Spatial verification techniques applied to high resolution models for an intense precipitation summer event in Greece

Presented in the: 14th International Conference on Meteorology,Climatology and Atmospheric Physics
October 15-17, 2018 Alexandroupolis, Greece

Boucouvala D.1*, Gofa F1., Samos I1

1 Hellenic National Meteorological Service, Hellinikon GR-16777, Athens, Greece

1 Introduction

Traditional precipitation verification metrics based on point-to-point comparison without providing information regarding spatial distribution are insufficient to evaluate precipitation forecasts, especially from high resolution mesoscale models. For example, when a small scale feature (also defined as an object) is correctly forecast but slightly displaced in time and space, the forecast will be penalized both for a miss and a false alarm (double penalty), especially for high resolution datasets (Ebert 2008).

Spatial verification methods that allow for some tolerance to reasonably small errors in space and time tend to resolve this problem (Cassola et al. 2015). The two main categories are: neighbourhood (or fuzzy) verification methods (Ebert 2008) based on a scale-dependent verification approach where the requirement for exact matches between forecasts and observations is relaxed and object oriented techniques which deal on with how well the forecast captures the overall structure of meteorological features by identifying and comparing precipitation features in the forecast and observations (Ebert and McBride 2000).

The aim of this study is to evaluate the relative model performance of the operational Numerical Weather Prediction (NWP) systems of the Hellenic National Meteorological Service (HNMS) (with different horizontal resolutions) for a rare summer precipitation event that affected almost the entire area of Greece by applying spatial verification methods.

2 Data and Methodology

An unusually strong precipitation event that occurred on the 16-17th of July 2017 was selected as a test case. The event, which was a combination of both dynamic and convective activity was accompanied by relatively low temperatures for the season and affected a large part of the country, causing hailstorms, flooding, property damage and unfortunately loss of human life. The event was preceded by a series of relatively warm days with 850hPa temperatures around 15-20°C. On 17/04 00UTC, a trough centered over Russia covering all of Eastern Europe moved southwards toward Greece, resulting in cold air masses (-15°C) at 500hPa (Fig 1a) moving slowly E-NE. The trough was accompanied by a low pressure system at the surface, which moved from west to east (Fig 1b). Initially, convective precipitation was observed over northern and western Greece which extended to the central and eastern parts of the country by the afternoon. This was accompanied by lightning (Fig 1c) and hail at several locations on the mainland.

2.1 Data

Spatial verification techniques require data defined continuously over a common spatial domain covering the area of interest. 3-hourly cumulative HSAF (EUMETSAT Satellite Application Facility on Support to Operational Hydrology and Water Management) gridded observations and forecasts from: a) the global scale ECMWF (IFS) model with a horizontal resolution of 9km and b) the local model COSMO-GR (Gofa et al. 2008) with horizontal resolutions of 7, 4 and 1km were used.

The data were regridded (interpolated or extrapolated) to a common grid spacing of 0.06° (4km) in order to facilitate comparison. This grid spacing is also consistent with the spatial frequency of observations. Despite

doi:10.5676/dwd_pub/nwv/cosmo-nl_19_07

the smoothing effect associated with upscaling, the high resolution model configuration preserves details of the precipitation structure while also featuring larger magnitude departures from the observations at some locations. This can possibly be attributed to the coarser resolution of the initial observed precipitation field.

2.2 Methodology

Neighborhood verification (or fuzzy) techniques evaluate forecast performance using more elastic conditions regarding the exact spatio-temporal match between observed and forecast fields. It is based on the principle of expanding the area of comparison to include data points nearby (“neighbors”), employing a spatial window, or "neighborhood", surrounding the forecast and/or observed points. A relaxing filter can be applied to both fields, and the penalty for differences between modeled and observed values is relaxed. The properties of the relaxed fields (mean values, maximum values, number of grid points exceeding a threshold) can then be compared using traditional statistical methods.

The size of this window starts at the smallest possible scale (neighborhood of one grid box) and is gradually increased in order to provide insight into the scales at which the model has the most skill. The method shows how forecast skill varies with neighborhood size and can be used to determine the smallest neighborhood size that provides a sufficiently skillful forecast. A more detailed review of neighborhood approaches is available in Ebert (2008). There is a variety of methods that fall within this category, differentiated by their treatment of the points within each window, depending on the neighborhood method used. In order to determine if a forecast is “useful” or “good enough”, decision models are applied such as: Upscaling, Minimum Coverage, Anywhere in the Window and Practically Perfect Hindcast. Traditional dichotomous scores are then calculated.

The Fractions Skill Score (FSS) is a decision model based on the comparison of frequency of forecast and observed events. In this study, the VAST (VERSUS Additional Statistical Techniques) software package, which was developed by the COSMO consortium and offers a number of neighborhood verification tools, was employed (Gofa at al. 2018). SAL (which stands for Structure, Amplitude and Location) is an object-based method developed by Wernli et al. (2008, 2009) to measure the quality of a forecast by identifying objects in both forecast and observed fields at a given time and provide information on object shape and location differences between the two fields. The score consists of three components which correspond to a global field measure of: Structure (S), Amplitude (A) and Location (L). The S parameter compares the volume of the normalized precipitation objects of the two fields. Positive S values indicate that modeled precipitation objects are too large or too flat (more stratiform precipitation), while a negative value indicates that objects are too sharp and too small (more convective type precipitation).

The A parameter represents the normalized difference of the domain-averaged precipitation fields and is independent of structural features. Positive (Negative) values of the A parameter indicate overestimation (underestimation) of total domain precipitation. The range of the S and A parameters is [-2, 2]. The L component combines information about the distance of predicted and forecast mass centers (L1) and the normalized distance between the precipitation objects (L2). L ranges from 0 to 2. A perfect forecast is characterized by zero values for all three SAL components. The S and L (specifically the L2 component) parameters require the identification of objects in observed and forecast fields. An object is defined when it exceeds a fixed or statistically defined threshold value. Wernli et al. (2009). Consequently, if no features are found in either or both forecast and observed fields, the SAL values cannot be defined. The SAL parameters are calculated with a SpatialVx based software package (Gilleland 2017).

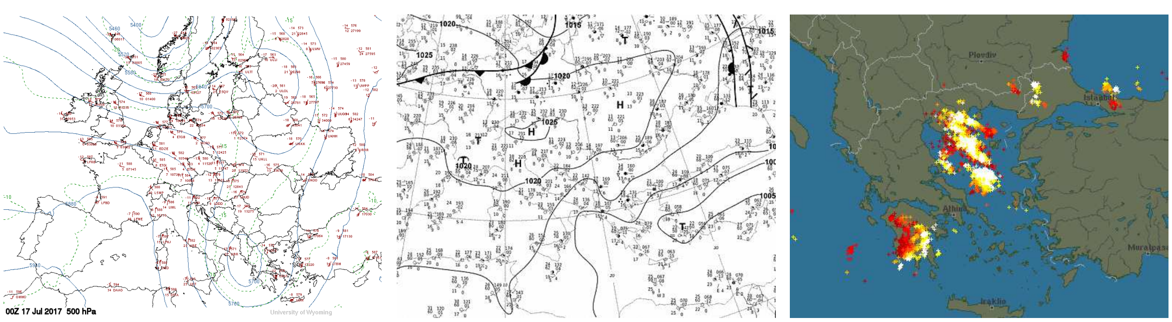

Figure 1: (left): 17/04 00UTC 500hPa (source: University of Wyoming), 1b. (center): Surface Analysis (source: UK Met Office), 1c. (right) Lightning activity at 10UTC (http://el.blitzortung.org) with dots indicating the location of lightning strikes where the color refers to the age of the strike (20 min intervals).

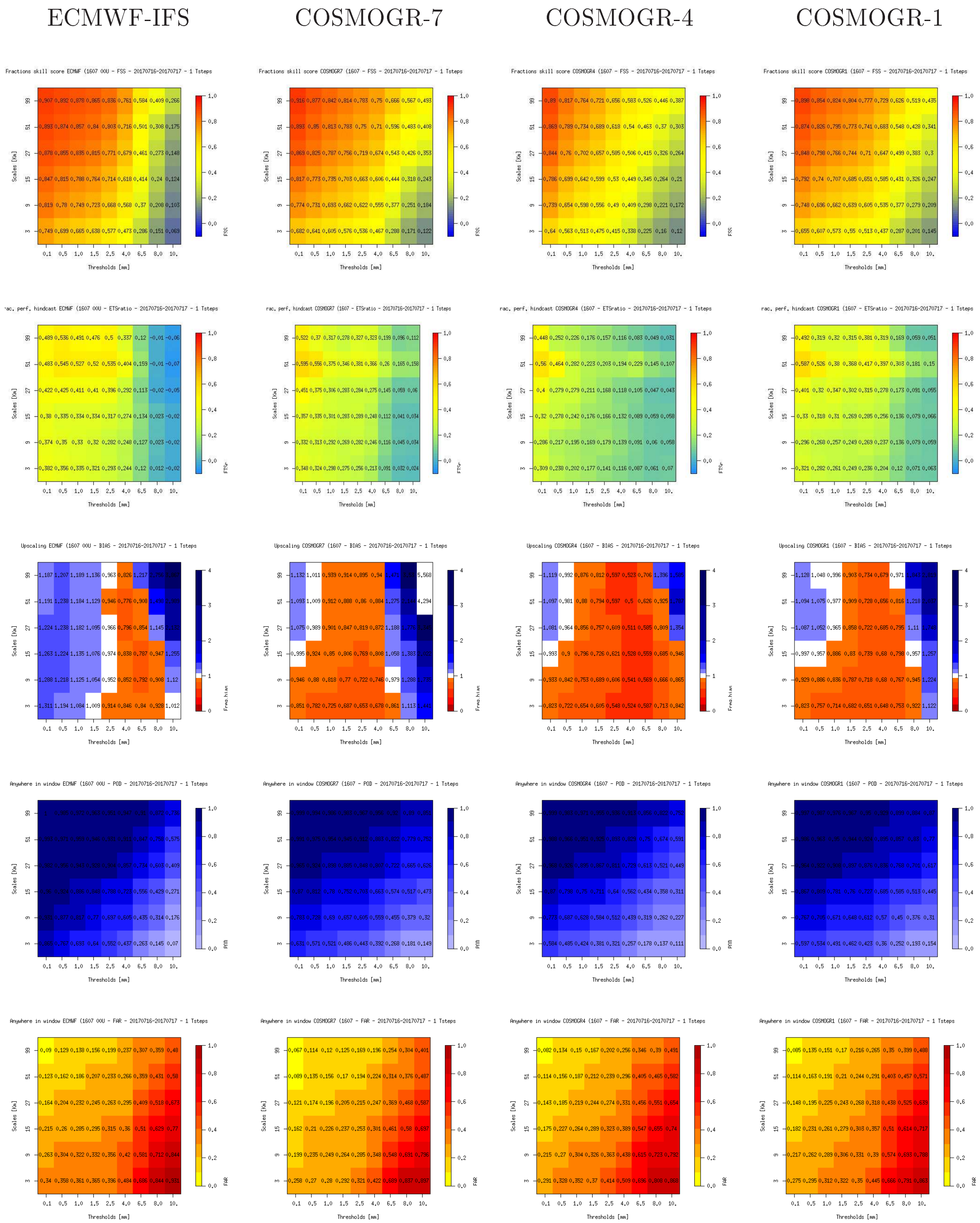

Figure 2: Neigborhood method plots for lead time 16/07 00 UTC derived for the various model setups. From top to bottom: FSS (Fractions Skill Score), ETS (Equitable Thread Score), Bias, POD (Probability of Detection) and FAR (False Alarm Rate)

3 Results

A selection of the evaluation plots for July 17th applying neighborhood methods to the various models and resolutions is presented in Fig 2. (time lead 1600 UTC). The scores are plotted as intensity-scale diagrams, where the intensity threshold and spatial scale averaging increase along the x and y axes respectively, and the color shade gives an indication of the value of the score (also plotted explicitly). By evaluating the color intensity (darkness), scales and thresholds at which a particular model system performs best, it is possible to evaluate model performance without focusing on the absolute value of each colored window. The forecast skill (as represented by the FSS score) does not differ significantly between models, but it does increase as window size (<15km) and precipitation thresholds (<3mm/3h) decrease. For high precipitation thresholds, on the other hand, forecast skill decreases.

ETS (Equitable Thread Score) index diagrams (Practically perfect Hindcast decision method) show that the forecast quality is better for window sizes <50km and thresholds 0.1-0.2mm. The indices for COSMO-GR1 and COSMO-GR4 are slightly better than those of ECMWF-IFS. However, significant differences appear in the Bias score (upscaling method) as ECMWF-IFS has the tendency to overestimate both the low thresholds (0.1-3mm) and high thresholds (>10mm/3h) while underestimating the remaining thresholds. The COSMO model generally overestimates rainfall for windows up to 27km for all thresholds, except for COSMO-GR7 which underestimates only the high thresholds. The POD (Probability of Detection) and FAR (False Alarm Rate) (calculated using the Anywhere in the Window method) show that ECMWF-IFS had more successful hits (dark red) but also more false alarms (dark blue). SAL parameters for the 24h accumulated precipitaion for July 17th are estimated for the 1600 UTC model run with different fixed thresholds (from lower to higher) (Fig 3).

The positive S parameter indicates that flatter objects (more stratiform precipitation) are calculated by the models for higher thresholds, while sharper objects (more convective) are produced at lower thresholds. COSMO7 predicts flatter objects versus sharper objects by COSMO1. The L parameter is constant and lower for ECMWF-IFS, while higher values are calculated for COSMO4. COSMO7 S values tend to be lower for higher precipitation thresholds. The A parameter, which is independent of objects and depends on the entire field, is positive, which means that for all models, especially for COSMO7, 24h precipitation is overestimated.

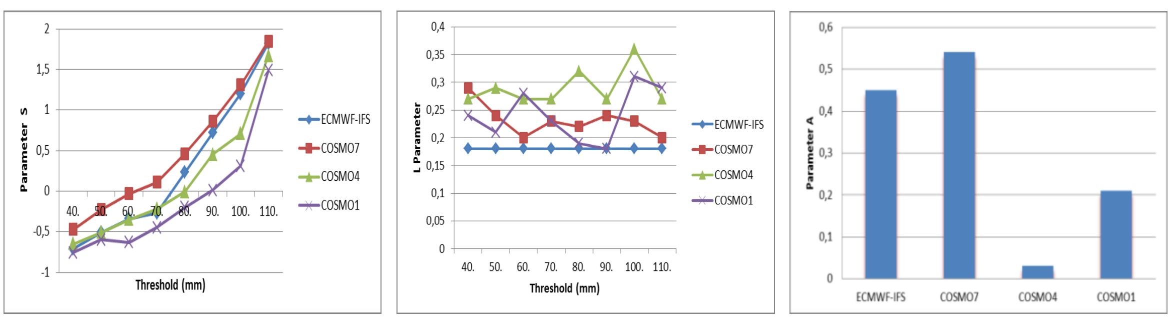

Figure 3: Left:S (Structure), Center: L(Location) parameters with threshold, Right: A (Amplitude) parameter for ECMWF-IFS, COSMOGR-7, COSMOGR-3 and COSMOGR-1.

4 Conclusions

The aim of this study was to compare the metrics of two spatial verification methods applied to the case of an intense summer precipitation event. Neighborhood verification results showed that for high rainfall rate thresholds and large spatial windows, the forecast skill and quality decreased for all models used in the study. Differences between the COSMO and ECMWF-IFS models at different scales and thresholds are mainly evident in Bias and ETS scores, with the latter model tending to overestimate precipitation for low thresholds and consequently producing more false alarms. Application of the SAL object-based method to 24h precipitation forecasts showed that finer resolution models led to prediction of sharper objects, that all models overestimate domain precipitation while location errors are more variable with threshold for finer resolution models. These results confirm that, when combined with traditional verification techniques, spatial verification methods enable more detailed and more complete assessment of model performance.

References

[1] Cassola F, Ferrari F, Mazzino A (2015) Numerical simulations of Mediterranean heavy precipitation events with the WRF model: analysis of the sensitivity to resolution and microphysics parameterization schemes. Atmos. Res. **164–165, 210–225**.

[2] Ebert E (2008) Fuzzy verification of high resolution gridded forecasts: A review and proposed framework. Meteorol. Appl. 15, **51-64**.

[3] Ebert E, McBride JL (2000) Verification of precipitation in weather systems: Determination of systematic errors. J. Hydrol. 239, **179-202**.

[4] Gilleland E (2017) *R package Version 0.6-1.***https://cran.r-project.org /package =SpatialVx**

[5] Gofa F, Pytharoulis I, Andreadis T, Papageorgiou I, Fragkouli P, Louka P, Avgoustoglou E, Tyrli V (2008) Evaluation of the operational numerical weather forecasts of the Hellenic National Meteorological Service. Proc. 9th COMECAP Conference of Meteorology, Thessaloniki, Greece, **51-58**.

[6] Wernli H, Hofmann C, Zimmer M (2009) Spatial Forecast Verification Methods Intercomparison Project: Application of the SAL Technique. Wea. Forecasting. 24, **1472–1484**.

[7] Wernli H, Paulat M, Hagen, Frei C (2008) SAL-A novel quality measure for the verification of quantitative precipitation forecasts. Mon. Wea. Rev. 136, **4470–4487**.