FEDERAL UNIVERSITY OF PELOTAS - UFPEL
School of Business Administration and Tourism
Eliseu Maciel School of Agronomy
Postgraduate Program in Territorial Development and Agribusiness Systems

Dissertation

Area-yield insurance feasibility in the context of remote sensing analysis

Roberto Mattes Horn

Pelotas, 2021.

ROBERTO MATTES HORN

Area-yield insurance feasibility in the context of remote sensing analysis

Dissertation presented to Postgraduate Program in Territorial Development and Agribusiness Systems from Federal University of Pelotas, as a partial requirement to Master Degree in Territorial Development and Agribusiness Systems.

Advisor: Rogério Costa Campos

Pelotas, 2021.

Universidade Federal de Pelotas / Sistema de Bibliotecas
Catalogação na Publicação

H813a Horn, Roberto Mattes

Area-yield insurance feasibility in the context of remote sensing analysis / Roberto Mattes Horn ; Rogério Costa Campos, orientador. — Pelotas, 2021.

40 f. : il.

Dissertação (Mestrado) — Programa de Pós-Graduação em Desenvolvimento territorial e sistemas agroindustriais, Centro de Ciências Socio-Organizacionais, Universidade Federal de Pelotas, 2021.

1. Cluster analysis. 2. Yield prediction. 3. Soybean. 4. Spacetime yield correlation. 5. Insurance utility optimization. I. Campos, Rogério Costa, orient. II. Título.

CDD : 658

Elaborada por Aline Herbstrith Batista CRB: 10/1737

ROBERTO MATTES HORN

Area-yield insurance feasibility in the context of remote sensing analysis

Dissertation presented as a partial requirement to Master Degree in Territorial Development and Agribusiness Systems, Postgraduate Program in Territorial Development and Agribusiness Systems from Federal University of Pelotas

Defense date: 14/04/2021

Examination board:

Prof. Dr. José Marinaldo Gleriani
Doctorate in: Remote Sensing, Instituto Nacional de Pesquisas Espaciais, INPE, Brasil

Prof. Dr. Mario Duarte Canever
Doctorate in: Administration - Emphasis on Agribusiness, Wageningen University, WAU, Netherlands

Prof. Dr. Rogério Costa Campos (Advisor)
Doctorate in: Remote Sensing, Instituto Nacional de Pesquisas Espaciais, INPE, Brasil

I dedicate this work to my
parents, my brother and sister.

Acknowledgments

To my family, that supported me in my journey and were always available whenever necessary.

To Rogério, my advisor, who is guiding me since the start of my graduation and made the path up to this point possible.

To the Graduated Program in Territorial Development Professors, Staff and Classmates for the lessons and good moments.

To UFPEL for the opportunity to develop my academic career.

To CAPES, which granted me a scholarship, providing the much-needed financial support for this work to happen.

Thank you.

Abstract

The conventional insurance was not effective on promoting massive insurance uptake in developing countries like Brazil. Its well-known operational problems and the lack of information on covariant risk prevent the insurers to remove risk from the system and so operate with risk independent portfolios. Since farmers do not find the expected utility in the conventional insurance, less than 15% of the crop area is currently insured in the country. We evaluated the synergism between actuary and remote sensing tools to outline an alternative insurance product based on area-yield concept. Empirical analyses were performed over 10 years of soybean growth inside a continuous region encompassing 54 counties at the North of Rio Grande do Sul state in Brazil. We evaluate the feasibility of insurance companies to offer a soybean crop area-yield insurance as the actuarial/operational flow is supported by yield correlation assessment from remote sensing data analysis. Results suggest the area-yield based insurance find a new attractive operational applying context as remote sensing provides yield inputs in high spatial and temporal resolutions. It was showed the systemic risk can be successfully managed through the space-time clustering of the estimated yields from soybean areas and it is a better strategy than the structural correlation from the less meaningful political county limits. The actuarial simulations of producer utility and premium loading showed that using per cluster correlation structure to set insurance parameters outperforms both using per county correlations structure and using no insurance. In contrast to county political limits, the extracted utility rose 39% with clusters correlation structure.

Keywords: Cluster analysis. Yield prediction. Soybean. Spacetime yield correlation. Insurance utility optimization.

Resumo

O seguro convencional não foi eficaz em promover a adoção massiva de seguros em países em desenvolvimento como o Brasil. Seus conhecidos problemas operacionais e a falta de informações sobre o risco covariante impedem as seguradoras de retirarem o risco do sistema e operarem com carteiras independentes de risco. Como os agricultores não encontram a utilidade esperada no seguro convencional, menos de 15% da área de cultivo está segurada atualmente no país. Avaliamos o sinergismo entre as ferramentas atuarias e de sensoriamento remoto para delinear um produto de seguro alternativo com base no conceito de rendimento por área. As análises empíricas foram realizadas ao longo de 10 anos de cultivo de soja em uma região contínua abrangendo 54 municípios no norte do estado do Rio Grande do Sul. Avaliamos a viabilidade das seguradoras em oferecer um seguro de produtividade da área de cultivo de soja, uma vez que o fluxo atuarial / operacional é apoiado pela avaliação da correlação de produtividade a partir da análise de dados de sensoriamento remoto. Os resultados sugerem que o seguro baseado em rendimento de área encontra um novo contexto de aplicação operacional atraente, pois o sensoriamento remoto fornece dados de rendimento em altas resoluções espaciais e temporais. Foi demonstrado que o risco sistêmico pode ser gerenciado com sucesso por meio do agrupamento espaço-temporal dos rendimentos estimados das áreas de soja e é uma estratégia melhor do que usar a correlação estrutural dos limites políticos de municípios, menos significativos. As simulações atuariais da utilidade do produtor e carregamento de prêmio mostraram que o uso da estrutura de correlação por cluster para definir parâmetros de seguro supera o desempenho tanto usando a estrutura de correlações por município quanto sem seguro. Em contraste com os limites políticos municipais, a utilidade extraída aumentou 39% com a estrutura de agrupamentos por correlação.

Palavras-chave: Análise de agrupamentos. Estimação produtividade. Soja. Correlação de rendimento espaço-temporal. Otimização da utilidade do seguro.

Figure List

Table List

Summary

1 Introduction

Agricultural insurance demand accurate risk evaluation along the process by which economic sectors will take the risk from producers. However, the lack of effective tools to classify risk level increases operational costs, prevents the long-term investments by the private sector and consequently turns products unsuitable for farmers. Conventional rural insurance products set the pricing of individual contracts on historical regional yields and indemnities are paid after a field verification of losses. It increases operational costs of the insurance and favors information asymmetries, moral hazard (where farmers have incentive to let their crops fail in order to receive a payout), and adverse selection (where those farmers less skilled at farming purchase the insurance, resulting in higher premium levels and more frequent payouts) (Miranda, 1991; Glauber, 2004; Ozaki, 2008). They also offer coverage against adverse weather effects in a generic way but weather information is not explicitly included at the rating or indemnity steps. Coverage is also limited to the availability of public subsidies, whose amount is variable from year to year (Glauber, 2004).

Traditional subsidized farm yield insurance suffers to achieve equity of premium and indemnities. When the subsidy is not high enough, only the more risk-exposed farmers tend to join the insurance programs. In this situation, equity of premiums and indemnities cannot be achieved with reasonable premium price. When the cost of insurance is diminished with an increase of subsidies, the farmers less prone to receive indemnities join the policies. This grants a possibility of equity but also makes the government subsidy marginal cost per hectare increase exponentially. This may have relation to the ignored systemic risk incorporated in these policies and can transfer the premium costs from the more to the less risk-exposed farmers as observed in USA (Glauber, 2004).

Risk assessment is the point at issue when economically sustainable crop insurance contracts are designed. Specifically, in crop insurance, due to correlated crop yield loss, companies are exposed to undiversifiable systemic risk over their region of actuation. Because crop default is a systemic risk, crop insurance market has not been actuarially efficient to deal with the risk exposure of their portfolios and the government ends up absorbing it (Miranda, 1997). Alternative methods to deal with systemic or undiversifiable risk have being investigated, but to be offered they all demand a reliable risk assessment tools (Miranda and Glauber, 1997; Carter, 2007).

Well-designed risk assessment products rely on data availability and quality to be used as a proxy to fluctuations in production (Miquelluti, 2019). It will be all accomplished if accurate risk scores can be delivered for the workflow and can be used as structured and pragmatic way of measuring risk exposure. In the last instance, risk classifying should be on dashboards to be used along with information that is readily available from individual contract at the farming level.

The interest for index-based weather insurance schemes is growing recently as they are pointed out as a promising risk transfer tool against weather related risks (e.g. Hellmuth *et al.*, 2009; Afriyie-Kraft *et al*., 2020; Fonta *et al*., 2018). These programs differ from traditional insurance in that payouts are directly tied to weather events rather than crop failure, which minimizes problems arising from asymmetric information (basis risk) and reduces the overhead cost of insurance as there is no need for in-field assessment of damage. Index-based weather insurance exploits the fact that weather observations can be used as proxies for crop losses. Weather observations obtained from meteorological services are used to determine payouts. Once a predefined threshold (trigger) of the underlying weather index has been reached during a specified time period, the contract starts to pay out.

Weather-based index insurance protects against specific events or risks, such as rainfall deficits and damage temperatures. Thus, it removes one or more production risks, but does not account for the crop loss itself which results from the whole local field (soil, plant, atmosphere, and technology) production system.

Since the index insurance only covers specific risks, it only provides partial protection and is therefore only one part of a complete risk management/adaptation package (Osgood *et al*., 2007). The popularity of this kind of insurance has led to the advent of a number of pilot projects, especially in lower income countries (Skees, 2008). However, its success strongly depends on a careful design under local conditions that takes into consideration aspects related to potential demand, identification and proper modeling of the best sources of weather information, suitable pricing, and risk assessment methodologies among other factors.

Although weather-based index insurance became attractive due to the lowering of operational costs, area-yield based insurance is more effective on minimizing the basis risk and making the farmers willing to pay by insurance product. In area-yield insurance, indemnities and premiums are based on a spatially localized yield within a previously defined area in which the yield is aggregated, instead on producer's individual yield. Indemnities paid to the individual producers equals the difference in aggregate yield and some agreed critical trigger value relative to the area (aggregated group) historical average yield. All producers in the defined area receive the same indemnity per insured acre regardless his actual crop yield and pay the same premium rate (Miranda, 1991).

Conversely to traditional crop insurance, in area-yield based contracts the insurer and farmer have the same amount of information about the aggregate yield, which goes in favor of the reduction of basis risk and adverse selection. It is also attractive by the fact of that a single farm cannot influence average area-yield effectively and its specific risk is not transferred to the system. Under area-yield based contracts, the farmer will aim his potential productivity and consequently the moral hazard is mitigated (Miranda and Glauber, 1997).

The presence of alternative methods (e.g. future and option markets) to compensate for yield/price fluctuation between sowing and harvest increase the area--yield insurance effectiveness in comparison to the traditional farm level insurance (Wang *et al*., 1998). This effect may be observed because crop insurance incorporates, in the lack of alternative methods to secure farm products price, a mechanism to ensure a higher payment based on expected price. When this mechanism is substituted by other means, conventional farm level insurance loses some competitiveness in comparison to area-yield insurance.

Area-yield insurance contracts stems from the idea that the productivity observed in a farm is related to the aggregated yields according to the Equation 1 (Miranda, 1991).

$$y_i = \mu_i^y + \beta_i^y (\bar{y} - \mu^y) + \varepsilon_i^y \tag{1}$$

y_i is the observed farm yield; $\bar{y}$ is the aggregate productivity observed in the area; μ_i^y is farmer expected (or average) yield; μ^y is the expected area-yield; β_i^y is a farmer-specific parameter giving the average or systematic relationship between the individual farmer's yield and the overall yield variation. ε_i^y represents idiosyncratic yield variation related to farmer-specific agrotechniques; β_i^y and ε_i^y have an inverse interpretation in the applicability of area-yield insurance for a target farmer.

Farmers with grater β_i^y access more benefits when contracting area-yield insurance. The reverse applies to ε_i^y. By definition, the majority of producers will have a positive β_i^y in the aggregate yield, but negative values are present and insurable as well (Miranda, 1991).

With the basis risk calculated for farmers it is possible to know what loading a farm level insurance premium must have to perform worse than the area-yield option. The loading refers to additional delivery, adverse selection, and moral hazard costs (Wang *et al*., 1998).

The implementation of area-yield based insurance in developing and emerging economy countries is constrained by the limited availability of detailed historical yield records longer than 10 years. In developing countries, historic data for crop productivity and weather are not easily available, if available at all (Ozaki, 2008). To mitigate the lack of data for risk prospection, remote sensing can be utilized as an alternative data source, partially supplementing the missing information. From the advent of satellite data, several variables became available to depict the crop status with high spatial and temporal frequency and coverage. By integrating satellite surface data and crop models one can feed crop-yield models in a per field/farming spatial scale and so regulate several insurance operational processes (eg., premium rating, subscription, fraud analysis, payouts) from the perspective of spatially disaggregated (area-based) yield realization.

Remote sensing can be used in several ways to aid insurance policies and companies. De Leeuw (2014) point out four of them: (i) making insurance affordable to low income households; (ii) reducing fraud, moral hazard, and adverse selection; (iii) eliminating the burden of costly verification of claims on-the-ground, and (iv) enabling faster and cheaper payouts to the insured.

It is usual to take arbitrary political limits to set contract yield aggregation groups and coverage level, however political limits are not necessarily compromised with yield spatial variation. Remote sensing allows to survey yield closed-related biophysical variables at high spatial resolution and so can support area-yield insurance by classifying farmers in more realistic spatially correlated groups than political limits do.

In this regard, remote sensing has a great potential to aid the start and expansion of area-yield insurance policies. One of the main aspects of competitiveness of area-yield crop insurance in contrast individual yield crop insurance is determined by the correlation of groups of farmers selected to measure aggregated yield (Wang *et al*., 1998). On the area-yield Insurance approach the number of sufficiently correlated farms ($\beta_i^y \sim 1$) (and so suitable for taking area-yield Insurance) can be maximized through yield prediction from remote sensing.

Previous works focus in compare area-yield insurance with farm level insurance (Barnett *et al*.,2005; Wang *et al*. 1998; Glauber, 2004; Schinithkey *et al*., 2002; Dismukes *et al*., 2013). In this paper we, instead, used remote sensing to try an innovative methodology to organize and classify farm groups of aggregation in order to improve the benefits of area-yield insurance.

1.1 Objectives

The overall objective is to evaluate through yield time series estimation the feasibility of remote sensing modeling to support area-yield insurance as a risk reduction tool for soybean when looking at capability of disaggregating spatial risk of yield loss at field scale. It was accomplished by checking for (i) data availability and model suitability when exploiting orbital remote sensing data to produce per field soybean yield estimates over a period of 10 years and (ii) actuarially compare area-yield insurance performance when correlations are taken from clustered yield against county limits criterion of yield aggregation.

2 Methodology

2.1 Study area

It is proposed to verify the feasibility of Area-Yield Based Insurance over a continuous region that encompass 54 municipalities in Rio Grande do Sul state (Figure 1). The region accounts for about 18% (~one million of hectares) of the total soybean cultivated area in the state. Soybean fields were mapped with 30 m of spatial resolution from Landsat 7 and 8 images acquired along the crop cycle from October - 2016 to March -2017. The mapping method took images time series as input in an automatic supervised classifier based on multinomial logistic function fitting. At final step the soybean map was validated by an independent procedure executed by experts on visual interpretation of spectral crop patterns. Table 1 presents the confusion matrix originated from the validation procedure.

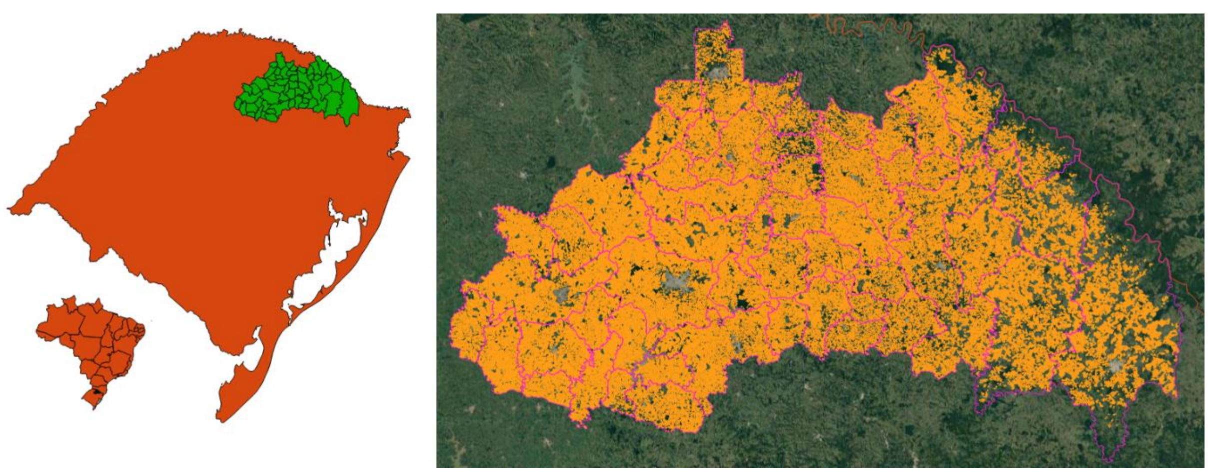

Figure 1 - Study area and soybean map.

Table 1 - Confusion matrix results from validation procedure.

		Predicted	
		Soybean	Other
Real	Soybean	144	4
	Other	5	145

2.2 Database modeling and satellite integration

Several variables became available from remote sensing and can be used to feed crop yield models (Sun *et al*., 2019; Wang *et al*., 2018). There are different levels of integration between them and ground level weather data. For instance, empirical yield prediction models use linear regression directly to output yield estimations from vegetation indices (VI) (Bolton *et al*., 2013; Esquerdo *et al*., 2011). Models that are more sophisticated simulate the plant growth physiology response to environment variables (Letort *et al*., 2007; Boote *et al*., 2021.; Eze *et al*., 2020). Many of the crop growth model variables are acquired or can be estimated with remote sensing tools (Setiono *et al*., 2018).

Here we are going to use surface reflectance products from MODIS sensor on Aqua and Terra platform acquired over the period from 2006-01-01 to 2017-12-31. Surface reflectance derived from 8-day composite MODIS products outline vegetation index time series to be linked to soybean phenology. Vegetation index calculated from different platforms were gathered to increase data availability as VI maximum values are preserved. The time series was denoised in an iterative mode under the rules of Whittaker filter (Vuolo *et al*., 2011).

Lastly, a stepwise logistic function was fit to the filtered VI time series and applied to estimate the soybean phenology step. Local maximum and minimum from the third derivative of fitted function located in the time series points were associated with crop emergence, reproductive, maturity, and harvest phenology (Zhang *et al*., 2003, 2006). Figure 2 present an example of time series from the study area and its phenology determined with this methodology.

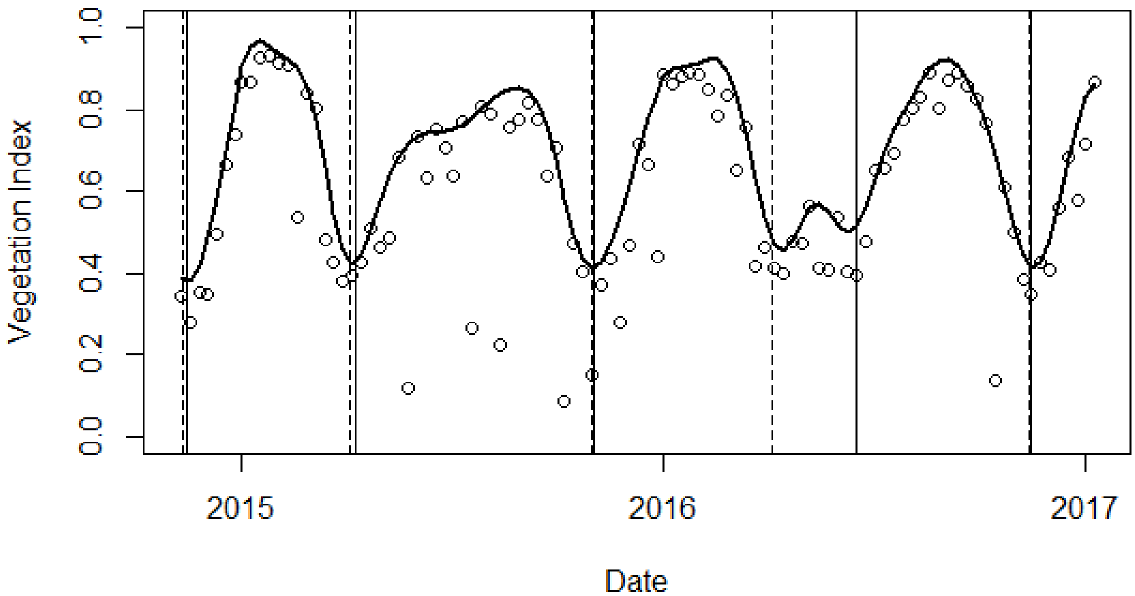

Figure 2 – Graphic example of vegetation index time series from study area. Points are the original vegetation index data from MODIS Terra platform. Solid line is the filtered time series. Vertical lines are the phenology extracted by the logistic fitting. Solid and dashed lines represent crop cycle start and end respectively

2.3 Analysis

Soybean yield estimation was performed by fitting an empirical model at the MODIS pixel (250m) level. The model was adjusted and crossvalidated with 150 yield datapoints from 30 farms in Paraná state over 5 years in the period from 2015 to 2019. The model explanatory variables were taken from the VI time series values from 24 to 96 days after the soybean growth start detection. The area under the VI curve was integrated to predict yield (Esquerdo *et al*., 2011). We adapted the proposed model from county level productivity data to the field scale due to calibration data availability and the MODIS spatial resolution being adequate in comparison with most of soybean field sizes. A full dataset crossvalidation showed the model performed with root mean squared error of 9 Sc/ha.

The fitted yield model was then used to construct 10 years series of estimate yields for the soybean mapped area in the Rio Grande do Sul state. The series of estimated yield was still subjected to a spatial process by which we could integrate the soybean map (30 meters of resolution) with the estimated yield (250 meters of resolution). The final dataset corresponded to 10 years of yields at farm field level (30m spatial resolution) and was used to evaluate correlated risk reduction and idiosyncratic risk. It was evaluated under different hypothetical situations regarding the value at risk and the stochastic dominance. The analysis ends up with an economic assessment of advantages/disadvantages of adopting area-yield insurance strategies by varying the criterion to perform the portfolio spatial aggregation.

We measured the willing to pay with a loading multiplying the premium price in order get the equity of utility extracted from soybean yield with and without the insurance strategies evaluated in the study. This was similar to what previous studies have approached to compare farm yield insurance to area-yield insurance (Wang *et al*., 1998).

A recursive and obvious criterion to set insurance parameters (average yield, coverage level, etc,.) in the crop space is given by spatial aggregation according to county political limits. It is convenient once crop statistics are disclosed in per county (IBGE – Instituto Brasileiro de Geografia e Estatística) database. However, the spatial yield autocorrelation may not be well represented according this simple criterion.

Roel and Plant (2004a and 2004b) developed an exploratory method for organizing spatiotemporal data by organizing the data into clusters in space and time. In this study, we utilized this method for generating and analyzing spatiotemporal yield clusters. We apply this approach to our ten years of estimated yield data. The spatiotemporal method works by first forming clusters of the response variable in the data spacetime (each year) of the yield data. The analysis uses the k-means clustering algorithm. This cluster analysis with spatiotemporal data allows to identify factors underlying the observed pattern in space and time of the response variable. It means, the clusters must be "biophysically meaningful." One criterion for being biophysically meaningful is that the clusters will form groups of similar farm productivity and will bring potential benefits to the insurance application if it can improve the level of correlated producers participating of the same portfolio with specific set of parameters. Our analysis relies on the spatiotemporal cluster aggregation method to contrast with simple county political division.

The risk aversion for farmers is defined by the Constant Relative Risk Aversion (CRRA) utility function:

$$U(c_{it}) = \frac{c_{it}^{1-g}}{1-g}, \text{ for } g > 0, g \neq 1 \tag{2}$$

Where *c* is farmer income; *g* indicates the level of risk aversion with values increasing with higher aversion; *i* varies among the farmers and *t* indicates the crop year. The parameter *g* used will be 2, which represent intermediate levels of risk aversion as observed in US (Wang *et al.*, 1998).

To maximize the utility function for each class of risk aversion, a monetary value that represents the willing to pay for premium (in order to reach a steady income from crop yields) will be calculated. The monetary value the farmer pays in order to maximize its utility function (given its class of risk aversion) is the willing to pay. The optimization was performed by setting 20% of payment frequency along the 10 years of the analysis. Consequently, every cluster (and county) was assigned to have the trigger (yield coverage level) that makes indemnities payments occur two out of ten years. It generates farmer groups with different levels of insurance trigger yields. For contract optimizations, we used fair premium, i.e. total payments from farmers is equal to the sum of indemnity delivered.

Following, we compared the area-yield insurance performance when the aggregate yields are taken from the counties political division against to the clustered farms dictated by the correlation of estimated yield. We measured the difference in total willing to pay as sum of monetary value, converted to soybean yield, directed to the acquisition of area-yield insurance with both methods of aggregation.

Willing to pay, in addition, was given as the premium load that equalizes the utility extracted from the yield with and without insurance when a payment frequency set to 20%. After the optimization is ready and premiums and indemnities are defined, we calculated both utilities, with and without insurance. After that, we verify if the utility with insurance is greater or smaller than without insurance and if it is greater, we multiply the premium with a coefficient > 1 until we equalize it with the utility without insurance. This coefficient represents the likely margin for the insurer cover its operational costs, makes profit, and not be paid back as indemnities. If the utility with insurance is smaller than without it, we will multiply it by a coefficient < 1 as to equalize it with the utility without insurance. This will be in order to measure how worse the insurance contract is in both yield aggregation methodologies than with no insurance coverage. Total sum of indemnities paid in the 10 years will be measured and compared between both methodologies. Figure 3 presents the main steps in the analysis.

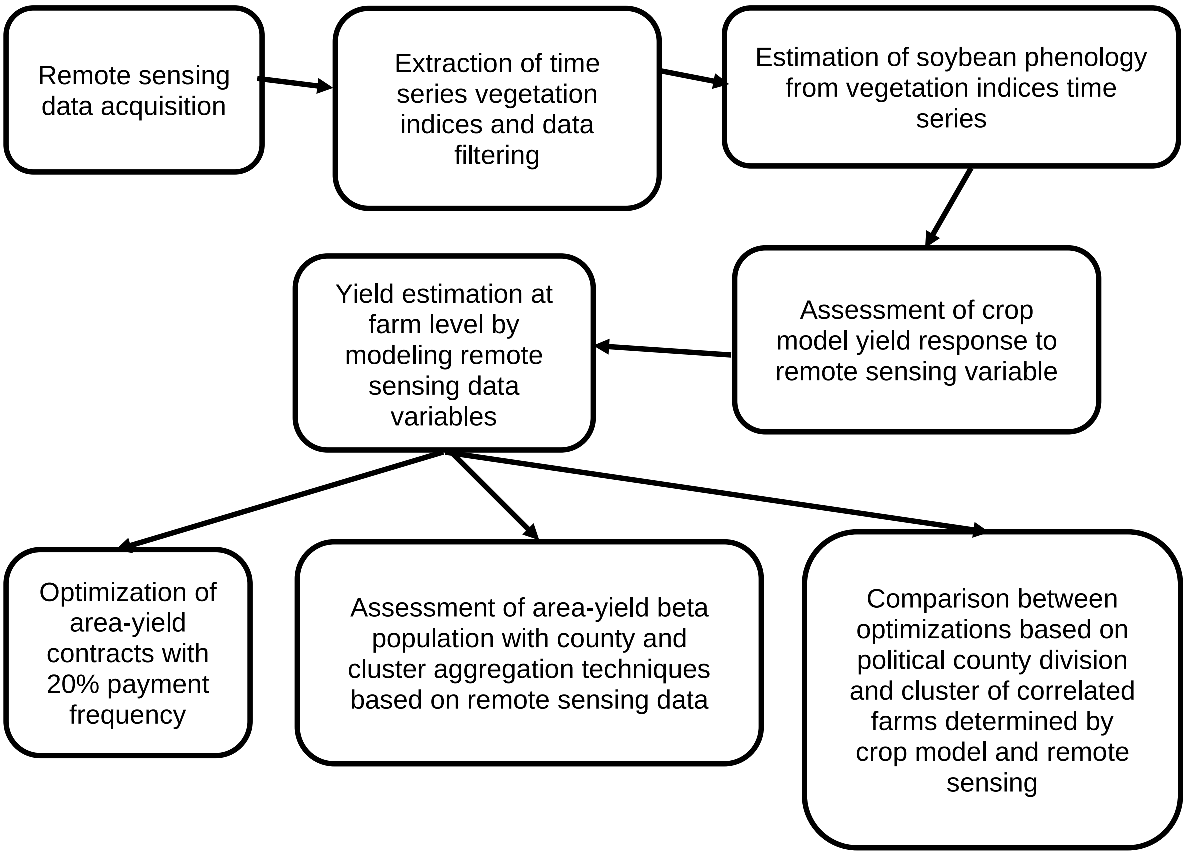

Figure 3 – Flowchart with the main analysis steps.

3 Results

The fitted yield model incorporates the detected phenology and indices derived from VI time series and estimates the yield with an acceptable and useful accuracy (RMSE = 588 Kg/ha). Figure 4 shows the mean estimated yield over the 10 years within the geographic space of the analysis. Figure 5 shows the standard deviation of yields for the same area and period. The mean is from the soybean summer season harvests that were estimated from 2007 to 2016.

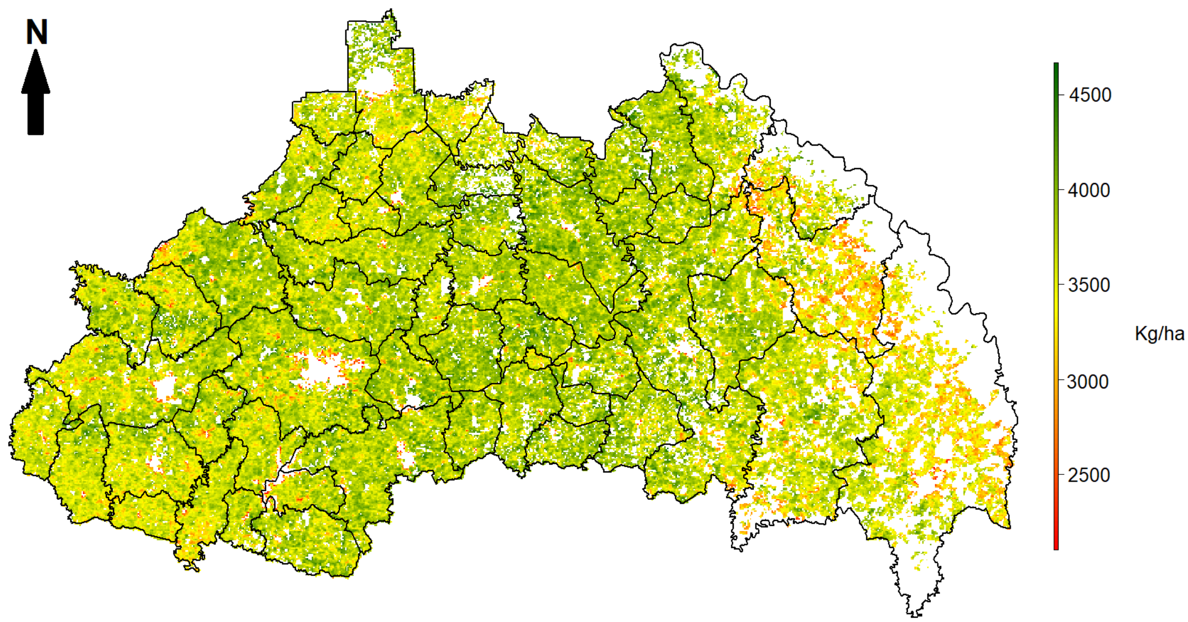

Figure 4 - Map of the mean yield in the 10 years series (Kg/ha).

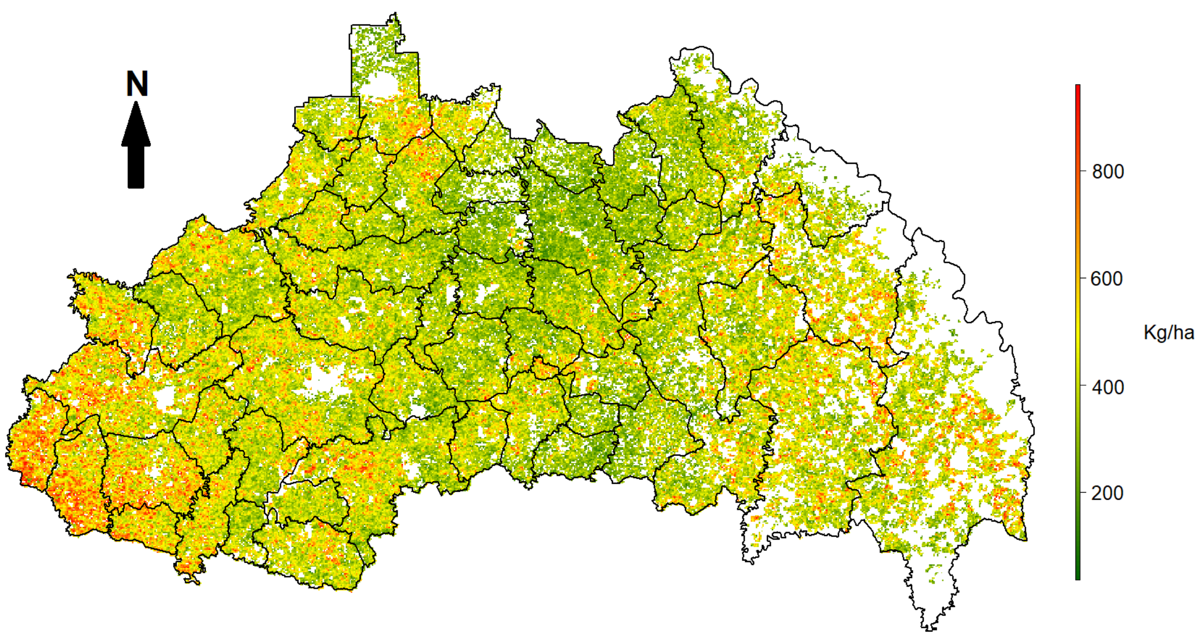

Figure 5 - Map of the standard deviation of yield in the 10 years of the series (Kg/ha).

We took the total yield variability within cluster to determine the number of clusters used in the analysis. We applied criteria of 50% of total squared sums to be due to cluster depicting. This criteria is chosen because of the nature of the exploratory analysis and lack of previous works on this issue. By analyzing the proportion of total variability within clusters (Figure 6) we observe that internal cluster variability clearly diminishes beyond eleven clusters.

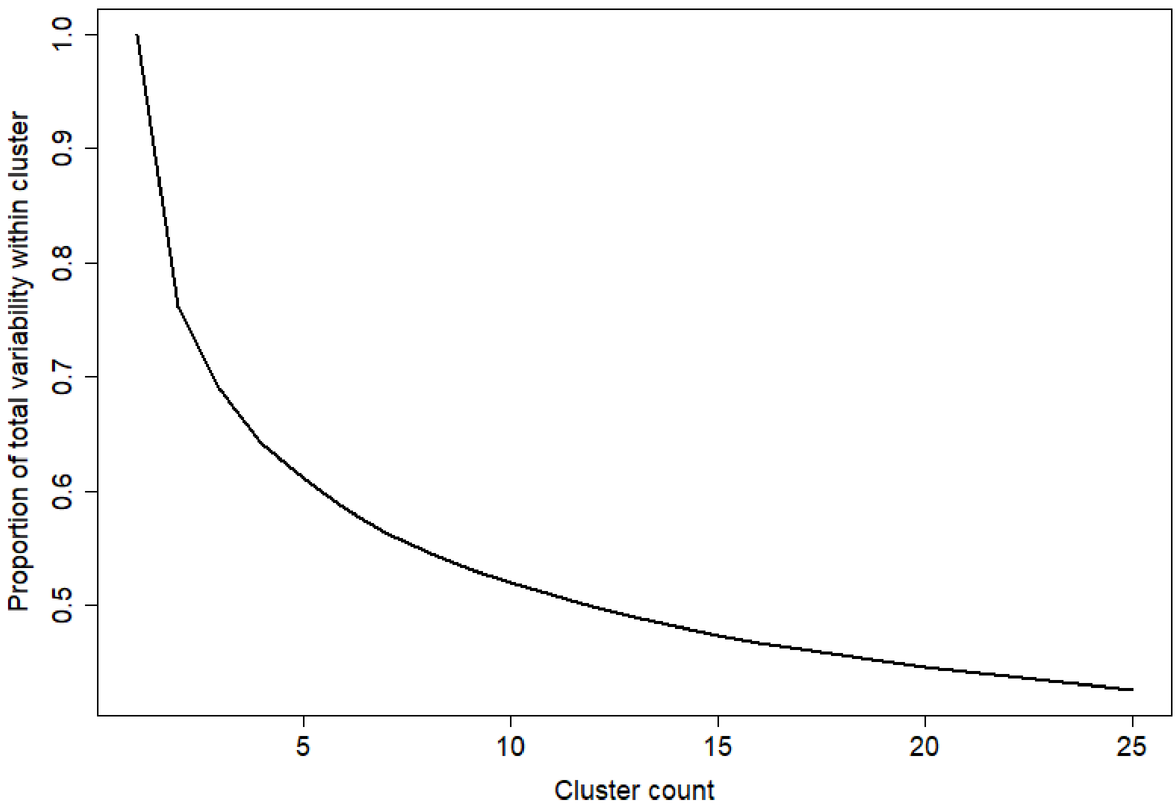

Figure 6 - Proportion of total variability within clusters and number of clusters used. In our analysis we use 11 clusters, which leaves about 50% of variability remaining inside them.

The resulting clusters have a pulverized spatial distribution over the study area. In figure 7 we can observe some predominant clusters forming in some areas but they surpass the counties political division. Figure 8 presents the clusters in context of the CAR (Cadastro Ambiental Rural) consolidated agriculture spatial database. It is possible to figure the CAR geocoded farming as belonging to its cluster and so having specific set of insurance parameters. It is an example to show that the methodology is readily to operate with the CAR database.

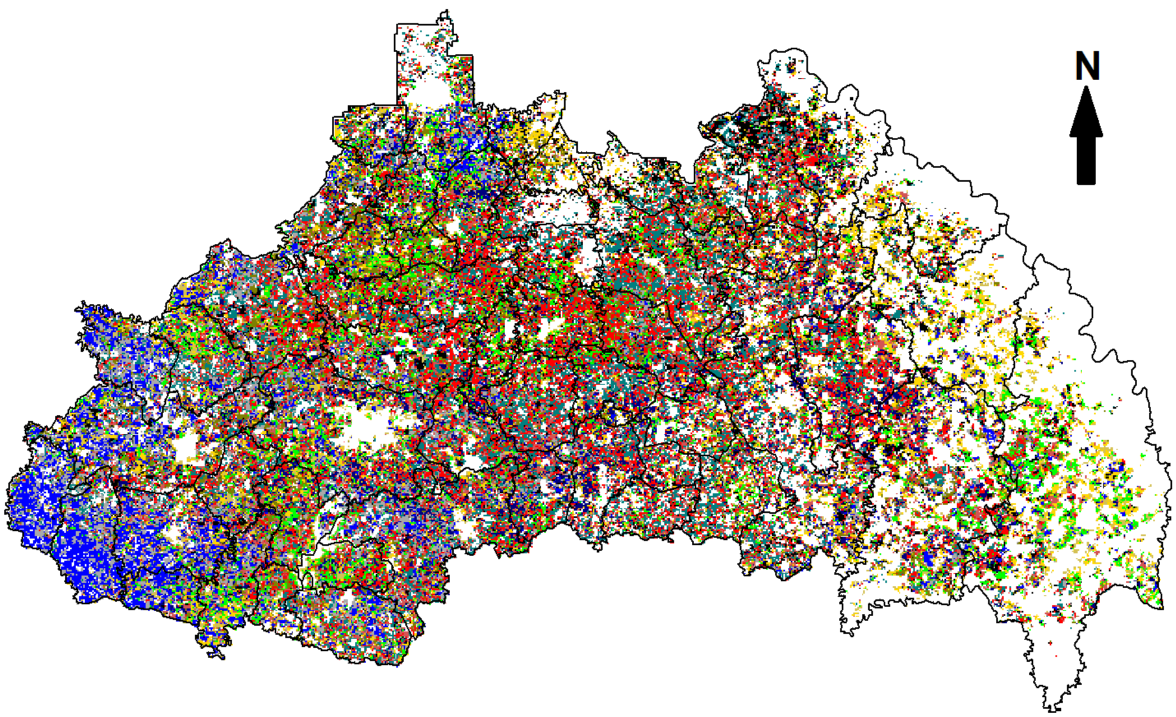

Figure 7 - Spatial disposition of the 11 clusters. Each color represents a different cluster.

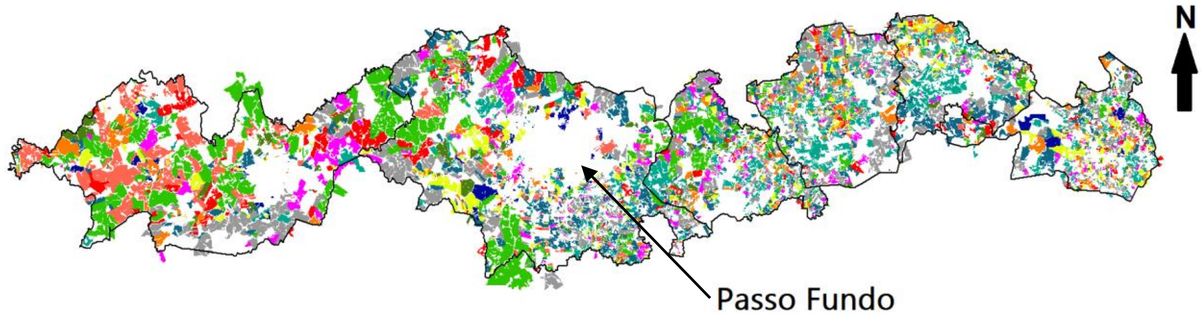

Figure 8 - Clusters spatially depicted according to CAR consolidated agriculture spatial database showed in 6 counties near Passo Fundo county. From left to right are the counties Carazinho, Passo Fundo, Mato Castelhano, Água Santa, Santa Cecília do Sul, and Caseiros. Crop fields delimited by CAR information.

By verifying the historic average productivity along the years on the aggregation groups (counties and cluster) we can see that two different counties clearly have a much more similar mean productivity than any two different clusters. Figure 9 and 10 show some of the more distinctive counties and cluster yield average time series. Although we are using only 11 clusters in comparison to 54 counties, the spectrum of different yields accounted by the clusters averages originates more opportunities for area-yield insurance than county aggregation does.

The change, from counties to cluster aggregation, in the distribution β coefficients (Equation 1) for all farms in the area is in agreement with what is punctuated above. The distribution of β depicts an evident mean shrinkage around 1 under clusters aggregation of yields in comparison to county aggregation (Figures 11 and 12). It goes in favor of the major requirement to apply area-yield based insurance by which the effectiveness of participants benefits increases as the individual yield became more correlated to the aggregated yield, $\beta_i \sim 1$. Table 2 presents some metrics to measure the effect of cluster aggregation in the beta distribution. The first quartile and third quartile, representing points that delimits 50% of the data, have increased by 0,11 and diminished by 0,10 respectively with the use of clusters.

Figure 9 - Time series of mean county soybean productivity for Água Santa, Colorado, Estação, Ibirapuitã, and Esmeralda in black, red, purple, green, and blue respectively.

To probe the effects of the beta distribution modification with clusters we choose to optimize the area-yield insurance contracts with 20% frequency of payment. Consequently, for every aggregation group (11 clusters or 54 counties) there will be two years that indemnities are paid. As every group has a different series of average soybean yield, groups have a different trigger yield to pay the indemnities. Table 3 shows statistics of those trigger yields as a proportion of the group average yield. We found the cluster aggregation does not change the mean trigger, but concentrates the trigger yields around the mean trigger value.

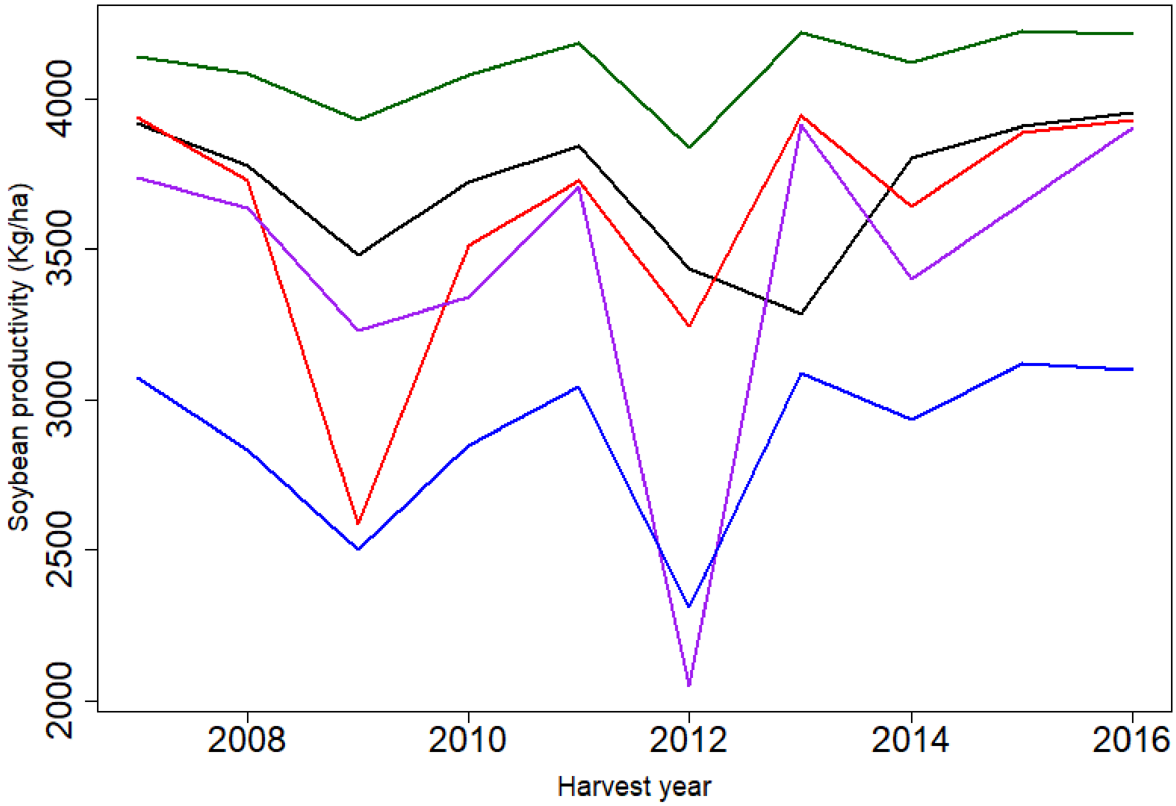

Figure 10 - Time series of mean soybean productivity for 5 of the 11 clusters.

The temporal distribution of indemnities paid by both methodologies of aggregation is shown in table 4. In the year 2012 a drought happened in the study region and productivities were overall lower. This caused 48,1% and 50% of the events of indemnities payment respectively for county and cluster aggregation. The remaining events of indemnities payments are somewhat more concentrated for county aggregation with 45.4% in 2009, 2.8% in 2010, and 0,9% in 2002, 2011 and 2014. For cluster aggregation indemnities payments were made 36.4% in 2009, and 4.5% in 2007, 2008, and 2013.

Figure 11 - Distribution of β resulting from county aggregation of yield.

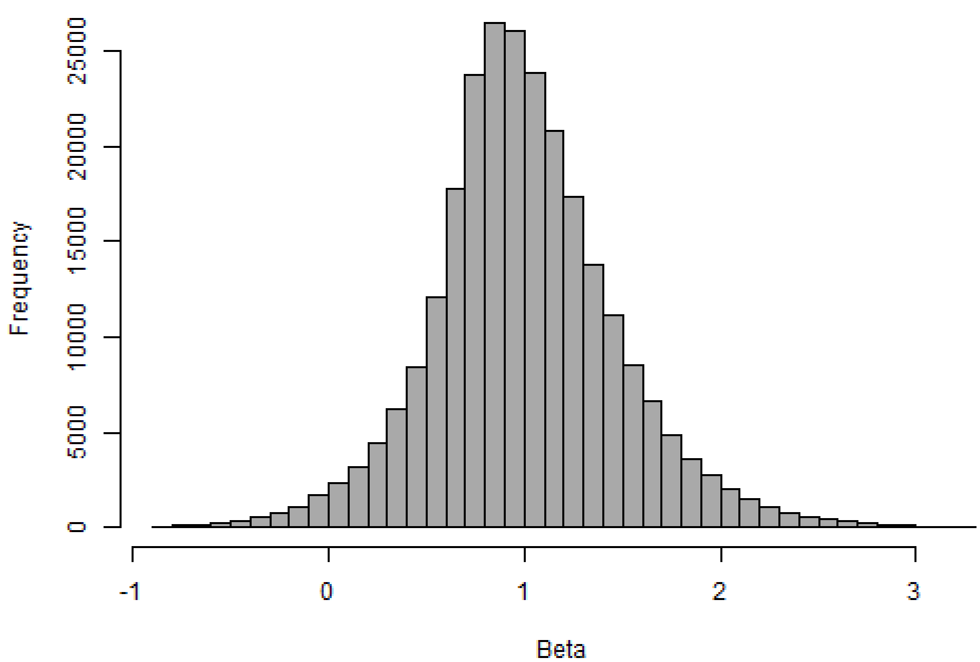

Figure 12 - Distribution of β resulting from the cluster yield aggregation method.

Table 2 - Statistics from beta distribution by county and cluster aggregation of productivity.

	Minimum	1º Quartile	Median	Mean	3º Quartile	Maximum
County	-3,30	0,60	0,97	1,00	1,37	9,30
Cluster	-2,50	0,71	0,97	1,00	1,27	4,17

Table 3 - Statistics of trigger yields by county political division and clusters as a proportion of mean productivity.

	Minimum	1º Quartile	Median	Mean	3º Quartile	Maximum
County	0,920	0,944	0,961	0,959	0,972	0,992
Cluster	0,930	0,937	0,955	0,952	0,959	0,986

Table 4 - Relative frequency of indemnities paid by year.

	Harvest year									
	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
County	0,000	0,009	0,454	0,028	0,009	0,481	0,009	0,009	0,000	0,000
Cluster	0,045	0,045	0,364	0,000	0,000	0,500	0,045	0,000	0,000	0,000

The effect of the area-yield insurance on farm income can be observed in Figures 13, 14 and 15. There was a reduction in the probability of income below 3000 Kg/ha when the producer contracts the area-yield insurance, it happens regardless the aggregation method utilized to produce the correlation structure. In relation to non-insured income, the reduction is more pronounced when contracts are parameterized using cluster aggregation of yields. Area-yield insurance increased the probability of incomes falling between 3000 – 4000 Kg/ha. As expected, the probabilities of high incomes above 4000 Kg/ha were reduced and the extreme yield incomes were shrunk towards the average yield. Cluster aggregation of yield was more efficient than the county aggregation in reducing probabilities of extreme incomes

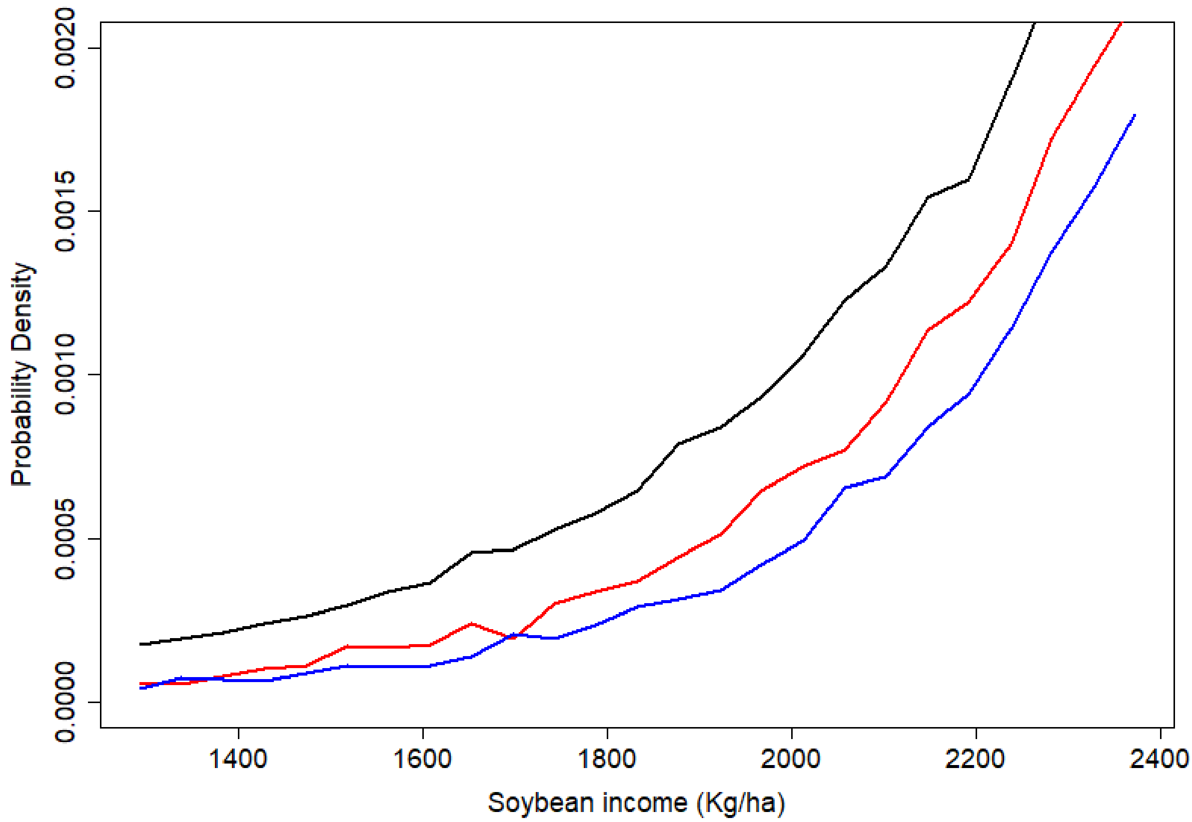

Figure 13 - Probability distribution of income. Black, red, and blue lines refer to no insurance, yield aggregation by county, and yield aggregation by cluster respectively.

Overall, cluster aggregation of yields increased the total indemnities paid to farmers in the period of analysis. Every single group created (11 clusters or 54 counties) has its series of average yields along the 10 years of analysis. This allow the total sum of indemnities paid along the 10 years period for each aggregation method to be calculated. Table 5 displays the differences of the indemnities paid in the two aggregation methods. The cluster aggregation increased the minimum paid by almost seven times. At the mean values the difference is +20,9%. Premium loads had minor improvement it increased 0,03 on average (table 6).

The combined factors promoted by cluster aggregation such as, gathering similar productivity farms, concentration of betas around 1, increasing of indemnities payment, and slightly increasing in premium loads has the effect of significantly improve the performance of area-yield insurance. With the county aggregation of yield, the excess in premium loading per year is a monetary value equal to 21,5 thousand tons of soybean. With cluster aggregation it becomes a value of 29,9 thousand tons of soybean, an increase of 39% in value.

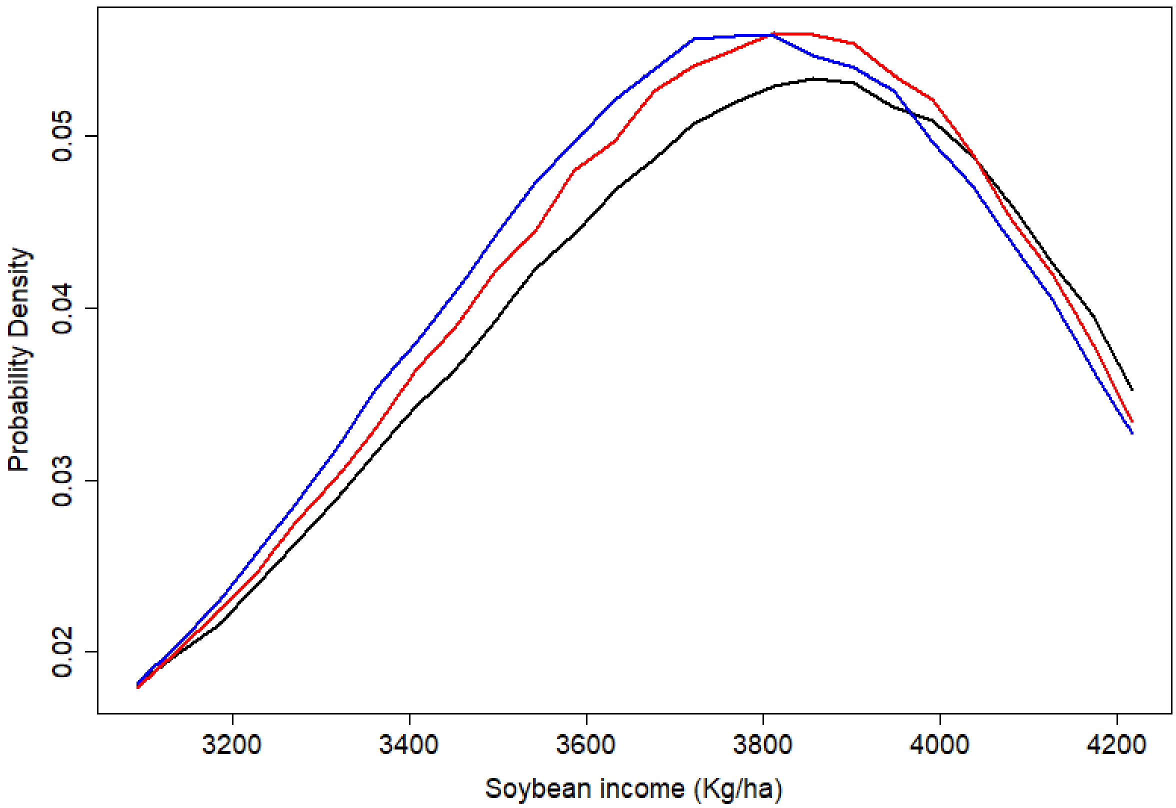

Figure 14 - Probability distribution of income. Lines are no insurance, yield aggregation by county, and yield aggregation by cluster in black, red, and blue respectively.

This rising value is due to k-mean clustering creating groups of correlated farmers. The constructed groups better capture the structure of yield variability than the arbitrary political county limits. K-mean clustering improves models predictions by classifying the desired variable in more coherent groups that are more homogeneous (Lu *et al*. 2017).

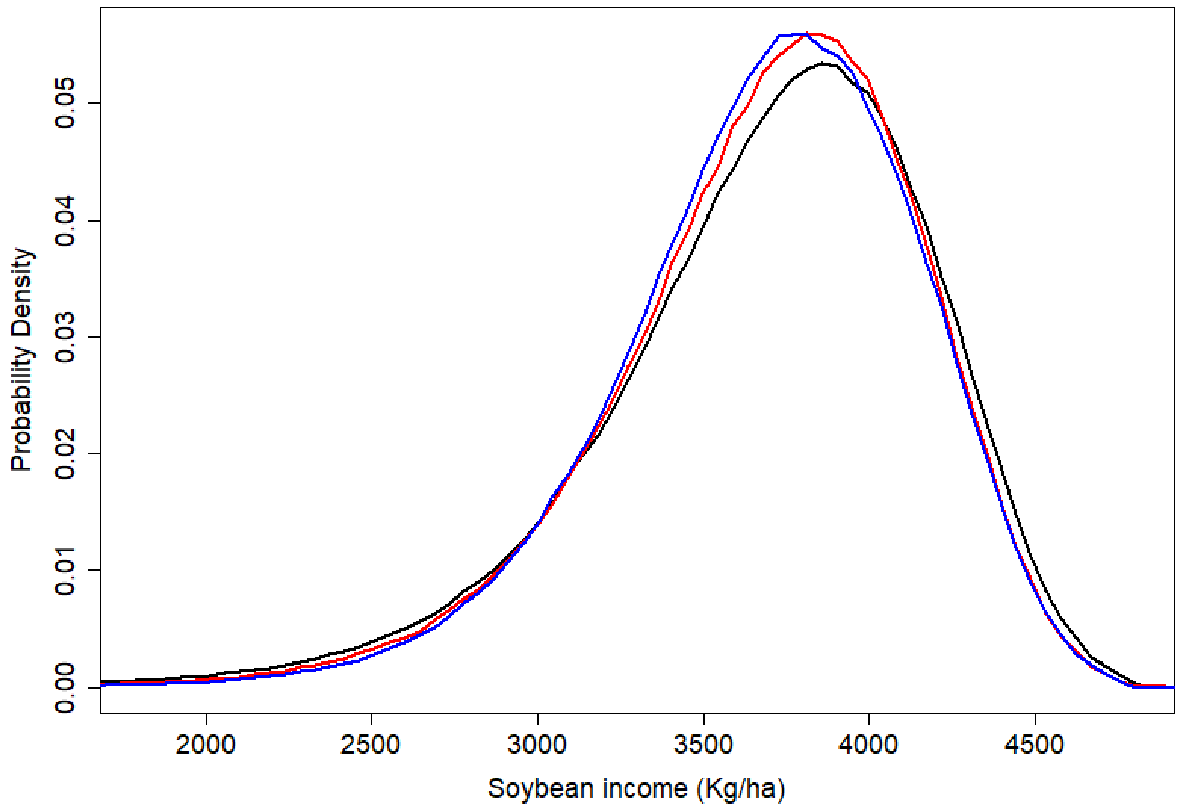

Figure 15 - Probability distribution of income. Black, red, and blue lines refer to no insurance, yield aggregation by county, and yield aggregation by cluster respectively.

Table 5 - Statistics of total indemnities paid in 10 years by county and cluster aggregation in Kg/ha.

	Minimum	1º Quartile	Median	Mean	3º Quartile	Maximum
County	32,4	294	442,2	464,4	575,4	1274,4
Cluster	222,6	306	401,4	561,6	763,8	1362

Table 6 - Statistics of premium load by county and cluster aggregation of productivity.

	Minimum	1º Quartile	Median	Mean	3º Quartile	Maximum
County	0,34	1,05	1,15	1,23	1,31	5,01
Cluster	0,36	1,08	1,19	1,26	1,34	4,68

Another benefit that originates from the use of k-means is the improvement on identification and compartmentalization of systemic risk. The resulting clusters are much less correlated to each other as we can see when comparing the time series between clusters and counties (Figures 11 and 12). Consequently, the likelihood that several clusters of productivity originated by k-means will receive indemnities at the same time decreases in relation to aggregation by county. Studies such as those carried out by Stigler and Lobel (2020) comparing the choice of producers between conventional and area-yield insurance, should be applied to the reorganization of producers by k-mean to verify the real appeal of the proposed technique.

Area-yield insurance appears to have more appeal for farmers with low productivity realizations. In the left tail of the yield distribution, even low indemnity payment from the area-yield insurance corresponds to a larger proportion of farm productivity.

The proposed methodology helps to mitigate the issue of subsidy equity, it increases the insurer's margin, lowering the premium and thereby the cost of producers, making area-yield insurance more attractive for farmers less exposed to production risks. This lowers the government investment needed in insurance subsidy. The lower price of the premium obtained by the aggregation methodology increases the accessibility of insurance to producers, which may cause greater market penetration and a larger portion of farmers acquiring insurance. This will provide protection against impoverishment increasing financial stability and preventing negative income cycles (Benami *et al*., 2021).

Our results highlight the benefits from detaching from currently available productivity reports at the county level and integrate statistical models and remote sensing data to predict productivity. Productivity estimation based on models and remote sensing data is much less costly to achieve than reliable productivity reports at the field level. This applies to area-yield insurance that does not incorporate remote sensing data in addition to conventional farm level insurance that in every indemnity payment requires a field verification of losses.

Predicting productivity by remote sensing when applying area-yield insurance covers several risks at the same time compared to insurance based on a weather index that usually covers a single risk. For this to occur, models that predict productivity must include as many variables that interfere with productivity as possible (Benami *et al*. 2021). In this work we focus on an empirical model that related productivity to a remote sensing time series index that indirectly captures the biomass of the soybean crop in the field and, in our tests, we obtained a reasonable result. This model would not be adequate in specific situations of loss of productivity (for example, a pest attack directly on the grain in the soybean pod, without indirect effects on biomass). Future studies could better explore the adequacy of empirical models at the field level for the application of insurance with different types of yield loss that would originate indemnities.

As the proposed area-yield insurance pays groups from hundreds to thousands of farmers, punctual model yield prediction errors do not compromise its utilization, because prediction of the average productivity of the group is what guides the indemnities payments. Isolated yield estimation errors end up canceling out in all observations used to compute the group's average productivity (on unbiased statistical models with normal error distribution around the average). The punctual errors in estimating productivity may not be crucial in calculating the average productivity of the aggregation group, however, the impact on classification of producers by k-mean in their most correlated groups has yet to be known.

With the use of remote sensing data to predict yield, we were able to overcome the lack of productivity data at the farm level and reconstruct the productivity for a period of 10 years. We reconstructed the yield series for a static soybean culture map that represents the area in the 2016/2017 harvest.

The calibration data used was from a period and region that does not fully represent the place and period of model application, however, as it is an exploratory-empirical analysis and these interferences are equally represented in the two forms of aggregation tested, the limitations do not make the results obtained in the study inconclusive. Future studies can use more appropriate calibration data to gain higher representativeness of the area of interest. For instance, to classify the soybean culture to produce maps for each time harvest the model is applied.

On the lack of extensive farm level data, the reconstruction of the productivity series from remote sensing models is very useful. For the application of area-yield insurance, the only requirements are punctual reliable farm level yield data for calibration and a computational structure for acquisition of remote sensing data and model application.

For the comparison between the two productivity aggregation techniques we use 20% payment frequency criterion. This does not necessarily optimize contracts offered in the region in the best way. Other parameters can be used in replacement, for example, a specific average productivity or an indemnity sum over 10 years. Insurers must adjust their contracts to make them as attractive as possible to potential customers. Another parameter that should be explored and better adjusted is the *g* of the constant relative risk aversion function (CRRA, Equation 2) that we assume as 2.

4 Conclusions

It was shown that remote sensing could provide the data required to parameterize area-yield insurance over 54 counties in northern region of Rio Grande do Sul, Brazil. We applied the yield estimation model to a previously mapped soybean fields to show that data and model are available to produce results in high spatial and temporal resolutions. It makes possible to use many other strategies to set parameter of crop insurance. For instance, once the yield time series became available at field scale, we can operate crop insurance under an analytics steps to maximize the portfolio returns for both insurance companies and farmers.

We compared two methodologies to perform yield aggregation and could conclude that spatiotemporal clustering outperforms the county limits due to the more effective β shrinkage towards the mean value equals to 1.

The perception of insurance value was measured with the CRRA utility function to suggest that spatiotemporal clustering technique greatly improves the performance of area-yield insurance in comparison to using county political limits. Clusters improved the correlation between farmers and their respective groups of aggregation, decreased the probabilities of low income, increased the total indemnities paid to makes farmer more will to pay by crop insurance and better distributed indemnities along the crop harvest years. Adding, the maximum potential margin of insurance companies rose 39%.

References

AFRIYIE-KRAFT, Lydia; ZABEL, Astrid; DAMNYAG, Lawrence. Index-based weather insurance for perennial crops: A case study on insurance supply and demand for cocoa farmers in Ghana. **World Development Perspectives**, v. 20, p. 100237, 2020.

BARNETT, Barry J.; BLACK, J. Roy; HU, Yingyao; *et al*. Is Area Yield Insurance Competitive with Farm Yield Insurance? **Journal of Agricultural and Resource Economics**, v. 30, n. 2, p. 285–301, 2005.

BENAMI, Elinor; JIN, Zhenong; CARTER, Michael R.; *et al*. Uniting remote sensing, crop modelling and economics for agricultural risk management. **Nature Reviews Earth & Environment**, v. 2, n. 2, p. 140–159, 2021.

BOLTON, Douglas K.; FRIEDL, Mark A. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. **Agricultural and Forest Meteorology**, v. 173, p. 74–84, 2013.

BOOTE, K J; JONES, J W; HOOGENBOOM, G. Incorporating Realistic Trait Physiology into Crop Growth Models to Support Genetic Improvement. **In silico Plants**, 2021.

CARTER, Michael R.; GALARZA, Francisco; BOUCHER, Stephen. Underwriting area-based yield insurance to crowd-in credit supply and demand. **Savings and Development**, p. 335-362, 2007.

DE LEEUW, Jan; VRIELING, Anton; SHEE, Apurba; *et al*. The Potential and Uptake of Remote Sensing in Insurance: A Review. **Remote Sensing**, v. 6, n. 11, p. 10888–10912, 2014.

DISMUKES, Robert; COBLE, Keith H.; MILLER, Corey; O'DONOGHUE, Erick. The effects of area-based revenue protection on producers' choices of farm-level revenue insurance. **Agricultural and Applied Economics Association 2013 Annual Meeting**, Washington, DC. August 4-6, 2013.

ESQUERDO, J. C. D. M.; ZULLO Júnior, J.; ANTUNES, J. F. G. Use of NDVI/AVHRR time-series profiles for soybean crop monitoring in Brazil. **International Journal of Remote Sensing**, v. 32, n. 13, p. 3711–3727, 2011.

EZE, Emmanuel; GIRMA, Atkilt; ZENEBE, Amanuel; *et al*. Exploring the possibilities of remote yield estimation using crop water requirements for area yield index insurance in a data-scarce dryland. **Journal of Arid Environments**, v. 183, p. 104261, 2020.

FONTA, William M.; SANFO, Safietou; KEDIR, Abbi M.; *et al*. Estimating farmers' willingness to pay for weather index-based crop insurance uptake in West Africa: Insight from a pilot initiative in Southwestern Burkina Faso. **Agricultural and Food Economics**, v. 6, n. 1, p. 11, 2018.

GLAUBER, J. W. Crop Insurance Reconsidered. **American Journal of Agricultural Economics**, v. 86, n. 5, p. 1179–1195, 2004.

INTERNATIONAL RESEARCH INSTITUTE FOR CLIMATE AND SOCIETY. **Index insurance and climate risk: prospects for development and disaster management.** Palisades, NY: International Research Institute for Climate and Society, 2009.

JONES, J.W; HOOGENBOOM, G; PORTER, C.H; *et al*. The DSSAT cropping system model. **European Journal of Agronomy**, v. 18, n. 3–4, p. 235–265, 2003.

LETORT, V.; MAHE, P.; COURNEDE, P.-H.; *et al*. Quantitative Genetics and Functional-Structural Plant Growth Models: Simulation of Quantitative Trait Loci Detection for Model Parameters and Application to Potential Yield Optimization. **Annals of Botany**, v. 101, n. 8, p. 1243–1254, 2007.

LU, Weixun; ATKINSON, David E.; NEWLANDS, Nathaniel K. ENSO climate risk: predicting crop yield variability and coherence using cluster-based PCA. **Modeling Earth Systems and Environment**, v. 3, n. 4, p. 1343–1359, 2017.

MIRANDA, M. J.; GLAUBER, J. W. Systemic Risk, Reinsurance, and the Failure of Crop Insurance Markets. **American Journal of Agricultural Economics**, v. 79, n. 1, p. 206–215, 1997.

MIRANDA, Mario J. Area-Yield Crop Insurance Reconsidered. **American Journal of Agricultural Economics**, v. 73, n. 2, p. 233, 1991.

MIQUELLUTI, Daniel Lima. **Weather index insurance design: a novel approach for crop insurance in Brazil**. Doutorado em Economia Aplicada, Universidade de São Paulo, Piracicaba, 2019. Disponível em: <http://www.teses.usp.br/teses/disponiveis/11/11132/tde-02082019-100224/>. Access in: 8 mar. 2021.

OSGOOD, Daniel E. et al. Designing Weather Insurance Contracts for Farmers in Malawi, Tanzania and Kenya: Final Report to the Commodity Risk Management Group, ARD, World Bank. 2007.

OZAKI, Vitor A. Em busca de um novo paradigma para o seguro rural no Brasil. **Revista de Economia e Sociologia Rural**, v. 46, n. 1, p. 97–119, 2008.

ROEL, Alvaro; PLANT, Richard E. Factors underlying yield variability in two California rice fields. **Agronomy Journal**, v. 96, n. 5, p. 1481-1494, 2004.

ROEL, Alvaro; PLANT, Richard E. Spatiotemporal analysis of rice yield variability in two California fields. **Agronomy Journal**, v. 96, n. 1, p. 77-90, 2004.

SCHNITKEY, Gary D.; SHERRICK, Bruce J.; IRWIN, Scott H. Evaluation of risk reductions associated with multi-peril crop insurance products. **Agricultural Finance Review**, v. 63, n. 1, p. 1–21, 2003.

SETIYONO, Tri; QUICHO, Emma; GATTI, Luca; *et al*. Spatial Rice Yield Estimation Based on MODIS and Sentinel-1 SAR Data and ORYZA Crop Growth Model. **Remote Sensing**, v. 10, n. 2, p. 293, 2018.

SKEES, Jerry R. Challenges for use of index-based weather insurance in lower income countries. **Agricultural Finance Review**, v. 68, n. 1, p. 197–217, 2008.

STIGLER, Matthieu; LOBELL, David. Suitability of index insurance: new insights from satellite data. p. 27, .

SUN, Jie; DI, Liping; SUN, Ziheng; *et al*. County-Level Soybean Yield Prediction Using Deep CNN-LSTM Model. **Sensors**, v. 19, n. 20, p. 4363, 2019.

VUOLO, Francesco; RICHTER, Katja; ATZBERGER, Clement. Evaluation of time-series and phenological indicators for land cover classification based on MODIS data. *In*: NEALE, Christopher M. U.; MALTESE, Antonino (Orgs.). Prague, Czech Republic: [s.n.], 2011, p. 81740E. Disponível em: <http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.898389>. Access in: 27 nov. 2019.

WANG, Anna X.; TRAN, Caelin; DESAI, Nikhil; *et al*. Deep Transfer Learning for Crop Yield Prediction with Remote Sensing Data. *In*: **Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies**. Menlo Park and San Jose CA USA: ACM, 2018, p. 1–5. Disponível em: <https://dl.acm.org/doi/10.1145/3209811.3212707>. Access in: 25 fev. 2021

WANG, H. H.; HANSON, S. D.; MYERS, R. J.; *et al*. The Effects of Crop Yield Insurance Designs on Farmer Participation and Welfare. **American Journal of Agricultural Economics**, v. 80, n. 4, p. 806–820, 1998.

ZHANG, Xiaoyang; FRIEDL, Mark A.; SCHAAF, Crystal B. Global vegetation phenology from Moderate Resolution Imaging Spectroradiometer (MODIS): Evaluation of global patterns and comparison with in situ measurements: GLOBAL PHENOLOGY FROM MODIS. **Journal of Geophysical Research: Biogeosciences**, v. 111, n. G4, 2006. Available in: <http://doi.wiley.com/10.1029/2006JG000217>. Access in: 27 nov. 2019.

ZHANG, Xiaoyang; FRIEDL, Mark A.; SCHAAF, Crystal B.; *et al*. Monitoring vegetation phenology using MODIS. **Remote Sensing of Environment**, v. 84, n. 3, p. 471–475, 2003.