

# Generation of soil moisture patterns at the catchment scale by EOF interpolation

M. A. Perry, J. D. Niemann

▶ **To cite this version:**

M. A. Perry, J. D. Niemann. Generation of soil moisture patterns at the catchment scale by EOF interpolation. Hydrology and Earth System Sciences Discussions, 2007, 4 (5), pp.2837-2874. ⟨hal-00298876⟩

**HAL Id: hal-00298876**

**https://hal.science/hal-00298876v1**

Submitted on 18 Jun 2008

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Hydrol. Earth Syst. Sci. Discuss., 4, 2837–2874, 2007
www.hydrol-earth-syst-sci-discuss.net/4/2837/2007/
© Author(s) 2007. This work is licensed
under a Creative Commons License.

Hydrology and Earth System Sciences Discussions

Papers published in *Hydrology and Earth System Sciences Discussions* are under open-access review for the journal *Hydrology and Earth System Sciences*

# Generation of soil moisture patterns at the catchment scale by EOF interpolation

**M. A. Perry**[1] **and J. D. Niemann**[2]

[1]Colorado Division of Water Resources, Pueblo, CO 81004, USA
[2]Department of Civil and Environmental Engineering, Colorado State University, Fort Collins, CO 80523, USA

Received: 21 August 2007 – Accepted: 21 August 2007 – Published: 7 September 2007

Correspondence to: J. D. Niemann (jniemann@engr.colostate.edu)

## Abstract

Spatial patterns of soil moisture cannot be adequately characterized by direct measurement for most practical applications, so interpolation between observations is required. Interpolation of soil moisture is complicated because multiple hydrologic processes can affect soil moisture and these processes can introduce distinct modes of variation into the soil moisture patterns. In this paper, a new method to interpolate soil moisture data is presented. This method accepts a dataset of soil moisture at widely-spaced locations on multiple dates and produces fine-scale patterns of soil moisture on the same dates. The method first uses Empirical Orthogonal Function (EOF) analysis to decompose the dataset into a set of time-invariant patterns of covariation (EOFs) and a set of associated time series (called expansion coefficients or ECs) that indicate the importance of the patterns on each date. The method then uses a statistical test to retain only the most important EOFs, and these EOFs are interpolated to the desired resolution using a standard estimation or interpolation method. The interpolated EOFs are finally combined with the spatial averages and the ECs to construct the fine-scale soil moisture patterns. Using the Tarrawarra dataset, the EOF-based interpolation method is shown to outperform analogous direct interpolation methods, and this improved performance is observed when as few as two observation dates are available. The improved performance occurs because EOF analysis decomposes soil moisture roughly according to the controlling processes and the most important EOFs exhibit distinct but more consistent spatial structures than soil moisture itself. Less predictable variation is also separated into higher order EOFs, which are discarded by the method.

## 1 Introduction

Spatial variability of soil moisture is important because it affects agriculture (Jaynes et al., 2003; Green and Erskine, 2004), climate (Delworth and Manabe, 1988; Liu, 2003), ecology (Moore et al., 1993), and hydrology (Bárdossy and Lehmann, 1998; Western et



HESSD

4, 2837–2874, 2007

al., 1999b). Processes like infiltration, evapotranspiration, vegetation growth, and energy balance are non-linearly related to soil moisture. For example, Wood (1997, 1999) showed that estimates of evapotranspiration based on spatially-averaged soil moisture tend to be too high when atmospheric demand is low and too low when atmospheric demand is high. Likewise, Jaynes et al. (2003) documented the complex response of crop yields to the spatial variations of soil moisture at the field scale. Crop growth can cease when locations are saturated or very dry, so using spatial average soil moisture can lead to poor estimates of crop yields for a field. Spatial organization of soil moisture has been shown to be important in predicting runoff at the catchment scale (Fitzjohn et al., 1998; Western et al., 1999c, 2001). Catchments have been observed to produce more discharge and erosion if areas of high soil moisture are well-connected to the channels.

Unfortunately, soil moisture measurements are not typically available at the spatial resolutions that are adequate to capture the variability that impacts these applications. Remote sensing techniques have great potential for measuring soil moisture (Bras, 1999; Entekhabi et al., 2004), but they typically observe the average soil moisture very close to the ground surface and over large geographical areas. Downscaling methods have been proposed to disaggregate remote sensing observations and produce realistic variations of soil moisture at finer spatial resolutions (Charpentier and Groffman, 1992; Hu et al., 1998; Kim and Barros, 2002). However, most disaggregation methods aim to produce realistic variability rather and accurate estimates at any given location. Limitations are also confronted when using ground-based measurements of soil moisture. Collection of ground-based data is labor-intensive and expensive, and these methods typically measure average soil moisture over a horizontal radius on the order of centimeters, making them essentially point measurements. Soil moisture can vary significantly between such measurements due to variations in soil, vegetation, and topographic characteristics (Seyfried and Wilcox, 1995; Western and Grayson, 1998; Lin et al., 2006a).



HESSD
4, 2837–2874, 2007

Numerous researchers have used process-based models and interpolation methods to estimate soil moisture patterns. Downer and Ogden (2003) used a distributed hydrologic model called GSSHA to estimate soil moisture patterns and compared the estimated patterns to point observations. Pellenq et al. (2003) coupled a soil vegetation atmospheric transfer (SVAT) model with Topmodel (Beven and Kirkby, 1979) to estimate soil moisture patterns that result from evapotranspiration and lateral redistribution of soil water within catchments. The use of numerical models to estimate soil moisture is limited by the fact that they require knowledge of a large number of spatially-distributed properties to realize their full potential and they require calibration based on past soil moisture or streamflow measurements. Interpolation methods are typically much less data-intensive in their application. In this approach, soil moisture values are estimated between sparse observations using spatial relationships to the observation points and/or correlations to other properties observed at a finer resolution. Bárdossy and Lehmann (1998) interpolated soil moisture at the catchment scale using variations of kriging. Thattai and Islam (2000) used kriging to show that remotely-sensed soil moisture from widely-spaced flight paths could be interpolated to produce a full soil moisture pattern. Wilson et al. (2005) estimated soil moisture patterns using a dynamic multiple linear regression that links the spatial variations of soil moisture to topographic attributes.

Estimation of soil moisture has had mixed results in the past in part because soil moisture patterns can exhibit different statistical characteristics at different times (Beven and Kirkby, 1979; Burt and Butcher, 1985; Western et al., 1999a; Mohanty et al., 2000). This tendency occurs because soil moisture patterns arise from the interaction of multiple hydrologic processes. At the catchment-scale, Grayson et al. (1997b) argued that two different processes combine to control much of the variation in soil moisture over seasonal time scales. One process is the lateral redistribution of surface and subsurface water, which is important when the soil is relatively wet. The other process is evapotranspiration, which controls the soil moisture pattern when the soil is relatively dry. These two processes tend to promote very different patterns of soil

moisture (Seyfried and Wilcox, 1995). Kachanoski and De Jong (1988) observed differences in soil moisture patterns at particular scales between different times. They found differences in the spectral densities of soil moisture on dry and wet days, which they attributed to the time varying roles of different hydrologic processes that act at distinct spatial scales. Such dynamic and multi-scale variations in soil moisture patterns make interpolation difficult (Western et al., 1999a; Florinsky et al., 2002).

One solution to this difficulty is the decomposition of soil moisture variation into underlying, time invariant patterns, which can be done with Empirical Orthogonal Function (EOF) analysis (Hu et al., 1998; Wilson et al., 2004). EOF analysis can decompose space-time datasets into a series of spatial patterns of underlying orthogonal covariation and associated time series that indicate the importance of each spatial pattern at each time. Kim and Barros (2002) and Jawson and Niemann (2007) used EOF analysis to decompose remotely-sensed soil moisture images from the SGP97 field campaign. Yoo and Kim (2004) used EOF analysis to decompose space-time patterns of ground-based soil moisture measurements for agricultural fields at the same site. The patterns of covariation that they identified are related to both topographic and soil properties. They identified a pattern of variation that was related to wet periods and another pattern related to dry periods, and they showed the time evolution of their relative importance. Perry and Niemann (2007) applied EOF analysis to the Tarrawarra catchment in Australia, where previous studies had shown that the soil moisture patterns depend on different topographic properties at different times (Western and Grayson, 1998; Western et al., 1999a). Perry and Niemann (2007) found that the most important pattern of covariation shows a clear dependence on hillslope and valley topography and is most important during wet periods. The second most important pattern of covariation exhibits strong aspect dependence, which is correlated with patterns of solar insolation and possibly evapotranspiration. They also developed a soil moisture forecasting method in which an EOF analysis of past data is used to forecast spatial patterns of soil moisture from the spatial average. Their method could also be used for downscaling of soil moisture.

Here, we consider the problem of interpolating a dataset of point soil moisture observations. Because previous studies have shown that EOF analysis is effective at decomposing soil moisture into distinct and physically meaningful modes of variation, we hypothesize that improved interpolation can be accomplished by decomposing the space-time variability of soil moisture into the patterns of covariation, interpolating those patterns separately, and then reassembling the interpolated patterns to determine the fine-scale soil moisture. In Sect. 2, we describe this proposed interpolation method in detail. In Sect. 3, we describe a dataset and study site where we test the method. In Sect. 4, we evaluate the performance of the proposed method and compare it to several traditional interpolation methods. In Sect. 5, we discuss the results, and in Sect. 6, we state our main conclusions.

## 2 Method

The problem we consider is outlined as follows. We assume that a set of widely-spaced point measurements of soil moisture are available at the same locations for at least two observation times (e.g., at least two days). This is a typical scenario when soil moisture is measured by permanent in-situ probes. In addition, auxiliary information such as topographic elevations or soil characteristics may or may not be available at a higher spatial resolution than the soil moisture measurements for the site. Our objective is to generate fine-scale patterns of soil moisture from the widely-space measurement and the auxiliary information, if available. We propose to estimate the fine-scale patterns using the following four-step procedure. First, the space-time dataset of soil moisture will be decomposed into the spatial average at each time, a set of spatial patterns of covariation (EOFs), and a set of time series that indicate the importance of each EOF to the soil moisture variation at each time. The time series are called the expansion coefficients (ECs). Second, the EOFs that identify statistically significant patterns of covariation based on a statistical test will be retained and the remaining EOFs will be discarded. Third, each retained EOF will be interpolated to the desired spatial

resolution using a standard interpolation method. Fourth, the interpolated EOFs will be combined with the original ECs and spatial averages to construct the fine-scale patterns of soil moisture. The remainder of this section describes these four steps in more detail.

The first step of the proposed technique is to perform an EOF decomposition on the sparse soil moisture data. Detailed mathematical treatment of EOF analysis is given in texts on the topic (Preisendorfer, 1988; Jolliffe, 2002; Jackson, 2003). Here, we briefly summarize its application in the soil moisture interpolation method. The analysis begins with a matrix of the space-time soil moisture observations:

$$S = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1n} \\ s_{21} & s_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ s_{m1} & \cdots & \cdots & s_{mn} \end{bmatrix} \quad (1)$$

where $s_{ij}$ corresponds to soil moisture at location $i$ and time $j$. $m$ is the number of sample locations, and $n$ is the number of sample times. Each row $i$ in matrix $S$ corresponds to a particular location $(x_i, y_i)$. Note that we use capital letters to denote matrices and lowercase letters to denote scalars. Next, the matrix of spatial anomalies of soil moisture $Z$ is computed from the original data by subtracting the spatial average for each time from all measurements at that time. Specifically,

$$z_{ij} = s_{ij} - \frac{1}{m}\sum_{k=1}^{m} s_{kj} \quad (2)$$

where $z_{ij}$ is the spatial anomaly at location $i$ and time $j$. The temporal covariance matrix $V$ $(n \times n)$ is then computed as:

$$V = \frac{1}{m} Z^T Z \quad (3)$$

where $T$ indicates the matrix transpose. In general, $V$ is not a diagonal matrix, that is, some covariance is observed between the soil moisture anomalies at different times. However, $V$ can be diagonalized using EOF or eigenanalysis. The results of this procedure are the diagonal matrix $L$ ($n \times n$), whose diagonal components are the eigenvalues of $V$, and a matrix $E$ ($n \times n$) that contains the eigenvectors as columns. Together these matrices satisfy:

$$VE = LE. \tag{4}$$

The eigenvectors define a new basis or coordinate system for the soil moisture data. The first eigenvector is oriented in the direction of maximum covariation, the second eigenvector is oriented perpendicular to the first eigenvector in the direction of maximum residual variation, and so on. The eigenvalues indicate the amount of covariation (in the original basis) that lies in the direction of each eigenvector. The transformation $E$ can be applied to the soil moisture anomalies to obtain $n$ new spatial patterns called EOFs. Mathematically, this is:

$$F = ZE \tag{5}$$

where $F$ is an $m \times n$ matrix containing the EOFs as columns. We define EOF$k$ as the values of the data projected onto the $k$th axis, so EOF$k$ is found in the $k$th column of $F$ in Eq. (5). Notice that EOF1 is associated with the largest eigenvalue, so it is associated with the axis that explains the most variation. One can think of the EOFs as patterns of covariation that are imbedded in the soil moisture anomalies. We define the $k$th expansion coefficient (EC$k$) as the unit vector of the $k$th axis, which can be found in the $k$th column of $E$ in Eq. (5). The expansion coefficients are time series that indicate the importance of the EOFs to the individual soil moisture patterns.

After the EOF decomposition is completed, the second step is to retain the subset of EOFs that are statistically significant. It is important to note that almost any space-time dataset can be decomposed using EOF analysis irrespective of whether statistically significant covariation is observed between the patterns at different observation times.

In general, the lower order EOFs are associated with stronger patterns of covariation and thus are more likely to be associated with true system variations. The higher order EOFs are associated with weaker covariation and thus are often associated with instrument error (Jackson, 2003). Including these higher order EOFs in the soil moisture interpolation method would introduce noise into the soil moisture patterns and likely increase the estimation errors. Numerous methods are available to judge whether EOFs represent statistically significant patterns of covariation, and unfortunately, these different methods can give rather different results. Because no method is clearly preferred from a theoretical basis, we propose calculating the number of significant EOFs using two different methods and averaging their results to determine the number of EOFs to retain. This approach will be evaluated later in the paper. The first of the two methods was proposed by Bartlett (1950). It assesses whether the EOFs describe statistically significant patterns of covariation by testing the hypothesis that the eigenvalues of the last $(n-d)$ EOFs are all equal. The relevant statistic for this test is $\chi^2_{\text{crit}}$, which is calculated:

$$\chi^2_{\text{crit}} = -(m-1)\sum_{j=d+1}^{n} \ln(l_j) + (m-1)(n-d)\ln\left[\frac{\sum_{j=d+1}^{n} l_j}{(n-d)}\right]. \quad (6)$$

Bartlett (1950) showed that $\chi^2_{\text{crit}}$ has a chi-squared distribution with $(1/2)(n-d-1)(n-d+2)$ degrees of freedom. If $\chi^2_{\text{crit}}$ is less than or equal to the standard tabulated $\chi^2$ variate, then the hypothesis cannot be rejected at the selected confidence level. In this case, the last $(n-d)$ EOFs would be discarded, and the first $d$ EOFs would be considered statistically significant (Jackson, 2003). The second test is presented in Johnson and Wichern (2002) and is based on Gaussian confidences limits for the eigenvalues. Assuming Gaussian random errors about the eigenvalues,

the confidence limits for the eigenvalue $l_{jj}$ can be written as:

$$\frac{l_{jj}}{1 \pm z_{1-\alpha}\sqrt{2/m}} \tag{7}$$

where $z_{1-\alpha}$ is the standard normal variate at the $1-\alpha$ confidence level. If the confidence limits of the eigenvalue $l_{jj}$ do not overlap with those for the next higher order eigenvalue, then $l_{jj}$ would be considered statistically significant. This argument is based on the observation that a dataset without significant covariation tends to exhibit the same amount of variation in all directions. As a result, the EOF coordinate transformations are not unique and the eigenvalues are expected to be nearly the same.

The third step of the proposed method is to interpolate the retained EOFs. The EOFs are defined only at the observation points, so interpolation is used to estimate values between the observation points. EOFs can be interpolated using any standard interpolation technique. In this paper, we demonstrate three methods: (1) multiple linear regression (MLR) against topographic attributes, (2) inverse distance weighting (IDW), and (3) ordinary kriging. Note that multiple linear regression is more correctly described as spatial estimation rather than interpolation because it does not use distance to determine the estimates. However, the term interpolation is used in reference to all of three methods to simplify the discussion. The MLR approach assumes that auxiliary data are available at the final spatial resolution for the soil moisture patterns. Such data might include soil characteristics or topographic attributes derived from elevation data (e.g., slope, curvature, etc.). In our application, topographic attributes are used based on previously-observed correlations with soil moisture data (Western et al., 1999a; Wilson et al., 2005). The MLR approach uses a stepwise partial correlation analysis to build a parsimonious model for each sparse EOF in terms of the topographic attributes (Salas et al., in development). In the partial correlation analysis, the marginal increase in the multiple correlation is checked at each step and only variables that produce a statistically significant improvement are added to the MLR. The MLR then estimates

the EOF value at an unobserved location using:

$$\hat{f}_{ik} = a_k + \sum_{j=1}^{l} T_{ij} b_{jk} \quad (8)$$

where $\hat{f}_{ik}$is the estimated $k$th order EOF at location $i$, $a_k$ and $b_{jk}$ are the parameters determined from the regression of the sparse $k$th order EOF against topographic attribute $T_j$, and $l$ is the total number of topographic attributes included in the MLR. Note that the topographic attributes included in the regression and the parameter values can differ for each EOF. The second interpolation technique is the IDW method, which assumes that the EOF values can be determined based on their proximity to observations. This method determines $\hat{f}_{ik}$ as:

$$\hat{f}_{ik} = \sum_{j=1}^{l} w_j f_{jk} \quad (9)$$

where $f_{jk}$ is the value of the $k$th order EOF at sparse sample location $j$ and $l$ is the number of neighboring observations used in the interpolation. In this paper, we use $l$=5, but the results are not sensitive to this choice. $w_j$ is the weight applied to each neighbor and is computed as:

$$w_j = \frac{h_j^{-2}}{\sum_{ll=1}^{l} h_{ll}^{-2}} \quad (10)$$

where $h_j$ is the horizontal distance between unobserved location $i$ and each neighboring observation $j$. The third interpolation technique used here is ordinary kriging (Cressie, 1991a). Kriging assumes that the soil moisture at an unobserved location is related to observations at neighboring locations according to an autocovariance that is



HESSD
4, 2837–2874, 2007



EGU

computed as a function of separation distance. Like IDW, kriging uses Eq. (9) to interpolate the soil moisture, but the weights $w_j$ are determined by solving the following system of equations:

$$\sum_{j=1}^{l} w_j \gamma_{rj}(h_{rj}) + \mu = \gamma_{ri}(h_{ri}) \quad r = 1, \ldots, l \tag{11}$$

$$\sum_{j=1}^{l} w_j = 1 \tag{12}$$

where $h_{rj}$ is horizontal distance between sparse measurement locations $r$ and $j$, $\gamma_{rj}$ is a model semi-variance at distance $h_{rj}$, $h_{ri}$ is the distance between sparse measurement location $r$ and the unobserved location $i$, and $\mu$ is a Lagrange-multiplier (Bárdossy and Lehmann, 1998).

The fourth and final step of the proposed method is to estimate soil moisture over the entire spatial domain using the interpolated EOFs along with the spatial averages and ECs from the original sparse measurements. By doing this, we assume that the spatial average from the sparse measurements is a suitable estimate for the spatial average at all locations, which implicitly assumes stationarity. Similarly, we assume that the ECs estimated from the sparse measurements are suitable estimates for the ECs at all locations. Recall that the ECs are spatially invariant and in theory apply to all locations within a stationary field. The soil moisture at any location of interest can be developed based on Eq. (5). In particular:

$$\hat{s}_{ij} = \bar{s}_j + \sum_{k=1}^{d} \hat{f}_{ik} e^T_{kj} \qquad \text{for} \qquad i=1, \ldots, m j=1, \ldots, n, \tag{13}$$

where $\hat{s}_{ij}$ is estimated soil moisture at location $i$ and time $j$, $\bar{s}_j$ is the spatial average of the sparse soil moisture measurements at time $j$, $\hat{f}_{ik}$ is the interpolated value of the $k$th order EOF at location $i$, $d$ is the number of EOFs considered statistically significant, and $e_{kj}$ is the $k$th order EC determined from the sparse measurements.

## 3 Study site

The proposed EOF-based interpolation method is tested by applying it to the well-known Tarrawarra Catchment soil moisture dataset. The Tarrawarra Catchment is located in southeast Australia (Western and Grayson, 1998). The site consists of two small valleys and surrounding hillslopes (Fig. 1). No channels occur within the site, and the land-use is pasture. Unfortunately, detailed soil information is not available at the same sample spacing as soil moisture observations. However, soils are fairly uniform over the site and consist of silty loam A horizon overlying a clay B horizon. The A horizon is approximately 15–40 cm deep, and saturated conditions form in the A horizon during wet periods. Annual precipitation is approximately 820 mm, and annual potential evapotranspiration (PET) is about 830 mm. A wet season occurs between April and September (austral winter) when precipitation exceeds PET, and a dry season occurs between October and March when PET exceeds precipitation (Grayson et al., 1997b; Western and Grayson, 1998; Kandel et al., 2004).

The soil moisture measurements were collected using time domain reflectometry (TDR) and are publicly available at the Global Soil Moisture Databank (Robock et al., 2000). Observations were collected on 13 dates spanning approximately 14 months, and they capture the normal range of seasonal soil moisture conditions (Western and Grayson, 1998). The TDR readings were taken on a 10 m×20 m grid, where the 10 m spacing is roughly in the north-south direction. Generally about 500 measurements were collected per sampling date. A typical set of sample locations are shown in Fig. 1. The TDR measurements report the average soil moisture in volume of water per volume of soil for a cylinder of soil approximately 30 cm deep and 5 cm in diameter. Thus, in comparison to the size of the catchment, the measurements provide essentially point soil moisture values. The exact TDR sampling locations varied slightly from day to day, so we interpolated the original TDR data onto a common 20 m×10 m grid using the cubic spline method. The purpose of this interpolation is simply to translate the original data from different days onto a common grid, which is necessary for the application

of the method, not to increase the spatial resolution of the data. In nearly all cases, the original TDR measurements were taken within 0.5 m of the final grid points, so the cubic-spline interpolation caused very little change in the patterns. Finally, the dataset for each sampling date was clipped to include only the aerial extent common to all sampling dates.

A detailed topographic survey is available for the catchment with point elevations available at approximately 10 m spacing. In order to facilitate comparisons with soil moisture, the elevation data were interpolated to a 10 m×10 m grid where every other column is coincident with the 20 m×10 m soil moisture grid.

## 4 Results

To test the proposed method, the 10 m×20 m soil moisture measurements from Tarrawarra are first sampled to produce a subset of data at a 30 m×60 m spacing. These data are considered observation points, while the remaining data are considered verification points. The observations include 52 soil moisture values on 13 dates or 11% of the available soil moisture data. These observations will be used to estimate soil moisture at 407 other locations. Later in this section, we will consider observation sets with closer and wider spacings.

The first step is to perform an EOF decomposition of the observations to determine the sparse EOFs and associated ECs. The assumption underlying this step is that the EOFs and ECs obtained from the sparse observations are good estimates of the EOFs and ECs for the full soil moisture pattern. The solid lines in Fig. 2 show the first two EOFs and associated ECs determined from the sparse observations, and the dashed lines show the EOF and EC values at the same locations when the EOF analysis is performed using the entire dataset. The figure shows that the EOF1 values from the full dataset are reproduced almost exactly by an EOF analysis of the sparse data (correlation coefficient, $r$=1.00). The EC1 values from the full dataset are also well reproduced by an EOF analysis of the sparse data ($r$=0.91). EOF2 and EC2 are

reproduced fairly well ($r$=0.75 and $r$=0.69, respectively), but the pattern of covariance identified by EOF2 is more difficult to distinguish from other variation when the observations are sparse. In general, it is expected that higher order EOFs are increasingly difficult to determine from sparse data, which justifies their exclusion from the interpolation method. However, the figure confirms that the most important EOFs and ECs can be identified from sparse observations.

Next, we evaluate the method to determine the number of EOFs that should be retained. Figure 3 shows the results of the tests described by Bartlett (1950) and Johnson and Wichern (2002). At the 95% confidence level, the Bartlett test indicates that the first five EOFs are significant, and the Johnson and Wichern test indicates that only the first EOF is significant. Averaging the two numbers suggests that we retain and interpolate 3 EOFs. To assess this element of the interpolation method, we determined the number of EOFs that should be retained to optimize the performance of the EOF-based interpolation method. The optimum number was determined by retaining every possible number of EOFs from a minimum of 1 to the maximum of 13. In each case, the retained EOFs were interpolated to produce fine scale soil moisture patterns, which were then compared to the observed patterns. For data at the 30 m×60 m spacing, the optimum number of retained EOFs is three. At other spacings, the optimum number is generally quite similar to the number of EOFs retained by the proposed method. It should be noted that this method may not perform as well with other datasets.

The next step is to interpolate the retained EOFs. The following topographic attributes were considered: elevation, slope, slope$^{-1}$, vertical topographic curvature ($k_v$), horizontal topographic curvature ($k_h$), specific drainage area ($a$), the natural log of the specific drainage area ln($a$), wetness index (WI) (Beven and Kirkby, 1979), potential solar radiation index (PSRI) (Moore et al., 1993), and the lowness index (L) (Roberts et al., 1997). These particular topographic attributes were selected based on results of previous studies relating soil moisture and topography (Western et al., 1999a; Florinsky et al., 2002; Lin et al., 2006b). Figure 4a–c shows the first three EOFs computed from the full soil moisture dataset. Figure 4d–f shows the first three EOFs computed

from the sparse observation points, and Fig. 4g–i shows the EOFs interpolated from the sparse observation points. Notice that the EOF values at the observation points are retained in lieu of using the MLR, which gives the speckled appearance to the interpolated EOFs. While much of the local variation in the EOF patterns shown in Fig. 4a–c is not reproduced by the interpolation, the large scale features are largely captured by the interpolation method. Linear correlations coefficients between the interpolated values and observed values are 0.77, 0.25 and 0.15 for EOF1, EOF2 and EOF3, respectively.

Finally, the soil moisture patterns can be generated by combining the interpolated EOFs with the observed ECs and spatial averages. Notice that the EOF-based interpolation method simultaneously generates soil moisture patterns on all 13 observation dates. Figure 5a–b shows the actual soil moisture patterns on two observation dates: 3 July 1996 and 20 September 1996. Comparing these two patterns, one observes that the soil moisture pattern on 20 September 1996 exhibits a stronger dependence on the topography. Wet locations tend to be located in the valley bottoms while dry locations are more common on the hillsides. On 3 July, this tendency is somewhat weaker. Figure 5c–d shows the soil moisture patterns generated by the EOF-based interpolation technique. As expected, the generated soil moisture patterns exhibit much less local variability than the actual observations, but the large scale features are reproduced in the patterns. In particular, the wet locations are more consistently located in the valley bottom on 20 September than 3 July. The differences between these dates arise from the time-varying weights on the EOFs (i.e., the ECs) that are estimated from the sparse observations. Figure 5 also shows the estimation error for both dates. The associated error patterns appear to be mainly local variations. The magnitudes of the errors are not insignificant, which implies that a substantial portion of the soil moisture variation is not captured by the interpolation method.

In order to generalize these results and compare the EOF-based interpolation method to standard interpolation methods, we developed 54 different observation sets from the Tarrawarra data. The data were resampled at four spacings: 20 m×40 m, 30 m×60 m, 40 m×80 m, and 50 m×100 m. For each spacing, multiple observation sets



HESSD
4, 2837–2874, 2007

were developed by translating the grid of observation points. For example, in the case of the 20 m×40 m grid, every other data point in each cardinal direction is treated as an observation. The 20 m×40 m grid of observations can be shifted 10 m to the right to obtain a different set of observations. Similarly, it can be shifted down and diagonally to obtain two more observation sets. Table 1 shows the number of possible observation sets and other characteristics for each sample spacing. For each observation set, the EOF-based interpolation method was performed using MLR, IDW, and Kriging to interpolate the EOFs. We refer to these as the EOF-MLR, EOF-IDW, and EOF-Kriging methods. For comparison, these three methods were also used to interpolate the soil moisture patterns directly. We refer to these as the MLR, IDW, and Kriging methods.

For each method and observation set, the interpolation performance was measured using the average Nash Sutcliffe Coefficient of Efficiency (NSCE) for the 13 days in the dataset, which is computed as:

$$\mathrm{NSCE}=\frac{1}{n}\sum_{i=1}^{n}\frac{\sigma^2_{\mathrm{obs},i}-\sigma^2_{\varepsilon,i}}{\sigma^2_{\mathrm{obs},i}} \tag{14}$$

where $\sigma^2_{\mathrm{obs},i}$ is the variance of the measured soil moisture at the verification locations on day $i$ and $\sigma^2_{\varepsilon,i}$ is the mean squared difference between measured and estimated soil moisture at the verification locations on day $i$. The term in the summation is the usual definition of the NSCE (Nash and Sutcliffe, 1970). The maximum possible value of the NSCE on any particular day is one, which would imply that the interpolated surface reproduces the unobserved values exactly. If the NSCE is zero, it indicates that the interpolation surface has the same error at the verification points as the spatial average of the observations would.

Figure 6a shows the average NSCE for the generated soil moisture patterns as a function of the spacing of the observations and the interpolation technique used. For each spacing, the symbol indicates the average NSCE for all observation sets. The error bars identify +/− one standard deviation to indicate the variation of the NSCE results



HESSD
4, 2837–2874, 2007

between observation sets at that spacing. Figure 6a shows that the performance of all of the interpolation methods decreases as the sample spacing increases. If the sample spacing exceeds 50 m×100 m, the interpolation methods typically do not perform better than using the spatial average of the observations. At the 20 m×40 m sample spacing, the EOF-IDW and EOF-Kriging interpolation methods outperform the EOF-MLR method. At this scale, neighboring soil moisture measurements contain more information about soil moisture than the local topographic attributes do. However, this behavior tends to reverse above the 30 m×60 m spacing as the information contained in neighboring soil moisture measurements decreases. Furthermore, the assumption that topographic data is available at high resolution means that the MLR technique can use more information than the IDW technique.

Figure 6b shows the difference between the NSCE computed using the EOF-based interpolation methods and the NSCE computed using the analogous direct soil moisture interpolation methods. Again, the symbols in Fig. 6b show the average difference in NSCE over all possible observation sets at a given spacing. When the symbol is above zero, it suggests that the EOF-based interpolation method outperforms the analogous direct interpolation methods on average. This is almost always the case. The NSCE differences in Fig. 6b appear to be small, but they are significant relative to the magnitudes of NSCE in Fig. 6a. The EOF-based method offers from around 3 percent to greater than 60 percent improvement over the direct interpolation method, depending on the spacing and specific interpolation method that is used. The figure also shows error bars, which identify +/−1 standard deviation. When the lower end of the error bar exceeds zero, it indicates that the EOF-based method outperforms the direct method more than 80% of the time. Again, this is usually the case. A traditional $t$-test cannot be used to evaluate the significance of the difference between the methods because the samples (i.e. sparse sample grid realizations) are not independent. However, the figure provides some confidence that the EOF-based method usually outperforms standard interpolation methods.

## 5 Discussion

In this section, we investigate why EOF-based interpolation outperforms direct soil moisture interpolation. One key difference between the two methods is the amount of information used. The EOF-based method uses sparse soil moisture observations on all 13 sampling dates to compute EOFs, which are then interpolated to produce the fine-scale soil moisture patterns on all 13 dates simultaneously. The direct interpolation methods determine the soil moisture values based only on the observations from the same day. To assess the role that the additional data plays, Fig. 7 shows the NSCE of the generated soil moisture patterns for one sampling date (28 March) as an increasing number of dates is included in the dataset for the direct MLR and EOF-MLR methods. For simplicity, successive dates are added in chronological order. For the direct MLR method, data from each additional date were standardized by subtracting the spatial average and dividing by the standard deviation and then combined into one dataset for the regression analysis. This standardization improves the performance of the direct MLR method. Both the direct MLR and EOF-MLR methods yield identical results when only one date is used because the only EOF pattern is the soil moisture pattern itself. When two sample dates are used in each method, the NSCE for the EOF-MLR method immediately doubles from 0.075 to about 0.15, while the NSCE for the direct MLR method actually decreases. The NSCE in both cases does not change significantly with the addition of more observation dates. The analysis was repeated numerous times by adding the successive dates in different orders, and the same general tendency was observed. This analysis shows that the EOF-MLR method outperforms direct interpolation even when the data requirements are forced to be the same. These results also suggest that the EOF-based method can produce improved results as soon as two sampling dates are used.

One reason for the improved performance of the EOF-based method can be seen by examining the correlations to the topographic attributes. Table 2 shows the correlation coefficients between the EOFs and topographic attributes and the correlation



HESSD
4, 2837–2874, 2007

coefficients between the individual soil moisture patterns and the same topographic attributes when all of the data are included. Blank entries in Table 2 indicate correlations that are not statistically significant at the 95% confidence level. Three interesting observations can be made from Table 2. First, the correlations support previous physical interpretations of the EOFs (Perry and Niemann, 2007). EOF1 has particularly high correlations with $k_h$, ln($a$), WI, and L, and this collection of attributes has been related to lateral redistribution of soil water by several authors (Burt and Butcher, 1985; Western et al., 1999a; Florinsky et al., 2002). EOF2 has a large correlation with PSRI, which supports its association with evapotranspiration. EOF3 exhibits a large negative correlation with elevation, which may be associated with transition times when the assumption of steady-state flow that underlies WI is violated (Perry and Niemann, 2007; Barling et al., 1994; Grayson et al., 1997a). Second, Table 2 shows that the most important EOFs also tend to exhibit the strongest correlations to topographic attributes. The EOFs are almost perfectly sorted according to their multiple correlation coefficients with the topographic attributes (the far right column in the table). EOFs 1–3 have fairly high multiple correlation coefficients, while the remaining EOFs have lower values. Third, Table 2 shows that the most important EOFs have stronger correlations to certain topographic attributes than the individual soil moisture patterns do. For example, EOF1 is more highly correlated to $k_v$, $a$, ln($a$), WI, and L than any individual soil moisture pattern is. Similarly, EOF2 is more highly correlated to PSRI than any individual soil moisture pattern is. The closer association between the EOFs and the topographic attributes suggests that an EOF-based interpolation may be more efficient at using topographic information. Table 3 examines the correlation between the estimation errors from the EOF-MLR and direct MLR methods versus the topographic attributes. The estimation errors were computed for a single 30 m×60 m observation set at the verification locations on every sampling date, and the multiple correlation coefficients between the estimation errors and the topographic attributes were calculated. On 10 out of 13 dates, the residual errors from the EOF-MLR method have a lower correlation with the topographic attributes than do the errors from the direct

MLR method. This result also suggests that the EOF-MLR method accounts for the topographic influence on soil moisture better than the direct MLR method does.

The EOF-based method also performs better when coupled with distance-based interpolation methods like IDW and kriging. This result suggests that nearby values of the most important EOFs may provide more information about the unobserved values of those EOFs than nearby soil moisture values provide about unobserved soil moisture values. To evaluate the spatial structure of the EOFs and soil moisture patterns, the sample semivariogram $\gamma(h)$ is calculated as:

$$\hat{\gamma}(h) = \frac{1}{2m(h)} \sum_{m(h)} \left[z(x_i) - z(x_j)\right]^2 \quad (15)$$

where $z(x_i)$ is the value of the pattern at location $x_i$, and $x_i$ and $x_j$ belong to the set of $m(h)$ pairs separated by horizontal distance $h$ (Cressie, 1991b). The sample semivariogram can be fit with a theoretical or model semivariogram curve. In a previous geostatistical analysis of the Tarrawarra soil moisture data, Western and Grayson (1998) found that an exponential model described the data well. The exponential semivariogram $\gamma_e$ can be written as:

$$\gamma_e(h) = a + (b - a)[1 - \exp(-h/c)] \quad (16)$$

where $a$, $b$, and $c$ are model parameters. $a$ is the so-called nugget and is interpreted as the variance at zero separation distance. The nugget is usually viewed as the result of instrument error or variation at a smaller scale than the sample spacing. $b$ is the sill, which corresponds to the semivariogram's asymptotic upper limit, and is related to the sample variance. $c$ determines the rate of increase from $a$ to $b$. Western et al. (1998) define the range or the distance to the sill as $3c$ for the exponential model.

Sample semivariograms were computed for the EOFs and soil moisture patterns at the finest available resolution, and the exponential model was fit in each case. To facilitate comparison, all data were standardized and normalized by subtracting the mean



HESSD
4, 2837–2874, 2007



EGU

and dividing by the standard deviation prior to computing the semivariograms. The parameters of the model semivariogram are shown in Table 4. The differences in the values indicate that the EOFs exhibit distinct spatial behaviors. The EOF1 semivariogram has a zero nugget and relatively small range, which suggest that the autocorrelation in EOF1 occurs at relatively small spatial scales. EOF2 and EOF3 have a progressively larger nuggets and ranges.

Interpolating with Kriging implicitly assumes that the autocovariance can be described as a function of separation distance. One way to test the strength of the dependence of the autocovariance on separation distance is to calculate the root mean square error (RMSE) of the data about the sample semivariogram. The RMSE measures the spread of the original data about the average (i.e., sample) semi-variogram. Table 4 shows that lower RMSE values are observed for the most important EOFs than for the soil moisture patterns themselves. The more consistent spatial structure of the EOFs explains why kriging is more effective with the EOF patterns than the soil moisture patterns. Table 4 also suggests that much of the noise in the original soil moisture measurements is filtered into higher order EOFs. The Tarrawarra measurements are known to contain measurement errors associated with the TDR measuring device (Western and Grayson, 1998), and such measurement errors would be expected to produce a non-zero nugget (Bárdossy and Lehmann, 1998). The model semivariogram nugget for EOF1 is zero, which suggests that measurement errors have largely been removed from EOF1. The nugget tends to increase as one considers EOFs 2 through 6. As one considers higher order EOFs, the amount of variation explained by each EOF is smaller, so it becomes increasingly difficult to distinguish meaningful system variation from random noise.

## 6 Conclusions

A simple EOF-based interpolation method was proposed for generating soil moisture patterns from widely-spaced observations available at multiple times. This method

determines the soil moisture patterns by decomposing the dataset into EOFs, interpolating the EOFs using any standard interpolation method (MLR, IDW, or kriging), and using the interpolated EOFs to construct the fine-scale soil moisture patterns at all times. When applied to soil moisture data from the well-known Tarrawarra catchment, the following conclusions can be made:

- The proposed method routinely outperforms standard interpolation methods. One typically obtains soil moisture patterns with a closer similarity to the actual soil moisture patterns by interpolating the EOFs than by interpolating the soil moisture patterns directly.
- The proposed interpolation method can outperform the other methods when as few as two observation times are available. Repeated observations allow the EOF-based method to identify underlying patterns of covariation, while direct interpolation methods have difficulty efficiently utilizing data from multiple times.
- If the available observations are closely-space (less than 30 m×60 m), the proposed method produces better results when coupled with a distance-based method like IDW or kriging. When the observations are more widely-space, the method performs better using a MLR with topographic data. This general tendency is expected to hold for other datasets given the limited correlation lengths of soil moisture patterns. However, the efficiency of a MLR with topographic attributes or other site characteristics is expected to vary widely.
- The EOF-based interpolation method captures the dependence on topography more efficiently than direct interpolation methods. In particular, the estimation errors of soil moisture patterns generated by the EOF-MLR method typically have lower correlations with topographic attributes than the estimation errors of the soil moisture patterns generated directly by a MLR.
- More of the semi-variance of the EOFs is explained as a function of separation distance than that of the soil moisture patterns. The more consistent spatial struc-



HESSD
4, 2837–2874, 2007

ture of the EOFs allows distance-based interpolation methods to perform better on the EOFs than on the individual soil moisture patterns.

Overall, if data are available from more than one observation time, then interpolation of soil moisture data usually appears to be improved by conducting an EOF analysis and interpolating the most important EOFs instead of direct interpolation of the soil moisture observations.

*Acknowledgements.* We gratefully acknowledge financial support from the U.S. Army Research Office Terrestrial Sciences Program, and we thank T. Green for his helpful input.

## References

Bárdossy, A. and Lehmann, W.: Spatial distribution of soil moisture in a small catchment. Part 1: geostatistical analysis, J. Hydrol., 206, 1–15, 1998.

Barling, R. D., Moore, I. D., and Grayson, R. B.: A Quasi-Dynamic Wetness Index For Characterizing The Spatial-Distribution Of Zones Of Surface Saturation And Soil-Water Content, Water Resour. Res., 30(4), 1029–1044, 1994.

Bartlett, M. S.: Tests of Significance in Factor Analysis, Br. J. Psych. Stat. Sec. 3, 77–85, 1950.

Beven, K. J. and Kirkby, M. J.: A physically based variable contributing area model of basin hydrology, Hydrol. Sci., 24, 43–69, 1979.

Bras, R. L.: A brief history of hydrology, The Robert E. Horton Lecture, B. Am. Meteorol. Soc., 80(6), 1151–1164, 1999.

Burt, T. P. and Butcher, D. P.: Topographic controls of soil moisture distributions, J. Soil Sci., 36, 469–486, 1985.

Charpentier, M. A. and Groffman, P. M.: Soil moisture variability within remote sensing pixels, J. Geophys. Res., 97(D17), 18 987–18 995, 1992.

Cressie, N. A. C.: Statistics for Spatial Data, Johh Wiley & Sons, New York, 1991a.

Cressie, N. A. C.: Statistics for Spatial Data, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, New York, 1991b.

Delworth, T. L. and Manabe, S.: The influence of potential evaporation on the variabilities of simulated soil wetness and climate, J. Climate, 1, 523–547, 1988.

Downer, C. W. and Ogden, F. L.: Prediction of runoff and soil moistures at the watershed scale: Effects of model complexity and parameter assignment, Water Resour. Res., 39(3), 1–13, 2003.
Entekhabi, D., Njoku, E. G., Houser, P., Spencer, M., Doiron, T. S., J., Girard, R., Belair, S., Crow, W., Kerr, Y. H., Kimball, J. S., McDonald, K. C., O'Neill, P. E., Pultz, T., Running, S. W., Shi, J., Wood, E., and van Zyl, J.: The Hydrosphere State (Hydros) Satellite Mission: An earth system pathfinder for global mapping of soil moisture and land freeze/thaw, IEEE T. Geosci. Remote, 42, 2184–2195, 2004.
Fitzjohn, C., Ternan, J. L., and Williams, A. G.: Soil moisture variability in a semi-arid gully catchment: implications for runoff and erosion control, Catena, 32, 55–70, 1998.
Florinsky, I. V., Eilers, R. G., Manning, G. R., and Fuller, L. G.: Prediction of soil properties by digital terrain modelling, Environ. Modell. Softw., 17, 295–311, 2002.
Grayson, R. B., Western, A. W., Chiew, F. H. S., and Bloschl, G.: Preferred states in spatial soil moisture patterns: Local and nonlocal controls, Water Resour. Res., 33(12), 2897–2908, 1997a.
Grayson, R. B., Western, A. W., Chiew, H. S., and Blöschl, G.: Preferred states in spatial soil moisture patterns: Local and nonlocal controls, Water Resour. Res., 33(12), 2897–2908, 1997b.
Green, T. R. and Erskine, R. H.: Measurement, scaling, and topographic analysis of spatial crop yield and soil water content, Hydrol. Processes, 18, 1447–1465, 2004.
Hu, Z., Yizong, C., and Islam, S.: Multiscaling properties of soil moisture images and decomposition of large- and small-scale features using wavelet transforms, Int. J. Remote Sens., 19(13), 2451–2467, 1998.
Jackson, J. E.: A User's Guide to Principal Components, Wiley Series in Probability and Statistics, John Wiley & Sons, Hoboken, New Jersey, 2003.
Jawson, S. and Niemann, J. D.: Spatial patterns from EOF analysis of soil moisture at at large scale and their dependence on soil, land use, and topographic properties, Adv. Water Resour., 30(3), 366–381, 2007.
Jaynes, D. B., Kaspar, T. C., Colvin, T. S., and James, D. E.: Cluster analysis of spatiotemporal corn yield patterns in an Iowa field, Agron. J., 95, 574–586, 2003.
Johnson, R. A. and Wichern, D. W.: Applied Multivariate Statistical Analysis, Prentice Hall, Upper Saddle River, New Jersey, 2002.
Jolliffe, I. T.: Principal Component Analysis, Springer Series in Statistics, Springer, New York,

2002.
Kachanoski, R. G. and De Jong, E.: Scale dependence and the temporal persistence of spatial patterns of soil water storage, Water Resour. Res., 24(1), 85–91, 1988.
Kandel, D. D., Western, A. W., Grayson, R. B., and Turra, H. N.: Process parameterization and temporal scaling in surface runoff and erosion modelling, Hydrol. Processes, 18(8), 1423–1446, 2004.
Kim, G. and Barros, A. P.: Downscaling of remotely sensed soil moisture with a modified fractal interpolation method using contraction mapping and ancillary data, Remote Sens. Environ., 83, 400–413, 2002.
Lin, H. S., Kogelmann, W., Walker, C., and Bruns, M. A.: Soil moisture patterns in a forested catchment: A hydropedological perspective, Geoderma, 131, 345–368, 2006a.
Lin, H. S., Kogelmann, W., Walker, C., and Bruns, M. A.: Soil moisture patterns in a forested catchment: A hydropedological perspective, Geoderma, 131, 345–368, 2006b.
Liu, Y.: Spatial patterns of soil moisture connected to monthly-seasonal precipitation variability in a monsoon region, J. Geophys. Res., 108(D22), 1–14, 2003.
Mohanty, B. P., Skaggs, T. H., and Famiglietti, J. S.: Analysis and mapping of field-scale soil moisture variability using high-resolution, ground-based data during the Southern Great Plains 1997 (SGP97) Hydrology Experiment, Water Resour. Res., 36(4), 1023–1031, 2000.
Moore, I. D., Norton, T. W., and Williams, J. E.: Modelling environmental heterogeneity in forested landscapes, J. Hydrol., 150, 717–747, 1993.
Nash, J. E. and Sutcliffe, J. V.: River forecasting through conceptual models, part I, a discussion of principle, J. Hydrol., 10, 282–290, 1970.
Pellenq, J., Kalma, J., Boulet, G., Saulnier, G. M., Wooldridge, S., Kerr, Y., and Chehbouni, A.: A disaggregation scheme for soil moisture based on topography and soil depth, J. Hydrol., 276, 112–127, 2003.
Perry, M. A. and Niemann, J. D.: Analysis and estimation of soil moisture at the catchment scale using EOFs, J. Hydrol., 334(3–4), 388–404, 2007.
Preisendorfer, R. W.: Principal Component Analysis in Meteorology and Oceanography, Developments in Atmospheric Science, 17, Elsevier, New York, 1988.
Roberts, D. W., Dowling, T. I., and Walker, J.: FLAG: A fuzzy landscape analysis GIS method for dryland salinity assessment, Technical Report No. 8/97, CSIRO Land and Water, Canberra, Australia, 1997..
Robock, A., Vinnikov, K. Y., Srinivasan, G., Entin, J. K., Hollinger, S. E., and Speranskaya, N.

A.: The global soil moisture databank, B. Am. Meteorol. Soc., 81, 1281–1299, 2000.
Seyfried, M. S. and Wilcox, B. P.: Scale and the nature of spatial variability: Field example having implications for hydrologic modeling, Water Resour. Res., 31, 173–184, 1995.
Thattai, D. and Islam, S.: Spatial analysis of remotely sensed soil moisture data, J. Hydrol. Eng., 5, 386–392, 2000.
Western, A. W., Bloschl, G., and Grayson, R. B.: Toward capturing hydrologically significant connectivity in spatial patterns, Water Resour. Res., 37(1), 83–97, 2001.
Western, A. W., Blöschl, G., and Grayson, R. B.: Geostatistical characterization of soil moisture patterns in the Tarrawarra catchment, J. Hydrol., 205, 20–37, 1998.
Western, A. W. and Grayson, R. B.: The Tarrawarra data set: Soil moisture patterns, soil characteristics, and hydrological flux measurements, Water Resour. Res., 34(10), 2765–2768, 1998.
Western, A. W., Grayson, R. B., Blöschl, G., Willgoose, G. R., and McMahon, T. A.: Observed spatial organization of soil moisture and its relation to terrain indices, Water Resour. Res., 35(3), 797–810, 1999a.
Western, A. W., Grayson, R. B., and Green, T. R.: The Tarrawarra project: high resolution spatial measurement, modelling and analysis of soil moisture and hydrologic response, Hydrol. Processes, 13, 633–652, 1999b.
Western, A. W., Grayson, R. B., and Green, T. R.: The Tarrawarra project: high resolution spatial measurement, modelling and analysis of soil moisture and hydrological response, Hydrol. Process., 13, 633–652, 1999c.
Wilson, D. J., Western, A. W., and Grayson, R. B.: Identifying and quantifying sources of variability in temporal and spatial soil moisture observations, Water Resour. Res., 40, 1–10, 2004.
Wilson, D. J., Western, A. W., and Grayson, R. B.: A terrain and data-based method for generating the spatial distribution of soil moisture, Adv. Water Resour., 28, 43–54, 2005.
Wood, E. F.: Effects of soil moisture aggregation on surface evaporation fluxes, J. Hydrol., 190, 397–412, 1997.
Wood, E. F.: Report on Dahlem Workshop on Integrating Hydrology, Ecosystem Dynamics, and Biogeochemistry in Complex Landscapes, Wiley, New York, 1999.
Yoo, C. and Kim, S.: EOF analysis of surface soil moisture field variability, Adv. Water Resour., 27, 831–842, 2004.

**Table 1.** Characteristics of the observation sets used for interpolation.

| Observation Spacing | Number of Observations | Percent of Catchment Observed | Number of Realizations Available for Analysis |
|---|---|---|---|
| 10 m×20 m | 459 | 100 | 1 |
| 20 m×40 m | 113 | 25 | 4 |
| 30 m×60 m | 52 | 11 | 9 |
| 40 m×80 m | 30 | 6.5 | 16 |
| 50 m×100 m | 19 | 4.1 | 25 |



HESSD
4, 2837–2874, 2007

**Table 2.** Comparison of EOF and soil moisture correlations with various topographic attributes. Blank entries correspond to statistically insignificant correlations at the 95% confidence level.

| | Simple Correlation Coefficient | | | | | | | | | | Multiple Correlation Coefficient |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Elev. | slope | slope$^{-1}$ | $k_h$ | $k_v$ | $a$ | ln($a$) | WI | PSRI | L | |
| EOF1 | −0.39 | −0.26 | 0.39 | −0.58 | −0.36 | 0.40 | 0.67 | 0.67 | −0.29 | 0.67 | 0.82 |
| EOF2 | | −0.40 | 0.40 | −0.13 | | 0.20 | | 0.12 | 0.56 | | 0.63 |
| EOF3 | −0.33 | −0.19 | 0.22 | | | 0.16 | 0.15 | 0.19 | −0.17 | | 0.50 |
| EOF4 | | 0.11 | | 0.13 | | | | −0.10 | | | 0.15 |
| EOF5 | 0.15 | | | | 0.10 | | | | | | 0.20 |
| EOF6 | | −0.09 | | −0.11 | −0.13 | | | | | | 0.13 |
| 27 Sep | −0.41 | −0.23 | 0.35 | −0.52 | −0.33 | 0.36 | 0.65 | 0.64 | −0.38 | 0.62 | 0.82 |
| 14 Feb | −0.10 | 0.28 | −0.23 | | −0.18 | | 0.14 | | −0.40 | 0.18 | 0.48 |
| 23 Feb | | 0.20 | −0.12 | | | | 0.11 | | −0.31 | 0.13 | 0.36 |
| 28 March | −0.21 | | | −0.23 | −0.23 | | 0.27 | 0.21 | −0.38 | 0.31 | 0.50 |
| 13 April | −0.16 | | | −0.36 | −0.31 | 0.12 | 0.41 | 0.35 | −0.38 | 0.43 | 0.58 |
| 22 April | −0.22 | −0.29 | 0.37 | −0.50 | −0.26 | 0.33 | 0.52 | 0.54 | | 0.51 | 0.62 |
| 2 May | −0.35 | −0.30 | 0.40 | −0.63 | −0.34 | 0.37 | 0.63 | 0.64 | −0.18 | 0.64 | 0.77 |
| 3 July | | −0.14 | 0.21 | −0.39 | | 0.22 | 0.31 | 0.32 | | 0.32 | 0.51 |
| 2 Sep | −0.30 | −0.24 | 0.38 | −0.35 | −0.24 | 0.34 | 0.43 | 0.46 | −0.13 | 0.46 | 0.57 |
| 20 Sep | −0.33 | −0.34 | 0.40 | −0.42 | −0.25 | 0.35 | 0.43 | 0.47 | | 0.45 | 0.56 |
| 25 Oct | −0.41 | −0.24 | 0.35 | −0.45 | −0.30 | 0.40 | 0.63 | 0.63 | −0.30 | 0.57 | 0.74 |
| 10 Nov | −0.35 | −0.18 | 0.29 | −0.40 | −0.27 | 0.31 | 0.51 | 0.51 | −0.46 | 0.49 | 0.75 |
| 29 Nov | −0.22 | | 0.18 | −0.36 | −0.21 | 0.21 | 0.43 | 0.40 | −0.32 | 0.46 | 0.60 |



HESSD
4, 2837–2874, 2007

**Table 3.** Correlations between soil moisture estimation errors and topographic attributes for the EOF-MLR and the direct MLR method. Interpolations were performed using a single observation set at the 30 m×60 m spacing. Estimation errors were computed only at verification points.

| | Multiple Correlation Coefficient between Residual Errors and Topographic Attributes | |
|---|---|---|
| Soil Moisture Sampling Date | MLR-EOF | Direct MLR of Soil Moisture |
| 27 Sep | 0.21 | 0.12 |
| 14 Feb | 0.17 | 0.13 |
| 23 Feb | 0.11 | 0.14 |
| 28 March | 0.12 | 0.25 |
| 13 April | 0.14 | 0.22 |
| 22 April | 0.16 | 0.22 |
| 2 May | 0.23 | 0.22 |
| 3 July | 0.18 | 0.21 |
| 2 Sep | 0.08 | 0.27 |
| 20 Sep | 0.09 | 0.18 |
| 25 Oct | 0.20 | 0.42 |
| 10 Nov | 0.17 | 0.20 |
| 29 Nov | 0.15 | 0.27 |



HESSD
4, 2837–2874, 2007

**Table 4.** Estimated parameters of the exponential semivariogram model for the most important EOFs and individual soil moisture patterns.

| Spatial Pattern | Nugget, $a$ $(V/V)^2$ | Sill, $b$ $(V/V)^2$ | Range, $3c$ (m) | RMSE |
|---|---|---|---|---|
| EOFs | | | | |
| 1 | 0.00 | 1.0 | 94 | 1.61 |
| 2 | 0.25 | 1.0 | 116 | 1.35 |
| 3 | 0.50 | 1.2 | 370 | 1.54 |
| 4 | 0.58 | 1.4 | 536 | 1.43 |
| 5 | 0.71 | 1.1 | 264 | 1.51 |
| 6 | 0.53 | 1.0 | 64 | 1.49 |
| Soil Moisture | | | | |
| 3 July | 0.37 | 1.1 | 138 | 1.81 |
| 20 Sep | 0.17 | 1.0 | 62 | 1.81 |



HESSD
4, 2837–2874, 2007



EGU

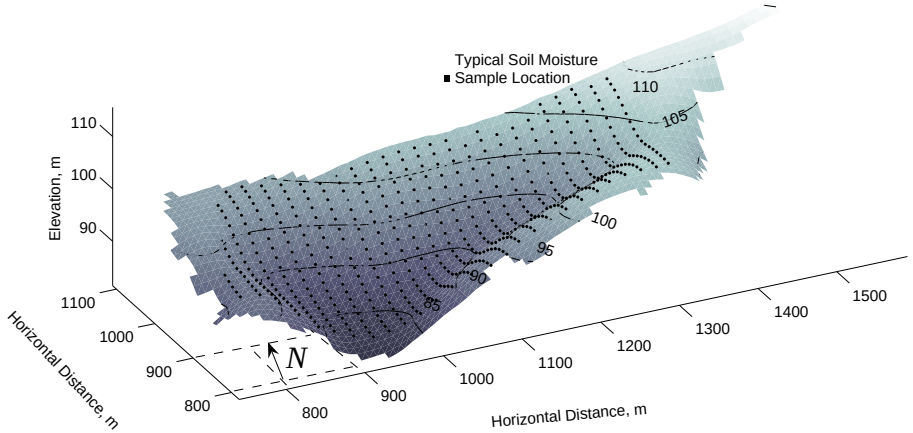



**Fig. 1.** Tarrawarra catchment topography. The overlaid dots show the soil moisture sampling grid for a typical observation date.



HESSD
4, 2837–2874, 2007

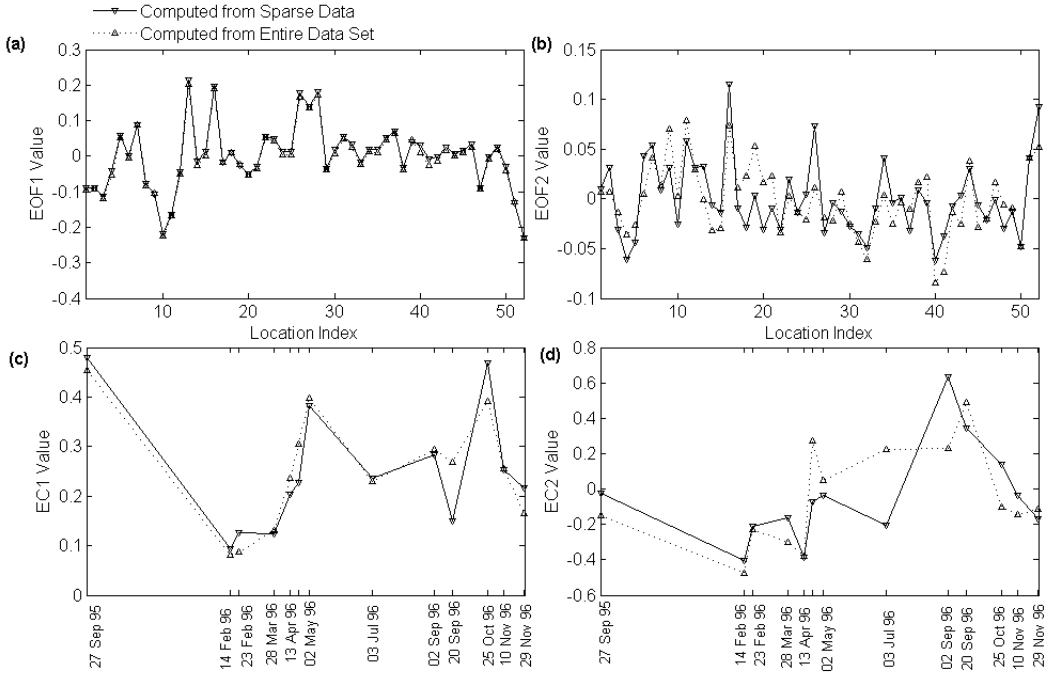



**Fig. 2.** Comparison of the EOFs and ECs determined from an observation set at a 30 m×60 m spacing and the EOFs and ECs at the same locations using the entire dataset. The lines connecting data points are for visual clarity only and do not to represent values between data points. The location index is an arbitrary index that identifies individual observation points.



HESSD
4, 2837–2874, 2007

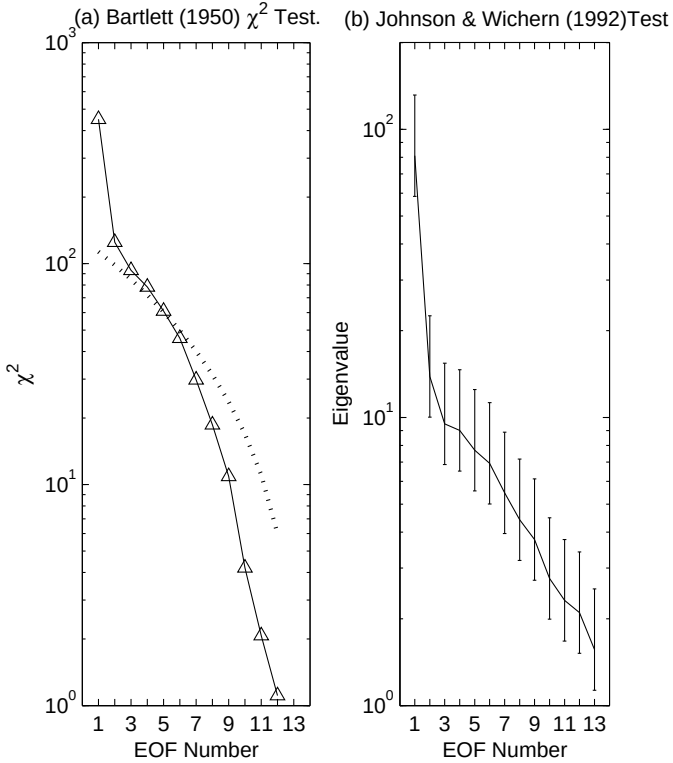



**Fig. 3.** Results from **(a)** the Bartlett (1950) and **(b)** the Johnson and Wichern (2002) tests for determining statistically significant EOFs. In (a) significant EOFs have chi-squared statistics that exceed the table value. In (b), significant EOFs have error bars that do not overlap with the next higher EOF. Both tests are shown for the 95% confidence level and a 30 m×60 m sample spacing.



HESSD
4, 2837–2874, 2007

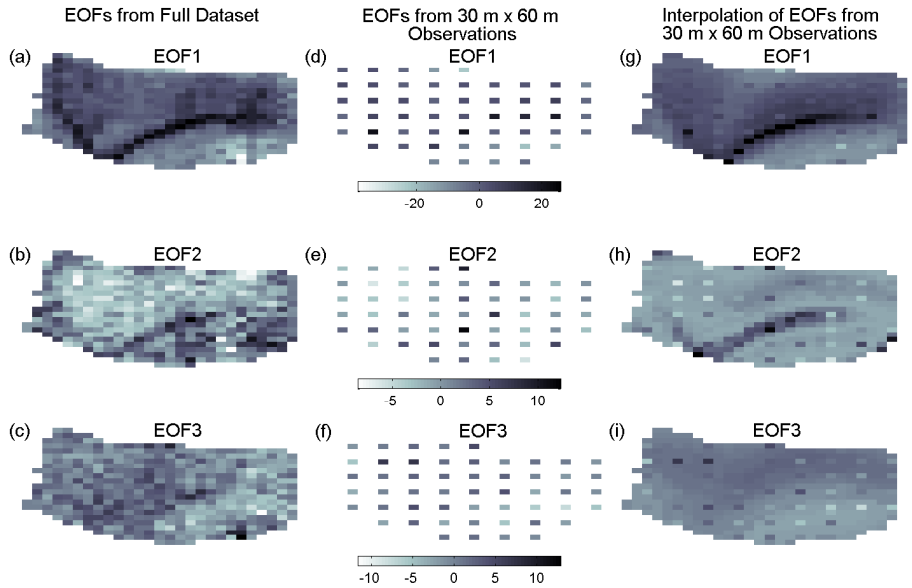



**Fig. 4. (a–c)** EOFs 1-3 computed from the full 10 m×20 m soil moisture dataset, **(d–f)** EOFs 1–3 computed from an observation set with a 30 m×60 m spacing, and **(g–i)** Interpolations of EOFs 1–3 from the 30 m×60 m spacing back to the 10 m×20 m spacing using a MLR against topographic attributes.



HESSD
4, 2837–2874, 2007



EGU

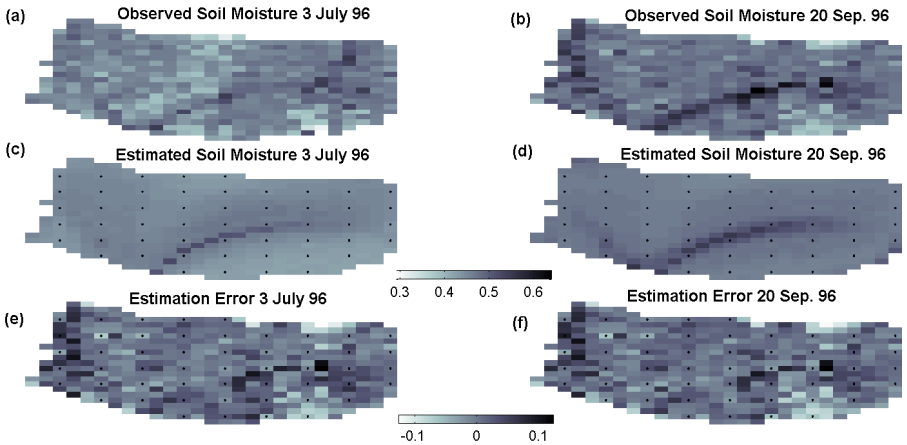



**Fig. 5. (a–b)** Observed soil moisture patterns on 3 July 1996 and 20 September 1996, **(c–d)** soil moisture patterns generated using the EOF-based interpolation method, and **(e–f)** differences between the actual and estimated soil moisture patterns (estimation errors). Units are volume of water per volume of soil.



HESSD
4, 2837–2874, 2007



EGU

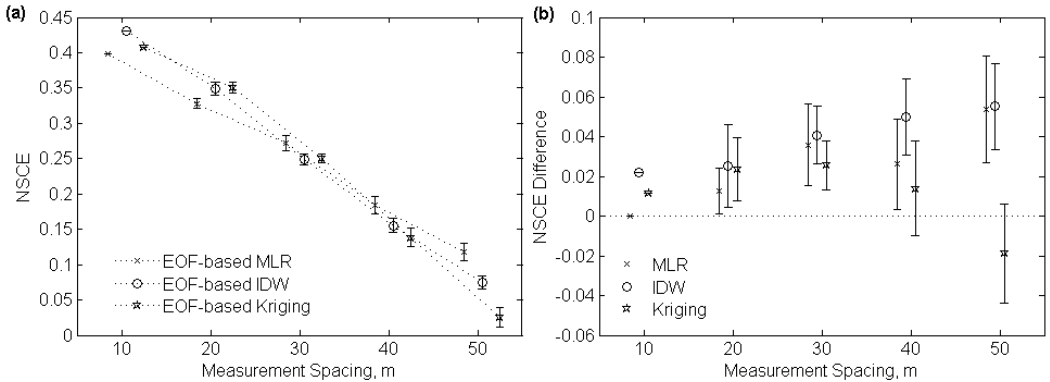



**Fig. 6. (a)** Average NSCE for estimated soil moisture patterns as a function of the observation spacing for different interpolation methods. Error bars show +/−1 standard deviation about the average NSCE. **(b)** Difference between the average NSCE from the EOF-based interpolation patterns and the analogous direct interpolation patterns. Error bars show +/−1 standard deviation about the average NSCE difference. In both (a) and (b), the numbers on the horizontal axis refer to the short dimension of observation spacing (e.g., 20 m corresponds to a 20 m×40 m spacing). In all cases, spacings are exact multiples of 10 (the symbols have been slightly offset horizontally for visual clarity).



HESSD
4, 2837–2874, 2007

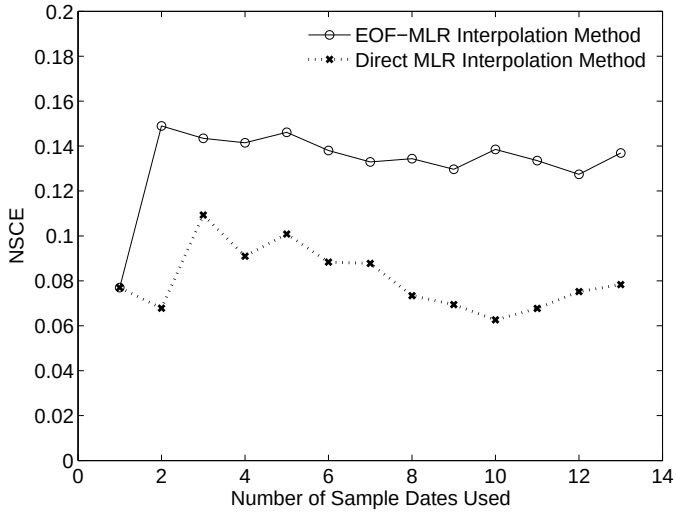



**Fig. 7.** The NSCE of the interpolated soil moisture patterns for 28 March as a function of the number of sample dates included. The plot compares EOF-MLR interpolation and direct MLR. In the latter case, the data from each date were standardized by removing the spatial average and dividing by the standard deviation before adding them to the regression dataset.



HESSD
4, 2837–2874, 2007