Journal of Japan Society of Civil Engineers, Ser.
B1 (Hydraulic Engineering), Vol. 70, No. 4, I_901-I_906, 2014.

STATISTICAL DOWNSCALING OF PRECIPITATION WITH A FORMATTED REGRESSION FRAME

Sunmin KIM [1], Eiichi NAKAKITA[2], Yasuto TACHIKAWA[3],
Michiharu SHIIBA[4] and Masataka INOUE[5]

[1]Member of JSCE, Dr. Eng., Associate Professor, DPRI, Kyoto University
(Gokasho, Uji, Kyoto 611-0011, Japan) E-mail:sunmin@hmd.dpri.kyoto-u.ac.jp
[2]Member of JSCE, Dr. Eng., Professor, Disaster Prevention Research Institute, Kyoto University
[3]Member of JSCE, Dr. Eng., Professor, Dept. of Civil and Earth Resources Eng., Kyoto University
[4]Member of JSCE, Dr. Eng., Emeritus Professor, Dept. of Civil and Earth Resources Eng., Kyoto University
[5]Member of JSCE, Master of Eng., Staff, Nagoya Station, Central Japan Railway Company

This paper illustrates a statistical downscaling technique considering the spatial correlation structure of precipitation. Downscaling target is 60-km resolution of daily precipitation for 20-km resolution data. We have considered a window having (3x60-km)x(3x60-km) of area, and the downscaling target is the 3x3 of 20-km resolution grids in the center of the downscaling window. For the evaluation of the proposed method, we have prepared 15 years (1979-1993) of observation data, and identify the parameters with the square root information filter scheme. We optimize the parameters on a monthly basis, and apply the regression model to 10 more years of testing period (1994-2004). The proposed regression model provides very effective and efficient results with a certain level of estimation error.

Key Words: *Statistical downscaling, precipitation, spatial correlation, regression model*

1. INTRODUCTION

The downscaling issue has been taking an important role to bridge research in climate change and impact assessment[1),2)]. Though many general circulation models (GCMs) have been developed and improved for the last several decades, their output is still too coarse for local impact assessment research. For example, many GCMs in the recent CMIP5[3)] (Coupled Model Inter-comparison Project, phase 5) provide spatial resolution that is larger than 100-km in longitudinal grid space at 35°N.

In the meantime, Japan Meteorological Agency (JMA) and the Meteorological Research Institute (MRI) of Japan has developed a couple of Atmospheric General Circulation Models (AGCMs) with 20-km and 60-km spatial resolutions[4)], within two nationwide climate change research projects, Kakushin (FY2007 - FY2011) and Sousei (FY2012 - FY2016 planned) Projects. Though these models focus on atmospheric simulation only, because of their ultra-fine resolution in the global sense, the 20-km AGCM has provided only two sets of output so far. (The ongoing Sousei project may produce more.) On the other hands, the rather coarse resolution of 60-km AGCM provides 24 sets of output with variant boundary conditions and model parameters, which gives us a glimpse of the uncertain future of climate projection[5)]. The 20-km AGCM output is able to provide reliable hydrologic impact assessment results in the major river basins of Japan[6),7)]. The goal of our research is to develop a simple yet efficient statistical downscaling (SDS) method to downscale 60-km AGCM output into 20-km resolution for the precipitation data.

The SDS issue has a long history of research and development in the field of hydrology[1)] and several types of SDS methods are already successful in other applications, such as SDSM[8)]. Basic categories of SDS include regression models (e.g. canonical correlation analysis by Schmidli et. al.[9)]), weather generators (e.g. Markov chain model by Wilks[10)]) and weather typing schemes (e.g. cluster analysis by Fowler et. al.[11)]).

The main advantage of SDS compared to DDS (dynamic downscaling) is that it does not take high computing resources, and can easily apply to any place with a minimum of observation data available[12)]. Even though the DDS method with a regional climate model (RCM) provides stable and reasonable output based on physical backgrounds, RCMs demand many computing resources, additional information, and difficult initial setup[13].

However, SDS also has limitations. Some statistical relationships between model variables are not strong enough to build a stable SDS model[12]. Most critically, we can not sure whether the statistical relationship developed with the present climate data can simulate the statistical relationship of the future climate. We do not have future data with which to evaluate the present statistical relationship or to establish the correct future one.

We have been developing an SDS method that can avoid the critical issue of the conventional SDS method and take as many advantages of DDS as possible, based on analyzing two different spatial resolutions of AGCM outputs, 20-km and 60-km. By establishing a statistical relationship between the 60-km and 20-km output for both present and future separately, and by applying the relationship to the ensemble output of 60-km AGCM, it can produce ensemble output at 20-km spatial resolution.

In this paper, we introduce an SDS method with the basic concept and evaluation results with the observed precipitation data, before it is applied to the AGCM output. Section 2 describes the basic concept of the proposed methodology and the data used in the experiments. Section 3 introduces the developed SDS method. Section 4 provides the results and related discussions. The last section concludes this paper with prospective application method to the 60-km AGCM output directly.

2. METHODOLOGY

The utilized observation data in this study is the gridded daily precipitation data of Japan, APHR_JP. The original spatial resolution of the data is 0.05 degrees in both longitude and latitude, and it was up-scaled into 0.2 degrees (around 20 km) and again into 0.6 degrees (around 60 km).

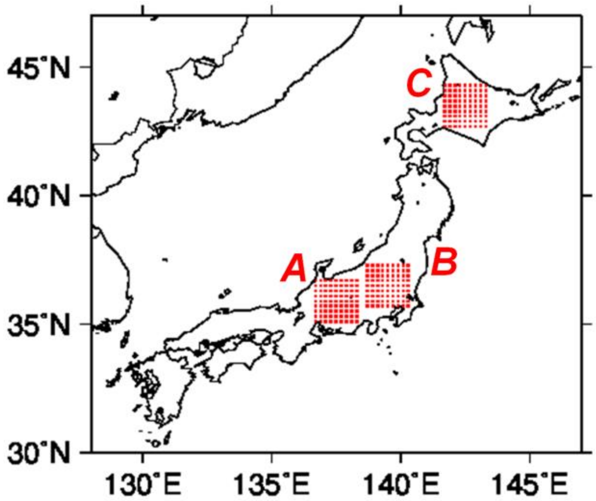

Fig. 1 Location map of the sample data. Each region is composed of 9x9 grids of 0.2 degrees (around 20-km).

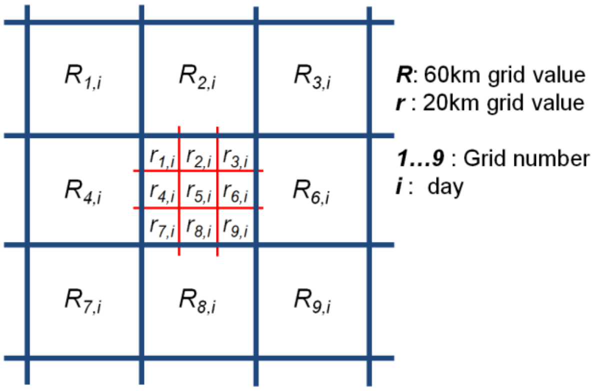

Fig. 2 Schematics of the downscaling target (20-km; r_1 … r_9) using the surrounded grids (60-km; R_1 … R_9).

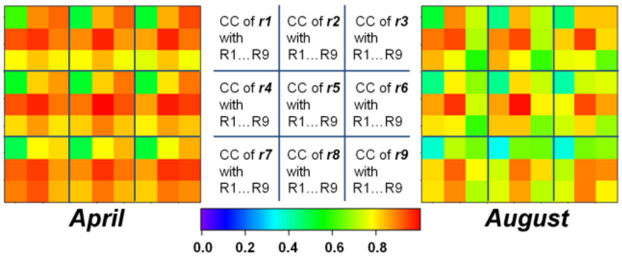

Fig. 3 Spatial correlation of the 9 grids (20-km) in the center with the surrounded grids (60-km) in the case of April and August of the region B in the Fig. 1. The correlation values are based on 25 years of observation (1979-2003).

We selected three representative regions to test the proposed method (Fig. 1). Each region is composed of 9x9 grids with 20-km resolution for each grid, or in other words, 3x3 grids with 60-km resolution (see Fig. 2). The target of downscaling is the center of the subject regions.

The basic concept of the developed downscaling method is to consider the spatial correlation pattern of precipitation, and to estimate the amount of a certain 20-km gridded precipitation (for example, r_1 in Fig. 2) based on its relationship with the surrounding 60-km gridded precipitation amount (for example, R_1, R_2 … R_9 in Fig. 2). To validate the basic concept of the method, we checked spatial correlation of r_k (k=1…9) with the surrounding R_k (k=1…9) (result from region B is shown in Fig. 3). Here, spatial correlation means the correlation of daily time series of precipitation.

As shown, r_k (k=1…9) correlates most strongly with R_5, which covers its own area, then with adjacent grids with certain tendencies. In region B, South-East directions show higher correlation than North-West directions, and the April precipitation pattern shows higher spatial correlation than the August one. These tendencies reflect regional, seasonal, and topographic characteristics, which would be sophisticated and variant in many cases.

Rather than analyzing the tendencies directly, we are focusing on generalizing a relationship that naturally includes the variant spatial correlation pattern and thus can easily apply to different regions with the same format.

3. FORMATTED REGRESSION FRAME

Considering the correlation of precipitation, we can estimate daily precipitation amount $r_{k,i}$ on a certain 20-km grid k, on a certain day i, using the surrounding 60-km gridded precipitation amount $R_{1,i}$, $R_{2,i}$ … $R_{9,i}$ and a multi-variable regression equation like Eq. 1.

$$r_{k,i} = C_{k,1}R_{1,i} + C_{k,2}R_{2,i} + \ldots + C_{k,9}R_{9,i} + \varepsilon_{k,i} \tag{1}$$

Here, $C_{k,1}\ldots C_{k,9}$ are the regression coefficients for the target grid k, and $\varepsilon_{k,i}$ is the regression residual. To estimate the regression coefficients, we can gather n days of data and reformulate the equations as a matrix, as shown in Eq. 2 or simple form as Eq. 3.

$$\begin{bmatrix} r_{k,1} \\ r_{k,2} \\ \vdots \\ r_{k,i} \\ \vdots \\ r_{k,n} \end{bmatrix} = \begin{bmatrix} R_{1,1} & R_{2,1} & \cdots & R_{9,1} \\ R_{1,2} & R_{2,2} & \cdots & R_{9,2} \\ \vdots & \vdots & \cdots & \vdots \\ R_{1,i} & R_{2,i} & \cdots & R_{9,i} \\ \vdots & \vdots & \cdots & \vdots \\ R_{1,n} & R_{2,n} & \cdots & R_{9,n} \end{bmatrix} \begin{bmatrix} C_{k,1} \\ C_{k,2} \\ \vdots \\ C_{k,9} \end{bmatrix} + \begin{bmatrix} \varepsilon_{k,1} \\ \varepsilon_{k,2} \\ \vdots \\ \varepsilon_{k,i} \\ \vdots \\ \varepsilon_{k,n} \end{bmatrix} \tag{2}$$

$$Z = Ax + v \tag{3}$$

Here, $\boldsymbol{Z}$ is the matrix with the n days of r_k values, $\boldsymbol{A}$ is the matrix with the n days of $R_{1\ldots9}$, $\boldsymbol{x}$ represents the regression coefficient $C_{1\ldots9}$, and $\boldsymbol{v}$ is the residuals. Our remaining job is to estimate the $\boldsymbol{x}$ that is minimizing $\boldsymbol{v}$, and it can be rewritten as Eq. 4, which is a typical form of the square root information filter (SRIF).

$$J(x) = \sum_{i=1}^{n} {v_i}^2 = (Z - Ax)^T (Z - Ax) \tag{4}$$

We estimated regression coefficients (or parameters) on a monthly basis using the household transformation technique on the calculation matrix for the observation data of 15 years (1979-1993), and the developed regression model was applied for the other 10 years of data from 1994 to 2003.

The efficiency of the parameters estimation was checking with the root mean square error (RMSE), and the RMSE was calculated including 0 mm/day value. Table 1 summarizes some of the efficiency values from region B, and shows successful estimation results with very small RMSE values. The most centered grid, r_5, generally shows the best results for every month and every region (see the bolded number in the Table 1).

The summer season, especially August, shows poorer efficiency than the rest of the year. But even in August, we successfully estimated the parameter, as all nine grids averaged RMSE of 0.26 mm/day for region A, 0.35 for region B, and 0.17 for region C. In the case of correlation coefficients (CC), every region shows very high correlation, even in August, as the 9 grids averaged CC of 0.95 for region A, 0.94 for region B, and 0.98 for the region C.

Table 1. Estimation efficiency of the Region B

Grids	RMSE (unit: mm/day)			
	April	**June**	**August**	**October**
r_1	0.10	0.11	0.31	0.09
r_2	0.19	0.21	0.55	0.19
r_3	0.10	0.15	0.41	0.16
r_4	0.08	0.15	0.33	0.09
$\boldsymbol{r_5}$	**0.04**	**0.08**	**0.23**	**0.05**
r_6	0.10	0.12	0.34	0.09
r_7	0.09	0.14	0.35	0.09
r_8	0.13	0.17	0.31	0.13
r_9	0.10	0.15	0.27	0.12
*MMP**	109	179	265	146

**Monthly Mean Precipitation (MMP; mm) of the R_5 grid for the 25 years is shown in the table to consider the relative magnitude of the RMSE values.*

4. RESULTS AND DISCUSSIONS

We evaluated the regression model on its applicability with the validation period of 1994-2004 by comparing simulated (downscaled) results with observed (true) values, both directly and by statistical characteristics.

Table 2. Simulation efficiency of the Region B

Grids	RMSE (unit: mm/day)			
	April	**June**	**August**	**October**
r_1	1.81	2.01	5.55	1.63
r_2	3.33	3.74	9.63	3.32
r_3	1.76	2.73	7.29	2.83
r_4	1.50	2.73	5.91	1.64
$\boldsymbol{r_5}$	**0.78**	**1.37**	**4.11**	**0.86**
r_6	1.83	2.09	5.92	1.62
r_7	1.62	2.45	6.09	1.56
r_8	2.30	2.92	5.49	2.22
r_9	1.83	2.74	4.80	2.04
*MMP**	109	179	265	146

** Monthly Mean Precipitation (MMP; mm) is shown as the Table 1.*

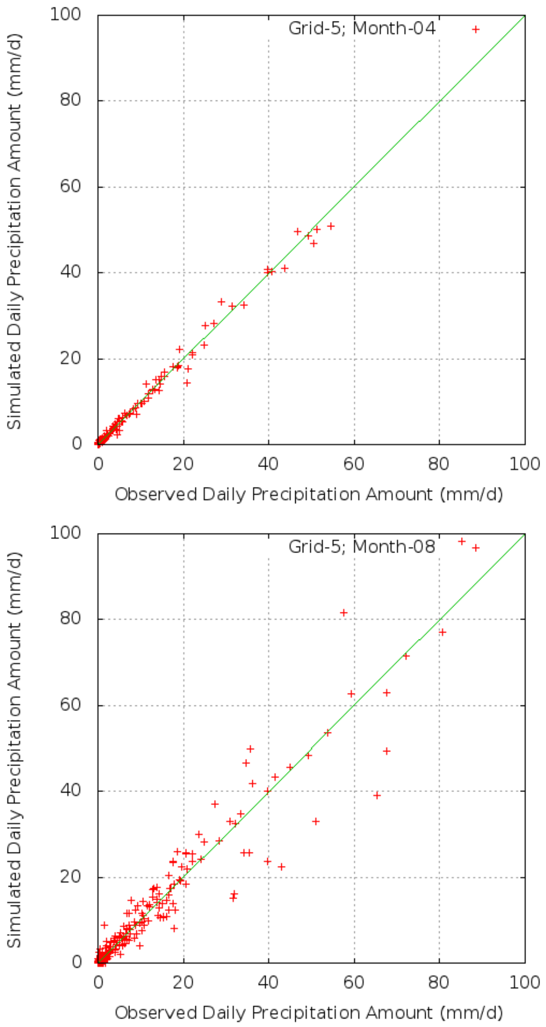

Fig. 4 Scatter plot of simulated 20-km daily precipitation compared with true values for region B, $\boldsymbol{r_5}$ (the center grid) for April (top), which is one of the best results, and August (bottom), which is one of the worst results.

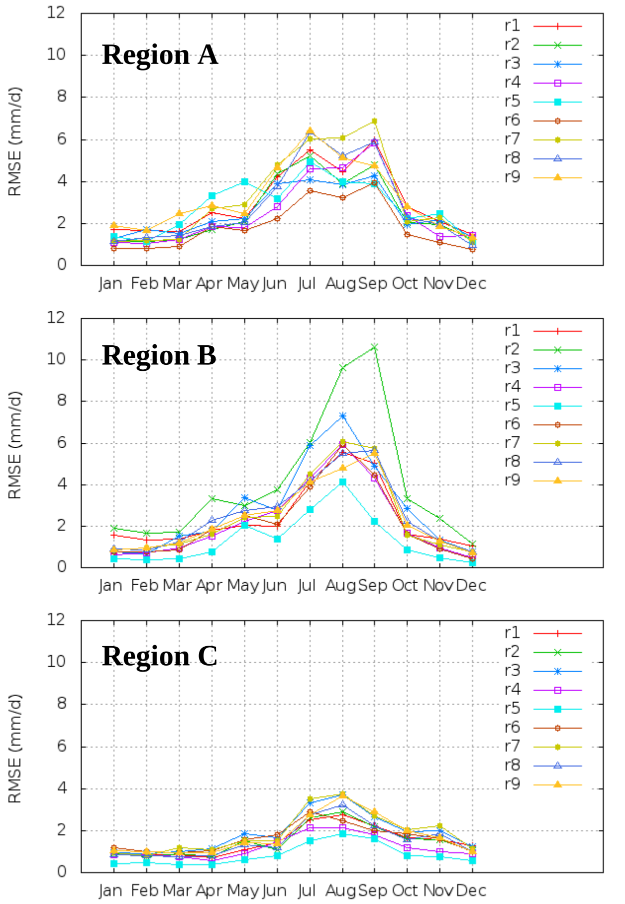

Fig. 5 RMSE of the SDS results for the observed values for the region A (top), B (middle) and C (bottom).

Figure 4 shows the representative results for region B, r_5 (the most centered grid) for April and August. In the figure, results for April shows quite successful results with 0.78 mm/day of RMSE, and the results from the August show rather dispersed pattern (RMSE: 4.11 mm/day).

Table 2 and Fig. 5 show summarized results for other seasons and regions. Table 2 shows the representative RMSE from the region B with the variation of each grid. In Table 2, the most centered grid shows the best results for every month, but region A of Figure 5 trends differently. Efficiency of the regression model introduced here is largely affected by correlation with the adjacent area's precipitation pattern. Some grids or regions might have a good pattern relationship with the precipitation of surrounding areas, making the regression model work better for them than others, due to particular topographic shape and/or atmospheric behavior of the region, and vice versa.

August shows the worst results in most cases. One reason for this is the poor spatial correlation of this month as shown in Fig. 3 (right). Another is the large amount of precipitation in August, which is in the middle of summer in Japan. As shown in Table 2, the RMSE magnitude is correlated with the monthly precipitation amount. Lastly, the pattern of the RMSE from the simulated results (Table 2) is very similar to the pattern of the parameter estimation efficiency (Table 1). In other words, the simulation efficiency in RMSE can be imagined with the parameter estimation efficiency.

In this study, we have developed an SDS method to downscale 60-km precipitation information into 20-km resolution on a daily basis. Simulation results do not need to match observation values perfectly. If they are good enough to show the statistical characteristics of the observation, we can approve the proposed regression model. To evaluate these statistical characteristics, we have checked the frequency of rainfall intensity with histograms (Fig. 6), and error percentage of monthly precipitation amount (Fig. 7). The rainfall intensity interval in Figure 6 is basically 2 mm/day except for the first bar of the histogram with the percentage of the 0 mm precipitation amount.

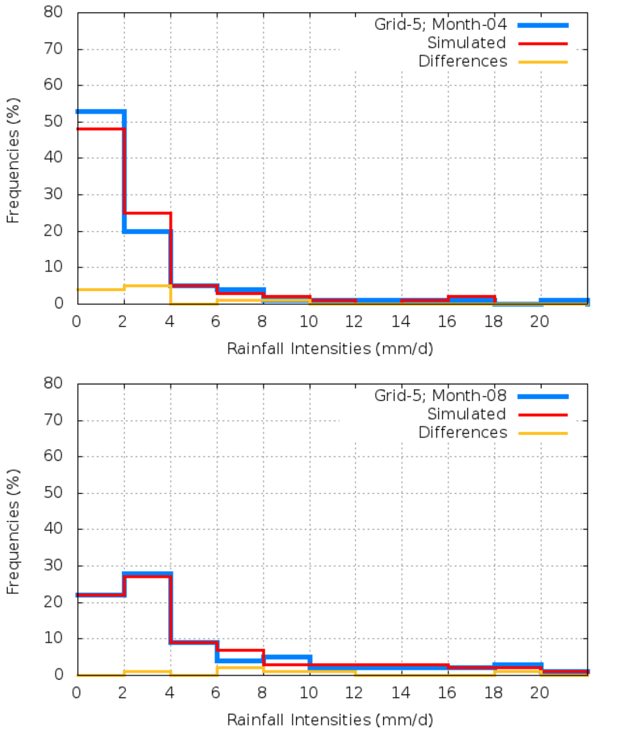

Fig. 6 Histogram of the SDS 20-km daily precipitation amount comparing with the true values, for region B, $\boldsymbol{r_5}$ (the center grid) for April (top) and August (bottom). Note that the first bar is the frequency of 0 mm/day, and the others are the frequencies of each 2 mm/day interval.

In Figure 6, the results from August show a very good match with observed rainfall intensities in terms of frequency, and the results from April also shows good match except for frequencies lower than 2 mm/day. This is an encouraging result, but the discrepancy of frequency under 2mm/day can make a significant difference in accumulated values over a long duration of simulation. Thus, we surveyed the error ratio of monthly precipitation and summarized the results (Fig. 7). For a downscaled grid i, for a certain month j, the error ratio E_{ij} is

$$E_{ij} = \frac{S_{ij} - O_{ij}}{O_{ij}} \times 100(\%) \qquad (5)$$

where, S_{ij} is the simulated monthly precipitation amount and O_{ij} is the observed one, during the examination periods (1994-2003).

The normalized error ratio with the monthly precipitation amount does not depend on season or month, and most of it is within ±10%. More specifically, the error ratio E_{ij} is showing a normal distribution around zero and the standard deviations are 4.4%, 5.4% and 5.1% for regions A, B and C, respectively. Surely there is much room for improvement, but this is a plausible result when we consider the simplicity of the proposed method.

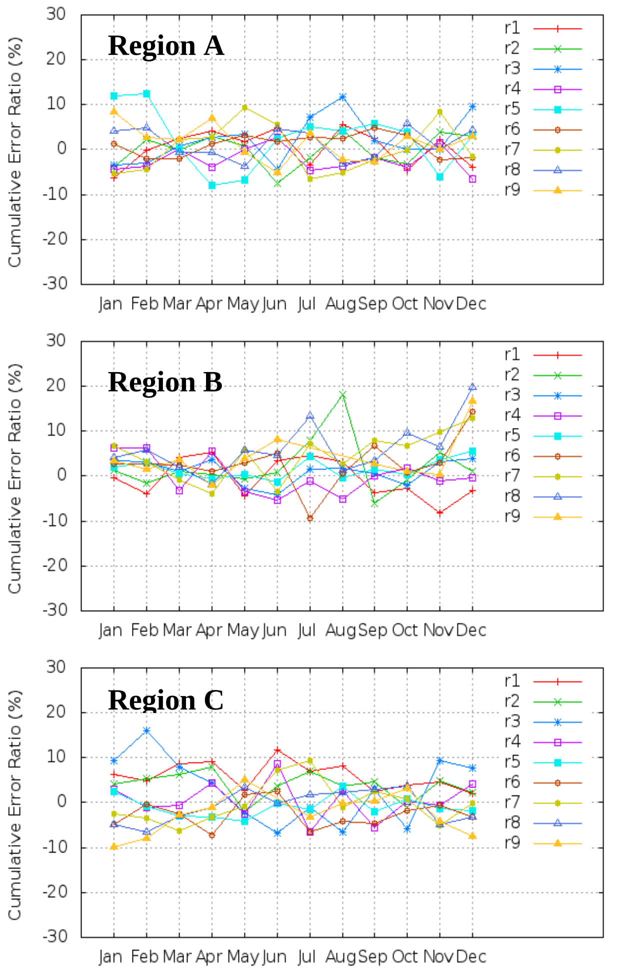

Fig. 7 Error percentage of monthly precipitation amounts for the region A (up), B (middle) and C (down).

Finally, representation of spatial pattern was evaluated with correlation of the simulated monthly mean precipitation with the observed one in the downscaled 9 grids. As shown in Table 3, the spatial pattern is successfully well represented in the downscaled results with high correlation coefficient.

Table 3. Correlation coefficients of simulated monthly mean precipi. with the observation

Month	Region A	Region B	Region C
Jan.	0.828	0.998	0.925
Feb.	0.912	0.993	0.929
Mar.	0.992	0.995	0.959
Apr.	0.944	0.984	0.976
May.	0.973	0.983	0.988
Jun.	0.963	0.964	0.941
Jul.	0.955	0.917	0.946
Aug.	0.944	0.963	0.981
Sep.	0.964	0.989	0.970
Oct.	0.973	0.978	0.974
Nov.	0.971	0.982	0.940
Dec.	0.932	0.985	0.963

5. SUMMARY AND FURTHER RESEARCHES

An SDS technique is proposed with a formalized regression frame considering spatial correlation structure of precipitation. Downscaling windows in a 3x3 60-km grid were considered to downscale the centered 60-km grid into 9 grids of 20-km spatial resolution. For the evaluation of the proposed method, we used 25 years of gridded observation data to calibrate regression model parameters (1979-1993) and evaluate downscaled results (1994-2004). We optimized the parameters on monthly basis, and found that the regression model provides very effective and efficient results except for the summer season in terms of root mean square error. However, statistical characteristics, such as the frequencies of rainfall intensities, show stable patterns for every season and region, with a certain level of mismatch. Overall there was around ±5% margin of error in the monthly precipitation amount.

The proposed SDS technique can be applied to the 60-km AGCM of MRI (officially, MRI-AGCM 3.1H & 3.2H) to downscale it into 20-km spatial resolution based on the analysis of the original 20-km resolution AGCM (MRI-AGCM3.1S & 3.2S). Application method can be 1) upscaling the original 20-km AGCM output into 60-km resolution, 2) estimating the regression parameters for both the present and future climate using the original 20-km one and the upscaled 60-km one, 3) applying the estimated parameters to the original 60-km ensemble output from the MRI-AGCM3.1H & 3.2H and downscaling them.

However, statistical characteristics and spatial pattern can differ between the original 60-km AGCM output and the upscaled one from the 20-km AGCM output. One simple solution is to modify the original 60-km data to have the characteristics of the upscaled one. In this case, we need an additional consideration while we handle the ensemble output of the 60-km AGCM: to preserve the original ensemble characteristics.

Secondly, a certain estimation bias from the proposed SDS method should be carefully considered. This may not be avoidable, however, information about the bias range and amount should be checked before it is applied to the secondary usage such as for impact assessment research.

For more generalized application of the proposed method, we are now considering some variation of the method. For example, the SDS method illustrated here can be modified to apply to different spatial and different temporal resolution of data. The downscaling window (3x3) can be modified for other dimensions, such as (5x5). Finally, the proposed method should be evaluated for its applicability to different regions and different climate conditions.

ACKNOWLEDGMENTS

This work was done under the framework of the “Precise impact assessments on climate change” SOUSEI Program of MEXT, Japan and supported by the Wakate-B of JSPS (Project No: 23760459).

REFERENCES

1) Fowler, H., Blenkinsop, S. and Tebaldi, C.: Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling, International Journal of Climatology, Vol. 27, pp.1547-1578, 2007.
2) Charles, S., Bari, M., Kitsios, A. and Bates, B.: Effect of GCM bias on downscaled precipitation and runoff projections for the Serpentine catchment, Western Australia, International Journal of Climatology, Vol. 27, pp.1673-1690, 2007.
3) Taylor, K.E., R.J. Stouffer, G.A. Meehl: An Overview of CMIP5 and the experiment design, Bulletin of the American Meteorological Society, Vol. 93, pp. 485-498, doi:10.1175/BAMS-D-11-00094.1, 2012.
4) Kitoh, A., Ose, T., Kurihara, K., Kusunoki, S., Sugi M. and KAKUSHIN Team-3 Modeling Group: Projection of changes in future weather extremes using super-high-resolution global and regional atmospheric models in the KAKUSHIN Program: Results of preliminary experiments, Hydrological Research Letters, Vol. 3, pp.49-53, 2009.
5) Kusunoki, S., Mizuta, R. and Matsueda, M.: Future changes in the East Asian rain band projected by global atmospheric models with 20-km and 60-km grid size, Climate Dynamics, doi:10.1007/s00382-011-1000x, 2011.
6) Kim, S., Tachikawa, Y., Nakakita, E. and Takara, T.: Hydrologic evaluation on the AGCM20 output using observed river discharge data, Hydrological Research Letters, Vol. 4, pp.35-39, 2010.
7) Kim, S., Tachikawa, Y., Nakakita, E., Yorozu, K.. and Shiiba, M.: Climate change impact on river flow of the tone river basin, Japan, Journal of Japan Society of Civil Engineers, Ser. B1 (Hydraulic Engineering), Vol. 67, No. 4, pp.I_85-I_90, 2011.
8) Wilby, R.L. and Dawson, C.S.: The statistical DownScaling Model (SDSM): Insights from one decade of application, International Journal of Climatology, Vol. 33, pp.1707-1719, 2013.
9) Schmidli, J., Goodess, C.M., Frei, C., Haylock, M.R., Hundecha, Y., Ribalaygua, J. and Schmith, T.: Statistical and dynamical downscaling of precipitation: An evaluation and comparison of scenarios for the European Alps, Journal of Geophysical Research, Vol. 112, doi: 10.1029/2005JD007026, 2007.
10) Wilks, D.S. and Wilby, R.L.: The weather generation game: a review of stochastic weather models, Progress in Physical Geography, Vol. 23 pp. 329-358, 1999.
11) Fowler, H.J., Kilsby, C.G., C’Connell, P.E. and Burton, A.: A weather-type conditioned multi-site stochastic rainfall model for the generation of scenarios of climatic variability and change, Journal of Hydrology, Vol. 308, pp. 50-66, 2005.
12) Wilby, R.L. and Wigley, T.M.L.: Downscaling general circulation model output: a review of methods and limitations, Progress in Physical Geography, Vol. 21 pp. 530-548, 1997.
13) Wilby, R.L., Hay, L.E., Gutowski Jr., W.J., Arritt, R.W., Takle, E.S., Pan, Z., Leavesley, G.H. and Clark, M.P.: Hydrological responses to dynamically and statistically downscaled climate model output, Geophysical Research Letters, Vol. 27, pp. 1199-1202, 2000.

(Received September 30, 2013)