

Research Article

Cropland Precipitation Prediction for Food Produce of the Weihe Basin Using Statistical Downscaling Model

¹Shibao Lu, ¹Xiao Bai, ²Liang Pei and ¹Xiaoqing Gan

¹School of Urban-Rural Planning and Management, Zhejiang University of Finance and Economics, Hangzhou 310018, China

²Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

Abstract: This study describes the establishment of deferent statistical downscaling equations, using the existing Shannxi hydrological data of the Weihe basin, to downscaly analyze its monthly rainfall and temperature for food produce. The combination of main element analysis with the multi linear regression of downscaling is to get the statistical relationship between the atmosphere circling factors as outputs of global weather model with the rainfall and temperature data of 8 hydrological stations. Using the gradual regression method to analog the rainfall tells the relative error in a range of $\pm 5\%$ but that of standard deviation within $-25\sim-40\%$, all smaller than the real time values and R^2 about $0.45\sim 0.7$. For simulating the temperature, it shows almost no error of the average monthly temperature and the standard deviation obviously less than that of the real time and R^2 ranging $0.85\sim 0.95$ most of the time, except about 0.7 in May, July and September. So, the statistical downscaling method of main element analysis has had a good result in imitating cropland precipitation and temperature for food produce.

Keywords: Cropland precipitation, food produce, gradual regression, statistical downscaling, weihe basin

INTRODUCTION

In order to get more scientific and convincing results, the hydrological study has to start to accept or integrate information of its neighbor disciplines so that it has extended from a single to a compound science, with the grand background of a sharp global population increase, serious fresh water shortage, severe environmental pollution and the world climate change (James, 2000; Mpelasoka *et al.*, 2001; Clark *et al.*, 2001; Stenseth *et al.*, 2002). The climate change has obvious effect on hydrological factors and so a new consideration has come out to combine climate models with hydrological models for a study of climate effect onto hydrology and water resources. But the scale matching problem exists in the combination, although the Global Circulation Mode (GCMs) provides information on future climate change based upon simulation of the current climate using existing data and the hydrological model plays an important role in study of hydrological element change, water circulation and water resource evaluation (Wilby *et al.*, 2002, 2003).

The three has a common point that they need the AOGCM provide large scale information. The first that can use the Regional Climate Model (RCM) coupling the AOGCM to predict future climate changes,

demonstrating an advantage of physical significance, say, that can be used in any place and does not affected by observation data, without requirement to resolution rate.

This study describes the establishment of deferent statistical downscaling equations, using the existing Shannxi hydrological data of the Weihe basin, to downscaly analyze its monthly rainfall and temperature for food produce. The combination of main element analysis with the multi linear regression of downscaling is to get the statistical relationship between the atmosphere circling factors as outputs of global weather model with the rainfall and temperature data of 8 hydrological stations.

STATISTICAL DOWNSCALING METHOD MODEL

Category of downscaling and their individual advantages: As mentioned, the limitation of C, CMs in regional climate change is amended by converting the low resolution outputs into underling surface information such as cropland precipitation and air temperature, for the downscaling method works on condition that the regional climate changes expressed in large scale (such as continental or even spatial scales).

Corresponding Author: Shibao Lu, School of Urban-Rural Planning and Management, Zhejiang University of Finance and Economics, Hangzhou 310018, China

This work is licensed under a Creative Commons Attribution 4.0 International License (URL: <http://creativecommons.org/licenses/by/4.0/>).

There are approximately three types of downscaling methods, i.e., first, the dynamic downscaling method (Dynamical Downscaling, DD); second, the statistical downscaling method (Statistical Downscaling, SD); and third, the downscale method combining Statistics and Dynamics (DD and SD). The difference is that the DD method has the RCMs and the LAMs imbedded in the GCMs while using effective boundary conditions provided by C, CMs. It can then obtain climate change information of local scale. The SD method tries to establish a statistical relationship of mean value between the regional variables and the grand or free convection variables. Based on the fundamental data, it uses the established relationship to analog climate change information of a river basin or acquisition method. The common point of the two is that they all need C, CMs to simulate the phenomenon that provides climate change information of large scale. The DD and SD method, easy, flexible and fast running in comparison to simple methods, has the advantages of the above two while avoiding their shortcomings, especially, much better than the DD in evaluation of weather impacts. In rainfall prediction, it can, with help of GIS spatial analysis technology, increase the precision, based upon the hydrological features and the temporal and spatial variation of rainfall.

Basic principles of SD method: To predict future phenomenon (rainfall or temperature) of a basin climate, the SD method, first to establish a statistical relationship between the large scale climosequence and basin climate factors employing many-year observations and is cross checked with independent observation data, can then put the relationship into the large scale climate information output from the C, CMs (Clark *et al.*, 2001; James, 2000; Mpelasoka *et al.*, 2001; Stenseth *et al.*, 2002). The statistical function equation is as Eq. (1):

$$Y = F(X)$$

In which, “y” is the variant of a basin weather forecast, “z” is the factor of large scale weather forecast and “f” is the statistical relationship of the two variants. Normally, the relationship is unknown and developed through statistics of observation data or simulation of the basin climate model. The core issues of the SD method include: 1 determination of the statistical relationship between the large scale forecast factors and site predictand. It is noted that selection of different forecast factors will generate different expectations of future climate phenomenon. The selection is mainly determined through seasonal correlation analysis, scatter diagram and partial correlation analysis. The selection should follow the following 4 principles: first, clear physical significance with predictand; second, strong correlation and consistency with predictand;

third, common factors for measured data and GCMs outputs; and fourth, factors fit for GCMs and good for simulation. The 3 basic assumptions are preconditions for the use of SD method: some statistical relationship between parameters of large scale climate field and basin climate field; effective GCMs to get downscaling relations at a designated scale; and effective statistical relationship at indefinite weather phenomena.

SD model and evaluation indexes:

SD model: In general, there are 5 important steps in SD application study: selection of large scale predictors; model determination and calibration; check mode with independent observation data; generation of future climate phenomenon with SD applied to AOGCM; and diagnosis analysis of future weather.

In order to evaluate visually the simulation results of the two methods, this study takes fitting degree R^2 , mean value relative error R_{mean} and standard deviation relative error R_{sd} evaluation indexes.

Mean value relative error R_{mean} :

$$R_{mean} = \frac{\overline{X_{sim}} - \overline{X_{obs}}}{\overline{X_{obs}}} \times 100\% \tag{1}$$

In which,

$\overline{X_{obs}}$ = The average of measured monthly rainfall

$\overline{X_{sim}}$ = The mean value of monthly rainfall simulated with SD method

Standard deviation relative error R_{sd} :

$$R_{sd} = \frac{\sigma_{sim} - \sigma_{obs}}{\sigma_{obs}} \times 100\% \tag{2}$$

In which σ_{obs} standard deviation of measured monthly rainfall and σ_{sim} Standard deviation of monthly rainfall simulated with SD method.

Gradual regression method: Firstly, it is to define an index that measures the importance of a factor to the forecast object, not like the way to put all forecast factors in the regression equation, so as to choose the factor that affects remarkably the “y”. Screening is done step by step and one step selects only one factor, in the process of establishing the regression equation, which can get maximized residual sum of squares and downscaling and it can pass significance testing of assigned credibility-F check. If the first selected factor is x_1 , the first-step transitional equation can be:

$$y = b_0^{(1)} + b_1^{(1)}x_1 \tag{3}$$

In which, superscript (1) is regression coefficient. Next step is to select second factor according to the

standard balancing the importance of factors. Therefore, this factor is assumed as x_2 , the second equation is:

$$y = b_0^{(2)} + b_1^{(2)}x_1 + b_2^{(2)}x_2 \quad (4)$$

The process continues until there is not factors obviously useful to “y” before they can be introduced into the regression equation. At the same time, factor removal should be carried out when an introduced factor would act less to the “y” as its followers come in and require cooperation between each other. The first three steps should not do any removals and checks. So, the basic concept of gradual regression goes like this: there should be no more put-in nor removal, otherwise, the regression is not yet finished.

ANALYSIS OF RAINFALL SIMULATION RESULTS

Calibration result of gradual regression algorithm:

The gradual regression algorithm is good in simulating the average of monthly rainfall series, with relative error less than $\pm 5\%$, but not so good in that of standard deviation, with relative error $-25\sim-40\%$, all smaller than the measured value; while R^2 is within the range of about $0.45\sim 0.7$.

Cross check with gradual regression algorithm:

To check SD model effects, the gradual regression algorithm is used, with monthly predictors of every station, 1961-1990 and calibrated coefficients combining with predictors of 1991-2000 to get the simulated cropland precipitation data of 1991-2000. The gradual regression algorithm can simulate remarkably the average value of the monthly cropland precipitation series, relative error within $\pm 10\%$ and not more than 25% of September and December; but the standard deviation is not good, the relative error $-25\sim-40\%$, averagely, less than the measured value; while the R^2 is controlled within $0.45\sim 0.8$ except those of January and September. Analogue of the verification period is not as good as that of the calibration period and no matter what period is, the verification or the calibration, the standard deviation is less than the measured value, meaning a bad simulation to the extreme air temperature.

CONCLUSION

The SD method has good physical significance in imitating the global weather, precise in climate information and downscaling from the grand to a basin scale through statistical relationship. In comparison with the DD method, the SD is easy in computation and

suitable to different climates. This study describes the establishment of different SD equations, using the hydrological data of the Weihe River basin, to downscale the monthly rainfall and temperature series and do down-dimensioning and compressing treatment of the NCEP output data through the employment of the principal component analysis. That way has grasped the main feature of the general atmospheric circulation factors and avoided a curse of dimensionality in downscaling the global weather. The evaluation indexes of this study are selected the establishment of the SD model based upon gradual regression algorithm, with predictors of 8 stations in Shannxi on the Weihe River and the measured monthly temperatures of 1961-1990. The gradual regression algorithm is good in simulating the mean value of the monthly temperature series, almost no error, but it is not good in standard deviation simulation, tat of temperature standard deviation less obviously than that of the measured value. R^2 is about 0.7 in May, July and September and about $0.85\sim 0.95$ of the other months. The use of principal component analysis tells a good result in simulation of rainfall and temperature.

ACKNOWLEDGMENT

The study is funded by the National Natural Science Foundation of China (Grant No.: 51379219), the National Social Science Foundation of China (Grant No.: 14BGL205) and the Social Science Foundation of Zhejiang province (Grant No.: 14YSXK02ZD-1YB).

REFERENCES

- Clark, J.S., S.R. Carpenter, M. Barber, S. Collins, A. Dobson *et al.*, 2001. Ecological forecasts: An emerging imperative. *Science*, 293: 657-660.
- James, M., 2000. Predictions of climate change over Europe using statistical and dynamical downscaling techniques. *Int. J. Climatol.*, 20: 489.
- Mpelasoka, F.S., A.B. Mullan and R.G. Heerdegen, 2001. New Zealand climate change information derived by multivariate statistical and artificial neural networks approaches. *Int. J. Climatol.*, 21: 1415-1433.
- Stenseth, N.C., A. Mysterud, G. Ottersen, J.W. Hurrell, K.S. Chan and M. Lima, 2002. Ecological effects of climate fluctuations. *Science*, 297: 1292-1296.
- Wilby, R.L., C.W. Dawson and E.M. Barrow, 2002. SDSM-a decision support tool for the assessment of regional climate change impacts. *Environ. Modell. Softw.*, 17: 147-159.
- Wilby, R.L., O.J. Tomlinson and C.W. Dawson, 2003. Multi-site simulation of precipitation by conditional resampling. *Climate Res.*, 23: 183-194.