A reduced order modeling approach to represent subgrid-scale hydrological dynamics for land-surface simulations: Application in a polygonal tundra landscape

G. S. H. Pau, G. Bisht, W. J. Riley

Earth Science Division, Lawrence Berkeley National Laboratory,

1 Cyclotron Road, Berkeley, California 94720, USA

Correspondence to: G. S. H. Pau (gpau@lbl.gov)

Abstract

Existing land surface models (LSMs) describe physical and biological processes that occur over a wide range of spatial and temporal scales. For example, biogeochemical and hydrological processes responsible for carbon (CO_2, CH_4) exchanges with the atmosphere range from molecular scale (pore-scale O_2 consumption) to tens of kilometer scale (vegetation distribution, river networks). Additionally, many processes within LSMs are nonlinearly coupled (e.g., methane production and soil moisture dynamics), and therefore simple linear upscaling techniques can result in large prediction error. In this paper we applied a reduced-order modeling (ROM) technique known as "Proper Orthogonal Decomposition mapping method" that reconstructs temporally-resolved fine-resolution solutions based on coarse-resolution solutions. We developed four different methods and applied them to four study sites in a polygonal tundra landscape near Barrow, Alaska. Coupled surface-subsurface isothermal simulations were performed for summer months (June-September) at fine (0.25 m) and coarse (8 m) horizontal resolutions. We used simulation results from three summer seasons (1998-2000) to build ROMs of the 4D soil moisture field for the study sites individually (single-site) and aggregated (multi-site). The results indicate that the ROM produced a significant computational speedup ($>10^3$) with very small relative approximation error (<0.1%) for two validation years not used in training the ROM. We also demonstrate that our approach: (1) efficiently corrects for coarse-resolution model bias and (2) can be used for polygonal tundra sites not included

in the training dataset with relatively good accuracy (< 1.7% relative error), thereby allowing for the possibility of applying these ROMs across a much larger landscape. By coupling the ROMs constructed at different scales together hierarchically, this method has the potential to efficiently increase the resolution of land models for coupled climate simulations to spatial scales consistent with mechanistic physical process representation.

1 Introduction

The terrestrial hydrological cycle strongly impacts, and is impacted by, atmospheric processes. Further, a primary control on terrestrial biogeochemical (BGC) dynamics and greenhouse gas (GHG) emissions from soils (e.g., CO_2, CH_4, N_2O) across spatial scales is exerted by the system's hydrological state (Schuur et al., 2008). Soil moisture also impacts soil temperature, which is another important controller of GHG emissions (Torn and Chapin, 1993). Since climate change is predicted to change the amount and temporal distribution of precipitation globally, there is a critical need for models to not only accurately capture subgrid heterogeneity of terrestrial hydrological processes, but also the impacts of subgrid hydrological heterogeneity on BGC fluxes.

Terrestrial hydrological states are important for climate prediction across a wide range of spatial scales, from soil pores to continental. The critical spatial scale relevant to soil moisture state and subsurface and surface fluxes may be as small as ~100 m (Wood et al., 2011), although there is vibrant disagreement about the relative increase in predictability when trying to explicitly simulate at such high resolutions with limited observational data to constrain parameter values (Beven and Cloke, 2012). However, the importance of representing fine-resolution spatial structure in hydrological states and fluxes has been demonstrated for surface evapotranspiration budgets (Vivoni et al., 2007; Wood, 1997), runoff and streamflow (Arrigo and Salvucci, 2005; Barrios and Francés, 2012; Vivoni et al., 2007), and atmospheric feedbacks (Nykanen and Foufoula-Georgiou, 2001). It remains unclear what the critical spatial scale is for biogeochemical dynamics, but it has been shown that 'hot spot' formation is important for wetland biogeochemistry at scales ~O(10 cm) (Frei et al., 2012) and for nitrogen cycle variations at ~O(m) (McClain et al., 2003). In contrast, the current suite of land surface models applicable at watershed (e.g., PAWS (Riley and Shen, 2014; Shen, 2009)), regional (Maxwell et al., 2012), or climate (Koven et al., 2013; Tang et al., 2013) scales typically represent hydrological or biogeochemical cycles at ~O(100 m - km) scales.

The methods to represent spatial heterogeneity in hydrological and biogeochemical dynamics differ between watershed and regional or climate-scale models. While many current watershed-scale models explicitly represent lateral inter-connectivity for subsurface and surface fluxes, regional and climate-scale models currently rely on a non-spatially explicit tiling approach. For example, CLM4.5 (Koven et al., 2013; Lawrence et al., 2012; Tang et al., 2013), the land model integrated in the Community Earth System Model (Hurrell et al., 2013), represents land-surface grid cells with the same horizontal extent as the atmospheric grid cells (which can range from ~1°×1° for climate change simulations to ~0.25°×0.25° for relatively short simulations (Bacmeister et al., 2013; Wehner et al., 2014)). These grid cells are disaggregated into a subgrid hierarchy of non-spatially explicit land units (e.g., vegetated, lakes, glacier, urban), columns (with variability in hydrological, snow, and crop management), and plant functional types (accounting for variations in broad categories of plants and bare ground). Therefore, we contend that representing the much smaller spatial scales now recognized to control hydrological and biogeochemical dynamics in regional and global-scale models will require a reformulation of the overall design of these models.

One potential approach to represent spatial heterogeneity in soil moisture fields at resolutions finer than represented in a particular modeling framework is to relate the statistical properties of the soil moisture field with the spatial scale. Hu et al (1997) showed that the variance (σ_θ^2) of the soil moisture (θ) field at different spatial averaging areas (*A*) can be related to the ratio of those areas raised to a scaling exponent (γ). They also showed that γ is related to the spatial correlation structure of the soil moisture field and that it decreases as soils dry. Observational studies have described a power law decay of variance as a function of the observation scale (Rodriguez-Iturbe et al., 1995; Wood, 1998), and several investigators have demonstrated that the relationship between σ_θ^2 and spatial scale is not 'simple' (i.e., not log-log linear across all spatial scales; e.g., Das and Mohanty (2008); Famiglietti et al. (1999); Joshi and Mohanty (2010); Mascaro et al. (2010, 2011); Nykanen and Foufoula-Georgiou (2001)).

A second potential approach to account for spatial heterogeneity in soil moisture states is to relate its higher-order moments to the mean, and then apply these relationships within a model that predicts the transient coarse-resolution mean. In many observationally-based studies, an upward convex relationship between the mean and variance has been reported (e.g., Brocca et al. (2010); Brocca et al. (2012); Choi and Jacobs (2011); Famiglietti et al. (2008); Lawrence and Hornberger (2007); Li and Rodell (2013); Pan and Peters-Lidard (2008); Rosenbaum et al. (2012); Tague et al. (2010); Teuling et al. (2007); Teuling and Troch (2005)). Theoretical analyses have also indicated that an upward convex relationship is consistent with current understanding of soil moisture dynamics (e.g., Vereecken et al. (2007)). However, as discussed in Brocca et al. (2007), the relationships between soil moisture mean and statistical moments have been reported to depend on many factors, including lateral redistribution, radiation, soil characteristics, vegetation characteristics, elevation above the drainage channel, downslope gradient, bedrock topography, and specific upslope area. These large number of observed controllers and the lack of an accepted set of dominant factors argue that substantial work remains before this type of information can be integrated with land models to represent subgrid spatial heterogeneity.

Modeling studies have also been performed to investigate spatial scaling properties of moisture and how these properties relate to ecosystem properties. For example, Ivanov et al. (2010) studied spatial heterogeneity in moisture on an idealized small hill slope, and found hysteretic patterns during the wetting-drying cycle and that the system response depends on precipitation magnitude. Riley and Shen (2014) used a distributed modeling framework to analyze relationships between mean and higher-order moments of soil moisture and ecosystem properties in a watershed in Michigan. They concluded that the strongest relationship between the observed declines in variance with increases in mean moisture (past a peak in this relationship) was with the gradient convolved with mean evapotranspiration. Other studies have focused on upscaling fine-resolution model parameters to effective coarser-resolution parameters. For example, Jana and Mohanty (2012) showed that power-law scaling of hydraulic parameters was able to capture subgrid topographic effects for four different hill slope configurations.

Theoretical work to explicitly include spatial heterogeneity in the hydrological governing equations has also been applied to this problem. Albertson and Montaldo (2003) and Montaldo and Albertson (2003) developed a relationship for the time rate of change of soil moisture variance based on the mean moisture and spatial covariances between soil moisture, infiltration, drainage, and ET. Teuling and Troch (2005) applied a similar approach to study the impacts of vegetation, soil properties, and topography on the controls of soil moisture variance. Kumar (2004) applied a Reynolds averaging approach, and ignoring second and higher order terms, derived a relationship for the time rate of change of the mean moisture field that depends on the moisture variance. Choi et al. (2007) applied the model to a ~25,000 km^2 Appalachian Mountain region for summer months of one year and found that subgrid variability significantly affected the prediction of mean soil moisture.

The approaches described above to capture fine-resolution spatial heterogeneity within a coarse-resolution modeling framework have some limitations. First, the soil moisture probability density function is often very non-normal (Ryu and Famiglietti, 2005), making the sole use of variance as a descriptor of moisture heterogeneity insufficient. A similar problem arises with the Reynolds averaging approach that does not include higher-order terms. This approach also requires a method to ‘close’ the solution (i.e., relate the higher-order terms to the mean moisture), and there is no generally accepted method to perform this closure. Perhaps the largest constraint of these approaches in the context of climate change and atmospheric interactions is that they cannot account for the temporal memory in the system that impacts biogeochemical transformations. In particular, the biogeochemical dynamics at a particular point in time depend on the state and dynamics that occurred in the past, and just knowing the statistical distribution of moisture at a particular time may not maintain the continuity required for accurate prediction. Therefore, for applications related to regional to global-scale interactions with the atmosphere, a method is required that allows for (1) computationally tractable simulations (i.e., relatively coarser resolution); (2) spatially explicit prediction of the temporal evolution of soil moisture at relatively finer resolutions; and (3) integration of the relatively finer resolution soil moisture predictions with representations of the relevant biogeochemical dynamics.

To that end, we describe a generally applicable reduced order modeling technique to reconstruct a fine-resolution heterogeneous 4D soil moisture solution from a coarse-resolution simulation, thereby resulting in significant computational savings. In this study, we built ROMs based on the Proper Orthogonal Decomposition Mapping Method (Robinson et al., 2012), which first involved training the ROMs using fine- and coarse-resolution simulations over multiple years. Hydrologic simulations of coupled surface and subsurface processes for an Alaska polygonal tundra system were performed using the PFLOTRAN model (Bisht and Riley, 2014; Hammond et al., 2012). Simulations were performed for four study sites in Alaska with distinct polygonal surface characteristics and individual ROMs were built for each site. The resulting ROMs were then applied over periods outside of the ROM training period.

In the Methods section we describe the polygonal tundra site used for our simulations, the PFLOTRAN hydrological simulations configuration, and the methods used to develop and evaluate the ROMs. In the Results and Discussion section, these methods are used under different scenarios to develop ROMs for the polygonal tundra site that increase in generality in the following order: single-site ROMs (limited to a single site), multisite ROMs (limited to sites included in the training data) and site-independent ROMs (applicable even for sites not included in the training data). For each of the above scenarios, different ROMs can be developed using methods that we propose in the Methods section; the applicability of a method to a given scenario is discussed in the Methods section. We then compare the accuracy of the different ROMs and end with a discussion of limitations of the approach, possible improvements, and methods to incorporate the proposed ROM approach within a global-scale hydrological and biogeochemical model.

2 Methods

2.1 Site Description and Hydrologic Simulation Setup

In this study, we developed ROMs for hydrological simulations performed at four sites in the Barrow Environmental Observatory (BEO) in Barrow, Alaska (71.3° N, 156.5° W). The BEO lies within the Alaskan Arctic Costal Plain, which is a relatively flat region, characterized by thaw lakes and drained basins (Hinkel et al., 2003; Sellmann et al., 1975) and polygonal ground features (Hinkel et al., 2001; Hubbard et al., 2013). The Department of Energy (DOE) Next-Generation Ecosystem Experiments (NGEE-Arctic) project has established four intensely monitored sites (A, B, C and D, shown in Figure 1.) within the BEO in 2012 to study impact of climate change in high-latitude regions. The four NGEE-Arctic study sites have distinct micro-topographic features, which include low-centered (A), high-centered (B), and transitional polygons (C, and D). The mean annual air temperature for our study sites is approximately -13°C (Walker et al., 2005) and the mean annual precipitation is 106 mm with the majority of precipitation falling during the summer season (Wu et al., 2013). The study site is underlain with continuous permafrost and the seasonally active layer depth ranges between 30-90 cm (Hinkel et al., 2003).

We applied a version of the three-dimensional subsurface reactive transport simulator PFLOTRAN, which was modified to include surface water flows, for simulating surface-subsurface hydrologic processes at the four NGEE-Arctic study sites. The subsurface flows in PFLOTRAN are solved with a finite volume and an implicit time integration scheme, and are sequentially coupled to a finite volume based surface flow solution that is solved explicitly in time. Simulations at the four study sites were conducted using meshes at horizontal resolutions of 0.25 m, 0.5 m, 1.0 m, 2.0 m, 4.0 m, and 8.0 m. A constant vertical resolution of 5 cm with a total depth of 50 cm was used for all simulations. The simulations were carried out for four summer months (July-Sept) of each year between 1998-2006. Evapotranspiration and effective precipitation boundary conditions for the PFLOTRAN simulations were obtained from offline simulations of the Community Land Model (CLM4.5; (Oleson, 2013)). Vertical heterogeneity in soil properties was prescribed using data from Hinzman et al. (1991). A static active layer depth of 50 cm, corresponding approximately to the maximum seasonal value, was assumed for all simulations. Details of model setup are provided in Bisht and Riley (2014). In the current study, the ROM was trained on three years of data (1998-2000) and ROM predictions for 2002 and 2006 were compared against fine-resolution simulations.

2.2 Development of the Reduced Order Modeling Approach

The multifidelity ROM approach used in this study is based on the gappy Proper Orthogonal Decomposition (POD) mapping approach (Robinson et al., 2012). Let $\mathbf{p}$ be a set of parameters that defines a particular solution or observation. The set of parameters could include system parameters (e.g., vegetation distribution, soil types, and topography), climate forcings, time, and other quantities that have an influence on the system response. In this paper, the parameters that vary in the simulations that we have performed for each site are time (days for summer seasons in a year) and the climate forcings (precipitation and evapotranspiration rates) prescribed at that particular time. Then, given a sample set $S_N = \{\mathbf{q}_1, \ldots, \mathbf{q}_N\}$, where $\mathbf{q}_i$ is a set of parameters $\mathbf{p}$ and N is the number of samples, we can compute the corresponding solution $\{\mathbf{f}(\mathbf{q}_1), \ldots, \mathbf{f}(\mathbf{q}_N)\}$. In this paper, $\mathbf{f}$ corresponded to a simulated fine-resolution three-dimensional soil moisture field, but in general, $\mathbf{f}$ can be any spatial quantity of interest (e.g., soil temperature or GHG emission).

2.2.1 POD method

The POD approximation of $\mathbf{f}$, $\mathbf{f}^{\mathrm{POD}}$, is given by

$$\mathbf{f}(\mathbf{p}) \approx \mathbf{f}^{\mathrm{POD}}(\mathbf{p}) = \mathbf{f}^{\mathrm{ref}} + \sum_{i=1}^{M} \alpha_i(\mathbf{p}) \zeta_i^{\mathrm{POD}}, \tag{1}$$

where $M \le N \ll \mathcal{N}$, $\mathcal{N}$ is the degree of freedom of $\mathbf{f}$, $\mathbf{f}^{\mathrm{ref}}$ is the reference basis (here, $\mathbf{f}^{\mathrm{ref}} = \overline{\mathbf{f}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{f}(\mathbf{q}_i)$), ζ_i^{POD} are the POD bases and M is the number of POD bases. The POD bases are determined through a singular value decomposition (SVD) of the data matrix given by $\mathbf{W}^{\mathrm{POD}} = \left[\mathbf{f}(\mathbf{q}_1) - \overline{\mathbf{f}}, \ldots, \mathbf{f}(\mathbf{q}_N) - \overline{\mathbf{f}}\right]$:

$$\mathbf{W}^{\mathrm{POD}} = \mathbf{UDV}^T \tag{2}$$

where $\mathbf{U} \in \mathbb{R}^{\mathcal{N}\times\mathcal{N}}$ are the left eigenvectors, $\mathbf{V} \in \mathbb{R}^{N\times N}$ are the right eigenvectors, and $\mathbf{D} = \mathrm{diag}(\lambda_1,\ldots,\lambda_N) \in \mathbb{R}^{\mathcal{N}\times N}$, with $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N > 0$. The POD bases $\zeta_i^{\mathrm{POD}}, 1 \leq i \leq N$ are thus given by $\mathbf{W}^{\mathrm{POD}}\mathbf{V}$ and λ_i are the associated eigenvalues with each POD basis. The POD method is similar to the principal component analysis (Jolliffe, 2002) and the Karhunen Loeve decomposition (Moore, 1981). We computed the POD bases based on the kernel eigenvalue approach (Everson and Sirovich, 1995).

The number of POD bases (denoted by M) used to reconstruct the approximate solution to a certain level of error (ε^{λ}) can be determined by finding M that satisfies

$$e_M^{\lambda} = 1 - \sum_{i=1}^{M} \lambda_i / \lambda_T \leq \varepsilon^{\lambda} \tag{3}$$

where $\lambda_T = \sum_{i=1}^{N} \lambda_i$. As mentioned in Wilkinson (2011), the dimensional reduction afforded by the POD method depends on the extent to which the components of $\mathbf{f}$ are correlated. We note that equation (1) only states how $\mathbf{f}$ is represented in a linear space spanned by the POD bases, but there are multiple approaches of determining $\boldsymbol{\alpha}(\mathbf{p}) = \{\alpha_1(\mathbf{p}),\ldots,\alpha_M(\mathbf{p})\}$ for a given $\mathbf{p}$. One optimal solution of $\boldsymbol{\alpha}$ that minimizes the least squares error between $\mathbf{f}(\mathbf{p})$ and $\mathbf{f}^{\mathrm{POD}}(\mathbf{p})$, denoted by $\boldsymbol{\alpha}^{\mathrm{POD}}(\mathbf{p})$, is given by:

$$\alpha_i^{\mathrm{POD}}(\mathbf{p}) = \zeta_i^{\mathrm{POD},T}(\mathbf{f}(\mathbf{p}) - \overline{\mathbf{f}}), \quad i = 1,\ldots,M\ . \tag{4}$$

However, $\boldsymbol{\alpha}^{\mathrm{POD}}(\mathbf{p})$ determined using equation (4) does not lead to any computational savings since $\mathbf{f}(\mathbf{p})$ is the quantity we would like to approximate. Determination of $\mathbf{f}(\mathbf{p})$ can be avoided by using the POD projection method (Willcox and Peraire, 2002), which discretizes the governing equations using the linear space spanned by ζ_i^{POD} and solves the resulting algebraic equations for $\boldsymbol{\alpha}^{\mathrm{POD}}(\mathbf{p})$. However, the POD projection method requires extensive modification of the existing code of the simulator, and is thus not suitable for existing LSMs. To demonstrate the limit of accuracy of POD-related methods presented in subsequent subsections (sections 2.2.2-2.2.5), we determine

$\boldsymbol{\alpha}^{\text{POD}}(\mathbf{p})$ based on equation (4) by evaluating $\mathbf{f}(\mathbf{p})$ explicitly and present the results in Results and Discussion section.

In subsequent sections, we describe 4 different methods of developing a ROM that reconstruct the fine resolution solution based on the coarse resolution solution. Each of the methods is a modification of the basic POD method, but uses a different reference basis, data matrix, or method to compute $\boldsymbol{\alpha}(\mathbf{p})$. The differences among the various methods for developing a ROM are summarized in Table 1.

2.2.2 POD mean method (POD-mean)

To overcome the difficulties associated with calculating $\boldsymbol{\alpha}^{\text{POD}}(\mathbf{p})$, we propose a POD-mean method (POD-mean). We first determine $\alpha^{\text{POD}}(\mathbf{q}), \forall \mathbf{q} \in S_N$ using equation (4) ; this step requires negligible computational overhead since construction of ROM based on POD method already requires the determination of $\mathbf{f}(\mathbf{q}), \forall \mathbf{q} \in S_N$. We then construct a polynomial fit between $\boldsymbol{\alpha}^{\text{POD}}(\mathbf{q})$ and the mean of $\mathbf{f}(\mathbf{q})$ (i.e., fine-resolution mean soil moisture, $\mu_f(\mathbf{q})$) which we denote as $\boldsymbol{\alpha}^{\text{fit}}(\mu_f)$. Then, for any given $\mathbf{p}$, we approximate $\mathbf{f}$ by

$$\mathbf{f}_{\Delta x_g}^{\text{POD-mean}}(\mathbf{p}) = \overline{\mathbf{f}} + \sum_{i=1}^{N} \alpha_i^{\text{fit}}(\mu_g(\mathbf{p}))\zeta_i^{\text{POD}} \tag{5}$$

where $\mu_g(\mathbf{p})$ is the mean of $\mathbf{g}(\mathbf{p})$, a coarse-resolution solution simulated at resolution $\Delta x_g > \Delta x_f$. This particular approach works well if: (1) the relationships between $\alpha_i^{\text{fit}}(\mu_f)$ and μ_f exist; and (2) μ_g is a good approximation of μ_f. For the Artic Tundra study sites, we will show that these conditions hold true for $i = 1$, and $\mathbf{f}_{\Delta x_g}^{\text{POD-mean}}$ is a good approximation of $\mathbf{f}$.

2.2.3 POD mapping method (POD-MM)

In the POD-mean method, we only used the mean of the coarse-resolution solution, $\mathbf{g}(\mathbf{p})$, to reconstruct the fine-resolution solution. The POD mapping method

(POD-MM) attempts to use all information in $\mathbf{g}(\mathbf{p})$ to efficiently and accurately reconstruct the fine-resolution solution. The POD-MM method is a modification of the gappy POD (Everson and Sirovich, 1995). For the same sample set S_N, we determine $\{\mathbf{g}(\mathbf{q}_1),\ldots,\mathbf{g}(\mathbf{q}_N)\}$, where $\mathbf{q}_i \in S_N$. As in Robinson et al. (2012), the multifidelity POD bases, $\zeta_i^{\text{POD-MM}}$, are then determined through a SVD of the data matrix $\mathbf{W}^{\text{POD-MM}}$:

$$\mathbf{W}^{\text{POD-MM}} = \begin{bmatrix} \mathbf{f}(\mathbf{q}_1)-\overline{\mathbf{f}} & \cdots & \mathbf{f}(\mathbf{q}_N)-\overline{\mathbf{f}} \\ \mathbf{g}(\mathbf{q}_1)-\overline{\mathbf{g}} & & \mathbf{g}(\mathbf{q}_N)-\overline{\mathbf{g}} \end{bmatrix}, \tag{6}$$

where $\overline{\mathbf{f}}$ is as defined before and $\overline{\mathbf{g}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{g}(\mathbf{q}_i)$. The POD bases $\zeta_i^{\text{POD-MM}}$ can be decomposed into

$$\zeta_i^{\text{POD-MM}} = \begin{bmatrix} \zeta_i^{\mathbf{f},\text{POD-MM}} \\ \zeta_i^{\mathbf{g},\text{POD-MM}} \end{bmatrix} \tag{7}$$

where $\zeta_i^{\mathbf{f},\text{POD-MM}}$ and $\zeta_i^{\mathbf{g},\text{POD-MM}}$ are components associated with the fine- and coarse-resolution models. Given a coarse-resolution solution $\mathbf{g}(\mathbf{p})$, we first determine

$$\boldsymbol{\alpha}^{\text{POD-MM}}(\mathbf{p}) = \arg\min_{\gamma} \| \mathbf{g}(\mathbf{p}) - \overline{\mathbf{g}} - \sum_{i=1}^{M}\gamma_i \zeta_i^{\mathbf{g},\text{POD-MM}} \|_2 \tag{8}$$

where $\|\cdot\|_2$ is the L_2 norm. We note that $\boldsymbol{\alpha}^{\text{POD-MM}}(\mathbf{p})$ is not simply given by equation (4) since $\zeta_i^{\mathbf{g},\text{POD-MM}}$ are not mutually orthogonal. The approximate solution, $\mathbf{f}_{\Delta x_g}^{\text{POD-MM}}(\mathbf{p})$, is then given by $\mathbf{f}_{\Delta x_g}^{\text{POD-MM}}(\mathbf{p}) = \overline{\mathbf{f}} + \sum_{i=1}^{M}\alpha_i^{\text{POD-MM}}(\mathbf{p})\zeta_i^{\mathbf{f},\text{POD-MM}}$, where Δx_g is the resolution at which $\mathbf{g}(\mathbf{p})$ is computed.

2.2.4 Second alternative formulation of the POD mapping method (POD-MM2)

We also introduce an alternative formulation of the POD-MM method (POD-MM2) to determine whether the number of POD bases required could be reduced for a fixed approximation error threshold. Instead of applying equation (6), we perform a SVD of the data matrix $\mathbf{W}^{\text{POD-MM2}}$:

$$\mathbf{W}^{\text{POD-MM2}} = \begin{bmatrix} \mathbf{h}(\mathbf{q}_1) - \overline{\mathbf{h}} & \cdots & \mathbf{h}(\mathbf{q}_N) - \overline{\mathbf{h}} \\ \mathbf{g}(\mathbf{q}_1) - \overline{\mathbf{g}} & & \mathbf{g}(\mathbf{q}_N) - \overline{\mathbf{g}} \end{bmatrix}, \tag{9}$$

where

$$\mathbf{h}(\mathbf{p}) = \mathbf{f}(\mathbf{p}) - \tilde{\mathbf{g}}(\mathbf{p}), \tag{10}$$

and $\tilde{\mathbf{g}}$ is the solution obtained from a piecewise constant mapping of $\mathbf{g}$ from the coarse-resolution grid of $\mathbf{g}$ onto the fine-resolution grid of $\mathbf{f}$. By using the deviation of $\mathbf{f}$ from the mapped coarse-resolution solution $\tilde{\mathbf{g}}$, we remove the bias resulting from the mismatch between the mean of $\mathbf{f}$ and $\mathbf{g}$. We note that this alternative POD mapping formulation is possible since our coarse- and fine-resolution grids are nested (which will always be the case for the types of applications we are developing here). For non-nested grids, a linear mapping is expected to work as well, although we do not analyze that approach here. We denote the resulting POD bases vector as

$$\zeta_i^{\text{POD-MM2}} = \begin{bmatrix} \zeta_i^{\mathbf{h},\text{POD-MM2}} \\ \zeta_i^{\mathbf{g},\text{POD-MM2}} \end{bmatrix} \tag{11}$$

where $\zeta_i^{\mathbf{h},\text{POD-MM2}}$ are the components associated with $\mathbf{h}$. Given a solution $\mathbf{g}(\mathbf{p})$, the approximate $\mathbf{f}_{\Delta x_g}^{\text{POD-MM2}}(\mathbf{p})$ is then given by

$$\mathbf{f}_{\Delta x_g}^{\text{POD-MM2}}(\mathbf{p}) = \tilde{\mathbf{g}} + \overline{\mathbf{h}} + \sum_{i=1}^{M} \alpha_i^{\text{POD-MM2}}(\mathbf{p}) \zeta_i^{\mathbf{h},\text{POD-MM2}} \tag{12}$$

where $\alpha^{\text{POD-MM2}}$ is determined analogously to $\alpha^{\text{POD-MM}}$ based on equation (8) with $\zeta_i^{\mathbf{g},\text{POD-MM}}$ replaced by $\zeta_i^{\mathbf{g},\text{POD-MM2}}$.

2.2.5 Third alternative formulation of the POD mapping method (POD-MM3)

When a solution is spatially highly correlated with a spatially-varying parameter $\mathbf{w}$, such as the topography, we may use this information in our reconstruction of the fine-resolution solution. This third alternative formulation of the POD-mapping method approximates $\mathbf{f}$ by

$$\mathbf{f}_{\Delta x_g}^{\text{POD-MM3}}(\mathbf{p}) = \hat{\mathbf{f}} + \sum_{i=1}^{M} \alpha_i^{\text{POD-MM3}}(\mathbf{p}) \zeta_i^{\mathbf{f},\text{POD-MM3}}, \tag{13}$$

where $\zeta_i^{\mathbf{f},\text{POD-MM3}}$ is the fine-resolution component of the POD basis $\zeta_i^{\text{POD-MM3}}$ constructed from

$$\mathbf{W}^{\text{POD-MM3}} = \begin{bmatrix} \mathbf{f}(\mathbf{q}_1) - \hat{\mathbf{f}} & \cdots & \mathbf{f}(\mathbf{q}_N) - \hat{\mathbf{f}} \\ \mathbf{g}(\mathbf{q}_1) - \hat{\mathbf{g}} & & \mathbf{g}(\mathbf{q}_N) - \hat{\mathbf{g}} \end{bmatrix}, \tag{14}$$

and $\boldsymbol{\alpha}^{\text{POD-MM3}}$ is determined analogously to $\boldsymbol{\alpha}^{\text{POD-MM}}$ based on equation (8) with $\overline{\mathbf{g}}$ and $\zeta_i^{\mathbf{f},\text{POD-MM}}$ replaced by $\hat{\mathbf{g}}$ and $\zeta_i^{\mathbf{f},\text{POD-MM3}}$ respectively. In above, the correlation between $\mathbf{w}$ and $\mathbf{f}$ ($\mathbf{g}$) is used to construct $\hat{\mathbf{f}}$ ($\hat{\mathbf{g}}$) based on the following:

$$\hat{\mathbf{f}} = \mathbf{w}_{\Delta x_f} \frac{\mu_f}{\mu_{\mathbf{w}_{\Delta x_f}}}, \quad \hat{\mathbf{g}} = \mathbf{w}_{\Delta x_g} \frac{\mu_g}{\mu_{\mathbf{w}_{\Delta x_g}}}, \tag{15}$$

where μ_f (μ_g) is $\mathbf{f}$ ($\mathbf{g}$) averaged over the domain and all the snapshots used in constructing the ROM, $\mathbf{w}_{\Delta x_f}$ ($\mathbf{w}_{\Delta x_g}$) is the model parameter evaluated at resolution Δx_f (Δx_g), and $\mu_{\mathbf{w}_{\Delta x_f}}$ ($\mu_{\mathbf{w}_{\Delta x_g}}$) is the mean of $\mathbf{w}_{\Delta x_f}$ ($\mathbf{w}_{\Delta x_g}$).

The POD-MM3 approach is developed to improve the performance of POD-MM method when one of the parameters is heterogeneous and spatially varying. This method is only applicable to site-independent ROM since the surface elevation is included as a parameter in the site-independent ROM but not in the single and multi-site ROMs.

2.2.6 Error definitions

We define the relative error of the POD method with respect to the true fine-resolution solution as:

$$e^{\text{POD}} = \frac{\| \mathbf{f}^{\text{POD}} - \mathbf{f} \|_2}{\| \mathbf{f} \|_2}. \tag{16}$$

This error measure gives the maximum theoretical accuracy achievable using POD-related methods. We also define $\bar{e}^{\mathrm{POD}}$ as the mean of e^{POD} evaluated over a specified number of days.

For POD-X methods, where POD-X stands for POD-mean, POD-MM, POD-MM2, or POD-MM3, the error measures can be constructed for each Δx_g, and are defined as:

$$e_{\Delta x_g}^{\text{POD-X}} = \frac{\| \mathbf{f}_{\Delta x_g}^{\text{POD-X}} - \mathbf{f} \|_2}{\| \mathbf{f} \|_2}. \tag{17}$$

Similarly, we define $\bar{e}_{\Delta x_g}^{\text{POD-X}}$ as the mean of $e_{\Delta x_g}^{\text{POD-X}}$ evaluated over a specified number of days.

3 Results and Discussion

As described in the Methods section, we developed the ROM models for the four NGEE-Arctic Barrow study sites chosen for detailed characterization. The four sites differ in their topographic characteristics and therefore each site has a different dynamic soil moisture response to the same meteorological forcings. In addition, since the parameters varied in this study are time and the magnitude of the forcing terms, historical data (prior-year simulations) can be used to construct the ROM. The resulting ROM is subsequently used to predict future responses. For more general cases involving system parameters, statistical or adaptive sampling techniques are needed to generate S_N (Pau et al., 2013a; Pau et al., 2013b). For all study years, domain average soil moisture decreased during the first half of the simulation time period due to losses associated with evapotranspiration, while soil moisture increased in the latter half due to increased rainfall. Sites A and B had the lowest mean soil moisture, followed by site C, and then the wettest site, D.

3.1 Single-site ROMs

3.1.1 Application of POD method

We first constructed four separate ROMs, one for each site, using the POD method and the finest resolution (Δx_f =0.25 m) soil moisture predictions from 1998-2000. Given the soil moisture data for 2002 and 2006, $\mathbf{f}^{\mathrm{POD}}$ is determined based on equations (1) and (4). The mean relative error of the POD method, $\overline{e}^{\mathrm{POD}}$, over 120 days in year 2002 and 2006 decreases with increasing M (Figure 2). The M values at which we evaluate $\overline{e}^{\mathrm{POD}}$ correspond to decreasing $\varepsilon^{\lambda} = 10^{-1}, 10^{-2}, \ldots, 10^{-8}$ in equation (3). There is no significant difference between the error budgets as a function of M for 2002 and 2006. The number of POD bases for a given $\overline{e}^{\mathrm{POD}}$ increases with sites in the following order: A, B, C, and D.

The above observation cannot be deduced based solely on the probability distribution functions (PDFs) of the DEM of the sites (Figure 3) even though DEM is the only quantity that is different between the models for the four study sites. For example, site D requires the most M although its DEM has the smallest standard deviation. The larger number of POD bases required by site D can be attributed to particularly non-smooth soil moisture PDFs under relatively saturated conditions. The POD method is more efficient when the approximated solution has more smoothness in the parameter space (i.e., a solution at a particular point varies smoothly with the parameters). Site D is relatively flat and at the end of the summer season it tends to get completely saturated, thereby resulting in a discontinuity in the parameter space and requiring larger M.

3.1.2 Application of POD-mean method

To determine whether we can use the POD-mean method, we first examine the relationship between $\alpha_i^{\mathrm{POD}}(\mathbf{q})$ and $\mu_f(\mathbf{q})$ for all $\mathbf{q} \in S_N$. For all four sites, we found α_1^{POD} to be linearly correlated to μ_f (Figure 4). For $i > 1$, a simple correlation between α_i^{POD} and μ_f cannot be found. We can thus approximate α_1^{POD} by $\alpha_1^{\mathrm{fit}}(\mu_f) = a_1(\mu_f) + a_2$, where a_1 and a_2 are determined from a least-square fit of $\alpha_1^{\mathrm{POD}}(\mathbf{q})$ and $\mu_f(\mathbf{q})$. In addition, μ_f is well approximated by μ_g, allowing us to use the POD-mean method. For $\Delta x_g = 8$ m, the maximum and mean of $e_{\Delta x_g}^{\text{POD-mean}}$ are, respectively, 0.013 and 0.0016 at site A, and 0.016 and 0.005 at site D. A mean error that is <1% can thus be achieved using POD-mean method.

3.1.3 Application of POD-MM method

As with the previous analysis, ROMs based on POD-MM were constructed using only soil moisture data from 1998-2000 and daily prediction of soil moisture at 0.25 m were made for 2002 and 2006 using only the ROMs and coarse-resolution solutions. We only present our analyses for site A and D for brevity but the results are consistent with those from the remaining sites; Figure 2 shows that site B should yield similar results to site A and site C to site D.

The mean error for the POD mapping method ($\overline{e}_{\Delta x_g}^{\text{POD-MM}}$) decreases monotonically with M for all $\Delta x_g = \Delta x > 0.25$ m up to $M = M_{\text{optimal}}$, after which $\overline{e}_{\Delta x_g}^{\text{POD-MM}}$ starts to increase and fluctuate (Figure 6). This behavior is consistent with results from Everson and Sirovich (1995) in their development of ROMs for face reconstruction. Although larger M improves the least square fit, it leads to overfitting and increases the uncertainty in the computed $\boldsymbol{\alpha}^{\text{POD-MM}}$. This increased uncertainty in $\boldsymbol{\alpha}^{\text{POD-MM}}$ introduces significant random noise into the reconstructed fine-resolution solution, leading to fluctuating $\overline{e}_{\Delta x_g}^{\text{POD-MM}}$. Compared to $\mathbf{f}_{\Delta x_g}^{\text{POD-mean}}$, the accuracy of $\mathbf{f}_{\Delta x_g}^{\text{POD-MM}}$ can be systematically improved by utilizing more POD bases in the approximation.

For a given $M \le M_{\text{optimal}}$, $\overline{e}_{\Delta x_g}^{\text{POD-MM}}$ decreases with Δx_g, which implies that increasing the number of bases leads to a more accurate reconstructed fine-resolution solution. For site D, M_{optimal} also increases with decreasing Δx_g since larger information content allows more $\boldsymbol{\alpha}^{\text{POD-MM}}$ for more POD bases to be determined accurately. For year 2006, $M_{\text{optimal}} = 33$ for Δx_g = 0.5 m and $M_{\text{optimal}} = 28$ for Δx_g = 8 m. For site A however, $M_{\text{optimal}} = 10$ for all Δx_g. As such, when the underlying dynamics that we want to capture are mild (as indicated by the small M_{optimal}), the dependence of M_{optimal} on Δx_g is weaker. For the results shown, $\overline{e}_{\Delta x_g}^{\text{POD-MM}}$ is less than 6×10^{-5} when $\mathbf{f}_{\Delta x_g}^{\text{POD-MM}}$ is evaluated at

M_{optimal} for site D. In addition, the mean of ($\mathbf{f}_{\Delta x_g}^{\text{POD-MM}} - \mathbf{f}$) is 1.15×10^{-5} and 6.72×10^{-6} for sites A and D, respectively, indicating that there is only a negligible bias in the ROM solution.

The above approach requires knowledge of the true fine-resolution solutions to determine M_{optimal}. Alternatively, we can determine M_{optimal} by examining the amount of variance represented by the first M POD bases. For $M < M_{\text{optimal}}$, there is a linear relationship between log($\overline{e}_{\Delta x_g}^{\text{POD-MM}}$) and log($e_M^{\lambda}$); the slope of the line is dependent on Δx_g (Figure 7). In addition, $e_M^{\lambda} < 10^{-6}$ appears to be a reasonable criterion for determining M_{optimal}. Choosing this value leads to $M = 10$ and $M = 25$ at sites A and D, respectively, for Δx_g = 8.0 m. These values are very close to the M_{optimal} values identified based on Figure 6.

We next analyze the daily variation of $e_{\Delta x_g}^{\text{POD-MM}}$ for years 2002 and 2006 (Figure 8). The error $e_{\Delta x_g}^{\text{POD-MM}}$ is typically larger in the wetter periods although its maximum is below 0.01. We further examined the relative point-wise error, given by

$$\varepsilon_{\Delta x_g}^{\text{POD-MM}} = \left| \frac{\mathbf{f}_{\Delta x_g}^{\text{POD-MM}} - \mathbf{f}}{\mathbf{f}} \right| \tag{18}$$

for the days with the largest $e_{\Delta x_g}^{\text{POD-MM}}$; they corresponded to day 1 of 2002 for site A (Figure 9) and day 106 of 2002 for site D (Figure 10). For site A, the maximum $\varepsilon_{\Delta x_g}^{\text{POD-MM}}$ is 2.77×10^{-3} and the locations of large errors are not discernable from Figure 9, indicating that large errors are only localized to small region of the domain, resulting in small average errors, $e_{\Delta x_g}^{\text{POD-MM}}$. For site D, the maximum $\varepsilon_{\Delta x_g}^{\text{POD-MM}}$ is 1.17×10^{-3}, but a larger region of the domain has a higher $\varepsilon_{\Delta x_g}^{\text{POD-MM}}$ compared to site A, resulting in a higher $e_{\Delta x_g}^{\text{POD-MM}}$ (Figure 10). In addition, the saturated portion of the solution has small fluctuating errors, as evident from $\varepsilon_{\Delta x_g}^{\text{POD-MM}}$ of the bottom layer (Figure 10). Future work will examine how we can remove these fluctuations by simultaneously taking into account both water content and saturation.

3.1.4 Application of POD-MM2 method

With the POD-MM2 method, the resulting error, $\overline{e}_{\Delta x_g}^{\text{POD-MM2}}$, is smaller than $\overline{e}_{\Delta x_g}^{\text{POD-MM}}$ for small M (Figure 11). For example, for $M = 1$ and $\Delta x_g = 0.5\,\text{m}$, $\overline{e}_{\Delta x_g}^{\text{POD-MM2}}$ is an order of magnitude smaller than $\overline{e}_{\Delta x_g}^{\text{POD-MM}}$. However, the convergence behavior of $\overline{e}_{\Delta x_g}^{\text{POD-MM2}}$ with M is less well behaved as compared to the POD-MM. As a result, the minimum achievable value of $\overline{e}_{\Delta x_g}^{\text{POD-MM2}}$ is larger than the minimum achievable value of $\overline{e}_{\Delta x_g}^{\text{POD-MM}}$, especially for larger Δx_g. The POD-MM method is thus preferred since it allows the error to be reduced systematically by increasing M, especially when Δx_g is large.

3.2 Multi-site ROM

To construct a multi-site ROM, we used daily snapshots from all four sites for 1998 – 2000 to construct a single ROM. This is a first step towards developing a ROM that is applicable to the entire NGEE-Arctic study region. Based on the analysis performed using the POD method, we conclude that the POD-related methods can theoretically perform very well even when all four sites are considered in aggregate (Figure 12). However, the number of POD bases needed to achieve similar accuracy is greater than when separate ROMs are constructed for each site (compare Figure 2 and Figure 12).

With the POD-MM method, $\overline{e}_{\Delta x_g}^{\text{POD-MM}} \le 10^{-3}$ when only a relatively small number of POD bases are used (Figure 13). For $\Delta x_g = 8$ m, the error is minimum when $M = 30$. The magnitude of the error is only slightly larger than single-site ROMs. Although this approach is less efficient since M is generally larger than M for the single-site ROM, it is still significantly faster than performing simulations at the finest-resolution. A multi-site ROM is a good alternative to multiple single-site ROMs when the number of sites becomes large. In addition, if the sites have some similar features, a smaller number of snapshots is required per site, leading to lower computational cost needed to construct a single multi-site ROM compared to multiple single-site ROMs. POD-MM2 method is not used to develop multi-site ROM for reasons given in our analysis of single-site ROMs.

3.3 Site-independent ROM

Here, we include the spatially heterogeneous surface elevation, as described by the DEM, in the parameter space during the construction of the ROM. We trained the ROM using the soil moisture solutions at sites B, C, and D and evaluated the performance of the ROM for soil moisture prediction at site A. The resulting ROM is denoted as a site-independent ROM since it is applied on a site that was excluded from training dataset.

For the POD method, the error $\overline{e}^{\mathrm{POD}}$ for the site-independent ROM decreases with increasing number of bases but not as rapidly as $\overline{e}^{\mathrm{POD}}$ of single- or multi-site ROMs (Figure 14). For the POD mapping method, the error ($\overline{e}_{\Delta x_g}^{\mathrm{POD\text{-}MM}}$) also decreases slowly with M when compared to single- or multi-site ROMs (Figure 15(a)). For $\Delta x_g = 8$ m, the minimum $\overline{e}_{\Delta x_g}^{\mathrm{POD\text{-}MM}}$ is 0.025, occurring at $M = 10$. For $\Delta x_g < 8$ m, $\overline{e}_{\Delta x_g}^{\mathrm{POD\text{-}MM}}$ has negligible decrease for $M > 10$. The PDF of $\mathbf{f}_{\Delta x_g}^{\mathrm{POD\text{-}MM}}$ for $0.4 \le \theta \le 0.6$ is reasonably close to the PDF of $\mathbf{f}$ (Figure 16, shown for day 20 of year 1998 for which $e_{\Delta x_g}^{\mathrm{POD\text{-}MM}}$ is approximately the minimum $\overline{e}_{\Delta x_g}^{\mathrm{POD\text{-}MM}}$). However, for $0.6 \le \theta \le 0.8$, the fit is poorer with the PDF of $\mathbf{f}_{\Delta x_g}^{\mathrm{POD\text{-}MM}}$ resembling a dual-mode Gaussian distribution centered on 0.69 and 0.74. These peaks also deviate slightly from that of $\mathbf{f}$. The three modes in the PDF in Figure 16 correspond to the three different soil material properties used to characterize subsurface structure of the polygonal landscape.

The predicted pointwise soil moisture errors at site A have a maximum relative error of 0.15 and a mean of 0.02 (Figure 17, shown for 20-25 cm layer solutions of day 20 of year 1998), and is substantially less accurate than for the site-dependent ROM (Figure 8). To improve the accuracy of the reconstructed fine-resolution solution, we study the use of the POD-MM3 method. Since Bisht and Riley (2014) demonstrated that the soil moisture at each soil layer is inversely correlated to elevation, we define $\hat{\mathbf{f}}$ and $\hat{\mathbf{g}}$ in equation (14) as

$$\hat{\mathbf{f}}_{\ell_i} = -\mathrm{DEM}^{\mathrm{A}}_{\Delta x_f} \frac{\mu_{f,\ell_i}}{\mu_{\mathrm{DEM}^{\mathrm{A}}_{\Delta x_f}}}, \quad \hat{\mathbf{g}}_{\ell_i} = -\mathrm{DEM}^{\mathrm{A}}_{\Delta x_g} \frac{\mu_{g,\ell_i}}{\mu_{\mathrm{DEM}^{\mathrm{A}}_{\Delta x_g}}}, \quad 1 \le i \le 10, \tag{19}$$

where $\hat{\mathbf{f}}_{\ell_i}$ ($\hat{\mathbf{g}}_{\ell_i}$) is the ℓ_i layer solution of $\hat{\mathbf{f}}$ ($\hat{\mathbf{g}}$), $\mathrm{DEM}^{\mathrm{A}}_{\Delta x_f}$ ($\mathrm{DEM}^{\mathrm{A}}_{\Delta x_g}$) is the DEM of site A at resolution Δx_f (Δx_g), $\mu_{\mathrm{DEM}^{\mathrm{A}}_{\Delta x_f}}$ ($\mu_{\mathrm{DEM}^{\mathrm{A}}_{\Delta x_g}}$) is the average of the elevation over site A, and μ_{f,ℓ_i} (μ_{g,ℓ_i}) is the average of all $\mathbf{f}$ ($\mathbf{g}$) in the training data. Since, the DEM is 2-dimensional dataset, and $\mathbf{f}$ ($\mathbf{g}$) is 3-dimensional soil moisture fields, thus $\hat{\mathbf{f}}$ ($\hat{\mathbf{g}}$) is constructed separately for each vertical layer of the 3-dimensional domain of our discrete models of the sites via equation (19).

At $\Delta x_g = 8$ m, a minimum of 0.017 is obtained for $\bar{e}^{\text{POD-MM3}}_{\Delta x_g}$ when $M = 21$ (Figure 15(b)), compared to 0.025 for $\bar{e}^{\text{POD-MM}}_{\Delta x_g}$. The PDF of $\mathbf{f}^{\text{POD-MM3}}_{\Delta x_g}$ is also a closer approximation of the PDF of $\mathbf{f}$ compared to the PDF of $\mathbf{f}^{\text{POD-MM}}_{\Delta x_g}$ (Figure 16) and the heterogeneous structure of $\mathbf{f}$ is approximately reproduced (Figure 17). In addition, for the 5$^{\text{th}}$ soil layer, the mean and variance of $\varepsilon^{\text{POD-MM3}}_{\Delta x_g}$, defined analogously to $\varepsilon^{\text{POD-MM}}_{\Delta x_g}$ with $\mathbf{f}^{\text{POD-MM}}_{\Delta x_g}$ in equation (18) replaced by $\mathbf{f}^{\text{POD-MM3}}_{\Delta x_g}$, are more uniformly smaller than $\varepsilon^{\text{POD-MM}}_{\Delta x_g}$.

Fine-resolution soil moisture fields retrieved using the site-independent ROM are quite accurate (< 1.5%) given the large topographic differences between Site A and the remaining three sites. In other words, this approach led to an accurate fine-scale soil moisture prediction for a site that was excluded from the training dataset, but did share some topographic features with sites that were part of the training dataset. Our hypothesis is that the level of error from the site-independent ROM is well below that required for accurate prediction of soil moisture impacts on BGC dynamics. For example, at a moisture content of 0.4%, a mean relative error of 0.017 corresponds to an error in moisture content of 0.007%, which will have negligible impacts on GHG emission predictions.

To improve the performance of the site-independent ROM, the topography of the subdomains must be more carefully parameterized and sampled, allowing the impact of topographic variations on soil moisture to be captured by the ROM. For the above example, larger number of sites needs to be included in the training data. More generally, the inclusion of any spatially heterogeneous parameter requires proper parameterization and sampling of that parameter. Appropriate parameterization and sampling of a heterogeneous parameter is a research question that will be addressed in our future work.

3.4 Application to larger-scale hydrological simulations

The POD mapping method shows great promise in allowing prediction of fine-resolution soil moisture dynamics using coarse-resolution simulations. Here we applied a factor of 2^5 difference in resolution and achieved soil moisture simulation errors of <0.06% during two years that were not included in the training dataset, with an effective decrease in computation time of more than a factor of 1000. If the above results hold for simulations that include more sources of heterogeneity in the subsurface (e.g., conductivity) and surface (vegetation) properties, integration of the relevant ROMs into a land model such as CLM will allow much finer representation of processes than is currently possible without a drastic increase in computational cost.

The results indicate that the POD-MM is insensitive to fine- versus coarse-resolution simulation biases. This result is potentially useful in cases where we know coarse-resolution solutions are biased. For example, Chen and Durlofsky (2006) showed that an adaptive upscaling technique for subsurface permeability was needed to correct for bias in coarse simulation of synthetic channelized reservoir. To demonstrate that POD-MM corrects bias in the coarse solution, let $\mathbf{g}^{\text{bias}} = (1+\delta)\mathbf{g}$ where δ is a prescribed perturbation. We showed that $\boldsymbol{\alpha}^{\text{POD-MM}}$ is determined by solving

$$\begin{aligned}\boldsymbol{\alpha}^{\text{POD-MM}}(\mathbf{p}) &= \arg\min_{\gamma} \| \mathbf{g}^{\text{bias}}(\mathbf{p}) - \bar{\mathbf{g}}^{\text{bias}} - \sum_{i=1}^{M} \gamma_i \zeta_i^{\mathbf{g}^{\text{bias}},\text{POD-MM}} \|_2 \\ &= \arg\min_{\gamma} (1+\delta) \| \mathbf{g}(\mathbf{p}) - \bar{\mathbf{g}} - \sum_{i=1}^{M} \gamma_i \zeta_i^{\mathbf{g},\text{POD-MM}} \|_2\end{aligned} \quad (20)$$

for which the solution is equivalent to solving equation (8). Therefore, a constant bias will not affect the accuracy of our approximation. To further support the above analysis, we constructed and validated single-site ROMs constructed using $\mathbf{g}^{\text{bias}}$ for 0.01, 0.05, 0.1, 0.2, and 0.3 and for all Δx_g studied earlier. The results agreed with our earlier analysis and the errors are the same as when there was no bias (Figure 18, shown for $\Delta x_g = 8.0$ m but similar behaviors were obtained for all other Δx_g). Small differences only emerge at large M due to overfitting, the same reason we observed fluctuations in Figure 7. However, the above analysis does not apply to $\mathbf{f}_{\Delta x_g}^{\text{POD-mean}}$, since $\mathbf{f}_{\Delta x_g}^{\text{POD-mean}}$ relies on the assumption that the coarse- and the fine-resolution means have negligible differences. Thus, any bias in the coarse-resolution mean will lead to a biased $\mathbf{f}_{\Delta x_g}^{\text{POD-mean}}$. Further study is needed to study the biases in site-independent ROMs and the effects of coarse-resolution bias due to the upscaling of heterogeneous soil properties across scales.

The Arctic Tundra sites that we have studied have spatial extents that are smaller and landscapes that are less heterogeneous than domains studied in typical regional- and climate-scale simulations. Although our results conceptually demonstrate that the POD mapping method can accurately reconstruct fine-resolution solutions from coarse-resolution solutions, further development is needed to generalize the technique to problems of larger extent and diversity. The development of a site-independent ROM is one of the first steps in achieving this goal.

For larger-scale simulations, the parameter space that we are interested in is expected to be significantly more diverse (e.g., larger variations in topography and multiple landscape types). A single ROM will typically be inefficient since a large number of bases would be needed to accurately approximate the response of a diverse parameter space. Partitioning of the parameter space will allow us to construct multiple ROMs that are tailored to each domain. Dividing the parameter space based on the landscape types is one possible approach. Partitioning strategies, such as treed partitioning (Gramacy and Lee, 2008), can also help minimize the number of ROMs that we need to build.

Directly downscaling from 10 km scale (climate-scale) to 0.01 m (BGC-scale) may not be possible, especially if simulation at the finest scale is infeasible on the spatial extent used to simulate the coarse-scale solution. The coarse-scale solution may also have insufficient information to accurately reconstruct the finest-scale solution. We propose a hierarchical approach that involves using POD-MM methods to develop ROMs at multiple scales; scales at which these ROMs are built may critically depend on scales of the different processes we are modeling. The POD-MM reconstruction procedure is then recursively applied to reconstruct solution at progressively finer scale, starting from the coarsest scale solution. In addition, proper parameterization (such as parameterizing the topography) will allow finer-scale simulations to be performed on subsets of the original domain.

As with any sampling-based technique, the POD mapping method performs well only if the snapshots of the solution used to construct the ROM form an approximation space that can reasonably represent the solution. In the cases that we examined here, the annual cycle of the climate forcing does not change drastically from year to year and the response of soil moisture to climate forcing was relatively smooth. We thus obtained good predicted solutions using only data from a period of 3 years to build the ROM. However, for a more diverse parameter space, relying solely on historical climate forcings is insufficient. Statistical or adaptive sampling techniques should be used to sample the parameter space to ensure that future conditions not represented by historical data are accounted for. Accurately defining the extent of the parameter space is crucial. In addition, just as with any data assimilation technique, the ROM must be updated when new information is available, or when the forcing moves outside of the phase space under which the ROM was developed. For example, if we are using the ROM at a parameter point far outside the convex hull of the parameter space used to construct the ROM, it is a clear indication that the ROM needs to be updated to reflect the change in the extent of the parameter space.

The current method can be efficiently deployed within the existing CESM framework. For the cases that we have examined, the ROMs for the subsurface processes can be developed without considering the full coupled system so that the fine-resolution solutions can be determined more efficiently. Once ROMs are constructed, coarse resolution predictions of soil moisture can be mapped onto a fine grid to predict biogeochemical processes at higher spatial resolution, while the land-atmosphere interactions can still be modeled at a coarser grid. We will explore how such a ROM framework can be robustly implemented within the CLM model in future work.

Finally, while the computational costs of evaluating the ROM are typically low, the initial computational overhead required to construct the ROM can be large. High performance computing resources are needed to simulate the potentially large number of simulations required. Storing and retrieving the simulated solutions will also require good database management system and efficient parallel IO.

4 Conclusions

In this paper, we describe the construction of ROMs for land surface models based on POD-related methods. ROMs were built for soil moisture predictions from the PFLOTRAN model for the four NGEE-Arctic sites. An initial analysis based on the POD method is first used to determine whether POD-related methods can be used to accurately approximate the soil moisture. We then use four different methods that utilize coarse-resolution solutions to reconstruct fine-resolution solutions to construct single-site, multi-site, and site-independent ROMs. We evaluate their performance against fine-resolution simulations. Both the single-site and multi-site ROMs are very accurate (< 0.1%) with a computational speedup greater than 10^3. The site-independent ROM has a relative error < 1.5% when it is used to assess a site that is not included in the ROM training. However, the overall error magnitude is still quite low given the large topographical differences across the sites, thereby giving creditability for using ROMs in larger-scale simulations. We provide several approaches by which we can generalize our methods to problems of larger extent and diversity in this paper. We thus conclude that the integration of ROMs into an Earth System Modeling framework is practical and can provide an accurate approach to spatial scaling,

Acknowledgements: This research was supported by the Director, Office of Science, Office of Biological and Environmental Research of the U.S. Department of Energy under Contract #DE-AC02-05CH11231 as part of the Early Career Research Program (Pau) and the Terrestrial Ecosystem Science Program including the Next-Generation Ecosystem Experiments (NGEE Arctic) project (Bisht and Riley).

References

Albertson, J. D. and Montaldo, N.: Temporal dynamics of soil moisture variability: 1. Theoretical basis, Water Resour Res, 39, 2003.

Arrigo, J. A. S. and Salvucci, G. D.: Investigation hydrologic scaling: Observed effects of heterogeneity and nonlocal processes across hillslope, watershed, and regional scales, Water Resour Res, 41, 2005.

Bacmeister, J. T., Wehner, M. F., Neale, R. B., Gettelman, A., Hannay, C., Lauritzen, P. H., Caron, J. M., and Truesdale, J. E.: Exploratory high-resolution climate simulations using the Community Atmosphere Model (CAM), Journal of Climate, doi: 10.1175/JCLI-D-13-00387.1, 2013. 2013.

Barrios, M. and Francés, F.: Spatial scale effect on the upper soil effective parameters of a distributed hydrological model, Hydrol Process, 26, 1022-1033, 2012.

Beven, K. J. and Cloke, H. L.: Comment on “Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water” by Eric F. Wood et al, Water Resour Res, 48, W01801, 2012.

Bisht, G. and Riley, W. J.: Topographic controls on soil moisture scaling properties in polygonal ground, 2014. In preparation.

Brocca, L., Melone, F., Moramarco, T., and Morbidelli, R.: Spatial-temporal variability of soil moisture and its estimation across scales, Water Resour Res, 46, W02516, 2010.

Brocca, L., Morbidelli, R., Melone, F., and Moramarco, T.: Soil moisture spatial variability in experimental areas of central Italy, J Hydrol, 333, 356-373, 2007.

Brocca, L., Tullo, T., Melone, F., Moramarco, T., and Morbidelli, R.: Catchment scale soil moisture spatial-temporal variability, J Hydrol, 422, 63-75, 2012.

Chen, Y. and Durlofsky, L.: Adaptive local-global upscaling for general flow scenarios in heterogeneous formations, Transport in Porous Media, 62, 157-185, 2006.

Choi, H. I., Kumar, P., and Liang, X. Z.: Three-dimensional volume-averaged soil moisture transport model with a scalable parameterization of subgrid topographic variability, Water Resour Res, 43, W04414, 2007.

Choi, M. and Jacobs, J. M.: Spatial soil moisture scaling structure during Soil Moisture Experiment 2005, Hydrol Process, 25, 926-932, 2011.

Das, N. N. and Mohanty, B. P.: Temporal dynamics of PSR-based soil moisture across spatial scales in an agricultural landscape during SMEX02: A wavelet approach, Remote Sens Environ, 112, 522-534, 2008.
Everson, R. and Sirovich, L.: Karhunen–Loeve procedure for gappy data, Journal of the Optical Society of America A, 12, 1657-1664, 1995.
Famiglietti, J. S., Devereaux, J. A., Laymon, C. A., Tsegaye, T., Houser, P. R., Jackson, T. J., Graham, S. T., Rodell, M., and van Oevelen, P. J.: Ground-based investigation of soil moisture variability within remote sensing footprints during the Southern Great Plains 1997 (SGP97) Hydrology Experiment, Water Resour Res, 35, 1839-1851, 1999.
Famiglietti, J. S., Ryu, D., Berg, A. A., Rodell, M., and Jackson, T. J.: Field observations of soil moisture variability across scales, Water Resour Res, 44, W01423, 2008.
Frei, S., Knorr, K. H., Peiffer, S., and Fleckenstein, J. H.: Surface micro-topography causes hot spots of biogeochemical activity in wetland systems: A virtual modeling experiment, J Geophys Res-Biogeo, 117, 2012.
Gramacy, R. B. and Lee, H. K. H.: Bayesian treed Gaussian process models with an application to computer modeling, Journal of the American Statistical Association, 103, 1119-1130, 2008.
Hammond, G. E., Lichtner, P. C., Lu, C., and R.T., M.: PFLOTRAN: Reactive flow and transport code for use on laptops to leadership-class supercomputers. In: Groundwater Reactive Transport Models Zhang, F., Yeh, G. T., and Parker, J. C. (Eds.), Bentham Science Publishers, Sharjah, UAE, 2012.
Hinkel, K. M., Doolittle, J. A., Bockheim, J. G., Nelson, F. E., Paetzold, R., Kimble, J. M., and Travis, R.: Detection of subsurface permafrost features with ground-penetrating radar, Barrow, Alaska, Permafrost and Periglacial Processes, 12, 179-190, 2001.
Hinkel, K. M., Eisner, W. R., Bockheim, J. G., Nelson, F. E., Peterson, K. M., and Dai, X.: Spatial extent, age, and carbon stocks in drained thaw lake basins on the Barrow Peninsula, Alaska, Arctic, Antarctic and Alpine Research, 35, 291-300, 2003.
Hinzman, L. D., Kane, D. L., Gieck, R. E., and Everett, K. R.: Hydrologic and thermal-properties of the active layer in the Alaskan Arctic, Cold Regions Science and Technology, 19, 95-110, 1991.

Hu, Z. L., Islam, S., and Cheng, Y. Z.: Statistical characterization of remotely sensed soil moisture images, Remote Sens Environ, 61, 310-318, 1997.

Hubbard, S. S., Gangodagamage, C., Dafflon, B., Wainwright, H., Peterson, J., Gusmeroli, A., Ulrich, C., Wu, Y., Wilson, C., Rowland, J., Tweedie, C., and Wullschleger, S. D.: Quantifying and relating land-surface and subsurface variability in permafrost environments using LiDAR and surface geophysical datasets, Hydrogeology Journal, 21, 149-169, 2013.

Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., Lamarque, J. F., Large, W. G., Lawrence, D., Lindsay, K., Lipscomb, W. H., Long, M. C., Mahowald, N., Marsh, D. R., Neale, R. B., Rasch, P., Vavrus, S., Vertenstein, M., Bader, D., Collins, W. D., Hack, J. J., Kiehl, J., and Marshall, S.: The Community Earth System Model: A framework for collaborative research, Bulletin of the American Meteorological Society, 94, 1339-1360, 2013.

Ivanov, V. Y., Fatichi, S., Jenerette, G. D., Espeleta, J. F., Troch, P. A., and Huxman, T. E.: Hysteresis of soil moisture spatial heterogeneity and the "homogenizing" effect of vegetation, Water Resour Res, 46, W09521, 2010.

Jana, R. B. and Mohanty, B. P.: A topography-based scaling algorithm for soil hydraulic parameters at hillslope scales: Field testing, Water Resour Res, 48, W02519, 2012.

Jolliffe, I. T.: Principal component analysis, Springer, New York, 2002.

Joshi, C. and Mohanty, B. P.: Physical controls of near-surface soil moisture across varying spatial scales in an agricultural landscape during SMEX02, Water Resour Res, 46, 2010.

Koven, C. D., Riley, W. J., Subin, Z. M., Tang, J. Y., Torn, M. S., Collins, W. D., Bonan, G. B., Lawrence, D. M., and Swenson, S. C.: The effect of vertically resolved soil biogeochemistry and alternate soil C and N models on C dynamics of CLM4, Biogeosciences, 10, 7109-7131, 2013.

Kumar, P.: Layer averaged Richard's equation with lateral flow, Adv Water Resour, 27, 521-531, 2004.

Lawrence, J. E. and Hornberger, G. M.: Soil moisture variability across climate zones, Geophys Res Lett, 34, L20402, 2007.

Lawrence, P. J., Feddema, J. J., Bonan, G. B., Meehl, G. A., O'Neill, B. C., Oleson, K. W., Levis, S., Lawrence, D. M., Kluzek, E., Lindsay, K., and Thornton, P. E.: Simulating the biogeochemical and biogeophysical impacts of transient land cover change and wood harvest in the Community Climate System Model (CCSM4) from 1850 to 2100, Journal of Climate, 25, 3071-3095, 2012.

Li, B. and Rodell, M.: Spatial variability and its scale dependency of observed and modeled soil moisture over different climate regions, Hydrol Earth Syst Sc, 17, 1177-1188, 2013.

Mascaro, G., Vivoni, E. R., and Deidda, R.: Downscaling soil moisture in the southern Great Plains through a calibrated multifractal model for land surface modeling applications, Water Resour Res, 46, 2010.

Mascaro, G., Vivoni, E. R., and Deidda, R.: Soil moisture downscaling across climate regions and its emergent properties, J Geophys Res-Atmos, 116, 2011.

Maxwell, R. M., Putti, M., Meyerhoff, S., Delfs, J.-O., Ferguson, I. M., Ivanov, V., Kim, J., O.Kolditz, Kollet, S. J., Kumar, M., Paniconi, C., Park, Y.-J., Phanikumar, M. S., Sudicky, E., and Sulis, M.: Surface-subsurface model intercomparison: A first set of benchmark results to diagnose integrated hydrology and feedbacks, Vienna, Austria, April 22-27 2012.

McClain, M. E., Boyer, E. W., Dent, C. L., Gergel, S. E., Grimm, N. B., Groffman, P. M., Hart, S. C., Harvey, J. W., Johnston, C. A., Mayorga, E., McDowell, W. H., and Pinay, G.: Biogeochemical hot spots and hot moments at the interface of terrestrial and aquatic ecosystems, Ecosystems, 6, 301-312, 2003.

Montaldo, N. and Albertson, J. D.: Temporal dynamics of soil moisture variability: 2. Implications for land surface models, Water Resour Res, 39, 2003.

Moore, B. C.: Principal component analysis in linear systems - controllability, observability, and model-reduction, Ieee Transactions on Automatic Control, 26, 17-32, 1981.

Nykanen, D. K. and Foufoula-Georgiou, E.: Soil moisture variability and scale-dependency of nonlinear parameterizations in coupled land-atmosphere models, Adv Water Resour, 24, 1143-1157, 2001.

Oleson, K. W., D.M. Lawrence, G.B. Bonan, B. Drewniak, M. Huang, C.D. Koven, S. Levis, F. Li, W.J. Riley, Z.M. Subin, S.C. Swenson, P.E. Thornton, A. Bozbiyik, R. Fisher, E. Kluzek, J.-F. Lamarque, P.J. Lawrence, L.R. Leung, W. Lipscomb, S. Muszala, D.M. Ricciuto, W. Sacks, Y. Sun, J. Tang, Z.-L. Yang: Technical description of version 4.5 of the Community Land Model (CLM), National Center for Atmospheric Research, Boulder, CO, 2013.

Pan, F. and Peters-Lidard, C. D.: On the relationship between mean and variance of soil moisture fields, J Am Water Resour As, 44, 235-242, 2008.

Pau, G. S. H., Zhang, Y., and Finsterle, S.: Reduced order models for many-query subsurface flow applications, Computational Geosciences, 17, 705-721, 2013a.

Pau, G. S. H., Zhang, Y., Finsterle, S., Wainwright, H., and Birkholzer, J.: Reduced order modeling in iTOUGH2, Computers & Geosciences, doi: 10.1016/j.cageo.2013.08.008, 2013b. 2013b.

Riley, W. J. and Shen, C.: Characterizing coarse-resolution watershed soil moisture heterogeneity using fine-scale simulations, Hydrol. Earth Syst. Sci. Discuss., 11, 1967-2009, 2014.

Robinson, T., Eldred, M., Willcox, K., and Haimes, R.: Strategies for multifidelity optimization with variable dimensional hierarchical models, 47th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, Reston, Virigina, 2012.

Rodriguez-Iturbe, I., Vogel, G. K., Rigon, R., Entekhabi, D., Castelli, F., and Rinaldo, A.: On the spatial-organization of soil-moisture fields, Geophys Res Lett, 22, 2757-2760, 1995.

Rosenbaum, U., Bogena, H. R., Herbst, M., Huisman, J. A., Peterson, T. J., Weuthen, A., Western, A. W., and Vereecken, H.: Seasonal and event dynamics of spatial soil moisture patterns at the small catchment scale, Water Resour Res, 48, 2012.

Ryu, D. and Famiglietti, J. S.: Characterization of footprint-scale surface soil moisture variability using Gaussian and beta distribution functions during the Southern Great Plains 1997 (SGP97) hydrology experiment, Water Resour Res, 41, W12433, 2005.

Schuur, E. A. G., Bockheim, J., Canadell, J. G., Euskirchen, E., Field, C. B., Goryachkin, S. V., Hagemann, S., Kuhry, P., Lafleur, P. M., Lee, H., Mazhitova, G., Nelson, F. E.,

Rinke, A., Romanovsky, V. E., Shiklomanov, N., Tarnocai, C., Venevsky, S., Vogel, J. G., and Zimov, S. A.: Vulnerability of permafrost carbon to climate change: Implications for the global carbon cycle, BioScience, 58, 701-714, 2008.
Sellmann, P. V., Brown, J., Lewellen, R. I., McKim, H., and Merry, C.: The classification and geomorphic implication of thaw lakes on the Arctic Coastal Plain, Alaska, U.S. Army Cold Reg. Res. Eng. Lab, Hanover, NH, 1975.
Shen, C.: A process-based distributed hydrologic model and its application to a Michigan watershed, Ph.D., Civil and Environmental Engineering, Michigan State University, East Lansing, MI, 270 pp., 2009.
Tague, C., Band, L., Kenworthy, S., and Tenebaum, D.: Plot- and watershed-scale soil moisture variability in a humid Piedmont watershed, Water Resour Res, 46, 2010.
Tang, J. Y., Riley, W. J., Koven, C. D., and Subin, Z. M.: CLM4-BeTR, a generic biogeochemical transport and reaction module for CLM4: model development, evaluation, and application, Geoscientific Model Development, 6, 127-140, 2013.
Teuling, A. J., Hupet, F., Uijlenhoet, R., and Troch, P. A.: Climate variability effects on spatial soil moisture dynamics, Geophys Res Lett, 34, 2007.
Teuling, A. J. and Troch, P. A.: Improved understanding of soil moisture variability dynamics, Geophys Res Lett, 32, L05404, 2005.
Torn, M. S. and Chapin, F. S.: Environmental and biotic controls over Methane flux from Arctic tundra, Chemosphere, 26, 357-368, 1993.
Vereecken, H., Kamai, T., Harter, T., Kasteel, R., Hopmans, J., and Vanderborght, J.: Explaining soil moisture variability as a function of mean soil moisture: A stochastic unsaturated flow perspective, Geophys Res Lett, 34, 2007.
Vivoni, E. R., Entekhabi, D., Bras, R. L., and Ivanov, V. Y.: Controls on runoff generation and scale-dependence in a distributed hydrologic model, Hydrol Earth Syst Sc, 11, 1683-1701, 2007.
Walker, D. A., Raynolds, M. K., Daniëls, F. J. A., Einarsson, E., Elvebakk, A., Gould, W. A., Katenin, A. E., Kholod, S. S., Markon, C. J., Melnikov, E. S., Moskalenko, N. G., Talbot, S. S., Yurtsev, B. A., and The other members of the, C. T.: The circumpolar Arctic vegetation map, Journal of Vegetation Science, 16, 267-282, 2005.

Wehner, M. F., Reed, K., Li, F., Prabhat, J. B., Chen, C.-T., Paciorek, C., Gleckler, P., Sperber, K., Collins, W. D., Gettelman, A., Jablonowski, C., and Algieri, C.: The effect of horizontal resolution on simulation quality in the Community Atmospheric Model, CAM5.1, Submitted to the Journal of Modeling the Earth System, 2014. 2014.

Wilkinson, R. D.: Bayesian calibration of expensive multivariate computer experiments, Large-Scale Inverse Problems and Quantification of Uncertainty, 707, 195-215, 2011.

Willcox, K. and Peraire, J.: Balanced model reduction via the proper orthogonal decomposition, AIAA journal, 40, 2323-2330, 2002.

Wood, E. F.: Effects of soil moisture aggregation on surface evaporative fluxes, J Hydrol, 190, 397-412, 1997.

Wood, E. F.: Scale analyses for land-surface hydrology. In: Scale Dependence and Scale Invariance in Hydrology, Sposito, G. (Ed.), Cambridge University Press, Cambridge, UK, 1998.

Wood, E. F., Roundy, J. K., Troy, T. J., van Beek, L. P. H., Bierkens, M. F. P., Blyth, E., de Roo, A., Doll, P., Ek, M., Famiglietti, J., Gochis, D., van de Giesen, N., Houser, P., Jaffe, P. R., Kollet, S., Lehner, B., Lettenmaier, D. P., Peters-Lidard, C., Sivapalan, M., Sheffield, J., Wade, A., and Whitehead, P.: Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water, Water Resour Res, 47, 2011.

Wu, Y., Hubbard, S. S., Ulrich, C., and Wullschleger, S. D.: Remote monitoring of freeze-thaw transitions in Arctic soils using the complex resistivity method, Vadose Zone J., 12, 2013.

TABLES:

Method	Reference basis	ith column of the data matrix	Determination of $\boldsymbol{\alpha}(\mathbf{p})$
POD	$\bar{\mathbf{f}}$	$\mathbf{f}(\mathbf{q}_i)-\bar{\mathbf{f}}$	Equation (4).
POD-mean	$\bar{\mathbf{f}}$	$\mathbf{f}(\mathbf{q}_i)-\bar{\mathbf{f}}$	Approximated by $\boldsymbol{\alpha}^{\text{fit}}(\mu_g(\mathbf{p}))$ where $\boldsymbol{\alpha}^{\text{fit}}$ is a polynomial fit between $\boldsymbol{\alpha}(\mathbf{q})$ and $\mu_f(\mathbf{q})$.
POD-MM	$\begin{bmatrix}\bar{\mathbf{f}}\\ \bar{\mathbf{g}}\end{bmatrix}$	$\begin{bmatrix}\mathbf{f}(\mathbf{q}_i)-\bar{\mathbf{f}}\\ \mathbf{g}(\mathbf{q}_i)-\bar{\mathbf{g}}\end{bmatrix}$	Equation (8) using $\mathbf{g}(\mathbf{p})$.
POD-MM2	$\begin{bmatrix}\bar{\mathbf{h}}\\ \bar{\mathbf{g}}\end{bmatrix}$	$\begin{bmatrix}\mathbf{h}(\mathbf{q}_i)-\bar{\mathbf{h}}\\ \bar{\mathbf{g}}\end{bmatrix}$	Equation (8), by substituting $\zeta_i^{\mathbf{g},\text{POD-MM}}$ by $\zeta_i^{\mathbf{g},\text{POD-MM2}}$.
POD-MM3	$\begin{bmatrix}\hat{\mathbf{f}}\\ \hat{\mathbf{g}}\end{bmatrix}$	$\begin{bmatrix}\mathbf{f}(\mathbf{q}_i)-\hat{\mathbf{f}}\\ \mathbf{g}(\mathbf{q}_i)-\hat{\mathbf{g}}\end{bmatrix}$	Equation (8), by substituting $\zeta_i^{\mathbf{g},\text{POD-MM}}$ by $\zeta_i^{\mathbf{g},\text{POD-MM3}}$ and $\bar{\mathbf{g}}$ by $\hat{\mathbf{g}}$.

Table 1. Summary of differences between various methods used for constructing ROM. In above, $\mathbf{f}$ is the fine resolution solution; $\mathbf{g}$ is the coarse resolution solution; $\mathbf{h}$ is given by equation (10); $\hat{\mathbf{f}}$ and $\hat{\mathbf{g}}$ are given by equation (15); $\mathbf{p}$ is any given parameter set; $\mathbf{q}_i$ is the ith parameter set in $\boldsymbol{S}_N$; $\mu_f(\mathbf{p})$ and $\mu_g(\mathbf{p})$ are spatially averaged $\mathbf{f}(\mathbf{p})$ and $\mathbf{g}(\mathbf{p})$; and $\zeta_i^{\mathbf{g},\text{POD-MM}}$, $\zeta_i^{\mathbf{g},\text{POD-MM2}}$ and $\zeta_i^{\mathbf{g},\text{POD-MM3}}$ are POD bases for POD-MM, POD-MM2 and POD-MM3 methods, respectively. (Please refer to each method's subsection in the Methods section for more details on the above variables.)

FIGURES:

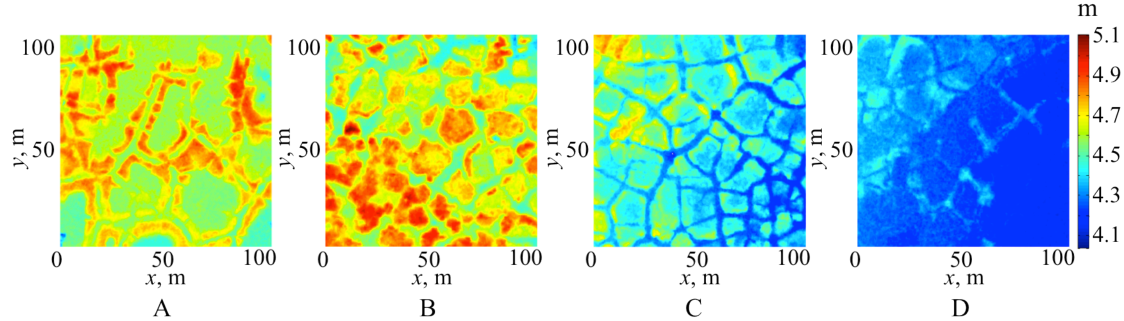

Figure 1. DEM for site A, B, C, and D. The spatial extent of each site is 104 m x 104 m.

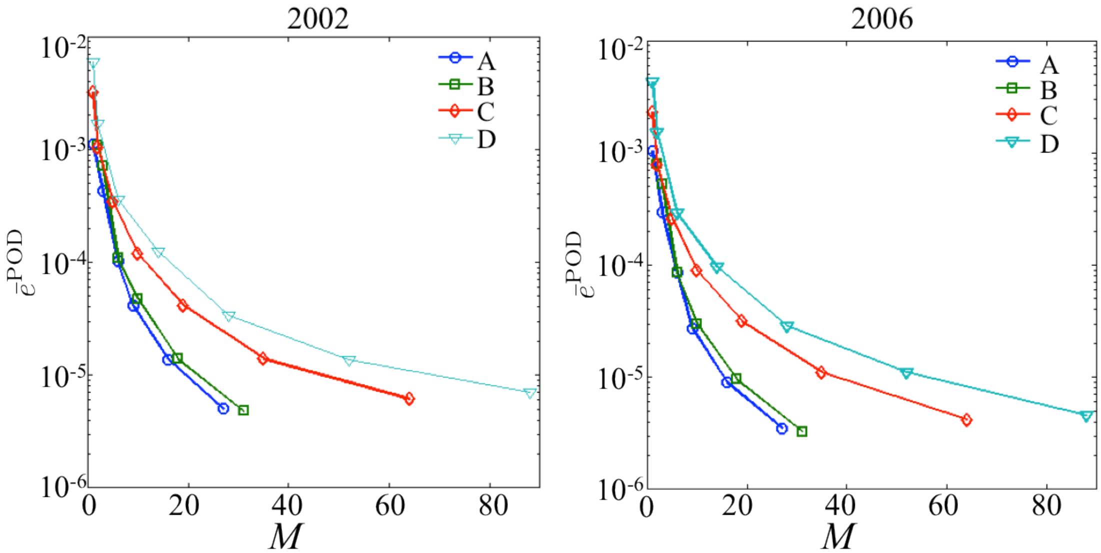

Figure 2. Variation of the mean POD error ($\bar{e}^{\mathrm{POD}}$), with respect to number of bases (M), in year 2002 and 2006 for single-site ROM constructed using POD method.

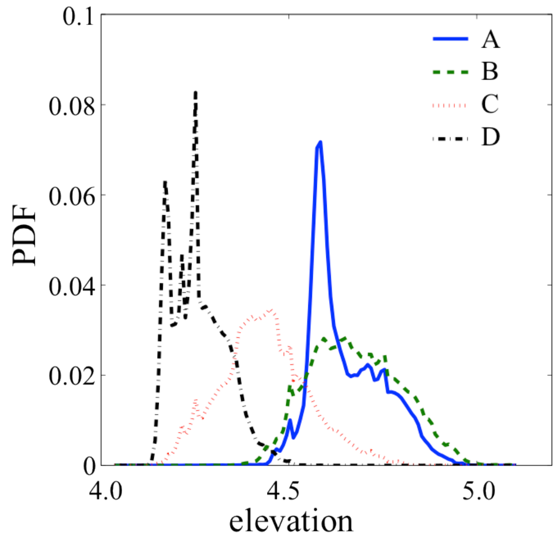

Figure 3. Elevation distributions of the DEM for sites A, B, C and D.

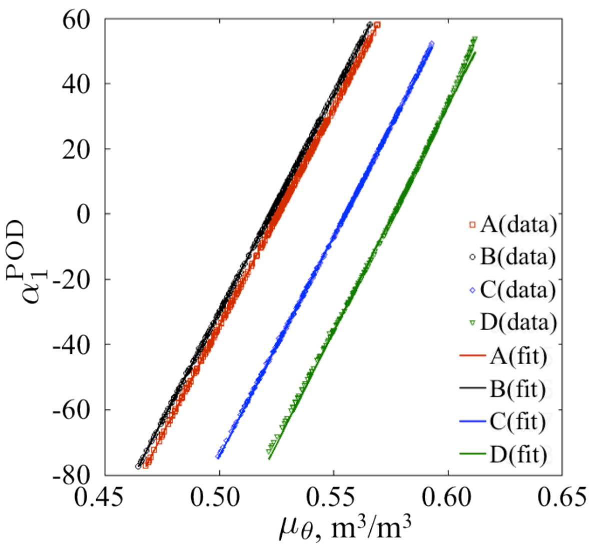

Figure 4. Relation between mean soil moisture $\mu_f(\mathbf{q}) = \mu_\theta(\mathbf{q})$ and $\alpha_1^{\mathrm{POD}}(\mathbf{q})$ for sites A, B, C, and D. The lines are linear fits to the data (symbols).

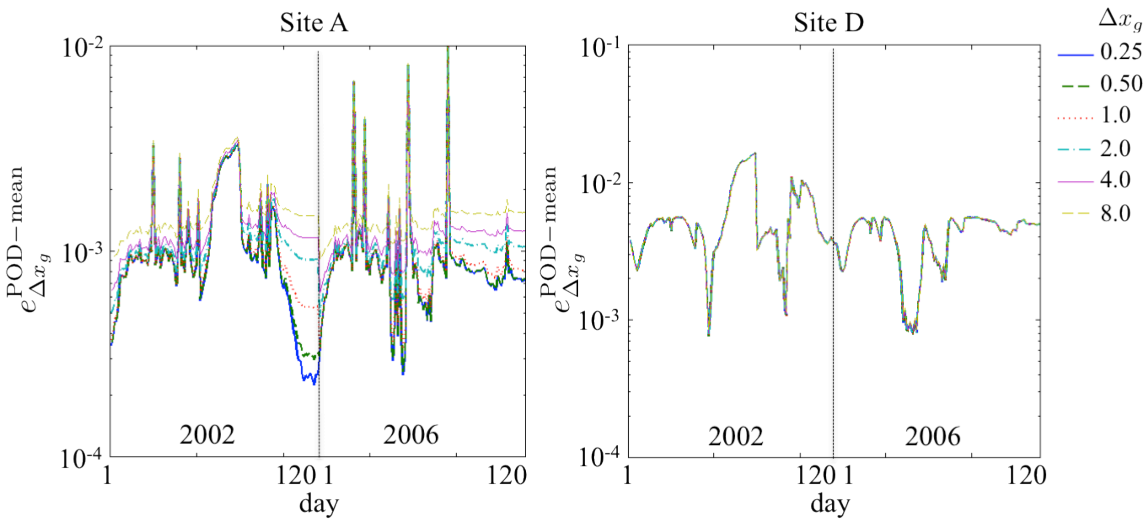

Figure 5. POD-mean error ($e_{\Delta x_g}^{\text{POD-mean}}$) at sites A and D in years 2002 and 2006 for single-site ROM constructed using POD-mean method.

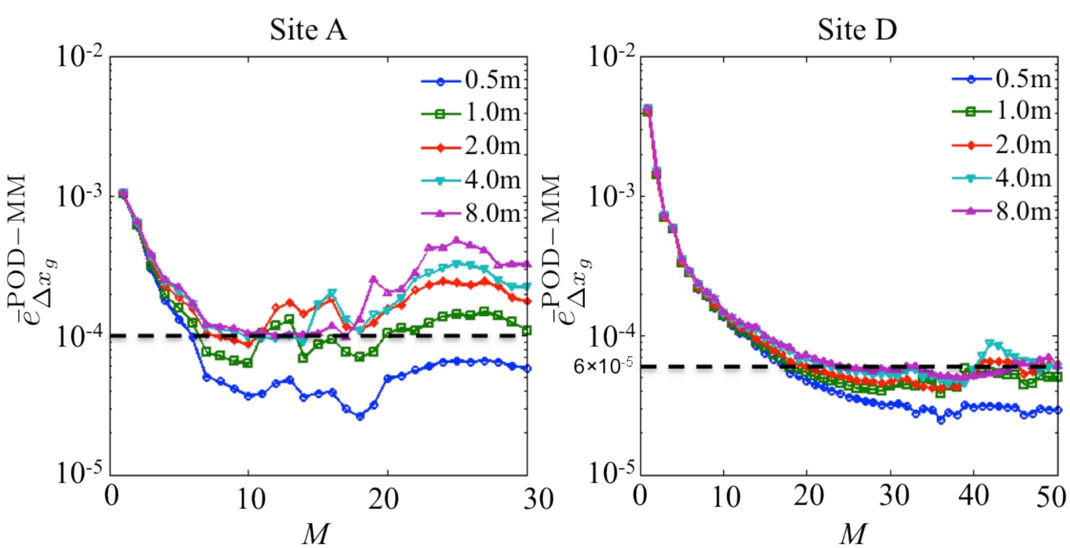

Figure 6. The variation of mean POD-MM error ($\bar{e}_{\Delta x_g}^{\text{POD-MM}}$) with respect to M for different Δx_g at sites A and D in 2006. Results are shown for single-site ROM constructed using POD-MM method.

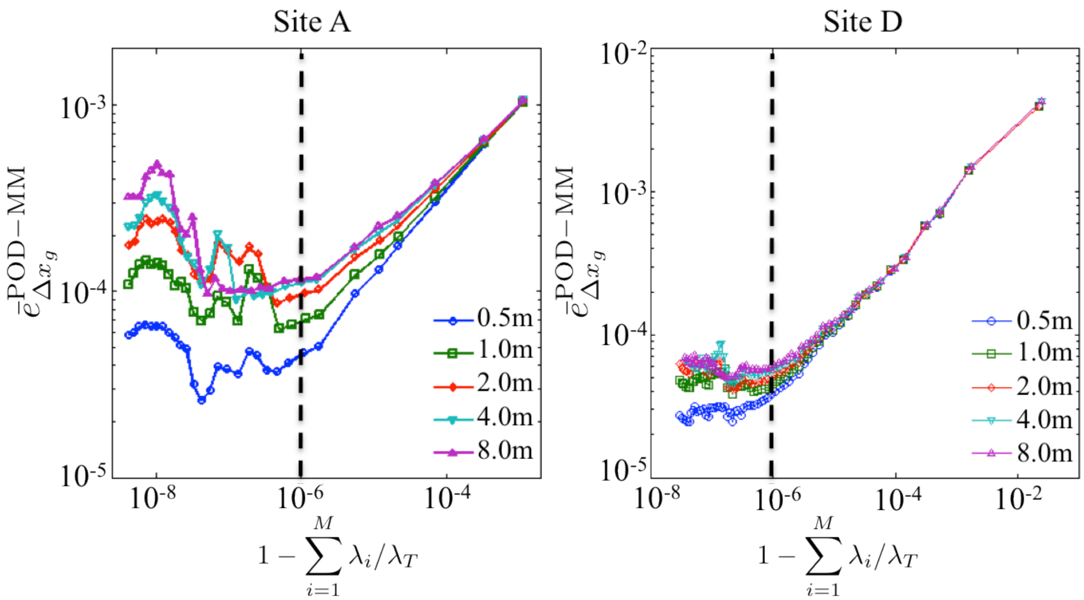

Figure 7. The variation of the mean POD-MM error ($\bar{e}_{\Delta x_g}^{\text{POD-MM}}$) with respect to $e_M^{\lambda}=1-\sum_{i=1}^{M}\lambda_i/\lambda_T$ for different Δx_g at sites A and D in 2006.

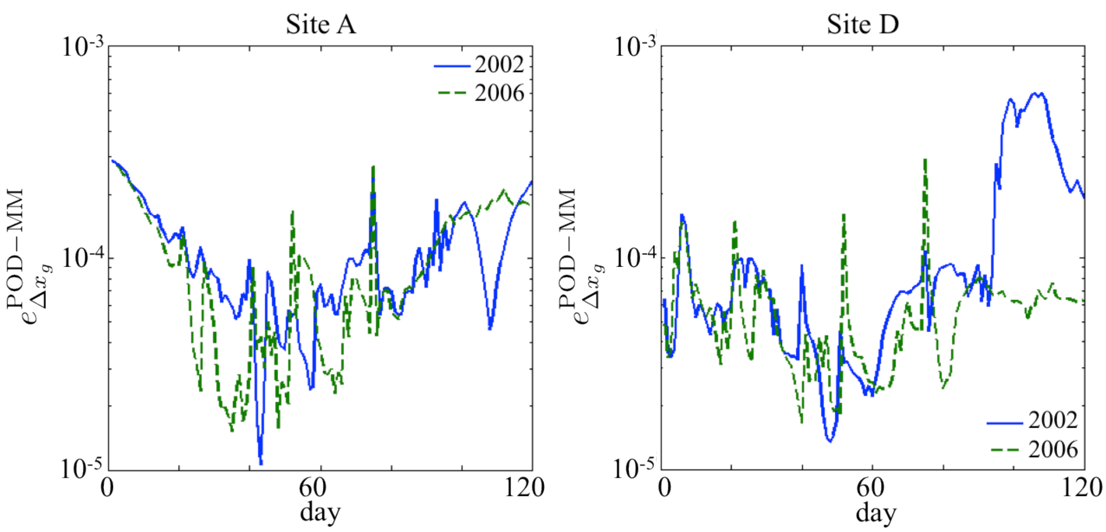

Figure 8. The POD-MM error ($e_{\Delta x_g}^{\text{POD-MM}}$) in 2002 and 2006 at sites A and D for single-site ROM constructed using POD-MM method.

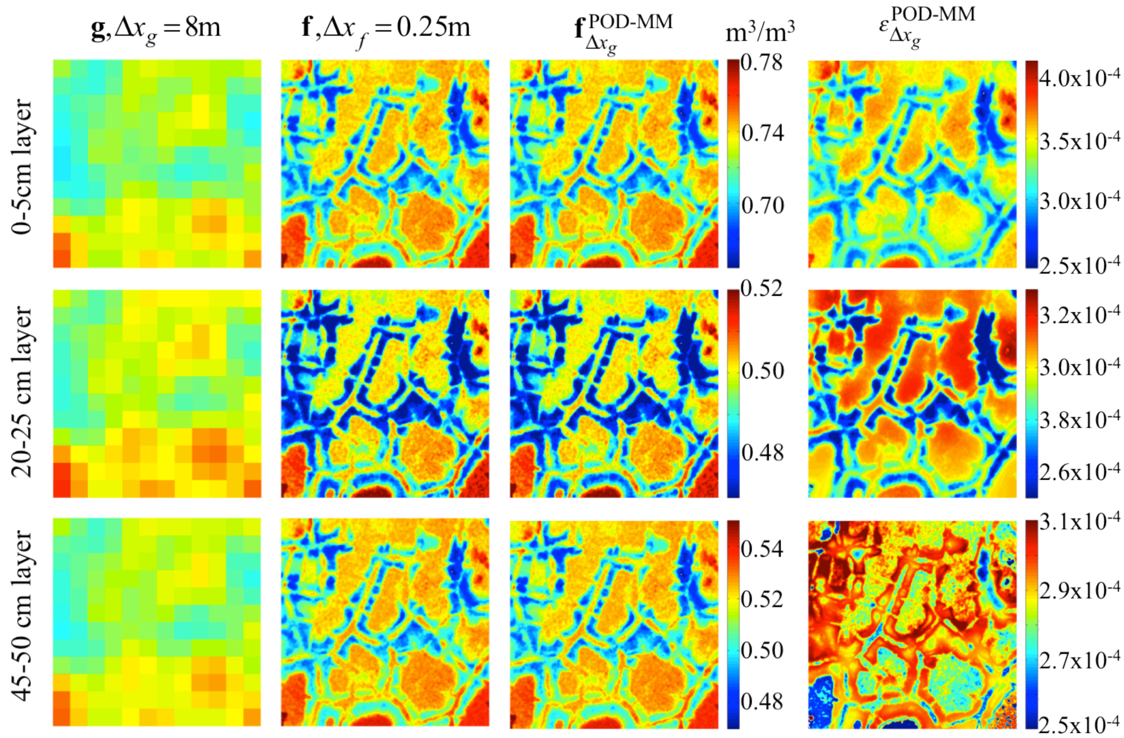

Figure 9. Solutions of $\mathbf{f}$, $\mathbf{g}$, and $\varepsilon_{\Delta x_g}^{\text{POD-MM}}$ for day 1 of year 2002 at site A and for three soil depths; the top, middle, and bottom rows correspond to layers 0-5 cm, 20-25 cm, and 45-50 cm, respectively, from the surface.

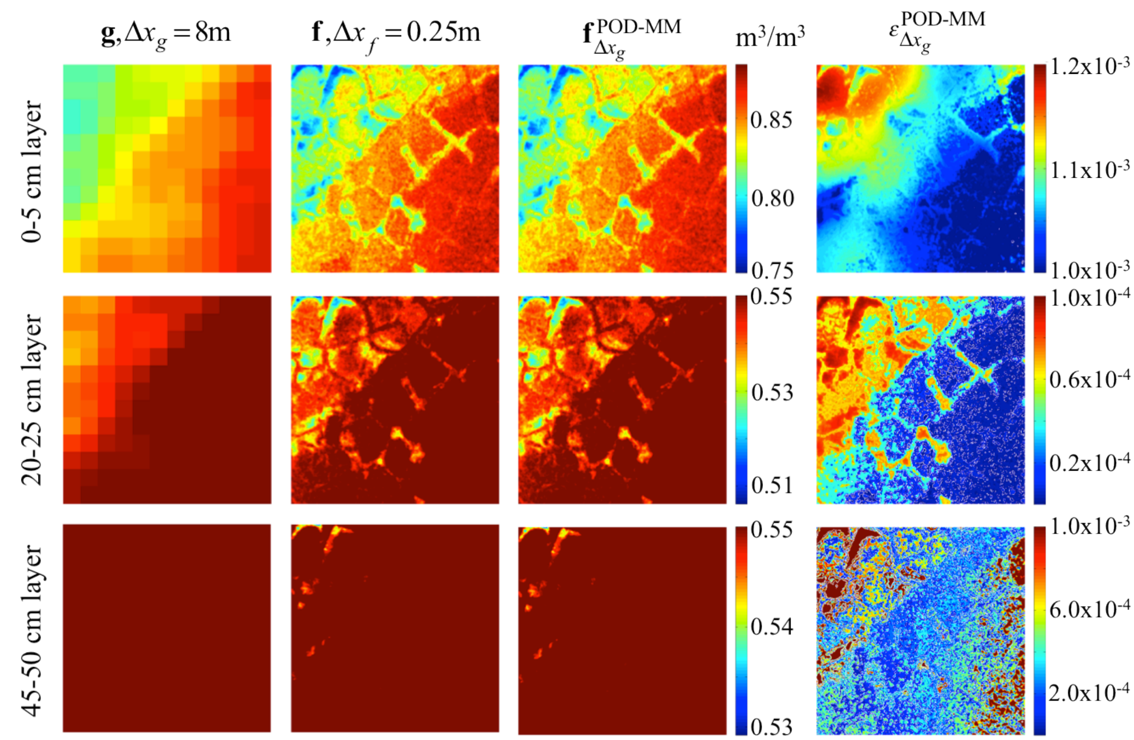

Figure 10. Solutions of $\mathbf{f}$, $\mathbf{g}$, and $\varepsilon_{\Delta x_g}^{\text{POD-MM}}$ for day 106 of year 2002 at site D and for three soil depths; the top, middle, and bottom rows correspond to layers 0-5 cm, 20-25 cm, and 45-50 cm, respectively, from the surface. Regions with homogeneous red color in the panels reflect the fact that large regions of the solutions are saturated.

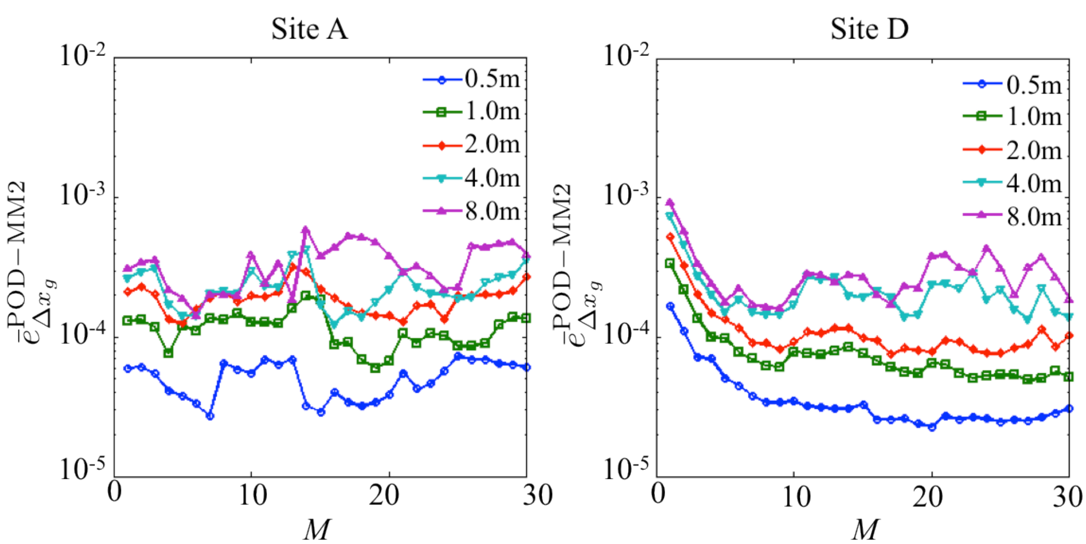

Figure 11. The variation of mean POD-MM2 error ($\bar{e}_{\Delta x_g}^{\text{POD-MM2}}$) with respect to M for different Δx_g at sites A and D in 2006. Results are shown for single-site ROM constructed using POD-MM2 method.

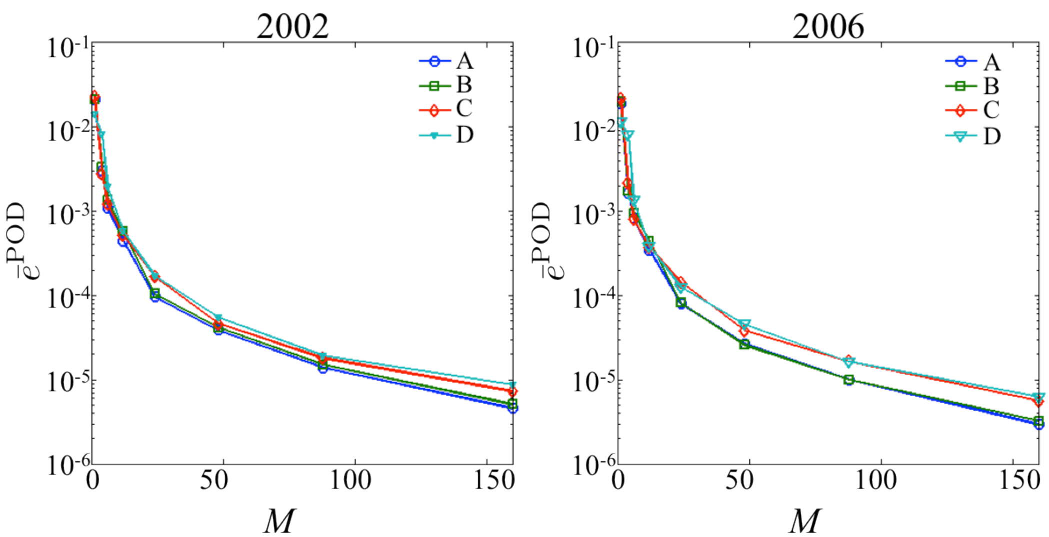

Figure 12. Variation of the mean of POD error ($\bar{e}^{\text{POD}}$) with respect to M in 2002 and 2006 for multi-site ROM constructed using POD method.

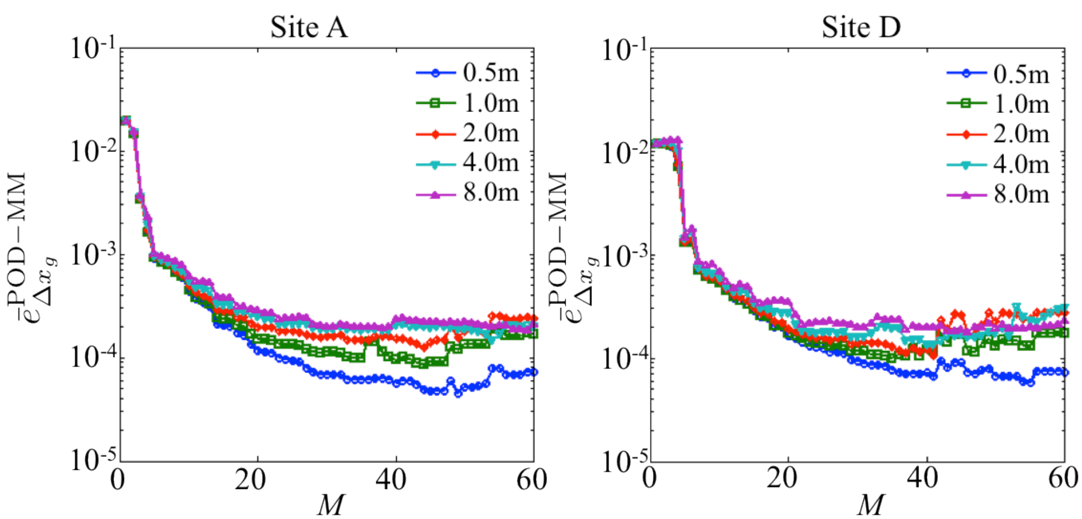

Figure 13. The variation of mean POD-MM error ($\bar{e}_{\Delta x_g}^{\text{POD-MM}}$) with respect to M for different Δx_g at sites A and D in 2006. Results are shown for multi-site ROM constructed using POD-MM method.

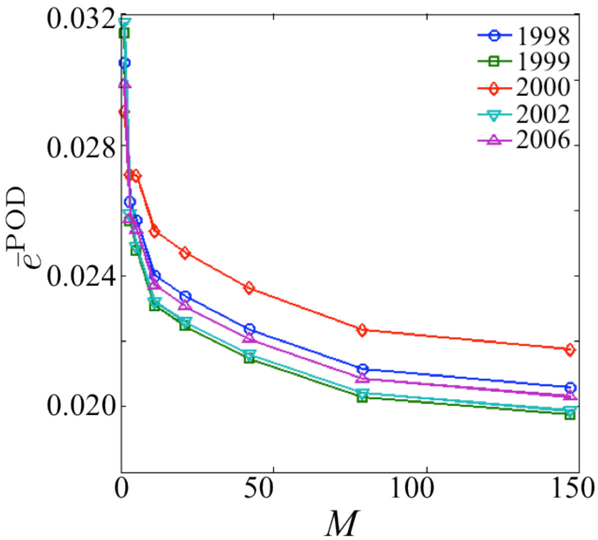

Figure 14. The mean POD error $\bar{e}^{\text{POD}}$ at site A based on a site-independent ROM constructed using POD method that only utilizes soil moisture solutions from sites B, C, and D.

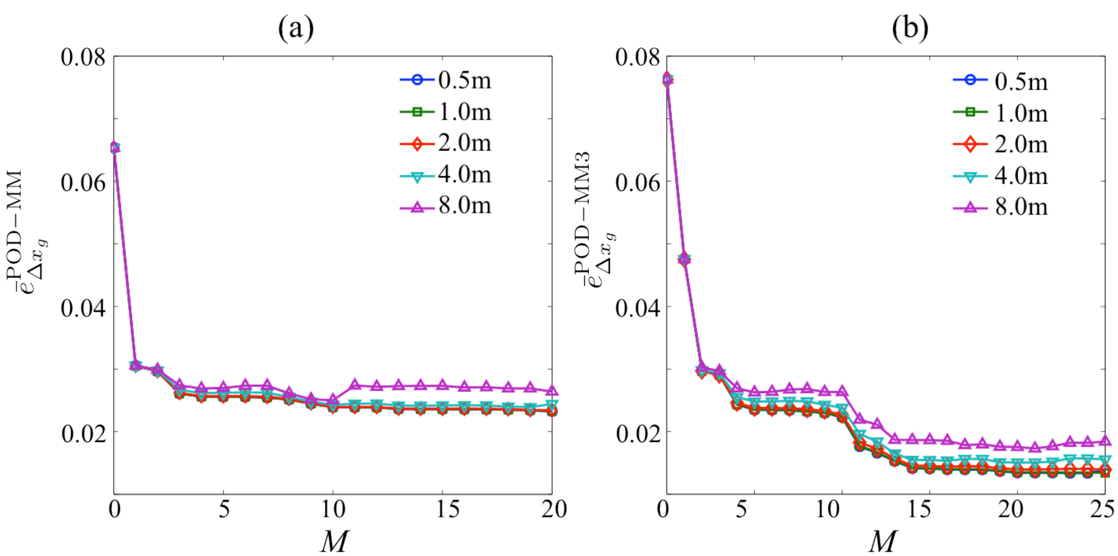

Figure 15. The errors (a) $\overline{e}_{\Delta x_g}^{\text{POD-MM}}$ and (b) $\overline{e}_{\Delta x_g}^{\text{POD-MM3}}$ versus M for different Δx_g at site A for site-independent ROM constructed using POD-MM and POD-MM3 methods respectively. The means are taken over 1998, 1999, 2000, 2002, and 2006.

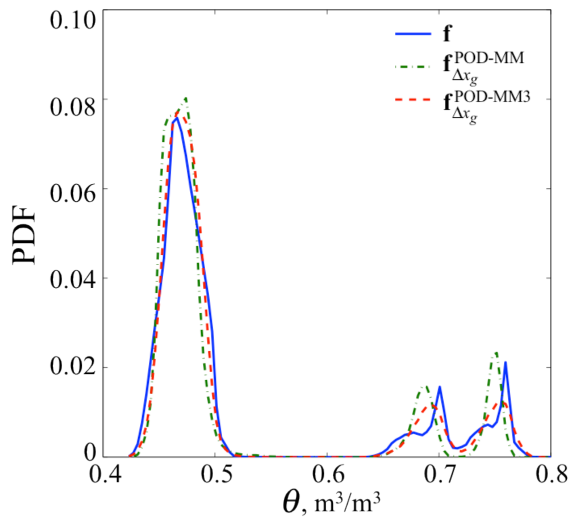

Figure 16. The probability density function (PDF) of $\mathbf{f}$, $\mathbf{f}_{\Delta x_g}^{\text{POD-MM}}$ and $\mathbf{f}_{\Delta x_g}^{\text{POD-MM3}}$ for day 20 of year 1998 at site A for which $e_{\Delta x_g}^{\text{POD-MM3}}$ is approximately equal to $\overline{e}_{\Delta x_g}^{\text{POD-MM3}}$; $\Delta x_g = 8\,\text{m}$.

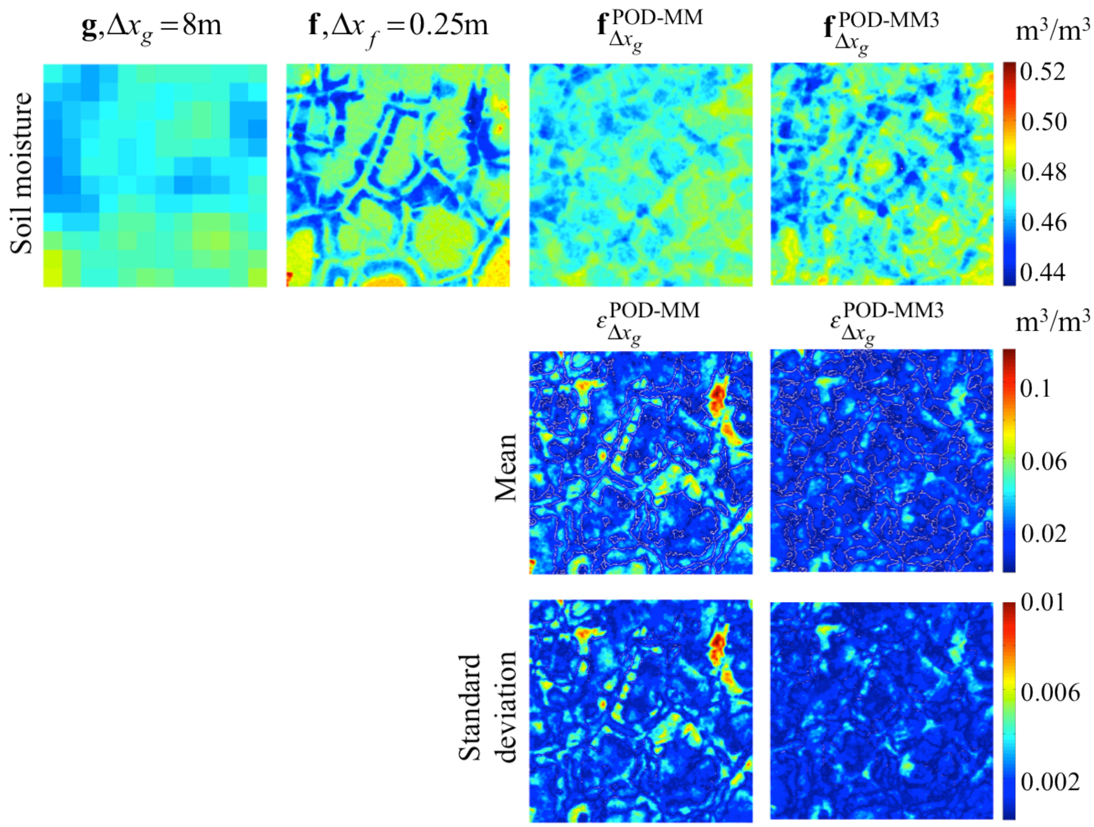

Figure 17. The top row shows the 20-25 cm layer solutions of $\mathbf{g}$, $\mathbf{f}$, $\mathbf{f}_{\Delta x_g}^{\text{POD-MM}}$ and $\mathbf{f}_{\Delta x_g}^{\text{POD-MM3}}$ for day 20 of year 1998 at site A for which $e_{\Delta x_g}^{\text{POD-MM3}}$ is approximately equal to $\overline{e}_{\Delta x_g}^{\text{POD-MM3}}$; $\Delta x_g = 8$ m. The second and third rows show the mean and standard deviation of $\boldsymbol{\varepsilon}_{\Delta x_g}^{\text{POD-MM}}$ and $\boldsymbol{\varepsilon}_{\Delta x_g}^{\text{POD-MM3}}$ for the 20-25 cm layer computed over 1998, 1999, 2000, 2002, and 2006.

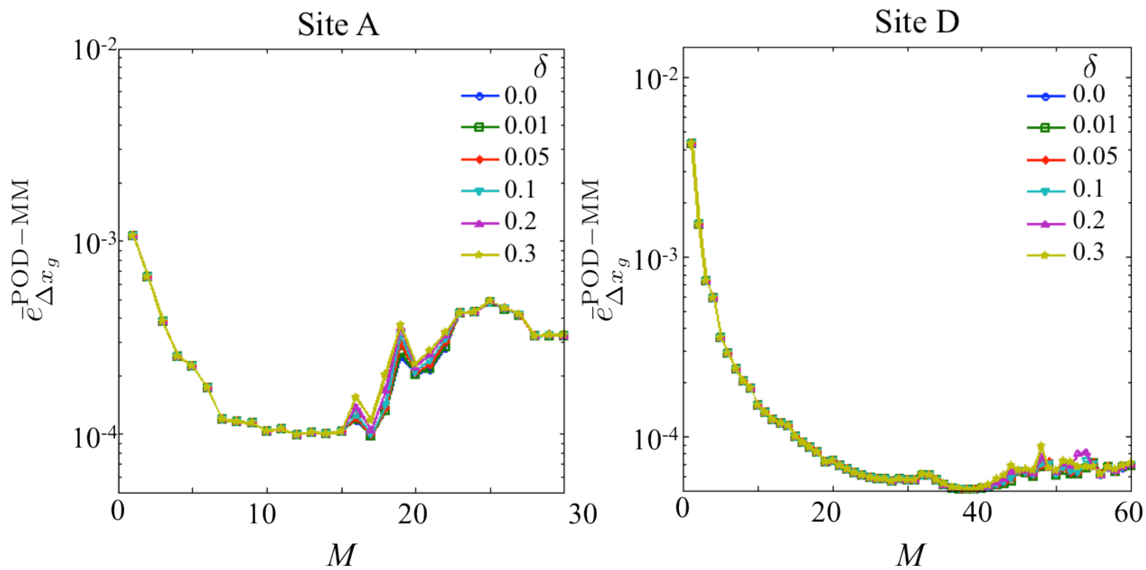

Figure 18. The variation of mean POD-MM error ($\bar{e}_{\Delta x_g}^{\text{POD-MM}}$) with respect to M for different δ and for $\Delta x_g = 8\,\mathrm{m}$; $\delta = 0$ is the reference case where there is no bias. Results are shown for sites A and D for year 2006.