Empirical Data Assimilation for Merging Total Electron Content Data with Empirical and Physical Models

Forootan, Ehsan; Kosary, Mona; Farzaneh, Saeed; Schumacher, Maike

Published in:
Surveys in Geophysics

DOI (link to publication from Publisher):
10.1007/s10712-023-09788-7

Publication date:
2023

Document Version
Accepted author manuscript, peer reviewed version

Empirical Data Assimilation for Merging Total Electron Content Data with Empirical and Physical Models

Ehsan Forootan[1], Mona Kosary[2], Saeed Farzaneh[2*] and Maike Schumacher[1]

[1]Geodesy Group, Department of Planning, Aalborg University, Rendburggade 14, 9000, Aalborg, Denmark.
[2*]School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, 113654563, Tehran, Iran.

*Corresponding author(s). E-mail(s): Saeed Farzaneh: farzaneh@ut.ac.ir;
Contributing authors: Ehsan Forootan: efo@plan.aau.dk; Mona Kosary: mona.kosary@ut.ac.ir; Maike Schumacher: maikes@plan.aau.dk;

Abstract

An accurate estimation of ionospheric variables such as the Total Electron Content (TEC) is important for many space weather, communication, and satellite geodetic applications. Empirical and physics-based models are often used to determine TEC in these applications. However, it is known that these models cannot reproduce all ionospheric variability due to various reasons such as their simplified model structure, coarse sampling of their inputs, and dependencies to the calibration period. Bayesian-based Data Assimilation (DA) techniques are often used for improving these model's performance but their computational cost is considerably large. In this study, first, we review the available DA techniques for upper atmosphere data assimilation. Then, we will present an empirical Decomposition-based Data Assimilation (DDA), based on the Principal Component Analysis (PCA) and the Ensemble Kalman Filter (EnKF). DDA considerably reduces the computational complexity of previous DA implementations. Its performance is demonstrated by updating the Empirical Orthogonal Functions (EOFs) of the empirical NeQuick and the physics-based TIEGCM models using the rapid Global

Ionosphere Map (GIM) TEC products as observation. The new models, respectively called 'DDA-NeQuick' and 'DDA-TIEGCM', are then used to predict TEC values for the next day. Comparisons of the TEC forecasts with the final GIM TEC products (that are available after 11 days) represent an average **42.46%** and **31.89%** Root Mean Squared Error (RMSE) reduction during our test period, September 2017.

Keywords: Data Assimilation (DA), Total Electron Content (TEC), Principal Component Analysis (PCA), Ensemble Kalman Filter (EnKF), NeQuick, TIEGCM

Article Highlight

- A new empirical Decomposition-based Data Assimilation (DDA) method is introduced
- DDA is applied to merge the Global Ionospheric Maps (GIMs) with empirical and physics-based models.
- The Empirical Orthogonal Functions (EOFs) of the empirical NeQuick and the physics-based TIEGCM models are updated through the DDA procedures.
- The Total Electron Content (TEC) forecasts after DDA are of the similar quality of the final GIM products.

1 Introduction

A comprehensive knowledge of the Earth's ionosphere and its 4-dimensional dynamics is necessary to support the effective operation, planning, and management of numerous radio communication, navigation, space weather, and surveying applications [1–4]. Satellite geodetic techniques provide a great opportunity to measure the ionosphere-related variables. For example, the dual frequency measurements of the Global Navigation Satellite System (GNSS) can be used to estimate the Total Electron Content [TEC, 5, 6] or electron density [7–9]. The Radio Occultation [RO, e.g., 10] technique makes use of the GNSS measurements of Low-Earth-Orbiting (LEO) satellites to measure the electron number along the ray-path between the GNSS and LEO satellites [11]. Satellite altimetry missions provide the opportunity to measure the two-way range between satellites and water bodies that can be used to estimate Vertical TEC (VTEC) between satellites and surface of the Earth [12]. Though these techniques are extremely helpful for monitoring the ionosphere, their spatial and temporal resolutions are limited by the mission design, e.g., satellite orbits [13, 14] or restricted to the missions' limited life time [15, 16].

Similar to other science communities, many models of the Earth's ionosphere have been developed in last years to simulate and forecast the density of electrons and TEC [17–20]. These models can be divided into four main groups: (1) empirical models that define the ionospheric electron density profiles and their global characteristics, for example, those related to modelling the critical frequencies and peak electron density in different regions such as the E layer from 110-140 km [21–23] that is often described by a simple Chapman theory [24], the F1 layer [that is located between 140 and 210 km and is tightly related to the F2-layer via the neutral composition, 25–27], and the F2 layer [that is above 210 km containing foF2 and NmF2 ionospheric parameters, 28–32]; (2) physical models that work based on the continuity and momentum equations for different ionospheric regions [33–36]; and (3) data assimilation systems that merge sparse real-world observations with model-based (regular) estimations, examples include the IRI Real-Time Assimilative Mapping (IRTAM), Advanced Ensemble electron density (Ne) Assimilation System (AENeAS) and TEC-based ionospheric data assimilation system (TIDAS) [see, e.g., 37–44].

The main idea behind developing the physical and empirical models (in 1 and 2) was to provide the community with tools to predict the 4D structure of ionosphere. Current physical models such as the Thermosphere-Ionosphere-Electrodynamics General Circulation Model [TIEGCM, 36, 45], the Coupled Thermosphere Ionosphere Plasmasphere Electrodynamics [CTIPe, 46–49], and the Global Ionosphere Thermosphere Model [GITM, 50] can numerically resolve differential continuity, momentum and energy equations on $5^\circ \times 5^\circ$ or $2.5^\circ \times 2.5^\circ$, $2^\circ \times 18^\circ$ and $2.5^\circ \times 5^\circ$ spatial resolutions in latitude and longitude, respectively. The quality of the now-casting and forecasting of these models depends on the initial states of the system and the reasonable definition of model parameters [51–54].However, a complete information to define them at specific times is rarely available. Moreover, both model states and observations contain uncertainties that prevent them to achieve the best possible performance [37, 55–60]. For example, ionosphere models generally fail to specify ionospheric weather [61–64], which can be likely due to the absence of accurate representation of thermospheric composition and winds [65], the equatorial and high-latitude electric fields, and the high-latitude particle precipitation [66–68].

Empirical models (in 1) are mostly used in operational applications thanks to their low computational needs (compared to physical models). Among the ionospheric models, NeQuick [69–71] is recommended by the International Telecommunication Union for Slant or Vertical TEC (STEC or VTEC) modeling [72]. In addition, this model is adapted for ionospheric corrections in the single-frequency operation of the European Galileo satellite navigation system [71, 73]. Other empirical models such as Klobuchar [74] is used in the GPS navigation messages. The International Reference Ionosphere [IRI, 75] describes almost all variables and related ionospheric data such as electron temperature, ion temperature and ion composition and, critical frequency, peak height and peak electron density in the F2 layer within the altitude range 50-2000 km, globally [76]. The NeQuick empirical model represents only up to 50–70% of the actual ionospheric activities at mid-latitude locations under typical (quiet) ionospheric conditions [77]. More accurate models are therefore needed for real-time and single-frequency GNSS positioning applications [78–82].

To mitigate existing limitations of empirical and physical models, and to take advantage of the real-world observation data, Data Assimilation (DA or known as data-model fusion) methods are applied in previous studies to spread information from remote sensing or geodetic observations to model variables (that are somehow connected to the observations). Through this implementation, one can interpolate, extrapolate, aggregate, and down-scale geodetic observations. Therefore, DA can be used to organize and merge redundant, conflicting, and conventional observations into a single best estimate [42, 65, 83–94].

Between the existing DA methods, sequential ensemble Kalman filter (EnKF)-based [95] frameworks are widely used in the atmosphere science community. EnKF-DA is formulated based on the Monte Carlo method [overall integration method, 96] to calculate predicted error covariance of the model states without linearizing the model or observation operators. However, considerable computational requirements of EnKF and the filter's convergence after some steps of the DA are among its major drawbacks [97–99]. To speed up the DA process, the reduced order modelling techniques such as Square Root Analysis [SQRA, 100], Singular Evolutive Interpolated Kalman filter [SEIK, 101], and the Ensemble Transform Kalman Filter [ETKF, 102] are used in previous studies [43, 103–110].

DA techniques based on the empirical orthogonal functions are introduced in previous studies [e.g., 111–116] for assimilating geodetic and remote sensing data into weather and atmosphere models. These studies took advantage of statistical decomposition techniques such as the Principal Component Analysis (PCA) or its equivalence Singular Vector Decomposition (SVD) techniques [117] to reduce the high dimensions and computational loads, as well as to improve the efficiency of the DA techniques. Generally speaking, empirical DA techniques modify the dominant statistical modes, derived from atmospheric model outputs, which are explained by sets of two-dimensional Empirical Orthogonal Functions (EOFs). Their associated time series, known as Principal Components (PCs), are then updated sequentially using, e.g., non-linear regression analysis, [for a 4D-Variational DA implementation, see, e.g., 115].

This view is followed in this paper by proposing an alternative Decomposition-based Data Assimilation (DDA) technique that takes advantage of [PCA, 117] for dimension reduction. This step can be replaced with more sophisticated techniques such as applying the Independent Component Analysis (ICA) as in [118–120]. Unlike many of previous studies [e.g., 111, 112, 114], the formulated DDA works based on the ensemble of model outputs and observations, thus, it contains the positive features of the EnKF-based techniques, which means that this new DDA formulation considers the uncertainty of model outputs.

The DDA is tested for merging the physical model of TIEGCM [36] and the empirical model of NeQuick [69], while as observation, the global VTECs from the Global Ionospheric Maps [GIM, 121] were used. Within the DDA, PCA is applied on the ensemble of model outputs and on the ensemble of observations perturbed by their covariance matrices. PCA produces EOFs that are spatially orthogonal base functions and are associated with temporally uncorrelated PCs. The GIM-VTEC observations are then used in an EnKF procedure to improve the spatial base functions (i.e., EOFs of PCA) of the empirical and physical models (i.e., chosen here to be NeQuick and TIEGCM). After performing the DDA, the combined data and models, called 'DDA-NeQuick' and 'DDA-TIEGCM', are used to simulate VTECs globally, and the assimilated EOFs of the previous day are applied for forecasting VTECs of the next day. Since DDA is implemented on the dominant modes of model/data outputs, the computational cost of this approach is relatively less than other global DA approaches [e.g., 122–126]. The DDA provides efficient forecasting skill, which is a feature that was missing in the previous DA studies. The entire month of September 2017 is chosen to perform the validation in terms of VTEC.

The proposed DDA can be applied for forecasting both global and regional VTECs, and thus, estimating ionospheric delays in the GNSS-based Standard Point Positioning (SPP) applications, where atmospheric corrections must be applied to the GNSS-derived pseudo-range measurements using models. Numerical experiments are performed using the global GIM data [127] as observation during quiet and active ionosphere conditions (September 26^{th} and 7^{th}, 2017 with $k_p = 2$ and 8, respectively) to assess the adaptability of the DDA approach. The rapid products of GIM (GIM/UQRG [128]) are used to estimate the assimilated EOFs within the DDA procedure. The updated EOFs are then used to replace those of TIEGCM and NeQuick models and to produce new VTEC maps. The predicted DDA VTEC fields are then compared with the final product of GIM (GIM/CODE [129]), which are produced by IGS with around 11 days delays. This means if the quality of the DDA derived TEC forecasts meets the accuracy of the final GIM products, they can replace them in (near) real-time applications.

This paper is organized as follows: in Section 2, we present the data and models of this study. The methodology related to PCA, DDA, and evaluation metrics is provided in Section 3. The main numerical results of this study are presented in Section 4, and finally, this study is concluded in Section 5.

2 Period of Study, Data, and Models

The DDA scheme is demonstrated during the Day of Year (DOY) 244 to 273 in 2017 (i.e., September 1-30, 2017). The chosen period contains both quiet days (during DOY 262-269) and geomagnetic storm (during DOY 250-251). Figure 1 represents the $F_{10.7}$ solar flux from ftp://ftp.ngdc.noaa.gov, K_p from ftp://ftp.ngdc.noaa.gov/STP/GEOMAGNETIC_DATA/INDICES/KP_AP, and the daily mean Disturbance Storm Time [DST, 130] from http://wdc.kugi.kyoto-u.ac.jp/dst_realtime/ to illustrate the space weather conditions during this month. Considering the solar activity, the $F_{10.7}$ index shows a high peak on September 4^{th}, 2017 with the value of 183 sfu, which this large spike is likely due a flaring event on the Sun and caused unrealistically large

$F_{10.7}$ observation, while K_p and DST indicate 8 and -88 nt on September 8^{th}, 2017.

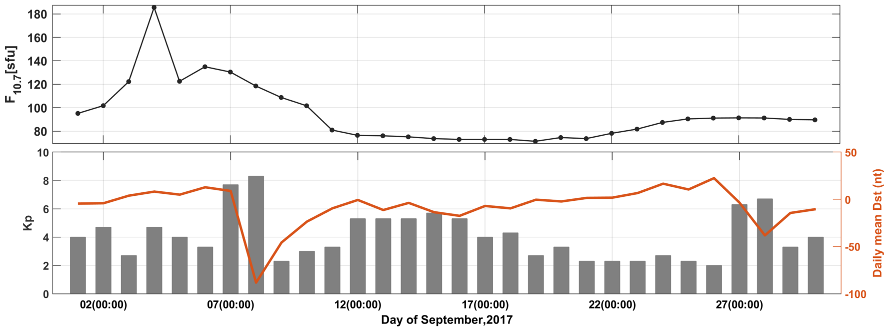

Fig. 1: Space weather conditions during September 2017 demonstrated by the solar ($F_{10.7}$), geomagnetic (K_p), and the Disturbance Storm Time (Dst) indices

2.1 Data

Since 1998, the GNSS dual-frequency code and phase measurements from the globally distributed International GNSS Service (IGS) tracking stations have been used to establish products known as the Global Ionosphere Maps (GIMs) in the IONEX (IONosphere EXchange) format and they are available from ftp://cddis.gsfc.nasa.gov/pub/gps/products/ionex/. GIMs contain global VTECs expanded in terms of the spherical harmonics up to degree and order 15 or in the grid domain with the spatial resolution of $2.5° \times 5°$ in latitude and longitude, respectively. Their temporal resolution is 15 minutes to 2 hours. The GIM products with 2-8 TECU accuracy are available with a latency of less than 24 hours and approximately 11 days in the rapid and final solution modes, respectively [121, 127].

In this study, the rapid global VTEC maps with 15 minutes time interval are obtained from the Technical University of Catalonia, called here 'GIM/UQRG', and these fields are ingested into the NeQuick and TIEGCM models through the DDA procedure. The final VTEC estimates from the CODE products, called here 'GIM/CODE', with 2 h time interval are used for validating the DDA results. The mean of VTEC and their Root Mean Squared (RMS) maps derived from GIM/UQRG are presented in Fig. (2).

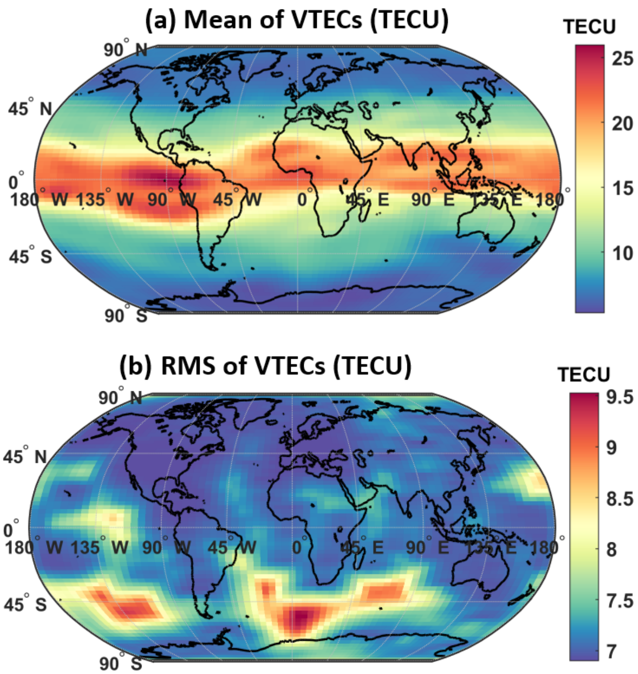

Fig. 2: The gridded mean of VTEC and their RMS derived from the GIM/UQRG during September 2017.

2.2 NeQuick2

Galileo adopts the NeQuick model, recommended by the International Telecommunication Union (ITU), for estimating the ionospheric corrections in single frequency positioning [69, 70]. NeQuick is a three-dimensional and time-dependent ionospheric model [131, 132], which is run by considering daily $F_{10.7}$ index as a proxy of the solar activity. Based on the inputs of position, time and this index, NeQuick evaluates both VTEC and Slant TEC (STEC) values along ground-to-satellite or satellite-to-satellite ray path by integrating the resulting electron density profiles. These measures can also be converted to the range measurement errors in the GNSS positioning experiments [70].

2.3 Thermosphere-Ionosphere-Electrodynamics General Circulation Model (TIEGCM)

The physics-based model TIEGCM is a coupled thermosphere-ionosphere model that uses a finite differential scheme to solve the nonlinear equations of conservation of mass, energy, and momentum for the neutral and ionized species [36, 45]. This study is based on the TIEGCM version 2.0 (released on March 21^{st}, 2016). The horizontal resolution of this model is set to $5° \times 5°$, and the vertical resolution consists of two levels per scale height. The altitude of the model extends from approximately 97 km to 450∼600 km depending on the solar activity [133].

In TIEGCM, the EUVAC (Extreme Ultraviolet Flux model for Aeronomic Calculations) empirical solar proxy model [134, 135] provides the solar irradiance inputs via the daily $F_{10.7}$ and its 81-day averaged ($F_{10.7A}$) time series. This model uses the K_p index[136] instead of the A_p index [137] to indicate the geomagnetic activity. Other forcing parameters in this model include cross-tail potential drop and hemispheric power, which represent the magnitude of auroral particle precipitation and the ionospheric convective electric fields imposed from the magnetosphere, respectively. Throughout this work, the Heelis model is used to specify the high latitudes ion convection [138]. The Global Scale Wave Model (GSWM) provides the lower boundary condition, which is related to the atmospheric tides [139].

To run TIEGCM, primary history files need to be introduced, which include the prognostic fields to start the model. These fields contain variables such as the neutral and ion temperature, neutral zonal and meridional wind, molecular and atomic Oxygen, Nitric Oxide, Helium, Argon, O+ ion, electron temperature and density, O2+ ion, vertical motion, geopotential height and electric potential [36]. Its is worth mentioning that the model runs in this study are performed after the 'spin-up' period of 15 days. The TIEGCM model has an upper boundary level of $\sim 500 - 700$ km altitude, while the VTEC estimates of GIM represent electron variability of up to $\sim 20,200$ km altitude. To reduce this inconsistency, the VTEC estimates above the upper boundary of models are added using the simulation of the NeQuick ionosphere model [43].

3 Method

The details of [PCA, 117], the formulation of DDA, and the evaluation criteria are described in Sections 3.1, 3.2 and 3.3, respectively.

3.1 A review of PCA for dimension reduction within the data assimilation

PCA is a useful statistical (data-driven) approach for dimension reduction, data compression, and noise reduction. Its application has been described in a number of papers with slightly different approaches and notations [e.g., 118, 140, 141]. The dimension of data set is reduced by replacing the original set of the correlated samples with a smaller number of uncorrelated components called Empirical Orthogonal Functions (EOFs) and their associated Principal Components [PCs, see, e.g., 119, chapter 4]. Here, we briefly summarize the PCA approach in the context of ionosphere modelling application.

Our data set, which can be grid maps of TEC or VTEC changes, consists of m time epoch and n grid points, which are arranged into an m by n data matrix (i.e., $\mathbf{O} = [\mathbf{o}_1, \ldots, \mathbf{o}_n]$). The temporal mean of data set is $\bar{\mathbf{o}}_{1,p} = \frac{1}{m}\sum_{i=1}^{m} \mathbf{o}_{i,p}$ where $p = 1, \ldots, n$, which is a raw vector with dimension n, and each element of $\bar{\mathbf{o}}$ is the mean value of all m observations for a given grid point. The deviations of all observations from their mean values are arranged into an $m \times n$ matrix, $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$, where each column of $\mathbf{V}$ stores the deviation of the time series of one grid point with respect to its temporal mean $\bar{\mathbf{o}}$ (i.e., $\mathbf{v}_i = \mathbf{o}_i - \bar{\mathbf{o}}_{1,i}$) and each row of $\mathbf{V}$ contains n observation deviations at each time epoch.

The (auto-)covariance matrix $\mathbf{C}_{m\times m}$ of matrix $\mathbf{V}$ can be written as:

$$\mathbf{C} = \frac{1}{m-1}\mathbf{V}\mathbf{V}^T, \tag{1}$$

where the superscript T denotes the matrix transpose. Through the eigenvalue decomposition [117], the covariance matrix $\mathbf{C}$ can be decomposed as:

$$\mathbf{C} = \mathbf{E}\Lambda\mathbf{E}^T, \tag{2}$$

where Λ is a diagonal matrix that contains the eigenvalues λ_i (the square form of the singular values) of $\mathbf{C}$ are arranged with respect to their magnitude, and $\mathbf{E} = [e_1 \ldots e_m]$ is an orthogonal matrix consisting of corresponding eigenvectors of $\mathbf{C}$ as column vectors, where $\mathbf{E}^T\mathbf{E} = I$ and I is the identity matrix. The matrix $\mathbf{E}$ contains EOFs that are spatially orthogonal vectors. The dimension of $\mathbf{E}$ is $m \times m$ and the each column of $\mathbf{E}$ contains the weights of time epochs. PCs are stored in $\mathbf{P}$ and they are computed by projecting $\mathbf{V}$ onto EOFs ($\mathbf{E}$) as:

$$\mathbf{P} = \mathbf{V}\mathbf{E}. \tag{3}$$

The original data set can be reconstructed from the EOFs and PCs as:

$$\hat{\mathbf{V}} = \mathbf{P}\mathbf{E}^T, \quad \hat{\mathbf{O}} = \mathbf{P}\mathbf{E}^T + \bar{\mathbf{o}}, \tag{4}$$

where $\bar{\mathbf{o}}$ contains the temporal mean field, $\hat{\mathbf{V}}$ contains the mean reduced reconstructed data field, and $\hat{\mathbf{O}}$ represents the reconstructed data with the mean values. The variance explained by the i^{th} PC and EOF is given by the eigenvalue associated with its λ_i. The proportion of variance explained by the i^{th} PC and EOF, or the variance ratio, is given by $\lambda_i / \sum_j \lambda_j$. λ_i decreases with increasing i, indicating that the majority of variance in the data set can be expressed using a smaller number of leading EOFs and PCs. Using only the first n_{pc} components, the data can approximated as:

$$\tilde{V} = \mathbf{P}_{n_{pc}}\mathbf{E}^T_{n_{pc}}, \quad \tilde{O} = \mathbf{P}_{n_{pc}}\mathbf{E}^T_{n_{pc}} + \bar{\mathbf{o}}, \tag{5}$$

where $\mathbf{P}_{n_{pc}}$ is an $m \times n_{pc}$ matrix with the first n_{pc} PCs as its columns and $\mathbf{E}_{n_{pc}}$ is a $n \times n_{pc}$ matrix.

3.2 Direct assimilation of EOFs within the EnKF procedure

The Decomposition-based Data Assimilation (DDA) technique is formulated to integrate VTEC of GIM/UQRG into NeQuick and TIEGCM models. The dimension of this type of GIM-VTECs for one day with the spatial resolution of $2.5° \times 5°$ in latitude and longitude, and temporal resolution of 15 minutes is 96×5183. The simulated VTECs from NeQuick and TIEGCM are determined at the same times and grid points of GIM/UQRG. Based on the PCA technique (Eq. (5)), the VTECs from NeQuick or TIEGCM model are mathematically represented as:

$$\text{Model (NeQuick or TIEGCM), i.e., } : \mathrm{F}(\mathbf{P}, \mathbf{E}, \bar{\mathbf{o}}) = \mathbf{P}\mathbf{E}^T + \bar{\mathbf{o}}, \tag{6}$$

where $\mathbf{P}$ contains the first n_{pc} PCs derived from NeQuick or TIEGCM (using Eq.(3)), $\mathbf{E}$ is based on the the first n_{pc} EOFs of NeQuick or TIEGCM (from Eq.(2)), and $\bar{\mathbf{o}}$ represents the temporal mean of the simulated VTECs from model.

Thus, the VTECs from models during one day are projected onto the 96 time epochs to produce modeled PCs and EOFs. This means that each column of EOFs derived from models has the length of $m = 96$. To reduce the computational load and possible noise, we will assimilate the first n_{pc} (here $n_{pc} = 30$) of EOFs (n_{pc} must be smaller than the rank of the data matrix m). The selection of 30 as n_{pc} corresponds to $\sim 99\%$ of the cumulative variance in global VTEC maps. This number might be changed in other DDA experiences.

To merge models with observations, we propose the use of the Ensemble Kalman Filter [EnKF, 142] with the highest rank of EOFs $(1, 2, 3, \ldots, n_{pc})$ that convey the most available information of the ionosphere state. For this purpose, the ensemble of background model $\mathbf{Y}^B$ and GIM/UQRG VTECs $\mathbf{Y}^{OBS}$ during a day are generated through adding random error. The Gaussian distribution is built using the VTEC estimates from model or IONEX product, which done by a Monte Carlo simulation that considers the i^{th} (i.e., $i = 1, ..., Ne$) ensemble members of the VTECs expressed as:

$$\mathbf{Y}^B = \mathbf{M}^B + \xi_i, \; i = 1, ..., Ne, \tag{7}$$

$$\mathbf{Y}^{OBS} = \mathbf{O}^{\mathrm{OBS}} + \eta_i, \; i = 1, ..., Ne, \tag{8}$$

where Ne is the ensemble member (Ne=90). In Eqs.(7)) and (8)), $\mathbf{M}^B_{96\times5183}$ and $\mathbf{O}^{\mathrm{OBS}}_{96\times5183}$ are VTEC estimates from models (NeQuick or TIEGCM) and GIM/UQRG. The $\xi_{i,96\times5183}$ vector contains random errors with the mean equal to zero and the standard deviation of 10 TECU while $\eta_{i,96\times5183}$ corresponds to the uncertainties of GIM/UQRG VTECs, given by the IONEX products. The standard deviation of GIM/UQRG changes globally and their values are smaller over land (where there is data). It should be mentioned here that the biases of VTECs that exist between the model estimates and observations, called here 'bias$_{\mathrm{VTECs}}$', are considered as unknowns and they will be calibrated throughout the DDA procedure along with the EOFs.

Within the DDA procedure, GIM/UQRG VTECs are used to update the EOFs and bias$_{\mathrm{VTECs}}$ of the model by minimizing the following cost function:

$$J(\mathbf{X}) = \frac{1}{2}[\mathbf{X} - \bar{\mathbf{X}}^B]^T(\mathbf{P}^B)^{-1}[\mathbf{X} - \bar{\mathbf{X}}^B] + \frac{1}{2}[\mathbf{H}\mathbf{X}^B - \mathbf{Y}^{OBS}]^T\mathbf{R}^{-1}(\mathbf{H}\mathbf{X}^B - \mathbf{Y}^{OBS}), \tag{9}$$

where $\mathbf{X}^B$ is the ensembles of background states and is composed of two parts: the ensemble of EOFs from models and the bias values (see Eq. (11)). In Eq. (9), $\bar{\mathbf{X}}^B$ is the ensemble mean vector and $\mathbf{P}^B$ is the background error covariance. Ensembles of observations are stored in Y^{OBS} (Eq. (8)), and $\mathbf{R}$ holds the uncertainty of these observations. The details of these variables are described in the following.

The core of DDA is selected to be the Ensemble Kalman Filter [EnKF as in, 93, 142, 143]. This technique uses the available observations to update the background state (model-derived EOFs and the $\mathrm{bias}_{\mathrm{VTECs}}$) and it decides how to update the states based on their error covariance estimates.

Each ensemble member of EOFs (model state) are generated by applying PCA on the each ensemble of VTECs from NeQuick or TIEGCM. The DDA procedure, which is based on the cost function in Eq. (9), has been evaluated at each grid point to estimate the assimilated EOFs and $\mathrm{bias}_{\mathrm{VTECs}}$ of that grid point. In the following, we stated the DDA technique for one grid point and this procedure is repeated for all grid points (i.e., in this study, the number of grid points that covers the globe with the spatial resolution of $2.5° \times 5°$ in latitude and longitude is 5183). The ensemble of EOFs for one grid point is expressed by $\mathbf{X}^B_{1,n_{pc}\times Ne}$ as:

$$\mathbf{X}^B_1 = [\mathbf{x}^B_{1,1}, \cdots, \mathbf{x}^B_{1,Ne}], \tag{10}$$

where the upper-index 'B' represents the model background and i^{th} ensemble member of $\mathbf{X}^B_1$ (i.e., $\mathbf{x}^B_{1,i}$) is the first n_{ps} of i^{th} EOF maps for the one grid point.

The ensemble of model states $\mathbf{X}^B_1$ and bias $\mathbf{X}^B_2$ for one grid point are integrated and denoted by $\mathbf{X}^B_{n_{pc}+1\times Ne}$ as:

$$\mathbf{X}^B = \begin{bmatrix} \mathbf{X}^B_{1,n_{pc}\times Ne} \\ ------ \\ \mathbf{X}^B_{2,1\times Ne} \end{bmatrix}, \tag{11}$$

where the ensembles of $\mathrm{bias}_{\mathrm{VTECs}}$ $\mathbf{X}^B_2$ are built based on the Gaussian distribution, whose mean value and standard deviation are set to 0 and 5 TECU, respectively. The ensemble mean vector ($\bar{\mathbf{x}}^B_{n_{pc}+1\times 1}$) of Eq. (11) and the error covariance matrix of the background model ($\mathbf{C}^B_{n_{pc}+1\times n_{pc}+1}$) are computed as follows:

$$\bar{\mathbf{x}}^B = \frac{1}{Ne}\sum_{i=1}^{Ne} \mathbf{x}^B_i, \text{ and} \tag{12}$$

$$\mathbf{C}^B = \frac{1}{Ne-1}(\mathbf{X}^B - \bar{\mathbf{x}}^B)(\mathbf{X}^B - \bar{\mathbf{x}}^B)^T. \tag{13}$$

The analysis step (shown by the upper-index a) corrects the model-derived EOFs and $\mathrm{bias_{VTECs}}$ value for one grid point, and predicts the states and their uncertainties using the GIM/UQRG VTEC as follows:

$$\mathbf{X}^{a}_{n_{pc}+1\times Ne} = \mathbf{X}^{B} + \mathbf{K}(\mathbf{Y}^{\mathrm{OBS}} - \mathbf{H}\mathbf{X}^{\mathrm{B}}), \tag{14}$$

and their ensemble mean and their uncertainties, shown by $\bar{\mathbf{x}}^{a}$ and $\mathbf{C}^{a}$, are computed as:

$$\bar{\mathbf{x}}^{a}_{n_{pc}+1\times 1} = \bar{\mathbf{x}}^{B} + \mathbf{K}(\bar{\mathbf{y}}^{\mathrm{OBS}} - \mathbf{H}\bar{\mathbf{x}}^{B}), \tag{15}$$

$$\mathbf{C}^{a}_{n_{pc}+1\times n_{pc}+1} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{C}^{B}, \tag{16}$$

where $\mathrm{Y}^{\mathrm{OBS}}_{m\times Ne}$ and $\bar{\mathbf{y}}^{\mathrm{OBS}}_{m\times 1}$ represent the ensembles (i.e., perturbed by the estimated noise derived from GIM products) and the ensemble mean of GIM/UQRG VTECs for the one grid point, respectively. Considering Eqs. (14-16), the analyzed states and their covariance matrix depend on differences between the real observations ($\mathrm{Y}^{\mathrm{OBS}}$) and model predictions ($\mathbf{H}\mathbf{X}^{\mathrm{B}}$), while considering their weights, which are reflected in the Kalman gain matrix ($\mathbf{K}_{n_{pc}+1\times m}$) that is defined as:

$$\mathbf{K} = \mathbf{C}^{\mathrm{B}}\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\mathbf{C}^{\mathrm{B}}\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right)^{-1}. \tag{17}$$

Here, $\mathbf{R}_{m\times m}$ represents the covariance matrix of observations (i.e., GIM/UQRG VTECs). By assuming the VTECs of one point during a day to be independent, this matrix will be diagonal, where the root of its diagonal elements is derived from the IONEX products. In Eqs. (14, 15, and 17), the design matrix $\mathbf{H}$ is defied as:

$$\mathbf{H}_{m\times n_{pc}} = [\mathbf{P}_{m\times n_{pc}} 1_{m\times 1}], \tag{18}$$

where $\mathbf{P}$ contains the first n_{pc} PCs derived from models (Eq. (2)), and $1_{m\times 1}$ is represented the impact of bias in simulating VTECs from model Eq. (19).

Therefore, by implementing the DDA procedure for all grid points, the ensemble mean and uncertainties of EOFs and $\mathrm{bias_{VTECs}}$ from the analysis step (Eqs. (15 and 16)) provide us with the global updated EOFs ($\hat{\mathbf{E}}$) and new bias ($\hat{bias}_{VTECs}$) estimates along with their uncertainties. The DDA model (in its general form) and the associated uncertainties can be derived from Eqs. (19) and (20), respectively. Thus, for NeQuick or TIEGCM, the model can be derived as Eqs. (21) and (22), respectively.

$$\text{DDA model, i.e., } : \mathrm{F1}(\hat{\mathbf{E}}, \mathbf{P}, \hat{bias}_{VTECs}) = \mathbf{P}\hat{\mathbf{E}}^{T} + \hat{bias}_{VTECs}, \tag{19}$$

$$\text{DDA Error, i.e., } : \mathrm{F2}(\mathbf{C}_{\hat{\mathbf{E}}}, \mathbf{P}, \mathbf{C}_{\hat{bias}_{VTECs}}) = \mathbf{P}\mathbf{C}_{\hat{\mathbf{E}}}\mathbf{P}^{T} + \mathbf{C}_{\hat{bias}_{VTECs}}, \tag{20}$$

$$\text{DDA-NeQuick} = \mathbf{P}_{NeQuick}\hat{\mathbf{E}}^{T}_{NeQuick} + \hat{bias}_{VTECs NeQuick}, \tag{21}$$

$$\text{DDA-TIEGCM} = \mathbf{P}_{TIEGCM}\hat{\mathbf{E}}^{T}_{TIEGCM} + \hat{bias}_{VTECs TIEGCM}, \tag{22}$$

To forecast VTEC for the next day, the original empirical or physics-based model VTECs of the next day can be generated either using the solar and geomagnetic indices of the previous day or by inserting them from available prediction products. The model runs of the next day can be computed, e.g., every 15 minutes. The new VTEC fields are then decomposed using PCA to estimate the EOFs ($\mathbf{E}_{d+1}$) and PCs ($\mathbf{P}_{d+1}$) of the next day. We replace the EOF of the forecasting day with the updated EOFs of the previous day $\hat{\mathbf{E}}_d$. Mathematically, the one-day VTEC forecasts of a general DDA model and their uncertainties can be estimated by Eqs (23) and (24), respectively. Particularly the DDA-NeQuick and DDA-TIE-GCM forecasts can be respectfully expressed as Eqs. (25) and (26).

$$\text{Predictor model, i.e., } : \mathrm{F1}(\hat{\mathbf{E}}_d, \mathbf{P}_{d+1}, \hat{bias}_{VTECs,d}) = \mathbf{P}_{d+1}\hat{\mathbf{E}}_d^T + \hat{bias}_{VTECs,d}, \tag{23}$$

$$\text{Predictor Error, i.e., } : \mathrm{F2}(\mathbf{C}_{\hat{\mathbf{E}},d}, \mathbf{P}_{d+1}, \mathbf{C}_{\hat{bias}_{VTECs},d}) = \mathbf{P}_{d+1}\mathbf{C}_{\hat{\mathbf{E}},d}\mathbf{P}_{d+1}^T + \mathbf{C}_{\hat{bias}_{VTECs},d}, \tag{24}$$

$$\text{Predictor DDA-NeQuick} = \mathbf{P}_{NeQuick,d+1}\hat{\mathbf{E}}_{NeQuick,d}^T + \hat{bias}_{VTECs_{NeQuick,d}}, \tag{25}$$

$$\text{Predictor DDA-TIEGCM} = \mathbf{P}_{TIEGCM,d+1}\hat{\mathbf{E}}_{TIEGCM,d}^T + \hat{bias}_{VTECs_{TIEGCM,d}}, \tag{26}$$

3.3 Evaluating the results

Various evaluation measures are applied to examine the performance of the original and DDA outputs compared to the observations, including 'Bias' (Eq. (27)), 'Relative Error' (RE, Eq. (28)), 'Standard deviation' (STD, Eq. (29)), 'Root Mean Squared of Error (RMSE, Eq. (30)), 'Improvement' (Eq. (31)), 'Average of Absolute Percentage Deviation (AAPD, Eq. (32))', 'Fit' (Eq. (33)), and 'Correlation Coefficients (CCs, Eq. (34))'. The details are provided in the Appendix.

4 Results

The DDA procedure is performed using 90 ensemble members and the first 30 of EOFs are found to represent 99% of the eigenvalues. Eventually, the assimilated EOFs Eq. (15) replace the model-derived EOFs in Eq. (19) for simulating VTECs of the same day (i.e., now-casting). This means that the now-casting of NeQuick is estimated using Eq. (21), and that of TIEGCM from Eq. (22). For forecasting VTECs during the next day the general model reads as Eq. (23), i.e., for forecasting based on the DDA NeQuick, we use Eq. (25), and the DDA TIEGCM forecasts follow Eq. (26).

An overview of the work-flow of this study to apply DDA on NeQuick or TIEGCM and testing its performance for forecasting global VTECs is presented in Fig. (3). In what follows, VTEC estimates from the DDA are assessed in different ways. In Section 4.1, the prediction of EOFs is presented. Then, the VTEC estimates from NeQuick, TIEGCM, DDA-NeQuick, and DDA-TIEGCM are compared with the VTECs derived from GIM/UQRG in the forecasting mode (Section 4.2). This is done to understand how the DDA changed the original modeled values during September 2017. In Section 4.3, the 6-hourly global maps of DDA in the forecasting mode during two days with high and low Kp are compared with those of GIM/UQRG to see whether the new model represents expected spatial-temporal as reflected in the global models. Finally, the time-series of VTECs from DDA are compared with the final IONEX GIM/CODE products over some selected IGS stations in Section 4.4.

4.1 Predicting EOFs in the forecasting mode

PCA is applied on the global VTEC fields (with spatial/temporal resolution $2.5^\circ \times 5^\circ$ in latitude and longitude / 15 minutes). Here, we use GIM/UQRG to derive the DDA-NeQuick Eq. (21) and DDA-TIEGCM Eq. (22) during September 2017. Plots in Fig. (4,a-e) represent the first EOF of VTECs. In addition, plots in Fig. (4,g-h) indicate the magnitude of singular values that correspond to all of the PCA modes. The amount of VTEC variability captured by first EOF is found to be 32.28%, 47.26%, 43.96%, 44.20% and 44.22% of the total variance for NeQuick, TIEGCM (i.e., TIEGCM and for the top level, height from $\sim 500-800$ km to $\sim 20,200$ km, we used NeQuick), DDA-NeQuick, DDA-TIEGCM and GIM/UQRG VTECs, respectively. The numerical results show that after implementing the DDA, the overall spatial correlation coefficient between the EOFs of GIM/UQRG and models are increased from 90.17% with NeQuick to 99.81% with DDA-NeQuick, and from 62.66% with TIEGCM to 99.84% with DDA-TIEGCM.

4.2 Comparison of VTEC predictions with GIM/UQRG

To assess whether the daily DDA improves the performance of empirical (i.e., NeQuick) or physic-based (TIEGCM) models in the forecasting mode, the assimilated EOF maps are used to forecast VTECs for the next day based on Eq. (23). Figure (5,left) presents the improvements in terms of RMSE of VTECs compared to the GIM/UQRG in the forecasting mode. The DDA results are found to agree well with the GIM/UQRG (e.g., the CC of 91% and 93% for DDA-NeQuick and DDA-TIEGCM, respectively). The average improvement is found to be 42.46% (in the range of $14.47-70.45\%$) and 31.89% (in the range of $6.43 - 59.65\%$) for the DDA-NeQuick and DDA-TIEGCM, respectively. In addition, the mean of global uncertainties of VTECs Eq. (24) derived from NeQuick (TIEGCM) in the DDA procedure is decreased from 5.4 (4.47) TECU to 0.08 (0.44) TECU in the forecasting step during September, 2017.

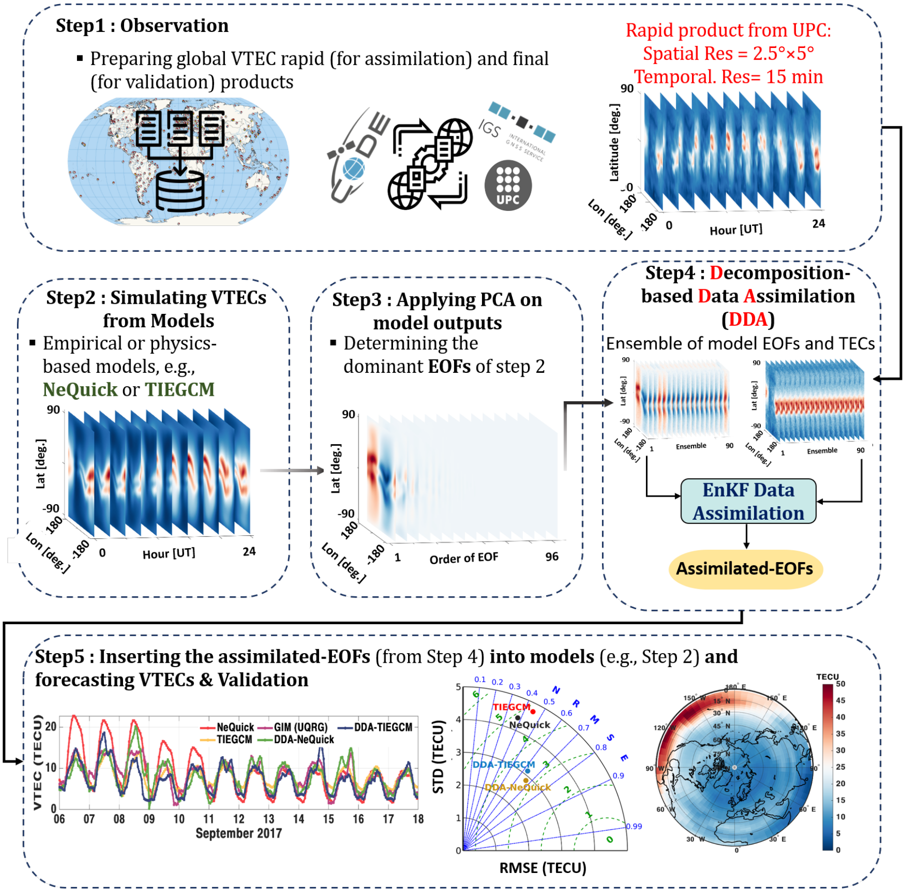

Fig. 3: An overview of the proposed DDA procedure and validation experiments. The procedure is divided into five steps: 1- Generating the ensemble of VTECs from GIM/UQRG, 2- Simulating VTECs and generating ensemble of them from empirical model NeQuick and physics-based model TIEGCM, 3- Applying PCA on each model ensemble and estimating the ensemble of EOFs, 4- Performing DDA for assimilating EOFs and at the same time computing bias_{VTECs}, and 5- Replacing the assimilated EOFs into original models and forecasting VTECs for the next day.

Figure (5,right) shows a Taylor [144, 145] diagram that compares the prediction values with those of GIM/UQRG during September 2017. The results indicate that after implementing the DDA on NeQuick, the RMSE values decreased from 5.33 TECU to 2.87 TECU. Using DDA for TIEGCM, the RMSE values decreased from 4.74 TECU to 3.09 TECU. Based on the statistical values shown in this figure, the DDA-NeQuick is found to provide better statistics, which are closer to the GIM/UQRG, compared to the DDA-TIEGCM model.

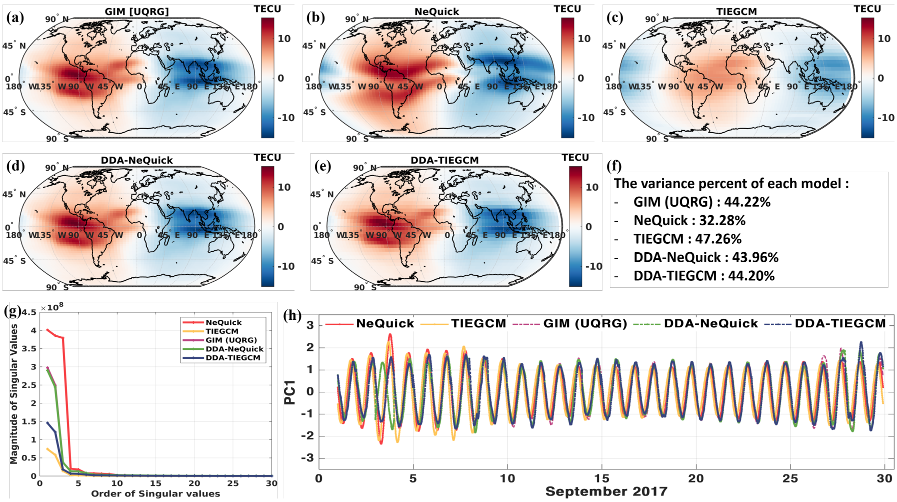

Fig. 4: In (a-e), the first EOF of the VTECs from GIM/UQRG, NeQuick, TIEGCM, DDA-NeQuick and DDA-TIEGCM, respectively is shown, in (f), the variance percentage of the first PCA mode of different models is presented, in (g), the magnitude of the singular values are shown, and in (h), the corresponding PC1 of the plots in a-e are presented. The results correspond to September 2017 using every 15 minutes data with $2.5° \times 5°$ spatial resolution in latitude and longitude, respectively.

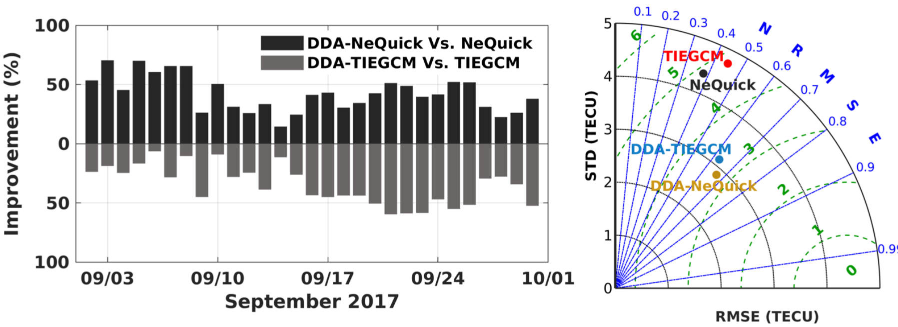

Fig. 5: Left panel: A comparison between evaluation statistics for forecasting VTECs after implementing DDA during September 2017. The improvement are estimated between the NeQuick (TIEGCM) and the DDA-NeQuick(DDA-TIEGCM) models relative to the GIM/UQRG. Right panel: an overview of the three performance measures (RMSE, Standard Deviation (STD), and NRMSE), which are used to assess the performance of the NeQuick, TIEGCM, DDA-NeQuick and DDA-TIEGCM compared to the GIM/UQRG during September 2017.

4.3 Validations in days with high and low K_p

Here, the effect of DDA in forecasting of VTECs derived from NeQuick and TIEGCM is shown during two days (see Figs. (6 and 7)), where from left to right, the 6-hourly snapshots of VTEC differences between (1) NeQuick and GIM/UQRG, (2) DDA-NeQuick and GIM/UQRG, (3) TIEGCM and GIM/UQRG, and (4) DDA-TIEGCM and GIM/UQRG are presented. For two days of 26^{th} (DOY=269) and 8^{th} (DOY=251) in September, 2017 with different level of geomagnetic activity index (i.e., the K_p values of +2 and +8, respectively). Comparing (1) and (2) in Fig. (6) indicates that the VTEC forecasts of DDA-NeQuick agree better with those of IGS (i.e., RMSE of 3.81, 3.76, 3.30, and 3.78 TECU for (1), while 1.85, 1.78, 1.73, and 1.79 TECU were found for (2)). The daily analysis represents a reduction in the range of 51.1% in the forecasting errors for DDA-NeQuick during a day with low geomagnetic activity. In addition, the results in column (3) and (4) of Fig. (6) illustrate that lower RMSEs of 2.45, 2.23, 2.25, and 2.38 TECU were found between the DDA-TIEGCM and GIM/UQRG compared to those of the original TIEGCM, i.e., 4.15, 4.92, 4.64, and 4.89 TECU. An average improvement of 49.86% is obtained for the DDA-TIEGCM on the same day.

Analogous to Fig. (6), in Fig. (7), 6-hourly maps of VTEC differences are presented in the forecasting phase during the day with high K_p. The RMSE between NeQuick and GIM/UQRG are decreased from 8.46, 8.58, 7.88, and 8.18 TECU to 3.06, 2.82, 2.64, and 2.56 TECU for the DDA-NeQuick against GIM/UQRG, and for TIEGCM, it is reduced from 7.35, 6.17, 5.37, and 5.85 TECU to 6.25, 5.63, 4.79, and 4.82 TECU for the DDA-TIEGCM against GIM/UQRG. In summary, the reduction of overall RMSE during September 8^{th}, 2017 is found to be 66.4 and 13.1% for NeQuick and TIEGCM models, respectively. Thus, we conclude that DDA is efficient during days with variable geomagnetic activities.

Figures (6 and 7) indicate that the maximum absolute differences in DDA-NeQuick against GIM/UQRG and DDA-TIEGCM against GIM/UQRG are found around $\pm 30°$ latitude during the two days, which may indicate that NeQuick and TIEGCM do not fully represent the Equatorial Ionosphere Anomaly (EIA) [146] region. It can be seen from the Figs. (6,b and d) and (7, b and d) that DDA decreases errors within the EIA region. The numerical results indicate that the maximum absolute differences of NeQuick and DDA-NeQuick with GIM/UQRG in September 26^{th} (low K_p) are $\sim$ 20 and 15 TECU around 06:00 and 12:00 UT (day time), respectively. These values are estimated to be $\sim$ 30 and 17 TECU for TIEGCM and DDA-TIEGCM. The results for September 8^{th} (high K_p) are found to be $\sim$ 30 and 18 TECU for NeQuick and DDA-NeQuick, while $\sim$ 23 and 19 for TIEGCM and DDA-TIEGCM, respectively.

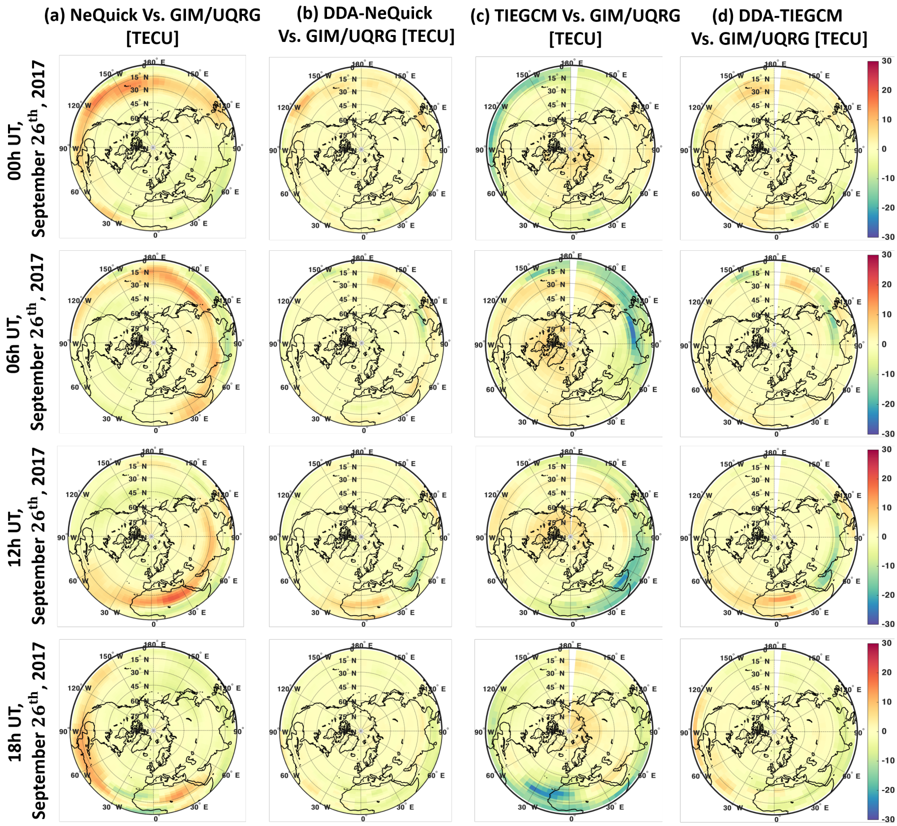

Fig. 6: An overview of the VTEC changes in the forecasting mode on September 26^{th}, 2017 (low geomagnetic activity with $K_p = 2$). The left to right maps : a) NeQuick against GIM/UQRG, b) DDA-NeQuick against GIM/UQRG, c) TIEGCM against GIM/UQRG and d) DDA-TIEGCM against GIM/UQRG, respectively.

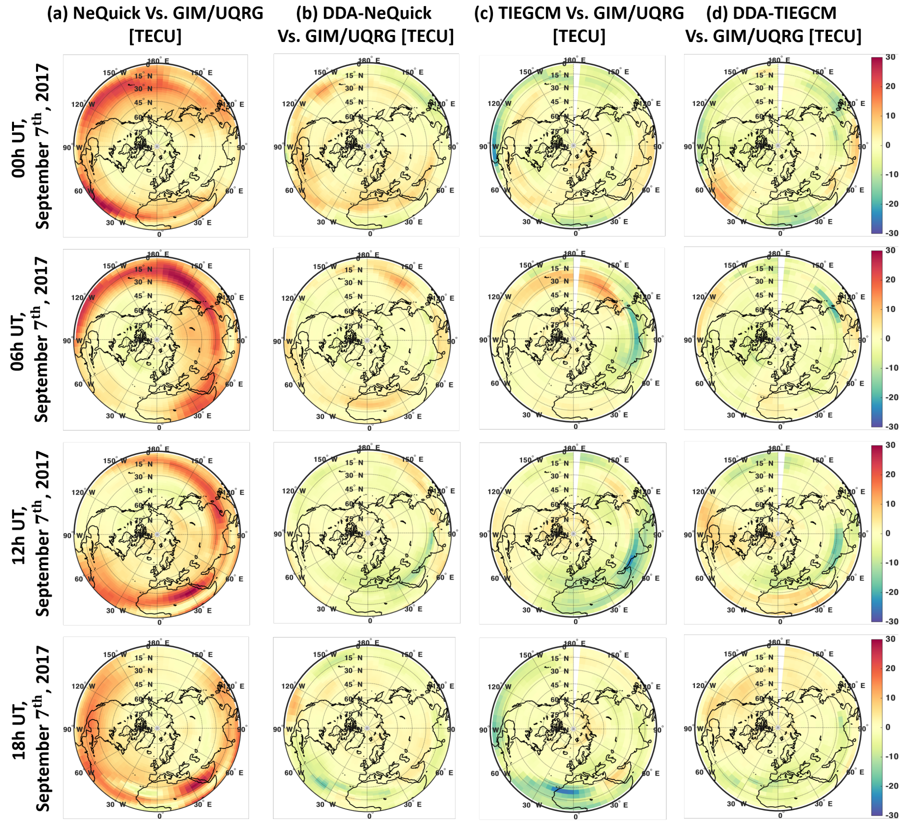

Fig. 7: An overview of VTEC changes in the forecasting mode on September 8^{th}, 2017 (high geomagnetic activity with $K_p = 8$). The left to right maps correspond to the a) NeQuick against GIM/UQRG, b) DDA-NeQuick against GIM/UQRG, c) TIEGCM against GIM/UQRG and d) DDA-TIEGCM against GIM/UQRG, respectively.

4.4 Global validation with the final GIM/CODE VTEC products

Global VTECs of the GIM/CODE products are compared with original and DDA models in Fig. (8). This figure represents the temporal average of bias Eq. (27) and STD Eq. (29)) between models and GIM/CODE in different latitudes. The results indicate that NeQuick overestimates VTECs. TIEGCM underestimates them around the low latitude (from $30°S$ to $40°N$) and overestimate in other latitudes. The maximum absolute biases are found to be 3.67, 6.42, 1.34 and 1.48 TECU for NeQuick, TIEGCM, DDA-NeQuick, and DDA-TIEGCM, respectively. In terms of STD, the models represent similar variations with changing the geographical latitudes. The maximum value of

STD appears in the north and south EIA regions and decreases with increasing latitudes in both northern and southern hemispheres. The maximum STD values are reduced from 7.90 and 3.40 to 1.52 and 2.93 after implementing the DDA approach on NeQuick and TIEGCM, respectively. From these results, we conclude that DDA is efficient in reducing the global errors.

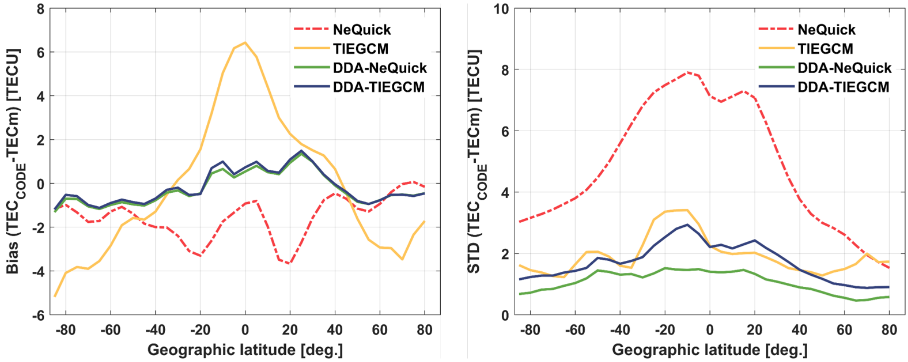

Fig. 8: Bias and STD of the differences between the model derived VTEC estimates and the IGS's GIM/CODE products. The statistics were generated in 2.5° geographic latitude bins for the entire September 2017.

The diurnal VTEC estimates from NeQuick, TIEGCM, DDA-NeQuick, and DDA-TIEGCM are compared with the GIM/CODE ionosphere estimates over some IGS stations. We selected 12 days of September 6^{th}-18^{th}, 2017 to perform the comparisons and the results are shown in Fig. (9). These days are selected because of changes in the geomagnetic index were considerable (see Fig. (1)). After implementing DDA on NeQuick (TIEGCM), the overall RMSE is reduced by 34.3% (30.1%), 57.8% (19.3%), 24.5% (18.9%), 20.8% (47.1%), 51.4% (10.2%) and 21.8% (13.9%) in the six IGS stations (FFMJ - latitude: 50.09° and longitude: 8.66°, Germany; URUM - latitude: 43.80° and longitude: 87.60°, China; SCRZ - latitude: $-17.80°$ and longitude: $-63.16°$, Bolivia; YELL - latitude: 62.48° and longitude: $-114.48°$, Canada; ZAMB - latitude: $-15.43°$ and longitude: 28.31°, Zambia); and NYAL - latitude: 79.83° and longitude: 11.86°, Norway). More statistical evidences of the DDA improvements are provided in Table.1.

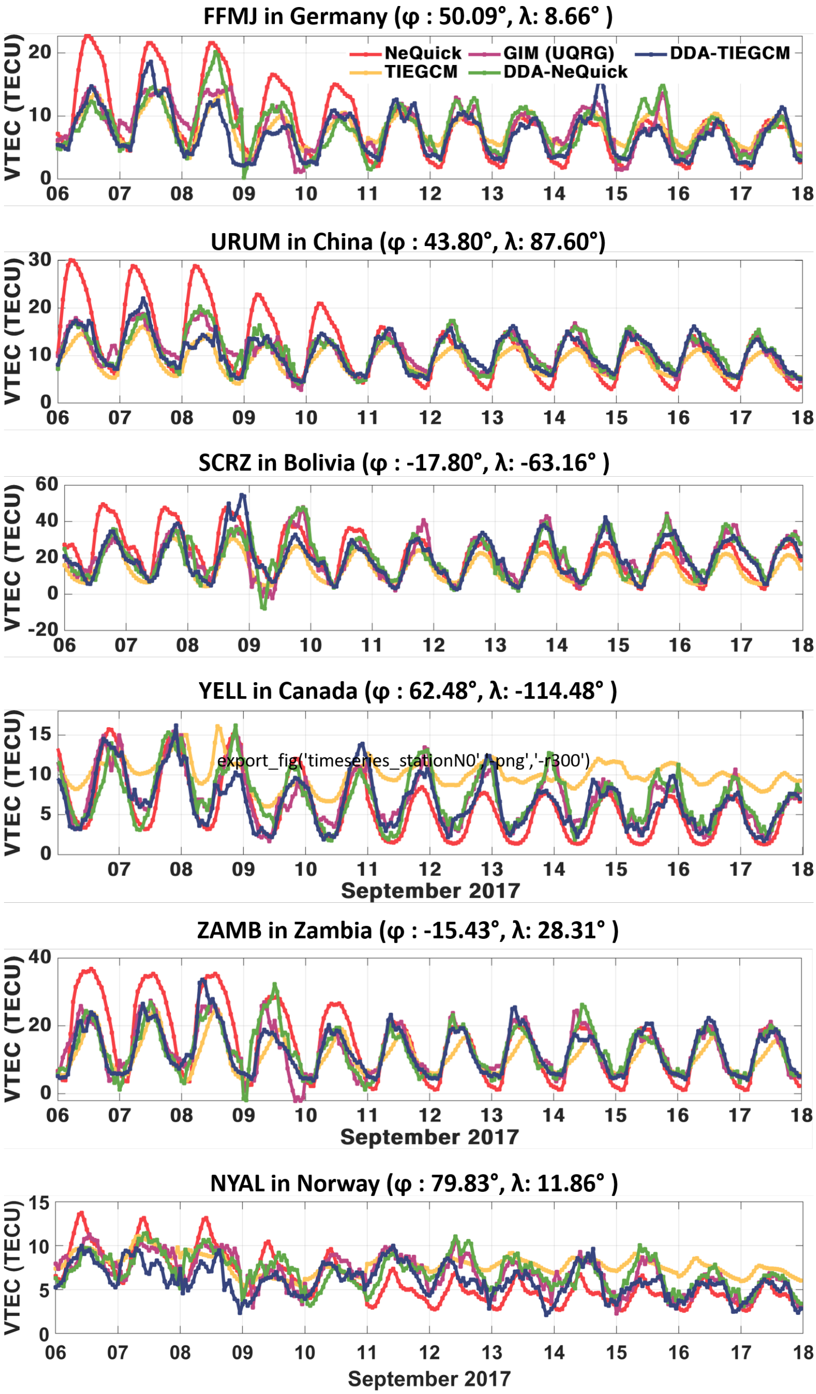

Fig. 9: Diurnal VTEC variations obtained from the NeQuick, TIEGCM, DDA-NeQuick, and DDA-TIEGCM, as well as GIM/CODE over the six selected IGS stations during 12 days in September 2017 (6^{th}-18^{th}).

Table 1: A summary of RMSE, AAPD and NRMSE measures to assess the impact of DDA in forecasting VTECs of 5 IGS stations (in Fig. (9)). These values correspond to the entire September 2017.

(a) The evaluation criteria based on NeQuick during September 2017

Stations (Lat [deg], Long [deg])	RMSE [TECU]		AAPD [%]		Fit	
	NeQuick Vs. GIM/CODE	DDA-NeQuick Vs. GIM/CODE	NeQuick Vs. GIM/CODE	DDA-NeQuick Vs. GIM/CODE	NeQuick Vs. GIM/CODE	DDA-NeQuick Vs. GIM/CODE
FFMJ (50.09 , 8.66)	4.18	2.23	31.72	25.68	-0.23	0.33
URUM (43.80 , 87.60)	5.09	1.91	27.24	13.19	-0.44	0.45
SCRZ (-17.80 , -63.16)	9.62	5.51	34.24	26.63	0.13	0.50
Yell (62.48 , -114.48)	2.25	1.92	27.76	26.64	0.29	0.39
ZAMB (-15.43 , 28.31)	6.96	3.73	39.23	23.01	0.01	0.46
NYAL (78.93, 11.86)	2.06	1.61	27.73	23.23	-0.16	0.09

(b) The evaluation criteria based on TIEGCM during September 2017

Stations (Lat [deg] , Long [deg])	RMSE [TECU]		AAPD [%]		Fit	
	TIEGCM Vs. GIM/CODE	DDA-TIEGCM Vs. GIM/CODE	TIEGCM Vs. GIM/CODE	DDA-TIEGCM Vs. GIM/CODE	TIEGCM Vs. GIM/CODE	DDA-TIEGCM Vs. GIM/CODE
FFMJ (50.09 , 8.66)	2.28	2.06	30.74	24.83	0.32	0.38
URUM (43.80 , 87.60)	2.19	1.67	15.80	11.52	0.37	0.52
SCRZ (-17.80 , -63.16)	8.89	5.47	33.19	26.65	0.20	0.51
Yell (62.48 , -114.48)	4.90	1.83	117.67	29.02	-0.52	0.42
ZAMB (-15.43 , 28.31)	4.08	2.98	20.33	19.25	0.41	0.57
NYAL (78.93, 11.86)	2.21	1.89	39.83	29.41	-0.23	0.05

A comprehensive comparison in terms of global RMSE (Eq. (30)) and Fit (Eq. (33)) are performed with the final VTEC products of GIM/CODE. These hourly measures are summarized in Figs. (10 and (11), which indicates that the main differences can be found as expected during high solar activity (i.e.,from September 7^{th} to 9^{th}, 2017). DDA can improve these differences by 47.2% and 26.6% for NeQuick and TIEGCM, respectively.

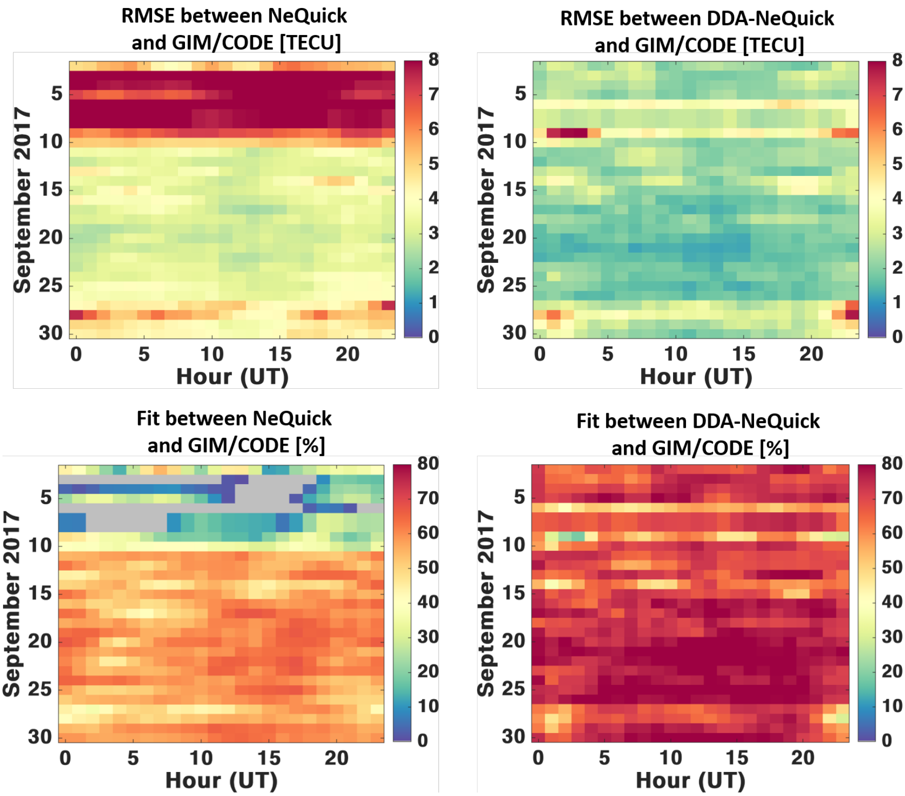

Fig. 10: Hourly global RMSE and Fit before and after performing the DDA. The specific UT hour are shown along the x-axis, each day of September 2017 is represented on the y-axis. The colored values show the RMSE and Fit values, and the gray color in the Fit maps are related to the negative fitting values.

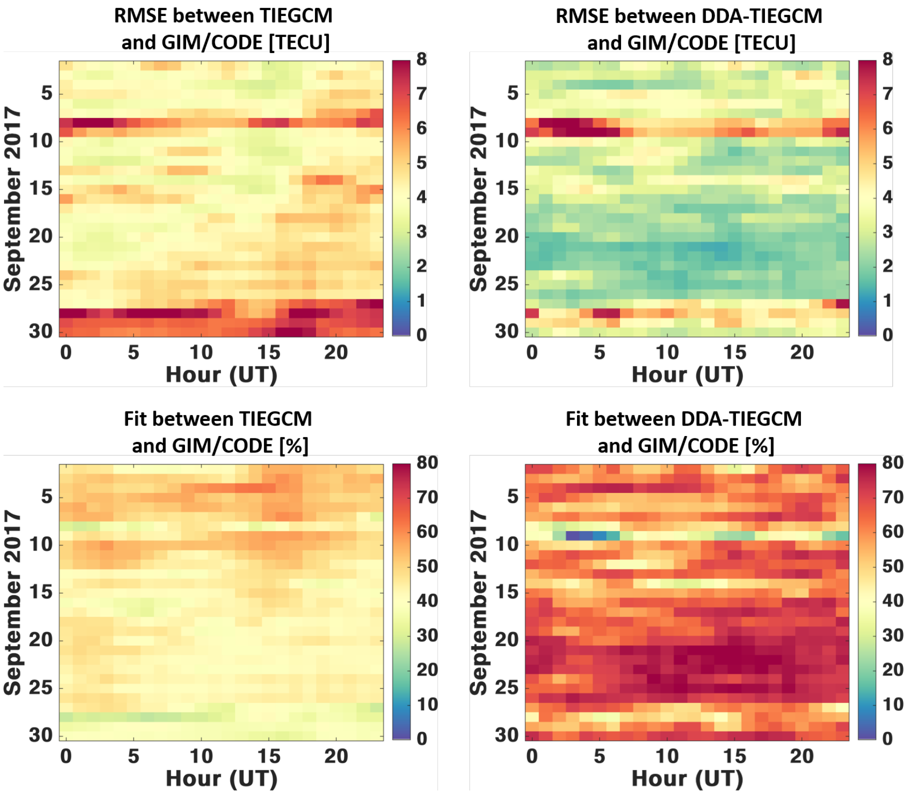

Fig. 11: Global RMSE and Fit calculated by Eqs. (30 and 33) for measurements taken at a specific UT hour (x axis), for a specific day of September 2017 (y axis), before and after applying the DDA approach on TIEGCM.

4.5 Validation with the VTECs derived from GPS measurements

In this section, the NeQuick, TIEGCM, DDA-NeQuick and DDA-TIEGCM are compared to the VTEC derived from GPS measurements for six selected IGS stations as in Fig.9. The VTEC determination based on the GPS measurements follows our previous paper [94].Based on the statistical results, after implementing the DDA, the overall RMSE for the stations during the entire month is reduced by 35.86 and 18.27% using DDA-NeQuick and DDA-TIEGCM compared to the original models, respectively. Also, the average of fitting parameters between models and GPS-VTECs are increased (in terms of NRMSE) from 0.002 and 0.19 to 0.38 and 0.31 for NeQuick and TIEGCM, respectively. The comparison of the models and GPS-VTECs in terms of correlation coefficients and normalized histogram are shown in Fig.12. The NeQuick, TIEGCM, DDA-NeQuick and DDA-TIEGCM are represented in red, yellow, green and blue colors, respectively. The left panel represents the higher correlations between of the DDA results and GPS-VTEC, i.e, 90% and 89%, while these value are about 80% and 79% for the original models. Normalized histograms of the VTEC modeling errors relative to GPS-VTECs are shown in the right panel of Fig. (12). They indicate that the mean of normalized errors of the DDA-NeQuick and DDA-TIEGCM are low, i.e., 0.76 and 0.6 TECU, respectively. The STD of DDA-TIEGCM and DDA-NeQuick model is also lower than that of the original TIEGCM and NeQuick (3.3 vs. 4.6 and 3.5 vs. 6.2 TECU) after implementing the proposed approach.

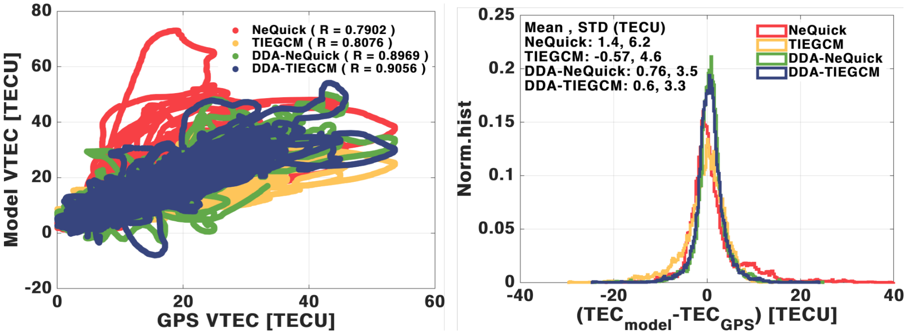

Fig. 12: Left panel: Corresponding scatter-plots of modeled (i.e., NeQuick, TIEGCM, DDA-NeQuick and DDA-TIEGCM) and measured GPS-VTECs values which the Pearson correlation coefficient of each model with observations are shown in the upper right corner. Right panel: Histogram of corresponding residuals between modeled and measured VTECs for six IGS stations during September 2017. The residuals mean and standard deviation are shown in the upper left corner of the histogram.

5 Conclusion

In this study, a Decomposition Data Assimilation (DDA) technique based on the PCA dimension reduction technique, and the EnKF as merger is proposed. DDA can be used to improve the VTEC estimates of available ionosphere models globally using the IGS GIM products. The method can be easily adopted to the regional case studies by changing the domain of the background model and observation fields. The numerical assessments of this study are performed based on the NeQuick and TIEGCM models and the GIM/UQRG as observation. The daily global VTECs obtained from GIM/UQRG are used in the DDA procedure to update the EOFs of models and the new models are shown as ‘DDA-NeQuick’ and ‘DDA-TIEGCM’. The main aim of this work is to show the forecasting skills of DDA for the next 24 hours during quiet and storm conditions during September 2017 was chosen as a test period with the K_p index being considerably changed see Fig. (1). Results are then evaluated against the rapid and final GIM VTEC products. The main findings of this study can be summarized as:

- The DDA is implemented here by considering 90 ensemble members and only the first 30 EOFs with the highest rank of each grid point are used for assimilation. After integrating the VTECs from GIM/UQRG with EOFs from models, the assimilated EOFs are used in the forecasting step. The new assimilated models (DDA-NeQuick and DDA-TIEGCM) provide better VTEC estimates than the original models especially in days (and at those times of the day) with more pronounced ionospheric dynamics, where considerable differences exist between the original models and GIM/CODE VTECs.
- Comparisons between DDA-NeQuick (DDA-TIEGCM) and original models against the VTEC estimates from GIM/UQRG represent the capability of the proposed model in simulating or forecasting VTECs in the EIA region. The differences between the NeQuick (TIEGCM) and DDA-NeQuick (DDA-TIEGCM) compared to the GIM/CODE indicate that the reduction of error around EIA is found to be 50 (30)% approximately.
- Statistical measures indicate that the DDA-NeQuick and DDA-TIEGCM perform better than the original models, compared to the final product GIM/CODE, in both now-casting and forecasting modes. For example, the monthly averages of RMSE, bias and fit parameters in the forecasting step are found to be improved from the original values of 6.62 (5.09) TECU, -1.51 (-0.31) TECU, 0.26 (0.43) to 3.90 (3.63) TECU, -0.30 (-0.22) TECU, 0.56 (0.59) after implementing the DDA procedure into the NeQuick (TIEGCM), respectively.

This work can be extended by performing other decomposition techniques such as Independent Component Analysis [ICA 118, 147]. The DDA can be tested on irregular observations (not-gridded) such as scatter GNSS-derived VTEC estimates from the IGS stations.

6 Acknowledgment

The authors would like to acknowledge the TEC estimates from IGS product https://cddis.nasa.gov/, which were freely available to us. The source code for the simulation models used in this study, the NeQuick and TIE-GCM, are freely available at https://t-ict4d.ictp.it/nequick2 and https://www.hao.ucar.edu/modelling/tgcm/, respectively.

7 Funding

E. Forootan acknowledges the financial support by the Danmarks Frie Forskningsfond [10.46540/2035-00247B].

8 Conflict of interest

The authors declare no conflicts of interest with respect to this work.

References

[1] Verhagen, S., Odijk, D., Teunissen, P., Huisman, L.: Performance improvement with low-cost multi-GNSS receivers. In: Proceedings of the 2010 5th ESA Workshop on Satellite Navigation Technologies and European Workshop on GNSS Signals and Signal Processing (NAVITEC), Noordwijk, The Netherlands, 8–10 December 2010, pp. 1–8 (2010)

[2] Dubey, S., Wahi, R., Gwal, A.: Ionospheric effects on GPS positioning. Advances in Space Research **38**(11), 2478–2484 (2006). https://doi.org/10.1016/j.asr.2005.07.030

[3] Kintner, P.M., Ledvina, B.M.: The ionosphere, radio navigation, and global navigation satellite systems. Advances in Space Research **35**(5), 788–811 (2005). https://doi.org/10.1016/j.asr.2004.12.076. Fundamentals of Space Environment Science

[4] Gu, S., Dai, C., Fang, W., Zheng, F., Wang, Y., Zhang, Q., Lou, Y., Niu, X.: Multi-gnss ppp/ins tightly coupled integration with atmospheric augmentation and its application in urban vehicle navigation. Journal of Geodesy **95**(6), 1–15 (2021). https://doi.org/10.1007/s00190-021-01514-8

[5] Dabbakuti, J.R.K.K., Peesapati, R., Panda, S.K., Thummala, S.: Modeling and analysis of ionospheric tec variability from GPS–TEC measurements using ssa model during 24th solar cycle. Acta Astronautica **178**, 24–35 (2021). https://doi.org/10.1016/j.actaastro.2020.08.034

[6] Ansari, K., Panda, S.K., Jamjareegulgarn, P.: Singular spectrum analysis of gps derived ionospheric tec variations over nepal during the low solar activity period. Acta Astronautica **169**, 216–223 (2020). https://doi.org/10.1016/j.actaastro.2020.01.014

[7] Juan, J.M., Rius, A., Hernández-Pajares, M., Sanz, J.: A two-layer model of the ionosphere using global positioning system data. Geophysical Research Letters **24**(4), 393–396 (1997). https://doi.org/10.1029/97GL00092

[8] Hernández-Pajares, M., Lyu, H., Garcia-Fernandez, M., Orus-Perez, R.: A new way of improving global ionospheric maps by ionospheric tomography: consistent combination of multi-gnss and multi-space geodetic dual-frequency measurements gathered from vessel-, leo-and ground-based receivers. Journal of Geodesy **94**(8), 1–16 (2020). https://doi.org/

10.1007/s00190-020-01397-1

[9] Prol, F.S., Kodikara, T., Hoque, M.M., Borries, C.: Global-scale ionospheric tomography during the 17 March 2015 geomagnetic storm. Space Weather **n/a**(n/a), 2021–002889. https://doi.org/10.1029/2021SW002889

[10] Hajj, G.A., Romans, L.J.: Ionospheric electron density profiles obtained with the global positioning system: Results from the GPS/MET experiment. Radio Science **33**(1), 175–190 (1998). https://doi.org/10.1029/97RS03183

[11] Shume, E.B., Vergados, P., Komjathy, A., Langley, R.B., Durgonics, T.: Electron number density profiles derived from radio occultation on the cassiope spacecraft. Radio Science **52**(9), 1190–1199 (2017). https://doi.org/10.1002/2017RS006321

[12] Seeber, G.: Satellite geodesy: Foundations, methods and applications. INTERNATIONAL HYDROGRAPHIC REVIEW **4**(3), 92–93 (2003)

[13] Montenbruck, O., Gill, E.: Satellite Orbits: Models, Methods and Applications. Springer, ??? (2012)

[14] Sebestyen, G., Fujikawa, S., Galassi, N., Chuchra, A.: Low Earth Orbit Satellite Design vol. 36. Springer, ??? (2018). https://doi.org/10.1007/978-3-319-68315-7

[15] Gonzalo, J., Domínguez, D., López, D.: On the challenge of a century lifespan satellite. Progress in Aerospace Sciences **70**, 28–41 (2014). https://doi.org/10.1016/j.paerosci.2014.05.001

[16] Sneeuw, N., Flury, J., Rummel, R.: Science Requirements On Future Missions And Simulated Mission Scenarios, pp. 113–142. Springer, ??? (2005). https://doi.org/10.1007/0-387-33185-9_10

[17] McCoy, R.P.: Space weather comes of age: new sensors and models for ionospheric specification and forecast. In: Huang, H.-L.A., Bloom, H.J. (eds.) Atmospheric and Environmental Remote Sensing Data Processing and Utilization: an End-to-End System Perspective, vol. 5548, pp. 341–347. SPIE, ??? (2004). International Society for Optics and Photonics. https://doi.org/10.1117/12.562786

[18] Decker, D.T., McNamara, L.F.: Validation of ionospheric weather predicted by global assimilation of ionospheric measurements (gaim) models. Radio Science **42**(4) (2007). https://doi.org/10.1029/2007RS003632

[19] Cander, L.R.: Ionospheric research and space weather services. Journal of Atmospheric and Solar-Terrestrial Physics **70**(15), 1870–1878 (2008). https://doi.org/10.1016/j.jastp.2008.05.010. Ionospheric Effects and Telecommunications

[20] Schunk, R.W., Scherliess, L., Eccles, V., Gardner, L.C., Sojka, J.J., Zhu, L., Pi, X., Mannucci, A.J., Komjathy, A., Wang, C., Rosen, G.: Challenges in specifying and predicting space weather. Space Weather **19**(2), 2019–002404 (2021). https://doi.org/10.1029/2019SW002404

[21] Kouris, S.S., Muggleton, L.M.: Diurnal variation in the e-layer ionization. Journal of Atmospheric and Terrestrial Physics **35**(1), 133–139 (1973). https://doi.org/10.1016/0021-9169(73)90221-3

[22] Chapman, S.: The absorption and dissociative or ionizing effect of monochromatic radiation in an atmosphere on a rotating earth part ii. grazing incidence. Proceedings of the Physical Society **43**(5), 483–501 (1931). https://doi.org/10.1088/0959-5309/43/5/302

[23] Bessarab, F.S., Korenkov, Y.N., Klimenko, V.V., Klimenko, M.V., Zhang, Y.: E-region ionospheric storm on may 1–3, 2010: Gsm tip model representation and suggestions for iri improvement. Advances in Space Research **55**(8), 2124–2130 (2015). https://doi.org/10.1016/j.asr.2014.08.003. INTERNATIONAL REFERENCE IONOSPHERE AND GLOBAL NAVIGATION SATELLITE SYSTEMS

[24] Davies, K.: Ionospheric Radio. Electromagnetic Waves. Institution of Engineering and Technology, ??? (1990). https://doi.org/10.1049/PBEW031E

[25] Bremer, J.: Trends in the ionospheric e and f regions over europe. Annales Geophysicae **16**(8), 986–996 (1998). https://doi.org/10.1007/s00585-998-0986-9

[26] Hochegger, G., Nava, B., Radicella, S., Leitinger, R.: A family of ionospheric models for different uses. Physics and Chemistry of the Earth **25**(4), 307–310 (2000). https://doi.org/10.1016/S1464-1917(00)00022-2

[27] Mikhailov, A.V.: Ionospheric f1 layer long-term trends and the geomagnetic control concept. Annales Geophysicae **26**(12), 3793–3803 (2008). https://doi.org/10.5194/angeo-26-3793-2008

[28] Torr, M.R., Torr, D.G.: The seasonal behaviour of the f2-layer of the ionosphere. Journal of Atmospheric and Terrestrial Physics **35**(12), 2237–2251 (1973). https://doi.org/10.1016/0021-9169(73)90140-2

[29] Rishbeth, H., Müller-Wodarg, I., Zou, L., Fuller-Rowell, T., Millward,

G., Moffett, R., Idenden, D., Aylward, A.: Annual and semiannual variations in the ionospheric f2-layer: Ii. physical discussion. In: Annales Geophysicae, vol. 18, pp. 945–956 (2000). https://doi.org/10.1007/s00585-000-0945-6. Springer

[30] Rishbeth, H., Mendillo, M.: Patterns of f2-layer variability. Journal of Atmospheric and Solar-Terrestrial Physics **63**(15), 1661–1680 (2001). https://doi.org/10.1016/S1364-6826(01)00036-0

[31] Zossi, B.S., Fagre, M., de Haro Barbás, B.F., Elias, A.G.: Ionospheric conductance using different iri f2 layer models. Journal of Atmospheric and Solar-Terrestrial Physics **225**, 105759 (2021). https://doi.org/10.1016/j.jastp.2021.105759

[32] Chen, Y., Liu, L., Le, H., Zhang, H.: Latitudinal dependence of daytime electron density bite-out in the ionospheric f2-layer. Journal of Geophysical Research: Space Physics **126**(1), 2020–028277 (2021). https://doi.org/10.1029/2020JA028277

[33] Sojka, J.J.: Global scale, physical models of the f region ionospere. Reviews of Geophysics **27**(3), 371–403 (1989). https://doi.org/10.1029/RG027i003p00371

[34] Anderson, D.N., Buonsanto, M.J., Codrescu, M., Decker, D., Fesen, C.G., Fuller-Rowell, T.J., Reinisch, B.W., Richards, P.G., Roble, R.G., Schunk, R.W., Sojka, J.J.: Intercomparison of physical models and observations of the ionosphere. Journal of Geophysical Research: Space Physics **103**(A2), 2179–2192 (1998). https://doi.org/10.1029/97JA02872

[35] Man-Lian, Z., She-Ping, S., *et al.*: A physical numerical ionospheric model and its simulation results. Communications in Theoretical Physics **41**(5), 795 (2004). https://doi.org/10.1088/0253-6102/41/5/795

[36] Qian, L., Burns, A.G., Emery, B.A., Foster, B., Lu, G., Maute, A., Richmond, A.D., Roble, R.G., Solomon, S.C., Wang, W.: The NCAR TIE-GCM: A community model of the coupled thermosphere/ionosphere system. Modeling the ionosphere-thermosphere system **201**, 73–83 (2014). https://doi.org/10.1002/9781118704417.ch7

[37] Wang, C., Hajj, G., Pi, X., Rosen, I.G., Wilson, B.: Development of the global assimilative ionospheric model. Radio Science **39**(1) (2004). https://doi.org/10.1029/2002RS002854

[38] Khattatov, B., Murphy, M., Cruikshank, B., Fuller-Rowell, T.: Ionospheric corrections from a prototype operational assimilation and forecast system. In: PLANS 2004. Position Location and Navigation

Symposium (IEEE Cat. No.04CH37556), pp. 518–526 (2004). https://doi.org/10.1109/PLANS.2004.1309037

[39] Angling, M.J., Cannon, P.S.: Assimilation of radio occultation measurements into background ionospheric models. Radio Science **39**(1) (2004). https://doi.org/10.1029/2002RS002819

[40] Scherliess, L., Schunk, R.W., Sojka, J.J., Thompson, D.C., Zhu, L.: Utah state university global assimilation of ionospheric measurements gauss-markov kalman filter model of the ionosphere: Model description and validation. Journal of Geophysical Research: Space Physics **111**(A11) (2006). https://doi.org/10.1029/2006JA011712

[41] Galkin, I.A., Reinisch, B.W., Huang, X., Bilitza, D.: Assimilation of GIRO data into a real-time IRI. Radio Science **47**(04), 1–10 (2012). https://doi.org/10.1029/2011RS004952

[42] Chen, C.-H., Lin, C., Chen, W.-H., Matsuo, T.: Modeling the ionospheric prereversal enhancement by using coupled thermosphere-ionosphere data assimilation. Geophysical Research Letters **44**(4), 1652–1659 (2017). https://doi.org/10.1002/2016GL071812

[43] Elvidge, Sean, Angling, Matthew J.: Using the local ensemble transform Kalman filter for upper atmospheric modelling. J. Space Weather Space Clim. **9**, 30 (2019). https://doi.org/10.1051/swsc/2019018

[44] Aa, E., Zhang, S.-R., Erickson, P.J., Wang, W., Coster, A.J., Rideout, W.: 3-D regional ionosphere imaging and SED reconstruction with a new TEC-Based ionospheric data assimilation system (TIDAS). Space Weather **20**(4), 2022–003055 (2022). https://doi.org/10.1029/2022SW003055

[45] Maute, A.: Thermosphere-ionosphere-electrodynamics general circulation model for the ionospheric connection explorer: TIEGCM-ICON. Space Science Reviews **212**(1), 523–551 (2017). https://doi.org/10.1007/s11214-017-0330-3

[46] Fuller-Rowell, T.J., Rees, D.: A three-dimensional time-dependent global model of the thermosphere. Journal of Atmospheric Sciences **37**(11), 2545–2567 (1980)

[47] Fuller-Rowell, T.J., Rees, D.: Derivation of a conservation equation for mean molecular weight for a two-constituent gas within a three-dimensional, time-dependent model of the thermosphere. Planetary and Space Science **31**(10), 1209–1222 (1983). https://doi.org/10.1016/0032-0633(83)90112-5

[48] Fuller-Rowell, T.J., Rees, D., Quegan, S., Moffett, R.J., Bailey, G.J.: Interactions between neutral thermospheric composition and the polar ionosphere using a coupled ionosphere-thermosphere model. Journal of Geophysical Research: Space Physics **92**(A7), 7744–7748 (1987). https://doi.org/10.1029/JA092iA07p07744

[49] Millward, G.H., Rishbeth, H., Fuller-Rowell, T.J., Aylward, A.D., Quegan, S., Moffett, R.J.: Ionospheric f 2 layer seasonal and semiannual variations. Journal of Geophysical Research: Space Physics **101**(A3), 5149–5156 (1996). https://doi.org/10.1029/95JA03343

[50] Ridley, A.J., Deng, Y., Tóth, G.: The global ionosphere–thermosphere model. Journal of Atmospheric and Solar-Terrestrial Physics **68**(8), 839–864 (2006). https://doi.org/10.1016/j.jastp.2006.01.008

[51] Wang, C., Shi, C., Fan, L., Zhang, H.: Improved modeling of global ionospheric total electron content using prior information. Remote Sensing **10**(1) (2018). https://doi.org/10.3390/rs10010063

[52] Elvidge, S., Godinez, H.C., Angling, M.J.: Improved forecasting of thermospheric densities using multi-model ensembles. Geoscientific Model Development **9**(6), 2279–2292 (2016). https://doi.org/10.5194/gmd-9-2279-2016

[53] Yao, Y., Liu, L., Kong, J., Zhai, C.: Global ionospheric modeling based on multi-gnss, satellite altimetry, and formosat-3/cosmic data. GPS Solutions **22**(4), 1–12 (2018). https://doi.org/10.1007/s10291-018-0770-6

[54] Pedatella, N.M., Anderson, J.L., Chen, C.H., Raeder, K., Liu, J., Liu, H.-L., Lin, C.H.: Assimilation of ionosphere observations in the whole atmosphere community climate model with thermosphere-ionosphere extension (waccmx). Journal of Geophysical Research: Space Physics **125**(9), 2020–028251 (2020). https://doi.org/10.1029/2020JA028251

[55] Schunk, R.W., Scherliess, L., Sojka, J.J.: Recent approaches to modeling ionospheric weather. Advances in Space Research **31**(4), 819–828 (2003). https://doi.org/10.1016/S0273-1177(02)00791-3

[56] Withers, P.: Prediction of uncertainties in atmospheric properties measured by radio occultation experiments. Advances in Space Research **46**(1), 58–73 (2010). https://doi.org/10.1016/j.asr.2010.03.004

[57] McNamara, L.F., Angling, M.J., Elvidge, S., Fridman, S.V., Hausman, M.A., Nickisch, L.J., McKinnell, L.-A.: Assimilation procedures for updating ionospheric profiles below the F2 peak. Radio Science **48**(2), 143–157 (2013). https://doi.org/10.1002/rds.20020

[58] Fu, N., Guo, P., Chen, Y., Wu, M., Huang, Y., Hu, X., Hong, Z.: The analysis of assumptions' error sources on assimilating ground-based/spaceborne ionospheric observations. Journal of Atmospheric and Solar-Terrestrial Physics **207**, 105354 (2020). https://doi.org/10.1016/j.jastp.2020.105354

[59] Nina, A., Nico, G., Mitrović, S.T., Čadež, V.M., Milošević, I.R., Radovanović, M., Popović, L.C.: Quiet ionospheric d-region (qiondr) model based on vlf/lf observations. Remote Sensing **13**(3) (2021). https://doi.org/10.3390/rs13030483

[60] Ren, X., Zhang, J., Chen, J., Zhang, X.: Global ionospheric modeling using multi-gnss and upcoming leo constellations: Two methods and comparison. IEEE Transactions on Geoscience and Remote Sensing, 1–15 (2021). https://doi.org/10.1109/TGRS.2021.3050413

[61] Gulyaeva, T.L., Arikan, F., Hernandez-Pajares, M., Stanislawska, I.: Gim-tec adaptive ionospheric weather assessment and forecast system. Journal of Atmospheric and Solar-Terrestrial Physics **102**, 329–340 (2013). https://doi.org/10.1016/j.jastp.2013.06.011

[62] Rodrigues, F., Wright, I., Moraes, A., Freitas, M.: ScintPi: On the use of low-cost sensors to monitor ionospheric weather and evaluate potential risks. In: 43rd COSPAR Scientific Assembly. Held 28 January - 4 February, vol. 43, p. 673 (2021)

[63] Stanislawska, I., Gulyaeva, T., Arikan, F.: Ionospheric weather risk mitigation challenges in deleterious impacts on ground and space based operational systems and infrastructure. In: 43rd COSPAR Scientific Assembly. Held 28 January - 4 February, vol. 43, p. 655 (2021)

[64] Rao, T.V., Sridhar, M., Ratnam, D.V., Harsha, P.B.S., Srivani, I.: A bidirectional long short-term memory-based ionospheric fof2 and hmf2 models for a single station in the low latitude region. IEEE Geoscience and Remote Sensing Letters, 1–5 (2021). https://doi.org/10.1109/LGRS.2020.3045702

[65] Forootan, E., Kosary, M., Farzaneh, S., Kodikara, T., Vielberg, K., Fernandez-Gomez, I., Borries, C., Schumacher, M.: Forecasting global and multi-level thermospheric neutral density and ionospheric electron content by tuning models against satellite-based accelerometer measurements. Scientific Reports **12 (1)**, 2095 (2022). https://doi.org/10.1038/s41598-022-05952-y

[66] Robinson, R., Zhang, Y., Garcia-Sage, K., Fang, X., Verkhoglyadova, O.P., Ngwira, C., Bingham, S., Kosar, B., Zheng, Y., Kaeppler, S., Liemohn, M., Weygand, J.M., Crowley, G., Merkin, V., McGranaghan,

R., Mannucci, A.J.: Space weather modeling capabilities assessment: Auroral precipitation and high-latitude ionospheric electrodynamics. Space Weather **17**(2), 212–215 (2019). https://doi.org/10.1029/2018SW002127

[67] Miller, K.L., Vondrak, R.R.: A high-latitude phenomenological model of auroral precipitation and ionospheric effects. Radio Science **20**(3), 431–438 (1985). https://doi.org/10.1029/RS020i003p00431

[68] Matsuo, T., Araujo-Pradere, E.A.: Role of thermosphere-ionosphere coupling in a global ionospheric specification. Radio Science **46**(06), 1–7 (2011). https://doi.org/10.1029/2010RS004576

[69] Benyassine, A., Shlomot, E., Su, H.-Y., Massaloux, D., Lamblin, C., Petit, J.-P.: Itu-t recommendation g.729 annex b: a silence compression scheme for use with g.729 optimized for v.70 digital simultaneous voice and data applications. IEEE Communications Magazine **35**(9), 64–73 (1997). https://doi.org/10.1109/35.620527

[70] Nava, B., Coisson, P., Radicella, S.: A new version of the nequick ionosphere electron density model. Journal of Atmospheric and Solar-Terrestrial Physics **70**(15), 1856–1862 (2008). https://doi.org/10.1016/j.jastp.2008.01.015

[71] Aragon-Angel, A., Zürn, M., Rovira-Garcia, A.: Galileo ionospheric correction algorithm: An optimization study of NeQuick-G. Radio Science **54**(11), 1156–1169 (2019). https://doi.org/10.1029/2019RS006875

[72] Series, P.: Ionospheric propagation data and prediction methods required for the design of satellite services and systems. Recommendation ITU-R, 531–613 (2016). https://doi.org/https://www.itu.int/dms_pubrec/itu-r/rec/p/R-REC-P.531-12-201309-S!!PDF-E.pdf

[73] Yuan, Y., Wang, N., Li, Z., Huo, X.: The BeiDou global broadcast ionospheric delay correction model (BDGIM) and its preliminary performance evaluation results. NAVIGATION **66**(1), 55–69 (2019). https://doi.org/10.1002/navi.292

[74] Klobuchar, J.A.: Ionospheric time-delay algorithm for single frequency GPS users. IEEE Transactions on Aerospace and Electronic Systems **AES-23**(3), 325–331 (1987). https://doi.org/10.1109/TAES.1987.310829

[75] Bilitza, D.: International reference ionosphere 2000. Radio Science **36**(2), 261–275 (2001). https://doi.org/10.1029/2000RS002432

[76] Bilitza, D.: Iri the international standard for the ionosphere. Advances in

Radio Science **16**, 1–11 (2018). https://doi.org/10.5194/ars-16-1-2018

[77] Montenbruck, O., Rodríguez, B.G.: Nequick-g performance assessment for space applications. GPS Solutions **24**(1), 1–12 (2020). https://doi.org/10.1007/s10291-019-0931-2

[78] Sanz Subirana, J., Juan Zornoza, J., Hernández-Pajares, M.: GNSS data processing book, vol. i: fundamentals and algorithms. Technical report, TM-23/1. Noordwijk: ESA Communications (2013)

[79] Rose, J.A., Watson, R.J., Allain, D.J., Mitchell, C.N.: Ionospheric corrections for GPS time transfer. Radio Science **49**(3), 196–206 (2014). https://doi.org/10.1002/2013RS005212

[80] Rovira-Garcia, A., Juan, J.M., Sanz, J., González-Casado, G.: A worldwide ionospheric model for fast precise point positioning. IEEE Transactions on Geoscience and Remote Sensing **53**(8), 4596–4604 (2015). https://doi.org/10.1109/TGRS.2015.2402598

[81] Su, K., Jin, S., Hoque, M.: Evaluation of ionospheric delay effects on multi-GNSS positioning performance. Remote Sensing **11**(2), 171 (2019). https://doi.org/10.3390/rs11020171

[82] Zhang, J., Gao, J., Yu, B., Sheng, C., Gan, X.: Research on remote GPS common-view precise time transfer based on different ionosphere disturbances. Sensors **20**(8), 2290 (2020). https://doi.org/10.3390/s20082290

[83] Hajj, G.A., Wilson, B.D., Wang, C., Pi, X., Rosen, I.G.: Data assimilation of ground gps total electron content into a physics-based ionospheric model by use of the kalman filter. Radio Science **39**(1) (2004). https://doi.org/10.1029/2002RS002859

[84] Bust, G.S., Garner, T.W., Gaussiran II, T.L.: Ionospheric data assimilation three-dimensional (ida3d): A global, multisensor, electron density specification algorithm. Journal of Geophysical Research: Space Physics **109**(A11) (2004). https://doi.org/10.1029/2003JA010234

[85] Scherliess, L., Thompson, D.C., Schunk, R.W.: Ionospheric dynamics and drivers obtained from a physics-based data assimilation model. Radio Science **44**(1) (2009). https://doi.org/10.1029/2008RS004068

[86] Chen, P., Chen, J.: The multi-source data fusion global ionospheric modeling software—ionogim. Advances in Space Research **53**(11), 1610–1622 (2014). https://doi.org/10.1016/j.asr.2014.02.025

[87] Chiang, K.Q., Psiaki, M.L.: Gps and ionosonde data fusion for ionospheric tomography. In: Proceedings of the 27th International Technical

Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2014), pp. 1163–1172 (2014)

[88] Matsuo, T.: Upper atmosphere data assimilation with an ensemble Kalman filter. Modeling the Ionosphere–Thermosphere System (eds J. Huba, R. Schunk and G. Khazanov), 273–282 (2014). https://doi.org/10.1002/9781118704417.ch22

[89] Chartier, A.T., Matsuo, T., Anderson, J.L., Collins, N., Hoar, T.J., Lu, G., Mitchell, C.N., Coster, A.J., Paxton, L.J., Bust, G.S.: Ionospheric data assimilation and forecasting during storms. Journal of Geophysical Research: Space Physics **121**(1), 764–778 (2016). https://doi.org/10.1002/2014JA020799

[90] Pilinski, M.D., Crowley, G., Sutton, E., Codrescu, M.: Improved orbit determination and forecasts with an assimilative tool for satellite drag specification. In: Advanced Maui Optical and Space Surveillance Technologies Conference, vol. 104 (2016). https://doi.org/2016amos.confE.104P

[91] Codrescu, S., Codrescu, M., Fedrizzi, M.: An ensemble Kalman filter for the thermosphere-ionosphere. Space Weather **16**(1), 57–68 (2018). https://doi.org/10.1002/2017SW001752

[92] Forsythe, V.V., Azeem, I., Crowley, G.: Ionospheric horizontal correlation distances: Estimation, analysis, and implications for ionospheric data assimilation. Radio Science **55**(12), 2020–007159 (2020). https://doi.org/10.1029/2020RS007159

[93] Forootan, E., Farzaneh, S., Kosary, M., Schmidt, M., Schumacher, M.: A simultaneous calibration and data assimilation (C/DA) to improve NRLMSISE00 using thermospheric neutral density (TND) from space-borne accelerometer measurements. Geophysical Journal International **224**(2), 1096–1115 (2020). https://doi.org/10.1093/gji/ggaa507

[94] Kosary, M., Forootan, E., Farzaneh, S., Schumacher, M.: A sequential calibration approach based on the ensemble kalman filter (c-enkf) for forecasting total electron content (tec). Journal of Geodesy **96**(4), 1–26 (2022). https://doi.org/10.1007/s00190-022-01623-y

[95] Epstein, E.S.: Stochastic dynamic prediction. Tellus **21**(6), 739–759 (1969). https://doi.org/10.3402/tellusa.v21i6.10143

[96] Metropolis, N., S.Ulam: The monte carlo method. Journal of the American Statistical Association **44**(247), 335–341 (1949). https://doi.org/10.1080/01621459.1949.10483310

[97] Luo, X., Bhakta, T., Jakobsen, M., Nævdal, G.: Efficient big data assimilation through sparse representation: A 3d benchmark case study in petroleum engineering. PLOS ONE **13**(7), 1–32 (2018). https://doi.org/10.1371/journal.pone.0198586

[98] Avasarala, S., Subramani, D.: A non-gaussian bayesian filter for sequential data assimilation with non-intrusive polynomial chaos expansion. International Journal for Numerical Methods in Engineering **122**(23), 7156–7181 (2021). https://doi.org/10.1002/nme.6827

[99] Hoang, T.-V., Krumscheid, S., Matthies, H.G., Tempone, R.: Machine learning-based conditional mean filter: A generalization of the ensemble kalman filter for nonlinear data assimilation. Foundations of Data Science **5**(1), 56–80 (2023). https://doi.org/10.3934/fods.2022016

[100] Evensen, G.: Sampling strategies and square root analysis schemes for the enkf. Ocean dynamics **54**(6), 539–560 (2004). https://doi.org/10.1007/s10236-004-0099-2

[101] Tuan Pham, D., Verron, J., Christine Roubaud, M.: A singular evolutive extended kalman filter for data assimilation in oceanography. Journal of Marine Systems **16**(3), 323–340 (1998). https://doi.org/10.1016/S0924-7963(97)00109-7

[102] Bishop, C.H., Etherton, B.J., Majumdar, S.J.: Adaptive sampling with the ensemble transform kalman filter. part i: Theoretical aspects. Monthly weather review **129**(3), 420–436 (2001). https://doi.org/10.1175/1520-0493(2001)129⟨0420:ASWTET⟩2.0.CO;2

[103] Cao, Y., Zhu, J., Navon, I.M., Luo, Z.: A reduced-order approach to four-dimensional variational data assimilation using proper orthogonal decomposition. International Journal for Numerical Methods in Fluids **53**(10), 1571–1583 (2007). https://doi.org/10.1002/fld.1365

[104] Rozier, D., Birol, F., Cosme, E., Brasseur, P., Brankart, J.M., Verron, J.: A reduced-order kalman filter for data assimilation in physical oceanography. SIAM Review **49**(3), 449–465 (2007). https://doi.org/10.1137/050635717

[105] Schumacher, M., Kusche, J., Döll, P.: A systematic impact assessment of grace error correlation on data assimilation in hydrological models. Journal of Geodesy **90**(6), 537–559 (2016). https://doi.org/10.1007/s00190-016-0892-y

[106] Meldi, M., Poux, A.: A reduced order model based on kalman filtering for sequential data assimilation of turbulent flows. Journal of Computational Physics **347**, 207–234 (2017). https://doi.org/10.1016/j.jcp.2017.06.042

[107] Khaki, M., Hoteit, I., Kuhn, M., Awange, J., Forootan, E., van Dijk, A.I.J.M., Schumacher, M., Pattiaratchi, C.: Assessing sequential data assimilation techniques for integrating grace data into a hydrological model. Advances in Water Resources **107**, 301–316 (2017). https://doi.org/10.1016/j.advwatres.2017.07.001

[108] Xiao, D., Du, J., Fang, F., Pain, C.C., Li, J.: Parameterised non-intrusive reduced order methods for ensemble kalman filter data assimilation. Computers and Fluids **177**, 69–77 (2018). https://doi.org/10.1016/j.compfluid.2018.10.006

[109] Zerfas, C., Rebholz, L.G., Schneier, M., Iliescu, T.: Continuous data assimilation reduced order models of fluid flow. Computer Methods in Applied Mechanics and Engineering **357**, 112596 (2019). https://doi.org/10.1016/j.cma.2019.112596

[110] Casas, C.Q., Arcucci, R., Wu, P., Pain, C., Guo, Y.-K.: A reduced order deep data assimilation model. Physica D: Nonlinear Phenomena **412**, 132615 (2020). https://doi.org/10.1016/j.physd.2020.132615

[111] Matsuo, T., Richmond, A.D., Nychka, D.W.: Modes of the High-Latitude Electric Field Variability Derived From DE-2 Measurements: Empirical Orthogonal Function (EOF) Analysis. In: AGU Fall Meeting Abstracts, vol. 2001, pp. 32–0689 (2001)

[112] Matsuo, T., Richmond, A.D., Lu, G.: Optimal interpolation analysis of high-latitude ionospheric electrodynamics using empirical orthogonal functions: Estimation of dominant modes of variability and temporal scales of large-scale electric fields. Journal of Geophysical Research: Space Physics **110**(A6) (2005). https://doi.org/10.1029/2004JA010531

[113] Collard, A.D., McNally, A.P., Hilton, F.I., Healy, S.B., Atkinson, N.C.: The use of principal component analysis for the assimilation of high-resolution infrared sounder observations for numerical weather prediction. Quarterly Journal of the Royal Meteorological Society **136**(653), 2038–2050 (2010). https://doi.org/10.1002/qj.701

[114] Matsuo, T., Fedrizzi, M., Fuller-Rowell, T.J., Codrescu, M.V.: Data assimilation of thermospheric mass density. Space Weather **10**(5) (2012). https://doi.org/10.1029/2012SW000773

[115] Matricardi, M., McNally, A.P.: The direct assimilation of principal components of iasi spectra in the ecmwf 4d-var. Quarterly Journal of the Royal Meteorological Society **140**(679), 573–582 (2014). https://doi.org/10.1002/qj.2156

[116] Lu, Y., Zhang, F.: Toward ensemble assimilation of hyperspectral

satellite observations with data compression and dimension reduction using principal component analysis. Monthly Weather Review **147**(10), 3505–3518 (2019). https://doi.org/10.1175/MWR-D-18-0454.1

[117] Jolliffe, I.: Principal Component Analysis. John Wiley and Sons, Ltd, ??? (2005). https://doi.org/10.1002/0470013192.bsa501. https://onlinelibrary.wiley.com/doi/abs/10.1002/0470013192.bsa501

[118] Forootan, E., Kusche, J.: Separation of global time-variable gravity signals into maximally independent components. Journal of Geodesy **86**(7), 477–497 (2012). https://doi.org/10.1007/s00190-011-0532-5

[119] Forootan, E.: Statistical signal decomposition techniques for analyzing time-variable satellite gravimetry data. PhD thesis, University of Bonn, https://bonndoc.ulb.uni-bonn.de/xmlui/handle/20.500.11811/1452 (2014)

[120] Forootan, E., Kusche, J., Talpe, M., Shum, C., Schmidt, M.: Developing a complex independent component analysis (cica) technique to extract non-stationary patterns from geophysical time series. Surveys in Geophysics **39**, 435–465 (2018). https://doi.org/10.1007/s10712-017-9451-1

[121] Hernández-Pajares, M., Juan, J., Sanz, J., Orus, R., Garcia-Rigo, A., Feltens, J., Komjathy, A., Schaer, S., Krankowski, A.: The IGS VTEC maps: a reliable source of ionospheric information since 1998. Journal of Geodesy **83**(3-4), 263–275 (2009). https://doi.org/10.1007/s00190-008-0266-1

[122] Goss, A., Schmidt, M., Erdogan, E., Seitz, F.: Global and regional high-resolution vtec modelling using a two-step b-spline approach. Remote Sensing **12**(7) (2020). https://doi.org/10.3390/rs12071198

[123] Liu, L., Zou, S., Yao, Y., Wang, Z.: Forecasting global ionospheric tec using deep learning approach. Space Weather **18**(11), 2020–002501 (2020). https://doi.org/10.1029/2020SW002501

[124] Forsythe, V.V., Azeem, I., Crowley, G.: Ionospheric horizontal correlation distances: Estimation, analysis, and implications for ionospheric data assimilation. Radio Science **55**(12), 2020–007159 (2020). https://doi.org/10.1029/2020RS007159

[125] Forsythe, V.V., Azeem, I., Blay, R., Crowley, G., Gasperini, F., Hughes, J., Makarevich, R.A., Wu, W.: Evaluation of the new background covariance model for the ionospheric data assimilation. Radio Science **56**(8), 2021–007286 (2021). https://doi.org/10.1029/2021RS007286

[126] Qiao, J., Liu, Y., Fan, Z., Tang, Q., Li, X., Zhang, F., Song, Y., He, F.,

Zhou, C., Qing, H., Li, Z.: Ionospheric tec data assimilation based on gauss–markov kalman filter. Advances in Space Research **68**(10), 4189–4204 (2021). https://doi.org/10.1016/j.asr.2021.08.004

[127] Feltens, J., Schaer, S.: IGS products for the ionosphere. In: Proceedings of the 1998 IGS Analysis Center Workshop Darmstadt, Germany, pp. 3–5 (1998)

[128] Orús, R., Hernández-Pajares, M., Juan, J.M., Sanz, J.: Improvement of global ionospheric VTEC maps by using kriging interpolation technique. Journal of Atmospheric and Solar-Terrestrial Physics **67**(16), 1598–1609 (2005). https://doi.org/10.1016/j.jastp.2005.07.017

[129] Schaer, S., helvétique des sciences naturelles. Commission géodésique, S.: Mapping and Predicting the Earth's Ionosphere Using the Global Positioning System vol. 59. Institut für Geodäsie und Photogrammetrie, Eidg. Technische Hochschule ..., ??? (1999)

[130] Gonzalez, W.D., Tsurutani, B.T., De Gonzalez, A.L.C.: Interplanetary origin of geomagnetic storms. Space Science Reviews **88**(3), 529–562 (1999). https://doi.org/10.1023/A:1005160129098

[131] Di Giovanni, G., Radicella, S.M.: An analytical model of the electron density profile in the ionosphere. Advances in Space Research **10**(11), 27–30 (1990). https://doi.org/10.1016/0273-1177(90)90301-F

[132] Radicella, S.M., Zhang, M.L.: The improved DGR analytical model of electron density height profile and total electron content in the ionosphere. http://hdl.handle.net/2122/1743

[133] Kodikara, T.: Physical understanding and forecasting of the thermospheric structure and dynamics. PhD thesis, RMIT University (2019)

[134] Richards, P., Fennelly, J., Torr, D.: EUVAC: A solar EUV flux model for aeronomic calculations. Journal of Geophysical Research: Space Physics **99**(A5), 8981–8992 (1994). https://doi.org/10.1029/94JA00518

[135] Solomon, S.C., Qian, L.: Solar extreme-ultraviolet irradiance for general circulation models. Journal of Geophysical Research: Space Physics **110**(A10) (2005). https://doi.org/10.1029/2005JA011160

[136] Webb, D.F., Howard, R.A.: The solar cycle variation of coronal mass ejections and the solar wind mass flux. Journal of Geophysical Research: Space Physics **99**(A3), 4201–4220 (1994). https://doi.org/10.1029/93JA02742

[137] Ahluwalia, H.S.: Ap time variations and interplanetary magnetic field

intensity. Journal of Geophysical Research: Space Physics **105**(A12), 27481–27487 (2000). https://doi.org/10.1029/2000JA900124

[138] Heelis, R., Lowell, J.K., Spiro, R.W.: A model of the high-latitude ionospheric convection pattern. Journal of Geophysical Research: Space Physics **87**(A8), 6339–6345 (1982). https://doi.org/10.1029/JA087iA08p06339

[139] Hagan, M.E., Roble, R.G., Hackney, J.: Migrating thermospheric tides. Journal of Geophysical Research: Space Physics **106**(A7), 12739–12752 (2001). https://doi.org/10.1029/2000JA000344

[140] Kositsky, A.P., Avouac, J.-P.: Inverting geodetic time series with a principal component analysis-based inversion method. Journal of Geophysical Research: Solid Earth **115**(B3) (2010). https://doi.org/10.1029/2009JB006535

[141] Dong, D., Fang, P., Bock, Y., Webb, F., Prawirodirdjo, L., Kedar, S., Jamason, P.: Spatiotemporal filtering using principal component analysis and karhunen-loeve expansion approaches for regional gps network analysis. Journal of Geophysical Research: Solid Earth **111**(B3) (2006). https://doi.org/10.1029/2005JB003806

[142] Evensen, G.: The ensemble Kalman filter for combined state and parameter estimation. IEEE Control Systems Magazine **29**(3), 83–104 (2009). https://doi.org/10.1109/MCS.2009.932223

[143] Schumacher, M.: Methods for assimilating remotely-sensed water storage changes into hydrological models. PhD thesis, Rheinische Friedrich-Wilhelms-Universität Bonn (2016). http://hdl.handle.net/20.500.11811/6630

[144] Taylor, K.E.: Summarizing multiple aspects of model performance in a single diagram. Journal of Geophysical Research: Atmospheres **106**(D7), 7183–7192 (2001). https://doi.org/10.1029/2000JD900719

[145] Elvidge, S., Angling, M.J., Nava, B.: On the use of modified taylor diagrams to compare ionospheric assimilation models. Radio Science **49**(9), 737–745 (2014). https://doi.org/10.1002/2014RS005435

[146] MacDougall, J.W.: The equatorial ionospheric anomaly and the equatorial electrojet. Radio Science **4**(9), 805–810 (1969). https://doi.org/10.1029/RS004i009p00805

[147] Forootan, E., Kusche, J.: Separation of deterministic signals using independent component analysis (ica). Studia Geophysica et Geodaetica **57**, 17–26 (2013). https://doi.org/10.1007/s11200-012-0718-1

Appendix - Evaluation measures

To numerically evaluate the performance of the original and DDA model compared to the observation, the following statistical measures are applied:

- 'Bias' is defined as:

$$\text{Bias} = \frac{1}{n}\sum_{i=1}^{n}(\text{Obs}_i - \text{Model}_i), \tag{27}$$

  where Obs and Model denote observation and model estimates, receptively, and n is the number of observations. The positive (negative) values of the bias demonstrate that the model underestimates (overestimates) compared to the observations.
- The expression of bias in percentage is computed based on the 'Relative Error (RE)' as:

$$\text{RE} = 100 \times \sum_{i=1}^{n}\left(\frac{|\text{Obs}_i - \text{Model}_i|}{\text{Obs}_i}\right), \tag{28}$$

  where $|.|$ represents an operator that returns the absolute values.
- Standard deviation (STD) determines the dispersion of a data-set relative to its mean and is calculated as:

$$\text{STD} = \sqrt{\frac{\sum_{i=1}^{n}(\text{Obs}_i - \bar{\text{Obs}})^2}{n}} \tag{29}$$

- 'Root Mean Squared of Error (RMSE)' is determined to assess how model estimates match with observations as:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\text{Obs}_i - \text{Model}_i)^2}{n}} \tag{30}$$

  The square term inside the RMSE equation highlights both positive and negative differences between the quantities.
- 'Improvement' is defined as percentage in the computed RMSEs after implementing DDA as:

$$\text{Improvement} = 100 \times \frac{\text{RMSE}_1 - \text{RMSE}_2}{\text{RMSE}_1}, \tag{31}$$

  where RMSE_1 is computed between the original NeQuick or TIEGCM and GIM-VTECs, and RMSE_2 is determined between those of DDA and GIM-VTECs.
- 'Average of Absolute Percentage Deviation (AAPD)' is expressed as the percentage of absolute difference between observation and model as:

$$\text{AAPD} = 100 \times \frac{\sum_{i=1}^{n}\left(\left|\frac{\text{Obs}_i - \text{Model}_i}{\text{Obs}_i}\right|\right)}{n}, \tag{32}$$

  Minimum (maximum) values of AAPD correspond to the average best (worst) performance of a model in estimating VTECs.

- 'Fit' is determined as the fraction of data variance that is predicted by the model as:

$$\text{Fit} = 1 - \frac{\sqrt{\sum_{i=1}^{n} (\text{Obs}_i - \text{Model}_i)^2}}{\sqrt{\sum_{i=1}^{n} (\text{Obs}_i - \bar{\text{Obs}})^2}}, \tag{33}$$

  where $\bar{\text{Obs}}$ is defined as the mean of observations. In contrast to AAPD, the minimum (maximum) values of fitting correspond to the average worst (best) performance of model in simulating VTECs.
- 'Correlation Coefficients (CCs)' are used as a unit-less measure to represent the overall agreement between model estimations and observations:

$$\text{CC} = \frac{\sum_{i=1}^{n} (\text{Model}_i - \bar{\text{Model}})(\text{Obs}_i - \bar{\text{Obs}})}{\sqrt{\sum_{i=1}^{n} (\text{Model}_i - \bar{\text{Model}})^2 \sum (\text{Obs}_i - \bar{\text{Obs}})^2}}. \tag{34}$$

  The range of CCs is from –1 to +1, where –1 indicates the perfect negative correlation, +1 corresponds to the 100% correspondence, and zero indicates no correlations.