Increasing forecast accuracy in the context of a multi-model ensemble framework

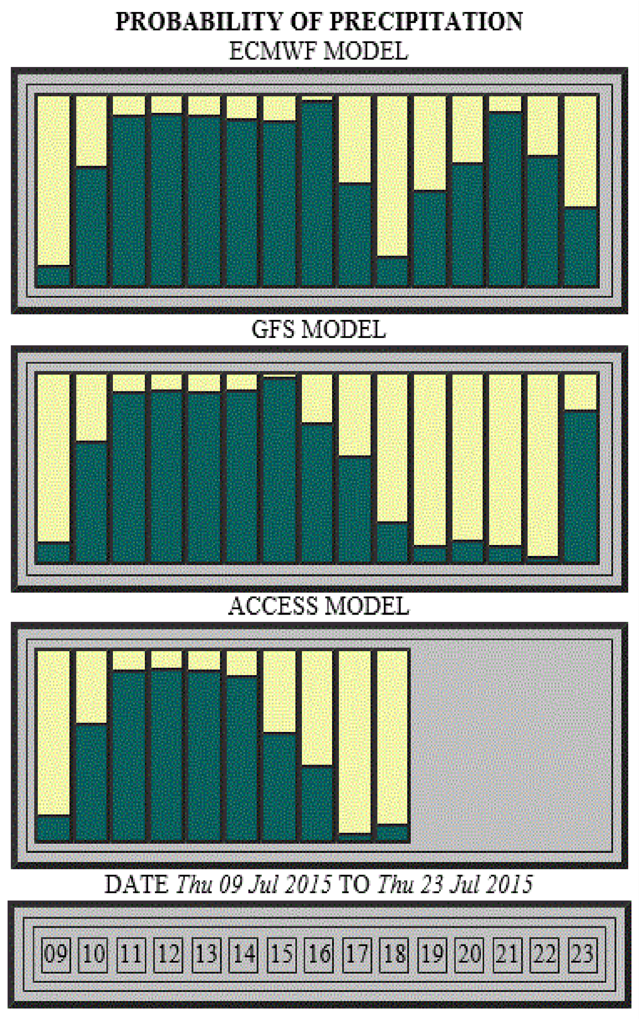
Abstract. The paper presents an assessment of the potential accuracy of day-to-day weather forecasts (for Week-1 and Week-2) that have been derived in the context of a multi-model ensemble framework (*ECMWF & GFS*, plus *Official Predictions* which are largely based on *ACCESS*).

The accuracy of very long range (real-time) day-to-day weather forecasts (Day-15 to Day-32) for Melbourne, Australia, derived by interpreting the output of the *ECMWF* ensemble control model, is also evaluated.

The evaluation is an update of an earlier assessment that was presented to the 2015 Annual Meeting of the American Meteorological Society:

https://ams.confex.com/ams/95Annual/webprogram/Paper267305.html

Results suggest that a very low level of skill exists for predictions Day-15 and beyond and also that the ensemble approach to weather forecasting increases the accuracy of Week-1 and Week-2 day-to-day predictions.



Background. The authors have recently completed a piece of work exploring trends in the skill of day-to-day weather prediction at lead times of 1 to 14 days for Melbourne, Australia.

http://onlinelibrary.wiley.com/doi/10.1002/qj.2559/abstract

The system that was used to establish these trends was, in part, based upon an algorithm that generates local weather forecasts by statistically interpreting the GFS model output.

Previously, the authors presented preliminary results about what was achieved during a six month trial Jul-14 to Dec-14.

The trial involved:

(1) Applying an algorithm to statistically interpret the output of the ECMWF control models in terms of day-to-day local weather out to Day-32, and the output of the GFS model out to Day-14;

(2) Averaging the ECMWF and GFS based output, the ECMWF control models being those applied in the ECMWF ensemble prediction system:

http://old.ecmwf.int/about/corporate_brochure/leaflets/EPS-2012.pdf

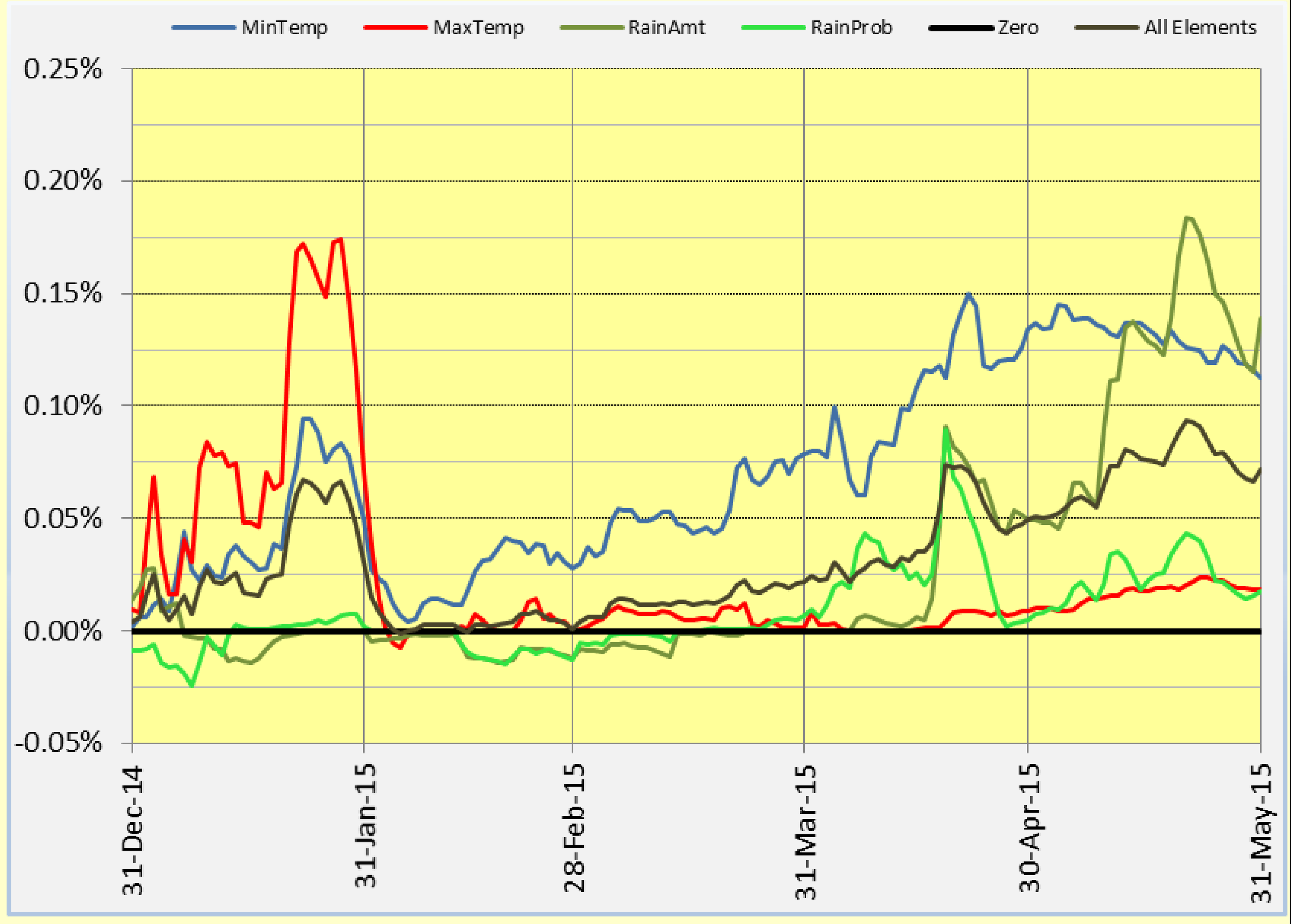
It was considered that it would be interesting to:

(1) Extend the aforementioned trial of day-to-day local weather forecasts out to Day-32;

(2) Assess what might be achieved by optimally combining the output of the models.

It is therefore the purpose of the current presentation to report upon the results of the extended Jul-14 to May-15 trial (Summary Figure 1) and also upon what potentially might be achieved by optimally combining the output of the models (Summary Figures 2 & 3).

Summary Figure 1.

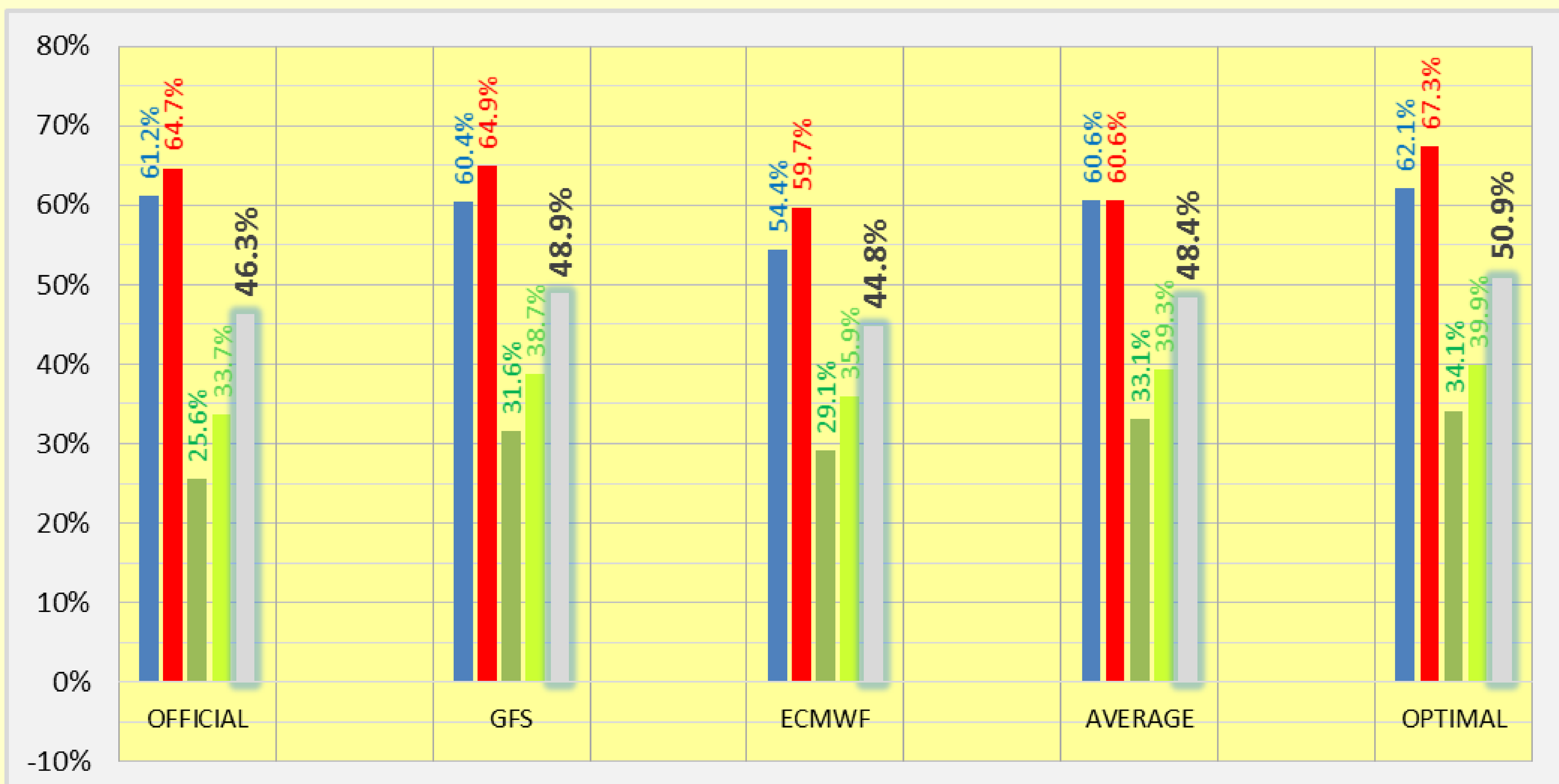


Accumulated (since the beginning of the trial in Jul-14) ***Percent Variance of the Observations Explained*** (***PVOE***)* by the day-to-day ECMWF Day 15-32 real-time predictions of minimum temperature (blue), maximum temperature (red), rainfall amount (dark green), & rainfall probability (light green). Predictions are generated by application of an interpretive algorithm to the model output. As the data base grows, the accumulated skill displayed by the sets of predictions all trend to values close to, but **slightly above**, zero.

*Where ***ACC***, the ***Anomaly Correlation Coefficient***, represents the correlation coefficient between the observed & forecast departure from the seasonal normal:

$$PVOE = (ACC^2) \times (|ACC|/ACC)$$

Summary Figure 2.



PVOE for predictions of minimum temperature (blue columns), maximum temperature (red columns), rainfall amount (dark green columns) and rainfall probability (light green columns), and overall (grey columns).

The accuracy of the official Week-1 predictions (represented by the group of columns on the left) is compared with that of those generated by the application of an interpretive algorithm to the output of the GFS model, the output of the ECMWF model, and an average of the outputs.

The accuracy of a set of predictions derived by using regression analysis to optimally combining the two outputs is represented by the ***adjusted PVOE***, ***PVOE$_{adj}$****.

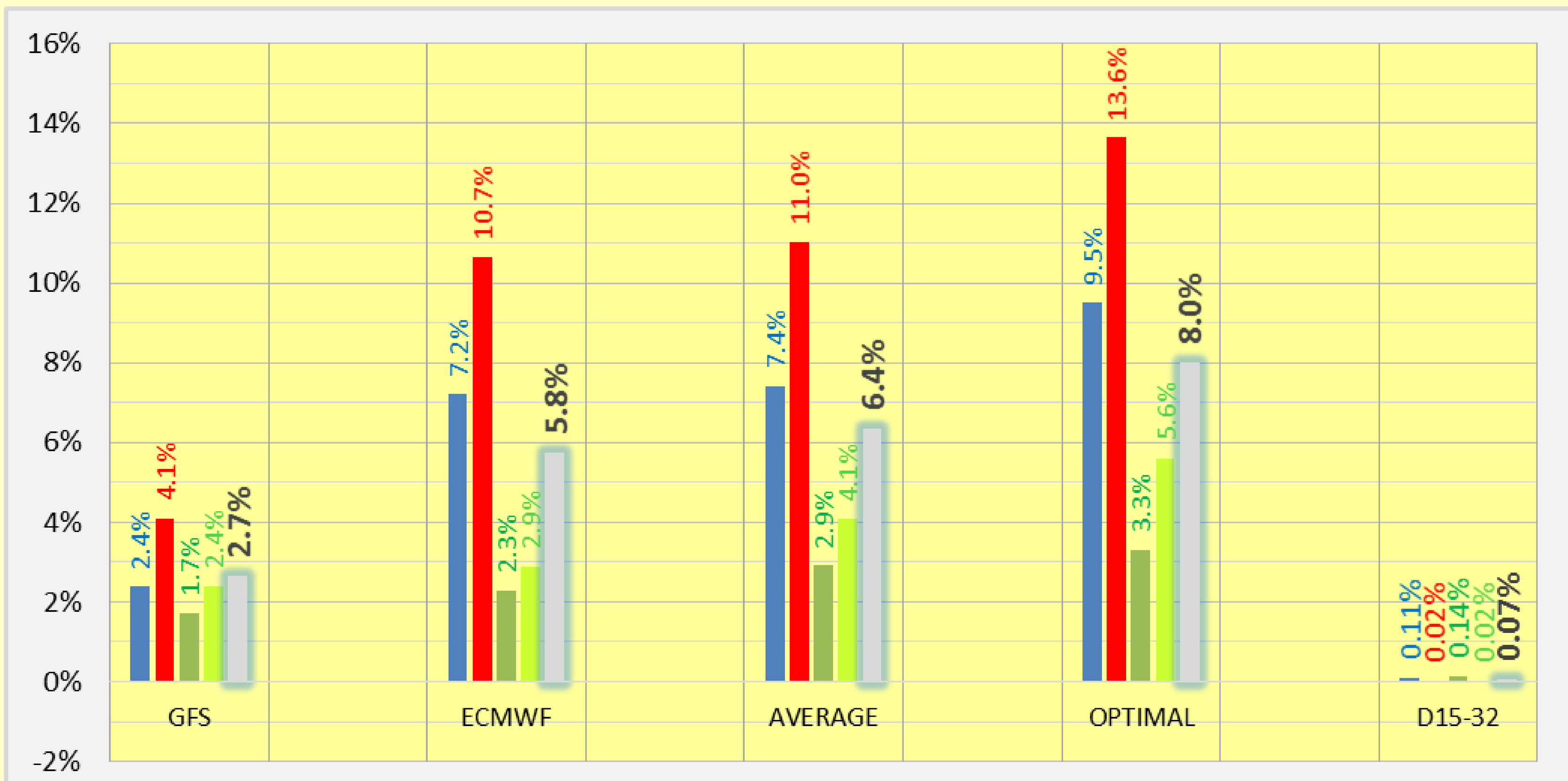
This is shown by the group of columns on the right labelled "OPTIMAL".

*$\boldsymbol{PVOE_{adj}}$ is related to ***PVOE*** in the same way that ***R-squared$_{adj}$***, used in a regression analysis context to avoid the appearance of "false" skill, is related to ***R-squared*** with:

$$PVOE_{adj} = PVOE - ((1\text{-}PVOE) \times (p))/(n\text{-}p\text{-}1)$$

where p is the number of predictors and n is the number of elements.

Summary Figure 3.



As for Summary Figure 2 but ***PVOE*** and ***PVOE$_{adj}$*** for Week-2 predictions, except for there being no official predictions and for the addition of the group of columns on the right which represents the actual ***PVOE*** for the real-time Day 15-32 predictions.

Conclusions. Optimally combining the output of the models leads to a level of overall accuracy of predictions in excess of that achieved by using a simple average of the forecast sets. The accuracy of the Week-1 predictions is increased from 48.4% to 50.9% whilst the accuracy of the Week-2 predictions is increased from 6.4% to 8.0%.

There appears to be some overall skill inherent in the Day 15-32 predictions, albeit of a very low level and almost entirely confined to the forecasts for Day 15-19. Over all four elements, the ***PVOE*** is 0.07%, the ***PVOE*** for Day-15 being 0.8%, for Day-16 being 1.2%, for Day-17 being 0.3%, for Day-18 being 0.4%, and for Day-19 being 1.8%.

Corresponding author: Harvey Stern, School of Earth Sciences, University of Melbourne, Parkville, Victoria, Australia, e-mail: hstern@unimelb.edu.au