Research Article

Spatial Downscaling of TRMM Precipitation Using Geostatistics and Fine Scale Environmental Variables

No-Wook Park

Department of Geoinformatic Engineering, Inha University, Incheon 402-751, Republic of Korea

Correspondence should be addressed to No-Wook Park; nwpark@inha.ac.kr

Received 6 September 2013; Revised 19 November 2013; Accepted 27 November 2013

A geostatistical downscaling scheme is presented and can generate fine scale precipitation information from coarse scale Tropical Rainfall Measuring Mission (TRMM) data by incorporating auxiliary fine scale environmental variables. Within the geostatistical framework, the TRMM precipitation data are first decomposed into trend and residual components. Quantitative relationships between coarse scale TRMM data and environmental variables are then estimated via regression analysis and used to derive trend components at a fine scale. Next, the residual components, which are the differences between the trend components and the original TRMM data, are then downscaled at a target fine scale via area-to-point kriging. The trend and residual components are finally added to generate fine scale precipitation estimates. Stochastic simulation is also applied to the residual components in order to generate multiple alternative realizations and to compute uncertainty measures. From an experiment using a digital elevation model (DEM) and normalized difference vegetation index (NDVI), the geostatistical downscaling scheme generated the downscaling results that reflected detailed characteristics with better predictive performance, when compared with downscaling without the environmental variables. Multiple realizations and uncertainty measures from simulation also provided useful information for interpretations and further environmental modeling.

1. Introduction

Precipitation information has been regarded as one of the important information sources for understanding hydrological, ecological, and environmental systems [1–3]. This information can be obtained from either rain gauge station data or remote sensing data. Although precise precipitation information can be obtained from rain gauge data, few rain gauge data are usually available, which may hinder the generation of reliable maps. Remote sensing data that can provide periodic and exhaustive information can be effectively used to map precipitation information. For example, the Global Precipitation Climatology Project (GPCP) [4] and the Tropical Rainfall Measuring Mission (TRMM) [5] provide precipitation data at regional and global scales [3]. As an international satellite mission, the Global Precipitation Measurement (GPM) mission, scheduled to launch in 2014, will also provide next-generation observations of rain and snow worldwide [6]. Although these missions or projects can provide time-series precipitation information, their spatial resolutions are too coarse to be applied to local analysis (e.g., the finest resolution from TRMM data is 0.25 degree).

In hydrological, ecological, and environmental modeling, various data sets acquired at the different scales are generally used as inputs in addition to precipitation data. Thus, scale conversion or change of support is usually required for the consistent analysis. When coarse scale precipitation data and other fine scale data sets are available, for example, downscaling of the coarse scale precipitation data is required before subsequent modeling is conducted. Downscaling can be regarded as the spatial prediction of unknown values at a finer scale from coarse scale data. In relation to this kind of change of support, several statistical downscaling schemes have been proposed and applied to various research fields [7–11]. The spatial downscaling of coarse scale data can be categorized into two cases according to data availability. In the first case, in which ground measurements are simultaneously available, calibration of the coarse scale data and integration for downscaling are enabled by using the ground measurement data. In the second case, in which no ground measurements are available, direct downscaling of the coarse scale data is only feasible by using spatial correlation information from the available data. If auxiliary environment
variables that are related to the primary attribute of interest are acquired at a relatively finer scale, these data can improve the quality of the downscaling result.

Regarding downscaling of coarse scale precipitation data, auxiliary environmental variables such as topography and vegetation have been used for downscaling with regression analysis [12–17]. Regression analysis can account for the statistical relationships between precipitation and environmental variables, but residuals that cannot be explained by the auxiliary variables usually remain after regression analysis. If the auxiliary variables fail to convey sufficient information on precipitation patterns, the residuals cannot be discarded, resulting in unreliability in the downscaling results. Conventional regression analysis follows the assumption that residuals are spatially independent or uncorrelated. If reasonable spatial correlation structures are observed in the residuals, this correlation information can improve the quality of the downscaling. Recently, Jia et al. [3] proposed a statistical scheme for the downscaling of TRMM data with fine scale elevation and normalized difference vegetation index (NDVI). They first conducted multiple regression analysis at various spatial scales then interpolated the residuals to the fine scale grid, and finally generated downscaling results by adding the trend component by regression analysis at an optimal spatial scale to the residuals at a fine scale. During the interpolation of the residuals, however, they neither considered the spatial patterns nor examined the reproduction of original TRMM values.

Geostatistics has been known as an effective tool for scale conversion and data integration [18, 19]. Spatial downscaling or disaggregation based on geostatistics has been applied to various fields such as land-cover mapping [20], population density estimation [21], and image sharpening [22]. Despite its great potential for spatial downscaling, geostatistics has not, to the author’s knowledge, been applied to the downscaling of coarse scale TRMM data with fine scale environmental variables.

In this paper, a geostatistical downscaling scheme is presented that incorporates multiple fine scale environmental variables into downscaling of coarse scale TRMM data in the absence of any ground measurement data. Geostatistical downscaling and data integration are implemented by using area-to-point residual kriging with regression. Elevation and NDVI at a fine scale, which are related to precipitation patterns, are used as auxiliary environmental variables for downscaling of TRMM data. Regression analysis is first conducted to derive statistical relationships between TRMM data and auxiliary environmental variables at an original coarse scale. Then residuals at the coarse scale are downscaled to a finer scale via area-to-point residual kriging by accounting for the spatial correlation information of the input residuals [23]. The final downscaling results are obtained by adding the downscaled residuals to the trend components estimated by regression analysis. This approach can reproduce the original TRMM precipitation values when the downscaling results at a fine scale are upscaled or aggregated to the coarse scale. To quantify the uncertainty of downscaling, stochastic simulation based on block sequential simulation is also implemented. A downscaling experiment using the data sets from South Korea is carried out to examine the potential and demonstrate the applicability of the presented geostatistical scheme for downscaling.

2. Study Area and Data Sets

For the downscaling experiment, the TRMM 3B43 data acquired in October 2005 over South Korea were experimentally used. The original TRMM data that provide precipitation rates (mm/h) at a 0.25° scale were geocoded to Transverse Mercator (TM) coordinates with a spatial resolution of 25 km, and converted to monthly cumulative precipitation at a cm scale (Figure 1(a)). Both digital elevation model (DEM) and NDVI data at 1 km resolution were used as fine scale auxiliary variables, for topography and a proxy for vegetation, respectively. The DEM was generated from digital topographic maps and MODIS monthly NDVI data were used (Figures 1(b) and 1(c)). DEM and NDVI values at land areas were only considered and sea areas were masked out. Independent precipitation data observed at 69 meteorological observation stations in October 2005 were used to quantitatively evaluate the predictive performance of the presented downscaling scheme. In this experiment, the target scale was set to 1 km, which is the same as that of DEM and NDVI.

3. Geostatistical Downscaling

The geostatistical downscaling scheme presented in this paper is summarized in Figure 2. The main key idea of this approach is that TRMM precipitation values are decomposed into deterministic trends and stochastic residuals. The trend components are estimated using auxiliary environmental variables such as DEM and NDVI. More specifically, statistical relationships between original precipitation values and multiple environmental variables at a coarse scale are first derived via multiple regression analysis. Under the assumption that attribute values at a coarse scale are linear averages of their constituent fine scale point values, these relationships are applied to environmental variables at a fine scale and finally trend components are estimated at a fine scale.

Suppose that a study area of interest consists of \( K \) TRMM precipitation blocks acquired at a coarse scale \( \{ z(v_k), k = 1, \ldots, K \} \), where \( v_k = (u_k) \) is the \( k \)th data with its centroid \( u_k \). In this study, the TRMM data at a coarse scale are regarded as block data. Other data sources are \( M \) auxiliary fine scale environmental variables \( \{ y^i_k(u_k), i = 1, \ldots, M, n = 1, \ldots, N \} \) within each \( k \)th block data. \( N \) denotes the number of discretizing points within each block and its determination depends on the predefined finer scale value.

If the environmental variables show linear relationships with the TRMM precipitation values, the TRMM precipitation data at both coarse and fine scales can be expressed in terms of the environmental variables via multiple linear regression analysis as

\[
\begin{align*}
z(v_k) &= a + \sum_{i=1}^{M} b_i \cdot y_i(v_k) + R(v_k), \\
\end{align*}
\]

\[
\begin{align*}
z^k(u_n) &= a + \sum_{i=1}^{M} b_i \cdot y^i_k(u_n) + R^k(u_n),
\end{align*}
\]
where are \( a \) and \( b_i \) are regression coefficients for the intercept and slope of the \( i \)th variable, respectively. \( z^k(u_n) \) is the downscaled precipitation value at a target finer scale within the \( k \)th block data. \( R(V_k) \) and \( R^k(u_n) \) denote the residual components at coarse and fine scales, respectively, which cannot be accounted for by environmental variables.

If the original TRMM data, auxiliary environmental variables, and the residual components at a coarse scale can be expressed by the average values of \( N \) fine scale data within each block, respectively, (1) can be reformulated as (2)

\[
    z(v_k) = a + \sum_{i=1}^{M} b_i \left[ \frac{1}{N} \sum_{n=1}^{N} y_i^k(u_n) \right] + \frac{1}{N} \sum_{n=1}^{N} R^k(u_n)
    = \frac{1}{N} \sum_{n=1}^{N} \left[ a + \sum_{i=1}^{M} b_i y_i^k(u_n) + R^k(u_n) \right]
\]

To obtain the downscaled precipitation value \( z^k(u_n) \), the residual component value \( R_k^k(u_n) \) at a fine scale needs to be determined from (2), because the regression coefficients have already been determined from (1). To predict residual components at a fine scale, area-to-point simple kriging proposed in Kyriakidis [23] is applied to the residual components available at a coarse scale. Area-to-point simple kriging predicts the residual component values at a fine scale by a linear combination of neighboring attribute values at a coarse scale [23, 24], as shown in (3):

\[
    R^k(u_n) = \sum_{k=1}^{K} \lambda_k(u_n) R(v_k),
\]
where $\lambda_k(u_n)$ is a simple kriging weight assigned to the neighboring block residual component $R(v_k)$ at a prediction location. Since the mean value of the residual component is zero, a constant mean value required for simple kriging is set to zero and does not appear in (3).

The simple kriging weight is computed by solving the following simple block kriging system:

$$
\sum_{k=1}^{K} \lambda_k(u_n) C(v_k, v_n) = C(v_k, u_n) \quad k = 1, \ldots, K,
$$

where $C(v_k, v_n)$ and $C(v_k, u_n)$ refer to block-to-block covariance and block-to-point covariance, respectively.

Unlike conventional punctual kriging algorithms, both block-to-block and block-to-point covariances are required to compute the simple kriging weights. These two covariance values can be computed by averaging point covariance values such as the conventional block kriging system [24, 25]. The block-to-block covariance can be computed by averaging the covariance values between any two points discretizing two blocks. The covariance values between the point location and a set of points discretizing the block are averaged to obtain the block-to-point covariance. Thus, a point-support covariance (equivalently, variogram) is required to obtain the above two covariance values. However, only residuals at a coarse scale are available, so it is not feasible to directly obtain the point-support variogram. In relation to this issue, Goovaerts [24] proposed an iterative deconvolution procedure to estimate the point-support variogram from block or areal data irrespective of their shapes. Variogram deconvolution iteratively finds an optimal point-support variogram model that minimizes the difference between the regularized variogram and the variogram of the coarse scale data [24]. More specifically, after defining an initial point-support variogram model, the theoretically regularized model is computed and compared with the variogram of the coarse scale data. By computing the difference statistic from those two models, the parameters of the point-support variogram model are adjusted to minimize that difference statistic. This procedure is iteratively repeated until the stop criteria are satisfied [24, pp.113–115].

After computing the point-support variogram, the simple kriging weight is computed using (4) and then residual component values at a fine scale are predicted using (3). The final downscaling results are obtained by adding the predicted residual component at a fine scale to the trend component that can be directly computed by applying regression relationships to the auxiliary environmental variable at a fine scale in (1).

An interesting property of area-to-point kriging is its ability for coherent predictions, which means that the average of the predicted values at the fine scale points discretizing a given block reproduces the original block data value [23]. Since the residual component at a fine scale is generated by preserving this coherence property, the final downscaling results could reproduce the original TRMM precipitation values when upscaled.

The downscaling results by area-to-point residual kriging with regression are optimal in the least-squares sense. As discussed in Boucher and Kyriakidis [20]; however, downscaling should be regarded as a kind of under-determined inversion process, since there are many possible attribute values at...
a fine scale that lead to the same original values at a coarse scale when upscaled. Thus, stochastic simulation, which can generate multiple alternative realizations of unknown truth at a fine scale, is also applied for downscaling of residual components at a coarse scale. Block sequential simulation proposed by Liu and Journel [26] is applied in this study. By adopting the concept of direct sequential simulation, simulation does not require any data transformation unlike sequential Gaussian simulation. Originally, this simulation algorithm was developed to combine kriging using both block and point data with direct sequential simulation [19, 26]. As only block data (i.e., TRMM precipitation data) are available in this downscaling study, block sequential simulation with block data is implemented to generate multiple downscaling realizations. The same point-support variogram model estimated from variogram deconvolution is also used for simulation. Like the prediction by area-to-point kriging, residuals at a coarse scale are first simulated at a fine scale and final simulation results are obtained by adding the simulated residual components to the trend components.

Simulation aims at generating a set of alternative realizations with the reproduction of spatial patterns, not a single downsampling result like kriging. By comparing the differences between multiple realizations, the uncertainty attached to the prediction can be quantified. In this study, conditional variance values are computed and used as an uncertainty measure.

4. Results and Discussion

4.1. Regression Analysis Results. First, regression analysis was applied to derive statistical relationships between precipitation and the two environmental variables at the original 25 km resolution. For this analysis, DEM and NDVI were upscaled to 25 km resolution data via linear averaging. The linear correlation coefficients between precipitation and environmental variables were 0.56 and 0.26 for DEM and NDVI, respectively. In addition to the two variables, the interaction of DEM with NDVI was included in the multiple regression analysis, since forest areas showing high NDVI values tend to be located in high altitude zones (the linear correlation coefficient between DEM and NDVI was 0.65). From backward stepwise elimination, DEM and the interaction of DEM with NDVI were finally chosen as statistically significant variables and the regression relation between precipitation and those two variables was modeled as

\[ z^{*} (v_k) = 17.508 + 0.513 \text{DEM}(v_k) \]

\[-0.592 \text{DEM}(v_k) \cdot \text{NDVI}(v_k), \quad (5)\]

where \( z^{*}(v_k) \), DEM\((v_k) \), and NDVI\((v_k) \) are the estimated mean (trend component) from regression analysis, DEM, and NDVI values at a block location, respectively. This regression relationship was also applied to the same variables to generate the trend components at 1 km resolution.

The coefficient of determination \( R^2 \) was 0.37, which means that about 37% of the variation of precipitation could be accounted for by these two terms. This value of \( R^2 \) implies that a simple application of multiple regression for downscaling could not generate reliable downscaling results and that the residuals should be considered for downscaling.

4.2. Variogram Deconvolution Results. After generating residual components from regression analysis at original 25 km resolution, variogram deconvolution was applied to estimate the unknown point-support variogram of the residuals and implemented using SpaceStat software (BioMedware). By considering the target resolution (i.e., 1 km), each block of TRMM data at 25 km resolution was discretized by 25 by 25 points at 1 km resolution.

The variogram deconvolution result is shown in Figure 3. The variogram of residual components at 25 km resolution had a bounded correlation structure with a range of about 266 km. The deconvoluted point-support variogram, which, once regularized, was the most similar to the variogram model of residuals at 25 km resolution, had a greater sill value than that of the residual components at 25 km resolution. It also showed a reasonable correlation structure with two spherical models showing short- and long-range correlation structures, respectively. This bounded variogram model implies that the residual components were spatially correlated unlike the assumption of spatial independence of residuals in regression. Such spatial correlation of the residual components was accounted for in the estimation and simulation of residuals.

4.3. Area-to-Point Residual Kriging Results. The residual components at 1 km resolution were estimated by area-to-point kriging using SpaceStat software and then added to the trend components. To investigate the effects of using auxiliary environmental variables on downscaling results, the original TRMM data were directly downscaled without DEM and NDVI by applying area-to-point kriging to the original block values, not to the residuals. The point-support variogram model of the original TRMM data was also estimated by variogram deconvolution.

The two downsampling results generated by area-to-point kriging are shown in Figure 4. Overall patterns of both downsampling results were very similar to those of the coarse scale
TRMM data in Figure 1(a) (e.g., high precipitation values in Kangwon province areas). As expected, however, the direct downscaling results of the original TRMM data without DEM and NDVI showed very smoothly varying patterns, compared with that of using DEM and NDVI. In addition, local details that reflect both topographic and vegetation patterns were observed in the downscaling result with DEM and NDVI. When both downscaling results were averaged within the original 25 km resolution block, the coherence property was well preserved (Figure 5). The downscaling result without DEM and NDVI perfectly satisfied the coherence property. In the downscaling result with DEM and NDVI, the averaged values at some blocks were slightly different from the original block value. This small discrepancy was observed at some blocks where only small portions were occupied by lands or islands, and elevation and NDVI values at a finer scale were only available at those land areas. Average values were computed from the land areas within such blocks and sea areas were not considered, which resulted in the slight discrepancy. However, this difference was negligible and the coherency property was well satisfied.

To quantitatively evaluate the effects of using environmental data for downscaling, precipitation values at 69 meteorological observation stations were compared with those of two downscaling results. The predictive performance was quantified using the mean absolute error (MAE) and the root mean square error (RMSE), and validation results are presented in Table 1. As expected, the incorporation
Table 1: Predictive performance evaluation results.

<table>
<thead>
<tr>
<th></th>
<th>Downscaling without DEM and NDVI using area-to-point kriging</th>
<th>Downscaling with DEM and NDVI using area-to-point residual kriging</th>
<th>Downscaling with DEM and NDVI using ordinary kriging of residuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAE</td>
<td>19.98</td>
<td>13.03</td>
<td>17.73</td>
</tr>
<tr>
<td>Relative improvement of MAE</td>
<td>—</td>
<td>34.78%</td>
<td>11.30%</td>
</tr>
<tr>
<td>RMSE</td>
<td>27.49</td>
<td>20.03</td>
<td>24.81</td>
</tr>
<tr>
<td>Relative improvement of RMSE</td>
<td>—</td>
<td>27.14%</td>
<td>9.75%</td>
</tr>
</tbody>
</table>

of DEM and NDVI showed smaller prediction errors and improved MAE and RMSE by 34.78% and 27.14%, respectively, compared to the downscaling without the auxiliary variables. In addition, point ordinary kriging of residuals at the centroid points of original TRMM blocks was also applied to highlight the advantage of area-to-point residual kriging. The variogram model of residuals at 25 km resolution was used for implementation of point ordinary kriging. The predictive performance of point ordinary kriging of residuals showed improved MAE and RMSE (11.30% and 9.75%, resp.), compared to the downscaling without DEM and NDVI via area-to-point kriging (Table 1). However, this result comes mainly from using the auxiliary variables such as DEM and NDVI, regardless of the kriging algorithms applied to the downscaling of residuals. Point ordinary kriging of residuals showed poorer predictive performance than area-to-point residual kriging. In addition, the downscaling result by point ordinary kriging of residuals could not reproduce the original TRMM precipitation values when upscaled.

4.4. Simulation Results. Block sequential simulation was implemented using SGeMS software [19] and 50 alternative realizations of precipitation at 1 km resolution were generated. Two of the 50 realizations are given in Figure 6. The overall patterns in the simulation results were similar to those in the kriging result, but local variations were much stronger in the simulation results, which was the typical property of the simulation. When averaged within original 25 km resolution blocks, all realizations well satisfied the coherence property (results are not shown here).

Two summary statistics can be computed from the 50 realizations, the E-type (average) estimate and conditional variance for the overall pattern and the uncertainty measure, respectively, and are given in Figure 7. The E-type estimate and the kriging result in Figure 4(b) showed very similar spatial patterns. Unlike kriging variance that cannot include the effects of surrounding data values [18], conditional variance computed from the 50 realizations, which reflect the differences among the 50 alternative realizations, can be used as a quantitative measure of uncertainty. The lower variance values, which imply few differences among the 50 realizations and, therefore, lower uncertainty, were mainly observed in the southern parts where negative residuals were computed and smaller precipitation values were predicted.

These quantitative evaluation results confirmed that using both fine scale auxiliary environmental variables and spatial correlation information of residuals within the presented geostatistical scheme can generate reliable downscaling results with better predictive performance, compared to the case without fine scale environmental variables. Regarding downscaling of residuals, the presented geostatistical scheme based on area-to-point residual kriging could account for the support difference between the target scale and the input data scale, and outperformed the simple application of conventional interpolation methods by treating coarse scale block data as point data.
It should be noted that the ultimate goal of stochastic simulation is to generate alternative realizations that can be used for quantification of uncertainty attached to the spatial prediction or downscaling [18, 19], rather than either to generate the E-type estimate as shown in Figure 7(a). More importantly, all realizations are equi-probable or equally important [27], and a single best realization showing the best predictive performance should not be looked for, as discussed in Boucher and Kyriakidis [20]. A set of alternative realizations can be used for modeling the uncertainty about the unknown true precipitation at a fine scale. For example, each downscaled precipitation realization can be fed into hydrological models to generate numerous model outputs. By analyzing these model outputs, the impacts of uncertainty about precipitation information on model outputs (i.e., error or uncertainty propagation) can be investigated. Although multiple alternative realizations and summary statistics were generated via simulation in this experiment, this kind of error or uncertainty propagation problem should be regarded as an important application using simulation.

5. Conclusions

This paper has demonstrated the application of geostatistical kriging and simulation to integrate auxiliary fine scale environmental variables for downscaling of precipitation values derived from coarse scale TRMM data. The environmental variables contribute to the determination of trend components via regression analysis. The residual components that have a reasonable spatial correlation structure are accounted for during downscaling.

An experiment on downscaling of TRMM precipitation data with DEM and NDVI produced reliable downscaling results that not only preserved overall patterns in the original TRMM data, but also presented local details from DEM and NDVI. Improved predictive performance was also obtained, compared to the downscaling without the auxiliary environmental variables. Stochastic simulation provided much richer information for uncertainty modeling by generating multiple alternative realizations, unlike the single downscaling result by kriging.

The geostatistical downscaling scheme presented in this paper can be applied to areas in which ground measurement data are very sparse or even unavailable, especially in developing countries. Fine scale DEM and NDVI can be obtained from the Shuttle Radar Topography Mission (SRTM) DEM and MODIS data, respectively, and hence used effectively for downscaling of precipitation data in such areas. In addition to downscaling of TRMM precipitation data, the presented scheme can also be extended and generalized to downscaling of other coarse scale remote sensing data, such as coarse scale soil moisture data from the Soil Moisture and Ocean Salinity (SMOS) mission.

The applicability of the presented geostatistical downscaling scheme has been tested on only one TRMM data in this study. The time-series data sets are usually considered in any precipitation patterns analyses. Therefore, more extensive experiments using time-series TRMM precipitation data sets will be required to confirm the major findings of this study. The impacts of variations in correlation strengths between precipitation data and auxiliary variables on downscaling performance should also be included in future work.

Acknowledgments

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT, and Future Planning (NRF-2012R1A1A1005024).

References


