Environmental Modelling & Software 93 (2017) 418–435

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Integrating hydrological modelling, data assimilation and cloud computing for real-time management of water resources

Wolfgang Kurtz [a, b, *], Andrei Lapin [c], Oliver S. Schilling [d], Qi Tang [a, b], Eryk Schiller [e], Torsten Braun [e], Daniel Hunkeler [d], Harry Vereecken [a, b], Edward Sudicky [f, g], Peter Kropf [c], Harrie-Jan Hendricks Franssen [a, b], Philip Brunner [d]

[a] Institute of Bio- and Geosciences (IBG-3): Agrosphere, Forschungszentrum Jülich GmbH, Jülich, Germany
[b] Centre for High-Performance Scientific Computing in Terrestrial Systems (HPSC-TerrSys), Geoverbund ABC/J, Jülich, Germany
[c] Computer Science Department (IIUN), Université de Neuchâtel, Neuchâtel, Switzerland
[d] Centre for Hydrogeology and Geothermics (CHYN), Université de Neuchâtel, Neuchâtel, Switzerland
[e] Communication and Distributed Systems (CDS), University of Bern, Bern, Switzerland
[f] Department of Earth and Environmental Sciences, University of Waterloo, Waterloo, Ontario, Canada
[g] Aquanty, Inc., Waterloo, Ontario, Canada

ARTICLE INFO

Article history:
Received 20 November 2016
Accepted 11 March 2017

Keywords:
Cloud computing
Integrated hydrological modelling
Data assimilation
Water resources management
HydroGeoSphere
Wireless sensor networks

ABSTRACT

Online data acquisition, data assimilation and integrated hydrological modelling have become more and more important in hydrological science. In this study, we explore cloud computing for integrating field data acquisition and stochastic, physically-based hydrological modelling in a data assimilation and optimisation framework as a service to water resources management. For this purpose, we developed an ensemble Kalman filter-based data assimilation system for the fully-coupled, physically-based hydrological model HydroGeoSphere, which is able to run in a cloud computing environment. A synthetic data assimilation experiment based on the widely used tilted V-catchment problem showed that the computational overhead for the application of the data assimilation platform in a cloud computing environment is minimal, which makes it well-suited for practical water management problems. Advantages of the cloud-based implementation comprise the independence from computational infrastructure and the straightforward integration of cloud-based observation databases with the modelling and data assimilation platform.

1. Introduction

Hydrological and hydrogeological systems are highly heterogeneous, and the temporal evolution of their spatially variable states is driven by dynamic forcing functions. Deterministic numerical models are an important tool for understanding and managing such systems. Such models can support the water management decision making process with predictions of the temporal evolution and the spatial distribution of target state variables. Groundwater management often relies on simulations with distributed, physically-based hydrological models, e.g., for well field operations adjacent to a river. The available numerical models have greatly improved in recent years. There are, for example, ongoing efforts towards a better description of the dynamic feedbacks between subsurface and surface water processes (Kollet and Maxwell, 2006; Brunner and Simmons, 2012). One of the advantages of such physically-based, fully-coupled (or integrated) surface-subsurface models is that the location of surface water features, such as the position of rivers, no longer needs to be predefined through boundary conditions. They are therefore very well-suited for simulating changing surface water conditions such as floods or droughts.

Deterministic models need to be calibrated based on existing observations. However, there is a growing awareness of the uncertainty related to such deterministic model predictions (Liu et al., 2012). The uncertainties stem from the limited knowledge about the spatial distribution and magnitude of important model parameters, such as hydraulic conductivity or porosity (Chen and Zhang, 2006; Hendricks Franssen and Kinzelbach, 2008). Also, the

* Corresponding author. Institute of Bio- and Geosciences (IBG-3): Agrosphere, Forschungszentrum Jülich GmbH, Jülich, Germany.
E-mail address: w.kurtz@fz-juelich.de (W. Kurtz).

http://dx.doi.org/10.1016/j.envsoft.2017.03.011
1364-8152/

high spatio-temporal variability of boundary conditions and input variables, such as precipitation, can highly affect the quality of model predictions. Moreover, the limited availability of spatial and temporal field data limits the reliability of the calibration process. Finally, the computational requirements of many numerical simulators, especially fully-coupled, physically-based models, exclude in most cases a solid uncertainty analysis. These uncertainties can be substantial and they often undermine the credibility of hydrological and hydrogeological models, especially once it comes to predicting highly dynamic systems.

However, a range of technological and mathematical advances allows overcoming some of the previous limitations. Above all, these advances are related to three key developments: data acquisition techniques, the increasing computational capacities of hydrological models, as well as the integration of measurement data in the modelling process.

Firstly, the acquisition of field data has been greatly facilitated. Traditionally, hydrological field measurements such as piezometer levels, precipitation, soil moisture, discharge or water quality indicators, were acquired either manually in the field at predefined measurement intervals or recorded with data loggers, which have to be read out on a weekly or monthly basis. Ongoing advances in sensor technology and telemetry make it now possible to obtain hydrological data shortly after their acquisition in the field, even for very remote field sites. Wireless sensor networks (WSNs) are increasingly applied in environmental studies, e.g., in the context of soil moisture monitoring (Robinson et al., 2008; Ritsema et al., 2009; Bogena et al., 2010) or surface water monitoring (Li et al., 2011), studies on wetland dynamics (Watras et al., 2014) or the acquisition of solute transport data for modelling purposes (Loden et al., 2009; Barnhart et al., 2010). Such WSNs consist of distributed sensors that transmit the measured data through wireless in-built radio modules to a set of router units that manage the communication within the network (Bogena et al., 2010). As a result, the measured data can be accessed by the user on a permanent storage device in near real-time. This can be of great advantage for water management purposes, especially when a system needs to be controlled and regulated continuously (such as pumps for river bank filtration), and the hydraulic forcings are highly transient (such as the water level in a river).

Secondly, the computational efficiency of hydrological models is continuously increasing. Recent advances in numerical mathematics lead to the development of more efficient solvers and preconditioning techniques (Herbst et al., 2008; Maxwell, 2013). Parallelisation of model codes (Ashby and Falgout, 1996; Vereecken et al., 1996; Jones and Woodward, 2001; Mills et al., 2007) makes it possible to solve hydrogeological problems with a high spatio-temporal resolution and on large scales. These advancements in computational efficiency also facilitated the usage of more sophisticated physical process descriptions in the modelling process. For example, state-of-the-art hydrological models now also provide a full 3D solution of the Richards equation, and a physically consistent coupling between surface and subsurface flow equations (Kollet and Maxwell, 2006; Brunner and Simmons, 2012).

Finally, the combination of sequential data assimilation techniques like the ensemble Kalman filter (EnKF) (Evensen, 1994; Burgers et al., 1998) with hydrogeological models now allows integrating real-time data into the modelling process. These methods can be used to effectively merge uncertain model predictions with uncertain observation data in a Bayesian sense. The uncertainty of model predictions is approximated through the forward simulation of an ensemble of model realisations, where each realisation can have a different combination of initial conditions, model forcings and model parameters. The uncertain model predictions are then sequentially updated with measurement data. In this updating step, the uncertainties in the model predictions and the uncertainties of the observations are optimally weighted and the model predictions are effectively adjusted towards the measured data. Besides the correction of state variables, it is also possible to use observation data to update model parameters jointly with the model states (Hendricks Franssen and Kinzelbach, 2008), which makes these methods very effective calibration tools. This methodology has already been applied to a variety of hydrogeological problems including assimilation of hydraulic head data (Chen and Zhang, 2006; Nowak, 2009), transport problems (Liu et al., 2008; Li et al., 2012), surface water-groundwater interactions (Kurtz et al., 2014; Rasmussen et al., 2015; Tang et al., 2015), assimilation of discharge data (Camporese et al., 2009), operational flood forecasting (Seo et al., 2009; Weerts et al., 2010) and integrated hydrological modelling (Shi et al., 2015; Rasmussen et al., 2015; Kurtz et al., 2016). An application in a hydrogeological setting was given by Hendricks Franssen et al. (2011) for groundwater management of the upper Limmat aquifer in Zurich (Switzerland). In this case, a groundwater model is run on a daily basis to support management decisions on groundwater abstraction, and the EnKF methodology is used to continuously correct the model predictions and model parameters with available hydraulic head data. These corrected model predictions are then used as input for the real-time optimisation of groundwater management activities (Bauser et al., 2010, 2012). In the particular case of the Limmat aquifer, the updated hydraulic head distribution from the groundwater model is used to optimally control the groundwater abstraction from a well field according to predefined management goals, which include the total abstraction rate and the maintenance of certain hydraulic conditions to prevent the leakage of contaminants to the well field from a close-by disposal site. Other applications of real-time optimisation of groundwater resources include the energy efficient operation of well fields (Hansen et al., 2012; Bauer-Gottwein et al., 2016) or the accounting for the thermal regime within an aquifer (Marti, 2014). In Schwanenberg et al. (2011), data assimilation methods are used in conjunction with optimal control algorithms for a large-scale river network.

Such methods, especially in combination with fully-coupled, physically-based hydrological models, are usually associated with a high computational burden due to the need to perform hundreds of model simulations in a Monte Carlo framework. This requires the availability of a dedicated computer infrastructure, which is not readily available for every end-user due to the high personal and financial effort for acquiring and maintaining such systems. This can, in part, be overcome by cloud-based services that provide computational resources on demand, which is seen as an upcoming solution for different environmental applications (Granell et al., 2016). Cloud computing has already been suggested as a future platform for hydrological modelling, model calibration and uncertainty analysis (Hunt et al., 2010; Bürger et al., 2012; Ercan et al., 2014; Zhang et al., 2016) and as a promising tool in the context of decision making in water management (Sun, 2013). Mure-Ravaud et al. (2016) recently also presented an example of a flood forecasting system which is hosted on a cloud server. Such cloud-based solutions are flexible with respect to the choice of the computing environment (operating system, CPU, main memory, etc.), and can thus host a variety of simulation platforms with different computational requirements. Furthermore, such services are paid according to the actually consumed computation time. Therefore, the costs to the end-user are effectively reduced by avoiding the financial overhead that is required for installing and maintaining an own in-house computer system.

This study presents a fully-operational architecture for a cloud-based stochastic real-time prediction and management system in the context of groundwater management. The proposed system