CSI India: Enhancing farmers’ adaptive capacity by developing Climate‐Smart Insurance for weather risk and Innovative Insurance Products for the Rural Sector

# Introduction

The main objective of this project was to help develop novel insurance products for weather‐related risks smart agriculture (CSA).

The following outlines the development of indices for use in insurance products, with an aim to support climate‐smart insurance products in Punjab and Haryana in India as well as in Uruguay. The International Food Policy Research Institute (IFPRI) hopes to use these findings to improve planned services for Indian wheat farmers. This component of the project provides core functionality to a larger portfolio of high potential insurance products.

The following report acts as the final deliverable report for the CSI India statement of work. The statement of work (SOW) outlined the following tasks for the final report:

* Description of indices proposed and methodology to estimate indices
* Methodology to estimate quantiles
* Results of quantile estimates (India)
* Datasets with indices and quintiles estimates (India)

# methodology

## Study Area & Time Period

We focus on the states of Haryana and Punjab in Northwestern India. Satellite data of interest from the Moderate Resolution Imaging Spectroradiometer (MODIS) satellite are available from 7/4/2002 to present. A longer time period, back to 1989, may be generated using the older Advanced Very High Resolution Radiometer (AVHRR). The ability to do so however is limited because AVHRR spatial resolution is 1 km and the data collected[[1]](#footnote-2) is different than that collected from MODIS. Extending the record is possible however “a perfect 1-to-1 correspondence between the two sensors is not expected due to the [collection] different dates and spectral response functions along with the lack of aerosol corrections for the AVHRR data.”

## Vegetation Indicies

MODIS’s Enhanced Vegetation Index (EVI) and Normalized Difference Vegetation Index (NDVI) are both sensitive to the amount of chlorophyll in any given pixel. EVI and similar indexes are commonly used to estimate plant productivity and health in agricultural applications [1–5]. EVI time series of croplands typically follow predictable patterns across the year, with a green-up following planting, and continuing rapidly through growth of the plant, a leveling off as the plant matures and the sudden decline in greenness as the plant is harvested and soils exposed again.

After preprocessing and the identification of agricultural pixels, a series of statistics can be produced for each year to summarize the properties of the growing season of interest. Properties of the year can be used in turn to estimate plant health and crop yields [1,3]. We can see the link between greenness and growth throughout the season in Figure 1. Here we can see some indication of the expected patterns by looking at the mean greenness (black dash line) as well as an example of an unusually productive year (red line).

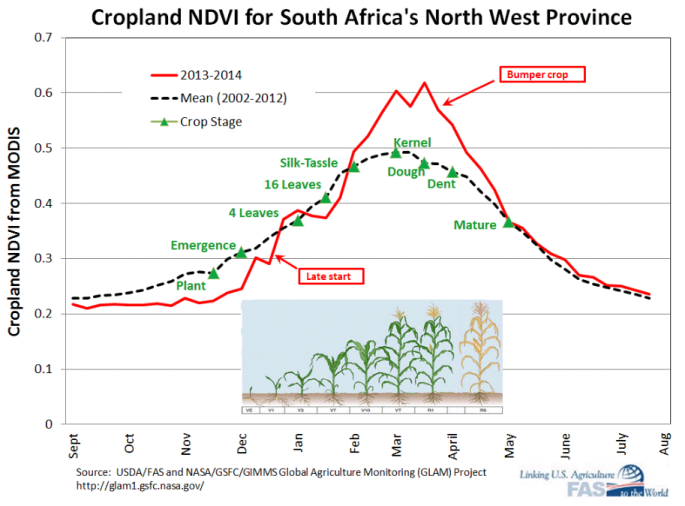


Figure 1 Cropland NDVI for South Africa's North West Province

Mean normalized difference vegetation index (NDVI), another metric for greenness, shown in black dashed line. Green up for the 2013 growing season (an unusually high yielding year) shown in red.

Figure 1 also provides an indication of the required task for this report­- we must extract key metrics of the growing season corresponding to plant health and productivity. These metrics should be able to predict crop yields or identify the timing of damages from a variety of sources (hail, frost, excess rain etc.).

## Vegetation Indices Metrics

A variety of metrics can be calculated for the growing season that correspond to crop health and yields. There are multiple classes of vegetation metrics of interest including: 1) basic summary statistics, 2) integration methods, and 3) comparisons to the norm. Each class of metric could provide important information and should be tested for its ability to describe spatial and temporal patterns in crop yields, or in identifying particular extreme events.

As currently specified, statistics are broken into two summary periods: the calendar year, and the crop growing season. Functions have been carefully developed to handle growing seasons that span more than one year. In these cases, values for a growing season spanning late 2014 and early 2015 are assigned to the 2014 growing season.

### Basic Summary Statistics

This class of function is the most basic but provides key attributes for the period of interest. Metrics include the mean, median, maxima and minima. These functions can be run on individual image pixels, points or polygons defining a region of interest[[2]](#footnote-3). They can also be run for specific time periods. Of particular interest is the period that defines the full extent of a growing season for one crop. Functions defined to summarize a growing season take estimates of planting and harvest dates as inputs. These dates are then adjusted on a pixel by pixel basis to find the actual pixel planting/harvest date for that location. The search is currently limited to 15 days before or after the planting and harvest dates recommended by farmer interviews (currently 11/23 and 4/30 for planting and harvesting respectively). Looking at the red NDVI curve in Figure 1 the planting date would be estimated sometime in mid-November and harvest mid-July.

|  |  |  |
| --- | --- | --- |
| Name | Alt Name | Description |
| A\_mn | EVI\_annual\_mean | Annual mean values of EVI (NDVI) |
| A\_min | EVI\_annual\_min | Annual min values of EVI (NDVI) |
| A\_max | EVI\_annual\_max | Annual maximum values of EVI (NDVI) |
| A\_sd | EVI\_annual\_sd | Annual standard deviation in values of EVI (NDVI) |
| A\_Qnt | EVI\_annual\_5th\_prct | Annual estimated lower 5th percentile values of EVI (NDVI) |
| A\_max\_Qnt | EVI\_annual\_max\_5th\_prct | Annual estimated lower 5th percentile values of the maximum EVI (NDVI) values observed across the historical record |
| G\_mn | EVI\_growing\_mean | Growing season mean values of EVI (NDVI) |
| G\_min | EVI\_growing\_min | Growing season min values of EVI (NDVI) |
| G\_max | EVI\_growing\_max | Growing season maximum values of EVI (NDVI) |
| G\_mx\_dates | EVI\_growing\_max\_date | Date of growing season maximum values of EVI (NDVI)[[3]](#footnote-4) |
| G\_sd | EVI\_growing\_sd | Growing season standard deviation of EVI (NDVI) |
| G\_Qnt | EVI\_growing\_5th\_prct | Annual estimated lower 5th percentile values of EVI (NDVI) |
| G\_mx\_Qnt | EVI\_growing\_max\_5th\_prct | Growing season estimated lower 5th percentile values of the maximum EVI (NDVI) values observed across the historical record |
| T\_G\_Qnt | EVI\_all\_growing\_5th\_prct | The estimated lower 5th percentile values of all growing seasons in the EVI (NDVI) historical record |
| plant\_dates | plant\_dates | Date of estimated planting date based on global minimum within 30 days of predefined plant date |
| harvest\_dates | harvest\_dates | Date of estimated harvest date based on global minimum within 30 days of predefined harvest date |

### Integration Methods

Another set of information can be extracted by integrating across a period of interest. As before, the period of interest may be the whole year or a particular growing season. One metric of interest is the length of the growing season. Unusually short growing seasons are often an indication of a delayed planting date (often due to late onset of a rainy season) or early harvesting (often undertaken to avoid addition damage from weather/pest/disease).

Another temporal integration method is area under the curve (AUC) estimates[[4]](#footnote-5). This method can be applied across a variety of time periods of interest. Of particular interest is the AUC for both the increasing and decreasing portions of the vegetation index curve for a particular growing season. We can see in Figure 3 that the increasing portion of the curve roughly corresponds to the periods of plant establishment to flowering, while the declining portion corresponds to the yield formation and ripening of the grain [7]. We can also see that low values for decreasing or increasing AUC portions of the curve likely correspond to poor plant health (year 2011 in Figure 2). Due to its importance, we develop two methods of estimating the AUC, one using splines and the other using the traditional trapezoidal method. Initial tests indicate that the trapezoidal method is more closely correlated with wheat yields likely due to the fact that day to day variation is not removed through smoothing.

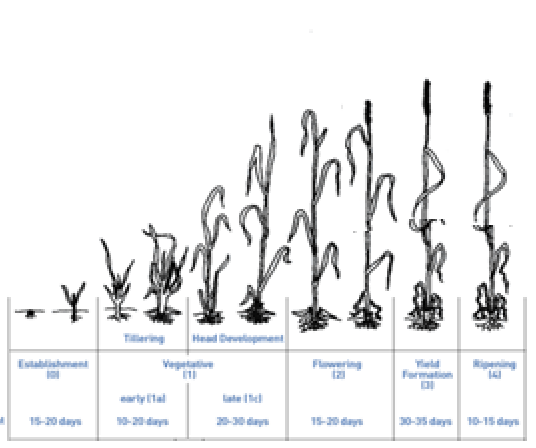
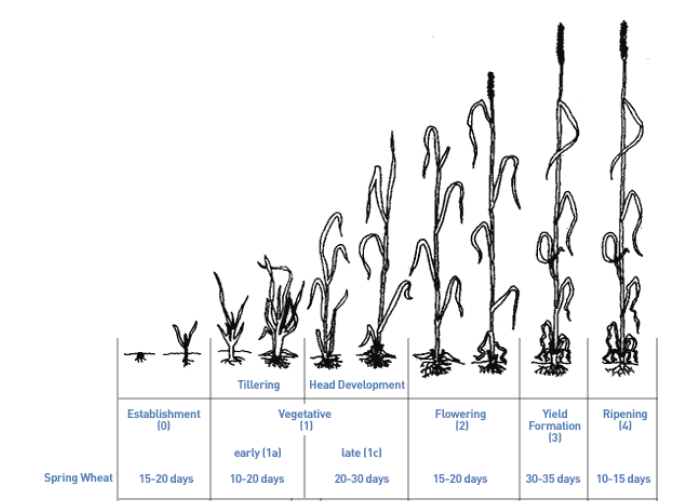
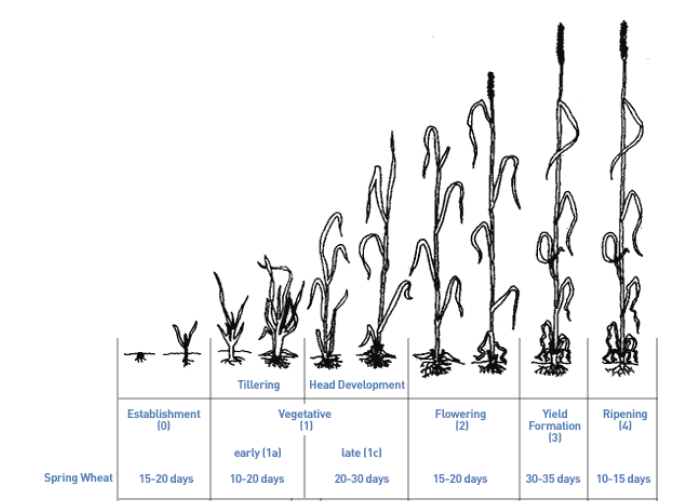
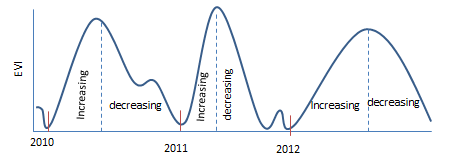


Figure 2: Area under the curve and wheat yields

Upper Panel: Illustration of area under the curve for three growing seasons (2010-2012), with estimated planting dates (vertical red dash), and labels for the ‘increasing’ and ‘decreasing’ portions of the EVI curve. Lower Panel: Illustration of simulated plant growth corresponding to observed EVI curves where 2011 experiences a significant shock, thereby lowering yields.

|  |  |  |
| --- | --- | --- |
| Name 1 | Name 2 | Description |
| A\_AUC | EVI\_annual\_AUC | Annual area under the curve of EVI (NDVI) |
| A\_AUC\_Qnt | EVI\_annual\_AUC\_5th\_prct | Estimated lower 5th percentile of annual area under the curve values for EVI (NDVI) |
| G\_AUC | EVI\_growing\_AUC | Growing season area under the curve of EVI (NDVI) |
| G\_AUC\_Qnt | EVI\_growing\_AUC\_5th\_prct | Estimated lower 5th percentile of growing season area under the curve values for EVI (NDVI) |
| G\_AUC2 | EVI\_growing\_AUC\_v2 | Growing season area under the curve of EVI (NDVI) estimated by splines |
| G\_AUC\_leading | EVI\_growing\_AUC\_leading | Area under the curve of EVI (NDVI) for the ascending part of the curve during the growing season |
| G\_AUC\_trailing | EVI\_growing\_AUC\_trailing | Area under the curve of EVI (NDVI) for the decreasing part of the curve during the growing season |
| season\_length | season\_length | Difference between plant\_dates and harvest\_dates in days | |

### Comparison to Norms

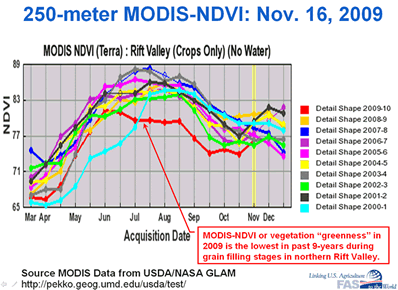
Another approach is to compare a given year to the characteristics of the expected growing season or mean year. Here departures from the mean may have positive or negative implications for crops. Looking at Figure 1 and we can see that the red line corresponds to an unexpectedly good growing season relative to the norm (dashed line). In Figure 4 we can see the effects of significant crop failures in the 2009 growing season in the Rift Valley in East Africa (red line). From these examples we can see that indications of good or bad seasons can be made by comparing any given year’s mean, maxima or AUC to the expected vegetation index curve. The mean season can be calculated by simply calculating the mean vegetation index value for each 8-day period. Comparisons to this norm can then be facilitated by functions described above, such as the difference between the AUC of the mean year and any individual year, or differences between growing season maxima or mean values.

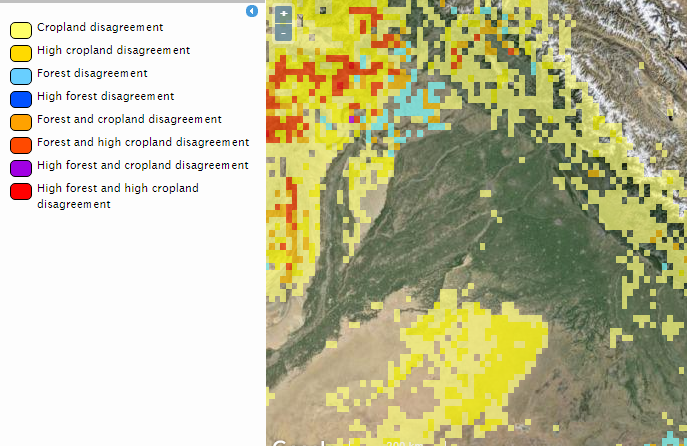
Figure 3: Vegetation index curves for 2000-2009

Representative vegetation index curves for the periods 2000-2009 for Africa’s rift valley, where 2009 (red line) experienced a significant crop failure

|  |  |  |
| --- | --- | --- |
| Name 1 | Name 2 | Description |
| G\_AUC\_diff\_mn | EVI\_growing\_AUC\_diff\_mn | Difference between growing season area under the curve for EVI (NDVI) and the mean area under the curve |
| G\_AUC\_diff\_90th | EVI\_growing\_AUC\_diff\_90th | Difference between growing season area under the curve for EVI (NDVI) and the 90th percentile area under the curve |

## Vegetation Indicies preprocessing

Remotely sensed vegetation indexes were obtained from the 16 day MODIS MOD13Q1 composite product[[5]](#footnote-6) at 250m resolution [8] for both the Aqua and Terra satellites from 07-04-2002 to 02-02-2016. Aqua and Terra products are staggered 16 day periods; therefore, we are able to construct a time series representing approximately every 8-day period. Each individual pixel from a large image records the day of the year of its observation during this 16-day period. However to reduce the computational complexity and avoid the need for data smoothing[[6]](#footnote-7) we assume all pixels in an image reflect the date at which the image is processed. We removed low quality pixels and non-crop land cover types, and allow for the use of smoothing splines to clean the time series6. Preliminary tests at the district level suggest however, that smooth splines may remove variation important to predicting wheat yeilds. Croplands were identified from the MODIS MCD12Q1 land cover product using the International Geosphere–Biosphere Programme (IGBP) classification scheme. The vegetation index signals observed for croplands were treated as a proxy for plots planted with wheat, the crop of primary interest. We have every reason to believe that global land cover products can effectively identify croplands in Haryana and Punjab with their extensive planting of wheat (Figure 4). Here we can see that there is little to no disagreement between land cover products for the area of interest [9]. The only major area of disagreement is along (and predominantly outside of the) the Southern border of Haryana where the Rajasthan Desert begins. Care should be taken however for communities in these semi-arid transition regions.



**Haryana**

**Punjab**

Figure 4: Estimates of cropland disagreement between multiple land cover products

Land cover products often suffer from disagreement between classification schemes. Disagreement between land cover products on the location of croplands can be seen above in yellow (low disagreement) and dark yellow (high disagreement).

## REmote Sensing Limitations

Remote sensing like most data suffers from a variety of limitations. These include the fact that satellites do not directly measure vegetation properties, rainfall, or land cover but instead infer them from reflected portions of the electromagnetic spectrum. Vegetation indexes have a variety of well-documented shortcomings including errors related to clouds, dust, solar illumination angles, and background soil properties. Many of the issues have been address by recent advances in image preprocessing, however many issues remain. Negative consequences can be avoided to rigorous pretesting and verification. Additionally, at a relatively course spatial resolution (250m) individual pixels often include a mix of crops, soil, trees or other natural vegetation. Therefore, vegetation indices must be interpreted carefully depending on whether it is looking at crops, grasslands or forest [13]. EVI (NDVI) values are also affected by changes in land cover. Some shifts in vegetation indices will not reflect changes in plant health but instead be the result of the removal of trees, the construction of a road, or some other shift in land cover to a portion of a pixel.

## Preliminary Yield Estimation Results

It is critical that the indices we develop here correlate with the losses we aim to indemnify. As a preliminary test, we estimate district level wheat yields using a panel regression against 17 metrics of vegetation and district fixed effects for the 2002 to 2016 period (Table 1). We find that we can explain 84% of the variance (65% within) in wheat yields. Although the course spatial scale of the data (district-level) leaves its applicability for farm-level analysis an open ended question, these findings are suggestive that the algorithms developed in this project provide meaningful indicators of plant health and productivity across a growing season.

Table 1: District level random-effects regression of wheat yields per hectare using EVI vegetation indices

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Random-effects regression** | **Dependent:** | **yield\_tn\_ha** |  |  |
| Group variable: i |  |  | Number of groups = | 28 |
|  |  |  | Number of obs = | 269 |
|  |  |  |  |  |
| R-sq: |  |  | Obs per group: |  |
| within = | 0.6545 |  | min = | 5 |
| between = | 1.0000 |  | avg = | 9.6 |
| overall = | 0.8468 |  | max = | 11 |
|  |  |  |  |  |
| **Coef.** |  | **Std. Err.** | **z** | **P>z** |
| District FE | - | - | - | - |
| year\_trend | .0864424 | 0.007399 | 11.68 | 0 |
| plant\_dates | 0.01946 | 0.0357317 | 0.54 | 0.586 |
| a\_mn | 0.000704 | 0.0001799 | 3.91 | 0 |
| a\_min | -0.00019 | 0.0000675 | -2.78 | 0.005 |
| a\_max | -0.00025 | 0.0000973 | -2.57 | 0.01 |
| a\_auc | -2.43E-12 | 2.40E-12 | -1.01 | 0.31 |
| a\_qnt | 6.19E-05 | 0.000199 | 0.31 | 0.756 |
| a\_sd | -0.00016 | 0.0003169 | -0.51 | 0.608 |
| g\_mx\_dates | -0.00045 | 0.0080921 | -0.06 | 0.956 |
| g\_mn | -0.00135 | 0.0015089 | -0.89 | 0.372 |
| g\_min | -4.3E-05 | 0.0001765 | -0.24 | 0.81 |
| g\_mx | -3.8E-05 | 0.0001569 | -0.24 | 0.808 |
| g\_auc | 0.000819 | 0.0004294 | 1.91 | 0.057 |
| g\_qnt | 6.99E-05 | 0.0003637 | 0.19 | 0.848 |
| g\_auc\_leading | -0.00075 | 0.000422 | -1.78 | 0.075 |
| g\_auc\_trailing | -0.00075 | 0.0004211 | -1.78 | 0.075 |
| g\_sd | 0.000388 | 0.0003813 | 1.02 | 0.308 |
| \_cons | -177.867 | 19.54596 | -9.1 | 0 |

# Creating An Index

Index insurance is linked to an index such as measures of rainfall, temperatures, or estimates of crop yields rather than losses observed by an insurance adjuster. Payments to farmers are triggered by pre-specified patterns of the index, as opposed to actual yields, which eliminates the need for in-field assessments. Insurance providers typically look to “historical burning cost analysis” to estimate the long term costs of the project by looking to the distribution of past index values to estimate the probability of a payout [12].

Often the risks faced by farmers are more complex than an easily defined weather event. Instead farms face risks from a variety of smaller and often covariate shocks such as pests, molds, and extreme weather event. However, our ability to understand the frequency and determine the causes of these losses remotely may be limited. In response, index insurance triggers can run the gamut from simple to complex.

### Examples

Simple triggers include indemnity against a particular weather related risk. Payments for simple index products are typically triggered by a specific and predefined event of interest, such as receiving less than 0.5 inches of rain during the growing season. The probability of that particular event can be determined by evaluating the distribution of the historical record of rainfall for any given location.

More complex triggers include crop models that simulate the growth of plants in response to changes in weather patterns. Examples of this include the World Food Programme’s (WFP) project in Ethiopia which provided rainfall based index insurance (against drought) as part of an integrated risk management project [11]. Historical rainfall data from the meteorological agency and a crop-water balance model were used to design a trigger that had an 80% correlation with the number of food aid beneficiaries from 1994 to 2004. The reliance on crop-water balance models lead to significant basis risk in early years as the modeled planting date did not correspond to actual farming practices [12].

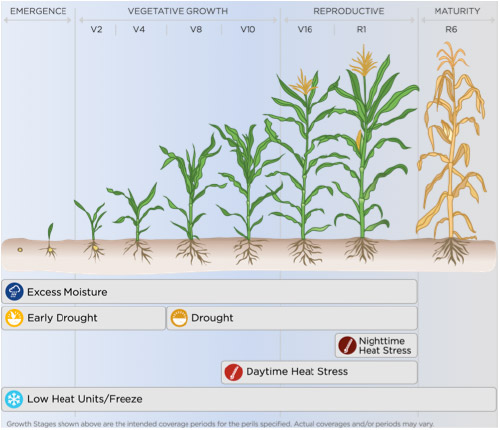
Another example is that of the Millennium Villages insurance program that used NDVI as the basis of its index. They use 8 km resolution images and evaluate the frequency of significant losses due to drought events, as proxied by vegetation greenness [12]. The index targeted the period of the growing season between plant flowering and harvesting because it showed the strongest correlation with historical yields. Despite its coarse spatial resolution (8 km), researchers found that drought stress observed on non-crop natural vegetation also corresponded to periods of significant reductions in crop yields [12]. To capture broader temporal trends, from changes in climate, the frequency of damaging droughts was calculated for a 100km area surrounding each village, thereby better capturing regional trends [12].

More complex indices including multiple sources of data should be considered, as the deficits of one data might be solved by the advantages of another [13]. For example, since the Millenium NDVI product insured against drought they used a weighted index of NDVI and rainfall estimates to trigger payments to farmers, thereby avoiding payout for losses not associated with drought [12]. Although this study points to the ability of remotely sensed data to predict shifts in wheat yields, care should be taken to identify short-coming and address them creatively through alternative data sources.

### CSI India Approach

Here we propose the use of EVI or NDVI vegetation indices and stationary camera data to underlie an insurance product that might cover a variety of risks including inadequate temperatures, excess rain, and possibly pests and diseases. The use of weekly farmer questionnaires and stationary cameras through a cellphone app provide us a unique view of both the farm plots but also the farmer’s perceptions of risk and damage. The approach proposed here has many advantages, but also some significant challenges.

An insurable index can be established through “historical burning cost analysis” for periods relevant to any given risk. In Figure 6 we can see that the growing season might be broken down into a series of insured risks (e.g. excess heat in the early growing season). The historical record of vegetation indices from 2002 onward would provide a distribution of values for a field, village, or district from which probabilities can be derived.



Pests / Molds

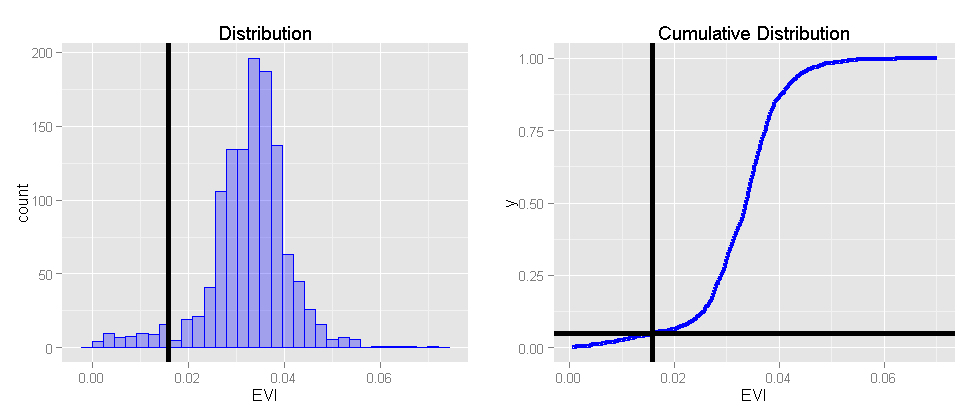
Excess Rain

Excess Heat

Figure 5: Timing of insurable events

Illustration of simulated plant growth and examples of the timing of insurable risks.

The accurate estimation of quantiles is of particular interest for this project. Insurers must be able to calculate the frequency of particular events over time and space. To address this issue we apply a set of functions that could be used to estimate the value of any percentile of the distribution. Importantly these can be estimated for a single location or for a polygon defining a village boundary/state/region etc. Additionally, the formulation of quantile estimation makes no assumption of normality and is median unbiased regardless of distribution [6], so quantiles can be estimated with confidence. However, for very small sample sizes bootstrapping with replacement should be considered to estimate our confidence in these estimates. We can see from Figure 5 below that the 5th percentile value can be accurately estimated from a skewed distribution with a heavy left tail.



5%

5%

Figure 6: Estimating quantiles, finding the 5% event

Left: Histogram of EVI values for a wheat field, within this sample values to the left of the vertical black line occur only 5% of the time. Right: Cumulative distribution of EVI values shown to the left, vertical black line corresponds to the same lower 5th percentile value.

One major limitation to this approach is the fact that the distribution can be significantly affected by one or two major events, while other risks, unobserved in the record, are left unevaluated [12]. However, this approach could be supplemented with simulated data to more accurately reflect long-term risks and trends. For instance, a longer time series of a vegetation index could be simulated using rainfall and climate reanalysis data, both of which have high quality data available back to the mid 1980’s or earlier. Another approach might be to run the “historical burning cost analysis” on an aggregated spatial scale, for instance at the district-level, thereby providing more observation (less prone to outliers) and a sample of index values more representative of the population.

The multifactorial risk approach (covering a broader set of farming risks) is unusual but has great potential. Because vegetation indices are indirect measurements of vegetation properties, they will reflect a variety of different shocks across a growing season. Ideally, each one would have a distinguishable temporal signature or fingerprint. However, our ability to attribute any change in a vegetation index to single extreme event or disease will likely be limited, particularly without access to high quality training data. This issue is partially ameliorated through the integration of farmer and professional evaluation of stationary camera data. Whereby we can assign labels (e.g. hail, flood damage) to particular time periods and locations. The ability of machine learning to observe and classify new unlabeled images (from field images and satellites) has great potential but remains an important as of yet unanswered question.

Another challenge is that of the non-stationary nature of the probabilities of extreme events under climatic change. Some effort can be made to de-trend the data, however the current ability to update insurance probabilities based on short- and long-term processes is still limited [12]. Advances in adaptive machine learning techniques, developed to specifically address issues of spatial and temporal non-stationarity are under development by Prof Monteleoni at George Washington University, and may address many of these concerns.

# Results of quantile estimates & Datasets

### Primary Data Outputs

As described above, the primary data output of this study will be crop relevant metrics summarizing a growing season or any other period of interest. This data in turn can be used to estimate crop yields or to provide the likelihood of observing a particular value of a vegetation index. Values of these metrics can be summarized on all pixels or aggregated at any other spatial scale. Aggregation can be completed by finding the mean EVI value for all pixels in the district for each 8-day period and then calculating the quantile estimate. Alternatively, as done in this exercise, pixel by pixel basis estimates can be found by calculating the quantiles based on the value of all individual pixels that fall within a polygon. We can see the lower 5th percentile of all EVI pixel values at the district level for Punjab and Haryana (Figure 7).

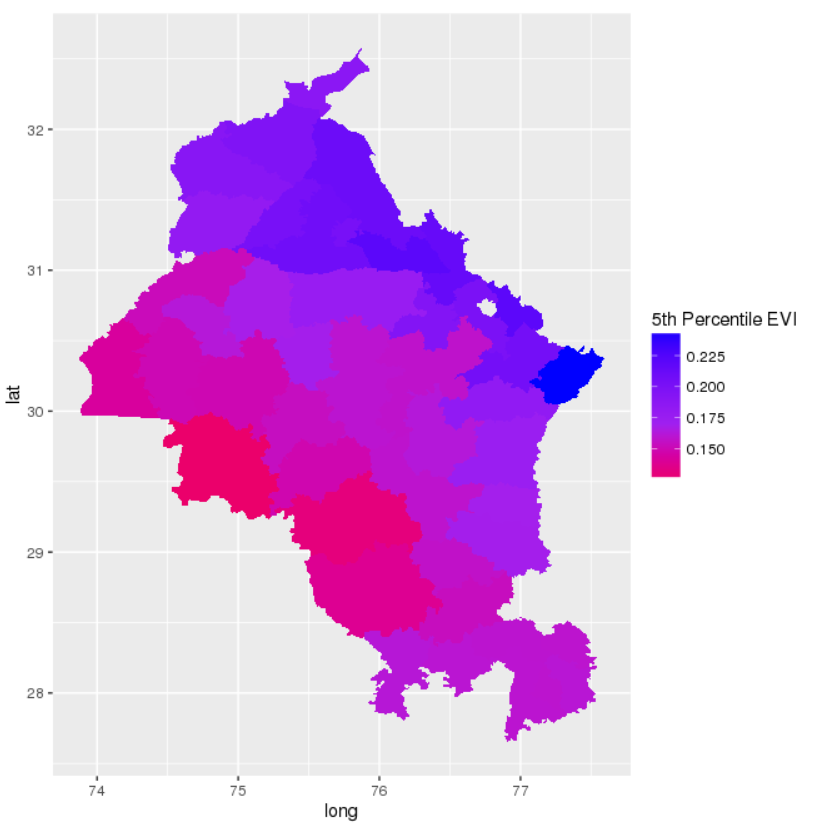


Figure 7: 5th Percentile Wheat Growing Season Enhanced Vegetation Index Values (2012-2016) by District

The lower fifth percentile values of EVI for all pixels within each district comprising Punjab and Haryana, spanning from low (red) to high (blue)

### Primary Software Outputs

One of the primary objectives of this project was to develop a clearly-specified and scalable workflow to estimate vegetation based indices in developing countries. All software was developed in the open-source language R and with freeware. The application of open-source languages and automation of processes should make all scripts accessible to GIS or programming experts, and therefore replicable by appropriate IFPRI staff.

The primary output of this project are scripts held in an open Github repository (<https://github.com/mmann1123/India-Index-Insurance-Code>). These scripts allow the user to calculate metrics for a time period of interest, for use in the proposed insurance index. To generate these statistics we developed 23 custom functions, and utilize parallel processing methods. Many of these functions provide summary statistics flexibly, taking other functions as parameters (for instance we can calculate the mean, median, maxima etc across one growing season using a single function). A great deal of effort was placed on creating scalable functions, in that we can run them for large spatial areas and or across long periods of time. Importantly these scripts allow the user to quickly run these calculations at a variety of temporal and spatial scales (cell, point, or polygon).

### Data and Metadata Standards

For large spatial data outputs, standard practice is to format them using Network Common Data Form (NetCDF), or standard raster formats such as geotiff. For disseminating smaller time-series data sets to the broader community, we will create comma-separated values (CSV) files, whenever possible. We will adhere to common metadata standards in order to provide clear descriptions of data sources and outputs, as well as instructions on citations.

### Plan for Data Sharing, Archiving and Preservation

The data collected during the project is available to participants and IFPRI. Much of the input data are stored on publically available websites and servers (and are freely available for download). Output data is archived and available. George Washington University’s Information Technology support services will provide high speed networking, database support and longer-term archival services.

# references

1. Mann, M.; Warner, J. *Ethiopian Wheat Yield and Yield Gap Estimation: A Small Area Integrated Data Approach*; Addis Ababa, Ethiopia, 2015.

2. Pan, Y.; L., L.; Zhang, J.; Liang, S.; Zhu, X.; Sulla-Menashe, D. Winter wheat area estimation from MODIS-EVI time series data using Crop Proportion Phenology Index. *Remote Sens. Environ.* **2012**, *119*, 232–242.

3. Fontana, A.; Potgieter, B.; Apan, A. *Assessing the relationship between shire winter crop yield and multi-temporal MODIS NDVI and EVI images*; Proceedings of SSC 2005 Spatial Intelligence, Innovation and Praxis: The national biennial conference of the spatial Sciences Institute; Melbourne, 2005.

4. Prasad, A. K.; Chai, L.; Singh, R. P.; Kafatos, M. Crop yield estimation model for Iowa using remote sensing and surface parameters. *Int. J. Appl. Earth Obs. Geoinf.* **2006**, *8*, 26–33.

5. Perry, E. M.; Morse-McNabb, E.; Nuttall, J.; O’Leary, G.; Clark, R. Managing Wheat From Space: Linking MODIS NDVI and Crop Models for Predicting Australian Dryland Wheat Biomass. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, 1–8.

6. Hyndman, R. J.; Fan, Y. Sample quantiles in statistical packages. *Am. Stat.* **1996**, *50*, 361–365.

7. FAO Crop Water Information: Wheat http://www.fao.org/nr/water/cropinfo\_wheat.html.

8. Didan, K.; Huete, A. MODIS Vegetation Index Product Series Collection 5 Change Summary 2006.

9. Fritz, S.; McCallum, I.; Schill, C.; Perger, C.; Grillmayer, R.; Achard, F.; Kraxner, F.; Obersteiner, M. Geo-Wiki. Org: The use of crowdsourcing to improve global land cover. *Remote Sens.* **2009**, *1*, 345–354.

10. Hazell, P. B. R. *The potential for scale and sustainability in weather index insurance for agriculture and rural livelihoods*; International Fund for Agricultural Development, 2010.

11. Greatrex, H.; Hansen, J.; Garvin, S.; Diro, R.; Le Guen, M.; Blakeley, S.; Rao, K.; Osgood, D. Scaling up index insurance for smallholder farmers: Recent evidence and insights. **2015**.

12. Hellmuth, M. E.; Osgood, D. E.; Hess, U.; Moorhead, A.; Bhojwani, H.; Singh, U.; Thornton, P. K.; Saka, A. R.; Barry Dent, J.; Mkandiwire, R. Index insurance and climate risk: prospects for development and disaster management. In *ACIAR Proceedings Series (Australia).*; Cornell Univ., Ithaca, NY (EUA)., 2009.

13. Brown, M. E.; Grace, K.; Shively, G.; Johnson, K.; Carroll, M. Using Satellite Remote Sensing and Household Survey Data to Assess Human Health and Nutrition Response to Environmental Change. *Popul. Environ.* **2014**, *36*, 48–72.

1. AVHRR channels are effectively ‘broader’ picking up larger portion of the red, green, blue, near-infrared therefore making comparison somewhat difficult and AVHRR data more sensitive to poor atmospheric conditions. [↑](#footnote-ref-2)
2. When summarizing raster values for an area of interest (polygon), statistics can be run on a pixel by pixel basis, or the function can be run on the mean values of EVI (NDVI) for the polygon [↑](#footnote-ref-3)
3. In the case of a tie the function chooses the first instance of the maxima for the max date. Ties separated by a single lower value, the middle value between the two instances of the maxima is chosen. Ties separated by two lower values, the middle left value is chosen. [↑](#footnote-ref-4)
4. This is the same as taking the integral of the EVI (NDVI) curve. By default integrals are estimated using the traditional trapezoidal method, except for the variable EVI\_growing\_AUC\_v2 which is estimated using the spline method. [↑](#footnote-ref-5)
5. Composite images compile data over a fixed number of days (typically 16 days) to create cloud-free images representative of that period. [↑](#footnote-ref-6)
6. Smoothing splines can be used to interpolate observations of EVI (NDVI) on a fixed 8-day calendar or used to minimize the effects of outliers. However, smoothing splines can remove a great deal of variability and may undermine the predictive power of the EVI time signature. [↑](#footnote-ref-7)