# Preconditioning Representer-based Variational Data Assimilation Systems: Application to NAVDAS-AR

Boon S. Chua, Liang Xu, Tom Rosmond and Edward D. Zaron

**Abstract** Assimilation of observations into numerical models has emerged as an essential modeling component in geosciences. This procedure requires the solution of large systems of linear equations. Solving these systems in "real-time" or "near-real-time" in a timely manner is still a computational challenge. This paper shows how new methods in computational linear algebra are used to "speed-up" the representer-based algorithm in a variety of assimilation problems, with particular application to the Naval Research Laboratory (NRL) Atmospheric Variational Data Assimilation System-Accelerated Representer (NAVDAS-AR) system.

## 1 Introduction

The representer-based algorithm is deployed in variational data assimilation systems such as (i) Inverse Ocean Modeling system (IOM) (Chua and Bennett, 2001; Bennett et al., 2008) and (ii) Naval Research Laboratory (NRL) Atmospheric Variational Data Assimilation System-Accelerated Representer (NAVDAS-AR) (Xu et al., 2005; Rosmond and Xu, 2006). Both weak and strong constraint variational assimilation (Sasaki, 1970) may be accomplished with this algorithm, which is formulated in the terms of dual variables (the so-called, "observation space"), the dimension of which is generally much smaller than the corresponding state space (Courtier, 1997). This restriction to the observation space is accomplished by discarding the unobservable degrees of freedom in the system (Bennett, 2002). The power of the representer algorithm as an iterative solver derives from this reduction of degrees of freedom, and results in considerable improvement in solver efficiency as compared to state space algorithms (Zaron, 2006). Despite its computational advantages over other assimilation algorithms, however, solving "real-time", "near-real-time", and "real-world" problems in a timely fashion using this algorithm is still a computational challenge. In this paper we explore some of the most recent methods in computational linear algebra (Simoncini and Szyld, 2007), and implement these methods to "speed-up" the representer-based algorithm.

B.S. Chua (✉)
SAIC, Monterey, CA, USA

S.K. Park, L. Xu, *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications*, DOI 10.1007/978-3-540-71056-1_16,

The IOM system is a Graphical User Interface (GUI)-driven system for configuring and running weak-constraint four-dimensional variational assimilation (W4DVAR) with any model. All of the many model-dependent steps in the optimization algorithms are automatically and correctly generated. The user simply enters the model details via the GUI, and then the master program written in Parametric Fortran (Erwig et al., 2006) will generate the optimization steps as customized Fortran-90 code. If the user supplies the forward nonlinear model coded in Parametric Fortran, the IOM will automatically generate its tangent linear and adjoint models in Fortran-90. The IOM generates the run scripts for partial or complete inversion, as selected by the user via the GUI. In addition to computing the ocean circulation estimates, the IOM also computes dynamical residuals, data residuals, significance tests and posterior error covariances. The IOM has been applied to a level model (free-surface Bryan and Cox), a sigma-coordinate model (ROMS), a finite-element model (ADCIRC) and a spectral-element model (SEOM). Users experiences with it are documented in Muccino et al. 2008. In this paper the methods tested in the IOM system are further extended to an atmospheric variational data assimilation system, NAVDAS-AR.

NAVDAS-AR is an observation space four-dimensional variational data assimilation system based on the accelerated representer algorithm (Xu and Daley, 2000; Chua and Bennett, 2001). It is designed to become the U.S. Navy's next generation operational atmospheric data assimilation system. A wide variety of observations have to be accurately assimilated in a timely fashion. The number of observations assimilated in the current pre-operational NAVDAS-AR tests is around 500,000 during every 6 hour data assimilation window. We anticipate the number of observations to be assimilated will be around 1,000,000 during each 6 hour data assimilation window in the near future. The observations being routinely assimilated are: conventional observations, aircraft observations, feature tracked winds, AMSU-A, SSM/I total precipitable water and wind speeds, scatterometer winds, and Australia synthetic observations. The major computational cost of NAVDAS-AR is the minimization of a weighted least-squares cost function, made up of three computational components: an adjoint model integration, a background error covariance calculation, and a tangent-linear model integration. The cost of these is primarily a function of the model resolution chosen, and surprisingly the cost is only weakly dependent on the number of observations. This is in stark contrast to the NRL Atmospheric Variational Data Assimilation System (NAVDAS), which is the U.S. Navy's current operational three-dimensional assimilation system. The computational cost of NAVDAS is quadratically dependent on the number of observations, a serious liability with anticipated increases in observation volume. Details of the differences between NAVDAS and NAVDAS-AR are discussed in Rosmond and Xu (2006).

To satisfy the operational time constraint, the minimization of the cost function has to be made extremely efficient without degrading the accuracy of the analysis. In this paper we explore the benefits of using an efficient solver and robust preconditioner for NAVDAS-AR.

The outline of this paper is as follows. Section 2 describes three general types of linear algebra problems arising from variational assimilation applications. A linear