Is the WRF better than the GFS?

Barry Lynn[1], Asher Meir, Yaakov Consor, and Guy Kelman
Weather It Is, LTD, Israel

1. Introduction

Weather It Is (WII) has used the WRF model to produce operational weather forecasts for the last two years over Israel, and for shorter time periods elsewhere. The WII forecasts have compared quite well against observations (Lynn, 2007). More recently, WII compared their forecasts with those of the Israeli Meteorological Service (IMS), and the Global Forecast Systems Model (GFS).

2. Method

The GFS global forecast data is used to provide both initial (soil, sea, atmospheric) and lateral (atmospheric) boundary conditions for the WRF simulations. Hence, the WRF forecasts "downscale" the GFS forecasts. The GFS boundary conditions (forecasts) are obtained from NCEP servers at 100 km grid resolution. The WRF forecasts are at 36 km, 12 km, and even 4 km grid resolution. When the forecasts are at 12 or 4 km grid resolution, the forecast output is from nested domains within the 36 km grid. In comparison, the IMS forecasts are obtained from MM5 12 km grid resolution model output using the global forecast data from the German Meteorological Service. They are available on the IMS website. WRF forecasts are usually produced four times daily.

3. Results

3a. Seasonal statistics

Figure 1 shows forecast comparisons from November 26th, 2007 to February 29th, 2008 (or winter). The correlations between same day forecast maximum temperatures and observed maximum temperatures are generally highest in the GFS;

[1] E-mail: barry-lynn@weather-it-is.com; U.S Phone: 914 432 3108.

but after applying a commonly used approach to the WII-WRF and GFS forecast data to correct for model forecast bias, the WRF 4 km (and 12 km) forecasts generally produced smaller biases than the GFS. The WRF simulations have higher correlations than the IMS and substantially lower biases (note the IMS forecasts are already corrected for station bias). Hence, the WRF forecasts are providing a better downscaling to the local-forecast level than the IMS.

The forecast comparisons for March 24^{th} to April 29^{th} are also shown in Fig. 1. This time period is when springtime sea-breezes and Sharav heat waves impact Israeli weather. The WRF produces much higher Temperature correlations than the IMS and GFS, and much smaller errors in the mean absolute temperatures than both.

The combined winter and spring forecasts results from 26 November 2007 to 29 April 2008 are shown in Figure. 2. The WRF 12 km same day maximum temperature forecasts have higher correlations than the GFS and IMS between predicted maximum temperatures and observed temperatures. The 4 km WRF did not produce higher correlations than the GFS. Nevertheless, the WRF 12 km and 4 km simulations had the lowest temperature biases in most cities. We did not produce WRF 4 km forecasts beyond one day, but the WRF 12 km forecasts were still consistently much better than the IMS and GFS for the next day and three day forecasts (Fig. 3).

A comparison was made between the predicted daily rainfall and observed rain (Fig. 4). For instance, for same day forecasts the GFS produced higher (winter-time) correlations than the WRF. But, the GFS had a positive rainfall bias at all stations. The WRF (36, 12, and 4 km resolution) simulations had generally smaller biases than the GFS. Moreover, the six station mean precipitation was much less for all the WRF grids than for the GFS.

Figure 5 compares WII-WRF forecasts with those of the IMS. Three categories were created: accuracy, precision, and specificity. Accuracy is defined as the ability of the forecast to predict whether it will rain or not at a particular location. Precision is fraction of times the forecast predicted rain at a particular location and it did rain (> 0.5 mm in 24 hours). Specificity is the fraction of times the forecast correctly predicted it would not rain. The WRF predictions are for rain amounts. A model prediction for rain was indicated when the model predicted greater than 0.5 mm in 24 hours. The IMS predictions on the IMS website are textual. The best results were obtained when we included any kind of rain prediction such as "occasional showers" for a predicted rain event. For the six cities (shown, for example, in Fig. 4) for same day, next day, and three days out, the WRF prediction gave better accuracy, precision, and specificity. Most importantly, when the WRF predicted rain at a particular location, it did so with almost 25% better accuracy.

Forecasted winds are important to both the public and private power generation industry. Figure 6 shows the forecasted wind direction from the WRF and GFS. The WRF produced a much more realistic simulation of three hourly wind directions than the GFS at both Jerusalem and Beer Sheva (a mountain and desert site, respectively).

WRF simulations were produced at 4 km grid resolution for Calcutta, India, for a one-week demonstration. Figure 7 shows that the WRF reproduced the diurnal cycle of moisture, while the GFS did not. The WRF correlation was 0.9, while the GFS correlation was 0.4. Figure 8 shows the WRF versus GFS predicted winds at the same city. The wind direction is crucial in this location for advecting moisture from nearby lakes, rivers, and even the ocean. The WRF's better simulation of wind direction explains quite well the better WRF simulation of dew points.

3b. Case Studies

During the January and February 2008, there were two winter storms. Heavy snow fell in the central mountains of Jerusalem, the Golan, and the northern mountains of Safed. The WRF 4 km (and 12 km) simulation of rainfall (as well as snow) was better than the GFS forecasted rainfall (Fig. 9). Moreover, the 12 km simulations, and especially the 4 km simulations produced much better minimum temperature forecasts in the cold and snowy cities of Jerusalem and Safed (Fig. 9).

During a late March heat wave, the WRF 12 km and 4 km simulations also produced much better same day forecasts of hourly temperatures. Figure 10 shows the hourly temperatures and their correlations. Figure 11 shows the mean average error on all days, hot days (30 ^{o}C or higher), and non-hot days. The WRF produced much higher correlations at the two coastal cities of Netanya and Tel Aviv, and the mountain city of Jerusalem than the GFS. At Lod (or Ben-Gurion Airport) the correlations were similar, but the GFS maximum temperatures greatly underestimated the observed maximum temperatures.

The WII Hurricane Ensemble Forecast System (WHEFS) was used to simulate Tropical Cyclone Nargis, which devastated Myanmar in early April 2008. The WRF 36 km simulations have somewhat higher grid resolution than the operational GFS model (~45 km). Figures 12 and 13 show surface wind speeds obtained from Ensemble Member #7 prior to and during landfall, which strengthened quite substantially during the 12 hours prior to landfall. The ensemble members at 36 km grid resolution simulated the track of the hurricane (from March 26th to April 1st) better than the Joint Tropical Warning Center, which uses the GFS among other models (Table 1). In fact, the best track (not shown) from the Joint Tropical Warning Center on April 1st, 2008 (the day before landfall) had Nargis about 100 to 200 km up the coast of the actual landfall. This forecast track was not nearly as good as those from the 36 km WRF grid simulations or WRF12 km simulations (made three days prior to the event).

If the WRF ensemble set can be considered like an ensemble of the GFS, then it is apparent that the coarse resolution simulations do not simulate the observed wind speed at time of landfall (Table 1), even though most of the simulated storms reached landfall near the observed landfall location. However, the WHEFS 12 km ensemble set simulates a much higher mean maximum wind speed at landfall, and some of the simulations are fairly close to the observed maximum wind speed at landfall. A single simulation of the WRF at 4 km grid resolution gave a maximum wind speed very close to the observed.

4. Conclusions

Is the WRF better than the GFS? It appears that the WRF higher resolution forecasts are essential for predicting daily surface maximum temperatures, dew points, and wind speed. The WRF also does much better than the IMS, which uses MM5 to make their predictions. The correlation for predicted rainfall do not standout for being better than the GFS, but the model has on average lower rainfall biases than the GFS. One explanation for this maybe that many precipitation events are during the winter in Israel are convectively driven. Hence, the higher 12 and 4 km WRF forecasts may predict that heavier rain will fall, but not in the correct location. Hence, the correlations are higher when using the 36 km WRF or even GFS. When comparing the WRF precipitation forecasts to those of the IMS, the WRF better predicted the occurrence or non-occurrence of rain.

Many public and private users of meteorological data are quite concerned with extreme events. Three case studies: two winter storms, a heat wave, and a hurricane demonstrate the superiority of the WII-WRF prediction system. In regard to predicting station rainfall amounts in the two winter storms, we postulate that the improvement in precipitation for the WRF occurred because the synoptic fields had a greater organizing impact on the mesocale precipitation. The heat waves were better predicted because the WRF, but not the GFS, can resolve the localized sea-breeze circulations and the Sharav Lows that cause such features in springtime in Israel. The WII Hurricane Ensemble Forecasting System simulated quite well the track of Nargis that devastated Myanmar; the WRF's high-resolution grid was essential for predicting the range of possible maximum winds at landfall, some of which were very close to observed.

5. Reference

Lynn, B. H., 2007: High resolution accurate WRF forecasts for the Middle East. 22nd Conference on Weather Analysis and Forecasting/18th Conference on Numerical Weather Prediction. Abstract ID#: 123944, Salt Lake City Utah.

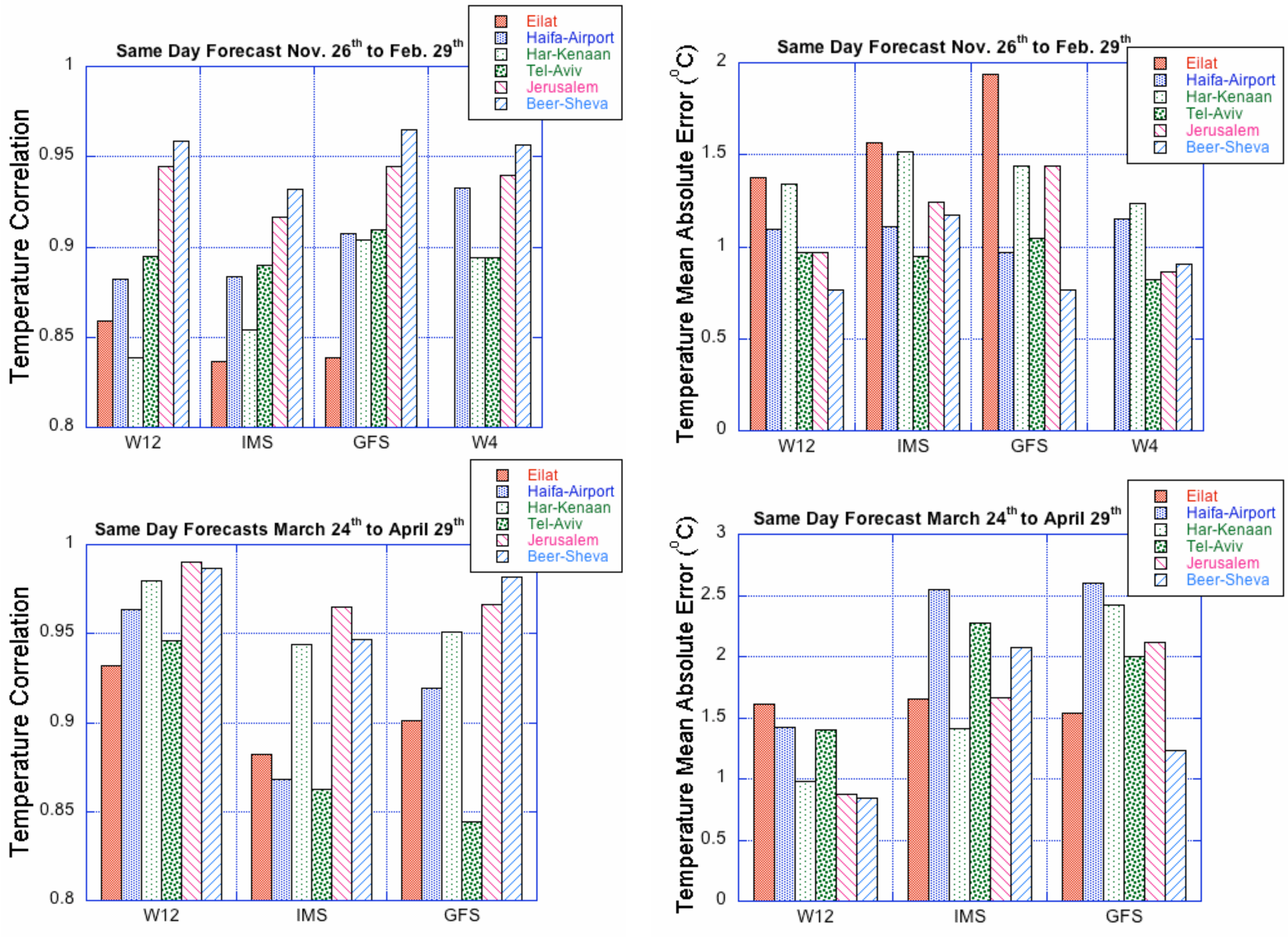

Figure 1: Temperature correlations and corrected mean absolute error from the WRF model at 12 km grid resolution (W12), the Israel Meteorological Service (IMS), the Global Forecast Systems Model (GFS), and the WRF at 4 km grid resolution (W4). The relevant dates are at the top of each graph.

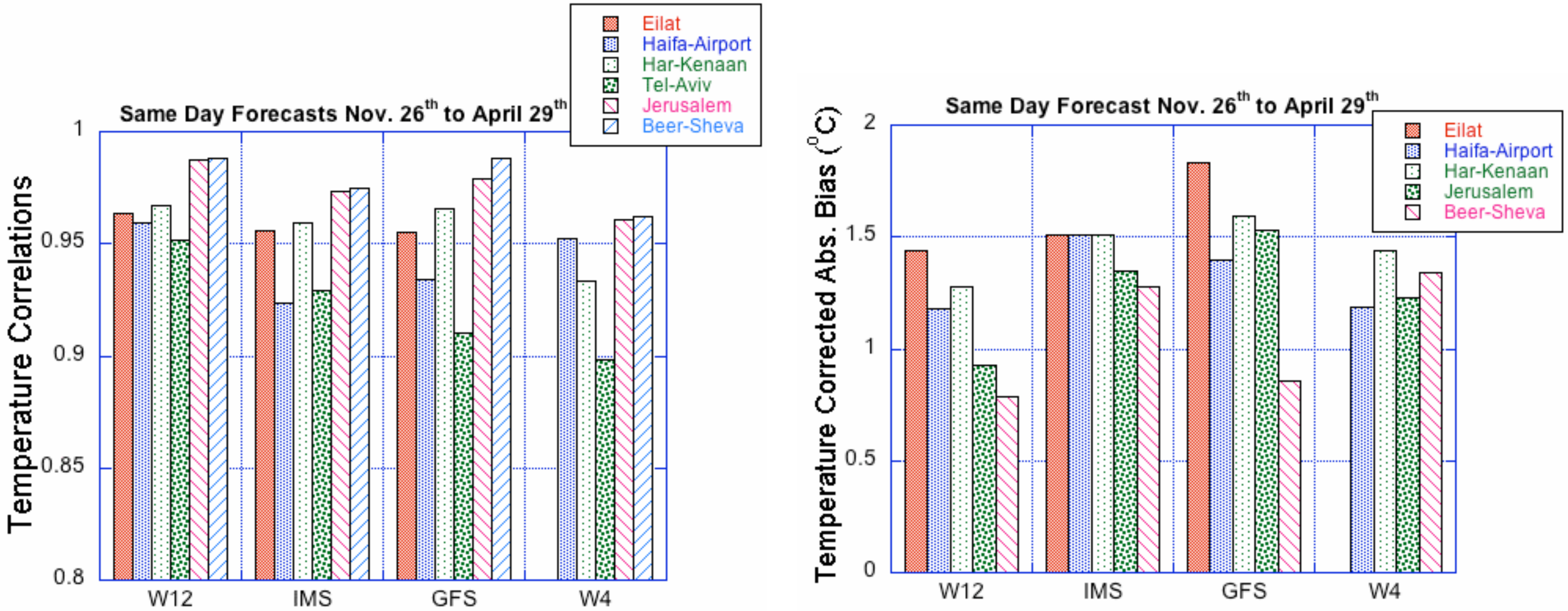

Figure 2: Temperature correlations and corrected mean absolute error from the WRF model at 12 km grid resolution (W12), the Israel Meteorological Service (IMS), the Global Forecast Systems Model (GFS), and the WRF at 4 km grid resolution (W4). The relevant dates are at the top of each graph.

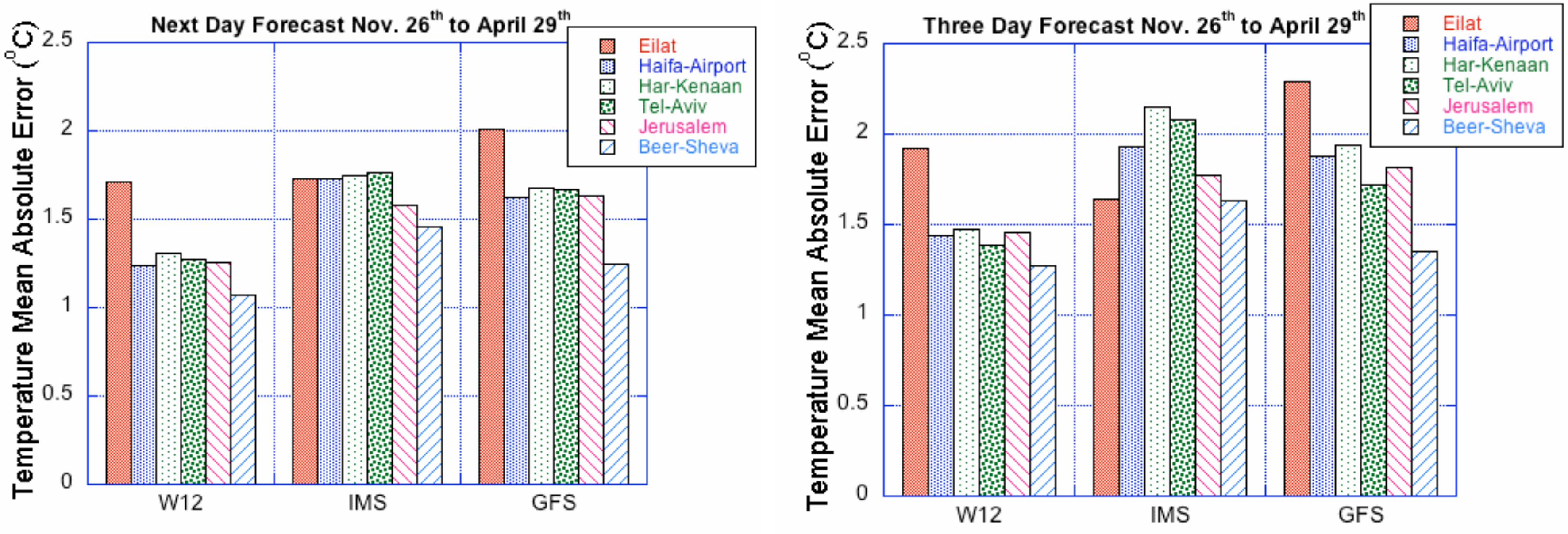

Figure 3: Temperature mean absolute error from the WRF model at 12 km grid resolution (W12), the Israel Meteorological Service (IMS), the Global Forecast Systems Model (GFS), and the WRF at 4 km grid resolution (W4). The relevant dates are at the top of each graph.

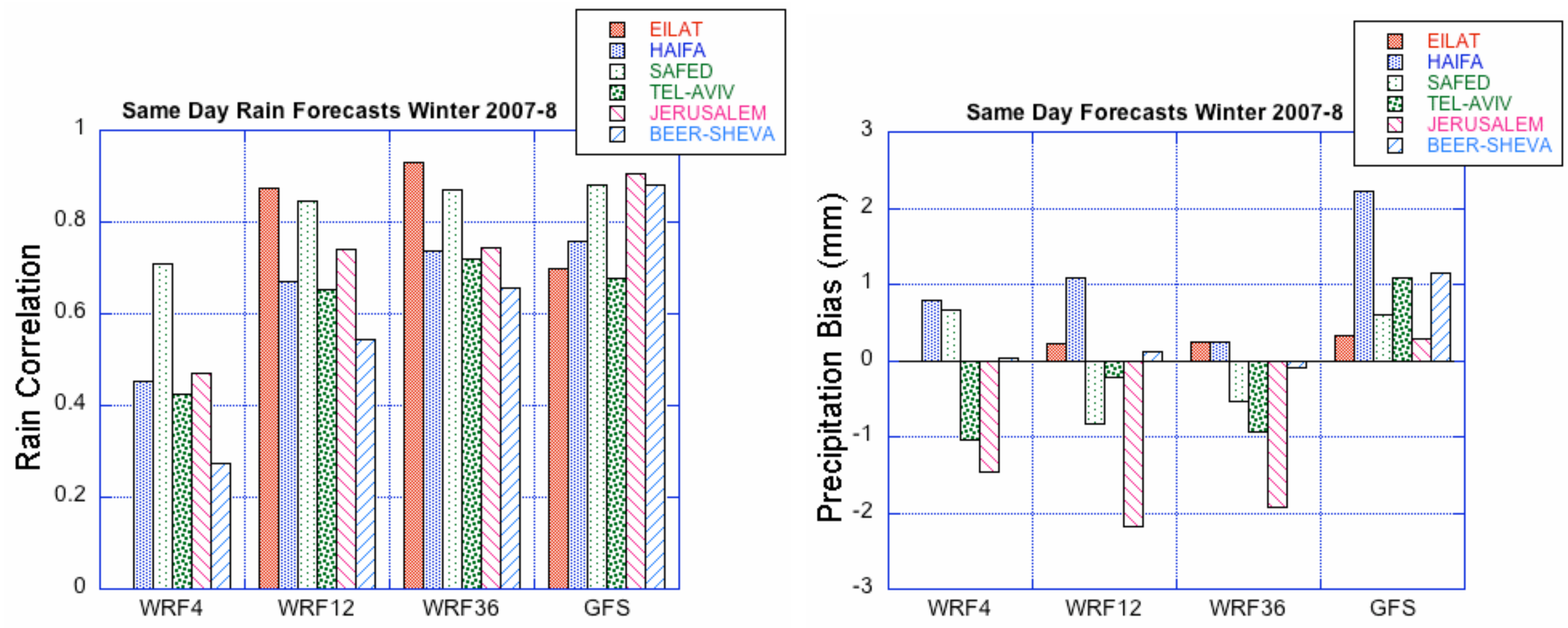

Figure 4: Rain correlations and biases from the WRF model at 4, 12, and 36 km grid resolution (WRF12), versus the Global Forecast Systems Model (GFS). The relevant dates are at the top of each graph.

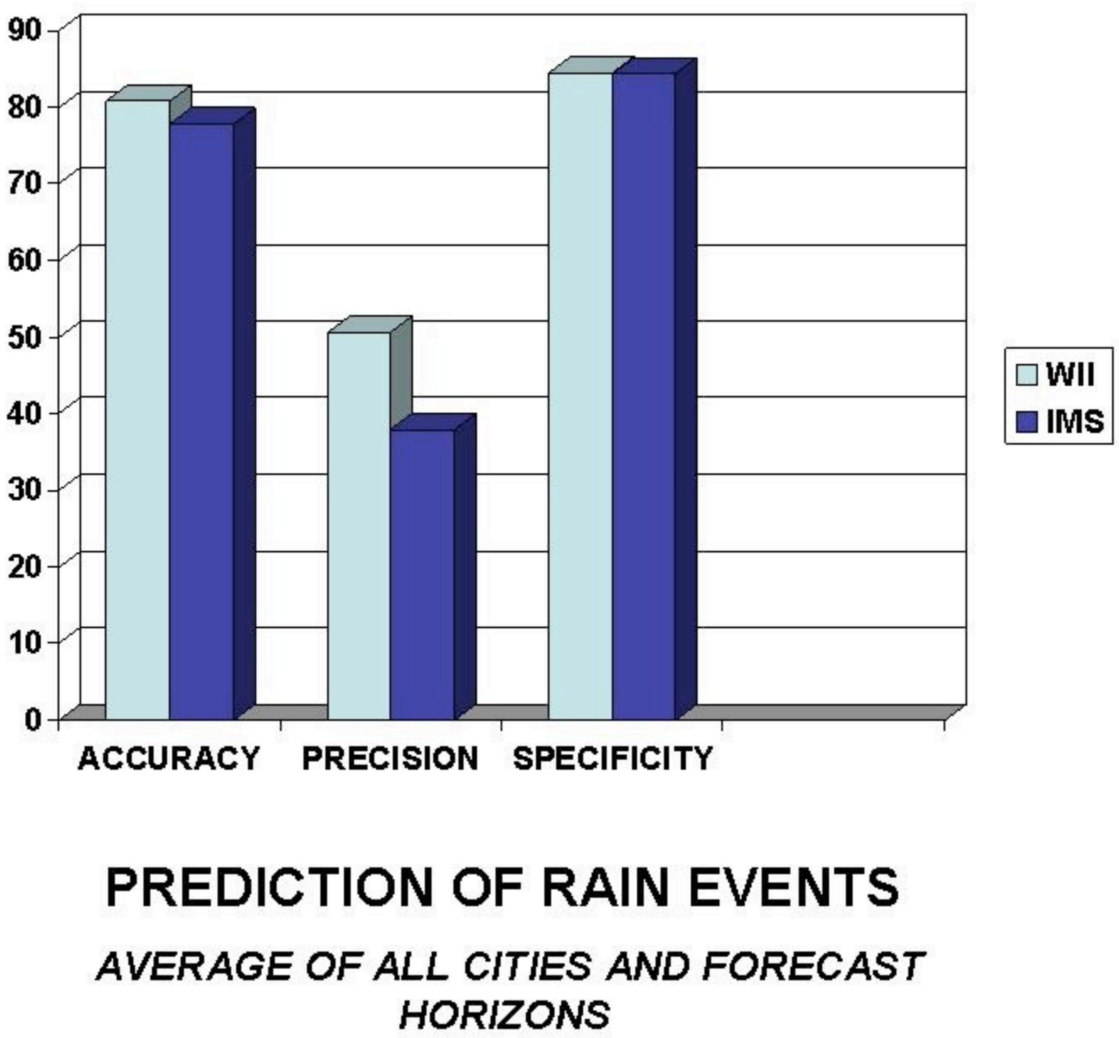

Figure 5: IMS vs WII-WRF(12 km) prediction of rain events for 2007-2008. Accuracy: what fraction of times did WII or the IMS predict it would or would not rain; precision: what fraction of rain events actually occurred; specificity: what fraction of times did either correctly predict it would not rain.

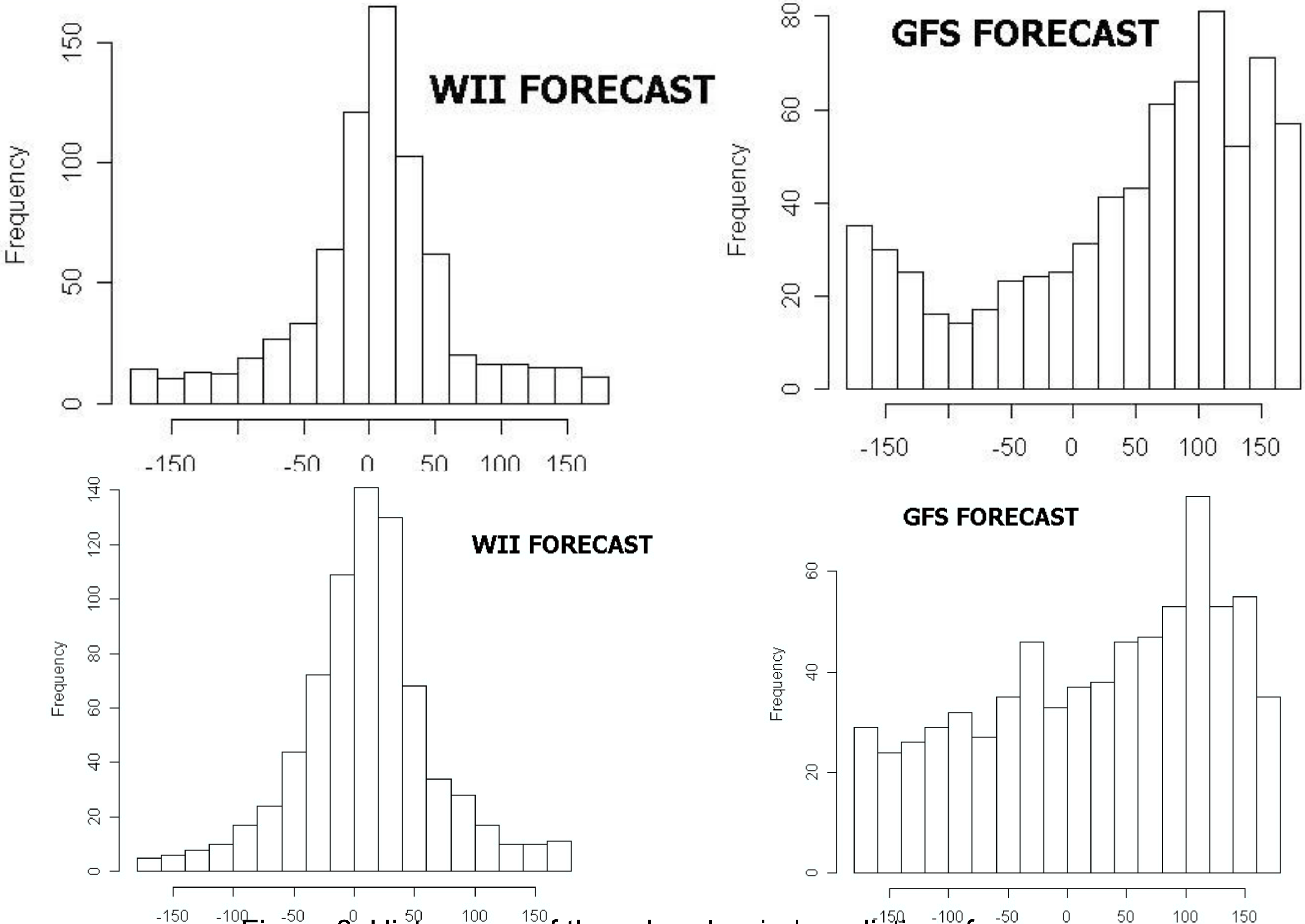

Figure 6: Histograms of three-hourly wind predictions for Jerusalem (top) and Beer-Sheva (bottom) from the 12km WRF and GFS. The dates were Nov. 26th 2007 to Feb. 29th 2008.

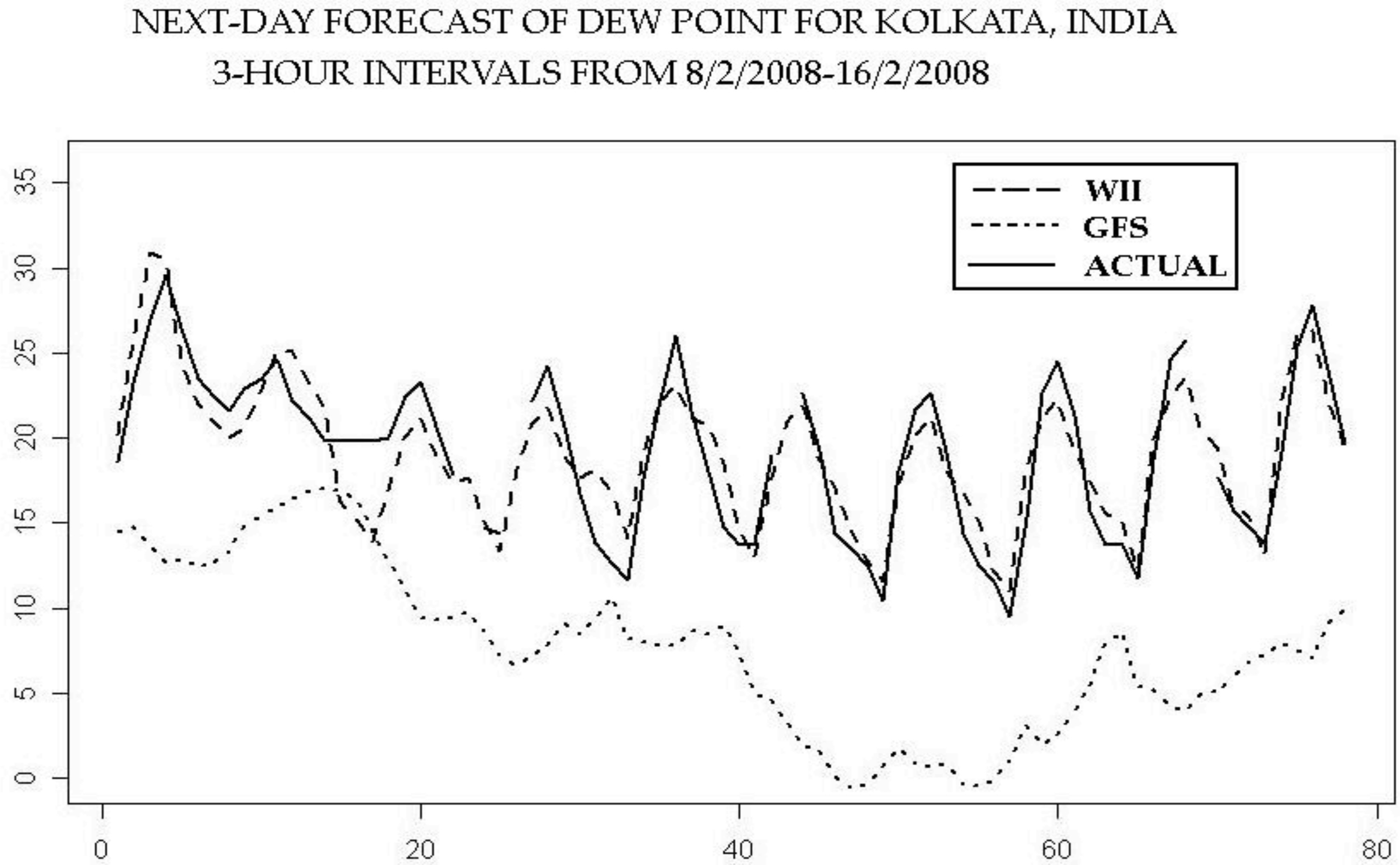

Figure 7: Observed, 4 km WRF-and GFS forecasted dew points. Note, the WRF dew point values were corrected for bias error.

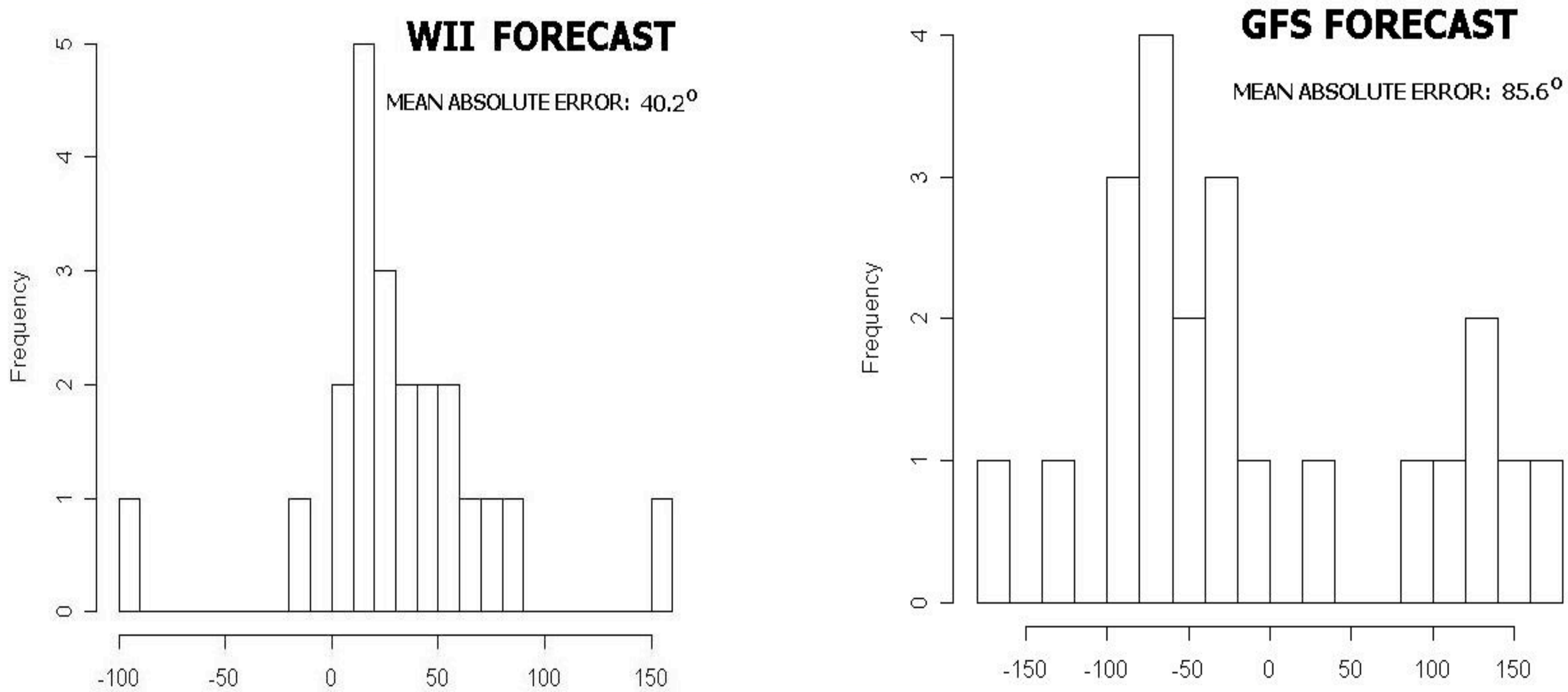

Figure 8: Histogram of three-hourly winds from the WRF (left) and GFS (right) for Calcultta. The dates were Feb. 10th to Feb. 16th, 2008.

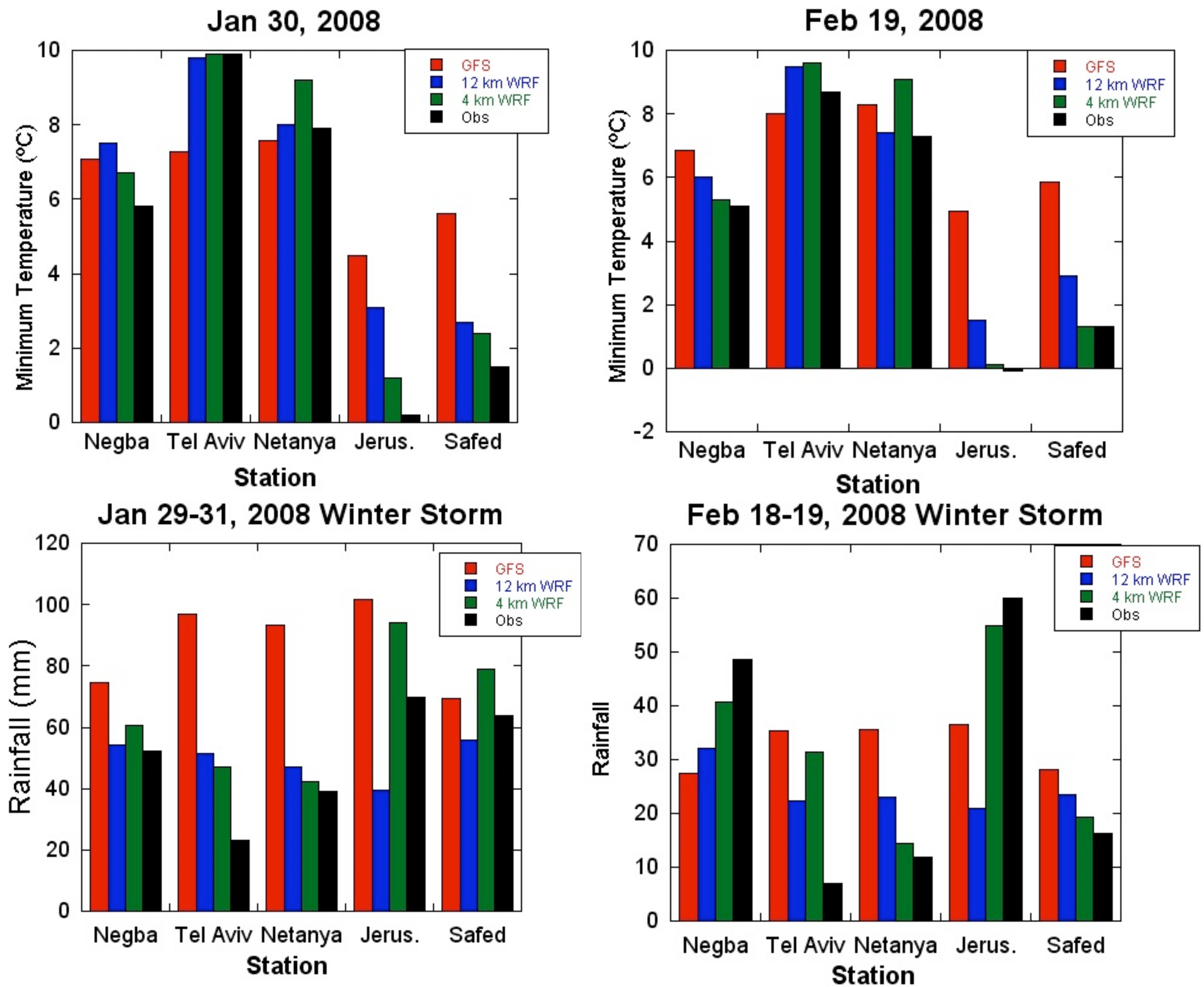

Figure 9: Comparison of same day forecasts of minimum temperature and precipitation amounts. Heavy snow fell in Jerusalem and Safed.

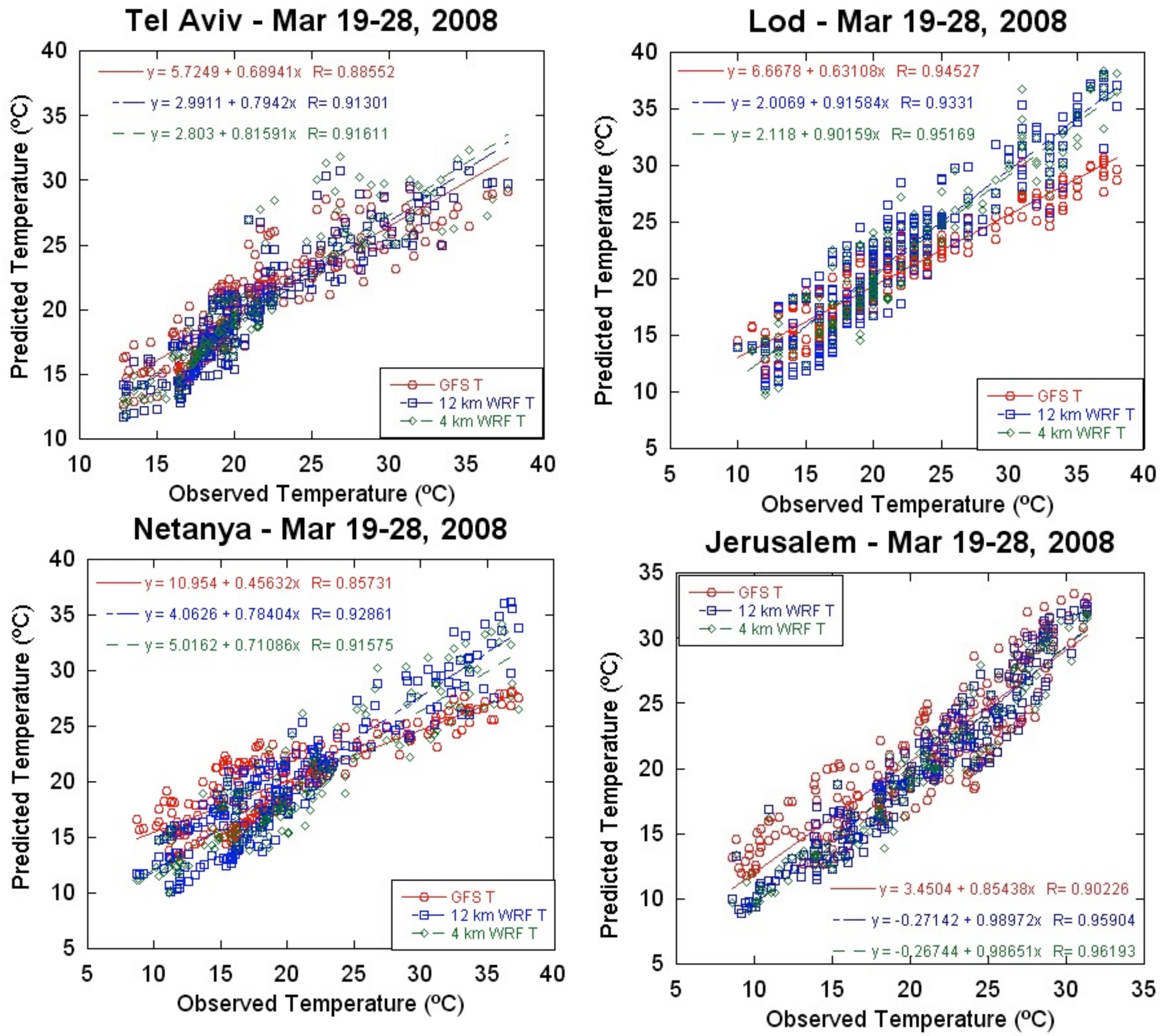

Figure 10: Predicted versus Observed two-meters temperatures at 4 Israeli cities before, during, and after a late March heat wave.

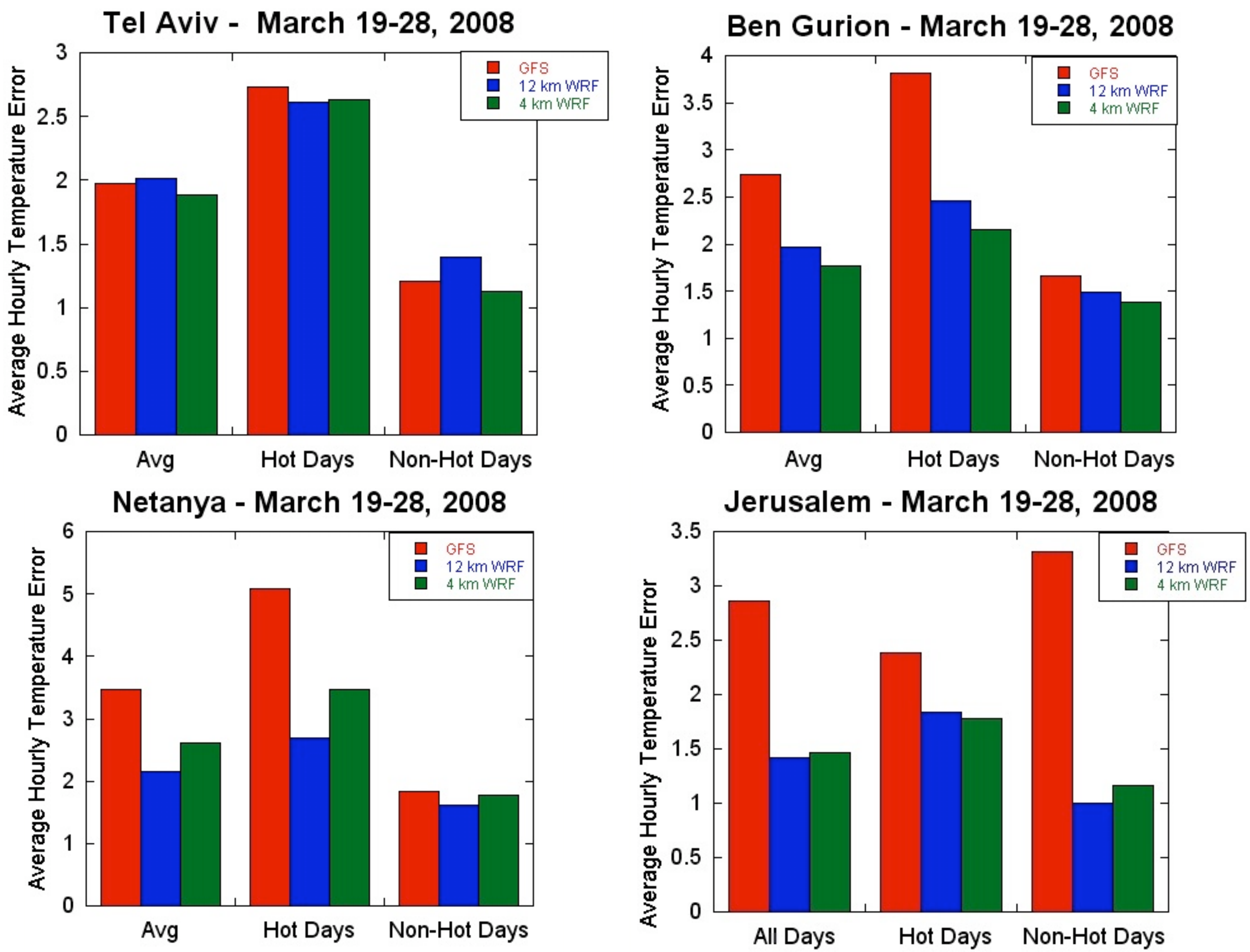

Figure 11: Predicted errors at 4 Israeli cities prior to, during, and immediately after a late March heat wave.

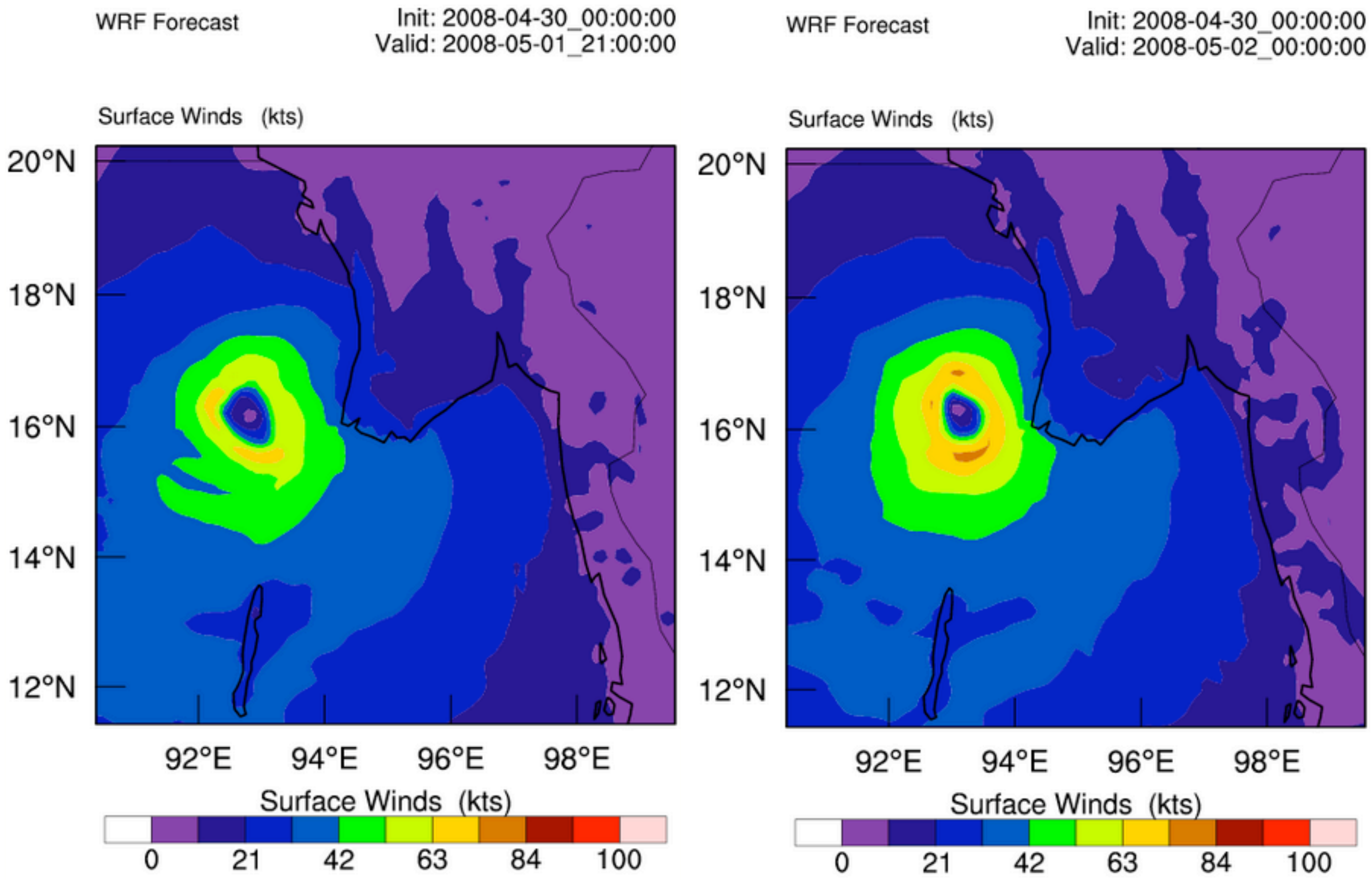

Figure 12: Surface wind speed simulated with the WHEFS from ensemble member #7. The grid resolution was 12 km. The figures on the right shows the hurricane beginning to strengthen (as observed) before reaching land.

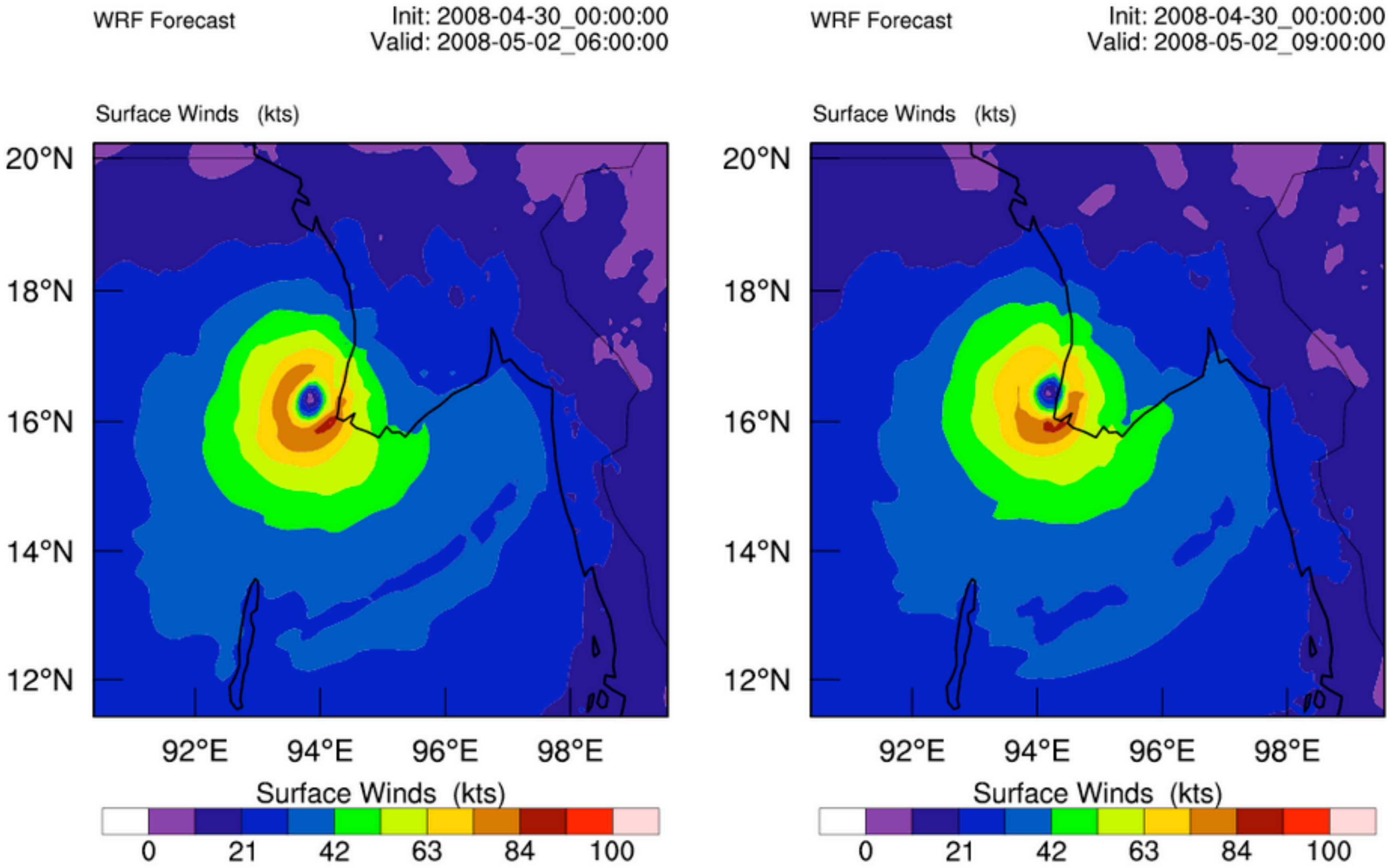

Figure 13: WRF 12 km grid resolution of Nargis at 6 and 9 GMT (at landfall) on April 2, 2008. The results are from ensemble member #7.

Table 1: WHEFS ensemble member's minimum distance in degrees from the landfall position of Tropical Cyclone Nargis, as calculated from a comparison of the WHEFS landfall position and actual landfall position in degrees latitude and longitude. The date the forecast was issued is indicated in the left column (for 0z). The mean of all ensemble members is in the second column from the right. The WHEFS forecasts are for seven days. The last column shows the forecast from the Joint Typhoon Warning Center (JTWC). The JTWC makes forecasts for 5 days. For day 6 and 7, the JTWC forecasts on 26 and 27 of August were extrapolated from the JTWC data and newspaper reports. Top table shows results from simulations at 36 km grid resolution; bottom table shows results from 12 km grid resolution simulations.

date	ens_0	ens_1	ens_2	ens_3	ens_4	ens_5	ens_6	ens_7	ens_8	mean	JTWC
26/4	2.9	3.3	3.8	2.1	2.4	2.7	3.6	3.2	4.1	3.1	6.5
27/4	2.7	1.7	0.8	5.2	2.0	3.0	0.3	2.4	3.9	2.4	6.3
28/4	2.3	1.4	1.7	2.7	3.8	1.4	2.6	2.3	2.0	2.2	3.1
29/4	1.0	1.7	1.4	0.4	1.5	1.7	1.1	2.6	0.5	1.3	2.4
30/4	0.5	0.7	0.8	0.5	0.1	1.0	0.5	0.1	1.7	0.7	1.9
1/5	0.2	0.1	0.1	0.2	0.1	0.1	0.2	0.1	0.3	0.2	1.3

date	ens_0	ens_1	ens_2	ens_3	ens_4	ens_5	ens_6	ens_7	ens_8	mean	4 km	JWTC
29/4	0.3	2.2	1.8	0.6	2.1	1.9	1.5	2.9	0.7	1.6	X	2.4
30/4	0.7	0.9	1.3	0.7	0.5	1.4	0.9	0.3	1.8	0.9	0.9	1.9
1/5	0.2	0.2	0.5	0.1	0.0	0.3	0.1	0.3	0.6	0.3	0.3	1.3

Table 2: WHEFS simulated maximum 10 meter wind speed in knots at time of landfall. The observed maximum speed was given at 105 knots. The table at the top shows the results from WHEFS using a single 36 km grid. The forecast winds vary from simulation to simulation by more than 30%, but are substantially less than observed. Below, the table continues with results from simulations using a nested 12 km grid as an ensemble, which produced wind speeds about 50% higher than those obtained on the more coarse, 36 km grid. The 4 km grid produced a 98.4 knot maximum wind at land fall from its 0z forecast on 1 May 2008. This suggests a category 3 storm, but the range of ensemble values suggests the possibility of a category 4 tropical cyclone on the Saffir-Simpson scale.

date	ens 0	ens 1	ens 2	ens 3	ens 4	ens 5	ens 6	ens 7	ens 8	mean	JTWC
29/4	60.	44.	41.	59.	54.	55.	54.	61.	58.	54.	85.
30/4	63.	56.	50.	63.	57.	49.	62.	62.	65.	58.	80.
1/5	61.	54.	64.	50.	60.	66.	58.	47.	58.	58.	80.

date	ens_0	ens_1	ens_2	ens_3	ens_4	ens_5	ens_6	ens_7	ens_8	mean	4 km	JTWC
29/4	76.	69.	67.	82.	75.	78.	79.	84.	72.	76.	X	80
30/4	84.	82.	73.	82.	80.	65.	81.	91.	82.	80.	77.1	80
5/1	79.	71.	93.	63.	67.	85.	79.	60.	81.	75.	98.4	80