Optimal design of surface networks for observation of soil moisture

Konstantin Y. Vinnikov
Department of Meteorology, University of Maryland, College Park

Alan Robock
Department of Environmental Sciences, Rutgers University, New Brunswick

Shuang Qiu\(^1\) and Jared K. Entin\(^2\)
Department of Meteorology, University of Maryland, College Park

Abstract. By analyzing in situ soil moisture data, we show that soil moisture variability consists of two components, one of which is related to large-scale atmospheric forcing, and the other related to small-scale land surface variability and hydrologic processes. We use empirically estimated spatial autocorrelation functions for Illinois to estimate errors of spatial averaging of soil moisture observations, using the method of statistically optimal averaging of meteorological fields. The estimated dependence of the root-mean-square errors of averaging on the soil moisture station network density can be used to analyze existing observational networks and for designing new ones. For the application of providing information on a regular grid for numerical models of weather and climate, we show that the new, relatively high density networks of soil moisture observations in Oklahoma, may not provide estimates with very much more accuracy than the relatively low density currently operational network in Illinois. This prediction must be tested when we receive sufficiently long time series of observations from Oklahoma.

1. Introduction

In this paper we show how long term observations of soil moisture in one region can be used to evaluate or improve the networks of soil moisture stations now being created in other regions with similar climatic conditions. The same approach can be used for every other meteorological variable. The traditional requirement for the density of meteorological stations is that the random error of spatial interpolation of the observed variable at the point farthest from the other stations should not exceed some critical value. This approach was first used by Drozdov and Sheplevskiy [1946], based on linear interpolation, and later by Gandin [1963], based on the optimal interpolation technique. The same principles, in combination with the statistically optimal averaging method, were applied to optimize snow course observations by Lysikhman and Kagan [1960] and Kagan [1979]. Vinnikov [1967, 1970] used estimates of random errors of optimal averaging to develop scientific criteria for the density of a global network of radiation stations.

\(^1\)Now at NOAA National Environmental Satellite, Data, and Information Service, Camp Springs, Maryland.

\(^2\)Now at NASA Goddard Space Flight Center, Greenbelt, Maryland.

Copyright 1999 by the American Geophysical Union.

Paper number 1999JD900060.

0148-0227/99/1999JD/000060S09.00

There are two principally different techniques for large-scale long-term monitoring of soil moisture variations. The future obviously belongs to remote sensing of soil moisture from satellites. The spatial resolution of such observations may be optimized if we use information on the statistical structure of the soil moisture field [Vinnikov et al., 1999].

Currently, the traditional technique for soil moisture monitoring is based on networks of surface observational stations. The first such networks were created many decades ago in the former Soviet Union (FSU) and later in a few surrounding countries, including China and Mongolia [Vinnikov and Yererkepova, 1991; Robock et al., 1998; Entin et al., 1999]. On average, the distance between soil moisture stations in the FSU was about 85 km (~3000 stations in a land area of 22,400,000 km²), but the network was much denser in the European part of the country than in Siberia. China does not run its soil moisture stations as a united network, so a network density estimate would not be meaningful. We analyzed data of 43 Chinese stations [Entin et al., 1999], but China has many more soil moisture stations, the data of which are still unavailable to the international scientific community. The average distance between soil moisture stations in Mongolia (40 stations in an area of 1,565,000 km²) is about 200 km [Erdenetsogt, 1996]. The best North American network of soil moisture stations was created by the Illinois State Water Survey in 1981 [Hollinger and Isard, 1994]. The average distance between soil moisture stations in Illinois is about 93 km. The Oklahoma Mesonet soil moisture network that is currently being created has 60 stations; 52 more Mesonet stations are scheduled for installation and calibration.
of soil moisture sensors [Brock et al., 1995; Basara et al., 1998; Elliott et al., 1998]. The network with 60 stations now and another 52 stations in the future in an area of 182,000 km² corresponds to an average distance between stations of about 55 km now and 35 km in the future. In addition, there are a few independent observational programs for soil moisture in Oklahoma, including observations at the Department of Energy Atmospheric Radiation Measurement program Cloud and Radiation Test site. This is an example of a very dense network of soil moisture stations.

Observations from individual stations in the network in the FSU and Russia are usually spatially averaged for particular catchments or administrative districts [Meshcherskaya et al., 1982; Zhukov, 1986; Kelchekaya, 1989] If soil moisture information is to be used as input into a weather prediction model, the observed data must be averaged over the model grids. These grids can be as small as 10 km for some mesoscale models and as large as a few hundred kilometers for global models.

It is obvious that the denser the network of observational stations in a region the smaller the errors of spatial averaging of the observed data. To estimate the root-mean-square errors of spatial averaging of observed soil moisture, information about the statistical structure of the soil moisture field must be used. These errors are a potential source of errors in meteorological, hydrological, and agricultural predictions. As a network becomes denser, after some limit is reached, further increase of the network density will not decrease the errors of predictions. If this were the main use for these observations, economical considerations would stop us from further increasing the network density in the region. Here we evaluate this limit for Illinois and show that further increasing of the number of soil moisture stations in that state would not noticeably increase the accuracy of soil moisture information for mesoscale and global weather prediction models.

In this paper we first explain the statistical technique of optimal averaging that we use for the analysis of soil moisture observations. Next, we present our analysis of the separation of scales of spatial variation of soil moisture. Then we apply the theory to observations in the state of Illinois, and use it to compare with a new network of soil moisture observations in the state of Oklahoma.

2. Optimal Averaging Technique

The optimal averaging technique [Kagan, 1979] has been shown to be efficient and convenient for regional and global averaging of meteorological observations [Vinnikov and Lugina, 1982; Vinnikov et al., 1991; Smith et al., 1994]. It has recently been successfully applied for regional averaging of soil moisture observations [Robock et al., 1998; Entin et al., 1999; Vinnikov et al., 1999].

The technique for optimal averaging over the area S of observations at n stations may be expressed by very simple formulae. Let \( f_i \) (i=1, ..., n) be the observed soil moisture at n stations. These stations may be inside and outside of the area. We assume that the field of anomalies

\[
 f'_i = f_i - \bar{f}
\]

is homogeneous and isotropic and has variance \( \sigma_f^2 \). The covariance function \( R(d_{ij}) \) is

\[
 R(d_{ij}) = \frac{\bar{f}_i \bar{f}_j}{\bar{f}^2} = \sigma_f \rho(d_{ij}),
\]

where \( \rho(d_{ij}) \) is the normalized autocorrelation function and \( d_{ij} \) is the distance between stations i and j. The spatial average \( F'_s \) of a variable \( f(x, y) \) over area S is

\[
 F'_s = \frac{1}{S} \int_S f(x, y) \, dx \, dy
\]

and can be approximated as a linear combination of the observed values at n stations

\[
 F'_s = \sum_{i=1}^{n} p_i f'_i.
\]

The optimal weights \( p_i \) (i=1, ..., n) should be determined by minimizing the variance of the random error of approximation (4)

\[
 \varepsilon^2 = \left( F'_s - \sum_{i=1}^{n} p_i f'_i \right)^2 = \min.
\]

The expressions for the weights and \( \varepsilon^2 \) are

\[
 \sum_{i=1}^{n} p_i R_p + \rho^2 = \frac{f'_s}{F'_s},
\]

\[
 \varepsilon^2 = \left( F'_s \right)^2 - \sum_{i=1}^{n} p_i F'_s.
\]

The covariance between local \( f'_i \) and spatially averaged variables \( F'_s \) and the variance of the averages, \( \bar{F}'_s \) and \( (\bar{F}'_s)^2 \) can be calculated preliminarily as multiple integrals of the autocorrelation function. Here \( \rho^2 \) is the variance of the random error of observation. Kagan [1979] describes the theory and technique of such calculations in detail. This technique cannot be used for observations of a variable without information on the statistical structure of that variable.

3. Statistical Structure of Soil Moisture

Meshcherskaya et al. [1987] used long-term observational data from the FSU to study the spatial autocorrelation functions of soil moisture for agricultural fields. Vinnikov et al. [1996, 1999], making additional interpretation of these estimates and original estimates of spatial autocorrelation in the soil moisture field in Illinois, showed that the spatial covariance function \( R(d) \) of a soil moisture field may be expressed as

\[
 R(d) = \sigma_f^2 \exp(-d/L_v) + \sigma_v^2 \exp(-d/L_v).
\]

where \( d \) is distance, \( \sigma_f^2 \) and \( L_v \) are the variance and scale of the land-surface-related component of variability, and \( \sigma_v^2 \) and \( L_v \) are the variance and scale of the soil moisture variability related to atmospheric forcing. This is illustrated in a log-linear graph in Figure 1, where the slope of the line labeled "meteorological scale" corresponds to \( L_v \). the slope of the line labeled "catchment hydrological scale" corresponds to \( L_v \).

According to empirical estimates for the European part of the FSU, the scale \( L_v \) depends on the time of year and depth of the soil layer (upper 20 cm or upper 1 m) and varies in the range of 530-880 km [Meshcherskaya et al., 1982].

Meshcherskaya et al. [1982] also showed that the dependence of the spatial autocorrelation on direction for distances up to a few hundred kilometers is rather weak and may be ignored. The field of standardized anomalies of soil
moisture may be considered to be relatively homogeneous and isotropic with respect to its spatial autocorrelation function. This means that the autocorrelation depends on the distance between points but not on their relative direction.

As opposed to \( L_{a} \), the other scale, \( L_{r} \), reflects the spatial variability of topography, soil properties, and vegetation and is much smaller, equal to about 10-20 m [Vaucaud et al., 1985]. The land-surface-related component of spatial variability of a soil moisture field may be interpreted as white noise added to the atmosphere-related signal that corresponds to a red noise statistical model. The variances of these two components are approximately the same order of magnitude, but their ratio \( \sigma_{r}^{2}/\sigma_{a}^{2} \) may vary quite significantly depending on the complexity of the landscape.

Vinnikov et al. [1996] showed, for three small catchments (with areas 0.015 km\(^2\), 0.36 km\(^2\), and 0.45 km\(^2\) and average distances between stations of 40, 180, and 200 m, respectively) at the Valday Research Water Balance Station in Russia, that the spatial autocorrelation coefficients between 9 and 11 points inside each of the catchments do not depend noticeably on the distances between the points. These observations are already at the “meteorological scale.” Only for much smaller catchments and much smaller distances between soil moisture observations should the “hydrological scale” of spatial autocorrelation in soil moisture be taken into account. Therefore for all practical purposes we may consider that \( L_{a} \gg L_{r} \rightarrow 0 \) and that \( \sigma_{r}^{2} \) may be interpreted as the variance of random errors of soil moisture measurements.

For soil moisture observations in Illinois, spatial autocorrelation functions of soil moisture estimated for the upper 1 m and 10 cm soil layers by Vinnikov et al. [1999] are shown in Figure 2. According to these estimates the part of the variance related to the white noise component for both soil layers is about 30-35%. The observed standard deviations of soil moisture of the upper 10 cm and 1 m layers, \( \sigma_{a}^{2} + \sigma_{r}^{2} \), are 0.83 and 4 cm of plant available water respectively. In the analysis that follows we will use the following parameters for the statistical structure of the soil moisture field in Illinois.

For the top 10 cm soil layer

\[
\sigma_{a} = 0.7 \text{ cm}, \quad L_{a} = 435 \text{ km}, \quad \delta^{2}/\sigma_{a}^{2} = 0.48.
\]

Analogous estimates of the parameters of spatial statistical structure of the soil moisture field for other climatic conditions, in Russia, China, and Mongolia, may be found in the work of Entin [1998]. These estimates may be used for optimal spatial averaging of soil moisture observations and for receiving a priori estimates of errors of spatial averaging.

We do not discuss here the problem of temporal variability of soil moisture. It also may be divided in two components, one of which is related to the atmospheric forcing and the other component related to short-term hydrological processes (infiltration, surface runoff, gravitational drainage) [Entin, 1998]. Atmospheric forcing is responsible for the long-term component of soil moisture variability with a scale of the order of a few months. In comparison with this scale, the other component may be interpreted as the random error of observation. More details on this subject may be found in the works of Delworth and Manabe [1988], Vinnikov and Yereskepova [1991], Robock et al. [1995], Vinnikov et al. [1996, 1999], and Entin [1998]. This information may be used to optimize the temporal frequency of soil moisture observations for different users. A detailed analysis is beyond the scope of this paper.

To monitor the meteorological component of soil moisture variability with a timescale of a few months, Russians make soil moisture observations 3 times per month, every 8-11 days. The shift of 1-3 days between times of observation at each station is assumed to be negligible. The Illinois soil moisture network has the same approach. To monitor the full spectrum of soil moisture variability, including their meteorological and hydrological components, the Oklahoma Mesonet makes soil moisture measurements every 30 min.
There is a lot of room for optimization between these two time intervals, 10 days and 30 min.

4. Dependence of Error of Spatially Averaged Soil Moisture for Illinois on Station Number

A map of the 17 Illinois soil moisture stations described by Hollinger and Isard [1994] is given in Figure 3. In the figure, a square box with sides of length 382 km, with an area equal to that of the State of Illinois, is also plotted. We use the optimal averaging technique to estimate the root-mean-square (RMS) errors of spatial averaging of soil moisture for the area of the box depending on the number of stations used, from 0 to 17. Volumetric soil moisture units (%) are used to compare error estimates for the upper 10 cm and upper 1 m soil layers. The results are shown in Figure 4. The first two stations in the experiment are Bondville and Brownstown. The order of the stations is not significant if the number of stations is more than 4-5. Here we use a square box for simplicity only. The optimal averaging technique may be used for any simply connected domain. The results of numerical experiments with different domains show that it should not be expected that averages for the state of Illinois and for the box shown in Figure 3 would differ significantly [Kagan, 1979].

The standard deviation of temporal variability (seasonal variations excluded) of the upper 10 cm soil layer is equal to 8.5% by volume (0.85 cm of water). This is shown in Figure 4 as the solid circle. The standard deviation of averages for the box area is equal to 7.2%. This value is the error of information in absence of observations. The data from the first seven stations decrease the RMS error of spatial averaging to 2.5%. The next 10 stations further decrease the error only a little, down to 2.2%. The same tendency may be seen in the estimates for 1 m layer soil moisture. If we are interested in monitoring averages for the entire state of Illinois, we do not need more than 10 stations. The effect of additional stations will be almost negligible.

5. Dependence of Errors of Spatial Averaging on Box Size and Network Density

Suppose that we would like to create a regular square network of soil moisture stations with observations to be used as input for weather forecast models with different spatial resolution (grid size). Using the statistical parameters estimated for the 1 m layer soil moisture in Illinois (equation 10) allows us to estimate a priori the RMS errors of optimal spatial averaging of hypothetical observations for different grid sizes. The results of such calculations are shown in Figures 5 and 6. In these calculations, the distance between stations varies from 0 to 1000 km, and model grid box size varies from about 30 km (spatial resolution of the best mesoscale models [Berbery et al., this issue]) to a few hundred kilometers (spatial resolution of a typical climate model). We assume that we use all the stations inside the grid box and those outside of a grid box that are at a distance less than the radius of autocorrelation from the box. This means that the number of stations that are used for averaging increases with increasing box size and network density.

Figure 5 shows the dependence of the RMS errors for different box sizes on the network density, the distance between stations. We have to distinguish two cases, when
one of the stations is in the center of the box, and when the four stations nearest to the center are at equal distances from the center of the box. We can see that for a regular network the RMS errors of grid-box averages are larger if there is no station in the center of the box, and the distance between stations is not very small.

In general, there will not be a station at the very center of a given domain, so we will not consider these cases. Figure 6 shows the dependence of RMS error for different distances between stations on model grid size, when there is no station at the box center. The dependence of standard deviations of soil moisture averages on box size is also given in Figure 5. This line is the limit of RMS errors of optimal averaging. We can see that the standard deviation of soil moisture spatial averages varies slowly, decreasing with increasing grid box size. In general, RMS errors of spatial averaging decrease with an increase of box size or station density. These estimates may be used as a source of quantitative and qualitative information for planning networks.

Figure 6. Dependence of RMS error of upper 1 m soil moisture optimal averaging for different observational network densities on box size, for the case with no station in the center of the box.
averaging of soil moisture. This does not mean that the Oklahoma soil moisture network is excessively dense. However, it should be recognized that after a very fast initial decrease of the errors in the soil moisture, a further increasing of network density does not provide much additional information and can be very expensive. It may be better to use part of the available resources to extend the network to cover a larger region.

7. Conclusions

We have presented an example of how information about the statistical structure of soil moisture in one region (Illinois) may be used for planning soil moisture networks in another region (Oklahoma). It may be very useful if such estimates are made before the network is actually created. This information may be also used for improving existing observational networks. What, however, may be the source of information about the statistical structure? For soil moisture we have only a few regions with well-developed observational networks. It is very significant to make information on all existing observations available to the scientific community and to use them to study the spatial and temporal statistical structure of the soil moisture fields. For regions without any available observations of soil moisture, estimates of scales of temporal autocorrelation may be based on the theory of Delworth and Manabe [1988]. This theory has been validated for Russian climatic conditions by Vinokov and Yserkerpo [1991]. In many cases, empirical estimates of scales of spatial autocorrelation of monthly precipitation may be used as the first guess for the scale of spatial autocorrelation of the soil moisture field [Meshcherskaya et al., 1982; Vinokov et al., 1996, Enuin, 1998].

When working on the problem of land surface model validation or calibration of satellite indices, we prefer to use spatially averaged observations of soil moisture. Information about the statistical structure of the soil moisture field and the optimal averaging technique allow us to estimate the RMS errors of spatially averaged observations of land-based networks [Robock et al., 1998, Enuin et al., 1999, Vinokov et al., 1999].

We should point out that taking into account this very specific type of statistical structure of soil moisture fields might produce a contradiction between meteorological and hydrological requirements for observations of soil moisture. Some hydrological models may need a very high density observational network that may be considered as excessive for meteorological applications. Meteorologists and most agrometeorologists have been satisfied with 10 day time resolution of soil moisture observations in the past. This time interval seems to be reasonable compared to the scale of temporal autocorrelation in soil moisture fields, which is a few months. Observations at the Oklahoma Mesonet have temporal resolution of 30 min to resolve the diurnal cycle of soil moisture variation at different soil depths. It is difficult to expect that the diurnal cycle is a significant component of temporal variability of upper 1 m soil layer water content, but temporal averaging of observations may be used to decrease the short-term variability and random errors in soil moisture observations.

On the basis of the above analysis we can make the following conclusions:

1. The spatial variability of a soil moisture field consists of two components, one of which is related to large-scale atmospheric forcing and the other is related to small-scale land surface variability.
2. Estimates of RMS errors of optimal averaging for model grids can be used for planning or improving observational networks.
3. The variance of soil moisture spatial averages changes slowly, decreasing with increasing grid box size. RMS errors of spatial averaging decrease with an increase of box size or station density.
4. For regular networks the RMS error of grid-box averages is largest if there is no station in the center of the box, and the distance between stations is large.
5. The Illinois and Oklahoma soil moisture networks are almost equally efficient at supplying information for mesoscale models.

Acknowledgments. Two anonymous reviewers made very useful suggestions, which significantly improved this paper. We thank Steven E. Hollinger for the Illinois soil moisture data. This work is supported by NOAA grant NAG56GP0212 and by NASA grant NAG55161. The views expressed herein are those of the authors and do not necessarily reflect the views of NOAA or NASA.

References


Drozdy, O. A., and A. A. Shepelevskiy, Theory of interpolation in stochastic field of meteorological elements and its application to weather maps and rationalization of the network (in Russian), Tr.


J. K. Entin, Hydrological Sciences Branch, NASA GSFC, Mail Code 974, Greenbelt, MD 20771. (e-mail: jentin@dao.gsfc.nasa.gov)

A. Robock, Department of Environmental Sciences, Rutgers - The State University of New Jersey, 14 College Farm Road, New Brunswick, NJ 08901-8551. (e-mail: robock@envsci.rutgers.edu)

S. Qiu, NOAA/NESDIS/ORA, 5200 Auth Road, Camp Springs, MD 20746. (e-mail: sqiu@nesdis.noaa.gov)

K. Y. Vinnikov (corresponding author), Department of Meteorology, University of Maryland, College Park, MD 20742. (e-mail: kopyta@atmos.umd.edu)

(Received September 21, 1998; revised December 18, 1998; accepted January 4, 1999.)