ORIGINAL ARTICLE

Data Assimilation

Two Methods for Data Assimilation of Wind Direction

Ian Grooms

Department of Applied Mathematics, University of Colorado, Boulder, Colorado, 80309, USA

[1]Department of Applied Mathematics, University of Colorado, Boulder, Colorado, 80309, USA

Correspondence
Ian Grooms, Department of Applied Mathematics, University of Colorado, Boulder, Colorado, 80309, USA
Email: ian.grooms@colorado.edu

Funding information
US National Science Foundation, Division of Mathematical Sciences, Grant Number: 2152814

Wind direction observations are instrumental weather records that hold promise for improving historical weather reanalyses and extending them deeper into the past. Two methods are developed for assimilating wind direction observations. The first uses a linear observation model with Gaussian additive error, and is thus amenable to use in standard EnKF and variational frameworks. The second is nonlinear and non-Gaussian, and is based on a two-step approach for sampling from the Bayesian posterior. Both methods are tested in the context of an idealized two-dimensional model of turbulent fluid dynamics. The nonlinear, non-Gaussian method assimilating only wind direction observations performs as well as an EnKF assimilating only pressure observations, whereas the first method based on the linear model provides no benefit when assimilating only wind direction observations. The method based on the linear model performs well when paired with other observations, e.g. of pressure, since it performs best when the forecast of wind direction is not far from correct.

KEYWORDS
data assimilation, wind direction, ensemble, historical observations

1 | INTRODUCTION

The study of how changing climate changes weather patterns, especially extreme weather, requires some knowledge of historical weather patterns. Historical reanalyses like the Twentieth Century Reanalysis (20CR Compo et al., 2011; Giese et al., 2016; Slivinski et al., 2019) and the European Centre for Medium-Range Weather Forecasts (ECMWF) reanalyses ERA-20C (Poli et al., 2016) and CERA-20C (Laloyaux et al., 2018) attempt to reconstruct global weather patterns as far back as the nineteenth century using data assimilation (DA) – a class of methods for combining observational data with a forecast model to estimate the state and parameters of a dynamical system.

The further back in time one goes, the fewer are the observations available for use in a reanalysis. One class of weather observations that extends deeply into the historical record comes from the written reports of mariners. Records from voyages of the English East India Company, for example, stretch back to the seventeenth century (Brohan et al., 2012). Many of these records have been digitized and are available through the International Comprehensive Ocean-Atmosphere Data Set (ICOADS Freeman et al., 2017). Wind speed and direction are common observations, but early wind speed measurements were reported using qualitative language. The introduction of the Beaufort scale for wind speed in the early nineteenth century enables qualitative descriptions of wind speed to be converted to a quantitative scale, but early wind speed records have significantly greater uncertainty than wind direction measurements, and are not considered instrumental (Prieto et al., 2005; de Paula Gomez-Delgado et al., 2019). Wind direction measurements were recorded with high precision using accurate magnetic compasses with as many as 64 points. Unfortunately, wind direction is a strongly nonlinear function of the wind field, which makes it difficult to assimilate using standard ensemble Kalman filter (EnKF) data assimilation methods that work best for linear observations with Gaussian errors (Evensen, 2009). The goals of this investigation are to develop ensemble methods that can assimilate wind direction measurements, and to demonstrate, at least in an idealized model, the value in assimilating such measurements. These goals are in service of the larger goal of ultimately improving historical reanalyses and perhaps motivating the digitization of more historical observations of wind direction.

The plan of the paper is as follows. Two ensemble-based approaches to assimilating wind direction are described in section 2. The idealized dynamical model used in the tests is described in section 3. The data assimilation experimental configuration is presented in section 4, and the results of those experiments are presented and discussed in 5. Conclusions are offered in section 6. Figure data and simulation code are available (Grooms, 2023).

2 | WIND DIRECTION ENSEMBLE DATA ASSIMILATION

We begin by setting notation. The state of the dynamical system is denoted $\boldsymbol{x}$, with a subscript j to denote the value of $\boldsymbol{x}$ at time t_j. Observational information at time t_j is denoted $\boldsymbol{y}_j$. Our uncertainty about the state of the system before taking observational information into account is described by a Bayesian prior distribution; the probability density function (pdf) associated with this distribution is denoted $[\boldsymbol{x}]$ and a random variable with pdf $[\boldsymbol{x}]$ is denoted $\boldsymbol{X}$. The observation $\boldsymbol{y}$ is a draw from an observational distribution with pdf $[\boldsymbol{y}|\boldsymbol{x}]$. Viewed as a function of $\boldsymbol{y}$ this is the pdf of the observational distribution; viewed as a function of $\boldsymbol{x}$ it is proportional to the Bayesian likelihood. The goal of ensemble data assimilation is to draw samples $\boldsymbol{x}^{(n)}$, which together form an ensemble $\{\boldsymbol{x}^{(n)}\}_{n=1}^{N}$, from some distribution that approximates the Bayesian posterior $[\boldsymbol{x}|\boldsymbol{y}]$.

Ensemble Kalman filters (EnKFs) approximate the joint distribution of $\boldsymbol{X}$ and $\boldsymbol{Y}$ as Gaussian, which implies that the Bayesian posterior (which is simply a conditional of the joint distribution) is also Gaussian with well-known formulas for the posterior mean and covariance. To wit, if the parameters of the joint distribution are

$$\begin{aligned} \mathbb{E}[\boldsymbol{X}] &= \boldsymbol{\mu}_x && \text{(1a)}\\ \mathbb{E}[\boldsymbol{Y}] &= \boldsymbol{\mu}_y && \text{(1b)}\\ \mathrm{Cov}[\boldsymbol{X}] &= \mathbf{C}_x && \text{(1c)}\\ \mathrm{Cov}[\boldsymbol{Y}] &= \mathbf{C}_y && \text{(1d)}\\ \mathrm{Cov}[\boldsymbol{X},\boldsymbol{Y}] &= \mathbf{C}_{xy} && \text{(1e)} \end{aligned}$$

then the posterior mean and covariance are

$$\mathbb{E}[\boldsymbol{X}|\boldsymbol{Y}=\boldsymbol{y}] = \boldsymbol{\mu}_x + \mathbf{C}_{xy}\mathbf{C}_y^{-1}(\boldsymbol{y}-\boldsymbol{\mu}_y) \tag{2}$$

$$\mathrm{Cov}[\boldsymbol{X}|\boldsymbol{Y}=\boldsymbol{y}] = \mathbf{C}_x - \mathbf{C}_{xy}\mathbf{C}_y^{-1}\mathbf{C}_{xy}^T. \tag{3}$$

The Kalman filter (KF) formulas on which EnKFs are commonly based make a further assumption about the relationship between the state and observations, namely that

$$\boldsymbol{Y} = \mathbf{H}\boldsymbol{X} + \boldsymbol{\epsilon} \tag{4}$$

where $\boldsymbol{\epsilon}$ is a centered Gaussian independent of $\boldsymbol{X}$ and with covariance $\mathbf{R}$, and $\mathbf{H}$ is a matrix. This assumption, which is necessary for the KF but not for EnKFs, implies

$$\boldsymbol{\mu}_y = \mathbf{H}\boldsymbol{\mu}_x \tag{5}$$

$$\mathbf{C}_y = \mathbf{H}\mathbf{C}_x\mathbf{H}^T + \mathbf{R} \tag{6}$$

$$\mathbf{C}_{xy} = \mathbf{H}\mathbf{C}_x \tag{7}$$

which allows the recovery of more familiar KF update formulas.

In the context of wind direction observations, $\boldsymbol{y}$ is an angle and $\boldsymbol{Y}$ is a circular random variable. The assumption of joint Gaussianity at the heart of EnKFs is a severe limitation in this context, since circular random variables are simply not Gaussian. Indeed, when the measured wind direction is recorded on a 16, 32, or 64 point compass, $\boldsymbol{Y}$ is also a discrete random variable, which further underscores that it is not Gaussian.

In principle an EnKF assimilating wind direction can be implemented using a nonlinear observation model

$$\boldsymbol{y} = \boldsymbol{h}(\boldsymbol{x}) + \boldsymbol{\epsilon} \tag{8}$$

where $\boldsymbol{h}$ is an arctangent that maps the wind vector to the wind direction. Though algorithmically straightforward, this approach implicitly treats $\boldsymbol{Y}$ as a Gaussian variable, rather than a circular one, by relying on the usual formulas for means and covariances of linear variables. The performance of this type of method is expected to be erratic at best. For example, consider an ensemble of two directions: $\pm 7\pi/8$ radians, i.e. west by northwest and west by southwest. A standard EnKF approach to nonlinear observation operators will treat the mean wind direction as eastward when it should be westward, and will consider the ensemble spread to be large when it is in fact small.

This section develops two ensemble data assimilation approaches for assimilating wind direction measurements.

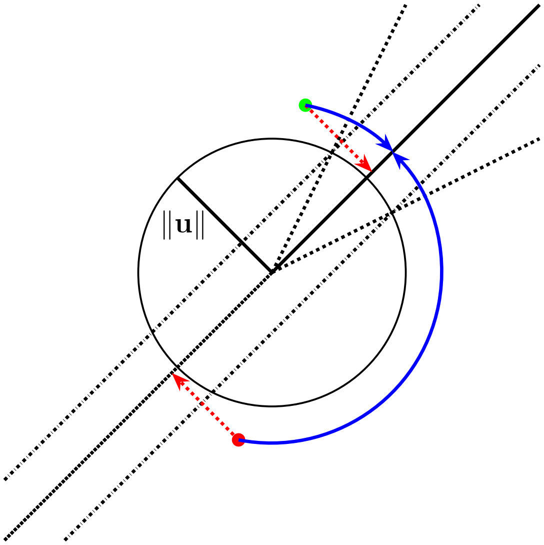

FIGURE 1 The solid line emanating from the center of the circle indicates the observed wind direction, while the dashed lines flanking the solid line indicate the uncertainty in the measurement of the wind direction. The dotted lines indicate the linear observation model: The central dotted line is parallel to the observed wind direction, and the flanking dotted lines represent the uncertainty in the magnitude of the wind perpendicular to the observed wind direction. The magnitude of the uncertainty in the linear model is chosen so that it matches the magnitude of the uncertainty in the wind direction at a particular wind speed; this wind speed where the uncertainties match is the radius $\|\boldsymbol{u}\|$ of the circle. The red lines indicate the way in which a wind vector (solid circles) is adjusted by the EnKF approach, while the blue lines indicate the way in which a wind vector is adjusted by the TSEF approach. The green solid circle is adjusted similarly by the EnKF and TSEF approaches, while the red solid circle is adjusted correctly by the TSEF and incorrectly by the EnKF.

The first, presented in section 2.1, is based on an observation model for wind direction that takes the form (4), which allows the use of standard EnKF or variational methods. The second, presented in section 2.2, uses a two-step approach to ensemble data assimilation based on the seminal approach of Anderson (2003), as expanded recently by Grooms (2022). The two approaches are fundamentally distinguished by the way in which they deal with the wind direction observations. In the first approach, the wind direction observations are assimilated by constraining to zero the component of wind perpendicular to the observed wind direction. In the second approach, the wind direction observations are assimilated by rotating the wind direction towards the observed direction. The mathematical structure of the first approach makes it amenable to implementation via standard EnKF methods, while the structure of the second approach requires a more sophisticated implementation via a two-step ensemble filter.

2.1 | EnKF for Wind Direction

The goal of this section is to develop an observation model that approximates the true likelihood associated with a wind direction measurement, and that is amenable to incorporation into standard EnKF-type methods. Let $\boldsymbol{u} = (u, v)^T$ be a two-component horizontal wind vector with eastward component u and northward component v. If $\theta \in [-\pi, \pi)$

is the direction of $\boldsymbol{u}$ measured in radians counterclockwise from east, then

$$\boldsymbol{u} = \|\boldsymbol{u}\| \left(\begin{array}{c} \cos(\theta) \\ \sin(\theta) \end{array} \right). \tag{9}$$

If $\boldsymbol{u}$ points in the direction θ, then

$$\boldsymbol{u} \cdot \left(\begin{array}{c} -\sin(\theta) \\ \cos(\theta) \end{array} \right) = 0, \tag{10}$$

although the converse is not true; (10) only guarantees that $\boldsymbol{u}$ has angle ±θ.

Equation (10) is reminiscent of the linear observation model (4) commonly used in EnKFs, where

$$\boldsymbol{y} = 0 \tag{11a}$$

$$\mathbf{H} = [-\sin(\theta),\ \cos(\theta)] \tag{11b}$$

$$\boldsymbol{x} = \left(\begin{array}{c} u \\ v \end{array} \right). \tag{11c}$$

It is interesting to note that this version of the observation model flips the usual paradigm upside down. In the usual paradigm the observation matrix $\mathbf{H}$ is fixed and the value of the observation is $\boldsymbol{y}$. In the version adopted here the value of $\boldsymbol{y}$ is always zero, while the observation matrix $\mathbf{H}$ depends on the observed angle θ.

What's missing here is an additive observation error $\boldsymbol{\epsilon}$. The choice of error variance is somewhat ad hoc, since the observation model is only an approximation. The approach to specifying $\boldsymbol{\epsilon}$ developed here is illustrated in Figure 1. The true angular likelihood is illustrated schematically in Figure 1 as follows. A solid line emanating from the origin indicates the observed wind direction; uncertainty in the true wind direction (as distinct from the observed wind direction which may contain slight errors) is illustrated by dashed lines emanating from the origin and flanking the solid line. The approximate likelihood associated with the linear observation model (11) says that the component of wind perpendicular to the observed direction is zero, with some uncertainty. This is illustrated in Figure 1 as follows. The central dotted line is parallel to the observed wind direction, while the flanking dash-dotted lines indicate uncertainty in the magnitude of the component of the wind vector perpendicular to the observed wind direction. The uncertainty in the two likelihoods intersects at a particular flow speed $\|\boldsymbol{u}\|$. In the figure this flow speed is the radius of the circle that passes through the points where the dashed lines (uncertainty in the true likelihood) intersect the dash-dotted lines (uncertainty in the linear model). Choosing an observation error variance in the linear model is thus equivalent to choosing a flow speed such that the uncertainty in the linear model and angular model match.

Based on this geometrical motivation, the observational uncertainty in the linear model is set as follows. In the context of an EnKF, an ensemble of prior (forecast) wind vectors $\{\boldsymbol{u}^{(n)}\}_{n=1}^{N}$ is available; in order to make the linear observation model accurate for a typical wind speed, the effective wind speed is set to

$$\|\boldsymbol{u}\| = \left(\frac{1}{N} \sum_{n=1}^{N} \left\| \boldsymbol{u}^{(n)} \right\|^2 \right)^{1/2}. \tag{12}$$

To get the observation error standard deviation in the linear observation model, we need the observational uncertainty in the angular measurement of wind direction. Naturally, historical observations of wind direction do not come with built-in quantification of their uncertainty, i.e. the true likelihood is unknown. The approach taken here, which can easily be modified, is to say that the standard error in measurements recorded on a compass with N_θ points is equal to half of a compass increment, i.e. π/N_θ. The arc length of a circular segment from the true wind direction to the true wind direction plus one standard error at a wind speed of $\|\boldsymbol{u}\|$ is $\|\boldsymbol{u}\| \times \pi/N_\theta$. The observational error variance in the linear model is thus set to the square of this arc length

$$\text{Var}[\epsilon] = \frac{1}{N}\left(\frac{\pi}{N_\theta}\right)^2 \sum_{n=1}^{N} \|\boldsymbol{u}^{(n)}\|^2. \tag{13}$$

With this linear observation model, the wind direction observation can be assimilated by any one of the wide variety of EnKF methods available in the literature, or even using variational methods. The EnKF experiments reported here use the serial square root assimilation scheme of Whitaker and Hamill (2002). To distinguish this specific EnKF algorithm from other EnKFs, the serial square root assimilation will hereafter be denoted ESRF (Ensemble Square Root Filter). In a situation where wind speed measurements are available but with a much lower precision than wind direction, e.g. because the wind speed observation is simply the phrase 'light wind' (Brohan et al., 2012), the component of wind in the observed direction could be assimilated using an appropriately large error variance; this would be equivalent to observing both components of wind, but with a non-diagonal 2×2 observation error covariance matrix. Some historical observations of wind direction are unclear about whether the recorded direction is the direction from which the wind blows or to which the wind blows (Freeman et al., 2017); since the linear observation model is exactly the same for angles $\theta \pm \pi$, this method can still use wind direction observations that are subject to this ambiguity.

2.2 | A Two-Step Ensemble Filter

Anderson (2003) developed a two-step approach to assimilating scalar observations that are nonlinearly related to the state by a relation similar to the linear observation model (4), namely

$$\boldsymbol{Y} = \boldsymbol{h}(\boldsymbol{X}) + \epsilon. \tag{14}$$

The first step of this two-step approach consists of a Bayesian estimation of $\boldsymbol{h}(\boldsymbol{X})$, while the second step consists of a simple linear regression from $\boldsymbol{h}(\boldsymbol{X})$ back to $\boldsymbol{X}$. In the original implementation, the first step used the ensemble adjustment Kalman filter (EAKF Anderson, 2001); more recently, methods like the Rank Histogram Filter (RHF Anderson, 2010) and the GIGG-EnKF (Bishop, 2016) have been developed for the first step to relax the Gaussian approximation of the original EAKF. The second step typically uses simple linear regression, although generalized regression methods have been developed (e.g. Anderson, 2019). A key aspect of these methods is that they are algorithmically similar to EnKFs, and can be used efficiently with large-scale geophysical models (Anderson et al., 2009).

Grooms (2022) showed how two-step ensemble filters (TSEFs) are related to Bayesian estimation. A new random variable $\boldsymbol{Z}$ is introduced that has the property

$$[\boldsymbol{y}|\boldsymbol{x},\boldsymbol{z}] = [\boldsymbol{y}|\boldsymbol{z}]. \tag{15}$$

In the context of the observation model (14), Anderson (2003) chose to use $\boldsymbol{Z} = \boldsymbol{h}(\boldsymbol{X})$, but this is not the only possible choice of $\boldsymbol{Z}$ that satisfies the property (15). An illustrative but impractical alternative is to set $\boldsymbol{Z} = \boldsymbol{X}$; another choice is made below in the specific context of observations of wind direction. The introduction of this new variable allows the pdf of the Bayesian posterior to be written as

$$[\boldsymbol{x}|\boldsymbol{y}] = \int [\boldsymbol{z}|\boldsymbol{y}][\boldsymbol{x}|\boldsymbol{z}]\mathrm{d}\boldsymbol{z}. \tag{16}$$

Two-step ensemble filters sample from this distribution by first sampling an ensemble $\{\boldsymbol{z}^{(n)}\}_{n=1}^N$ from $[\boldsymbol{z}|\boldsymbol{y}]$, and then sampling an ensemble $\{\boldsymbol{x}^{(n)}\}_{n=1}^N$ where $\boldsymbol{x}^{(n)}$ is a sample from $[\boldsymbol{x}|\boldsymbol{z}^{(n)}]$.

In the context of wind direction assimilation, let $\boldsymbol{z} = \boldsymbol{u}$ and $\boldsymbol{y} = \theta$. It is convenient to change to polar coordinates for the wind vector, $(u, v) \mapsto (\rho, \phi)$ where the notation ϕ has been adopted to distinguish a general angle ϕ from the observed value θ. Note that in the context of the first step of the two-step filter, which is being described here, the choices $\boldsymbol{z} = \boldsymbol{u}$ and $z = \phi$ work equally well; the choice $\boldsymbol{z} = \boldsymbol{u}$ is made here because it leads to a better second step, which is described at the end of section 4. The posterior in polar coordinates is

$$[\rho, \phi|\theta] = \frac{[\theta|\rho, \phi]}{[\theta]}[\rho, \phi]. \tag{17}$$

While in general there is no reason to assume that ρ and ϕ are independent in the prior distribution $[\rho, \phi] = [\phi][\rho]$, it is a highly convenient approximation, which is adopted here. Further assuming that the likelihood of observing the wind direction θ does not depend on the wind speed ρ, the posterior simplifies to

$$[\rho, \phi|\theta] = \left(\frac{[\theta|\phi]}{[\theta]}[\phi]\right)[\rho] = [\phi|\theta][\rho]. \tag{18}$$

The assumption that ρ and ϕ are independent in the prior implies that the observation of wind direction has no impact on the distribution of wind speed. This assumption could potentially be relaxed by using the methods of Murphy et al. (2022), who, in a context different from data assimilation, model the conditional distribution of wind speed given wind direction $[\rho|\phi]$ as a Weibull distribution whose parameters depend on ϕ.

With these simplifying assumptions, the first step of the two-step process updates the direction of the wind vectors at the location of the observation while the second step uses regression to push the local changes in the wind vector back to all the other state variables. In this paper, the second step of the two-step process uses linear regression, as in the two-step EnKF of Anderson (2003). The first step is accomplished using a probability integral transform, which is the same idea that underlies the RHF and the Quantile-Conserving Ensemble Filter Framework (QCEFF; Anderson, 2022). If F_- and F_+ are the cumulative distribution functions (cdfs) of the prior and posterior, respectively, then the random variable $X_+ = F_+^{-1}(F_-(X_-))$ is the probability integral transform of X_- (which is distributed according to the prior), and X_+ is distributed according to the posterior. To build a scalar filter for the first step of a two-step ensemble filter based on the probability integral transform, one uses the prior ensemble and the likelihood to approximate the cdfs, and then applies the resulting approximate transform $F_+^{-1} \circ F_-$ to the prior ensemble.

In the context of wind direction, which is a circular random variable, the probability integral transform can still be used, though the definition of the cdf requires a small amount of care. For a random variable X taking values in $\mathbb{R}$, the cdf is defined to be

$$F_X(x) = \int_{-\infty}^{x} [X = \xi]\mathrm{d}\xi \tag{19}$$

where ξ is a dummy integration variable and $[X = \xi]$ is the pdf of X evaluated at ξ. For a circular random variable ϕ, one can define the cdf as

$$F_\phi(\phi) = \int_{\phi_0}^{\phi} [\phi = \xi] \mathrm{d}\xi \tag{20}$$

for any ϕ_0. The choice of ϕ_0 determines the domain of F_ϕ, which is $[\phi_0, \phi_0 + 2\pi]$.

To approximate the prior pdf $[\phi]$, we use kernel density estimation with a von Mises kernel (Mardia, 1975)

$$K(\phi) = \frac{\exp(\kappa \cos(\phi))}{2\pi I_0(\kappa)} \tag{21}$$

where I_0 is the modified Bessel function of the first kind and order 0. The parameter κ controls the width of the kernel; for large κ and for $\phi \in (-\pi, \pi)$ the kernel approximates a normal distribution with mean zero and variance κ^{-1}. The prior pdf is thus approximated from the prior ensemble as

$$[\phi] \approx \frac{1}{N} \sum_{n=1}^{N} K\left(\phi - \phi^{(n)}\right) \tag{22}$$

where $\{\phi^{(n)}\}_{n=1}^{N}$ is the ensemble of prior wind directions. The kernel bandwidth parameter κ is set using a standard bandwidth selection scheme for Gaussian kernels (Silverman, 1998), adapted to the von Mises context:

$$\kappa = -\frac{N^{2/5}}{2 \times 1.06^2 \ln(R)} \tag{23}$$

where

$$R = \left| \frac{1}{N} \sum_{n=1}^{N} e^{\mathrm{i}\phi^{(n)}} \right| . \tag{24}$$

Improved bandwidth selection schemes for von Mises kernel density estimation are discussed by Tenreiro (2022).

The domain of the prior cdf F_- that is required for the probability integral transform is chosen to be $[\theta - \pi, \theta + \pi]$ where θ is the observed wind direction. The prior cdf F_- is approximated using trapezoid rule quadrature on an equispaced grid of 257 points in $[\theta - \pi, \theta + \pi]$. Linear interpolation is used to approximate F_- between these 257 points. To obtain the posterior cdf F_+ that is required for the probability integral transform, the posterior pdf is first evaluated as the product of the prior pdf and the likelihood on the same set of 257 points in $[\theta - \pi, \theta + \pi]$. Trapezoid rule quadrature is then used to obtain an approximation to F_+ on the grid, and linear interpolation is used to fill in the intermediate values. The cdf is inverted by simply swapping the role of the ordinate and abscissa in the linear interpolation scheme used to evaluate F_+. The form of the likelihood used in the experiments is discussed in section 4.

2.3 | Example

This section presents a simple example to illustrate the differences in the first step of the EnKF and TSEF methods. In practice there would be many observations of wind direction at different locations, each of which would be serially assimilated into the state vector. The example presented here illustrates only the first step of the assimilation for a single observation of wind direction, and in this example the observed wind direction is very different from the wind

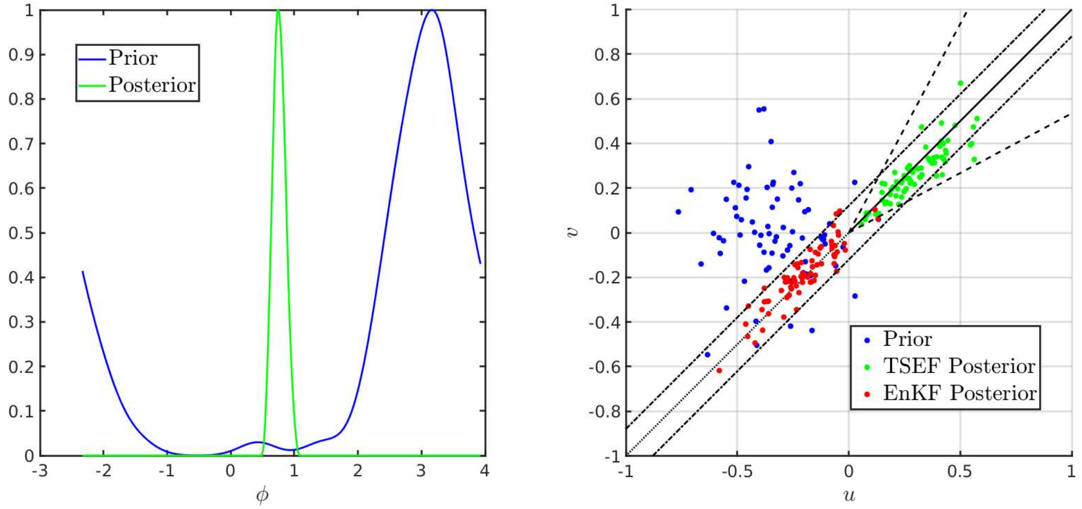

FIGURE 2 Left: The prior (blue) and posterior (green) pdfs of angle ϕ. Both are normalized to have unit maximum height so their shapes can be compared more easily. Right: A prior ensemble of wind vectors (blue), together with the posterior ensemble of wind vectors produced by the probability integral transform approach (green) and the EnKF approach (red). The solid black line is in the direction of the observed wind; the dotted black line is opposite to the direction of the observed wind; the dashed lines enclose the directions with nonzero likelihood; and the dash-dotted lines enclose the 95% confidence interval associated with the EnKF observation model.

directions in the forecast ensemble. An ensemble of $N = 72$ wind vectors are drawn from a random distribution – specifically, from randomly-chosen spatial location in the initial distribution of the experiments described in section 4. This initial ensemble is shown as blue dots in the right panel of Figure 2. The estimate of the prior pdf produced by the von Mises kernel density estimation is shown as a blue line in the left panel of Figure 2. For ease of visual comparison with the posterior, the prior pdf has been scaled to have a maximum height of 1.

The observed wind direction is set to $\theta = \pi/4$, i.e. towards the northeast. This is shown in the right panel of Figure 2 by a solid black line emanating from the origin, while the dotted black line points in the opposite direction $5\pi/4$. The posterior pdf, normalized to a maximum height of 1, is shown in green in the left panel of Figure 2; the likelihood used to form the posterior is given by Equation (31). The example uses a $N_\theta = 32$ point compass so that the likelihood is nonzero only over an interval of directions of width $3\pi/16$. The posterior is nearly equal to the likelihood in this example because the prior is so widely spread compared to the likelihood.

The probability integral transform rotates the prior wind ensemble members into posterior wind ensemble members, shown in green in the right panel of Figure 2, which is concentrated around the observed wind direction. The likelihood is identically zero for directions far from the observed value of $\pi/4$. The range of directions with nonzero likelihood is shown by a pair of dashed lines emanating from the origin in the right panel of Figure 2: The likelihood is zero outside these lines, and all of the posterior wind ensemble members lie between these lines.

The EnKF approach instead reduces the component of wind that is orthogonal to the observed wind direction; in this example the result is a posterior ensemble, shown in red in the right panel of Figure 2, whose members mostly point in the direction opposite to the observed direction. The EnKF uses an approximate observation model with an additive observation error set by Equation (13). The dash-dotted lines in the right panel of Figure 2 are located two standard deviations of the observation error above and below the observed direction, and most of the EnKF posterior ensemble members lie between these lines. This example illustrates that the EnKF approach does not work well for prior ensemble members that are pointing in the wrong direction.

To understand this result, Figure 1 illustrates how the ESRF and TSEF approaches update the wind direction vector in the first step of the two-step process. For a prior wind vector that is already close to the true observed wind direction (green dot) the TSEF rotates the wind vector without changing its amplitude (blue arc), while the ESRF projects the wind vector towards the right direction while minimally changing its amplitude (red arrow). The difference between the two methods is small when the prior wind vector is already close to the right direction. For a prior wind vector that is far from the observed direction (red dot), the TSEF again rotates the wind vector towards the observed direction without changing its amplitude (blue arc). The ESRF instead projects the wind vector in the wrong direction (red arrow).

3 | IDEALIZED MODEL CONFIGURATION

The wind direction data assimilation methods developed in the preceding section are applied here in the context of nondimensional two-dimensional incompressible vorticity dynamics on a β-plane. The vorticity ω evolves according to

$$\partial_t \omega + \mathrm{J}[\psi, \omega] + \beta \partial_x \psi = F - r_d \omega + \nu \nabla^2 \omega. \tag{25}$$

The streamfunction ψ is proportional to a geostrophic pressure anomaly and will henceforth be referred to as pressure; it is related to vorticity by $\omega = \nabla^2 \psi$. Vorticity advection is represented via the Jacobian $\mathrm{J}[\psi, \omega] = \boldsymbol{u} \cdot \nabla \omega$ where $(u, v) = (-\partial_y \psi, \partial_x \psi)$. The coefficient $\beta = 10$ is the nondimensional meridional gradient of planetary vorticity. The domain is a periodic square with a nondimensional width of 2π. To convert the model state, i.e. vorticity, to wind direction at a given location requires a sequence of steps. First the Poisson equation $\nabla^2 \psi = \omega$ is solved to find the streamfunction; then the components of velocity are obtained from the gradient of ψ via $(u, v) = (-\partial_y \psi, \partial_x \psi)$; finally, the wind direction is obtained as the argument (sometimes called the phase or angle) of the complex number $u + \mathrm{i}v$.

The forcing F is stochastic. It is Gaussian and white in time. Its spatial Fourier coefficients are nonzero only for wavenumbers with $10 \le \sqrt{k_x^2 + k_y^2} \le 12$. On these wavenumbers the amplitude of forcing is constant; the value is chosen so that the net enstrophy injection rate is unity (nondimensional). Energy and enstrophy injected by the forcing are dissipated by a linear drag term with coefficient $r_d = 0.01$ and viscosity with coefficient $\nu = 4 \times 10^{-4}$.

The spatial discretization is a Fourier spectral method with 256 Fourier modes in each direction; the Jacobian is dealiased using the 3/2-rule, i.e. 384 Fourier modes in each direction are used in the computation of the Jacobian. The spatial discretization uses the fourth-order adaptive Runge-Kutta method ARK4(3)6L[2]SA of Kennedy and Carpenter (2003). All terms except the Jacobian are treated implicitly, while the stochastic forcing is added explicitly at the end of each Runge-Kutta step. The stepsize is adjusted using a PI.3.4 control (Soderlind, 2002) with a tolerance of 10^{-3}. The code is publicly available (Grooms, 2023).

Figure 3(a) shows a snapshot of vorticity from the reference simulation. Eddies are mingled with zonal bands that correspond to zonal jets; the time- and zonal-mean velocity is shown in Figure 3(b). The energy spectrum is shown in Figure 3(c); a short inverse cascade produces a shallow spectrum between the range of forcing wavenumbers (shaded) and the peak of the energy spectrum at wavenumber 3, while the combination of linear drag and viscosity conspires to create a steep spectrum falling off from the forcing wavenumbers to the viscous dissipation range.

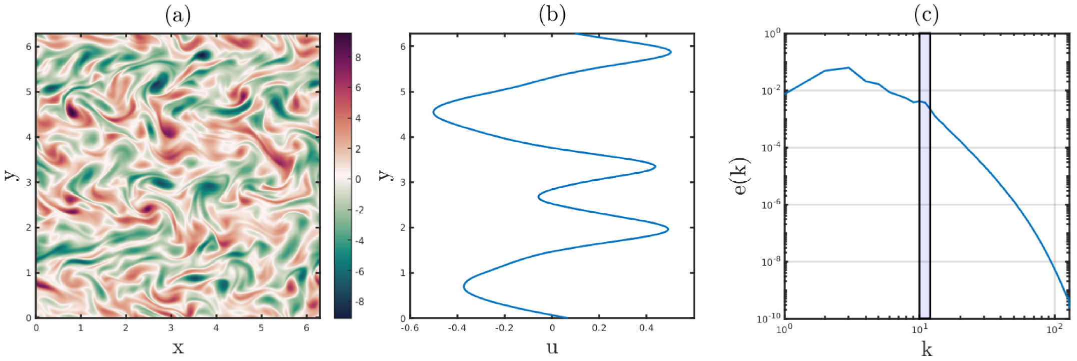

FIGURE 3 Properties of the reference experiment. (a) A snapshot of vorticity ω. (b) The time-mean and zonal-mean velocity profile. (c) The kinetic energy spectrum; the range of forcing wavenumbers is shaded. All variables are nondimensional.

4 | EXPERIMENTAL CONFIGURATION

The data assimilation experiments reported here are designed to extrapolate, as far as possible for such an idealized model, to the setting of historical reanalysis. To that end, the observing system is spatially sparse: Observations are only available at 64 locations in the domain. In one set of experiments called 'grid' the observations are available on an equispaced 8×8 grid of points. The Nyquist wavenumber for this observing grid is wavenumber 4, which is between the forcing scale at wavenumbers 10–12 and the peak of the energy spectrum at wavenumber 3. In another set of experiments called 'random' the observations are taken at 64 randomly drawn locations throughout the domain, where the locations are drawn independently at every assimilation cycle.

The reference simulation was spun up from rest to a statistically steady state, at which point 1,024 reference states separated by 0.01 nondimensional time units were saved. Observations are assimilated every 0.01 nondimensional time units for the entire sequence of 1,024 reference states. These reference states constitute the 'nature run' used to make the synthetic observations and to evaluate the accuracy of the analyses. A fixed ensemble size of $N = 72$ was used, and the ensemble members were initialized as random draws from the time series of the reference state. The initial ensemble was the same for all experiments.

We perform a baseline set of experiments that assimilate only observations of pressure ψ. The observational error variance for ψ is 0.002, which is small compared to the climatological variance of 0.03. To assess the value of assimilating wind direction, we perform a set of experiments assimilating observations of wind direction in addition to observations of ψ, and another set of experiments assimilating only observations of wind direction. All experiments in the 'grid' configuration use exactly the same observations, and all experiments in the 'random' configuration share the same observation locations and values. For all configurations we assimilate wind direction using both the EnKF and TSEF schemes described in sections 2.1 and 2.2.

Localization is accomplished by multiplying the ensemble increments with a Gaussian localization function having a radius (i.e. standard deviation) of L. The radius L is chosen adaptively in a manner motivated by the approach taken in the third version of 20CR (Slivinski et al., 2019). Slivinski et al. (2019) argued that if an observation leads to a large increment of the state variable, then one might want to allow the increments to be spread over a wide region, whereas if an observation leads to a small increment of the state variable, then one might prefer to confine the increments to a small region. One motivation for such an approach is that a large increment confined to a small location is likely to produce large gradients and consequently dynamical instability. In the experiments developed here, the localization radius is set using

$$L = L_0 \frac{\rho + 0.0064}{\rho + 0.0256} \tag{26}$$

where

$$\rho = \max_n |\psi_+^{(n)} - \psi_-^{(n)}| \tag{27}$$

is the maximal increment to the pressure (subscripts − and + denoting prior and posterior, respectively). The smallest possible localization radius is four times smaller than the largest possible localization radius. Some initial tuning suggested that the localization radius should not be larger than 16 grid points, i.e. $\pi/8$ nondimensional units, so the maximal radius was set to

$$L_0 = \frac{\pi}{8}. \tag{28}$$

This maximal localization radius is half the shortest distance between the observation locations in the grid configuration. To standardize the different experiments, all experiments use the same formula for localization radius based only on observations of ψ. In experiments where observations of ψ are not assimilated, the increments to ψ that would result from observations of ψ are still computed solely for the purpose of setting the localization radius.

Also following 20CR, inflation used the 'relaxation to prior spread' (RTPS) inflation scheme of Whitaker and Hamill (2012). In this scheme the multiplicative inflation coefficient r_{inf} is set to

$$r_{\text{inf}} = (1 - \alpha) + \alpha \frac{\sigma_b}{\sigma_a} \tag{29}$$

where α is the RTPS parameter and σ_b and σ_a are the background (prior) and analysis (posterior) spreads, respectively. The relaxation coefficient α was manually tuned in each configuration to achieve optimal results. Inflation is applied to the analysis ensemble at each cycle.

Since historical observations do not come with a likelihood, we assume that the observation θ is obtained from

$$\theta = \mathbb{P}\left[\phi + \epsilon\right] \tag{30}$$

where $\mathbb{P}$ denotes projection onto a compass with $N_\theta = 32$ points and ϵ is the observation error, which is independent of $\boldsymbol{u}$. This is clearly not in the form (14), but the TSEF framework of Grooms (2022) does not require (14), it only requires the likelihood $[\theta|\phi]/[\phi]$. This is obtained, up to a normalization constant, by convolution of an indicator function $I(\phi)$ with the pdf of ϵ:

$$[\theta|\phi] = \frac{N_\theta}{2\pi}[I * \pi_\epsilon](\theta - \phi) \tag{31}$$

where π_ϵ denotes the pdf of ϵ and $*$ indicates convolution. The indicator function $I(\phi)$ is zero for $|\phi| > \pi/N_\theta$ and one for $|\phi| \leq \pi/N_\theta$. The observation errors ϵ are here drawn from a symmetric triangular distribution peaked at zero and with a radius of $\pi/16$, i.e. the radius equals the precision of the 32-point compass. With this configuration of ϵ

and a 32-point compass the likelihood is a quadratic B spline. The likelihood associated with this observation model could potentially be used to assimilate wind direction observations with local particle filters Penny and Miyoshi (2016); Poterjoy (2016).

The TSEF assimilates the observations serially, i.e. one at a time. For each observation there are two steps. The first step of the TSEF produces an analysis ensemble of wind vectors $\boldsymbol{u}_+^{(n)}$ at the observation location. The second produces an analysis ensemble of vorticity fields $\omega_+^{(n)}$. The update to the vorticity is accomplished using linear regression, following the approach of Anderson (2003), and explained in the Bayesian context by Grooms (2022). Let the vorticity values on the computational grid be unrolled into a vector $\boldsymbol{x}$. A simple linear model is posited for the relationship between the vorticity vector $\boldsymbol{x}$ and the components u and v of the velocity vector $\boldsymbol{u}$ at the observation location

$$\boldsymbol{x} = \boldsymbol{a}_0 + \boldsymbol{a}_u u + \boldsymbol{a}_v v + \boldsymbol{\eta} \tag{32}$$

where $\boldsymbol{a}_{0,u,v}$ are regression coefficients and $\boldsymbol{\eta}$ is the regression residual. The regression coefficients are obtained by plugging the prior ensemble into the simple linear model, and solving for the unknown regression coefficients using ordinary least squares. The estimates of the regression coefficients produced in this way are denoted $\hat{\boldsymbol{a}}_{0,u,v}$, to distinguish them from the true regression coefficients. An ensemble of regression residuals can be defined using the estimated regression coefficients and the prior ensemble

$$\boldsymbol{\eta}^{(n)} = \boldsymbol{x}_-^{(n)} - \hat{\boldsymbol{a}}_0 - \hat{\boldsymbol{a}}_u u_-^{(n)} - \hat{\boldsymbol{a}}_v v_-^{(n)} \tag{33}$$

where the subscript $-$ serves to indicate that these values come from the prior ensemble. The analysis vorticity ensemble is then defined using the analysis velocity ensemble and the regression residuals as follows

$$\boldsymbol{x}_+^{(n)} = \hat{\boldsymbol{a}}_0 + \hat{\boldsymbol{a}}_u u_+^{(n)} + \hat{\boldsymbol{a}}_v v_+^{(n)} + \boldsymbol{\eta}^{(n)}. \tag{34}$$

This update can be written in incremental form as

$$\boldsymbol{x}_+^{(n)} = \boldsymbol{x}_-^{(n)} + \Delta\boldsymbol{x}^{(n)} \tag{35}$$

where the increment is

$$\Delta\boldsymbol{x}^{(n)} = \hat{\boldsymbol{a}}_u \left(u_+^{(n)} - u_-^{(n)}\right) + \hat{\boldsymbol{a}}_v \left(v_+^{(n)} - v_-^{(n)}\right). \tag{36}$$

Localization is accomplished by multiplying the increments by a localization function so that vorticity values far from the observation location are not updated.

Note that the choice $\boldsymbol{z} = \boldsymbol{u}$ implies a regression problem in this second step where $\boldsymbol{u}$ is used as a predictor variable in the regression. The alternative choice $z = \phi$, which works equally well in the first step, would require a second step where ϕ is used as a predictor variable in the regression. A first-order trignonometric polynomial model for the relationship between ϕ and $\boldsymbol{x}$ would take the form

$$\boldsymbol{x} = \boldsymbol{a}_0 + \boldsymbol{a}_c \cos(\phi) + \boldsymbol{a}_s \sin(\phi) + \boldsymbol{\eta}. \tag{37}$$

This is similar but not equivalent to the model (32). The model (32) posits a linear relationship between the vorticity field $\boldsymbol{x}$ and the wind field $\boldsymbol{u}$ at a single point, whereas the model (37) posits a linear relationship between the vorticity field $\boldsymbol{x}$ and a unit vector $(\cos(\phi), \sin(\phi))$ in the same direction as the wind field $\boldsymbol{u}$ at a single point. The former model is more realistic (vorticity is related to wind speed and direction, not just wind direction), so the choice $\boldsymbol{z} = \boldsymbol{u}$ is better than $z = \phi$.

5 | RESULTS

The data assimilation results presented here are all given in terms of the root mean squared error (RMSE) in the vorticity posterior ensemble mean. Denoting the posterior ensemble in vorticity by $\{\omega^{(n)}\}_{n=1}^{N}$, the posterior mean is

$$\bar{\omega} = \frac{1}{N}\sum_{n=1}^{N}\omega^{(n)} \tag{38}$$

and the RMSE is

$$\text{RMSE } = \left[\left\langle(\omega - \bar{\omega})^2\right\rangle\right]^{1/2} \tag{39}$$

where the $\langle\cdot\rangle$ denotes an average in space or time and ω is the reference state from the nature run. Results showing the spatial pattern of RMSE use a time average, and results showing time series of RMSE use a spatial average. The analysis spread is defined to be

$$\text{Spread } = \left[\frac{1}{N}\left\langle\sum_{n=1}^{N}\left(\omega^{(n)} - \bar{\omega}\right)^2\right\rangle\right]^{1/2}. \tag{40}$$

In all experiments the RMSE was only weakly sensitive to the RTPS parameter α, so it was tuned such that the spread matched the RMSE. Presenting results in terms of vorticity is convenient, but also somewhat more stringent of a test than considering alternatives like the pressure ψ or velocity $\boldsymbol{u}$, since the latter are smoother fields than vorticity. Below we present results first for the grid observations, and then for the random observations.

5.1 | Grid Observations

Results for the experiments with gridded observations are shown in Figure 4. The upper left panel shows the standard deviation of climatological variability in the reference simulation vorticity, with white crosses marking the locations at which observations are taken. The time- and space-mean standard deviation for the climatological variability is 2.17, but the variability is distributed non-uniformly through the domain in patterns associated with the zonal jets shown in Figure 3(b).

The results for the ESRF and TSEF filters assimilating only pressure observations are essentially identical, and are not shown. Their RMSE is 1.85 at an RTPS value of $\alpha = 0.1$. When assimilating only pressure observations, the TSEF is effectively a two-step EnKF, though not identical to the ESRF used in the other experiment, so it is not surprising that they yield indistinguishable results.

The lower panels of Figure 4 show the RMSE for the ESRF (left) and TSEF (right) filters assimilating both pressure and wind direction. Optimal results for both filters are obtained at an RTPS value of $\alpha = 0.3$. The ESRF performs no

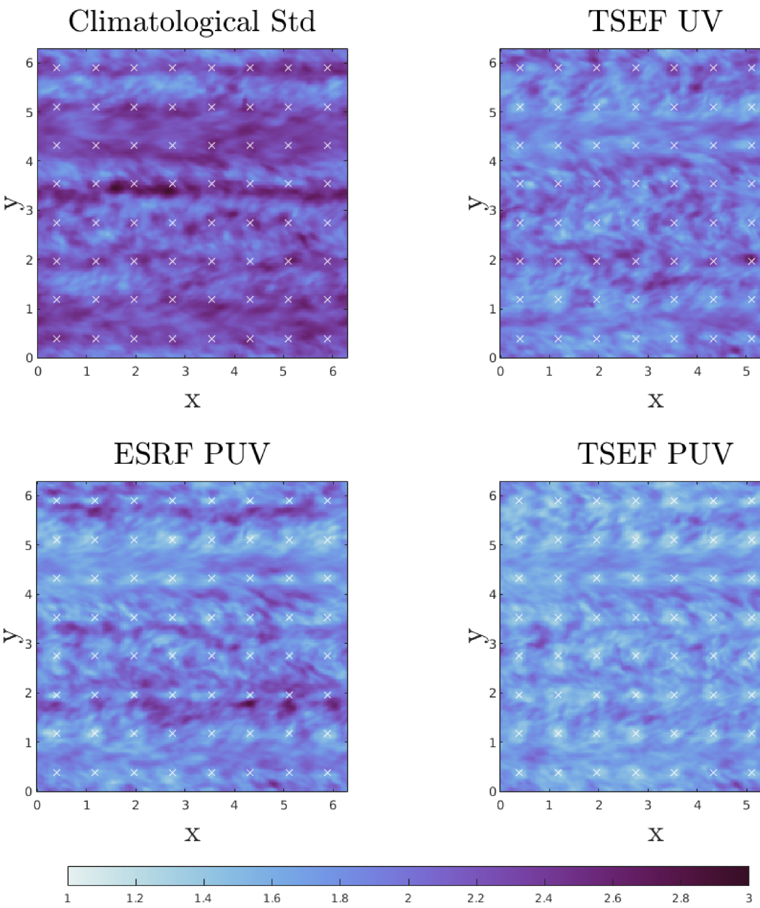

FIGURE 4 Upper left: Local standard deviation of climatological variability in the reference simulation vorticity. Upper Right: RMSE for the TSEF filter assimilating only wind direction. Lower Left: RMSE for the ESRF filter assimilating geostrophic streamfunction and wind direction. Lower Right: RMSE for the TSEF filter assimilating geostrophic streamfunction and wind direction. The white crosses in each figure are the locations of the observations.

better in this example than the ESRF that assimilates only pressure: they both have RMSE of 1.85. In contrast, the TSEF performs slightly better than the ESRF, with an RMSE of 1.7; this slight improvement is visible in the spatial pattern of RMSE shown in the lower right panel of Figure 4.

When assimilating only wind direction, the ESRF filter remains diverged (not shown). Across a range of RTPS parameters α from 0.1 to 0.9 the RMSE remains high, with an optimal RMSE of 2.34 at an RTPS value of 0.9. In contrast, the TSEF filter is able to perform nearly as well with only wind observations as it does with only pressure observations; the optimal RMSE of 1.9 is obtained at an RTPS value of $\alpha = 0.2$. The spatial pattern of RMSE for the TSEF filter with only wind observations, shown in the upper right panel of Figure 4, is slightly better than the climatological pattern shown in the upper left panel.

In all cases the moderate performance of the filters, insofar as they improve only slightly over climatology, is because the observing system has been deliberately chosen to be sparse so as to be reminiscent of the sparsity of observations in the early centuries of a historical reanalysis.

5.2 | Random observations

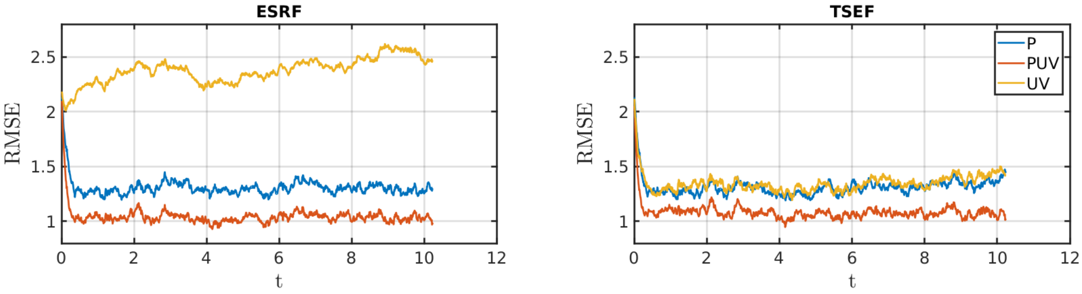

FIGURE 5 Time series of RMSE for the ESRF (left) and TSEF (right) filters with random observation locations. In the legend, P denotes experiments assimilating only observations of pressure, PUV denotes experiments assimilating observations of pressure and wind direction, and UV denotes experiments assimilating only observations of wind direction. The climatological standard deviation of vorticity is 2.17, for reference.

Figure 5 shows time series of the RMSE for both ESRF (left) and TSEF (right) filters with all three sets of observations (pressure only, pressure and wind direction, or only wind direction) with the random observation locations. As with the gridded observations, ESRF and TSEF perform indistinguishably when assimilating only pressure observations. Unlike the gridded observations, the ESRF and TSEF perform indistinguishably when assimilating observations of pressure and wind direction; for both filters, assimilating wind direction improves performance compared to just assimilating pressure. The big difference comes when assimilating only observations of wind direction. In this case, the ESRF diverges with RMSE above climatology, as it did with gridded observations, but the TSEF performs as well with only wind direction as it does with only pressure.

The results presented in Figure 5 use the following values of the RTPS parameter α: ESRF and TSEF assimilating only pressure $\alpha = 0.3$; ESRF and TSEF assimilating pressure and wind direction $\alpha = 0.1$; ESRF assimilating only wind direction $\alpha = 0.8$; TSEF assimilating only wind direction $\alpha = 0.1$.

5.3 | Discussion

The main conclusion is that the TSEF performs as well using only wind direction observations as the ESRF does using only pressure. 20CR relies heavily on surface pressure observations since they are available deeply into the historical record; the results here suggest that similar performance might be possible using only wind direction observations. In contrast, the EnKF approach to wind direction provides no benefit when wind direction is the only observation.

To understand why the EnKF approach diverges when using only wind direction observations, assume that the prior uncertainty in the wind direction is high, as it would be in the early stages of a historical reanalysis. Figure 2 illustrates that the EnKF approach simply removes the component of wind orthogonal to the observed direction; for some ensemble members this is an improvement, while for others it is the opposite. With only this kind of observation, the EnKF approach is unable to reduce the uncertainty in wind direction, and remains diverged.

When other observation types are also available (e.g. pressure), these other observations can indirectly improve the accuracy of the prior wind direction. Then, with a reasonably-accurate prior, the EnKF approach can extract further value from the wind direction observations; this is most evident in the random-location configuration, where the EnKF performs better with wind direction and pressure than with pressure alone. Of course the TSEF approach is also able to perform better with wind direction and pressure than with wind direction alone.

The gridded observing system does not produce large differences in performance between the methods because none of the methods are able to perform very well. The limitation of that configuration is that the observations are sparse in the domain, and while the methods are able to accurately estimate the state close to the observation locations, the spatial correlations are not sufficient to spread the accuracy to the entire domain. The random observing system still has sparsely-spaced observations, but over time the observations cover the spatial domain uniformly, which leads to improved overall accuracy as well as a greater difference in the performance of the different methods.

As a result of this overall improved accuracy using randomly located observations, the prior ensemble in wind directions is more accurate than using gridded observations. This explains why the TSEF is slightly better than the ESRF with gridded observations while the methods perform similarly with randomly located observations: The prior wind ensemble is far more accurate with the randomly located observations, and the ESRF approach to assimilating wind direction works well with an accurate prior ensemble.

6 | CONCLUSIONS

Two methods for assimilating wind direction observations have been developed for the purpose of enabling future historical reanalyses to make use of historical observations of wind direction. The first method uses a linear observation model and can be used with EnKF or variational approaches, while the second method is inherently nonlinear and non-Gaussian and requires an ensemble approach. The first step of the nonlinear TSEF approach uses a nonparametric ensemble approximation of a probability integral transform, and is thus an example of a Quantile Conserving Ensemble Filter (QCEF; Anderson, 2022). The nonlinear TSEF approach is amenable to implementation within the Data Assimilation Research Testbed software suite (DART; Anderson et al., 2009).

The two methods were tested in the context of an idealized two-dimensional fluid model. The main result is that the TSEF approach using only wind direction observations performs as well as an EnKF method using only pressure observations. Although the performance parity seen here depends on the details of the observing system, this is a clear demonstration that the new method can unlock latent value in historical measurements of wind direction. In contrast, the linear observation model provides no benefit at all when assimilating only wind direction observations.

The linear observation model is primarily valuable when used in concert with other observation types, e.g. pressure observations. If enough observational data is available to produce a reasonably-accurate forecast of wind direction, then the linear observation model for wind direction observations can be used to further improve the accuracy of the posterior estimate.

Acknowledgements

This research was funded by US National Science Foundation grant number DMS 2152814. The author is grateful to G. Compo and P. Brohan for introducing the author to the challenges of historical reanalyses, and to an anonymous reviewer for their constructive criticism. Figure data and code are available (Grooms, 2023).

references

Anderson, J., Hoar, T., Raeder, K., Liu, H., Collins, N., Torn, R. and Avellano, A. (2009) The data assimilation research testbed: A community facility. *Bulletin of the American Meteorological Society*, **90**, 1283–1296.

Anderson, J. L. (2001) An ensemble adjustment Kalman filter for data assimilation. *Mon. Weather Rev.*, **129**, 2884–2903.

— (2003) A local least squares framework for ensemble filtering. *Mon. Weather Rev.*, **131**, 634–642.

— (2010) A non-Gaussian ensemble filter update for data assimilation. *Mon. Weather Rev.*, **138**, 4186–4198.

— (2019) A nonlinear rank regression method for ensemble Kalman filter data assimilation. *Mon. Weather Rev.*, **147**, 2847–2860.

— (2022) A quantile-conserving ensemble filter framework. Part I: Updating an observed variable. *Mon. Weather Rev.*, **150**, 1061–1074.

Bishop, C. H. (2016) The GIGG-EnKF: ensemble Kalman filtering for highly skewed non-negative uncertainty distributions. *Q. J. Roy. Meteor. Soc.*, **142**, 1395–1412.

Brohan, P., Allan, R., Freeman, E., Wheeler, D., Wilkinson, C. and Williamson, F. (2012) Constraining the temperature history of the past millennium using early instrumental observations. *Climate of the Past*, **8**, 1551–1563.

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P. et al. (2011) The twentieth century reanalysis project. *Q. J. Roy. Meteor. Soc.*, **137**, 1–28.

Evensen, G. (2009) *Data Assimilation: The Ensemble Kalman Filter*. Springer.

Freeman, E., Woodruff, S. D., Worley, S. J., Lubker, S. J., Kent, E. C., Angel, W. E., Berry, D. I., Brohan, P., Eastman, R., Gates, L. et al. (2017) ICOADS Release 3.0: a major update to the historical marine climate record. *Int. J. Climatology*, **37**, 2211–2232.

Giese, B. S., Seidel, H. F., Compo, G. P. and Sardeshmukh, P. D. (2016) An ensemble of ocean reanalyses for 1815–2013 with sparse observational input. *J. Geophys. Res.-Oceans*, **121**, 6891–6910.

Grooms, I. (2022) A comparison of nonlinear extensions to the ensemble Kalman filter. *Computational Geosciences*, 1–18.

— (2023) iangrooms/Wind_Direction_DA: Two methods for data assimilation of wind direction. URL: `https://doi.org/10.5281/zenodo.7534894`.

Kennedy, C. and Carpenter, M. (2003) Additive Runge-Kutta schemes for convection-diffusion-reaction equations. *Appl. Numer. Math.*, **44**, 139–181.

Laloyaux, P., de Boisseson, E., Balmaseda, M., Bidlot, J.-R., Broennimann, S., Buizza, R., Dalhgren, P., Dee, D., Haimberger, L., Hersbach, H. et al. (2018) CERA-20C: A coupled reanalysis of the twentieth century. *J. Adv. Model. Earth Syst.*, **10**, 1172–1195.

Mardia, K. V. (1975) Statistics of directional data. *Journal of the Royal Statistical Society: Series B (Methodological)*, **37**, 349–371.

Murphy, E., Huang, W., Bessac, J., Wang, J. and Kotamarthi, R. (2022) Joint modeling of wind speed and wind direction through a conditional approach. URL: `https://arxiv.org/abs/2211.13612`.

de Paula Gomez-Delgado, F., Gallego, D., Peña-Ortiz, C., Vega, I., Ribera, P. and Garcia-Herrera, R. (2019) Long term variability of the northerly winds over the eastern mediterranean as seen from historical wind observations. *Global and Planetary Change*, **172**, 355–364.

Penny, S. G. and Miyoshi, T. (2016) A local particle filter for high-dimensional geophysical systems. *Nonlinear Proc. Geoph.*, **23**, 391–405.

Poli, P., Hersbach, H., Dee, D. P., Berrisford, P., Simmons, A. J., Vitart, F., Laloyaux, P., Tan, D. G., Peubey, C., Thépaut, J.-N. et al. (2016) ERA-20C: An atmospheric reanalysis of the twentieth century. *J. Climate*, **29**, 4083–4097.

Poterjoy, J. (2016) A localized particle filter for high-dimensional nonlinear systems. *Mon. Weather Rev.*, **144**, 59–76.

Prieto, M., Gallego, D., García-Herrera, R. and Calvo, N. (2005) Deriving wind force terms from nautical reports through content analysis. the spanish and french cases. *Climatic Change*, **73**, 37–55.

Silverman, B. (1998) *Density estimation for statistics and data analysis*. CRC Press.

Slivinski, L. C., Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Giese, B. S., McColl, C., Allan, R., Yin, X., Vose, R., Titchner, H. et al. (2019) Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system. *Q. J. Roy. Meteor. Soc.*, **145**, 2876–2908.

Soderlind, G. (2002) Automatic control and adaptive time-stepping. *Numer. Algorithms*, **31**, 281–310.

Tenreiro, C. (2022) Kernel density estimation for circular data: a Fourier series-based plug-in approach for bandwidth selection. *Journal of Nonparametric Statistics*, **34**, 377–406.

Whitaker, J. S. and Hamill, T. M. (2002) Ensemble data assimilation without perturbed observations. *Mon. Weather Rev.*, **130**, 1913–1924.

— (2012) Evaluating methods to account for system errors in ensemble data assimilation. *Mon. Weather Rev.*, **140**, 3078–3089.