J. Space Weather Space Clim. 2023, **13**, 21
© D. Hodyss et al., Published by EDP Sciences 2023
https://doi.org/10.1051/swsc/2023019

JSWSC

Available online at:
www.swsc-journal.org

RESEARCH ARTICLE OPEN ACCESS

The effects of estimating a photoionization parameter within a physics-based model using data assimilation

Daniel Hodyss[1,*], Douglas R. Allen[1], Daniel Tyndall[2], Peter Caffrey[1], and Sarah E. McDonald[3]

[1] Remote Sensing Division, Naval Research Laboratory, 4555 Overlook Ave SW, Washington, DC 20375, USA
[2] Marine Meteorology Division, Naval Research Laboratory, Monterey, CA 93943, USA
[3] Space Science Division, Naval Research Laboratory, Washington, DC 20375, USA

Received 11 October 2022 / Accepted 4 July 2023

Abstract – Data assimilation (DA) is the process of merging information from prediction models with noisy observations to produce an estimate of the state of a physical system. In ionospheric physics-based models, the solar ionizing irradiance is commonly estimated from a solar index like F10.7. The goal of this work is to provide the fundamental understanding necessary to appreciate how a DA algorithm responds to estimating an external parameter driving the model's interpretation of this solar ionizing irradiance. Therefore, in this work we allow the DA system to find the F10.7 value that delivers the degree of photoionization that leads to a predicted electron density field that best matches the observations. To this end, we develop a heuristic model of the ionosphere along the magnetic equator that contains physics from solar forcing and recombination/plasma diffusion, which allows us to explore the impacts of strongly forced system dynamics on DA. This framework was carefully crafted to be both linear and Gaussian, which allows us to use a Kalman filter to clearly see how: (1) while recombination acts as a sink on the information in the initial condition for ionospheric field variables, recombination does not impact the information in parameter estimates in the same way, (2) when solar forcing dominates the electron density field, the prior covariance matrix becomes dominated by its leading eigenvector whose structure is directly related to that of the solar forcing, (3) estimation of parameters for forcing terms leads to a time-lag in the state estimate relative to the truth, (4) the performance of a DA system in this regime is determined by the relative dominance of solar forcing and recombination to that of the smaller-scale processes and (5) the most impactful observations on the electron density field and on the solar forcing parameter are those observations on the sunlit side of the ionosphere. These findings are then illustrated in a full physics-based ionospheric model using an ensemble Kalman filter DA scheme.

Keywords: Specification / Data assimilation / Kalman filter / Ensemble Kalman filter / Parameter estimation

1 Introduction

The ionosphere is a dynamic plasma enveloping the Earth that interacts directly with the neutral components of the atmosphere. The dynamical evolution of this mixture of ions, electrons and neutral molecules is well-known to be strongly governed by the external forcing being imposed on the ionosphere (Siscoe & Solomon, 2006, Codrescu et al., 2018). In physics-based ionospheric modelling, the various external forcings critical to predicting the evolution of the ionosphere are thought to be due to solar radiation and solar wind, geomagnetic and auroral dynamics, electrodynamic processes, and importantly wave motions in the neutral atmosphere (Rishbeth & Mendillo, 2001; Fang et al., 2018). In addition to these, there are also solar flares, solar cycle activity, and solar rotation that impact in one way or another the ionosphere through ionization. Gravity waves, thermospheric tides, and planetary waves in the neutral atmosphere influence the ionosphere directly through the transport of the ion/electron density fields and indirectly through electrodynamic processes. In addition to this direct impact through the neutrals, the thermospheric winds, temperature, and density also respond to solar and geomagnetic activity, which leads to changes in the amount of ionization as well as the efficiency of recombination processes.

The fact that ionospheric modelling is so strongly controlled by external forcing is the focus of this work, while at the same time, our overarching goal is to further the understanding of ionospheric prediction by focusing on problems related to the impact

*Corresponding author: daniel.hodyss@nrl.navy.mil

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

of the specification of external model parameters on ionospheric simulations. In general, the forecasting of a given state consists of three main components. First, there is a physical model that simulates all of the known physics, chemistry, dynamics, etc. of the ionosphere. This model consists of computer code that solves mathematical equations that describe known physical laws. This model of the ionosphere has a state vector defined as containing all the variables within the model as well as all the parameters used to specify external and internal processes. This model then takes as input both initial conditions as well as parameter values and propagates the ionospheric state vector forward in time. Second, there are observations of the state in question. These may be *in situ* observations or remotely sensed from satellite or ground-based measurements, and they may directly observe the model state or may observe variables that can be derived from the model state. Third, there is a data assimilation (DA) system that combines short-term model forecasts with these observations to produce initial conditions for the next forecast, as explained further below. The overall goal is to minimize the difference between future observations and the forecasts of the model state. It is important to keep all of these aspects in mind when building and evaluating a prediction system.

The focus of this manuscript is DA. DA, or what is sometimes referred to as "specification" in the space weather community, is a set of mathematical techniques that merge information from an ionospheric model with noisy observations to obtain an estimate of the state of the ionosphere that is generally better than either individually. DA occurs whenever a set of observations are available and at each of these times, the DA proceeds to use or "assimilate" this set of observations. This step in the process where observations are used is referred to as a DA "cycle". At each of these DA cycles, the key statistical quantity, that DA exploits to merge information from the ionospheric model with the latest set of noisy observations, is the correlation between the elements of the model's state vector and the observations. In ensemble-based DA, an ensemble of model runs, each with slightly different initial conditions (e.g., $E \times B$ drifts, ion densities, ion temperatures, electrostatic potential, etc.) and parameter values (e.g. F10.7, Ap, etc.), is used to predict this instantaneous correlation of state-vector elements with each of the observations. We emphasize that the use of this correlation between state-vector elements and the observed quantities allows for observations of one variable (say, total electron content [TEC]) to update all other variables, even those that are of a different type with different units. Once this information from correlations is available to the present cycle the DA system proceeds to update all parts of the state vector where there exists a non-zero correlation with the observations. We emphasize here that this correlation structure is not a fixed, climatological one, but an evolving, instantaneous estimate that is recalculated at every single cycle of the process. These correlation magnitudes may be stronger in some cycles and weaker or even non-existent in others. For example, in the ionosphere, strong correlations are often observed between TEC and the parameter we use to predict solar irradiance, i.e. F10.7. Bergot et al. (2013) showed correlations as a function of latitude of greater than 0.7, with a maximum correlation of 0.93 at latitudes of 10 and 20 N (see their Fig. 4). However, this correlation can vary with local time, with the season, with the phase of the solar cycle, and with geomagnetic activity. Allen et al. (2023) provide further examination of the correlations of F10.7 and TEC using the SAMI3 model and the Jet Propulsion Laboratory (JPL) Global Ionospheric Maps (GIM). The point here is that a good ensemble DA system that correctly includes the correlation of F10.7 and TEC will be able to discriminate the level of correlation in both time and space thereby properly adjusting the degree to which it updates state variables with the information from the observations.

Now, with all of that said, DA works best when the observations are used to inform the state variables that most strongly influence the evolution of the system. In tropospheric weather, there exists a sensitive dependence on forecasts to the variables constituting the initial condition such that using observational information to inform the initial condition provides enormous information content to the prediction of relevant field variables. By contrast, and as discussed earlier, the ionosphere is largely controlled by external forcing and not the initial condition. Hence, DA in the context of space weather should focus on using observational information to update those parameters in the model that most strongly influence its evolution through external forcing.

Parameter estimation for the external forcing within an ensemble DA system using an ionospheric physics-based model has been explored in several studies (Solomentsev et al., 2012; Matsuo et al., 2013; Morozov et al., 2013; Codrescu et al., 2018). Solomentsev et al. (2012) used an ensemble Kalman filter (EnKF) in which the state vector was augmented with additional parameters that describe an imposed $E \times B$ drift and neutral wind velocity on the electron density field. They used an iterative procedure applied to their EnKF and showed that they could recover these parameters using observations of total electron content (TEC). Matsuo et al. (2013) use the Ensemble Adjustment Kalman Filter (EAKF), which is a part of the Data Assimilation Research Testbed (DART), with the Thermosphere-Ionosphere-Electrodynamics General Circulation Model (TIEGCM) to assimilate Challenging Minisatellite Payload (CHAMP) neutral mass density measurements and electron density profiles from the COSMIC/FORMOSAT-3 mission. They updated several parameters including the F10.7 index and showed some improvement in certain configurations of their system. Morozov et al. (2013) also used the EAKF and DART with the Global Ionosphere-Thermosphere Model (GITM) to assimilate CHAMP neutral mass density measurements. They found that while the state estimates of F10.7 did not always converge to the true values, the GITM output did in fact become substantially closer to CHAMP measurements. Codrescu et al. (2018) set up a unique ensemble Kalman filter in which the posterior perturbations generating the ensemble were fixed at each cycle and were applied to the parameters of F10.7, the interplanetary magnetic field, as well as the speed and density of the solar wind. In addition to perturbing these parameters, Codrescu et al. (2018) estimated these same parameters using their ensemble Kalman filter and similar to Morozov et al. (2013) they showed that performance improvements can be obtained in the quality of state variables even when the parameters do not converge to observed values.

A fascinating result of this work is the performance improvements seen when the parameters did not actually converge to their true values. This pathological behaviour is well-known to occur when the forecast model used in the DA is not identical to the true model evolving reality forward in time. There are two ways a model can be flawed in this way. First, it could be that we are not certain of all of the laws of physics governing the evolution of a particular physical system of

interest. In this case, we are led to build a model that is missing some physics that should have been included in order to evolve the system correctly. Second, it could be that we do know the laws of physics perfectly but because of computational constraints, we cannot employ sufficient resolution to resolve all the phenomena of interest. When our model applies to either or both of these cases, simulations of the ionosphere will not evolve correctly even when the initial condition and parameters, like F10.7, are specified correctly. In fact, in this case, where the model is flawed, finding another value for a parameter like F10.7 that is not the true value is likely to produce a better simulation of the ionospheric field variables (e.g. electron density), which is precisely what Morozov et al. (2013) and Codrescu et al. (2018) found.

This manuscript aims to provide the fundamental understanding necessary to appreciate how and why the properties of estimating one of the drivers of an ionospheric model, namely the photoionization through the irradiance proxy, F10.7, matters to the prediction of the evolution of the ionospheric field variables. We view this problem as composed of two main questions. On the one hand, there is the question of how, in principle, the DA should behave when estimating photoionization from solar irradiance using a parameter like F10.7 and then using it in a *perfect* forecast model. On the other hand, there is the question of what it means to use DA to find an F10.7 value that differs from the true one yet still can produce a better mean prediction of the degree of photoionization happening in a *flawed* model. While both questions are of great interest, this manuscript focuses on the former by carefully constructing a "perfect" model scenario where the only model error present is the misspecification of the solar ionizing irradiance through the F10.7 parameter. This configuration implicitly assumes that the only error is that we have the wrong F10.7 value available, but the physical relationship between photoionization and that F10.7 value is correctly specified in the model. This is in contrast to the other scenario whereby the model is flawed and there is an error associated with using F10.7 as a proxy for the solar ionizing irradiance (Chamberlin et al., 2008). We then go on to show the behaviour and limitations of a DA system in this *best-case* scenario. We focus here on this best-case scenario of using a perfect forecast model because it is crucial to understand exactly how a DA system should behave in the perfect model scenario before one can fully understand and appreciate its behaviour when the model is flawed.

The rest of this manuscript is organized as follows. In Section 2 we build a heuristic model of the low-latitude ionosphere in order to study the estimation of parameters in a simple setting that allows full control of all aspects of the problem. In Section 3 we show that what we see in the simplified setting is robust by using an ensemble Kalman filter with a full physics-based ionospheric model. Lastly, in Section 4 we close the manuscript with a discussion of our main results and their implication for data assimilation in the ionosphere.

2 A Heuristic model of the equatorial ionosphere

In this section, we develop a heuristic model of vertical total electron content (vTEC) along a latitude circle in the equatorial region. The goal is to reveal the basic principles we will see in the more complex, physics-based model of Section 3.

2.1 Model

We begin by imagining a latitude circle along the magnetic equator and focus attention on the vTEC field. Further, imagine that vTEC is being impacted by solar forcing, recombination, and a myriad of second-order, essentially random processes that might move plasma in and out of the equatorial column. We note that while technically solar forcing and recombination act locally to influence the electron density field itself, this heuristic model of the ionosphere considers the vertically integrated influence of these processes.

We hereby denote the vTEC along this latitude circle as a column vector of length N, which we label $\mathbf{q} = \mathbf{q}(x, t)$, and write its evolution as a stochastic differential equation of the form

$$\mathrm{d}\mathbf{q} = F(t)g(\mathbf{x} - ct\mathbf{1})\mathrm{d}t - \frac{1}{\tau}\mathbf{q}\mathrm{d}t + \alpha_s \mathrm{d}\mathbf{S}. \tag{1}$$

The first term on the right-hand side is meant as a heuristic model of the production of electrons from solar ionizing radiation, where $\mathrm{d}t$ is the infinitesimal time increment, $F = F(t)$ is some proxy for solar irradiance that is to be thought of as something akin to an F10.7 index value, $g \geq \mathbf{0}$ is a structure-function represented in the form of a column vector that is N long that describes the sunlit region of the latitude circle (an example will be provided below), $\mathbf{x}$ is an N-vector denoting the distance along the latitude circle, and c is the speed of the sunlit region as it moves across the latitude circle. The second term on the right-hand side of (1) represents a very simple dissipation term that is intended to grossly model the recombination of ions and electrons and the dissipative effects of plasma diffusion along field lines. Both of these processes are thought of as combined into a dissipation term implying an exponential decay with rate τ. While a recombination term is more properly thought of as quadratic, writing it here as linear allows us to make our points in the clearest possible way because the statistical character of the system remains Gaussian and, as we will see, this allows for a careful analytic analysis of the problem. The impact of representing the recombination as quadratic in this model is described in Appendix A. In addition, we emphasize that this recombination term has a rate coefficient that is independent of the solar forcing parameter, F. In some regards, this is a significant assumption as the recombination is typically controlled in one way or another by the solar forcing through the thermosphere and hence this is yet another route that could lead the recombination term to be nonlinear in the prognostic variables. We chose a constant rate coefficient to again keep the physical system linear and therefore to allow for the analysis of Section 2.2, where we implement a Kalman filter which requires a linear/Gaussian system of equations. We will return to this assumption at several places below in order to comment on where there exist some minor qualitative differences from this form of recombination versus one that is a function of F. Lastly, the last term in equation (1) is meant to represent the collection of second-order processes that might move plasma in and out of an equatorial column. This heuristically includes the effects of the east-west $E \times B$ drifts as well as smaller-scale processes such as vertically propagating gravity waves in the

thermosphere. α_s is a parameter governing the amplitude of these processes and we simply represent this as Gaussian random noise of the form $\mathrm{d}s_j = w_j \mathrm{d}t$, where $\mathrm{d}s_j$ is the jth element of the N-vector $\mathrm{d}\mathbf{S}$ and w_j is drawn from $N(0, 1/\mathrm{d}t)$.

We also write the solar forcing parameter as varying through time according to a random walk as

$$\mathrm{d}F = \alpha_F w \mathrm{d}t \tag{2}$$

where α_F is a parameter that governs the size of the steps in (2) and w here is also drawn from $N(0, 1/dt)$.

We may solve (1) and (2), respectively, to obtain a broad understanding of the general characteristics of the solution to such equations, viz.

$$\mathbf{q}(t) = \mathbf{q}(0)\exp\left(-\frac{t}{\tau}\right) + \int_0^t \exp\left(\frac{t'-t}{\tau}\right) F(t') g(\mathbf{x} - ct'\mathbf{1})\mathrm{d}t' + \alpha_s \int_0^t \exp\left(\frac{t'-t}{\tau}\right)\mathrm{d}\mathbf{S}' \tag{3}$$

$$F(t) = F(0) + \alpha_F \int_0^t w \mathrm{d}t'. \tag{4}$$

First, note that in (3) the initial condition for the vTEC field is in $\mathbf{q}(0)$, which therefore would contain all the information derived from the observations up to time 0 in a potential DA system. Hence, if the time between DA cycles is longer than the recombination time, τ, then the information being provided from observations will be largely lost between each cycle. This shows that recombination is essentially a sink in the information in the initial condition that we are attempting to extract from the observations and place into the model's time evolution. On the other hand, the initial condition for the solar forcing parameter, $F(0)$, does not have a dissipation factor associated with recombination. Therefore, information derived from observations about the solar forcing parameter is likely to last longer than in the vTEC field. Second, the second term on the right-hand side shows how the solar forcing term drives the diurnal pattern on the vTEC field. The integral shows that this term has a memory of both past solar forcing values as well as where the sun was shining through the structure-function, $g(\mathbf{x} - ct\mathbf{1})$, but nevertheless the exponential weighting through time implies that it quickly forgets past values further back than the recombination time, τ. The last term on the right-hand side of (3) is a modified random walk in which the random walk forgets where it has been over the same time scale τ. Lastly, equation (4) is a standard random walk because equation (2) does not include any physical damping or forcing terms as in (1).

2.2 The Kalman Filter

We assume it reasonable to discretize (1) and (2) in time, denoted with index, n, and use the standard forward Euler scheme

$$\mathbf{q}_{n+1} = D\mathbf{q}_n + \Delta F_n g(\mathbf{x} - ct_n\mathbf{1}) + \Delta\alpha_s \mathbf{u}_n \tag{5a}$$

$$\mathrm{F}_{n+1} = F_n + \Delta\alpha_F \mathrm{v}_n \tag{5b}$$

where $\Delta \leq \tau$ is the time step and the noise terms in (5a) and (5b) are defined such that $\mathbf{u}_n$ is an N-vector and v_n is a scalar whose elements are re-drawn from $\mathrm{N}(0, 1/\Delta)$ at each time, n, and the parameters α_s and α_F determine the size of the noise forcing. Finally, the dissipation rate parameter is

$$D = 1 - \frac{\Delta}{\tau}. \tag{6}$$

Equations (5a) and (5b) are both linear and Gaussian. This implies that we have quite a few analytic tools available for understanding DA in this context. In the linear Gaussian setting we know that the Kalman filter is the optimal DA system. DA through the Kalman filter is comprised of two parts: the estimation of the state and the propagation of the error variance through time. These two parts are separated here and analyzed in detail in the next two sub-sections.

2.2.1 Mean state

Upon applying an expectation to (5a) and (5b) we find the evolution equation for the prior mean state (i.e. step before observations are assimilated) as

$$\langle \mathbf{q}_{n+1} \rangle = D\langle \mathbf{q}_n \rangle + \Delta\langle F_n \rangle g(\mathbf{x} - ct_n\mathbf{1}) \tag{7a}$$

$$\langle \mathrm{F}_{n+1} \rangle = \langle F_n \rangle. \tag{7b}$$

The first term on the right of (7a) reveals how information in the previous mean state moves from time n to n+1 and the inherent dissipation through the factor D. Note that when the time between observations is equal to the recombination time ($\Delta = \tau$) no information from the previous time moves to the next step. This fact that no information at all moves from one time to the next is simply a result of the approximation of this solution through the forward Euler scheme. If a higher-order numerical method is applied here a very small impact from the previous step can be felt even when the time step is approximately the size of the recombination time. In any event, the second term on the right of (7a) is the mean of the solar forcing term whose impact is simply from the prior mean solar forcing value at that time step. Also, note that the noise term in (5a) makes no contribution to the prior mean in (7a) because that term has zero mean and there are no nonlinear terms in (5a) and (5b) to "fold" that information back into the mean. Lastly, because of the simplicity of the random walk in (5b) equation (7b) shows that the mean of the solar forcing parameter is constant with time, but we emphasize that this is only true here because we have yet to add DA, which we do next.

For ease of presentation, we assume that observations are available at each time step of our model integration in (5a) and (5b). We will assimilate observations at the n+1time to obtain the posterior mean (indicated by superscript "s" for specification) as

$$\langle \mathbf{q}^s_{n+1} \rangle = \langle \mathbf{q}_{n+1} \rangle + \mathbf{G}_{n+1}\left[\mathbf{y}_{n+1} - \mathbf{H}\langle \mathbf{q}_{n+1} \rangle\right] \tag{8a}$$

$$\langle F^s_{n+1} \rangle = \langle F_{n+1} \rangle + \mathbf{Q}_{n+1}\left[\mathbf{y}_{n+1} - \mathbf{H}\langle \mathbf{q}_{n+1} \rangle\right] \tag{8b}$$

where there are N_o observations contained in the vector $\mathbf{y}_{n+1}$, $\mathbf{H}$ is an $N_o \times N$ matrix referred to as the observation operator, which describes the relationship between the observation and the state vector, and

$$\mathbf{G}_{n+1} = \mathbf{B}_{n+1}\mathbf{H}^T\left[\mathbf{H}\mathbf{B}_{n+1}\mathbf{H}^T + \mathbf{R}\right]^{-1} \tag{9a}$$

$$\mathbf{Q}_{n+1} = \mathbf{C}_{n+1}\mathbf{H}^T\left[\mathbf{H}\mathbf{B}_{n+1}\mathbf{H}^T + \mathbf{R}\right]^{-1} \tag{9b}$$

where we define the relevant covariance matrices as

$$\mathbf{B}_{n+1} = \left\langle \left(\mathbf{q}_{n+1} - \langle \mathbf{q}_{n+1}\rangle\right)\left(\mathbf{q}_{n+1} - \langle \mathbf{q}_{n+1}\rangle\right)^T \right\rangle \tag{9c}$$

$$\mathbf{C}_{n+1} = \left\langle \left(F_{n+1} - \langle F_{n+1}\rangle\right)\left(\mathbf{q}_{n+1} - \langle \mathbf{q}_{n+1}\rangle\right)^T \right\rangle. \tag{9d}$$

The matrix $\mathbf{B}_{n+1}$ is the covariance matrix of the vTEC ($\mathbf{q}_{n+1}$) and the row-vector $\mathbf{C}_{n+1}$ describes the covariance between F_{n+1} and $\mathbf{q}_{n+1}$ at each grid point in the domain.

Given (8a) and (8b) we may use (7a) and (7b) to obtain the time evolution equations for the posterior means, viz.

$$\langle \mathbf{q}^s_{n+1}\rangle = D(\mathbf{I} - \mathbf{G}_{n+1}\mathbf{H})\langle \mathbf{q}^s_n\rangle + \Delta(\mathbf{I} - \mathbf{G}_{n+1}\mathbf{H})g(\mathbf{x} - ct_n\mathbf{1})\langle F^s_n\rangle + \mathbf{G}_{n+1}\mathbf{y}_{n+1} \tag{10a}$$

$$\langle F^s_{n+1}\rangle = [1 - \Delta\mathbf{Q}_{n+1}\mathbf{H}g(\mathbf{x} - ct_n\mathbf{1})]\langle F^s_n\rangle + \mathbf{Q}_{n+1}\left[\mathbf{y}_{n+1} - D\mathbf{H}\langle \mathbf{q}^s_n\rangle\right]. \tag{10b}$$

Equations (10a) and (10b) constitute the evolution equations for the posterior means upon assimilating a given set of observations. What is interesting about these equations is that they broadly take the form of an autoregressive process of lag-1 [AR(1)]. An AR(1) process is characterized by the presence of a term 1-step in the past and a noise term, which in this case takes the form of noise in the observations. This is particularly interesting for the update of the solar forcing in (10b) because, for sufficient numbers of observations, the quantity $\Delta\mathbf{Q}_{n+1}\mathbf{H}g(\mathbf{x} - ct_n\mathbf{1})$ is approximately a constant after the spin-up of the DA system (not shown). The fact that it is approximately constant with time implies a phase lag in the estimate of the solar forcing, which we will see later in the numerical experiments. See Appendix B for further discussion.

Another critical component of the update of the solar forcing, F, in (10b) is the prior covariance between the solar forcing and the vTEC field in (9d), which is a major component of $\mathbf{Q}_{n+1}$, (9b). Note that a perturbation from the mean in vTEC evolves according to

$$\boldsymbol{\varepsilon}^q_{n+1} = D\boldsymbol{\varepsilon}^q_n + \Delta\varepsilon^F_n g(\mathbf{x} - ct_n\mathbf{1}) + \alpha_s\Delta\mathbf{u}_n \tag{11}$$

where $\boldsymbol{\varepsilon}^q_n = \mathbf{q}_n - \langle \mathbf{q}_n\rangle$ and $\varepsilon^F_n = F_n - \langle F_n\rangle$. After the use of (11), and the use of similar equations for $\boldsymbol{\varepsilon}^q_n$ at other past times, in (9d) we find the prior covariance between the state and the solar forcing if no DA has yet happened as

$$\mathbf{C}_{n+1} = \Delta\sigma^2_n\left[g(\mathbf{x} - ct_n\mathbf{1})^T + D\frac{\sigma^2_{n-1}}{\sigma^2_n}g(\mathbf{x} - ct_{n-1}\mathbf{1})^T + D^2\frac{\sigma^2_{n-2}}{\sigma^2_n}g(\mathbf{x} - ct_{n-2}\mathbf{1})^T + \cdots\right]. \tag{12}$$

Note that because $D < 1$ and because a random walk has the property that $\sigma^2_n > \sigma^2_{n-1} > \sigma^2_{n-2} > \cdots > \sigma^2_0$ (as will be seen in equation (13b)) the series of terms in brackets in (12) are rapidly decaying for large n. Each of these terms corresponds to the state of the solar forcing at times starting at n and backward from there. This implies that the amplitude and structure of the covariance between the solar forcing at time n+1 and the vTEC at time n+1 is determined by the solar forcing in the past starting at time n and looking backwards in time. Note further that each of these terms is greater than or equal to zero and therefore the covariance between the solar forcing and the vTEC is either zero or positive. This immediately implies that increases in solar forcing always increase vTEC and decreases in solar forcing always decrease vTEC. We note here however that numerical experiments (not shown here) with this heuristic model but where we modify the recombination term to have a rate coefficient that is proportional to F leads to negative regions within the vector $\mathbf{C}_{n+1}$. We will see later on in Section 3 that the physics-based model, which does in fact have a recombination rate that is a function of the F10.7, also creates negative correlations between the F10.7 and the vTEC and this suggests that the reason is because of the F10.7 dependence in the recombination rate.

2.2.2 Error covariance

The prior error covariance equations for the models in (5a) and (5b) are,

$$\mathbf{B}_{n+1} = D^2\mathbf{B}_n + \mathbf{S}_n + \mathbf{N}_n + \mathbf{M}_n \tag{13a}$$

$$\sigma^2_{n+1} = \sigma^2_n + \alpha^2_F\Delta \tag{13b}$$

where

$$\mathbf{C}_{n+1} = D\mathbf{C}_n + \sigma^2_n\Delta g(\mathbf{x} - ct_n\mathbf{1})^T \tag{13c}$$

$$\mathbf{S}_n = \sigma^2_n\Delta g(\mathbf{x} - ct_n\mathbf{1})g(\mathbf{x} - ct_n\mathbf{1})^T \tag{14a}$$

$$\mathbf{N}_n = \alpha^2_s\Delta\mathbf{I} \tag{14b}$$

$$\mathbf{M}_n = \Delta D\left[g(\mathbf{x} - ct_n\mathbf{1})\mathbf{C}_n + \mathbf{C}^T_n g(\mathbf{x} - ct_n\mathbf{1})^T\right] \tag{14c}$$

$$\sigma^2_n = \left\langle (F_n - \langle F_n\rangle)^2\right\rangle. \tag{14d}$$

There are four important features of (13a) that we would like to emphasize. First, the dissipation factor, D, is now squared, while for the prior mean in (7a) dissipation is linear in D. Because $D < 1$ this means the dissipation induced by recombination is faster here than in the prior mean evolution equation. Second, the solar forcing drives the matrix, $\mathbf{S}_n$, which is constructed from the outer product of the structure-function, $g(\boldsymbol{x} - ct_n\mathbf{1})$. A matrix constructed from an outer product of a vector can have only one non-zero eigenvalue. Hence, $\mathbf{S}_n$ has a rank of 1 and therefore when solar forcing dominates it drives the prior covariance matrix to have a single dominant eigenvalue. This is important because the number of non-zero eigenvalues in the prior covariance matrix is one factor that impacts the amount of information derived from a given set of observations and having few dominant eigenvalues leads to very strong impacts from observations. Third, the covariance between the solar forcing and the vTEC comes in through the matrix, $\mathbf{M}_n$. Fourth, while the noise term in (5a) did not contribute directly to the prior or posterior mean the same noise does contribute directly to the prior covariance in (13a) through $\mathbf{N}_n$. Because we constructed this noise as white in space it results in essentially the opposite of the solar forcing term as this matrix is

full-rank and of the form of a scaled identity. Hence, the resulting structure in the prior covariance matrix is a delicate balance between the strength of the recombination versus the solar forcing and this noise in the system.

The evolution of the prior variance of the solar forcing in (13b) is considerably simpler than that of (13a). In a random walk, the variance increases linearly with time as can be seen in (13b) as the variance forcing term is linear in the time step. However, while (13b) is independent of the variance of the vTEC field, the vTEC variance in (13a) depends on the variance of solar forcing through the solar forcing term, $\mathbf{S}_n$. Lastly, (13c) describes the covariance between the vTEC and the solar forcing, F, and that covariance structure is largely determined by the structure-function, g, and again because $g \geq 0$ this implies that $\mathbf{C}_{n+1} \geq 0$.

As we will see in the numerical experiments to come, the variances in this problem setup are approximately constant after an initial spin-up period. This is a commonly understood feature of the solutions to the Kalman equations for autonomous systems of equations (i.e. system equations with time-independent coefficients). The equation in (1) is in general non-autonomous because of the solar forcing term, but nevertheless, there are interesting symmetries that allow us to see situations where the covariance matrices could become steady. We examine some of those situations next.

Equation (13a) has an interesting symmetry if we transform to a moving frame ($\mathbf{x}' = \mathbf{x} - ct\mathbf{1}$) and subsequently search for a steady-state ($\mathbf{B} = \mathbf{B}_{n+1} = \mathbf{B}_n$), viz.

$$\mathbf{B} = \frac{1}{1-D^2}\left[\mathbf{S}(\mathbf{x}') + \mathbf{N} + \mathbf{M}(\mathbf{x}')\right] \tag{15}$$

where we have dropped the subscripts on **S**, **N,** and **M** because **S** and **M** are independent of time in the moving frame and, as seen in (14b), **N** is always independent of time. Hence, while the prior covariance is constantly evolving with the diurnal cycle, in a frame moving with the diurnal cycle this model's resulting prior covariance is purely a constant through time. Note further that the presence of this translating steady-state also implies that the trace of (13a) is also a constant even without the transformation to a moving frame. Hence, global summary statistics as we will see in the numerical experiments below are likely to find steady states.

The Kalman posterior error covariance equations for vTEC and the solar forcing, respectively, are

$$\mathbf{P}^s_{n+1} = \mathbf{B}_{n+1} - \mathbf{G}_{n+1}\mathbf{H}\mathbf{B}_{n+1} \tag{16a}$$

$$\Sigma^s_{n+1} = \sigma^2_{n+1} - \mathbf{Q}_{n+1}\mathbf{H}\mathbf{C}^T_{n+1} \tag{16b}$$

$$\mathbf{C}^s_{n+1} = \mathbf{C}_{n+1} - \mathbf{Q}_{n+1}\mathbf{H}\mathbf{B}_{n+1} \tag{16c}$$

where we emphasize that the result of (16b) is a scalar and (16c) is a row vector. These equations describe the expected reduction in error from a given set of observations, where (16a) describes the reduction in error in vTEC, (16b) describes the reduction in error in the parameter F, and (16c) reveals how the covariance between F and vTEC changes owing to the corrections made by the latest set of observations.

Again, the DA process is to replace $\mathbf{B}_n$ with $\mathbf{P}^s_n$ at each assimilation time. We do this to (16a) while making use of

$$\mathbf{P}^s_{n+1}\mathbf{H}^T = \mathbf{G}_{n+1}\mathbf{R} \tag{17}$$

to obtain

$$\mathbf{P}^s_{n+1} = \left[D^2\mathbf{P}^s_n + \mathbf{S}_n + \mathbf{N}_n + \mathbf{M}_n\right] \left[\mathbf{I} + \left(D^2\mathbf{P}^s_n + \mathbf{S}_n + \mathbf{N}_n + \mathbf{M}_n\right)\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right]^{-1}. \tag{18}$$

While this equation is complicated by the inverse we can reveal the presence of a steady state in the limit where the dissipation is strong ($D \ll 1$) and again using the frame moving with the Sun. In that limit, we see that the observation locations in this equation, as described by the observation operator, **H**, control the emergence of a steady state posterior covariance. In a regime where $D \ll 1$, we find a steady-state when we observe all possible state locations ($\mathbf{H} = \mathbf{I}$) to find

$$\mathbf{P}^s = \left[\mathbf{S} + \mathbf{N} + \mathbf{M}\right]\left[\mathbf{I} + (\mathbf{S} + \mathbf{N} + \mathbf{M})\mathbf{R}^{-1}\right]^{-1} \tag{19}$$

which in the limit of large observation error reduces to (15) because in that limit the observations provide no information and therefore the posterior covariance must reduce to the prior covariance.

In terms of the posterior variance of the solar forcing parameter, F, it will prove illuminating to examine (16b) in two different observation configurations. In the first, we examine the resulting posterior variance from a vTEC observation at a single point along the latitude circle. Note that in this case, **H** is a row vector. Hence, $\mathbf{H}\mathbf{C}_{n+1}$ selects a single element at the location, x_0, of the observation. In this case, the posterior variance is

$$\Sigma^{s1}_{n+1} = \sigma^2_{n+1} - \frac{\left(\Delta\sigma^2_n\beta\right)^2}{b_0 + r} \tag{20}$$

where r is the observation error variance for this observation, b_0 is the scalar $\mathbf{H}\mathbf{B}_{n+1}\,\mathbf{H}^T$, and

$$\beta = g(x_0 - ct_n) + D\frac{\sigma^2_{n-1}}{\sigma^2_n}g(x_0 - ct_{n-1}) + D^2\frac{\sigma^2_{n-2}}{\sigma^2_n}g(x_0 - ct_{n-2}) + \cdots. \tag{21}$$

We further assume that dissipation is strong ($D \ll 1$) and the noise term, $\mathbf{N}_n$ is weak such that the prior covariance reduces to the solar forcing, $\mathbf{S}_n$ and (21) is reasonably approximated as $\beta \approx g(x_0 - ct_n)$. In this scenario, and with $r = 0$, (20) reduces to

$$\Sigma^{s1}_{n+1} = \sigma^2_{n+1} - \left(\Delta\sigma^2_n\right)^2. \tag{22}$$

In the second observation configuration we observe every point ($\mathbf{H} = \mathbf{I}$). In this situation we have

$$\Sigma^s_{n+1} = \sigma^2_{n+1} - \mathbf{C}_{n+1}[\mathbf{B}_{n+1} + \mathbf{R}]^{-1}\mathbf{C}^T_{n+1}. \tag{23}$$

Again, assume that dissipation is strong ($D \ll 1$) and the noise term, $\mathbf{N}_n$ is weak such that the prior covariance reduces to the solar forcing, $\mathbf{S}_n$. In this case (13c) is reasonably approximated as

$$\mathbf{C}_{n+1} \approx \Delta\sigma^2_n g(\mathbf{x} - ct_n\mathbf{1})^T. \tag{24}$$

Again, in the scenario where $r = 0$, and we interpret the inverse in (23) as the pseudo-inverse, we see (23) reduce precisely to (22). Hence, in terms of the update of the solar forcing

parameter, F, when solar forcing dominates then observing at one location will be approximately the same as observing everywhere, since in this case, the vTEC field will have strong spatial correlations.

2.3 Numerical experiments

In this section, we perform numerical experiments to illustrate the behaviors we have just discussed as well as thoroughly explore the possible solution space. We first integrate the mean and covariances forward to the next set of observations using (7a) and (7b) for the mean and (13a)–(13c) for the covariances. The result of these integrations is the prior means and covariances required to evaluate the Kalman filter equations of (8a) and (8b) and (16a)–(16c) for the posterior means and covariances. Once we have these updated values we return to (7a) and (7b) and (13a)–(13c) and we repeat the process. Note that the amplitude of the solution to (5a) and (5b) depends on the recombination time, τ. To compare simulations with different recombination times, we set the observation error variance as, $\mathbf{R} = r^2\mathbf{I}$, where r is a fixed constant determined as a percentage of the spatial average of the true vTEC at the final time. This allows for a self-consistent comparison of DA experiments with different recombination times. We note that the initial condition for the forcing parameter, F, is arbitrarily assigned an initial value of 1 and represents the production in units of vTEC over time. In addition, the time step is also an arbitrary value of 1. The subsequent units for $\mathbf{q}$ are then based on these relatively arbitrary values and therefore do not correspond to common units of TECU. We remind the reader that this is a heuristic model and we are interested in the general behaviors and interrelationships among variables. Lastly, in the experiment shown here, we use a state vector of 100 points and an observational network of 10 equally spaced point observations along the latitude circle starting at the prime meridian.

Figure 1 shows an example DA experiment using the model of (7a) and (7b) and (13a)–(13c) for the assimilation and (5a) and (5b) to generate the truth. In this experiment, we set τ = 6-time units, Δ = 1-time unit, $\alpha_s = 0.05$, $\alpha_F = 0.025$, and the observation error is set at 100% of the spatial average of the true final time field, which equates to setting the observation error variance to $r = 2.35$. The structure function, g, is defined as the uniformly westward translating, positive portion of the wave-1 sine function on the latitude circle, which effectively creates the daylit portion of the latitude circle as can be seen in Figure 1a. While the time step is of course also heuristic we set it to 1 unit to be broadly equivalent to 1 h, such that there are 24 time steps in a day, and we integrate out to 10 days. In Figure 1a we see the shape of the simulated vTEC field as it looks at the final time ($t = 240$). This structure is the shape of the vTEC field after the initial spin-up (i.e. after it has come into equilibrium between solar forcing and recombination), but nevertheless, this structure rotates with the Sun and traverses the globe from east to west. An important feature of the vTEC field is that it follows the solar forcing with its peak always to the east of local noon and, in additional experiments not shown here, we have seen that the degree to which it is shifted to the east of local noon is dependent on the recombination time, τ. We will see that the prior and posterior covariances have this phase lag with respect to local noon as well. Phase lags in ionospheric variables have been previously found in several studies (Schmölter & Bedermann, 2021; Vaishnav et al., 2021a) and shown to be controlled by eddy diffusivity rates (Vaishnav et al., 2021b). In any event, the prior and posterior means generally have the structure of the true state except that they are missing the small scales as was already discussed in regard to equations (7a) and (10a).

Solar forcing impacts the amplitude of the vTEC field. This can be seen in Figures 1c and 1d. In Figures 1c (1d) we show the maximum value of $\mathbf{q}$ (the solar forcing, F) for the truth as well as the prior and posterior means. There are two points to take away from Figures 1c and 1d. First, the maximum value of $\mathbf{q}$ and the parameter F largely follow each other's movements. Second, there exists a phase lag between the truth and the prior/posterior means in both the maximum $\mathbf{q}$ and the F. Note however that a careful examination of Figures 1a and 1b does not reveal any phase lag in the structure of $\mathbf{q}$ for the truth relative to the prior or posterior mean even though data assimilation is being done simultaneously for the $\mathbf{q}$-field as it is for the parameter F. For example, the structure seen in Figure 1a is uniformly translated from right to left. If a phase lag were to be present it would be seen as a shift to the right by the prior (blue) or posterior (red) with respect to the truth (black).

This phase lag is therefore only found in the amplitude of $\mathbf{q}$ and because the parameter F determines the amplitude of $\mathbf{q}$ this phase lag arises from the F through the posterior update equation (10b). Please see Appendix B for an analysis showing how this type of equation leads to a phase lag. Lastly, in Figures 1e and 1f we show both the domain-averaged prior and posterior variances as well as the mean-squared error with respect to the truth. Note that the error in both are quite similar and peak at very similar times, which suggests that the largest errors are dominated by those errors in F. Additionally, note that the prior/posterior variances are very nearly constant after the initial spin-up period. As mentioned previously, in a Kalman filter this is normally the case with autonomous physical systems, which is technically not a property of equation (5a) because it has the time-dependent solar forcing term. However, in this system what is important is having a sufficient number of observations such that there are always several within the sunlit side of the domain (i.e. within the non-zero portion of the function g). Experiments not shown here using a single observation in the centre of the domain show that the prior and posterior variances then become a strong function of time and this functional dependence is clearly seen to be determined by whether the observation location is presently within the daylit side of the domain. This controlling influence of the function g can be most easily seen by noting that in equation (12), and the small D limit, the covariance between $\mathbf{q}$ and F reduces to (13c), which implies that the covariance between $\mathbf{q}$ and F is small on the nighttime side of the domain. Even when D is relatively large, D is nevertheless less than 1 and so the remaining terms in the expansion (12) are still small, which again results in a small covariance on the nighttime side of the domain. Therefore, the most impactful observations in the estimation of F are those on the sunlit side of the domain. Lastly, under these same conditions, a careful examination of (13a) reveals that when $\mathbf{S}_n$ dominates, observations of $\mathbf{q}$ on the nighttime side of the domain also will have very little impact through the gain, $\mathbf{G}$, on the update of $\mathbf{q}$ itself. Therefore, the most impactful observations for both the vTEC field, $\mathbf{q}$, and the solar forcing, F, are those that are presently on the daylit side of the domain.

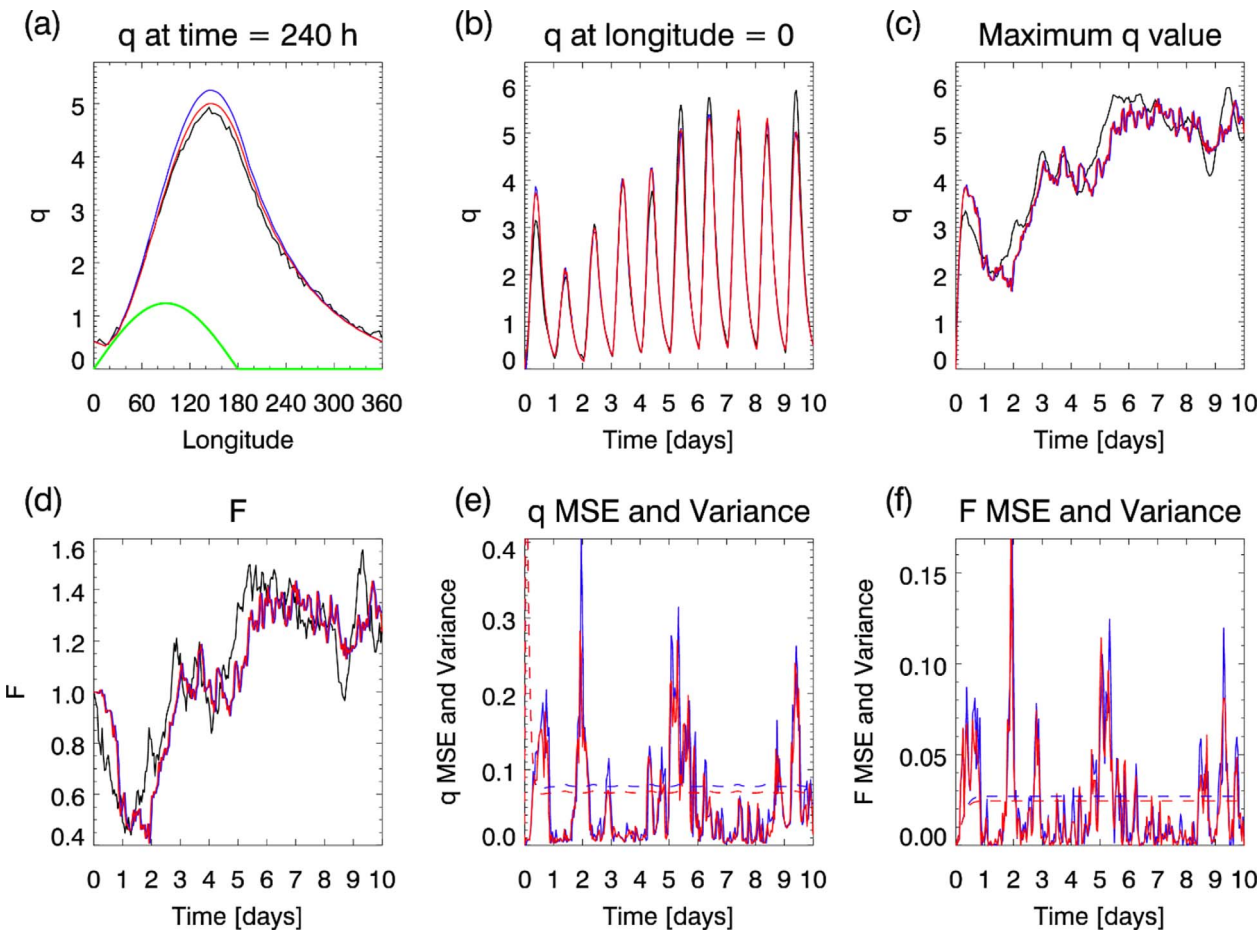

Fig. 1. DA with the Kalman Filter. a) The true state q (black), prior mean q (blue), and posterior mean q (red) at the final assimilation time. The green curve is the solar forcing function, g, also at the final time. b) The state for the grid point at longitude = 0 as a function of time; same colour scheme as in (a). c) The maximum value of q as a function of time; same colour scheme as in (a). d) Solar forcing, F, as a function of time; same colour scheme as in (a). e) Spatially averaged error squared error with respect to the truth for q (solid) for the prior (blue) and posterior (red). The spatially averaged variance prediction from the Kalman filter (dashed) for the prior (blue) and posterior (red). f) Same as (e) but for the solar forcing parameter, F.

Another way to see the interplay between the movement of the solar forcing and the DA is to look at the prior covariance matrices after the system has been assimilating observations for some time. In Figure 2 we show the prior covariance matrices for $\tau = 6$ and $r = 0.235$, 1.18, and 2.35, which corresponds to 10, 50, and 100% of the final time truth, respectively. These matrices show the two main properties we see in equation (13a): large scales from the solar forcing term and smaller scales from the noise term. This is most easily seen in Figure 2a, where we can see that the daylit portion of the domain has the main mass of contours, but the nighttime side of the domain has only a narrow diagonal. The larger scale portion arises from (14a) while the smaller scale, diagonal component arises from (14b). This reveals a strong diurnal dependence on the correlation length scales, with the daytime length scales being much longer than those at night. In Figures 2d–2f we show the function, g, denoting the daylit portion of the domain. By comparing the peak of the function g to the prior covariance matrix in Figures 2a–2c one can clearly see that the peak uncertainty lags local noon and that the lag increases with observation error, r. Lastly, in Figures 2d–2f is shown the leading eigenvector of each prior covariance matrix. The leading eigenvector [sometimes called an Empirical Orthogonal Function (EOF)] is peaked at the location of the peak in the prior covariance matrix itself and is therefore lagged with respect to local noon. Each of these EOFs represents more than 95% of the variability in the system as described by the prior covariance. Furthermore, for small observation error, the EOF structure is very nearly that of, g, while for larger observation errors there exists a "tail" of past solar forcing remaining in the system.

This phase lag that we see in the various components of the DA algorithm manifests itself in the posterior means of the solar forcing (F) estimates as already shown in Figures 1c and 1d. In Figure 3 we quantify this phase lag as a function of recombination time and observation error. In Figure 3a we show the time of the peak correlation between the posterior mean and the truth, while in Figure 3b we show the value of the correlation at its peak. Figure 3a shows that the time lag in the state estimate of the parameter increases as the observation error increases and as the recombination time increases. We have confirmed this phase lag by also creating this same figure by calculating the mean squared error (MSE) at various time lags and finding

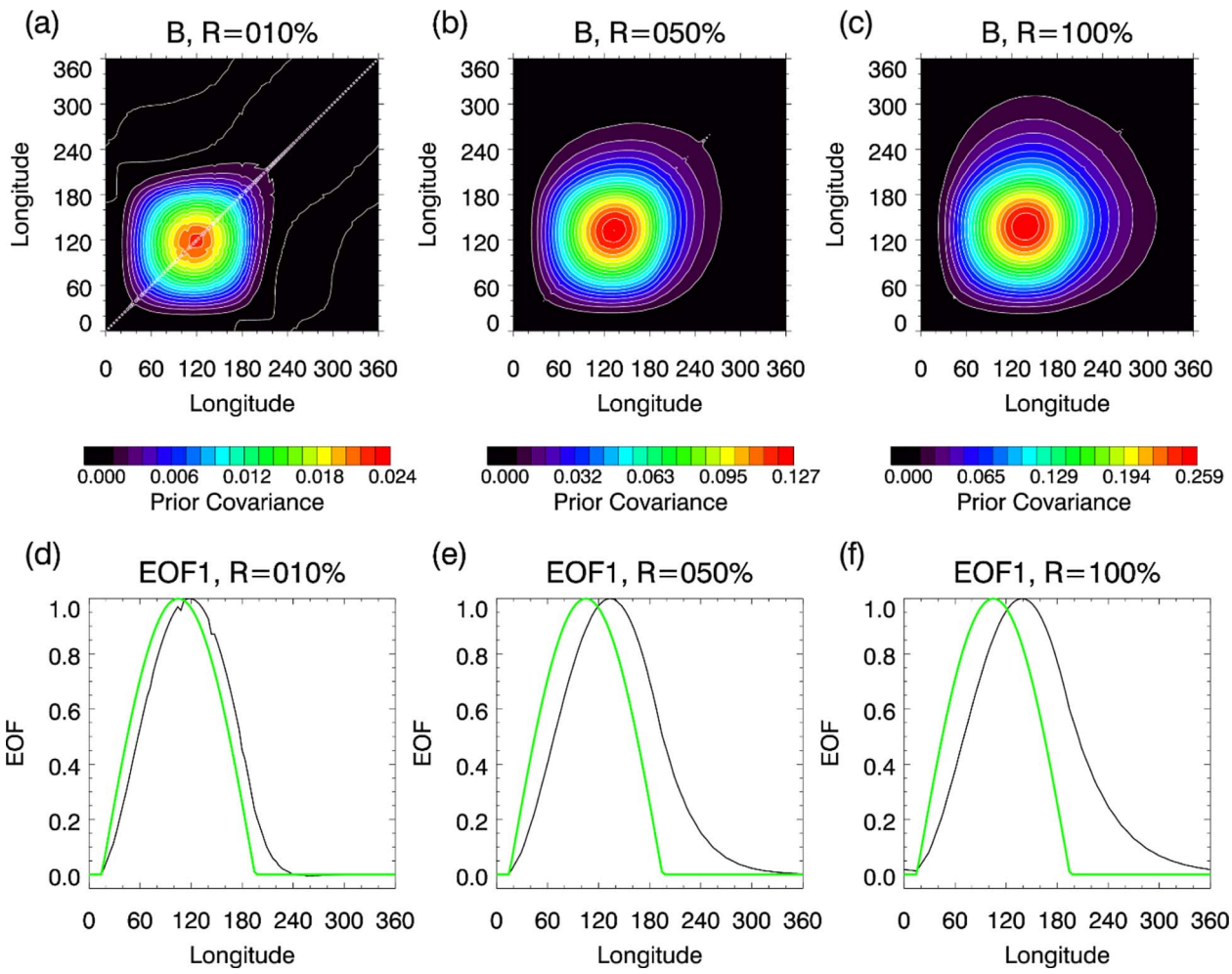

Fig. 2. Prior covariance matrices and EOFs at $t = 240$ from the Kalman Filter. In (a), (b), and (c) are the prior covariance matrices for $\tau = 6$ and observation errors of 10, 50, and 100% of the truth, respectively. In (d), (e), and (f) the black line is the leading eigenfunction of the prior covariance matrix for $\tau = 6$ and observation error of 10, 50, and 100%, respectively, while the green line is the function, g, at the time that these covariance matrices are valid.

nearly identical curves (not shown). Lastly, we again emphasize that in Figures 1a and 1b that there is no phase lag in longitude between the analysis and the truth. Hence, this phase lag in the solar forcing parameter leads to an inconsistency between the vTEC field, which does not have a longitudinal phase lag, and the solar forcing parameter.

3 Physics-based model and DA

In this section, we describe the development of an ensemble-based Kalman filter data assimilation system coupled to a physics-based ionospheric model.

3.1 System description

The physics-based ionospheric model we use is that of Huba et al. (2000) and is commonly referred to as Sami is Also a Model of the Ionosphere (SAMI3). SAMI3 simulates the transport and chemistry of seven ion species and solves for the electron/ion temperatures for three species on a grid aligned along the geomagnetic field lines. This grid is an eccentric dipole system aligned along geomagnetic field lines such that there are 160 points along a field line along with two more coordinates orthogonal to this field line where the approximately "longitudinal" direction has 90 points and the approximately vertical direction has 160. The momentum equation is split into motions along and across these geomagnetic field lines. Further details on the solver can be obtained from previous studies (see Zawdie et al., 2020 and references therein). The neutral temperature and composition are from NRLMSISE-00 (Picone et al., 2002) and the thermospheric winds are from the Horizontal Wind Model (HWM14; Drob et al., 2015).

In order to clearly make a connection to Section 2 it is important to understand how the SAMI3 model produces its estimate of the solar flux which is then used as a predictor of the photoionization. First, the estimated F10.7 value is used as a proxy to form

$$P = \frac{F + F_a}{2} \tag{25}$$

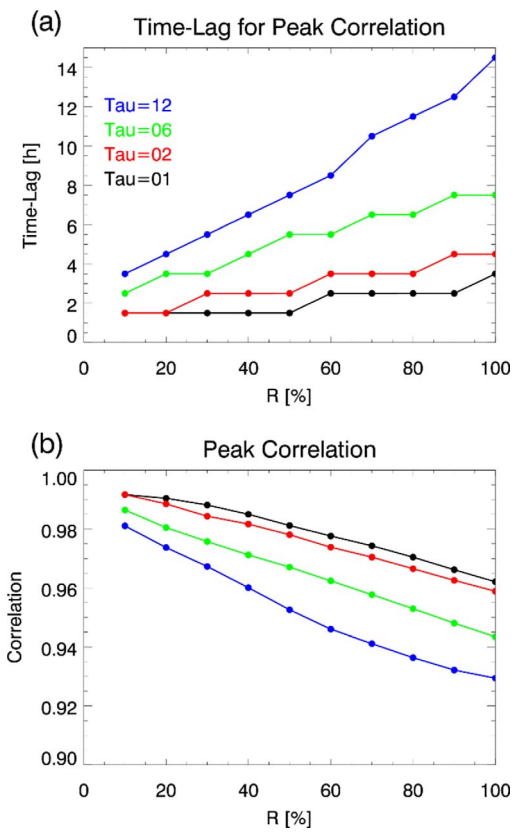

Fig. 3. (a) Time lag between the peak correlations between truth and the posterior mean as a function of recombination time and observation error. (b) Peak correlation value as a function of recombination time and observation error.

where F is the F10.7 value and F_a is the 81-day averaged F10.7 value. This proxy is then used in the EUVAC model developed by Richards et al. (1994) to estimate the solar flux, s_i, as

$$s_i = F74113_i[1 + A_i(P - 80)] \tag{26}$$

where $F74113_i$ is the modified solar minimum reference flux, A_i is a scaling factor, and the subscript i denotes one of 37 wavelength bands. Note that the solar flux is simply a linear function of this F10.7 value. Hence, from the perspective of the DA, there is little difference between directly estimating the solar flux in the DA algorithm or simply estimating the F10.7 parameter in the DA algorithm (as we do here) and subsequently using (26) to find the solar flux value.

The DA system we use is an ensemble-based Kalman filter, which has two main components. The first component is the update of the posterior mean using the observations, viz.

$$\bar{\mathbf{z}}^s_{n+1} = \bar{\mathbf{z}}_{n+1} + \mathbf{G}_n\left[\mathbf{y}_{n+1} - \mathbf{H}\bar{\mathbf{z}}_{n+1}\right] \tag{27}$$

where $\bar{\mathbf{z}}_{n+1}$ denotes the prior mean and $\bar{\mathbf{z}}^s_{n+1}$ denotes the posterior mean. The expression for the Kalman gain used here is identical to that in (9a) but now we obtain the prior covariance from an ensemble of SAMI3 runs in which we denote the entire state vector as a column vector, $\mathbf{z}^{(i)}_{n+1}$, where

$$\mathbf{B}_{n+1} = \frac{1}{N_e - 1}\mathbf{Z}_{n+1}\mathbf{Z}^T_{n+1}, \tag{28a}$$

$$\mathbf{Z}_{n+1} = \left[\mathbf{z}^{(1)}_{n+1} - \bar{\mathbf{z}}_{n+1} \quad \mathbf{z}^{(2)}_{n+1} - \bar{\mathbf{z}}_{n+1} \quad \cdots \quad \mathbf{z}^{(N_e)}_{n+1} - \bar{\mathbf{z}}_{n+1}\right]. \tag{28b}$$

The superscript in (28b) identifies an ensemble member out of N_e total members. For SAMI3, the vector $\mathbf{z}^{(i)}_{n+1}$ is constructed from the concatenation of the following fields: electrostatic potential, ion density for O^+, NO^+, O^+_2, N^+, N^+_2, H^+, and He^+, ion velocity along the field lines for each ion, ion temperature for each ion, electron temperature, the two components of the "$E \times B$ drift" (the cross-product of the electric and geomagnetic field vectors), and of course the F10.7 value, which results in a state-vector of approximately 50×10^6 elements. Lastly, we emphasize that we have replaced the expectation operator ($\langle z \rangle$) used in (8a) and (8b) with the overbar used in (27) to emphasize that these quantities are now sample estimates.

The second component of an ensemble DA system is the ensemble generation step. We examined two different types of ensemble generation schemes. One type of ensemble generation scheme was that of the Ensemble Transform Kalman Filter (ETKF; Bishop et al., 2001) and the other was that of stochastic observations (van Leeuwen, 2020). Because we saw very little difference between the predictions from either method we choose here to show our implementation of the ETKF.

In the ETKF scheme, the new posterior perturbations are calculated using a basis of eigenvectors from

$$\frac{\mathbf{Z}^T_{n+1}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{Z}_{n+1}}{N_e - 1} = \mathbf{V}\Lambda\mathbf{V}^T \tag{29}$$

where the columns of $\mathbf{V}$ represent a basis of eigenvectors and Λ is a diagonal matrix of corresponding eigenvalues. Note that the eigenvectors $\mathbf{V}$ also serve as the right singular vectors of

$$\mathbf{R}^{-1/2}\mathbf{H}\mathbf{Z}_{n+1} = \mathbf{U}\Lambda^{1/2}\mathbf{V}^T \tag{30}$$

where $\mathbf{U}$ are the left singular vectors. Given (30) we may rewrite (16a) as

$$\mathbf{P}^s_{n+1} = \mathbf{Z}_{n+1}\mathbf{T}\mathbf{T}^T\mathbf{Z}^T_{n+1} \tag{31}$$

where $\mathbf{T}$ is referred to as a "transformation matrix" that scales and rotates the prior perturbations into those consistent with the posterior and takes the form

$$\mathbf{T} = \mathbf{V}(\Lambda + \mathbf{I})^{-1/2}\mathbf{V}^T \tag{32}$$

and is written here including the mean-preserving rotation of Wang et al. (2004). Hence, an ensemble of perturbations consistent with the posterior error covariance matrix may be made by defining those perturbations from the columns of

$$\mathbf{Z}^s_{n+1} = \mathbf{Z}_{n+1}\mathbf{T}. \tag{33}$$

On the other hand, in order to model the F10.7 evolution here in a way that is self-consistent with Section 2 we developed the following extension of equation (2),

$$\mathrm{d}F = \beta\left(F^E - F\right)\mathrm{d}t + \alpha_F w \mathrm{d}t \tag{34}$$

where F^E is the actual 24-hourly time series of the Earth's observed F10.7. The goal of the addition of the first term on the right-hand side of (34), when compared against (2), is to require the evolution of F to remain in the vicinity of the

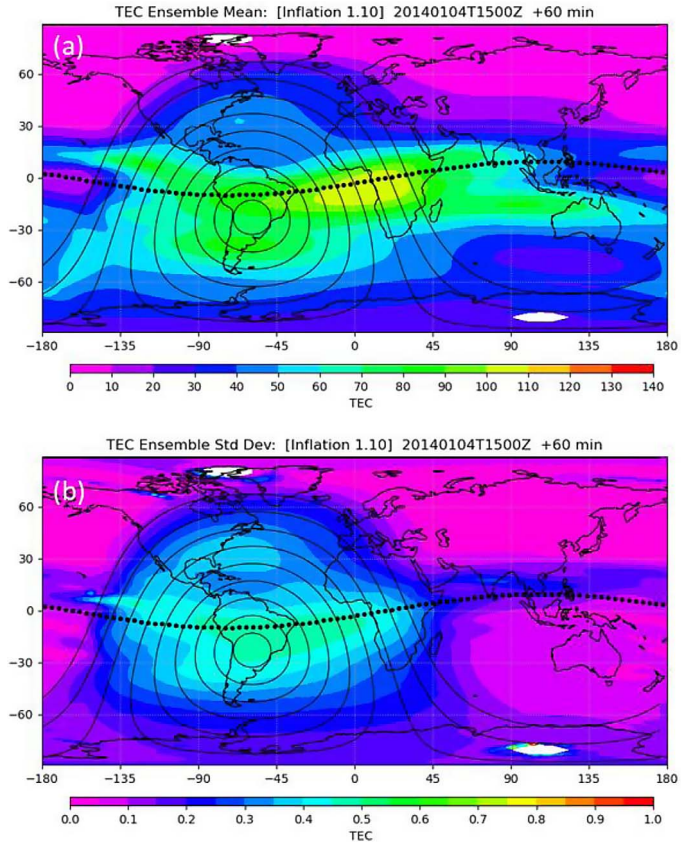

Fig. 4. In (a) is the prior ensemble mean (in TECU) and in (b) is the ensemble standard deviation both valid at 12 Z on 20140104 and for the 1-h forecast. Solid black contours are the solar zenith angle at 10°, 20°, . . . , 90°, denoting both local noon and the daylit region. The dotted line is the magnetic equator.

Earth's observed F10.7 value to ensure physical realism. Note that this term pushes F towards F^E, while the second term on the right-hand side is a random walk term that induces a drift away from F^E at a rate governed by α_F. The degree to which the solution is pushed towards F^E is governed by the size of β. The larger β is the greater the push towards F^E. This competition will produce the diversity we need in our ensemble members while maintaining the F in a physically realistic regime.

The experimental setup will use a truth run, which will constitute the hidden truth from which we will draw noisy observations. We emphasize here that the ensemble members as well as the truth will make use of (34) for its F10.7 value, and therefore the Earth's actual F10.7 value is not set here as the particular value for the simulation we label truth. The reason we did this is that we want to examine the ability of the DA to determine the F10.7 value at a temporal frequency shorter than what the Earth's observed F10.7 values are generally available, for which observed F10.7 is generally available daily while we are looking for those values hourly.

3.2 Experiment

The experiment described here will use simulated vertical vTEC observations at 30° geographic longitude spacing along 30° S, 20° S, 10° S, 0°, 10° N, 20° N, and 30° N degrees geographic latitude for a total of 84 observations. The DA will use an hourly cycling interval starting at 1 January 2014 at 0 Z and these observations will be constructed from the truth run at this 1-h frequency. The F10.7 evolution equation (32) will use parameters of $\beta = 2$ and $\alpha_F = 10$.

A critical aspect of starting the DA process is to get the initial ensemble variance correct during the "spin-up" of the system such that we replicate random draws from a climatological distribution that is self-consistent with the truth run in the sense that the variance of the ensemble predicts the actual error in the ensemble mean. We accomplish this in the following way. The first day of the period for which we would like to assimilate observations is 1 January 2014 at 0 Z. We start 3 days before that on 29 December 2013 with a "cold" start at 0 Z in which all ensemble members have an identical initial condition, but a different F10.7 value through time as derived from (34). We use 30 ensemble members. We integrate these 30 ensemble members using the SAMI3 model for the 3 days needed to get an ensemble that is valid at 0 Z on 1 January 2014. Self-consistency with the truth run is obtained by generating the truth in exactly the same way. The key features of this procedure are 1) giving the system enough time (3 days) to develop its own natural variability and 2) we generate the truth run in the exact same way as the initial members such that we can be absolutely certain that truth and the ensemble members are random draws from the same distribution.

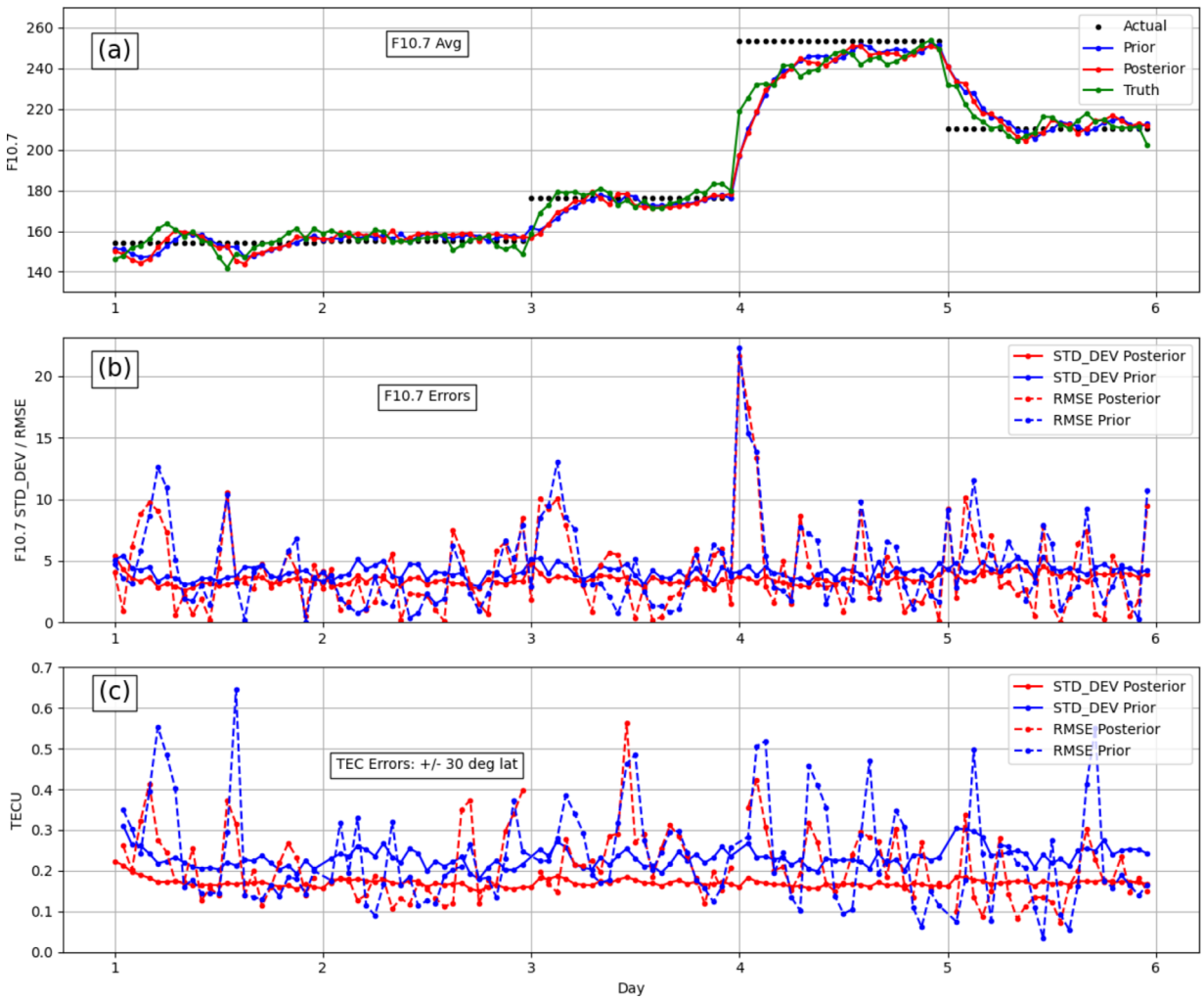

Fig. 5. F10.7 and the error in F10.7/vTEC. In (a) is the Earth's actual F10.7 value, F^E, which is only available daily (black dots), the F10.7 value, F, used as a proxy for truth (green), as well as the prior (blue) and posterior (red) mean estimates of the truth. In (b) is the root mean square error (RMSE) of the prior and posterior mean as compared with the standard deviation of the ensemble for the F10.7 prediction (in TECU). In (c) is the RMSE of the prior and posterior mean as compared with the standard deviation of the ensemble for the vTEC prediction averaged between −30° S and 30° N degrees latitude.

It is common in an ensemble DA approach for the prior and posterior variance to be too small as compared to the MSE of the ensemble mean, and this is usually due to inaccuracies in the model as well as sampling error from small ensemble sizes. To help ameliorate this issue we applied posterior inflation to the ensemble posterior perturbations in equation (33) by multiplying that equation by a factor slightly greater than one. We used a separate factor for the state variables versus the F10.7 value. Upon tuning these inflation factors we found that the best performance (i.e. smallest MSE of the posterior mean) was found when there was no inflation for the state variables and an inflation factor of 1.1 for the F10.7 value. The presentation to follow shows the behaviour of the ensemble using this posterior inflation factor of 1.1.

In Figure 4 we show an example of the prior mean and standard deviation of the ensemble vTEC field at the 1-h forecast time. The mean shows the strongest vTEC values are within the daylit portion of the globe and slightly to the east of local noon. The standard deviation shows that the strongest differences between ensemble members also occur over the daylit portion of the globe and has its peak slightly to the east of local noon. Note that both of these features are represented in the vTEC field of the heuristic model of Section 2 (see Figs. 1 and 2).

In Figure 5 we show the behaviour of this DA system for predicting the F10.7 value as well as the ensemble's prediction of the uncertainty in both F10.7 and vTEC for both the prior and posterior. Figure 5a shows the F10.7 values from equation (34) that we use as the truth (green), as well as the prior (blue) and posterior (red) mean estimates of the truth. The sharp jump at the beginning of day 4 is caused by changes in the daily F10.7 values used for F^E in equation (34) as the Earth's F10.7 value jumped significantly from 3 to 4 January 2014. Note that the predictions by the prior or posterior lag the truth anytime there is a sudden change in the true F10.7 value, which is similar to Figure 1. This phase lag between the truth and the prediction can be most easily seen at the beginning of

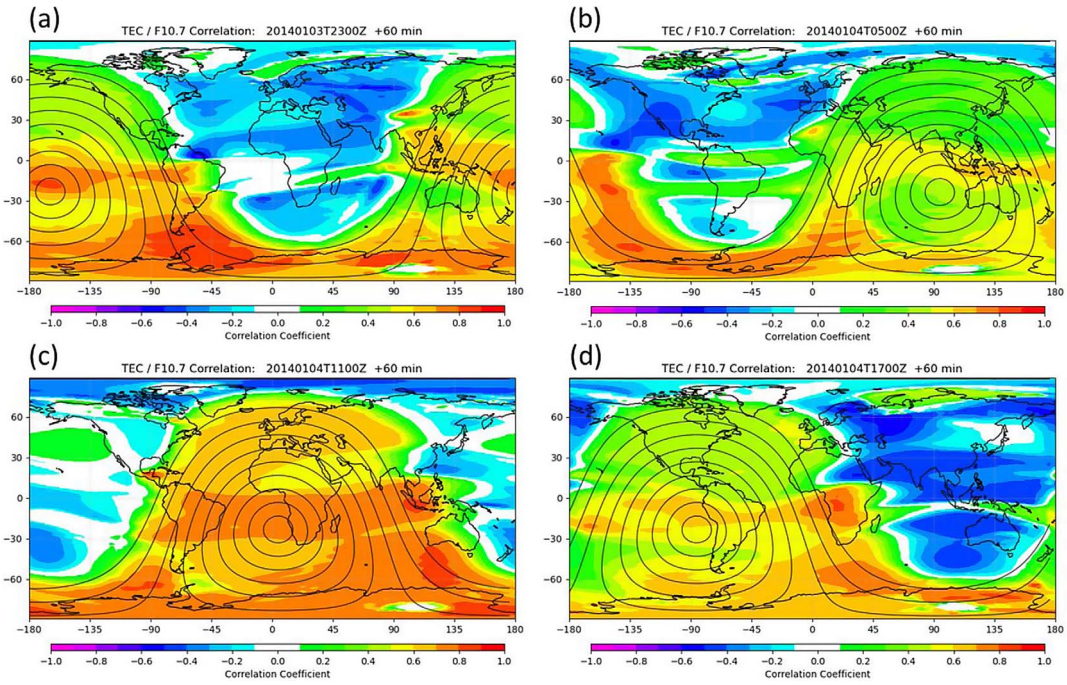

Fig. 6. The prior ensemble correlation between F10.7 and the prior vTEC on 4 January 2014 at (a) 0000 Z, (b) 0600 Z, (c) 1200 Z, (d) 1800 Z. Solid black contours are the solar zenith angle at 10°, 20°, 30°, . . . , 90°, denoting both local noon and the daylit region.

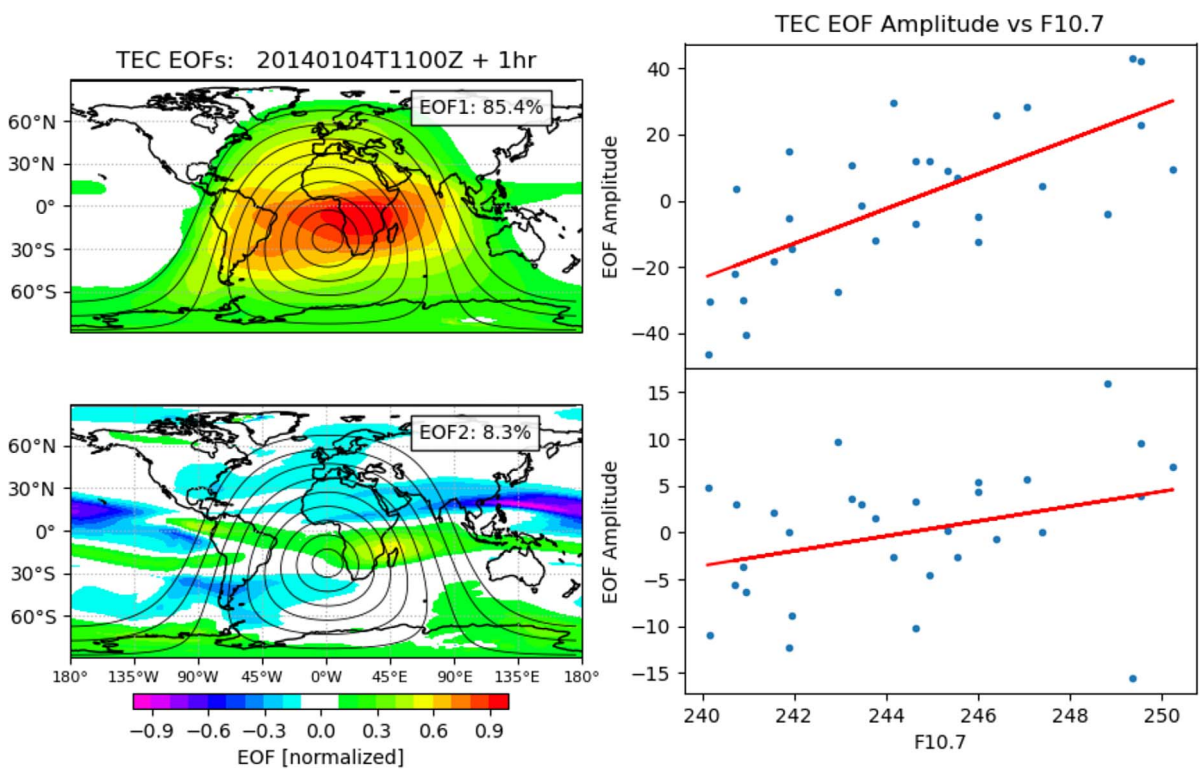

Fig. 7. (a) The leading EOFs for the prior VTEC on 4 January 2014 at 1200 Z and their projection onto the ensemble. Solid black contours are the solar zenith angle at 10°, 20°, 30°, . . . , 90°, denoting both local noon and the daylit region. (b) Projection of the leading EOF onto the ensemble as a function of the ensemble F10.7 values. (c), (d) Same as (a), (b) but for the second EOF.

days 3 and 4. Otherwise, the prediction of the F10.7 value is quite accurate.

Another important quantity to examine is the degree to which the ensemble standard deviation matches the prediction of the error. Figure 5b plots the error as compared to the standard deviation of the ensemble for the F10.7 and in Figure 5c we see the same comparison but this time for the vTEC averaged over the region −30 to 30 degrees latitude. In general, we see a reasonable agreement between the time-averaged average error and the ensemble's prediction of the error. In addition, it is interesting that the peak error in F10.7 can be seen to lead the peak error in vTEC. For example, note that the error spikes immediately at the beginning of day 4 in F10.7 but that error does not peak in vTEC until several hours later. We believe that this is due to the fact that the increased F10.7 takes time to modify the vTEC field and this then leads to a time lag in the error.

In order to update the F10.7 from observations of vTEC we must have a correlation between those two quantities. Figure 6 shows the correlation between those two variables from the ensemble on 4 January 2014. It is clear from this figure that the correlation between F10.7 and vTEC is strongest over the daylit portion of the globe. This shows that observations of vTEC in the daylit portion of the globe provide large amounts of information towards the estimate of the F10.7 value, which was a fact also deduced from the heuristic model of Section 2. An interesting feature of Figure 6 is the generally negative correlation found at night. These negative nighttime correlations between F10.7 and vTEC imply that increases (decreases) in F10.7 leads to less (more) vTEC at night. Recall that equation (12) from the low-latitude heuristic model shows the covariance between the solar forcing parameter and the vTEC is either zero or positive. Hence, these negative nighttime correlations are suggesting that there is physics missing from our heuristic model and is apparently not described simply by a solar forcing term balanced against linear recombination, where that recombination rate is independent of the F10.7. To this end, we performed some DA experiments (not shown) in the low-latitude heuristic model replacing the linear recombination term with a recombination term that was a linear function of the F10.7. This did in fact result in negative correlations between our proxy for F10.7, F, and the vTEC in that model. Therefore, it would appear that these negative nighttime correlations between F10.7 and vTEC arise from the modulation of the recombination rate by F10.7.

Another way to see this relationship between vTEC and F10.7 is to examine the EOFs of the prior covariance matrix. In Figure 7 we show the leading two EOFs of the prior covariance matrix on 4 January 2014 at 12 Z. These two eigenvectors account for 94% of the global variability in vTEC. Recall from Section 2 that in a system dominated by solar forcing it was seen that the prior covariance matrix should be of very low rank and here we are seeing that manifest itself in the prior covariance matrix in our full physics model. Additionally, we project each eigenvector back onto each ensemble member and subsequently produce a scatter plot of this EOF amplitude versus the F10.7. From this, it is then seen that there is a strong relationship between the amplitude of each member as represented by these EOFs and the F10.7 value of the ensemble.

Lastly, we take the vTEC fields in magnetic coordinates from the ensemble on 4 January 2014 at 16 Z, extract the equatorial latitude circle, and then make a prior covariance matrix. This matrix can be seen in Figure 8 and compares remarkably favourably to Figure 2. The main difference between Figures 2 and 8 appears to be the length scale of the nighttime covariances, which are much longer here than in Figure 2. In the heuristic model from which Figure 2 was derived the nighttime covariances are effectively modelled by the residual small-scale variability term. While we could have modified the noise term in the heuristic model to produce correlated noise to better match that of Figure 8 we do not believe this would provide any new insights into the behaviours we have seen. In any event, we show in Figure 8b the structure of the eigenvector from this prior covariance matrix and it too has a striking resemblance to those shown in Figure 2. This eigenvector explains 99.3% of the variance, which shows that the vTEC along the magnetic equator results in a prior covariance matrix that is very nearly rank-1. The heuristic model of Section 2 shows that this is a result of the dominance of solar forcing over the other physical processes in the SAMI3 model.

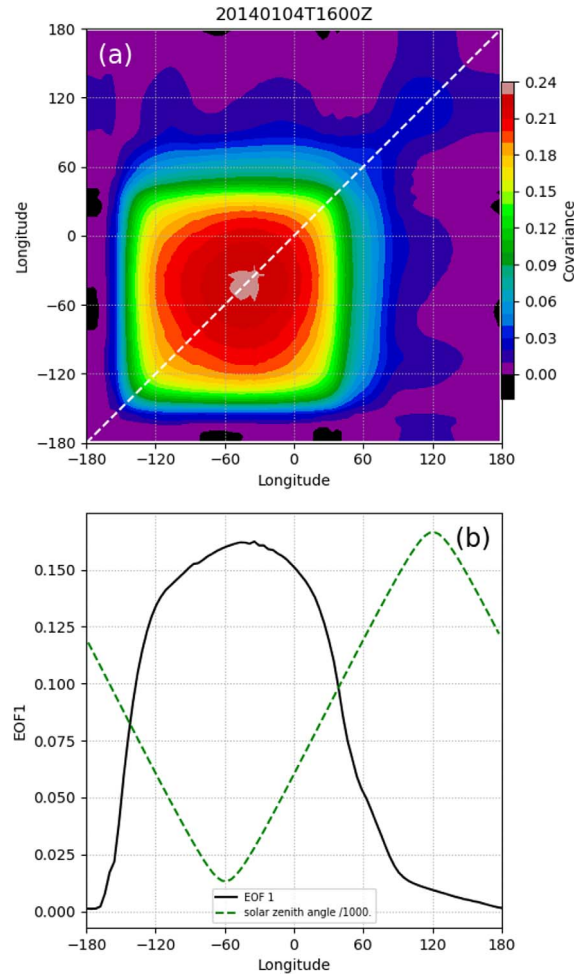

Fig. 8. In (a) is the covariance matrix for the equatorial latitude circle as defined in magnetic coordinates. In (b) is the leading eigenvector of the matrix in the top panel. The green dashed line is the solar zenith angle along the magnetic equator, where we emphasize that local noon is at a minimum in the solar zenith angle.

4 Summary and conclusions

In this work we have sought to reveal the fundamental principles governing data assimilation (DA) in a physical system being strongly impacted by external forcing. The specific application here was to understand what happens when one drives the predicted state of the ionosphere through the estimation of the solar ionizing irradiance from observations of the state of the ionosphere. To this end, we developed a heuristic model of the ionosphere along the magnetic equator that contains physics from solar forcing, recombination, and smaller-scale random processes and showed that its general behaviours were quite close to the full-physics ionospheric model referred to as SAMI3. Because this heuristic framework was both linear and statistically Gaussian this allowed us to use the Kalman filter to clearly reveal several important features of the problem. We showed how recombination acts as a sink on the information in the initial condition about ionospheric field variables (such as vTEC or electron density) obtained from DA because dissipation in a model, such as recombination, leads to the dissipation of the information in the initial condition. On the other hand, we also saw that the initial condition for the solar forcing parameter is not damped in the same way and therefore it is likely that observational information in these parameters will last longer than for ionospheric field variables. This may mean that state estimation for parameters, more so than field variables, will lead to better forecasts from ionospheric modelling systems. We also saw that when solar forcing dominates the prior covariance matrix of the vTEC that this matrix becomes dominated by a single eigenvector whose structure is directly related to that of the solar forcing. This is important because low-dimensional prior covariance matrices typically require that DA use fewer observations to make substantial impacts. We showed an example where the parameter estimation for forcing terms leads to a phase error in the state estimate relative to the truth. This phase error arises from the specific structure of the Kalman equation and the way it averages the prior mean with the observations. We also saw that the performance of a DA system in this regime is determined by the relative dominance of solar forcing and recombination to that of the smaller-scale processes. This means given the same observations and when the solar forcing dominates, the ability for DA to constrain the state is much higher than when the smaller scales dominate. Lastly, we saw that the most impactful observations on vTEC and the solar forcing parameter are those observations on the sunlit side of the Earth.

One remarkable finding of this work was that the global SAMI3 ionospheric system was constrained using 84 observations and 30 ensemble members. An ensemble-based Kalman filter is known to have the property that the dimension of the vector space that it can constrain is set by the size of the ensemble. Therefore, even though the dimension of the state vector in this system was of the order of 50×10^6 it must be true that the actual dimension of the error space in SAMI3 is much lower, and apparently no more than 30. This low-dimensional character that we see from the SAMI3 model has two important implications. First, low-dimensional systems are typically highly correlated (e.g. McNamara & Wilkinson, 2009; Forsythe et al., 2020a; Allen et al., 2023) and therefore few observations are required to constrain them, which of course is what we saw happen here. This system (all 50×10^6 variables) was constrained by 84 observations of vTEC, which we emphasize is not a description of the apparent information content in vTEC but in fact a description of the highly correlated nature of the model. (See Reid et al., 2022 for a similar example.) Second, this low dimensionality in both our heuristic model as well as SAMI3 is directly arising from our assumptions about the amplitude of the small-scale noise terms relative to solar forcing. In our low-latitude heuristic model, we simply set the noise level to be fairly low and representative of the SAMI3 variability. However, in SAMI3 small-scale neutral variability from such atmospheric processes as vertically propagating gravity waves are entirely missing because the climatological thermosphere is set from HWM14/NRLMSISE-00 (Picone et al., 2002, Drob et al., 2015), which only models the broadest scale atmospheric tides. For this reason, it could very well be that the structure of the covariance matrices seen here, and the relative dominance of solar forcing to the smaller scale processes, needs to be revisited in the presence of a thermosphere with more realistic variability at all scales (Forsythe et al., 2020b, 2021). Indeed, Allen et al. (2023) compared correlations of vTEC from SAMI3 forecasts driven by NRLMSISE-00 with SAMI3 forecasts driven by more realistic neutral winds from a general circulation model and found that the latter had significantly smaller correlation lengths that better agreed with those correlation length scales calculated using JPL GIM vTEC fields.

Lastly, in this work, we only looked at how perturbations in solar forcing through F10.7 affects the evolution of uncertainty and the state estimation process. It would be interesting to understand exactly what types of structural changes arise in the ensemble statistics (i.e. the covariance and correlation structure of the prior covariance matrices, etc.) of the electron density fields from variability in other parameters, like Ap, a measure of global geomagnetic activity, or the rate coefficients in the photoproduction routines. For example, the physical processes that are driven by Ap in the SAMI3 model (such as the storm-time component of the neutral winds) are different from that of the F10.7 parameter and therefore it is likely that the impact on DA is different. Work in this direction is already underway.

Acknowledgements. We would like to thank the Office of Naval Research for supporting this work through grant N0001422WX00451. We thank Dr. Douglas Drob and an anonymous reviewer for pointing out to us that the lack of F10.7 in the recombination term of our heuristic model could lead to negative nighttime correlations. The editor thanks David Themens and an anonymous reviewer for their assistance in evaluating this paper.

References

Allen DR, Hodyss D, Forsythe VV, McDonald SE. 2023. Low latitude monthly total electron composite correlations. *J Space Weather Space Clim* **13**: 7. https://doi.org/10.1051/swsc/2023005.

Bergot N, Tsagouri I, Bruyninx C, Legrand J, Chevalier J-M, Defraigne P, Baire Q, Pottiaux E. 2013. The influence of space weather on ionospheric total electron content during the 23rd solar cycle. *J Space Weather Space Clim* **3**: A25. https://doi.org/10.1051/swsc/2013047.

Bishop CH, Etherton BJ, Majumdar SJ. 2001. Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon Wea Rev* **129**: 420–436. https://doi.org/10.1175/1520-0493(2001)129 %3C0420:ASWTET%3E2.0.CO;2.

Chamberlin PC, Woods TN, Eparvier FG. 2008. Flare Irradiance Spectral Model (FISM): Flare component algorithms and results. *Space Weather* **6**: S05001. https://doi.org/10.1029/2007SW000372.

Codrescu SM, Codrescu MV, Fedrizzi M. 2018. An ensemble Kalman filter for the thermosphere-ionosphere. *Space Weather* **16**: 57–68. https://doi.org/10.1002/2017SW001752.

Drob DP, Emmert JT, Meriwether JW, Makela JJ, Doornbos E, Conde M, Hernandez G, Noto J, Zawdie KA, McDonald SE, Huba JD, Klenzing JH. 2015. An update to the Horizontal Wind Model (HWM): The quiet time thermosphere. *Earth Space Sci* **2(7)**: 301–319. https://doi.org/10.1002/2014ea000089.

Fang T-W, Fuller-Rowell T, Yudin V, Matsuo T, Viereck R. 2018. Quantifying the sources of ionosphere day-to-day variability. *J Geophys Res: Space Phys* **123**: 9682–9696. https://doi.org/10.1029/2018JA025525.

Forsythe VV, Azeem I, Crowley G. 2020a. Ionospheric horizontal correlation distances: Estimation, analysis, and implications for ionospheric data assimilation. *Rad Sci* **55**: e2020RS007159. https://doi.org/10.1029/2020RS007159.

Forsythe VV, Azeem I, Crowley G, Makarevich RA, Wang C. 2020b. The global analysis of the ionospheric correlation time and its implications for ionospheric data assimilation. *Rad Sci* **55**: e2020RS007181. https://doi.org/10.1029/2020RS007181.

Forsythe VV, Azeem I, Blay R, Crowley G, Gasperini F, et al. 2021. Evaluation of the new background covariance model for the ionospheric data assimilation. *Rad Sci* **56**: e2021RS007286. https://doi.org/10.1029/2021RS007286.

Huba JD, Joyce G, Fedder JA. 2000. Sami2 is Another Model of the Ionosphere (SAMI2): A new low-latitude ionosphere model. *J Geophys Res: Space Phys* **105(A10)**: 23035–23053. https://doi.org/10.1029/2000JA000035.

Matsuo T, Lee I, Anderson JL. 2013. Thermospheric mass density specification using an ensemble Kalman filter. *J Geophys Res: Space Phys* **118**: 1339–1350. https://doi.org/10.1002/jgra.50162.

McNamara LF, Wilkinson PJ. 2009. Spatial correlations of foF2 deviations and their implications for global ionospheric models: 1. Ionosondes in Australia and Papua New Guinea. *Rad Sci* **44**: RS2016. https://doi.org/10.1029/2008RS003955.

Morozov AV, Ridley AJ, Bernstein DS, Collins N, Hoar TJ, Anderson JL. 2013. Data assimilation and driver estimation for the global ionosphere-thermosphere model using the ensemble adjustment Kalman filter. *J Atmos Sol-Terr Phys* **104**: 126–136. https://doi.org/10.1016/j.jastp.2013.08.016.

Picone JM, Hedin AE, Drob DP, Aikin AC. 2002. NRLMSISE-00 empirical model of the atmosphere: Statistical comparisons and scientific issues. *J Geophys Res* **107(A12)**: 1468. https://doi.org/10.1029/2002JA009430.

Reid B, Themens DR, McCaffrey AM, Jayachandran PT, Johnsen MG, Ulich T. 2022. A-CHAIM: Near-real-time data assimilation of the high latitude ionosphere with a particle filter. *Space Weather* **21**: e2022SW003185. https://doi.org/10.1002/essoar.10511639.1.

Richards PG, Fennelly JA, Torr DG. 1994. EUVAC: A solar EUV flux model for aeronomic calculations. *J Geophys Res* **99**: 8981. https://doi.org/10.1029/94JA00518.

Rishbeth H, Mendillo M. 2001. Patterns of F2-layer variability. *J Atmos Sol-Terr Phys* **63**: 1661–1680. https://doi.org/10.1016/S1364-6826(01)00036-0.

Schmölter EJ, Bedermann M Codrescu. 2021. The delayed ionospheric response to the 27-day solar rotation period analyzed with GOLG and IGS TEC data. *J Geophys Res: Space Phys* **126**: e2020JA028861. https://doi.org/10.1029/2020JA028861.

Siscoe G, Solomon SC. 2006. Aspects of data assimilation peculiar to space weather forecasting. *Space Weather* **4**: S04002. https://doi.org/10.1029/2005SW000205.

Solomentsev DV, Khattatov BV, Codrescu M, Titov AA, Yudin V, Khattatov VU. 2012. Ionosphere state and parameter estimation using the Ensemble Square Root Filter and the global three-dimensional first-principle model. *Space Weather* **10**: S07004. https://doi.org/10.1029/2012SW000777.

Vaishnav R, Jacobi C, Berdermann J, Codrescu M, Schmölter E. 2021a. Role of Eddy diffusion in the delayed ionospheric response to solar flux changes. *Ann Geophys* **39**: 641–655. https://doi.org/10.5194/angeo-39-641-2021.

Vaishnav R, Schmölter E, Jacobi C, Berdermann J, Codrescu M. 2021b. Ionospheric response to solar extreme ultraviolet radiation variations: Comparisons based on CTIPe model simulations and satellite measurements. *Ann Geophys* **39**: 341–355. https://doi.org/10.5194/angeo-39-341-2021.

van Leeuwen PJ. 2020. A consistent interpretation of the stochastic version of the ensemble Kalman filter. *Quart J Roy Met Soc* **146**: 2815–2825. https://doi.org/10.1002/qj.3819.

Wang X, Bishop CH, Julier SJ. 2004. Which is better, an ensemble of positive-negative pairs or a centered spherical simplex ensemble? *Mon Wea Rev* **132**: 1590–1605. https://doi.org/10.1175/1520-0493(2004)132 %3C1590:WIBAEO%3E2.0.CO;2.

Zawdie KA, Dhadly MS, McDonald SE, Sassi F, Coker C, Drob D. 2020. Day-to-day variability of the bottomside ionosphere. *J Atmos Sol-Terr Phys* **205**: 105299. https://doi.org/10.1016/j.jastp.2020.105299.

Appendix A

Quadratic recombination in a Heuristic model

Imagine that we are interested in an N-vector, $\mathbf{q}$, evolving through time as

$$\frac{\mathrm{d}\mathbf{q}}{\mathrm{d}t} = F(t)g(\mathbf{x} - ct\mathbf{1}) - \frac{a}{\tau}\mathbf{q}^2 \tag{A.1}$$

where $\frac{a}{\tau}$ is the rate of recombination and all other variables are the same as Section 2.1. Note that this equation has a steady-state solution when $c = 0$, i.e.

$$\mathbf{q} = \sqrt{\frac{\tau}{a}Fg}. \tag{A.2}$$

Our goal here is to reveal the impact of the movement implied by $|c| > 0$. We note that the solution to (A.1) is independent on a grid-point by grid-point basis. Therefore, in the interest of clarity, we describe the solution procedure at a representative grid point, x_0.

We will use the WKB approximation below and that methodology works most clearly with a non-dimensional equation. Therefore, we define the following re-scaling:

$$(x, t) \rightarrow (Lx,\ Tt) \tag{A.3}$$

$$q \rightarrow Qq \tag{A.4}$$

$$F \rightarrow \frac{Q}{T}F \tag{A.5}$$

$$c \to \frac{L}{T}c. \tag{A.6}$$

Using these scales, we obtain the following non-dimensional equation

$$\frac{\mathrm{d}q}{\mathrm{d}t} = F(t)g(x_0 - ct) - \theta q^2 \tag{A.7}$$

where we have reduced the problem to a single non-dimensional parameter

$$\theta = \frac{aQT}{\tau}. \tag{A.8}$$

Equation (A.7) is a nonlinear Ricatti equation. To solve (A.7), we look for a solution, v, such that

$$q = \frac{1}{\alpha}\frac{1}{v}\frac{\mathrm{d}v}{\mathrm{d}t} \tag{A.9}$$

which upon using this in (A.7) obtains

$$\frac{1}{\theta}\frac{\mathrm{d}^2 v}{\mathrm{d}t^2} = Fgv. \tag{A.10}$$

We assume that

$$\varepsilon^2 = \frac{1}{\theta} \ll 1 \tag{A.11}$$

and, therefore, we have a WKB problem whose solution can be written as

$$v(t) = \exp\left[\frac{1}{\varepsilon}S_0(t) + S_1(t) + \ldots\right]. \tag{A.12}$$

The standard solution method is to insert this into (A.10), balance like orders in ε, and solve the sequence of resulting problems. We complete these steps out to the second term of the series to obtain

$$v(t) = c_1 M^{-1/4}\exp\left[\theta^{1/2}\int_0^t \sqrt{Q(t')}\mathrm{d}t'\right] + c_2 M^{-1/4} \times \exp\left[-\theta^{1/2}\int_0^t \sqrt{Q(t')}\mathrm{d}t'\right] \tag{A.13}$$

where $Q(t) = Fg$. While there are two possible solutions, when we insert (A.13) into (A.9) we see that one obtains a negative vTEC, which is obviously unphysical and we therefore discard. The remaining solution obtains

$$q = \theta^{-1/2}\sqrt{Fg} - \frac{1}{4\theta F}\frac{\mathrm{d}F}{\mathrm{d}t} - \frac{1}{4\theta g}\frac{\mathrm{d}g}{\mathrm{d}t}. \tag{A.14}$$

Note that the first term of this expression is the steady state (A.2), which implies that the next two terms are the first-order corrections to that steady state owing to the variability and the movement of the solar forcing.

As a concrete example, imagine the structure-function takes the following form

$$g(x_0 - ct) = \exp\left[-\frac{L^2}{2L_g^2}(x_0 - ct)^2\right] \tag{A.15}$$

where L_g is the Gaussian half-width of the solar forcing structure function. Using this in (A.14) obtains

$$q = \theta^{-1/2}\sqrt{F}\exp\left[-\frac{L^2}{4L_g^2}(x_0 - ct)^2\right] - \frac{1}{4\theta F}\frac{\mathrm{d}F}{\mathrm{d}t} - \frac{cL^2}{4\theta L_g^2}(x_0 - ct). \tag{A.16}$$

This solution reveals several interesting features of the VTEC field and the impact of solar forcing balanced by *quadratic* recombination. The first term varies in time according to the square root of the F10.7 and whose structure is wider in scale than the structure-function, g. Note that in the case where F is a constant then the vTEC field, when forced by linear recombination (as seen in Eq. (3)), would depend linearly on the solar forcing, while in the quadratic recombination case, the vTEC field would be more weakly forced as $\sqrt{F}$. The second term shows that the next order correction to the time variations in the solar forcing is to reduce (increase) the amplitude of the vTEC field when the solar forcing increases (decreases) with time. This non-intuitive result is a direct result of the quadratic nature of this recombination term. As solar forcing increases it causes an increase in amplitude in the first term. However, this increase in amplitude causes a quadratically greater impact in dissipation through recombination as the solar forcing changes, which leads to this damping in the second term. The third term is the most interesting as it implies a pedestal of vTEC behind the main Gaussian vTEC mass as that mass moves across the latitude circle. This pedestal is longest when the speed is largest and the width of the structure-function is narrow. This asymmetry in the vTEC field can also be seen in the presence of linear recombination as in Figure 1a as the asymmetry of the vTEC mass and its elongation to the east.

Appendix B

Time-lag from equation (10b)

We wish to discuss how equation (10b) results in a time-lag. Equation (10b) is broadly of the following form

$$F_{n+1} = (1-\alpha)F_n + \alpha d_n \tag{B.1}$$

where α plays the role of $\Delta\mathbf{Q}_{n+1}\mathbf{H}g(\mathbf{x} - ct_n\mathbf{1})$ in (10b), d_n plays the role of the observations and we have discretized time as $t_{n+1} = t_n + \Delta$. We can determine an expression for the time lag in (B.1) most easily if we choose d_n to be sinusoidal, i.e.

$$d_n = e^{i\omega t_n} \tag{B.2}$$

which upon assuming a solution of the form

$$F_n = ae^{i\omega t_n}. \tag{B.3}$$

Obtains

$$a = \frac{\alpha[\alpha - 1 + \cos(\omega\Delta)]}{1 - (1-\alpha)\cos(\omega\Delta) + (1-\alpha)^2} - i \times \frac{\alpha\sin(\omega\Delta)}{1 - (1-\alpha)\cos(\omega\Delta) + (1-\alpha)^2}. \tag{B.4}$$

Because the parameter a is a complex number we choose to write it in phase-modulus form, viz.

$$a = re^{-i\theta} \tag{B.5}$$

such that (B.3) maybe re-written as

$$F_n = re^{i(\omega t_n - \theta)} \tag{B.6}$$

where

$$\theta = \tan^{-1}\left(\frac{\sin(\omega\Delta)}{\alpha - 1 + \cos(\omega\Delta)}\right). \tag{B.7}$$

Note that it is clear from (B.6) that when $\theta > 0$ that this is a time lag. As an example, we define a frequency $\omega = 2\pi/24$ and time-step of $\Delta = 1$ in order to plot θ versus α in Figure B.1. We see that the phase lag is inversely proportional to α, and is 6 time-units for a value of $\alpha = 0.035$.

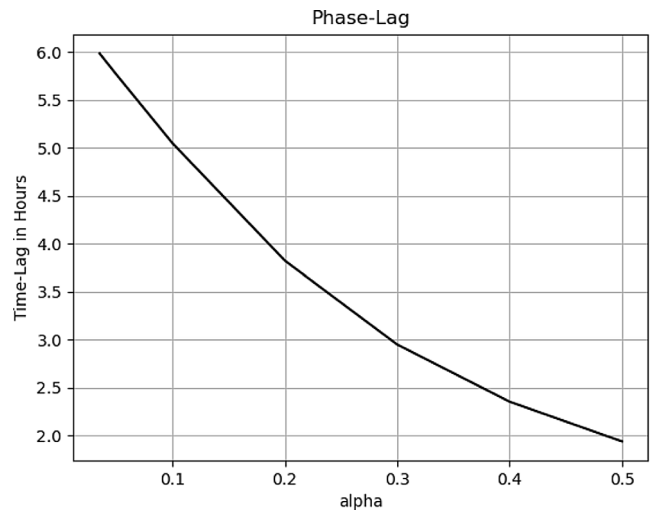

Fig. B.1. The phase lag in units of non-dimensional "hours" as arising from equation (B.1).

Cite this article as: Hodyss D, Allen DR, Tyndall D, Caffrey P & McDonald SE, 2023. The effects of estimating a photoionization parameter within a physics-based model using data assimilation. *J. Space Weather Space Clim.* **13**, 21. https://doi.org/10.1051/swsc/2023019.