Comparison of Empirical and Deep Learning Models for Solar Wind Speed Prediction

Seungwoo Ahn[1] , Jihyeon Son[2] , Yong-Jae Moon[1,3] , and Hyun-Jin Jeong[4,1]
[1] School of Space Research, Kyung Hee University, Yongin, 17104, Republic of Korea; moonyj@khu.ac.kr
[2] Center for Heliophysics Research, Division of Fundamental Astronomy & Space Science, Korea Astronomy and Space Science Institute, Daejeon, 34055, Republic of Korea
[3] Department of Astronomy and Space Science, Kyung Hee University, Yongin, 17104, Republic of Korea
[4] Centre for mathematical Plasma Astrophysics, Department of Mathematics, KU Leuven, Celestijnenlaan 200B, 3001 Leuven, Belgium
Received 2025 April 2; revised 2025 May 28; accepted 2025 June 2; published 2025 July 8

Abstract

In this study, we compare representative empirical models with a deep learning model for predicting solar wind speed at 1 au. The empirical models are the Wang–Sheeley–Arge–ENLIL model, which combines empirical methods with a magnetohydrodynamic model, and the empirical solar wind forecast model, which uses the relationship between the fractional coronal hole area and solar wind speed. Our deep learning model predicts solar wind speed over 3 days ahead using extreme-ultraviolet images and up to 5 days of solar wind speed before the prediction date. We evaluate the models over the test period (October–December in each year from 2012 to 2020) in view of solar activity phases and the entire period. To validate the model's performance, we use two evaluation methods: a statistical approach and an event-based approach. For statistical verification during the entire period, our model outperforms the other empirical models, with a much lower mean absolute error of $51.4 \, \mathrm{km \, s^{-1}}$ and rms error of $68.6 \, \mathrm{km \, s^{-1}}$, along with a much higher correlation coefficient of 0.69. For the event-based verification for high-speed solar wind streams, our model has superior performance in most of the six metrics evaluated within a ± 1 day time window. In particular, it achieves a high success ratio of 0.82, emphasizing the model's stable performance and ability to minimize false alarms. These results show that our deep learning model has strong potential for practical application as a reliable tool for fast solar wind forecasting with its high accuracy and stability.

Unified Astronomy Thesaurus concepts: Solar wind (1534); Neural networks (1933); Space weather (2037)

1. Introduction

Space weather is a field that studies the phenomena in which energy and particles emitted from the Sun interact with the Earth, affecting modern technological infrastructure such as satellites, communication systems, and power grids. Plasma and magnetic fields emitted by the Sun interact with Earth's magnetosphere, triggering phenomena such as substorms and geomagnetic storms. One of the main causes of these phenomena is the solar wind, a plasma flow emitted from the Sun. It is classified into fast solar wind and slow solar wind on the basis of its speed. The fast solar wind is well known to originate from coronal holes (CHs) with open magnetic fields (D. McComas et al. 2000) and to travel at speeds exceeding $650 \, \mathrm{km \, s^{-1}}$. In contrast, the origin of the slow solar wind, which travels at speeds below $400 \, \mathrm{km \, s^{-1}}$ (S. Bravo & G. A. Stewart 1997), remains an outstanding question (L. Abbo et al. 2016). The interaction caused by the speed difference between these types of solar wind forms a region known as the corotating interaction region (CIR). A CIR is where plasma density increases, magnetic fields are compressed, and plasma is accelerated, affecting the Earth's magnetosphere. As a result, its structure is altered, which affects space-based technologies such as satellites. Therefore, accurate prediction of the solar wind's arrival time at Earth remains a critical challenge in space weather forecasting.

To address this challenge, empirical models have been proposed to predict solar wind speed using several approaches.

These models rely on empirical formulae derived from observational data. Furthermore, comparative studies have been conducted to evaluate the performance of various solar wind speed forecasting models, thus identifying the most effective approach (P. Riley et al. 2015; M. A. Reiss et al. 2016). Representative examples include the Wang–Sheeley–Arge (WSA)–ENLIL model (Y.-M. Wang & N. Sheeley 1990a, 1990b, 1992; C. Arge & V. Pizzo 2000), operated by the National Oceanic and Atmospheric Administration/Space Weather Prediction Center, and the emperical solar wind forecast (ESWF) model (D. Milošić et al. 2023), operated by the European Space Agency (ESA)/Heliospheric Weather Expert Service Centre (H-ESC), both of which have outperformed other models in the above comparative studies.

Recently, various deep learning (DL)-based models have been developed in the field of space weather. For example, models that use the convolutional layer and Long-Short-Term Memory (LSTM; V. Upendran et al. 2020), the two-dimensional attention mechanism (Y. Sun et al. 2021), and the convolution neural network (H. Raju & S. Das 2021) offer new possibilities to predict solar wind speed. In addition, models using attention mechanisms (E. J. Brown et al. 2022) have been introduced, further expanding the range of techniques used in space weather forecasting. In particular, J. Son et al. (2023) have developed a DL model for solar wind speed prediction. This model combines convolution layers and dense layers in its architecture using extreme-ultraviolet (EUV) solar images and solar wind speed data from the previous 5 days to predict solar wind speed for up to 3 days ahead.

In this study, we make an extensive comparison of our DL model (J. Son et al. 2023) for solar wind speed forecasting

with the following focuses. First, we consider two representative empirical models (WSA–ENLIL and ESWF 3.2) for comparison. Second, our analysis covers a broad range of Carrington rotation (CR) periods: October to December in each year from 2012 to 2020. Third, we perform statistical verification as well as event-based verification to evaluate the model's predictive performance for high-speed solar wind streams (HSS). This paper is organized as follows. In Section 2, we introduce the empirical models and the DL model used for comparison and describe the data preprocessing. In Section 3, we detail two verification methods. In Section 4, we present the verification results. Finally, in Section 5, we summarize the main findings of this study.

2. Forecasting Models

To evaluate the performance of the DL model for solar wind speed prediction, we compare it with two empirical models: the WSA–ENLIL and ESWF models. The WSA–ENLIL model is provided by NASA/Community Coordinated Modeling Center (CCMC), and the ESWF model is provided by the ESA/H-ESC. To compare the performance of the DL model with those of empirical models, empirical model results are processed to match the DL model's 6 hr cadence (00:00, 06:00, 12:00, and 18:00). The comparison is conducted over the test period of October–December for the years 2012–2020. In addition, data corresponding to interplanetary coronal mass ejection (ICME) periods are excluded from the analysis. These periods, as listed by I. G. Richardson & H. V. Cane (2010), represent the intervals from when a shock reaches Earth to the expected end of the corresponding ICME. Information about ICME periods is taken from the Richardson and Cane ICME catalog.[5] In this section, we provide an overview of both the empirical models and the DL model.

2.1. Wang–Sheeley–Arge (WSA)–ENLIL Model

The WSA–ENLIL model is a well-known solar wind speed prediction model that combines the empirical WSA model and the physics-based ENLIL model into a hybrid structure. The WSA model estimates solar wind speed at 30 solar radii ($R_\odot$) using the expansion factor and the CH boundary distance of the solar coronal magnetic field. To derive these parameters, we combine the potential-field source surface model (M. D. Altschuler & G. Newkirk 1969; K. H. Schatten et al. 1969), which extrapolates the coronal magnetic field up to $2.5\,R_\odot$ from Global Oscillation Network Group (GONG) magnetograms (National Solar Observatory, GONG; http://gong.nso.edu/), with the Schatten current sheet model (K. H. Schatten 1971), which extends them up to $30\,R_\odot$. These estimated solar wind speeds are extended into the heliosphere using the ENLIL model, which simulates their propagation through interplanetary space to predict solar wind speed. The ENLIL model is a 3D time-dependent heliospheric model based on ideal magnetohydrodynamic equations, designed to simulate the physical properties of the solar wind. It is often combined with WSA or magnetohydrodynamic algorithm outside a sphere (MAS) coronal models (e.g., D. Odstrcil et al. 2002, 2004; P. Riley et al. 2002) to simulate the propagation and structural evolution of the solar wind. Applying this hybrid approach, the WSA–ENLIL model is a

large-scale physics-based heliospheric prediction system that provides 1–4 days advance warnings of solar wind structures and Earth-directed coronal mass ejections (CMEs) that cause geomagnetic storms.

In this study, the results of the WSA–ENLIL model are obtained from CORHEL-MAS_WSA_ENLIL[6] of NASA/CCMC. We use GONG synoptic maps as input data because, as shown in a previous study (L. K. Jian et al. 2015), they outperform other data sets. For the CR periods corresponding to the verification period, we select the WSA model as the coronal model and the ENLIL model as the heliospheric model. The default configuration is used for this study, which is provided by CORHEL-MAS_WSA_ENLIL. The WSA–ENLIL model provides solar wind speed data at Earth from the CR start time to 5 days beyond the CR end time. However, the data for the official start time do not account for the delay in solar wind propagation, which is approximately 3–5 days from the Sun to Earth. Therefore, to account for this delay, we adjust the WSA–ENLIL data to begin 3 days after the CR start time and extend 3 days beyond the CR end time (L. K. Jian et al. 2015). This 3 days delay setting enables an accurate evaluation of the prediction performance specifically for HSS events. For verification, we use model output at a 6 hr cadence by selecting the values corresponding to the times closest to 00:00, 06:00, 12:00, and 18:00.

2.2. Empirical Solar Wind Forecast Model

The ESWF model predicts the solar wind speed at 1 au by calculating the fractional CH areas within a meridional slice (S. Robbins et al. 2006; B. Vršnak et al. 2007) from Solar Dynamics Observatory/Atmospheric Imaging Assembly (SDO/AIA) EUV images at 193 Å, based on the empirical relationship between these areas and the measured solar wind speed (D. Milošić et al. 2023). Specifically, it uses an empirical formula to perform predictions while dynamically considering changes in CH areas (T. Rotter et al. 2015). Therefore, the propagation delay depends on the change in the fractional CH area. The ESWF model has been updated to ESWF 3.2 (D. Milošić et al. 2023), and this latest version demonstrates improvements in various aspects compared to previous versions. In the ESWF 3.2 model, the pileup algorithm is adopted to better capture the asymmetric profiles of HSS events. This approach considers interactions and the effects of compression and rarefaction occurring during the propagation of the solar wind between the Sun and Earth, effectively considering the asymmetry in velocity profiles. Additionally, the model uses dynamic thresholding to optimally extract CH areas. Specifically, instead of using a fixed threshold intensity as a constant percentage of the median intensity of the solar disk, a dynamic thresholding method based on statistical analysis is introduced. Moreover, to consider the relationship between CH areas and the maximum solar wind speed observed at Earth, the model divides the meridional slice into 12 latitudinal sectors and assigns colatitude-based weights to the CH area in each sector when calculating the solar wind speed. These improvements enable a more accurate prediction of the speed variations associated with HSS events.

In this study, the results of the ESWF 3.2 model are obtained from ESA H-ESC. This model provides solar wind speed data at 1 au for the period 2012–2020.

2.3. Deep Learning Model

In this study, to evaluate the performance of the solar wind prediction, we use a DL model developed by J. Son et al. (2023). This model predicts solar wind speeds up to 3 days ahead with a 6 hr cadence. The input data consist of 211 Å and 193 Å images from the SDO/AIA along with OMNI solar wind speed data measured at 1 au. Additionally, the input data include up to 5 days, which corresponds to the longest traveling time from the Sun to Earth, of data prior to the prediction time. The architecture of the DL model is designed to process and concatenate two types of data. EUV images are processed through a convolutional block, named Inception, followed by an LSTM layer that captures long-term dependences. Meanwhile, solar wind speed data are processed through dense layers to make them suitable for prediction. These two processed data types are then concatenated, and additional dense layers are applied to predict solar wind speeds up to 3 days ahead.

The data set for the DL model spans from 2010 May to 2020 December and is divided into three subsets: a training set (January–August), a validation set (September), and a test set (October–December), taking into account the solar cycle. The comparisons with empirical models are conducted using the test set for the period 2012–2020.

3. Methods

To validate the performance of the DL model for forecasting solar wind speed, two verification methods are applied. First, statistical verification evaluates the model's accuracy in predicting solar wind speed by analyzing prediction errors and provides an overview of its predictive performance. Second, event-based verification evaluates the model's ability to detect HSS events by evaluating its predictions of the arrival time and intensity of fast solar winds. These two verification methods complement each other: statistical verification is useful for evaluating the model's overall prediction accuracy, and event-based verification is essential for ensuring the model's reliability in predicting critical solar wind events. This comprehensive validation process identifies the strengths and weaknesses of each model, providing a basis for improving the accuracy of solar wind forecasting.

3.1. Statistical Verification

For statistical verification, the models' performance is evaluated using OMNI data, which indicate solar wind speed at 1 au. The performance of the models is assessed using metrics including mean absolute error (MAE), rms error (RMSE), and the correlation coefficient (CC). These metrics are defined as follows:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|f_i - y_i|, \tag{1}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(f_i - y_i)^2}{N}}, \tag{2}$$

$$CC = \frac{\sum_{i=1}^{N}(f_i - \bar{f})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(f_i - \bar{f})^2}\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}}. \tag{3}$$

Here, f_i, y_i, $\bar{f}$, $\bar{y}$, and N denote the ith predicted value, the ith target value, their average values, and the total number of data, respectively. MAE is an intuitive metric that indicates how close the predicted values are to the target values. Since it considers all errors equally, it is less affected by outliers and provides a simple way to evaluate the accuracy of the model. RMSE is more sensitive to large errors, which makes it suitable for assessing how well a model handles significant discrepancies. It emphasizes the overall magnitude of the prediction errors. When used together with MAE, it offers a more comprehensive analysis of the model's performance. CC measures the relationship between the predicted and target values, evaluating how well the model captures the patterns of data variation. A CC value close to 1 indicates that the predicted values closely follow the patterns of the target values, demonstrating the model's ability to accurately capture the characteristics of the data. These three metrics offer diverse perspectives on model performance, enabling a comprehensive assessment of its strengths and weaknesses.

To analyze how the model performance varies with the solar activity cycle, statistical verification is conducted not only for the entire period but also for four distinct phases of the solar activity cycle. As solar activity levels vary across these phases, the model performance is also expected to fluctuate accordingly. Solar activity phases, classified according to the October–December period, include the ascending phase (2012), the maximum phase (2013–2014), the descending phase (2015–2018), and the minimum phase (2019–2020).

We use a Taylor diagram (K. E. Taylor 2001) to visualize the performance of various prediction models for different solar activity phases (P. Riley et al. 2013). This diagram allows us to compare different models and track changes in their performance over a period. In the Taylor diagram, the radial distance from the origin represents the standard deviation, the azimuth (angle from the horizontal axis) represents the CC, and the Euclidean distance from the reference point (OMNI data, placed on the x-axis at CC = 1) represents the RMSE. Consequently, when a model accurately predicts the OMNI data, its standard deviation is similar to that of the reference, its CC is high, and its RMSE is low, positioning the model's results close to the reference star on the diagram.

3.2. Event-based Verification

Event-based verification evaluates how accurately solar wind speed prediction models can forecast actual HSS events identified from OMNI data. For this purpose, we use the algorithm for automatic detection (L. K. Jian et al. 2015) based on the methods proposed by M. J. Owens et al. (2005) and P. MacNeice (2009). This algorithm detects peaks in HSS events in both the OMNI data and the model predictions, excluding ICME events. Figure 1 shows the application of the algorithm for automatic detection, highlighting the identified peaks in HSS events with inverted triangle markers in both the observed and predicted solar wind speed data.

The model verification process is conducted by checking whether the predicted events belong to a ±1 day time window around the peaks in the OMNI data. If a predicted event matches an OMNI peak within the specified time window, it is classified as "hits." Conversely, if no predicted event occurs within the time window, it is classified as "misses."
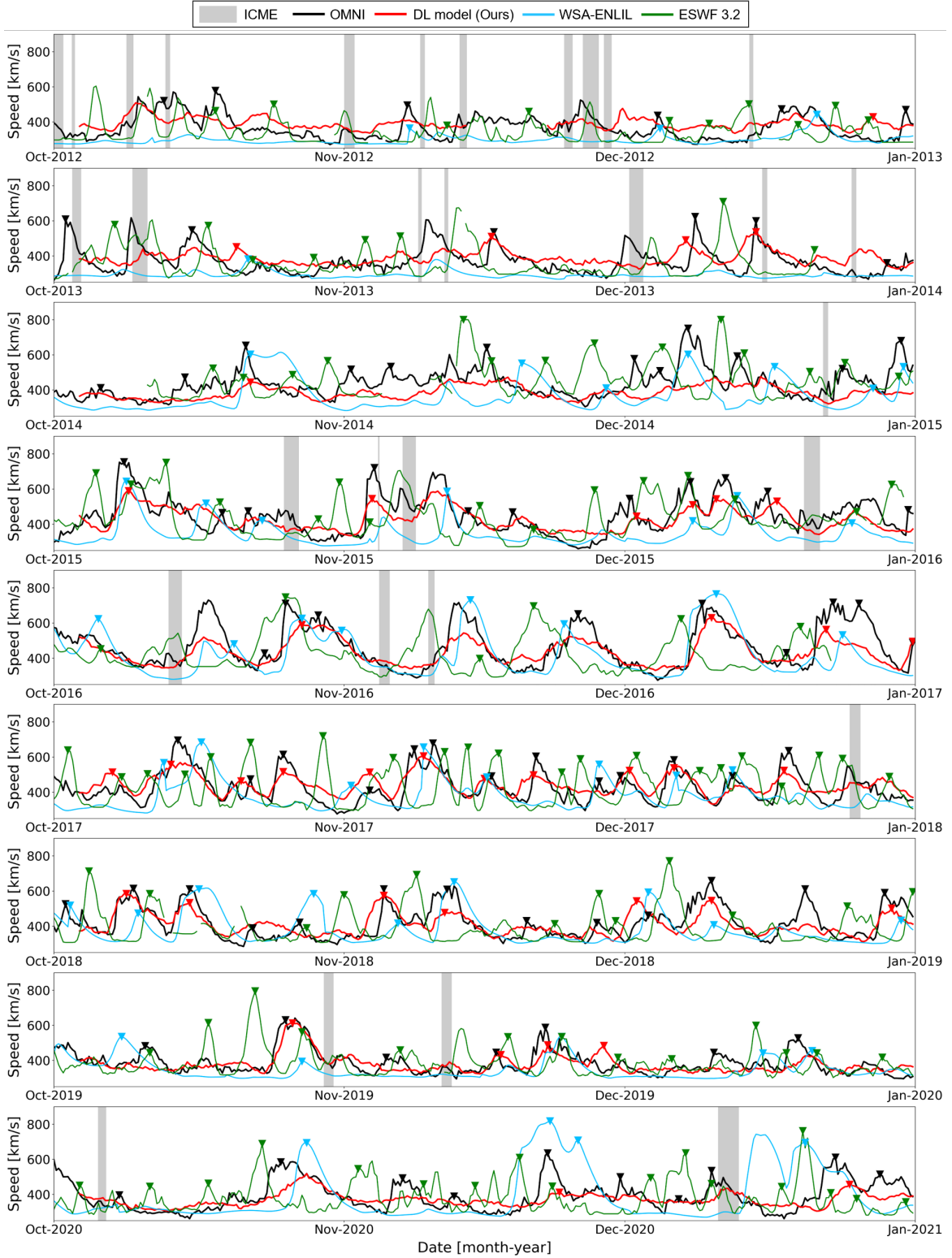
Figure 1. Solar wind speed at 1 au during the period October–December for the years 2012–2020 by OMNI (black), our work (red), WSA–ENLIL (sky blue), and ESWF 3.2 (green). Here inverted triangles indicate HSS events detected by the algorithm for automatic detection. Intervals corresponding to ICMEs are shaded in gray, with these periods based on the Richardson and Cane ICME catalog.

Additionally, if a model erroneously predicts an event when no actual HSS event has occurred, it is categorized as "false alarms." Using this methodology, hits (true positives, TPs), misses (false negatives, FNs), and false alarms (false positives, FPs) are counted to construct a confusion matrix, as shown in Table 1. Since this analysis focuses on predicted and observed events, cases with no predicted or observed events (true negatives, TNs) are not considered.

Using the calculated TP, FN, and FP, we derive the following metrics to evaluate model performance. The

Table 1
Confusion Matrix for Predicted and Observed HSS Events

	Observed	
Predicted	HSS	No HSS
HSS	TP	FP
No HSS	FN	TN

Table 2
Statistical Verification Results

CR Period	Ascending Phase			Maximum Phase			Descending Phase			Minimum Phase			Entire Period		
Metrics	MAE	RMSE	CC	MAE	RMSE	CC	MAE	RMSE	CC	MAE	RMSE	CC	MAE	RMSE	CC
DL model (Ours)	**57.6**	**67.3**	**0.49**	**60.5**	**83.3**	0.35	**51.0**	**66.2**	**0.77**	**41.7**	**58.6**	**0.65**	**51.4**	**68.6**	**0.69**
WSA–ENLIL	67.7	92.9	0.35	98.0	119.2	0.37	86.2	114.8	0.49	86.8	125.7	0.24	86.9	116.3	0.41
ESWF 3.2	74.5	94.9	0.12	91.5	122.2	0.19	117.4	147.8	−0.02	86.5	116.1	−0.09	100.4	130.9	0.08
Persistence (3 days)	63.4	82.5	0.32	82.6	105.7	0.23	107.6	135.0	0.11	74.3	97.1	0.19	90.1	116.3	0.22
Persistence (4 days)	65.9	85.8	0.26	85.8	109.3	0.18	116.9	146.4	−0.04	80.3	105.8	0.06	96.7	125.0	0.11
Persistence (5 days)	69.9	89.7	0.2	84.6	107.9	0.22	120.4	151.4	−0.1	85.4	111.6	0.0	99.7	128.9	0.06
Persistence (27 days)	67.1	86.2	0.22	64.1	88.5	**0.39**	68.3	90.9	0.61	59.5	77.8	0.55	65.2	87.0	0.57

Note. The performance of the DL model, WSA–ENLIL model, ESWF 3.2 model, and persistence models is evaluated by statistical metrics such as MAE, RMSE, and CC for different solar activity phases. Bold values indicate the best performance for each metric.

probability of detection (POD), which is the fraction of observed events successfully detected by the model, is given by

$$POD = \frac{TP}{TP + FN}. \quad (4)$$

The false negative rate (FNR), which is the fraction of observed events that the model failed to detect, is given by

$$FNR = \frac{FN}{TP + FN}. \quad (5)$$

The success ratio (SR), which is the fraction of predicted events that match the observed events, is given by

$$SR = \frac{TP}{TP + FP}. \quad (6)$$

The false alarm ratio (FAR), which is the fraction of predicted events that fail to occur among all predicted events, is given by

$$FAR = \frac{FP}{TP + FP}. \quad (7)$$

The critical success index (CSI), which is the ratio of correctly predicted events to the total number of recorded events, is given by

$$CSI = \frac{TP}{TP + FN + FP}. \quad (8)$$

These metrics range from 0 to 1, where for POD, SR, and CSI, 0 indicates poor performance and 1 indicates perfect performance. In contrast, for FNR and FAR, the reverse is true. The bias score (BS), which is the ratio of predicted events to observed events, is given by

$$BS = \frac{TP + FP}{TP + FN}. \quad (9)$$

Although this does not measure the correlation between predictions and observations, it shows whether the model underestimates (BS < 1) or overestimates (BS > 1) observations. These six metrics (POD, FNR, SR, FAR, CSI, and BS) are used to evaluate the performance of solar wind speed prediction models in detecting HSS events. A detailed discussion on the verification matrix is given by D. S. Wilks (2011) and I. T. Jolliffe & D. B. Stephenson (2012).

4. Results and Discussion

The performance of each model—the DL model and the two empirical models—in predicting solar wind speed is evaluated using both statistical and event-based verification. Figure 1 shows the solar wind speed predictions at 1 au from 2012 to 2020. The DL model shows fluctuations similar to the OMNI data, especially during 2016 and 2017, although it generally underestimates the solar wind speed. The WSA–ENLIL model shows relatively smooth and stable predictions, but deviates significantly from the OMNI data during HSS events. The ESWF 3.2 model tends to predict an earlier arrival of HSS events than observations; however, its predictions are overly variable, resulting in large discrepancies with the OMNI data at certain intervals.

Table 2 presents a comparison of the statistical metrics for the DL model, the two empirical models, and the four persistence models for different solar activity phases, as well as over the entire period. Persistence is the naive forecasting model that carries forward the most recently observed solar wind speed unchanged (M. J. Owens et al. 2013). For all four solar activity phases, the DL model consistently shows lower errors than the empirical models and persistence models, both within each phase and over the entire period. Over the entire period, the DL model achieves an MAE of 51.4 km s^{-1} and an RMSE of 68.6 km s^{-1}, demonstrating lower errors than the empirical models, with a high CC of 0.69. In particular, during
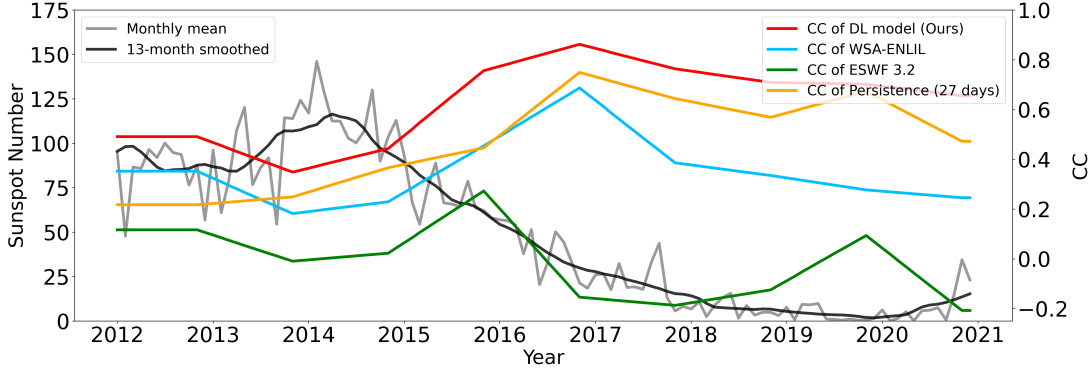
Figure 2. Variations in the solar sunspot number and the DL model's CC over the entire period (2012–2020). The gray line represents the monthly mean total sunspot number, obtained by taking the simple arithmetic mean of the daily sunspot numbers for each calendar month. The black line indicates the monthly smoothed total sunspot number, derived by applying a 13 months tapered-boxcar running mean. The red line represents the CC of the DL model, the sky blue line represents the CC of WSA–ENLIL, the green line represents the CC of ESWF 3.2, and the orange line represents the CC of the 27 days persistence model.
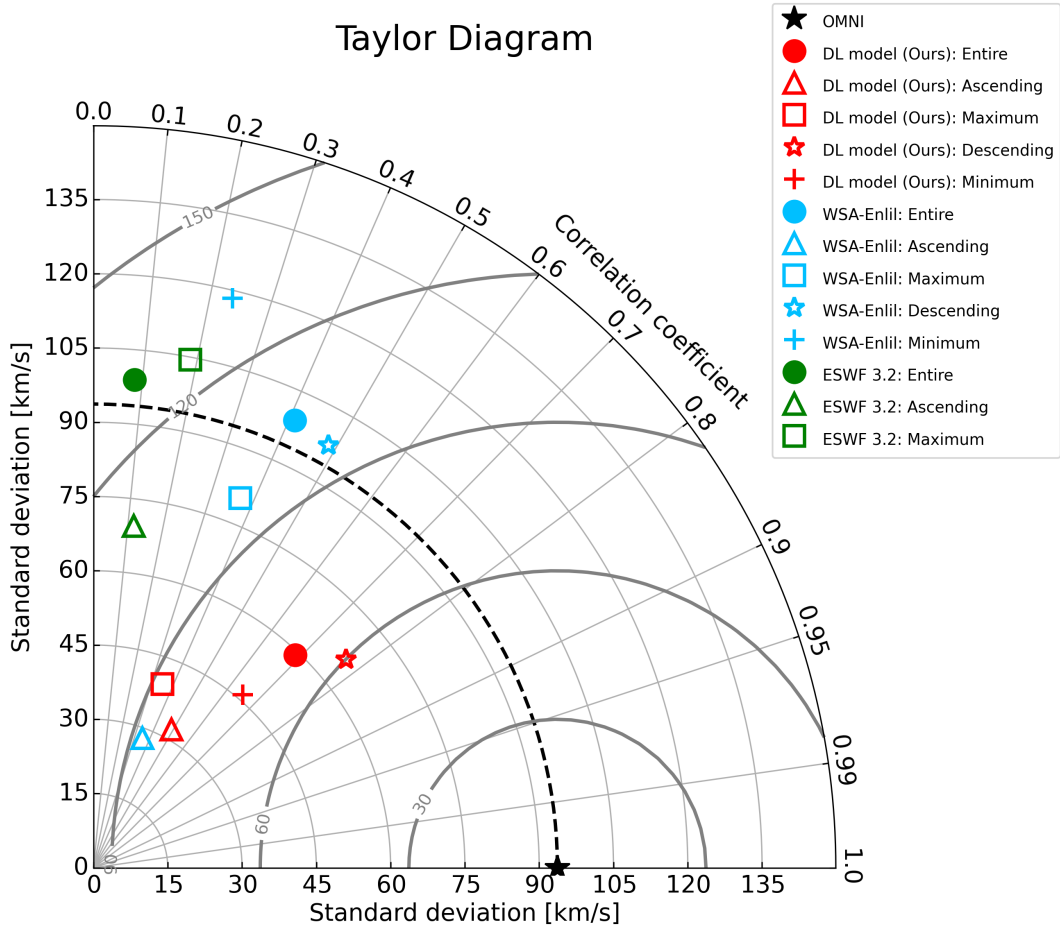
Figure 3. Taylor diagram showing the statistical comparison of RMSE, CC, and standard deviation for solar wind speed prediction models for different solar activity phases and over the entire period. Note that because the CC values for the descending and minimum phases in the ESWF 3.2 model are negative, they are not displayed in the figure.

the descending phase, the DL model records notably lower errors, with an MAE of $51.0 \, \mathrm{km \, s^{-1}}$ and an RMSE of $66.2 \, \mathrm{km \, s^{-1}}$. Moreover, the DL model achieves a higher CC of 0.77, demonstrating a good agreement with the observations. In contrast, the ESWF 3.2 model shows negative CC during the descending and minimum phases, which is probably

caused by insufficient detection of the CH regions. The results show that the DL model consistently outperforms the four persistence models.

Figure 2 shows how the CC of the models and the solar sunspot numbers from the Solar Influences Data Analysis Center (SIDC)/Sunspot Index and Long-term Solar
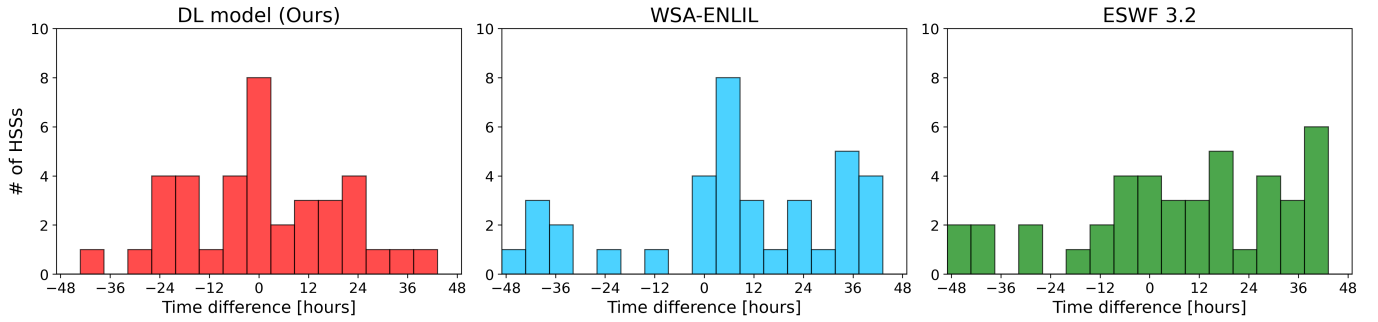
Figure 4. Histograms showing the number of HSS events within a ±2 days time window for the DL model (left, red), the WSA–ENLIL model (center, blue), and the ESWF 3.2 model (right, green). The histogram, using a bin size of 6 hr, shows the time difference on the horizontal axis calculated as (the predicted peak times – the observed peak times for HSS events).

Table 3
Event-based Verification Results

Model	Totals	Entire Period (2012–2020)								
		TP	FP	FN	POD	FNR	SR	FAR	CSI	BS
DL model (ours)	40	33	7	54	0.38	0.62	**0.82**	**0.18**	**0.35**	0.46
WSA–ENLIL	52	21	31	66	0.24	0.76	0.4	0.6	0.18	0.6
ESWF 3.2	116	25	91	62	0.29	0.71	0.22	0.78	0.14	1.33
Persistence (27 days)	94	41	53	46	**0.47**	**0.53**	0.44	0.56	0.29	**1.08**

Note. Comparison of metrics (POD, FNR, SR, FAR, CSI, and BS) for the DL model, the WSA–ENLIL model, the ESWF 3.2 model, and the 27 days persistence model based on ±1 day time windows over the period of October–December for the years 2012–2020. Bold values indicate the best performance for each metric.

Observations (SILSO) vary over the entire period of solar cycle 24. As seen in Table 2, the DL model outperforms the other models in all phases, suggesting that the DL model captures the CH areas more effectively than the other models. In addition, the DL model has the best CC value (0.86) during the descending phase (around 2016) and maintains good performance during the minimum phase, which may be related to the CH areas that are dominant during these phases as the source of fast solar winds. In contrast, during the maximum phase, when the CH area remains primarily at high latitudes and solar eruptive activities such as CMEs become more frequent, the performance of the DL model is reduced compared to other phases. This result is presumed to arise because the DL model cannot adequately learn information about the magnetic field structure.

Figure 3 shows the performance of the models for different solar activity phases and over the entire period using a Taylor diagram. The diagram shows RMSE, CC, and standard deviation, providing an intuitive comparison of the models' performance. Note that the CCs for the descending and minimum phases in the ESWF 3.2 model are negative; consequently, they are not displayed in the figure. Consistent with the results in Table 2, the DL model (red) has lower RMSE values and higher CC than the empirical models, indicating better predictive performance. The lower RMSE values indicate that the values predicted by the DL model are closer to the observations than those from the empirical models. The higher CC values imply that the pattern of predicted solar wind speeds is more similar to the observations than those of the empirical models. However, the DL model's standard deviations are notably lower than those of the empirical models.

Figure 4 shows the histograms of identified HSS events within a ±2 days time window for the DL model and two empirical models. The WSA–ENLIL and ESWF 3.2 models demonstrate a wide distribution of time differences between their predicted peak times for HSS events and the observed peak times for them. In contrast, the DL model shows a distribution similar to a normal distribution, with time differences concentrated around zero, indicating that the predicted peak times for HSS events align more closely with the observed peak times for them.

Table 3 shows the results of the event-based verification, including the counts for TP, FP, and FN, as well as six metrics: POD, FNR, SR, FAR, CSI, and BS. This method identifies HSS events from the OMNI data using a ±1 day time window to more accurately detect event peaks. Unlike the POD of the WSA–ENLIL model of 0.24 and the POD of the ESWF 3.2 model of 0.29, the DL model records a higher POD of 0.38, thus outperforming these empirical models in accurately predicting HSS events. It also maintains the highest SR of 0.82, underscoring its reliability. Moreover, it attains the highest CSI of 0.35, demonstrating the best overall performance among the models. In comparison, the WSA–ENLIL model records an SR of 0.4 and a CSI of 0.18, and the ESWF 3.2 model records an SR of 0.22 and a CSI of 0.14, further highlighting the advantages of the DL model. However, the DL model shows the lowest BS value, suggesting a tendency to underestimate the occurrence of HSS events compared to the empirical models. Although the 27 days persistence model achieves the highest POD of 0.47, its SR of 0.44 and CSI of 0.29 are below those of the DL model, indicating lower reliability.

5. Summary

In this study, we have compared a DL model for solar wind speed prediction with two representative empirical models: the WSA–ENLIL and ESWF 3.2 models. For statistical verification, the DL model outperforms the empirical models throughout the evaluation period (2012–2020), achieving the lowest MAE of 51.4 km s^{-1}, RMSE of 68.6 km s^{-1}, and the highest CC of 0.69. In particular, during the descending phase (2015–2018), it records an MAE of 51.0 km s^{-1}, an RMSE of 66.2 km s^{-1}, and a CC of 0.77, indicating impressive performance. For event-based verification, histogram analysis reveals that the DL model's identified HSS events within a ±2 days time window follow a distribution similar to a normal distribution, indicating that the predicted peak times align more closely with the observed peaks. Moreover, the DL model achieves the highest SR of 0.82 and CSI of 0.35 in detecting HSS events within a ±1 day time window. These findings indicate that the model is effective in reducing false alarms and improving prediction accuracy, although it tends to underestimate the frequency of events.

The DL model, which serves as a practical tool for space weather forecasting, provides highly accurate and reliable predictions. In particular, it offers near-real-time solar wind speed forecasts with an inference time of less than a minute. Our DL model still has the limitation that, although its POD is higher than those of other empirical models, it remains low due to a high FN rate and is lower than that of the persistence models. This fact suggests that the DL model does not fully capture the dynamic variability of solar wind speed. During the solar maximum phase, both the empirical models and the DL model struggle to predict solar wind speed, underscoring the need to incorporate information on magnetic field structure. For example, adding a magnetogram or a squashing factor that quantifies magnetic topology would help the model learn the source regions of the slow solar wind more accurately and thus enhance its predictive performance. In addition, the study focuses on event-based verification of HSS events, as CME data are currently not incorporated into the model. In an effort to address this, attempts to integrate CME data into the model have resulted in only minimal improvements, suggesting that a deeper analysis of the relationship between CMEs and solar wind speed is necessary for future research. Lastly, models that combine DL with empirical methods, such as hybrid frameworks or those that integrate attention mechanisms, are expected to enhance the sensitivity of event detection. This would improve the overall accuracy and reliability of space weather forecasting.

Acknowledgments

ORCID iDs

Seungwoo Ahn https://orcid.org/0009-0007-8498-9985
Jihyeon Son https://orcid.org/0000-0003-2678-5718
Yong-Jae Moon https://orcid.org/0000-0001-6216-6944
Hyun-Jin Jeong https://orcid.org/0000-0003-4616-947X

References

Abbo, L., Ofman, L., Antiochos, S., et al. 2016, SSRv, 201, 55
Altschuler, M. D., & Newkirk, G. 1969, SoPh, 9, 131
Arge, C., & Pizzo, V. 2000, JGRA, 105, 10465
Bravo, S., & Stewart, G. A. 1997, ApJ, 489, 992
Brown, E. J., Svoboda, F., Meredith, N. P., Lane, N., & Horne, R. B. 2022, SpWea, 20, e2021SW002976
Jian, L. K., MacNeice, P. J., Taktakishvili, A., et al. 2015, SpWea, 13, 316
Jolliffe, I. T., & Stephenson, D. B. 2012, Forecast Verification: A Practitioner's Guide in Atmospheric Science (New York: Wiley)
MacNeice, P. 2009, SpWea, 7, S12002
McComas, D., Barraclough, B., Funsten, H. O., et al. 2000, JGR, 105, 10419
Milošić, D., Temmer, M., Heinemann, S., et al. 2023, SoPh, 298, 45
Odstrcil, D., Linker, J. A., Lionello, R., et al. 2002, JGRA, 107, 1493
Odstrcil, D., Pizzo, V. J., Linker, J. A., et al. 2004, JASTP, 66, 1311
Owens, M. J., Arge, C. N., Spence, H. E., & Pembroke, A. 2005, JGRA, 110, A12105
Owens, M. J., Challen, R., Methven, J., Henley, E., & Jackson, D. 2013, SpWea, 11, 225
Raju, H., & Das, S. 2021, SoPh, 296, 134
Reiss, M. A., Temmer, M., Veronig, A. M., et al. 2016, SpWea, 14, 495
Richardson, I. G., & Cane, H. V. 2010, SoPh, 264, 189
Riley, P., Linker, J. A., & Arge, C. N. 2015, SpWea, 13, 154
Riley, P., Linker, J. A., & Mikić, Z. 2013, JGRA, 118, 600
Riley, P., Linker, J. A., Mikić, Z., et al. 2002, ApJ, 578, 972
Robbins, S., Henney, C. J., & Harvey, J. W. 2006, SoPh, 233, 265
Rotter, T., Veronig, A. M., Temmer, M., & Vršnak, B. 2015, SoPh, 290, 1355
Schatten, K. H. 1971, CosEl, 2, 232
Schatten, K. H., Wilcox, J. M., & Ness, N. F. 1969, SoPh, 6, 442
Son, J., Sung, S.-K., Moon, Y.-J., Lee, H., & Jeong, H.-J. 2023, ApJS, 267, 45
Sun, Y., Xie, Z., Chen, Y., Huang, X., & Hu, Q. 2021, SpWea, 19, e2020SW002707
Taylor, K. E. 2001, JGRD, 106, 7183
Upendran, V., Cheung, M. C., Hanasoge, S., & Krishnamurthi, G. 2020, SpWea, 18, e2020SW002478
Vřšnak, B., Temmer, M., & Veronig, A. M. 2007, SoPh, 240, 315
Wang, Y.-M., & Sheeley, N., Jr. 1990a, ApJ, 355, 726
Wang, Y.-M., & Sheeley, N., Jr. 1990b, ApJ, 365, 372
Wang, Y.-M., & Sheeley, N., Jr. 1992, ApJ, 392, 310
Wilks, D. S. 2011, Statistical Methods in the Atmospheric Sciences (New York: Academic)