NRC Research Press

REVIEW

# Application of dynamic data driven application system in environmental science

Jingwei Song, Bo Xiang, Xinyuan Wang, Li Wu, and Chun Chang



**Abstract:** The paradigm of dynamic data driven application system (DDDAS) has been proposed as a framework to analyze and predict the character and behavior of complex systems that influence computational models significantly. Its accuracy and efficiency lies in its ability to integrate observations on different temporal and spatial scales from real-time sensors, and in its measurement steering and controlling capabilities. Many problems in environmental sciences are nonlinear and complex, impossible to solve by using input/output sequence flows without feedback control. Nonlinear system efficiency depends on measurement control and steering, on-line data assimilation, and model selection with dynamic optimization. Compared with traditional methods, DDDAS possesses the capacity to overcome these limitations. This paper discusses DDDAS and classifies typical cases of its application in environmental sciences into three levels of paradigm. Short reviews of multi-model simulation and data assimilation are provided for practical use. Recent developments and future perspectives are reviewed. Future work may address determining automatically where, when, and how to acquire real-time data, and its integration with GIS, to improve efficiency and accuracy. User-generated content will find wide application in the future. Considering the differences between DDDAS and other data-driven methods in solving the same nonlinear complex system problems, a combination of nonlinear science and chaos theory is advocated.

*Key words:* dynamic data driven application system, GIS, nonlinear system, environmental science.

**Résumé :** On a proposé le paradigme de système d'application dynamique géré par les données (SADGD) comme cadre pour analyser et prédire le caractère et le comportement de systèmes complexes influençant significativement les modèles informatiques. Sa précision et son efficacité résident dans sa capacité à intégrer les observations à différentes échelles spatio-temporelles provenant de senseurs en temps réel, et dans ses mesures dirigeant et contrôlant ses capacités. Plusieurs problèmes en environnement sont non linéaires et complexes, impossibles à résoudre en utilisant le flux de séquences entrées/sorties sans mécanisme rétroactif de contrôle. L'efficacité des systèmes non linéaires dépend du contrôle et de la gestion des mesures, de l'assimilation des données en ligne et du modèle de sélection avec optimisation dynamique. Comparativement aux modèles traditionnels, le SADGD possède la capacité de surmonter ces limitations. Les auteurs discutent du SADGD et classifient des cas typiques de ses applications en sciences de l'environnement, selon trois degrés de paradigmes. Les auteurs proposent de courtes revues de simulation multi modèles et d'assimilation de données, pour utilisation pratique. On passe en revue les développements récents et les perspectives futures. Les progrès futurs pourraient viser à déterminer automatiquement où, quand et comment obtenir les données en temps réel et son intégration avec les GIS pour améliorer l'efficacité et la précision. Le contenu généré par les usagers trouvera de larges applications dans le futur. Considérant les différences entre le DDDAS et les autres méthodes dirigées par les données pour résoudre les mêmes problèmes de systèmes complexes non linéaires, on préconise une combinaison de la science non linéaire avec la théorie du chaos. [Traduit par la Rédaction]

*Mots-clés :* système d'application dynamique géré par les données, GIS, système non linéaire, science de l'environnement.

## 1. Introduction

The computational concept dynamic data driven application system (DDDAS) was first introduced in the early 1980s while using the Monte Carlo and discrete ordinates methods to compute radiation transport for simulations and measurements relating to oil exploration. Originally, it was envisioned to accelerate computation by using additional experimental data in selective places (Darema 2004).

Later, when utilized in various fields, the content of DDDAS was enriched and expanded. The National Science Foundation (NSF) defines DDDAS as a paradigm with "the ability to dynamically incorporate data into an executing application simulation, and in reverse, the ability of applications to dynamically steer measurement processes" (Darema et al. 2005). Among the various choices for solving nonlinear complex problems, DDDAS is attractive because of its flexibility with respect to on-line data assimilation, measurement control, and model selection and optimization. It is an ideal method by which to analyze, manipulate, and simulate complex systems.

Traditionally, DDDAS has been applied in meteorology (Kanamitsu 1989; Maraun et al. 2010; Pagowski et al. 2010), oceanography (Cummings 2005; Carton and Giese 2008), hydrology (Vermeulen et al. 2005; Blaas et al. 2007; El Serafy et al. 2007; Li 2007), geography (Dey and Singh 1999; Li et al. 2004; Liang and Qin 2008), and other environmental systems. Even though the names of the methods might be different they convey the same para-

Received 12 November 2013. Accepted 12 March 2014.

**J. Song, B. Xiang, and C. Chang.** Key Laboratory of Digital Earth Sciences, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China; University of Chinese Academy of Sciences, Beijing 100049, China.
**X. Wang.** Key Laboratory of Digital Earth Sciences, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China.
**L. Wu.** College of Territorial Resources and Tourism, Anhui Normal University, Wuhu 241002, China.
**Corresponding author:** Xinyuan Wang (e-mail: wangxy@radi.ac.cn).

digm, as described by NSF DDDAS (2013; www.cise.nsf.gov/dddas). Later, novel applications and technologies were achieved in other fields. DDDAS was introduced as a main framework with detailed techniques, such as fast Fourier transform, ensemble Kalman filter, and genetic algorithms for coupling data from diverse sources in simulating and predicting the spread of wildfire (Mandel et al. 2007; Denham et al. 2008; Mandel et al. 2010, 2012). These researchers defined and refined an integral workflow covering data acquisition, data assimilation, model selection, and web publishing that can be used as references for environmental monitoring and decision making. Furthermore, owing to its flexibility, DDDAS has been used in emergency systems to integrate different media and sensors via real-time communication, feedback control, and real-time processing (Schoenharl and Adviser-Madey 2007; Chen et al. 2011; Hawe et al. 2012). Recently, a DDDAS has been successfully introduced in monitoring volcanic ash propagation (Patra et al. 2012) and hazard analysis and real-time estimation of total phosphorus load in a Mississippi Delta stream (Ouyang et al. 2013). DDDAS has also been applied in many other fields, such as earth observation sensor web architecture enhancement (Moghaddam et al. 2007), subsurface water spill problems (Douglas et al. 2006), agro-ecosystem modeling (Dorigo et al. 2007), and even in industrial engineering (Koyuncu et al. 2007; Williams et al. 2013). However, a complete implementation of DDDAS is demanding of both time and effort; therefore, some examples mentioned above are incomplete applications that will be addressed in detail. This paper is structured as follows: (*i*) introduction to the basic structure of the DDDAS paradigm; (*ii*) description of three levels of the DDDAS paradigm and their pros and cons; and (*iii*) an overview of future prospects of DDDAS with regard to both technology and application.

## 2. Concept and structure

The concept of DDDAS comprises three basic components: data, measurement, and algorithm. DDDAS data do not just refer to traditional off-line data, but also to real-time or near real-time data, which are incorporated instantly into the simulation system. On-line or off-line data can be either continuous or discrete depending on the actual situation. Traditional paradigms have predetermined measurements and data acquisition sensors, frequencies, and processing algorithms that are fixed once the system starts operating. DDDAS, however, can adjust dynamically and steer the measurement component to gather data at the optimum time and site. DDDAS can also adapt or choose the optimum algorithm according to the output and real-time data, whereas a traditional model is fixed and cannot steer the measurement or modify the algorithm by dynamic compilation.

The core feature of DDDAS lies in its ability to inject data dynamically, to steer the measurement, and to select the optimum algorithm. This provides DDDAS with a high level of freedom that is crucial in the simulation of complex systems. Examples can be found in meteorology in which systems are highly nonlinear and complex (Brutsaert 1982). A complex system is often described by a numerical simulation method, and simulation errors are caused by model error and errors in initial conditions (Harlim 2006). Furthermore, a complex system comprises both intrinsic and extrinsic stimuli interacting with each other, which makes the resulting system model so complex that we have to simplify the system representation in the model so that the numerical simulations are tractable. In some cases, the observations are describing the system more realistically than the simulation itself. However, simulation is essential if we want to examine and analyze the system, and make predictions. Data assimilation (DA) is raised to solve initial error propagation and its impact on the fidelity of the results of the simulation model. For example, the typical approach of three/four dimensional variational (3/4-D VAR) DA has been introduced successfully in meteorology, and errors have been restricted to acceptable levels (Gustafsson et al. 2001; Rawlins et al. 2007).

**Fig. 1.** Schematic structure of dynamic data driven application system (DDDAS) comprising five basic components: measurement steering, measurement selection, application model selecting, runtime environment, and prediction. These parts can adjust themselves to form a self-adaptive DDDAS. Real-time data are acquired from a sensor web consisting of ground observations, satellite images, and other sensors.

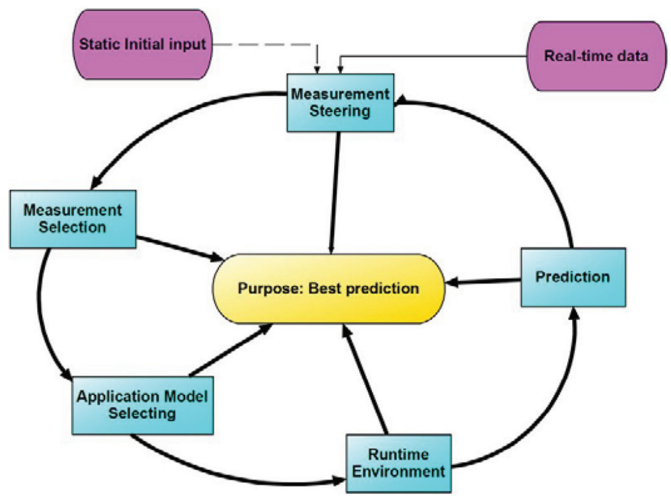
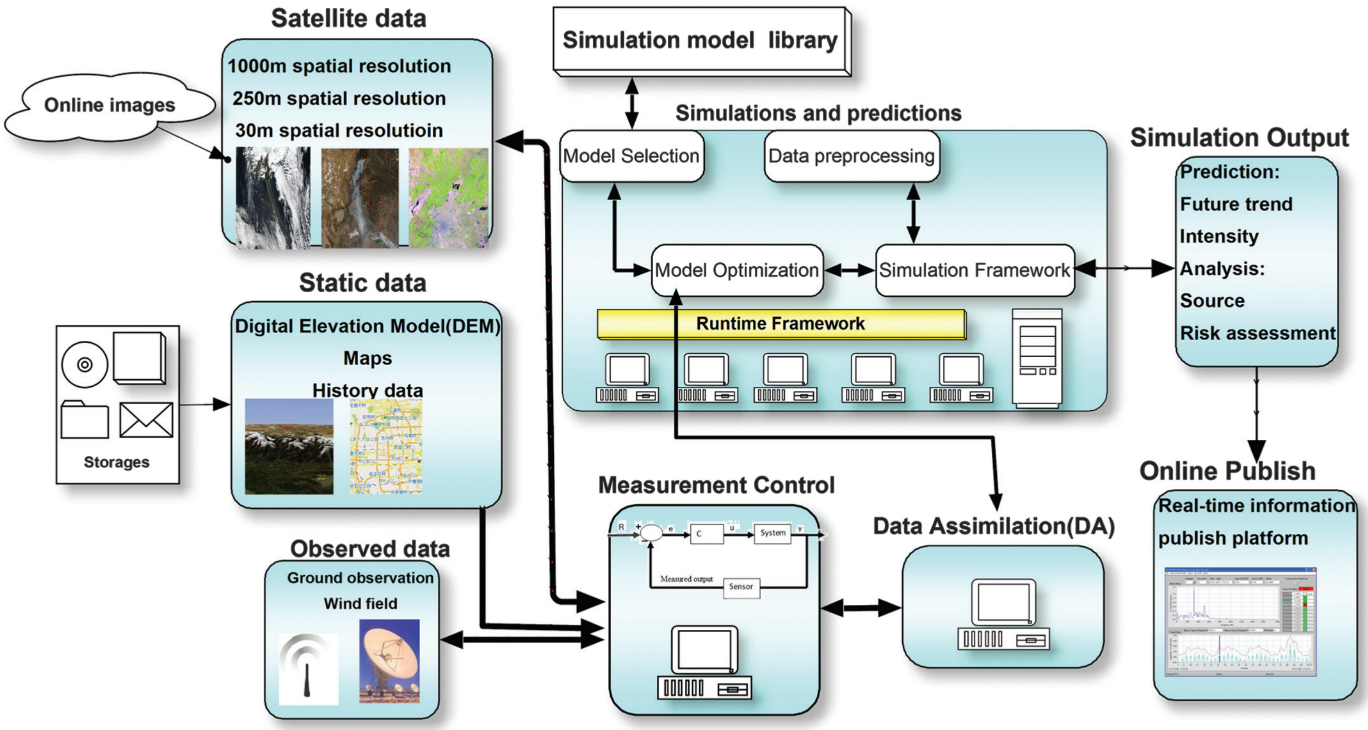



Figure 1 is an abstract schematic structure showing the basic symbiotic components and functions of DDDAS, which includes five parts: measurement steering, measurement selection, application model selecting, runtime environment, and prediction. Real-time data act like a propeller injecting new data into the dynamic model, keeping it looping and iterating for the purpose of obtaining the best prediction. However, it is also steered and controlled by the measurement steering component for obtaining the best data source. All components of this model are dynamic and adjust to the ever-changing nonlinear complex system, except for the initial input data. In most cases, the simulations exhibit sensitive dependence on the initial values and if not controlled, the simulation system and the object system will bifurcate and their orbits will separate. If new measurements are injected into the simulation system at the appropriate time, they will synchronize with the real system before bifurcation. In this way, the error propagation of the simulation system is restricted to a certain level and errors caused by the butterfly effect in the nonlinear system, as discovered first by Lorenz (1963), can be closely mitigated.

Figure 2 is a typical example of a DDDAS of a real-time natural hazards monitoring system chosen after consulting several existing systems (Mandel et al. 2007, 2012; Denham et al. 2008; Moghaddam et al. 2010; Allaire et al. 2012). The five parts in the abstract model (see Fig. 1) are all presented in this system with measurement steering, measurement control, and prediction corresponding to measure control, DA, and simulation output, respectively. Application model selecting and the runtime environment framework are included as the simulations and predictions model. Satellite data and static data are either downloaded from the internet or are loaded from storage devices, but they cannot be controlled by the system. Observed data are flexible, allowing the system to decide dynamically the measurement rate or indicators. Currently, this feature is more important in mechanisms such as unmanned vehicles, so that intelligent machines are equipped with the ability to make decisions based on the actual situation (Allaire et al. 2012). After one loop, feedback is sent back to the simulation and prediction module to determine whether to

**Fig. 2.** Typical dynamic data driven application system (DDDAS)-based paradigm.



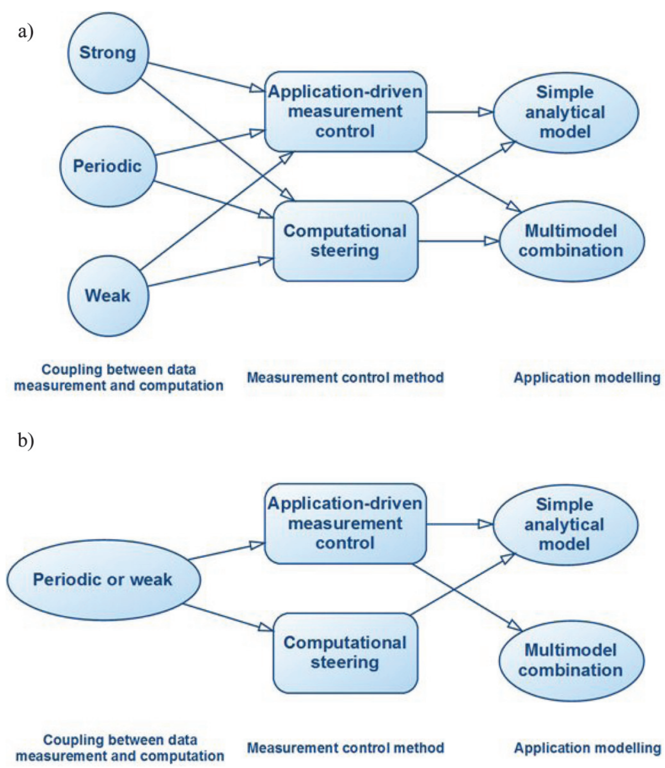

change to an alternate simulation model, refine the parameters of the current model, or to adjust the measurements. When using data in environmental sciences, DA and simulation processes may run on remote computers without human interference. Further analysis can be published on a website, such that end users have access to its origin, status, and future trends.

## 3. Levels of DDDAS in current applications

The synergistic coupling of computation and measurement of a DDDAS using the W-tuple (S, D, M, R, A) has been presented previously by Darema (2011). In her theory, a DDDAS comprises static data inputs (S), dynamic data inputs (D), measurement steering (M), application modeling (A), and runtime environment (R). Data available to the application when commencing the execution is defined as static data inputs (S). Data at execution time is defined as dynamic data inputs (D). Measurement steering (M) is real-time adaptation of heterogeneous and varying time-scales, modalities, and formats of measurements. Application modeling (A) designates the application simulation model with many degrees of freedom and multimodal components interfaces for mathematical or statistical representation. Runtime environment (R) is the general system software supporting heterogeneous computational and measurement environments.

Various types of instantiations of DDDAS can be categorized by measurement and computation coupling, measurement control method, and model type from the W-tuple theory. We summarize the W-tuple theory from Darema (2011) and present the decision tree in Fig. 3*a*. Strong coupling between measurement and computation means the on-line or real-time data are continually injected into the system, while periodic or weak coupling indicates the data are injected periodically or at sparse frequency. Application-driven measurement control is used to dynamically adjust measurement method; it could degrade to computational steering only if simple parametric adjustments are induced by the user. Application model is classified as a simple analytical model and model library. The difference between these two models is whether or not there are multiple models.

**Fig. 3.** W-tuple based decision tree of general dynamic data driven application system (DDDAS) (*a*) and one adapted to an environmental system (*b*).




Cases in environmental sciences have their own characteristics. Their study objects are often macro-systems involving atmosphere, water, forest, soil, and animal or plants species. Observed data from these complex data sources are heterogeneous with widely varying quality, and it is hard to unify them for system processing. Based on these facts, we simplified the general DDDAS decision tree to a concise decision tree (Fig. 3*b*) suitable for describing environmental problems applications. Strong coupling between measurement and computation are impossible because satellite and many ground sensors cannot provide continuous observations. Periodic and weak coupling are merged considering these two measurement intervals can be switched based on actual demand in environmental cases. And if the simulation system is consisted of multi-model combination, in most cases, it will require measurement control to satisfy different model's input needs.

This simplified definition more accurately covers current DDDAS applications and highlights the important features in environmental sciences and engineering. As Darema (2011) points out, many systems implement only part of the definition that happens especially in the environmental DA system (Li et al. 2004; Carton and Giese 2008; Liang and Qin 2008; Pagowski et al. 2010). A large sum of incomplete DDDAS applications suggests that the paradigm maybe more suitably described by levels of models ranging from simple to complex. After having undertaken an in-depth analysis of general DDDAS applications with the W-tuple theory, the DDDAS paradigm fit for current environmental applications as shown in Fig. 3*b* can be classified into three levels.

### 3.1. The basic level

Figure 4 is the flow paradigm of a basic level of DDDAS with only a data stream. Its W-tuple is reduced to (S, "DataAssim"/p, –, A, R). Where A is just a single traditional analytical model. Its dynamic input can either be just parameter adjustment or DA or both of them. After the initial input is set, the model functions to produce predictions, which are then merged with on-line data in the DA module. As on-line data are injected continuously into the system, new observation data are assimilated into the system, and errors caused by the initial field and the simulation's defects are limited to an acceptable level. In this module, various types of DA algorithm are employed to integrate the on-line data (Ma and Qin 2012). Thus, the parameters of the model are improved to fit the complex system, and the error bars are limited to acceptable levels. The new input data with the DA method synchronize the simulation and real system orbits. This simple level is applicable, although it comes at a price because its measurement sampling intervals cannot be modified. This simulation requires that the bifurcation point be reached before new data are assimilated into the system (Lorenz 1993). After the bifurcation point, the simulation becomes unreliable. Furthermore, its accuracy is also affected by the simulation algorithm, because where a major factor or an interaction process is not considered when designing the algorithm, the orbit of the simulation could soon separate from that of the real system. However, this simulation model is suitable in geospatial information related studies, where data are acquired via a satellite or airborne sensor and the data acquisition time lag is fixed, and thus, measurement steering is not applicable (Barnes et al. 1998). Most simulation or prediction needs where remote sensing data are used can be sufficiently met at the basic level of DDDAS.

### 3.2. The measurement control level

In some systems, dynamic measurement steering plays an important role in the simulation, especially in unmanned vehicles and aircraft. Its W-tuple is represented as (S, "DataAssim"/p, M, A, R). Other simulations will obtain better performance if the sensor networks have the ability to determine dynamically the appropriate method of measurement (e.g., what, when, and how to measure). To cope with the ever-changing environment, the simulation system adjusts its measurements based on the needs of the algorithm.

**Fig. 4.** Basic level of dynamic data driven application system (DDDAS). Data assimilation (DA) is responsible for on-line data into the model. No control stream is included in this system.

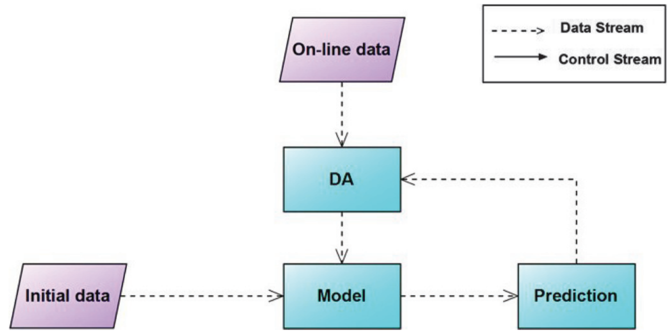



Figure 5 shows the paradigm of adding measurement control into the basic-level DDDAS with a data stream and control stream. After assimilating new real-time data, it will adjust automatically its measurement method to acquire new data appropriate for the situation. Thus, the effort of the sensors or machinery is reduced and accurate useful information gathered. In environmental sciences, where many observations are made from ground-based stations, it is possible for a system to determine the time and method of data acquisition. Ouyang et al. (2011) have employed simulations to guide real-time data measurement (download data dynamically from the USGS) in estimating real-time N (nitrogen) load in surface water. Moghaddam et al. (2010) have adopted measurement control, by modifying the sampling rate and other parameters of in situ sensors, to achieve minimum energy costs in optimal measurement of surface-to-depth profiles of soil moisture. Data acquired by sensors can be classified as continuous or discrete data. For discrete data, the sampling intervals are very important because intervals that are too sparse lead to unreliable results, whereas intervals that are too frequent may lead to redundant data or a rise in energy consumption. Measurement control of DDDAS will adjust the measurements dynamically according to the requirements of balancing the sampling rate with other related parameters.

### 3.3. The complete level

Figure 6 is the complete-level DDDAS, which fits the definition of Darema (2004, 2011, 2012) and the NSF. Its W-tuple is the complete DDDAS as (S, "DataAssim"/p, M, model_lib, R). In this level, a model library is added to the system, and model selection and optimization strategies are employed. It is steered and controlled by the DA module to select the ideal model and to adjust the parameters to describe and simulate the complex system. The basic hypothesis is that, in some situations, the systems could be very complex and nonlinear with multiple factors for which no single physical method can achieve good performance. With a model base, different models are able to be compared for the selection of the most suitable and its parameters can be adjusted. The system can select dynamically the best temporal model to simulate the complex system. Allaire et al. (2012) have designed a self-aware aerospace vehicle that can adapt its performance dynamically by gathering information and responding intelligently. Real-time execution of online models and the exploitation of dynamic data streams are achieved by employing a multi-fidelity approach. Bazilevs et al. (2013) have executed computational models from measurement data and used control strategy to steer the measurements. However, at the time of writing this paper, only one report (Patra et al. 2012) on the application of a complete-level DDDAS, as shown in Fig. 6, in an environmental science problem had been found. However, many more opportunities for new applications of complete-level DDDAS for addressing environmental sciences problems can be envisioned; we discuss such perspectives next.

**Fig. 5.** A measurement control level of dynamic data driven application system (DDDAS). Compared with Fig. 4, this DDDAS possesses a control stream and measurement control module. They cooperate to obtain the best data source for the needs of the simulation model.

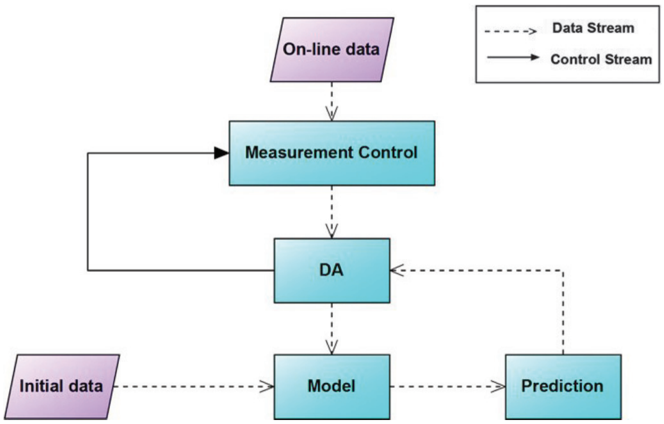
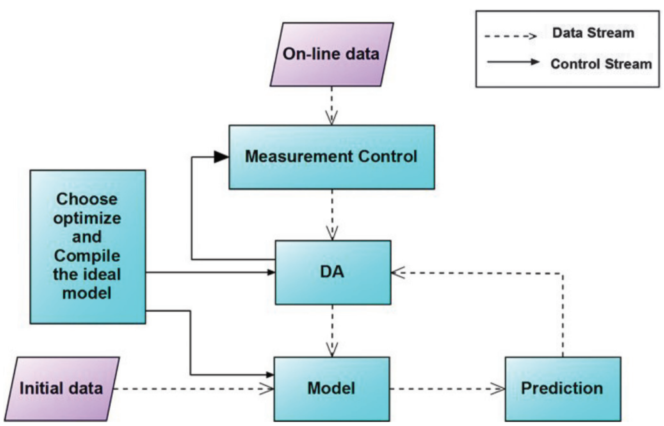



**Fig. 6.** The complete level of dynamic data driven application system (DDDAS). This paradigm is more complex with a dynamic compiling function. Parameters or functions can be adjusted and compiled according to the latest performance.




## 4. Application strategy

### 4.1. Multi-model simulation and forecast

The three-level DDDAS paradigm provides the alternative structure for a specific environmental application. Even though current mainstream applications stop at the measurement control level (Fig. 5), the complete level with multi-model (Fig. 6) will be used in the near future to overcome the complexity of the environmental problems. There are two strategies for multi-model DDDAS application and management: hybrid forecast and multi-model adaptive control (MMAC).

Hybrid forecast is a method of improving each single model by combining all the candidate models. As different models have different mechanism and parameters, combining them into a single model may compensate disturbance and limit the error bar. Various ways of assigning weight deciding include equal weights, unequal weights, and other deep integration methods. Unequal weight is often carried out by a training data optimization algorithm like particle swarm (Huang et al. 2005) or genetic algorithm (Colorni et al. 1996). Other deep integration methods are only available with regard to specific environmental problem. For example, Xu et al (2013) built a hybrid forecast model in predicting municipal solid waste by taking time scale into consideration. Anctil and Tape (2004) used wavelet decomposition method to decompose the time series data into three subsets, and artificial neural network is then used to forecast each sub-series. These deep integrations are often built based on specific cases and are not universally significant as references. As for weight-based combinations, Smith and Wallis (2009) assert that optimal combinations of equal weights or weights close to equality often outperform more complicated weighting schemes. We believe for a specific problem, before choosing the best method, all the available weights assigning methods should be tested on the training data first.

MMAC was first proposed by Lainiotis (1971). This method is based on the hypothesis that the current condition can be simulated by one or a set of relatively simple models. Unlike hybrid forecast, it has a supervisory controller (control model) responsible for switching between different models. Three different switching strategies are as follows: direct multiple models adaptive control, indirect multiple models adaptive control, and weighted multiple models adaptive control algorithms. Direct multiple models adaptive control labels candidate models and compare their performances to determine the optimal model and corresponding parameters. As Zhivoglyadov et al (2000) have pointed out, this exhaustive search may converge very slowly, resulting in excessive transients that will lead the system "unstable" in a practical sense. However, most environmental problems do not have high near-real-time demand and sparse data will not allow a continuous model. Indirect multiple models adaptive control establishes sets of candidate models with their parameter ranges as given (off line). The on-line controller switching is based on the performance evaluation of the model sets. This strategy is more efficient but requires prior knowledge of the models and their parameter ranges. Weighted multiple models adaptive control algorithm is similar to direct adaptive method, but the output is the weighted result of the candidate. This method is most flexible because the weight for each candidate can be dynamically adjusted or even abandoned temporally.

### 4.2. Data assimilation

The European Centre for Medium-Range Weather Forecasts (Bouttier and Courtier 1999) defines DA as an "analysis technique in which the observed information is accumulated into the model state by taking advantage of consistency constraints with laws of time evolution and physical properties." Through the definition, to describe the time evolution of the study object and its future trend within a certain range of precision requirement, DA requires constant observation (usually large sum of heterogeneous data in environmental problems) absorption in each analysis cycle and produces the "analysis" as the best estimated current state, and this state can be applied for accurate future prediction. Global data assimilation system (GDAS) (Data Assimilation Team 2014) from the National Oceanic and Atmospheric Administration (NOAA) of USA and the land data assimilation system (LDAS) (NASA 2014) from the National Aeronautics and Space Administration (NASA) are the two most famous DA systems for climatology, hydrology, oceanography, ecology, and other related environmental sciences.

Commonly used distinction for DA algorithms are sequential and nonsequential. Sequential means taking historic and current observations to infer the current and future state, while nonsequential refers to taking all the available data to infer the past true state. Various types of DA algorithm are developed including 4-D optimum interpolation, variational, data assimilation, filtering, Bayesian method, generalized solution data assimilation, adjoint equation method, and artificial neural network. Table 1 lists the most commonly used DA algorithms with their characteristics.

**Table 1.** Commonly used data assimilation algorithms.

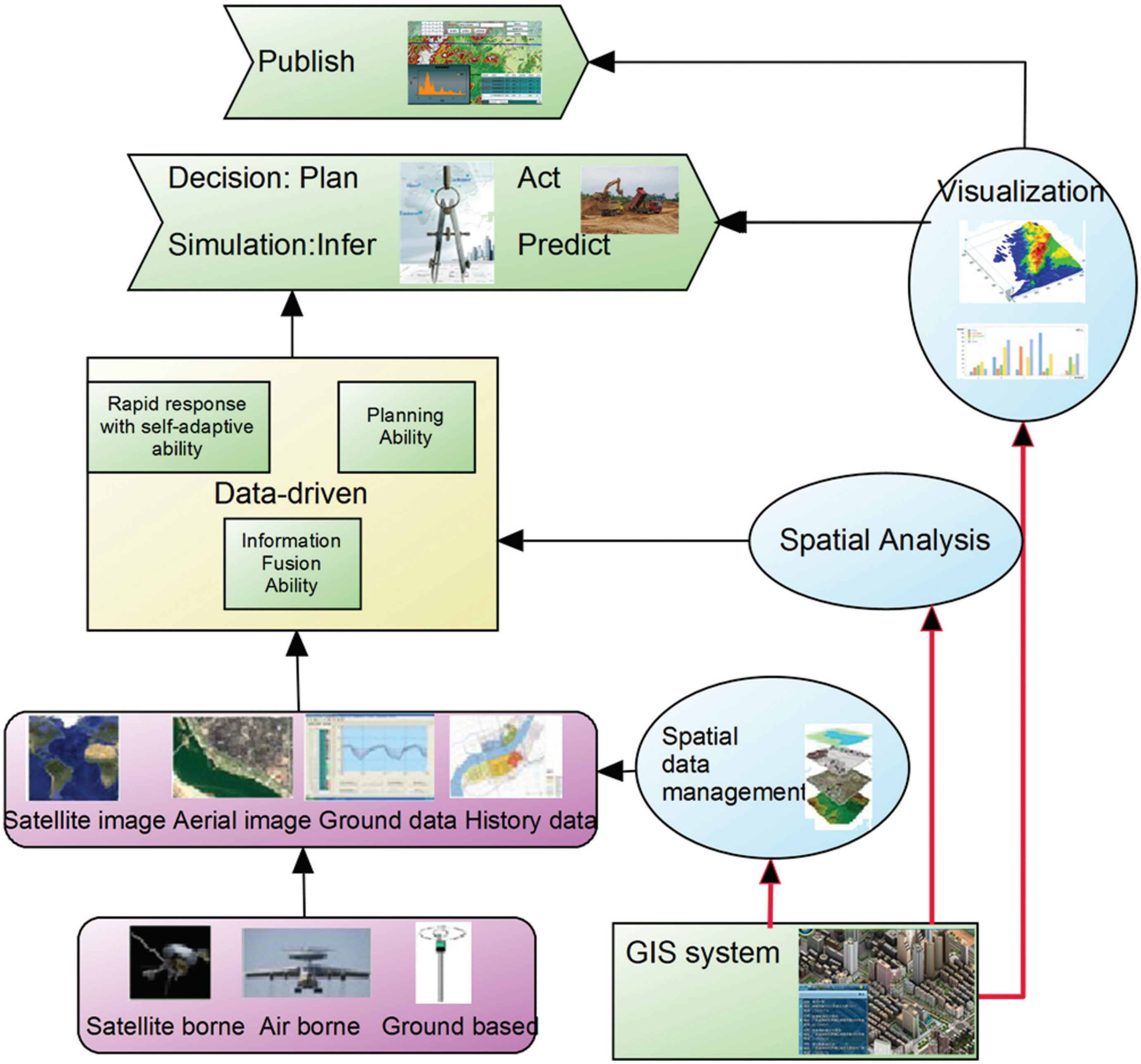
| Algorithm | Mechanism and character | Applications |
|---|---|---|
| 3-D VAR | Find the minimum of a penalty function that measures the size of a control variable and the misfit between observations and corresponding prediction in a time-window. No account is taken of the actual time of each observation. Adjoint models and the functional gradients are often required in calculating penalty function. | Weather simulation (Gustafsson et al. 2001). A typhoon bogusing case study by fifth-generation Mesoscale weather model (Barker et al. 2004). Ozone and fine particulate matter observations (Pagowski et al. 2010). |
| 4-D VAR | Improves 3-D VAR by taking time into consideration. Increment is accumulated at each iteration in the time-window. Adjoint models and the functional gradients are also needed. Requires longer computation time. It usually has better performance than 3-D VAR though it takes more time (Lorenc and Rawlins 2005). | Meteorological system simulation (Dimet and Talagrand 1986; Rawlins et al. 2007). Soil moisture estimation (Reichle et al. 2001). |
| Kalman filter | Estimate statically optimum state of the system from predicted state and observed state using minimum mean square error. The optimum state is used as the initial value for the model to make forecast. This algorithm is not suitable for models with nonlinear transferring function. | Navigation and positioning (Cooper and Durrant-Whyte, 1994). Spatio-temporal prediction of snow water equivalent (Huang and Cressie 1996). Ocean climate (Carton and Giese 2008). |
| Ensemble Kalman filter (EnKF) | A hybrid model of ensemble forecast and Kalman Filter. Basic assumption is that all probability distributions involved are Gaussian. This method is suitable for nonlinear complex models and consumes relatively less time. It is currently the most widely used model in environmental science and is now available in OpenDA (OpenDA Association 2014). | (Houtekamer and Mitchell 2001). Hydrology modelling (Reichle et al. 2002). Atmospheric simulation land surface variable estimation (Liang and Qin 2008). |
| Particle filter | Ensemble members (particles) are set first for sampling and probability distribution function is analyzed accordingly. It is suitable for nonlinear models. Different from EnKF, this algorithm is free of Gaussian distribution assumption. The random process makes it easy to be for parallel computing. | Hydrologic simulation (Moradkhani et al. 2005). Soil moisture simulation (Qin et al. 2009). |
| Artificial neural network (ANN) | Using history observations and corresponding next step predictions in a time-window to train a neural network. Make future predictions according to current observation and model prediction. Artificial intelligent models are black-box model and its performance highly depends on its data. | 1-D shallow water simulation (Härter and de Campos Velho 2008). Storm surge prediction (Siek and Solomatine 2011). |

## 5. Future perspectives

### 5.1. Technical breakthroughs

The integration of remote sensing data and ground-based observations in GIS with an intelligent model base is a potential application, especially in hydrology and meteorology. For example, in most places, MODIS (moderate resolution imaging spectroradiometer) collects images four times per day when combining the two Terra and Aqua EOS satellites (Freeborn et al. 2011). Compared with satellite imagery, in situ measurement is time consuming, inconvenient, and can only retrieve data at specific locations. However, a few studies have reported that the quantitative retrieval process from satellite images is unreliable owing to incomplete retrieval algorithms, cloud contamination, and aerosol impact (Kutser 2004; Zhang and Reid 2009). Because of the uncertainty of remote sensing images, complementary ground-based observations enable the algorithm to make better predictions. Ouyang et al. (2011) and Moghaddam et al. (2010) have employed successfully a method to control ground observations to improve the performance of sensors, while maintaining prediction precision. As Fig. 7 shows, GIS can provide a spatial data management service, a spatial analysis tool, and visualization methods as aids for massive data storage, model optimization, and decision making, respectively. GIS spatial database can offer a significant capability of organizing, storing, and managing multi-source data. GIS spatial analysis method can provide spatial data mining tool for model optimization and refinement. The visualization techniques of GIS can assist in decision making and multimedia publishing. However, there are still three principal problems with these space-air-ground sensor systems. (*i*) Because there are many numerical models available, collating them into a single model base and organizing it dynamically to yield optimum results systematically remains a difficult task. Efforts have been made, for example by the OpenMI Association (2013), to define an interface that allows different models to exchange data simultaneously at runtime. Successful applications of OpenMI can be found in many application areas (Becker and Schuttrumpf 2011; Betrie et al. 2011; Bulatewicz et al. 2012). Section 4.1 provides three model controlling strategies (see Figs. 4, 5, and 6) that can be carried out in common applications. But for those applications with high demand in efficiency and accuracy, the technical details for OpenMI type of applications are very sophisticated and require breakthroughs in cybernetics and systems theory. (*ii*) Because GIS is built for spatial data storage and spatial analysis, it is necessary to synchronize the steps of GIS and DDDASs with regard to measurement control and model selection. Methods for achieving satisfactory levels of synchronization are needed. (*iii*) Different data acquisition methods have different spatial and temporal scales (Hu 2009; Shang et al. 2011), which lead to different conclusions, and fusing them into one unified system remains difficult. Figure 7 illustrates a GIS-based DDDAS in an environmental science application. GIS can provide the spatial database to assimilate and normalize the massive amount of data acquired for future management and geocomputing. Spatial analysis can optimize and configure the geocomputing and simulation system. The visualization method can be used for both decision making and on-line publishing. The result of geocomputing and the efficiency of the

**Fig. 7.** GIS-based dynamic data driven application system (DDDAS).




simulation are largely dependent on the cooperative work of GIS spatial analysis, the physical model, and the data-driven models. Because spatial expertise knowledge is stored within the GIS and physical expertise knowledge lies within the DDDAS, an intelligent processing system is essential for harmonizing the different types of knowledge. Currently, a GIS provides a runtime framework for simulation and modeling. Madey et al. (2007) have adopted GIS to map and visualize scenarios of disaster and emergency. Chen et al. (2011) have used a GIS (NetLogo GIS) directly as a framework to implement their simulation and integrate buffer analysis into their model.

As many research subjects relating to DDDAS have chaotic behavior, nonlinear data-driven methodologies have been introduced. Originally, these data-driven methodologies have been used for nonlinear time series forecasting in fields such as finance (Hsieh 1991; Kim 2003), meteorology (Shukla 1998), power load (Niu et al. 2010), and landslide forecasting (Huaming et al. 2003). Among the many theories, artificial neural networks and the phase space reconstruction method stand out as the most useful. In this regard, pioneering work was performed by Härter et al. (2008) by employing a radial-basis function neural network for assimilating data by emulating an ensemble Kalman filter to improve a 1-D shallow water model, and by Siek and Solomatine (2011) who used a nonlinear autoregressive exogenous inputs neural network for DA of chaotic storm surge models. There is a natural connection between physical-based model and chaos theory. For example, the physically based Navier–Stokes equations, which are used widely in meteorology and hydrology, have proven to be chaotic and have Lorenz attractors (Lorenz 1963). Siek (2011) listed several DA methods and chaotic prediction methods, and proposed the idea of combining data (time series) from observations, using the well-developed methods of nonlinear dynamics and chaos theory, while assimilating new data into the system to improve performance. Because chaos is the essence of complex systems, such as weather and water-bodies, the potential applications of nonlinear dynamics and chaos theory should be of great interest to scientists working in these fields.

User-generated content (UGC) is a tool for collecting data from the Web, news media, or volunteers. Data from UGC are flexible, abundant, and timely with low fidelity and high levels of redundancy. They can be used as complementary data in some cases where data acquisition is difficult or for data validation in many applications. In some cases public participation can provide comprehensive and in-time data at very low costs, and if integrated into the DDDAS, the efficiency and performance of the simulating system will be enhanced. Schweik et al. (2005) raised the idea OS/OC (open source/open content) for collaborative approaches for social–ecological research in general and mentioned an example in land-use/land-cover change. Maisonneuve et al. (2009) built a mobile phone based citizen noise monitoring prototype implementation and tested the feasibility and credibility of this system by evolving volunteers providing the noise and corresponding position data. Taking advantage of this data acquisition method, Hasenfratz et al. (2012) built a DDDAS with multiple stationary and mobile sensors web for ozone pollution and achieved reasonably good results.

### 5.2. Future applications

Applications in meteorology, oceanography, hydrology, geography, and limnology will progress technically in GIS-integrated DDDAS, advanced machine learning algorithm, and chaotic model, UGC strategy described above. Additional spatial and temporal data will be injected into the system, the DA algorithms will be refined, and the GIS will be integrated to handle the massive amount of data required for improving the performance of geocomputing. Nonlinear dynamics and chaos theory will be considered and potential application will be advocated further.

DDDAS has great potential in social and management science and can be applied widely in the fields of natural disaster rapid response (Madey et al. 2007), health monitoring (Cortial et al. 2007), and crime prevention or other related social sciences (Kennedy and Theodoropoulos 2006). Social sciences have more nonlinearity and complexity, and thus, they rely on comprehensive understanding, dynamic measurement steering, and behavioral analysis. In social sciences, massive amounts of information are gathered from multimedia sources and the Internet. Without a good paradigm, the "big data" era may just be data rich, but information poor and inconsistencies in the massive amounts of data may lead to confusion. Such defects could be eliminated by designing a DDDAS that incorporates new data and limits uncertainty.

Air pollution has become increasingly important in urban areas, particularly in rapidly growing economies such as China, and is closely linked to public health. During the first half of 2013 in Beijing, only 38.9% percent of days met government standards (Ministry of Environmental Protection, People's Republic of China 2013), for which $PM_{2.5}$ was the primary pollutant. Harrison et al. (2012) provided a study on the temporal and spatial patterns of $PM_{2.5}$ and found that the density of $PM_{2.5}$ has so-called diurnal patterns or steep rises and falls during the day, and complex spatial variation due to wind direction and speed. Even though methods of simulation and prediction have provided valuable insights into the spatial and temporal distribution of air pollutants, inadequate input data with associated uncertainties hinders the effectiveness and accuracy of the models. In addition, an increasing number of air pollutant monitoring stations are being built to provide a better picture of $PM_{2.5}$ distribution. If the DDDAS paradigm is introduced to these sensor webs, the monitoring system will possess the capabilities of sampling interval control and sensor spatial distribution control, reducing sampling effort and improving treatment efficiency. Similar work was done by Hasenfratz et al. (2012) who used DDDAS for targeted data gathering from multiple stationary and mobile sensors for ozone pollution. Their framework can be implemented easily into other air quality monitoring and assessment system with portable hardwares, low-cost devices, and reliable results.

Other applications include biological systems, which need comprehensive understanding using various types of models. For example, when building symbiotic simulation systems to describe various kinds of symbiotic relationships, more than one model is used to build a hybrid simulation system. DDDAS is an appropriate choice because it is a paradigm that has the ability to incorporate new data dynamically and in reverse, steer the measurement processes (Aydt et al. 2008). Biodiversity, invasions of exotic species, and floods and droughts are all suitable research areas for applying DDDAS to analyze the complex processes and to gain insights into such phenomena. The basic assumption of using DDDAS in ecology and biology has been described when the DDDAS initiative was launched at NSF, but real implementations of complete-level DDDAS in those areas are seldom reported. This could be because building a sensor network to monitor species is a very difficult task; and because there are many technical problems associated with DDDAS that have not yet been solved completely. Some applications have introduced DDDAS without measurement control or model selection and optimization strategies (Szewczyk et al. 2004; Nagendra et al. 2012; Maffey et al. 2013). UGC has also been introduced into wildlife data acquisition. Zhang et al. (2012) mined data from photos collected from social networks like Flickr and Twitter as complementary data to ground stations and satellite images. Their studies could further be integrated into ecological modelling and monitoring.

## 6. Conclusion

This paper presents a review of DDDAS focusing on their basic concept and structure, three different levels of paradigm, and future perspectives. A synthesis of the literature reveals the wide range of applications of DDDAS in environmental sciences, especially with regard to complex nonlinear systems. As DDDAS is an integrated system covering many types of hardware and software, integrating different types of methods and algorithms is a problem that needs to be solved. Furthermore, the methods of chaos theory and machine learning have become established tools in environmental sciences, whereas the combination of model-driven and data-driven methods is a new analytical approach. UGC may be widely applied into environmental sciences for providing complementary data and increase credibility. Dynamic data-driven methods bring additional dimensions and capabilities into these approaches. Because of the spatial and temporal variation in the different data acquisition methods, unifying these data remains difficult. Further studies could investigate and analyze the cooperative work of model-based and data-based models, and design targeted controlling flow-work in solving environmental problems. Chaos and nonlinear theories could be considered and applied in some specific areas, such as meteorology and hydrology.

With the step of increasing spatial and temporal resolution, the DDDAS paradigm has the potential capability of assimilating and coordinating remote sensing data and ground-based observation data into a single system for better prediction of nonlinear complex phenomena. As data acquisition techniques are improved, and advanced models and simulations provided, the fields of application may well expand beyond meteorology, oceanography, hydrology, geography, limnology, biology, and environmental management. However, unwanted and unused data in this "big data" era may contaminate the model and simulation. DDDAS provides methods to dynamically and adaptively select and manage data, thus mitigating the "data deluge" of "big data" and allowing the efficient exploitation of such data. Therefore, to avoid dilemmas due to the massive amounts of data, advanced algorithms, intelligent sensor webs, and fundamental and multidisciplinary research are all needed to implement a full-level paradigm for better performance of DDDAS.

## Acknowledgements

This review was supported by the program of International S&T Cooperation "Fined Earth Observation and Recognition of the Impact of Global Change on World Heritage Sites" (grant No. S2013GR0477), the National Natural Science Foundation of China (grant No. 41271427), and the Foundation of IRSD (Institute of remote sensing and digital earth, Chinese Academy of Sciences) "Application of remote sensing in field water resource detection" (grant No. YIC006101B). The authors are very grateful to Natarajan Ishwaran, Wanchang Zhang, and Chuansheng Liu for many valuable comments and critical editorial guidance. The authors would also like to thank two anonymous reviewers for their helpful suggestions.

## References

Allaire, D., Biros, G., Chambers, J., Ghattas, O., Kordonowy, D., and Willcox, K. 2012. Dynamic data driven methods for self-aware aerospace vehicles. Procedia Computer Science, **9**: 1206–1210. doi:10.1016/j.procs.2012.04.130.

Anctil, F., and Tape, D.G. 2004. An exploration of artificial neural network rainfall-runoff forecasting combined with wavelet decomposition. J. Environ. Eng. Sci. **3**(S1): S121–S128. doi:10.1139/s03-071.

Aydt, H., Turner, S.J., Cai, W., and Low, M.Y.H. 2008. Symbiotic simulation systems: an extended definition motivated by symbiosis in biology. *In* Principles of Advanced and Distributed Simulation, 2008. PADS '08. 22nd Workshop on IEEE. pp. 109–116. doi:10.1109/PADS.2008.17.

Barker, D.M., Huang, W., Guo, Y.R., Bourgeois, A.J., and Xiao, Q.N. 2004. A three-dimensional variational data assimilation system for MM5: implementation and initial results. Mon. Weather Rev. **132**(4): 897–914. doi:10.1175/1520-0493(2004)132<0897:ATVDAS>2.0.CO;2.

Barnes, W.L., Pagano, T.S., and Salomonson, V.V. 1998. Prelaunch characteristics of the moderate resolution imaging spectroradiometer (MODIS) on EOS-AM1. IEEE Trans. Geosci. Remote Sens. **36**(4): 1088–1100. doi:10.1109/36.700993.

Bazilevs, Y., Hsu, M.C., and Bement, M.T. 2013. Adjoint-based control of fluid-structure interaction for computational steering applications. Procedia Computer Science, **18**: 1989–1998. doi:10.1016/j.procs.2013.05.368.

Becker, B., and Schuttrumpf, H. 2011. An OpenMI module for the groundwater flow simulation programme Feflow. J. Hydroinf. **13**(1): 1–12. doi:10.2166/hydro.2010.039.

Betrie, G.D., Van Griensven, A., Mohamed, Y.A., Popescu, I., Mynett, A.E., and Hummel, S. 2011. Linking SWAT and SOBEK using open modeling interface (OPENMI) for sediment transport simulation in the Blue Nile River basin. Trans. ASABE, **54**(5): 1749–1757. doi:10.13031/2013.39847.

Blaas, M., El Serafy, G.Y., van Kessel, T., De Boer, G.J., Eleveld, M.A., and Van der Woerd, H.J. 2007. Data model integration of spm transport in the Dutch coastal zone. *In* Proceedings of the Joint 2007 EUMETSAT/ASM Conference. pp. 1–8. Available from https://www.eumetsat.int/cs/idcplg?IdcService=GET_FILE&dDocName=pdf_conf_p50_s5_07_blaas_v&allowInterrupt=1&noSaveAs=1&RevisionSelectionMethod=LatestReleased [accessed 23 September 2013].

Bouttier, F., and Courtier, P. 1999. Data assimilation concepts and methods [online]. Available from http://www.ecmwf.int/newsevents/training/rcourse_notes/DATA_ASSIMILATION/ASSIM_CONCEPTS/Assim_concepts2.html [accessed 26 Februray 2014].

Brutsaert, W. 1982. Evaporation into the atmosphere: theory, history, and applications. Reidel, Dordrecht. pp. 1–20.

Bulatewicz, T., Allen, A., Peterson, J.M., Staggenborg, S., Welch, S.M., and Steward, D.R. 2012. The simple script wrapper for openMI: enabling interdisciplinary modeling studies. Environ. Modell. Software. **39**: 283–294. doi:10.1016/j.envsoft.2012.07.006.

Carton, J.A., and Giese, B.S. 2008. A reanalysis of ocean climate using simple ocean data assimilation (SODA). Mon. Weather Rev. **136**(8): 2999–3017. doi:10.1175/2007MWR1978.1.

Chen, F., Zhai, Z., and Madey, G. 2011. Dynamic adaptive disaster simulation: developing a predictive model of emergency behavior using cell phone and GIS data. *In* Proceedings of the 2011 Workshop on Agent-Directed Simulation. Society for Computer Simulation International. pp. 5–12. Available from http://dl.acm.org/citation.cfm?id=2048356 [accessed 23 September 2013].

Colorni, A., Dorigo, M., Maffioli, F., Maniezzo, V., Righini, G., and Trubian, M. 1996. Heuristics from nature for hard combinatorial optimization problems. Int. Trans. Oper. Res. **3**(1): 1–21. doi:10.1111/j.1475-3995.1996.tb00032.x.

Cooper, S., and Durrant-Whyte, H. 1994. A Kalman filter model for GPS navigation of land vehicles. *In* Intelligent Robots and Systems '94. Advanced Robotic Systems and the Real World, IROS '94. Proceedings of the IEEE/RSJ/GI International Conference on IEEE. Vol. 1, pp. 157–163. doi:10.1109/IROS.1994.407396.

Cortial, J., Farhat, C., Guibas, L.J., and Rajashekhar, M. 2007. Compressed sensing and time-parallel reduced-order modeling for structural health monitoring using a DDDAS. *In* Computational Science–ICCS 2007. Springer Berlin Heidelberg. pp. 1171–1179. doi:10.1007/978-3-540-72584-8_153.

Cummings, J.A. 2005. Operational multivariate ocean data assimilation. Q. J. R. Meteorol. Soc. **131**(613): 3583–3604. doi:10.1256/qj.05.105.

Darema, F. 2004. Dynamic data driven applications systems: a new paradigm for application simulations and measurements. *In* Computational Science-ICCS 2004. Springer Berlin Heidelberg. pp. 662–669. doi:10.1007/978-3-540-24688-6_86.

Darema, F. 2011. DDDAS computational model and environments. J. Algorithms Comput. Technol. **5**(4): 545–560. doi:10.1260/1748-3018.5.4.545.

Darema, F. 2012. Dynamic data driven application systems (DDDAS). Available from http://www.dddas.org/nsf-workshop-2006/talks/NSF/Darema.pdf [accessed 23 September 2013].

Darema, F., Rotea, M., Goldberg, M., Newlon, D.H., Cherniavsky, J.C., Figueroa, J.E., Hudson, J.E., Friedman, C., Lyster, P., and Bohn, R. 2005. DDDAS: dynamic data driven applications systems [online]. Available from http://www.nsf.gov/pubs/2005/nsf05570/nsf05570.htm [accessed 23 September 2013].

Data Assimilation Team. 2014. Grid-point statistical interpolation [online]. Available from http://www.emc.ncep.noaa.gov/gmb/gdas/ [accessed 26 February 2014].

Denham, M., Cortés, A., Margalef, T., and Luque, E. 2008. Applying a dynamic data driven genetic algorithm to improve forest fire spread prediction. *In* Computational Science–ICCS 2008. Springer Berlin Heidelberg. pp. 36–45. doi:10.1007/978-3-540-69389-5_6.

Dey, S., and Singh, R.P. 1999. Surface latent heat flux as an earthquake precursor. Nat. Hazards Earth Syst. Sci. **3**(6): 749–755. doi:10.5194/nhess-3-749-2003.

Dimet, F.X.L., and Talagrand, O. 1986. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. Tellus A, **38**(2): 97–110. doi:10.1111/j.1600-0870.1986.tb00459.x.

Dorigo, W.A., Zurita-Milla, R., de Wit, A.J., Brazile, J., Singh, R., and Schaepman, M.E. 2007. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. Int. J. Appl. Earth Obs. Geoinf. **9**(2): 165–193. doi:10.1016/j.jag.2006.05.003.

Douglas, C., Efendiev, Y., Ewing, R., Ginting, V., and Lazarov, R. 2006. Dynamic data driven simulations in stochastic environments. Computing, **77**(4): 321–333. doi:10.1007/s00607-006-0165-3.

El Serafy, G.Y., Blaas, M., Eleveld, M.A., and van der Woerd, H.J. 2007. Data assimilation of satellite data of suspended particulate matter in Delft3D-WAQ for the North Sea. *In* Proceedings of the Joint EUMETSAT/AMS Conference, Darmstadt, Germany. pp. 1–8. Available from https://www.eumetsat.int/cs/idcplg?IdcService=GET_FILE&dDocName=pdf_conf_p50_s5_02_elserafy_p&allowInterrupt=1&noSaveAs=1&RevisionSelectionMethod=LatestReleased [accessed 23 September 2013].

Freeborn, P.H., Wooster, M.J., and Roberts, G. 2011. Addressing the spatiotemporal sampling design of MODIS to provide estimates of the fire radiative energy emitted from Africa. Remote Sens. Environ. **115**(2): 475–489. doi:10.1016/j.rse.2010.09.017.

Gustafsson, N., Berre, L., Hörnquist, S., Huang, X.-Y., Lindskog, M., Navascues, B., Mogensen, K.S., and Thorsteinsson, S. 2001. Three-dimensional variational data assimilation for a limited area model. Tellus A, **53**(4): 425–446. doi:10.1111/j.1600-0870.2001.00425.x.

Harlim, J. 2006. Errors in the initial conditions for numerical weather prediction: a study of error growth patterns and error reduction with ensemble filtering. Available from http://drum.lib.umd.edu/handle/1903/3430 [accessed 23 September 2013].

Harrison, R.M., Laxen, D., Moorcroft, S., and Laxen, K. 2012. Processes affecting concentrations of fine particulate matter ($PM_{2.5}$) in the UK atmosphere. Atmos. Environ. **46**: 115–124. doi:10.1016/j.atmosenv.2011.10.028.

Härter, F.P., and de Campos Velho, H.F. 2008. New approach to applying neural network in nonlinear dynamic model. Appl. Math. Modell. **32**(12): 2621–2633. doi:10.1016/j.apm.2007.09.006.

Hasenfratz, D., Saukh, O., Sturzenegger, S., and Thiele, L. 2012. Participatory air pollution monitoring using smartphones. Mobile Sensing. Available from ftp://yosemite-ife.ee.ethz.ch/pub/people/hdavid/HSST2012.pdf [accessed 3 February 2014].

Hawe, G.I., Coates, G., Wilson, D.T., and Crouch, R.S. 2012. Agent-based simulation for large-scale emergency response: a survey of usage and implementation. ACM Computing Surveys (CSUR), **45**(1), 8. Available from http://dl.acm.org/citation.cfm?id=2379784 [accessed 23 September 2013].

Houtekamer, P.L., and Mitchell, H.L. 2001. A sequential ensemble Kalman filter for atmospheric data assimilation. Mon. Weather Rev. **129**(1): 123–137. doi:10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2.

Hsieh, DA. 1991. Chaos and nonlinear dynamics: application to financial markets. J. Finance, **46**(5): 1839–1877. doi:10.1111/j.1540-6261.1991.tb04646.x.

Hu, C. 2009. A novel ocean color index to detect floating algae in the global oceans. Remote Sens. Environ. **113**(10): 2118–2129. doi:10.1016/j.rse.2009.05.012.

Huaming, L., Huan, Q., and Zhiqiang, C. 2003. Nonlinear chaotic model of landslide forecasting. Chin. J. Rock Mech. Eng. **22**(3): 434–437. Available from http://en.cnki.com.cn/Article_en/CJFDTOTAL-YSLX200303019.htm [accessed 23 September 2013].

Huang, C.M., Huang, C.J., and Wang, M.L. 2005. A particle swarm optimization to identifying the ARMAX model for short-term load forecasting. IEEE Trans. **20**(2): 1126–1133. doi:10.1109/TPWRS.2005.846106.

Huang, H.C., and Cressie, N. 1996. Spatio-temporal prediction of snow water equivalent using the Kalman filter. Computational Statistics and Data Analysis, **22**(2): 159–175. doi:10.1016/0167-9473(95)00047-X.

Kanamitsu, M. 1989. Description of the NMC global data assimilation and forecast system. Weather and Forecasting, **4**(3): 335–342. doi:10.1175/1520-0434(1989)004<0335:DOTNGD>2.0.CO;2.

Kennedy, C., and Theodoropoulos, G. 2006. Intelligent management of data driven simulations to support model building in the social sciences. *In* Computational Science–ICCS 2006. Springer Berlin Heidelberg. pp. 562–569. doi:10.1007/11758532_74.

Kim, K.J. 2003. Financial time series forecasting using support vector machines. Neurocomputing, **55**(1): 307–319. doi:10.1016/S0925-2312(03)00372-2.

Koyuncu, N., Lee, S., Vasudevan, K.K., Son, Y.J., and Sarfare, P. 2007. DDDAS-based multi-fidelity simulation for online preventive maintenance scheduling in semiconductor supply chain. *In* Simulation Conference, 2007 Winter. pp. 1915–1923. doi:10.1016/S0925-2312(03)00372-2.

Kutser, T. 2004. Quantitative detection of chlorophyll in cyanobacterial blooms by satellite remote sensing. Limnol. Oceanogr. **49**(6): 2179–2189. doi:10.4319/lo.2004.49.6.2179.

Lainiotis, D.G. 1971. Optimal adaptive estimation: structure and parameter adaption. IEEE Trans. Autom. Control, **16**(2): 160–170. doi:10.1109/TAC.1971.1099684.

Li, X., Koike, T., and Pathmathevan, M. 2004. A very fast simulated re-annealing (VFSA) approach for land data assimilation. Comput. Geosci. **30**(3): 239–248. doi:10.1016/j.cageo.2003.11.002.

Li, Y. 2007. An integrated water quality modeling system with dynamic remote sensing feedback. Ph.D. thesis, Rochester Institute of Technology. Available from http://www.cis.rit.edu/~cnspci/references/theses/phd/li2007.pdf [accessed 23 September 2013].

Liang, S., and Qin, J. 2008. Data assimilation methods for land surface variable estimation. *In* Advances in Land Remote Sensing. Springer Netherlands. pp. 313–339. doi:10.1007/978-1-4020-6450-0_12.

Lorenc, A.C., and Rawlins, F. 2005. Why does 4D-Var beat 3D-Var? Q. J. R. Meteorol. Soc. **131**(613): 3247–3257. doi:10.1256/qj.05.85.

Lorenz, E.N. 1963. Deterministic nonperiodic flow. J. Atmos. Sci. **20**(2): 130–141. doi:10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.

Lorenz, E.N. 1993. The Essence of Chaos. The University of Washington Press, Washington. p. 25.

Ma, J., and Qin, S. 2012. Recent advances and developments of data assimilation algorithms. Adv. Earth Sci. **27**(7): 747–757. Available from http://mall.cnki.net/magazine/Article/DXJZ201207007.htm [accessed 23 September 23, 2013]. [In Chinese with English summary.]

Madey, G.R., Barabási, A.L., Chawla, N.V., Gonzalez, M., Hachen, D., Lantz, B., Pawling, A., Schoenharl, T., Szabo, G., Wang, Pu., and Yan, P. 2007. Enhanced situational awareness: application of DDDAS concepts to emergency and disaster management. *In* Computational Science–ICCS 2007. Springer, Berlin Heidelberg. pp. 1090–1097. doi:10.1007/978-3-540-72584-8_143.

Maffey, G., Reed, M., Irvine, R.J., and van der Wal, R. 2013. Habitat monitoring in the wider countryside: a case study on the pursuit of innovation in red deer management. J. Environ. Manage. **128**: 779–786. doi:10.1016/j.jenvman.2013.06.022. PMID:23860380.

Maisonneuve, N., Stevens, M., Niessen, M.E., Hanappe, P., and Steels, L. 2009. Citizen noise pollution monitoring. *In* Proceedings of the 10th Annual International Conference on Digital Government Research: Social Networks: Making Connections between Citizens, Data and Government. pp. 96–103. Available from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.159.1782&rep=rep1&type=pdf [accessed 22 February 2014].

Mandel, J., Beezley, J.D., Bennethum, L.S., Chakraborty, S., Coen, J.L., Douglas, C.C., Hatcher, J., Kim, M., and Vodacek, A. 2007. A dynamic data driven wildland fire model. *In* Computational Science–ICCS 2007. Springer, Berlin Heidelberg. pp. 1042–1049. doi:10.1007/978-3-540-72584-8_137.

Mandel, J., Beezley, J.D., and Kondratenko, V.Y. 2010. Fast Fourier transform ensemble Kalman filter with application to a coupled atmosphere-wildland fire model. *In* Proceedings of the MS '10 International Conference. pp. 777–784. doi:10.1142/9789814324441_0089.

Mandel, J., Beezley, J.D., Kochanski, A.K., Kondratenko, V.Y., and Kim, M. 2012. Assimilation of perimeter data and coupling with fuel moisture in a wildland fire–atmosphere DDDAS. Procedia Computer Science, **9**: 1100–1109. doi:10.1016/j.procs.2012.04.119.

Maraun, D., Wetterhall, F., Ireson, A.M., Chandler, R.E., Kendon, E.J., Widmann, M., Brienen, S., Rust, H.W., Sauter, T., Themeßl, M., Venema, V.K.C., Chun, K.P., Goodess, C.M., Jones, R.G., Onof, C., Vrac, M., and Thiele-Eichand, I. 2010. Precipitation downscaling under climate change: recent developments to bridge the gap between dynamical models and the end user. Rev. Geophys. **48**(3). RG3003. doi:10.1029/2009RG000314.

Ministry of Environmental Protection, The People's Republic of China. 2013. Available from http://www.mep.gov.cn/gkml/hbb/qt/201308/t20130802_256884.htm [accessed 23 September 2013].

Moghaddam, M., Entekhabi, D., Liu, M., and Teneketzis, D. 2007. Integrating data assimilation and optimal control for enhancing sensor web architectures. *In* NASA Earth Science Technology Office, Advanced Information Systems Sensor Web PI Meeting, February, pp. 13–14. Available from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.99.660&rep=rep1&type=pdf [accessed 23 September 2013].

Moghaddam, M., Entekhabi, D., Goykhman, Y., Li, K., Liu, M., Mahajan, A., and Teneketzis, D. 2010. A wireless soil moisture smart sensor web using physics-based optimal control: concept and initial demonstrations. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **3**(4): 522–535. doi:10.1109/JSTARS.2010.2052918.

Moradkhani, H., Hsu, K.L., Gupta, H., and Sorooshian, S. 2005. Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. Water Resour. Res. **41**(5): W05012. doi:10.1029/2004WR003604.

Nagendra, H., Lucas, R., Honrado, J.P., Jongman, R.H., Tarantino, C., Adamo, M., and Mairota, P. 2012. Remote sensing for conservation monitoring: assessing protected areas, habitat extent, habitat condition, species diversity, and threats. Ecol. Indic. **33**: 45–59. doi:10.1016/j.ecolind.2012.09.014.

NASA. 2014. Land data assimilation system [online]. Available from http://ldas.gsfc.nasa.gov/ [accessed 26 February 2014].

Niu, D.X., Wang, Y.L., and Ma, X.Y. 2010. Optimization of support vector machine power load forecasting model based on data mining and Lyapunov exponents. J. Central South University of Technology, **17**: 406–412. doi:10.1007/s11771-010-0060-0.

NSF DDDAS. 2013. Dynamic data driven applications systems, NSF program solicitation [online]. Available from www.cise.nsf.gov/dddas [accessed 23 September 2013].

OpenDA Association. 2014. Integrating models and observations [online]. Available from http://www.openda.org/joomla/ [accessed 25 February 2014].

OpenMI Association. 2013. About the OpenMI association [online]. Available from http://www.openmi.org [accessed 12 October 2013].

Ouyang, Y., Luo, S.M., Cui, L.H., Wang, Q., and Zhang, J.E. 2011. Estimation of real-time N load in surface water using dynamic data-driven application system. Ecol. Eng. **37**(4): 616–621. doi:10.1016/j.ecoleng.2010.12.023.

Ouyang, Y., Leininger, T.D., and Hatten, J. 2013. Real-time estimation of TP load in a Mississippi Delta stream using a dynamic data driven application system. J. Environ. Manage. **122**: 37–41. doi:10.1016/j.jenvman.2013.02.047. PMID:23542568.

Pagowski, M., Grell, G.A., McKeen, S.A., Peckham, S.E., and Devenyi, D. 2010. Three-dimensional variational data assimilation of ozone and fine particulate matter observations: some results using the Weather Research and Forecasting-Chemistry model and Grid-point Statistical Interpolation. Q. J. R. Meteorol. Soc. **136**(653): 2013–2024. doi:10.1002/qj.700.

Patra, A., Bursik, M., Dehn, J., Jones, M., Pavolonis, M., Pitman, E.B., Singh, T., Singla, P., and Webley, P. 2012. A DDDAS framework for volcanic ash propagation and hazard analysis. Procedia Computer Science, **9**: 1090–1099. doi:10.1016/j.procs.2012.04.118.

Qin, J., Liang, S., Yang, K., Kaihotsu, I., Liu, R., and Koike, T. 2009. Simultaneous estimation of both soil moisture and model parameters using particle filtering method through the assimilation of microwave signal. J. Geophys. Res.: Atmos. **114**(D15): 16. doi:10.1029/2008JD011358.

Rawlins, F., Ballard, S.P., Bovis, K.J., Clayton, A.M., Li, D., Inverarity, G.W., and Payne, T.J. 2007. The Met Office global four-dimensional variational data assimilation scheme.Q. J. R. Meteorol. Soc. **133**(623): 347–362. doi:10.1002/qj.32.

Reichle, R.H., Entekhabi, D., and McLaughlin, D.B. 2001. Downscaling of radio brightness measurements for soil moisture estimation: A four-dimensional variational data assimilation approach. Water Resources Research, **37**(9): 2353–2364. doi:10.1029/2001WR000475.

Reichle, R.H., McLaughlin, D.B., and Entekhabi, D. 2002. Hydrologic data assimilation with the ensemble Kalman filter. Mon. Weather Rev. **130**(1): 103–114. Available from http://nml.yonsei.ac.kr/front/bbs/paper/das/DAS_2002-2_Reichle_et_al.pdf [accessed 23 February 2013].

Schoenharl, T.W., and Adviser-Madey, G. 2007. Creating, updating and validating simulations in a dynamic data-driven application system. University of Notre Dame. Available from http://dl.acm.org/citation.cfm?id=1925475 [accessed 23 September 2013].

Schweik, C., Evans, T., and Grove, J.M. 2005. Open source and open content: a framework for global collaboration in social-ecological research. Ecol. Soc. **10**(1): 33. Available from http://www.nrs.fs.fed.us/pubs/jrnl/2005/nrs_2005_schweik_002.pdf [accessed 23 February 2014].

Shang, L., Ma, R., Duan, H., Jiang, G., and Zhou, L. 2011. Scale analysis of cyanobacteria bloom in Lake Taihu from MODIS observations. J. Lake Sci. **23**(6): 847–854. Available from http://www.cnki.com.cn/Article/CJFDTOTAL-FLKX201106003.htm [accessed 23 September 2013]. [In Chinese with English summary.]

Shukla, J. 1998. Predictability in the midst of chaos: a scientific basis for climate forecasting. Science, **282**(5389): 728–731. doi:10.1126/science.282.5389.728. PMID:9784127.

Siek, M. 2011. Predicting storm surges: chaos, computational intelligence, data assimilation, ensembles. Ph.D. thesis, Delft University of Technology. Available from http://repository.tudelft.nl/assets/uuid:fb1b8de3-776f-4d4e-8d2b-b01e768bd58c/UNESCO-IHE_PhD_SIEK_THESIS.pdf [accessed 23 September 2013].

Siek, M., and Solomatine, D.P. 2011. Real-time data assimilation for chaotic storm surge model using narx neural network. *In* Proceedings of the 11th International Coastal Symposium. pp. 1189–1194. Available from http://

www.ics2011.pl/artic/SP64_1189-1194_M.Siek.pdf [accessed 23 September 2013].
Smith, J., and Wallis, K.F. 2009. A simple explanation of the forecast combination puzzle. Oxford Bulletin of Economics and Statistics, **71**(3): 331–355. doi:10.1111/j.1468-0084.2008.00541.x.
Szewczyk, R., Osterweil, E., Polastre, J., Hamilton, M., Mainwaring, A., and Estrin, D. 2004. Habitat monitoring with sensor networks. Commun. ACM, **47**(6): 34–40. doi:10.1145/990680.990704.
Vermeulen, C.J.M., Barneveld, H.J., Huizinga, H.J., and Havinga, F.J. 2005. Data-assimilation in flood forecasting using time series and satellite data. *In* International Conference on Innovation, Advances and Implementation of Flood Forecasting Technology. Tromsø, Norway. Available from http://hkvlijninwater.nl/documenten/Data-assimilationinfloodforecastingusingtimeseriesandsatellitedata_CJV_HJB_JH_FHa(1).pdf [accessed 23 September 2013].
Williams, M.O., Li, F., Kevrekidis, P., Daraio, C., and Kevrekidis, Y. 2013. Equation-free computations as DDDAS protocols in the study of engineered granular crystals. Procedia Computer Science, **18**: 2638–2642. doi:10.1016/j.procs.2013.05.456.
Xu, L., Gao, P., Cui, S., and Liu, C. 2013. A hybrid procedure for MSW generation forecasting at multiple time scales in Xiamen City, China. Waste Manage. **33**(6): 1324–1331. doi:10.1016/j.wasman.2013.02.012. PMID:23490364.
Zhang, H., Korayem, M., Crandall, D.J., and LeBuhn, G. 2012. Mining photo-sharing websites to study ecological phenomena. *In* Proceedings of the 21st International Conference on World Wide Web, pp. 749–758. doi:10.1145/2187836.2187938.
Zhang, J., and Reid, J.S. 2009. An analysis of clear sky and contextual biases using an operational over ocean MODIS aerosol product. Geophys. Res. Lett. **36**(15): L15824. doi:10.1029/2009GL038723.
Zhivoglyadov, P.V., Middleton, R.H., and Fu, M. 2000. Localization based switching adaptive control for time-varying discrete-time systems. IEEE Trans. Autom. Control, **45**(4): 752–755. doi:10.1109/9.847116.