



# Ensemble Kalman Filter Data Assimilation into the Surface Flux Transport Model to Infer Surface Flows: An Observing System Simulation Experiment

Soumyaranjan Dash<sup>1</sup> , Marc L. DeRosa<sup>2</sup> , Mausumi Dikpati<sup>3</sup> , Xudong Sun (孙旭东)<sup>1</sup> , Sushant S. Mahajan<sup>4</sup> , Yang Liu (刘扬)<sup>4</sup> , and J. Todd Hoeksema<sup>4</sup>

<sup>1</sup> Institute for Astronomy, University of Hawai'i at Mānoa, Pukalani, HI 96768, USA; [dashs@hawaii.edu](mailto:dashs@hawaii.edu)

<sup>2</sup> Lockheed Martin Solar and Astrophysics Laboratory, Palo Alto, CA 94306, USA

<sup>3</sup> High Altitude Observatory, NSF-NCAR, 3080 Center Green Drive, Boulder, CO 80301, USA

<sup>4</sup> W.W. Hansen Experimental Physics Laboratory, Stanford University, Stanford, CA 94305-4085, USA

Received 2024 June 25; revised 2024 September 9; accepted 2024 September 21; published 2024 November 8

## Abstract

Knowledge of the global magnetic field distribution and its evolution on the Sun's surface is crucial for modeling the coronal magnetic field, understanding the solar wind dynamics, computing the heliospheric open flux distribution, and predicting the solar cycle strength. As the far side of the Sun cannot be observed directly and high-latitude observations always suffer from projection effects, we often rely on surface flux transport (SFT) simulations to model the long-term global magnetic field distribution. Meridional circulation, the large-scale north-south component of the surface flow profile, is one of the key components of the SFT simulation that requires further constraints near high latitudes. Prediction of the photospheric magnetic field distribution requires knowledge of the flow profile in the future, which demands reconstruction of that same flow at the current time so that it can be estimated at a later time. By performing Observing System Simulation Experiments, we demonstrate how the ensemble Kalman filter technique, when used with an SFT model, can be utilized to make "posterior" estimates of flow profiles into the future that can be used to drive the model forward to forecast the photospheric magnetic field distribution.

Unified Astronomy Thesaurus concepts: Solar magnetic fields (1503); Solar meridional circulation (1874)

## 1. Introduction

Magnetic fields on the Sun drive a diverse range of phenomena that span different timescales. Solar flares, coronal mass ejections, and radiation from energetic particles influence our space environment in the short term, whereas the emergence of sunspots and magnetic flux transport on the photosphere control the solar magnetic activity cycle that covers a longer timescale. Overall, these events impact Earth's space environment, affecting satellite operations and telecommunication, and contribute to space weather variations. Knowledge of the surface magnetic field distribution and evolution is crucial for predicting both short-term space weather and long-term climate variations. The Sun's surface magnetic field distribution is extremely important and is used in multiple applications as a boundary condition for modeling coronal magnetic fields, an initial condition for solar wind models, as well as to compute the heliospheric open solar flux distribution. Polar magnetic field evolution is also closely linked to the behavior of the solar cycle. It allows us to make predictions about the amplitude and timing of solar cycles (L. Upton & D. H. Hathaway 2014; P. Bhowmik & D. Nandy 2018; D. Nandy 2021; B. K. Jha & L. A. Upton 2024; S. Pal & D. Nandy 2024), which have implications on long-term climate variability.

The solar dynamo is essentially an interplay between the poloidal and toroidal components of the Sun's magnetic field. During a solar activity minimum, the global magnetic field is dominated by the poloidal component. The solar differential rotation (i.e., the longitudinal component of the large-scale flow) stretches this poloidal field longitudinally to generate the

toroidal field component in the interior. E. N. Parker (1955) suggests that due to magnetic buoyancy, these amplified toroidal flux ropes manifest as bipolar magnetic regions (BMRs) on the surface. While the toroidal flux rope rises through the solar convection zone, Coriolis forces typically tilt it such that one of the sunspots emerges leading in longitude and closer to the equator while the other one is behind (following) in longitude and closer to the pole of the respective hemisphere (S. D'Silva & A. R. Choudhuri 1993). Near the Sun's equator, the leading BMR polarities of two opposite hemispheres cancel each other, and the remaining magnetic flux drifts toward the respective poles via large-scale meridional flows (north-south component of the large-scale flow) and diffusion.

In order to study this flux transport process, we require global magnetic field observations spanning all the longitudes and latitudes. However, current ground- and space-based observing facilities are largely restricted to the Sun-Earth line, which limits our observation window to one-half of the Sun's surface. Such observations are not reliable near high latitudes and limb extrema due to high projection effects. To address this issue, we often utilize the surface flux transport (SFT) model, which solves the radial component of the magnetic induction equation (Equation (1)) with imposed large-scale flow parameters, surface magnetic diffusivity, and observed BMR properties (more details in the review paper by A. R. Yeates et al. 2023). Some of the SFT models also use data assimilation (DA) to incorporate the observed magnetic flux to generate a global photospheric magnetic field map. In order to accurately model the magnetic flux transport we require an observationally constrained description of global flow parameters, i.e., differential rotation and meridional circulation as well as magnetic diffusivity. Meridional flow has been observed in the



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

photosphere and within the upper convection zone, spanning from the equatorial region to approximately  $60^\circ$  latitude in each hemisphere (R. W. Komm et al. 1993; D. H. Hathaway & L. Rightmire 2010; S. S. Mahajan et al. 2021). Nevertheless, comprehensive knowledge regarding the flow speed, profile, and temporal fluctuations over multiple solar cycles on the surface and within the deeper convection zone remains elusive based on current observational data owing in part to the increased uncertainty near the polar latitudes due to projection effects. Recently, S. S. Mahajan et al. (2023) analyzed variations in the global meridional flow pattern for solar cycle 24 using the time-distance helioseismology technique utilizing data from Helioseismic Magnetic Imager (HMI; P. H. Scherrer et al. 2012a; J. Schou et al. 2012b) on board the Solar Dynamics Observatory and found that the meridional flow does show a small variation on the order of  $2 \text{ m s}^{-1}$  far away from active regions and much larger variation close to active regions due to inflows. The study, however, is still limited to  $62^\circ$  latitude due to projection effects.

Despite these limitations, SFT modeling has proven remarkably effective in calculating the distribution of photospheric magnetic flux. Both large-scale surface-flow components, i.e., differential rotation and meridional circulation, primarily show latitude-dependent variation. SFT simulations are sensitive to parameterized flow profiles. Modeled surface magnetic field distribution can be calibrated to the observations considering the available observations of the polar field or its proxy. In order to achieve this, different SFT models consider slightly different yet solar-cycle-independent meridional flow profiles and magnetic diffusivity to mimic the flux cancellation and transport process. This raises the question, is it possible to infer the parameter value, e.g., the peak flow speed of an assumed meridional flow profile, using the magnetic field observations? In this paper, we build an Observation System Simulation Experiment (OSSE) system with the SFT model, which will guide us further in building a model using observed magnetic fields to infer the different physical variables (e.g., flow parameters, magnetic diffusivity).

Ensemble Kalman filter (EnKF) data assimilation is ideal for our purpose. The EnKF computational technique uses observations sequentially obtained at a current time together with associated “prior knowledge” from the numerical computation of physical models to provide Bayesian estimates of the “posterior” states of the system (model+observations) while accounting for uncertainties in the observations and using physical model outputs of the evolved states. Although the motivation for the development of such a technique arose from the need for short-to-medium range weather forecasting in geophysics (G. Evensen 1994), now this technique has been used extensively in many scientific fields such as Earth upper-atmosphere models (T. Matsuo et al. 2013; N. M. Pedatella et al. 2013), weather prediction models (S. Ha et al. 2017; I. Hoteit et al. 2021), and ocean circulation models (Y. Chen et al. 2022). EnKF is used for estimating or constraining otherwise unknown parameters of a physical model, as well as producing better forecasts that are updated through the most recent observations while integrating the model forward in time rather than deriving from pure model outputs. The filtering, in brief, is a statistical perspective for obtaining the posterior distributions of a system’s states and their uncertainties at the current time based on all accumulated observations so far. The Kalman filter estimates the state of a system using two steps: (i)

estimating the state and uncertainty in that state, which are adjusted to newly available observations, and (ii) forecasting the updated state and uncertainty by propagating them forward in time. A model, such as Gaussian error, is used to obtain the first guess of the states, and then from the error covariance matrices of the observations and the theoretically generated observations from the first guesses of the states, the corrections to the first guess of the state are estimated. The process of combining the first guesses in an ensemble of models with observations to derive the new states is essentially the EnKF data assimilation. A detailed review can be found in G. Evensen (2003).

Numerous studies have explored the application of data assimilation in solar physics (A. S. Brun 2007); for example, the reconstruction of dynamic solar corona using ensemble Kalman filtering and tomography methods with observations (M. D. Butala et al. 2010), forecasting solar cycle behavior by combining real observations with a reduced  $\alpha$ - $\Omega$  dynamo model (I. Kitiashvili & A. G. Kosovichev 2008), using the four-dimensional variational method and a cellular-automaton-based avalanche model to predict simulated solar-flare data (E. Bélanger et al. 2007), and generating ensemble members of the photospheric magnetic field distribution with localized EnKF techniques driven by available observations (C. N. Arge et al. 2010; K. S. Hickmann et al. 2015). L. Jouve et al. (2011) developed a variational data assimilation framework based on a solar  $\alpha$ - $\Omega$  dynamo model and validated it through synthetic data. A. Svedin et al. (2013) showed how a three-dimensional variational data assimilation methodology can be utilized to obtain accurate model estimates given a set of observations. A. Fournier et al. (2013) employed the EnKF framework in a three-dimensional, convection-driven geodynamo model, using surface poloidal magnetic fields for full-state estimation. M. Dikpati et al. (2014, 2016) utilized EnKF sequential data assimilation to reconstruct the time variations in the meridional flow speed over multiple solar cycles, utilizing poloidal and toroidal magnetic fields as observational data in a two-dimensional kinematic Babcock-Leighton solar dynamo model.

Utilizing EnKF data assimilation with a one-dimensional SFT model, we aim to reconstruct the parameters of the meridional flow profile. With statistical multidimensional regression analysis using a Bayesian approach, we can compare the magnetic field observations to the model-generated magnetic field for a random initial guess of the meridional peak flow speed. Subsequently, the model parameter (peak flow speed) is adjusted to minimize the deviation between the observed and modeled magnetic field, thereby eventually converging at a parameter value inspired by direct observations.

We build a DA experiment with a 1D SFT model to study how sensitive the reconstruction of the peak meridional flow speed is to different tunable DA parameters: (a) magnetic field observations near high latitudes—how does the error in reconstruction vary with increasing or decreasing the assimilation interval? (b) possible error associated with the observations—how well can the model parameter be reconstructed with different values of error associated with the observations? In the following sections, we describe our SFT model and different cases for the reconstruction of the peak meridional flow speed. In Section 2, we explain our SFT model and mathematical formulation of the EnKF methodology, and Section 3 describes our results. A comprehensive discussion and summary of our study are presented in Section 4.

## 2. Models

### 2.1. Surface Flux Transport Model

The SFT model is based on the idea that radial magnetic flux on the solar surface is carried around by horizontal plasma flows with no back reaction on these flows (R. B. Leighton 1964). This can be modeled by solving the radial component of the magnetic induction equation on the solar surface,

$$\frac{\partial B_r}{\partial t} + \nabla \cdot (u_h B_r) = \eta \nabla^2 B_r + S, \quad (1)$$

where  $\eta$  denotes the effect of supergranular diffusion and  $u_h$  describes the imposed horizontal plasma flows, i.e., meridional circulation and differential rotation. The supergranular diffusion is assumed to be a constant quantity; hence it is placed outside of the spatial derivative operator.  $B_r(s, \phi, t)$  represents the large-scale mean radial magnetic field defined on a sine-latitude ( $s = \sin \lambda$ ) and longitude ( $\phi$ ) grid for the time step  $t$ . The source terms, i.e., BMRs emerging on the solar surface, are denoted by  $S(s, \phi, t)$ . Since the evolution of the axial dipole moment depends only on the longitude-averaged field (C. R. DeVore et al. 1984; R. Cameron & M. Schüssler 2007; H. Iijima et al. 2017; K. Petrovay & M. Talafha 2019; A. R. Yeates 2020), we simplify our SFT model by averaging the mean field over the longitude  $\bar{B}(s, t) = \frac{1}{2\pi} \int_0^{2\pi} B_r(s, \phi, t) d\phi$ . The simplified SFT model can be described by

$$\frac{\partial \bar{B}}{\partial t} = \frac{\partial}{\partial s} \left[ \frac{\eta}{R_\odot^2} (1 - s^2) \frac{\partial \bar{B}}{\partial s} - \frac{v_s(s)}{R_\odot} \sqrt{1 - s^2} \bar{B} \right], \quad (2)$$

where  $v_s(s)$  denotes the meridional circulation and  $R_\odot$  is the solar radius. Meridional circulation is parameterized as

$$v_s(s) = D_u s (1 - s^2)^{(p/2)}. \quad (3)$$

We can compute the axial dipole moment using the longitudinally averaged magnetic field as,

$$DM = \frac{3}{2} \int_{-1}^1 s \bar{B}(s, dt) ds. \quad (4)$$

A. R. Yeates (2020) introduced such a mathematical formulation of the SFT model. We developed our numerical model following A. R. Yeates (2020) in FORTRAN for this study. We solve the model equations numerically using the finite-volume method. We adopt their meridional flow profile shape of  $p = 2.33$  and use  $D_u = 0.085 \text{ km s}^{-1}$ , which corresponds to a meridional peak flow speed of  $v_0 = 30.7 \text{ m s}^{-1}$ . In the sine-latitude grid, there are 180 grid points considered for our simulation. The source terms are modeled as idealized BMRs using the observed sunspot properties taken from the HMI database spanning a time frame from 2010 June 17 to 2023 September 4 following the formalism provided by A. R. Yeates (2020). Figure 1 shows the time-latitude distribution of selected BMRs, where the color indicates the unsigned flux magnitude and the size of the markers denotes the separation between opposite polarities. There are 1643 BMRs identified for incorporation into our SFT model. The BMR properties are extracted from the `hmi.sharp_cea_720s` series available

through Joint Science Operations Center using the script distributed by A. R. Yeates (2020). It is important to note that all the BMR properties are considered at their maximum area coverage during disk passage. Utilizing these properties we model the source terms for our SFT model following the algorithm described in the same paper. Our SFT model can be accessed on GitHub<sup>5</sup> with a copy archived to Zenodo: doi:10.5281/zenodo.13831085 (S. Dash et al. 2024).

We evolve the SFT model for 14 yr to generate a butterfly diagram of  $\bar{B}(s, t)$ . The initial condition for our simulation is the magnetic field distribution for Carrington map 2097. For our experiment, we used  $\eta = 450 \text{ km}^2 \text{s}^{-1}$  and  $v_0 = 30.7 \text{ m s}^{-1}$  based on several trial runs, as these values best reproduce the observed polar fields. An analogous butterfly diagram of  $B_r$  is plotted from HMI Carrington maps for comparison. Figure 2 shows the comparison between the HMI-observed radial component of the surface magnetic field and the SFT-generated butterfly diagram of the surface radial magnetic field. The brown dashed line toward the right of Figure 2(b) indicates the last BMR insertion into the SFT model. Beyond this line, the model only shows the forward run of the SFT model without new active regions.

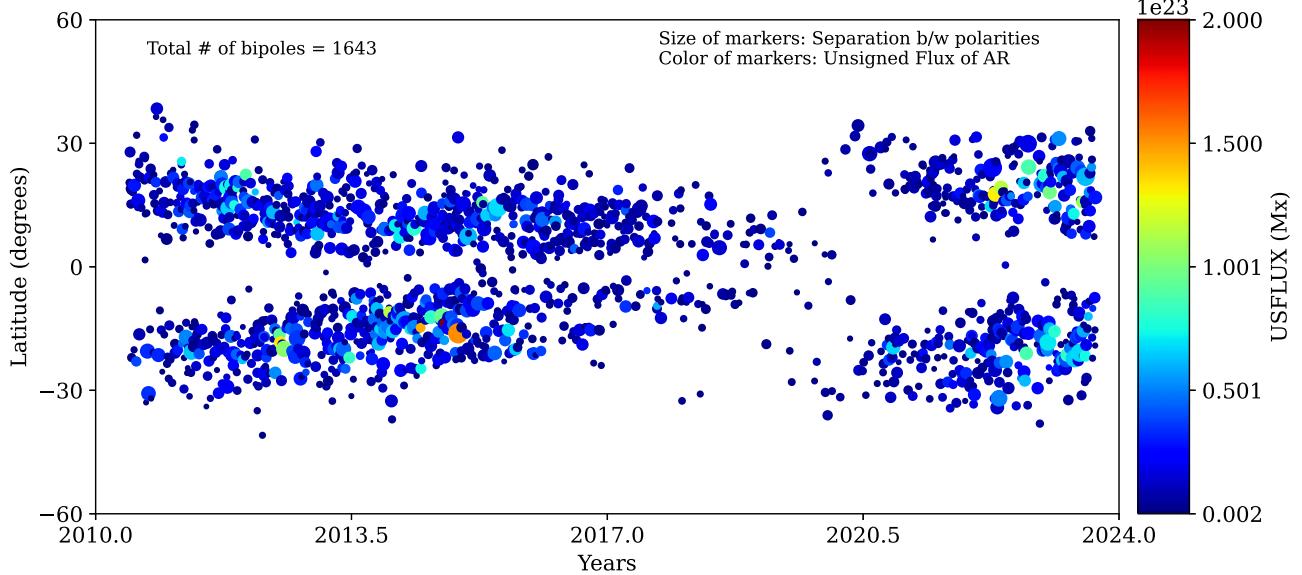
It is important to note that for our DA experiment, the numerical model must have “restart” functionality. This is because the data assimilation modifies the model parameter (peak meridional flow speed for our case) constrained by the surface magnetic field observations at a user-defined assimilation cadence. After each modification, the simulation is restarted with the new value of peak meridional flow speed for the next iteration of the assimilation window. We model the restart functionality in our simulation and compare the resulting magnetic field at random time frames to ensure the correctness of magnetic fields at the last frame and continuation in the next frame. Figure 2(b) denotes the uninterrupted SFT evolution for 14 yr (5111 days, considering the leap years). Panels (c), (d), and (e) show partial SFT model evolution starting from 0 days, 1500 days, and 3500 days, respectively.

In order to check the magnetic field at the end of an SFT run and the start of the next instance, the latitudinal variation in  $B_r$  is independently plotted and compared with the corresponding time step of panel (b). The continuity in the butterfly diagram in panels (c), (d), and (e) also indicates the evolution of surface magnetic field distribution. The magnetic field profile and amplitude at these iterations do not show a significant difference when compared with the corresponding time step of panel (b), which validates the restart functionality of our model. Now the numerical model is ready to be coupled with the EnKF algorithm to infer various model parameters.

### 2.2. Coupling the Ensemble Kalman Filter Algorithm with the Surface Flux Transport Model

Data assimilation is performed by the EnKF as implemented in the Data Assimilation Research Testbed (DART; J. Anderson et al. 2009, <https://dart.ucar.edu>). The algorithm operates through a two-phase process comprising a prediction phase and an update phase. During the prediction phase, the Kalman filter generates estimates of the current state variables along with their associated uncertainties. Then for the update phase, these parameter estimates are used to compute the observed

<sup>5</sup> <https://github.com/sr-dash/SFT-1D>



**Figure 1.** Time–latitude distribution of BMRs obtained from HMI observations. There are a total of 1643 BMRs identified at their maximum disk-passage area. The color of the markers denotes the unsigned flux and size corresponds to the separation between opposite polarities.

quantities and the corresponding uncertainties. In an iterative process, state variables are “nudged” closer to the true state that is inspired by the observations. An OSSE designed with EnKF in this context helps to estimate the potential value of state variables guided by the simulated observations.

Here, we provide a concise description of the Kalman Filter algorithm starting from a Bayesian formulation, as it applies to our DA experiment. For a detailed and more generalized mathematical derivation, readers can refer to M. Dikpati et al. (2016). The surface flux transport process can be numerically modeled by solving the equations of the SFT model (Equation (2)) that require the prescription of meridional flow (Equation (3)). One of the parameters in the presumed flow profile is the maximum flow speed in both hemispheres, which can be considered as our state variable ( $D_u$  in Equation (3)). One can use  $x_i$  for the range of state variables involved in modeling a physical system, where  $i$  is the index of variable type and location in the range [1,  $n$ ]. Here  $n$  denotes the dimension of state space. Ideally, the state vector ( $\mathbf{x}_i$ ) can evolve in a time-dependent fashion that can be expressed as

$$\mathbf{x}_t = \mathbb{M}(\mathbf{x}_{t-1}, t). \quad (5)$$

In data assimilation terminology, the physical SFT model is the forward operator, whereas  $\mathbb{M}$  is the “model,” which is a random-walk operator that generates the prior flow speed from prior time step  $t-1$ . In our case, the state vector  $\mathbf{x}_t$  is the peak meridional flow speed ( $D_u$ ), which is time-independent.

For our experiment, we assume a prescribed meridional flow profile (Equation (3)) that does not vary with time. Using the vector notation, we can describe the state vector  $\mathbf{x} \equiv \mathbf{x}^i$  and the corresponding observations  $\mathbf{y} \equiv \mathbf{y}^j$ , where the index  $i$  denotes the state vector space that has a range [1,  $n$ ] and the index  $j$  shows the observations with a range [1,  $m$ ]. As stated earlier,  $n$  is the dimension of the state space, whereas  $m$  is the dimension of the observation space. While reconstructing the time-independent flow speed by

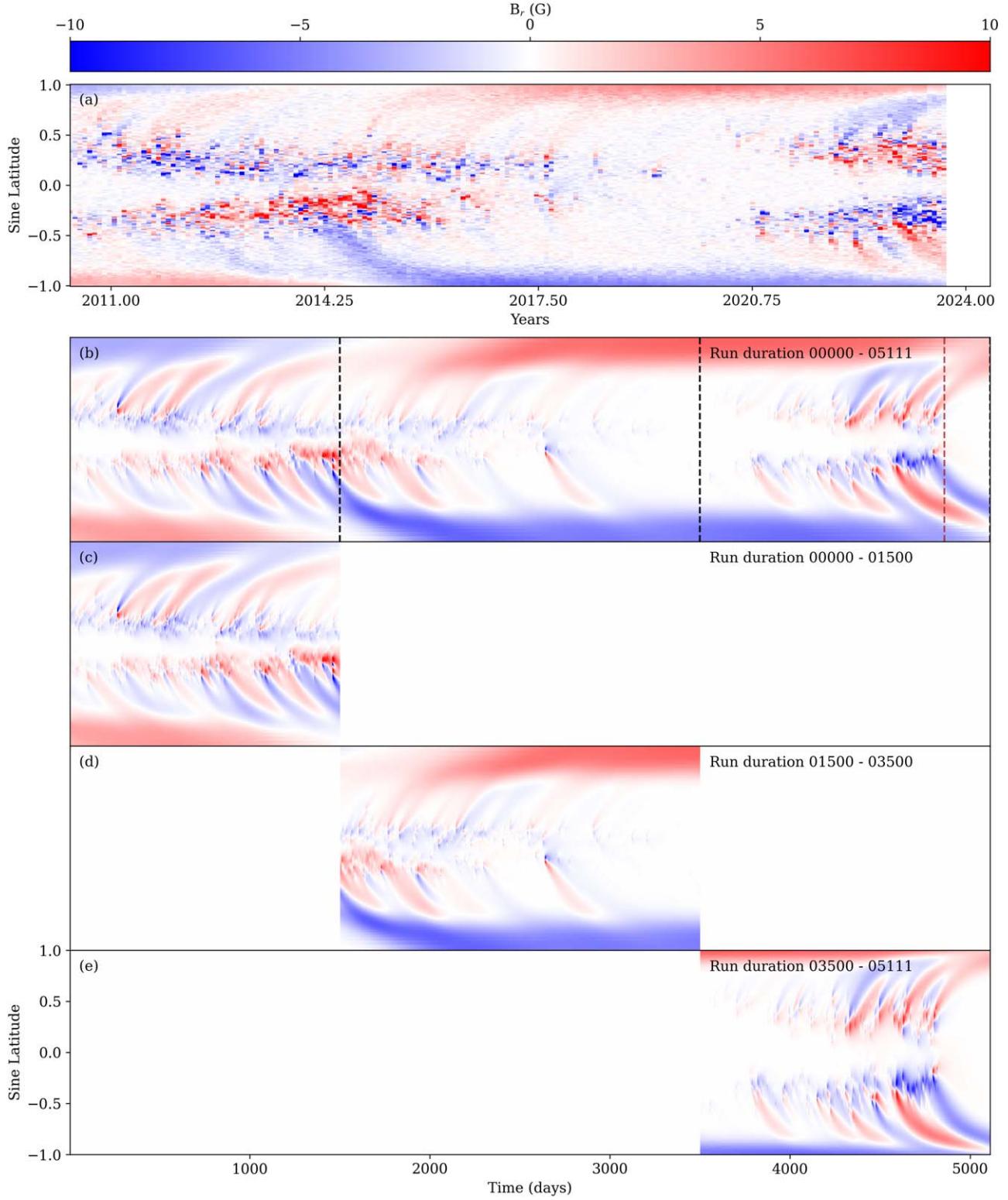
assimilating data in the SFT model, the state vector dimension is 1, i.e.,  $n = 1$  in our case. As per EnKF-DART requirements, the observation space dimension will have to be less than or equal to the dimension of the state vector. Thus,  $m$  is also 1. In our OSSE, the state vector  $\mathbf{x}^i$  is inferred from the observations  $\mathbf{y}^j$  at different assimilation steps. A detailed description of the integration of SFT in the EnKF framework is provided in Section 2.2.1.

For an assimilation time window  $m_t$ , a set of observations  $\mathbf{y}_t^i$  are available that can be generated using OSSE. Usually  $m_t$  can vary with time and is much smaller than  $n$ . The prior estimates of the set of observations can be generated by a forward operator  $\mathbf{h}(\mathbf{x}, t)$  as,

$$\mathbf{y}^0 = \mathbf{h}(\mathbf{x}, t) + \epsilon^0(\mathbf{x}, t), \quad (6)$$

where the forward operator ( $\mathbf{h}$ ) and state variable ( $\mathbf{x}$ ) are assumed to model the observations accurately and  $\epsilon^0$  is the uncertainty associated with the observations. For our DA experiment, the SFT model equation (Equation (2)) is the forward operator and meridional peak flow speed ( $v_0$ ) is the state variable that can be expressed as Equation (3), which is analogous to Equation (5). In our OSSE system, we choose the value of the magnetic field at 60°N latitude to be the observations ( $\mathbf{y}^0$ ).

The uncertainty associated with the simulated system may be due to (i) limitations of the forward operator or (ii) errors in the initial conditions. In the EnKF experiment, we can only estimate the type (ii) uncertainty. In the context of our SFT model, the model equations are averaged over the longitude. Hence, the impact of differential rotation on the flux transport process is not modeled. We also assume that the magnetic flux is only advected in response to the large-scale flow properties. Such assumptions limit us from modeling the small-scale flux distribution and the EnKF algorithm will not be able to model the uncertainty associated with such assumptions. On the other



**Figure 2.** Butterfly diagram of the surface radial magnetic field from observations and the SFT model, along with demonstrations of the “restart” functionality. Panel (a) shows the HMI-observed butterfly diagram interpolated onto the SFT model resolution. A total of 14 yr of SFT model evolution is plotted in (b), where the brown dashed line toward the right indicates the time beyond which no new BMRs are incorporated into the SFT model. In all the panels, the magnetic field saturates at  $\pm 10$  G. Panel (b) shows the uninterrupted SFT run for 14 yr, where the black dashed lines indicate the time stamps at which the computation is stopped and restarted, the results of which are shown below. Panel (c) shows the SFT evolution from  $t = 0$  to  $t = 1500$  days. Panel (d) shows the SFT evolution from  $t = 1500$  to  $t = 3500$  days. Panel (e) shows the SFT evolution from  $t = 3500$  to  $t = 5111$  days. EnKF data assimilation requires the numerical model to have restart functionality.

hand, we can estimate the errors associated with the initial condition of the model, i.e., a synoptic map of magnetic field distribution and properties of the BMRs used to generate the

source terms from the distribution of  $\mathbf{x}$  in the ensemble. If  $\mathbf{x}^{\text{Tr}}$  is the true state, i.e., the true state variable, and  $\mathbf{x}^p$  is the prior estimate of the state variable with EnKF (Equation 5), then the

error in the prior estimate ( $\epsilon^p$ ) can be written as

$$\epsilon^p = \mathbf{x}^p - \mathbf{x}^{\text{Tr}}. \quad (7)$$

Similarly, the posterior state  $\mathbf{x}^a$  and the error ( $\epsilon^a$ ) associated with it can be computed by,

$$\epsilon^a = \mathbf{x}^a - \mathbf{x}^{\text{Tr}}. \quad (8)$$

Here the index  $a$  denotes quantities of the analysis states, i.e., the posterior states. In order to find the optimal  $\mathbf{x}^a$ , we require observation ( $\mathbb{R}$ ) and model ( $\mathbb{P}$ ) error covariance matrices,

$$\mathbb{R} = \overline{\epsilon^0 \epsilon^{0T}}, \quad (9)$$

$$\mathbb{P} = \overline{\epsilon^p \epsilon^{pT}}, \quad (10)$$

where the bar denotes the statistical average and  $T$  indicates the transpose of the corresponding column vector. We also evaluate the estimated analysis error covariance ( $\mathbb{A}$ ),

$$\mathbb{A} = \overline{\epsilon^a \epsilon^{aT}}. \quad (11)$$

Now with this system of prior and analysis variables, we pose the question: What would be the most likely state vector  $\mathbf{x}^a$  and its estimated variance  $\epsilon^a$ , considering the prior state vector  $\mathbf{x}^p$  and available observations  $\mathbf{y}^0$ ? In probability notation,  $P(\mathbf{x}^a|\mathbf{x}^p, \mathbf{y}^0)$  is the probability of occurrence of  $\mathbf{x}^a$  provided  $\mathbf{x}^p$  and  $\mathbf{y}^0$  are known. According to Bayes' theorem, the probability distribution of  $\mathbf{x}$  given  $\mathbf{y}^0$  is

$$P(\mathbf{x}|\mathbf{y}^0) = \frac{P(\mathbf{y}^0|\mathbf{x})P(\mathbf{x})}{P(\mathbf{y}^0)}. \quad (12)$$

Assuming the state vectors ( $\mathbf{x}$ ) are drawn from a multivariate Gaussian distribution, prior state  $P(\mathbf{x})$  can be written as,

$$P(\mathbf{x}) = \frac{1}{(2\pi)^{1/2} |\mathbb{P}^{-1}|} e^{-\frac{1}{2} |(\mathbf{x}^p - \mathbf{x})^T \mathbb{P}^{-1} (\mathbf{x}^p - \mathbf{x})|}, \quad (13)$$

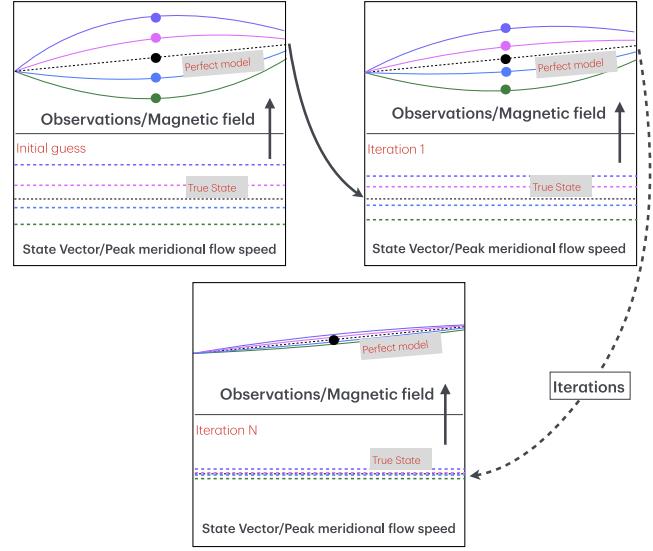
and  $P(\mathbf{y}^0|\mathbf{x})$  as,

$$P(\mathbf{y}^0|\mathbf{x}) = \frac{1}{(2\pi)^{1/2} |\mathbb{R}^{-1}|} e^{-\frac{1}{2} |(\mathbf{y}^0 - \mathbf{h}(\mathbf{x}))^T \mathbb{R}^{-1} (\mathbf{y}^0 - \mathbf{h}(\mathbf{x}))|}. \quad (14)$$

Substituting Equations (13) and (14) in Equation (12), we can reduce the problem into finding an optimum distribution of  $\mathbf{x}$ . By linearizing  $\mathbf{h}$  around  $\mathbf{x}^p$ , we can calculate an extremum of  $\mathbf{x}$  as,

$$\mathbf{x}^a = \mathbf{x}^p + (\mathbb{P} + \mathbf{h}^T \mathbb{R}^{-1} \mathbf{h})^{-1} \mathbf{h}^T \mathbb{R}^{-1} [\mathbf{y}^0 - \mathbf{h}(\mathbf{x}^p)]. \quad (15)$$

This is the expression for the Kalman Filter. A more detailed mathematical description in a solar context is provided in M. Dikpati et al. (2016). For EnKF, the posterior state/update is constructed by creating an ensemble distribution of state variables from a multivariate Gaussian distribution and using these prior states to compute the ensemble of posterior states. If the errors associated with the observations are not expected to be correlated at different time steps, we can employ the sequential Bayesian analysis to accelerate the computation of EnKF. It is worthy of note that sequential Bayesian analysis is a statistical method used for updating the probability estimate of a state vector as new observations become available. After each update, the new posterior probability becomes the prior probability for the next iteration of the analysis. This is known as the ensemble adjustment Kalman filter proposed by J. L. Anderson (2001). We have used this variant of EnKF for our experiment.



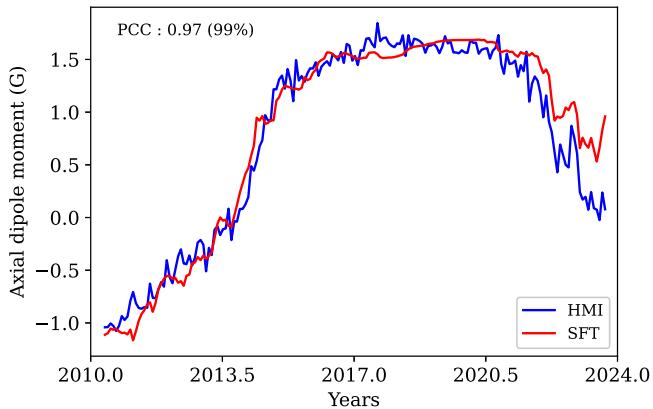
**Figure 3.** The illustration shows how the surface flux transport model is coupled to the ensemble Kalman filter data assimilation algorithm. Initially, an ensemble of state vectors, i.e., the peak meridional flow speed is generated and utilized in the SFT model evolution to calculate the corresponding observations, i.e., the magnetic fields. In the next iteration, the state vectors were “nudged” toward the true state. The process drives the resulting observations toward the perfect model. The iteration time step is decided by the assimilation cadence of the DA experiment. This iterative process is continued for the whole time duration of the global SFT calculations; e.g., if the global SFT simulation is set to evolve for 300 days, for an assimilation cadence of 30 days the iteration  $N$  would be 100.

### 2.2.1. Applying Ensemble Kalman Filter Data Assimilation to the Surface Flux Transport Model

In the EnKF framework, the state vector ( $\mathbf{x}_t$  in Equation (5)) prescribes the meridional flow in our SFT model where the peak flow speed (defined by Equation (3)) is one of the model parameters. In the following sections, we will be performing DA experiments to infer the value of peak meridional flow speed  $v_0$ , which can be calculated from  $D_u$  following A. R. Yeates (2020),

$$v_0 = \pm D_u p^{p/2} (1 + p)^{-(1+p)/2}. \quad (16)$$

For  $D_u = 0.085 \text{ km s}^{-1}$  and  $p = 2.33$ , the peak meridional flow speed is  $v_0 = 30.7 \text{ m s}^{-1}$ . With our state vector prescription, the observations ( $\mathbf{y}^0$  in Equation (6)), i.e., the magnetic field, are computed using the forward operator ( $\mathbf{h}(x, t)$  in Equation (6)). A perfect SFT model is computed that provides the values of magnetic fields at all latitudes for the above flow parameters. In the data assimilation framework, an ensemble of the initial guesses for the state vectors is generated and is used to evolve the SFT model to the next iteration. Since the initial guesses for the state vectors are drawn from a multivariate Gaussian distribution, the resulting observations do not necessarily generate identical observations as the perfect model. In the next iteration, the state vectors are adjusted according to Bayes' theorem (Equation (12)). This iterative process is continued until the end of the SFT evolution. In this iterative process, the state vector is adjusted according to the observations as depicted by the illustration in Figure 3.



**Figure 4.** Axial DM from HMI observations and the SFT model. The blue solid line denotes the DM time series calculated using HMI observations (butterfly diagram) and the red line indicates a similar calculation for the SFT-modeled magnetic field. A Pearson correlation coefficient value of 0.97 between the HMI- and SFT-modeled axial DM indicates that the numerical model performs very well.

### 3. Results

Using the SFT model described in Section 2.1, we evolve the surface radial magnetic field for about 14 yr using a peak meridional flow speed  $v_0 = 30.7 \text{ m s}^{-1}$ . The resulting butterfly diagram is compared to the observations in Figure 2. In order to test our model output, we compare the calculated axial dipole moment (DM; Equation (4)) with the observations in Figure 4. The Pearson correlation coefficient between these two time series is 0.97, which indicates that the SFT model output is reasonable. For our current experiment, we choose the value of the magnetic field at 60°N latitude, noting that the result of the DA experiment does not depend on the location of the reference magnetic field value. In principle, one can choose to utilize the values of the magnetic field at all latitudes as their true observations. However, this makes the setup computationally expensive due to increased dimensionality.

The resulting butterfly diagram from the SFT model (Figure 2(b)) is used as our “true observations” for the data assimilation experiment with a true state (peak meridional flow speed) value of  $v_0 = 30.7 \text{ m s}^{-1}$ . In order to check the sensitivity and robustness of the assimilation results we explore variations in the: (a) assimilation cadence, (b) error associated with the true observations, and (c) ensemble size. We define our perfect model with a peak meridional flow speed of  $v_0 = 30.7 \text{ m s}^{-1}$ . Depending upon the assimilation cadence, the perfect model observations (i.e., the magnetic fields at 60°N latitude) for the corresponding time step from the original SFT model are considered for our DA experiment. For different experiments with different ensemble sizes and error variances, the perfect model calculations are performed with the same parameters. Only the magnetic field at 60°N latitude generated by the perfect model is used for computing the state vectors for these different cases. Consequently, the state vectors are “nudged” based only on observations sampled at that latitude point. Under the current setup of EnKF data assimilation, each experimental case takes approximately 45 minutes to complete for an assimilation cadence of 60 days and an ensemble size of 20. Reconstruction of the state vector at multiple spatial locations for a larger ensemble size would be more computationally demanding.

### 3.1. Impact of Different Assimilation Cadence on Data Assimilation

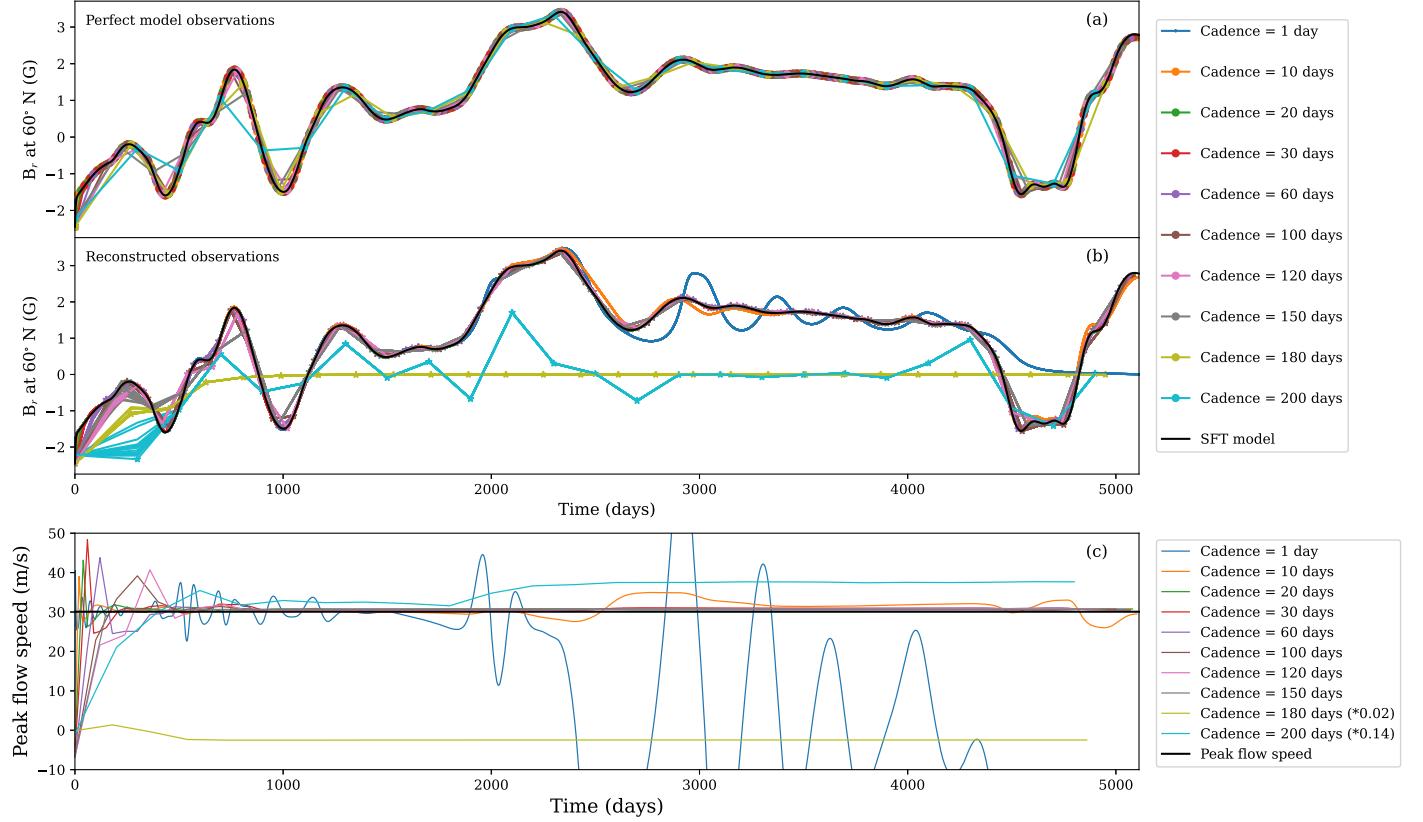
Assimilation cadence is probably one of the most important parameters in a data assimilation experiment. This corresponds to a time after which the DA simulation is interrupted, the value of the state vector is compared with the true states, and a posterior state vector is created using Equation (15). The ensemble size chosen for this study is 20 with an error variance value of  $10^{-6}$ . Our SFT model has a time step of  $dt = 4 \text{ hr}$ . We run our DA experiment for assimilation cadences of 1, 10, 20, 30, 60, 100, 120, 150, 180, and 200 days. The choice of these cadence values is arbitrary. For all model runs, the assimilation cadence denotes the time interval after which the state vector (i.e., the peak meridional flow speed) is modified and the observation (i.e., the magnetic field) is sampled.

Figure 5(a) shows the true observations (value of the magnetic field from the SFT model) in different colors. Figure 5(b) indicates the reconstructed observations (value of the magnetic field from the SFT model after the DA). The initial deviation from the perfect model in this plot shows the impact of the data assimilation experiment where the state vector (peak meridional flow speed) is iteratively “nudged” toward the perfect model. Figure 5(c) shows the evolution of one of the state vectors (value of peak meridional flow speed) from the ensemble. For all the cases, the baseline value of the magnetic field and  $v_0$  is solid black. In some cases, we observe a deviation from the true observations. To summarize, we show the final mean value of the ensemble of state variables in Figure 6. Here the markers denote the ensemble mean peak meridional flow speed at the end of the DA experiment for different assimilation cadence instances. For assimilation cadences between 10 and 150 days, the reconstructed observations and corresponding peak meridional flow speed are close to the true state values. We discuss the possible cause of the spread from the true state for a few cases of high assimilation cadence value within the plot in Section 4.

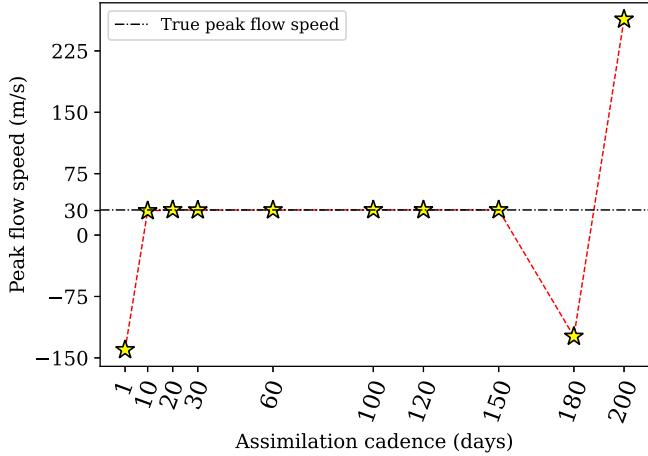
To explore the reason for deviations at low and high assimilation cadences, we inspect the results for the case of 1 day, that is, we modify the ensemble of state vectors after each day. In Figure 7 we plot the true observations (SFT magnetic field at 60°N latitude), reconstructed observations, and state variable (peak meridional flow speed) in panels (a), (b), and (c), respectively. Initially, the state variable and the reconstructed observations oscillate around the true state; however, at later times they diverge drastically. It is apparent that for such a high cadence of assimilation, the reconstructed observations are not converging toward the true value at the end of the simulation. This is commonly known as “ensemble collapse,” which is discussed in Section 4.

### 3.2. Impact of Different Error Variance on Data Assimilation

Uncertainty associated with the modeled surface magnetic field may originate from the initial conditions or accuracy of the numerical schemes. In our case, the output of the SFT model is an “exact” numerical solution of a partial differential equation within the limit of our numerical discretization scheme, which does not have any uncertainties. However, in the case of direct magnetic field observations, there can be associated uncertainties. Therefore we explore a parameter space of different error variances associated with the true observations. For this study, we choose an assimilation cadence



**Figure 5.** True observations, reconstructed observations, and state vectors are plotted for different assimilation cadences. (a) Perfect model observations (magnetic field at  $60^\circ$  N latitude) for different assimilation cadence intervals. (b) Reconstructed magnetic field at  $60^\circ$  N when the meridional flow is iteratively “nudged” toward the true state vector (peak meridional flow speed). (c) Evolution of peak meridional flow speed (state vector) for these cases where the variable is approaching the parameter value of the perfect model calculation. The solid black line indicates the outputs from the original SFT model and the different colors correspond to various assimilation cadences. Note that for very high (more than 150 days) and very low (less than 10 days) assimilation cadences, the state vector values do not converge toward the true state. State vectors for assimilation cadences of 180 days and 200 days are rescaled to fit in the plot.



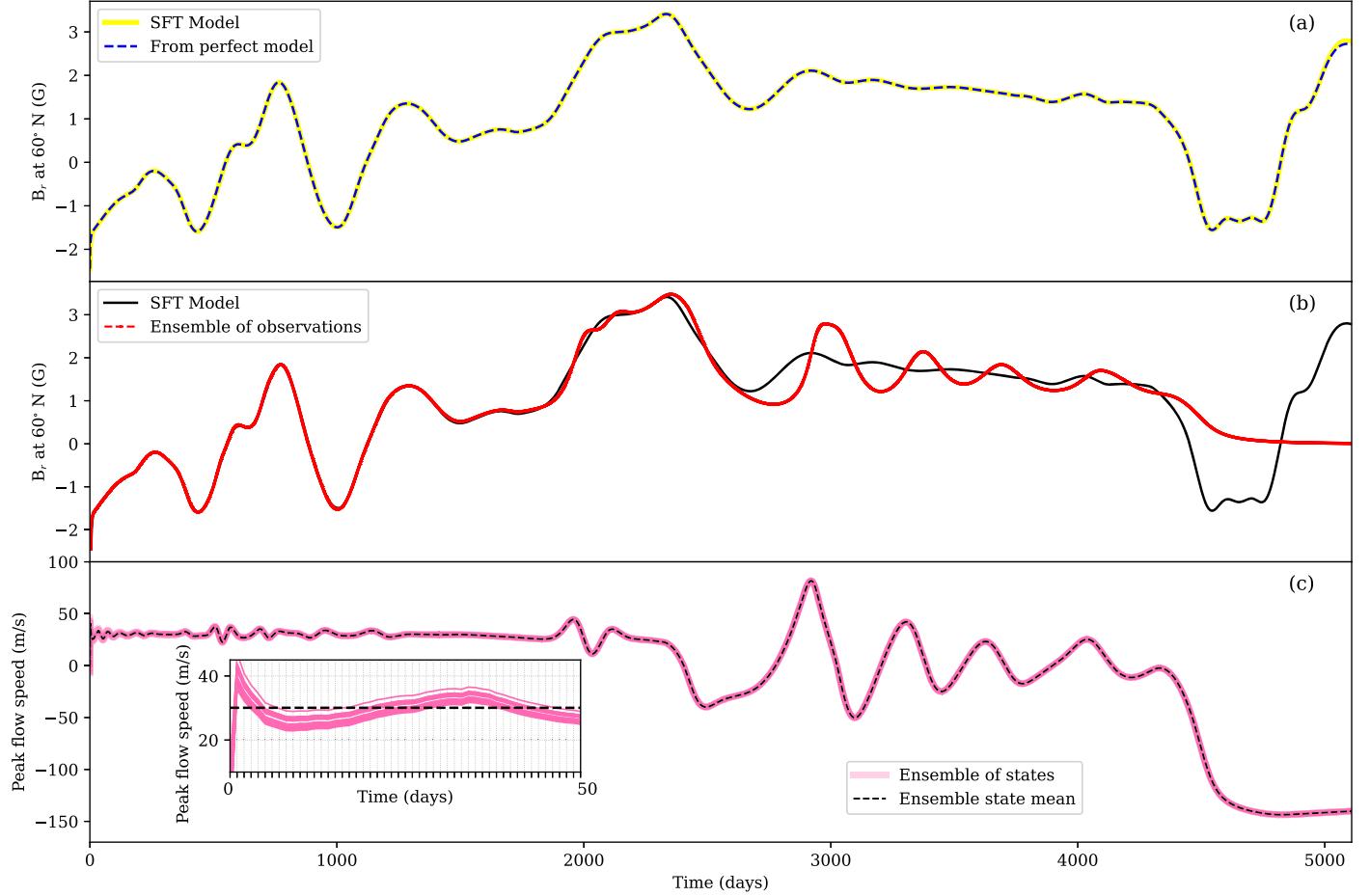
**Figure 6.** Ensemble mean of the state vector at the end of the DA experiment for different assimilation cadences. The horizontal axis denotes the different cadence values and the vertical axis indicates the value of the ensemble mean of state vectors at the end of the DA experiment. It is essential to choose an appropriate value of data assimilation frequency in order to correctly reconstruct the observations. The black dashed curve depicts the true state value for reference.

of 60 days and an ensemble size of 20. We run our DA experiment for relative error variance values of  $10^{-6}$ ,  $10^{-4}$ ,  $10^{-3}$ , and  $10^{-2}$ . Figure 8 shows a comparison between the perfect model observations and the reconstructed observations along with the state vectors for different cases of error variance

value. For most of the cases, the reconstruction is reasonably good with minor deviation from the perfect model. Since the ensemble of state vectors for the assimilation process is drawn from a multivariate Gaussian distribution, the algorithm shows an oscillatory pattern around the true state before finally converging, as highlighted by the inset plot in Figure 8(c). In Figure 9 we plot the final ensemble mean value of the state vector at the end of the DA experiment. In all cases, the inferred meridional peak flow speed from our DA experiment is close to the true state value; the maximum difference is less than  $1 \text{ m s}^{-1}$ . The assimilation cadence chosen here is 60 days. We intentionally choose this assimilation cadence to highlight the fact that even though there are fewer observations available to assimilate, the degree of error associated with those observations can modify the inferred state vector.

### 3.3. Impact of Different Ensemble Sizes on Data Assimilation

In our data assimilation scheme, the size of the ensemble is one of the controlling parameters that can impact the robustness of the assimilation process. On one hand, in order to produce a meaningful reconstruction the size should not be too small. On the other hand, the size of the ensemble plays a key role in the effective use of computational resources, as the forward operator has to compute the observations for a given state vector. J. L. Anderson (2010) shows that choosing a larger ensemble size does not necessarily improve the assimilation results. For our DA experiment, we explored the impact of



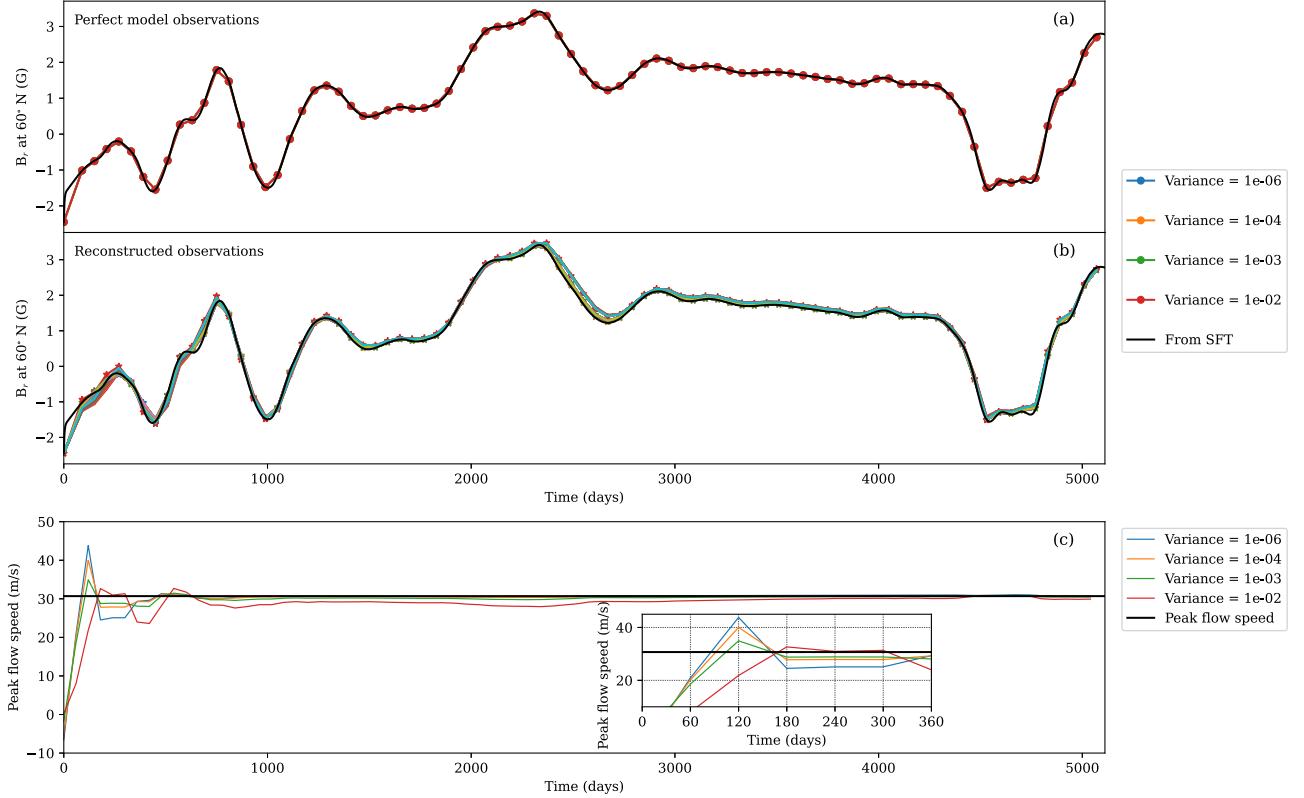
**Figure 7.** Data assimilation with an assimilation frequency of 1 day. True observations (magnetic field at  $60^\circ\text{N}$  latitude), reconstructed magnetic fields in red solid line, and the variation in the peak meridional flow speed are plotted in panels (a), (b), and (c), respectively. Such a high assimilation cadence does not provide appropriate relaxation time to the forward operator; hence the resulting reconstruction deviates from the perfect model. The inset plot in (c) denotes the first 50 days of the assimilation process indicating the oscillatory behavior around the true state (dashed black curve) before approaching the true state. Such oscillations are found for all the cases of the DA experiment.

different ensemble sizes, e.g., 2, 5, 10, 15, 20, 25, 40, 60, and 70. Figure 10 shows a comparison between the perfect model and the reconstructed observations along with the state vectors for computations with different ensemble sizes. We have chosen an assimilation cadence of 60 days and an error variance of  $10^{-6}$  for this study. In all the cases, initially the algorithm produces state vectors that oscillate around the true state before converging with minor deviations from the perfect model. In Figure 11, we plot the final ensemble mean value of the state vector at the end of the DA experiment. Here the black dashed curve denotes the true state for reference. As we increase the ensemble size the assimilation result seems to improve while saturating toward the end. In our DA experiment, we have prior knowledge of the true state parameter (i.e., the peak meridional flow speed); however, when such an assimilation technique is applied to observations of the magnetic field to extract the peak meridional flow speed, the choice of a suitable ensemble size is crucial for meaningful interpretation of the results. The difference in the state vector for ensemble size 40 is only  $0.04 \text{ m s}^{-1}$ , which does not significantly alter the modeled magnetic field. However, we believe that for an assimilation process involving multi-dimensional spatial reconstruction of global flow characteristics, such a degree of sensitivity of the algorithm will be helpful.

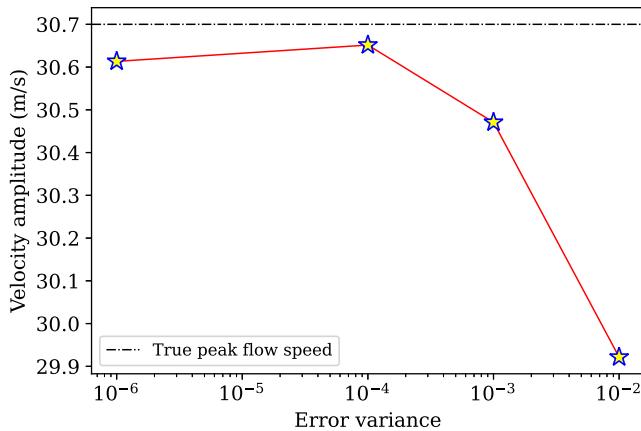
#### 4. Discussion and Conclusion

In this study, we utilized the SFT model to infer the peak meridional flow speed within a data assimilation framework of EnKF. For this purpose, we developed a longitudinally averaged SFT model experiment and evolved the model for  $\approx 14$  yr. In order to check the accuracy of the SFT evolution, we compared the axial dipole moment from our model to the HMI observations. We noticed that adjusting the DA parameters, e.g., assimilation cadence, the error associated with the perfect model can modify the reconstruction of the observations.

Direct observation of the photospheric magnetic field has its challenges (e.g., for high latitudes and the far side), but these measurements are essential for driving global coronal models (J. A. Linker et al. 1999; C. J. Schrijver & M. L. De Rosa 2003; C. J. Schrijver et al. 2013; B. van der Holst et al. 2014; J. T. Hoeksema et al. 2020), solar wind simulations (M. J. Owens et al. 2017), reconstruction of three-dimensional inner heliospheric states (M. J. Owens & R. J. Forsyth 2013), and space weather applications (C. N. Arge & V. J. Pizzo 2000). The polar fields are difficult to measure accurately, but are particularly important for solar cycle prediction studies (J. Jiang et al. 2010; L. Upton & D. H. Hathaway 2014; P. Bhowmik & D. Nandy 2018). Global surface-flow characteristics play an important role in modulating



**Figure 8.** Probable variation in the error associated with the perfect model and its implications for the reconstruction process. We have chosen an assimilation cadence of 60 days and an ensemble size of 20 for this experiment. Panel (a) shows the perfect model observations, i.e., magnetic fields at \$60^\circ\$N latitude for different error variance values while (b) denotes the reconstructed observations. In panel (c), we plot the corresponding state vectors for different error variance values. We notice a deviation in the reconstruction from the perfect model when the associated error variance is higher. The inset plot highlights the initial oscillatory pattern of the state vectors for different experimental cases before converging toward the true state (denoted by the solid black curve).



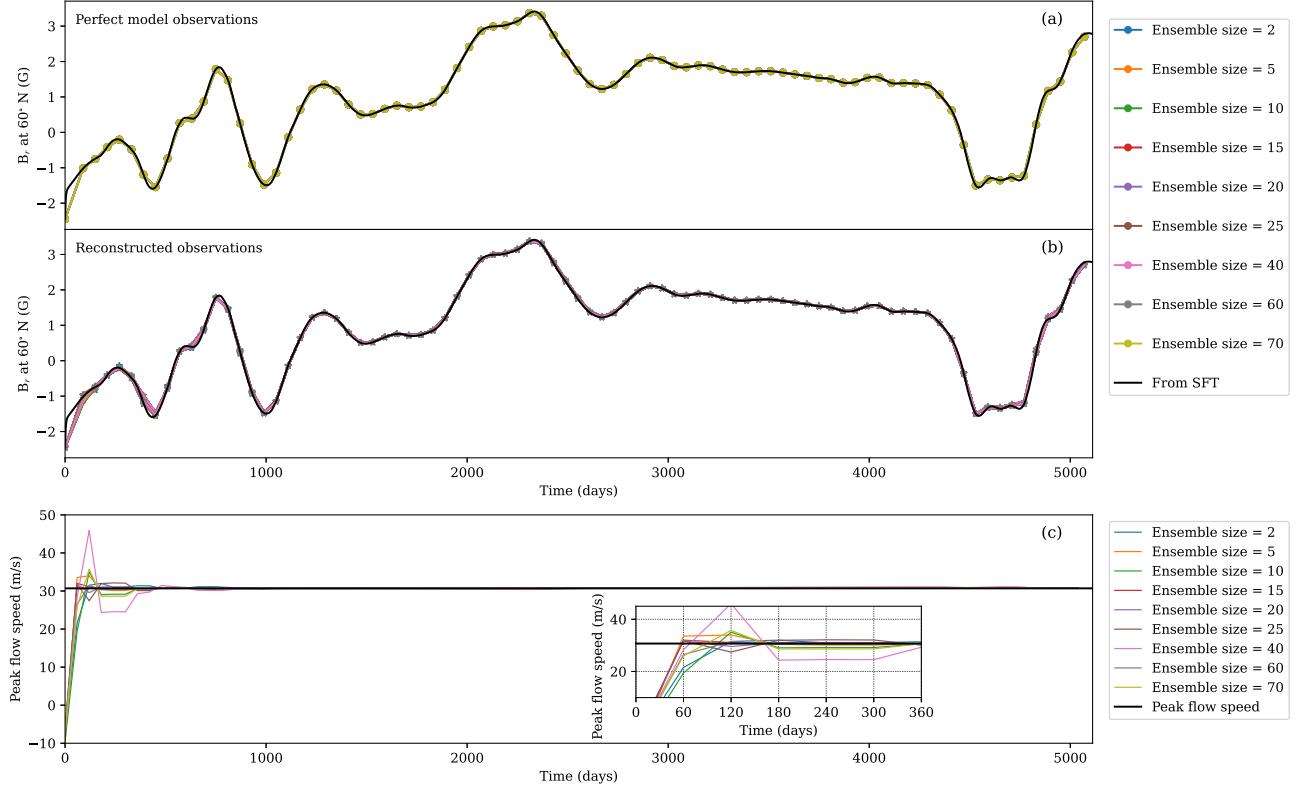
**Figure 9.** Ensemble mean value of the state vector at the end of the DA experiment for different error variances. The assimilation cadence is chosen to be 60 days and the ensemble size is 20 for this experiment. DA experiment is sensitive to the degree of error associated with the perfect model. The black dashed curve shows the true state value for reference.

the polar fields. This warrants the usage of SFT simulations constrained with available magnetic field observations—utilizing the theoretically modeled flow distribution that best matches the measurements—to produce true global photospheric magnetic field distributions that include polar field estimates. Apart from the global flow characteristics, the active region flux and the mid-latitude poleward flow speed show a negative correlation, which might be attributed to the field-dependent converging flow toward flux concentrations (X. Sun et al. 2015). This firmly establishes the

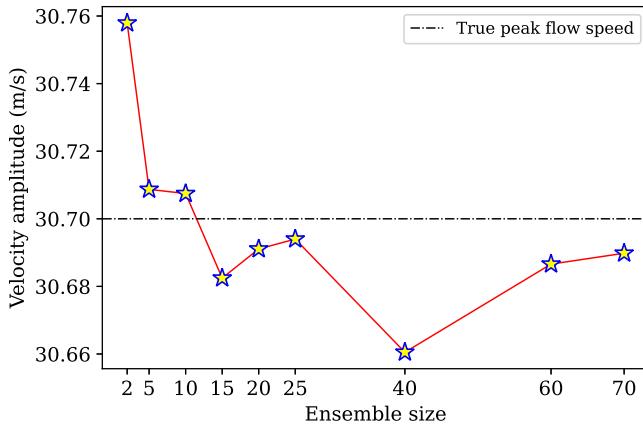
requirement for techniques to infer the flow characteristics from observational proxies.

Utilizing our EnKF + SFT method, we are able to constrain the surface-flow properties with magnetic field measurements, which will improve the current global surface flux distribution modeling efforts. Therefore, we built an OSSE setup where one of the flow parameters, the peak meridional flow speed, is estimated from synthetic observations of magnetic fields.

Several control parameters within the assimilation framework can be adjusted to improve the assimilation process. When such an algorithm is applied to a system of observed magnetic fields, there are constraints on the observation cadence (e.g., Carrington maps of magnetic fields are only available at a cadence of 28 days) and sometimes there are data gaps from the observing stations. With the EnKF setup, we can assimilate available observations having different or irregular cadences without compromising the reconstruction process. It is important to note that the forward operator (SFT model in our case) requires sufficient time to relax. Hence, if the flow properties are modified within a shorter timescale (e.g., each day) it will lead to inconsistent results. This is known as ensemble collapse as illustrated in Figure 7. Hence the optimal assimilation cadence varies depending upon the properties of the forward operator. Apart from the assimilation cadence, we can also control the reconstruction process for different degrees of error associated with the observations. The algorithm also provides further constraints in terms of ensemble size for computing the state vector. In our experiment, the EnKF algorithm reconstructs the state vector within the first few



**Figure 10.** Reconstruction using different ensemble sizes. We have chosen an assimilation cadence of 60 days and an error variance of  $10^{-6}$  for this experiment. (a) Perfect model observations, i.e., magnetic field at  $60^\circ$ N latitude with varying ensemble sizes. (b) Reconstructed magnetic fields after the DA experiment for respective ensemble sizes. (c) Corresponding evolution of state vectors, i.e., the peak meridional flow speed as they approach toward the true state. Although the assimilation cadence is 60 days, increasing ensemble size improves the reconstructed observations. In the inset plot shown in the bottom panel, we plot the initial iterations of the DA experiment showing the oscillatory pattern of the state vectors around the true state (denoted by the solid black curve).



**Figure 11.** Ensemble mean value of the state vector at the end of the DA experiment for different ensemble sizes. The assimilation cadence is chosen to be 60 days and the error variance is  $10^{-6}$  for this experiment. The final state vector approaches the true state for a larger ensemble size even for a higher assimilation cadence. The black dashed curve denotes the true state value for reference.

months of simulation where the impact of the choice of ensemble size on the assimilation process is amplified. The final values of the state vector corresponding to different experiment cases show negligible deviation from the true state. However, the SFT-generated magnetic fields for different such choices still closely track the perfect model output. We believe, for a global flow reconstruction, such minor changes (e.g., for the case of ensemble size 40 in Figure 11) may impact the modeling of magnetic field distribution locally.

Provided the versatility of the EnKF technique, usage of the same within the solar physics community has not been widely explored. These tools have been employed by various authors in a few contexts, such as solar cycle prediction, parameter reconstruction, and estimating solar wind properties (M. Dikpati et al. 2016; I. N. Kitashvili 2016; M. Lang et al. 2017). However, a golden era for data assimilation in solar models is coming soon. In a recent ground-breaking effort, H. Turner et al. (2023) demonstrated that EnKF assimilation of near-real-time solar wind data into the Wang–Sheeley–Arge (WSA) model can improve the 5 days lead-time forecast by 15% if L5 data are included in the assimilation along with L1 data.

The versatility and applicability of the EnKF technique across a range of scientific and practical problems demonstrates its effectiveness and calls for ongoing development in these areas. Nevertheless, the method does have several limitations. The EnKF assumes that the distributions of the forecast and observational errors are Gaussian, which can be a significant limitation when dealing with non-Gaussian processes. With a limited ensemble size, the EnKF can suffer from sampling errors, which can lead to inaccurate estimates of the state and error covariances. To improve the accuracy, a large ensemble size can be used, but then the EnKF can be computationally expensive, especially for large-scale systems, because it requires the integration of a large number of ensemble members through the model. Two assumptions in the EnKF tool, namely, (i) the model is “perfect” (i.e., all physics are sufficiently included to reproduce the observations) and (ii) the covariance coefficients of the state vectors are considered to be

linearly independent, may lead to inaccuracies in a highly nonlinear system (A. Carrassi et al. 2018).

In this paper, we demonstrated the successful implementation of the EnKF algorithm with an SFT model with a parameterized meridional flow profile. In future work, we plan to explore the usage of our technique to infer two-dimensional spatiotemporal descriptions of the global flows on the photosphere inferred from the magnetic field observations.

### Acknowledgments

This work is supported by NASA LWS award 80NSSC20K0184 to Stanford University and a subaward to the University of Hawai'i. S.D. and X.S. additionally acknowledge support from the State of Hawai'i. This work is also supported by the NSF National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under cooperative agreement 1852977. M.D. acknowledges partial support from various NASA grants, such as NASA LWS award 80NSSC20K0355 to NCAR and NASA-HSR award 80NSSC21K1676 to NCAR. S.S.M., Y.L., M.L.D., and J.T.H. were also partly supported by the COFFIES: NASA-DRIVE Science Center award 80NSSC22M0162. S.S.M., Y.L., and J.T.H. also acknowledge support from NASA contract NAS5-02139 (HMI) to Stanford University. Data used in this work are courtesy of NASA/SDO and the HMI science teams.

### ORCID iDs

- Soumyaranjan Dash <https://orcid.org/0000-0003-0103-6569>  
 Marc L. DeRosa <https://orcid.org/0000-0002-6338-0691>  
 Mausumi Dikpati <https://orcid.org/0000-0002-2227-0488>  
 Xudong Sun (孙旭东) <https://orcid.org/0000-0003-4043-616X>  
 Sushant S. Mahajan <https://orcid.org/0000-0003-1753-8002>  
 Yang Liu (刘扬) <https://orcid.org/0000-0002-0671-689X>  
 J. Todd Hoeksema <https://orcid.org/0000-0001-9130-7312>

### References

- Anderson, J., Hoar, T., Raeder, K., et al. 2009, *BAMS*, 90, 1283  
 Anderson, J. L. 2001, *MWRv*, 129, 2884  
 Anderson, J. L. 2010, *MWRv*, 138, 4186  
 Arge, C. N., Henney, C. J., Koller, J., et al. 2010, in AIP Conf. 1216, 12th Int. Solar Wind Conf., ed. M. Maksimovic et al. (Melville, NY: AIP), 343  
 Arge, C. N., & Pizzo, V. J. 2000, *JGR*, 105, 10465  
 Bélanger, E., Vincent, A., & Charbonneau, P. 2007, *SoPh*, 245, 141  
 Bhowmik, P., & Nandy, D. 2018, *NatCo*, 9, 5209  
 Brun, A. S. 2007, *AN*, 328, 329  
 Butala, M. D., Hewett, R. J., Frazin, R. A., & Kamalabadi, F. 2010, *SoPh*, 262, 495  
 Cameron, R., & Schüssler, M. 2007, *ApJ*, 659, 801  
 Carrassi, A., Bocquet, M., Bertino, L., & Evensen, G. 2018, *WIRCC*, 9, e535  
 Chen, Y., Shen, Z., & Tang, Y. 2022, *JAMES*, 14, e2022MS003106  
 Dash, S., Derosa, M. L., Dikpati, M., et al. 2024, sr-dash/SFT-1D: Final Updates to the repo. v0.0.2, Zenodo, doi:[10.5281/zenodo.13831085](https://doi.org/10.5281/zenodo.13831085)  
 DeVore, C. R., Boris, J. P., & Sheeley, N. R. 1984, *SoPh*, 92, 1  
 Dikpati, M., Anderson, J. L., & Mitra, D. 2014, *GeoRL*, 41, 5361  
 Dikpati, M., Anderson, J. L., & Mitra, D. 2016, *ApJ*, 828, 91  
 D'Silva, S., & Choudhuri, A. R. 1993, *A&A*, 272, 621  
 Evensen, G. 1994, *JGR*, 99, 10  
 Evensen, G. 2003, *OcDyn*, 53, 343  
 Fournier, A., Nerger, L., & Aubert, J. 2013, *GGG*, 14, 4035  
 Ha, S., Snyder, C., Skamarock, W. C., Anderson, J., & Collins, N. 2017, *MWRv*, 145, 4673  
 Hathaway, D. H., & Rightmire, L. 2010, *Sci*, 327, 1350  
 Hickmann, K. S., Godinez, H. C., Henney, C. J., & Arge, C. N. 2015, *SoPh*, 290, 1105  
 Hoeksema, J. T., Abbott, W. P., Bercik, D. J., et al. 2020, *ApJS*, 250, 28  
 Hoteit, I., Abualnaja, Y., Afzal, S., et al. 2021, *BAMS*, 102, E99  
 Iijima, H., Hotta, H., Imada, S., Kusano, K., & Shiota, D. 2017, *A&A*, 607, L2  
 Jha, B. K., & Upton, L. A. 2024, *ApJL*, 962, L15  
 Jiang, J., İşik, E., Cameron, R. H., Schmitt, D., & Schüssler, M. 2010, *ApJ*, 717, 597  
 Jouve, L., Brun, A. S., & Talagrand, O. 2011, *ApJ*, 735, 31  
 Kitiashvili, I., & Kosovichev, A. G. 2008, *ApJL*, 688, L49  
 Kitiashvili, I. N. 2016, *ApJ*, 831, 15  
 Komm, R. W., Howard, R. F., & Harvey, J. W. 1993, *SoPh*, 147, 207  
 Lang, M., Browne, P., van Leeuwen, P. J., & Owens, M. 2017, *SpWea*, 15, 1490  
 Leighton, R. B. 1964, *ApJ*, 140, 1547  
 Linker, J. A., Mikić, Z., Biesecker, D. A., et al. 1999, *JGR*, 104, 9809  
 Mahajan, S. S., Hathaway, D. H., Muñoz-Jaramillo, A., & Martens, P. C. 2021, *ApJ*, 917, 100  
 Mahajan, S. S., Sun, X., & Zhao, J. 2023, *ApJ*, 950, 63  
 Matsuo, T., Lee, I. T., & Anderson, J. L. 2013, *JGRA*, 118, 1339  
 Nandy, D. 2021, *SoPh*, 296, 54  
 Owens, M. J., & Forsyth, R. J. 2013, *LRSP*, 10, 5  
 Owens, M. J., Lockwood, M., & Riley, P. 2017, *NatSR*, 7, 41548  
 Pal, S., & Nandy, D. 2024, *MNRAS*, 531, 1546  
 Parker, E. N. 1955, *ApJ*, 121, 491  
 Pedatella, N. M., Raeder, K., Anderson, J. L., & Liu, H. L. 2013, *GeoRL*, 40, 4469  
 Petrovay, K., & Talafha, M. 2019, *A&A*, 632, A87  
 Scherrer, P. H., Schou, J., Bush, R. I., et al. 2012a, *SoPh*, 275, 207  
 Schou, J., Scherrer, P. H., Bush, R. I., et al. 2012b, *SoPh*, 275, 229  
 Schrijver, C. J., & De Rosa, M. L. 2003, *SoPh*, 212, 165  
 Schrijver, C. J., Title, A. M., Yeates, A. R., & DeRosa, M. L. 2013, *ApJ*, 773, 93  
 Sun, X., Hoeksema, J. T., Liu, Y., & Zhao, J. 2015, *ApJ*, 798, 114  
 Svedin, A., Cuellar, M. C., & Brandenburg, A. 2013, *MNRAS*, 433, 2278  
 Turner, H., Lang, M., Owens, M., et al. 2023, *SpWea*, 21, e2023SW003457  
 Upton, L., & Hathaway, D. H. 2014, *ApJ*, 780, 5  
 van der Holst, B., Sokolov, I. V., Meng, X., et al. 2014, *ApJ*, 782, 81  
 Yeates, A. R. 2020, *SoPh*, 295, 119  
 Yeates, A. R., Cheung, M. C. M., Jiang, J., Petrovay, K., & Wang, Y.-M. 2023, *SSRv*, 219, 31