Met Office

Metrics for assessing the impact of observations in NWP: a theoretical study. Part I: optimal systems

Forecasting Research
Technical Report No: 643

9 March 2021

John Eyre

Abstract

Two methods are widely used to assess the impact of observations in global numerical weather prediction (NWP): data denial experiments (DDEs) and the forecast-sensitivity-to-observations-impact (FSOI) method. A DDE measures the impact on forecast accuracy of removing an observation type from the system, whereas FSOI measures the amount by which an observation type reduces the short-range forecast error (within a system containing all observation types). It is usually found, in global NWP experiments in which both methods are used, that the FSOI impact of a given observation type is greater than its DDE impact. However, it can be shown that in some theoretical limits the two should be equal. The aim of this paper is to present the theory behind the DDE and FSOI metrics that are commonly used and to explore factors that cause DDE and FSOI metrics to give different results. The general theory is presented for an optimal analysis and forecast system, and then it is applied to a simple model with two state variables. In order to explain the commonly found result of FSOI impact greater than DDE impact, the following system properties are found to be important: mixing of information (and error) by the forecast model between state variables, the rate of forecast error growth, the presence of forecast model error, the way in which observational information is distributed between state variables and denied from them, and the presence of error in the data used for forecast verification. The FSOI metric uses an energy norm and so assesses impacts in terms of error variance, whereas DDE impacts are often expressed in terms of root-mean-square errors. This difference alone accounts for a factor ~2 between the two metrics. These results provide insight into why NWP systems are resilient to the removal of observational information; they are shown to be resilient if information is denied primarily from well observed variables but not when denied primarily from poorly observed variables.

1. Introduction

The systems that provide the observations used in all the applications that support weather and climate services are expensive to implement and maintain. They are, however, crucial to the utility of these applications and to the accuracy of the products that they provide: "…behind every weather, water and climate condition forecast, every disaster mitigated, and every prediction debated, are the observational data" (WMO 2010). There are continual activities to assess the impact of different observation types, as part of the process of research and development to improve the impact of the observations available today, and to aid decisions about the observing systems to be maintained or developed for the future. Perhaps more than for any other application, the numerical weather prediction (NWP) community has been active in this field and has developed methods for quantitative assessment of the impact of observations on the accuracy of its products.

The assessment approach used in NWP for decades has been through observing system experiments (OSEs), particularly in a mode known as data denial experiments (DDEs). These use real observations, and they measure the effect of removing them from an NWP system. Usually they make use of an NWP data assimilation (DA) system that is the same as or similar to an operational system, assimilating the full range of observation types that are currently used operationally, and then examining the impact of removing one observation type at a time. Impact is assessed using a range of measures available to operational centres, but principally by assessing the impact on the skill of short/medium-range forecasts, using either observations or NWP analyses as the "truth" against which the forecasts are verified. Operational NWP centres have the facilities to do such experiments, as they are necessary tools for testing the assimilation of any new observational dataset before it is used operationally. Occasionally, a

consistent set of OSEs/DDEs is run to assess the impact of a range of observation types in a systematic way (e.g. see Bouttier and Kelly 2001; Bormann *et al.* 2019 and its references; WMO 2016).

The OSE/DDE method is the "cleanest" way of assessing impact on NWP; it measures what we really want to know. However, it can only be used with observations that are available from existing observing systems; it cannot be used to estimate the impact of future observing systems. The latter question can be addressed using observing system simulation experiments (OSSEs). These are similar to OSEs, except that the whole observing system (with both existing and proposed new components) is now simulated from a "nature run", a hypothetical "true" state of the atmosphere generated by running an NWP model without assimilating any observations for several months (e.g. see Errico *et al*., 2013).

The results of OSEs (and OSSEs) are generally noisy. One primary reason for this is because even "good" observations, with small observation errors that are well understood, will degrade the NWP analysis (and subsequent forecasts) in some areas, i.e. those areas where the NWP background field is, by chance, more accurate than the observations. This is a consequence of the inherently statistical nature of the DA problem. Because of these considerations, experience shows that global OSEs/DDEs need to be run for several months before they provide estimates of impact that are statistically significant. Indeed Geer (2016) found that an experiment of ~6 months was the minimum period required to achieve statistical significance for medium-range forecasts for the impact of a typical single instrument in the ECMWF system. The lower the impact of a given observation type, the longer the experiment must be run before results become statistically significant. When a major component of the observing system is removed, shorter experiments should be sufficient.

Consequently, although OSEs/DDEs are straightforward for an operational NWP centre to perform, they can be demanding in computational cost, particularly if many different combinations of observation type or of observation processing method need to be tested. OSSE systems are even more difficult and expensive to build, maintain and run. For more detail on these and other methods for studying the impact of observations on NWP, see Eyre (2018).

A method developed in the last two decades, providing a different approach for assessing observation impact, is the technique now known as the forecast-sensitivity-to-observations impact (FSOI) method (Langland and Baker 2004; Gelaro et al. 2007; Cardinali 2009; Lorenc and Marriott 2014). This technique takes advantage of the components available within a variational DA system to calculate the impact of any observation, or any group of observations, on the accuracy of a forecast (typically a 24-hour forecast for a global NWP system). The FSOI method has the advantage of being relatively cheap in computational terms compared with OSEs; FSOI results can be obtained almost "for free" when running a four-dimensional DA system, and they allow the impact of different observation types, and different sub-divisions of observation types, to be assessed more quickly.

Both OSE/DDEs and FSOI encounter problems associated with the errors in the data used for verification. These have been addressed by several authors, including Daescu (2009), Todling (2013) and Privé *et al.* (2020).

OSE/DDEs and FSOI address different questions. A DDE assesses the degradation in forecast accuracy when a given observation type is removed from the system. The FSOI method assesses how much a given observation type contributes to the reduction of analysis error, and hence forecast error, within a system including all observation

types. One might naively expect these two approaches to give similar results, or at least results related to each other in a simple way but, in general, they do not. Most studies have concluded that the rankings of impact, whether measured by DDE or FSOI, are broadly similar but that the absolute magnitudes of impacts are different.

An example using data from recent DDEs at the Met Office (Candy *et al.* 2021) is shown in Figure 1, which illustrates how FSOI impacts are generally larger than DDE impacts. Here the DDE result is expressed in terms of a percentage increase in forecast error variance when an observation type is denied. It is compared with the percentage FSOI impact, calculated using the method described by Lorenc and Marriott (2014), for the same observation type from the control experiment for the DDE. In both cases, only the T+24 forecasts are considered, and the DDE results are averaged over a range of forecast variables and altitudes, to approximate the FSOI metric as closely as possible.

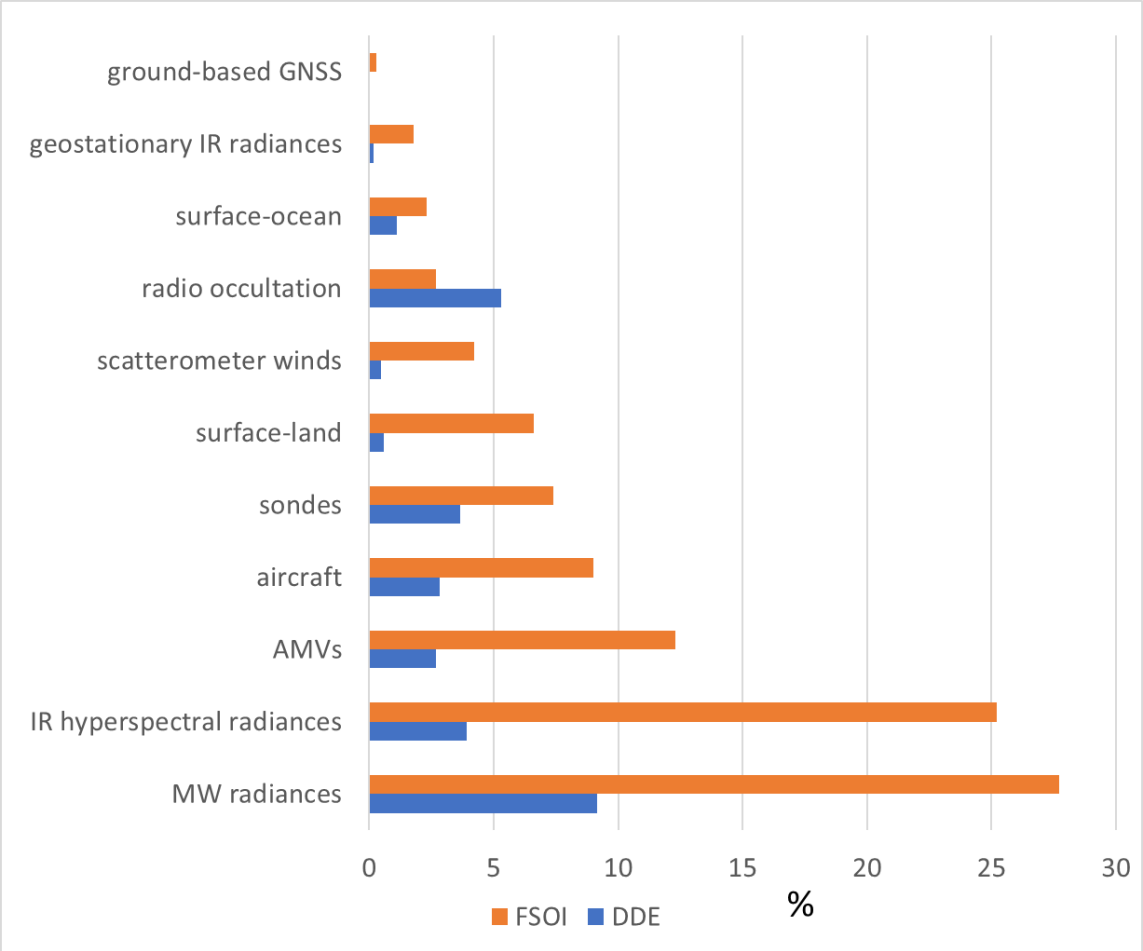

Figure 1. A comparison of DDE and FSOI results from global NWP experiments at the Met Office for August-October 2019 (from Candy et al. 2021). See text for details.

The purpose of this study is to shed some light on what these two approaches really measure and how they are related.

One might naively expect that when (say) 1% of the observations, or perhaps 1% of the observational information, is removed from an NWP system, then the forecast accuracy would be degraded by ~1%. This is found through experience not to be the case; forecast accuracy is degraded by much less than 1%. Why is this? This result is sometimes explained away on the grounds that the NWP system is “resilient”, i.e. that

when one observation type is withdrawn, others step in to fill the gap. Whilst this must be true in some sense, it is not a satisfactory explanation as to why it happens. Also, we can create counter examples, as will be shown in this paper. In fact, we will propose an alternative explanation of NWP system "resilience".

There are also counter examples in the literature. Gelaro and Zhu (2009) show results of DDE and FSOI impacts that are broadly comparable overall, with some examples of FSOI impacts greater than DDE impacts, and also of DDE impacts of all observing systems summing to greater than one (i.e. the opposite of a "resilient" observing system).

It can be argued that the lack of proportionality between observational information removed and the resulting forecast degradation is the result of real NWP systems being suboptimal. This is clearly true in many cases; if the observation error characteristics (their magnitudes and/or their correlations) assumed by the NWP system do not correspond to their true values, then the removal of some observations may not degrade the forecast by as much as expected and might even improve it. The widespread use of observation thinning in NWP is a good example of a response to this problem. There is continuing research within the NWP/DA community to address problems of this type.

In this paper we will consider only optimal systems; some aspects of the problem for suboptimal systems will be considered in a companion study (Part II). We will look here principally at how we would expect the different impact assessment methods, and the metrics they use, to compare if the assimilation and forecast systems were truly optimal.

In section 2 we present the theoretical basis of the study. We show how DDE and FSOI metrics are related to forecast error statistics and to the impact of observations on these statistics. We also prepare the ground for applying this theory to a very simple assimilation and forecast system, in which the forecast model has only two variables. In section 3 we present the details of the simple system used in the study, and the results of running it over a range of system parameters. In section 4 we summarise and discuss these results, and in section 5 present some conclusions and suggestions for future work.

2. Theory

In this paper we try to stay close to the standard DA notation proposed by Ide *et al.* (1997).

2.1 The NWP analysis and its error

An NWP DA system provides an analysis of the atmospheric state which can be expressed as:

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}.(\mathbf{y}^o - H[\mathbf{x}^b]) \ , \qquad (2.1.1)$$

where $\mathbf{x}^a$ is a vector containing the analysed state,
$\mathbf{x}^b$ is a vector containing the background state,
$\mathbf{y}^o$ is a vector containing the observations,
$H[...]$ is a generally nonlinear "observation operator", which maps from the state space to the observation space,

and $\mathbf{K}$ is the analysis operator (Kalman gain matrix), which maps observation increments ("innovations") into analysis increments.

This equation is valid in the "almost linear" case, for which $H[\mathbf{x}^b]$ is linear for typical perturbations of the state in the region of $\mathbf{x}^b$.

For an optimal system,

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} = (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1} \ , \tag{2.1.2}$$

where $\mathbf{B}$ is the background error covariance,
$\mathbf{R}$ is the observation error covariance,
and $\mathbf{H}$ is the Jacobian of $H[\mathbf{x}^b]$, i.e. the gradient of the observation operator with respect to the state space variables.

These equations are derived on the assumption that both the observations and the background are unbiased. An unbiased background also implies an unbiased forecast model (see section 2.2).

The equivalent equation relating the errors in the quantities in (2.1.1) is:

$$\boldsymbol{\varepsilon}^a = \boldsymbol{\varepsilon}^b + \mathbf{K}.(\boldsymbol{\varepsilon}^o - \mathbf{H}\boldsymbol{\varepsilon}^b) \ , \tag{2.1.3}$$

where $\boldsymbol{\varepsilon}^a$ is the error in the analysis,
$\boldsymbol{\varepsilon}^b$ is the error in the background,
and $\boldsymbol{\varepsilon}^o$ is the error in the observations.

Taking ensemble averages of (2.1.3) and assuming the observation errors are uncorrelated with background errors, we find that the mean analysis error is zero (by definition) and the covariance of analysis error, $\mathbf{A}$, is given by

$$\mathbf{A} = \mathbf{B} - \mathbf{K}\mathbf{H}\mathbf{B} = \mathbf{B} - \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{B} \ . \tag{2.1.4}$$

It can also be shown (e.g. see Rodgers, 1976, equations 18 and 22) that

$$\mathbf{A}^{-1} = \mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} \ . \tag{2.1.5}$$

If we define the "precision" of a quantity as the inverse of its error covariance, then (2.1.5) shows that, in an optimal system, the precision of the analysis is the sum of the precision of the background and the precision of the observations mapped into state space. Let us define this latter quantity as

$$\mathbf{Z} = \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}. \tag{2.1.6}$$

Then,

$$\mathbf{A}^{-1} = \mathbf{B}^{-1} + \mathbf{Z} \, . \tag{2.1.7}$$

Note that, even if the observation errors are uncorrelated with each other (i.e. $\mathbf{R}$ is diagonal), $\mathbf{Z}$ will tend to be non-diagonal, because of the effect of $\mathbf{H}$ in mapping $\mathbf{R}^{-1}$ into the different state space variables.

The role of different observations – either individual observations or observation types – can be seen by disaggregating these equations as follows: (2.1.1) becomes

$$\mathbf{x}^a = \mathbf{x}^b + \sum_j \mathbf{K}_j(\mathbf{y}_j^o - H_j[\mathbf{x}^b]) \ , \tag{2.1.8}$$

where j is the jth observation or group of observations,
and $\mathbf{K}_j$ and $H_j[\mathbf{x}^b]$ and the corresponding rows and elements of $\mathbf{K}$ and $H[\mathbf{x}^b]$.

We have assumed here that the errors in observation j (or group of observations j) are uncorrelated with those in other observations.

(2.1.5) then becomes:

$$\mathbf{A}^{-1} = \mathbf{B}^{-1} + \sum_j \mathbf{H}_j{}^T \mathbf{R}_j{}^{-1} \mathbf{H}_j = \mathbf{B}^{-1} + \sum_j \mathbf{Z}_j \ , \tag{2.1.9}$$

where we define

$$\mathbf{Z}_j = \mathbf{H}_j{}^T \mathbf{R}_j{}^{-1} \mathbf{H}_j . \tag{2.1.10}$$

When j refers to a group of observations, $\mathbf{R}$ is assumed to be block diagonal with $\mathbf{R}_j$ representing the jth block.

Equation (2.1.9) is informative; it emphasises further the additive properties of precision matrices in state space.

2.2 The forecast model and the forecast error

The forecast $\mathbf{x}_{n+1}^f$ at time t_{n+1} is generated from the forecast $\mathbf{x}_n^f$ at time t_n using the forecast model $M_n[\mathbf{x}_n^f]$:

$$\mathbf{x}_{n+1}^f = M_n[\mathbf{x}_n^f] . \tag{2.2.1}$$

The equivalent equation relating the errors in the quantities in (2.2.1) is:

$$\boldsymbol{\varepsilon}_{n+1}^f = \mathbf{M}_n \boldsymbol{\varepsilon}_n^f + \boldsymbol{\eta}_n \ , \tag{2.2.2}$$

where $\boldsymbol{\varepsilon}_n^f$ is the error in the forecast $\mathbf{x}_n^f$,
$\boldsymbol{\eta}_n$ is the model error generated in the period from t_n to t_{n+1},
and $\mathbf{M}_n$ is the Jacobian of $M_n[\dots]$.

Taking ensemble averages of (2.2.2), we find that the mean forecast error is zero (by definition) and the covariance of forecast error evolves through:

$$\mathbf{P}_{n+1} = \mathbf{M}_n \mathbf{P}_n \mathbf{M}_n^T + \mathbf{Q}_n \ , \tag{2.2.3}$$

where $\mathbf{Q}_n$ is the model error covariance for the period t_n to t_{n+1}.

We initialise the forecast with an analysis so that, for the first forecast period following the analysis,

$$\mathbf{P}_1 = \mathbf{M}_0 \mathbf{A} \mathbf{M}_0^T + \mathbf{Q}_0 \ . \tag{2.2.4}$$

If $\Delta t = t_{n+1} - t_n$, the period of an assimilation cycle (e.g. 6 hours), then $\mathbf{x}_1^f$ becomes the background for the next assimilation cycle with an estimated error covariance of $\mathbf{P}_1$. (For simplicity, we assume here that the observations are assimilated only at analysis times, at intervals of Δt, and not continuously over an assimilation window.)

In this study we will assume the system evolves until $\mathbf{P}_1$ is equal to the background error covariance used in the preceding assimilation step. In other words, we have a system in equilibrium, in which $\mathbf{Z}$ is the same in each assimilation cycle, as are $\mathbf{M} = \mathbf{M}_0$ and $\mathbf{Q} = \mathbf{Q}_0$. Therefore, the error covariances, $\mathbf{B} = \mathbf{P}_1$, $\mathbf{A}$ and $\mathbf{R}$, are also static and do not change from one assimilation cycle to the next:

$$\mathbf{B} = \mathbf{P}_1 = \mathbf{M}\mathbf{A}\mathbf{M}^T + \mathbf{Q} \,. \tag{2.2.5}$$

where $\mathbf{A}$ is, in turn, derived from $\mathbf{B}$ via (2.1.6).

2.3 The system used in this study

The analysis error covariance for the system in equilibrium is calculated by iteration of (2.1.7) and (2.2.5) to convergence, i.e. by solving for $\mathbf{A}$ the system:

$$\mathbf{A}^{-1} = \mathbf{B}^{-1} + \mathbf{Z}, \tag{2.3.1}$$
$$\mathbf{B} = \mathbf{P}_1 = \mathbf{M}\mathbf{A}\mathbf{M}^T + \mathbf{Q} \,. \tag{2.3.2}$$

From $\mathbf{A}$, we then calculate the error covariances for the forecasts at time intervals of Δt, using (2.2.3):

$$\mathbf{P}_{n+1} = \mathbf{M}\mathbf{P}_n\mathbf{M}^T + \mathbf{Q} \,, \tag{2.3.3}$$

where n is now the index of forecast cycle, i.e. the forecast time $t_n = n\Delta t$. For example, for an assimilation cycle period of 6 hours, $\mathbf{P}_4$ is the error covariance of the 24-hour forecast.

In sections 2.5 and 2.7, we discuss further the appropriate values of $\mathbf{Q}$ to use for each of the different observation impact metrics.

(2.3.1) and (2.3.2) can be combined to give:

$$\mathbf{A}^{-1} = (\mathbf{M}\mathbf{A}\mathbf{M}^T + \mathbf{Q})^{-1} + \mathbf{Z} \,. \tag{2.3.4}$$

In general, this does not have an explicit analytic solution for $\mathbf{A}$, but in the theoretical limit of $\mathbf{Q}$=0 and $\mathbf{M} = a\mathbf{I}$, where $\mathbf{I}$ is a unit matrix, then

$$\mathbf{A}^{-1} = (1 - a^{-2})^{-1}\mathbf{Z} \,. \tag{2.3.5}$$

In this simple case, the precision of the analysis is proportional to the precision of the observations mapped into state space, with the constant of proportionality determined by the forecast error growth factor, a.

2.4 The DDE metrics

DDE impacts can be assessed in many ways: for different forecast lead times, different regions, different heights in the atmosphere, different geophysical variables, etc. In this

study we look at two specific DDE metrics that can easily be compared with FSOI impacts: the percentage change in forecast error variance averaged over many variables, and an energy metric analogous to that used by Gelaro and Zhu (2009).

For the system described above, the DDE metrics are evaluated by comparing the forecast error for the full observing system with the forecast error for the system in which one observation type has been removed. This involves running the system described in section 2.3 twice, once with the full observing system and once with the degraded system. In each case, eq.(2.3.3) is evaluated to give the forecast error covariance for a sequence of forecast lead times.

The first metric is a mean percentage error metric, and this is often used to assess DDE results. For this metric, the difference in error variance between the two runs for each variable is computed as a percentage of its value for the full system. This quantity is then averaged over all (or a subset of) the elements of $\mathbf{x}$. In this study, we will average over all the diagonal elements of $\mathbf{P}_n$:

$$\%DDE_n = 100\ \frac{1}{I}\sum_{i=1}^{I}\frac{\mathbf{P}_{ii_n}(deg)-\mathbf{P}_{ii_n}(full)}{\mathbf{P}_{ii_n}(full)}\ , \tag{2.4.1}$$

where $full$ denotes the full observing system and deg the degraded system. We will focus on the DDE score for the 24h forecast, $\%DDE_4 = \%DDE$. We choose, in this study, to define a DDE score that is positive when the forecast error is increased by the removal of observations, i.e. that the observations removed were having a beneficial impact on the system.

It is important also to note that the metric defined by (2.4.1) is a fractional reduction in forecast error variance. We choose this for comparison with the FSOI metric defined below, which is also an energy/variance metric. However, DDE results are often presented in terms of changes in rms errors. For small fractional changes in forecast errors, there is simple factor of 2 between these two results; a 1% change in rms error is approximately equivalent to a 2% change in error variance. This immediately accounts for some of the difference between DDE and FSOI results as usually presented.

The second DDE metric, an energy metric, is given by:

$$\%DDE_n^e = 100\ \frac{trace\{\mathbf{P}_{ii_n}(deg)\}-trace\{\mathbf{P}_{ii_n}(full)\}}{trace\{\mathbf{P}_{ii_n}(full)\}}\ . \tag{2.4.2}$$

As argued by Gelaro and Zhu (2009), this metric is analogous to the FSOI metric in its use of an energy norm (see section 2.5), which leads to an interesting comparison. However, it is rarely used in DDE assessments at most NWP centres.

It should be noted that DDEs are usually run in a suboptimal way: assuming that the background error covariance matrix, **B**, is close to optimal for the full system, it is not normally retuned for the reduced system. When the observational information denied in a DDE is small (i.e. <~10%) this effect can be shown to be very small, and smaller than the usual uncertainties in **B**.

2.5 The FSOI metric

The FSOI method (as originally proposed by Langland and Baker, 2004) involves computing both the 24h forecast from the analysis and the 24h forecast from the background used for this analysis. In a DA system with a 6h cycling period, this is

equivalent to comparing a 24h forecast from the analysis with a 30h forecast from the previous analysis.

Let us denote the 6h forecast from the analysis by $\mathbf{P}_1$ and the 6h forecast from the background for this analysis $\mathbf{P}_1^B$. These can be obtained from (2.3.3):

$$\mathbf{P}_1 = \mathbf{MAM}^T + \mathbf{Q} \,, \tag{2.5.1}$$

$$\mathbf{P}_1^B = \mathbf{MBM}^T + \mathbf{Q} \,. \tag{2.5.2}$$

Therefore

$$\mathbf{P}_1^B - \mathbf{P}_1 = \mathbf{M}(\mathbf{B} - \mathbf{A})\mathbf{M}^T \,. \tag{2.5.3}$$

Similarly, the difference between the error covariances of the two forecasts at the nth forecast step can be shown to be:

$$\mathbf{P}_n^B - \mathbf{P}_n = \mathbf{M}^n(\mathbf{B} - \mathbf{A})\mathbf{M}^{nT} \,. \tag{2.5.4}$$

For a system in equilibrium, $\mathbf{P}_n^B = \mathbf{P}_{n+1}$, and so

$$\mathbf{P}_{n+1} - \mathbf{P}_n = \mathbf{M}^n(\mathbf{B} - \mathbf{A})\mathbf{M}^{nT} \,. \tag{2.5.5}$$

For the 24h and 30h forecast usually used in the FSOI method and a 6h assimilation cycle, n=4. Note that $\mathbf{Q}$ does not appear in this equation. The role of $\mathbf{Q}$ in DDE and FSOI calculations is discussed further in section 2.7.

From (2.1.7) and (2.1.9), pre-multiplying by $\mathbf{A}$ and post-multiplying by $\mathbf{B}$, we obtain

$$\mathbf{B} - \mathbf{A} = \mathbf{AZB} = \mathbf{A}\textstyle\sum_j \mathbf{Z}_j\, \mathbf{B} \,. \tag{2.5.6}$$

This gives the reduction in analysis error covariance (compared with the background error covariance) caused by all the observations. We can then disaggregate this total reduction in analysis error into the contributions from the different observation types: we define $(\mathbf{B} - \mathbf{A})_j$ as the reduction caused by the jth subset of observations:

$$(\mathbf{B} - \mathbf{A})_j = \mathbf{AZ}_j\mathbf{B} \,. \tag{2.5.7}$$

Substituting (2.5.6) into (2.5.5), we obtain

$$\mathbf{P}_{n+1} - \mathbf{P}_n = \mathbf{M}^n(\mathbf{A}\textstyle\sum_j \mathbf{Z}_j\, \mathbf{B})\mathbf{M}^{nT} \,. \tag{2.5.8}$$

It can be shown (Appendix A) that the FSOI method is equivalent, in some limiting cases (including the one used in this study), to evaluating the trace of this matrix:

$$\delta e_j = trace[\mathbf{M}^n(\mathbf{B} - \mathbf{A})_j\mathbf{M}^{n\boldsymbol{T}}] = trace[\mathbf{M}^n\big(\mathbf{AZ}_j\mathbf{B}\big)\mathbf{M}^{n\boldsymbol{T}}] \,, \tag{2.5.9}$$

and $$\delta e = \textstyle\sum_j \delta e_j = trace[\mathbf{M}^n\big(\mathbf{A}\sum_j \mathbf{Z}_j\, \mathbf{B}\big)\mathbf{M}^{n\boldsymbol{T}}] \;=\; \sum_j trace[\mathbf{M}^n\big(\mathbf{AZ}_j\mathbf{B}\big)\mathbf{M}^{n\boldsymbol{T}}] \,, \tag{2.5.10}$$

where δe is the FSOI energy metric and δe_j is its contribution from observation subset j.

From this, we can define the FSOI percentage error reduction attributable to observation subset j as:

$$\%FSOI = 100\ \delta e_j / \delta e\ . \qquad (2.5.11)$$

In this study $\%FSOI$ is calculated through evaluation of eq.(2.5.9), (2.5.10) and (2.5.11) for n=4.

An important property of $\%FSOI$ is the way in which the impact of observations is effectively normalised; the sum for all observations = 100%. This will be true no matter how effectively the assimilation uses observations. Also, although in this paper we consider only optimal systems, this result will also hold for suboptimal systems, unlike the DDE metrics.

Note that, for an optimal system, $\%FSOI$ cannot be negative. In a suboptimal system, it can be negative, and this is a useful diagnostic for erroneous observations or poorly tuned error statistics. However, when this is the case for one observation type, $\%FSOI$ for other types will compensate such that their total is 100%.

As with DDE scores, we choose in this study to define a positive FSOI score when the observations are having a beneficial impact on the system.

We also define ratios between this FSOI metric and the DDE metrics define above:

$$RAT = \%FSOI / \%DDE\ . \qquad (2.5.12)$$

$$RAT^e = \%FSOI / \%DDE_n^e\ . \qquad (2.5.13)$$

The main focus of this study is an exploration of how and why these ratios depart from unity.

Before leaving the discussion of the FSOI metric, we note the link between this metric and the degrees of freedom for signal (DFS) metric, which is also used to assess the potential impact (i.e. the impact in the optimal case) of observations within an NWP system. The relationship between these two metrics is presented in Appendix B.

2.6 The scalar case

When there is only one analysis variable, $\mathbf{M}$ becomes a and eq.(2.3.4) simplifies to

$$A^{-1} = (a^2 A + Q)^{-1} + Z\ . \qquad (2.6.1)$$

where

$$Z = \sum_j Z_j = \sum_j R_j^{-1}. \qquad (2.6.2)$$

Eq.(2.6.1) can be written as

$$A^2 a^2 Z + A(1 - a^2 + QZ) - Q = 0\ , \qquad (2.6.3)$$

which can be solved analytically for A. In the limit of $Q = 0$, A^{-1} (a scalar) simplifies further; it becomes proportional to Z (also a scalar) through eq.(2.3.5). Therefore, a small increase in Z, δZ, will result in a proportionate increase in A^{-1}: $\delta(A^{-1})/A^{-1} = \delta Z/Z$. Additionally, the same change, δZ**,** will result in a decrease in A approximately proportional to Z: $\delta A/A \simeq -\delta Z/Z$ for $\delta Z \ll Z$. Consequently, in the same limit, changes in

P_n will also be approximately proportional to the changes in Z, through equations (2.3.2) and (2.3.3). It follows also that the change in forecast error variance, as measured by the DDE metric (eq.(2.4.1)), will be proportional to small changes in Z to the same extent.

Turning to the FSOI metric, in the scalar case A and B are fixed scalar quantities for a given observing system, and so $B - A$ is proportional to Z through eq.(2.5.6). Therefore e is proportional Z and e_j to Z_j through equations (2.5.9) and (2.5.10).

So, in the scalar case, a reduction of 0.01 (1%) in Z will result in $\%FSOI$=1 and also $\%DDE_n = \%DDE_n^e \simeq 1$ for all n, and hence $RAT = RAT^e \simeq 1$. This is the result obtained by Eyre and Weston (2014), who also showed that the effect of reasonable non-zero values of Q on this result is small. Because of the approximate proportionality between Z and DDE impacts, scalar systems are not "resilient" to the removal of observations in the way that NWP systems are often described (see section 1).

Although scalar systems can be used to demonstrate some interesting properties of observing system design, they clearly do not replicate the results found with real-world NWP systems, in which $\%DDE$ tends to be significantly less than $\%FSOI$, and so $RAT > 1$; hence the need to look at systems of more than one variable.

2.7 The effects of model error on observation impact metrics

Considering firstly the FSOI metric: Langland and Baker (2004) state that, while the errors in forecasts computed from the analysis and from its background are caused by inaccuracies in both the initial conditions and the forecast model, the difference between the forecast errors used in the FSOI metric is due solely to the differences in the initial conditions caused by the observations assimilated in a single DA cycle. In support of this statement they point out that, if there are no observations in this cycle, then the two forecasts follow the same trajectory through state space. Whilst in practice the two forecasts in question do not follow exactly the same trajectory, and so will not experience exactly the same model errors, they will experience similar model errors, particularly as these are likely to be strongly correlated locally (in time and space). It follows therefore that the usual approximation for the FSOI method of $\mathbf{Q} = 0$ is likely to be good; as can be seen from the equations in section 2.5, $\mathbf{Q}$ can be eliminated.

For the DDE method, the two forecasts being compared do not originate from the same analysis/background pair but from forecasts that evolve independently. They therefore experience model errors that are not so strongly correlated. Nevertheless, if only a small amount of observational information is denied in a DDE, then the short-range forecast trajectories will be similar and the associated model errors may be correlated to some extent. In this study we examine both limits for a DDE: the one in which $\mathbf{Q}$ takes its full value as given by eq.(2.3.3), and the one in which $\mathbf{Q} = 0$.

2.8 The effects of errors in the verification data

A forecast error is the difference between the forecast and the true atmospheric state. Unfortunately, in a DDE or in the computation of FSOI, the true state is not known; a proxy for the truth must be found, and it is usual to use either observations or an analysis at the time for which the forecast is valid. These proxies contain their own errors, and the effect of these errors on the two metrics should be taken into account.

Because of this problem we never measure a forecast error directly but always the difference $\mathbf{d}_v^f$ between a forecast $\mathbf{x}^f$ and the verifying data $\mathbf{x}^v$:

$$\mathbf{d}_v^f = \mathbf{x}^f - \mathbf{x}^v = \boldsymbol{\varepsilon}^f - \boldsymbol{\varepsilon}^v \ , \tag{2.8.1}$$

where $\boldsymbol{\varepsilon}^f$ and $\boldsymbol{\varepsilon}^v$ are the errors in the forecast and the verifying data, both of which we assume to be unbiased (for the purposes of this study – see section 2.1).

Taking ensemble averages of (2.8.1), we find that the covariance of the difference is given by

$$\mathbf{D}_n = \mathbf{P}_n + \mathbf{V} \ , \tag{2.8.2}$$

where $\mathbf{D}_n$ is the covariance of $\mathbf{d}_v^f$,
and $\mathbf{V}$ is the error covariance of $\mathbf{x}^v$.

It is assumed here that the errors in the forecast and the verifying data are uncorrelated. This may or may not be a valid approximation for a 24h forecast. In practice it will be compromised by biases in either the forecast or the verifying data. These effects are considered by Privé *et al.* (2020) using an OSSE framework. They are not considered further in this study.

We examine the limiting assumption that the error in the verifying data is zero, but also the effect of using an independent analysis as the verifying data, and here we assume that this has an error covariance equal to that of the system with the full set of observations. We consider this limit for the DDE metric, noting that verification error is eliminated in the FSOI metric; it can be seen from eq.(2.5.8) that the addition of equal verification error to both forecast errors does not affect the result (but noting again the assumption that the errors in neither forecast are correlated with those in the verifying analysis).

3. Theoretical study of a simple system

3.1 Experimental design

In this study we apply the equations presented in section 2.3 to a system of two state variables. Such a system is clearly very different from the systems of millions of variables used in NWP, but it is nevertheless a non-scalar system and displays some interesting properties which (we will argue) are relevant to understanding some aspects of the behaviour of systems of larger dimension.

What do these two “variables” represent? They can be taken to stand for various aspects of an NWP system: multiple geophysical variables (temperature, wind, etc.), or the control variables of an NWP DA system (e.g. wind field, unbalance mass field, etc.), or different spatial grid-points both vertically or horizontally, or different scales of motion (large scale and small scale), or combinations of them all. In real-world systems, some of these variables are coupled to each other more strongly than others, and on different timescales. Also, in real-world systems, some “variables” (using the rather loose definition above) are observed better than others. In a DDE, information is denied preferentially from some variables rather than from others, and the effect of removing information from better observed variables may be different from removing it from poorly observed variables. Let us consider two examples. Radiosondes observe variables close to model grid points very well and they do not observe variables away from these

grid points. If we remove some radiosondes, we deny information primarily from well observed variables. Similarly, for radiances from passive satellite sounders, large scales in the vertical are well observed and small scales are not observed. If we deny some of these radiances, we deny information primarily from well observed variables. Overall, some variables are not observed at all, and they are analysed in a DA system via the B-matrix, which is correlated because of mixing via the model (see section 3.2).

The model is run with various sets of parameter values. The parameters represent the observations and their error statistics, and how they project on to the state variables. They also include parameters representing the generation and propagation of forecast error – $\mathbf{M}$ and $\mathbf{Q}$. To avoid the problem of representing specific observation types with specific values of $\mathbf{H}$, we describe the observational information in terms of $\mathbf{Z}$, the observation precision matrix mapped into state space, and we vary this matrix and also the way in which information is "denied" from this matrix to simulate a DDE. Note that $\mathbf{Z}$ has the same dimensions as $\mathbf{A}$ and $\mathbf{B}$ but the observations themselves may be of different dimension.

The model is initialised with a very large value of $\mathbf{B}$ (small value of $\mathbf{B}^{-1}$) and then iterated until the value of $\mathbf{A}$ converges (i.e. iterating equations (2.3.1) and (2.3.2)). 20 iterations are found to be sufficient for the sets of parameters used in this study. When $\mathbf{A}$ has converged, the forecast model is then run forward for 5 steps. With an assimilation cycle corresponding to 6 hours, this provides the 24- and 30-hour forecast, which allows computation of the FSOI metric for the 24h forecast.

For each set of parameters, the model is run twice, once with the full set of observational information and then with a degraded set to simulate a DDE (as described in section 3.2). From the results of these two runs, the DDE diagnostics are calculated and, from the run with the full set, the FSOI diagnostics are calculated, using the equations presented in sections 2.4 and 2.5.

3.2 System parameters

The observations and their errors are represented by the observation precision matrix mapped into state space, $\mathbf{Z}$ – see eq.(2.1.6). For simplicity, $trace[\mathbf{Z}]$, is set to 2 for the control in every experiment. However, the ratio of the diagonal elements of $\mathbf{Z}$ is varied to represent a system in which one variable is observed better than the other. The ratio is given by ρ:

$$\rho = Z_{11}/Z_{22}\ , \tag{3.2.1}$$

under the constraint that $Z_{11} + Z_{22} = 2$. (Note that an "experiment" in this context involves two runs of the model: one with $trace[\mathbf{Z}] = 2$, and one with $trace[\mathbf{Z}]$ reduced as explained below.)

In some experiments, off-diagonal elements are added to $\mathbf{Z}$ such that the associated correlation coefficient is $-c$ (i.e. the correlation coefficient for $\mathbf{Z}^{-1}$ is c). This represents the inevitable correlation of the errors in $\mathbf{Z}$ through $\mathbf{H}$, even if the observation errors themselves are uncorrelated (i.e. diagonal $\mathbf{R}$).

To construct a DDE, observational information must be denied (i.e. $\mathbf{Z}$ reduced) in one variable or both. In all experiments we reduce the observation "information" (quantified by $trace[\mathbf{Z}]$) by 1%, i.e. change it by -0.02 in total. In most experiments we reduce $\mathbf{Z}$

only in the first variable, Z_{11}. In others we reduce it only in the second, Z_{22}; or in both equally, $\delta Z_{11} = \delta Z_{22} = -0.01$; or in proportion to $\mathbf{Z}$: $\delta Z_{11}/Z_{11} = \delta Z_{22}/Z_{22} = -0.01$.

The Jacobian of the forecast model is represented as constant in time (i.e. $\mathbf{M}$ is not state dependent) and in the form

$$\mathbf{M} = a\begin{pmatrix} 1-m & m \\ m & 1-m \end{pmatrix}. \qquad (3.2.2)$$

a represents the average amplification of error by the model, and m the mixing of error between the two variables by the forecast process. The use of a scalar a implies the same error growth rate in each variable, which is a simplification adopted for the purposes of this study.

a is set to 1.2, which represents a doubling time for forecast error amplitude of ~24h, or for forecast error variance of ~12h. This is based on the work of Simmons and Hollingsworth (2002) for the average values found in global NWP systems. It is consistent with the values used by Eyre and Weston (2014), where it is discussed in more detail. Other experiments are run with a is set to 1.4 and 1.1, to test the sensitivity to this parameter.

Some experiments are run with $m = 0$. These are called "simple" systems. They exhibit some interesting behaviour, but the results show that these systems do not exhibit the characteristics of $\%DDE$, $\%FSOI$ and RAT found in real-world systems. Therefore, for most experiments, we use a non-zero value of m. For most experiments we use $m = 0.05$ and for one we use $m = 0.1$. These lead to correlations between the errors in the two variables in the 6h forecast of ~25% and ~60% respectively. Larger values of m are found to lead to rapid "saturation" of $\mathbf{P}$, i.e. correlation of errors between the two variables >90% at 6 hours and higher at 24 hours. The behaviour of the system in this limit has not been explored further.

Some experiments are run with $\mathbf{Q} = 0$ and some with a value of $\mathbf{Q}$ set to give approximately the situation reported by Raynaud *et al.* (2012): that variance of the 6h forecast error is found to be ~20% higher with real data than with simulated data for $\mathbf{Q} = 0$. $\mathbf{Q}$ is approximated as diagonal, with equal values on the diagonal, which again is a simplification adopted for the purposes of this study. A value of $Q_{ii} = 0.02$ has been found empirically to reproduce the conditions reported by Raynaud *et al*., giving increases in 24h forecast error variance in the range 18-25% depending on the value of ρ, and with $m = 0.05$. $Q_{ii} = 0.02$ has been used for the baseline system (see below). One experiment has been run with $Q_{ii} = 0.04$.

In order to explore the parameter space, we define a baseline configuration of: $a = 1.2$, $m = 0.05$, $Q_{ii} = 0.02$, $c = 0$. Then, we explore departures from this configuration. In all experiments we explore values of ρ = 1, 2, 4 and 8, and in some we include additionally ρ = 16, 32, 64 and 128.

The set of experiments exploring the parameter space is summarised in Table 3.2.1. The parameter "Z-red" describes how $\mathbf{Z}$ is reduced in the DDE: either in variable 1 only, or in variable 2 only, or equally in both variables, or in proportion to the value of $\mathbf{Z}$ in each variable, as explained above. In all experiments we use $trace[\mathbf{Z}] = 2$.

Expt	A	M	Q_{ii}	c	Z-red	Description
1	1.2	0	0	0	1	Simple (no mixing, diagonal $\mathbf{Z}$, $\mathbf{Q} = \mathbf{0}$)
2	1.2	0	0	0.1	1	Simple with non-diagonal $\mathbf{Z}$
3	1.2	0	0.02	0	1	Simple with non-zero $\mathbf{Q}$
4	1.2	0.05	0	0	1	As Expt.1, but with mixing (= baseline with $\mathbf{Q} = 0$)
5	1.2	0.05	0.02	0	1	**Baseline**
6	1.2	0.1	0.02	0	1	Baseline with double mixing
7	1.2	0.05	0.04	0	1	Baseline with double $\mathbf{Q}$
8	1.2	0.05	0.02	0.1	1	Baseline with non-diagonal $\mathbf{Z}$
9	1.4	0.05	0.02	0	1	Baseline with increased a
10	1.1	0.05	0.02	0	1	Baseline with decreased a
11	1.2	0.05	0.02	0	2	Baseline, $\mathbf{Z}$ reduced in 2nd variable
12	1.2	0.05	0.02	0	=	Baseline, $\mathbf{Z}$ reduced equally both variables
13	1.2	0.05	0.02	0	prop	Baseline, $\mathbf{Z}$ reduced in proportion to $\mathbf{Z}$
14	1.2	0.05	0.02	0	1	Baseline with verification error (DDE only)

Table 3.2.1 Summary of experiments. See text for explanation of Z-red.

3.3 Results

3.3.1 Comparisons of $\%FSOI$ and $\%DDE$

In this section we discuss results for $\%DDE$, $\%FSOI$ and their ratio, RAT. The primary purpose of these experiments is to explore how and why the ratio differs from one. The results of all experiments are given in the tables in Appendix C, for completeness. Figures in this section illustrate key results. Results for $\%FSOI$, $\%DDE^e$ and their ratio, RAT^e, are also presented in the tables, and they are discussed in section 3.3.2.

Expt.1: "Simple" – no mixing, no model error, no correlation in $\mathbf{Z}$.

The results of this experiment are given in Table C.1 and in Figure 2 (thick/red lines).

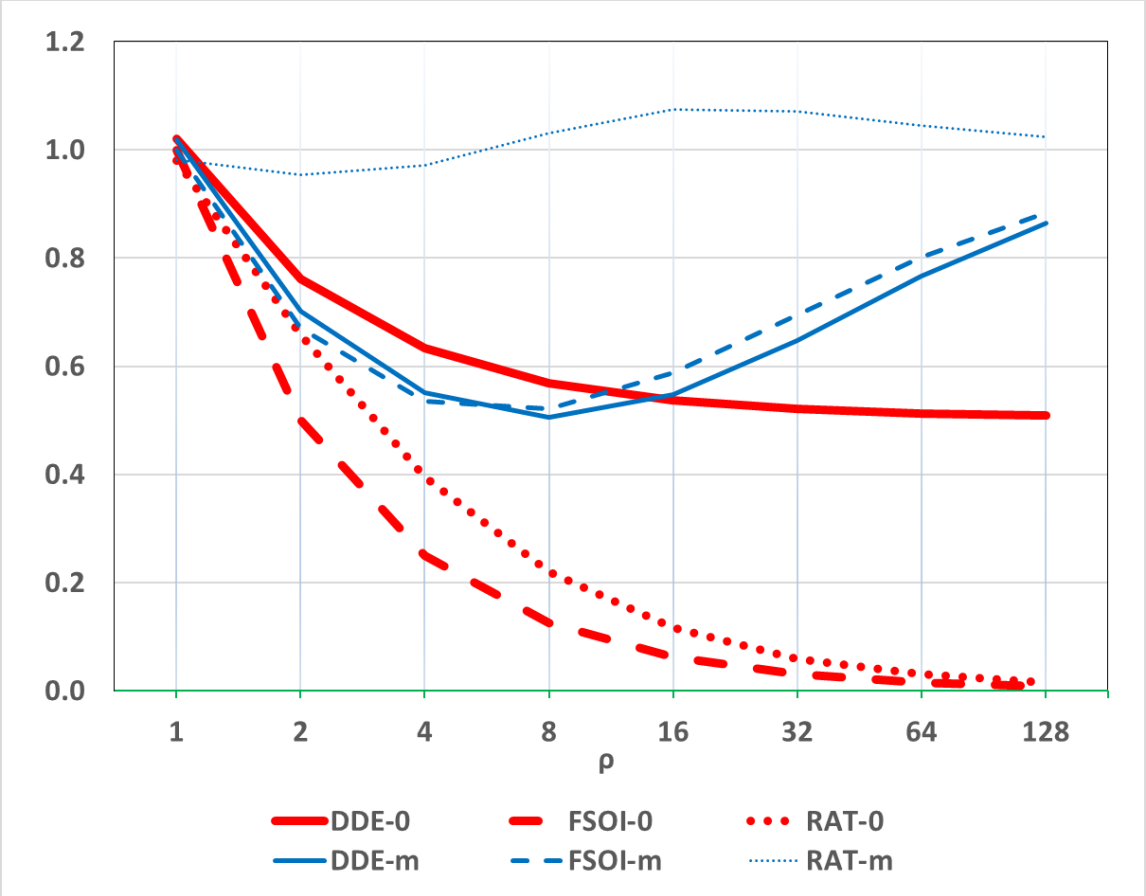

Figure 2. Results of $\%DDE$, $\%FSOI$ and RAT for Expt.1 ("-0") and Expt.4("-m"), illustrating the effect of introducing mixing.

This system is equivalent to two uncoupled scalar systems each with the properties discussed in section 2.6.

Looking first at the FSOI results, we find that, for $\rho = 1$, $\%FSOI = 1$. (We also find this result for all the experiments in Table 3.2.1 for $\rho = 1$). This can be understood by considering equations (2.5.6) and (2.5.7) – when $Z_{11} = Z_{22}$, the elements of $\mathbf{A}$ and $\mathbf{B}$ will be equal for both variables, and then $(\mathbf{B} - \mathbf{A})_j$ will be proportional to $\mathbf{Z}_j$, no matter into which variable the observational information projects.

When $\rho > 1$, we find $\%FSOI = \rho^{-1}$. This can also be understood from eq.(2.3.5) – $\mathbf{A}^{-1}$, and hence $\mathbf{B}^{-1}$, are proportional $\mathbf{Z}$, and so in eq.(2.5.6) the diagonal elements of $(\mathbf{B} - \mathbf{A})$ are proportional to the diagonal elements of $\mathbf{Z}^{-1}$, with a ratio between them of ρ^{-1}.

Looking now at the $\%DDE$ results, we find that $\%DDE$ is approximately proportional to $(1 + \rho)/2\rho$. It can be shown that this is expected from eq.(2.4.1) in the limit of small changes in $\mathbf{Z}$. For $\rho = 1$, $\%DDE$ is not exactly equal to 1 because the change in $\mathbf{Z}$ is finite, and $\mathbf{A}^{-1}$ (not $\mathbf{A}$) is exactly proportional to $\mathbf{Z}$.

Consequently, $RAT = \%FSOI/\%DDE$ is approximately proportional to $2/(1 + \rho)$, which is 1 when $\rho = 1$ and becomes progressively smaller as ρ increases. This is not the result found with real-world NWP systems, where RAT is usually found to be greater than 1 (see Fig.1).

So, both $\%FSOI$ and $\%DDE$ decrease as ρ increases. Another way of understanding these results is to recall that, when $\rho > 1$, the first variable is the better observed variable. For this variable, the background and analysis and subsequent forecast errors are lower than in the second variable. Both metrics involve averages or summations over both variables; a 2% improvement in the forecast of the well observed variable, i.e. the one with the smaller forecast error, will result in a change of less than 1% in the average over both variables (because a 2% increase in the smaller number and 0% increase in the larger number causes an increase of <1% in their mean). Note also that the result of $\%DDE < 1$ for $\rho > 1$ is the first evidence we have for why and under what conditions a system is "resilient" to the removal of observations. This issue is discussed further in section 4.

Expt.2: "Simple", but with some correlation in $\mathbf{Z}$.

In this configuration, the introduction of correlation in $\mathbf{Z}$ makes no difference to either $\%FSOI$ or $\%DDE$; they are exactly the same as in Expt.1 (and so are not repeated in Table C.1). This is because, when the variables are uncoupled and with zero model error, the correlations in $\mathbf{Z}$ create equal correlations in $\mathbf{A}^{-1}$ but do not affect its diagonal elements. (However, it is found that, as $c \rightarrow 1$, the system of equations can become unstable or slow to converge for some combinations of system parameters.)

Expt.3: "Simple", but with some model error

The results of this experiment are also given in Table C.1 and Figure 3 (thin/red lines).

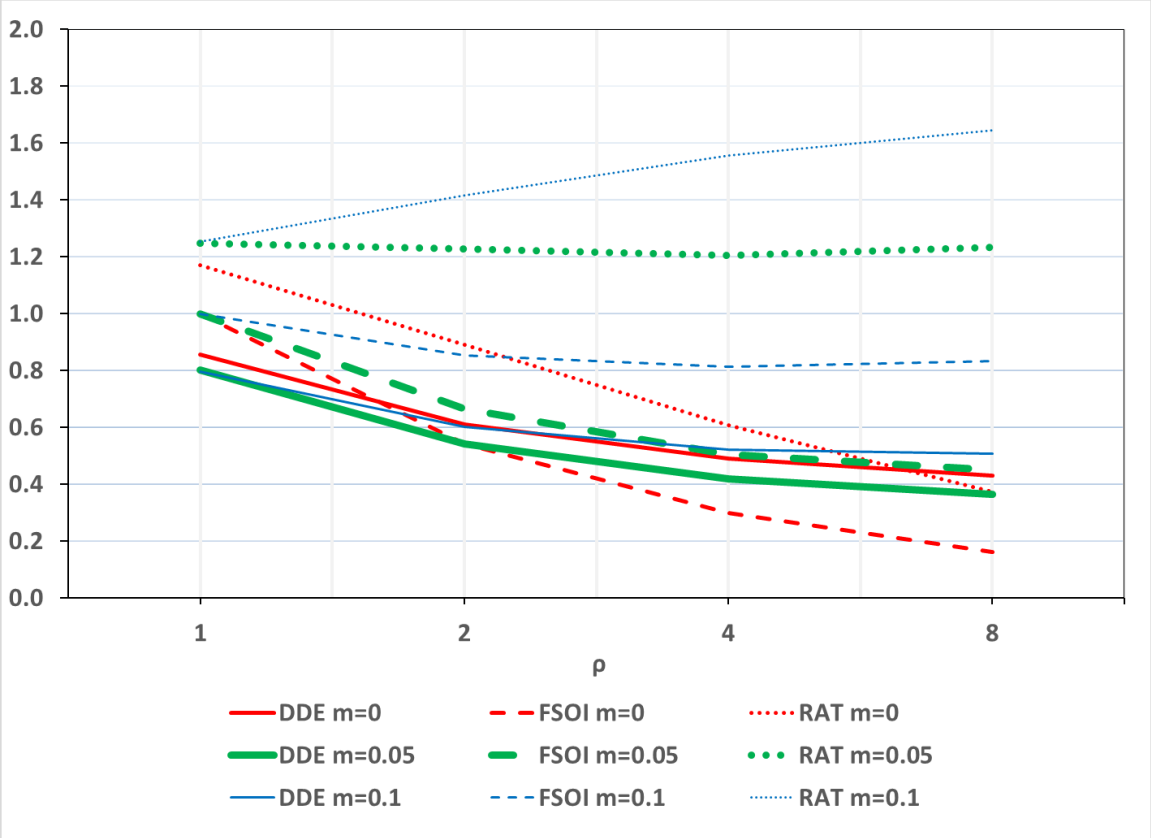

Figure 3. Results of $\%DDE$, $\%FSOI$ and RAT for Expt.3 ("m=0"), Expt.5=baseline ("m=0.05") and Expt.6("m=0.1") illustrating the effect of different levels of mixing in the presence of model error.

Compared with Expt.1, there are small increases in $\%FSOI$ for $\rho > 1$. These changes occur because the introduction of model error changes the ratio of the analysis error variances in the two variables. There are substantial decreases in $\%DDE$, and hence values of RAT are increased relative to Expt.1, but RAT is still < 1 (i.e. $\%DDE > \%FSOI$) except for $\rho = 1$. With a non-zero value of model error, we see the effect on $\%DDE$ of the system "forgetting" observational information more quickly. The effect is much less pronounced for $\%FSOI$ because of the way in which this quantity is normalised (see section 2.5).

Expt.4. As Expt.1, but with mixing (or baseline with no model error)

Results are also shown in Table C.1 and Figure 2 (thin/blue lines).

The results are discussed relative to the baseline below, but we first consider them relative to Expt.1, as this provides some insight into the role of mixing in these experiments. When $\rho = 1$, $\%FSOI = 1$ and $\%DDE \simeq 1$, as in Expt.1 and for the same reasons. However now, as ρ tends to ∞, both $\%FSOI$ and $\%DDE$ tend to 1. This is because, in the limit of $\mathbf{Z} = \begin{pmatrix} Z & 0 \\ 0 & 0 \end{pmatrix}$, all the elements of $\mathbf{B} - \mathbf{A}$, $\mathbf{A}$ and $\mathbf{B}$ become proportional to Z. Consequently, as ρ increases, both $\%FSOI$ and $\%DDE$ first decrease and then they both subsequently increase again and asymptote towards 1. Compare this with Expt.1 where, in the limit of no mixing and as ρ increases, $\%FSOI$ tends to 0 and $\%DDE$ tends to 0.5. The value of RAT is close to 1 in this experiment, and it is substantially increased compared with Expt.1. Therefore, it seems that mixing is necessary for increasing the value of RAT but not necessarily sufficient for maintaining it well above 1.

Expt.5: **Baseline** – forecast error growth rate, $a = 1.2$; mixing coefficient, $m = 0.05$; model error, $Q_{ii} = 0.02$; uncorrelated $\mathbf{Z}$.

For illustration, Table 3.3.1 gives some intermediate results of the computation for this combination of parameters and for $\rho = 4$.

			Full	Deg
		$\mathbf{Z}$	$\begin{pmatrix} 1.600 & 0 \\ 0 & 0.400 \end{pmatrix}$	$\begin{pmatrix} 1.580 & 0 \\ 0 & 0.400 \end{pmatrix}$
		$\mathbf{B}^{-1}$	$\begin{pmatrix} 4.101 & -0.885 \\ -0.885 & 1.459 \end{pmatrix}$	$\begin{pmatrix} 4.063 & -0.881 \\ -0.881 & 1.458 \end{pmatrix}$
		$\mathbf{A}^{-1}$	$\begin{pmatrix} 4.701 & -0.885 \\ -0.885 & 1.859 \end{pmatrix}$	$\begin{pmatrix} 5.643 & -0.881 \\ -0.881 & 1.858 \end{pmatrix}$
		$\mathbf{A}$	$\begin{pmatrix} 0.189 & 0.090 \\ 0.090 & 0.581 \end{pmatrix}$	$\begin{pmatrix} 0.191 & 0.091 \\ 0.091 & 0.581 \end{pmatrix}$
forecast error covariance	6h	$\mathbf{B} = \mathbf{P}_1$	$\begin{pmatrix} 0.281 & 0.170 \\ 0.170 & 0.788 \end{pmatrix}$	$\begin{pmatrix} 0.283 & 0.171 \\ 0.171 & 0.788 \end{pmatrix}$
	24h	$\mathbf{P}_4$	$\begin{pmatrix} 0.869 & 0.773 \\ 0.773 & 1.973 \end{pmatrix}$	$\begin{pmatrix} 0.875 & 0.776 \\ 0.776 & 1.975 \end{pmatrix}$
	30h	$\mathbf{P}_5$	$\begin{pmatrix} 1.262 & 1.201 \\ 1.201 & 2.693 \end{pmatrix}$	$\begin{pmatrix} 1.271 & 1.206 \\ 1.206 & 2.696 \end{pmatrix}$

Table 3.3.1. Intermediate results of the computation for the baseline case for $\rho = 4$. "Full" is the full observing system and "Deg" is the degraded observing system, with Z_{11} reduced by 0.02. (Values are retained at higher precision than those shown here for subsequent calculations.)

The results of this experiment are given in Table C.2 and Figure 3 (thick/green lines). Compared with Expt.3 (in which Q_{ii} also equals 0.02), the introduction of mixing has a large effect on $\%FSOI$, which now decreases more slowly as ρ increases. This is

because the mixing tends to equalise the forecast errors in the two variables, both during the assimilation and in the forecast steps. Therefore, the effects of $\mathbf{A}$ and $\mathbf{B}$ in eq.(2.5.6) become less pronounced.

The mixing has a comparatively small effect on $\%DDE$, for which the introduction of model error has a larger effect for small values of ρ, as in Expt.3. Therefore RAT now has a very different form, being around 1.2 for small values of ρ and increasing for larger values. So, RAT is now greater than unity, as found in real-world systems.

Expt.6. Baseline, with doubled mixing coefficient.

The results of this experiment are also given in Table C.2 and Figure 3 (very thin / blue lines), together with those of Expt.3 and 5.

Comparing experiments 3, 5 and 6, in which the mixing increases from 0 to 0.05 to 0.1 respectively, we see how the $\%FSOI$ increases towards 1 as the mixing increases and the forecast errors in the different variables become more similar. The effect on the $\%DDE$ is smaller. It is almost more complex; with increasing ρ, $\%DDE$ initially decreases but then increases again when mixing is present, as in Expt.4. The overall effect is for RAT to increase with increasing m, for $\rho > 1$.

Expt.7. Baseline, with double model error.

Results are shown in Table C.3 and Figure 4 (very thin / blue lines), along with those of Expt.4 and 5, to show the effects of increasing values of model error. The effects on $\%FSOI$ are very small. $\%DDE$ decreases significantly with increasing model error and with increasing ρ, thus increasing RAT for all ρ.

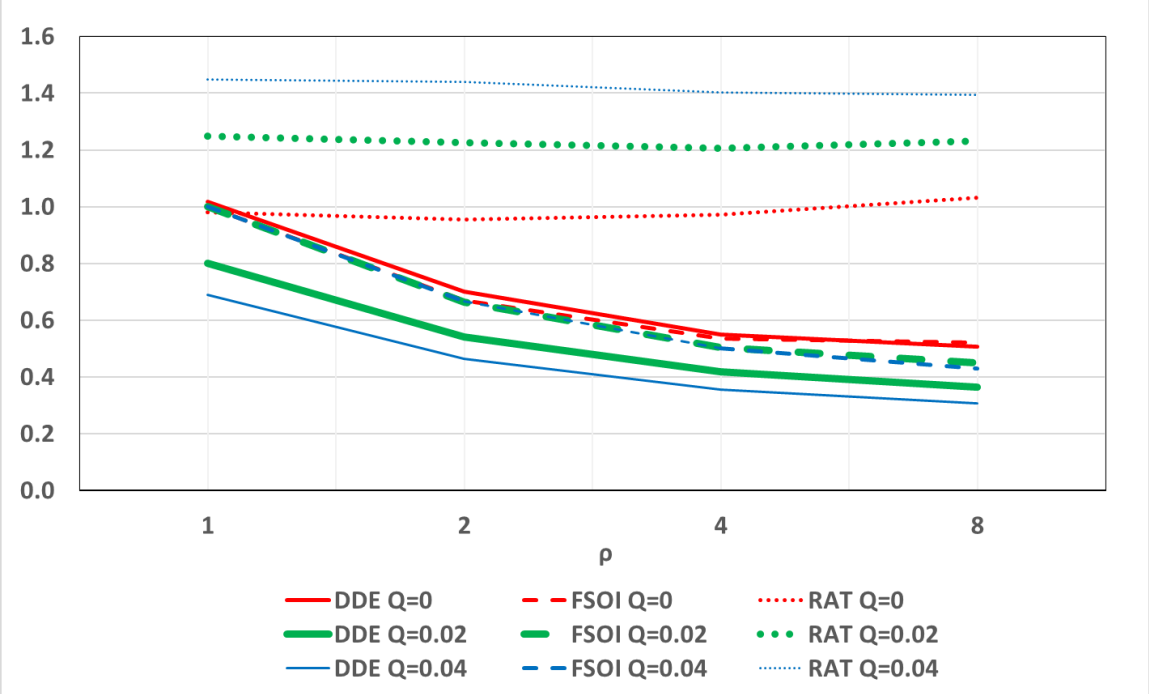

Figure 4. Results of $\%DDE$, $\%FSOI$ and RAT for Expt.4 ("Q=0"), Expt.5=baseline ("Q=0.02") and Expt.7("Q=0.04") illustrating the effect of different levels of model error in the presence of mixing.

Expt.8. Baseline, with correlated error in $\mathbf{Z}$

Results are shown in Table C.4, where they are compared with the baseline.

Compared with Expt.1 and 2, there is now some effect of introducing correlations in $\mathbf{Z}$ but it is small in both $\%DDE$ and $\%FSOI$. (Note that we have modelled here both the effects of correlated errors in $\mathbf{Z}$ and of optimal assimilation under these conditions.)

Expt.9. Baseline, with increased rate of forecast error growth.

Results are shown in Table C.5 and Figure 5 (very thin / blue lines), along with those of Expt.5 and 10, to show the effects of changing the value of forecast error growth rate.

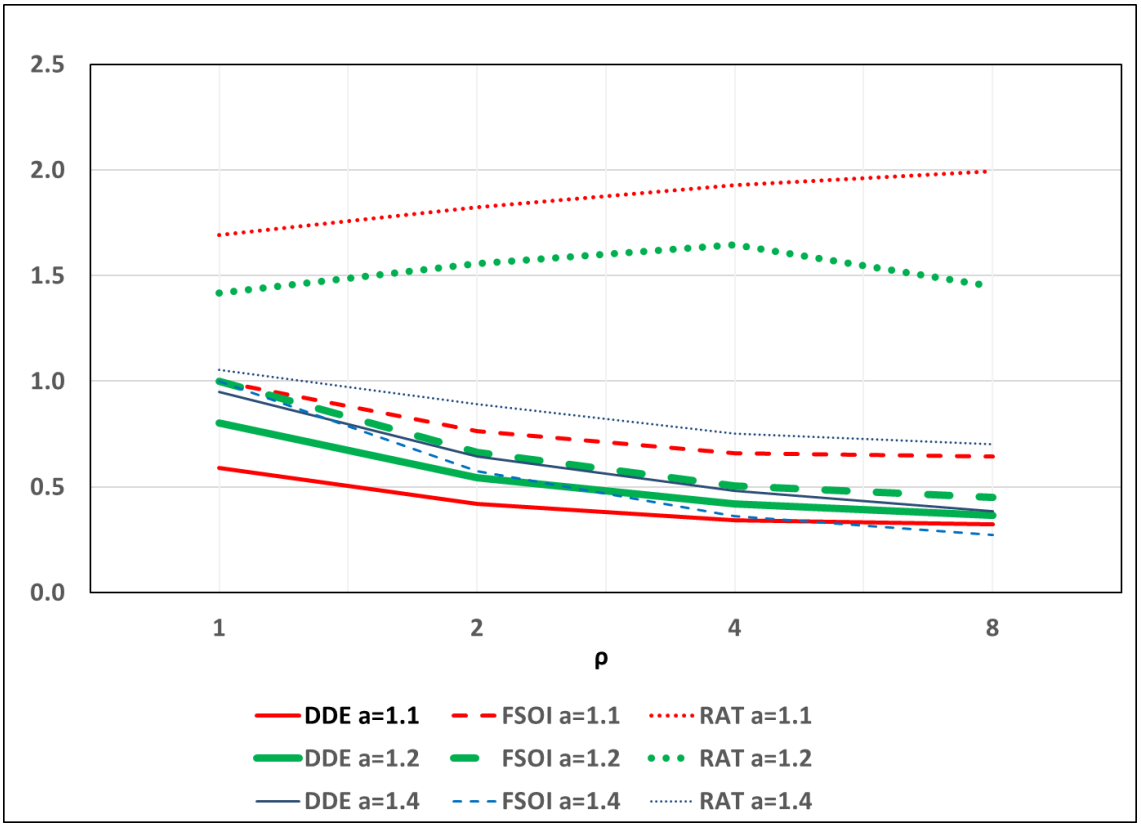

Figure 5. Results of $\%DDE$, $\%FSOI$ and RAT for Expt.10 ("a=1.1"), Expt.5=baseline ("a=1.2") and Expt.7("a=1.4") illustrating the effect of different levels of forecast error growth rate.

$\%FSOI$ is reduced for $\rho > 1$. $\%DDE$ is increased and hence RAT is reduced, to values <1 for all $\rho > 1$.

Expt.10. Baseline, with decreased rate of forecast error growth.

Results are also shown in Table C.6 and Figure 5 (thin/red lines), along with those of Expt.5 and 9. At this reduced rate of forecast error growth, the $\%FSOI$ is increased and the $\%DDE$ is decreased, leading to substantially increased RAT for all ρ. Overall, Expts 9 and 10 show that decreasing the rate of forecast error growth substantially increases the value of RAT.

Expt.11. Baseline, but with reduction in $\mathbf{Z}$ switched from variable 1 to variable 2.

Results are shown in Table C.6 and Figure 6 (thin/red lines), where they are compared with the baseline.

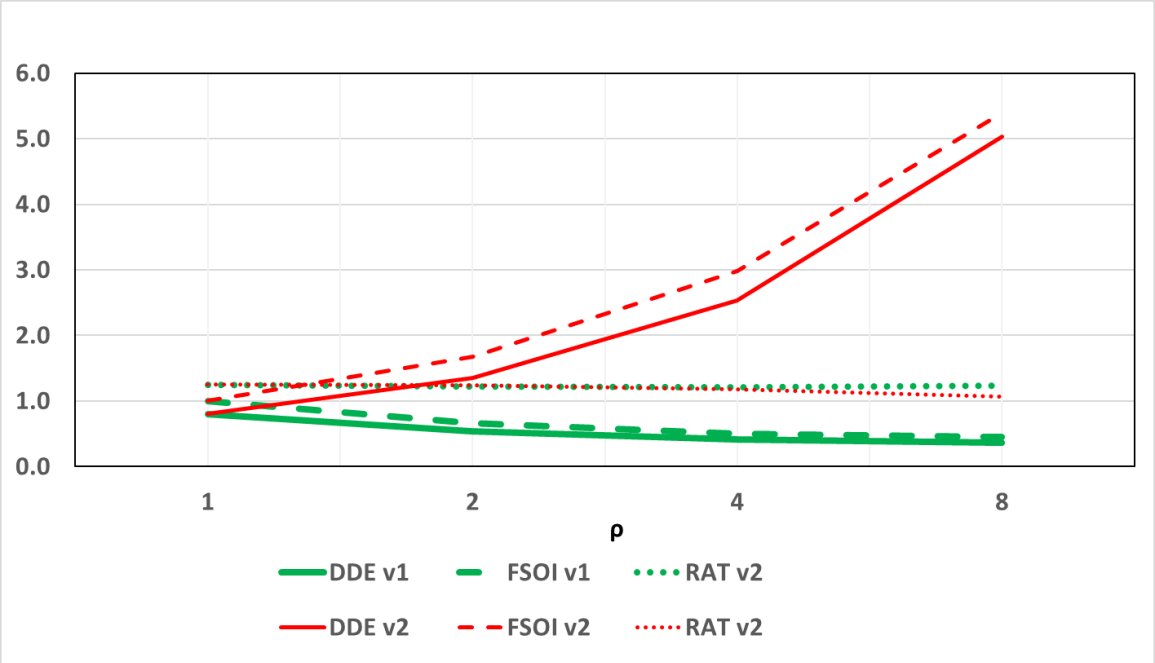

Figure 6. Results of $\%DDE$ for Expt.11 and Expt.5=baseline, illustrating the effect of denying observation information from the poorly observed variable (v2) and well observed variable (v1) respectively.

Results for $\rho = 1$ are the same, because the observation and forecast errors in both variables are equal here. However now, as ρ increases, both $\%DDE$ and $\%FSOI$ increase, and in such a way that RAT is almost unchanged at around 1.2.

Expt.12. Baseline, but with equal reduction in $\mathbf{Z}$ between the two variables.

Results are also shown in Table C.6. This experiment is effectively intermediate in $\mathbf{Z}$ reduction between Expts 5 and 11, and the results are broadly intermediate.

Expt.13. Baseline, but with reductions in $\mathbf{Z}$ in the two variables proportionate to the value of $\mathbf{Z}$ (see section 3.2).

Results are also shown in Table C.6. $\%FSOI$ is now identically 1 for all values of ρ. Values of $\%DDE$ are also comparatively insensitive to ρ, and hence values of RAT are similarly insensitive. The difference between $\%FSOI$ and $\%DDE$ is caused largely by model error.

The most striking result from Expt.5, 10, 11 and 12 is that it is only when observational information is denied more from the poorly observed variable (in proportion to the well observed variable) that we find values of $\%DDE$ and $\%FSOI$ greater than 1. In fact, of all the experiments reported here, it is only in Expt.10 and 11 that this behaviour is found; it is only in these cases that the forecast impacts are greater, as a percentage, than the percentage of observational information denied. In other words, these are the only conditions under which the system is not resilient to the denial of observations. This behaviour is discussed further in Section 4.

Expt.14. Baseline, with verification error included for DDE

Results are also shown in Table C.7 and Figure 7 (thin/red lines).

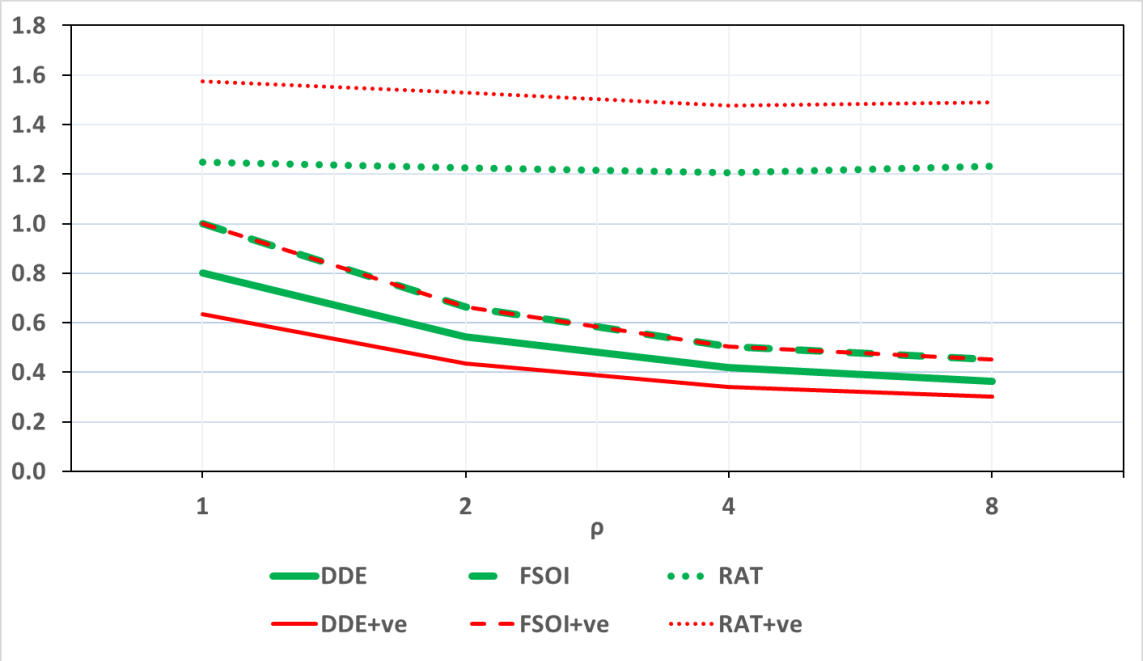

Figure 7. Results of $\%DDE$, $\%FSOI$ and RAT for Expt.5=baseline and Expt.14 ("+ve") illustrating the effect of allowing for verification error in the DDE.

$\%DDE$ is substantially reduced compared with the baseline. $\%FSOI$ remains the same (because the computation has not change – see section 2.8), and hence RAT is increased substantially. The results suggest that verification error makes a significant contribution to the values of RAT found in real-world systems. The $\%DDE$ results can be understood by recalling that the forecast error variance doubles in ~12 hours, and so the error variance in the 24h forecast is ~4 times the analysis error variance. Therefore ~20% of the apparent "error" in the 24h forecast results from error in the verification data.

3.3.2 Comparisons of $\%FSOI$ and $\%DDE^e$

Results are given in the tables in Appendix C.

For Expt.1 ("simple": no mixing, no model error, no correlation in $\mathbf{Z}$), when $\rho = 1$, $\%DDE^e = \%DDE \simeq \%FSOI = 1$, and when $\rho > 1$, $\%DDE^e \simeq \%FSOI$. Therefore, in this case, the nature of the DDE metric is dominant, recalling that both $\%FSOI$ and $\%DDE^e$ use an energy-based norm whereas $\%DDE$ does not. The results are the same for Expt.2 ("simple", but with some correlation in $\mathbf{Z}$).

For Expt.3 ("simple", but with model error), when $\rho = 1$, $\%DDE^e = \%DDE$, and when $\rho > 1$, $\%DDE^e < \%FSOI < \%DDE$. Again, the nature of the metric is dominant.

When mixing is introduced, for Expt.4 to 10, we obtain this result: when $\rho = 1$, $\%DDE^e = \%DDE$, and generally, when $\rho > 1$, $\%DDE^e < \%DDE < \%FSOI$. Also, the ratio between $\%FSOI$ and $\%DDE^e$ is always increased relative to that between $\%FSOI$ and $\%DDE$. Therefore, the ratio changes but its form does not; where it is greater than 1 in $\%DDE$ it remains greater than 1 in $\%DDE^e$.

Similarly, with Expt.11-13 in which the distribution of data denied between variables is changed, the behaviour of $\%DDE^e$ follows closely the behaviour of $\%DDE$, and so the ratio with $\%FSOI$ is also similar.

In Expt.14, the impact on $\%DDE^e$ of including verification error is similar to that on $\%DDE$, increasing the ratio with $\%FSOI$.

In summary, the change to an energy metric for DDE only changes the form of the ratio when there is no mixing; in other experiments it tends to increase the ratio if already greater than one or reduce it if already less than one. It does not therefore change any of the main conclusions about the factors affecting the ratio between the FSOI and DDE metrics.

4. Summary and discussion

The main finding from these experiments is the importance of mixing of forecast error between variables (i.e. non-diagonal $\mathbf{M}$) to the form of the results. The simple experiments with no mixing are useful for illustrating some interesting properties of the system, but the results give a ratio of FSOI impact expressed as a percentage ($\%FSOI$) to the DDE impact expressed as a percentage ($\%DDE$) opposite to that found in real-world NWP systems (see Fig.1); mixing of information between variables is crucially important for obtaining a ratio greater than 1.

The inclusion of model error is found to be important in reducing $\%DDE$ scores. This can be understood by considering model error as one way in which the NWP system progressively "forgets" observational information. FSOI is comparatively little affected, because the sum of each $\%FSOI$ score is normalised to 100%. Consequently, model error also acts to increase the $\%FSOI$:$\%DDE$ ratio.

Results are sensitive to the way in which observational "information" is distributed, via the observation precision mapped into state space, $\mathbf{Z}$, between the two variables. If both variables are observed equally, the $\%FSOI$ is equal to percentage contribution to $trace[\mathbf{Z}]$, whatever the values of other system parameters. So, in these experiments, because $trace[\mathbf{Z}]$ is constant (= 2) and its reduction is constant (= 0.02), $\%FSOI$ remains constant at 1 when both variables are observed equally. Also, in this case, the $\%DDE$ scores are $\simeq$1 only when the model error is zero; otherwise they are reduced.

If one variable is observed better than the other, then reduction of information in the well-observed variable causes the $\%FSOI$ score and the $\%DDE$ score to be reduced (Fig.6). The opposite effect is found when information is removed from the variable that is less well observed. This suggests that most real-world DDEs are exploring removal of observations from relatively well observed variables or that other effects are dominant. In real-world NWP systems, any given observing system will observe some variables well and others less well and many not at all. It is therefore very plausible that, when one of these systems is removed in a DDE, it will preferentially affect the better observed variables.

The experiments show the relative unimportance of correlations in the observations, as represented by the matrix $\mathbf{Z}$, which is the precision matrix of the observations mapped into state space. There are two caveats here: the correlations in $\mathbf{Z}$ are not simply related to the correlations in $\mathbf{R}$; they are also affected by the $\mathbf{H}$ and, in this study, we have not explored hypothetical vales of $\mathbf{H}$. Also, $trace[\mathbf{Z}]$ has been held constant, which

is not equivalent to holding constant the information content of the observations; the two are related through the determinant of $\mathbf{Z}$.

Accounting for verification error has a large effect on $\%DDE$ but no effect on $\%FSOI$ (assuming the errors in the forecasts used are uncorrelated with those in the verifying analysis), and hence a large effect on the ratio between them (increasing it). Therefore, verification error provides another reason why $\%DDE$ scores are significantly lower than $\%FSOI$ scores in real-world experiments, and also a reason why DDEs will underestimate the impact of observations, particularly for short-range forecasts. The effects on $\%FSOI$ scores of correlations because forecast errors and those in the verifying analysis, caused by systematic errors in the model and/or the observations (as highlighted by Privé *et al.* 2020) are not explored in this study but are also expected to be important in real-world systems.

As explained in section 2.4, FSOI results represent contributions to forecast error expressed with an energy norm and DDE results are often expressed as fractional changes in rms forecast error. This difference alone can account for a factor of 2 between FSOI and DDE results.

$\%DDE$ measures percentage change in forecast error averaged over a range of variables. We have additionally computed results in terms of $\%DDE^e$, an energy metric analogous to that used in the FSOI method. In the presence of mixing and compared to $\%DDE$, it is found that $\%DDE^e$ tends to enhance the differences relative to $\%FSOI$ but does not change the form of the results.

If we consider the results for $\%DDE$ or $\%DDE^e$ or $\%FSOI$ over all experiments, it is only when observational information is denied disproportionately from the poorly observed variable that we find values significantly greater than 1 for any of these forecast impact metrics. Note that the percentage of observational information denied in these experiments, as measured by the percentage change in $trace[\mathbf{Z}]$, is always 1. This suggests a definition for the **resilience** of an NWP system to the removal of observed information: a system is resilient if the percentage impact on forecast accuracy, as measured by a DDE, is less that the percentage of observational information removed, in terms of $trace[\mathbf{Z}]$, and it is **not resilient** otherwise.

Therefore, using this definition, the system will be resilient to the removal of observational information when it is denied from well observed variables, and it will not be resilient when it is denied from poorly observed variables. This is an interesting result; it contradicts the commonly held view that resilience results from a rebalancing of the analysis, i.e. from the analysis weights changing to generate this effect. Of course, when observations are denied, the weights given to the remaining observations will change, but this alone will not give rise to a resilient system.

It can also be seen that verification error introduces the appearance of resilience, but this can be misleading. Similarly, model error increases forecast error and leads to reduced DDE impacts, but this is really just a decreasing sensitivity of the forecast to observations in general, and it cannot usefully be called “resilience”.

Other potentially important effects have not been explored in these experiments. Firstly, the system is static; $\mathbf{M}$, $\mathbf{B}$ and $\mathbf{A}$ do not evolve with time. Secondly, the system parameters of forecast error growth, error mixing and model error are set equal for the (two) model variables; in real-world systems they would differ between variables. For example, errors in some variables and on some scales would grow more rapidly initially and may then saturate. Finally, this study has modelled only optimal systems, and all

real-world systems are suboptimal to some extent. Suboptimal systems will be considered in Part II.

As discussed earlier, it is often found that, although DDEs and FSOI give different results, the way in which they rank the impacts of different observing system are usually similar. This is consistent with the finding that, for a given set of system parameters and with the approximations and simplifications inherent in this study, the impact assessed by each metric will be proportional to the change in $\mathbf{Z}$ represented by the observation type denied in a DDE or measured by FSOI, but that the constant of proportionality will differ between DDE and FSOI as indicated by the results reported here.

An obvious weakness of this study is its use of only two variables, which limits the range of properties that can be explored. However, it is expected that the major characteristics described above will also be found in systems of more than two variables.

5. Conclusions and suggestions for further work

We have developed a model for analysis and forecast error covariances of a very simple, two-variable system. Using this system, we have computed observation impacts for metrics analogous to those used in DDEs and with the FSOI method. We have compared the results of these $\%DDE$ and $\%FSOI$ scores for a range of system parameters and, in particular, we have considered which sets of system parameters give rise to the usual result found in real-world observing system impact studies, namely that FSOI impacts are considerably greater that DDE impacts. We find that the mixing of information (and error) between different model variables is crucial for obtaining this real-world result, and that increasing model error and decreasing forecast error growth rate usually act to enhance this result. The effects of errors in the verification data also act to decrease $\%DDE$ scores and hence to increase the ratio between $\%FSOI$ and $\%DDE$ scores.

The way in which the observational information projects on to the different state variables is also important; in practice some state variables will be observed better than others, and the results of this study are consistent with DDEs normally denying information from the better observed variables.

Defining a “resilient” system as one in which the percentage impact on forecast accuracy of removing observations is less that the percentage of observational information denied, we find that the system will be resilient if observations are denied primarily from well observed variables, and not resilient if denied from poorly observed variables. This represents a revised understanding of the reason for NWP systems being resilient to the removal of observations.

Only optimal systems have been modelled in this study. Real-world systems are always suboptimal to some extent, and these effects will be considered in Part II.

It would be interesting to develop this approach further: to explore a system of ~100 state variables with the same equation set. This would permit the study of aspects that cannot be studied adequately with two variables, such as the spectrum of observed information in projecting on to different variables and differential rates of mixing, of forecast error growth and of model error between different variables. It would also be interesting to apply the same approach to a toy OSSE using a nonlinear forecast model and (for example) an ensemble Kalman filter, rather than through static error covariances.

The results of this study provide some insight into the factors that drive DDE and FSOI scores and also the resilience of NWP systems to the removal of observations. It is therefore hoped that these results will guide the critical examination of DDE and FSOI results from real-world NWP experiments.

Acknowledgements

I am grateful to Alison Fowler (University of Reading), Nancy Baker (NRL) and my Met Office colleagues, Neill Bowler, Chawn Harlow, Heather Lawrence, Stefano Migliorini and Chiara Piccolo for their helpful comments. I also acknowledge contributions from Peter Weston (now ECMWF) in the early stages of this work.

Appendix A. Derivation of FSOI metric

Langland and Baker (2004) define the FSOI metric (using our notation):

$$\delta e_f^g =< \left(\mathbf{y}^o - H[\mathbf{x}^b]\right), \frac{\partial J_f^g}{\partial \mathbf{y}^o} > , \tag{A.1}$$

which they derive from the equation

$$\delta e_f^g \simeq \Delta e_f^g = e_f - e_g , \tag{A.2}$$

where $2J_f = e_f =< \left(\mathbf{x}^f - \mathbf{x}^t\right), \mathbf{C}\left(\mathbf{x}^f - \mathbf{x}^t\right) >$,
$2J_g = e_g =< (\mathbf{x}^g - \mathbf{x}^t), \mathbf{C}(\mathbf{x}^g - \mathbf{x}^t) >$,
$\mathbf{C}$ defines an energy norm,
f and g denote, respectively, the forecast from the analysis and the forecast from the background for this analysis.

$\frac{\partial J_f^g}{\partial \mathbf{y}^o}$ is the observation sensitivity obtained using both trajectories, f and g.

With a unit norm, $\mathbf{C} = \mathbf{I}$, and recalling that $\boldsymbol{\varepsilon}^f = \mathbf{x}^f - \mathbf{x}^t$, $\boldsymbol{\varepsilon}^f = \mathbf{x}^g - \mathbf{x}^t$,

$$\delta e_f^g \simeq< {\boldsymbol{\varepsilon}^f}^T \boldsymbol{\varepsilon}^f - {\boldsymbol{\varepsilon}^g}^T \boldsymbol{\varepsilon}^g > . \tag{A.3}$$

This is for pairs of forecasts f and g which, we recall, are represented in this study (and by Langland and Baker) by 24h and 30h forecasts respectively (see section 2.5).

Taking averages over an ensemble of K forecasts,

$$\begin{aligned} \delta e = \frac{1}{K}\sum_{k=1}^{K} \delta e_{f\ k}^{g} &\simeq \frac{1}{K}\sum_{k=1}^{K} \left({\boldsymbol{\varepsilon}^f}^T \boldsymbol{\varepsilon}^f - {\boldsymbol{\varepsilon}^g}^T \boldsymbol{\varepsilon}^g\right)_k \\ &= -trace[\mathbf{P}_{n+1} - \mathbf{P}_n] , \end{aligned} \tag{A.4}$$

where n=4 for a 24h forecast.

This is the same as equations (2.5.8)-(2.5.10):

$$\begin{aligned} \delta e = \sum_j \delta e_j = trace[\mathbf{P}_{n+1} - \mathbf{P}_n] &= trace\left[\mathbf{M}^n\left(\mathbf{A}\sum_j \mathbf{Z}_j\, \mathbf{B}\right)\mathbf{M}^{n\boldsymbol{T}}\right] \\ &= \sum_j trace[\mathbf{M}^n\left(\mathbf{A}\mathbf{Z}_j\mathbf{B}\right)\mathbf{M}^{n\boldsymbol{T}}] . \end{aligned} \tag{A.5}$$

except with the opposite sign, because Langland and Baker choose to define a positive impact of an observation as a negative FSOI score.

Note that the vector product in (A.1) is effectively a summation over all observations j, which can be grouped instead as a summation over observation subsets j. (A.5) is also a summation over observation subsets j, where the contribution from the jth subset can be similarly written.

$$\delta e_j = trace[\mathbf{M}^n(\mathbf{A}\mathbf{Z}_j\mathbf{B})\mathbf{M}^{n\boldsymbol{T}}] \ . \tag{A.6}$$

When $\mathbf{C} \neq \mathbf{I}$, equations (A.3)-(A.6) will take modified forms. In fact, $\mathbf{C} \neq \mathbf{I}$ will be necessary when the state vector has mixed units (i.e. contains wind and mass and perhaps moist variables). The exact choice of $\mathbf{C}$, particularly for the moist terms, is somewhat arbitrary and hence also is the value of the FSOI energy metric. However, with the very simple model adopted for this study, this problem is avoided.

Appendix B. The relationship between FSOI and DFS metrics

The degrees of freedom for signal (DFS) metric can be shown (e.g. Rodgers, 2000, p.31, eq.2.56), using the notation of this paper, to be given by

$$DFS = trace[(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}] \ . \tag{B.1}$$

Therefore, using equations (2.1.5) and (2.1.6),

$$DFS = trace[\mathbf{A}\mathbf{Z}] \ . \tag{B.2}$$

Disaggregating $\mathbf{Z}$, as discussed in section 2.1, we obtain an expression for the contribution to the DFS of the jth observation type:

$$DFS_j = trace[\mathbf{A}\boldsymbol{Z_j}] \ . \tag{B.3}$$

Compare this with the equivalent expression for FSOI, i.e. eq.(A.6). It can be seen that the two metrics are closely related but that they differ as follows: DFS measures the impact on analysis whereas FSOI measures the impact of forecast (i.e. the analysis error covariance propagated forward in time via the model matrix $\mathbf{M}^n$), and the FSOI metric contains an additional multiplier $\mathbf{B}$. Hence FSOI is an energy metric (usually expressed in J/Kg) whereas DFS is dimensionless. However, it is also usual to present both as fractional scores: $\delta e_j/\delta e$ and DFS_j/DFS.

The two metrics also have another important difference in their application: DFS assumes an optimal system – it assesses the impact that the observations would have if used optimally. FSOI measures the impact of observations as used in a (generally) suboptimal system. However, the FSOI measure is derived using elements of optimal estimation theory and can be misleading within a suboptimal system, as will be discussed in Part II.

Appendix C. Tables of results

Expt	a	m	c	Q_{ii}	ρ	$\%DDE$	$\%FSOI$	RAT	$\%DDE^e$	RAT^e
1	1.2	0	0	0	1	1.020	1.000	0.980	1.020	0.980
					2	0.761	0.500	0.657	0.508	0.984
					4	0.633	0.250	0.395	0.253	0.988
					8	0.569	0.125	0.220	0.126	0.992
					16	0.537	0.063	0.117	0.063	1.000
					32	0.521	0.031	0.060	0.031	1.000
					64	0.513	0.016	0.031	0.016	1.000
					128	0.509	0.008	0.016	0.008	1.000
3	1.2	0	0	0.02	1	0.855	1.000	1.170	0.855	1.170
					2	0.609	0.542	0.890	0.436	1.243
					4	0.490	0.298	0.608	0.228	1.307
					8	0.431	0.161	0.374	0.120	1.342
4	1.2	0.05	0	0	1	1.019	1.000	0.981	1.019	0.981
					2	0.702	0.670	0.954	0.585	1.145
					4	0.551	0.535	0.971	0.406	1.318
					8	0.506	0.522	1.032	0.384	1.359
					16	0.547	0.588	1.075	0.463	1.270
					32	0.648	0.694	1.071	0.599	1.159
					64	0.767	0.801	1.044	0.739	1.084
					128	0.864	0.884	1.023	0.850	1.040

Table C.1. Results for "simple" experiments (with no mixing) and for an experiment with mixing.

Expt	a	m	C	Q_{ii}	ρ	$\%DDE$	$\%FSOI$	RAT	$\%DDE^e$	RAT^e
3	1.2	0	0	0.02	1	0.855	1.000	1.170	0.855	1.170
					2	0.609	0.542	0.890	0.436	1.243
					4	0.490	0.298	0.608	0.228	1.307
					8	0.431	0.161	0.374	0.120	1.342
5	**1.2**	**0.05**	**0**	**0.02**	**1**	**0.801**	**1.000**	**1.248**	**0.801**	**1.248**
					2	**0.542**	**0.665**	**1.227**	**0.454**	**1.465**
					4	**0.418**	**0.504**	**1.206**	**0.293**	**1.720**
					8	**0.365**	**0.450**	**1.233**	**0.232**	**1.940**
					16	**0.359**	**0.476**	**1.326**	**0.238**	**2.000**
					32	**0.396**	**0.563**	**1.422**	**0.292**	**1.928**
					64	**0.460**	**0.682**	**1.483**	**0.373**	**1.828**
					128	**0.529**	**0.796**	**1.505**	**0.454**	**1.753**
6	1.2	0.10	0	0.02	1	0.797	1.000	1.255	0.797	1.255
					2	0.603	0.854	1.416	0.589	1.450
					4	0.522	0.813	1.557	0.501	1.623
					8	0.507	0.834	1.645	0.483	1.727

Table C.2. Results for experiments with different values of forecast error mixing. The baseline experiment is highlighted in bold.

Expt	a	m	c	Q_{ii}	ρ	$\%DDE$	$\%FSOI$	RAT	$\%DDE^e$	RAT^e
4	1.2	0.05	0	0	1	1.019	1.000	0.981	1.019	0.981
					2	0.702	0.670	0.954	0.585	1.145
					4	0.551	0.535	0.971	0.406	1.318
					8	0.506	0.522	1.032	0.384	1.359
5	**1.2**	**0.05**	**0**	**0.02**	**1**	**0.801**	**1.000**	**1.248**	**0.801**	**1.248**
					2	**0.542**	**0.665**	**1.227**	**0.454**	**1.465**
					4	**0.418**	**0.504**	**1.206**	**0.293**	**1.720**
					8	**0.365**	**0.450**	**1.233**	**0.232**	**1.940**
7	1.2	0.05	0	0.04	1	0.690	1.000	1.449	0.690	1.449
					2	0.464	0.668	1.440	0.394	1.695
					4	0.357	0.501	1.403	0.253	1.980
					8	0.308	0.429	1.393	0.188	2.282

Table C.3. Results for experiments with mixing, but with different values of model error.

Expt	a	m	c	Q_{ii}	ρ	$\%DDE$	$\%FSOI$	RAT	$\%DDE^e$	RAT^e
5	**1.2**	**0.05**	**0**	**0.02**	**1**	**0.801**	**1.000**	**1.248**	**0.801**	**1.248**
					2	**0.542**	**0.665**	**1.227**	**0.454**	**1.465**
					4	**0.418**	**0.504**	**1.206**	**0.293**	**1.720**
					8	**0.365**	**0.450**	**1.233**	**0.232**	**1.940**
8	1.2	0.05	0.1	0.02	1	0.813	1.000	1.230	0.813	1.230
					2	0.553	0.683	1.235	0.471	1.450
					4	0.430	0.534	1.242	0.311	1.717
					8	0.377	0.489	1.297	0.248	1.972

Table C.4. Results for experiments with mixing, but with different values of correlation in $\mathbf{Z}$.

Expt	a	m	c	Q_{ii}	ρ	$\%DDE$	$\%FSOI$	RAT	$\%DDE^e$	RAT^e
10	1.1	0.05	0	0.02	1	0.591	1.000	1.692	0.591	1.692
					2	0.418	0.763	1.825	0.380	2.008
					4	0.343	0.661	1.927	0.285	2.319
					8	0.322	0.642	1.994	0.251	2.558
5	**1.2**	**0.05**	**0**	**0.02**	**1**	**0.801**	**1.000**	**1.248**	**0.801**	**1.248**
					2	**0.542**	**0.665**	**1.227**	**0.454**	**1.465**
					4	**0.418**	**0.504**	**1.206**	**0.293**	**1.720**
					8	**0.365**	**0.450**	**1.233**	**0.232**	**1.940**
9	1.4	0.05	0	0.02	1	0.948	1.000	1.055	0.948	1.055
					2	0.645	0.574	0.890	0.492	1.167
					4	0.483	0.363	0.752	0.273	1.330
					8	0.386	0.271	0.702	0.178	1.522

Table C.5. Results for experiments with different values of forecast error growth rate.

Expt	a	m	c	Q_{ii}	ρ	$\%DDE$	$\%FSOI$	RAT	$\%DDE^e$	RAT^e
5	**1.2**	**0.05**	**0**	**0.02**	**1**	**0.801**	**1.000**	**1.248**	**0.801**	**1.248**
					2	**0.542**	**0.665**	**1.227**	**0.454**	**1.465**
					4	**0.418**	**0.504**	**1.206**	**0.293**	**1.720**
					8	**0.365**	**0.450**	**1.233**	**0.232**	**1.940**
11	1.2	0.05	0	0.02	1	0.801	1.000	1.248	0.801	1.248
					2	1.352	1.671	1.236	1.545	1.082
					4	2.539	2.982	1.174	3.131	0.952
					8	5.038	5.396	1.071	6.327	0.853
12	1.2	0.05	0	0.02	1	0.794	1.000	1.259	0.794	1.259
					2	0.936	1.168	1.248	0.988	1.182
					4	1.449	1.743	1.203	1.676	1.040
					8	2.604	2.924	1.123	3.157	0.926
13	1.2	0.05	0	0.02	1	0.794	1.000	1.259	0.794	1.259
					2	0.802	1.000	1.247	0.807	1.239
					4	0.824	1.000	1.214	0.838	1.193
					8	0.847	1.000	1.181	0.862	1.160

Table C.6. Results for experiments with different distribution of observation reduction between variables.

Expt	a	m	c	Q_{ii}	P	$\%DDE$	$\%FSOI$	RAT	$\%DDE^e$	RAT^e
5	**1.2**	**0.05**	**0**	**0.02**	**1**	**0.801**	**1.000**	**1.248**	**0.801**	**1.248**
					2	**0.542**	**0.665**	**1.227**	**0.454**	**1.465**
					4	**0.418**	**0.504**	**1.206**	**0.293**	**1.720**
					8	**0.365**	**0.450**	**1.233**	**0.232**	**1.940**
14	1.2	0.05	0	0.02	1	0.635	1.000	1.575	0.635	1.575
					2	0.435	0.665	1.529	0.359	1.852
					4	0.341	0.504	1.478	0.231	2.182
					8	0.302	0.450	1.490	0.181	2.486

Table C.7. Results for experiments with verification error included in the DDE calculation. Results of baseline Expt.5 are included for reference.

Acronyms

AMV	atmospheric motion vector
DA	data assimilation
ECMWF	European Centre for Medium-range Weather Forecasts
IR	infra-red
DDE	data denial experiment
DFS	degrees of freedom for signal
FSOI	forecast-sensitivity-to-observation-impact
GNSS	global navigation satellite system
MW	microwave
NRL	US Naval Research Laboratory
NWP	numerical weather prediction
OSE	observing system experiment
OSSE	observing system simulation experiment

rms	root mean square
WMO	World Meteorological Organisation

References

Bormann N, Lawrence H, Farnan J. 2019. *Global observing system experiments in the ECMWF assimilation system*. ECMWF Technical Memorandum 839. ECWMF, Reading, UK. Available at: https://www.ecmwf.int/en/elibrary/18859-global-observing-system-experiments-ecmwf-assimilation-system

Bouttier F, Kelly G. 2001. Observing-system experiments in the ECMWF 4D-Var data assimilation system. *Q J R Meteorol Soc*, **127**: 1469-1488. https://doi.org/10.1002/qj.49712757419

Candy B, Cotton J, Eyre J. 2021. *Recent results of observation data denial experiments*. Met Office Forecasting Research Technical Report 641. Met Office, Exeter, UK. Available at: https://www.metoffice.gov.uk/research/library-and-archive/publications/science/weather-science-technical-reports

Cardinali C. 2009. Monitoring the observation impact on the short-range forecast. *Q J R Meteorol Soc*, **135**: 239-250. https://doi.org/10.1002/qj.366

Daescu D. 2009. On the deterministic observation impact guidance: a geometrical perspective. *Monthly Weather Review*, **137**: 3567-3574. https://doi.org/10.1175/2009MWR2954.1

Errico RM, Yang R, Privé NC, Tai K-S, Todling R, Sienkiewicz ME, Guo J. 2013. Development and validation of observing-system simulation experiments at NASA's Global Modeling and Assimilation Office. *Q J R Meteorol Soc*, **139**: 1162-1178. https://doi.org/10.1002/qj.2027

Eyre J. 2018. *Observing system studies proposed for GNSS-R ocean surface wind data.* Met Office Forecasting Research Technical Report 629. Met Office, Exeter, UK. Available at https://www.metoffice.gov.uk/research/library-and-archive/publications/science/weather-science-technical-reports

Geer AJ. 2016. Significance of changes in medium-range forecast scores. *Tellus* A, **68**:1. https://doi.org/10.3402/tellusa.v68.30229

Gelaro R, Zhu Y, Errico R. 2007. Examination of various-order adjoint-based approximations of observation impact. Meteorologische Zeitschrift, **16**: 685-692. https://doi.org/10.1127/0941-2948/2007/0248

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

Ide K, Courtier P, Ghil M and Lorenc AC. 1997. Unified notation for data assimilation: operational, sequential and variational. *J Meteorol Soc Japan*, **75**: 181-189. https://doi.org/10.2151/jmsj1965.75.1B_181

Langland RH, Baker NL. 2004. Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. *Tellus* A, **56**, 189-201. https://doi.org/10.1111/j.1600-0870.2004.00056.x

Lorenc AC, Marriott R. 2014. Forecast sensitivity to observations in the Met Office global numerical weather prediction system. *Q J R Meteorol Soc*, **140**: 209–224. https://doi.org/10.1002/qj.2122

Privé NC, Errico RM, Todling R, El Akkraoui A. 2020. Evaluation of adjoint-based observation impacts as a function of forecast length using an Observing System Simulation Experiment. *Q J R Meteorol Soc*, **147**: 121-138. https://doi.org/10.1002/qj.3909

Raynaud L, Berre L, Desroziers G. 2012. Accounting for model error in the Météo-France ensemble data assimilation system. *Q J R Meteorol Soc*, **138**: 249-262. https://doi.org/10.1002/qj.906

Rodgers CD. 1976. Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation. *Rev. Geophys. Space Phys*. **14**: 609–624. https://doi.org/10.1029/RG014i004p00609

Rodgers CD. 2000. Inverse methods for atmospheric sounding: theory and practice. World Scientific. Series on Atmospheric, Oceanic and Planetary Physics: Volume 2. https://doi.org/10.1142/3171

Simmons AJ, Hollingsworth A. 2002. Some aspects of the improvement in the skill of numerical weather prediction. *Q J R Meteorol Soc*, **128**: 647-677. https://doi.org/10.1256/003590002321042135

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

WMO. 2010. Regional Association V (South-West Pacific) Fifteenth session: Abridged final report with resolutions. World Meteorological Organization, WMO No-1056, Geneva, Switzerland. https://library.wmo.int/?lvl=notice_display&id=9087#.XvM-E-d7mUk

WMO. 2016. 6th Workshop on the Impact of Various Observing Systems on NWP. Shanghai, China; 10-13 May 2016. https://www.wmo.int/pages/prog/www/WIGOS-WIS/reports/6NWP_Shanghai2016/WMO6-Impact-workshop_Shanghai-May2016.html