Title of the paper: Creating agricultural drought statistics for developing countries using historic data from satellite images

Last Name, First Name (1st Author): Dr. Ms. Kumari Gurusamy,
Institution, Department: Faculty Department of Economics, BITS Pilani Goa Campus,
Street address: Near NH 17B Bye Pass Road,
City (with postcode): Zuari Nagar, Goa. 403726 INDIA
E-mail: gvkumari@yahoo.com

Abstract: Greenness indices (eg.NDVI) from satellite images have been well demonstrated for their use in detecting drought in regional studies from various parts of the world. A comprehensive drought database at national or continental scale using 20 years of available satellite imagery greenness data can be easily created using well established formulas like Vegetation Condition Index (Kogan 1998) and Percent Carrying Capacity Index (Gurusamy, 2005). The TRMM satellite data also provides excellent rainfall information based on satellite imagery for the past 10 years. The utility of these satellite data and derived indices as proxies of rainfall data has not been verified systematically. In this study we show the utility of freely available satellite imagery in providing good spatial and temporal resolution drought statistics to a developing country (India) where rarity of spatially and temporally continuous rainfall data hinders disaster management decision processes. Drought disaster management involves introduction of “insurance” schemes, which rely on historic rainfall data at meaningful spatial resolution for fixing appropriate premiums based on probability of drought proneness of an area. Currently weather stations are located hundreds of kilometers away from each other and provide discontinuous data, whereas free satellite data are available at 1 km, 8 km, and 16 km spatial resolutions at every 10 day interval for more than 10 years. Insurance claims are also made based on the available current rainfall data, which can vary significantly within 100 square kilometer area which a weather station would normally cover. The availability of free satellite imagery would come as a boon to the farmers who are seated far away from insurance agency trusted weather station data source. Creating drought statistics for developing countries using historic satellite images would help the farmer, insurance agent and the government to make more meaningful decisions for drought disaster management.

1. **Introduction:**

Probability of drought occurrence in a particular spatial unit is a quintessential agricultural statistical measure that aids in agricultural development policy and research. This statistical measure remains as an inherited indigenous knowledge within the farming communities spread in remote rural areas and has not entered the national statistical knowledge bases of many developing countries. The national level policy makers of many developing countries rely on disaster reporting systems to disseminate drought relief mechanisms and cannot pro-actively implement drought management policies due to the lack of spatiotemporally continuous drought probability information. All that is available readily is information of a drought disaster that has had an extreme economic impact. Such information are haphazard in nature and do not render themselves in a spatiotemporally continuous and meaningful formats for rigorous analysis of drought prone-ness.

International development workers who are seated in remote areas are also relying on disaster relief information databases for arriving at drought probability scenarios for their development target countries. A well exposed field worker or a regional scientist can predict drought probabilities of a spatial unit based on the observation of the prevailing traditional cropping systems. Traditional unaltered cropping systems tend to include drought resistant crops to the extent of drought proneness, as an indigenously-evolved insurance scheme or drought management policy. However due to the introduction of mono-cropping mechanisms encouraged through corporate agricultural processing industries, and due to the decay of indigenous knowledge transfer mechanisms in formally educated societies of developing countries, the rural knowledge base of drought and indigenous drought management techniques are rapidly fading away. The developing country agricultural sector is heading towards a scenario where indigenous drought probability knowledge and drought management systems have been eroded and the national policy makers don't have a statistical database to develop effective drought management policies. This could lead to extreme economic vulnerability of a developing country to a natural disaster like drought.

There has been a tendency to rely on weather models that use weather station rainfall data to show deviation from normal rainfall in particular spatial units in order to predict drought proneness. The biggest drawback of this methodology is that it assumes all spatial units to be equally drought prone, since the data is normalized across all spatial units. The problem is not readily visible since these research methods always produce results for individual spatial units for specific time periods. When you integrate them across time and space to calculate drought proneness based on deviation from normality we will statistically arrive at equal probabilities for all spatial units, because of the nature of the assumption. Non-normalized methods have to be employed to calculate drought prone areas, since clearly drought prone spatial units would have "positively skewed" rainfall data structure.

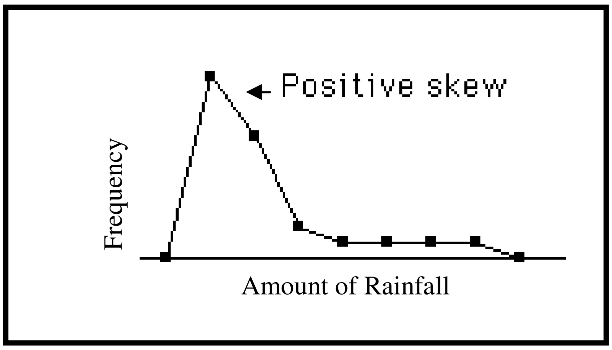


Figure (1) Drought prone areas with positively skewed rainfall

By definition, rainfall data of drought prone areas will have higher frequency of less than optimal rainfall than the optimal and more than optimal rainfall. Drought proneness is analytically equivalent to the extent of positive skew-ness in the rainfall data. Hence normalizing such skewed data and calculating deviation from normal rainfall is clearly not a rigorous method to thematically map relative drought proneness of different spatial units.

Weather stations are located far from each other (about 100km) and hence the best theoretical spatial resolution of rainfall data derived from developing country weather station data has been approximately 100 square kilometer. The rainfall data are not available continuously through long periods of time, since the capacity for infrastructural maintenance is pretty low in developing countries. If we were to develop a workable method using the available length of continuous rainfall data, the time period for which such continuity is existent is not the same for all weather stations, which is a hindrance for a comparative evaluation of drought proneness across the spatial units.

1.1 Background on remote sensing methodologies for drought statistics development:
The researchers who hitherto had access only to poor quality rainfall data available at low spatial and temporal resolutions, in discontinuous formats to delineate spatially drought occurrence in developing countries, have benefited from the remote sensing methods that provide better data that can serve as proxies for rainfall data. Greenness indices such as Normalized Difference Vegetation Index (NDVI) data derived from satellite imagery has been tested by various scientists on the ability to predict drought and serve as a proxy to rainfall data. The preliminary research publications that used these methods were using data across small time periods, in small spatial extents, since satellite imagery data were expensive.

The recent availability of free data on greenness indices such as NDVI data at 8 square kilometer spatial resolution from AVHRR (USA) for 19 years and 1 square kilometer resolution from SPOT (France) for 9 years, both at a very high temporal resolution (every 10 days) has completely changed the potential for drought prone area delineation for developing country researchers. The current study deals with a specific methodology called percent carrying capacity method (PCC) that uses this freely available satellite data to measure drought proneness for developing countries at high spatial and temporal resolution.

Such freely available satellite based greenness indices data have been extensively used by developed country nations and drought observation and prediction data based on NDVI are distributed freely online for the geographic extent of the developed countries like USA[1], Canada and Australia and are updated in real time. The Africa Data Dissemination Service[2] also provides free access to NDVI related data for African countries along with rainfall data from weather stations. Among the developing countries though there is plenty of need and opportunity to develop similar drought prone area maps such activity hasn’t been implemented. This is possibly due to the exorbitant costs of GIS software, the novelty of the technology and remote sensing methodologies. The increasing availability of open source and free GIS and remote sensing tools will help the developing country institutions participate more actively in utilizing satellite based data sources for developing agricultural drought statistics measures.

1.2 Literature Review:
NDVI has been used to detect drought in various parts of the world in the recent decade. The literature relevant to developing countries are only listed here, since the current project concerns about the enhancement of drought statistics data in developing counties using NDVI data. Among the developing countries, in the North East Brazil, the drought due to El Nino Southern Oscillation was studied using NDVI data from 1981-1993(Liu and Juarez, 2001). In South East Asia (Song et. al., 2004), and in Bolivian Alti plano (Washington-Allen et.al, 1998) studies have been conducted using NDVI for delineating drought. In Papua New Guinea the drought of 1997 was studied using

[1] Web address: http://earthobservatory.nasa.gov/Library/MeasuringVegetation/measuring_vegetation_3.html; last accessed August 2007)
[2] Web address: http://earlywarning.usgs.gov/adds/; last accessed August 2007

NDVI and rainfall data (McVicar and Bierwirth, 2001). In Sudan a dissertation (Kassa, 1999) has been written on using NDVI data from 1982-1993 for drought prediction. There are several types of vegetation indices (Jensen, 2005) similar to NDVI that are also being more popularly used in the field of remote sensing. Drought delineation using PCC method has been completed using AVHRR 8km * 8km data from NOAA for the three continents – South America, Africa and Asia as a part of the PhD dissertation work (Gurusamy, 2005). In India (Singh et al, 2003) drought was studied using 11 years of AVHRR NDVI data at 8km * 8km resolution. The current work uses SPOT data at 1.1km * 1.1km resolution data for South East Asia which covers the Indian sub-continent, the focus of the current project.

1.4 Specific objectives of the project:

To create drought prone area maps corresponding to each NDVI image using PCC method

- To delineate drought prone areas (1.1 km * 1.1 km) for every 10 days.
- Alternatively, using the same resource, to predict drought prone decads (three parts of each month) for every square kilometer area.

To create drought probability maps (probability in space and time) by accumulating the drought prone area maps of 9 years

2. Methodology:

The percent carrying capacity (PCC) method estimates the deviation from optimal rainfall, unlike the deviation from the normal rainfall which neutralizes the spatial differences in drought proneness. Optimal rainfall in the current project is defined as a phenomenon that gives rise to maximum greenness for the specific ecosystem of a spatial unit. It has been assumed that if the rainfall was less than optimal, the lack of water availability causes reduction in visible greenness of the ecosystem and if the rainfall was more than optimum, it causes flooding, affecting the greenness reflected by the ecosystem. Hence in this project it has been assumed that the maximum greenness captured by the satellite is an indicator of the operational definition of optimal rainfall. Deviation from optimal rainfall is thus given by the deviation from maximum greenness. The maximum greenness also varies between each spatial unit, and for each temporal unit. It would be irrational to compare the greenness of different ecosystems and the greenness of the same ecosystems in different time periods to predict the occurrence of a drought in a spatial unit in a particular time period. The methodology doesn't compare the maximum greenness of a desert with the maximum greenness of a forest. Spatial identity is respected and hence every ecosystem underlying the spatial unit is dealt with a different criterion. Hence in this methodology we run the analysis to specific spatial units for specific decads separately, maintaining the spatial and temporal identity of drought proneness.

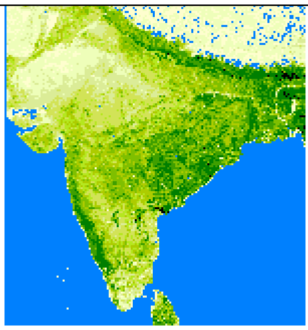
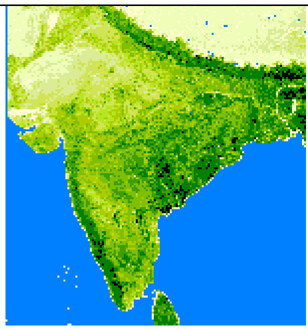
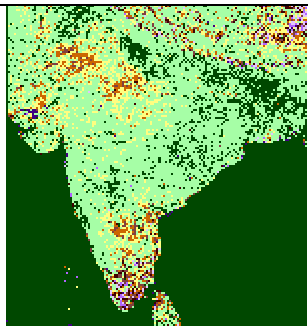
a. Original Data Oct 21 2006	**b. Maximum for Oct 21 2006**	**c. PCC for October 21**	**d. PCC Legend**
			0 - 1 % PCC 2 - 20 % PCC 21 - 30 % PCC 31 - 40 % PCC 41 - 50 % PCC 51 - 60 % PCC 61 - 70 % PCC 71 - 80 % PCC 81 - 100 % PCC

Figure (1) (a) Current Greenness in October 21 2006 data, (b) Maximum Capacity for third decad of October, and (c) the corresponding Percent Carrying Capacity Map of Drought

Formula for the percent carrying capacity method is:

PCC_{st} = (Max Greenness $_{st}$ – Current Greenness $_{st}$) * 100/ Max Greenness $_{st}$

Where s = spatial identity and t = temporal identity

2.1 Steps in the method:

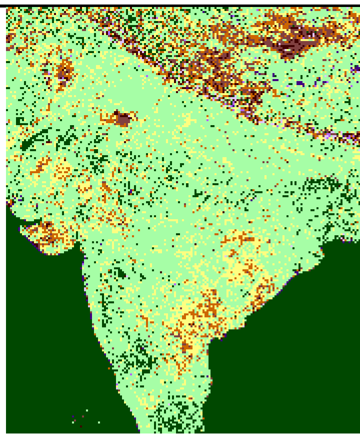
1. Download zip files of NDVI for the region of interest (in this project it was South East Asia since it included India the study focus area) for all the available decads. For this project the source data NDVI files were downloaded from SPOT free vegetation data website. One could create a batch process to automatically download the files using wget commands, since we are downloading 100s of files.
2. Unzip them and store only the NDVI file in HDF format for the analysis. The unzipping process also can be automated using shell scripts for batch processing.
3. Change the NDVI file names to names that identify the space and time it represents. (For example, SEAsia_1998Mar21_NDVI.HDF represents the NDVI file of the third decad of March 1998 for the geographic extent of South East Asia)
4. Convert the HDF format files into a format that is helpful for GIS analysis and raster algebra. In this project FAO's free WINDISP program was used to do the analysis. Hence all the HDF format files were converted into WINDISP format files.

5. Add, edit and verify that the header of all the files contains the geographic co-ordinates and the projection systems recorded properly.
6. Make lists of files corresponding to each decad, to conduct the analysis, maintaining the temporal identity. In WinDISP list files can be used to do raster algebra on specific sets of files. For example, “SEAsia_Mar21.ls” will contain the list of names of the third decad of March files for all 9 years starting from 1998 to 2007.
7. Raster algebra functions can be used to find the pixel among the 9 years, with the maximum NDVI value, for each spatial unit. This calculation is done for each spatial unit (pixel) separately and hence the spatial identity is kept intact. The calculation also considers each decad separately and hence the temporal identity is kept intact. For example if you explore the “SE_Asia_Mar21Max_NDVI” file each pixel in that map will inform you the maximum greenness potential of that particular pixel for the third decad of March derived from the 9 years of Mar 21 NDVI data. The maximum NDVI map of March 21 represents the maximum theoretical greenness South East Asia can attain for the decad of Mar 21, given the historical records. Since one region of South East Asia could have got optimal rainfall in a year in which another region suffered a severe drought, the maximum greenness map doesn’t represent any specific year. The study compares current reality with the maximum potential of the ecosystem (pixel) for a given season (decad).
8. Quality check for the data files have to be done atleast at this point of time, because if the original NDVI data files had stripes in them, the maximum NDVI files would have picked them up.
9. Once the maximum NDVI pixels have been stored in the max NDVI files of the corresponding decads, they can be used to calculate percent carrying capacity of each pixel for each decad in the 9 years. The only cautionary note during this step is that the original data NDVI should be compared with the maximum NDVI for that specific decad. Using raster algebra the PCC formula can be input into the GIS program and the results can be stores as a PCCMap file for each decad corresponding to each original NDVI data file. Macro commands can be written within the GIS software for repeating the steps for all the decads and thus the whole process can be automated.
10. Collect ground truths for verification of calculated data. From a drought disaster relief database collect a few known drought observations with clear indications of time and space. The higher the resolution of the spatial unit in the observed data, more rigorous is the verification process. Hence select drought disaster information at least at admin 2 (state/province) level, if not at admin 3(district/county) level. Information at country level is not very useful. The time stamp on the drought information may indicate the time at which the relief money was sent. It may not correspond to the time period of drought occurrence. There is no specific or uniform time gap between actual occurrence of drought and relief record time stamp to use in the verification process. One could assume that drought occurred during a crop calendar if a drought relief was sent in the end of a cropping season.
11. The PCC calculated maps should display PCC thematically in different colors for varying values of PCC. During manual verification process for a random number of records from the drought relief database, one could figure out the delineating PCC value that causes a drought relief process to begin. In other words, the verification process will point to a range of PCC value in which it is possible to relate to drought relief records. For example, if the PCC value fell above the range of 30% for all the verified drought records, it means when the deviation from maximum NDVI (optimal rainfall) caused more than 30 percent reduction in carrying capacity, drought is recorded. It should also be noted that values about 60 could be due to “clouds.” The most meaningful range in which drought can be observed is possibly between 30 and 60. This range would get more precise and more accurate with rigorous verification methods. In this project 30%-60% loss in carrying capacity is assumed to be drought.

3. **Results and Discussion:**

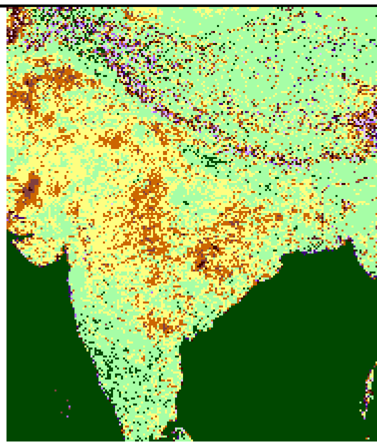
The following two maps are a representative sample of real world drought records[3] compared against the result maps using PCC methods. The colors yellow to brown show drought intensity. The colors violet and blue represent clouds or snow cover.

Table (1) Table of comparison between the drought record 2000-000222 and the corresponding PCC map

Event:	DR Drought	
Number:	2000-000222	
Country:	IND India	
Location:	Gujarat, Rajasthan, Madhya Prasdesh, Andhra Pradesh, Orissa, Maharashtra	
Date (Y-M-D):	2000-4-	
Time:		
Duration:		
Magnitude:	km?	
Information Source:		
Comments:	(Drought)	Figure (2) PCC Image of Feb 21 2000

Drought has been recorded in April. Drought has been observed in February in Orissa, Andhra Pradesh, Gujarat, Rajasthan and some eastern parts of Madhya Pradesh. Not much drought evident in Maharashtra in the month of February.

Table (2) Table of comparison between the drought record 2001-000279 and the corresponding PCC map

Event:	DR Drought	
Number:	2001-000279	
Country:	IND India	
Location:	New Delhi, Rajasthan, Gujarat, Orissa	
Date (Y-M-D):	2001-5-	
Time:		
Duration:		
Magnitude:	km?	
Information Source:		
Comments:	(Drought)	Figure (3) PCC Image of May 11 2001

Drought has been recorded in May possibly during the occurrence of drought, since it covers the capital city (New Delhi) area. Drought has been observed in the PCC Map of May 2001 in Gujarat, Rajasthan, New Delhi and Orissa. The map also points to drought in some areas of Andhra Pradesh and Uttar Pradesh which have not been recorded.

[3] Web address: http://www.glidenumber.net/glide/public/search/search.jsp, last accessed August 2007

4. Concluding Remarks:

Spatiotemporally continuous drought information can be generated using satellite imagery. Agricultural drought statistics in developing countries can be improved significantly with the use of time series of greenness indices derived from satellite image data. The deviation from maximum greenness for a spatial unit, during a particular decad, can be a good proxy for optimal rainfall necessary for the ecosystem in that spatial unit for that respective time of the year. Weather station rainfall data, drought disaster relief data and agricultural census data, give sets of observations and indications of drought which cannot be used for rigorous analysis of drought proneness, but they can be effectively used for verification of the remote sensing methods used to delineate drought.

Percent carrying capacity method is based on the assumption that deviation from greenness is equivalent to deviation from optimal rainfall. But deviation from maximum greenness of a spatial unit could be due to various other factors such as deforestation, urbanization, frost, pests and diseases. Hence in conclusion it could be said that potentially almost all of the drought prone areas are included in the percent carrying capacity method, but all of the areas with reduced percent carrying capacity need not be due to drought. The drought prone area mapping using percent carrying capacity method hence is a precise method, but not an accurate method to point out drought. Accuracy can be significantly improved by removing other causes using sources of data that can filter out other causes, for example temperature data can be used to remove abiotic stress caused in an ecosystem by frost.

Web GIS technologies enable presentation of these maps online with interactive capacities. When the local experts zoom into the data of their area, they would be able to get information relevant to their area which they can cross verify against their local knowledge of the drought phenomenon. It is possible to establish effective feedback systems in web enabled interfaces, which will help harvest critical assessment of the drought delineation work from the local experts. The feedback can then be used for detection of systematic errors in the processes. The drought mapping project hence will evolve over time with continuous feedback between local reality and remotely sensed methods.

References:

1. Gurusamy, Kumari (2005). “Using NDVI-based Measures to Derive Geographic Information on Drought-Prone Areas for Developing Countries,” Dissertation, University at Albany, State University of New York, USA
2. Jensen, J. R. (2005). Introductory Digital Image Processing: A Remote Sensing Perspective. Upper Saddle River, NJ 07458, Prentice Hall.
3. Kassa, A. (1999). Drought Risk Monitoring for the Sudan using NDVI 1982- 1993. Geomatic Engineering. London, University College London: 47.
4. Kogan, F. N. (1998). "Global drought and flood watch from NOAA polar orbitting satellites." Advances in Space Research 21(3): 477-480.
5. Liu, W. T. and R. I. N. Juarez (2001). "ENSO drought onset prediction in northeast Brazil using NDVI." International Journal of Remote Sensing 22 Part 17: 3483-3502.
6. McVicar, T. R. and P. N. Bierwirth (2001). "Rapidly assessing the 1997 drought in Papua New Guinea using composite AVHRR imagery." International Journal of Remote Sensing 22 Part 11: 2109-2128.
7. Singh, R., S. Roy and F.Kogan. (2003). "Vegetation and temperature condition indices from NOAA AVHRR data for drought monitoring over India." International Journal of Remote Sensing 24, no 22: 4393-4402 (10 pages).
8. Washington-Allen, R. A., R. D. Ramsey, B.E.Norton and N.E.West. (1998). "Change detection of the effect of severe drought on subsistence agropastoral communities on the Bolivian Altiplano." International Journal of Remote Sensing 19(7): 1319-1333.
9. Song, X., G. Saito, M.Kodama and H.Sawada. (2004). "Early detection system of drought in East Asia using NDVI from NOAA/AVHRR data." International Journal of Remote Sensing 25, no 16: 3105-3111 (7 pages).