University of Reading

Analog ensemble forecasts of solar wind parameters: quantification of the predictability and time-domain spectral performance

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Simon, P. A. ORCID: https://orcid.org/0000-0003-2091-2093, Chen, C. H. K., Owens, M. J. ORCID: https://orcid.org/0000-0003-2061-2453 and Sishtla, C. ORCID: https://orcid.org/0000-0003-4236-768X (2025) Analog ensemble forecasts of solar wind parameters: quantification of the predictability and time-domain spectral performance. Space Weather, 23 (7). e2025SW004473. ISSN 1542-7390 doi: 10.1029/2025SW004473 Available at https://centaur.reading.ac.uk/123504/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1029/2025SW004473

Publisher: American Geophysical Union

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other

University of Reading

copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Check for updates

Space Weather

RESEARCH ARTICLE
10.1029/2025SW004473

Analog Ensemble Forecasts of Solar Wind Parameters: Quantification of the Predictability and Time-Domain Spectral Performance

Pauline A. Simon[1], Christopher H. K. Chen[1], Mathew J. Owens[2], and Chaitanya Sishtla[1]

[1]Department of Physics and Astronomy, Queen Mary University of London, London, UK, [2]Department of Meteorology, University of Reading, Reading, UK

Key Points:

- The analog ensemble (AnEn) method can forecast mesoscale structures of solar wind velocity and magnetic field for resolutions higher than 1 hr
- AnEn forecasts can be more time-accurate than persistence, climatology, and synodic recurrence for solar wind velocity and magnetic field
- An ensemble reduction based on the spectra of the AnEn forecasts preserves the forecast of the fluctuations in the reduction process

Correspondence to:

P. A. Simon,
pauline.simon@qmul.ac.uk

Citation:

Simon, P. A., Chen, C. H. K., Owens, M. J., & Sishtla, C. (2025). Analog ensemble forecasts of solar wind parameters: Quantification of the predictability and time-domain spectral performance. *Space Weather*, *23*, e2025SW004473. https://doi.org/10.1029/2025SW004473

Received 3 APR 2025
Accepted 13 JUN 2025

Author Contributions:

Conceptualization: Pauline A. Simon, Christopher H. K. Chen, Mathew J. Owens
Data curation: Pauline A. Simon
Formal analysis: Pauline A. Simon
Funding acquisition: Christopher H. K. Chen
Investigation: Pauline A. Simon, Christopher H. K. Chen
Methodology: Pauline A. Simon, Christopher H. K. Chen, Mathew J. Owens
Project administration: Christopher H. K. Chen
Resources: Pauline A. Simon
Software: Pauline A. Simon
Supervision: Christopher H. K. Chen
Validation: Pauline A. Simon
Visualization: Pauline A. Simon, Christopher H. K. Chen, Chaitanya Sishtla
Writing – original draft: Pauline A. Simon

© 2025. The Author(s).
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Abstract Forecasting multiscale properties of the solar wind is one of the important aspects of space weather prediction as mesoscales, larger than 1 min, can affect the magnetosphere. Amongst forecasting techniques, the analog ensemble (AnEn) method allows the forecast of a quantity from its past behavior, is easy and quick to implement, and results in an ensemble of time series. A comparison of optimal AnEn forecasts of *Wind* spacecraft observations of near-Earth solar wind properties with the persistence and climatology baselines allows a quantification of the predictability of the magnetic and velocity components and magnitude. The AnEn predictions were found to be as accurate as persistence for short-term forecasts and climatology for long-term ones, and performed better than both baselines for more than 60% of the samples for a particular lead time. Furthermore, using an AnEn instead of the baselines enables prediction of the full spectrum of solar wind fluctuations. However, using the standard averaging method to generate a unique forecast from the AnEn ensemble results in a loss of power in the small-scale fluctuations. To prevent this loss, a new spectral reduction method is proposed and compared to the standard averaging method as well as the synodic recurrence baseline. The AnEn spectral-reduced forecast is shown to be more time-accurate than the synodic baseline and more frequency-accurate than the mean-reduced forecasts. Such a reduced forecast is then confirmed to be useful as a comparative baseline in performance diagnostics of space weather models.

Plain Language Summary Forecasting solar wind behavior is crucial for space weather applications, as it impacts the Earth's magnetic environment and our society. The role of fluctuations of the velocity and the magnetic field, of duration larger than 1 min, is important in this impact. Amongst forecasting techniques, the analog ensemble (AnEn) method can forecast a quantity from its past behavior, is easy and quick to implement, and results in an ensemble of time series. Comparing the AnEn forecasts of the solar wind observations at half-minute resolution with alternate forecasts shows the effectiveness of AnEn in forecasting more than half of our data set and estimating temporal performance. However, smaller-scale fluctuations are lost while making an average to reduce the forecast ensemble to one time series. To preserve such fluctuations, we propose a methodology based on the spectral properties of the ensemble and confirm its performance with a novel spectral diagnostic. The new reduced forecast can be useful to replace missing data in solar wind measurements, to inform solar wind models or to forecast the magnetospheric activity. The post-treatment spectral techniques discussed here are also applicable beyond the field of Space Weather.

1. Introduction

The recent geomagnetic storm of 10 May 2024, triggered by a series of interplanetary coronal mass ejections (ICME) (Liu et al., 2024), confirmed that current space weather models fall short on multiple fronts (Hayakawa et al., 2025). Inaccurate predictions during the storm, including poor forecasts of geomagnetic indices (Parker & Linares, 2024), and the post-application of the currently used models to that event—for instance magnetospheric (Tulasi Ram et al., 2024)—highlight shortcomings in our understanding of the formation and acceleration of solar wind plasma, its evolution in the interplanetary space, its interactions with the magnetosphere, and the magnetic structure of ICMEs. Such improvements are critical given our increasing reliance on technology sensitive to space weather disruptions (Eastwood et al., 2017; Oughton et al., 2019).

Space weather forecasting can be viewed as a coupling between different systems. Measurements at the first Lagrangian point (L1), 0.01AU (astronomical unit) upstream of the Earth, are used by inner-magnetospheric and ionospheric models to forecast geomagnetic activity. Thus, lead times of such forecasts are limited to less than an

Writing – review & editing: Pauline A. Simon, Christopher H. K. Chen, Mathew J. Owens, Chaitanya Sishtla

hour as solar wind plasma at L1 reaches Earth in less than an hour. To increase that lead time, L1 measurements need to be forecast. Coronal model output initialized with remote measurements of the solar activity (Owens et al., 2008), and in-situ measurements of sub-L1 solar wind parameters (Lugaz et al., 2025), propagated to L1 by solar wind models can increase the lead time to several days. However, improving the accuracy of the forecasts requires improvements in each step of the modeling pipeline. MacNeice et al. (2018) provides an overview of the state-of-the-art modeling efforts to forecast solar wind conditions at 1AU. The currently obtained errors of L1 forecasts come from various sources: the boundary conditions close to the sun, the background solar wind structures and ongoing interactions (Hinterreiter et al., 2019) or the under-resolved grids used by the physics-based models (Regnault et al., 2024). The physics-based models of the solar wind reflects mostly large scales in accordance to an extensible literature associated to the geomagnetic impact of large scales structures (Dremukhina et al., 2018). However, the multi-scale and non-linear behavior of the solar wind can be observed in the Fourier spectral space and is spread through more than 8 decades of frequencies (Roberts et al., 1987; Sahraoui et al., 2020), from days and hours (scales of ICME and Corotating Interaction Regions, here referred to large scales), to the scales associated to proton and electrons smaller than 10s (here referred to small scales). In particular, mesoscale structures larger than a minute have the right size to affect Earth's environment in a quasi-stationary way (Viall et al., 2021). For instance, the periodic density structures are associated to various spatial mesoscales (Kepko et al., 2020) and can be correlated to enhancements of the magnetic power spectra (Di Matteo et al., 2024). Furthermore, as the coupling between the solar wind and magnetosphere is not linear, with energy loading and abrupt release, an accumulation of small perturbations can generate geomagnetic activity (D'Amicis et al., 2020). Owens et al. (2014) demonstrate the efficiency of a well-fitted noise added to solar wind forecasts in the improvement of a magnetospheric model's responses. A recent study (Ala-Lahti et al., 2024) concludes that 15min and longer fluctuations (mesoscale) are needed to well estimate the starting point of a substorm and capture 40% of the energy transfer, while higher frequency fluctuations (2 to 8min Ultra-Law-Frequency (ULF) waves) contribute up to 15% of the energy transfer. Note that these large and mesoscale (>15min) fluctuations are what is defined as "low frequency" fluctuation by Sibeck and Murphy (2025). One way to introduce mesoscale fluctuations to large-scale L1 forecasts would be to use statistical forecasts based on the history of the solar wind at L1. Here, we would then want to verify if the spectral behavior of solar wind velocity and magnetic field can be well reproduced through such empirical models. To facilitate such a study, we use the Analogue Ensemble model (AnEn) that is fast and easy-to-implement.

AnEn has been introduced in Space Weather by Owens et al. (2017) and Riley et al. (2017), and builds an ensemble of forecasts from historical satellite measurements. The basic principle is that historical observations can provide a good analog, or "similar day," of the current conditions, and of the future. Lorenz (1969) proposed this method for meteorological forecasts and it is one of the paradigms used in neural network models (Burov et al., 2020). It is based on the idea that with a large enough data set, we can find an ensemble of time periods in the past that display similar values and variations to current conditions. It is then assumed that the future evolution will be similar enough to make meaningful predictions. Such meaningful predictions could be used to fill observational data gaps that are pathological in space weather empirical geomagnetic models that couple the solar wind parameters to the geomagnetic indices. Following this aim, Lockwood et al. (2019) analyzed the auto-correlation of various solar wind parameters at 1AU and demonstrated that the reliability of persistence forecasts is inherently limited. Their study showed that while solar wind parameters exhibit high autocorrelation over short timescales (several hours), the correlation decays significantly over longer periods, reducing the accuracy of persistence-based forecasts for extended lead times, that is, how long in the future, the forecast is accurate and can be used. One objective of this paper, is then to compare the AnEn forecasts to persistence, climatology and synodic recurrence baselines, to quantify its performance in terms of lead time accuracy and the associated relative predictability of the solar wind velocity and magnetic fields.

Owens et al. (2017) and Riley et al. (2017) applied the AnEn method to 1-hr near-Earth solar wind observations to forecast solar wind parameters (density, velocity, magnetic field). Haines et al. (2021) also applied this method to forecast geomagnetic indices. Empirically, these ensembles of forecasts contain all of the fluctuation information that has been measured, and hence the information to predict the solar wind Fourier spectra. However, that information can be lost when using bulk statistics of the ensemble (such as mean or median) for visualization, performance diagnostics and prediction purposes. Using spectral diagnostics, this paper tackles the question of the accuracy of this forecast method at higher resolution, the quantification of the loss of information in the fluctuations, and suggests an improved method to preserve the smaller-scale information in the analysis.

15427390, 2025, 7, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025SW004473 by Test, Wiley Online Library on [15/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

10.1029/2025SW004473

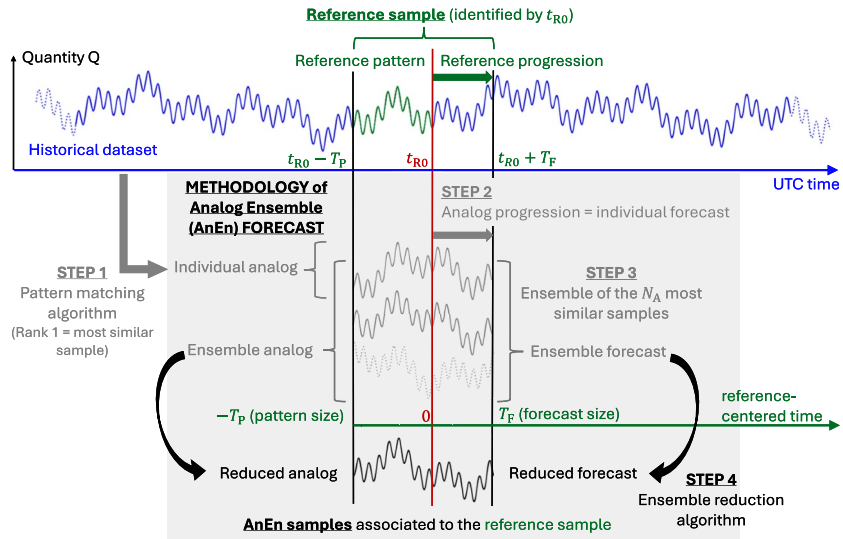

Figure 1. Summary of the Analog Ensemble method used in this article. The historical data set (blue) provides individual analogs (STEP 1, gray, left side of the red line). Ensemble forecasts (STEP 3) are obtained from the progression of these analogs (STEP 2, gray, right side of the red line), and are reduced with STEP 4 (black). More details in Section 2.1.

2. Methods and Data

2.1. The Analog Ensemble (AnEn) Method

The analysis and forecast method follows the Analog Ensemble (AnEn) technique proposed by Owens et al. (2017) and Riley et al. (2017), which is schematically illustrated in Figure 1. Consider a quantity Q with the aim of making a forecast of the progression of Q into the future, from the current, or "reference," time, t_{R0}, to some maximum forecast lead time or "forecast size," T_F.

We begin by defining a "reference pattern" formed from the recent observations of Q, spanning a time window from $t_{R0}-T_P$ to the reference time t_{R0}, where T_P is referred to as the "pattern size." The rest of the data set, that is, prior to $t_{R0}-T_P$ and after t_{R0}, is used as a "historical data set," from which analogs are drawn with the following procedure of pattern-matching. Each data point t_A in this historical data set is associated with an interval of length T_P, from t_A-T_P to t_A, and each of these intervals, or "individual analogs," is compared with the reference pattern. The metric used to quantify the similarity between the reference pattern and the individual analogs is the mean square distance,

$$\mathrm{MSD} = \left\langle \left(Q_{\text{individual analog}} - Q_{\text{reference pattern}}\right)^2 \right\rangle_{T_P}. \quad (1)$$

We also require that individual analogs and reference patterns contain less than 10% of missing values in the computation of the MSD. Minimizing the MSD allows us to attribute a rank to every individual analog from the most to the least similar to the reference pattern. A set of the N_A most similar time series is extracted and forms the "ensemble analog" of rank N_A.

The progression of the time series after the end of an individual analog is the "individual forecast." Thus, an "ensemble forecast" of rank N_A will be obtained from the progression of the individuals in the ensemble analog of rank N_A. While the ensemble forecast can be used as a probabilistic forecast, it is often reduced to a single time series through an ensemble reduction algorithm, such as computing the ensemble mean or median at each time step. The reduction of that ensemble to a single time series (as a prediction of the future behavior) will be called the "reduced forecast" of rank equal to the rank of the least similar individual analog, N_A, considered in the ensemble analog. The duration of these time series will be noted T_F for "forecast size." They will be compared to the progression of the reference pattern, after the reference time t_{R0}, the "reference progression." The set of an analog and the associated forecast will be called the "output sample" while the set of the reference pattern and progression will be the "reference sample."

15427390, 2025, 7, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025SW004473 by Test, Wiley Online Library on [15/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

15427390, 2025, 7, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025SW004473 by Test, Wiley Online Library on [15/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.2. Parameters of the Statistical Analysis

A statistical analysis will be presented on the performance of output samples from $N_R = 200$ reference samples and for several pattern sizes. $N_R = 200$ is chosen for computational reasons. The pattern sizes T_P are chosen from 96s (4 datapoints) to 24h (3600 datapoints). $T_P = 96$s effectively tests if the progression depends on the most recent small-scale fluctuation, while $T_P = 24$h contains the information from the previous days. However, note that the MSD will dilute small-scale fluctuations except for outlier values. An improvement of the forecast performance with a decrease of T_P means that short-term past information is sufficient to model the future. A decrease in the performance of the forecast is expected with an increase of the forecast size T_F, as analogs increasingly diverge from the reference progression due to small initial differences growing with time and to unpredictable changes in the stream structure. Ensembles of up to $N_A = 2000$ individuals will be considered. N_A will also affect the performance. We expect the appearance of unpredictable events (ICME, stream variations, phase differences) in the ensemble forecast to be compensated when N_A is large enough but that also means that progressions from less similar individual analogs will be accounted for.

2.3. Data Set

The *WIND* spacecraft (Acuña et al., 1995) has been the only spacecraft at L1 providing sub-minute resolution measurements of the velocity and the magnetic field, resulting in long-term and high-resolution data sets (Wilson et al., 2021). Its strategic position at L1 has been held since 2004. For computational reasons, the analysis is limited to the years from 2004 to 2009, corresponding to the decrease of activity of solar cycle 23 and the first year of solar cycle 24. Using L1 data to forecast L1 data restricts the analysis to solar wind historical occurrences in similar states of evolution, as its non-linear behavior evolves during its propagation in the interplanetary space (Chen et al., 2020). The *WIND PLSP* data set (Lin et al., 2021) provides ground computation of proton velocity amplitude and components in the geocentric solar ecliptic (GSE) coordinate system, resolved at 24s, based on measures of the 3DP (Three-Dimensional Plasma and Energetic Particle Investigation, Lin et al., 1995) instrument, and synchronized with magnetic field amplitude and component measurements from Magnetic Field Instrument (MFI, Lepping et al., 1995) instrument, also in the GSE coordinate system. 24s-resolution provides most of the mesoscale or inertial range where no dispersion and dissipation occurs according to turbulence theories. Due to large gaps in 2007 and 2008, 3 sub-datasets are considered—from 3 May 2004 to 21 November 2006, from 1 October 2007 to 30 April 2008, from 1 June 2008 to 31 August 2009—in which anomalous values found in the velocity have been removed after comparison with *WIND* Solar Wind Experiment (SWE, Ogilvie et al., 1995) measurements resolved at 94s. One can note a slight shift in value affecting the velocity estimations of the PLSP data set by comparison with the SWE one. As this shift varies slowly over multiple months that potentially affects the ranking of the analogs. This bias, due to the onboard calibration process, makes the analysis more realistic for a future application of the AnEn method for operational forecasting. The proportion of invalid values is estimated at $\sim$10% for the first data subset, $\sim$27% for the second subset and $\sim$35% for the third subset. However, most gaps have a duration less than 10 points (4 min).

495 potential reference times t_{R0} were selected carefully such that the error due to datagaps on the pattern matching algorithm (and hence on the performance diagnostics) are reduced and the ensemble is as representative as possible of the solar wind activity of 2004 and 2005. The requirements are:

- no more than 5% of missing values in patterns of several sizes T_P log-regularly chosen between 96s (4 datapoints) and 24h (3600 datapoints),
- no gaps larger than 4% of the T_P,
- no gaps in the progression larger than 4% of $T_F = 3600$ points,
- at least 12 h-interval between two consecutive t_{R0}.

Then $N_R = 200$ reference times t_{R0} are randomly drawn from this ensemble of potential t_{R0}. They are spread from 4 May 2004 03:26:24 to 24 December 2005 18:07:36.

2.4. Metrics and Other Forecasts Used to Diagnose the Performances of the AnEn Forecasts

The performance of the AnEn forecast is computed by comparison to the observed variations (reference progression) and to several traditional forecasts. These are: a climatology baseline, wherein the future value of Q is forecast to always be the historical average value, $\langle Q \rangle_t$; a persistence baseline, wherein the future value of Q is

forecast to be equal to the value at the reference time, $Q(t = t_{R0})$; and a synodic cycle recurrence forecast, where the future value of Q is forecast to be equal to that one solar synodic rotation period (t_{syn}) earlier, $Q(t - t_{syn})$. We use $t_{syn} = 27.125$days, following Owens et al. (2013) that observed that the correlation peak is closer to 27.125days for the velocity and the magnetic field main components and hence this value gives better forecasts than the synodic period of 27.27days.

To estimate the agreement between forecasts and observations, a number of metrics are considered. The first, associated to time-domain accuracy, is the Normalized Root-Mean-Square error:

$$\mathrm{NRMSE} = \sqrt{\frac{\left\langle \left(Q_{\text{forecast}} - Q_{\text{reference progression}}\right)^2 \right\rangle_t}{\left\langle Q^2_{\text{reference progression}} \right\rangle_t}}. \tag{2}$$

The lower the NRMSE, the better the forecast performance. By computing the AnEn NRMSE relative to that of a baseline forecast, it is possible to compute the AnEn skill:

$$\mathrm{Skill} = 1 - \frac{\mathrm{NRMSE(AnEn\ forecast)}}{\mathrm{NRMSE(baseline)}} \tag{3}$$

that quantified the performance of an AnEn forecast relative to a baseline. If $\mathrm{Skill} > 0$, NRMSE(AnEn forecast) < NRMSE(baseline), and the performance of the AnEn forecast is better than the performance of the baseline.

The last metric is a spectral ratio that quantifies the scale-by-scale performance in the Fourier space:

$$\mathrm{SR}(f) = \log_{10}\left(\frac{\|\mathrm{FFT}(Q_{\text{forecast}})\|}{\|\mathrm{FFT}(Q_{\text{reference progression}})\|}\right). \tag{4}$$

It is fitted linearly from frequencies above 10^{-4}Hz, assuming the frequency dependency is logarithmic:

$$\mathrm{SR}(f) \sim \mathrm{f_{SR}} = \gamma_{\mathrm{SR}}\left(\log_{10} f + 2\right) + \delta_{\mathrm{SR}}. \tag{5}$$

Hence, it is possible to reduce the analysis of spectral performance at small scales to two parameters: the slope, γ_{SR}, and the estimated spectral ratio at $f = 10^{-2}$Hz, δ_{SR}. The sign and the value of δ_{SR} indicate the gain or loss of power of the small-scale fluctuations. γ_{SR} indicates if this gain or loss is constant for all frequencies above 10^{-4}Hz or if the amplitude of the spectrum of $Q_{\text{output sample}}$ diverges from the one of $Q_{\text{reference sample}}$. The value 10^{-4}Hz is empirically chosen from the observed behavior of the particular case introduced in Figure 6 and to give the same importance to all small-scale frequencies in the fit, SR(f) is binned log-regularly in frequency.

3. Results

The forecast of a particular reference time t_{R0}, as a case study, is analyzed in Sections 3.1 and 3.2. The choice of algorithms for the reduction of the ensemble analogs, statistically compared in Section 3.3, is established from these results. They also illustrate the relationship between the ensemble analog paradigm and the baselines that will be used to define the predictability in Section 3.4.

3.1. Case Study: 6 October 2004 14:58:00 UT and Ensemble-Reduction Algorithm

Figure 3 is an example of forecasts for $t_{R0} = 6$ October 2004 14:58:00 UT with a pattern of size $T_P = 24$h and two forecast sizes, $T_F = 24$h (panel a) and $T_F = 1000$h $\sim$ 42days (panel b). The individual analogs (gray) obtained from applying STEP 1 and 2 of Figure 1 and the baseline forecasts, see Figure 2, are displayed. Panel a shows that while the individual analogs (gray) follow the reference pattern (green) quite closely, the individual forecasts (gray) tend to diverge from the reference progression (green) around 20,000 s from $t = 0$. This is around the same time as the persistence forecast (red dash-dotted) diverges from the reference progression. On longer time scales shown on panel b, both the reference progressions and individual forecasts vary around the climatological value

15427390, 2025, 7, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025SW004473 by Test, Wiley Online Library on [15/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

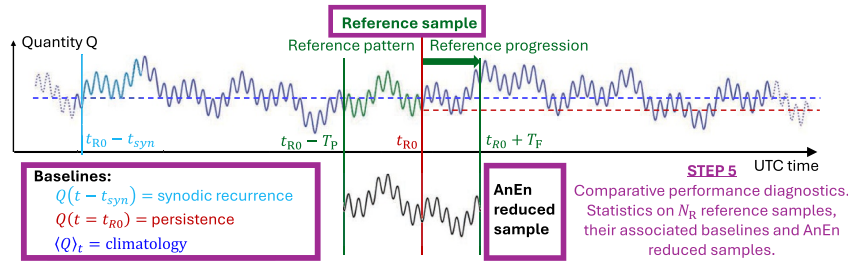

Figure 2. Summary of the progressions compared to quantify the performance of the Analog Ensemble forecasts. More details in Section 2.4.

(blue dash-dotted). This diverging behavior from a similar initial state is expected in a complex system following the rules of deterministic chaos (Yao & Tong, 1995), but on the longer timescales will result from changing conditions in the global solar wind structure too. A quasi-cyclic variation, or mean-reversibility, around the climatological value is also expected as the large-scale behavior of the solar wind follows multiple cycles. However, the synodic period of 27 days does not seem associated with this mean-reversibility. The first convergence to the climatology baseline seems to coincide with the first main period observed by Oloketuyi et al. (2020), a period of 5–8 days, which was found in *Wind* data during the time period of the present study using wavelet analysis. This periodicity is displayed with a gray vertical area (panel b). For this interval, the previous synodic period recurrence is not even in the first 500 analogs, too dissimilar to the reference pattern. However, long-term analysis reveals that it seems to reflect the main variations of the reference quite well.

Figure 4 shows the long-term, $T_F = 1000\text{h} \sim 42\text{days}$, time series (panel a) and spectra (panel b) of individual (gray) and synodic recurrence (cyan) samples along side reduced AnEn samples (STEP 4 of Figure 1). Table 1 summarizes the time-accuracy NRMSE and Skill metrics of the baselines and the AnEn reduced analogs and forecasts. The first ensemble reduction algorithm applied is the traditional mean reduction (black dashed). The amplitude of the spectra of the individual samples and the previous synodic recurrence follow the reference one. However, while the AnEn mean-reduced forecast (black dashed line) seems to be quite accurate in terms of

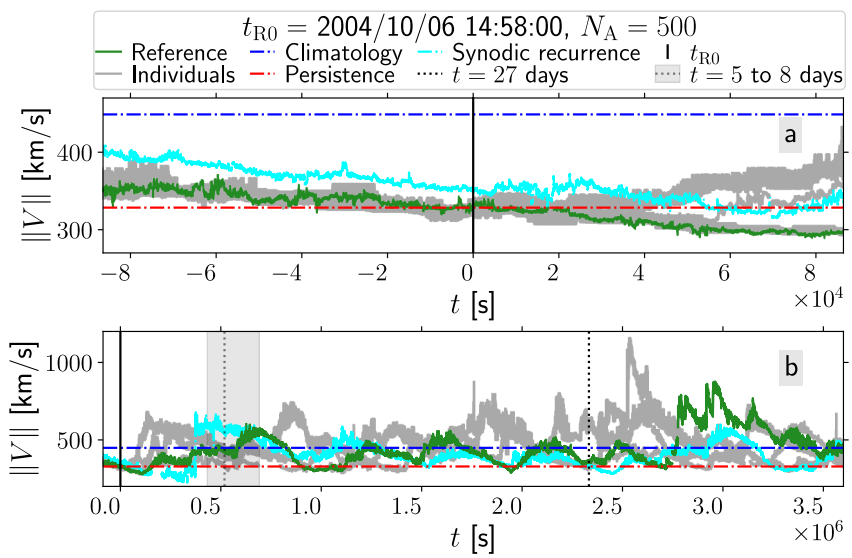

Figure 3. Forecasts of proton velocity magnitude for the reference time t_{R0} = 6 October 2004 14:58:00 UT (black vertical line). Short-term progression of size $T_F = 24\text{h}$ (a) and long-term progression $T_F = 42\text{days}$ (b) of the reference sample (green solid), $N_A = 500$ AnEn individual samples (gray solid) obtain with a pattern size $T_P = 24$ h, the persistence (red dash-dotted), the climatology (blue dash-dotted), and the synodic rotation period recurrence (cyan dash-dotted). $t_{R0} + 27$ days (vertical dotted black) is the synodic rotation period, $t_{R0} + 5$–8 days (dotted vertical gray line and area) is the period observed by Oloketuyi et al. (2020).

15427390, 2025, 7, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025SW004473 by Test, Wiley Online Library on [15/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

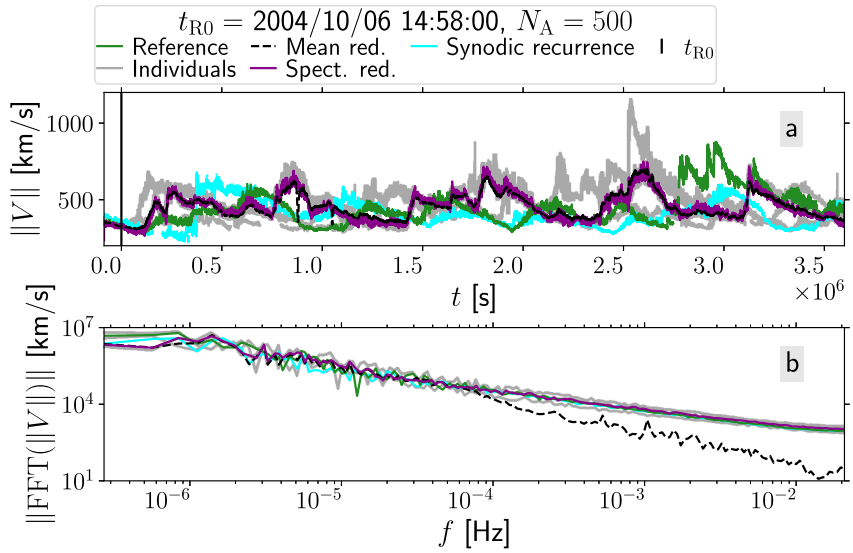

Figure 4. Forecasts of proton velocity magnitude for t_{R0} = 6 October 2004 14:58:00 UT (black vertical line). Long-term progression T_F = 42days (a) and spectra (b) of the reference sample (green solid), N_A = 500 analog ensemble individual (gray solid) obtain with a pattern size T_P = 24 h, mean-reduced (dashed black dashed) and spectral-reduced (purple solid) samples, and the synodic rotation period recurrence (cyan dash-dotted).

NRMSE by comparison with the baselines, the spectrum reveals a systematic loss of small-scale fluctuations (here shown for the interval $[-T_P; T_F]$ with T_P = 24h). Note that the frequency-by-frequency geometric average of the amplitude of the spectra of the individual samples (purple solid line) provides a closer level to the reference.

Table 1
NRMSE *Performances and Skill for the AnEn Reduced Samples and Baselines of* t_{R0} = *6 October 2004 14:58:00 UT With* N_A = 500

Intervals	$[-T_P; 0]$	$[0; T_P]$	$[0; T_F]$
NRMSE			
Climatology	0.3151	0.4556	0.2385
Persistence	0.0470	0.0756	0.3428
Synodic recurrence	0.0991	0.1065	0.2490
Mean red.	0.0162	0.0357	0.3090
Spect. red. to T_P	0.0199	0.0288	N/A
Spect. red. to T_F	0.0305	0.0289	0.3073
Skill over climatology			
Mean red.	0.9487	0.9217	−0.2956
Spect. red. to T_P	0.9370	0.9367	N/A
Spect. red. to T_F	0.9031	0.9367	−0.2883
Skill over persistence			
Mean red.	0.6557	0.5281	0.0985
Spect. red. to T_P	0.5772	0.6184	N/A
Spect. red. to T_F	0.3495	0.6182	0.1036
Skill over synodic recurrence			
Mean red.	0.8370	0.6652	−0.2410
Spect. red. to T_P	0.7997	0.7293	N/A
Spect. red. to T_F	0.6919	0.7292	−0.2340

15427390, 2025, 7, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025SW004473 by Test, Wiley Online Library on [15/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

10.1029/2025SW004473

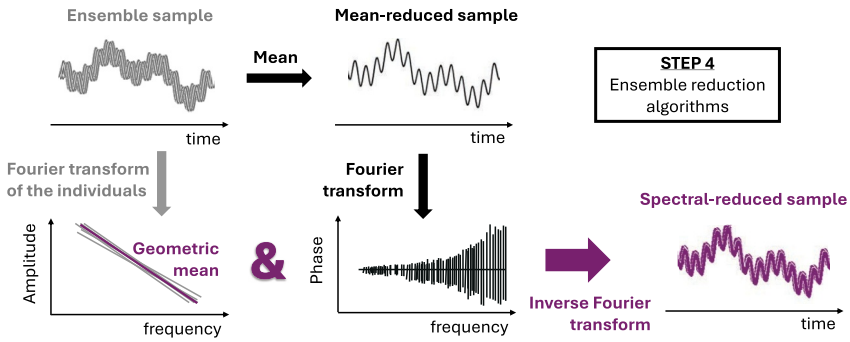

Figure 5. Summary of the reduction algorithm used to obtain the analog ensemble reduced forecast in STEP 4 of Figure 1. More details in Section 3.1.

Thus, we define a new reduction algorithm, here called "spectral reduction" (Spect. red.). The reduced time series is produced following the algorithm of Figure 5. It uses the inverse Fourier transform of a composite spectrum whose amplitude is the geometric average of the amplitude of the spectra of the individual sample (purple spectrum), and the phase is the phase φ_{FFT} of the mean-reduced sample:

$$\text{Spec. red.} = \text{invFFT}\left(\sqrt[N_A]{\prod_{N_A} \left\| \text{FFT}(Q_{\text{individual sample}}) \right\|} \cdot e^{i\varphi_{\text{FFT}}\left(\langle Q_{\text{individual sample}} \rangle_{N_A}\right)} \right). \tag{6}$$

The resulting time series is displayed in purple and the NRMSE and Skill measures associated are in the table (rows Spect. red.). The spectral reduction algorithm is based on the spectra of the whole samples instead of just the forecasts or the analogs. That ensures the coincidence at the reference time t_{R0} as edge effects are induced by the inverse Fourier transform of the composite spectra. In the following sections, the spectra used to do the reduction are computed on time series of forecast size equal to the analyzed interval. Indeed, for an interval $]0; T_{\mathrm{P}}]$, the reduction will be applied on samples of size $[-T_{\mathrm{P}}; T_{\mathrm{P}}]$ even if using a larger sample does seem to only affect slightly the Skill performance in Table 1 (comparison between rows Spect. red. and column $[0; T_{\mathrm{P}}]$).

3.2. Quantification of the Small Scale Performance With the Case Study: 6 October 2004 14:58:00

Figure 6 illustrates the quantification of the spectral performances with the spectral ratio SR applied on short-term (panel a) and long-term (panel b) forecasts. A value of 0 corresponds to the ideal case when the forecasted spectral amplitude is the same as the reference, a value greater than 0 indicates an overestimation, and a value less than 0 indicates an underestimation (equivalent to a loss of power). At all spectral scales for short-term progression and large scales for long-term progression, the AnEn mean reduction (black), the spectral reduction (purple) and the synodic recurrence (cyan) give similar results. These forecasts fall within the span of the minimum and maximum values of the spectral amplitude of the ensemble forecast (gray area). For all forecast times, this span is centered on the 0-value, meaning that the ensemble forecast reflects well the spectral properties of the reference progression.

However, the AnEn mean-reduced forecast diverges to smaller values of the spectral ratio SR at small scales, slightly for short-term progression and more obviously for long-term. This divergence is quantified in Table 2 with the slope γ_{SR} and the value at 10^{-2} Hz δ_{SR} of the small-scale fits. It happens for frequencies higher than 10^{-4} Hz, with a break-point between 10^{-5} Hz and 10^{-4} Hz. This is also approximately the value at which the AnEn begins to diverge from the persistence forecast. This could be a signature of phase shifts between individual samples at small scales, whereas similar phase shifts at the largest scales do not result in a loss of power in the associated fluctuations. As illustrated by Figure 3, most individual samples are consecutive shifts (small-scale shifts whose maximum is defined by the correlation time $\tau_{\mathrm{c}} = 1.2 \times 10^5$s, a diagnostic of the auto-similarity of the time series) of the higher-ranked one and have superposed spectra. Considering a large AnEn size N_{A} seems to have effectively randomized the ensemble distribution of smaller-scale fluctuations, and flatten their amplitude through the mean reduction, while the spectral reduction preserves them. The gray area, larger for short-term progression, reveals a statistical convergence of the spectra of the individual forecasts—statistics on the

15427390, 2025, 7, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025SW004473 by Test, Wiley Online Library on [15/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

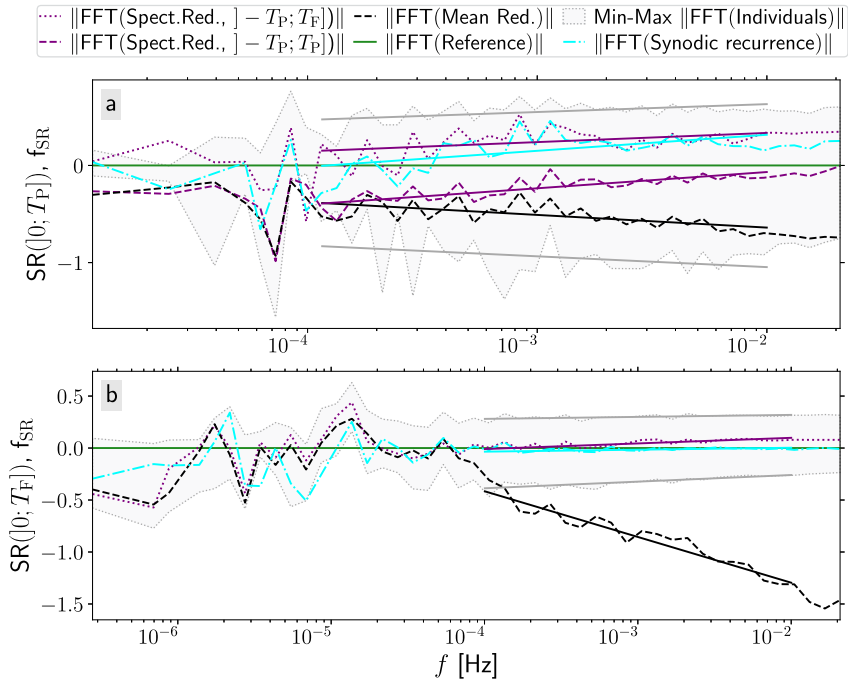

Figure 6. Spectral ratio of short-term (panel a) and long-term (panel b) forecasts and respective small-scale fits (straight lines). Green: 0-value, the threshold between over- and under-estimation. Cyan: spectral ratio SR applied to the spectral amplitude of the synodic recurrence forecast. Gray: range between the minimum and the maximum spectra obtained from the analog ensemble (AnEn) ensemble forecast. Black: SR applied on the spectra of the AnEn mean-reduced forecast. Purple: SR applied on the spectra of the spectral-reduced forecasts obtained for AnEn individual samples of size $[-T_P; T_P]$ (dashed) or $[-T_P; T_F]$ (dotted). The results are binned on log-regular 50 frequencies for readability. Example of the proton velocity magnitude.

duration of the time series. In spectral and other turbulence statistical investigation, this duration is usually compromised with the number of individuals (that here would be equivalent to N_A, and the purple long-term results) (Isaacs et al., 2015).

3.3. Effect of the Rank of the Ensemble Forecast on Statistical Performances Over 200 Reference Samples

The forecast performance is now assessed by computing median values of the time-accuracy metric NRMSE, and the frequency-accuracy metrics δ_{SR}, and γ_{SR}. The metrics are applied on AnEn reduced forecasts computed for three forecast sizes and five pattern sizes. In order to limit computational cost, the AnEn forecasts are estimated at $N_R = 200$ reference samples. Figure 7 summarizes this study according to the rank of the ensemble forecast, N_A that is, the rank of similarity of the less similar analog used to obtain the ensemble, from 1 to 2,000, for the proton velocity magnitude (similar behaviors have been found for the magnetic field magnitude and their components).

Table 2
Slope, γ_{SR}, and Values at 10^{-2}Hz, δ_{SR}, Associated to the Spectral Ratio Fits f_{SR} of Figure 6

	SR([0, T_P])		SR([0, T_F])	
Type	δ_{SR}	γ_{SR}	δ_{SR}	γ_{SR}
Mean red.	−0.643	−0.132	−1.294	−0.453
Spect. red. to T_P	−0.069	0.166	N/A	N/A
Spect. red. to T_F	0.363	0.111	0.101	0.055
Min_{N_A} ||FFT(individuals)||	−1.022	−0.087	−0.248	−0.071
Max_{N_A} ||FFT(individuals)||	0.628	0.078	0.321	−0.020

Note. Short-term [0, T_P] and long-term [0, T_F] AnEn forecasts of t_{R0} = 6 October 2004 14:58:00 with $N_A = 500$ individual forecasts, a pattern size $T_P = 24$ h and a maximal forecast size $T_F = 42$days.

For the NRMSE (panels a1–a3), mean-reduced forecast performance is, on average, better than spectral-reduced. So the large fluctuations brought by the spectral correction seem to worsen this accuracy metric. As expected, the accuracy increases with the size of the ensemble; a larger ensemble size means a smoothing of extreme events (coronal mass ejection, etc.). Even with $N_A = 2,000$ individuals in the ensembles, we are still not able to include sufficiently dissimilar and varying individuals in the AnEn ensembles due to the high resolution of the data set limiting the computational load. The largest pattern size and smaller forecast size make an exception: as the values closer to the reference time t_{R0} are lost in the MSD, chosen as recognition criteria, the AnEn ensembles can include more dissimilar individuals. For the forecast

15427390, 2025, 7, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025SW004473 by Test, Wiley Online Library on [15/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

10.1029/2025SW004473

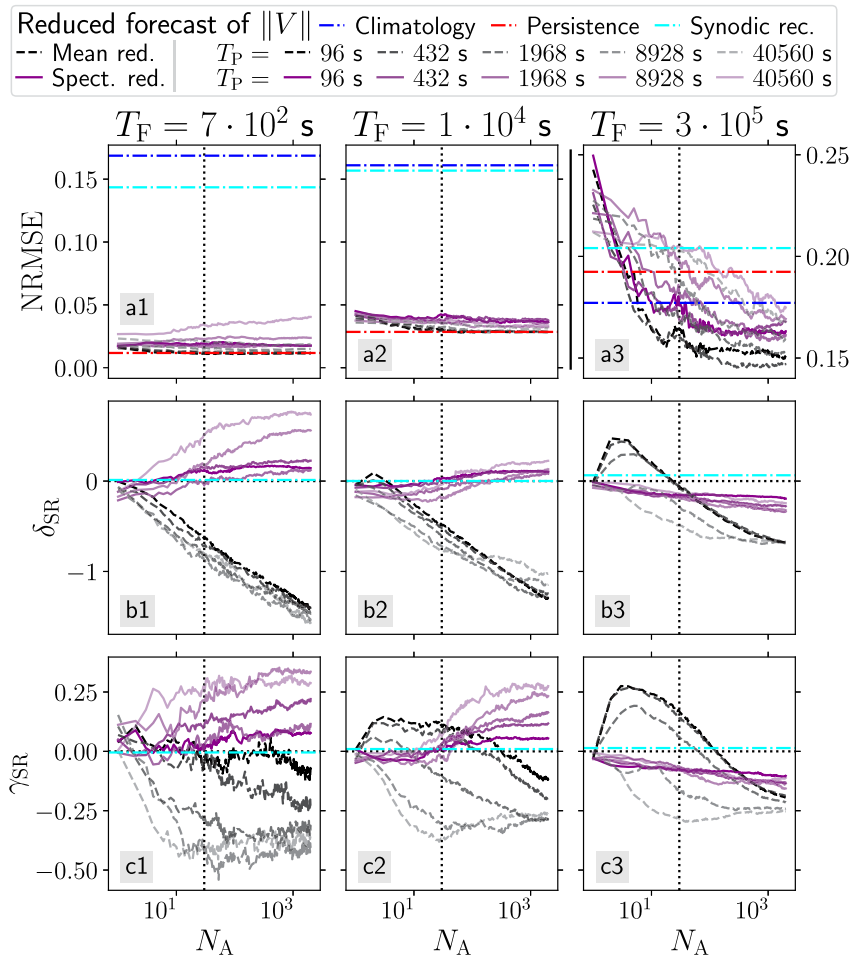

Figure 7. Summary of statistical performances (median on $N_R = 200$ reference times) of AnEn mean-reduced (dashed black) and spectral-reduced (purple) forecasts, according to the forecast duration T_F (columns), the pattern size used for the forecasts T_P (lightness) and the size of the forecast ensemble N_A (horizontal axis). Example for proton velocity magnitude. Panels (a1–a3): the time-accuracy NRMSE better if minimal. Panels (b1–b3): the high frequency level δ_{SR} better close to 0 (dotted black horizontal line). Panels (c1–c3): the slope γ_{SR} better close to 0 (dotted black horizontal line). Dotted vertical line: value $N_A = 30$ selected for Section 3.4. Dash-dotted lines: baseline forecasts.

sizes $T_F = 7 \times 10^2$ s and 1×10^4 s, the performances are close to that of persistence, while for 3×10^5 s they are better than all other models (a more rigorous study according to T_F is proposed in Section 3.4).

δ_{SR} (panels b1–b3) confirms a general underestimation of small-scale power for the mean-reduced forecast, except for small ensemble size N_A and small forecast size T_P, most likely due to a lack of statistics in the ensemble forecasts to effectively smooth the small-scale fluctuations (similarly to the $T_F = T_P$ case of Section 3.1). The slope γ_{SR} (panels c1–c3) is mostly negative too, indicating a divergence of the AnEn mean-reduced forecast spectral amplitude from the spectra of the reference progression. For small pattern sizes, γ_{SR} can even be positive (divergence is the high-frequency value is positive/overestimation, convergence in the other scenario) but still close to zero, indicating that the spectra are more or less parallel. δ_{SR} and γ_{SR} of the spectral-reduced forecasts closer to 0 than the mean-reduced ones prove the effectiveness of the spectral reduction. However, a divergence from zero is still observed with the increase of N_A.

Consequently, the choice of an optimal ensemble size N_A has to be a balance between absolute accuracy (low NRMSE) and minimizing loss of fluctuation power (values of γ_{SR} and δ_{SR} close to zero), whatever the quantity and the choice of reduction method. $N_A = 30$ seems to be a good compromise for all quantities. Riley et al. (2017) chose $N_A = 50$, while Owens et al. (2017) chose $N_A = 100$, compromising the accuracy with numerical

15427390, 2025, 7, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025SW004473 by Test, Wiley Online Library on [15/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

15427390, 2025, 7, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025SW004473 by Test, Wiley Online Library on [15/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 3
Summary of Predictability and Performances of Analog Ensemble (AnEn) Mean- and Spectral-Reduced Forecasts and Synodic Recurrence for the Velocity and the Magnetic Fields

	$\|\|V\|\|$	V_x	V_y	V_z	$\|\|B\|\|$	B_x	B_y	B_z
Π [%]								
Mean red.	66	65	61	64	58	66	63	63
Spect. red.	59	58	56	52	47	56	54	53
T_L [s]								
Mean red.	3.8×10^5	3.8×10^5	3.9×10^4	3.3×10^4	1.7×10^5	1.3×10^5	2.5×10^5	1.4×10^4
Spect. red.	3.1×10^5	3.1×10^5	3.3×10^4	3.9×10^4	1.7×10^5	1.7×10^5	2.5×10^5	0.7×10^4
τ_c [s]	1.2×10^5	1.2×10^5	1.4×10^4	1.5×10^4	0.4×10^5	0.4×10^5	0.3×10^5	0.4×10^4
λ_c [km]	6.3×10^7	6.3×10^7	6.8×10^6	7.4×10^6	2.2×10^7	2.3×10^7	1.7×10^7	2.1×10^6
NRMSE								
Mean red.	0.152	0.153	0.747	0.940	0.317	0.934	0.957	0.918
Spect. red.	0.165	0.159	0.804	1.017	0.358	0.965	1.020	0.963
Synodic rec.	0.199	0.199	1.197	1.348	0.461	1.211	1.149	1.390
δ_{SR}								
Mean red.	−0.048	−0.053	−0.485	−0.513	−0.220	−0.457	−0.484	−0.484
Spect. red.	−0.145	−0.137	−0.163	−0.163	−0.118	−0.201	−0.209	−0.210
Synodic rec.	0.041	0.040	−0.011	−0.002	0.002	0.023	0.016	0.020

Note. AnEn best predictability Π and associated forecast size, that is, optimal lead time T_L, for mean-reduced and spectral-reduced forecasts of the magnitude and components of the velocity and the magnetic fields. The lead times correspond to several autocorrelation times τ_c (associated with autocorrelation lengths λ_c). Spectral reduction offers a compromise between time series accuracy (median of NRMSE at $T_F = T_L$ better than the one of the synodic recurrence baseline) and the loss of high-frequency fluctuations (median of δ_{SR} at $T_F = T_L$ better than one of mean-reduced forecasts, except for $\|V\|$ and V_x in accordance with Figure 7). Values obtained for $N_A = 30$ individuals, the pattern size $T_P = 192$s and statistics over $N_R = 200$ reference times.

computation reasons. Such values would also work quite well for our data set resolved at 24s instead of 1h according to Figure 7.

3.4. Predictability and Statistical Performances for the Velocity and Magnetic Field

One crucial question is the predictability of future solar wind evolution. Two quantities are of interest: how many intervals of the solar wind can be predicted? and how long in the future can we forecast? To investigate this, we use AnEn reduced forecasts of rank $N_A = 30$ obtained for a pattern size $T_P = 192$s and associated with $N_R = 200$ references, parameters chosen from the results of Section 3.3. Section 3.3 reveals that small values of T_P are better, and similar performance for $T_P = 96$s and $T_P = 432$s. For these parameters, the maximal dissimilarity of the considered analogs $\left(\sqrt{\text{MSD}}_{N_A=30}\right)$ is statistically 1km/s for the magnitude and the components of the velocity and 0.1nT for the magnetic field. The predictability, Π, is defined as the proportion of reference samples better forecast by AnEn than the baseline. It is contained in the positive Skill diagnostic depending on the ratio between a reduced forecast and the baseline forecast. $\Pi > 50\%$ means that the reduced forecast performs better for more than half of reference samples, a sign of a good predictability relative to the NRMSE diagnostic. The forecast size associated to the best predictability, T_L, corresponds to the optimal lead time, characteristic of the progression duration such that the AnEn is the optimal choice of forecast model (relative to persistence and climatology). The optimal lead time for the persistence baseline for instance, is the correlation time τ_c and its spatial equivalent the correlation length λ_c (estimations are provided in Table 3).

The NRMSE for each forecast type (panels a and b) and the proportion of positive skill Π (panel c) for different forecasts are displayed in Figure 8 for the proton velocity magnitude. The performance of the AnEn mean-reduced forecasts follows the persistence baseline for the forecast sizes $T_F < 2 \times 10^5$ s and then converges to the

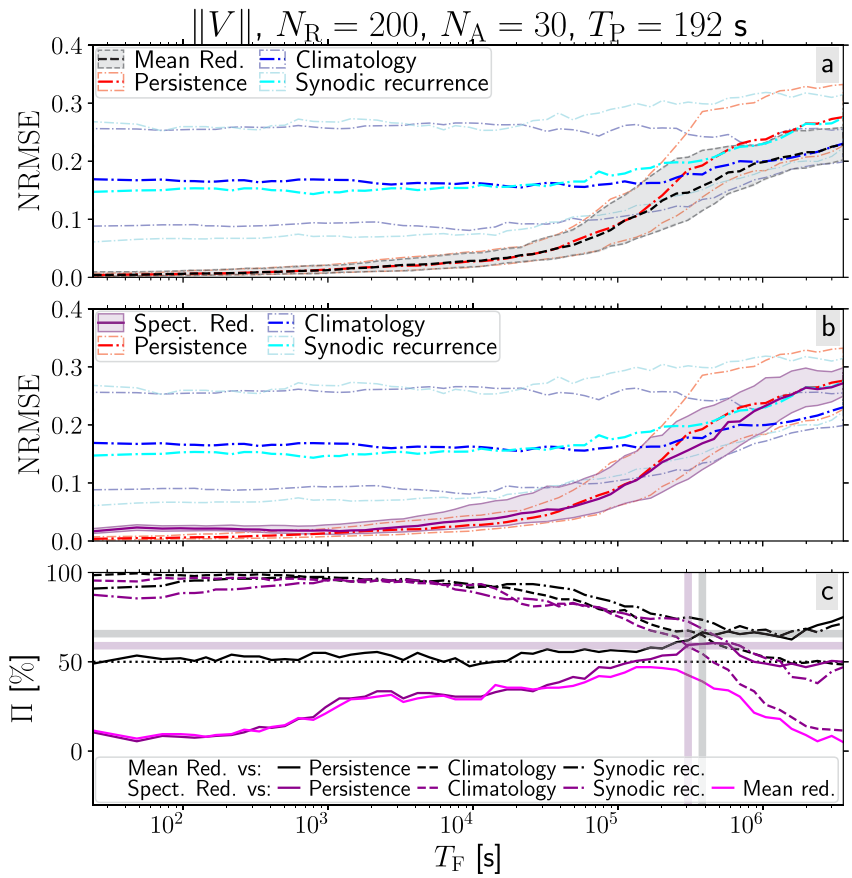

Figure 8. Performance of mean-reduced (black, panels a and c) and spectral-reduced (purple, panels b and c) forecasts of proton velocity magnitude as a function of the size of the forecast, T_F. Results are shown for $N_A = 30$ individuals, $N_R = 200$ reference times and a pattern size $T_P = 192$s. Panels (a, b): Median and quartiles (lighter) of NRMSE. Comparison with the persistence (red dash-dotted), the climatology (blue dash-dotted), and the synodic recurrence (cyan dash-dotted). Panel c: Proportion of positive Skill relative to persistence (straight), to climatology (dashed), and to synodic recurrence (dash-dotted), and of spectral-reduced relative to mean-reduced forecasts (magenta). The crosses indicate the predictability $\Pi = {}_{N_R}(\text{Skill} > 0)$ where $\Pi_{\text{persistence}} = \Pi_{\text{climatology}}$ and the associated optimal lead time T_L for mean-reduced (gray) and spectral-reduced (pink) forecasts.

climatology. This behavior has been observed by Owens et al. (2017). The spectral reduced forecasts diverge slightly from the AnEn mean reduced forecast at small and large T_F. At large T_F, it seems to converge to the performance of either the synodic cycle forecast or persistence. However, the latter behavior varies according to the investigated quantity. As expected, the distribution of the NRMSE accuracy of the N_R AnEn reduced forecasts, revealed by the inter-quartiles ranges surrounding the median in Figures 8a and 8b, is converging at small T_F similarly to the persistence inter-quartile. It diverges slightly at larger T_F due to the divergence of the forecast from the reference progression. The latter is within the climatology inter-quartile range. The synodic cycle forecast for the velocity magnitude is similar to climatology according to the median of the NRMSE for small T_F, however the quartiles indicate that the distribution is very broad and slightly larger for the synodic cycle. Depending on the historical data set and not minimizing the error with the reference last point value, they are less relevant than the persistence that is based on the last value of the pattern reference. For Skill (panel c), the mean reduced and the spectral reduced forecasts can perform better than the three baselines with optimal values Π and T_L referenced in Table 3.

The table of Figure 3 summarizes the predictability Π and the optimal lead time T_L for every quantity. The magnetic field and velocities are on average predictable at 60% with AnEn, with slightly better results for AnEn mean-reduced forecasts. The main difference between the quantities is the optimal lead time that corresponds to several autocorrelation time, τ_c. This optimal lead time is similar for both reduction algorithms. τ_c has been

15427390, 2025, 7, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025SW004473 by Test, Wiley Online Library on [15/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

15427390, 2025, 7, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025SW004473 by Test, Wiley Online Library on [15/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

computed on chunks smaller than 22days over the whole the historical data set, and correspond to the smallest scale such that the normalized autocorrelation has dropped below $1/e$. For $||V||$, V_x, $||B||$, B_x, and B_y, the lead time is around 2–3 days, while V_y, V_z, and B_z give the worst results of lead times, around several hours. As observed previously, the best performance of AnEn happened when the persistence performance are decreasing in favor of the climatology. As V_y, V_z and B_z are not the main components of their field and contain primarily fluctuations, their last pattern value is not expected to last long in the future. Indeed, V_x contains most of the magnitude as the flow is predominantly radial direction, while **B** is mainly in the ecliptic plane (x, y) but not radial direction due to the Parker spiral. These discrepancies agree with auto-correlations studies implying that the values of B_z persist less than the velocity and, hence, that this quantity is harder to predict (Lockwood et al., 2019; Wicks et al., 2010). The estimated values of correlation length, λ_c, obtained by multiplying the estimation of τ_c associated to each chunk by its average proton velocity magnitude, are 10 times above the estimation of Wicks et al. (2010) done for 2004 to 2009 with L1 observations resolved at 1min. However, such estimations vary heavily according to the size of the considered chunks (closer values have been obtained by using accordingly 1day intervals instead of 22days).

The time-accuracy NRMSE and frequency-accuracy δ_{SR} applied on the AnEn mean-reduced and spectral-reduced forecasts and the synodic recurrence with $T_F = T_L$ are also provided in Table 3. For all quantities, the spectral reduction provides time accuracy slightly worse than the mean reduction but not as much as the synodic recurrence. For V_y, V_z, $||B||$, B_x, B_y, and B_z, the frequency accuracy is improved relative to the mean-reduced forecasts. $||V||$ and V_x provide the opposite behavior as the optimal lead time for the chosen parameters, $N_A = 30$ individuals and the pattern size $T_P = 192$s, is close to the T_F investigated in Figure 7b3. For all quantities, the chosen parameters do not provide a better frequency accuracy than the synodic recurrence. However, these values provide insight on how the AnEn spectral-reduced forecast compromised the time and frequency accuracies.

4. Discussion and Conclusion

We have presented a study of the analog ensemble (AnEn) method (which is based on pattern matching of past behavior of a system to make predictions about its future), with a new emphasis on its ability to capture the multi-scale nature of the solar wind accurately. It is one of the few space weather modeling tools that can provide ensemble forecasts capturing and predicting the solar wind mesoscale variations. These scales are important for estimating the geomagnetic response to solar activity (Ala-Lahti et al., 2024; Owens et al., 2014).

In this study, we have demonstrated several techniques that can be applied for performance estimation and operational forecasts. The performance of the AnEn method with varying ensemble size and forecast duration is quantified using a high-resolution data set (24s). A spectral performance diagnostic is introduced to investigate the small-scale statistical accuracy of the forecasts. A new way for reducing the ensemble of forecasts to a future prediction of the time series is introduced through a spectral reduction method that (unlike previous methods) preserves small-scale fluctuation amplitudes.

The AnEn method is found to provide forecasts of performance close to the persistence at short-term progression and the climatology at long-term, but better than both at intermediate progression lengths. It can also, according to the desired aspect (for instance, for short-term forecast and time-accuracy), be significantly better than the synodic cycle forecast. Statistically, the AnEn provides better results in about 60% of the reference intervals of the velocity and the magnetic field-related quantities than persistence and climatology. That enhanced predictability is mostly reached for lead times associated with transitions from persistence-like to climatology-reversible durations. For the main components of the velocity (Vx) and the magnetic field (Bx and By), the lead time is around 2–3 days, and around several hours for the other components agreeing with correlation-time analyses done in L1. While large scales and the short-term are better forecasted with AnEn, the synodic recurrence gives a more accurate forecast of small-scale variations in the Fourier space. However, AnEn spectral reduced forecasts provide a better compromise between time and frequency accuracy than the mean-reduced forecasts, maximizing the time-accuracy, or the synodic recurrence, maximizing the frequency accuracy.

In addition, there are some caveats to consider regarding this study. The first concerns data gaps or missing values, which can affect the reduction algorithm output. For the mean reduction, the missing values are omitted. However, before applying the Fourier transform needed by the spectral reduction, the missing values in the

15427390, 2025, 7, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025SW004473 by Test, Wiley Online Library on [15/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

samples must be linearly interpolated. This results in the discrepancies between the spectral reduced and mean reduced forecast performances of Figure 7 at small ensemble size N_A. In the ideal case of no missing values, the performance would be the same for $N_A = 1$. They converge for spectral performances as the diagnostic also interpolates the missing values of the mean reduced forecasts. Data gaps also affect the pattern matching algorithm. To mitigate their effect on the forecast performance analysis, several precautions have been applied, as described in Section 2. However, missing values are the reality of real-time forecasts. So a rigorous investigation of their effects on the AnEn forecasts would be needed to support a proof of concept of this method for operational applications. Such an analysis has been proposed for persistence and auto-correlation by Lockwood et al. (2019).

The second caveat is the historical data set. Here, around four years are used, spanning the transition from the solar cycle 23 to the solar cycle 24. Thus, we test the usefulness of the AnEn methodology during low solar activity period. A comparison with high solar activity period, such as the cycle 25 that provided the 10 May 2024 geomagnetic storm is left for future studies. Operational applications would use all available historical data set, and potentially pattern matching over multiple quantities and/or spacecraft. An example of pattern-matching algorithm using multiple quantities is given by Owens et al. (2017). Other biases associated with the historical data set come from its quality, in terms of calibration. However, not trying to compensate the miscalibrations in the PLSP data set that have surely affected the pattern matching ranking is more realistic for operational purposes: what forecast can we provide with the data at our disposition?

The final caveat is about the use of the mean over the median. The MSD criterion is fast, but it loses significance for large pattern sizes, as more potentially irrelevant information is considered, and it is sensible to extreme values due to random localized events or instrumental nonphysical fluctuations. Owens et al. (2017) and Riley et al. (2017) used a median reduction algorithm to build the reduced forecast.

To conclude, the AnEn methodology can provide forecasts that include small-scale fluctuations, which could enhance solar wind physics-based models or data assimilation methods (Lang et al., 2021) by providing well-forecasted data to assimilate. It also provides more accurate upstream conditions to inject into magnetospheric response models, for instance, complementing large scale physics-based forecasts through downscaling (Owens et al., 2014). This work investigates the balance between the time- and frequency-domain accuracy associated with the AnEn reduction methods and the interpretation of such ensemble forecasts.

Data Availability Statement

The original spacecraft observation dataset published by NASA Space Physics Data Facility and the Space Sciences Laboratory has been accessed by the Coordinated Data Analysis Web (CDAWeb) through the Python module cdasws (Lin et al., 2021). Figures were made with Matplotlib version 3.9.1 (Caswell et al., 2021; Hunter, 2007), available under the Matplotlib license at https://matplotlib.org/. The Python software and Jupyter notebooks used to obtain and analyze the AnEn forecasts are available on Simon (2025b). The datasets, forecasts and analysis data are available on Simon (2025a). Schematics have been built with Microsoft 365 PowerPoint.

Acknowledgments

P. A. Simon, C. H. K. Chen, and C. Sishtla are supported by UKRI Future Leaders Fellowship MR/W007657/1. C. H. K. Chen is also supported by STFC Consolidated Grant ST/X000974/1. M.J. Owens is part-funded by Science and Technology Facilities Council (STFC) Grant ST/V000497/1 and Natural Environment Research Council (NERC) Grant NE/Y001052/1. P. A. Simon acknowledges useful comments and suggestions from the space plasma group at Queen Mary University of London. We would like to thank the NASA Space Physics Data Facility and the Space Sciences Laboratory that provide access to the *Wind* dataset.

References

Acuña, M. H., Ogilvie, K. W., Baker, D. N., Curtis, S. A., Fairfield, D. H., & Mish, W. H. (1995). The global geospace science program and its investigations. *Space Science Reviews*, *71*(1), 5–21. https://doi.org/10.1007/BF00751323

Ala-Lahti, M., Pulkkinen, T. I., Brenner, A., Keebler, T., Al Shidi, Q., Hill, S., & Welling, D. (2024). The impact of solar wind magnetic field fluctuations on the magnetospheric energetics. *Geophysical Research Letters*, *51*(24), e2024GL112922. https://doi.org/10.1029/2024GL112922

Burov, D., Giannakis, D., Manohar, K., & Stuart, A. (2020). Kernel analog forecasting: Multiscale test problems. *arXiv*. Retrieved from http://arxiv.org/abs/2005.06623(arXiv:2005.06623[physics,stat])

Caswell, T. A., Droettboom, M., Lee, A., Sales De Andrade, E., Hoffmann, T., Hunter, J., et al. (2021). matplotlib/matplotlib: Rel: V3.5.0. *Zenodo*. https://doi.org/10.5281/zenodo.5706396

Chen, C. H. K., Bale, S. D., Bonnell, J. W., Borovikov, D., Bowen, T. A., Burgess, D., et al. (2020). The evolution and role of solar wind turbulence in the inner Heliosphere. *Astrophysical Journal*, *246*(2), 53. https://doi.org/10.3847/1538-4365/ab60a3

D'Amicis, R., Telloni, D., & Bruno, R. (2020). The effect of solar-wind turbulence on magnetospheric activity. *Frontiers in Physiology*, *8*. https://doi.org/10.3389/fphy.2020.604857

Di Matteo, S., Katsavrias, C., Kepko, L., & Viall, N. M. (2024). Azimuthal size scales of solar wind periodic density structures. *Astrophysical Journal*, *969*(1), 67. https://doi.org/10.3847/1538-4357/ad479e

Dremukhina, L. A., Lodkina, I. G., & Yermolaev, Y. I. (2018). Relationship between the parameters of various solar wind types and geomagnetic activity indices. *Cosmic Research*, *56*(6), 426–433. https://doi.org/10.1134/S0010952518060011

Eastwood, J. P., Biffis, E., Hapgood, M. A., Green, L., Bisi, M. M., Bentley, R. D., et al. (2017). The economic impact of space weather: Where do we stand? *Risk Analysis*, *37*(2), 206–218. https://doi.org/10.1111/risa.12765

Haines, C., Owens, M. J., Barnard, L., Lockwood, M., Ruffenach, A., Boykin, K., & McGranaghan, R. (2021). Forecasting occurrence and intensity of geomagnetic activity with pattern-matching approaches. *Space Weather*, *19*(6), e2020SW002624. https://doi.org/10.1029/2020SW002624

Hayakawa, H., Ebihara, Y., Mishev, A., Koldobskiy, S., Kusano, K., Bechet, S., et al. (2025). The solar and geomagnetic storms in 2024 May: A flash data report. *Astrophysical Journal*, *979*(1), 49. https://doi.org/10.3847/1538-4357/ad9335

Hinterreiter, J., Magdalenic, J., Temmer, M., Verbeke, C., Jebaraj, I. C., Samara, E., et al. (2019). Assessing the performance of EUHFORIA modeling the background solar wind. *Solar Physics*, *294*(12), 170. https://doi.org/10.1007/s11207-019-1558-8

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Isaacs, J. J., Tessein, J. A., & Matthaeus, W. H. (2015). Systematic averaging interval effects on solar wind statistics. *Journal of Geophysical Research: Space Physics*, *120*(2), 868–879. https://doi.org/10.1002/2014JA020661

Kepko, L., Viall, N. M., & Wolfinger, K. (2020). Inherent length scales of periodic mesoscale density structures in the solar wind over two solar cycles. *JGR Space Physics*, *125*(8), e2020JA028037. https://doi.org/10.1029/2020JA028037

Lang, M., Witherington, J., Turner, H., Owens, M. J., & Riley, P. (2021). Improving solar wind forecasting using data assimilation. *Space Weather*, *19*(7), e2020SW002698. https://doi.org/10.1029/2020SW002698

Lepping, R. P., Acũna, M. H., Burlaga, L. F., Farrell, W. M., Slavin, J. A., Schatten, K. H., et al. (1995). The WIND magnetic field investigation. *Space Science Reviews*, *71*(1), 207–229. https://doi.org/10.1007/BF00751330

Lin, R. P., Anderson, K. A., Ashford, S., Carlson, C., Curtis, D., Ergun, R., et al. (1995). A three-dimensional plasma and energetic particle investigation for the wind spacecraft. *Space Science Reviews*, *71*(1), 125–153. https://doi.org/10.1007/BF00751328

Lin, R. P., Bale, S. D., Wilson, L. B., & Schroeder, P. (2021). *Wind 3DP 24-sec ion OmniDirectional fluxes and moments*. NASA Space Physics Data Facility; Space Sciences Laboratory, University of California Berkeley. https://doi.org/10.48322/C4MF-9N52

Liu, Y. D., Hu, H., Zhao, X., Chen, C., & Wang, R. (2024). A pileup of coronal mass ejections produced the largest geomagnetic storm in two decades. *Astrophysical Journal Letters*, *974*(1), L8. https://doi.org/10.3847/2041-8213/ad7ba4

Lockwood, M., Bentley, S. N., Owens, M. J., Barnard, L. A., Scott, C. J., Watt, C. E., & Allanson, O. (2019). The development of a space climatology: 1. Solar wind magnetosphere coupling as a function of timescale and the effect of data gaps. *Space Weather*, *17*(1), 133–156. https://doi.org/10.1029/2018SW001856

Lorenz, E. N. (1969). Atmospheric predictability as revealed by naturally occurring analogues. *Journal of the Atmospheric Sciences*, *26*(4), 636–646. https://doi.org/10.1175/1520-0469(1969)26⟨636:APARBN⟩2.0.CO;2

Lugaz, N., Al-Haddad, N., Zhuang, B., Möstl, C., Davies, E. E., Farrugia, C. J., et al. (2025). The need for a sub-L1 space weather Research mission: Current knowledge gaps on coronal mass ejections. *Space Weather*, *23*(2), e2024SW004189. https://doi.org/10.1029/2024SW004189

MacNeice, P., Jian, L. K., Antiochos, S. K., Arge, C. N., Bussy-Virat, C. D., DeRosa, M. L., et al. (2018). Assessing the quality of models of the ambient solar wind. *Space Weather*, *16*(11), 1644–1667. https://doi.org/10.1029/2018SW002040

Ogilvie, K. W., Chornay, D. J., Fritzenreiter, R. J., Hunsaker, F., Keller, J., Lobell, J., et al. (1995). SWE, a comprehensive plasma instrument for the WIND spacecraft. *Space Science Reviews*, *71*(1), 55–77. https://doi.org/10.1007/BF00751326

Oloketuyi, J., Liu, Y., Amanambu, A. C., & Zhao, M. (2020). Responses and periodic variations of cosmic ray intensity and solar wind speed to sunspot numbers. *Advances in Astronomy*, *2020*(1), 3527570-10. https://doi.org/10.1155/2020/3527570

Oughton, E. J., Hapgood, M., Richardson, G. S., Beggan, C. D., Thomson, A. W. P., Gibbs, M., et al. (2019). A risk assessment framework for the socioeconomic impacts of electricity transmission infrastructure failure due to space weather: An application to the United Kingdom. *Risk Analysis*, *39*(5), 1022–1043. https://doi.org/10.1111/risa.13229

Owens, M. J., Challen, R., Methven, J., Henley, E., & Jackson, D. R. (2013). A 27 day persistence model of near-Earth solar wind conditions: A long lead-time forecast and a benchmark for dynamical models. *Space Weather*, *11*(5), 225–236. https://doi.org/10.1002/swe.20040

Owens, M. J., Horbury, T. S., Wicks, R. T., McGregor, S. L., Savani, N. P., & Xiong, M. (2014). Ensemble downscaling in coupled solar wind-magnetosphere modeling for space weather forecasting. *Space Weather*, *12*(6), 395–405. https://doi.org/10.1002/2014SW001064

Owens, M. J., Riley, P., & Horbury, T. S. (2017). Probabilistic solar wind and geomagnetic forecasting using an Analogue ensemble or "similar Day" Approach. *Solar Physics*, *292*(5), 69. https://doi.org/10.1007/s11207-017-1090-7

Owens, M. J., Spence, H. E., McGregor, S., Hughes, W. J., Quinn, J. M., Arge, C. N., et al. (2008). Metrics for solar wind prediction models: Comparison of empirical, hybrid, and physics-based schemes with 8 years of L1 observations. *Space Weather*, *6*(8), S0-8001. https://doi.org/10.1029/2007SW000380

Parker, W. E., & Linares, R. (2024). Satellite Drag analysis during the May 2024 Gannon geomagnetic storm. *Journal of Spacecraft and Rockets*, *61*(5), 1412–1416. https://doi.org/10.2514/1.A36164

Regnault, F., Al-Haddad, N., Lugaz, N., Farrugia, C. J., Yu, W., Zhuang, B., & Davies, E. E. (2024). Discrepancies in the properties of a coronal mass ejection on scales of 0.03 au as revealed by simultaneous measurements at solar orbiter and wind: The 2021 November 3–5 event. *ApJ*, *962*(2), 190. https://doi.org/10.3847/1538-4357/ad1883

Riley, P., Ben-Nun, M., Linker, J. A., Owens, M. J., & Horbury, T. S. (2017). Forecasting the properties of the solar wind using simple pattern recognition. *Space Weather*, *15*(3), 526–540. https://doi.org/10.1002/2016SW001589

Roberts, D. A., Klein, L. W., Goldstein, M. L., & Matthaeus, W. H. (1987). The nature and evolution of magnetohydrodynamic fluctuations in the solar wind: Voyager observations. *Journal of Geophysical Research*, *92*(A10), 11021–11040. https://doi.org/10.1029/JA092iA10p11021

Sahraoui, F., Hadid, L., & Huang, S. (2020). Magnetohydrodynamic and kinetic scale turbulence in the near-Earth space plasmas: A (short) biased review. *Reviews of Modern Plasma Physics*, *4*(1), 4. https://doi.org/10.1007/s41614-020-0040-2

Sibeck, D. G., & Murphy, K. R. (2025). Quantifying the effects of solar wind fluctuations on the solar wind-magnetosphere interaction. *Geophysical Research Letters*, *52*(6), e2025GL114954. https://doi.org/10.1029/2025GL114954

Simon, P. (2025a). Analog ensemble forecast and performance analysis applied on WIND PLSP [Dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.15169310

Simon, P. (2025b). Analog ensemble forecast and performance analysis [workflow]. *Zenodo*. https://doi.org/10.5281/zenodo.15169244

Tulasi Ram, S., Veenadhari, B., Dimri, A. P., Bulusu, J., Bagiya, M., Gurubaran, S., et al. (2024). Super-intense geomagnetic storm on 10–11 May 2024: Possible mechanisms and impacts. *Space Weather*, *22*(12), e2024SW004126. https://doi.org/10.1029/2024SW004126

Viall, N. M., DeForest, C. E., & Kepko, L. (2021). Mesoscale structure in the solar wind. *Frontiers in Astronomy and Space*, *8*, 735034. https://doi.org/10.3389/fspas.2021.735034

15427390, 2025, 7, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025SW004473 by Test, Wiley Online Library on [15/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Wicks, R. T., Owens, M. J., & Horbury, T. S. (2010). The variation of solar wind correlation lengths over three solar cycles. *Solar Physics*, *262*(1), 191–198. https://doi.org/10.1007/s11207-010-9509-4

Wilson, L. B., Brosius, A. L., Gopalswamy, N., Nieves-Chinchilla, T., Szabo, A., Hurley, K., et al. (2021). A quarter century of wind spacecraft discoveries. *Reviews of Geophysics*, *59*(2), e2020RG000714. https://doi.org/10.1029/2020rg000714

Yao, Q., & Tong, H. (1995). On initial-condition sensitivity and prediction in nonlinear stochastic systems. *Bulletin of the International Statistical Institute*, *50*(4), 395–412. Retrieved from http://isi.cbs.nl/bulletin.htm

15427390, 2025, 7, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025SW004473 by Test, Wiley Online Library on [15/07/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License