

Lavoro pervenuto in redazione il 02/02/2009

# Use of exploratory spatial data analysis and geostatistics to verify the skill of atmosphere forecast models with improved data initialisation schemes

*Charles Galdies\* & D. N. M. Donoghue\*\**



**Sommario** - La capacità di previsione di un modello numerico dell'atmosfera con uno schema di inizializzazione migliorato è stato comparato con le osservazioni da satellite utilizzando nuovi metodi di verifica diagnostico-spaziale. Questi metodi utilizzano schemi per la misura della similarità/dissimilarità e dell'accordo delle caratteristiche spaziali tra previsioni e osservazioni utilizzando il processamento di immagini e l'analisi geostatistica. Nel presente studio l'analisi spaziale mostra un incremento della similarità tra le previsioni generate dal modello atmosferico ad alta risoluzione Eta e le osservazioni da satellite acquisite per mezzo

**Summary** - The forecasting skill of a numerical atmosphere model with an improved initialisation scheme was evaluated against remote sensing observations using new spatial diagnostic verification methods. These methods use schemes to measure similarity/dissimilarity and spatial matching features between forecast and observations using image processing and geostatistical analysis. In this study spatial analysis showed an enhanced similarity between the experimental forecasts produced by a high resolution Eta atmosphere model and collocated remote sensing observations acquired by the orbiting Tropical Microwave



del Tropical Microwave Imager a bordo del satellite Tropical Microwave Rainfall Measuring Mission (TRMM). Questo studio consente di effettuare la validazione dei modelli numerici da un nuovo punto di vista utilizzando il concetto di analisi esplorativa spaziale dei dati e geo-statistica per la validazione dei modelli numerici. Questo tipo di approccio non era stato mai applicato per verificare i miglioramenti apportati ai modelli atmosferici di previsione numerica. D'altra parte, la geo-statistica viene utilizzata piuttosto comunemente per lo studio e la derivazione delle distribuzioni, dei campioni spaziali e dell'analisi strutturale dei fenomeni naturali.

Imager on board the Tropical Microwave Rainfall Measuring Mission (TRMM) satellite. This study approaches numerical model validation from a new angle by using the concepts of exploratory spatial data analysis and geo-statistics for numerical model validation. This approach has never been applied to verify improvements made to numerical atmosphere models. On the other hand, geostatistics is a fairly common approach to study and derive the distribution, spatial patterns and texture analysis of natural phenomena.

## 1. Introduction

The positive impact of assimilating SST data acquired by passive microwave remote sensing into a high resolution, one-way coupled primitive equation high resolution atmosphere-ocean model has been recently demonstrated (Galdies & Donoghue, 2009). The advantage of assimilating remotely sensed sea surface observation is attributed to their almost continuous data information over wide geographic areas and with significant meso-scale resolution. Remotely sensed SST data provides an opportunity to initialise the surface condition of high-resolution atmosphere models on a continuous and quasi-synoptic basis. The assimilation of high-resolution SST data within the prognostic equations of dynamic models, such as atmosphere models, can profoundly influence the accuracy of modeled/forecasted air-sea fluxes (Gill, 1982). Geostatistics and exploratory data spatial analysis are often found within the realm of geographic information system (GIS). This approach is nowadays used multidisciplinary studies but so far has been rarely used for the verification analysis of geophysical models. These methods can well complement the application of the conventional statistical measures so far used by the numerical weather prediction community. They offer a way to understand the spatial structure of residual fields and can translate this variation in numerical and graphical terms, thus providing an analysis of the spatial variation between the model output and collocated observations. This paper demonstrates how this new exploratory spatial data analysis and geostatistics are able to assess the impact of data initialisation schemes of atmosphere forecast models using realistic surface boundary conditions (i.e. remotely-sensed SST).

Improvement in forecasting capability would much benefit other applications, such as warning systems in case of oil spillages and extreme climatic events in the Mediterranean region.

## 2. Geographical area

The geographical domain of the study covers the Mediterranean region, and to a high detail over part of the Ionian basin ( $15.78^{\circ}$  E,  $33.18^{\circ}$  N -  $19.18^{\circ}$  E,  $35.74^{\circ}$  N) in the Eastern Mediterranean Sea. This region constitutes a suitable test area for studies related to air-sea interaction where climatic characteristics drive the dynamics and dense water formations. The strong forcing by the local north-westerly winds and by the inflow of cool and less saline Atlantic waters lead to significant air-sea interactions and sub-basin scale oceanic features that eventually affect the basin-wide circulation (Robinson et al., 2001). Since this study uses remotely sensed data retrieved by the TMI sensor, the highest latitude of the geographical area does not exceed  $40^{\circ}$  N because of the low inclination orbit of the satellite.

## 3. Datasets

The time period for this study was determined by the availability of an annual set of meteorological data used to initialise and verify the enhanced skill of the atmosphere model.

### 3.1. Remotely sensed SST and 10 m wind magnitude

Each daily coverage of the sea surface retrieved by the Tropical Microwave Imager (TMI) onboard the TRMM Satellite is organised into seven ascending and descending datasets as follows: time of data retrieval, sea surface temperature, 10m surface wind speed detected at 11 GHz, 10 m surface wind speed detected at 37 GHz, integrated precipitable water vapour, cloud liquid water, and precipitation rate. Daily TMI data covering a full year period (January-December 1999) was retrieved by ftp from <http://www.ssmi.com>. The data is provided as binary data and

cover a global region extending from 40° S to 40° N, and each daily data file consisted of fourteen 0.25° x 0.25° grid (1440, 320) byte maps. Two FORTRAN programmes were written to decode, process and format the TMI data into Grid Analysis and Display System (GrADS)-format monthly set of sequential, daily ascending and descending data. Since precipitation induces error in the retrieval of the geophysical parameters collected by the TMI sensor (Gentemann et al., 2004), rain raster grid points were used to mask off same-area raster grid points from the other collocated ascending and descending geophysical field maps. Valid geophysical TMI data lay between 0 and 250 and pixels were scaled according to the calibration information supplied with the data as to obtain meaningful geophysical fields. TMI monthly datasets were displayed to note orbit time and the integrity of the SST and wind fields over the model domain area. This enabled the selection of the appropriate sets of SST data to initialise the Eta model at 00:00 UTC and the identification of co-temporal remotely-sensed geophysical fields to similar field elements predicted by the model.

### 3.2. Quality control of the remotely sensed datasets

Special attention was given to the data used to initialise the model in order to ensure the insertion of good quality initial SST fields in the numerical model. The QC method considered the following elements: (a) time of retrieval  $\pm$  3 hours from 00:00 UTC initialisation time; (b) no data gaps due to presence of precipitation in the study area; (c) wind speeds over model domain higher than 2  $\text{ms}^{-1}$  but not exceeding 12  $\text{ms}^{-1}$  (reflecting the main wind regime over the regional area); (d) spatial auto-correlation (Moran's Index I) of SST raster fields exceeding 0.76 to eliminate erroneous pixels and (e) full-data integrity over the area of interest.

	valid	dates								
Jan	18	22	24							
Feb	-									
Mar	7	8	10	11	14					
Apr	21	22	23	26	27	28				
May	3									
June	-									
Jul	22	23	24	25	26	27	28	29	30	31
Aug	1	2	3	4	5	6				
Sep	16	18	20	21						
Oct	22	23	24	26	28	29				
Nov	-									
Dec	9	13	14	15	16	17	19	20		

**Table 1. Use of valid TMI-derived SST data acquired during 1999 and used to initialise the atmosphere model.**

### 3.3. Eta Model initial and boundary conditions

Topography data and sea-land masks obtained from USGS were re-interpolated over the regional atmosphere model domain in both vertical and horizontal dimension. Soil texture information (source: NOAA) and the full 10" by 10" vegetation cover (source: USGS/EROS) were also re-interpolated to the regional model's grid points. This initial condition was only relevant for the Mediterranean-wide regional model. Daily World Area Forecast System (WAFS; source: NCEP, Washington) predictions pertaining to octant grid 37 (I) were downloaded from the United States National Weather Service gateway ftp server and used to initialise the lateral boundary conditions of the regional atmosphere model. The data contains 1.25° by 1.25° latitude-longitude gridded forecasts of many meteorological parameters, including air temperature, u- and v- wind components, relative humidity and pressure at standard atmospheric levels. Forecasts of each meteorological parameters are available up to 72 hours in advance the time they are issued. Daily SST data produced by the Global Data Assimilation System (GDAS; source: NCEP, Washington) were also archived. This dataset is routinely used by operational numerical weather prediction agencies in the US to initialise their local area atmosphere models to produce daily atmospheric forecasts. The necessary lateral and surface boundary conditions for all days during 1999 were downloaded and archived as GRIB products with 6-hr time resolution. Table 1 shows valid initialisation dates depending on the timely availability of lateral and surface boundary conditions and remotely sensed observations used for model verification.

## 4. Methodology

The Eta atmosphere model is one of NOAA's weather prediction model with a variety of unique features in its numerical formulation and a comprehensive physical package. Some of these features are its step-mountain ('eta') vertical coordinate (Mesinger 1984), its Arakawa-type horizontal momentum advection scheme (Janjic 1984); and its algebraic conservation of energy in transformations between the kinetic and potential energy in both space and time differencing (Janjic et al. 1995). Within the model's physical package, some of the special features are its modified Betts-Miller convection scheme (Janjic 1994), its Mellor-Yamada level 2.5



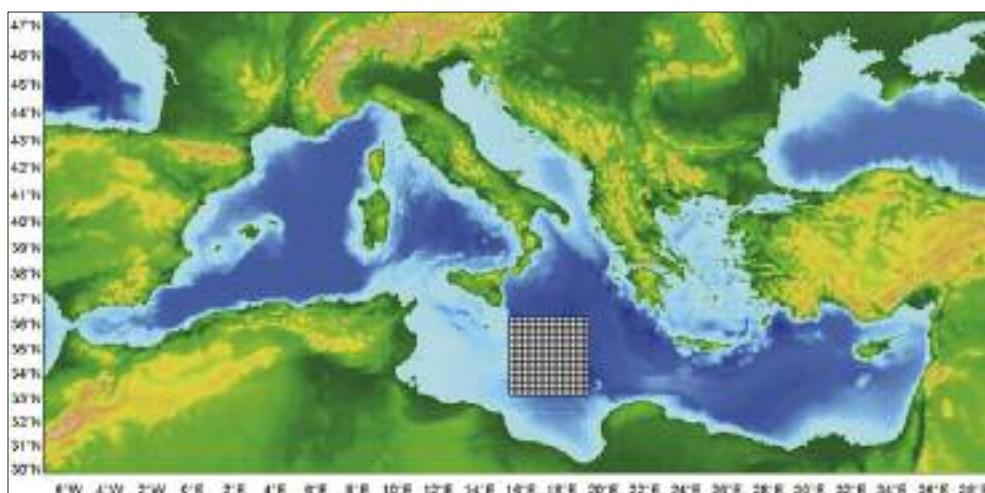
turbulence closure with improved treatment of the realisability problem (Janjic 1996a), its viscous sublayer scheme over both water and land surfaces (Janjic 1996b), and its prognostic cloud water/ice scheme (Zhao & Carr 1997). The Eta model had source codes and related executable files that: (1) decoded WAFS and GDAS binary data, land cover and topography data and re-interpolated to the Eta coordinate system; (2) constituted the numerical schemes of the model; (3) calculated the output of all relevant forecasted fields and their conversion into standard coordinates for displaying. Scripts unified the operations of all these three main groups of codes, according to a secondary file that specified the location of data files.

#### 4.1. Atmosphere model setup

A Mediterranean-wide atmosphere model with a horizontal resolution of  $0.17^\circ \times 0.17^\circ$  was driven by a 1-year long dataset consisting of daily lateral and surface boundary conditions forecasted by a global model at  $1.25^\circ$  by  $1.25^\circ$  horizontal resolution. This regional model was used to drive another Eta model nested within it at a horizontal resolution of  $0.04^\circ$  by  $0.04^\circ$  covering part of the Ionian basin (fig. 1).

#### 4.2. Mediterranean-wide model runs

The modelling process was divided into three stages: pre-processing, processing and post-processing. Model pre-processing started with the definition of the model boundaries, horizontal ( $0.17^\circ$ ) and vertical (32 layers) resolutions, with a model time step of 75s, followed by the preparation of constant input data such as topography and land cover data. The lateral (WAFS) and surface (SST) boundary conditions were decoded and used by the regional Eta model to initialise its boundary conditions. The processing calculated the future state of the atmosphere using the numerical schemes embedded in the model for a number of pre-defined time steps that eventually lead to a set of short-range at-



**Figure 1. Domains of the limited area, Mediterranean-wide Eta model (LAT:  $29.00^\circ$  N –  $47.50^\circ$  N; LON:  $-10.00^\circ$  W –  $42.00^\circ$  E). Nested within its integration domain is the high-resolution Eta model (LAT:  $33.24^\circ$  N –  $35.74^\circ$  N; LON:  $15.74^\circ$  E –  $19.17^\circ$  E).**

mospheric forecasts. The full-year, daily integration of the regional Eta model generated mesoscale, 3-hourly forecasts for 36 hours; the prognostic variables were in turn used as lateral boundary conditions to initialise the nested Eta model.

#### 4.3. Nested, high resolution model runs

Model pre-processing again started with the definition of the model boundaries, horizontal ( $0.04^\circ$ ) and vertical (24 layers) resolutions, with a model time step of 24s. The lateral (produced by the regional model) and surface (SST) boundary conditions were decoded and used by the regional Eta model to initialise its boundary conditions. The experimental work consisted of a parallel series of high resolution, hind-cast model simulations. The reference nested Eta model used GDAS-SST as its initial lower surface boundary data at 00 UTC. An identical experimental nested Eta model was set up with exactly the same numerics as the reference model but using, instead of GDAS-SST data, the SST observations collected by the TMI sensor between the period January to December 1999. This high resolution, hind-cast model output was referred to as the "Experimental forecast". Both nested models used the same lateral boundary conditions as predicted by the regional atmosphere model for each consecutive 36-hourly run starting at 00:00 UTC. The switching between the two sources of SST data was done by making available and accessing the appropriate SST dataset. Computations were made to the predicted fields of both models and diagnostics and meteorological parameters were derived from the forecast variables. Model variables were interpolated vertically to isobaric and constant al-



titude surfaces and interpolated horizontally to output grids. This generated a set of high resolution, 36-hour, 3-hourly air-sea surface flux forecasts which included the latent and sensible heat fluxes, outgoing and incoming long- and short-wave radiation, u- and v-component of momentum flux, moisture heat flux, surface air temperature and surface pressure. In total, 98 forecast runs were processed, each taking around 6 hours to complete. The 36-hr forecasted air-sea surface fluxes from both the reference and experimental models were extracted for statistical analysis.

#### 4.4. Model performance analysis

Descriptive and inferential model verification measures were applied to verify model improvement based on the relative accuracy between the forecasts and collocated, remotely sensed observations. Descriptive statistics, related to the understanding of the variations between forecast and observations, included the calculation of the mean, standard deviation, bias error and mean square error. Inferential statistics consisted of methods and procedures used to understand the underlying processes and quality of the generated results. Traditional, non-spatial methods were used to quantify the skill score of each model system to forecast the geophysical field. The Skill Score (SS) was used to reflect the average accuracy of the experimental forecasts relative to those produced by the reference setup (Murphy, 1997). SS was based on the mean square error, defined as:

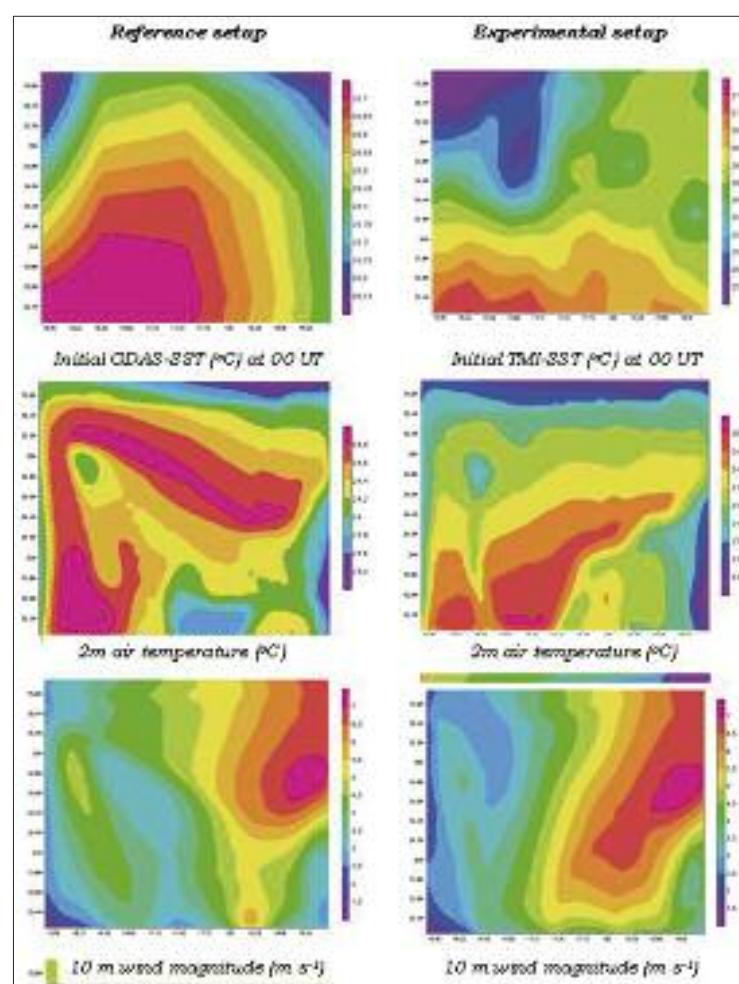
$$SS = 1 - (MSE_{exp}/MSE_{ref})$$

where  $MSE_{exp}$  is the mean square error of the experimental forecasts and  $MSE_{ref}$  relates to that of the reference forecasts.

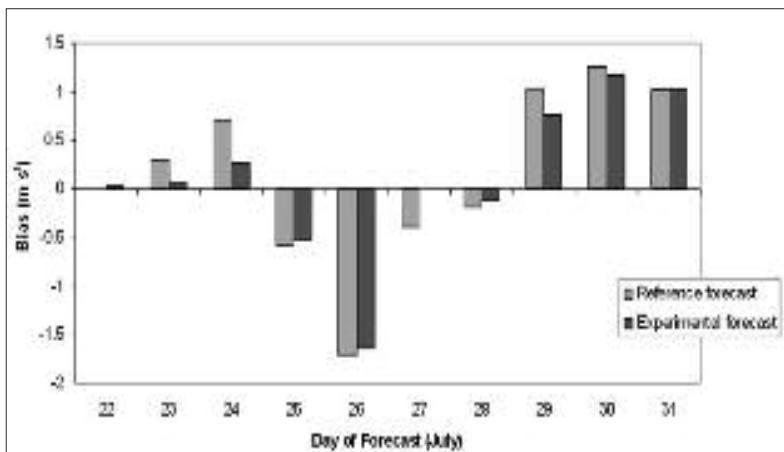
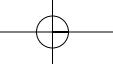
Exploratory Spatial Data Analysis (ESDA) and geo-statistics were used to detect and assess improved forecasting of spatial patterns. This was performed using two independent approaches to analyse the performance of the competing models in spatial terms. These were (1) Spatial similarity analysis and feature matching analysis and (2) Residual semi-variogram

analysis.

Spatial similarity was performed to analyse the spatial relationship between the two sets of forecasts and observations. This study consisted of the combination of information fields to determine the position of similar clusters between pairs of forecasted and their collocated fields. The raster information between observed and predicted fields was recoded into a set of normalised class values that assisted the calculation of spatial similarity. By using proximity analysis, the observation field was assigned as a reference while the forecasted field as the source to measure degree of similarity. These two recoded fields were then overlayed and the difference between the coinciding classes was related to the similarity between them. An



**Figure 2. Some of the geophysical fields predicted by the Reference (left) and Experimental (right) setup, initialised by the GDAS-SST and TMI-SST respectively. The predicted geophysical fields correspond to the 27th hour fields starting on Jul 27, 1999 00:00 UT. (y axis: latitude; x axis: longitude).**



**Figure 3. Bias between the two sets of forecasted fields and collocated remotely sensed 10 m wind magnitude ( $m s^{-1}$ ) observations during the month of July. The dates represent the starting point of the model run at 00 UT and verification was performed at the 21<sup>st</sup>, 24<sup>th</sup> and 27<sup>th</sup> forecast hour depending on the closest collocated observations.**

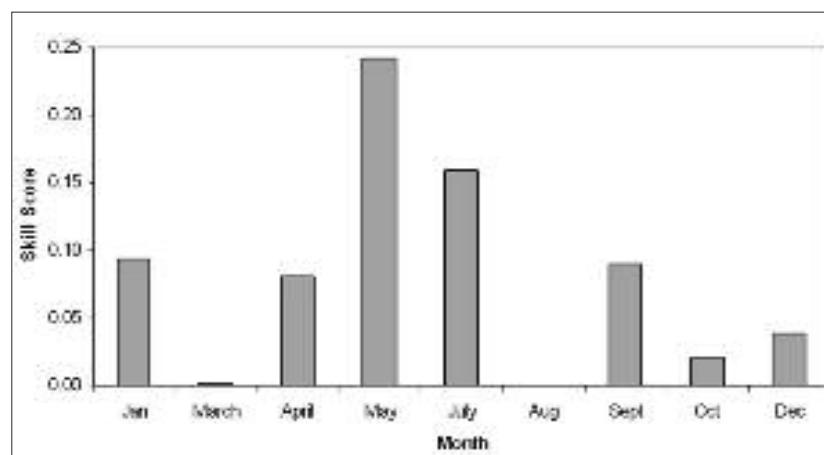
ISOCLASS, unsupervised classification, processed the resulting raster field and generated a classified map consisting of a single band, byte raster dataset, with each data point holding an integer value that corresponded to the class number (or degree of similarity) for that point. The processing algorithm also displayed a look-up-table indicating the degree of spatial similarity through a matching and two-way ranking measure.

Semi-variance analysis was performed on raster maps produced from the residuals between the forecasted 10m wind magnitude fields (originating from the reference and experimental setup) and collocated observations. Residual maps were generated by subtracting the individual forecasted ascii datasets from co-temporal observations. The residual spatial structure between the two sets of data (i.e. forecast and observation) were modelled on the basis of autocorrelation. Pixel values corresponding to the matrix of the geophysical field (in this case the 10m wind magnitude) that had similar spatial attributes were considered more similar than pixel values further apart. This was used to model the structure of the residual data, i.e. its variability as a function of space by means of variography. The semi-variograms displayed the relation between the semi-variance and the spatial separation (lag distances), and is a quantitative descriptive sta-

tistic that can be graphically represented in a manner which characterises the spatial continuity (i.e. roughness) of the residual data sets. GS+ (<http://www.gammadesign.com>) was used for variogram modelling at the full resolution of these residuals. Five types of isotropic models were produced for each residual map, each of which described according to three parameters: Nugget Variance, Sill, and Range. The Range parameter was used to define the best-fit line and the best model fit was selected and using the Reduced Sums of Squares value, the best model fit was selected. These semi-variograms translated the texture information according to the idealised relationships in the form N (nugget variance; which is the non-spatial variation due to measurement

error and variations in the data that relate to shorter ranges than the minimum sampled data spacing), C (sill minus the nugget; which is total semivariance between the two datasets), R (range; which is the increase in the semivariogram values with increasing lags) and h (distance). Analyses proceeded with the calculation of the anisotropic semi-variance surface or variogram map for each residual dataset.

This map provided a visual picture of the semi-variance of the residuals in every compass direction as to find the most appropriate principal axis that defines the anisotropic variogram model.



**Figure 4. Skill score between the two competing models showing an overall positive skill score of the experimental over the reference setup, based on the mean square error.**

## 5. Results and discussion

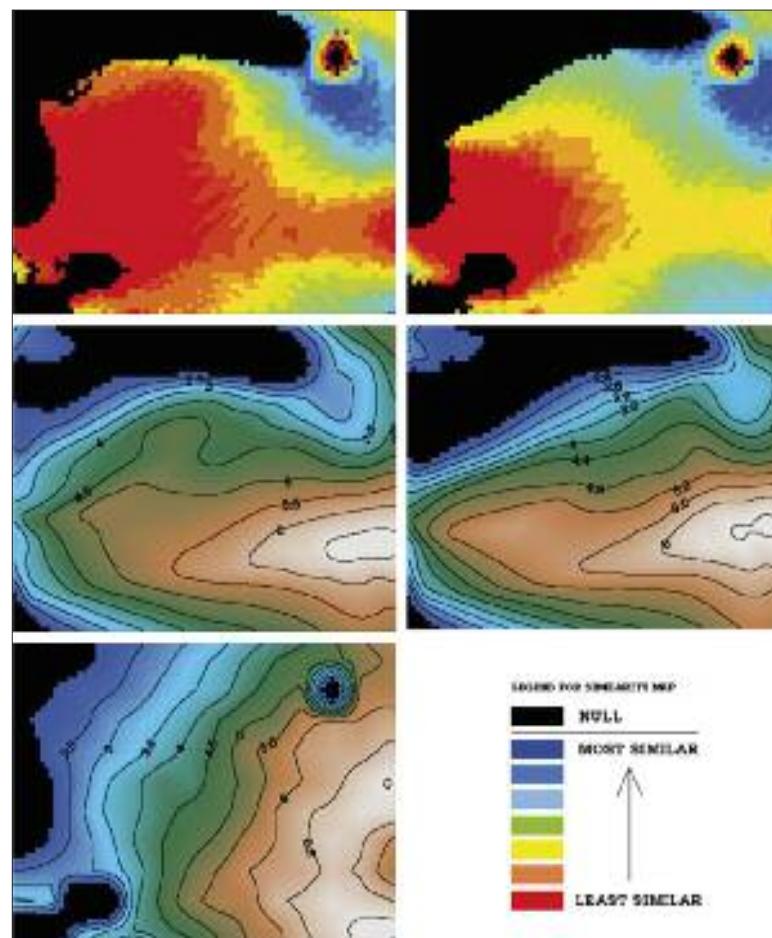
### 5.1. The numerical experiments

The statistical measures were derived from "single-forecast" analysis which is ideal to address (1) the accuracy of model initialisation of both forecasting systems as well as (2) the resulting forecast performance. Some examples of the forecasted geophysical fields derived by the reference and experimental setup are shown in figure 2. The most significant observations derived from these two sets of output is that the different initial SST lower boundary conditions always led to spatially different forecasts. This shows the influence of SST on the prediction of these fields, in particular to the surface geophysical fields. The spatial pattern of the two sets of the surface atmospheric fields are different, especially for the 2m air temperature and total heat flux. A look at the basic statistical measures of the two forecasting systems in relation to observations shows that both systems produce very similar 10m wind magnitude fields. This small difference between model-generated fields and observations is indicative of a high accuracy of both systems to predict the surface wind speed at forecasting steps ranging between the 21<sup>st</sup> and 27<sup>th</sup> hour after model initialisation. However, closer inspection at the biases of the two sets of forecasts against observations shows that the experimental setup is more precise than the reference one throughout the year without any seasonal trend. An example of the overall better performance of the experimental model setup is shown during the month of July as shown in figure 3. It is also important to note that the 10 m wind magnitude forecasts were verified only when the wind speed observations valid at the time of the forecast was greater than or equal to 2 ms<sup>-1</sup>. This prevented wind magnitude forecasts associated with light and variable winds from degrading verification scores. The same approach is taken by major numerical weather prediction centers such as NOAA for its model-based output statistics wind

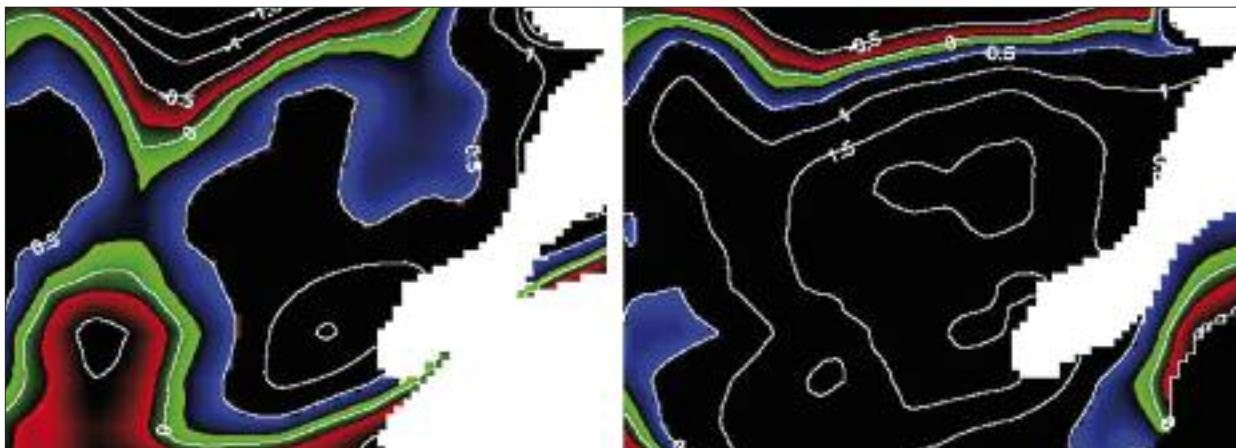
guidance over the US territory using the Nested Grid Model (NGM).

### 5.2. Performance measures

The mean square error (MSE) provided a clear picture of the performance between the two model setup, showing an overall improvement when TMI-derived SST is used to initialise the lower boundary conditions at the start of the model run. The Skill Score shows a noticeable improvement reached by the experimental over the reference setup by an average of 10% in the forecast accuracy of the 10m wind magnitude (fig. 4). The



**Figure 5 a-e. Case study: January 22, 1999:** (a) similarity map between predicted surface wind speed by the reference system and observations; (b) similarity map between predicted surface wind speed by the experimental system and observations; (c) predicted wind speed (in ms<sup>-1</sup>) by the reference system; (d) predicted wind speed (in ms<sup>-1</sup>) by the experimental system; (e) observed wind speed (in ms<sup>-1</sup>) by the tropical microwave imager on January 23, 1999 at 02:08 UT. Black pixels represent nulled pixels that are below 2.0 ms<sup>-1</sup>. Small circular contours with central black pixels coincide with precipitation and are nulled.



**Figure 6 a-b.** The residual map of the differences between the 10 m wind magnitude ( $ms^{-1}$ ) forecasted by the (a) reference setup and (b) experimental setup, and collocated observations derived from the Tropical Microwave Imager on July 25 at 24:00 UT (or July 26th at 00 UT).

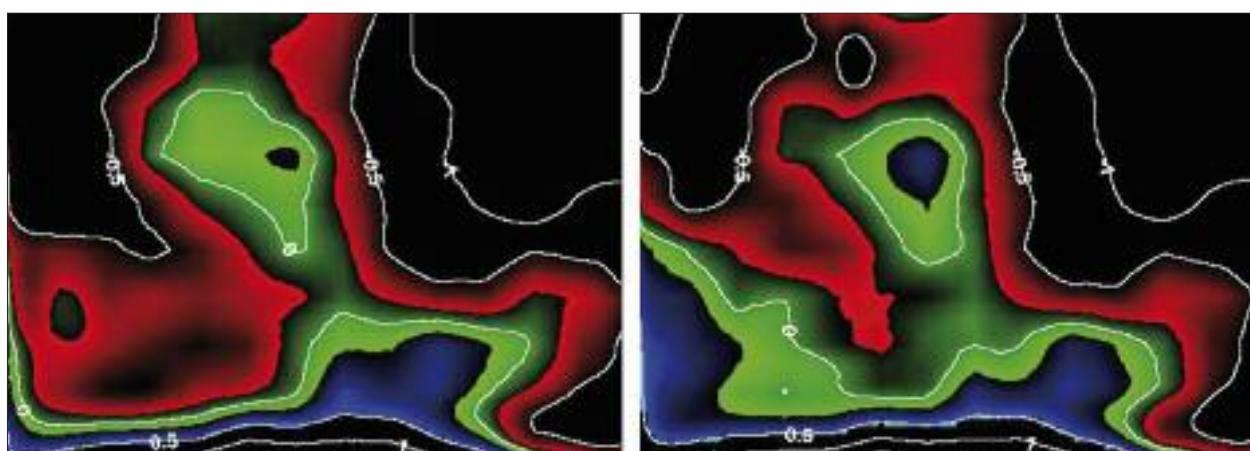
use of averaged monthly skill scores for the entire analysis provide only a partial picture of this improvement. Results show that highest scores are observed during the spring and summer seasons, although this period shows the greatest variability in terms of standard errors. Lesser variability accompanied by lower skill scores occurs during the colder months, when the wind magnitude is more intense.

The choice of single dates for verification throughout the entire analysis corresponded to the availability of observations to initialise and validate the experimental model setup. Lack of matching observations to verify the forecasts during February, June and November led to the omission of these three months in the analysis. As for the months analysed, it needs to be taken into account that the single-date verification does not reflect the aver-

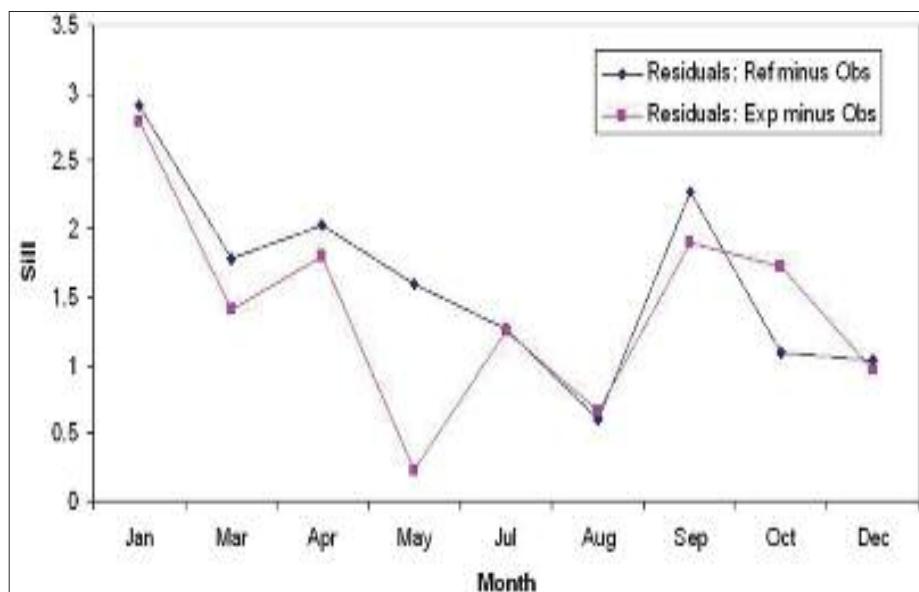
age behaviour of the model for the entire month and therefore this verification tends to become more sensitive to model errors with respect to the general monthly weather systems.

### 5.3. ESDA

The Spatial Similarity index was used by this study to assess the degree of shared attributes between the two model systems and the remotely-sensed observations. In order to best illustrate the spatial analysis, two case studies are presented that best describe the results over the area of interest. Case study January 22, 1999: This case study example presents a wind speed regime that is typical for January. Figure 5e shows relatively moderate winds reaching a maximum of  $6.5 ms^{-1}$  within the area of interest. A more or



**Figure 7 a-b** showing the residual map of the differences between the 10 m wind magnitude ( $ms^{-1}$ ) forecasted by the (a) reference setup and (b) experimental setup, and collocated observations derived from the Tropical Microwave Imager on December 15 at 24:00 UT.



**Figure 8. Model fit for isotropic semi-variograms of the residuals from January till December. These were best described by the Gaussian model.**

less constant gradient is observed starting with moderate wind gradients of about  $7 \text{ ms}^{-1}$  to very light wind conditions towards the East, reaching a minimum of  $2 \text{ ms}^{-1}$ . A close inspection of the similarity maps shows that the experimental (fig. 5b) setup gives an overall higher similarity index than the reference setup (fig. 5a). The individual forecasted wind fields show that the wind gradients generated by the experimental setup (fig. 5d) are closer in pattern to the observed wind fields than the reference ones. This is especially true for the wind magnitude range of  $4.5$  to  $5.0 \text{ ms}^{-1}$ , which bulges out towards the north in the reference forecast as opposed to a calmer condition given by both the experimental forecast and observation. Minor dissimilarity between the reference setup and observations is also shown by the low intensity wind fields of around  $2.4 \text{ ms}^{-1}$ . The patterns of such fields produced by the reference setup do not agree with those observed, unlike the fields produced by the experimental one. Spatial similarity of individual classes of wind magnitude, rather than the individual pixel values, proved to be a more convenient and realistic way of assessing the tendency of the spatio-temporal attributes of the predicted fields. These results bring out the usefulness of relational spatial similarity rather than the total scalar comparison between the two datasets for ocean forecasting systems. The system allows results to be displayed that indicate the degree of similarity through a matching and ranking measure. This facility allows the user to search for a set of textural and spatial parameters to derive the similarity between the background information and the

new parameters entered into the analysis. This study shows how similarity assessment can be a useful concept for retrieving and analysing spatial information as it may help numerical modelers describe and explore their forecasts, their immediate environment and relationships to observations.

In this study, geostatistics complemented the application of the other statistical measures so far presented. The reason behind using geostatistical analysis was to model the spatial structure of the residual fields of the predictions and observations and translate the degree of spatial correlation between these two datasets in numerical and graphical terms. In

doing so, residual variography provided an analysis of the spatial variation on the differences between the model output and collocated observations. The sensitivity of this tool focuses on the spatial variation between the competing models with collocated observations. Figures 6 and 7 are examples of residual maps made up of the differences between the forecast  $10 \text{ m}$  wind magnitude ( $\text{ms}^{-1}$ ) (generated by the reference and experimental models respectively) and collocated observations derived from the TMI on July 25 and December 15 1999. Only those residual pixel groups ranging from  $-0.5$  to  $0.5$  are shown; the remaining gradients are shown by means of contours for ease of clarity. A greater range of pixels showing minimal difference occurs in figures (b). Areas shown in white correspond to nulled filtered pixels due to either low variable, or high wind speeds. The best overall fit for the isotropic semi-variogram plots derived for the entire residual dataset was the Gaussian model. The following general observations were made:

1. These model variograms suggested that in the most significant portion of the semi-variogram model, a single, long-range process dominates. This preference towards the gaussian behaviour was probably due to the way the particular geophysical field behaves, giving a general smooth description of wind field gradients.
2. The semi-variogram models exhibit a strong spatial dependence with very limited random variation. In most cases, the value for the nugget was always low except on few occasions when 'extreme events' occurred, such as extensive precipitation or during



strong wind events (e.g. January 24, March 17, and December 13). The presence of random variability included in the numerical forecasts is therefore extremely low and that most of the variation observed can be attributed to the nature of the geophysical data.

3. The lower value for both the sill and range values of the residuals resulting from the experimental forecasts and collocated observations point towards a closer similarity that the ones produced by the reference system (figures 8).

Geostatistical analysis provided this study with additional textual information on the behaviour of the two 'competing' models. The results for the anisotropic semi-variogram analysis, for example, very often indicated the existence of directional trends. The cause of anisotropy may be due to the prevailing variability in the 10m wind direction over the area of interest, and if so, this could mean that the residuals show a certain degree of dependence on the climatology of the area. Residual analysis however, proved to be quite insensitive to catch the subtleties that were identified using the spatial similarity method. A case in point is the single forecast analysis for December 15<sup>th</sup>. Visual inspection of the two spatial variograms indicate no significant differences and both reveal least semi-variance in the 45° direction. A similar relationship was observed for January 22, January 24, and April 22. This approach has never been applied to verify improvements made on numerical atmosphere models. On the other hand, geostatistics is a fairly common approach to study and derive the distribution, spatial patterns and texture analysis of natural phenomena ranging from insect population (Liebhold et al., 1996), ozone (Liu and Rossini, 1997), forests (Treitz, 2001), mineral resources (e.g. Reis et al., 2003), remotely sensed images (Atkinson and Lewis, 2000), down to microbial patterns (Franklin and Mills, 2003). It is interesting to note that analysis of residual variograms has been done by Holdaway (1996) for the modelling and interpolation of monthly temperature and Aranuvachapun and Maskell (1997) to study oceanic temperature frontal fields.

### 5.5. Spatial variography

A representation in 2D and 3D space of the behaviour of the variogram was made by drawing a map of the iso-variogram lines as a function of the vector  $h$ .

This was also an excellent way to check for anisotropy by means of a contour plot of semi-variogram values by direction. The majority of the case studies showed that the iso-variogram lines were approximated by minor ellipses having a mild degree of anisotropy in the 45° direction. This indicates a small degree of directional drift in the residual wind magnitude, possibly

brought about by systematic weaknesses of both forecasting systems to predict the 10 m wind magnitude over the ocean. The cause of this may be due to the model's predictive skill to spatially forecast better categories of wind scales as was highlighted by the spatial similarity analysis when the viscous sublayer model scheme was fine-tuned for events when strong surface winds are prevalent. Similar characterisation is required. The spatial similarity analysis for January 22 and December 15 for example, revealed a better prediction of stronger wind field gradients by the experimental model. On the other hand, dissimilarity is observed when wind fields tend to get weaker as supported by the collocated surface wind field maps.

### 6. Summary and conclusions

This study showed that the provision of microwave-derived SST to initialise a high resolution atmosphere model can result in an overall improved skill to forecast one of the most important air-sea surface fluxes in both quantitative and spatial terms. The extent of the improvement was consistent throughout the study period and it is evident that the mesoscale spatial information offered by the novel remotely sensed dataset is the main cause for such an improvement. The diagnostic verification routines developed by this study were applied to evaluate the spatial context of the improved model predictions. Both the spatial match score (that measures the precise spatial hit of wind speed categories) and the spatial similarity index, highlighted the superiority of the microwave data to initialise the surface boundary condition of the Eta model. Semi-variance analysis was used to model the spatial structure of the residual fields between predictions and observations, and to translate the degree of spatial correlation in numerical and graphical form. This approach supported the results showed by the standard and diagnostic verification, providing further evidence of the improved correspondence between the experimental forecasts and collocated observations. Geostatistics described a prevalent gaussian model for the residuals for both the reference and experimental setup that were characterised by low noise, and mild degree of anisotropy. This study has successfully demonstrated the application of exploratory spatial statistical methods to verify the output of a weather forecasting model. The geostatistical analysis of residuals was used as a convenient tool to assess the improved skill of the atmosphere model that was initialised using realistic surface boundary conditions (i.e. remotely-sensed SST).

This results of this method complemented those obtained by both standard statistical routines (average, standard deviation, bias, means square error and skill score) and the exploratory spa-

ial statistical methods to verify the output of a weather forecasting model. The geostatistical analysis of residuals was used as a convenient tool to assess the improved skill of the atmosphere model that was initialised using realistic surface boundary conditions (i.e. remotely-sensed SST). This results of this method complemented those obtained by both standard statistical routines (average, standard deviation, bias, means square error and skill score) and the exploratory spatial data analysis (Match score analysis and spatial similarity), in that the use of microwave-derived SST as the initial surface boundary condition for the Eta model leads to an overall improved skill. With the availability of orbiting microwave sensors, such as the AMSR series of instruments on AQUA and TERRA, the availability of even more accurate retrievals of sea surface temperature data on a global scale is now a reality. This availability is leading to the generation of fused SST products as part of the GODAE High-Resolution SST Pilot Project (GHRSST-PP), consisting of twice-daily, global SST fields originally derived from the AMSR-E orbiting sensors (<http://ghrsst-pp.metoffice.com/>).

### Acknowledgements

The author is grateful to Remote Sensing Systems ([www.remss.com](http://www.remss.com)), sponsored by the NASA Earth Science REASoN DISCOVER Project, for making available important data used in this work.

to contact  
cgaldies@yahoo.co.uk

\* Malta Meteorological Office, MIA plc, Malta

\* \* Dept. of Geography, University of Durham, UK

