



Probabilistic forecasting of remotely sensed cropland vegetation health and its relevance for food security

Ahmed T. Hammad^a, Giacomo Falchetta^{b,c,*}

^a Decatab Pte Ltd., Singapore

^b International Institute for Applied Systems Analysis, Schlossplatz, 1, Laxenburg A-2361, Austria

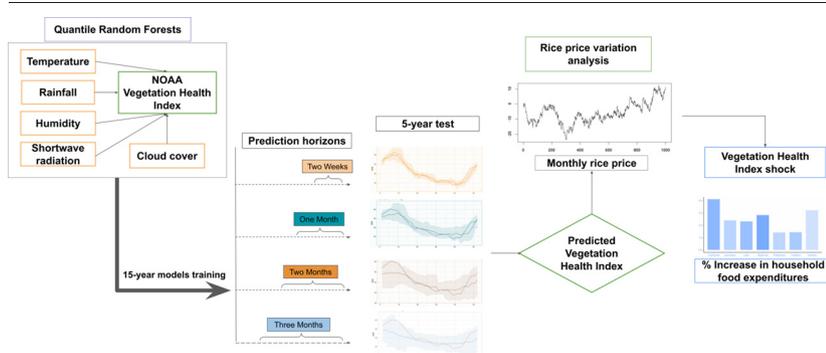
^c Centro Euro-Mediterraneo sui Cambiamenti Climatici, Università Ca'Foscari Venezia, RFF-CMCC European Institute on Economics and the Environment, Italy



HIGHLIGHTS

- Important to monitor weather, crop health and food security nexus
- Probabilistic ML framework to anticipate cropland vegetation health
- Most forecast reliable even months ahead, but increasing uncertainty
- Probabilistic vegetation health forecasts statistically linked to food prices
- Framework implemented in nine countries of South-East Asia

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Fernando A.L. Pacheco

Keywords:

Vegetation health
Food security
Remote sensing of the environment
Probabilistic machine learning
Vulnerability analysis

ABSTRACT

In a world where climate change, population growth, and global diseases threaten economic access to food, policies and contingency plans can strongly benefit from reliable forecasts of agricultural vegetation health. To inform decisions, it is also crucial to quantify the forecasting uncertainty and prove its relevance for food security. Yet, in previous studies both these aspects have been largely overlooked. This paper develops a methodology to anticipate the agricultural Vegetation Health Index (VHI) while making the underlying prediction uncertainty explicit. To achieve this aim, a probabilistic machine learning framework modelling weather and climate determinants is introduced and implemented through Quantile Random Forests. In a second step, a statistical link between VHI forecasts and monthly food price variations is established. As a pilot implementation, the framework is applied to nine countries of South-East Asia (SEA) with consideration of national monthly rice prices. Model benchmarks show satisfactory accuracy metrics, suggesting that the probabilistic VHI predictions can provide decision-makers with reliable information about future cropland health and its impact on food price variation weeks or even months ahead, albeit with increasing uncertainty as the forecasting horizon grows. These results - ultimately allowing to anticipate the impact of weather shocks on household food expenditure - contribute to advancing the multidisciplinary literature linking vegetation health, probabilistic forecasting models, and food security policy.

1. Introduction

Since the introduction of agriculture, 7000 to 10,000 years ago, the link between weather and crop yield has been a constant reason for concern due

to the potential impact of yield variability on food security (Trnka et al., 2016; Black and Thompson, 1978; Frieler et al., 2017; McKeown et al., 2006). Whilst technological change, such as irrigation, new seeds varieties, fertilisers, greenhouses, and land management techniques have

* Corresponding author at: International Institute for Applied Systems Analysis, Schlossplatz, 1, Laxenburg A-2361, Austria.
E-mail address: falchetta@iiasa.ac.at (G. Falchetta).

dramatically mitigated the direct impact of weather on agriculture (Rasmussen, 1962; Parayil, 1992; Sullivan, 1984), weather shocks remain a major reason for concern for stakeholders in the agricultural value chain (Jägermeyr et al., 2021; Wang et al., 2021). This is particularly true in lower income regions of the world, where technology adoption and agricultural knowledge accumulation have lagged behind substantially compared to industrialised countries (Lybbert and Sumner, 2012). Besides weather variability, several studies have highlighted how anthropogenic climate change is and will increasingly be exerting a substantial, mostly negative impact on crops growth and health (Anderson et al., 2020; Mbow, 2020). Climate change is expected to negatively affect global crop yields, and ultimately food supply and accessibility, especially in the most vulnerable regions with fragile agricultural systems (Mora et al., 2015; Porter, 2014). Under these changing conditions and negative outlook, it has become of utmost importance to monitor and provide reliable predictions of vegetation health to better allocate scarce resources, design early warning systems, and ensure food security.

Remote sensing techniques enable collecting large amounts of granular information over vast areas and at different spatio-temporal resolution scales. By characterising natural features on the ground and monitoring their changes over time, remote sensing applications can ultimately support food-related policy-making at different levels (Sishodia et al., 2020; Vroege et al., 2021). The scientific community has made extensive use of satellite imagery for mapping and monitoring changes in land cover and estimating geophysical and biophysical characteristics of the soil (Shanmugapriya et al., 2019; Weiss et al., 2020), as well as to develop and validate vegetation health indicators (Kogan, 2002; Kogan et al., 2004; Kogan, 2019; Kogan et al., 2005; Xue and Su, 2017). The Vegetation Health Index (VHI) (Kogan, 1997; Kogan, 1987) is one such indicators and it has largely been used to monitor crop vegetation over large areas and predict crop yield (Kogan, 1990; Orlovsky et al., 2010; Rahman et al., 2009; Zuhro et al., 2020). In parallel, learning-based statistical algorithms have opened up new frontiers and enabled the development of tools for analysing satellite imagery, providing better and more nuanced insights thanks to their ability to find patterns underlying the complex nonlinear relations that characterise environmental variables. With regards to forecasting, Probabilistic Machine Learning (ML) (Duan, 2020; Ghahramani, 2015; Gneiting and Katzfuss, 2014; Palmer, 2012) has highlighted the importance of accompanying any statement about the future with a quantification of its uncertainty to give a complete picture of the possible scenarios and better inform decision-making at all levels.

In this context, previous research has applied a variety of statistical methods to anticipate vegetation health. Yet, virtually all studies neither analyse the prediction uncertainty nor seek to investigate the relevance of forecasts for food security. For instance, Nay et al. (2018) developed a gradient-boosted machine learning approach to predict the Enhanced Vegetation Index (EVI) based on MODIS satellite data. A similar strategy is followed by Perera et al. (2020), who expand the approach to a larger set of vegetation indexes and ML models for the case of the MENA region. Lees et al. (2022) test Long Short Term Memory Deep Learning architectures for Vegetation Condition Index (VCI) forecasting in Kenya. The only notable exception that explicitly models vegetation health forecasts uncertainty is the recent work by Salakpi et al. (2021), who apply a Bayesian Auto-regressive Distributed Lags (BARDL) model in the context of arid territories in Kenya to forecast the VCI up to 10 weeks ahead while providing a probabilistic prediction.

Irrespective of this evidence, no previous multi-country, large-scale study has sought to analyse the probability distribution of vegetation health forecasts. Yet, to inform policy it is crucial to quantify uncertainty and demonstrate the relevance of such vegetation health forecasts for food security. To fill those gaps, this paper introduces a probabilistic framework seeking to anticipate remotely sensed agricultural vegetation health based on previous weather conditions while making the underlying uncertainty explicit. This represents an important element of novelty compared to studies based on point estimate prediction of vegetation health. The approach is implemented and validated in nine countries of South-East Asia (namely,

Cambodia, Indonesia, Laos, Malaysia, Myanmar, Philippines, Thailand, Timor-Leste and Vietnam) over the 2000–2021 period to demonstrate the suitability of probabilistic ML to predict crop vegetation health at increasing forecasting horizons between the measurement of the weather predictors and the VHI response. The choice of this particular region is motivated by several factors, including but not limited to the vast potential for agricultural production in the region, the high vulnerability to climate change and climate extremes and, the related strong dependency on climatic conditions of their food supply chains. In our study, VHI probabilistic predictions are then statistically linked with historical rice price monthly time series in each country to provide evidence of the relevance of VHI for economic policies for food security and evaluate the food price and expenditure impact of a vegetation health shock. This is of utmost relevance for the development and policy-oriented application of data-driven early-warning systems. The analysis thus contributes to advancing the multidisciplinary literature linking remotely sensed vegetation health monitoring, probabilistic ML, and food security objectives.

The remainder of the paper is structured as follows: Section 2 provides a theoretical background on vegetation health monitoring, the role of remote sensing for food security, and on probabilistic machine learning. It also describes SEA, the pilot region under analysis. Section 3 then describes the methodology designed and implemented to generate probabilistic VHI forecasts and link them to food prices time series. Section 4 presents the results of the analysis, which are then commented in Section 5. Finally, conclusions and future work are discussed in Section 6.

2. Background and theory

2.1. Vegetation health measurement

The literature on the monitoring of the agricultural sector has taken great advantage of satellite technologies, and several indices have been developed to characterise land cover with a specific focus on vegetation health. As defined by Kogan (Kogan, 2019) [p.52]:

Vegetation health is the method designed to derive vegetation condition (favorable, unfavorable, normal, etc.) or health (healthy, unhealthy, stressed, etc.) in response to changing weather (precipitation, temperature, and others) in each ecosystem and climate.

Remote sensing instruments detect light within the electromagnetic spectrum of an object when it is illuminated. Depending on the source of energy used to illuminate the object, there are two ways to collect remote data. Actively, that is, a sensor device emits a signal on the object of interest, and the sensor captures its reflection. Passively, when the natural sunlight reflection of the object is used (Woodhouse, 2017). An important implication for measuring vegetation health using remote sensing instruments is that plants absorb most of the light as part of photosynthesis. Specifically, vegetation absorbs the second-highest amount of solar energy in the visible range of the solar spectrum, which turns to increase the photosynthetic rate and green mass accumulation. Hence healthy plants reflect more near-infrared (NIR) than visible red (RED). When a plant becomes unhealthy, it reflects more RED and less of the NIR. This finding brought to the introduction of the Normalized Difference Vegetation Index (NDVI), measured as the ratio of the difference of the red RED and NIR radiance over their sum as in Eq. (1) (Kriegler et al., 1969).

$$NDVI = \frac{(NIR - RED)}{(NIR + RED)} \quad (1)$$

Based on the NDVI, (Kogan, 1987; Kogan, 1990) proposed three alternative indexes. The Vegetation Condition Index (VCI), the Temperature Condition Index (TCI) and the Vegetation Health Index (VHI). The VCI is expressed as the relation between the value of the NDVI during a chosen composite period of analysis to the long-term minimum NDVI, normalized by the range of NDVI values calculated from the long-term record of the

period of interest (Eq. (2)). VCI evaluates the current vegetation health compared to the historical trends and is designed to disentangle the weather-related component of the NDVI from the other environmental components. The TCI is similar to the VCI, but it is based on the brightness temperature (BT) (Eq. (3)) extracted from the thermal band (Ch 4) of the Advanced Very High-Resolution Radiometer (AVHRR)¹ and it captures the vegetation stress induced by abnormal temperature fluctuations and excessive wetness. Finally, the VHI combine VCI and TCI to assess the total moisture and thermal impacts on vegetation health (Eq. (4)). The three indices range from 0 (very unhealthy) to 100 (very healthy), with values from 40 to 0 indicating increased vegetation stress and 60 to 100 suggesting better vegetation health conditions.

$$VCI = 100 \times (NDVI - NDV_{min}) / (NDV_{max} - NDV_{min}) \quad (2)$$

$$TCI = 100 \times (BT_{max} - BT) / (BT_{max} - BT_{min}) \quad (3)$$

$$VHI = a \times VCI + (1 - a) \times TCI \quad (4)$$

The three Vegetation Health indexes developed by Kogan have been used extensively in the literature for monitoring the vegetation activity in response to weather-related drivers such as drought (Zuhro et al., 2020; Baniya et al., 2019; Kamble et al., 2019; Liang et al., 2017; Marufah et al., 2017; Masitoh and Rusydi, 2019; Pei et al., 2018) and to evaluate crop production (Kogan, 1990; Orlovsky et al., 2010; Rahman et al., 2009). The findings in the literature highlight that VHI correlates well with meteorological drought and agricultural drought in monsoonal rainfall areas (Marufah et al., 2017) and, most importantly, is useful to predict the yield of several grain crops such as corn in China (Kogan et al., 2005), wheat in the USA (Salazar et al., 2007), and rice in Bangladesh (Rahman et al., 2009) several months in advance of the harvest with considerable implications for food security. On the other hand, VHI was found to perform poorly in high latitude regions due to the assumption of inverse correlation between NDVI and BT that is not met in northern ecosystems where increasing temperatures support plants growth (Karnieli et al., 2006).

2.2. Food security and remote sensing

In the comprehensive definition of the Food and Agriculture Organization (FAO) of Food Security (FAO, 2003), some key features of what is to be in a "state of food security", are far from having being globally reached. A constant, undeniable and powerful force that limits access to "sufficient, safe, and nutritious food", is represented by extreme climate variability. In a recent work Hasegawa et al. (2021), suggest that regions heavily affected by climate extremes such as South Asia, might require to triple the current food production to compensate for the impact of climate change. An increase in production is challenging considering that since 1961 the total agricultural productivity has decrease by 21% globally and between 26 and 34% for regions closer to the tropics (Ortiz-Bobea et al., 2021). On the top of this, the "physical, social, and economic access" to food has been additionally hindered by the global COVID-19 pandemic, stressing the fragility of the world food supply chain as a consequence of the protracted crisis (Aday and Aday, 2020; FAO, 2020).

The complexity of these forces and the difficulties in providing reliable measurements and predictions of food security, have proven to be a stimulus for the development of a wide variety of systems for food security prediction and early warning systems based on different proxies for food insecurities. Using a variety of methodologies such as expert judgment (Funk et al., 2019) and data-driven approaches (Westerveld et al., 2021), the scientifically literature has contributed significantly to the way the status of food security is monitored in any given region of the world. Remote sensing have been widely used by practitioners and researcher as a cost-effective quantitative and non-destructive source of data, to monitor and

detect the evolution and health of crops over large areas and correlate it with crop yields as primary proxy for food security.

As highlighted by the in-depth review of Karthikeyan et al. (2020), there is a clear and robust relationship between crop yield and remotely sensed vegetation indices with correlations ranging from 60% to 80%. Aside from the use of VHI, other important indices have been explored in the literature such as the well-known (NDVI) (Rouse et al., 1974) and its variations such as the enhanced Vegetation Index (EVI) (Liu and Huete, 1995), the Green-Red Vegetation Index (GRVI) (Tucker, 1979), the Soil Adjusted Vegetation Index (SAVI) (Huete, 1988) and the Red Edge Position (REP) (Filella and Penuelas, 1994). In a recent work Jeong et al. (2022), used a combination of vegetation indices, weather and geographic variables to predict rice yields in North and South Korea within a Deep Learning framework. The target variables was reported rice yields collected at different scales by the local government and international organisations such as FAO as a part of a crop process-based model designed to simulate crop growth. The resulting model reached an R-squared of 0.86. In line with the results in Korea, the study of Son et al. (2020), demonstrated the effectiveness of ML algorithms for regional rice yield predictions based on Moderate Resolution Imaging Spectroradiometer (MODIS) NDVI data in Taiwan. Using a less computational intensive approach Nazir et al. (2021), found results ranging from 0.62 to 0.83 based on a Partial Least Square Regression in Pakistan on multiple indices including NDVI, EVI, SAVI and REP). On the other hand, in the review of Wen et al. (2021), the authors points out how some of these indices were found to be inaccurate in characterising crop health under drought and salinity stress and plant functional traits to be better in characterising the impact of these stressors.

It has to be noted that while all these studies indicate the robustness of the relation between crop yield and remote sensing and its direct relation with food security monitoring and prediction, none of them have explored the economic implication and relevance for further food security analysis. This goes in line with the review of Kubitzka et al. (2020), where the authors stress the preponderance of studies limited to the biophysical sphere.

2.3. Probabilistic machine learning

Uncertainty and its quantification are fundamental aspects of any environmental analysis, and they are now crucial aspects in the face of climate change and its complicated consequences. Epistemic uncertainty (caused by limited data and knowledge) and aleatory uncertainty (randomness or variability in the underlying variables) are virtually present in all projections of a future phenomenon (Helton et al., 2010). In the last decade, researchers in the field of Probabilistic Machine Learning have developed more sophisticated algorithms that provide reliable predictions and account for the inherent uncertainties of Machine Learning algorithms (Duan, 2020; Ghahramani, 2015; Gneiting and Katzfuss, 2014; Kabir et al., 2018; Meinshausen and Ridgeway, 2006). By shifting from a point-wise estimation focused only at the conditional mean of a response variable to a distribution estimation aiming at providing information about the full conditional distribution of the outcome variable of interest, the notion of uncertainty is explicitly integrated in the prediction while providing more comprehensive information to policy-making.

To clarify how decision making and our understanding of a phenomenon of interest could benefit from moving beyond the conditional mean, it is worth introducing some notations.²

Let Y and X denote a real-valued outcome variable and a set of covariates, respectively. A typical ML algorithm, produce a point estimate as in Eq. (5).

$$E(Y|X = x) \quad (5)$$

² In the discussion below, the presentation of (Meinshausen and Ridgeway, 2006) is closely followed

¹ <https://www.eumetsat.int/avhrr>

Eq. (5) can be interpreted as way of answering the question “What value Y would take on average, given X ?”. Answering this question can provide important insights, but - at the same time - such insights will regard exclusively the mean of the outcome of interest Y , while other information inherited from the data will be disregarded. A relatively simple, yet highly informative approach to overcome this limitation is to focus on the quantiles distribution of Y give X instead of the simple mean. Formally this is equivalent to Eq. (6).

$$Q_\alpha(x) = \inf \{y : F(y|X = x) \geq \alpha\} \tag{6}$$

Where the conditional distribution function $F(y | X = x)$ is equivalent to the α quantile (Q_α) given $X = x$. This leads to prediction intervals as a way to encapsulate information about uncertainty into the model. Indeed, a prediction interval for Y given X , with a desired width α , can now be obtained directly from the data. For example, if the 90% prediction interval $I(x)$ of Y for $X = x$ is sought, then:

$$I(x) = [Q_{0.05}(x), Q_{0.95}(x)] \tag{7}$$

In the classic parametric quantile regression, obtaining the quantities of interest is achieved by minimizing the expected loss $E(L_\alpha)$ (the analogous of minimizing the expected squared error loss for classic regression). If the loss function L_α for $0 < \alpha < 1$ is defined by the weighted absolute deviations as in Eq. (8), then the objective is to minimize the expected loss $E(L_\alpha)$ as per Eq. (9).

$$L_\alpha(y, q) = \begin{cases} \alpha|y - q| & y > q \\ (1 - \alpha)|y - q| & y \leq q \end{cases} \tag{8}$$

$$Q_\alpha(x) = \arg \min_q E\{L_\alpha(Y, q) | X = x\} \tag{9}$$

In a non-parametric setting, [Meinshausen and Ridgeway \(2006\)](#) have shown an alternative approach to retrieve the conditional quantile distribution using the Random Forests (RF) algorithm of [Breiman \(2001\)](#). Quantile Random Forests (QRF) generalise RF by storing all observed responses at the leaf level of each tree, instead of just the mean, allowing to measure empirical quantile estimates. In other words, the classic conditional mean estimated by RF as the weighted mean of the observed outcome of interest, is replaced with its weighted distribution.

The predicted intervals give us an important additional information regarding the possible variation around a new predicted value. The correspondent width of the intervals can be interpreted as a simple yet highly informative measure of accuracy of the model, with wider intervals indicating a higher degree of uncertainty in the predicted value. A more sophisticated measure of performance of a distribution prediction can be found in the family of scoring rules ([Gneiting and Katzfuss, 2014](#); [Brier et al., 1950](#); [Gneiting and Raftery, 2007](#); [Krueger et al., 2016](#)). A scoring rule function $S(F, Y)$ is a measure of accuracy of a distribution prediction F given the observed outcome Y . In this paper, the *Interval Score* is considered, a proper scoring rule designed to score quantile predictions based on the work of [Gneiting and Raftery \(2007\)](#) which is computed as per Eq. (10).

$$S = (u - l) + 2/\alpha * (l - Y) * 1(Y < l) + 2/\alpha * (Y - u) * 1(Y > u) \tag{10}$$

Where Y is the true value of the outcome of interest, l and u denotes respectively the lower and the upper quantiles of the range defined by the value of α .

The usefulness of the QRF algorithm has been investigated in a number of contribution in the environmental literature. [Sanderman et al. \(2018\)](#) used QRF to provide probabilistic predictions of carbon stock in Mangrove finding that the highest level of uncertainty in the predicted carbon stocks at 1 m depth was on average 40.4% of the organic carbon stocks. [Taillardat et al. \(2016\)](#) compared QRF with a parametric probability density function model to forecast surface temperature and wind speed, using hourly data

and lagged features from 3 up to 54 h. Their results highlight the suitability and better performances of QRF with sharp and reliable probabilistic predictions. [Vaysse and Lagacherie \(2017\)](#) Compared Regression Kriging for sparse digital soil mapping with QRF. The authors suggest that while the two algorithms reached similar performances on the point-wise prediction of soil properties, QRF is preferable over Regression Kriging since the former provides more accurate and interpretable probabilistic prediction than latter.

2.4. Study area

The SEA region is characterised by a tropical climatic with monsoons and a substantial amount of rainfall alternated with period of dryness. Variation in rainfall and temperature between countries ([Table 1](#)) can be observed and have been related with El Niño/Southern Oscillation (ENSO), causing differences in the timing of the rainfall phase between the north-west of the region and the rest of the area.

According to [Eckstein et al. \(2021\)](#) the average temperatures in the SEA have been increasing rapidly since 1960 bringing Myanmar, the Philippines, and Thailand among the most affected countries in world by climate change in the period 2000–2019. Panel A of [Fig. 1](#) shows the 2010–2020 average temperature change with respect to the long-term historical average in each country, highlighting that in most countries the one °C warming threshold has already been surpassed, with a peak of 1.3 °C in Myanmar.

Turning to food production, the main commodity are represented by rice, maize, coffee, cocoa and palm oil. Rice is the single most important crop in the region. SEA countries have already experienced extensive losses due to extreme weather and climate change ([Zhongming et al., 2021](#)). Panel B of [Fig. 1](#) shows the Coefficient of variation of rice yield coefficient of variation of rice yield (units of production per land cultivated) according to the country-level statistics for the 2010–2020 period provided by the FAOSTAT database ([FAO, 2022](#)). This reveals significant heterogeneity in the stability of rice productivity, with the greatest fluctuations observed in Cambodia, Laos, and Malaysia. Among the listed countries, Cambodia, Myanmar and Laos have been labeled as the “*poorest of the poor*” ([Food & of the United Nations, 1998](#)). Overall, the percentage of population living on less than \$5.50 a day, ranges from 2.7% in Malaysia to as high as 94% in Timor-Leste ([Bank, 2021](#)). The number of undernourished people in 2020 increased by 6% in the region most probably due to the pandemic ([Zhongming et al., 2021](#)). Finally, four of the country under analysis have a Global Food Security Index (GFSI) score ([The Economist, 2021](#)) falling below 60 with Laos, Cambodia and Myanmar below 50.

3. Materials and methods

3.1. Modelling framework

[Fig. 2](#) introduces the framework of the analysis, which is detailed in the sections below. A probabilistic prediction framework is designed and four prediction horizon scenarios are considered, indicating the temporal

Table 1

Summary statistic for temperature (in Celsius degrees) and rainfall (in mm), 20-year long-term average, by country.

Country	Mean rain	Min. rain	Max. rain	SD rain	Mean temp.	Min. temp.	Max. temp.	SD temp.
Cambodia	3.660	0	20.879	3.847	29.014	23.448	38.604	2.427
Indonesia	4.240	0	14.000	3.757	27.618	24.336	33.156	1.066
Laos	3.429	0	28.427	4.645	25.597	17.633	31.004	2.110
Malaysia	5.300	0	21.565	3.556	27.379	23.359	30.761	0.989
Myanmar	2.322	0	25.231	3.677	26.388	18.103	35.870	3.323
Philippines	3.538	0	18.077	3.870	27.879	23.442	31.926	1.329
Thailand	2.549	0	17.398	3.136	29.175	23.274	37.603	2.383
Timor-Leste	3.851	0	32.356	5.600	28.634	23.434	38.201	2.786
Vietnam	2.335	0	18.937	3.060	27.425	17.147	32.236	2.350

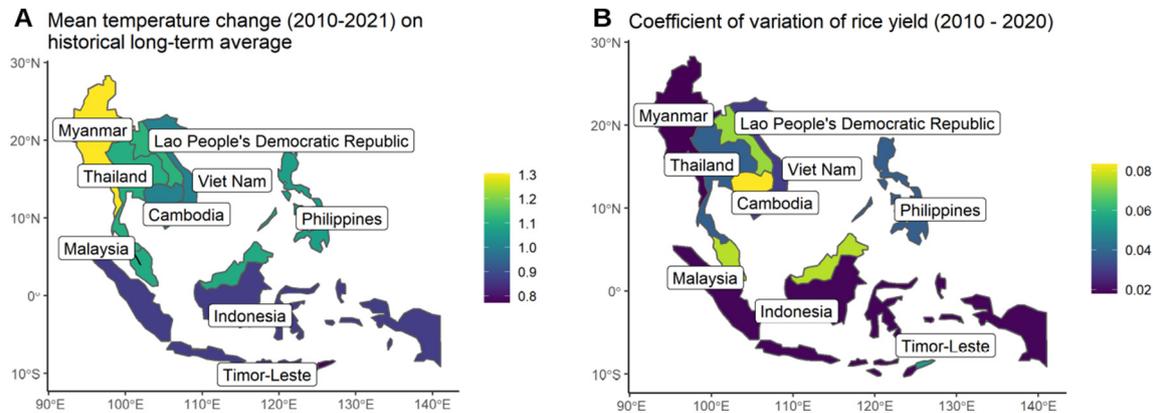


Fig. 1. Panel A: Average temperature change (2010 – 2020) with respect to the long-term historical average in the SEA countries considered in the analysis. Source: FAOSTAT. Panel B: Coefficient of variation of rice yield (2010–2020) in the SEA countries considered in the analysis.

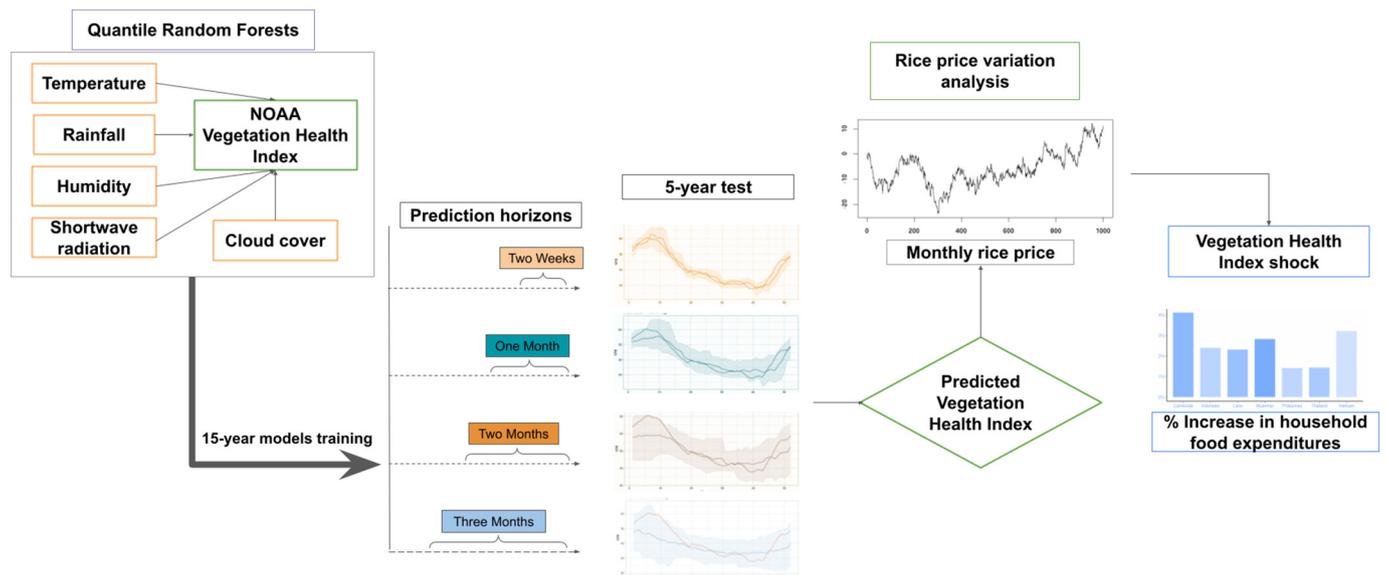


Fig. 2. Schematic representation of the modelling framework. The flowchart describes the application of a probabilistic ML approach to predict the VHI in nine SEA countries under different prediction horizons and the assessment of the capacity of the VHI predictions to determine food prices and thus household food expenditure.

distance between the target VHI measurement and the predictor variables measurement: two-week, one-month, two-month and three-month ahead. The QRF models (a model for each of the nine pilot countries in SEA, for each prediction horizon) is trained on 15 years of weekly data (2000–2015) and tested on the following five years of sample (2016–2021). Finally, a panel regression model is estimated to appraise the statistical link between country-specific variations in the VHI predictions and monthly rice prices. The estimated coefficients are used to derive implications for food expenditure and security.

3.2. Data sources and processing

3.2.1. VHI, climate, and satellite data

The analysis is based on a combination of remotely sensed and atmospheric data detailed in Table 2 from the first week of April 2000 to the last week of July 2021. The primary outcome of the analysis is the VHI over cropland produced from the National Environmental Satellite, Data, and Information (NESDIS) part of the National Oceanic and Atmospheric Administration (NOAA). The data is available from 1981 with 4 km spatial and 7-day composite temporal resolution. To capture the long-term contribution of climate on vegetation health, yearly means of precipitation and

temperature are included as covariates. For the short-term weather contribution, monthly rainfall, temperature, humidity, solar radiation means, cloud coverage and the total amount of rain by month are used. Raw covariates data are masked over cropland pixels on the Google Earth Engine platform (Gorelick et al., 2017) using the GFSAD1000 gridded cropland product (Teluguntla, 2015); then, the mean value of each variable is extracted for each day of the study period over SEA.

The masked data are then imported into the R scientific programming environment (Team, 2020), where additional processing steps are carried out to generate a set of features that could enhance the accuracy of the model for each prediction scenario, and capture trends and seasonality in the time series. This includes lagged features of each variable, rolling sum of rain and rolling mean of rain, temperature, humidity and solar radiation over the previous month. Year, month, week of the year, and sine and cosine function of time (for week, month, and year), are included as additional predictors (see Table 3).

3.2.2. Food security analysis data

To examine the relevance of the VHI indicator for food security forecasting purposes, additional statistical analysis based on monthly time series of the wholesale prices of rice in seven countries of East-Asia is carried out.

Table 2
Input data for VHI prediction.

Data product	Variable	Source
GFSAD1000: Cropland Extent 1 km Multi-Study Crop Mask	Landcover	(Teluguntla, 2015)
STAR - Global Vegetation Health Products	VHI	(STAR, 2018)
CHIRPS Daily: Climate Hazards Group InfraRed Precipitation with Station Data (version 2.0 final)	Precipitation	(Funk et al., 2015)
MOD11A1.006: Terra Land Surface Temperature and Emissivity Daily Global 1 km	Temperature	(Wan et al., 2015)
CFSV2: NCEP Climate Forecast System Version 2, 6-Hourly Products	Maximum specific humidity 2 m above ground, 6-hour interval; Downward Short-Wave Radiation Flux surface 6 Hour Average	(Saha, 2011)
NOAA CDR PATMOSX: Cloud Properties, Reflectance, and Brightness Temperatures, Version 5.3	Cloud Coverage	(Heidinger et al., 2014)
World Bank official boundaries	SEA shapefile and Administrative boundaries	(world-bank-official-boundaries)

Table 3
Covariates used to predict VHI.

Variable	Lag	Rolling Sum	Rolling Mean	Year Mean	Month Mean	Month Sum
VHI	x					
Rain	x	x	x	x	x	x
Temperature	x		x	x	x	
Humidity	x		x		x	
Cloud %	x					
Solar Radiation	x		x		x	

The price data is drawn from the FAO GIEWS Food Price Monitoring and Analysis (FPMA) tool³ historical records, retrieving the longest possible span of time for each country corresponding to the VHI period under analysis. In this food security analysis, rice is considered as it is the first staple crop produced and in the region (FAO, 2022), as well as one where historical monthly price data is consistently available across the countries considered. The rice price time series are visualised in Fig. SI-3.

Then, for each of the countries investigated, statistics on the yearly average consumption of rice per capita are retrieved from the OECD-FAO Agricultural Outlook (OECD, 2020) and on the yearly average food expenditure from the USDA Economic Research Service. Combining these statistics with the results of the VHI-rice price statistical analysis, the potential impact of a shock of the VHI (such as a standard deviation decrease from its average long-run value) is simulated to evaluate its impact on household food expenditure.

³ <https://fpma.apps.fao.org/giews/food-prices/tool/public/#/dataset/domestic>

3.3. Statistical modelling

3.3.1. Probabilistic forecasting of VHI

To explore the suitability of the QRF algorithm (Meinshausen and Ridgeway, 2006) to produce reliable probabilistic predictions of crop health for use in early-warning systems, a modelling framework based on increasing prediction horizons between the environmental predictors and the VHI is defined, as depicted in Fig. 2. More specifically, instead of providing a single estimation over the full SEA region as whole and using a single prediction horizon period, for each country four different prediction leads are considered, corresponding to multiple forecast horizons, namely, two weeks, one month, two months and three months ahead prediction. These lead to a total of thirty-six models trained over about sixteen years and tested over a period of five years. While our main interest is over the predicted quantiles intervals at a probability of 0.1 and 0.9, also provide point-wise predictions of RF are provided.

In this study, the QRF algorithm is implemented using the R packages **ranger** (Wright and Ziegler, 2015) and **caret** (Kuhn, 2008). To find the optimal set of hyperparameters, the resample procedure for times series proposed by (Hyndman and Athanasopoulos, 2013) is adopted. To score the result of the probabilistic prediction, the quantiles scoring rule metric as implemented in the package **scoringutils** (Bosse et al., 2020) is considered, while for the point-wise predictions classic metrics including R-squared (R^2), root mean square percentage error (RMSPE), and mean absolute percentage error (MAPE) are adopted.⁴

3.3.2. Linking VHI forecasts with food prices and expenditure

To link time series of wholesale rice prices with the VHI probabilistic predictions, a monthly unbalanced panel (due to price data availability varying across countries, see Fig. SI-3) dataset is assembled. Based on these data:

1. Fixed effects panel data regression is carried out to absorb unobserved country-specific and time-specific heterogeneity from the effect estimation. In particular, both country fixed effects, and month and year fixed effects, as well as their interaction are included. Here:
 - country fixed effects absorb national market specificities;
 - month fixed effects capture the seasonality dynamics of prices;
 - year fixed effect are representative of regional to global variations in prices which affect all countries in the panel;
 - finally, month by year interactions capture the interplay between the two latter dynamics.

Fixed effects regression thus aims at isolating the true link between VHI predictions and rice prices while controlling for additional underlying trends affecting prices. The models are run on a dataset ensemble of two-week up to three-month ahead VHI predictions.

2. As predicting variables, the middle quantile of the predicted VHI is included, as well as an array of lagged values (1-3-6-12-24-48-72 months) to control for the short- to long-term impact of vegetation health on prices.
3. Different dependent variable timings are tested for, starting from the current value of prices, going back to lagged values of prices by 1, 2, and 3 months with respect to the VHI prediction. The reason for testing specifications with lagged dependent variables is that lags can capture the anticipation effect of prices, such as market expectations over future food prices
4. Finally, a control variable measuring the prediction interval width at each t is also included.

⁴ Additional packages used in this work include **data.table** (Dowle and Srinivasan, 2021), **tidyverse** (Wickham et al., 2019), **ggplot2**, (Wickham, 2016), **lubridate** (Grolemund and Wickham, 2011), **purrr** (Henry and Wickham, 2020), **stargazer** (Hlavac, 2015), **patchwork** (Pedersen, 2020), and **ggsci** (Xiao et al., 2018).

In mathematical terms, the following is estimated:

$$P_{cmy} = VHI_{cmy} + VHI_{cm-1y} + VHI_{cm-3y} + VHI_{cm-6y} + VHI_{cm-12y} + VHI_{cm-24y} + VHI_{cm-48y} + VHI_{cm-72y} + width_{cmy} + \gamma_c + \psi_m + \mu_y + \psi_m \cdot \mu_y + \epsilon_{cmy} \quad (11)$$

where:

- P are wholesale rice prices
- c, m and y represent country, month and year, respectively
- VHI is the middle quantile of the Vegetation Health Index prediction
- $width$ is the VHI prediction interval width
- γ are country fixed-effects
- ψ are monthly fixed-effects
- μ are yearly fixed-effects
- ϵ is the error term

Finally, the estimated regression coefficients are used to evaluate the impact of a shock in the (predicted) VHI on rice prices, and thus on rice expenditure given a representative yearly level of per-capita consumption based on historical statistics. To achieve this, a shock S_c is simulated:

$$S_c = n_c \cdot SD_c \quad (12)$$

Here, for each country c , n describes the number of SD (standard deviations) from the mean VHI values to the lowest percentile of each distribution, i.e. for the vegetation to become stressed (falling in the range 0 to 40). The VHI shock S_c is then multiplied by the regression coefficient from Table 4 (considering the mean value of VHI regression coefficients

Table 4

Panel fixed-effects regression results for the impact of the VHI predictions on national monthly rice prices. Estimated regression coefficients of Eq. (11) with standard errors in parentheses. L1-L2-L3: one, two, and three-month lagged variables.

	Dependent variable			
	Rice price	L1 rice price	L2 rice price	L3 rice price
	(1)	(2)	(3)	(4)
VHI	-0.002*** (0.0003)	-0.003** (0.002)	-0.005*** (0.002)	-0.008*** (0.002)
VHI (lag 1 month)	0.0002 (0.0002)	0.005*** (0.002)	0.004** (0.002)	0.006*** (0.002)
VHI (lag 3 months)	0.001** (0.0002)	0.004*** (0.001)	0.010*** (0.002)	0.009*** (0.001)
VHI (lag 6 months)	-0.0003 (0.0002)	-0.008*** (0.001)	-0.015*** (0.002)	-0.008*** (0.001)
VHI (lag 12 months)	-0.0004* (0.0002)	0.0002 (0.001)	-0.001 (0.002)	-0.001 (0.001)
VHI (lag 24 months)	0.002*** (0.0002)	-0.002 (0.002)	-0.002 (0.002)	0.005*** (0.001)
VHI (lag 48 months)	0.0003 (0.0002)	0.002* (0.001)	0.002 (0.002)	-0.002 (0.001)
VHI (lag 72 months)	0.0004* (0.0002)	-0.001 (0.001)	-0.001 (0.002)	0.003** (0.001)
Prediction interval width	-0.0001 (0.0002)	0.0003 (0.001)	0.002 (0.001)	0.005*** (0.001)
Constant	0.302*** (0.032)	0.588*** (0.203)	0.934*** (0.228)	0.514*** (0.199)
Observations	1568	1568	1568	1568
Country fixed-effects	YES	YES	YES	YES
Month fixed-effects	YES	YES	YES	YES
Year fixed-effects	YES	YES	YES	YES
Year by month fixed-effects	YES	YES	YES	YES
R ²	0.986	0.415	0.260	0.435
Adjusted R ²	0.985	0.387	0.224	0.407
Residual Std. Error (df = 1494)	0.038	0.243	0.274	0.239
F Statistic (df = 73; 1494)	1391.725***	14.528***	7.184***	15.742***

* $p < 0.1$.
 ** $p < 0.05$.
 *** $p < 0.01$.

from specifications 1–4) to derive the impact on rice prices, and divide it by the most recent yearly mean price in each country to get the relative (%) price impact. Finally, based on country-specific average household yearly rice consumption and food expenditure statistics, the potential impact of the shock S_c on total household food expenditure due to this potential increase in rice prices is calculated. Table SI-3 reports, for each country, the yearly average household food expenditure, the mean rice price for the most recently available year, and the value of one standard deviation in the VHI. In addition, it is useful to compare this quantity with the month and country specific observed variations in the VHI index between 2000 and 2021 plotted in Figure SI-1.

4. Results

4.1. Probabilistic forecasts of cropland vegetation health

Contrarily to point-wise estimates, a prediction interval offer a way to assess the reliability of a prediction by its width. Tighter prediction intervals indicate less fluctuations of the response variable and consequentially, a more reliable prediction. Fig. 3, displays the range of potential outcome values for VHI against the observed values (yellow line) for the four prediction horizon for each of the nine countries analysed in the five-year validation period (see Methods). Wider fluctuations are observed especially for Malaysia, Myanmar and Thailand even on the two weeks prediction horizon (dark blue ribbon). Overall, the intervals for two weeks and one month prediction horizons are tight around the observed values of VHI while the intervals for two and 3 months are very large with a tendency to under-predict the observed outcome especially for Malaysia and Myanmar. On the other hand, over-prediction is more pronounced for Laos in the period 2019–2020 and at the end of the time series for Timor-Leste.

To quantify the prediction intervals goodness-of-fit, the Interval Score (IS) metric is considered (Gneiting and Raftery, 2007). The IS has attractive properties because it balances coverage and length of the prediction intervals with lower value of the metric indicating better estimated prediction intervals. Fig. 4A reports the mean value of the calculated IS metric, across the 5-year testing period in the analysis, while Fig. SI-2 shows the complete time series for each week in the testing period.

With a prediction horizon of two weeks, for most of the countries under analysis the MIS is below the value of one, suggesting high reliability on average. For each increase in the prediction horizon, all the predictions depart from 0, but Indonesia and the Philippines are the best performing. The predicted intervals for Myanmar, Cambodia, Laos and Thailand are the most sensitive with the one for Cambodia triplicating from the initial prediction horizon of two weeks to the three months horizon.

While the focus and the core novelty of our approach is its probabilistic nature (allowing to produce prediction intervals) the conventional point-wise prediction accuracy metrics are also considered for comparability with the existing literature. Fig. 4B reports the training and testing accuracy metrics -as measured by the R²- of the point-wise prediction of the QRF models,⁵ by country and prediction horizon. While the results on the training are all very close to 100%, it is evident that increasing the prediction horizon negatively impacts the accuracy of the prediction. Within a prediction horizon of maximum one month the results on the test are still within the range of 60% to 80% which is in line with the results found in the existing literature. An exception is Indonesia, where the accuracy remain constantly above 75% even with a prediction horizon of 3 months. On the other hand, the worst performing prediction is for Malaysia with an accuracy as low as 35% and 18% for two and three months prediction horizons respectively (additional metrics including RMSPE and MAPE are displayed in Table SI-1).

⁵ These are equivalent to the mean predictions given by the standard version of RF.

5-year predicted Quantile Intervals

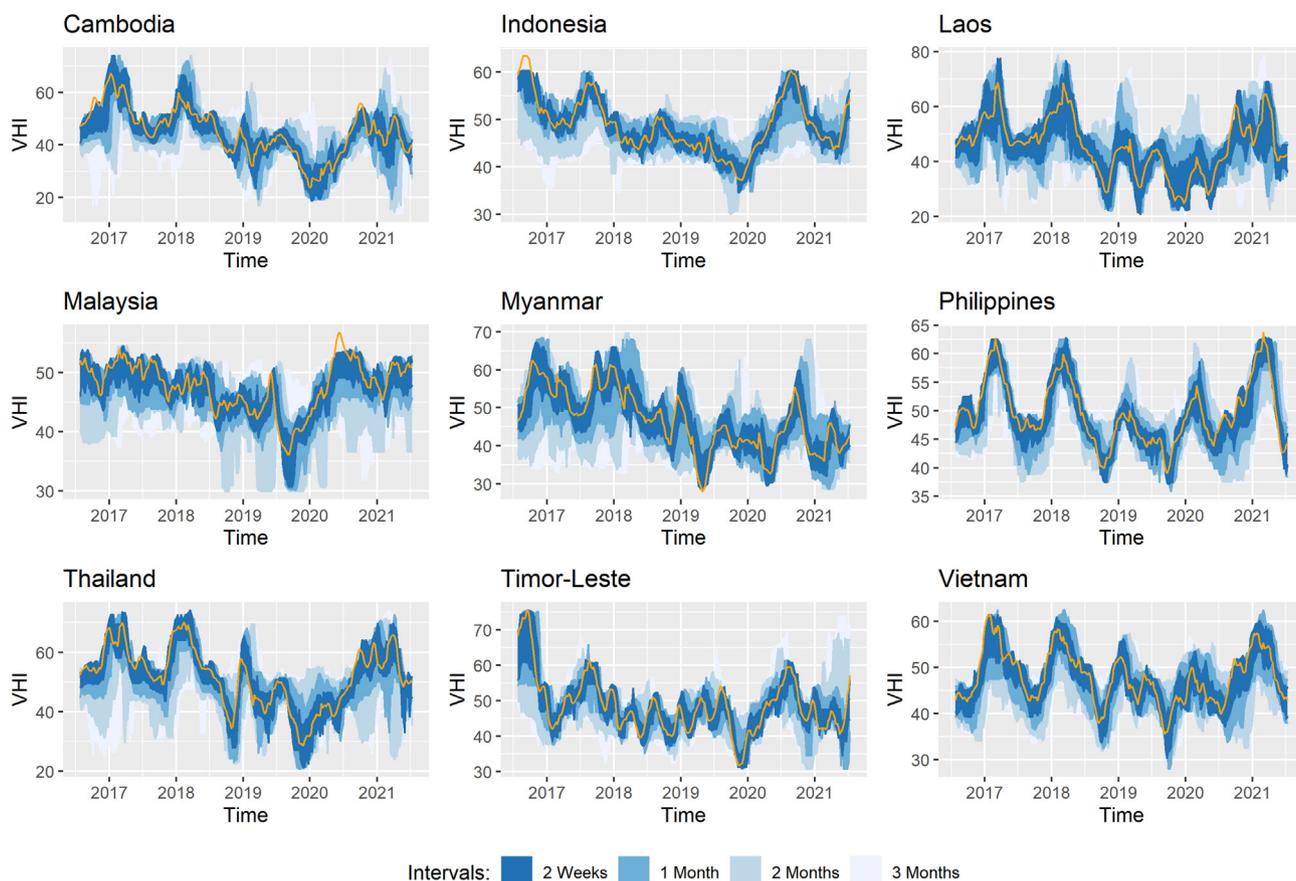


Fig. 3. Predicted Vegetation Health Index (VHI) intervals for the four prediction horizons vs. observed value on the test dataset (2016–2021), by country. *Note:* the upper and lower intervals for all prediction scenarios, correspond to a probabilities of 0.1 and 0.9, respectively.

4.2. Linking cropland health predictions and food security

Table 4 presents the results of the fixed-effects regression models linking VHI predictions with monthly national rice price time series (see Methods). In particular, models with four different dependent variables are tested: rice prices, and the one, two, and three-month lag of rice prices. The lagged specifications are considered to test whether an anticipation effect exists, i.e. if prices adjust in advance to future variations in the VHI index, such as based on future yield expectations. In the regression table, the variable VHI identifies the middle quantile of the probabilistic VHI estimates, whilst a control variable measuring the prediction interval width at each t is also added (see Methods).

The results show that the VHI has a strong and significant impact on rice prices in all the considered specifications, and that this effect increases in magnitude as the order of the lag of the dependent variable increases. This means that rice prices are adjusting in advance based on future vegetation health (and, presumably, yield) expectations. Additional results based on the actual VHI values, rather than the predictions, are found in Table SI-2. An additional important finding is that the results based on our VHI ahead predictions find good agreement with the results of the models based on the observed VHI values in terms of their magnitude and statistical significance. This result suggests that VHI predictions are suitable for food price change prediction analysis, such as in early warning systems.

To complement the analysis and quantify the regression coefficients in Table 4, Fig. 5 reports the results of the simulation of a potential shock S_c in the VHI (see Methods for details on the definition of the shock magnitude and Table SI-3 for numerical details) and its impact on rice prices, and thus on total household food expenditure. The graphs show that the simulated

shock can have a large impact on national rice prices (an average increase of 16.5% of the current mean price), although with great variability, with a minimum of 7% in Laos and a maximum of up to 25% in Myanmar and Vietnam. This price change could have a potential direct impact on total household food expenditure: on average, a 2.5% growth, ranging between a minimum of 1.4% in the Philippines and Thailand, and a maximum of 4.1% in Cambodia.

5. Discussion and policy implications

Most of the studies in the literature have focused on point-wise estimations, setting aside considerations on the inevitable uncertainty that comes with any prediction. Moreover, compared to most previous studies, the results of this study are based on a large and heterogeneous region of the world in both vegetation, climate and food insecurity terms.

Given the innovative methodological aspects and the wider geographical scope of this work, it is challenging to compare its results with previous studies. To mention the most similar ones, the analysis of Nay et al. (2018) achieved R^2 values ranging from 0.76 to 0.86 with a two-week ahead prediction horizon, but it was limited to agricultural land in California and Sri Lanka. Similarly, Perera et al. (2020) achieved an accuracy of 0.93 for the MENA region, but their analysis solely focuses on near-real-time prediction. Lees et al. (2022) reached an average accuracy of 0.83 for one-month ahead prediction and 0.95 for three-month ahead prediction using a Deep Learning architecture but focusing on a circumscribed area in Kenya. The most directly comparable study is the work of Salakpi et al. (2021), which compares different prediction horizons within a Bayesian framework that naturally provide uncertainty in the predicted values. More specifically

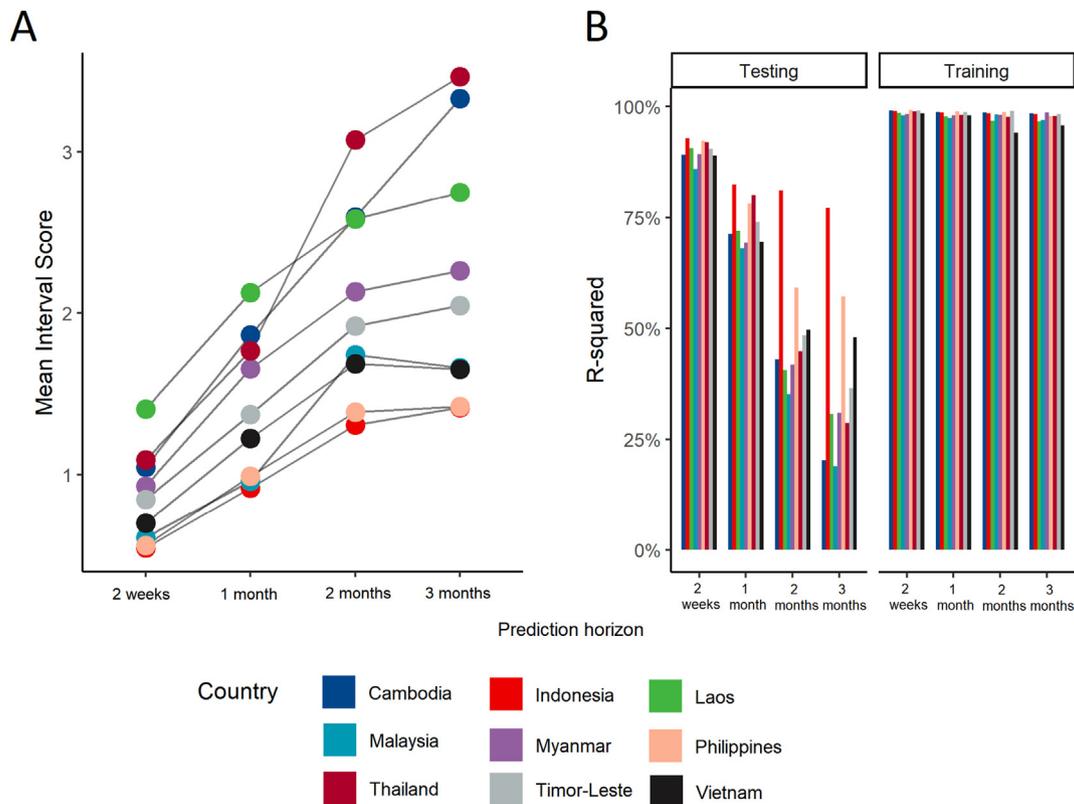


Fig. 4. ML accuracy metrics. (A) Mean Interval Score (MIS), by country and prediction horizon; (B) Training and testing accuracy (R^2), by country and prediction horizon.

their 1.5 to 3 months-ahead prediction R^2 ranges between 0.94 and 0.62 but also in this case, the focus remain on a limited region.

In this context, the findings of this paper thus have important implications from a policy-making point of view as they provide a proof-of-concept for the design and implementation of early-warning systems and contingency plans for crop health and food prices based on satellite data and probabilistic machine learning. In particular, the results are based on openly available remotely sensed data and they are able to capture non-linear and accumulation dynamics of weather variables. In addition, the proposed framework can be adapted to virtually any geography and crop type. These are attractive features for decision-makers looking for readily updated, inexpensive food-related forecasts (Ehlers et al., 2021). Such system can substantially reduce the need for continuous field data collection to keep track of current and expected changes in cropland vegetation health. In turn, the ability to anticipate national food price variation following weather shocks can have important implications for governmental food trade and stocking decisions, as well as for short-run risk hedging decisions in the commodity stock market. Additional considerations are related with the trade-offs between the short and the long run impact of food security policies. The proposed framework and results are mostly relevant for the short term policy response to support households against food prices shocks (Ansah et al., 2021). On the other hand, this strategy might eventually converge to more structural measures and prove to have long-term moderating effects on climate shocks on household dietary and health conditions.

6. Conclusions

In this paper, a statistical framework to improve the understanding of the nexus between weather variability, crop health, and food security is designed and implemented. QRF - a ML algorithm designed to estimate quantile values and thus produce probabilistic predictions - is used to train a set of models with twenty years of weekly multi-source and high-resolution environmental satellite data in nine countries of SEA.

The work stresses the importance of accounting for uncertainty in prediction when using ML and how this is needed for more actionable insights

for policy-making under rapidly changing climate conditions. To concretely show the added value of the probabilistic prediction of future crop health conditions, the paper examines the impact of weather shocks on household food expenditure and finds that the VHI predictions are useful to anticipate food price and expenditure shocks.

Results shows that the algorithm can accurately learn and reproduce the historical patterns of vegetation health based on weather variability, and thus successfully anticipate crop health under different prediction horizons. In a second step, a follow-up econometric analysis to evaluate the link between the vegetation health predictions and food security, relating the VHI dynamics with time series of national monthly rice prices is carried out. Results show that VHI predictions are suitable to anticipate food price and expenditure shocks.

Irrespective of this strong policy-support potential, a number of remarks must be discussed in relation to the results. First, while findings are robust and in line with the performance and results of the wider literature that links satellite imagery with crop health, a limitation of the analysis and its findings stems from the potential error in remotely sensed variables used in the analysis and generally in the data granularity. It also has to be noted that smooth effects between the set of variables under analysis are not well captured by RF hence any further analysis could benefit from the use of a local linear correction as proposed by Friedberg et al. (2020) but at the cost of higher computation and time requirements. Second, in the VHI shock analysis, the estimated household food expenditure increases are solely based on rice prices variations. Hence, they are best interpreted as lower-bound values of the overall impact of a VHI shock on food security. Conversely, these rice price growth impacts on food expenditure do not consider potential substitution dynamics, e.g. temporary switches to other foods less affected by a potential VHI shocks. In other words, the price elasticity of rice demand is assumed to be inelastic ($\epsilon = 0$).

Overall, this analysis contributes to the growing literature focused on using remotely sensed data and ML algorithms to improve food security analysis. The proposed framework sets the ground for the development of early warning food security systems inclusive of an uncertainty component that could support programs and policies aiming at boosting the resilience

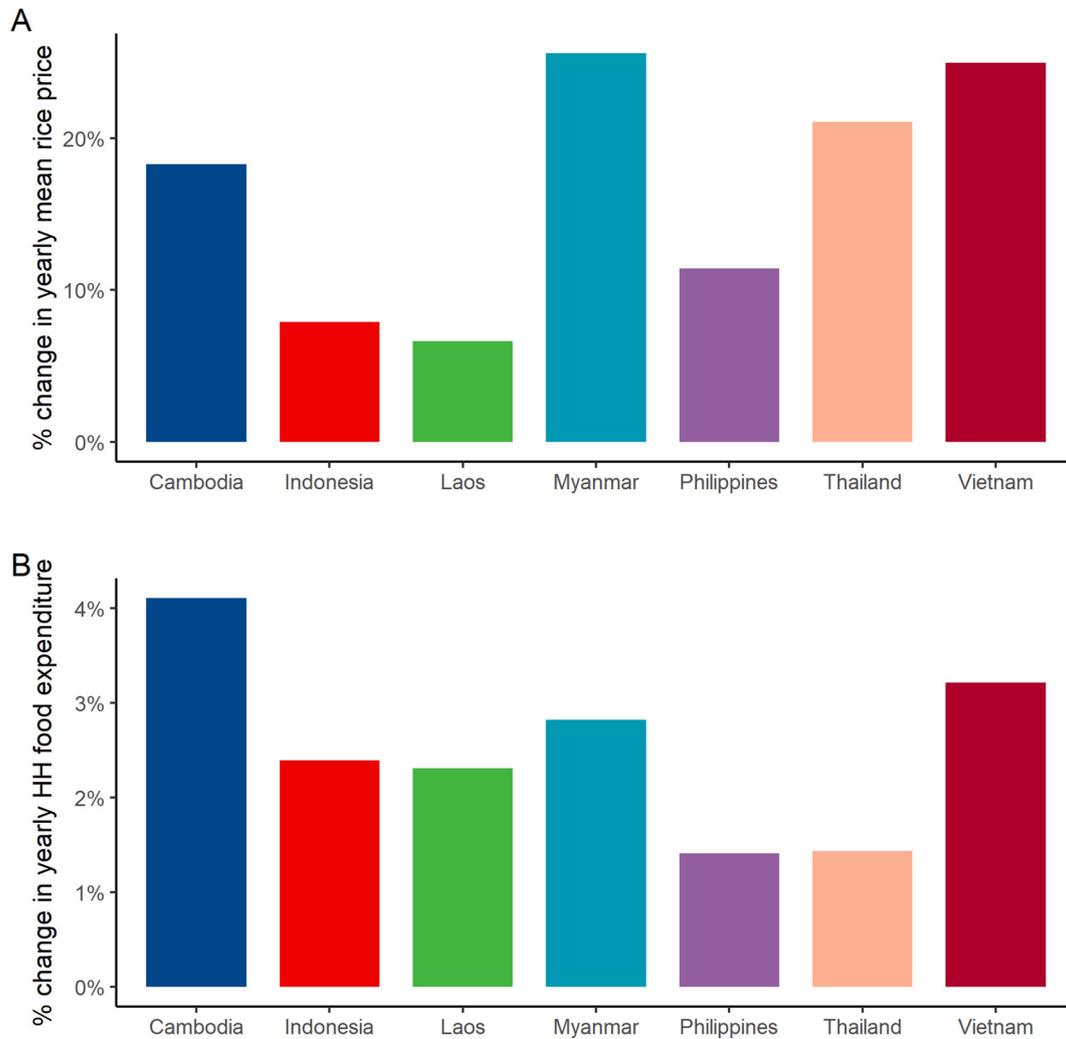


Fig. 5. Country-level impact of a VHI shock of magnitude S_c (see Methods) on (A) rice prices and (B) yearly household food expenditure.

and adaptation to climate change in vulnerable countries. In this perspective, future research could extend the here proposed framework to other geographies, consider additional crops, and represent food trade and crop substitution dynamics.

Code and data availability

The code to replicate the analysis and the figures are publicly hosted at https://github.com/athammad/VHI_SEA. The repository includes references to retrieve the input data.

CRediT authorship contribution statement

Ahmed T. Hammad: Conceptualization, Methodology, Formal analysis, Software, Visualization, Writing – original draft, Writing – review & editing. **Giacomo Falchetta:** Conceptualization, Methodology, Formal analysis, Software, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.scitotenv.2022.156157>.

References

- Aday, S., Aday, M.S., 2020. Impact of covid-19 on the food supply chain. *Food Qual. Saf.* 4, 167–180. <https://doi.org/10.1093/fqsafe/fyaa024>.
- Anderson, R., Bayer, P.E., Edwards, D., 2020. Climate change and the need for agricultural adaptation. *Curr. Opin. Plant Biol.* 56, 197–202. <https://doi.org/10.1016/j.pbi.2019.12.006>.
- Ansah, I.G.K., Gardebroeck, C., Ihle, R., 2021. Shock interactions, coping strategy choices and household food security. *Clim. Dev.* 13, 414–426. <https://doi.org/10.1080/17565529.2020.1785832>.
- Baniya, B., Tang, Q., Xu, X., Haile, G.G., Chhipi-Shrestha, G., 2019. Spatial and temporal variation of drought based on satellite derived vegetation condition index in Nepal from 1982–2015. *Sensors* 19, 430. <https://doi.org/10.3390/s19020430>.
- Bank, W., 2021. *Poverty Headcount Ratio at \$5.50 a Day (2011 ppp) (of Population)*.
- Black, J.R., Thompson, S.R., 1978. Some evidence on weather-crop-yield interaction. *Am. J. Agric. Econ.* 60, 540–543. <https://doi.org/10.2307/1239954>.
- Bosse, N., Abbott, S., EpiForecasts, F.S., 2020. *scoringutils: Utilities for Scoring and Assessing Predictions*.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32. https://doi.org/10.1007/978-3-662-56776-0_10.
- Brier, G.W., et al., 1950. Verification of forecasts expressed in terms of probability. *Mon. Weather Rev.* 78, 1–3. [https://doi.org/10.1175/1520-0493\(1950\)078<0001:vofeit>2.0.co;2](https://doi.org/10.1175/1520-0493(1950)078<0001:vofeit>2.0.co;2).
- Dowle, M., Srinivasan, A., 2021. *data.table: Extension of 'data.frame'*. R Package Version 1.14.2.

- Duan, T., 2020. Ngboost: natural gradient boosting for probabilistic prediction. In: PMLR (Ed.), *International Conference on Machine Learning*, vol. 119. PMLR, pp. 2690–2700.
- Eckstein, D., Künzel, V., Schäfer, L., 2021. Global climate risk index 2021. *Who Suff. Most From Extrem. Weather. Events 2000–2019*.
- Ehlers, M.-H., Huber, R., Finger, R., 2021. Agricultural policy in the era of digitalisation. *Food Policy* 100, 102019. <https://doi.org/10.1016/j.foodpol.2020.102019>.
- FAO, 2003. Food and Agriculture Organization of the United Nations Trade Reforms and Food Security.
- FAO, 2020. Covid-19 and the Risk to Food Supply Chains: How to Respond? <https://doi.org/10.4060/ca8388en>
- FAO, 2022. Faostat Food Balance Sheets.
- Filella, I., Penuelas, J., 1994. The red edge position and shape as indicators of plant chlorophyll content, biomass and hydric status. *Int. J. Remote Sens.* 15, 1459–1470. <https://doi.org/10.1080/01431169408954177>.
- Food & the United Nations, A.O., 1998. *Poverty Alleviation and Food Security in Asia: Lessons and Challenges*.
- Friedberg, R., Tibshirani, J., Athey, S., Wager, S., 2020. Local linear forests. *J. Comput. Graph. Stat.* 30, 1–15. <https://doi.org/10.1080/10618600.2020.1831930>.
- Frieler, K., et al., 2017. Understanding the weather signal in national crop-yield variability. *Earth's Future* 5, 605–616. <https://doi.org/10.1002/2016ef000525>.
- Funk, C., et al., 2015. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Sci. Data* 2, 1–21. <https://doi.org/10.1038/sdata.2015.66>.
- Funk, C., et al., 2019. Recognizing the famine early warning systems network: over 30 years of drought early warning science advances and partnerships promoting global food security. *Bull. Am. Meteorol. Soc.* 100, 1011–1027. <https://doi.org/10.1175/bams-d-17-0233.1>.
- Ghahramani, Z., 2015. Probabilistic machine learning and artificial intelligence. *Nature* 521, 452–459. <https://doi.org/10.1038/nature14541>.
- Gneiting, T., Katzfuss, M., 2014. Probabilistic forecasting. *Annu. Rev. Stat. Its Appl.* 1, 125–151. <https://doi.org/10.1146/annurev-statistics-062713-085831>.
- Gneiting, T., Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. *J. Am. Stat. Assoc.* 102, 359–378. <https://doi.org/10.1198/016214506000001437>.
- Gorelick, N., et al., 2017. Google earth engine: planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* 202, 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>.
- Grolemund, G., Wickham, H., 2011. Dates and times made easy with lubridate. *J. Stat. Softw.* 40, 1–25.
- Hasegawa, T., et al., 2021. Extreme climate events increase risk of global food insecurity and adaptation needs. *Nat. Food* 2, 587–595. <https://doi.org/10.1038/s43016-021-00335-4>.
- Heidinger, A., Foster, M., Walther, A., Zhao, X., Program, N., 2014. climate data record (cdr) of cloud properties from avhrr pathfinder atmospheres-extended (patmos-x), version 5.3. NOAA Natl. Centers for Environ. Information: NOAA CDR Program.
- Helton, J.C., Johnson, J.D., Oberkampf, W.L., Sallaberry, C.J., 2010. Representation of analysis results involving aleatory and epistemic uncertainty. *Int. J. Gen. Syst.* 39, 605–646. <https://doi.org/10.1080/03081079.2010.486664>.
- Henry, L., Wickham, H., 2020. purrr: functional programming tools. R Package Version 0.3.4.
- Hlavac, M., 2015. stargazer: Beautiful Latex, html and ascii Tables From r Statistical Output. Huetz, A.R., 1988. A soil-adjusted vegetation index (savi). *Remote Sens. Environ.* 25, 295–309. [https://doi.org/10.1016/0034-4257\(88\)90106-x](https://doi.org/10.1016/0034-4257(88)90106-x).
- Hyndman, R.J., Athanasopoulos, G., 2013. *Forecasting: Principles and Practice*.
- Jägermeyr, J., et al., 2021. Climate impacts on global agriculture emerge earlier in new generation of climate and crop models. *Nat. Food* 2, 873–885. <https://doi.org/10.1038/s43016-021-00400-y>.
- Jeong, S., Ko, J., Yeom, J.-M., 2022. Predicting rice yield at pixel scale through synthetic use of crop and deep learning models with satellite data in south and North Korea. *Sci. Total Environ.* 802, 149726. <https://doi.org/10.1016/j.scitotenv.2021.149726>.
- Kabir, H.D., Khosravi, A., Hosen, M.A., Nahavandi, S., 2018. Neural network-based uncertainty quantification: a survey of methodologies and applications. *IEEE Access* 6, 36218–36234. <https://doi.org/10.1109/access.2018.2836917>.
- Kamble, D.B., Gautam, S., Bisht, H., Rawat, S., Kundu, A., 2019. Drought assessment for kharif rice using standardized precipitation index (spi) and vegetation condition index (vci). *J. Agrometeorology* 21, 182–187. <https://doi.org/10.54386/jam.v21i2.230>.
- Karnieli, A., et al., 2006. Comments on the use of the vegetation health index over Mongolia. *Int. J. Remote Sens.* 27, 2017–2024. <https://doi.org/10.1080/01431160500121727>.
- Karthikeyan, L., Chawla, I., Mishra, A.K., 2020. A review of remote sensing applications in agriculture for food security: crop growth and yield, irrigation, and crop losses. *J. Hydrol.* 586, 124905. <https://doi.org/10.1016/j.jhydrol.2020.124905>.
- Kogan, F., 1987. Vegetation health index for areal analysis of ndvi in monitoring crop conditions. Preprint 18th Conf. on Agricultural and Forest Meteorology, pp. 103–114.
- Kogan, F.N., 1990. Remote sensing of weather impacts on vegetation in non-homogeneous areas. *Int. J. Remote Sens.* 11, 1405–1419. <https://doi.org/10.1080/01431169008955102>.
- Kogan, F.N., 1997. Global drought watch from space. *Bull. Am. Meteorol. Soc.* 78, 621–636. [https://doi.org/10.1175/1520-0477\(1997\)078<0621:gdwfs>2.0.co;2](https://doi.org/10.1175/1520-0477(1997)078<0621:gdwfs>2.0.co;2).
- Kogan, F., 2002. World droughts in the new millennium from avhrr-based vegetation health indices. *EOS Trans. Am. Geophys. Union* 83, 557–563. <https://doi.org/10.1029/2002EO000382>.
- Kogan, F., 2019. *Remote Sensing for Food Security*. Springer.
- Kogan, F., et al., 2004. Derivation of pasture biomass in Mongolia from avhrr-based vegetation health indices. *Int. J. Remote Sens.* 25, 2889–2896. <https://doi.org/10.1080/01431160410001697619>.
- Kogan, F., Yang, B., Wei, G., Zhiyuan, P., Xianfeng, J., 2005. Modelling corn production in China using avhrr-based vegetation health indices. *Int. J. Remote Sens.* 26, 2325–2336. <https://doi.org/10.1080/01431160500034235>.
- Kriegler, F., Malila, W., Nalepka, R., Richardson, W., 1969. Preprocessing transformations and their effects on multispectral recognition. *Remote Sens. Environ.* 97.
- Krueger, F., Lerch, S., Thorarindottir, T.L., Gneiting, T., 2016. Probabilistic Forecasting and Comparative Model Assessment Based on Markov Chain Monte Carlo Output. *arXiv preprint arXiv:1608.06802*, p. 12.
- Kubitzka, C., Krishna, V.V., Schulthess, U., Jain, M., 2020. Estimating adoption and impacts of agricultural management practices in developing countries using satellite data. A scoping review. *Agron. Sustain. Dev.* 40, 1–21. <https://doi.org/10.1007/s13593-020-0610-2>.
- Kuhn, M., 2008. Building predictive models in r using the caret package. *J. Stat. Softw.* 28, 1–26.
- Lees, T., Tseng, G., Atzberger, C., Reece, S., Dadson, S., 2022. Deep learning for vegetation health forecasting: a case study in Kenya. *Remote Sens.* 14, 698. <https://doi.org/10.3390/rs14030698>.
- Liang, L., et al., 2017. Long-term spatial and temporal variations of vegetative drought based on vegetation condition index in China. *Ecosphere* 8, e01919. <https://doi.org/10.1002/ecs2.1919>.
- Liu, H.Q., Huete, A., 1995. A feedback based modification of the ndvi to minimize canopy background and atmospheric noise. *IEEE Trans. Geosci. Remote Sens.* 33, 457–465. <https://doi.org/10.1109/36.377946>.
- Lybbert, T.J., Sumner, D.A., 2012. Agricultural technologies for climate change in developing countries: policy options for innovation and technology diffusion. *Food Policy* 37, 114–123. <https://doi.org/10.1016/j.foodpol.2011.11.001>.
- Marufah, U., Hidayat, R., Prasasti, I., 2017. Analysis of relationship between meteorological and agricultural drought using standardized precipitation index and vegetation health index. *IOP Conf. Ser.: Earth Environ. Sci.* 54, 012008. <https://doi.org/10.1088/1755-1315/54/1/012008>.
- Masitoh, F., Rusydi, A., 2019. Vegetation health index (vhi) analysis during drought season in brantas watershed. 012033IOP Conference Series: Earth and Environmental Science. vol. 389. IOP Publishing. <https://doi.org/10.1088/1755-1315/389/1/012033> IOP Publishing.
- Mbow, C., 2020. *Food Sec.* 66, 281–282. <https://doi.org/10.1080/10357718.2012.658620>.
- McKeown, A., Warland, J., McDonald, M., 2006. Long-term climate and weather patterns in relation to crop yield: a minireview. *Botany* 84, 1031–1036. <https://doi.org/10.1139/b06-080>.
- Meinshausen, N., Ridgeway, G., 2006. Quantile regression forests. *J. Mach. Learn. Res.* 7, 983–999.
- Mora, C., et al., 2015. Suitable days for plant growth disappear under projected climate change: potential human and biotic vulnerability. *PLoS Biol.* 13, e1002167. <https://doi.org/10.1371/journal.pbio.1002167>.
- Nay, J., Burchfield, E., Gilligan, J., 2018. A machine-learning approach to forecasting remotely sensed vegetation health. *Int. J. Remote Sens.* 39, 1800–1816. <https://doi.org/10.1080/01431161.2017.1410296>.
- Nazir, A., et al., 2021. Estimation and forecasting of rice yield using phenology-based algorithm and linear regression model on sentinel-ii satellite data. *Agriculture* 11, 1026. <https://doi.org/10.3390/agriculture11101026>.
- OECD, F., 2020. *Oecd-fao Agricultural Outlook 2020–2029*. <https://doi.org/10.1787/1112c23b-en>.
- Orlovsky, L., Kogan, F., Eshed, E., Dugarjav, C., 2010. Monitoring droughts and pastures productivity in mongolia using noaa-avhrr data. Use of Satellite and In-Situ Data to Improve Sustainability. Springer, Dordrecht, The Netherlands, pp. 69–79. https://doi.org/10.1007/978-90-481-9618-0_8.
- Ortiz-Bobea, A., Ault, T.R., Carrillo, C.M., Chambers, R.G., Lobell, D.B., 2021. Anthropogenic climate change has slowed global agricultural productivity growth. *Nat. Clim. Change* 11, 306–312. <https://doi.org/10.1038/s41558-021-01000-1>.
- Palmer, T., 2012. Towards the probabilistic earth-system simulator: a vision for the future of climate and weather prediction. *Q. J. R. Meteorol. Soc.* 138, 841–861. <https://doi.org/10.1002/qj.1923>.
- Parayil, G., 1992. The green revolution in India: a case study of technological change. *Technol. Cult.* 33, 737–756. <https://doi.org/10.2307/3106588>.
- Pedersen, T.L., 2020. *patchwork: The Composer of Plots*.
- Pei, F., et al., 2018. Monitoring the vegetation activity in China using vegetation health indices. *Agric. For. Meteorol.* 248, 215–227. <https://doi.org/10.1016/j.agrformet.2017.10.001>.
- Perera, S., Li, W., Linstead, E., El-Askary, H., 2020. Forecasting vegetation health in the mena region by predicting vegetation indicators with machine learning models. In: IEEE (Ed.), *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, pp. 4642–4645. <https://doi.org/10.1109/igarss39084.2020.9323747>.
- Porter, J.R., 2014. Food Security and Food Production Systems. <https://doi.org/10.1017/cbo9781107415379.012>.
- Rahman, A., Roytman, L., Krakauer, N.Y., Nizamuddin, M., Goldberg, M., 2009. Use of vegetation health data for estimation of aus rice yield in Bangladesh. *Sensors* 9, 2968–2975. <https://doi.org/10.3390/s90402968>.
- Rasmussen, W.D., 1962. The impact of technological change on American agriculture, 1862–1962. *J. Econ. Hist.* 22, 578–591. <https://doi.org/10.1017/S0022050700066778>.
- Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., 1974. Monitoring Vegetation Systems in the Great Plains With Ertis. 351. NASA special publication, p. 309.
- Saha, S., 2011. Ncep climate forecast system version 2 (cfsv2) 6-hourly products. Res. Data Arch. at Natl. Cent. for Atmospheric Res. Comput. Inf. Syst. Lab.
- Salakpi, E.E., et al., 2021. Forecasting vegetation condition with a bayesian auto-regressive distributed lags (bardl) model. *Nat. Hazards Earth Syst. Discuss.* 1–31. <https://doi.org/10.5194/nhess-2021-223>.
- Salazar, L., Kogan, F., Roytman, L., 2007. Use of remote sensing data for estimation of winter wheat yield in the United States. *Int. J. Remote Sens.* 28, 3795–3811. <https://doi.org/10.1080/01431160601050395>.
- Sanderman, J., et al., 2018. A global map of mangrove forest soil carbon at 30 m spatial resolution. *Environ. Res. Lett.* 13, 055002. <https://doi.org/10.1088/1748-9326/aabe1c>.
- Shannugapriya, P., Rathika, S., Ramesh, T., Janaki, P., 2019. Applications of remote sensing in agriculture—a review. *Int. J. Curr. Microbiol. Appl. Sci.* 8, 2270–2283. <https://doi.org/10.20546/ijcmas.2019.801.238>.

- Sishodia, R.P., Ray, R.L., Singh, S.K., 2020. Applications of remote sensing in precision agriculture: a review. *Remote Sens.* 12, 3136. <https://doi.org/10.3390/rs12193136>.
- Son, N.-T., et al., 2020. Machine learning approaches for rice crop yield predictions using time-series satellite data in Taiwan. *Int. J. Remote Sens.* 41, 7868–7888. <https://doi.org/10.1080/01431161.2020.1766148>.
- STAR, N., 2018. Global vegetation health products. *Cent. for Satell. Appl. Res. NOAA, USA*.
- Sullivan, R.J., 1984. Measurement of English farming technological change, 1523–1900. *Explor. Econ. Hist.* 21, 270–289. [https://doi.org/10.1016/0014-4983\(84\)90009-3](https://doi.org/10.1016/0014-4983(84)90009-3).
- Taillardat, M., Mestre, O., Zamo, M., Naveau, P., 2016. Calibrated ensemble forecasts using quantile regression forests and ensemble model output statistics. *Mon. Weather Rev.* 144, 2375–2393. <https://doi.org/10.1175/mwr-d-15-0260.1>.
- Team, R.C., 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Teluguntla, P., 2015. Global Cropland Area Database (gcad) Derived From Remote Sensing in Support of Food Security in the Twenty-first Century: Current Achievements and Future Possibilities.
- The Economist, 2021. Global Food Security Index 2021: an annual measure of the state of global food security. *Economist Impact*. <https://impact.economist.com/sustainability/project/food-security-index/Resources>.
- Trnka, M., et al., 2016. Changing regional weather crop yield relationships across Europe between 1901 and 2012. *Clim. Res.* 70, 195–214. <https://doi.org/10.3354/cr01426>.
- Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* 8, 127–150. [https://doi.org/10.1016/0034-4257\(79\)90013-0](https://doi.org/10.1016/0034-4257(79)90013-0).
- Vaysse, K., Lagacherie, P., 2017. Using quantile regression forest to estimate uncertainty of digital soil mapping products. *Geoderma* 291, 55–64. <https://doi.org/10.1016/j.geoderma.2016.12.017>.
- Vroege, W., Vrieling, A., Finger, R., 2021. Satellite support to insure farmers against extreme droughts. *Nat. Food* 2, 215–217. <https://doi.org/10.1038/s43016-021-00244-6>.
- Wan, Z., Hook, S., Hulley, G., 2015. *Mod11a1 modis/terra Land Surface Temperature/emissivity Daily 13 Global 1 km Sin Grid v006 [Data Set]*. nasa eosdis lp daac.
- Wang, C., et al., 2021. Occurrence of crop pests and diseases has largely increased in China since 1970. *Nat. Food* 3, 1–9. <https://doi.org/10.1038/s43016-021-00428-0>.
- Weiss, M., Jacob, F., Duveiller, G., 2020. Remote sensing for agricultural applications: a meta-review. *Remote Sens. Environ.* 236, 111402. <https://doi.org/10.1016/j.rse.2019.111402>.
- Wen, W., Timmermans, J., Chen, Q., van Bodegom, P.M., 2021. A review of remote sensing challenges for food security with respect to salinity and drought threats. *Remote Sens.* 13, 6. <https://doi.org/10.3390/rs13010006>.
- Westerveld, J.J., et al., 2021. Forecasting transitions in the state of food security with machine learning using transferable features. *Sci. Total Environ.* 786, 147366. <https://doi.org/10.1016/j.scitotenv.2021.147366>.
- Wickham, H., 2016. *ggplot2: Elegant Graphics for Data Analysis*. vol. 174. Springer-Verlag, New York.
- Wickham, H., et al., 2019. Welcome to the tidyverse. *J. Open Source Softw.* 4, 1686. <https://doi.org/10.21105/joss.01686>.
- Woodhouse, I.H., 2017. *Introduction to Microwave Remote Sensing*. CRC press.
- world-bank-official-boundaries.
- Wright, M.N., Ziegler, A., 2015. ranger: A Fast Implementation of Random Forests for High Dimensional Data in c++ and r. arXiv preprint arXiv:1508.04409 <https://doi.org/10.18637/jss.v077.i01>.
- Xiao, N., Li, M., Xiao, M.N., LazyData, T., 2018. Package 'ggsci'.
- Xue, J., Su, B., 2017. Significant remote sensing vegetation indices: a review of developments and applications. *J. Sensors* 1–17, 2017. <https://doi.org/10.1155/2017/1353691>.
- Zhongming, Z., Linong, L., Wangqiang, Z., Wei, L., 2021. *Weather and Climate Extremes in Asia Killed Thousands, Displaced Millions and Cost Billions in 2020*.
- Zuhro, A., Tambunan, M.P., Marko, K., 2020. Application of vegetation health index (vhi) to identify distribution of agricultural drought in indramayu regency, west java province. *IOP Conf. Ser.: Earth Environ. Sci.* 500, 012047. <https://doi.org/10.1088/1755-1315/500/1/012047>.