Coupled data assimilation in climate research: A brief review of applications in ocean and land

Kazuyoshi Suzuki¹, Milija Zupanski²
¹ Institute of Arctic Climate and Environment Change Research (JACE), Japan Agency for Marine-Earth Science and Technology (JAMSTEC)
² Cooperative Institute for Research in the Atmosphere, Colorado State University

Abstract: Regions of the cryosphere, including the poles, that are currently unmonitored are expanding, therefore increasing the importance of satellite observations for such regions. With the increasing availability of satellite data in recent years, data assimilation research that combines forecasting models with observational data has begun to flourish. Coupled land/ice-atmosphere/ocean models generally improve the forecasting ability of models. Data assimilation plays an important role in such coupled models, by providing initial conditions and/or empirical parameter estimation. Coupled data assimilation can generally be divided into three types: uncoupled, weakly coupled, or strongly coupled. This review provides an overview of coupled data assimilation, introduces examples of its use in research on sea ice-ocean interactions and the land, and discusses its future outlook. Assimilation of coupled data constitutes an effective method for monitoring cold regions for which observational data are scarce and should prove useful for climate change research and the design of efficient monitoring networks in the future.

Keywords: coupled atmosphere-ocean/land; uncoupled data assimilation; weakly coupled data assimilation; strongly coupled data assimilation; error covariance

1. Introduction

To date, research on data assimilation has mainly been conducted by large-scale local climate monitoring organizations for the purpose of obtaining initial data and/or optimal empirical parameters for meteorological forecasts. However, the scope of data assimilation research has begun to expand in recent years, thanks to the dramatic increase in computing power. Moreover, increasing availability of data measured by satellites has led to a flourishing of data assimilation research focused on the effective use of such data by linking forecasting models with satellite data.

Responding to a recommendation from the Committee on Space Research (COSPAR), Simmons et al.[1] developed a road map for Earth system scientific research to be conducted under the lead of the European Centre for Medium-Range Weather Forecasts (ECMWF) from 2016 to 2025. The road map emphasizes the importance of data assimilation research in creating data sets to monitor Earth systems using satellite data and in evaluating and obtaining initial values for forecasting models. Among such efforts, research on coupled monitoring systems and data assimilation to support such systems is particularly prioritized and is believed to be an area that will be the target of much international research over the next decade.

The use of coupled monitoring systems improves and expands our ability to predict coupled processes. The components of coupled modeling systems include atmospheric, marine, terrestrial, biosphere, cryosphere, and hydrological processes, along with processes related to atmospheric chemistry such as aerosol movement and other processes that are of interest. Coupled models use diverse data related to these model components. Accordingly, there is a substantial role to be played by data assimilation in coupled modeling.

Copyright © 2018 Kazuyoshi Suzuki et al.
doi: 10.18063/som.v3i2.599
This is an open-access article distributed under the terms of the Creative Commons Attribution Unported License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Furthermore, data assimilation systems can be used to evaluate the impact of observational data in observation system experiments (OSEs), to design effective monitoring networks, and to demonstrate the utility of observational data (e.g., Inoue et al.\(^2\)). Such systems also provide the means to verify in advance the utility of satellite data to be generated by satellites scheduled for launch. Because hypothetical data are used as observational data in such cases, the experiments are referred to as observation system simulation experiments (OSSEs). The assimilation of coupled data can be used to design monitoring networks that take into account the interaction between two or more model components (e.g., snow/ice and ocean) and to demonstrate the utility of observational data.

This review focuses on data assimilation research involving coupled atmosphere-ocean/land data and global coupled data that include cryosphere data. Specifically, after providing an overview of coupled data assimilation in Section 2, this paper reviews existing data assimilation research focused on ocean, sea ice, and land, which are main components of the Earth system, in Section 3, and discusses the future outlook of data assimilation research in Section 4.

2. Coupled data assimilation

2.1 Overview of coupled-data assimilation

The primary motivation for developing coupled data assimilation methods is to improve the transfer of information within coupled modeling systems and to improve their predictive performance in terms of their ability to capture state quantities and to reduce their uncertainties. The advantage of coupled modeling systems is their ability to assimilate data; some components of a system may function to limit data assimilation of other components. In other words, the analysis of assimilated coupled data is expected to more closely match what occurs in the real world than separate analyses of assimilated data for each individual component.

![Figure 1. A schematic diagram of the analysis/forecast cycling for general data.](image)

**Figure 1.** A schematic diagram of the analysis/forecast cycling for general data. The flow is as follows. When forecast values and observed values are provided, an analysis value is calculated through data assimilation. This is known as the analysis step. This analysis value is then used as the initial value for the model, and the next forecasting cycle begins. Prediction conducted up to the point where the next observed data are provided is known as the forecasting step.

Coupled data assimilation systems can generally be divided into three types: uncoupled, weakly coupled, or strongly coupled.

Uncoupled data assimilation means that each system component is completely independent. Modeling and assimilation of one component does not impact other system components. An example of uncoupled data assimilation
involving two separate models (an atmospheric model and hydrologic model) would be to create a precipitation dataset using forecasts and data assimilation in the atmospheric model and then using that precipitation dataset as a forcing dataset for the hydrologic model that employs a completely different data assimilation system.

Weakly coupled data assimilation enables interaction between model components (e.g., atmosphere and land). In such cases, data assimilation is still performed independently for each component, as in the analysis step of uncoupled data assimilation. However, initial conditions and parameters obtained from independent data assimilation systems are fed back into the coupled model. Therefore, in weakly coupled data assimilation model components are coupled in the forecasting step but are not coupled in the analysis step. The critical point is that the cross-component covariance between coupled components is not used in assimilation.

Strongly coupled data assimilation refers to the use of a single system that combines information from all components to perform coupled forecasting and coupled data assimilation for all components. Because the cross-component error covariance is used, forecasts and observations of each individual component have the potential to affect all other components.

2.2 Forecast error covariance

Here, error covariance is discussed, which is the key element of strongly coupled data assimilation.

Forecast error covariance plays a fundamental role in Kalman filter, variational, ensemble, hybrid, and other related data assimilation methods. Forecast error covariance plays an even more important role in coupled systems. This is because the cross-component forecast error covariances describes the information shared by the components, and effectively represents a mechanism for exchanging the information between components and observational data within the model. Consequently, observations of one component can impact the analysis of other components. This clearly increases the utility of observational data. There are at least two possible situations worth mentioning: (1) one component has very sparse or non-existing observations, while other components have a relatively good coverage, and (2) all components have a similar, relatively good geographical coverage. In the first situation the exchange mechanism described above transfers the observation information from well-observed component to the analysis of the under-observed component, eventually allowing the increase of information content of the under-observed component. In the second situation the transfer of information between observations from different components still exists, but now it serves as an additional constraint for analysis adjustment in data assimilation.

However, one should also be aware that a reliable cross-component error covariance may be difficult to obtain. In pure variational methods, that include only static background error covariance, one needs to develop cross-correlation functions that represents statistical behavior of uncertainties between the components. This is often impossible or unreliable since such statistics is generally not available. In ensemble methods, on the other hand, there is a straightforward way of obtaining cross-component covariances from the ensemble forecasting of coupled models. Depending on the number of ensembles and on the true covariances between components, however, this estimate may also be unreliable. Since it is generally easier to adjust ensemble cross-covariances than to develop new cross-correlation functions from unknown statistics, the use of ensemble error covariances is preferred in strongly coupled data assimilation, implying an advantage of ensemble and hybrid data assimilation methods.

Following the method proposed by Thépaut et al\textsuperscript{(3)}, forecast error covariance using a Kalman filtering framework can be simplified by assimilating a single observation and can be expressed by the following equation:

\[
x_a = x_f + P_f H^T \left( \frac{y - H x_f}{\sigma_f^2 + \sigma_o^2} \right)_k
\]

where \( H^T \) is the transposed matrix-vector product of the observation operator \( H \) for a single observation, \( y \) is the observed value, and the subscripts \( a \) and \( f \) indicate the analysis and forecast values, respectively. The standard deviations of the forecast error and observation error are given by \( \sigma_f \) and \( \sigma_o \), respectively, and \( k \) is the location index for the observation. The equation is further simplified by assuming that the observation is located at a model grid point.
\[ x_a = x_f + P_f w \]
\[ w = (0 \cdots w_k \cdots 0)^T \]
\[ w_k = \left[ \frac{y - x_f}{\sigma_f^2 + \sigma_o^2} \right] \]

Error covariance for a two-component coupled system can be expressed in block-matrix form as

\[
P_f = \begin{pmatrix} P_{11} & (P_{21})^T \\ P_{21} & P_{22} \end{pmatrix}
\]

Here, the subscripts indicate coupled components. If it is assumed that only the first component is observed \((y_1)\) and that the second component is not observed, the Kalman filter analysis equation for a two-component coupled system becomes

\[
\begin{pmatrix} \hat{x}_{a1} \\ \hat{x}_{a2} \end{pmatrix} = \begin{pmatrix} x_{f1} + P_{11} w_f \\ x_{f2} + P_{21} w_f \end{pmatrix}
\]

where \(w_1\) is the first component of vector \(w\). If the cross-component covariance is non-zero \((P_{21} \neq 0)\), the analysis equation for the second component allows the analysis to be altered by the observation of the first component. Because the forecast error covariance for a two-component coupled system is typically non-zero, coupled data assimilation provides a mechanism by which the observed value of one component can influence the analysis of other components. To summarize, an observation has the potential to affect all components of a coupled system, improving the processing efficiency of observational data.

3. Coupled data assimilation research

This section introduces examples of uncoupled, weakly coupled, and strongly coupled data assimilation research focused on changes in the cryosphere, including sea-ice ocean processes, snow cover, and frozen soil, which have received much attention in recent years.

3.1 Ocean processes

3.1.1 Uncoupled data assimilation

The state of sea ice and state of oceans are important as initial values for coupled atmosphere-ocean models. Uncoupled data assimilation research in this area focuses on the assimilation of sea-ice extent data to generate initial values for coupled atmosphere-ocean models.

By assimilating sea-ice extent data, Tietsche et al.\(^{(4)}\) demonstrated the dependence of a global climate model on the initialization of sea-ice concentration in the Northern Hemisphere. The study also revealed the importance of sea-ice thickness in such models. The assimilated data can be used as initial values for global climate models to forecast changes in seasonal and decadal climate.

Toyoda et al.\(^{(5)}\) conducted a data assimilation experiment using global data including sea-ice extent data. The method used entailed correcting forcing data for ocean parameters and ocean temperature distribution (sea-ice extent) using three-dimensional variational assimilation. The study demonstrated that the reanalysis of sea-ice extent in both hemispheres is improved by incorporating corrected ocean forcing data and ocean temperature distribution. Coupled atmosphere-ocean-sea ice forecasting experiments suggest that ocean-sea ice areas using reanalyzed values calculated by this method as initial values provide suitable initial conditions.

3.1.2 Weakly coupled data assimilation

Fujii et al.\(^{(6)}\) demonstrated that in climate model simulations, using assimilated ocean data in a coupled system yields better precipitation forecasts than using ocean surface temperatures analyzed separately from the atmospheric model.
In the United States, the National Oceanic and Atmospheric Administration and the National Centers for Environmental Prediction (NCEP) are performing coupled data assimilation for Climate Forecast System Reanalysis (CFSR). Data assimilation, carried out separately for atmospheric and oceanic data, has generated reanalysis datasets for ocean surface temperature and sea-ice extent.

### 3.1.3 Strongly coupled data assimilation

One early example of coupled data assimilation that included the transmission of error information between atmosphere and ocean components was provided by Sugiura et al., who developed a four-dimensional variational (4D-Var) data assimilation system using a coupled ocean-atmosphere wide-area model to clarify the mechanical state of global climate on a seasonal timescale over multiple years. The coupled data assimilation system involved optimization of bulk coefficients for transfer between the atmosphere and ocean and resulted in improved estimates of mass, momentum, and heat exchange fluxes between the atmosphere and ocean. Sugiura et al. demonstrated that the resolution of structures related to a number of important phenomena, especially those occurring in the tropical Pacific and the Indian Ocean (El Nino, Indian Ocean Dipole, and Asian summer monsoons), could be substantially improved by applying this system to state estimations of climate processes from 1996 to 1998. They also explained that the forecasting performance for phenomena that change seasonally or annually can be improved by utilizing the coupled data assimilation system.

Coupled data assimilation research is beginning to flourish in the United States and Europe. The most advanced coupled data assimilation methods use forecast error covariance between atmosphere and ocean to correct the state of atmosphere-ocean coupling based on observational data.

The ECMWF introduced a certain degree of coupling in analysis values based on a 4D-Var method developed for the atmosphere (Laloyaux et al.). This coupled data assimilation system assimilates a wide range of oceanic and atmospheric observations and performs coupled ocean-atmosphere reanalysis using a coupled model. By applying the constraints of the coupled model in the analysis, the assimilated ocean observations directly influence the estimated atmospheric state, and, in the opposite direction, atmospheric observations directly influence the estimated ocean state. Although the 4D-Var method had a neutral effect overall in comparison with the uncoupled data assimilation system, the estimate ocean temperature value was slightly higher.

Sluka et al. used the cross-component forecast error covariance from the ensemble background estimate of the coupled model and performed strongly coupled data assimilation in a coupled atmosphere-ocean model. The results indicate that strongly coupled data assimilation reduces ocean analysis error relative to the weakly coupled data assimilation and demonstrate that the greater the reproducibility of ocean forecasts, the greater the improvement in atmosphere forecasting.

### 3.2 Terrestrial land

#### 3.2.1 Uncoupled data assimilation

First, uncoupled data assimilation research is introduced. There are examples of data assimilation research aimed at combining a land surface model and satellite data to compensate for the shortcomings of each. One such example is a paper by Rodell et al. that describes NASA Global Land Data Assimilation Project (GLDAS). Andreasis and Lettenmaier improved the reproducibility of outflow by assimilating snow cover retrieved from the Moderate Resolution Imaging Spectroradiometer (MODIS) and snow water equivalent retrieved from the Advanced Microwave Scanning Radiometer for EOS (AMSR-E) into a land surface model.

The above cases are involving the assimilation of satellite observation data. There are also examples of research entailing the assimilation of brightness temperature data itself. The majority of such research is related to soil moisture (Pathmathevan et al.; Reichle et al.). There are fewer examples of research using brightness temperature data assimilation for snow cover data assimilation than for soil moisture. Pulliainen demonstrated that the assimilation of brightness temperature data from the Special Sensor Microwave/Imager (SSM/I) and AMSR can reduce error in snow cover and snow water equivalent compared to spatial interpolation. In addition, Durand et al. showed that the assimilation of brightness temperature data into a land surface model yields better estimates of snow cover.
characteristics than the assimilation of snow cover data retrieved from the satellite brightness temperature data.

The above research on snow cover data assimilation was conducted via offline simulations in which atmospheric conditions serve as forcing and the magnitude of land surface correction does not influence the atmosphere. However, the atmosphere and land are closely linked and influence each other (e.g., Yasunari et al.\[16\]; Matsumura et al.\[17\]).

3.3.2 Weakly coupled data assimilation

In this section examples of research involving weakly coupled data assimilation are introduced. Using a global climate model, Rasmy et al.\[18\] developed a data assimilation system that analyzes atmosphere and land separately and then couples the results. The assimilation system yielded analysis values for soil moisture on the Tibetan Plateau that were closer to observed values. ECMWF (de Rosnay et al.\[19\]) developed a system to simultaneously analyze atmospheric and soil moisture and demonstrated that inconsistencies arising from offline data assimilation can be avoided and, further, that forecasts of ground surface temperature and soil moisture can be improved by using the system. The above research accomplishes coupled atmosphere-land surface data assimilation by enabling exchange (i.e., flux) between atmosphere and land surface.

3.3.3 Strongly coupled data assimilation

A number of examples of strongly coupled data assimilation research related to atmosphere-land processes have emerged in recent years. They involve methods that are more natural than combining information from models with observational data. Here, research by Suzuki et al.\[20\] is introduced that investigated the forecast error covariance and correlation structure of snow cover and atmosphere in Siberia.

Suzuki et al.\[20\] show the east-west cross-sectional error covariance structure of snow temperatures (TSNO) obtained from a single-observation of 2-m air temperature (T2) at two different locations (P1 and P2). It is evident that the forecast error covariance differs substantially at the two locations, with large forecast error covariance occurring at point P2. Meanwhile, evaluation of forecast error correlations at the same locations indicates that T2 and TSNO errors are more strongly correlated at P1 than at P2. These results indicate the need to quantify both the magnitude of coupling and the correlation of the two components based on forecast error covariance.

Further examination of Suzuki et al.\[20\] reveals the existence of even more complex (anisotropic) covariance and correlation structure than at P1 and P2, especially in the vertical direction. The major difference between the two locations in terms of weather is that P1 was experiencing snowfall, whereas the weather was clear at P2. The more complex covariance structure can be attributed to the impact of snowfall on the accumulated snow layer.

In this research, the fact that the atmosphere and land surface, and consequently the flux between atmosphere and land surface, are corrected by data assimilation suggests that the forecast error covariance in coupled systems will be complex. Detailed evaluation of forecast error covariance in different locations under a broad range of atmospheric and land surface conditions is needed.

In addition, recently Suzuki and Zupanski\[21\] showed more sophisticated land surface model can improve forecast skill in the land surface and lower atmosphere, and the selection of land surface model could be important for atmosphere-land coupled data assimilation.

4. Future outlook

In this review, the current state of coupled data assimilation research was discussed in relation to the cryosphere. In this final section, the future outlook of such research is discussed within the broader context of coupled data assimilation research.

Creation of long-term coupled data for reanalysis: At present, no data exist that can be used for coupled snow/ice-atmosphere/ocean reanalysis. Research aimed at creating datasets from coupled reanalysis of diverse observational data for snow ice, atmosphere, and ocean data is important. It is anticipated that such research will serve as a foundation for the long-term monitoring of cryosphere change and will contribute to breakthroughs in identifying key factors governing cryosphere change.

Forecast error covariance: Forecast error covariance plays a fundamental role in Kalman filter, variational, ensemble, hybrid, and other related data assimilation methods. The use of forecast error covariance enables forecasting
and data assimilation of coupled components to be carried out in a single system that combines information from all system components. In strongly coupled data assimilation, because the cross-component forecast error covariance is used, forecasts and observations for any given component have the potential to affect all other components. As the cross-component forecast error covariances between different components are generally unknown, characterization of their structure is a necessary preliminary step before coupled data assimilation can be performed. Research on the use of cross-component forecast error covariance is still in its infancy, and further research in this area is needed.

Differences in timescale: Forecast error covariance enables information from one system component to be efficiently transferred to another system component and enables analysis that is more physically consistent. That said, the timescale of atmospheric measurements is typically short, while the timescale of ocean and snow/ice-measurements is typically longer. When assimilating observational data of different system components at the same time, it is important to handle the different timescales of different observational data in a suitable manner. Accordingly, standard methods for doing so must be developed.

OSEs and OSSEs: Long-term field observations and the launching of new satellites to monitor the cryosphere are extremely expensive. As such, the design of efficient observation systems is a top priority. Coupled data assimilation systems can aid in the design of observation systems. It is anticipated that improvements in coupled data assimilation systems and OSEs utilizing coupled data assimilation systems will prove useful in developing efficient observation networks for the cryosphere.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgements

This research is supported by "Computational Joint Research Program (Collaborative Research Project on Computer Science with High-Performance Computing)” at the Institute for Space-Earth Environmental Research, Nagoya University and is partly carried out by the Joint Research Program of the Institute for Space-Earth Environmental Research, Nagoya University. Parts of this study were supported by a Grant-in-Aid for Scientific Research (C) (No. 16K00581), and a Grant-in-Aid for Challenging Exploratory Research (No. 25550022). In addition, the second author acknowledges partial support from the Office of Naval Research under contract N000149169192040.

References

11. Andreadis KM., Lettenmaier DP. 2006. Assimilating remotely sensed snow observations into a macroscale
hydrology model. Advances in Water Resources, 29, 872–886