Science of the Total Environment 869 (2023) 161716

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Exploring the use of Sentinel-2 datasets and environmental variables to model wheat crop yield in smallholder arid and semi-arid farming systems

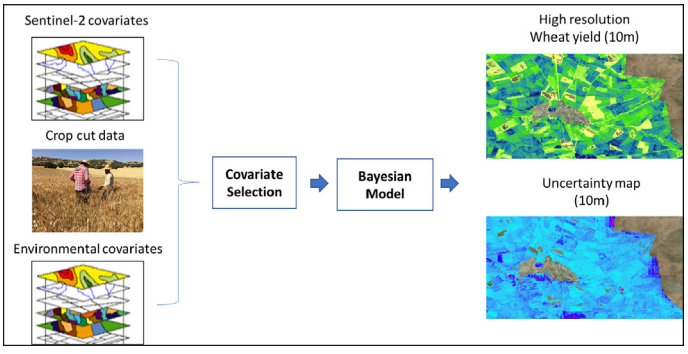
Sarchil Hama Qader [a,b,*], Chigozie Edson Utazi [a], Rhorom Priyatikanto [a,c], Peshawa Najmaddin [b], Emad Omer Hama-Ali [d], Nabaz R. Khwarahm [e], Andrew J. Tatem [a], Jadu Dash [a]

[a] School of Geography and Environmental Science, University of Southampton, Southampton, UK
[b] Natural Resources Department, College of Agricultural Engineering Sciences, University of Sulaimani, Sulaimani, Kurdistan Region, Iraq
[c] Research Center for Space, National Research and Innovation Agency, Bandung 40173, Indonesia
[d] Biotechnology and Crop Science Department, College of Agricultural Engineering Sciences, University of Sulaimani, Sulaimani 46001, Kurdistan Region, Iraq
[e] Department of Biology, College of Education, University of Sulaimani, Sulaimani 46001, Kurdistan Region, Iraq

HIGHLIGHTS

- Agriculture ground data is often scarce in conflict zones
- Yield variability is mapped at 10m using Sentinel-2 data
- Yield is estimated using Bayesian model trained with harvested data
- Environmental variables can strengthen the model predicyive power

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Martin Drews

ABSTRACT

Low levels of agricultural productivity are associated with the persistence of food insecurity, poverty, and other socio-economic stresses. Mapping and monitoring agricultural dynamics and production in real-time at high spatial resolution are essential for ensuring food security and shaping policy interventions. However, an accurate yield estimation might be challenging in some arid and semi-arid regions since input datasets are generally scarce, and access is restricted due to security challenges. This work examines how well Sentinel-2 satellite sensor-derived data, topographic and climatic variables, can be used as covariates to accurately model and predict wheat crop yield at the farm level using statistical models in low data settings of arid and semi-arid regions, using Sulaimani governorate in Iraq as an example. We developed a covariate selection procedure that assessed the correlations between the covariates and their relationships with wheat crop yield. Potential non-linear relationships were investigated in the latter case using regression splines. In the absence of substantial non-linear relationships between the covariates and crop yield, and residual spatial autocorrelation, we fitted a Bayesian multiple linear regression model to model and predict crop yield at 10 m resolution. Out of the covariates tested, our results showed significant relationships between crop yield and mean cumulative NDVI during the growing season, mean elevation, mean end of the season, mean maximum temperature and mean the start of the season at the farm level. For in-sample prediction, we estimated an R^2 value of 51 % for the model, whereas for out-of-sample prediction, this was 41 %, both of which indicate reasonable predictive performance. The calculated root-mean-square error for out-of-sample prediction was 69.80, which is less than the standard deviation of 89.23 for crop yield, further showing that the model performed well by reducing prediction variability. Besides crop yield estimates, the model produced uncertainty metrics at 10 m resolution. Overall, this study

* Corresponding author at: School of Geography and Environmental Science, University of Southampton, Southampton, UK.
E-mail address: S.Qader@soton.ac.uk (S.H. Qader).

http://dx.doi.org/10.1016/j.scitotenv.2023.161716
Received 13 September 2022; Received in revised form 3 January 2023; Accepted 15 January 2023
Available online 20 January 2023

showed that Sentinel-2 data can be valuable for upscaling field measurement of crop yield in arid and semi-arid regions. In addition, the environmental covariates can strengthen the model predictive power. The method may be applicable in other areas with similar environments, particularly in conflict zones, to increase the availability of agricultural statistics.

1. Introduction

Over recent decades, humanity has faced several natural and anthropogenic related events which have posed a major risk to current and future global food production. The consequences of these events have escalated the level of poverty on national and global scales. In particular, the impact of the recent Covid-19 pandemic has increased global poverty rates from 8.4 % in 2019 to 9.3 % in 2020 resulting in pushing >70 million people into extreme poverty by the end of 2020 (World Bank, 2022a). One of the most powerful tools to eradicate poverty and feed a projected 9.7 billion people by 2050 is agricultural development (United Nations, 2019; World Bank, 2020). In 2021, agriculture accounted for 4.3 % of the global gross domestic product (GDP), however, its contribution to the total GDP was higher than 25 % in many developing countries (World Bank, 2022b). In addition, more than two-thirds of the population in poor countries work in agriculture (Roser, 2013). Therefore, increasing agricultural productivity could alleviate global hunger and food insecurity (Uphaus, 2008). However, there is inconsistency in the availability of input data among countries to help improve agriculture productivity.

Currently, the lack of full coverage of ground agriculture data and difficulties with access to many areas in the developing world due to security issues and other obstacles mean that traditional approaches to collecting agricultural data are challenging (Eklund et al., 2017; Qader et al., 2021). For example, obtaining survey estimates at the farm level for an entire district requires a large number of crops cut data which is highly resource-intensive (Murthy et al., 1997; Lobell et al., 2018). In addition, with the increasing need for improving food insecurity and livelihood issues, reliable farm-level estimates of yield are becoming ever more important (Sapkota et al., 2016; Karst et al., 2020). Commonly, farmer self-reports and crop cuts are used to measure crop productivity. Although the self-report approach is much easier and faster for collecting crop data, the reliability of the approach can be questionable (Baumeister et al., 2007). These concerns are due to the lack of standard measurement units, various conditions at which the crop is harvested, a tendency to provide inaccurate numbers and potential bias (Carletto et al., 2015; Gourlay et al., 2019; Wahab, 2020). A second common approach to measure crop production is crop cut, where yield is harvested from a randomly selected portion of a farmer's land (Fermont and Benson, 2011). However, adopting such an approach is resource and time intensive, and it can lead to uncertainties since yield can exhibit large spatial heterogeneities within farmland (Lobell et al., 2020; Paliwal and Jain, 2020). In addition, Kosmowski et al. (2021) found that protocol choices to select samples are important as alternative protocols could result in various accuracies relative to the whole plot. Therefore, neither self-report nor crop cut approaches alone are able to offer a reliable complete picture of crop productivity because of their limitations and incomplete spatial coverage. In addition, such data collection might not be possible in some areas due to limited resources, security challenges and access issues.

Advances in remote sensing technology present opportunities for innovation in local and global agricultural statistics systems (Carfagna and Gallego, 2005; GSARS, 2017). However, not all the regions in the world are able to benefit from such technological advancement. In Sub-Saharan Africa, Bégué et al. (2020) concluded that to take advantage from this technological advancement and bridge the gap between technical analysts and policy makers some key points such as capacity building, public-private partnership, political will, proofs of concept and institutional commitment are fundamental. With the combination of satellite-derived information and ground data, statistical approaches can be used to predict crop yield at the farm level (Burke and Lobell, 2017; Engen et al., 2021; Segarra et al., 2022). For instance, a combination of high-resolution satellite imagery and field data collected from thousands of smallholder plots in Kenya showed that satellite imagery can be used to estimate and understand yield variation at the field scale and provide potential capabilities that include a broader characterization of the source and extent of yield gaps and the development of financial products aimed at African smallholders (Burke and Lobell, 2017). Recently, the potential of Sentinel-2 and PlanetScope were evaluated to estimate maize yield in intercropped smallholder fields in Malawi (Li et al., 2022). The results indicate that maize yield with moderate accuracy (R^2 = 0.51, nRMSE = 19.95 %) can be estimated using Sentinel-2 red-edge vegetation index (VI), while PlanetScope (3 m) only showed a marginal improvement in performance (R^2 = 0.52, nRMSE = 19.95 %). However, remote sensing-based crop yield models require either crop cut data or comparable high-quality measurements to produce better prediction results, than low-quality training datasets (Sida et al., 2021). Quantifying farm-level crop production in advance could help policymakers, scientists, and decision-makers to improve agricultural management and food security under a variety of environmental conditions.

In general, remote sensing data can be used for yield estimation in two ways. First as an input to crop models where the remotely sensed information is incorporated into simulating models for crop development and growth. Examples of these models are the World Food Studies (WOFOST) (Vandiepen et al., 1989), Crop Systems Simulation (CROPSYST) models (Van Evert and Campbell, 1994) and Simulateur mulTIdisciplinaire pour les Cultures Standard (STICS) (Brisson et al., 1998). Despite their ability to obtain the soil-environment-plant interactions, these models are generally challenging to calibrate particularly in data-sparse environments due to their complexity and high data demand (Moriondo et al., 2007). Other researchers have combined remote sensing and crop growth models in case of ground data scarce environments, where outputs of crop models were used to calibrate remote sensing-based models (Lobell et al., 2015; Leroux et al., 2019). On the other hand, yield prediction can be made directly from remotely sensed measurements through biophysical variables quantified within the season and their relationship with ground yield data using simple linear models (Mkhabela et al., 2011; Huang et al., 2013; Bolton and Friedl, 2013). These approaches are widely used particularly in developing countries because of their low demand for data and simplicity of implementation.

Researchers have employed multiple satellite datasets such as those produced from the Advanced Very High-Resolution Radiometer (AVHRR) (Doraiswamy and Cook, 1995), Moderate Resolution Imaging Spectroradiometer (MODIS) (Lopresti et al., 2015), Medium Resolution Imaging Spectrometer (MERIS) (Dash and Curran, 2007) and Landsat (Kang and Özdoğan, 2019) to map crop yield in different locations across the world. However, considering the common cloud coverage issues and lack of temporal or spatial resolutions of these datasets, mapping smallholder yield is a challenge. Whitcraft et al. (2015) suggested that at early and mid-agricultural season, which are crucial periods for crop type mapping and crop yield forecasting, optical, polar-orbiting imaging might not be efficient for operational monitoring due to cloud issues, and alternatives (e.g., microwave synthetic aperture radar, SAR) should be considered. With the improvement of satellite products, several attempts have been made to explore the possibility of using remote sensing data to map crop yield and its variability within smallholdings (Jain et al., 2013; Burke and Lobell, 2017). Although the outputs of these studies are promising, many of them depend on commercial satellite data (Burke and Lobell, 2017) or a combination of freely available and commercial satellite data (Jeffries et al., 2020). Considering the financial aspect and limited resources, particularly in developing countries, obtaining such data is infeasible. In addition, using remote sensing data to map crop yield has been recognized as

a data-intensive process. Furthermore, this process is more challenging where agricultural practices are heterogeneous across neighbouring fields (Jain et al., 2013) and field sizes are typically <2 ha (Lowder et al., 2016).

The recent improvements in temporal, spatial and spectral resolutions from sensors such as the Sentinel-2 Multispectral imager (MSI) sensor datasets with its global coverage and free access have opened new possibilities for monitoring small smallholder farms. Recent studies have shown the potential of Sentinel-2 to estimate crop yield (Lambert et al., 2018; Hunt et al., 2019; Mehdaoui and Anane, 2020; Segarra). In a comparative analysis between Sentinel 2 and Landsat 8 in semi-arid region of Iraq, Sentinel-2 derivative covariates produced the strongest relationship with the actual grain yield (R^2 = 0.77) (Faqe Ibrahim et al., 2022). In other examples of semi-arid smallholder schemes, various data configuration of Sentinel-1, Sentinel-2 and their derivatives were tested to obtain an accurate yield estimation and the results are promising (Ouattara et al., 2020; Chahbi et al., 2022).

Despite providing predictive crop yield parameters, satellite images can be used to derive other environmental factors such as elevation, slop and climatic variables which are directly and indirectly related to the land productivity. For instance, high-altitude areas have similar characteristics as high latitudes including low temperature, high wind velocity and poor soil which constraint the crop productivity (Dhillon, 2004). In addition, soil formation, soil-water availability, climate and water drainage can be influenced by the slope of landscape and ultimately affect the land fertility. Furthermore, crop production is heavily affected by climatic condition including long-term trends in average temperature and rainfall, extreme weather events, interannual climate variability and shocks during growing period (FAO, 2013; IPCC, 2022). These environmental factors could be employed in the remote sensing-based crop yield modelling to improve the predictive power. In the UK, an accurate map of within wheat yield variation at 10 m resolution was possible using Sentinel-2 data (RMSE 0.66 t/ha), and when environmental data were incorporated, further improvement in the accuracy was achieved ((RMSE 0.61 t/ha) (Hunt et al., 2019). In the complex land use where crops are intercropped under a dense tree canopy in Burkina Faso, Sentinel-2 data was able to enlighten between 41 and 80 % of the variation in the ground crop production measurement (Karlson et al., 2020). Similarly, the inclusion of parkland structure has improved the explanation of observed yield variability from 46 % to 70 % in Senegal's smallholder context (Leroux et al., 2020). However, Sentinel-2s potential and the inclusion of environmental factors to map crop yield in areas that are prone to drought and known as conflict zones have not been fully explored.

With the advantage of remote sensing techniques, various machine learning models have been studied for crop yield estimation using satellite images at various spatial scales (Cunha and Silva, 2020; Xu et al., 2021; Khaki et al., 2021; Pham et al., 2022). Based on recent systematic literature reviews, Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN) are the most widely used deep learning approaches for crop yield prediction (Oikonomidis et al., 2022; Muruganantham et al., 2022). However, to the best of our knowledge, no study has been carried out to investigate the potential of a combination of Sentinel-2 and environmental data to model and predict crop yield in arid and semi-arid regions of Middle Eastern countries using a robust statistical methodology. Bayesian modelling approaches, in particular, have a greater capacity to accurately capture uncertainty compared to the more common discriminative models in machine learning (Shirley et al., 2020). An empirical Bayesian approach was applied to the state-level maize yield and meteorological data for the US Corn Belt from 1981 to 2014 and the results showed that temperature and precipitation had the largest impact on yield in the six months prior to the harvest (Shirley et al., 2020). A county-scale corn yield prediction model was built in the US based on a Bayesian neural network (BNN), and they found that the proposed BNN model had the best performance among all the six approaches for the ten testing years 2010–2019 achieving an average R^2 of 0.77 (Ma et al., 2021). In addition, besides its accurate corn yield estimation in a normal year, the model was able to accurately estimate corn yield in abnormal years when extreme weather events happened.

For the first time, here we explore the potential of various input datasets including crop phenology and different vegetation indices derived from Sentinel-2, climatic and topographic data, as well as crop cut to estimate wheat crop yield at 10 m resolution. In addition, we adopt a Bayesian modelling approach to predict crop yield and associated uncertainties in arid and semi-arid regions of Iraq where extreme weather events and security issues are prevalent. Furthermore, the research investigated the added value of environmental covariates to improving the predictive power.

2. Materials and methods

2.1. Study area

Iraq consists of eighteen administrative governorates. This research focuses on the Sulaimani governorate, which is located in the northeastern part of Iraq and is the largest governorate in the Kurdistan Region, bordering Iran from the east (Fig. 1a). Its elevation ranges from 176 to 3411 m above sea level (Fig. 1b). It covers an administrative area of 17,023 km^2 (Zakaria et al., 2013) with ten main districts including Pshdar, Rania, Dokan, Sharbazher, Panjwin, Al-Sulaymaniyah, Chamchamal, Derbendikhan, Halabja and Kalar (Fig. 1) (HDX, 2022). The climate of the Sulaimani is driven by the Mediterranean zone with hot and dry summer and cold and wet winter. The annual average temperature and rainfall ranging from 12 to 20 °C and from 400 to 1000 mm, respectively.

2.2. Data and pre-processing

For the model development and validation, a large set of variables were derived from five types of data sources including time-series Sentinel-2 imagery, land cover, climatic data, topographic data and crop cuts. A summary of the input datasets is presented in Table 1.

2.2.1. Sample stratification and crop masking

Crop yield can exhibit large spatial heterogeneity within a region due to variations in ecological zones, climatic, topographic conditions and agriculture management. To assure the randomness within the crop cut data collection and sample representativeness, the study area was stratified. Stratified sampling with strata for this work was defined by the combination of various attributes, including vegetation, topographic and climatic variables. To do this, various input datasets were extracted and pre-processed. The monthly rainfall data from October 2019 to July 2020 were extracted from the center for Hydrometeorology and Remote Sensing (CHRIS) (Nguyen et al., 2019). The average and cumulative rainfall for the corresponding period were computed. The mean, minimum and maximum temperatures over the study period were extracted from MOD11B3 Version 6 product (Wan et al., 2015). The average NDVI over the study period was extracted from the NOAA Climate Data Record (CDR) of Advanced Very High-Resolution Radiometer (AVHRR) Surface Reflectance (Vermote et al., 2014). The Shuttle Radar Topography Mission (SRTM) was used to derive topographic variables including Digital Elevation Model and slope (NASA, 2013). All the datasets were pre-processed and standardised in terms of spatial resolution and spatial extent. The nearest neighbour resampling approach was employed to resample all the datasets to 5 km spatial resolution for the purpose of sample stratification.

To generate the crop mask for the study area, the GlobeLand30, which is a 30 m resolution global land cover data product developed by the National Geomatics Center of China, was obtained from the GlobeLand30 website (GlobeLand30, 2022) (Fig. 2b). The GlobeLand30 datasets encompass 10 land cover classes in total, including cultivated land, grassland, shrubland, water bodies, forest, wetland, artificial surface, tundra, ice, bare land and perennial snow.

Once the input datasets were ready and standardised, all of them were compiled and the k-means unsupervised classifier (Lloyd, 1957) was used to stratify the study area into five strata (Fig. 2a). These five sample strata were considered during crop data collection to make sure the collected crop cut data represents the study area.

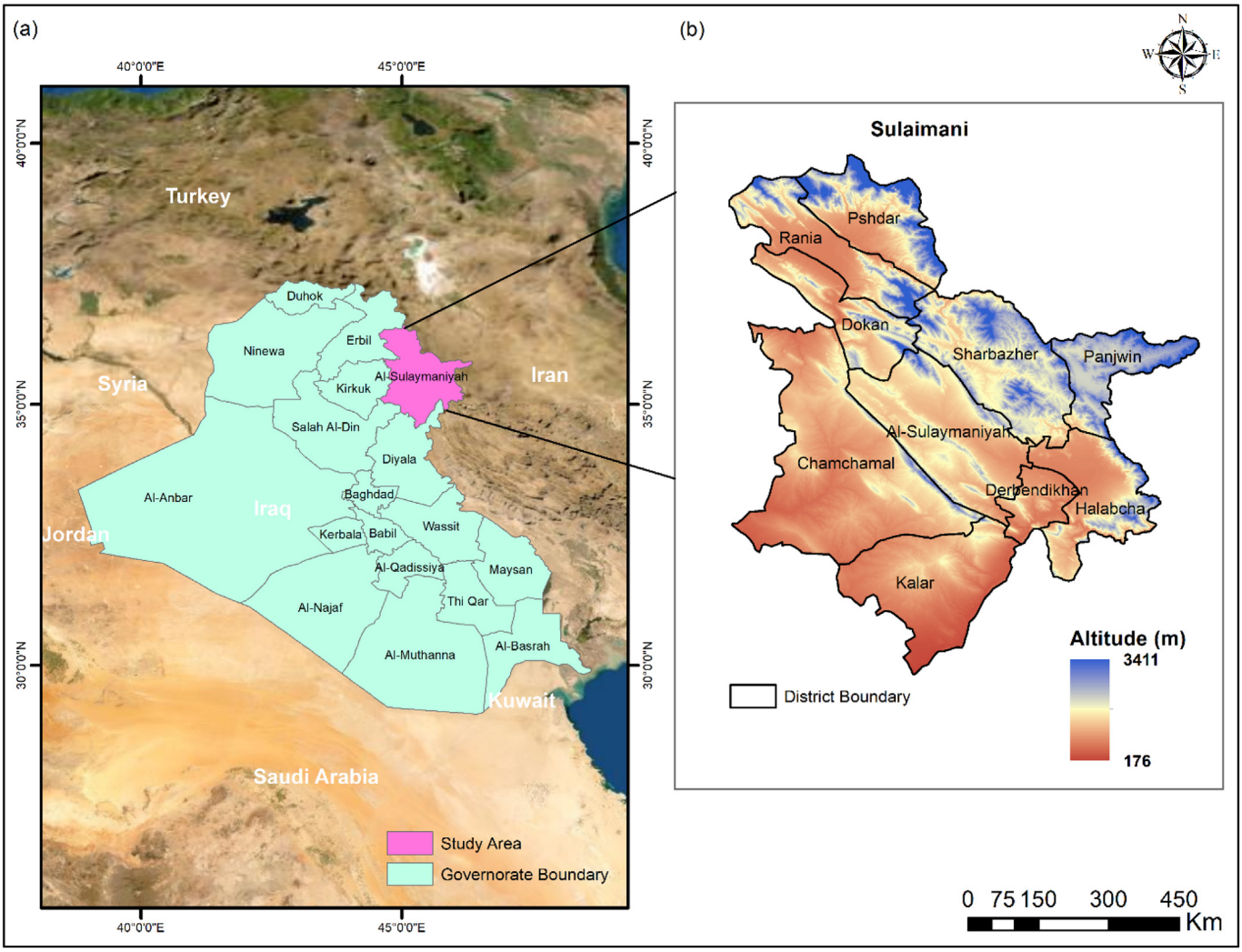


Fig. 1. The boundary of Iraqi governorates where the study area is highlighted (a), and the ten main districts in the study area (HDX, 2022) with their altitudes (NASA, 2013) (b).

2.2.2. Crop cut data

Crop cut data were collected for 138 randomly selected fields across our study area in 2020. To collect crop data, the field team visited each farmer's field at the time of crop harvest and selected a 1 × 1 m^2 plot within each field, harvesting the crop from these sub-plots, and weighting the crop grain (yield) in the field. The lack of resources limited us to conduct more than one crop cut in each field. However, in 17 farm fields, the crop cuts were conducted by selecting three 1 × 1 m^2 sup-plots at random from each farmer's field to assess the spatial variation of the yield in the selected field. To compute the mean yield of a field, the yield values from all 1 × 1 m^2 crop cut plots were averaged within each field. In addition, five GPS points for each of these fields were collected: four in each corner of the field and one at the center of the field. The collected GPS locations were used to create polygons for each farm and spatially linked to Sentinel imagery.

To assess the accuracy of the yield data that were collected through the crop cutting experiments, the average yield for each local area was compared to the agricultural statistics in Sulaimani governorate. The Kurdistan Region Statistics Office (KRSO) provides average yield in different local areas in the Sulaimani governorate including Sulaimani, Halabja and Garmian administrations.

Table 1
Summary of the source of input datasets.

Category	Variables	Spatial resolution	Source	Temporal resolution	Purpose
Satellite imagery	$NDVI_1$	0.05°	AVHRR	Average Year	Stratification
	$NDVI_2$	10 m	Sentinel-2	5 days	Crop yield Estimation
	EVI	10 m	Sentinel-2	5 days	Crop yield estimation
	MTCI	10 m	Sentinel-2	5 days	Crop yield estimation
	LST	5600 m	MODIS	Monthly	Stratification/ Crop yield estimation
Climatic	Rainfall	0.05°	CHRIS	Monthly	Stratification/ Crop yield estimation
Topographic	Elevation	30 m	SRTM	–	Stratification/ Crop yield estimation
	Slope	30 m	SRTM	–	Stratification/ Crop yield estimation
Crop cut	Yield	1 m	field	–	Crop yield estimation

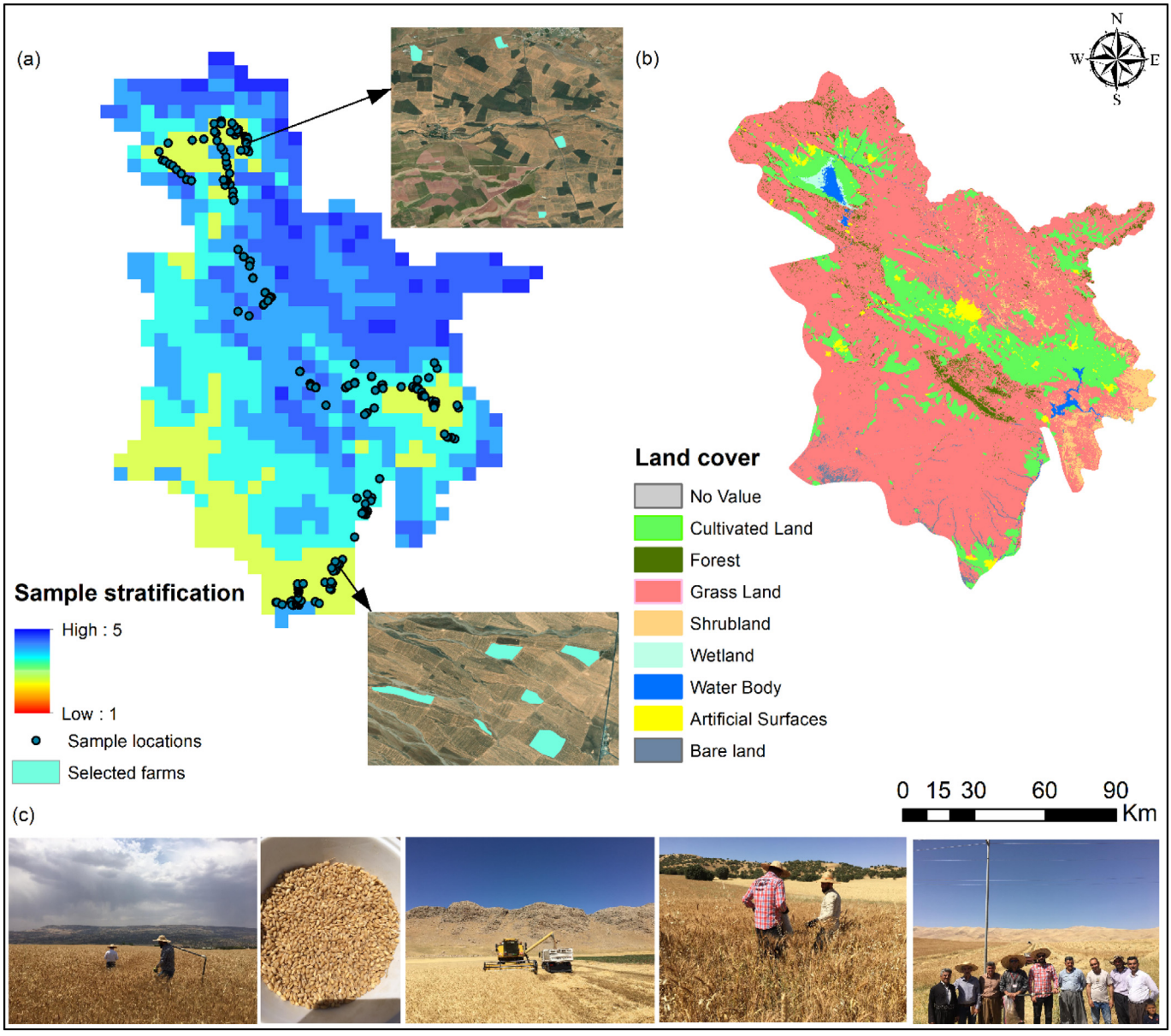


Fig. 2. (a) sample stratification for the study area and sample locations, (b) land cover type with a spatial resolution of 30 m for 2020 (GlobeLand30, 2022), (c) field photos during crop cut data collection.

2.2.3. Sentinel data and crop phenology

A collection of imagery from the Sentinel-2 MultiSpectral Instrument (MSI) (Drusch et al., 2012) was used in this study. Since 2015, Sentinel-2A dan 2B have been providing high-spatial resolution multi-band images with an average revisit time of 5 days, when combined. This satellite constellation enables environmental monitoring with diverse purposes, including crop type classification (Campos-Taberner et al., 2019), yield prediction (Hunt et al., 2019) and precision agriculture planning (Segarra et al., 2020). MSI captures the Earth's surface in 12 spectral bands ranging from the blue (443 nm) to the short-wave infrared part of the electromagnetic spectrum (2190 nm). Typical Sentinel-2 images for crop monitoring have 10–20 m resolution which is adequate for field-level monitoring though smallholder farmer fields may require a higher resolution. To be noted that most of the wheat crops in Sulaimani region have an area of >0.5 ha (more specifically, the median area of the 138 selected fields is 0.7 ha) such that the Sentinel-2 dataset is sufficient. The high temporal resolution of the dataset provided by Sentinel-2 provides a broader opportunity for crop phenology studies which can be exploited for crop monitoring with higher precision (Segarra et al., 2020).

The Sulaimani Governorate is covered by 6 tiles of Sentinel 2, namely tiles 38SMD, 38SME, 38SMF, 38SND, 38SNE, and 38SNF. Those tiles are based on the US-Military Grid Reference System (MGRS). Sentinel-2 has been providing data since June 2015 and the images from 2019 to 07-31 to 2020-10-31 were acquired for the current study. This time window was determined to capture the cropping season of the main commodities, including possible early start and late harvest dates.

Initial quality assurance was conducted by using the cloud mask band (QA10) provided by the Sentinel-2 data. Besides pixel-wise masking, tile-wise filtering was also performed by omitting the tiles with a cloud cover percentage of <30 %. These initial filtering steps produced time-series data containing 50–300 data points stored in every pixel.

To capture the greenness of the crops, several vegetation indices were derived from top-of-atmosphere reflectance images. They are the normalized difference vegetation index (NDVI) (Rouse et al., 1973), Enhanced Vegetation Index (EVI) (Huete et al., 2002), and Medium Resolution Imaging Spectrometer (MERIS) Terrestrial Chlorophyll Index (MTCI) (Dash and Curran, 2004). NDVI is the most common index for vegetation monitoring though this index tends to saturate at regions with high biomass (Huete

et al., 2002). EVI provides a better sensitivity at high biomass by adjusting the canopy background signal and utilizing the blue band to correct the atmospheric effects. Nevertheless, NDVI and EVI tend to correlate well in semi-arid regions (Huete et al., 2002). Lastly, MTCI has a strong connection with the leaf chlorophyll content (Frampton et al., 2013) which is proportional to crop productivity.

The following formulae define the indices from Sentinel-2 bands. It is noteworthy that the B6 band was resampled to match with other bands.

$$\mathrm{NDVI_{raw}} = \frac{\mathrm{B_8} - \mathrm{B_4}}{\mathrm{B_8} + \mathrm{B_4}}, \tag{1}$$

$$\mathrm{EVI_{raw}} = \frac{2.5(\mathrm{B_8} - \mathrm{B_4})}{\mathrm{B_8} + 6\mathrm{B_4} - 7.5\mathrm{B_2} + 10000}, \tag{2}$$

$$\mathrm{MTCI_{raw}} = \frac{(\mathrm{B_6} - \mathrm{B_5})}{\mathrm{B_5} - \mathrm{B_4}} \tag{3}$$

NDVI and EVI values range from −1 to 1, while MTCI has a typical range of 1 to 6. The maximum value of each index was computed to be used as covariates. Additionally, the phenology extraction was performed using the NDVI through the following steps.

Firstly, the NDVI time series was passed to the Fourier analysis sequence (Jakubauskas et al., 2001; Wagenseil and Samimi, 2006) for denoising dan sampling the data in a homogeneous time interval. Essentially, a harmonic function was fitted iteratively to the upper envelope of the raw data using least squares regression.

$$NDVI_{smoothed} = f(t) = A + Bt + \sum_{n=1}^{n=n_H} C_n \sin\left(2\pi \frac{nt}{T}\right) + \sum_{n=1}^{n=n_H} D_n \cos\left(2\pi \frac{nt}{T}\right) \tag{4}$$

Here, t is the observation epoch expressed in year fraction, $T = 2$ years is the maximum period of variability, $n_H = 6$ is the number of harmonics. A, B, C, and D are the regression coefficients. In total, there are 14 regression coefficients obtained in this step. Afterward, fluctuations with higher frequencies (higher number of harmonic) were filtered leaving more prominent seasonal variations. Fourier analysis performs well when implemented to the noisy time-series data and it shows superiority over the double logistic model when dealing with data with multiple cropping seasons (Atkinson et al., 2012).

In order to account for the inter-annual variation of crop practices (e.g., active crop in the first year and no crop in the second year), we use 2 years as the maximum period of variability and 6 as the number of harmonics. This is equivalent to the same fitting procedure but using 1 year period and 3 harmonics which is sufficient even for detecting double cropping practices.

The ordinary Fourier analysis tends to underestimate the NDVI since it involves least-square fit of a harmonic function such that the upper envelope fit was implemented to overcome this issue (Bradley et al., 2007). To get the upper envelope, the NDVI values were altered in every iteration using the following rules.

$$NDVI_{raw,i+1} = \max\left(NDVI_{raw,i}, NDVI_{smoothed,i}\right) \tag{5}$$

After three iterations, the smoothed NDVI that fits the upper envelope was obtained. To achieve evenly spaced time series data for the phenology extraction, the smoothed NDVI was reconstructed/recalculated for every 5-day time interval. The procedures were implemented to every pixel in the image collection and after this process, an image collection (time series data) containing 93 images with appropriate timestamps from 2019 to 07-31 to 2020-10-31 was generated and ready for phenological parameters extraction. Fig. 2 shows examples of NDVI time series from one pixel that represents a crop and other pixels that is associated with natural vegetation.

The essence of phenological parameters extraction is to determine the start, peak, and end of the season of interest (Fig. 3). This can be performed using thresholding-based methods, derivative-based methods, and autoregressive moving average methods (Caparros-Santiago et al., 2021). In the current study, the first derivative of the smoothed NDVI was calculated and the inflection points were identified. Instead of using a root finding algorithm, the approximate location of the inflection point was identified as the point with the absolute value of the derivative of <0.01. The peak of the season was defined as the inflection point with positive curvature and significantly high local maximum which is >80 % of the global maximum and >0.20 NDVI. This criterion will be more important when dealing with multiple cropping seasons or when inspecting areas with relatively low peaks in NDVI. Two valleys or inflection points with negative curvatures located before and after the peak was considered as the identification of the start and end of the season. To avoid misidentification due to

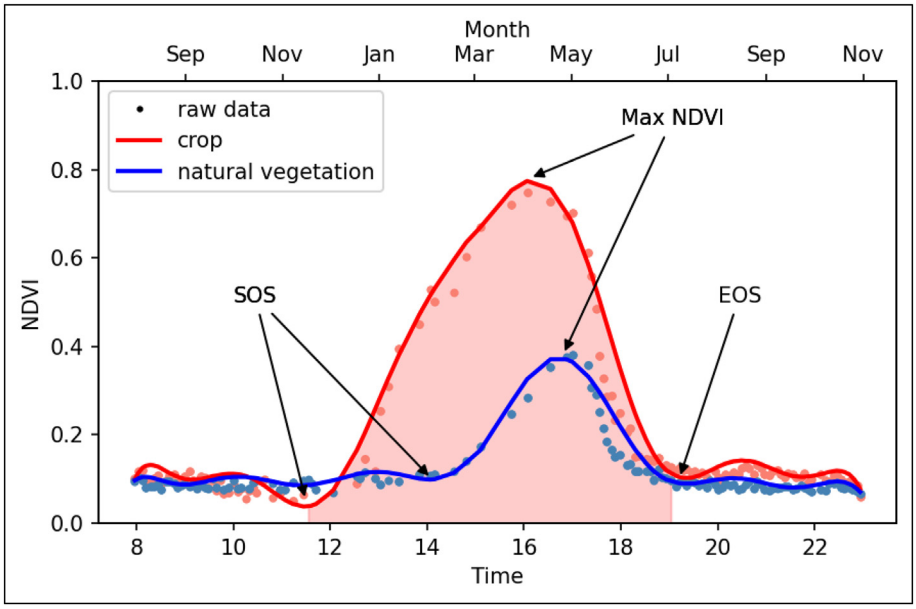


Fig. 3. Time series plot of the raw (circles) and smoothed (lines) NDVI from pixels associated with crop (black) and with natural vegetation (red).

local fluctuation, the start and end of the season should have NDVI values of <80 % of the associated peak. Additionally, valid peaks should be preceded by at least 6 consecutive points with positive derivative (consecutive ups) while the valleys should be preceded by 6 consecutive downs or more. When applied to the data with a 5-day time interval, this criterion can be translated into consecutive ups or downs for more than a month which is comparable to a previous study by Qader et al. (2015). After the identification of the start and end of the season, more parameters were extracted (Table 2, Fig. 4). Pixels with incomplete seasonal parameters were masked.

All these processes were performed in Google Earth Engine (GEE) (Gorelick et al., 2017) which is a web platform providing a huge collection of satellite imagery and vector/tabular data, cloud computing resources and algorithms available in JavaScript and Python. For the current study, a web-based interactive development environment provided by GEE was used to run the JavaScript codes.

2.3. Statistical analysis

2.3.1. Exploratory data analysis and covariate selection.

We first verified that the dependent variable, i.e., crop yield at the farm level, was approximately normally distributed as is required in a linear regression context. We, however, opted to model crop yield on the log scale to ensure that our modelled estimates were positive when these are back-transformed. We then evaluated the correlations between the covariates and their relationships with crop yield using scatter plot diagrams. We also fitted regression splines (natural splines) to obtain smooth curves to assess potential non-linear relationships between the covariates and crop yield. We considered log transformations of the covariates that appeared to have non-linear relationships with crop yield, but this did not produce any improvements in linearity. We also considered accounting for the non-linear relationships using piecewise linear functions, but this did not yield any improvements in predictive power (perhaps due to insufficient data to capture the non-linear relationships), hence we did not pursue this further.

In order to investigate potential (multi)collinearity among the covariates, we fitted simple linear regression models to rank the covariates based on their out-of-sample predictive ability. This was determined using the predictive R^2 statistic computed during a Monte Carlo cross-validation exercise (other cross-validation techniques are also possible) in which we used 80 % of the data for training and 20 % for validation. We chose between highly correlated pairs of covariates (Pearson correlation $\geq$ 0.8 using their ranks. We calculated the variance inflation factor (VIF) of the remaining covariates to further test for (multi)collinearity and potential overfitting using the multiple linear regression model as shown in Eq. (1). All covariates with VIF values <4.0 were retained in the analysis (Hair et al., 1995). Next, we assessed the presence of spatial correlation in the residuals resulting from the multiple linear regression model by calculating their empirical variogram using the 'gstat' package in R.

Also, we have tested the added value of the environmental covariates (Rainfall, Land Surface temperature, Elevation and Slope) in our analysis. To do so, we excluded the environmental covariates from the analysis and implemented the entirety of the analysis using only Sentinel-2 covariates – from covariate selection to model-fitting and validation. Then both modelling outputs were compered.

Table 2
Name of the Sentinel 2 and environmental parameters used as covariates in the study.

Parameter	Description
Start of the season (SOS)	Valid valley before the peak of the season
Peak of the season (TNDVI)	Time at which NDVI reaches a maximum in a year
End of the season (EOS)	Valid valley after the peak of season
Length of the season (LOS)	Time difference between EOS and SOS
Maximum NDVI (MaxNDVI)	Maximum value of NDVI between SOS and EOS
Maximum EVI (MaxEVI)	Maximum value of EVI between SOS and EOS.
Maximum MTCI* (MaxMTCI)	Maximum value of MTCI between SOS and EOS
Cumulative NDVI (CumNDVI)	Accumulation of NDVI from SOS to EOS. The value depends on the time interval between consecutive data points (e.g., 5 days).
MaxLST	Maximum Land Surface Temperature between October 2019 to July 2020
CumRain	Cumulative Rainfall between October 2019 to July 2020
MinLST	Minimum Land Surface Temperature between October 2019 to July 2020
Slope	Slope
Elevation	Elevation
MeanRain	Mean Rainfall between October 2019 to July 2020
MeanLST	Mean Land Surface Temperature between October 2019 to July 2020

2.3.2. Crop yield model

The final model used for prediction is given by:

$$\log\,(Yield_i) = \beta_0 + \Sigma_{j=1}^{p}\beta_j x_{ij} + \epsilon_i, i = 1, \ldots, n = 138, \quad (6)$$

where $x_{i1}, \ldots, x_{ip}$ are the covariates, $\beta_0, \ldots, \beta_p$ are regression coefficients and ϵ_i is an error term assumed to be normally distributed with mean 0 and variance σ^2. We fitted the model in a Bayesian framework and placed non-informative priors on the parameters: $\beta_j \sim N(0, 10^3)(j = 1, \ldots, p)$ and $\log(\sigma^{-2}) \sim N(0, 10)$. The model was implemented in R using the R-INLA package (Lindgren and Rue, 2015; R Core Team, 2022). Using the fitted model, we generated 1000 samples from the posterior predictive distributions of crop yield for each of the prediction locations, which were summarized to produce the 10 m × 10 m gridded estimates and associated uncertainties.

2.3.3. Model validation

We evaluated the predictive performance of the model using a *k*-fold cross-validation exercises, with the folds created as random subsets of the data and *k* set equal to 3. Using the true and predicted values, we computed the R-squared statistic and the root-mean-square error (RMSE) to evaluate predictive performance, both of which were averaged over the *k* folds. We also examined the plot of the standardised residuals against the predicted values to assess the adequacy of the model.

3. Results

3.1. Crop mask assessment and ground yield validation

With the advantage of GPS coordinates that were collected for the crop cut data, the generated crop map extracted from GlobeLand30 2020 was validated. Out of 138 samples, 119 crop fields were correctly classified as cropland which yielded an overall accuracy of around 86 %. To extend the crop map to the crop fields that were not classified as cropland, the crop map was manually modified.

Fig. 5a shows the locations of the selected crop fields with their ground wheat crop yield (gm/m^2) in Sulaimani governorate. Overall, the same pattern can be seen in both estimates in which yield is generally high in Pshdar, Sulaimani and Halabja compared to Garmian administration (Fig. 5a). Regarding the yield variation among the sub-plots within the same crop field, three crop cut replications were carried out in 17 crop fields (Fig. 5b). Fig. 5b illustrates the variation of the yield among the sub-plots within crop fields. It can be seen from the figure that yield variations among the sub-plots within a crop field are generally low. Table 3 shows the comparison between yield estimates collected for this study and KRSO data (KRSO, 2022). Regarding the quantitative comparison between the two estimates, close agreement was observed in Sulaimani and Halabja while the values are slightly different in the Garmian administration (Table 3).

3.2. Covariate selection

In Fig. 6a, we show the correlations between the covariates and the plots of the covariates against crop yield (on the log scale) (Fig. 6b). We

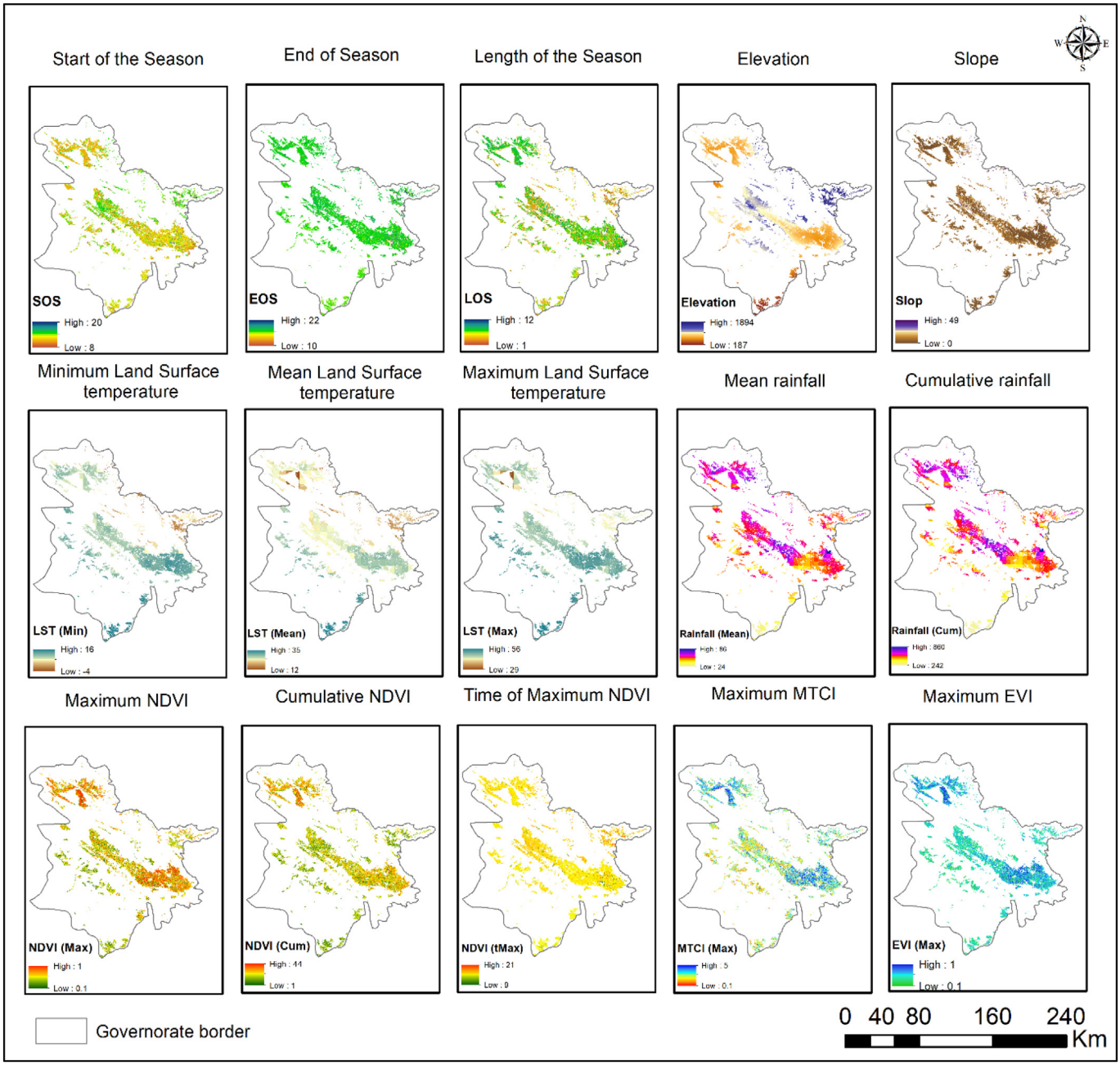


Fig. 4. Maps of phenological, climatic and topographic parameters used as covariates in the study.

also show the histograms of crop yield on the original and logarithmic scales (see supplementary information Fig. S1), both of which showed no substantial evidence of departure from normality. We note that even though some covariates (e.g., MEAN_SOS, MEAN_LOS, MEAN_MaxLST and MEAN_MeanLST) appeared to exhibit non-linear relationships with crop yield, accounting for these (using regression splines) (James et al., 2013) did not lead to any improvements in predictive power as mentioned previously. The ranks of the covariates based on their predictive R^2 produced during covariates selection is shown in supplementary Tables S1and S2.

In all, eight covariates were selected for the analysis. These include: MEAN_SOS, MEAN_EOS, MEAN_MaxMtci, MEAN_CumNDVI, MEAN_TNDVI, MEAN_MaxLST, MEAN_Slop and MEAN_Elevation.

We did not observe any evidence of residual spatial autocorrelation in crop yield after accounting for covariate effects (see supplementary information Fig. S2), justifying the use of the model in Eq. (1) to predict crop yield.

3.3. Crop yield model fitting, validation and prediction

The estimates of the parameters of the model are presented in Table 4. The covariates that had significant relationships with crop yield were MEAN_CumNDVI, MEAN_Elevation, MEAN_EOS, MEAN_MaxLST and MEAN_SOS.

These covariates had significant positive relationships with (log-transformed) crop yield except MEAN_EOS. On average, a unit increase in MEAN_CumNDVI will increase crop yield by 12 % (=100 × (exp (0.1156) − 1)) over a 10-m square area, holding other variables constant. Similar interpretations can also be made for other covariates.

Fig. 7 compares the in-sample and out-of-sample predictions (based on a *k*-fold cross-validation exercise) produced by the model with the observed values. Both plots show that the model produced reasonable predictions of the data in both cases. For in-sample prediction, we estimated an R^2

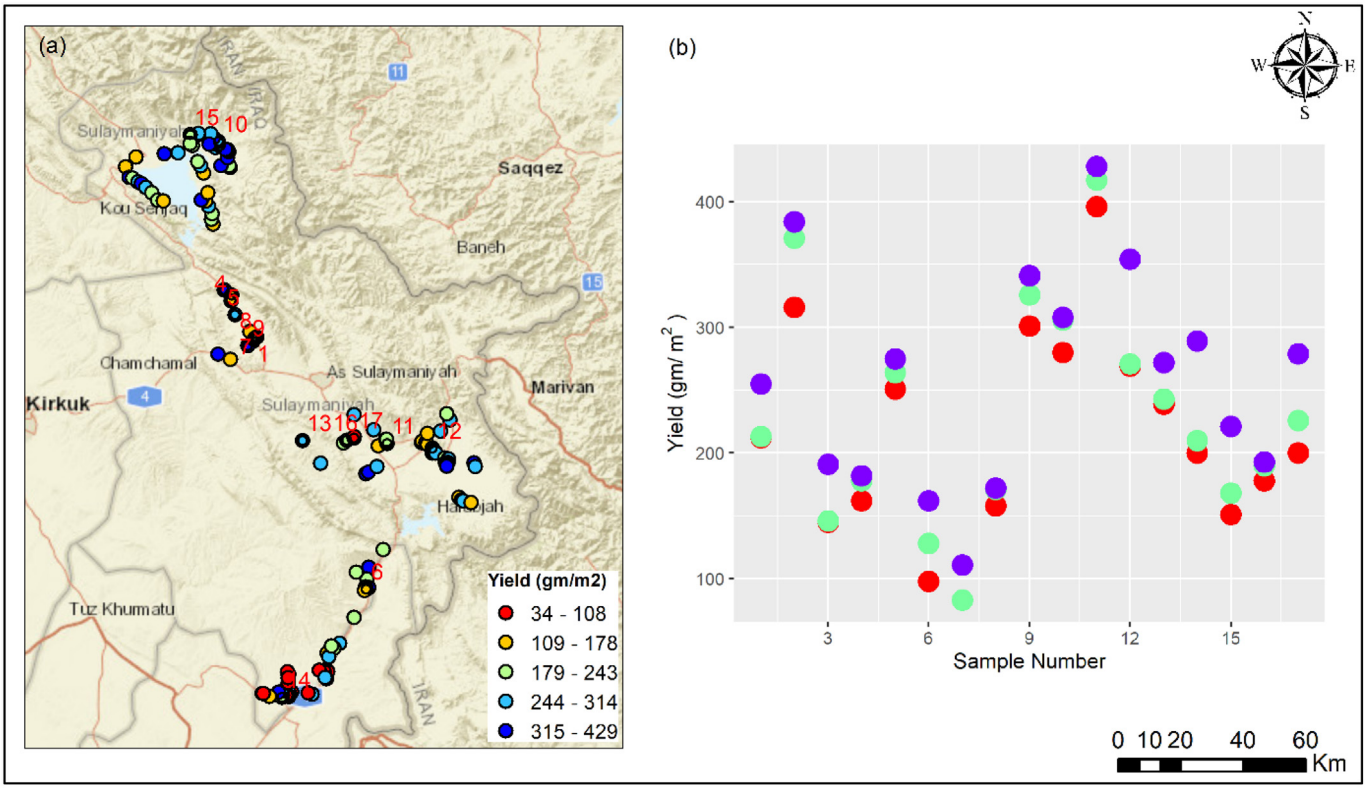


Fig. 5. (a) Locations of the selected crop fields with their ground wheat crop yield (gm/m^2) in Sulaimani governorate and (b) shows the yield variation for the 17 crop fields where the crop cuts were conducted by selecting three 1 × 1 m^2 sub-plots at random from each farmer's field. The distribution of the selected 17 crop fields is highlighted in Fig. 5a by their numbers in Fig. 5b x-axis.

value of 51 % whereas for out-of-sample prediction, this was 41 %, both of which indicate reasonable predictive performance.

Further, the calculated root-mean-square error for out-of-sample prediction was 69.80 g/m^2 (= 63.69 g/m^2 for in-sample prediction) which is less than the standard deviation of 89.23 g for crop yield for 1-m square area, further showing that the model performed well by reducing prediction variability.

To examine the role of the environmental covariates (Land Surface Temperature, Rainfall, Elevation and Slope) in the model, we re-ran the analysis (from covariate selection to model-fitting) excluding these covariates and using only the Sentinel-2 covariates. Consequently, the selected Sentinel-2 covariates were MEAN_CumNDVI, MEAN_MaxMtci, MEAN_EOS, MEAN_TNDVI and MEAN_SOS, which are also subset of the covariates included in the full analysis. The model fitted using these covariates had an in-sample R^2 value of 43 % and an in-sample RMSE value of 69.08 %. Following a cross-validation exercise, as with the full model, we obtained an out-of-sample R^2 value of 35 % and an out-of-sample RMSE of 69.08. These results clearly show that the inclusion of the environmental covariates in the full analysis greatly improved the predictive ability of the model (for the covariate selection, please see Tables S1and S2).

Next, we present the predicted crop yield and associated uncertainties at 10 m resolution in Fig. 8.

These maps reveal substantial heterogeneities in crop yield, which does not appear to exhibit any strong spatial structure. At the local level, it can be seen from the map that higher yield estimates were recorded for Pshdar and Halabja while lower yield estimates were observed in Garmian administrative. Theses spatial patterns are in line with KRSO data (KRSO, 2022). The uncertainties (or standard deviations) associated with these predictions are generally low, which further evidences the accuracy of the predictions.

Agriculture management is generally heterogeneous in the country and farmers are using different techniques and wheat crop species which results in substantial heterogeneities in final yield among the crop fields. In addition, variations in soil fertility and climatic variables might also cause heterogeneities in crop yield in the region. Larger errors in yield estimates were observed for small areas that are either misclassified or close to the Dukan and Darbandikhan lakes in which their reflectance might have been disturbed by water.

4. Discussion

The traditional approach of crop monitoring and estimation requires a complete enumeration of each district where crop-cutting experiments are performed to estimate crop yield (Kosmowski et al., 2021). However, relying only on the traditional approach to estimate crop yield at the farm level may not be sufficient to provide yield estimates for all the crop areas since unsampled areas are left unpredicted. In addition, crop cut data collection may not be possible in some regions or countries because of insecurity challenges. This work has collated various satellite-based covariates and climatic and topographic variables that have relation to crop yield and explored a statistical methodology to estimate yield at the farm level

Table 3
Crop yield estimates for 2019/2020 from KRSO agricultural statistics data and crop cut data collected in this study for the local areas in Sulaimani governorate.

Winter crop yield data for 2019/2020	Sulaimani	Range	Halabja	Range	Garmian administration	Range
Weight	(Kgm/Dounm)					
KRSO data	700	NA	600	NA	345	NA
Crop cut data	621	(100–429)	617	(263–1038)	442	(85–940)

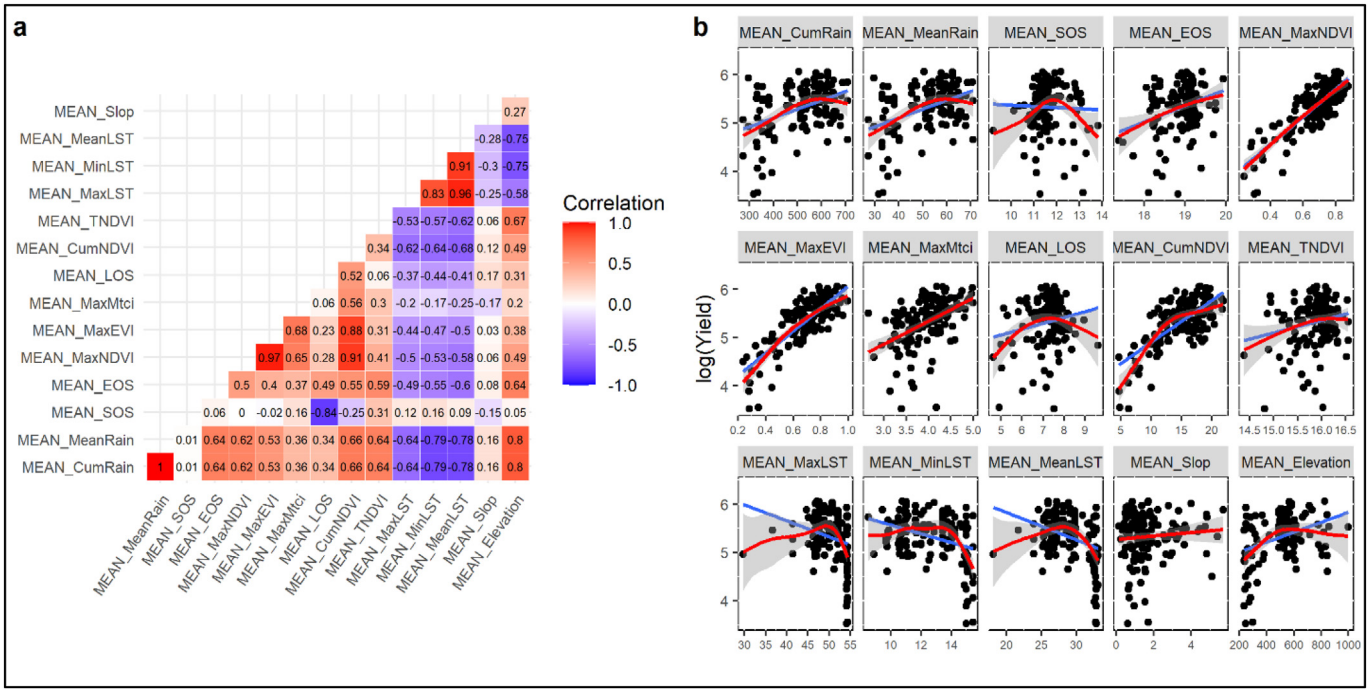


Fig. 6. (a) Plots of the correlations between the covariates. (b) Scatter plots of crop yield (on the log scale) against the covariates considered in the study. The blue lines are simple linear regression fits while the red lines are from natural splines, with the corresponding uncertainties shown in grey.

in a sparse data setting data environment. Such a model can be expanded to other areas with similar environments, particularly where access is restricted due to security issues.

Overall, there is a strong agreement and correlation between predicted and observed yield in the study area. Agriculture management is generally heterogeneous in the country and farmers are using different techniques and wheat crop species which results in substantial heterogeneities in final yield among the crop fields. In addition, variations in soil fertility and climatic variables might also cause heterogeneities in crop yield in the region. Larger errors in yield estimates were observed for small areas that are either misclassified or close to the Dukan and Darbandikhan lakes in which their reflectance might have been disturbed by water.

The reliability of our modelling outputs relies on the quality of ground yield and extracted variables. Our analysis reveals that the derived spectral bands are on the top of the list in the ranking of the covariates in terms of predictive power during covariate selection. This is in line with other findings where strong relationships between remote sensing spectral indices and crop yield were reported (Hunt et al., 2019; Skakun et al., 2017; Cavalaris et al., 2021). Although, the climatic datasets used in our work have very low spatial resolution compared to the Sentinel-2 covariates, they have produced a reasonable positive correlation with final yield. In particular, maximum temperature is one of the covariates significantly correlated with crop yield. This indicates that rainfall and temperature could be considered as one of the main crop yield drivers in the region. Topographically, the main fertile plains such as Halabja and Pshdar are located in higher altitudes of the crop area in the region while Garmian administration, which exhibits low yield is situated in low altitudes. Such distribution might have caused the significant positive relationship between yield and elevation. Interestingly, the phenological parameters such as SOS and EOS also produced a significant relationship with crop yield. This might be because the study area is dominantly rainfed and the start and end of the season are mainly driven by rainfall and temperature. Therefore, the date of the phenological parameters may change due to climatic condition and their variation could have different implications on the final crop yield. However, these interactions and correlations merit further investigation using study designs that can enable causal interpretations since our modelling did not consider possible interaction between the covariates.

Among the used spectral indices and climatic and topographic variables, this study found that maximum NDVI had the strongest relationship with wheat crop yield. These findings are in line with that of Qader et al.

Table 4
Estimates of parameters of the fitted model showing the posterior means, standard deviations and the lower (2.5 % quantile) and upper (97.5 % quantile) limits of the 95 % credible interval.

Parameter	Mean	Std. dev.	2.5 %	97.5 %
Intercept	6.8571	2.6384	1.6662	12.0410
MEAN_CumNDVI[a]	0.1156	0.0153	0.0855	0.1457
MEAN_MaxMtci	0.0627	0.0909	−0.1162	0.2414
MEAN_Elevation[a]	0.0006	0.0003	0.0000	0.0012
MEAN_EOS[a]	−0.2227	0.1093	−0.4377	−0.0079
MEAN_MaxLST[a]	0.0412	0.0130	0.0157	0.0667
MEAN_TNDVI	−0.0533	0.1384	−0.3255	0.2188
MEAN_Slop	0.0227	0.0297	−0.0358	0.0812
MEAN_SOS[a]	0.1255	0.0585	0.0105	0.2404
Variance $\left(\tilde{\sigma}^2\right)$	0.1475	0.0200	0.1131	0.1915

[a] Significant predictors of crop yield at 5 % level.

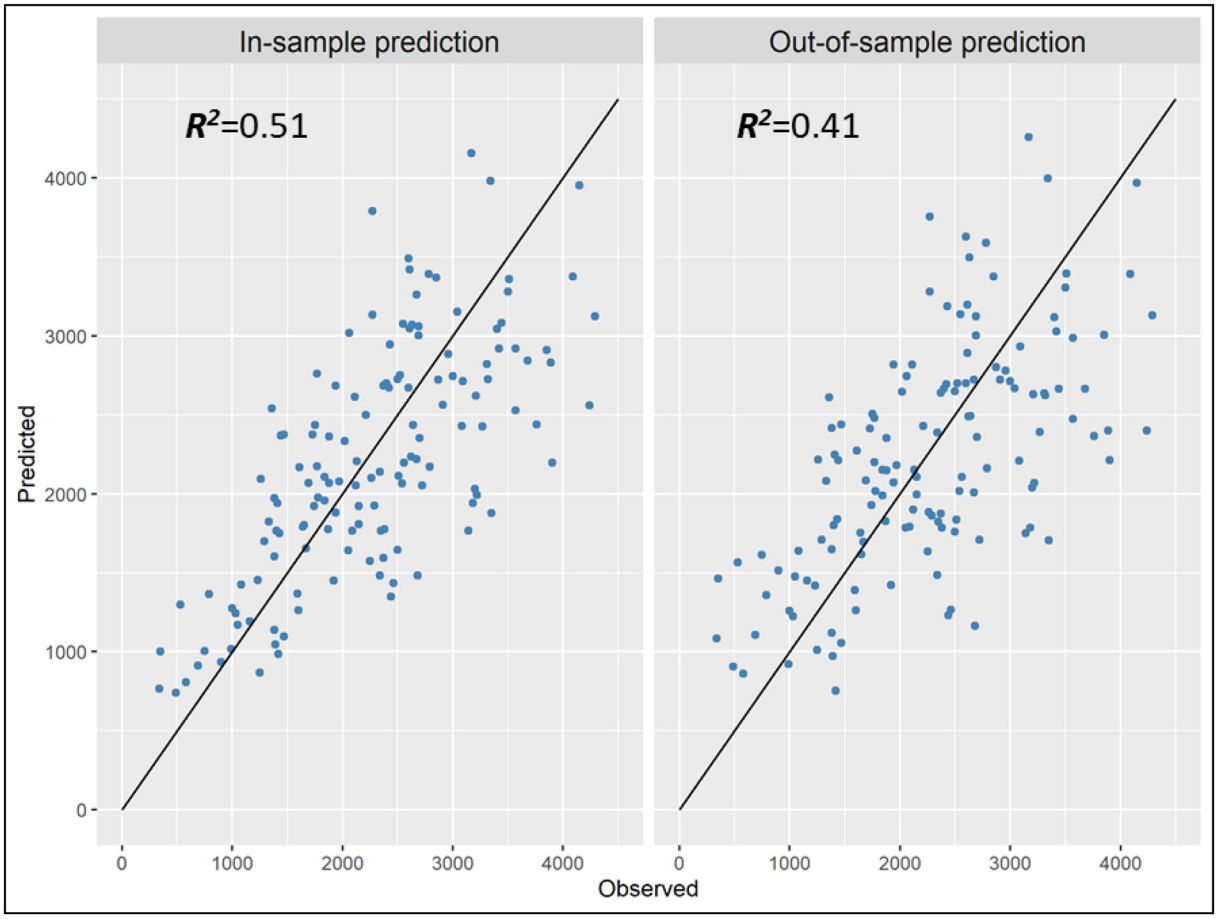


Fig. 7. Plots of observed versus predicted yield for 10-m square areas showing the in-sample and out-of-sample predictive performance of the fitted model. The out-of-sample predictions were generated using a *k*-fold cross-validation exercise.

(2018) who found a strong positive correlation between crop production and maximum NDVI at the governorate level in Iraq. Similarly, strong correlations between NDVI and EVI with crop yield have been reported in neighbouring countries including Syria and Iran (Jaafar and Ahmad, 2015; Mirasi et al., 2021). In addition, Gianquinto et al. (2011) reported that different physiological processes essential to obtaining yield can be indicated by the NDVI index. Although previous remote sensing models that are based on a spectral index sensitive to chlorophyll content reported more accurate crop yield prediction (Zhang and Liu, 2014; Jin et al., 2017; Ryu et al., 2020), in this study, the coefficient of determination between maximum NDVI and EVI covariates with crop yield was double that of the MTCI, indicating that NDVI and EVI are more sensitive to winter wheat crop yield than MTCI in the study area. This could be due to the involvement of lower spatial resolution band (20 m) in computing MTCI index compared to the bands that are used in NDVI and EVI (10 m). In addition, researchers have suggested different band combination in Sentinel-2 dataset to calculate MTCI (Karlson et al., 2020 suggested B8; Segarra et al., 2022 suggested B6; Li et al., 2022 suggested B7) and it is not clear which formula can produce the best results.

The lack of accuracy in the input datasets might have increased uncertainties in overall yield prediction and affected their estimated relationships with crop yield in this study. Although our model was shown to exhibit a reasonable predictive performance, this can be improved in future analyses through the inclusion of additional covariates. We will also compare the Bayesian approach used here with machine learning approaches which are particularly suitable for exploiting potential non-linear relationships to improve predictive performance.

This work has focused entirely on winter wheat which is a major winter crop type that is widely practised in the Sulaimani governorate (Eklund et al., 2017; KRSO, 2022). Within a similar crop calendar, some farmers are planting barley. However, the crop mask that was used for this work was not able to separate the crop types. This study adopted the GlobeLand30 for 2020 and defined the crop mask through its cultivated land cover type which may represent more than one crop type in the region. This might have added more uncertainties in the overall prediction since the relationship between predictive variables and crop yield is species specific (KRSO, 2022). In addition, although the cultivated land produced high accuracy based on the collected crop data locations, it failed to exclude cropland fallow, which is widely practised in the region (Eklund et al., 2017). This was concluded based on some visual comparison between the crop mask and the crop phenology parameters. Furthermore, the overall accuracy of GlobeLand30 2020 is 85 % and the Kappa coefficient is 0.78 (GlobeLand30, 2022), but this accuracy might not be achievable when the product is validated at the country level. For instance, the older version of the GlobeLand30 2020 were assessed in a similar environment and different accuracies were achieved (Sun et al., 2016; Jokar Arsanjani et al., 2016). Therefore, the lack of classification accuracy may result in mixing different land cover types which can lead to inaccurate yield prediction.

Another uncertainty may emerge from the lack of replications in crop cut data collection in each crop field. Besides overall low variation of the yields among the sub-plots in the selected 17 crop fields where more than one plots was taken, more replications should have been considered across the study area to assure the crop field representativeness in terms of intra-spatial yield variation. This is mainly because, agricultural management in this region is heterogeneous and environmental factors such as soil type, climate condition, soil moisture and soil fertility may vary within a crop field (Hilmi, 2018). However, insufficient

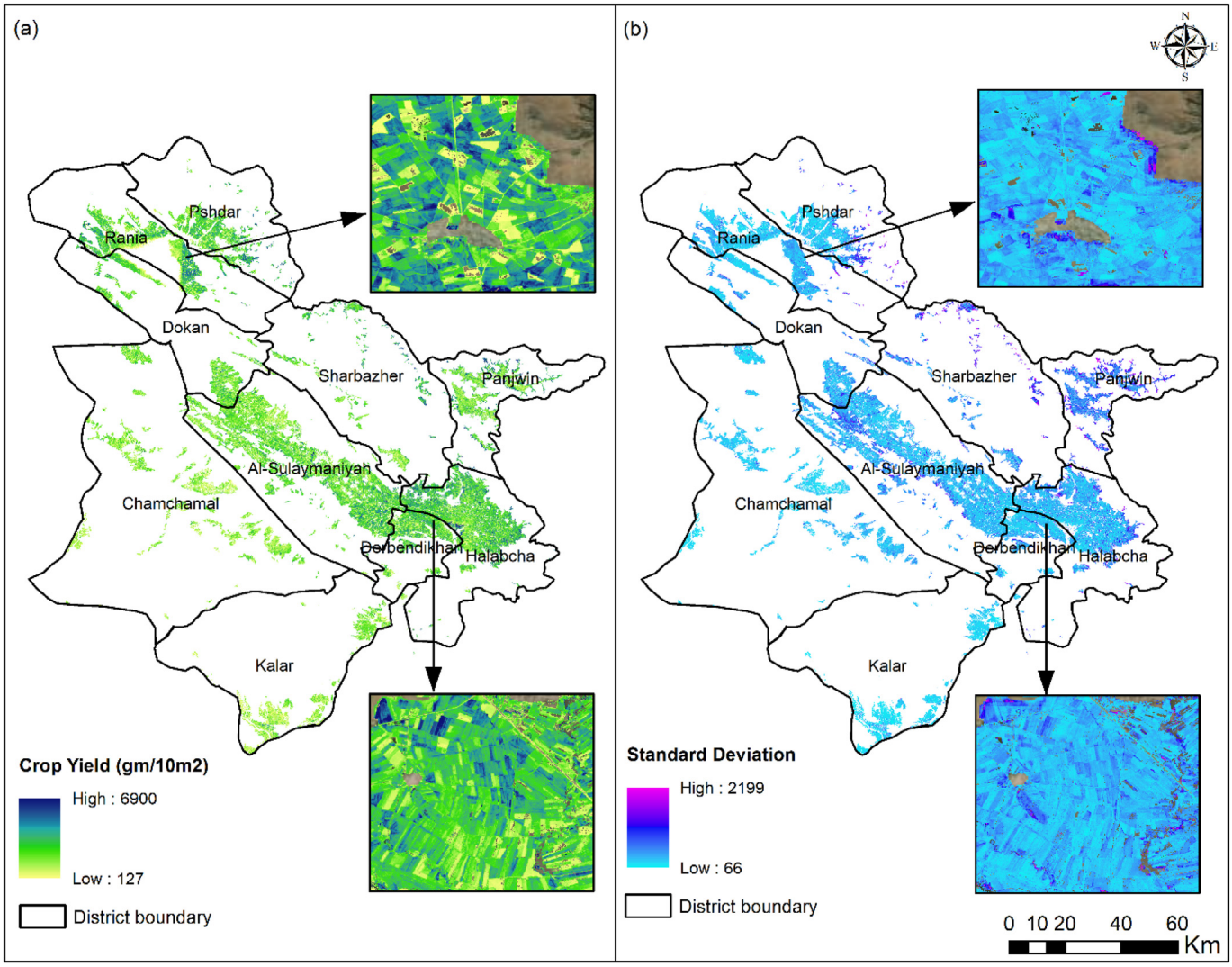


Fig. 8. Predicted crop yield (a) and associated uncertainties shown as standard deviations (b) at 10 m resolution for Sulaimani governorate in Iraq.

resources and limited time during the pandemic were the major obstacles to conducting more than one replication of crop cut data in each crop field. In addition, the crop prediction was made at 10 m resolution which are generally smaller than the agriculture field size and it can account for the variations within the field.

5. Conclusion

High resolution crop yield mapping is required in smallholder arid and semi-arid farming systems for a wide range of applications including monitoring food security and precision agriculture. However, high resolution crop yield maps in smallholder arid and semi-arid farming systems are scarce due to a lack of appropriate calibrated and validated crop yield models. This work is one of the first to examine how well Sentinel-2-derived information, topographic and climatic variables, can be used as covariates to accurately model and predict crop yield at the farm level in low data settings of arid and semi-arid regions, using Sulaimani governorate in Iraq as an example. The Bayesian multiple linear regression model was fitted to model and predict crop yield at 10 m resolution in the absence of substantial non-linear relationships between the covariates and crop yield, and residual spatial autocorrelation. In the ranking of covariates based on the predictive power, maximum NDVI produced the largest relationship with crop yield. Reasonable crop yield predictions were produced for in-sample and out-sample predictions (in-sample R^2 = 51 % and out-sample R^2 = 41 %). Our work shows that with freely accessible satellite-derived data, climatic and topographic covariates, crop yield can be predicted accurately using statistical models in complex small holding arid and semi-arid farming systems. This points to the potential for a broadly approach suitable for the data-poor setting environments, particularly areas with insecurity challenges.

CRediT authorship contribution statement

Conceptualization, S.H.Q.; Data curation, S.H.Q., C.E.U. and R.P.; Formal analysis, S.H.Q. and C.E.U.; Funding acquisition, S.H.Q.; Investigation, S.H.Q.; Methodology, S.H.Q. C.E.U. and R.P.; Field Data collection, P.N. and E.O.H.; Fieldwork design, S.H.Q.; Project administration, S.H.Q.; Resources, S.H.Q.; Supervision, S.H.Q.; Validation, S.H.Q. C.E.U. and R.P.; Visualization, S.H.Q. and R.P.; Writing—original draft preparation, S.H.Q., C.E.U. R.P.; Review and editing, N.R.K., A.J.T., P.N., E.O.H. and J.D. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by UK Research and Innovation GCRF 323036/ARCP011217.

Data availability

Data will be made available on request.

Declaration of competing interest

The authors declare no competing interests.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2023.161716.

References

Atkinson, P.M., Jeganathan, C., Dash, J., Atzberger, C., 2012. Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology. Remote Sens. Environ. 123, 400–417.

Baumeister, R.F., Vohs, K.D., Funder, D.C., 2007. Psychology as the science of self-reports and finger movements: whatever happened to actual behavior? Perspect. Psychol. Sci. 2 (4), 396–403. https://doi.org/10.1111/j.1745-6916.2007.00051.x.

Bégué, A., Leroux, L., Soumaré, M., Faure, J.-F., Diouf, A.A., Augusseau, X., Tonneau, J.-P., 2020. Remote sensing products and services in support of agricultural public policies in Africa: overview and challenges. Front.Sustain.Food Syst. 4. https://doi.org/10.3389/fsufs.2020.00058.

Bolton, D.K., Friedl, M.A., 2013. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. Agric. For. Meteorol. 173, 74–84. https://doi.org/10.1016/j.agrformet.2013.01.007.

Bradley, B.A., Jacob, R.W., Hermance, J.F., Mustard, J.F., 2007. A curve fitting procedure to derive inter-annual phenologies from time series of noisy satellite NDVI data. Remote Sens. Environ. 106 (2), 137–145. https://doi.org/10.1016/j.rse.2006.08.002.

Brisson, N., Mary, B., Ripoche, D., Jeuffroy, M.H., Ruget, F., Nicoullaud, B., Delecolle, R., 1998. STICS: a generic model for the simulation of crops and their water and nitrogen balances. I. Theory and parameterization applied to wheat and corn. Agronomie 18 (5–6), 311–346. https://doi.org/10.1051/agro:19980501.

Burke, M., Lobell, D.B., 2017. Satellite-based assessment of yield variation and its determinants in smallholder African systems. Proc. Natl. Acad. Sci. U. S. A. 114 (9), 2189–2194. https://doi.org/10.1073/pnas.1616919114.

Campos-Taberner, M., García-Haro, F.J., Martínez, B., Sánchez-Ruiz, S., Gilabert, M.A., 2019. A copernicus sentinel-1 and sentinel-2 classification framework for the 2020+ European common agricultural policy: a case study in València (Spain). Agronomy 9, 556.

Caparros-Santiago, J.A., Rodriguez-Galiano, V., Dash, J., 2021. Land surface phenology as indicator of global terrestrial ecosystem dynamics: a systematic review. ISPRS J. Photogramm. Remote Sens. 171, 330–347.

Carfagna, E., Gallego, F.J., 2005. Using remote sensing for agricultural statistics. Int. Stat. Rev. 73 (3), 389–404.. http://www.jstor.org/stable/25472682.

Carletto, C., Jolliffe, D., Banerjee, R., 2015. From tragedy to renaissance: improving agricultural data for better policies. J. Dev. Stud. 51 (2), 133–148. https://doi.org/10.1080/00220388.2014.968140.

Cavalaris, C., Megoudi, S., Maxouri, M., Anatolitis, K., Sifakis, M., Levizou, E., Kyparissis, A., 2021. Modeling of durum wheat yield based on Sentinel-2 imagery. Agronomy 11 (8), 1486.

Chahbi, A., Zribi, M., Shil, E., Lili-Chabaane, Z., 2022. Characterization of cereals in a semi-arid context based on remote sensing indicators from high spatial resolution images from the Sentinel 1 and Sentinel 2 satellite in central Tunisia. EGU General Assembly 2022, Vienna, Austria, 23–27 May 2022, EGU22-9150 https://doi.org/10.5194/egusphere-egu22-9150.

United Nations Population Division, Department of Economic and Social Affairs, 2019. World Population Prospects 2019: Highlights (ST/ESA/SER.A/423).

Cunha, R.L.F., Silva, B., 2020. ESTIMATING CROP YIELDS WITH REMOTE SENSING AND DEEP LEARNING. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., IV-3/W2-2020, pp. 59–64 https://doi.org/10.5194/isprs-annals-IV-3-W2-2020-59-2020.

Dash, J., Curran, P.J., 2004. The MERIS terrestrial chlorophyll index. Int. J. Remote Sens. 25, 5403–5413.

Dash, J., Curran, P.J., 2007. Relationship between the MERIS vegetation indices and crop yield for the stateof South Dakota, USA. Proceedings Envisat Symposium, Montreux, April 23–27 (ESA SP-636, July 2007).

Dhillon, S.S., 2004. Agricultural Geography. Tata McGraw-Hill.

Doraiswamy, P.C., Cook, P.W., 1995. Spring wheat yield assessment using NOAA AVHRR data. Can. J. Remote. Sens. 21 (1), 43–51. https://doi.org/10.1080/07038992.1995.10874595.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygert, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's optical high-resolution Mission for GMES operational services. Remote Sens. Environ. 120, 25–36.

Eklund, L., Abdi, A., Islar, M., 2017a. From producers to consumers: the challenges and opportunities of agricultural development in Iraqi Kurdistan. Land 6 (2), 44.

Engen, M., Sandø, E., Sjølander, B.L.O., Arenberg, S., Gupta, R., Goodwin, M., 2021. Farm-Scale Crop Yield Prediction from Multi-Temporal Data Using Deep Hybrid Neural Networks. Agronomy 11 (12), 2576.

Faqe Ibrahim, G.R., Rasul, A., Abdullah, H., 2022. Sentinel-2 Accurately Estimated Wheat Yield in a Semi-arid Region Compared With Landsat-8. Preprints https://doi.org/10.20944/preprints202209.0416.v1 2022090416.

Fermont, A., Benson, T., 2011. Estimating Yield of Food Crops Grown by Smallholder Farmers: A Review in the Uganda Context; Uganda Strategy Support Program Working Paper No. USSP 05. IFPRI, Washington, DC, USA, pp. 1–57.

Food and Agriculture Organization of the United Nations (FAO), 2013. Climate Smart Agriculture Sourcebook Rome. (Accessed: 04/12/2022).

Frampton, W.J., Dash, J., Watmough, G., Milton, E.J., 2013. Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. ISPRS J. Photogramm. Remote Sens. 82, 83–92. https://doi.org/10.1016/j.isprsjprs.2013.04.007.

Gianquinto, G., Orsini, F., Fecondini, M., Mezzetti, M., Sambo, P., Bona, S., 2011. A methodological approach for defining spectral indices for assessing tomato nitrogen status and yield. Eur. J. Agron. 35 (3), 135–143. https://doi.org/10.1016/j.eja.2011.05.005.

Global Strategy to improve Agricultural and Rural Statistics (GSARS), 2017. Handbook on Remote Sensing for Agricultural Statistics. GSARS Handbook: Rome.

GlobeLand30, 2022. Global Land Cover Mapping at 30m Resolution (2020). . http://www.globallandcover.com/.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google earth engine: planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27.

Gourlay, S., Kilic, T., Lobell, D.B., 2019. A new spin on an old debate: errors in farmer-reported production and their implications for inverse scale - productivity relationship in Uganda. J. Dev. Econ. 141, 102376. https://doi.org/10.1016/j.jdeveco.2019.102376.

Hair Jr., J.F., Anderson, R.E., Tatham, R.L., Black, W.C., 1995. Multivariate Data Analysis. 3rd ed. Macmillan, New York.

Hilmi, M., 2018. Entrepreneurship in farming: what is the current status of knowledge in the Kurdistan Region of Iraq? <journal-title><sb:contribution><sb:title>Middle East</sb:title></sb:contribution><sb:host><sb:issue><sb:series><sb:title>J. Agric.</sb:title></sb:series></sb:issue></sb:host></journal-title><sb:host><sb:issue><sb:series><sb:title> </sb:title></sb:series></sb:issue></sb:host> 07 (03), 858–875.a.

Huang, J.F., Wang, X.Z., Li, X.X., Tian, H.Q., Pan, Z.K., 2013. Remotely sensed rice yield prediction using multi-temporal NDVI data derived from NOAA's-AVHRR. Plos One 8 (8). https://doi.org/10.1371/journal.pone.0070816.

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ. 83, 195–213.

Humanitarian Data Exchange (HDX), 2022. Iraq - Subnational Administrative Boundaries. Accessed 24 06 2022 https://data.humdata.org/dataset/cod-ab-irq.

Hunt, M.L., Blackburn, G.A., Carrasco, L., Redhead, J.W., Rowland, C.S., 2019. High resolution wheat yield mapping using Sentinel-2. Remote Sens. Environ. 233, 111410. https://doi.org/10.1016/j.rse.2019.111410.

IPCC, 2022. Summary for policymakers. In: Pörtner, H.-O., Roberts, D.C., Poloczanska, E.S., Mintenbeck, K., Tignor, M., Alegría, A., Craig, M., Langsdorf, S., Löschke, S., Möller, V., Okem, A. (Eds.), Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [H.-O. Pörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama (eds.)]. Cambridge University Press, Cambridge, UK and New York, NY, USA, pp. 3–33 https://doi.org/10.1017/9781009325844.001 (Accessed: 15/01/2021).

Jaafar, H.H., Ahmad, F.A., 2015. Crop yield prediction from remotely sensed vegetation indices and primary productivity in arid and semi-arid lands. Int. J. Remote Sens. 36 (18), 4570–4589. https://doi.org/10.1080/01431161.2015.1084434.

Jain, M., Mondal, P., DeFries, R.S., Small, C., Galford, G.L., 2013. Mapping cropping intensity of smallholder farms: a comparison of methods using multiple sensors. Remote Sens. Environ. 134, 210–223. https://doi.org/10.1016/j.rse.2013.02.029.

Jakubauskas, M.E., Legates, D.R., Kastens, J.H., 2001. Harmonic analysis of time-series AVHRR NDVI data. Photogramm. Eng. Remote Sens. 67, 461–470.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. Unsupervised learning. In: James, G., Witten, D., Hastie, T., Tibshirani, R. (Eds.), An Introduction to Statistical Learning: With Applications in R. Springer, New York, New York, NY, pp. 373–418.

Jeffries, G.R., Griffin, T.S., Fleisher, D.H., Naumova, E.N., Koch, M., Wardlow, B.D., 2020. Mapping sub-field maize yields in Nebraska, USA by combining remote sensing imagery, crop simulation models, and machine learning. Precis. Agric. 21 (3), 678–694. https://doi.org/10.1007/s11119-019-09689-z.

Jin, Z., Azzari, G., Burke, M., Aston, S., Lobell, D.B., 2017. Mapping smallholder yield heterogeneity at multiple scales in Eastern Africa. Remote Sens. 9 (9), 931.

Jokar Arsanjani, J., Tayyebi, A., Vaz, E., 2016. GlobeLand30 as an alternative fine-scale global land cover map: challenges, possibilities, and implications for developing countries. Habitat Int. 55, 25–31. https://doi.org/10.1016/j.habitatint.2016.02.003.

Kang, Y., Özdoğan, M., 2019. Field-level crop yield mapping with landsat using a hierarchical data assimilation approach. Remote Sens. Environ. 228, 144–163. https://doi.org/10.1016/j.rse.2019.04.005.

Karlson, M., Ostwald, M., Bayala, J., Bazie, H.R., Ouedraogo, A.S., Soro, B., Reese, H., 2020. The potential of Sentinel-2 for crop production estimation in a smallholder agroforestry landscape,Burkina Faso. Front. Environ. Sci. 8. https://doi.org/10.3389/fenvs.2020.00085.

Karst, I.G., Mank, I., Traoré, I., Sorgho, R., Stückemann, K.-J., Simboro, S., Sié, A., Franke, J., Sauerborn, R., 2020. Estimating yields of household fields in rural subsistence farming systems to study food security in Burkina Faso. Remote Sens. 12 (11), 1717. https://doi.org/10.3390/rs12111717 License: Creative Commons Attribution CC BY 3.0 IGO.

Khaki, S., Pham, H., Wang, L., 2021. Simultaneous corn and soybean yield prediction from remote sensing data using deep transfer learning. Sci. Rep. 11 (1), 11132. https://doi.org/10.1038/s41598-021-89779-z.

Kosmowski, F., Chamberlin, J., Ayalew, H., Sida, T., Abay, K., Craufurd, P., 2021. How accurate are yield estimates from crop cuts? Evidence from smallholder maize farms in Ethiopia. Food Policy 102, 102122. https://doi.org/10.1016/j.foodpol.2021.102122.

Kurdistan Region Statistics Office (KRSO), 2022. Winter crops planted area and yield survey in Kurdistan Region, 1969 – 2020. https://krso.gov.krd/ku/statistics [Accessed: 27/04/2022].

Lambert, M.J., Traore, P.C.S., Blaes, X., Baret, P., Defourny, P., 2018. Estimating smallholder crops production at village level from Sentinel-2 time series in Mali's cotton belt. Remote Sens. Environ. 216, 647–657. https://doi.org/10.1016/j.rse.2018.06.036.

Leroux, L., Castets, M., Baron, C., Escorihuela, M.-J., Bégué, A., Lo Seen, D., 2019. Maize yield estimation in West Africa from crop process-induced combinations of multi-domain remote sensing indices. Eur. J. Agron. 108, 11–26. https://doi.org/10.1016/j.eja.2019.04.007.

Leroux, L., Falconnier, G.N., Diouf, A.A., Ndao, B., Gbodjo, J.E., Tall, L., Roupsard, O., 2020. Using remote sensing to assess the effect of trees on millet yield in complex parklands of Central Senegal. Agric. Syst. 184, 102918. https://doi.org/10.1016/j.agsy.2020.102918.

Li, C., Chimimba, E.G., Kambombe, O., Brown, L.A., Chibarabada, T.P., Lu, Y., Dash, J., 2022. Maize yield estimation in intercropped smallholder fields using satellite data in southern Malawi. Remote Sens. 14 (10), 2458.

Lindgren, F., Rue, H., 2015. Bayesian Spatial Modelling with R-INLA. J. Stat. Softw. 63 (19), 1–25. https://doi.org/10.18637/jss.v063.i19.

Lloyd, S.P., 1957. Least Squares Quantization in PCM. Technical Report RR-5497, Bell Lab, September 1957.

Lobell, D.B., Thau, D., Seifert, C., Engle, E., Little, B., 2015. A scalable satellite-based crop yield mapper. Remote Sens. Environ. 164, 324–333. https://doi.org/10.1016/j.rse.2015.04.021.

Lobell, David B., Azzari, George, Marshall, Burke, Gourlay, Sydney, Jin, Zhenong, Kilic, Talip, Murray, Siobhan, 2018. Eyes in the Sky, Boots on the Ground: Assessing Satellite- and Ground-Based Approaches to Crop Yield Measurement and Analysis in Uganda. Policy Research Working Paper;No. 8374. © World BankWorld Bank, Washington, DC License: CC BY 3.0 IGO https://openknowledge.worldbank.org/handle/10986/29554.

Lobell, D.B., Di Tommaso, S., You, C., Yacoubou Djima, I., Burke, M., Kilic, T., 2020. Sight for sorghums: comparisons of satellite- and ground-based sorghum yield estimates in Mali. Remote Sens. 12 (1), 100.

Lopresti, M.F., Di Bella, C.M., Degioanni, A.J., 2015. Relationship between MODIS-NDVI data and wheat yield: a case study in northern Buenos Aires province,Argentina. Inf. Process. Agric. 2 (2), 73–84. https://doi.org/10.1016/j.inpa.2015.06.001.

Lowder, S.K., Skoet, J., Raney, T., 2016. The number, size, and distribution of farms, smallholder farms, and family farms worldwide. World Dev. 87, 16–29. https://doi.org/10.1016/j.worlddev.2015.10.041.

Ma, Y., Zhang, Z., Kang, Y., Özdoğan, M., 2021. Corn yield prediction and uncertainty analysis based on remotely sensed variables using a Bayesian neural network approach. Remote Sens. Environ. 259, 112408. https://doi.org/10.1016/j.rse.2021.112408.

Mehdaoui, R., Anane, M., 2020. Exploitation of the red-edge bands of Sentinel 2 to improve the estimation of durum wheat yield in Grombalia region (Northeastern Tunisia). Int. J. Remote Sens. 41 (23), 8984–9006. https://doi.org/10.1080/01431161.2020.1797217.

Mirasi, A., Mahmoudi, A., Navid, H., Valizadeh Kamran, K., Asoodar, M.A., 2021. Evaluation of sum-NDVI values to estimate wheat grain yields using multi-temporal Landsat OLI data. Geocarto Int. 36 (12), 1309–1324. https://doi.org/10.1080/10106049.2019.1641561.

Mkhabela, M.S., Bullock, P., Raj, S., Wang, S., Yang, Y., 2011. Crop yield forecasting on the Canadian prairies using MODIS NDVI data. Agric. For. Meteorol. 151 (3), 385–393. https://doi.org/10.1016/j.agrformet.2010.11.012.

Moriondo, M., Maselli, F., Bindi, M., 2007. A simple model of regional wheat yield based on NDVI data. Eur. J. Agron. 26 (3), 266–274. https://doi.org/10.1016/j.eja.2006.10.007.

Murthy, C.S., Thiruvengadachari, S., Jonna, S., Raju, P.V., 1997. Design of crop cutting experiments with satellite data for crop yield estimation in irrigated command areas. Geocarto Int. 12 (2), 5–11. https://doi.org/10.1080/10106049709354580.

Murugananthan, P., Wibowo, S., Grandhi, S., Samrat, N.H., Islam, N., 2022. A systematic literature review on crop yield prediction with deep learning and remote sensing. Remote Sens. 14 (9), 1990.

NASA Shuttle Radar Topography Mission (SRTM), 2013. Shuttle Radar Topography Mission (SRTM) Global. Distributed by OpenTopography. https://doi.org/10.5069/G9445JDF Accessed: 2022-04-26.

Nguyen, P., Shearer, E.J., Tran, H., Ombadi, M., Hayatbini, N., Palacios, T., Huynh, P., Updegraff, G., Hsu, K., Kuligowski, B., Logan, W.S., Sorooshian, S., 2019. The CHRS data portal, an easily accessible public repository for PERSIANN global satellite precipitation data. Nat.Sci.Data 6, 180296. https://doi.org/10.1038/sdata.2018.296.

Oikonomidis, A., Catal, C., Kassahun, A., 2022. Deep learning for crop yield prediction: a systematic literature review. N. Z. J. Crop. Hortic. Sci., 1–26 https://doi.org/10.1080/01140671.2022.2032213.

Ouattara, B., Forkuor, G., Zoungrana, B.J.B., Dimobe, K., Danumah, J., Saley, B., Tondoh, J.E., 2020. Crops monitoring and yield estimation using sentinel products in semi-arid smallholder irrigation schemes. Int. J. Remote Sens. 41 (17), 6527–6549. https://doi.org/10.1080/01431161.2020.1739355.

Paliwal, A., Jain, M., 2020. The accuracy of self-reported crop yield estimates and their ability to train remote sensing algorithms. Front.Sustain.Food Syst. 4. https://doi.org/10.3389/fsufs.2020.00025.

Pham, H.T., Awange, J., Kuhn, M., Nguyen, B.V., Bui, L.K., 2022. Enhancing crop yield prediction utilizing machine learning on satellite-based vegetation health indices. Sensors 22 (3), 719. https://doi.org/10.3390/s22030719.

Qader, S.H., Atkinson, P.M., Dash, J., 2015. Spatiotemporal variation in the terrestrial vegetation phenology of Iraq and its relation with elevation. Int. J. Appl. Earth Obs. Geoinf. 41, 107–117. https://doi.org/10.1016/j.jag.2015.04.021.

Qader, S.H., Dash, J., Atkinson, P.M., 2018. Forecasting wheat and barley crop production in arid and semi-arid regions using remotely sensed primary productivity and crop phenology: a case study in Iraq. Sci. Total Environ. 613, 250–262. https://doi.org/10.1016/j.scitotenv.2017.09.057.

Qader, S.H., Dash, J., Alegana, V.A., Khwarahm, N.R., Tatem, A.J., Atkinson, P.M., 2021. The role of earth observation in achieving sustainable agricultural production in arid and semi-arid regions of the world. Remote Sens. 13 (17), 3382.

R Core Team, 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Roser, M., 2013. Employment in Agriculture. Published online at OurWorldInData.org. Retrieved from: [Online Resource] (Accessed: 15/01/2021) https://ourworldindata.org/employment-in-agriculture.

Rouse, J.W., Haas, R.H., Schell, J.A., Deering, W.D., 1973. Monitoring vegetation systems in the Great Plains with ERTS. Third ERTS Symposium, NASA SP–351, pp. 309–317.

Ryu, J.-H., Jeong, H., Cho, J., 2020. Performances of vegetation indices on paddyRice at elevated air temperature, heat stress, and herbicide damage. Remote Sens. 12 (16), 2654.

Sapkota, T.B., Jat, M.L., Jat, R.K., Kapoor, P., Stirling, C., 2016. Yield estimation of food and non-food crops in smallholder production systems. In: Rosenstock, T., Rufino, M., Butterbach-Bahl, K., Wollenberg, L., Richards, M. (Eds.), Methods for Measuring Greenhouse Gas Balances and Evaluating Mitigation Options in Smallholder Agriculture. Springer, Cham https://doi.org/10.1007/978-3-319-29794-1_8.

Segarra, J., Buchaillot, M.L., Araus, J.L., Kefauver, S.C., 2020. Remote sensing for precision agriculture: Sentinel-2 improved features and applications. Agronomy 10 (5), 641.

Segarra, J., Araus, J.L., Kefauver, S.C., 2022. Farming and Earth observation: Sentinel-2 data to estimate within-field wheat grain yield. Int. J. Appl. Earth Obs. Geoinf. 107, 102697. https://doi.org/10.1016/j.jag.2022.102697.

Shirley, R., Pope, E., Bartlett, M., Oliver, S., Quadrianto, N., Hurley, P., Bacon, J., 2020. An empirical, Bayesian approach to modelling crop yield: maize in USA. Environ.Res. Commun. 2 (2), 025002. https://doi.org/10.1088/2515-7620/ab67f0.

Sida, T.S., Chamberlin, J., Ayalew, H., Kosmowski, F., Craufurd, P., 2021. Implications of intra-plot heterogeneity for yield estimation accuracy: evidence from smallholder maize systems in Ethiopia. Field Crops Res. 267, 108147.

Skakun, S., Vermote, E., Roger, J.C., Franch, B., 2017. Combined use of Landsat-8 and Sentinel-2A images for winter crop mapping and winter wheat yield assessment at regional scale. AIMS Geosci. 3, 163.

Sun, B., Chen, X., Zhou, Q., 2016. Uncertainty assessment of GlobeLand30 land cover data set over central Asia. ISPRS Int Arch Photogramm Remote Sens Spat Inf Sci XLI-B8, pp. 1313–1317 https://doi.org/10.5194/isprsarchives-XLI-B8-1313-2016.

Uphaus, C., 2008. Ending hunger: the role of agriculture. Bread for the World Institute, Washington, DC (Accessed: 15/01/2021) https://www.issuelab.org/resources/1124/1124.pdf.

Van Evert, F.K., Campbell, G.S., 1994. Cropsyst - a collection of object-oriented simulation-models of agricultural systems. Agron. J. 86 (2), 325–331. https://doi.org/10.2134/agronj1994.00021962008600020022x.

Vandiepen, C.A., Wolf, J., Vankeulen, H., Rappoldt, C., 1989. Wofost - a simulation-model of crop production. Soil Use Manag. 5 (1), 16–24.

Vermote, Eric, Justice, Chris, Csiszar, Ivan, Eidenshink, Jeff, Myneni, Ranga B., Baret, Frederic, Masuoka, Ed, Wolfe, Robert E., Claverie, Martin, NOAA CDR Program, 2014. NOAA Climate Data Record (CDR) of Normalized Difference Vegetation Index (NDVI), Version 4. [indicate subset used]. NOAA National Centers for Environmental Information https://doi.org/10.7289/V5PZ56R6 Accessed [date].

Wagenseil, H., Samimi, C., 2006. Assessing spatio-temporal variations in plant phenology using Fourier analysis on NDVI time series: results from a dry savannah environment in Namibia. Int. J. Remote Sens. 27, 3455–3471.

Wahab, I., 2020. In-season plot area loss and implications for yield estimation in smallholder rainfed farming systems at the village level in sub-Saharan Africa. GeoJournal 85 (6), 1553–1572. https://doi.org/10.1007/s10708-019-10039-9.

Wan, Z., Hook, S., Hulley, G., 2015. MOD11B3 MODIS/Terra Land Surface Temperature/Emissivity Monthly L3 Global 6km SIN Grid V006 [Data set]. NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/MODIS/MOD11B3.006.

Whitcraft, A.K., Vermote, E.F., Becker-Reshef, I., Justice, C.O., 2015. Cloud cover throughout the agricultural growing season: impacts on passive optical earth observations. Remote Sens. Environ. 156, 438–447. https://doi.org/10.1016/j.rse.2014.10.009.

World Bank, 2020. Agriculture and Food. (Accessed: 15/01/2021) https://www.worldbank.org/en/topic/agriculture/overview#1.

World Bank, 2022a. Poverty and Shared Prosperity. Correcting Course. World Bank, Washington, DC https://doi.org/10.1596/978-1-4648-1893-6 License: Creative Commons Attribution CC BY 3.0 IGO.

World Bank, 2022b. Agriculture, forestry, and fishing, value added (% of GDP). https://data.worldbank.org/indicator/NV.AGR.TOTL.ZS.

Xu, W., Chen, P., Zhan, Y., Chen, S., Zhang, L., Lan, Y., 2021. Cotton yield estimation model based on machine learning using time series UAV remote sensing data. Int. J. Appl. Earth Obs. Geoinf. 104, 102511. https://doi.org/10.1016/j.jag.2021.102511.

Zakaria, S., Mustafa, Y., Mohammed, D., Ali, S., Al-Ansari, N., Knutsson, S., 2013. Estimation of annual harvested runoff at Sulaymaniyah Governorate, Kurdistan region of Iraq. Nat. Sci. 5, 1272–1283. https://doi.org/10.4236/ns.2013.512155.

Zhang, S., Liu, L., 2014. The potential of the MERIS terrestrial chlorophyll index for crop yield prediction. Remote Sens.Lett. 5 (8), 733–742. https://doi.org/10.1080/2150704X.2014.963734.