilful prediction of regional climate requires not only the skilful prediction of the climate drivers, but also an accurate depiction of their relationship, or teleconnection, to Australian climate. Lim et al. (2016) found that the teleconnection strength of ENSO and the IOD with Australian rainfall respectively, is substantially weaker in ACCESS-S1 than observed. Furthermore, the teleconnection between the Southern Annular Mode (SAM) and south-eastern Australian rainfall is deficient in early winter to winter (Lim et al. 2016). The latter will affect the skill of the rainfall forecasts for the Mildura, Werribee and Adelaide regions in particular. Therefore, although the skill for forecasting spring rainfall is reasonably good (Figures A1 and A2), there is still room for improvement (Lim et al. 2016).

Figure 2 shows how the correlation skill for rainfall differs by season and region for the nine horticulture regions. Skill generally decreases as the forecast lead time increases. However, differences between the 0-month (Figure 2 top row) and 1-month (Figure 2 bottom row) lead forecasts also reflect the slightly different seasons being verified. For the eastern (Bowen, Bundaberg, Sydney) and south-eastern (Werribee, Mildura, Devonport, Adelaide) regions, spring (SON) is clearly the most skilful time of year for forecasts of rainfall. For the two western Australian regions, there is some skill for forecasting early spring (ASO) rainfall at short lead times, particularly for Carnarvon, but not spring (SON) rainfall at longer leads. The Bowen region exhibits the best skill of all the regions for forecasts of early spring (ASO) and spring (SON) rainfall; although these forecasts may be of limited use since rainfall totals are low at that time of year.

The Carnarvon region, in north-west Western Australia, has reasonable skill (statistically significant) for the short-lead predictions of all the seasons (i.e., FMA, MJJ, ASO, NDJ). Carnarvon exhibits good skill (r>0.5) for forecasts of early summer (NDJ) rainfall, as well as summer (DJF) rainfall at the longer lead time. Bowen, in the north-east, also has reasonably skilful forecasts of early summer rainfall (NDJ), but the skill drops off markedly for forecasts of summer rainfall (DJF) at the longer lead time. For early winter (MJJ) and winter (JJA) rainfall the skill is in general poor, apart from the early winter (MJJ) forecast for Carnarvon. Forecasts of early autumn (FMA) rainfall at short lead times are skilful for Carnarvon and reasonably
skilful over the southern regions of Sydney, Mildura and Adelaide. The forecast of autumn (MAM) rainfall at 1-month lead is skilful over Bowen.

Figure 2: Correlation skill for 0-month lead (top) forecasts of seasonal mean rainfall anomalies for the FMA (red), MJJ (blue), ASO (green) and NDJ (orange) seasons; and for 1-month lead (bottom) forecasts of the MAM (red), JJA (blue), SON (green) and DJF (orange) seasons. Skill is shown for the horticulture regions (Figure 1), indicated from left to right: Bowen, Bundaberg, Sydney Basin, Werribee/Cranbourne/Gippsland; Mildura; Devonport; Adelaide Plains; south western WA and Carnarvon. Statistically significant correlations exceed 0.4 (at the 95% confidence level, n=23).

3.1.2 Maximum temperature

Seasonal forecasts of Tmax are usually more skilful than forecasts of rainfall. This can be seen in general from the Australia-wide maps (Figures A3 and A4) and for the horticulture regions (Figure 3), particularly for the 0-lead forecasts.
Figure 3: As for Figure 2, but for Tmax.

There is notably poor (negative in some areas) skill over eastern Australia for Tmax seasonal forecasts of early winter (MJJ) (Figure A3) and winter (JJA) (Figure A4). This impacts the eastern horticulture regions (Bowen, Bundaberg and Sydney), which exhibit no skill in these seasons (Figure 3). This poor performance is largely related to the UKMO using climatological land surface initial conditions (soil moisture) to initialise the forecasts, rather than realistic initial conditions. We ran an experiment to test the impact of initialising the land surface with realistic initial conditions. The details of the experiment are provided in Appendix B, and will be reported on fully in a separate report (Zhao et al., in preparation). The results show that the skill for Tmax forecasts for May (the first month of the forecast) are significantly improved when the land surface is initialised with realistic initial conditions and that the largest impact is seen over eastern Australia (Figure B1). For the next version of ACCESS-S (version 2), we will not be reliant on the UKMO initialisation strategy, instead implementing the Bureau's data assimilation, whereby the land surface will be initialised with realistic initial conditions. The
experiment described above suggests that this will significantly improve the skill of Tmax forecasts in the winter half of the year.

Forecasts of early summer (NDJ) Tmax exhibit good skill for all the horticulture regions (all regions have statistically significant skill and $r>0.6$ in 5 regions) (Figure 3). However, the skill drops away markedly for the longer lead time forecasts for summer (DJF), particularly for the southern regions (Werribee, Mildura, Devonport, Adelaide). In contrast, the longer lead spring forecasts (SON) are more skilful than the shorter lead early spring (ASO) forecasts (Figure 3). All regions exhibit skill for seasonal forecasts of spring (SON) season Tmax at a 1-month lead time, with the eastern regions (Bowen, Bundaberg, Sydney) showing highest skill.

### 3.1.3 Minimum temperature

In general, the skill for Tmin (Figures A5 and A6) is less than that for Tmax.

For the 0-month lead seasonal forecast of Tmin, the northern regions (Bowen, Bundaberg, Carnarvon) have good skill ($r>0.6$) for forecasts of early spring (ASO) (Figure 4 and Figure A5). This season is less skilful for the more southern horticulture regions (apart from Devonport). Similarly, the Bowen region, and to a lesser degree the Bundaberg and Carnarvon regions, also have reasonable skill for the early winter (MJJ) season (Figure 4). Skill for this season is particularly poor for the south-eastern mainland regions (Sydney, Werribee, Mildura, Adelaide). The climatology that the UKMO use to initialise soil moisture (rather than realistic initial conditions) does not have as much negative impact on the skill of Tmin as it does for Tmax, but there is nonetheless an impact, particularly over south-eastern Australia (Zhao et al., in preparation).

In general, the southern horticulture regions exhibit good skill for Tmin forecasts of early summer (NDJ) and to a lesser degree early autumn (FMA) (Figure 4 and Figure A5). For forecasts of early autumn (FMA), the south eastern regions (Sydney, Werribee, Mildura, Devonport, Adelaide) exhibit the most skill.

The skill tends to drop away sharply for 1-month lead seasonal forecasts of Tmin compared to the 0-month lead forecasts (Figure 4 and Figure A6). For the 1-month lead forecasts there is good skill ($r>0.5$) for forecasts of winter (JJA) for the north eastern regions (Bowen, Bundaberg). With the possible exception of Devonport, forecasts of winter (JJA) Tmin are poor for the other horticulture regions (Figure 4). There is good skill ($r>0.5$) for forecasts of summer (DJF) Tmin for the western regions (south western WA, Carnarvon) and the Sydney region (Figure 4 and Figure A6). The skill is generally low for forecasts of autumn (MAM) Tmin.
3.2 Multi-week forecasts

The correlation skill for the whole of Australia for forecasts of fortnightly mean rainfall, Tmax and Tmin anomalies for fortnight 1 (week 1+2), fortnight 2 (weeks 2+3) and fortnight 3 (weeks 3+4) are shown in Appendix A (Figure A7), using all available forecast start dates. There are statistically significant positive correlations Australia-wide for Tmax and Tmin at all lead times, and for rainfall for the first two fortnights (except for isolated small regions in fortnight 2). The correlation skill for rainfall for fortnight 3 (weeks 3+4) of the forecast is not statistically significant over large parts of eastern Australia (Figure A7). Multi-week skill for Tmax and Tmin is generally higher than that for rainfall at all lead times. There is a clear reduction in skill with forecast lead time, i.e. going from fortnight 1 to fortnight 3.

The multi-week skill for the horticulture regions is shown for rainfall, Tmax and Tmin in Figures 5, 6 and 7 respectively. However, for each variable, instead of showing the overall skill for all start dates, as in Figure A7, we show the skill for forecast start dates in the summer half
of the year (i.e., forecasts initialised on the 25 Oct, 1 Nov, 25 Jan, 1 Feb) and the winter half of the year (i.e., forecasts initialised on the 25 Apr, 1 May, 25 Jul, 1 Aug) separately. For rainfall, the multi-week forecasts are more skilful in the winter half of the year compared to the summer half of the year in virtually all the regions (Figure 5).

For Tmax, there are high correlations for all regions and both halves of the year for the first and second fortnights and all are statistically significant, including for fortnight 3 (weeks 3+4) (Figure 6). Correlations for fortnight 1 (weeks 1+2) mostly exceed 0.7. For fortnight 3 (weeks 3+4) the correlations are higher in the summer half of the year compared to the winter half for all regions except Carnarvon (Figure 6). This is likely due to the use of climatological land surface initial conditions, which had a detrimental impact on forecasts in the winter half of the year, as discussed in Section 3.1.2. The experiment conducted to test the impact of the land surface initialisation (Appendix B) found that if realistic land surface initial conditions are used, then the benefit in forecast performance for Tmax becomes evident from week 4 onwards (Zhao et al., in preparation). In the first couple of weeks of the forecast, the atmosphere initial conditions are likely dominating the predictability of the forecast, but as the lead time increases the impact of the land initial conditions become more important.

The Tmin multi-week forecasts show an interesting seasonality in performance. The south eastern regions (Sydney, Werribee, Mildura, Devonport, Adelaide) have good skill for all three lead times during the summer half of the year, but have relatively lower skill in the winter half of the year (Figure 7). As mentioned in Section 3.1.3, the poorer skill in winter is likely due, in part, to the unrealistic initialisation of soil moisture. Apart from the two north-eastern regions (Bowen and Bundaberg) the skill for forecasting fortnight 3 (weeks 3+4) is higher in the summer half of the year. In contrast, the north-eastern regions (Bowen and Bundaberg) are more skilful in the winter half of the year at all lead times.
Figure 5: Correlation skill for multi-week rainfall anomalies for the summer half of the year (i.e., forecasts initialised on the 25 Oct, 1 Nov, 25 Jan, 1 Feb) (top) and the winter half of the year (i.e., forecasts initialised on the 25 Apr, 1 May, 25 Jul, 1 Aug) (bottom). Skill is shown for fortnight 1 (weeks 1+2; Red), fortnight 2 (weeks 2+3; Blue), and fortnight 3 (weeks 3+4; Green) of the forecast for each of the horticulture regions (Figure 1), indicated from left to right: Bowen, Bundaberg, Sydney Basin, Werribee/Cranbourne/Gippsland; Mildura; Devonport; Adelaide Plains; Southern Western Australia and Carnarvon. Statistically significant correlations exceed 0.2 (at the 95% confidence level, n=92)
Figure 6: As for Figure 5, but for Tmax.
In the sections above, we verified forecasts of the ensemble mean anomaly using a simple metric, the correlation. There are other types of multi-week or seasonal forecasts that we could have verified, for example probabilistic forecasts (such as the probability of having above normal rainfall) or category forecasts (yes/no for falling in a particular category, like above normal rainfall), and each of these would require a different verification metric. Typically, we find that the choice of verification metric does not change the overall conclusion of which regions or seasons have skill (e.g. the regions that show good correlation skill are also the regions that show good hit rates or good Relative Operating Characteristic scores). However, in this section we show an alternative view of forecast performance, since it is useful for highlighting some additional concepts (as suggested in Abawi et al. 2008).

3.3 Performance of three-category seasonal forecasts

Figure 7: As for Figure 5, but for Tmin.
Here, the forecast is expressed in tercile categories i.e., the probability that rainfall (or Tmax/Tmin) will be below-normal, normal or above normal. For example, a forecast for a particular region and season may be as follows: a 20% chance the season will be drier than normal, a 30% chance it will be normal and a 50% chance it will be wetter than normal. Because this is a probabilistic forecast, all three outcomes are possible, but in practice, a users’ expectation is that the most likely outcome is the tercile with the highest probability. If the observed rainfall was in the same tercile as the tercile with the highest probability, then it is termed a consistent forecast, as shown in the schematic in Figure 8. If the observed rainfall was in the neighbouring category, then it is a near consistent forecast and if it is two categories away, then it is an inconsistent forecast (e.g., if the highest probability forecast category is for below-normal rainfall, but above-normal rainfall was observed) (Figure 8).

Figure 8: Schematics to describe the tercile scoring used in Figures 9-12. Panel a) indicates a hypothetical time series of tercile forecasts and observations for rainfall for a given region (it applies equally to temperature, except the y-axis would read "colder than normal" and "hotter than normal") (after Abawi et al. 2008). "O" denotes the tercile with the observed value of rainfall e.g., in 1990 the rainfall was wetter than normal and in 1992 it was drier than normal. "P" denotes the tercile with the highest forecast probability e.g., the model predicted that it would most likely be wetter than normal in 1990 and most likely be normal in 1992. The year is shaded in blue if the observed rainfall was in the same tercile as the tercile with the highest probability (consistent forecast). Red indicates that the observed rainfall was different by two categories from the most probable forecast category (inconsistent forecast) and green indicates that the observed was in the neighbouring category (near consistent forecast). This colour scoring is summarised in the table in b).

Although the conversion of a probabilistic forecast to a category forecast is often how a probabilistic forecast is interpreted, it is not desirable. The simple attribute of forecast accuracy, such as "how often are the forecasts correct" is not actually applicable to a probabilistic forecast. Translating the probability forecast into a categorical forecast, as done in this section, can be misleading, and lot of potentially useful forecast information is lost when doing the conversion to a category forecast. For example, in the case of the following tercile probabilities - T1 30%, T2 33%; T3 37% - the top tercile (T3) has the highest probability, but only marginally i.e., there are not strong odds for the top tercile. If the observations fell into the bottom tercile, then by the accuracy metric it would be an inconsistent forecast (or a "miss"), yet clearly it is not necessarily a bad forecast. There are other forecast verification metrics that take into account the issued probabilities (such as Relative Operating Characteristic scores and the Brier score), but they are more complex to convey to users. Bearing in mind the caveats noted above, the advantage of the analysis which follows, is that it is simple to understand, useful for conveying variation in forecast performance over time and can be used to highlight potential inconsistent forecasts (for further investigation) which could have resulted in negative impacts for the industry if action had been taken on their advice.
Figure 9 shows a time series of consistent (blue), near consistent (green) and inconsistent (red) forecasts for each of the horticulture regions for forecasts of the early spring season (ASO) at 0-month lead time for 1990-2012. The red squares highlight those years and regions where the forecast was inconsistent, and therefore times when the user may have made a decision that could have resulted in losses (e.g., money, inputs, yield) or lost opportunities. Figure 9 also demonstrates that forecast performance (and hence "usefulness") is not uniform across all years. For example, there are clusters of successive years where performance is good (e.g., 2006-2010) and clusters of years where performance is not as good (2000-2003). Some years have virtually all consistent forecasts across the horticulture regions (e.g. 1994, 2006), whereas in other years there is a high proportion of near-consistent or inconsistent forecasts across the regions (e.g. 1993, 2001). Figure 9 also highlights the relatively small number of forecasts (23 years) upon which the forecast performance is being judged.

As per Figure 9, graphs were made for seasonal forecasts of the other 0-month lead seasons and variables (Tmax, Tmin) (not shown) and their outcomes are summarised in Figures 10-12. For each horticulture region, Figures 10-12 show the percentage of the 23 years that have consistent (blue), near consistent (green) and inconsistent (red) forecasts for a given season for rainfall, Tmax and Tmin respectively. To have skill better than a climatological/random forecast, the consistent forecasts (or "hits"; blue bars) need to exceed 33.3% (i.e., over a long period of time you would expect 33.3% hits by chance if you made random forecasts – blue squares in Figure 8b i.e., 3 divided by 9, given that each box has a 33.3% chance of occurring) and the inconsistent forecasts need to be less than 22.2% (i.e. over a long period of time you would expect 22% inconsistent forecasts by chance if you made random forecasts - red squares in Figure 8b i.e., 2 divided by 9). Refer to Wilks (2006) for a discussion on how to calculate proportion correct for multi-category forecasts. Ideally, the proportion of consistent forecasts
(blue bars) will be large, at the expense of the proportion of near-consistent (green bars) and inconsistent (red bars) forecasts.

For rainfall, as was concluded in the evaluation of performance using correlation (Section 3.1.1) the early spring (ASO) season is the most skilful for most of the regions (Figure 10). In this season all regions have skill better than a climatological forecast, with relatively high proportions of consistent forecasts (>33.3%, with more than half the regions having at least 50%) and relatively low proportions of inconsistent forecasts (for most regions it is around 10%).

For Tmax, the poor forecast performance in terms of correlation (Section 3.1.2) that was seen over the north-eastern Australian regions (Bowen, Bundaberg) for the forecast of early winter (MJJ) is also clear for the categorical forecasts (Figure 11). The percentage of consistent forecasts is low (< 33.3%) and the percentage of inconsistent forecasts is high (> 22.2%) (Figure 11). As was also apparent for the correlation skill, the forecasts of the early summer season (NDJ) are good for most regions, with high hit rates (consistent forecasts) and generally low percentages of inconsistent forecasts.

For Tmin, as was shown in Section 3.1.3, the skill of forecasts of the early winter (MJJ) season for the south-eastern mainland regions (Sydney, Werribee, Mildura, Adelaide) is poor (Figure 12). In fact, forecasts for MJJ for these regions are the only instances of any region, season or variable (apart from Bowen Tmax for MJJ) where the number of inconsistent forecasts exceeds the number of consistent forecasts. This again relates back to using climatology to initialise the soil moisture, as mentioned in Section 3.1.3.

As shown in correlation analysis for Tmin, the northern regions (Bowen, Bundaberg, Carnarvon) have good skill for forecasts of early spring (ASO) for the 3-category forecasts, with a high percentage of consistent forecasts (>50%). Similarly, the south-eastern horticulture regions (Sydney, Mildura, Devonport, Adelaide) exhibit good skill for the early autumn (FMA) season (Figure 12).
Figure 10: The percentage of the 23 years that have consistent (blue), near consistent (green) and inconsistent (red) forecasts for 0-month lead forecasts of rainfall for the seasons indicated on the panels. The horticulture regions are indicated from left to right: Bowen, Bundaberg, Sydney Basin, Werribee/Cranbourne/Gippsland; Mildura; Devonport; Adelaide Plains; south western WA and Carnarvon.
Figure 11: As for Figure 10, but for Tmax.
4. DISCUSSION AND CONCLUSIONS

ACCESS-S1 will be the next version of the Bureau of Meteorology's seasonal prediction system, due to become operational in 2017. In this report we have evaluated the performance of ACCESS-S1 for forecasts of Australian climate on seasonal and multi-week timescales, focusing on 9 key regions of importance to the horticultural industry.

For the seasonal forecasts, we have evaluated 3-month mean rainfall, Tmax and Tmin forecasts for 0-month and 1-month lead times. The 0-month lead forecasts verify in the early autumn (FMA), early winter (MJJ), early spring (ASO) and early summer (NDJ) seasons. The 1-month lead forecasts verify in the autumn (MAM), winter (JJA), spring (SON) and summer (DJF) seasons. For all three variables, the skill generally decreases as forecast lead time increases (i.e., 0-month lead compared to 1-month lead). The performance for the horticulture regions is summarised as follows.

For forecasts of *seasonal mean rainfall*:

- For the eastern (Bowen, Bundaberg, Sydney) and south-eastern (Werribee, Mildura, Devonport, Adelaide) regions, forecasts of spring rainfall (ASO and SON) are generally skilful and exhibit more skill than other seasons.
- For the two western Australian regions (Carnarvon and south-western WA), there is some skill for forecasting early spring (ASO) rainfall at short lead times, particularly for Carnarvon, but less so for spring (SON) rainfall at longer leads.

![Seasonal Tmin: 0-month lead time](image-url)
The Carnarvon region, in the north-west, has good skill for the short-lead predictions of all the seasons (i.e., FMA, MJJ, ASO and NDJ).

For summer rainfall, the two northern regions of Carnarvon (north-west) and Bowen (north-east) exhibit good skill for forecasts of early summer (NDJ) rainfall. Carnarvon also has good skill for summer (DJF) rainfall at the longer lead time, but the skill for summer (DJF) rainfall for Bowen drops off markedly.

The skill for forecasts of early winter (MJJ) (at 0-month lead) and winter (JJA) (at 1-month lead) rainfall is in general poor, apart from the early winter (MJJ) forecast for Carnarvon.

Forecasts of early autumn (FMA) rainfall at short lead times are skilful for Carnarvon and reasonably skilful over the over southern regions of Sydney, Mildura and Adelaide. The forecast of autumn (MAM) rainfall at 1-month lead is skilful over Bowen.

For forecasts of seasonal mean Tmax:

- In general Tmax forecasts are more skilful than rainfall
- All regions exhibit skill for seasonal forecasts of the spring (SON) season at a 1-month lead time, with the eastern regions (Bowen, Bundaberg, Sydney) showing the highest skill.
- Forecasts of early summer (NDJ) exhibit good skill for all the horticulture regions. However, the skill drops away markedly for the longer lead time forecasts for summer (DJF), particularly for the southern regions (Werribee, Mildura, Devonport, Adelaide).
- There is notably poor (negative) skill for the eastern horticulture regions (Bowen, Bundaberg and Sydney) for forecasts of early winter (MJJ) and winter (JJA) Tmax.

For forecasts of seasonal mean Tmin:

- In general, the skill for Tmin is less than that for Tmax.
- The skill tends to drop away sharply for 1-month lead seasonal forecasts of Tmin compared to the 0-month lead forecasts.
- The northern regions (Bowen, Bundaberg, Carnarvon) have good skill for short lead forecasts of early spring (ASO). This season is less skilful for the more southern horticulture regions.
- The northern regions (Bowen and to a lesser degree the Bundaberg and Carnarvon) have reasonable skill for short lead forecasts of early winter (MJJ). Skill for this season is particularly poor for the south-eastern mainland regions.
- The southern horticulture regions exhibit good skill for forecasts of early summer (NDJ) and early autumn (FMA).
- For the 1-month lead forecasts there is good skill for forecasts of winter (JJA) for the north eastern regions (Bowen, Bundaberg) and for summer (DJF) for the western regions (south western WA, Carnarvon) and the Sydney region.

The evaluation of the multi-week forecasts focusses on the skill of fortnightly mean rainfall, Tmax and Tmin anomalies for fortnight 1 (week 1+2), fortnight 2 (weeks 2+3) and fortnight 3 (weeks 3+4) of the forecast.

For forecasts of multi-week rainfall, Tmax and Tmin:

- The skill in fortnight 1 is high for all regions and variables.
- There is a clear reduction in skill with forecast lead time, i.e. going from fortnight 1 to 3.
- Beyond fortnight 1, the multi-week forecasts of rainfall are more skilful in the winter half of the year compared to the summer half of the year in virtually all the regions. In contrast, for Tmax and Tmin the skill tends to be higher in the summer half of the year compared to the winter half, particularly for the southern regions for Tmin and the eastern regions for Tmax.
The skill for the Tmax multi-week forecasts in particular, is high in all regions and lead times.

There is still scope for improvement in the skill of the ACCESS-S1 forecasts. Some of the identified shortfalls relevant to this study include:

- weak/deficient teleconnections between ENSO, the IOD and the SAM and Australian rainfall over eastern and south eastern Australia respectively, particularly in the winter half of the year (Lim et al 2016).
- initialisation of soil moisture with climatology, rather than realistic soil initial conditions that vary year-to-year. Initialising using climatology impacts negatively on the skill of the multi-week (from week 4) and seasonal forecasts, particularly for Tmax over eastern Australia and to a lesser degree on Tmin over south-eastern Australia for forecasts in early winter and winter. The next version of ACCESS-S (version 2) will not be reliant on the UKMO initialisation strategy; instead the Bureau's data assimilation will be used, whereby the land surface will be initialised with realistic initial conditions.

A caveat of the current study is the relatively small sample size (n=23) used for the evaluation of forecast performance, particularly for the seasonal forecasts. Whilst 23 years is better than the UKMO's 14-year hindcast period (Shi et al. 2016), it is still too short to obtain statistically reliable results, particularly for aspects of the prediction system that are inherently noisy, such as for forecasts of relatively small regions of Australian climate (in contrast to less noisy, more large-scale aspects like ENSO). In addition, a long enough hindcast set is necessary for including sufficient cases of the low frequency influences on Australian climate, like the different phases of ENSO, and for knowing how the skill may vary based on the state of these climate drivers. The size of the hindcast for ACCESS-S1 is limited due to the availability and reliance on the UKMO's initial conditions. However, for ACCESS-S2 we plan to create a longer hindcast set, spanning at least 30 years.

The evaluation of forecast performance has focused on key regions of importance to vegetable growers. Going forwards, to get optimal use out of climate forecasts, there is a need to partner with the vegetable industry to determine both the level of skill that is useful and what management decisions could be made on the basis of the forecasts at their given level of skill. For each region and crop there may be different climate related risks and management decisions at different times of the year. The key question is whether the forecasts lead to better decision-making.

**ACKNOWLEDGEMENTS**

This project has been funded by Horticulture Innovation Australia Limited with co-investment from the Bureau of Meteorology and funds from the Australian Government (Project code VG13092). We would like to thank Andrew Marshall and Eunpa Lim for their useful reviews of this report.

**REFERENCES**


Weedon, G. P. and Coauthors, 2011: Creation of the WATCH Forcing Data and its use to assess global and regional reference crop evaporation over land during the twentieth century. *J. Hydrometeor.*, 12, 823–848


APPENDIX A: SKILL MAPS FOR AUSTRALIA

Figure A 1: Correlation skill for 0 month-lead forecasts of seasonal mean rainfall anomalies for the FMA, MJJ, ASO and NDJ seasons. The dashed line indicates the correlation value above which the correlations are statistically significant (i.e. r>0.4 is significant, 95% confidence level, n=23).

Figure A 2: Correlation skill for 1 month-lead forecasts of seasonal mean rainfall anomalies for the MAM, JJA, SON and DJF seasons. The dashed line indicates the correlation value above which the correlations are statistically significant (i.e. r>0.4 is significant, 95% confidence level, n=23).
Figure A 3: As for Figure A1, but for Tmax.

Figure A 4: As for Figure A2, but for Tmax
Figure A 5: As for Figure A1, but for Tmin

Figure A 6: As for Figure A2, but for Tmin
Figure A 7: Correlation skill for the fortnights comprising weeks 1 and 2 (left column), weeks 2 and 3 (middle column) and weeks 3 and 4 (right column) for rainfall (top row), Tmax (middle row) and Tmin (bottom row) using all of the 8 forecast start dates noted in section 2.2. The dashed line indicates the correlation value above which the correlations are statistically significant (i.e. $r>0.14$ is significant, 95% confidence level, $n=184$). If there is no dashed line shown on a map, then all the correlations are significant.
APPENDIX B: LAND SURFACE INITIAL CONDITIONS EXPERIMENT

A sensitivity experiment was conducted to assess the impact of land surface initialisation on forecast skill.

In the UKMO seasonal prediction system and in ACCESS-S1, the soil moisture is initialised with a monthly climatology of a land surface reanalysis using the JULES model (run in stand-alone mode) forced with the Integrated Project Water and Global Change Forcing Data methodology applied to ERA-Interim (Dee et al. 2011) data (Weedon et al. 2011, MacLachlan et al. 2014). The soil temperature and snow cover are initialised with time-varying initial conditions from ERA-Interim (i.e., not a climatology).

The sensitivity experiment aimed to determine the impact on forecast skill of initialising using time varying soil moisture. To this aim, the stand-alone version of JULES (the same as that used in the coupled ACCESS-S1 model) was run forced by bias corrected 3-hourly ERA-interim data (2m temperature, 2m specific humidity, downwards longwave radiation, downwards shortwave radiation, surface pressure, total precipitation, 10m wind speed) to obtain a land reanalysis (1990-2014, with a 10-year spin-up period starting in 1980). The forcing precipitation data is a hybrid between ERA-Interim and monthly GPCP version 2.2 data (Adler et al. 2003; Huffman et al. 2009), following the method described by Zhao and Dirmeyer (2003). The time varying fields of soil moisture from this land reanalysis were used to initialise the forecasts of the sensitivity experiment.

The sensitivity experiment was for start dates on the 1st May for 1990-2012 and an 11-member ensemble was used, as in ACCESS-S1. The results of this experiment were compared to the ACCESS-S1 forecasts for the same period and start dates. Figure B1 shows that the skill of maximum temperature forecasts for May (the first month of the forecast) are significantly improved when the land surface is initialised with realistic initial conditions and that the largest impact is seen over eastern Australia.

![Figure B1: Correlation skill for Tmax for 0-month lead forecasts of the month of May (1990-2012) from (a) the sensitivity experiment where the forecasts are initialised with realistic soil moisture conditions, and (b) ACCESS-S1, where the forecasts are initialised using climatological soil moisture.](image-url)