Title

Assessment of weather-yield relations of starchy maize at different scales in Peru to support the NDC implementation

Author names and affiliations

Rahel Laudien[a], Bernhard Schauberger[a], Stephanie Gleixner[a], Christoph Gornott[a]

[a]Potsdam Institute for Climate Impact Research (PIK)
Member of the Leibniz Association
P.O. Box 60 12 03
D-14412 Potsdam
Germany

Corresponding author

Rahel Laudien (laudien@pik-potsdam.de, +49 331 28820771)

Co-authors

Christoph Gornott (gornott@pik-potsdam.de, +49 331 2882655)

Bernhard Schauberger (schauber@pik-potsdam.de, +33 169087724)

Stephanie Gleixner (gleixner@pik-potsdam.de, +49 331 28820818)

Abstract

Climate change poses a substantial risk to agricultural production in Peru. Nationally Determined Contributions (NDCs) are currently developed and outline Peru's mitigation actions and adaptation plans to climate change in various sectors. To support the implementation of adaptation measures in the agricultural sector, information on weather-related risks for crop production and the effectiveness of adaptation options on the local scale are needed. We assess weather influences on starchy maize yields on different scales in Peru based on statistical crop models and a machine learning algorithm. The models explain 91% of yield variability (55% based on the cross-validation) on the regional scale. On the local scale, weather-related yield variation can be explained in some areas, but to a lower extent. Based on these models, we assess the effectiveness of adaptation measures which increase water availability to protect against negative impacts from dry weather conditions. The results show that a higher water availability of 77mm in the growing season would have regionally different effects, ranging from an increase of 20% to a decrease of 17% in maize yields. This large range underlines the importance of a local assessment of adaptation options. With this example, we illustrate how a statistical approach can support a risk-informed selection of adaptation measures at the local scale as suggested in Peru's NDC implementation plan.

Keywords

starchy maize, yield, Peru, NDCs, climate adaptation, statistical crop model

1. Introduction

The Nationally Determined Contributions (NDCs) implementation plan is currently developed in Peru, as a follow-up to the COP21 in Paris in 2015. It provides the basis for Peru's contributions to reducing greenhouse gas emissions and outlines how Peru plans to adapt to climate change impacts in various sectors. Climate change projections show that the agricultural sector in Peru is expected to be confronted with higher temperatures and fewer rainy days but more intense rainfall events (Christensen et al., 2013; Giorgi et al., 2014). In addition to changes in precipitation patterns, the melting of Peru's glaciers (Rabatel et al., 2013) diminishes an important water source for agriculture.

To support the development of the NDCs in Peru, information on the expected impacts of climate change and altered weather conditions on agriculture is required to formulate effective adaptation measures. In addition to information on long-term climatic changes, information on production risks for the near future is needed to provide guidance on how to

---

[1] Abbreviations: NDCs, Nationally Determined Contributions; COP21, 21st session of the Conference of the Parties; RRM, regional regression model; PDM, panel data model, DT, decision tree; ENSO, El Niño–Southern Oscillation; GDD, growing degree days; AIC, Akaike Information Criterion; MSE, mean squared error; NSEe (NSEv), Nash–Sutcliffe efficiency coefficient of the estimation (validation) result; CART, Classification and Regression Tree; SENAMHI, Servicio Nacional de Meteorología e Hidrología del Perú; PISCO, Peruvian Interpolation Data of the SENAMHI's Climatological and Hydrological Observations; JRA-55, Japanese 55-year Reanalysis; MODIS, Moderate Resolution Imaging Spectroradiometer, cloud fraction;

reduce weather risks by short-term adaptation planning and how to deal with already altered climate conditions.

To test if an adaptation measure is appropriate to reduce risks in crop production, process-based and statistical crop models are important tools to resolve the impact of weather on crop yields (Challinor et al., 2014), and the particular effect of adaptation measures under different environmental conditions. A statistical model, which is trained on past weather and yield data, can provide valuable empirical support for informing adaptation measures in the near future (Iglesias et al., 2010; Lobell et al., 2008; Lobell and Burke, 2010; Mu et al., 2013).

As outlined in the NDCs of Peru, the local scale is of particular importance when it comes to developing and implementing adaptation measures as often ways to anticipate and protect against climate change impacts need to be found on the local scale (Gobierno del Perú, 2018). To support an effective selection of appropriate adaptation measures at this level, a model assessment needs to be able to provide information about yield impacts and effects of adaptation at this scale. Apart from a global assessment by Ray et al., (2015) that describes weather-related yield variation on the regional level in Peru, there is – to our best knowledge – no local or regional assessment of the complex weather influences on crop production covering the whole country of Peru. Moreover, there is no assessment of suitable adaptation measures informed by risk assessments of current or future weather conditions.

In this study, we assess the influence of weather on starchy maize yields. Together with potato (*Solanum tuberosum L.*), starchy maize (*Zea mays L ssp. amiláceo*) belongs to the main food crops in Peru. It is mostly produced for self-consumption, which makes its production particularly important for the food security of 42% of the rural population who live in poverty in Peru (INEI, 2019).

We assess weather influences on starchy maize yields on different scales in Peru with a regional regression model, a panel data model and the machine learning algorithm decision tree. Based on these models, we evaluate the effectiveness of adaption measures which increase water availability (e.g. irrigation, water storage and harvesting or soil management, as named in Peru's NDCs) as action options under unfavourable weather conditions. With this example, we illustrate how a statistical approach can support a risk-informed selection of adaptation measures suggested in Peru's NDCs. This can support the implementation process of the NDCs and the design of effective adaptation measures for current and future periods at the local scale.

Weather variability and climate change in Peru

Already today, Peru is affected by high weather variability – both spatially and temporally. Peru has 15 Köppen-Geiger climate zones (Fig. A.1 of the supplementary information, SI), which range from hot and cold desert climate at the coast to subtropical highland climate in the Andes and tropical rainforest climate in the Amazon (Peel et al., 2007). This spatial climatic diversity results in different growing conditions across the country. Even within the same climate zones, the inter-annual weather variability is high because of periodically recurring El Niño–Southern Oscillation (ENSO) events. El Niño particularly impacts the coastal region of the Andes in Peru (Bourrel et al., 2015). The recent Coastal El Niño event in 2016/17 led to severe floods and landslides along the northern and central coastal region (Rodríguez-Morata et al., 2019). The El Niño in 2015/16 was associated with exceptionally dry conditions in the south of Peru. Agriculture in Peru needs to be able to adapt to these varying weather conditions for which early warnings are still difficult to obtain, despite progress in the prediction lead time (Ludescher et al., 2014).

In addition to these already prevalent weather risks on agriculture, climate change will amplify risks for Peru's agricultural sector. In the last four decades, the strongest warming within Peru occurred in the southern Andes of Peru (warming rates reached up to 0.3°C per decade since 1981; SI Fig. A.2). This warming trend is projected to continue to an additional increase of 1° to 2°C until mid-century depending on the emission scenario (Collins et al., 2013; Marengo et al., 2009). Apart from the direct impact of increased temperatures on agriculture, higher temperatures also increase the atmospheric water demand and consequently evapotranspiration rates (Allen et al., 1998). This in turn reduces the amount of water available to agriculture (Lobell et al., 2013).

With regards to future precipitation projections, a general tendency towards fewer rainy days, but more intense rainfall events, is expected (Christensen et al., 2013; Giorgi et al., 2014). Whereas the northern Coast and the Amazon region are expected to experience fewer consecutive dry days throughout the 21st century, the number of dry days and consequently dry spells are projected to increase over southern Peru (Giorgi et al., 2014; Sörensson et al., 2010) where the main starchy maize producing regions can be found.

Many glaciers have already lost up to 30% of their surface area since the 1960s (Rabatel et al., 2013) such that an important water source for agriculture is diminishing. Melting glaciers, dryer conditions in southern Peru and more erratic rainfall suggest that adaptation options (e.g. irrigation, water storage and harvesting or soil management) that address water scarcity and precipitation vagaries will be important to maintain and increase agricultural productivity in Peru.

2. Material and methods

In this study, we used two yield data sources and three different statistical models to assess the influence of weather on starchy maize yields in Peru. Based on the models, we evaluated the effectiveness of artificially increased water availability to increase yields.

2.1 Input data

2.1.1 Yield data

The first source of crop yield data is obtained from official statistics from the Ministry of Agriculture of Peru (MINAGRI, 2018). The time series covers 13 years (from 2005 to 2017) and is available on region level. The dataset comprises 19 out of the 25 regions in Peru.

The second yield data source is a household survey on plot level carried out by the Peruvian national statistical institute "Instituto Nacional de Estadística e Informática" (INEI, 2017). The first survey wave started in 2014. However, the size assessment of the agricultural plots changed in 2016 (source: personal communication with INEI), so that the first two and the last two waves are not comparable. Therefore we only considered the last two waves (2016 and 2017) of the survey to have a consistent measurement of the plot sizes in the data set. This is important as yield is calculated as production over area. The data set of waves 2016 and 2017 comprises 19,704 yield observations in total (SI Fig. A.3). We aggregated the plot level data to cluster level. The clusters were derived from the IV national census from 2012 and are the primary sampling unit of the survey. Clusters are below districts, i.e. the lowest administrative unit in Peru (SI Fig. A.4 shows a map of regions and clusters in Peru). The aggregation to cluster level was necessary to limit a potential sampling bias or reporting error on plot level and because the GPS locations of the plots are not publically available, thus precluding a correlation with weather data. We averaged the yields per cluster with weights based on harvest area.

2.1.2 Weather data

We used the following climate data for our analysis. For precipitation, we used a data set provided by the Peruvian meteorological service SENAMHI (Servicio Nacional de Meteorología e Hidrología del Perú) called PISCO (the Peruvian Interpolation Data of the SENAMHI's Climatological and Hydrological Observations, Aybar et al., 2019). The data set provides daily precipitation sums at a spatial resolution of 0.1° and has a robust performance in particular in the coastal regions and the western flank of the Andes (Aybar et al., 2019), which are particularly used for maize production. While PISCO also includes temperature data, this was not available for the whole study period. Therefore, we used surface temperature data from JRA-55 (Japanese 55-year Reanalysis; Kobayashi et al., 2015). JRA-55 provides 6-hourly temperature at a resolution of ca. 0.6°. To represent the influence of solar radiation on plant growth, we used cloud fraction derived from MODIS (Moderate Resolution Imaging Spectroradiometer; Platnick et al., 2015) due to its high spatial and temporal resolution for a long time series. MODIS provides daily raster imageries at a resolution of 5km. We used cloud fraction instead of solar radiation due to reliability issues in the data quality of solar radiation over the Amazon (Dutra et al., 2015) and because cloud fraction can be regarded as tantamount to solar radiation (Muneer and Gul, 2000). Even though the combination of different climate data sources bears the risk of incorporating physical inconsistencies, the different sources were considered the best available choices for Peru in particular in regard of spatial resolution, which is crucial in complex terrain. Moreover, physical correlations are less relevant when aggregating climate data temporally and spatially to administrative levels. For the regional assessment, we calculated the mean over all grid points per region. For the local assessment, we aggregated the weather data to cluster scale. Due to the small size of the clusters, we used the weather grid point that has the shortest distance to the cluster center calculated by the Euclidean distance.

2.1.3 Crop calendar

To represent the weather conditions during the growing season, we followed the crop calendar provided by MINAGRI (MINAGRI, 2017). Even though the survey also contains information about the growing season, we preferred the MINAGRI crop calendar to allow for a better comparison between the two yield data sources and to avoid possible errors in the survey-based seasons (section 2.2.1). Per region the most frequent sowing and harvesting month were selected. If the previous (next) sowing month (harvesting month) had a share higher than 25%, it was also included in the growing season. For the regions Lambayeque and Ica no clear growing season could be found, because several months have a similarly high share. Therefore, we used the dates of the provinces within these regions with the highest production of starchy maize.

2.2 Modelling approaches

We used three modelling approaches (Fig. 1). To assess weather-yield relations on the regional level, we used a regional regression model (RRM). The data set on region level covering 2005 to 2017 was used as input yield data. For each of the 19 regions, a RRM (Eq. 1) with different variables and parameters was constructed to account for the diverse climatic conditions within the country.

Due to the short time series of the household survey (covering the harvesting years 2015 to 2017), we used a panel data model (PDM) to analyse weather influences on maize yields at the local level. The PDM uses one parameter set for all the considered spatial units (Eq. 2). Both the RRM and PDM follow the approach of Gornott and Wechsung (2016).

$$y_{it} = \sum_{k=1}^{K} \beta_{ki} X_{kit} + \alpha_i + \varepsilon_{it} \quad (1)$$

$$y_{it} = \sum_{k=1}^{K} \beta_{k} X_{kit} + \alpha_i + \varepsilon_{it} \quad (2)$$

with β as parameters, X as explanatory input variable, α as unobserved time-invariant individual effects, ε as error term for K variables ($k = 1, \ldots, K$), N spatial units ($i = 1, \ldots, N$) and T years ($t = 1, \ldots, T$)

In addition to these two regression models, we used a decision tree (DT) to cross-check the results obtained from the PDM. The PDM is based on fewer input data, which bears the risk of generating less robust results. We therefore apply a machine learning algorithm as an additional validation. The DT is a non-parametric machine learning algorithm which does not require distributional assumptions and is robust to outliers (Song and Lu, 2015). For our analysis, we used the CART (Classification and Regression Tree) algorithm (Breiman et al., 1984). To avoid overfitting, we pruned the tree by defining the minimum amount of observations required to perform a split to 50 and the minimum amount of observations for an end node to 25. This is referred to as pre-pruning and stops the tree from growing completely until it perfectly classifies the training set (Patel and Upadhyay, 2012).

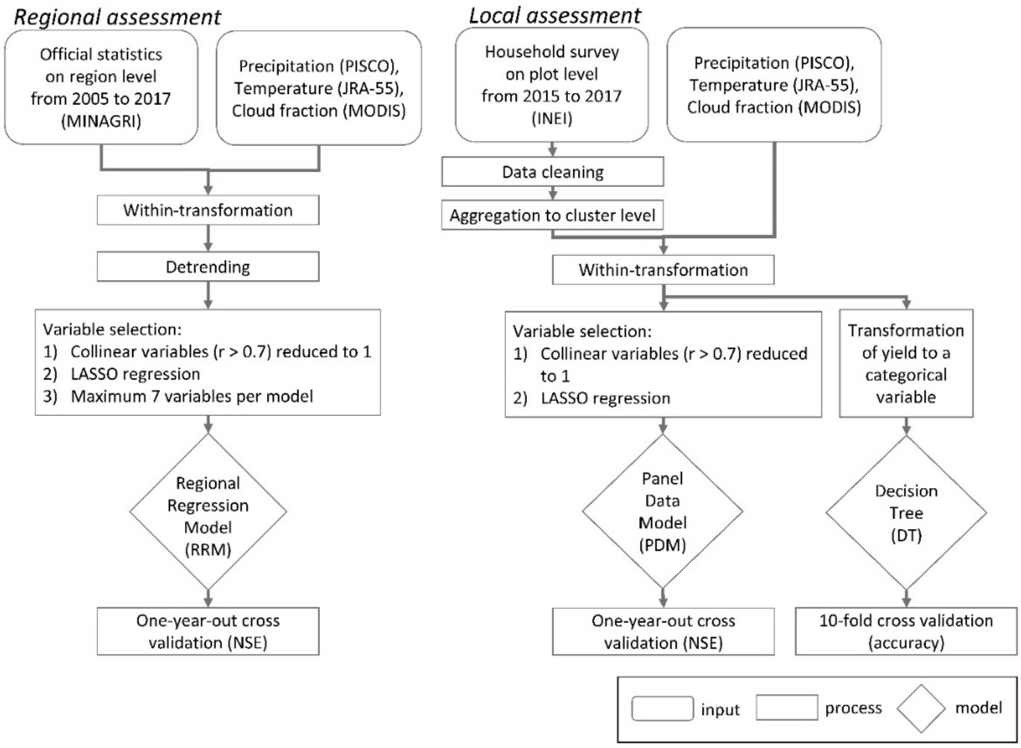

Fig. 1: Work flow of the analysis for regional (left) and local assessment (right)

For our analysis, we used R (R Core Team, 2014) with the package glmnet (Friedman et al., 2008) to perform LASSO regression, the package rpart (Therneau et al., 2019) to perform the decision tree and the package ggplot2 (Wickham, 2009) to produce the figures.

2.2.1 Pre-processing

The household survey contained obvious inconsistencies that required data cleaning. Some observations in the survey have higher harvest areas than planted areas or unrealistically high production on small plots (e.g. > 300kg production on < 0.03ha harvest area). The sowing and harvesting dates show questionable entries for many observations, i.e. either strong deviations from the median dates per region or a growing season longer than a year. These inconsistencies could originate from differing techniques applied by enumerators or wrong statements made by the interviewees. Different units used for production (kg, arroba, bag, quintal and other units), and plot sizes (m^2, ha, acre and other units) make the data set prone to errors and inaccuracies, because the conversion factors used to standardize production and plot size might be inaccurate. To ensure a proper data base, we deemed it essential to clean the data before modelling. First, we removed observations with missing values for yield. Second, we omitted yields outside the 1st and 99th percentiles to guard against outliers, leaving only yields between 0.24 and 3.75 t/ha. Third, we only kept clusters that have yield data for at least three harvesting years (the survey in 2016 also asked for harvests in 2015, with equal plot size calculations as in 2016 and 2017). Fourth, we removed those clusters that show very little yield variability (standard deviation < 0.01 percentile), as constant yields are unlikely and probably reporting errors. In sum, after data cleaning (remaining n=19,253), aggregating to cluster level (n=1,994), only considering clusters with three harvesting years (n=324) and sufficient variability, we used 291 aggregated yield observations (SI Fig. A.3).

In addition, we applied the following transformations: For all three models, we applied a within-transformation to remove time-constant effects from the variables (Wooldridge, 2014). For the RRM that is based on a time series of 13 years, we removed the trend in the yield data by testing different de-trending methods (none, linear, quadratic) and then applying the one that resulted in the lowest Akaike Information Criterion (AIC; Bozdogan, 1987). For the DT, we transformed quantitative maize yields into a categorical variable by splitting it into five groups (using the 0th, 20th, 40th, 60th, 80th, 100th percentiles as group limits).

2.2.2 Variable creation and selection

Following the climatic envelope for maize growth (SI Text A.1), we generated the following variables (formulas are provided in SI Text A.2): To represent temperature conditions, we included mean temperature (T.mean), the mean and maximum of daily maximum temperatures (T.max and maxOfmax, respectively) and the mean and minimum of daily minimum temperatures (T.min and minOfmin, respectively). The two extreme values were included to account for single high and low temperature events that could already induce crop damages. To represent cumulative extreme temperature influences, we included Freezing Degree Days (FDD) to account for the harmful influence of cold temperatures below 0°C and Heat Degree Days above 30°C (HDD) to cover too high temperatures. Variations in temperature within the season are represented by the coefficient of variation (CV, defined as standard deviation over mean). We calculated the CV of T.mean, T.max and T.min.

We included the precipitation sum in the growing season (P.sum) to represent the overall water availability for maize. We also included the number of days with precipitation above a threshold of 5 (PA5), 10 (PA10), 15 (PA15) and 20mm (PA20) per day as well as the maximum daily precipitation (P.max) to cover the influence of single precipitation events. The number of days without precipitation (DWP) accounts for dry conditions. Because the distribution of rainfall within the growing season is of particular importance for plant growth, we included dry spells of more than 5 (cdd5) and 10 consecutive dry days (cdd10) and wet spells of more than 5 (cwd5) and 10 consecutive wet days (cwd10). The coefficient of variation of the precipitation sum (P.cv) represents the variability of rainfall within the growing season.

To cover the influence of solar radiation, we included the mean cloud fraction (C.mean), the minimum cloud fraction (C.min) and the coefficient of variation of the mean cloud fraction (C.cv).

The variables were separately calculated for the vegetative and reproductive phase of the growing season. The separation between vegetative and reproductive phase was based on the sum of growing degree days (GDD; SI Text A.2 Eq. 22). The days in the growing season until 50% of the full-season GDD sum was reached were allocated to the vegetative phase and the remainder to the reproductive phase, following Schauberger et al. (2017).

Since there is a large number of potentially relevant variables, a selection process was applied for the RRM and the PDM to elucidate important influences. 1) To avoid multicollinearity, only those variables were selected that are not strongly collinear (i.e. Pearson's $r > 0.7$) with another explanatory variable. For the RRM: if a set of variables was strongly collinear, then from this set only one variable was included, choosing the one with the highest correlation with yield. For the PDM: variables were selected for which more clusters have a high correlation with yield. 2) Least Absolute Shrinkage and Selection Operator (LASSO) regression was performed for the final feature selection. Through regularization, LASSO performs a co-variate selection, which improves both the prediction accuracy and the interpretability (Tibshirani, 1996). For the RRM: to select the optimum lambda (the regularization penalty for the LASSO regression), we used the lowest cross-validation (years were omitted subsequently) mean squared error (MSE). For the PDM: because of a higher amount of input data, we used the lowest cross-validation MSE of a 10-fold cross-validation to select the optimal lambda. Because in this case the folds were selected across time and space, we used the mean minimum lambda value of 30 model runs. 3) As an extra safeguard against overfitting in the RRM, we included one further restriction, allowing only half as many variables as there are observations. For the DT, the described variable selection process was not applied as the algorithm explicitly selects the variables that achieve the highest information gain and pre-pruning avoids overfitting.

2.2.3 Validation

To validate the RRM and the PDM results, we performed a one-year-out cross-validation. For each year subsequently, all observations in that year are removed from the dataset and the remaining observations from the other years are used to estimate the model and predict yield changes for the removed year. The goodness of fit was then evaluated based on the Nash–Sutcliffe efficiency coefficient (NSE; Nash and Sutcliffe, 1970), for the combined out-of-sample predictions. In contrast to R^2, NSE does not only evaluate similarities in variability, but also the mean model bias, which makes it a robust quality measure. In addition, we tested the significance of the models based on the F-statistics and the performance of the models compared to a constant model that only takes the mean yield per region as a predictor. With the Breusch–Godfrey test we assessed autocorrelation and with the Breusch–Pagan test we tested for heteroscedasticity. As highly co-linear variables were removed in the variable selection process, a test for multicollinearity was not necessary.

The DT is validated based on a 10-fold cross-validation and the model fit is assed based on accuracy, which is defined as the share of correctly predicted values of all predicted values.

We compare the model fit of the different modelling approaches on region level. Therefore, the observed and modelled yields of the PDM and the DT were aggregated to region level.

2.3 Assessment of adaptation options

We assessed the effect of increased water availability on crop yields. To get an estimate of how much more water is going to be needed in future with ongoing climate change, we focused in our analysis on changes in the atmospheric water demand. A warming climate increases the atmospheric water demand (Allen et al., 1998). This in turn reduces the amount of water available to agriculture (Konings et al., 2017; Konings and Gentine, 2017; Lobell et al., 2013; Novick et al., 2016). The actual water availability depends also on other factors, most importantly on precipitation and soil water holding capacity. However, as climate models do not agree on a sign of the change in precipitation in the future in Peru (SI Fig. A.5), we concentrated on the change in atmospheric moisture demand.

For an estimate of future changes in the atmospheric water demand, we focused on potential evaporation in the ISIMIP2b simulations from the global hydrological model WaterGAP2 (Müller Schmied et al., 2016). The model was driven with climate data from four global climate models (namely GFDL-ESM2M, HadGEM2-ES, IPSL-CM5A-LR, MIROC5), which were bias-adjusted and provided within ISIMIP2b (Warszawski et al., 2014). We focused on the RCP8.5 scenario, which is the RCP emission scenarios without climate policy interventions, to take into account the range from conditions similar to today to unmitigated climate change. The spatial variation of current potential evapotranspiration over Peru and expected changes by mid-century are shown in Fig. 2. The ensemble mean change of potential evaporation under RCP8.5 in Peru is an increase of ca. 77mm for the growing season in the middle of this century. This amount of water is estimated to feed the atmospheric water demand and as such is not available to crops. Therefore, we assume this amount to be the minimum necessary additional water in the near future to compensate for the negative temperature effects on water availability. We used this estimated change in potential evapotranspiration and assessed the effect of 77mm more water per growing season on starchy maize yields.

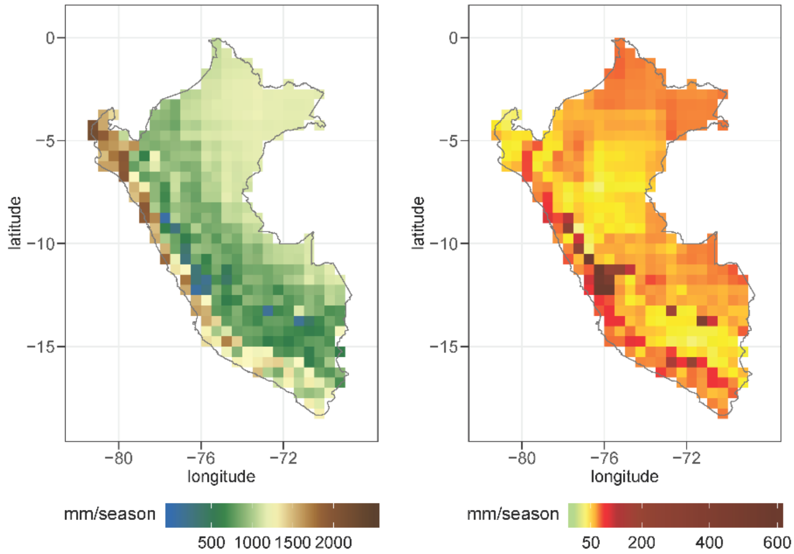

Fig. 2. Mean potential evapotranspiration from 2000 to 2020 (left) and the change in mean evapotranspiration in the period from 2040 to 2060 compared to the period from 2000 to 2020 (right), both figures show modelled data for consistency

To distribute the additional water evenly over the growing season, we assumed that the farmer would irrigate twice per month (on day 1 and 15 of each month) by an amount of 77mm divided by the number of months times two. As most growing seasons are around nine months, this means ca. 9mm of irrigation per month. We compared the original model output (the estimation with currently available water) with the adaptation scenario outputs and evaluated the changes in simulated maize yields.

3. Results

3.1 Starchy maize yields in Peru

In Peru, 273,868 tons of starchy maize were produced in 2017. The main producing regions were Cusco, Apurimac, Huancavelica, La Libertad, Ayacucho and Cajamarca (Fig. 3). These six regions accounted for 70% of the total national production in 2017.

Between 2005 and 2017, yields for starchy maize in Peru were 1.7t/ha on average. In 2017, the highest yields could be found in the southern coastal regions in Peru (Ica, Arequipa and Tacna) with the highest yields in Ica (4.5t/ha). The lowest yields were obtained in Piura, Amazonas and Cajamarca (0.8 to 1t/ha). Apart from Cusco, which had relatively high yields (2.5t/ha), most regions with high total production only showed medium yield levels (Apurimac = 1.9t/ha, Huancavelica = 1.6t/ha, La Libertad = 1.6t/ha, Ayacucho = 1.3t/ha). An exception is Cajamarca, which had the lowest yield level (0.8t/ha) compared to the other regions even though it was one of the most producing regions. Accordingly the harvest area was one of the highest compared to the other regions, whereas the high-yield regions had lower harvest areas (Ica, Lima, Tacna and Moquegua; MINAGRI, 2018). The vast majority of starchy maize production was based on smallholder agriculture with an average harvest areas of 0.13ha. Even in Piura and Lambayeque, which were the regions with the highest harvest area for starchy maize in Peru, the harvest areas was only 0.28ha on average (INEI, 2017).

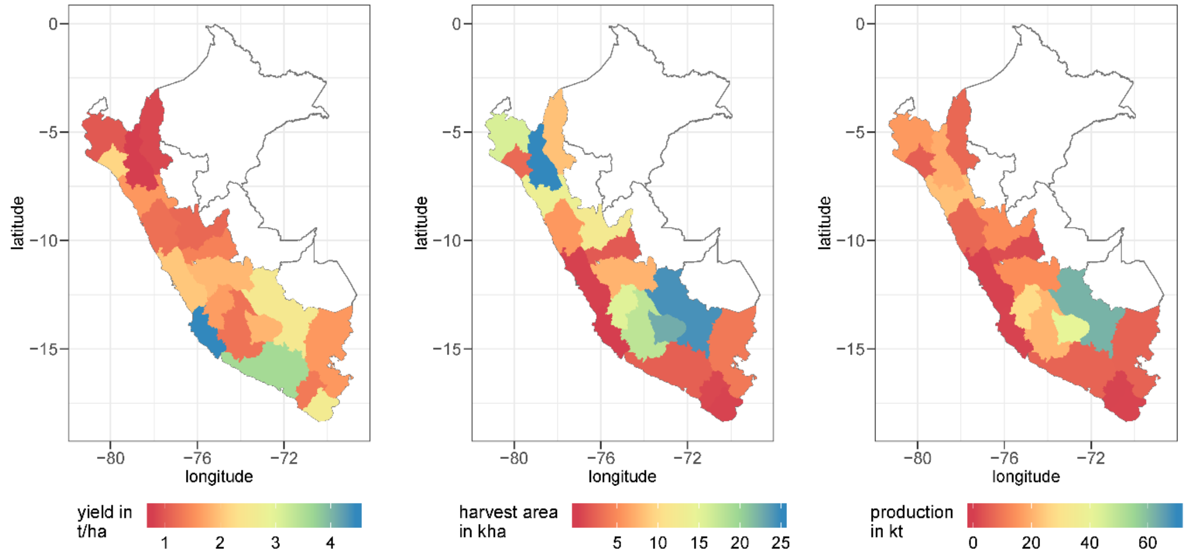

Fig. 3. Yield, harvest area and production of starchy maize in Peru in 2017

As there is no spatially explicit information in the data set when maize is planted or harvested, we chose the entire official growing period (Fig. 4) as reported by the Ministry of Agriculture (MINAGRI, 2017). The length of the growing season ranged from 204 (5th Percentile) to 306 (95th Percentile) days and the main growing season is from September to June. Exemptions were Piura and Lima with relatively late sowing in January/February and late harvest in August. Lambayeque and Ica have a high amount of larger farms with irrigation facilities available, which is why the growing season in these two regions is less dependent on the start of the rainy season. In these two regions, starchy maize is produced almost all year long and a clear growing season cannot be found. The sowing and harvesting months in these two regions do not represent most of farmers, but represent the month with the highest share of farmers who planted/harvested.

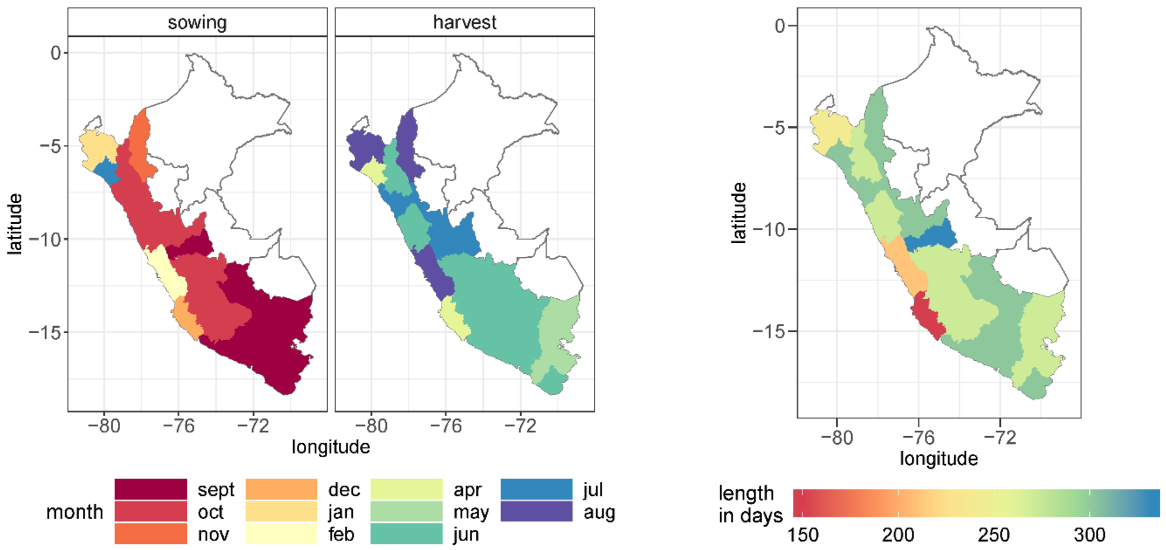

Fig. 4. Sowing and harvesting month (left) and the length of the growing season (right) for starchy maize yields on region level in Peru

3.2 The influence of weather on starchy maize yields

3.2.1 Regional assessment

We used a RRM to assess weather influences on yield variability from 2005 to 2017 on region level. The model shows a good performance for almost the whole country (Fig. 6). The median NSE for the estimation with all data (coined ‘NSEe’ hereinafter) is 0.91, which corresponds to an explained variability of 91%. The exhaustive one-year-out cross-validation produces a median NSE (coined ‘NSEv’ for ‘validation’) of 0.55 (i.e. an explained variability of 55%). The model is also able to reproduce most of the extreme years as can be seen with the high harvests in Amazonas in 2014 and in Lambayeque in 2011 or the low harvests in 2016 in Huánuco and in Tacna (Fig. 5). The yield variability in Lima cannot be reproduced by the model (NSE of -0.75; SI Fig. A.6), which can be explained by the diverging agricultural conditions and the high degree of urbanization in this region. Also the models for Apurimac (NSE of -0.24) and Huancavelica (NSE of -0.45) show an overall weak performance, even if some years can be reproduced well by the models.

We tested whether the assumptions in linear regression are fulfilled. Whereas heteroscedasticity does not occur in any of the RRMs, a few models show autocorrelation, which is why we used robust standard errors in these cases (Zeileis, 2004). The significance of the models on the 0.05 significance level and the lower RMSE of the RRMs compared to a constant model (SI Table A.1) suggest that the model results are useful (apart from the models for Lima, Apurimac and Huancavelica) and that weather explains a substantial share of observed maize yield variation.

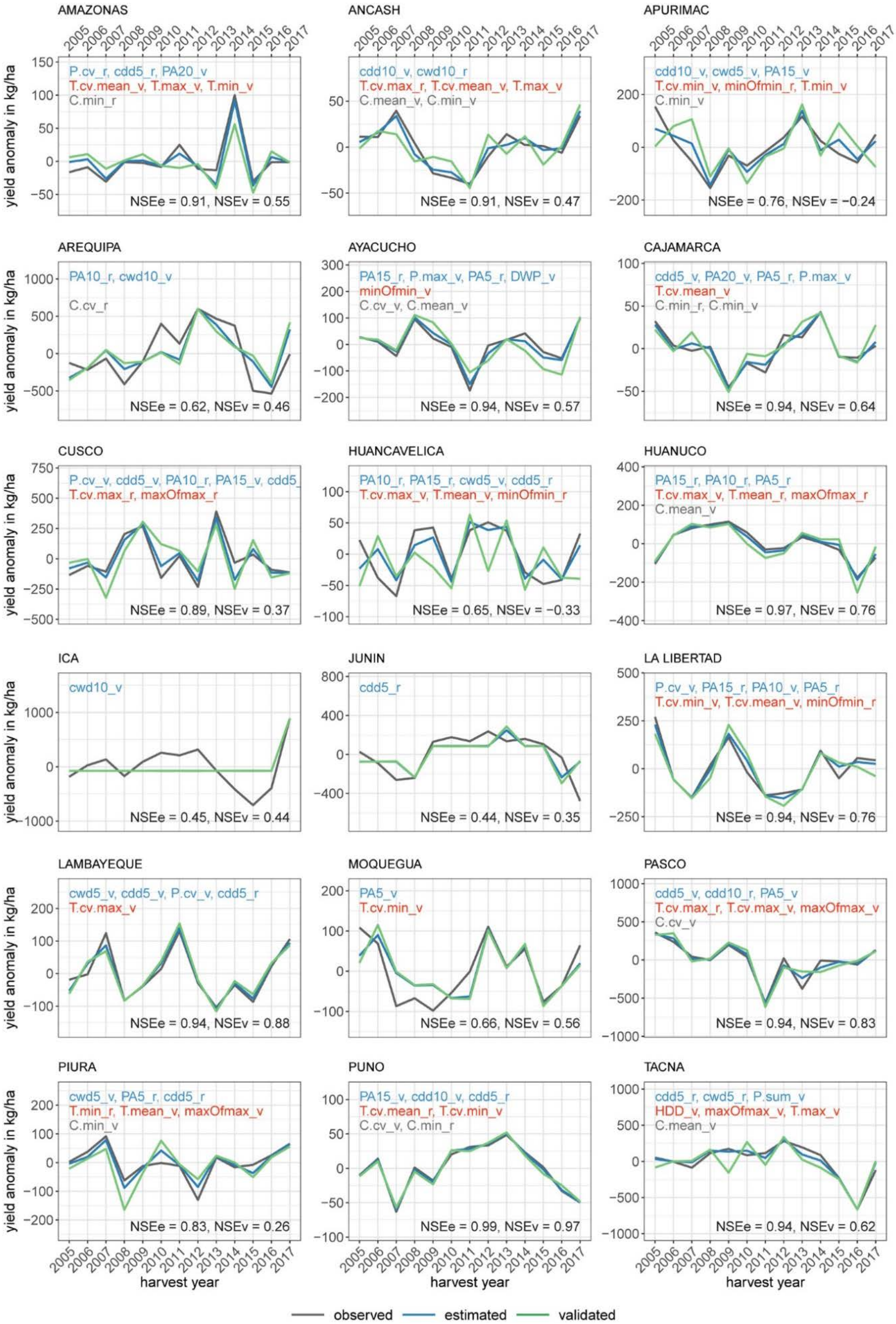

Fig. 5. Observed and simulated maize yield anomalies in Peruvian regions. Black lines show observed yield anomalies, blue lines show anomalies estimated with the full model and green lines those estimated out-of-sample. Region-specific selected variables are shown in blue for precipitation, red for

temperature and grey for cloud fraction. The abbreviations of the variable names are explained in Table 1. Lima is not shown here for space reasons and because it is not used for further applications, but it can be seen in SI Fig. A.6. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The variable selection reflects the high climatic variability within Peru. Precipitation seems to determine yields most evidenced by the most frequent selection of precipitation-based variables when compared to others. The variable cdd5 is selected most often. Strong variations in both maximum and mean temperatures influence maize yields in most regions (mostly negatively). Selected variables related to cloud fraction are most often minimum cloud fraction and the coefficient is usually positive, which may be due to a higher cloud fraction often being synonymic with more precipitation. A comparative description of the variable selection in different regions in Peru can be found in SI Text A.3 and SI Table A.1 provides the region-specific coefficients of selected variables.

Precipitation-related variables			Temperature-related variables			Cloud-related variables		
Name	Explanation	Count	Name	Explanation	Count	Name	Explanation	Count
cdd5 (cdd10)	Consecutive dry days of more than 5 (10) days	11 (3)	*T.cv.max*	Coefficient of variation of maximum temperature	6	*C.min*	Mean minumum cloud fraction [%]	6
PA5, PA10, PA15, PA20	Precipitation events above 5, 10, 15 or 20[mm]	7, 4, 5,2	*T.cv.mean*	Coefficient of variation of mean temperature	5	*C.cv*	Coefficient of variation of mean cloud fraction	4
P.cv	Coefficient of variation of daily precipitation sum	4	*maxOfmax*	Maximum of the maximum temperature [°C]	5	*C.mean*	Mean cloud fraction [%]	4
cwd5 (cwd10)	Consecutive wet days of more than 5 (10) days	3 (3)	*T.cv.min*	Coefficient of variation of minimum temperature	3			
P.max	Maximum precipitation [mm]	2	*T.max*	Mean maximum temperature [°C]	3			
DWP	Days without precipitation	1	*T.min*	Mean minimum temperature [°C]	2			
P.sum	Precipitation sum [mm]	1	*T.mean*	Mean temperature [°C]	2			
			minOfmin	Minimum of the minimum temperature [°C]	2			
			HDD	Heat degree days	1			

Table 1. Abbreviations of selected variables. The column “count” denotes the frequency of selected variables over all RRMs (excluding the model for Lima). The variables are generated for the vegetative (indicated by the suffix “_v”) and the reproductive phase (indicated by “_r”) of the growing season

3.2.2 Local assessment

As stated by the Peruvian NDCs, the local scale is of particular importance in adaptation planning (Gobierno del Perú, 2018). Therefore, we also analysed the weather influence on starchy maize yields on the local level. For this assessment, we used a PDM that takes a survey covering the harvesting years 2015 to 2017 as input data. Over all observations (n=291), the model has an NSE of 0.20 in the estimation. Autocorrelation and heteroscedasticity do not impair the model. The non-exhaustive one-year-out cross-validation produces an NSE of 0.11, which is a lower model performance compared to the RRM. However, in 33% of the clusters the NSE in the validation is higher than 0.25 and the model is significant and has a higher performance compared to a constant model, indicating a detectable impact of weather on crop yields beyond random influence (SI Table A.2). The following descriptions only relate to these clusters (Fig. 6).

The variable selection process revealed an influence of 16 weather variables on starchy maize yields for the harvesting years 2015 to 2017 and the considered observations. Like in the RRM, precipitation seems to have a stronger influence on maize yields than temperature and cloud fraction. Overall, the model suggests a positive influence of moderately wet conditions on starchy maize yields. Consecutive dry spells of more than 5 and 10 days influenced yields negatively in the reproductive phase, whereas consecutive wet days of more than 5 days showed a positive impact in the whole growing season. Maximum precipitation showed a positive impact, but precipitation above 20mm had a negative impact. Mean cloud fraction, which is often related with rainy weather conditions, showed a positive impact. The negative influence of variations in cloud fraction can be explained by the deviation from high cloud fractions, which are related with dryer weather conditions. Temperature-related influences were most pronounced for variations in temperature (mean, maximum and minimum). Heat degree days and minimum temperature proved to be detrimental for starchy maize yields. For the considered observations, the model suggests that too dry conditions had a stronger negative impact on yields than too wet weather conditions.

To compare the results with the RRM, we aggregated the results to region level. Whereas the RRM is able to explain yield variability in almost all regions in Peru, the PDM captures yield variability in nine out of 16 regions (Moquegua, Huancavelica, Lambayeque, Cajamarca, Ayacucho, Junín, Tacna, Arequipa and La Libertad), evidenced by an NSEv higher than 0.25. The model seems to be particularly suitable to capture the weather influences on starchy maize yields in the cold and hot desert climate that is found in the coastal region of Peru. Wet conditions, on the other hand, like those in the tropical wet and dry or savanna climate, the tropical rainforest and monsoon climate like in Ancash, Amazonas and Pasco are not well captured by the model. The regional differences in the model performances indicate that the PDM is better in representing dry weather influences on maize yields than too wet weather conditions.

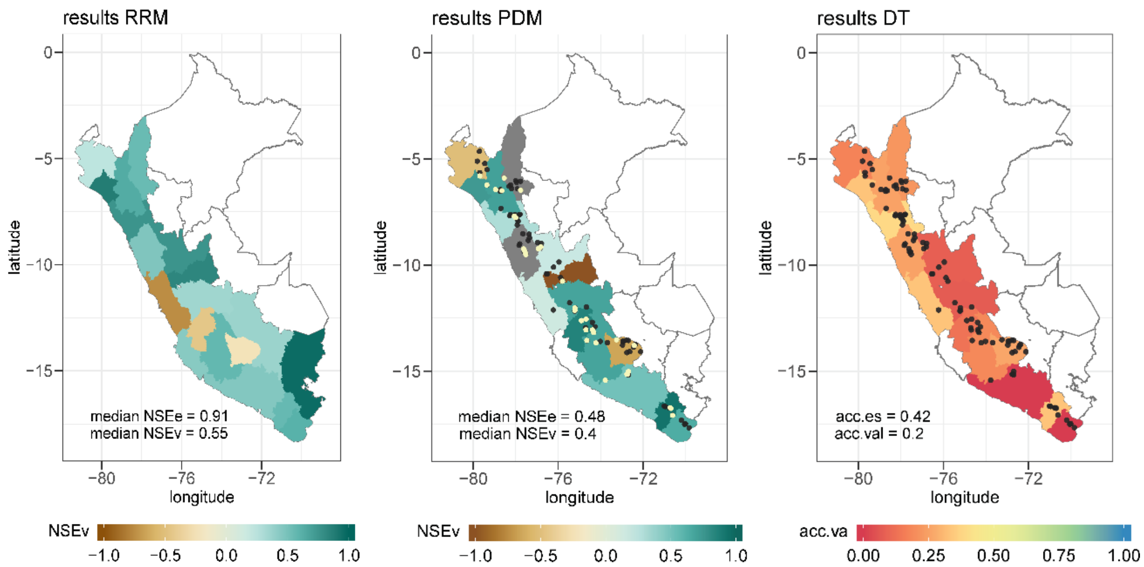

Fig. 6. Model performance in the out-of-sample validation for the RRM (left), the PDM (middle) and the DT (right). Points show the location of the considered cluster centroids (points in yellow in the PDM show clusters with a higher model performance than an NSEv of 0.25, black points show all other clusters). Grey colour in the PDM panel indicates NSEv values < -1. Regions in white have no observations in our data base. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Because the PDM on the local level can explain yield variability for fewer regions compared to the RRM, we used another modelling approach to corroborate our results. We therefore applied a decision tree to the same input data as for the PDM (Fig. 7). The accuracy of the DT is 42% in the training set and 20% in the test set (Fig. 6).

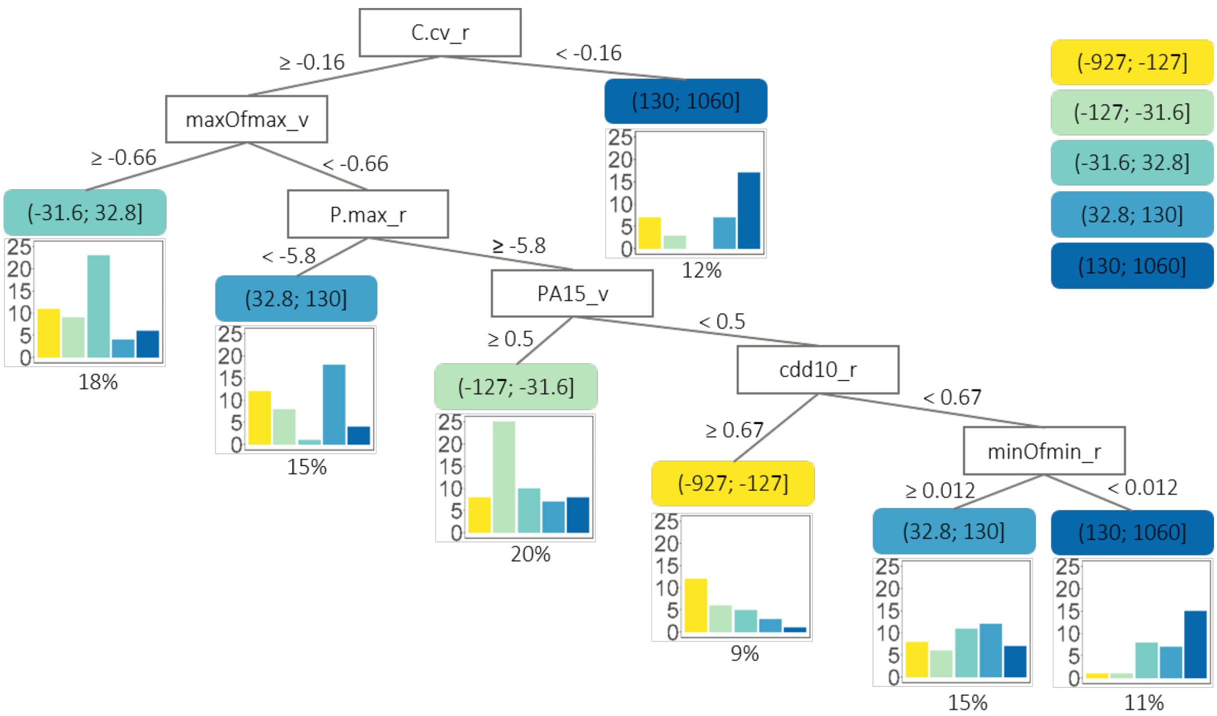

Fig. 7. A decision tree for starchy maize yield anomalies on the local scale in Peru. Yield anomalies in kg/ha are split into five categories (each comprising ca. 58 observations, with 291 observations in

total), denoted by different colours. For each end node, the majority category is indicated by colour, while the number of observations in each category are displayed in the histogram below the nodes. Percentages show the fractions of all observations contained in the node. Variable explanations are provided in Table 1.

In both the PDM and the DT, the variables C.cv, P.max, cdd10 and minOfmin were selected. In the DT, the first split variable on the top node was cloud fraction, which shows that major yield differences can be explained by deviations from high cloud fractions that are usually related with dryer weather conditions. 12% of observations had high yield levels when cloud variation is smaller. Whereas 18% of the observations showed a positive relation with high maximum temperatures, in most cases (the remaining 70%) there was a negative impact of maximum temperatures. The influence of precipitation depends on the yield range and suggests non-linear influences. Whereas low maximum precipitation was related to higher yields in 15% of the observations and precipitation above 15mm with lower yields in 20% of the observations, more dry spells of 10 days were connected with low yields for 9% of observations and with higher yields for the remaining 26% of the observations.

3.3 Assessing an adaptation option

Due to the high performance of the RRM for large parts of the country, we used the RRM to assess the effect of higher water availability on starchy maize yields. This is relevant to gauge expectations on adaptation measures that aim to compensate the water loss due to increased potential evapotranspiration under climate change in the middle of this century (section 2).

Our analysis shows that 77mm more water availability in the growing season would have regionally different effects (Fig. 8). The model suggests an increase of 17% (i.e. 160kg/ha) in Piura and an increase of 15% (i.e. 405kg/ha) in Tacna. In contrast, the model result for Lambayeque shows a drop in yields by 21% (i.e. 380kg/ha). Even though Lambayeque is a neighbouring region of Tacna, yields are influenced more strongly by the negative influence of excessive rain, such that a higher water availability in the growing season would have a negative influence on yields. Half of the regions show only a slight change (-2% to +2%), and partly an insignificant change in yields (SI Table A.3).

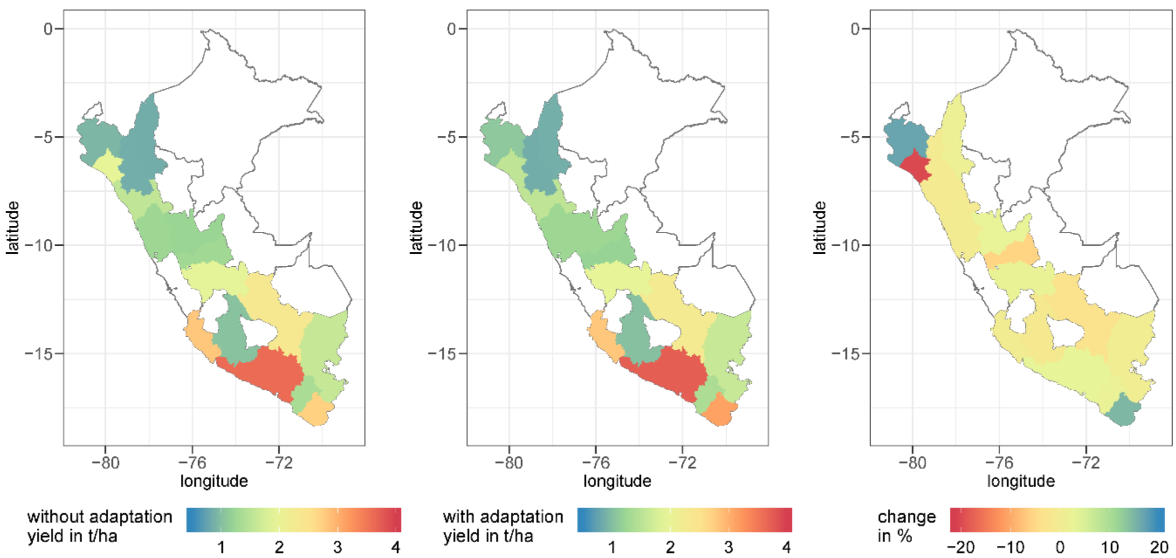

Fig. 8. The effect of 77mm more available water during the growing season on starchy maize yields

4. Discussion

We have shown that three different statistical crop models can explain a substantial fraction of starchy maize yield variation in Peru by variations in weather. Model performances diverge, with a regional regression model showing the highest skill in yield estimation while local-scale panel data and decision tree models account for less of the observed variation. The effect of more plant-available water – to counterbalance possibly higher atmospheric demands under climate change – was assessed, resulting in mostly limited, regionally distinct yield effects.

Whereas the RRMs on the regional scale showed a high performance, the PDM on the local scale could capture yield variability only to a limited extent. There are also divergences in the importance of climatic variables between the two scales. We cannot exactly decipher the underlying reasons for the difference in the model performances and the variable selections due to different input data and different modelling approaches used. Differences between the RRM and the PDM could be related to issues in data quality, the modelling approach or a lacking influence of weather on crop yields at the local scale caused by higher importance of management factors.

The yield input data for the local assessment is a survey and we identified inconsistencies in the data set (detailed in section 2.2.1). Despite the applied filters, some uncertainties remain such as those stemming from the normalization of different units for production and plot sizes or the high inner-regional differences in the growing season length. The PDM performs better on an aggregated (region) level, which also indicates that the survey might contain reporting errors that are averaged by the spatial aggregation and that other influences on crop yields may also play an important role at local level.

Due to the short time series of the survey data, we had to apply a different modelling approach for the local assessment, i.e. a pooled PDM instead of an RRM. The PDM uses one parameter set to explain weather-yield relations for entire Peru. Because of the highly diverse climatic conditions within Peru, one parameter set may only be able to capture the complex yield-weather relations in Peru to a limited extent.

Moreover, existing adaptation efforts, e.g. irrigation possibilities, or individual farmers' decision on sowing dates or cultivar choices could contribute to a low model performance in some regions. The high inter-regional differences in weather-responses underline the need for a spatially distinct assessment of adaptation options to climate change.

In addition to regression analysis (RRM and PDM), we used a machine learning algorithm. The regression models conform to the assumptions of linear regression, apart from a few RRMs that showed autocorrelation. In this case, we used robust standard errors. To explicitly address autocorrelation in those regions, a multi-variate time series model could be applied in future research. To avoid overfitting, we restricted the maximum number of selected variables per RRM to half of the amount of available observations. This is a conservative approach, because in LASSO regression the number of nonzero coefficients is an unbiased estimate for the degrees of freedom (Zou et al., 2007). Though it could still be argued that the number of variables is high with respect to the number of observations (13 years), the comparable performance of the model in the cross-validation clearly points to a robust measurement of the influence of weather on maize yields that can be transferred to unknown data sets. The F-statistics and the lower RMSE values compared to a constant model suggest that the models are robust. Due to the lower model performance of the PDM, we applied a machine learning algorithm as an additional validation for the variable selection. The decision tree has the advantage of being less dependent on rigid model assumptions as found in regression analysis (e.g. it is also able to capture non-linear weather influences). Despite the limited amount of observations, the DT revealed largely similar weather-yield relations, underlining that there are large-scale patterns of weather impacts on maize yields in Peru and that the variable selection can be considered robust.

Given the higher performance of the RRM, we assessed the effect of more plant-available water on starchy maize yields on the regional scale. Higher plant-available water could be realised by several adaptation measures, but these need to be implemented with diligence as irrigation has already led to unsustainable water withdrawal rates (Drenkhan et al., 2015). Water consumption from the agricultural sector accounts for 89% of total fresh water use in Peru (ANA, 2013). In addition, with on-going climate change the water sources coming from the Andes are diminishing (Rabatel et al., 2013). Therefore, a focus should be on improving irrigation efficiency and other adaptation measures, such as improved water storage, water harvesting and better soil management practices, as proposed by the NDCs (Gobierno del Perú, 2018).

The result suggests that in some regions more water availability would neither decrease nor increase yields substantially (Section 3.3 and SI Table A.3). Apart from possible uncertainties stemming from the model coefficients or the model structure, this can be related to the presence of favourable rain conditions or sufficient adaptation practices already in place that regulate water availability, such as irrigation. In this case, other adaptation measures need to be tested to enable an increase in yields. The NDCs emphasize the need for good soil fertilization practices, erosion and flood control, salinity management, diversification of the production system, pest and disease control, improved seed varieties and the implementation of early warning and agricultural risk transfer systems. Also the necessity to provide information services in the agricultural sector, better access to markets and adding value to agricultural products are emphasised. Based on this list of possible adaptation options, local adaptation strategies need to be developed together with relevant stakeholders that take account of the interconnectedness of adaptation options in various sectors (Goosen et al., 2014). Our study provides an example of how a statistical modelling approach can inform this process.

As the results show, there are strong regional differences within Peru. Despite the high societal relevance of local weather-yield relations (Gobierno del Perú, 2018), a quantitative assessment of weather impacts on maize yield at the local level has not yet been done for Peru, to our best knowledge. To increase the pertinence of such modelling efforts, in particular for adaptation planning, comprehensive and consistent local yield data are required. The currently available data base seems limiting for this high-resolution modelling. Thus, once longer time series of high-quality yield and weather data become available on the local scale, future research should be directed towards supporting local adaptation planning by quantitative analyses of weather influences on crops.

Particularly in regions that show a low model performance, a further evaluation is needed and results should be tested on the ground. Despite the discussed uncertainties, we consider the assessment of more plant-available water useful as it provides a first indication of the effectiveness of adaptation measures suggested by the NDCs. Our results can be used to prioritize regions in which more water availability would potentially increase or decrease maize yields through the implementation of appropriate adaptation measures.

5. Conclusion

In this study, we assessed the influence of weather on starchy maize yields on the regional and the local scale in Peru based on a regional regression model, a panel data model and a machine learning algorithm. Based on these models, we assessed the effect of higher water availability on starchy maize yields, which is suggested by the Peruvian NDCs to adapt to climate change.

To our best knowledge, this is the first paper assessing weather-yield relations in Peru in such temporal and spatial detail and our study underlines the importance of a spatially-distinct assessment of adaptation options. Under such diverse climatic conditions as can be found in Peru, a local assessment is needed to account for the complex weather-yield relations. This study shows how a statistical approach can support the implementation of NDCs by providing quantitative information about the effectiveness of adaptation measures that can be used to identify priority areas for adaptation efforts.

Acknowledgements

The research was supported by the projects AgRATI, funded by Climate-KIC, and by the project EPICC. EPICC is part of the International Climate Initiative (IKI). The Federal Ministry for the Environment, Nature Conservation and Nuclear Safety (BMU) supports this initiative on the basis of a decision adopted by the German Bundestag.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.agrformet.2020.108154.

References

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration - Guidelines for computing crop water requirements - FAO Irrigation and drainage paper 56. Rom.

ANA, 2013. Plan Nacional De Recursos Hídricos Del Perú [WWW Document]. URL https://www.onemi.gov.cl/wp-content/themes/onemi-bootstrap-master/library/doc/plan_nacional_0_0.pdf (accessed 7.11.19).

Aybar, C., Fernández, C., Huerta, A., Lavado, W., Vega, F., Felipe-Obando, O., 2019. Construction of a high-resolution gridded rainfall dataset for Peru from 1981 to the present day. Hydrol. Sci. J. 00, 1–16. https://doi.org/10.1080/02626667.2019.1649411

Bourrel, L., Rau, P., Dewitte, B., Labat, D., Lavado, W., Coutaud, A., Vera, A., Alvarado, A., Ordoñez, J., 2015. Low-frequency modulation and trend of the relationship between ENSO and precipitation along the northern to centre Peruvian Pacific coast. Hydrol. Process. 29, 1252–1266. https://doi.org/10.1002/hyp.10247

Bozdogan, H., 1987. MODEL SELECTION AND AKAIKE'S INFORMATION CRITERION (AIC): THE GENERAL THEORY AND ITS ANALYTICAL EXTENSIONS. Psychometrika 52, 345–370.

Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. Classification and Regression Trees. Taylor & Francis Ltd.

Challinor, A.J., Watson, J., Lobell, D.B., Howden, S.M., Smith, D.R., Chhetri, N., 2014. A meta-analysis of crop yield under climate change and adaptation. Nat. Clim. Chang. 4, 287–291. https://doi.org/10.1038/nclimate2153

Christensen, J.H., Kumar, K.K., Aldria, E., An, S.-I., Cavalcanti, I.F. a., Castro, M. De, Dong, W., Goswami, P., Hall, A., Kanyanga, J.K., Kitoh, A., Kossin, J., Lau, N.-C., Renwick, J., Stephenson, D.B., Xie, S.-P., Zhou, T., 2013. Climate Phenomena and their Relevance for Future Regional Climate Change, in: Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M. (Eds.), Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1217–1308. https://doi.org/10.1017/CBO9781107415324.028

Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., Gao, X., Gutowski, W.J., Johns, T., Krinner, G., Shongwe, M., Tebaldi, C., Weaver, A.J., Wehner, M., 2013. Long-term Climate Change: Projections, Commitments and Irreversibility, in: Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M. (Eds.), Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1029–1136. https://doi.org/10.1017/CBO9781107415324.024

Drenkhan, F., Carey, M., Huggel, C., Seidel, J., Oré, M.T., 2015. The changing water cycle: climatic and socioeconomic drivers of water-related changes in the Andes of Peru. Wiley Interdiscip. Rev. Water 2, 715–733. https://doi.org/10.1002/wat2.1105

Dutra, E., Balsamo, G., Calvet, J.-C., 2015. Report on the current state-of-the-art Water Resources Reanalysis [WWW Document]. URL http://earth2observe.eu/files/Public Deliverables/D5.1_Report on the WRR1 tier1.pdf (accessed 7.1.19).

Friedman, J., Hastie, T., Rob Tibshirani, B.N., Tay, K., Simon, N., Qian, J., 2008. Regularization Paths for Generalized Linear Models via Coordinate Descent [WWW Document]. URL https://web.stanford.edu/~hastie/Papers/glmnet.pdf

Giorgi, F., Coppola, E., Raffaele, F., Diro, G.T., Fuentes-Franco, R., Giuliani, G., Mamgain, A., Llopart, M.P., Mariotti, L., Torma, C., 2014. Changes in extremes and hydroclimatic regimes in the CREMA ensemble projections. Clim. Change 125, 39–51. https://doi.org/10.1007/s10584-014-1117-0

Gobierno del Perú, 2018. Grupo de Trabajo Multisectorial de naturaleza temporal encargado de generar información técnica para orientar la implementación de las Contribuciones Nacionalmente Determinadas (GTM-NDC) [WWW Document]. URL http://www.minam.gob.pe/cambioclimatico/wp-content/uploads/sites/127/2019/01/190107_Informe-final-GTM-NDC_v17dic18.pdfPAÑOL.pdf (accessed 3.5.19).

Goosen, H., de Groot-Reichwein, M.A.M., Masselink, L., Koekoek, A., Swart, R., Bessembinder, J., Witte, J.M.P., Stuyt, L., Blom-Zandstra, G., Immerzeel, W., 2014. Climate Adaptation Services for the Netherlands: An operational approach to support spatial adaptation planning. Reg. Environ. Chang. 14, 1035–1048. https://doi.org/10.1007/s10113-013-0513-8

Gornott, C., Wechsung, F., 2016. Statistical regression models for assessing climate impacts on crop yields: A validation study for winter wheat and silage maize in Germany. Agric. For. Meteorol. 217, 89–100. https://doi.org/10.1016/j.agrformet.2015.10.005

Iglesias, A., Quiroga, S., Schlickenrieder, J., 2010. Climate change and agricultural adaptation: assessing management uncertainty for four crop types in Spain. Clim. Res. 44, 83–94. https://doi.org/10.3354/cr00921

INEI, 2019. Evoluacion de la Pobreza Monetaria - 2007-2018 - Informe Técnico [WWW Document]. URL https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1646/libro.pdf (accessed 6.22.19).

INEI, 2017. Perú - Encuesta Nacional Agropecuaria [WWW Document]. URL https://webinei.inei.gob.pe/anda_inei/index.php/catalog/654/ (accessed 11.30.18).

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K., Kiyotoshi, T., 2015. The JRA-55 reanalysis: General specifications and basic characteristics. J. Meteorol. Soc. Japan 93, 5–48. https://doi.org/10.2151/jmsj.2015-001

Konings, A.G., Gentine, P., 2017. Global variations in ecosystem-scale isohydricity. Glob. Chang. Biol. 23, 891–905. https://doi.org/10.1111/gcb.13389

Konings, A.G., Williams, A.P., Gentine, P., 2017. Sensitivity of grassland productivity to aridity controlled by stomatal and xylem regulation. Nat. Geosci. 10, 284–288. https://doi.org/10.1038/ngeo2903

Lobell, D.B., Burke, M.B., 2010. On the use of statistical models to predict crop yield responses to climate change. Agric. For. Meteorol. 150, 1443–1452. https://doi.org/10.1016/j.agrformet.2010.07.008

Lobell, D.B., Burke, M.B., Tebaldi, C., Mastrandrea, M.D., Falcon, W.P., Naylor, R.L., 2008. Prioritizing climate change adaptation needs for food security in 2030. Science (80-.). 319, 607–610. https://doi.org/10.1126/science.1152339

Lobell, D.B., Hammer, G.L., McLean, G., Messina, C., Roberts, M.J., Schlenker, W., 2013. The critical role of extreme heat for maize production in the United States. Nat. Clim. Chang. 3, 497–501. https://doi.org/10.1038/nclimate1832

Ludescher, J., Gozolchiani, A., Bogachev, M.I., Bunde, A., Havlin, S., Schellnhuber, H.J., 2014. Very early warning of next El Niño. Proc. Natl. Acad. Sci. 111, 2064–2066. https://doi.org/10.1073/pnas.1323058111

Marengo, J.A., Jones, R., Alves, L.M., Valverde, M.C., 2009. Future change of temperature and precipitation extremes in South America as derived from the PRECIS regional climate modeling system. Int. J. Climatol. 29, 2241–2255. https://doi.org/10.1002/joc.1863

MINAGRI, 2018. Anuario Estadístico de Producción Agrícola [WWW Document]. URL http://siea.minagri.gob.pe/siea/?q=publicaciones/anuario-de-produccion-agricola (accessed 12.19.18).

MINAGRI, 2017. Calendario de Siembras y Cosechas [WWW Document]. URL http://siea.minagri.gob.pe/calendario/ (accessed 10.3.18).

Mu, J.E., McCarl, B.A., Wein, A.M., 2013. Adaptation to climate change: Changes in farmland use and stocking rate in the U.S. Mitig. Adapt. Strateg. Glob. Chang. 18, 713–730. https://doi.org/10.1007/s11027-012-9384-4

Müller Schmied, H., Adam, L., Eisner, S., Fink, G., Florke, M., Kim, H., Oki, T., Theodor Portmann, F., Reinecke, R., Riedel, C., Song, Q., Zhang, J., Doll, P., 2016. Variations of global and continental water balance components as impacted by climate forcing uncertainty and human water use. Hydrol. Earth Syst. Sci. 20, 2877–2898. https://doi.org/10.5194/hess-20-2877-2016

Muneer, T., Gul, M.S., 2000. Evaluation of sunshine and cloud cover based models for generating solar radiation data. Energy Convers. Manag. 41, 461–482. https://doi.org/10.1016/S0196-8904(99)00108-9

Nash, J.E., Sutcliffe, J. V., 1970. River flow forecasting through conceptual models part I - A discussion of principles. J. Hydrol. 10, 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Novick, K.A., Ficklin, D.L., Stoy, P.C., Williams, C.A., Bohrer, G., Oishi, A.C., Papuga, S.A., Blanken, P.D., Noormets, A., Sulman, B.N., Scott, R.L., Wang, L., Phillips, R.P., 2016. The increasing importance of atmospheric demand for ecosystem water and carbon fluxes. Nat. Clim. Chang. 6, 1023–1027. https://doi.org/10.1038/nclimate3114

Patel, N., Upadhyay, S., 2012. Study of Various Decision Tree Pruning Methods with their Empirical Comparison in WEKA. Int. J. Comput. Appl. 60, 20–25. https://doi.org/10.5120/9744-4304

Peel, M.C., Finlayson, B.L., McMahon, T.A., 2007. Updated world map of the Köppen-Geiger climate classification. Hydrol. Earth Syst. Sci. 11, 1633–1644. https://doi.org/10.5194/hess-11-1633-2007

Platnick, S., Ackerman, S., King, M., Wind, G., Menzel, P., Frey, R., 2015. MODIS Atmosphere L2 Cloud Product (06_L2). NASA MODIS Adaptive Processing System, Goddard Space Flight Center, USA.

R Core Team, 2014. R: A language and environment for statistical computing [WWW Document]. URL http://www.r-project.org/

Rabatel, A., Francou, B., Soruco, A., Gomez, J., Cáceres, B., Ceballos, J.L., Basantes, R., Vuille, M., Sicart, J., Huggel, C., Scheel, M., Lejeune, Y., Arnaud, Y., Collet, M., Condom, T., Consoli, G., Favier, V., Jomelli, V., Galarraga, R., Ginot, P., Maisincho, L., Mendoza, J., Ménégoz, M., Ramirez, E., Ribstein, P., Suarez, W., Villacis, M., Wagnon, P., Palmas, L.,

2013. Current state of glaciers in the tropical Andes: a multi-century perspective on glacier evolution and climate change. Cryosph. 7, 81–102. https://doi.org/10.5194/tc-7-81-2013

Ray, D.K., Gerber, J.S., Macdonald, G.K., West, P.C., 2015. Climate variation explains a third of global crop yield variability. Nat. Commun. 6, 1–9. https://doi.org/10.1038/ncomms6989

Rodríguez-Morata, C., Díaz, H.F., Ballesteros-Canovas, J.A., Rohrer, M., Stoffel, M., 2019. The anomalous 2017 coastal El Niño event in Peru. Clim. Dyn. 52, 5605–5622. https://doi.org/10.1007/s00382-018-4466-y

Schauberger, B., Gornott, C., Wechsung, F., 2017. Global evaluation of a semiempirical model for yield anomalies and application to within-season yield forecasting. Glob. Chang. Biol. 23, 4750–4764. https://doi.org/10.1111/gcb.13738

Song, Y., Lu, Y., 2015. Decision tree methods: applications for classification and prediction. Shanghai Arch. Psychiatry 27, 130–135.

Sörensson, A.A., Menéndez, C.G., Ruscica, R., Alexander, P., Samuelsson, P., Willén, U., 2010. Projected precipitation changes in South America: A dynamical downscaling within CLARIS. Meteorol. Zeitschrift 19, 347–355. https://doi.org/10.1127/0941-2948/2010/0467

Therneau, T., Atkinson, B., Ripley, B., 2019. rpart: Recursive Partitioning and Regression Trees [WWW Document]. URL https://github.com/bethatkinson/rpart

Tibshirani, R., 1996. Regression Shrinkage and Selection Via the Lasso. J. R. Stat. Soc. Ser. B 58, 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O., Schewe, J., 2014. The inter-sectoral impact model intercomparison project (ISI-MIP): Project framework. Proc. Natl. Acad. Sci. U. S. A. 111, 3228–3232. https://doi.org/10.1073/pnas.1312330110

Wickham, H., 2009. ggplot2: elegant graphics for data analysis. Springer, New York.

Wooldridge, J.M., 2014. Introductory Econometrics A Modern Approach, 5th ed. Cenage Learning.

Zeileis, A., 2004. Econometric Computing with HC and HAC Covariance Matrix Estimators. J. Stat. Softw. 11, 128–129. https://doi.org/10.18637/jss.v011.i10

Zou, H., Hastie, T., Tibshirani, R., 2007. On the “degrees of freedom” of the lasso. Ann. Stat. 35, 2173–2192. https://doi.org/10.1214/009053607000000127