AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE
10.1002/2015WR016944

# Optimal averaging of soil moisture predictions from ensemble land surface model simulations

**W. T. Crow[1], C.-H. Su[2], D. Ryu[2], and M. T. Yilmaz[3]**

[1]USDA Hydrology and Remote Sensing Laboratory, Beltsville, Maryland, USA, [2]Department of Infrastructure Engineering, University of Melbourne, Parkville, Victoria, Australia, [3]Department of Civil Engineering, Middle East Technical University, Ankara, Turkey

**Key Points:**
- Soil moisture can be predicted from a mulit-model ensemble
- Interpretation of the ensemble requires model error covariance information
- Such information can be obtained using an instrumental variable approach

**Correspondence to:**
W. Crow,
Wade.Crow@ARS.USDA.GOV

**Citation:**
Crow, W. T., C.-H. Su, D. Ryu, and M. T. Yilmaz (2015), Optimal averaging of soil moisture predictions from ensemble land surface model simulations, *Water Resour. Res.*, *51*, 9273–9289, doi:10.1002/2015WR016944.

Received 15 JAN 2015
Accepted 5 NOV 2015
Accepted article online 11 NOV 2015
Published online 28 NOV 2015

© 2015. American Geophysical Union. All Rights Reserved.

**Abstract** The correct interpretation of ensemble information obtained from the parallel implementation of multiple land surface models (LSMs) requires information concerning the LSM ensemble's mutual error covariance. Here we propose a technique for obtaining such information using an instrumental variable (IV) regression approach and comparisons against a long-term surface soil moisture data set acquired from satellite remote sensing. Application of the approach to multimodel ensemble soil moisture output from Phase 2 of the North American Land Data Assimilation System (NLDAS-2) and European Space Agency (ESA) Soil Moisture (SM) Essential Climate Variable (ECV) data set allows for the calculation of optimal weighting coefficients for individual members of the NLDAS-2 LSM ensemble and a biased-minimized estimate of uncertainty in a deterministic soil moisture analysis derived via optimal averaging. As such, it provides key information required to accurately condition soil moisture expectations using information gleaned from a multimodel LSM ensemble. However, existing continuity and rescaling concerns surrounding the generation of long-term, satellite-based soil moisture products must likely be resolved before the proposed approach can be applied with full confidence.

## 1. Introduction

The use of ensemble predictions generated from multiple land surface models (LSMs) is becoming increasingly popular [see e.g., *Xia et al.*, 2014; *Wang et al.*, 2009; *Sheffield et al.*, 2012]. A number of recent studies have demonstrated that ensemble-averaged output obtained from multimodel ensembles is generally superior to results from any single model within the ensemble [see e.g., *Gao and Dirmeyer*, 2006; *Crow et al.*, 2012]. However, in order to obtain an optimized deterministic prediction, or properly interpret the ensemble as a probabilistic forecast, the error covariance properties of individual ensemble members must be considered [*Bohn et al.*, 2010].

Assume $\mathbf{x}_i$ is a $M$-member time series vector of daily soil moisture output acquired from model $i$ and can be represented in a linear form as:

$$\mathbf{x}_i = \alpha_i \hat{\mathbf{x}} + \epsilon_i \tag{1}$$

where the $M$-member vector $\hat{\mathbf{x}}$ represents the unknown true value of the output variable; $\alpha_i$ is a constant scalar gain, and $\epsilon_i$ is a time series vector of orthogonal random error. A priori, only $\mathbf{x}_i$ is assumed to be known. Note that (1) represents the most general linear form for a time series and therefore represents the logical starting form for any model ensemble analysis. In addition, $\mathbf{x}_i$ is assumed to represent mean-zero anomalies calculated after the removal of a seasonal climatology.

Given that each model member within the ensemble is likely generated from overlapping forcing data and model physics, they will almost certainly be afflicted by cross-correlated modeling errors. Therefore, it must be assumed that:

$$\mathrm{E}[\epsilon_i \epsilon_j^T] \neq 0 \tag{2}$$

for any given combination of models $i$ and $j$.

A common goal is obtaining an optimal (i.e., error-variance minimized) weighted average of $\mathbf{x}_i$ across $N$ different models via weighted linear averaging. Prior to such averaging, all time series $\mathbf{x}_i$ must be rescaled so that they have equivalent scalar gains (i.e., $\alpha$) with respect to the truth [*Yilmaz and Crow*, 2013]. This assures that differencing the two time series will yield a time series composed solely of random errors. Assuming that any additive bias has already been removed, this is generally based on arbitrarily selecting one model time series vector ($\mathbf{x}_1$) to serve as a reference and re-scaling all other model output by:

$$\tilde{\mathbf{x}}_i = \frac{\alpha_1}{\alpha_i}\mathbf{x}_i. \tag{3}$$

A $M$-member time series can be defined as the weighted averaged of these rescaled model products:

$$\mathbf{x}^* = \sum_{i=1}^{N} \lambda_i \tilde{\mathbf{x}}_i \tag{4}$$

where:

$$\sum_{i=1}^{N} \lambda_i = 1. \tag{5}$$

This weighted average will be optimal in a least-squared-error sense if:

$$\lambda_i(\mathbf{C}) = \frac{\sum_j C_{ij}^{-1}}{\sum_i \sum_j C_{ij}^{-1}} \tag{6}$$

where $\mathbf{C}$ is the $N \times N$ error covariance matrix for $[\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \ldots, \tilde{\mathbf{x}}_N]$; $C_{ij}^{-1}$ is a particular component within the matrix inverse of $\mathbf{C}$, and $\mathbf{x}^*$ is a $M$-member vector containing a time series prediction with minimized error variance. In addition, $\mathbf{C}$ can also be used to obtain the scalar error variance of $\mathbf{x}^*$ as:

$$\mathrm{Var}[\epsilon_{\mathbf{x}}^*] = \left[\sum_{i,j} C_{ij}^{-1}\right]^{-1}. \tag{7}$$

Therefore, $\mathbf{C}$ is central to understanding how multimodel ensemble output should condition expectations concerning the true signal $\hat{\mathbf{x}}$ [*Perrons and Cooper*, 1992]. In particular, it provides the fundamental statistical information required by any sophisticated attempt to optimally merge multimodel soil moisture state estimates.

The practical challenge is therefore providing an accurate estimate of the $N \times N$ $\mathbf{C}$ matrix as well as the scalar $\alpha_1/\alpha_i$ ratios required by (3). These quantities can be directly calculated for cases in which high-quality reference observations are available as an absolute point of comparison for LSM predictions [see e.g., *Gao and Dirmeyer*, 2006; *Xia et al*., 2014]. However, such observations are typically available only for limited geographic areas. Over broader spatial scales, remote-sensing retrievals provide a viable alternative; however, they contain their own unique random errors which will bias estimates of $\mathbf{C}$ obtained from direct comparisons of modeling and remote sensing products. In addition, the typically short historical length of remote sensing data sets may introduce unacceptably high levels of sampling uncertainty.

In the theory of instrumental variables (IV), a variable $\mathbf{z}$ is referred to as an instrument if it is linearly related to the same unknown truth as $\mathbf{x}_i$ yet still contains independent errors [*Bowden and Turkington*, 1990]. Therefore, the basis of a soil moisture IV analysis is the assumption that an additional estimate of surface soil moisture is available and can be represented as:

$$\mathbf{z} = \alpha_z \hat{\mathbf{x}} + \eta \tag{8}$$

where:

$$\mathrm{E}[\epsilon_i \eta^T] = 0 \tag{9}$$

for all $i$. For consistency with $\mathbf{x}$ in (1), $\mathbf{z}$ is also assumed to represent an estimate of soil moisture daily climatological anomalies.

Recent work has introduced the concept of a lagged-variable IV analysis to estimate $\alpha_1/\alpha_i$ and soil error variance information via the cross-comparison of remotely sensed and modeled soil moisture products [*Su et al*., 2014]. This approach is notable because it effectively eliminates the impact of (significant) random errors in each product on the estimation of these quantities. However, the approach in *Su et al*. [2014] can only be used to capture the diagonal of the **C** matrix. Due to the error cross-dependence expressed in (2), off-diagonal components of **C** are expected to have a significant impact on the estimation of $\lambda_i$ and/or Var $[\epsilon_{\mathbf{x}}^*]$ for soil moisture products generated within a multimodel LSM ensemble [*Bohn et al*., 2010]. Here we will expand upon earlier results presented in *Su et al*. [2014] by introducing a novel strategy for obtaining both the diagonal and off-diagonal components of **C** necessary for the optimal calculation of $\lambda_i$ and Var$[\epsilon_{\mathbf{x}}^*]$ for soil moisture estimates obtained from multiple LSMs. While our presentation will focus exclusively on an application involving LSM and remotely sensed surface soil moisture estimates, it should be stressed that the approach is broadly applicable to any auto-correlated geophysical variable which can be simultaneously estimated via a multimodel ensemble and independently measured via either ground-based instrumentation or remote sensing.

Section 2 outlines our strategy for using an IV analysis to estimate $\alpha_1/\alpha_i$, **C**, $\lambda_i$ and Var$[\epsilon_{\mathbf{x}}^*]$, and sections 3–5 apply the approach to the case of multimodel soil moisture ensemble output generated by LSMs within Phase 2 of the North American Land Data Assimilation System (NLDAS-2) [*Xia et al*., 2012a, 2012b] and remotely sensed surface soil moisture retrievals obtained from the multisatellite European Space Agency (ESA) Soil Moisture (SM) Essential Climate Variable (ECV) data set [*Wagner et al*., 2012].

## 2. Theoretical Approach

As noted above, the first step in an optimal averaging approach—assuming all additive biases have been removed—is rescaling all time series vectors such that their gains match and all relative multiplicative biases have been removed. Such calibration can be accomplished given the availability of an additional instrumental variable with independent errors [*Su et al*., 2014]. In this particular case, if $\mathbf{x}_i$ and **z** satisfy the assumptions listed in (1), (8) and (9), then **z** can be used as an instrumental variable to estimate the required scaling ratio in (3) as:

$$\frac{\alpha_1}{\alpha_i} = \frac{\mathbf{x}_1 \mathbf{z}^T}{\mathbf{x}_i \mathbf{z}^T}. \tag{10}$$

This result can be directly obtained by taking the cross-product of the vectors introduced in (1) and (8) under the condition that $\eta$ is independent of $\epsilon$ and $\hat{\mathbf{x}}$ [*Stoffelen*, 1998].

The term triple collocation (TC) is commonly used to refer to the special case of IV in which all three time series products are derived from independent sources [*Vogelzang and Stoffelen*, 2012; *Scipal et al*., 2008]. However, *Su et al*. [2014] also describe the application of IV to a lagged variable (LV) case where only two independent data sources are used and a third variable is constructed by temporally lagging one variable. It is based on the premise that $\hat{\mathbf{x}}$ is strongly auto-correlated in time and is being sampled (by $\mathbf{x}_i$ and **z**) at temporal repeat scales which are smaller than its signal auto-correlation length.

Our particular approach is based on lagging the *M*-member time series vector **z** by *L* days to produce the time series $\mathbf{z}_L$. In this LV framework, the required rescaling factors for **z** and $\mathbf{z}_L$ can then be estimated and applied via:

$$\tilde{\mathbf{z}} = \frac{\mathbf{x}_1 \mathbf{z}_L^T}{\mathbf{z}\mathbf{z}_L^T} \mathbf{z} \tag{11}$$

$$\tilde{\mathbf{z}}_L = \frac{\mathbf{x}_1 \mathbf{z}^T}{\mathbf{z}_L \mathbf{z}^T} \mathbf{z}_L. \tag{12}$$

The relationship between **z** and **x** is resolved using $\mathbf{z}_\mathbf{L}$ as the instrument in (11). Likewise, the relationship between $\mathbf{z}_\mathbf{L}$ and **x** is resolved using **z** as the instrument in (12). However, while IV sampling is predicated on the assumption of an auto-correlated signal, it can become biased by auto-correlated errors in the instrumental variable [*Su et al*., 2014]. Specifically, if the assumption in (9) holds, inserting (1) and (8) into (11) and (12) yields:

$$\frac{\mathbf{x}_1\mathbf{z}_L^T}{\mathbf{z}\mathbf{z}_L^T} = \frac{\sigma_{\hat{x}}^2\alpha_1\alpha_{zL}}{\sigma_{\hat{x}}^2\alpha_z\alpha_{zL}+\alpha_z\alpha_{zL}\mathrm{E}[\eta\eta_L^T]} \tag{13}$$

$$\frac{\mathbf{x}_1\mathbf{z}^T}{\mathbf{z}_L\mathbf{z}^T} = \frac{\sigma_{\hat{x}}^2\alpha_1\alpha_z}{\sigma_{\hat{x}}^2\alpha_{zL}\alpha_z+\alpha_{zL}\alpha_z\mathrm{E}[\eta_L\eta^T]} \tag{14}$$

where $\sigma\hat{x}^2$ is the temporal variance of $\hat{\mathbf{x}}$ and $\alpha_{zL}$ the scaling constant obtained when decomposing $z_L$ according to (1). Note that (13) and (14) converge to unbiased estimates of the required scaling factors ($\alpha_1/\alpha_{zL}$ and $\alpha_1/\alpha_z$, respectively) only if $\mathrm{E}[\eta_L\eta^T]$ is negligible. Therefore the presence of temporally correlated errors in the instrumental variable $\mathbf{z}$ will introduce biases into our proposed LV analysis. This will have important implications below.

After rescaling, the IV approach can be applied to estimate $\mathbf{C}$. This approach is based on defining a general IV operator to be:

$$\mathrm{IV}[\mathbf{a},\mathbf{b},\mathbf{c}] \equiv \tilde{M}^{-1}(\mathbf{b}-\mathbf{a})(\mathbf{b}-\mathbf{c})^T \tag{15}$$

where $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ are arbitrary time series vectors and $\tilde{M}$ is the total number of time steps at which a complete triplet of time series data is available. When applied to time series vectors, the operator will produce scalar outputs which can be used to define components of $\mathbf{C}$. For example, to estimate the diagonal components of $\mathbf{C}$ ($C_{ii}$), (15) can be applied to estimate the error variance of a particular rescaled model product ($\tilde{\mathbf{x}}_i$) using both an original ($\tilde{\mathbf{z}}$) and a lagged ($\tilde{\mathbf{z}}_L$) version of the remotely sensed surface soil moisture product. Following *Yilmaz and Crow* [2014], the results of this IV operation can be expressed as:

$$\mathrm{IV}[\tilde{\mathbf{z}},\tilde{\mathbf{x}}_i,\tilde{\mathbf{z}}_L] = C_{ii}+\delta_{LS}+\delta_{XCE} \tag{16}$$

where

$$\delta_{LS}(L) = \sigma_{\hat{x}}^2\left(\alpha_1 - \frac{\sigma_{\hat{x}}^2\alpha_1\alpha_z\alpha_{zL}}{\sigma_{\hat{x}}^2\alpha_z\alpha_{zL}+\alpha_z\alpha_{zL}\mathrm{E}[\eta\eta_L^T]}\right)^2 \tag{17}$$

and

$$\delta_{XCE}(L) = \frac{\alpha_1^2}{\alpha_z\alpha_{zL}}\mathrm{E}[\eta\eta_L^T]. \tag{18}$$

Our goal here is to isolate un-biased estimates of diagonal components in the $N \times N$ matrix $\mathbf{C}$ (or $C_{ii}$). Due to the error independence assumption in (9) and the IV-based averaging strategy, (16) is not impacted by the effects of modeling error. However, $C_{ii}$ in (16) is obscured by two separate additive scalar terms associated with errors in the remote-sensing product $\mathbf{z}$. The "leaked signal" (LS) term in (17) describes the effect of the $\hat{\mathbf{x}}$ signal "leaking" into an IV-based analysis of noise due to bias in rescaling coefficients calculated via (11 and 12). In contrast, the "cross-correlated error" (XCE) term defined in (18) captures the direct IV impact of $\tilde{\mathbf{z}}$ and $\tilde{\mathbf{z}}_L$ containing cross-correlated error and is nonzero even in the case of perfect rescaling. Despite this difference, both terms vanish for the case of white observing errors (i.e., $\mathrm{E}[\eta\eta_L^T]=0$). However, past research has demonstrated that, in general, $\mathrm{E}[\eta\eta_L^T] > 0$ for existing satellite-derived surface soil moisture products [*Crow and van den Berg*, 2010]. Therefore, it is reasonable to expect $\delta_{LS}$ and $\delta_{XCE}$ in (16) to be non-negligible for most soil moisture applications. This is potentially problematic given the difficulty in estimating $\mathrm{E}[\eta\eta_L^T]$, and therefore, the magnitude of both $\delta_{LS}(L)$ and $\delta_{XCE}(L)$.

Off-diagonal terms in $\mathbf{C}$ ($C_{ij}$) can be estimated by applying the operator in (15) to model products $\tilde{\mathbf{x}}_j$ and $\tilde{\mathbf{x}}_i$ and a nonlagged remote sensing product $\tilde{\mathbf{z}}$. Again following *Yilmaz and Crow* [2014], the resulting IV expression $\mathrm{IV}[\tilde{\mathbf{x}}_j,\tilde{\mathbf{x}}_i,\tilde{\mathbf{z}}]$ can be written as:

$$\mathrm{IV}[\tilde{\mathbf{x}}_j,\tilde{\mathbf{x}}_i,\tilde{\mathbf{z}}] = C_{ii} - \frac{\alpha_1^2}{\alpha_i\alpha_j}\mathrm{E}[\epsilon_i\epsilon_j^T] = C_{ii} - C_{ij}. \tag{19}$$

The second term in (19) captures the expected nonzero error covariance between models $i$ and $j$ and is equal to the negative of the off-diagonal matrix component $C_{ij}$. As a result, it is analogous to the cross-correlated-error bias term in (18). In contrast, *Yilmaz and Crow* [2014] demonstrate that the leaked signal term converges to zero whenever two out of the three IV inputs have been adequately rescaled. Based on

(10), this condition is met for both modeling products $\tilde{\mathbf{x}}_i$ and $\tilde{\mathbf{x}}_j$. As a result, no leaked signal term is required in (19).

Subtracting (19) from (16) leads to:

$$\mathrm{IV}[\tilde{\mathbf{z}}, \tilde{\mathbf{x}}_i, \tilde{\mathbf{z}}_L] - \mathrm{IV}[\tilde{\mathbf{x}}_j, \tilde{\mathbf{x}}_i, \tilde{\mathbf{z}}] = C_{ij} + \delta_{LS} + \delta_{XCE}. \tag{20}$$

Therefore, based on (16) and (20), it is possible to express components of the **C** matrix as:

$$C_{ij} = \begin{cases} \mathrm{IV}[\tilde{\mathbf{z}}, \tilde{\mathbf{x}}_i, \tilde{\mathbf{z}}_L] - \delta_{LS}(L) - \delta_{XCE}(L) & i = j \\ \mathrm{IV}[\tilde{\mathbf{z}}, \tilde{\mathbf{x}}_i, \tilde{\mathbf{z}}_L] - \mathrm{IV}[\tilde{\mathbf{x}}_j, \tilde{\mathbf{x}}_i, \tilde{\mathbf{z}}] - \delta_{LS}(L) - \delta_{XCE}(L) & i \neq j. \end{cases} \tag{21}$$

It is also possible to exploit the symmetric nature of **C** by independently sampling both $\mathrm{IV}[\tilde{\mathbf{z}}, \tilde{\mathbf{x}}_i, \tilde{\mathbf{z}}_L] - \mathrm{IV}[\tilde{\mathbf{x}}_j, \tilde{\mathbf{x}}_i, \tilde{\mathbf{z}}]$ and $\mathrm{IV}[\tilde{\mathbf{z}}, \tilde{\mathbf{x}}_j, \tilde{\mathbf{z}}_L] - \mathrm{IV}[\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j, \tilde{\mathbf{z}}]$ and then averaging. This leads to estimates of the form:

$$\tilde{C}_{ij} = \begin{cases} \mathrm{IV}[\tilde{\mathbf{z}}, \tilde{\mathbf{x}}_i, \tilde{\mathbf{z}}_L] & i = j \\ 0.5(\mathrm{IV}[\tilde{\mathbf{z}}, \tilde{\mathbf{x}}_i, \tilde{\mathbf{z}}_L] - \mathrm{IV}[\tilde{\mathbf{x}}_j, \tilde{\mathbf{x}}_i, \tilde{\mathbf{z}}] + \mathrm{IV}[\tilde{\mathbf{z}}, \tilde{\mathbf{x}}_j, \tilde{\mathbf{z}}_L] - \mathrm{IV}[\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j, \tilde{\mathbf{z}}]) & i \neq j. \end{cases} \tag{22}$$

The sampled matrix $\tilde{\mathbf{C}}$ in (22) converges to an unbiased estimate of **C** (21) only when $\mathrm{E}[\eta\eta_L^T] = 0$ and the $\delta_{LS}$ or $\delta_{XCE}$ terms are negligible with respect other terms in (21). In all other cases, individual components of the $\tilde{\mathbf{C}}$ matrix will be equally biased by $\delta_{LS} + \delta_{XCE}$.

As discussed above, the term $\delta_{LS} + \delta_{XCE}$ in (21) is likely to be both nonnegligible and challenging to quantify. Therefore, instead of attempting to directly estimate or neglect it, we will instead attempt to develop effective strategies for mitigating its impact. For instance, (21) demonstrates that, for the specific case of applying **z** as the lagged variable in an IV analysis, neither $\delta_{LS}$ nor $\delta_{XCE}$ varies as a function of matrix row or column (i.e., model type). This lack of component-to-component variation is important for the calculation of optimal weights since:

$$\lambda_i(\mathbf{C}) = \lambda_i(\mathbf{C} + \delta\mathbf{J}) = \lambda_i(\tilde{\mathbf{C}}) \tag{23}$$

if **J** is an $N \times N$ matrix containing only ones and $\delta$ any real scalar (here equal to the negative of $\delta_{LS} + \delta_{XCE}$ for consistency with (21)). See Appendix A for a proof of (23). As a result, estimated $\lambda_i$ acquired from $\tilde{\mathbf{C}}$ via (6) will be unaffected by nonzero $\delta_{LS}$ and $\delta_{XCE}$ terms, and uniform bias in $\tilde{\mathbf{C}}$ components will not propagate into $\lambda_i$ estimates.

The invariance captured by (23) also has implications for the optimal selection of $L$. A small value for lag $L$ increases the amount of lagged auto-covariance in the remotely sensed time series ($\mathrm{E}[\eta\eta_L^T]$) and, therefore, both $\delta_{LS} + \delta_{XCE}$ and the amount of bias in sampled $\tilde{\mathbf{C}}$ components. However, since time series signal autocorrelation generally decreases for larger sampling lags, increasing $L$ will gradually erode the strength of the instrumental variables $\mathbf{z}_L$ and **z** in (11 and 12) and increase the level of sampling error in LV-based estimates. In addition, if **z** is irregularly sampled in time (as is typically the case for satellite-derived data sets), the $L = 0$ case will also contain the largest number of viable data pairs. Therefore, smaller values of $L$ are generally preferable when estimating $\lambda_i$ since bias concerns are effectively eliminated by (23).

However, it is important to recognize that the invariance in (23) does not extend to all possible uses of the **C** matrix. For example, estimates of $\mathrm{Var}[\epsilon_{\mathbf{x}}^*]$ derived from (7) will be biased by nonzero $\delta_{LS}(L)$ and $\delta_{XCE}(L)$ terms (even if such bias is equally distributed among all components of the **C** matrix). In such cases, the bias-minimizing advantages of high $L$ must be balanced against the sampling advantages of low $L$. In addition, the additive invariance expressed in (23) is not relevant for other potential LV strategies. For example, model-based soil moistures $\mathbf{x}_i$ cannot be used as the lagged variable in (22) since that would result in $\delta_{XCE}$ terms varying along the diagonal of the $\tilde{\mathbf{C}}$ matrix (since each model presumably possesses its own unique error auto-correlation attributes). As a result, (23) is only relevant for the particular case of lagging the remote-sensing variable **z**.

## 3. Numerical Example

As a numerical example, we applied the IV approach outlined in section 2 to estimate **C**, $\lambda_i$ and $\mathrm{Var}[\epsilon_{\mathbf{x}}^*]$ for long-term, 0.25°-resolution surface soil moisture predictions acquired from an ensemble of four LSM's over

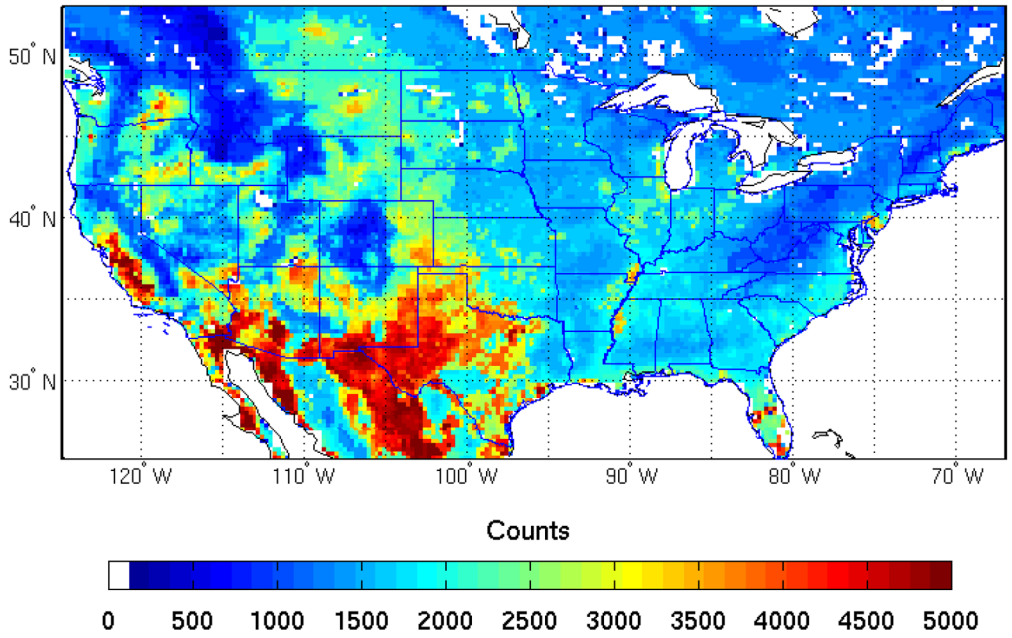



**Figure 1.** Number of separate daily surface soil moisture retrievals contained within the 30 year ESA SM ECV data set.

North America. This ensemble was based on daily (12 UTC) 1979–2008 surface (0–10 cm) soil moisture estimates generated from the Noah, Mosaic, Variable Infiltration Capacity (VIC) and Sacramento Soil Moisture Accounting Model (SAC) LSM models as part of NLDAS-2 activities [*Xia et al*., 2012a, 2012b]. These simulations were forced using gauge-based daily precipitation estimates disaggregated into hourly estimates using ground-based rain radar measurements [*Cosgrove et al*., 2003] and topographically corrected via monthly climatological information acquired from the Parameter-elevation Regressions on Independent Slopes Model (PRISM) [*Daly et al*., 1994]. Remaining nonprecipitation forcing data for NLDAS-2 simulations were obtained from National Center for Environmental Prediction (NCEP) North American Regional Reanalysis estimates. In all results below, the Noah model was used as the re-scaling reference. The impact of this arbitrary choice will be discussed below.

An additional independent data set, corresponding to **z** in (8), was based on daily 1979–2008 surface soil moisture retrievals obtained from the multisatellite European Space Agency (ESA) Soil Moisture (SM) Essential Climate Variable (ECV) data set. This integrated, multisensor data set was generated via the merger of a series of satellite-based C-band scatterometers and multifrequency radiometers [*Wagner et al*., 2012]. Individual retrieval products from each sensor were independently scaled to match a seasonal climatology derived from the Noah LSM [*Liu et al*., 2011, 2012].

Figure 1 shows the total number of retrievals available within each North American 0.25° grid cell during the 30 year duration of the multisensor ESA SM ECV data set. Note that data length numbers in Figure 1 are reduced by the masking of soil moisture retrievals due to: active precipitation, frozen soil, surface snow, and dense vegetation cover. In addition, the retrieval products comprising the first decade on the ESA SM ECV product tend to be available at significantly reduced temporal frequencies relative to more recent satellite products [*Wagner et al*., 2012]. As a result, the analysis places relatively more weight on recent decades characterized by more frequent soil moisture retrievals. Since our aim here is the derivation of soil moisture error covariance information, all values are reported in dimensions of squared volumetric soil moisture (VSM) percentages (i.e., $(100*\mathrm{VSM})^2$) or $(\%^2)$) where VSM is defined as the volume of water contained per (bulk) volume of soil.

The ESA SM ECV data set was selected because it provides the longest available remotely sensed surface soil moisture data product. However, its multisensor basis potentially complicates its application as an IV (see section 6). In order to examine this possibility, selected results were also generated for a shorter 9 year (2002–2011) analysis based on a single-sensor soil moisture product generated by applying the Land Parameter Retrieval Model (LPRM) to descending orbit (1:30 AM) brightness temperature products obtained

from the Advanced Microwave Scanning Radiometer—EOS (AMSR-E) sensor aboard the NASA Aqua satellite [*Owe et al*., 2008]. It should be noted that this same LPRM AMSR-E product is incorporated into the last 6 years (2002–2008) of the ESA SM ECV data set. In order to take advantage of the entire 2002–2011 AMSR-E period, daily LSM products were extended through October 2011 by utilizing Noah, Mosaic, VIC and SAC simulations generated as part of the real-time NLDAS-2 activity.

Prior to the application of the IV approach introduced in section 2, all surface soil moisture data sets were transformed into climatological anomalies via the subtraction of a soil moisture seasonal climatology. This product-specific seasonal climatology was based on averaging surface soil moisture estimates acquired during all years of the data record within a 31 day time window centered on a particular day-of-year. In this way, a separate climatological expectation was calculated for each day-of-year for each soil moisture product. This climatological expectation was then subtracted from the raw soil moisture analysis to construct a time series of daily soil moisture anomalies relative to a fixed seasonal climatology. This transformation was performed for two reasons. First, past research has demonstrated that the application of IV-type approaches to soil moisture data sets is more effective if climatological expectations are removed from time series data [*Miralles et al*., 2010; *Draper et al*., 2012]. Second, a large number of soil moisture applications (including agricultural drought monitoring) are based on the analysis of climatological anomalies [see e.g., *Bolten and Crow*, 2012].

## 4. Impact of *L*

Prior to the estimation of $\mathbf{C}$, $\lambda_i$ and $\mathrm{Var}[\epsilon_{\mathbf{x}}^*]$, it is instructive to consider the role of the time-lag $L$ (days) applied in the lagged-variable analysis. Figure 2 illustrates the impact of $L$ for the case of applying (22) to obtain $\tilde{\mathbf{C}}$ using the multisensor ESA SM ECV data set and surface soil moisture estimates acquired from the four individual NLDAS-2 LSMs. Plotted values are spatial medians for all 0.25° pixels in the North American modeling domain shown in Figure 1. As described above, two opposing considerations must be taken into account when selecting $L$. Low values of $L$ increase $\mathrm{E}[\eta\eta_L^T]$ and therefore bias in estimated components of the $\tilde{\mathbf{C}}$ matrix. This is seen in Figure 2b where median estimates for diagonal components of the $\mathbf{C}$ matrix tend to decrease as a function of $L$. However, due to the irregular temporal sampling of the ESA SM ECV data set, significantly more data pairs are available for the $L = 0$ case than for the $L > 0$ (days) cases (Figure 2a). In addition, smaller values of $L$ reduce sampling uncertainly in sampled covariance estimates required for the calculation of scaling coefficients and $\tilde{\mathbf{C}}$ components. To examine this issue, a 100-member boot-strapping analysis was conducted to estimate such sampling uncertainty. As expected, the median sampling uncertainty in diagonal components of the $\tilde{\mathbf{C}}$ matrix steadily increases as a function of $L$ (Figure 2c).

As described above, the optimal interplay between these opposing consequences of increased $L$ depends on the specific goal of the analysis. Since derived $\lambda_i$ values are not biased by nonzero $\mathrm{E}[\eta\eta_L^T]$, the primary consideration in the estimation of $\lambda_i$ should be the reduction of random sampling error. In addition, there is no restriction on the use of multiple (integer) values of $L$. To examine this possibility, sampling uncertainty in $\tilde{\mathbf{C}}$ for each integer value of $L$ in the range $0 \leq L \leq 5$ (days) were estimated using a boot-strapping approach (applied separately to each LSM within a particular 0.25° spatial pixel) and then averaged across all four models to minimize sampling error. Next, individual components of $\tilde{\mathbf{C}}$ (for each choice of $L$) were averaged into a single value using weighting derived from this boot-strapping uncertainty analysis and the application of (4–6).

Note this optimal weighting (across various integer values of $L$) naturally places more weight on lower values of $L$ since small $L$ tend to be associated with lower sampling errors (see Figure 2c). However, while $L = 0$ represents the *single* integer value of $L$ with the lowest sampling error (Figure 2c), sampling errors can be further reduced by aggregating results obtained for integer values of $L$ in the range $0 \geq L \leq 5$ (days) (Figure 2d).

In sharp contrast to $\lambda$, bias associated with nonzero $\mathrm{E}[\eta\eta_L^T]$ (and low $L$) will *not* cancel out when estimating $\mathrm{Var}[\epsilon_{\mathbf{x}}^*]$ (Figure 2b). Therefore, steps must be taken to minimize $\mathrm{E}[\eta\eta_L^T]$ by avoiding the use of small $L$. Here this is done by limiting $L$ to the truncated integer range $3 \leq L \leq 5$ (days) when calculating $\tilde{\mathbf{C}}$ for the purposes of obtaining $\mathrm{Var}[\epsilon_{\mathbf{x}}^*]$. Estimates for components of the $\tilde{\mathbf{C}}$ matrix associated with $L = 3$, 4 and 5 (days) were aggregated into a single value based on error-covariance information sampled from a 100-member bootstrapping analysis and (4–6). The resulting $\tilde{\mathbf{C}}$ matrix was then converted into $\mathrm{Var}[\epsilon_{\mathbf{x}}^*]$ via (7). The reduced sampling uncertainty for Noah results in Figures 2c and 2d is attributable to its selection as the scaling

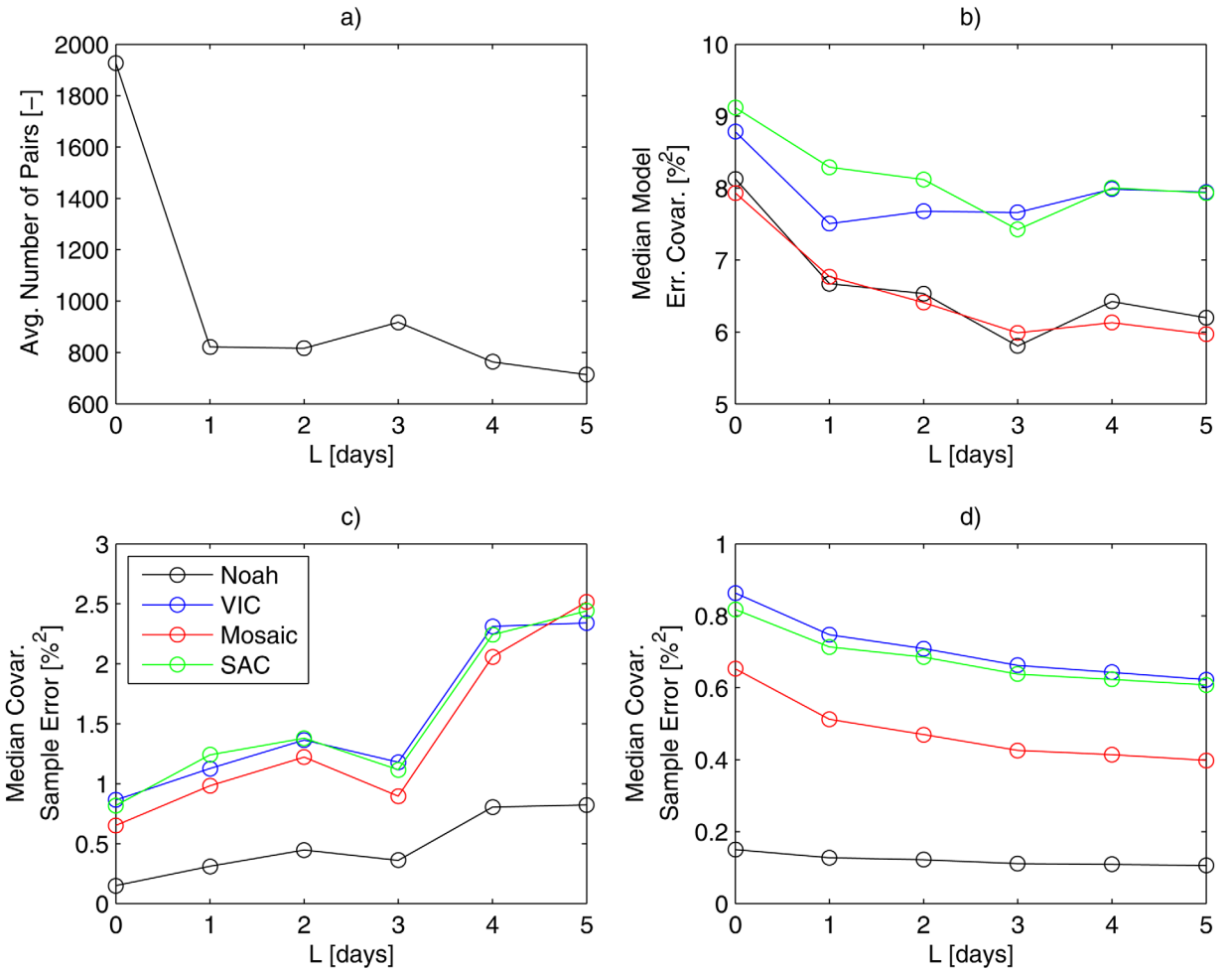



**Figure 2.** (a) Average number of available data pairs separated by an $L$-day lag in the 30 year ESA SM ECV data set, (b) median error covariance for each LSM (i.e., diagonal components of the $\tilde{\mathbf{C}}$ matrix) as a function of $L$, (c) median $1\sigma$ sampling error in error covariance estimates when utilizing a single value of $L$, and (d) median $1\sigma$ sampling error in error covariance estimates when utilizing a combination of all nonnegative lags $\leq L$ (days). The Y axis in Figures 2b–2d are plotted in units of squared volumetric soil moisture (VSM) percentages ($\%^2$).

reference in the analysis. As a result of this designation, it does not require the estimation of a re-scaling coefficient and therefore avoids the incremental increase in sampling uncertainty associated with the estimation of this coefficient.

## 5. Results

Based on (22), Figure 3 shows 0.25°-resolution maps spatial maps for each component of the $\tilde{\mathbf{C}}$ error covariance matrix (corresponding to Noah, VIC, Mosaic and SAC surface soil moisture predictions) derived via the application of the ESA SM ECV data set as the instrumental variable defined in (8). In order to minimize bias in $\tilde{\mathbf{C}}$, results in Figure 3 are based on the weighted averaging of $\tilde{\mathbf{C}}$ obtained for integer lags $3 \leq L \leq 5$ (days) (see section 3). As expected, the largest $\tilde{\mathbf{C}}$ components reflect model error variances along the matrix diagonal. However, generally positive error cross-covariances are also found in off-diagonal positions—suggesting the presence of error cross-correlation in predictions obtained from multiple LSMs. Along the diagonal of the matrix, Noah and Mosaic have the lowest error variances.

Spatial variations associated with political boundaries (e.g., the United States-Mexico border) are likely reflective of contrasts in the quality of precipitation products used to force the LSMs (and—potentially—trans-border changes in land cover/land use). Other geographic patterns can be used to glean information concerning the relative performance of each model. For example, Mosaic demonstrates relatively low error variance values in the third diagonal element of $\tilde{\mathbf{C}}$ in Figure 3—especially over the Western United States.

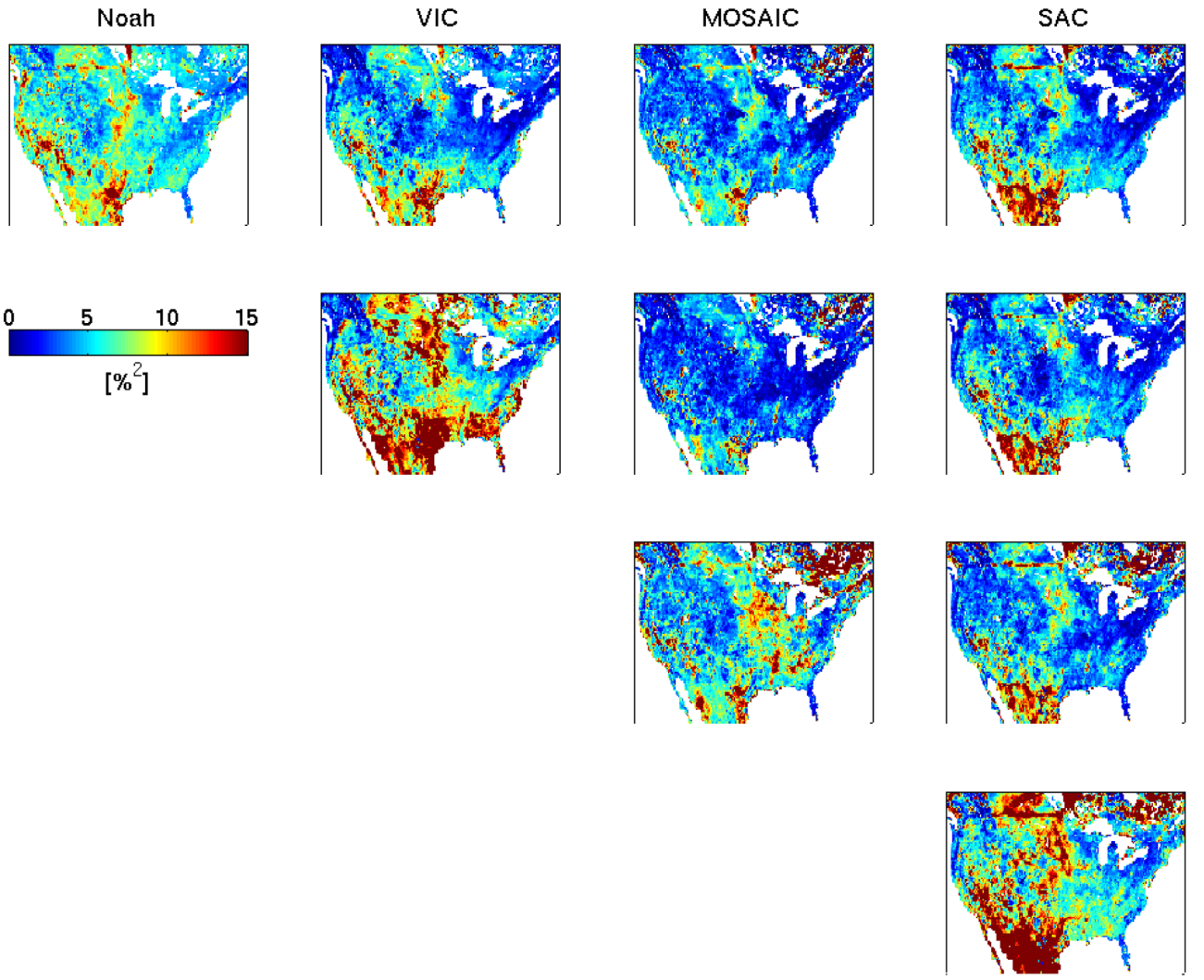



**Figure 3.** Spatially distributed components of the $\tilde{\mathbf{C}}$ matrix in units of squared volumetric soil moisture (VSM) percentages $(\%^2)$. To minimize bias, plotted results are based on the weighted averaging of sampled $\tilde{\mathbf{C}}$ matrix components for integer values $3 \leq L \leq 5$ (days).

This suggests that it does the best job of characterizing rapid surface soil moisture dynamics in relatively dry climates and, all other things being equal, should be weighted more heavily in these areas. In contrast, both VIC and SAC perform relatively worse in semiarid areas like Northern Mexico and in agricultural areas of the North Central United States and Southern Canada—perhaps reflecting a reduced emphasis on characterizing surface soil moisture and/or evapotranspiration dynamics in these models. The Noah model performs consistently well throughout the North American domain. However, relative to Mosaic, Noah also displays a relatively higher error cross-correlation with both the VIC and SAC models (especially in the Western United States). This tendency will have important implications for the calculation of optimal relative weights below.

Figure 4 plots $1\sigma$ sampling uncertainties for each $\tilde{\mathbf{C}}$ element derived from a 100-member boot-strapping analysis. Due to the relatively high number of data points available for sampling (Figure 1), and the efficient use of multiple $L$ (Figure 2d), absolute uncertainties in Figure 4 are low relative to the magnitude of sampled estimates in Figure 3. However, much larger sampling errors are found in areas exhibiting low skill in either the remotely sensed soil moisture product (e.g., Eastern Canada) or the LSM's (e.g., Mexico). In these areas, relatively low correlation between modeled and remotely sensed surface soil moisture products reduces the sampling efficiency of $\tilde{\mathbf{C}}$ estimates and leads to larger sampling uncertainties. In addition, the increased masking of retrievals over northern and/or heavily vegetated portions of the domain (Figure 1) further erodes the statistical power of our time series sampling. As discussed above, relatively lower sampling error

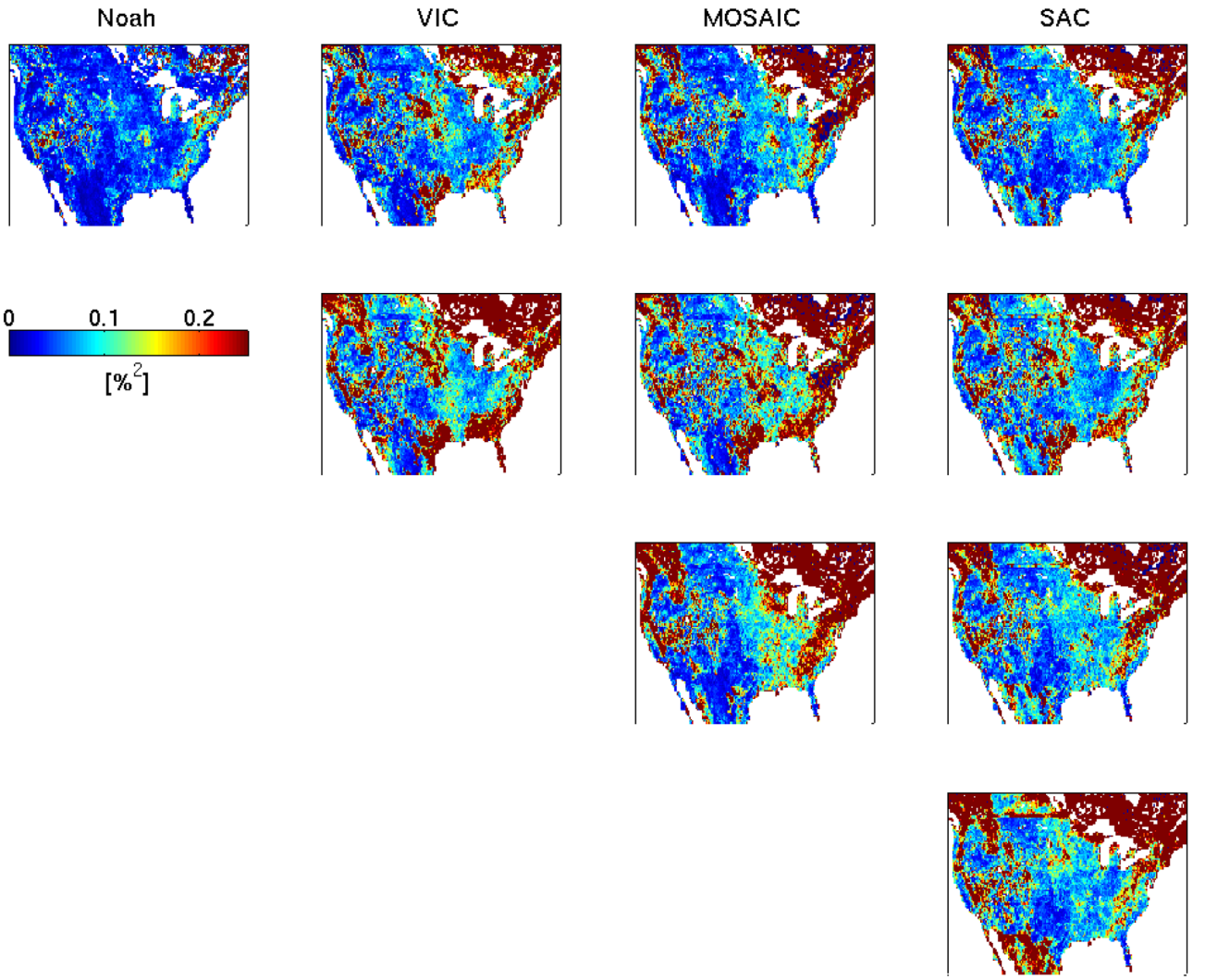



**Figure 4.** Standard sampling errors ($1\sigma$) in spatially distributed components of the $\tilde{\mathbf{C}}$ matrix in units of squared volumetric soil moisture (VSM) percentages ($\%^2$).

for the error variance of the Noah model is an expected artifact of its arbitrary selection as the reference data set during IV processing described in section 2.

Figure 5 plots optimal relative weights ($\lambda_i$) for all four LSMs calculated using (6). Note that, based on (23) and Appendix A, uniform bias in $\tilde{\mathbf{C}}$ will not propagate into $\lambda_i$ estimates. Therefore, sampled $\lambda_i$ in Figure 5 can be considered statistically unbiased. In addition, since bias in not an issue, the entire integer range of $L$ ($0 \leq L \leq 5$ (days)) can be utilized to minimize sampling error (section 3). As a result, the $\tilde{\mathbf{C}}$ values applied to estimate $\lambda_i$ are based on a different set of $L$ values than (bias-minimized) $\tilde{\mathbf{C}}$ results in Figure 3.

The consideration of off-diagonal components of the $\tilde{\mathbf{C}}$ axis has a number of interesting consequences for derived $\lambda_i$ fields plotted in Figure 5. For example, utilizing model error-cross correlation information via (6) leads to a recommendation of relatively high weighting for the Mosaic model in the Western United States. This emphasis on the Mosaic model is not readily apparent from consideration of only error variance information contained in the diagonal components of $\tilde{\mathbf{C}}$ (see Figure 3). However, it is underscored by the relative independence of Mosaic errors versus corresponding VIC and Noah errors. Therefore, significant weighting of the Mosaic model comes not just from its relatively low error variance but also its tendency to provide a nonredundant source of soil moisture anomaly information. The net impact of off-diagonal terms in $\tilde{\mathbf{C}}$ is summarized in Figure 6 which plots the difference between $\lambda_i$ calculated by applying (6) to the entire $\tilde{\mathbf{C}}$ matrix versus $\lambda_i$ calculated using only error variance information (i.e., forcing all off-diagonal components of $\tilde{\mathbf{C}}$ to zero prior to applying (6)). The magnitude of the observed differences in Figure 5 underscores the

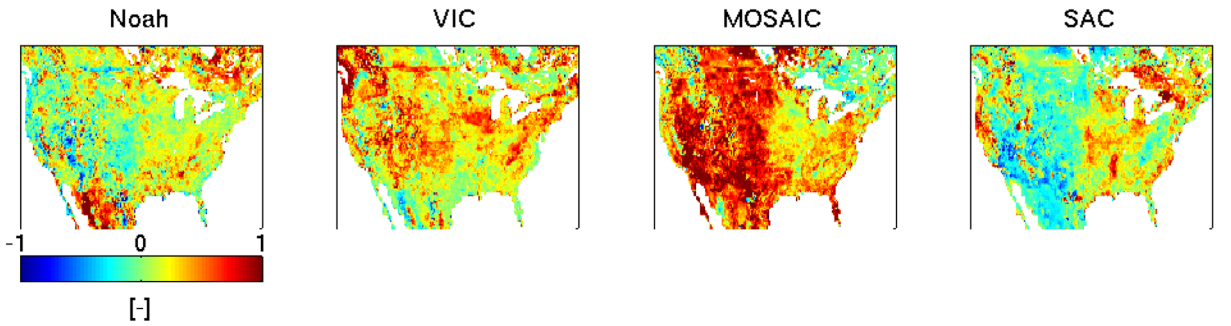



**Figure 5.** Dimensionless optimal relative weights $\lambda_i$ calculated for each model. To minimize sampling error, plotted results are based on the weighted averaging of sampled $\tilde{\mathbf{C}}$ matrix components for $0 \leq L \leq 5$ (days).

importance of error cross-covariances for the calculation of $\lambda_i$. In particular, it highlights the importance of accounting for off-diagonal $\tilde{\mathbf{C}}$ components on the relative weighting of Noah and SAC results in the Western United States (see above).

A second surprising result in Figure 5 is the (occasional) recommendation of negative weighting. While somewhat counter-intuitive, such weighting is a predicable consequence of a particular LSM having a combination of relatively low skill and high error cross-correlation with other (more accurate) LSMs (see Appendix B). The strongest example of this is the SAC model which performs relatively poorly in the Southwestern United States and exhibits errors which are highly cross-correlated with the other (more accurate) models (see the right-most column in Figure 3). As a result, limited areas of negative $\lambda_i$ values are observed for the SAC model in Figure 5. These negative $\lambda_i$ tend to require compensating values of $\lambda_i > 1$ for other LSMs in order to respect the constraint in (5).

Based on (7) and bias-minimized $\tilde{\mathbf{C}}$ results plotted in Figure 3, Figure 7 plots the error in surface soil moisture retrievals obtained via the optimal averaging of all four LSMs ($\mathrm{Var}[\epsilon^*_{\mathbf{x}}]$). Note that unlike $\lambda_i$, uniform bias in components of $\tilde{\mathbf{C}}$ will impact the estimation of $\mathrm{Var}[\epsilon^*_{\mathbf{x}}]$. Therefore, (7) is applied to $\tilde{\mathbf{C}}$ calculated using $L$ in the integer range $3 \leq L \leq 5$ (days) in order to avoid bias associated with the $L = 0$, 1 or 2 (days) cases. In general, estimated levels of $\mathrm{Var}[\epsilon^*_{\mathbf{x}}]$ are well below the individual error variances plotted in Figure 3 along the diagonal of the $\tilde{\mathbf{C}}$ matrix. This reduction reflects the reduced uncertainty associated with the optimal linear averaging of uncertain surface soil moisture predictions obtained from multiple LSMs.

## 6. Discussion

There are a number of issues which can potentially impact the objectivity of this analysis. First, it should be demonstrated that the selection of the Noah model as the reference scaling target in (3) does not skew key analysis results. Figure 8 examines this issue directly by comparing $\lambda_i$ results obtained via the (baseline) use of Noah as the scaling reference target (Figure 8a) to comparable results obtained by using the Mosaic model as the scaling reference (Figure 8b). Outside of very minor differences associated with sampling error

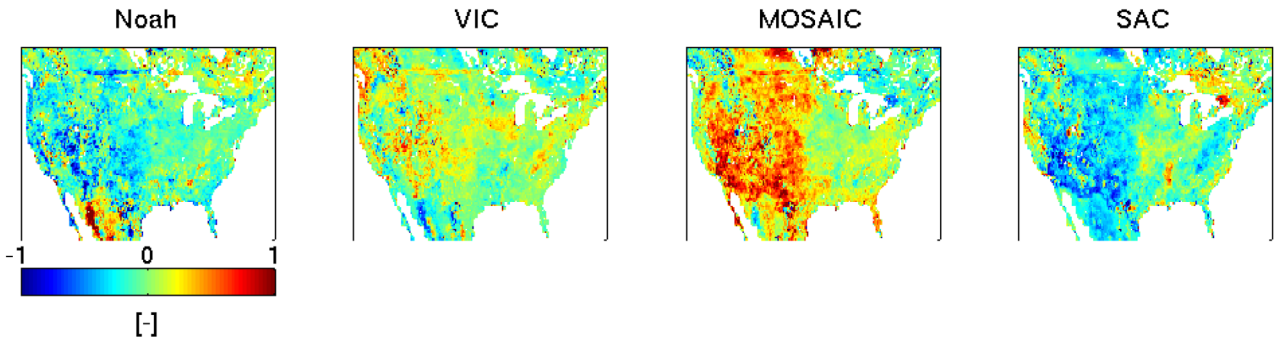



**Figure 6.** Net impact of off-diagonal components of $\tilde{\mathbf{C}}$ on the calculation of optimal modeling weights $\lambda_i$ via (6). To minimize sampling error, plotted results are based on the weighted averaging of sampled $\tilde{\mathbf{C}}$ matrix components for $0 \leq L \leq 5$ (days).

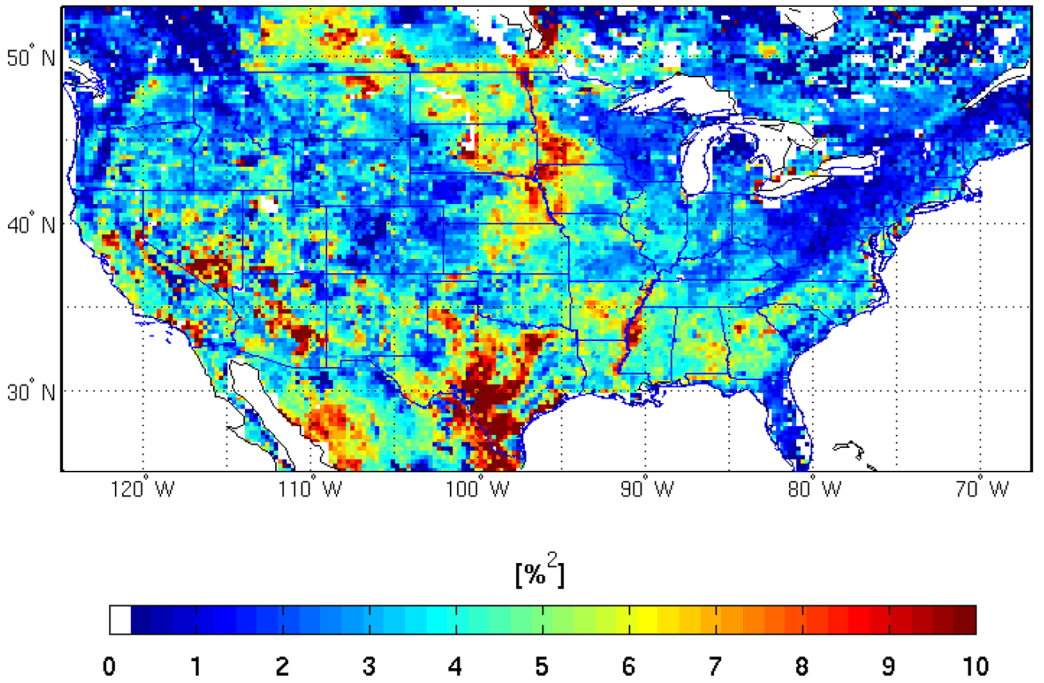



**Figure 7.** Spatially distributed plot of error variance in optimally averaged ensemble output ($\mathrm{Var}[\epsilon_x^*]$) in units of squared volumetric soil moisture (VSM) percentages ($\%^2$). To minimize bias, plotted results are based on the weighted averaging of sampled $\bar{\mathbf{C}}$ matrix components for $3 \leq L \leq 5$ (days) only.

incurred during the rescaling process, the two $\lambda_i$ results are essentially identical. Similar results are found when utilizing either SAC or VIC as the scaling reference model (not shown) Therefore, the requirement to establish an arbitrary reference LSM for rescaling purposes has little or no impact on the estimation of optimal scaling weights.

The 30 year ESA SM ECV product is attractive for this application since the sampling efficiency of IV calculations in section 2 is relatively low. As a result, long data sets are required to reduce sampling errors down to acceptable levels. However, another area of concern is that the multisensor, multisatellite nature of the ESA SM ECV data set may compromise its value as an IV.

One aspect of this issue is the presence of large temporal variations in the temporal frequency, spatial resolution, and accuracy of multisensor soil moisture retrievals used to construct the ESA SM ECV product. In general, soil moisture retrievals become more frequent, finer-resolution and more reliable in later years of the ESA SM ECV data set [*Wagner et al*., 2012]. Therefore, it should be clarified that the IV analysis presented in section 2 does not require that soil moisture retrievals have temporally homogeneous error statistics or are obtained at regular temporal intervals. However, sparse temporal sampling in the ESA SM ECV data set will likely exacerbate sampling challenges associated with IV calculations since larger temporal sampling gaps in $\mathbf{z}$ estimates will weaken the basis for applying $\mathbf{z_L}$ as an instrumental variable.

A related concern is that the use of Noah LSM soil moisture output to systematically rescale and combine various remotely sensed soil moisture products into the unified ESA SM ECV data set [*Liu et al*., 2011, 2012]. This use of Noah LSM output to process the ESA SM ECV data set raises obvious concerns about the LSM/remote sensing independence assumption asserted in (9). One potential problem is that the use of Noah as a scaling target could lead to the spurious favoring of Noah by the IV-analysis. However, such a bias is not readily apparent in $\lambda_i$ results (Figure 5). One potential reason for this is that Noah LSM output is used only to modify the long-term climatology characteristics of the ESA SM EV data set and has little impact on its relative temporal characteristics [*Loew et al*., 2013]. As a result, the anomaly based focus of this particular analysis (in which the long-term climatology of each soil moisture product is explicitly removed) will partially inoculate it against the possible cross-contamination of the ESA SM ECV data set by Noah-based climatological information. Nevertheless, some residual impact of this rescaling on results cannot be completely

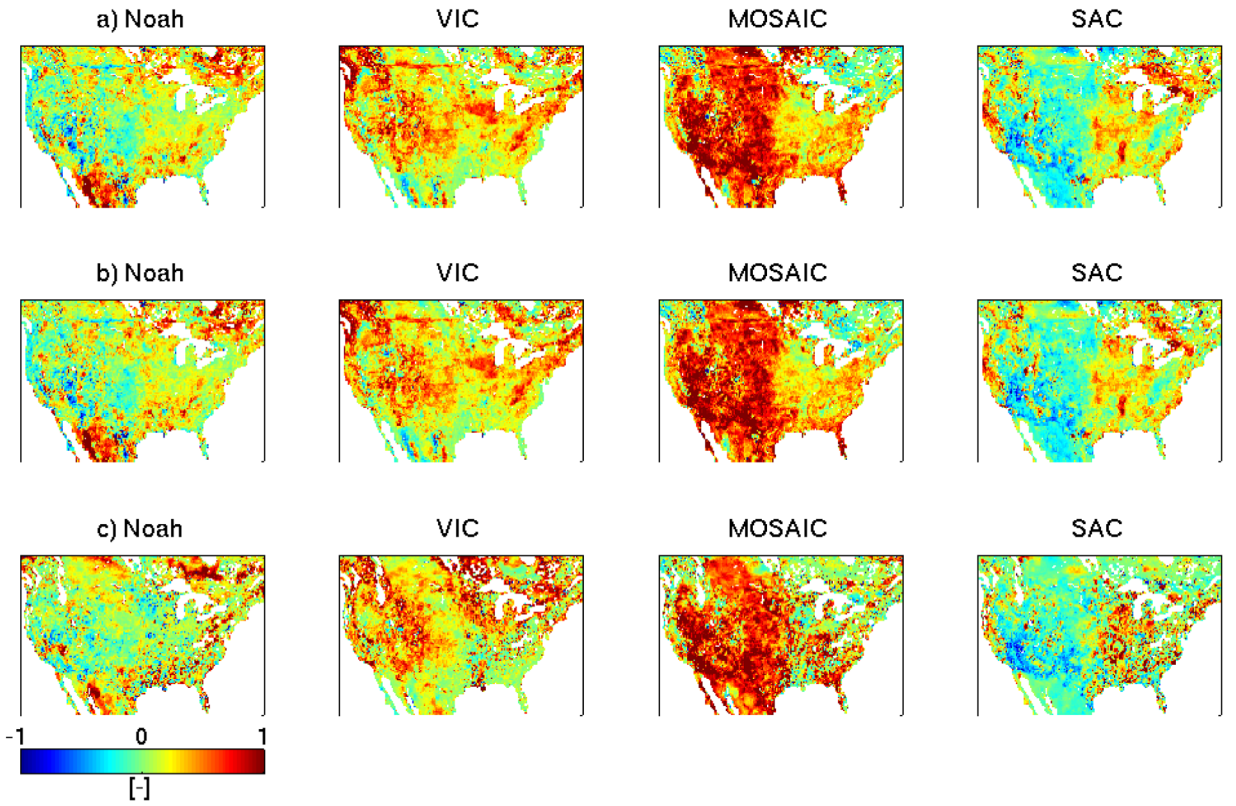



**Figure 8.** Dimensionless optimal relative weights $\lambda_i$ calculated for each model based on: (a) the baseline case of applying ESC ECV SM as the IV and Noah as the scaling reference (top row), (b) modifying the baseline case to use Mosaic as the scaling reference (middle row), and (c) modifying the baseline case by using the LPRM AMSR-E SM data set as the IV (bottom row).

ruled yet. Another concern is the presence of systematic shifts in the soil moisture time series associated with changes in rescaling strategy (occurring during transitions between the availability of various satellite-derived data products) [*Loew et al*., 2013]. By preventing the calculation of a stable soil moisture climatology, such shifts could imperil the orthogonality of errors in (1) and produce cases in which modeling errors are cross-correlated with true soil moisture.

The aggregate impact of these issues on the use of the ESA SM ECV data set is difficult to evaluate based solely on theoretical considerations. A more effective approach is empirically comparing results obtained from the multisensor ESA SM ECV data set to comparable results obtained via the use of a single-sensor product. To this end, Figure 8c shows $\lambda_i$ results re-generated via the application of the single-sensor, descending-only, 9 year (2002–2011) LPRM AMSR-E soil moisture product. Note that since the LPRM AMSR-E product constitutes a major subcomponent of the 30 year ESA SM ECV data set [*Wagner et al*., 2012], these comparisons should isolate the impact of multisensor inconsistencies and Noah-based rescaling on the application of the ESA SM ECV data set as an LV.

Despite large difference in temporal extent, 2002–2011 LPRM AMSR-E and 1979–2008 ESA SM ECV results predict highly similar patterns in optimized weights (compare Figure 8a to Figure 8c). This implies that the multisensor, multisatellite nature of the ESA SM ECV data set does not systematically impact its application as an IV. The only significant difference noted between $\lambda_i$ results obtained from single-sensor AMSR-E results and those associated with the multisensor, multisatellite ESA SM ECV product is that the ESA SM ECV appears to contain less spatial speckling than the AMSR-E results. This is likely attributable to reduced sampling noise associated with the longer length of the ESA SM ECV data set (40 years versus only 9 years for AMSR-E).

Finally, it must be noted that, subtle differences in the vertical support of various soil moisture products will complicate the interpretation of results. The ESA SM ECV data set is based on satellite estimates with vertical measurement supports of around 1–3 cm—although the exact depth can vary significantly as a function of near-surface soil moisture content. In contrast, the top-most soil layer in the NLDAS-2 LSM ensemble is generally held fixed at a 10 cm depth in order to preserve the numerical stability of prognostic soil water

balance calculations. Therefore, subtle temporal memory differences in individual soil moisture time series may imperil the baseline assumption—captured in (1) and (8)—that all products can be linearly related to a single time-series representation of soil moisture truth.

In addition, even if these relative differences in vertical support can be ignored, the optimized weights plotted in Figure 5 will apply only to surface (0–10 cm) soil moisture predictions made by a LSM. It will take an additional leap of faith to also apply them to deeper—and potentially more valuable—root-zone soil moisture predictions. However, recent work has argued that the low-pass filtering of surface soil moisture estimates is sufficient for replicating the information content of root-zone soil moisture variations [*Albergel et al*., 2008; *Qiu et al*., 2014]. This implies that the informed weighting of LSM surface-zone information may be adequate for constraining root-zone soil moisture variations; however, this extension remains speculative and must be examined in greater detail through follow-on research.

## 7. Summary and Conclusions

Due to commonalities in model physics, parameterizations and forcing input, LSM output products are likely to contain mutually correlated errors when generated within a multimodel ensemble. Therefore, the proper interpretation of this ensemble requires knowledge of mutual error covariance information among various LSMs. The recent development of the long-term, remotely sensed ESA SM ECV data set provides a unique opportunity to better leverage multimodel ensemble predictions for the retrospective analysis of soil moisture anomaly conditions. Specifically, it allows for the application of a lagged variable IV approach to estimate model error covariance information required to produce an optimally weighted average of ensemble LSM soil moisture predictions via (6) and a bias-minimized estimation of merged analysis error via (7). Both of these estimates are critical for conditioning soil moisture likelihoods based on multimodel LSM ensemble output. For the case of Gaussian modeling errors, such conditioning will be identical to that provided by a full Bayesian merger of multimodel ensemble output.

Since the ESA SM ECV is based on the merger of both passive and active remote sensing products, and model products cannot be assumed to be mutually independent, combining models with the ESA SM ECV data set does not provide the necessary triplet of independent soil moisture estimates required for the application of a standard triple collocation analysis [see e.g., *Scipal et al*., 2008; *Draper et al*., 2012]. The approach presented here circumvents this limitation by expanding upon *Su et al*. [2014] and using a lagged-version of the ESA SM ECV data set as the third member of an instrumental variable (IV) triplet analysis aimed at estimating the entire error covariance matrix for surface soil moisture predictions obtained from a multimodel LSM ensemble (Figure 3). While this will cause biased error covariances estimates in the case of auto-correlated errors in the remotely sensed soil moisture time series (section 2), it can be shown that such biases are uniform across all elements of the covariance matrix and, therefore, will not induce comparable biases in estimates of optimal model weights (Appendix A and Figure 5). For applications in which biased error covariance estimates are of concern (e.g., the estimation of error variances for optimally averaged ensemble values), avoiding smaller temporal lags (i.e., less than about 2 days) appears to significantly reduce the magnitude of potential biases (Figures 2b). In addition, the use of a range of integer $L$ values up to 5 days is shown to reduce sampling uncertainty in error covariance estimates (Figure 2d).

While the multisensor, multisatellite nature of the 30 year ESA SM ECV data set is a theoretical concern (see section 6), comparisons with comparable results derived from the 2002–2009, single-sensor LPRM AMSR-E data set (Figure 8) suggest that the multisensor nature of the ESA ECV SM data set does not imperil its application as a robust IV.

As a result, a lagged-variable IV analysis appears well-suited for leveraging the long-term ESA SM ECV data set for improved retrospective analysis of LSM soil moisture products obtained from a multimodel ensemble. Nevertheless, it should be stressed that the presence of temporal discontinuities and the use of LSM-based rescaling in the current ECV SM ECV data set could still imperil assumptions at the heart of this approach (see section 6). Therefore, minimizing the future role of LSM-based rescaling in long-term soil moisture data sets and maximizing the temporal consistency of any applied re-scaling should remain a priority for the construction of future long-term, satellite-based soil moisture data sets. Finally, the approach presented here is also potentially attributable to other types of ensemble estimates derived from geophysical models provided that they are adequately auto-correlated and a (suitably independent) observational record is available.

## Appendix A: Inversion of a Uniformly Biased Covariance Matrix

This appendix contains a proof for the assertion in (23) that:

$$\lambda_i(\mathbf{C}) = \lambda_i(\mathbf{C} + \delta\mathbf{J}) \tag{A1}$$

is true for a general symmetric matrix $\mathbf{C}$ where $\mathbf{J}$ is a matrix of ones with dimensions of $\mathbf{C}$ and $\delta$ is an arbitrary real number. The proof will be based on a lemma from *Miller* [1981] which states:

$$(\mathbf{A}+\mathbf{B})^{-1} = \mathbf{A}^{-1} - \frac{1}{1+g}\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1} \tag{A2}$$

where $\mathbf{A}$ and $\mathbf{A}+\mathbf{B}$ are invertible matrices; $\mathbf{B}$ is rank 1, and $g = \mathrm{Trace}(\mathbf{B}\mathbf{A}^{-1})$. Note that, as a matrix containing only a single constant value, $\delta\mathbf{J}$ in (A1) is always rank 1.

For ease of notation, we rename the inverse of $\mathbf{C}$ as $\mathbf{D}$: So, based on (6), the left-hand-side of (A1) can be written as

$$\lambda_i(\mathbf{C}) = \frac{\sum_j D_{ij}}{\sum_j \sum_k D_{jk}}. \tag{A3}$$

We will now demonstrate the right-hand-side of (A1) is identical to (A3).

First, define $\mathbf{F}$ as the inverse of $\mathbf{C}+\delta\mathbf{J}$ and apply the lemma in (A2) to it:

$$\mathbf{F} = (\mathbf{C}+\delta\mathbf{J})^{-1} = \mathbf{C}^{-1} - \frac{\delta}{1+g}\mathbf{C}^{-1}\mathbf{J}\mathbf{C}^{-1}. \tag{A4}$$

Given that $\mathbf{J}$ is a matrix of ones, the components of $\mathbf{F}$ can be re-written as:

$$F_{ij} = D_{ij} - \eta\left(\sum_k D_{ik}\right)\left(\sum_l D_{jl}\right) \tag{A5}$$

where $\eta = \delta/(1+g)$. Also note that the application of (6) to $\mathbf{F}^{-1} = \mathbf{C}+\delta\mathbf{J}$ can be written as:

$$\lambda_i(\mathbf{C}+\delta\mathbf{J}) = \lambda_i(\mathbf{F}^{-1}) = \frac{\sum_j F_{ij}}{\sum_j \sum_k F_{jk}}. \tag{A6}$$

Next, substitute (A5) into the nominator of (A6):

$$\begin{aligned}\sum_j F_{ij} &= \sum_j D_{ij} - \eta\sum_j\left(\sum_k D_{ik}\right)\left(\sum_l D_{jl}\right)\\ &= \sum_j D_{ij} - \eta\sum_k D_{ik}\sum_j\left(\sum_l D_{jl}\right)\\ &= \left[1 - \eta\sum_j\sum_k D_{jk}\right]\sum_j D_{ij}.\end{aligned} \tag{A7}$$

Do the same for the denominator of (A6):

$$\begin{aligned}\sum_j\sum_k F_{jk} &= \sum_j\sum_k D_{jk}\\ &\quad -\eta\sum_j\sum_k\left(\sum_{l'} D_{jl'}\right)\left(\sum_l D_{kl}\right)\\ &= \sum_j\sum_k D_{jk} - \eta\left(\sum_j\sum_{l'} D_{jl'}\right)\left(\sum_k\sum_l D_{kl}\right)\\ &= \sum_j\sum_k D_{jk}\,1 - \eta\sum_l\sum_{l'} D_{ll'}.\end{aligned} \tag{A8}$$

Substituting (A7) and (A8) into (A6), we find that square bracket terms in (A7) and (A8) cancel out—leaving only:

$$\lambda_i(\mathbf{C}+\delta\mathbf{J})=\frac{\sum_j D_{ij}}{\sum_j\sum_k D_{jk}}. \tag{A9}$$

Comparing (A9) to (A3) establishes the equality asserted in (A1).

## Appendix B: Existence of Negative Optimal Weights

Consider an arbitrary 2 × 2 error covariance matrix **M** for soil moisture output obtained from two separate LSMs:

$$\begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

By definition of a covariance matrix:

$$ac\rho=b^2 \tag{B1}$$

where $0 \leq \rho \leq 1$ is a cross-correlation coefficient for errors in the two models (assumed here to be nonnegative). Solving (B1) for $b$ and plugging it into (6) gives the following expression for the optimal weighting of the first model:

$$\lambda_1(\mathbf{M})=\frac{c-\sqrt{ac\rho}}{c+a-2\sqrt{ac\rho}}. \tag{B2}$$

If $a$, $b$ and $\rho$ are positive real numbers, then the numerator of (B2) will be negative if $a > c/\rho$; however, the denominator of (B2) will *always* be positive (except for the uninteresting case where $a = c$ and $\rho = 1$). Therefore, (B2) will predict negative weights when $a > c/\rho$. This condition cannot be met for noncorrelated errors ($\rho = 0$); however, it is readily met in cases when one retrieval is considerably worse than the other (i.e., $a \gg c$), and $\rho$ is relatively large. Analogous, although more complex, conditions exist for higher dimensional cases.

**Acknowledgments**
The NLDAS-2 data sets used in this analysis can be downloaded from http://www.emc.ncep.noaa.gov/mmb/nldas. The ESA ECV SM data set can be obtained at http://www.esa-soilmoisture-cci.org.

## References

Albergel, C., C. Rudiger, T. Pellarin, J. C. Calvet, N. Fritz, F. Froissard, D. Suquia, A. Petitpa, B. Piguet, and E. Martin (2008), From near-surface to root-zone soil moisture using an exponential filter: An assessment of the method based on in-situ observations and model simulations, *Hydrol. Earth Syst. Sci.*, *12*, 1323–1337, doi:10.5194/hess-12-1323-2008.

Bohn, T. J., M. Y. Sonessa, and D. P. Lettenmaier (2010), Seasonal hydrologic forecasting: Do multi-model ensemble averages always yield improvements in forecast skill?, *J. Hydrometeorol.*, *11*(6), 1357–1371, doi:10.1175/2010JHM1267.1.

Bolten, J. D., and W. T. Crow (2012), Improved prediction of quasi-global vegetation conditions using remotely-sensed surface soil moisture, *Geophys. Res. Lett.*, *39*, L19406, doi:10.1029/2012GL053470.

Bowden, R. J., and D. A. Turkington (1990), *Instrumental Variables*, Cambridge Univ. Press, Cambridge, N. Y.

Cosgrove, B. A., et al. (2003), Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project, *J. Geophys. Res.*, *108*(D22), 8842, doi:10.1029/2002JD003118.

Crow, W. T., and M. J. van den Berg (2010), An improved approach for estimating observation and model error parameters for soil moisture data assimilation, *Water Resour. Res.*, *46*, W12519, doi:10.1029/2010WR009402.

Crow, W. T., S. V. Kumar, and J. D. Bolten (2012), On the utility of land surface models for agricultural drought monitoring, *Hydrol. Earth Syst. Sci.*, *16*, 3451–3460, doi:10.5194/hess-16-3451-2012.

Daly, C., R. P. Neilson, and D. P. Phillips (1994), A statistical-topographic model for mapping climatological precipitation over mountainous terrain, *J. Appl. Meteorol.*, *33*, 140–158.

Draper, C., R. Reichle, R. de Jeu, V. Naeimi, R. Parinussa, and W. Wagner (2012), Estimating root mean square errors in remotely sensed soil moisture over continental scale domains, *Remote Sens. Environ.*, *137*, 288–298.

Gao X., and P. A. Dirmeyer (2006), A multimodel analysis, validation, and transferability study of global soil wetness products, *J. Hydrometeorol.*, *7*, 1218–1236.

Liu, Y. Y., R. M. Parinussa, W. A. Dorigo, R. A. M. de Jeu, W. Wagner, A. I. J. N. van Dijk, M. F. McCabe, and J. P. Evans (2011), Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals, *Hydrol. Earth Syst. Sci.*, *15*, 425–436, doi:10.5194/hess-15-425-2011.

Liu, Y. Y., W. A. Dorigo, R. M. Parinussa, R. A. M. de Jeu, W. Wagner, M. F. McCabe, J. P. Evans, and A. I. J. M. van Dijk (2012), Trend-preserving blending of passive and active microwave soil moisture retrievals, *Remote Sens. Environ.*, *123*, 280–297, doi:10.1016/j.rse.2012.03.014.

Loew, Q., T. Stackle, W. Dorigo, R. de Jeu, and S. Hageman (2013), Potential and limitations of multidecadal satellite soil moisture observations for select climate model evaluation studies, *Hydrol. Earth Syst. Sci.*, *17*, 3523–3542, doi:10.5194/hess-17-3523-2013.

Miller, K. S. (1981), On the inverse of the sum of matrices, *Math. Mag.*, *54*(2), 67–72.

Miralles, D. G., W. T. Crow, and M. H. Cosh (2010), Estimating spatial sampling errors in coarse-scale soil moisture estimates derived from point-scale observations, *J. Hydrometeorol.*, *11*(6), 1404–1410, doi:10.1175/2010JHM1285.1.

Owe, M., R. A. M. De Jeu, and T. R. H. Holmes (2008), Multi-sensor historical climatology of satellite-derived global land surface moisture, *J. Geophys. Res.*, *113*, F01002, doi:10.1029/2007JF000769.

Perrons, M. P., and L. N. Cooper (1992), When networks disagree: Ensemble methods for hybrid neural networks, in *Neural Networks for Speech and Image Processing*, edited by R. J. Mammone, pp. 126–142, Chapman and Hall, N. Y.

Qiu, J., W. T. Crow, G. S. Nearing, X. Mo, and S. Liu (2014), The impact of vertical measurement depth on the information content of soil moisture times series data, *Geophys. Res. Lett.*, *41*, 4997–5004, doi:10.1002/2014GL060017.

Scipal, K., T. Holmes, R. A. M. de Jeu, V. Naeimi, and W. Wagner (2008), A possible solution for the problem of estimating the error structure of global soil moisture data sets, *Geophys. Res. Lett.*, *35*, L24403, doi:10.1029/2008GL035599.

Sheffield, J., et al. (2012), Drought monitoring with the North American Land Data Assimilation System (NLDAS): A framework for merging model and satellite data for improved drought monitoring, in *Remote Sensing of Drought: Innovative Monitoring Approaches*, edited by B. Wardlow, M. Anderson, and J. Verdin, 270 pp., Taylor and Francis, London.

Stoffelen, A. (1998), Towards the true near surface wind speed: Error modeling and calibration using triple collocation. *J. Geophys. Res.*, *103*(C4), 7755–7766.

Su, C.-H., D. Ryu, W. T. Crow, and A. W. Western (2014), Beyond triple collocation: Applications to satellite soil moisture, *J. Geophys. Res. Atmos.*, *119*, 6419–6439, doi:10.1002/2013JD021043.

Vogelzang, J., and A. Stoffelen (2012), Triple collocation, *EUMETSAT Report*. [Available at https://nwpsaf.eu/deliverables/scatterometer/TripleCollocation_NWPSAF_TR_KN_021_v1_0.pdf.]

Wagner, W., W. Dorigo, R. de Jeu, D. Fernandez, J. Benveniste, E. Haas, and M. Ertl (2012), Fusion of active and passive microwave observations to create an Essential Climate Variable data record on soil moisture, ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS Annals), vol I-7, pp. 315–321, XXII ISPRS Congress, Melbourne, Australia, 25 August–1 September 2012.

Wang A., T. J. Bohn, S. P. Mahanama, R. D. Koster, and D. P. Lettenmaier (2009), Multimodel ensemble reconstruction of drought over the continental United States, *J. Clim.*, *22*, 2694–2712, doi:10.1175/2008JCLI2586.1.

Xia, Y., et al. (2012a), Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products, *J. Geophys. Res.*, *117*, D03109, doi:10.1029/2011JD016048.

Xia, Y., et al. (2012b), Continental-scale water and energy flux analysis and validation for North American Land Data Assimilation System project phase 2 (NLDAS-2): 2. Validation of model-simulated streamflow, *J. Geophys. Res.*, *117*, D03110, doi:10.1029/2011JD016051.

Xia, Y., M. B. Ek, D. Mocko, C. D. Peters-Lidard, J. Sheffield, J. Dong, and E. F. Wood (2014), Uncertainties, correlations, and optimal blends of drought indices from the NLDAS multiple land surface model ensemble, *J. Hydrometeorol.*, *15*, 1636–1650, doi:10.1175/JHM-D-13-058.1

Yilmaz, M. T., and W. T. Crow (2013), The optimality of potential rescaling approaches in land data assimilation, *J. Hydrometeorol.*, *14*, 650–660, doi:10.1175/JHMD12052.1.

Yilmaz, M. T., and W. T. Crow (2014), Evaluation of assumptions in soil moisture triple collocation analysis, *J. Hydrometeorol.*, *15*, 1293–1302, doi:10.1175/JHM-D-0158.1.