Performance Analysis of a Strong Constraint 4DVar and 4DEnVar on Regional Ionosphere Imaging

***Nicholas Ssessanga[1], Wojciech Jacek Miloch[1], Lasse Boy Novock Clausen[1], and Daria Kotova [1]**

[1]4DSpace, Department of Physics, University of Oslo, Blindern, N-0316 Oslo, Norway.

Corresponding author: Nicholas Ssessanga (nikizxx@gmail.com)

Key Points:

- Specifying regional ionosphere with strong constraint 4DVar and 4DEnVar while ingesting ground GNSS data.
- 4DEnVar is computationally efficient at handling linearization and provides better 3D analyses in severe conditions and remote areas.
- 4DEnVar is recommended for regional ionosphere short-time forecast.

Abstract

Data assimilation (DA) techniques have recently gained traction in the ionospheric community, particularly at regional operational centers where more precise data are becoming prevalent. At centre stage is the argument over which technique or scheme merits realization. At 4DSpace, we have in-house developed and assessed the performance of two regional flavors of short-term forecast strong constraint four-dimensional (4D, space and time) variational (SC4DVar) DA schemes; the orthodox incremental (SC4DVar-Inc) and ensemble-based (SC4DEnVar) approach. SC4DVar-Inc is bottled-necked by expensive Tangent Linear Models (TLMs) and model Ad-joints (MAs), while SC4DEnVar design mitigates these limitations. Both schemes initialize from the same background (IRI-2016), and electron densities forward propagated (30-min) by a Gauss Markov filter- the densities take on a log-normal distribution to assert the mandatory ionosphere density positive definiteness. Preliminary assimilation is performed only with ubiquitous Global Navigation Satellite System observables from ground-based receivers, with a focus on moderately stable mid-latitudes, specifically the Japanese archipelago and neighboring areas. Using a simulation analysis, we find that under model space localization, 30 member Ensembles are sufficient for regional SC4DEnVar. Verification of reconstructions is with independent observations from ground-based ionosonde and satellite radio occultations: the performance of both schemes is fairly adequate during the quiet period when the background has a better estimation of the hmF2. SC4DVar-Inc is slightly better over areas densely populated with measurements, but SC4DEnVar estimates the overall 3D ionosphere picture better, particularly in remote areas and during severe conditions. These results warrant SC4DEnVar as a better candidate for precise short-time regional forecasts.

Plain Language Summary

We have developed and assessed the performance of two flavors of short-term forecast four-dimensional (space and time) variational schemes. The first scheme relies on the quality of Tangent Linear Models and model Ad-joints in its design. The second scheme adopts an ensemble-based approach. Both schemes initialize from the same background and electron densities propagated by a Gauss Markov filter. Our verification analysis indicates that the ensemble approach approximates peak density-height variations better and offers improved estimates of the overall 3D ionosphere picture, particularly in remote areas and during severe conditions. Therefore, we recommend employing the ensemble-based approach for precise short-time regional forecasts.

1 Introduction

In most applications the direct and precise measurement of a required state of a system is generally not feasible. More so, the measurements are usually imperfect and spotty in space (three-dimensional) and time. Data assimilation (DA) techniques optimally combine measurements with prior information (via models, either physics-based model or empirical model), to obtain a consistent picture (completeness) of the state while taking into account the

theoretical underpinnings. DA techniques have strong roots in the meteorological and oceanographic communities, where they have continuously matured and successfully tuned for applications suited to these communities. By contrast, in the ionospheric community, DA techniques were often neglected and have only recently taken centre stage due to the growing number of synergistic ionospheric observations and advances in computation power. More specifically, the advent of cost-effective dual-frequency, multichannel and high temporal resolution GNSS (Global Navigation Satellite System) receivers has drastically increased the volume of ionospheric data available for scientific analysis. There is a growing interest to borrow and customize DA techniques from meteorology and oceanography for ionospheric applications; for example, optimal interpolation, weighted least squares, Kalman filters and variational methods (Hajj et al., 2004; Pi et al., 2004; Schunk et al., 2004, 2016; Scherliess et al., 2004, 2006; Thompson et al., 2006; Matsuo and Araujo-Pradere, 2011; Lee et al., 2012, 2013; Hsu et al., 2014; Ssessanga et la., 2019; Mengist et al., 2023). Variational methods, heavily based on the calculus of variations, have advantages over other techniques: a) the analysis formulation can globally include modeled 3D background model error covariances. b) there is flexibility to include nonlinear observation operators - this is a limitation for optimal interpolation methods mainly restricted to linear observation operators. c) the mismatch between a model state and the measurements can be formulated into a global error/cost function (J) and minimized. Variational methods exist in a multitude of flavors based on the level of complexity or simplification. Here, the focus is on the strong-constraint four-dimensional Variational (SC4DVar) technique: a smoother, which in theory, seeks a solution that optimizes the errors in 3D space and trajectory within a specified time window on the assumption of a perfect prognostic model (Evensen et al., 2022). The temporal component edges 4DVar over observation filter variation schemes such as

3DVar that are restricted per analysis assimilation cycle to observations within a window less than the time scales associated with the model error. Moreover, the 4DVar scheme limits the averaging ionospheric dynamics because the time dimension allows for the simultaneous assimilation of asynchronous observations at their appropriate times.

For the past few decades, a variant of SC4DVar, based on the incremental (-Inc) approach that involves the non-trivial requirement to compute the Tangent Linear Model (TLM) and Models Ad-joints (MA), has been the standard application in most fields, but TLMs and MAs implementation and maintenance is generally a bottleneck. The recent introduction of the ensemble-based variational schemes, SC4DEnVar, has alleviated some of these hindrances and thus motivated this study; using the readily available ground GNSS data, we analyze the fidelity of SC4DVar-Inc and SC4DEnVar in reconstructing a 4D regional mid-latitude ionosphere. (Note that, for most cases, the abbreviation for both full-field 4DVar and the incremental approach is "4DVar", and the prefix (SC) and suffix(-Inc) appended in this work are coined to avoid ambiguity in interpretation.) The performance analysis is during quiet (days before the storm) and severe conditions during the St. Patrick's Day 2015 geomagnetic storm.

The paper is organized as follows: section 2 describes the mathematical formulation of a SC4DVar while assuming a simplified Gauss Markov propagation (Gelb, 1974) and log-normal distribution for ionosphere electron densities (Garner et al., 2005) and summaries the SC4DVar based on Lagrangian formulation. It further provides a derivation of the SC4DVar-Inc and SC4DEnVar. In the latter, an extended emphasis is on tuning the ensemble covariance central in defining the scheme's fidelity. Section 3 details the experimental setup and compares reconstructions to independent observations from ionosondes and low earth orbit satellites (COSMIC radio occultation densities). Section 4 discusses and summarizes the findings of this

work. The section also includes recommendations and future work to improve the SC4DEnVar scheme for ionospheric applications.

2 Strong constrained 4DVar formulation (SC4DVar)

We assume the volumetric regional ionosphere to be composed of P finite elements/voxels, $\vec{\mathrm{n}} = (n_1, n_2, \cdots, n_i, \cdots, n_p)$ each with a uniform electron density distribution. Element (n_i) are assumed to be random variables sampled from independent identical random distributions, such that $\vec{n}$ is a random vector.

Usually, a schematic/simplified assumption is that $\vec{n}$ follows Gaussian statistics, however in the ionosphere perspective, this has a few unrealistic implications: i) A Gaussian distribution is symmetric, and can take values between $(-\infty,\infty)$. If n_i is close to zero (particularly electron densities in E-region and topside/plasmasphere), then the Gaussian distribution can assign probability to $n_i \leq 0$ - hence violating the positive definite ($\vec{n} > 0$) requirement for ionosphere electron densities. ii) In reality, ionospheric electron densities are better represented by other probability density functions rather "off the shelf" Gaussian statistics. Moreover, in Kotova et al. (2022), distributions of electron densities measured by the Swarm satellite mission (Olsen et al., 2013) in the Northern and Southern hemispheres were best described mainly by exponentiated Weibull, log-normal, and chi distributions. Therefore, to still utilize Gaussian statistics in our work and better approximate density distributions, we assume n_i to take on a log-normal distribution - whose natural logarithm then assumes Gaussian statistics (Garner et al., 2005); that is, we take the natural logarithm of each element, $x_i = \log(n_i)$, such that the new random vector is expressed as $\vec{\mathrm{X}} = (x_1, x_2, \cdots, x_i, \cdots, x_P)$, and at convergence, the analysis log densities are projected back into n_i space via exp^{x_i}. This argument also has a flaw - in

derivations that follow, the analysis is the mode or maximum likelihood in the Gaussian space. However, if the variance is large enough, the assumption that the mode in Gaussian space maps onto the mode in log-normal space ceases to hold (Fletcha 2006a; 2006b). The atmospheric community faces the same issue, and there is literature and developing studies proposing a better representation with mixed Gaussian-log-normal distributions. Such proposals are worth a future investigation in the ionospheric community but fall out of the scope of the current study. The assertion here is that even with the flaws, the log-normal distribution is a better statistical representation than a direct Gaussian approach.

Forward Model: In a 4DVar assimilation schemes, the propagation of the state from time t_k to t_{k+1} within an assimilation window, $\boldsymbol{t}_0 \leq \boldsymbol{t}_k \leq \boldsymbol{t}_N$, is assumed to be governed by a prognostic model $M_k : \mathbb{R}^P \rightarrow \mathbb{R}^P$. That is to say

$$\vec{\mathrm{X}}_{\mathrm{k+1}} = M_{\mathrm{k}}(\vec{\mathrm{X}}_{\mathrm{k}}) + \vec{\omega}_{\mathrm{k}}, \qquad (1)$$

where $\vec{\omega}_{\mathrm{k}} \in \mathbb{R}^P$ are model errors. In a strong constrained (perfect model) 4DVar, these errors are considered negligible ($\vec{\omega}_{\mathrm{k}} = 0$). Therefore, the accuracy of all density states ($\vec{\mathrm{X}}_{\mathrm{k}}$, k= 0,1,2,…N) within the assimilation window are entirely dependent on the accuracy of M_{k} and the control vector ($\vec{\mathrm{X}}_{\mathrm{k=0}}$). For the perfect model assumption to hold, assimilation must be limited to short-time windows. Otherwise, the integrated $\vec{\omega}_{\mathrm{k}}$ could grow inexorably. There are sophisticated weak constraint 4DVar schemes that cater for $\vec{\omega}_{\mathrm{k}}$ and can forecast over long periods; however, their realization is still challenging in the ionospheric community because of a lack of a basic understanding of the size and structure of the error covariance matrix associated with $\vec{\omega}_{\mathrm{k}}$.

In the ionosphere community, M_{k} is generally a first principle physics-based models (Fuller-Rowell and Rees, 1980; Huba et al., 2000; Pi et al., 2003; Qian et al., 2014). However, this approach is still challenging or limited especially during geomagnetic storms conditions due to

the difficulty in tracking the rapid changes of forcings such as electric fields, auroral particle precipitations, and Joule heating. Moreover, the dynamics that govern physical and chemical processes are far from being fully understood and represented in physics-based models. To circumvent some of the above limitations, Bust et al. (2004); Bust and Datta-Barua (2014), developed different DA flavour, Ionospheric Data Assimilation Four-Dimensional (IDA4D), based on a three-dimensional variational data assimilation (3DVar) derived in Daley (1991) and references therein. In IDA4D, rather than using a Physics based forward propagation or forecast, densities are drawn from an empirical model (for example NeQuick (Nava et al., 2006) or International reference ionosphere (IRI, Bilitza et al., 2022)), but then assumed to propagate for short time periods (15 minutes) following a Gauss Markov filter. We follow the same approach in this study with densities at t_{k+1} define as

$$\vec{X}_{k+1} = \vec{X}_b^{k+1} + (\vec{X}_k - \vec{X}_b^k) * \exp(-dT/\tau) \quad (2)$$

where dT is the length of the time sample ($t_{k+1} - t_k$). τ is the correlation time along the time trajectory that determines how much predictions at time sample k influence the predictions at k+1. All entities with subscript b are a background state (prior information from the empirical model) at that time sample. In this setup, the scheme is not confined to the grid of the background state: the upside is that the grid can easily get computed externally and the background interpolated or generated from previous analyses or any other ionospheric model (either empirical or physics) without modifications to the whole setup.

Cost function: consider at each time sample k, to have a random vector $\vec{Y}_k \in \mathbb{R}^M$ of noisy observations sampled from a Gaussian distributions, and h_k: $\mathbb{R}^{P \to M}$ a non linear operator that maps ($\vec{X}_k$) into the observational space at time t_k. Then we can write

$$\vec{Y}_k = h_k(\vec{X}_k) + \vec{\varepsilon}_k \qquad (3)$$

where $\vec{\varepsilon}_k$ is a combination of instrumental and the representativeness errors, assumed to be Gaussian, unbiased and with covariance $\mathbf{R}_k \in \mathbb{R}^{M\times M}$ (a square positive definite matrix). A typical assumption in data assimilation applications, and hereafter, is that $\vec{\varepsilon}_k$ elements are uncorrelated ($\mathbf{R}_k$is diagonal): this assumption has a consequence similar in a way to the effect of the ratio of background to observation variances; if the background errors are more correlated than the observation errors (the case here), the consequence in spectral space, is that scheme mainly corrects large scales while ignoring the small scales (Stonebridge et al., 2018). Besides, even when the observational correlations are known, their implementation in a scheme is usually challenging and better off ignored (mostly horizontal correlations) - most applications somewhat avoid large-scale over-fitting by compensating for the neglected off-diagonal terms by inflating the $\mathbf{R}_k$ diagonal elements.

If the multivariate $\vec{X}_{k=0}$ is considered to have a background error covariance $\mathbf{B} \in \mathbb{R}^{P\times P}$ (a positive definite matrix), such that $\vec{X}_{k=0} \sim N(\vec{X}^b_{k=0}, \mathbf{B})$, then the goal is to find $\vec{X}_{k=0}$ that minimizes the global scalar function

$$J(\vec{X}_o) = \frac{1}{2}\left[(\vec{X}_o - \vec{X}^b_o)^T \mathbf{B}^{-1}(\vec{X}_o - \vec{X}^b_o)\right] + \frac{1}{2}\sum_{k=0}^{k=N}\left[\left(\vec{Y}_k - h_k(\vec{X}_k)\right)^T \mathbf{R}_k{}^{-1}\left(\vec{Y}_k - h_k(\vec{X}_k)\right)\right] \qquad (3)$$

where, the superscripts "–1" and "T" denote the inverse and the transpose of a matrix, respectively, and the observational and the background model errors are assumed to be independent. In the nutshell, 4DVar minimizes the mismatch between model state and the measurements, in both space and time: the background term (first term) acts as constraint to the "inverse problem problem" with no unique solution because the observations (second term) are not sufficient to cover all degrees of freedom.

2.1 Lagrange multipliers approach to the solution (SC4Dvar-L)

In preparation for the discussion on SC4DVar-Inc and SC4DEnVar, we lay a foundation with a strong constraint 4DVar solved through a Lagrange multiplier approach. Results from this approach are not presented or discussed here - in the sequel we only exploit the terminologies and bottlenecks to justify why the incremental and ensemble flavors are necessary. The 4DVar equations and formulation follow those in Ssessanga et al. (2019) and the references therein, but with an added complexity that we solve for log densities. For the analysis to strictly follow model dynamics, the perfect model constraints are appended to equation (3), with the resultant Lagrangian to solve as

$$\left.\begin{aligned} \mathrm{L} &= \mathrm{J}(\vec{\mathrm{X}}_\mathrm{o}) + \sum_{\mathrm{k}=1}^{\mathrm{k}=\mathrm{N}} \lambda_\mathrm{k}^\mathrm{T} \mathrm{g}(\vec{\mathrm{X}}_\mathrm{k}) \\ \mathrm{g}(\vec{\mathrm{X}}_\mathrm{k}) &= (\vec{\mathrm{X}}_\mathrm{k} - M_{\mathrm{k}-1}(\vec{\mathrm{X}}_{\mathrm{k}-1})) \end{aligned}\right\} \tag{4}$$

The multipliers ($\lambda_\mathrm{k} \in \mathbb{R}^P$) are associated with the constraint to satisfy, $\mathrm{g}(\vec{\mathrm{X}}_\mathrm{k})$. Asserting a boundary condition $\lambda_{\mathrm{N}+1} = 0$, and through variational calculus, it is easy to show that at extreme when $\nabla_\alpha \mathrm{L} = 0$, where α: $\vec{\mathrm{X}}_\mathrm{o}$, $\vec{\mathrm{X}}_\mathrm{k}$, λ_k (e.g., Zou et al., 1997; Ssessanga et al., 2018;2019), then

$$\lambda_\mathrm{N} = \tilde{\mathbf{H}}_\mathrm{N}{}^\mathrm{T} \mathbf{R}_\mathrm{N}^{-1} \left(\vec{\mathrm{Y}}_\mathrm{N} - \mathrm{h}_\mathrm{N}(\vec{\mathrm{X}}_\mathrm{N})\right) \tag{5}$$

$$\lambda_\mathrm{k} = \tilde{\mathbf{M}}_\mathrm{k}{}^\mathrm{T} \lambda_{\mathrm{k}+1} - \tilde{\mathbf{H}}_\mathrm{k}{}^\mathrm{T} \mathbf{R}_\mathrm{k}^{-1} (\mathrm{h}_\mathrm{k}(\vec{\mathrm{X}}_\mathrm{k}) - \vec{\mathrm{Y}}_\mathrm{k}),\ \mathrm{k} = \mathrm{N}-1, \mathrm{N}-2, \cdots 0 \tag{6}$$

$$\vec{\mathrm{X}}_\mathrm{o} = \vec{\mathrm{X}}_\mathrm{b} + \mathbf{B}\lambda_0 \tag{7}$$

Where $\tilde{\mathbf{M}}_\mathrm{k} \in \mathbb{R}^{\mathrm{P}\times\mathrm{P}}$ and $\tilde{\mathbf{H}}_\mathrm{k} \in \mathbb{R}^{\mathrm{M}\times\mathrm{P}}$ are the Jacobians of the forward operator M_k and the observation operator h_k, evaluated at $\vec{\mathrm{X}}_\mathrm{k}$,

$$\widetilde{\mathbf{M}}_{\mathrm{k}} = \left.\frac{\partial M_{\mathrm{k}}}{\partial \vec{\mathrm{X}}}\right|_{\vec{\mathrm{X}}_{\mathrm{k}}}, \widetilde{\mathbf{H}}_{\mathrm{k}} = \left.\frac{\partial \mathrm{h}_{\mathrm{k}}}{\partial \vec{\mathrm{X}}}\right|_{\vec{\mathrm{X}}_{\mathrm{k}}}$$

Note that to make corrections to the background in (7) the whole process is run backwards in time. That is, the prognostic model is run forward in time $M^{\mathrm{o}\to\mathrm{N}}(\vec{\mathrm{X}}_{\mathrm{o}})$, to obtain λ_{N}; and the information about disagreements between estimates and observations integrated back to $\mathrm{t}_{\mathrm{k}=0}$. λ_{k}'s are somewhat a measure the sensitivity to changes in the solutions $\vec{\mathrm{X}}_{\mathrm{k}}$ at t_k. The whole process is repeated until convergence.

Certainly, it is obvious that in cases of high dimensions (P) SC4DVar-L is expensive, since every iteration, we have to compute $M^{\mathrm{o}\to\mathrm{N}}(\vec{\mathrm{X}}_{\mathrm{o}})$ along the trajectory, store all $\vec{\mathrm{X}}_{\mathrm{k}}$, and then perform a backward integration. In addition, without preconditioning and M_{k} and h_{k} non-linear, in real-time applications quick convergence is generally uncertain due to ellipsoidal cost function iso-surfaces; the path to convergence is short if iso-surfaces are spherical.

2.2 Incremental approach to the solution (SC4DVar-Inc)

The incremental approach addresses some of the SC4DVar-L shortfalls. Rather than solving for the full-field, the control vector $\vec{\mathrm{X}}_{\mathrm{o}}$ is expressed as an increment to the background $\vec{\mathrm{X}}_{\mathrm{o}}^{\mathrm{b}}$ at $\mathrm{t}_{\mathrm{k}=0}$,

$$\vec{\mathrm{X}}_{\mathrm{o}} = \vec{\mathrm{X}}_{\mathrm{o}}^{\mathrm{b}} + \partial\vec{\mathrm{X}}_{\mathrm{o}} \qquad (8)$$

The intuition is that we have a good approximation to the true trajectory from a nonlinear prognostic model, and what is sought is a small change or increment in the initial condition ($\partial\vec{\mathrm{X}}_{\mathrm{o}}$) to minimize a cost function through a period of time (Courtier et al., 1994). An advantage of this approach is that $\partial\vec{\mathrm{X}}_{\mathrm{o}}$ can be represented in terms of modes that satisfy linear evolution (and neglect small scales that are nonlinear), which is a requirement for the Gaussian probability distribution assumptions (see equation 9). In a way, this is preconditioning of the increment or a

control variable transform (CVT). Our choice of precondition here is to use the square root of the background covariance matrix **B** (corresponding to time $t_{k=0}$)

$$\partial\vec{X}_o = \mathbf{B}^{1/2}\vec{W} \tag{9}$$

Substituting the above back into the original cost function

$$\begin{aligned} J(\vec{W}) = \frac{1}{2}\left[(\vec{W})^T(\vec{W})\right] \\ + \frac{1}{2}\sum_{k=0}^{k=N}\left[\left(\vec{Y}_k - h_k\left(M^{0\to k}\left(\vec{X}_o^b + \mathbf{B}^{1/2}\vec{W}\right)\right)\right)^T \mathbf{R}_k^{-1}\left(\vec{Y}_k\right.\right. \\ \left.\left. - h_k\left(M^{0\to k}\left(\vec{X}_o^b + \mathbf{B}^{1/2}\vec{W}\right)\right)\right)\right] \end{aligned} \tag{10}$$

Through Taylor expansion to the first order, we can express

$$M^{0\to k}\left(\vec{X}_o^b + \mathbf{B}^{1/2}\vec{W}\right) \approx M^{0\to k}\left(\vec{X}_o^b\right) + \widetilde{\mathbf{M}}_k\mathbf{B}^{1/2}\vec{W};\ \partial\vec{X}_k \approx \widetilde{\mathbf{M}}_k\mathbf{B}^{1/2}\vec{W}$$

$$\vec{Y}_k - h_k\left(M^{0\to k}\left(\vec{X}_o^b + \mathbf{B}^{1/2}\vec{W}\right)\right) \approx \vec{Y}_k - \left[h_k\left(M^{0\to k}\left(\vec{X}_o^b\right)\right) + \widetilde{\mathbf{H}}_k\widetilde{\mathbf{M}}_k\mathbf{B}^{1/2}\vec{W}\right]$$

where, in the second equation, the extreme right expression is TLM approximation to a composite of two functions. In taking the above approximations and remembering that this is a strong-constraint formulation, a condition required is that $\partial\vec{X}_k = \mathbf{M}^{0\to k}\left(\partial\vec{X}_o\right) \approx \widetilde{\mathbf{M}}_\mathbf{k}\mathbf{B}^{1/2}\vec{W}$ must be sufficiently small so that $\widetilde{\mathbf{M}}_k$ seamlessly approximates the evolution of these perturbations throughout the assimilation window.

If we let the departure of the observations with respect to the evolution of the background guess at t_k equal to $\vec{D}_k = \vec{Y}_k - h_k\left(M^{0\to k}\left(\vec{X}_o^b\right)\right)$ then

$$J(\vec{W}) = \frac{1}{2}\left[(\vec{W})^T(\vec{W})\right] + \frac{1}{2}\sum_{k=0}^{k=N}\left[\left(\vec{D}_k - \widetilde{\mathbf{H}}_k\widetilde{\mathbf{M}}_k\mathbf{B}^{1/2}\vec{W}\right)^T\mathbf{R}_k^{-1}\left(\vec{D}_k - \widetilde{\mathbf{H}}_k\widetilde{\mathbf{M}}_k\mathbf{B}^{1/2}\vec{W}\right)\right] \tag{11}.$$

Obvious, in the first term of equation (9), is the omission of an expensive $\mathbf{B}^{-1}$ (with a condition number $\kappa \gg 1$) - to an identity matrix (I) with the best condition number ($\kappa = 1$) and errors in $\vec{W}$

uncorrelated. Computational-wise, rather than dealing with a full field $\vec{X}_o$ with all sorts of correlations, as in SC4DVar-L, the cost function is now explicitly expressed in the affordable control variable $\vec{W}$ - under the assumption that the full field analysis increments, $\partial\vec{X}_o$, are the linear combinations of the column vectors of $\mathbf{B}^{1/2}$. The gradient to the quadratic in equation (11),

$$\nabla J(\vec{W}) = \vec{W} + \sum_{k=0}^{k=N} \left[\left(\widetilde{\mathbf{H}}_k \widetilde{\mathbf{M}}_k \mathbf{B}^{1/2} \right)^T \mathbf{R}_k^{-1} \left(\vec{D}_k - \widetilde{\mathbf{H}}_k \widetilde{\mathbf{M}}_k \mathbf{B}^{1/2} \vec{W} \right) \right] \tag{12}$$

and at extreme when $\nabla J(\vec{W}) = 0$

$$\vec{W} = \frac{\sum_{k=0}^{k=N} \left(\widetilde{\mathbf{H}}_k \widetilde{\mathbf{M}}_k \mathbf{B}^{1/2} \right)^T \mathbf{R}_k^{-1} \vec{D}_k}{I + \sum_{k=0}^{k=N} \left[\left(\widetilde{\mathbf{H}}_k \widetilde{\mathbf{M}}_k \mathbf{B}^{1/2} \right)^T \mathbf{R}_k^{-1} \left(\widetilde{\mathbf{H}}_k \widetilde{\mathbf{M}}_k \mathbf{B}^{1/2} \right) \right]} \tag{13}$$

Then, $\vec{X}_o = \vec{X}_o^b + \mathbf{B}^{1/2}\vec{W}$ at convergence, after a few of iterations.

In equation 13, the merit consequent to preconditioning is reflected in the denominator term, to which an inverse is required to solve for the control variables $\vec{W}$: the additive identity matrix, I, has all eigenvalues ($\lambda_i = 1$), and the term in the brackets takes on ($\lambda_i \geq 0$) due its positive definiteness (or semi-positive definiteness) following $\mathbf{R}_k$ being symmetric and positive definite. Therefore, the inverse of the resulting matrix is more stable with all $\lambda_i \geq 1$. In fact, if the assimilation is limited to time (k=0), it can easily be shown that the denominator term in equation 13 is the Hessian matrix in a preconditioned (using $\mathbf{B}^{1/2}$) 3DVar scheme; reducing Hessian matrix condition number accelerates the convergence rate. Further, because all the eigenvalues are greater than 0, the Gaussian curvature (product of all eigenvalues) and mean curvature (mean of all eigenvalues) are constrained to greater than zero, implying that the

curvature in any direction is positive and the iso-surfaces of the cost function are nearly spherical (or the shape of the $J(\overrightarrow{W})$ is convex).

Still in equation 13, there is an implicit propagation of Background errors through $\left(\widetilde{\mathbf{M}}_k \mathbf{B} \widetilde{\mathbf{M}}_k^{\mathbf{T}}\right)$. This leads to the quasi generation of flow-dependent covariance structures through the data assimilation window (Lorenc, 2003). This is considered as one of the reason why 4DVar is expected to perform better than 3DVar schemes that strictly implement static time-invariant background error covariance. At high spatial resolution, with enough computation power, such flow structures could be vital in resolving small scale ionosphere density structures: the downside in the SC4DVar-Inc is that the evolved structures are not utilized in the next assimilation cycle, which initializes with a new static **B**. Hybrid data assimilation techniques (in which **B** is a linear combination of climatological and ensemble covariances) have been proposed as remedy and are worth investigating in the future. In this work, we try to estimate the analysis errors (B_a) and forecast of background errors(B_f) for the next cycle by using the traditional Kalman filter approach and a Gauss Markov filter (Bust et al., 2004):

$$\left.\begin{aligned} \mathbf{B}_a &= \mathbf{B} + \mathbf{B}\widetilde{\mathbf{H}}_{k=0}^{T}\left[\widetilde{\mathbf{H}}_{k=0}\mathbf{B}\widetilde{\mathbf{H}}_{k=0}^{T} + \mathbf{R}_{k=0}\right]^{-1}\widetilde{\mathbf{H}}_{k=0}\mathbf{B} \\ \mathbf{B}_f &= [1 - \exp(-2dT/\tau)]\mathbf{B} + \exp(-2dT/\tau)\mathbf{B}_a \end{aligned}\right\} \quad (14)$$

The inverse in equation 14 is efficiently computed after a Cholesky decomposition of the matrix. The process of estimating B_f is continued for a 2-hour analysis period, during which the ionosphere temporal correlation coefficient falls to 70% (Jonathan, 1997; Klobuchar 1980; Ssessanga et al., 2019). After this period, the covariance and background initial guess are initialized to climatological settings.

Covariance design (B)

The result from the preconditioning requires that column of $\mathbf{B}^{1/2}$ span the analysis model space - which implies that an incorrect **B** would have a direct negative impact on the final analysis solution. However, primarily, the actual structure of **B** is not know, and for ionospheric studies, 3-D analyses are mainly of high dimension such that **B** in most cases is too large to be stored or even calculate explicitly. In addition, a full field **B** matrix is quite dense such that optimized large parse matrix solvers are inapplicable. Therefore, the **B** matrix utilized in this work is an approximation with some a prior information (dynamics and physics we know about the system). For example, the plasma structure is based on an a climatological model with knowledge that the plasma is magnetized and any existing correlations imposed in the geomagnetic coordinate system; we presume the variance and correlation fields are separable and define the entries of the former as

$$\Sigma^{ij} = \left(S_b^i * S_b^j\right) \qquad (15)$$

where i and j are indices of the grid points in geomagnetic coordinate system using AACGM (Altitude-adjusted corrected geomagnetic coordinates, Shepherd, (2014)); S_b^{α} $(\alpha = i, j)$, is the standard deviation at grid points α, assumed to be 40% of the background density (X_b^{α}). Although the S_b^{α} values seems high, from our experience it is better to start off with inflated background errors such that in situations where the observation errors are overestimated the scheme does not entirely ignore the observations and revert the analysis to X_b^{α}.

To build 3D correlations, the horizontal $\left(h(d_{ij})\right)$ and vertical $\left(V(z_i, z_j)\right)$ components between any two grid points are also assumed separable and modeled independently

$$C_{Vh}^{ij} = h(d_{ij}) * V(z_i, z_j) \qquad (15)$$

where d_{ij} is the horizontal distance between the two points and z_i, z_j are the associated vertical levels. d_{ij} is measured in great circle distance on a sphere.

To define the shapes of $h(d_{ij})$ and $V(z_i, z_j)$, the correlation is assumed to decrease with increasing distance of separation between points i and j. The tapering off in the horizontal and vertical is described with elliptical and normal Gaussian, respectively;

$$\left.\begin{aligned} h(d_{ij}) &= \exp\left(-\frac{(d_{ij})^2}{\left(L_H^{ij}\right)^2}\right) \\ V_h(z_i, z_j) &= \exp\left(\frac{-(z_i - z_j)^2}{\left(L_z^{ij}\right)^2}\right) \end{aligned}\right\} \tag{16}$$

L_H^{ij} and L_z^{ij} are the horizontal and vertical scale lengths, with L_H^{ij} defined as

$$\frac{1}{\left(L_H^{ij}\right)^2} = \left(\frac{(\sin\alpha)^2}{L_\emptyset^2} + \frac{(\cos\alpha)^2}{L_\theta^2}\right) \tag{17}$$

where α is azimuth angle between i and j points, and L_θ and $L_\emptyset$ are latitude and longitude scale lengths (Bust et al., (2004)). For more sparseness, we also add a filter that forces the covariance to go to zero if C_{Vh}^{ij} is less than 0.05. There exits other forms of decorrelations that need future investigation, particularly the second and third order auto regressive models (Gaspari and Cohn (1999)), generally because, compared to the Gaussian, the former have a quicker taper off, hence introducing more sparseness and less variability.

Based on the parameters in Bust et al. (2004) together with experience from our previous analyses, we fined tuned $L_\theta(L_\emptyset)$ for mid-latitudes as 8 (15) degrees. Vertically, Lz is set to vary from 100 km in the E and F regions to 500 km in the plasmasphere. Albeit in reality these length scales should vary with changing ionospheric conditions or external deriving factors, we could not find literature extensively discussing ionospheric correlations in 4D-space under different ionospheric dynamics: such an extensive study would be outside the scope of this work. Thus,

what is presented is the static component of in-house tuned parameters learned from experience. Of course a severe mismatch is expected to arise when the ionosphere is populated with small scale density irregularities or in low density ionosphere with sharp gradients particularly in E-region; an algorithm weakness that needs to be addressed in future investigations.

The final entries of **B** are defined as

$$B^{ij} = \left(S_b^i * S_b^j\right) * C_{Vh}^{ij} \tag{18}$$

2.3 Strong constrained 4D ensemble -Var (SC4DEn-Var) formulation

In both the Lagrangian and incremental approach, the designing, realization and maintenance of the TLMs and MAs is complicated and computationally expensive: operational wise, these are some of the major drawbacks of the SC4DVar-Inc. The notion here is to circumvent these cumbersome steps: the increment $\partial\vec{X}_o$ in (8), is assumed to be made up of a weighted linear combination of perturbations computed as scaled differences between ensemble members and the ensemble mean, as discussed below.

Consider an ensemble ($\mathbf{X}_o^b$) made up of "q" regional electron density ionospheres at time $t_{k=0}$

$$\mathbf{X}_o^b = \left[\vec{X}_{o1}^b, \vec{X}_{o2}^b, \cdots, \vec{X}_{oj}^b, \cdots, \vec{X}_{oq}^b\right] \tag{19}$$

The normalized ensemble of perturbations is expressed as

$$\widehat{\mathbf{X}}_o^b = \left[\frac{\vec{X}_{o1}^b - \overline{X}}{\sqrt{q-1}}, \frac{\vec{X}_{o2}^b - \overline{X}}{\sqrt{q-1}}, \cdots, \frac{\vec{X}_{oj}^b - \overline{X}}{\sqrt{q-1}}, \cdots, \frac{\vec{X}_{oq}^b - \overline{X}}{\sqrt{q-1}}\right] \text{ where } \overline{X} = \frac{1}{q}\sum_{j=1}^{q} \vec{X}_{oj}^b \tag{20}$$

The normalization is by $\sqrt{q-1}$ because one degree of freedom is lost (the sum of perturbations now equal to zero (unbiased)). If the q column vectors of rectangular matrix, $\widehat{\mathbf{X}}_o^b$, are random samples from a distribution with a covariance $\mathbf{B_{en}}$, then, we can estimate the later (at $t_{k=0}$) as

$$\left.\begin{array}{c} \mathbf{B_{en}} \approx \widehat{\mathbf{X}}_{\mathrm{o}}^{\mathrm{b}} \widehat{\mathbf{X}}_{\mathrm{o}}^{\mathrm{b}^{\mathrm{T}}} \\ \mathbf{B}_{\mathbf{en}}^{\frac{1}{2}} \approx \widehat{\mathbf{X}}_{\mathrm{o}}^{\mathrm{b}} \end{array}\right\} \tag{20}$$

This is where SC4DVar-Inc and SC4DEnVar precisely start to differ. Subscript "en" indicates that this is ensemble covariance. Substituting these new entities into (13)

$$\overrightarrow{\mathrm{W}} = \frac{\sum_{\mathrm{k}=0}^{\mathrm{k}=\mathrm{N}} \left(\widehat{\mathbf{Y}}_{\mathrm{k}}^{\mathrm{b}}\right)^{\mathrm{T}} \mathbf{R}_{\mathrm{k}}^{-1} \overrightarrow{\mathrm{D}}_{\mathrm{k}}}{\mathrm{I} + \sum_{\mathrm{k}=0}^{\mathrm{k}=\mathrm{N}} \left[\left(\widehat{\mathbf{Y}}_{\mathrm{k}}^{\mathrm{b}}\right)^{\mathrm{T}} \mathbf{R}_{\mathrm{k}}^{-1} \widehat{\mathbf{Y}}_{\mathrm{k}}^{\mathrm{b}}\right]} \tag{21}$$

where we have assumed

$$\widetilde{\mathbf{H}}_{\mathrm{k}} \widetilde{\mathbf{M}}_{\mathrm{k}} \widehat{\mathbf{X}}_{\mathrm{o}}^{\mathrm{b}} \approx \widetilde{\mathbf{H}}_{\mathrm{k}} \widehat{\mathbf{X}}_{\mathrm{k}}^{\mathrm{b}} = \widehat{\mathbf{Y}}_{\mathrm{k}}^{\mathrm{b}} \tag{22}$$

and the ensemble of perturbations $\widehat{\mathbf{Y}}_{\mathrm{k}}^{\mathrm{b}}$ can in a simplified form be estimated as follows

$$\left.\begin{array}{c} \widehat{\mathbf{Y}}_{\mathrm{k}}^{\mathrm{b}} = \left[\frac{\mathrm{h_k}(\mathrm{M}^{0\to\mathrm{k}}(\vec{\mathrm{X}}_{\mathrm{o1}}^{\mathrm{b}})) - \bar{\mathrm{Y}}}{\sqrt{\mathrm{q}-1}}, \frac{\mathrm{h_k}(\mathrm{M}^{0\to\mathrm{k}}(\vec{\mathrm{X}}_{\mathrm{o2}}^{\mathrm{b}})) - \bar{\mathrm{Y}}}{\sqrt{\mathrm{q}-1}}, \cdots, \frac{\mathrm{h_k}(\mathrm{M}^{0\to\mathrm{k}}(\vec{\mathrm{X}}_{\mathrm{oj}}^{\mathrm{b}})) - \bar{\mathrm{Y}}}{\sqrt{\mathrm{q}-1}}, \cdots, \frac{\mathrm{h_k}(\mathrm{M}^{0\to\mathrm{k}}(\vec{\mathrm{X}}_{\mathrm{oq}}^{\mathrm{b}})) - \bar{\mathrm{Y}}}{\sqrt{\mathrm{q}-1}}\right] \\ \bar{\mathrm{Y}} = \frac{1}{\mathrm{q}} \sum_{\mathrm{j}=1}^{\mathrm{q}} \mathrm{h_k}(\mathrm{M}^{0\to\mathrm{k}}(\vec{\mathrm{X}}_{\mathrm{oj}}^{\mathrm{b}})) \end{array}\right\} \tag{23}$$

The advantage of utilizing ensemble perturbations are eminent in equation 21: the computation becomes inexpensive with no need to compute the TLMs and MAs.

In particular, in equation 23 (used in 21) $\mathrm{h_k}$ is applied directly to each ensemble member or background fields ($\vec{\mathrm{X}}_{\mathrm{oj}}^{\mathrm{b}}$), implying that different types of observations with a non-linear relationship with electron density can now easily get incorporated into the algorithm (e.g., satellite UV radiances related to the O+ density that is a major ion of the F-layer). Flow-dependence: note that in equation 22, each ensemble member, ($\vec{\mathrm{X}}_{\mathrm{oj}}^{\mathrm{b}}$), is propagated to time sample $\mathrm{t_k}$ through the full-field prognostic model, $\mathrm{M}^{0\to\mathrm{k}}(\vec{\mathrm{X}}_{\mathrm{oj}}^{\mathrm{b}})$, and the associated covariance $\left(\mathbf{B}_{\mathrm{en}}^{\mathrm{k}}\right)$ at $\mathrm{t_k}$ is expressible as $\left(\mathbf{X}_{\mathrm{k}}^{\mathrm{b}} - \overline{\mathbf{X}}_{\mathrm{k}}^{\mathrm{b}}\right)\left(\mathbf{X}_{\mathrm{k}}^{\mathrm{b}} - \overline{\mathbf{X}}_{\mathrm{k}}^{\mathrm{b}}\right)^{\mathrm{T}} / \mathrm{q}-1 = \widehat{\mathbf{X}}_{\mathrm{k}}^{\mathrm{b}} \widehat{\mathbf{X}}_{\mathrm{k}}^{\mathrm{b}}$: this is a significant difference from the SC4DVar-Inc that implicitly evolves the covariance through $\left(\widetilde{\mathbf{M}}_{\mathrm{k}} \mathbf{B} \widetilde{\mathbf{M}}_{\mathrm{k}}^{\mathbf{T}}\right)$; this difference is

apparent when nonlinear processes dominate the ionosphere (and the SC4DVar-Inc TLM approximations become erroneous) or when SC4DEnVar ensemble members are very diverse. Indeed, results from our experimental run show SC4DVar-Inc as having an inferior performance in reconstructing ionosphere densities during severe storm conditions (see section 3).

The disadvantage of SC4DEnVar approach is the loss of information on the sensitivities (we no-longer compute MA's), which are vital when analysing the sensitivity of the model state to an arbitrary change to input parameters (particularly vital if the prognostic model is switched to a physics-based model with multiple input drivers).

Ensemble preparation: to obtain the j^{th} ensemble member, we run the climatological empirical International reference ionosphere model (IRI; version-2016, Bilitza et al., 2017) using a set of parameters/indices (pr) randomly sampled from Gaussian distributions: the original parameters contained in IRI-2016 are set as the best-guess or mean (μ_j^{pr}) and the variance specified as $\left(\sigma_j^{pr}\right)^2 = \left(0.3 * \mu_j^{pr}\right)^2$. The value 0.3 was determined from experience. Because IRI-2016 only extends to a maximum of 2,000 km in altitude, the extension to GNSS orbital altitudes ($\approx 22{,}000$ km) is covered by the NeQuick model. The perturbed parameters (pr) include solar (Rz12, F10.7), magnetic (Kp), and ionospheric indices (IG12).

R12 is the 12-month running mean of sunspot number used to describe the solar activity: utilized in estimating hmF2, foF2 (when using the CCIR (1967) recommended for continent areas), foF1, the bottom-side thickness and topside (if using NeQuick option).

F10.7 is a measure of solar radio flux at 10.7 cm wavelength and correlates better with EUV irradiance from the Sun. Among the many uses in IRI, a variation of the F10.7 index is used in computing hmF2 with the model of choice as SDMF2 (Satellite and Digisonde Model of the F2-Layer Height; developed based on a large amount of radio occultation (RO) and data from

digisondes (Shubin et al., 2013; Shubin, 2015)). Evaluations of the different hmF2 models in IRI-2016 have found SDMF2 a better candidate (Bilitza et al., 2021; Moses et al., 2021; Huang et al., 2021 and Mengist et al., 2020) and therefore selected for use in this work.

IG12 is a 12-month running mean index of IG (Ionosonde Global) determined with data from four reference stations distributed in the Northern (UK (Port Stanley) and Canada (Canberra)) and Southern (Japan (Kokubunji) and UK (Chilton)) Hemispheres. In IRI-2016, IG12 is crucial in estimating foF2 since it is from ionospheric measurements and thus includes solar cycle changes in the F-region ionosphere.

Kp is the planetary activity index indicating the level of global geomagnetic field disturbance. It is a 3-hr index derived from measurements collected at 13 sub-auroral magnetic observatories. IRI-2016 uses Kp when specifying auroral boundaries: it is vital to include the Kp index variations because, during extreme geomagnetic storms, the aurora boundaries can extend to mid-latitudes latitudes introducing electron density gradient at those latitudes.

Localization: equation 21 has highlighted some of the advantages of using ensembles. However, the same has problems associated rank deficiency and under-sampling noise. Rank deficiencies- in the denominator of equation 21, the maximum associated rank of the term in the brackets is the minimum between m and q. Therefore, if we consider that the assimilated observations are only available at time sample k, and the number of observations (m) is greater than q, $\left(\widehat{\mathbf{Y}}_{\mathrm{k}}^{\mathrm{b}}\right)^{\mathrm{T}}\mathbf{R}_{\mathrm{k}}^{-1}\widehat{\mathbf{Y}}_{\mathrm{k}}^{\mathrm{b}}$ will be a singular matrix.

Under-sampling- the ionosphere hosts a wide range of frequencies, which for the most part, if the imperfect observations and an inaccurate model (probably because the ensemble sample size is too small to capture all the dynamics) describe the evolution of high-frequency modes, the

384 resultant uncontrolled spurious modes can inexorably grow and introduce artefacts in the
385 analysis.
386 The simple cure to rank deficiency and under-sampling noise is to increase the ensemble sample
387 size (q). However, the dimensionality scales with increasing q, and thus for operational
388 applications, the computation expense and time constraints tied to this simple approach limit its
389 realization. An elegant way is through localization (Houtekamer and Mitchell (2001)); a matrix
390 $\mathbf{L} \in \mathbb{R}^{P \times P}$ is employed such that in $\mathbf{B_{en}}$, correlations between grid points are gradually dumped
391 with increasing geomagnetic distance of separation (the general physical interpretation is that
392 grid points far apart should have no statistical relationship; although their could be situations
393 where distant points are correlated):

$$\mathbf{B_{en}} \approx \left(\widehat{\mathbf{X}}_o^b \widehat{\mathbf{X}}_o^{b\,T}\right) \circ \mathbf{L} \qquad (23)$$

394 where ∘ symbolizes the Schur (element-wise) product. To maintain the positive definiteness of
395 $\mathbf{B_{en}}$, $\mathbf{L}$ is a square matrix that is symmetric and positive definite, here designed with entries
396 similar to C_{Vh}^{ij} as in the SC4DVar-Inc. The tuning of $\mathbf{L}$ or selection of length-scales is discussed
397 later in subsection 2.3.1.
398 ***Effects of localization:*** the top panel in Figure 1 illustrates an example of $\mathbf{B_{en}}$ non-localized (left)
399 and localized (right) versions; the insets are a zoom-in to the first 200 × 200 grid points. The
400 colour bar is in the far right bottom corner. Unlike in the non-localized version, a taper-off of
401 long-distance correlations is eminent in the right panel, with maximum power (variance) along
402 the diagonal where correlation is maximum. The middle panel is an extraction of the associated
403 singular values in descending order, with red (blue) corresponding to localized (non-localized) $\mathbf{B}$;
404 broken axes in the left subplot indicate the extreme ends of the spectrum. The left subplot shows
405 the cumulative percentage of the variance captured by a given number of modes. Before

localization, a large proportion of the total power is concentrated probably within the first six modes and then a sharp decay into the high frequency or noisy regime; indeed, the cumulative curve indicates that the first two modes describe ~ 98% (blue dashed vertical lines) of the variance in $\mathbf{B_{en}}$ **-** in reference to equations 8 and 9, this implies that the freedom of choice on the number of modes that could span our regional ionosphere is significantly limited. After localization, the decay in the variance of the leading modes is slower, and the total number contributing 98% (red dashed vertical lines) of total variance has increased to more than 100; thus extending the flexibility. Since $\widehat{\mathbf{X}}_0^b\widehat{\mathbf{X}}_0^{b^T}$ and $\mathbf{L}$ are positive definite matrices and later has all diagonal entries as 1, the trace of $\mathrm{Tr}\left(\left(\widehat{\mathbf{X}}_0^b\widehat{\mathbf{X}}_0^{b^T}\right)\circ\mathbf{L}\right) = \mathrm{Tr}\left(\widehat{\mathbf{X}}_0^b\widehat{\mathbf{X}}_0^{b^T}\right)$ is also equivalent to the sum of all singular values of $\widehat{\mathbf{X}}_0^b\widehat{\mathbf{X}}_0^{b^T}$. Therefore, the $\mathbf{L}$ matrix redistributes or balances the energy across the spectrum - not altering the total signal strength but taking energy from the trailing modes and "pouring" it into major leading mod; this explains the drop in the amplitudes of the first two and trailing singular values. With a flattened spectrum, the distance between the singular values significantly reduces, and $\mathbf{B_{en}}$ becomes more numerically stable, as indicated here with condition number (κ) reducing nearly by two orders of magnitude (see upper right hand corner of the left subplot). The downside to localization is the significant reduction in the weight of the leading modes; that is, the approximated $\mathbf{B_{en}}$ after localization is now of lower amplitude if only estimated from the first few modes; for our application, this implies an underestimation of electron densities, and actually observed in our simulation analysis (sub-subsection, 2.3.1).

The bottom panel shows the first three leading orthogonal modes. Left: is the vertical extraction, and for clarity, we limit the extent to 1000km in altitude. Right: is horizontal extraction. Modes in the upper (lower) plate are from a non-localized (localized) $\mathbf{B_{en}}$. In the non-localized $\mathbf{B_{en}}$, the modes are diverse in frequency and nearly infiltrate the whole domain, pointing to the aspect of

429 the presence of various modes that can negatively impact the analysis. After localization,
430 infiltration is only to a few grid points, and the structure of the modes tends toward the shape of
431 the slow manifold. Therefore, localization acts as a filter for high frequencies and bends the
432 analysis towards the low frequencies. In the localized vertical modes, diverse variation is quasi-
433 limited to the F-region (200~700km) and nearly zero elsewhere. Thus, we expect minor
434 corrections to the E-region and the plasmasphere. With this understanding, the aim is to select
435 the appropriate number (r) of modes that can resolve low-frequency waves of interest and neglect
436 the noisy high-frequency variations, possibly due to under-sampling.

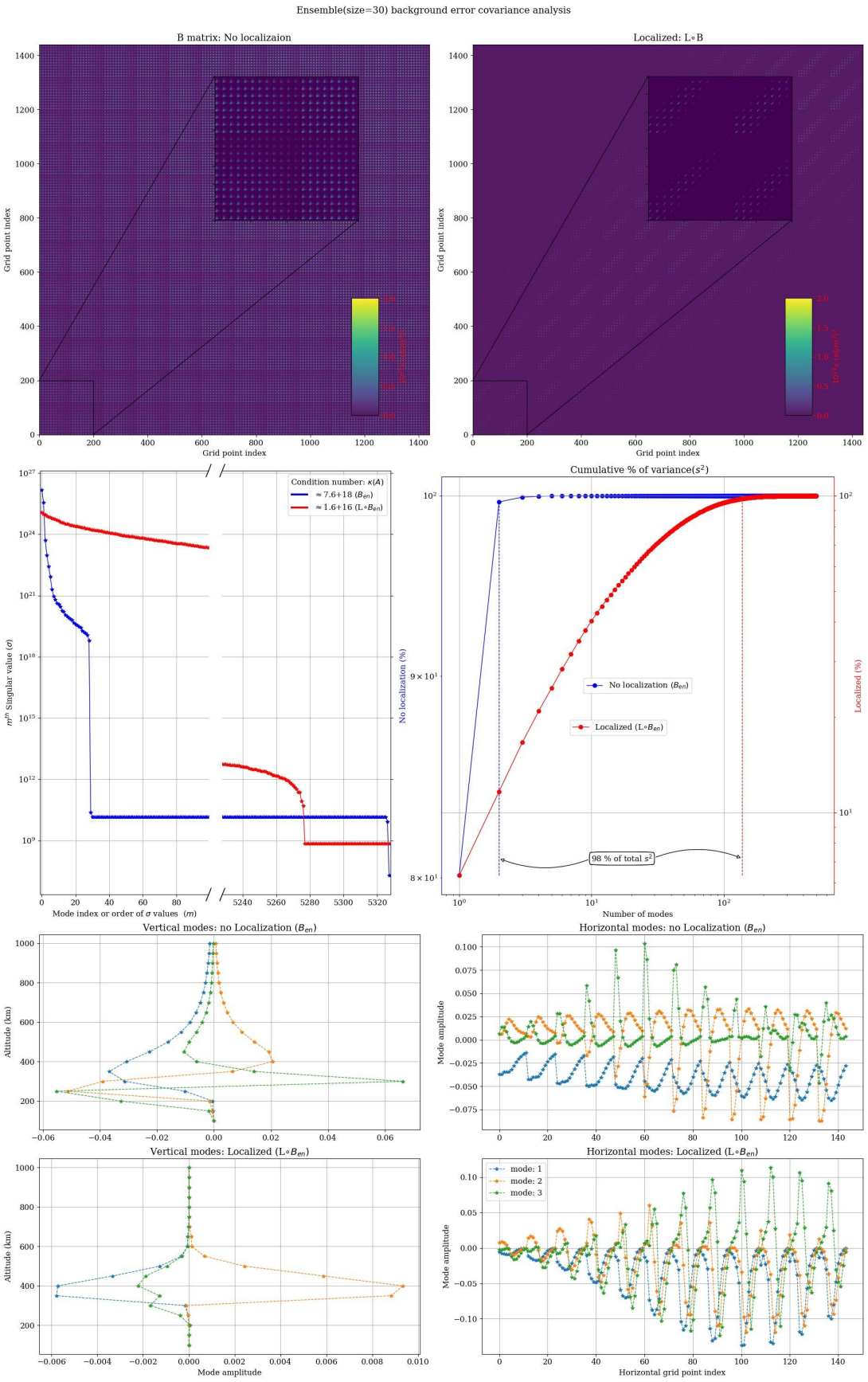

Ensemble(size=30) background error covariance analysis
B matrix: No localizaion
Localized: L∘B
Grid point index
Grid point index
Condition number: κ(A)
≈ 7.6+18 (B_en)
≈ 1.6+16 (L∘B_en)
m^th Singular value (σ)
No localization (%)
Mode index or order of σ values (m)
Cumulative % of variance(s²)
No localization (B_en)
Localized (L∘B_en)
98 % of total s²
Localized (%)
Number of modes
Vertical modes: no Localization (B_en)
Horizontal modes: no Localization (B_en)
Vertical modes: Localized (L∘B_en)
Horizontal modes: Localized (L∘B_en)
Altitude (km)
Mode amplitude
mode: 1
mode: 2
mode: 3
Horizontal grid point index

Figure 1: Ensemble covariance (**B**) analysis: top panel illustrates an example of a non-localized (left) and localized (right) **B** matrix; inset axes zoom in to the first 200 × 200 grid points. Middle panel: left is an extract of **B** matrix singular values, with broken axes indicating extreme ends of the spectrum - Blue (red) is non-localized (localized). Right is captured variance in **B** (represented in percentage) as the number of representative orthogonal modes increases - left(right) y-scale is for non-localized (localized) **B**. Blue (red) dashed vertical lines indicate the points where the variance captured is at 98% in non-localized (localized). Bottom panel: an example of the three leading orthogonal modes of **B**; left (right) is vertical (horizontal) modes - bottom and top represented modes from localized and non-localized **B**, respectively.

For brevity, we will not give a mathematical derivation, but after localization

$$\mathbf{B}_{\mathbf{en}}^{1/2} = \widehat{\mathbf{X}}_{\mathrm{o_L}}^{\mathrm{b}} = \left[\mathrm{diag}(\vec{\mathrm{X}}_{\mathrm{o1}}^{\mathrm{b}})\mathbf{L}^{1/2}, \mathrm{diag}(\vec{\mathrm{X}}_{\mathrm{o1}}^{\mathrm{b}})\mathbf{L}^{1/2}, \cdots, \mathrm{diag}(\vec{\mathrm{X}}_{\mathrm{oq}}^{\mathrm{b}})\mathbf{L}^{1/2}\right] \tag{24}$$

where $\mathbf{L}^{1/2} \in \mathbb{R}^{\mathrm{P}\times\mathrm{r}}$ and $\widehat{\mathbf{X}}_{\mathrm{o_L}}^{\mathrm{b}} \in \mathbb{R}^{\mathrm{P}\times\mathrm{qr}}$, subscript L indicates a localized version of the ensemble and $\mathrm{diag}(\cdot)$ represents a diagonal matrix with diagonal elements as entries of the vector (Buehner, 2005). Notice that the increase in dimension from "q" to q*r abates the rank deficiency but also increases computation expense. However, because q*r is generally much less than the size of the ensemble sufficient to capture all dynamics without localization, this approach is more computationally efficient. Another note is the implementation of the same $\mathbf{L}^{1/2}$ through the assimilation window. This unique localization is a common approach that lowers the expense, but correlations should have a spatial-temporal change as the errors evolve through the assimilation window - this weakness needs to be addressed if the SC4DEnVar scheme is to resolve small-scale fast-evolving ionosphere density structures. Next, we discuss realization of $\mathbf{L}^{1/2}$ and the selection of appropriate modes.

Computing the square root B: in equation 13 and 24 we are required to compute $\mathbf{B}^{1/2}$ and $\mathbf{L}^{1/2}$, respectively. Since **B** and **L** are square symmetric positive definite matrices, then there should exist non-negative unique square roots matrices $\mathbf{B}^{1/2}$ and $\mathbf{L}^{1/2}$; to compute the square roots we decompose a matrix $\mathbf{A}\in \mathbb{R}^{\mathrm{P}\times\mathrm{P}}$ into left (right) $\mathbf{U}\in \mathbb{R}^{\mathrm{P}\times\mathrm{P}}$ **(**$\mathbf{V}\in \mathbb{R}^{\mathrm{P}\times\mathrm{P}}$**)** singular vectors (e.g., Figure

1 bottom panel) and a symmetric diagonal matrix ($\mathbf{D} \in \mathbb{R}^{P\times P}$) comprising of singular values (e.g., Figure 1 middle panel)- using singular value decomposition (SVD); $\mathrm{SVD}(\mathbf{A}) = \mathbf{UDV}^{\mathrm{T}}$. For square symmetric positive definite matrices $\mathbf{U} = \mathbf{V}$, $\mathbf{U}^{-1}$=$\mathbf{U}^{\mathrm{T}}$such that

$$\left.\begin{aligned} \mathrm{SVD}(\mathbf{A}) &= \left(\mathbf{UD}^{1/2}\right)\left(\mathbf{UD}^{1/2}\right)^{\mathrm{T}} \\ \text{and}\quad \mathbf{A}^{1/2} &= \mathbf{U}_{P\times P}\mathbf{D}_{P\times P}^{1/2} \end{aligned}\right\} \tag{25}$$

As mentioned earlier, only a subset of the spectrum that captures the low frequencies is required. Then, $\mathbf{L}^{1/2} \approx \mathbf{U}_{p\times r}\,\mathbf{D}_{r\times r}^{1/2}$.

The SVD algorithm is numerically stable and freely available in most scientific computation packages, but the high computation cost limits its utilization when large dimensions are involved. As a mitigation, we use randomized linear algebra (hereinafter referred to as randomized SVD; Halko et al., 2011; Boullé & Townsend, 2021): the idea is to find an orthonomal matrix $\mathbf{Q} \in \mathbb{R}^{P\times r}$, with columns that form an approximate basis for the column space of $\mathbf{A}$. In other words, $\mathbf{A} \approx \mathbf{QQ}^{*}\mathbf{A}$, and "r" singular values and right singular vectors of $\mathbf{Q}^{*}\mathbf{A}$ are the same as those of $\mathbf{A}$; where $\mathbf{Q}^{*}$is the conjugate transpose of $\mathbf{Q}$. Notice that the computational burden is now shifted from SVD of a high dimensional $\mathbf{A} \in \mathbb{R}^{P\times P}$ to SVD of much lower dimensional $\mathbf{Q}^{*}\mathbf{A} \in \mathbb{R}^{r\times p}$ matrix. Also, in terms of storage, because $\mathbf{Q} \in \mathbb{R}^{P\times r}$ and $\mathbf{Q}^{*}\mathbf{A} \in \mathbb{R}^{r\times p}$, only space for r∗(P+P) elements is required instead of space for (P∗P) elements in full $\mathbf{A}$. Randomness becomes important when establishing the $\mathbf{Q}$ matrix: a high probability exist that randomly selected vectors are linearly independent, and, as such, these vectors are used to randomly sample the column space of $\mathbf{A}$, with the sketch matrix computed as $\mathbf{Y} = \mathbf{AO}$; $\mathbf{O} \in \mathbb{R}^{p\times r}$ consists of r column vectors randomly selected from a stand normal distribution. Finally, the required orthonomal matrix $\mathbf{Q}$ is obtained from the QR-decomposition of $\mathbf{Y} = \mathbf{QR}$. If the selected number of modes r is not enough to describe $\mathbf{A}$, r is updated by a factor d, r = r+d and algorithm run again. To

automatically terminate the algorithm, the following crude method is hard-wired into the our algorithm: we normalize all the extracted singular values by the maximum value, and the curve described by these values is shifted by a constant value above 1. From experience, the consequent cumulative will start to saturate when the total captured variance is more than 98 %, and the natural log of the absolute gradient to curve starts to get negative - this is the termination point. The randomized SVD approach, reduces computation complexity of extracting the major modes from $O\,(P^2 * r)$ for traditional SVD, to $O\,(P^2 * \log r)$. Figure 2 shows an example of singular values extracted using traditional SVD (red) and Randomized SVD (black). The singular values are nearly equal, but a zoom-in (see the insert axes; the blue dashed square curve with the scale on the right-hand side is the relative change in the singular values ($\Delta\sigma_r$)) shows that singular values produced by randomization tend to be a bit noisy, though from our analysis, the noise amplitude decreases as "r" is increased and the inaccuracy does not affect our final results.

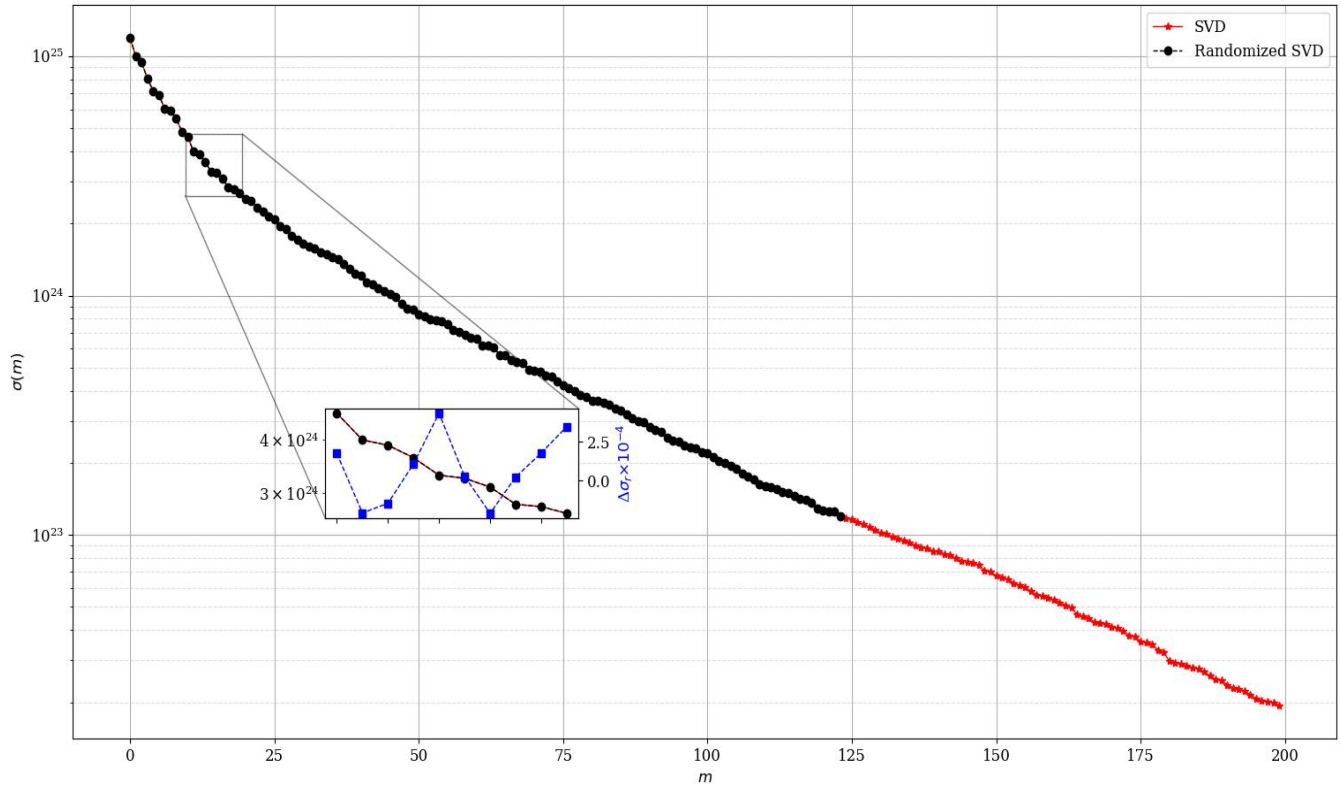


Figure 2: A comparison of singular values obtained using Randomized Singular Value Decomposition (SVD) used in the algorithms against the traditional SVD (red, considered as the "truth"). The zoom-in shows how close the two SVDs match- the blue square-dashed line is the relative change $\left(\frac{[\text{Randomized SVD} - \text{truth}]}{\text{truth}}\right)$.

2.3.1 Tuning of parameters

Formally under covariance analysis, we underscored the importance of localization but a correct choice of scale lengths is still a primary challenge in ionospheric data assimilation: the observation data available on a global and regional scale are insufficient to precisely resolve localized correlation parameters in spatial (horizontal, vertical) and temporal domains, particularly during geomagnetic storms when the ionosphere gets highly chaotic due to an injection of a broad spectrum of energy.

Alternatively, we presume that an understanding from a simulation analysis, in addition to prior knowledge, can guide us and shorten the tedious intricate tuning process: we simulate and analyse the fidelity of reconstructing a dynamic parcel of density in the region of interest while disregarding the underlying physics and chemistry that could seed such an ionospheric response. The interpretation is that the optimal parameters in reconstructing the induced synthetic density gradients are transferable to the real world within an acceptable error range. Because the ionosphere is a chaotic system, it is complex to tune both horizontal and vertical parameters; for simplicity, we limit ourselves to the horizontal parameters and the vertical scales set as in SC4DVar-Inc. The vertical turning will be addressed elsewhere when the focus is on perfecting the performance of SC4DEnVar (the future idea is to estimate localized vertical scales by augmenting on a new set of vertical correlations published by Forsythe et al., 2020;2021, and recently applied in the work of Mengist et al., 2023).

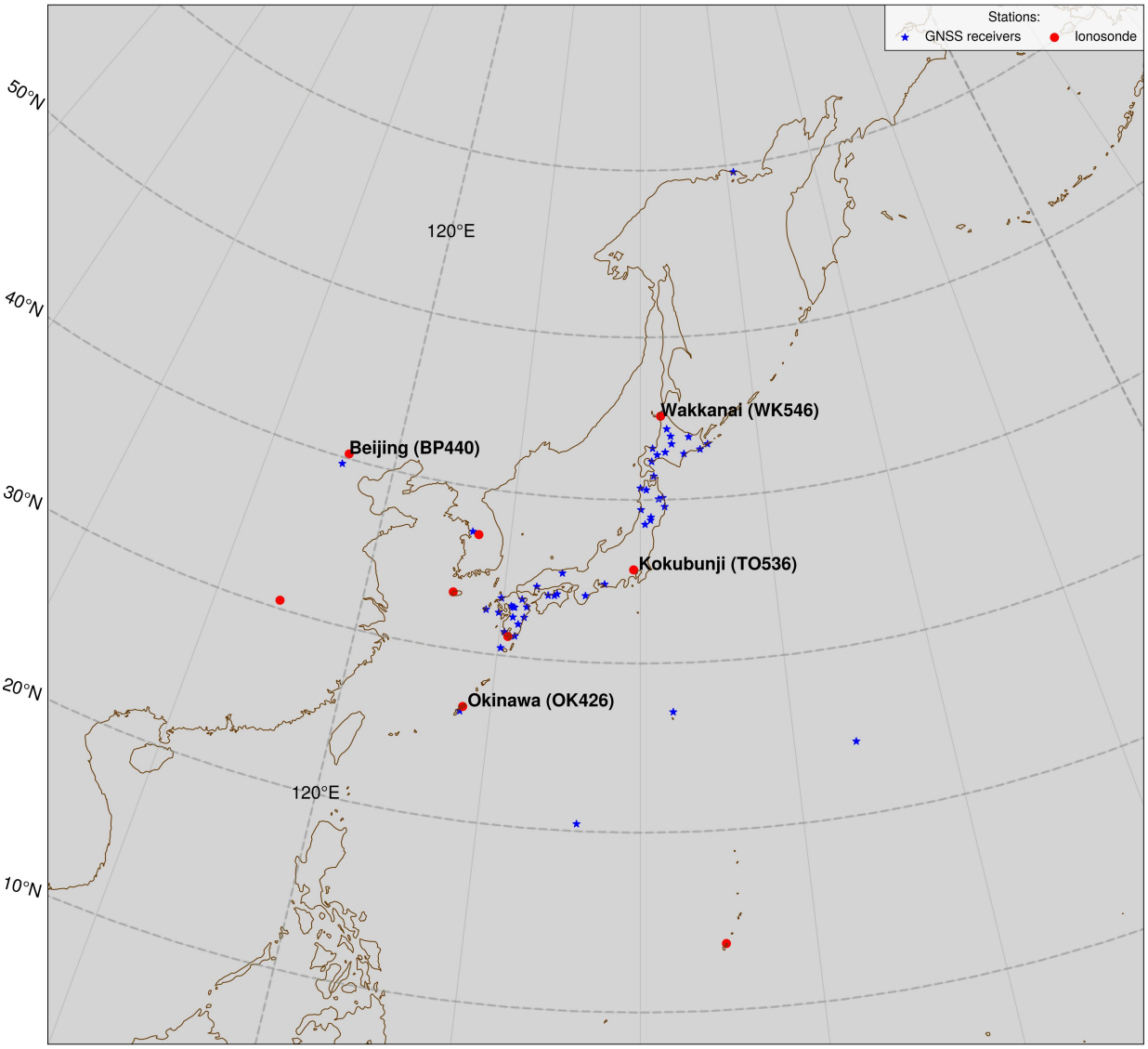


Figure 3: Blue asterisks: geographic location of 50 ground GNSS receivers considered in this study. The minimum separation distance between receivers is 1 degree. Red solid circles: ground ionosondes station in the region of interest. Only labeled stations (location and four-character code) are adopted in our analysis.

***Simulation setup*:** the utilized grid extends 110°-160° in longitude and 20°-65° in latitude at a resolution of 5° degrees. Vertically the resolution is irregular 25 km, and ≧ 200km, for height ranges 100-1000km, 1000-20,000 km, respectively. The resolution is highest in the region where we expect the most electron density variability. To facilitate the reconstructions, we use the exact geometry of GNSS and ground receivers to generate synthetic STEC measurements by

integrating densities along different lines of sight. White noise (standard deviation = 10% of the synthetic STEC) is added to each STEC value to mimic real-world observations. Blue asterisks in Figure 1 show the distribution of ground GNSS receivers in use. The Japanese archipelago has a dense network of more than 1000 ground GNSS receivers, but only use 50 stations with a minimum separation distance of 1 degree: because our focus is on resolving low frequencies on a coarse grid, it is better to use a few well-distributed stations and avoid data redundancy. In a way, this also acts as a spatial data thinning process that lowers the chances of observation correlations and mitigates some of the difficulties associated with misrepresentation of the observation covariance, $\mathbf{R}_k$ (earlier mentioned in section 2). Moreover, neglecting some data speeds up the quality control process and may act to improve the stability of the inverse matrix in equations 13 and 21.

The initial condition is from IRI-2016, and the simulation is run for two 2 hours starting at 04:00 UT 2015-03-14. Figure 4 upper subplots show the target (parcel of density) to reconstruct, with the background removed. The structure is in 3-D, but only the peak layer is shown here with snapshots taken at early, intermediate and late stages of development: at the base (Figure 4 upper left subplot), a density enhancement (30% more than the background densities) develops at low-latitudes F-region with the vertical centre horizontally varying with hmF2-indicated as solid blue contour lines. Outward the parcel extends vertically 100 km and 10 degrees horizontally. With time (Figure 5 middle and far-right subplots), the parcel quasi-spirally migrates northwestward through the moderate densities at mid-latitudes while diffusing out at 2% per 2.5 minutes. As the parcel diffuses, sharp edges become eminent; purposefully designed in this way to monitor the extent to which localization affects fidelity.

554 Reconstructions were performed for different horizontal scale lengths and ensemble sample sizes,
555 and the root mean square error (RMSE) determined as a function of time. The algorithm was
556 defined to have converged if the chi-squared value was less or equal to 0.5. Otherwise, from our
557 preliminary trials, the maximum number of iterations is set to 20; from experience, the optimal
558 solution is generally within the first six iterations. Figure 5 shows the results; in each panel
559 different colours indicate a change in ensemble size from 3 to 33 (in step size of 6). Vertical bars
560 and appended numbers indicate the extent of the spread in RMSE. In nearly all plots, specifically
561 in the left upper panel, RMSE generally increases for 1 hour and then slightly decreases: the
562 increase is due to the compounding errors as the prognostic model fails to describe the correct
563 trajectory of the parcel. As the parcel densities diffuse to other locations with good data coverage,
564 we observe a slight decrease in RMSE. At this point, the prognostic model errors are immense
565 such that the impact of good data coverage is not enough to reproduce the parcel. Therefore, our
566 first draw from the simulation is that the current forward model is not applicable for assimilation
567 windows longer than 1 hour. As a balance between the time-integrated impact of the
568 observations and a limitation of model errors, in this work, we set all assimilation windows to 30
569 minutes.
570 In Figure 5 bottom subplots, the increase in horizontal scale lengths beyond 5(8) degrees in
571 latitude (longitude) exacerbates the reconstruction errors; the RMSE at the base is nearly equal or
572 greater than the RMSE after the structure has evolved for 1.5 hours in Figure 5 top panels.
573 Therefore, we can disregard the bottom panel scale lengths because the base errors should be
574 minimal before prognostic model errors infiltrate the analysis. Besides, the poor estimates in
575 bottom panels can also be associated with spurious correlations due to a weak localization of the
576 ensemble.

In Figure 5 left upper subplot, the scales are smaller than the assigned horizontal grid size - implying $\mathbf{B_{en}}$ is a diagonal matrix with little or no correlation among grid points. Although this case offers the lowest RMSE in the simulation, such a severe localization is not desirable at the operational level; for exactness and completeness, we need the observation at a given location to impact data-void areas. Therefore, the intermediate scales selected as optimal for analysis were 5(8) degrees in latitude(longitude); see the top right panel.

Ensemble size analysis: the focus is on Figure 5 top right panel and a maximum time window of 30 minutes; we observe poor skill when the ensemble size is small (q=3). We anticipate this result following the earlier discussed under-sampling limitations. When the ensemble size increases, we notice an improvement in RMSE, but more than a sample size of 30 does not have a significant improvement on the analysis; the effect of localization is eminent, with the whole pool of ensemble sizes (9-33) nearly falling within the same error range. Thus, with localization, there is no advantage in using large ensemble sample sizes. In application, we chose an ensemble size of 30 members as a balance between the computation expense and sufficiently capturing the ionospheric variability.

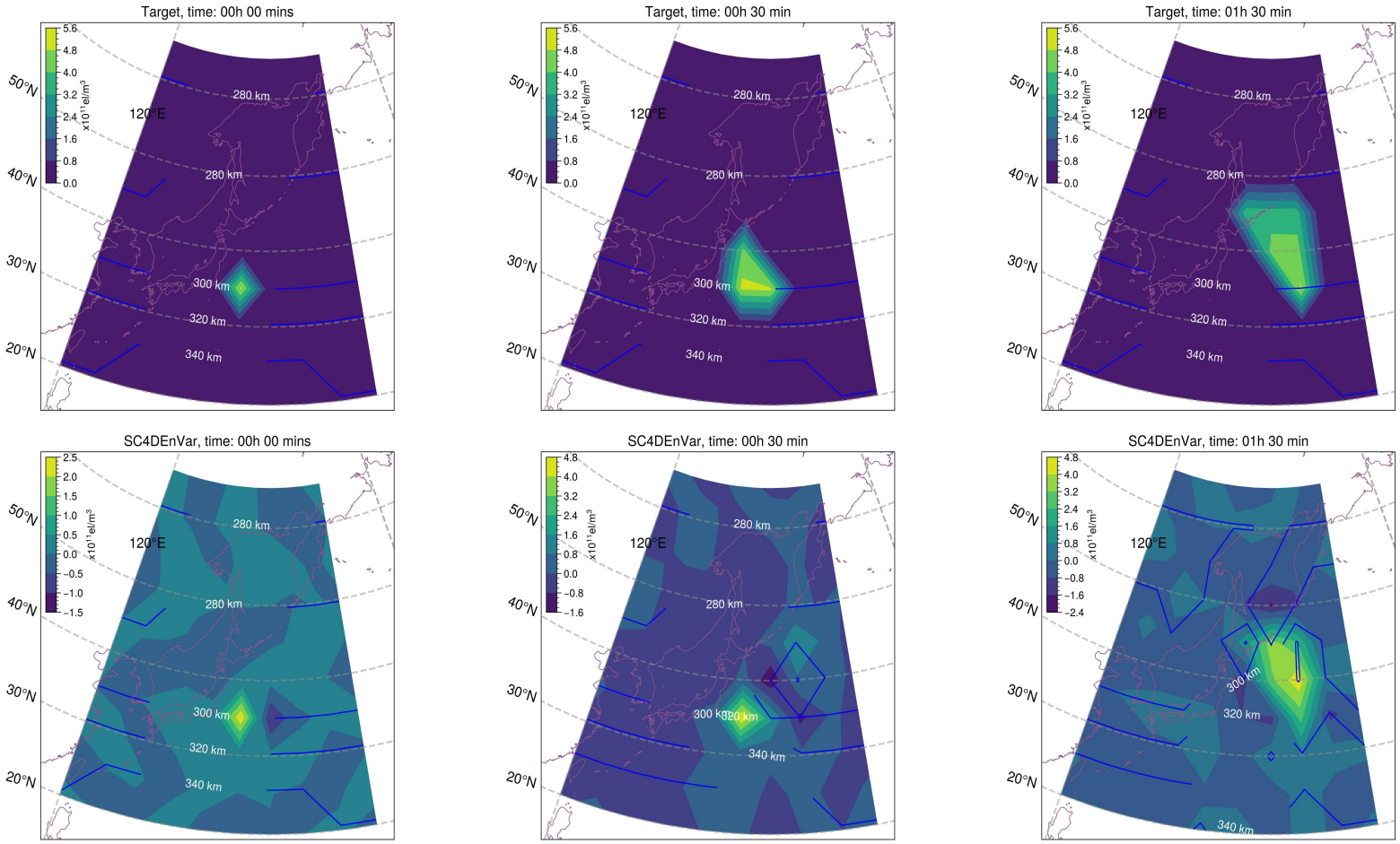


Figure 4: SC4DEnVar simulation reconstruction analysis; top panels show the target to reconstruct; a density parcel extending vertically 100 km and 10 degrees horizontally spirally migrates northwestward through the moderate densities at mid-latitudes while diffusing out. Blue contours represent hmF2 variation through the region. The bottom panels are reconstructs from a tuned SC4DEnVar.

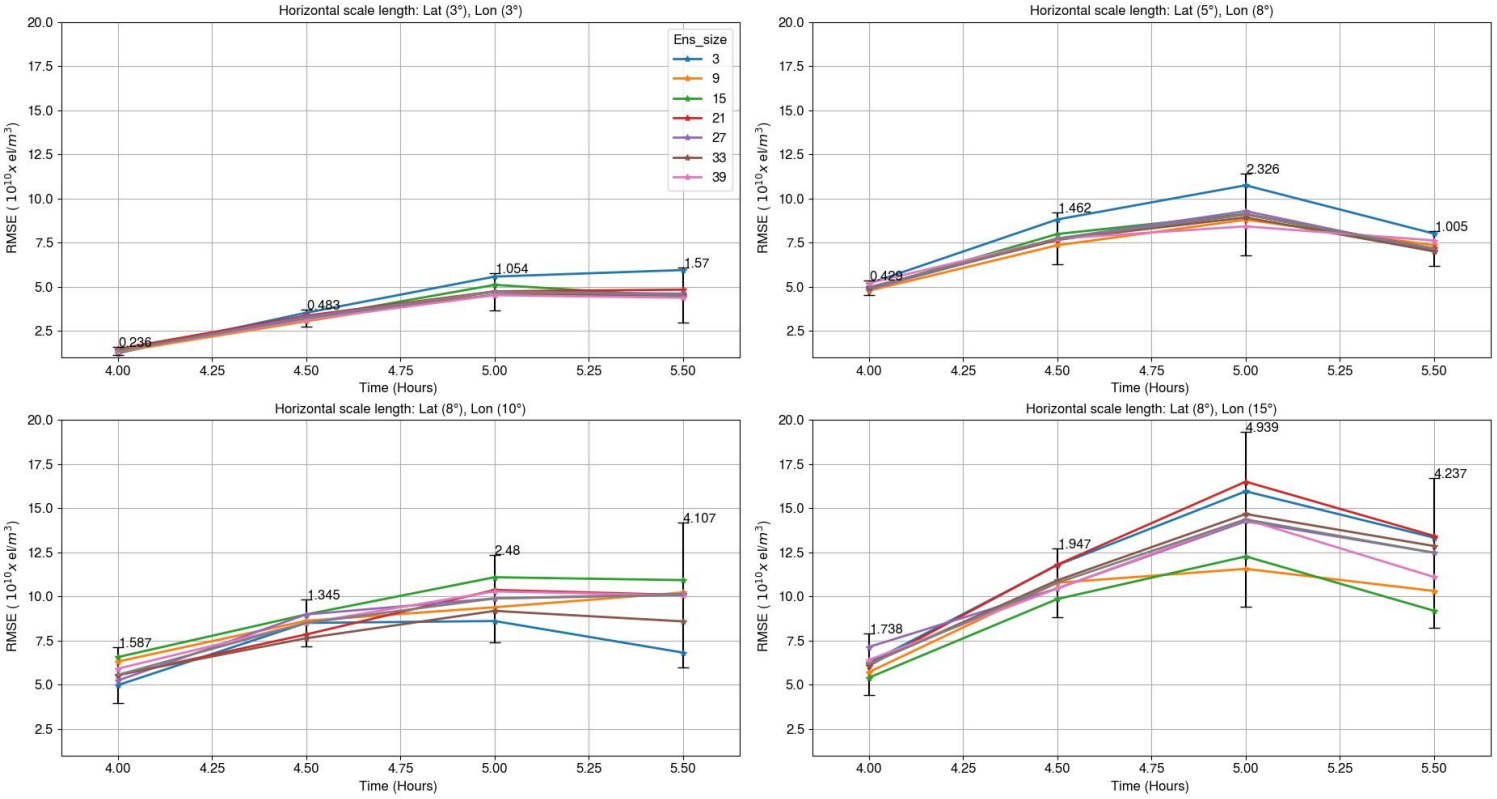


Figure 5: Ensemble tuning using simulation data to reconstruct the target structure in Figure 4 upper panels: each subplot shows a variation of RMSE as the assimilation time window is extended from 0 to 1.5 hours; different colors indicate the ensemble sizes varied under a fixed set of horizontal scale parameters.Vertical bars and appended numbers illustrate the measure of the spread in RMSE at that assimilation window length.

602 ***Simulation results:*** back to Figure 4; results in the lower panels are reconstructions of the parcel
603 while using the optimal settings from analyses; assimilation window length is 30 minutes,
604 horizontal length scales are 5 × 8 degrees, and ensemble size is 30. The background has been
605 subtracted for easy comparison with the target in the top panels. In each image, the intensity
606 colour bar is in the far upper left-hand corner and solid blue lines indicate contours of estimated
607 hmF2. We can ignore the random ghost structures attributed to reconstruction errors. At the base,
608 the assimilation captures the parcel shape but underestimates the maximum amplitude by ~ 45%:
609 at this stage, most parcel densities are concentrated at the centre within a tiny region, not
610 adequately traversed by the spotty GNSS rays (observations) used in the reconstruction. Because
611 the fitting is not exact, the algorithm compensates by spreading energy over a wider area than in
612 the target, hence the inconsistency in amplitude estimation. In addition, ratio of the background
613 to observation variance can influences quality or the amplitude of the analysis.
614 In the second row plots, when the parcel has evolved for 30 minutes and densities partially
615 diffused, the assimilation horizontally relatively tracks the changing density but falls short of
616 reproducing the sharp edges; the current localization is filtering or suppressing a subset of high
617 frequencies that could be representative of the sharp edges. Therefore, there is still room for
618 some future gross tuning, and the current settings are only utilizable when relatively tracking the
619 position of large-density structures but not defining the exact shape. In the north-eastward
620 direction, the reconstruction starts to suffer from ghost pockets of hmF2; this is a well-known
621 error in ionosphere ground-based GNSS data 3D reconstructions - mainly due to the luck
622 horizontal rays to constrain the vertical structure. In northern high latitudes where least GNSS
623 rays are expected, the scheme reverts to the background values, and indeed the hmF2 values are

well reproduced - pointing to the critical importance of utilizing an accurate background in analysis.

In the panels on the right, the parcel has evolved for 1.5 hours and aligned northwest-southeast. The sharp edges are still not well recovered, but the gradient in density and orientation of the structure are noticeable. The location of the maximum density is now displaced toward higher latitudes, and the hmF2 error is ubiquitous throughout the image. An indication that better results are in the early stages of analysis and marginally worse towards the later stages; the exacerbation might be the influence of the non-adaptive localization matrix (**L**) as the errors evolve. That is to say, at the initial stage, nearly all ensemble spurious correlations are eliminated since **L** is well suited. On the contrary, at the later stages, the spurious correlations might have shifted beyond the width of a static **L**; in such instances, deleterious effects of non-suppressed spurious correlations start emanating in the analysis.

3 Experimental run

To qualitatively analyse and assess the performance of the two schemes, SC4DVar-Inc and SC4DEnVar, we attempt to reconstruct the ionosphere 3-D structure during quiet and strongly disturbed periods and compare the results to independent observations from ionosondes and radio occultation. The experimental setup follows the grid setting and optimal parameters discussed in section 2. Although both schemes are generalized to accept different types of observations (linear and non-linear), we only assimilate ground-based GNSS observations. Of course, having different types of measurements that complement each other can ameliorate the analysis. However, the complexity of defining, understanding and balancing the errors to spread the correct information in space makes the process intricate. For instance, Forsythe et al. (2022) show that if horizontal gradients exist in the ionosphere, the ingestion of radio occultation STEC

can introduce a fictitious F3-layer in the analysis. Thus, sequentially assessing the impact of the different types of observations can highlight the areas to improve; here, ground GNSS data are the base starting point to improve as the scheme(s) mature. Specifically, only observables from the GPS (Global Positioning System) constellation are utilized, though most receivers are multichannel and can track other constellations (e.g., BDS, Galileo, and GLONASS). The GPS satellites transmit signals at two L-band frequencies (f1 = 1.57542 GHz; f2 = 1.22760 GHz). To derive the relative STEC utilized in our assimilation, we use the L1 and L2 phase measurements: in a continuous transmitter-receiver arc that spans a maximum length of 2.5 minutes, the associated biases and phase observation ambiguities are assumed invariant and removed by subtracting the arc's average STEC from each STEC value in that arc. This approach has advantages: 1) phase observations have high precision. 2) because each continuous arc is analyzed independently, we do not have to worry about fixing cycle slips.

Quality control: *a*ll GNSS data are subject to data quality control (QC) procedures before assimilation. However, the decision on which observations to retain or throw away is one of the challenging steps in ionospheric data assimilation; when the ionosphere is highly non-linear, it is difficult to decide whether observations considered as outliers are representative of an event (maybe instabilities) or errors as a result of inherent measurement processes. Therefore, the decisions made at the QC step can determine the success or failure of the scheme.

Here, first we, check the GNSS observable (L1 and L2) against the corresponding signal to noise ratio projected into the interval 1-9; if the flag is set to a value below 2, that ray is rejected. This is followed up with a simple "buddy check" procedure previously utilized in Ssessanga et al. (2019). An observation or datum is an outlier if the magnitude of the innovation, $|\Delta\vec{Y}_k - \Delta h_k(\vec{X}_k^b)|$, exceeds a threshold $C * (\sigma_k^o + \sigma_k^b)$, where C is a predefined multiple (set to 2;

this is a coverage of 95% of the total area under the distribution, for an inherent assumption that error statistics are Gaussian) and σ_k^o, σ_k^b are observation and model (in observation space) error variances, respectively, at time t_k (see, e.g., Dee et al., 2011). At this stage the outliers,are not entirely rejected, but the corresponding observational variance is inflated (5 times more than the surrounding observations).

Analysis period: the analysis period covers the 15 - 20th of March 2015, which includes the strongest geomagnetic storm in the solar cycle 24 known as the St. Patrick's Day geomagnetic storm in the ionospheric community. There is already a multitude of literature detailing and attempting to explain the dynamics and physics relating to the St. Patrick's Day geomagnetic storm (e.g., Astafyeva et al., 2015; Cherniak et al., 2015; Sripathi et al., 2015; Chen et al., 2016; Nava et al., 2016; Huang et al., 2016; Wu et al., 2016; Joshi et al., 2016; Patra et al., 2016; Venkatesh et al., 2017; Dmitriev et al., 2017; Amaechi et al., 2018; Yasyukevich et al., 2018; Tulasi et al., 2016;2019). Therefore, the focus here is on the fidelity of the schemes rather than an explanation for the observed dynamics.

3.1 Bottom side assessment with Ionosonde data

Ground Ionosondes measurements are prime on probing the ionosphere vertical bottom-side plasma density structure and are typically used for calibrating other complex methods such as tomography and data assimilation (e.g., Decker and McNamara, 2007; McNamara et al., 2011; Ssessanga et al., 2017). Solid red circles in Figure 1 indicate the ionosondes stations in the region of interest, but we only select four for analysis: Okinawa, Kokubunji, Wakannai and Beijing. Refer to Table 1 for the locations, five-character code names, geographic coordinates and geomagnetic latitude of each ionosonde used in the analysis. Hereafter, we refer to stations by

692 the five-character code names. OK426-TO536-WK546 (TO536-BP440) offer a sample of the
693 latitudinal (longitudinal) bottom-side ionosphere variations.

694 **Table 1.** A summary of ionosonde stations used when verifying variational scheme reconstructs.

Location	Five character Code	Geographic location (degree)	Mag Latitude (degree)
Okinawa	OK426	26.68 , 128.15	17.43
Wakannai	WK546	45.16 , 141.75	36.75
Kokubunji	TO536	35.71 , 139.49	27.18
Beijing	BP440	40.30 , 116.20	30.80

695

696 Comparison is with automatically scaled profile data since raw data or manually-scaled profiles
697 were not readily available. The data are accessible at ftp://ftp.ngdc.noaa.gov/ionosonde/data/. It
698 is cautionary, however, to interpret the results while putting into consideration the error bounds
699 associated with auto-scaled data; Bamford et al. (2008) and Stankov et al. (2012) have reported
700 on these error bounds with a 95% probability - foF2 (–0.75, +0.85 MHz), foF1(–0.25, +0.35
701 MHz), foE (–0.35, +0.40 MHz), h′F2 (–68, +67 km), h′F (–38, +32 km), and h′E (–26, +2 km).

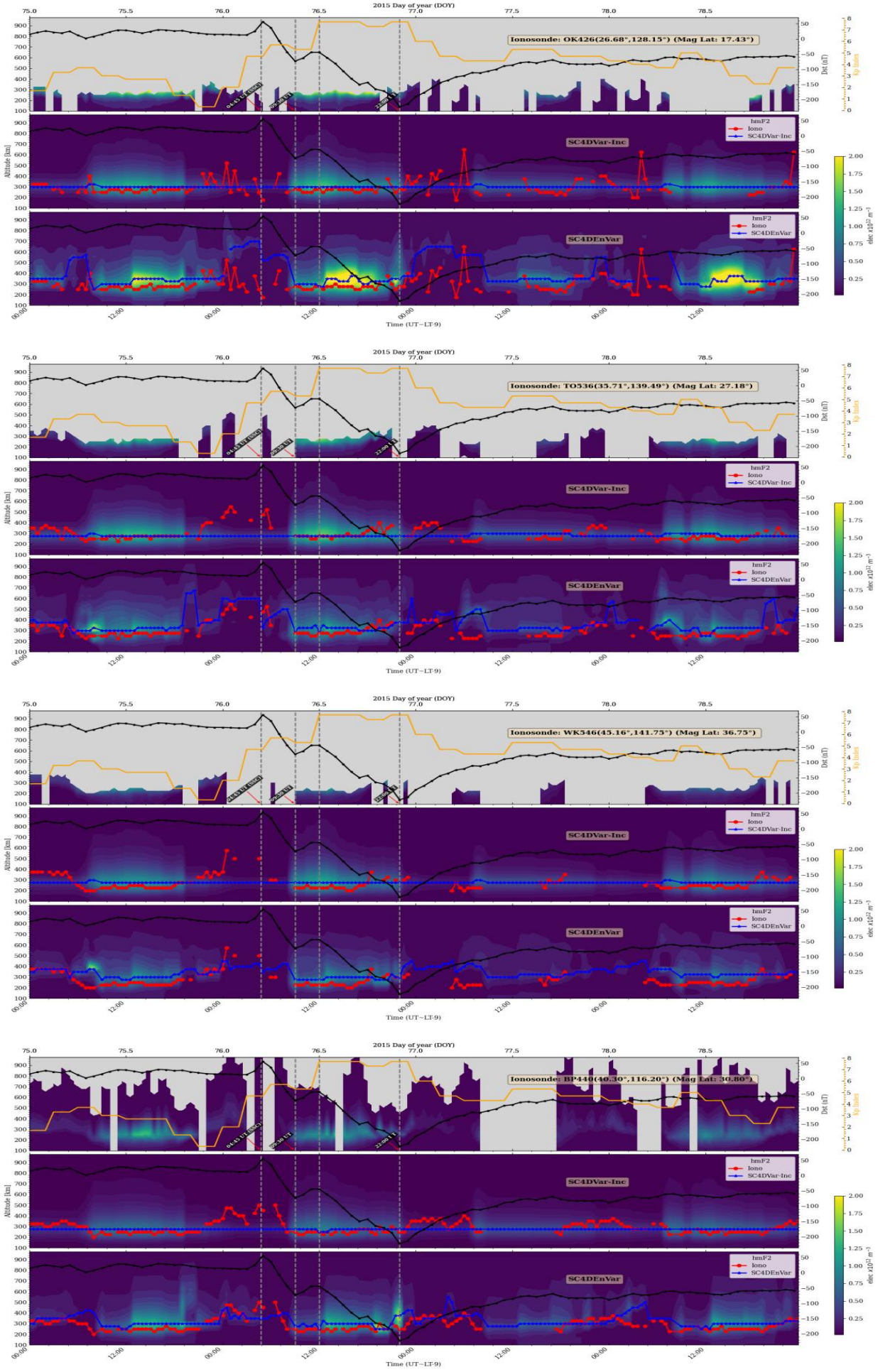

2015 Day of year (DOY)
Ionosonde: OK426(26.68°,128.15°) (Mag Lat: 17.43°)
Ionosonde: TO536(35.71°,139.49°) (Mag Lat: 27.18°)
Ionosonde: WK546(45.16°,141.75°) (Mag Lat: 36.75°)
Ionosonde: BP440(40.30°,116.20°) (Mag Lat: 30.80°)
Dst (nT)
kp Index
hmF2
Iono
SC4DVar-Inc
SC4DEnVar
Altitude [km]
elec x10^12 m^-3
Time (UT−LT·9)

Figure 6: Time series images of the vertical profiles at each ionosonde location in Table 1, covering a period of four days; before (DOY 75), during (DOY 76-77) and recovery (DOY 78) after the 2015 St. Patrick's Day geomagnetic storm. In each subplot, the top, middle and bottom are ionosonde observations, SC4DVar-Inc and SC4DEnVar reconstructs. For clarity, the images only extend to 1000 km. Each ionosonde panel includes the variation in Dst (asterisks black curve) and Kp (solid orange line) indices to track the ionosphere response due to energy injected into the ionosphere system during this period; the scales are on the right-hand side, and vertical dashed lines mark the main events of the storm. In SC4DVar-Inc and SC4DEnVar reconstructions, the red (blue) line is a time variation of ionosonde (scheme) hmF2 at that location. Gaps in ionosonde hmF2 correspond to periods when observations were non-available or failed to pass the quality control test. In the reconstructs, hmF2 values above 700km are considered void. The colour scale for all reconstructs is in the far right-hand corner.

Figure 6 presents time series images of the vertical density profiles recorded at the selected ionosonde station. The sample rate is 15 minutes. On the x-axis, tick labels at the top and bottom indicate time expressed in DOY (day of the year) and UT (universal time, the East Asian sector local time (LT) is ~ 9 h ahead of UT), respectively. In each ionosonde plot, observed bottomed-side recordings are in the top panel; on the left y-scale is altitude in km; on the right-hand is Kp (orange) and Dst (black) indices that show a measure of geomagnetic variations and strength of the ring current, respectively, in response to the energy injected into the terrestrial system (both indices are accessible at http://wdc.kugi.kyoto-u.ac.jp/ or https://omniweb.gsfc.nasa.gov/form/dx1.html). We assume a quiet terrestrial ionosphere when Kp < 4 and the magnitude of Dst < 50 nT. Vertical dashed grey lines indicate the main events of the storm. The middle and bottom panels in each subplot are reconstructions from SC4DVar-Inc and SC4DEnVar, respectively, at that ionosonde location: ionosonde observations (frequency in MHz) have been converted to electrons densities per cubic meters (elec/m^3) for easy comparison with reconstructions. The colour scale is in the far right bottom corner. Albeit our reconstructions extend to GPS orbital altitudes (22,000 km), for clarity, the altitude range is limited to the region of uttermost interest (~ 100-1000 km). Red dots and blue asterisks lines are variations of ionosonde and reconstructed hmF2. Ssessanga et al. (2021) generated analogous images for the

same storm period, but in that work, assimilation (using a 3DVar) is with ionosonde data into tomography reconstructions at only those voxels intersected by GNSS rays. The downside in their results is the lack of completeness due to empty patches in areas where measurements are lacking.

Before discussing the reconstructed results, a brief review of the chronological highlights of the storm is necessary: on the DOY 075, before the storm, the ionosphere was quiet, with Kp and Dst below 4 and 50 nT, respectively. Early, at 04:45 UT, on DOY 076, the Dst increases > 50 nT; this is the SSC (storm sudden commencement) following a coronal mass ejection (CME) that was observed erupting between ~ 00:30 UT and 00:40 UT the day before and predicted to encounter the Earth's magnetosphere on DOY 076. The main phase of the storm starts at ~ 07:30 UT, and Dst continuously decreases to a local minimum of ~ − 80 nT; previous analyses of the same storm have illustrated that the interplanetary magnetic field (IMF) Bz component was southward during this period (see for example, Cherniak et al., 2015; Ssessanga et al., 2021). Between ~ 9:30 and 12:20 UT, the Dst index has a short-lived recovery to ~ − 50nT; due to the IMF Bz component turning Northward (Astafyeva et al., 2015). Thereon, Dst (Kp index) decreases (increases) to a minimum (maximum) of approximately − 223 nT (8) at 22:00 UT. The recovery phase follows through the next few days, with the Kp index below 6.

Quiet period: in Figure 6, on DOY 075, both SC4DVar-Inc and SC4DEnVar reproduce the expected ionosphere variations during quiet conditions (Kp<4 and |Dst| < 50 nT): particularly, F-region densities (150 - 600 km in altitude) decrease from low to high latitudes; see stations OK426 - TO536 - WK546. Longitudinal-wise (TO536 -BP440), neglecting the slight difference in geomagnetic latitude (~ 3 degrees), high F-region densities are observed first at TO536 and ~ 2 hours later at BP440, mainly due to the solar zenith angle variations. At low- and mid-latitudes,

hmF2 time variation is stable around 350 km, while towards high latitudes values drop below the 300 km mark. A cursory glance at the F-region maximum density intensity suggests that the SC4DEnVar picture correlates better with ionosondes observations - also visible in the days that follow the quiet period. Interestingly, particularly at OK426, in SC4DEnVar results, is the post-sunset up-welling of the plasma sludge to approximately 500km due to the prereversal enhancement in E × B drift.

Disturbed period: in theory, at the onset of the storm, eastward prompt penetration electric fields (PPEF) were induced on the day side, with after-effects dominant at equatorial and low-latitude and dumped towards high latitudes (Maruyama et al., 2004). Indeed this is observed in ionosonde and SC4DEnVar reconstructions: in reference to the quiet period, the F-region plasma has risen more than 200 km at OK426 (low-latitudes) and a mere ~ 100 km at Wakannai (~high latitudes). Specifically at OK426, notice the erratic pattern in ionosonde hmF2 in situations when the peak uplifts to altitudes > 600 km and shortly after gradually falls to the average peak altitude of ~ 300 km. An insight from SC4DEnVar reconstructions shows that as the plasma rises, the intensity levels reduce. We estimated the vertical drift velocity (Vz) of the F-layer at this station as the rate of change in the peak height (dh/dt) from a base value of ~ 350 km. On DOY 75 (Vz ≈ 150km/1.5hr ≈ 27m/s), and on DOY 76 (Vz ≈ 300km/1hr ≈ 83 m/s). These values are comparable to Joshi et al. (2016) Doppler sounding measurements over the Indian sector, using a Canadian Advanced Digital Ionosonde (CADI). This agreement lends confidence to our analyses, even in the absence of ionosonde observations. Consequently, we hypothesize that the rise in altitude by the plasma sludge creates conducive conditions for spread-F leading to the erratic nature of ionosonde observations.

778 Contrary to SC4DEnVar tracking of observed hmF2, SC4DVar-Inc reconstructions continuously
779 reproduce hmF2 at levels of the quiet period (~300 km). This is the first noteworthy advantage
780 that SC4DEnVar has over SC4DVar-Inc, with the following justification: 1) the ground GNSS
781 measurements currently assimilated have little information on the vertical structuring of the
782 densities. 2) The **B** matrix in SC4DVar-Inc, which spans the analysis solution, is based on the IRI
783 model that is well known to underperform during geomagnetic storm conditions, even when
784 optimized to storm condition settings. Conversely, in SC4DEnVar, the ensemble that spans the
785 analysis solution, though generated from IRI, the perturbations include variations in hmF2, from
786 which the scheme estimates the optimal value.
787 On the day after the storm (DOY 077), at all stations, both schemes show that the plasma density
788 (frequency) reduced ~ 60% below the values observed during the quiet period. Again,
789 reconstructions show that the F-region plasma frequency was ranging below 4 MHz, hence the
790 total or partial absence of data at all ionosonde stations: the reduction in plasma intensity is
791 typical of ionosphere response after a major geomagnetic storm- courtesy of the composition of a
792 composition bulge from high altitudes, that can persist for several hours (e.g., Fuller-Rowell et
793 al., 1994; Prölss 1995; Tsagouri et al., 2000; Ssessanga et al., 2021a). Indeed, analyses of the
794 thermospheric column integrated O/N2 ratio changes, measured by the GUVI (global ultraviolet
795 imager) instrument onboard the TIMED satellite, found that the Asian sector had significant
796 composition changes that could have led to the negative ionosphere response (Astafyeva et al.,
797 (2015) and Nava et al., (2016)).
798 During the recovery phase, on DOY 078, specifically at low latitudes (OK426), SC4DVar-Inc
799 reconstructs densities as a moderate recovery from the previous day, while SC4DEnVar shows
800 reinforced ionization. The paucity of ionosonde observations during this period, does not allow

for a concrete verification of such intensification. However, Ssessanga et al. (2021a) found the same while using different techniques and observation types. Klimenko et al. (2018) analyzed the same storm using a first principle physics-based model and TEC measurements; they attributed these after-storm daytime electron density increases to neutral composition changes; the n(O) increases and n (N2) depletes leading to a significant enhancement of n(O)/n (N2) ratio at low and equatorial latitudes. Note that in Klimenko et al. (2018), these positive effects are discussed as occurring on DOY 081-082, but our reconstructions show the enhancements starting earlier, on DOY 078. Because OK426 is close to the minimum bounding region (MBR), not easily populated with GNSS measurements, and the reconstructions from SC4DEnVar seem to match earlier independent reconstructions and theoretical analysis, it is therefore not premature to suggest that SC4DEnVar performs better than SC4DVar-Inc, in remote region lacking observations: we revisit this argument below when looking at individual vertical profiles.

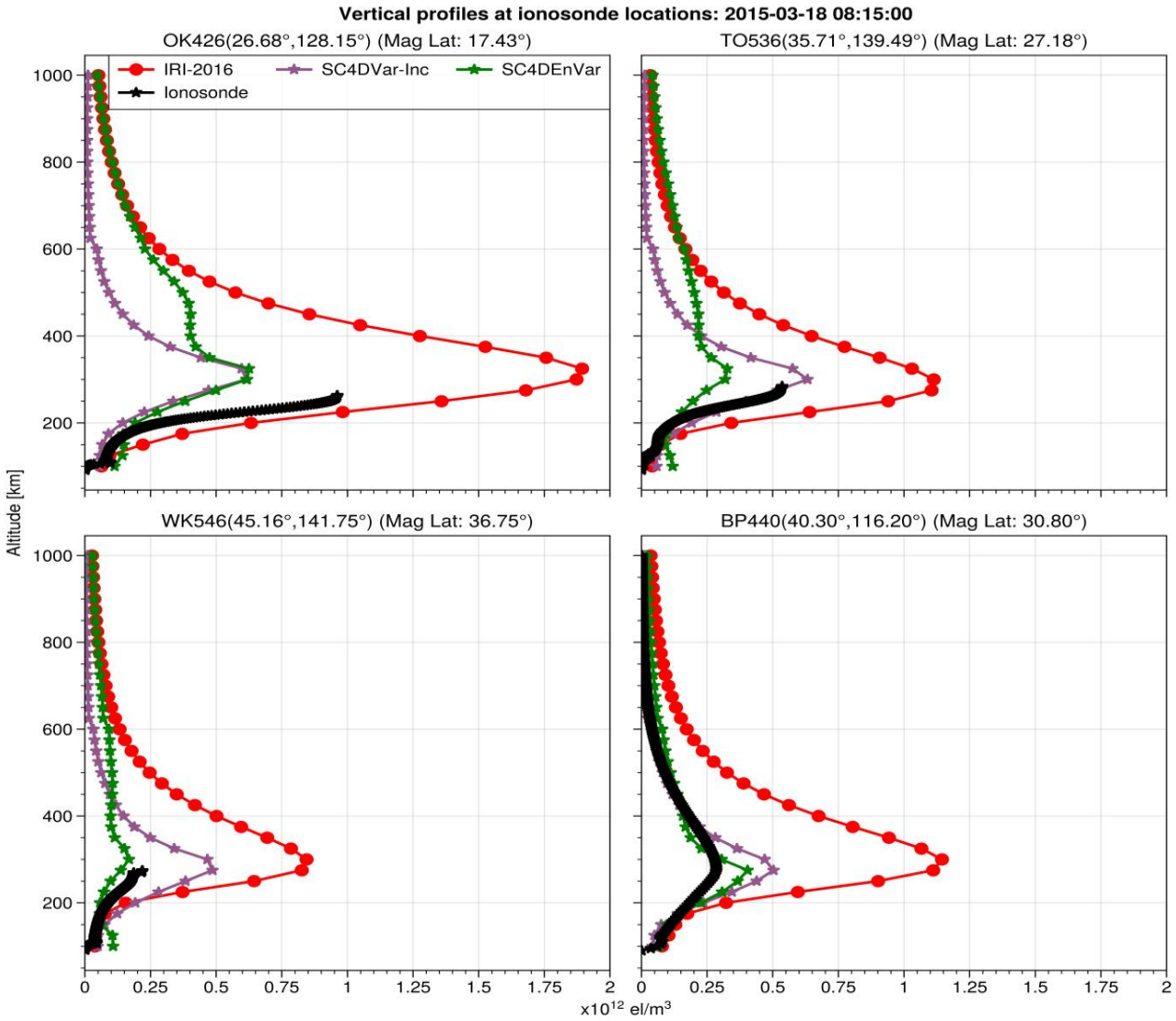


Figure 7: Example of vertical profiles extracted at the four ionosonde stations in Table 1; the timestamp is within the period of the immense depletion in mid-latitude electron densities after the 2015 St. Patrick's Day geomagnetic storm. IRI-2016 has a good fit for the mid-latitude average electron density profile and is included here as a reference for the extent of the depletion - at all stations, a near 60% reduction in plasma densities is apparent.

Individual vertical profiles: in Figure 7, we have extracted vertical profiles at the ionosonde locations. The timestamp is within the depletion period when the ionosphere is most perturbed, and observations available for all ionosondes. Black, red, purple and green represent ionosonde observations, background from IRI-2016, SC4DVar-Inc and SC4DEnVar, respectively. IRI is the

824 reference because of its good approximation of the average profile - both schemes portray the
825 earlier mentioned ~ 60% reduction in densities.
826 At low latitudes (OK426), both schemes nearly identically track the relative deviation of the
827 bottom-side F-layer from the average profile but underestimate the bottom-side thickness and
828 NmF2 densities. Note that during this period, the plasma peak density at this location had
829 returned to the average altitude, well estimated by the background model central in designing the
830 climatological **B** matrix employed in SC4DVar-Inc; suggesting that under the current
831 experimental settings, to match SC4DVar-Inc and SC4DEnVar performance in data void areas,
832 our prior knowledge (in SC4DVar-Inc) of the location of the peak density has to be nearly exact.
833 We argue that the inconsistencies in estimating the density by both schemes may be a combined
834 outcome of a paucity of observations at these latitudes and improper vertical correlation lengths.
835 Besides, during major geomagnetic storms, the horizontal extremes of the different geomagnetic
836 ionospheric regions are highly altered; the low-equatorial and high-latitude aurora boundaries
837 extend into mid-latitudes, thus introducing horizontal regional transitional zones not catered for
838 in the current covariance matrices - in future, there is a possibility of using the new hourly or
839 half-hourly planetary (Hp60/Hp30) index to estimate these boundary extensions; this index is
840 similar to the Kp index in quantifying magnetic activity but is better suited for near-real-time
841 applications because of the higher time resolution and no upper limit on the quantifiable level of
842 geomagnetic activity (Yamazaki et al., 2022).
843 In the E-region, both schemes tend towards the background (IRI) that agrees with observed
844 values. We expect this behaviour because, 1) as mentioned earlier, this area is generally data void,
845 and if any, due to the elevation angle limitations(>=20 degrees), compared to the F-region, the E-
846 region is generally porously crossed by ground receiver - GNSS ray links. 2) the vertical

847 localization set in the E-region and F-region are small to the extent that observations in the F-
848 region almost have no effect on the E-region (refer to localized vertical modes in the bottom
849 panel of Figure 1). 3) the contribution of the E-region to GNSS STEC is minute (also see Saito et
850 al., 2017) and can easily fall within the error range. In such situations, the algorithm dictates that
851 the schemes revert to the background.
852 Because the bottom-side layer is crucial in HF-trans-ionospheric applications and is known to
853 contribute ~40% to TEC, it is essential to integrate another sensor in the network at these
854 locations (for example the OK426 ionosonde). Ssessanga et al., 2021a; 2021b has already shown
855 that the assimilation of ionosonde data significantly (>60%) improves the bottom side structure.
856 Despite these inconsistencies, both schemes are better candidates at these locations than a stand-
857 alone IRI model.
858 At mid-latitudes (TO536), the region with the most data points, SC4DVar-Inc estimates the
859 bottom-side structure better. SC4DEnVar underestimates the NmF2/F-region and overestimates
860 the E-region. An obvious corollary is at WK546 (high latitudes) and BP440 (mid-latitude): at
861 these locations, observations are scanty, yet SC4DEnVar has a superior estimation of the bottom-
862 side structure. SC4DVar-Inc overestimates the F-region densities at both locations. This
863 underscores our earlier argument that SC4DEnVar is well-suited for regions with spotty
864 measurements.
865 A possible argument for the SC4DEnVar under-performing in areas populated with observations
866 is that we might have overestimated (underestimated) the background (observational)
867 covariances. Nonetheless, these results confer a vital future aspect; for an overall improvement in
868 accuracy, our scheme designs might need to evolve to hybrid covariance matrices to capture the
869 advantages of SC4DVar-Inc and SC4DEnVar in one picture. Above the NmF2 point, the two

schemes suggest a different picture, with SC4DVar-Inc portraying a thinner layer; ground ionosondes do not sound beyond the NmF2 point, so no conclusive comment is made about the two pictures at this stage. However, since earlier analyses have already shown SC4DEn-Var as a better scheme at locations without observations, we are persuaded to choose SC4DEn-Var as the better picture.

3.2 Radio occultation (RO) data assessment

In this set of verification, the comparison is with COSMIC (constellation observing system for meteorology, climate, and ionosphere; orbital altitude ~800 km) radio occultation (RO) density profiles and STEC. RO data are accessible in level 2 format at https://cdaac-www.cosmic.ucar.edu/cdaac/. Again, the analysis is during quiet and disturbed periods.

RO density profiles are derived from RO STEC using the Abel inverse technique that assumes spherical symmetry in the ionosphere. The profiles have good accuracy, except in the E-region and low latitudes: in the E-region, rays have asymmetric contributions from the F-region portions of the rays and in the low latitudes large density gradients exist (Garcia-Fernandez et al., 2003; Wu et al., 2009; Yue et al., 2010). Consequently, this RO density profiles comparison is limited to reconstructions away from the low latitudes and above the 200 km altitude mark.

In Figure 8, the area of the solid circles is proportional to the square of residuals between COSMIC (considered as “truth”) and reconstructed densities, normalized to the “truth”. The comparison is along the trace of RO tangent (ROT) points, which contribute the most density to the RO STECs used in the Abel inversion. The colour shade of the circles represents different altitude levels from 100-800km; see the colour bar in the upper corner of the left subplot. Therefore, the size of the circles corresponds to the size of the error for that scheme at that height

level and location. The left (right) subplot represents SC4DVar-Inc (SC4DEnVar). For reference purposes, we have added white circles (centered on ROT points) indicating a 3% error value.

Quiet period (DOY 074): this day is not shown in time series images (in Figure 6), but |Dst| and Kp were less than 50nT and 5, respectively; selected for analysis here because that is when enough RO profile data that passed our quality control check were available. On average, circles in the left subplot have small radius; thus, SC4DVar-Inc has a better performance than SC4DEnVar, with a worse performance noticed in the bottom altitudes (~ < 300km) particularly in South East.

In the right subplot, SC4DEnVar has a degradation in performance, with the errors worse mainly in the topside (500~800 km) and over the region where ground receiver GNSS network observations are dense (see Figure 1). To analyse and assess why this degradation occurs, we need a sense of whether the estimate is above or below the "background"; we use an extract of a sample density profile along the southeast ROT traces indicated in the zoomed-in windows - black asterisks indicate the original high resolution RO profile density data from COSMIC. Cyan squares are average values from the original profile that match the current grid. Red, purple and green represent the background, SC4DVar-Inc and SC4DEnVar, respectively: SC4DEnVar (SC4DVar-Inc) densities (~ 520-680 km) are larger (smaller) than the background values (IRI-2016). Thus, the analysis increments are positive (negative). Because the background is overestimating the “truth”, we are of the assumption that the innovation vector is generally negative, which implies that for the increments in the two schemes to have different signs, the error must arise from the vertical covariance terms that can take on either a positive or negative sign. In addition, in areas where observations are lacking, the analysis increments are proportional to the magnitude of the covariance term (between the grid point with and those

without the measurement). Since the vertical correlation lengths increase with height, and the error in SC4DEnVar follows the same trend, high-altitude vertical correlation lengths are probably the likely candidates to readjust. Also, because **L** is SC4DEnVar persistent (kept the same; refer to equation 24) through the assimilation window without including the error auto-correlations, this can introduce inappropriate cancellation of some analysis increments down the assimilation window, hence introducing errors in final analyses.

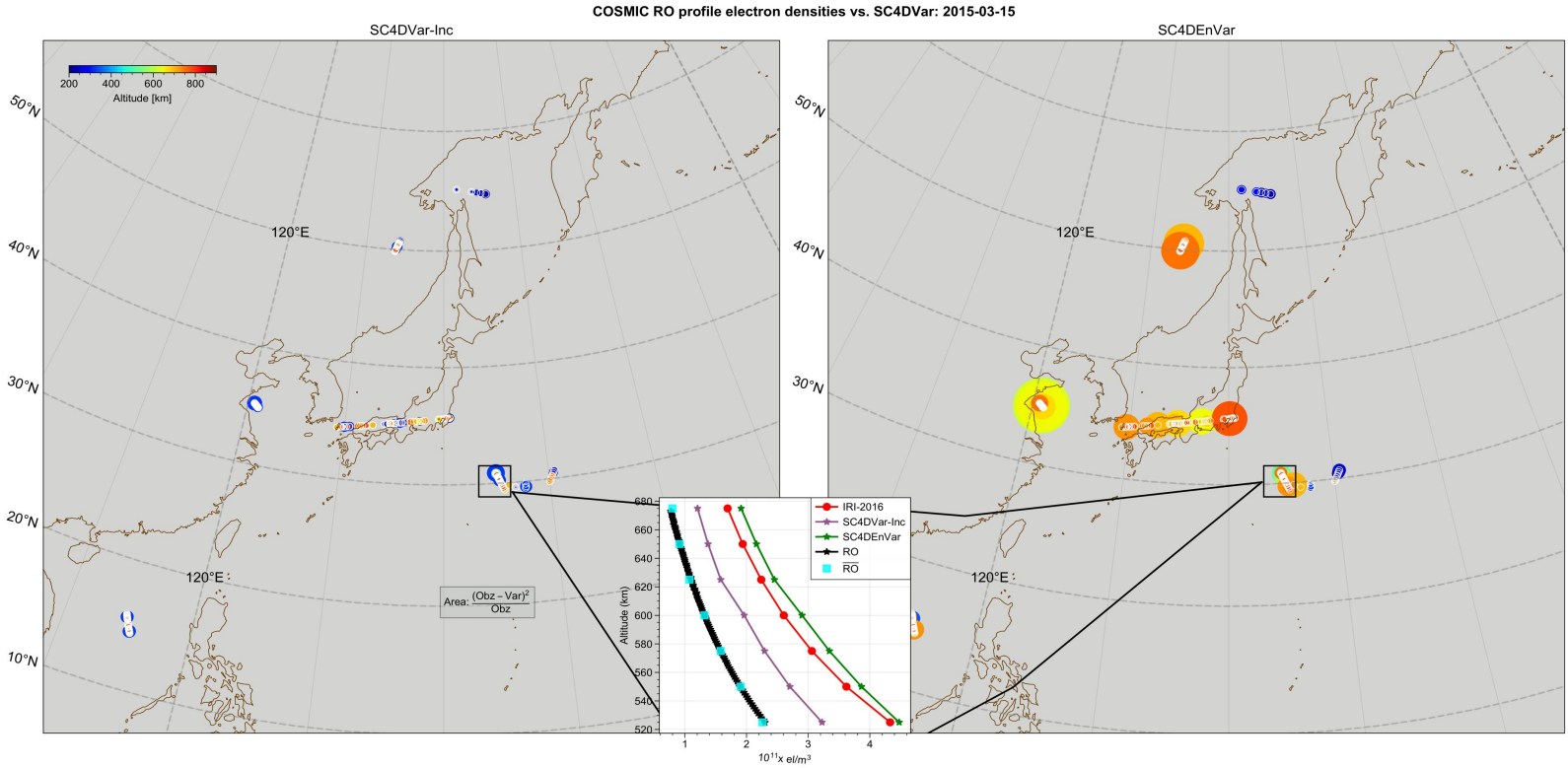
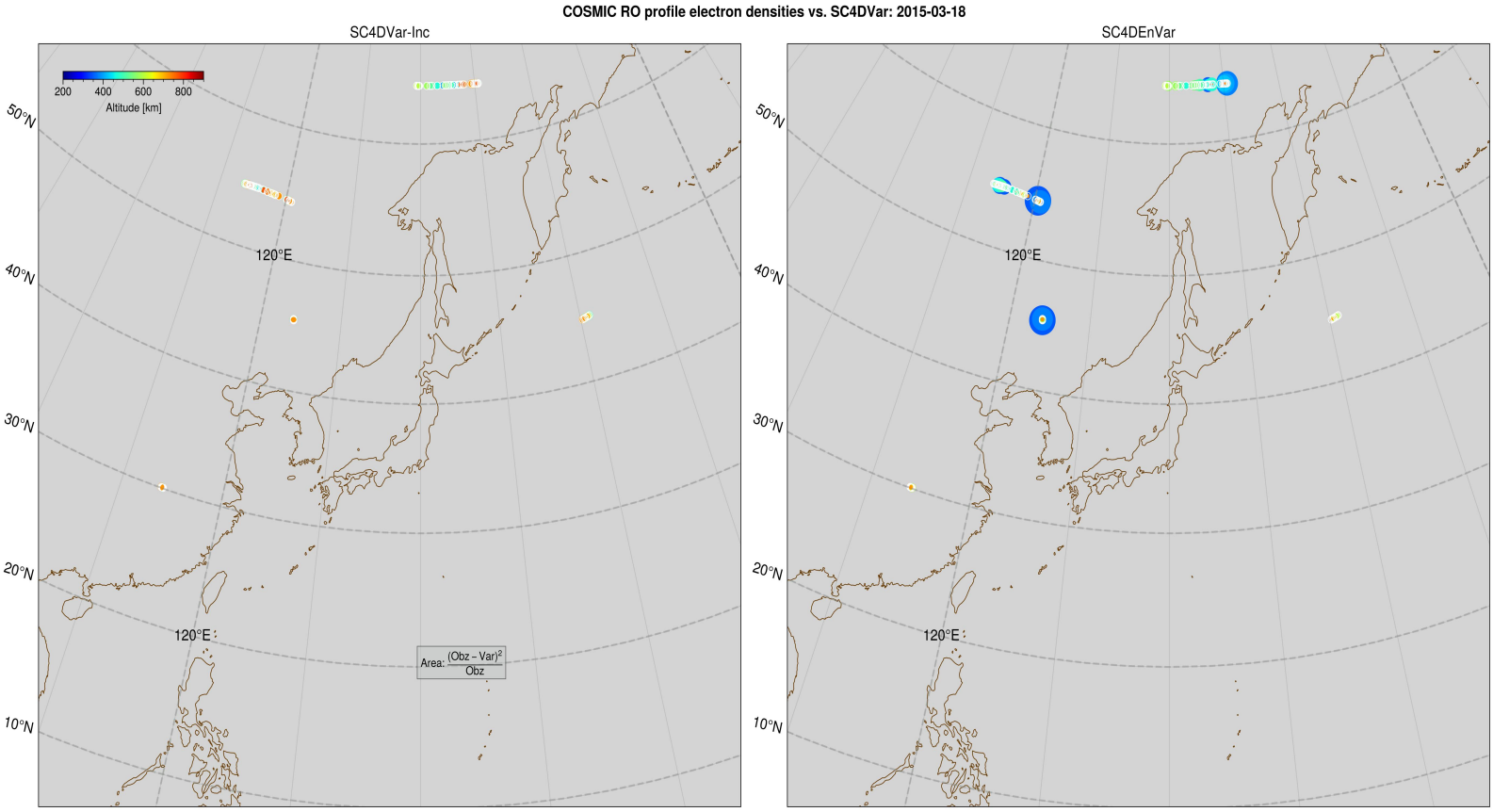


Figure 8: Quiet period error analysis, DOY 75. A chi-square fit of reconstructs to COSMIC radio occultation (RO) electron densities from the Abel inversion scheme. Comparison is along the RO tangent point geographic locations. The colour indicates the altitude level; the scale is in the far upper left corner. The area of the solid circles is the extent of the error at that horizontal location and altitude level. For reference, white circles indicate a 3% error level. Left (right) is SC4DVar-Inc (SC4DEnVar): for further analysis, the in-set axes zoom into a set of profile points; black asterisks are high-resolution electron densities from Abel inversion, and cyan points are average points along the profile used in the analysis. Red is the background (IRI-2016), and purple (green) represent SC4DVar-Inc (SC4DEnVar) reconstructed densities.



930 Figure 9: Same as Figure 8 but for the ionosphere disturbed period, DOY 077, 2015
931

932 ***Disturbed period:*** Figure 9 is the same as Figure 8 but for the storm period. The ROT traces are
933 biased to the North West away from the region of dense data points. SC4DVar-Inc still performs
934 better, with the error size nearly equal at the different altitudes. On the other hand, SC4DEnVar
935 has good performance on the topside, but now the errors are pronounced in the height range of
936 200~ 400km. In the North East, the few ROT traces show SC4DEnVar slightly performing better
937 than SC4DVar-Inc.

938 ***ROSTEC:*** remember that in Figures 6-9, assessments are at a particular grid point within the
939 radius of influence of a given single observation point (either along ionosonde or occultation
940 density profile), which limits the accuracy evaluation of the 3-D picture as whole. Also, from the
941 simulation analysis, we saw that the SC4DEnVar scheme reconstructs approximate the correct

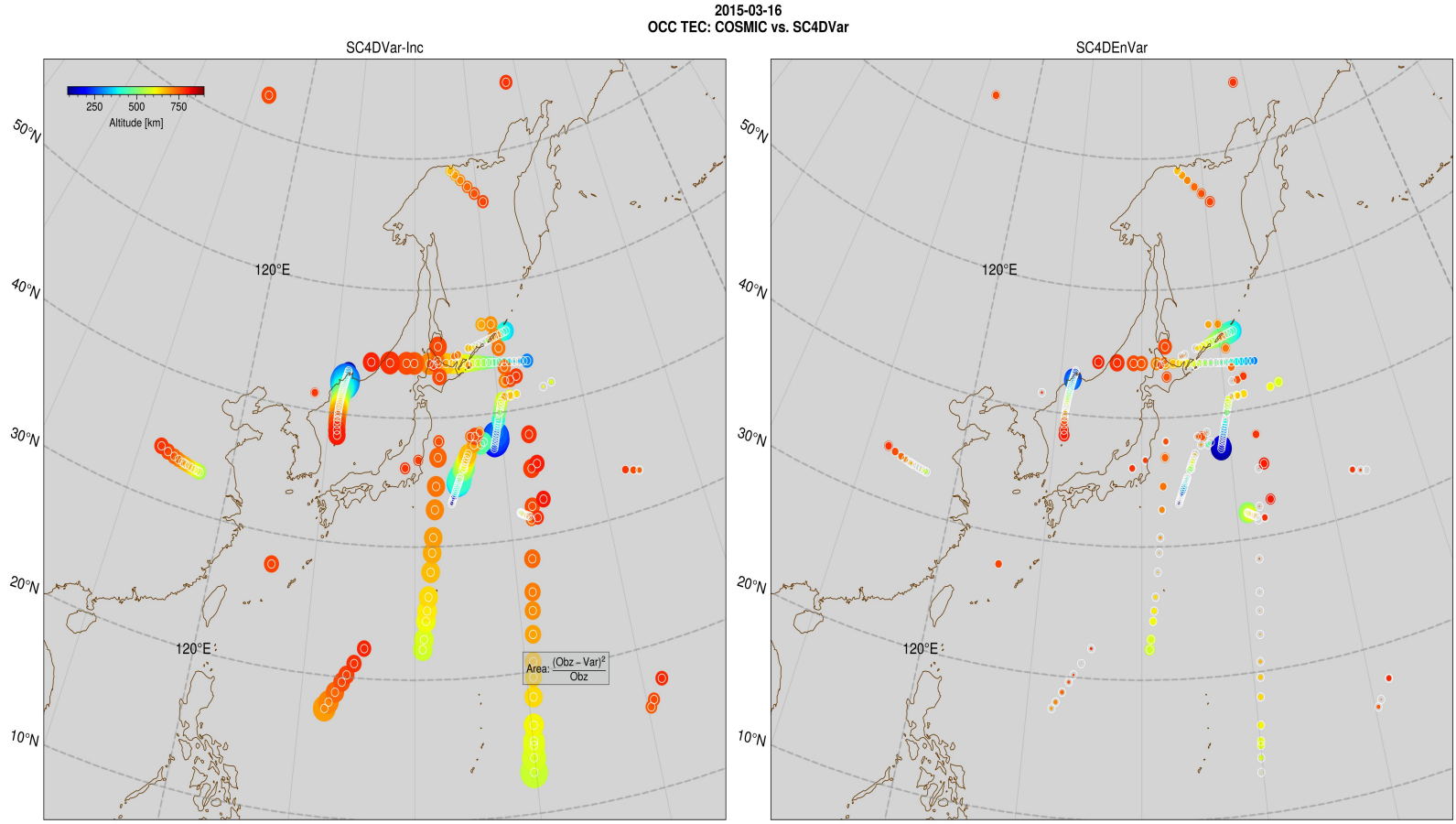
densities but overspreads or displaces the density structures from the actual position. Therefore, a point-wise verification might not be an appropriate conclusive score assessment. In what follows, rather than specific grid point evaluations, the comparison is with ROSTEC- an integration density along the different GNSS - LEO lines of sight (LOS) through the 3D picture, with different projection angles offering snip views into the 3D volume.

In Figure 10 and 11, the calculation of the errors is the same as in Figures 8 and 9, but for ROSTEC. Figure 10(11) is the during the quiet (disturbed) period. Only ROSTEC data with ROT points within the 3D volume are analyzed. For some LOS, the transmitters or receivers are outside the MBR, but as mentioned earlier, the ROT point contributes the most density to ROSTEC, such that the extra contribution from outside the MBR can be considered negligible.

In Figures 10 and 11, the SC4DEnVar error sizes are smaller, which is an improvement in the SC4DEnVar analyses relative to the SC4DVar-Inc results for quiet and disturbed periods. In both cases, the impact of SC4DEnVar was overall > 70% better than SC4DVar-Inc. This is a drastic change in the SC4DEnVar performance score, compared to RO density profile analyses: an indication that despite inconsistencies at specific locations, overall, the ensemble set has a better representation of the ionosphere 4D density distribution, with a significantly improved approach to the representation and evolution of the error covariances.

In Figure 10 (during quiet period), the impact to the far North high latitudes (>50^{o}N) is neutral between the two schemes. In the mid latitudes, there is tendency for the impact to get better with ROT traces towards Japanese archipelago; the area with the most ground GNSS data points. For a particular scheme, at mid-latitudes, at high altitudes (> 600 km), the impact is quasi-equal; this is because, at these altitudes, during quiet periods, the plasma smoothly varies without prevalent

964 sharp dynamical changes, and the correlations are high with vertical localization nearly non-
965 existent. Subsequently, the assimilation impact is similar over a wide range.
966 The larger error values are attained below ~ 500 km; this region engulfs the F-region, where
967 large density gradients are ubiquitous. Therefore the under-performance of SC4DVar-Inc is likely
968 due to the climatological background-error covariance that does not fully span the dynamics.
969 Indeed, though SC4DEnVar is generally better than SC4DVar-Inc at these altitudes, the error is
970 persistently sizeable compared to other altitude levels; thus, the source of error should be the
971 shared climatological background and prognostic model between the two schemes, but probably
972 with SC4DEnVar having a better handling of linearization (TLMs and MAs) errors than
973 SC4DVar-Inc.



974 Figure 10: Same as Figure 8 but for RO STEC (quiet period DOY 75) - SC4DEnVar and
975 SC4DVar-Inc STEC are an integration of electron densities along the GNSS - LEO (Low Earth
976 Orbit) satellite link through the reconstructed 3D picture. The highest contribution to the STECs
977 comes from the tangent point locations. The superiority of SC4DEnVar is distinct throughout the
978 whole 3D region.

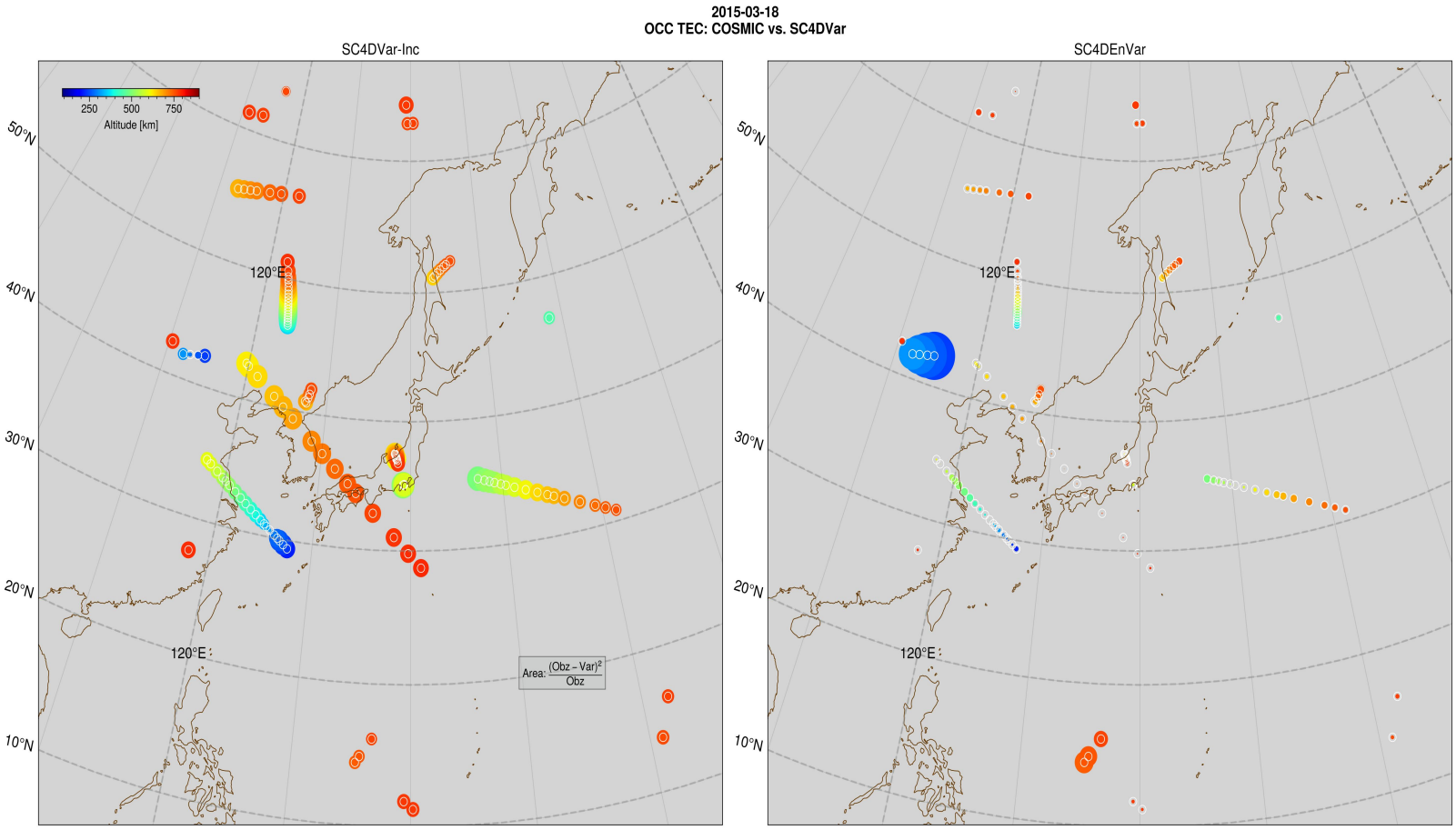


Figure 11: Same as Figure 10 but for the disturbed period (DOY 077, 2015).

On the storm day, the impact of SC4DVar-Inc is comparable to the quiet period (the error size is nearly the same), while SC4DEnVar shows a drastic positive change in performance (smaller error size). We are of the view that the superior performance of SC4DEnVar is a consequence of the scheme's sensitiveness to altitude variations as evidenced by the analyses of vertical bottom-side structure using ionosondes (absent in SC4DVar-Inc); during storm conditions, spatial-temporal density altitude variations are ubiquitous and can lead to sharp density gradients resulting in sizeable ensemble covariance with an impact as noticed in the difference between SC4DEnVar quiet and storm results.

Further, on the storm day, the impact of SC4DVar-Inc at altitudes >500km is nearly consistently the same throughout the region. Again the worst performance is at altitudes below 500 km. At

high latitudes (>50°N) SC4DEnVar impact is better than SC4DVar-Inc. The impact even gets better towards the mid-latitudes, with the exception of the lower altitudes (< 300 km) on westerly side. For some traces, the fit to observations is nearly exact, that error representations are smaller than white circles, as seen in the longest trace in SC4DVar-Inc results - aligned from North East to South West. An explanation may be that the forecasting of ensemble members increases the background errors, and in return, the impact of observations increases over the radius of influence, hence strongly improving the analysis.

Toward the low latitudes (<20°N), at about ~ 600 km in altitude, SC4DEnVar impact is strongest on the far South East; this remote area is dominated by the oceans without any sensors and close to the MBR (the schemes rejects any rays crossing the MBR). The error in such a region is a more cardinal measure of the effectiveness of a scheme than the fit at locations where observations are readily available (for example, the quasi-central continent area). Therefore, it is likely that flow dependence in the ensemble setup introduces adequate weather throughout the region.

However, there are inconsistencies within the SC4DEnVar picture, probably due to the ensemble sample space not fully capturing some nonlinear geomagnetic storm ionosphere responses. Moreover, the assumption that observational error statistics, $\mathbf{R}_k$, are well-known and defined is ideal - yet in regions without observations, the analysis can be sensitive to infiltration of observation noise if $\mathbf{R}_k$ and $\mathbf{B}_{en}$ are not well balanced or weighted (Johnson et al., 2006). Furthermore, in the simulation analysis in section 3, even under a short time assimilation window, our simple Gauss Markov prognostic model was found to introduce errors which distort the analysis. A further improvement would be to trade-off some of the computational cost with

improved accuracy and use a simple physics-based model that allows 3D spatial density evolution.

4 Discussion and conclusions

We have conducted a performance analysis of two in-house developed flavors of strong constraint 4DVar assimilation schemes; SC4DVar-Inc and SC4DEnVar.

We ingested only ground-based GNSS observations that are ubiquitous and generally readily available, serving as a base starting point for future improvements. The aim was to assess whether the SC4DEnVar is sufficiently accurate in reconstructing the 4D regional ionosphere during quiet and geomagnetic conditions, compared to the popular SC4DVar-Inc approach. Table 2 includes a schematic summary of the comparison between SC4DVar-Inc and SC4DEnVar, with check marks indicating a better scheme.

Table 2. Summary of SC4Var-Inc Vs. SC4DEnVar

Comparison Factor	SC4DVar-Inc	SC4DEnVar
Height variation	☒	☑
Electron density estimation in:		
Areas populated with data	☑	☒
Remote areas	☒	☑
Severe conditions (geomagnetic storm)	☒	☑
The whole 3D structure	☒	☑
Computation expense	☒	☑
Maintenance	☒	☑
Parallelization	☒	☑
Under-sampling and Rank deficiency	☑	☑
Sensitivities	☑	☒
* This makes SC4DEnVar a favourable candidate for real-time applications and remote areas		

Through verification with ionosonde data, we first noticed that the SC4DEnVar scheme is surprisingly good at tracking the hmF2 variations, even when the assimilated data contain little information about the ionosphere vertical structuring. This is a critical result since most

ionosphere inverse problems that rely solely on ground-based GNSS observations fall short of this realization; the NmF2 values are well estimated but with an incorrect hmF2. Indeed, there are considerations to use the SC4DEnVar analyses as prior information in tomography algorithms intended to add regional fine-scale structures - the accuracy of the background, mainly the peak location, highly influences the fidelity and convergence rate of most of such tomography algorithms (e.g., Ssessanga et al., 2015; 2017).

During the quiet period, SC4DVar-Inc and SC4DEnVar performances in reconstructing the vertical structure are nearly identical, but there is a caveat for SC4DVar-Inc that the background state should have a good approximation of hmF2. Since there was no attempt to fine tune the vertical correlations, it is probable that better analyses are attainable with the specification of more localized vertical correlations.

In locations where observations are prevalent, specifically at high altitudes in mid-latitudes, SC4DVar-Inc shows a larger impact than SC4DEnVar. This result could be due to a combination of effects. First, the utilized vertical correlations are not optimal (not tuned) for the SC4DEnVar covariance. Second, the 3D covariances in both schemes might suitably be localized at the start of the assimilation window but evolve differently over time; that is, in SC4DVar-Inc, as long as the linear model ($\widetilde{\mathbf{M}}_{\mathrm{k}}$) is correct, the time correlations are correct with implicit ($\widetilde{\mathbf{M}}_{\mathrm{k}}\mathbf{B} \circ \mathbf{C}\, \widetilde{\mathbf{M}}_{\mathrm{k}}$) propagating of 3D variances ($\mathbf{B}$) and correlations ($\mathbf{C}$). On the other hand, in SC4DEnVar, the same localization $\mathbf{L}$ is persistently used throughout the window even after the ensemble members evolve, $\mathbf{B}_{\mathrm{en}}^{\mathrm{k}} = \mathbf{L} \circ \widehat{\mathbf{X}}_{\mathrm{k}}^{\mathrm{b}}\, \widehat{\mathbf{X}}_{\mathrm{k}}^{\mathrm{b}}$. This renders the time correlations incorrect, and in areas with dense observations, these discrepancies might as well have a deleterious effect on the analysis increments.

1052 During severe conditions and in remote areas, SC4DEnVar is superior to SC4DVar-Inc, reducing
1053 errors by nearly 70%. Because both schemes assume a strong-constraint formulation under the
1054 same type of prognostic model, and the corrected background ($\vec{X}_0^b$) is the same, the difference in
1055 performance is probably a consequence of SC4DEnVar's better representation of the error
1056 covariances and non-linearities (TLMs and MAs). Specifically, the flow-dependency when
1057 evolving error covariances introduces adequate information through the 3D electron density field.
1058 Computational-wise, SC4DEnVar abates the major SC4DVar-Inc bottlenecks, which majorly
1059 motivated this investigation: there is no need to compute and maintain the cumbersome and high-
1060 expense TLMs and MAs. The major inhibition in implementing SC4DEnVar is the cost of
1061 forecasting each ensemble member for the length of the assimilation window (see equation 22),
1062 but with the current computation resources, this task is efficiently parallelizable. With the spared
1063 computation resources (time), a resolve of medium scales structures using a finer grid is possible
1064 and reconstructs can be extended to a much larger regional spatial space, which engulfs more
1065 observations to assimilate. Since the computation expense scales with the grid size, ensemble
1066 size and rank of the localizing matrix, any changes to the grid warrant a further diagnostic to
1067 understand whether the current 30-member ensemble is optimal to capture the desired dynamics.
1068 Additionally, as the ionosphere becomes more an-isotropic with region expansion, the ensemble
1069 members may need to be weighted differently in different local regions. An option is to divide
1070 the large area into sub-local regions and parallelly apply the current SC4DEnVar settings, but
1071 handling local boundary discontinuities when merging analysis pictures may pose challenges.
1072 In conclusion, it is not clear-cut that the SC4DEnVar is always better than SC4DVar-Inc, but it
1073 generally provides better ionosphere analyses under extreme stresses and in remote area. These

advantages recognized in the SC4DEnVar system warrant an accelerated upgrade from the orthodox SC4DVar-Inc scheme.

Future recommendations: though not mentioned earlier in the text, the aim in maturing these schemes is to transfer the technical know-how and understanding of their limitations from moderately quiet mid-latitudes (pilot project) to high latitudes where ionospheric non-linearities are prevalent. Therefore, the following limitations need redress before extending the SC4DEnVar scheme to high latitudes:

- Simplified strong-constraint prognostic model: we have noticed that our simplified model is easily error-infiltrated over short-time assimilation windows, which limits our ability to exploit the full advantages of the 4DVar scheme; the assimilation window should be long enough to allow ample evolution of the background covariances. Additionally, 4DVar performs better with observations widely distributed across the long assimilation window.
- Model error correction: with a longer assimilation window, the scheme must account for model errors ($\vec{\omega}_k$ in equation 1); therefore, in the advanced stage, a weak 4DEnVar formulation could be a suitable candidate. Desroziers et al. (2014) have shown that 4DEnVar formulation has an appropriate way to account for different representations of model errors - albeit in ionospheric applications, a description of the $\vec{\omega}_k$ error covariances is still challenging, particularly for high-resolution models that scale to millions of elements with insufficient data for validation.
- Spatial-temporal localization or correlation matrices: of course, under an extensive assimilation window, the assumption of a persistent **L** can no longer hold, and the need to propagate or compute different spatial-temporal localization or correlation matrices

for time samples "t_k" is necessary; this is crucial in applications that will involve the tracking of small-scale structures. Moreover, when applying the **L** matrix, the length scales were the same through the quiet and disturbed periods. This setup might be unrealistic for operational data assimilation involving an ionospheric density field that continuously responds to an injection of a wide spectrum of energy. It could be worth investigating machine learning or deep learning applications to ascertain these correlation parameters for the different solar activities, seasons, day and night, and geomagnetic regions.

- Background model inadequacy: the climatological background model (IRI-2016) incorporated into the Gauss Markov filter may be inept at high latitudes, where processes such as particle precipitation are vital drivers of the ionosphere dynamics; a further option is to combine the climatological model with a highly simplified ionosphere physics-based model that allows for 3D advection and processes like precipitation. The expense will remain moderate but with extended applicability. The other feature to consider, especially at high latitudes, is the atmosphere-ionosphere-magnetosphere coupling; lower atmosphere and magnetosphere dynamics can propagate and deposit energy and momentum into the ionosphere system. For a more appropriate representation, the scheme must include such dynamics; we think that interpolating or imposing the current grid boundary conditions to outputs from the observation-driven global magnetosphere and lower atmosphere models might quasi-introduce some of the dynamics in the system but also mitigate the cumbersome process of performing assimilation using a fully coupled system. If the regional and global models target different scales, it would be necessary to monitor if the large-scale dynamics introduced

by the global model do not infiltrate the region domain and shadow or suppress the fine scales.

- Ensemble preparation: the current ensembles are a perturbed version of the climatological model. In extreme ionosphere weather conditions (when deriving parameters significantly deviate from climatological normals), such ensembles might inadequately represent the ionosphere variations. Further thorough research is needed to determine the optimal level of stochastic perturbation for different prognostic model driving parameters, the ensemble sizes and associated distributions for different ionospheric conditions.
- Hybridization: there are some cases where SC4DVar-Inc outperforms the SC4DEnVar, particularly in areas with dense data points. To exploit the advantages of both schemes, SC4DEnVar needs to extend to a hybrid form: wherein the static-climatological and flow-dependent ensemble covariances are optimally combined (in a weighted form that depends on how well we design and improve each covariance).
- Synergistic data assimilation: now that the potential of SC4DEnVar is evident after alleviating the use of TLMS and MAs, as the scheme matures, future assimilation needs to extend beyond ground GNSS data to include synergistic data types such as ionosonde data (as frequency vs virtual height) for the bottom side improvement particularly the E-region, over the satellite TEC to cover the topside and plasmasphere dynamics, ROSTEC - to constraint the vertical structure and also cover the topside up-to LEO satellite orbital altitudes and in situ satellite measurements (such as densities from SWAM satellites).

Acknowledgment

WJM and DK acknowledge funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (ERC Consolidator Grant agreement No. 866357, POLAR-4DSpace). This research is a part of the 4DSpace Strategic Research Initiative at the University of Oslo.

Open Research

The ionosonde data are accessible at ftp://ftp.ngdc.noaa.gov/ionosonde/data/. RO data are accessible in level 2 format at https://cdaac-www.cosmic.ucar.edu/cdaac/. Kp and Dst indices are accessible at http://wdc.kugi.kyoto-u.ac.jp/ or https://omniweb.gsfc.nasa.gov/form/dx1.html. GNSS data from GEONET were provided by the Geospatial Information Authority of Japan and the University NAVSTAR Consortium (UNAVCO) network (ftp://data-out.unavco.org).

References:

Amaechi, P. O., Oyeyemi, E. O., & Akala, A. O. (2018). The response of African equatorial/low-latitude ionosphere to 2015 St. Patrick's Day geomagnetic storm. Space Weather, 16, 601-618, doi:10.1029/2017SW001751.

Astafyeva, E., Zakharenkova, I., & Förster, M. (2015). Ionospheric response to the 2015 St. Patrick's Day storm: a global multi-instrumental overview. J. Geophys. Res. Space Physics, 120, 9023-9037, doi:10.1002/2015JA021629.

Bilitza, D., Altadill, D., Truhlik, V., Shubin, V., Galkin, I., Reinisch, B., & Huang, X. (2017). International Reference Ionosphere 2016: From ionospheric climate to real-time weather predictions. Space Weather, 15, 418-429, doi:10.1002/2016SW001593.

Bilitza, D., Pezzopane, M., Truhlik, V., Altadill, D., Reinisch, B. W., & Pignalberi, A. (2022). The International Reference Ionosphere model: A review and description of an ionospheric benchmark. Reviews of Geophysics, 60, e2022RG000792, doi:10.1029/2022RG000792.

Boullé, N., & Townsend, A. (2021). A generalization of the randomized singular value decomposition. arXiv preprint arXiv:2105.13052.

Buehner, M. (2005). Ensemble-derived stationary and flow dependent background-error covariances: Evaluation in a quasi-operational NWP setting. Q. J. R. Meteorol.Soc., 131, 1013-1043, doi:10.1256/qj.04.15.

Bust, G.S., & Datta-Barua, S. (2014). Scientific Investigations Using IDA4D and EMPIRE. In Modeling the Ionosphere-Thermosphere System (eds J. Huba, R. Schunk and G. Khazanov), doi:10.1002/9781118704417.ch23.

Bust, G. S., Garner, T. W., & Gaussiran, T. L. (2004). Ionospheric Data Assimilation Three-Dimensional (IDA3D): A global, multisensor, electron density specification algorithm, J. Geophys. Res., 109, A11312, doi:10.1029/2003JA010234.

Chen, C. H., Lin, C. H., Matsuo, T., & Chen, W. H. (2016). Ionosphere data assimilation modeling of 2015 St. Patrick's Day geomagnetic storm, J. Geophys. Res. Space Physics, 121, 11,549-11,559, doi:10.1002/2016JA023346.

Cherniak, I., Zakharenkova, I., & Redmon, R. J. (2015). Dynamics of the high-latitude ionospheric irregularities during the 17 March 2015 St. Patrick's Day storm: ground-based GPS measurements. Space Weather, 13, 585-597, doi:10.1002/2015SW001237.

Courtier, P., Thépaut, J.-.-N., & Hollingsworth, A. (1994). A strategy for operational implementation of 4D-Var, using an incremental approach. Q. J. R. Meteorol.Soc.,120, 1367-1387, doi:10.1002/qj.49712051912.

Daley, R. (1991). Atmospheric data analysis, Cambridge atmospheric and space science series. Cambridge University Press. Cambridge, pp. 809- 822.

Decker, D. T., & McNamara, L. F. (2007). Validation of ionospheric weather predicted by Global Assimilation of Ionospheric Measurements (GAIM) models, Radio Sci., 42, RS4017, doi:10.1029/2007RS003632.

Desroziers, G., Camino, J.-T., & Berre, L. (2014). 4DEnVar: Link with 4D state formulation of variational assimilation and different possible implementations. Q. J. R. Meteorol.Soc., 140, 2097-2110, doi:10.1002/qj.2325.

Dmitriev, A. V., Suvorova, A. V., Klimenko, M. V., Klimenko, V. V., Ratovsky, K. G., Rakhmatulin, R. A., & Parkhomov, V. A. (2017). Predictable and unpredictable ionospheric disturbances during St. Patrick's Day magnetic storms of 2013 and 2015 and on 8-9 March 2008, J. Geophys. Res. Space Physics, 122, 2398-2423, doi:10.1002/2016JA023260.

Evensen, G., Vossepoel, F.C., & van Leeuwen, P.J. (2022). Strong-Constraint 4DVar. In: Data Assimilation Fundamentals. Springer Textbooks in Earth Sciences, Geography and Environment. Springer, Cham., doi:10.1007/978-3-030-96709-3_4.

Forsythe, V. V., Azeem, I., Blay, R., Crowley, G., Gasperini, F., Hughes, J., Makarevich, R.A., & Wu, W. (2021). Evaluation of the new background covariance model for the ionospheric data assimilation. Radio Science, 56, e2021RS007286, doi:10.1029/2021rs007286.

Forsythe, V. V., Azeem, I., Crowley, G., & Themens, D. R. (2020). Ionospheric vertical correlation distances: Estimation from ISR data, analysis and implications for ionospheric data assimilation. Radio Science, 56, e2020RS007177, doi:10.1029/2020RS007177

Fuller-Rowell, T. J., & Rees, D. (1980). A three-dimensional time-dependent global model of the thermosphere. Journal of the Atmospheric Sciences, 37, 2545-2567, doi:10.1175/1520-0469(1980)037<2545:Atdtdg>2.0.Co;2.

Fuller-Rowell, T.J., Codrescu, M.V., Moffett, R.J., & Quegan, S. (1994). Response of the thermosphere and ionosphere to geomagnetic storms. J. Geophys. Res. Space Physics, 99, 3893-3914, doi:10.1029/93JA02015.

Garner, T. W., Taylor, B. T., Gaussiran, T. L., Coley, W. R., & Hairston, M. R. (2005). On the distribution of ionospheric electron density observations, Space Weather, 3, S10002, doi:10.1029/2005SW000169.

Garcia-Fernandez, M., Hernandez-Pajares, M., Juan, M., & Sanz, J. (2003). Improvement of ionospheric electron density estimation with GPSMET occultations using Abel inversion and VTEC information. J. Geophys. Res. Space Physics,108,1338, doi:10.1029/2003JA009952.

Gaspari, G., & Cohn, S. E. (1999). Construction of correlation functions in two and three dimensions. Q. J. R. Meteorol.Soc., 125, 723-757, doi:10.1002/qj.49712555417.

Gelb, A. (1974). Applied Optimal Estimation, MIT Press, Cambridge, Mass.

Hajj, G. A., Wilson, B. D., Wang, C., Pi, X., & Rosen, I. G. (2004). Data assimilation of ground GPS total electron content in to a physics-based ionospheric model by use of the Kalman filter. Radio Science, 39, RS1S05, doi:10.1029/2002RS002859.

Halko, N., Martinsson, P. G., & Tropp, J. A. (2011). Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. SIAM review, 53, 217-288, doi:10.1137/090771806.

Houtekamer, P. L., & Mitchell, H. L. (2001). A sequential ensemble Kalman filter for atmospheric data assimilation. Mon. Weather Rev. 129, 123-137, doi:10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2.

Huang, H., Moses, M., Volk, A. E., Elezz, O. A., Kassamba, A. A., & Bilitza, D. (2021). Assessment of IRI-2016 hmF2 model options with digisonde, COSMIC observations for low and high solar flux conditions during 23rd solar cycle. Advances in Space Research, 68, 2093-2103, doi:10.1016/j.asr.2021.01.033.

Huang, C. S., Wilson, G. R., Hairston, M. R., Zhang, Y., Wang, W., & Liu, J. (2016). Equatorial ionospheric plasma drifts and O+ concentration enhancements associated with disturbance dynamo during the 2015 St. Patrick's Day magnetic storm. J. Geophys. Res. Space Physics, 121, 7961-7973, doi:10.1002/2016JA023072.

Huba, J. D., Joyce, G., & Fedder, J. A. (2000). Sami2 is Another Model of the Ionosphere (SAMI2): A new low-latitude ionosphere model, J. Geophys. Res., 105, 23035-23053, doi:10.1029/2000JA000035.

Hsu, C.-T., Matsuo, T., Wang, W., & Liu, J.-Y. (2014). Effects of inferring unobserved thermospheric and ionospheric state variables by using an Ensemble Kalman Filter on global ionospheric specification and forecasting, J. Geophys. Res. Space Physics, 119, 9256- 9267, doi:10.1002/2014JA020390.

Johnson, C., Hoskins, B. J., Nichols, N. K., & Ballard S. P. (2006). A singular vector perspective of 4DVAR: The spatial structure and evolution of baroclinic weather systems. Mon. Weather Rev., 134, 3436-3455, doi:10.1175/MWR3243.1.

Jonathan P. A. (1997). TOPEX dual frequency altimeter studies: Ionospheric corrections and ocean surface measurements, PhD thesis of University of Leicester, Leicester, UK.

Joshi, L. M., Sripathi, S., & Singh, R. (2016). Simulation of low-latitude ionospheric response to 2015 St. Patrick's Day super geomagnetic storm using ionosonde-derived PRE vertical drifts over Indian region. J. Geophys. Res. Space Physics, 121, 2489-2502, doi:10.1002/2015JA021512.

Klimenko, M., Klimenko, V., Despirak, I., Zakharenkova, I., Kozelov, B., Cherniakov, S., & Ratovsky, K. (2017). Disturbances of the thermosphere-ionosphere-plasmasphere system and auroral electrojet at 30° E longitude during the St. Patrick's Day geomagnetic storm on 17-23 March 2015. J. Atmos. Sol.-Terr. Phys., 180, 78-92. doi:10.1016/j.jastp.2017.12.017.

Klobuchar, J. A. (1980). Present state of ionospheric time delay prediction, paper presented at Beacon Satellite Symposium, Warszawa, Poland, 1980.

Kotova, D., Jin, Y., & Miloch, W. J. (2022). Interhemispheric variability of the electron density and derived parameters by the swarm satellites during different solar activity. J. Space Weather Space Clim., 12, doi:10.1051/swsc202207.

Lee, I. T., Matsuo, T., Richmond, A. D., Liu, J. Y., Wang, W., Lin, C. H., Anderson, J. L., & Chen, M. Q. (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.

Lee, I. T., Tsai, H. F., Liu, J. Y., Lin, C. H., Matsuo, T., & Chang, L. C. (2013). Modeling impact of FORMOSAT-7/COSMIC-2 mission on ionospheric space weather monitoring. J. Geophys. Res. Space Physics, 118, 6518-6523. doi:10.1002/jgra.50538.

Lorenc, A. C., (2003). Modelling of error covariances by 4D-Var data assimilation. Q. J. R. Meteorol.Soc., 129, 3167-3182, doi:10.1256/qj.02.131.

Maruyama, T., Ma, G., & Nakamura, M. (2004). Signature of TEC storm on 6 November 2001 derived from dense GPS receiver network and ionosonde chain over Japan. J. Geophys. Res., 109, A10302, doi:10.1029/2004JA010451.

Matsuo, T., & Araujo-Pradere, E. A. (2011). Role of thermosphere-ionosphere coupling in a global ionospheric specification, Radio Science, 46, RS0D23, doi:10.1029/2010RS004576.

McNamara, L. F., Bishop, G. J., & Welsh, J. A. (2011). Assimilation of ionosonde profiles into a global ionospheric model. Radio Science, 46, RS2006, doi:10.1029/2010RS004457.

Mengist, C. K., Yadav, S., Kotulak, K., Bahar, A., Zhang, S.-R., & Seo, K.-H. (2020). Validation of International Reference Ionosphere model (IRI-2016) for F-region peak electron density height (hmF2): Comparison with incoherent scatter radar (ISR) and ionosonde measurements at Millstone Hill. Advances in Space Research, 65, 2773-2781, doi:10.1016/j.asr.2020.03.017.

Mengist, C. K., Seo, K.-H., Kim, Y. H., Eswaraiah, S., Ssessanga, N., & Kwak, Y.-S. (2023). 3-D regional imaging of ionosphere over Africa through assimilating satellite and ground-based data. J. Geophys. Res. Space Physics, 128, e2022JA030859, doi:10.1029/2022JA030859.

Moses, M., Bilitza, D., Kumar Panda, S., & Ochonugor, B. J. (2021). Assessment of IRI-2016 hmF2 model predictions with COSMIC observations over the African region. Advances in Space Research, 68, 2115-2123, doi:10.1016/j.asr.2020.10.029.

Nava, B., Radicella, S. M., Leitinger, R., & Coïsson, P. (2006). A near-real-time model-assisted ionosphere electron density retrieval method, Radio Science, 41, RS6S16, doi:10.1029/2005RS003386.

Nava, B., Rodríguez-Zuluaga, J., Alazo-Cuartas, K., Kashcheyev, A., Migoya-Orué, Y., Radicella, S.M., Amory-Mazaudier, C., & Fleury, R. (2016). Middle- and low-latitude ionosphere response to 2015 St. Patrick's Day geomagnetic storm. J. Geophys. Res. Space Physic, 121, 3421-3438, doi:10.1002/2015JA022299.

Olsen, N., Friis-Christensen, E., Floberghagen, R., Alken, P., Beggan., C. D., Chulliat, A., Doornbos, E., da Encarnao, J.T., Hamilton, B., Hulot, G., Van Ijssel, J., Kuvshinov, A, Lesur V, Lhr H, Macmillan S, Maus S, Noja M, Olsen PEH, Park J, Plank G, Pthe C., Rauberg, J., Ritter, P., Rother, M., Sabaka, T.J, Schachtschneider, R., Sirol, O., Stolle, C., Thbault, E., Thomson, A.W., Tffner-Clausen, P., Velmsk, L., Vigneron, J., Visser, P. (2013). The Swarm satellite

constellation application and research facility (SCARF) and Swarm data products. Earth Planets Space 65, 1189-1200, doi:10.5047/eps.2013.07.001.

Pi, X., Hajj, G. A., Wilson, B. D., Mannucci, A. J., Komjathy, A., Mandrake, L., Wang, C. & Rosen, I. G. (2004). Three-dimensional assimilative ionospheric modeling for regions of large TEC gradient, paper presented at National Technical Meeting, Inst. of Navig., San Diego, Jan.

Prölss, G.W., (1995) Ionospheric F-region storms. In: Volland H (ed) Handbook of atmospheric electrodynamics, 2nd edn. CRC Press, Boca Raton.

Patra, A. K., Chaitanya, P. P., Dashora, N., Sivakandan, M., & Taori, A. (2016). Highly localized unique electrodynamics and plasma irregularities linked with the 17 March 2015 severe magnetic storm observed using multi-technique common-volume observations from Gadanki, India. J. Geophys. Res. Space Physics, 121, 11518-11527, doi:10.1002/2016JA023384.

Pi, X., Wang, C., Hajj, G. A., Rosen, G., Wilson, B. D., & Bailey, G. J. (2003). Estimation of E × B drift using a global assimilative ionospheric model: An observation system simulation experiment, J. Geophys. Res., 108, 1075, doi:10.1029/2001JA009235.

Qian, L., Burns, A.G., Emery, B.A., Foster, B., Lu, G., Maute, A., Richmond, A.D., Roble, R.G., Solomon, S.C., & Wang, W. (2014). The NCAR TIE-GCM: A community model of the coupled thermosphere/ionosphere system. Modeling the ionosphere-thermosphere system, pp.73-83.

Sripathi, S., Sreekumar, S., Banola, S., Emperumal, K., Tiwari, P., & Kumar, B. S. (2015). Low-latitude ionosphere response to super geomagnetic storm of 17/18 March 2015: Results from a chain of ground-based observations over Indian sector, J. Geophys. Res. Space Physics, 120, 10,864– 10,882, doi:10.1002/2015JA021509.

Scherliess, L., Schunk, R. W., Sojka, J. J., & Thompson, D. C. (2004). Development of a physics-based reduced state Kalman filter for the ionosphere. Radio Science, 39, RS1S04, doi:10.1029/2002RS002797.

Scherliess, L., Schunk, R. W., Sojka, J. J., Thompson, D. C., & 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, A11315, doi:10.1029/2006JA011712.

Schunk, R.W., Scherliess, L., Sojka, J.J., Thompson, D. C., Anderson, D.N., Codrescu, M., Minter, C., Fuller-Rowell, T. J., Heelis, R. A., Hairston, M., & Howe, B.M. (2004). Global Assimilation of Ionospheric Measurements (GAIM), Radio Science, 39, RS1S02, doi:10.1029/2002RS002794.

Schunk, R.W., Scherliess, L., Eccles, V., Gardner, L.C., Sojka, J.J., Zhu, L., Pi, X., Mannucci, A.J., Komjathy, A., Wang, C., & Rosen, G., (2021). Challenges in specifying and predicting space weather. Space Weather, 19, e2019SW002404, doi:10.1029/2019SW002404.

Schunk, R.W., Scherliess, L., Eccles, V., Gardner, L.C., Sojka, J.J., Zhu, L., Pi, X., Mannucci, A.J., Butala, M., Wilson, B. D., & Komjathy, A. (2016). Space weather forecasting with a Multimodel Ensemble Prediction System (MEPS), Radio Science, 51, 1157-1165, doi:10.1002/2015RS005888.

Shepherd, S. G. (2014). Altitude-adjusted corrected geomagnetic coordinates: Definition and functional approximations. J. Geophys. Res. Space Physics, 119, 7501-7521. doi:10.1002/2014ja020264.

Ssessanga, N., Kim, Y. H., Habarulema, J. B., & Kwak, Y.-S. (2019). On imaging South African Regional Ionosphere using 4D-var technique. Space Weather, 17, 1584-1604, doi:10.1029/2019SW002321.

Ssessanga, N., Kim, Y. H., & Kim, E. (2015). Vertical structure of medium-scale traveling ionospheric disturbances, Geophys. Res. Lett., 42, 9156-9165, doi:10.1002/2015GL066093.

Ssessanga, N., Kim, Y. H., & Jeong, S.-H. (2017). A statistical study on the F2-layer vertical variation during night-time medium scale travelling ionospheric disturbances. J. Geophys. Res. Space Physics, 122, 3586-3601, doi:10.1002/2016JA023463.

Ssessanga, N., Kim, Y. H., Choi, B., & Chung, J.-K., (2018). The 4D-var estimation of North Korean rocket exhaust emissions into the ionosphere. J. Geophys. Res. Space Physics,123, 2315-2326, doi: 10.1002/2017JA024596.

Ssessanga, N., Yamamoto, M., & Saito, S. (2021a). Assessing the performance of a Northeast Asia Japan-centered 3-D ionosphere specification technique during the 2015 St. Patrick's day geomagnetic storm. Earth Planets Space, 73, 124, doi:10.1186/s40623-021-01447-8.

Ssessanga, N., Yamamoto, M., Saito, S., Saito, A., & Nishioka, M. (2021). Complementing regional ground GNSS-STEC computerized ionospheric tomography (CIT) with ionosonde data assimilation. GPS Solutions, 25, 93, doi:10.1007/s10291-021-01133-y.

Stonebridge, G., Scott, K. A., & Buehner, M. (2018). Impacts on sea ice analyses from the assumption of uncorrelated ice thickness observation errors: Experiments using a 1D toy model, Tellus A, Dynamic Meteorology and Oceanography, 70, 1-13, doi:10.1080/16000870.2018.1445379.

Thompson, D. C., Scherliess, L., Sojka, J. J., & Schunk, R. W. (2006). The Utah State University Gauss-Markov Kalman filter of the ionosphere: The effect of slant TEC and electron density profile data on model fidelity. J. Atmos. Sol.-Terr. Phys., 68, 947-958. https://doi.org/10.1016/j.jastp.2005.10.011.

Tsagouri, I., Belehaki, A., Moraitis, & G., Mavromichalaki, H. (2000). Positive and negative ionospheric disturbances at middle latitudes during geomagnetic storms. Geophys. Res. Lett., 27, 3579-3582, doi:10.1029/2000GL003743.

Tulasi Ram, S., Nilam, B., Balan, N., Zhang, Q., Shiokawa, K., Chakrabarty, D., Xing, Z., Venkatesh, K., Veenadhari, B., & Yoshikawa,A. (2019). Three different episodes of prompt equatorial electric field perturbations under steady southward IMF Bz during St. Patrick's Day storm. J. Geophys. Res. Space Physics, 124, 10428-10443, doi:10.1029/2019JA027069.

Tulasi Ram, S., Yokoyama, T., Otsuka, Y., Shiokawa, K., Sripathi, S., Veenadhari, B., Heelis, R. Ajith, K. K., Gowtam, V. S., Gurubaran, S., Supnithi, P., & Le Huy, M. (2016). Duskside enhancement of equatorial zonal electric field response to convection electric fields during the St. Patrick's Day storm on 17 March 2015. J. Geophys. Res. Space Physics, 121, 538-548, doi:10.1002/2015JA021932.

Venkatesh, K., Tulasi Ram, S., Fagundes, P. R., Seemala, G. K., & Batista, I. S. (2017). Electrodynamic disturbances in the Brazilian equatorial and low-latitude ionosphere on St. Patrick's Day storm of 17 March 2015. J. Geophys. Res. Space Physics, 122, 4553-4570, doi:10.1002/2017JA024009.

Wu, C. C., Liou, K., Lepping, R. P., Hutting, L., Plunkett, S., Howard, R. A., & Socker, D. (2016). The first super geomagnetic storm of solar cycle 24: "The St. Patrick's day event (17 March 2015)". Earth, Planets and Space, 68, 151, doi:10.1186/s40623-016-0525-y.

Wu, X., Hu, H., Gong, X., Zhang, X., & Wang, X. (2009). Analysis of inversion errors of ionospheric radio occultation. GPS Solutions, 13, 231-239, doi:10.1007/s10291-008-0116-x.

Yamazaki, Y., Matzka, J., Stolle, C., Kervalishvili, G., Rauberg, J., Bronkalla, O., Morschhauser, A., Bruinsma, S., Shprits, Y. Y., & Jackson, D. R. (2022). Geomagnetic activity index Hpo. Geophys. Res. Lett., 49, e2022GL098860, doi:10.1029/2022GL098860.

Yasyukevich, Y., Astafyeva, E., Padokhin, A., Ivanova, V., Syrovatskii, S., & Podlesnyi, A. (2018). The 6 September 2017 X-class solar flares and their impacts on the ionosphere, GNSS, and HF radio wave propagation. Space Weather, 16, 1013-1027. doi:10.1029/2018SW001932.

Yue, X., Schreiner, W. S., Lei, J., Sokolovskiy, S. V., Rocken, C., Hunt, D. C., & Kuo, Y.-H. (2010). Error analysis of Abel retrieved electron density profiles from radio occultation measurements. Ann. Geophys, 28, 217-222. doi:10.5194/angeo-28-217-2010.

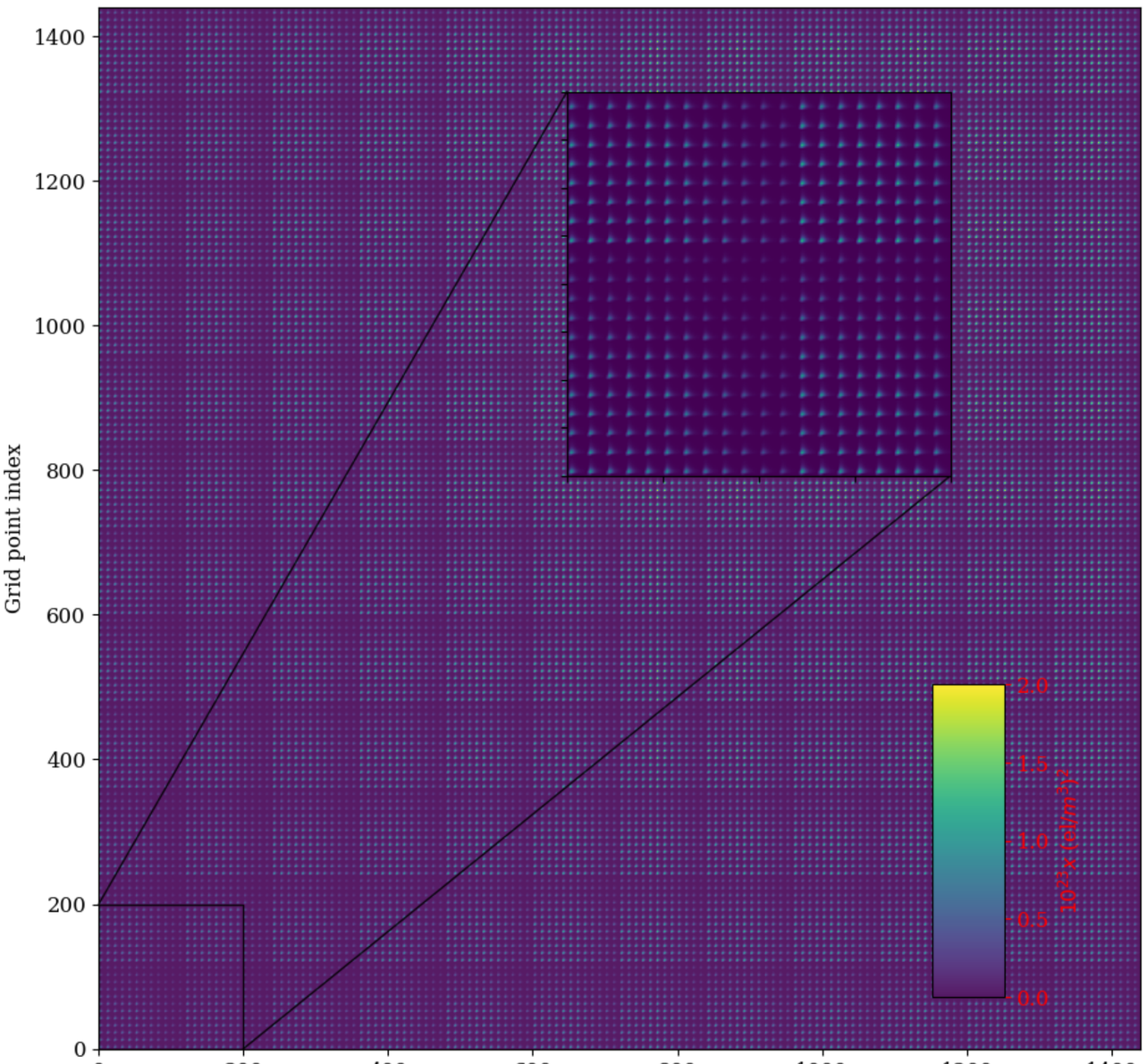
Yue, X., Wan, W., Liu, L., Fei, Z., Lei, J., Zhao, B., Xu, G., Zhang, S.-R., & Zhu, J. (2007). Data assimilation of incoherent scatter radar observation into a one-dimensional midlatitude ionospheric model by applying ensemble Kalman filter. Radio Science, 42, RS6006, doi:10.1029/2007RS003631.

1491 Zou, X., Vandenberghe, F., Pondeca, M., Kuo, Y., (1997). Introduction to adjoint techniques and
1492 the vmm5 adjoint modeling system. Tech. rep., NCAR Technical note.
1493

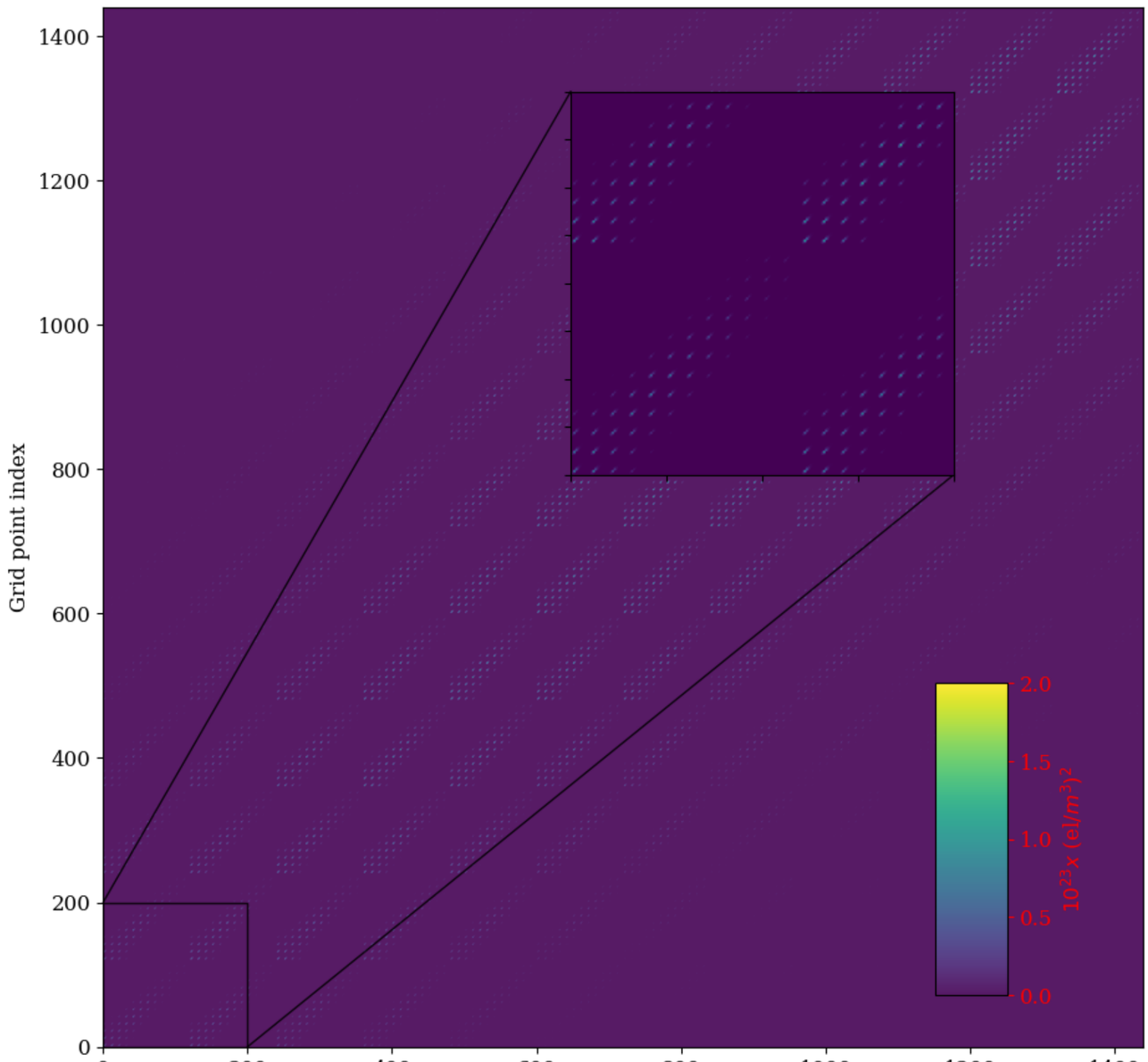
Figure 1.

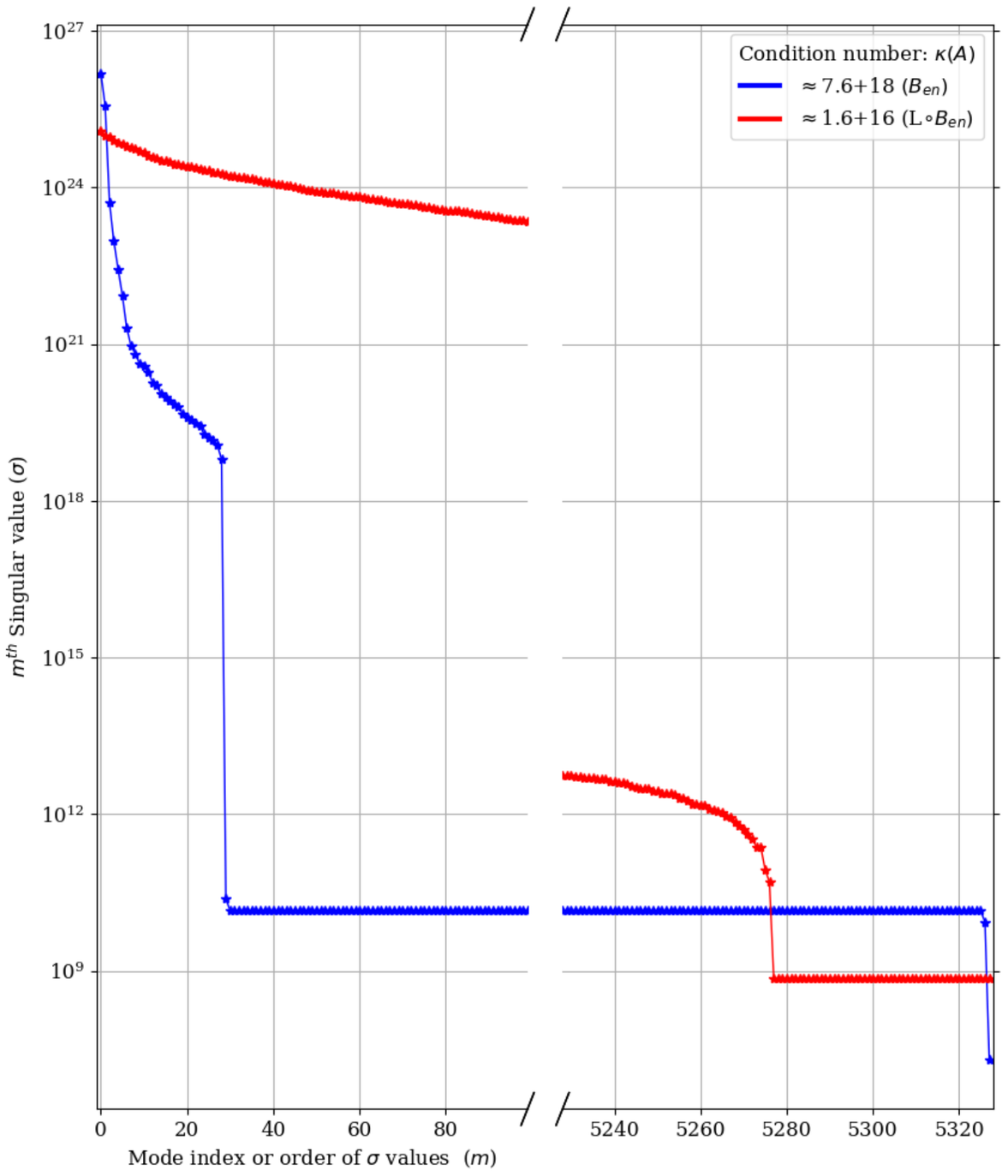
Ensemble(size=30) background error covariance analysis

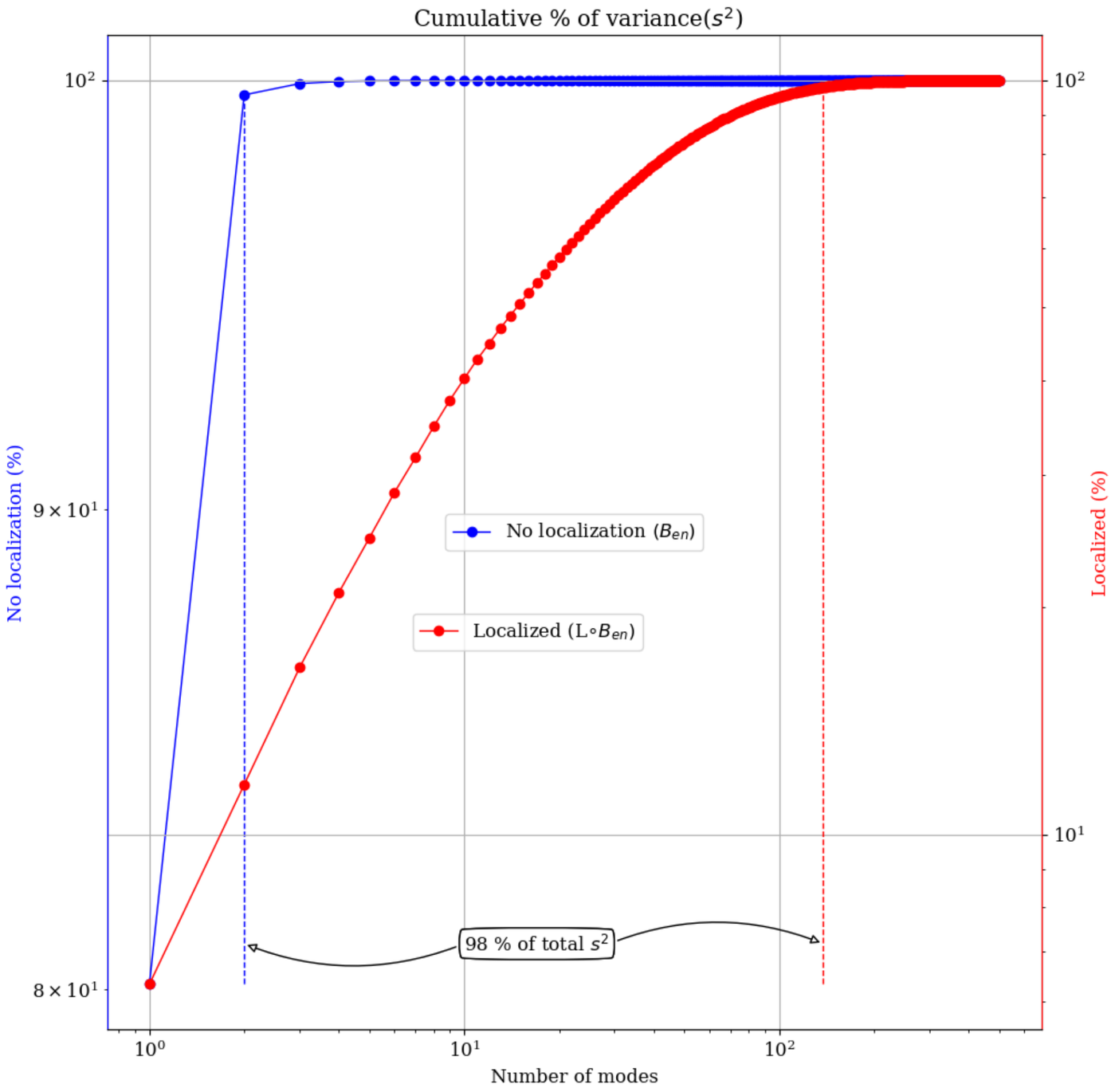
B matrix: No localizaion

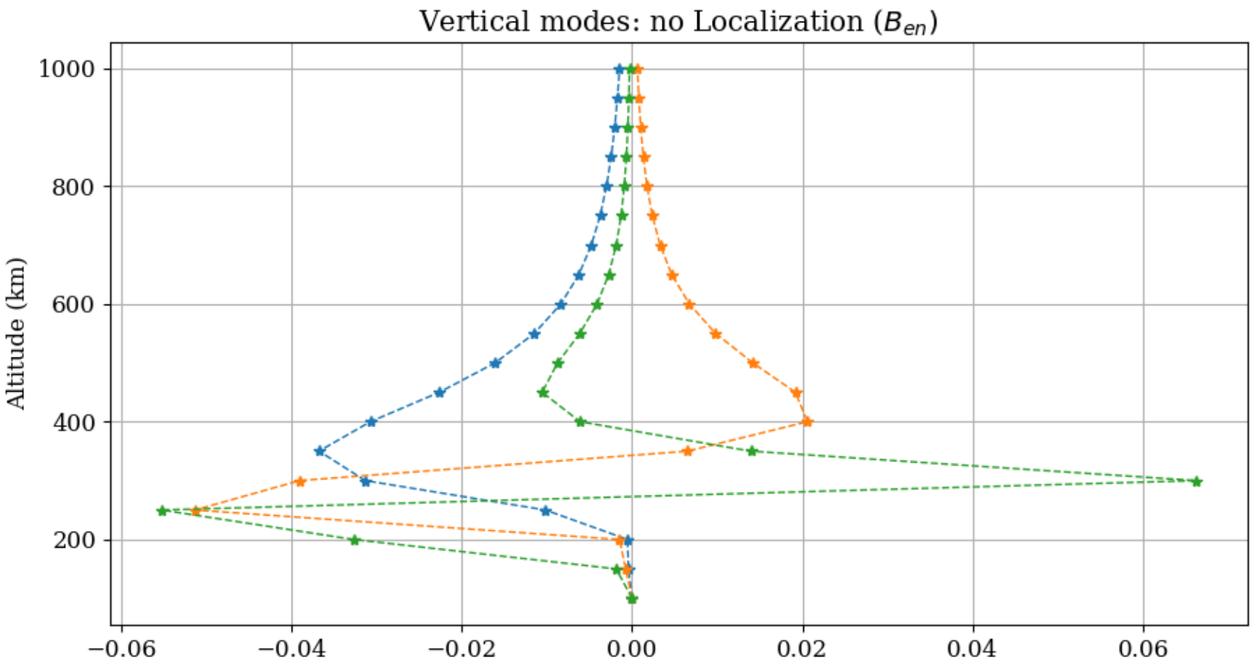


Localized: L∘B









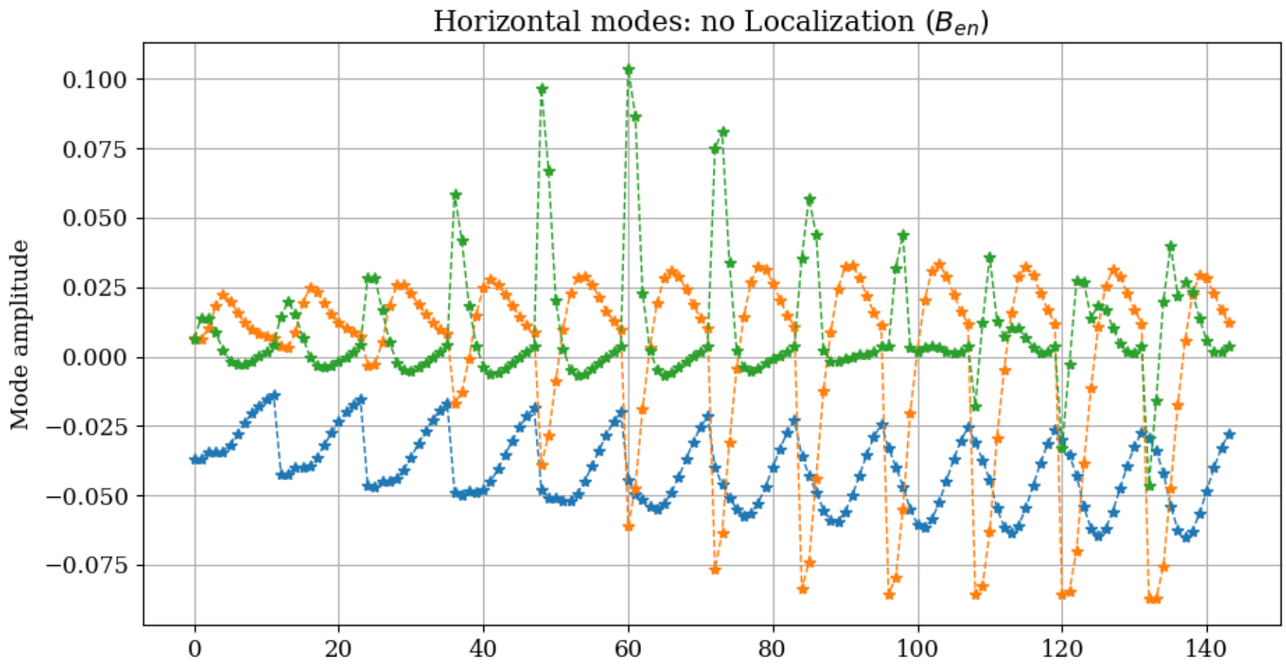


Figure 2.

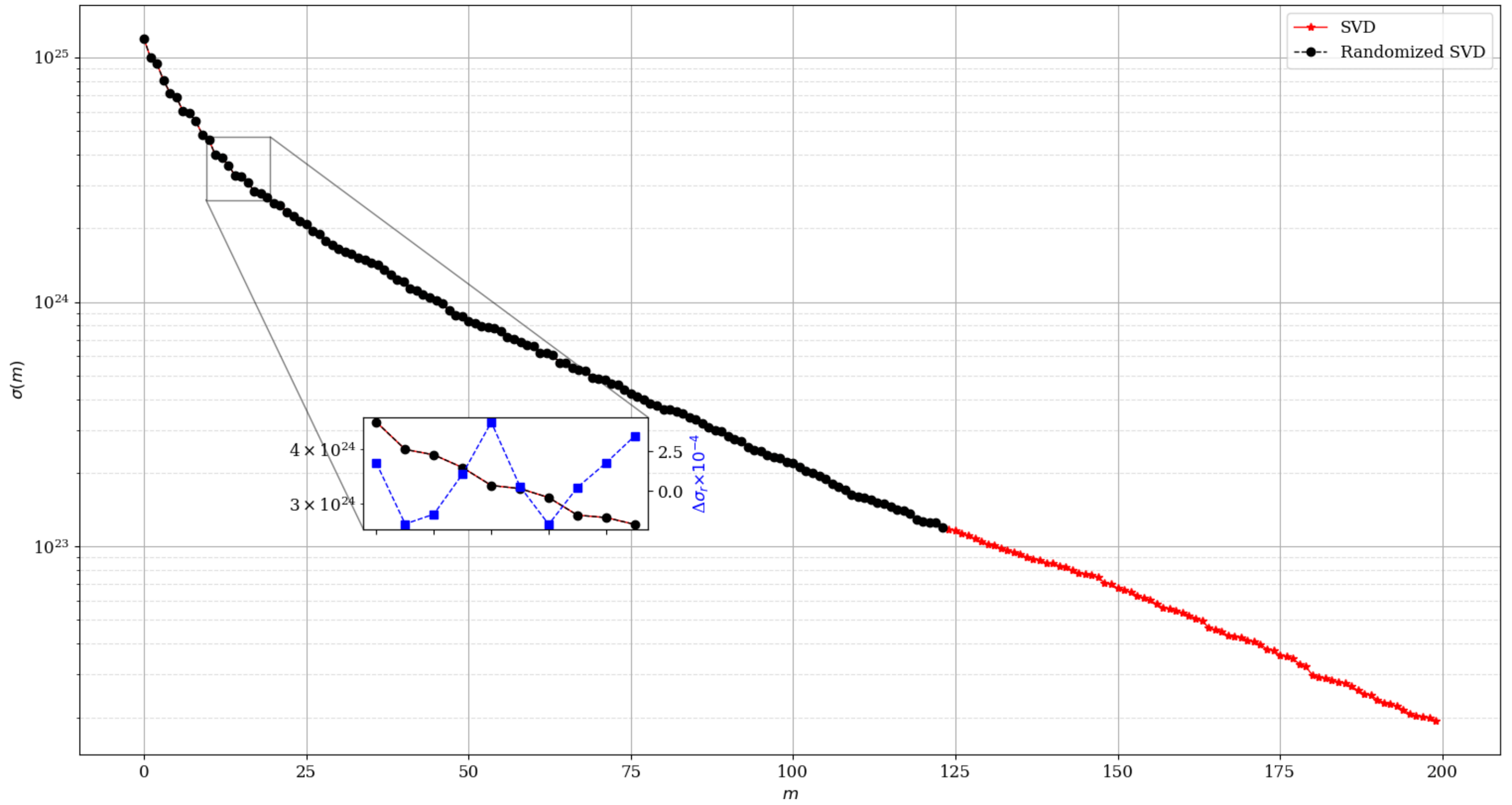

SVD
Randomized SVD
10^{25}
10^{24}
10^{23}
$\sigma(m)$
0
25
50
75
100
125
150
175
200
m
4×10^{24}
3×10^{24}
2.5
0.0
$\Delta\sigma_r \times 10^{-4}$

Figure 3.

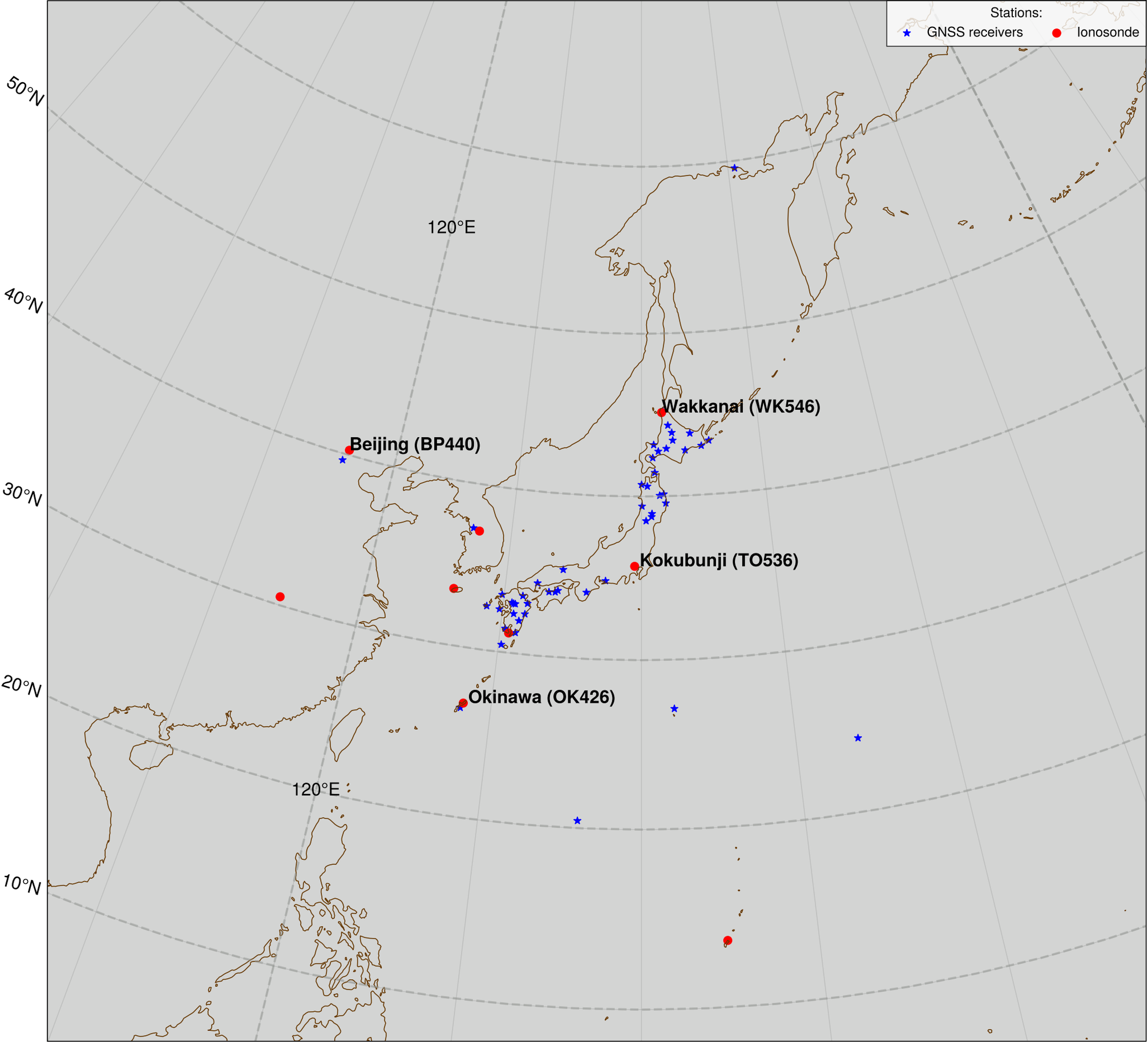
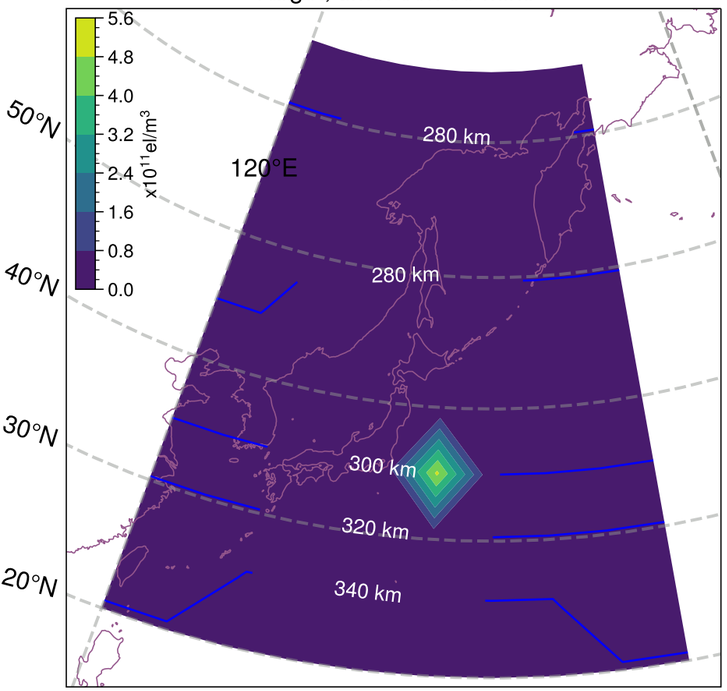
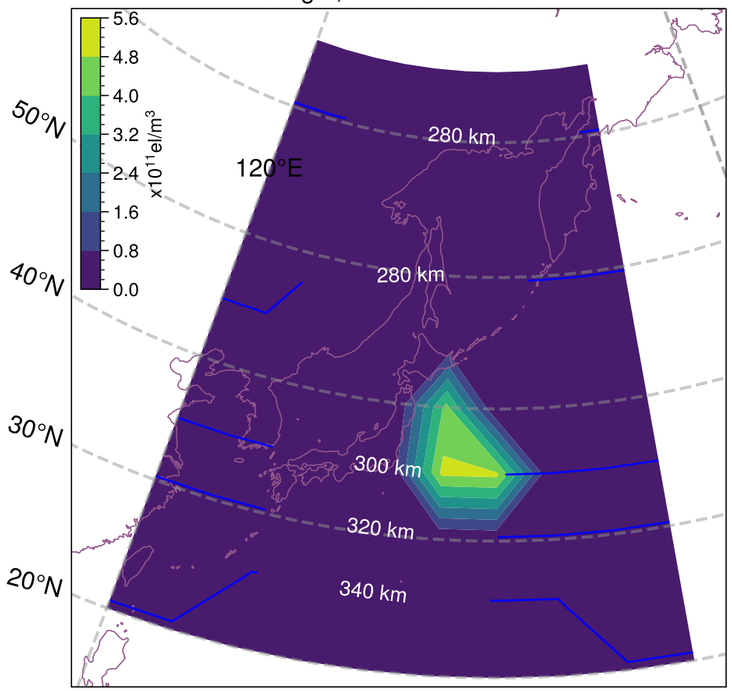

Stations:
GNSS receivers
Ionosonde
50°N
40°N
30°N
20°N
10°N
120°E
120°E
Wakkanai (WK546)
Beijing (BP440)
Kokubunji (TO536)
Okinawa (OK426)

Figure 4.

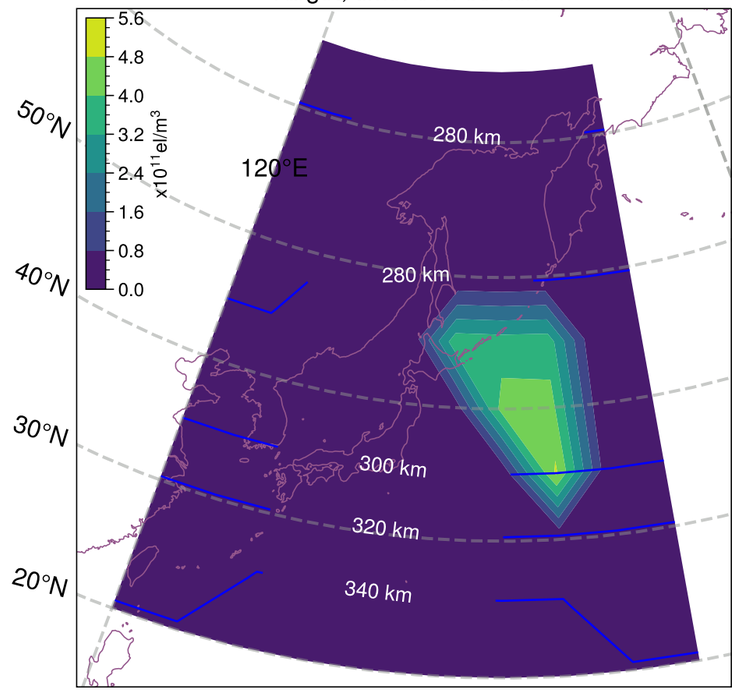
Target, time: 00h 00 mins



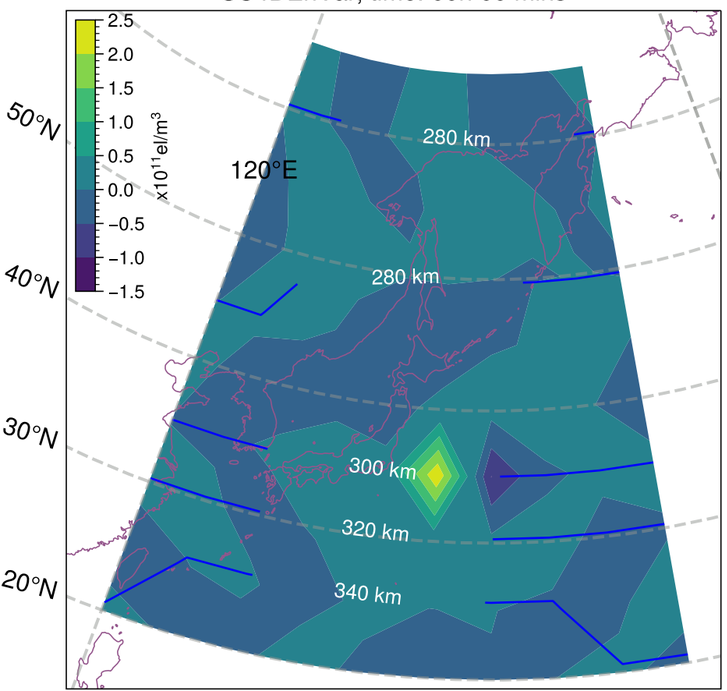
Target, time: 00h 30 min



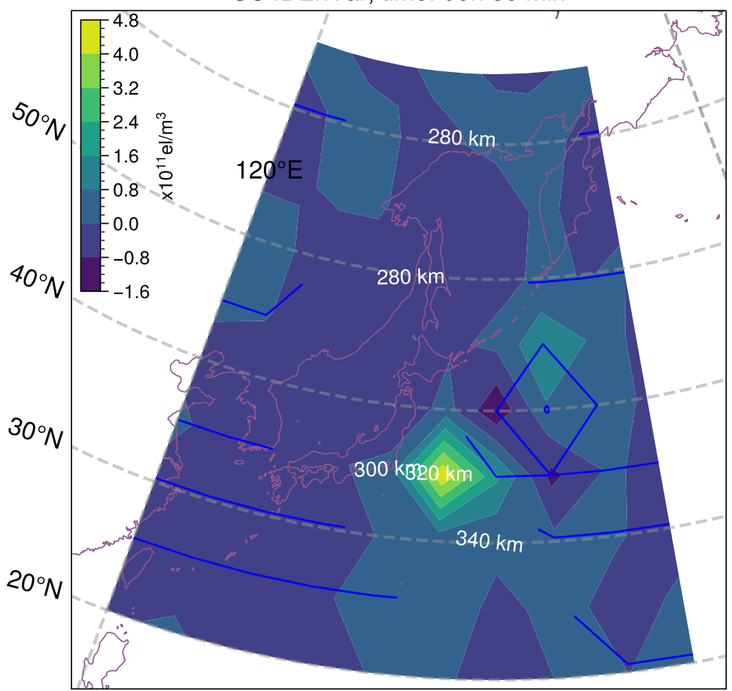
Target, time: 01h 30 min



SC4DEnVar, time: 00h 00 mins



SC4DEnVar, time: 00h 30 min



SC4DEnVar, time: 01h 30 min

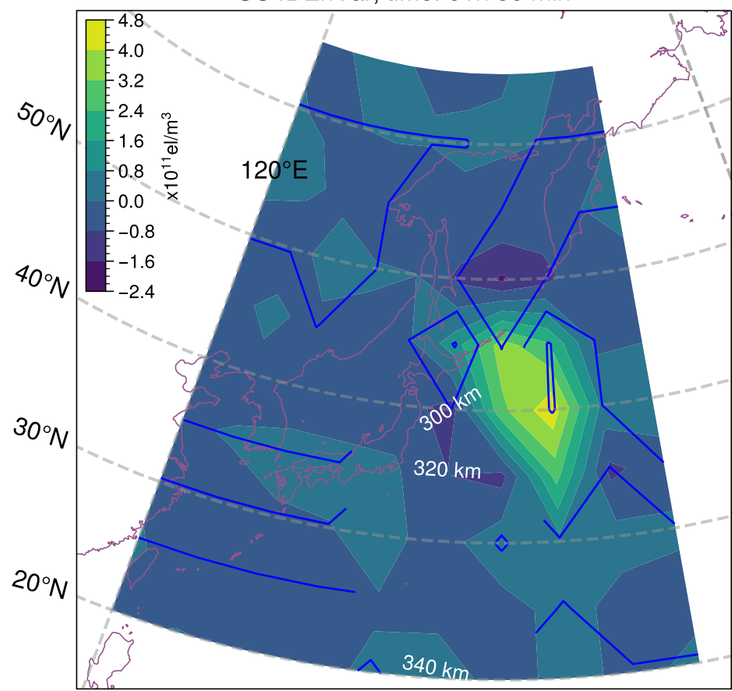


Figure 5.

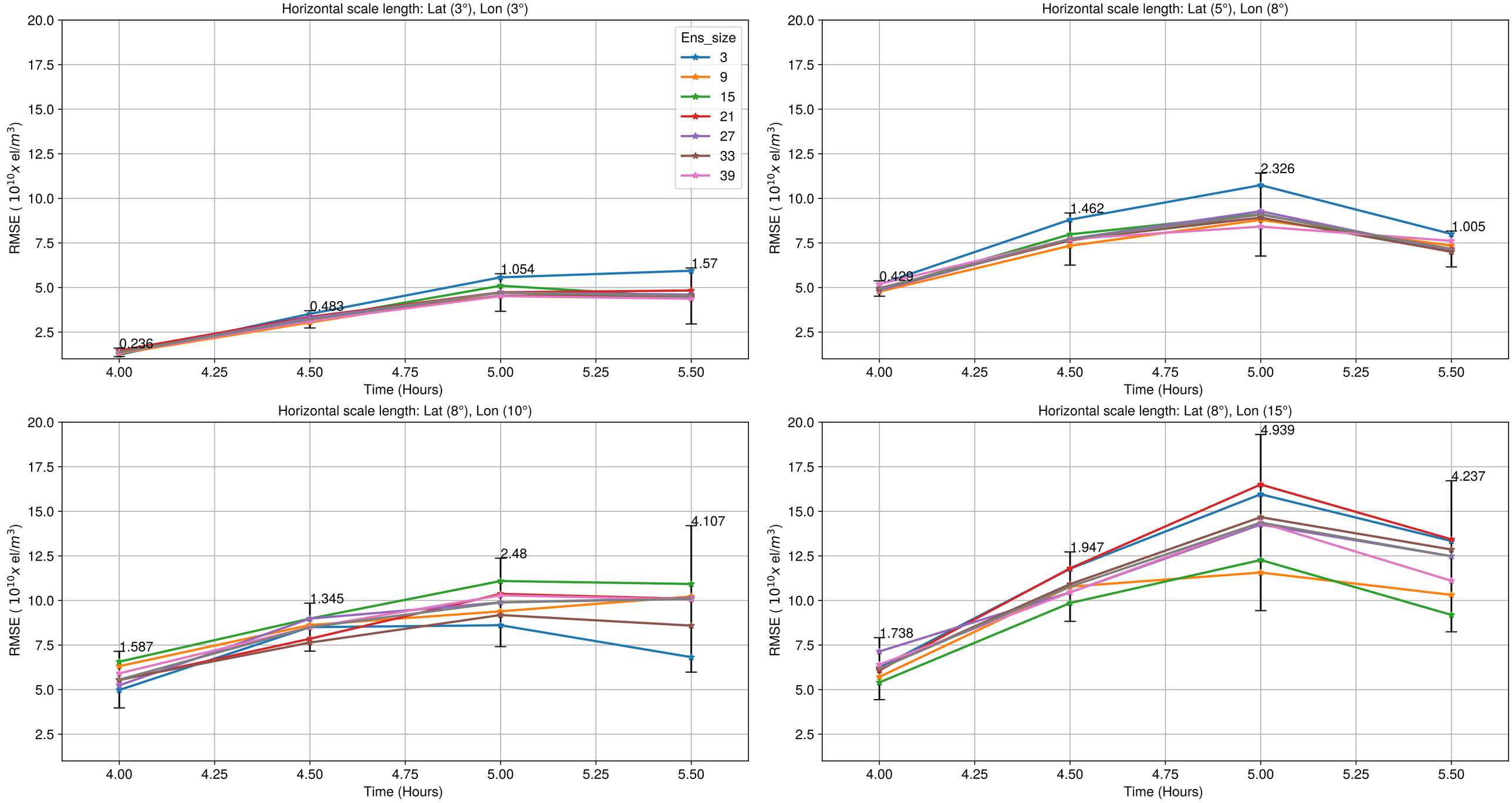
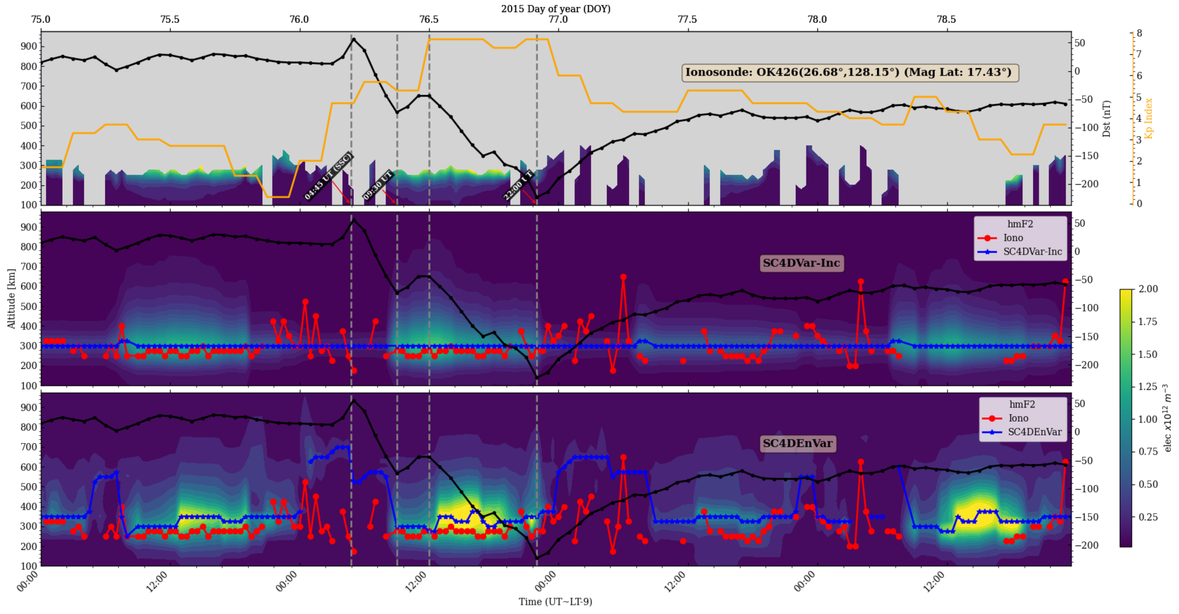
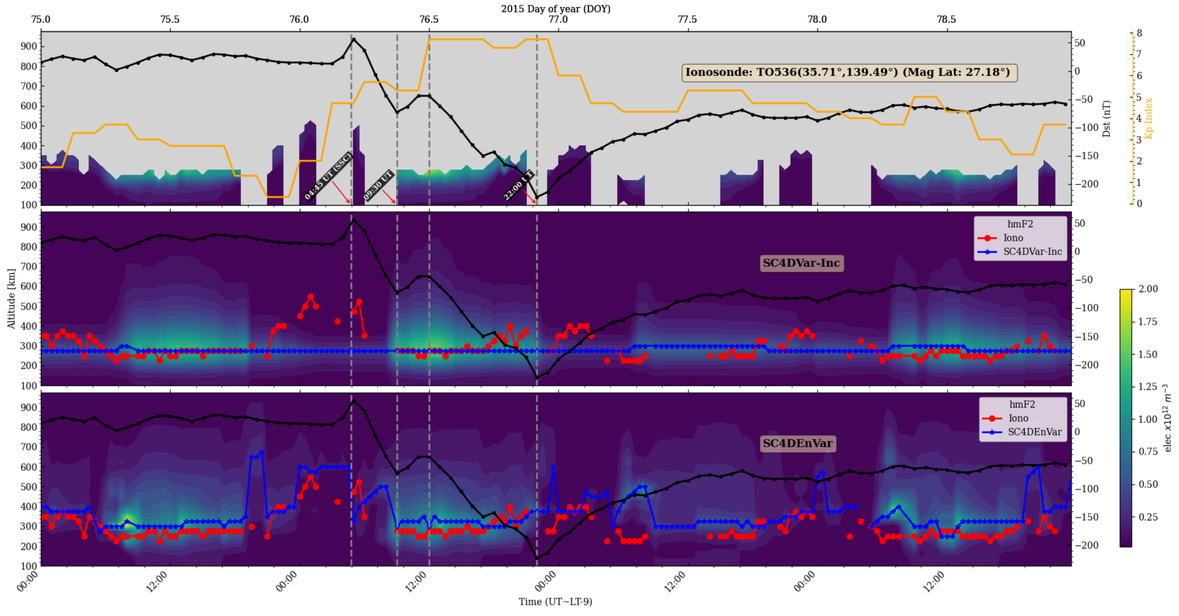
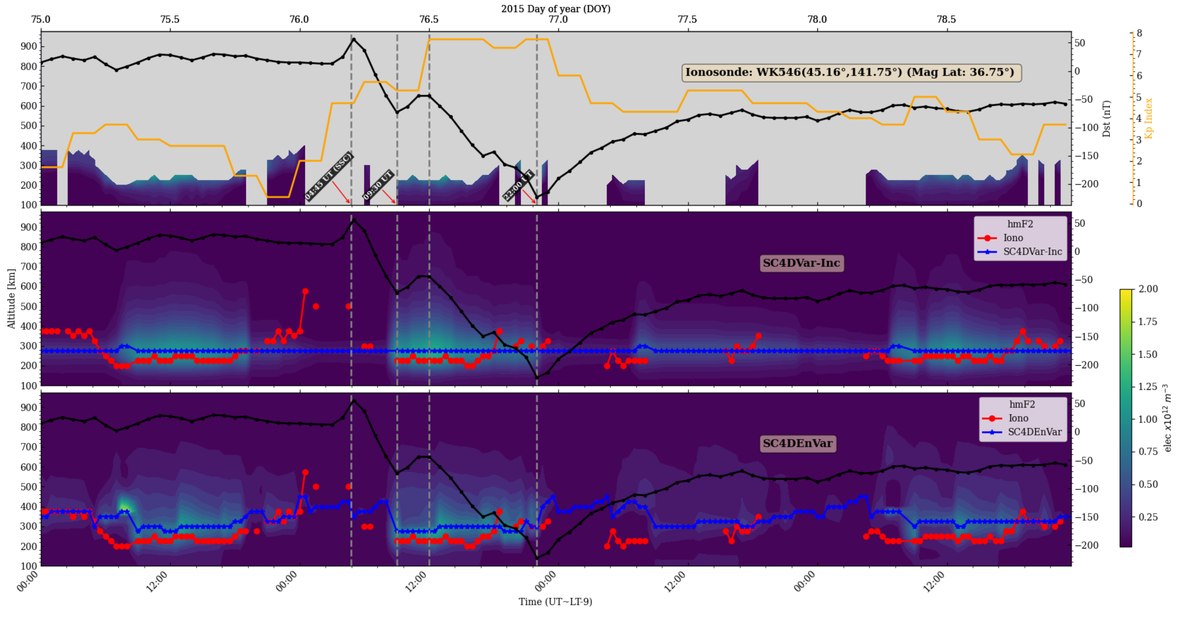

Horizontal scale length: Lat (3°), Lon (3°)
Horizontal scale length: Lat (5°), Lon (8°)
Horizontal scale length: Lat (8°), Lon (10°)
Horizontal scale length: Lat (8°), Lon (15°)
Ens_size
3
9
15
21
27
33
39
RMSE (10^{10}x el/m^3)
Time (Hours)
0.236
0.483
1.054
1.57
0.429
1.462
2.326
1.005
1.587
1.345
2.48
4.107
1.738
1.947
4.939
4.237

Figure 6.


2015 Day of year (DOY)
Ionosonde: OK426(26.68°,128.15°) (Mag Lat: 17.43°)
04:45 LT (ESC)
09:30 LT
22:00 LT
Dst (nT)
Kp Index
SC4DVar-Inc
SC4DEnVar
hmF2
Iono
Altitude [km]
elec x10^12 m^-3
Time (UT~LT-9)


2015 Day of year (DOY)
Ionosonde: TO536(35.71°,139.49°) (Mag Lat: 27.18°)
04:45 LT (ESC)
09:30 LT
22:00 LT
Dst (nT)
Kp Index
SC4DVar-Inc
SC4DEnVar
hmF2
Iono
Altitude [km]
elec x10^12 m^-3
Time (UT~LT-9)


2015 Day of year (DOY)
Ionosonde: WK546(45.16°,141.75°) (Mag Lat: 36.75°)
04:45 LT (ESC)
09:30 LT
22:00 LT
Dst (nT)
Kp Index
SC4DVar-Inc
SC4DEnVar
hmF2
Iono
Altitude [km]
elec x10^12 m^-3
Time (UT~LT-9)

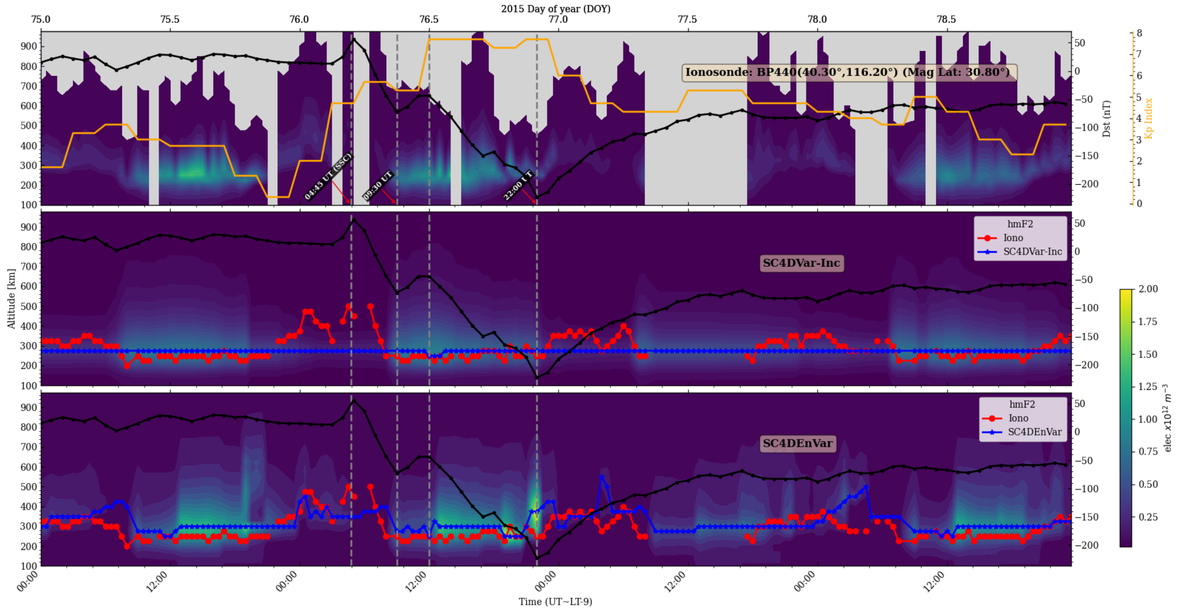

2015 Day of year (DOY)
Ionosonde: BP440(40.30°,116.20°) (Mag Lat: 30.80°)
04:45 LT (ESC)
09:30 LT
22:00 LT
Dst (nT)
Kp Index
SC4DVar-Inc
SC4DEnVar
hmF2
Iono
Altitude [km]
elec x10^12 m^-3
Time (UT~LT-9)

Figure 7.

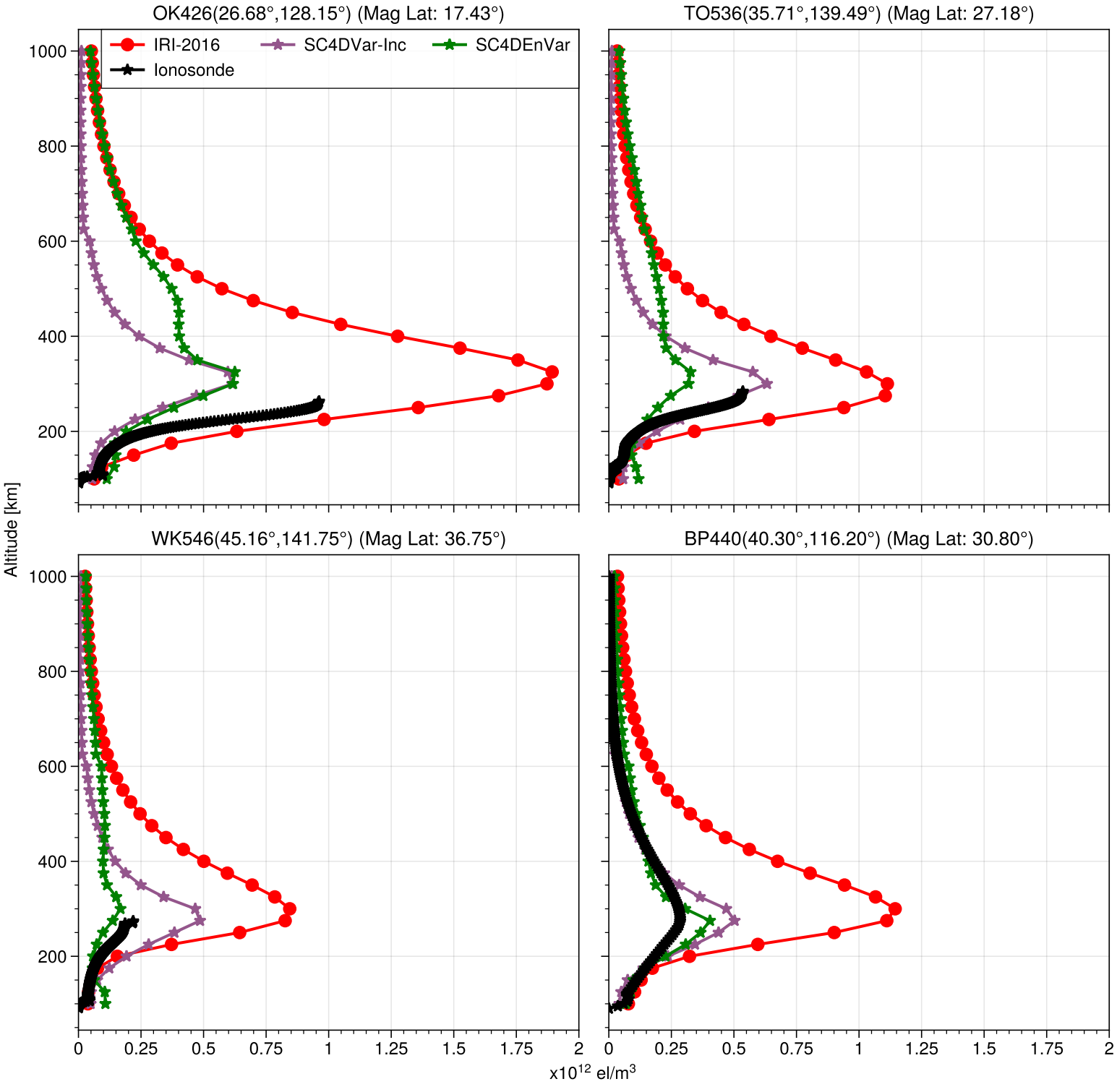

Vertical profiles at ionosonde locations: 2015-03-18 08:15:00
OK426(26.68°,128.15°) (Mag Lat: 17.43°)
TO536(35.71°,139.49°) (Mag Lat: 27.18°)
WK546(45.16°,141.75°) (Mag Lat: 36.75°)
BP440(40.30°,116.20°) (Mag Lat: 30.80°)
IRI-2016
SC4DVar-Inc
SC4DEnVar
Ionosonde
Altitude [km]
x10^12 el/m^3

Figure 8.

COSMIC RO profile electron densities vs. SC4DVar: 2015-03-15

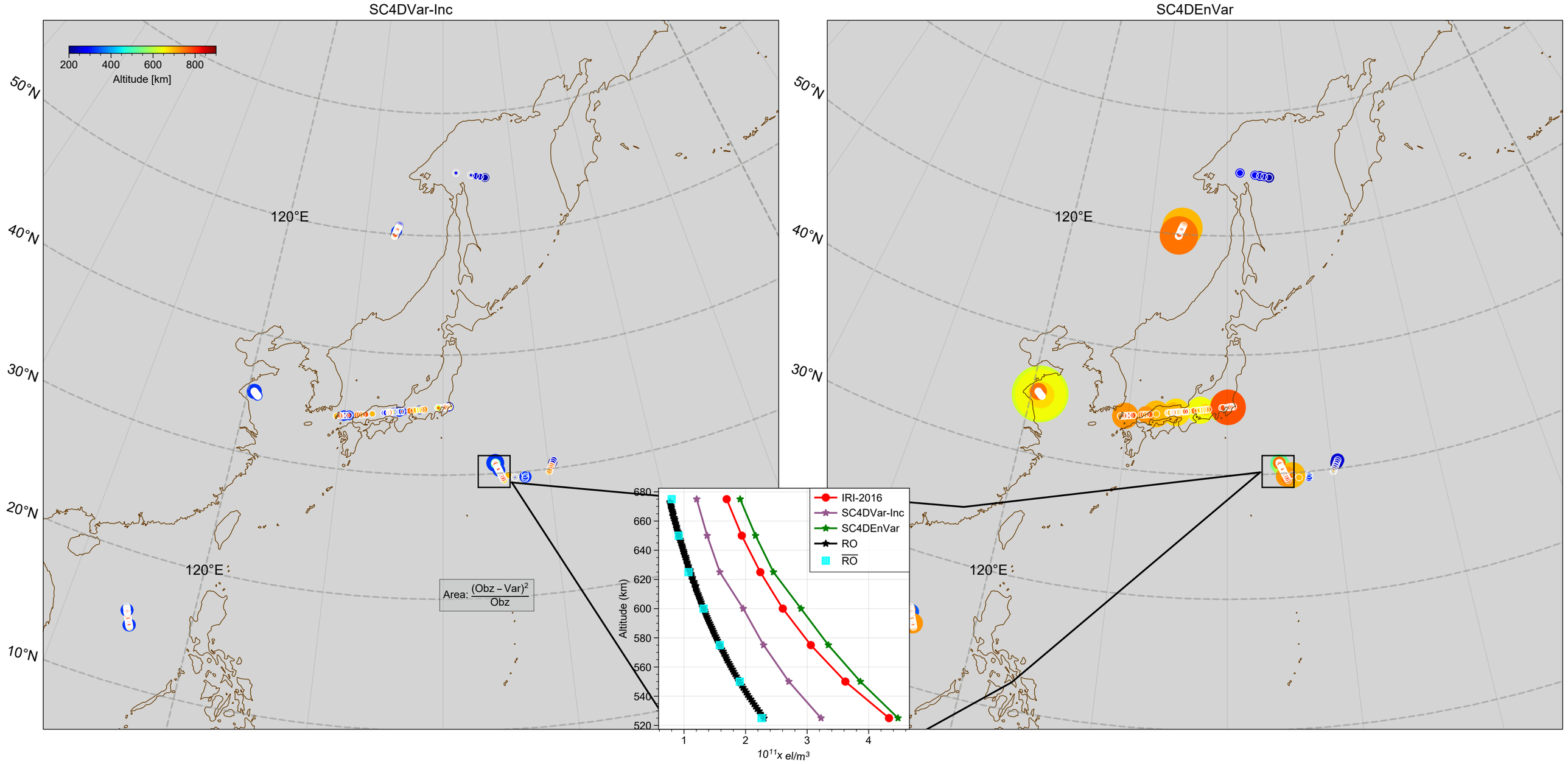
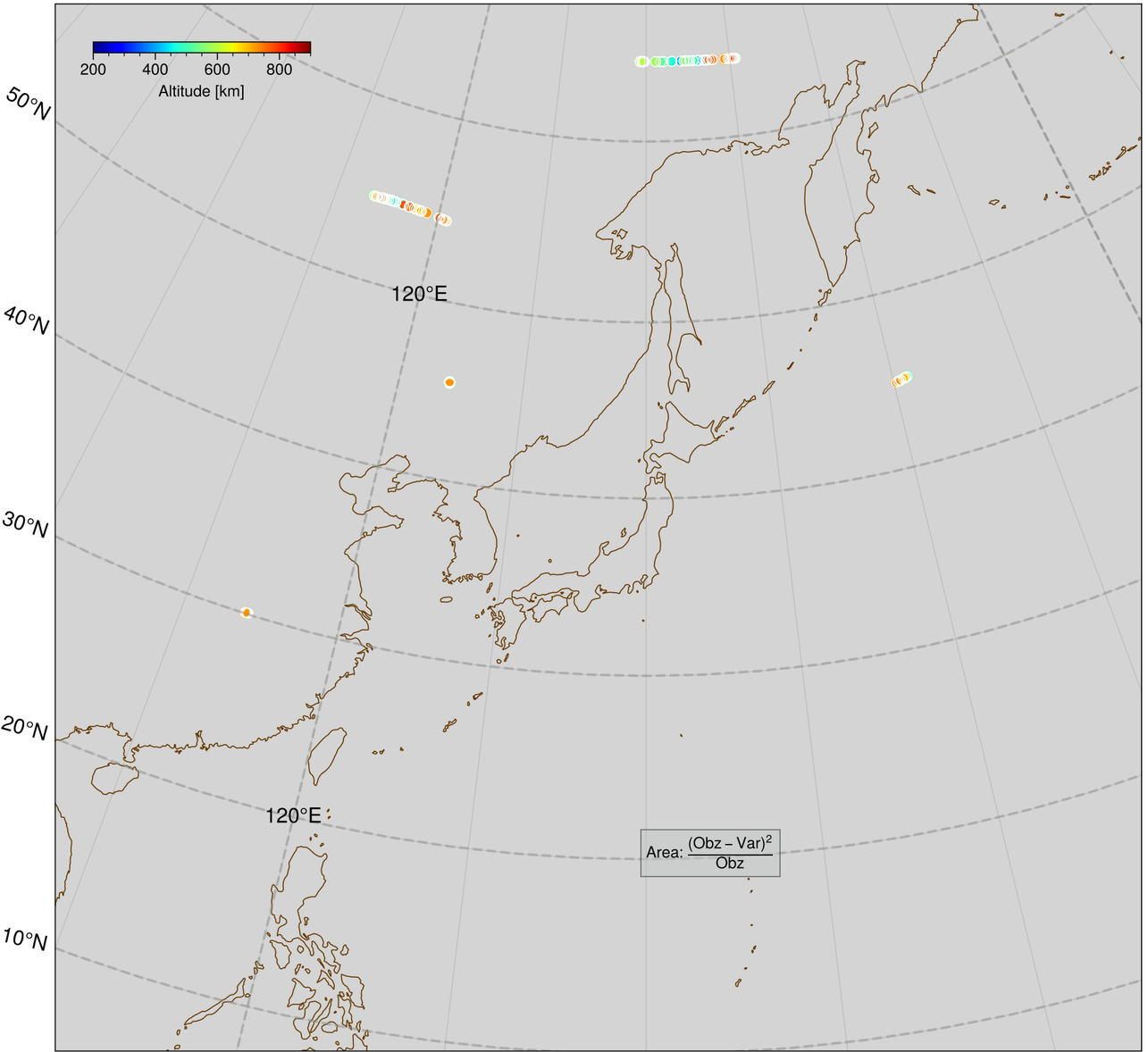


Figure 9.

COSMIC RO profile electron densities vs. SC4DVar: 2015-03-18

SC4DVar-Inc



SC4DEnVar

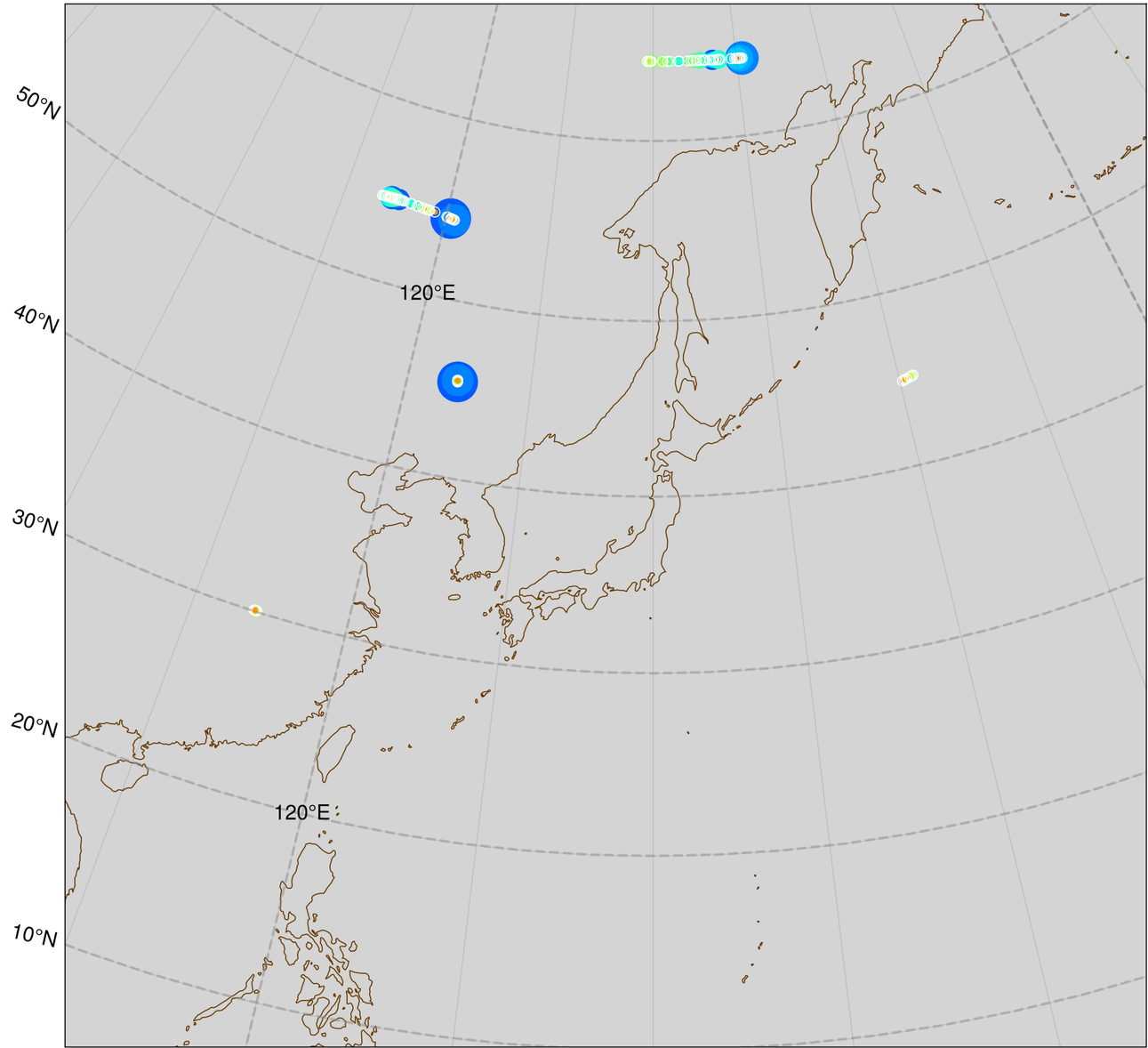


Figure 10.

2015-03-16
OCC TEC: COSMIC vs. SC4DVar

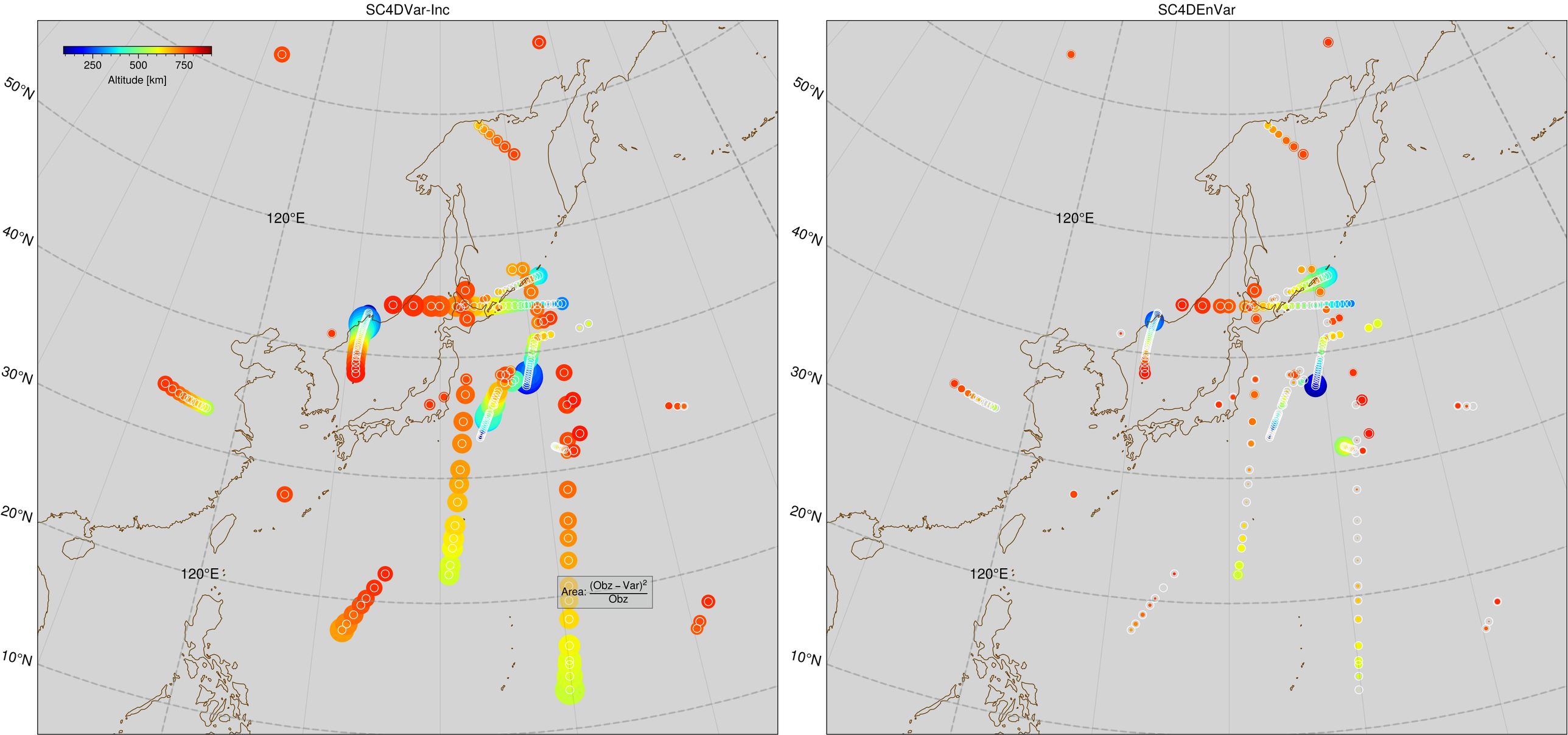


Figure 11.

2015-03-18
OCC TEC: COSMIC vs. SC4DVar