

Evaluation of the performance of DIAS ionospheric forecasting models

Ioanna Tsagouri*

National Observatory of Athens, Institute for Space Applications and Remote Sensing, Metaxa and Vas. Pavlou, 15236 P. Penteli, Greece

*corresponding author: e-mail: tsagouri@noa.gr

Received 14 January 2011 / Accepted 30 June 2011

ABSTRACT

Nowcasting and forecasting ionospheric products and services for the European region are regularly provided since August 2006 through the European Digital upper Atmosphere Server (DIAS, <http://dias.space.noa.gr>). Currently, DIAS ionospheric forecasts are based on the online implementation of two models: (i) the Solar Wind driven autoregression model for Ionospheric short-term Forecast (SWIF), which combines historical and real-time ionospheric observations with solar-wind parameters obtained in real time at the L1 point from NASA ACE spacecraft, and (ii) the Geomagnetically Correlated Autoregression Model (GCAM), which is a time series forecasting method driven by a synthetic geomagnetic index. In this paper we investigate the operational ability and the accuracy of both DIAS models carrying out a metrics-based evaluation of their performance under all possible conditions. The analysis was established on the systematic comparison between models' predictions with actual observations obtained over almost one solar cycle (1998–2007) at four European ionospheric locations (Athens, Chilton, Juliusruh and Rome) and on the comparison of the models' performance against two simple prediction strategies, the median- and the persistence-based predictions during storm conditions. The results verify operational validity for both models and quantify their prediction accuracy under all possible conditions in support of operational applications but also of comparative studies in assessing or expanding the current ionospheric forecasting capabilities.

Key words. 2447: modeling and forecasting – 2441: ionospheric storms – 2435: ionospheric disturbances – 2431: ionosphere/magnetosphere interactions – 2443: midlatitude ionosphere

1. Introduction

Within space weather framework, ionospheric predictions, especially in the short and the medium term, received much of the attention as the state of the ionosphere influences significantly the performance of widely used systems, e.g. HF communication, Earth Observation and satellite positioning and navigation systems. The emphasis was given on the prediction of the F region peak electron density (NmF2) or the critical foF2 frequency and on the total electron content as these parameters specify the conditions in the topside ionosphere and plasmasphere, where a large number of space assets reside.

An overview of ionospheric models available for space weather purposes has been recently published (Belehaki et al. 2009a). Focusing in the prediction of the foF2, significant contribution for operational applications comes from data-driven empirical and semi-empirical models that exploit either time series forecasting techniques such as the standard autocorrelation as well as auto-covariance and neural networks (e.g. Koutroumbas et al. 2008; Koutroumbas & Belehaki 2005; Tulunay et al. 2004a, 2004b; Cander 2003; Stanislawska & Zbyszynski 2001, 2002; McKinnell & Poole 2001; Wintoft & Cander 2000a, 2000b) or space weather indices and proxies as drivers of the ionospheric response during disturbed conditions (e.g. Mikhailov et al. 2007; Muhtarov et al. 2002; Araujo-Pradere et al. 2002; Kutiev & Muhtarov 2001, 2003; Muhtarov & Kutiev 1999; Pietrella & Perrone 2008; Tsagouri & Belehaki 2008; Tsagouri et al. 2009). In addition, the existing

modeling background can take significant advantage from the use of real-time networks of ionospheric ground-based sounding systems (Galkin et al. 2006; Belehaki et al. 2007) that are currently available to sufficiently drive the development of operational ionospheric forecasting services.

The European Digital Upper Atmosphere Server (DIAS; Belehaki et al. 2006, 2007) is based on the European real-time network of ionosondes that feed the system with real-time information as well as historical data and it aims to provide reliable information on the current conditions of the ionosphere over middle latitude European region and accurate forecasting information in long-term and short-term time scales. DIAS system operates regularly since August 2006 and delivers a full range of ionospheric products, through the address <http://dias.space.noa.gr>. DIAS forecasting products and services are currently supported by the implementation of two ionospheric forecasting models that were designed following current trends in the field: (i) the Geomagnetically Correlated Autoregression Model (GCAM; Muhtarov et al. 2002), which is a time series forecasting method driven by a synthetic geomagnetic index and (ii) the Solar Wind driven autoregression model for Ionospheric short-term Forecast (SWIF; Tsagouri et al. 2009), which combines an autoregression forecasting algorithm, called Time Series Autoregressive (TSAR; Koutroumbas et al. 2008), with the empirical Storm-Time Ionospheric Model (STIM; Tsagouri & Belehaki 2006, 2008) that formulates the ionospheric storm-time response based on IMF disturbances monitored at L1 point by NASA ACE spacecraft. Based on the implementation of the

two models, DIAS system releases ionospheric forecasts up to 24 h ahead over single DIAS locations in a plot format and regional forecasts over Europe in a map format.

In principle, the successful transition from research to operational models requires among others the systematic assessment of the models' performance under all possible conditions. Indeed, for the evaluation of operational models quantitative tests should be carefully designed to drive the efforts beyond the conventional scientific assessment of the models. Ideally, for the independent and unbiased evaluation of the models' realistic and standardized test suites should be implemented (Hesse et al. 2001). The determination and the maintenance of metrics is fundamental component of this approach as metrics-based testing is necessary to establish the absolute models' performance as well as the models' relative performance in respect to other models with comparable output. The latter makes metrics-based testing valuable in tracking specification capabilities over time. Because of its significance, the metrics-based assessment of the performance of existing space weather models available for operational applications is centrally addressed within COST ES0803 Action "Developing Space Weather Products and Services in Europe" (<http://www.costes0803.noa.gr>) to drive the improvement of the existing models and the introduction of new ones (Belehaki et al. 2009b).

To meet the requirement for the assessment of existing ionospheric forecasting models for operational use but also to support the prediction efficiency of DIAS ionospheric forecasting products and services, this paper is focused on a metrics-based evaluation of the performance of the GCAM and SWIF models aiming: (i) to provide evidence of the operational validity and the forecasting efficiency of DIAS ionospheric forecasting models and (ii) to quantify their prediction accuracy under all possible conditions. The results presented here may be efficiently exploited for current operational applications, but they could also be used to facilitate comparisons either with similar results obtained previously contributing to the overall assessment of our current ionospheric forecasting capabilities or with future investigations to justify improvement in ionospheric forecasting capabilities over time.

2. Models' description and DIAS implementation plan

2.1. The Geomagnetically Correlated Autoregression Model (GCAM)

GCAM (Muhtarov et al. 2002) incorporates the cross-correlation between the foF2 and the Ap index into the auto-correlation analysis. In GCAM approach, the hourly time series of foF2 is considered as composed of a periodic component and a random component. The periodic component contains the average diurnal variation, which in ionospheric studies is traditionally represented by monthly medians. GCAM uses, instead of the foF2 itself, the relative deviations from the medians so that the periodic component is removed and the model predicts the correction to the median values for each hour of the prediction period as a function of the geomagnetic activity level that is expressed by a geomagnetic activity index. It is an extrapolation model based on weighted past data and the model's predictions are driven by (i) current and recent past foF2 measurements and records of the geomagnetic activity index, (ii) estimates of the reference ionosphere and (iii) prediction of the geomagnetic activity index for each hour of the prediction period. A synthetic

geomagnetic activity index G of hourly resolution based on the Ap index was also introduced by Muhtarov et al. (2002) to ensure forecasting ability of the model. In particular, the index G is extracted from the daily Ap index for which a 45-day prediction is issued by the Space Weather Prediction Center of NOAA, Boulder (<http://www.swpc.noaa.gov>). Parametric expressions of the auto- and cross-correlation functions ensure statistical sufficiency of the model, while the parameters are obtained by data fitting at each ionospheric location.

For the implementation of GCAM in DIAS system, model parameters were calculated for each DIAS location (Athens, Rome, Ebre, El Arenosilo, Chilton, Pruhonice and Juliusruh) by using ionospheric observations from at least one and a half solar cycle. The reference ionosphere is determined by the running 25-day median while the recent history of the ionospheric activity is considered by autoscaled foF2 values obtained from the last 24 h. GCAM provides foF2 forecasts up to 24 h ahead.

2.2. The Solar Wind driven autoregression model for Ionospheric short-term Forecast (SWIF)

SWIF (Tsagouri et al. 2009) combines real-time and past ionospheric observations with solar-wind parameters obtained in real time at the L1 point through the cooperation of an autoregression forecasting algorithm, the TSAR with the empirical STIM that formulates the ionospheric storm-time response based on solar wind input.

In particular, TSAR is an autoregressive based (AR) model (Koutroumbas et al. 2008) where the prediction of the foF2 value in a particular time is given as a linear combination of the most recent foF2 values. The number of the recent values taken into account depends on the order of the AR model. TSAR modifies the order of the AR model and estimates model's coefficients at the beginning of each calendar month using the foF2 measurements from the previous month. Typically, the AR needs the foF2 values up to 24 h before. TSAR provides ionospheric forecasts up to 24 h ahead during all possible conditions.

STIM, on the other hand, is an empirical model designed to provide a correction factor to the quiet time ionospheric variation during storm conditions in the middle latitude ionosphere (Tsagouri & Belehaki 2006, 2008). Storm predictions are triggered by an alert signal for upcoming ionospheric disturbances obtained from the analysis of the real-time IMF's observations (in particular, observations of the total magnitude B and the IMF-Bz component) obtained from ACE spacecraft. STIM estimates the time delay in the ionospheric storm onset for each geographic location as a function of the local time (LT) of the station at the storm onset. The ionospheric storm time response is obtained by empirical expressions applied to the reference ionospheric variation. The empirical expressions have been derived by superimposed epoch analysis applied to ionospheric observations obtained from eight European ionosondes during 30 storm-time intervals occurred between 1998 and 2005. They follow a polynomial function of sixth order and the coefficients are dependent on the LT of the station under question and its latitude (Tsagouri & Belehaki 2008), as STIM identifies two latitudinal zones: the middle-to-low latitudinal zone for latitudes less than 45° and the middle-to-high latitudinal zone for latitudes greater than 45°.

The SWIF algorithm was established through the comparative analysis of STIM and TSAR predictions obtained during storm disturbances (Tsagouri et al. 2009) so that in no alert/

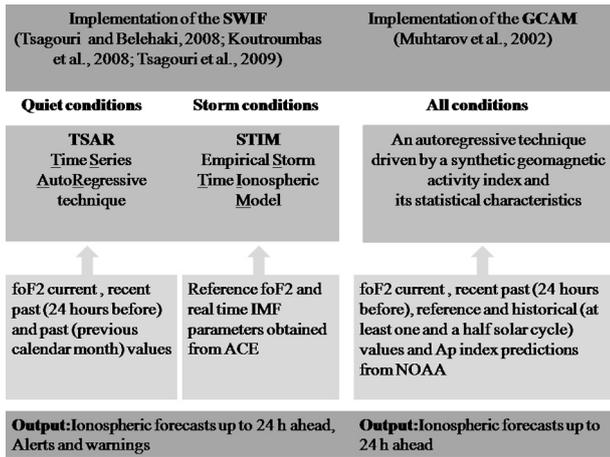


Fig. 1. A schematic diagram describing the implementation plan for SWIF (on the left) and GCAM (on the right) model.

storm conditions, SWIF performs like TSAR, while for storm ionospheric conditions STIM's predictions are progressively adopted for the whole of the disturbance as well as for 24 h after its end. Here it is important to note that STIM's predictions for post-storm conditions are equivalent to the quiet time variation.

For SWIF implementation purposes, quiet time variation is determined by the running 25-day median. SWIF recovers the full set of TSAR's predictions 24 h after the end of the ionospheric storm disturbance, which is determined by STIM algorithm. Recent history of the ionospheric activity is considered again by autoscaled foF2 values obtained the last 24 h. SWIF provides foF2 forecasts up to 24 h ahead as well as alerts and warnings.

A schematic diagram describing the implementation plan for each model is given in [Figure 1](#).

3. Validation and verification plans and results

Validation aims to determine the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model while verification aims to determine if a model implementation accurately represents the developer's conceptual description of the model and its solution ([Thacker et al. 2004](#)). Therefore, model's verification includes both the computerized model validation and the evaluation of the model's operational validity. Both validation and verification may be evaluated through the comparison of the model's prediction with actual measurements and with other models' predictions, but as they are designed to test the model's efficiency in the representation of the actual conditions, the emphasis tends to be given to the systematic comparison with actual measurements ([Doggett 1996](#); [Vassiliadis 2007](#)). This comparison determines the absolute performance of each model and verifies the operational validity of the codes and algorithms. Nevertheless, the investigation of the model's relative performance against simple prediction strategies (e.g. median- and persistence-based estimations) is highly recommended as it serves the assessment of improvement in the prediction capabilities and facilitates comparisons between models with comparable output.

In this investigation, the evaluation tests are established on the systematic comparison of the two DIAS models' predictions

with actual measurements during the time interval 1998–2007 that fairly corresponds to the duration of the previous solar cycle. This made possible the exhaustive investigation of the models' operational ability and the quantification of their prediction accuracy during all possible conditions. The operational implementation of both models in DIAS system aims to support ionospheric forecasts over Europe, so the evaluation tests are designed to cover the European region. The analysis is based on autoscaled foF2 values of hourly resolution obtained from four European stations: Athens (38.1°N, 23.8°E), Rome (41.8°N, 12.5°E), Chilton (51.6°N, 358.7°E) and Juliusruh (54.6°N, 13.4°E). Data from Athens station are obtained for the time interval 2001–2007. Especially for intense storm conditions, the two models' performance was investigated against median (climatological) and persistence values, through the calculation of a skill score.

3.1. Operational validity and general trends

As a first test, [Figures 2a](#) and [2b](#) present the scatter plots of the SWIF (top panel) and GCAM (bottom panel) predictions versus actual measurements for several prediction steps (1, 3, 6, 12, 24 h ahead) for the time interval 1998–2007 and two DIAS locations: Chilton ([Fig. 2a](#)) as a representative example of the middle-to-high latitude stations and Rome ([Fig. 2b](#)) as a representative example of middle-to-low latitude stations. The linear regression line is also plotted in all cases. Visual inspection of [Figures 2a](#) and [2b](#) indicates that: (i) the two models' predictions are satisfactorily correlated with actual observations in all cases, (ii) there is no latitudinal dependence of the models' performance and (iii) the prediction efficiency tends to depend on the prediction step. This tendency is more evident in SWIF case for which the prediction efficiency seems systematically poorer for predictions provided 3 and 6 h ahead. These arguments are also supported by [Figure 3](#) where the correlation coefficient r (left panel) and the coefficient of determination r^2 (right panel) comparing models' predictions with actual measurements over the period 1998–2007 are presented in respect with the prediction step for four DIAS locations: Chilton, Juliusruh, Rome and Athens. Moreover, [Figure 3](#) makes the dependence of the prediction efficiency on the prediction step more evident: both models' performance is very successful for predictions provided 1 h ahead ($r > 0.95$ and $r^2 > 0.9$ in all cases) and tends to be poorer as the prediction step increases. The minimum in the prediction efficiency for SWIF is recorded for predictions provided 6 h ahead (except for Athens where the minimum corresponds to predictions provided 3 h ahead) tending to be substantially recovered for prediction steps greater than 6 h ahead. For GCAM, the prediction efficiency becomes systematically poorer as the prediction step increases, although the variation is rather small. It is important to note that the dependence of the prediction efficiency on the prediction step is more pronounced in SWIF case.

Next, the prediction error distributions are examined for all cases. The prediction error is defined as the difference between observed and modeled values. To facilitate comparisons the results are presented in a boxplot format in [Figure 4](#). This includes a box and whisker plot for each case. The box has lines at the lower quartile, median (red line) and upper quartile values. Whiskers extend from each end of the box to the adjacent values in the data; in our case to the most extreme values within 1.5 times the interquartile range from the ends of the box. Outliers (e.g. data with values beyond the ends of the whiskers) are not displayed for visualization purposes, but they are distributed

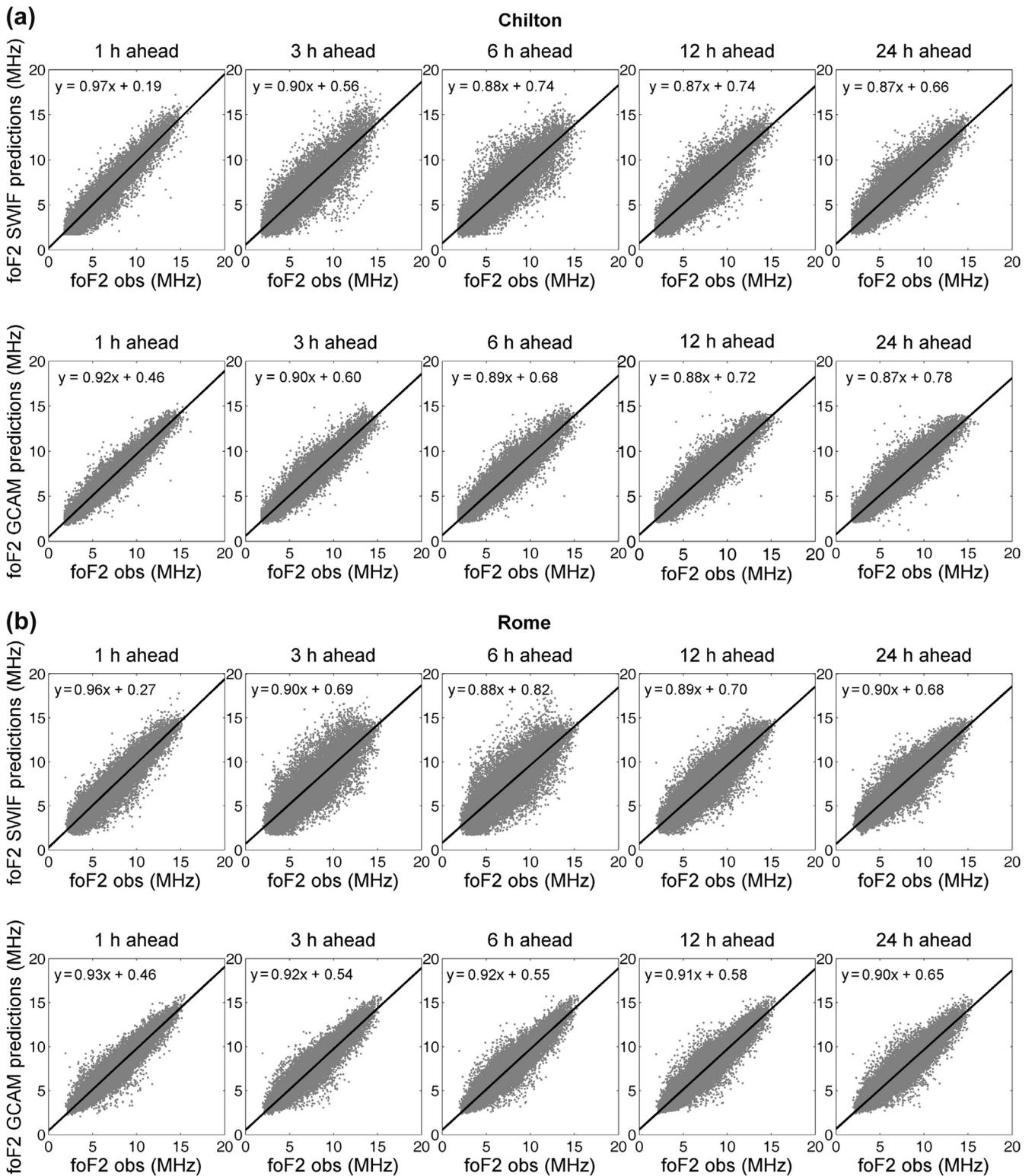


Fig. 2. Scatter plots of the SWIF (top panel) and GCAM (bottom panels) predictions for several prediction steps (1, 3, 6, 12, 24 h ahead) versus actual measurements for the time interval 1998–2007 over (a) Chilton and (b) Rome. The linear regression line is also plotted in all cases.

symmetrically in respect to the median in each case. In overall, the median error is close to zero and distributions appear to be symmetric in all of the cases indicating a “normal” and unbiased response of the forecasts in respect to actual observations. For the complete description of the error distributions, the mean errors (in MHz) and the standard deviations (in MHz) for all cases are also provided in Table 1. The mean error is also close

to zero indicating statistical accuracy for both models. The greater of the variance and consequently the minimum of precision is recorded for SWIF predictions provided 6 h ahead (except for Athens where the minimum is recorded at prediction step 3 h ahead). The variance of GCAM error increases slightly as the prediction step increases up to 12 h ahead and then remains constant.

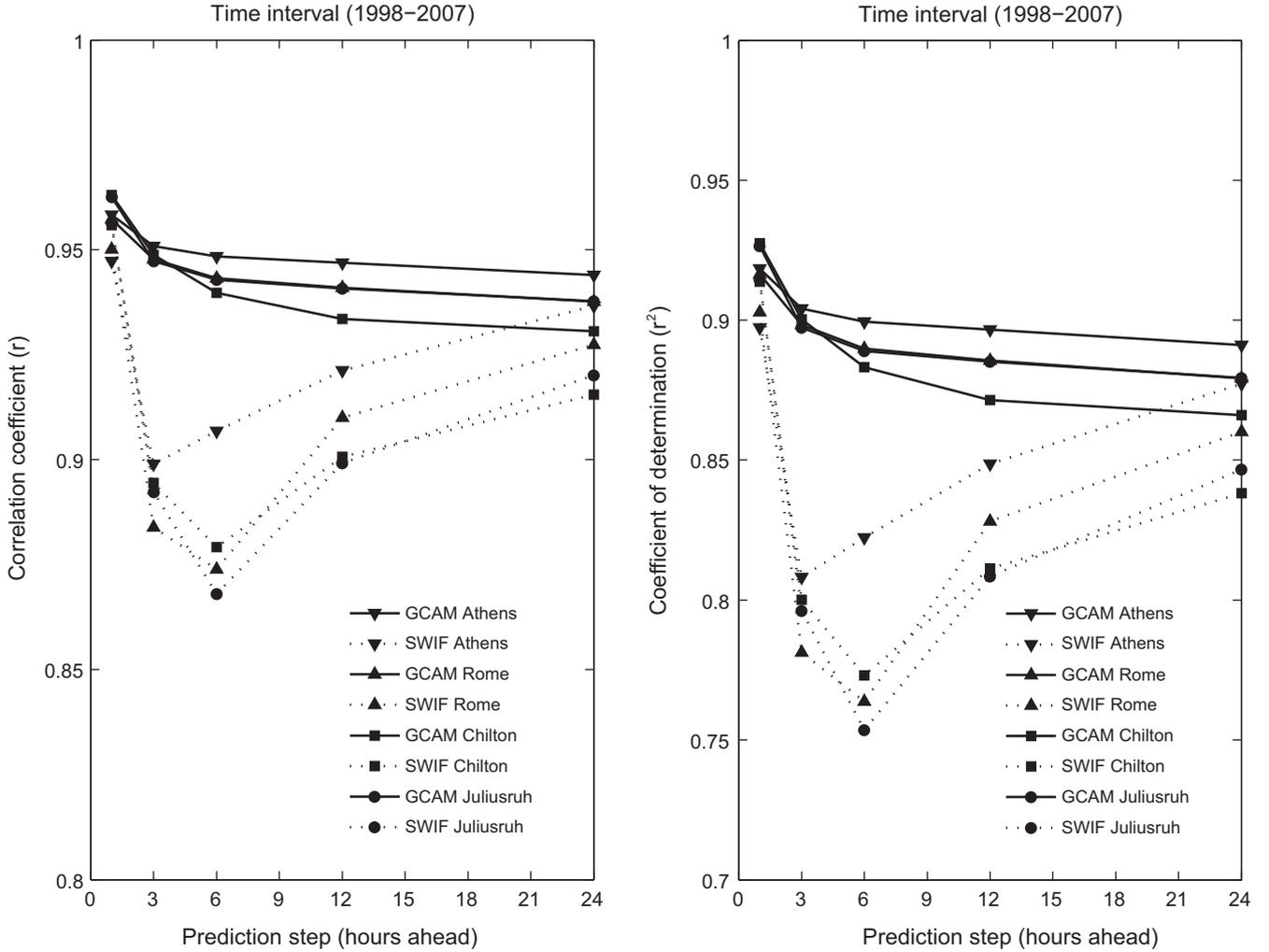


Fig. 3. The correlation coefficient r (left panel) and the coefficient of determination r^2 (right panel) comparing models' predictions with actual measurements over the period 1998–2007 in respect with the prediction step for four DIAS locations: Chilton, Juliusruh, Rome and Athens.

In summary, the above analysis provides evidence for: (i) consistent performance of each model in respect to the latitude (one may note a slightly different response of SWIF predictions in Athens, but as Athens data are limited to the time interval 2001–2007 and the differences are small, they are not considered significant), (ii) sufficient accuracy for both models' predictions, as the median and the mean errors range around zero and (iii) dependence of the accuracy and the precision of both models' predictions on the prediction step that seems more pronounced in SWIF case: the minimum of SWIF accuracy and precision is expected for predictions provided 6 h ahead.

The tests performed so far provide a good first look at the correspondence between forecasts and observations. More detailed diagnostics are based on the evaluation of complementary metric parameters over time. For this purpose, Figure 5 presents the coefficient of determination r^2 (top panel), the mean absolute error, MAE, defined as the averaged difference between model predictions and actual measurements in absolute values (second panel), the root mean square error, RMSE (third panel) and the mean relative error, MRE, defined as $MRE = \frac{1}{N} \sum_{i=1}^N \frac{|foF2_{obsi} - foF2_{modi}|}{foF2_{obsi}}$ (bottom panel), in respect to the prediction step and the year (from 1998 up to 2007). This test helps also the investigation of the models' performance in respect to the solar activity level. As no dependence of the models' performance on the latitude was suggested from the

previous analysis, this part of the analysis tests the overall performance of each model over Europe and the metrics estimates are averaged over all stations.

Based on Figure 5, the temporal variation of all metrics parameters exhibits a characteristic – for each model – pattern. This pattern is maintained for all prediction steps although some of its characteristics depend on the prediction step. Most importantly, SWIF's and GCAM's patterns present noticeable similarities regarding the solar cycle dependence. More precisely: (i) The linear correlation between predictions and observations is higher for both models during solar maximum years. This dependence is clear and significant (r^2 varies more than 15% between solar maximum and solar minimum years) for both models and all prediction steps. The correlation is higher for predictions provided 1 h ahead for both models and poorer for predictions provided 6 and 24 h ahead for SWIF and GCAM, respectively. (ii) The MAE and the RMSE tend to increase during solar maximum years. This trend is more pronounced for SWIF and more evident for predictions provided 3 and 6 h ahead (MAE and RMSE vary almost 40% between solar maximum and solar minimum years) while for GCAM is clear for predictions provided 12 and 24 h ahead (MAE and RMSE vary almost 30% between solar maximum and solar minimum years). Moreover, in SWIF case the greater values of MAE and RMSE are recorded for predictions provided 3 and

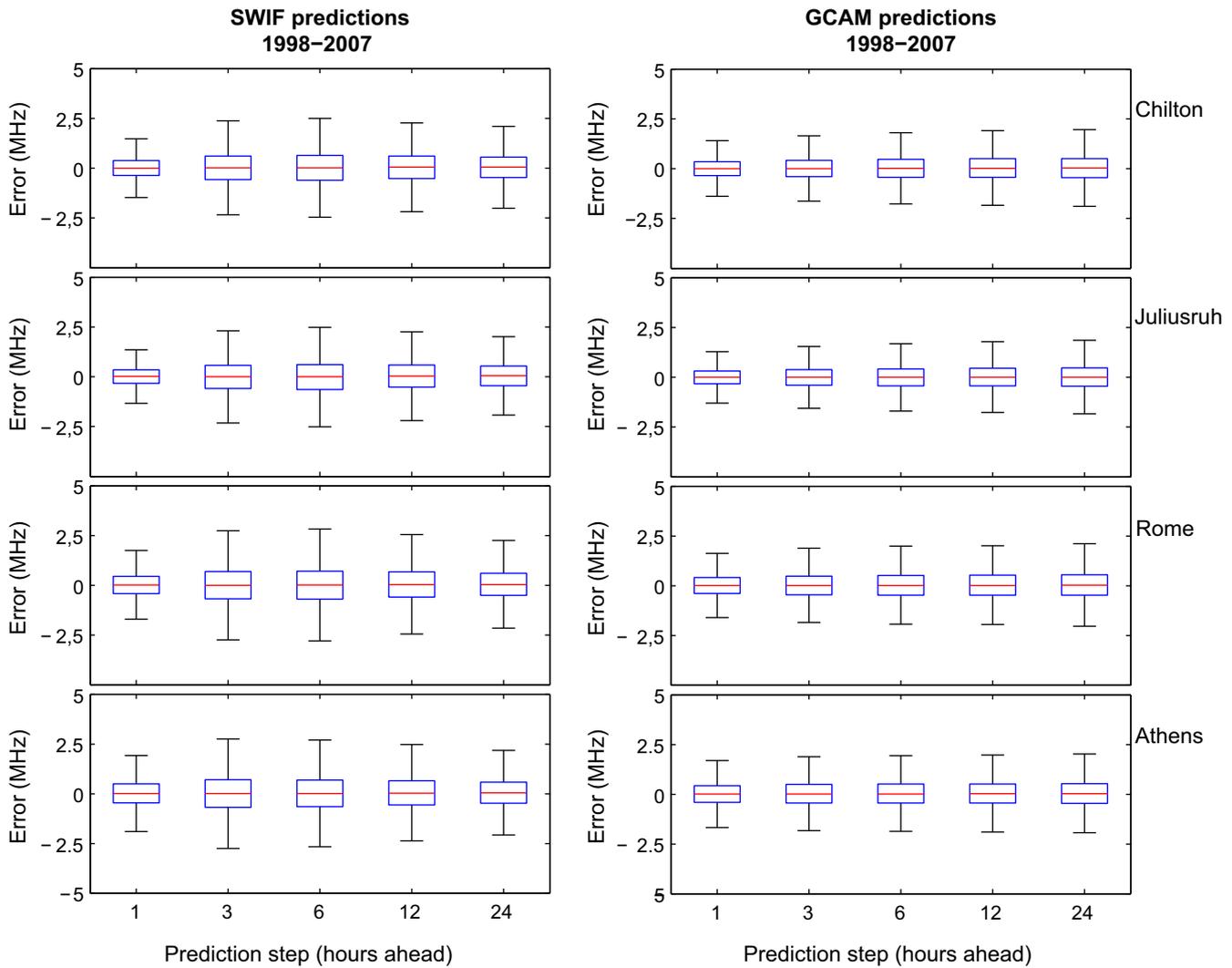


Fig. 4. Boxplots representing the models’ error distributions in respect to the prediction step. The box has lines at the lower quartile, median (red line) and upper quartile values. Whiskers extend from each end of the box to the adjacent values in the data; in our case to the most extreme values within 1.5 times the interquartile range from the ends of the box. Outliers (e.g. data with values beyond the ends of the whiskers) are not displayed for visualization purposes, but they are distributed symmetrically in respect to the median in each case and may reach ± 6 MHz for both models.

Table 1. Mean errors (ME in MHz) and standard deviations (SD in MHz) comparing the models’ predictions and actual measurements at all stations and several prediction steps.

	Chilton		Juliusruh		Rome		Athens									
	SWIF (ME, SD)	GCAM (ME, SD)														
1 h	0.01	0.69	0.00	0.62	0.01	0.72	-0.02	0.65	0.02	0.80	0.01	0.73	0.02	0.87	0.02	0.76
3 h	0.02	1.06	0.01	0.73	-0.01	1.12	-0.02	0.77	0.00	1.23	0.02	0.81	0.01	1.21	0.02	0.82
6 h	0.01	1.13	0.02	0.79	-0.04	1.23	-0.01	0.80	0.00	1.28	0.02	0.84	0.02	1.15	0.03	0.84
12 h	0.04	1.01	0.02	0.83	0.02	1.06	0.00	0.82	0.05	1.07	0.02	0.86	0.05	1.04	0.03	0.86
24 h	0.03	0.93	0.03	0.84	0.03	0.95	0.00	0.84	0.05	0.95	0.04	0.88	0.06	0.94	0.04	0.88

6 h ahead while for GCAM are recorded for predictions provided 12 and 24 h ahead. (iii) MRE is free of solar activity dependence for both models, indicating that the models’ performance is also free of solar activity dependence. However, it tends to keep the trends in the dependence of each model’s performance on the prediction step: greater for predictions provided 3 and 6 h ahead for SWIF (MRE around 15%) and 12 and 24 h ahead for GCAM (around 12%).

In the US National Space Weather Program Implementation Plan (2000), it is recommended to explore ionospheric forecast-

ing capabilities in terms of LT. It is further identifies four LT sectors, 03:00–09:00, 09:00–15:00, 15:00–21:00 and 21:00–03:00, and suggests the RMSE as the basic criterion for this test. To meet this requirement, the variation of the RMSE is presented in Figure 6a, in respect to the prediction step and the LT sector for two years: 2001 for high solar activity conditions and 2007 for low solar activity conditions. It is clear from Figure 6a that during solar maximum years RMSE receives also LT dependence being higher in the daytime hours (09:00–15:00 LT) for both models. In general, the RMSE variation seems to follow

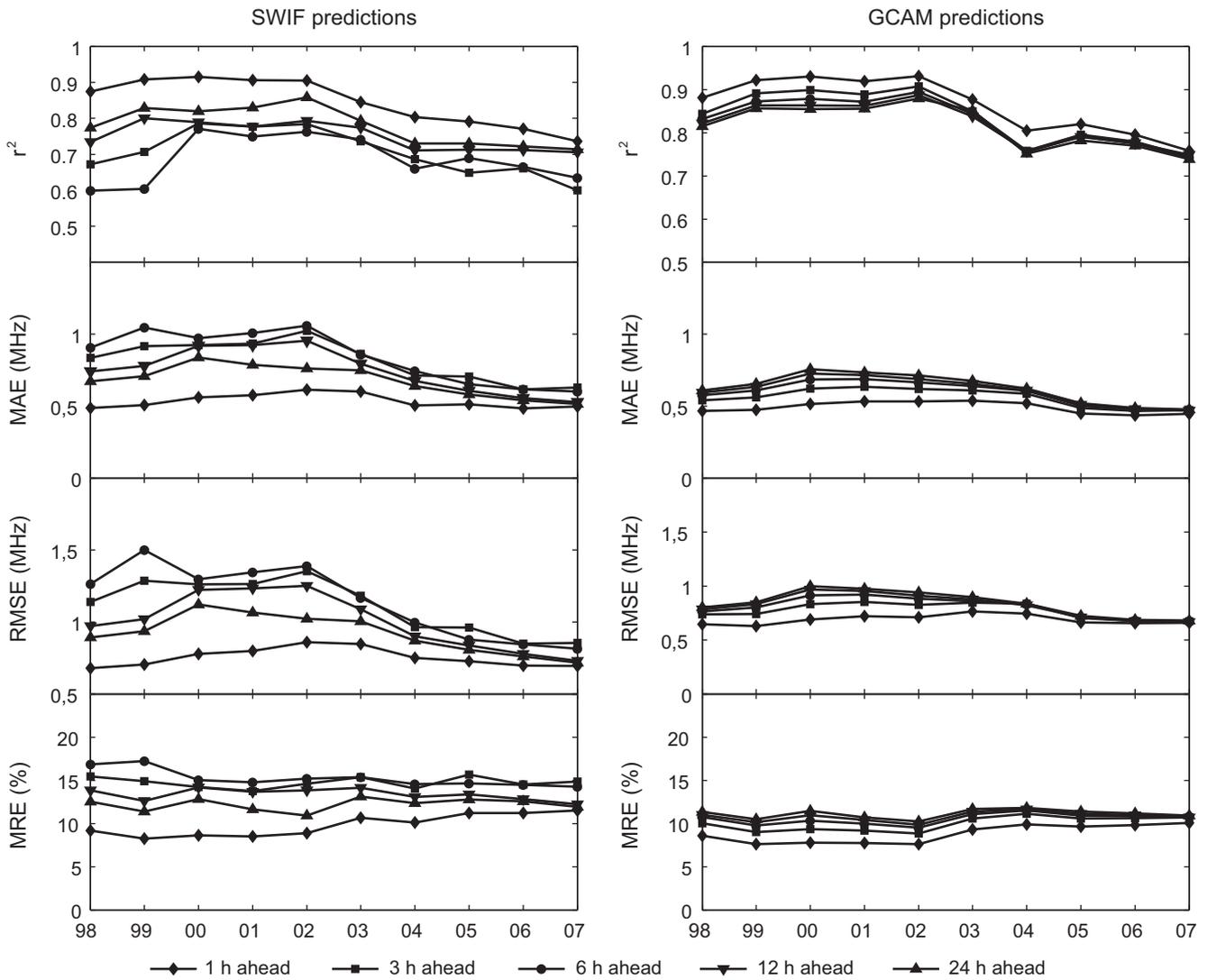


Fig. 5. The coefficient of determination r^2 (top panel), the mean absolute error – MAE (second panel), the root mean square error – RMSE (third panel) and the mean relative error – MRE (bottom panel), in respect to the prediction step and the year.

the diurnal and solar cycle variation of the foF2, being greater for solar maximum years and noon hours, indicating that the RMSE variation over solar cycle and LT reflects rather normal aspects of the foF2 variability than models' forecasting efficiency.

To further explore the models' forecasting capability over LT, Figure 6b also presents the variation of the MRE in respect to the prediction step and the LT sector for the years 2001 and 2007. The MRE is proved once again free of solar cycle dependence but the variation of the MRE in respect to the LT appears to be reversely proportional to the corresponding variation of the RMSE, tending to be smaller during daytime hours (09:00–15:00 LT). This indicates more successful performance of the models during daytime hours, but as the MRE varies less than 10% through the four LT sectors, its LT dependence and consequently the LT dependence of the models' performance are considered weak and rather negligible. The dependence on the prediction step is maintained in both MRE and RMSE local time variations.

At this point, it is worthy to discuss the prevailing geophysical conditions. For our test period (1998–2007) the MRE and the standard deviation comparing median estimates and actual

measurements is about 12.5% indicating quiet ionospheric conditions (Araujo-Pradere et al. 2003). Therefore, the results presented in this part of the analysis are mainly valid for quiet conditions.

The prediction efficiency tends to systematically depend on the prediction step, but at least for quiet conditions this dependence is rather weak: the MRE varies from 12% up to 15% for SWIF and from 8% up to 12% for GCAM. This dependence should be further explored during disturbed conditions. As the RMSE itself receives solar cycle and LT dependence, it may be treated with caution in comparative analyses. Alternatively, one may use the normalized RMSE, which is assigned to the range of actual measurements. During quiet conditions the normalized RMSE ranges between 5% and 6% for GCAM (maximum for predictions provided 12–24 h ahead) and between 5.5% and 8.5% for SWIF (maximum for predictions provided 6 h ahead).

As the predictions of both models are comparable to the climatological estimates (averaged MRE around 12.5% for median, 13.5% for SWIF and 10% for GCAM), no improvement of the models' predictions over climatology is noted during quiet conditions. Another test includes the evaluation of a model's performance against persistence. For ionospheric

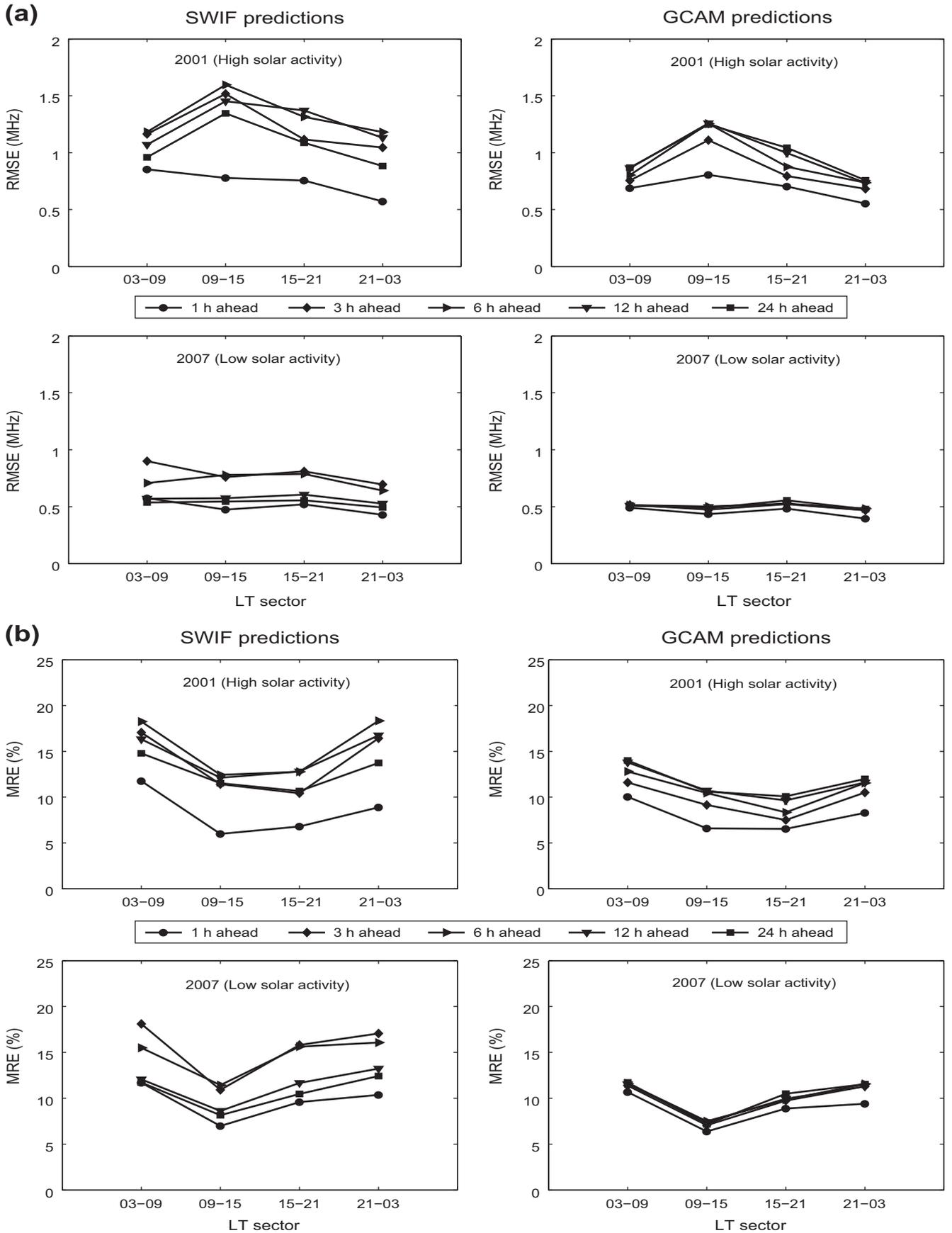


Fig. 6. The variation of the (a) RMSE of SWIF and GCAM and (b) MRE of SWIF and GCAM in respect of the prediction step and the LT sector for two years: 2001 for high solar activity conditions and 2007 for low solar activity conditions.

purposes, we consider important two persistence approaches: Per-1 in which the last observation is used as prediction for the current time and Per-24, in which observations from the previous day are used as prediction for the current day at the same UT time. For our test period, the MRE of persistence predictions from observations are calculated to be about 11% and 13.5% for Per-1 and Per-24, respectively. This indicates that during quiet conditions both SWIF and GCAM perform comparably also to persistence.

We believe that all the above verifies operational validity for both models. To conclude on the accuracy of the models, one has to investigate further their performance under disturbed conditions. This investigation is attempted in the next section.

3.2. Models' performance during disturbed ionospheric conditions

GCAM is driven by geomagnetic activity level while SWIF incorporates special component (e.g. STIM) for the prediction of storm time ionospheric disturbances. This suggests that both models were designed to make successful predictions especially during periods of enhanced geomagnetic activity. In addition, for operational purposes it is important to test the models' prediction efficiency during extremely hard conditions, which for ionospheric applications correspond to intense storm effects. This part of the analysis focuses mainly in the investigation of the two models' performance under such conditions.

Testing of the models was performed by using all available foF2 observations obtained at four European stations (Athens, Chilton, Juliusruh and Rome) during strongly disturbed periods occurred in the time interval 1998–2007. The selection of the test intervals was initially driven by the determination of all disturbed periods over each single location. Only significant disturbances ($> 20\%$ over median) of long duration (e.g. greater than 3 h) were considered in this selection. Then the disturbed periods were assigned to the occurrence of geomagnetic storm events (based on Ap and Dst indices) to finally keep disturbed periods related to the occurrence of intense ($-200 \text{ nT} \leq \text{Dst} < -100 \text{ nT}$) or big ($\text{Dst} < -200 \text{ nT}$) storm events (Gonzalez et al. 1999). Trying to apply the correct criteria for model evaluation (e.g. Pittock 1978; Araujo-Pradere et al. 2003), we kept out of this database the events used for the development of SWIF (Tsagouri & Belehaki 2008; Tsagouri et al. 2009). Such a treatment was not possible for GCAM case, since all available observations up to 2006 were used for the determination of the model empirical expression parameters for the implementation in DIAS system. In that respect, the assessment of GCAM model performance discussed in this paper may not be considered as an objective one but serves rather the investigation of the internal consistency of the model. Nevertheless, as this is also a critical aspect of a model's performance and in addition may provide a valuable background for comparison purposes with SWIF and/or other existing or future models, the analysis of GCAM predictions is included in this investigation. The most disturbed days in the course of each storm, which are defined as storm days, were extensively analyzed. As the ionospheric storm-time response exhibits latitudinal dependence, this part of the analysis was performed by distinguishing two latitudinal zones, the middle-to-high and the middle-to-low latitude zones following STIM's classification (see Sect. 2.2).

As for the quiet conditions, the correlation between observed and modeled values was first investigated. The results are provided in Figure 7 in terms of the coefficient of determi-

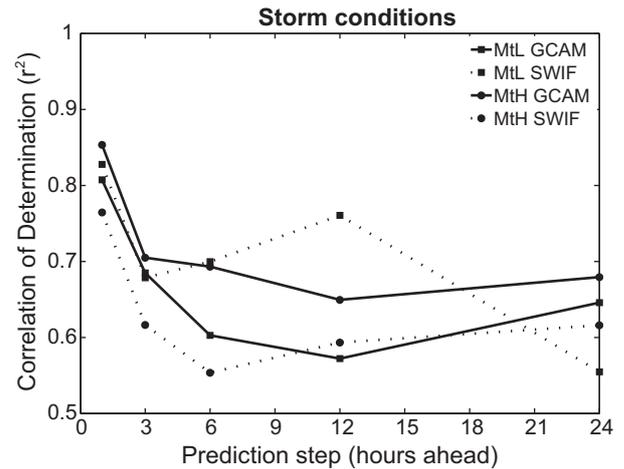


Fig. 7. The coefficient of determination, r^2 comparing models' predictions with actual measurements during storm conditions in respect to the prediction step, separately for middle-to-high and middle-to-low latitudes.

nation, r^2 . According to our findings, the models' prediction efficiency during storm conditions is still strongly dependent on the prediction step, although the pattern of the dependence is now slightly different: the minimum efficiency for the GCAM is recorded for predictions provided 12 h ahead and then it is partially recovered. For SWIF the minimum of the efficiency tends to be recorded for predictions provided 3 and 6 h ahead, but the correlation is also significantly poorer for predictions provided 24 h ahead at lower latitudes. It is also evident that both models' performance is subject to latitudinal dependence during storm conditions. The dependence tends to be more significant for predictions provided 6 to 12 h ahead. In general, the correlation between observed and modeled values during storm conditions seems to be lower than in quiet conditions, indicating that under disturbances the prediction error is higher. The potential effect of systematic offsets is investigated below.

To further investigate the last point, the distributions of the models' prediction errors were examined. The results are presented in Figure 8 in a boxplot format (following the format of Fig. 4), separately for middle-to-high and middle-to-low latitudes. For comparison purposes, Figure 8 includes also the corresponding results received for the comparison between median-based predictions with actual observations. This aims to test critical aspects of the models' performance in terms of the disturbances' features. Concerning the median-based predictions, the distribution of their prediction error is mainly located within negative values' domain in both latitudinal zones. This indicates the occurrence of mainly negative storm effects over Europe during the selected periods. The distributions of models' prediction errors seem to be rather symmetric, but they are now displaced toward negative values in respect to the corresponding ones received for quiet conditions, giving further evidence that the prediction error is increased during disturbed conditions. Taking into account the occurrence of negative storm effects, the last point indicates that the models' predictions tend to systematically underestimate the intensity of the disturbances. The accuracy and the precision of both models' performance depend on the prediction step and the latitude of the observation point following the pattern determined by the variation of r^2 . An interesting point is that the distributions of the models' errors are clearly displaced upwards in respect to the distributions of the median estimates' errors,

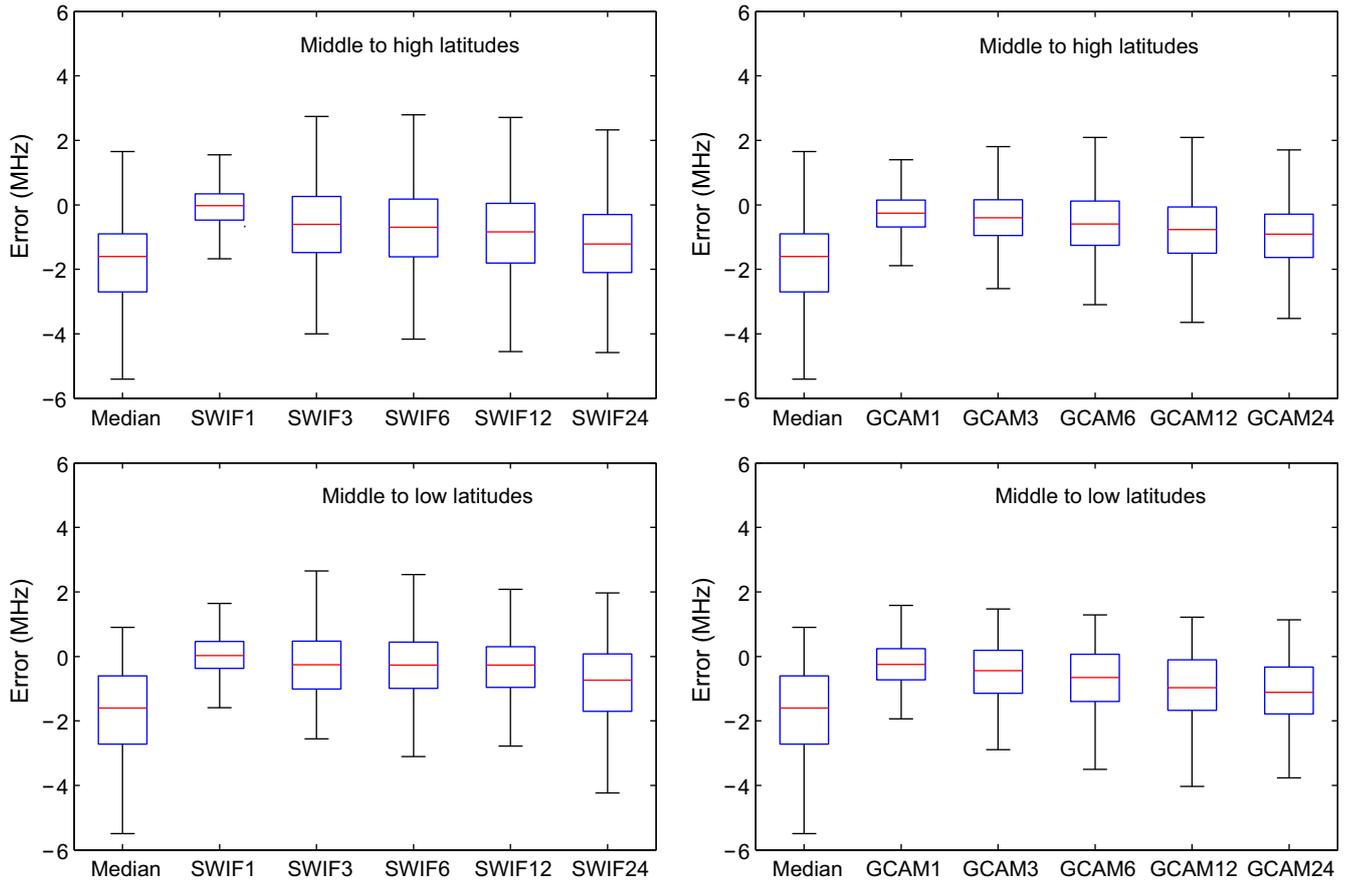


Fig. 8. Boxplots representing the models’ error distributions for storm conditions in respect to the prediction step (following the format of Fig. 4), separately for middle-to-high and middle-to-low latitudes. The corresponding results for the median estimates in comparison to actual observations are included in all cases.

Table 2. The mean errors (ME in MHz), the normalized RMSE (e.g. 0.48 represents a RMSE of 48%) and the standard deviation (SD in MHz) comparing the monthly median (MM), the persistence (Per-1 and Per-24) and models’ predictions with ionosonde observations.

	MM	Per-1	Per-24	GCAM					SWIF				
				Prediction step (hours ahead)					Prediction step (hours ahead)				
				1	3	6	12	24	1	3	6	12	24
(A) Middle to high latitudes													
ME (MHz)	-1.87	-0.05	-0.79	-0.26	-0.42	-0.55	-0.75	-0.95	-0.11	-0.71	-0.85	-0.98	-1.26
RMSE	0.48	0.16	0.41	0.16	0.23	0.24	0.28	0.31	0.14	0.32	0.34	0.34	0.38
SD (MHz)	1.08	0.49	1.02	0.47	0.66	0.63	0.70	0.72	0.54	0.81	0.85	0.88	0.96
(B) Middle to low latitudes													
ME (MHz)	-1.69	0.03	-1.01	-0.34	-0.57	-0.79	-1.08	-1.23	0.06	-0.10	-0.14	-0.29	-0.92
RMSE	0.42	0.15	0.40	0.17	0.23	0.27	0.31	0.31	0.15	0.20	0.20	0.22	0.29
SD (MHz)	1.09	0.43	1.13	0.52	0.67	0.80	0.92	0.86	0.46	0.58	0.57	0.71	0.76

tending to decrease the median-based prediction error. This indicates that both models’ predictions capture the disturbances and provide improvement over climatological predictions. The latter will be further explored below.

The mean errors (ME), the normalized daily RMSE (%) and the standard deviations (in MHz) comparing models’ predictions with actual measurements were next estimated as metric parameters to quantify the models’ performance during storm days. The averaged results over all storm days separately for middle-to-high and middle-to-low latitudes are presented in Table 2. For comparison purposes, the respective results

obtained for monthly median but also for persistence predictions are also included in Table 2. The MRE of the median estimates in respect to the actual measurements for the testing days are 38% and 27% for middle-to-high and middle-to-low stations, respectively. This verifies very disturbed ionospheric conditions during the selected days (Araujo-Pradere et al. 2003).

The results provided in Table 2 reveal three major trends: (i) considerable improvement of both models’ predictions over climatology and Per-24 approaches, while persistence performs comparable to GCAM and SWIF for predictions provided 1 h

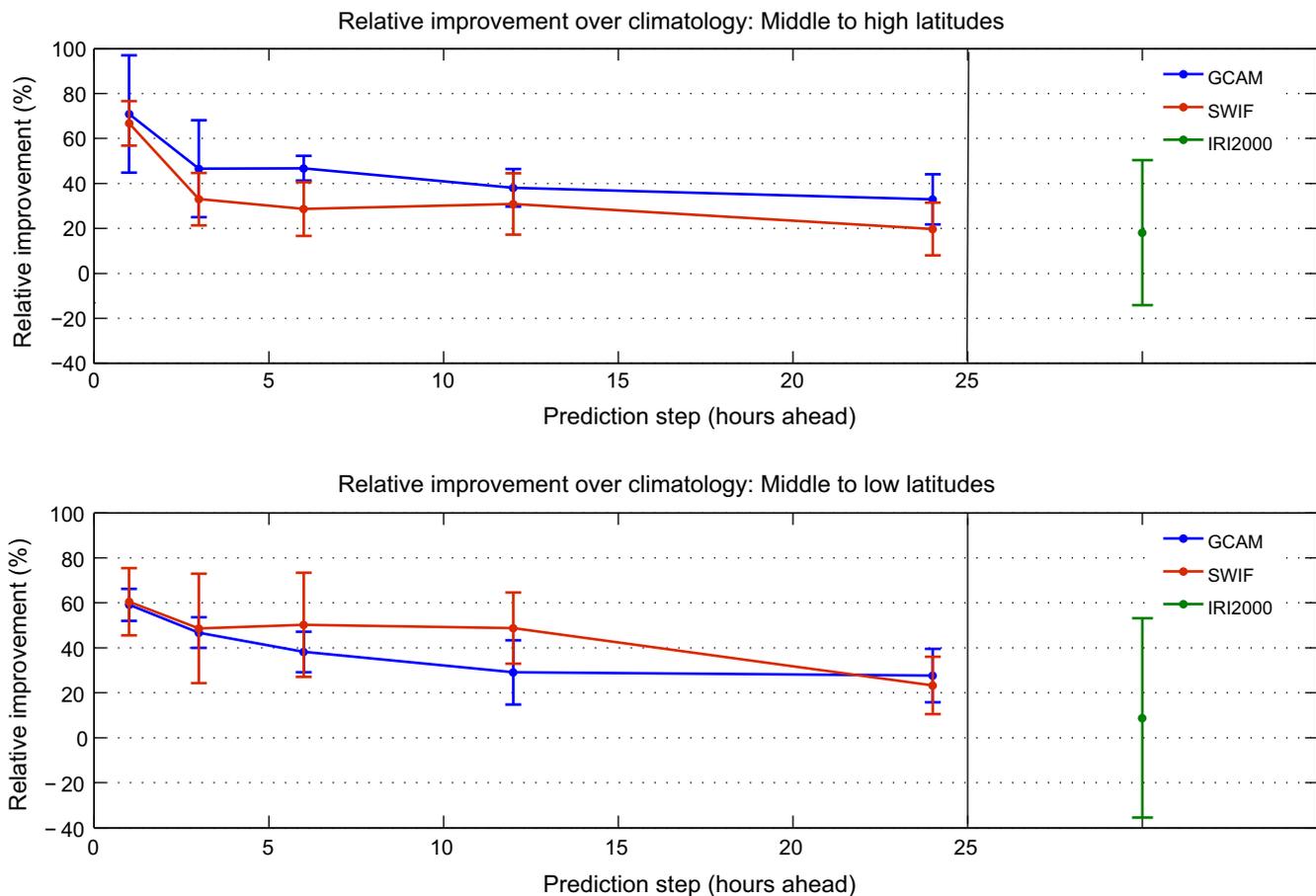


Fig. 9. The improvement over climatology for GCAM and SWIF separately for middle-to-high (top) and middle-to-low (bottom) latitudes in respect to the prediction step for storm days. Error bars denote standard deviations from the mean. For comparison purposes, the relative improvement provided by IRI2000 predictions over climatology for the same set of events is also provided.

ahead (Per-1 case), (ii) dependence of the relative improvement on the prediction step (in particular, the improvement decreases as the prediction step increases) and (iii) dependence of the relative improvement on the latitude: at middle-to-high latitudes GCAM exhibits greater improvement over both climatology and persistence, while at middle-to-low latitudes SWIF provides more successful results.

To quantify the relative improvement of the models' accuracy over climatology and persistence, the % improvement parameter is used given by the generalized formula:

$$\% \text{ improvement} = \frac{\text{RMSE}_Y - \text{RMSE}_X}{\text{RMSE}_Y} \times 100,$$

where X stands for the model predictions and Y for median or persistence predictions (Tsagouri et al. 2009).

The results obtained for all combinations and prediction steps regarding the comparison between model and climatological estimates are given in Figure 9 in terms of the averaged improvement over all storm days. The error bars indicate the standard deviation of the estimates in respect to the mean value. The results verify considerable improvement (greater than 10%) in all cases (Araujo-Pradere & Fuller-Rowell 2002) and the general trends reported previously. The averaged improvement ranges: (i) for GCAM case between 70% and 30% at middle-to-high latitudes and between 60% and 30% at middle to low latitudes, and (ii) for SWIF case between 70% and 20% at middle-to-high latitudes and between 60% and 25% at middle-to-

low latitudes. For comparison purposes, Figure 9 includes also the results obtained under the current investigation from the analysis of IRI2000 (Bilitza 2001) predictions for the same set of test days. The IRI2000 version of IRI (<http://ulcar.uml.edu/>) includes the STORM model (Araujo-Pradere et al. 2002) as the storm-time correction, which contains a non-linear dependence of the ionospheric response on the integrated time history (over 33 h) of Ap index. With this addition, IRI2000 is able to capture much of the repeatable characteristics of the storm-time ionospheric response (Araujo-Pradere et al. 2003). The IRI model is recommended by the Center for Integrated Space weather Modeling (CISM) as the “baseline model” for a metrics-based ionospheric model evaluation purposes (Spence et al. 2004). In general, although the performance of the IRI2000 varies significantly within the selected set of storm days, the averaged results indicate more successful performance of GCAM and SWIF than IRI for the selected periods.

The improvement over persistence is investigated in Figure 10. As both models perform comparable to persistence for predictions provided 1 h ahead (see Tab. 2), Figure 10 includes the results for the comparison of models' predictions for greater prediction steps with Per-24 predictions. This comparison was performed for storm days but also for post-storm days (the days that follow the storm days in each event). The latter is considered important in order to investigate the models' efficiency in following the recovery to normal conditions. Although the improvement varies considerably around the

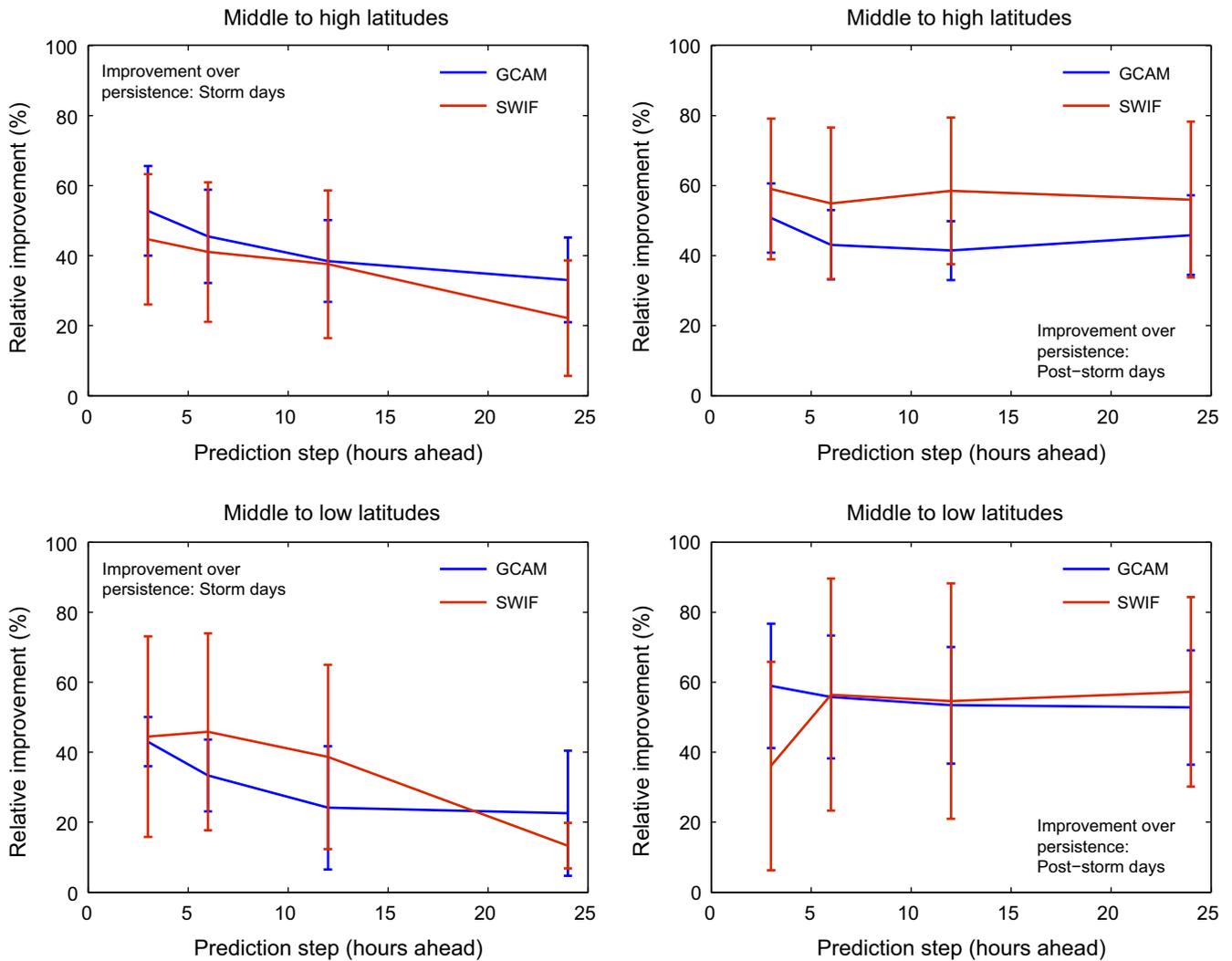


Fig. 10. The improvement over persistence (Per-24) for GCAM and SWIF separately for middle-to-high (top) and middle-to-low (bottom) latitudes and storm (left) and post-storm days (right) in respect to the prediction step. Error bars denote standard deviations from the mean.

averaged estimates, on average the results verify significant improvement over persistence in almost all cases. Marginal improvement is noted only for predictions provided 24 h ahead during storm days especially for SWIF case. The improvement is greater during post-storm days, which is reasonable as Per-24 predictions for post-storm days “carry” the storm-time variation. During storm days, Per-24 may provide more successful predictions under the occurrence of successive storm days.

Finally, to comment also on the precision of the models’ predictions one may consider the standard deviation estimates. Visual inspection of Table 2 gives evidence that both models reduce significantly the standard deviations of MM and Per-24 (in comparison to models’ predictions provided more than 3 h ahead) predictions. For predictions provided 1 h ahead, standard deviation estimates are comparable to Per-1. Following the definition of the % improvement parameter provided above and by using the SD as the metric criterion, the precision improvement over climatological predictions is estimated to range from 39% to 33% for GCAM and from 25% to 11% for SWIF at middle to high latitudes, while at lower latitudes range from 39% to 21% and from 47% to 30%, for GCAM and SWIF, respectively. Slightly lower and higher values are obtained for higher and lower latitudes, respectively, against Per-24. These results are applied to predictions provided more than 1 h ahead.

3.3 Determination of models’ accuracy

The two models’ accuracy is further explored through the variation of the MRE between models’ predictions and actual measurements in respect to the prediction step under all possible geophysical conditions. The MRE is chosen as the metric criterion for this part of the analysis as it is free of any other dependence (e.g. LT and solar cycle) but it keeps the prediction step dependence that was clearly seen in all tests. This gives MRE significant advantage in assessing ionospheric forecasting capabilities as it may be directly comparable to similar results obtained in other studies. The results for both models are demonstrated in Figure 11 and include: (i) MRE estimates for quiet conditions (labeled as q) obtained from the analysis of models’ predictions and observations for the time interval 1998–2007 (see Sect. 3.1), (ii) MRE estimates for very disturbed conditions (labeled as d) obtained from the analysis of models’ predictions and observations for storm days as it is described above and (iii) MRE estimates for moderately disturbed conditions (labeled as m) obtained from the analysis of models’ predictions and observations for all disturbed periods that were initially selected, regardless of their relation to a storm occurrence (excluding the storm cases used above). This dataset may include quiet time disturbances or disturbances associated to storms of weak

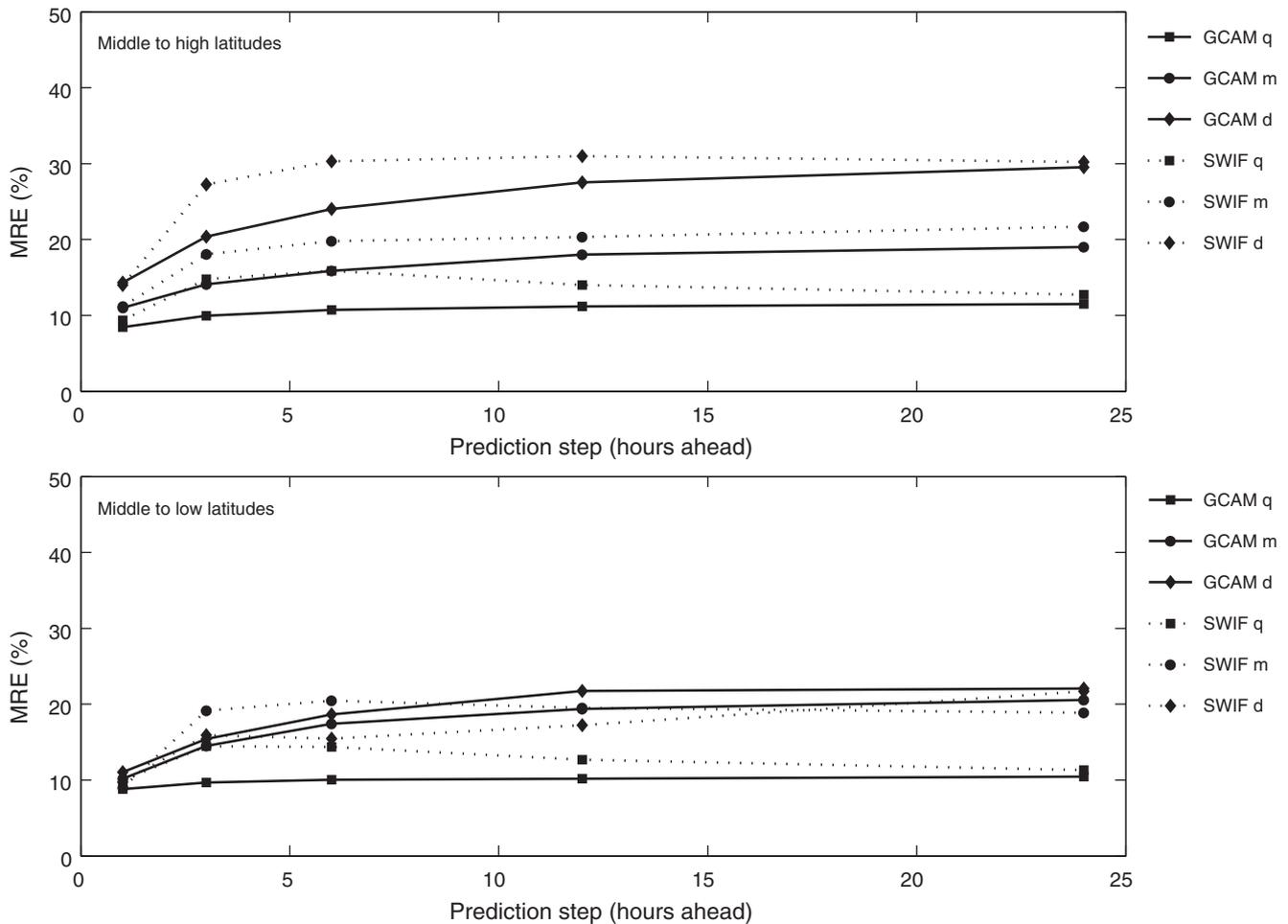


Fig. 11. MRE estimates for both models in respect to the prediction step and the ionospheric activity level: q stands for quiet, m for moderately disturbed and d for extremely disturbed conditions (see text for more explanation).

or moderate intensity. The MRE of the median estimates in respect to actual observations for the latter case is about 22% for all latitudes indicating marginally significant disturbances.

For both models, the MRE depends on the prediction step, the level of the ionospheric activity and the latitude of the observation point. It is relatively small (10–13%) for predictions provided 1 h ahead and may reach 30% at middle-to-high latitudes for high ionospheric activity level and predictions provided 24 h ahead. The dependence on the ionospheric activity follows similar pattern for both models: for higher latitudes the errors systematically increase as the ionospheric activity level increases while for lower latitude the prediction errors during very disturbed and moderately disturbed conditions are comparable for both models. The dependence on the prediction step differs slightly for two models: for GCAM the MRE increases constantly with the prediction step up to 12 h and remains constant afterwards, while for SWIF the greater MRE is usually recorded for predictions provided 3 and 6 h ahead. Then it remains constant or it is partially recovered. However, as the differences are not significant, one may consider this finding only as evidence of a general trend.

4. Summary and discussion

In this paper, a metrics-based evaluation of two ionospheric forecasting models, the SWIF (Tsagouri et al. 2009) and GCAM (Muhtarov et al. 2002), implemented online in DIAS

system (<http://dias.space.noa.gr>) to provide ionospheric forecasts of the foF2 parameter over Europe is attempted. As metrics measure performance rather than provide specifics about models' deficiencies (Spence et al. 2004), the validation and verification plans were designed to test the operational validity of the two models and to quantify the models' prediction accuracy. In that respect, the results presented here do not include any component of purely scientific assessment of the models' performance (e.g. case studies) and they are not suggestive for any improvements. A more scientifically oriented validation plan was implemented during the development phase of each model (see Tsagouri et al. 2009; Muhtarov et al. 2002). In addition, as the operational implementation of the two models in DIAS aims to provide ionospheric forecasts for the European middle latitude ionosphere the evaluation tests were carried out for this specific region and the validity of the results is partially limited to this region.

The analysis was established on the systematic comparison of models' predictions with actual observations obtained at four European ionospheric locations (e.g. Athens, Chilton, Juliusruh and Rome) over the period 1998–2007 that fairly corresponds to the duration of one solar cycle. This allowed testing of the models' operational and prediction efficiency under all possible conditions. In addition, the models' performance was extensively assessed by complementary analysis of several metrics parameters. Overall, the results demonstrate consistency and sufficient accuracy of both models' performance over the years. Moreover,

despite the quantitative differences, appreciable qualitative similarities were detected in the solar cycle and LT trends of the estimated metric parameters for the two models. These findings strongly support the operation validity of both DIAS models.

Although the solar cycle and the LT dependence were clear for most of the metrics parameters estimated here, the complementary analysis of all of them gives evidence that the performance of the two DIAS models is free of solar cycle and LT dependence. The main trend survived through all tests is the dependence of both models' performance on the prediction step: for GCAM the prediction error is progressively increased for predictions provided up to 12 h ahead tending to remain constant for greater prediction steps, while for SWIF it is rapidly increased for predictions provided up to 3 and 6 h ahead and then it remains constant or it is partially recovered. The dependence of the models' performance on the prediction step is rather weak during quiet conditions and all middle latitudes but it is becoming stronger as the level of the ionospheric activity increases. Here it is worthy to note that the dependence of ionospheric forecasting capabilities on the prediction step is a common feature in many similar studies (e.g. Mikhailov et al. 2007) and may be attributed to the profound diurnal variation of the foF2 parameter.

The models' prediction efficiency depends also on the level of the ionospheric activity and the latitude of the observation point especially during extremely disturbed conditions. In general, the error tends to be higher for higher latitudes and higher ionospheric activity levels. This tendency seems to follow the general trends of the ionospheric variability (Araujo-Pradere et al. 2004). The models' accuracy in respect to the prediction step was quantified under all levels of ionospheric activity in terms of the MRE that was proven to be free of diurnal and solar cycle dependence. During quiet conditions MRE ranges from 12% up to 15% for SWIF and from 8% up to 10% for GCAM. During disturbed conditions, it is still relatively small (10–13%) for predictions provided 1 h ahead and may reach 20% and 30% at middle-to-low and middle-to-high latitudes, respectively, for high ionospheric activity level and predictions provided 24 h ahead.

During quiet and moderately disturbed conditions the two models perform comparable to climatological predictions, expressed by monthly median estimates, and persistence predictions provided either 1 or 24 h before. During disturbed ionospheric storm conditions considerable improvement (greater than 10%) over climatology and Per-24 is gained for all prediction steps and at all latitudes. The improvement depends on the latitude and the prediction step tending to be reduced as the prediction step is increased. The precision improvement over climatology ranges between 21% and 39% for GCAM and between 11% and 47% for SWIF. The averaged accuracy improvement over climatology ranges from 30% to 70% for GCAM and from 25% to 70% for SWIF. The testing of the models' performance against climatology gave evidence of the models' efficiency to capture the disturbances, although models' predictions seem to underestimate the intensity of the disturbances.

The relative improvement over Per-24 was also investigated, separately for storm and post-storm days and it was proved to be significant at almost all cases. In particular, for storm days it ranges from 20% to 55% for GCAM and from 10% to 45% for SWIF. The improvement is higher during post-storm days, ranging from 40% to 60% for both models. This investigation verified the models' efficiency to capture successfully the transition from disturbed to quiet conditions.

A fair comparison with other models' performance ideally requires testing of all models' prediction efficiency with comparable databases and analysis of comparable metric parameters. In an effort to roughly assess GCAM'S and SWIF'S forecasting efficiency in respect to widely accepted ionospheric prediction models, the two models' performance was also compared with that of IRI2000 during storm conditions in terms of the relative accuracy improvement over climatology. The IRI2000 version of IRI includes the STORM model (Araujo-Pradere et al. 2002) as the storm-time correction and it is able to capture much of the repeatable characteristics of the storm-time ionospheric response (Araujo-Pradere et al. 2003). The IRI model is usually recommended as the "baseline model" for a metrics-based ionospheric model evaluation purposes (Spence et al. 2004). Based on our findings, the averaged relative improvement over climatology for IRI2000 over Europe is estimated to be about 20% and 10% for middle-to-high and middle-to-low latitudes, respectively, indicating that GCAM and SWIF capture the disturbances more successfully than IRI2000 during the selected periods for all prediction steps. Similar results reported in the literature include accuracy gain of 29% over climatology obtained by Kutiev & Muhtarov (2001) for their prediction model and gain of 33% for STORM model (Araujo-Pradere & Fuller-Rowell 2002). In a recent review, Mikhailov et al. (2007) tested their model's performance in terms of MRE estimates. For quiet conditions typical MRE up to 10–15% were received, while during severe storm events the tests were performed in comparison to the performance of IRI2000. Their results report MRE estimates up to 24% depending on the prediction step and the season for their model and an averaged MRE of 29% for IRI2000. Although one has to keep in mind that the results reported in the literature were derived from a different than our set of test events and stations, they may support the argument that SWIF and GCAM perform successfully in alignment with current ionospheric capabilities, but in fully operational mode.

Finally, it should be noted that the testing database used in this investigation was included in the one used for the calculation of the GCAM'S coefficients before its implementation in the DIAS system. Therefore, the evaluation of GCAM performance, as it is discussed in this paper, is not actually objective and the obtained results are not directly comparable to the ones received from the evaluation of SWIF'S performance so that the comparison between the two models is out of the scope of this paper. Here, the assessment of GCAM performance serves better the investigation of the model's internal consistency, but this is also critical in the evaluation of a model's performance. In addition, it provides a valuable background for comparison purposes. An interesting outcome of the current investigation is that similar trends were detected in both models' performance. Moreover, comparable results in quantitative aspect were obtained in several cases. Based on this discussion one may argue that the current study provides the research community with a well-established pattern regarding current ionospheric forecasting capabilities that could be efficiently exploited as the basis for comparative studies either for operational applications or for future developments. In that respect, the results obtained here may fulfill the objectives of COST ES0803 Action regarding the assessment of existing ionospheric forecasting models.

Acknowledgements. This work is supported by the European Action COST ES0803 "Developing Space Weather Products and Services in Europe". The author is grateful to Dr. A. Mikhailov

and Dr. L. Perrone for their contribution in the selection of the test events used in this investigation to evaluate the models' performance during disturbed ionospheric conditions.

References

- Araujo-Pradere, E.A., and T.J. Fuller-Rowell, STORM: An empirical storm-time ionospheric correction model, 2. Validation, *Radio Sci.*, **37**, 1071, 2002.
- Araujo-Pradere, E.A., T.J. Fuller-Rowell, and D. Bilitza, Validation of the STORM response in IRI2000, *J. Geophys. Res.*, **108** (A3), 1120, DOI: [10.1029/2002JA009720](https://doi.org/10.1029/2002JA009720), 2003.
- Araujo-Pradere, E.A., T.J. Fuller-Rowell, and D. Bilitza, Ionospheric variability for quiet and disturbed conditions, *Adv. Space Res.*, **34**, 1914–1921, 2004.
- Araujo-Pradere, E.A., T.J. Fuller-Rowell, and M. V. Codrescu, STORM: An empirical storm-time ionospheric correction model, 1. Model description, *Radio Sci.*, **37**, 1070, DOI: [10.1029/2001RS002447](https://doi.org/10.1029/2001RS002447), 2002.
- Belehaki, A., L.R. Cander, B. Zolesi, J. Bremer, C. Juren, I. Stanislawski, D. Dialetis, and M. Hatzopoulos, Monitoring and forecasting the ionosphere over Europe: The DIAS project, *Space Weather*, **4**, S12002, DOI: [10.1029/2006SW000270](https://doi.org/10.1029/2006SW000270), 2006.
- Belehaki, A., L.R. Cander, B. Zolesi, J. Bremer, C. Juren, I. Stanislawski, D. Dialetis, and M. Hatzopoulos, Ionospheric specification and forecasting based on observations from European ionosondes participating in DIAS project, *Acta Geophys.*, **55** (3), 398–409, 2007.
- Belehaki, A., I. Stanislawski, and J. Lilensten, An overview of ionosphere-thermosphere models available for space weather purposes, *Space Sci. Rev.*, **147** (3–4), 271–313, DOI: [10.1007/s11214-009-9510-0](https://doi.org/10.1007/s11214-009-9510-0), 2009a.
- Belehaki, A., J. Watermann, A. Lilensten, M. Glover, M. Hapgood, M. Messerotti, R. van der Linden, and H. Lundstedt, Renewed support dawns in Europe: An action to develop space weather products and services, *Space Weather – Int. J. Res. Appl.*, **7**, art. no. S03001, 2009b.
- Bilitza, D., International reference ionosphere 2000, *Radio Sci.*, **36** (2), 261–276, 2001.
- Cander, L.R., Toward forecasting and mapping ionospheric space weather under the COST actions, *Adv. Space Res.*, **31** (4), 957–964, 2003.
- Doggett, K., (ed.), *Proceedings of a Workshop at Boulder*, Colorado, June 19–21, 1996, NOAA/SEC, Boulder, 1996.
- Galkin, I.A., G.M. Khmyrov, A. Kozlov, B.W. Reinisch, X. Huang, and D.F. Kitrosser, Ionosonde networking, databasing, and Web serving, *Radio Sci.*, **41** (5): art. no. RS5S33, 2006.
- Gonzalez, W.D., B.T. Tsurutani, and A.L.C. de Gonzalez, Interplanetary origin of geomagnetic storms, *Space Sci. Rev.*, **88**, 529–562, 1999.
- Hesse, M., P. Bellaire, and R. Robinson, Community Coordinated Modeling Center: A new approach to space weather modeling, *Proceedings of the Space Weather Workshop: Looking Towards a European Space Weather Programme*, 17–19 December 2001, ESTEC, Noordwijk, The Netherlands, 2001.
- Koutroumbas, K., and A. Behlaki, One-step ahead prediction of foF2 using time series forecasting techniques, *Ann. Geophys.*, **23**, 3035–3042, 2005.
- Koutroumbas, K., I. Tsagouri, and A. Behlaki, Time series autoregression technique implemented on-line in DIAS system for ionospheric forecast over Europe, *Ann. Geophys.*, **26** (2), 371–386, 2008.
- Kutiev, I., and P. Muhtarov, Modeling of midlatitude F region response to geomagnetic activity, *J. Geophys. Res.*, **106** (A8), 15501–15509, 2001.
- Kutiev, I., and P. Muhtarov, Empirical modeling of global ionospheric foF(2) response to geomagnetic activity, *J. Geophys. Res.*, **108** (A1): art. no. 1021, 2003.
- McKinnell, L.A., and A.W.V. Poole, Ionospheric variability and electron density profile studies with neural networks, *Adv. Space Res.*, **27** (1), 83–90, 2001.
- Mikhailov, A.V., V.H. Depuev, and A.H. Depueva, Short-term foF2 forecast: Present day state of art, in space weather: Research towards applications in Europe, *Astrophys. Space Sci. Libr.*, **344**, 169–184, 2007.
- Muhtarov, P., and I. Kutiev, Autocorrelation method for temporal interpolation and short-term prediction of ionospheric data, *Radio Sci.*, **34** (2), 459–464, 1999.
- Muhtarov, P., I. Kutiev, and L.R. Cander, Geomagnetically correlated autoregression model for short-term prediction of ionospheric parameters, *Inverse Prob.*, **18**, 49–65, 2002.
- NSWP – National Space Weather Program Implementation Plan, 2nd edition, July 2000 (<http://www.ofcm.gov/>).
- Pietrella, M., and L. Perrone, A local ionospheric model for forecasting the critical frequency of the F2 layer during disturbed geomagnetic and ionospheric conditions, *Ann. Geophys.*, **26**, 323–334, 2008.
- Pitcock, A.B., A critical look at long-term sun-weather relationships, *Rev. Geophys.*, **16** (3) 400–420, 1978.
- Spence, H., D. Baker, A. Burns, T. Guild, C.-L. Huang, G. Siscoe, and R. Weigel, Center for integrated space weather modeling metrics plan and initial model validation, *J. Atmos. Solar-Terr. Phys.*, **66/15–16**, 1499, 2004.
- Stanislawski, I., and Z. Zbyszynski, Forecasting of the ionospheric quiet and disturbed foF2 values at a single location, *Radio Sci.*, **36** (5), 1065–1071, 2001.
- Stanislawski, I., and Z. Zbyszynski, Forecasting of the ionospheric characteristics during quiet and disturbed conditions, *Ann. Geophys.*, **45** (1), 169–175, 2002.
- Thacker, B.H., S.W. Doebeling, F.M. Hemez, M.C. Anderson, J.E. Pepin, and E.A. Rodriguez, Concepts of model verification and validation, *Technical Report*, Los Alamos National Lab, Los Alamos, NM, USA, DOI: [10.2172/835920](https://doi.org/10.2172/835920), 2004.
- Tsagouri, I., and A. Behlaki, A new empirical model of middle latitude ionospheric response for space weather applications, *Adv. Space Res.*, **37**, 420–425, 2006.
- Tsagouri, I., and A. Behlaki, An upgrade of the solar wind driven empirical model for the middle latitude ionospheric storm time response, *J. Atmos. Solar-Terr. Phys.*, **70**, 2061–2076, DOI: [10.1016/j.jastp.2008.09.010](https://doi.org/10.1016/j.jastp.2008.09.010), 2008.
- Tsagouri, I., K. Koutroumbas, and A. Behlaki, Ionospheric foF2 forecast over Europe based on an autoregressive modeling technique driven by solar wind parameters, *Radio Sci.*, **44**, RS0A35, DOI: [10.1029/2008RS004112](https://doi.org/10.1029/2008RS004112), 2009.
- Tulunay, Y., E. Tulunay, and E.T. Senalp, The neural network technique – 1: A general exposition, *Adv. Space Res.*, **33**, 983–987, 2004a.
- Tulunay, Y., E. Tulunay, and E.T. Senalp, The neural network technique – 2: An ionospheric example illustrating its application, *Adv. Space Res.*, **33**, 988–992, 2004b.
- Vassiliadis, D., Forecasting space weather, *Space Weather – Physics and Effects*, Springer Praxis Books, 403–425, DOI: [10.1007/978-3-540-34578-7_14](https://doi.org/10.1007/978-3-540-34578-7_14), 2007.
- Wintoft, P., and L.R. Cander, Twenty-four hour predictions of foF2 using time delay neural networks, *Radio Sci.*, **35** (2), 395–408, 2000a.
- Wintoft, P., and L.R. Cander, Ionospheric foF2 storm forecasting using neural networks, *Phys. Chem. Earth*, **25** (4) 267–273, 2000b.