

# Using the local ensemble transform Kalman filter for upper atmospheric modelling

Elvidge, Sean; Angling, Matthew

DOI:

[10.1051/swsc/2019018](https://doi.org/10.1051/swsc/2019018)

License:

Creative Commons: Attribution (CC BY)

*Document Version*

Publisher's PDF, also known as Version of record

*Citation for published version (Harvard):*

Elvidge, S & Angling, M 2019, 'Using the local ensemble transform Kalman filter for upper atmospheric modelling', *Journal of Space Weather and Space Climate*, vol. 9, no. 2019, A30.  
<https://doi.org/10.1051/swsc/2019018>

[Link to publication on Research at Birmingham portal](#)

## General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

## Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.

# Using the local ensemble Transform Kalman Filter for upper atmospheric modelling

Sean Elvidge\*, and Matthew J. Angling

Space Environment and Radio Engineering, University of Birmingham, Birmingham B15 2TT, UK

Received 19 June 2018 / Accepted 19 April 2019

**Abstract**—The Advanced Ensemble electron density (Ne) Assimilation System (AENeAS) is a new data assimilation model of the ionosphere/thermosphere. The background model is provided by the Thermosphere Ionosphere Electrodynamics General Circulation Model (TIE-GCM) and the assimilation uses the local ensemble transform Kalman filter (LETKF). An outline derivation of the LETKF is provided and the equations are presented in a form analogous to the classic Kalman filter. An enhancement to the efficient LETKF implementation to reduce computational cost is also described. In a 3 day test in June 2017, AENeAS exhibits a total electron content (TEC) RMS error of 2.1 TECU compared with 5.5 TECU for NeQuick and 6.8 for TIE-GCM (with an NeQuick topside).

**Keywords:** data assimilation / ensemble Kalman filter / LETKF / ionosphere / thermosphere

## 1 Introduction

### 1.1 Background

Comprehensive, global and timely specifications of the Earth's upper atmosphere are required to ensure the effective operation, planning and management of both radio frequency (RF) systems and satellites. The atmospheric layers most effected by space weather are the ionosphere and thermosphere. The ionosphere/thermosphere is a coupled system which spans an altitude of approximately 90–2000 km above the Earth's surface. In this region, radiation from the Sun causes photoionization of neutral molecules (thermosphere) creating a plasma (ions and free electrons, ionosphere). The thermosphere causes drag on satellites and is the dominant error in propagating low Earth orbits (below approximately 600 km). Furthermore, the ionosphere affects all RF systems that operate via or through the ionosphere at frequencies below approximately 2 GHz. Thus, systems such as global navigation satellite systems (GNSS), high frequency (HF) communications, space-based Earth observation radars and space situational awareness radars can be impaired.

### 1.2 Modelling

Approaches for modelling the Earth's upper atmosphere (ionosphere/thermosphere) can be broadly split into three categories: empirical modelling, physics-based modelling and data

assimilation techniques. Empirical models are based upon observations and the relationships between model variables are usually found via curve fitting techniques (Llewellyn & Bent, 1973; Bilitza et al., 1988). Physics-based models solve the underlying equations of the physical processes in the atmosphere. Often these are in the form of coupled ionospheric-thermospheric physics models. Finally, data assimilation models combine measurements of the atmosphere with a background model.

Data assimilation (DA) is a broad term covering a number of techniques. These include: weighted least squares (Plackett, 1950), Kalman filters (Kalman, 1960), optimal interpolation (OI) (Gandin, 1963; Eddy, 1967) and variational methods (Le Dimet & Talagrand, 1986). The background model for the DA approach can be either an empirical or physics-based model.

Previous model comparison work has shown that, in terms of specifying the current ionosphere (nowcasting), DA models do best, followed by empirical and then physics-based models (Shim et al., 2011, 2012; McNamara et al., 2013; Elvidge et al., 2014). In particular, McNamara et al. (2013) showed that two DA models (both with an empirical background model (the Electron Density Assimilative Model, EDAM; Angling & Jackson-Booth, 2011) and the GPS Ionospheric Inversion model, GPSII; Fridman et al., 2009) perform close to the minimum achievable errors for nowcasting parameters related to the peak ionospheric density.

Since both EDAM and GPSII rely on an empirical background model (the International Reference Ionosphere 2007 (IRI2007) (Bilitza & Reinisch, 2008), in the absence of data

\*Corresponding author: [s.elvidge@bham.ac.uk](mailto:s.elvidge@bham.ac.uk)

they revert back to this model. This is a weakness suffered by all empirically-based data assimilation models and means that they are not well suited to forecasting. This weakness can be overcome by using physics-based DA models, which, when forecasting, can propagate the ionosphere/thermosphere state forward in time.

An example of a physics-based DA model is the Utah State University Global Assimilation of Ionospheric Measurements (USU-GAIM) models. These models encompass a range of different modelling approaches (Schunk et al., 2016). In particular the GAIM Gauss-Markov (GAIM-GM) (Scherliess et al., 2006) and GAIM Full Physics (GAIM-FP) (Scherliess et al., 2009) models both use physics-based background models, and whilst GAIM-GM uses a Gauss Markov Kalman filter for assimilation, GAIM-FP relies upon an ensemble Kalman filter. The background model for GAIM-FP includes six ion species ( $\text{NO}^+$ ,  $\text{O}_2^+$ ,  $\text{N}_2^+$ ,  $\text{O}^+$ ,  $\text{He}^+$ ,  $\text{H}^+$ ), as well as ion and electron temperatures and plasma drifts (Schunk et al., 2005). However in this model the thermosphere is modelled empirically and this limits the model's ability to model satellite drag.

Coupled models such as the Thermosphere Ionosphere Electrodynamics General Circulation Model (TIE-GCM) (Qian et al., 2014) have also been used in DA schemes. TIE-GCM was developed at the National Center for Atmospheric Research (NCAR) and solves for neutral and ion species in the upper atmosphere. Using TIE-GCM in an ensemble Kalman filter modelling scheme has been shown to increase skill in specifying ionospheric parameters (Lee et al., 2012; Chartier et al., 2013, 2016).

This paper introduces the Advanced Ensemble electron density (Ne) Assimilation System (AENeAS). AENeAS uses TIE-GCM as its background model and uses a variant of the ensemble Kalman filter (the local ensemble transform Kalman filter, LETKF) to assimilate data. As well as describing the theory behind the model, initial benchmark results are also described.

## 2 Kalman Filters

### 2.1 Introduction

Kalman filters (Kalman, 1960) have found widespread use in applications ranging from signal processing to spacecraft navigation and control. They have also been used in a variety of assimilative ionospheric models. For a full derivation of the classic Kalman filter equations see Hamilton (1994).

Given a series of states at times  $t$ ,  $\mathbf{x}^t$ , the Kalman filter gives the optimal (i.e. the minimum mean square error) estimate of the combination (analysis,  $\mathbf{x}_a^t$ ) of a background model ( $\mathbf{x}_b^t$ ) and observations  $\mathbf{y}_o^t$ . This optimality is dependent on the process being linear and error statistics being Gaussian. The filter works in two repeating steps:

1. Update;
2. Prediction.

#### Update step

The filter starts with a background state vector and its associated error covariance matrix ( $\mathbf{x}_b^t$  and  $\mathbf{B}^t$ ). The analysis (i.e. the

updated state vector,  $\mathbf{x}_a^t$ ) and its associated covariance matrix ( $\mathbf{A}^t$ ) are given by

$$\mathbf{A}^t = (\mathbf{I} - \mathbf{K}^t \mathbf{H}) \mathbf{B}^t, \quad (1)$$

$$\mathbf{x}_a^t = \mathbf{x}_b^t + \mathbf{K}^t (\mathbf{y}_o^t - \mathbf{H} \mathbf{x}_b^t), \quad (2)$$

where

$$\mathbf{K}^t = \mathbf{B}^t \mathbf{H}^T (\mathbf{H} \mathbf{B}^t \mathbf{H}^T + \mathbf{O})^{-1}. \quad (3)$$

$\mathbf{K}$  is commonly called the Kalman gain matrix,  $\mathbf{B}$  is the background error covariance matrix,  $\mathbf{O}$  is the observation error covariance matrix,  $\mathbf{H}$  is the observation operator,  $\mathbf{x}_b$  and  $\mathbf{x}_a$  are the background and analysis vectors respectively,  $\mathbf{y}_o$  is the vector of observations and  $\mathbf{I}$  is the identity matrix.

#### Prediction step

Using the values from the current update step, the prediction step is used to find the background values  $\mathbf{x}_b^{t+1}$  and  $\mathbf{B}^{t+1}$  for use in the next update step:

$$\mathbf{x}_b^{t+1} = \mathbf{M}^{t,t+1} \mathbf{x}_a^t, \quad (4)$$

$$\mathbf{B}^{t+1} = \mathbf{M}^{t,t+1} \mathbf{A} (\mathbf{M}^{t,t+1})^T + \mathbf{Q}^t, \quad (5)$$

where  $\mathbf{M}$  is the forecast model which moves the state from one time step to the next and  $\mathbf{Q}^t$  is the associated error covariance matrix. This process is repeated  $t$  times.

The classic Kalman filter (Kalman, 1960) specifically defines a covariance matrix for the system (Eq. (1)), and consequently the computational cost of the filter scales with the square of the dimension of the model. For a fully coupled ionosphere/thermosphere model the dimension is approximately  $10^7$ . This is too large for most computers to be able to manipulate in a reasonable amount of time. The ensemble Kalman Filter (EnKF) (Evensen, 2009) is a Monte Carlo approximation to the Kalman filter for high dimensional or non-linear models. Since its introduction it has gained wide spread use throughout a variety of data assimilation applications such as weather forecasting (Etherton, 2007), fireline modelling (Mandel et al., 2009), power system state tracking (Li et al., 2012) and soil moisture estimation (Reichle et al., 2002). The EnKF replaces the covariance matrix (Eq. (1)) by a sample covariance found from a set of state vectors (called the ensemble) (Evensen, 1994). The ensemble members are independent until data is assimilated, at which point all of the members are modified. This approach is useful for both large dimensional systems (where estimating the covariance matrix is necessary to reduce the computational cost) and also in complex physical systems where the exact covariances between model variables are unknown.

There is no unique, optimal, method for implementing the EnKF equations. Previous EnKF work using TIE-GCM as a background model (e.g. Lee et al., 2012; Chartier et al., 2013, 2016), has used the NCAR's Data Assimilation Research Testbed (DART) (Anderson, 2009). DART provides a set of tools for applying DA techniques to existing models across a range of geophysics disciplines. As well as the EnKF, DART includes the Ensemble Adjustment Kalman Filter (EAKF) (Anderson, 2001) and the Rank Histogram filter (Anderson, 2010; Metref et al., 2014).

## 2.2 Local ensemble transform Kalman filter

In this work, the local ensemble transform Kalman filter (LETKF) is used as the DA scheme. An outline of the LETKF equations are presented here; for a complete derivation see [Hunt et al. \(2007\)](#).

As with the classic Kalman filter, background values and an associated error covariance matrix are required to start the process (we drop the time subscript  $t$  from the variables for ease of reading). The single background state vector used in the classic Kalman filter is replaced by a set of  $k$  background ensemble members  $\{\mathbf{x}_{b(i)}; i = 1, 2, \dots, k\}$  where each  $\mathbf{x}_{b(i)}$  contains  $m$  variables. It is assumed that the best estimate of the true state (before observations are taken into account) is the background mean:

$$\bar{\mathbf{x}}_b = k^{-1} \sum_{i=1}^k \mathbf{x}_{b(i)}. \quad (6)$$

The associated error covariance matrix  $\mathbf{B}$  can then be found in the following way. Define a matrix  $\mathbf{X}_b$  to be an  $(m \times k)$  matrix of background perturbations; that is, a matrix where the  $i$ th column is defined by  $\mathbf{x}_{b(i)} - \bar{\mathbf{x}}_b$  (i.e. each column of  $\mathbf{X}_b$  is an ensemble member with the ensemble mean removed). The background error covariance matrix is then estimated by

$$\mathbf{B} = (k - 1)^{-1} \mathbf{X}_b (\mathbf{X}_b)^T. \quad (7)$$

This has at most rank  $k - 1$ , and only accounts for uncertainty in  $k - 1$  directions of the model space. This can cause issues if  $k$  is small; however this can be somewhat overcome by localizing the model space (described later in this section). After an assimilation time step an analysis ensemble ( $\{\mathbf{x}_{a(i)}; i = 1, 2, \dots, k\}$ ) is returned. The analysis mean and error covariance matrix are:

$$\bar{\mathbf{x}}_a = k^{-1} \sum_{i=1}^k \mathbf{x}_{a(i)}, \quad (8)$$

$$\mathbf{A} = (k - 1)^{-1} \mathbf{X}_a (\mathbf{X}_a)^T, \quad (9)$$

where  $\mathbf{X}_a$  is the matrix of analysis perturbations (defined equivalently as  $\mathbf{X}_a$ ). It is required that  $\bar{\mathbf{x}}_a$  and  $\mathbf{A}$  should estimate the best analysis of the system and its associated uncertainty. Therefore to construct the analysis (post assimilation) ensemble  $\mathbf{x}_{a(i)}$  the matrix  $\mathbf{X}_a$  must satisfy Equation (9). Let  $\mathbf{y}_o \in \mathbb{R}^s$  be a vector of  $s$  observations,  $\mathbf{O} \in \mathbb{R}^{s \times s}$  be the observation error covariance matrix and  $\mathbf{H} : \mathbb{R}^m \rightarrow \mathbb{R}^s$  be an observation operator that maps from the model space to the observation space. The Kalman gain for the LETKF is found by substituting the definition of the background error covariance (Eq. (7)) into the classic Kalman gain matrix definition (Eq. (3)):

$$\mathbf{K} = (k - 1)^{-1} \mathbf{X}_b (\mathbf{H} \mathbf{X}_b)^T [(k - 1)^{-1} \mathbf{H} \mathbf{X}_b (\mathbf{H} \mathbf{X}_b)^T + \mathbf{O}]^{-1}. \quad (10)$$

The matrix  $\mathbf{H} \mathbf{X}_b$  is the background perturbation matrix mapped into observation space. For simplicity this is rewritten as  $\mathbf{Y}_b$ , defined similarly to  $\mathbf{X}_b$ , i.e., the  $i$ th column is  $\mathbf{H}(\mathbf{x}_{b(i)} - \bar{\mathbf{y}}_b)$ . Using this, the identity  $\mathbf{P}^T (\mathbf{P} \mathbf{P}^T + \mathbf{Q})^{-1} =$

$(\mathbf{I} + \mathbf{P}^T \mathbf{Q} \mathbf{P})^{-1} \mathbf{Q}^T \mathbf{P}^{-1}$ , and setting  $\mathbf{P} = (k - 1)^{\frac{1}{2}} \mathbf{Y}_b$  and  $\mathbf{Q} = \mathbf{O}$ , Equation (10) can be rewritten as:

$$\mathbf{K} = (k - 1)^{-1} \mathbf{X}_b [\mathbf{I} + (k - 1)^{-1} (\mathbf{Y}_b)^T \mathbf{O}^{-1} \mathbf{Y}_b]^{-1} (\mathbf{Y}_b)^T \mathbf{O}^{-1}. \quad (11)$$

To find the analysis covariance matrix the LETKF Kalman gain and background error covariance matrix is substituted into Equation (1):

$$\mathbf{A} = (\mathbf{I} - \mathbf{K} \mathbf{H}) \mathbf{B} (k - 1)^{-1} \mathbf{X}_b (\mathbf{I} - (k - 1)^{-1} \times [\mathbf{I} + (k - 1)^{-1} (\mathbf{Y}_b)^T \mathbf{O}^{-1} \mathbf{Y}_b]^{-1} (\mathbf{Y}_b)^T \mathbf{O}^{-1} \mathbf{Y}_b) (\mathbf{X}_b)^T \quad (12)$$

where again  $\mathbf{Y}_b = \mathbf{H} \mathbf{X}_b$ . This can be simplified using the identity  $\mathbf{I} - (\mathbf{I} + \mathbf{P})^{-1} \mathbf{P} = (\mathbf{I} + \mathbf{P})^{-1}$  where  $\mathbf{P} = (k - 1)^{-1} (\mathbf{Y}_b)^T \mathbf{O}^{-1} \mathbf{Y}_b$ :

$$\mathbf{A} = \mathbf{X}_b [(k - 1) \mathbf{I} + (\mathbf{Y}_b)^T \mathbf{O}^{-1} \mathbf{Y}_b]^{-1} (\mathbf{X}_b)^T, \quad (13)$$

the analysis error covariance matrix. For ease of writing, define

$$\tilde{\mathbf{A}} = [(k - 1) \mathbf{I} + (\mathbf{Y}_b)^T \mathbf{O}^{-1} \mathbf{Y}_b]^{-1}, \quad (14)$$

so Equation (13) becomes:

$$\mathbf{A} = \mathbf{X}_b \tilde{\mathbf{A}} (\mathbf{X}_b)^T. \quad (15)$$

Using  $\tilde{\mathbf{A}}$  in the Kalman gain definition for the LETKF (Eq. (11)) gives:

$$\mathbf{K} = \mathbf{X}_b [(k - 1) \mathbf{I} + (\mathbf{Y}_b)^T \mathbf{O}^{-1} \mathbf{Y}_b]^{-1} (\mathbf{Y}_b)^T \mathbf{O}^{-1} \mathbf{X}_b \tilde{\mathbf{A}} (\mathbf{Y}_b)^T \mathbf{O}^{-1}, \quad (16)$$

and using Equations (15) and (16) one can define the ensemble analysis mean (Eq. (2)) as:

$$\bar{\mathbf{x}}_a = \bar{\mathbf{x}}_b + \mathbf{K} (\mathbf{y}_o - \mathbf{H} \bar{\mathbf{x}}_b) \bar{\mathbf{x}}_b + \mathbf{X}_b \tilde{\mathbf{A}} (\mathbf{Y}_b)^T \mathbf{O}^{-1} (\mathbf{y}_o - \bar{\mathbf{y}}_b). \quad (17)$$

Let  $\mathbf{w}_a = \tilde{\mathbf{A}} (\mathbf{Y}_b)^T \mathbf{O}^{-1} (\mathbf{y}_o - \bar{\mathbf{y}}_b)$ , which defines the analysis increment in observation space, and thus:

$$\bar{\mathbf{x}}_a = \bar{\mathbf{x}}_b + \mathbf{X}_b \mathbf{w}_a. \quad (18)$$

Letting

$$\mathbf{W}_a = [(k - 1) \tilde{\mathbf{A}}]^{-\frac{1}{2}}, \quad (19)$$

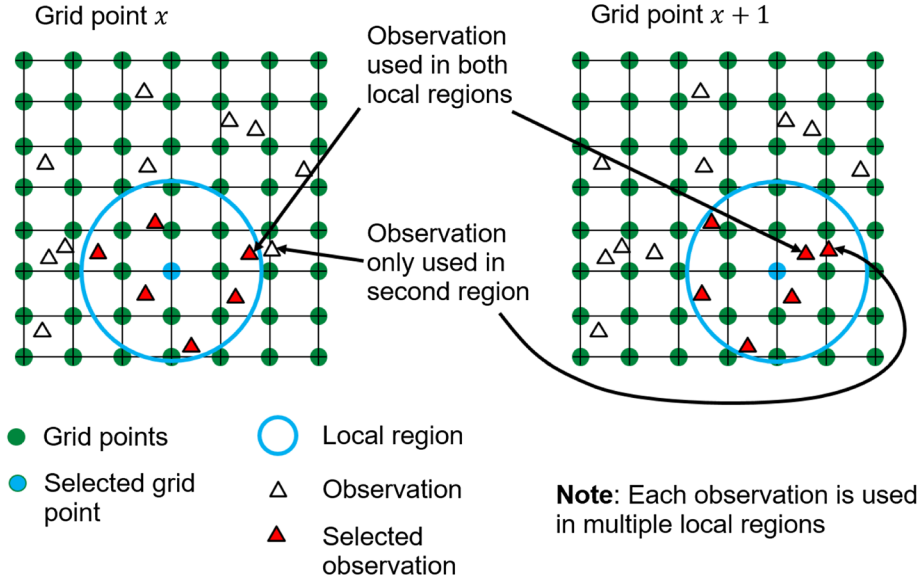
or equivalently  $\mathbf{W}_a \mathbf{W}_a^T = (k - 1) \tilde{\mathbf{A}}$ , then using this in Equation (15) gives the required definition of the analysis covariance, Equation (9). As such the the analysis ensemble can then be constructed by:

$$\mathbf{X}_a = \mathbf{X}_b \mathbf{W}_a. \quad (20)$$

For  $\bar{\mathbf{x}}_a$  to be the mean of the analysis ensemble, the sum across the columns of  $\mathbf{X}_a$  must be zero, i.e.  $\mathbf{X}_a \mathbf{1}^T = 0$  (where  $\mathbf{1} = (1, 1, \dots, 1)$ , a vector of 1's). By the definition of  $\mathbf{X}_b$ ,  $\mathbf{X}_b \mathbf{1}^T = 0$ . Therefore for  $\mathbf{X}_a \mathbf{1}^T = 0$  it is required that  $\mathbf{1}$  is an eigenvector of  $\mathbf{W}_a$ . [Wang et al. \(2004a\)](#) showed that this is true if the symmetric solution of the square root is chosen.

The complete set of LETKF equations (in form analogous to the update step of the classic Kalman filter) are:

$$\mathbf{K} = \mathbf{X}_b \tilde{\mathbf{A}} (\mathbf{Y}_b)^T \mathbf{O}^{-1}, \quad (21)$$



**Fig. 1.** Cartoon of how the LETKF works. For each grid point (green dots, selected point is blue) a localisation region is drawn (blue circle) and only those observations (red triangles) which are in the localised region are included in the analysis. Cartoon based on that by Kalnay (2005).

$$\mathbf{A} = \mathbf{X}_b \tilde{\mathbf{A}} (\mathbf{X}_b)^T, \quad (22)$$

$$\bar{\mathbf{x}}_a = \bar{\mathbf{x}}_b + \mathbf{X}_b \mathbf{w}_a, \quad (23)$$

$$\mathbf{X}_a = \mathbf{X}_b ((k-1) \tilde{\mathbf{A}})^{\frac{1}{2}}, \quad (24)$$

where

$$\tilde{\mathbf{A}} = [(k-1)\mathbf{I} + (\mathbf{Y}_b)^T \mathbf{O}^{-1} \mathbf{Y}_b]^{-1}, \quad (25)$$

$$\mathbf{w}_a = \tilde{\mathbf{A}} (\mathbf{Y}_b)^T \mathbf{O}^{-1} (\mathbf{y}_o - \bar{\mathbf{y}}_b). \quad (26)$$

It can be seen that all the matrix operations take place in ensemble space (i.e. on matrices defined by the ensemble) with the largest matrix dimension being  $s$  (the number of observations). For the AENeAS implementation of the LETKF it is assumed that the observations are independent and thus the observation error covariance matrix,  $\mathbf{O}$ , is diagonal.

The prediction step of the classic Kalman filter (Eqs. (4) and (5)) is not needed. The forecast model (which moves the state from one time step to the next) is replaced by a (not necessarily linear) model which moves each ensemble member forward in time.

The final component of the LETKF is localisation. The localisation used by the LETKF is similar to that described by Ott et al. (2004), where for each model grid point a subset of the global matrices are used. First the background perturbation matrices in both model and observation space are calculated globally. Then for each grid point the observation vector and associated background perturbation matrix mapped into observation space ( $\mathbf{Y}_b$ ) are restricted to only include observations from a region around that point (Fig. 1). Similarly the perturbation matrix and associated mean vector are restricted to only include model variables from the grid point. The LETKF equations (Eqs. (21)–(26) from Sect. 2.2) are then

applied to each region. The output of the equations for each region only updates the centre grid point.

At each grid point an optimum combination of ensemble members are calculated (compared to the data). These represent local multidimensional unstable manifolds (Szunyogh, 2014). The global analysis is constructed by smoothly joining each of these together, i.e. forming a global unstable manifold. Szunyogh (2014) describes how this combination can be made using a model propagator; however in the LETKF this combination relies upon the choice of the matrix square root in the construction of the analysis (Wang et al., 2004b). The choice of the symmetric square root transformation forces consistency between adjacent local analyses since the transformation is a continuous function of the analysis error covariance matrix (Hunt et al., 2007). This consistency does not occur if the transformation suggested in Bishop et al. (2001) is used. As long as subsets of observations used for neighbouring grid points overlap heavily, the local weight vectors for the grid points are very similar. When weight vectors do not change much, the analysis ensemble members are roughly linear combinations of the ensemble members, and so the changes between neighbouring grid points remain reasonably physical. The effectiveness of this smoothing as compared to the global approach (i.e. LETKF compared to ETKF) is highly dependent upon the choice of region size. The regions around each model grid point should be chosen such that their intersection is sufficiently large (in comparison to the number of observations in each region).

One issue with all EnKFs is undersampling, whereby the ensemble is too small to statistically represent the error covariance matrix. Consequences of undersampling are spurious covariances and filter divergence (Anderson, 2001). Spurious covariances between variables are ones where the variables are not physically related. These may be in parts of the state vector that are physically a large distance apart. Filter divergence is the term used to describe when the assimilation stops

improving the analysis state (Anderson, 2001). This can occur if the model variances become too large or too small in relation to the observation variance.

The localisation component of the the LETKF helps to address filter divergence by improving the specification of the covariance matrix. One of the underlying assumptions of ensemble modelling is that the set of ensemble members span the space of the model variables. In reality the whole space is often not fully spanned, but it is necessary that it is a reasonable approximation. By reducing the ensemble analysis to local regions, the ensemble members only need to span the region within those regions. The dimensionality of the reduced region is significantly less than the whole space and the approximation that the ensemble spans the space is a better one. Hunt et al. (2007) used a Lorenz-96 model to investigate the effectiveness of this localisation system by comparing the LETKF and ETKF and varying the ensemble sizes for models of dimension 40 and 80. The results showed that the global ETKF required an ensemble size of half the model dimension to be effective. In contrast doubling the dimension of the model had little effect on the performance of the LETKF, and a good performance was obtained for ensemble sizes greater than or equal to 10.

### 3 Advanced Ensemble electron density (Ne) Assimilation System (AENeAS)

#### 3.1 Introduction

The Advanced Ensemble electron density (Ne) Assimilation System (AENeAS) is a new upper atmosphere DA model that combines a physics background model and the LETKF. Currently, the physics model is provided by TIE-GCM (described in Sect. 3.2). However AENeAS is modular so that in future other physics models can be used instead of, or as well as, TIE-GCM.

The primary motivation for developing AENeAS has been for ionospheric/thermospheric forecasting. Ionospheric/thermospheric physics models rely upon knowledge of electron, ion and neutral densities as well as temperatures and wind speeds. However the most abundant data source in the upper atmosphere is an integrated measurement of electron density called the total electron content (TEC) (described in Sect. 4.3). The LETKF is used so that the unknown covariances can be estimated from the TIE-GCM ensemble. The optimal number of ensemble members (balancing performance and computational cost) is still to be determined. However a 15 min AENeAS timestep can be run with 100 ensemble members in real-time on a 32-core computer. Benchmarking test results have shown that using 32 ensemble members provides good agreement with independent observations.

#### 3.2 Background Models

##### 3.2.1 TIE-GCM

TIE-GCM is a global three-dimensional model of the coupled thermosphere ionosphere system (Richmond et al., 1992). At each time step the continuity, energy, and momentum equations are solved for neutral and ion species using a fourth-order,

centred finite difference scheme (Roble et al., 1988). The model is driven by the daily F10.7 (the solar flux at a wavelength of 10.7 cm at the Earth's orbit which is used as a proxy for solar output) (Tapping, 2013), the 81-day F10.7 average and the Kp (which indicates the severity of the magnetic disturbances in near-Earth space) (Bartels, 1957). Instead of the Kp the model can use the Earth's magnetic field cross-tail potential (CTP) and the hemispheric power (HP) which are measures of auroral activity. These values are related to Kp via (Neale, 2010):

$$HP = \begin{cases} 16.82e^{0.32Kp} - 4.86, & \text{if } Kp \leq 7 \\ 153.13 + \frac{146.87(Kp-7)}{2}, & \text{otherwise} \end{cases} \quad (27)$$

$$CTP = 0.8Kp^2 + 15Kp + 15. \quad (28)$$

The lower boundary condition (atmospheric tides) is given by the Global Scale Wave Model (GSWM) (Hagan et al., 1999).

##### 3.2.2 NeQuick

TIE-GCM only extends to 600 km in altitude (depending on solar conditions). Whereas in reality the ionosphere/plasmasphere extends much further than this and for accurate specification of the domain higher altitudes need to be considered. To achieve this in AENeAS, a higher reaching model of the ionosphere is attached to the topside of the TIE-GCM profiles which extends the electron density grids to 25,000 km. The other species considered by the model also have their grids extended to the same height (to ensure the altitude grid is the same across all species); however all values above the TIE-GCM maximum altitude are set to zero. The current ionospheric topside model used by AENeAS is NeQuick (Nava et al., 2008).

NeQuick is a monthly median ionospheric electron density model developed at the Aeronomy and Radiopropagation Laboratory (now Telecommunications/ICT for Development Laboratory) of the Abdus Salam International Centre for Theoretical Physics (ICTP), Trieste, Italy, and at the Institute for Geophysics, Astrophysics and Meteorology (IGAM) of the University of Graz, Austria (Nava et al., 2008). The model is based on the Di Giovanni-Radicella (DGR) model (Giovanni & Radicella, 1990). It has been designed to have continuously integrable vertical profiles which allows for rapid calculation of the total electron content for trans-ionospheric propagation applications. For use in AENeAS, the NeQuick bottomside profile is driven by parameters (height and peak density of the E, F1 and F2 ionospheric layers) taken from TIE-GCM. At altitudes above the TIE-GCM maximum the electron densities from NeQuick are used. By driving NeQuick with parameters derived from TIE-GCM ensures that the bottom and top-side profiles are smoothly connected.

As well as time and location, the model is also driven by F10.7 (like TIE-GCM). The topside of NeQuick, which is used in AENeAS, is a simplified approximation to a diffusive equilibrium. A semi-Epstein layer represents the model topside with a height-dependent thickness parameter that has been empirically determined (Nava et al., 2008). For AENeAS, the University of Birmingham (UoB) Python version of NeQuick has been used (Angling et al., 2018).

### 3.3 Ensemble generation

To use the background model in the LETKF an ensemble of the model must be generated. The spread of the ensemble members should be sufficient to estimate the background covariance matrix (Sect. 3.1). For AENeAS, the F10.7, cross-tail potential (CTP) and hemispheric power (HP) are perturbed to generate a set of different background models. The CTP and HP values are chosen to be Normally distributed, where the mean of the distribution is the observed Kp converted to HP (Eq. (27)) and CTP (Eq. (28)) and the standard deviation is set at 40% of the observed value. The F10.7 values are generated using a random walk. At each time step the walk value has a uniformly distributed random number in the range  $[-3, 3]$  added to it but restricted to the range  $[-100, 100]$ . By construction the mean of the F10.7 values (across the ensemble members) is close to the observed F10.7, ensuring the long term trends of F10.7 are captured. F10.7 values are updated every 3 h (which is in line with the cadence of the Kp data) and ensures the ensemble members have time to respond to changes in conditions. The internal model time step is set to 60 s and each assimilation window is set at 15 min.

### 3.4 Localisation

In AENeAS localisation regions of  $\pm 10^\circ$  latitude and  $\pm 20^\circ$  longitude are used. These values have been chosen based on the estimated correlation lengths in the ionosphere from McNamara (2009) (Table 1).

### 3.5 Covariance inflation

Any EnKF depends on the accuracy of the estimated background and observation error covariances and errors can effect the optimality of the data assimilation scheme (Liang et al., 2012). Unfortunately, using the ensemble to estimate the covariance matrices can result in sampling errors as the number of ensemble members is limited. Previous work has shown that small ensembles lead to underestimation of the covariances which eventually result in filter divergence (Anderson & Anderson, 1999; Constantinescu et al., 2007). Specifically the ensemble members become too similar. To mitigate this covariance inflation has been used. First proposed by Anderson and Anderson (1999) covariance inflation artificially increases the uncertainty in the background model. The required inflation value to use depends on the specific case (both domain and ensemble size) Hamill et al. (2005). Whilst there are various strategies for implementing covariance inflation (e.g. Hamill & Whitaker, 2005; Anderson, 2007, 2009; Whitaker et al., 2008; Li et al., 2009; Miyoshi, 2011; Kang et al., 2012), here multiplicative inflation Anderson and Anderson (1999) has been used.

Multiplicative inflation is more suited to the ionosphere-thermosphere problem than other inflation schemes due to the dynamic range of values involved. Currently the Anderson and Anderson (1999) fixed multiplicative inflation is implemented in AENeAS. This involves multiplying the background error covariance matrix,  $B$ , by an inflation factor  $\rho > 1$ . Since  $B = (k-1)^{-1} \mathbf{X}_b \mathbf{X}_b^T$  this is equivalent to multiplying  $\mathbf{X}_b$  by  $\sqrt{\rho}$ . This can be achieved efficiently by replacing  $\tilde{\mathbf{A}} = [(k-1)\mathbf{I} + (\mathbf{Y}_b)^T \mathbf{O}^{-1} \mathbf{Y}_b]^{-1}$  with  $\tilde{\mathbf{A}} = [(k-1)\mathbf{I}/\rho + (\mathbf{Y}_b)^T \mathbf{O}^{-1} \mathbf{Y}_b]^{-1}$  in Equation (15) (Hunt et al., 2007).

**Table 1.** Ionospheric (F region) correlation lengths (km) from McNamara (2009).

Solar activity	North-South	East-West
High	1000	1500
Low	700	1000

### 3.6 Sporadic Forcing

It is well known that physics-based general circulation models (GCMs) suffer in data assimilation schemes when sporadically forced by data (Baker et al., 1987; Lorenc et al., 1991; Macpherson, 1991; Bloom et al., 1996; Fox-Rabinovitz, 1996). The sporadic nature of the assimilated data (be it temporally or spatially) can upset the intrinsic balance of the equations (Ham et al., 2016) leading to discontinuities and high-frequency oscillations (Baker et al., 1987).

Two main approaches have been developed to deal with this problem. The first relies on the repeated insertion (RI) of data (also known as nudging). This involves using each observation multiple times in an observation window (which in the lower atmosphere community can be as long as 6 h). The observations are weighted throughout the window so that the largest weights are applied at the observation time, with decreasing weights before and after (Lorenc et al., 1991; Macpherson, 1991). This has the impact of slowly moving the background model state towards the analysis state over time. Secondly, incremental analysis updating (IAU) has been used (Baker et al., 1987). IAU distributes one analysis increment throughout the assimilation window rather than repeatedly assimilating differently weighted observations. This acts as a low pass filter to reduce high frequency oscillations (Ham et al., 2016). The key difference between IAU and RI is that IAU only modifies the analysis increment and not the background state, as RI does (Polavarapu et al., 2004).

Whilst both approaches have been used extensively in the lower atmosphere they have been developed in schemes which use long (approximately 6 h) assimilation windows. AENeAS relies on a much shorter assimilation window of 15 min, so neither RI or IAU are directly applicable. Therefore a different approach been developed; at each grid point the amount which any variable can change has been limited. For example for a single species, after calculating  $X_a$  in the LETKF ( $\mathbf{X}_a = \mathbf{X}_b((k-1)\tilde{\mathbf{A}})^{\frac{1}{2}}$ ; Eq. (25)) the actual value of  $X_a$  (at each grid point) used to update the model is determined by:

$$X_a^{\text{used}} = \begin{cases} \frac{X_b}{\alpha}, & \text{if } \frac{X_a}{X_b} < \frac{1}{\alpha} \\ X_a, & \text{if } \frac{1}{\alpha} \leq \frac{X_a}{X_b} \leq \alpha, \\ \alpha X_b, & \text{if } \frac{X_a}{X_b} > \alpha \end{cases} \quad (29)$$

where  $\alpha$  is a predefined constant set to the largest allowable change (i.e.  $\alpha = 1.1$  limits the amount of changes from  $-9.1\%$  to  $+10\%$ ). This causes the background variables to move “in the direction” of the analysis, but slowly so as to ensure the background model equations remain stable. Initially, many grid points changes are limited by this procedure (since the difference between the background and analysis is the largest; Lee et al., 2006), however this reduces over time as the background model is moved nearer the observations.

### 3.7 Implementation

Hunt et al. (2007) described a computationally efficient implementation of the LETKF in seven steps. Steps 4 and 5 involve calculating  $\tilde{\mathbf{A}}$  (Eq. (14)) and  $\mathbf{W}_a$  (Eq. (19)). The computational cost of both calculations is proportional to the cube of the ensemble size. However the number of calculations can be reduced by first calculating

$$\tilde{\mathbf{A}}^* = [(k-1)\mathbf{I} + \mathbf{C}\mathbf{Y}_b], \quad (30)$$

which is the pre-inverse of  $\tilde{\mathbf{A}}$  defined in Equation (14). This can then be factorized to

$$\tilde{\mathbf{A}}^* = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1}, \quad (31)$$

where  $\mathbf{Q}$  is the matrix of eigenvalues of  $\tilde{\mathbf{A}}^*$  and  $\mathbf{\Lambda}$  is a diagonal matrix whose elements are the eigenvalues of  $\tilde{\mathbf{A}}^*$ . The number of operations to do this is, at most, the order of the cube of the ensemble size (since it is known that  $\tilde{\mathbf{A}}^*$  is symmetric). Since  $\mathbf{\Lambda}$  is diagonal it is easy to calculate both  $\mathbf{\Lambda}^{-1}$  and  $\mathbf{\Lambda}^{-\frac{1}{2}}$  (with the number of operations linearly proportional to size of the ensemble) and so

$$\tilde{\mathbf{A}} = \mathbf{Q}\mathbf{\Lambda}^{-1}\mathbf{Q}^{-1}, \quad (32)$$

$$\mathbf{W}_a = \mathbf{Q}\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{Q}^{-1}. \quad (33)$$

This approach also allows for a simple check that  $\tilde{\mathbf{A}}$  is positive definite (as required by the LETKF) since all the eigenvalues will be  $> 0$ .

## 4 Initial results – AENeAS benchmarking

### 4.1 Overview

During the AENeAS development cycle a short benchmark test has been developed for investigating changes in performance as improvements are made. A more detailed statistical study to quantify the performance of the model is necessary in the future. This initial test compares AENeAS with its background model, TIE-GCM+ (TIE-GCM with an NeQuick topside), NeQuick and the IGS global TEC map.

### 4.2 Test scenario

The time period for the test scenario is between June 4th and June 7th, 2017. Figure 2 shows the solar and geomagnetic conditions during the test period. It is a solar and geomagnetically quiet time with the F10.7 only varying between 75 and 79 and the Kp between 0 and 2 (well below the storm threshold of Kp 4). This is beneficial for the benchmarking test as the models do not have to respond to any rapidly varying geophysical conditions.

### 4.3 Data

Signals from global navigation satellite systems (GNSS) can be used to determine the total electron content (TEC) between the satellite and a receiver on the ground (or in space for radio occultation, RO). TEC is the line integral of electron density on

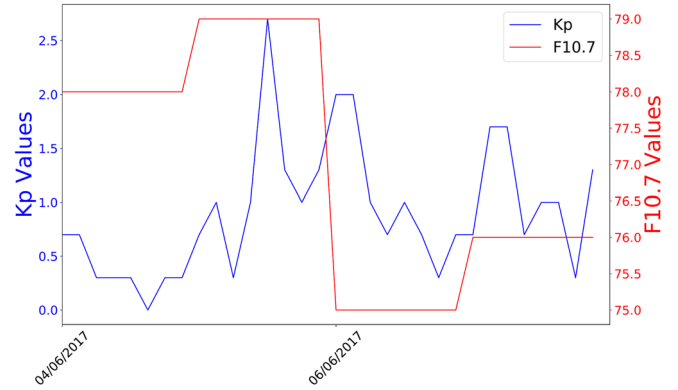


Fig. 2. Kp and F10.7 values for the AENeAS benchmarking test (June 4th 2017–June 7th 2017).

a path between two points (i.e. between the satellite and the receiver)

$$\text{TEC} = \int N_e dl, \quad (34)$$

where  $N_e$  is the electron density. TEC is often expressed in TECU units (TECU), which is  $10^{16}$  electrons/m<sup>2</sup>. A full derivation of how TEC is estimated can be found in Elvidge (2014).

The TEC data has been provided by the Ionospheric determination and Navigation based on Satellite And Terrestrial (IonSAT) systems research group at the Universitat Politècnica de Catalunya (UPC). The data has been calibrated (i.e. biases have been removed) by leveling the data to the UQRG model which is produced by combining tomographic modelling of the ionosphere with Kriging interpolation (Hernández-Pajares et al., 1999; Orus et al., 2002). This model has been shown to provide excellent performance compared to independent observations and other Global Ionosphere Maps (GIMs) (Hernández-Pajares et al., 2017).

An example map of the data is shown in Figure 3. This figure shows the TEC values (which are assimilated) mapped to vertical TEC (VTEC) values (in order to more easily visualize the data). The mapping function assumed a thin-shell ionosphere at a fixed altitude of 300 km.

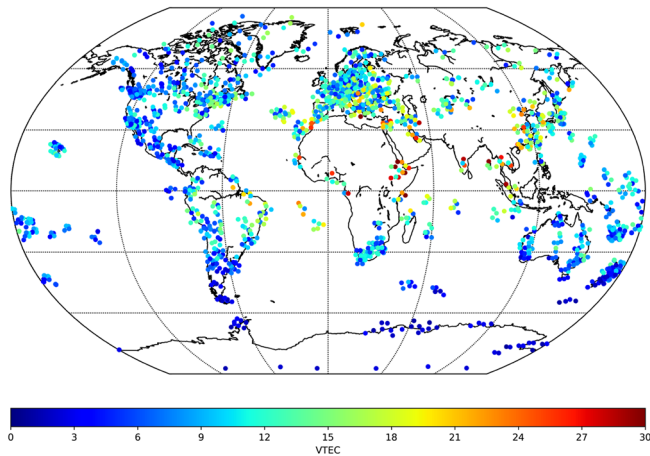
For testing, the models are compared in terms of their ability to reproduce the total electron content (TEC) from a receiver that has not been assimilated.

### 4.4 Results

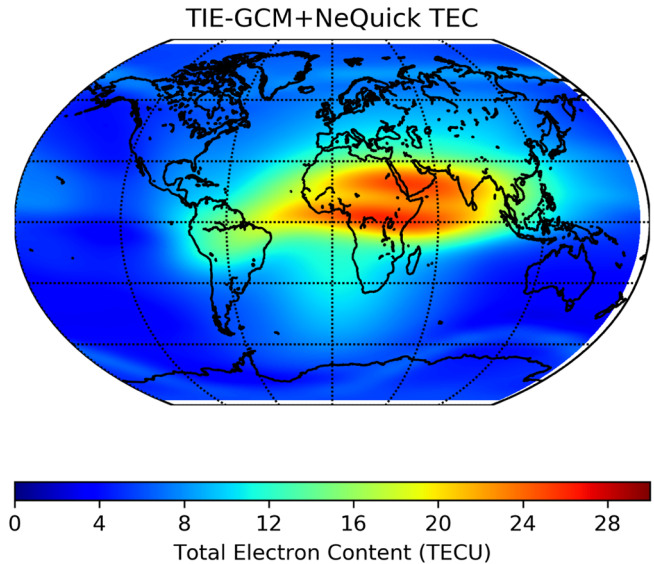
Assimilating the data causes AENeAS to make global changes to the background model. A sample AENeAS post assimilation TEC map (at 1230 on June 5th) is shown in Figure 4. The corresponding TIE-GCM+ output is shown in Figure 5.

The changes result in the AENeAS TEC being generally higher than TIE-GCM+ in the northern hemisphere and lower or unchanged in the south. The exception to this is Australasia that has a higher TEC than the background model. Additional small scale structure is also created and can be seen especially clearly around the equator. Figures 3–5 are plotted on the same colour scale for direct comparison. It can be seen that the AENeAS results are consistent with the input data. For reference the IGS Global (IGSG) TEC map is shown in Figure 6

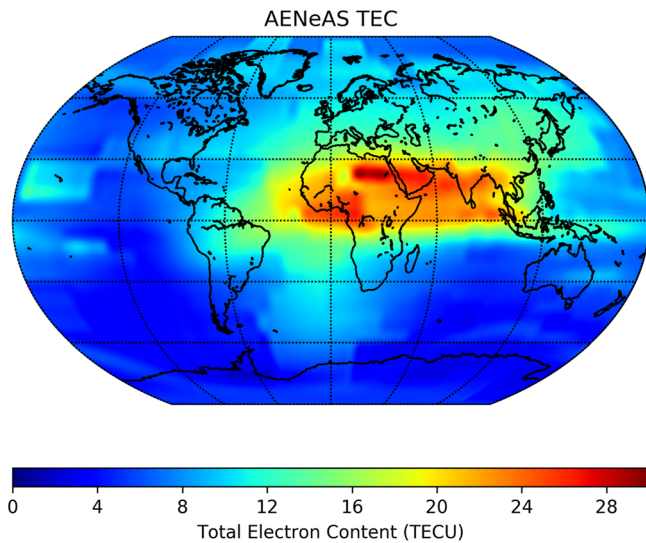




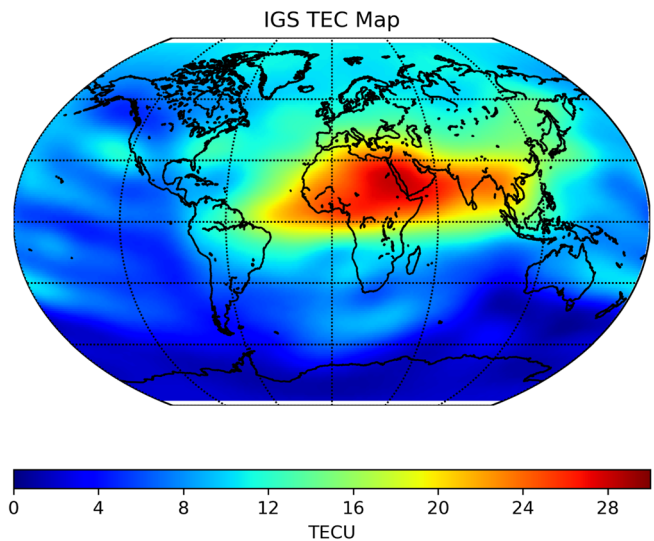
**Fig. 3.** Example map of vertical TEC values derived from the UPC dataset of slant TEC (which is assimilated) on June 5th 2017 at 1230 UT. The colour bar shows the VTEC value (in TECU).



**Fig. 5.** TEC map for June 5th at 1230 from TIE-GCM+.



**Fig. 4.** TEC map for June 5th at 1230 from AENeAS.



**Fig. 6.** TEC map for June 5th at 1230 from the IGS Global map.

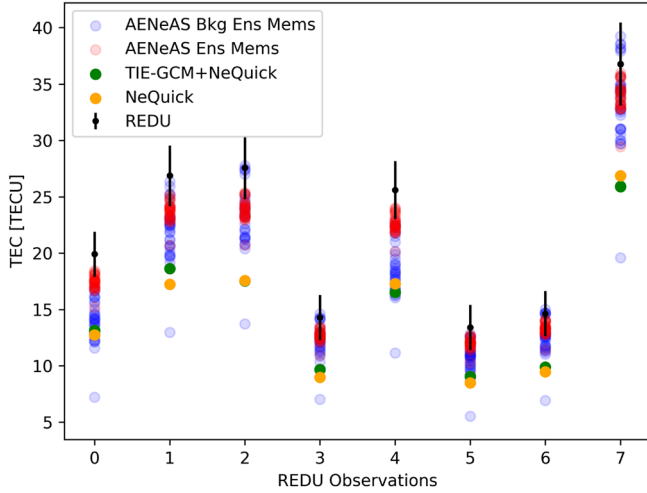
(Hernández-Pajares et al., 2009). It should be observed that the general structure more closely agrees with the AENeAS output rather than TIE-GCM+. In particular the areas of increased density over Australasia, the South Atlantic and California.

The GPS station “REDU” (based in Belgium) was excluded from the assimilation so that it could be independently compared against the AENeAS results. At 1230 on June 5th the REDU receiver could see 8 satellites and therefore returned 8 slant TEC observations. As with the assimilated data, these TEC values are levelled to UQRG to provide calibrated TEC. For comparison, the modelled TEC is then found by integrating along each ray path through the electron density grids.

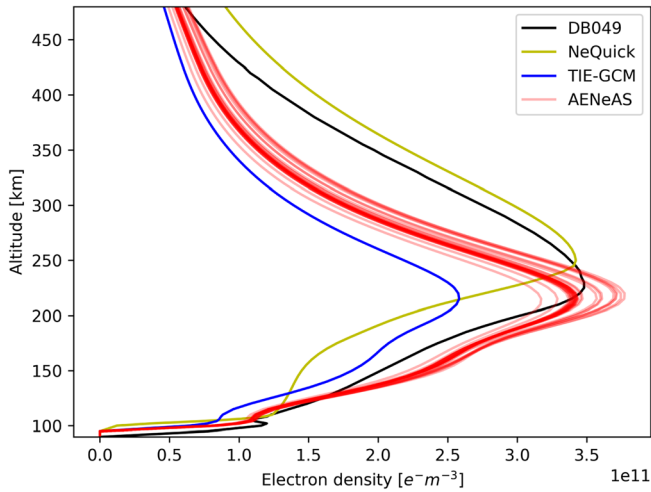
The observations have been compared with AENeAS (pre- and post-assimilation), TIE-GCM+ and NeQuick (Fig. 7). The REDU observations are plotted with black vertical lines representing the uncertainty of the observations. TIE-GCM+ is shown in green and is significantly different from the

observations. On average the TIE-GCM integrated TECs are approximately 70% of the observation. The NeQuick results (yellow) are very similar to the TIE-GCM+ values and are lower than all of the observations. The pre-assimilation AENeAS results (blue) shows that the model has successfully “moved” the background TIE-GCM+ values to nearer the observed values. The pre-assimilation results include all of the AENeAS history upto the previous timestep. Assimilating data has the effect of moving the mean closer to the observation and reducing the spread of the ensemble members (shown in red). For each observation at least one fifth of the analysis ensemble member is within the uncertainty of the observation.

Figure 8 shows a snapshot of electron density profiles from June 5th at 1230 from Dourbes, Belgium (longitude: 4.6, latitude: 50.1). Each individual AENeAS ensemble member (red) is shown, as well as TIE-GCM (blue) and NeQuick (yellow).



**Fig. 7.** Observed TEC from REDU and the same TEC values modelled by AENeAS, TIE-GCM+ and NeQuick at 1230, 5 June 2017. The x-axis numbering is the receiver channel number and the ordering is arbitrary.



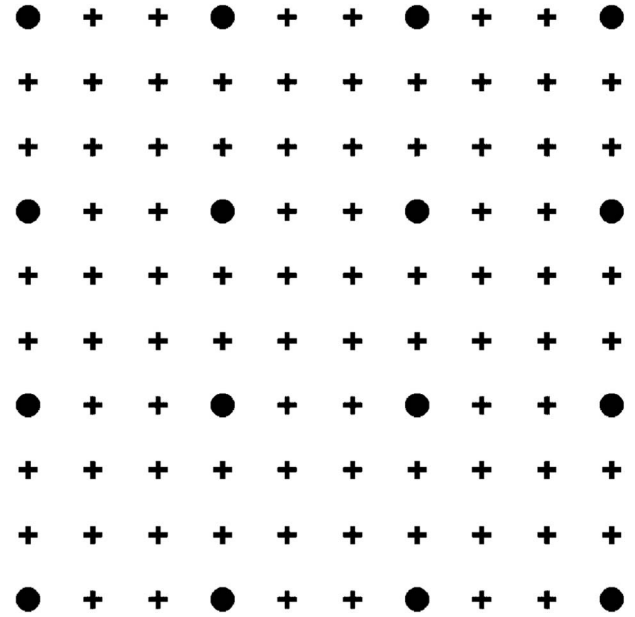
**Fig. 8.** Example electron density profiles. Observed profile (DB049) is from Dourbes, Belgium at 1230, 5 June 2017. Profiles from TIE-GCM, NeQuick and each individual AENeAS member is plotted.

For reference the profile derived from the Dourbes ionosonde (DB049) is shown in black. It can be seen that AENeAS does very well moving from the background model (TIE-GCM) towards the Dourbes ionosonde. It is important to remember that the electron density profile has not been assimilated into AENeAS, and only slant TEC has. Even with the lack of height information from the TEC measurements, the reconstructed peak electron density from AENeAS closely matches the value from the ionosonde.

The overall statistics for the full 3 day test are shown in Table 2. It can be seen that AENeAS performs very well, especially compared to its background model, TIE-GCM+. Both the mean and standard deviation of the model errors (model minus observation) are much less for AENeAS than the other two models. NeQuick, the empirical model, also performs better

**Table 2.** Overall statistics of the model test for June 4th–7th, 2017.

Models	Error mean (TECU)	Error standard deviation (TECU)	Root mean square error
NeQuick	−5.3	1.6	5.5
TIE-GCM+	−6.1	2.9	6.8
IGSG	−0.6	4.3	4.3
AENeAS	−1.9	0.8	2.1



**Fig. 9.** The circles are where the LETKF assimilation is performed. The higher resolution grid (crosses) is then found by interpolation of the weight vectors. Image from Yang et al. (2009).

than the physics-based TIE-GCM+. For further reference the statistics from the IGSG model are also included. Two important things should be noted with regards to the IGS global output. Firstly the REDU observations have not been removed from the IGSG model, and, unsurprisingly, the mean error is significantly, 1.5 TECU units, less than AENeAS. Secondly, the IGS global map is only output every 2 h, and the map for 1230 is linearly interpolated from the 1200 and 1400 maps. Whilst the mean error of the IGSG is the smallest of all of the models, it has the largest error standard deviation. This could be a result of comparing the smoothed IGSG output to the 30 s observations (Hernández-Pajares et al., 2009).

Although AENeAS performs well there are some issues with artificial boundary effects in the assimilation. This is most obvious in the region around Hawaii in Figure 4. Whilst Section 2.2 presents approaches for improving smoothness these are clearly insufficient in areas of very sparse data. Another method for ensuring smoothness has been described by Yang et al. (2009). This approach involves performing the LETKF calculations on a coarse grid and then interpolating to the required grid resolution (Fig. 9). Specifically the ensemble weight vectors are interpolated to find the high resolution grid point values. Yang et al. (2009) use a smooth bivariate

interpolation scheme by locally fitting quintic polynomials to the zonal and meridional values of the coarse analysis grid. This scheme insures that the analysis ensemble perturbations maintain the zero mean property as required for the LETKF. Whilst this method is a sub-optimal assimilation scheme it has the advantage that the smoothness is guaranteed. It will be implemented in the next iteration of AENeAS.

## 5 Conclusions

A new model, the Advanced Ensemble electron density (Ne) Assimilation System (AENeAS) has been developed. AENeAS is a physics-based data assimilation model of the Earth's upper atmosphere. Its background model is TIE-GCM and the model uses the local ensemble transform Kalman filter (LETKF) for the assimilation scheme. The LETKF is an efficient implementation of the ensemble Kalman filter which defines local regions where the assimilation is performed. The advantage of this is that it reduces the state space of the model and brings it closer to the space spanned by the ensemble members. The algorithm iterates through each grid point independently and so is naturally suitable for parallelization. A computationally efficient implementation using eigendecomposition has also been presented.

Data assimilation with physics-based models can upset the intrinsic balance of the model equations. Nudging or incremental analysis updating is used in the meteorological community to slowly move the background model to the analysis state. However in the lower atmosphere the assimilation window is usually 6 h and these approaches have been developed with that in mind. For the upper atmosphere the assimilation windows are usually much shorter (15 min) and therefore a new sporadic forcing technique is described.

Results of initial testing show that AENeAS can effectively assimilate GPS TEC data and reduce TEC errors compared to TIE-GCM and NeQuick. However a more detailed statistical study is required to quantify the performance of the model. Further methods will be required to ensure model smoothness in areas with very sparse data.

*Acknowledgements.* AENeAS has been developed as part of a QinetiQ project, funded by the Dstl Space programme. We thank Manuel Hernández Pajares for the cleaned GNSS data used in the benchmarking test.

The editor thanks Sergey Fridman and an anonymous referee for their assistance in evaluating this paper.

## References

- Anderson JL. 2001. An ensemble adjustment Kalman filter for data assimilation. *Mon Weather Rev* **129**: 2884–2903.
- Anderson JL. 2007. Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. *Physica D* **230**(1–2): 99–111. DOI: [10.1016/j.physd.2006.02.011](https://doi.org/10.1016/j.physd.2006.02.011).
- Anderson JL. 2009. Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus A* **61**(1): 72–83. DOI: [10.1111/j.1600-0870.2008.00361.x](https://doi.org/10.1111/j.1600-0870.2008.00361.x).
- Anderson JL. 2010. A non-Gaussian ensemble filter update for data assimilation. *Mon Weather Rev* **138**: 4186–4198.
- Anderson JL, Anderson SL. 1999. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon Weather Rev* **127**(12): 2741–2758. DOI: [10.1175/1520-0493\(1999\)127<2741:AMCIOT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO;2).
- Angling MJ, Elvidge S, Healy SB. 2018. Improved model for correcting the ionospheric impact on bending angle in radio occultation measurements. *Atmos Meas Tech* **11**: 2213–2224. DOI: [10.5194/amt-11-2213-2018](https://doi.org/10.5194/amt-11-2213-2018).
- Angling MJ, Jackson-Booth N. 2011. A short note on the assimilation of collocated and concurrent GPS and ionosonde data into the Electron Density Assimilative Model. *Radio Sci* **46**: DOI: [10.1029/2010RS004566](https://doi.org/10.1029/2010RS004566).
- Baker WE, Bloom SC, Woollen JS, Nestler MS, Brin E, Schlatter TW, Branstator GW. 1987. Experiments with a three-dimensional statistical objective analysis scheme using FGGE data. *Mon Weather Rev* **115**(1): 272–296. DOI: [10.1175/1520-0493\(1987\)115<0272:EWATDS>2.0.CO;2](https://doi.org/10.1175/1520-0493(1987)115<0272:EWATDS>2.0.CO;2).
- Bartels J. 1957. The geomagnetic measures for the time-variations of solar corpuscular radiation, described for use in correlation studies in other geophysical fields. *Ann. Intern. Geophys. Year* **4**: 227–236.
- Bilitza D, Rawer K, Pallaschke S. 1988. Study of ionospheric models for satellite orbit determination. *Radio Sci* **23**(3): 223–232.
- Bilitza D, Reinisch BW. 2008. International reference ionosphere 2007: Improvements and new parameters. *Journal of Advanced Space Research* **42**(4): 599–609. DOI: [10.1016/j.asr.2007.07.048](https://doi.org/10.1016/j.asr.2007.07.048).
- Bishop CH, Etherton BJ, Majumdar SJ. 2001. Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon Weather Rev* **129**: 420–436.
- Bloom SC, Takacs LL, da Silva AM, Ledvina D. 1996. Data assimilation using incremental analysis updates. *Mon Weather Rev* **124**(6): 1256–1271. DOI: [10.1175/1520-0493\(1996\)124<1256:DAUIAU>2.0.CO;2](https://doi.org/10.1175/1520-0493(1996)124<1256:DAUIAU>2.0.CO;2).
- Chartier AT, Jackson DR, Mitchell CN. 2013. A comparison of the effects of initializing different thermosphere-ionosphere model fields on storm time plasma density forecasts. *J Geophys Res (Space Phys)* **118**(11): 7329–7337. DOI: [10.1002/2013JA019034](https://doi.org/10.1002/2013JA019034).
- Chartier AT, Matsuo T, Anderson JL, Collins N, Hoar TJ, Lu G, Mitchell CN, Coster AJ, Paxton LJ, Bust GS. 2016. Ionospheric data assimilation and forecasting during storms. *J Geophys Res (Space Phys)* **121**: 764–778. DOI: [10.1002/2014JA020799](https://doi.org/10.1002/2014JA020799).
- Constantinescu EM, Sandu A, Chai T, Carmichael GR. 2007. Ensemble-based chemical data assimilation. I: General approach. *Quart J Roy Meteor Soc* **133**(626): 1229–1243. DOI: [10.1002/qj.76](https://doi.org/10.1002/qj.76).
- Eddy A. 1967. The statistical objective analysis of scalar data fields. *J Appl Meteorol* **6**(4): 597–609.
- Elvidge S. 2014. On the use of multi-model ensemble techniques for ionospheric and thermospheric characterisation. *PhD Thesis*. University of Birmingham.
- Elvidge S, Angling MJ, Nava B. 2014. On the use of modified Taylor diagrams to compare ionospheric assimilation models. *Radio Sci* **49**: 737–745. DOI: [10.1002/2014RS005435](https://doi.org/10.1002/2014RS005435).
- Etherton BJ. 2007. Preemptive forecasts using an ensemble Kalman filter. *Mon Weather Rev* **135**(10): 3484–3495. DOI: [10.1175/MWR3480.1](https://doi.org/10.1175/MWR3480.1).
- Evensen G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J Geophys Res* **99**(C5): 10143. DOI: [10.1029/94JC00572](https://doi.org/10.1029/94JC00572).
- Evensen G. 2009. *Data assimilation, the ensemble Kalman filter*, 2nd edn. Springer-Verlag, Berlin, Heidelberg. ISBN 9783642037108.
- Fox-Rabinovitz MS. 1996. Diabatic dynamic initialization with an iterative time integration scheme as a filter. *Mon Weather Rev* **124**(7): 1544–1557. DOI: [10.1175/1520-0493\(1996\)124<1544:DDIWA1>2.0.CO;2](https://doi.org/10.1175/1520-0493(1996)124<1544:DDIWA1>2.0.CO;2).

- Fridman SV, Nickisch LJ, Hausman M. 2009. Personal-computer-based system for real-time reconstruction of the three-dimensional ionosphere using data from diverse sources. *Radio Sci* **44** (RS3008): 1–12. DOI: [10.1029/2008RS004040](https://doi.org/10.1029/2008RS004040).
- Gandin L. 1963. *Objective analysis of meteorological fields*. Gridromet, Leningrad. English translation (Jerusalem: Israel Program for Scientific Translation), 1965.
- Giovanni GD, Radicella S. 1990. An analytical model of the electron density profile in the ionosphere. *Adv Space Res* **10**(11): 27–30. DOI: [10.1016/0273-1177\(90\)90301-F](https://doi.org/10.1016/0273-1177(90)90301-F).
- Hagan ME, Burrage MD, Forbes JM, Hackney J, Randel WJ, Zhang X. 1999. GSWM-98: Results for migrating solar tides. *J Geophys Res (Space Phys)* **104**(A4): 6813–6827. DOI: [10.1029/1998JA900125](https://doi.org/10.1029/1998JA900125).
- Ham Y-G, Song H-J, Jung J, Lim G-H. 2016. Development of the nonstationary incremental analysis update algorithm for sequential data assimilation system. *Adv Meteorol* **2016**: 4305204. DOI: [10.1155/2016/4305204](https://doi.org/10.1155/2016/4305204).
- Hamill TM, Whitaker JS. 2005. Accounting for the error due to unresolved scales in ensemble data assimilation: a comparison of different approaches. *Mon Weather Rev* **133**(11): 3132–3147. DOI: [10.1175/MWR3202.1](https://doi.org/10.1175/MWR3202.1).
- Hamill TM, Whitaker JS, Snyder C. 2001. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon Weather Rev* **129**(11): 2776–2790. DOI: [10.1175/1520-0493\(2001\)129<2776:DDFOBE>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2).
- Hamilton JD. 1994. *Time series analysis*, Princeton University Press, Princeton, NJ.
- Hernández-Pajares M, Juan JM, Sanz J. 1999. New approaches in global ionospheric determination using ground GPS data. *J Atmos Solar Terr Phys* **61**: 1237–1247.
- Hernández-Pajares M, Juan JM, Sanz J, Orus R, Garcia-Rigo A, Feltens J, Komjathy A, Schaer SC, Krankowski A. 2009. The IGS VTECmaps: a reliable source of ionospheric information since 1998. *Special IGS issue of J Geodesy* **83**(3–4): 263–275. DOI: [10.1007/s00190-008-266-1](https://doi.org/10.1007/s00190-008-266-1).
- Hernández-Pajares M, Roma-Dollase D, Krankowski A, García-Rigo A, Orús-Pérez R. 2017. Methodology and consistency of slant and vertical assessments for ionospheric electron content models. *J Geodesy* **91**(12): 1405–1414. DOI: [10.1007/s00190-017-1032-z](https://doi.org/10.1007/s00190-017-1032-z).
- Hunt BR, Kostelich EJ, Szunyogh I. 2007. Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D* **230**(1–2): 112–126. DOI: [10.1016/j.physd.2006.11.008](https://doi.org/10.1016/j.physd.2006.11.008).
- Kalman RE. 1960. A new approach to linear filtering and prediction problems. *Trans ASME J Basic Eng* **82**: 34–45.
- Kalnay E. 2005. *The future of data assimilation: 4D-Var or ensemble Kalman filter*. Available at <https://www.atmos.umd.edu/~ekalnay/pubs/4DVarEnKFKalnaySAMSI.ppt.pdf>.
- Kang J-S, Kalnay E, Miyoshi T, Liu J, Fung I. 2012. Estimation of surface carbon fluxes with an advanced data assimilation methodology. *J Geophys Res (Atmos)* **117**(D24): D24101. DOI: [10.1029/2012JD018259](https://doi.org/10.1029/2012JD018259).
- Le Dimet F-X, Talagrand O. 1986. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus* **38A**: 97–110.
- Lee IT, Matsuo T, Richmond AD, Liu JY, Wang W, Lin CH, Anderson JL, Chen MQ. 2012. Assimilation of FORMOSAT-3/COSMIC electron density profiles into a coupled thermosphere/ionosphere model using ensemble Kalman filtering. *J Geophys Res* **117**: A10318. DOI: [10.1029/2012JA017700](https://doi.org/10.1029/2012JA017700).
- Lee M-S, Kuo Y-H, Barker DM, Lim E. 2006. Incremental analysis updates initialization technique applied to 10-km MM5 and MM5 3DVAR. *Mon Weather Rev* **134**(5): 1389–1404. DOI: [10.1175/MWR3129.1](https://doi.org/10.1175/MWR3129.1).
- Li H, Kalnay E, Miyoshi T. 2009. Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. *Quart J Roy Meteor Soc* **135**(639): 523–533. DOI: [10.1002/qj.371](https://doi.org/10.1002/qj.371).
- Li Y, Huang Z, Zhou N, Lee B, Diao R, Du P. 2012. Application of ensemble Kalman filter in power system state tracking and sensitivity analysis. *PES T&D* **1–8**: 2012. DOI: [10.1109/TDC.2012.6281499](https://doi.org/10.1109/TDC.2012.6281499).
- Liang X, Zheng X, Zhang S, Wu G, Dai Y, Yong L. 2012. Maximum likelihood estimation of inflation factors on error covariance matrices for ensemble Kalman filter assimilation. *Quart J Roy Meteor Soc* **138**: 263–273.
- Llwelllyn SK, Bent RB. 1973. Documentation and description of the bent ionospheric model. *Tech. Rep.* Air Force Geophysics Laboratory, Hanscom AFB, MA.
- Lorenc AC, Bell RS, Macpherson B. 1991. The Meteorological Office analysis correction data assimilation scheme. *Quart J Roy Meteor Soc* **117**(497): 59–89. DOI: [10.1002/qj.49711749704](https://doi.org/10.1002/qj.49711749704).
- Macpherson B. 1991. Dynamic initialization by repeated insertion of data. *Quart J Roy Meteor Soc* **117**(501): 965–991. DOI: [10.1002/qj.49711750105](https://doi.org/10.1002/qj.49711750105).
- Mandel J, Beezley JD, Coen JL, Kim M. 2009. *Data assimilation for wildland fires*. **29**: 47–65.
- McNamara LF. 2009. Spatial correlations of foF2 deviations and their implications for global ionospheric models: 2. Digisondes in the United States, Europe, and South Africa. *Radio Sci* **44**: DOI: [10.1029/2008RS003956](https://doi.org/10.1029/2008RS003956).
- McNamara LF, Angling MJ, Elvidge S, Fridman SV, Hausman MA, Nickisch LJ, McKinnell L-AA. 2013. Assimilation procedures for updating ionospheric profiles below the F2 peak. *Radio Sci* **48**(2): 143–157. DOI: [10.1002/rds.20020](https://doi.org/10.1002/rds.20020).
- Metref S, Cosme E, Snyder C, Brasseur P. 2014. A non-Gaussian analysis scheme using rank histograms for ensemble data assimilation. *Nonlin Process Geophys* **21**: 869–885.
- Miyoshi T. 2011. The Gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform Kalman filter. *Mon Weather Rev* **139**(5): 1519–1535. DOI: [10.1175/2010MWR3570.1](https://doi.org/10.1175/2010MWR3570.1).
- Nava B, Coisson P, Radicella S. 2008. A new version of the NeQuick ionosphere electron density model. *J Atmos Sol Terr Phys* **70**(15): 1856–1862. DOI: [10.1016/j.jatosp.2008.01.015](https://doi.org/10.1016/j.jatosp.2008.01.015).
- Neale RB. 2010. *Description of the NCAR community atmosphere model*. TN-486+STR, National Centre for Atmospheric Research, Boulder, CO.
- Orus R, Hernandez-Pajares M, Juan JM, Sanz J, Garcia-Fernandez M. 2002. Performance of different TEC models to provide TEC ionospheric corrections. *J Atmos Solar Terr Phys* **64**: 2055–2062.
- Ott E, Hunt B, Szunyogh I, Zimin A, Kostelich E, Corazza M, Kalnay E, Yorke J. 2004. A local ensemble Kalman filter for atmospheric data assimilation. *Tellus* **56**: 415–428.
- Plackett RL. 1950. Some theorems in least squares. *Biometrika* **37**(1–2): 149–157. Available at <http://www.ncbi.nlm.nih.gov/pubmed/15420260>.
- Polavarapu S, Ren S, Clayton AM, Sankey D, Rochon Y. 2004. On the relationship between incremental analysis updating and incremental digital filtering. *Mon Weather Rev* **132**(10): 2495–2502. DOI: [10.1175/1520-0493\(2004\)32<2495:OTRBLA>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)32<2495:OTRBLA>2.0.CO;2).
- Qian L, Burns AG, Emery BA, Foster B, Lu G, Maute A, Richmond AD, Roble RG, Solomon SC, Wang W. 2014. The NCAR

- TIE-GCM: A community model of the coupled thermosphere/ionosphere system. In: *Modeling the ionosphere-thermosphere system*, Huba J, Schunk R, Khazanov G (Eds.), Geophysical Monograph Series, John Wiley & Sons, Washington. pp. 73–84. DOI: [10.1002/9781118704417.ch7](https://doi.org/10.1002/9781118704417.ch7).
- Reichle RH, Walker JP, Koster RD, Houser PR. 2002. Extended versus ensemble Kalman filtering for land data assimilation. *J Hydrometeorol* **3**: 728–740.
- Richmond AD, Ridley EC, Roble RG. 1992. A thermosphere/ionosphere general circulation model with coupled electrodynamics. *Geophys Res Lett* **19**: 601–604.
- Roble RG, Ridley EC, Richmond AD. 1988. A coupled thermosphere/ionosphere general circulation model. *Geophys Res Lett* **15** (88): 1325–1328.
- Scherliess L, Schunk RW, Sojka JJ, Thompson DC, Zhu L. 2006. Utah State University global assimilation of ionospheric measurements Gauss-Markov Kalman filter model of the ionosphere: Model description and validation. *J Geophys Res* **111**(A11): 315. DOI: [10.1029/2006JA011712](https://doi.org/10.1029/2006JA011712).
- Scherliess L, Thompson DC, Schunk RW. 2009. Ionospheric dynamics and drivers obtained from a physics-based data assimilation model. *Radio Sci* **44**: DOI: [10.1029/2008RS004068](https://doi.org/10.1029/2008RS004068).
- Schunk R, Scherliess L, Sojka JJ, Thompson DC, Zhu L. 2005. Ionospheric weather forecasting on the horizon. *Space Weather* **3** (8): S08007. DOI: [10.1029/2004SW000138](https://doi.org/10.1029/2004SW000138).
- Schunk RW, Scherliess L, Eccles V, Gardner LC, Sojka JJ, et al. 2016. Space weather forecasting with a Multimodel Ensemble Prediction System (MEPS). *Radio Sci* **51**(7): 1157–1165. DOI: [10.1002/2015RS005888](https://doi.org/10.1002/2015RS005888).
- Shim JS, Kuznetsova M, Rastaetter L, Bilitza D, Butala M, et al. 2012. CEDAR Electrodynamic Thermosphere Ionosphere (ETI) Challenge for systematic assessment of ionosphere/thermosphere models: Electron density, neutral density, NmF2, and hmF2 using space based observations. *Space Weather* **10**: DOI: [10.1029/2012SW000851](https://doi.org/10.1029/2012SW000851).
- Shim JS, Kuznetsova M, Rastaetter L, Hesse M, Bilitza D, et al. 2011. Electrodynamic Thermosphere Ionosphere (ETI) Challenge for systematic assessment of ionosphere/thermosphere models: NmF2, hmF2, and vertical drift using ground-based observations. *Space Weather* **9**. DOI: [10.1029/2011SW000727](https://doi.org/10.1029/2011SW000727).
- Szunyogh I. 2014. *Applicable atmospheric dynamics: Techniques for the exploration of atmospheric dynamics*. World Scientific, Singapore. ISBN 9814630578.
- Tapping KF. 2013. The 10.7 cm solar radio flux (F10.7). *Space Weather* **11**(7): 394–406.
- Wang C, Hajj GA, Pi X, Rosen G, Wilson B. 2004a. Development of the global assimilative model. *Rad Sci* **39**(RS1S06). DOI: [10.1029/2002RS002854](https://doi.org/10.1029/2002RS002854).
- Wang X, Bishop CH, Julier SJ. 2004b. Which is better, an ensemble of positive-negative pairs or a centered spherical simplex ensemble? *Mon Weather Rev* **132**: 1590–1605.
- Whitaker JS, Hamill TM, Wei X, Song Y, Toth Z. 2008. Ensemble data assimilation with the NCEP global forecast system. *Mon Weather Rev* **136**(2): 463–482. DOI: [10.1175/2007MWR2018.1](https://doi.org/10.1175/2007MWR2018.1).
- Yang S-C, Kalnay E, Hunt B, Bowler NE. 2009. Weight interpolation for efficient data assimilation with the local ensemble transform Kalman filter. *Quart J Roy Meteor Soc* **135**(638): 251–262. DOI: [10.1002/qj.353](https://doi.org/10.1002/qj.353).

**Cite this article as:** Elvidge S & Angling MJ 2019. Using the local ensemble Transform Kalman Filter for upper atmospheric modelling. *J. Space Weather Space Clim.* **9**, A30.