

# Flying the Satellite into Your Model

Thomas Kaminski and Pierre-Philippe Mathieu

## Abstract

The vehicles that fly the satellite into a model of the Earth System are *observation operators*. They provide the link between the quantities simulated by the model and quantities observed from space, either directly (spectral radiance) or indirectly estimated through a retrieval scheme (bio-geophysical variables). By doing so, observation operators enable modellers to properly compare, evaluate and constrain their models with the model-analogue of the satellite observations. This paper provides the formalism and a few examples of how observation operators can be used, in combination with data assimilation techniques, to better ingest satellite products in models in a dynamically consistent manner. It describes communalities and potential synergies between assimilation and classical retrievals. The paper explains how the combination of observation operators and their derivatives (linearisations) form powerful research tools. It introduces a technique called automatic differentiation that greatly simplifies both development and maintenance of derivative code.

## Index Terms

IEEEtran, journal, L<sup>A</sup>T<sub>E</sub>X, paper, template.  
Draft: January 19, 2012

## I. INTRODUCTION

Earth System Models (ESMs) are complex software capturing our knowledge of how the ocean, atmosphere, land and ice operate and interact. ESMs provide scientists with powerful tools to better understand our global environment, its evolution, and the potential impact of human activities (e.g. analyses of relevant processes, their interaction and feedback mechanisms). ESM applications range from numerical weather prediction (NWP) over seasonal forecasting [see, e.g., *Stockdale et al.*, 2011] to climate projections on centennial scale [*Solomon et al.*, 2007] or longer [*Jungclaus et al.*, 2010].

Before being used for predictions, ESMs and their components should be confronted with observations in order to assure their realism (validation). Such validation procedures can be extended to standardised assessments of model performance in so-called benchmarking systems [see, e.g., *Blyth et al.*, 2011]. This involves the definition of metrics that quantify the model performance through the fit to observations. A further step towards the rigorous use of the observations is their ingestion in formal data assimilation procedures, e.g. to constrain the model's initial state (initialisation) or tunable parameters in the model's process representations (calibration).

T. Kaminski is with FastOpt, Hamburg, e-mail: Thomas.Kaminski@FastOpt.com.

P.P. Mathieu is with the European Space Agency, .

Manuscript received .. ; revised ...

Such confrontation with observations is hampered by the fact that observed and modelled quantities typically differ in nature or scale. For example, a flask sample of the atmospheric carbon dioxide concentration provides a value at a specific point in space and time, whereas an atmospheric tracer model operates in a discretised representation of space and time, i.e. on values that refer to a box in the atmosphere and a particular period of time. Any comparison of the two quantities (modelled and observed) must, hence, take the uncertainty arising from this representation error into account [see, e.g., *Heimann and Kaminski, 1999*]. Another example is a vertical profile of the ocean temperature and salinity provided by a floating buoy (ARGO, see <http://www.argo.ucsd.edu>). Again the spatial scales of the observation and the model do not match (in the horizontal dimension). In addition, ocean models are formulated in terms of potential temperature rather than temperature. Since we can only compare quantities of the same nature, some form of transformation is required before any comparison can take place. Such a difference in nature is intrinsic to observations from space, where the raw quantities measured by satellites, i.e. spectral radiance (photon counts [*Mathieu and O'Neill, 2008*]) are by nature only indirectly (through radiative transfer processes) related to the model quantities of interest.

The link from the model to the observations is provided through a set of relationships expressed in terms of an *observation operator*. We can think of an observation operator as an arm, which enables the ESM to access a particular type of observation (see Figure 1). Observation operators are ranging in complexity from a simple interpolation or integral scheme up to a chain of sophisticated non-linear radiative transfer models.

The layout of the remainder of this paper is as follows. Section II introduces the concept of an observation operator and presents examples. The role of observation operators in applications is presented in section III. Section IV highlights the use of derivatives of observation operators and introduces *automatic differentiation*, a technique to provide these derivatives. Finally, section V draws conclusions.

## II. OBSERVATION OPERATORS

### A. Definition

Mathematically the observation operator is defined as a mapping  $H$  from the vector of state variables  $s$  onto the vector of observed radiances  $d$ :

$$H : s \mapsto d \quad (1)$$

The vector of state variables (*state vector*) defines the simulated system for a given time step at all points in space, and the evolution of the system is described by a sequence of state vectors, forming a trajectory through the state space. The state variables are also called prognostic variables, to contrast them with diagnostic variables, which are computed from the state and evolve only indirectly through the evolution of the state. For example the albedo of the land surface is diagnosed from the state of the vegetation-soil system. Hence, if we want to change the trajectory of the model (e.g. to improve the fit to observations), we must arrange for a change of the state. The model will then propagate this change of state forward in time. This means, to bring observational information into the model, we

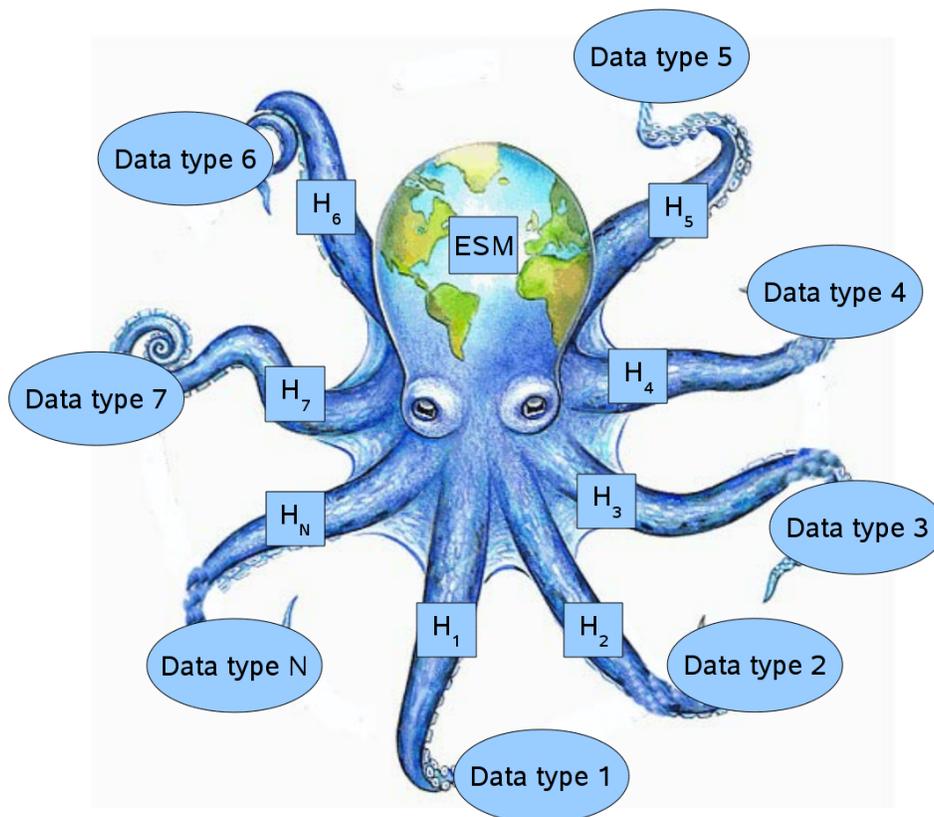


Fig. 1. Schematic of an ESM assessing several of data types via observation operators  $H_1, \dots, H_n$ .

must link the observations to the state: In other words the state vector constitutes the interface between the model and the observation operator.

The solid path in Figure 2 sketches how an observation operator ( $H_1$ ) enables the comparison of simulated and observed values at the sensor level, i.e. at the level of spectral radiances, typically referred to as level 1 data products [Arvidson *et al.*, 1986].

Another way to make EO data accessible to dynamical models is by feeding the satellite observations into a retrieval algorithm and so derive a biophysical variable. Such EO products are usually called level 2 data products. Internally, the retrieval algorithm also relies on a functional relationship that maps the biophysical variable(s) of interest onto the spectral radiance. This mapping is similar, if not identical to the observation operator  $H_1$ , although the term used by the EO community is *forward model*. The retrieval can be regarded as an inversion of  $H_1$ . As the examples below will illustrate, the retrieved level 2 product will typically not exactly coincide with a component

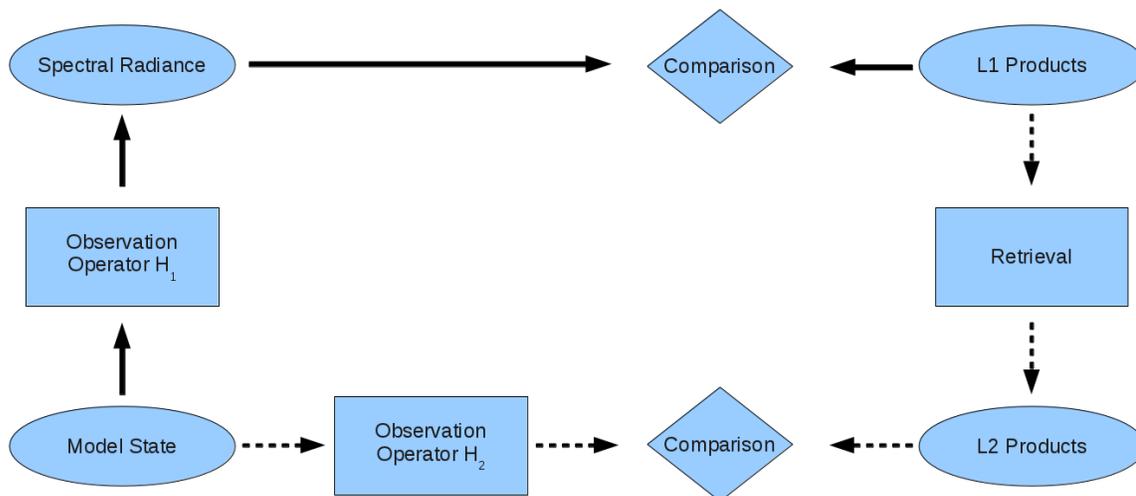


Fig. 2. Model-Data Comparison at the sensor level (Level 1, solid arrows) and at the level of biophysical variables (Level 2, dotted arrows). Ovals denote data, rectangulars some form of processing.

of the model state vector. Hence, the confrontation of level 2 data with the model (dotted path in figure 2) also requires an observation operator (denoted by  $H_2$ ).

### B. Examples

Figure 3 attempts to sketch a generic observation operator  $H_1$ , which links a model's state vector to observed spectral radiance. For the sake of clarity the figure focuses on processing steps that map one variable onto another and omits further important steps that involve transformations in space and time, i.e. interpolation, averaging, or orbit simulation.

The simulation of spectral radiances at the sensor level requires information from the atmosphere and the land/ocean surface, including covers by ice or snow. Hence, the observation operator typically consists of various modules. First, from the model state the relevant electromagnetic signatures are simulated. For example, for a passive optical sensor observing the terrestrial vegetation this would be the reflected sun light, and it would be computed by a model of the radiative transfer within the canopy, for examples see *Pinty et al.* [2006] or *Knorr et al.* [2010]. For a passive microwave sensor that observes sea ice and snow, this would be the thermal emission, and it would involve a model of the radiative transfer within the sea ice and snow pack [see, e.g., *Wiesmann and Mätzler*, 1999; *Tonboe et al.*, 2006]. In the atmosphere this could be a model for the emissivity of clouds as a function of the atmospheric state. The next step covers the path through the atmosphere from the observed components to the sensor and requires a model of the radiative transfer through the atmosphere. Prime examples are the Radiative

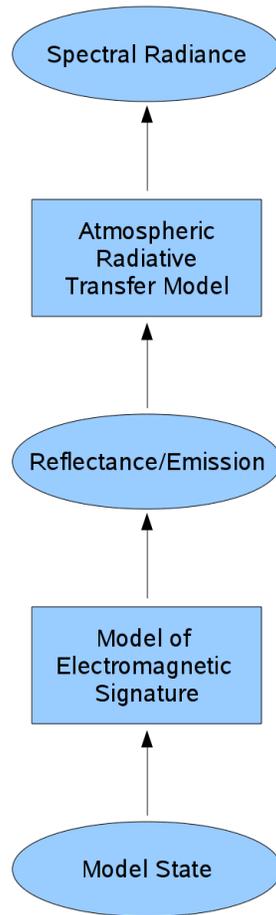


Fig. 3. Generic scheme of an observation operator for spectral radiance. Oval boxes denote data, rectangular boxes denote processing.

Transfer for Tiros Operational Vertical Sounder (RTTOV, [Eyre, 1991; Saunders *et al.*, 1999]) for the microwave and infra-red domain or 6S [Vermote *et al.*, 1997] for the solar domain. The output of the radiative transfer model can be compared with a level 1 product.

Each type of observation requires its own observation operator in order to be accessible to models. The complexity of the observation operator typically reflects a compromise between the accuracy required for the application at hand and the available computational resources. In a space mission, the observation operator depends on characteristics such as the orbit of the platform or the measuring principle and, thus, spectral sensitivity of the sensor. The observation operator also depends on the dynamical model. One aspect is the state space, which depends on the model formulation. For example, an atmospheric model can either diagnose clouds or include them in the state space [Chevallier *et al.*, 2004]. In the former case the diagnostic cloud model is part of the observation operator

in the latter it is not. Even though parts of an observation operator are usually model-dependent, it is desirable to implement the observation operator in a modular form with carefully designed interfaces. This modularity maximises the flexible use for assimilation and retrievals and the adaptation to new models or observations, i.e. it assures multi-functionality.

The crucial role of observation operators is reflected in comparison exercises such as the radiation transfer model intercomparison (RAMI) initiative for the transfer of radiation in plant canopies and over soil surfaces [Pinty *et al.*, 2001; Widlowski *et al.*, 2007]. A similar activity for the atmosphere is the international Intercomparison of 3D Radiation Codes (I3RC) project [Cahalan *et al.*, 2005]. The I3RC focuses on the interaction of solar and thermal radiation with cloudy atmospheres. Another activity in this domain is the Cloud Feedback Model Intercomparison Project (CFMIP) which has set up the CFMIP Observation Simulator Package (COSP) [Bodas-Salcedo, 2011]: The modular package includes a set of observation operators that map model output consisting of “vertical profiles of temperature, humidity, hydrometeor (clouds and precipitation) mixing ratios, cloud optical thickness and emissivity, along with surface temperature and emissivity” onto a set of level 2 products retrieved from “the following instruments: CloudSat radar, CALIPSO lidar, ISCCP, the MISR, and the Moderate Resolution Imaging Spectroradiometer (MODIS)”. The above-mentioned “fast radiative transfer code RTTOV can also be linked to COSP to produce clear-sky brightness temperatures for many different channels of past and current infrared and passive microwave radiometers.” Not only does COSP greatly simplify the comparison of model output with EO products. Using standardised interfaces it allows to compare multiple models through the same observation operators with the same EO data, and thus facilitates the attribution of a model-data mismatch to aspects of the model, the observation operator, or the observations.

### III. APPLICATIONS OF OBSERVATION OPERATORS

This section starts with an introduction of the formalism behind advanced data assimilation and retrieval schemes. The details of the formalism are useful to understand the application examples in this section and the need for derivative information that is discussed in section IV.

#### A. Formalism of Data Assimilation and Retrieval

Data assimilation is a procedure to combine the information from observations with the information in a dynamical model. There is a range of data assimilation techniques with varying degree of sophistication. The simplest techniques try and replace the model state by the observed state, or, more precisely, some average of the two. More advanced approaches can assimilate observations  $d$  which are linked to the state through an observation operator  $H$ .  $H$  can be an observation operator for in situ data or for EO data, for example the operators  $H_1$  and  $H_2$  introduced in section II (see equation (1) and figure 3)). The assimilation problem is typically formalised as a minimisation problem for a misfit function

$$J(x) = \frac{1}{2} (H(x) - d)^T \mathbf{C}_d^{-1} (H(x) - d) + \frac{1}{2} (x - x_{pr})^T \mathbf{C}_{xpr}^{-1} (x - x_{pr}) \quad , \quad (2)$$

where we introduce the symbol  $x$  for the vector of unknowns instead of reusing the above defined symbol for the state,  $s$ . This is more convenient for later use where  $x$  is more general than the state.

The function  $J(x)$  is composed of two terms. The first term quantifies the misfit between the observations and their simulated counterpart (observational term). The second term quantifies the deviation of the model state from the prior information  $x_{pr}$  (prior term often also called background). Both terms are weighted in inverse proportion to the respective uncertainties, i.e. the combined uncertainty in the observations and observation operator,  $\mathbf{C}_d$ , and the uncertainty in the prior information,  $\mathbf{C}_{xpr}$ . The superscript  $T$  denotes the transposed. Note that the equation does not require the observations to be provided in the space time grid of the model. The observations can come in any spatio-temporal distribution, e.g. the above mentioned point measurements or orbits, as long as we can formulate the appropriate observation operator.

Equation (2) formalises what in numerical weather prediction is called three dimensional variational assimilation (3D-Var) [Courtier *et al.*, 1998], more precisely its analysis step, which is then followed by a forecast step. Operationally the assimilation scheme is run in cycling mode through these two steps. In such a cyclic assimilation scheme, the prior information is provided by the previous forecast, i.e. it is consistent with the dynamical information from the model, and at the same time suffers from errors in the model.

The model dynamics are even more emphasised when the scheme of equation (2) is extended to contain observations  $d_i$  at different time steps ( $i = 1, \dots, n$ ) to constrain the initial state  $s_0 = x$  through:

$$J(x) = \frac{1}{2} \sum_{i=1,n} (H(s_i(x)) - d_i)^T \mathbf{C}_{d_i}^{-1} (H(s_i(x)) - d_i) + \frac{1}{2} (x - x_{pr})^T \mathbf{C}_{xpr}^{-1} (x - x_{pr}) \quad (3)$$

This is the setup of the analysis step in four dimensional variational assimilation (4D-Var) schemes, where the dynamical model  $M$  is used as a constraint that links the states at all observation times via

$$s_{i+1} = M(s_i) \quad (4)$$

to the initial state  $s_0 = x$ . For convenience the notation suppresses the time dependent nature of  $H$  and  $M$ , and it also assumes that the data uncertainties at different time steps are uncorrelated. While 4D-Var solves a single minimisation problem to find a (dynamically consistent) model trajectory, 3D-Var is a sequential approach, i.e. it solves a sequence of minimisation problems, which yield a dynamically inconsistent sequence of model states.

In the 4D-Var approach, the vector of unknowns  $x$  can be extended from the initial state to boundary values and process parameters (model calibration). Since these are external controls to the dynamical system,  $x$  is also called control vector, a term taken from control theory [Lions, 1971]. We usually try to select the control vector such that it comprises the fundamental unknowns of the system at hand.

The *Kalman Filter* is another sequential approach. Its analysis step solves a slightly simplified form of equation (2), in which  $H$  is replaced by its linearisation  $\mathbf{H}'$  (Jacobian matrix) around the prior. This allows an analytic solution  $x_{po}$  of equation (2):

$$x_{po} = x_{pr} - \mathbf{C}_{xpo} \mathbf{H}'^T \mathbf{C}_d^{-1} (\mathbf{H}' x_{pr} - \mathbf{d}_i) \quad (5)$$

the evaluation of which involves the inversion of the (typically high dimensional) matrix

$$\mathbf{C}_{\mathbf{x}_{po}} = (\mathbf{H}'^T \mathbf{C}_d^{-1} \mathbf{H}' + \mathbf{C}_{\mathbf{x}_{pr}}^{-1})^{-1} \quad (6)$$

which expresses the uncertainty range in  $x_{po}$  that is consistent with uncertainty ranges in the data and the prior values.

In case of linear  $H$  and Gaussian probability densities for the prior and the data, the solution of equation (2) is Gaussian as well, and completely described by its mean (equation (5)) and covariance (equation (6)). This formalism is, for example, applied in inverse modelling of the atmospheric transport of carbon dioxide, where an atmospheric transport model takes the role of  $H$ , and the space-time distribution of the surface fluxes takes the role of  $x$ . Note that the cost function's second derivative (Hessian matrix)  $\mathbf{J}''$  is related to  $\mathbf{C}_{\mathbf{x}_{po}}$  through:

$$\mathbf{C}_{\mathbf{x}_{po}} = \mathbf{J}''(x)^{-1} \quad (7)$$

In the non-linear case we cannot solve equation (2) or equation (3) analytically, but via the cost function's Hessian we can use equation (7) to approximate  $\mathbf{C}_{\mathbf{x}_{po}}$ . Via a linearisation  $\mathbf{N}'$  of the model that links the control variables to model outputs of interest  $f$  we can approximate the uncertainty range of these model outputs  $\mathbf{C}_f$  by:

$$\mathbf{C}_f = \mathbf{N}' \mathbf{C}_{\mathbf{x}_{po}} \mathbf{N}'^T \quad (8)$$

The alternative to the above assimilation approaches are ensemble methods such as Markov Chain Monte Carlo [see, e.g., *Metropolis et al.*, 1953] or Ensemble Kalman Filter [*Evensen*, 2003] techniques, which rely on forward simulations to sample the control space. The feasible ensemble size is limited by the computational demands which essentially reflect the complexity of the underlying model.

We used equation (2) to introduce the formalism of data assimilation. The *same* equation also plays a central role in retrievals. Minimisation of equation (2) describes a retrieval algorithm for the entire state. The prior term regularises what is otherwise an underdetermined inverse problem: Several of the unknown variables that influence the observed signal vary continuously with altitude (continuous vertical profiles). Even though we formulate our observation operators on a vertical grid, there are typically fewer measurements than unknowns. Consequently there are many sets of unknown variables that yield an equal fit to the observations. The prior term provides additional information on every unknown and helps the retrieval algorithm to find a unique solution. Further equation (6) or equation (7) are used to furnish the retrievals with uncertainty ranges.

Another perspective on the assimilation of level 1 data is to regard it as an advanced form of retrieval, and the assimilation system as an advanced retrieval algorithm that optimally combines the information from remote sensing, radiative transfer and dynamical model. The other point to note is that  $H$  is usually not constant in space and time. For instance the radiative transfer in the optical domain is affected by atmospheric water vapour and aerosols. A retrieval of, say, a land surface variable requires information on clouds and aerosols. In a coupled atmosphere-land model these are available in a form that is dynamically consistent with the state of the land surface but, on the other hand, also affected by errors in the model.

### B. Data Assimilation examples

The prime example of an atmospheric 4D-Var system is the one operated at the European Centre of Medium-Range Weather Forecasts (ECMWF) [Rabier *et al.*, 2000]. Currently about 98% of the assimilated observations are remotely sensed, and their majority are level 1 data [IFS, 2010]. Sensors include ATOVS, AIRS, IASI, SSMI, and AMSR-E as well as geostationary water vapour clear-sky and rain affected SSMI radiances, together with the appropriate observation operators [IFS, 2010]. The system is used with a 12 hour assimilation window to initialise the operational forecast. Several other weather services (including those of Canada, France, and the UK) are running similar 4D-Var systems.

A prominent example of a variational ocean assimilation system was set up by the ECCO consortium (see <http://www.ecco-group.org>) around the MITgcm [Marshall *et al.*, 1997]. The system [Stammer *et al.*, 2002] uses a combination of in-situ observations and level 2 remote sensing products (including sea surface height, sea surface temperature, wind-stress, and geoid) for decadal-scale assimilation windows [Wunsch and Heimbach, 2006]. Owing to these long assimilation windows the prescribed exchange fluxes with the atmosphere are a major source of uncertainty in their model trajectory. Hence, this boundary condition is included in the control vector along with the initial state. Various applications of the assimilation product require closed property budgets over the entire assimilation period which are achieved via variational approaches in contrast to sequential approaches. Examples are mechanistic or diagnostic studies of climate variability or oceanic tracer transport problems [Wunsch *et al.*, 2009].

A recent example of a regional variational assimilation system for the coupled ocean sea-ice system in the Northern latitudes was developed by Kauker *et al.* [2009]. This system is operated for assimilation windows from a few months to a few years. Their control vector combines (depending on the application) the initial state, boundary conditions and process parameters. The system is constrained by in-situ observations and level 1 products of sea ice concentration, thickness, and displacement.

An example for the global terrestrial vegetation is provided by the Carbon Cycle Data Assimilation System (CCDAS). Initially set up for the assimilation of in-situ observations of the atmospheric carbon dioxide concentration [Rayner *et al.*, 2005], the system was extended with observation operators for two types of level 2 products, namely Fraction of Absorbed Photosynthetically Active Radiation (FAPAR) products [Knorr *et al.*, 2010] and for the column-integrated atmospheric carbon dioxide concentration (XCO<sub>2</sub>) [Kaminski *et al.*, 2010b]. The observation operator for FAPAR was a considerable extension to the previous system, because it required modules for the simulation of vegetation phenology and of hydrology, which were previously provided by an off-line calculation. The observation operator for XCO<sub>2</sub> consists of a model of the atmospheric transport that solves the continuity equation for carbon dioxide [Heimann and Körner, 2003]. The assimilation window ranges from years to decades. Considering uncertain values of the parameters (constants) in process formulations as the major source of uncertainty in the model trajectory, the control vector is composed of (depending on the setup) in the order of 50-100 (in extreme cases up to 1000) process parameter values. This type of application is called parameter estimation or

model calibration.

### C. Retrieval Examples

The integrated retrieval of *Toudal* [1994] or *Melsheimer et al.* [2009] solves simultaneously for biophysical variables (level 2 data) of the atmosphere (wind speed, total water vapour, cloud liquid water) the ocean (sea surface temperature) and the sea ice (ice surface temperature, total sea ice concentration, multi-year ice fraction). Technically they use equation (5) in an iterative procedure ( $x_{po}$  from one step is provided as  $x_{pr}$  to the subsequent step), which recomputes  $\mathbf{H}'$  by linearisation around the current  $x_{pr}$ . Upon convergence they deliver posterior uncertainties via equation (6). Their input are radiances (brightness temperatures) observed by the Advanced Microwave Scanning Radiometer for EOS (AMSR-E). Their prior values are taken from a range of sources including analysis data provided by an NWP assimilation system, or separate univariate retrievals. This integrated retrieval is performed individually for each observed point in space and time, at 12.5 km horizontal resolution. Though the use of the same level 1 data in an assimilation system (assuring dynamical consistency between the atmosphere, ocean and sea ice components) appears desirable, it is highly challenging in various respects: From a software development perspective, because it would require an assimilation system built around a coupled atmosphere, ocean and sea ice model. From a computational perspective, because a single run of the coupled model at 12.5 km resolution is already computationally expensive, let alone an iterative assimilation scheme.

Another example is the Joint Research Centre-Twostream Inversion Package (JRC-TIP) [*Pinty et al.*, 2007], which solves equation (2) for model parameters controlling the radiation transfer regime in vegetation canopies, namely the effective Leaf Area Index (LAI) and the spectral scattering properties of the vegetation and the soil. The latter information is then used to estimate the spectral fluxes scattered by, absorbed in and transmitted through the vegetation layer as well the fluxes absorbed in the background (radiant fluxes). Further the system uses equation (7) to infer the uncertainty in the retrieved parameters and equation (8) to propagate these forward to uncertainties in the simulated radiant fluxes. In its typical setup the system uses observed albedos in two broad wavebands (visible and near infrared) [*Pinty et al.*, 2007, 2011a, b].

The JRC-TIP is constructed around a one dimensional two stream model, which takes three-dimensional radiative transport effects into account [*Pinty et al.*, 2006]. As a consequence the retrieved vegetation parameters are effective parameters (i.e. their values are only meaningful within this model) and are determined such that the radiant fluxes are simulated as accurately as possible. This illustrates a crucial point when confronting retrieved level 2 variables with their ESM counterparts: One has to make sure that the variables have the same meaning in the forward model that is used for the retrieval and the observation operator that is used for its assimilation. For the JRC-TIP products this is the case for the radiant fluxes and soil parameters, for the effective vegetation parameters it requires the use of the same two-stream model in the observation operator.

A serious practical difficulty in data assimilation is the specification of  $\mathbf{C}_d$ . In the case of level 2 data  $\mathbf{C}_d$  is the posterior uncertainty of the retrieval. Since the retrieval is typically carried out point by point, uncertainty correlations in space and time are difficult to assess. Another issue is the data volume required by the uncertainty

information: For a product of a retrieved biophysical variable at  $n$  points in space and time  $\mathbf{C}_d$  contains (taking its symmetry into account)  $n*(n+1)/2$  different values, a volume that is usually prohibitive in real world applications. The challenge is to develop ways of providing (approximations of)  $\mathbf{C}_d$  that retain the essential information in a minimal data volume. The contribution by the observational uncertainty  $\mathbf{C}_d$  is certainly easier to specify in the case of level 1 data. Their direct assimilation automatically propagates, through the observation operator  $H_1$ , the full information content of  $\mathbf{C}_d$  into the model.

#### *D. Observing System Simulation Experiments and Quantitative Network Design*

Observing System Simulation Experiments (OSSEs) and Quantitative Network Design (QND) are two methodologies that rely on assimilation systems. By an observing system or observational network we understand the superset of all observations that are made available to an assimilation system.

An OSSE (for an overview see *Böttger et al.* [2004]) uses a model plus observation operators to simulate model analogues of observations that would be collected by a potential observing system (often the current observing system extended by a potential new data stream). The model is also used to simulate a surrogate of reality, i.e. a reference trajectory over the period of investigation. Then an assimilation/forecast system (often built around the same model) is used to evaluate some measure of the performance of the potential observing system and its subsystems. In NWP, the performance of an observing system is usually quantified by the quality (skill) of a forecast from the initial value that was constrained by the observation system. Via this procedure one can, for example, assess the added value of a planned mission in terms of an increment in forecast skill.

QND (for an overview see *Kaminski and Rayner* [2008]) relies on the ability of an assimilation system to evaluate posterior uncertainties on target quantities of interest via equation (7) and equation (8). For a linear model, this propagation of uncertainty is independent of the observational value, it just depends (via equation (7) and equation (3)) on the data and prior uncertainties, the sensitivity of the observations with respect to the control variables and (via equation (8)) on the sensitivity of the target quantity to the control variables. A first application to mission design was presented by *Rayner and O'Brien* [2001], who ran an inversion system built around a linear model of the atmospheric transport of carbon dioxide in QND mode. They assessed the utility of remotely sensed carbon dioxide in constraining its surface fluxes. Their benchmark was the in-situ flask sampling network. *Kaminski et al.* [2010b] generalised the method to the above mentioned CCDAS, and assessed the utility of XCO<sub>2</sub> observations by an active LIDAR instrument. The performance of the observing system is quantified by posterior uncertainty of surface fluxes and compared to the performance of the in-situ network. *Kaminski et al.* [2010a] use CCDAS to assess the performance of potential optical sensor configurations in constraining the vegetation's carbon and water fluxes. Their benchmark was the MERIS sensor.

For both approaches, the importance of suitable observation operators is obvious. A disadvantage is that the result depends on the model in the assimilation system. Both techniques require the specification of data uncertainties for the hypothetical data streams to be evaluated.

#### IV. DERIVATIVES OF OBSERVATION OPERATORS

This section first summarises how the capability to evaluate derivatives of the observation operator is used in efficient schemes for retrieval, assimilation or QND and then introduces a technique for providing derivative information.

In variational assimilation, equation (2) or equation (3) are typically minimised in an iterative procedure that varies  $x$ . To do this efficiently so-called gradient algorithms are employed. They rely on the capability of evaluating the gradient of  $J$  with respect to  $x$  to define a search direction in a (usually) high-dimensional space of unknowns. The gradient is useful, because it yields the direction of steepest ascent. For equation (2) the gradient takes the form

$$\nabla J(x) = \mathbf{H}'(x)^T \mathbf{C}_d^{-1} (H(x) - d) + \mathbf{C}_{\mathbf{xpr}}^{-1} (x - x_{pr}), \quad (9)$$

and we see that its evaluation requires the capability to multiply the transposed of  $\mathbf{H}'$  with a vector. The uncertainty estimation via equation (7) based on  $\mathbf{J}''$  requires, in addition, second derivative information on  $H$ . This second derivative expresses the curvature of (the components) of  $H$ , i.e. the change of its linearisation corresponding to a unit change of  $x$ .

Likewise the Kalman filter requires derivatives of  $H$ : In equation (5) it multiplies the matrix  $\mathbf{H}'$  and its transposed with vectors, and for the evaluation of equation (6) it needs to invert a matrix that contains  $\mathbf{H}'$  and its transposed. One can do this inversion by precomputing  $\mathbf{H}'$  or by so-called matrix-free methods that repeatedly multiply  $\mathbf{H}'$  and its transposed with vectors.

As mentioned, advanced retrieval algorithms are based on the same equations, i.e. they typically solve equation (2) either via gradient methods, or via equation (5), and use either equation (7) or its approximation equation (6) to derive uncertainties. Hence, they benefit in the same way on derivatives of  $H$  as data assimilation systems. The same holds for QND schemes which also rely on uncertainties via equation (7) or its approximation, equation (6).

Traditionally derivatives were approximated by multiple forward runs (finite difference approximation) [see, e.g., *Toudal, 1994; Melsheimer et al., 2009*]. This discretised procedure has two disadvantages: The first is the limited accuracy which degrades the performance of the above listed algorithms. For example, incorrect gradient information will slow down or prematurely stop the iterative minimisation of  $J$ . The other disadvantage is that the computational cost of this approximation grows linearly with the length of the control vector.

Both disadvantages can be avoided by Automatic differentiation (AD [*Griewank, 1989*]). AD is a procedure which generates source code for evaluation of derivatives from the code of the underlying function. In the current case this function is the observation operator mapping the state variables onto remote sensing products. The function code is decomposed into elementary functions (such as  $+$ ,  $-$ ,  $\sin(\cdot)$ ), for which the derivative (local Jacobian) is straightforward to derive. The derivative of the composite function is then constructed via the chain rule as the product of all local Jacobians. According to the associative law, this multiple matrix product can be evaluated in arbitrary order without changing the result. The *tangent linear* code (or just tangent code) does this evaluation in the same order as the function is evaluated, which is called forward mode of automatic differentiation. The *adjoint*

code uses exactly the opposite order, which is called reverse mode of automatic differentiation. Even though both modes yield the same derivative, depending on the dimensions of the function to be differentiated, there may be large differences in their computational efficiency: The CPU time required by tangent code is proportional to the number of the function's input variables but independent on the number of output variables. By contrast, the CPU time required by the adjoint code is proportional to the number of output variables and independent of the number of input variables. Both the tangent and adjoint codes use values from the function evaluation (*required values*, see *Giering and Kaminski [1998]*). Providing required values to the adjoint code is more complicated than to the tangent code. Being an application of the chain rule, AD provides derivatives that are accurate up to rounding error.

For variational assimilation we require the derivative of the scalar-valued cost function  $J(x)$  of equation (2) or equation (3) with respect to a usually high-dimensional vector  $x$ . For a state-of-the-art model, only the adjoint can provide this derivative with sufficient efficiency. A product  $\mathbf{H}'v$  of  $\mathbf{H}'$  with a vector  $v$  yields the directional derivative of  $\mathbf{H}'$  in the direction defined by  $v$ , i.e. the derivative of the function  $H(x + tv)$  of a scalar unknown  $t$ . Hence, this type of product is evaluated most efficiently in forward mode, i.e. by the tangent linear code of  $H$ . By contrast a product of the form  $\mathbf{H}'^T v$  is the (transposed of the) derivative of the scalar valued function  $v^T H(x)$ , which is evaluated most efficiently in reverse mode, i.e. by the adjoint of  $H$ . The scalar forward and reverse modes required for efficient evaluation of the above Jacobian-vector products are the standard forms of derivative code. The scalar mode is contrasted by the vector mode. In forward mode the vector mode simultaneously computes the sensitivities with respect to multiple input quantities, and in reverse mode simultaneously the sensitivity of multiple output quantities. Experience shows that the vector mode is considerably more efficient than multiple runs in scalar mode [see, e.g., *Kaminski et al., 2003*]. We use the vector mode for applications that require the entire Jacobian,  $\mathbf{H}'$ . Here the sensible choice between forward and reverse modes depends on the relative dimensions of state and observation spaces.

A particular advantage of AD is that it can guarantee readability and locality [*Talagrand, 1991*], i.e. every statement in the derivative code belongs to a particular statement in the function code. As a consequence, if the function code is modular, the same modularity is preserved in the derivative code. Another distinct advantage of the AD approach is that it simplifies the maintenance of the derivative code, because it can be quickly updated after any modification of the function code.

Since an AD tool operates at the code level, it is restricted to a particular programming language. Luckily there are AD tools for the most frequently used programming languages in Earth System Science, namely Fortran and C. It is a considerable effort to develop and maintain an AD tool up to a level robust enough for relevant scientific applications. Over the last decade, tool development has made good progress and there is a long list of successful AD applications to component models of the Earth System. Meanwhile AD can generate a highly efficient adjoint of a general circulation model [*Heimbach et al., 2005*] with more than 100.000 code lines (excluding comments) in less than a minute. The capability of running on parallel architectures is preserved [see, e.g., *Heimbach et al., 2005*; *Giering et al., 2005*].

The AD process can be mimicked by hand [see, e.g., *Rabier et al., 2000*; *Weaver et al., 2003*; *Moore et al., 2004*;

*Kleespies et al.*, 2004; *O'Dell et al.*, 2006; *Barrett and Renzullo*, 2009], i.e. a human transforms the function code line by line into derivative code following the same recipes [*Giering and Kaminski*, 1998] that are implemented in AD tools. The advantage of hand-coding derivatives is that a human can be more flexible than a software tool. On the other hand the hand-coding approach is tedious and error prone. As a consequence this approach suffers from a considerable development and maintenance effort, and is restricted to first derivatives. The large assimilation systems in the above list [*Rabier et al.*, 2000; *Weaver et al.*, 2003; *Moore et al.*, 2004] were set up before AD tools were mature enough to handle the respective function codes.

## V. CONCLUSIONS

EO products are only accessible to Earth System models via suitable observation operators. There are overlaps between observation operators used to confront dynamical models with EO data (validation, benchmarking, assimilation) and forward models used for retrievals of biophysical products. To allow a most flexible use, observation operators should be designed in modular form with carefully constructed interfaces. Both advanced retrieval algorithms and advanced assimilation techniques (Kalman Filter, 3D-Var, and 4D-Var) rely on first derivatives (linearisations) of the observation operators, i.e. their tangent and adjoint versions. Assessment of uncertainties and quantitative network design in addition require second derivatives of observation operators. To maximise their application range, these derivative codes should be developed and maintained together with their underlying observation operators. This procedure is, for example, applied at the European Centre for Medium range Weather Forecasting. Automatic Differentiation (AD) provides a means to minimise the development and maintenance effort for these derivative codes. There is an ever-increasing list of successful AD applications to large-scale Earth sciences codes, including many observation operators. Meanwhile there is a tendency among code developers to achieve and preserve compliance with an automatic differentiation tool and thus enhance the functionality of their modelling system through the availability of derivative information [see, e.g., *Heimbach et al.*, 2005; *Kaminski et al.*, 2003; *Pinty et al.*, 2007; *Kauker et al.*, 2009; *Lewis et al.*, 2011]. To maximise sustainability of such a modelling system, it is essential that the automatic differentiation tool is permanently maintained and adapted to user needs by an experienced development team.

January 19, 2012

## ACKNOWLEDGMENT

The authors would like to thank Laurent Bertino, Patrick Heimbach, Christian Melsheimer, and Bernard Pinty for fruitful discussions.

## REFERENCES

IFS technical documentation, *Tech. rep.*, European Centre for Medium-Range Weather Forecasts, Reading, UK, 2010.

- Arvidson, R., et al., *Report of the EOS Data Panel on the Data and Information System*, vol. IIA of NASA TM-87777, National Aeronautics and Space Administration, Washington, D.C., 1986.
- Barrett, D. J., and L. J. Renzullo, On the Efficacy of Combining Thermal and Microwave Satellite Data as Observational Constraints for Root-Zone Soil Moisture Estimation, *J. Hydrometeorol.*, *10*, 1109–1127, 2009.
- Blyth, E., D. B. Clark, R. Ellis, C. Huntingford, S. Los, M. Pryor, M. Best, and S. Sitch, A comprehensive set of benchmark tests for a land surface model of simultaneous fluxes of water and carbon at both the global and seasonal scale, *Geoscientific Model Development*, *4*, 255–269, 2011.
- Bodas-Salcedo, A. e. a., COSP: satellite simulation software for model assessment, *Bull. Amer. Meteor. Soc.*, *92*, 1023–1043, 2011.
- Böttger, H., P. Menzel, and J. Pailleux (Eds.), *Proceedings of the third WMO Workshop on the Impact of Various Observing Systems on Numerical Weather Prediction*, WMO, 2004.
- Cahalan, R., et al., The 13RC—Bringing Together the Most Advanced Radiative Transfer Tools for Cloudy Atmospheres, *Bulletin of the American Meteorological Society*, *86*, 1275–1293, 2005.
- Chevallier, F., P. Lopez, A. Tompkins, M. Janisková, and E. Moreau, The capability of 4D-Var systems to assimilate cloud-affected satellite infrared radiances, *Quarterly Journal of the Royal Meteorological Society*, *130*, 917–932, 2004.
- Courtier, P., E. Andersson, W. Heckley, D. Vasiljevic, M. Hamrud, A. Hollingsworth, F. Rabier, M. Fisher, and J. Pailleux, The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation, *Quarterly Journal of the Royal Meteorological Society*, *124*, 1783–1807, 1998.
- Evensen, G., The ensemble kalman filter: Theoretical formulation and practical implementation, *Ocean Dynamics*, *53*, 343–367, 2003.
- Eyre, J., A fast radiative transfer model for satellite sounding systems, *ECMWF Research Dept. Tech. Memo*, *176*, 28, 1991.
- Giering, R., and T. Kaminski, Recipes for Adjoint Code Construction, *ACM Trans. Math. Software*, *24*, 437–474, 1998.
- Giering, R., T. Kaminski, R. Todling, R. Errico, R. Gelaro, and N. Winslow, Generating tangent linear and adjoint versions of NASA/GMAO’s Fortran-90 global weather forecast model, in *Automatic Differentiation: Applications, Theory, and Implementations*, edited by H. M. Bücker, G. Corliss, P. Hovland, U. Naumann, and B. Norris, Lecture Notes in Computational Science and Engineering, pp. 275–284, Springer, 2005.
- Griewank, A., On automatic differentiation, in *Mathematical Programming: Recent Developments and Applications*, edited by M. Iri and K. Tanabe, pp. 83–108, Kluwer Academic Publishers, Dordrecht, 1989.
- Heimann, M., and T. Kaminski, Inverse modeling approaches to infer surface trace gas fluxes from observed atmospheric mixing ratios, in *Approaches to scaling of trace gas fluxes in ecosystems*, edited by A. F. Bouwman, chap. 14, pp. 275–295, Elsevier, Amsterdam, 1999.
- Heimann, M., and S. Körner, The global atmospheric tracer model TM3, *Tech. Rep. 5*, Max-Planck-Institut für Biogeochemie, Jena, Germany, 2003.

- Heimbach, P., C. Hill, and R. Giering, An efficient exact adjoint of the parallel MIT general circulation model, generated via automatic differentiation, *Future Generation Computer Systems*, 21, 1356–1371, 2005.
- Jungclaus, J. H., et al., Climate and carbon-cycle variability over the last millennium, *Climate of the Past*, 6, 723–737, 2010.
- Kaminski, T., and P. J. Rayner, Assimilation and network design, in *Observing the continental scale Greenhouse Gas Balance of Europe*, edited by H. Dolman, A. Freibauer, and R. Valentini, Ecological Studies, chap. 3, pp. 33–52, Springer-Verlag, New York, 2008.
- Kaminski, T., R. Giering, M. Scholze, P. Rayner, and W. Knorr, An example of an automatic differentiation-based modelling system, in *Computational Science – ICCSA 2003, International Conference Montreal, Canada, May 2003, Proceedings, Part II*, edited by V. Kumar, L. Gavrilova, C. J. K. Tan, and P. L’Ecuyer, vol. 2668 of *Lecture Notes in Computer Science*, pp. 95–104, Springer, Berlin, 2003.
- Kaminski, T., W. Knorr, M. Scholze, N. Gobron, B. Pinty, R. Giering, and P.-P. Mathieu, Assimilation of MERIS FAPAR into a Terrestrial Vegetation Model and Mission Design, in *Proceedings of ESA, iLEAPS, EGU joint Conference, Frascati, Italy*, European Space Agency, 2010a.
- Kaminski, T., M. Scholze, and S. Houweling, Quantifying the Benefit of A-SCOPE Data for Reducing Uncertainties in Terrestrial Carbon Fluxes in CCDAS, *Tellus B*, doi:10.1111/j.1600-0889.2010.00483.x, 2010b.
- Kauker, F., R. Gerdes, M. Karcher, T. Kaminski, R. Giering, and M. Voßbeck, Retrospective Summary Comments SIO 2009 - AWI/FastOpt/OASys, Sea Ice Outlook web page, 2009.
- Kleespies, T., P. van Delst, L. McMillin, and J. Derber, Atmospheric transmittance of an absorbing gas. 6. OPTRAN status report and introduction to the NESDIS/NCEP community radiative transfer model, *Applied Optics*, 43, 3103–3109, 2004.
- Knorr, W., T. Kaminski, M. Scholze, N. Gobron, B. Pinty, R. Giering, and P.-P. Mathieu, Carbon cycle data assimilation with a generic phenology model, *J. Geophys. Res.*, doi:10.1029/2009JG001119, 2010.
- Lewis, P. E., J. Gomez-Dans, T. Kaminski, J. Settle, T. Quaipe, N. Gobron, J. Styles, and M. Berger, An Earth Observation Land Data Assimilation System (EO-LDAS), in *revision at Remote Sensing of the Environment*, 2011.
- Lions, J. L., *Optimal Control of Systems Governed by Partial Differential Equations*, Springer, 1971.
- Marshall, J., A. Adcroft, C. Hill, L. Perelman, and C. Heisey, A finite-volume, incompressible navier stokes model for studies of the ocean on parallel computers, *J. Geophys. Res.*, 102, 5753–5766, 1997.
- Mathieu, P., and A. O’Neill, Data assimilation: From photon counts to Earth System forecasts, *Remote Sensing of Environment*, 112, 1258–1267, 2008.
- Melsheimer, C., G. Heygster, N. Mathew, and L. Pedersen, Retrieval of sea ice emissivity and integrated retrieval of surface and atmospheric parameters over the arctic from amsr-e data, *Journal of the Remote Sensing Society of Japan*, 29, 236–241, 2009.
- Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Equation of state calculations for fast computing machines, *Journal of Chemical Physics*, 21, 1087–1092, 1953.

- Moore, A., H. Arango, E. Di Lorenzo, B. Cornuelle, A. Miller, and D. Neilson, A comprehensive ocean prediction and analysis system based on the tangent linear and adjoint of a regional ocean model, *OCEAN MODELLING*, 7, 227–258, 2004.
- O’Dell, C. W., A. K. Heidinger, T. Greenwald, P. Bauer, and R. Bennartz, The successive-order-of-interaction radiative transfer model. Part II: Model performance and applications, *Journal of Applied Meteorology and Climatology*, 45, 1403–1413, 2006.
- Pinty, B., T. Lavergne, R. Dickinson, J. Widlowski, N. Gobron, and M. Verstraete, Simplifying the interaction of land surfaces with radiation for relating remote sensing products to climate models, *J. Geophys. Res.*, 2006.
- Pinty, B., I. Andreadakis, M. Clerici, T. Kaminski, M. Taberner, M. M. Verstraete, N. Gobron, S. Plummer, and J.-L. Widlowski, Exploiting the modis albedos with the two-stream inversion package (jrc-tip): 1. effective leaf area index, vegetation, and soil properties, *Journal of Geophysical Research Atmospheres*, 116, D09,105, 2011a.
- Pinty, B., M. Clerici, I. Andreadakis, T. Kaminski, M. Taberner, M. M. Verstraete, N. Gobron, S. Plummer, and J.-L. Widlowski, Exploiting the modis albedos with the two-stream inversion package (jrc-tip): 2. fractions of transmitted and absorbed fluxes in the vegetation and soil layers, *Journal of Geophysical Research Atmospheres*, 116, D09,106, 2011b.
- Pinty, B., et al., Radiation transfer model intercomparison (RAMI) exercise, *J. Geophys. Res.*, 106, 11,937–11,956, 2001.
- Pinty, B., et al., Retrieving surface parameters for climate models from MODIS-MISR albedo products, *J. Geophys. Res.*, 112, 2007.
- Rabier, F., H. Jarvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons, The ECMWF operational implementation of four-dimensional variational assimilation. Part I: Experimental results with simplified physics, *Quart. J. Roy. Meteor. Soc.*, 126, 1143–1170, 2000.
- Rayner, P., M. Scholze, W. Knorr, T. Kaminski, R. Giering, and H. Widmann, Two decades of terrestrial Carbon fluxes from a Carbon Cycle Data Assimilation System (CCDAS), *Global Biogeochemical Cycles*, 19, doi:10.1029/2004GB002,254, 2005.
- Rayner, P. J., and D. M. O’Brien, The utility of remotely sensed CO<sub>2</sub> concentration data in surface source inversions, *Geophys. Res. Lett.*, 28, 175–178, 2001.
- Saunders, R., M. Matricardi, and P. Brunel, An improved fast radiative transfer model for assimilation of satellite radiance observations, *Quarterly Journal of the Royal Meteorological Society*, 125, 1407–1425, 1999.
- Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. Averyt, M. Tignor, and H. Miller, IPCC, 2007: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, 2007.
- Stammer, D., C. Wunsch, R. Giering, C. Eckert, P. Heimbach, J. Marotzke, A. Adcroft, C. N. Hill, and J. Marshall, The global ocean circulation during 1992–1997, estimated from ocean observations and a general circulation model, *J. Geophys. Res.*, 107, doi:10.1029/2001JC000,888, 2002.
- Stockdale, T., D. Anderson, M. Balmaseda, F. Doblas-Reyes, L. Ferranti, K. Mogensen, T. Palmer, F. Molteni, and

- F. Vitart, Ecmwf seasonal forecast system 3 and its prediction of sea surface temperature, *Climate Dynamics*, 37, 455–471, 2011, 10.1007/s00382-010-0947-3.
- Talagrand, O., The use of adjoint equations in numerical modelling of the atmospheric circulation, in *Automatic Differentiation of Algorithms: Theory, Implementation, and Application*, edited by A. Griewank and G. F. Corliss, pp. 169–180, SIAM, Philadelphia, Penn., 1991.
- Tonboe, R., S. Andersen, L. Toudal, and G. Heygster, Sea ice emission modelling. In: Mtzler, C., P.W. Rosenkranz, A. Battaglia and J.P. Wigneron (eds.), *Thermal Microwave Radiation - Applications for Remote Sensing*, IET Electromagnetic Waves Series 52, London, UK., 2006.
- Toudal, L., Merging microwave radiometer data and meteorological data for improved sea ice concentrations, *EARSeL Advances in Remote Sensing*, 3, 81–89, 1994.
- Vermote, E., D. Tanré, J. Deuze, M. Herman, and J. Morcette, Second simulation of the satellite signal in the solar spectrum, 6s: An overview, *Geoscience and Remote Sensing, IEEE Transactions on*, 35, 675–686, 1997.
- Weaver, A., J. Vialard, and D. Anderson, Three- and four-dimensional variational assimilation with a general circulation model of the tropical Pacific Ocean. Part I: Formulation, internal diagnostics, and consistency checks, *Monthly weather review*, 131, 1360–1378, 2003.
- Widlowski, J., et al., Third Radiation Transfer Model Intercomparison (RAMI) exercise: Documenting progress in canopy reflectance models, *J. Geophys. Res.*, 112, 28 PP, 2007.
- Wiesmann, A., and C. Mätzler, Microwave emission model of layered snowpacks, *Remote Sensing of Environment*, 70, 307–316, 1999.
- Wunsch, C., and P. Heimbach, Estimated decadal changes in the north atlantic meridional overturning circulation and heat flux 1993–2004, *Journal of Physical Oceanography*, 36, 2012, 2006.
- Wunsch, C., P. Heimbach, R. Ponte, and I. Fukumori, The global general circulation of the ocean estimated by the ECCO-Consortium, *Oceanography*, 22, 88–103, 2009.



**Thomas Kaminski** received his Diploma in Mathematical Physics from Technical University Berlin and his PhD in Biogeochemistry from University of Hamburg. He worked at the Max-Planck Institutes for Meteorology and Biogeochemistry, and at Arthur Andersen’s Financial and Commodity Risk Consulting group. He is co-owner and managing director of FastOpt. Thomas is working on automatic differentiation, inverse problems, and assimilation of observations in numerical models in a range of application fields such as climate research, observing network/experimental design, remote sensing, shape design, and computational finance. He is the author/co-author of over 40 refereed publications.



**Pierre-Philippe Mathieu is an Earth Observation Applications Engineer in the Earth Observation Science & Applications Department of the European Space Agency in ESRI (Frascati, Italy). He spent 10 years working in the field of environmental modelling, weather risk management and remote sensing. He has a degree in mechanical engineering and M.Sc from University of Liege (Belgium), a Ph.D. in oceanography from the University of Louvain (Belgium), and a Management degree from the University of Reading Business School (UK).**