Columbia Water Center
EARTH INSTITUTE | COLUMBIA UNIVERSITY

PEPSICO

Columbia Water Center White Paper

Toward Hedging Climate Risk in Corporate Value Chains

April 2013

Naresh Devineni
Shama Perveen
Upmanu Lall

Summary

Global climate risks that manifest as droughts and floods are an ever present concern. The associated regional variability in water supply poses risks to large multinational companies, especially those who rely heavily on agricultural commodities. The agricultural supply chains of such corporations are often managed through contract farming. Corporations are usually better placed than farmers to buffer financial risks and facilitate innovation on managing climate induced production risk through prognostic approaches and insurance. Examples of proactive corporate approaches include an analysis of what should be planted; where, when and how, considering productivity, market conditions, labor and water constraints/risks at local to regional levels. ***A scientific analysis facilitates the success of such risk hedging mechanisms in value chains***.

PepsiCo is at the forefront of water risk identification and strategic management of its internal operations and the supply chain based on the prognostic information. In collaboration with PepsiCo, the Columbia Water Center developed a prototype corporate water risk and sustainability framework for quantifying and analyzing climate induced water risks. The climate risk tool is based on (a) developing specific indicators for assessment of climate induced water risk as aggregated seasonal water deficits; (b) investigating the sources of predictability for these indicators; and (c) developing statistically verifiable models for issuing season ahead probabilistic forecasts for potential water deficits that imply regional production shortfalls. Although work to date has focused on water-related impacts, weather extremes will be included in forthcoming work. Temperature extremes are of particular interest.

The above developments are part of a joint effort to build a global capacity to forecast near-term climate patterns and water availability across PepsiCo's agricultural sourcing areas, to positively impact PepsiCo's revenue and goals of sustainable agriculture, contributions to global food security, and the livelihood of their core production partners, the farmers.

Contract Farming and PepsiCo

PepsiCo pioneered contract farming in India in 1989. Currently, the company engages directly with farmers across the country to grow and supply a variety of crops to support their beverage, snack food and exports business. Specifically, the potato contract farming structure is sustained through a program that provides the farmers with the best quality potato seeds, agronomy advice, training, help with capital investments like drip irrigation and mitigation of market price risk through contracts, in order to supply the Frito Lay foods division's three state of the art manufacturing plants in Punjab, West Bengal and Maharashtra. These improved practices allow the farmers to dramatically improve productivity and their income. Through contracts and other relationships, more than 11,000 farmers across the states of Punjab, Uttar Pradesh, Karnataka, Bihar, West Bengal, Gujarat and Maharashtra supply PepsiCo with world-class chip grade potatoes (source: PepsiCo, India http://pepsicoindia.co.in).

This setting provided an interesting opportunity to explore whether there was the possibility to provide advance predictions of significant rainfall deficits or other extreme weather events that could adversely impact either potato production or underground water reserves and lead to higher energy costs for pumping groundwater for irrigation to maintain yield. Since the farmers are under contract to supply potatoes at a price point that has a limited spread, it is unclear whether a farmer faced with a significant water shortage would invest in expensive groundwater pumping to sustain production. On the other hand, if such a situation could be anticipated, the contracting corporation could take measures to insure such risks for the contracted farmers, diversify their supply chain to reduce their risk, or take other precautionary steps, such as helping with the installation or operation of additional irrigation capacity or other field-based solutions.

During 2010-2012, the methodology for assessing potential water risks at the district level or higher resolution in India, considering all water demands and supply within that region was developed. **A pilot experiment was initiated in one district using one crop, in the state of Maharashtra, to see if an advance forecast during the 2012 monsoon season could be developed. This white paper summarizes the results of this pilot experiment.**

The 2012 PepsiCo-CWC Pilot

The Satara district in Maharashtra is one of the primary regions for sourcing potatoes during Kharif, the peak rainfall season (June - September), in India. Satara supplies the majority of the potatoes processed by the Frito Lay's manufacturing plant in Pune, Maharashtra. The average annual rainfall in this arid to semi-arid region is around 350mm with high inter-annual variability. The region has experienced four droughts (seasonal rainfall below long term average) since 2001. The ability to predict such droughts, with a reasonable probability at lead times of a season to six months, could suggest ways to adapt existing operations to the anticipated conditions and minimize the impacts of droughts on the companies supply chain.

Climate Induced Water Stress Indicator

A modified version of the water stress index introduced by Devineni et al. (2013) was developed for Satara. The index is derived by accumulating differences in supply (rainfall) and demand (crop water requirement measured through regional reference crop evapo-transpiration) over time to assess the maximum cumulative deficit that is likely to occur. This cumulative deficit index is a primary determinant of the water stress faced by the crop and hence of the dependence of the crop yield on water availability.

The *Seasonal Crop Stress Indicator* (*SCSI*) developed here computes the maximum cumulative deficit over a season between the daily water requirement for optimal crop growth and the daily rainfall. The Index improves on metrics derived using monthly or seasonal rainfall and crop demand because it focuses on the rainfall distribution within the season relative to the crop water demand. Therefore, the *SCSI* accounts for the timing of planting, different stages of crop growth, and the timing and distribution of rainfall in the season. Thus, it is able to discriminate between two monsoon seasons which have the same total rainfall, but differ in that one may have rainfall distributed uniformly over the season through modest rainfall events, while the other may have a few intense rain events separated by long dry periods. The latter gives rise to a much higher SCSI. The computation of SCSI is illustrated in Figure 1.

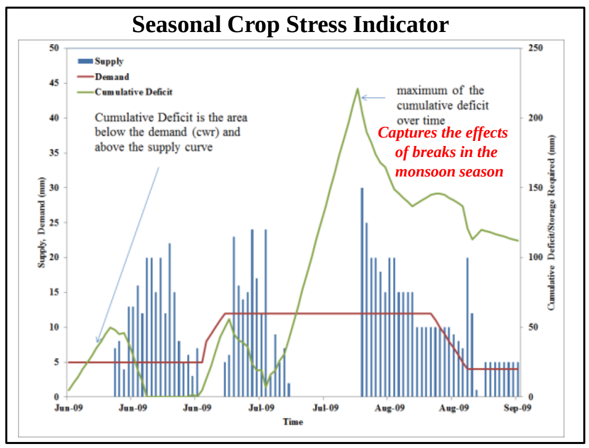


Figure 1: Illustration of *SCSI* computation. Daily rainfall (blue bars), crop water requirement (red line) and the cumulative deficit (green line) over the season.

The annual time series of the *SCSI* computed for the Kharif season in Satara (based on a 112 year daily rainfall data Rajeevan et al. (2006) is presented in Figure 2. We have standardized the SCSI values as the percentage difference each year from the 112-year average of *SCSI*. The long term average *SCSI* for growing potatoes in Satara is 250 mm. This is equivalent to approximately 271,500 gallons of water used for irrigating a one-acre farm of potatoes on average throughout the season. The percent differences in Figure 2 refer to percentages of this number, i.e. a 10% increase in SCSI indicates an additional requirement of 27,150 gallons.

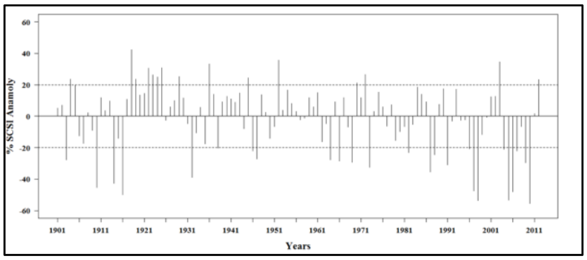


Figure 2: 112 year time variations of *SCSI* expressed as percent anomaly from long-term average *SCSI*.

From Figure 2 it is clear that (a) Satara experiences recurrent droughts with intermediate wet periods and (b) there is year-to-year persistence in the incidence of these droughts. Such variations and epochal changes are typically modulated through large scale global climate patterns. Investigating the relations between monsoon rainfall/deficit and the large scale climate tele-connections could enable the development of models that can be used to understand and hence predict the variability in the SCSI in the region.

The Climate Precursors

India has an extensive history of developing long range predictions of monsoon rainfall that are based on identifying various regional to large scale climate predictors diagnosed from historical data collected over the years (Walker 1924; Thapliyal 1987). A variety of seasonal forecasts of the All India monsoon are documented and available for reference (Gadgil et al. 2007; Kumar et al. 1995). It is well established that inter-annual climate modes such as the El Nino-Southern Oscillation (ENSO) associated with anomalous Sea Surface Temperature (SST) conditions in the tropical Pacific Ocean influence the inter-annual variability of monsoonal rainfall (Parthasarathy and Pant 1985; Shukla and Paolino 1983).

The El Nino (La Nina) phase of ENSO is associated with an increased probability of experiencing drought (good rainfall) over the Indian sub-continent (Sikka 1980; Parthasarathy and Panth, 1984; Rasmusson and Carpenter, 1983). Various indices of the ENSO phenomena have been developed. Some of these based on averages of regional SST are known as the Nino12, Nino3, Nino4 and Nino34 indices (Figure 3). Several other regional to global predictors of Indian monsoon have been documented by Kumar et al. (1995). Other climate phenomena that are related to the Indian monsoon have also been discussed by them.

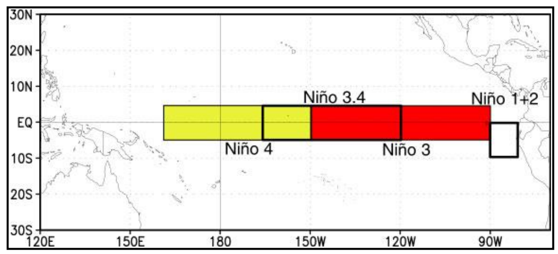


Figure 3: Regions in the tropical Pacific over which various ENSO indices are developed.

One goal of our work was to develop a simple statistical model for predicting *SCSI* for growing potatoes in Satara, as opposed to developing a statistical model for the rainfall in Satara or the All India monsoon. The generalized climate forecast models available are not specific enough for this task. Consequently, the first objective was to identify appropriate climate predictors within a season or more before the monsoon starts in June. After initial statistical investigations and a review of the key literature associated with Indian monsoon prediction, the following predictors were identified:

1. The average of the NINO3 index from the previous winter (December to February)
2. The change in the NINO12 index from December to March, as a directional

indicator. (as suggested in Parthasarathy et al. 1988)

3. Concurrent season (June-Sept.) SSTs over the eastern Indian Ocean (see Figure 4). This corresponds to enhanced (suppressed) atmospheric convection during the anomalous warming (cooling) of the Indian Ocean waters, which serves as a continuous source of moisture as the monsoon season develops.

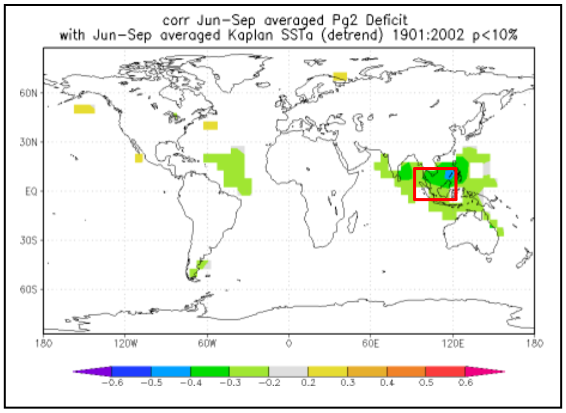


Figure 4: Spearman's Rank Correlation between SCSI in Satara and SST during the same season. SST region in the Indian Ocean (red box) that influences the *SCSI* has a statistically significant correlation at the 95% level.

Statistical Forecast Model

Season ahead *SCSI* forecasts based on the k-nearest neighbors (*k-nn*) semi-parametric resampling approach were developed. This is a data driven approach due to Souza and Lall (2003) that develops a conditional probability distribution of *SCSI* given the predictors by first identifying the k historical climate conditions that are most similar to the current values of the climate predictors and then randomly drawing the *SCSI* values in the historical data that correspond to these k neighbors. The neighbors are weighted so that the closer or more similar neighbors are chosen more often than those further away. For the applications to Satara, 50 neighbors were considered and 1000 samples were drawn from them at each time to develop a forecast. A regression based approach is used first to weight the predictors based on their "importance" in explaining the future *SCSI*.

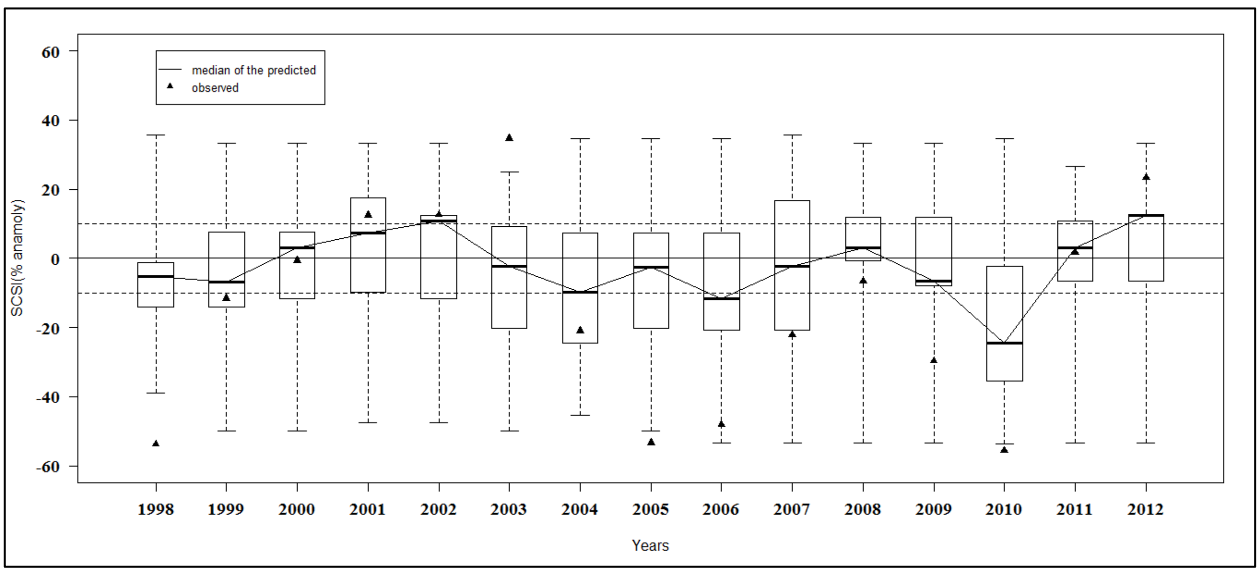
The forecast procedure is tested using the leave-one-out cross validation method. Each historical observation is omitted in turn, and the model is developed using the remaining 111 years of data. A prediction of the observation that was not kept in the model building set is then made and compared with the actual outcome for that year. For the 112-year data set this leads to 111 prediction comparisons.

Results from a variant of this approach are presented in Figure 5. The cross-validated forecasts for the most recent years, 1998 – 2012 are illustrated in Figure 5. Here, the *SCSI* for the 1998 Kharif season is predicted using the model developed based on data from 1901 – 1997. Similarly, the deficit for 1999 is predicted based on the model that is developed using the data from 1901 – 1998. Thus, we always use only the historical data and update the model each year with the information of the previous year, much as a normal user of the forecast system would have to do. Hence, as we move from year to year, we update the model observations and predict the future state. It is important to note that while the Nino3 and Nino12 predictors are from the previous season, the predictors from the Eastern Indian Ocean are based on concurrent season (JJAS SSTa). Hence in the forecasting scheme, we used the JJAS forecasted SST state issued in May from ECHAM4.5 operational forecasting center (available from IRI data library (http://iridl.ldeo.columbia.edu/SOURCES/.IRI/.FD/.ECHAM4p5/.Forecast/.ca_sst/); Li and Goddard, 2005; van den Dool 2007). Skillful forecasts for the tropical SSTs based on coupled ocean – atmospheric general circulation models have been in operation from various climate centers since 1998. Hence, in our forecasting scheme for Satara, we develop and demonstrate the forecasts of *SCSI* from 1998 – present.

The box and whiskers shown in Figure 5 for each year illustrate the range of possible values of *SCSI* for that year. The box shows the range of forecast values that are likely to occur at least 25% of the time and at most 75% of the time. The solid line in the middle of the box gives the median, i.e. the value for which 50% of the forecasts are above or below. So the spread between the 25% and the 75% indicates the uncertainty in the forecast, and the median indicates a middle value that may be useful to compare with the subsequent *SCSI* observation. Note that the directional indication of the forecast is generally quite accurate, while the uncertainty varies from year to year. Knowing the uncertainty is useful since years in which the uncertainty in the forecast is low and there is a strong directional indication for SCSI may lead to different risk management actions than years in which the forecast has a strong directional change but is also marked by high uncertainty.

To provide insight as to how this could be approached, we consider arbitrary thresholds of departures in SCSI as deviations from its long-term average, and use the forecast probability distribution for each year to assess the probability of a positive or negative departure of that magnitude. In reality a decision maker could prescribe their own action threshold, evaluate the production and hence monetary consequences of the forecast relative to the uncertainty and the threshold prescribed, as part of the decision process.

Figure 5: Cross-Validated Performance of the model in predicting *SCSI* during 1998 – 2012 for Satara. The boxplots for each year represent the 25th, 50th and 75th percentiles, in the middle with the whiskers extending to the largest or smallest value generated in 1000 random trials. The triangle plotted each year is the observation for *SCSI* for that year, and the line connecting the medians represents the center of the forecast. Triangles represent observed *SCSI* in that year as percentage departures from the long term average. The three horizontal lines represent average deficit and ± 10% of the average *SCSI* representing above average deficit (or drought) and below average deficit (good rainfall).



This probabilistic evaluation of model performance is illustrated in Table 1. For each year we compute the number of forecasts (out of the 1000) that are greater than the long term average of *SCSI* to obtain the probability of being in drought. The years when the model forecasts a high probability (0.6) of a good rainfall year (low *SCSI*) and the actual rainfall is good corresponding to a low *SCSI*, values are highlighted in blue. Similarly years when the model forecasted a drought (high *SCSI*) with high probability and the actual observation was also a high *SCSI* values are highlighted in orange. It is interesting to note that **8 of the 12 years when the departure in *SCSI* is greater than 10% in either direction, the indicated forecast probability is in the correct direction**. If we increase the departure to 20% (40%) of the long term average this is still 7 out of 9 (3 out of 4) years. Thus, it seems that **extreme departures of *SCSI* may have some predictability**. However, the odds in many of these years are not very far from 50/50, which is what we expect by chance.

If one considers acting only if the probability associated with this forecast of above or below average *SCSI* is at least 60%, then the forecast is in the right direction in 8 of 12 years. The two years that are misclassified have modest departures of the opposite sign. **Raising this to 66% leads to 5 of 5 years being properly classified.** The forecasts consequently appear to be well calibrated.

Table 1: Model forecasts issued as probability of occurrence or non-occurrence of *SCSI* for a year being above or below its long term values. The table also shows the actual observations as percentage departures from the long term average for evaluating the forecasts.

Year	Probability of good rain	Probability of drought	Actual Observed SCSI Anamoly (%)
1998	0.76	0.24	-54
1999	0.56	0.44	-12
2000	0.45	0.55	-1
2001	0.45	0.56	12
2002	0.40	0.60	13
2003	0.63	0.37	35
2004	0.66	0.34	-21
2005	0.51	0.49	-53
2006	0.70	0.30	-48
2007	0.60	0.40	-22
2008	0.44	0.57	-7
2009	0.64	0.36	-30
2010	0.77	0.23	-56
2011	0.38	0.62	2
2012	0.33	0.67	23

Future Direction

Our future work will be focused on developing an integrated regional climate-weather forecast system covering precipitation, temperature, humidity, etc., over the year to benefit the farmers in the context of a specific decision time table for irrigation scheduling along with the pre-season crop choices. These multi-scale risk attributes will include mutually dependent, spatially disaggregated statistics such as total rainfall, average temperature, growing degree days, relative humidity, total number of rainfall days/dry spell length, and cumulative water deficits that inform the potential irrigation water requirements for crops etc. Given that these attributes exhibit mutual dependence across space and time, we will explore common ocean-atmospheric conditions from the observations and the state of the art Global Circulation Models (GCMs) that can be utilized as the predictor variables for the forecasting system. Hierarchical Bayesian methods that can easily handle the high dimensionality of such problems will be used to develop the integrated forecast system. The developed multivariate forecasts will be adapted and disseminated as decision tools for the farmers.

References

Devineni, N., S. Perveen, and U. Lall (2013), Assessing Chronic and Climate Induced Water Risk Through Spatially Distributed Storage Indices: A New Picture of Water Sustainability in India, *in press*, *Water Resources Research*.

Gadgil, S., Rajeevan, M., Francis, P.A., 2007. Monsoon variability: links to major oscillations over the equatorial Pacific and Indian oceans. Curr. Sci. 93, 182–194.

Krishna Kumar, K., Soman, M. K. and Rupa Kumar, K., Seasonal forecasting of Indian summer monsoon rainfall: a review. *Weather*, 1995 **50**, 449.

Li, S. and L. Goddard 2005: Retrospective Forecasts with the ECHAM4.5 AGCM IRI Tech. Report 05-02 December 2005

Parthasarathy, B. and Pant, G.B., 1984. The spatial and temporal relationships between the Indian summer monsoon rainfall and the southern oscillation. *Tellus*, 36A, 269-277.

Parthasarathy, B. and Pant, G.B., 1985. Seasonal relationships between Indian summer monsoon rainfall and the southern oscillation, *J. Climatol.*, 5, 369-378.

Parthasarathy, B., Diaz, H.F. and Eischeid, J.K. 1988: Prediction of all/India summer Monsoon Rainfall with Regional and large-scale parameters. *J. Geophys. Res.*, 93, 5, 5341-5350.

Rajeevan, M., J. Bhate, J. Kale, and B. Lal (2006), High resolution daily gridded rainfall data for the Indian region: Analysis of break and active monsoon spells, *Curr. Sci.*, 91:296–306. http://www.agri.mah.nic.in/

Rasmusson, E.M. and Carpenter, T.H., 1983. The relationship between eastern equatorial Pacific sea surface temperature and rainfall over India and Sri Lanka, Mon.Wea.Rev. 111, 517-528.

Shukla, J. and Paolino, J.A., 1983. The southern oscillation and the long-range forecasting of summer monsoon rainfall over India. *Mon. Wea. Rev.*, 111, 1830-1853.

Sikka, D.R., 1980. Some aspects of the large-scale fluctuations of summer monsoon rainfall over India in relation to fluctuations in the planetary and regional scale circulation parameters. Proc. Ind. Acad. Sci. (Earth. Planet. Sci.), 89, 179-195.

Souza, F. A., and U. Lall (2003), Seasonal to interannual ensemble streamflow forecasts for Ceara, Brazil: Applications of a multivariate, semiparametric algorithm, Water Resour. Res., 39(11), 1307, doi:10.1029/ 2002WR001373.

Thapliyal, V., 1987. Prediction of Indian monsoon variability evaluation and prospects including development of a new model. Climate of China and Global Climate, D. Ye : C Fu; J.Chano, and M. Yoshino, Eds. China Ocean Press, 397-416.

Van den Dool, H. M., 2007: Empirical Methods in Short-Term Climate Prediction. Oxford University Press, 215 pp.

Walker, G.T., 1924. Correlations in seasonal variations of weather, IX, A further study of world weather (World weather II). *Memoirs of India Meteorological Department*, 24, 275-332.