APA: Remote Sensing Applications for Insurance: A Predictive Model for Pasture Yield in the Presence of Systemic Weather

C. Brock Porth[*], Lysa Porth[†], Wenjun Zhu[‡], Milton Boyd[§], Ken Seng Tan[¶], and Kai Liu[‖]

Abstract

A robust predictive model for crop yield is essential for designing a commercially viable index-based insurance policy. Index insurance for crops is still at a relatively infant stage, and more research and development is needed in order to address a main limitation, which is referred to as basis risk (the mismatch between the loss indicated by the index and the actual loss suffered by the insured). Traditionally ground weather station measurements have been the most common approach used in weather indices, and this approach has often led to high levels of basis risk. Recent advances in satellite-based remote sensing provides new opportunities to use publicly available and transparent "big data," to potentially make index-based insurance policies more relevant by reducing basis risk. This is the first paper to provide a comprehensive comparison of thirteen Pasture Production Indices (PPI's), including those developed based on satellite-derived vegetation and biophysical parameter indices using data products from the Moderate Resolution Imaging Spectroradiometer - MODIS. A validation protocol is established, and a unique dataset covering the period from 2002 to 2016 from a network of pasture clip sites in the province of Alberta, Canada is used to demonstrate new applications for insurance based on remote sensing derived data. The results of the satellite-derived PPI's are compared to those PPI's based on ground weather station data as benchmarks,

[*]brockporth@gmail.com. University of Manitoba, Department of Biosystems Engineering.
[†]lysa.porth@umanitoba.ca. Warren Centre for Actuarial Studies and Research, University of Manitoba, Canada.
[‡]wjzhu@ntu.edu.sg. Nanyang Business School, Nanyang Technological University, Singapore.
[§]boyd@ms.umanitoba.ca. Warren Centre for Actuarial Studies and Research, University of Manitoba, Canada.
[¶]kstan@uwaterloo.ca. Department of Statistics and Actuarial Science, University of Waterloo, Canada.
[‖]k26liu@gmail.com, University of Prince Edward Island

which to date are the most common design for index insurance. Overall, the satellite-based indices report higher correlations with the ground truth forage yield data compared to the weather station PPI's. As an example, in 2015 and 2016 the best performing satellite-based PPI's produce correlations of close to 90%. When considering the whole sample period, the highest overall average correlation is 62.0% for the biophysical parameter PPI based on FPAR 500 meter MODIS, and the lowest overall average correlation is 43.8% for the vegetation PPI based on EVI 250 meter MODIS. Comparatively, the highest correlation reported for the weather station indices is for precipitation at 12.51%. A straightforward, yet effective PCA-based prediction model is discussed in this paper as an example model to address the issues of variable selection and dimension reduction in insurance design. This research makes an important contribution to the field of actuarial science and insurance, as it highlights potential new opportunities for insurance design and predictive analytics using large and comprehensive satellite datasets that remain relatively unexplored to date in insurance practice. While agriculture is used as the example here, the research could be extended to other crops, and other areas in the Property & Casualty sector, including for fire and flood.

Keywords: Crop Yield Prediction; Forage Insurance; Remote Sensing; Index-Based Insurance; Principal Component Analysis; Predictive Analytics.

1 Introduction

Forage includes plant material eaten by grazing livestock, or pasture land and also similar plants cut for fodder. The forage industry is heterogeneous and has several distinct sectors based on the end use of the forage crop. Forage includes both annual and perennial crops, which are usually harvested as silage, greenfeed or swath grazing, and grazed as pasture, harvested as greenfeed, stored as hay or silage, processed into pellets (alfalfa, etc.), cubes or compressed hay for domestic and export markets, respectively.

In Canada, forage crops have received coverage under AgriInsurance (crop insurance) since 1967. The forage insurance's participation rates, however, have been relatively low at approximately 10% to 20% of the potential land area. In contrast, the insurance participation rates for other crops (such as wheat, canola, corn, etc.) are approximately 70% to 85%. This has left the forage sector particularly vulnerable in times of substantial forage shortages due to adverse weather, leaving producers unable to produce or purchase the forage they need to feed their livestock, or to re-establish their forage. As a result of such forage and livestock industry risk, there has been pressure on Federal, Provincial and Territorial governments to provide disaster-related assistance. Since 2008, in Canada there have been four AgriRecovery initiatives in response to pasture or forage related issues, with government providing payments to producers totaling $148 million. Therefore, improving forage and pasture insurance, and increasing participation rates among producers has been identified as a priority for federal and provincial governments.

Traditional insurance programs used for other crops do not always work well for forage. For example, forage management systems may vary greatly from one farm to another, and may also change according to annual weather conditions. In addition, forage may be harvested multiple times throughout the growing season and/or grazed by livestock. When forage is produced directly for feed, it is often not accounted for by the producer because the producer may let their livestock feed on forage fields as necessary and do not calculate how much feed is consumed. Due to these reasons, accurate insurance loss estimates for forage are difficult to achieve even when relying on costly, and time-consuming, human expertise. Collectively, these challenges make it difficult to design a relevant forage insurance program and to determine actuarially fair premium rates. As a result, index-based insurance is thought to be an alternative framework for developing forage

insurance. With index-based insurance, the insurance payout is linked to an index, such as rainfall, temperature, or satellite-derived data, rather than the actual loss on the farm. Therefore, an index-based insurance approach may solve some of the problems that limit the application of traditional crop insurance in the case of forage, including for pasture.

1.1 Index Insurance Benefits and Challenges

Index-insurance may provide several advantages over traditional insurance, including lower transaction costs, fast and transparent settlement, and less adverse selection and less moral hazard. Cost efficiencies may be realized in theory since the insurance payouts are determined automatically from the index, rather than on-farm loss adjustments, which can be costly. The index should be constructed from underlying data that is transparent, reliable and representative. Information asymmetry problems, including adverse selection and moral hazard, can also largely often be avoided given that the individual loss characteristics of the producer often cannot influence the underlying index and the producer's ability to collect an insurance payout. Despite the many benefits of index insurance, it is very challenging to develop an index that is representative and suitable for the individual producer. One key difficulty is attributed to the basis risk which is triggered by the discrepancy between the insurance payout dictated by index and the actual. This can refer to situations where a producer suffers a loss, yet, the index does not trigger an insurance payout, or alternatively when the index determines an insurance payout, but, the producer does not suffer a loss on the farm.

Weather stations have traditionally been the primary data source for agricultural weather index insurance programs (e.g. temperature, rainfall, etc.). However, in many cases there are spatial data limitations in terms of the density of weather stations, and temporal data limitations due to the frequency of observations (time series) available from each weather station. Another challenge with weather insurance is that it does not directly cover other risks such as disease and pests, and may not include other weather risks, such as wind, snow, hail, etc. All of these additional risks may result in basis risk. Therefore, a main focus of this research is exploring alternative approaches to index insurance, including satellite/remote sensing imagery.

The Normalized Difference Vegetation Index (NDVI) is commonly used in satellite-based grassland insurance programs, however, there are a number of limitations (Roumiguié et al., 2015b).

For example, NDVI is quite sensitive to sensor effects, lighting conditions, atmospheric conditions and soil effects. This means that it may not always be stable over time and in space, possibly making it more difficult from an operational point of view in terms of incorporating new satellites and obtaining a sufficient and reliable historical data time series. As an alternative to NDVI, satellite-derived biophysical parameters such as the Fraction of Absorbed Photosynthetically Active Radiation (fAPAR), or the Fraction of Green Cover (fCover), provide a physically meaningful measure of the vegetation. These biophysical parameter measures may be more independent of the sensor, and may help to overcome the challenges associated with NDVI regarding data accuracy and reliability.

1.2 Research Objectives

The overall objective of this research is to demonstrate new applications for insurance based on remote sensing using the example of pasture crops in the province of Alberta, Canada. This is the first paper to provide a comprehensive comparison of thirteen Pasture Production Indices (PPI's), including those developed based on satellite-derived vegetation and biophysical parameter indices using data products from the Moderate Resolution Imaging Spectroradiometer - MODIS. A validation protocol is established, and a unique and proprietary dataset covering the period from 2002 to 2016 from a network of pasture clip sites managed by the government crop insurance agency of the province of Alberta, Canada is used to demonstrate new applications for insurance based on remote sensing derived data. A Principal Component Analysis (PCA)-based prediction model is provided as an example to demonstrate how predictive techniques can be used to combine the strengths of multiple satellite-based indices to enhance the design of index insurance and reduce basis risk.

The remainder of this paper is organized as follows. In the next section, a background and literature review on index insurance is presented, including an overview of alternative data sources for the construction of an index-based insurance policy. Section 3 provides a description of the data sources used in this research, including satellite-derived vegetation and biophysical parameter indices, ground weather station indices, and the ground truth forage yield data used for validation. In Section 4, the analysis is presented, which includes an overview of the PPI's, a discussion on aggregating the daily index values, and the assumed start and end dates of the PPI's. This section

concludes with correlation results for each of the PPI's relative to the ground truth forage yield data. The predictive analysis is provided in Section 5, including a discussion on the proposed PCA-based predictive model, and empirical results, as well as the out-of-sample predictions. Section 6 concludes the paper and Appendix A describes the alternative PPI's that are studied in the paper.

2 Background and Literature Review

Index-based insurance provides coverage based on an aggregate index that conveys information about the (individual) losses, making use of variables that are exogenous to the individual policy holder. An index linked to an objective weather event that is highly correlated to farm-level loss is an example of an index-based approach (Ceccato et al., 2005; Hazell and Hess, 2010). For index insurance to be successful, the risk faced by agricultural producers should be spatially correlated within a defined geographical region. In the case of agriculture, weather risk is the dominant cause of yield loss, with some estimates that as much as 70 to 90 percent of crop loss is attributed to adverse weather. Since weather patterns, such as droughts or floods, often impact the yields of many farmers within a region (i.e. yield losses are spatially correlated), index insurance represents an effective alternative to traditional farm-level crop insurance.

While index insurance addresses many of the shortcomings associated with traditional crop insurance as mentioned above, including higher administration costs, moral hazard, and adverse selection, the demand for index insurance overall has been relatively low. For example, despite substantial premium subsidy (often in excess of 60 percent of the actuarial value) pilot index insurance in Malawi (Giné, 2015) and India (Cole et al., 2011) reveal participation rates of only 20-30 percent. In developed countries, such as Canada, participation rates for traditional crop insurance (i.e. farm-level loss adjusting) is generally quite high, with approximately 80-85 percent of crop acres insured in many of the provinces. However, in the case of forage insurance in Canada, participation rates are generally quite low, and in many provinces index-based forage insurance has participation rates of only about 10 percent.

A number of possible explanations have been presented to help explain the relatively low demand for index insurance. For example, low demand may be attributed to actuarial challenges regarding data, product design, and ratemaking. However, basis risk is a dominant cause (Chantarat

et al., 2013; Deng et al., 2007; Jensen et al., 2016a), which refers to the risk that the underlying index is not perfectly correlated to actual losses. This will therefore lead to situations in which farmers are not indemnified for actual losses, or are paid indemnities despite having no actual losses (Skees et al., 1997). Therefore, minimizing basis risk is crucial because farmers will not fully insure if basis risk is present, even when premium rates are actuarially sound (Mobarak and Rosenzweig, 2012).

2.1 Alternative Data Sources for the Construction of an Index-Based Insurance Policy

Insurance indices based on weather station data have traditionally been the most common approach to index-insurance in agriculture. However, in many cases there are limitations in terms of the density of weather stations, and possibly the weather time series available from each station. Most research in this area has examined various interpolation techniques to help address this in order to enable the prediction of values at an unmeasured location using known data belonging to its neighborhoods. One such technique is the Inverse Distance Weighting (IDW) approach, which averages weather data from all nearby sample stations. This method has been used extensively in geospatial analysis primarily due to its simplicity, and in general this approach has been successful (Chen and Liu, 2012). However, a drawback of the IDW method is that its accuracy relies strongly on the density of the meteorological stations (Dirks et al., 1998; Paulson and Hart, 2006). A more complex approach to help reduce spatial basis risk is kriging, which overcomes many of the barriers of IDW. One of the benefits of this method is that spatial variation is taken into consideration. In recent years, a number of researchers have examined and employed the kriging method for meteorological studies and weather index insurance. Sun et al. (2003) use kriging to provide the delineation of raining areas in Australia. Paulson and Hart (2006) use a spatial kriging model using Markov Chain Monte Carlo methods to interpolate rainfall for farm pasture owners in Iowa. While Paulson and Hart (2006) find that the kriging approach can reduce spatial basis risk to a similar extent as IDW, other researchers find that kriging provides more accurate precipitation estimations compared to other methods (Li et al., 2006; Cao et al., 2015).

Despite the various interpolation techniques that can be used to help address the situation of limited or missing data, a remaining problem is that the low density of stations has been proven

to systemically underestimate extreme values, which are precisely those extreme events that the insurance program is intended to cover. A potential alternative to weather station data could be the use of satellite imagery. Satellite based crop yield estimation has been improving over time, and will continue to improve rapidly with advances in satellite technology. Satellites continue to improve with more bands, better sensors, and better resolution. Also, software and image processing capability continues to improve, along with more computing power (e.g. cloud computing), and more data storage is available to deal with big data at lower cost. As well, advances in machine learning, such as neural networks, can improve the computing and processing capability. In general, there are two main types of remote sensing index approaches to estimate crop yield, including vegetation indices and biophysical variable indices. The remaining of this subsection further elaborates these two indices.

The Normalized Difference Vegetation Index (NDVI) (Turvey and Mclaurin, 2012a; Pineiro et al., 2006) is a satellite based remote sensing approach, referred to as a vegetation index, that can be used to make vegetation estimates (e.g. forage yield estimates). NDVI is the most commonly used vegetation index, and is widely used for index-based insurance design (Chantarat et al., 2013; Jensen et al., 2016a). Several operational commercial grassland insurance programmes have adopted this approach, including in Spain, Mexico, Canada and the U.S. NDVI became very popular in the 1980's and 1990's and has since been expanded upon and improved (Mulla, 2013). NDVI is computed from visible (VIS) light and near-infrared (NIR) light, reflected by vegetation (NASA, 2017). The VIS light used in the NDVI formula is most often Red (R). Though sometimes Green (G) light is used instead of Red (R), as mentioned later below, and this is a more recent development called GNDVI. The value of NDVI is calculated as follows:

$$NDVI = \frac{(NIR - VIS)}{(NIR + VIS)}. \tag{1}$$

Higher NDVI indicates greener vegetation (higher yield), whereas lower NDVI indicates less green vegetation (lower yield). NDVI is based on the normalized difference between NIR and VIS. NDVI ranges from -1 to 1. It is commonly well above zero for moderate vegetation, and nearer to 1 for very dense vegetation. NDVI will be typically slightly above zero for no vegetation (bare soil), and may be negative for clouds, snow, or water, which have high reflectance.

Due to the challenges associate with NDVI (such as soil reflectance), numerous other im-

proved methods building upon NDVI have been developed for yield estimation (Mulla, 2013; Sripada et al., 2008). These more recent improved NDVI versions include adding the Green band (G), in addition to NIR and Red (R) bands, as adding the Green band (G) focuses on pigments other than chlorophyll absorbing radiation. For example, the Green Normalized Difference Vegetation Index (GNDVI) is defined as:

$$GNDVI = \frac{(NIR - G)}{(NIR + G)}. \tag{2}$$

The next group of NDVI improvements below were developed primarily to compensate for the effects of soil reflectance, given for example, that NDVI will tend to produce too high a value and overestimate crop yield when soil contains more water. For example, EVI (Enhanced Vegetation Index), attempts to use the Blue band (B) to correct for atmosphere, as the aerosol influences the Red band (R). EVI also attempts to address the problem of underestimation of yield, for high yield cases. As well, compared to NDVI, EVI is less sensitive to chlorophyll, and more sensitive to canopy structure variations, such as Leaf Area Index (LAI).

$$EVI = G * \frac{(NIR - R)}{(NIR + C1 * R - C2 * B + L)}, \tag{3}$$

where for example $L = 1, C1 = 6, C2 = 7.5$, and G (gain factor) $= 2.5$.

In addition to EVI, the Modified Soil Adjusted Vegetation Index (MSAVI2) and the Optimized Soil Adjusted Vegetation Index (OSAVI) are commonly used. As the names imply, these indices attempt to correct for the effect of exposed soil in the reflectance signal, and they are defined as follows:

$$MSAVI2 = 0.5 \left[2\,(NIR + 1) - \sqrt{(2NIR + 1)^2 - 8(NIR - R)} \right], \tag{4}$$

$$OSAVI = (1 + 0.16)\frac{(NIR - R)}{(NIR + R + 0.16)} \tag{5}$$

An alternative to a vegetation index is an approach that uses biophysical variables (parameters). LAI and the fraction of photosynthetically active radiation (FPAR) are popular biophysical variables, and these can be used to estimate yield. Some research suggests that biophysical variable approaches are superior to vegetation indices, such as NDVI, for quantifying biomass (Baret and Weiss, 2010; Camacho and Torralba, 2011). The biophysical parameter values are usually obtained by fitting a semi-empirical model (such as PROSAIL) to spectral data.

The yield of a crop depends on photosynthesis, photosynthetically active radiation (PAR) absorbed (e.g. sunlight), along with temperature, water, nutrients, and other factors. The amount of absorbed sunlight (PAR) depends on the amount of incoming sunlight and the plant's ability to absorb the sunlight. The amount of incoming sunlight can be measured, and the ability of the plant to absorb sunlight can be measured by the plant leaf area, or leaf area index (LAI), which is a biophysical variable that can be used to estimate crop yield. Related to LAI is FPAR, which refers to the fraction of absorbed PAR (APAR), to incoming PAR. FPAR = APAR/PAR, and is between 0 and 1. This measurement depends mainly on the leaf area of the plant (canopy), and is therefore related to the LAI (Rembold et al., 2013). Some of the main factors influencing PAR (and FPAR) in order of importance are: sun zenith angle, soil optical properties, leaf angle distribution, leaf optical properties (e.g. leaf pigment concentration), and leaf area index (Atzberger, 1997).

2.2 Review of Basis Risk Literature for Agricultural Insurance Applications

There is quite extensive literature that discusses the feasibility of weather index insurance for agriculture (Norton et al., 2012; Lin et al., 2015; Odening and Shen, 2014; Jensen et al., 2016b; Black et al., 2016). The majority of this research examines linkages between the weather index and crop yield at aggregated levels, such as for a district or county, rather than at the farm-level. A main reason for aggregating the index to a larger area is to improve the performance of the index, as the averaging effect of this aggregation tends to result in higher correlations between the index and crop yield. For example, correlations between rainfall and forage yield at the farm level are shown to be weakly dependent, and all are negative and statistically insignificant (Maples et al., 2016). When the scale of the rainfall index is expanded to cover a larger area, such as a township or county or municipality, the correlation results slightly improve, with typical values ranging from 0.1 to approximately 0.5 (Norton et al., 2012). While increasing the spatial scale of the index may result in improved correlations, in the end the accuracy of the insurance index relative to the farm-level loss is critical in order to provide adequate coverage of downside risk, and in general, correlations at an aggregated level (i.e. county) tend to underestimate the presence of basis risk. Given the various model assumptions that can be made, and in particular the different spatial scales the index can cover, it is difficult to directly compare the correlation results across research studies.

Other researchers have examined the feasibility of satellite-based vegetation indices, typically NDVI, in an attempt to strengthen the relationship between the insurance index and the crop yield. At the individual farmer-level, Rowley et al. (2007) report correlations of only 0.15 with NDVI for grasslands. In the same study, aggregating the NDVI to the county level results in correlations of 0.65. In another study, Makaudze and Miranda (2010b) find that the relationship between NDVI and corn and cotton yield in Zimbabwe at the district level range from 0.42 to 0.70. Some researchers have reported improved success with satellite-based biophysical parameter indices, such as LAI and FPAR, with correlations of approximately 0.80 (Bala and Islam, 2009).

This paper seeks to contribute to the literature in several ways. First, a comprehensive comparison of insurance indices computed from ground weather stations, satellite-based vegetation and biophysical parameter indices is conducted, and this allows the various approaches to be compared relative to one another based on standardized assumptions that hold across all of the models. Second, a unique and proprietary data set at the farm level is used to validate the various approaches, and this is expected to provide a more realistic and accurate evaluation of the potential basis risk for the design of an insurance index. Third, this paper provides an example of a simple, yet, effective predictive technique to combine the strengths of several satellite-based insurance indices into one multivariable index to enhance the performance of the index and to reduce the basis risk.

3 Data Sources and Descriptions

In this section, an overview of the sources of data used in this paper is provided. This includes the publicly available data used to construct the Pasture Production Indices (PPI's), which is based on satellite imagery as well as ground weather stations. In addition, ground truth forage yield data corresponding to improved and native pasture yield data from 25 clip sites in Alberta, Canada are used for validation. The same data is also used for the empirical analysis of the proposed PCA-based predictive model in Section 5. Included in the discussion of the data are possible limitations, which are important considerations when interpreting the results, as well as for future research and development.

3.1 Pasture Production Indices (PPI's)

In Subsection 2.1, we described two main types of remote sensing indices, namely vegetation indices (such as NDVI, GNDVI, EVI, MASVII2, OSAVI) and biophysical variable indices (such as LAI, FPAR). These indices, in turn, can be used to construct Pasture Production Indices (PPI's) for estimating crop yield. In particular we consider 11 PPI's that are constructed from the vegetation indices and two PPI's that are produced from the biophysical variable indices. The variants of PPI's depend on the type of indices, resolution of the satellite imagery, source of the satellite imagery, as well as the sampling observation frequency. These PPI's, which are labeled as $PPI_i, i = 1, \ldots, 13$, are described in greater details in Appendix A. It should be emphasized that each of these PPI's can be computed from the publicly available data.

In addition to the above 13 PPI's that are developed from either the vegetation indices or the biophysical variable indices, we consider three other PPI's that are based on Accumulated Precipitation (AccPcpn), Heating Degree Days (HDD), and Cooling Degree Days (CDD). The data required for producing these weather-based PPI's is available from the ground weather stations. A detailed description of these weather-based PPI's, which are labeled respectively as $PPI_{13}, PPI_{14}, PPI_{15}$, can be found in Appendix A.

3.2 Ground Truth Forage Yield Data

To evaluate PPI's, ground truth forage yield data is used for validation. The term "ground truth" originates from the geological and earth sciences disciplines to describe the validation of data by going out in the field and checking "on the ground." This term has been adopted in other disciplines to explain the idea of data that is "known" to be correct.

In Alberta there are currently three main insurance programs (policies) that provide coverage for hay and pasture producers. These include 1) Individual Farm Loss-Adjusted Hay program, as well as two insurance alternatives for Pasture; i.e. 2) Moisture Deficiency Insurance (MDI) program and a 3) Satellite Yield (SAT) program. The MDI program is an area-based index program that uses weather station and spring soil moisture information. The SAT program is also an area-based index program but it uses NDVI data at the township level. Producers can be insured under either the MDI or SAT program.

The ground truth forage yield data used in Alberta corresponds to the SAT program. Typically index-based insurance programs face challenges in terms of validation data. This is because payments to producers are based on the index rather than farm-level loss adjusting, therefore, actual farm-level historical forage yield data is often not tracked. However, in the case of Alberta, a network of 99 cage sites (clip locations) are available, and each site has several years of historical data from 2002 to 2016. The clip sites are used to monitor and estimate pasture plant growth, and are used to supplement the SAT information to revise payment rates. Therefore, pasture data from the network of cage sites are the focus of the validation in this paper.

Figure 1 provides an overview of the network of cage sites around the province. Most samples correspond to the South and South East part of Alberta, which is consistent with the current area of enrolment in the SAT program. It is estimated that there are approximately 24 million acres of pasture in the province of Alberta, and the clip site areas represent about 8 million acres; i.e. about 33.3%. Most of these clip locations are for native pasture, however, some observations are for improved pasture. Native pastures can be described as simple grazing environments that usually are dominated by native grasses and contain many other native herbs and shrubs. Management of native pastures are typically in the form of grazing management, and can also be used for hay production as a complement. In contrast, improved pastures are regularly seeded with specific species (crop mix), usually grasses in combination with legumes. Improved pastures are typically more productive than the local native pastures, and once sown they require more management in terms of fertilization, harvest and in some cases irrigation.

From the total 99 locations in the clip network in Alberta, 25 locations were chosen for the analysis, which includes historical observations from 2002 to 2016. Test sites were chosen to be representative, with 20 native pasture sites and five improved pasture sites. For each location there are six sites, with two cages per site. Currently, clipping at each site is done twice in the growing season (June and August). For clarity, the first clip corresponds to growth from season start until June, and the second clip corresponds to growth from season start until August (i.e. not regrowth from June to August). For the analysis in this paper, the yield measurement at each of the six sites are averaged to obtain an average area yield for that location for that measurement period, which is then compared to the PPI's.

Details regarding the data recorded at each clip site are outlined below.

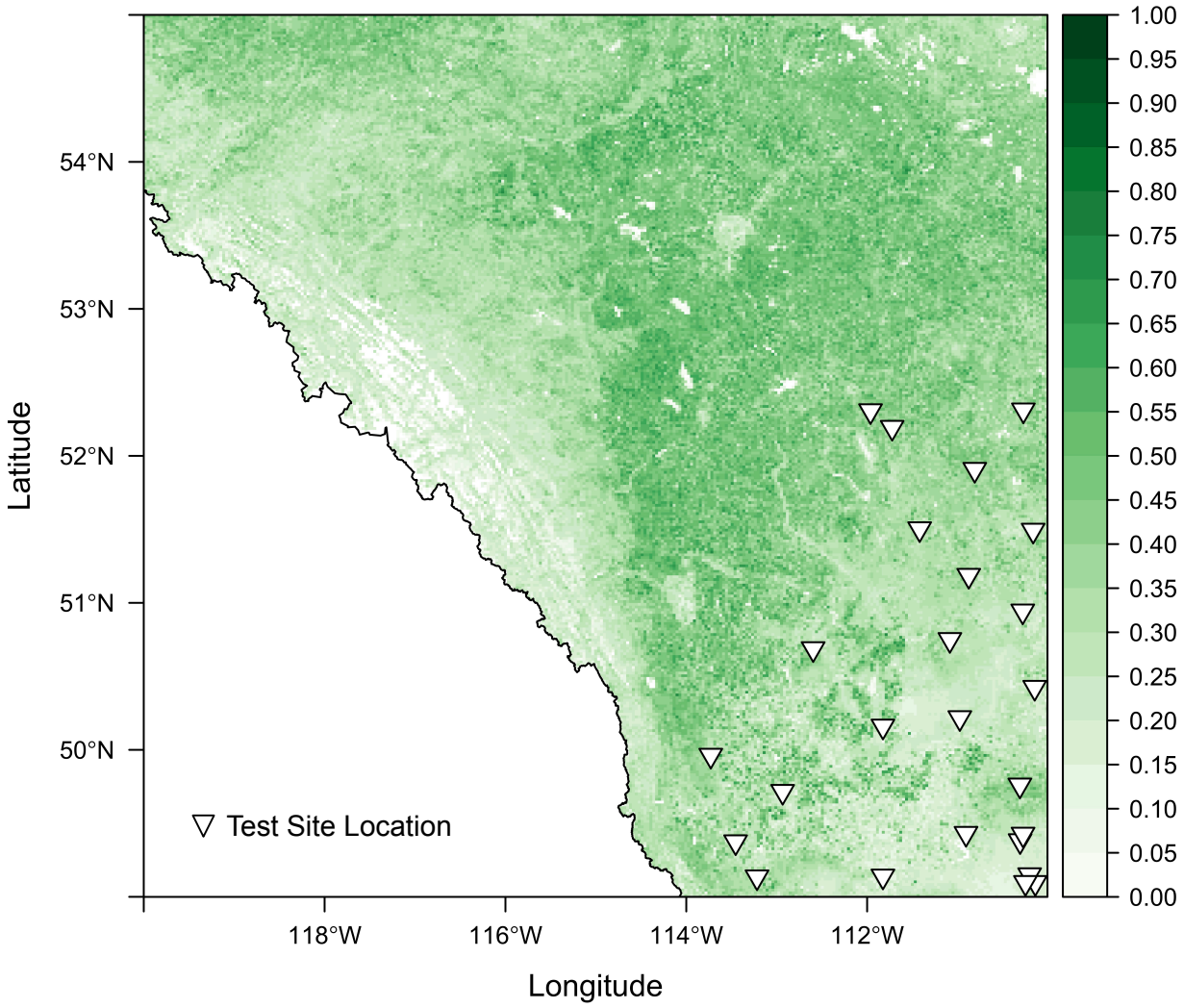


Figure 1: The network of 25 test site locations in southern Alberta, Canada is shown. As an example, the figure also shows the modified soil adjusted vegetation index (MSAVI2) at a 500 m per pixel resolution, which is calculated on August 15, 2016.

1. Latitude and Longitude: For each cage of the 25 locations the corresponding coordinates are recorded (latitude and longitude).

2. Yield is measured as follows:

   (a) 1 m by 0.5 m clip frame area and clipped at 1/4 inch (close to ground) for six cages at many of the 25 clip sites.

   (b) Live clipped grass and forbs (which refer to any herb that is not a grass or grasslike) sorted in field or office and dried to constant dry weight. The cage grass and forb weight is converted from the measurement unit of g/m^2 to a unit of $lbs/acre$ (dry weights of grass and forbs each multiplied by a factor of 17.84 to get grass and forbs weights in $lbs/acre$ and total (grass + forbs) in $lbs/acre$).

   (c) Two measurements are taken over the growing season in the months of June and August for the years 2012 to 2016, and four measurements are taken over the growing season in the months of May, June, July and August for the years 2011 and prior.

3. Grass Type: for each site, the grass type is recorded, including Dry Mixed Grass, Mixed Grass, Central Parkland, Foothills Fescue and Norther Fescue.

The polygon layer for the Alberta test sites were manually classified to include only acres that were determined to produce forage crops. This manual classification is possible given the relatively small number of test sites and the consistency of the geographic locations from year to year.

3.3 Observations of the Ground Truth Data

Preference would be to use the dataset in its entirety, however, some data cleaning is necessary for a number of different reasons, including data consistency from year to year, and variability in the pasture production measured at each clipping site. The data filtering process begins with the entire ground truth yield dataset, consisting of measurements from 25 test sites, each of which has six clipping cages with distinct locations identified by their longitude and latitude. After cleaning, we continue with 15 of the test sites. Next, each cage in the dataset is linked to the corresponding

index grid cell. All cage yield measurements are averaged, and a polygon layer is created that corresponds to the forage at the cage site locations. The polygon layer is used to sample the underlying alternative index, which is at a spatial resolution of either 250 m or 500 m, depending on the PPI.

Pasture production in Alberta is highly volatile across the province. This volatility may be due to a number of different factors, such as the location of the farm (e.g. altitude, latitude, soil, local weather, etc.), the type of pasture (native, improved), as well as differences due to farm management (e.g. applying nitrogen, carry over management, etc.). In some specific cases, known events (e.g. drought, flood) do not seem to be captured by the clipping data. This could be due to clipping location or to clipping protocol (and especially the date at which the measurement was made). It should also be noted that the ground truth forage yields are divided into two parts, including grass and forbs, where the sum of the two gives the total pasture production. Since the indices constructed in this study provide an estimate of total pasture production, the sum of grass and forbs is considered as a single yield value for each cage.

4 Empirical Analysis: Validation of PPI's to the Ground Truth Yield Data

In this section, an overview of the statistical validation methods regarding the assessment of the indices is provided. Next, background information relevant to the overall analysis is summarized, including an overview of the alternative PPI's, aggregation of the PPI's, and the measurement periods of the ground truth forage yield data. Following this, analysis of the alternative PPI's is then presented.

4.1 Statistical Validation Overview

To verify that the underlying index is a good representation of the loss experienced on the farm, it is necessary to quantitatively assess the relationship between the estimated forage yield computed via the PPI's and the ground truth forage yield data measured at the individual farm level. Correlation analysis is one of the most commonly used methods to understand how accurate the index can

predict the yield relative to the ground truth measured yield, and this is done by calculating the correlation coefficient between the two (Makaudze and Miranda, 2010a; Wehlage et al., 2016; Leblois and Quirion, 2013; Berg et al., 2009). Another statistical approach commonly used in validating vegetation indices is regression analysis (Turvey and Mclaurin, 2012b; Wehlage et al., 2016; Wall et al., 2008; Rojas, 2007; Quarmby et al., 1993; Potdar et al., 1999; Manjunath et al., 2002; Rasmussen, 1998; Labus et al., 2002; Huang et al., 2013; Atzberger et al., 2010), which involves identifying the relationship between a dependent variable (e.g. ground truth forage yield data) and one (or more) independent variable(s) (e.g., the various PPI's). With this approach a model of the relationship is hypothesized, and estimates of the parameter values are used to develop an estimated regression equation. The least square regression method is one of the most commonly used methods.

In general, we can first develop a model describing the relationship between the ground truth forage yield data and the index. By denoting y_t and $\hat{PPI}_t$ as the year t level of yield and PPI, respectively, the regression model in general can be expressed as

$$y_t = f(\hat{PPI}_t) + \epsilon_t \tag{6}$$

where $f(\hat{PPI}_t)$ is the functional form of the regression model and ϵ_t is the residuals of the model. The validity of the above regression model can be assessed by resorting to standard measures such as the goodness of fit test, coefficient of determination R^2, and root mean square error (RMSE). More specifically, large R^2 or small $RMSE$ values indicate a better PPI design.

Several studies have used the linear regression model to describe the relationship between NDVI and wheat yield in different regions. For example Lopresti et al. (2015) study the relationship between NDVI and wheat yield using a linear regression model, where the dependent variable is represented by wheat yield and the independent variable by NDVI. Aparicio et al. (2000) use Pearson correlation coefficients to study the relationship between radiometric indices and biological variables. The percentage of grain yield variation explained by the progressive addition of each spectral reflectance index measured at different growth stages is assessed by means of the coefficient of determination of a multilinear fitting. Hansen and Schjoerring (2003) compare the reflectance measurement of canopy biomass and nitrogen status in wheat crops using NDVI and partial least squares regression. Validation of the models is performed using regression analysis

and root mean square error (RMSE). Correlation analysis between the actual and predicted values was also calculated to assess the goodness of fit. The coefficient of determination R^2 was used as a relative index of model performance, and RMSE was used to compare the observed and predicted crop coefficients. These provide an indication of both bias and variance.

Further examples of the use of the linear regression model as a validation method can be found in Kamble et al. (2013), where NDVI based on MODIS is compared to a crop coefficient. Similarly, Wehlage et al. (2016) study crop yield forecasting in the Canadian Prairies using NDVI data from MODIS. Regression analysis is used with power function models to explain the crop yield variability in barley, canola, field peas, and spring wheat, and their findings are reported using RMSE values. Further, Roumiguié et al. (2015a) validate a forage index based on fCover derived from MODIS time series data. They use an upscaled approach based on direct validation, which compares a Grass Production Index with field-collected biomass data and high spatial resolution (HR) time series images. They perform validation of their models by comparing differences in R^2 and RMSE.

4.2 Analysis of Remote Sensing Based PPI's

Recall that Appendix A provides a description of the 13 PPI's that are derived from either the vegetation indices or the biophysical variable indices. In this subsection, we provide some empirical evidence on the relative efficiency of these PPI's by resorting to a correlation analysis. However, to properly define these PPI's there are two additional considerations which are described in the following two subsections.

4.2.1 Aggregation of Daily Index Values

A main consideration when analyzing the PPI's is how the values are aggregated over the desired measurement period. This process begins by interpolating the PPI to daily observations using the moving average method. Next, the daily PPI are integrated during the period interval of $[t_1, t_2]$ for the year. More specifically, let $\hat{PPI}_{t,n}$ denote the daily PPI in year t for $n = January\ 1, January\ 2, \ldots, December\ 31$. Then the PPI of year t is obtained by aggregating

the daily $\hat{PPI}_{t,n}$ over the period $[t_1, t_2]$; i.e.

$$\hat{PPI}_t = \sum_{n=t_1}^{t_2} \hat{PPI}_{t,n}. \tag{7}$$

An example of the daily observation time series is shown in Figure 2 for the MSAVI2 500 m resolution data by year from 2002 to 2016.

4.2.2 Start and End Dates for the PPI's

For the validation of the various PPI's to the ground truth forage yield data, another important consideration is the start and end dates that correspond to the measurement values. In the province of Alberta, the measurement periods are roughly known based on the recorded values at the various test sites. It is assumed that the season start date is May 15, and the end dates correspond to the two measurement periods, which are either June 30 or August 15. It should be noted that the time of the measurement likely changes from one year to the next, and this may introduce error into the subsequent analysis. Simple correlation analysis is conducted based on the two measurement periods, and the August measurement period produces the highest values. In addition, the earlier years of the dataset have relatively fewer observations for the June measurement compared to August. Therefore, the analysis proceeds using only the August measurement.

4.2.3 Correlation Results

Table 1 provides a summary of the thirteen PPI's based on vegetation and biophysical parameter approaches considered in this study. For each PPI, the average correlation results over the years 2002 to 2016 are shown for the measurement period defined as the growth from May 15 to August 15. The results show that the highest overall average correlation is 0.620 for PPI_{13} (i.e. FPAR MODIS 500 m - Daily Observations), closely followed by PPI_{10} (i.e._OSAVI MODIS 500 m - Daily Observations) at 0.614. The lowest correlation is 0.438 for PPI_7 (i.e. EVI MODIS 250 m - 8 Day Observations). It is interesting to note that the PPI's with the lowest resolution (i.e. 500 m) have the highest average correlations, and this is likely explained by the fact that the index is sampled daily, compared to higher resolutions (i.e. 250 m) where the index is sampled every 8 days. The daily sampling likely leads to more accurate index values. It is also interesting to note that both biophysical variable PPI's have average correlations exceeding 0.60.

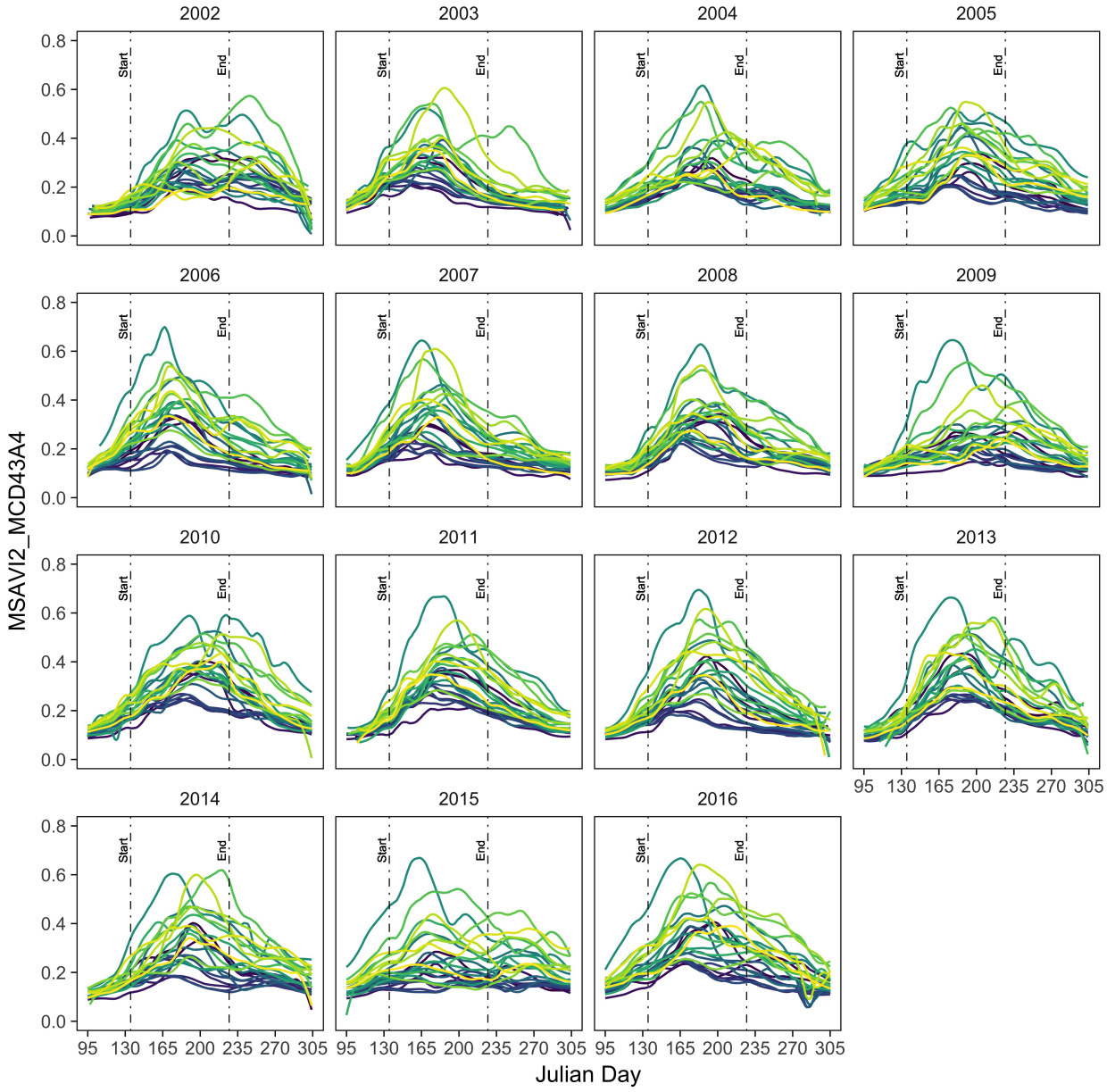


Figure 2: MSAVI2 time series data (growth profile) at 500 m resolution by year for each test site location. The dashed lines indicated a start date of May 15 and end date of August 15.

Table 1: Forage yield correlation results of the PPI's based on various vegetation index and biophysical parameter approaches, with average values reported for the years 2002 to 2016. The measurement period is defined as the growth period from May 15 to August 15.

Alternative PPI's	**Measurement Period**	**Ave. Correlation**
PPI_1: NDVI 1 km AVHRR	May 15 - Aug 15	0.512
PPI_2: NDVI MODIS 500 m - Daily	May 15 - Aug 15	0.613
PPI_3: NDVI MODIS 250 m - 8 Day	May 15 - Aug 15	0.567
PPI_4: GNDVI MODIS 500 m - Daily	May 15 - Aug 15	0.605
PPI_5: GNDVI MODIS 250 m - 8 Day	May 15 - Aug 15	0.564
PPI_6: EVI MODIS 500 m - Daily	May 15 - Aug 15	0.603
PPI_7: EVI MODIS 250 m - 8 Day	May 15 - Aug 15	0.438
PPI_8: MSAVI2 MODIS 500 m - Daily	May 15 - Aug 15	0.604
PPI_9: MSAVI2 MODIS 250 m - 8 Day	May 15 - Aug 15	0.543
PPI_{10}: OSAVI MODIS 500 m - Daily	May 15 - Aug 15	0.614
PPI_{11}: OSAVI MODIS 250 m - 8 Day	May 15 - Aug 15	0.559
PPI_{12}: LAI MODIS 500 m - Daily	May 15 - Aug 15	0.600
PPI_{13}: FPAR MODIS 500 m - Daily	May 15 - Aug 15	0.620

In addition to the average correlation results across all of the sample years, it is interesting to explore the year-by-year correlations for each of the satellite-based PPI's. Figure 3 provides the correlation results for all of the satellite-based PPI's for comparison over the years 2002 to 2016 for the August measurement period (May 15 and end date of August 15). For example, in 2015 and 2016 the best performing PPI's have correlations as high as 90%. All of the PPI's appear to be highly related to one another, moving in similar directions and exhibiting similar magnitudes. EVI appears to behave differently in 2005 and 2006, as well as 2010, 2011 and 2012. Overall, 2003, 2008, and 2011 have the lowest correlations over the time series. To the best of the author's knowledge, these years do not appear to correspond to a particular weather event, such as excessive drought or moisture, as examples. These lower correlations could be due to several of the data limitations previously discussed, such as timing of the measurement period, representativeness of the sample, etc. It is emphasized again that the correlations reported here should not be directly compared to those reported in other studies, given the many different model assumptions that are made. Instead, the indices should be directly compared more relative to one another based on the analysis in this paper.

4.3 Analysis of Weather Station Based PPI's

The remaining three PPI's considered in this paper are based on ground weather station data. The Forage Rainfall Insurance Program (FRIP) in the neighbouring province of Saskatchewan is used as the motivation for the assumptions regarding the weighting of the variables in each month to construct the index. The FRIP (policy) is available for native and tame grazing acres, providing protection for pastureland in the event that seasonal precipitation is below the long-term average. The current FRIP policy provides producers with several alternatives regarding the weighting of accumulated precipitation (AccPcpn) by month during the growing season. This is intended to provide flexibility to best match the producer's own growing and management experience. The weighting assumptions used for the analysis in Alberta are shown in Table 2. Therefore, in this subsection three weather station PPI's are considered, including heating degree days (HDD), cooling degree days (CDD) and accumulated precipitation (AccPcpn). Further, for each weather station PPI three weighting alternatives are considered using the current FRIP policy specifications as a guideline.

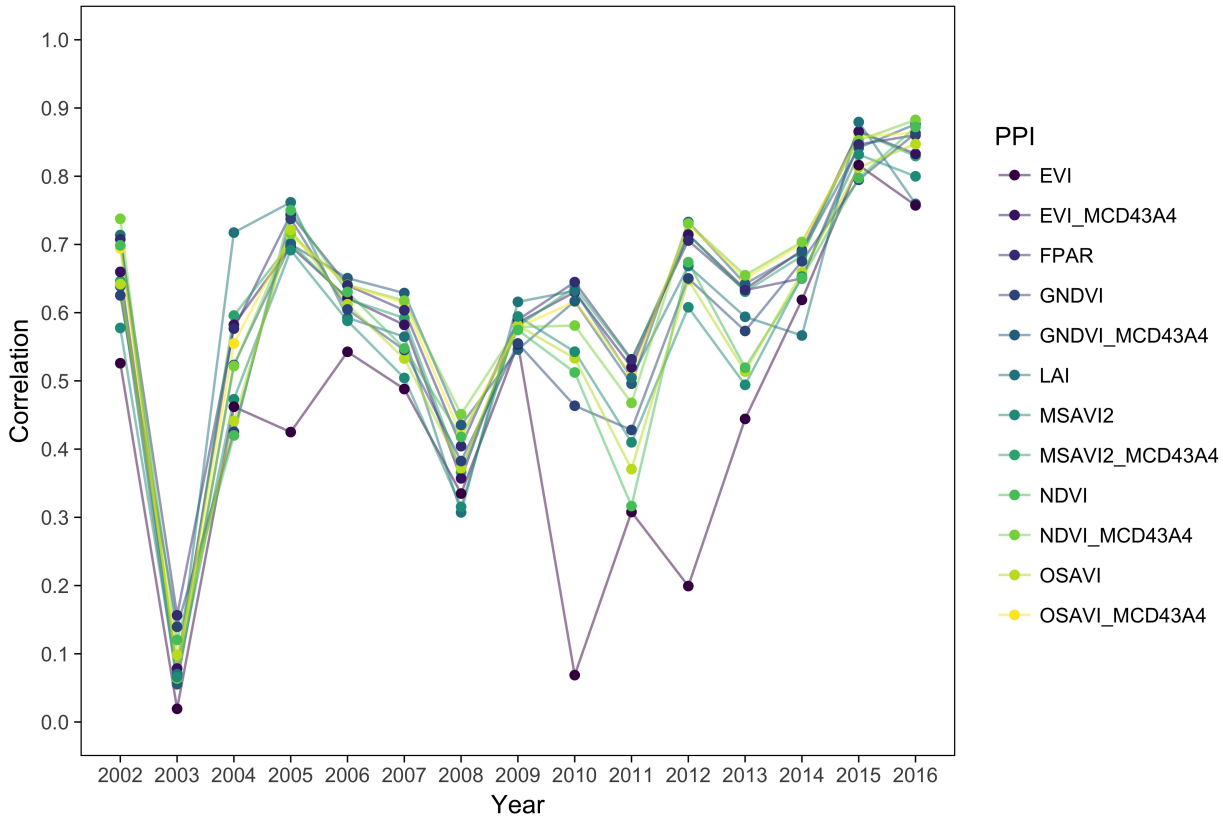


Figure 3: Results for the Forage Yield Correlation Analysis for each remote sensing based PPI for the years 2002 to 2016 for the August measurement period (May 15 and end date of August 15).

Table 2: Monthly Precipitation Weighting Options for the Ground Weather Station PPI's

Monthly Weather Weighting Options									
	June Cut				August Cut				
	April	May	June	Total	April	May	June	July	Total
Option 1	1/3	1/3	1/3	100%	30%	30%	30%	10%	100%
Option 2	1/3	1/3	1/3	100%	10%	40%	40%	10%	100%
Option 3	1/7	3/7	3/7	100%	10%	30%	30%	30%	100%

Table 3 shows the correlation results for each of the three PPI's and for each of the three weighting options over the samples from 2006 to 2016. The correlation results clearly demonstrate the inefficiency of the weather-based PPI's in capturing the ground truth yield. The highest attainable correlation is only 0.1251 from the precipitation variable with second weighting option (Pcpn02). The lowest (absolute value of) correlation is 0.04 and is produced from the CDD weather variable with third weighting option (CDDO3). When the correlations are examined on a year-by-year basis, the analysis shows a great deal of variation. This indicates that the PPI's based on the ground weather station data is not very robust. This variation may present a concern when used solely to construct the underlying insurance index, compared to some of the PPI's that appear to be more stable and with much higher correlation.

Table 3: Correlation Matrix of the Ground Truth Yield and Weather Variables (2006-2016). Last Digit of Each Weather Variable Indicates Weather Weighting Option.

	Yield	PcpnO1	HDDO1	CDDO1	PcpnO2	HDDO2	CDDO2	PcpnO3	HDDO3	CDDO3
Yield	1.0000									
PcpnO1	0.1007	1.0000								
HDDO1	−0.0274	0.7533	1.0000							
CDDO1	−0.1067	0.2682	0.4692	1.0000						
PcpnO2	0.1251	0.9931	0.7424	0.2875	1.0000					
HDDO2	−0.0666	0.7329	0.9780	0.3849	0.7048	1.0000				
CDDO2	−0.0981	0.2555	0.4472	0.9943	0.2787	0.3534	1.0000			
PcpnO3	0.0831	0.9891	0.7614	0.2559	0.9803	0.7455	0.2466	1.0000		
HDDO3	−0.0628	0.7446	0.9813	0.3899	0.7187	0.9963	0.3610	0.7574	1.0000	
CDDO3	−0.0400	0.2739	0.4208	0.9489	0.3049	0.3074	0.9455	0.2414	0.3142	1.0000

5 Predictive Analysis

In the preceding section, we conducted a correlation analysis on the 13 PPI's that are based satellite-derived vegetation and biophysical parameter, as well as weather data. Despite the popularity of the weather-based PPI's, it was found that their respective correlation with the ground truth yield is unacceptably low, hence casting doubt on the effectiveness of an index insurance based on the weather-based PPI's. The satellite-derived indices, especially that based on the biophysical

parameter, gave a much higher correlation. Large variability among these PPI's, however, were also observed. Prompted by their potential and their limitation, the objective of this section is to investigate how advanced predictive techniques can help reduce basis risk in designing index insurance from the satellite-derived PPI's. The method discussed in this section is a model based on principal component analysis (PCA). While the idea of PCA is not new, this is the first time it is applied to the satellite-derived PPI's. The PCA-based model is briefly described first, followed by empirical results based on data from the province of Alberta. Two important factors for predicting forage yield are identified and these correspond to the "Level factor" and "Tilt factor" of PCA. Finally, an illustrative predictive model based on PCA is analyzed with a special focus on the out-of-sample prediction ability.

5.1 PCA-Based Predictive Model

It is of interest to examine whether forage yield can be better explained by combining several satellite-based indices into a single predictive model. Information important for predicting forage yield may be contained in more than one PPI, however, the PPI's may be highly correlated and subject to collinearity. Therefore, a main challenge in constructing a robust predictive model using PPI data is to balance the task of variable selection while avoiding collinearity. In the case of crop insurance, historical data is often limited (due to only one growing season and therefore one data observation per year per location, among other factors). Therefore, forecasting results can be sensitive to variable selection, which is the focus of this subsection.

PCA is an efficient data-rotation and dimension-reduction technique, which replaces the highly correlated original data with a smaller number of uncorrelated linear combinations, reducing the dimensionality by keeping as much variability of the data as possible. Denote by $\widetilde{\boldsymbol{x}} = (\widetilde{\boldsymbol{x}}_1', ..., \widetilde{\boldsymbol{x}}_T')'$ the standardized $J \times T$ data matrix, i.e., the mean in $\widetilde{\boldsymbol{x}}$ is normalized to be 0, and the standard deviation is normalized to be 1. Then the principal components transform of $\widetilde{\boldsymbol{x}}$ is defined to be,

$$\boldsymbol{f} = \boldsymbol{\Gamma}'\widetilde{\boldsymbol{x}}, \tag{8}$$

which equivalently means that the i-th principal component (PC_i) is expressed as

$$\boldsymbol{f}_i = \boldsymbol{\gamma}_i'\widetilde{\boldsymbol{x}}, \tag{9}$$

where $\boldsymbol{\gamma}_i'$ is normalized so that $\boldsymbol{\gamma}_i'\boldsymbol{\gamma}_i = 1$. The objective of PCA is to find uncorrelated PC's of $\widetilde{\boldsymbol{x}}$, which are ordered by variance $\mathsf{Var}\,(\boldsymbol{f}_i) = \lambda_i$, as follows:

1. The first principal component, PC_1 is the standardized linear combination of $\widetilde{\boldsymbol{x}}$ with the maximal variance, i.e., $\mathsf{Var}\,(\boldsymbol{f}_i) = \lambda_1 = \max\{\mathsf{Var}\,(\boldsymbol{a'x}) : \boldsymbol{a'a} = 1\}$;

2. the i-th, $i = 2, 3, \ldots, r$, principal component, PC_i, is the standardized linear combination that maximizes the i-th variance that are orthogonal to the first $i-1$ principal components.

In empirical application of the PCA-based model, usually the number of PC's selected, r, satisfies the important condition that $r \ll J$, and therefore, substantial dimension reduction can be achieved.

5.2 PCA-Based Predictive Model Results

In this subsection, the empirical results of the proposed PCA-based model are presented. The PCA algorithm in Subsection 5.1 is applied to all satellite-based PPI's as described in Appendix A (with the exception of the NDVI AVHRR 1 km index)[1]. The analysis reveals two important explanatory and predictive factors for forage modeling, which are termed the "Level factor" and "Tilt factor".

Table 4 displays the variance explained by the first and second principle components (PC's), and Figure 4 shows the variance explained by each of the PC's as a bar plot. From the table it is apparent that the first and the second PC's explain more than 95% of the variance, and therefore, are identified as the most important factors.

In order to explore the meaning of the first two PC's, the loading vectors $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$, are examined. Table 5 displays the loading coefficients for each individual PPI for the first and second PC's for measurement periods for the months of June, July, and August. The table shows that the loading coefficients for the first PC is very similar for each PPI variable, remaining at a stable level

[1]To ensure a comprehensive analysis, the five PPI's based on vegetation indices (NDVI, EVI, GNDVI, MSAVI2, and OSAVI) include both 250 m per pixel and 500 m per pixel resolutions. In addition, two biophysical parameter indices (FPAR and LAI) are considered, which gives 12 types of satellite generated PPI's for the PCA. Three measurement periods of the ground truth forage yield is considered, including June, July, and August cuts.

Table 4: PCA variance explained by the first and second principal components (PC_1 and PC_2), for measurement periods for the months of June, July, and August.

	June	July	August
PC_1	0.9332	0.9291	0.9497
PC_2	0.0238	0.0326	0.0191
PC_1 + PC_2	0.9569	0.9617	0.9688

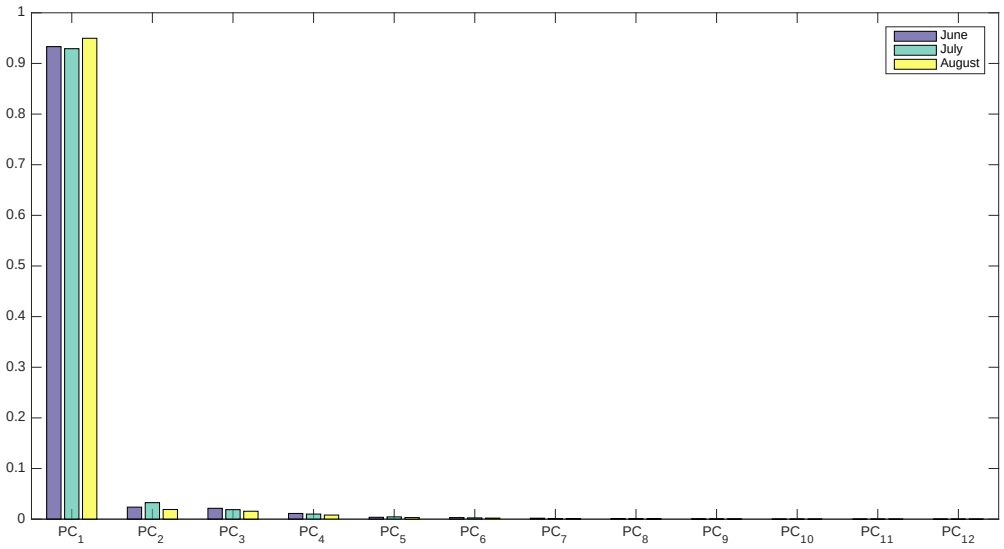


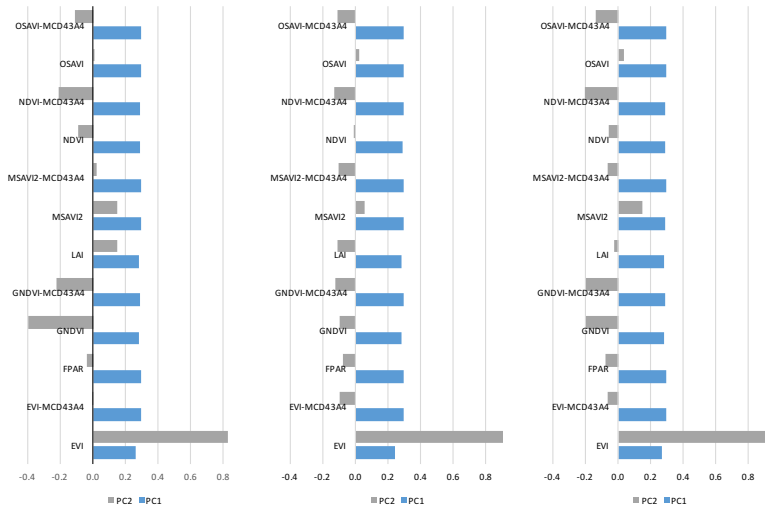
Figure 4: PCA variance for all principal components, $\lambda_i / \sum_{i=1}^{i=12} \lambda_i$, for the measurement periods for the months of June, July, and August.

Table 5: The loading vectors, $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$, for the first and second principal components (PC's), for measurement periods for the months of June, July, and August.

Index	PC_1			PC_2		
	June	July	August	June	July	August
EVI	0.2640	0.2401	0.2658	0.8259	0.9507	0.9065
EVI-MCD43A4	0.2947	0.2956	0.2933	-0.0067	-0.1008	-0.0666
FPAR	0.2943	0.2961	0.2936	-0.0351	-0.0744	-0.0749
GNDVI	0.2792	0.2827	0.2837	-0.3958	-0.0952	-0.1961
GNDVI-MCD43A4	0.2912	0.2935	0.2914	-0.2257	-0.1255	-0.1947
LAI	0.2805	0.2852	0.2829	0.1456	-0.1095	-0.0251
MSAVI2	0.2932	0.2941	0.2917	0.1483	0.0574	0.1487
MSAVI2-MCD43A4	0.2943	0.2950	0.2926	0.0200	-0.1027	-0.0619
NDVI	0.2898	0.2914	0.2897	-0.0941	-0.0136	-0.0587
NDVI-MCD43A4	0.2919	0.2937	0.2912	-0.2100	-0.1288	-0.2053
OSAVI	0.2934	0.2949	0.2927	0.0067	0.0237	0.0385
OSAVI-MCD43A4	0.2958	0.2971	0.2943	-0.1121	-0.1118	-0.1347

of approximately 0.3. However, the loading coefficients for the second PC differ in magnitude and sign across PPI's, displaying both positive and negative values. The biophysical parameter indices (LAI and FPAR) mainly have positive loadings for PC_2, while for the vegetation indices, the sign of the loading coefficient appears to be determined based on the resolution of the data (i.e. 250 m per pixel or 500 m per pixel). For example, several of the 250 m vegetation indices (such as EVI, MSAVI2, and OSAVI) have positive loading coefficients. All the 500 m resolution vegetation indices have negative loading coefficients. Therefore, it is suggested that PPI's with higher resolution (250 m per pixel) may produce positive loadings, while lower resolution (500 m per pixel) may produce negative loadings. This finding holds for both vegetation and biophysical parameter indices. The loading coefficients for PC_1 and PC_2 are illustrated in Figure 5.

The first principal component loading coefficients are similar for all indices, indicating that a movement in PC_1 induces a parallel movement in the entire production level. Therefore, PC_1 is termed "Level factor". The second principal component loading coefficients have positive loadings for higher resolution indices, and negative loadings for lower resolution indices. As a result, a movement in PC_2 induces a change in slope in the production level, which is a tilting effect. Therefore, this component is termed "Tilt factor".



(a) June (b) July (c) August

Figure 5: The loading coefficients $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$, for the first and second principal components (PC's) for measurement periods for the months of June, July, and August.

To more clearly visualize the impact of the "Level factor" and "Tilt factor", the observations are grouped by year and displayed in Figure 6 for average values of PC_1 and PC_2. The results show that PC_1 captures the shift in the same direction as the original PPI's, and is therefore termed the "Level factor" because it reflects the level/magnitude of the PPI's. PC_2 shows the slope of the indices, and moves in the same direction as the original PPI's for the higher resolution indices, however, for the lower resolution indices it moves in the opposite direction. In other words, PC_2 tilts up the higher resolution indices and tilts down the lower resolution indices, and therefore is termed the "Tilt factor".

5.3 Out-of-Sample PCA-Based Prediction Model Results

In order to analyze the predictive ability of the PCA-based model discussed above, in this subsection a simple PCA-regression model is constructed to forecast forage yield and test the models' out-of-sample predictive ability. As mentioned above, the PCA-based model has the advantage of effectively reducing the dimensionality of the original data, while also addressing the challenge of variable selection given limited degrees of freedom. The results show that the PCA-based predictive model is more accurate and robust, and produces smaller out-of-sample forecasting errors when compared to forage yield models constructed using individual PPI's.

In the formula below, $y_{i,t,s}$ denotes the forage ground truth yield data for test site i, in year t, for the measurement period in month s. The entire sample period $t = 2002, 2003, \ldots, 2016$; and measurement periods in months $s = June, July$, or $August$ is considered as:

$$y_{i,t,s} = \alpha + \boldsymbol{PC} \cdot \boldsymbol{\beta} + \xi_1 \mathbb{I}_{i,t,s}^{Jun} + \xi_2 \mathbb{I}_{i,t,s}^{Jul} + \varepsilon_{i,t,s}, \tag{10}$$

where α is the intercept term; $\boldsymbol{PC}$ is a vector of factors estimated from PCA; $\boldsymbol{\beta}$ is a vector of coefficients for each PC; $\mathbb{I}_{i,t,s}^{Jun}$ and $\mathbb{I}_{i,t,s}^{Jul}$ are dummy variables indicating June and July cuts, respectively[2] ; ξ_1 and ξ_2 are the corresponding coefficients; $\varepsilon_{i,t,s}$ are i.i.d. normal errors. Table 6

[2]Only two of the three measurement periods need to be considered as dummy variables, since including all three dummy variables would result in a singular matrix.

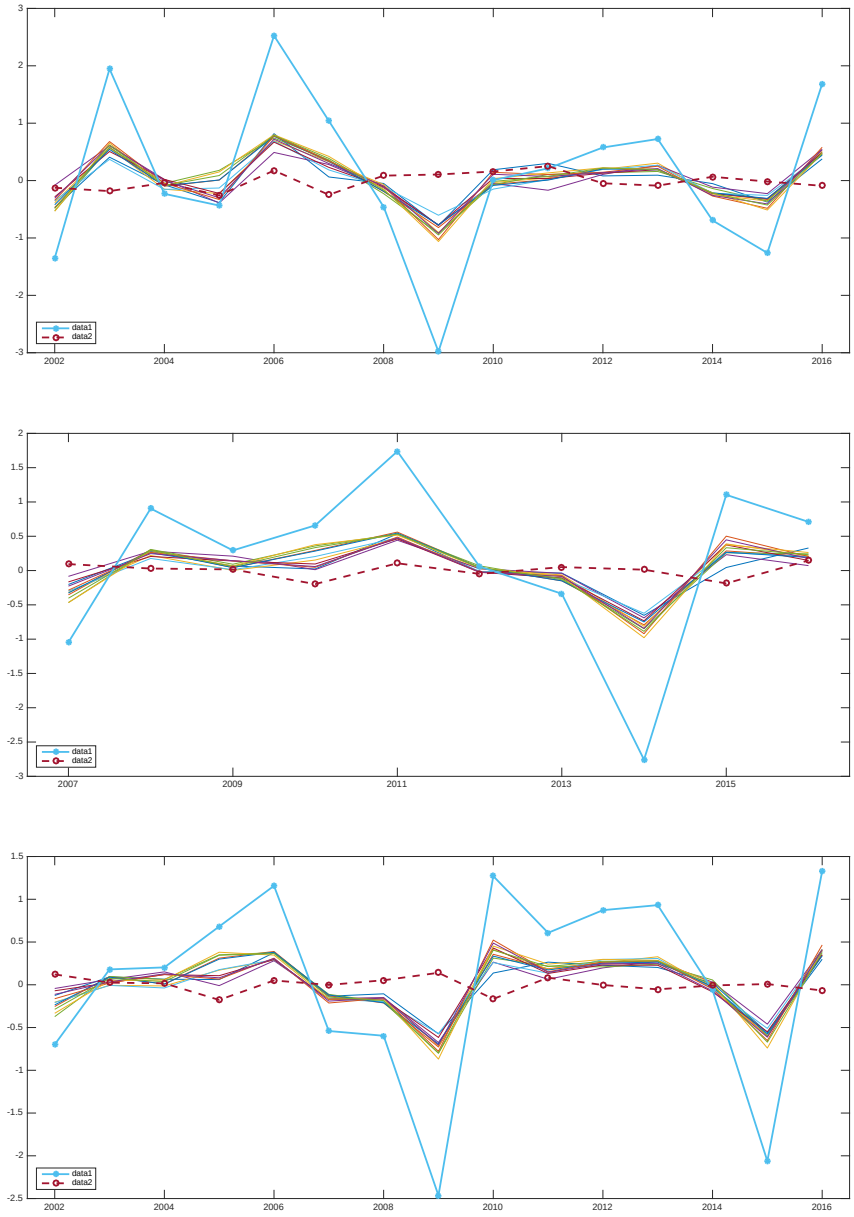


Figure 6: Resulting principal components values. This figure shows first and second principal component values for each year. The values are calculated based on average values of the two principal components grouped by years. Top to bottom are results for June, July, and August cuts, respectively.

compares the R^2 and adjusted R^2 (Adj-R^2) of the PCA-based model relative to the models from the individual PPI's, as well as the weather index with the best correlation results in Section 4.3 (i.e., precipitation with weighting option 2.). The R^2 (Adj-R^2) of the PCA-based model is 44.50% (43.83%), compared to the average R^2 (Adj-R^2) of the individual PPI 37.77% (37.45%), and 3.71% (1.07%) for the ground weather station index. Hence the PCA-based model is shown to fit better compared to all individual PPI models, particularly when compared to the benchmark model based on ground weather station data, which is currently the most common approach to weather index insurance in practice. We also compare the results to some shrinkage methods, including Lasso, Ridge regression, and Elastic Net,[3] and the results are also displayed in Table 6. We can see that the PCA-based method performs similarly compared to the other models.. As discussed above, forage crops present a number of insurability challenges, which make traditional insurance methods not possible. Therefore, governments around the world have identified the need to improve the methods for developing forage insurance. The results of this paper provide supporting evidence for the development of new applications for insurance based on satellite remote sensing,, and it is shown that integrating advanced predictive techniques, including the PCA and other shrinkage methods, can help to reduce basis risk and increase the commercial viability of the insurance.

A good in-sample fit does not necessarily imply good out-of-sample behaviour. Therefore, the out-of-sample predictive abilities of the PCA-based model in Equation (10) is also examined. Using 2014, 2015, and 2016 as the jump-off periods, the one-year-ahead out-of-sample predictive power is calculated by evaluating the root mean squared relative errors (RMSRE), defined as follows,

$$RMSRE = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(\frac{y_j - \hat{y}_j}{y_j}\right)^2}, \tag{11}$$

where $\hat{y}_j$ is the predicted yield value for observation j based on Equation (10), and N is the sample size. Table 7 and Figure 7 summarize the out-of-sample prediction results. The RMSREs for the PCA-based model constructed from the satellite-based PPI's are 0.9406, 1.2942, and 1.1315 for years 2014, 2015, and 2016, respectively. In comparison, the average RMSREs based on the 12 individual PPI for the respective years are 0.9553, 1.5917, and 1.227, which, respectively, are 1.6%, 23.0%,

[3]For a detailed introduction of the shrinkage methods, refer to Hastie et al. (2005).

Table 6: The R^2 and Adj-R^2 values for the PCA-based predictive model and the competing models. The first panel shows the results of the PCA model in Equation (10); the second panel shows results for individual PPI Models; the third panel summaries the average and standard deviations of the 12 individual index predictive models; and the forth panel shows results for shrinkage methods, including Lasso regression, Ridge regression, and Elastic Net. The last panel shows the results of the ground weather station index that has the highest correlation in Table 3, i.e., PcpnO2.

	R^2	Adj-R^2
PCA Model	0.4450	0.4383
EVI	0.2698	0.2660
EVI-MCD43A4	0.4130	0.4100
FPAR	0.4160	0.4130
GNDVI	0.3277	0.3243
GNDVI-MCD43A4	0.4016	0.3985
LAI	0.3855	0.3823
MSAVI2	0.3482	0.3449
MSAVI2-MCD43A4	0.4115	0.4085
NDVI	0.3605	0.3572
NDVI-MCD43A4	0.4179	0.4149
OSAVI	0.3582	0.3549
OSAVI-MCD43A4	0.4219	0.4189
Avg (Satellite Index)	0.3777	0.3745
Std (Satellite Index)	0.0465	0.0467
Lasso	0.4329	0.4250
Ridge	0.4458	0.4343
Elastic Net	0.3692	0.3594
Weather Index	0.0371	0.0107

and 8.4% larger. Hence the PCA-based model on average is more accurate with smaller relative forecasting errors compared to the individual PPI forecasting models. The predictive capability of the PCA-based model is further supported by comparing the results with the other with the shrinkage methods shown in Table 7. These models produce similar out-of-sample forecasting performance. In particular, Ridge regression slightly outperform the others in the year 2014 and 2015, while the PCA-based model is the best one in year 2016. Another potential complication with relying on an individual PPI is that there is no single PPI that consistently outperforms the other PPI's. In fact high variability among these PPI's are observed.

The deficiency of the weather-based PPI's is again highlighted in their much larger RMSREs. Comparing to the PCA-model, the forecasting errors from the benchmark PPI's based on ground weather station data for years 2014, 2015, and 2016 are, respectively, 68.8%, 59.3%, and 41.7% larger. Hence this demonstrates the superiority of using remote sensing data to improve the design of the insurance and reduce basis risk. Future research should focus on integrating additional information into the PCA-based model or other advanced predictive models, including ground weather station data, soil quality, soil moisture, and possibly ground water.

6 Summary

This research is focused on demonstrating new applications for insurance based on remote sensing using the example of pasture crops in the province of Alberta, Canada. A comprehensive comparison of Pasture Production Indices (PPI's) was provided, including those developed based on satellite-derived vegetation and biophysical parameter indices using data products from the Moderate Resolution Imaging Spectroradiometer - MODIS, and these are compared to benchmark indices based on ground weather station data. The paper also introduced a Principal Component Analysis (PCA)-based prediction model to address difficulties regarding basis risk, and two important factors for predicting forage yield were identified, termed a "Level factor" and a "Tilt" factor. To validate the PPI's, a unique and proprietary data set of a network of pasture clip sites at the farm level maintained by the provincial government crop insurance agency in Alberta was used to validate the insurance indices. Overall, the satellite-based indices reported higher correlations with the ground truth forage yield data compared to the weather station PPI's. As an example, in 2015 and 2016

Table 7: Out-of-sample prediction results. This table shows the one-year-ahead out-of-sample relative mean square errors (RMSE) for 2014, 2015, and 2016 jump-off years. The first panel shows the results of the PCA model in Equation (10); the second panel shows the results for individual PPI Models; the third panel summaries the average and standard deviations of the 12 individual index predictive models; the forth panel shows results for shrinkage methods, including Lasso regression, Ridge regression, and Elastic Net. The last panel shows the results of the ground weather station index that has the highest correlation in Table 3, i.e., PcpnO2.

Predict Year	2014	2015	2016
PCA regression	0.9406	1.2942	1.1315
EVI	1.2036	2.0630	1.2675
EVI-MCD43A4	0.9036	1.4690	1.1602
FPAR	0.7977	1.3342	1.1501
GNDVI	1.1720	1.9172	1.4056
GNDVI-MCD43A4	0.9523	1.4335	1.2344
LAI	1.0534	1.7904	1.2370
MSAVI2	0.9199	1.8222	1.2017
MSAVI2-MCD43A4	0.8946	1.4863	1.1447
NDVI	0.8984	1.5872	1.3281
NDVI-MCD43A4	0.9216	1.2370	1.1877
OSAVI	0.8748	1.6640	1.2525
OSAVI-MCD43A4	0.8720	1.2959	1.1542
Avg (Satellite Index)	0.9553	1.5917	1.2270
Std (Satellite Index)	0.1238	0.2623	0.0793
Lasso	0.9327	1.6984	1.2997
Ridge	0.7028	0.9033	1.2180
Elastic Net	1.2304	2.3030	1.4291
Weather Index	1.5879	2.0611	1.6036

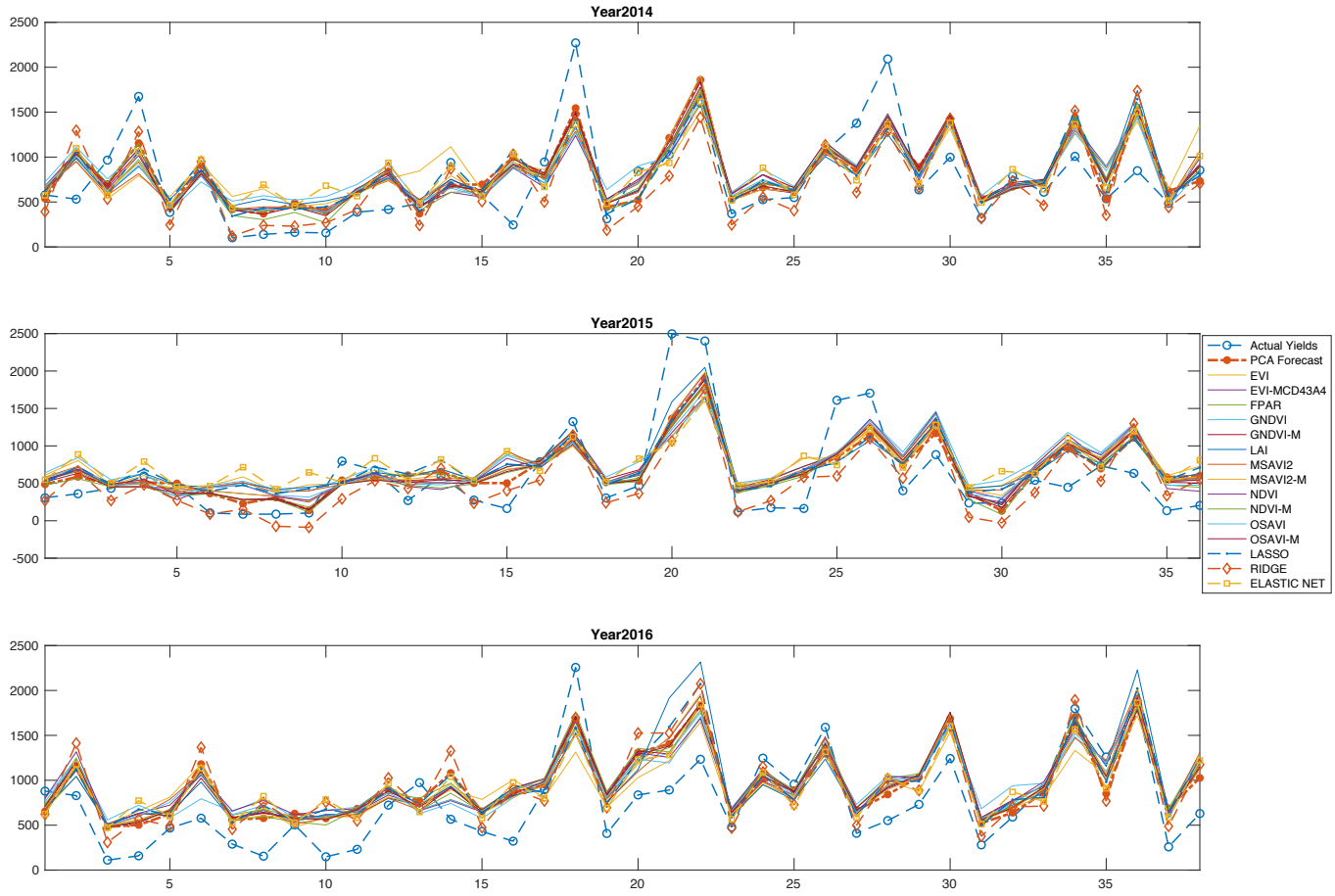


Figure 7: Out-of-sample prediction results. This figure displays the one-year-ahead out-of-sample prediction results for 2014, 2015, and 2016 jump off years.

the best performing satellite-based PPI's produced correlations of close to 90% at the farm level. When considering the average correlation results over the whole sample period (2002-2016), the biophysical parameter PPI based on FPAR produced the highest overall correlation of 0.620, closely followed by the vegetation index based on OSAVI MODIS 500 m - Daily Observations. The lowest average correlation was 0.438 for EVI MODIS 250 m - 8 Day Observations. It is interesting that the lower resolution (i.e. 500 m per pixel) indices have higher correlations with the ground truth forage yield data compared to the higher resolution (i.e. 250 m per pixel) indices. This may be due to the fact that the 500 m per pixel indices are sampled daily and may more accurately capture the temporal growth curve during the season resulting in more accurate integrated values for biomass. This is compared to the 250 m per pixel indices that are sampled every eight days, which likely results in less accuracy. The two biophysical parameter indices, which include LAI and FPAR, are based on lower resolutions (i.e. 500 m per pixel), and are produced daily. The three alternative PPI's based on ground weather station variables produce substantially weaker correlations. Of concern is that the correlations appear to be highly variable from one year to the next, and the highest overall correlation is for the PPI based on accumulated precipitation (Precpn) at 12.51%.

In an effort to combine the predictive strengths of several satellite-based PPI's, a simple and effective PCA-based model was introduced. Given recent advancements in computing power, and the increased accessibility of free satellite data, such as the MODIS data considered in this study, more advanced predictive techniques for developing insurance indices may prove valuable for the insurance industry, addressing the issues of variable selection and dimension reduction. The empirical out-of-sample prediction results of the PCA-based model constructed from the satellite-based PPI's shown as an example in this paper are found to be more accurate with smaller relative forecasting errors compared to the individual PPI forecasting models. For example, the R^2 (Adj-R^2) of the PCA-based model is 44.50% (43.83%), compared to the average R^2 (Adj-R^2) of the individual PPI 37.77% (37.45%), and 3.71% (1.07%) for the ground weather station index. While some may argue that the overall average correlations between the PPI and the farm level ground truth forage yield data are too low to be commercially viable, compared to related studies the results presented here are substantially improved and provide a more accurate assessment of basis risk at an insurance scale at the farm level (rather than insurance scales at a more aggregate level, such as a county).

Future research should focus on integrating additional information into the PCA-based model, including ground weather station data, soil quality, soil moisture, and possibly ground water. Accurate yield forecasting is central to effective forage risk management for many stakeholders, including farmers, insurers, reinsurers and government. As a result, this research makes an important contribution to the field of actuarial science, as it highlights potential new opportunities for insurance design and predictive analytics using large and comprehensive satellite datasets that remain relatively unexplored to date in insurance practice. This includes applications to areas of agricultural insurance, such as is the focus of this paper, as well as to other sectors within the Property & Casualty sector, such as for floods, fire, etc.

References

Aparicio, N., Villegas, D., Casadesus, J., Araus, J. L., and Royo, C. (2000). Spectral vegetation indices as nondestructive tools for determining durum wheat yield. Agron. J., 92(1):83–91.

Atzberger, C. (1997). Estimates of Winter Wheat Production through Remote Sensing and Crop Growth Modelling: A Case Study on the Camargue Region. Ph.D. Thesis. PhD thesis, Verlag für Wissenschaft und Forschung, Berlin, Germany.

Atzberger, C., Guérif, M., Baret, F., and Werner, W. (2010). Comparative analysis of three chemometric techniques for the spectroradiometric assessment of canopy chlorophyll content in winter wheat. Computers and Electronics in Agriculture, 73(2):165 – 173.

Bala, S. and Islam, A. (2009). Correlation between potato yield and modis-derived vegetation indices. International Journal of Remote Sensing, 30(10):2491–2507.

Baret, F. and Weiss, M. (2010). Towards an operational gmes land monitoring core service - biopar methods compendium - lai, fapar, f cover ndvi. Technical Report 1.5, European Research Project Geoland 2 (EC Proposal Reference No. FP-7-218795).

Berg, A., Quirion, P., and Sultan, B. (2009). Weather-index drought insurance in burkina-faso: Assessment of its potential interest to farmers. Weather, Climate, and Society, 1(1):71–84.

Black, E., Tarnavsky, E., Maidment, R., Greatrex, H., Mookerjee, A., Quaife, T., and Brown, M. (2016). The use of remotely sensed rainfall for managing drought risk: a case study of weather index insurance in zambia. Remote Sensing, 8(4):342.

Camacho, F. and Torralba, I. (2011). Towards an operational gmes land monitoring core service – validation report high resolution vegetation parameters. Technical Report 1.1, European Research Project Geoland 2 (EC Proposal Reference No. FP-7-218795).

Cao, X., Okhrin, O., Odening, M., and Ritter, M. (2015). Modelling spatio-temporal variability of temperature. Computational Statistics, 30(3):745–766.

Ceccato, P., Connor, S., Jeanne, I., and Thomson, M. (2005). Application of geographical information systems and remote sensing technologies for assessing and monitoring malaria risk. Parassitologia, 47(1):81–96.

Chantarat, S., Mude, A. G., Barrett, C. B., and Carter, M. R. (2013). Designing index-based livestock insurance for managing asset risk in northern kenya. Journal of Risk and Insurance, 80(1):205–237.

Chen, F.-W. and Liu, C.-W. (2012). Estimation of the spatial rainfall distribution using inverse distance weighting (idw) in the middle of taiwan. Paddy and Water Environment, 10(3):209–222.

Cole, S., Sampson, T., and Zia, B. (2011). Prices or knowledge? what drives demand for financial services in emerging markets? The Journal of Finance, 66(6):1933–1967.

Deng, X., Barnett, B. J., and Vedenov, D. V. (2007). Is there a viable market for area-based crop insurance? American Journal of Agricultural Economics, 89(2):508–519.

Dirks, K., Hay, J., Stow, C., and Harris, D. (1998). High-resolution studies of rainfall on norfolk island: Part ii: Interpolation of rainfall data. Journal of Hydrology, 208(3):187 – 193.

Giné, X. (2015). Innovations in insuring the poor: Experience with weather index-based insurance in india and malawi.

Hansen, P. and Schjoerring, J. (2003). Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. Remote Sensing of Environment, 86(4):542 – 553.

Hastie, T., Tibshirani, R., Friedman, J., and Franklin, J. (2005). The elements of statistical learning: data mining, inference and prediction. The Mathematical Intelligencer, 27(2):83–85.

Hazell, P. B. and Hess, U. (2010). Drought insurance for agricultural development and food security in dryland areas. Food Security, 2(4):395–405.

Huang, J., Wang, X., Li, X., Tian, H., and Pan, Z. (2013). Remotely sensed rice yield prediction using multi-temporal ndvi data derived from noaa's-avhrr. PLOS ONE, 8(8):1–13.

Jensen, N. D., Barrett, C. B., and Mude, A. G. (2016a). Index insurance quality and basis risk: Evidence from northern kenya. American Journal of Agriculture Economics, 98(5):1450–1469.

Jensen, N. D., Barrett, C. B., and Mude, A. G. (2016b). Index insurance quality and basis risk: evidence from northern kenya. American Journal of Agricultural Economics, 98(5):1450–1469.

Kamble, B., Kilic, A., and Hubbard, K. (2013). Estimating crop coefficients using remote sensing-based vegetation index. Remote Sensing, 5(4):1588–1602.

Labus, M. P., Nielsen, G. A., Lawrence, R. L., Engel, R., and Long, D. S. (2002). Wheat yield estimates using multi-temporal ndvi satellite imagery. International Journal of Remote Sensing, 23(20):4169–4180.

Leblois, A. and Quirion, P. (2013). Agricultural insurances based on meteorological indices: realizations, methods and research challenges. Meteorological Applications, 20(1):1–9.

Li, J., Zhang, J., Zhang, C., and Chen, Q. (2006). Analyze and compare the spatial interpolation methods for climate factor. Pratacultural Science, 8:001.

Lin, J., Boyd, M., Pai, J., Porth, L., Zhang, Q., and Wang, K. (2015). Factors affecting farmers? willingness to purchase weather index insurance in the hainan province of china. Agricultural Finance Review, 75(1):103–113.

Lopresti, M. F., Di Bella, C. M., and Degioanni, A. J. (2015). Relationship between modis-ndvi data and wheat yield: A case study in northern buenos aires province, argentina. Information Processing in Agriculture, 2(2):73 – 84.

Makaudze, E. and Miranda, M. J. (2010a). Catastrophic drought insurance based on the remotely sensed normalized difference vegetation index for smallholder farmers in zimbabwe. 2010 AAAE Third Conference/AEASA 48th Conference, September 19-23, 2010, Cape Town, South Africa 96183, African Association of Agricultural Economists (AAAE).

Makaudze, E. M. and Miranda, M. J. (2010b). Catastrophic drought insurance based on the remotely sensed normalised difference vegetation index for smallholder farmers in zimbabwe. Agrekon, 49(4):418–432.

Manjunath, K. R., Potdar, M. B., and Purohit, N. L. (2002). Large area operational wheat yield model development and validation based on spectral and meteorological data. International Journal of Remote Sensing, 23(15):3023–3038.

Maples, J. G., Brorsen, B. W., and Biermacher, J. T. (2016). The rainfall index annual forage pilot program as a risk management tool for cool-season forage. Journal of Agricultural and Applied Economics, 48(1):29–51.

Mobarak, A. and Rosenzweig, M. (2012). Selling formal insurance to the informally insured.

Mulla, D. J. (2013). Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. Biosystems Engineering, 114(4):359–371.

NASA (2017). Measuring Vegetation (NDVI & EVI) : Feature Articles.

Norton, M. T., Turvey, C., and Osgood, D. (2012). Quantifying spatial basis risk for weather index insurance. The Journal of Risk Finance, 14(1):20–34.

Odening, M. and Shen, Z. (2014). Challenges of insuring weather risk in agriculture. Agricultural Finance Review, 74(2):188–199.

Paulson, N. D. and Hart, C. E. (2006). Editors, "a spatial approach to addressing weather derivative basis risk: A drought insurance example. In Iowa State University, American.

Pineiro, G., Oesterheld, M., and Paruelo, J. (2006). Seasonal variation in aboveground production and radiation-use efficiency of temperate rangelands estimated through remote sensing. Ecosystems, 9:357–373.

Potdar, M. B., Manjunath, K. R., and Purohit, N. L. (1999). Multi-season atmospheric normalization of NOAA AVHRR derived NDVI for crop yield modeling. Geocarto International, 14(4):52–57.

Quarmby, N. A., Milnes, M., Hindle, T. L., and Silleos, N. (1993). The use of multi-temporal ndvi measurements from avhrr data for crop yield estimation and prediction. International Journal of Remote Sensing, 14(2):199–210.

Rasmussen, M. S. (1998). Developing simple, operational, consistent ndvi-vegetation models by applying environmental and climatic information. part ii: Crop yield assessment. International Journal of Remote Sensing, 19(1):119–139.

Rembold, F., Atzberger, C., Savin, I., and Rojas, O. (2013). Using low resolution satellite imagery for yield prediction and yield anomaly detection. Remote Sens., 5:1704–1733.

Rojas, O. (2007). Operational maize yield model development and validation based on remote sensing and agro-meteorological data in kenya. Int. J. Remote Sens., 28(17):3775–3793.

Roumiguié, A., Jacquin, A., Sigel, G., Poilvé, H., Hagolle, O., and Daydé, J. (2015a). Validation of a forage production index (fpi) derived from modis fcover time-series using high-resolution satellite imagery: Methodology, results and opportunities. Remote Sensing, 7(9):11525–11550.

Roumiguié, A., Jacquin, A., Sigel, G., Poilvé, H., Lepoivre, B., and Hagolle, O. (2015b). Development of an index-based insurance product: validation of a forage production index derived from medium spatial resolution fcover time series. GIScience & Remote Sensing, 52(1):94–113.

Rowley, R. J., Price, K. P., and Kastens, J. H. (2007). Remote sensing and the rancher: Linking rancher perception and remote sensing. Rangeland Ecology & Management, 60(4):359–368.

Skees, J. R., Black, J. R., and Barnett, B. J. (1997). Designing and rating an area yield crop insurance contract. American journal of agricultural economics, 79(2):430–438.

Sripada, R., J.P., S., Dellinger, A., and Beegle, D. (2008). Evaluating multiple indices from a canopy reflectance sensor to estimate corn n requirements. Agronomy Journal, 100:1553–1561.

Sun, X., Manton, M., and Ebert, E. E. (2003). Regional rainfall estimation using double-kriging of raingauge and satellite observations. Bureau of Meteorology.

Turvey, C. G. and Mclaurin, M. K. (2012a). Applicability of the normalized difference vegetation index (ndvi) in index-based crop insurance design. Weather, Climate, and Society, 4(4):271–284.

Turvey, C. G. and Mclaurin, M. K. (2012b). Applicability of the normalized difference vegetation index (ndvi) in index-based crop insurance design. Weather, Climate, and Society, 4(4):271–284.

Wall, L., Larocque, D., and Léger, P. (2008). The early explanatory power of ndvi in crop yield modelling. International Journal of Remote Sensing, 29(8):2211–2225.

Wehlage, D. C., Gamon, J. A., Thayer, D., and Hildebrand, D. V. (2016). Interannual variability in dry mixed-grass prairie yield: A comparison of modis, spot, and field measurements. Remote Sensing, 8(10).

A Pasture Production Indices (PPI's)

This appendix provides a more detailed explanation on the alternative PPI's that are studied in this paper. The first eleven PPI's are developed from the vegetation indices, the next two PPI's are based on the biophysical variable indices, and the final three PPI's depend on the weather station data.

1. PPI_1: NDVI AVHRR 1 km
   - Uses satellite data at the 1 km resolution from the Advanced Very High Resolution Radiometer (AVHRR) of the National Oceanic and Atmospheric Administration (NOAA) series of Earth observation (EO) satellites, and obtained from Statistics Canada.
   - The current pasture insurance program in Alberta uses this 1km AVHRR-NDVI, therefore, this serves as an important consideration in the valuation of the various PPI's.
2. PPI_2: NDVI MODIS 500 m - Daily observations
   - Uses 500 m resolution satellite imagery data obtained from the Moderate Resolution Imaging Spectroradiometer (MODIS), onboard the multi-national Terra scientific research satellite operated by the National Aeronautics and Space Administration (NASA).
   - The MODIS Reflectance MCD43A4 product provides 500-meter reflectance data adjusted using a bidirectional reflectance distribution function (BRDF) to model the values as if they were taken from nadir view.
   - The MCD43A4 product is a level-3 gridded data set in Sinusoidal projection and is provided daily.
3. PPI_3: NDVI MODIS 250 m - 8 Day Observations
   - Similar to "NDVI MODIS 500 m - Daily observations" except its satellite imagery data has 250 m resolution, is observed once every 8 days, and uses MODIS Reflectance MOD09GQ product to provide 250-meter reflectance data, and is used in conjunction with the MOD09GA product and provides top of atmosphere corrected surface reflectances.

4. PPI_4: GNDVI MODIS 500 m - Daily observations
   - Uses 500 m resolution satellite imagery obtained from MODIS, onboard NASA.
   - This PPI uses the MCD43A4 product to provide 500-meter reflectance data adjusted using a bidirectional reflectance distribution function (BRDF) to model the green and near-infrared bands of the electromagnetic spectrum, and is provided each day.
5. PPI_5: GNDVI MODIS 250 m - 8 Day Observations
   - Uses 250 m resolution satellite imagery data based on the MOD09GQ product in conjunction with the MOD09GA product, and provides top of atmosphere corrected surface reflectances.
6. PPI_6: EVI MODIS 500m - Daily observations
   - Uses the MCD43A4 product to provide 500-meter reflectance data adjusted using a BRDF to model the blue, near-infrared, and red bands of the electromagnetic spectrum, and is provided each day.
7. PPI_7: EVI MODIS 250 m - 8 Day Observations
   - Uses 250 m resolution satellite imagery data based on the MOD09GQ product in conjunction with the MOD09GA product, and provides top of atmosphere corrected surface reflectances.
8. PPI_8: MSAVI2 MODIS 500m - Daily observations
   - Uses the MCD43A4 product to provide 500-meter reflectance data adjusted using a BRDF to model the near-infrared and red bands, and adjusts for the influence of the soil background, and is provided each day.
9. PPI_9: MSAVI2 MODIS 250 m - 8 Day Observations
   - Uses 250 m resolution satellite imagery data based on the MOD09GQ product in conjunction with the MOD09GA product, and provides top of atmosphere corrected surface reflectances.
10. PPI_{10}: OSAVI2 MODIS 500m - Daily observations
    - Uses the MCD43A4 product to provide 500-meter reflectance data adjusted using a BRDF to model the near-infrared and red bands, and is optimized for agricultural modelling and adjusts for the influence of the soil background.
11. PPI_{11}: OSAVI2 MODIS 250 m - 8 Day Observations
    - Uses 250 m resolution satellite imagery data based on the MOD09GQ product in conjunction with the MOD09GA product, and provides top of atmosphere corrected surface reflectances.

12. PPI_{12}: LAI MODIS 500m - Daily observations
    - Uses the MOD15A2H version 6 MODIS Level 4 product to provide Leaf Area Index (LAI) for an 8-day composite data set with 500-meter pixel size.
    - LAI is defined as the one-sided green leaf area per unit ground area in broadleaf canopies and as one-half the total needle surface area per unit ground area in coniferous canopies.
13. PPI_{13}: FPAR MODIS 500m - Daily observations
    - Uses the MOD15A2H version 6 MODIS Level 4 product to provide Fraction of Incident Photosynthetically Active Radiation (FPAR) for an 8-day composite data set with 500-meter pixel size.
    - FPAR is defined as the fraction of incident photosynthetically active radiation (400-700nm) absorbed by the green elements of a vegetation canopy.
14. PPI_{14}: Accumulated Precipitation (AccPcpn) Weather Station.
    - This is index is constructed based on the daily precipitation.
    - Daily precipitation data is summed to create an index of accumulated values, which are matched to the ground truth forage yield data based on selecting the weather station geographically closest to the location of the test site.
15. FPI_{15}: Heat Degree Days (HDD) Weather Station.
    - Daily temperature data is used to construct a HDD index.
    - The index considers when the daily mean temperature falls below 65 °F. The daily mean temperature is found by adding together the high and low temperature for the day and dividing by two. When the mean temperature is above 65 °F, the HDD total is zero. If the mean temperature is below 65 °F, the HDD amount is the difference between 65 °F and the mean temperature.
16. FPI_{16}: Cooling Degree Days (CDD) Weather Station.
    - The CDD index is related to the HDD index in that it uses the daily temperature data.
    - The index considers when the number of degrees that a day's average temperature is above 65 °F, which is the temperature above which buildings may need to be cooled.