

Chapter 3

Review of Literature

Abstract This chapter covers the various research works carried out in the field of analyzing climatic variability induced due to climate change scenarios and its impact on agriculture. The review of works is classified into four domains, viz. climate change, downscaling techniques, multiple linear regression, and principal component analysis and regression.

Keywords Review · Climate change · Downscaling techniques · MLR · PCA · PCR

3.1 Review of Works on Climate Change

The predictions of climate change under various emission scenarios were highly uncertain but it was expected to affect agricultural crop production in the twenty-first century. However, we know very little about future changes in specific cropping systems under climate change in California's Central Valley. Lee and Six (2010) used DAYCENT to simulate changes in yield and fluxes of greenhouse gases under A2 (medium-high) and B1 (low) emission scenarios. In total, 18 climate change predictions for the two scenarios were considered by applying different climate models and downscaling methods. The following crops were selected: alfalfa (hay), cotton, maize, winter wheat, tomato, rice, and sunflower. The simulations suggest that future climate change under the different emission scenarios will lead to a broad range of impacts on crop yields. By 2007, yields under A2 decreased in comparison to the 2009 baseline in the following order: cotton (29 %) > sunflower (27 %) > wheat (17 %) > rice (12 %) > tomato (9 %) > maize (8 %). Yields were between 5 (alfalfa) and 21 % (cotton) lower under A2 compared to B1. Under A2, soil carbon (C) storage tended to decrease under climate change due to a decrease in C inputs to the soil and an increase in soil C decomposition. However, differences in nitrous oxide (N₂O) flux between A2 and B1 were not clear.

Global warming, climate change, and tourism of late have taken the center stage of academic research. A raging debate was on, apart from the popular writings and research articles published on the theme. According to the Intergovernmental Panel

on Climate Change “Warming of the climate system is unequivocal as is now evident from observations of increases in global average air and ocean temperatures, widespread melting of snow and ice since the mid-20th century.” The conceptual paper by Ramasamy and Swamy (2012) is carried out by the contributions of 30 selected papers published in tourism-related journals. The approaches of this manuscript are conceptual and are self-oriented to bring readers an all-encompassed state of the art. The purpose of this study was to identify and understand the extent of research carried out to assess the impact of global warming and climate change on tourism. A three-pronged approach was adopted to collect data. First, a literature search was conducted on Google search engine; second, referred research journals in the areas of global warming, climate change, and tourism are consulted, and third, published reports of national and international scientific organizations and government organizations are examined. The fortunes of tourism industry, given the nature of activity, obviously depend on the magnitude and impact of global warming and climate change. Countries like USA, China, Russia, India, and Australia are largely attributed for the growing pollution and the consequent changes in the global climate. Sector-wise, aviation accounts for 40 %, automobiles 32 %, accommodations 21 %, and others 7 % are found to be the major contributors. Incidentally, all these sectors are related both directly and indirectly to the tourism industry.

Kumar and Sharma (2013) analysed the impact of climate change on agricultural productivity in quantity terms, value of production in monetary terms, and food security in India. The study undertook state-wise analysis based on secondary data for the duration of 1980–2009. Climate variation affects food grain and non-food grain productivity and both these factors along with other socioeconomic and government policy variables affect food security. Food security and poverty are interlinked with each other as cause and effect and vice versa, particularly, for a largely agrarian economy of India. Regression results for models proposed in this study show that for most of the food grain crops, non-food grain crops in quantity produced per unit of land, and in terms of value of production, climate variation causes negative impact. The adverse impact of climate change on the value of agricultural production and food grains indicates food security threat to small and marginal farming households. The state-wise food security index was also generated in this study; and econometric model estimation reveals that the food security index itself also gets adversely affected due to climatic fluctuations.

3.2 Review of Works on Downscaling Techniques

The impact of global warming on the temperature regime of a single site is explored by Trigo and Palutikof (1999) with reference to Coimbra in Portugal. The basis of the analysis is information taken from a climate change simulation performed with a state-of-the-art general circulation model (the Hadley Centre model). First, it is shown that the model is unable to reproduce accurately the statistics of daily

maximum and minimum temperature at the site. Second, using a reanalysis data set, downscaling models are developed to predict site temperature from large-scale free atmosphere variables derived from the sea-level pressure and 500 hPa geopotential height fields. In particular, the relative performances of linear models and nonlinear artificial neural networks are compared using a set of rigorous validation techniques. It is shown that even a simple configuration of a two-layer nonlinear neural network significantly improves the performance of a linear model. Finally, the nonlinear neural network model is initialized with general circulation model output to construct scenarios of daily temperature at the present day (1970–1979) and for the future decade (2090–2099). These scenarios are analyzed with special attention to the comparison of the frequencies of heat waves (days with maximum temperature greater than 35 °C) and cold spells (days with minimum temperature below 5 °C).

Schoof and Pryor (2001) had carried out study with comparison of two statistical downscaling methods for daily maximum and minimum surface air temperature, total daily precipitation, and total monthly precipitation at Indianapolis, in USA. The analysis is conducted for two seasons, the growing season and the nongrowing season, defined based on variability of surface air temperature. The predictors used in the downscaling are indices of the synoptic scale circulation derived from rotated principal components analysis (PCA) and cluster analysis of variables extracted from an 18-year record from seven rawinsonde stations in the Midwest region of the United States. PCA yielded seven significant components for the growing season and five significant components for the nongrowing season. These PCs explained 86 and 83 % of the original rawinsonde data for the growing and nongrowing seasons, respectively. Cluster analysis of the PC scores using the average linkage method resulted in 8 growing season synoptic types and 12 non-growing synoptic types. The downscaling of temperature and precipitation are conducted using PC scores and cluster frequencies in regression models and artificial neural networks (ANNs). Regression models and ANNs yielded similar results, but the data for each regression model violated at least one of the assumptions of regression analysis. As expected, the accuracy of the downscaling models for temperature was superior to that for precipitation. The accuracy of all temperature models was improved by adding an autoregressive term, which also changed the relative importance of the dominant anomaly patterns as manifest in the PC scores. Application of the transfer functions to model daily maximum and minimum temperature data from an independent time series resulted in correlation coefficients of 0.34–0.89. In accord with previous studies, the precipitation models exhibited lesser predictive capabilities. The correlation coefficient for predicted versus observed daily precipitation totals was less than 0.5 for both seasons, while that for monthly total precipitation was below 0.65. The downscaling techniques are discussed in terms of model performance, comparison of techniques, and possible model improvements.

Researchers are aware of certain types of problems that arise when modeling interconnections between general circulation and regional processes, such as prediction of regional, local-scale climate variables from large-scale processes, e.g., by means of general circulation model (GCM) outputs. A statistical downscaling approach to monthly total precipitation over Turkey, which is an integral part of

system identification for analysis of local-scale climate variables, is investigated by Tatli et al. (2004). Based on perfect prognosis, a new computationally effective working method is introduced by the proper predictors selected from the National Centers for Environmental Prediction–National Center for Atmospheric Research reanalysis data sets, which are simulated as perfectly as possible by GCMs during the period of 1961–1998. The Sampson correlation ratio is used to determine the relationships between the monthly total precipitation series and the set of large-scale processes (namely 500 hPa geopotential heights, 700 hPa geopotential heights, sea-level pressures, 500 hPa vertical pressure velocities, and 500–1000 hPa geopotential thicknesses). In the study, statistical preprocessing is implemented by independent component analysis rather than principal component analysis or principal factor analysis. The proposed downscaling method originates from a recurrent neural network model of Jordan that uses not only large-scale predictors, but also the previous states of the relevant local-scale variables. Finally, some possible improvements and suggestions for further study are mentioned.

Monthly mean temperatures at 562 stations in China are estimated using a statistical downscaling technique. The technique used by Li-Jun (2009) was multiple linear regressions (MLRs) of principal components (PCs). A stepwise screening procedure is used for selecting the skilful PCs as predictors used in the regression equation. The predictors include temperature at 850 hPa (T), the combination of sea-level pressure and temperature at 850 hPa (P + T), and the combination of geopotential height and temperature at 850 hPa (H + T). The downscaling procedure is tested with the three predictors over three predictor domains. The optimum statistical model is obtained for each station and month by finding the predictor and predictor domain corresponding to the highest correlation. Finally, the optimum statistical downscaling models are applied to the Hadley Centre Coupled Model, version 3 (HadCM3) outputs under the Special Report on Emission Scenarios (SRES) A2 and B2 scenarios to construct local future temperature change scenarios for each station and month. The results show that (1) statistical downscaling produces less warming than the HadCM3 output itself; (2) the downscaled annual cycles of temperature differ from the HadCM3 output, but are similar to the observation; (3) the downscaled temperature scenarios show more warming in the north than in the south; and (4) the downscaled temperature scenarios vary with emission scenarios, and the A2 scenario produces more warming than the B2, especially in the north of China.

Ojha et al. (2010) studied downscaling models using a linear multiple regression (LMR) and artificial neural networks (ANNs) for obtaining projections of mean monthly precipitation to lake-basin scale in an arid region in India. The effectiveness of these techniques was demonstrated through application to downscale the predict and (precipitation) for the Pichola lake region in Rajasthan state in India, which was considered to be a climatically sensitive region. The predictor variables are extracted from (1) the National Centers for Environmental Prediction (NCEP) reanalysis dataset for the period 1948–2000, and (2) the simulations from the third-generation Canadian Coupled Global Climate Model (CGCM3) for emission scenarios A1B, A2, B1, and COMMIT for the period 2001–2100. The scatter plots

and cross correlations were used for verifying the reliability of the simulation of the predictor variables by the CGCM3. The performance of the linear multiple regression and ANN models was evaluated based on several statistical performance indicators. The ANN-based models are found to be superior to LMR-based models and subsequently, the ANN-based model was applied to obtain future climate projections of the predict and (i.e., precipitation). The precipitation is projected to increase in future for A2 and A1B scenarios, whereas it is least for B1 and COMMIT scenarios using predictors. In the COMMIT scenario, the emissions are held the same as in the year 2000.

An extensive statistical ‘downscaling’ study is done to relate large-scale climate information from a general circulation model (GCM) to local-scale river flows in SW France for 51 gaging stations ranging from nival (snow-dominated) to pluvial (rainfall-dominated) river systems. Tisseuil et al. (2010) studied to select the appropriate statistical method at a given spatial and temporal scale to downscale hydrology for future climate change impact assessment of hydrological resources. The four proposed statistical downscaling models use large-scale predictors (derived from climate model outputs or reanalysis data) that characterize precipitation and evaporation processes in the hydrological cycle to estimate summary flow statistics. The four statistical models used are generalized linear (GLM) and additive (GAM) models, aggregated boosted trees (ABT), and multi-layer perceptron neural networks (ANN). These four models were each applied at two different spatial scales, namely at that of a single flow gaging station (local downscaling) and that of a group of flow gaging stations having the same hydrological behavior (regional downscaling). For each statistical model and each spatial resolution, three temporal resolutions were considered, namely the daily mean flows, the summary statistics of fortnightly flows, and a daily “integrated approach.” The results show that flow sensitivity to atmospheric factors is significantly different between nival and pluvial hydrological systems which are mainly influenced, respectively, by shortwave solar radiations and atmospheric temperature. The non-linear models (i.e., GAM, ABT, and ANN) performed better than the linear GLM when simulating fortnightly flow percentiles. The aggregated boosted trees method showed higher and less variable R^2 values to downscale the hydrological variability in both nival and pluvial regimes. Based on GCM cnrm-cm3 and scenarios A2 and A1B, future relative changes of fortnightly median flows were projected based on the regional downscaling approach. The results suggest a global decrease of flow in both pluvial and nival regimes, especially in spring, summer, and autumn, whatever the considered scenario. The discussion considers the performance of each statistical method for downscaling flow at different spatial and temporal scales as well as the relationship between atmospheric processes and flow variability.

Aksornsingchai and Srinilta (2011) studied three statistical downscaling methods to predict temperature and rainfall at 45 weather stations in Thailand. Methods under consideration are multiple linear regressions (MLR), support vector machine with polynomial kernel (SVM-POL), and support vector machine with radial basis function kernel (SVM-RBF). Large-scale data are from Geophysical Fluid Dynamics Laboratory (GFDL). Five predictor variables are chosen: (1) temperature,

(2) pressure, (3) precipitation, (4) evaporator, and (5) net short wave. Accuracy is assessed by tenfold cross-validation in terms of root-mean-squared error (RMSE) and correlation coefficient (R). SVM-RBF is the most accurate model. Prediction accuracy of monthly average rainfall and temperature is satisfying in most part of the country. Lastly, downscaling models can project long-term trends of monthly average rainfall and temperature.

The summer rainfall over the middle-lower reaches of the Yangtze River valley (YRSR) has been estimated by Yan et al. (2011) with a multi-linear regression model using principal atmospheric modes derived from a 500 hPa geopotential height and a 700 hPa zonal vapor flux over the domain of East Asia and the West Pacific. The model was developed using data from 1958 to 1992 and validated with an independent prediction from 1993 to 2008. The independent prediction was efficient in predicting the YRSR with a correlation coefficient of 0.72 and a relative root-mean-square error of 18 %. The downscaling model was applied to two general circulation models (GCMs) of Flexible Global Ocean-Atmosphere-Land System Model (FGOALS) and Geophysical Fluid Dynamics Laboratory coupled climate model version 2.1 (GFDL-CM2.1) to project rainfall for present and future climate under B1 and A1B emission scenarios. The downscaled results provided a closer representation of the observation compared to the raw models in the present climate. In addition, compared to the inconsistent prediction directly from different GCMs, the downscaled results provided a consistent projection for this half-century, which indicated a clear increase in the YRSR. Under the B1 emission scenario, the rainfall could increase by an average of 11.9 % until 2011–2025 and 17.2 % until 2036–2050 from the current state; under the A1B emission scenario, rainfall could increase by an average of 15.5 % until 2011–2025 and 25.3 % until 2036–2050 from the current state. Moreover, the increased rate was faster in the following decade (2011–2025) than the latter of this half-century (2036–2050) under both emissions.

Downscaling is a technique for obtaining local-scale hydrological variables from coarser-scale atmospheric variables that are generated by general circulation models. Mainly there are two downscaling methods, i.e., dynamic downscaling and statistical downscaling. Statistical downscaling offers less computational work compared to dynamic downscaling and it also provides a platform to use ensemble GCM outputs. Devak and Dhanya (2014), in their paper, compared the results generated from two methods, i.e., by support vector machine (SVM) and k-Nearest Neighbor (KNN), which covers some parts of Chhattisgarh, Orissa, Bihar, and Maharashtra state. The above two models are applied at five different locations in Mahanadi Basin. Bias correction by equidistant CDF matching method is also applied to the future projection. Calibration and validation of the model incorporates the result from Canadian global climate model (CanCM4) for historical scenario and future projections are done using predictors from RCP 4.5 scenario. Various performance measures like, normalized mean square error (NMSE) and correlation coefficient is also taken into account. Kolmogorov Smirnov test is also performed for the two models.

3.3 Review of Works on Multiple Linear Regressions

Meteorological data mining is a form of data mining concerned with finding hidden patterns inside largely available meteorological data, so that the information retrieved can be transformed into usable knowledge. Weather is one of the meteorological data that is rich in important knowledge. The most important climatic element which impacts on agricultural sector is rainfall. Thus, rainfall prediction becomes an important issue in agricultural country like India. Dutta and Tahbilder (2014) used data mining technique in forecasting monthly rainfall of Assam. This was carried out using traditional statistical technique—multiple linear regression. The data include six-year period (2007–2012) collected locally from Regional Meteorological Center, Guwahati, Assam, India. The performance of this model is measured in adjusted R-squared. Our experiment results show that the prediction model based on multiple linear regression indicates acceptable accuracy.

Agrarian sector in India is facing rigorous problem to maximize the crop productivity. More than 60 % of the crop still depends on monsoon rainfall. Recent developments in information technology for agriculture field has become an interesting research area to predict the crop yield. The problem of yield prediction is a major problem that remains to be solved based on available data. Data mining techniques are the better choices for this purpose. Different data mining techniques are used and evaluated in agriculture for estimating the future year's crop production. Ramesh and Vardhan (2015) presented a brief analysis of crop yield prediction using multiple linear regression (MLR) technique and density-clustering technique for the selected region, i.e., East Godavari district of Andhra Pradesh in India.

3.4 Review of Works on Principal Component Analysis and Principal Component Regression

Tenderness is the most important factor affecting consumer prediction of eating quality of meat. Park et al. (2001) developed the principal component regression (PCR) models to relate near-infrared (NIR) reflectance spectra of raw meat to Warner–Bratzler (WB) shear force measurement of cooked meat. NIR reflectance spectra with wavelengths from 1100 to 2498 nm were collected on 119 longissimus dorsi meat cuts. The first principal component (or factor) from the absorption spectra $\log(1/R)$ showed that the most significant variance from the spectra of tough and tender meats were due to the absorptions of fat at 1212, 1722, and 2306 nm and water at 1910 nm. The distinctive fat absorption peaks at 1212, 1722, 1760, and 2306 nm were found in the second factor of the second derivative spectra of meat. In addition, the local minima in the second principal component of the second derivative spectra showed the importance of water absorption at 1153 nm and protein absorption at 1240, 1385, and 1690 nm. When the absorption spectra

between 1100 and 2498 nm were used, the coefficient of determination (R^2) of the PCR model to predict WB shear force tenderness was 0.692. The R^2 was 0.612 when the spectra between 1100 and 1350 nm were analysed. When the second derivatives of the spectral data were used, the R^2 of the PCR model to predict WB shear force of the meat was 0.633 for the full spectral range of 1100–2498 nm and 0.616 for the spectral range of 1100–1350 nm.

Webster (2001) studied principal component regression analysis to examine the relative contributions of 11 ranking criteria used to construct the U.S. News & World Report (USNWR) tier rankings of national universities. The main finding of the study was that the actual contributions of the 11 ranking criteria examined difference substantially from the explicit USNWR weighting scheme because of severe and pervasive multicollinearity among the ranking criteria. USNWR assigns the greatest weight to academic reputation. However, generated first principal component eigenvalues of tier rankings indicate that the most significant ranking criterion was the average SAT scores of enrolled students. This result was significant since admission requirements are policy variables that indirectly affect, for example, admission applications, yields, enrollment, retention, tuition-based revenues, and alumni contributions.

Principal component analysis is one of the most widely applied tools in order to summarize common patterns of variation among variables. Several studies have investigated the ability of individual methods, or compared the performance of a number of methods, in determining the number of components describing common variance of simulated data sets. Peres Neto et al. (2005) identified a number of shortcomings related to these studies and conducted an extensive simulation study where they compared a larger number of rules available and developed some new methods. In total, we compare 20 stopping rules and propose a two-step approach that appears to be highly effective. First, a Bartlett's test is used to test the significance of the first principal component, indicating whether or not at least two variables share common variation in the entire data set. If significant, a number of different rules can be applied to estimate the number of non-trivial components to be retained. However, the relative merits of these methods depend on whether data contain strongly correlated or uncorrelated variables. Also estimated the number of non-trivial components for a number of field data sets so that, one can evaluate the applicability of our conclusions based on simulated data.

As a useful alternative to the Cox proportional hazards model, the linear regression survival model assumes a linear relationship between the covariates and a known monotone transformation, for example, logarithm of an event time of interest. Ma (2007), in their article, studied the linear regression survival model with right censored survival data, when high-dimensional microarray measurements are present. Such data may arise in studies investigating the statistical influence of molecular features on survival risk. They proposed using the principal component regression (PCR) technique for model reduction based on the weight least squared Stute estimate. Compared with other model reduction techniques, the PCR approach was relatively insensitive to the number of covariates and hence suitable for high-dimensional microarray data. Component selection based on the

nonparametric bootstrap, and model evaluation using the time-dependent ROC (receiver operating characteristic) technique are investigated. They demonstrated the proposed approach with datasets from two microarray gene expression profiling studies of lymphoma cancers.

The main purpose of this study by Mendes (2011) was to show that how one can use multivariate multiple linear regression analysis (MMLR) based on principal component scores to investigate relations between two data sets (i.e., pre- and postslaughter traits of Ross 308 broiler chickens). Principal component analysis (PCA) was applied to predictor variables to avoid multicollinearity problem. According to results of the PCA, out of 7 principal components, only the first three components (PC1, PC2, and PC3) with eigenvalue greater than 1 were selected (explained 89.45 % of the variation) for MMLR analysis. Then, the first three principal component scores were used as predictor variables in MMLR. The results of MMLR analysis showed that shank width, breast circumference, and body weight had a similar linear effect on predicting the post-slaughter traits ($P = 0.746$). As a result, since the animals had high value of shank width, breast circumference, and body weight, it might be probable that their post-slaughter traits namely heart weight, liver weight, gizzard weight, and hot carcass weight were also expected to be high.

Principal component analysis (PCA) and multiple linear regressions were applied on the surface water quality data by Mustapha and Abdu (2012) with the aim of identifying the pollution sources and their contribution toward water quality variation. Surface water samples were collected from four different sampling points along Jakara River. Fifteen physico-chemical water quality parameters were selected for analysis: dissolved oxygen (DO), biochemical oxygen demand (BOD₅), chemical oxygen demand (COD), suspended solids (SS), pH, conductivity, salinity, temperature, nitrogen in the form of ammonia (NH₃), turbidity, dissolved solids (DS), total solids (TS), nitrates (NO₃), chloride (Cl), and phosphates (PO₄³⁻). PCA was used to investigate the origin of each water quality parameters and yielded five varimax factors with 83.1 % total variance, and in addition PCA identified five latent pollution sources namely: ionic, erosion, domestic, dilution effect, and agricultural run-off. Multiple linear regressions identified the contribution of each variable with significant values.

In recent decades, particulate matter is one of the prevalent pollutants recorded throughout Malaysia. The development of models to predict particulate matter less than and equal 10 μm (PM₁₀) concentration is thus very useful because it can provide early warning to the population and for input into decision regarding abatement measures and air quality management. The aim of the study by Ul-Saufie et al. (2011) was to improve the predictive power of multiple linear regression models using principal components as input for predicting PM₁₀ concentration for the next day. The developed model was compared with multiple linear regression models. Performance indicator such as prediction accuracy (PA), coefficient of determination (R^2), index of agreement (IA), normalized absolute error (NAE) and root-mean-square error (RMSE) were used to measure the accuracy of the models. Results showed that the use of principal component as inputs improved multiple

linear regression models prediction by reducing their complexity and eliminating data collinearity.

Kelechi (2012), in their paper, used the regression analysis and principal component analysis (PCA) to examine the possibility of using few explanatory variables to explain the variation in the dependent variable. It applied regression analysis and principal component analysis (PCA) to assess the yield of turmeric, from National Root Crop Research Institute Umudike in Abia State, Nigeria. This was done by estimating the coefficients of the explanatory variables in the analysis. The explanatory variables involved in this analysis show a multiple relationship between them and the dependent variable. A correlation table was obtained from which the characteristic roots were extracted. Also, the orthonormal basis was used to establish the linearly independent relationships of the variables. The regression analysis shows in details the constant and the coefficients of the three explanatory variables. On the other hand, the principal component analysis estimates the first principal component and second principal component, and both components accounted for 71.4 % of the total variation. The regression analysis and principal component analysis (PCA) yielded good estimates, which lead to the structural coefficient of the regression model. The study shows that regression analysis and principal component analysis (PCA) use few explanatory variables to explain variations in a dependent variable and are therefore efficient tools for assessing turmeric yield depending on the set objective. But that PCA is more efficient, since it uses fewer variables to achieve the same result.

Accurate forecast of water demand is very crucial in developing a water resource management strategy to check the balance of future water supply and demand to ensure proper water supplies to the people. In order to forecast water demand, different models have been adopted in the literature. Among these the multiple regression analysis was quite popular for long term water demand forecasting. In spite of their evident success in modeling water demands, it can face difficulties in the case of multicollinearity, which implies highly correlated variables. Since water demand depends on many factors such as population, household size, rainfall, temperature, age of population, education, water price, and policy, a multicollinearity problem may arise during the development of a multiple regression model which may lead to the incorrect estimation of future water demand. To avoid multicollinearity problem, principal component regression analysis has been used in several environmental studies which demonstrated its ability to eliminate the multicollinearity problem and to produce better model results. However, application of principal component regression in water demand forecasting is limited. In their study, Haque et al. (2013) developed principal component regression model by combining multiple linear regression and principal component analysis to forecast future water demand in the Blue Mountains Water Supply systems in New South Wales, Australia. In addition, performances of the developed principal component regression model were compared with multiple linear regression models by adopting several model evaluation statistics such as relative error, bias, Nash-Sutcliffe efficiency, and accuracy factor. It was found that the developed principal component regression model was able to predict future water demand with a higher

degree of accuracy with an average relative error, bias, Nash-Sutcliffe efficiency, and accuracy factor values of 3.4, 2.92, 0.44, and 1.04 %, respectively. Moreover, it was found that the principal component regression model performed better than the multiple linear regression models and could be used to eliminate the multicollinearity problem. The method presented in their paper can be adapted to other cities in Australia and the world.

Application of principal component analysis in developing statistical models for forecasting crop yield has been demonstrated. The time series data on wheat yield and weekly weather variables, viz., minimum and maximum temperature, relative humidity, wind- velocity and sunshine hours pertaining to the period 1990–2010 in Faizabad district of Uttar Pradesh have been used in this study. Weather indices have been constructed using weekly data on weather variables. Yadav et al. (2014) developed four models using principal component analysis as regressor variables including time trend and wheat yield as regressand. The model 1 and 3 have been found to be most appropriate on the basis of R^2_{adj} , percent deviation of forecast, RMSE (%), and PSE for the forecast of wheat yield 2 months before the harvest of the crop.