**Abstract:** Wheat yield variability will increase in the future due to the projected increase in extreme weather events and long-term climate change effects. Currently, regional agricultural statistics are used to monitor wheat yield. Remotely sensed vegetation indices have a higher spatio-temporal resolution and could give more insight into crop yield. In this paper, we (i) evaluate the possibility to use Normalized Difference Vegetation Index (NDVI) time series to estimate wheat yield in Latvia and (ii) determine which weather variables impact wheat yield changes using both ALARO-0 and REMO Regional Climate Models (RCM) output. The integral from NDVI series (aNDVI) for winter and spring wheat fields is used as a predictor to model regional wheat yield from 2014 to 2018. A correlation analysis between weather variables, wheat yield and aNDVI was used to elucidate which weather variables impact wheat yield changes in Latvia. Our results indicate that high temperatures in June for spring wheat and in July for winter wheat had a negative correlation with yield. A linear regression yield model explained 71% of the variability with a residual standard error of 0.55 Mg/ha. When RCM data were added as predictor variables to the wheat yield empirical model a random forest approach resulted in better results compared to a linear regression approach, the explained variance increased up to 97% and the residual standard error decreased to 0.17 Mg/ha. We conclude that NDVI time series and RCM output enabled regional crop yield and weather impact monitoring at higher spatio-temporal resolutions than regional statistics.
Keywords: yield estimation; NDVI; regional climate model; winter wheat; spring wheat; Latvia; PROBA-V; ALARO-0; REMO; weather impact

1. Introduction

Extreme weather events and long-term climate change effects will have a large impact on the yield of agriculture crops in the future [1–4]. A negative impact on major food and feed crops is expected in temperate and tropical regions under the climate change scenario of global warming equal to 2 °C or more compared to the pre-industrial era [4]. In some regions, the negative effect of increased temperature and occurrence of extreme events (e.g., drought and pests) on crop production might be offset by the carbon dioxide (CO₂) fertilization effect [4,5]. However, the magnitude of the CO₂ fertilization effect and the importance of other interacting factors is still unclear [4,5]. Considering these scenarios, it will be a challenge to ensure global food security in the future. In Europe, particularly in southern Europe, extreme weather events have already resulted in crop yield losses [2,6,7]. The extreme drought of 2018 resulted in a drop of 8% in European cereal harvest compared to the five-year average [6]. Years with extreme heat waves between 1964 and 2007 resulted in national cereal production deficits of 9.1% worldwide [8].

In order to guarantee food supply, it is crucial to monitor crop yield and understand how weather variability and weather extremes influence crop yield. Accurate and timely crop yield data are indispensable for decision making, as reliable estimations of yield changes are required for strategic planning [9]. Currently, crop yield statistics are predominantly used to monitor and evaluate the causes of crop yield changes. However, these crop yield statistics have their limitations as the spatial resolution and temporal frequency of these data are limited. Data are, for example, often only available at county or regional levels and assessed yearly [10].

Remotely sensed vegetation indices (VIs) have a large scope to fill the crop yield data gap [11]. VIs capture various properties of the Earth’s vegetation by combining the reflectance of specific spectral bands and can be used to monitor crop performance. An advantage of VIs, compared to crop yield statistics, is that information of the crop can be delivered at a high spatio-temporal resolution. The spatial resolution of VIs can be as high as a few meters and some VIs are available at a daily resolution. There is, however, often a tradeoff between the spatial and temporal resolution of VIs. Time series of VIs enable us to evaluate crop development throughout the growing season. In addition, remotely sensed VIs are globally available, allowing us to monitor crop yield at the field level as well as at the global level. Several studies have already demonstrated the possibility to use VIs to monitor yield at different geographical locations and at different scales i.e., from the field to global level [12–17]. One of the VIs that is often and successfully used to estimate crop yield is the Normalized Difference Vegetation Index (NDVI) [14,18,19].

Deriving crop yield from VIs can be performed using a mechanistic approach or an empirical approach [20]. When a mechanistic approach is applied, growth parameters such as: VI, soil characteristics, and management information are used to run and calibrate plant growth models from which crop yield is derived [21]. When an empirical approach is used, a VI is related to crop yield by using statistical methods without describing the causality between the VI and crop yield as performed in mechanistic-based models [14,22]. A simple linear regression model where a VI is modeled as a function of crop yield is an example of an empirical based model. Whereas empirical-based approaches depend on the availability and quality of the data i.e., VI and yield data, mechanistic approaches rely on the assumptions made in the models [20]. The main limiting factor of empirical crop yield models is that the models cannot be easily extrapolated to other areas [12,20].

Wheat is one of the most important cereal crops in Europe, in 2017 wheat constituted 52% of the total harvested cereal production in Europe [23]. European wheat production is also important on a global scale, as Europe accounted for 32.6% of the global wheat production for the period from 1994 to
2017, and wheat yield in Europe is high (i.e., in 2017 wheat yield in Europe was 44 kg/ha compared to 35 kg/ha globally) [23]. However, the occurrence of adverse weather conditions in European wheat production areas is expected to increase due to climate change which will lead to a larger wheat yield variability in Europe [1,24–26]. This advocates wheat yield monitoring in Europe. The aim of this paper is to (i) evaluate the possibility to use NDVI time series to estimate yields and (ii) determine which weather variables impact yield changes. We tested the methodology for spring and winter wheat yield in Latvia. NDVI series at the field level were combined with crop yield statistics at the regional level from 2014 until 2018 using an empirical modelling strategy. In addition, we explored which weather variables impact wheat yield changes in Latvia, using output from the Regional Climate Models ALARO-0 and REMO2015 (henceforth REMO), and subsequent correlation analysis between weather variables, wheat yield, and NDVI.

2. Materials and Methods

2.1. Study Area

Spring and winter wheat yield of Latvia for the period from 2014 to 2018 was studied. For this period, geospatial field data of 2384 wheat fields in Latvia were available through the Land Parcel Information System (LPIS) (Figure 1). Only wheat fields with an area equal to or bigger than 0.1 km$^2$ were used in order to be able to extract pure NDVI pixels from PROBA-V, following [14]. The majority of the fields were located in the Zemgale region (Figure 1) which has the most fertile soils, the dominant World Reference Base Soil Group is Luvisol in this region (Figure 2). In the Latgale region where less fertile Albeluvisols dominate (now named Retisol according to [27]), there were not many wheat fields located (Figures 1 and 2). In the following sections, we describe the methods used to extract NDVI data for the wheat fields and the available crop yield statistics.

![Map of the location and shape of the considered spring (purple) and winter (green) wheat fields in Latvia for each year from 2014 to 2018. The selected fields have an area equal to or bigger than 0.1 km$^2$.](attachment:figure1.png)
2.2. Remote Sensing Data: NDVI

The 10-daily NDVI synthesis images with a resolution of 300 by 300 m from PROBA-V were used. The geospatial field data were combined with the NDVI raster data to extract the pixels for each field with the terrascope platform (https://terrascope.be/en). To ensure that the extracted pixels were located within the field boundaries, an inward buffer of 150 m was applied before extracting the pixels. The extracted pixels were averaged at the field level, resulting in NDVI series consisting of 10 daily observations per field.

The NDVI integral (hereafter referred to as aNDVI) was calculated using the trapezoidal rule. The calculated aNDVI was used to predict spring and winter wheat yield. To ensure that the calculated aNDVI represented the wheat yield growing season, NDVI values measured after the 11th of August (i.e., harvest) and below the threshold of 0.2 were discarded. The NDVI threshold of 0.2 resulted in the best wheat yield predictions [14]. To make sure that the calculated aNDVI captured the different phenological stages (from green-up to harvest) which are important for wheat yield, only the NDVI curves that had at least ten continuous observations were retained for the calculation of aNDVI series (Figure 3). For the calculation of aNDVI, 339 fields, i.e., 14% of the initial number of fields, were discarded because their NDVI series had less than ten observations. Thus, in total, 2045 wheat fields (1561 winter wheat fields and 484 spring wheat fields) were used for the calculation of aNDVI.
Figure 2. The fitted linear regression model can be written as

\[ Y_i = \alpha + \beta_1 \text{crop} + \beta_2 \text{region} + \beta_3 \text{NDVI}_{\text{mean}} + \beta_4 \text{NDVI}_{\text{integral}} + \beta_5 \text{NDVI}_{\text{max}} + \beta_6 \text{NDVI}_{\text{median}} + \epsilon_i \]

where \( Y_i \) is the yield of field \( i \), \( \text{crop} \) is a factor indicating if the crop was winter or spring wheat, \( \text{region} \) is a factor indicating the region where the field was located, \( \text{NDVI}_{\text{mean}}, \text{NDVI}_{\text{integral}}, \text{NDVI}_{\text{max}}, \text{NDVI}_{\text{median}} \) are NDVI series of spring and winter wheat to estimate and evaluate the yield of the fields. Two factors were added to the linear model. First, a factor indicating if the crop was winter or spring wheat. Second, a factor indicating the region where the field was located. By adding this factor to the model, regional differences such as the dominant Reference Soil Group in each region were taken into account (Figure 2). The fitted linear regression model can be written as

2.3. Crop yield Statistics Data

Yearly regional crop yield statistics from 2014 to 2018 for spring and winter wheat in Latvia for the NUTS 3 (Nomenclature of Territorial Units for Statistics level 3) regions: Pierīga, Vidzeme, Kurzeme, Zemgale, and Latgale were provided by the Central Statistical Bureau of Latvia (https://www.csb.gov.lv/) (Figure 1). The regional crop yield statistics were used to determine the yield of each field based on its location (i.e., in which region it was located). If a field was located in more than one region, the field centroid was used to determine the region.

2.4. Data Analysis

2.4.1. Wheat Yield Empirical Model

A linear regression model with aNDVI as a predictor variable and statistical yield (i.e., crop yield statistics reported by the Central Statistical Bureau of Latvia) as a response variable was made for each field to evaluate the possibility of using NDVI series of spring and winter wheat to estimate and evaluate the yield of the fields. Two factors were added to the linear model. First, a factor indicating if the crop was winter or spring wheat. Second, a factor indicating the region where the field was located. By adding this factor to the model, regional differences such as the dominant Reference Soil Group in each region were taken into account (Figure 2). The fitted linear regression model can be written as

\[ Y = \beta_0 + \beta_1 \text{crop} + \beta_2 \text{region} + \beta_3 \text{NDVI}_{\text{mean}} + \beta_4 \text{NDVI}_{\text{integral}} + \beta_5 \text{NDVI}_{\text{max}} + \beta_6 \text{NDVI}_{\text{median}} + \epsilon \]

where \( \beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6 \) are the coefficients of the model.
\[ \text{Yield}_i = \alpha + \beta_1 \text{aNDVI}_i + \beta_2 \text{crop}_i + \beta_3 R1_i + \beta_4 R2_i + \beta_5 R3_i + \beta_6 R4_i + \epsilon_i \]  

(1)

with \( \text{Yield}_i \) = the yield of field \( i \); \( \text{aNDVI}_i \) = the calculated aNDVI of field \( i \); \( \text{crop}_i \) = 1 if crop on field \( i \) is winter wheat and \( \text{crop}_i \) = 0 if crop on field \( i \) is spring wheat; \( R1_i = 1 \), \( R2_i = 1 \), \( R3_i = 1 \), or \( R4_i = 1 \) if field \( i \) is located in the Latgale, Pierīga, Vidzeme, Zemgale, region, respectively, or \( R1_i = 0 \), \( R2_i = 0 \), \( R3_i = 0 \), \( R4_i = 0 \) if it is located in another region. \( \alpha \), \( \beta_1 \), \( \beta_2 \), \( \beta_3 \), \( \beta_4 \), \( \beta_5 \), and \( \beta_6 \) are the coefficients of the model, and \( \epsilon_i \) is the error of the model. The linear regression model performance was evaluated with the explained variance, the residual standard error, and the Akaike Information Criterion (AIC) [28].

In addition to a linear regression model, a random forest regression was applied. The latter was built based on 100 trees and evaluated with the explained variance. Similar to the linear regression model, the variables aNDVI, a factor crop indicating if the crop was winter or spring wheat, and a factor indicating the region where the field was located were used in the random forest model to predict wheat yield. The random forest model was build using the randomForest function, which implements the Breiman’s algorithm, from the R package randomForest [29]. The results of the two modeling approaches were compared.

2.4.2. Regional Climate Modeling

We used the field level monthly mean values of minimum and maximum temperatures (henceforth T\( \text{Min} \) and T\( \text{Max} \)) from the REMO [30,31] and ALARO-0 Regional Climate Models (RCM) [32,33], both driven by lateral boundary conditions from the ERA-Interim reanalysis [34] using the Davies procedure [35], as prepared according to [35,36]. T\( \text{Min} \) and T\( \text{Max} \) were selected based on preliminary analysis using climate station data which indicated that wheat yield and aNDVI were more strongly correlated with T\( \text{Min} \) and T\( \text{Max} \), and less with precipitation, reference evapotranspiration, or frost occurrence (defined as number of days with T\( \text{Min} \) < \(-15 \) °C). Validation of the variables for the period from 1980 to 2017 over Latvia showed that the modeled T\( \text{Min} \) and T\( \text{Max} \) contained biases compared to station observations [37]. For example, REMO showed a cold T\( \text{Min} \) bias in winter and in February in particular, and underestimated T\( \text{Max} \) in summer in June and July. However, the spatial patterns for temperature were similar between RCMs and observations. [37] concluded that both ALARO and REMO could be applied in impact modelling. These high-resolution climate data can be freely downloaded from the Earth System Grid Federation (ESGF) data nodes (http://esgf.llnl.gov/). Monthly mean values of T\( \text{Min} \) and T\( \text{Max} \) were available from 2014 to 2018 for the REMO and ALARO-0 models. RCM climate data were preferred over climate station data as their spatial resolution is much higher. In addition, RCM data form the basis for the analysis of climate projections and could thus be used to evaluate the effect of future climate on wheat yield.

2.4.3. Evaluating the Effect of Weather on Wheat Yield

To evaluate the effect of weather on wheat yield, the Pearson correlation between temperatures (T\( \text{Min} \) and T\( \text{Max} \)) and aNDVI, and the correlation between temperatures (T\( \text{Min} \) and T\( \text{Max} \)) and wheat yield, were calculated. Considering the growing season of wheat in Latvia, the correlation was calculated for the months from January to August.

Finally, we tested if REMO and ALARO-0 RCM could be used to improve wheat yield modeling. Therefore, in a second modeling stage, monthly mean RCM temperatures which showed a high correlation with wheat yield and aNDVI were added as predictors to the wheat yield linear and random forest models. Separate linear regressions and random forest models were built for the selected REMO and ALARO-0 temperature data.
3. Results

3.1. NDVI Series of Winter and Spring Wheat

The average NDVI series from 2014 to 2018 for the studied winter and spring wheat fields in Latvia are visualized in Figure 4. For winter wheat, the NDVI peak was around mid-May, while for spring wheat the NDVI peak was later in the growing season—around mid-June. For both winter and spring wheat, the NDVI dropped around the end of August which coincided with the harvest time of winter and spring wheat in Latvia.

![Average NDVI series for winter wheat (blue line) and spring wheat (red line) for 2045 fields (1561 winter wheat fields and 484 spring wheat fields) in Latvia from 2014 to 2018.](image)

3.2. aNDVI and Crop Yield Statistics

The NDVI series used to calculate the aNDVI for spring and winter wheat for each of the five NUTS 3 regions are visualized in Figure 5. The Zemgale region had the highest number of wheat fields, whereas the Latgale region had the lowest number of wheat fields. The number of fields used for the calculation of aNDVI for each year for spring and winter wheat is presented in Figure 6. Note that the number of fields in the years 2014 and 2018 was lower compared to the other years. The number of winter wheat fields was higher than the number of spring wheat fields, except in 2014. This is because a large proportion of winter wheat fields were damaged by frost due to a cold spell in mid-January in 2014. These fields were subsequently re-sown with spring wheat [38].
Figure 5. NDVI series used to calculate the aNDVI for spring and winter wheat for the NUTS 3 regions of Latvia: Kurzeme, Latgale, Pieriša, Vidzeme, and Zemgale.

In Figure 7, the yield at the regional level, as reported by the Central Statistical Bureau of Latvia for spring and winter wheat for the years from 2014 to 2018, is visualized. For spring wheat, a low yield was reported in 2018. For winter wheat there was a low yield reported in the years 2014 and 2018 (Figure 7).

Similar patterns were visible when the aNDVI calculated for the studied fields was plotted for each year for spring and winter wheat (Figure 8). Nonetheless, the variation in aNDVI between the years (Figure 8) was not as pronounced as compared to the variation in yield between the years (Figure 7).
Figure 6. Number of spring and winter wheat fields for the years from 2014 to 2018 for which aNDVI was calculated.

Figure 7. Boxplot of the yield (in Mg/ha) for spring and winter wheat for the years from 2014 to 2018. Yield data were provided by the Central Statistical Bureau of Latvia at the regional level for the regions Pieriņa, Vidzeme, Kurzeme, Zemgale, and Latgale.
3.3. Wheat Yield Empirical Model

The statistics of the coefficients (estimate, standard error, t-value, and p-value) of the wheat yield empirical linear regression model (Equation (1)) are presented in Table 1. All the coefficients of the linear regression model were significant ($p < 0.001$). The explained variance of the linear regression model was 71% and the residual standard error was 0.55 Mg/ha. When a random forest regression model was used instead of a linear regression model, the residual standard error of the random forest model was equal to 0.58 Mg/ha and the explained variance by the model decreased to 62%. Therefore, the linear regression model performed better. The explained variance of the linear regression model decreased by 23% when the region factor was not included in the model. The AIC increased from 3400 to 4574 when region was not included in the model. When neither crop nor region factors were included in the model, the explained variance decreased by 27%—from 71% to 44%. In addition, the AIC increased from 3400 to 4724 when neither crop nor region were included in the model. Allowing the slope of the linear regression model to change depending on the region, i.e., including an interaction term between aNDVI and region, did not result in the improvement of the model accuracy and resulted in non-significant parameter estimates. The proposed linear regression model with the factor crop, indicating if the crop is spring or winter wheat, and the factor region, indicating in which region the field is located, explains the wheat yield variability well. In Figure 9, the yield, as predicted by the empirical linear regression model, is plotted versus the observed yield for spring and winter wheat. The adjusted $R^2$ ($R^2 = 0.71$) and the residual standard error of the linear regression model fit (0.55 Mg/ha) resulted in good estimates of the model parameters (Table 1). However, the linear regression model overestimated the yield of winter wheat slightly in the years 2014 and 2018 (Figure 9) and slightly underestimated the 2015 yield of spring wheat.
Table 1. Statistics of the coefficients of the wheat yield empirical linear regression model. See Equation (1) for the model description.

| Coefficient | Estimate | Std. Error | t-Value | Pr(>|t|) |
|-------------|----------|------------|---------|---------|
| α           | 1.794    | 0.067      | 26.697  | < 0.001 |
| β₁          | 0.033    | 0.001      | 29.203  | < 0.001 |
| β₂          | 0.661    | 0.035      | 18.705  | < 0.001 |
| β₃          | -0.702   | 0.070      | -10.049 | < 0.001 |
| β₄          | 0.174    | 0.048      | 3.630   | < 0.001 |
| β₅          | -0.403   | 0.053      | -7.634  | < 0.001 |
| β₆          | 0.846    | 0.034      | 24.556  | < 0.001 |

Figure 9 indicates that the overestimation of winter wheat yield by the linear regression model is similar in all regions in 2014. However, in 2018 the overestimation of winter wheat yield was mostly for the fields located in the Kurzeme region. The underestimation of spring wheat yield in 2015 was similar in all regions.

Figure 9. The yield predicted by the empirical linear regression model versus the observed yield for each year for spring (top row) and winter wheat (bottom row). The colors indicate the different NUTS 3 regions. The black line indicates the 1:1 reference line.

3.4. Effect of Tmin and Tmax on Wheat Yield

In Figure 10, the Pearson correlation between the monthly means of temperature (Tmax and Tmin) and aNDVI, and the monthly means of temperature (Tmax and Tmin) and yield for spring wheat is visualized for the months from January to August. The most negative correlation between monthly means of temperature and yield, and temperature and aNDVI, was found in June, indicating that high temperatures in June resulted in lower aNDVI and yield values for spring wheat. In the month of March, there was a high positive correlation between Tmin and yield, and Tmax and yield. The correlations between aNDVI and Tmin, and aNDVI and Tmax of March, showed similar high positive correlations, indicating that high temperatures in March resulted in high aNDVI and yield values for spring wheat.
Figure 10. Pearson correlation between monthly means of minimum and maximum temperatures (Tmax and Tmin) and aNDVI and yield for spring wheat. Monthly means of Tmin and Tmax of the months from January to August are used to calculate the correlation. Top graph: Pearson correlation between TminALARO and aNDVI (red circles), TminREMO and aNDVI (blue circles), TminALARO and yield (red triangles), and TminREMO and yield (blue triangles). Bottom graph: Pearson correlation between TmaxALARO and aNDVI (red circles), TmaxREMO and aNDVI (blue circles), TmaxALARO and yield (red triangles), and TmaxREMO and yield (blue triangles).

In Figure 11, the Pearson correlation between the monthly means of temperature (Tmax and Tmin) and aNDVI, and the monthly means of temperature (Tmax and Tmin) and yield for winter wheat, is visualized for the months from January to August. The most negative correlation between temperature and aNDVI, and temperature and yield, was observed in July, indicating that high temperatures in July resulted in lower aNDVI and yield values for winter wheat. The most positive correlations between temperature and yield, and between temperature and aNDVI, were observed in January.
Figure 11. Pearson correlation between monthly means of temperature (Tmax and Tmin) and aNDVI and yield for winter wheat. Monthly means of Tmin and Tmax of the months from January to August are used to calculate the correlation. Top graph: Pearson correlation between TminALARO and aNDVI (red circles), TminREMO and aNDVI (blue circles), TminALARO and yield (red triangles), and TminREMO and yield (blue triangles). Bottom graph: Pearson correlation between TmaxALARO and aNDVI (red circles), TmaxREMO and aNDVI (blue circles), TmaxALARO and yield (red triangles), and TmaxREMO and yield (blue triangles).

In Figure 12, wheat yield is plotted as a function of maximum temperature in June and March, and July and January, for spring and winter wheat yield, respectively. These are the two months which showed the strongest correlation with spring and winter wheat. The highest spring wheat yields were found in the years with low temperatures in June and high temperatures in March. The highest winter wheat yields were found in the years with low temperatures in July and high temperatures in January.
The explained variance by the random forest model was higher compared to the linear regression model. When a random forest modeling approach was used, the explained variance was equal to 97% for the ALARO-0 model and 94% for the REMO model, when including July maximum temperatures in the random forest models.

The residual standard error decreased from 0.55 to 0.47 Mg/ha for the ALARO-0 model and 0.23 Mg/ha for the REMO model, when including the July maximum temperature in the random forest models. The residual standard error was equal to 0.17 Mg/ha for the ALARO-0 model and 0.23 Mg/ha for the REMO model, when including July maximum temperatures in the random forest models.

The highest July temperatures were observed in 2014 and 2018 (Figure 13), but the correlation of the monthly means of temperature and yield for winter wheat suggests that high temperatures in July resulted in lower winter wheat yields (Figure 11). The highest July temperatures were observed in 2014 and 2018 (Figure 13), but their negative impact on winter wheat yield did not show in aNDVI. This might explain the overestimation of modeled winter wheat in the years 2014 and 2018. Considering this observation, the wheat yield model could improve by adding explanatory variables which are known to affect wheat yield such as Tmin or Tmax of July. Indeed, when Tmax of July from the ALARO-0 model or REMO model were added to the linear regression model the explained variance increased by 10%—from 71% to 79% and 11%—from 71% to 80%, respectively. The AIC of the wheat linear regression model decreased from 3400 to 2615 for the ALARO-0 model and to 2702 for the REMO model, when including the July maximum temperatures. The residual standard error decreased from 0.55 to 0.47 Mg/ha and to 0.46 Mg/ha, respectively, when Tmax of July from ALARO-0 or REMO were added to the linear regression model. However, when July maximum temperatures were included in the model, the explained variance by the random forest model was higher compared to the linear regression model. When a random forest modeling approach was used, the explained variance was equal to 97% for the ALARO-0 model and 94% for the REMO model, when including July maximum temperatures. The residual standard error was equal to 0.17 Mg/ha for the ALARO-0 model and 0.23 Mg/ha for the REMO model, when including July maximum temperatures in the random forest models.

Figure 12. Top row: scatter plot of maximum temperature in June versus maximum temperature in March for the spring wheat fields. Each point represents a field, the color of the points indicates the yield of the spring wheat field. Bottom row: scatter plot of maximum temperature in July versus maximum temperature in January for the winter wheat fields. Each point represents a field, the color of the points indicates the yield of the winter wheat field. Monthly maximum temperature as modeled by both the REMO (left graphs) and the ALARO-0 (right graphs) models were used.

3.5. Regional Wheat Yield Modelling

For winter wheat, the modeled yield was overestimated in 2014 in all regions and in 2018 for fields located in the Kurzeme region (Figures 9 and 13). The years with the lowest yields were also 2014 and 2018 (Figure 7). This suggests that aNDVI did not capture one or more factors affecting low winter wheat yields observed in 2014 and 2018. The correlation of the monthly means of temperature and yield for winter wheat suggests that high temperatures in July resulted in lower winter wheat yields (Figure 11). The highest July temperatures were observed in 2014 and 2018 (Figure 13), but their negative impact on winter wheat yield did not show in aNDVI. This might explain the overestimation of modeled winter wheat in the years 2014 and 2018. Considering this observation, the wheat yield model could improve by adding explanatory variables which are known to affect wheat yield such as Tmin or Tmax of July. Indeed, when Tmax of July from the ALARO-0 model or REMO model were added to the linear regression model the explained variance increased by 10%—from 71% to 79% and 11%—from 71% to 80%, respectively. The AIC of the wheat linear regression model decreased from 3400 to 2615 for the ALARO-0 model and to 2702 for the REMO model, when including the July maximum temperatures. The residual standard error decreased from 0.55 to 0.47 Mg/ha and to 0.46 Mg/ha, respectively, when Tmax of July from ALARO-0 or REMO were added to the linear regression model. However, when July maximum temperatures were included in the model, the explained variance by the random forest model was higher compared to the linear regression model. When a random forest modeling approach was used, the explained variance was equal to 97% for the ALARO-0 model and 94% for the REMO model, when including July maximum temperatures. The residual standard error was equal to 0.17 Mg/ha for the ALARO-0 model and 0.23 Mg/ha for the REMO model, when including July maximum temperatures in the random forest models.
Adding explanatory variables which are known to affect wheat yield also improved the explained variance of the individual spring and winter wheat linear regression models. The explained variance of the winter wheat linear regression model increased from 59% to 77% for the ALARO-0 model and to 84% for the REMO model, when including the July maximum temperatures. The AIC of the winter wheat linear regression model decreased from 2690 to 1794 for the ALARO-0 model and to 1258 for the REMO model, when including July maximum temperatures. Both linear regression models indicated that July maximum temperatures had a negative effect on winter wheat yield. The residual standard error decreased from 0.57 to 0.43 Mg/ha and to 0.36 Mg/ha, respectively, when July maximum temperatures from ALARO-0 or REMO models were added to the winter wheat linear regression model. Likewise, the explained variance of the spring wheat linear regression model increased from 59% to 83% for the ALARO-0 model and to 77% for the REMO model, when including the June maximum temperatures. The AIC of the spring wheat linear regression model decreased from 661 to 246 for the ALARO-0 model and to 387 for the REMO model, when including the June maximum temperatures. Both linear regression models indicated that June maximum temperatures had a negative effect on spring wheat yield. The residual standard error decreased from 0.48 to 0.31 Mg/ha and to 0.36 Mg/ha, respectively, when June maximum temperatures from ALARO-0 or REMO models were added to the spring wheat linear regression model.

The explained variance was higher when a random forest regression modelling approach was used for both the winter and spring wheat model when including July and June maximum temperatures, respectively. For the winter wheat random forest model, the explained variance was equal to 98% for the ALARO-0 model and to 97% for the REMO model, when including July maximum temperatures. The residual standard error was equal to 0.12 Mg/ha for the ALARO-0 model and to 0.16 Mg/ha for the REMO model, when including July maximum temperatures in the random forest winter wheat model. For the spring wheat random forest model, the explained variance was equal to 92% for the ALARO-0 model and to 93% for the REMO model, when including June maximum temperatures. The residual standard error was equal to 0.21 Mg/ha for the ALARO-0 model and to 0.20 Mg/ha for the REMO model, when including June maximum temperatures in the random forest spring wheat model.
4. Discussion

The statistics of the empirical models showed that there is a scope to use 10-daily NDVI series with a 300 m resolution provided by satellites with a PROBA-V configuration to estimate spring and winter wheat yield in Latvia (Table 1 and Figure 9). The variance explained by the wheat linear regression model decreased by 22% when the factor region was not included as a predictor factor in the wheat yield model. This is likely related to the fact that yield data (the response variable) reflect the regional environment. Wheat yield variability may not be fully captured by aNDVI and may need additional information, such as, for example, differences in soil characteristics, used wheat varieties, and management practices between the regions, although the analysis by [14] suggests that NDVI time series are sufficient to estimate yields on a parcel to regional scale.

In general, the model performance is in the same range as that of other studies using PROBA-V NDVI data to predict wheat yield [14,21,22]. Zheng, Y. et al. used NDVI to estimate the aboveground biomass with a light use efficiency model [21], and subsequently predicted wheat yield. Durgun, Y.Ö. et al. fitted an asymmetric double sigmoid model on the NDVI series and used the integration of the fitted function as a predictor for wheat yield [14]. Compared to these studies, our methodology of computing the integral of the NDVI series (i.e., aNDVI) as a predictor of wheat yield is straightforward. However, a prerequisite is the availability of continuous NDVI series. Methodologies which combine different remotely sensed data—such as vegetation indicators and microwave vegetation optical depth data—can be used to estimate crop yield in regions where continuous NDVI series are scarce due to the frequent presence of clouds [17]. The developed wheat modelling method can be used in other wheat producing areas. In addition, NDVI, aNDVI, and meteorological variables may capture only part of the important wheat yield-influencing variables. In other regions’ local soil fertility conditions, the effects of variety and weather extremes may explain a much larger proportion of the yield variation. The explained variance by the random forest regression models that included RCM data as a predictor variable was higher compared to the linear regression models. Therefore, in regions where RCM data are available, a random forest regression modeling approach is recommended for estimating regional wheat yields.

Durgun, Y.Ö. et al. studied whether better wheat yield estimations could be obtained with thermal instead of calendar time for the integration of PROBA-V NDVI series at spatial resolutions of 100 m, 300 m, and 1 km [14]. Thermal time outperformed calendar time for estimating winter wheat, but a notable exception occurred for fields with a pixel purity above 65% at 300 m spatial resolution. However, due to low sample size at 300 m resolution no general conclusion could be made by [14]. The present study seems to confirm the observation of [14], since we obtained good wheat yield estimations with pure pixel NDVI data at 300 m resolution projected on calendar time (i.e., aNDVI). The use of thermal time did not improve wheat yield estimations. The explained variance decreased by 4% when thermal time instead of calendar time was used in the linear regression wheat yield model.

In general, the variation in correlation from January to August between monthly means of temperature and wheat yield on the one hand, and monthly means of temperature and aNDVI on the other hand, were quite similar.

The variation in correlation of monthly mean temperature and winter and spring wheat yield and aNDVI between the different months from January to August indicated that temperature during some months influenced the yield and aNDVI more than others. For spring wheat, high temperatures in June resulted in lower yield and lower aNDVI, while for winter wheat the same pattern was found in July. The negative effect of temperature in June (for spring wheat) and July (for winter wheat) on aNDVI can be linked to important phenological stages of winter and spring wheat. This observation indicates that the variability in wheat yield in Latvia is likely to increase in the future. Indeed, the average temperature as well as the occurrence of extremely hot days will further increase due to climate change. The negative effect of high temperatures on wheat yield in June (for spring wheat) and July (for winter wheat) confirms similar observations in Europe [39,40] and should, therefore, be considered in climate change adaptation strategies [2,26,41]. From the observed positive correlation between temperature in January and winter wheat yield, we could conclude that projected increases in winter
temperatures could have a positive effect on winter wheat yield. However, dehardening (i.e., loss of freezing tolerance) of wheat is also triggered by higher temperatures [42] and may negatively impact crop performance. Winter wheat growing at high temperatures in January generally loses its freezing tolerance and will be more susceptible to low temperatures in spring [43]. In addition, in the case of, due to global warming, both January and July temperatures increase, the positive effect of high January temperatures on winter wheat yield could be offset by the negative effect of high July temperatures on winter wheat yield. A similar remark can be made for spring wheat, as the positive effect of high March temperatures on spring wheat yield is likely to be offset by the negative effect of high June temperatures on spring wheat yield. These phenomena are already visible at present. In years with high temperatures for both March and June the spring wheat yield is moderate, and in years with high temperatures for both January and July, winter wheat yield is moderate (Figure 12).

5. Conclusions

Wheat yield was estimated from NDVI time series for spring and winter wheat in Latvia. The wheat yield was derived from the integral of the NDVI series (i.e., aNDVI) using a simple methodology which can be applied to other regions and crops where continuous NDVI series are available. We can conclude that adding (i) region as a factor, (ii) both the region and crop type, and (iii) explanatory weather variables from regional climate models (RCM) to the model improved wheat yield prediction. The importance of adding region as a factor to the model might be explained by regional differences in the dominant Reference Soil Group. Allowing the slope of the model to change depending on the region did not improve wheat yield prediction. Our results suggest that there is a large scope to use NDVI-derived crop yield estimates, such as aNDVI, and that RCM output plays a valuable role in understanding regional weather impacts on yield. In the case that RCM data are available, a random forest model approach is recommended to model regional wheat yields. The high spatial extent of the VI-derived crop yield estimates is an important benefit compared to the commonly used crop yield statistics that are often only available at the regional level. Likewise, the spatio-temporal availability of the regional climate model output is superior to sparsely available climate stations. Using aNDVI and the RCMs ALARO-0 and REMO, we found that high temperatures during June (for spring wheat) and July (for winter wheat) influenced the wheat yield negatively in Latvia. High temperatures in March and January had a positive effect on spring and winter wheat yield, respectively. However, this positive effect of high temperatures was found to be offset when temperatures were high in both March and June, and January and July for spring and winter wheat, respectively. This information should be considered in weather impact analysis and climate change adaptation strategies such as, for example, wheat breeding programs.

Author Contributions: Conceptualization and methodology, A.G. and A.V. (Astrid Vannoppen); validation, formal analysis, writing—original draft preparation, A.V. (Astrid Vannoppen); data curation and writing—review and editing, A.G., A.V. (Astrid Vannoppen), S.C., L.B. (Lars Buntemeyer), A.R.R., L.D.C., L.K., S.T., R.D.T., R.H., B.V.S., P.T., S.A., L.B. (Leonid Bobylev), N.G., A.S., H.V.D.V. and A.V. (Andris Vīksna); supervision, A.G., funding acquisition, A.G. All authors have read and agreed to the published version of the manuscript.

Funding: The project is granted by the ERA.Net RUS Plus Initiative, ID 166. VITO received funding from the Research Foundation-Flanders (FWO). NIERSC received funding from the Russian Foundation for Basic Research (RFBR)-grant № 18-55-76004. HZG-GERICS received funding from the Federal Ministry of Education and Research (BMBF).

Acknowledgments: The computational resources and services for the ALARO-0 regional climate simulations were provided by the Flemish Supercomputer Center (VSC), funded by the Research Foundation-Flanders (FWO) and the Flemish Government department EWI. The CORDEX-CORE REMO simulations were performed under the HZG-GERICS share at the German Climate Computing Centre (DKRZ).

Conflicts of Interest: The authors declare no conflict of interest.
References

1. Kahiluoto, H.; Kaseva, J.; Balek, J.; Olesen, J.E.; Ruiz-Ramos, M.; Gobin, A.; Kersebaum, K.C.; Takáč, J.; Ruget, F.; Ferrise, R.; et al. Decline in climate resilience of European wheat. *Proc. Natl. Acad. Sci. USA* 2019, 116, 123–128. [CrossRef]

2. Mäkinen, H.; Kaseva, J.; Trnka, M.; Balek, J.; Kersebaum, K.C.; Nendel, C.; Gobin, A.; Olesen, J.E.; Bindi, M.; Ferrise, R.; et al. Sensitivity of European wheat to extreme weather. *Field Crop. Res.* 2018, 222, 209–217. [CrossRef]

3. Moore, F.C.; Lobell, D.B. The fingerprint of climate trends on European crop yields. *Proc. Natl. Acad. Sci. USA* 2015, 112, 2670–2675. [CrossRef] [PubMed]

4. Porter, J.R.; Xie, L.; Challinor, A.J.; Cochrane, K.; Howden, S.W.; Iqbal, M.M.; Lobell, D.B.; Travasso, M.I. Food security and food production systems. In *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; Cambridge University Press: Cambridge, UK; New York, NY, USA, 2014; pp. 485–533.

5. Balkovič, J.; Skalský, R.; Folberth, C.; Khabarov, N.; Schmid, E.; Madaras, M.; Obersteiner, M.; van der Velde, M. Impacts and Uncertainties of +2 °C of Climate Change and Soil Degradation on European Crop Calorie Supply. *Earth’s Future* 2018, 6, 373–395. [CrossRef]

6. European Commission Short-term Outlook for EU Agricultural Markets in 2018 and 2019 Report nr 22. Available online: https://ec.europa.eu/info/sites/info/files/food-farming-fisheries/farming/documents/agri-short-term-outlook-reports_2018.zip (accessed on 12 April 2019).

7. Jacobs, C.; Berglund, M.; Kurnik, B.; Dworak, T.; Marras, S.; Mereu, V.; Michetti, M. Climate Change Adaptation in the Agriculture Sector in Europe. European Environment Agency: Luxembourg, 2019; ISBN 978-92-9480-072-5.

8. Lesk, C.; Rowhani, P.; Ramankutty, N. Influence of extreme weather disasters on global crop production. *Nature* 2016, 529, 84–87. [CrossRef] [PubMed]

9. Doraiswamy, P.C.; Akhmedov, B.; Beard, L.; Stern, A.; Mueller, R. Operational prediction of crop yields using MODIS data and products. In Proceedings of the International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, International Society for Photogrammetry and Remote Sensing (ISPRS) working group VIII/10 Workshop, Stresa, Italy, 30 November–1 December 2006.

10. Milesi, C.; Samanta, A.; Hashimoto, H.; Kumar, K.K.; Ganguly, S.; Thenkabail, P.S.; Srivastava, A.N.; Nemani, R.R.; Myneni, R.B. Decadal Variations in NDVI and Food Production in India. *Remote Sens.* 2010, 2, 758–776. [CrossRef]

11. Rembold, F.; Atzberger, C.; Savin, I.; Rojas, O. Using Low Resolution Satellite Imagery for Yield Prediction and Yield Anomaly Detection. *Remote Sens.* 2013, 5, 1704–1733. [CrossRef]

12. Becker-Reshef, I.; Vermote, E.; Lindeman, M.; Justice, C. A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data. *Remote Sens. Environ.* 2010, 114, 1312–1323. [CrossRef]

13. Bolton, D.K.; Friedl, M.A. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. *Agric. For. Meteorol.* 2013, 173, 74–84. [CrossRef]

14. Durgun, Y.Ö.; Gobin, A.; Duveiller, G.; Tychon, B. A study on trade-offs between spatial resolution and temporal sampling density for wheat yield estimation using both thermal and calendar time. *Int. J. Appl. Earth Obs. Geoinf.* 2020, 86, 101988. [CrossRef]

15. Durgun, Y.Ö.; Gobin, A.; Gilliams, S.; Duveiller, G.; Tychon, B. Testing the Contribution of Stress Factors to Improve Wheat and Maize Yield Estimations Derived from Remotely-Sensed Dry Matter Productivity. *Remote Sens.* 2016, 8, 170. [CrossRef]

16. Ferencz, C.; Bognár, P.; Lichtenberger, J.; Hamar, D.; Tarcsei, G.; Timár, G.; Molnár, G.; Pásztor, S.; Steinbach, P.; Székely, B.; et al. Crop yield estimation by satellite remote sensing. *Int. J. Remote Sens.* 2004, 25, 4113–4149. [CrossRef]

17. Mateo-Sanchois, A.; Piles, M.; Muñoz-Mari, J.; Adua, J.E.; Pérez-Suay, A.; Camps-Valls, G. Synergistic integration of optical and microwave satellite data for crop yield estimation. *Remote Sens. Environ.* 2019, 234, 111460. [CrossRef] [PubMed]
18. Genovese, G.; Vignolles, C.; Negrè, T.; Passera, G. A methodology for a combined use of normalised difference vegetation index and CORINE land cover data for crop yield monitoring and forecasting. A case study on Spain. *Agronomic 2001*, 21, 91–111. [CrossRef]

19. Rojas, O. Operational maize yield model development and validation based on remote sensing and agro-meteorological data in Kenya. *Int. J. Remote Sens.* **2007**, 28, 3775–3793. [CrossRef]

20. Weiss, M.; Jacob, F.; Duveiller, G. Remote sensing for agricultural applications: A meta-review. *Remote Sens. Environ.* **2020**, 236, 111402. [CrossRef]

21. Zheng, Y.; Zhang, M.; Zhang, X.; Zeng, H.; Wu, B. Mapping Winter Wheat Biomass and Yield Using Time Series Data Blended from PROBA-V 100- and 300-m S1 Products. *Remote Sens. 2016*, 8, 824. [CrossRef]

22. Meroni, M.; Fasbender, D.; Balaghi, R.; Dali, M.; Haffani, M.; Haythem, I.; Hooker, J.; Lahlou, M.; Lopez-Lozano, R.; Mahyou, H.; et al. Evaluating NDVI Data Continuity Between SPOT-VEGETATION and PROBA-V Missions for Operational Yield Forecasting in North African Countries. *IEEE Trans. Geosci. Remote Sens. 2016*, 54, 795–804. [CrossRef]

23. FAOSTAT. Production Share of Wheat by Region. Average 1994–2017. Available online: [http://www.fao.org/faostat/en/#data/QC/visualize](http://www.fao.org/faostat/en/#data/QC/visualize) (accessed on 12 April 2019).

24. Asseng, S.; Ewert, F.; Martre, P.; Rötter, R.P.; Lobell, D.B.; Cammarano, D.; Kimball, B.A.; Ottman, M.J.; Wall, G.W.; White, J.W.; et al. Rising temperatures reduce global wheat production. *Nat. Clim. Chang.* **2015**, 5, 143–147. [CrossRef]

25. Schils, R.; Olesen, J.E.; Kersebaum, K.–C.; Rijk, B.; Oberforster, M.; Kalyada, V.; Khitrykau, M.; Gobin, A.; Kirchev, H.; Manolova, V.; et al. Cereal yield gaps across Europe. *Eur. J. Agron.* **2018**, 101, 109–120. [CrossRef]

26. Trnka, M.; Rötter, R.P.; Ruiz-Ramos, M.; Kersebaum, K.C.; Olesen, J.E.; Žalud, Z.; Semenov, M.A. Adverse weather conditions for European wheat production will become more frequent with climate change. *Nat. Clim. Chang.* **2016**, 4, 637–643. [CrossRef]

27. FAO. *World Reference Base for Soil Resources 2014: International Soil Classification System for Naming Soils and Creating Legends for Soil Maps*; FAO: Rome, Italy, 2014; ISBN 978-92-5-108369-7.

28. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, 19, 716–723. [CrossRef]

29. Liaw, A.; Wiener, M. Classification and Regression by randomForest. *R News* **2002**, 2, 18–22.

30. Jacob, D.; Elizalde, A.; Haensler, A.; Hagemann, S.; Kumar, P.; Podzun, R.; Rechid, D.; Remedio, A.R.; Saeed, F.; Sieck, K.; et al. Assessing the Transferability of the Regional Climate Model REMO to Different COordinated Regional Climate Downscaling Experiment (CORDEX) Regions. *Atmosphere* **2012**, 3, 181–199. [CrossRef]

31. Remedio, A.R.; Teichmann, C.; Buntemeyer, L.; Sieck, K.; Weber, T.; Rechid, D.; Hoffmann, P.; Nam, C.; Kotova, L.; Jacob, D. Evaluation of New CORDEX Simulations Using an Updated Köppen–Trewartha Climate Classification. *Atmosphere 2019*, 10, 726. [CrossRef]

32. Giot, O.; Termonia, P.; Degrawue, D.; Troch, R.D.; Caluwaerts, S.; Smet, G.; Berckmans, J.; Deckmyn, A.; Cruz, I.D.; Meutter, P.D.; et al. Validation of the ALARO-0 model within the EURO-CORDEX framework. *Geosci. Model Dev.* **2016**, 9, 1143–1152. [CrossRef]

33. Termonia, P.; Fischer, C.; Bazile, E.; Bouyssel, F.; Brožková, R.; Bénard, P.; Bochenek, B.; Degrawue, D.; Derkóvá, M.; Khatib, R.E.; et al. The ALADIN System and its canonical model configurations AROME CY41T1 and ALARO CY40T1. *Geosci. Model Dev.* **2018**, 11, 257–281. [CrossRef]

34. Dee, D.P.; Uppala, S.M.; Simmons, A.J.; Berrisford, P.; Poli, P.; Kobayashi, S.; Andrae, U.; Balmaseda, M.A.; Balsamo, G.; Bauer, P.; et al. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Q. J. Royal Meteorol. Soc.* **2011**, 137, 553–597. [CrossRef]

35. Davies, H.C. A lateral boundary formulation for multi-level prediction models. *Q. J. Royal Meteorol. Soc.* **1976**, 102, 405–418. [CrossRef]

36. Kotova, L.; Aniskevich, S.; Bobylev, L.; Caluwaerts, S.; De Cruz, L.; De Troch, R.; Gniatik, N.; Gobin, A.; Hamdi, R.; Sakalli, A.; et al. A new project AFTER investigates the impacts of climate change in the Europe-Russia-Turkey region. *Clim. Serv.* **2018**, 12, 64–66. [CrossRef]

37. Top, S.; Kotova, L.; Cruz, L.D.; Aniskevich, S.; Bobylev, L.; Troch, R.D.; Gniatik, N.; Gobin, A.; Hamdi, R.; Kriegsmann, A.; et al. Evaluation of regional climate models ALARO-0 and REMO2015 at 0.22° resolution over the CORDEX Central Asia domain. *Geosci. Model Dev. Discuss. Rev.* **2020**, 1, 1–38. [CrossRef]
38. European Commission. Crop Monitoring in Europe. In Monitoring Agricultural Resources (MARS) Bulletins; Publications Office of the European Union Luxembourg: Luxembourg, 2014; Volume 22, p. 20.

39. Gobin, A. Impact of heat and drought stress on arable crop production in Belgium. *Natl. Hazards Earth Syst. Sci.* **2012**, *12*, 1911–1922. [CrossRef]

40. Semenov, M.A.; Shewry, P.R. Modelling predicts that heat stress, not drought, will increase vulnerability of wheat in Europe. *Sci. Rep.* **2011**, *1*, 1–5. [CrossRef] [PubMed]

41. Gobin, A. Weather related risks in Belgian arable agriculture. *Agric. Syst.* **2018**, *159*, 225–236. [CrossRef]

42. Gusta, L.; Fowler, D. Effect of temperature on dehardening and rehardening of winter cereals. *Can. J. Plant Sci.* **1976**, *56*, 673–678. [CrossRef]

43. Li, X.; Cai, J.; Liu, F.; Zhou, Q.; Li, X.; Cao, W.; Jiang, D. Wheat plants exposed to winter warming are more susceptible to low temperature stress in the spring. *Plant Growth Regul.* **2015**, *77*, 11–19. [CrossRef]