Agricultural and Forest Meteorology 217 (2016) 89–100

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

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

Statistical regression models for assessing climate impacts on crop yields: A validation study for winter wheat and silage maize in Germany

Christoph Gornott*, Frank Wechsung

Potsdam Institute for Climate Impact Research, Telegraphenberg A62, P.O. Box 601203, D-14412 Potsdam, Germany

ARTICLE INFO

Article history:
Received 12 June 2015
Received in revised form
27 September 2015
Accepted 1 October 2015
Available online 9 December 2015

Keywords:
Statistical crop yield models
Climate impacts
Yield changes
Winter wheat
Silage maize
Germany

ABSTRACT

For agriculture in Germany and generally all around the world, yield variability due to uncertain climate conditions represents an increasing production risk. Regional assessments of future yield changes can diminish this risk. For Germany's two most important crops winter wheat (*Triticum aestivum* L.) and silage maize (*Zea mays* L.), we investigate three regression models estimating relative climate impacts on relative crop yield changes: the separate time series model (STSM), the panel data model (PDM) and the random coefficient model (RCM). These regression models use the Cobb–Douglas function to capture climatic and non-climatic impacts on yields (e.g., changing prices or inventory management). The yield influencing climatic impacts contain the potential growth and stress factors during vegetative and reproductive plant development. The models are estimated and validated at the county scale. To improve the robustness and goodness of fit, the models are aggregated at the scale of German federal states, river basins and at the national scale. The observed yield changes are satisfactorily reproduced by all models for all aggregated scales (measured by the Nash–Sutcliffe efficiency (NSE)). According to their NSE values, the methodically simple STSMs reproduce extreme yield changes better (0.85) than the RCMs (0.79) and PDMs (0.72) at the national scale. This order can be also found across all scales when considering the models' goodness of fit. Generally, spatial aggregation increases the goodness of fit by +0.16 for federal states and river basins and by +0.29 for entire Germany compared to the county scale. The mean NSE increase is lowest for STSMs (+0.11), followed by RCMs (+0.13) and PDMs (+0.25) for federal states and river basins, which is opposite to the goodness of fit order. The model parameters show clear spatial patterns, which reflect regional differences of climate and soil. Within its methodological limits, our approach can directly be combined with the output of climate models and is suitable for assessing short- and medium-term yield effects for the current agronomic practice. It requires neither bias correction of the climate variables nor explicit modeling of crop yield trends.

1. Introduction

1.1. Statistical crop models for yield assessments

Crop yield assessments for upcoming climate anomalies or a systematic climate shift are of general interest for farmers, traders (e.g., grain mills, retailers), insurance companies, and policy makers. Statistical models (Ray et al., 2015; Iizumi et al., 2013; Mueller et al., 2012; Roberts et al., 2012; Schlenker and Roberts, 2009) and process based models (Asseng et al., 2013; Angulo et al., 2013; Palosuo et al., 2011) are model types for such assessments. Both model types are parametrized for past weather records. For future projections, they need weather records from climate simulation models. These climate models very often require a bias correction of the simulated output before they allow a reasonable yield projection (Lobell, 2013).

Process based crop models may not include all climate related effects on crop yields. There are many yield effects, which simply cannot be captured in process based models, because of limited spatial information about these effects. Examples are climate triggered effects on agronomic adaptation (irrigation, crop varieties, agronomic technics) or on pests, weeds, and diseases (Mueller et al., 2012). These climate triggered effects can be collinear with the climate variables. Since crop yields also contain climate triggered

* Corresponding author. Tel.: +49 331 288 2655.
E-mail address: gornott@pik-potsdam.de (C. Gornott).

http://dx.doi.org/10.1016/j.agrformet.2015.10.005
0168-1923/

effects, statistical crop yield models estimate in their parameter values not only the sole, but also the triggered effect of the climate variable. Process based models do not capture these climate triggered effects as long as they are not explicitly embedded in the models (Estes et al., 2013; Lobell and Burke, 2010). In the assessment of farm scale yield effects, this is an important disadvantage of process based models in comparison to statistical models.

Statistical yield models also allow relating inter-annual yield and yield factor changes (i.e., first order temporal ratios) instead of absolute values to each other (You et al., 2009; Lobell, 2007; Lobell and Asner, 2003). Considering changes instead of absolute values eliminates the trend of the variables and it allows neglecting systematic biases for exogenous variables for example when using simulated climate data from circulation models (Lobell, 2013). However, the neglected absolute level by using changes ignores a possible level dependency of yield and climate conditions. This limits the suitability to climate change assessments for changes within the range of recent climate variability. For yield projections beyond the yield variability of the dataset used for model estimation, process based models might be more appropriate (Rötter et al., 2011). At least, process based models should complement the statistical assessments under such circumstances.

The impact of climate on crop yields can be subdivided into two variable groups: variables that primarily determine potential growth and those that can be related to stress influences. The distinction is not disjunctive, overlaps might exist. We focus on the main influences that can contribute to a statistical explanation of the yield variability. The potential yield is determined mainly by the incoming solar radiation (Monteith, 1977; Long et al., 2006). The best usage of this incoming solar radiation requires an optimal mix of agronomic measures to establish the crop, to supply the necessary nutrients and water, and to keep biotic stress factors under control. Any divergence from this optimal mix will result in stress that reduces the potential yield. For these potential stress factors, we distinguish two groups: climate and management driven stress factors.

Among all possible climate driven stress factors, we hypothesize water stress as the most relevant stress factor for German winter wheat and silage maize yields (Wessolek and Asseng, 2006; Kersebaum and Nendel, 2014; Wolf and Diepen, 1994). Other possible influences, like temperature stress, might also exist in single years (Rötter and van de Geijn, 1999; Lobell et al., 2013), but are less generally associable with German climatic conditions. Management driven stress factors, like the crop variety, fertilizer, plant protection, and machinery, are reflected in the mean yield level and the yield trend. However, there are also economic conditions, e.g., statutory set-aside quotas or renewable energy subsidies for biogas and biodiesel, which influence the annual yield variability (Krause, 2008; Bakker et al., 2005). We use the fertilizer price and the acreage of the respective crops as proxy variables to control the economic yield impacts in the models. The fertilizer price represents the varying profitability of production factor inputs (e.g., seeds, plant protection, fuel, and fertilizer) and may directly affect the yield variability. The acreage of winter wheat and silage maize represents changes in the Common Agricultural Policy (CAP) of the European Union. An expanded acreage might generally suppress the yield level of both crops due to the inclusion of marginal productive land.

1.2. Modeling approach

In our approach, we follow the modeling concept introduced by Wechsung et al. (2008) and the validation scheme of Gornott and Wechsung (2015), who expanded the concept by two other statistical approaches. A level neutralizing transformation is applied for all variables, i.e., the crop yield, the climatic and the non-climatic variables. We utilize first order ratios $y'_t = y_t/y_{t-1}$ and $x'_t = x_t/x_{t-1}$, for the years $t = 2, \ldots, M$ of the endogenous variable crop yield y_t and the exogenous climatic and non-climatic variables x_t. As functional form, we use the Cobb–Douglas function analogous to Oury (1965). The function is proven in both economic (You et al., 2009) and agronomic applications (Lee et al., 2013) and considers yield impacts arising from substitution and interaction between the exogenous variables. The first order ratios are transformed to logarithmized first order ratios of yields and yield-factors, hereafter expressed as yield and factor changes. These changes allow an intercomparison of the effects of different variables.

We test three alternative ways to incorporate the spatial heterogeneity of yield changes and yield factor changes: by separately estimated time series models (STSMs), panel data models (PDMs), and random coefficient models (RCMs). All three approaches refer to a spatial dataset consisting of N discrete subunits and M years. In our case, the subunits are German counties within a federal state, river basin, or Germany as a whole. The methodically simple STSMs are estimated independently for the N subunits resulting in N parameter sets (Butler and Huybers, 2013; Lobell and Burke, 2010). In contrast, PDMs capture directly the temporal and spatial variability by one parameter set for all of the considered N subunits (You et al., 2009). RCMs can be ranked between PDMs and STSMs. They allow individual parameter variations per subunit and a parameter set for the entire unit (Reidsma et al., 2007). The results of the estimations will be presented and evaluated at two scales: the original spatial data scale, i.e., the German county yields, and the aggregated data scale, i.e., federal states, river basins, and entire Germany. Due to the aggregation, county- and farm-individual influences are largely averaged out, which might have biased the model results otherwise (Woodard and Garcia, 2008).

We restricted the temporal and spatial resolution of all variables to a division, which is accessible for climate simulations. The model results are evaluated at a larger scale than the estimation scale. Thus, we make explicit use of spatial aggregation effects. We test and apply the approach in respect to its possible suitability for fast impact assessment of seasonal- and medium-term projections (up to 30 years) from climate models. The approach is conducted for winter wheat and silage maize because these are the major winter and summer annual crops in Germany.

2. Materials and methods

2.1. Data

We use a spatial dataset of German crop yields per county for winter wheat and silage maize from 1991 to 2010. The dataset is supplied by the Statistical Offices of the Federation and the Länder (2013b). Climate data are available for the same period from 1218 German weather stations (DWD, 2011). The data were averaged per county to match the spatial resolution of the crop yield data. The total acreage of winter wheat and silage maize is taken from the datasets of the Statistical Offices of the Federation and the Länder (2013a) [1991–2008] and the Federal Statistical Office (2013) [2008–2010]. The fertilizer price index is published by the Statistical Offices of the Federation and the Länder (2013c). Ideally, all variables would be estimated at the county scale. However, the economic variables were only available on a national scale, so we applied the national values to all counties. A detailed description of the data is contained in the supplemental information (SI) S.1.