

Data Assimilation for Coupled Modeling Systems

Milija Županski

Abstract Coupled numerical models address the interaction between processes in the atmosphere, ocean, land surface, biosphere, chemistry, cryosphere, and hydrology. Including interaction between such processes can potentially extend the predictability and eventually help in reducing the uncertainty of the prediction. Coupled data assimilation is a branch of data assimilation that deals with coupled modeling systems. In this article the fundamentals of coupled data assimilation are described. Challenges of coupled data assimilation are addressed in terms of the variational and ensemble methods, with implications for hybrid data assimilation methods. Several illustrative examples of coupled data assimilation of a single observation with realistic regional coupled modeling systems are included as well.

1 Introduction

Use of coupled modeling systems is important for improving the predictability of coupled processes. The components of coupled modeling system may include atmosphere, ocean, land surface, biosphere, cryosphere, hydrology, chemistry, aerosol, as well as other processes of interest. The role of data assimilation for coupled modeling systems (e.g., coupled data assimilation) is also important as it has a capability to utilize the information from various observation types across the components to improve the prediction by coupled models.

Due to their increased complexity, coupled modeling systems are potentially more sensitive to the initial conditions than the individual components, and thus require special attention in data assimilation (Sakaguchi et al. 2012). In general one

M. Županski (✉)

Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO 80523-1375, USA

e-mail: milija.zupanski@colostate.edu

can distinguish three main regimes of coupled data assimilation systems: (a) uncoupled, (b) weakly coupled, and (c) strongly coupled.

Uncoupled data assimilation implies that each system is completely independent: the model and data assimilation for each individual component is done separately from each other. Using as an example atmosphere and hydrology, the atmospheric model forecast and data assimilation are used to produce the precipitation, which can be then used as forcing in the hydrology model with its own, independent data assimilation system.

Weakly coupled data assimilation allows some interaction between components. Typically, this means that data assimilation is performed completely separately for each component, but their initial conditions and parameters are fed back into the coupled model. In terms of sequential data assimilation, which consists of the forecast and the analysis steps, the weakly coupled data assimilation implies coupling in the forecast step, but no coupling in the analysis step. Most importantly, there is no use of the cross-component part of the forecast error covariance in data assimilation.

Strongly coupled data assimilation means that both the coupled forecast and coupled data assimilation are done in a single system that combines the information from all components. The cross-component part of the forecast error covariance is used in data assimilation, thus the forecast and observations of each component have a potential to influence all other components. In this paper we focus on strongly coupled data assimilation, as it offers a more profound, but also a more natural way of combining the information from models and observations.

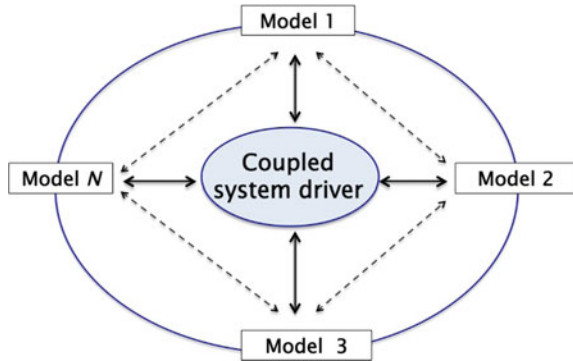
Although limited in number, the coupled data assimilation systems have already been developed (Sugiura et al. 2008; Rasmy et al. 2012; Han et al. 2013). They are mostly based on variational data assimilation, as most operational weather and climate data assimilation systems are variational, but there are also few ensemble coupled data assimilation systems being developed (Zhang et al. 2007; Tardif et al. 2014).

The paper is organized as follows. We begin by describing the motivation for coupled data assimilation in Sect. 2, followed by the challenges of coupled data assimilation in Sect. 3. An example of the two-component coupled data assimilation is presented in Sect. 4, followed by realistic applications of coupled data assimilation to investigate the coupled forecast error covariance and uncertainty interaction in Sect. 5. The summary and future development are presented in Sect. 6.

2 Motivation

Main motivation for developing coupled data assimilation is to improve the knowledge and/or improve the prediction of a coupled modeling system, in terms of the state and its uncertainty. The benefit of having a combined modeling system, as opposed to having the individual systems only, is that the model components are

Fig. 1 Schematic representation of an N -component coupled modeling system. The *full lines* represent the mandatory interaction between the model component and the driver, while the *dashed lines* represent the optional interaction between components



acting as “constraints” that restrict possible adjustments of the state to those that are “allowed” by other system components. In other words, it is anticipated that the analysis adjustment in the coupled data assimilation is in better physical agreement with the real world than the analysis adjustments in data assimilation of individual components.

A coupled modeling system is schematically represented in Fig. 1. While the interaction between the driver and the model components is always present, the interaction between model components is optional, related to the degree of dependence between relevant physical processes.

Another useful view of coupled data assimilation that supports its use can be described in terms of Shannon information theory (Shannon and Weaver 1949). Recall that the entropy is defined as

$$H(X) = - \int_x p(x) \log p(x) dx \quad (2.1)$$

where p denotes the probability density function (pdf) and x is a random variable. The entropy is closely associated with uncertainty, which is important for data assimilation. In coupled systems we are interested in the joint entropy of at least two processes, X_1 and X_2 ,

$$H(X_1, X_2) = - \int_{x_1} \int_{x_2} p(x_1, x_2) \log p(x_1, x_2) dx_1 dx_2 \quad (2.2)$$

where $p(x_1, x_2)$ is the joint pdf. An important quantification of the information exchange in coupled systems can be defined in terms of mutual information

$$I(X_1; X_2) = - \int_{x_1} \int_{x_2} p(x_1, x_2) \log \left[\frac{p(x_1, x_2)}{p(x_1)p(x_2)} \right] dx_1 dx_2 \quad (2.3)$$

where $p(x_1)$ and $p(x_2)$ are the marginal probability density functions. Mutual information measures the information shared by multi-component coupled system. It can be conveniently represented in terms of entropy (e.g., Cover and Thomas 2006)

$$I(X_1; X_2) = H(X_1) + H(X_2) - H(X_1, X_2) \quad (2.4)$$

Since $H(X_1, X_2) \leq H(X_1) + H(X_2)$, with equality sign true for independent processes X_1 and X_2 , the mutual information is non-negative with zero value corresponding to independent processes. Given that the motivation of developing coupled systems modeling is to improve the interaction between related physical processes, the component processes of the coupled system have well developed dependence and are intrinsically characterized by positive mutual information. This implies that the information from one component enhances the information about other components. This is quite important for data assimilation of coupled systems since it effectively reduces the dimension of the forecast error covariance and allows cross-component impacts during data assimilation, as will be shown in detail in the next sections.

When coupling exists, one can further partition it in terms of its strength: low intensity coupling referring to having only marginal interaction between the components, and high intensity coupling implying a profound interaction between the components. This qualitative distinction can be in principle quantified by calculating the mutual information of coupled modeling system, i.e. using their prior and joint probability density functions. Using as an example the Gaussian pdf, one can define the mutual information of multivariate Gaussian probability distribution (Silva and Quiroz 2003)

$$I(X_1, X_2) = -\frac{1}{2} \ln \frac{\det(\Sigma)}{\det(\Sigma_{11}) \det(\Sigma_{22})} \quad (2.5)$$

where Σ , Σ_{11} , and Σ_{22} are the joint, and the marginal covariances of variables X_1 and X_2 , respectively, and $\det$ denotes the determinant of a matrix. Since the joint covariance is

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_{22} \end{pmatrix} \quad (2.6)$$

and (e.g., Arellano-Valle et al. 2012)

$$\det(\Sigma) = \det(\Sigma_{11}) \det(\Sigma_{22}) \det(I - \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{12}^T) \quad (2.7)$$

one can define the mutual information as

$$I(X_1, X_2) = -\frac{1}{2} \ln [\det(I - \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{12}^T)] \quad (2.8)$$

The above equation implies that the mutual information can be calculated from the auto- and cross-components of the coupled system error covariance. It is also clear that, if the cross-component covariance does not exist in Eq. (2.6) ($\Sigma_{12} = 0$), the mutual information is zero.

Depending on the coupled system's complexity and dimensions, in some applications it may be possible to calculate the exact value of the mutual information according to (2.8), or an approximate value using only the diagonal elements of the matrices. By defining a mutual information threshold for low/high intensity of coupling, it may be possible to have a quantitative measure of the strength of coupling.

3 Challenges

There are numerous challenges of coupled data assimilation, some of which will be discussed here. They include: control variable, spatiotemporal scales, forecast error covariance, high state dimensions, and non-Gaussian errors.

3.1 Control Variable

In principle, control variable refers to modeling system parameters that impact its prediction. In particular, the prediction has to be sensitive to the choice of these parameters in the spatiotemporal limits of the data assimilation window. In practical data assimilation one can define control variable different from the state variable. For example, in some data assimilation systems it is assumed that the control variable includes stream function and velocity potential, while the state (model) variable includes the east-west and north-south wind components (e.g., u and v , respectively). However, for simplicity of presentation, we will assume that the control variable is a simple subset of the state variable, without requiring any transformation.

In geophysical modeling based on using partial differential equations the control variable typically includes (i) initial conditions, (ii) lateral boundary conditions (for regional systems), (iii) model empirical parameters, and (iv) model and observation operator biases (systematic errors). Therefore, for the k -th modeling component the control variable can be written as

$$x_k = \begin{pmatrix} x_k^{ic} & x_k^{bias} & x_k^{par} \end{pmatrix}^T \quad k = 1, \dots, N \quad (3.1)$$

where the superscripts *ic*, *bias*, and *par* refer to the initial conditions, bias, and empirical parameters, respectively, and N is the number of coupled system components. Note that this formulation includes both the single and the coupled systems, $N=1$ corresponding to the single system, and $N \geq 2$ corresponding to the coupled system. The general form of coupled data assimilation control variable is

$$x = (x_1 \ x_2 \ \cdots \ x_N)^T. \quad (3.2)$$

The coupled data assimilation challenge comes from the fact that practical data assimilation for individual components does not include all possible control parameters as in (3.2). Most often the control variable is defined as the initial conditions. However, there are data assimilation systems that are focused on adjusting empirical parameters (e.g., climate), empirical parameters and initial conditions (hydrology), model bias (carbon cycle, biosphere), rather than the initial conditions only. This creates a need for developing a coupled data assimilation system that is general enough so that it can assimilate any type of control variable. In practice, the control variable would include relevant control variables for each component. For example, if the coupled system includes the following components: atmospheric—denoted *a*, carbon transport—denoted *c*, land surface—denoted *l*, and hydrological—denoted *h*, one could define control variable as

$$x = (a^{ic} \ c^{bias} \ l^{ic} \ l^{par} \ h^{par})^T. \quad (3.3)$$

Such control variable may be a very efficient way of updating the given coupled system, but it does require a mathematical apparatus that can handle it in practice. If the available algorithmic structure of the coupled data assimilation does not allow the inclusion of all these variables, one may reduce the list to include only the feasible parameters. A better solution would be to update the coupled data assimilation algorithm by adding the capability to augment the control variable.

3.2 Forecast Error Covariance

The forecast error covariance is intrinsically related to the choice of control variable. The forecast error covariance represents the uncertainty of the control variable and thus includes all control variable components. Forecast error covariance has a fundamental role in data assimilation (e.g., Bannister 2008a, b) since all analysis adjustments are projected onto the subspace of this matrix. There are numerous papers addressing the role of error covariance in variational and ensemble data assimilation (Hollingsworth and Lonnberg 1986; Derber and Bouttier 1999; Buehner 2005; Bello Pereira and Berre 2006; Berre and Desrozier 2010).

Given a general coupled control variable (3.2), the forecast error covariance is a matrix

$$P_f = \begin{pmatrix} P_{11} & P_{12} & \cdots & P_{1N} \\ P_{12}^T & P_{22} & & P_{2N} \\ \vdots & & \ddots & \vdots \\ P_{1N}^T & P_{2N}^T & \cdots & P_{NN} \end{pmatrix} \quad (3.4)$$

where the subscripts refer to the coupled system component. Note that, according to (3.1), each of the inputs in (3.4) is also a matrix. For example

$$P_{ij} = \begin{pmatrix} P_{ij}^{ic,ic} & P_{ij}^{ic,bias} & P_{ij}^{ic,par} \\ (P_{ij}^{ic,bias})^T & P_{ij}^{bias,bias} & P_{ij}^{bias,par} \\ (P_{ij}^{ic,par})^T & (P_{ij}^{bias,par})^T & P_{ij}^{par,par} \end{pmatrix} \quad (3.5)$$

where i and j define the system components. It is clear that defining the elements of such complex matrix becomes challenging.

In variational data assimilation, the error covariance is modeled using some previous knowledge about correlations and variances. Knowing a priori the correlations between initial conditions, empirical model parameters, and model biases is very difficult. In ensemble data assimilation the correlations are obtained directly from the ensemble forecast, without the need for modeling the correlation functions. However, ensemble data assimilation requires at least some knowledge about the true correlations in order to know if low-dimensional approximation of the ensemble error covariance is acceptable.

Such challenges exist even in assimilation of individual components; in coupled systems this issue is magnified and thus made more difficult. The most challenging aspect of defining the error covariance in coupled data assimilation is to define the cross-component correlations, since the cross-variable correlations, and in particular the cross-component correlations are the least known.

3.3 High-Dimensional State Vector

The dimension of the state vector can considerably impact the design and performance of coupled data assimilation. Realistic atmospheric data assimilation systems have the control variable of the order of hundreds of millions. Adding chemistry component, for example, can further double the dimension of state vector since there may be several chemical species that are adjusted in data assimilation. Similarly, the aerosol can also add considerably to the total state dimension since it is defined at all model grid points, and there can be a dozen of aerosol species, such as dust, sea salt, black carbon, etc. Some other components, such as land surface, may

add only a smaller dimension to the augmented state since typically there are fewer soil layers than atmospheric layers.

In general, coupled data assimilation system has to be able to deal with such high dimensions. This may imply that new approximations may be necessary when extending the single-component assimilation to the coupled data assimilation. Practical data assimilation algorithms often include approximations such as dual-resolution (i.e. coarse resolution adjoint or ensemble and a high-resolution control), or a related incremental variational approach with additional linearization. In cases when one coupled component is more nonlinear than others, or there is a highly nonlinear observation operator of one of the components, it may be necessary to re-validate the assumptions in the context of the coupled system.

3.4 *Non-gaussian Errors*

It is well known that some control variables can have non-Gaussian errors, although typical data assimilation is based on Gaussian error assumption. If there is a small number of such variables, this may impact the results of assimilation only locally. However, this may become more relevant for coupled systems, as one component can have all non-Gaussian variables. For example, typical atmospheric chemistry and aerosol variables are strictly positive definite, which effectively excludes the Gaussian distribution with infinite tails. Therefore, in case of coupled atmosphere-chemistry-aerosol data assimilation, there may be a need to revisit the Gaussian error assumptions. In more extreme cases, it may be necessary to design a mixed Gaussian-non-Gaussian data assimilation system to accommodate all coupling components.

3.5 *Spatiotemporal Scales*

Another potential issue in coupled data assimilation is related to characteristic spatiotemporal scales of control variables. For example, atmospheric variables likely have shorter temporal scales than ocean variables, or compared to some chemistry variables such as stratospheric ozone. In the coupled system forecast, this can be resolved by using different time steps for different components of the system. For data assimilation, however, this may be more complicated as it should involve the coupled forecast error covariance, given its fundamental role in data assimilation.

In variational data assimilation the modeled error covariance has to include some knowledge of spatiotemporal scales in order to produce a dynamically balanced increments of both components. Since the knowledge of such cross-component correlations is very limited in general and possibly situation-dependent, the use of complete cross-component correlations in practical variational data assimilation is

not very likely. The use of ensemble error covariance may be simpler since the information about different spatiotemporal scales is already included in the coupled forecast models used in ensemble forecasting. However, since both of these approaches can produce only an approximate error covariance, there is still a need for verifying the validity of the coupled error covariance in terms of the spatiotemporal scales. The resolution may be more intuitive in data assimilation with four-dimensional error covariance, since than one can impose the temporal aspect of error covariance more directly.

4 Two-Component Coupled System Data Assimilation

The role of coupled forecast error covariance is now examined in the context of an idealized two-component coupled modeling system. Let define the two state components as x_1 and x_2 , thus forming a two dimensional state vector of the coupled system

$$x = (x_1 \quad x_2)^T \quad (4.1)$$

In this system each component is represented by a value at a single grid point. For the land-atmosphere system, for example, one grid point would be located in the atmosphere and one in the soil. For the atmosphere-chemistry coupled system, both components can represent the value at the same grid point. The forecast step of such system includes a coupled prediction model (denoted m)

$$x^n = m(x^{n-1}) \quad (4.2)$$

where the superscript n defines time. We are now interested in performing the analysis at time n . In coupled data assimilation, the cost function formally appears the same as in any other data assimilation. Assuming sequential data assimilation for simplicity, the cost function for a general two-component system can be formally defined in terms of its components

$$\begin{aligned} J(x_1, x_2) = & \frac{1}{2} \begin{pmatrix} x_1 - x_1^f \\ x_2 - x_2^f \end{pmatrix}^T \begin{pmatrix} P_{11} & P_{12} \\ P_{12}^T & P_{22} \end{pmatrix}^{-1} \begin{pmatrix} x_1 - x_1^f \\ x_2 - x_2^f \end{pmatrix} + \\ & + \frac{1}{2} [y_1 - h_1(x_1)]^T R_1^{-1} [y_1 - h_1(x_1)] + \frac{1}{2} [y_2 - h_2(x_2)]^T R_2^{-1} [y_2 - h_2(x_2)] \end{aligned} \quad (4.3)$$

where h is the nonlinear observation operator, y is the observation and R is the observation error covariance. The subscripts 1 and 2 refer to the first and second state vector components, respectively. It is implied in (4.3) that forecast error covariance can have cross-component correlations, but that observation errors are

independent. A general linear solution to the problem (4.3) can be obtained after setting $\nabla J = 0$ (e.g., Lorenc 1986)

$$x^a = (I + P_f H^T R^{-1} H)^{-1} (x^f + P_f H^T R^{-1} y) \quad (4.4)$$

with

$$x^a = \begin{pmatrix} x_1^a \\ x_2^a \end{pmatrix} \quad x^f = \begin{pmatrix} x_1^f \\ x_2^f \end{pmatrix} \quad y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \quad (4.5)$$

where H is the Jacobian of the observation operator. Now let assume that there is only a single observation of one of the components, say the component 1, defined at that grid point. Then the cost function (4.3) becomes

$$J(x_1, x_2) = \frac{1}{2} \begin{pmatrix} x_1 - x_1^f \\ x_2 - x_2^f \end{pmatrix}^T \begin{pmatrix} (\sigma_f^2)_1 & \rho_{12} \\ \rho_{12} & (\sigma_f^2)_2 \end{pmatrix}^{-1} \begin{pmatrix} x_1 - x_1^f \\ x_2 - x_2^f \end{pmatrix} + \frac{1}{2} [y_1 - x_1]^T (\sigma_o^2)^{-1} [y_1 - x_1] \quad (4.6)$$

where σ denotes the standard deviation, the cross-component error covariance (directly related to the correlation between the coupling components) is denoted ρ , and indexes f and o refer to the forecast and the observation, respectively. The analysis solution (4.4) and (4.5) for the cost function (4.6) in terms of the components is

$$x_1^a = \frac{1}{1 + \varepsilon_1^2} x_1^f + \frac{\varepsilon_1^2}{1 + \varepsilon_1^2} y_1 \quad (4.7)$$

$$x_2^a = x_2^f + \frac{1}{1 + \varepsilon_1^2} \frac{\rho_{12}}{(\sigma_o^2)_1} (y_1 - x_1^f) \quad (4.8)$$

where ε represents the ratio between forecast and observation standard deviations

$$\varepsilon_1 = \frac{(\sigma_f)_1}{(\sigma_o)_1}. \quad (4.9)$$

The solution (4.7) and (4.8) illustrates several important aspects of coupled data assimilation. When observing only one (the first) component, the analysis is identical to the de-coupled assimilation of the first component (e.g., (4.7)). This means that, although the coupled covariance has cross-component correlations, the analysis of the first component does not benefit from the coupling. Although the second component is not observed, the analysis of the second component is impacted by the observation of the first component (e.g., (4.8)). A closer inspection reveals that this was possible only because of the non-zero cross-component error

covariance ρ_{12} , assuming $y_1 \neq x_1^f$. This is a very important result of coupled data assimilation, as it indicates that it is possible to change the analysis of a coupled component even if it is not observed. There are often instances when one component, e.g., land surface, is not well observed, while the other component, e.g., atmosphere, is well observed. The above results suggest that the land surface control variables can still be adjusted in the analysis. Note that this would not be possible in the stand-alone land surface analysis system.

5 Structure of Coupled Forecast Error Covariance

In this section we conduct single observation experiments in order to understand and illustrate the structure of coupled forecast error covariance. As suggested above, the true power of coupled data assimilation comes from the cross-component correlations. They allow a more efficient use of observations by impacting all control variables, and thus produce a more balanced change of control variables. The structure of forecast error covariance can be assessed using a single observation experiment setup, where only one observation at a single point is assimilated (e.g., Thepaut et al. 1996; Whitaker et al. 2009). We apply the Maximum Likelihood Ensemble Filter (MLEF—Županski 2005; Županski et al. 2008) as a coupled data assimilation system. In all experiments we use 32 ensembles. Since this system is employing the ensemble coupled error covariance, there is no need to model the cross-components as they are created automatically by the ensemble forecasting.

As a first example we consider a land-atmosphere coupling by using the NASA Unified Weather Research and Forecasting (WRF) model (Peters-Lidard et al. 2015) that has an implicit coupling between Noah land surface model and the WRF atmospheric component. The horizontal model resolution of the parent domain is 27 km with the nest at 9 km. There are 31 vertical layers in the atmosphere, and 4 vertical layers in the soil. The control variable includes atmospheric variables (perturbation surface pressure, perturbation potential temperature, perturbation geopotential, horizontal winds, specific humidity) with clouds (cloud ice, cloud snow, cloud rain, cloud water, graupel) as well as land surface variables (soil moisture, soil temperature). We assimilate a pseudo-observation of cloud rain at 700 hPa and evaluate its impact in the analysis, in particular on land surface variables.

The analysis increments ($x^a - x^f$) for cloud rain and soil moisture are shown in Fig. 2. One can note that auto-correlation of cloud rain is adequately represented by the ensemble, indicating the maximum near the observed location and the response in the lower troposphere spreading down to surface (Fig. 2a). From the coupled data assimilation point of view, the analysis increments of soil variables are more interesting (e.g., Fig. 2b). First, one can see that the soil moisture analysis increments are non-zero, implying the existence of the cross-component correlation between cloud rain and soil moisture. The impact of cloud rain pseudo-observation has spread into all soil layers, with stronger impact at the near-surface layers.

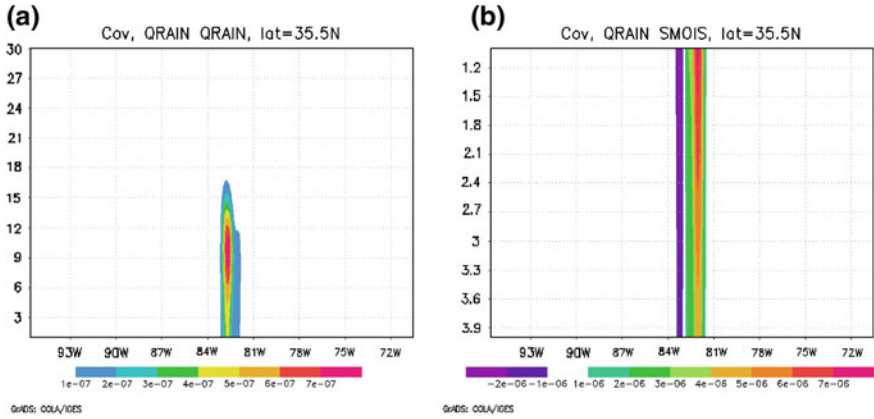


Fig. 2 Vertical cross-section of analysis increments $x^a - x^f$ for **a** cloud rain (kg/kg), and **b** soil moisture (kg/kg), as a response to a single pseudo-observation of cloud rain at 700 hPa. Note that the vertical index increases downwards for soil moisture, thus defines the depth. The vertical location of single observation is near level 11

A possible physical interpretation of these results is that an increase of cloud rain at 700 hPa induces an increase of moisture at the surface, which eventually spreads into the soil causing an increase of soil moisture. A careful investigation of the location of these impacts shows that the analysis increments of soil moisture lags the analysis increment of the atmospheric variable by about 60 km, with a shift towards the east. This suggests that the forecast error does not represent an instantaneous response, which would be characterized by a response at the same location. Rather, it reflects the interactions between atmosphere and soil as represented by the coupled modeling system, in this case characterized by a delay of land surface response. This potentially relates to the issue of spatiotemporal scale interaction in the coupled system (Sect. 3.5), but certainly requires further evaluation.

As a second example we consider a coupled atmosphere-chemistry model WRF-Chem (Grell et al. 2005). The control variables include the standard atmospheric variables and five species of chemical constituents (o3, so2, so3, no2, no3). The single observation experiments that examine the structure of the coupled atmosphere-chemistry error covariance have been published in Park et al. (2015). Here we present several additional cross-component correlations that confirm the inherent complexity of the coupled forecast error covariance. As in Park et al. (2015), we assimilate the Ozone Monitoring Instrument (OMI) total column ozone at a single point. In Fig. 3 we present the additional cross-component correlation that such atmosphere-chemistry coupled system describes, in terms of specific humidity (Fig. 3a) and east-west wind component (Fig. 3b). One can notice that both atmospheric components have well-defined analysis increments in both vertical and horizontal directions. Specific humidity has somewhat smaller magnitudes than the wind. It is also interesting to note that analysis responses for atmospheric

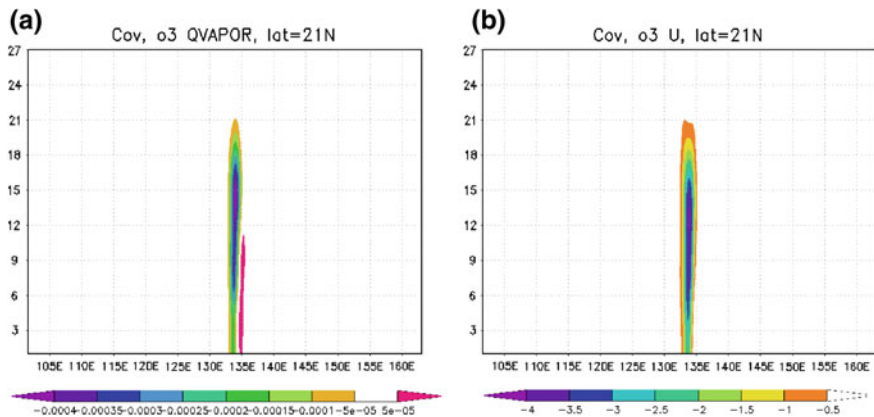


Fig. 3 Vertical cross-section of analysis increments $x^a - x^f$ for **a** specific humidity (kg/kg), and **b** east-west wind component (m/s), as a response to a single observation of total column ozone. The single observation location is near level 24

variables are located in middle and lower troposphere, while the maximum ozone response is at 250 hPa (e.g., Park et al. 2015). These analysis increments confirm that the cross-component covariance holds important information that could potentially benefit coupled atmosphere-chemistry data assimilation.

Lastly, we consider the experiment with coupled atmosphere-aerosol-chemistry model, which is the WRF-Chem model with the Goddard Chemistry Aerosol Radiation and Transport (GOCART) aerosol model (Chin et al. 2000) that includes the prediction of sulfates, black carbon, organic carbon, sea salt and dust. In particular, we investigate the correlations between ozone and dust. These correlations are largely unknown, but this example suggests that they do contain information that could be made useful in coupled data assimilation. The augmented control variable includes atmospheric variables, chemistry variables, and the aerosol variables (dust) at 0.5, 1.4, 2.4, 4.5 and 8.0 μm . As in the previous example we assimilate a single point OMI total column ozone observation. The analysis increments for selected dust variables are shown in Fig. 4. The cross-component correlation between ozone and dust indicates that the maximum analysis response of dust to ozone observation is at 300–400 hPa. One can again notice well-defined analysis increments, which suggests that the cross-component correlations are not a noise; rather, they appear to be a signal with important information about the coupled system uncertainties. A closer inspection of Fig. 4a, b shows that, although the responses are similar, larger dust particles have the maximum response at lower levels than small particles. Also, the magnitude of the increments is larger for smaller dust particles.

In overview, the above results indicate that the structure of coupled ensemble forecast error covariance is complex and contains important cross-component correlations, with potential benefit for coupled data assimilation.

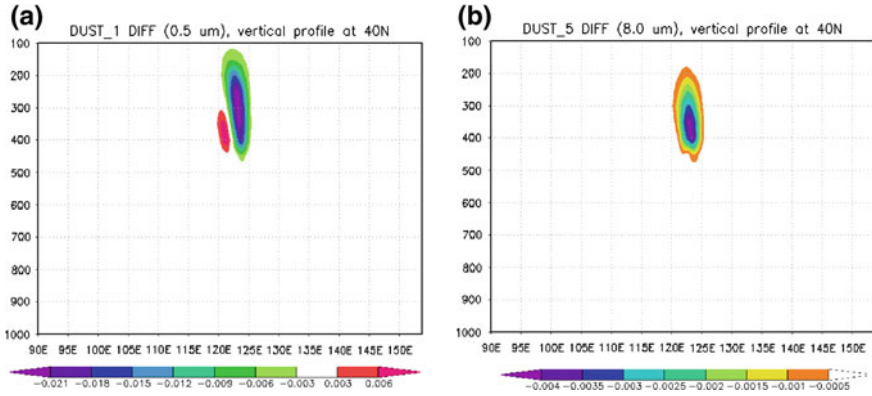


Fig. 4 Vertical cross-section of analysis increments $x^a - x^f$ in response to a single OMI total column ozone observation for **a** dust-1 (0.5 μm), and **b** dust-5 (8.0 μm)

6 Summary and Future

Coupled data assimilation is an important component of the coupled system modeling. It holds a great potential for improving the forecast of various Earth science components, as well as for better understanding of the coupled systems' state and uncertainty.

The coupled data assimilation formally represents a system very similar to single component data assimilation, however with increased difficulty. Main challenges that make coupled data assimilation difficult are associated with the structure of the forecast error covariance, the augmented control variable, the increased state dimensions, potentially non-Gaussian errors, and the interactions between the coupled components characterized by different spatiotemporal scales.

A simple two-component coupled system analysis indicates the relevance of the cross-component correlations. In particular, the cross-component correlations have a potential to increase the utility of observations in data assimilation by spreading the information throughout the components. The conducted single observation assimilation experiments confirm that the structure of cross-component correlations is complex and clearly related to the dynamical links between the coupled components and their control variables.

Numerous challenges of coupled data assimilation are still remaining, in particular related to the use of hybrid variational-ensemble systems, as the cross-component correlations of hybrid error covariance, coming from both the variational and ensemble methodologies, need to be reconciled. However, given the relevance and the increased interest in performing the forecasts with coupled modeling systems, the role of coupled data assimilation and its benefits will likely steadily increase.

Acknowledgements The author gratefully acknowledges support from the NASA Modeling, Analysis, and Prediction (MAP) Program Grant NNX13AO10G, the NASA Precipitation Measurement Mission (PMM) Program Grant NNX10AG92G, and the National Science Foundation Collaboration in Mathematical Geosciences Grant 0930265. The author would also like to acknowledge the computational support of NASA Advanced Supercomputing (NAS), and extend gratitude to the computing support from Yellowstone provided by NCAR's Computational and Information System Laboratory, sponsored by the National Science Foundation.

References

- Arellano-Valle RB, Contreras-Reyes JE, Genton MG (2012) Shannon entropy and mutual information for multivariate skew-elliptical distributions. *Scand J Statistics* 40:42–62
- Bannister RN (2008a) A review of forecast error covariance statistics in atmospheric variational data assimilation. I: characteristics and measurements of forecast error covariances. *Q J R Meteorol Soc* 134:1951–1970
- Bannister RN (2008b) A review of forecast error covariance statistics in atmospheric variational data assimilation. II: modelling the forecast error covariance statistics. *Q J R Meteorol Soc* 134:1971–1996
- Belo Pereira MB, Berre L (2006) The use of an ensemble approach to study the background error covariances in a global NWP model. *Mon Weather Rev* 134:2466–2489
- Berre L, Desroziers G (2010) Filtering of background error variances and correlations by local spatial averaging: a review. *Mon Weather Rev* 138:3693–3720
- Buehner M (2005) Ensemble derived stationary and flow dependent background error covariances: evaluation in a quasi-operational NWP setting. *Q J R Meteorol Soc* 131:1013–1043
- Chin M, Rood RB, Lin S-J, Muller JF, Thompson AM (2000) Atmospheric sulfur cycle in the global model GOCART: model description and global properties. *J Geophys Res* 105:24671–24687
- Cover TM, Thomas JA (2006) *Elements of information theory*. 2nd edn. John Wiley & Sons, Hoboken, New Jersey, 776 pp
- Derber J, Bouttier F (1999) A reformulation of the background error covariance in the ECMWF global data assimilation system. *Tellus* 51A:195–221
- Grell GA, Peckham SE, Schmitz R, McKeen SA, Frost G, Skamarock WC, Eder B (2005) Fully coupled “online” chemistry within the WRF model. *Atmos Environ* 39:6957–6975
- Han G, Wu X, Zhang S, Li W (2013) Error covariance estimation for Coupled Data Assimilation using a lorenz atmosphere and a simple pycnocline ocean model. *J Clim* 26:10218–10231
- Hollingsworth A, Lonnberg P (1986) The statistical structure of short-range forecast errors as determined from radiosonde data. Part I Wind Field *Tellus* 38A:111–136
- Lorenc A (1986) Analysis methods for numerical weather prediction. *Q J R Meteorol Soc* 112:1177–1194
- Park SK, Lim S, Županski M (2015) Structure of forecast error covariance in coupled atmosphere–chemistry data assimilation. *Geosci Model Dev* 8:1315–1320
- Peters-Lidard CD, Kemp EM, Matsui T, Santanello JA Jr, Kumar SV, Jacob JP, Clune T, Tao W-K, Chin M, Hou A, Case JL, Kim D, Kim K-M, Lau W, Liu Y, Shi J-J, Starr D, Tan Q, Tao Z, Zaitchik BF, Zavodsky B, Zhang SQ, Županski M (2015) Integrated modeling of aerosol, cloud, precipitation and land processes at satellite-resolved scales. *Environ Model. Softw.* 67:149–159
- Rasmy M, Koike T, Kuria D, Mirza CR, Li X, Yang K (2012) Development of the Coupled Atmosphere and Land Data Assimilation System (CALDAS) and Its Application Over the Tibetan Plateau. *IEEE Trans Geosci Rem Sen* 50:4227–4242
- Sakaguchi K, Zeng X, Brunke MA (2012) The hindcast skill of the CMIP ensembles for the surface air temperature trend. *J Geophys Res* 117:D16113. doi:[10.1029/2012JD017765](https://doi.org/10.1029/2012JD017765)

- Shannon CE, Weaver W (1949) The mathematical theory of communication. University of Illinois Press, 144 pp
- Silva C, Quiroz A (2003) Optimization of the atmospheric pollution monitoring network at Santiago de Chile. *Atmos Environ* 37:2337–2345
- Sugiura N, Awaji T, Masuda S, Mochizuki T, Toyoda T, Miyama T, Igarashi H, Ishikawa Y (2008) Development of a four-dimensional variational coupled data assimilation system for enhanced analysis and prediction of seasonal to interannual climate variations. *J Geophys Res* 113:C10017. doi:[10.1029/2008JC004741](https://doi.org/10.1029/2008JC004741)
- Tardif R, Hakim GJ, Snyder C (2014) Coupled atmosphere–ocean data assimilation experiments with a low-order climate model. *Clim Dyn* 43:1631–1643
- Thepaut J-N, Courtier P, Belaud G, Lemaitre G (1996) Dynamical structure functions in a four-dimensional variational assimilation: a case study. *Q J R Meteorol Soc* 122:535–561
- Whitaker JS, Compo GP, Thepaut J-N (2009) A comparison of variational and ensemble-based data assimilation systems for reanalysis of sparse observations. *Mon Weather Rev* 137:1991–1999
- Zhang S, Harrison MJ, Rosati A, Wittenberg A (2007) System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies. *Mon Weather Rev* 135:3541–3564
- Županski M (2005) Maximum likelihood ensemble filter: theoretical aspects. *Mon Weather Rev* 133:1710–1726
- Županski M, Navon IM, Županski D (2008) The maximum likelihood ensemble filter as a non-differentiable minimization algorithm. *Q J R Meteorol Soc* 134:1039–1050

Data Assimilation for Atmospheric, Oceanic and
Hydrologic Applications (Vol. III)

Park, S.K.; Xu, L. (Eds.)

2017, XXXVI, 553 p. 216 illus., 155 illus. in color.,
Hardcover

ISBN: 978-3-319-43414-8