



AALBORG UNIVERSITY
DENMARK

Aalborg Universitet

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](https://doi.org/10.1007/s10712-023-09788-7)

Publication date:
2023

Document Version
Accepted author manuscript, peer reviewed version

[Link to publication from Aalborg University](#)

Citation for published version (APA):
Forootan, E., Kosary, M., Farzaneh, S., & Schumacher, M. (2023). Empirical Data Assimilation for Merging Total Electron Content Data with Empirical and Physical Models. *Surveys in Geophysics*, 44(6), 2011-2041. <https://doi.org/10.1007/s10712-023-09788-7>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

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

Ehsan Forootan¹, Mona Kosary², Saeed Farzaneh^{2*}
and Maike Schumacher¹

¹Geodesy 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](mailto:Saeed.Farzaneh@ut.ac.ir);

Contributing authors: [Ehsan Forootan: efo@plan.aau.dk](mailto:Ehsan.Forootan@plan.aau.dk); [Mona
Kosary: mona.kosary@ut.ac.ir](mailto:Mona.Kosary@ut.ac.ir); [Maike Schumacher:
maikes@plan.aau.dk](mailto: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

73 their global characteristics, for example, those related to modelling the critical
74 frequencies and peak electron density in different regions such as the E layer
75 from 110–140 km [21–23] that is often described by a simple Chapman theory
76 [24], the F1 layer [that is located between 140 and 210 km and is tightly related
77 to the F2-layer via the neutral composition, 25–27], and the F2 layer [that is
78 above 210 km containing foF2 and NmF2 ionospheric parameters, 28–32]; (2)
79 physical models that work based on the continuity and momentum equations
80 for different ionospheric regions [33–36]; and (3) data assimilation systems
81 that merge sparse real-world observations with model-based (regular) estima-
82 tions, examples include the IRI Real-Time Assimilative Mapping (IRTAM),
83 Advanced Ensemble electron density (Ne) Assimilation System (AENeAS) and
84 TEC-based ionospheric data assimilation system (TIDAS) [see, e.g., 37–44].

85 The main idea behind developing the physical and empirical models (in 1
86 and 2) was to provide the community with tools to predict the 4D structure
87 of ionosphere. Current physical models such as the Thermosphere-Ionosphere-
88 Electrodynamics General Circulation Model [TIEGCM, 36, 45], the Coupled
89 Thermosphere Ionosphere Plasmasphere Electrodynamics [CTIpe, 46–49], and
90 the Global Ionosphere Thermosphere Model [GITM, 50] can numerically
91 resolve differential continuity, momentum and energy equations on $5^\circ \times 5^\circ$ or
92 $2.5^\circ \times 2.5^\circ$, $2^\circ \times 18^\circ$ and $2.5^\circ \times 5^\circ$ spatial resolutions in latitude and longitude,
93 respectively. The quality of the now-casting and forecasting of these models
94 depends on the initial states of the system and the reasonable definition of
95 model parameters [51–54]. However, a complete information to define them at
96 specific times is rarely available. Moreover, both model states and observations
97 contain uncertainties that prevent them to achieve the best possible perfor-
98 mance [37, 55–60]. For example, ionosphere models generally fail to specify
99 ionospheric weather [61–64], which can be likely due to the absence of accurate
100 representation of thermospheric composition and winds [65], the equatorial
101 and high-latitude electric fields, and the high-latitude particle precipitation
102 [66–68].

103 Empirical models (in 1) are mostly used in operational applications thanks
104 to their low computational needs (compared to physical models). Among the
105 ionospheric models, NeQuick [69–71] is recommended by the International
106 Telecommunication Union for Slant or Vertical TEC (STEC or VTEC) mod-
107 eling [72]. In addition, this model is adapted for ionospheric corrections in the
108 single-frequency operation of the European Galileo satellite navigation system
109 [71, 73]. Other empirical models such as Klobuchar [74] is used in the GPS nav-
110 igation messages. The International Reference Ionosphere [IRI, 75] describes
111 almost all variables and related ionospheric data such as electron temperature,
112 ion temperature and ion composition and, critical frequency, peak height and
113 peak electron density in the F2 layer within the altitude range 50–2000 km,
114 globally [76]. The NeQuick empirical model represents only up to 50–70% of
115 the actual ionospheric activities at mid-latitude locations under typical (quiet)
116 ionospheric conditions [77]. More accurate models are therefore needed for
117 real-time and single-frequency GNSS positioning applications [78–82].

118 To mitigate existing limitations of empirical and physical models, and to
119 take advantage of the real-world observation data, Data Assimilation (DA
120 or known as data-model fusion) methods are applied in previous studies to
121 spread information from remote sensing or geodetic observations to model
122 variables (that are somehow connected to the observations). Through this
123 implementation, one can interpolate, extrapolate, aggregate, and down-scale
124 geodetic observations. Therefore, DA can be used to organize and merge redun-
125 dant, conflicting, and conventional observations into a single best estimate
126 [42, 65, 83–94].

127 Between the existing DA methods, sequential ensemble Kalman filter
128 (EnKF)-based [95] frameworks are widely used in the atmosphere science com-
129 munity. EnKF-DA is formulated based on the Monte Carlo method [overall
130 integration method, 96] to calculate predicted error covariance of the model
131 states without linearizing the model or observation operators. However, consid-
132 erable computational requirements of EnKF and the filter’s convergence after
133 some steps of the DA are among its major drawbacks [97–99]. To speed up
134 the DA process, the reduced order modelling techniques such as Square Root
135 Analysis [SQRA, 100], Singular Evolutive Interpolated Kalman filter [SEIK,
136 101], and the Ensemble Transform Kalman Filter [ETKF, 102] are used in
137 previous studies [43, 103–110].

138 DA techniques based on the empirical orthogonal functions are introduced
139 in previous studies [e.g., 111–116] for assimilating geodetic and remote sensing
140 data into weather and atmosphere models. These studies took advantage of
141 statistical decomposition techniques such as the Principal Component Analy-
142 sis (PCA) or its equivalence Singular Vector Decomposition (SVD) techniques
143 [117] to reduce the high dimensions and computational loads, as well as to
144 improve the efficiency of the DA techniques. Generally speaking, empirical DA
145 techniques modify the dominant statistical modes, derived from atmospheric
146 model outputs, which are explained by sets of two-dimensional Empirical
147 Orthogonal Functions (EOFs). Their associated time series, known as Princi-
148 pal Components (PCs), are then updated sequentially using, e.g., non-linear
149 regression analysis, [for a 4D-Variational DA implementation, see, e.g., 115].

150 This view is followed in this paper by proposing an alternative
151 Decomposition-based Data Assimilation (DDA) technique that takes advan-
152 tage of [PCA, 117] for dimension reduction. This step can be replaced with
153 more sophisticated techniques such as applying the Independent Component
154 Analysis (ICA) as in [118–120]. Unlike many of previous studies [e.g.,
155 111, 112, 114], the formulated DDA works based on the ensemble of model
156 outputs and observations, thus, it contains the positive features of the EnKF-
157 based techniques, which means that this new DDA formulation considers the
158 uncertainty of model outputs.

159 The DDA is tested for merging the physical model of TIEGCM [36] and the
160 empirical model of NeQuick [69], while as observation, the global VTECs from
161 the Global Ionospheric Maps [GIM, 121] were used. Within the DDA, PCA is
162 applied on the ensemble of model outputs and on the ensemble of observations

163 perturbed by their covariance matrices. PCA produces EOFs that are spatially
164 orthogonal base functions and are associated with temporally uncorrelated
165 PCs. The GIM-VTEC observations are then used in an EnKF procedure to
166 improve the spatial base functions (i.e., EOFs of PCA) of the empirical and
167 physical models (i.e., chosen here to be NeQuick and TIEGCM). After per-
168 forming the DDA, the combined data and models, called ‘DDA-NeQuick’ and
169 ‘DDA-TIEGCM’, are used to simulate VTECs globally, and the assimilated
170 EOFs of the previous day are applied for forecasting VTECs of the next day.
171 Since DDA is implemented on the dominant modes of model/data outputs,
172 the computational cost of this approach is relatively less than other global DA
173 approaches [e.g., 122–126]. The DDA provides efficient forecasting skill, which
174 is a feature that was missing in the previous DA studies. The entire month of
175 September 2017 is chosen to perform the validation in terms of VTEC.

176 The proposed DDA can be applied for forecasting both global and regional
177 VTECs, and thus, estimating ionospheric delays in the GNSS-based Standard
178 Point Positioning (SPP) applications, where atmospheric corrections must
179 be applied to the GNSS-derived pseudo-range measurements using models.
180 Numerical experiments are performed using the global GIM data [127] as obser-
181 vation during quiet and active ionosphere conditions (September 26th and 7th,
182 2017 with $k_p = 2$ and 8, respectively) to assess the adaptability of the DDA
183 approach. The rapid products of GIM (GIM/UQRG [128]) are used to esti-
184 mate the assimilated EOFs within the DDA procedure. The updated EOFs
185 are then used to replace those of TIEGCM and NeQuick models and to pro-
186 duce new VTEC maps. The predicted DDA VTEC fields are then compared
187 with the final product of GIM (GIM/CODE [129]), which are produced by
188 IGS with around 11 days delays. This means if the quality of the DDA derived
189 TEC forecasts meets the accuracy of the final GIM products, they can replace
190 them in (near) real-time applications.

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

195 2 Period of Study, Data, and Models

196 The DDA scheme is demonstrated during the Day of Year (DOY) 244 to
197 273 in 2017 (i.e., September 1-30, 2017). The chosen period contains both
198 quiet days (during DOY 262-269) and geomagnetic storm (during DOY 250-
199 251). Figure 1 represents the $F_{10.7}$ solar flux from <ftp://ftp.ngdc.noaa.gov>,
200 K_p from [ftp://ftp.ngdc.noaa.gov/STP/GEOMAGNETIC_DATA/INDICES/](ftp://ftp.ngdc.noaa.gov/STP/GEOMAGNETIC_DATA/INDICES/KP_AP)
201 http://wdc.kugi.kyoto-u.ac.jp/dst_realtime/ to illustrate the space weather condi-
202 tions during this month. Considering the solar activity, the $F_{10.7}$ index shows
203 a high peak on September 4th, 2017 with the value of 183 *sfu*, which this large
204 spike is likely due a flaring event on the Sun and caused unrealistically large
205

6 *Empirical Data Assimilation for Ionosphere*

206 $F_{10.7}$ observation, while K_p and DST indicate 8 and -88 nt on September
 207 8th, 2017.

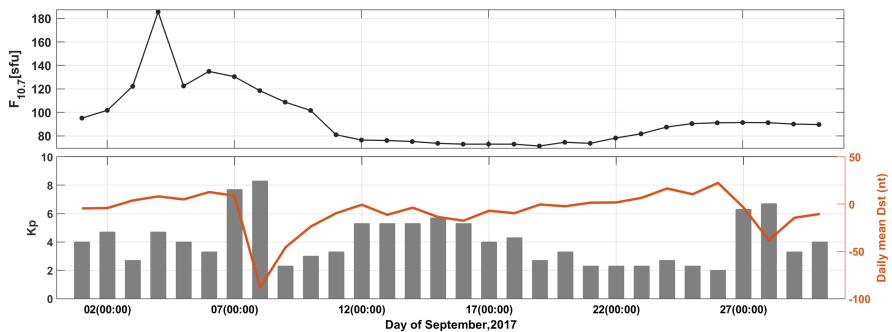


Fig. 1: Space weather conditions during September 2017 demonstrated by the solar ($F_{10.7}$), geomagnetic (K_p), and the Disturbance Storm Time (Dst) indices

2.1 Data

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

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

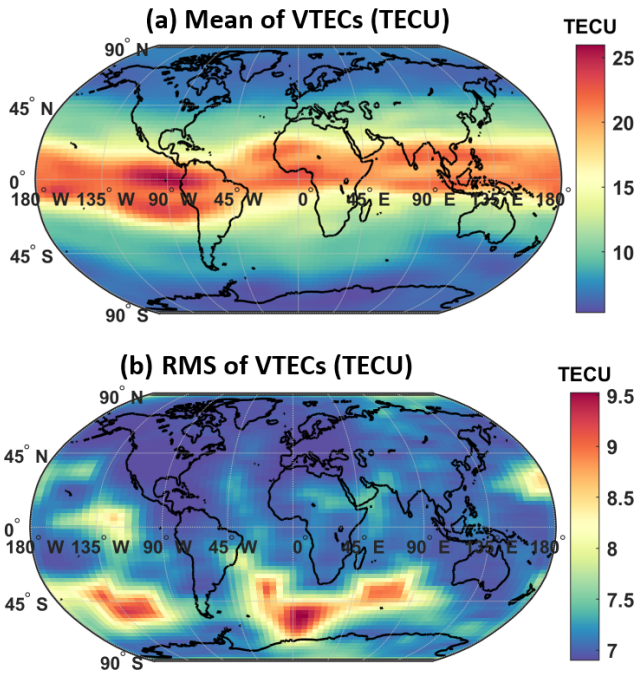


Fig. 2: The gridded mean of VTEC and their RMS derived from the GIM/UQRG during September 2017.

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 21st, 2016). The horizontal resolution of this model is set to $5^\circ \times 5^\circ$, 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 Aeronomical 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]. It 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 ~ 500 – 700 km altitude, while the VTEC estimates of GIM represent electron variability of up to ~ 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., $\mathbf{O} = [\mathbf{o}_1, \dots, \mathbf{o}_n]$). The temporal mean of data set is $\bar{\mathbf{o}}_{1,p} = \frac{1}{m} \sum_{i=1}^m \mathbf{o}_{i,p}$ where $p = 1, \dots, n$, 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, $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_n]$, 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:

$$\mathbf{C} = \frac{1}{m-1} \mathbf{V} \mathbf{V}^T, \quad (1)$$

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

$$\mathbf{C} = \mathbf{E} \mathbf{A} \mathbf{E}^T, \quad (2)$$

where \mathbf{A} 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 $\mathbf{E} = [e_1 \dots e_m]$ is an orthogonal matrix consisting of corresponding eigenvectors of \mathbf{C} as column vectors, where $\mathbf{E}^T \mathbf{E} = \mathbf{I}$ and \mathbf{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:

$$\mathbf{P} = \mathbf{V} \mathbf{E}. \quad (3)$$

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

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

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 be approximated as:

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

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 $2.5^\circ \times 5^\circ$ in latitude and longitude, and temporal resolution of 15 minutes is 96×5183 . The simulated VTECs from NeQuick and TIEGCM are determined

319 at the same times and grid points of GIM/UQRG. Based on the PCA technique
 320 (Eq. (5)), the VTECs from NeQuick or TIEGCM model are mathematically
 321 represented as:

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

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

326 Thus, the VTECs from models during one day are projected onto the 96
 327 time epochs to produce modeled PCs and EOFs. This means that each col-
 328 umn of EOFs derived from models has the length of $m = 96$. To reduce the
 329 computational load and possible noise, we will assimilate the first n_{pc} (here
 330 $n_{pc} = 30$) of EOFs (n_{pc} must be smaller than the rank of the data matrix m).
 331 The selection of 30 as n_{pc} corresponds to $\sim 99\%$ of the cumulative variance in
 332 global VTEC maps. This number might be changed in other DDA experiences.

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

$$\mathbf{Y}^B = \mathbf{M}^B + \xi_i, \quad i = 1, \dots, Ne, \quad (7)$$

$$\mathbf{Y}^{OBS} = \mathbf{O}^{OBS} + \eta_i, \quad i = 1, \dots, Ne, \quad (8)$$

341 where Ne is the ensemble member ($Ne=90$). In Eqs.(7) and (8)), $\mathbf{M}_{96 \times 5183}^B$
 342 and $\mathbf{O}_{96 \times 5183}^{OBS}$ are VTEC estimates from models (NeQuick or TIEGCM) and
 343 GIM/UQRG. The $\xi_{i,96 \times 5183}$ vector contains random errors with the mean
 344 equal to zero and the standard deviation of 10 TECU while $\eta_{i,96 \times 5183}$ cor-
 345 responds to the uncertainties of GIM/UQRG VTECs, given by the IONEX
 346 products. The standard deviation of GIM/UQRG changes globally and their
 347 values are smaller over land (where there is data). It should be mentioned here
 348 that the biases of VTECs that exist between the model estimates and obser-
 349 vations, called here 'bias_{VTECs}', are considered as unknowns and they will be
 350 calibrated throughout the DDA procedure along with the EOFs.

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

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

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 \mathbf{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 bias_{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 bias_{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 $2.5^\circ \times 5^\circ$ in latitude and longitude is 5183). The ensemble of EOFs for one grid point is expressed by $\mathbf{X}_{1,n_{pc} \times Ne}^B$ as:

$$\mathbf{X}_1^B = [\mathbf{x}_{1,1}^B, \dots, \mathbf{x}_{1,Ne}^B], \quad (10)$$

where the upper-index ‘ B ’ represents the model background and i^{th} ensemble member of \mathbf{X}_1^B (i.e., $\mathbf{x}_{1,i}^B$) is the first n_{ps} of i^{th} EOF maps for the one grid point.

The ensemble of model states \mathbf{X}_1^B and bias \mathbf{X}_2^B for one grid point are integrated and denoted by $\mathbf{X}_{n_{pc}+1 \times Ne}^B$ as:

$$\mathbf{X}^B = \begin{bmatrix} \mathbf{X}_{1,n_{pc} \times Ne}^B \\ \mathbf{X}_{2,1 \times Ne}^B \end{bmatrix}, \quad (11)$$

where the ensembles of bias_{VTECs} \mathbf{X}_2^B 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 ($\bar{\mathbf{x}}_{n_{pc}+1 \times 1}^B$) of Eq. (11) and the error covariance matrix of the background model ($\mathbf{C}_{n_{pc}+1 \times n_{pc}+1}^B$) are computed as follows:

$$\bar{\mathbf{x}}^B = \frac{1}{Ne} \sum_{i=1}^{Ne} \mathbf{x}_i^B, \text{ and} \quad (12)$$

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

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

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

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

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

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

385 where $\mathbf{Y}_{m \times N_e}^{\text{OBS}}$ and $\bar{\mathbf{y}}_{m \times 1}^{\text{OBS}}$ represent the ensembles (i.e., perturbed by the
 386 estimated noise derived from GIM products) and the ensemble mean of
 387 GIM/UQRG VTECs for the one grid point, respectively. Considering Eqs.
 388 (14-16), the analyzed states and their covariance matrix depend on differ-
 389 ences between the real observations (\mathbf{Y}^{OBS}) and model predictions ($\mathbf{H}\mathbf{X}^B$),
 390 while considering their weights, which are reflected in the Kalman gain matrix
 391 ($\mathbf{K}_{n_{pc}+1 \times m}$) that is defined as:

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

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

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

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

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

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

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

$$\text{DDA-NeQuick} = \mathbf{P}_{\text{NeQuick}}\hat{\mathbf{E}}_{\text{NeQuick}}^T + \hat{\text{bias}}_{\text{VTECsNeQuick}}, \quad (21)$$

$$\text{DDA-TIEGCM} = \mathbf{P}_{\text{TIEGCM}}\hat{\mathbf{E}}_{\text{TIEGCM}}^T + \hat{\text{bias}}_{\text{VTECsTIEGCM}}, \quad (22)$$

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

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

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

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

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

417 3.3 Evaluating the results

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

425 4 Results

426 The DDA procedure is performed using 90 ensemble members and the first
 427 30 of EOFs are found to represent 99% of the eigenvalues. Eventually, the
 428 assimilated EOFs Eq. (15) replace the model-derived EOFs in Eq. (19) for
 429 simulating VTECs of the same day (i.e., now-casting). This means that the
 430 now-casting of NeQuick is estimated using Eq. (21), and that of TIEGCM
 431 from Eq. (22). For forecasting VTECs during the next day the general model
 432 reads as Eq. (23), i.e., for forecasting based on the DDA NeQuick, we use Eq.
 433 (25), and the DDA TIEGCM forecasts follow Eq. (26).

434 An overview of the work-flow of this study to apply DDA on NeQuick or
435 TIEGCM and testing its performance for forecasting global VTECs is pre-
436 sented in Fig. (3). In what follows, VTEC estimates from the DDA are assessed
437 in different ways. In Section 4.1, the prediction of EOFs is presented. Then,
438 the VTEC estimates from NeQuick, TIEGCM, DDA-NeQuick, and DDA-
439 TIEGCM are compared with the VTECs derived from GIM/UQRG in the
440 forecasting mode (Section 4.2). This is done to understand how the DDA
441 changed the original modeled values during September 2017. In Section 4.3,
442 the 6-hourly global maps of DDA in the forecasting mode during two days
443 with high and low Kp are compared with those of GIM/UQRG to see whether
444 the new model represents expected spatial-temporal as reflected in the global
445 models. Finally, the time-series of VTECs from DDA are compared with the
446 final IONEX GIM/CODE products over some selected IGS stations in Section
447 4.4.

448 4.1 Predicting EOFs in the forecasting mode

449 PCA is applied on the global VTEC fields (with spatial/temporal resolution
450 $2.5^\circ \times 5^\circ$ in latitude and longitude / 15 minutes). Here, we use GIM/UQRG
451 to derive the DDA-NeQuick Eq. (21) and DDA-TIEGCM Eq. (22) during
452 September 2017. Plots in Fig. (4,a-e) represent the first EOF of VTECs. In
453 addition, plots in Fig. (4,g-h) indicate the magnitude of singular values that
454 correspond to all of the PCA modes. The amount of VTEC variability captured
455 by first EOF is found to be 32.28%, 47.26%, 43.96%, 44.20% and 44.22% of
456 the total variance for NeQuick, TIEGCM (i.e., TIEGCM and for the top level,
457 height from $\sim 500-800$ km to $\sim 20, 200$ km, we used NeQuick), DDA-NeQuick,
458 DDA-TIEGCM and GIM/UQRG VTECs, respectively. The numerical results
459 show that after implementing the DDA, the overall spatial correlation coeffi-
460 cient between the EOFs of GIM/UQRG and models are increased from 90.17%
461 with NeQuick to 99.81% with DDA-NeQuick, and from 62.66% with TIEGCM
462 to 99.84% with DDA-TIEGCM.

463 4.2 Comparison of VTEC predictions with GIM/UQRG

464 To assess whether the daily DDA improves the performance of empirical (i.e.,
465 NeQuick) or physic-based (TIEGCM) models in the forecasting mode, the
466 assimilated EOF maps are used to forecast VTECs for the next day based
467 on Eq. (23). Figure (5,left) presents the improvements in terms of RMSE of
468 VTECs compared to the GIM/UQRG in the forecasting mode. The DDA
469 results are found to agree well with the GIM/UQRG (e.g., the CC of 91%
470 and 93% for DDA-NeQuick and DDA-TIEGCM, respectively). The average
471 improvement is found to be 42.46% (in the range of 14.47 – 70.45%) and 31.89%
472 (in the range of 6.43 – 59.65%) for the DDA-NeQuick and DDA-TIEGCM,
473 respectively. In addition, the mean of global uncertainties of VTECs Eq. (24)
474 derived from NeQuick (TIEGCM) in the DDA procedure is decreased from 5.4
475 (4.47) TECU to 0.08 (0.44) TECU in the forecasting step during September,
476 2017.

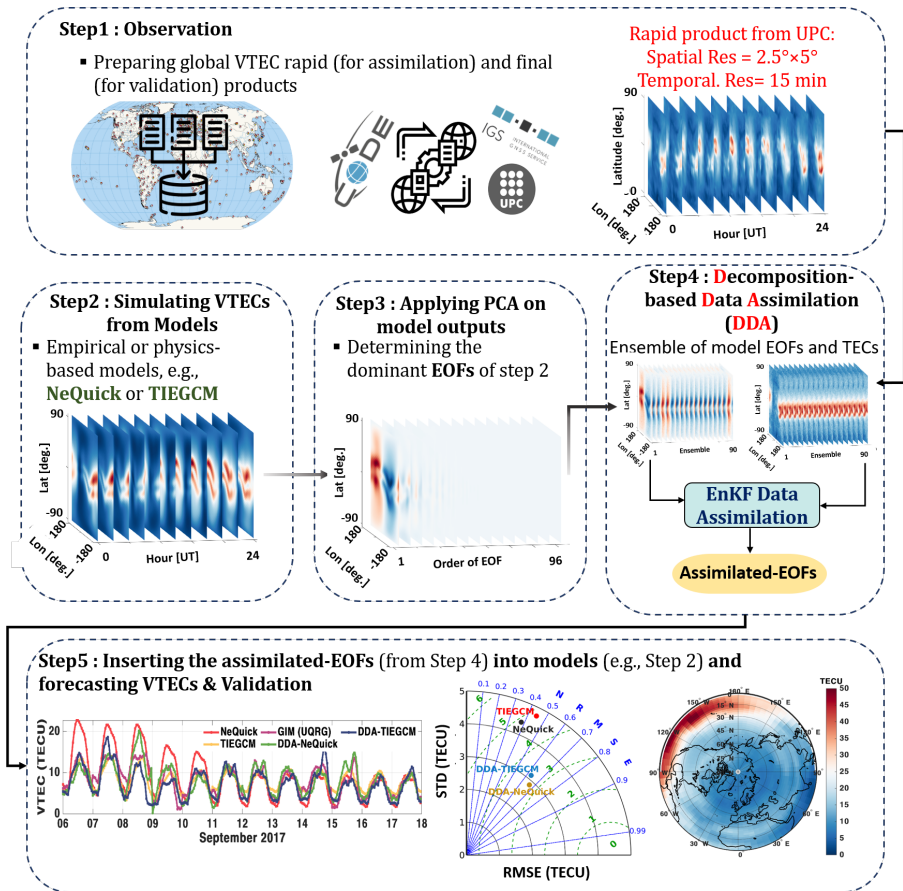


Fig. 3: An overview of the proposed DDA procedure and validation experiments. The procedure is divided into five steps: 1- Generating the ensemble of VTECs from GIM/UQRG, 2- Simulating VTECs and generating ensemble of them from empirical model NeQuick and physics-based model TIEGCM, 3- Applying PCA on 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.

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

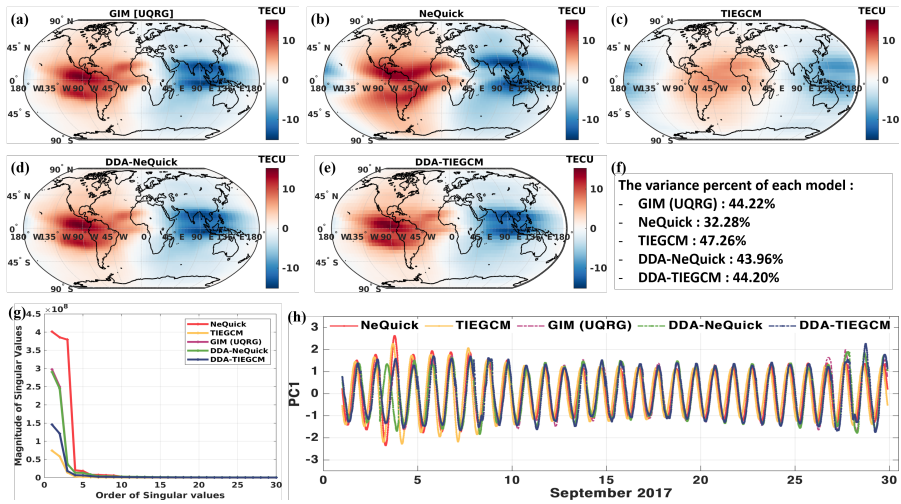


Fig. 4: In (a-e), the first EOF of the VTECs from GIM/UQRG, NeQuick, TIEGCM, DDA-NeQuick and DDA-TIEGCM, respectively is shown, in (f), the variance percentage of the first PCA mode of different models is presented, in (g), the magnitude of the singular values are shown, and in (h), the corresponding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

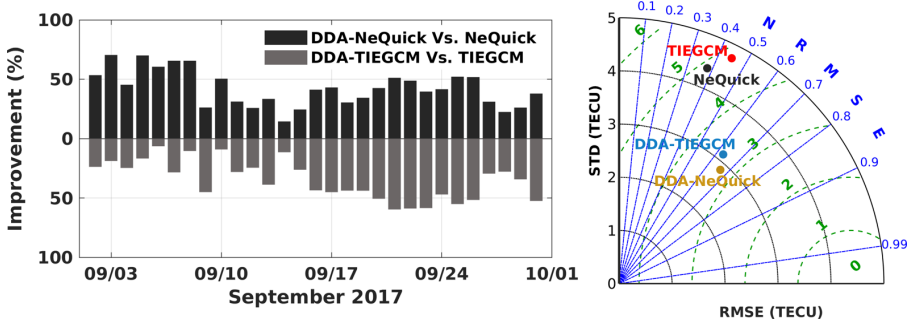


Fig. 5: Left panel: A comparison between evaluation statistics for forecasting VTECs after implementing DDA during September 2017. The improvement are estimated between the NeQuick (TIEGCM) and the DDA-NeQuick (DDA-TIEGCM) models relative to the GIM/UQRG. Right panel: an overview of the three performance 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 26th (DOY=269) and 8th (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 8th, 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 $\pm 30^\circ$ 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 26th (low K_p) are ~ 20 and 15 TECU around 06:00 and 12:00 UT (day time), respectively. These values are estimated to be ~ 30 and 17 TECU for TIEGCM and DDA-TIEGCM. The results for September 8th (high K_p) are found to be ~ 30 and 18 TECU for NeQuick and DDA-NeQuick, while ~ 23 and 19 for TIEGCM and DDA-TIEGCM, respectively.

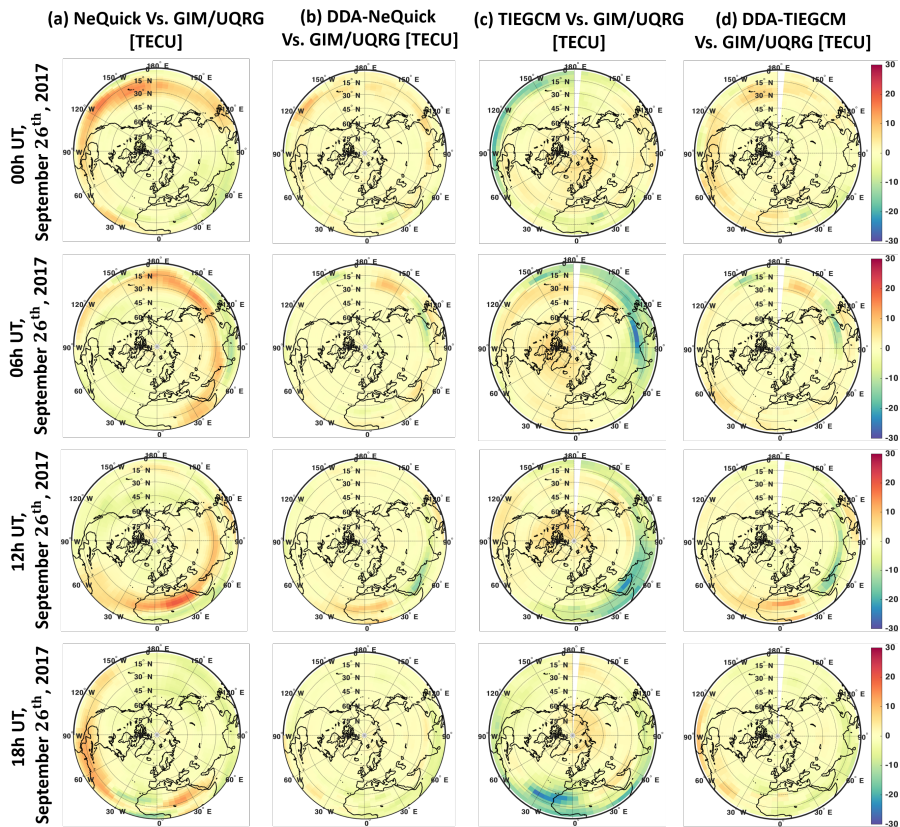


Fig. 6: An overview of the VTEC changes in the forecasting mode on September 26th, 2017 (low geomagnetic activity with $K_p = 2$). The left to right maps : a) NeQuick against GIM/UQRG, b) DDA-NeQuick against GIM/UQRG, c) TIEGCM against GIM/UQRG and d) DDA-TIEGCM against GIM/UQRG, respectively.

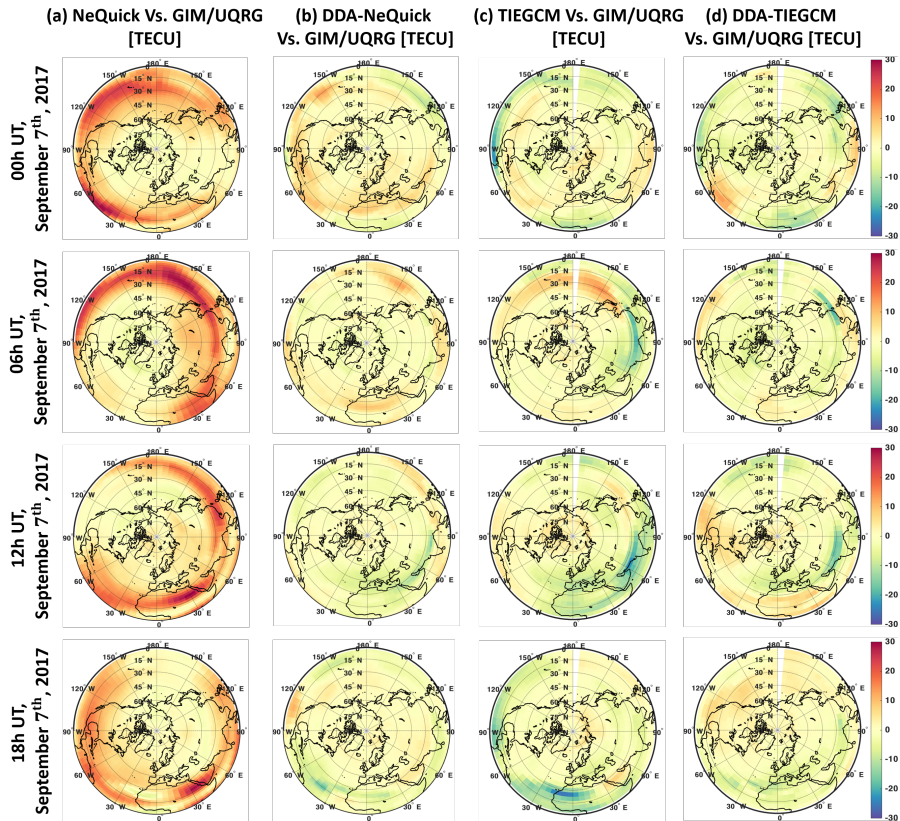


Fig. 7: An overview of VTEC changes in the forecasting mode on September 8th, 2017 (high geomagnetic activity with $K_p = 8$). The left to right maps correspond to the a) NeQuick against GIM/UQRG, b) DDA-NeQuick against GIM/UQRG, c) TIEGCM against GIM/UQRG and d) DDA-TIEGCM against GIM/UQRG, respectively.

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

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

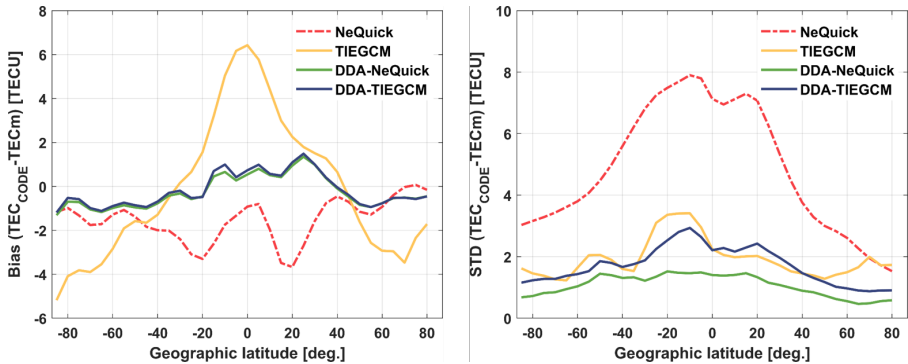


Fig. 8: Bias and STD of the differences between the model derived VTEC estimates and the IGS's GIM/CODE products. The statistics were generated in 2.5° geographic latitude bins for the entire September 2017.

542 The diurnal VTEC estimates from NeQuick, TIEGCM, DDA-NeQuick,
 543 and DDA-TIEGCM are compared with the GIM/CODE ionosphere estimates
 544 over some IGS stations. We selected 12 days of September 6th-18th, 2017 to
 545 perform the comparisons and the results are shown in Fig. (9). These days are
 546 selected because of changes in the geomagnetic index were considerable (see
 547 Fig. (1)). After implementing DDA on NeQuick (TIEGCM), the overall RMSE
 548 is reduced by 34.3% (30.1%), 57.8% (19.3%), 24.5% (18.9%), 20.8% (47.1%),
 549 51.4% (10.2%) and 21.8% (13.9%) in the six IGS stations (FFMJ - latitude:
 550 50.09° and longitude: 8.66°, Germany; URUM - latitude: 43.80° and longi-
 551 tude: 87.60°, China; SCRZ - latitude: -17.80° and longitude: -63.16°, Bolivia;
 552 YELL - latitude: 62.48° and longitude: -114.48°, Canada; ZAMB - latitude:
 553 -15.43° and longitude: 28.31°, Zambia); and NYAL - latitude: 79.83° and lon-
 554 gitude: 11.86°, Norway). More statistical evidences of the DDA improvements
 555 are provided in Table.1.

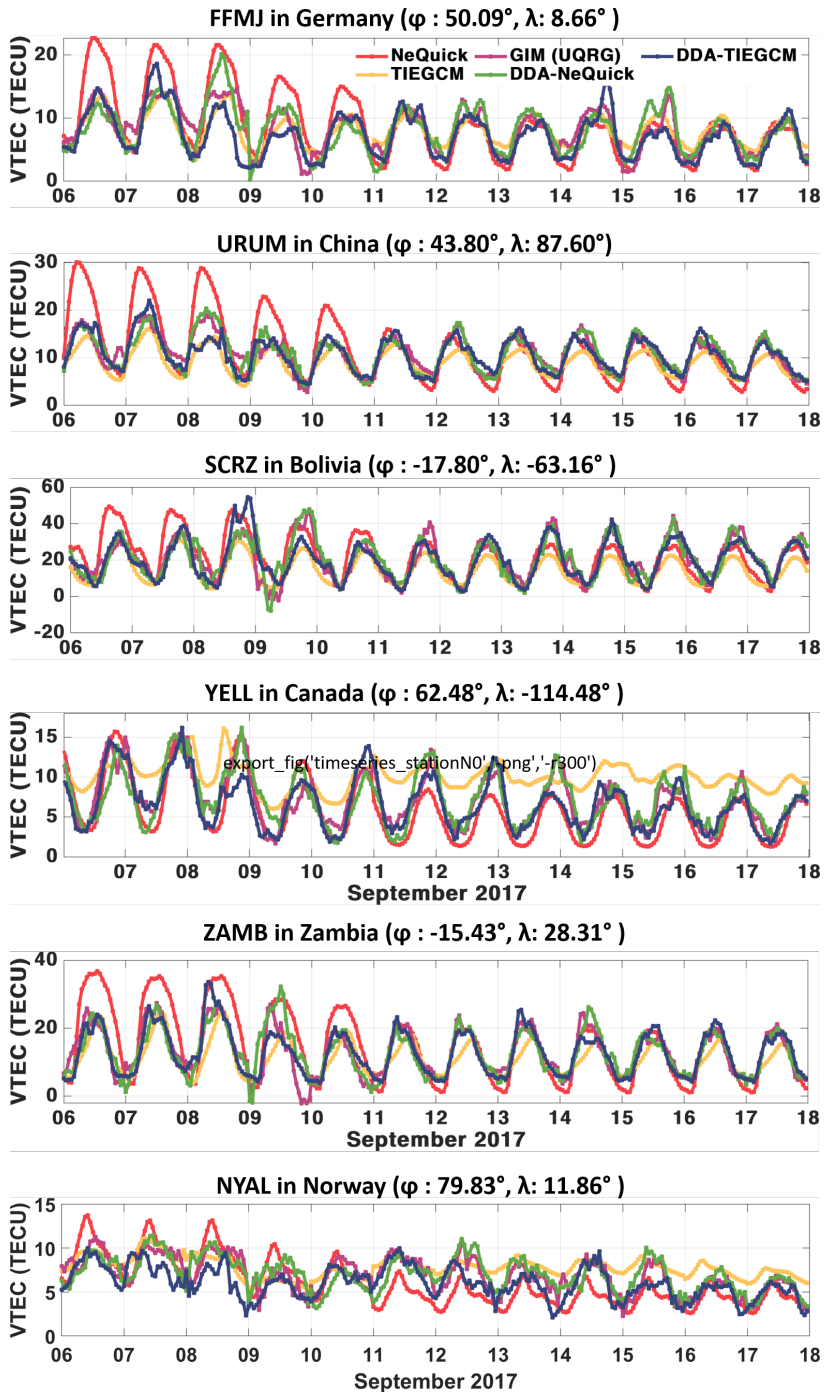


Fig. 9: Diurnal VTEC variations obtained from the NeQuick, TIEGCM, DDA-NeQuick, and DDA-TIEGCM, as well as GIM/CODE over the six selected IGS stations during 12 days in September 2017 (6th-18th).

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
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
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

556 A comprehensive comparison in terms of global RMSE (Eq. (30)) and Fit
557 (Eq. (33)) are performed with the final VTEC products of GIM/CODE. These
558 hourly measures are summarized in Figs. (10 and (11)), which indicates that the
559 main differences can be found as expected during high solar activity (i.e., from
560 September 7th to 9th, 2017). DDA can improve these differences by 47.2% and
561 26.6% for NeQuick and TIEGCM, respectively.

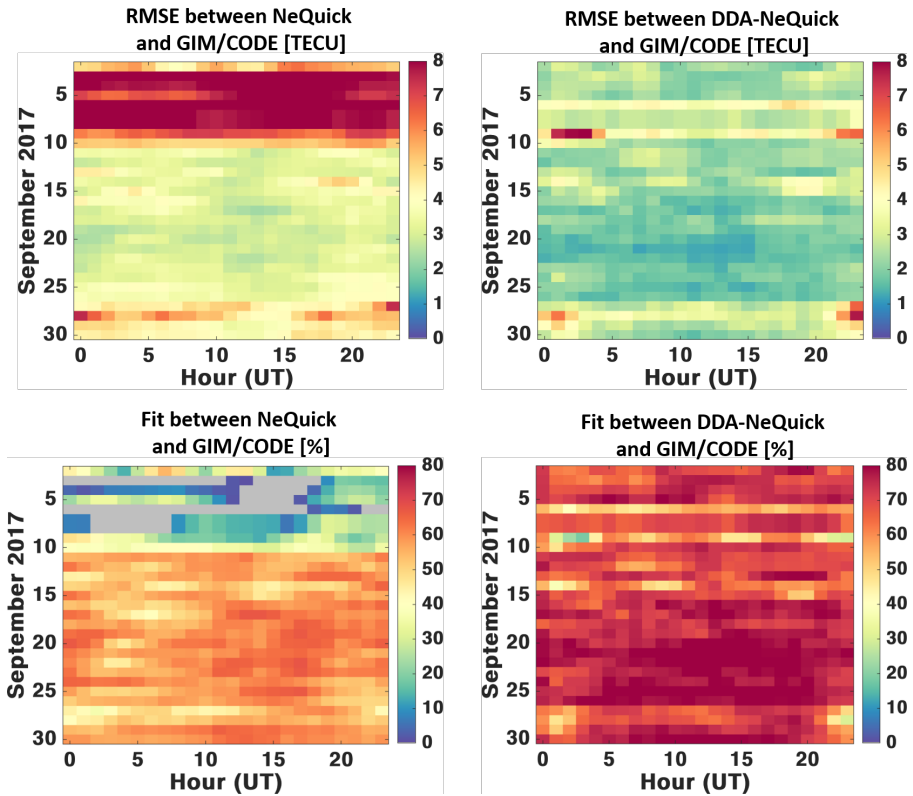


Fig. 10: Hourly global RMSE and Fit before and after performing the DDA. The specific UT hour are shown along the x-axis, each day of September 2017 is represented on the y-axis. The colored values show the RMSE and Fit values, and the gray color in the Fit maps are related to the negative fitting values.

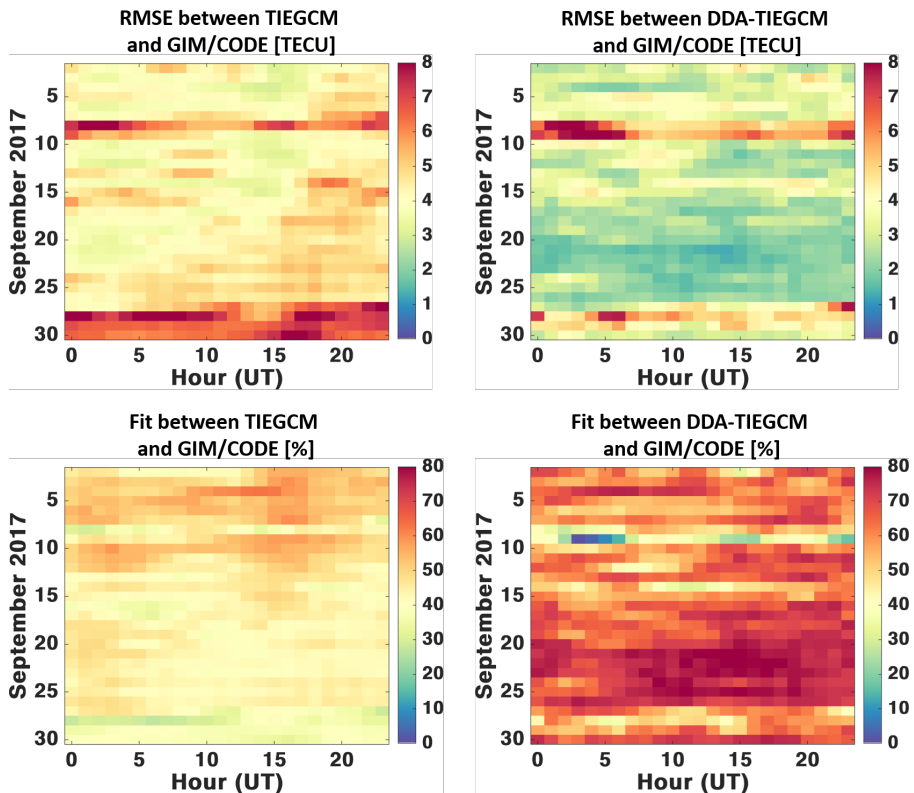


Fig. 11: Global RMSE and Fit calculated by Eqs. (30 and 33) for measurements taken at a specific UT hour (x axis), for a specific day of September 2017 (y axis), before and after applying the DDA approach on TIEGCM.

4.5 Validation with the VTECs derived from GPS measurements

562

563

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,

575

576 green and blue colors, respectively. The left panel represents the higher correlations
 577 between of the DDA results and GPS-VTEC, i.e., 90% and 89%, while
 578 these value are about 80% and 79% for the original models. Normalized his-
 579 tograms of the VTEC modeling errors relative to GPS-VTECs are shown in
 580 the right panel of Fig. (12). They indicate that the mean of normalized errors
 581 of the DDA-NeQuick and DDA-TIEGCM are low, i.e., 0.76 and 0.6 TECU,
 582 respectively. The STD of DDA-TIEGCM and DDA-NeQuick model is also
 583 lower than that of the original TIEGCM and NeQuick (3.3 vs. 4.6 and 3.5 vs.
 584 6.2 TECU) after implementing the proposed approach.

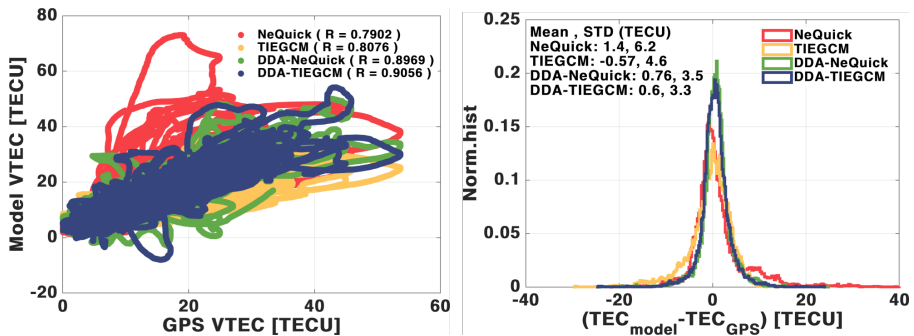


Fig. 12: Left panel: Corresponding scatter-plots of modeled (i.e., NeQuick, TIEGCM, DDA-NeQuick and DDA-TIEGCM) and measured GPS-VTECs values which the Pearson correlation coefficient of each model with observations are shown in the upper right corner. Right panel: Histogram of corresponding residuals 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.

585 5 Conclusion

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

599 the rapid and final GIM VTEC products. The main findings of this study can
600 be summarized as:

- 601 • The DDA is implemented here by considering 90 ensemble members and
602 only the first 30 EOFs with the highest rank of each grid point are used
603 for assimilation. After integrating the VTECs from GIM/UQRG with EOFs
604 from models, the assimilated EOFs are used in the forecasting step. The
605 new assimilated models (DDA-NeQuick and DDA-TIEGCM) provide bet-
606 ter VTEC estimates than the original models especially in days (and at
607 those times of the day) with more pronounced ionospheric dynamics, where
608 considerable differences exist between the original models and GIM/CODE
609 VTECs.
- 610 • Comparisons between DDA-NeQuick (DDA-TIEGCM) and original models
611 against the VTEC estimates from GIM/UQRG represent the capability of
612 the proposed model in simulating or forecasting VTECs in the EIA region.
613 The differences between the NeQuick (TIEGCM) and DDA-NeQuick (DDA-
614 TIEGCM) compared to the GIM/CODE indicate that the reduction of error
615 around EIA is found to be 50 (30)% approximately.
- 616 • Statistical measures indicate that the DDA-NeQuick and DDA-TIEGCM
617 perform better than the original models, compared to the final product
618 GIM/CODE, in both now-casting and forecasting modes. For example, the
619 monthly averages of RMSE, bias and fit parameters in the forecasting step
620 are found to be improved from the original values of 6.62 (5.09) TECU, -
621 1.51 (-0.31) TECU, 0.26 (0.43) to 3.90 (3.63) TECU, -0.30 (-0.22) TECU,
622 0.56 (0.59) after implementing the DDA procedure into the NeQuick (TIE-
623 GCM), respectively.

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

628 **6 Acknowledgment**

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

634 **7 Funding**

635 E. Forootan acknowledges the financial support by the Danmarks Frie Forskn-
636 ingsfond [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/>

674 [10.1007/s00190-020-01397-1](https://doi.org/10.1007/s00190-020-01397-1)

- 675 [9] Prol, F.S., Kodikara, T., Hoque, M.M., Borries, C.: Global-scale
676 ionospheric tomography during the 17 March 2015 geomagnetic
677 storm. *Space Weather* **n/a**(n/a), 2021–002889. [https://doi.org/10.1029/
678 2021SW002889](https://doi.org/10.1029/2021SW002889)
- 679 [10] Hajj, G.A., Romans, L.J.: Ionospheric electron density profiles obtained
680 with the global positioning system: Results from the GPS/MET exper-
681 iment. *Radio Science* **33**(1), 175–190 (1998). [https://doi.org/10.1029/
682 97RS03183](https://doi.org/10.1029/97RS03183)
- 683 [11] Shume, E.B., Vergados, P., Komjathy, A., Langley, R.B., Durgonics, T.:
684 Electron number density profiles derived from radio occultation on the
685 cassiope spacecraft. *Radio Science* **52**(9), 1190–1199 (2017). [https://doi.
686 org/10.1002/2017RS006321](https://doi.org/10.1002/2017RS006321)
- 687 [12] Seeber, G.: Satellite geodesy: Foundations, methods and applications.
688 *INTERNATIONAL HYDROGRAPHIC REVIEW* **4**(3), 92–93 (2003)
- 689 [13] Montenbruck, O., Gill, E.: *Satellite Orbits: Models, Methods and
690 Applications*. Springer, ??? (2012)
- 691 [14] Sebestyen, G., Fujikawa, S., Galassi, N., Chuchra, A.: *Low Earth Orbit
692 Satellite Design* vol. 36. Springer, ??? (2018). [https://doi.org/10.1007/
693 978-3-319-68315-7](https://doi.org/10.1007/978-3-319-68315-7)
- 694 [15] Gonzalo, J., Domínguez, D., López, D.: On the challenge of a century
695 lifespan satellite. *Progress in Aerospace Sciences* **70**, 28–41 (2014). [https:
696 //doi.org/10.1016/j.paerosci.2014.05.001](https://doi.org/10.1016/j.paerosci.2014.05.001)
- 697 [16] Sneeuw, N., Flury, J., Rummel, R.: *Science Requirements On Future
698 Missions And Simulated Mission Scenarios*, pp. 113–142. Springer, ???
699 (2005). https://doi.org/10.1007/0-387-33185-9_10
- 700 [17] McCoy, R.P.: Space weather comes of age: new sensors and models for
701 ionospheric specification and forecast. In: Huang, H.-L.A., Bloom, H.J.
702 (eds.) *Atmospheric and Environmental Remote Sensing Data Processing
703 and Utilization: an End-to-End System Perspective*, vol. 5548, pp. 341–
704 347. SPIE, ??? (2004). International Society for Optics and Photonics.
705 <https://doi.org/10.1117/12.562786>
- 706 [18] Decker, D.T., McNamara, L.F.: Validation of ionospheric weather pre-
707 dicted by global assimilation of ionospheric measurements (gain) mod-
708 els. *Radio Science* **42**(4) (2007). <https://doi.org/10.1029/2007RS003632>

- 709 [19] Cander, L.R.: Ionospheric research and space weather services. Jour-
710 nal of Atmospheric and Solar-Terrestrial Physics **70**(15), 1870–1878
711 (2008). <https://doi.org/10.1016/j.jastp.2008.05.010>. Ionospheric Effects
712 and Telecommunications
- 713 [20] Schunk, R.W., Scherliess, L., Eccles, V., Gardner, L.C., Sojka, J.J., Zhu,
714 L., Pi, X., Mannucci, A.J., Komjathy, A., Wang, C., Rosen, G.: Chal-
715 lenges in specifying and predicting space weather. Space Weather **19**(2),
716 2019–002404 (2021). <https://doi.org/10.1029/2019SW002404>
- 717 [21] Kouris, S.S., Muggleton, L.M.: Diurnal variation in the e-layer ionization.
718 Journal of Atmospheric and Terrestrial Physics **35**(1), 133–139 (1973).
719 [https://doi.org/10.1016/0021-9169\(73\)90221-3](https://doi.org/10.1016/0021-9169(73)90221-3)
- 720 [22] Chapman, S.: The absorption and dissociative or ionizing effect of
721 monochromatic radiation in an atmosphere on a rotating earth part ii.
722 grazing incidence. Proceedings of the Physical Society **43**(5), 483–501
723 (1931). <https://doi.org/10.1088/0959-5309/43/5/302>
- 724 [23] Bessarab, F.S., Korenkov, Y.N., Klimenko, V.V., Klimenko, M.V.,
725 Zhang, Y.: E-region ionospheric storm on may 1–3, 2010: Gsm tip
726 model representation and suggestions for iri improvement. Advances in
727 Space Research **55**(8), 2124–2130 (2015). [https://doi.org/10.1016/j.asr.](https://doi.org/10.1016/j.asr.2014.08.003)
728 [2014.08.003](https://doi.org/10.1016/j.asr.2014.08.003). INTERNATIONAL REFERENCE IONOSPHERE AND
729 GLOBAL NAVIGATION SATELLITE SYSTEMS
- 730 [24] Davies, K.: Ionospheric Radio. Electromagnetic Waves. Institution
731 of Engineering and Technology, ??? (1990). [https://doi.org/10.1049/](https://doi.org/10.1049/PBEW031E)
732 [PBEW031E](https://doi.org/10.1049/PBEW031E)
- 733 [25] Bremer, J.: Trends in the ionospheric e and f regions over europe.
734 Annales Geophysicae **16**(8), 986–996 (1998). [https://doi.org/10.1007/](https://doi.org/10.1007/s00585-998-0986-9)
735 [s00585-998-0986-9](https://doi.org/10.1007/s00585-998-0986-9)
- 736 [26] Hochegger, G., Nava, B., Radicella, S., Leitinger, R.: A family of iono-
737 spheric models for different uses. Physics and Chemistry of the Earth
738 **25**(4), 307–310 (2000). [https://doi.org/10.1016/S1464-1917\(00\)00022-2](https://doi.org/10.1016/S1464-1917(00)00022-2)
- 739 [27] Mikhailov, A.V.: Ionospheric f1 layer long-term trends and the geomag-
740 netic control concept. Annales Geophysicae **26**(12), 3793–3803 (2008).
741 <https://doi.org/10.5194/angeo-26-3793-2008>
- 742 [28] Torr, M.R., Torr, D.G.: The seasonal behaviour of the f2-layer of the
743 ionosphere. Journal of Atmospheric and Terrestrial Physics **35**(12),
744 2237–2251 (1973). [https://doi.org/10.1016/0021-9169\(73\)90140-2](https://doi.org/10.1016/0021-9169(73)90140-2)
- 745 [29] Rishbeth, H., Müller-Wodarg, I., Zou, L., Fuller-Rowell, T., Millward,

- 746 G., Moffett, R., Idenden, D., Aylward, A.: Annual and semiannual vari-
 747 ations in the ionospheric f2-layer: Ii. physical discussion. In: *Annales*
 748 *Geophysicae*, vol. 18, pp. 945–956 (2000). [https://doi.org/10.1007/](https://doi.org/10.1007/s00585-000-0945-6)
 749 [s00585-000-0945-6](https://doi.org/10.1007/s00585-000-0945-6). Springer
- 750 [30] Rishbeth, H., Mendillo, M.: Patterns of f2-layer variability. *Journal of*
 751 *Atmospheric and Solar-Terrestrial Physics* **63**(15), 1661–1680 (2001).
 752 [https://doi.org/10.1016/S1364-6826\(01\)00036-0](https://doi.org/10.1016/S1364-6826(01)00036-0)
- 753 [31] Zossi, B.S., Fagre, M., de Haro Barbás, B.F., Elias, A.G.: Ionospheric
 754 conductance using different iri f2 layer models. *Journal of Atmospheric*
 755 *and Solar-Terrestrial Physics* **225**, 105759 (2021). [https://doi.org/10.](https://doi.org/10.1016/j.jastp.2021.105759)
 756 [1016/j.jastp.2021.105759](https://doi.org/10.1016/j.jastp.2021.105759)
- 757 [32] Chen, Y., Liu, L., Le, H., Zhang, H.: Latitudinal dependence of day-
 758 time electron density bite-out in the ionospheric f2-layer. *Journal of*
 759 *Geophysical Research: Space Physics* **126**(1), 2020–028277 (2021). [https:](https://doi.org/10.1029/2020JA028277)
 760 [//doi.org/10.1029/2020JA028277](https://doi.org/10.1029/2020JA028277)
- 761 [33] Sojka, J.J.: Global scale, physical models of the f region ionosphere.
 762 *Reviews of Geophysics* **27**(3), 371–403 (1989). [https://doi.org/10.1029/](https://doi.org/10.1029/RG027i003p00371)
 763 [RG027i003p00371](https://doi.org/10.1029/RG027i003p00371)
- 764 [34] Anderson, D.N., Buonsanto, M.J., Codrescu, M., Decker, D., Fesen,
 765 C.G., Fuller-Rowell, T.J., Reinisch, B.W., Richards, P.G., Roble,
 766 R.G., Schunk, R.W., Sojka, J.J.: Intercomparison of physical models
 767 and observations of the ionosphere. *Journal of Geophysical Research:*
 768 *Space Physics* **103**(A2), 2179–2192 (1998). [https://doi.org/10.1029/](https://doi.org/10.1029/97JA02872)
 769 [97JA02872](https://doi.org/10.1029/97JA02872)
- 770 [35] Man-Lian, Z., She-Ping, S., *et al.*: A physical numerical ionospheric
 771 model and its simulation results. *Communications in Theoretical Physics*
 772 **41**(5), 795 (2004). <https://doi.org/10.1088/0253-6102/41/5/795>
- 773 [36] Qian, L., Burns, A.G., Emery, B.A., Foster, B., Lu, G., Maute, A.,
 774 Richmond, A.D., Roble, R.G., Solomon, S.C., Wang, W.: The NCAR
 775 TIE-GCM: A community model of the coupled thermosphere/iono-
 776 sphere system. *Modeling the ionosphere-thermosphere system* **201**,
 777 73–83 (2014). <https://doi.org/10.1002/9781118704417.ch7>
- 778 [37] Wang, C., Hajj, G., Pi, X., Rosen, I.G., Wilson, B.: Development of
 779 the global assimilative ionospheric model. *Radio Science* **39**(1) (2004).
 780 <https://doi.org/10.1029/2002RS002854>
- 781 [38] Khattatov, B., Murphy, M., Cruikshank, B., Fuller-Rowell, T.: Iono-
 782 spheric corrections from a prototype operational assimilation and
 783 forecast system. In: *PLANS 2004. Position Location and Navigation*

- 784 Symposium (IEEE Cat. No.04CH37556), pp. 518–526 (2004). <https://doi.org/10.1109/PLANS.2004.1309037>
785
- 786 [39] Angling, M.J., Cannon, P.S.: Assimilation of radio occultation measure-
787 ments into background ionospheric models. *Radio Science* **39**(1) (2004).
788 <https://doi.org/10.1029/2002RS002819>
- 789 [40] Scherliess, L., Schunk, R.W., Sojka, J.J., Thompson, D.C., Zhu, L.: Utah
790 state university global assimilation of ionospheric measurements gauss-
791 markov kalman filter model of the ionosphere: Model description and
792 validation. *Journal of Geophysical Research: Space Physics* **111**(A11)
793 (2006). <https://doi.org/10.1029/2006JA011712>
- 794 [41] Galkin, I.A., Reinisch, B.W., Huang, X., Bilitza, D.: Assimilation of
795 GIRO data into a real-time IRI. *Radio Science* **47**(04), 1–10 (2012).
796 <https://doi.org/10.1029/2011RS004952>
- 797 [42] Chen, C.-H., Lin, C., Chen, W.-H., Matsuo, T.: Modeling the ionospheric
798 prereversal enhancement by using coupled thermosphere-ionosphere data
799 assimilation. *Geophysical Research Letters* **44**(4), 1652–1659 (2017).
800 <https://doi.org/10.1002/2016GL071812>
- 801 [43] Elvidge, Sean, Angling, Matthew J.: Using the local ensemble transform
802 Kalman filter for upper atmospheric modelling. *J. Space Weather Space*
803 *Clim.* **9**, 30 (2019). <https://doi.org/10.1051/swsc/2019018>
- 804 [44] Aa, E., Zhang, S.-R., Erickson, P.J., Wang, W., Coster, A.J., Ride-
805 out, W.: 3-D regional ionosphere imaging and SED reconstruction
806 with a new TEC-Based ionospheric data assimilation system (TIDAS).
807 *Space Weather* **20**(4), 2022–003055 (2022). <https://doi.org/10.1029/2022SW003055>
808
- 809 [45] Maute, A.: Thermosphere-ionosphere-electrodynamics general circula-
810 tion model for the ionospheric connection explorer: TIEGCM-ICON.
811 *Space Science Reviews* **212**(1), 523–551 (2017). <https://doi.org/10.1007/s11214-017-0330-3>
812
- 813 [46] Fuller-Rowell, T.J., Rees, D.: A three-dimensional time-dependent global
814 model of the thermosphere. *Journal of Atmospheric Sciences* **37**(11),
815 2545–2567 (1980)
- 816 [47] Fuller-Rowell, T.J., Rees, D.: Derivation of a conservation equation
817 for mean molecular weight for a two-constituent gas within a three-
818 dimensional, time-dependent model of the thermosphere. *Planetary*
819 *and Space Science* **31**(10), 1209–1222 (1983). [https://doi.org/10.1016/0032-0633\(83\)90112-5](https://doi.org/10.1016/0032-0633(83)90112-5)
820

- 821 [48] Fuller-Rowell, T.J., Rees, D., Quegan, S., Moffett, R.J., Bailey, G.J.:
822 Interactions between neutral thermospheric composition and the polar
823 ionosphere using a coupled ionosphere-thermosphere model. *Journal of*
824 *Geophysical Research: Space Physics* **92**(A7), 7744–7748 (1987). <https://doi.org/10.1029/JA092iA07p07744>
825
- 826 [49] Millward, G.H., Rishbeth, H., Fuller-Rowell, T.J., Aylward, A.D., Que-
827 gan, S., Moffett, R.J.: Ionospheric f 2 layer seasonal and semiannual
828 variations. *Journal of Geophysical Research: Space Physics* **101**(A3),
829 5149–5156 (1996). <https://doi.org/10.1029/95JA03343>
- 830 [50] Ridley, A.J., Deng, Y., Tóth, G.: The global ionosphere–thermosphere
831 model. *Journal of Atmospheric and Solar–Terrestrial Physics* **68**(8), 839–
832 864 (2006). <https://doi.org/10.1016/j.jastp.2006.01.008>
- 833 [51] Wang, C., Shi, C., Fan, L., Zhang, H.: Improved modeling of global iono-
834 spheric total electron content using prior information. *Remote Sensing*
835 **10**(1) (2018). <https://doi.org/10.3390/rs10010063>
- 836 [52] Elvidge, S., Godinez, H.C., Angling, M.J.: Improved forecasting of
837 thermospheric densities using multi-model ensembles. *Geoscientific*
838 *Model Development* **9**(6), 2279–2292 (2016). [https://doi.org/10.5194/](https://doi.org/10.5194/gmd-9-2279-2016)
839 [gmd-9-2279-2016](https://doi.org/10.5194/gmd-9-2279-2016)
- 840 [53] Yao, Y., Liu, L., Kong, J., Zhai, C.: Global ionospheric modeling based on
841 multi-gnss, satellite altimetry, and formosat-3/cosmic data. *GPS Solu-*
842 *tions* **22**(4), 1–12 (2018). <https://doi.org/10.1007/s10291-018-0770-6>
- 843 [54] Pedatella, N.M., Anderson, J.L., Chen, C.H., Raeder, K., Liu, J., Liu,
844 H.-L., Lin, C.H.: Assimilation of ionosphere observations in the whole
845 atmosphere community climate model with thermosphere-ionosphere
846 extension (waccmx). *Journal of Geophysical Research: Space Physics*
847 **125**(9), 2020–028251 (2020). <https://doi.org/10.1029/2020JA028251>
- 848 [55] Schunk, R.W., Scherliess, L., Sojka, J.J.: Recent approaches to modeling
849 ionospheric weather. *Advances in Space Research* **31**(4), 819–828 (2003).
850 [https://doi.org/10.1016/S0273-1177\(02\)00791-3](https://doi.org/10.1016/S0273-1177(02)00791-3)
- 851 [56] Withers, P.: Prediction of uncertainties in atmospheric properties mea-
852 sured by radio occultation experiments. *Advances in Space Research*
853 **46**(1), 58–73 (2010). <https://doi.org/10.1016/j.asr.2010.03.004>
- 854 [57] McNamara, L.F., Angling, M.J., Elvidge, S., Fridman, S.V., Haus-
855 man, M.A., Nickisch, L.J., McKinnell, L.-A.: Assimilation procedures for
856 updating ionospheric profiles below the F2 peak. *Radio Science* **48**(2),
857 143–157 (2013). <https://doi.org/10.1002/rds.20020>

- 858 [58] Fu, N., Guo, P., Chen, Y., Wu, M., Huang, Y., Hu, X., Hong, Z.:
859 The analysis of assumptions' error sources on assimilating ground-
860 based/spaceborne ionospheric observations. *Journal of Atmospheric and*
861 *Solar-Terrestrial Physics* **207**, 105354 (2020). [https://doi.org/10.1016/j.](https://doi.org/10.1016/j.jastp.2020.105354)
862 [jastp.2020.105354](https://doi.org/10.1016/j.jastp.2020.105354)
- 863 [59] Nina, A., Nico, G., Mitrović, S.T., Čadež, V.M., Milošević, I.R.,
864 Radovanović, M., Popović, L.C.: Quiet ionospheric d-region (qiondr)
865 model based on vlf/lf observations. *Remote Sensing* **13**(3) (2021). [https:](https://doi.org/10.3390/rs13030483)
866 [//doi.org/10.3390/rs13030483](https://doi.org/10.3390/rs13030483)
- 867 [60] Ren, X., Zhang, J., Chen, J., Zhang, X.: Global ionospheric modeling
868 using multi-gnss and upcoming leo constellations: Two methods and
869 comparison. *IEEE Transactions on Geoscience and Remote Sensing*, 1–15
870 (2021). <https://doi.org/10.1109/TGRS.2021.3050413>
- 871 [61] Gulyaeva, T.L., Arikan, F., Hernandez-Pajares, M., Stanislawska, I.:
872 Gim-tec adaptive ionospheric weather assessment and forecast system.
873 *Journal of Atmospheric and Solar-Terrestrial Physics* **102**, 329–340
874 (2013). <https://doi.org/10.1016/j.jastp.2013.06.011>
- 875 [62] Rodrigues, F., Wright, L., Moraes, A., Freitas, M.: ScintPi: On the use
876 of low-cost sensors to monitor ionospheric weather and evaluate poten-
877 tial risks. In: 43rd COSPAR Scientific Assembly. Held 28 January - 4
878 February, vol. 43, p. 673 (2021)
- 879 [63] Stanislawska, I., Gulyaeva, T., Arikan, F.: Ionospheric weather risk mit-
880 igation challenges in deleterious impacts on ground and space based
881 operational systems and infrastructure. In: 43rd COSPAR Scientific
882 Assembly. Held 28 January - 4 February, vol. 43, p. 655 (2021)
- 883 [64] Rao, T.V., Sridhar, M., Ratnam, D.V., Harsha, P.B.S., Srivani, I.: A
884 bidirectional long short-term memory-based ionospheric fo2 and hmf2
885 models for a single station in the low latitude region. *IEEE Geoscience*
886 *and Remote Sensing Letters*, 1–5 (2021). [https://doi.org/10.1109/LGRS.](https://doi.org/10.1109/LGRS.2020.3045702)
887 [2020.3045702](https://doi.org/10.1109/LGRS.2020.3045702)
- 888 [65] Forootan, E., Kosary, M., Farzaneh, S., Kodikara, T., Vielberg, K.,
889 Fernandez-Gomez, I., Borries, C., Schumacher, M.: Forecasting global
890 and multi-level thermospheric neutral density and ionospheric electron
891 content by tuning models against satellite-based accelerometer measure-
892 ments. *Scientific Reports* **12** (1), 2095 (2022). [https://doi.org/10.1038/](https://doi.org/10.1038/s41598-022-05952-y)
893 [s41598-022-05952-y](https://doi.org/10.1038/s41598-022-05952-y)
- 894 [66] Robinson, R., Zhang, Y., Garcia-Sage, K., Fang, X., Verkhoglyadova,
895 O.P., Ngwira, C., Bingham, S., Kosar, B., Zheng, Y., Kaeppler, S.,
896 Liemohn, M., Weygand, J.M., Crowley, G., Merkin, V., McGranaghan,

- 897 R., Mannucci, A.J.: Space weather modeling capabilities assessment:
898 Auroral precipitation and high-latitude ionospheric electrodynamic-
899 ics. *Space Weather* **17**(2), 212–215 (2019). [https://doi.org/10.1029/
900 2018SW002127](https://doi.org/10.1029/2018SW002127)
- 901 [67] Miller, K.L., Vondrak, R.R.: A high-latitude phenomenological model
902 of auroral precipitation and ionospheric effects. *Radio Science* **20**(3),
903 431–438 (1985). <https://doi.org/10.1029/RS020i003p00431>
- 904 [68] Matsuo, T., Araujo-Pradere, E.A.: Role of thermosphere-ionosphere cou-
905 pling in a global ionospheric specification. *Radio Science* **46**(06), 1–7
906 (2011). <https://doi.org/10.1029/2010RS004576>
- 907 [69] Benyassine, A., Shlomot, E., Su, H.-Y., Massaloux, D., Lamblin, C.,
908 Petit, J.-P.: Itu-t recommendation g.729 annex b: a silence compression
909 scheme for use with g.729 optimized for v.70 digital simultaneous voice
910 and data applications. *IEEE Communications Magazine* **35**(9), 64–73
911 (1997). <https://doi.org/10.1109/35.620527>
- 912 [70] Nava, B., Coisson, P., Radicella, S.: A new version of the nequick
913 ionosphere electron density model. *Journal of Atmospheric and Solar-
914 Terrestrial Physics* **70**(15), 1856–1862 (2008). [https://doi.org/10.1016/
915 j.jastp.2008.01.015](https://doi.org/10.1016/j.jastp.2008.01.015)
- 916 [71] Aragon-Angel, A., Zürn, M., Rovira-Garcia, A.: Galileo ionospheric cor-
917 rection algorithm: An optimization study of NeQuick-G. *Radio Science*
918 **54**(11), 1156–1169 (2019). <https://doi.org/10.1029/2019RS006875>
- 919 [72] Series, P.: Ionospheric propagation data and prediction methods required
920 for the design of satellite services and systems. Recommendation ITU-R,
921 531–613 (2016). [https://doi.org/https://www.itu.int/dms_pubrec/itu-r/
922 rec/p/R-REC-P.531-12-201309-S!!PDF-E.pdf](https://doi.org/https://www.itu.int/dms_pubrec/itu-r/rec/p/R-REC-P.531-12-201309-S!!PDF-E.pdf)
- 923 [73] Yuan, Y., Wang, N., Li, Z., Huo, X.: The BeiDou global broadcast
924 ionospheric delay correction model (BDGIM) and its preliminary per-
925 formance evaluation results. *NAVIGATION* **66**(1), 55–69 (2019). [https:
926 //doi.org/10.1002/navi.292](https://doi.org/10.1002/navi.292)
- 927 [74] Klobuchar, J.A.: Ionospheric time-delay algorithm for single frequency
928 GPS users. *IEEE Transactions on Aerospace and Electronic Sys-*
929 *tems* **AES-23**(3), 325–331 (1987). [https://doi.org/10.1109/TAES.1987.
930 310829](https://doi.org/10.1109/TAES.1987.310829)
- 931 [75] Bilitza, D.: International reference ionosphere 2000. *Radio Science* **36**(2),
932 261–275 (2001). <https://doi.org/10.1029/2000RS002432>
- 933 [76] Bilitza, D.: Iri the international standard for the ionosphere. *Advances in*

- 934 Radio Science **16**, 1–11 (2018). <https://doi.org/10.5194/ars-16-1-2018>
- 935 [77] Montenbruck, O., Rodríguez, B.G.: Nequick-g performance assessment
936 for space applications. *GPS Solutions* **24**(1), 1–12 (2020). <https://doi.org/10.1007/s10291-019-0931-2>
937
- 938 [78] Sanz Subirana, J., Juan Zornoza, J., Hernández-Pajares, M.: GNSS data
939 processing book, vol. i: fundamentals and algorithms. Technical report,
940 TM-23/1. Noordwijk: ESA Communications (2013)
- 941 [79] Rose, J.A., Watson, R.J., Allain, D.J., Mitchell, C.N.: Ionospheric cor-
942 rections for GPS time transfer. *Radio Science* **49**(3), 196–206 (2014).
943 <https://doi.org/10.1002/2013RS005212>
- 944 [80] Rovira-Garcia, A., Juan, J.M., Sanz, J., González-Casado, G.: A world-
945 wide ionospheric model for fast precise point positioning. *IEEE Trans-*
946 *actions on Geoscience and Remote Sensing* **53**(8), 4596–4604 (2015).
947 <https://doi.org/10.1109/TGRS.2015.2402598>
- 948 [81] Su, K., Jin, S., Hoque, M.: Evaluation of ionospheric delay effects on
949 multi-GNSS positioning performance. *Remote Sensing* **11**(2), 171 (2019).
950 <https://doi.org/10.3390/rs11020171>
- 951 [82] Zhang, J., Gao, J., Yu, B., Sheng, C., Gan, X.: Research on remote GPS
952 common-view precise time transfer based on different ionosphere distur-
953 bances. *Sensors* **20**(8), 2290 (2020). <https://doi.org/10.3390/s20082290>
- 954 [83] Hajj, G.A., Wilson, B.D., Wang, C., Pi, X., Rosen, I.G.: Data assimi-
955 lation of ground gps total electron content into a physics-based ionospheric
956 model by use of the kalman filter. *Radio Science* **39**(1) (2004). <https://doi.org/10.1029/2002RS002859>
957
- 958 [84] Bust, G.S., Garner, T.W., Gaussiran II, T.L.: Ionospheric data assimi-
959 lation three-dimensional (ida3d): A global, multisensor, electron density
960 specification algorithm. *Journal of Geophysical Research: Space Physics*
961 **109**(A11) (2004). <https://doi.org/10.1029/2003JA010234>
- 962 [85] Scherliess, L., Thompson, D.C., Schunk, R.W.: Ionospheric dynamics and
963 drivers obtained from a physics-based data assimilation model. *Radio*
964 *Science* **44**(1) (2009). <https://doi.org/10.1029/2008RS004068>
- 965 [86] Chen, P., Chen, J.: The multi-source data fusion global ionospheric mod-
966 eling software—ionogim. *Advances in Space Research* **53**(11), 1610–1622
967 (2014). <https://doi.org/10.1016/j.asr.2014.02.025>
- 968 [87] Chiang, K.Q., Psiaki, M.L.: Gps and ionosonde data fusion for iono-
969 spheric tomography. In: *Proceedings of the 27th International Technical*

- 970 Meeting of the Satellite Division of The Institute of Navigation (ION
971 GNSS+ 2014), pp. 1163–1172 (2014)
- 972 [88] Matsuo, T.: Upper atmosphere data assimilation with an ensemble
973 Kalman filter. Modeling the Ionosphere–Thermosphere System (eds J.
974 Huba, R. Schunk and G. Khazanov), 273–282 (2014). [https://doi.org/
975 10.1002/9781118704417.ch22](https://doi.org/10.1002/9781118704417.ch22)
- 976 [89] Chartier, A.T., Matsuo, T., Anderson, J.L., Collins, N., Hoar, T.J., Lu,
977 G., Mitchell, C.N., Coster, A.J., Paxton, L.J., Bust, G.S.: Ionospheric
978 data assimilation and forecasting during storms. *Journal of Geophys-
979 ical Research: Space Physics* **121**(1), 764–778 (2016). [https://doi.org/10.
980 1002/2014JA020799](https://doi.org/10.1002/2014JA020799)
- 981 [90] Pilinski, M.D., Crowley, G., Sutton, E., Codrescu, M.: Improved orbit
982 determination and forecasts with an assimilative tool for satellite drag
983 specification. In: *Advanced Maui Optical and Space Surveillance Tech-
984 nologies Conference*, vol. 104 (2016). [https://doi.org/2016amos.confE.
985 104P](https://doi.org/2016amos.confE.104P)
- 986 [91] Codrescu, S., Codrescu, M., Fedrizzi, M.: An ensemble Kalman filter
987 for the thermosphere-ionosphere. *Space Weather* **16**(1), 57–68 (2018).
988 <https://doi.org/10.1002/2017SW001752>
- 989 [92] Forsythe, V.V., Azeem, I., Crowley, G.: Ionospheric horizontal corre-
990 lation distances: Estimation, analysis, and implications for ionospheric
991 data assimilation. *Radio Science* **55**(12), 2020–007159 (2020). [https:
992 //doi.org/10.1029/2020RS007159](https://doi.org/10.1029/2020RS007159)
- 993 [93] Forootan, E., Farzaneh, S., Kosary, M., Schmidt, M., Schumacher, M.:
994 A simultaneous calibration and data assimilation (C/DA) to improve
995 NRLMSISE00 using thermospheric neutral density (TND) from space-
996 borne accelerometer measurements. *Geophysical Journal International*
997 **224**(2), 1096–1115 (2020). <https://doi.org/10.1093/gji/ggaa507>
- 998 [94] Kosary, M., Forootan, E., Farzaneh, S., Schumacher, M.: A sequential
999 calibration approach based on the ensemble kalman filter (c-enkf) for
1000 forecasting total electron content (tec). *Journal of Geodesy* **96**(4), 1–26
1001 (2022). <https://doi.org/10.1007/s00190-022-01623-y>
- 1002 [95] Epstein, E.S.: Stochastic dynamic prediction. *Tellus* **21**(6), 739–759
1003 (1969). <https://doi.org/10.3402/tellusa.v21i6.10143>
- 1004 [96] Metropolis, N., S.Ulam: The monte carlo method. *Journal of the Ameri-
1005 can Statistical Association* **44**(247), 335–341 (1949). [https://doi.org/10.
1006 1080/01621459.1949.10483310](https://doi.org/10.1080/01621459.1949.10483310)

- 1007 [97] Luo, X., Bhakta, T., Jakobsen, M., Nævdal, G.: Efficient big data assim-
1008 ilation through sparse representation: A 3d benchmark case study in
1009 petroleum engineering. *PLOS ONE* **13**(7), 1–32 (2018). [https://doi.org/](https://doi.org/10.1371/journal.pone.0198586)
1010 [10.1371/journal.pone.0198586](https://doi.org/10.1371/journal.pone.0198586)
- 1011 [98] Avasarala, S., Subramani, D.: A non-gaussian bayesian filter for sequen-
1012 tial data assimilation with non-intrusive polynomial chaos expansion.
1013 *International Journal for Numerical Methods in Engineering* **122**(23),
1014 7156–7181 (2021). <https://doi.org/10.1002/nme.6827>
- 1015 [99] Hoang, T.-V., Krumscheid, S., Matthies, H.G., Tempone, R.: Machine
1016 learning-based conditional mean filter: A generalization of the ensem-
1017 ble kalman filter for nonlinear data assimilation. *Foundations of Data*
1018 *Science* **5**(1), 56–80 (2023). <https://doi.org/10.3934/fods.2022016>
- 1019 [100] Evensen, G.: Sampling strategies and square root analysis schemes for
1020 the enfk. *Ocean dynamics* **54**(6), 539–560 (2004). [https://doi.org/10.](https://doi.org/10.1007/s10236-004-0099-2)
1021 [1007/s10236-004-0099-2](https://doi.org/10.1007/s10236-004-0099-2)
- 1022 [101] Tuan Pham, D., Verron, J., Christine Roubaud, M.: A singular evo-
1023 lutive extended kalman filter for data assimilation in oceanography.
1024 *Journal of Marine Systems* **16**(3), 323–340 (1998). [https://doi.org/10.](https://doi.org/10.1016/S0924-7963(97)00109-7)
1025 [1016/S0924-7963\(97\)00109-7](https://doi.org/10.1016/S0924-7963(97)00109-7)
- 1026 [102] Bishop, C.H., Etherton, B.J., Majumdar, S.J.: Adaptive sampling with
1027 the ensemble transform kalman filter. part i: Theoretical aspects.
1028 *Monthly weather review* **129**(3), 420–436 (2001). [https://doi.org/10.](https://doi.org/10.1175/1520-0493(2001)129(0420:ASWTET)2.0.CO;2)
1029 [1175/1520-0493\(2001\)129\(0420:ASWTET\)2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129(0420:ASWTET)2.0.CO;2)
- 1030 [103] Cao, Y., Zhu, J., Navon, I.M., Luo, Z.: A reduced-order approach to
1031 four-dimensional variational data assimilation using proper orthogonal
1032 decomposition. *International Journal for Numerical Methods in Fluids*
1033 **53**(10), 1571–1583 (2007). <https://doi.org/10.1002/fld.1365>
- 1034 [104] Rozier, D., Birol, F., Cosme, E., Brasseur, P., Brankart, J.M., Verron, J.:
1035 A reduced-order kalman filter for data assimilation in physical oceanog-
1036 raphy. *SIAM Review* **49**(3), 449–465 (2007). [https://doi.org/10.1137/](https://doi.org/10.1137/050635717)
1037 [050635717](https://doi.org/10.1137/050635717)
- 1038 [105] Schumacher, M., Kusche, J., Döll, P.: A systematic impact assessment
1039 of grace error correlation on data assimilation in hydrological mod-
1040 els. *Journal of Geodesy* **90**(6), 537–559 (2016). [https://doi.org/10.1007/](https://doi.org/10.1007/s00190-016-0892-y)
1041 [s00190-016-0892-y](https://doi.org/10.1007/s00190-016-0892-y)
- 1042 [106] Meldi, M., Poux, A.: A reduced order model based on kalman filtering for
1043 sequential data assimilation of turbulent flows. *Journal of Computational*
1044 *Physics* **347**, 207–234 (2017). <https://doi.org/10.1016/j.jcp.2017.06.042>

- 1045 [107] Khaki, M., Hoteit, I., Kuhn, M., Awange, J., Forootan, E., van Dijk,
1046 A.I.J.M., Schumacher, M., Pattiaratchi, C.: Assessing sequential data
1047 assimilation techniques for integrating grace data into a hydrological
1048 model. *Advances in Water Resources* **107**, 301–316 (2017). [https://doi.
1049 org/10.1016/j.advwatres.2017.07.001](https://doi.org/10.1016/j.advwatres.2017.07.001)
- 1050 [108] Xiao, D., Du, J., Fang, F., Pain, C.C., Li, J.: Parameterised non-intrusive
1051 reduced order methods for ensemble kalman filter data assimilation.
1052 *Computers and Fluids* **177**, 69–77 (2018). [https://doi.org/10.1016/j.
1053 compfluid.2018.10.006](https://doi.org/10.1016/j.compfluid.2018.10.006)
- 1054 [109] Zerfas, C., Rebholz, L.G., Schneier, M., Iliescu, T.: Continuous data
1055 assimilation reduced order models of fluid flow. *Computer Methods in
1056 Applied Mechanics and Engineering* **357**, 112596 (2019). [https://doi.
1057 org/10.1016/j.cma.2019.112596](https://doi.org/10.1016/j.cma.2019.112596)
- 1058 [110] Casas, C.Q., Arcucci, R., Wu, P., Pain, C., Guo, Y.-K.: A reduced order
1059 deep data assimilation model. *Physica D: Nonlinear Phenomena* **412**,
1060 132615 (2020). <https://doi.org/10.1016/j.physd.2020.132615>
- 1061 [111] Matsuo, T., Richmond, A.D., Nychka, D.W.: Modes of the High-Latitude
1062 Electric Field Variability Derived From DE-2 Measurements: Empirical
1063 Orthogonal Function (EOF) Analysis. In: *AGU Fall Meeting Abstracts*,
1064 vol. 2001, pp. 32–0689 (2001)
- 1065 [112] Matsuo, T., Richmond, A.D., Lu, G.: Optimal interpolation analysis
1066 of high-latitude ionospheric electrodynamics using empirical orthogonal
1067 functions: Estimation of dominant modes of variability and temporal
1068 scales of large-scale electric fields. *Journal of Geophysical Research:
1069 Space Physics* **110**(A6) (2005). <https://doi.org/10.1029/2004JA010531>
- 1070 [113] Collard, A.D., McNally, A.P., Hilton, F.I., Healy, S.B., Atkinson, N.C.:
1071 The use of principal component analysis for the assimilation of high-
1072 resolution infrared sounder observations for numerical weather predic-
1073 tion. *Quarterly Journal of the Royal Meteorological Society* **136**(653),
1074 2038–2050 (2010). <https://doi.org/10.1002/qj.701>
- 1075 [114] Matsuo, T., Fedrizzi, M., Fuller-Rowell, T.J., Codrescu, M.V.: Data
1076 assimilation of thermospheric mass density. *Space Weather* **10**(5) (2012).
1077 <https://doi.org/10.1029/2012SW000773>
- 1078 [115] Matricardi, M., McNally, A.P.: The direct assimilation of principal com-
1079 ponents of iasi spectra in the ecmwf 4d-var. *Quarterly Journal of the
1080 Royal Meteorological Society* **140**(679), 573–582 (2014). [https://doi.org/
1081 10.1002/qj.2156](https://doi.org/10.1002/qj.2156)
- 1082 [116] Lu, Y., Zhang, F.: Toward ensemble assimilation of hyperspectral

- 1083 satellite observations with data compression and dimension reduction
1084 using principal component analysis. *Monthly Weather Review* **147**(10),
1085 3505–3518 (2019). <https://doi.org/10.1175/MWR-D-18-0454.1>
- 1086 [117] Jolliffe, I.: *Principal Component Analysis*. John Wiley and Sons,
1087 Ltd, ??? (2005). <https://doi.org/10.1002/0470013192.bsa501>. [https://](https://onlinelibrary.wiley.com/doi/abs/10.1002/0470013192.bsa501)
1088 onlinelibrary.wiley.com/doi/abs/10.1002/0470013192.bsa501
- 1089 [118] Forootan, E., Kusche, J.: Separation of global time-variable gravity sig-
1090 nals into maximally independent components. *Journal of Geodesy* **86**(7),
1091 477–497 (2012). <https://doi.org/10.1007/s00190-011-0532-5>
- 1092 [119] Forootan, E.: *Statistical signal decomposition techniques for analyz-*
1093 *ing time-variable satellite gravimetry data*. PhD thesis, University
1094 of Bonn, [https://bonndoc.ulb.uni-bonn.de/xmlui/handle/20.500.11811/](https://bonndoc.ulb.uni-bonn.de/xmlui/handle/20.500.11811/1452)
1095 [1452](https://bonndoc.ulb.uni-bonn.de/xmlui/handle/20.500.11811/1452) (2014)
- 1096 [120] Forootan, E., Kusche, J., Talpe, M., Shum, C., Schmidt, M.: Developing
1097 a complex independent component analysis (cica) technique to extract
1098 non-stationary patterns from geophysical time series. *Surveys in Geo-*
1099 *physics* **39**, 435–465 (2018). <https://doi.org/10.1007/s10712-017-9451-1>
- 1100 [121] Hernández-Pajares, M., Juan, J., Sanz, J., Orus, R., Garcia-Rigo,
1101 A., Feltens, J., Komjathy, A., Schaer, S., Krankowski, A.: The IGS
1102 VTEC maps: a reliable source of ionospheric information since 1998.
1103 *Journal of Geodesy* **83**(3-4), 263–275 (2009). [https://doi.org/10.1007/](https://doi.org/10.1007/s00190-008-0266-1)
1104 [s00190-008-0266-1](https://doi.org/10.1007/s00190-008-0266-1)
- 1105 [122] Goss, A., Schmidt, M., Erdogan, E., Seitz, F.: Global and regional high-
1106 resolution vtec modelling using a two-step b-spline approach. *Remote*
1107 *Sensing* **12**(7) (2020). <https://doi.org/10.3390/rs12071198>
- 1108 [123] Liu, L., Zou, S., Yao, Y., Wang, Z.: Forecasting global ionospheric
1109 tec using deep learning approach. *Space Weather* **18**(11), 2020–002501
1110 (2020). <https://doi.org/10.1029/2020SW002501>
- 1111 [124] Forsythe, V.V., Azeem, I., Crowley, G.: Ionospheric horizontal corre-
1112 lation distances: Estimation, analysis, and implications for ionospheric
1113 data assimilation. *Radio Science* **55**(12), 2020–007159 (2020). [https://](https://doi.org/10.1029/2020RS007159)
1114 doi.org/10.1029/2020RS007159
- 1115 [125] Forsythe, V.V., Azeem, I., Blay, R., Crowley, G., Gasperini, F., Hughes,
1116 J., Makarevich, R.A., Wu, W.: Evaluation of the new background covari-
1117 ance model for the ionospheric data assimilation. *Radio Science* **56**(8),
1118 2021–007286 (2021). <https://doi.org/10.1029/2021RS007286>
- 1119 [126] Qiao, J., Liu, Y., Fan, Z., Tang, Q., Li, X., Zhang, F., Song, Y., He, F.,

- 1120 Zhou, C., Qing, H., Li, Z.: Ionospheric tec data assimilation based on
 1121 gauss–markov kalman filter. *Advances in Space Research* **68**(10), 4189–
 1122 4204 (2021). <https://doi.org/10.1016/j.asr.2021.08.004>
- 1123 [127] Feltens, J., Schaer, S.: IGS products for the ionosphere. In: Proceedings
 1124 of the 1998 IGS Analysis Center Workshop Darmstadt, Germany, pp.
 1125 3–5 (1998)
- 1126 [128] Orús, R., Hernández-Pajares, M., Juan, J.M., Sanz, J.: Improvement of
 1127 global ionospheric VTEC maps by using kriging interpolation technique.
 1128 *Journal of Atmospheric and Solar-Terrestrial Physics* **67**(16), 1598–1609
 1129 (2005). <https://doi.org/10.1016/j.jastp.2005.07.017>
- 1130 [129] Schaer, S., helvétique des sciences naturelles. Commission géodésique,
 1131 S.: Mapping and Predicting the Earth’s Ionosphere Using the Global
 1132 Positioning System vol. 59. Institut für Geodäsie und Photogrammetrie,
 1133 Eid. Technische Hochschule . . . , ??? (1999)
- 1134 [130] Gonzalez, W.D., Tsurutani, B.T., De Gonzalez, A.L.C.: Interplanetary
 1135 origin of geomagnetic storms. *Space Science Reviews* **88**(3), 529–562
 1136 (1999). <https://doi.org/10.1023/A:1005160129098>
- 1137 [131] Di Giovanni, G., Radicella, S.M.: An analytical model of the electron
 1138 density profile in the ionosphere. *Advances in Space Research* **10**(11),
 1139 27–30 (1990). [https://doi.org/10.1016/0273-1177\(90\)90301-F](https://doi.org/10.1016/0273-1177(90)90301-F)
- 1140 [132] Radicella, S.M., Zhang, M.L.: The improved DGR analytical model
 1141 of electron density height profile and total electron content in the
 1142 ionosphere. <http://hdl.handle.net/2122/1743>
- 1143 [133] Kodikara, T.: Physical understanding and forecasting of the thermo-
 1144 spheric structure and dynamics. PhD thesis, RMIT University (2019)
- 1145 [134] Richards, P., Fennelly, J., Torr, D.: EUVAC: A solar EUV flux model for
 1146 aeronomic calculations. *Journal of Geophysical Research: Space Physics*
 1147 **99**(A5), 8981–8992 (1994). <https://doi.org/10.1029/94JA00518>
- 1148 [135] Solomon, S.C., Qian, L.: Solar extreme-ultraviolet irradiance for gen-
 1149 eral circulation models. *Journal of Geophysical Research: Space Physics*
 1150 **110**(A10) (2005). <https://doi.org/10.1029/2005JA011160>
- 1151 [136] Webb, D.F., Howard, R.A.: The solar cycle variation of coronal
 1152 mass ejections and the solar wind mass flux. *Journal of Geophysical*
 1153 *Research: Space Physics* **99**(A3), 4201–4220 (1994). [https://doi.org/10.](https://doi.org/10.1029/93JA02742)
 1154 [1029/93JA02742](https://doi.org/10.1029/93JA02742)
- 1155 [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., Kutsche, 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>

1193 Appendix - Evaluation measures

1194 To numerically evaluate the performance of the original and DDA model
1195 compared to the observation, the following statistical measures are applied:

1196 • ‘Bias’ is defined as:

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

1197 where Obs and Model denote observation and model estimates, respectively,
1198 and n is the number of observations. The positive (negative) values of the
1199 bias demonstrate that the model underestimates (overestimates) compared
1200 to the observations.

1201 • The expression of bias in percentage is computed based on the ‘Relative
1202 Error (RE)’ as:

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

1203 where $|\cdot|$ 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:

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

1204 • ‘Root Mean Squared of Error (RMSE)’ is determined to assess how model
1205 estimates match with observations as:

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

1206 The square term inside the RMSE equation highlights both positive and
1207 negative differences between the quantities.

1208 • ‘Improvement’ is defined as percentage in the computed RMSEs after
1209 implementing DDA as:

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

1210 where RMSE_1 is computed between the original NeQuick or TIEGCM and
1211 GIM-VTECs, and RMSE_2 is determined between those of DDA and GIM-
1212 VTECs.

1213 • ‘Average of Absolute Percentage Deviation (AAPD)’ is expressed as the
1214 percentage of absolute difference between observation and model as:

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

1215 Minimum (maximum) values of AAPD correspond to the average best
1216 (worst) performance of a model in estimating VTECs.

- 1217 • ‘Fit’ is determined as the fraction of data variance that is predicted by the
 1218 model as:

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

1219 where $\bar{\text{Obs}}$ is defined as the mean of observations. In contrast to AAPD,
 1220 the minimum (maximum) values of fitting correspond to the average worst
 1221 (best) performance of model in simulating VTECs.

- 1222 • ‘Correlation Coefficients (CCs)’ are used as a unit-less measure to represent
 1223 the overall agreement between model estimations and observations:

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

1224 The range of CCs is from -1 to $+1$, where -1 indicates the perfect negative
 1225 correlation, $+1$ corresponds to the 100% correspondence, and zero indicates
 1226 no correlations.