Advances in Water Resources 86 (2015) 327–339

Contents lists available at ScienceDirect

Advances in Water Resources

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

Impact of sensor failure on the observability of flow dynamics at the Biosphere 2 LEO hillslopes

Damiano Pasetto[a,*], Guo-Yue Niu[b,c], Luke Pangle[b], Claudio Paniconi[d], Mario Putti[a], Peter A. Troch[b,c]

[a] Department of Mathematics, University of Padua, Padua, Italy
[b] Biosphere 2—University of Arizona, Tucson, USA
[c] Department of Hydrology and Water Resources, University of Arizona, Tucson, USA
[d] Institut National de la Recherche Scientifique, Centre Eau Terre Environnement (INRS-ETE), Quebec City, Canada

ARTICLE INFO

Article history:
Available online 21 May 2015

Keywords:
Sensor failure analysis
Data assimilation
Landscape Evolution Observatory
Soil moisture
Hydraulic conductivity
Richards' equation

ABSTRACT

The Biosphere 2 Landscape Evolution Observatory (LEO) has been developed to investigate hydrological, chemical, biological, and geological processes in a large-scale, controlled infrastructure. The experimental hillslopes at LEO are instrumented with a large number of different sensors that allow detailed monitoring of local and global dynamics and changes in the hydrological state and structure of the landscapes. Sensor failure, i.e., a progressive reduction in the number of active or working sensors, in such an evolving system can have a dramatic impact on observability of flow dynamics and estimation of the model parameters that characterize the soil properties. In this study we assess the retrieval of the spatial distributions of soil water content and saturated hydraulic conductivity under different scenarios of heterogeneity (different values of correlation length of the random field describing the hydraulic conductivity) and a variable number of active sensors. To avoid the influence of model structural errors and measurement bias, the analysis is based on a synthetic representation of the first hydrological experiment at LEO simulated with the physically-based hydrological model CATHY. We assume that the true hydraulic conductivity is a particular random realization of a stochastic field with lognormal distribution and exponential correlation length. During the true run, we collect volumetric water content measurements at an hourly interval. Perturbed observations are then used to estimate the total water storage via linear interpolation and to retrieve the conductivity field via the ensemble Kalman filter technique. The results show that when less than 100 out of 496 total sensors are active, the reconstruction of volumetric water content may introduce large errors in the estimation of total water storage. In contrast, retrieval of the saturated hydraulic conductivity distribution allows the CATHY model to reproduce the integrated hydrological response of LEO for all sensor configurations investigated.

1. Introduction

Determination of the number and location of sensors needed to monitor a real-world hydrological process is a classical problem in experimental design, where the best compromise between maximum amount of information and minimum number of sensors is sought (e.g., [1,2]). In this framework, an aspect that is rarely taken into consideration is that sensors may fail during long-term experiments, thereby putting at risk the observability of the system since it may not always be possible to replace broken sensors. The lifetime of sensors is thus a crucial unknown in experiments of long duration, and it becomes important to be able to predict how the information obtained from the active sensors changes over time as the sensor network deteriorates.

This is the premise for the present study, which is based on the setup of the Landscape Evolution Observatory (LEO) of the Biosphere 2 facility near Tucson, Arizona. The three synthetic, controlled hillslopes at LEO were constructed with the aim of improving our predictive understanding of the coupled physical, chemical, biological, and geological processes at Earth's surface in changing climates [3]. Each hillslope is 30 m long and 11.15 m wide and has an average slope of 10°. The 1 m deep soil consists of basaltic tephra, ground to homogeneous loamy sand texture. For the first years of LEO operation, vegetation is not present and the research is focused on the characterization of the hydrological response of the hillslopes in terms of water transit times, generation of seepage and overland flow, internal dynamics of soil moisture, and evaporation. The second part of the experiment envisages the presence of plants growing on the hillslopes

* Corresponding author. Tel.: +39 049 827 1319; fax: +39 049 827 1499.
E-mail address: damiano.pasetto@unipd.it, pasetto@math.unipd.it, pase84@gmail.com (D. Pasetto).

http://dx.doi.org/10.1016/j.advwatres.2015.04.014
0309-1708/

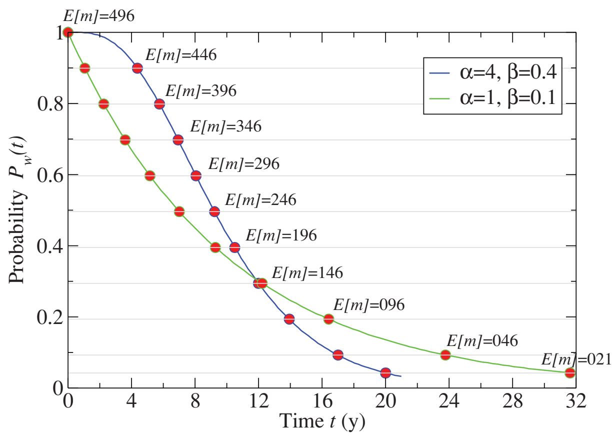

Fig. 1. The probability of a sensor being active at time t for two distributions of failure time. The two red circles along each horizontal line give, for each distribution, the time at which the number of active sensors is expected to have dropped to the indicated value of $E[m]$. For instance, the expected value of active sensors is 46 ($E[m] = P_w n = 0.093 \times 496$) after 17 years for the blue distribution and after 23.8 years for the green distribution. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and aims to monitor the oxygen and carbon cycles inside LEO, as well as the impact of vegetation on the spatial distribution of soil water content and on changes in the soil hydraulic properties [4,5].

To monitor these processes, each hillslope is equipped with a dense network of soil sensors (496 locations) that measures volumetric water content (496 sensors), soil water potential, and soil temperature. These local observations of the internal state of the soil are combined with measurements of the global system response, such as the total weight of the infrastructure (and thus the water storage), the rate of irrigation/evaporation, and the water outflow at the seepage face. Finally, geochemical analysis of irrigation water, soil water, and seepage outflow are available to monitor solute transport processes along the hillslopes.

As sensors fail, the number of active sensors, m, will decrease in time. For example, assuming that the time of failure of a sensor, t_f, follows a Gamma distribution with shape parameter α and rate parameter β, the probability that a sensor is working at time t is

$$P_w(t) = P(t_f > t) = 1 - \int_0^t g(\tau; \alpha, \beta)\, d\tau \qquad (1)$$

where $g(\cdot;\ \alpha,\ \beta)$ is the probability density function (pdf) of the Gamma distribution. With the further assumption that the times of failure of the sensors are independent and identically distributed random variables, the number of active sensors at time t has a binomial distribution with parameters $p = P_w(t)$ and $n = 496$, and the expected value of active sensors at time t is $E[m] = np$. Fig. 1 shows the probability of the lifetime of a sensor, $P_w(t)$, for two possible combinations of parameters α and β (the expected value of the failure time in this example is $E[t_f] = \alpha/\beta = 10$ years).

In this study we assess the impact of the number of active sensors on the observability of the LEO hillslopes. The physically-based hydrological model CATHY [6] is employed to numerically simulate the water dynamics on the LEO landscapes. CATHY couples a finite element solver of the Richards equation for subsurface flow developed by Paniconi and Putti [7] with a surface routing scheme developed by Orlandini and Rosso [8]. Surface flow occurs along a conceptual channel network derived from the digital elevation model (DEM) of the landscape [9], and the coupling between the surface and subsurface modules is resolved via a boundary condition-based partitioning of the atmospheric inputs into soil infiltration and land surface ponding. To account for heterogeneities in the LEO soil [10], we represent the saturated hydraulic conductivity as a three-dimensional random field with a lognormal probability distribution and an anisotropic exponential covariance function.

We use two different approaches to quantitatively assess the information associated with the network of active sensors of volumetric water content. In the first approach we are interested in knowing if LEO's sensor network allows us to accurately retrieve the spatial and temporal distribution of the water content in the entire landscape. To assess the accuracy of the retrieval, we compare the integral of the computed water content over the entire domain with the measured variation of water storage in the landscape. In the second approach we assess the sensor network's ability to allow retrieval of the saturated hydraulic conductivity of the soil, a critical parameter for numerical modeling of the future hydrological experiments at LEO. To account for parameter and measurement uncertainties, the ensemble Kalman filter (EnKF) [11–14] is used to compute the posterior probability distribution of the saturated hydraulic conductivity. EnKF performs a Gaussian approximation of sequential Bayesian inversion, thereby extending the Kalman filter to nonlinear models. The evolution in time of the state pdf is simulated using a Monte Carlo (MC) technique. The ensemble of model solutions is associated with random realizations of the unknown parameters. These MC realizations are then used in the update step to compute the covariance matrices required in the Kalman filter. Due to its straightforward implementation and its computational efficiency [15], EnKF is largely employed in engineering applications for measurement assimilation in real time. Moreover, since EnKF seeks a probability distribution of the parameters, this approach reduces the issues associated with non-uniqueness of the solution that typically occurs in inverse problems (e.g., [16]).

One of the major drawbacks of the EnKF technique is the so-called ensemble inbreeding (i.e., the strong reduction of the ensemble variance after few updates). For this reason, Drecourt et al. [17] and De Lannoy et al. [18] suggest that it is important to ensure that the ensemble spread is large enough at the assimilation time. Recent enhancements to the EnKF technique for estimation of two-dimensional stochastic parameters include introduction of a damping parameter [19] to reduce ensemble inbreeding, and covariance localization to clean the ensemble covariance matrices of spurious terms [20,21]. Sun et al. [22,23] combine EnKF with grid-based localization and Gaussian mixture model clustering techniques to estimate a multimodal parameter distribution. Panzeri et al. [24] couple EnKF with the ensemble moment equation of the transient groundwater flow equation to circumvent the MC simulation. Alzraiee et al. [25] compare centralized and decentralized fusion to invert the measurements generated with different pumping tests. Amongst applications of EnKF for estimating the spatial distribution of parameters in three-dimensional hydrological models, Chen and Zhang [26] showed that EnKF provides a satisfactory estimation of the three-dimensional hydraulic conductivity field assimilating measurements of pressure head in a synthetic example of saturated flow.

2. Problem representation

We represent the hillslope (the three LEO hillslopes are identical) as a three-dimensional domain Ω with the DEM depicted in Fig. 2 and a 1 m deep soil. The bottom of the hillslope, the two side boundaries (the edges along the y axis in Fig. 2), and the upslope boundary are impermeable, while the downslope boundary (at $y = 0$ m, hereafter denoted by Γ) is the outflow face, and is modeled as a seepage face boundary condition. Let $\theta(t, \vec{x})$ be the soil water content [−] at a time t [T] at a point $\vec{x} = (x, y, z) \in \Omega$. Given a spatial distribution of θ at a reference time $t_0 = 0$ (initial condition), rainfall and evaporation boundary conditions are imposed at the surface, and θ responds according to this forcing term and to the soil hydraulic properties.

The dense sensor network allows the system to be monitored every 15 min from the reference time t_0 (times t_i with $t_i - t_{i-1} = 15$