URSI AP-RASC, Sydney, Australia, 17 – 22 August 2025

Improving Space Weather Specification Using Effective Indices

Benjamin Reid
SERENE, University of Birmingham, Birmingham, UK

Abstract

The variability of the near-Earth space environment is greater than that described by existing empirical or physics-based models. Accordingly, data assimilation models are growing in both sophistication and importance in forecasting space weather. While traditional data assimilation uses the ionospheric state produced by a model as a background state, other approaches exist which modify the input drivers provided to the background as a mean to propagate information. This study will examine a simple data assimilation experiment where traditional data assimilation, a driver-only approach, and a mixed approach are compared. This experiment suggests that a mixed approach, where both the ionospheric state and the background drivers are modified, is able to outperform other options.

1 Introduction

The ionosphere is a region of ionized plasma in the upper atmosphere extending from approximately 70 kilometres in altitude to over 2000 kilometres. This ionization occurs due to extreme ultraviolet photoionization from the sun, from auroral precipitation, and other sources. The behaviour of the ionosphere is affected by solar activity, changes to the neutral atmosphere, via coupling with the magnetosphere and solar wind, and is influenced by gravity waves propagating upward from the lower atmosphere. As a plasma, the ionosphere affects electromagnetic radiation which propagates through it. This affects technological systems on Earth and in space, and is also the primary means available to characterize and study the ionosphere.

This study will use ionosonde data from the Global Ionospheric Radio Observatory (GIRO) [1]. Ionosondes broadcast a sweep of High Frequency (HF) radio waves, and record the reflected signals. The HF signals are reflected at an altitude where the plasma frequency matches the signal frequency, allowing a profile of the bottom half of the ionosphere to be constructed. Two important parameters which can be obtained using this technique are the critical frequency at the F2 layer peak $foF2$, and the altitude of the peak $hmF2$.

Both $foF2$ and $hmF2$ are highly variable, and have a significant impact on HF and trans-ionospheric radio propagation. Modelling and predicting $foF2$ and $hmF2$ are thus important capabilities of any ionospheric space weather model. The data assimilation models used in this study are simplified, essentially toy models, however the techniques used are analogous to true space weather models such as A-CHAIM [2]. The experiment in this study will help to characterize what benefit, if any, can be seen by adjusting the driving parameters of a background model in a data assimilation alongside the rest of the ionospheric state.

2 Method

The experiment will compare several models, three of which are data assimilation models using a particle filtering scheme (Section 2.1), and the other being the International Reference Ionosphere (IRI) [3]. All models assimilate $foF2$ and $hmF2$ data from autoscaled ionosondes. 6 of the ionosondes, PF765, JR055, PQ052, AT138, and WP937, were not included in the assimilated dataset, and held aside for validation and analysis. The predicted $foF2$ and $hmF2$ at each of these locations was determined by taking the model value at the nearest assimilated ionosonde. These are EI764 for PF765, DB049 for JR055 and PQ052, VT139 for AT138, and MHJ45 for WP937. Any data which does not meet the selection criteria from Themens *et al.* [4] is discarded, both for the assimilation and the analysis.

2.1 Particle Filters

Let $\mathscr{S}$ be some physical system of interest, such as the ionosphere. The properties of this system are described using some choice of parameterization, for example the electron density at a set of grid points of fixed latitude, longitude, and altitude. A specific configuration of $\mathscr{S}$ is then described by a n_X-dimensional state vector $\mathbf{x}$, which exists in a n_X-dimensional state space $X \subseteq \mathbb{R}^{n_X}$. The set of all possible configurations of this system, for a fixed parameterization, is therefore X.

To describe the time evolution of the system, let $t_{1:n} = \{t_1, t_2, ..., t_{n-1}, t_n\}$ be a sequence of times during which the system $\mathscr{S}$ is presumed to be static. The system then evolves through a sequence of discrete state vectors $\mathbf{X}_{1:n} = \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{n-1}, \mathbf{X}_n\}$. This discrete-time assumption is common across many data assimilation techniques, including the family Kalman filters [5]. For this assumption to be valid, the difference between successive times in $t_{1:n}$ must be smaller than the characteristic timescales of $\mathscr{S}$.

The transition between one state $\mathbf{X}_k$ and another $\mathbf{X}_{k+1}$ is modelled as a random process with a probability distribution $p(\mathbf{x} = \mathbf{X}_{k+1}|\mathbf{X}_k, \theta, t)$. This does not depend on any states before $\mathbf{X}_k$, and is therefore a Markov process. The true configuration of $\mathscr{S}$ at any instant t_k is unknown (or hidden), and is only partially observed through a set of observations $\mathbf{y}_k$. For a state $\mathbf{X}_k$, the likelihood of obtaining any given observation $\mathbf{y}_k$ is $p(\mathbf{y}_k|\mathbf{X}_k)$. Thus, the core problem which data assimilation attempts to solve is

$$p(\mathbf{X}_{1:n}|\mathbf{y}_{1:n}) = \frac{p(\mathbf{X}_{1:n})p(\mathbf{y}_{1:n}|\mathbf{X}_{1:n})}{p(\mathbf{y}_{1:n})} \tag{1}$$

Particle filters are a Monte Carlo technique to solve Equation 1, using a weighted ensemble of representative states $\mathbf{X}^i_{1:n} = \{\mathbf{X}^i_{1:n}|i = 1,\ldots,N\}$. Each particle is assigned a weight w^i, which is updated at each assimilation step using $p(\mathbf{y}_k|\mathbf{X}^i_k)$ and $p(\mathbf{X}_k|\mathbf{X}_{k-1})$, and another density function called the importance density $q_k(\mathbf{X}^i_k|\mathbf{X}^i_{k-1})$.

$$w^i_n = w^i_1 \prod_{k=2}^{n} \frac{p(\mathbf{X}^i|\mathbf{X}^i_{k-1})p(\mathbf{y}_k|\mathbf{X}^i_k)}{q_k(\mathbf{X}^i_k|\mathbf{X}^i_{k-1})} \tag{2}$$

The probability density $p(\mathbf{x}_{1:n}|\mathbf{y}_{1:n})$ can then be approximated by

$$\hat{p}(\mathbf{x}_{1:n}|\mathbf{y}_{1:n}) = \frac{\sum_{i=1}^{N} w^i_n \delta_{\mathbf{X}^i_{1:n}}}{\sum_{i=1}^{N} w^i_n} \tag{3}$$

For more detail on particle filtering in an ionospheric context, see Reid *et al.* [2].

2.2 The Models

2.2.1 The IRI

The IRI is a climatological empirical ionospheric model [3]. It is used as the background for all data assimilation models in this study. The IRI is driven by a number of geophysical indices, including the IG ionospheric index, and the sunspot number Rz. The IG index is produced by fitting observed $foF2$ at a number of reference ionosondes, and controls $foF2$ in the IRI. The sunspot number Rz is determined using solar observations, and controls $hmF2$.

2.2.2 The Control Data Assimilation

This model uses the conventional particle filter-based data assimilation technique in Reid *et al.* [2], using the IRI as a background model. The ionospheric state is represented as a vector of both $foF2$ and $hmF2$ at each individual ionosonde (Equation 4). This assimilation uses $N = 100$ ensemble members.

$$x = \begin{bmatrix} foF2_{\mathrm{THJ76}} \\ \vdots \\ foF2_{\mathrm{CAJ2M}} \\ hmF2_{\mathrm{THJ76}} \\ \vdots \\ hmF2_{\mathrm{CAJ2M}} \end{bmatrix} \tag{4}$$

2.2.3 The New Data Assimilation

This model works much like the previous data assimilation, but adds two new global parameters. Rather than each particle using the same background model, the IRI using the true IG and Rz indices, instead each particle uses an effective index IG_{eff} and Rz_{eff}. These effective indices are now part of the ionospheric state x (Equation 5). Like the other models, this model uses an ensemble size of $N = 100$ particles.

$$x = \begin{bmatrix} foF2_{\mathrm{THJ76}} \\ \vdots \\ foF2_{\mathrm{CAJ2M}} \\ hmF2_{\mathrm{THJ76}} \\ \vdots \\ hmF2_{\mathrm{CAJ2M}} \\ Rz_{\mathrm{eff}} \\ IG_{\mathrm{eff}} \end{bmatrix} \tag{5}$$

2.2.4 Index-Only Data Assimilation

This data assimilation model has been constructed using only the effective IG and Rz indices as a state vector (Equation 6). This will allow some estimation of how much of the improvement of the new technique comes from the effective indices alone, and how much comes from combining the effective indices with the rest of the ionospheric state. Like the control assimilation, this model uses $N = 100$ ensemble members. There are other data assimilation models which act entirely in the effective parameter space, such as the IRI-UP model [6].

$$x = \begin{bmatrix} Rz_{\mathrm{eff}} \\ IG_{\mathrm{eff}} \end{bmatrix} \tag{6}$$

3 Results

All models were run through the period of June 20th to June 30th, 2024 using all available ionosonde data, excluding those instruments held in reserve for validation.

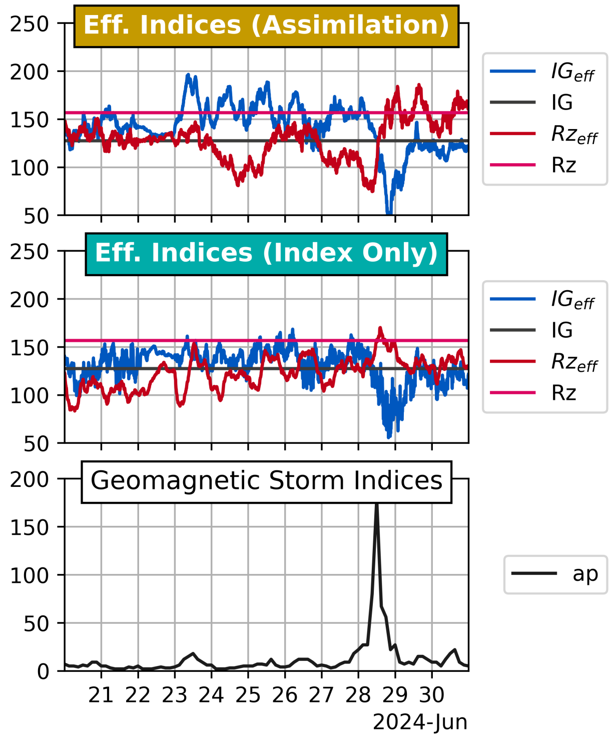

Figure 1. Effective and true IG and Rz indices for both the new data assimilation (Section 2.2.3) and the index-only data assimilation (Section 2.2.4). The bottom plot shows the ap storm index, showing a moderate storm on June 28th, 2024.

3.1 Effective Indices

Figure 1 shows the resulting IG_{eff} and Rz_{eff} produced by both models able to adjust the indices of the IRI. The bottom plot shows the ap index, a measure of geomagnetic storm activity. There was a moderate storm on June 28th, which caused a dramatic shift in both IG_{eff} and Rz_{eff}, for both models. Both models tend to show a pretty similar behaviour in IG_{eff}, with a correlation coefficient of $R = 0.80$. The behaviour of Rz_{eff} is less similar, with a correlation coefficient of only $R = 0.24$. For the index-only assimilation in particular, there is a strong diurnal variation in Rz_{eff}, although all effective indices here seem to show some diurnal variation.

3.2 Ionosonde $foF2$

The overall root mean square error (RMSE) in $foF2$ for each ionosonde and model is shown in Table 1. The ionosondes are ordered by geographic latitude. The reference ionosondes are in bold type. The background colour of each RMSE entry is graded based on the relative performance to the control data assimilation, with red indicating worse performance (greater RMSE) and blue indicating improvement (lower RMSE).

The performance of the IRI depends sharply on latitude, with lower latitude ionosondes generally reporting larger errors than midlatitude or higher ionosondes. This is likely due in part to the higher electron densities at low latitudes, as a similar proportional error would result in a larger absolute error. The traditional data assimilation is able to equilibrate the errors, bringing most well below 0.8 MHz. The new data assimilation shows broad improvement over the traditional method, showing that the use of effective indices has some benefit. The index-only assimilation, meanwhile, shows some improvement over the control assimilation at high latitudes, but suffers poor performance at lower latitudes, in line with the IRI itself.

Table 1. $foF2$ RMSE performance for each ionosonde and model. Rows in bold font are unassimilated ionosondes. Entries are coloured based on their relative RMSE compared to the control data assimilation model.

station	glat	glon	IRI	Control	Assim.	Index
THJ76	76.5	291.6	0.74	0.53	0.63	0.72
TR169	69.6	19.2	0.80	0.77	0.57	0.70
PF765	**65.1**	**212.6**	**0.88**	**0.88**	**0.69**	**0.81**
EI764	64.7	212.9	0.85	0.83	0.64	0.79
GA762	62.4	215.0	0.87	0.85	0.69	0.81
JR055	**54.6**	**13.4**	**0.83**	**0.81**	**0.76**	**0.68**
EA653	52.7	185.9	1.00	0.67	0.70	0.88
FF051	51.7	358.5	0.88	0.55	0.36	0.69
DB049	50.1	4.6	0.74	0.42	0.42	0.59
PQ052	**50.0**	**14.6**	**0.70**	**0.69**	**0.69**	**0.55**
SO148	47.6	16.7	0.83	0.45	0.43	0.73
AL945	45.1	276.4	0.95	0.55	0.50	0.85
MHJ45	42.6	288.5	0.97	0.57	0.48	0.85
EB040	40.8	0.5	1.04	0.67	0.62	0.80
VT139	40.6	17.8	0.87	0.67	0.49	0.74
AT138	**38.0**	**23.5**	**0.94**	**0.90**	**1.08**	**0.90**
WP937	**37.9**	**284.5**	**1.15**	**1.12**	**0.78**	**0.97**
IC437	37.1	127.5	1.01	0.56	0.55	1.00
EA036	37.1	353.3	1.15	0.57	0.64	0.98
EG931	30.5	273.5	1.11	0.60	0.64	0.99
AU930	30.4	262.3	1.31	0.88	0.89	1.31
AW426	26.3	127.8	1.33	0.65	0.49	1.29
LL721	21.4	201.8	1.49	0.57	0.66	1.49
WA619	19.3	166.6	1.17	0.48	0.49	1.22
AS00Q	-8.0	345.6	1.46	0.59	0.48	1.47
CAJ2M	-22.7	315.0	1.28	0.62	0.50	1.33

3.3 Ionosonde $hmF2$

Table 2 shows the hmF2 RMSE for every ionosonde and model in the study. Like the previous table, the background of each entry is coloured according to its performance relative to the control assimilation. Here, the control assimilation is able to eke out a marginal improvement over the IRI in most cases, but the difference is on the order of a few kilometres at most. Notably, the unassimilated ionosondes show no real improvement over the IRI. Both data assimilation models show a strong improvement over both the IRI and the control assimilation, with the index-only assimilation performing best of all.

Table 2. *hmF*2 RMSE performance for each ionosonde and model. Rows in bold font are unassimilated ionosondes. Entries are coloured based on their relative RMSE compared to the control data assimilation model.

station	glat	glon	IRI	Control	Assim.	Index
THJ76	76.5	291.6	46.0	46.4	37.4	34.4
TR169	69.6	19.2	30.0	29.1	27.9	23.6
PF765	**65.1**	**212.6**	**38.9**	**36.2**	**25.8**	**23.1**
EI764	64.7	212.9	38.7	37.0	27.3	25.2
GA762	62.4	215.0	31.7	31.8	27.7	23.4
JR055	**54.6**	**13.4**	**28.1**	**28.7**	**25.5**	**23.7**
EA653	52.7	185.9	38.9	36.9	32.2	25.7
FF051	51.7	358.5	23.0	22.7	23.9	21.2
DB049	50.1	4.6	28.7	28.7	24.3	19.4
PQ052	**50.0**	**14.6**	**29.3**	**29.6**	**24.0**	**19.7**
SO148	47.6	16.7	48.7	38.9	39.5	32.2
AL945	45.1	276.4	39.2	38.7	30.4	26.5
MHJ45	42.6	288.5	32.2	31.7	30.5	26.1
EB040	40.8	0.5	26.5	25.8	27.9	26.5
VT139	40.6	17.8	27.3	26.6	24.5	22.6
AT138	**38.0**	**23.5**	**40.7**	**41.0**	**41.3**	**28.2**
WP937	**37.9**	**284.5**	**35.6**	**35.5**	**25.7**	**23.4**
IC437	37.1	127.5	37.0	36.2	28.8	23.6
EA036	37.1	353.3	27.4	28.1	29.2	29.4
EG931	30.5	273.5	32.3	31.1	28.3	24.2
AU930	30.4	262.3	44.0	42.6	39.8	36.7
AW426	26.3	127.8	35.5	33.6	32.5	30.8
LL721	21.4	201.8	50.8	49.4	42.3	39.4
WA619	19.3	166.6	43.7	42.9	41.9	39.3
BLJ03	**1.4**	**311.6**	**50.9**	**50.5**	**44.0**	**37.2**
AS00Q	-8.0	345.6	42.2	41.7	36.0	32.8
CAJ2M	-22.7	315.0	47.3	46.5	42.5	39.2

4 Discussion and Conclusion

This study sought to examine whether including effective driving indices in the ionospheric state of a data assimilation model provides benefit. As can be seen in Tables 1 and 2, the answer is a definitive yes. There is always a trade-off in data assimilation between adding more dimensions to the state, which allows more phenomena to be captured by the model, and the resulting increase in complexity. In the case of particle filters, each dimension added to the state vector x decreases the sampling efficiency of the ensemble. It is therefore not a foregone conclusion that adding the indices to the state will provide a benefit.

This may be the reason why the performance of the index-only assimilation was superior in *hmF*2. With only two dimensions in the state space, the sampling efficiency of the small $N = 100$ ensemble is excellent. With the high degree of noise in autoscaled *hmF*2, there may not be enough information to constrain the F2 layer height beyond a crude global optimization. The new data assimilation presented here would then perform worse, as it has additional elements of the state to try and fit, despite the lack of information.

The limitations of a global optimization are clearly evident in the $foF2$ results. A single parameter is insufficient to describe the entire global variability in $foF2$, and so the index-only assimilation performs only marginally better than the IRI, and worse than any other data assimilation presented here. The addition of a global parameter to optimize did result in a clear benefit over the control data assimilation. This suggests that global optimization of driving parameters is a worthwhile consideration in the design of future ionospheric data assimilation models.

5 Acknowledgements

This research was undertaken with the support of the Office of Naval Research, BAA number N00014-23-S-B001.
The International Reference Ionosphere (IRI) is available from `irimodel.org/`
This study makes use of data from the Global Ionospheric Radio Observatory (GIRO) data `spase.info/SMWG/Observatory/GIRO`

References

[1] B. W. Reinisch and I. A. Galkin, "Global Ionospheric Radio Observatory (GIRO)," *Earth, Planets and Space*, vol. 63, no. 4, pp. 377–381, Apr. 2011, ISSN: 1880-5981. DOI: 10.5047/eps.2011.03.001.

[2] B. Reid *et al.*, "A-CHAIM: Near-real-time data assimilation of the high latitude ionosphere with a particle filter," *Space Weather*, vol. 21, no. 3, 2023. DOI: 10.1029/2022SW003185.

[3] D. Bilitza *et al.*, "The International Reference Ionosphere model: A review and description of an ionospheric benchmark," en, *Reviews of Geophysics*, vol. 60, no. 4, Oct. 2022. DOI: 10.1029/2022rg000792.

[4] D. R. Themens *et al.*, "Artist ionogram autoscaling confidence scores: Best practices," *Radio Science Letters*, Mar. 2022. DOI: 10.46620/22-0001.

[5] G. Evensen *et al.*, *Data Assimilation Fundamentals*. Springer International Publishing, 2022. DOI: 10.1007/978-3-030-96709-3.

[6] A. Pignalberi *et al.*, "Improvements and validation of the IRI UP method under moderate, strong, and severe geomagnetic storms," *Earth, Planets and Space*, vol. 70, no. 1, p. 180, Nov. 2018, ISSN: 1880-5981. DOI: 10.1186/s40623-018-0952-z.