On methods for assessment of the value of observations in convection-permitting data assimilation and numerical weather forecasting

Hu, G. ORCID: https://orcid.org/0000-0003-4305-3658, Dance, S. L. ORCID: https://orcid.org/0000-0003-1690-3338, Fowler, A. ORCID: https://orcid.org/0000-0003-3650-3948, Simonin, D., Waller, J., Auligne, T., Healy, S., Hotta, D., Löhnert, U., Miyoshi, T., Privé, N. C., Stiller, O., Wang, X. and Weissmann, M. (2025) On methods for assessment of the value of observations in convection-permitting data assimilation and numerical weather forecasting. Quarterly Journal of the Royal Meteorological Society, 151 (768). e4933. ISSN 1477-870X doi: 10.1002/qj.4933 Available at https://centaur.reading.ac.uk/119996/

Received: 21 September 2023 | Revised: 18 September 2024 | Accepted: 20 December 2024

DOI: 10.1002/qj.4933

Quarterly Journal of the Royal Meteorological Society RMetS

REVIEW ARTICLE

On methods for assessment of the value of observations in convection-permitting data assimilation and numerical weather forecasting

Guannan Hu[1,2] | **Sarah L. Dance**[1,2,3] | **Alison Fowler**[1,2] | **David Simonin**[4] | **Joanne Waller**[4] | **Thomas Auligne**[5] | **Sean Healy**[6] | **Daisuke Hotta**[6,7] | **Ulrich Löhnert**[8,9] | **Takemasa Miyoshi**[10,11,12] | **Nikki C. Privé**[13,14] | **Olaf Stiller**[15] | **Xuguang Wang**[16] | **Martin Weissmann**[17]

[1]Department of Meteorology, University of Reading, UK

[2]National Centre for Earth Observation, University of Reading, UK

[3]Department of Mathematics and Statistics, University of Reading, UK

[4]Met Office, Reading, UK

[5]University Corporation for Atmospheric Research, Joint Center for Satellite Data Assimilation, Boulder, Colorado, USA

[6]European Centre for Medium-Range Weather Forecasts, Reading, UK

[7]Meteorological Research Institute, Japan Meteorological Agency, Tsukuba, Japan

[8]Institute for Geophysics and Meteorology, University of Cologne, Cologne, Germany

[9]Hans-Ertel-Centre for Weather Research, Climate Monitoring and Diagnostics, Cologne/Bonn Germany

[10]RIKEN Center for Computational Science, Kobe, Japan

[11]RIKEN Cluster for Pioneering Research, Kobe, Japan

[12]RIKEN Interdisciplinary Theoretical and Mathematical Sciences Program (iTHEMS), Kobe, Japan

[13]Morgan State University/GESTAR II, Baltimore, Maryland, USA

[14]Global Modeling and Assimilation Office, NASA, Greenbelt, Maryland, USA

[15]Deutscher Wetterdienst (DWD), Offenbach, Germany

[16]School of Meteorology, University of Oklahoma, Norman, Oklahoma, USA

[17]Institut für Meteorologie und Geophysik, Universität Wien, Vienna Austria

Correspondence
Guannan Hu, Department of Meteorology, University of Reading, UK.
Email: guannan.hu@reading.ac.uk

Funding information
National Centre for Earth Observation (NCEO); Met Office; German Federal Ministry for Transportation and Digital Infrastructure, Grant/Award Number: BMVI/DWD 4818DWD5B; NOAA, Grant/Award Number: NA19OAR0220154

Abstract

In numerical weather prediction (NWP), a large number of observations are used to create initial conditions for weather forecasting through a process known as data assimilation. An assessment of the value of these observations for NWP can guide us in the design of future observation networks, help us to identify problems with the assimilation system, and allow us to assess changes to the assimilation system. However, assessment can be challenging in convection-permitting NWP. This is because verification of convection-permitting forecasts is not easy, the forecast model is strongly nonlinear, a

limited-area model is used, and the observations used often contain complex error statistics and are often associated with nonlinear observation operators. We compare methods that can be used to assess the value of observations in convection-permitting NWP and discuss operational considerations when using these methods. We focus on their applicability to ensemble forecasting systems, as these systems are becoming increasingly dominant for convection-permitting NWP. We also identify several future research directions, which include comparing results from different methods, comparing forecast validation using analyses versus using observations, applying flow-dependent covariance localization, investigating the effect of ensemble size on the assessment, and generating and validating the nature run in observing-system simulation experiments.

KEYWORDS

convection-permitting, numerical weather prediction, observation impact, observation influence, observing systems

1 | INTRODUCTION

Convection-permitting numerical weather prediction (NWP) is essential for predicting high-impact weather events, such as heavy precipitation, storms, floods, wind gusts, and fog (Baldauf *et al.*, 2011; Brousseau *et al.*, 2016; Clark *et al.*, 2016; Dance *et al.*, 2019; Milan *et al.*, 2020; Scheck *et al.*, 2020; Schraff *et al.*, 2016; Seity *et al.*, 2011). Convection-permitting models typically have a small horizontal grid length (around 1–4 km), allowing convection to be modelled explicitly rather than parameterized (Hu *et al.*, 2022), although research models do run at a smaller grid length (Miyamoto *et al.*, 2013; Waller *et al.*, 2021). Crucial to the realism of NWP at any scale is the routine assimilation of data. The data assimilation (DA) process combines observations with a previous short-range model forecast (called the background) to create the initial conditions (called the analysis) for the NWP model to produce weather forecasts (Bouttier & Courtier, 2002; Nichols, 2010).

Despite the assimilation of observations being essential to forecast skill, it is known that not every observation assimilated reduces the forecast error. This is due to the nature of the random errors present in the observations and the suboptimality in the assimilation system (see Lorenc & Marriott, 2014 for some suggestions for the causes). In addition, the same type of observations may be ranked differently in terms of their impact on global and regional convection-permitting forecasts. For global NWP, the observation types that have the largest overall positive impact on forecast skill are usually satellite microwave and infrared sounders as well as radiosondes, while for regional NWP radar, aircraft, and convectional observations can play a more important role (Boukabara *et al.*, 2020; Randriamampianina *et al.*, 2021). In addition, novel and unused observations (e.g., ground-based and satellite microwave links or uncrewed aerial systems, UAS) bear a high potential for further impact on regional forecasts. It should also be noted that impact estimates and rankings depend strongly on the chosen verification metric and forecast lead time. Therefore, the value of different observations (e.g., different types, or from different sensors or channels), as well as the value of different deployments of observations (location, spatial density, and temporal frequency) for convection-permitting NWP, needs to be properly assessed.

Generally speaking, there are four distinct needs for assessing the value of observations in NWP.

- First, there is a need to provide guidance for the maintenance and improvement of current observation networks.
- Second, there is a need to provide guidance for the design of future observation networks: for example, the design of new observing systems and new networks of currently available, but also potential, observations.
- Third, there is a need to identify problems with the assimilation system: for example, incorrectly specified background and observation-error statistics and inaccurate observation operators.
- Fourth, there is a need to evaluate changes to the assimilation system: for example, tuning background and observation-error variances and optimizing localization functions.

These types of scientific assessments can be used to help guide decisions about the deployment of novel observation systems under limited financial frameworks,

that is, to specify the cost–benefit of an observation system.

The observing-system experiment (OSE) is currently the mainstay of assessing the value of observations at many operational NWP centres (Bouttier & Kelly, 2001; Cress & Wergen, 2001). OSEs can reliably assess the actual improvement or degradation of forecast skill caused by the assimilation of observations. However, they are computationally expensive. In addition, OSEs can only provide insights into how the current observing system should evolve, and cannot measure the impact of future observing systems. The forecast sensitivity to observation impact (FSOI) is a computationally less expensive method than the OSE, but may be difficult to apply to strongly nonlinear systems due to the use of the tangent linear approximation (Langland & Baker, 2004). The observing-system simulation experiment (OSSE) is similar to the OSE, except that it uses synthetic observations to assess the value of future observing systems (Errico & Privé, 2018; Hoffman & Atlas, 2016). The establishment and maintenance of a reliable OSSE framework is, however, not an easy task.

Since convection-permitting DA differs in many aspects from global DA (Bauer *et al.*, 2011; Dance, 2004; Dance *et al.*, 2019; Gustafsson *et al.*, 2018; Hu *et al.*, 2022), assessing the value of observations in convection-permitting NWP presents unique challenges. Recently, there have been a number of methods (including methodologies and metrics) developed specifically for convection-permitting NWP, which overcome some of the drawbacks of the methods used for global NWP. In this work, we review these current state-of-the-art methods, as well as widely used ones. We also provide guidance and recommendations for their future use and development.

Throughout this article, we use the term observation *impact* to represent the value of observations to the *forecast* and observation *influence* to represent the contribution of observations to the *analysis*.

The outline of this review is as follows. In Section 2, we describe the general challenges in the assessment of the value of observations in convection-permitting NWP. In Section 3, we introduce currently available methods and discuss their similarities and differences. In Section 4, we compare the advantages and disadvantages of the different methods in terms of their ability to address the challenges described in Section 2. In Section 5, we consider the practical aspects of using these methods and make recommendations for their future use. In Section 6, we summarize the key remarks and open questions for each method, and provide some general considerations in assessing the value of observations in convection-permitting NWP.

2 | ASSESSMENT CHALLENGES FOR CONVECTION-PERMITTING NWP

Numerous studies have been carried out to assess the value of observations in global NWP (e.g., Cardinali, 2018; Gelaro *et al.*, 2010; Lorenc & Marriott, 2014; Privé, Errico, Todling *et al.*, 2021). However, less attention has been paid to convection-permitting NWP. In this section, we discuss some general challenges of assessing the value of observations in convection-permitting NWP.

2.1 | Verification reference and metric

The validation of convective-permitting forecasts can be challenging due to the selection of appropriate verification references and metrics. Since the true state is never known exactly, we need to use a proxy, such as the analysis or observations. Ideally, the error in the verification reference should be unbiased and statistically independent from the error in the forecast (Daescu, 2009). The analysis is commonly used to verify global forecasts because it provides complete and uniform spatial coverage and, by definition, is the optimal estimate of the atmospheric state. However, in convection-permitting NWP, the analysis may contain non-negligible systematic errors (Necker *et al.*, 2018). More importantly, forecast and analysis errors are unlikely to be independent for short-range forecasts (e.g., less than 6 hours; Privé, Errico, Todling *et al.*, 2021). A positive correlation between the forecast and analysis errors may lead to an overestimation of forecast skill (Hotta *et al.*, 2023).

Compared with the analysis, observations are independent of model forecasts and can be a better choice for verifying convection-permitting forecasts. However, there are still challenges in using observations. First, observations can contain much larger random and systematic errors than the analysis. Second, observations can be sparse and heterogeneous in space and time. Third, we may need appropriate observation operators that calculate the anticipated values of some types of observations (e.g., cloud-affected satellite radiances and simulated satellite images for a given model state: (Kostka *et al.*, 2014; Scheck *et al.*, 2016)). The error in the observation operator affects the verification.

It should be noted that, when using the OSSE, we avoid the problem of selecting the verification reference because a nature run of the NWP model is used as the truth (see Section 3.2.2). In addition, if we assess the impact of observations on forecast uncertainty (e.g., ensemble spread) instead of forecast skill, we do not need to know the truth (see Section 3.3). However, these methods may sometimes

fail to provide the actual impact of observation on forecast skill.

The verification challenge also lies in the choice of verification metrics. Traditional metrics, such as the root-mean-square error (RMSE) and anomaly correlation, may not be good fits for convection-permitting forecasts. These metrics provide point-to-point verification of forecasts against reference values. However, convective-scale features often have sharper gradients than synoptic-scale features. This can lead to the so-called "double penalty" issue if the predicted location of an event is slightly off its exact location (Lledó *et al.*, 2023). Therefore, we need to select metrics that are suitable for small-scale features and provide meaningful information. For example, when verifying precipitation, we need to distinguish displacement, magnitude, and form of features. The choice of metrics may significantly affect the estimation of the value of observations.

2.2 | Stronger nonlinearities

The atmospheric processes modelled in convection-permitting NWP are generally more nonlinear than in global models. For example, Hohenegger and Schar (2007) have shown that the tangent linear approximation breaks down at 1.5 h forecast lead time when using a regional model with a horizontal resolution of 2.2 km and at 54 h when using a global model with a horizontal spectral resolution of 80 km.

The stronger nonlinearity of error growth imposes three challenges in assessing the value of observations. First, it makes methods that rely on tangent linear approximation numerically unstable, unless there is a skilled perturbation model available at convective scale. Second, it may result in greater variability in the impact (influence) estimates for individual forecasts (analyses), which increases the difficulty of achieving a statistically significant estimate. To distinguish whether changes in the estimates are caused by modifications to the observing system or by the chaotic nature of the weather forecast, we need to make sufficient spatial and temporal averaging of the estimates (Geer, 2016). Third, it may lead to a non-Gaussian distribution of the forecast error. Previous studies have shown that approximating non-Gaussian error distributions as Gaussian when quantifying the influence of observations using the entropy reduction and degrees of freedom for signal (DFS) can potentially lead to significantly erroneous estimates (Fowler & Leeuwen, 2013; Fowler & Van Leeuwen, 2012).

Convection-permitting DA often uses remote-sensing observations related to clouds and hydrometeors, such as radar reflectivities and cloud-affected satellite radiances, which are nonlinearly related to model state variables. Therefore, the observation operator (sometimes called the observation forward operator/model) used to calculate the model equivalent of observations from a model state is potentially more nonlinear in convection-permitting NWP (e.g., Geiss *et al.*, 2021; Johnson *et al.*, 2023; Kostka *et al.*, 2014; Scheck *et al.*, 2016; Scheck *et al.*, 2018). Thus, we need methods that can account for the nonlinearity in the observation operator when assessing the value of observations.

2.3 | Limited-area model domain

Convection-permitting NWP usually uses a limited-area model (LAM; Dance, 2004; Gustafsson *et al.*, 2018), which also increases the difficulty of achieving statistically significant estimates of the value of observations. With a global model, we have different weather types happening in different regions at the same time. Therefore, we can easily average the estimates of the value of observations over different weather types. However, with a LAM we usually only have one weather type at one time, which means we need to run the model for a longer period of time to collect enough samples of different weather types.

In convection-permitting NWP, where the focus is often more on forecasting high-impact weather accurately than for global NWP, there is the additional challenge that, in order to assess the value of observations on a particular weather type, we need to average over cases of that weather. However, high-impact weather events usually occur at very low frequencies, making it difficult to gather a sufficient number of events of interest, especially over a LAM.

Lateral boundary conditions (LBCs) given to the LAM may have a large effect on the skill of regional forecasts (e.g., Gustafsson *et al.*, 1998), thus the impact estimates can be different when different LBCs are used. The effect of the LBC on the estimation of the observation impact may need to be considered explicitly. Furthermore, the observations used in global NWP affect the accuracy of the LBC provided to the LAM, and thus these observations also impact the skill of regional forecasts. In some cases, improvements in LBCs may have a larger impact on regional forecasts than the regional assimilation of some observations does (Atlas *et al.*, 2015; Randriamampianina *et al.*, 2021). In this case, the value of globally assimilated observations for regional forecasts may need to be taken into account (Milan *et al.*, 2023).

Another point related to the domain size and LBCs is that, after a certain length of time, the observation information will be advected out of the model domain, so this limits the forecast lead time that observations can have impact for. The short forecast lead time also causes some

problems with using analyses for verification, as for short forecasts the analysis is strongly correlated with the forecast (see Section 2.1).

The last point is that the LAM domain may imply a limited coverage of observations (e.g., Arctic and tropical regions). This may result in inadequate observations for validating the forecast, as described in Section 2.1. In addition, this makes it more difficult to determine coefficients for observation-bias correction reliably (Randriamampianina, 2005). The bias correction, however, affects the impact that observations can have.

2.4 | Complex observation-error statistics

Convection-permitting NWP is subject to the lack of high-resolution observations in the atmospheric boundary layer. Studies have identified the wind profile, temperature profile, humidity profile, precipitation, snow water equivalent, and soil moisture as variables currently not measured adequately (Leuenberger *et al.*, 2020; Teixeira *et al.*, 2021). It should also be noted that the breakdown of the geostrophic balance at convective scales enhances the need for wind observations, because we can no longer infer wind increments from pressure gradients (Bannister, 2021; Vetra-Carvalho *et al.*, 2012).

New observing systems have been suggested to fill the observation gap, including ground-based profiling networks (Chipilski *et al.*, 2022; Geerts *et al.*, 2017), UAS, high-resolution (geostationary) satellites (Scheck *et al.*, 2020; Waller *et al.*, 2016a), constellations of small satellites (Lean *et al.*, 2022), and a network collecting crowdsourced observations (Bell *et al.*, 2022; Hintz *et al.*, 2019), These observations usually have a high spatial and temporal resolution and can therefore provide information at appropriate scales (WMO OSCAR, 2023). However, they often contain complicated observation-error statistics, such as errors due to the observation operator (Janjić *et al.*, 2018), inter-channel error correlations (Bormann *et al.*, 2010; Bormann & Bauer, 2010; Geer, 2019; Stewart *et al.*, 2014), and spatial and temporal error correlations (Cordoba *et al.*, 2017; Michel, 2018; Waller *et al.*, 2016a, 2016b, 2019; Zeng *et al.*, 2021).

It is important that observation-error statistics are specified adequately in DA. Expert knowledge accumulated over decades provides us with a reasonable understanding of how to specify them in global DA. However, more work is required on how to specify observation-error statistics in convection-permitting DA. In global NWP, inter-channel correlations have been used for hyperspectral sounders for many years (Bormann *et al.*, 2010; Bormann & Bauer, 2010; Geer, 2019; Weston *et al.*, 2014), whilst spatial correlations are typically neglected because we can thin the observations in space to exclude correlations and avoid overfitting the data. However, with spatial thinning, we lose the fine-scale information that is important for convection-permitting NWP. To retain information at appropriate scales, we need to reduce the thinning distance and account for spatial correlations explicitly (Fowler *et al.*, 2018; Stewart *et al.*, 2008; Stewart *et al.*, 2013). This has been found to improve analysis quality and forecast skill in convection-permitting NWP (Fujita *et al.*, 2020; Simonin *et al.*, 2019; Yeh *et al.*, 2022).

Synthetic observations are used when assessing the value of observations that are not currently available. This is useful before or during the design, development, and deployment of an observing system. Simulated observations should have realistic observation-error statistics and realistic spatial distribution and temporal frequency (Privé, Errico, McCarty *et al.*, 2021). The complex observation-error statistics, however, complicate simulating the observations meaningfully (Hoffman & Atlas, 2016). For recommendations on the simulation of observations, see Section 5.3.

3 | AVAILABLE METHODS

In this section, we briefly introduce existing methods for assessing the value of observations in convection-permitting NWP, including those widely used in global NWP but also applicable to convection-permitting NWP (albeit with some additional considerations) and those developed recently specifically for convection-permitting NWP. We classify the methods into three main categories:

1. methods for quantifying observation influence on the analysis (Section 3.1);
2. methods for quantifying observation impact on forecast skill (Section 3.2);
3. methods for quantifying observation impact on ensemble forecast spread (Section 3.3).

These methods can provide useful information not only on how the observation network should evolve, but also on how the DA system can be improved to assimilate observations better. The methods in the third category are particularly used to assess the impact of *additional* observations. There are also methods that are proposed specifically for identifying problems with the assimilation of observations. They are extensions of methods described in the first two categories, and we introduce them in Section 3.4.

We note that our categorization is not strict. For example, we introduce the variance reduction (VR)

method in the first category, but it can also be used to assess the impact of observations like the methods in the third category. In addition, the OSSE presented in the second category can be used to assess the influence of observations.

3.1 | Methods for quantifying observation influence on the analysis

We first introduce methods for assessing the influence of observations on the analysis, including widely used information content methods and a recently developed method, the partial analysis increment (PAI; Diefenbach *et al.*, 2023). Since the true state is never known, most influence methods measure how much the analysis is *changed* due to the assimilation of observations, rather than directly measuring the *accuracy* of the analysis. Furthermore, a larger influence of observations on the analysis does not necessarily imply a larger impact of observations on forecasts. This means that we may need to assess the value of observations on the analysis and forecast separately.

3.1.1 | Variance reduction (VR)

Information content methods are used to measure how much information an analysis has retrieved from observations. The same information content can be retrieved from a single observation with very high accuracy or from several observations with lower accuracy (Rodgers, 1998). A simple way to measure the information content is to calculate the reduction of error variance (Brousseau *et al.*, 2014; Desroziers *et al.*, 2005b), which is

$$\delta\sigma^2 = \mathrm{tr}(\mathbf{B}) - \mathrm{tr}(\mathbf{A}), \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{n\times n}$ and $\mathbf{B} \in \mathbb{R}^{n\times n}$ are the analysis- and background-error covariance matrices, respectively, and $\mathrm{tr}(\cdot)$ denotes the trace of a matrix. A larger variance reduction ($\delta\sigma^2$) indicates that the analysis has retrieved more information from the assimilated observations, and thus these observations have a larger influence on the assimilation system. In the linear case, we have

$$\delta\sigma^2 = \mathrm{tr}(\mathbf{KHB}), \tag{2}$$

where $\mathbf{K} \in \mathbb{R}^{n\times m}$ is the Kalman gain matrix and $\mathbf{H} \in \mathbb{R}^{m\times n}$ the linearized observation operator.

In ensemble-based DA systems, the matrices $\mathbf{B}$, $\mathbf{A}$, and $\mathbf{K}$ can be estimated using ensemble perturbations (see Appendix A and Equation 7) and thus Equations (1) and (2) can be used directly. However, in variational DA systems these matrices are typically not formed explicitly, so it is not possible to estimate $\delta\sigma^2$ using these equations. Instead, a randomized approach (see Appendix B) has been used in practice (e.g., Brousseau *et al.*, 2014).

Let $\mathbf{M}$ denotes a tangent linear model (TLM) of the forecast model and $\mathbf{M}^\top$ the adjoint model of the TLM. The error at the analysis time can be projected to the error at a forecast time by $\mathbf{MAM}^\top$, so that the VR method can be extended to measure the impact of observations on the forecast (see Desroziers *et al.*, 2005b).

3.1.2 | Degrees of freedom for signal (DFS)

Another method that belongs to the class of information content methods is the DFS. The name of the DFS comes from the idea that the model state can be spanned by p orthogonal vectors, and hence it can vary statistically independently in p directions ("degrees of freedom": section 2.4 of Rodgers, 2000). If the error uncertainty in one direction is constrained well by the observations, then it can be considered to represent one "degree of freedom for signal". Directions that are not well constrained represent the "degrees of freedom for noise". Therefore, a larger DFS indicates a larger influence of the observations on the analysis. The DFS can be formulated as

$$\mathrm{DFS} = \mathrm{tr}(\mathbf{HK}) \tag{3}$$

for the linear case (Cardinali *et al.*, 2004; Chapnik *et al.*, 2006;Fowler *et al.*, 2020; Lupu *et al.*, 2011). There are other expressions for the DFS, but they can all lead to the same equation (see Appendix C). Compared with the VR, the DFS is normalized by the matrix $\mathbf{B}$ and thus dimensionless, which makes it easy to compare across different DA systems. However, the result of the DFS cannot be separated in model space (see Section 4.8).

Similarly to the VR, the DFS can be estimated using ensemble perturbations or the randomized approach (Appendix B). In addition, the DFS can also be estimated using the assimilation residual approach (Appendix D). The DFS has been widely used for satellite meteorology (e.g., Collard, 2007; Healy & Thépaut, 2006; Rodgers, 1998) and has been used with some success in optimizing the observation error (Chapnik *et al.*, 2006; Desroziers & Ivanov, 2001).

3.1.3 | Entropy reduction and relative entropy

There are also other methods related to information content. The calculation of the VR neglects the non-diagonal

elements of the matrices. An alternative method that includes the off-diagonal elements is the reduction in entropy (also called mutual information), which is expressed as

$$\delta S = \int p(\mathbf{x}) \ln p(\mathbf{x})\, d\mathbf{x} - \int p(\mathbf{y}) \int p(\mathbf{x}|\mathbf{y}) \ln p(\mathbf{x}|\mathbf{y})\, d\mathbf{x}\, d\mathbf{y},$$

where $p(\mathbf{x})$ and $p(\mathbf{y})$ denote the probability distributions of the model state variables and observations, respectively, $p(\mathbf{x}|\mathbf{y})$ denotes the probability distribution of $\mathbf{x}$ given $\mathbf{y}$, and the logarithm may be taken to any convenient base (section 2.2 of Shannon & Weaver, 1964; chapter 2 of Cover & Thomas, 1991). For Gaussian distributions, we have

$$\delta S = \frac{1}{2} \ln |\mathbf{B}| - \frac{1}{2} \ln |\mathbf{A}|, \tag{4}$$

where $|\cdot|$ denotes the determinant of a matrix (Fowler, 2017; Fowler & Leeuwen, 2013; Rodgers, 1998; Rodgers, 2000). In variational DA systems, Equation (4) can be simplified by using a control variable transform (Fisher, 2003). The entropy reduction, VR, and DFS are linked to each other (Xu *et al.*, 2009), and a comparison between them in the presence of spatially correlated observation errors is made in Fowler (2019). As with the DFS, an important use of the entropy reduction is for the channel selection of hyperspectral satellite instruments (Fowler, 2017; Rodgers, 2000).

The relative entropy (also called Kullback–Leibler distance) is

$$\%S = \int p(\mathbf{x}|\mathbf{y}) \ln \frac{p(\mathbf{x}|\mathbf{y})}{p(\mathbf{x})}\, d\mathbf{x},$$

and its integral over all possible realizations of the observations is the entropy reduction (Fowler, 2017; Xu, 2007). Compared with the other methods described previously, the relative entropy is more difficult to compute and interpret in operational systems.

3.1.4 | Partial analysis increments (PAIs)

Diefenbach *et al.* (2023) proposed to use the PAIs as a three-dimensional (3-D) influence measure and a diagnostic for ensemble-based DA systems. The analysis increment on the ensemble mean of the model state can be expressed as

$$\delta\mathbf{x} = \mathbf{K}\bar{\mathbf{d}}, \tag{5}$$

where $\bar{\mathbf{d}}$ is the ensemble mean of the innovation vector,

$$\mathbf{d} = \mathbf{y} - H(\mathbf{x}_\mathrm{b}), \tag{6}$$

with $\mathbf{y} \in \mathbb{R}^m$ being the observation vector, $\mathbf{x}_\mathrm{b} \in \mathbb{R}^n$ the background state vector and $H : \mathbb{R}^n \to \mathbb{R}^m$ the nonlinear observation operator. In ensemble-based DA systems, the Kalman gain can be formed explicitly by

$$\mathbf{K} = \frac{1}{N-1} \mathbf{X}_\mathrm{a} \mathbf{Y}_\mathrm{a}^\top \mathbf{R}^{-1}, \tag{7}$$

where N is the number of ensemble members, $\mathbf{X}_\mathrm{a} \in \mathbb{R}^{n\times N}$ the analysis-ensemble perturbation matrix (Equation A1), $\mathbf{Y}_\mathrm{a} \in \mathbb{R}^{m\times N}$ the analysis-ensemble perturbation matrix in observation space (Equation A2), and $\mathbf{R} \in \mathbb{R}^{m\times m}$ the observation-error covariance matrix (e.g., Liu *et al.*, 2009). It should be noted that Equation (7) is only exact in the absence of localization. However, in practice, the approximation error in the case of localization may be minor (e.g., Diefenbach *et al.*, 2023).

Equation (5) shows that the analysis increment for each model state variable is determined by a weighted sum of the elements of the vector $\bar{\mathbf{d}}$ (with the weights given by the elements of the Kalman gain, $\mathbf{K}$). Then, the PAI of the ith model state variable induced by the jth observation is

$$\left[\delta x^{(j)}\right]_i = K_{ij} d_j, \tag{8}$$

where $\left[\delta x^{(j)}\right]_i$ denotes the ith element of the PAI vector ($\delta\mathbf{x}^{(j)} \in \mathbb{R}^n$), K_{ij} the (i,j)th element of the matrix $\mathbf{K}$, and d_j the jth element of the vector $\bar{\mathbf{d}}$. The sum of PAIs for all observations is the analysis increment vector, that is, $[\delta x]_i = \sum_{j=1}^m \left[\delta x^{(j)}\right]_i$.

As shown by Diefenbach *et al.* (2023), statistics of the PAIs (e.g., mean, standard deviation, and absolute mean) can be used to examine the systematic increments resulting from a certain type of observations. In addition, by detecting whether different observations draw the analysis in opposite directions, the PAIs may be used to identify the suboptimality in the assimilation of observations (see also Section 3.4.2). Lastly, the relative magnitude of the PAI vector can represent the overall influence of an observation.

If we compare the PAI method with the previously introduced information content measures (the VR, DFS, and entropy reduction), a difference is that the PAI method assesses the influence of observations on the analysis state vector, while the others assess the influence on the variance of the analysis error.

3.2 | Methods for quantifying observation impact on forecast skill

We now introduce methods for assessing the impact of observations on forecast skill. Due to the stronger

nonlinearity of the error growth, the forecast lead time at which observations impact is much shorter in convection-permitting NWP than in global NWP.

3.2.1 | Observing-system experiments (OSEs)

The OSE is a common tool that has been used in operational NWP centres for decades. An OSE is often framed as a data denial experiment (DDE), which is carried out by first running the NWP system with a baseline of the observing system (e.g., an operational setup of the observing system) and then the system without assimilating the subset of observations for which we want to derive the impact (Candy *et al.*, 2021; Eyre, 2021a; Gelaro & Zhu, 2009). The impact of the removed observations is assessed by the difference in forecast skill between the two runs. We may also conduct a data addition experiment by adding new observations to the baseline observing system (e.g., Duncan *et al.*, 2021), or a data replacement experiment by replacing some observations in operational use with new observations.

In OSEs, forecast skill is measured by comparing the forecasts with a proxy for the truth, which is typically a verifying analysis or observations (e.g., Lawrence *et al.*, 2019). As discussed in Section 2.1, the analysis provides complete and uniform data coverage, but it may not be an independent verifying reference, which is particularly an issue in the case of systematic model error. In contrast, the error in the observation is independent of the error in the forecast, but observations suffer from uneven distribution and insufficient coverage (e.g., paucity of upper-air observations: Hagelin *et al.*, 2021). For a discussion of how to choose the verification reference in practice, see Section 5.4.

The OSE provides reliable impact estimates and can serve as a standard for validating the results of other methods that may use synthetic observations (e.g., OSSEs), a tangent linear approximation (e.g., the FSOI), or a proxy for forecast skill (e.g., the ensemble sensitivity analysis (ESA) method). Many studies have used OSEs to study the impact of observations in the context of convection-permitting NWP (Caumont *et al.*, 2016; Chipilski *et al.*, 2020, 2022; Degelia *et al.*, 2020; Johnson *et al.*, 2022).

3.2.2 | Observing-system simulation experiments (OSSEs)

An OSSE is similar to an OSE, except that the OSSE uses synthetic observations and a hypothetical true state that is given by a "nature run" (NR) from an NWP model. In order to reflect the model error realistically, the NR should be generated by running a more sophisticated model (e.g., with a higher resolution and/or an improved physical parameterization scheme) than the model used for the assimilation and forecasting (Hoffman & Atlas, 2016). Synthetic observations are generated by interpolation of the NR fields and sometimes the accompanying use of an observation operator, such as a radiative transfer model. Simulated observation errors should be added to the synthetic observations.

The OSSE offers greater flexibility than the OSE. It can be used to assess the impact of a future observation type and an alternate deployment of an existing observation type (e.g., Privé *et al.*, 2022). In addition, since the forecast is validated against the NR, there is no issue with selecting a verification reference. Using the NR, we can also calculate the analysis and background errors and assess the influence of observations (Cucurull *et al.*, 2018; Privé *et al.*, 2022; Privé & Errico, 2013).

Care must be taken to ensure that the OSSE framework replicates the real world sufficiently, including the behaviour of the NR and the full simulation of all observation types used in operational forecasting. However, in real applications for LAMs, OSSEs are sometimes simplified by assuming a perfect model (i.e., using the same model for the NR and the NWP experiment) and simulating only a part of the observing system (e.g., omitting radiance observations and adding surface observations), arguing that targeted variables are not sensitive to the neglected observations, which may lead to overoptimistic impact results (Bachmann *et al.*, 2019, 2020; Huang *et al.*, 2022; Kugler *et al.*, 2023; Maejima *et al.*, 2022; Schröttle *et al.*, 2020).

The OSSE systems should be calibrated (e.g., using the corresponding OSE) to ensure that their results are comparable with those that would be obtained using real observing systems. They should also keep pace with the development of operational NWP systems and real observing systems. However, it should be noted that the results of the OSSE are dependent on current DA capabilities and do not provide a projection of the capabilities of future versions of DA, including scientific, technical, and computing improvements that may be in place when a new observing system becomes operational.

3.2.3 | Sensitivity observation-system experiments (SOSEs)

The sensitivity observation-system experiment (SOSE) developed by Marseille *et al.* (2008) is another method for assessing the impact of a *future* observing system. In

contrast to an OSSE, the baseline of the observing system in a SOSE is given by real observations, and only future observations are simulated in a SOSE. A problem of using real and simulated observations together is that they are inconsistent with respect to the underlying truth. The SOSE tries to solve this problem by generating the simulated observations using an adapted analysis.

The adapted analysis is created by the following steps: (1) run a forecast from the background, (2) compute the forecast error with respect to a verifying analysis, (3) project the forecast error back to the initial time through a forecast sensitivity computation (Rabier *et al.*, 1996) and use the resulting key background errors to correct the background, and (4) create a new analysis using the corrected background and all existing observations. Readers are referred to fig. 1 of Marseille *et al.* (2008) for a schematic diagram of the SOSE analysis scheme. Marseille *et al.* (2008) have shown in an OSSE that the adapted analysis is generally closer to the truth than the original analysis. Since the SOSE is not as widely used as the OSSE, it is not yet clear how well the SOSE performs in convection-permitting NWP.

3.2.4 | Ensemble forecast sensitivity to observation impact (EFSOI)

The ensemble forecast sensitivity to observation impact (EFSOI: (Kalnay *et al.*, 2012; Li *et al.*, 2010; Liu & Kalnay, 2008)) was developed based on the FSOI. They are computationally inexpensive alternatives to the DDE. They measure how much a subset of observations contributes to the reduction of short-range forecast error within a system assimilating all observations (e.g., Eyre, 2021b). Observations that contribute to a greater reduction could be considered to have a greater impact. The reduction of the forecast error due to the assimilation of observations is measured by

$$\delta e = \left(\mathbf{x}_{\mathrm{a}}^{\mathrm{f}} - \mathbf{x}_{\mathrm{t}}\right)^{\top}\mathbf{C}\left(\mathbf{x}_{\mathrm{a}}^{\mathrm{f}} - \mathbf{x}_{\mathrm{t}}\right) - \left(\mathbf{x}_{\mathrm{b}}^{\mathrm{f}} - \mathbf{x}_{\mathrm{t}}\right)^{\top}\mathbf{C}\left(\mathbf{x}_{\mathrm{b}}^{\mathrm{f}} - \mathbf{x}_{\mathrm{t}}\right), \quad (9)$$

where $\mathbf{x}_{\mathrm{a}}^{\mathrm{f}}$ is the forecast generated from the analysis, $\mathbf{x}_{\mathrm{b}}^{\mathrm{f}}$ the forecast generated from the background (equivalent to the forecast generated from the previous analysis), $\mathbf{x}_{\mathrm{t}}$ the verifying analysis, and $\mathbf{C}$ a symmetric matrix of energy weighting coefficients often representing total energy (Fourrié *et al.*, 2002). The choice of the matrix $\mathbf{C}$ (Ehrendorfer *et al.*, 1999; Ehrendorfer & Errico, 1995; Rabier *et al.*, 1996; Rosmond, 1997) affects the value of δe and may thus have a large effect on the impact estimates (Janiskova & Cardinali, 2016). For a discussion of the selection of the matrix $\mathbf{C}$ in convection-permitting NWP, see Section 5.5.

The EFSOI is given by a linear approximation of δe (see Appendix E.1). After applying localization, we obtain

$$\delta e \approx \mathrm{EFSOI} = \frac{1}{N-1}\bar{\mathbf{d}}^{\top}\mathbf{R}^{-1}\mathbf{L}_{\mathrm{mn}} \circ \left(\mathbf{Y}_{\mathrm{a}}\mathbf{X}_{\mathrm{a}}^{\mathrm{f}^{\top}}\right)\mathbf{C}\left[\left(\bar{\mathbf{x}}_{\mathrm{a}}^{\mathrm{f}} - \mathbf{x}_{\mathrm{t}}\right) + \left(\bar{\mathbf{x}}_{\mathrm{b}}^{\mathrm{f}} - \mathbf{x}_{\mathrm{t}}\right)\right], \quad (10)$$

where $\circ$ denotes the element-wise product, $\mathbf{X}_{\mathrm{a}}^{\mathrm{f}}$ the analysis forecast ensemble perturbation matrix (Equation A1), and $\mathbf{L}_{\mathrm{mn}} \in \mathbb{R}^{m\times n}$ the observation to model localization matrix, which decides which model grid points can be impacted by each observation (Hotta *et al.*, 2017a, 2017b; Kalnay *et al.*, 2012). For more information on the localization, see Section 5.2.1.

We may consider Equation (10) as a product of $\bar{\mathbf{d}}^{\top}$ and a sensitivity vector given by

$$\left[\frac{\partial \delta e}{\partial \mathbf{y}}\right] = \frac{1}{N-1}\mathbf{R}^{-1}\mathbf{L}_{\mathrm{mn}} \circ \left(\mathbf{Y}_{\mathrm{a}}\mathbf{X}_{\mathrm{a}}^{\mathrm{f}^{\top}}\right)\mathbf{C}\left[\left(\bar{\mathbf{x}}_{\mathrm{a}}^{\mathrm{f}} - \mathbf{x}_{\mathrm{t}}\right) + \left(\bar{\mathbf{x}}_{\mathrm{b}}^{\mathrm{f}} - \mathbf{x}_{\mathrm{t}}\right)\right],$$

which describes how small changes in observations will change the forecast-error reduction. The major difference between the FSOI and EFSOI lies in the sensitivity calculation: the FSOI (Equation E3) uses the adjoint model (the transpose) of the TLM of the forecast model (adjoint sensitivity); the EFSOI uses ensemble perturbation matrices at initial and forecast times (ensemble sensitivity). We focus on the EFSOI because of a lack of TLMs useful at the convective scale and because the EFSOI is more computationally robust to the strong nonlinearity in the forecast-error growth (Lorenc & Marriott, 2014).

Since we are often interested in the impact of a subset of observations, we obtain the contribution from a single observation j to Equation (10) by

$$\mathrm{EFSOI}^{(j)} = d_j\left[\frac{\partial \delta e}{\partial \mathbf{y}}\right]_j, \quad (11)$$

where j denotes the index of the vectors $\bar{\mathbf{d}}$ and $\left[\partial \delta e/\partial \mathbf{y}\right]$ that corresponds to the jth observation.

In Equation (10), the analysis is used to verify the forecast. If observations are used, we obtain the observation-based EFSOI (Appendix E.2), the sensitivity vector of which has a different expression from the analysis-based EFSOI. We discuss whether the analysis or observations should be used to verify the forecast in convection-permitting NWP in Section 5.4.

The EFSOI has been used in convection-permitting NWP to assess the impact of observations (Gasperoni *et al.*, 2024; Necker *et al.*, 2018; Sommer & Weissmann, 2016). In addition, the EFSOI has been used

to identify the problems with the assimilation system (Ota *et al.*, 2013) and provide a fully flow-dependent quality control (Chen & Kalnay, 2019, 2020; Hotta *et al.*, 2017a).

3.2.5 | Comparison between EFSOI and DDE

Several differences between EFSOI and DDE should be noted when comparing their results (e.g., Eyre, 2021a). First, the EFSOI is based on a linear projection between the initial and forecast errors as it uses a sensitivity vector, although this sensitivity vector is calculated using nonlinear forecasts. In comparison, a DDE allows for fully nonlinear forecast-error growth. Second, the EFSOI is calculated using the error covariance matrices ($\mathbf{HBH}^{\mathrm{T}}$ and $\mathbf{R}$) of the full observing system only, whereas a DDE consists of two forecast runs, one using the error covariance matrices of the full observing system and the other using adjusted error covariance matrices as some observations are removed. Third, the EFSOI measures non-cycled observation impacts, whilst a DDE can be used to assess the accumulated impact of observations in a cycled DA environment (see Section 4.9).

Because of these differences, the EFSOI of two observation subsets should be equal to the sum of the EFSOI of each observation subset, which is, however, not the case for a DDE. In addition, the localization issue in the EFSOI and its treatment (see Section 5.2.1) can make a significant difference between the results of the EFSOI and DDE.

3.3 | Methods for quantifying observation impact on forecast spread

The spread of ensemble forecasts (e.g., calculated as the standard deviation between ensemble members) is an estimate of the uncertainty in the forecast and should be negatively correlated with forecast skill. Due to this spread–skill relationship, assessing the impact of observations on the ensemble spread is qualitatively equivalent to assessing the impact of observations on forecast skill. Specifically, a larger reduction in the ensemble spread indicates a larger improvement in forecast skill.

In this section, we introduce two methods that assess the impact of observations by measuring the change in ensemble spread. The successful use of these methods is dependent on a clear spread–skill relationship. In practice, the ensemble is often underdispersed, which affects the assessment of the absolute impact but has less effect on the assessment of the relative impact (see Section 5.1). These methods are advantageous for studying the impact of future observations because they only focus on the impact on the ensemble spread, not the ensemble mean. This means that, unlike an OSSE, they only require the simulation of the observations to be assessed or their error statistics, and the simulated observations do not need to be consistent with an underlying truth, so a separate nature run is not needed. However, it should be noted that, when adding new observations, we can reduce forecast spread but not actually improve forecast skill (Healy *et al.*, 2024).

3.3.1 | Ensemble of data assimilations (EDA) spread reduction

The ensemble of data assimilations (EDA: (Bonavita *et al.*, 2016; Isaksen *et al.*, 2010)) represents an ensemble of independent variational data assimilations, which is designed primarily to estimate flow-dependent background- and analysis-error statistics for hybrid four-dimensional variational (4D-Var) systems. The EDA may also be used to provide initial conditions for ensemble forecasting. When assessing the value of observations, the EDA allows for a statistical estimate of uncertainty reduction in analyses and short-range forecasts due to the assimilation of additional real or synthetic observations (Harnisch *et al.*, 2013; Healy *et al.*, 2024; Tan *et al.*, 2007). If the new observations added to the DA system result in a larger reduction in the EDA spread, they are considered to have a greater value.

In principle, spread reduction can be used as a measure of the value of observations in any ensemble-based DA systems. However, in practice, the necessity to apply strong covariance inflation to counteract the underdispersion issue may make it difficult to use. For example, in most ensemble Kalman filters for convection-permitting models, the posterior ensemble spread (after the assimilation) is nudged toward the prior spread (before the assimilation; Whitaker & Hamill, 2012; Zhang *et al.*, 2004), making the difference between them obscured. In comparison, the EDA does not require substantial covariance inflation, so that spread reduction can be a more reliable indicator of the value of observations (see Section 5.1 for a discussion of the underdispersion of EDA spread).

3.3.2 | Ensemble sensitivity analysis (ESA)

The ESA (Ancell & Hakim, 2007; Hakim & Torn, 2008; Torn & Hakim, 2008) provides a computationally cheap way of estimating the reduction in the ensemble variance of a forecast metric due to the assimilation of additional observations (Griewank *et al.*, 2023; Nomokonova *et al.*, 2022). Let $J = f(\mathbf{x})$ be a scalar function of model state variables, quantifying aspects of the forecasting system of interest: for example, cumulative

precipitation for a region or time period or averaged wind components over a region (Nomokonova *et al.*, 2022). Let $\mathbf{J} \in \mathbb{R}^N$ be a row vector containing a set of J computed for each ensemble member. The reduction in the ensemble variance of J due to the assimilation of additional observations is

$$\delta\sigma_J^2 = \left[\frac{\partial J}{\partial \mathbf{x}}\right]^{\top} (\mathbf{P}_a - \mathbf{P}_b)\left[\frac{\partial J}{\partial \mathbf{x}}\right], \tag{12}$$

where $\left[\partial J/\partial \mathbf{x}\right] \in \mathbb{R}^n$ is a sensitivity vector that describes how small changes in model state variables at the initial time will change J at a forecast lead time, and $\mathbf{P}_a$ and $\mathbf{P}_b$ are analysis and background ensemble covariance matrices (Equation A3), respectively. Since we are assessing the impact of new observations when added to the current observing system, the matrix $\mathbf{P}_b$ in this equation is the analysis ensemble covariance matrix for the current observing system, whereby the matrix $\mathbf{P}_a$ is calculated from the matrix $\mathbf{P}_b$ based on using the additional observations. For a derivation of Equation (12) and how $\delta\sigma_J^2$ is estimated in practice, see Appendix F.

Generally speaking, the value of $\delta\sigma_J^2$ can be calculated without knowing the sensitivity vector explicitly (Equation F7). However, this implicit calculation does not allow the covariance localization (Section 5.2) used in the assimilation to be easily included in the computation of $\delta\sigma_J^2$ (for details see Appendix F.1). If the sensitivity vector is calculated explicitly, then $\delta\sigma_J^2$ can be calculated using Equation (12), and the covariance localization can be included in the same way as in the assimilation system (see Appendix F.2).

Assuming the sensitivity vector and the background-error covariance matrix are available, the calculation of Equation (12) only requires the matrix $\mathbf{P}_a$, which can be calculated if we know the error statistics of the observations to be assessed and have the corresponding observation operator (e.g., Equation F11). Thus, without requiring any actual observations, the ESA method, like the classical OSSE, is ideally suited to assess the impact of not-yet-existing future observations (e.g., Nomokonova *et al.*, 2022). The advantage over the OSSE is that the impacts of arbitrary, potential observations can be derived from a once-only ensemble calculation.

3.3.3 | Comparison between the ESA method and EFSOI

A similarity between the ESA and EFSOI is that they both use ensemble sensitivity. However, the major difference between them is that the ESA method assesses the impact of changes in the ensemble perturbation matrix, $\mathbf{X}_a - \mathbf{X}_b$, on forecast spread, whereas the EFSOI can be seen as assessing the impact of the analysis increment, $\mathbf{x}_a - \mathbf{x}_b$, on forecast skill. Moreover, the EFSOI uses a fixed scalar forecast *error* metric, whereas the scalar forecast metric, J, used in the ESA method is an arbitrary function of the forecast state, and a different J can be used at different locations (Tardif *et al.*, 2022). This makes the ESA method attractive for specific applications (e.g., quantitative precipitation forecasts or forecasting for renewable energy applications). Another difference is that the ESA method does not require knowledge of the observation value, whereas the EFSOI does.

3.3.4 | Comparison between the ESA, EDA, and VR methods

The ESA, EDA, and VR methods all measure uncertainty reduction. However, there are some differences between them. First, the EDA and VR methods calculate the reduction in the ensemble spread/variance of model state variables, while the ESA method calculates the reduction in the ensemble variance of the metric J, which is a scalar function of model state variables. Second, the EDA method allows for a fully nonlinear growth in the ensemble spread of model state variables, whereas the ESA and VR methods use a linear projection. In the special case where $J = x_i$ is given by a single model state variable, the ESA method measures the uncertainty reduction of the variable x_i at a forecast lead time, as does the VR method. However, the difference is that the ESA method uses the ensemble sensitivity, whereas the VR method uses the TLM and adjoint model. Third, the EDA method requires new observations and their error statistics, while the ESA method requires only the error statistics. However, it should be noted that the EDA spread reduction is primarily sensitive to the error statistics and less sensitive to the actual value of the new observations.

Since the EDA system produces an estimate of analysis and background-error covariance matrices (**A** and **B**), it enables an estimate of the VR using Equation (1) or an estimate of the DFS using Equation (3). However, this is not how the EDA method is used; the EDA method requires the EDA system to be run twice (once using the current observing system and once adding new observations) and measures the difference in the two **A** matrices (or the two **B** matrices).

3.4 | Identification of problems in assimilation systems

How well observations are assimilated affects their value for the NWP system. For example, if background- and

observation-error statistics are specified incorrectly in the assimilation (i.e., they do not represent the true error statistics accurately), or if the observation operator is inaccurate, then the observations may not be optimally assimilated and thus their values may not be represented correctly. In this section, we discuss how to identify issues related to the assimilation of observations using methods based on the DFS and EFSOI. We note that methods described in Sections 3.1 to 3.3 may also be used to identify problems in the assimilation system.

3.4.1 | The actual value of the DFS in suboptimal DA systems

Fowler *et al.* (2020) have shown that the value of DFS calculated under the assumption of optimality of a DA system (Equation 3) may differ from the actual value of the DFS in a suboptimal system, and this difference can be used to identify the suboptimality in the assimilation of observations. Let

$$\mathbf{D} = \mathbf{HBH}^{\top} + \mathbf{R} \tag{13}$$

denote the innovation covariance matrix used in the assimilation and

$$\mathbf{D}_{\mathrm{t}} = \mathbb{E}\left[\mathbf{d}\mathbf{d}^{\top}\right] \tag{14}$$

the true innovation covariance matrix. If $\mathbf{D} \neq \mathbf{D}_{\mathrm{t}}$, then Equation (3) does not hold, and the actual DFS should be calculated as

$$\mathrm{DFS}_{\mathrm{Actual}} = \mathrm{tr}\left(\mathbf{HKD}_{\mathrm{t}}\mathbf{D}^{-1}\right). \tag{15}$$

In practice, the mismatch between $\mathbf{D}$ and $\mathbf{D}_{\mathrm{t}}$ can be caused by (1) misspecification of the background- and observation-error statistics, (2) linearization of a nonlinear observation operator, or (3) mutual correlations between observation errors and background errors (Fowler *et al.*, 2020; Lupu *et al.*, 2011).

A comparison between the theoretical and actual DFS allows us to identify the suboptimality in the assimilation of observations. However, it should be noted that the source of the discrepancy cannot be diagnosed uniquely from the difference between the actual and theoretical DFS alone (Fowler *et al.*, 2020). In a practical application, Fowler *et al.* (2020) calculated the two values of the DFS using assimilation residuals (see Appendix D) and found in the Met Office's convection-permitting NWP system that poor assumptions in the observation operator led to an actual DFS up to three times larger than its theoretical value when assimilating Doppler radial winds.

3.4.2 | Cross-validation (C-V) of observations

Observations may have a suboptimal influence/impact if they are given the wrong weight during the assimilation or, more severely, if they pull the model state in the wrong direction due to systematic errors or incorrect processing. To distinguish between these two cases, Stiller (2022) partitions the observation-based EFSOI (Equation E7) into two terms,

$$\mathrm{EFSOI}_{\mathrm{y}} = -(2J_{\mathrm{b}} - J_{\mathrm{ab}}), \tag{16}$$

where

$$J_{\mathrm{b}} = \frac{1}{N-1}\left(\bar{\mathbf{d}}\right)^{\top}\mathbf{R}^{-1}\mathbf{Y}_{\mathrm{a}}\left(\mathbf{Y}^{\mathrm{f}}_{\mathrm{av}}\right)^{\top}\mathbf{R}_{\mathrm{v}}^{-1}\left(\mathbf{y}_{\mathrm{v}} - \mathbf{y}^{\mathrm{f}}_{\mathrm{bv}}\right), \tag{17}$$

$$J_{\mathrm{ab}} = \frac{1}{N-1}\left(\bar{\mathbf{d}}\right)^{\top}\mathbf{R}^{-1}\mathbf{Y}_{\mathrm{a}}\left(\mathbf{Y}^{\mathrm{f}}_{\mathrm{av}}\right)^{\top}\mathbf{R}_{\mathrm{v}}^{-1}\left(\mathbf{y}^{\mathrm{f}}_{\mathrm{av}} - \mathbf{y}^{\mathrm{f}}_{\mathrm{bv}}\right), \tag{18}$$

with $\mathbf{y}_{\mathrm{v}}$, $\mathbf{R}_{\mathrm{v}}$, and H_{v} being the observation vector, observation-error covariance matrix, and observation operator for verifying observations, while $\mathbf{y}^{\mathrm{f}}_{\mathrm{av}} = \overline{H_{\mathrm{v}}(\mathbf{x}^{\mathrm{f}}_{\mathrm{a}})}$, $\mathbf{y}^{\mathrm{f}}_{\mathrm{bv}} = \overline{H_{\mathrm{v}}(\mathbf{x}^{\mathrm{f}}_{\mathrm{b}})}$, and $\mathbf{Y}^{\mathrm{f}}_{\mathrm{av}}$ is the analysis-ensemble perturbation matrix in verification space. Alternatively, following the same approximations that led to Equation (16) (see Appendix E.2), one can write

$$J_{\mathrm{b}} \approx \left(\mathbf{y}^{\mathrm{f}}_{\mathrm{av}} - \mathbf{y}^{\mathrm{f}}_{\mathrm{bv}}\right)^{\top}\mathbf{R}_{\mathrm{v}}^{-1}\left(\mathbf{y}_{\mathrm{v}} - \mathbf{y}^{\mathrm{f}}_{\mathrm{bv}}\right)$$

and

$$J_{\mathrm{ab}} \approx \left(\mathbf{y}^{\mathrm{f}}_{\mathrm{av}} - \mathbf{y}^{\mathrm{f}}_{\mathrm{bv}}\right)^{\top}\mathbf{R}_{\mathrm{v}}^{-1}\left(\mathbf{y}^{\mathrm{f}}_{\mathrm{av}} - \mathbf{y}^{\mathrm{f}}_{\mathrm{bv}}\right),$$

which shows that J_{b} is positive if the assimilation of the assessed observations pulls the model state towards the verifying observations (i.e., the analysis increment, $\mathbf{y}^{\mathrm{f}}_{\mathrm{av}} - \mathbf{y}^{\mathrm{f}}_{\mathrm{bv}}$, and background departure, $\mathbf{y}_{\mathrm{v}} - \mathbf{y}^{\mathrm{f}}_{\mathrm{bv}}$, have the same sign). In addition, since J_{ab} is non-negative, we can see directly from Equation (16) that $J_{\mathrm{b}} > 0$ is necessary (but insufficient) for observations to have a beneficial impact (i.e., $\mathrm{EFSOI}_{\mathrm{y}} < 0$).

In addition to checking the sign of $\mathbb{E}[J_{\mathrm{b}}]$, a more sensitive test is to compare $\mathbb{E}[J_{\mathrm{b}}]$ with a reference value (Stiller, 2022), which is given by

$$\widetilde{J}_{\mathrm{b}} = \frac{1}{(N-1)^2}\mathrm{tr}\left(\mathbf{R}^{-1}\mathbf{Y}_{\mathrm{a}}\left(\mathbf{Y}^{\mathrm{f}}_{\mathrm{av}}\right)^{\top}\mathbf{R}_{\mathrm{v}}^{-1}\mathbf{Y}^{\mathrm{f}}_{\mathrm{bv}}(\mathbf{Y}_{\mathrm{b}})^{\top}\right). \tag{19}$$

For a derivation of $\widetilde{J}_{\mathrm{b}}$ see Appendix G.1. In practice, since we are interested in assessing a subset of observations, we calculate the contributions $J_{\mathrm{b}}^{(j)}$ and $J_{\mathrm{ab}}^{(j)}$ from the assessed observations to J_{b} and J_{ab}. When using the equality $\mathbb{E}\left[J_{\mathrm{b}}^{(j)}\right] = \widetilde{J}_{\mathrm{b}}^{(j)}$ to assess a subset of observations, results of cross-validating two trusted observation types can be used as a benchmark. This gives us an idea of

what kind of agreement between $\mathbb{E}\left[J_b^{(j)}\right]$ and $\widetilde{J}_b^{(j)}$ should be expected in our DA system. For example, Stiller (2022) used radiosonde and aircraft data as trusted observations and found that for atmospheric motion vectors, although $\mathbb{E}\left[J_b^{(j)}\right] > 0$, the agreement between $\mathbb{E}\left[J_b^{(j)}\right]$ and $\widetilde{J}_b^{(j)}$ is significantly worse than the benchmark.

In Equation (19), $(N-1)^{-1}\mathbf{Y}_{bv}^{f}(\mathbf{Y}_b)^{\top}$ is the background ensemble cross-covariance between observation and verification space. Therefore, when using two trusted observation types, the equality $\mathbb{E}\left[J_b^{(j)}\right] = \widetilde{J}_b^{(j)}$ can also be used to assess the quality of the ensemble-based error covariances in terms of representing the true error statistics (see Equation G8), or different choices of parameters used for background covariance localization (Vural *et al.*, 2024).

In practice, samples of $J_b^{(j)}$ can be collected at low numerical cost in a great variety of different types of bins (e.g., latitudes, distance between observations or particular aircraft or radiosonde types), which facilitates identification of the cause of the sub-optimal impact. In addition, when testing observations that are not yet used routinely in the assimilation, we may, at a similarly low computational cost, compute the value of $J_b^{(j)}$ from single-observation experiments. This can provide reliable results regarding the suboptimality in the assimilation of observations (Stiller, 2022).

For the other diagnostic, $J_{ab}^{(j)}$, it is shown in Appendix G.2 that, in an optimal DA system, $\mathbb{E}\left[J_{ab}^{(j)}\right]$ should have the same magnitude as $\mathbb{E}[J_b]$. This yields a test for the optimality of the assumptions made about the DA system, and particularly the extent to which the weights given to the observations are optimal. However, it should be noted that one needs the cross-validation term J_b to assess how far this influence is beneficial and whether it is close to optimal.

3.4.3 | Ensemble forecast sensitivity to observation-error covariance (EFSR)

The ensemble forecast sensitivity to the matrix $\mathbf{R}$ (EFSR) was proposed by Hotta *et al.* (2017b) based on the EFSOI and forecast sensitivity to observation-error covariance (Daescu, 2008; Daescu & Langland, 2013, 2017). How the forecast error changes due to small changes in the observation-error covariance matrix is expressed as

$$\frac{\partial e_a}{\partial R_{ij}} = -\left[\mathbf{R}^{-1}\mathbf{r}\right]_j \left[\frac{\partial e_a}{\partial \mathbf{y}}\right]_i, \tag{20}$$

where i and j are matrix or vector indices,

$$e_a = \left(\mathbf{x}_a^f - \mathbf{x}_t\right)^{\top}\mathbf{C}\left(\mathbf{x}_a^f - \mathbf{x}_t\right)$$

measures the forecast error (the first part of Equation 9), and

$$\mathbf{r} = \mathbf{y} - H(\mathbf{x}_a) \tag{21}$$

is the observation-minus-analysis residual vector. Using the chain rule for partial derivatives, the sensitivity vector is calculated by

$$\left[\frac{\partial e_a}{\partial \mathbf{y}}\right] = \frac{2}{N-1}\cdot\mathbf{R}^{-1}\mathbf{Y}_a\left(\mathbf{X}_a^f\right)^{\top}\mathbf{C}\left(\overline{\mathbf{x}}_a^f - \mathbf{x}_t\right), \tag{22}$$

as in Hotta *et al.* (2017b). The EFSR can be used to diagnose whether the prescribed observation-error variance is optimal. A positive value of $\partial e_a/\partial R_{ij}$ means that e_a increases with R_{ij}. In this case, we should decrease the observation-error variance to reduce the forecast error. In contrast, if $\partial e_a/\partial R_{ij}$ is negative, we should increase the observation-error variance to reduce the forecast error.

4 | ADVANTAGES AND DISADVANTAGES OF DIFFERENT METHODS

In this section, we compare the advantages and disadvantages of different methods in terms of their ability to address the challenges described in Section 2. We summarize the results of the comparison in Table 1, listing the features that can be considered advantages and indicating which methods have which features. Since the C-V and EFSR methods are based on the EFSOI, they share similar features to the EFSOI and are therefore not listed in the table. Each feature is discussed in detail in the following subsections.

4.1 | Applicability for ensemble forecasting systems

The first step in choosing a suitable method is to check whether that method can be used in a given NWP system. Since many current operational convection-permitting NWP systems are either already operating ensemble forecasting systems or are moving toward this direction, it is important to have suitable methods that can be used for ensemble forecasting systems.

It is unclear how to perform a SOSE using an ensemble forecasting system. One idea is to use the ensemble sensitivity instead of the adjoint sensitivity (Marseille *et al.*, 2008). Generally speaking, other methods listed in Table 1 are applicable in principle to ensemble forecasting systems. However, how to use them with different ensemble-based DA systems may require further research.

TABLE 1 Features of different methods can be considered as advantages when assessing the value of observations in convection-permitting NWP.

Features/methods	VR	DFS	PAI	OSE	OSSE	SOSE	EFSOI	ESA	EDA
Applicable for ensemble forecasting systems	✓	✓	✓	✓	✓	○	✓	✓	✓
Targetable via user-specified verification metric	✓	–	–	✓	✓	✓	✓	✓	✓
Applicable for future observations	○	○	○	✗	✓	✓	○	✓	✓
Computationally inexpensive	✓	✓	✓	✗	✗	✗	✓	✓	✗
Able to consider the nonlinearity of observation operators	○	○	○	✓	✓	✓	○	○	✓
Without the assumption of linear error growth	✗	–	–	✓	✓	✗	○	○	✓
Able to be separated in observation space	✓	✓	✓	✗	✗	✗	✓	✗	✗
Able to be separated in model space	✓	✗	✓	✓	✓	✓	✓	✓	✓
Able to assess the accumulated value of observations	✗	✗	✗	✓	✓	○	✗	✗	✓
Able to assess the value of anchoring	✗	✗	✓	✓	✓	✓	✓	✗	✗

The methods are the following: variance reduction (VR; Section 3.1.1), degrees of freedom for signal (DFS; Section 3.1.2), partial analysis increment (PAI; Section 3.1.4), observing-system experiment (OSE; Section 3.2.1), observing-system simulation experiment (OSSE; Section 3.2.2), sensitivity observation-system experiment (SOSE; Section 3.2.3), ensemble forecast sensitivity to observation impact (EFSOI; Section 3.2.4), ensemble sensitivity analysis (ESA; Section 3.3.2), and ensemble of data assimilations (EDA; Section 3.3.1). The symbol ✓ means "yes", the symbol ✗ means "no", the symbol ○ means somewhere in between, and the symbol – means "not applicable".

For example, the VR and DFS have been estimated in the EDA (Desroziers *et al.*, 2009), but how to estimate them in ensemble Kalman filters is largely an unexplored area. In addition, the PAI and EFSOI were originally developed for the local ensemble transform Kalman filter (LETKF) system (Hunt *et al.*, 2007); the EFSOI has been used in other deterministic ensemble Kalman filters (e.g., in the ensemble square root filter [EnSRF] system; Whitaker & Hamill, 2002; Ota *et al.*, 2013). Conceptually, these methods should also work for stochastic ensemble Kalman filters (Houtekamer & Mitchell, 2001), but there may be practical issues to consider: for example, how accurate the Kalman gain estimated using the analysis ensemble perturbations is (Equation 7), how to compute the matrix $\mathbf{Y}_a$, and whether to calculate the PAI and EFSOI for the ensemble mean or the control member. Note that the EDA method is only applicable for ensemble forecasting systems that happen to be an EDA.

4.2 | Targetability via user-specified metric

The use of forecast verification metrics in a method allows us to focus on quantifying forecast errors for features relevant to high-impact weather events (e.g., quantitative precipitation forecasting, high winds, or fog), or forecasting for particular user requirements (such as renewable energy forecasting).

The OSE, OSSE, SOSE, and ESA are able to use any user-specified metrics. For the EFSOI, users can also specify different verification norms (e.g., matrix $\mathbf{C}$ in Equation 9). The VR and EDA methods can be targeted on model variables that are relevant to convection-permitting forecasts.

4.3 | Applicability for future observations

Convection-permitting NWP requires new observation types that may not be currently available. Therefore, we need to use synthetic observations that can represent the error statistics of the new observations realistically.

By definition, an OSE is only used to assess the value of observations from existing observing systems. In contrast, the OSSE, SOSE, EDA, and ESA methods are specifically used to assess the value of future observations. One difference between an OSSE and the other methods is that in an OSSE we simulate all observations (including existing and future observations), while in the other methods we simulate only future observations (or observation-error statistics). A merit of the ESA method is that it does not require observations (only observation-error statistics and observation operators are required).

The EFSOI was developed as a computationally inexpensive alternative to the OSE, so it is usually used with real observations. However, it can be calculated using synthetic observations in the context of an OSSE. Such experiments can be used to study how different factors (e.g., forecast length) affect the accuracy of the impact estimated by the EFSOI. For example, Privé, Errico, Todling

et al. (2021) evaluated the FSOI as a function of forecast length in an OSSE. Similarly to the EFSOI, the VR, DFS, and PAI methods can be used in an OSSE.

4.4 | Computational expense

Computational cost is an important factor to consider when choosing a method to assess the value of observations, especially for convection-permitting NWP, where a statistically significant estimate is more difficult to obtain (see Sections 2.2 and 2.3). The actual computational cost depends on the complexity of the convection-permitting NWP system, the size of the problem (e.g., the size of the model domain and the number of observations), and the computational resources (e.g., the power of the supercomputer). Here, the methods that use outputs or by-products of existing NWP runs are considered to be the computationally inexpensive ones.

An OSE, an OSSE, or a SOSE is expensive due to the need to run a full operational NWP system for assessing the impact of a change to the observing system. The EDA method is also expensive because we need to run the EDA systems with and without additional observations. It is possible to use simpler configurations of the EDA (e.g., lower resolution and fewer members) to assess the value of observations in global NWP (Lang *et al.*, 2019). However, given the importance of resolution in convection-permitting NWP, this may not be appropriate for assessing the value of observations in convection-permitting NWP.

Other methods can be calculated at a relatively low cost, making use of outputs from the NWP system. The VR and DFS can be estimated using observation and analysis perturbations (Appendix B) or the background and analysis ensemble perturbation matrices. The PAI, EFSOI, and ESA methods also make use of ensemble perturbation matrices. It should also be noted that the VR, DFS, PAI, and EFSOI methods do not require the NWP system to be rerun when assessing a different set of existing observations (i.e., observations already assimilated in the system) because their results can be partitioned in observation space (see Section 4.7).

4.5 | Ability to consider nonlinearity in observation operators

When carrying out an OSE, an OSSE, or a SOSE, or calculating the ensemble spread reduction, we can use linear or nonlinear observation operators, consistent with the DA system used. The PAI, EFSOI, and ESA methods are used for ensemble-based DA systems (e.g., LETKF). The ESA method uses the background ensemble perturbation in observation space ($\mathbf{Y}_b$), which is calculated using generally nonlinear observation operators. The PAI and EFSOI methods are based on the Kalman gain given by Equation (7), where the matrix $\mathbf{Y}_a$ is obtained by a linear transformation of the matrix $\mathbf{Y}_b$ (obtained using nonlinear observation operators). However, it should be noted that ensemble-based DA systems are commonly based on the implicit assumption of linear observation operators. This means that the analysis update is not optimal in the case of nonlinear observation operators (e.g., Kugler & Weissmann, 2024b). Nevertheless, the PAI, EFSOI, and ESA methods match the analysis update in ensemble-based DA systems exactly, so they reflect the actual value of observations in such systems. The nonlinearity of observation operators can also be taken into account when estimating the VR and DFS, but how it can be done depends on the approach used to estimate them (see Appendices B and D).

4.6 | Absence of the assumption of linear error growth

When assessing the impact of observations, all methods except the VR method use fully nonlinear forecasts. However, in the EFSOI, the difference between the two forecasts ($\overline{\mathbf{x}}^{\mathrm{f}}_{\mathrm{a}}$ and $\overline{\mathbf{x}}^{\mathrm{f}}_{\mathrm{b}}$) is linearly projected back to the difference at the initial time. Similarly, in the ESA method, initial ensemble perturbations of model state variables are linearly projected onto the perturbations of the forecast metric J, as shown by Equation (F2). However, due to the use of ensemble sensitivity, the EFSOI and ESA methods are computationally robust to the strong nonlinear error growth, making them useful for convection-permitting NWP. A tangent linear approximation is used in the SOSE for the generation of synthetic observations (see Section 3.2.3) and in the VR method for the calculation of forecast error. In general, the violation of the assumption of linear error growth becomes more significant as forecast lead time increases and scales reduce (Ancell & Coleman, 2022).

4.7 | Ability to be separated in observation space

An advantage of assuming a linear forecast-error growth in assessing the impact of observations (e.g., when using the EFSOI) is that the impact estimate can be partitioned into contributions from individual observations, or the contributions from many observations can be aggregated to give an impact estimate for a set of observations (Kalnay *et al.*, 2012; Stiller, 2022; Tardif *et al.*, 2022). Partitioning and aggregation make it easy to assess the value of

different subsets of observations (e.g., without the need to rerun the NWP system). Influence estimates obtained by the VR, DFS, and PAI methods can also be separated in observation space by type, height, location, satellite channel, etc. (Desroziers *et al.*, 2005b; Diefenbach *et al.*, 2023; Hotta & Ota, 2021).

It should be noted that the value of a single observation estimated by partitioning the value of a group of observations is different from the value of a single observation estimated using a single-observation DA experiment. A single-observation experiment measures the value of a single observation when it is assimilated alone, while partitioning gives the value of a single observation when it is assimilated with other observations. In addition, the same observations may have different values when assimilated together with different observations (e.g., Kugler & Weissmann, 2024a).

For a similar reason, when observations are assimilated serially by group (Anderson, 2001; Dance, 2004), the value of each group needs to be assessed after all observations have been processed. Assessing each group directly after its assimilation will only give the value of that group when assimilated with previous observations. This also implies that, when progressively denying observations in a DDE, the order in which observations are excluded affects the estimated impacts (Healy *et al.*, 2024).

4.8 | Ability to be separated in model space

It is useful to separate the influence or impact estimates in model space by model variable and grid point. This allows us to target the variables of most interest to convection-permitting NWP (e.g., surface wind). This also allows for analyzing the influence or impact of observations in detail (e.g., the influence on other non-observed variables or influence on distant locations), which can be useful for optimizing assimilation system settings, e.g. localization. In convection-permitting systems, one would also expect the value of observations to be assessed in terms of spatial scales (Brousseau *et al.*, 2014).

Except for the DFS, all other methods allow for the assessment of the value of observations on selected model variables or grid points. For the DFS, we could possibly rewrite Equation (3) as tr(**KH**) and consider that each diagonal element of the matrix **KH** corresponds to a model variable, but it is hard to understand the meaning of this value.

Separation in the model space can be achieved to some extent through the separation of the DA system. Domain localization is a practical technique for ensemble-based data assimilation. It separates the DA problem into smaller independent local processes, each producing the analysis for a single model grid point (or several: (Janjić *et al.*, 2011; Schraff *et al.*, 2016)). In the presence of domain localization, we can estimate the DFS in each local DA process and obtain the influence of observations on the grid points that are updated in that process.

4.9 | Ability to assess the accumulated value of observations over analysis steps

The operational DA process is typically cycled at regular time intervals: for example, every 1 h for convection-permitting NWP (Milan *et al.*, 2020). The forecast produced from the analysis of the current DA cycle is used to create the background for the next DA cycle. This means that the information from the observations assimilated in one DA cycle is carried forward to the following cycles. However, the VR, DFS, PAI, EFSOI, and ESA methods can only be used to assess the value of observations for the analysis and forecast of the same cycle. In other words, they cannot be used to assess the value of the observations assimilated in previous cycles for the analyses and forecasts of the current cycle.

The OSE, OSSE, and EDA method can be carried out in a cycling environment, allowing the assessment of the accumulated values of the previous and latest observations. A benefit of this is that we can assess the value of observations for the model variables that are not directly sensitive to these observations: for example, we can show how Global Navigation Satellite System Radio Occultation (GNSS-RO) or microwave radiances will impact the wind field (Harnisch *et al.*, 2013; Healy *et al.*, 2024). The SOSE implementation is described as a single-cycle experiment, but it has the potential to be applied in a cycling environment (Marseille *et al.*, 2008).

4.10 | Ability to assess the value of anchoring

In NWP systems, relatively precise observations that can be simulated accurately by the system (e.g., GNSS-RO and radiosonde observations) are used as a reference to minimize the effect of biases in the model and other observations (e.g., Auligné *et al.*, 2007; Chandramouli *et al.*, 2022; Francis *et al.*, 2023). Bias correction is important for convection-permitting NWP, as observational and model biases are potentially larger at small scales.

When using the VR, DFS, ESA, and EDA methods, it is difficult to isolate the additional value of observations as "anchor" observations (see e.g., Healy *et al.*, 2024). To

assess the value of anchoring, methods should involve the innovation vector. For example, when using the EFSOI, we can calculate the sensitivity of the bias coefficients to the anchoring sensors, as well as the impact of the bias correction on FSOI-type diagnostics. The C-V and EFSR methods are not used to assess the value of anchoring. Another point is that, since the updated bias is cycled to the next DA cycle, we may need to calculate the impact of the bias correction in a cycling environment.

5 | OPERATIONAL CONSIDERATIONS AND FUTURE RECOMMENDATIONS

In this section, we discuss operational considerations for assessing the value of observations in convection-permitting NWP. While focusing on convection-permitting NWP, we also include some general considerations applicable to both global and convection-permitting NWP. We also provide some future recommendations for assessing the value of observations.

5.1 | Ensemble resolution, size, and spread

The EFSOI, EFSR, C-V, and ESA methods are suitable for use in convection-permitting NWP for several reasons: (1) they are computationally inexpensive, (2) they are computationally robust to the nonlinearity of error growth, and (3) they are designed specifically for ensemble forecasting systems (assuming that future convection-permitting NWP systems are ensemble-based). However, their successful use relies on the availability of suitable ensembles concerning size (number of ensemble members), domain, resolution, and output. The resolution of the ensemble should be high enough to resolve the convective-scale atmospheric processes of interest, so that we can assess the value of observations for these scales.

The ideal ensemble size is not yet clear. In an idealized study, Griewank *et al.* (2023) found that the ensemble variance reduction calculated using the ESA method reached stability when the ensemble size was about an order of magnitude smaller than the number of state variables. However, the ensemble size of current operational regional ensemble forecasting systems is typically less than 100 (TIGGE-LAM dataset, Richardson *et al.*, 2019), which is many orders of magnitude smaller than the number of state variables. In practice, we may consider using a statistical approach to correct the sampling error caused by an insufficient ensemble size (Hacker & Lei, 2015; Necker *et al.*, 2020), which may improve the accuracy of the ensemble sensitivity.

The ensemble spread can greatly affect the calculation of the sensitivity vector in the ESA method; an ensemble spread that is too small can inflate the raw sensitivity substantially (Ancell & Coleman, 2022). This can be seen from Equation (F10). In addition, if the EDA spread is underdispersed (Bonavita *et al.*, 2016), then the EDA spread reduction may underestimate the absolute value of observations on forecast skill. However, the underdispersion may not be an issue when comparing the relative impact of two observing strategies. A simple solution to the underdispersion is to apply a scaling factor to the raw EDA spread reduction. This scaling factor may be different from the covariance inflation factor used for operational forecasting. How the scaling factor should vary with the observing system requires further research (Healy *et al.*, 2024). Another solution is to improve the reliability of the EDA by modelling the sources of uncertainty better, such as improving model uncertainty parameterizations and modelling observation errors better (Healy *et al.*, 2024).

Easily accessible regional ensemble forecast datasets (e.g., TIGGE-LAM dataset, Richardson *et al.*, 2019) may be useful for observation impact studies. For example, when using the ESA method, we only need to know the background ensemble perturbation matrix if the sensitivity vector, the observation operator, and the observation uncertainty are known (see Section 3.3.2). However, when using regional ensembles, we should bear in mind that different ensembles will be optimized for the weather type of their region. Consequently, observation impact studies conducted in one region may not be informative for other regions.

5.2 | Covariance localization

In ensemble-based DA, spatial covariance localization is used mainly to remove spurious long-range error correlations between model variables when a small ensemble (that is much smaller than the model dimension) is used to estimate these correlations (Hamill *et al.*, 2001; Houtekamer & Mitchell, 2001). Covariance localization also brings some computational side benefits such as the increase in effective ensemble size (Oke *et al.*, 2007) and the decrease in computational costs (Petrie & Dance, 2010). When assessing the value of observations, covariance localization should also be applied to the methods that use ensemble covariance matrices (e.g., the PAI, EFSOI, EFSR, C-V, and ESA methods). In addition to the covariance localization, the covariance inflation used in ensemble-based DA can also affect the estimation

of the value of observations (Hotta & Ota, 2021; Kotsuki *et al.*, 2019).

5.2.1 | Flow-dependent localization

The impact of assimilated observations will evolve as the forecast proceeds, and this evolution should be reflected in the localization function (Kalnay *et al.*, 2012). This can also be explained as follows: the model grid points that an observation can impact vary with the forecast lead time (Sommer & Weissmann, 2014). How the localization function should evolve depends on the observations to be assessed and the atmospheric processes at the time of assessment (Griewank *et al.*, 2023). In general, a flow-dependent localization is needed, which should be as consistent as possible with the dynamical flow (Bocquet, 2016).

The idea of using flow-dependent localization in the assessment of observation impact is the same as the idea of using flow-dependent localization in DA. In a DA system such as 4DEnVar, we need to calculate cross-covariances between perturbations at different times within the assimilation window. A fixed-in-time localization of the cross-covariance matrices is inappropriate, as the localization should adapt to variations in the cross-covariance matrices (Bishop & Hodyss, 2007; Desroziers *et al.*, 2016).

An inappropriate treatment of the localization in the EFSOI and ESA methods may introduce large errors (Hill *et al.*, 2020; Kalnay *et al.*, 2012). For example, in the EFSOI (Equation 10), the matrix $\mathbf{Y}_a\mathbf{X}_a^\top$ is propagated by the adjoint model as

$$\mathbf{Y}_a\mathbf{X}_a^\top\mathbf{M}^\top \approx \mathbf{Y}_a\mathbf{X}_a^{f^\top},$$

and if the matrix $\mathbf{Y}_a\mathbf{X}_a^\top$ is localized, then $\mathbf{Y}_a\mathbf{X}_a^\top\mathbf{M}^\top$ needs a propagated localization. Mathematically,

$$\left(\mathbf{L}_{mn}\circ\mathbf{Y}_a\mathbf{X}_a^\top\right)\mathbf{M}^\top \neq \mathbf{L}_{mn}\circ\left(\mathbf{Y}_a\mathbf{X}_a^\top\mathbf{M}^\top\right).$$

The ESA method with an implicit sensitivity vector has a similar problem (for details see Appendix F.1).

Simply speaking, when assessing the impact of observations, different localizations are needed to assess the impact on forecasts at different lead times. Strictly speaking, a variable-dependent localization may also be needed, as tracer information (e.g., humidity and hydrometeors) will be advected, while waves (wind and temperature) propagate differently. Several studies have addressed flow-dependent localization in the calculation of the EFSOI. Kalnay *et al.* (2012) has shown that observation impact estimates could be improved by simply displacing the localization function with a mean group velocity. Ota *et al.* (2013) found that moving the centre of the localization function proportionally to the average of the analysis and forecast horizontal wind at each vertical level can account for the effect of the propagation of the observation impact by the mean flow. Gasperoni and Wang (2015) employed a Monte Carlo "group filter" technique to estimate the optimal localization from a large ensemble and found remarkable improvement for longer forecast times and at midlatitudes.

The exact calculation of the time evolution of the localization function in the computation of the sensitivity vector is not an easy task. For an operational NWP model, a precise calculation is not feasible due to the complexity and nonlinearity of model dynamics (Griewank *et al.*, 2023). The importance of flow-dependent localization depends on forecast lead times and atmospheric processes (Sommer & Weissmann, 2014). Nevertheless, flow-dependent localization may be important for quantifying the impact of observations in convection-permitting NWP, and further research is needed.

5.2.2 | The effect of localization on the rank of the ensemble covariance matrix

Covariance localization can increase the rank of the ensemble covariance matrix, which also affects the estimates of the value of observations. For example, Hotta and Ota (2021) have shown that in, optimal DA systems, an upper bound of the DFS is

$$\begin{aligned}\mathrm{DFS} &\leq \min\{\mathrm{rank}(\mathbf{R}), \mathrm{rank}(\mathbf{H}), \mathrm{rank}(\mathbf{B})\}\\ &\leq \min\{m, n\},\end{aligned}$$

where m is the number of observations assimilated and n is the number of model state variables. Furthermore, if the DFS is computed with any ensemble-based DA systems that use N ensemble members to estimate the background-error covariance matrix, then

$$\mathrm{rank}(\mathbf{P}_b) \leq N-1,$$

and hence

$$\mathrm{DFS} < N-1,$$

which indicates that the DFS can be underestimated if the ensemble size is too small (Hotta & Ota, 2021). Applying covariance localization to the background-error covariance matrix can increase its rank, and therefore the upper bound of the DFS increases to

$$\mathrm{DFS} < N\cdot\mathrm{rank}(\mathbf{L}_{nn}) - 1,$$

where $\mathbf{L}_{nn} \in \mathbb{R}^{n \times n}$ is the model-space localization matrix (Hotta & Ota, 2021).

5.3 | Simulation of observations

Observation types required for convection-permitting NWP may require the consideration of complex error statistics, such as spatial and temporal error correlations (Section 2.4). Horizontal spatial error correlations are not easy to estimate accurately. A widely used method is the Desroziers et al. method (Desroziers *et al.*, 2005a), but this requires many assumptions that are often violated in practice (e.g., Hu & Dance, 2024). In addition, including correlated observation-error statistics may increase the computational cost of the methods that require the inverse of the observation-error covariance matrix (e.g., the EFSOI).

Simulating the spatial and temporal distribution of observations is also a challenge. In an OSSE (Section 3.2.2), a NR of an NWP model is defined as the truth. When simulating observations, the NR is first interpolated to the times and locations of the real observations, and then the simulated observation errors are added. To capture convection-permitting scales, the NR needs to be available at both a very high spatial resolution and a very high temporal frequency. For a NR with very high spatiotemporal resolution, the input/output and storage requirements can be very large, especially for global models or regional models covering a large area. Alternative approaches to generating the NR and synthetic observations may be necessary depending on available computational and storage resources. There is an additional layer of complexity, in that, to make full use of very high-resolution NR, some observations should be treated differently than they have been in the past. For example, the footprint of radiance observations should be considered during the simulation, rather than doing a simple point interpolation of the NR.

Since it is not possible to produce an exact simulacrum of real observations, decisions must be made about what aspects of the real observations need to be represented faithfully by the synthetic observations. This may require testing and experimentation to determine what characteristics of observations are important when working with convection-permitting DA.

5.4 | Selection of verification reference

As discussed in Section 2.1, both analyses and observations have their own issue when used as the verification reference. The main problem with the analysis is that its error is unlikely to be independent of the forecast error. The main problem with observations is that they are often inadequate.

Using the analysis from a different NWP system may remove the error correlations, but it may introduce other problems, such as inconsistency between model grids and reduced accuracy in the analysis. We may only want to verify the forecasts using analysis from an NWP system of better or similar quality. Using the analysis produced by an independent run with the same NWP system can reduce the error correlation, but adds computational cost. Practical approaches to generating twin analysis in an operational environment are required (Hotta *et al.*, 2023).

Observations could be the most reliable verification reference for short-range forecasts. The key point in using observations is to choose observations that are well distributed in space and time and that represent the domain of interest and the (most) critical atmospheric variables for assessing a successful forecast. For convection-permitting forecasts, the atmospheric variables that are of most interest may include precipitation, wind gusts, surface wind and temperature, and total hours of sunshine (Necker *et al.*, 2018).

In an OSSE, the truth is fully known, so model fields can be compared directly with the NR to determine analysis and forecast errors, and observation impacts can be calculated without the issue of self-analysis verification. However, to verify convection-permitting forecasts, we need a NR that can reproduce convective scales realistically, especially when the NWP model has drifted into its own preferred climatology. This NR needs to be carefully validated against the real world, especially for behaviour at small scales in the geographical area of interest. Such validation requires datasets that have spatiotemporal distributions sufficient to observe this type of behaviour, similar to the challenges of verifying forecasts using observations. How the validation of the NR would occur, and what datasets would be available to affect such a validation, is the major issue.

5.5 | Selection of verification metric

In convection-permitting NWP, we are more interested in forecast skill of high-impact weather events, such as heavy precipitation and tornadoes. The verification metrics used for these events should provide spatial information (Dey, 2016; Ebert, 2008) and measure multiple aspects such as intensity, location, timing, and structure (Gilleland *et al.*, 2009; Johnson *et al.*, 2013; Johnson & Wang, 2012). For example, precipitation forecasts can be verified using the equitable threat score (Ebert *et al.*, 2003), the fractions skill score (FSS: (Necker *et al.*, 2024; Roberts & Lean, 2008)), or statistical methods measuring the spatial and temporal dependence of precipitation events (Hu & Franzke, 2020). For the evaluation of ensemble forecasts, we also need to consider the effect of the

ensemble size on the verification metric (Clark *et al.*, 2011) and combine ensemble information with spatial information (Dey *et al.*, 2016a, 2016b). In addition to measuring the forecast skill of a weather event, the overall forecast skill can be summarised by using a weighted average of skill scores for a group of variables at different forecast times (such as the Met Office UK NWP index; Simonin *et al.*, 2019).

The EFSOI and EFSR methods use a verification norm (Equation 9): for example, a dry or a moist energy norm. Compared with the dry norm, the moist norm has an additional term that measures specific humidity explicitly and thus may be more suitable for convection-permitting NWP (Necker *et al.*, 2018). However, the dry norm has a simpler form and can be effective in assessing the forecast skill of tropical cyclones (Kunii *et al.*, 2012). In general, we may need to use different norms for different regions and the assessment of different forecast properties.

In an OSSE, the choice of verification metrics ties into the validation of the NR (see Section 5.4), and also the validation of the performance of the OSSE framework as a whole. Ideally, the convective-scale model error growth in an OSSE would be representative of the error growth in the real world. However, Yu *et al.* (2019) have shown that even though an OSSE may have realistic error growth rates, this is not necessarily sufficient to achieve accurate observation impacts.

5.6 | Lateral boundary conditions

In addition to the general considerations in Section 2.3, considerations should also be given to the provision of appropriate lateral boundary conditions for a regional OSSE. A proper regional OSSE will have a regional NR embedded within a global NR, which is a challenging enterprise that is rarely done (Atlas *et al.*, 2015). If a regional NR is used with boundary conditions from a "real world" forecast of a global model (Duruisseau *et al.*, 2017), then one is constrained to looking at limited case studies, with the forecast length limited in part by the time it takes for information to propagate from the boundary conditions. This type of NR is best suited to localized observing networks and analysis or very short forecasts, which is mostly what we are interested in for convection-permitting NWP.

5.7 | Retuning of the background-error covariance matrix

When carrying out an OSE, an OSSE, or a SOSE, an open question is whether we should retune the background-error covariance matrices for each combination of assimilated observations (Boukabara *et al.*, 2020; Duncan *et al.*, 2021). Theoretically, changes in observations will affect the estimation of the background-error covariance matrix. For example, when using the EDA to estimate the background-error covariance matrix, changes in observations will alter the EDA spread and hence the resultant matrix. Practically, Duncan *et al.* (2021) found that the effect of retuning the background-error covariance matrices is small and should not change the conclusions drawn from a typical OSE (without retuning the background-error covariance matrix).

Given the additional computational cost of rerunning the EDA, it is worth discussing whether the background-error covariance matrix needs to be updated, and which kind of changes to the observing system might require the retuning of the background-error covariance matrix to ensure reliable conclusions (Healy *et al.*, 2024). In low-baseline experiments, where large proportions of observations are to be denied, the use of the background-error covariance matrix obtained from the full observing system may be more problematic (Kelly *et al.*, 2008).

6 | SUMMARY

Assessing the value of observations in convection-permitting NWP can be challenging due to factors such as verification of convective-scale features, strong nonlinearities in the forecast model and observation operator, the LAM domain, and the complexity of observation-error statistics. These factors can create difficulties in selecting verification references and metrics, make the use of the adjoint sensitivity inappropriate (unless we have skilful TLMs at convective scale), introduce errors in influence and impact estimates, increase the difficulty of obtaining statistically significant results, and complicate the simulation of observations.

To provide guidance for the assessment, we have compared existing methods in terms of their ability to address the challenges (see Table 1), discussed issues that should be considered in their operational use, and identified future research directions. Table 2 summarizes the key comments and research directions for each method. General research directions include comparing results from different methods (e.g., on different model variables and different spatial scales), applying methods to different DA systems (e.g., applying the PAI and EFSOI methods to stochastic ensemble Kalman filters), comparing the use of observations versus analyses as the verification reference, applying

TABLE 2 Key remarks and future research directions for each method: Variance reduction (VR; Section 3.1.1), degrees of freedom for signal (DFS; Section 3.1.2), partial analysis increment (PAI; Section 3.1.4), observing system experiment (OSE; Section 3.2.1), observing system simulation experiment (OSSE; Section 3.2.2), sensitivity observation system experiment (SOSE; Section 3.2.3), ensemble forecast sensitivity to observation impact (EFSOI; Section 3.2.4), cross-validation (C-V; Section 3.4.2), ensemble forecast sensitivity to observation error covariance (EFSR; Section 3.4.3), ensemble of data assimilations (EDA; Section 3.3.1) and ensemble sensitivity analysis (ESA; Section 3.3.2).

Methods	Key remarks	Future research directions
VR	Measure the difference between the background and analysis error variance.	Applications in ensemble-based DA systems.
DFS	• Measure the normalized difference between the background and analysis error variance. • Widely used in satellite meteorology.	• Applications in ensemble-based DA systems. • Domain localization, covariance localization and covariance inflation.
PAI	• Recently developed for studying 3-D observation influence in ensemble-based DA systems. • Measure the partial analysis increments caused by single observations or a subset.	• Application and evaluation (e.g., optimizing localization). • Comparison to other methods (e.g., VR and DFS).
OSE	• The mainstay of assessing the value of existing observations at operational NWP centres. • Measure the difference in forecast skill due to removing, adding, or replacing a subset of observations.	• Verification reference (analyses versus observations) • Retuning of the background error covariance matrix when observations are changed.
OSSE	• Widely used to assess the impact of future observing systems. • Performance depends on the quality of the OSSE system.	• Generating and validating the NR for convective scales. • Simulating high-resolution observations with complicated error statistics. • Lateral boundary conditions for regional OSSEs.
SOSE	• A less common method for assessing the impact of future observations. • Use adjoint sensitivity.	• Application and evaluation. • Comparison to other methods (e.g., the OSSE).
EFSOI	• A computationally efficient alternative to an OSE for ensemble forecasting systems. • Use ensemble sensitivity.	• Verification reference (analyses versus observations). • Flow-dependent localization. • Effect of ensemble size, spread and resolution. • Application to other ensemble-based DA systems than the LETKF.
EFSR	• A relatively new method based on the EFSOI. • Assess the impact of observation errors	Application and evaluation.
C-V	• New diagnostics obtained by partitioning the observation-based EFSOI. • Cross-validation between assimilated and verifying observations	Application and evaluation (e.g., when assimilated and verifying observations are at different times).
EDA	• Measure the reduction in ensemble spread due to the assimilation of additional observations. • Rely on the spread-skill relationship.	• Application and evaluation in convection-permitting NWP. • Deal with under dispersion of ensemble spread.
ESA	• Measure the reduction in the ensemble variance of a forecast metric due to the assimilation of additional observations. • Use ensemble sensitivity. • Rely on the spread-skill relationship.	• Effect of ensemble size, spread and resolution. • Flow-dependent localization.

flow-dependent localization, simulating high-resolution observations with complicated error statistics, and investigating the effect of ensemble quality on assessing the value of observations.

Generally, we need to be mindful of several things when assessing the value of observations in convection-permitting NWP. The first thing to note is that different methods may have a different focus for use. For example, the OSSE, EDA, and ESA methods are specifically used to assess the impact of future observations. The DFS, PAI, C-V, and EFSR methods are particularly suitable for identifying the problems with the assimilation system and assessing changes to the assimilation system. An OSE assesses actual improvement (or degradation) in forecast

skill and analysis quality in operational NWP. Therefore, its result is often used to validate the results from other methods. Decisions on which methods to use should be based on the user's specific purpose and the resources available at the NWP centre. For example, establishing and maintaining an OSSE system is no easy task. In general, it is sensible to use more than one method to avoid the misinterpretation of the result from any single method. This is particularly important when using synthetic observations to assess the value of future observations.

The second thing to note is that the same observations can have different values in different convection-permitting NWP systems. Convection-permitting NWP typically uses a LAM, which can vary considerably between tropical and midlatitude regions due to differences in the dominant atmospheric processes (Gelaro *et al.*, 2010; Privé, Errico, Todling *et al.*, 2021). Methods for assessing the value of observations are usually operationally oriented and provide us with statistical estimates of the value of observations. However, we may need to understand the underlying dynamics that cause the results of these methods. Furthermore, the quality of convection-permitting NWP systems can vary considerably due to factors such as differences in local observation networks (e.g., large parts of the world have hardly any conventional observations and no radar) and differences in NWP models. A small added value should be expected when adding observations to a convection-permitting NWP system that already has good quality and assimilates extensive observations.

To summarize, appropriate selection and sensible use of the available methods, and careful interpretation of the results, can give us a reliable assessment of the value of observations, which can help us to improve the observing system and the DA system, and thus improve the forecast skill in convection-permitting NWP.

ACKNOWLEDGEMENTS

We acknowledge the discussions that led to this article via the World Meteorological Organization (WMO) World Weather Research Programme (WWRP) Data Assimilation and Observing Systems (DAOS), Predictability, Dynamics, and Ensemble Forecasting (PDEF), and Nowcasting and Mesoscale Research (NMR) working groups.

G. Hu was funded in part by the Natural Environment Research Council (NERC) National Centre for Earth Observation (NCEO) and UK Met Office. S. L. Dance and A. Fowler were funded in part by the NERC NCEO. U. Löhnert has been supported by the Hans-Ertel-Centre for Weather Research funded by the German Federal Ministry for Transportation and Digital Infrastructure (Grant number BMVI/DWD 4818DWD5B). X. Wang is supported by NOAA (NA19OAR0220154).

CONFLICT OF INTEREST STATEMENT

We declare we have no competing interests.

DATA AVAILABILITY STATEMENT

No data were used or created.

ORCID

Guannan Hu https://orcid.org/0000-0003-4305-3658
Sarah L. Dance https://orcid.org/0000-0003-1690-3338
Alison Fowler https://orcid.org/0000-0003-3650-3948
David Simonin https://orcid.org/0000-0002-2623-3828
Joanne Waller https://orcid.org/0000-0002-7783-6434
Sean Healy https://orcid.org/0000-0003-4810-9593
Daisuke Hotta https://orcid.org/0000-0003-2287-0608
Ulrich Löhnert https://orcid.org/0000-0002-9023-0269
Olaf Stiller https://orcid.org/0000-0001-7953-7393

REFERENCES

Ancell, B. & Hakim, G.J. (2007) Comparing Adjoint- and Ensemble-Sensitivity Analysis with Applications to Observation Targeting. *Monthly Weather Review*, 135(12), 4117–4134. Available from: https://doi.org/10.1175/2007MWR1904.1

Ancell, B.C. & Coleman, A.A. (2022) New Perspectives on Ensemble Sensitivity Analysis with Applications to a Climatology of Severe Convection. *Bulletin of the American Meteorological Society*, 103(2), E511–E530. Available from: https://doi.org/10.1175/BAMS-D-20-0321.1

Anderson, J.L. (2001) An Ensemble Adjustment Kalman Filter for Data Assimilation. *Monthly Weather Review*, 129(12), 2884–2903. Available from: https://doi.org/10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2

Atlas, R., Hoffman, R.N., Ma, Z., Emmitt, G.D., Wood, S.A., Greco, S. et al. (2015) Observing System Simulation Experiments (OSSEs) to Evaluate the Potential Impact of an Optical Autocovariance Wind Lidar (OAWL) on Numerical Weather Prediction. *Journal of Atmospheric and Oceanic Technology*, 32(9), 1593–1613. Available from: https://doi.org/10.1175/JTECH-D-15-0038.1

Auligné, T., McNally, A.P. & Dee, D.P. (2007) Adaptive bias correction for satellite data in a numerical weather prediction system. *Quarterly Journal of the Royal Meteorological Society*, 133(624), 631–642. Available from: https://doi.org/10.1002/qj.56

Bachmann, K., Keil, C., Craig, G.C., Weissmann, M. & Welzbacher, C.A. (2020) Predictability of Deep Convection in Idealized and Operational Forecasts: Effects of Radar Data Assimilation, Orography, and Synoptic Weather Regime. *Monthly Weather Review*, 148(1), 63–81. Available from: https://doi.org/10.1175/MWR-D-19-0045.1

Bachmann, K., Keil, C. & Weissmann, M. (2019) Impact of radar data assimilation and orography on predictability of deep convection. *Quarterly Journal of the Royal Meteorological Society*, 145(718), 117–130. Available from: https://doi.org/10.1002/qj.3412

Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M. & Reinhardt, T. (2011) Operational Convective-Scale Numerical Weather Prediction with the COSMO Model: Description and Sensitivities. *Monthly Weather Review*, 139(12), 3887–3905. Available from: https://doi.org/10.1175/MWR-D-10-05013.1

Bannister, R.N. (2021) Balance conditions in variational data assimilation for a high-resolution forecast model. *Quarterly Journal of the Royal Meteorological Society*, 147(738), 2917–2934. Available from: https://doi.org/10.1002/qj.4106

Bauer, P., Auligné, T., Bell, W., Geer, A., Guidard, V., Heilliette, S. et al. (2011) Satellite cloud and precipitation assimilation at operational NWP centres. *Quarterly Journal of the Royal Meteorological Society*, 137(661), 1934–1951. Available from: https://doi.org/10.1002/qj.905

Bell, Z., Dance, S.L. & Waller, J.A. (2022) Exploring the characteristics of a vehicle-based temperature dataset for kilometre-scale data assimilation. *Meteorological Applications*, 29(3), e2058. Available from: https://doi.org/10.1002/met.2058

Bernstein, D.S. (2009) *Matrix Mathematics: Theory, Facts, and Formulas*, 2nd edition. Princeton: Princeton University Press. Available from: https://doi.org/10.1515/9781400833344

Bishop, C.H. & Hodyss, D. (2007) Flow-adaptive moderation of spurious ensemble correlations and its use in ensemble-based data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 133(629), 2029–2044. Available from: https://doi.org/10.1002/qj.169

Bocquet, M. (2016) Localization and the iterative ensemble Kalman smoother. *Quarterly Journal of the Royal Meteorological Society*, 142(695), 1075–1089. Available from: https://doi.org/10.1002/qj.2711

Bonavita, M., Hólm, E., Isaksen, L. & Fisher, M. (2016) The evolution of the ECMWF hybrid data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 142(694), 287–303. Available from: https://doi.org/10.1002/qj.2652

Bormann, N. & Bauer, P. (2010) Estimates of spatial and interchannel observation-error characteristics for current sounder radiances for numerical weather prediction. I: Methods and application to ATOVS data. *Quarterly Journal of the Royal Meteorological Society*, 136(649), 1036–1050. Available from: https://doi.org/10.1002/qj.616

Bormann, N., Collard, A. & Bauer, P. (2010) Estimates of spatial and interchannel observation-error characteristics for current sounder radiances for numerical weather prediction. II: Application to AIRS and IASI data. *Quarterly Journal of the Royal Meteorological Society*, 136(649), 1051–1063. Available from: https://doi.org/10.1002/qj.615

Boukabara, S., Park, S., Riishojgaard, L.P. & Eyre, J. (Eds.). (2020) *The 7th WMO Workshop on the Impact of Various Observing Systems on NWP: Final Report*. Geneva, Switzerland: WMO. Available from: https://community.wmo.int/en/meetings/NWP-7

Bouttier, F. & Courtier, P. (2002) Data Assimilation Concepts and Methods. https://www.ecmwf.int/sites/default/files/elibrary/2002/16928-data-assimilation-concepts-and-methods.pdf

Bouttier, F. & Kelly, G. (2001) *Observing systems experiments in the ECMWF 4D-Var data assimilation system*. Technical Report 313, Reading: ECMWF, Shinfield Park. Available from: https://www.ecmwf.int/node/8351

Brousseau, P., Desroziers, G., Bouttier, F. & Chapnik, B. (2014) A posteriori diagnostics of the impact of observations on the AROME-France convective-scale data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 140(680), 982–994. Available from: https://doi.org/10.1002/qj.2179

Brousseau, P., Seity, Y., Ricard, D. & Léger, J. (2016) Improvement of the forecast of convective activity from the AROME-France system. *Quarterly Journal of the Royal Meteorological Society*, 142(699), 2231–2243. Available from: https://doi.org/10.1002/qj.2822

Buehner, M., Du, P. & Bédard, J. (2018) A New Approach for Estimating the Observation Impact in Ensemble–Variational Data Assimilation. *Monthly Weather Review*, 146(2), 447–465. Available from: https://doi.org/10.1175/MWR-D-17-0252.1

Candy, B., Cotton, J. & Eyre, J. (2021) *Recent results of observation data denial experiments*. Met Office Forecasting Research Technical Report 641. Exeter, United Kingdom: Met Office. Available from: https://digital.nmla.metoffice.gov.uk/IO_441cf4bf-3ca9-4f52-8e66-f640ac8eadd1/

Cardinali, C. (2018) Forecast sensitivity observation impact with an observation-only based objective function. *Quarterly Journal of the Royal Meteorological Society*, 144(716), 2089–2098. Available from: https://doi.org/10.1002/qj.3305

Cardinali, C., Pezzulli, S. & Andersson, E. (2004) *Influence matrix diagnostic of a data assimilation system*. Technical Report 450. Reading: ECMWF, Shinfield Park. Available from: https://www.ecmwf.int/node/8592

Caumont, O., Cimini, D., Löhnert, U., Alados-Arboledas, L., Bleisch, R., Buffa, F. et al. (2016) Assimilation of humidity and temperature observations retrieved from ground-based microwave radiometers into a convective-scale NWP model. *Quarterly Journal of the Royal Meteorological Society*, 142(700), 2692–2704. Available from: https://doi.org/10.1002/qj.2860

Chandramouli, K., Wang, X., Johnson, A. & Otkin, J. (2022) Online Nonlinear Bias Correction in Ensemble Kalman Filter to Assimilate GOES-R All-Sky Radiances for the Analysis and Prediction of Rapidly Developing Supercells. *Journal of Advances in Modeling Earth Systems*, 14(3), e2021MS002711. Available from: https://doi.org/10.1029/2021MS002711

Chapnik, B., Desroziers, G., Rabier, F. & Talagrand, O. (2006) Diagnosis and tuning of observational error in a quasi-operational data assimilation setting. *Quarterly Journal of the Royal Meteorological Society*, 132(615), 543–565. Available from: https://doi.org/10.1256/qj.04.102

Chen, T.-C. & Kalnay, E. (2019) Proactive Quality Control: Observing System Simulation Experiments with the Lorenz'96 Model. *Monthly Weather Review*, 147(1), 53–67. Available from: https://doi.org/10.1175/MWR-D-18-0138.1

Chen, T.-C. & Kalnay, E. (2020) Proactive Quality Control: Observing System Experiments Using the NCEP Global Forecast System. *Monthly Weather Review*, 148(9), 3911–3931. Available from: https://doi.org/10.1175/MWR-D-20-0001.1

Chipilski, H.G., Wang, X. & Parsons, D.B. (2020) Impact of Assimilating PECAN Profilers on the Prediction of Bore-Driven Nocturnal Convection: A Multiscale Forecast Evaluation for the 6 July 2015 Case Study. *Monthly Weather Review*, 148(3), 1147–1175. Available from: https://doi.org/10.1175/MWR-D-19-0171.1

Chipilski, H.G., Wang, X., Parsons, D.B., Johnson, A. & Degelia, S.K. (2022) The Value of Assimilating Different Ground-Based Profiling Networks on the Forecasts of Bore-Generating Nocturnal Convection. *Monthly Weather Review*, 150(6), 1273–1292. Available from: https://doi.org/10.1175/MWR-D-21-0193.1

Clark, A.J., Kain, J.S., Stensrud, D.J., Xue, M., Kong, F., Coniglio, M.C. et al. (2011) Probabilistic Precipitation Forecast Skill as a Function of Ensemble Size and Spatial Scale in a Convection-Allowing Ensemble. *Monthly Weather Review*, 139(5), 1410–1418. Available from: https://doi.org/10.1175/2010MWR3624.1

Clark, P., Roberts, N., Lean, H., Ballard, S.P. & Charlton-Perez, C. (2016) Convection-permitting models: a step-change in rainfall forecasting. *Meteorological Applications*, 23(2), 165–181. Available from: https://doi.org/10.1002/met.1538

Collard, A.D. (2007) Selection of IASI channels for use in numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 133(629), 1977–1991. Available from: https://doi.org/10.1002/qj.178

Cordoba, M., Dance, S.L., Kelly, G.A., Nichols, N.K. & Waller, J.A. (2017) Diagnosing atmospheric motion vector observation errors for an operational high-resolution data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 143(702), 333–341. Available from: https://doi.org/10.1002/qj.2925

Courtier, P., Andersson, E., Heckley, W., Vasiljevic, D., Hamrud, M., Hollingsworth, A. et al. (1998) The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation. *Quarterly Journal of the Royal Meteorological Society*, 124(550), 1783–1807. Available from: https://doi.org/10.1002/qj.49712455002

Cover, T.M. & Thomas, J.A. (1991) *Elements of Information Theory (Wiley Series in Telecommunications)*. New York, NY: John Wiley & Sons.

Cress, A. & Wergen, W. (2001) Impact of profile observations on the German Weather Service's NWP system. *Meteorologische Zeitschrift*, 10(2), 91–101. Available from: https://doi.org/10.1127/0941-2948/2001/0010-0091

Cucurull, L., Atlas, R., Li, R., Mueller, M.J. & Hoffman, R.N. (2018) An Observing System Simulation Experiment with a Constellation of Radio Occultation Satellites. *Monthly Weather Review*, 146(12), 4247–4259. Available from: https://doi.org/10.1175/MWR-D-18-0089.1

Daescu, D.N. (2008) On the Sensitivity Equations of Four-Dimensional Variational (4D-Var) Data Assimilation. *Monthly Weather Review*, 136(8), 3050–3065. Available from: https://doi.org/10.1175/2007MWR2382.1

Daescu, D.N. (2009) On the Deterministic Observation Impact Guidance: A Geometrical Perspective. *Monthly Weather Review*, 137(10), 3567–3574. Available from: https://doi.org/10.1175/2009MWR2954.1

Daescu, D.N. & Langland, R.H. (2013) Error covariance sensitivity and impact estimation with adjoint 4D-Var: theoretical aspects and first applications to NAVDAS-AR. *Quarterly Journal of the Royal Meteorological Society*, 139(670), 226–241. Available from: https://doi.org/10.1002/qj.1943

Daescu, D.N. & Langland, R.H. (2017) *Toward New Applications of the Adjoint Sensitivity Tools in Data Assimilation*. Cham: Springer International Publishing, pp. 361–382. Available from: https://doi.org/10.1007/978-3-319-43415-5.16

Dance, S.L. (2004) Issues in high resolution limited area data assimilation for quantitative precipitation forecasting. *Physica D: Nonlinear Phenomena*, 196(1), 1–27. Available from: https://doi.org/10.1016/j.physd.2004.05.001

Dance, S.L., Ballard, S.P., Bannister, R.N., Clark, P., Cloke, H.L., Darlington, T. et al. (2019) Improvements in Forecasting Intense Rainfall: Results from the FRANC (Forecasting Rainfall Exploiting New Data Assimilation Techniques and Novel Observations of Convection) Project. *Atmosphere*, 10(3), 125. Available from: https://doi.org/10.3390/atmos10030125

Degelia, S.K., Wang, X., Stensrud, D.J. & Turner, D.D. (2020) Systematic Evaluation of the Impact of Assimilating a Network of Ground-Based Remote Sensing Profilers for Forecasts of Nocturnal Convection Initiation during PECAN. *Monthly Weather Review*, 148(12), 4703–4728. Available from: https://doi.org/10.1175/MWR-D-20-0118.1

Desroziers, G., Arbogast, E. & Berre, L. (2016) Improving spatial localization in 4DEnVar. *Quarterly Journal of the Royal Meteorological Society*, 142(701), 3171–3185. Available from: https://doi.org/10.1002/qj.2898

Desroziers, G., Berre, L., Chabot, V. & Chapnik, B. (2009) A Posteriori Diagnostics in an Ensemble of Perturbed Analyses. *Monthly Weather Review*, 137(10), 3420–3436. Available from: https://doi.org/10.1175/2009MWR2778.1

Desroziers, G., Berre, L., Chapnik, B. & Poli, P. (2005a) Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 131(613), 3385–3396.

Desroziers, G., Brousseau, P. & Chapnik, B. (2005b) Use of randomization to diagnose the impact of observations on analyses and forecasts. *Quarterly Journal of the Royal Meteorological Society*, 131(611), 2821–2837. Available from: https://doi.org/10.1256/qj.04.151

Desroziers, G. & Ivanov, S. (2001) Diagnosis and adaptive tuning of observation-error parameters in a variational assimilation. *Quarterly Journal of the Royal Meteorological Society*, 127(574), 1433–1452. Available from: https://doi.org/10.1002/qj.49712757417

Dey, S.R. (2016) *A spatial approach to the analysis of convective scale ensemble systems*. PhD thesis. Reading, United Kingdom: University of Reading.

Dey, S.R.A., Plant, R.S., Roberts, N.M. & Migliorini, S. (2016a) Assessing spatial precipitation uncertainties in a convective-scale ensemble. *Quarterly Journal of the Royal Meteorological Society*, 142(701), 2935–2948. Available from: https://doi.org/10.1002/qj.2893

Dey, S.R.A., Roberts, N.M., Plant, R.S. & Migliorini, S. (2016b) A new method for the characterization and verification of local spatial predictability for convective-scale ensembles. *Quarterly Journal of the Royal Meteorological Society*, 142(698), 1982–1996. Available from: https://doi.org/10.1002/qj.2792

Diefenbach, T., Craig, G., Keil, C., Scheck, L. & Weissmann, M. (2023) Partial Analysis Increments as Diagnostic for LETKF Data Assimilation Systems. *Quarterly Journal of the Royal Meteorological Society*, 149(752), 740–756. Available from: https://doi.org/10.1002/qj.4419

Duncan, D.I., Bormann, N. & Hólm, E.V. (2021) On the addition of microwave sounders and numerical weather prediction skill. *Quarterly Journal of the Royal Meteorological Society*, 147(740), 3703–3718. Available from: https://doi.org/10.1002/qj.4149

Duruisseau, F., Chambon, P., Guedj, S., Guidard, V., Fourrié, N., Taillefer, F. et al. (2017) Investigating the potential benefit to a mesoscale NWP model of a microwave sounder on board a geostationary satellite. *Quarterly Journal of the Royal Meteorological Society*, 143(706), 2104–2115. Available from: https://doi.org/10.1002/qj.3070

Ebert, E.E. (2008) Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework. *Meteorological Applications*, 15(1), 51–64. Available from: https://doi.org/10.1002/met.25

Ebert, E.E., Damrath, U., Wergen, W. & Baldwin, M.E. (2003) The WGNE Assessment of Short-term Quantitative Precipitation Forecasts. *Bulletin of the American Meteorological Society*, 84, 481–492. Available from: https://doi.org/10.1175/BAMS-84-4-481

Ehrendorfer, M. & Errico, R.M. (1995) Mesoscale Predictability and the Spectrum of Optimal Perturbations. *Journal of Atmospheric Sciences*, 52(20), 3475–3500. Available from: https://journals.ametsoc.org/view/journals/atsc/52/20/1520-0469_1995_052_3475_mpatso_2_0_co_2.xml

Ehrendorfer, M., Errico, R.M. & Raeder, K.D. (1999) Singular-Vector Perturbation Growth in a Primitive Equation Model with Moist Physics. *Journal of the Atmospheric Sciences*, 56(11), 1627–1648. Available from: https://doi.org/10.1175/1520-0469(1999)056<1627:SVPGIA>2.0.CO;2

Errico, R.M. (2007) Interpretations of an adjoint-derived observational impact measure. *Tellus A: Dynamic Meteorology and Oceanography*, 59(2), 273–276. Available from: https://doi.org/10.1111/j.1600-0870.2006.00217.x

Errico, R.M. & Privé, N.C. (2018) Some general and fundamental requirements for designing observing system simulation experiments (OSSEs). https://ntrs.nasa.gov/api/citations/20190025338/downloads/20190025338.pdf

Eyre, J.R. (2021a) Observation impact metrics in NWP: A theoretical study. Part I: Optimal systems. *Quarterly Journal of the Royal Meteorological Society*, 147(739), 3180–3200. Available from: https://doi.org/10.1002/qj.4123

Eyre, J.R. (2021b) *Metrics for assessing the impact of observations in NWP: a theoretical study. Part II: suboptimal systems*. Technical Report. Exeter, United Kingdom: Met Office. Available from: https://digital.nmla.metoffice.gov.uk/IO_143718fb-a7a5-4220-9635-e4985192a097

Fisher, M. (2003) *Estimation of entropy reduction and degrees of freedom for signal for large variational analysis systems*. Technical Report 397. Reading: ECMWF, Shinfield Park. Available from: https://www.ecmwf.int/node/9402

Fourrié, N., Doerenbecher, A., Bergot, T. & Joly, A. (2002) Adjoint sensitivity of the forecast to TOVS observations. *Quarterly Journal of the Royal Meteorological Society*, 128(586), 2759–2777. Available from: https://doi.org/10.1256/qj.01.167

Fowler, A. (2019) Data compression in the presence of observational error correlations. *Tellus A: Dynamic Meteorology and Oceanography*, 71(1), 1634937. Available from: https://doi.org/10.1080/16000870.2019.1634937

Fowler, A. & Leeuwen, P.J.V. (2013) Observation impact in data assimilation: the effect of non-Gaussian observation error. *Tellus A: Dynamic Meteorology and Oceanography*, 65(1), 20035. Available from: https://doi.org/10.3402/tellusa.v65i0.20035

Fowler, A. & Van Leeuwen, P. (2012) Measures of observation impact in non-Gaussian data assimilation. *Tellus A: Dynamic Meteorology and Oceanography*, 64(1), 17192. Available from: https://doi.org/10.3402/tellusa.v64i0.17192

Fowler, A.M. (2017) A Sampling Method for Quantifying the Information Content of IASI Channels. *Monthly Weather Review*, 145(2), 709–725. Available from: https://doi.org/10.1175/MWR-D-16-0069.1

Fowler, A.M., Dance, S.L. & Waller, J.A. (2018) On the interaction of observation and prior error correlations in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 144(710), 48–62. Available from: https://doi.org/10.1002/qj.3183

Fowler, A.M., Simonin, D. & Waller, J.A. (2020) Measuring Theoretical and Actual Observation Influence in the Met Office UKV: Application to Doppler Radial Winds. *Geophysical Research Letters*, 47(24), e2020GL091110. Available from: https://doi.org/10.1029/2020GL091110

Francis, D.J., Fowler, A.M., Lawless, A.S., Eyre, J. & Migliorini, S. (2023) The effective use of anchor observations in variational bias correction in the presence of model bias. *Quarterly Journal of the Royal Meteorological Society*, 149(754), 1789–1809. Available from: https://doi.org/10.1002/qj.4482

Fujita, T., Seko, H., Kawabata, T., Ikuta, Y., Sawada, K., Hotta, D. et al. (2020) Variational Data Assimilation with Spatial and Temporal Observation Error Correlations of Doppler Radar Radial Winds. In: *Research activities in Earth system modelling. Working Group on Numerical Experimentation Report No 50*. Geneva: WMO. Available from: https://wgne.net/bluebook/

Gasperoni, N.A. & Wang, X. (2015) Adaptive localization for the ensemble-based observation impact estimate using regression confidence factors. *Monthly Weather Review*, 143(6), 1981–2000. Available from: https://doi.org/10.1175/MWR-D-14-00272.1

Gasperoni, N.A., Wang, X., Brewster, K.A. & Carr, F.H. (2024) Exploring Ensemble Forecast Sensitivity to Observations for a Convective-Scale Data Assimilation System over the Dallas–Fort Worth Testbed. *Monthly Weather Review*, 152(2), 571–588. Available from: https://doi.org/10.1175/MWR-D-23-0091.1

Geer, A.J. (2016) Significance of changes in medium-range forecast scores. *Tellus A: Dynamic Meteorology and Oceanography*, 68(1), 30229. Available from: https://doi.org/10.3402/tellusa.v68.30229

Geer, A.J. (2019) Correlated observation error models for assimilating all-sky infrared radiances. *Atmospheric Measurement Techniques*, 12(7), 3629–3657. Available from: https://doi.org/10.5194/amt-12-3629-2019

Geerts, B., Parsons, D., Ziegler, C.L., Weckwerth, T.M., Biggerstaff, M.I., Clark, R.D. et al. (2017) The 2015 Plains Elevated Convection at Night Field Project. *Bulletin of the American Meteorological Society*, 98(4), 767–786. Available from: https://doi.org/10.1175/BAMS-D-15-00257.1

Geiss, S., Scheck, L., de Lozar, A. & Weissmann, M. (2021) Understanding the model representation of clouds based on visible and infrared satellite observations. *Atmospheric Chemistry and Physics*, 21(16), 12273–12290. Available from: https://doi.org/10.5194/acp-21-12273-2021

Gelaro, R., Langland, R.H., Pellerin, S. & Todling, R. (2010) The THORPEX Observation Impact Intercomparison Experiment. *Monthly Weather Review*, 138(11), 4009–4025. Available from: https://doi.org/10.1175/2010MWR3393.1

Gelaro, R. & Zhu, Y. (2009) Examination of observation impacts derived from observing system experiments (OSEs) and adjoint models. *Tellus A*, 61(2), 179–193. Available from: https://doi.org/10.1111/j.1600-0870.2008.00388.x

Gilleland, E., Ahijevych, D., Brown, B.G., Casati, B. & Ebert, E.E. (2009) Intercomparison of Spatial Forecast Verification Methods. *Weather and Forecasting*, 24(5), 1416–1430. Available from: https://doi.org/10.1175/2009WAF2222269.1

Girard, D.A. (1989) A Fast 'Monte-Carlo Cross-Validation' Procedure for Large Least Squares Problems with Noisy Data. *Numerische Mathematik*, 56(1), 1–24. Available from: http://eudml.org/doc/133381

Griewank, P.J., Weissmann, M., Necker, T., Nomokonova, T. & Löhnert, U. (2023) Ensemble-based estimates of the impact of potential observations. *Quarterly Journal of the Royal Meteorological Society*, 149(754), 1546–1571. Available from: https://doi.org/10.1002/qj.4464

Gustafsson, N., Janjić, T., Schraff, C., Leuenberger, D., Weissmann, M., Reich, H. et al. (2018) Survey of data assimilation methods for convective-scale numerical weather prediction at operational centres. *Quarterly Journal of the Royal Meteorological Society*, 144(713), 1218–1256. Available from: https://doi.org/10.1002/qj.3179

Gustafsson, N., Källén, E. & Thorsteinsson, S. (1998) Sensitivity of forecast errors to initial and lateral boundary conditions. *Tellus A: Dynamic Meteorology and Oceanography*, 50(2), 167–185. Available from: https://doi.org/10.3402/tellusa.v50i2.14519

Hacker, J.P. & Lei, L. (2015) Multivariate Ensemble Sensitivity with Localization. *Monthly Weather Review*, 143(6), 2013–2027. Available from: https://doi.org/10.1175/MWR-D-14-00309.1

Hagelin, S., Azad, R., Lindskog, M., Schyberg, H. & Körnich, H. (2021) Evaluating the use of Aeolus satellite observations in the regional numerical weather prediction (NWP) model Harmonie–Arome. *Atmospheric Measurement Techniques*, 14(9), 5925–5938. Available from: https://doi.org/10.5194/amt-14-5925-2021

Hakim, G.J., Bumbaco, K.A., Tardif, R. & Powers, J.G. (2020) Optimal Network Design Applied to Monitoring and Forecasting Surface Temperature in Antarctica. *Monthly Weather Review*, 148(2), 857–873. Available from: https://doi.org/10.1175/MWR-D-19-0103.1

Hakim, G.J. & Torn, R.D. (2008) Ensemble Synoptic Analysis. *Meteorological Monographs*, 33, 147–162. Available from: https://doi.org/10.1175/0065-9401-33.55.147

Hamill, T.M., Whitaker, J.S. & Snyder, C. (2001) Distance-Dependent Filtering of Background Error Covariance Estimates in an Ensemble Kalman Filter. *Monthly Weather Review*, 129(11), 2776–2790. Available from: https://doi.org/10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2

Harnisch, F., Healy, S.B., Bauer, P. & English, S.J. (2013) Scaling of GNSS Radio Occultation Impact with Observation Number Using an Ensemble of Data Assimilations. *Monthly Weather Review*, 141(12), 4395–4413. Available from: https://doi.org/10.1175/MWR-D-13-00098.1

Healy, S., Bormann, N., Geer, A., Holm, E., Ingleby, B., Lean, K. et al. (2024) *Methods for assessing the impact of current and future components of the global observing system. ECMWF Technical Memoranda*, 916. Reading, United Kingdom: ECMWF. Available from: https://doi.org/10.21957/2f240fe55f

Healy, S.B. & Thépaut, J.-N. (2006) Assimilation experiments with CHAMP GPS radio occultation measurements. *Quarterly Journal of the Royal Meteorological Society*, 132(615), 605–623. Available from: https://doi.org/10.1256/qj.04.182

Hill, A.J., Weiss, C.C. & Ancell, B.C. (2020) Factors Influencing Ensemble Sensitivity–Based Targeted Observing Predictions at Convection-Allowing Resolutions. *Monthly Weather Review*, 148(11), 4497–4517. Available from: https://doi.org/10.1175/MWR-D-20-0015.1

Hintz, K.S., O'Boyle, K., Dance, S.L., Al-Ali, S., Ansper, I., Blaauboer, D. et al. (2019) Collecting and utilising crowdsourced data for numerical weather prediction: Propositions from the meeting held in Copenhagen, 4–5 December 2018. *Atmospheric Science Letters*, 20(7), e921. Available from: https://doi.org/10.1002/asl.921

Hoffman, R.N. & Atlas, R. (2016) Future observing system simulation experiments. *Bulletin of the American Meteorological Society*, 97(9), 1601–1616. Available from: https://doi.org/10.1175/BAMS-D-15-00200.1

Hohenegger, C. & Schar, C. (2007) Atmospheric Predictability at Synoptic Versus Cloud-Resolving Scales. *Bulletin of the American Meteorological Society*, 88(11), 1783–1794. Available from: https://doi.org/10.1175/BAMS-88-11-1783

Hotta, D., Chen, T.-C., Kalnay, E., Ota, Y. & Miyoshi, T. (2017a) Proactive QC: A Fully Flow-Dependent Quality Control Scheme Based on EFSO. *Monthly Weather Review*, 145(8), 3331–3354. Available from: https://doi.org/10.1175/MWR-D-16-0290.1

Hotta, D., Kadowaki, T., Yonehara, H. & Ishibashi, T. (2023) Twin-analysis verification: A new verification approach to alleviate pitfalls of own-analysis verification. *Quarterly Journal of the Royal Meteorological Society*, 149(752), 924–932. Available from: https://doi.org/10.1002/qj.4441

Hotta, D., Kalnay, E., Ota, Y. & Miyoshi, T. (2017b) EFSR: Ensemble Forecast Sensitivity to Observation Error Covariance. *Monthly Weather Review*, 145(12), 5015–5031. Available from: https://doi.org/10.1175/MWR-D-17-0122.1

Hotta, D. & Ota, Y. (2021) Why does EnKF suffer from analysis overconfidence? An insight into exploiting the ever-increasing volume of observations. *Quarterly Journal of the Royal Meteorological Society*, 147(735), 1258–1277. Available from: https://doi.org/10.1002/qj.3970

Houtekamer, P.L. & Mitchell, H.L. (2001) A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation. *Monthly Weather Review*, 129(1), 123–137. Available from: https://doi.org/10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2

Hu, G. & Dance, S.L. (2024) Sampling and misspecification errors in the estimation of observation-error covariance matrices using observation-minus-background and observation-minus-analysis statistics. *Quarterly Journal of the Royal Meteorological Society*, 150(762), 3052–3077. Available from: https://doi.org/10.1002/qj.4750

Hu, G., Dance, S.L., Bannister, R.N., Chipilski, H.G., Guillet, O., Macpherson, B. et al. (2022) Progress, challenges, and future steps in data assimilation for convection-permitting numerical weather prediction: Report on the virtual meeting held on 10 and 12 November 2021. *Atmospheric Science Letters*, 24(1), e1130. Available from: https://doi.org/10.1002/asl.1130

Hu, G. & Franzke, C.L.E. (2020) Evaluation of Daily Precipitation Extremes in Reanalysis and Gridded Observation-Based Data Sets Over Germany. *Geophysical Research Letters*, 47(18), e2020GL089624. Available from: https://doi.org/10.1029/2020GL089624

Huang, Y., Wang, X., Mahre, A., Yu, T.-Y. & Bodine, D. (2022) Impacts of Assimilating Future Clear-Air Radial Velocity Observations from Phased Array Radar on Convection Initiation Forecasts: An Observing System Simulation Experiment Study. *Monthly Weather Review*, 150(7), 1563–1583. Available from: https://doi.org/10.1175/MWR-D-21-0199.1

Hunt, B.R., Kostelich, E.J. & Szunyogh, I. (2007) Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena*, 230(1),

112–126. Available from: https://doi.org/10.1016/j.physd.2006.11.008

Hutchinson, M.F. (1990) A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines. *Communications in Statistics - Simulation and Computation*, 19(2), 433–450. Available from: https://doi.org/10.1080/03610919008812866

Isaksen, L., Bonavita, M., Buizza, R., Fisher, M., Haseler, J., Leutbecher, M. et al. (2010) *Ensemble of data assimilations at ECMWF*. Technical Report 636. Reading, United Kingdom: ECMWF. Available from: https://www.ecmwf.int/node/10125

Janiskova, M. & Cardinali, C. (2016) *On the impact of the diabatic component in the forecast sensitivity observation impact diagnostics*. Technical Report 786. Reading, United Kingdom: ECMWF. Available from: https://www.ecmwf.int/node/16716

Janjić, T., Bormann, N., Bocquet, M., Carton, J.A., Cohn, S.E., Dance, S.L. et al. (2018) On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 144(713), 1257–1278. Available from: https://doi.org/10.1002/qj.3130

Janjić, T., Nerger, L., Albertella, A., Schröter, J. & Skachko, S. (2011) On Domain Localization in Ensemble-Based Kalman Filter Algorithms. *Monthly Weather Review*, 139(7), 2046–2060. Available from: https://doi.org/10.1175/2011MWR3552.1

Johnson, A. & Wang, X. (2012) Verification and Calibration of Neighborhood and Object-Based Probabilistic Precipitation Forecasts from a Multimodel Convection-Allowing Ensemble. *Monthly Weather Review*, 140(9), 3054–3077. Available from: https://doi.org/10.1175/MWR-D-11-00356.1

Johnson, A., Wang, X. & Jones, T. (2022) Impacts of Assimilating GOES-16 ABI Channels 9 and 10 Clear Air and Cloudy Radiance Observations With Additive Inflation and Adaptive Observation Error in GSI-EnKF for a Case of Rapidly Evolving Severe Supercells. *Journal of Geophysical Research: Atmospheres*, 127(11), e2021JD036157. Available from: https://doi.org/10.1029/2021JD036157

Johnson, A., Wang, X., Kong, F. & Xue, M. (2013) Object-Based Evaluation of the Impact of Horizontal Grid Spacing on Convection-Allowing Forecasts. *Monthly Weather Review*, 141(10), 3413–3425. Available from: https://doi.org/10.1175/MWR-D-13-00027.1

Johnson, B.T., Dang, C., Stegmann, P., Liu, Q., Moradi, I. & Auligne, T. (2023) The Community Radiative Transfer Model (CRTM): Community-Focused Collaborative Model Development Accelerating Research to Operations. *Bulletin of the American Meteorological Society*, 104(10), E1817–E1830. Available from: https://doi.org/10.1175/BAMS-D-22-0015.1

Kalnay, E., Ota, Y., Miyoshi, T. & Liu, J. (2012) A simpler formulation of forecast sensitivity to observations: application to ensemble Kalman filters. *Tellus A: Dynamic Meteorology and Oceanography*, 64(1), 18462. Available from: https://doi.org/10.3402/tellusa.v64i0.18462

Kelly, G., Bauer, P., Geer, A.J., Lopez, P. & Thepaut, J.-N. (2008) Impact of SSM/I observations related to moisture, clouds, and precipitation on global NWP forecast skill. *Monthly Weather Review*, 136(7), 2713–2726.

Kostka, P.M., Weissmann, M., Buras, R., Mayer, B. & Stiller, O. (2014) Observation Operator for Visible and Near-Infrared Satellite Reflectances. *Journal of Atmospheric and Oceanic Technology*, 31(6), 1216–1233. Available from: https://doi.org/10.1175/JTECH-D-13-00116.1

Kotsuki, S., Kurosawa, K. & Miyoshi, T. (2019) On the properties of ensemble forecast sensitivity to observations. *Quarterly Journal of the Royal Meteorological Society*, 145(722), 1897–1914. Available from: https://doi.org/10.1002/qj.3534

Kugler, L., Anderson, J.L. & Weissmann, M. (2023) Potential impact of all-sky assimilation of visible and infrared satellite observations compared with radar reflectivity for convective-scale numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 149(757), 3623–3644. Available from: https://doi.org/10.1002/qj.4577

Kugler, L. & Weissmann, M. (2024a) The synergy of assimilating visible and infrared radiances and radar observations. https://doi.org/10.22541/au.170870251.15344379/v1

Kugler, L. & Weissmann, M. (2024b) The effect of nonlinear observation operators for visible and infrared radiances in ensemble data assimilation. https://doi.org/10.22541/au.172244493.31683657/v1

Kunii, M., Miyoshi, T. & Kalnay, E. (2012) Estimating the Impact of Real Observations in Regional Numerical Weather Prediction Using an Ensemble Kalman Filter. *Monthly Weather Review*, 140(6), 1975–1987. Available from: https://doi.org/10.1175/MWR-D-11-00205.1

Lang, S., Hólm, E., Bonavita, M. & Tremolet, Y. (2019) A 50-member ensemble of data assimilations. https://www.ecmwf.int/node/18883

Langland, R.H. & Baker, N.L. (2004) Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. *Tellus A: Dynamic Meteorology and Oceanography*, 56(3), 189–201. Available from: https://doi.org/10.3402/tellusa.v56i3.14413

Lawrence, H., Bormann, N., Sandu, I., Day, J., Farnan, J. & Bauer, P. (2019) Use and impact of Arctic observations in the ECMWF Numerical Weather Prediction system. *Quarterly Journal of the Royal Meteorological Society*, 145(725), 3432–3454. Available from: https://doi.org/10.1002/qj.3628

Lean, K., Bormann, N., Healy, S. & English, S. (2022) *Final Report: Study to assess earth observation with small satellites and their prospects for future global numerical weather prediction*. ESA Contract Report. Reading, United Kingdom: ECMWF. Available from: https://doi.org/10.21957/kp7z1sn1n

Leuenberger, D., Haefele, A., Omanovic, N., Fengler, M., Martucci, G., Calpini, B. et al. (2020) Improving High-Impact Numerical Weather Prediction with Lidar and Drone Observations. *Bulletin of the American Meteorological Society*, 101(7), E1036–E1051. Available from: https://doi.org/10.1175/BAMS-D-19-0119.1

Li, H., Liu, J. & Kalnay, E. (2010) Correction of 'Estimating observation impact without adjoint model in an ensemble Kalman filter'. *Quarterly Journal of the Royal Meteorological Society*, 136(651), 1652–1654. Available from: https://doi.org/10.1002/qj.658

Liu, J. & Kalnay, E. (2008) Estimating observation impact without adjoint model in an ensemble Kalman filter. *Quarterly Journal of the Royal Meteorological Society*, 134(634), 1327–1335. Available from: https://doi.org/10.1002/qj.280

Liu, J., Kalnay, E., Miyoshi, T. & Cardinali, C. (2009) Analysis sensitivity calculation in an ensemble Kalman filter. *Quarterly Journal of the Royal Meteorological Society*, 135(644), 1842–1851. Available from: https://doi.org/10.1002/qj.511

Livings, D.M., Dance, S.L. & Nichols, N.K. (2008) Unbiased ensemble square root filters. *Physica D: Nonlinear Phenomena*, 237(8),

1021–1028. Available from: https://doi.org/10.1016/j.physd.2008.01.005

Lledó, L., Haiden, T., Schroettle, J. & Forbes, R. (2023) Scale-dependent verification of precipitation and cloudiness at ECMWF. *ECMWF Newsletter*, 174, 18–22. Available from: https://doi.org/10.21957/c92loli749

Lorenc, A.C. & Marriott, R.T. (2014) Forecast sensitivity to observations in the Met Office Global numerical weather prediction system. *Quarterly Journal of the Royal Meteorological Society*, 140(678), 209–224. Available from: https://doi.org/10.1002/qj.2122

Lupu, C., Gauthier, P. & Laroche, S. (2011) Evaluation of the Impact of Observations on Analyses in 3D- and 4D-Var Based on Information Content. *Monthly Weather Review*, 139(3), 726–737. Available from: https://doi.org/10.1175/2010MWR3404.1

Maejima, Y., Kawabata, T., Seko, H. & Miyoshi, T. (2022) Observing System Simulation Experiments of a Rich Phased Array Weather Radar Network Covering Kyushu for the July 2020 Heavy Rainfall Event. *SOLA*, 18, 25–32. Available from: https://doi.org/10.2151/sola.2022-005

Marseille, G.-J., Stoffelen, A. & Barkmeijer, J. (2008) Sensitivity Observing System Experiment (SOSE)—a new effective NWP-based tool in designing the global observing system. *Tellus A*, 60(2), 216–233. Available from: https://doi.org/10.1111/j.1600-0870.2007.00288.x

Michel, Y. (2018) Revisiting Fisher's approach to the handling of horizontal spatial correlations of observation errors in a variational framework. *Quarterly Journal of the Royal Meteorological Society*, 144(716), 2011–2025. Available from: https://doi.org/10.1002/qj.3249

Milan, M., Clayton, A., Lorenc, A., Macpherson, B., Tubbs, R. & Dow, G. (2023) Large-scale blending in an hoy 4D-Var framework for a numerical weather prediction model. *Quarterly Journal of the Royal Meteorological Society*, 149(755), 2067–2090. Available from: https://doi.org/10.1002/qj.4495

Milan, M., Macpherson, B., Tubbs, R., Dow, G., Inverarity, G., Mittermaier, M. et al. (2020) Hoy 4D-Var in the Met Office UKV operational forecast model. *Quarterly Journal of the Royal Meteorological Society*, 146(728), 1281–1301.

Miyamoto, Y., Kajikawa, Y., Yoshida, R., Yamaura, T., Yashiro, H. & Tomita, H. (2013) Deep moist atmospheric convection in a subkilometer global simulation. *Geophysical Research Letters*, 40(18), 4922–4926.

Necker, T., Weissmann, M., Ruckstuhl, Y., Anderson, J. & Miyoshi, T. (2020) Sampling Error Correction Evaluated Using a Convective-Scale 1000-Member Ensemble. *Monthly Weather Review*, 148(3), 1229–1249. Available from: https://doi.org/10.1175/MWR-D-19-0154.1

Necker, T., Weissmann, M. & Sommer, M. (2018) The importance of appropriate verification metrics for the assessment of observation impact in a convection-permitting modelling system. *Quarterly Journal of the Royal Meteorological Society*, 144(714), 1667–1680. Available from: https://doi.org/10.1002/qj.3390

Necker, T., Wolfgruber, L., Kugler, L., Weissmann, M., Dorninger, M. & Serafin, S. (2024) The fractions skill score for ensemble forecast verification. *Quarterly Journal of the Royal Meteorological Society*, 150(764), 4457–4477. Available from: https://doi.org/10.1002/qj.4824

Nichols, N.K. (2010) Mathematical Concepts of Data Assimilation. In: Lahoz, W., Khattatov, B. & Menard, R. (Eds.) *Data Assimilation: Making Sense of Observations*. Berlin Heidelberg: Springer, pp. 13–39. Available from: https://doi.org/10.1007/978-3-540-74703-1.2

Nomokonova, T., Griewank, P.J., Löhnert, U., Miyoshi, T., Necker, T. & Weissmann, M. (2022) Estimating the benefit of Doppler wind lidars for short-term low-level wind ensemble forecasts. *Quarterly Journal of the Royal Meteorological Society*, 149(750), 192–210. Available from: https://doi.org/10.1002/qj.4402

Oke, P.R., Sakov, P. & Corney, S.P. (2007) Impacts of localisation in the EnKF and EnOI: experiments with a small model. *Ocean Dynamics*, 57(1), 32–45. Available from: https://doi.org/10.1007/s10236-006-0088-8

Ota, Y., Derber, J.C., Kalnay, E. & Miyoshi, T. (2013) Ensemble-based observation impact estimates using the NCEP GFS. *Tellus A: Dynamic Meteorology and Oceanography*, 65(1), 20038. Available from: https://doi.org/10.3402/tellusa.v65i0.20038

Petrie, R.E. & Dance, S.L. (2010) Ensemble-based data assimilation and the localisation problem. *Weather*, 65(3), 65–69. Available from: https://doi.org/10.1002/wea.505

Privé, N., Errico, R.M., Todling, R. & El Akkraoui, A. (2021) Evaluation of adjoint-based observation impacts as a function of forecast length using an Observing System Simulation Experiment. *Quarterly Journal of the Royal Meteorological Society*, 147(734), 121–138. Available from: https://doi.org/10.1002/qj.3909

Privé, N.C. & Errico, R.M. (2013) The role of model and initial condition error in numerical weather forecasting investigated with an observing system simulation experiment. *Tellus A: Dynamic Meteorology and Oceanography*, 65(1), 21740. Available from: https://doi.org/10.3402/tellusa.v65i0.21740

Privé, N.C., Errico, R.M. & Akkraoui, A.E. (2022) Investigation of the Potential Saturation of Information from Global Navigation Satellite System Radio Occultation Observations with an Observing System Simulation Experiment. *Monthly Weather Review*, 150(6), 1293–1316. Available from: https://doi.org/10.1175/MWR-D-21-0230.1

Privé, N.C., Errico, R.M. & McCarty, W. (2021) The importance of simulated errors in observing system simulation experiments. *Tellus A: Dynamic Meteorology and Oceanography*, 73(1), 1–17. Available from: https://doi.org/10.1080/16000870.2021.1886795

Rabier, F., Klinker, E., Courtier, P. & Hollingsworth, A. (1996) Sensitivity of forecast errors to initial conditions. *Quarterly Journal of the Royal Meteorological Society*, 122(529), 121–150. Available from: https://doi.org/10.1002/qj.49712252906

Randriamampianina, R. (2005) Radiance-bias correction for a limited area model. *Idojaras*, 109, 143–155 https://api.semanticscholar.org/CorpusID:123990784

Randriamampianina, R., Bormann, N., Køltzow, M.A.Ø., Lawrence, H., Sandu, I. & Wang, Z.Q. (2021) Relative impact of observations on a regional Arctic numerical weather prediction system. *Quarterly Journal of the Royal Meteorological Society*, 147(737), 2212–2232. Available from: https://doi.org/10.1002/qj.4018

Ren, S., Lei, L., Tan, Z.-M. & Zhang, Y. (2019) Multivariate Ensemble Sensitivity Analysis for Super Typhoon Haiyan (2013). *Monthly Weather Review*, 147(9), 3467–3480. Available from: https://doi.org/10.1175/MWR-D-19-0074.1

Richardson, D., Raoult, B., Fuentes, M., Mladek, R. & Buizza, R. (2019) TIGGE-LAM. https://confluence.ecmwf.int/display/TIGL

Roberts, N.M. & Lean, H.W. (2008) Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of

Convective Events. *Monthly Weather Review*, 136(1), 78–97. Available from: https://doi.org/10.1175/2007MWR2123.1

Rodgers, C.D. (1998) Information content and optimisation of high spectral resolution remote measurements. *Advances in Space Research*, 21(3), 361–367. Available from: https://doi.org/10.1016/S0273-1177(97)00915-0

Rodgers, C.D. (2000) *Inverse Methods for Atmospheric Sounding*. Singapore: World Scientific. Available from: https://www.worldscientific.com/worldscibooks/10.1142/3171#t=aboutBook

Rosmond, T.E. (1997) *A technical description of the NRL adjoint modeling system*. Technical report. Monterey, California: Naval Research Lab Monterey CA Marine Meteorology Division.

Scheck, L., Frèrebeau, P., Buras-Schnell, R. & Mayer, B. (2016) A fast radiative transfer method for the simulation of visible satellite imagery. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 175, 54–67. Available from: https://doi.org/10.1016/j.jqsrt.2016.02.008

Scheck, L., Weissmann, M. & Bach, L. (2020) Assimilating visible satellite images for convective-scale numerical weather prediction: A case-study. *Quarterly Journal of the Royal Meteorological Society*, 146(732), 3165–3186. Available from: https://doi.org/10.1002/qj.3840

Scheck, L., Weissmann, M. & Mayer, B. (2018) Efficient Methods to Account for Cloud-Top Inclination and Cloud Overlap in Synthetic Visible Satellite Images. *Journal of Atmospheric and Oceanic Technology*, 35(3), 665–685. Available from: https://doi.org/10.1175/JTECH-D-17-0057.1

Schraff, C., Reich, H., Rhodin, A., Schomburg, A., Stephan, K., Periáñez, A. et al. (2016) Kilometre-scale ensemble data assimilation for the COSMO model (KENDA). *Quarterly Journal of the Royal Meteorological Society*, 142(696), 1453–1472. Available from: https://doi.org/10.1002/qj.2748

Schröttle, J., Weissmann, M., Scheck, L. & Hutt, A. (2020) Assimilating Visible and Infrared Radiances in Idealized Simulations of Deep Convection. *Monthly Weather Review*, 148(11), 4357–4375. Available from: https://doi.org/10.1175/MWR-D-20-0002.1

Seity, Y., Brousseau, P., Malardel, S., Hello, G., Bénard, P., Bouttier, F. et al. (2011) The AROME-France Convective-Scale Operational Model. *Monthly Weather Review*, 139(3), 976–991. Available from: https://doi.org/10.1175/2010MWR3425.1

Shannon, C.E. & Weaver, W. (1964) *The Mathematical Theory of Communication*. Urbana: University of Illinois Press.

Simonin, D., Waller, J.A., Ballard, S.P., Dance, S.L. & Nichols, N.K. (2019) A pragmatic strategy for implementing spatially correlated observation errors in an operational system: An application to Doppler radial winds. *Quarterly Journal of the Royal Meteorological Society*, 145(723), 2772–2790. Available from: https://doi.org/10.1002/qj.3592

Sommer, M. & Weissmann, M. (2014) Observation impact in a convective-scale localized ensemble transform Kalman filter. *Quarterly Journal of the Royal Meteorological Society*, 140(685), 2672–2679. Available from: https://doi.org/10.1002/qj.2343

Sommer, M. & Weissmann, M. (2016) Ensemble-based approximation of observation impact using an observation-based verification metric. *Tellus A: Dynamic Meteorology and Oceanography*, 68(1), 27885. Available from: https://doi.org/10.3402/tellusa.v68.27885

Stewart, L.M., Dance, S.L. & Nichols, N.K. (2008) Correlated observation errors in data assimilation. *International Journal for Numerical Methods in Fluids*, 56(8), 1521–1527. Available from: https://doi.org/10.1002/fld.1636

Stewart, L.M., Dance, S.L. & Nichols, N.K. (2013) Data assimilation with correlated observation errors: experiments with a 1-D shallow water model. *Tellus A: Dynamic Meteorology and Oceanography*, 65(1), 19546.

Stewart, L.M., Dance, S.L., Nichols, N.K., Eyre, J.R. & Cameron, J. (2014) Estimating interchannel observation-error correlations for IASI radiance data in the Met Office system. *Quarterly Journal of the Royal Meteorological Society*, 140(681), 1236–1244. Available from: https://doi.org/10.1002/qj.2211

Stiller, O. (2022) New impact diagnostics for cross-validation of different observation types. *Quarterly Journal of the Royal Meteorological Society*, 148(747), 2853–2876. Available from: https://doi.org/10.1002/qj.4339

Tan, D.G.H., Andersson, E., Fisher, M. & Isaksen, L. (2007) Observing-system impact assessment using a data assimilation ensemble technique: application to the ADM–Aeolus wind profiling mission. *Quarterly Journal of the Royal Meteorological Society*, 133(623), 381–390. Available from: https://doi.org/10.1002/qj.43

Tardif, R., Hakim, G.J., Bumbaco, K.A., Lazzara, M.A., Manning, K.W., Mikolajczyk, D.E. et al. (2022) Assessing observation network design predictions for monitoring Antarctic surface temperature. *Quarterly Journal of the Royal Meteorological Society*, 148(743), 727–746. Available from: https://doi.org/10.1002/qj.4226

Teixeira, J., Piepmeier, J.R., Nehrir, A.R., Ao, C.O., Chen, S.S., Clayson, C.A. et al. (2021) *Toward a global planetary boundary layer observing system: The NASA PBL incubation study team report*. NASA PBL Incubation Study Team. Available from: https://ntrs.nasa.gov/api/citations/20230001633/downloads/AFridlindPBLTowardsReport.pdf

Tikhonov, A.N. (1965) Ill-conditioned problems in linear algebra and their robust method solution. *Doklady Akademii Nauk SSSR*, 163, 591–594.

Todling, R. (2013) Comparing two approaches for assessing observation impact. *Monthly Weather Review*, 141(5), 1484–1505. Available from: https://doi.org/10.1175/MWR-D-12-00100.1

Torn, R.D. (2014) The Impact of Targeted Dropwindsonde Observations on Tropical Cyclone Intensity Forecasts of Four Weak Systems during PREDICT. *Monthly Weather Review*, 142(8), 2860–2878. Available from: https://doi.org/10.1175/MWR-D-13-00284.1

Torn, R.D. & Hakim, G.J. (2008) Ensemble-Based Sensitivity Analysis. *Monthly Weather Review*, 136(2), 663–677. Available from: https://doi.org/10.1175/2007MWR2132.1

Vetra-Carvalho, S., Dixon, M., Migliorini, S., Nichols, N.K. & Ballard, S.P. (2012) Breakdown of hydrostatic balance at convective scales in the forecast errors in the Met Office Unified Model. *Quarterly Journal of the Royal Meteorological Society*, 138(668), 1709–1720. Available from: https://doi.org/10.1002/qj.1925

Vural, J., Merker, C., Löffler, M., Leuenberger, D., Schraff, C., Stiller, O. et al. (2024) Improving the representation of the atmospheric boundary layer by direct assimilation of ground-based microwave radiometer observations. *Quarterly Journal of the Royal Meteorological Society*, 150(759), 1012–1028. Available from: https://doi.org/10.1002/qj.4634

Wahba, G., Johnson, D.R., Gao, F. & Gong, J. (1995) Adaptive Tuning of Numerical Weather Prediction Models: Randomized GCV

in Three- and Four-Dimensional Data Assimilation. *Monthly Weather Review*, 123, 3358–3370. Available from: https://journals.ametsoc.org/view/journals/mwre/123/11/1520-0493_1995_123_3358_atonwp_2_0_co_2.xml

Waller, J.A., Ballard, S.P., Dance, S.L., Kelly, G., Nichols, N.K. & Simonin, D. (2016a) Diagnosing horizontal and inter-channel observation error correlations for SEVIRI observations using observation-minus-background and observation-minus-analysis statistics. *Remote Sensing*, 8(7), 581. Available from: https://doi.org/10.3390/rs8070581

Waller, J.A., Bauernschubert, E., Dance, S.L., Nichols, N.K., Potthast, R. & Simonin, D. (2019) Observation Error Statistics for Doppler Radar Radial Wind Superobservations Assimilated into the DWD COSMO-KENDA System. *Monthly Weather Review*, 147(9), 3351–3364. Available from: https://doi.org/10.1175/MWR-D-19-0104.1

Waller, J.A., Dance, S.L. & Lean, H.W. (2021) Evaluating errors due to unresolved scales in convection-permitting numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 147(738), 2657–2669. Available from: https://doi.org/10.1002/qj.4043

Waller, J.A., Simonin, D., Dance, S., Nichols, N. & Ballard, S. (2016b) Diagnosing observation error correlations for Doppler radar radial winds in the Met Office UKV model using observation-minus-background and observation-minus-analysis statistics. *Monthly Weather Review*, 144(10), 3533–3551. Available from: https://doi.org/10.1175/MWR-D-15-0340.1

Weston, P.P., Bell, W. & Eyre, J.R. (2014) Accounting for correlated error in the assimilation of high-resolution sounder data. *Quarterly Journal of the Royal Meteorological Society*, 140(685), 2420–2429. Available from: https://doi.org/10.1002/qj.2306

Whitaker, J.S. & Hamill, T.M. (2002) Ensemble Data Assimilation without Perturbed Observations. *Monthly Weather Review*, 130, 1913–1924. Available from: https://doi.org/10.1175/1520-0493(2002)130

Whitaker, J.S. & Hamill, T.M. (2012) Evaluating Methods to Account for System Errors in Ensemble Data Assimilation. *Monthly Weather Review*, 140(9), 3078–3089. Available from: https://doi.org/10.1175/MWR-D-11-00276.1

World Meteorological Organization (WMO). (2023) Observing Systems Capability Analysis and Review (OSCAR). https://space.oscar.wmo.int/

Xu, Q. (2007) Measuring information content from observations for data assimilation: relative entropy versus shannon entropy difference. *Tellus A: Dynamic Meteorology and Oceanography*, 59(2), 198–209. Available from: https://doi.org/10.1111/j.1600-0870.2006.00222.x

Xu, Q., Wei, L. & Healy, S. (2009) Measuring information content from observations for data assimilations: connection between different measures and application to radar scan design. *Tellus A*, 61(1), 144–153. Available from: https://doi.org/10.1111/j.1600-0870.2008.00373.x

Yeh, H.-L., Yang, S.-C., Terasaki, K., Miyoshi, T. & Liou, Y.-C. (2022) Including observation error correlation for ensemble radar radial wind assimilation and its impact on heavy rainfall prediction. *Quarterly Journal of the Royal Meteorological Society*, 148(746), 2254–2281. Available from: https://doi.org/10.1002/qj.4302

Yu, L., Fennel, K., Wang, B., Laurent, A., Thompson, K.R. & Shay, L.K. (2019) Evaluation of nonidentical versus identical twin approaches for observation impact assessments: an ensemble-Kalman-filter-based ocean assimilation application for the Gulf of Mexico. *Ocean Science*, 15(6), 1801–1814. Available from: https://doi.org/10.5194/os-15-1801-2019

Zeng, Y., Janjic, T., Feng, Y., Blahak, U., de Lozar, A., Bauernschubert, E. et al. (2021) Interpreting estimated Observation Error Statistics of Weather Radar Measurements using the ICON-LAM-KENDA System. *Atmospheric Measurement Techniques Discussions*, 2021, 1–28. Available from: https://doi.org/10.5194/amt-2021-95

Zhang, F., Snyder, C. & Sun, J. (2004) Impacts of Initial Estimate and Observation Availability on Convective-Scale Data Assimilation with an Ensemble Kalman Filter. *Monthly Weather Review*, 132(5), 1238–1253. Available from: https://doi.org/10.1175/1520-0493(2004)132<1238:IOIEAO>2.0.CO;2

How to cite this article: Hu, G., Dance, S.L., Fowler, A., Simonin, D., Waller, J., Auligne, T. *et al.* (2025) On methods for assessment of the value of observations in convection-permitting data assimilation and numerical weather forecasting. *Quarterly Journal of the Royal Meteorological Society*, 1–36. Available from: https://doi.org/10.1002/qj.4933

APPENDIX A. ENSEMBLE PERTURBATION AND COVARIANCE MATRICES

This Appendix introduces the ensemble perturbation matrix, the ensemble perturbation matrix in observation space, and the ensemble covariance matrix. These matrices are used throughout this work.

The ensemble perturbation matrix is the $n \times N$ matrix defined by

$$\mathbf{X} = \left[\mathbf{x}_1 - \bar{\mathbf{x}} \quad \mathbf{x}_2 - \bar{\mathbf{x}} \quad \dots \quad \mathbf{x}_N - \bar{\mathbf{x}}\right], \tag{A1}$$

where $\mathbf{x}_i \in \mathbb{R}^n$ is the model state vector for the ith ensemble member and

$$\bar{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i$$

is the ensemble mean of $\mathbf{x}_i$. The ensemble perturbation matrix can be calculated from the analysis ensemble ($\mathbf{X}_\mathrm{a}$), background ensemble ($\mathbf{X}_\mathrm{b}$), and forecast ensemble generated from the analysis ensemble ($\mathbf{X}_\mathrm{a}^\mathrm{f}$).

The ensemble perturbation matrix in observation space is the $m \times N$ defined by

$$\mathbf{Y} = \left[H(\mathbf{x}_1) - \overline{H(\mathbf{x})} \quad H(\mathbf{x}_2) - \overline{H(\mathbf{x})} \quad \dots \quad H(\mathbf{x}_N) - \overline{H(\mathbf{x})}\right], \tag{A2}$$

where $H(\mathbf{x}_i) \in \mathbb{R}^m$ is the model state vector in observation space for the ith ensemble member and

$$\overline{H(\mathbf{x})} = \frac{1}{N}\sum_{i=1}^{N} H(\mathbf{x}_i)$$

is the ensemble mean of $H(\mathbf{x}_i)$ (e.g., Livings *et al.*, 2008). If the observation operator is linear, then $\mathbf{Y} = \mathbf{HX}$. The matrix $\mathbf{Y}$ can be calculated from the analysis ensemble ($\mathbf{Y}_a$), background ensemble ($\mathbf{Y}_b$), and forecast ensemble generated from the analysis ensemble ($\mathbf{Y}_a^f$).

The last matrix to introduce is the ensemble covariance matrix, which is the $n \times n$ matrix defined by

$$\mathbf{P} = \frac{1}{N-1}\mathbf{X}\mathbf{X}^\top. \tag{A3}$$

It provides an estimate of the background-error covariance matrix ($\mathbf{P}_b$) and the analysis-error covariance matrix ($\mathbf{P}_a$).

APPENDIX B. THE RANDOMIZED APPROACH

This approach has been used to estimate the VR and DFS in variational DA systems (Desroziers *et al.*, 2005b; Desroziers & Ivanov, 2001; Fisher, 2003; Wahba *et al.*, 1995) and the EDA system (Desroziers *et al.*, 2009), but should also be applicable to stochastic ensemble Kalman filters. It exploits a randomized estimation of the matrix trace (Girard, 1989; Hutchinson, 1990). Let $\mathbf{G}$ be a square matrix, then an unbiased estimate of tr($\mathbf{G}$) is

$$T_{\mathbf{G}}(\boldsymbol{\eta}) = \boldsymbol{\eta}^\top \mathbf{G}\boldsymbol{\eta}, \tag{B1}$$

where $\boldsymbol{\eta}$ is a random vector with elements satisfying $\mathbb{E}[\eta_i] = 0$, $\mathbb{E}\left[\eta_i^2\right] = 1$ and $\mathbb{E}\left[\eta_i\eta_j\right] = 0$ for $i \neq j$ (Girard, 1989). The statistical expectation of $T_{\mathbf{G}}(\boldsymbol{\eta})$ is tr($\mathbf{G}$). If the elements of $\boldsymbol{\eta}$ are independent random values from the standard Gaussian distribution, that is, $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, then the variance of $T_{\mathbf{G}}(\boldsymbol{\eta})$ is $2\sum_i \sum_j G_{ij}^2$, where G_{ij} denotes the (i, j)th element of the matrix $\mathbf{G}$ (Girard, 1989).

Using the cyclic property of the matrix trace (section 2.2 of Bernstein, 2009), we may rewrite Equation (1) as

$$\delta\sigma^2 = \mathrm{tr}\left(\mathbf{R}^{1/2}\mathbf{R}^{-1}\mathbf{HBKR}^{1/2}\right).$$

Then, using Equation (B1), the VR is estimated by

$$\delta\sigma^2 = \mathbb{E}\left[\delta\mathbf{y}^\top\mathbf{R}^{-1}\mathbf{HB}\delta\mathbf{x}_a\right],$$

where

$$\delta\mathbf{y} = \mathbf{R}^{1/2}\boldsymbol{\eta}$$

is a vector of observation perturbations and

$$\delta\mathbf{x}_a = \mathbf{K}\delta\mathbf{y}$$

is the difference between the analysis increments obtained with perturbed and unperturbed observations. Similarly, the DFS can be estimated by

$$\mathrm{DFS} = \mathbb{E}\left[\delta\mathbf{y}^\top\mathbf{R}^{-1}\mathbf{H}\delta\mathbf{x}_a\right],$$

where the matrix $\mathbf{B}$ disappears.

The randomized approach requires the observation and analysis perturbations ($\delta\mathbf{y}$ and $\delta\mathbf{x}_a$), which are readily available in the EDA system and stochastic ensemble Kalman filters. A key is that the analysis perturbations should be obtained using the same Kalman gain.

It is also possible to perturb the background state vector rather than the observation vector (Desroziers & Ivanov, 2001; Fisher, 2003). In addition, assuming that $\partial[H(\mathbf{x}_a)]/\partial\mathbf{y}$ varies smoothly with respect to observations, the randomized approach can be used for nonlinear observation operators (Chapnik *et al.*, 2006).

APPENDIX C. DIFFERENT FORMULATIONS FOR THE DFS

In this Appendix, we present several different expressions for the DFS. They are all equivalent to Equation (3). The DFS can be formulated as

$$\mathrm{DFS} = p - \sum_{i=1}^{p} \lambda_i\left(\mathbf{UAU}^\top\right), \tag{C1}$$

where $\mathbf{U}$ is a transform matrix that satisfies $\mathbf{UU}^\top = \mathbf{B}^{-1}$ (e.g., $\mathbf{U} = \mathbf{B}^{-1/2}$: Courtier *et al.*, 1998), $\mathbf{UAU}^\top$ the analysis-error covariance matrix of transformed variable $\mathbf{Ux}_a$, and $\lambda_i(\cdot)$ the ith eigenvalue of a matrix (Fisher, 2003; Rodgers, Rodgers, 1998). In Equation (C1), p represents the total number of statistically independent directions, and the sum of eigenvalues estimates the number of directions that are not constrained by the observations (Fisher, 2003). Since the trace of a matrix is equal to the sum of its eigenvalues, Equation (C1) becomes DFS = p − tr$\left(\mathbf{UAU}^\top\right)$. Further, using $\mathbf{A} = \mathbf{B} - \mathbf{KHB}$ (e.g., Bouttier & Courtier, 2002) and the properties of the trace of a sum and product of matrices (section 2.2 of Bernstein, 2009), we obtain Equation (3).

The DFS can also be written as

$$\mathrm{DFS} = \mathbb{E}\left[(\mathbf{x}_a - \mathbf{x}_b)^\top\mathbf{B}^{-1}(\mathbf{x}_a - \mathbf{x}_b)\right], \tag{C2}$$

which is related to the background penalty term of the cost function of DA (section 2.4 of Rodgers, 2000). This

equation has been used to estimate the DFS for global positioning system radio occultation (Healy & Thépaut, 2006). The derivation from Equation (C2) to Equation (3) can be found in appendix A of Fowler *et al.* (2020).

Another expression for the DFS is

$$\mathrm{DFS} = \mathrm{tr}\left(\frac{\partial(\mathbf{H}\mathbf{x}_\mathrm{a})}{\partial \mathbf{y}}\right) = \sum_{i=1}^{m} \frac{\partial([\mathbf{H}\mathbf{x}_\mathrm{a}]_i)}{\partial y_i}, \tag{C3}$$

which is the trace of the derivative of the analysis in observation space with respect to observations (Cardinali *et al.*, 2004; Chapnik *et al.*, 2006; Liu *et al.*, 2009). In other terms, the DFS can be considered as the self-sensitivity of the analysis with respect to observations (i.e., the influence of the ith observation on the analysis at the ith location). Since we have $\mathbf{x}_\mathrm{a} = \mathbf{x}_\mathrm{b} + \delta\mathbf{x}$, where the analysis increment vector $\delta\mathbf{x}$ is given by Equation (5), we can easily prove the equivalence between Equations (C3) and (3).

APPENDIX D. THE ASSIMILATION RESIDUAL APPROACH

This approach has been used to estimate the DFS in variational DA (Fowler *et al.*, 2020; Lupu *et al.*, 2011), but it may also be applied to ensemble-based DA systems. It uses the innovation vector $\mathbf{d}$ (Equation 6), the residual vector $\mathbf{r}$ (Equation 21), and the observation-space increment vector,

$$\mathbf{v} = H(\mathbf{x}_\mathrm{a}) - H(\mathbf{x}_\mathrm{b}).$$

The theoretical DFS (Equation 3) is estimated by

$$\mathrm{DFS} = \mathrm{tr}\left(\mathbb{E}\left[\mathbf{R}^{-1/2}\mathbf{v}\left(\mathbf{R}^{-1/2}\mathbf{r}\right)^\top\right] \left(\mathbb{E}\left[\mathbf{R}^{-1/2}\mathbf{d}\left(\mathbf{R}^{-1/2}\mathbf{r}\right)^\top\right]\right)^{-1}\right),$$

and the actual DFS (Equation 15) is estimated by

$$\mathrm{DFS}_\mathrm{Actual} = \mathbb{E}\left[\left(\mathbf{R}^{-1/2}\mathbf{r}\right)^\top \mathbf{R}^{-1/2}\mathbf{v}\right],$$

where the matrix $\mathbf{R}^{-1/2}$ is used for normalization and the statistical expectation $\mathbb{E}[\cdot]$ is estimated by averaging over a sample of the vectors obtained from cycling the DA system (Fowler *et al.*, 2020).

This approach is derived by assuming a linear observation operator. To take into account the nonlinearity of the observation operator, see the supporting information for Fowler *et al.* (2020). In variational DA systems, the DFS can also be estimated by estimating the trace of a function of the Hessian matrix of the analysis cost function (Fisher, 2003) or using a low-rank approximation of the analysis-error covariance matrix (Cardinali *et al.*, 2004).

APPENDIX E. DERIVATION OF THE EFSOI

This Appendix shows the derivation of the EFSOI with analysis-based verification (Equation 10), and presents the EFSOI with observation-based verification.

E.1 Analysis-based verification

Equation (9) is an expression of the difference between two squares, which can be factorized as

$$\begin{aligned}\delta e &= \left[\left(\mathbf{x}_\mathrm{a}^\mathrm{f} - \mathbf{x}_\mathrm{t}\right) - \left(\mathbf{x}_\mathrm{b}^\mathrm{f} - \mathbf{x}_\mathrm{t}\right)\right]^\top \mathbf{C}\left[\left(\mathbf{x}_\mathrm{a}^\mathrm{f} - \mathbf{x}_\mathrm{t}\right) + \left(\mathbf{x}_\mathrm{b}^\mathrm{f} - \mathbf{x}_\mathrm{t}\right)\right] \\ &= \left(\delta\mathbf{x}^\mathrm{f}\right)^\top \mathbf{C}\left[\left(\mathbf{x}_\mathrm{a}^\mathrm{f} - \mathbf{x}_\mathrm{t}\right) + \left(\mathbf{x}_\mathrm{b}^\mathrm{f} - \mathbf{x}_\mathrm{t}\right)\right],\end{aligned} \tag{E1}$$

where $\delta\mathbf{x}^\mathrm{f} = \mathbf{x}_\mathrm{a}^\mathrm{f} - \mathbf{x}_\mathrm{b}^\mathrm{f}$ is the difference between two forecasts (Kalnay *et al.*, 2012). Using a tangent linear approximation, the forecast difference can be written as a linearized function of the analysis increments,

$$\delta\mathbf{x}^\mathrm{f} = \mathbf{M}\delta\mathbf{x}, \tag{E2}$$

where $\mathbf{M}$ is the TLM (or a perturbation forecast model). Sequentially substituting Equations (E2) and (5) into Equation (E1) gives

$$\delta e \approx \mathbf{d}^\top \mathbf{K}^\top \mathbf{M}^\top \mathbf{C}\left[\left(\mathbf{x}_\mathrm{a}^\mathrm{f} - \mathbf{x}_\mathrm{t}\right) + \left(\mathbf{x}_\mathrm{b}^\mathrm{f} - \mathbf{x}_\mathrm{t}\right)\right], \tag{E3}$$

which is an expression for the FSOI. (There is another expression that uses two different TLMs for background and analysis trajectories respectively: Errico, 2007; Langland & Baker, 2004). Further substituting Equation (7) into Equation (E3) and assuming

$$\mathbf{X}_\mathrm{a}^\mathrm{f} = \mathbf{M}\mathbf{X}_\mathrm{a}, \tag{E4}$$

we obtain

$$\delta e \approx \frac{1}{N-1}\bar{\mathbf{d}}^\top \mathbf{R}^{-1}\mathbf{Y}_\mathrm{a}\left(\mathbf{X}_\mathrm{a}^\mathrm{f}\right)^\top \mathbf{C}\left[\left(\bar{\mathbf{x}}_\mathrm{a}^\mathrm{f} - \mathbf{x}_\mathrm{t}\right) + \left(\bar{\mathbf{x}}_\mathrm{b}^\mathrm{f} - \mathbf{x}_\mathrm{t}\right)\right]$$

as the equation of the EFSOI without covariance localization. After applying localization, we finally obtain Equation (10) as requested.

The calculation of the EFSOI requires the ensemble perturbation matrices, which are not produced in variational DA systems. In such systems, the sensitivity should be estimated by iteratively minimizing a modified cost function (Lorenc & Marriott, 2014). In the 4DEnVar system, the sensitivity should be estimated in a similar

way, but making use of Equation (E4) (Buehner *et al.*, 2018).

E.2 Observation-based verification

In addition to using the analysis as the verification reference, we may use observations (Cardinali, 2018; Necker *et al.*, 2018; Sommer & Weissmann, 2016; Todling, 2013). In this case, Equation (9) becomes

$$\delta e_{\mathrm{y}} = \left(\mathbf{y}_{\mathrm{av}}^{\mathrm{f}} - \mathbf{y}_{\mathrm{v}}\right)^{\top}\mathbf{R}_{\mathrm{v}}^{-1}\left(\mathbf{y}_{\mathrm{av}}^{\mathrm{f}} - \mathbf{y}_{\mathrm{v}}\right) - \left(\mathbf{y}_{\mathrm{bv}}^{\mathrm{f}} - \mathbf{y}_{\mathrm{v}}\right)^{\top}\mathbf{R}_{\mathrm{v}}^{-1}\left(\mathbf{y}_{\mathrm{bv}}^{\mathrm{f}} - \mathbf{y}_{\mathrm{v}}\right), \tag{E5}$$

where the verification-space vectors and matrix are defined in Section (3.4.2). Instead of using the matrix $\mathbf{R}_{\mathrm{v}}$, we can use other observation-based verification norms (e.g., FSS score for precipitation: Necker *et al.*, 2018). Following Equation (E1) and using a linearized H_{v}, we obtain

$$\delta e_{\mathrm{y}} \approx \left(\mathbf{H}_{\mathrm{v}}\delta\mathbf{x}^{\mathrm{f}}\right)^{\top}\mathbf{R}_{\mathrm{v}}^{-1}\left[\left(\mathbf{y}_{\mathrm{av}}^{\mathrm{f}} - \mathbf{y}_{\mathrm{v}}\right) + \left(\mathbf{y}_{\mathrm{bv}}^{\mathrm{f}} - \mathbf{y}_{\mathrm{v}}\right)\right]. \tag{E6}$$

Sequentially substituting Equations (E2), (5), and (7) into Equation (E6), we obtain the equation of the observation-based EFSOI, which is

$$\mathrm{EFSOI}_{\mathrm{y}} = \frac{1}{N-1}\bar{\mathbf{d}}^{\top}\mathbf{R}^{-1}\mathbf{Y}_{\mathrm{a}}\left(\mathbf{Y}_{\mathrm{av}}^{\mathrm{f}}\right)^{\top} \times \mathbf{R}_{\mathrm{v}}^{-1}\left[\left(\mathbf{y}_{\mathrm{av}}^{\mathrm{f}} - \mathbf{y}_{\mathrm{v}}\right) + \left(\mathbf{y}_{\mathrm{bv}}^{\mathrm{f}} - \mathbf{y}_{\mathrm{v}}\right)\right]. \tag{E7}$$

The C-V method is based on splitting this equation into two parts (Section 3.4.2). Covariance localization can be applied to Equation (E7) in a similar way as in Equation (10).

In addition to Equation (E7), Sommer and Weissmann (2016) suggested calculating the observation-based EFSOI as a linearization around the analysis. As a result, the forecast from the background is no longer needed in the calculation. However, it should be noted that the mathematical expressions from Sommer and Weissmann (2016)'s method are not exactly the same as the ones from the $\mathrm{EFSOI}_{\mathrm{y}}$ given by Equation (E7).

APPENDIX F. THE ESA METHOD

This Appendix provides more information on the use of the ESA method, including how to apply localization in the case of implicit and explicit sensitivity vectors. Let $\mathbf{J} \in \mathbb{R}^N$ be a vector containing a set of J (defined in Section 3.3.2) computed for each ensemble member, then the variance of $\mathbf{J}$ about the ensemble mean value is

$$\sigma_{\mathrm{J}}^2 = \frac{1}{N-1}\delta\mathbf{J}\delta\mathbf{J}^{\top}, \tag{F1}$$

where $\delta\mathbf{J} \in \mathbb{R}^N$ is the ensemble perturbation calculated by subtracting the ensemble mean from $\mathbf{J}$. We may relate $\delta\mathbf{J}$ to ensemble perturbations of model state variables using the first-order Taylor expansion about the ensemble mean, which gives

$$\delta\mathbf{J} \approx \left[\frac{\partial J}{\partial \mathbf{x}}\right]^{\top}\mathbf{X}_{\mathrm{b}}. \tag{F2}$$

Substituting Equation (F2) into Equation (F1) gives

$$\sigma_{\mathrm{Jb}}^2 = \left[\frac{\partial J}{\partial \mathbf{x}}\right]^{\top}\mathbf{P}_{\mathrm{b}}\left[\frac{\partial J}{\partial \mathbf{x}}\right]. \tag{F3}$$

Similarly, we may calculate $\delta\mathbf{J}$ using the analysis perturbation matrix, and obtain

$$\sigma_{\mathrm{Ja}}^2 = \left[\frac{\partial J}{\partial \mathbf{x}}\right]^{\top}\mathbf{P}_{a}\left[\frac{\partial J}{\partial \mathbf{x}}\right]. \tag{F4}$$

Finally, subtracting Equation (F3) from Equation (F4) gives the equation for the ESA method (Equation 12).

F.1 Implicit sensitivity vector

The ESA method with an implicit sensitivity vector is derived as follows. Substituting

$$\mathbf{P}_{\mathrm{a}} - \mathbf{P}_{\mathrm{b}} = -\mathbf{P}_{\mathrm{b}}\mathbf{H}^{\top}\left(\mathbf{H}\mathbf{P}_{\mathrm{b}}\mathbf{H}^{\top} + \mathbf{R}\right)^{-1}\mathbf{H}\mathbf{P}_{\mathrm{b}}$$

(e.g., Nichols, 2010) into Equation (12) gives

$$\delta\sigma_{\mathrm{J}}^2 = -\left[\frac{\partial J}{\partial \mathbf{x}}\right]^{\top}\mathbf{P}_{\mathrm{b}}\mathbf{H}^{\top}\left(\mathbf{H}\mathbf{P}_{\mathrm{b}}\mathbf{H}^{\top} + \mathbf{R}\right)^{-1}\mathbf{H}\mathbf{P}_{\mathrm{b}}\left[\frac{\partial J}{\partial \mathbf{x}}\right]. \tag{F5}$$

Right-multiplying both sides of Equation (F2) by $(N-1)^{-1}\mathbf{X}_{\mathrm{b}}^{\top}$, we obtain

$$\left[\frac{\partial J}{\partial \mathbf{x}}\right]^{\top}\mathbf{P}_{\mathrm{b}} \approx \frac{1}{N-1}\delta\mathbf{J}\mathbf{X}_{\mathrm{b}}^{\top}. \tag{F6}$$

Now, substituting Equation (F6) into Equation (F5) gives

$$\delta\sigma_{\mathrm{J}}^2 = -\frac{1}{(N-1)^2}\delta\mathbf{J}\mathbf{X}_{\mathrm{b}}^{\top}\mathbf{H}^{\top}\left(\mathbf{H}\mathbf{P}_{\mathrm{b}}\mathbf{H}^{\top} + \mathbf{R}\right)^{-1}\mathbf{H}\mathbf{X}_{\mathrm{b}}\delta\mathbf{J}^{\top},$$

where the matrix $\mathbf{P}_{\mathrm{b}}$ can be expressed in terms of the matrix $\mathbf{X}_{\mathrm{b}}$ (Equation A3). Assuming $\mathbf{Y}_{\mathrm{b}} = \mathbf{H}\mathbf{X}_{\mathrm{b}}$, we have

$$\delta\sigma_{\mathrm{J}}^2 = -\frac{1}{(N-1)^2}\left(\delta\mathbf{J}\mathbf{Y}_{\mathrm{b}}^{\top}\right) \times \left(\frac{1}{N-1}\mathbf{Y}_{\mathrm{b}}\mathbf{Y}_{\mathrm{b}}^{\top} + \mathbf{R}\right)^{-1}\left(\delta\mathbf{J}\mathbf{Y}_{\mathrm{b}}^{\top}\right)^{\top}, \tag{F7}$$

which provides an estimate of $\delta\sigma_{\mathrm{J}}^2$ without knowing the sensitivity vector (Griewank *et al.*, 2023; Hakim *et al.*, 2020; Tardif *et al.*, 2022; Torn, 2014).

After applying covariance localization, Equation (F7) becomes

$$\delta\sigma_J^2 = -\frac{1}{(N-1)^2}\left[\frac{\partial J}{\partial \mathbf{x}}\right]^\top \mathbf{L}_{\mathrm{nm}} \circ\left(\mathbf{X}_\mathrm{b}\mathbf{Y}_\mathrm{b}^\top\right)\left(\frac{\mathbf{L}_{\mathrm{mm}}\circ\left(\mathbf{Y}_\mathrm{b}\mathbf{Y}_\mathrm{b}^\top\right)}{N-1}+\mathbf{R}\right)^{-1}\left(\delta\mathbf{J}\mathbf{Y}_\mathrm{b}^\top\right)^\top, \quad \text{(F8)}$$

where $\mathbf{L}_{\mathrm{nm}} \in \mathbb{R}^{n\times m}$ is the model-to-observation localization matrix and $\mathbf{L}_{\mathrm{mm}} \in \mathbb{R}^{m\times m}$ the observation-space localization matrix (Griewank *et al.*, 2023). Equation (F8) shows that, when localization is used, the ESA method generally requires the sensitivity vector to be known.

A possible solution to include localization without calculating the sensitivity vector explicitly is to have individual forecast metrics at each model grid point. Letting J_i denote the forecast metric for the ith grid point and assuming $J = \sum_{i=1}^{n} J_i$, we have

$$\left[\frac{\partial J}{\partial \mathbf{x}}\right]^\top \mathbf{L}_{\mathrm{nm}}\circ\left(\mathbf{X}_\mathrm{b}\mathbf{Y}_\mathrm{b}^\top\right) = \sum_{i=1}^{n}\mathbf{L}_i\circ\left(\delta\mathbf{J}_i\mathbf{Y}_\mathrm{b}^\top\right), \quad \text{(F9)}$$

where $\mathbf{L}_i \in \mathbb{R}^m$ is a vector containing the localization weights determined by the spatial distances between J_i and observations, and $\delta\mathbf{J}_i$ is a row vector containing the ensemble perturbations of J_i (appendix A.1 of Griewank *et al.*, 2023). This kind of localization ($\mathbf{L}_i$) can be considered as sensitivity localization, which has been used in Hakim *et al.* (2020) and in Tardif *et al.* (2022) with a slightly different implementation.

A potential problem of using $\mathbf{L}_i$ in Equation (F9) is that background ensemble perturbations in observation space at one location can influence the forecast metric at different grid points, and this may vary with the forecast lead time. As discussed in Section 5.2.1, the time evolution of the localization can be approximated by, for example, group velocity and advection, but it is unlikely to calculate it accurately in NWP models. The biggest problem is that signal propagation varies for different variables: humidity and hydrometeors will be advected, while wind, pressure, and temperature perturbations can be advected and/or propagate, with group velocities for different wave types (e.g., Rossby or gravity waves). This motivates the use of an explicit sensitivity vector. In this case, localization can be implemented in the same way as in the DA algorithm.

F.2 Explicit sensitivity vector

To obtain an implicit sensitivity vector, we solve the linear system given by Equation (F6). The problem is that this linear system is usually ill-conditioned, as the matrix $\mathbf{P}_\mathrm{b}$ typically has a large condition number. Nomokonova *et al.* (2022) and Griewank *et al.* (2023) have shown that Tikhonov regularization (also known as Ridge regression: Tikhonov, 1965) can be used to improve the conditioning of the linear system and enables a direct numerical solution of the sensitivity vector. In addition, Hacker and Lei (2015) have used the method of Lagrange multipliers and QR factorization to calculate the sensitivity vector.

A particular case of solving the linear system is to ignore the non-diagonal elements of $\mathbf{P}_\mathrm{b}$, which gives

$$\left[\frac{\partial J}{\partial \mathbf{x}}\right]_i = \frac{1}{N-1}\cdot\frac{\left[\mathbf{X}_\mathrm{b}\delta\mathbf{J}^\top\right]_i}{[X_\mathrm{b}]_i^2}, \quad \text{(F10)}$$

where i is the vector index. This diagonal approximation has been proven to be a useful way to calculate the sensitivity vector for large-scale dynamics (Torn & Hakim, 2008). However, it may lead to an overestimation of the sensitivity, especially at small scales, in the presence of model error or with sparse observation networks (Hacker & Lei, 2015; Ren *et al.*, 2019). Therefore, it may not be suitable for convection-permitting NWP. Considering the sensitivity vector as the slope of a linear regression between the metric J and state variables, the diagonal approximation represents a univariate regression and the other a multivariate regression (Ancell & Coleman, 2022).

With an explicit sensitivity vector, $\delta\sigma_J^2$ can be computed directly using Equation (12). The analysis-ensemble perturbation matrix is calculated in the same way as in DA. For example, it can be calculated by

$$\mathbf{X}_\mathrm{a} = \mathbf{X}_\mathrm{b} - \hat{\mathbf{K}}\mathbf{Y}_\mathrm{b}, \quad \text{(F11)}$$

with

$$\hat{\mathbf{K}} = \mathbf{X}_\mathrm{b}\mathbf{Y}_\mathrm{b}^\top\left(\mathbf{Y}_\mathrm{b}\mathbf{Y}_\mathrm{b}^\top+\mathbf{R}\right)^{-\top/2}\left[\left(\mathbf{Y}_\mathrm{b}\mathbf{Y}_\mathrm{b}^\top+\mathbf{R}\right)^{1/2}+\mathbf{R}^{1/2}\right]^{-1}$$

being the modified Kalman gain matrix (Whitaker & Hamill, 2002). We can apply the covariance localization in the calculation of $\mathbf{X}_\mathrm{a}$ by replacing $\hat{\mathbf{K}}$ with

$$\hat{\mathbf{K}}_{\mathrm{loc}} = \mathbf{L}_{\mathrm{nm}}\circ(\mathbf{X}_\mathrm{b}\mathbf{Y}_\mathrm{b})^\top\left(\mathbf{L}_{\mathrm{mm}}\circ(\mathbf{Y}_\mathrm{b}\mathbf{Y}_\mathrm{b})^\top+\mathbf{R}\right)^{-\top/2} \times\left[\left(\mathbf{L}_{\mathrm{mm}}\circ(\mathbf{Y}_\mathrm{b}\mathbf{Y}_\mathrm{b})^\top+\mathbf{R}\right)^{1/2}+\mathbf{R}^{1/2}\right]^{-1}.$$

Thus, by using an explicit sensitivity vector, we can easily include the covariance localization in the calculation of $\delta\sigma_J^2$, without the need to have individual forecast metrics for each grid point and sensitivity localization. This makes the ESA method more flexible in the choice of forecast metrics and more suitable for convection-permitting NWP where flow-dependent localization is needed.

APPENDIX G. THE C-V DIAGNOSTICS

In this Appendix, we provide more information about the C-V diagnostics (Stiller, 2022). To simplify the equations, we define

$$\mathbf{W} = \frac{1}{N-1}\mathbf{R}^{-1}\mathbf{Y}_{\mathrm{a}}\left(\mathbf{Y}_{\mathrm{av}}^{\mathrm{f}}\right)^{\top}\mathbf{R}_{\mathrm{v}}^{-1} \tag{G1}$$

for this Appendix. Furthermore, as stated in Section 3.4.2, the new diagnostics are the contributions $J_{\mathrm{b}}^{(j)}$ and $J_{\mathrm{ab}}^{(j)}$ from a subset of observations to J_{b} and J_{ab}, which can be expressed as

$$J_{\mathrm{b}}^{(j)} = \left(\mathbf{\Pi}_j\bar{\mathbf{d}}\right)^{\top}\mathbf{W}\left(\mathbf{y}_{\mathrm{v}} - \mathbf{y}_{\mathrm{bv}}^{\mathrm{f}}\right), \tag{G2}$$

$$J_{\mathrm{ab}}^{(j)} = \left(\mathbf{\Pi}_j\bar{\mathbf{d}}\right)^{\top}\mathbf{W}\left(\mathbf{y}_{\mathrm{av}}^{\mathrm{f}} - \mathbf{y}_{\mathrm{bv}}^{\mathrm{f}}\right), \tag{G3}$$

where $\mathbf{\Pi}_j$ is the projection operator that sets all vector components to zero, leaving only the component corresponding to the subset j unchanged.

Below, Appendix G.1 presents the reference value $\widetilde{J}_{\mathrm{b}}^{(j)}$ for $J_{\mathrm{b}}^{(j)}$ and shows under which conditions $\widetilde{J}_{\mathrm{b}}^{(j)}$ can be considered as the expectation value $\mathbb{E}\left[J_{\mathrm{b}}^{(j)}\right]$, and Appendix G.2 shows in which way $\mathbb{E}\left[J_{\mathrm{ab}}^{(j)}\right] = \mathbb{E}\left[J_{\mathrm{b}}^{(j)}\right]$ can be considered as an optimality condition.

G.1 Diagnostic 1

Here we show that the reference value for $J_{\mathrm{b}}^{(j)}$ (Equation 19) can indeed be regarded as the expectation value for $J_{\mathrm{b}}^{(j)}$. Noting that $J_{\mathrm{b}}^{(j)} = \mathrm{tr}(J_{\mathrm{b}}^{(j)})$ and using the fact that the trace is invariant under cyclic permutations, Equation (G2) becomes

$$J_{\mathrm{b}}^{(j)} = \mathrm{tr}\left[\mathbf{W}\left(\mathbf{y}_{\mathrm{v}} - \mathbf{y}_{\mathrm{bv}}^{\mathrm{f}}\right)\left(\mathbf{\Pi}_j\bar{\mathbf{d}}\right)^{\top}\right].$$

Then, taking the expectation value, we obtain

$$\mathbb{E}\left[J_{\mathrm{b}}^{(j)}\right] = \mathrm{tr}\left[\mathbf{W}\,\mathbb{E}\left[\left(\mathbf{y}_{\mathrm{v}} - \mathbf{y}_{\mathrm{bv}}^{\mathrm{f}}\right)\left(\mathbf{\Pi}_j\bar{\mathbf{d}}\right)^{\top}\right]\right]. \tag{G4}$$

Now, we can find that $\widetilde{J}_{\mathrm{b}}^{(j)} = \mathbb{E}\left[J_{\mathrm{b}}^{(j)}\right]$ if the condition

$$\frac{1}{N-1}\mathbf{Y}_{\mathrm{bv}}^{\mathrm{f}}(\mathbf{Y_b})^{\top} = \mathbb{E}\left[\left(\mathbf{y}_{\mathrm{v}} - \mathbf{y}_{\mathrm{bv}}^{\mathrm{f}}\right)\bar{\mathbf{d}}^{\top}\right] \tag{G5}$$

holds. Here, the left-hand side is an ensemble estimate of the cross-covariances between assimilation and verification space, and the right-hand side is an innovation-based estimate for these covariances. For the case where the forecast lead time is zero, Equation (G5) corresponds to the well-known result that the covariance of innovations should (ideally) be just the sum of background and observation errors, that is, $\mathbf{D}_{\mathrm{t}} = \mathbf{D}$ (Equations 13 and 14). More precisely, Equation (G5) represents the off-diagonal components of the innovation covariance matrix for which the contribution from the matrix $\mathbf{R}$ vanishes, since errors in the assessed and verifying observations are assumed to be mutually uncorrelated.

To show this more clearly, we write the verifying observation vector as

$$\mathbf{y}_{\mathrm{v}} = H_{\mathrm{v}}(\mathbf{x}_{\mathrm{t}}) + \boldsymbol{\varepsilon}_{\mathrm{v}},$$

where $\boldsymbol{\varepsilon}_{\mathrm{v}}$ is the vector containing the error in $\mathbf{y}_{\mathrm{v}}$. Using a linear assumption, we may also write

$$\mathbf{y}_{\mathrm{bv}}^{\mathrm{f}} = H_{\mathrm{v}}(\mathbf{x}_{\mathrm{t}}) + \mathbf{H}_{\mathrm{v}}\boldsymbol{\varepsilon}_{\mathrm{b}}^{\mathrm{f}},$$

where $\boldsymbol{\varepsilon}_{\mathrm{b}}^{\mathrm{f}}$ is the vector containing the error in $\overline{\mathbf{x}}_{\mathrm{b}}^{\mathrm{f}}$. Therefore, we obtain

$$\mathbf{y}_{\mathrm{v}} - \mathbf{y}_{\mathrm{bv}}^{\mathrm{f}} = \boldsymbol{\varepsilon}_{\mathrm{v}} - \mathbf{H}_{\mathrm{v}}\boldsymbol{\varepsilon}_{\mathrm{b}}^{\mathrm{f}}. \tag{G6}$$

Similarly, for the assessed observations we have

$$\bar{\mathbf{d}} = \boldsymbol{\varepsilon}_{\mathrm{o}} - \mathbf{H}\boldsymbol{\varepsilon}_{\mathrm{b}}, \tag{G7}$$

where $\boldsymbol{\varepsilon}_{\mathrm{o}}$ is the vector containing the error in $\mathbf{y}$ and $\boldsymbol{\varepsilon}_{\mathrm{b}}$ the vector containing the error in $\overline{\mathbf{x}}_{\mathrm{b}}$. We assume all observations and their model equivalents are bias-free, background and observation errors are mutually uncorrelated, and errors in verifying and assimilated observations are also mutually uncorrelated. Now, substituting Equations (G6) and (G7) into the right-hand side of Equation (G5), we obtain

$$\mathbb{E}\left[\left(\mathbf{y}_{\mathrm{v}} - \mathbf{y}_{\mathrm{bv}}^{\mathrm{f}}\right)\bar{\mathbf{d}}^{\top}\right] = \mathbf{H}_{\mathrm{v}}\mathbb{E}\left[\boldsymbol{\varepsilon}_{\mathrm{b}}^{\mathrm{f}}\boldsymbol{\varepsilon}_{\mathrm{b}}^{\top}\right]\mathbf{H}^{\top}.$$

If the ensemble's capability of representing model error is perfect, then the right-hand side of this equation can be estimated by

$$\mathbf{H}_{\mathrm{v}}\mathbb{E}\left[\boldsymbol{\varepsilon}_{\mathrm{b}}^{\mathrm{f}}\boldsymbol{\varepsilon}_{\mathrm{b}}^{\top}\right]\mathbf{H}^{\top} = \frac{1}{N-1}\mathbf{Y}_{\mathrm{bv}}^{\mathrm{f}}\mathbf{Y}_{\mathrm{b}}^{\top}. \tag{G8}$$

This proves Equation (G5) and thus the equality $\mathbb{E}\left[J_{\mathrm{b}}^{(j)}\right] = \widetilde{J}_{\mathrm{b}}^{(j)}$ as requested.

G.2 Diagnostic 2

Here we show that if the analysis state, $\mathbf{x}_{\mathrm{a}}$, is optimal in the sense that δe_{y} (Equation E5) has a global minimum for this analysis state, then

$$\mathbb{E}\left[J_{\mathrm{ab}}^{(j)}\right] = \mathbb{E}\left[J_{\mathrm{b}}^{(j)}\right] \tag{G9}$$

has to hold. In particular, this means that, in the optimal case, the term J_{ab} cancels out exactly half of the total contribution of the term J_b to $EFSOI_y$ in Equation (16). A smaller J_{ab} is less optimal due to the nonlinear nature of the EFSOI where the term J_b varies linearly, while the term J_{ab} varies quadratically with the weights given to the observations (see Section 3.4.2).

To prove Equation (G9), we show that if it is not fulfilled we can construct another analysis state, $\hat{\mathbf{x}}_a$, which gives a smaller value of δe_y. The more optimal analysis state is constructed as

$$\hat{\mathbf{x}}_a = \mathbf{x}_a + \alpha_j \mathbf{K}\mathbf{\Pi}_j \bar{\mathbf{d}},$$

where

$$\alpha_j = \frac{\mathbb{E}\left[J_b^{(j)} - J_{ab}^{(j)}\right]}{\mathbb{E}\left[J_{ab}^{(j,j)}\right]}$$

with

$$J_{ab}^{(j,j)} = \left(\mathbf{H}_v\mathbf{M}\mathbf{K}\mathbf{\Pi}_j\bar{\mathbf{d}}\right)^\top \mathbf{R}_v^{-1}\left(\mathbf{H}_v\mathbf{M}\mathbf{K}\mathbf{\Pi}_j\bar{\mathbf{d}}\right).$$

By definition, $J_{ab}^{(j,j)} \leq 0$. If Equation (G9) is fulfilled, then $\alpha_j = 0$ and $\hat{\mathbf{x}}_a = \mathbf{x}_a$. Using the linear approximation (which is central to the Kalman filter), the forecast (in verification space) initialized from $\hat{\mathbf{x}}_a$ takes the form

$$\overline{H_v(\hat{\mathbf{x}}_a^f)} = \overline{H_v(\mathbf{x}_a^f)} + \alpha_j \mathbf{H}_v\mathbf{M}\mathbf{K}\mathbf{\Pi}_j\bar{\mathbf{d}}.$$

Substituting this equation into

$$\hat{e}_{y,a} = \left(\overline{H_v(\hat{\mathbf{x}}_a^f)} - \mathbf{y}_v\right)^\top \mathbf{R}_v^{-1}\left(\overline{H_v(\hat{\mathbf{x}}_a^f)} - \mathbf{y}_v\right),$$

we obtain

$$\begin{aligned}\mathbb{E}\left[\hat{e}_{y,a}\right] &= \mathbb{E}\left[e_{y,a}\right] + \alpha_j^2\mathbb{E}\left[J_{ab}^{(j,j)}\right] - 2\alpha_j\mathbb{E}\left[J_b^{(j)} - J_{ab}^{(j)}\right] \\ &= \mathbb{E}\left[e_{y,a}\right] - \alpha_j\mathbb{E}\left[J_b^{(j)} - J_{ab}^{(j)}\right] \\ &= \mathbb{E}\left[e_{y,a}\right] - \frac{\left(\mathbb{E}\left[J_b^{(j)} - J_{ab}^{(j)}\right]\right)^2}{\mathbb{E}\left[J_{ab}^{(j,j)}\right]}.\end{aligned}$$

Thus, we have

$$\mathbb{E}\left[\hat{e}_{y,a}\right] < \mathbb{E}\left[e_{y,a}\right]$$

for $\alpha_j \neq 0$, which shows that, if Equation (G9) does not hold, then $\hat{\mathbf{x}}_a$ leads to a smaller value of δe_y. Therefore, Equation (G9) (or equivalently $\alpha_j = 0$) is a necessary condition for $\mathbf{x}_a$ being a global minimum of δe_y. Furthermore, Equation (G9) is also a necessary condition for a *local* minimum at $x = x_a$ (see Stiller, 2022; a local minimum requires that any derivative of δe_y with respect to the initial conditions is zero at $x = x_a$). Generally, Equation (G9) holds if the error covariance matrices employed in the assimilation are completely correct, and there is no bias in the observations and the model states (in which case the analysis state is known to be optimal: Stiller, 2022).