UPSCALING SOIL HYDRAULIC PARAMETERS IN THE PICACHO MOUNTAIN REGION USING BAYESIAN NEURAL NETWORKS

R. B. Jana, B. P. Mohanty, Z. Sheng

ABSTRACT. *A multiscale Bayesian neural network (BNN) based algorithm was applied to obtain soil hydraulic parameters at multiple scales in the Rio Grande basin (near Picacho Mountain, approximately 11 km northwest of Las Cruces, New Mexico). Point-scale measurements were upscaled to 30 m and 1 km resolutions. These scaled parameters were used in a physically based hydrologic model as inputs to obtain soil moisture states across the study area. The test sites were chosen to provide variety in terrain, land use characteristics, vegetation, soil types, and soil distribution patterns. In order to validate the effectiveness of the upscaled soil water retention parameters, and thus the soil hydraulic parameters, hydrologic simulations were conducted using the HYDRUS-3D hydrologic simulation software. Outputs from the hydrologic simulations using the scaled parameters were compared with those using data from SSURGO and STATSGO soil maps. The BNN-based upscaling algorithm for soil retention parameters from point-scale measurements to 30 m and 1 km, resolutions performed reasonably well (Pearson's R > 0.6) at both scales. High correlations (>0.6) between the simulated soil moisture values based on the upscaled and the soil map-derived soil hydraulic parameters show that the methodology is applicable to semi-arid regions to obtain effective soil hydraulic parameter values at coarse scales from fine-scale measurements of soil texture, structure, and retention data.*

Keywords. *Bayesian neural network, Multiscale pedotransfer functions, Rio Grande basin, Soil moisture.*

The Rio Grande is one of the major rivers on the North American continent. Unfortunately, in recent years, this river has been drying up, and the Rio Grande and Rio Bravo system has been classified as one of the world's top ten rivers at risk (Wong et al., 2007). There has been a steady decrease in the flow in the Rio Grande basin due to prolonged drought and surface water depletion by groundwater pumping. It has now reached a stage where serious water conservation operations have to be put in place to use the limited water resources more efficiently. In order to sustain the population dependent on the Rio Grande, a number of initiatives have been taken up by various organizations to secure future water supplies. Characterizing the river basin area is a critical effort to aid in the multiscale, multi-resolution hydroclimatic modeling of the river basin. Models to simulate and predict surface and subsurface flow components require soil hydrologic properties as input parameters. Hence, techniques to estimate these properties at various spatial resolutions must be developed. This would go a long way toward understanding the hydrologic and climatic feedback processes that occur at the fine, medium, and coarse scales and in turn help in reviving the river by minimizing losses through adoption of appropriate management practices, including capturing storm water to recharge the underlying aquifer using retention ponds.

Extensive use of pedotransfer functions (PTFs) has been made in the last two decades to derive certain complex soil hydraulic parameters, which would be difficult and expensive to measure directly, from other available or easily measurable soil properties (e.g., Cosby et al., 1984; Rawls et al., 1991; van Genuchten and Leij, 1992; Schaap and Bouten, 1996; Schaap et al., 1998; Schaap and Leij, 1998a, 1998b; Pachepsky et al., 1999; Wösten et al., 2001; Sharma et al., 2006, Jana et al., 2007, 2008; Jana and Mohanty, 2011). The traditional methods of using soil texture (sand, silt, and clay percentages) and bulk density as inputs have recently been augmented by the use of supplementary data, such as topography and vegetation parameters, which have been shown to enhance the predictive estimates of soil hydraulic parameters by PTFs to some extent (Pachepsky et al., 2001, Leij et al., 2004; Sharma et al., 2006; Jana et al., 2008). Increasing the number of model input parameters also means increasing the complexity of the model by introducing the inherent uncertainties associated with the input data and, consequently, the PTF estimates.

Artificial neural networks (ANNs) have been a mainstay for parameter estimation by PTFs in hydrology (e.g., Schaap and Bouten, 1996; Schaap et al., 1998; Schaap and Leij, 1998a; Sharma et al., 2006; Jana et al., 2007, 2008; Jana and Mohanty, 2011). However, a major drawback of conventional ANNs is the lack of uncertainty estimates. Conventionally, the weights of an ANN are obtained during

Submitted for review in September 2012 as manuscript number SW 9389; approved for publication by the Soil & Water Division of ASABE in February 2012.

The authors are **Raghavendra B. Jana,** Post-Doctoral Research Associate, and **Binayak P. Mohanty, ASABE Member,** Professor, Department of Biological and Agricultural Engineering, Texas A&M University, College Station, Texas; and **Zhuping Sheng, ASABE Member,** Associate Professor, Texas Agrilife Research Center, El Paso, Texas. **Corresponding author:** Raghavendra B. Jana, Department of Biological and Agricultural Engineering, 2117 TAMU, Texas A&M University, College Station, Texas 77843-2117; phone: 979-458-4421; e-mail: raghujana@tamu.edu

ISSN 2151-0032

training by iteratively adjusting the values until a single "optimal" set is obtained. Naturally occurring processes, and their defining parameters, are almost always stochastic (Kingston et al., 2005) and can seldom be described by a single deterministic set of parameters. However, the ANN methodology is not based on any physical processes underlying the hydrology. Rather, the training of the weights in ANNs is a statistical process that is totally dependent on the input values. Hence, the accuracy of the ANN predictions can be questioned. Schaap et al. (1998a) provided *a posteriori* estimates of the prediction uncertainties by generating multiple realizations of the ANN output. The resultant outputs were then bootstrapped and analyzed to provide confidence levels. A better approach to explicitly provide uncertainty estimates for predicted soil hydraulic properties is by using Bayesian neural networks (Jana et al., 2008; Jana and Mohanty, 2011).

Bayesian neural networks (BNNs) are designed to overcome the deficiency in conventionally trained ANNs by obtaining a range of weights. A distribution of values is predicted, explicitly accounting for the uncertainty in the predictions. Markov chain Monte Carlo (MCMC) simulation techniques, which form a part of the BNN training, also reduce the possibility of the training becoming stuck in local minima and overtraining of the network. BNNs incorporate the best features of conventional ANNs, such as their ability to form functional relationships between the inputs and the targets, while addressing some of the drawbacks, such as the ability to provide stochastic limits. Thus, BNNs may be considered the next generation of neural network models.

While the use of BNNs in the field of water resources modeling is still new, relatively little has been done toward using them for PTF development in the vadose zone. The utility of BNNs has mostly been in surface hydrology applications, where they have been used for forecasting river salinity (Kingston et al., 2005), rainfall-runoff (Khan and Coulibaly, 2006), and oxygen demand in estuaries and coastal regions (Borsuk et al., 2001). Zhang et al. (2009, 2011) studied the influence of uncertainties in the BNN model structure, inputs, and model parameters on the predictive capability for streamflow simulations. Most previous PTF studies have derived and adopted soil hydraulic parameters at the same spatial scale as the input and target data. Jana et al. (2007, 2008) demonstrated the usability of ANN- and BNN-based PTFs to estimate soil water contents at a scale different from that of the training data. The objective of this study is to develop and test a BNN-based PTF methodology to derive soil water retention values (at saturation, θ_{0bar}, and field capacity, $\theta_{0.3bar}$) at different scales using ground-based and remotely sensed data, including soil texture, bulk density, elevation, and leaf area index (LAI), in the Rio Grande basin. Remotely sensed data such as brightness temperatures have been used to derive soil state variables such as soil moisture (Chang and Islam, 2000; Das and Mohanty, 2006).

In this study, we applied the BNN methodology to obtain soil water retention parameters and saturated hydraulic conductivity at multiple scales in the Rio Grande basin. Point-scale measurements were upscaled to 30 m and 1 km resolutions. These scaled parameters were provided to a physically based hydrologic model as inputs to obtain soil moisture states across a large area.

Study Areas and Data Collection

The Bayesian training methodology was applied to data obtained from the region of Picacho Mountain near Las Cruces, New Mexico. This region is part of the Rio Grande valley in southern New Mexico. The region has a semi-arid climate, and the natural vegetation is scrub. Pecans and chilies are also grown under localized irrigation closer to the river channel. The test sites were chosen to provide variety in terrain, land use characteristics, vegetation, soil types, and soil distribution patterns. At the same time, sufficient data at the fine scale were available to validate the BNN predictions. A brief description of the test locations is given below.

Apache Canyon

Apache Canyon (fig. 1) is located northwest of Las Cruces, at the foot of Picacho Mountain, in the northeastern quadrant. The canyon is about 3 km long and 1 km wide on average. The soil is a sandy loam with a significant amount of gravel. The only vegetation in this canyon is scrub plants. Using *in situ* and laboratory methods, we developed a database of fine-scale soil properties using 14 disturbed soil samples and associated soil cores. The data set included saturated hydraulic conductivity, soil water retention function, and textural information for each sampling location. Soil cores of 8.70 cm diameter and 5.60 cm depth were collected at the surface layer. The highest sampling location was at an elevation of 1238.4 m, and the lowest was at an elevation of 1181.12 m, thus providing a topographic relief of 57.28 m across the canyon. Quarrying is carried out in certain portions of the canyon.

Box Canyon

Box Canyon is adjacent to Apache Canyon to the south (fig. 1). This is a smaller canyon of about 2 km in length and average width of about 0.6 km. The soil is finer loamy sand, as compared to Apache Canyon. Again, the sole vegetation is scrub. Soil samples and cores were collected from nine locations within this canyon. The highest sampling point in this canyon was at 1211 m, and the lowest at 1179.24 m, accounting for a relief of 31.76 m.

Lower Picacho Region

The region to the east of the Apache and Box canyons (fig. 1) up to the Rio Grande is designated the Lower Picacho region. This region has a gently sloping topography, covered with loamy sand, loam, and clay loam soil types. Farmland and homesteads cover this area, with pecan and chilies being the dominant crops. Soil cores and samples were obtained from seven locations within this region. The Lower Picacho region has the Rio Grande passing through it, as well as an arroyo and irrigation networks (canals, laterals, and drains). The highest located sampling point in this region was at an elevation of 1183.98 m, and the lowest

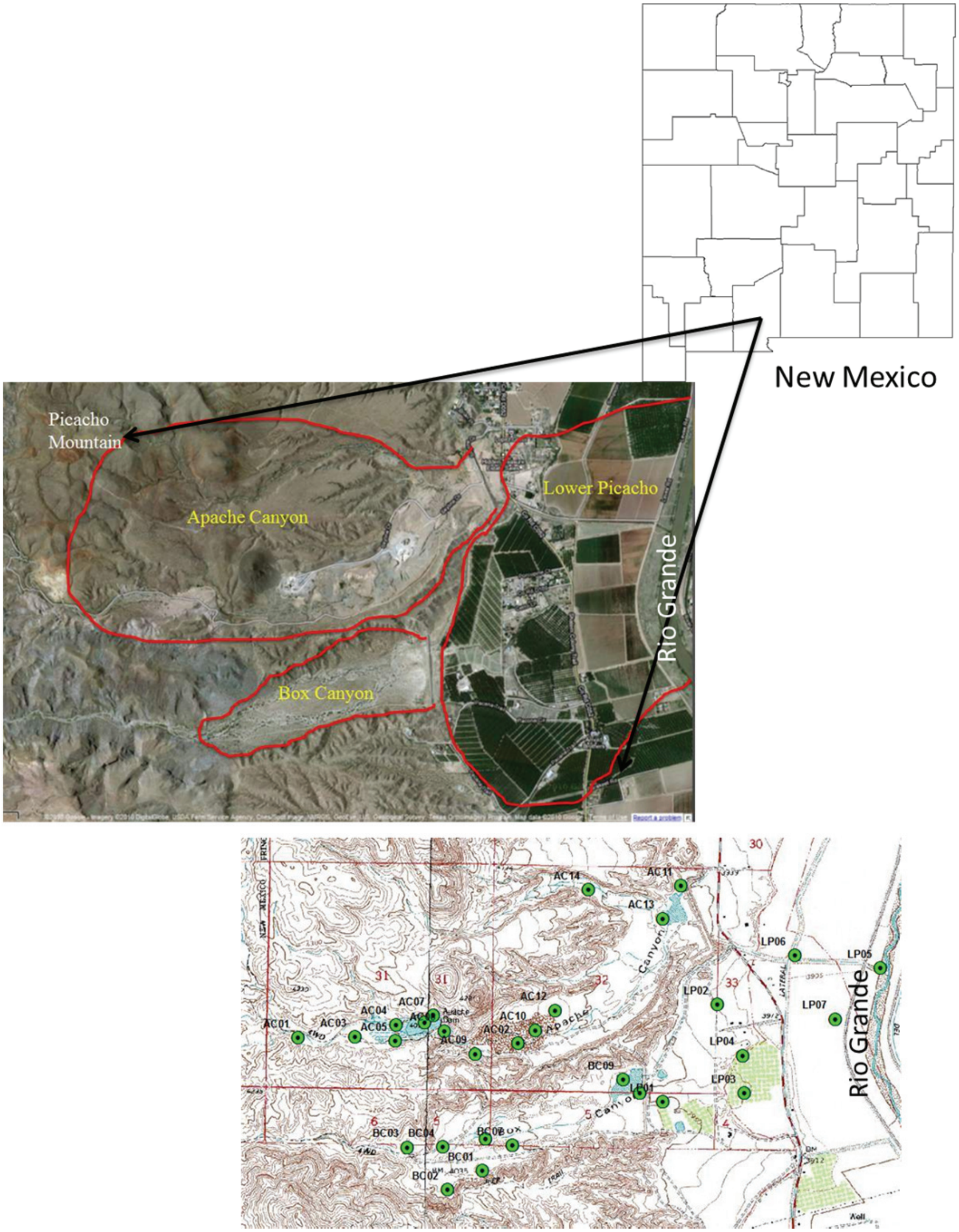

Figure 1. Rio Grande basin study area, New Mexico.

was at 1164.52 m. As can be expected from the farmed land, the topographic relief was small at only 19.46 m, generally sloping toward the Rio Grande.

Modeling data for the test locations were obtained from a variety of sources. The soil texture, bulk density, saturated hydraulic conductivity and water content details at the point scale were obtained by laboratory experiments conducted on core samples from the field. Soil property data at the 30 m resolution were obtained from the USDA-NRCS Soil Survey Geographic (SSURGO) database (http://soildatamart.nrcs.usda.gov). SSURGO is the most detailed soil map produced by the NRCS containing geo-referenced spatial and attribute data for soils. Since these surveys cover a large extent, the soil property data are based on the soil type rather than the spatial location. The SSURGO database is created by field methods, using observations along soil delineation boundaries and traverses, and determining map unit composition by field transects. Aerial photographs are interpreted and used as the field map base. Multiple readings are taken for each property within each map unit. The number of readings taken differs between map units based on factors such as the size of the soil polygon, the variation in topography, and the change in vegetation, among others. At the 1 km resolution, data were obtained from the Conterminous U.S. Multilayer Soil Characteristics Dataset for Regional Climate and Hydrology Modeling (CONUS-SOIL), a database of soil characteristics for the conterminous U.S. based on the USDA-NRCS State Soil Geographic Database (STATSGO) (Miller and White, 1998).

The SSURGO and STATSGO geo-databases are available in shapefile format. Conversion of shapefile to raster is a straightforward procedure in ArcGIS and hence is not described in detail here. Rasters were created from the SSURGO and STATSGO shapefiles for each input variable (sand, silt, and clay percentages, and bulk density) at the 30 m and 1 km cell sizes, respectively. The mean values of all included soil components within the cell were used. While the averaging of the soil components in the cells generates some uncertainty, the procedure is generally accepted as a means of aggregating the soil components to coarser resolutions. Further, the SSURGO and STATSGO documentation also refer to 30 m and 1 km, respectively, as the base resolutions for the data. Hence, it is felt that this procedure, and the resultant uncertainty, is acceptable.

Elevation data at the 30 m resolution were obtained from the National Elevation Dataset. Elevation data at the 1 km resolution were obtained from the GTOPO30 global digital elevation model provided by the U. S. Geological Survey (USGS) Earth Resources Observation and Science (EROS; http://eros.usgs.gov/products/elevation/gtopo30.html). The data are available at a resolution of 30 arcseconds, which corresponds to approximately 1 km grids. At the point scale, data from a hand-held GPS with sub-meter accuracy were used to record the coordinates and elevation information.

Multiscale BNN Analysis

Conventionally trained ANNs, as used in most previous PTF applications, form a relationship between the inputs and the targets during the training. Given *y* as the training target and *x* the input data, the relationship between *x* and *y* can be described as:

$$y = f(x \mid w) + E \tag{1}$$

where $f(x|w)$ is the functional approximation of the relationship between the input and the target as described by the ANN, *w* is the vector of weights and biases for the layers of ANN neurons, and *E* is the error term. Here, *w* is a single deterministic set of weights that provide outputs that best match the targets (i.e., least mean square error between outputs and targets). However, many such combinations of input and layer weights could exist that provide best-match outputs.

Unlike conventional ANNs, BNNs generate a probability distribution of the weights, which is dependent on the given input data. From Bayes' theorem:

$$P(w \mid y, X) = \frac{P(y \mid w, X)P(w)}{P(y \mid X)} \tag{2}$$

where *X* is the input vector (x_1, x_2,..., x_n); $P(y|X) = \int P(y \mid w, X)P(w)dw$; $P(w)$ is the prior distribution of weights; and $P(y|w, X)$ is the likelihood function (Gelman et al., 2004). As described by Kingston et al. (2005), the predictive distribution of y_{n+1} is given by:

$$\begin{aligned} &P(y_{n+1} \mid x_{n+1}, y, X) \\ &= \int P(y_{n+1} \mid x_{n+1}, w)P(w \mid y, X)dw \end{aligned} \tag{3}$$

where the subscript n+1 for *x* connotes new data that have not been used in the training of the BNN. This integral can be solved by numerical integration using Markov chain Monte Carlo (MCMC) methods (Neal, 1992).

MCMC methods are used to generate multiple samples from a continuous target density (Bates and Campbell, 2001). The posterior weight distribution is generally complex and difficult to sample. Hence, a simpler symmetrical distribution is used to generate the weight vectors. This is called the "proposal" distribution and is considered to be locally Gaussian. This proposal distribution depends only on the weights from the previous iteration in a random-walk Markov chain implementation. Arbitrary values are chosen for the weight vector *w* to start with. A series of values *w** are then proposed by the Markov chain, which are accepted with a probability given by:

$$\alpha = min\left\{\begin{matrix} 1 \\ \\ \dfrac{P(y \mid X, w^*)P(w^*)}{P(y \mid X, w_{prev})P(w_{prev})} \end{matrix}\right\} \tag{4}$$

where w_{prev} is the previous value of the weight vector. If w^* is accepted, the previous value w_{prev} is replaced by the proposed value w^* and the procedure is iterated over again. An acceptance rate between 30% and 70% is generally considered optimal (Bates and Campbell, 2001). Generating a large number of iterations ensures that the Markov chain is forced to converge to a stationary distribution. At that point, the weight vectors may be considered to have been generated from the posterior distribution itself. Detailed descriptions and discussions of the Metropolis algorithm for the MCMC method used in this study, the convergence criterion (Gelman-Rubin R-value), and the computation of the convergence efficiency of the MCMC algorithm are given by Gelman et al. (2004) and Kingston et al. (2005). We generated 50,000 Markov chain iteration samples and discarded the first 15,000 samples as burn-in. This was done to allow the network suitable time to "understand" the relationship between the inputs and the outputs and attain stability. Thus, 35,000 possible weight combinations, of which each satisfies the neural network's training requirements, were generated. Zhang et al. (2009) suggested that using multiple model structures for the BNN may improve the uncertainty estimation for the outputs. However, such a variable model structure has not been implemented in our study since the interactions between the different sources of

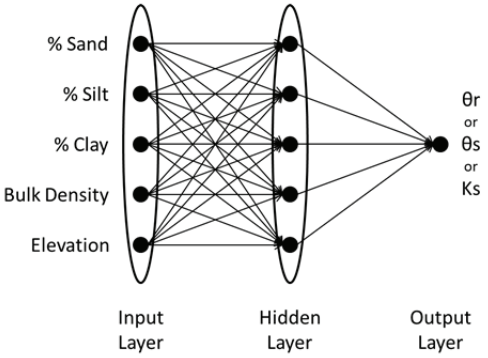

Figure 2. Neural network inputs and outputs.

uncertainty for the soil parameters and the BNN model are not really well understood at this stage (Zhang et al., 2011). A fixed-structure model (fig. 2) was implemented instead.

In this study, we adopted the BNN for upscaling the soil water content. Training data were obtained from the laboratory analyses of field samples at a point scale. The input data (soil texture and structure) for the prediction stage of the BNN analysis were obtained from the coarser resolution SSURGO (30 m) and STATSGO (1 km) databases, as mentioned earlier. Since neither the SSURGO nor the STATSGO databases report values for the van Genuchten parameters α and *n*, these were estimated from the ROSETTA database (Schaap et al., 2001) within HYDRUS-3D for all scales (point, 30 m, and 1 km). For this, the ROSETTA pedotransfer model was used given the soil texture and structure data (sand, silt, and clay percentages and bulk density) from SSURGO/STATSGO, and the BNN-predicted soil water contents at 0.33 and 15 bar as the inputs. A schematic diagram of the neural network layers, with inputs and outputs, is given in figure 2. BNNs with one input layer, one hidden layer with five neurons, and one output layer were used, with the tangent hyperbolic transfer function for all cases. The input layer had five neurons representing percent sand, percent silt, percent clay, bulk density, and elevation.

Hydrologic Simulation

In order to validate the effectiveness of the upscaled soil water retention parameters, and thus the soil hydraulic parameters, hydrologic simulations were conducted using the HYDRUS-3D hydrologic simulation software (Šimůnek et al., 2006). In this model, the Richards equation for water flow in unsaturated domains is solved numerically. HYDRUS-3D allows the user to analyze water flow through saturated or unsaturated regions with irregular boundaries and composed of non-uniform soils. HYDRUS-3D allows for three-dimensional flow representations in the unsaturated zone. The governing flow equation, a modified form of the Richards equation, is given by:

$$\frac{\partial\theta}{\partial t} = \frac{\partial}{\partial x_i}\left[K\left(K_{ij}^{A}\frac{\partial h}{\partial x_j} + K_{iz}^{A}\right)\right] - S \tag{5}$$

where θ is the volumetric water content, *h* is the pressure head, *S* is a sink term, x_i are the spatial coordinates (i = 1, 2), *t* is time, K_{ij}^{A} are components of a dimensionless anisotropy tensor K^A, and *K* is the unsaturated hydraulic conductivity, given by:

$$K(h, x, y, z) = K_s(x, y, z)K_r(h, x, y, z) \tag{6}$$

where K_r and K_s are the relative and saturated hydraulic conductivities, respectively.

The entire domain was created in HYDRUS-3D with finite element node spacing of 30 m using GIS software. A total of 19,314 pixels at 30 m resolution were obtained for the domain. Elevation data at 30 m resolution was extracted for each grid block using the GIS software. The data were then input to the HYDRUS-3D platform for creating the geometry of the domain. A minimum soil depth of 1 m was maintained across all pixels.

Four different sets of simulation were carried out. In the first set, soil water retention parameters were upscaled from the point scale to the 30 m resolution based on the SSURGO soil textural data. These were then fed to ROSETTA to obtain the van Genuchten fitting parameters. The soil hydraulic parameters thus obtained were used to simulate the water flow in the domain. In the second set, the soil water retention data from the SSURGO database were directly fed to ROSETTA and the resulting soil hydraulic parameters designated in HYDRUS-3D. This gave us a point of comparison for the hydrologic responses between the two sets of soil parameters at the 30 m resolution. Similarly, at the 1 km resolution, two sets of simulations were carried out using the STATSGO data.

As mentioned earlier, only the Lower Picacho region had any significant vegetative cover. In order to simulate this, an area equivalent to the Lower Picacho region (~2 km^2) was assigned to have a root water uptake. The Feddes model (Feddes et al., 1974) for root water uptake was applied for a deciduous trees scenario since the majority of the vegetation in the region consisted of pecan trees. In this scenario, roots extract water from the soil below a pressure head (depth) of 0.1 m, and extraction at maximum possible rate occurs below 0.25 m.

The top surface of the domain was assigned a time-dependent atmospheric boundary condition, and the sides of the domain had a seepage face. The lower boundary, at a depth of 1 m from the surface, had a free drainage condition assigned. Studies by the USGS (Nickerson, 1995, 2006) show that the Rio Grande is a losing stream in this stretch, with groundwater levels in the region being deep. The studies show that the groundwater level near the river fluctuates between about 1.5 and 3 m below the land surface, depending on the season. At the higher elevations, the groundwater levels are as deep as 10 m below land surface. In our study, we are more interested in the near-surface soil moisture; as such, the water table was considered to be beyond the depth of the model domain. Under such conditions of the water table being much below the domain boundary, the HYDRUS-3D manual (Šimůnek et al., 2006) recommends the use of the "free drainage" condition for the bottom boundary.

Due to lack of reliable rainfall data at the site, a comparable precipitation pattern was applied to the domain (fig. 3) based on rainfall data from the nearby Las Cruces city weather station. The total period of simulation was 365

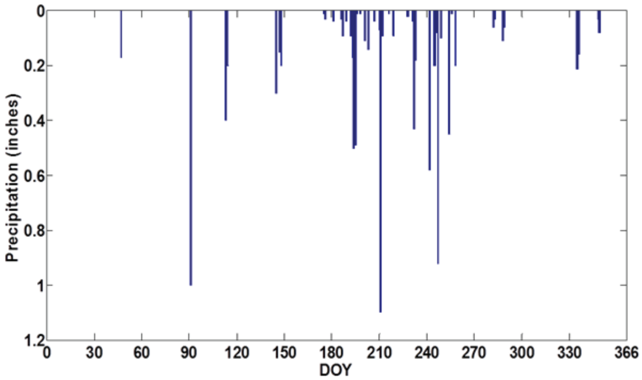

Figure 3. Precipitation pattern for the study domain (DOY = day of year).

days, and the first three months (90 days) were considered as model spin-up time to allow the domain characteristics to stabilize. The last nine months (275 days) of results were used for analysis.

Results and Discussion

Soil water retention parameters from the point scale were upscaled to 30 m and 1 km resolutions using the multiscale BNN methodology. The BNNs were trained with data from the point scale and asked to predict soil retention parameters and saturated hydraulic conductivity at the coarser resolutions. The saturation water content (θ_s), residual water content (θ_r), and saturated hydraulic conductivity (K_s) measured from the point-scale samples are plotted in figure 4. The corresponding values from the SSURGO and STATSGO databases for the soil map units surrounding the sampling location are also plotted. As can be seen, the three datasets differ quite significantly from each other. The scaling outputs from the BNN at the 30 m resolution are plotted along with their corresponding SSURGO values (fig. 5). The error bars, obtained from the MCMC simulations, represent the uncertainty in the neural network predictions. BNNs, as mentioned earlier, generate a distribution of training weights instead of a single set. The final predicted soil water content value is an average of all such possible values from the 35,000 Monte Carlo simulations. The uncertainty band (error bars) shows the limits to which the predictions could have varied based on the combination of weights used. Thirty sets of point-scale inputs and outputs are available for training the BNN on each parameter. The SSURGO values are taken from the 30 m pixel surrounding the point-scale sampling location. As can be seen from the comparative statistics, the upscaled retention parameters match the SSURGO data very well.

For the saturation water content, the mean prediction from the BNN ensemble matched the SSURGO-derived values very well for the Lower Picacho region (fig. 5, last seven data points) and Apache Canyon (fig. 4, data points 10 to 23) regions. The data points corresponding to Box Canyon (fig. 5, first nine data points) showed the most deviation between the BNN-predicted and SSURGO values. This discrepancy may be explained by the difference in the soil textural information. The USDA textural triangles for

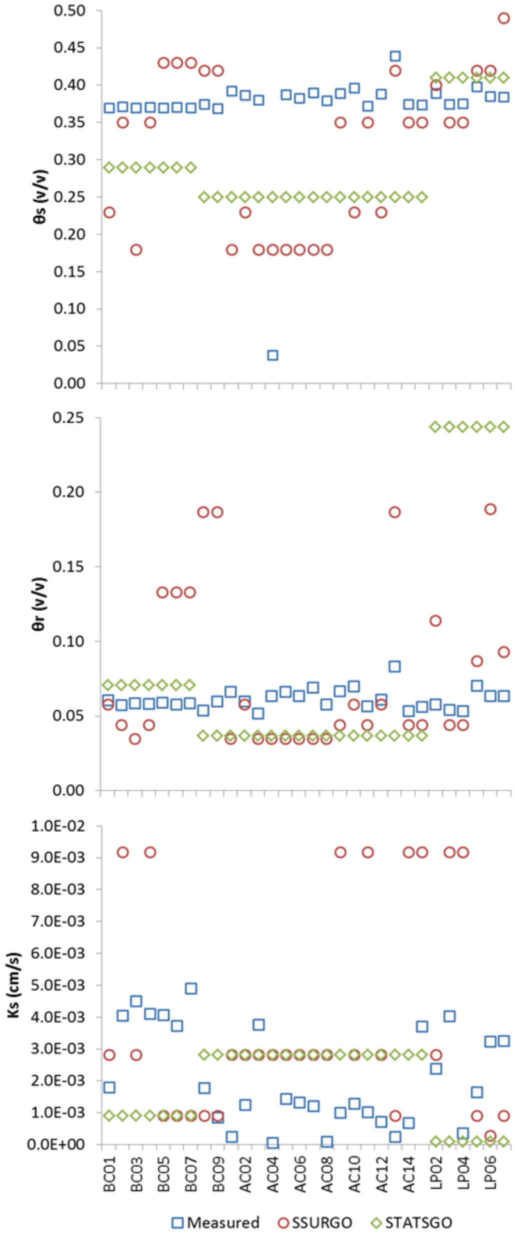

Figure 4. Measured, SSURGO, and STATSGO values for soil water retention and saturated hydraulic conductivity (θ_s = saturation water content, θ_r = residual water content, K_s = saturated hydraulic conductivity, SSURGO = 30 m resolution soil map, and STATSGO = 1 km resolution soil map).

the point-scale laboratory-measured data and the corresponding SSURGO data are shown in figure 6. The texture analysis of the field samples showed that the samples from Box Canyon were mostly clustered in the loamy sand classification, with a solitary point in the sandy loam. However, the SSURGO data have the corresponding soils evenly distributed among sandy loam, loam, and clay loam. Although the entire range of data from the point scale is used to train the BNN, there are hardly any training points in the clay loam and loam textures. Twarakavi et al. (2010) demonstrated the similarities between a textural classification based on the percentages of sand, silt, and clay and a classification based on the hydraulic properties of the soil. They showed that for sand- and loam-dominant soils, the textural

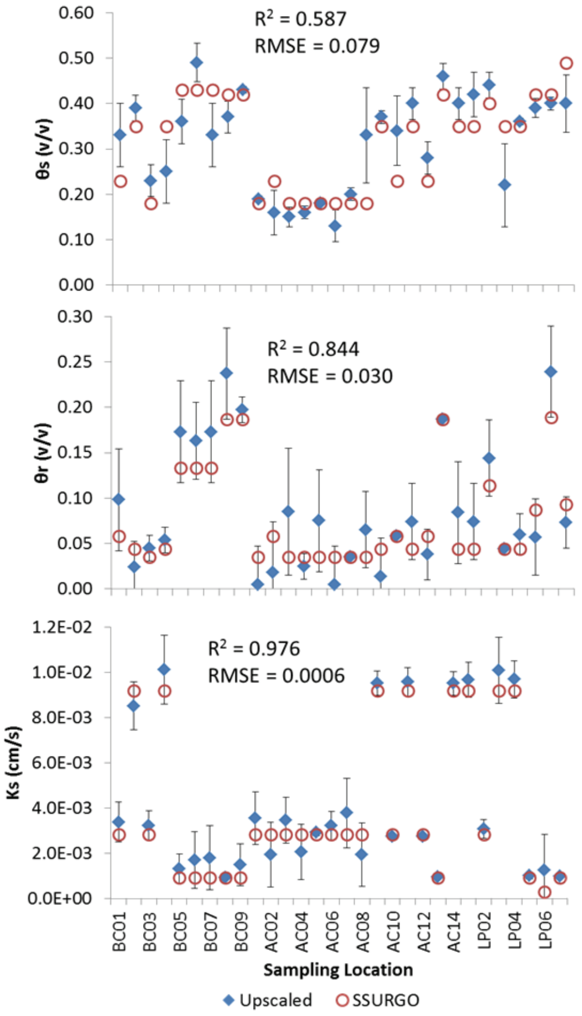

Figure 5. Upscaled and SSURGO values for soil water retention and saturated hydraulic conductivity (θ_s = saturation water content, θ_r = residual water content, K_s = saturated hydraulic conductivity, and SSURGO = 30 m resolution soil map).

and hydraulic classifications are similar. This means that the hydraulic properties of the soil are uniquely described by the texture. Hence, having only a few training points in the clay loam and loam textures makes it difficult for the BNN to reach a closer prediction.

For the residual water content (fig. 5), uncertainty in the predictions is larger, as can be expected from the greater variability of the soil water content at the drier end of the retention curve. However, this means that most SSURGO values are captured within the uncertainty zone of the predictions. The predicted values also show greater variability as compared to the SSURGO data. This may again be explained by the greater spread of the textural values at the point scale. A similar trend of better predictions for the drier water content was also observed by Jana and Mohanty (2011).

The K_s predictions from the multiscale BNN have a close match ($R^2 > 0.95$) with the SSURGO values. However, it must also be noted that while the error bars appear small, the percentage of uncertainty is greater for K_s as compared to the other two parameters. The hydraulic conductivity also depends on factors such as the pore connectivity and presence of macropores in the domain. This makes the estimation of K_s from only the texture, bulk density, and elevation data inaccurate. However, obtaining the pore connectivity and macropore densities of the field soils is not a trivial task, and these properties are generally not measured.

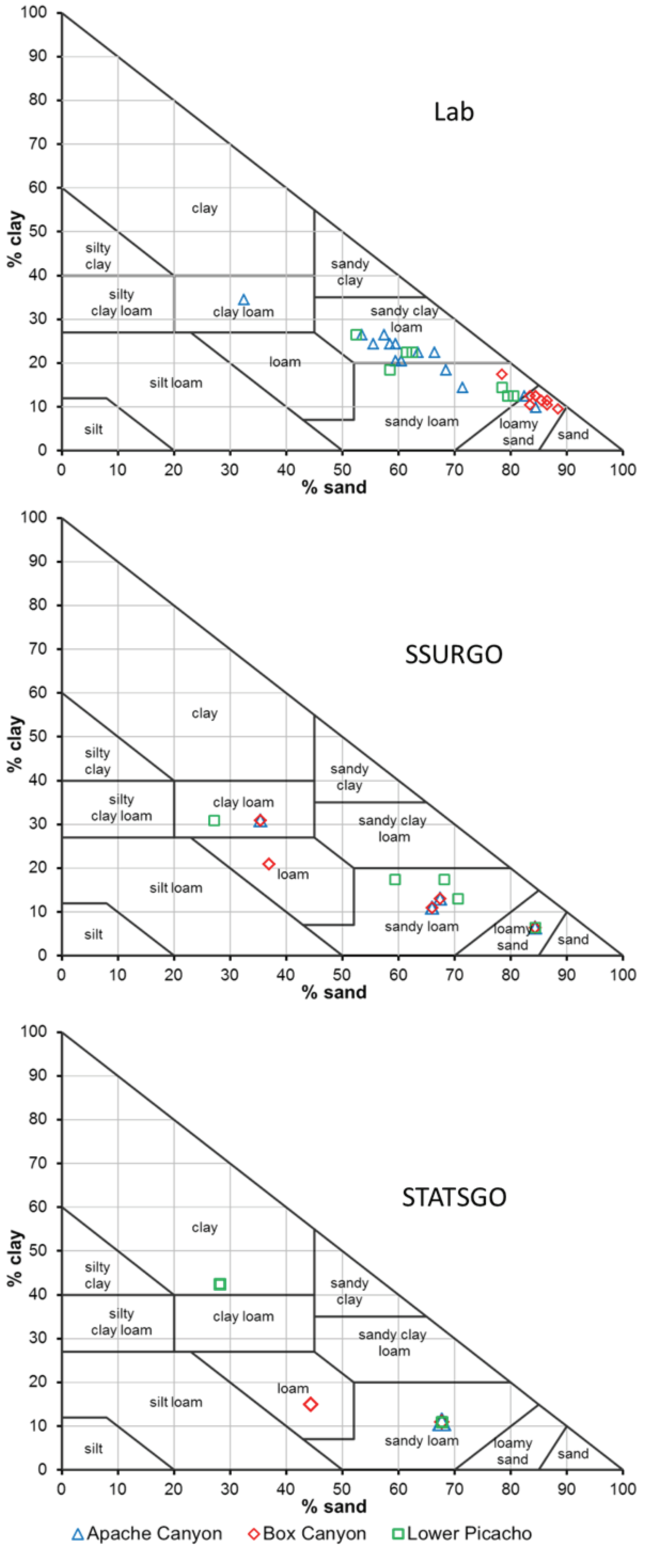

Figure 6. Soil textural triangles for the laboratory-measured samples and corresponding SSURGO and STATSGO soil map units.

Figure 7 shows the corresponding predicted and

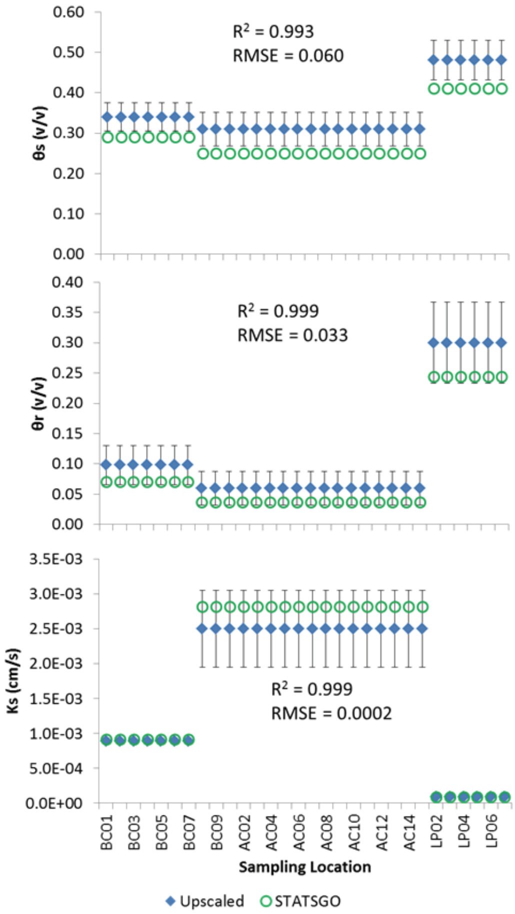

Figure 7. Upscaled and STATSGO values for soil water retention and saturated hydraulic conductivity (θ_s = saturation water content, θ_r = residual water content, K_s = saturated hydraulic conductivity, and STATSGO = 1 km resolution soil map).

STATSGO data for the soil parameters at 1 km resolution. The STATSGO data show three very distinct clusters for the retention and conductivity, and this is borne out by the textural classification. The 1 km resolution STATSGO pixels depict just three well defined clusters in the clay, loam, and sandy loam classifications (fig. 6). As can be expected with such few textures, the comparative statistics at the 1 km resolution are very good ($R^2 > 0.99$). However, more than the correlation, the RMSE must be observed in this case. It is seen that the error magnitude is comparable to that at the 30 m resolution.

The entire study domain (Apache Canyon, Box Canyon, and Lower Picacho region) was simulated in HYDRUS-3D to obtain the surface soil moisture states under different inputs of the soil hydraulic parameters, as explained earlier. Of the 275 days used for the analysis, days on which the average soil moisture over all 19,314 pixels when simulated with the SSURGO parameters was above 0.3 v/v were designated wet days. Similarly, days with an average soil moisture value below 0.2 v/v were designated dry days, and days having values in between were termed intermediate days. It was found that out of the 276 analysis days, 41 were wet days, 63 were intermediate days, and 172 were dry days. Histograms depicting the distribution of surface soil moisture across the domain for the wet, intermediate, and dry days are shown in figure 8. The comparative statistics for the soil moisture distributions are given in table 1. As can be seen, the soil moisture values across the domain for the three phases have high (>0.6) values of Pearson's correlation. In addition, as expected, the correlation for the wet phase was higher in comparison with the intermediate and dry phases, while the dry phase had the lowest correlation values. This is expected because the variability of soil moisture at the dryer end of the characteristic curve is higher as compared to the wetter end. Hence, as the domain dries out, more variability is observed, and thus a lower correlation exists between the upscaled and SSURGO parameter dependent soil moistures. It was seen that, on average, the upscaled soil hydraulic parameters resulted in underprediction of the values in the wet and dry phases but oveprediction in the intermediate phase.

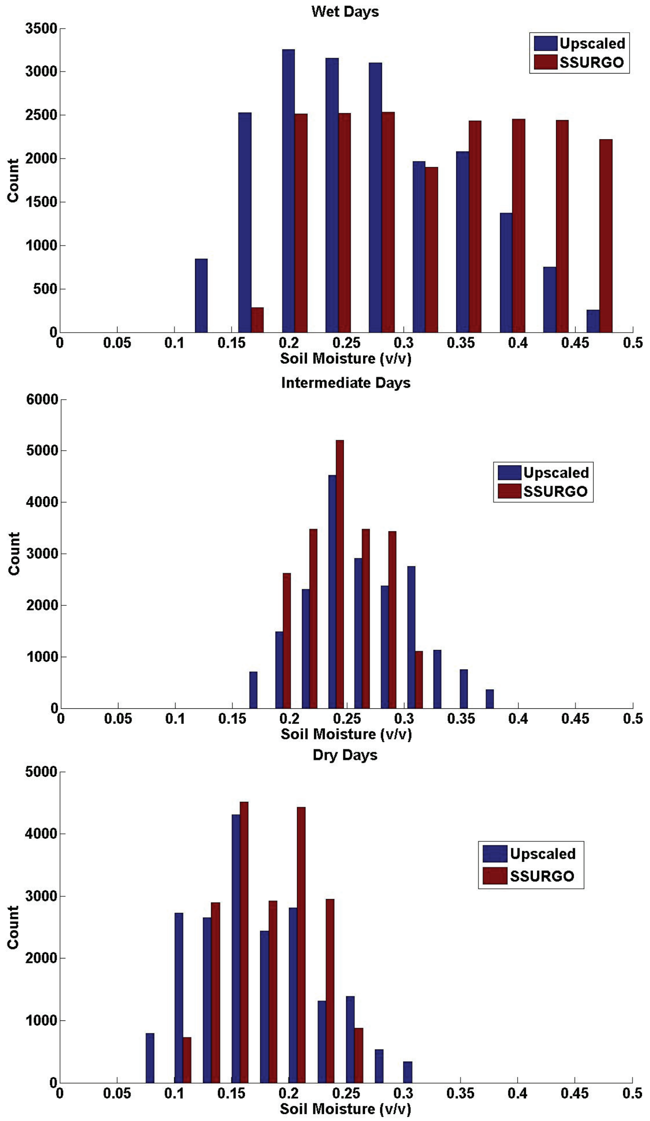

Figure 8. Soil moisture histograms for wet, intermediate, and dry days at 30 m resolution from upscaled and SSURGO soil hydraulic parameters.

Figure 9 shows histograms for the soil moisture distributions at the 1 km resolution. It is immediately apparent that the range of variability in the soil moisture values is much lower as compared to the 30 m resolution. This is because there are very few different soils at the 1 km resolution. What variability is seen in the soil moisture is due more to the topography, vegetation, and irrigation differences than to the soil. Table 2 contains the comparative statistics for the wet, intermediate, and dry phase soil moistures at the 1 km scale. Due to the reduced variability in soil moisture, the Pearson's correlation values between the upscaled and STATSGO parameter dependent soil moistures are significantly higher. As also deduced from figure 7, as scale increases, variability decreases, and correlation increases. The RMSE values at the 1 km resolution are also lower than those at the 30 m resolution.

It may appear counter-intuitive that the BNN-based upscaling algorithm performs better at the 1 km resolution than at the 30 m resolution. To better understand this result, we need to consider the inputs to the BNN at the simulation stage. At the 30 m resolution, each of the 19,314 domain pixels has a different value for the percent sand, percent silt, percent clay, bulk density, and elevation inputs. However, at the 1 km resolution, the number is brought down to just six pixels, with some soils being similar. This means all of the 30 m spaced nodes of the finite element domain mesh in the 1 km area have the same inputs. Hence, the apparently better upscaling performance at the 1 km resolution is observed.

While earlier implementations of the multiscale BNN methodology (Jana et al., 2008; Jana and Mohanty, 2011) concentrated on the retention parameters, in this study, we used the algorithm to estimate the effective saturated hydraulic conductivity at multiple scales. Earlier studies (Mohanty et al., 1994; Das et al., 2008) showed that it is difficult to scale the saturated hydraulic conductivity due to its high variability, which is caused by a number of factors in the field. However, our multiscale BNN approach has proven to be robust enough to provide very good estimates (Pearson's R > 0.97; RMSE < 0.001) at both coarse scales. This suggests that the BNN structure has been able to "understand" the non-linear relationship between the inputs and the saturated hydraulic conductivity.

The estimates of the soil hydraulic parameters were fed to a hydrologic model (HYDRUS-3D) to assess what im-

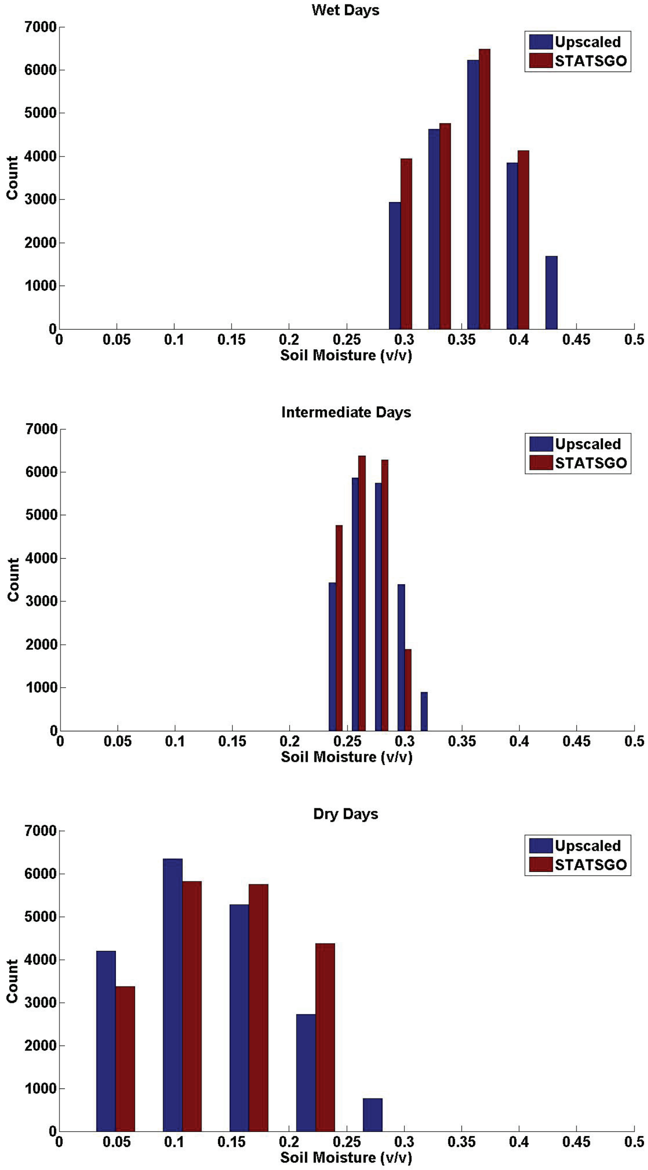

Figure 9. Soil moisture histograms for wet, intermediate, and dry days at 1 km resolution from upscaled and STATSGO soil hydraulic parameters.

Table 1. Comparative statistics of wet, intermediate, and dry phase soil moistures from upscaled and SSURGO soil parameters at 30 m resolution. All correlations are significant at the 0.01 level.

	Wet Phase (41 days)		Intermediate Phase (63 days)		Dry Phase (172 days)		Total (276 days)	
Statistic	Upscaled	STATSGO	Upscaled	STATSGO	Upscaled	STATSGO	Upscaled	STATSGO
Average soil moisture	0.27	0.33	0.26	0.24	0.18	0.19	0.21	0.22
Standard deviation	0.08	0.09	0.05	0.03	0.05	0.04	0.06	0.05
Pearson's correlation	0.874		0.723		0.675		0.715	
RMSE	0.078		0.038		0.040		0.045	

Table 2. Comparative statistics of wet, intermediate, and dry phase soil moistures from upscaled and STATSGO soil parameters at 1 km resolution. All correlations are significant at the 0.01 level.

	Wet Phase (41 days)		Intermediate Phase (63 days)		Dry Phase (172 days)		Total (276 days)	
Statistic	Upscaled	STATSGO	Upscaled	STATSGO	Upscaled	STATSGO	Upscaled	STATSGO
Average soil moisture	0.36	0.35	0.28	0.27	0.13	0.14	0.20	0.20
Standard deviation	0.04	0.03	0.02	0.02	0.06	0.06	0.04	0.04
Pearson's correlation	0.911		0.821		0.871		0.866	
RMSE	0.018		0.014		0.032		0.026	

pacts their variability has on the prediction of the soil moisture states of the test domain. The absence of real precipitation, soil moisture, and streamflow data make the current exercise open to questions of validation. However, we believe that the question of precision, at least with respect to soil moisture, has been answered. Data such as the discharge from the arroyo or the streamflow data at the Rio Grande, which runs on the eastern edge of the domain, could strengthen claims to the suitability of the BNN-based scaling algorithm in this region. While the local irrigation district is currently installing gauging stations to monitor such information, no reliable data is currently available.

From the above results, it may be inferred that the BNN-based upscaling algorithm performs reasonably well for soil retention parameters at both scales from point-scale measurements to 30 m and 1 km. High correlations (>0.6) between the simulated soil moisture values based on the upscaled and the soil map-derived soil hydraulic parameters show that the methodology is applicable to this region. The current study region in the Rio Grande basin of New Mexico is in contrast to the two study sites used by Jana and Mohanty (2011) in that it is a mostly arid region with a patch of irrigated land. This suggests that, while the BNN methodology is inherently site-specific, it may be applied in different hydro-climatic regions with comparable efficiency. This methodology for upscaling soil retention parameters, due to its demonstrated performance in different test sites, shows promise to be a generic scaling algorithm.

While the SSURGO and STATSGO databases have been used in this study to validate the upscaling algorithm, the multiscale BNN methodology can be applied with any dataset for which the input variables are available to generate the water retention parameters and the saturated hydraulic conductivity at the target scale. SSURGO and STATSGO data are available for much of the continental U.S., but such comprehensive data may not be available for other parts of the world. In such cases, any available aggregate soil dataset, such as the UNSODA database (Nemes et al., 2001), may also be used to obtain satisfactory outputs.

It may be observed that the bias correction was not applied in this upscaling study as in Jana and Mohanty (2011). It was not considered necessary, since upscaling is an interpolative exercise for the BNN. Neural networks perform better at interpolation than at extrapolation. This inherent property of BNNs means that they are naturally better at upscaling exercises than downscaling, where an additional bias correction step would be necessary to account for the scale disjoint.

Conclusions

Using point-scale soil property data from ground-based measurements, we have shown that a Bayesian neural network (BNN) can be applied across spatial scales to approximate coarse-scale soil hydraulic properties. The study was conducted for a semi-arid region of New Mexico from which the point data was collected. The upscaled parameters were fed to a physically based hydrologic model of the region to simulate surface soil moisture states. The results were compared with those obtained using parameters from published soil maps. The results show good match (Pearson's R > 0.6; RMSE < 0.001) between the soil moistures for the domain obtained using two different parameter sets at 30 m and 1 km resolutions. Previous implementations of the BNN methodology were mainly as a tool for downscaling of the soil water retention parameters. This study establishes the utility of the BNN method as a multiscale algorithm suitable for upscaling soil water retention parameters, as well as the highly uncertain saturated hydraulic conductivity. This study also highlights the applicability of the multiscale BNN algorithm in regions where there are drastic changes in vegetation, topography, and soil properties within the domain. As such, the BNN-based upscaling methodology may be applied at different regions to obtain effective soil hydraulic parameter values at coarse scales from fine-scale measurements of soil texture, structure, and retention data.

Acknowledgements

We acknowledge the partial support from the U.S. Army Corps of Engineers (Grant No. W912HZ-10-2-0038), the National Science Foundation (CMG/DMS Grants No. 062113 and C10-00021) and NASA (THP Grants No. NNX08AF55G and NNX09AK73G). We thank Dr. Phillip King for his assistance in arranging the site investigation and sample collection.

References

Bates, B. C., and E. P. Campbell. 2001. A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling. *Water Resour. Res.* 37(4): 937-947.

Borsuk, M. E., D. Higdon, C. A. Stow, and K. H. Reckhow. 2001. A Bayesian hierarchical model to predict benthic oxygen demand from organic matter loading in estuaries and coastal areas. *Ecol. Modelling* 143(3): 165-181.

Chang, D.-H., and S. Islam. 2000. Estimation of soil physical properties using remote sensing and artificial neural network. *Remote Sensing of Environ.* 74(3): 534-544.

Cosby, B. J., G. M. Hornberger, R. B. Clapp, and T. R. Ginn. 1984. A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils. *Water Resour. Res.* 20(6): 682-690.

Das, N. N., and B. P. Mohanty. 2006. Root zone soil moisture assessment using remote sensing and vadose zone modeling. *Vadose Zone J.* 5(1): 296-307.

Das, N. N., B. P. Mohanty, and E. G. Njoku. 2008. A Markov chain Monte Carlo algorithm for upscaled soil-vegetation-atmosphere-transfer modeling to evaluate satellite-based moisture measurements. *Water Resour. Res.* 44: W05416, doi: 10.1029/2007WR006472.

Feddes, R. A., E. Bresler, and S. P. Neuman. 1974. Field test of a modified numerical model for water uptake by root systems. *Water Resour. Res.* 10(6): 1199-1206.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 2004. *Bayesian Data Analysis*. Boca Raton, Fla.: CRC Press.

Jana, R. B., and B. P. Mohanty 2011. Enhancing PTFs with remotely sensed data for multi-scale soil water retention estimation. *J. Hydrol.* 399(3-4): 201-211.

Jana, R. B., B. P. Mohanty, and E. P. Springer. 2007. Multiscale pedotransfer functions for soil water retention. *Vadose Zone J.* 6(4): 868-878.

Jana, R. B., B. P. Mohanty, and E. P. Springer. 2008. Multiscale Bayesian neural networks for soil water content estimation. *Water Resour. Res.* 44(8): W08408, doi: 10.1029/ 2008wr006879.

Khan, M. S., and P. Coulibaly. 2006. Bayesian neural network for rainfall-runoff modeling. *Water Resour. Res.* 42(7): W07409, doi: 10.1029/2005wr003971.

Kingston, G. B., M. F. Lambert, and H. R. Maier. 2005. Bayesian training of artificial neural networks used for water resources modeling. *Water Resour. Res.* 41: W12409, doi:12410.11029/ 12005WR004152.

Leij, F. J., N. Romano, M. Palladino, M. G. Schaap, and A. Coppola. 2004. Topographical attributes to predict soil hydraulic properties along a hillslope transect. *Water Resour. Res.* 40(2): W02407, doi: 10.1029/2002wr001641.

Miller, D. A., and R. A. White. 1998. A conterminous United States multi-layer soil characteristics data set for regional climate and hydrology modeling. *Earth Interactions* 2(2): 1-16. Available at: http://EarthInteractions.org.

Mohanty, B. P., M. D. Ankeny, R. Horton, and R. S. Kanwar. 1994. Spatial analysis of hydraulic conductivity measured using disc infiltrometers. *Water Resour. Res.* 30(9): 2489-2498.

Neal, R. M. 1992. Bayesian training of back-propagation networks by the hybrid Monte Carlo method. Tech. Report CRG-TR-92-1. Toronto, Ontario, Canada: University of Toronto, Department of Computer Science.

Nemes, A., M. G. Schaap, F. J. Leij, and J. H. M. Wösten. 2001. Description of the unsaturated hydraulic database UNSODA version 2.0. *J. Hydrol* 251(3-4): 151-162.

Nickerson, E. L. 1995. Selected hydrologic data for the Mesilla groundwater basin, 1987 through 1992 water years, Dona Ana County, New Mexico, and El Paso County, Texas. USGS Open-File Report 95-111. Reston, Va.: U.S. Geological Survey.

Nickerson, E. L. 2006. Description of piezometers and groundwater quality characteristics at three new sites in the lower Mesilla Valley, Texas, and New Mexico, 2003. USGS Scientific Investigations Report 2005-5248. Reston, Va.: U.S. Geological Survey.

Pachepsky, Y. A., W. J. Rawls, and D. J. Timlin. 1999. The current status of pedotransfer functions, their accuracy, reliability, and utility in field- and regional-scale modeling. In *Assessment of Nonpoint-Source Pollution in the Vadose Zone*, 223-234. D. L. Corwin, K. Loague, and T. R. Ellsworth, eds. Washington, D.C.: American Geophysical Union.

Pachepsky, Y. A., D. J. Timlin, and W. J. Rawls. 2001. Soil water retention as related to topographic variables. *SSSA J.* 65(6): 1787-1795.

Rawls, W. J., T. J. Gish, and D. L. Brakensiek. 1991. Estimating soil water retention from soil physical properties and characteristics. In *Advances in Soil Science*, Vol. 16: 213-234. B. A. Stewart, ed. New York, N.Y.: Springer.

Schaap, M. G., and W. Bouten. 1996. Modeling water retention curves of sandy soils using neural networks. *Water Resour. Res.* 32(10): 3033-3040.

Schaap, M. G., and F. J. Leij. 1998a. Using neural networks to predict soil water retention and soil hydraulic conductivity. *Soil and Tillage Res.* 47(1-2): 37-42.

Schaap, M. G., and F. J. Leij. 1998b. Database-related accuracy and uncertainty of pedotransfer functions. *Soil Sci.* 163(10): 765-779.

Schaap, M. G., F. J. Leij, and M. Th. van Genuchten. 1998. Neural network analysis for hierarchical prediction of soil hydraulic properties. *SSSA J.* 62(4): 847-855.

Schaap, M. G., F. J. Leij, and M. Th. van Genuchten. 2001. Rosetta: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *J. Hydrol.* 251(3-4): 163-176.

Sharma, S. K., B. P. Mohanty, and J. Zhu. 2006. Including topography and vegetation attributes for developing pedotransfer functions. *SSSA J.* 70(5): 1430-1440.

Šimůnek, J., M. T. van Genuchten, and M. Šejna. 2006. The HYDRUS software package for simulating two- and three-dimensional movement of water, heat, and multiple solutes in variably saturated media: Technical manual. Version 1.0, edited. Prague, Czech Republic: PC-Progress.

Twarakavi, N. K. C., J. Šimůnek, and M. G. Schaap. 2010. Can texture-based classification optimally classify soils with respect to soil hydraulics? *Water Resour. Res.* 46: W01501, doi:10.1029/2009WR007939.

van Genuchten, M. T., and F. J. Leij. 1992. On estimating the hydraulic properties of unsaturated soils. In *Indirect Methods for Estimating the Hydraulic Properties of Unsaturated Soils, Proc. Intl. Workshop on Indirect Methods for Estimating the Hydraulic Properties of Unsaturated Soils*, 1-14. M. T. van Genuchten, F. J. Leij, and L. J. Lund, eds. Riverside, Cal.: University of California, Department of Soil and Environmental Sciences.

Wong, C. M., C. E. Williams, J. Pittock, U. Collier, and P. Schelle. 2007. World's top ten rivers at risk. Gland, Switzerland: WWF International.

Wösten, J. H. M., Y. A. Pachepsky, and W. J. Rawls. 2001. Pedotransfer functions: Bridging the gap between available basic soil data and missing soil hydraulic characteristics. *J. Hydrol.* 251(3-4): 123-150.

Zhang, X., F. Liang, R. Srinivasan, M. Van Liew. 2009. Estimating uncertainty of streamflow simulation using Bayesian neural networks. *Water Resour. Res.* 45: W02403, doi: 10.1029/2008WR007030.

Zhang, X., F. Liang, B. Yu, Z. Zong. 2011. Explicitly integrating parameter, input, and structure uncertainties into Bayesian neural networks for probabilistic hydrologic forecasting. *J. Hydrol.* 409(3): 696-709.