AFRL-AFOSR-VA-TR-2016-0102

A Framework for Quantifying and Reducing Uncertainty

Adrian Sandu
VIRGINIA POLYTECHNIC INST AND STATE UNIVERSITY

02/29/2016
Final Report

DISTRIBUTION A: Distribution approved for public release.

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/ RTA2
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
12/12/2015	Final	08/01/2012-09/30/2015

4. TITLE AND SUBTITLE
A Framework for Quantifying and Reducing Uncertainty in InfoSymbiotic Systems Arising in Atmospheric Environments

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-12-1-0293

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Adrian Sandu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Virginia Polytechnic Institute and State University, Blacksburg, VA 24060

8. PERFORMING ORGANIZATION REPORT NUMBER
N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Office of Scientific Research

10. SPONSOR/MONITOR'S ACRONYM(S)
AFOSR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release. There are no distribution limitations of this report (the the best of knowledge of the PI). DISTRIBUTION A

13. SUPPLEMENTARY NOTES
N/A.

14. ABSTRACT
This project focuses on large scale dynamic data driven applications systems (DDDAS, or InfoSymbiotic systems) governed by partial differential equations (PDEs), e.g., arising in atmospheric environments. Specifically, our main interests are data assimilation and the configuration of sensor networks. During this project we have developed a rigorous framework for quantifying and reducing uncertainty in the context of InfoSymbiotic systems. This includes a goal-oriented aposteriori error estimation methodology for the impact of different errors on the variational solutions of inverse problems; an optimization-constrained optimization problem approach to find the optimal configuration of the DDDAS system; new parallel-in-time algorithms to speed up variational inference; a trust-region approach to perform inference in an ensemble space; new nonlinear filtering and smoothing algorithms that sample directly from the posterior PDF using a Hybrid Markov-Chain Monte Carlo (HMCMC) approach; and a solid theoretical basis for optimization with reduced order models.

15. SUBJECT TERMS
Infosymbiotic systems; atmospheric environments; uncertainty quantification and reduction; aposteriori error estimates; parallel computing; numericla optimization.

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
					19b. TELEPHONE NUMBER *(Include area code)*

Reset

INSTRUCTIONS FOR COMPLETING SF 298

1. REPORT DATE. Full publication date, including day, month, if available. Must cite at least the year and be Year 2000 compliant, e.g. 30-06-1998; xx-06-1998; xx-xx-1998.

2. REPORT TYPE. State the type of report, such as final, technical, interim, memorandum, master's thesis, progress, quarterly, research, special, group study, etc.

3. DATES COVERED. Indicate the time during which the work was performed and the report was written, e.g., Jun 1997 - Jun 1998; 1-10 Jun 1996; May - Nov 1998; Nov 1998.

4. TITLE. Enter title and subtitle with volume number and part number, if applicable. On classified documents, enter the title classification in parentheses.

5a. CONTRACT NUMBER. Enter all contract numbers as they appear in the report, e.g. F33615-86-C-5169.

5b. GRANT NUMBER. Enter all grant numbers as they appear in the report, e.g. AFOSR-82-1234.

5c. PROGRAM ELEMENT NUMBER. Enter all program element numbers as they appear in the report, e.g. 61101A.

5d. PROJECT NUMBER. Enter all project numbers as they appear in the report, e.g. 1F665702D1257; ILIR.

5e. TASK NUMBER. Enter all task numbers as they appear in the report, e.g. 05; RF0330201; T4112.

5f. WORK UNIT NUMBER. Enter all work unit numbers as they appear in the report, e.g. 001; AFAPL30480105.

6. AUTHOR(S). Enter name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. The form of entry is the last name, first name, middle initial, and additional qualifiers separated by commas, e.g. Smith, Richard, J, Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES). Self-explanatory.

8. PERFORMING ORGANIZATION REPORT NUMBER. Enter all unique alphanumeric report numbers assigned by the performing organization, e.g. BRL-1234; AFWL-TR-85-4017-Vol-21-PT-2.

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES). Enter the name and address of the organization(s) financially responsible for and monitoring the work.

10. SPONSOR/MONITOR'S ACRONYM(S). Enter, if available, e.g. BRL, ARDEC, NADC.

11. SPONSOR/MONITOR'S REPORT NUMBER(S). Enter report number as assigned by the sponsoring/ monitoring agency, if available, e.g. BRL-TR-829; -215.

12. DISTRIBUTION/AVAILABILITY STATEMENT. Use agency-mandated availability statements to indicate the public availability or distribution limitations of the report. If additional limitations/ restrictions or special markings are indicated, follow agency authorization procedures, e.g. RD/FRD, PROPIN, ITAR, etc. Include copyright information.

13. SUPPLEMENTARY NOTES. Enter information not included elsewhere such as: prepared in cooperation with; translation of; report supersedes; old edition number, etc.

14. ABSTRACT. A brief (approximately 200 words) factual summary of the most significant information.

15. SUBJECT TERMS. Key words or phrases identifying major concepts in the report.

16. SECURITY CLASSIFICATION. Enter security classification in accordance with security classification regulations, e.g. U, C, S, etc. If this form contains classified information, stamp classification level on the top and bottom of this page.

17. LIMITATION OF ABSTRACT. This block must be completed to assign a distribution limitation to the abstract. Enter UU (Unclassified Unlimited) or SAR (Same as Report). An entry in this block is necessary if the abstract is to be limited.

Project Title: A Framework for Quantifying and Reducing Uncertainty in InfoSymbiotic Systems Arising in Atmospheric Environments

(Project AFOSR FA9550–12–1–0293–DEF)

Technical point of contact: Dr. Frederica Darema, E-mail: frederica.darema@us.af.mil

P.I.: Professor Adrian Sandu (Virginia Tech), E-mail: sandu@cs.vt.edu, Phone: 540-231-2193, Fax: 540 - 231-9218.

Contents

I Summary

This project focuses on large scale dynamic data driven applications systems (DDDAS, or *InfoSymbiotic systems*) governed by partial differential equations (PDEs), e.g., arising in atmospheric environments. Specifically, our main interests are data assimilation and the configuration of sensor networks. *Data assimilation* is the process to dynamically integrate information from measurements into models. In a variational approach the data assimilation problem is posed as an inverse problem, where parameters are adjusted such that the model predictions best fit the measurements. *Sensor network configuration* is the process of using model results to dynamically steer the measurement process.

InfoSymbiotic applications are inherently subject to uncertainties associated with imperfect models and with noisy data. There is an urgent need to quantify, and control, the effect of model and data errors on the overall DDDAS results, and to fill the gap between the state-of-the-art modeling techniques (capable of quantifying uncertainty in modeling results) and the computational tools currently available for InfoSymbiotic applications.

During the first three years of this project (2012–2015) we started to address this need and developed a rigorous framework for quantifying and reducing uncertainty in the context of InfoSymbiotic systems [1, 2, 4, 6, 10, 11] (details are given in Section II).

DDDAS integrates computational simulations and physical measurements in a symbiotic feedback control system. Inverse problems in this framework use data from measurements along with a numerical model to estimate the parameters or state of a physical system of interest. Uncertainties in both measurements and the computational model lead to inaccurate estimates. We developed a goal-oriented aposteriori error estimation methodology for the impact of different errors on the variational solutions of inverse problems. In the goal-oriented approach we are interested in estimating the impact of observation and model errors on the quantity of interest, i.e., on the aspect of interest of the optimal solution.

The variational data assimilation problem optimizes the model states and parameters in order to obtain predictions that fit best the measurements. In this project we solved the complementary problem of optimizing the DDDAS process. The strategies used to collect and process the data are considered to be parameters of the inference system, and are themselves improved via an additional optimization process. Specifically, we seek to improve the parameters of the data assimilation system. We have formulated the optimal configuration of the DDDAS system as the following "optimization-constrained optimization problem??. Our algorithm is based on first and second order adjoints, and on the solution of large linear systems. We also proposed efficient methods for computing observation impact, including low-rank approximations of observation-to-analysis sensitivities.

DDDAS variational inference in real time is hindered by costly forward and adjoint model runs. We proposed a new parallel-in-time algorithm to speed up the solution process. The original 4D-Var problem is solved in the augmented Lagrangian framework. To expose time parallelism the assimilation window is divided into several sub-intervals. This formulation allows to integrate the forward and the adjoint models over different subintervals in parallel.

The construction and validation of an adjoint model is an extremely labor-intensive process. To address this challenge we proposed a derivative-free 4D-Var(-like) smoothing algorithm that approximately solves DDDAS variational inference without the need to construct adjoint models. Specifically, our TR-4D-EnKF algorithm uses a trust-region approach to optimize in ensemble space.

Variational DDDAS inference solution does not include posterior uncertainty estimates. To address this challenge we developed new nonlinear filtering and smoothing algorithms that sample directly from the posterior PDF using a Hybrid Markov-Chain Monte Carlo (HMCMC) approach. The sampling smoother is implemented efficiently using the same adjoint computational infrastructure used in 4D-Var.

Consistency of the reduced-order Karush-Kuhn-Tucker conditions with the full-order optimality conditions is a key ingredient for successful reduced order data assimilation problems. This translates into accurate low-rank approximations of the both adjoint and forward models leading to reduced bases constructed from the dominant eigenvectors of the correlation matrix of the aggregated snapshots of full forward and the adjoint models. Our work underlines the importance of incorporating the adjoint information into the construction of reduced order basis for performing reduced order 4D-Var data assimilation.

II Results From the DDDAS Project AFOSR FA9550–12–1–0293–DEF (2012–2015)

We summarize here the main results obtained during the first three years of this project, 2012–2015. The research presented below was either fully or partially funded by this project.

II. 1 Mathematical framework

We work in a variational framework and regard the inference problem as an inverse problem, as follows.

- The real (physical) system is described by a state $\mathbf{x}^{\mathrm{true}}$ (e.g., the spatio-temporal distribution of wind velocities) and a vector of model parameters of interest θ^{true} (e.g., the fields at the initial time). We do not know the real state or the real parameters, and our goal is to derive information about θ^{true} from measurements of $\mathbf{x}^{\mathrm{true}}$.
- The prior information encapsulates our current knowledge of the system. Usually the prior information is contained in a background estimate of the state $\mathbf{x}^{\mathrm{b}}$ and the corresponding background error covariance matrix $\mathbf{B}$.
- The reality is described by a computer model that captures our knowledge about the physical laws that govern the evolution of the system:

$$\mathbf{x}_{\mathrm{k+1}} = \mathcal{M}_{\mathrm{k,k+1}}\left(\mathbf{x}_{\mathrm{k}}\right) , \quad \mathrm{k} = 0, 1, \cdots, N-1 \, , \tag{1}$$

where $\mathcal{M}_{\mathrm{k,k+1}}$ represents the model solution operator that propagates the state $\mathbf{x}_k$ at t_k to the state $\mathbf{x}_{k+1}$ at t_{k+1}.

- The sensor network provides observations of some aspects of the real state. Observations are noisy snapshots of reality available at discrete time instances t_k, $\mathrm{k} = 1, \cdots, N$. The model state is related to observations by the following relation:

$$\mathbf{y}_\mathrm{k} \;=\; \mathcal{H}\left(\mathbf{x}_\mathrm{k}\right) - \varepsilon_\mathrm{k}^\mathrm{obs}, \quad \varepsilon_\mathrm{k}^\mathrm{obs} = \varepsilon_\mathrm{k}^\mathrm{representativeness} + \varepsilon_\mathrm{k}^\mathrm{measurement}\,. \tag{2}$$

  The observation operator $\mathcal{H}$ maps the model state space onto the observation space. The observation error term $\left(\varepsilon_\mathrm{k}^\mathrm{obs}\right)$ accounts for both measurement and representativeness errors. Measurement errors are due to imperfect sensors. The representativeness errors are due to the inaccuracies of the mathematical and numerical approximations inherent to the model.

The inference problem. To simplify ideas consider (without loss of generality) that the model parameters are the initial conditions, $\theta = \mathbf{x}_0$. *The inference (data assimilation) problem* is formalized as a model-constrained optimization problem:

$$\mathbf{x}_0^\mathrm{a} \;=\; \arg\min_{\mathbf{x}_0}\; \mathcal{J}\left(\mathbf{x}_0\right) \qquad \text{subject to (1)}, \tag{3a}$$

$$\mathcal{J}\left(\mathbf{x}_0\right) = \frac{1}{2}\,\left\|\mathbf{x}_0 - \mathbf{x}_0^\mathrm{b}\right\|^2_{\mathbf{B}_0^{-1}(\mathbf{u})} + \frac{1}{2}\sum_{\mathrm{k}=1}^{N}\left\|\mathcal{H}_\mathrm{k}\left(\mathbf{x}_\mathrm{k},\mathbf{u}\right) - \mathbf{y}_\mathrm{k}\right\|^2_{\mathbf{R}_\mathrm{k}^{-1}(\mathbf{u})}\,. \tag{3b}$$

The first term of the sum (3b) quantifies the departure of the solution $\mathbf{x}_0$ from the background state $\mathbf{x}_0^\mathrm{b}$ at the initial time t_0. The second term measures the mismatch between the forecast trajectory (model solutions $\mathbf{x}_\mathrm{k}$) and observations $\mathbf{y}_\mathrm{k}$ at all times $t^{\{\ell\}}$ in the assimilation window. The weighting matrices $\mathbf{B}_0$ and $\mathbf{R}_\mathrm{k}$ need to be predefined, and their quality influences the accuracy of the resulting analysis. The vector $\mathbf{u}$ represents the parameters of the DDDAS system, e.g., sensor locations and weights attributed to various data points.

Weak constraint 4D-Var avoids the assumption of a perfect model [16], implicit in the traditional strong constraint formulation (3), at the expense of solving a larger optimization problem. The state $\mathbf{x}_k$ at t_k is allowed to differ from the model prediction $\mathcal{M}_{k-1,k}(\mathbf{x}_{k-1})$. The weak constraint 4D-Var estimates of the states $\mathbf{x} = [\mathbf{x}_0, \ldots, \mathbf{x}_N]$ are the unconstrained minimizer of the following cost function:

$$\min\; \mathcal{J}^\mathrm{w}\left(\mathbf{x}\right) = \frac{1}{2}\left\|\mathbf{x}_0 - \mathbf{x}_0^\mathrm{b}\right\|^2_{\mathbf{B}_0^{-1}} + \frac{1}{2}\sum_{k=1}^{N}\left\|\mathcal{H}(\mathbf{x}_k) - \mathbf{y}_k\right\|^2_{\mathbf{R}_k^{-1}} + \frac{1}{2}\sum_{k=1}^{N}\left\|\mathbf{x}_k - \mathcal{M}_{k-1,k}(\mathbf{x}_{k-1})\right\|^2_{\mathbf{Q}_k^{-1}}. \tag{4}$$

The last term in the cost function (4) corresponds to the contribution of model error to changing the analysis. The model is not imposed exactly; rather, it is treated as a weak constraint (i.e., the differences $\mathbf{x}_k - \mathcal{M}_{k-1,k}(\mathbf{x}_{k-1})$ are penalized in the cost function). The control variables in (4) can be not only the model states $\mathbf{x}_k$ at each time step, but also the model biases β_k [16].

The dynamic configuration of the observation network problem. The inverse problem is posed as a PDE-constrained nonlinear *optimization* problem, where model states and parameters are tuned in order to obtain predictions that fit best the measurements. The complementary problem is to optimize the strategies used to collect and process the data, such as to improve the performance of the inversion. Formally, dynamic configuration of the observation network is achieved by improving the parameters $\mathbf{u}$ of the system (3b), as explained in [4]. We discuss this in detail in Section II. 3.

II. 2 A-posteriori error estimates for the solution of variational inverse problems

DDDAS integrates computational simulations and physical measurements in a symbiotic feedback control system. Inverse problems in this framework use data from measurements along with a numerical model to estimate the parameters or state of a physical system of interest. Uncertainties in both measurements and the computational model lead to inaccurate estimates.

Specifically, in practice the evolution of the physical system is described by an imperfect model

$$\widehat{\mathbf{x}}_{k+1} = \mathcal{M}_{k,k+1}(\widehat{\mathbf{x}}_k) + \Delta\widehat{\mathbf{x}}_{k+1}(\widehat{\mathbf{x}}_k), \quad k = 0, 1, \ldots, N-1\,., \tag{5}$$

where $\Delta\widehat{\mathbf{x}}_{k+1}(\widehat{\mathbf{x}}_k, \theta)$ represents the (additive) model error at time t_{k+1}. The observations collected by the sensors are also imperfect and contain data errors $\Delta\mathbf{y}_k$. In practice one solves a perturbed inverse problem of the form:

$$\widehat{\mathbf{x}}_0^{\mathrm{a}} = \arg\min_{\mathbf{x}_0 \in \mathbb{R}^n} \widehat{\mathcal{J}}\left(\mathbf{x}_0\right) \quad \text{subject to } (5), \tag{6a}$$

$$\widehat{\mathcal{J}}\left(\mathbf{x}_0\right) = \frac{1}{2}\, \left\|\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\right\|^2_{\mathbf{B}_0^{-1}(\mathbf{u})} + \frac{1}{2}\sum_{\mathrm{k}=1}^{N} \left\|\mathcal{H}_{\mathrm{k}}\left(\widehat{\mathbf{x}}_{\mathrm{k}}, \mathbf{u}\right) - \mathbf{y}_{\mathrm{k}} - \Delta\mathbf{y}_k\right\|^2_{\mathbf{R}_{\mathrm{k}}^{-1}(\mathbf{u})}. \tag{6b}$$

In [10, 11] we developed a goal-oriented aposteriori error estimation methodology for the impact of different errors on the variational solutions of inverse problems. Consider a quantity of interest (QOI) defined by a scalar functional $\mathcal{E} : \mathbb{R}^m \to \mathbb{R}$ that measures a certain aspect of the the optimal solution value

$$\text{QOI} = \mathcal{E}\left(\mathbf{x}_0^{\mathrm{a}}\right)\,. \tag{7}$$

In the goal-oriented approach we are interested in estimating the impact of observation and model errors on the QOI, i.e., on the aspect of interest of the optimal solution. The error in the QOIis

$$\Delta\mathcal{E} = \mathcal{E}(\widehat{\mathbf{x}}_0^{\mathrm{a}}) - \mathcal{E}\left(\mathbf{x}_0^{\mathrm{a}}\right) \tag{8}$$

where $\mathbf{x}_0^{\mathrm{a}}$ and $\widehat{\mathbf{x}}_0^{\mathrm{a}}$ are the solutions of the ideal inverse problem (3) and of the perturbed inverse problem (6), respectively.

We have shown in [10, 11] that the error in the QOI is approximated to first order by the sum of contributions of errors in the forward model, adjoint model, and optimality equation:

$$\Delta\mathcal{E} \;\approx\; \Delta\mathcal{E}^{\rm est} = \Delta\mathcal{E}_{\rm fwd} + \Delta\mathcal{E}_{\rm adj} + \Delta\mathcal{E}_{\rm opt}. \tag{9}$$

Moreover, the contributions of errors are

$$\begin{aligned}
\Delta\mathcal{E}_{\rm fwd} &= \textstyle\sum_{k=1}^{N} \nu_k^{\rm T} \cdot \Delta\widehat{\mathbf{x}}_k \,, && \text{(10a)}\\
\Delta\mathcal{E}_{\rm adj} &= -\textstyle\sum_{k=0}^{N} \mu_k^{\rm T} \cdot \left(\mathbf{H}_k^{\rm T}\mathbf{R}_k^{-1}\Delta\mathbf{y}_k\right) + \sum_{k=0}^{N-1} \mu_k^{\rm T} \cdot \left(\Delta\widehat{\mathbf{x}}_{k+1}\right)_{\mathbf{x}_k}^{\rm T} \widehat{\lambda}_{k+1} \,, && \text{(10b)}\\
\Delta\mathcal{E}_{\rm opt} &= -\zeta^{\rm T} \, \left(\Delta\widehat{\mathbf{x}}_1\right)_{\mathbf{x}_0}^{\rm T} \, \widehat{\lambda}_1. && \text{(10c)}
\end{aligned}$$

We see that data errors contribute to the error in adjoint model. The forward model errors contribute to errors in both the forward and the adjoint equations. The "impact factors" $\zeta \in \mathbb{R}^m$, $\mu_k \in \mathbb{R}^n$ for $k = 0, \ldots, N$, and $\nu_k \in \mathbb{R}^n$ for $k = 0, \ldots, N$ are calculated by the following algorithm:

$$\begin{aligned}
\text{Linear system:} &\quad \left(\nabla^2_{\mathbf{x}_0,\mathbf{x}_0} j\right) \cdot \zeta = \nabla_{\mathbf{x}_0}\mathcal{E} \,; && \text{(11a)}\\
\text{Tangent linear model:} &\quad \mu_0 = -\zeta \,; \quad \mu_{k+1} = \mathbf{M}_{k,k+1}\, \mu_k, \quad k = 0, \ldots, N-1 \,; && \text{(11b)}\\
\text{Second order adjoint:} &\quad \nu_N = \mathbf{H}_N^{\rm T}\mathbf{R}_N^{-1}\mathbf{H}_N\, \mu_N \,, && \text{(11c)}\\
&\quad \nu_k = \mathbf{M}_{k,k+1}^{T}\, \nu_{k+1} + \left(\mathbf{M}_{k,k+1}^{\rm T}\, \lambda_{k+1}\right)_{\mathbf{x}_k}^{T} \mu_k \\
&\qquad\quad + \mathbf{H}_k^{\rm T}\mathbf{R}_k^{-1}\mathbf{H}_k\, \mu_k \,, \quad k = N-1, \ldots, 0.
\end{aligned}$$

We applied this methodology to real scenarios. Figure 1 illustrates a data assimilation calculation carried out using the Weather Research and Forecast Model (WRF-VAR). The errors in the meridional wind component observation from GEOAMV and their impact on the QoIare shown.

II. 3 Dynamic configuration of sensor networks via optimization of DDDAS parameters

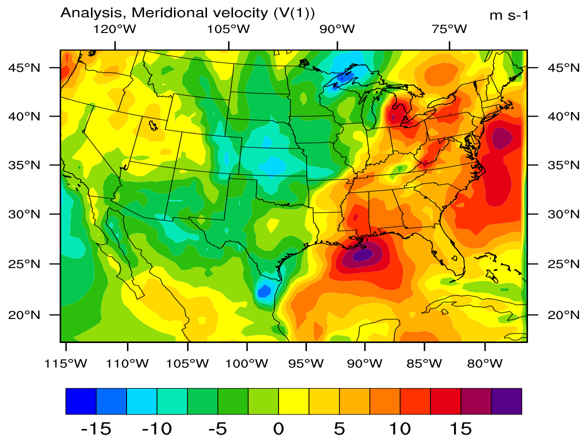
The variational data assimilation problem (3) optimizes the model states and parameters in order to obtain predictions that fit best the measurements.

In our work [4] we solve the complementary problem of optimizing the DDDAS process. The strategies used to collect and process the data are considered to be parameters of the inference system, and are themselves improved via an additional optimization process. Specifically, we seek to improve the parameters $\mathbf{u}$ of the data assimilation system (3b).

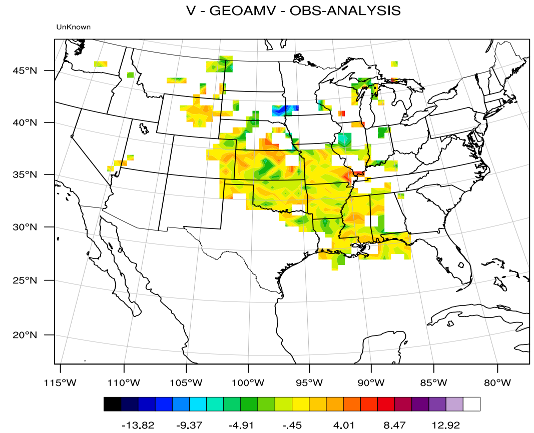
We measure the quality of the inverse solution (3) by the discrepancy between the model forecast (initialized from the analysis $\mathbf{x}_0^{\rm a}$) and a set of high-quality verification data $\mathbf{y}_v^{\rm verif}$ collected at verification time t_v. This discrepancy can be measured by the quadratic "verification" cost function

$$\Psi(\mathbf{u}) = \Psi\left(\mathbf{x}_v^{\rm a}(\mathbf{u})\right) = \frac{1}{2}\, \left\|\mathcal{H}_v(\mathbf{x}_v^{\rm a}(\mathbf{u})) - \mathbf{y}_v^{\rm verif}\right\|_{\mathbf{C}_v}^2 \,. \tag{12}$$

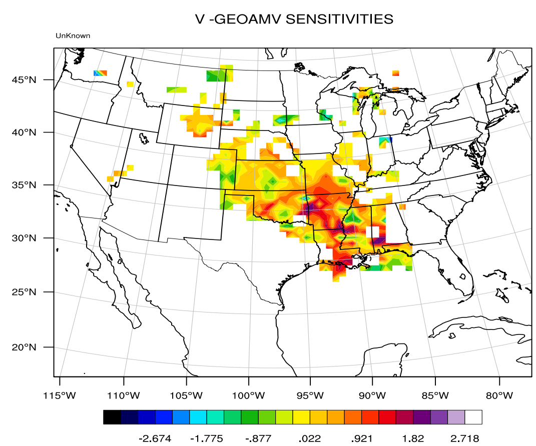
The function Ψ depends directly on $\mathbf{x}_0^{\rm a}(\mathbf{u})$, and indirectly on the system parameters $\mathbf{u}$.



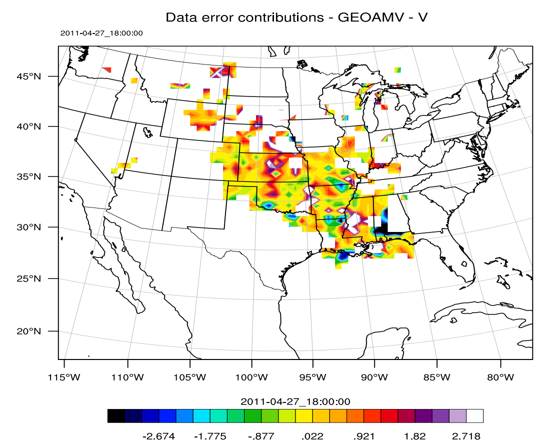
(a) Optimal initial V (meridional component of the wind field) at ground level (V component of $\mathbf{x}_0^{\mathrm{a}}$).



(b) Data errors in GEOAMV observations ($\Delta \mathbf{y}_k$).



(c) Data error impact factors (scaled μ_k).



(d) Contribution of data errors to error in analysis QOI($\mu_k^{\mathrm{T}} \mathbf{H}_k^{\mathrm{T}} \mathbf{R}_k^{-1} \Delta \mathbf{y}_k$).

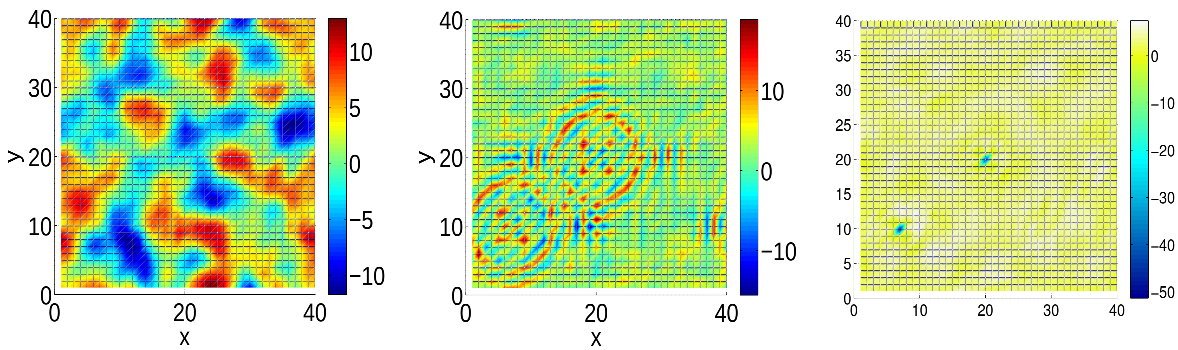
Figure 1: Our aposteriori error estimator [10, 11] applied to a real scenario: data assimilation with WRFDA. The assimilation window is from 18:00@04/27/2011 to 00:00@04/28/2011. The simulation domain covers the continental U.S. with a horizontal grid resolution of 60Km. We consider the meridional wind component (V) observations taken by the geostationary satellite GEOAMV (geostationary atmospheric motion vector).

In [4] we have formulated the optimal configuration of the DDDAS system as the following "optimization-constrained optimization problem"

$$\mathbf{u}_{\rm opt} = \arg\min_{\mathbf{u}} \Psi\big(\mathbf{x}_v^{\rm a}\big) \quad \text{subject to} \begin{cases} \mathbf{x}_0^{\rm a} = \arg\min_{\mathbf{x}_0} \mathcal{J}\big(\mathbf{x}_0, \mathbf{u}\big)\,, \\ \mathbf{x}_v^{\rm a} = \mathcal{M}_{t_0 \to t_v}(\mathbf{x}_0^{\rm a})\,. \end{cases} \tag{13}$$

Our algorithm to solve (13) is based on first and second order adjoints, and on the solution of large linear systems. In [5, 6] we proposed efficient methods for computing observation impact, including low-rank approximations of observation-to-analysis sensitivities.

Figure 2 illustrates the application of our methodology to detect faulty sensors. They are not not visible from the inference solution, but are detectable via our sensitivity to data approach.



(a) The smooth 4D-Var increment $(\mathbf{x}_0^{\rm a} - \mathbf{x}_0^{\rm b})$ does not indicate any problem with data.

(b) Supersensitivity field exhibits a clear structure.

(c) Sensitivity to observations field clearly identifies the locations of the two faulty sensors.

Figure 2: The observation impact methodology to identify possibly faulty sensors. From [6].

Figure 3 illustrates how the optimization-constrained optimization process can dynamically adjust the data covariances (weights) and the can define optimal sensor network configurations. Both procedures lead to considerable decrease of forecast error (12), and therefore to sconsiderably improved performance of the DDDAS system.

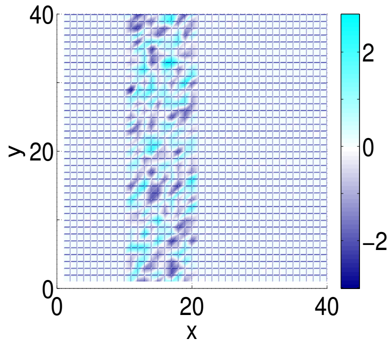
II. 4 Parallel-in-time algorithm for fast variational DDDAS inference

DDDAS variational inference in real time is hindered by costly forward and adjoint model runs. In [12] we proposed a new parallel-in-time algorithm to speed up the solution process. The concept is illustrated in Figure 4a.

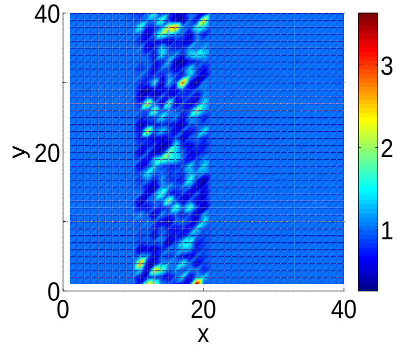
The original 4D-Var problem in (3a) is solved in the augmented Lagrangian framework [9, Section 17.3]. To expose time parallelism the assimilation window is divided into N sub-intervals, namely,

$$[t_0, t_N] = [t_0, t_1] \cup \ldots \cup [t_{N-1}, t_N]. \tag{14}$$

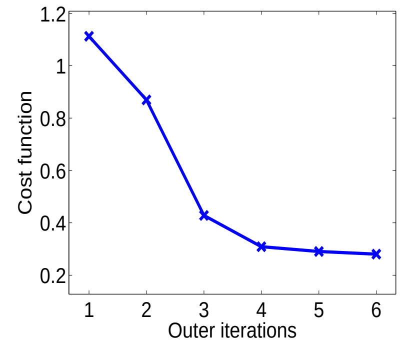
The optimization variables are the forward model and adjoint model states at the interval bound-



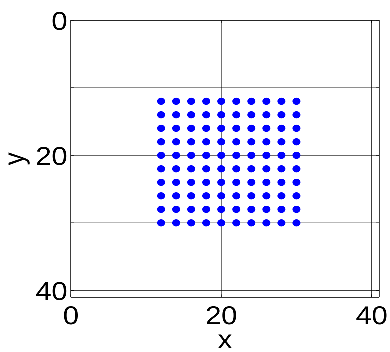
(a) The DDDAS system specifies initially equal noise levels, unaware of larger observation errors in area.



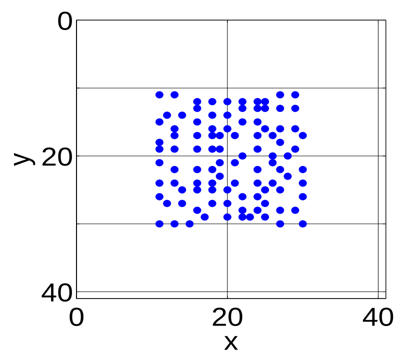
(b) Covariances optimized via solving (13) automatically reduce the weight given to noisy data.



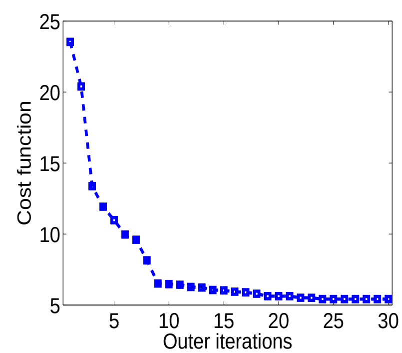
(c) Decrease of forecast error (12) shows a considerably improved performance of the DDDAS system.



(d) Initial sensor locations.



(e) Optimized sensor locations via solving (13).

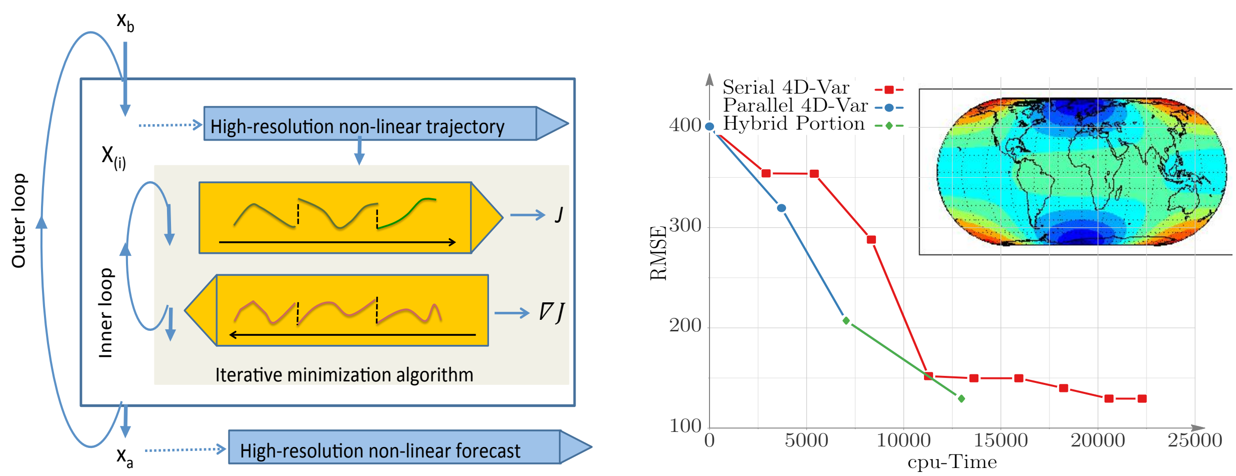


(f) Decrease of forecast error (12) $\equiv$ improved DDDAS performance.

Figure 3: Optimization of DDDAS parameters for a two dimensional shallow water system. From [4].

aries $\mathbf{x} = [\mathbf{x}_0, \cdots, \mathbf{x}_N]$ and $\boldsymbol{\lambda} = [\lambda_0, \cdots, \lambda_N]$, respectively. Solution continuity equations across interval boundaries are added as constraints. *This formulation allows to integrate the forward and the adjoint models over different subintervals in parallel.* The optimization proceeds in cycles of inner and outer iterations and updates alternatively $\mathbf{x}$ and $\boldsymbol{\lambda}$ variables. The augmented Lagrangian approach leads to a different formulation of the variational data assimilation problem than weakly constrained 4D-Var.

Results from applying parallel-in-time 4D-var data assimilation to the shallow water on the sphere model is illustrated in Figure 4b. A speedup factor of two is obtained by a combination of parallel and traditional approaches. Note that this factor of two is obtained on top of the parallel speedup due to traditional spatial domain decomposition parallelization applied to forward and adjoint models [13].



(a) Concept: the parallel-in-time solution of 4D-Var data assimilation problem.

(b) Data assimilation with the shallow water on the sphere model: 2x overall speed-up.

Figure 4: Parallel-in-time 4D-Var applied to 2D shallow water equations [12].

II. 5 Adjoint-free variational inference

The construction and validation of an adjoint model is an extremely labor-intensive process. To address this challenge in [8] we proposed a derivative-free 4D-Var(-like) smoothing algorithm that approximately solves DDDAS variational inference without the need to construct adjoint models. Specifically, our TR-4D-EnKF algorithm uses a trust-region approach to optimize in ensemble space.

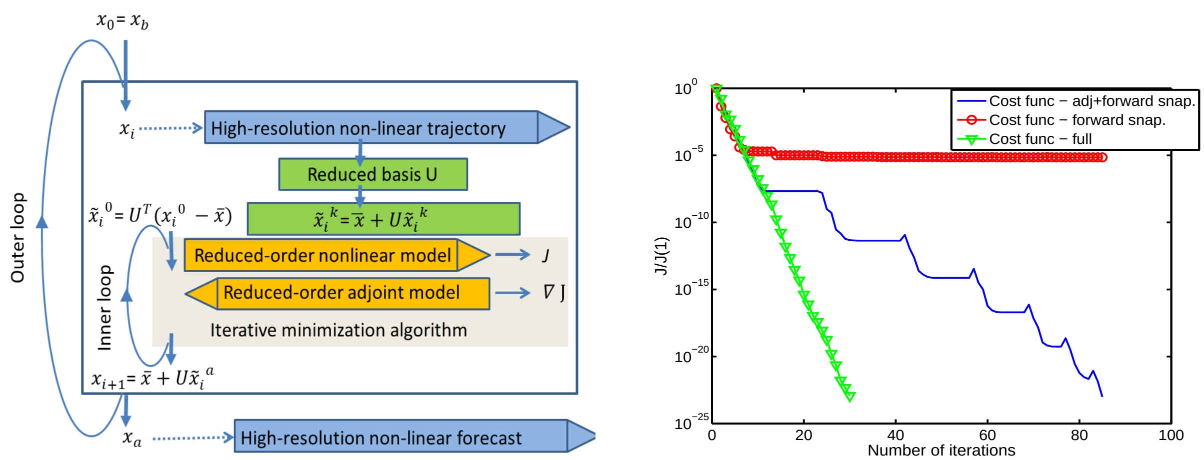
II. 6 Hamiltonian Monte-Carlo sampling filter and smoother

Variational DDDAS inference solution does not include posterior uncertainty estimates. To address this challenge we developed new nonlinear filtering [2] and smoothing [1] algorithms that sample directly from the posterior PDF using a Hybrid Markov-Chain Monte Carlo (HMCMC) approach. The sampling smoother is implemented efficiently using the same adjoint computational infrastructure used in 4D-Var.

II. 7 Optimization with reduced order model surrogates

Consistency of the reduced-order Karush-Kuhn-Tucker conditions with the full-order optimality conditions is a key ingredient for successful reduced order data assimilation problems. This translates into accurate low-rank approximations of the both adjoint and forward models (see [14]) leading to reduced bases constructed from the dominant eigenvectors of the correlation matrix of the aggregated snapshots of full forward and the adjoint models.

Our recent work [14, 15] underlines the importance of incorporating the adjoint information into the construction of reduced order basis for performing reduced order 4D-Var data assimilation (see red versus blue lines in Figure 5b). The new shallow water ROM data assimilation system provides analyses similar to those produced by the full resolution data assimilation system in one tenth of the computational time (see black versus blue lines in Figure 5b). Another order of magnitude in savings is expected with three dimensional models.



(a) Concept: the proposed solution of 4D-Var data assimilation problem uses ROMs as surrogates in an inner optimization loop.

(b) Cost function decrease for 4D-Var applied to 2D shallow water equations [15]. It is essential to incorporate adjoint information in reduced basis.

Figure 5: Reduced order 4D-Var applied to 2D shallow water equations [15]. Optimization based on traditional reduced order models - constructed from the forward solution snapshots - is inaccurate (red line). However, reduced order analysis (blue line) is as accurate as the full model one (green line) when adjoint information is incorporated in the reduced bases. The reduced order analysis is ten times faster than the full order one.

References

[1] A. Attia, V. Rao, and A. Sandu. A hybrid Monte-Carlo sampling smoother for four dimensional data assimilation. Technical Report CSL-TR-18/2014, Computational Science Laboratory, Virginia Tech, 2014.

[2] A. Attia and A. Sandu. A hybrid Monte–Carlo filter for non-Gaussian data assimilation. Technical Report CSL-TR-4/2014, Computational Science Laboratory, Virginia Tech, 2014.

[3] K. Bryan. Seeing the unseeable: the mathematics of inverse problems. Presentation at Rose-Hulman Institute of Technology, 2011.

[4] A. Cioaca and A. Sandu. An optimization framework to improve 4D-Var data assimilation system performance. Journal of Computational Physics, 275:377–389, 2014.

[5] A. Cioaca, A. Sandu, and E. de Sturler. Efficient methods for computing observation impact in 4D-Var data assimilation. Computational Geosciences, 17(6):975–990, 2013.

[6] Alexandru Cioaca and Adrian Sandu. Low-rank approximations for observation impact in 4D-Var data assimilation. Computers and Mathematics with Applications, 67(12):2112–2126, 2014.

[7] A. Jameson. Airplane design with aerodynamic shape optimization. In TFSA Optimization Workshop, Stanford University, 2011.

[8] E. Nino and A. Sandu. A derivative-free trust region framework for variational data assimilation. Journal of Computational and Applied Mathematics, 293:164–179, 2016.

[9] J. Nocedal and S. J. Wright. Numerical optimization. Springer Series in Operations Research. Springer-Verlag, 2nd edition, 2006.

[10] V. Rao and A. Sandu. A posteriori error estimates for DDDAS inference problems. In International Conference on Computational Science (ICCS-2014), volume 29, pages 1256–1265, 2014.

[11] V. Rao and A. Sandu. A posteriori error estimates for inverse problems. SIAM/ASA Journal on Uncertainty Quantification, In print., 2015.

[12] V. Rao and A. Sandu. A time-parallel approach to strong-constraint four-dimensional variational data assimilation. Journal of Computational Physics, Submitted, 2015.

[13] A. Sandu, D. Daescu, G.R. Carmichael, and T. Chai. Adjoint sensitivity analysis of regional air quality models. Journal of Computational Physics, 204:222–252, 2005.

[14] R. Stefanescu, A. Sandu, and I.M. Navon. Comparison of POD reduced order strategies for the nonlinear 2D shallow water equations. International Journal for Numerical Methods for Fluids, Accepted., 2014.

[15] R. Stefanescu, A. Sandu, and I.M. Navon. POD/DEIM strategies for reduced data assimilation systems. Journal of Computational Physics, 295:569–595, 2015.

[16] Yannick Tr'emolet. Accounting for an imperfect model in 4d-var. Quarterly Journal of the Royal Meteorological Society, 132(621):2483–2504, 2006.

[17] T. Varslot, J.H. Morales, and M. Cheney. Synthetic-aperture radar imaging through dispersive media. Inverse Problems, 26(025008):27, 2010.

1.

1. Report Type

Final Report

Primary Contact E-mail
Contact email if there is a problem with the report.

sandu@cs.vt.edu

Primary Contact Phone Number
Contact phone number if there is a problem with the report

540-231-2193

Organization / Institution name

Virginia Tech

Grant/Contract Title
The full title of the funded effort.

A Framework for Quantifying and Reducing Uncertainty in InfoSymbiotic Systems Arising in Atmospheric Environments

Grant/Contract Number
AFOSR assigned control number. It must begin with "FA9550" or "F49620" or "FA2386".

FA9550-12-1-0293

Principal Investigator Name
The full name of the principal investigator on the grant or contract.

Adrian Sandu

Program Manager
The AFOSR Program Manager currently assigned to the award

Dr. Frederica Darema

Reporting Period Start Date

08/01/2012

Reporting Period End Date

09/30/2015

Abstract

This project focuses on large scale dynamic data driven applications systems (DDDAS, or InfoSymbiotic systems)
governed by partial differential equations (PDEs), e.g., arising in atmospheric environments. Specifically, our main interests are data assimilation and the configuration of sensor networks.
Data assimilation is the process to dynamically integrate information from measurements into models. In a variational approach the data assimilation problem is posed as an inverse
problem, where parameters are adjusted such that the model predictions
best fit the measurements. Sensor network configuration is the process of using model results
to dynamically steer the measurement process.

InfoSymbiotic applications are inherently subject to uncertainties associated with imperfect models and with noisy data.
There is an urgent need to quantify, and control, the effect of model and data errors on the overall DDDAS

results, and to fill the gap between the state-of-the-art modeling techniques (capable of quantifying uncertainty in modeling results) and the computational tools currently available for InfoSymbiotic applications.

During the first three years of this project (2012--2015) we started to address this need and developed a rigorous framework for quantifying and reducing uncertainty in the context of InfoSymbiotic systems.

DDDAS integrates computational simulations and physical measurements in a symbiotic feedback control system. Inverse problems in this framework use data from measurements along with a numerical model to estimate the parameters or state of a physical system of interest. Uncertainties in both measurements and the computational model lead to inaccurate estimates. We developed a goal-oriented aposteriori error estimation methodology for the impact of different errors on the variational solutions of inverse problems. In the goal-oriented approach we are interested in estimating the impact of observation and model errors on the quantity of interest, i.e., on the aspect of interest of the optimal solution.

The variational data assimilation problem optimizes the model states and parameters in order to obtain predictions that fit best the measurements. In this project we solved the complementary problem of optimizing the DDDAS process. The strategies used to collect and process the data are considered to be parameters of the inference system, and are themselves improved via an additional optimization process. Specifically, we seek to improve the parameters of the data assimilation system. We have formulated the optimal configuration of the DDDAS system as the following ``optimization-constrained optimization problem''.
Our algorithm is based on first and second order adjoints, and on the solution of large linear systems. We also proposed efficient methods for computing observation impact, including low-rank approximations of observation-to-analysis sensitivities.

DDDAS variational inference in real time is hindered by costly forward and adjoint model runs. We proposed a new parallel-in-time algorithm to speed up the solution process. The original 4D-Var problem is solved in the augmented Lagrangian framework. To expose time parallelism the assimilation window is divided into several sub-intervals. This formulation allows to integrate the forward and the adjoint models over different subintervals in parallel.

The construction and validation of an adjoint model is an extremely labor-intensive process.
To address this challenge we proposed a derivative-free 4D-Var(-like) smoothing algorithm that approximately solves DDDAS variational inference without the need to construct adjoint models. Specifically, our TR-4D-EnKF algorithm uses a trust-region approach to optimize in ensemble space.

Variational DDDAS inference solution does not include posterior uncertainty estimates. To address this challenge we developed new nonlinear filtering and smoothing algorithms that sample directly from the posterior PDF using a Hybrid Markov-Chain Monte Carlo (HMCMC) approach. The sampling smoother is implemented efficiently using the same adjoint computational infrastructure used in 4D-Var.

Consistency of the reduced-order Karush-Kuhn-Tucker conditions with the full-order optimality conditions is a key ingredient for successful reduced order data assimilation problems. This translates into accurate low-rank approximations of the both adjoint and forward models leading to reduced bases constructed from the dominant eigenvectors of the correlation matrix of the aggregated snapshots of full forward and the adjoint models. Our work underlines the importance of incorporating the adjoint information into the construction of reduced order basis for performing reduced order 4D-Var data assimilation.

Distribution Statement

This is block 12 on the SF298 form.

Distribution A - Approved for Public Release

Explanation for Distribution Statement

If this is not approved for public release, please provide a short explanation. E.g., contains proprietary information.

SF298 Form

Please attach your SF298 form. A blank SF298 can be found here. Please do not password protect or secure the PDF The maximum file size for an SF298 is 50MB.

AFD-070820-035.pdf

Upload the Report Document. File must be a PDF. Please do not password protect or secure the PDF . The maximum file size for the Report Document is 50MB.

report.pdf

Upload a Report Document, if any. The maximum file size for the Report Document is 50MB.

Archival Publications (published) during reporting period:

V. Rao and A. Sandu: ``The Impact of Errors in Dynamic Data Driven Applications Systems". Dynamic Data Driven Application Systems, 2014.

A. Cioaca, A. Sandu, E. de Sturler, and E. Constantinescu: ``Efficient computation of observation impact in 4D-Var data assimilation", in Uncertainty Quantification in Scientific Computing. A.M. Dienstfrey and R.E. Boisvert, editors. IFIP Advances in Information and Communication Technology. Springer Heidelberg Dordrecht London New York, pp. 250--263, 2012. 319 p., ISBN: 978-3-642-32676-9.

V. Rao, A. Sandu, M. Ng, and E. Nino: ``Robust Data Assimilation Using L$_1$ and Huber Norms". Submitted to SIAM Journal on Scientific Computing, \#M104591, Oct. 2015.

V. Rao and A. Sandu: ``A Time-parallel Approach to Strong-constraint Four-dimensional Variational Data Assimilation". Submitted to Journal of Computational Physics, May 2015.

E. Nino and A. Sandu: `Ensemble Kalman Filter Implementations Based on Shrinkage Covariance Matrix Estimation". Ocean Dynamics, 2015.

V. Rao and A. Sandu: ``A-posteriori Error Estimates for Inverse Problems". SIAM/ASA Journal on Uncertainty Quantification, Vol. 3, No. 1, pp. 737--761, 2015.

E. Nino and A. Sandu: ``A Derivative-Free Trust Region Framework for Variational Data Assimilation". Journal of Computational and Applied Mathematics, Vol. 293, pp. 164--179, 2016.

A. Cioaca and A. Sandu: ``An Optimization Framework to Improve 4D-Var Data Assimilation System Performance". Journal of Computational Physics, Vol. 275, pp. 377--389, 2014.

E. Nino, A. Sandu, and J.L. Anderson: ``An Efficient Implementation of the Ensemble Kalman Filter Based on an Iterative Sherman-Morrison Formula". Statistics and Computing, Vol. 25, Issue 3, pp. 561-577, 2015.

A. Cioaca and A. Sandu: ``Low-rank Approximations for Observation Impact in 4D-Var Data Assimilation". Computers and Mathematics with Applications, Vol. 67, Issue 12, pp. 2112--2126, 2014.

V. Rao and A. Sandu: ``An Adjoint-Based Scalable Algorithm for Time-Parallel Integration". Journal of Computational Science, Vol. 5, Issue 2, pp. 76-84, 2014.

A. Cioaca, A. Sandu, and E. de Sturler: ``Efficient Methods for Computing Observation Impact in 4D-Var Data Assimilation". Computational Geosciences, Vol. 17, Issue 6, pp. 975--990, 2013.

A. Cioaca, M. Alexe, and A. Sandu: ``Second Order Adjoints for Solving PDE-Constrained Optimization Problems". Optimization Methods and Software, Volume 27, Issue 4-5, pp. 625--653, 2012.

Changes in research objectives (if any):

N/A.

Change in AFOSR Program Manager, if any:

N/A.

Extensions granted or milestones slipped, if any:

N/A.

AFOSR LRIR Number

LRIR Title

Reporting Period

Laboratory Task Manager

Program Officer

Research Objectives

Technical Summary

Funding Summary by Cost Category (by FY, $K)

	Starting FY	FY+1	FY+2
Salary			
Equipment/Facilities			
Supplies			
Total			

Report Document

Report Document - Text Analysis

Report Document - Text Analysis

Appendix Documents

2. Thank You

E-mail user

Dec 12, 2015 21:47:22 Success: Email Sent to: sandu@cs.vt.edu