SOME ASPECTS OF THE VERIFICATION OF WEATHER FORECASTS FOR MELBOURNE, AUSTRALIA[1]

Harvey Stern[2] & Noel Davidson[3]

1. BACKGROUND.

1.1 Introduction

The authors have recently completed a piece of work exploring trends in the skill of day-to-day weather predictions at lead times of 1 to 14 days for Melbourne, Australia. A summary of this work is presented at:

http://www.weather-climate.com/MelbourneForecastAccuracy.pdf

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 Global Forecasting System (GFS) NWP model output.

1.2 Verifying very long range predictions

As a follow-up to this work, it was considered that it would be interesting to assess what might be achieved using the output of other global NWP models, both by themselves and in the context of a multi-model ensemble framework.

This is the main focus of the current paper, which presents an analysis of preliminary results from a seven-month trial (Jul-14 to Jan-15).

The trial involved applying an algorithm to statistically interpret the output of the ECMWF (EC) NWP control models in terms of day-to-day local weather for Week 1 (Days 1-7) and Week 2 (Days 8-14), and also covering the Day 15-32 period.

We shall now address this feature of the work in detail. Later, in the Appendix, several other aspects of the verification of weather forecasts will be briefly considered.

2. DAY 15-32 PREDICTIONS

The Day 15-32 day-to-day predictions have been derived using an algorithm that statistically interprets the output of the EC control models in terms of local weather.

The EC control models are those applied in the EC ensemble prediction system:

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

Figure 1 represents the accumulated skill by the Percent Variance of the Observations Explained (PVOE[4]) by the predictions.

Figure 1 shows that, as the Day 15-32 data base grows, the accumulated skill displayed by the various sets of predictions, minimum and maximum temperature, rain amount and probability, trend towards zero for each of the four weather elements.

Specifically, as of 07-Feb-2015, the accumulated PVOE by each of the four sets of forecasts, the minimum temperature forecasts, the maximum temperature forecasts, the rainfall amount forecasts and the rainfall probability forecasts were all 0.0%.

An account of what has been achieved by averaging the Week-1 and Week-2 day-to-day predictions derived from the EC and GFS based output is now presented.

It will be seen in what follows that some forecasts derived from the EC model output are less accurate than corresponding forecasts derived from the GFS model output. This may be a consequence of the algorithm applied to interpret the EC model using fewer predictors than that used to interpret the GFS model.

3. WEEK 1 DAY-TO-DAY PREDICTIONS

Figure 2 compares the Week 1 performance of the algorithms utilised to interpret the GFS model output with those utilised to interpret the EC model output.

---

[1] *2015 Annual AMS Meeting, Phoenix, AZ, Harry R Glahn Symposium, Poster 444 [https://ams.confex.com/ams/95Annual/webprogram/Paper267305.html]*

Note that, in preparing the graphics presented in this paper, some later data is used in addition to the data used to prepare the graphics in Poster 444.

[2] *School of Earth Sciences, University of Melbourne, Parkville, Australia.*
hstern@unimelb.edu.au

[3] *Centre for Australian Weather and Climate Research, Bureau of Meteorology, Melbourne, Australia.*
n.davidson@bom.gov.au

[4] The PVOE is directly related to the Anomaly Correlation Coefficient (ACC). To explain, where ACC represents the correlation coefficient between the observed & forecast departure from the seasonal normal:

$$PVOE = (ACC^2)*(|ACC|/ACC))$$

The predictions are generated by the application of interpretive algorithms to the output of the GFS model, far left-hand columns (labelled Wk1GFS), and the EC model, second from left columns (labelled Wk1EC).

The performance of a set of forecasts based upon the *average* of the aforementioned predictions is illustrated by the set of columns third from left (labelled Wk1GFS/EC).

The three aforementioned sets of forecasts have been generated in *real-time*.

Figure 2 shows that the algorithm utilised to interpret the GFS model output yields more accurate Week-1 (Day 1-7) minimum temperature predictions (PVOE=65.4%) than does the algorithm utilised to interpret the EC model output (PVOE=60.4%). The GFS model output also yields more accurate maximum temperature predictions (70.2% vs 64.4%), more accurate rainfall probability predictions (33.3% vs 32.1%), but less accurate rainfall amount predictions (21.0% vs 21.7%). Overall, across all four elements, the GFS PVOE is 47.5%, whilst the EC PVOE is 44.6%.

Combining, by averaging the forecasts associated with the two models, increases the accuracy of forecasts for three of the weather elements to a level superior to those of either model - for minimum temperature to 66.4% from 65.4% & 60.4%, for rainfall amount to 24.1% from 21.0% & 21.7% and for rainfall probability to 35.2% from 33.3% & 32.1%). However, the maximum temperature GFS PVOE is slightly higher than that of the combined forecasts (70.2% vs 70.0%). Nevertheless, overall, across all four elements, combining increases skill to 48.9% from 47.5% & 44.6%.

The potential for achieving a further increase in accuracy, over that achieved by any of the three aforementioned sets of predictions, is illustrated by the set of columns on the far right (labelled Wk1Ensemble).

This set of columns represents the PVOE by a regression relationship derived on data from the outcome of the aforementioned *real-time* sets of predictions *after* the event.

Terms in the regression relationship include number of days ahead (NUM) and the GFS, EC and Official (OFF) predictions, and combinations thereof (NUM*GFS, NUM*EC and NUM*OFF).

To illustrate, the Week-1 regression relationships are presented in Table 1 for each of the four elements.

For minimum temperature, the predictor with the most highly significant partial correlation coefficient with the observed minimum temperature is OFF (0.01% level of significance), underlying the importance of the official forecasts. The second most highly significant predictor is the product, NUM*EC (0.55%), underlying the enhanced importance of the EC model output, especially for longer lead times.

For maximum temperature, the predictor with the most highly significant partial correlation coefficient with the observed maximum temperature is the same as that for minimum temperature, OFF (0.005% level of significance). The second most highly significant predictor is *also* the same as that for minimum temperature, NUM*EC (0.02%).

For rainfall amount, the predictor with the most highly significant partial correlation coefficient with the observed rainfall amount is NUM (0.05%), its negative sign suggesting the need to correct for some over-forecasting at longer lead times. The second most highly significant predictor is GFS (0.21%), reflecting the value of the GFS model output.

For rainfall probability, the predictor with the most highly significant partial correlation coefficient with the observed rainfall probability is the product, NUM*OFF (0.56%), its negative sign reflecting the enhanced importance of the official forecasts for shorter lead times. The second most highly significant predictor is OFF (1.02%), once again reflecting the importance of the official forecasts.

The set of columns on the far right of Figure 2 underlines the potential for further increases in the accuracy of predictions for each of the four elements – for minimum temperature to 68.5%, for maximum temperature to 71.5%, for rainfall amount to 26.1%, for rainfall probability to 35.7%, and overall, across all four elements, to 50.5%.

Given that this set of columns represents the PVOEs of regression relationships that are derived from *past* data, one would need to establish the validity of the relationships in a *real-time* trial, in order to establish the *true* extent of their potential to increase forecast accuracy.

4. WEEK 2 DAY-TO-DAY PREDICTIONS[5]

Figure 3 compares the Week 2 performance of the algorithm utilised to interpret the GFS model output with that utilised to interpret the EC model output. As for the Week 1 predictions, the Week 2 predictions are generated by the application of interpretive algorithms to the output of the GFS model - their performance is illustrated by the far left-hand columns (labelled Wk2GFS), by application to the output of the EC model - their performance is illustrated by the second from left columns (labelled Wk2EC), and an *average* of the two, illustrated by the third from left columns (labelled Wk2GFS/EC).

The potential for achieving a further increase in accuracy, over that achieved by any of the three aforementioned sets of predictions, is illustrated by the set of columns on the far right (labelled Wk2Ensemble) – subject to the qualification presented in Section 3.

[5]Note that there are no day-to-day official predictions for Week 2 at this time.

Terms in the regression relationship include number of days ahead (NUM), the GFS, EC predictions, and combinations thereof, NUM*GFS and NUM*EC.

To illustrate, the Week-2 regression relationships are presented in Table 2 for each of the four elements.

5. WEEKS 1&2 DAY-TO-DAY PREDICTIONS

Figure 4 summarises the outcome of the verification across Weeks 1&2 (Day 1-14). It shows that combining, by *averaging* the forecasts associated with the two models, increases the accuracy of the predictions to a level superior to that of either model for all four of the weather elements.

Subject, once again, to the qualification presented in Section3, the potential for achieving further increases in accuracy via a more comprehensive ensemble approach than simply averaging forecasts from two models (overall PVOE=26.3%), is illustrated by the set of columns fourth from left, labelled Wk1&2Ensemble, which suggests a potential overall PVOE as great as 29.8%.

As reported in Section 2, and now illustrated in Figure 4 by the set of columns on the far right (labelled Day15-32EC), the very long range EC predictions display little forecasting skill.

6. OFFICIAL PREDICTIONS

Figure 5 shows that combining, by *averaging* the forecasts associated with the two models, increases the accuracy of the predictions to a level superior to that of the official forecasts for most elements.

For maximum temperature, the increase is to 70.0% from 68.9%; for rainfall amount, the increase is to 27.9% from 19.8%; for rainfall probability, the increase is to 38.8% from 33.3%; however, for minimum temperature, the combined forecasts have the same PVOE (66.4%) as the official forecasts. Overall, across all four elements, combining increases skill to 50.8% from 47.1%.

Furthermore, the analysis presented earlier suggests that there exists the potential for achieving further increases in accuracy via a more comprehensive ensemble approach than simply averaging forecasts from two models.

7. DAY-TO-DAY PREDICTIONS – ALL LEAD TIMES

Figure 6 summarises the overall PVOE by individual day-to-day predictions across the complete range of lead times (Day-1, Day-2 … Day-32).

That the overall PVOE is positive for all lead times out to Day-19, suggests that there may be some marginal skill in the Day 15-19 component of the Day 15-32 predictions. Beyond Day-19, there is a mixture of positive and negative PVOEs: seven are positive, whilst six are negative.

8. SUMMARY

The paper focuses on what might be achieved by combining the output of global NWP models, both by themselves and in the context of a multi-model ensemble framework.

The results suggest that an ensemble approach to weather forecasting, applied via a process of combining forecasts from various sources, increases the accuracy of day-to-day weather predictions.

During the (albeit) short trial, very long range day-to-day weather forecasts (Day 15-32) for Melbourne, Australia, derived by interpreting the output of the EC control models, displayed little overall forecasting skill. However, there is evidence of a very small amount of skill in the Day 15-19 component of these forecasts.

APPENDIX

Several other aspects of the verification of weather forecasts will now briefly be considered.

A.1 Verifying the official précis

The general public's first impression of the forecast weather is provided by the official précis of that forecast.

An algorithm, which interprets the words contained in the précis in terms of precipitation probability and amount, has been developed. The algorithm was derived via the establishment of two regression relationships on more than 8 years of data (Sep-2005 to May-2014).

The first expresses the *occurrence/non-occurrence* of precipitation as a function of the occurrence/non-occurrence of particular words in the official précis.

The second expresses the *amount* of observed precipitation as a function of the *occurrence/non-occurrence* of particular words in the official précis.

By way of illustration, Figure A.1 demonstrates the effectiveness of the regression relationship that expresses the *occurrence/non-occurrence* of precipitation as a function of the *occurrence/non-occurrence* of particular words in the official précis for June 2014.

A.2 Verifying the prediction of unusual weather

Capability at predicting unusual weather events is of importance. Figure A.2 shows how effectively the GFS model predicts particularly wet days (2.5 mm or greater) in Melbourne well in advance.

Using the Critical Success Index as the evaluation measure, it may be seen that the

GFS model predictions outperforms randomly generated (no-skill) predictions almost out to the end of Week 2.

A.3 Verifying the accuracy of predictions at Melbourne's new observation site

The forecasting system that was referred to in section 1.1 was applied to the prediction of differences between temperatures recorded at the Melbourne Central Business District (CBD) site, and those at a new observation site just outside the CBD (at Olympic Park).

Figure A.3 shows that the Olympic Park site is somewhat cooler than those at the CBD site, particularly during the summer half of the year.

However, the differences in temperatures at the two sites are not just a function of season. Regression relationships (including predictors related to synoptic type) have been derived on 18 months (Jun-2013 to Nov-2014) of data for the purpose of forecasting the temperature differences between the two sites.

For example, Table A.1 presents the equations describing the relationships between the CBD minimum (MIN) and maximum (MAX) temperature, and those at Olympic Park (OP), as a function of SIN and COSIN day of the year, and cyclonicity (CYC), northerly component (N), westerly component (W), and strength (STR) of the surface flow.

Table A.2 compares minimum and maximum temperatures observed at the Olympic Park site since 6-Jan-15, when the CBD site was closed, with those suggested by the minimum and maximum temperature relationships that might have been observed at the CBD site, had the CBD site not been closed.

A *real-time* trial, conducted from Jun-2013 to Nov-2014, saw the correlation coefficient between forecast and observed differences between the minimum temperatures recorded at the two sites of +0.21.

The correlation coefficient between forecast and observed differences between the maximum temperatures recorded at the two sites was +0.47.

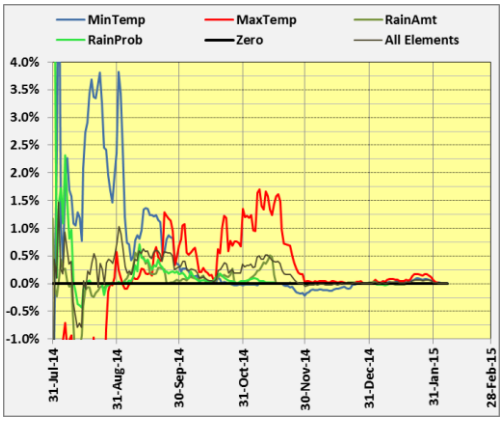

FIGURE 1 Accumulated (Jul-14 to Feb-14) Percent Variance of the Observations Explained (PVOE) by the day-to-day EC Day 15-32 predictions of minimum temperature (blue), maximum temperature (red), rainfall amount (dark green), rainfall probability (light green), and overall – for all elements (thin black line). 'Zero' is indicated by the thick black line.

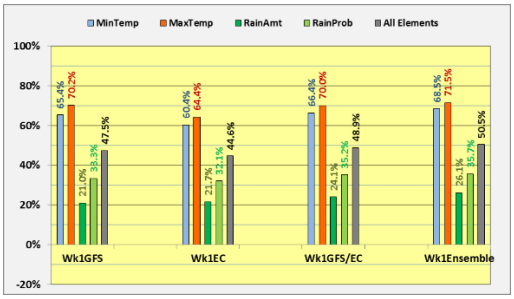

FIGURE 2 PVOE for Week-1 (Day 1-7) predictions of minimum temperature (blue columns), maximum temperature (red columns), rainfall amount (dark green columns) and rainfall probability (light green columns), and overall (grey columns).

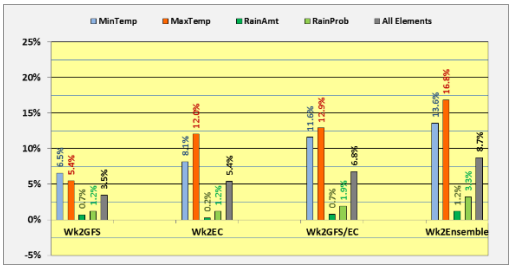

FIGURE 3 PVOE for Week-2 (Day 8-14) predictions of minimum temperature (blue columns), maximum temperature (red columns), rainfall amount (dark green columns) and rainfall probability (light green columns), and overall (grey columns).

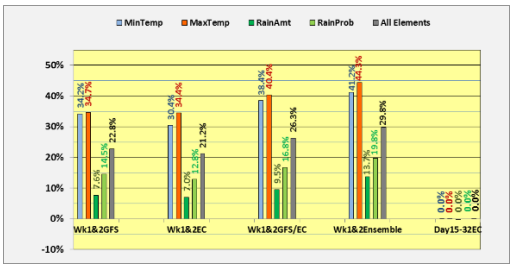

FIGURE 4 PVOE for Weeks 1&2 (Day 1-14), and also Day 15-32, predictions of minimum temperature (blue columns), maximum temperature (red columns), rainfall amount (dark green columns) and rainfall probability (light green columns), and overall (grey columns).

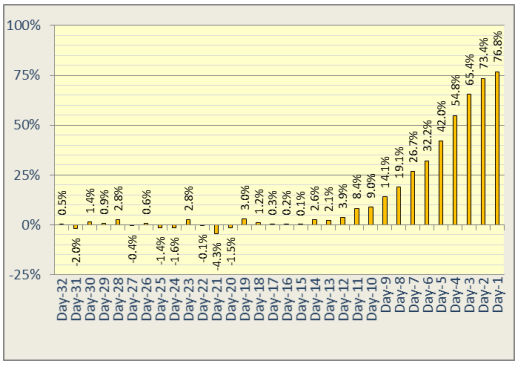

FIGURE 5 Overall PVOE by individual day-to-day predictions (Day-1, Day-2 … Day-32). The Day-1 to Day-14 predictions are the GFS/EC averages, whilst the Day 15-32 predictions are those of the EC alone.

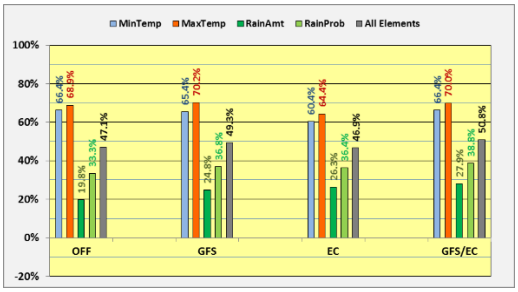

FIGURE 6 PVOE by the official (OFF) predictions of minimum & maximum temperature (Day 1-7), and rainfall amount & probability (Day 1-6)[6], in comparison with corresponding predictions generated by the application of an interpretive algorithm to the output of the GFS model, the EC model, and the average of GFS and EC forecasts.

[6]Note that official forecasts of rainfall amount & probability are only issued out to Day-6 at this time.

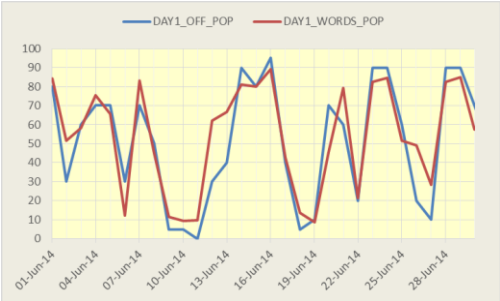

FIGURE A.1 Official Day-1 Probability of Precipitation (DAY1_OFF_POP) forecasts (Jun-2014) compared with those derived using the regression relationship's interpretation of the words in the précis (DAY1_WORDS_POP).

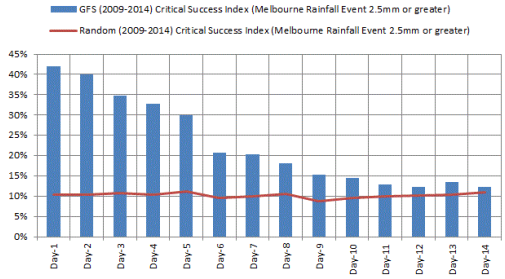

FIGURE A.2 Critical Success Index for predictions of daily rainfall of 2.5 mm or greater in Melbourne for the GFS model set of predictions, and also for a randomly generated (no skill) set.

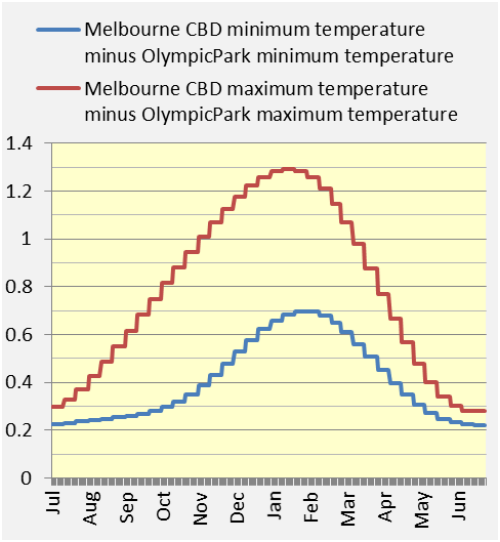

FIGURE A.3 Differences in minimum and maximum temperature between those at the Melbourne Central Business District (CBD) site, and those at a new observation site just outside the CBD (at Olympic Park).

TABLE 1 Week-1 regression relationships for each of the four elements, with regression coefficients that are significant at the 5% level shown in bold red type.

MIN=
0.13**+0.05***NUM**+0.35***GFS-0.02*EC
+0.65*OFF-0.02*NUM*GFS
+0.08*NUM*EC-0.03*NUM*OFF

MAX=
-0.18**+0.12***NUM**+0.57***GFS-**0.42***EC
+0.86*OFF+0.02*NUM*GFS
+0.12*NUM*EC-**0.12***NUM*OFF

AMT=
0.40-0.14*NUM**+0.64***GFS**+0.58***EC
+0.00*OFF-0.02*NUM*GFS
+0.01*NUM*EC-0.09*NUM*OFF

PROB=
-1.97-0.45*NUM+0.21*GFS+0.36*EC
+0.76*OFF+0.11*NUM*GFS
+0.05*NUM*EC-**0.19***NUM*OFF

TABLE 2 Week-2 regression relationships for each of the four elements, with regression coefficients that are significant at the 5% level shown in bold red type.

MIN=
1.36-0.07*NUM**+1.14***GFS**+1.54***EC
-0.06*NUM*GFS-**0.10***NUM*EC

MAX=
1.16-0.06*NUM+0.56*GFS**+1.77***EC
-0.03*NUM*GFS-**0.12***NUM*EC

AMT=
0.19-**0.11***NUM+0.37*GFS**+0.64***EC
-0.02*NUM*GFS-**0.05***NUM*EC

PROB=
-9.64+0.45*NUM**+1.22***GFS**+1.27***EC
-0.09*NUM*GFS-**0.10***NUM*EC

TABLE A.1 The relationships between the CBD minimum (MIN) and maximum (MAX) temperature, and those at Olympic Park (OP), as a function of SIN and COSIN day of the year, and cyclonicity (CYC), northerly component (N), westerly component (W) and strength (STR) of the surface flow. Regression coefficients that are significant at the 5% level are shown in bold red type.

CBDMIN=
(1/**0.9602**)*(OPMIN-0.0312-0.0418*SIN
+0.0028*COSIN+0.0052*CYC**+0.0214***N
-0.0230*W-0.0005*STR)

CBDMAX=
(1/0.9932)*(OPMAX+0.7529+0.0254*SIN
+0.2852*COSIN**+0.0133***CYC-**0.0767***N
-0.0293*W-0.0104*STR)

TABLE A.2 A comparison between maximum temperatures observed at the Olympic Park site (OP) since 6-Jan-15, when the CBD site was closed, with those suggested for the CBD site by the maximum temperature relationship.

DATE	CBDMIN	OPMIN	CBDMAX	OPMAX
06-Jan-15	18.6	17.7	27.4	26
07-Jan-15	19.9	18.9	38.4	37.4
08-Jan-15	24.0	23.1	29.2	28.2
09-Jan-15	17.2	16.5	19.4	18
10-Jan-15	14.1	13.6	20.6	18.9
11-Jan-15	18.6	17.7	29.1	27.5
12-Jan-15	17.0	16.1	32.3	31.3
13-Jan-15	22.8	21.7	26.2	25.6
14-Jan-15	18.4	18.3	24.6	22.7
15-Jan-15	16.1	15.8	21.7	20
16-Jan-15	13.3	13.1	28.4	27.3
17-Jan-15	16.1	15.9	26.5	25
18-Jan-15	15.7	15.3	21.2	19.5
19-Jan-15	14.2	13.5	25.1	23.6
20-Jan-15	13.7	13	30.9	30.4
21-Jan-15	20.5	19.7	34.2	33.1
22-Jan-15	19.5	18.7	36.9	35.8
23-Jan-15	21.3	20.6	26.3	25
24-Jan-15	16.4	15.8	27.4	26.3
25-Jan-15	14.9	14.7	23.6	22
26-Jan-15	12.0	11.6	21.1	19.7
27-Jan-15	12.2	11.7	20.6	19.1
28-Jan-15	14.1	13.3	27.2	25.4
29-Jan-15	13.4	12.8	24.2	22.2
30-Jan-15	13.5	13.1	22.7	20.5
31-Jan-15	14.7	14.3	22.2	20.2
01-Feb-15	13.6	13.2	19.7	19
02-Feb-15	16.3	15.6	22.0	21.1
03-Feb-15	16.7	16.1	20.8	20
04-Feb-15	15.5	14.8	19.8	19
05-Feb-15	14.1	13.3	25.4	24.2
06-Feb-15	15.4	14.5	37.5	35.7
07-Feb-15	23.6	22.7	36.5	35
Means	**16.5**	**15.9**	**25.9**	**24.5**