# What makes a 'good' seasonal forecast? Delivering actionable climate outlooks for grains farming

Patrick J. Mitchell[1], Jaclyn N. Brown[1]

[1] CSIRO Agriculture and Food, College Rd, Sandy Bay, Tasmania, 7005, Patrick.Mitchell@csiro.au

## Abstract
The prediction of climate patterns and weather conditions at the farm scale represents an important innovation for managing within season and year-to-year variability in crop production. Assessing skill and potential value of long-range, seasonal climate forecasts hinges on answering the fundamental questions: "Should I use this forecast when making my decision and how 'good' is it?" Here, we use model output from the new seasonal forecasting system, ACCESS-S1 to compare forecast approaches for deriving relevant and credible seasonal climate information for Australia's cropping regions. This evaluation addresses the role of two important components: categorisation of the model output and anchoring the forecast using antecedent conditions (fallow season rainfall). Overall, the model had relatively low accuracy at predicting correct forecasts across much of the forecast locations and seasons, whereas it had greater skill in the avoidance of false alarms i.e. false negative outcomes. The percentile categories used to derive the expected forecast had a large effect on the skill in terms of the rate of false alarms and the choice of categories can be matched to user requirements of both accuracy and resolution. Anchoring rainfall forecasts on antecedent conditions can reduce false alarms across the growing season and may be a useful guide when presented alongside a forecast based solely on in-season predicted rainfall. The next generation of climate data products and services for agriculture need to consider how a forecast system interacts with both on-farm biophysical drivers of yield and decision-making preferences of the user.

## Key Words
Grains, seasonal climate, climate forecast, decision-making

## Introduction
Seasonal climate outlooks have long offered the promise of de-risking agricultural decisions across a range of enterprises. However, the inherent challenges of providing locally relevant and accurate forecasts has meant adoption is patchy and slow (Hayman et al., 2007). Assessing skill and potential value of a particular model to predict rainfall for broad acre cropping is complex. Answering the fundamental questions: "Should I use this forecast when making my decision and how 'good' is it?" requires an understanding of the biophysical context within which a forecast is received and the decision-making behaviours that dictate a producer's uptake and response to this information. For example, social research showed that 30-50% of farmers considered seasonal climate forecasts as an important factor for farm management decision-making, yet challenges arising from its accuracy, communication and local-scale relevance to the user can often present significant barriers to adoption (Hayman et al., 2007).

Recent advancements in seasonal forecasting capabilities includes the new system from the Bureau of Meteorology (the Bureau): ACCESS-S1 (Australian Community Climate and Earth-Systems Simulator seasonal prediction system version 1) (Hudson et al., 2017). This new forecast system has the potential to provide forecasts at greater spatial and temporal resolutions and accuracy over previous seasonal climate models (Hudson et al., 2017).

Here we provide an assessment of different aspects of generating forecasts that are critical to providing relevant and credible information for grains producers. We evaluate the forecast system in terms of the influence of two important components of forecast generation and model skill. 1) Categorisation – identifying the consequences of applying different probabilistic categories to derive a forecast. 2) Anchoring – assessing forecast outcomes of the growing season forecast by incorporating antecedent rainfall i.e. fallow season rainfall used as a proxy for stored soil water.

## Methods
*Seasonal climate model*

This study assesses the new seasonal forecasting capability developed by the Bureau based around a coupled-model seasonal prediction system called ACCESS-S1 (Hudson et al., 2017). This forecast system will replace the earlier POAMA system (Hudson et al. 2013) that has been in operation at the Bureau since 2002. Full details of ACCESS-S1 are provided in (Hudson et al., 2017).

The model analysis is based on the ACCESS-S1 hindcast set generated for 23 years from 1990 to 2012. Eleven ensemble members, generated on the first of each month were assessed for forecast skill between one and six months lead times. Downscaling of the native ACCESS-S1 grid resolution was done using the quantile mapping approach similar to McIntosh and Brown (2016) to generate point based forecasts across a grid with a horizontal resolution of 0.05° (approximately 5 km at mid-latitudes).
A representative sample of 52 locations across the dryland cropping belt were selected (Figure 2 C), to test the forecast skill for rainfall using the method of Hochman and Horan (2018). These locations span the climatic and environmentali gradients across the three major grain growing regions of Australia (Grains Research and Development Corporation, 2018). Weather station data managed by the Bureau were extracted at these locations and used as observations to assess model skill.

*Model analysis and forecast approach*

The verification of the model skill for the 52 study locations was done using a categorical approach where a forecast was defined when >50% of model ensembles fell into a particular category. Instances where this condition was not met was categorised as an 'inconclusive' forecast. The conclusive forecast categories included: 'correct' – forecast matches the category of the observation, 'close' – forecast is one category from the observation e.g. forecast of tercile 1, where a tercile 2 event was observed and, 'false alarm' – forecast is two categories from the observation e.g. forecast of tercile 1, where a tercile 3 event was observed. Differences in forecast skill among different forecast approaches was evaluated using the proportion of forecast outcomes within each category. Previous work identified the incidence of false alarms as being a useful metric to assess skill, given the relatively low proportion of correct forecasts across the study sites (see below). Further, we maintain that identifying forecast outcomes that are highly unlikely is a useful forecast outcome and can inform risks around potential decision strategies.
Forecast skill derived for different percentile categories was compared using three different thresholds: median (above or below 50$^{th}$ percentile), terciles (33$^{rd}$ and 66$^{th}$ percentiles) and asymmetric categories (20$^{th}$ and 66$^{th}$ percentile). The latter category was chosen because it provides a better demarcation of 'dry' and 'average' (first and second percentile categories) given the skewed distribution of rainfall across most sites and months.
The role of defining forecast categories during a typical winter growing season was assessed by comparing forecast of cumulative rainfall between April and September (forecast months April, June and August) based on categories defined with either: cumulative rain between April and September, or cumulative rain between April and September plus total rainfall during the fallow period (December to March). The fallow period rainfall was used as proxy for stored soil water at the beginning of the season. This approach is akin to applying simple grain yield equations e.g. French and Schultz (1984) to estimate yield as fallow season and in-season rainfall are the key determinants of yield for a given location.

## Results and Discussion

The tercile approach used to categorise forecasts provides a 'base case' to assess overall model skill (

Figure 1 B).Figure 2 Across the three growing regions, the model predicted the correct category less than pure chance i.e. randomly choosing any tercile, for the majority of forecast months and lead times with a mean of 30, 24, and 27% for northern, western and southern regions respectively (data not shown). Forecasts categorised as false alarms, tended to be lower than pure chance across almost every forecast month and lead time and were particularly low for spring months (

Figure 1 B). The tercile approach also produced relatively large percentage of inconclusive forecasts with a mean of 28, 38, and 33% for northern, western and southern regions respectively (data not shown).
These results provide an indication of accuracy and suggest that overall winter and spring months provide greater certainty in delivering a reliable forecast. While the relatively low percentage of correct forecasts are concerning, the forecast system presented here provided potentially useful insights into the season ahead. Identifying those conditions that are highly unlikely is a useful forecast outcome and these results indicate relatively good skill in avoiding false alarms. The inclusion of an inconclusive category can also be instructive because it informs the user of instances of low model consensus and where using the default historic climate may be more instructive.

*Categorising forecasts*

The choice of forecast categories (based on percentiles) had a large effect on the rate of false alarms across the 23 year hindcast period. The median category had a higher percentage of false alarms with a mean of 42% across seasons and growing regions (

Figure 1 A). This approach also produced a higher rate of correct forecasts (mean of 43%, data not shown). By comparison, false alarms were lower for a tercile-based forecast and were between 3 and 17% (mean of 13 %) across the different forecast seasons and growing regions (

Figure 1 B). The asymmetric forecast categories (20$^{th}$ and 66$^{th}$ percentile) yielded the lowest false alarms with a mean of 6% (Figure 2 C). This category also produced a higher percentage of correct forecasts (data not shown).

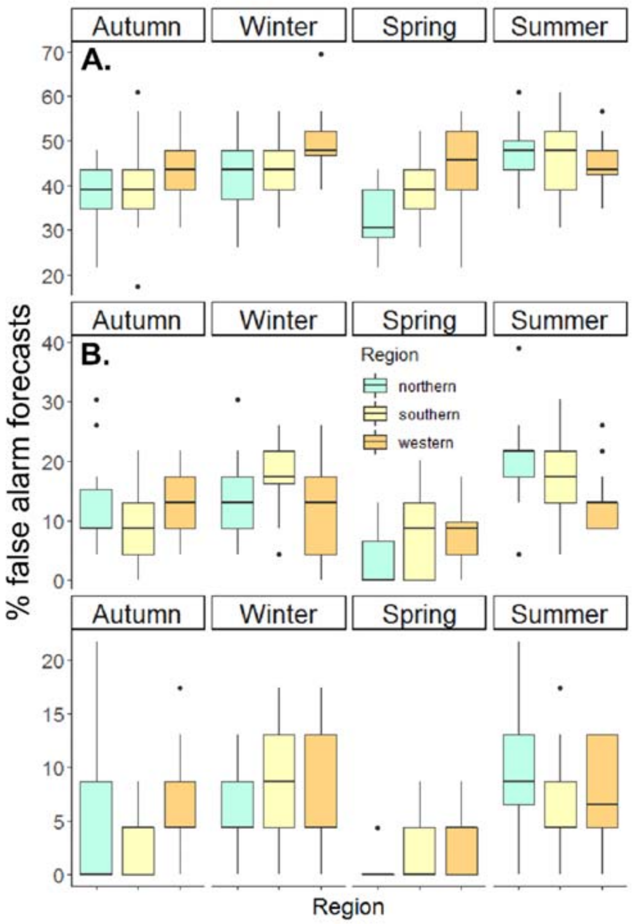


**Figure 1 Impact of different forecast categories on percentage of false alarms across the three growing regions and seasons. Forecast skill defined using A) median categories (50$^{th}$ percentile), B) tercile categories (33$^{rd}$ and 66$^{th}$ percentiles) and C) asymmetric categories (20$^{th}$ and 66$^{th}$ percentiles).**

The categorical approach used to define forecasts needs to balance considerations regarding the underlying statistical nature of the observed and forecast data with how useful the approach might be in a user's decision framework. The use of median-based categories is problematic in terms of the higher rates of false alarms and the absence of an inconclusive category. Thus the model always produces a seemingly conclusive forecast yet lacks resolution in separating seemingly average rainfall conditions with those close to the extremes. The use of terciles or the asymmetric forecast categories performed better with respect to false alarms, while the latter approach produce a spread of forecasts closer to the expected frequency (data not shown).

*Anchoring the forecast on stored soil water*

The inclusion of fallow season rainfall (December to March) as a proxy for stored soil water had a significant effect on the percentage of false alarm forecasts initiated in April (1-6 month lead times), June (1-4 month lead times) and September (1-2 month lead times) (Figure 2 A and B). Based on tercile forecast categories, forecasts with stored soil water were lower by 6-10 percentage points across the forecast months and growing regions (Figure 2 C). False alarms tended to diminish through the season reflecting the improved model skill both in terms of moving through the Autumn predictability barrier (Duan and Wei, 2013) and the tendency of cumulative forecasts to 'lock in' an outcome as the lead time reduces (Figure 2 C). This test of forecast skill is based on generalised conditions of winter season cropping patterns and obviously ignores regional differences in crop calendars and biophysical conditions. However, it provides a useful comparison of how forecasts can be interpreted and offers a preliminary assessment of skill of ACCESS-S1 for predicting crop yield using more sophisticated modelling frameworks (Brown et al., 2018).

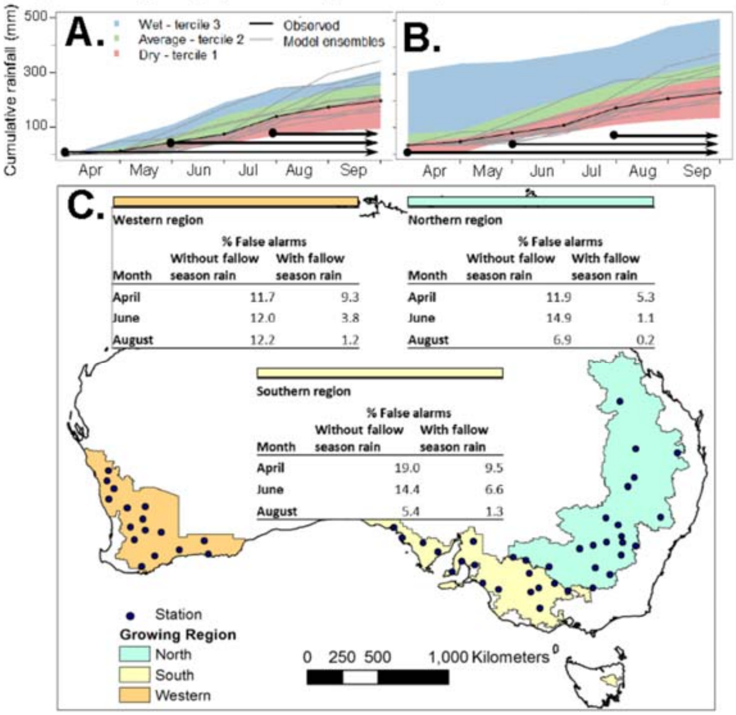


**Figure 2 The role of anchoring forecasts using fallow season rainfall (December to March) as a basis for categorising in-season rainfall outcomes. Example of an April forecast for Corrigin WA, 2007 where A) categories are based on April to September cumulative rainfall and B) categories are based on April to September cumulative rainfall plus stored soil water. Lines and arrows denote the forecast months and lead times used for the comparison in C). C) Map of 52 study locations and growing regions used for the analysis overlayed with results from the 'anchoring' comparison showing mean rate of false alarms for the two forecasting approaches for April, June and August forecast months**

## Conclusions

The new seasonal climate model system from the Bureau of Meteorology offers potential for improving climate information for grains farming provided the interpretation and communication of this somewhat complex probabilistic information can match the requirements of the users. The presentation of model skill can be defined using the likelihood of predicting potentially damaging information (false alarms) alongside identifying instances where the model offers no emphatic forecast i.e. inconclusive. Forecast categories can define the range of expected outcomes and should be matched to user requirements of both accuracy and resolution. Anchoring rainfall forecasts on antecedent conditions can reduce false alarms across the growing season and may be a useful guide when presented alongside a forecast. The next generation of climate data products and services for agriculture need to consider how a forecast system interacts with both on-farm biophysical drivers of yield and decision-making preferences of the user.

Acknowledgements: Funding for this work was provided by the Grains and Research Development Corporation (Project ref. CMA00003).

## References

Brown, J.N., Hochman, Z., Holzworth, D. and Horan, H., 2018. Seasonal climate forecasts provide more definitive and accurate crop yield predictions. Agricultural and Forest Meteorology, 260-261: 247-254.

Duan, W. and Wei, C., 2013. The 'spring predictability barrier' for ENSO predictions and its possible mechanism: results from a fully coupled model. International Journal of Climatology, 33(5): 1280-1292.

French, R. and Schultz, J., 1984. Water use efficiency of wheat in a Mediterranean-type environment. I. The relation between yield, water use and climate. Aust J Agric Res, 35(6): 743-764.

Grains Research and Development Corporation, 2018. Growing Regions. GRDC.

Hayman, P., Crean, J., Mullen, J. and Parton, K., 2007. How do probabilistic seasonal climate forecasts compare with other innovations that Australian farmers are encouraged to adopt? Aust J Agric Res, 58(10): 975-984.

Hochman, Z. and Horan, H., 2018. Causes of wheat yield gaps and opportunities to advance the water-limited yield frontier in Australia. Field Crops Res, 228: 20-30.
Hudson, D. et al., 2017. ACCESS-S1: The new Bureau of Meteorology multi-week to seasonal prediction system. Journal of Southern Hemisphere Earth Systems Science, 67(3): 132-159.
McIintosh, P.C. and Brown, J., 2016. Calibration and bias correction of seasonal climate forecasts for use in agricultural models CSIRO.