Available online at www.sciencedirect.com
SCIENCE DIRECT®

Agriculture, Ecosystems and Environment 110 (2005) 195–209

Agriculture Ecosystems & Environment

www.elsevier.com/locate/agee

Variability in regional wheat yields as a function of climate, soil and economic variables: Assessing the risk of confounding

Martha M. Bakker [a,*], Gerard Govers [b], Frank Ewert [c], Mark Rounsevell [a], Robert Jones [d]

[a] *Département de Géographie, Université Catholique de Louvain, Place Pasteur 3, B-1348 Louvain-la-Neuve, Belgium*
[b] *Laboratory for Experimental Geomorphology, K.U.Leuven, Redingenstraat 16, B-3000 Leuven, Belgium*
[c] *Plant Production Systems, Wageningen University, P.O. Box 430 6700 AK Wageningen, The Netherlands*
[d] *European Soil Bureau (CEC DG JRC), Environment Institute, TP 280 Via Enrico Fermi, Ispra (VA) IT 21020, Italy*

Received 5 October 2004; received in revised form 25 March 2005; accepted 4 April 2005
Available online 6 June 2005

Abstract

Mechanisms that explain spatial variability and trends in agricultural productivity at the regional scale are not well understood. Statistical approaches may be used to relate crop yields and trends in crop yields to changes in the economic and bio-physical environment. However, potential yield-explaining variables tend to confound at the regional scale due to strong correlations between these variables, which complicates the interpretation of such empirically derived relationships. In this paper, we assess relationships between different physical and economic variables and yields and trends in yields at the regional scale along a climatic gradient in Europe. We assess the extent of confounding (i.e. confusing the roles of different variables due to strong correlations) among these variables and the associated risk for explaining yield variability and trends in yields.

We analyze regional wheat yield data at NUTS3 and NUTS2 level for most of the EU-countries. Soft wheat (*Triticum aestivum* L.) is chosen as an indicator crop as it is grown over a wide climatic gradient, and its physiology has been subject to many agronomical studies. Time series were used to derive trends in yields. Data of important climatic, soil and economic variables were also derived at NUTS3 and NUTS2 level. Correlation coefficients were calculated between these variables and yields and yield trends. Confounding was assessed by comparing the R^2 values of the regression models with and without (groups of) variables.

Soft wheat productivity could be described to a very satisfying level. High R^2 values were obtained, partly due to aggregation of spatially autocorrelated in- and output data. High correlations were found between all variables, which indicates a risk of misinterpretation of results from statistical models when only few (groups of) variables are considered in the analysis. At a higher aggregation level (NUTS2) both the model fit and the risk of confounding increase. Validation by an independent dataset

* Corresponding author. Tel.: +32 10 478506; fax: +32 10 472877.
E-mail address: bakker@geog.ucl.ac.be (M.M. Bakker).

0167-8809/$ – see front matter
doi:10.1016/j.agee.2005.04.016

does not lead to exclusion of confounding. This paper serves as a basis for further research on spatial variability and trends in agricultural productivity at the regional scale, indicating both the possibilities and the risks of such research.

Keywords: Wheat yields; Productivity; Climate; Soils; Economics; Regression analysis

1. Introduction

Primary production and trends in productivity vary widely within the European Union, mainly due to variations in climate, soils and economic factors. However, understanding of relationships that determine this variability is limited and estimation of yield responses to changes in environmental conditions remains problematic. Considerable progress has been made in the past decades to improve process-based modeling of crop productivity at the plot and field scale (Brisson et al., 2003; Jamieson et al., 1998; Jones et al., 2003; Keating et al., 2003; Stockle et al., 2003; van Ittersum et al., 2003). However, research to examine the variability of crop yields at the regional scale has received little attention, since it has long been assumed that simple aggregation or extrapolation of results from fine-scale process-based crop models to broader scales would suffice (Brown and Rosenberg, 1997; Easterling et al., 1993, 2001; Nonhebel, 1996; Rosenzweig and Parry, 1994; Wolf, 1993). The few studies that specifically addressed the issue of scale for the regional application of crop models (Easterling et al., 1998; Olesen et al., 2000; Wassenaar et al., 1999) mainly focused on the required detail of input data to represent the heterogeneity of climatic and soil conditions within a study region. A comprehensive analysis of relationships that determine variability in regional yields in Europe has not been performed.

Available crop models use well-tested relationships to simulate responses of crop growth and yield at the plot and field level to variation in climatic conditions including water and/or nitrogen supply. However, at broader scales other yield-determining factors such as pests or economical factors emerge, which are often not considered (Ewert et al., 2002; Jamieson et al., 1999; Landau et al., 1998). Landau et al. (2000) used an extensive set of sophisticated climate variables with which they could predict up to 41% in regional wheat variability, leaving 59% unexplained, indicating that other, non-climate factors may play an important role. They assumed that indirect weather effects on grain filling, for example negative effects of rainfall, became more important than the direct weather effects. Veldkamp and Fresco (1997), on the other hand, found that labor force emerges as a yield-determining factor at regional scales.

Much of the yield increase achieved in the past decades was due to improved crop management and varieties through progress in plant breeding (Evans, 1997; Reynolds et al., 1999). These factors are probably related to the socio-economic conditions of a region but are generally not accounted for in process-based crop models. Hafner (2003) found that on a global scale, per-capita gross domestic product (GDP) has significant value as a descriptor for trends in wheat yields. Together with latitude, GDP explained approximately 50% of the observed variation in trends in yields (estimated from 64 to 67% concordant pairs derived from logistic regression).

Lack of process-knowledge and scarcity of detailed input data at broader scales force scientists to work with empirical models. However, even though significant relationships can often be detected (e.g. Hafner, 2003; Lobell and Asner, 2003) their interpretation remains problematic. Often, the basic assumptions of the scientist largely determine the conclusions drawn from a statistical analysis: an economist will find a reasonable relationship between GDP with yields, whereas a climatologist will find a good relationship between growing season length and yields, while both GDP and growing season may act as an alias for soil quality for example. Taking into account only one group of variables, whereby the impact of other variables is not taken into account, may lead to an overestimation of the reported effects. For example, Lobell and Asner (2003) concluded that climate was responsible for 25–32% of variability in yields trends in the continental USA. However, they did not control for variations in soil depth, which may vary systematically with climate, so that soil depth may have been responsible for at least a part of this 25–33%.