TRANSFER FUNCTION APPROACHES

Transfer functions are downscaling techniques that rely upon the statistical relationships among the large scale predictors and local scale predictands. They are regression based downscaling techniques that are simple and requires very less computation compared to other methods. The regression based methods however are constrained to the places wherein properly predictor-predictand relationships are established. The transfer function techniques can however be different and depends on the mathematical transfer function, predictor variables, or statistical fitting procedure used. The methods can be linear and non-linear regression, artificial neural networks, canonical correlation, and principal component analysis or independent component analysis.

Jeong et al., 2012 compared three linear models (multiple linear regression (MLR) with ordinary least squares (OLS) estimates, robust regression and ridge regression) and one non-linear model (artificial neural networks (ANNs)), to determine the best suited transfer function in statistical downscaling (SD) models for the temperature (daily maximum and minimum) and precipitation (occurrence and amount of rainfall).

Regression equations are used to establish predictor-predictand relationships in transfer functions, which uses linear relationships and non-linear relationship, classified as linear transfer functions and Non-linear transfer functions (Hadipour et al., 2016).

Trigo & Palutikof, 1999 developed transfer functions to predict temperature (minimum and maximum) at local scales, from the GCM outputs using linear ANN models and non-linear ANN models. The performances of non-linear ANN models were assessed to find out their advantages over non-linear models.

The determination of the most appropriate transfer function for a particular downscaling problem is the most important step in Statistical Downscaling using transfer functions. Linear transfer functions are quite simpler than the complex non-linear functions and for this reason, most statistical downscaling studies uses linear transfer functions to model the relationship between AOGCM predictors and local predictands. Also linear transfer functions that are calibrated separately for different seasons or months has shown better results as compared to annually calibrated models. Quite contrary to this, several other studies shows that linear transfer functions fails to capture the predictor-predictand in statistical downscaling relationship appropriately, and suggests using complex non-linear transfer functions. The advantage of non-linear transfer function over linear transfer functions are that, non- linear transfer functions are not limited by the linearity of the predictor-predictand relationship.

The Artificial Neural Networks (ANNs) are non-parametric transfer functions with their main advantage that they can establish a strong predictor-predictand relationship in downscaling using statistical methods. At the same time, ANN has drawbacks such as; it can get trapped in the local minima and the selection of model architecture/training algorithm. There are chances of over fitting, which can be avoided if sample sizes are large enough. ANNs however requires high operational costs and are very complex in nature (Jeong et al., 2012).

Ghosh & Mujumdar, 2006 combined weather typing and transfer function methods, for downscaling mean sea level pressure (MSLP), using principal component analysis, fuzzy clustering, and linear regression along with introducing a seasonal component to the model to project monthly precipitation over Orissa. This method, based on linear regression and fuzzy clustering is comparatively simple in terms of computation. Fuzzy clusters have an advantage over the limitations of the rigidity of hard clusters. The use of fuzzy cluster is demonstrated later in this chapter.

Ghosh & Mujumdar, 2007 states, downscaling using transfer functions are the most popular methods of downscaling. A number of methods are used including linear and nonlinear regressions, artificial neural network, fuzzy rule-based system, support vector machine, analogue methods. The methods differ according to the choice of mathematical transfer function, predictor variable, and statistical fitting procedure chosen.

Chen et al., 2014 used quantile mapping method (transfer function method) to downscale monthly precipitation in two steps. First, the transfer function was derived by fitting the first and third order polynomials the observed station and the precipitation simulated by the climate model for the reference period. Secondly the transfer functions were used to downscale the climate model simulated monthly precipitation for the future or validation period. They also observed that the transfer function’s ability to reproduce the probability distribution of the monthly precipitation as well as the correcting the bias of the grids and matching them.

The following example is taken from the book, 'Floods in a changing climate: Hydrologic modeling', by Mujumdar & Nagesh Kumar, 2010, which shows the transfer function approach through linear regression downscaling technique.

A transfer function approach for projection of future monthly precipitation based on fuzzy clustering and linear regression is explained through an example for the Orissa meteorological subdivision, India, with reanalysis data of MSLP as predictor and observed precipitation as predictand, from (Ghosh & Mujumdar, 2006). Gridded MSLP data used in the downscaling are obtained from NCEP reanalysis (E Kalnay, M Kanamitsu, R Kistler, W Collins, D Deaven, L Gandin, M Iredell, S Saha, G White, J Woollen, Y Zhu, M Chelliah, W Ebizusaki, W Higgins, J Janowiak, K C Mo, C Ropelewski, J Wang, A Leetmaa, R Reynolds, Roy Jenne, 1995). Monthly average MSLP outputs from 1948 to 2002 were obtained for a region spanning 15–25◦ N in latitude and 80–90◦ E in longitude that encapsulates the study region. Table 1 gives a list of GCMs with available scenarios used in the study.

Solution:

A regression relationship is established between NCEP reanalysis data for circular patterns and observed precipitation which is used to project future precipitation from GCM projections. The method consists of Principal Component Analysis (PCA), fuzzy clustering and linear regression with seasonality terms. The Figure 1 shows the downscaling based on Fuzzy clustering.

Step 1: Interpolation and standardization

The GCM grid points and NCEP grid points does not overlap and thus to find out the GCM output values at NCM grid points, interpolation is required, carried out with linear inverse square procedure using spherical distances. This interpolation will result in systematic biases in the mean and variance of GCM predictors, reduced by employing standardization. Standardization consists of subtracting the mean and then dividing it by standard deviation of the predictor variable for a baseline period of 1960-1990, for the NCEP/NCAR as well as GCM outputs.

Table 1: GCMs and the scenarios used

GCM	Scenario used
CCSR/NIES coupled GCM	A1, A2, B1, B2
Coupled global climate model (CGCM2)	IS92a, A2, B2
HadCM3	IS95a, (GHG + ozone + sulphate), A2
ECHAM4/OPYC3	IS92a, A2, B2
CSIRO-MK2	(IS92a + sulphate), IS92a, A1, A2, B1, B2

Step 2: Preprocessing – PCA

Pre-processing involves application of PCA and fuzzy clustering to the predictors before they are actually used in the downscaling model (PCA has been explained in the Data preprocessing section). PCA when applied on the 25 NCEP grid point predictors, showed that the 98.1% variability is caused by the first ten Principal Components, thus the first ten components are selected to model the stream flow.

Step 3: Preprocessing – Fuzzy clustering

Fuzzy clustering classifies the PCs into different classes or clusters and a membership value or rank is given to every data point. The parameters used in this step are number of clusters (c) and fuzzification parameter (m). The fuzzification parameter (m) controls the degree of fuzziness of the PCs after classification. The fuzzification is the degree of overlap between the clusters. The parameters for fuzzy clustering are decided according to the cluster validity indices, namely fuzziness performance index (FPI) and normalized classification entropy (NCE). The FPI determines the fuzziness caused by the different classes.

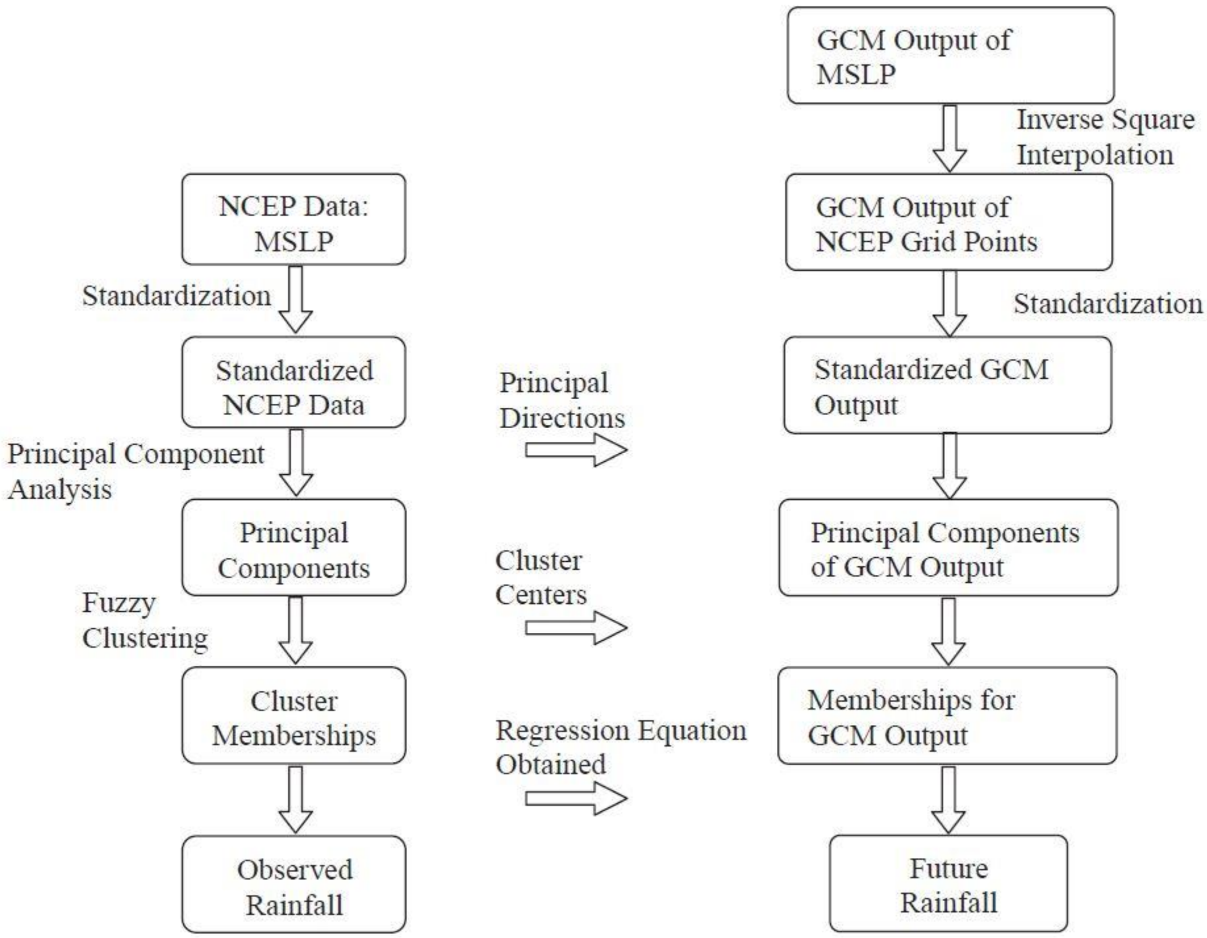

Figure 1: Fuzzy Clustering based downscaling (Source: Ghosh & Mujumdar, 2007)

$$FPI = 1 - \frac{cF-1}{c-1} \quad (1)$$

Where,

$$F = \frac{1}{T}\sum_{i=1}^{c}\sum_{t=1}^{T}(\mu_{it})^2 \quad (2)$$

μ_{it} is the membership in cluster 'i' of the PCs at time t. NCE determines the degree of disorganization generated by a given number of classes as:

$$NCE = \frac{H}{\log c} \quad (3)$$

Where,

$$H = \frac{1}{T}\sum_{i=1}^{c}\sum_{t=1}^{T} -\mu_{it} * \log(\mu_{it}) \quad (4)$$

The equations (1) and (2) are used to optimize the number of classes/clusters. Here, FPI was obtained as 0.25 for m= 2.0 and c= 2.

Step 4: Linear regression

In this step, monthly precipitation is modeled with PCs, their membership values in each cluster and the cross product of membership values and PCs. A seasonality term is introduced to represent the seasonality. The linear regression equation is given by:

$$P_t = C + \sum_{i=1}^{I-1} \beta_i * \mu_{it} + \sum_{k=1}^{K} \gamma k * pc_{kt} + \sum_{i=1}^{c}\sum_{t=1}^{T} \rho_{ik} * \mu_{it} * pc_{kt} \quad (5)$$

With,

$$C = C^0 + C^1 * \sin\left(\frac{2\pi p}{12}\right) + C^2 * \cos(\frac{2\pi p}{12}) \quad (6)$$

$$\beta_i = {\beta_i}^0 + {\beta_i}^1 * \sin\left(\frac{2\pi p}{12}\right) + {\beta_i}^2 * \cos(\frac{2\pi p}{12}) \quad (6)$$

$$\gamma_k = {\gamma_k}^0 + {\gamma_k}^1 * \sin\left(\frac{2\pi p}{12}\right) + {\gamma_k}^2 * \cos(\frac{2\pi p}{12}) \quad (7)$$

$$\rho_{ik} = {\rho_{ik}}^0 + {\rho_{ik}}^1 * \sin\left(\frac{2\pi p}{12}\right) + {\rho_{ik}}^2 * \cos(\frac{2\pi p}{12}) \quad (8)$$

Where,

P_t: Precipitation at time 't'

pc_{kt}: k^{th} PC of the CP at time 't'

μ_{it}: Membership in cluster 'i' of the PCs at time 't'

K: number of PCs used

i: number of clusters

β_i, γ_k, ρ_{ik}: coefficients of μ_{it}, pc_{kt} and their product terms respectively

C: constant

In each cluster, membership values μ_{it} are allocated to the different points, on the basis of fuzzy c-means algorithm. Seasonality is introduced in the form of ‘p’, where the value of p represents the serial number of the month within a year, i.e., p = 1, 2, ...12.

Step 5: Validating the regression

For a regression model, the goodness of fit is the correlation coefficient (r) between the observed and the predicted variables (here, precipitation). For validation, k-fold cross validation is adopted (k=10), and r value was obtained as 0.94 for running the model and 0.922 for testing the model. The Nash and Sutcliffe (1970) coefficient, which was also used to test the goodness of fit of the model. The Nash and Sutcliffe (1970) coefficient which lies between 0 and 1, is given by,

$$E = 1 - \frac{\sum_t (P_{ot} - \overline{P_{pt}})^2}{\sum_t (P_{ot} - P_o)^2} \qquad (9)$$

Where,

P_{ot}: Observed precipitation at time ‘t’

$\overline{P_{pt}}$: Predicted precipitation at time ‘t’

P_o: Mean observed precipitation

Here, the value of E was obtained as 0.83, which is satisfactory.

Step 6: Results for projection

The future projection of precipitation was obtained for wet period (JJAS) and dry period was obtained separately as shown in the Figure 2. It is found that the model underestimates the inter-annual variability of monsoon to a large extent. This is a common drawback of regression models. For the current scenario, the wet period precipitation was found to be slightly increased and dry period precipitation is heavily decreased.

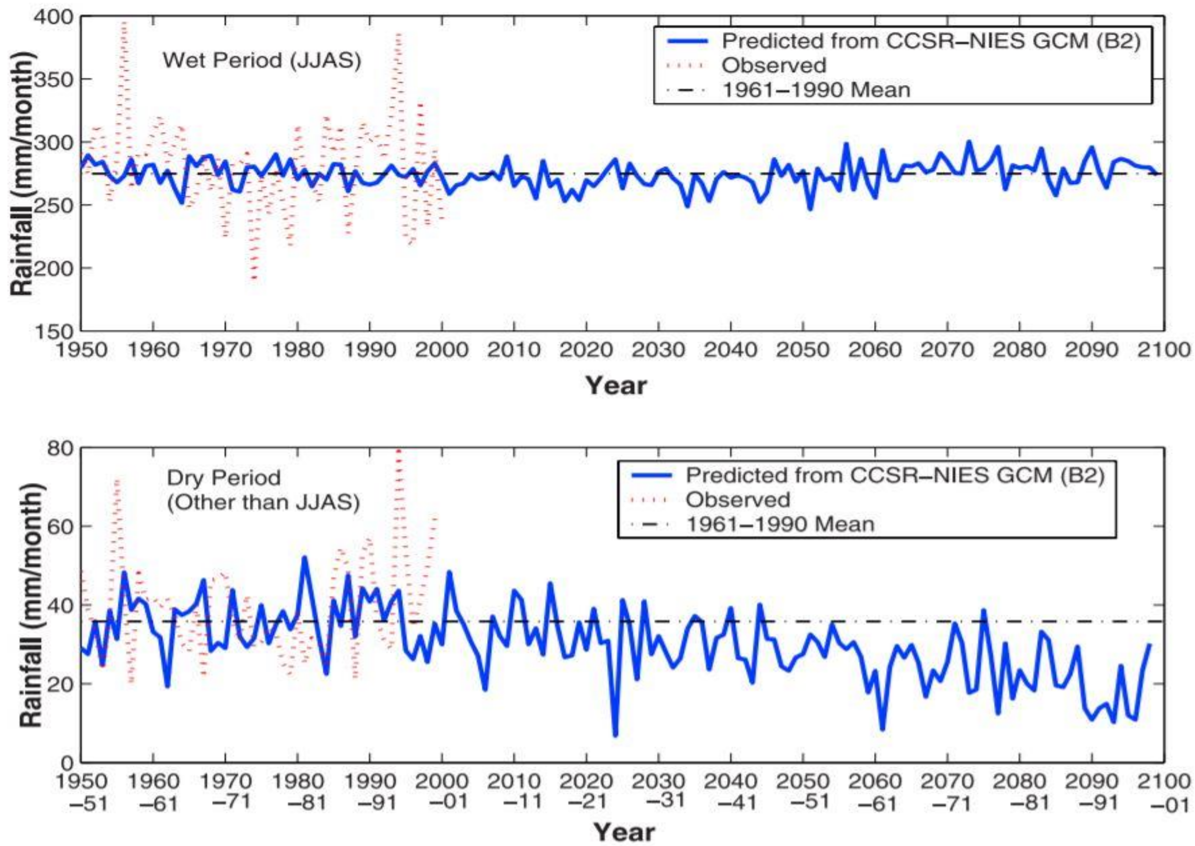

Figure 2: Rainfall for wet and dry periods with CCSR/NIES-B2 projection (Source: Ghosh & Mujumdar, 2007)

REFERENCES

Chen, J., Zhang, X. J., & Brissette, F. P. (2014). Assessing scale effects for statistically downscaling precipitation with GPCC model. *International Journal of Climatology*, *34*(3), 708–727. https://doi.org/10.1002/joc.3717

E Kalnay, M Kanamitsu, R Kistler, W Collins, D Deaven, L Gandin, M Iredell, S Saha, G White, J Woollen, Y Zhu, M Chelliah, W Ebizusaki, W Higgins, J Janowiak, K C Mo, C Ropelewski, J Wang, A Leetmaa, R Reynolds, Roy Jenne, and D. J. (1995). The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American Meteorological Society*. https://journals.ametsoc.org/doi/10.1175/1520-0477%281996%29077%3C0437%3ATNYRP%3E2.0.CO%3B2

Ghosh, S., & Mujumdar, P. P. (2006). Future rainfall scenario over Orissa with GCM projections by statistical downscaling. *Current Science*, *90*(3), 396–404. http://civil.iisc.ernet.in/~pradeep/current_science.pdf

Ghosh, S., & Mujumdar, P. P. (2007). Nonparametric methods for modeling GCM and scenario uncertainty in drought assessment. *Water Resources Research*, *43*(7). https://doi.org/10.1029/2006WR005351

Hadipour, S., Harun, S., Arefnia, A., & Alamgir, M. (2016). Transfer function models for statistical downscaling of monthly precipitation. *Jurnal Teknologi*, *78*(9–4), 55–62. https://doi.org/10.11113/jt.v78.9695

Jeong, D. I., St-Hilaire, A., Ouarda, T. B. M. J., & Gachon, P. (2012). Comparison of transfer functions in statistical downscaling models for daily temperature and precipitation over Canada. *Stochastic Environmental Research and Risk Assessment*, *26*(5), 633–653. https://doi.org/10.1007/s00477-011-0523-3

Mujumdar, P. P., & Nagesh Kumar, D. (2010). Floods in a changing climate: Hydrologic modeling. *Cambridge University Press*. Cambridge University Press. https://doi.org/10.1017/CBO9781139088428

Nash and Sutcliffe. (1970). River Flow forecasting through conceptual models, Part 1 - A discussion of principles. *Journal of Hydrology*, *10*, 282–290. https://doi.org/10.1080/00750770109555783

Trigo, R. M., & Palutikof, J. P. (1999). Simulation of daily temperatures for climate change scenarios over Portugal: A neural network model approach. *Climate Research*, *13*(1), 45–59. https://doi.org/10.3354/cr013045