

Verification of operational weather forecasts from the POSEIDON system across the Eastern Mediterranean

A. Papadopoulos¹ and P. Katsafados²

¹Institute of Inland Waters, Hellenic Centre for Marine Research, Anavyssos Attikis, Greece

²Department of Geography, Harokopio University of Athens, Greece

Received: 14 April 2009 – Revised: 13 July 2009 – Accepted: 20 July 2009 – Published: 31 July 2009

Abstract. The POSEIDON weather forecasting system became operational at the Hellenic Centre for Marine Research (HCMR) in October 1999. The system with its nesting capability provided 72-h forecasts in two different model domains, i.e. 25- and 10-km grid spacing. The lower-resolution domain covered an extended area that included most of Europe, Mediterranean Sea and N. Africa, while the higher resolution domain focused on the Eastern Mediterranean. A major upgrade of the system was recently implemented in the framework of the POSEIDON-II project (2005–2008). The aim was to enhance the forecasting skill of the system through improved model parameterization schemes and advanced numerical techniques for assimilating available observations to produce high resolution analysis fields. The configuration of the new system is applied on a horizontal resolution of $1/20^\circ \times 1/20^\circ$ (~ 5 km) covering the Mediterranean basin, Black Sea and part of North Atlantic providing up to 5-day forecasts. This paper reviews and compares the current with the previous weather forecasting systems at HCMR presenting quantitative verification statistics from the pre-operational period (from mid-November 2007 to October 2008). The statistics are based on verification against surface observations from the World Meteorological Organization (WMO) network across the Eastern Mediterranean region. The results indicate that the use of the new system can significantly improve the weather forecasts.

1 Introduction

The POSEIDON project was established at the Hellenic Centre for Marine Research (HCMR) in 1997 aiming at the development and implementation of a real time monitoring and an operational forecasting system for the marine environmental conditions of the Greek Seas. Within the framework of this project a weather forecasting system has been developed in collaboration with the University of Athens. The basic concept was to design a reliable and computationally efficient system that produces high accuracy weather forecasts, particularly useful for predicting local atmospheric conditions and for forcing the wave and ocean hydrodynamic models of the POSEIDON system with surface fluxes of momentum, moisture, heat, radiation and precipitation rates. The system consists of (a) the data acquisition and pre-processing system, (b) the core numerical weather prediction (NWP) model, (c) the dust cycle model and (d) the visualization/post-processing system. The NWP model is based on the SKIRON/Eta model (Kallos et al., 1997), which is a modified version of the regional Eta Model of the National Centers for Environmental Prediction (NCEP). It is fully-coupled with a module for the simulation and the operational forecast of the major phases of the atmospheric dust cycle life such as production, up taking, diffusion, advection and removal (Nickovic et al., 2001), in which dust substance is treated as a passive tracer. A detailed description of the Eta model dynamics and physics packages can be found in previous studies (e. g. Janjic, 1984; Mesinger et al., 1988; Janjic, 1994). The main products of the POSEIDON weather forecasting system are precipitation, snowfall, cloud coverage, near-surface air temperature and winds, sea level pressure, and dust concentration and deposition.



Correspondence to: A. Papadopoulos
(tpapa@ath.hcmr.gr)

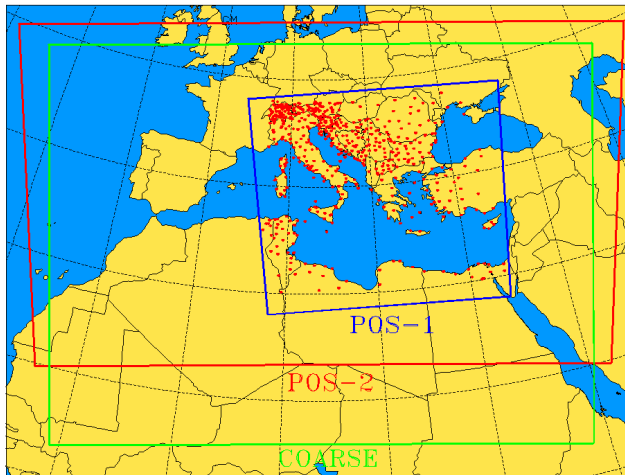


Fig. 1. The three model domains of the POSEIDON weather forecasting systems. The 25-km (COARSE), the 10-km (POS-1) and the 5-km (POS-2) grid spacing. The red cycles indicate the locations of the meteorological stations used in the study.

Maps of these data are available through the website <http://www.poseidon.hcmr.gr>.

This paper reviews the HCMR's weather forecasting systems and presents the quantitative verification statistics calculated from model outputs during the pre-operational period (from mid-November 2007 to October 2008). The statistics presented in this study are based on verification against surface observations from the World Meteorological Organization (WMO) network, covering the extended area of Eastern Mediterranean region. The ground reference data are derived from in situ measurements available from the European Center for Medium-range Weather Forecasts (ECMWF) database.

2 The POSEIDON weather forecasting system

The hydrostatic SKIRON/Eta model that constitutes the core NWP model of the POSEIDON-I weather forecasting system (Papadopoulos et al., 2002) is a limited area model requiring appropriate data to define its initial and boundary conditions (IC and BC). For this purpose, the World Aviation Forecast System (WAFS) form of the NCEP Aviation (AVN) model analysis and up to 72 h forecast data with 6-hourly interval were downloaded daily from NCEP. The WAFS/AVN model output were available with horizontal resolution of $1.25^{\circ} \times 1.25^{\circ}$ and a vertical resolution of 11 levels, which were quite coarse for the needs of the POSEIDON operations. To compensate for this, a one-way nesting technique was developed. Nesting within a domain, rather than simply running the higher resolution model using directly the WAFS data, offers advantages in the resolution and time updating of its lateral BC. The first model run (COARSE) was carried

out on a grid mesh with a resolution of $1/4^{\circ} \times 1/4^{\circ}$ (~ 25 km) and 32 vertical levels up to 100 hPa ($\sim 15,8$ km). In this simulation the WAFS analysis and the 6-hourly WAFS forecasts provided the IC and BC, respectively. In addition, the prediction of the dust cycle was obtained through the integration of the COARSE model. The second run (POS-1) was performed using a resolution of $1/10^{\circ} \times 1/10^{\circ}$ (~ 10 km) and the same vertical structure as in the COARSE model configuration, while the IC and BC were supplied from the COARSE model output. Therefore, the POS-1 model was initialized by the $1/4^{\circ} \times 1/4^{\circ}$ COARSE data and updated its BC every hour. For the sea surface temperature (SST), the daily NCEP SST data of $1^{\circ} \times 1^{\circ}$ were used. For both runs, the simulation period was 72 h. The system has been fully operational since 1999 and it had been designed to operate in an optimum way; while the COARSE model was running, the pre-processor of the POS-1 model was preparing the appropriate fields to feed the integration of the POS-1 model. The geographic areas covered by the two POSEIDON-I runs (COARSE and POS-1) are shown in Fig. 1.

Periodic improvements to the POSEIDON-I weather forecasting system have been made since then, including modifications and better tuning of the physical parameterization packages of the NWP model. However, the major upgrade of the system was carried out through the POSEIDON-II project (2005–2008). Recent advances offer the opportunity to achieve significant improvements of the NWP model forecasting skill. In general, the increase in computational resources supports the higher resolution operational model runs, the more detailed treatment of physical processes and the implementation of advanced numerical techniques for assimilating available observations. However, high resolution simulations allow the representation of processes for which the hydrostatic assumption ceases to be valid. For this reason, the formulation and the implementation of a non-hydrostatic NWP model was considered a matter of priority for the needs of the POSEIDON forecasts. In this end, a new generation weather forecasting system has been developed (based on the latest non-hydrostatic version of the SKIRON model). The physics of the SKIRON/Eta system were upgraded including non-hydrostatic dynamics through an add-on module as suggested by Janjic et al. (2001). Furthermore, a mesoscale 3-D meteorological data assimilation package, the Local Analysis Prediction System (LAPS), has been implemented to produce high resolution analysis fields (Albers, 1995). LAPS uses near-to-analysis forecasts obtained from NCEP's Global Forecast System (GFS) on $0.5^{\circ} \times 0.5^{\circ}$ horizontal grid increment to generate a 3-D first guess (background) field. It employs all available real-time observations obtained from METAR (MÉTéorologique Aviation Régulière; aviation routine weather reports typically generated once an hour), SYNOP (Surface sYNOPtic) observations and RAOBS (RADiosonde OBServations) and finally generates a realistic, fine resolution analysis ($1/6^{\circ} \times 1/6^{\circ}$). On a daily basis, using approximately 650 METAR hourly

reports, 40 SYNOP 3-hourly observations and 15 RAOBS twice per day, LAPS is able to resolve high frequency atmospheric features resulting in a better definition of the IC of the NWP model. For the BC the $0.5^\circ \times 0.5^\circ$ GFS/NCEP global forecasts are used. The new version has been tested extensively during 2007 and the final configuration is applied on a horizontal resolution of $1/20^\circ \times 1/20^\circ$ (~ 5 km) over a domain that covers the Mediterranean basin, Black Sea region and the main part of Europe and North Africa (depicted as POS-2 domain in Fig. 1). In the vertical, 50 levels are available up to 25 hPa (~ 25 km). It uses higher resolution NCEP SST data ($1/2^\circ \times 1/2^\circ$) and high resolution snow depth and ice cover analysis data, which are updated on a daily basis. The simulation period has been extended to 114 h (almost 5 days).

The change of the operational status took place in November 2008, when the POSEIDON-I weather forecasting system ceased to be operational and the POSEIDON-II became the new operational weather forecasting system of HCMR.

3 Data and methodology

This study focused on the performance assessment of the higher-resolution POSEIDON-I (10 km grid spacing; POS-1) and the POSEIDON-II (5 km grid spacing; POS-2) weather forecasts using as reference the surface measurements available from the WMO network. The comparison of the POSEIDON weather forecasts against observations was made across the Eastern Mediterranean where the available surface stations are depicted in Fig. 1. Surface observations from more than 500 conventional stations were used to verify and compare categorical model forecasts of the 10-m wind field, 2-m air temperature and sea level pressure every 3 h and the accumulated 6-h precipitation every 6 h for the pre-operational period (from mid-November 2007 to October 2008). Quality control has been applied to remove erroneous measurements, based on checking the physical range of each parameter being verified, the allowable rate of change in time and the stationarity. Despite the known issues associated with comparing point measurements with area-averaged estimates, the measurements from this network are valuable for the study due to their coverage and the continuous recording.

Based on the operational procedures, the POS-1 model was integrated for 72 h initialized at 12:00 UTC each day, while the POS-2 model run using the LAPS analysis at 18:00 UTC with a forecast window of 114 h ahead. In this study, the goal is to compare the forecasts produced by the two operational systems directly by using the same verification hours and observations, rather than computing the statistics as a function of the forecast length of each operational cycle. Therefore, to validate the forecasts of 10-m wind, 2-m air temperature and sea level pressure for the verification period (15 November 2007–31 October 2008)

the 9–72-h POS-1 forecast and the 3–66-h POS-2 forecast were used to match each 3-hourly verification hour, from 21:00 UTC of day 0 through 12:00 UTC of day 3 (22 verification hours). Thus, the 9-h POS-1 forecast and the 3-h POS-2 forecast were compared to the observed variables recorded at 21:00 UTC of day 0 (the day of the initialization), the 12-h POS-1 forecast and the 6-h POS-2 forecast were compared to the observations recorded at 00:00 UTC of the day 1, and so on. Likewise, for the accumulated 6-h precipitation, the 12–72-h POS-1 forecast and the 6–66-h POS-2 forecast were also compiled in the relevant statistical methods. Thus, the model-generated precipitation for the 12–18-h POS-1 forecast period and for the 6–12-h POS-2 forecast period were compared to the observed precipitation accumulated between 00:00 and 06:00 UTC of the day 1, the 18–24-h POS-1 and the 6–12-h POS-2 model precipitation accumulation were compared to the observed precipitation accumulated between 06:00 and 12:00 UTC of the day 1, and so on. These gridded forecasts were interpolated to each station location using bilinear interpolation and more than 110 000 pairs of forecasts and observations have been produced for each forecasting system and for each verification hour. Since fewer stations provide observations at night, the verification pairs (forecasts and observations) were increased up to 130 000 for the daytime.

The evaluation methodology was based on the point-to-point comparison between model-generated variables and observations. For the variables of wind speed, air temperature and sea level pressure the scores produced are the standard mean error (BIAS) and the root mean square error (RMSE). For the wind direction a different approach was considered, based on correctly counting the model direction if it ranged 22.5° from the observed value. Thus, the directional forecast skill score was computed by $D=E/N$, where E was the number of correct estimations and N the total number of observations. The verification scores used for the precipitation were derived using the contingency table approach (Wilks, 1995). This is a two-dimensional matrix where each element counts the number of occurrences in which the gauge measurements and the model forecasts exceeded or failed to reach a certain threshold for a given forecast period. The table elements are defined as: A-model forecast and gauge measurement exceeded the threshold; B-model forecast exceeded the threshold but measurement not; C-model forecast did not reach the threshold but measurement exceeded it; and D-model forecast and measurement did not reach the threshold. Considering the above elements the forecast skill can be measured by evaluating the bias score (BS) and the equitable threat score (ETS). The bias score is defined as $BS=(A+B)/(A+C)$, while the ET score defined as $ETS=(A-E)/(A+B+C-E)$ where E is defined by $E=[(A+B)-(A+C)]/N$, with N holding the total number of observations being verified ($N=A+B+C+D$). The introduction of the E term (Mesinger, 1996) is an enhancement to the normal threat score (as defined in Wilks, 1995); since it reduces

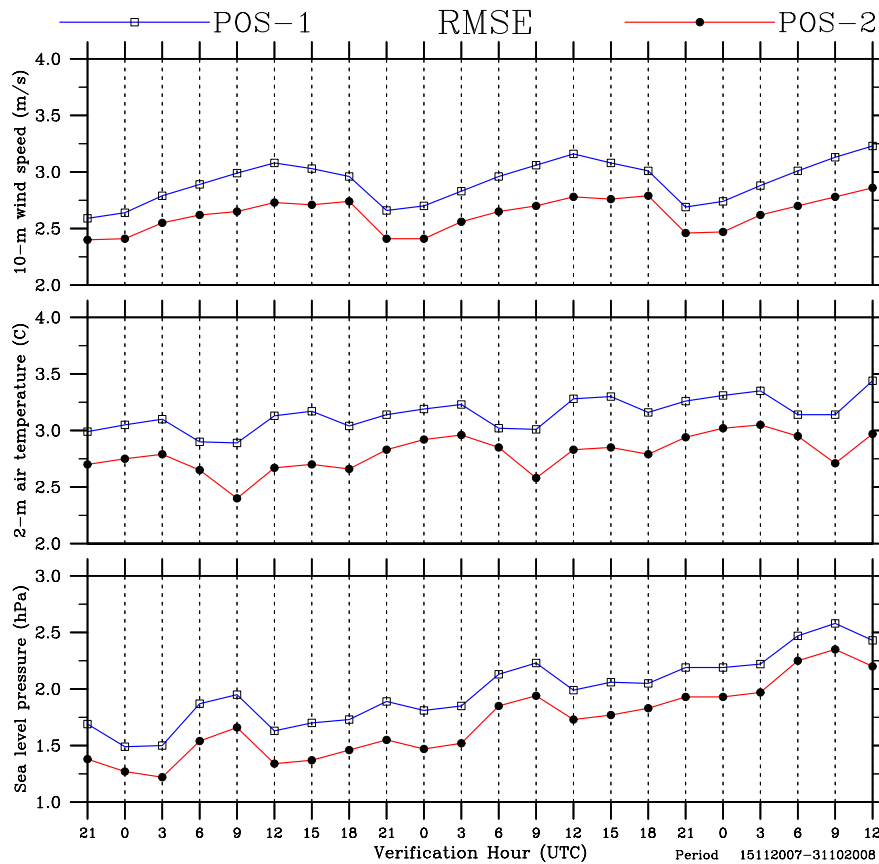


Fig. 2. The root mean square error (RMSE) of 10-m wind speed, 2-m air temperature and sea level pressure for the POS-1 (blue line) and POS-2 (red line) as a function of verification hour.

it by excluding the number of randomly forecast “hits”. To measure the magnitude of the difference between model forecast and observed precipitation the root mean square error (RMSE) was also calculated as follows (Colle et al., 2000; Mass et al., 2002):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{NOBS} (MP_i - OP_i)^2}{NOBS}} \quad (1)$$

where MP_i and OP_i are the model estimated and the observed precipitation, respectively, and the NOBS is the total number of observations at a specific location reaching or exceeding a certain threshold amount. The aforementioned statistical criteria have been combined in order to provide a comprehensive evaluation of model performance and comparison. For example, a greater ETS will represent a significant model improvement only if it is accompanied by a BS with value closer to one and a lowering RMSE.

4 Discussion of the verification results

Surface statistics of the forecast errors for the two POSEIDON weather forecasting systems are displayed in Figs. 2, 3, 4 and 5.

Figure 2 displays the RMSE of the fields of 10-m wind speed, 2-m air temperature and sea level pressure, respectively. For both sets of forecasts there is a slow increase of the RMS error with the time. On this error, a diurnal signal is evident for the wind speed and the near-surface air temperature, while for the sea level pressure a periodic component is observed during the early morning hours. In general, the POS-2 RMS errors are lower (better) than the POS-1 errors, ranges from 9 to 13% for the wind speed (top panel), about 6–18% for the air temperature (middle panel) and reaches 20% for the sea level pressure (bottom panel). Additionally, Fig. 3 shows the mean differences (or bias) for the three verified surface parameters. It is noted that POS-2 air temperature (middle panel) is associated with significant lower biases compared to the POS-1 forecasts. Moreover, POS-2 sea level pressure (bottom panel) forecast error properties reveal a decreasing trend of the significant semidiurnal

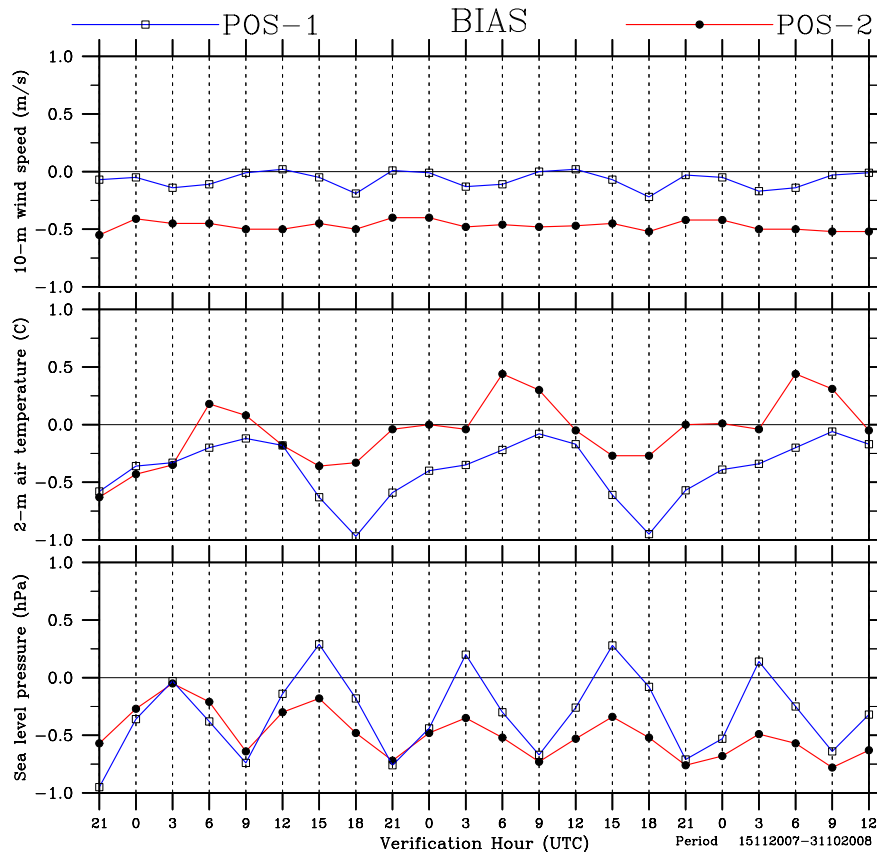


Fig. 3. The standard mean error (BIAS) of 10-m wind speed, 2-m air temperature and sea level pressure for the POS-1 (blue line) and POS-2 (red line) as a function of verification hour.

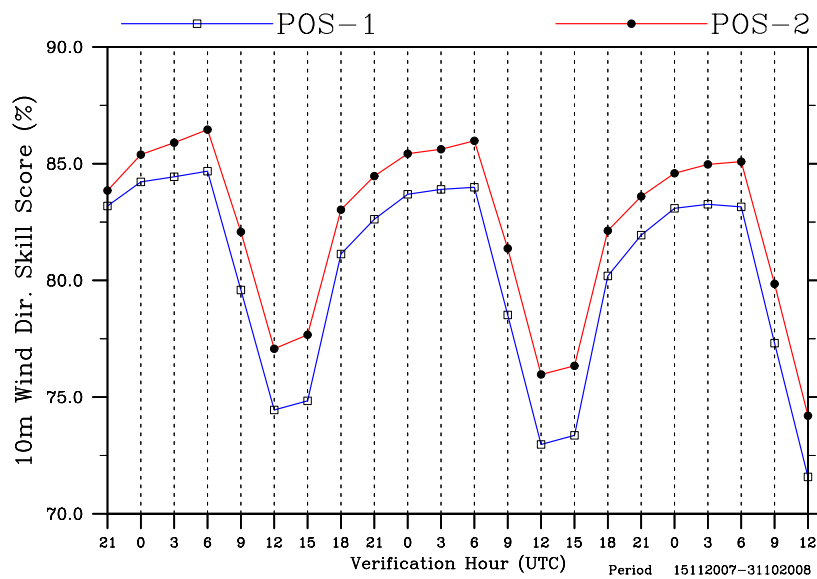


Fig. 4. The directional forecast skill score of 10-m wind field as a function of verification hour.

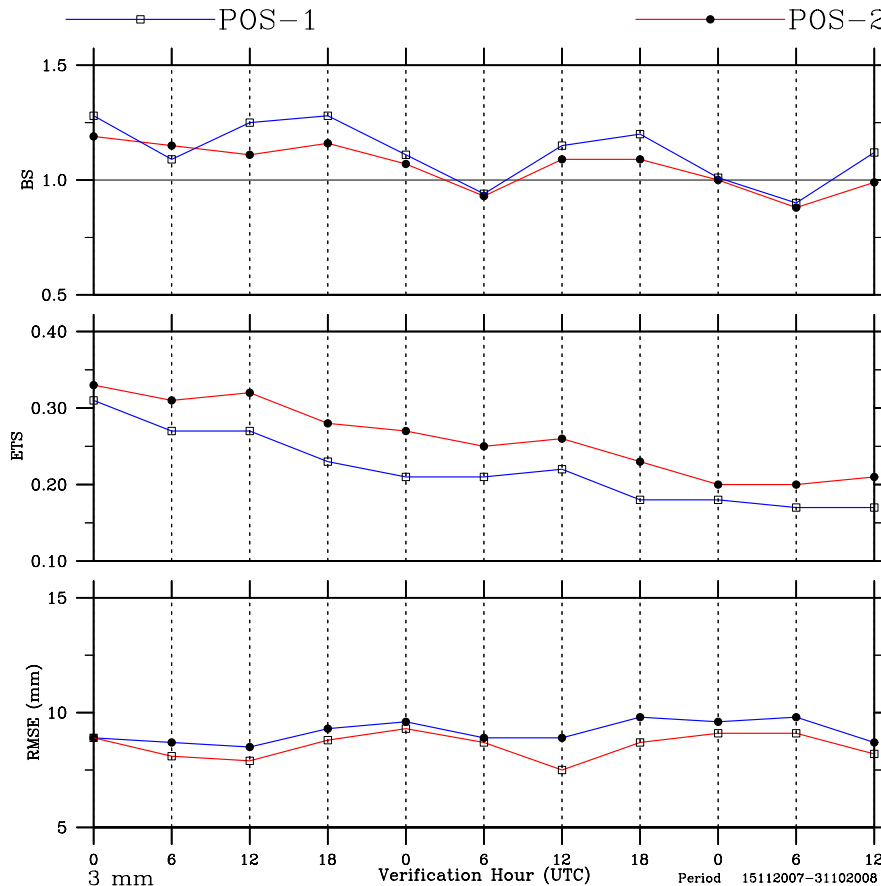


Fig. 5. The 6-h bias scores, 6-h equitable threat scores, and 6-h root mean square errors at the 3 mm threshold for the POS-1 (blue line) and the POS-2 (red line) precipitation forecasts as a function of verification hour.

variation that characterized the POS-1 forecasts, even though a negative bias still remains. Figure 3 also indicates that the POS-1 wind speed bias (top panel) has a bias close to zero while the POS-2 wind speed has almost a slight negative bias (~ 0.5 m/s). Furthermore, Fig. 4 shows that the wind directional forecast skill score from the 5-km (POS-2) wind directions is better than that calculated from the 10-km grid (POS-1) products. Also, there is a noticeable diurnal variation in which wind direction errors increase during midday (12:00–15:00 UTC) when terrain-induced and thermal circulations are more pronounced.

In addition to these surface parameters, precipitation fields produced by both weather forecasting systems were also verified. The time-dependent bias and equitable threat scores based on the contingency table were computed at 6-h intervals during the period from mid-November 2007 to October 2008. Figures 5 and 6 present the statistics for a low (3 mm) and a high (12 mm) threshold, respectively. As shown in Fig. 5, the 6-h bias scores are comparable and indicate a small overestimation for both systems. Nevertheless, the POS-2 bias scores are lower than those of the POS-1

forecasts and vary closer to 1 (nearly un-biased). The POS-2 has also better bias scores for the higher threshold (as depicted in Fig. 6). In agreement with the 6-h bias scores, the 6-h equitable threat scores for the POS-2 forecasts are generally greater (10–25% improvement for the lower and 10–50% for the higher threshold) than those from the POS-1 forecasts. This suggests the importance of terrain resolution and use of non-hydrostatic approach for capturing higher orographic precipitation systems. For the lower threshold, the rms errors show that the POS-2 has a steady smaller quantitative error. On the other hand, the rms errors for the higher threshold indicate that even though the POS-2 quantitative precipitation forecasts are overall better than the POS-1, during the first half of the verification period exhibit larger rms errors.

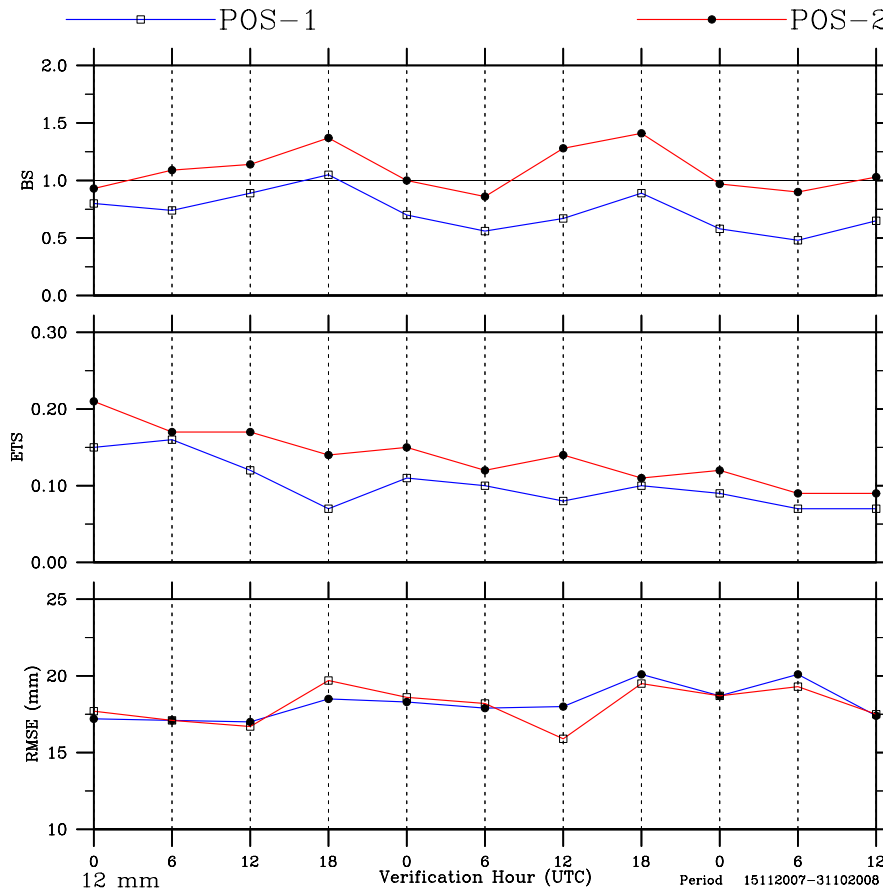


Fig. 6. The 6-h bias scores, 6-h equitable threat scores, and 6-h root mean square errors at the 12 mm threshold for the POS-1 (blue line) and the POS-2 (red line) precipitation forecasts as a function of verification hour.

5 Conclusions

The goal of this study was to evaluate and compare the performance of the 10-km/hydrostatic (POS-1) versus 5-km/non-hydrostatic (POS-2) POSEIDON forecasts as produced during the pre-operational period (from mid-November 2007 to October 2008). The assessment was performed using as reference surface data from conventional weather observing stations across Europe. On the basis of traditional objective verification techniques (like bias, RMSE, threat scores) preliminary results show within the 95% confidence level that the combined effect of the new model parameterization schemes and the advanced numerical techniques for assimilating available observations enhance the accuracy of POSEIDON forecasts. This is in agreement with the notion that smaller horizontal grid spacing produce better forecasts (e.g. Mass et al., 2002). However, the negative bias of POS-2 wind speed forecasts implies that the new system tends to predict weaker winds, even though exhibiting a better RMS error than POS-1. This behavior indicates the limitations that are imposed on the standard

verification techniques when applied over yearlong periods. Such techniques can be valuable to forecasters but the overall synopsis of the traditional verification metrics can hide the added value of the high resolution forecasts for highly significant events; because such information is wiped out by the very large number of ordinary forecast situations. Moreover, they provide little information on the timing errors and structures of transient meteorological phenomena. Among others, Colle et al. (2001) suggest that the advantage of the forecast to resolve meso- α and meso- β features (i.e. orographically forced circulations or sea breezes) will likely be poorly scored by traditional verification methods, because such metrics sharply penalize forecasts with small temporal or spatial errors of predicted features. Therefore, feature- and event-based verification approaches have been examined (e.g. Colle et al., 2001; Rife and Davis, 2005). Future work will include event-based approaches for the verification of temporal and spatial objects to further investigate the variation of forecast quality provided by the new POSEIDON weather forecasting system.

Acknowledgements. This study was supported by the POSEIDON-II project funded by the European Economic Area (EEA) Financial Mechanism and the Greek Ministry of National Economy and by the PreWEC project (MEXT-CT-2006-038331) funded by the European Union and the Greek General Secretariat for Research and Technology. ECMWF is also acknowledged for the provision of surface observational data from the WMO network.

Edited by: S. Michaelides, K. Savvidou, and F. Tymvios

Reviewed by: I. Pytharoulis and another anonymous referee

References

- Albers, S. C.: The LAPS wind analysis, *Weather Forecast.*, 10, 342–352, 1995.
- Colle, B. A., Mass, C. F., and Westrick, K. J.: MM5 precipitation verification over the Pacific Northwest during 1997–1999 cool seasons, *Weather Forecast.*, 15, 730–744, 2000.
- Colle, B. A., Mass, C. F., and Ovens, D.: Evaluation of the Timing and Strength of MM5 and Eta Surface Trough Passages over the Eastern Pacific, *Weather Forecast.*, 16, 553–572, 2001.
- Janjic, Z. I.: Non-linear advection schemes and energy cascade on semi-staggered grids, *Mon. Weather Rev.*, 112, 1234–1245, 1984.
- Janjic, Z. I.: The step-mountain Eta coordinate model: Further developments of the convection, viscous sublayer and turbulence closure schemes, *Mon. Weather Rev.*, 122, 927–945, 1994.
- Janjic, Z. I., Gerrity Jr., J. P., and Nickovic, S.: An Alternative Approach to Nonhydrostatic Modeling, *Mon. Weather Rev.*, 129, 1164–1178, 2001.
- Kallos, G., Nickovic, S., Papadopoulos, A., Jovic, D., Kakaliagou, O., Misirlis, N., Boukas, L., Mimikou, N., Sakellaridis, G., Papageorgiou, J., Anadranistakis, E., and Manousakis, M.: The regional weather forecasting system SKIRON: An overview, in: *Proceedings of the Symposium on Regional Weather Prediction on Parallel Computer Environments*, Athens, Greece, 15–17 October 1997, 109–122, 1997.
- Mass, C. F., Ovens, D., Westrick, K. J., and Colle, B. A.: Does increasing horizontal resolution produce more skillful forecasts?, *B. Am. Meteorol. Soc.*, 83, 407–430, 2002.
- Mesinger, F.: Improvements in quantitative precipitation forecasts with the Eta regional model at the National Centers for Environmental Prediction: The 48-km upgrade, *B. Am. Meteorol. Soc.*, 77, 2637–2649, 1996.
- Mesinger, F., Janjic, Z. I., Nickovic, S., Gavrilov, D., and Deaven, D. G.: The step-mountain coordinate: Model description and performance for cases of Alpine lee cyclogenesis and for a case of an Appalachian redevelopment, *Mon. Weather Rev.*, 116, 1493–1518, 1988.
- Nickovic, S., Kallos, G., Papadopoulos, A., and Kakaliagou, O.: A model for prediction of desert dust cycle in the atmosphere, *J. Geophys. Res.*, 106, 18113–18129, 2001.
- Papadopoulos, A., Katsafados, P., Kallos, G., and Nickovic, S.: The weather forecasting system for POSEIDON – An overview, *GAOS*, 8(2–3), 219–237, 2002.
- Rife, D. L. and Davis, C. A.: Verification of Temporal Variations in Mesoscale Numerical Wind Forecasts, *Mon. Weather Rev.*, 133, 3368–3381, 2005.
- Wilks, D. S.: *Statistical methods in the atmospheric sciences: An introduction*, Academic Press, San Diego, California, 467 pp., 1995.