Weather Together – Building the weather forecasts farmers want.

Jaclyn N. Brown[1], Andrew Schepen[2], David Robertson[3], Simon Allen[1], Chris Sharman[4], Christopher Nunn[5], Alister Hawksford[6], Michael Bange[5], Vicki Lane[7] and Alicia Garden[8].

[1] CSIRO Agriculture and Food, 15 College Rd, Sandy Bay, 7005.
[2] CSIRO Land and Water, 41 Boggo Road, Dutton Park, QLD 4102
[3] CSIRO Land and Water, Research Way, Clayton VIC, 3168
[4] CSIRO Data61, 15 College Rd, Sandy Bay, 7005.
[5] CSIRO Agriculture and Food, Narrabri
[6] Bureau of Meteorology Agriculture, Parkes Place West, Parkes, ACT, 2600
[7] Director, Square V Software for Agriculture, 2/12 Condon St, Kennington VIC 3550.
[8] CEO Discovery Ag, Trading as GoannaAg.

Abstract
Weather forecasts are crucial to on-farm decision making. Engagements with farmers and AgTech companies has revealed a significant opportunity for providing easier access to localised farm-scale weather and 7 day forecasts for improving timely decision making. Both a multidisciplinary and multiagency (CSIRO and the BoM) team are delivering these services with Agtech companies as an effective and efficient means to reach the users. Specifically a common application programming interface (API) service has been developed. Here we describe the localisation techniques and their skill, the agrometeorological variables we've been supporting. We are also exploring possible business models for developing the service where all stakeholders clearly benefit without concerns for privacy. Moving forward we are collecting on farm weather data which we can then be use in other applications to build even better climate and forecast systems as part of our business case.

Key Words
weather forecasts, down-scaling, AgTech

Introduction
The Bureau of Meteorology (Bureau) provides seven day weather forecasts across Australia. The forecasts are available at the resolution of 6km, but are often only accessed as forecasts for regional centres. These forecasts represent the average conditions for 6 km square grids. Despite this coverage there is still a desire for a forecast 'in my paddock.'

Differences in local paddock weather compared to Bureau's spatially averaged forecasts happen primarily due to topography. For example, winds can be funnelled along hillsides, rain shadows occur on the lee side of mountains, and minimum temperatures will decrease in valleys and hollows.

To address this need we have brought together a multi-disciplinary team from CSIRO including research agronomists, data analysts, computing platform developers and meteorologists from the Bureau. Most importantly we have also partnered with some Australian Agtech companies that provide weather stations and services to the end users to discover how we can understand the most efficient and effective delivery of this service   A key part of developing the initiative is to have each stakeholder benefit from the partnerships, what is sometimes called a 'daisy chain of benefits'

Methods
Partnerships and Data Sharing
In many instances in there is a hesitancy to share information unless the benefits are clear to the provider. Data is extremely valuable and without access it can impair progress, especially in research and development. Our current approach looks to overcome concerns by ensuring all stakeholders benefit from sharing their data with each other (Figure 1). Farmers with on-farm weather stations have their data uploaded automatically to AgTech companies who are in many instances better positioned to manage local on-farm weather stations and deliver the data effectively. In return they receive personalised weather forecasts for the specific paddock where they have a weather station. AgTech companies aggregate the data from many users and upload this to CSIRO. Using this on-farm data CSIRO is able to develop localisation algorithms for paddock scale forecasts. The forecasts are sent back to the AgTech companies to integrate into their larger offerings. CSIRO transfers the weather station data to the Bureau of Meteorology who can use this to

improve their own climate products. Inside CSIRO and Bureau the data is de-identified in climate products protecting the farmers' privacy. Each player has their own value proposition in this process. For farmers it is the perceived value of a localised forecast; for forecast providers, the use of the API and makes localised forecasts easily accessible and easy to update; and the data supply and exchange to CSIRO and BoM meets their independent research and service agendas.

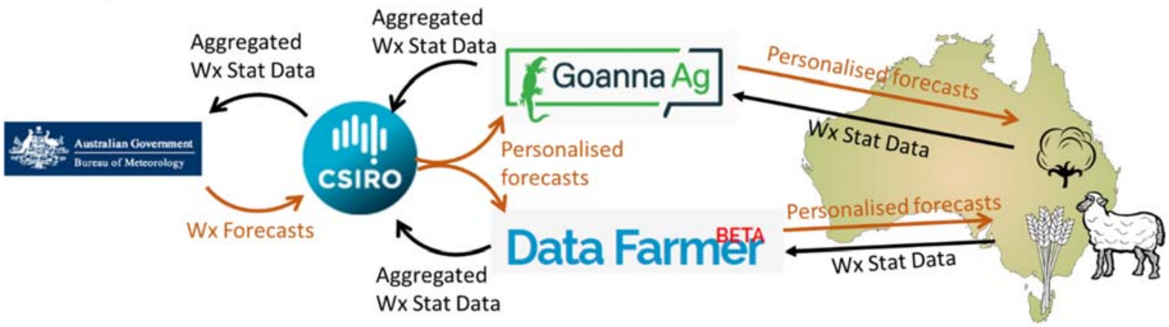

Figure 1. Daisy chain of data sharing to generate personalised weather forecasts.

Localisation algorithms

Key forecast quantities provided by the Bureau of Meteorology are: maximum daily temperature; minimum daily temperature; chance of any rain; and possible rainfall amounts. These are the quantities that we adjust in this first pass, with the addition of hourly temperatures. Our initial goal is to apply a 'light-touch' to the official forecasts to minimise biases at station level.

Historical and real-time 7-day rainfall and temperature forecasts are obtained from the Bureau of Meteorology's Australian Digital Forecast Database (ADFD) gridded forecast product. The forecasts are gridded at 6km resolution nationwide. Meteorologists in regional forecasting centres update the grid cells in their jurisdiction twice a day, at approximately 6 am and 6 pm EST. For this study, we focus on the morning update.

ADFD temperature forecasts are provided as a deterministic time series with an hourly time-step. The maximum daily temperature and minimum daily temperatures are derived from the hourly time series using the Bureau of Meteorology's definitions. The maximum temperature is calculated between 6 am and 9 pm local time. The minimum temperature is calculated between 6 pm and 9 am local time.

ADFD rainfall forecasts are probabilistic. The forecast probability distribution of rainfall is characterised by a probability of rainfall occurring and the rainfall amount corresponding to the 0.50, 0.25 and 0.10 exceedance probabilities (Q_{50}, Q_{25}, and Q_{10}, respectively). ADFD daily rainfall corresponds to a standard 15Z-15Z period, which represents a compromise between eastern and western parts of the country.

Historical and real-time observations for temperature and rainfall from our Agtech partners will be accessed from CSIRO's Senaps platform and processed to match the ADFD forecasts. Senaps is a cloud hosted data platform for managing near real-time spatio-temporal data and analytics. To be able to align the forecasts and observations adequately, it is important that the stations report at least hourly. Quality control is performed to remove erroneous and missing data.

Given the fundamental differences between the temperature and rainfall products, i.e., deterministic vs probabilistic, the calibration strategies necessarily differ. However, there are some common requirements. Weather stations often have short records. The calibration methods needs to perform acceptably with potentially less than 1 year of training data and be able to adapt as the forecast products or observations evolve over time. Observations will not be available on all days. The calibration methods need to handle missing data. We presently impose a minimum of 6 months training data for rainfall, and three months for temperature, noting that better adjustments will be possible with greater amounts of data and coverage of all seasons.

For temperature, we apply an exponential moving average (EMA) bias correction (otherwise known as a decaying average algorithm (Glahn 2014) to minimum and maximum daily temperatures separately. We also apply an independent bias correction for different lead times. The rationale for applying independent corrections is that the biases can differ in magnitude and even be in opposite directions. The EMA bias correction puts more weight on recent errors, allowing the bias adjustment to change over time. Once the minimum and maximum temperature forecasts have been adjusted across the 7-day forecast horizon, the hourly temperature series is rescaled to match the new data points. A non-uniform scaling is applied so that the adjustments reflect the relative biases in Tmin and Tmax. The adjusted hourly time series preserve the main features of the official forecasts, such as the timing of cool changes.

For rainfall, we take an approach that respects the methods used to generate the ADFD rainfall forecasts. The forecast rainfall probability distribution is characterised by 3 parameters: the probability of rainfall (PoP), the expected rainfall and a shape factor (Foley 2011). For each forecast the PoP is a standard ADFD product and the expected rainfall and distribution shape factor can be inferred from Q_{50}, Q_{25}, and Q_{10}. Where there are discrepancies between the probabilistic rainfall forecasts and observed rainfall over a long time window we apply a correction, separately for each lead time. The PoP is scaled by the ratio of the observed frequency of rainfall to the average forecast PoP for similar forecasts over a preceding period. The expected rainfall is scaled by the ratio of average observed rainfall to the average expected rainfall forecasts over the same similar forecasts in the preceding period. The scaled PoP, scaled expected rainfall, and inferred shape factor are then used to re-estimate Q_{50}, Q_{25}, and Q_{10}.

Results and Discussion

As an example of the service and outcomes generated we applied the calibration methods to ADFD forecasts to an existing Bureau weather station Ayr in northern Queensland as an example of the way this will be applied to private weather stations. Here we present example forecasts and verification results for temperature (Figure 2) and rainfall (Figure 3) forecasts. Data is used for Jul 2016 – Jun 2018.

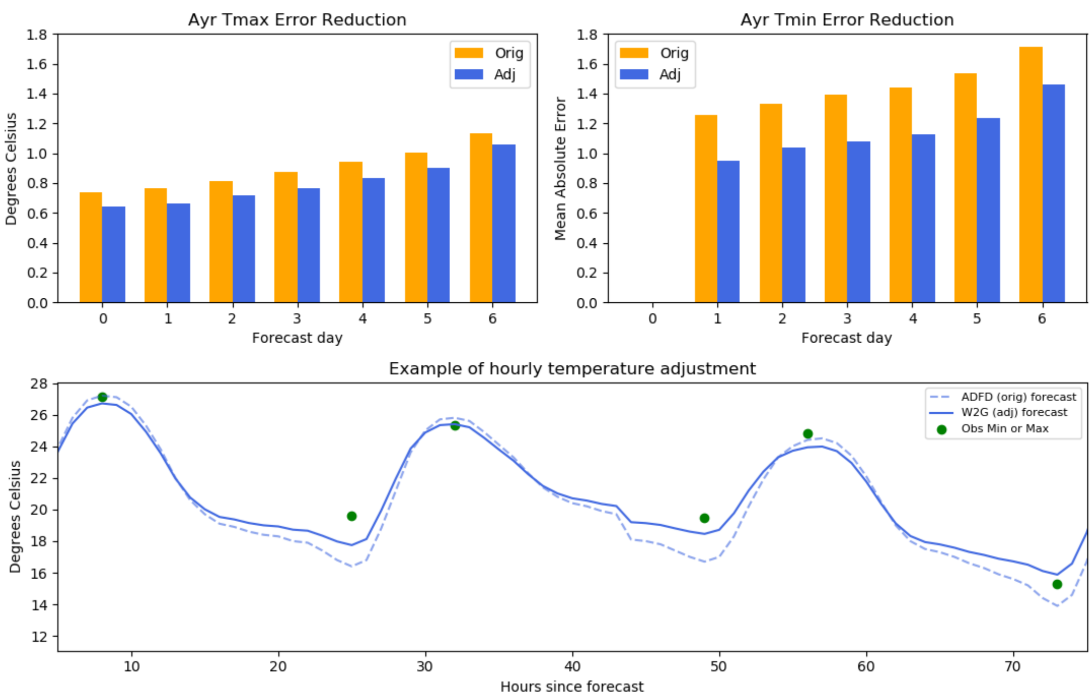

Figure 2 Verification for daily minimum and maximum temperature forecasts (top panels) and example of hourly temperature adjustment over three days (bottom panel). The top two panels show the reduction in Mean Absolute Error (MAE) achieved by applying the dynamic EMA bias correction.

Figure 2 shows that the reduction in error for Tmin forecasts is of the order of 20-30%. For Tmax, the reduction in error is smaller, which suggests less bias in the original forecasts. The bottom panel of Figure 2, which serves as an example only because it is difficult to comment on the performance of a single forecast, shows that for the Ayr location, Tmax requires a negative bias correction, whereas Tmin requires a positive bias correction of greater magnitude. Figure 3 show that the Ayr rainfall forecasts require relatively large adjustments to the expected rainfall. For example, when the forecast probability of precipitation is 50%, the likelihood is closer to 30% (right panel). For amount of rainfall forecasts (left panel) the adjusted probabilistic forecasts more closely resemble the performance of a perfect set of forecasts (black lines) than the uncalibrated forecasts, and thus are more reliable.

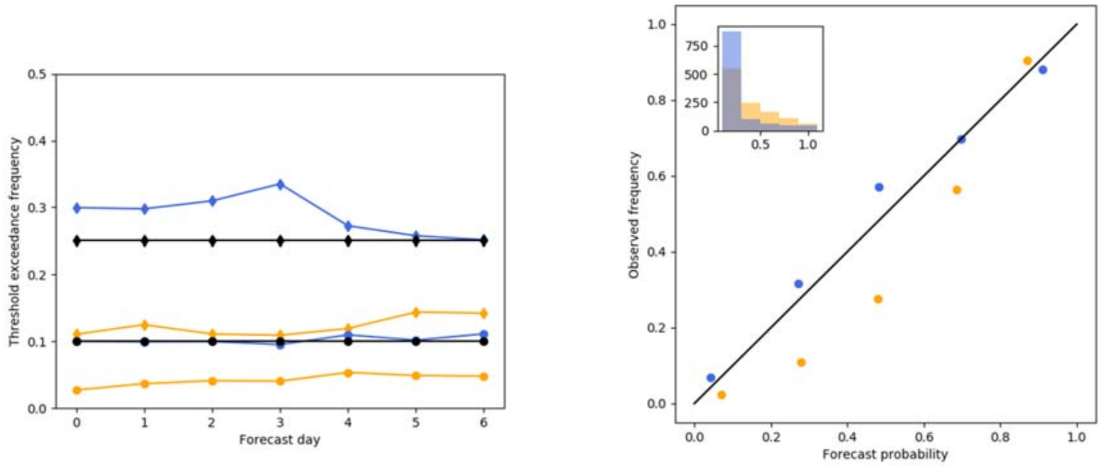

Figure 3 Verification for rainfall forecasts at Ayr in Queensland, before (orange) and after (blue) calibration compared to the perfect set of forecasts (black). The left panel shows the frequency the forecast Q_{25} (diamonds) and Q_{10} (circles) are exceeded at each lead time. The right panel plots the frequency rainfall is observed for forecasts with varying PoP for forecasts with a lead time of 1 day.

The results are in accordance with the design of the localisation methods, which correct for systematic error. Hence it is expected that the average error of the localised forecast will always be similar to or less than the original ADFD forecasts. The true "skill" of the forecast remains largely dependent on the skill of the original Bureau of Meteorology forecasts. We expect that other weather stations will require larger adjustments.

Future work will seek to provide uncertainty estimates on temperature forecasts, calibrate other forecast variables, determine/harness the skill of alternative sources of raw forecasts, and extend the forecast horizon beyond the current seven days.

Looking forward we plan to run user analysis of the value introduced by these paddock scale forecasts – both from the difference to the forecast from downscaling and the appetite of the end users for such a product. We expect that there may be circumstances where the downscaling offers no significant change to the local forecast and others where the local information could be extremely valuable. Our relationships with Agtech companies will also help to strengthen these outcomes.

Conclusion

We have built a 'localisation' algorithm to provide weather forecasts at the paddock scale to users with their own historical weather station records. In addition to this we are working with industry to find innovative pathways and business models to deliver these forecasts in a way that is sustainable and beneficial to all parties. As data privacy is an important 21st Century concern through trialling this process we hope to explore the sensitivities and find workable solutions to benefit agricultural production for the nation.

References

Foley, M., Cooper, S., Riley, P. and Bally, J. (2011). Australian approaches to probabilistic precipitation forecasting. 24th AMS Conference on Weather and Forecasting, Seattle, WA. Extended Abstract 6A.1. https://ams.confex.com/ams/91Annual/webprogram/Paper179054.html

Glahn, Bob. (2104) Determining an optimal decay factor for bias-correcting MOS temperature and dewpoint forecasts. Weather and Forecasting 29.4 (1076-1090).