

DATA ASSIMILATION, TREATMENT OF UNCERTAINTY AND OBSERVATION IMPACT

1. ABSTRACT

Throughout the other themes of the NCEO, data assimilation will be used to combine new Earth Observation measurements with different models of the Earth system components. A few illustrative examples of the use of data assimilation include the initialisation of atmosphere and ocean models for seasonal prediction (Theme 1), the initialisation of land surface models (Theme 1), parameter and flux estimation in carbon models (Theme 2), assimilation into chemical transport models (Theme 3) and initialisation of storm-scale forecasting models (Theme 4). Applications of assimilation to models of glacial flows and sea-ice interactions (Theme 5) and to models of volcanic activity (Theme 6) are also now emerging.¹ Current assimilation techniques have largely been developed for large-scale atmospheric and oceanographic flows and are based on simplifying assumptions that may no longer hold for these models and data. In this theme we will develop novel techniques for assimilating data into different Earth system models within the NCEO. The key challenges to be addressed are

- The use of very high resolution models and complex models.
- The presence of coupled systems with very different scales in space and time.
- The presence of very highly nonlinear systems.
- The desire to estimate surface fluxes and sources from EO data.
- The need to assess quantitatively the impact of novel observing systems.

Essential to this theme will be the development of methods to understand and represent uncertainty in the observations and the models. By concentrating resources in this separate data assimilation theme we will develop a national capability in data assimilation which will provide expertise to other scientists in the NCEO and will foster collaboration with operational centres, in particular the Met Office and the European Centre for Medium-range Weather Forecasting.

In coordination with the other themes, we will identify and abstract the essential components of the assimilation problem in the different applications. New data assimilation techniques for these applications will be developed and tested on simple models, before being transferred back into the other themes for implementation in full Earth system models. The theme will be resourced through 17 FTEs of PDRA time working with the other themes, robustly testing currently used methods and developing new data assimilation approaches that can feed directly into the application areas. At the same time 4 PhD studentships will be used to study novel approaches to longer term questions, building links with the wider mathematics and EPSRC communities. A fundamental objective for this theme is the training of scientists in the theory and practice of data assimilation.

2. BACKGROUND AND MOTIVATION

Earth Observation (EO) measurements from satellites provide exciting new possibilities for understanding and predicting the behaviour of the Earth system. The range of current and planned satellite missions is giving scientists information on many different aspects of the system behaviour. However, converting such measurements into scientific understanding is not a

¹ The Themes are listed for reference in the annex.

straightforward task. In many cases the observations are only indirectly related to the physical quantities of interest. Furthermore the observations may be irregularly distributed in space and time and will contain measurement errors. In order to extract the maximum understanding from the observations it is necessary to combine them with our understanding of the physical systems, as embodied in prediction models, taking into account uncertainties in both the data and the models. The technique for combining EO data with numerical models in this way is data assimilation.

Traditional data assimilation has matured in the application of numerical weather and ocean prediction to provide initial conditions for a model forecast. Within the NCEO, data assimilation will be applied to a much wider range of applications in order to make full use of the available EO and enhance our understanding of the whole Earth system. The assimilation techniques that have been developed thus far in synoptic-scale meteorology cannot, however, be applied in the same way to other components. Current assimilation techniques depend on simplifying assumptions that are difficult to justify in the context of these new application areas. Hence for each Earth system component the assimilation problem must be reconsidered and techniques need to be developed that are appropriate for each system. Whereas much of this work in developing different data assimilation systems will reside in the application themes of the NCEO, there are many questions and problems that arise that are common to more than one application. To solve these problems satisfactorily, it is necessary for those working on different applications of data assimilation to collaborate, ensuring dialogue and cross-fertilization of the basic techniques and theory of data assimilation. The assignment of such collaboration to this cross-cutting theme is conceived as providing a resource to enable this interaction to occur.

Scientists funded by this theme will engage with those working in the other themes, using expertise in data assimilation theory to help design assimilation systems and experiments for different Earth system components. At the same time they will be able to abstract from the particular application the general theoretical questions that need to be tackled, to develop novel data assimilation approaches for Earth system science and to spread best practice in data assimilation throughout the NCEO.

The different building blocks of a general data assimilation system are shown schematically in Figure 1. In order to obtain successful assimilation systems for different components of the Earth system we will need to consider the design of each of these building blocks for the specific application. In this theme we will develop

- (i) general theory and techniques for representing the different inputs and their uncertainties;
- (ii) novel assimilation algorithms for combining the inputs under different assumptions;
- (iii) methods for interpreting the outputs in the light of the input assumptions and for quantifying their uncertainty.

General methods will be developed which will then be adapted for the different application areas within the other themes of the NCEO.

In designing assimilation systems for these new applications we are presented with several major challenges that have not previously been faced in the context of synoptic-scale meteorology.

Some of these challenges include

- The use of very high resolution models and complex models.
- The presence of coupled systems with very different scales in space and time.
- The presence of very highly nonlinear systems.

- The desire to estimate surface fluxes and sources from EO data.
- The need to assess quantitatively the impact of novel observing systems.

It is essential that we begin to evaluate the power and limitations of current assimilation techniques in the context of these problems and to develop new approaches to data assimilation which better respond to these challenges. To achieve these aims it is recognized that the resources allocated to this theme are not sufficient to answer all of the questions that arise. It would be possible to concentrate all the available resources on one particular application of data assimilation. We have instead chosen to spread the resources throughout the different application areas of the NCEO. Hence the resources allocated to this cross-cutting theme are seen as seed corn funding, to provide a base of expertise in data assimilation theory and practice that will be extended through wider collaborations. The aim will be to ensure that a coherent programme in data assimilation is developed, with best practice being spread from one application area to another. Of particular importance will be to engage with the mathematical sciences community, to provide further resources for studying the longer-term research questions. Such engagement will be encouraged through joint PhD studentships with UK mathematicians, the setting up of a network on the mathematics of data assimilation and interactions with related projects such as VISDEM² and MUCM³.

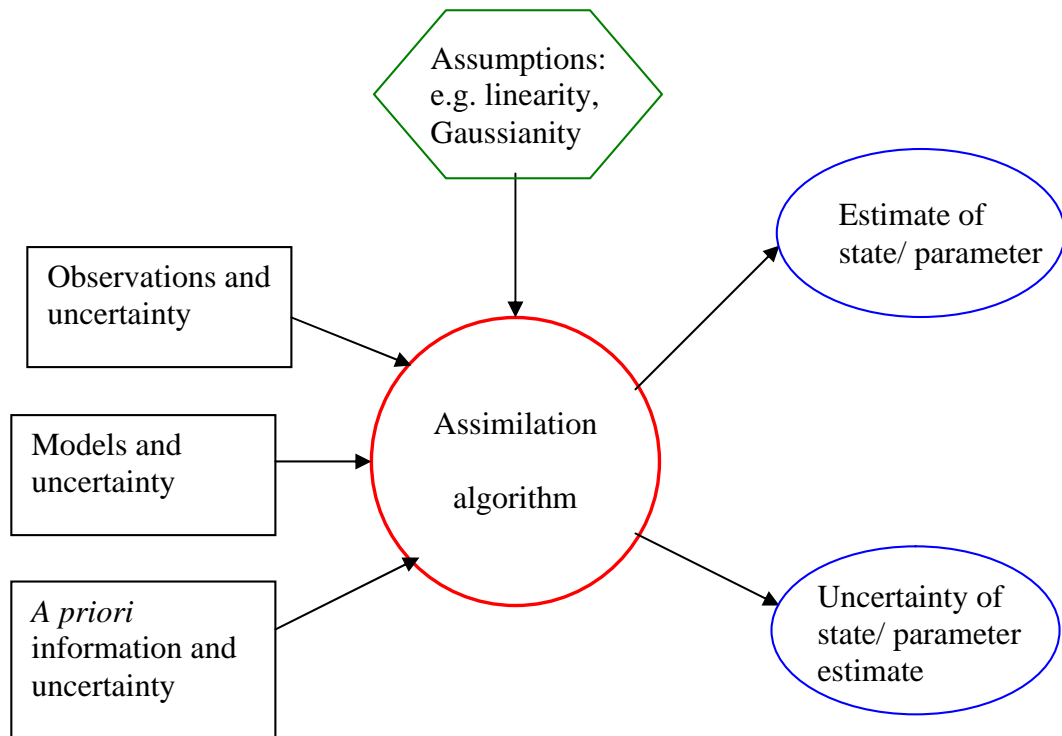


Figure 1: Schematic diagram of the building blocks of a data assimilation system.

In considering the structure of this theme, we have identified five sub-themes chosen to address the key scientific questions that arise in the different applications of data assimilation in the NCEO. A further sub-theme on training will address the skills shortage in data assimilation. The sub-themes are now listed, with an indication of their importance to the rest of the NCEO:

² Variational Inference for Stochastic Dynamic Environmental Models, EPSRC/ NERC funded project, 1/2/06-31/1/10.

³ Managing Uncertainty in Complex Models, RCUK funded project, 1/6/06-31/5/10.

- 3.1: *Analysis of coupled and multi-scale systems.*
The other themes include data assimilation applied to many different coupled and multi-scale systems, including coupled atmosphere-ocean and land-atmosphere models (Theme 1), coupled ocean physics-biology models (Themes 1 and 2), coupled atmosphere-chemistry models (Theme 3) and nested storm-scale models (Theme 4).
- 3.2: *Quantifying and representing uncertainty.*
A good representation of the uncertainty in data and models is a key foundation of any assimilation system and the work in this sub-theme will link to all areas of the NCEO where data assimilation techniques are being used.
- 3.3: *Using observations to estimate model parameters.*
In the NCEO, parameters will be estimated from EO data for land surface models (Theme 1), carbon models (Theme 2), chemistry models (Theme 3) and hydrological models (Theme 4).
- 3.4: *Estimating surface fluxes using data assimilation.*
Techniques for estimating surface fluxes and sources of atmospheric gases will be used for scientific investigation in Themes 1, 2 and 3.
- 3.5 *Quantifying observation impact.*
This sub-theme will develop methods for measuring the actual or potential impact of observations, initially using examples from Themes 1, 2 and 4, but with the aim of spreading these techniques to all assimilation systems within the NCEO.
- 3.6: *Training.*
Of fundamental importance in the NCEO will be the training of personnel in the theory and practice of data assimilation. This is required both for junior staff at the start of their careers and for more experienced staff who wish to apply data assimilation to their models.

The scientific content of each sub-theme is explained in more detail in the next section.

3. SPECIFIC OBJECTIVES

The specific objectives of this theme are:

- To develop data assimilation techniques for use in coupled, multi-scale Earth system models.
- To develop techniques to represent uncertainty in observations and models in data assimilation, including the representation of stochastic processes, formulation of observation operators and treatment of observational errors of representation.
- To develop mathematical theory of parameter estimation in Earth system models.
- To develop inverse modelling techniques to estimate surface fluxes of trace gases.
- To provide expertise to the wider NCEO in the development of assimilation systems for specific applications.
- To develop and improve techniques for quantifying the impact of observations.
- To develop, experiment with, and offer to the wider community a framework for carrying out observing system simulation experiments (OSSEs).
- To train scientists in the theory and practice of data assimilation.

In each sub-theme the basic theory will be developed using a hierarchy of models, from simple models to the more complex application models utilized by the NCEO. The objectives will be pursued under six sub-themes as follows.

3.1 Analysis of coupled and multi-scale systems

Motivation

The integrating theme for the NCEO is the exploitation of EO data for climate diagnosis and prediction, as described in Theme 1. Prediction on seasonal to inter-annual and decadal timescales requires consideration of the whole Earth system, using atmospheric models coupled to models of the land surface, the cryosphere and the ocean. The necessity of coupling models of Earth system components in this way is recognized throughout the NCEO and many coupled systems will be studied, including atmosphere-ocean and land-atmosphere models (Theme 1), ocean physics-biology models (Themes 1 and 2), atmosphere-chemistry models (Theme 3) and nested storm-scale models (Theme 4). Accurate analysis and prediction using such systems will rely on good initialisation of each of the system components in a way that is consistent between them. Whereas data assimilation techniques have previously been applied to some of these systems individually, very little is understood about how to perform data assimilation when these systems are coupled together. Each of the Earth system components evolves on different temporal and spatial scales and new theoretical developments are required to be able to assimilate information on different scales simultaneously. Furthermore, the coupling of solutions from independent assimilations leads to unrealistic features developing near the coupling boundaries. For example, in coupled ocean-atmosphere models used in seasonal forecasting, shocks may form in the system at the start of an integration (Lee et al., 2000). However the development of assimilation methods for coupled systems is still very much in its infancy. In this sub-theme novel assimilation techniques will be developed to assimilate data into multi-scale, coupled systems. Using the examples of atmosphere-ocean coupling and nested storm-scale models we will develop general techniques for assimilating data into coupled systems. The techniques will then be made available for implementation in other applications within the NCEO.

Work description and methodology

Current data assimilation methods have been developed to analyse features mainly on the large, slow scales in the atmosphere and oceans. However, in many of the Earth component systems important for longer term prediction, there are a wide range of spatial and temporal scales and it is important to initialise each of them carefully. The first stage in this work will be to develop four-dimensional variational data assimilation (4D-Var) schemes for application to systems with two distinct time scales. The 4D-Var method is a smoothing approach and is preferred over Kalman filter methods since for the present it is more practicable for the large systems used in climate prediction. Methods will be developed for understanding and using efficiently the information available to the assimilation at different scales and the 4D-Var methodology will be adapted to extract the maximum amount of information from the observations at each scale. Amongst other techniques, the inclusion of scale-dependent constraints in the variational assimilation problem to control the initialization of system components at different scales will be investigated in low order systems, such as the Lorenz 5-component model and the 3-body problem, expanding on our previous work reported in Watkinson (2006).

The work on multi-scale systems will then be extended to study the problem of assimilation for fully coupled systems with widely separated temporal or spatial scales. Using a one-dimensional coupled atmosphere-ocean model we will investigate the application of variational data assimilation methods to coupled systems. Different degrees of coupled assimilation will be compared, from completely separate assimilations in the atmosphere and ocean components, to a fully coupled assimilation in which observations of one component can affect the state of the other. Using results from the first part of this sub-theme, different methods for imposing weak constraints to control the coupling interface will be studied.

The problem of combining different scales also arises in storm-scale data assimilation, where small-scale features must be analysed while taking into account the synoptic scale information. In collaboration with Theme 4 of this proposal we will study methods for correctly incorporating information from the synoptic scales into a storm-scale model nested within a larger-scale forcing model. A one-dimensional multi-scale model, the Kuramoto-Sivashinski equation, will be used to investigate the control of different scales within a nested model assimilation. This is a nonlinear partial differential equation that is amenable to analysis (Protas et al., 2004). Particular attention will be paid to the formulation of the lateral boundary condition within the assimilation problem. High-resolution models bring with them the opportunity to use high-resolution EO measurements with highly correlated errors, which are not assimilated in current systems. To exploit the information contained in such data, advanced variational assimilation schemes will be developed for convective-scale models and for other high-resolution applications as part of the research.

Much effort has been, and continues to be, put into creating global reanalyses going back several decades (see also Themes 1 and 2 in this proposal). In reanalysis, state of the art methods are applied to get improved analyses from historical observations. Analyses for a particular time can also be improved by using information about how the system evolved after the analysis time. In coupled systems with widely ranging timescales a current analysis of the slow modes will depend on the estimated impact of many cycles of the fast modes. In this context retrospective analysis of the fast modes needs to be considered (for example, analysis of rainfall over the last 24 hours). The smoothing approach appears best suited to this problem, as the Kalman Filter and its variations do not allow information about subsequent evolution to affect the analysis. However, the Ensemble Kalman Filter has other advantages. Using the 5-component Lorenz model and simplified models of convection, the benefits of the retrospective potential of smoothers will be quantified.

An important component of this sub-theme will be to ensure that the assimilation problem for coupled and multi-scale systems can be solved efficiently and accurately at all relevant scales. As computer power continues to grow, coupled climate simulations for seasonal to decadal prediction and storm-scale forecast models will continue to be run at higher resolutions, incorporating a wider range of spatial and temporal scales. New techniques will be developed to exploit the multi-scale, coupled nature of these systems in order to generate accurate analyses economically in real time with given computational resources. In particular new pre-conditioners and new iteration procedures for coupled and multi-scale models will be derived for solving the assimilation problem.

Deliverables

- Characterisation of how different assimilation approaches deal with coupled and multi-scale systems.
- Novel assimilation methods for using observations in treating and estimating the state of such systems.
- Efficient methods for solving the assimilation problem in large climate prediction systems and in high resolution systems.

3.2 Quantifying and representing uncertainty

Motivation

Most assimilation methods are based on a statistical framework or can be interpreted in such a way. The optimal combination of observations with models within assimilation schemes depends on a good knowledge and representation of the uncertainty present in the inputs. The input statistics of errors in the a priori information strongly influence how information from observations is spread spatially in the analysis and how observations of one physical quantity may influence the analysis of a different quantity. Thus good representation of these forecast errors is important for making the best use of observations of the Earth system. At the same time it is necessary to characterise well the errors in the observations themselves in order to relate them to model quantities. Of particular importance is to consider how representative the observation is of the quantity being modelled on a discrete grid. Furthermore, we know that our models themselves are not an exact representation of reality. Including a representation of errors in our models will enable known model errors to be captured and the fit of the analysis to the observations to be improved, but good estimates of the model error covariances are important for success (Tremolet, 2007). In this sub-theme we will develop methods for quantifying and representing uncertainty in forecasts, observations and models. General methods will be developed in simple systems and demonstrated in the context of two applications, (i) coupled atmosphere-ocean models and (ii) chemistry transport models. These applications embody the problems of estimating forecast and model error covariances and the ill-posed nature of the inverse problem.

Work description and methodology

An essential component of fully coupled data assimilation systems will be the correct representation of forecast errors, since these will determine how information from observations is spread through the coupling boundary. The effect of coupling models together on the error covariances will be studied by using an ensemble Kalman filter to generate covariance statistics, using low order systems with different time scales and the one-dimensional coupled atmosphere-ocean model developed in sub-theme 3.1. Particular attention will be given to understanding how errors in one part of the coupled system are linked to errors in the other part. This work will require a formulation of the ensemble Kalman filter which ensures that useful information is obtained about the coupled covariances. Various formulations will be investigated in collaboration with Theme 4 to understand the most appropriate methods for generating these covariances. Other methods of sampling the probability density function which do not assume Gaussianity will also be applied, to understand how well the Gaussian assumption holds for coupled systems.

In practice it is not possible to use forecast errors derived in this way directly in a data assimilation system, due to the large size of the matrices. As a next step in this sub-theme we will develop methods to represent the forecast errors of coupled systems within the data assimilation problem. This work will be in collaboration with Theme 4, which will develop models for multi-scale covariances in storm-scale forecasting. Here we will apply similar methods to model the cross-covariances in coupled, multi-scale systems. Different covariance models will be developed based on the statistics calculated in the first part of this sub-theme. The covariance models will be compared numerically through implementation in the one-dimensional coupled atmosphere-ocean assimilation system developed in sub-theme 3.1.

A further uncertainty in our assimilation systems is the uncertainty in the models themselves. Such uncertainty consists of both random error and biases. It is known that biases become more

apparent when systems are coupled together. However most data assimilation schemes as currently implemented assume that the numerical model is perfect and do not allow for model error. The work of Griffith and Nichols (2001) and Dee (2005) will be extended to develop general methods of assimilating observations into different Earth system models in the presence of model bias. Particular attention will be paid to coupled models, where the error characteristics in each model component may be very different. Methods for characterising model error will be studied using the shadowing filter developed by Judd and Smith (2004).

Of particular importance in representing uncertainty in data assimilation is the underlying statistical model. For the majority of assimilation schemes currently used in practice all errors are assumed to follow a Gaussian probability distribution. However, in many of the highly nonlinear problems faced in the NCEO, such as assimilation for high resolution models or assimilation of tracer fields, this assumption is no longer valid. If we are to capture the correct interaction between high and low frequency modes then it is necessary to develop assimilation methods that do not rely on Gaussian statistics. We will develop new nonlinear filtering methods based on the theory of stochastic differential equations, extending the work of Crisan and Lyons 1997 and Stuart et al. 2004.

A key issue in the data assimilation problem for Earth observation is that many aspects of the problem are ill-posed; that is, small errors in the input data can lead to large changes in the analysed and predicted states. Resolving this issue is particularly important in the context of flux estimation from EO data (see sub-theme 3.4). Currently data assimilation problems deal with this by regularizing with a priori constraints and, in the case of variational methods, by the premature termination of the solution procedure. However the regularization properties of these two methods and their effect on the uncertainty in the outputs are poorly understood. These properties will be studied theoretically and numerically to understand and explain the behaviour of current data assimilation systems. Other regularization techniques will also be studied, to determine the most appropriate method to improve the solution procedure of variational data assimilation problems. The methods developed will be tested using simplified systems and in the flux estimation system with the TOMCAT chemical transport model being developed in sub-theme 3.4. Further details of this model are given under that sub-theme.

Deliverables

- Analysis of forecast covariance information for coupled systems.
- New methods for representing covariances in coupled models.
- New methods for treating random model error and bias in data assimilation.
- Methods for regularising four-dimensional variational assimilation and flux estimation problems.

3.3 Using observations to estimate model parameters

Motivation

Many models of the Earth system are very highly parameterized. These parameters appear in the models themselves and in the formulation of forward operators that provide the link between model and observation. For example, the terrestrial biosphere model BETHY used in the Carbon Cycle Data Assimilation System (CCDAS) contains 21 controlling parameters for 13 different plant functional types (Rayner et al., 2005). Often these parameters are poorly known and Earth observation data provide a new opportunity to estimate the parameter values more accurately. Throughout the NCEO observational data will be used to estimate model parameters in different

models, including models of the land-surface (Theme 1), terrestrial carbon (Theme 2), atmospheric chemistry (Theme 3), the ecosystem (Theme 1) and flood models (Theme 4). In this sub-theme we will develop and test techniques for estimating model parameters from observations. General assimilation techniques will be studied and methods for understanding the value of different observation types for parameter retrieval will be developed. The theory developed will be tested thoroughly in the context of a carbon cycle model using the CCDAS.

Work description and methodology

An important element of using Earth observation for parameter estimation is to determine which parameters can be retrieved and with what accuracy, given the available observations. We shall develop general techniques to quantify how much information is available in the observations assimilated into, typically, a carbon cycle model. However, in accordance with our overall approach in this theme, we shall study these issues using a hierarchy of simple models, from the 3-component Lorenz equations to the CCDAS. We will consider how many observations of different observation types are required to guarantee that the assimilation problem is well-posed and has a unique solution. In collaboration with sub-theme 3.5 we will examine the effect of removing certain observation types or making new kinds of measurements.

We shall then examine the required accuracy of the observations to retrieve particular parameters in the different models. The sensitivity of model outputs to the change in parameter value will be calculated using adjoint techniques and compared with the accuracy with which we are able to observe such outputs. We expect that the observational error should be less than the model output sensitivity if the observation is to provide any useful information towards the retrieval of the parameter. A theoretical analysis of the sensitivity to observations will be explored, extending the approach of Cardinali et al (2004).

In the main, two data assimilation approaches, sequential and variational, are currently used for parameter estimation in terrestrial ecosystem models. The ensemble Kalman Filter (sequential approach) has been used for estimating parameters in rather simple box models of the terrestrial biosphere operating on a site level assimilating flux data from eddy correlation measurements (e.g. Williams et al., 2005). The variational approach making use of the adjoint of the modeling system has proved to be a powerful tool in carbon cycle research on a global scale for assimilating atmospheric CO₂ concentration measurements from the global station network (e.g. CCDAS).

An essential commonality among the sequential and variational methods is that for both approaches data uncertainties are as important as data values themselves and have a comparable role in determining the outcome. Given the importance of data uncertainties, there is an urgent need for soundly based uncertainty characterizations for the main kinds of data used in carbon cycle observations and for the specification of the main properties of the error covariance matrix. Methods for weighting different data types in relation to each other will also be studied, for example to decide how to weight an integrated CO₂ concentration measurement against a single eddy flux measurement at a local site.

On a similar issue, special emphasis will also be given to the impact of biases in the observational data on parameter estimation and their uncertainties in a CCDAS type set up. As an example, the upcoming space based observations of atmospheric CO₂ (by OCO and GOSAT) are likely to have sampling biases, which are spatially coherent resulting from cloud conditions, atmospheric aerosol layers etc. However, the vast amount of global satellite observations of CO₂ will have the

potential to significantly improve our quantitative understanding of land surface processes by complementing the existing surface observation network. Therefore some fundamental theory on the treatment of biased data in parameter estimation will be developed. This work will link directly to the general problem of bias in data assimilation studied in sub-theme 3.2 and to the OSSE studies of sub-theme 3.5.

Deliverables

- Methods for quantifying the information content of observations for parameter estimation.
- Robust methods for estimating model parameters using EO data.

3.4 Estimating fluxes using data assimilation

Motivation

An important component of climate monitoring and prediction is the quantification of the global distribution, sources and sinks of key atmospheric constituents, in particular of CH₄ and ultimately CO₂. Earth observation is increasingly being used to measure the concentrations of these gases and two targeted missions, OCO and GOSAT, will be launched within the lifetime of the NCEO (see Themes 2 and 3 for further details and examples). Of particular interest is to use these measurements to determine sources and sinks of atmospheric gases, for example to determine natural carbon fluxes to quantify better the global carbon budget, or to monitor regional anthropogenic emissions of pollutants. The estimation of surface fluxes of these gases is a key goal of Theme 2 of the NCEO. However this estimation problem is very ill-conditioned and small errors in the input data or assumptions can lead to large errors in the solution. The estimation of CO₂ fluxes is an especially big challenge, since this gas is well mixed in the atmosphere. In this sub-theme we will develop robust methods for estimating surface fluxes from Earth observations using an adjoint technique. The validity of the methods will be demonstrated on the problem of determining CH₄ sources using the TOMCAT chemical transport model. Results will be compared with the estimates from the Kalman filter technique being developed as part of Theme 3. New methods will be provided to Theme 2 to develop systems for estimating fluxes of CH₄, CO₂ and other gases.

Work description and methodology

The identification of sources and sinks of atmospheric trace gases is a fundamental part of modelling global atmospheric change and the attribution of pollution. New satellite observations provide measurements from which we can derive concentrations of different gases. The challenge is to determine the surface fluxes of these gases as a function of space and time, using the concentration measurements together with atmospheric winds and temperatures. However inverting measurements of concentrations into estimates of sources and sinks is a non-trivial problem. The inversion problem is very ill-conditioned, such that errors in the inputs are subject to large amplification (Enting 2002, Section 1.3). The problem may be regularised by the specification of a priori information about the sources, but the uncertainty on this information must also be specified. A particularly important problem is to understand the effect of different spatial and temporal aggregations of the data on the flux estimates, in the context of a changing observing system.

In this project we will develop four-dimensional variational techniques to estimate surface fluxes from satellite measurements of passive tracers. Initial work will consider a simple transport model, such as the one-dimensional advection-diffusion equation. Different methods for regularising the inverse problem will be investigated, for example the inclusion of a Tikhonov

term or a maximum entropy principle (Bocquet 2005). This work will link directly to the work on regularisation of ill-posed problems in sub-theme 3.2 of this theme. Methods for understanding the information content of the data will be investigated in collaboration with sub-themes 3.3 and 3.5.

Techniques developed in this simple system will then be implemented in the NCEO chemical transport model TOMCAT (see Theme 3). We will use the TOMCAT model with a linearized chemistry scheme, where the OH radical is not part of the active chemistry, but is a specified field (Chipperfield 2006). The adjoint of this model will be derived and will be used to develop variational methods for estimating surface fluxes of CH₄ from satellite data. This project will be in collaboration with Theme 3, which will implement an ensemble Kalman filter method of flux estimation using the same transport model. Characterisation of the input uncertainties from Theme 3 will be used here, to enable a direct comparison between the two systems. Particular attention will be paid to the sensitivity of the flux estimates with respect to

- assumptions on the a priori information and its uncertainty;
- size of observational errors;
- representation of spatial correlations in the data; and
- errors in the chemical transport model.

Idealised experiments will be used to compare flux estimates from the adjoint method with those from the ensemble Kalman filter approach. The most promising inversion system will be extended to the problem of estimating fluxes of CO₂, in collaboration with Theme 2.

Collaborations will be developed with the Environmental Flow Facility at the University of Surrey (EnFlo), with the aim of carrying out idealised laboratory experiments to validate the techniques on idealised problems.

Deliverables

- 4D-Var flux estimation systems for a hierarchy of models.
- The adjoint of the TOMCAT model incorporated within a flux estimation system.
- A comparison of methods for regularising the flux inversion problem.

3.5 Quantifying observation impact (OSEs and OSSEs)

Motivation

Of particular importance in developing data assimilation systems is to understand and quantify the actual and/or potential impact of different observation types within the assimilation process. In meteorological applications this has led to the development of observing system experiments (OSEs) and observing system simulation experiments (OSSEs). Applied work using OSSEs is already planned in Theme 4 of the NCEO. In this sub-theme we will develop the methodology of OSEs and OSSEs, with the aim of understanding their benefits and limitations and extending their use to different applications within the NCEO. In particular we will develop

- methods to assess the benefit of different observing systems within an assimilation system;
- methods to quantify the impact of planned new satellite instruments;
- a national capability in OSSEs which can be drawn upon by ESA and by the new NERC Centre for Earth Observation Instrumentation in order to decide which new technologies should be developed.

Work description and methodology

A method for assessing the potential benefit of future observations, which has been applied in meteorology and oceanography, is to carry out an observing system assimilation experiment (OSSE). This method uses a ‘nature run’ of an atmospheric or ocean model, on the basis of which simulated observations of both current and future observations are generated. These simulated observations are then assimilated into a different numerical model on the basis of a state-of-the-art assimilation scheme and the added benefit of the future observing system is quantified (Atlas 1997, Arnold and Dey 1986). Such experiments have been widely used to assess the likely impact of new satellite data before a satellite is launched. However, great care must be taken in the design and validation of these experiments to ensure that they give meaningful results (Masutani et al. 2006). For example, if the error properties of the assimilating model do not properly represent mismatches with the nature run, or if error characteristics used for simulating observations are different from the real-world observation error characteristics, then conclusions can be misleading. Nevertheless, when carefully designed and calibrated, results obtainable from OSSEs do clearly outweigh their potential drawbacks. More recently, Tan et al. (2007) proposed a new method for measuring observation impact based on an ensemble technique.

The work on data impact and OSSEs is envisaged here along three lines:

- (i) the design of OSSEs in close collaboration with scientists elsewhere in the NCEO on very specific questions, extending the methodology to different Earth system applications;
- (ii) an investigation of the advantages and shortcomings of OSSEs and the development of methods to overcome the shortcomings;
- (iii) a proactive approach that will make NCEO responsive to invitations-to-tender (ITT) concerning new data and measurement techniques coming from outside agencies (e.g., ESA), as well as to provide OSSE capabilities for training and studying data assimilation techniques. The capability to respond to ITTs will build also on the relevant expertise in other areas within the NCEO.

Work on (i) will initially focus on collaboration with Theme 4, aiming at the assessment of novel observations of differential phase shift from the Met Office polarimetric radar at Thurnham for estimating the location and timing of small-scale precipitation events. Specific questions that will be examined are the impact of indirect precipitation measurements for the high-quality analysis of moist fields and the relative importance of high-quality analyses of the dynamical fields versus moisture analyses (Errico et al. 2004). These questions can also be addressed through an OSE approach by performing data-denial experiments given that the data are available.

The second example of collaborative OSSE work within the NCEO concerns the importance of CO₂ measurements for land-surface data assimilation. The launch of the OCO satellite in 2008 will provide column-integrated measurements of CO₂. In collaboration with Theme 1 we will develop OSSEs to understand the relevance and usefulness of such measurements for the estimation of fluxes in the JULES land-surface model. The usefulness of OCO measurements for estimating parameters in terrestrial carbon models will be studied in collaboration with sub-theme 3.3 of this theme.

In a further example we will use an OSSE-type experiment to assess the potential impact of the recently proposed ESA BIOMASS mission to recover biomass from space using long wavelength radar (P-band), in collaboration with the carbon theme (Theme 2). Using a land surface model developed in that theme we will design a suitable experiment to understand the effect of

radiometric and polarimetric distortion on our ability to recover biomass values from such an instrument.

Work on (ii) will focus on the theory of data assimilation and OSSEs. It is anticipated to carry out this work with simplified atmospheric models that allow extensive simulation at a fundamental level addressing, for example, the extent to which specific targeted observations are useful in predicting hazardous weather, and the role and impact of methodological changes made to the state-of-the-art assimilation system that is used in the OSSE experimentation phase. The specific aim of this work will be to critically examine benefits and limitations of OSSEs with regard to the estimation of observation impact, as well as to study different OSSE methodologies (e.g., Tan et al. 2007). For this work, the creation of a self-contained OSSE system with simplified models, such as quasigeostrophic models or simplified primitive-equation models, is planned.

Work area (iii) is aimed at ensuring that data assimilation and best practice capabilities developed within the NCEO are available to the wider research community. In this context, the creation of a modular OSSE is envisaged that is suitable to be responsive to outside calls asking for the assessment of benefits of new observation types. As an example for the modularity, a widely used nature run for February 1993, together with simulated observations will be used. An important component of this work area will be to interact with the Met Office and the European Centre for Medium-range Weather Forecasting (ECMWF) to test the impact of new research data in their operational systems. The personnel of the Data Assimilation Research Centre (DARC) have over 6 years of experience in running these operational systems, which will be invaluable in ensuring the transfer of knowledge to operational practice. Other DARC experience in performing OSSEs will also be greatly advantageous in carrying out this work (Lahoz et al. 2005, Ehrendorfer 1992).

Deliverables

- OSSE systems for measuring observation impact in storm-scale atmospheric models, land surface models and terrestrial carbon models.
- A simplified OSSE system for theoretical investigations of observation impact and for training.
- An NCEO-wide modular OSSE capability enhancing the responsiveness of NCEO to ITTs.
- Tests of the impact of new and proposed research data in the operational forecast systems of the Met Office and ECMWF.

3.6 Training and Education

An essential part of maintaining a national capability in data assimilation is to address the current skills shortage in this area through the continual training of current and future scientists. This has previously been a core element of the programme of the Data Assimilation Research Centre through involvement in summer schools, PhD and masters programmes and through the provision of web-based tools. There is a requirement both to train current Earth system scientists who wish to apply data assimilation to their models and to train new scientists in this area. In this sub-theme we will continue to train scientists in the theory and practice of data assimilation through the following means:

- *Specialised data assimilation courses for NCEO members.*
In order to aid NCEO scientists embarking on data assimilation projects short, specialised courses of 2-3 days will be developed to teach the basic theory and practice of data

assimilation. The courses will be adapted according to the requirements of those attending, ensuring that the content is directly relevant to the application areas the scientists are working in.

- *Summer schools.*
The Data Assimilation Research Centre has previously contributed to many different national and international summer schools, including a NATO Advanced Study Institute on data assimilation and ESA summer schools. We will set up an annual NCEO summer school, subsuming the highly successful RAL/ Oxford Spring School, which will include training in data assimilation theory and practice. This will be aimed at young scientists of PhD/ early postdoctoral level in the NCEO, but with further places available for other UK and international young scientists. We will aim to run at least one EPSRC/LMS course on the mathematics of data assimilation in order to engage the mathematical community.
- *PhD and PDRA supervision.*
The supervision of PhD students and new PDRA is an essential part of training new scientists in the area of data assimilation and building up the national capability of the NCEO in this area.
- *Provision of postgraduate courses.*
The departments of Mathematics and Meteorology at the University of Reading have developed a joint MSc. in Mathematical and Numerical Modelling for the Atmosphere and Oceans which has been running since October 2001. This contains a module on data assimilation (theory and practice), taught by staff of the Data Assimilation Research Centre. The University of Surrey offers an MSc course in Hydroinformatics, which contains modules on weather prediction and hydrological engineering. We will continue to provide these modules at Reading and Surrey, while seeking ways to make them more widely available to students in other universities as distance-learning courses.
- *Provision of a web-based learning environment.*
Within the Data Assimilation Research Centre a suite of simplified models to illustrate the principles of data assimilation has been made available through a web site. In collaboration with the NCEO Training Manager this concept will be extended to develop new simplified models and assimilation systems relevant to the different application areas of the NCEO. These will be made available through a central web site, together with course notes and suggested exercises, to provide a unique web-based environment for learning about data assimilation and related techniques in Earth observation, e.g. satellite retrieval methods. This resource will be made available to other international training schools such as the summer school run by ESA.

4. OBSERVATIONAL REQUIREMENTS FOR THE WHOLE THEME

The extension of techniques to the different application areas requires a wide range of EO data, dependent on the particular application. This work will be carried out in collaboration with the other themes and observational requirements are covered in the description of those themes. Studies with simplified models will use synthetic data and so have no observational requirements.

5. RESOURCE REQUIREMENTS, JUSTIFICATION AND MANAGEMENT

Staff: The work requires experienced staff with backgrounds in EO, data assimilation, environmental modelling and mathematics. Additional funds will be sought for projects involving CASE students with the Met Office, Rutherford Appleton Laboratory etc. At Oxford and Warwick, where the work will initially be carried out through PhD studentships, provision has been made for the Co-Is to spend time collaborating with others in the Theme. This activity will be an essential part of ensuring effective cross-fertilization of ideas between the mathematics community and the EO community.

Travel: In addition to presentation of work at conferences and workshops, a vital activity of this theme will be to liaise closely with other themes and with operational centres. Therefore provision has been made in the budget for the additional travel that will be necessary to ensure we are a nucleus for data assimilation activity within the NCEO community.

Computing: Most of the work in this theme can be carried out with PC-based systems and software. Some additional resources have been allocated to sub-themes 3.3 and 3.5 where we will work with larger models. The resources required for any implementation work will be provided by the other themes as appropriate.

Management and Reporting:

It is essential that personnel of this theme work together as one team. The research of the PDRAs will be managed by the Co-Investigators, with the Theme Leader, helped by a senior PDRA, acting as research coordinator for the theme. Team meetings for the whole theme will be held every six months. Each PDRA will spend up to 50% of their time working in a particular application area and will also attend theme meetings for the theme they are working most closely with.

For a breakdown of staff/institutions and FTEs over the sub-themes, please refer to the Table in the Annex.

6. KNOWLEDGE TRANSFER

We have a close working relationship with the Met Office and ECMWF through the current Data Assimilation Research Centre. We will engage with these centres to understand their priorities for OSSE-type experiments and to communicate results on OSSEs with research instruments. An important part of knowledge transfer will be through continual interaction with the other parts of the NCEO, ensuring that the techniques developed in this theme can be implemented in the models used in the other themes. By engagement with the U.K. applied mathematics community we will act as a bridge for knowledge transfer from the mathematical theory to the practice of data assimilation. We will redress the shortage of data assimilation skills by means of a comprehensive training programme, providing trained data assimilation scientists to work in different areas of Earth observation.

7. RISK ASSESSMENT

The development of theoretical ideas using simplified models will ensure that there are minimal risks associated with this theme in terms of dependency on other projects or initiatives. Some of

the OSE work in sub-theme 3.5 will depend on the availability of appropriate observation datasets, but in the event of the expected data not being available the resources will be concentrated on OSSE experiments using different systems.

Recruitment of PDRAs and students with a good mathematical background has not always proven easy in recent years, but the departments involved with this theme have excellent track records in postgraduate research.

8. KEY REFERENCES

- Arnold, C. P. and C. H. Dey (1986), Observing-Systems Simulation Experiments: Past, Present and Future. *Bull. Amer. Met. Soc.*, **67**, 687-695.
- Atlas, R. (1997), Observing system simulation experiments: Methodology, examples and limitations. *CGC/WMO Workshop, Geneva 7-9 April 1997, WMO TD No. 868*, World Meteorological Organization, Geneva, Switzerland.
- Bocquet, M. (2005a), Reconstruction of an atmospheric tracer source using the principle of maximum entropy. I: Theory. *Q. J. R. Meteorol. Soc.*, **131**, 2191-2208.
- Cardinali, C., Pezzulli, S. and Andersson, E. (2004), Influence-matrix diagnostic of a data assimilation system. *Q. J. R. Meteorol. Soc.*, **130**, 2767-2786.
- Chipperfield, M.P. (2006), New Version of the TOMCAT/SLIMCAT Off-Line Chemical Transport Model: Intercomparison of Stratospheric Tracer Experiments, *Q. J. R. Meteorol. Soc.*, **132**, 1179-1203.
- Dee, D.P. (2005) Bias and Data assimilation. *Q. J. R. Meteorol. Soc.*, **131**, 3323-3343.
- Crisan, D. and Lyons, T., (1997), Nonlinear filtering and measure-valued processes, *Probability Theory and Related Fields*, **109**, 217-244.
- Ehrendorfer, M. (1992), Four-dimensional data assimilation: comparison of variational and sequential algorithms. *Q. J. R. Meteorol. Soc.*, **118**, 673-713.
- Enting, I.G. (2002), *Inverse problems in atmospheric constituent transport*, Cambridge University Press.
- Errico, R. M., Raeder, K. and Ehrendorfer, M. (2004), Singular vectors for moisture-measuring norms. *Q. J. R. Meteorol. Soc.*, **130**, 963-987.
- Griffith, A. K., and Nichols, N.K. (2001), Adjoint techniques in data assimilation for treating systematic model error. *J. Flow, Turbulence and Combustion*, **65**, 469-488.
- Judd, K. and Smith, L.A. (2004), Indistinguishable states II : imperfect model scenarios. *Physica D*, **196**, 224-242.
- Lahoz, W.A., Brugge, R., Jackson, D.R., Migliorini, S., Swinbank, R., Lary, D. and Lee, A. (2005), An observing system simulation experiment to evaluate the scientific merit of wind and ozone measurements from the future SWIFT instrument. *Q. J. R. Meteorol. Soc.*, **131**, 503-523.
- Lee, T., Boulanger, J.P., Foo, A., Fu, L.-L. and Giering, R. (2000), Data assimilation by an intermediate coupled ocean-atmosphere model: Application to the 1997-1998 El Nino. *J. Geophys. Res.*, **105**, 26,063-26,087.
- Masutani, M. et al. (2006). Observing system simulation experiments at NCEP. *NCEP Office Note 451*, 34 pp.
- Protas, B., Bewley, T. R. and Hagen, G. (2004), A comprehensive framework for the regularization of adjoint analysis in multiscale PDE systems. *J. Comp. Phys.*, **195**, 49-89.
- Rayner, P.J., Scholze, M., Knorr, W., Kaminski, T., Giering, R. and Widmann, H. (2005), Two decades of terrestrial Carbon fluxes from a Carbon Cycle Data Assimilation System (CCDAS). *Global Biogeochemical Cycles* **19**, doi:10.1029/2004GB002254.
- Stuart, A.M., Voss, J. and Wiberg, P. (2004), Conditional Path Sampling of SDEs and the Langevin MCMC Method. *Communications in Mathematical Sciences*, **2**, 685-697.
- Tan, D.G.H., Andersson, E., Fisher, M. and Isaksen, L. (2007) Observing-system impact assessment using a data assimilation ensemble technique: application to the ADM-Aeolus wind profiling mission, *Quart. J. Roy. Met. Soc.*, **131**, 381-390.
- Tremolet, Y. (2007), Model error estimation in 4D-Var, ECMWF Tech Memorandum 520.
- Watkinson, L.R. (2006), *Four-dimensional variational data assimilation for Hamiltonian problems*. PhD thesis, Department of Mathematics, University of Reading, U.K.
- Williams, M., Schwarz, P.A., Law, B.E., Irvine, J. and Kurpius, M.R. (2005), An improved analysis of forest carbon dynamics using data assimilation. *Global Change Biology*, **11**, 89-105.

ANNEX

1. Interactions across NCEO

The research in this Theme (T7) will interact vitally with the other themes in the main NCEO proposal. These themes are:

- T1:** EO for climate diagnosis and prediction
- T2:** Monitoring, diagnosis, re-analysis and prediction of the global carbon cycle
- T3:** Atmospheric composition: air quality and climate
- T4:** High resolution predictions of hazardous weather, floods and water resources
- T5:** Cryosphere and polar oceans
- T6:** Dynamic earth and geo-hazards

The sub-themes in this Theme (T7) link to specific projects in the other themes as follows:

- *3.1: Analysis of coupled and multi-scale systems.*
Theme 1: coupled atmosphere-ocean models, land-atmosphere models, ocean physics-biology models.
Theme 2: ocean physics-biology models.
Theme 3: coupled atmosphere-chemistry models.
Theme 4: nested storm-scale models.
Theme 5: sea-ice interaction models.
- *3.2: Quantifying and representing uncertainty.*
A good representation of the uncertainty in data and models is a key foundation of any assimilation system and the work in this sub-theme will link to all areas of the NCEO where data assimilation techniques are being used.
- *3.3: Using observations to estimate model parameters.*
Theme 1: land surface models.
Theme 2: terrestrial carbon models.
Theme 3: chemical transport models.
Theme 4: hydrological models.
Theme 5: glacial sheets and glacial flow models.
Theme 6: models of volcanic activity.
- *3.4: Estimating surface fluxes using data assimilation.*
Theme 1: estimation of CO₂ fluxes.
Theme 2: estimation of CO₂ fluxes.
Theme 3: estimation of CH₄ fluxes.
- *3.5 Quantifying observation impact.*
This sub-theme will link to all other themes where observation impact experiments may

be useful, but taking as particular examples:

Theme 1: Usefulness of OCO measurements to estimate parameters in JULES.

Theme 2: OSSEs for BIOMASS measurements and for using OCO measurements to estimate carbon model parameters.

Theme 4: OSSEs for radar measurements.

- *3.6 Training.*

Training will be made available to all personnel in the NCEO working with data assimilation systems.

2. Staffing

The distribution of staff resources between the different sub-themes is shown in the table. All the Co-Is will also contribute to sub-theme 3.6 on training.

	ST 3.1	ST 3.2	ST 3.3	ST 3.4	ST 3.5
Reading					
Ehrendorfer		X			X
Lawless*	2 FTE	2 FTE	0.25 FTE	0.25 FTE	
Nichols**	0.5 FTE	0.5 FTE			
PDRA 1	3 FTE				
PDRA 2					3 FTE
Surrey					
Roulstone	X	X		X	
PDRA				3 FTE	
Bristol					
Scholze			X		
PDRA			3 FTE		
Oxford					
Lyons		X			
Norbury	X	X	X		
Warwick					
Stuart		X		X	

*Lawless will also contribute 0.25 FTE to sub-theme 3.6 and 0.25 FTE to act as Research Coordinator for the theme.

**Nichols will also contribute 0.25 FTE to sub-theme 3.6.