Neumann, K., Verburg, P. H., Stehfest, E., Müller, C. (2010): The yield gap of global grain production: a spatial analysis. - Agricultural Systems, 103, 5, 316-326

DOI: 10.1016/j.agsy.2010.02.004

Elsevier Editorial System(tm) for Agricultural Systems
Manuscript Draft

Manuscript Number: AGSY1602R1

Title: The yield gap of global grain production: A spatial analysis

Article Type: Research Paper

Keywords: Grain production; yield gap; land management; intensification; inefficiency; frontier analysis

Corresponding Author: M.Sc. Kathleen Neumann,

Corresponding Author's Institution:

First Author: Kathleen Neumann

Order of Authors: Kathleen Neumann; Peter H Verburg; Elke Stehfest; Christoph Müller

Abstract: Global grain production has increased dramatically during the past 50 years, mainly as a consequence of intensified land management and introduction of new technologies. For the future, a strong increase in grain demand is expected, which may be fulfilled by further agricultural intensification rather than expansion of agricultural area. Little is known, however, about the global potential for intensification and its constraints. In the presented study we analyze to what extent the available spatially explicit global biophysical and land management-related data are able to explain the yield gap of global grain production. We combined an econometric approach with spatial analysis to explore the maximum attainable yield, yield gap, and efficiencies of wheat, maize, and rice production. Results show that the actual grain yield in some regions is already approximating its maximum possible yields while other regions show large yield gaps and therefore tentative larger potential for intensification. Differences in grain production efficiencies are significantly correlated with irrigation, accessibility, market influence, agricultural labor, and slope. Results of regional analysis show, however, that the individual contribution of these factors to explaining production efficiencies strongly varies between world-regions.

The yield gap of global grain production: A spatial analysis

Kathleen Neumann, Peter H. Verburg, Elke Stehfest, Christoph Müller

Kathleen Neumann (corresponding author)
Land Dynamics Group, Wageningen University,
P.O. Box 47, 6700 AA, Wageningen, The Netherlands
Email: kathleen.neumann@wur.nl
Phone: phone: 0031-317-482430, fax: 0031-317-419000

Peter H. Verburg
Institute for Environmental Studies,VU University Amsterdam,
De Boelelaan 1087, 1081 HV, Amsterdam, The Netherlands
Email: Peter.Verburg@ivm.vu.nl

Elke Stehfest
Netherlands Environmental Assessment Agency (PBL),
P.O. Box 303, 3720 AH, Bilthoven, The Netherlands
Email: elke.stehfest@pbl.nl

Christoph Müller
Netherlands Environmental Assessment Agency (PBL),
P.O. Box 303, 3720 AH, Bilthoven, The Netherlands
and
Potsdam Institute for Climate Impact Research (PIK), Telegrafenberg,
PO Box 601203, 14412 Potsdam, Germany
Email: christoph.mueller@pik-potsdam.de

Abstract

Global grain production has increased dramatically during the past 50 years, mainly as a consequence of intensified land management and introduction of new technologies. For the future, a strong increase in grain demand is expected, which may be fulfilled by further agricultural intensification rather than expansion of agricultural area. Little is known, however, about the global potential for intensification and its constraints. In the presented study we analyze to what extent the available spatially explicit global biophysical and land management-related data are able to explain the yield gap of global grain production. We combined an econometric approach with spatial analysis to explore the maximum attainable yield, yield gap, and efficiencies of wheat, maize, and rice production. Results show that the actual grain yield in some regions is already approximating its maximum possible yields while other regions show large yield gaps and therefore tentative larger potential for intensification. Differences in grain production efficiencies are significantly correlated with irrigation, accessibility, market influence, agricultural labor, and slope. Results of regional analysis show, however, that the individual contribution of these factors to explaining production efficiencies strongly varies between world-regions.

Keywords: Grain production, yield gap, land management, intensification, inefficiency, frontier analysis

1 Introduction

Human diets strongly rely on wheat (*Triticum aestivum L.*), maize (*Zea mays L.*), and rice (*Oryza sativa L.*). Their production has increased dramatically during the past 50 years, partly due to area extension and new varieties but mainly as a consequence of intensified land management and introduction of new technologies (Cassman, 1999; Wood *et al.*, 2000; FAO, 2002a; Foley *et al.*, 2005). For the future, a continuous strong increase in the demand for agricultural products is expected (Rosegrant and Cline, 2003). It is highly unlikely that this increasing demand will be satisfied by area expansion because productive land is scarce and also increasingly demanded by non-agricultural uses (Rosegrant *et al.*, 2001; DeFries *et al.*, 2004). The role of agricultural intensification as key to increasing actual crop yields and food supply has been discussed in several studies (Ruttan, 2002; Tilman *et al.*, 2002; Barbier, 2003; Keys and McConnell, 2005). However, in many regions, increases in grain yields have been declining (Cassman, 1999; Rosegrant and Cline, 2003; Trostle, 2008). Inefficient management of agricultural land may cause deviations of actual from potential crop yields: the yield gap. At the global scale little information is available on the spatial distribution of agricultural yield gaps and the potential for agricultural intensification. There are three main reasons for this lack of information.

First of all, little consistent information of the drivers of agricultural intensification is available at the global scale. Keys and McConnell (2005) have analyzed 91 published studies of intensification of agriculture in the tropics to identify factors important for agricultural intensification. They emphasize that a plentitude of factors drive changes in agricultural systems. The relative contribution of them varies greatly between regions. This problem was confirmed by a number of studies that have investigated grain yields, and tried to identify factors that either support or hamper grain production at different scales (Kaufmann and Snell, 1997; Timsina and Connor, 2001; FAO, 2002a; Reidsma *et al.*, 2007). These studies also indicate that most of these factors are locally or regionally specific, which makes it difficult to derive a generalized set of factors that apply to all countries. A second reason for the absence of reliable information on the global yield gap is the limited availability of consistent data at the global scale. Especially land management data are lacking. When it comes to quantifying potential changes in crop yields often only biophysical factors, such as climate are considered while constraints for increasing actual crop yields are often neglected or captured by a simple management factor that is supposed to include all factors that cause a deviation from potential yields (Alcamo *et al.*, 1998; Harris and Kennedy, 1999; Ewert *et al.*, 2005; Long *et al.*, 2006). Finally, lack of data also leads to another difficulty. Many yield gap analyses have in common that they apply crop models for simulating potential crop yields which are compared to actual yields (Casanova *et al.*, 1999; Rockstroem and Falkenmark, 2000; van Ittersum *et al.*, 2003). Potential yields, however, are a concept describing crop yields in absence of any limitations. This concept requires assumptions on crop varieties and cropping periods. While such information is easily attainable at the field scale it is not available at the global scale. Moreover, different simplifications of crop growth processes exist between the models. This may result in uncertainties of globally simulated potential yields, and makes an appropriate model calibration essential for global applications. Comparing simulated global crop yields to actual yields therefore bears the risk of dealing with error ranges and uncertainties of different data sources (i.e., observations and simulation results) which might even outrange the yield gap itself.

Consequently, available knowledge about the yield gap is rather inconsistent and regional and global levels of agricultural production have hardly been studied together.

The aim of this paper is to overcome some of the mentioned shortcomings by analyzing actual yields of wheat, maize, and rice production at both regional and global scale accounting for biophysical and land management-related factors. We propose a methodology to explain the spatial variation of the potential for intensification and identifying the nature of the constraints for further intensification. We estimated a stochastic frontier production function to calculate global datasets of maximum attainable grain yields, yield gaps, and efficiencies of grain production at a spatial resolution of 5 arc minutes (approximately 9.2 x 9.2 km on the equator). Applying a stochastic frontier production function facilitates estimating the yield gap based on the actual grain yield data only, instead of using actual and potential grain yield data from different sources. Therefore, the method allows for a robust and consistent analysis of the yield gap. The factors determining the yield gap are quantified at both global and regional scales.

2 Methodology

2.1 The Stochastic Frontier Production Function

Stochastic frontier production functions originate from economics where they were developed for calculating efficiencies of firms (Aigner *et al.*, 1977; Meeusen and Broeck, 1977). Since agricultural farms are a special form of economic units this econometric methodology can also be used to calculate farm efficiencies and efficiencies of agricultural production in particular. In our global analysis, the agricultural production within one grid cell (5 arc minute resolution) is considered as one uniform economic unit. The stochastic frontier production function represents the maximum attainable output for a given set of inputs. Hence, it describes the relationship between inputs and outputs. The frontier production function is thus “a regression that is fit with the recognition of the theoretical constraint that all actual productions lie below it” (Pesaran and Schmidt, 1999). In case of agricultural production the frontier function represents the highest observed yield for the specified inputs. Inefficiency of production causes the actual observations to lie below the frontier production function. The stochastic frontier accounts for statistical noise caused by data errors, data uncertainties, and incomplete specification of functions. Hence, observed deviations from the frontier production function are not necessarily caused by the inefficiency alone but may also be caused by statistical noise (Coelli *et al.*, 2005).

The frontier production function to be estimated is a Cobb-Douglas function as proposed by Coelli *et al.* (2005). Cobb-Douglas functions are extensively used in agricultural production studies to explain returns to scale (Bravo-Ureta and Pinheiro, 1993; Bravo-Ureta and Evenson, 1994; Battese and Coelli, 1995; Reidsma *et al.*, 2009b). If the output increases by the same proportional change in input then returns to scale are constant. If output increases by less than the proportional change in input the returns decrease. The main advantage of Cobb-Douglas functions is that returns to scale can be increasing, decreasing or constant, depending of the sum of its exponent terms. In agricultural production decreasing returns to scale are common. The Cobb-Douglas function is specified as following:

$$\ln(q_i) = \beta_1 x_i + v_i - u_i \qquad \textbf{Equation 1}$$

where $\ln(q_i)$ is the logarithm of the production of the *i*-th grid cell (i = 1, 2,…N), x_i is a $(1 \times k)$ vector of the logarithm of the production inputs associated with the *i*-th grid cell, *ß* is a $(k \times 1)$ vector of unknown parameters to be estimated and v_i is a random (i.e., stochastic) error to account for statistical noise. Statistical noise is an inherit property of the data used in our study resulting from reporting errors and inconsistencies in reporting systems. The error can be positive or negative with a mean zero. The non-negative variable u_i represents inefficiency effects of production and is independent of v_i. Figure 1 illustrates the frontier production function.

Insert Figure 1 here

Stochastic frontier analyses are widely used for calculating efficiencies of firms and production systems. The most common measure of efficiency is the ratio of the observed output to the corresponding frontier output (Coelli *et al.*, 2005):

$$E_i = \frac{q_i}{\exp(x'_i ß + v_i)} = \frac{\exp(x'_i ß + v_i - u_i)}{\exp(x'_i ß + v_i)} = \exp(-u_i) \qquad \textbf{Equation 2}$$

where E_i is the efficiency in the *i*-th grid cell. The efficiency is an index without a unit of measurement. The observed output at the *i*-th grid cell is represented by q_i while $x'_i ß$ is the frontier output. The efficiency E_i determines the output of the *i*-th grid cell relative to the output that could be produced if production would be fully efficient given the same input and production conditions. The efficiency ranges between zero (no efficiency) and one (fully efficient).

Kudaligama and Yanagida (2000) applied stochastic frontier production functions to study inter-country agricultural yield differences at the global scale. However, that study disregards spatial variability within countries, which can be very large. To our knowledge, our study presents the first application of a stochastic frontier function to grid cell specific crop yield data at the global scale. At the national and regional scale a number of authors have applied frontier production functions to calculate both efficiencies of grain productions and frontier grain productions (Battese, 1992; Battese and Broca, 1997; Tian and Wan, 2000; Verburg *et al.*, 2000). Each of these studies contribute significantly to the understanding of variation in grain yields and agricultural production efficiencies. However, most of these studies lack a comprehensive analysis and discussion of the spatial variations of the yield gap and production efficiencies within the region considered.

2.2 Global level estimation of frontier yields and efficiencies

We applied a stochastic frontier production function to calculate frontier yields, yield gaps, and efficiencies of wheat, maize, and rice production. Thereby, we integrated both biophysical and land management-related factors. In our analysis the actual grain yield is defined as observed grain yield expressed in tons per hectare. The frontier yield is indicative for the highest observed yield for the combination of conditions. Global data on actual grain yields were obtained from Monfreda *et al.* (2008). These datasets comprise information on harvested areas and actual yields of 175 crops in 2000 at a 5 arc minute resolution and are based on a combination of national-, state-, and county-level census statistics as well as information on global cropland area (Ramankutty *et al.*, 2008).

The vector of independent variables in the frontier production function contains several crop growth factors. Crop growth factors can be classified as *growth-defining*, *growth-*

limiting, and *growth-reducing* factors (van Ittersum *et al*., 2003). According to van Ittersum *et al*. (2003) growth-defining factors determine the potential crop yield that can be attained for a certain crop type in a given physical environment. Photosynthetically Active Radiation (PAR), carbon dioxide (CO_2) concentration, temperature and crop characteristics are the major growth-defining factors. Growth-defining factors themselves cannot be managed but management adapts to these conditions, for example by choosing the most productive growing season. Growth-limiting factors consist of water and nutrients and determine water- and nutrient-limited production levels in a given physical environment. Availability of water and nutrients can be controlled through management to increase actual yields towards potential levels. Growth-reducing factors, such as pests, pollutants, and diseases reduce crop growth. Effective management is needed to protect crops against these growth-reducing factors. The interplay of growth-defining, growth-limiting, and growth-reducing factors determines the actual yield level.

The stochastic frontier production function was composed in such a way that the frontier grain yield is defined by growth-defining factors, precipitation and soil fertility constraints. Hence, frontier yields may be below potential yields because they consider growth-limiting factors for their calculation. Factors that determine the deviation from the frontier grain yield, and hence lead to the actual grain yield, are called inefficiency effects and are considered in the inefficiency function u_i. According to our definition this yield gap is caused by inefficient land management. The stochastic frontier production function to be estimated for each grain type:

$$\ln(q_i) = ß_0 + ß_1\ln(temp_i) + ß_2\ln(precip_i) + ß_3\ln(par_i) + ß_4\ln(soil\_const_i) + v_i - u_i$$

Equation 3

where q_i is the actual grain yield, specified per grain type. The most important crop growth-defining factors are PAR (par_i) and temperature. The relation between temperature and grain yield is not log-linear as it is implied by the Cobb-Douglas stochastic frontier model. Increasing temperature first leads to an optimum grain yield before the yield declines again. We therefore defined the variable $temp_i$ as the deviation from the optimal monthly mean temperature. The optimal monthly mean temperature is the mean monthly temperature at which the highest crop yields are observed according the observed actual crop yields. CO_2 concentration, another growth-defining factor, was not included in our production function because only slight CO_2 concentration differences exist between the Northern and Southern Hemisphere and local CO_2 concentrations show hardly any spatial variability. Precipitation ($precip_i$) and soil fertility constraints ($soil\_const_i$) represent growth-limiting factors, which can be controlled by management. Rather than using annual averages for each climatic variable, monthly mean temperature, precipitation, and PAR data were integrated over the grain type specific growing period (Table 1). The growing period is defined as the period between sowing date and harvest date which differs between grain type and climatic conditions and thus location. Using growing period specific climate data allows us to account for only those climate conditions which contribute significantly to grain development. A similar approach is also used in many crop modeling approaches (for examples see Kaufmann and Snell, 1997; Jones and Thornton, 2003; Parry *et al.*, 2004; Stehfest *et al.*, 2007). Empirical data on growing season were available for irrigated rice (Portmann *et al.*, 2008), while we obtained grain specific growing period information for wheat and maize from the LPJmL model (Bondeau *et al.*, 2007). Cropping periods for rice are based on irrigated rice and the same growing period was applied for both irrigated and non-irrigated rice production areas because data on non-irrigated rice were not available. A full sensitivity analysis of the effect of cropping period choice was beyond the scope of this paper. A description of all variables used is given in Table 1.

The influence of land management on the actual grain yield was considered in the inefficiency function u_i. Several regional and global studies have identified factors which determine land management and intensification (Tilman, 1999; Kerr and Cihlar, 2003; Keys and McConnell, 2005; Reidsma *et al.*, 2007). Only a few of these factors are available as spatially explicit global datasets. Therefore, proxies of these factors for which global datasets are available were used instead as determinants of land management. The inefficiency function is specified as:

$$u_i = \delta_1(irrig_i) + \delta_2(slope_i) + \delta_3(agr\_pop_i) + \delta_4(access_i) + \delta_5(market_i)$$

Equation 4

Irrigation $(irrig_i)$ as a traditional management technique for improving actual grain yields was taken into account. Slope $(slope_i)$ might restrict actual grain yield because it hinders accessing land with machinery, leads to surface runoff of (irrigation) water, and supports soil erosion which limits soil fertility. Nevertheless, adverse slope conditions can, to a certain extent, be offset by effective management and were therefore considered in the inefficiency function. The importance of labor as determinant of agricultural production has been discussed and analyzed in several studies (Battese and Coelli, 1995; Mundlak *et al.*, 1997; Hasnah *et al.*, 2004; Keys and McConnell, 2005). A proper consideration of agricultural labor at the global scale remains, however, challenging with limited data availability as a major obstacle. For this reason we used non-urban population data as proxy for agricultural population and hence labor availability *(agr_pop$_i$)*. Market accessibility *(access$_i$)* gives an indication of the attractiveness of regions for grain production in terms of the time-costs to reach the closest market. We considered the accessibility of the nearest markets, including large harbors, which are the door to distant markets as well. A proxy for the market influence *(market$_i$)* was included in the inefficiency function as it is assumed that regions with stronger markets are better suited for investments in yield increases of agricultural production than regions with less strong markets. *Market$_i$* and *access$_i$* are at the same time proxies for the availability of fertilizers, pesticides and machinery.

Fertilizer application, one of the most important management options to increase actual grain yields (Tilman *et al.*, 2002; Alvarez and Grigera, 2005) could not be included in the inefficiency function due to lack of appropriate data. Globally consistent and comparable fertilizer application data are only available at the national scale. We obtained grain type specific fertilizer application rates per country from the International Fertilizer Industry Association (IFA) (FAO, 2002b). A correlation analysis to identify the relationship between fertilizer application and efficiency of grain production was done with these data at the national level.

We computed a globally consistent grain yield frontier under the assumption of globally uniform relations with the growth-defining, growth-limiting, and growth-reducing factors. This consistency allows us to directly compare estimated frontier yields, efficiencies and yield gaps between grid cells across the globe. Only 5 arc minute grid cells with a cropping area of at least 3% coverage of the particular grain type were considered in the analysis to prevent an overrepresentation of marginal cropping areas. From these grid cells a random sample of 10% with a minimum distance of two grid cells between each sampled grid cell was chosen to allow efficient estimations and reduce spatial autocorrelation, which may have been caused by the characteristics of the data that were derived from administrative units of varying size (Monfreda *et al*., 2008). We tested the robustness of this 10% sample to verify the appropriateness of the sample size. Maximum-likelihood estimates of the model parameters were estimated using the software FRONTIER 4.1 (Coelli, 1996).

Insert Table 1 here

2.3 Regional level estimation of frontier yields and efficiencies

The importance of the variables explaining the efficiencies is hypothesized to be different between world-regions. For example, the conclusion that slope is a determining factor for efficiencies of global wheat production does not rule out the possibility that in some world-regions slope does not influence efficiency of wheat production while other variables do. To uncover such differences, we conducted a second analysis at the scale of world-regions. World-regions consist of countries with strong cultural and economic similarities. We distinguish 26 world-regions for the regional analysis.

If frontier yields and efficiencies are calculated for each world-region individually inconsistencies may be introduced since some world-regions may not contain grid cells with actual yields close to the frontier yields. Such analysis can lead to an underestimation of the frontier yield. Efficiencies were therefore calculated at the global scale to retrieve globally comparable frontier yields. However, in this case efficiencies were calculated without synchronously estimating the inefficiency effects contrary to the global approach in section 2.2. The applied stochastic frontier production function remains the same (Equation 3); however, the inefficiency effects are not synchronously estimated. In our regional analysis, forward stepwise regressions were applied to identify the statistically significant inefficiency effects (independent variables) and to determine their relative contribution to the overall efficiency of grain production (dependent variable) per world-region (Equation 5).

$$ln(eff_i) = ß_0 + ß_1(irrig_i) + ß_2(slope_i) + ß_3(agr\_pop_i) + ß_4(access_i) + ß_5(market_i)$$

Equation 5

where eff_i is the efficiency in each grid cell. Again, efficiency in our study is defined as the actual yield in relation to the frontier yield. The percentage of grain area within a grid cell was used as weighting factor. The natural logarithm was calculated for the efficiency in order to account for non-linear relations. The variance inflation factor (VIF) was calculated to ensure independence amongst the variables. Variables with a VIF of 10 or higher were removed from the analysis.

3 Results

3.1 Global frontier yields and efficiencies

All coefficients in the stochastic frontier production function are significant at 0.05 level (Table 2). The deviation from optimal monthly mean temperature *(temp)* has a negative coefficient for all grain types, meaning that the frontier grain yield decreases with an increasing deviation from the optimal monthly mean temperature. The relationship is strong indicated by the large t-ratios (Table 2). *Precip* and *soil_const* also determine a significant share explaining the frontier production. The positive coefficients for *precip* for all three grain types indicate that with an increased precipitation sum the grain yield increases. The negative coefficient for *par* for all three grain types may be related to cloudiness which is closely related to precipitation. Another reason for the negative coefficient for *par* may be that the higher PAR (and consequently energy influx), the higher potential evapo-transpiration, which causes water stress and might therefore decrease frontier grain yields. Furthermore, a relationship between the temperature sum over the growing period and *par* for all three grain types (Pearson correlation coefficient r>= 0.67) is potentially causing multicollinearity. While frontier yields of maize and rice are negatively correlated to *soil_const*, a positive coefficient for *soil_const* for wheat is obtained. Highest actual wheat yields are found in countries with highly mechanized and capital intensive agriculture, such as Denmark and Germany. Soil fertility constraints in these countries can be reduced by an effective land management, especially fertilizer application. Hence, soil fertility constraints are only up to a certain level not an obstacle for wheat production in those countries. Because these countries supply a large share of global wheat production this may explain the positive coefficient for wheat. It is unlikely that there is a causal relation underlying this observation.

In the inefficiency function, a positive coefficient indicates that the respective variable has a negative influence on efficiency. *Irrig* and *market* have negative coefficients for all grain types. Hence, the absence of irrigation and a low market influence reduce efficiency. The coefficient for *slope* is positive for wheat and maize but negative for rice. Steeper slopes indicate lower efficiencies in wheat and maize production. The negative coefficient for rice may be explained by the large amount of global rice that is produced on terraces in sloped areas, especially in the core production regions in South-

East Asia. The production on terraces is very intensive and may explain high actual yields and efficiencies. Furthermore, in many hilly regions rice is produced on the valley bottoms. Due to the limited spatial resolution of the analysis these locations are represented as sloping, leading to a possible negative association with inefficiency. The positive coefficients for *access* are all as expected. Hence, the more hours needed to reach the next city, the lower the efficiency of grain production. According to the theory of von Thuenen (1966), who concludes that crop production is only profitable within certain distances from a market, crop production becomes less productive and less efficient in more remote regions. Somewhat surprising results are achieved for *agr_pop*. While the coefficient for wheat is negative as expected it is positive for maize and rice. It can be argued that for many less developed countries the more labor is available the lower is the technology level and, therefore, the efficiency. This applies for many rice and maize growing countries as shown with our results. Furthermore, the percentage of agricultural population as part of the non-urban population tends to be smaller nearby urban agglomerations. In those regions agricultural activities provide often only a small contribution to the non-urban household income whereas off-farm activities are the primary income source, which tends to be associated with lower agricultural efficiencies (Verburg *et al.*, 2000; Goodwin and Mishra, 2004; Paul and Nehring, 2005).

The correlations (Pearson coefficients) for fertilizer application and the grain production efficiency at country level are r = 0.67 for wheat, r = 0.59 for maize and r = 0.27 for rice. Countries with lower fertilizer application rates therefore achieve lower efficiencies in grain production than countries with higher fertilizer application rates.

Results of the obtained likelihood-ratio tests are shown in Table 2. The likelihood ratio (LR) statistics for wheat (LR = 4307), maize (LR = 3695) and rice (LR = 1558) exceed the 1% critical values of 21.67 for 6 degrees of freedom and therefore indicate high statistical significance (Kodde and Palm, 1986). A Wald test was conducted to test the significance of all included variables. Results indicate that we can only explain about half of the efficiencies in wheat production ($\gamma = 0.47$). This means that the other half of the variation cannot be explained by inefficiency effects but rather by statistical noise. The γ-values for maize and rice are much higher: 0.91 for both. Hence, a major part of the error term is due to inefficiency rather than statistical noise. Reasons for the remarkable differences between the obtained γ-values are diverse. Statistical noise in our study is an inherent data property possibly introduced by data errors or data uncertainties. The large variation of sources and years of validity of the grain yield data and the different size of the administrative units that underlie these datasets are likely to cause high uncertainties. Input data are not validated and it can be expected that some of them are more accurate than others with large differences between regions. Statistical noise may also be caused by variances within the data. For example, variability of climate within a particular month may influence crop management but cannot be captured by mean monthly climate data. Furthermore, actual yields are likely to reflect large inter-annual variations due to climate variation which is not captured by the long-term average climate parameters used in this study. Uncertainties in cropping periods may also add to the statistical noise. Furthermore, we considered only a limited number of inefficiency effects to explain spatial variation in efficiencies.

The mean efficiencies for wheat, maize and rice are 0.637, 0.501 and 0.638, respectively (Table 2). Hence, the highest efficiencies at global scale are obtained for production of wheat and rice, while maize production is the least efficient.

Insert Table 2 here

Frontier grain yields show a wide variation across the globe. Exemplary regions with high frontier yields are Northwest Europe, central USA, and parts of China, while central Asia, Mexico, and West Africa show low frontier yields for wheat, maize, and rice production respectively (Figure 2).

Insert Figure 2 here

Figure 2 and 3 illustrate that some regions produce grain close to the estimated frontier yields while others show a large yield gap. These yield gaps are an indication for the potential to increase actual grain yields. The maximum yield gaps are 7.5 t/ha for wheat, 8.4 t/ha for maize and 6.4 t/ha for rice. If we express the global aggregated yield gap in total production (i.e. in tons) we can show that the yield gap equals 43%, 60%, and 47% of the actual global production of wheat, maize and rice, respectively.

Insert Figure 3 (Maps 1-3) here

3.2 Regional determinants of efficiencies

We present and discuss only the most important results of the region-specific analysis of factors that explain efficiencies. Two world-regions per grain type, which are characterized by a different agricultural, cultural and economical background, were selected and are presented in Table 3. Results show that the individual contribution of determinants of efficiencies varies strongly between world-regions and grain types (Figure 4).

Insert Table 3 here

The results indicate that regional efficiencies of grain production can be explained by irrigation *(irrig)* in five of the six presented world-regions. The coefficients for *irrig* are all positive, but the individual contributions vary between world-regions. For example, in the Thailand region intensive irrigation is only applied in some rice growing regions, e.g. in the surroundings of Bangkok and in the Mekong Delta while rain-fed rice production mostly faces severe constraints in obtaining a highly efficient production. *Irrig* explains most of the variance in efficiency of rice production in the Thailand region. Market accessibility *(access)* can explain efficiencies of grain production in the USA, Southern Africa, Indonesia and the Thailand region. For all regions poor accessibility mean lower efficiency of grain production but the contribution of *access* differs between world-regions. For example, the USA is the world's main wheat exporter and *access* can explain most of the variability in wheat efficiency. In the more remote regions land prices are lower and inputs are therefore often substituted by land leading to lower efficiencies. China's wheat export is minor with less than 1% of its total production (FAOSTAT, 2009) and within the densely populated wheat production areas generally little time is needed for reaching markets. *Access* can therefore not explain the variance in efficiency of Chinese wheat production. Market influence *(market)*, as a proxy for land rent indicating the investments in machinery, pesticides and fertilizer, has a positive coefficient for most grain types and regions: especially for maize production. A large part of the variance in efficiency of maize production in Mexico and Southern African can be explained by the variation in market influence while it can neither explain efficiencies of wheat production in the USA nor efficiencies of rice production in the Thailand region. Agricultural population *(agr_pop)* as proxy for agricultural labor has a positive contribution to efficiencies of rice production in the Thailand region, Indonesia, and wheat production in the USA and China, while its contribution is negative for maize production in Southern Africa. For both Indonesia and the Thailand region these results can be traced back to the labor intensity of rice production with large number of people engaged in rice production and post-production activities including processing, storage, and transport. Also Chinese cereal production is well-known for being labor intensive. Farmers try to substitute capital and land with labor which explains the positive coefficient as also confirmed by Tian and Wan (2000). *Slope* explains most of the variability in efficiency of Chinese wheat production. Actual wheat yields in China are significantly higher in flat areas (yellow river valley) as these areas are easier to access and allow for better use of machinery. China's rapid urbanization has, however, forced wheat farmers to also produce in less productive, for example more hilly regions to meet the food demand (Chen, 2007; Xin *et al.*, 2009). *Slope* coefficients are also positive for rice production in Indonesia and the Thailand region and for Mexico. Mexican maize is largely produced in the highlands of Mexico. However, *slope* adds less to the explanation of efficiency of maize production than most of the other inefficiency effects.

Insert Figure 4 (Maps 1-3) here

4 General discussion

4.1 Evaluation of data and methodology

Agricultural production efficiency, yield, and intensification are closely linked (de Wit, 1992; Matson *et al.*, 1997; Cassman, 1999; Reidsma *et al.*, 2009b). In this paper we have shown how to disentangle actual grain yields from production efficiencies by using stochastic frontier production functions. The strength of our approach lies in its integration of biophysical and land management-related determinants of grain yields. Kaufmann and Snell (1997) showed that climate variables alone account for only a minor part of the variation in US maize yield while socio-economic variables, such as farm size, technology, and loan rates, account for the main part of yield variation. This example underpins the necessity to include socio-economic variables when exploring crop yields. The selection of land management-related factors included as inefficiency effects in our analysis was, however, heavily restricted by data availability. Additional aspects related to agricultural production that may be considered are for example stimulation of alternative management options, applied technology, land ownership, farm size, and land degradation. All these factors may affect the yield gap but their consideration was beyond the scope of our study as consistent spatially explicit data are not available at the global scale.

The presented approach combines econometric methods with concepts applied in crop sciences. The Cobb-Douglas function implies a log-linear relation between dependent and independent variables. This may, however, be inappropriate to present the relation between yield, growth-defining, and growth-limiting factors as some of these factors may not have such a relationship. Yet, the data did not provide an indication that another functional form would be more appropriate.

A big advantage of the frontier production approach is the consistent use of one dataset of observed yields. Observed grain yield data were derived from different national censuses and partly show constant values for each grid cell belonging to the same administrative unit (Monfreda *et al.*, 2008). We minimized this effect (that causes spatial autocorrelation of observations) by excluding all minor cropping areas from the analysis and using a sample with a minimum distance between the sampled grid cells. Alternatively, observed yields may be compared to simulated potential yields. However, only few model results of potential yields at the global scale are available. A simple comparison of published maps of potential yields originating from different models indicates large deviations between the simulated potential yields. The deviation between simulated potential yields is often larger than the yield gap itself, which makes a reliable yield gap analysis impossible based on these simulated yields (MNP, 2006; Bondeau *et al.*, 2007).

4.2 Closing the yield gap

Potential yields were explored in many studies. One of the first studies carried out at the global scale was published by Buringh *et al.* (1975) who assessed maximum grain production per soil region. The authors calculated the highest total production levels for Asia and Africa with up to 14.000 Mio tons/ yr but did not explore variability of grain yields within each soil region. In recent studies, Reidsma *et al.* (2009a) has simulated water-limited potential maize yields for Europe and observes a gradient from the North-East of Europe to the South-West. Our frontier yields confirm this trend, although the gradient is weaker and the frontier yields tend to be higher than the model results. The same is observed for frontier wheat yields for the North China Plain which are tentative higher (up to 10 tons/ ha) than potential wheat yields simulated by Wu *et al.* (2006) which do not exceed 8 tons/ ha. Peng *et al.* (1999) have conducted several field level experiments and conclude potential rice yields of about the 10 tons/ ha for the tropics. We can, however, not confirm such high frontier rice yields for the tropics, those we have only estimated for Central China where hybrid rice technology has been widely adopted (Cassman, 1999).

We define the process of closing the yield gap as intensification. To increase actual grain yields through intensification a catalyst is needed to initialize the intensification process. Lambin et al. (2001) have identified three trigger of agricultural intensification: 1) land scarcity, 2) investments in crops and livestock, and 3) intervention in state-, donor-, or non-governmental organization (NGO)-sponsored projects to further push development in a region or economic sector. For exploring potential temporal dynamics of intensification it is essential to know whether these triggers exist and how these interact with local constraints. The results of our analysis have confirmed that the factors explaining inefficiencies in production widely vary by region. Furthermore, factors explaining efficiencies are related to complex social, economic, and political processes. Taking this into account it is debatable to what extent the calculated yields gaps can and will be closed. Particularly developing and transition countries often lack capital investments, infrastructure, education, and effective agricultural policies and agricultural expansion is practiced instead to increase grain yield (Reardon *et al.*, 1999; Swinnen and Gow, 1999; Coxhead *et al.*, 2002). The presented frontier yields illustrate what currently could be achieved while breeding improvements may lead to higher yielding varieties in the future. Several authors have discussed the role of technological development to further increase potential crop yields (Cassman, 1999; Evans and Fischer, 1999; Huang *et al.*, 2002) but its specific contribution remains difficult to determine (Ewert *et al.*, 2005).

Another aspect to be considered when exploring grain yields is the effect of climate change. Climate change is expected to have different impacts on agricultural yields in different parts of the world and for different crop types (Parry *et al.*, 2004; Erda *et al.*, 2005; Thornton *et al.*, 2009; Wei *et al.*, 2009). The presented methodology and results may be used for assessing the impact of climate change on actual and potential grain yields as well as for investigating possible adaptation strategies. A negative aspect often associated with intensification is environmental damage. Many studies have shown that agricultural intensification may lead to air and water pollution, loss of biodiversity, soil degradation and erosion (Harris and Kennedy, 1999; Donald *et al.*, 2001; Foley *et al.*, 2005) and more and more authors emphasize the need for a more efficient use of natural resources and ecological intensification (Cassman, 1999; Tilman, 1999).

5 Conclusions

In this study we explored factors associated to grain production efficiencies and yield gaps of global grain production. We explained the spatial variation across the globe to explore the potential for intensification and the nature of the constraints given the current technological development. Results show that on average the present actual yields of wheat, maize, and rice are 64%, 50%, and 64% of their frontier yields, respectively. Based on these results it appears tempting to conclude a tremendous potential for intensification of global grain production. In fact, quantitative assessment of intensification potential remains challenging as intensification has multiple pathways and often goes parallel with agricultural expansion. Minimizing the yield gap requires understanding the nature and strength of region-specific constraints. From our results we can conclude that, while some factors can explain efficiencies of global grain production the same factors may not be relevant at the world-regional scale. Hence, the efficiency of grain production is the result of several processes operating at different spatial scales but the influence of each of these processes differs between the scales. From the comparison of our global results with the regional results we can conclude that these processes do not necessarily behave linearly across these scales. Drawing conclusions from the global results about factors explaining grain production efficiencies at the regional scale would therefore be wrong. Hence, region-specific identified constraints need to be assessed separately to provide a basis for increasing actual grain yields. This paper has provided a first global overview of the spatial distribution of the influence of some of these factors.

Acknowledgement

This research is contributing to the Global Land Project (GLP). The authors acknowledge the BSIK RvK IC2 project ‘Integrated analysis of emission reduction over regions, sectors, sources and greenhouse gases’ for the funding of the research leading to the present publication. We especially thank Tom Kram for critical discussions about the developed methodology as well as Stefan Siebert and Felix Portmann for processing the MIRCA2000 data for our purposes.

References

Aigner, D., Lovell, C.A.K., Schmidt, P., 1977. Formulation and estimation of stochastic frontier production function models. Journal of Econometrics 6, 21-37.

Alcamo, J., Leemans, R., Kreileman, E., 1998. Global Change Scenarios of the 21st Century, Kidlington, Oxford, UK.

Alvarez, R., Grigera, S., 2005. Analysis of Soil Fertility and Management Effects on Yields of Wheat and Corn in the Rolling Pampa of Argentina. Journal of Agronomy and Crop Science 191, 321-329.

Barbier, E.B., 2003. Agricultural Expansion, Resource Booms and Growth in Latin America: Implications for Long-run Economic Development. World Development 32, 137-157.

Battese, G.E., 1992. Frontier production functions and technical efficiency: a survey of empirical applications in agricultural economics. Agricultural Economics 7, 185-208.

Battese, G.E., Coelli, T.J., 1995. A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data. Empirical Economics 20, 325-332

Battese, G.E., Broca, S.S., 1997. Functional Forms of Stochastic Frontier Production Functions and Models for Technical Inefficiency Effects: A Comparative Study for Wheat Farmers in Pakistan. Journal of Productivity Analysis 8, 395-414.

Bondeau, A., Smith, P.C., Zaehle, S., Schaphoff, S., Lucht, W., Cramer, W., Gerten, D., Lotze-Campen, H., Mueller, C., Reichstein, M., Smith, B., 2007. Modelling the role of agriculture for the 20th century global terrestrial carbon balance. Global Change Biology 13, 679–706.

Bravo-Ureta, B.E., Pinheiro, A.E., 1993. Efficiency Analysis of Developing Country Agriculture: A Review of the Frontier Function Literature. Agricultural and Resource Economics Review 22, 88-101.

Bravo-Ureta, B.E., Evenson, R.E., 1994. Efficiency in agricultural production: The case of peasant farmers in eastern Paraguay. Agricultural Economics 10, 27-37.

Buringh, P., H.D.J., v.H., Staring, G.J., 1975. Computation of the absolute maximum food production of the world. Department of Tropical Soil Science, University of Wageningen, Netherlands.

Casanova, D., Goudriaan, J., Bouma, J., Epema, G.F., 1999. Yield gap analysis in relation to soil properties in direct-seeded flooded rice. Geoderma 91, 191-216.

Cassman, K.G., 1999. Ecological intensification of cereal production systems: Yield potential, soil quality, and precision agricultureNational Academy of Sciences colloquium ‘‘Plants and Population: Is There Time?’’, Arnold and Mabel Beckman Center in Irvine, CA.

Chen, J., 2007. Rapid urbanization in China: A real challenge to soil protection and food security. CATENA 69, 1-15.

Coelli, T., Rao, P., O’Donnell, C.J., Battese, G.E., 2005. An Introduction to Efficiency and Productivity Analysis. Kluwer Academic Publishers, Norwell, Massachusetts.

Coelli, T.J., 1996. A Guide to Frontier Version 4.1: A Computer Program for Stochastic Frontier Production and Cost Function EstimationCEPA Working Paper. University of New England, Departement of Econometrics.

Coxhead, I., Shively, G., Shuai, X., 2002. Development policies, resource constraints, and agricultural expansion on the Philippine land frontier. Environment and Development Economics 7, 341- 363.

de Wit, C.T., 1992. Resource use efficiency in agriculture. Agricultural Systems 40, 125-151.

DeFries, R.S., Foley, J.A., Asner, G.P., 2004. Land-use choices: balancing human needs and ecosystem function. Frontiers in Ecology and the Environment 2, 249–257.

Donald, P.F., Green, R.E., Heath, M.F., 2001. Agricultural Intensification and the Collapse of Europe's Farmland Bird Populations. Biological Sciences 268, Pp. 25-29.

Ellis, E.C., Ramankutty, N., 2008. Putting people in the map: anthropogenic biomes of the world. Frontiers in Ecology and the Environment 6, 439-447.

Erda, L., Wei, X., Hui, J., Yinlong, X., Yue, L., Liping, B., Liyong, X., 2005. Climate change impacts on crop yield and quality with CO2 fertilization in China. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 360, 2149–2154.

Evans, L.T., Fischer, R.A., 1999. Yield Potential: Its Definition, Measurement, and Significance. Crop Science 39, Pp. 1544–1551.

Ewert, F., Rounsevell, M.D.A., Reginster, I., Metzger, M.J., Leemans, R., 2005. Future scenarios of European agricultural land use I. Estimating changes in crop productivity. Agriculture, Ecosystems and Environment 107, 101-116.

FAO, 2002a. Bread wheat Improvement and Production. In: Curtis, B.C., Rajaram, S., Macpherson, H.G. (Eds.), FAO Plant Production and Protection Series.

FAO, 2002b. Fertilizer use by crop. Food and Agriculture Organization of the United Nations (FAO), Rome.

Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., Coe, M.T., Daily, G.C., Gibbs, H.K., Helkowski, J.H., Holloway, T., Howard, E.A., Kucharik, C.J., Monfreda, C., Patz, J.A., Prentice, I.C., Ramankutty, N., Snyder, P.K., 2005. Global Consequences of Land Use. Science 309, 570-574.

Goodwin, B.K., Mishra, A.K., 2004. Farming Efficiency and the Determinants of Multiple Job Holding by Farm Operators. American Journal of Agricultural Economics 86, 722-729.

Harris, J.M., Kennedy, S., 1999. Carrying capacity in agriculture: global and regional issues. Ecological Economics 29, 443-461.

Hasnah, Fleming, E., Coelli, T., 2004. Assessing the performance of a nucleus estate and smallholder scheme for oil palm production in West Sumatra: a stochastic frontier analysis. Agricultural Systems 79, 17-30.

Huang, J., Pray, C., Rozelle, S., 2002. Enhancing the crops to feed the poor. Nature 418, 678-684.

Jones, P.G., Thornton, P.K., 2003. The potential impacts of climate change on maize production in Africa and Latin America in 2055. Global Environmental Change 13, 51-59.

Kaufmann, R.K., Snell, S.E., 1997. A Biophysical Model of Corn Yield: Integrating Climatic and Social Determinants. American Journal of Agricultural Economics 79, 178-190.

Kerr, T., Cihlar, J., 2003. Land use and cover with intensity of agriculture for Canada from satellite and census data. Global Ecology & Biogeography 12, 161-172.

Keys, E., McConnell, W.J., 2005. Global change and the intensification of agriculture in the tropics. Global Environmental Change 15, 320-337.

Kodde, D.A., Palm, F.Z., 1986. Wald Criteria for Jointly Testing Equality and Inequality Restrictions Econometrica 54, 1243-1248.

Kudaligama, V.P., Yanagida, J.F., 2000. A Comparison of Intercountry Agricultural Production Functions: A Frontier Function Approach. Journal of Economic Development 25, 57-74.

Lambin, E.F., Turner, B.L., Geist, H.J., Agbola, S.B., Angelsen, A., Bruce, J.W., Coomes, O.T., Dirzo, R., Fischer, G., Folke, C., George, P.S., Homewood, K., Imbernon, J., Leemans, R., Li, X., Moran, E.F., Mortimore, M., Ramakrishnan, P.S., Richards, J.F., Skanes, H., Steffen, W., Stone, G.D., Svedin, U., Veldkamp, T.A., Vogel, C., Xu, J., 2001. The causes of land-use and land-cover change: moving beyond the myths. Global Environmental Change 11, 261-269.

Long, S.P., Ainsworth, E.A., Leakey, A.D.B., Noesberger, J., Ort, D.R., 2006. Food for Thought: Lower-Than-Expected Crop Yield - Stimulation with Rising CO 2 Concentrations. Science 312, 1918 - 1921.

Matson, P.A., Parton, W.J., Power, A.G., Swift, M.J., 1997. Agricultural Intensification and Ecosystem Properties. Science 277, 504-509.

Meeusen, W., Broeck, J.v.D., 1977. Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error. International Economic Review 18, 435-444.

MNP (Ed.) 2006. Integrated modelling of global environmental change. An overview of IMAGE 2.4. Netherlands Environmental Agency (MNP), Bilthoven, The Netherlands.

Monfreda, C., Ramankutty, N., Foley, J.A., 2008. Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. Global Biogeochem. Cycles 22.

Mundlak, Y., Larson, D., Butzer, R., 1997. The determinants of agricultural production : a cross-country analysis. In: Bank, T.W. (Ed.), Policy Research Working Paper Series

Parry, M.L., Rosenzweig, C., Iglesias, A., Livermore, M., Fischer, G., 2004. Effects of climate change on global food production under SRES emissions and socio-economic scenarios. Global Environmental Change 14, 53-67.

Paul, C.J.M., Nehring, R., 2005. Product diversification, production systems, and economic performance in U.S. agricultural production. Journal of Econometrics 126, 525-548.

Peng, S., Cassman, K.G., Virmani, S.S., Sheehy, J., Khush, G.S., 1999. Yield Potential Trends of Tropical Rice since the Release of IR8 and the Challenge of Increasing Rice Yield Potential. Crop Sci 39, 1552-1559.

Pesaran, M.H., Schmidt, P., 1999. Handbook of Applied Econometrics. Blackwell Publishers.

Portmann, F., Siebert, S., Bauer, C., Döll, P., 2008. Global data set of monthly growing areas of 26 irrigated cropsFrankfurt Hydrology Paper 06. Institute of Physical Geography, University of Frankfurt, Frankfurt am Main, Germany.

Ramankutty, N., Evan, A.T., Monfreda, C., Foley, J.A., 2008. Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. Global Biogeochem. Cycles 22.

Reardon, T., Barrett, C., Kelly, V., Savadogo, K., 1999. Policy Reforms and Sustainable Agricultural Intensification in Africa. Development Policy Review 17, 375-395.

Reidsma, P., Ewert, F., Oude Lansink, A., 2007. Analysis of farm performance in Europe under different climatic and management conditions to improve understanding of adaptive capacity. Climatic Change 84, 403-422.

Reidsma, P., Ewert, F., Boogaard, H., Diepen, K.v., 2009a. Regional crop modelling in Europe: The impact of climatic conditions and farm characteristics on maize yields. Agricultural Systems 100, 51-60.

Reidsma, P., Oude Lansink, A., Ewert, F., 2009b. Economic impacts of climatic variability and subsidies on European agriculture and observed adaptation strategies. Mitigation and Adaptation Strategies for Global Change 14, 35-59.

Rockstroem, J., Falkenmark, M., 2000. Semiarid Crop Production from a Hydrological Perspective: Gap between Potential and Actual Yields. Critical Reviews in Plant Sciences 19, 319 - 346.

Rosegrant, M.W., Paisner, M.S., Meijer, S., Witcover, J., 2001. 2020 Global Food Outlook Trends,Alternatives, and Choices. International Food Policy Research Institute, Washington, DC.

Rosegrant, M.W., Cline, S.A., 2003. Global Food Security: Challenges and Policies. Science 302, 1917-1919.

Ruttan, V.W., 2002. Productivity Growth in World Agriculture: Sources and Constraints. Journal of Economic Perspectives 16, 161–184.

Stehfest, E., Heistermann, M., Priess, J.A., Ojima, D.S., Alcamo, J., 2007. Simulation of global crop production with the ecosystem model DayCent. Ecological Modelling 209, 203-219.

Swinnen, J.F.M., Gow, H.R., 1999. Agricultural credit problems and policies during the transition to a market economy in Central and Eastern Europe. Food Policy 24, 21-47.

Thornton, P.K., Jones, P.G., Alagarswamy, G., Andresen, J., 2009. Spatial variation of crop yield response to climate change in East Africa. Global Environmental Change 19, 54-65.

Tian, W., Wan, G.H., 2000. Technical Efficiency and Its Determinants in China’s Grain Production. Journal of Productivity Analysis 13, 159–174.

Tilman, D., 1999. Global environmental impacts of agricultural expansion: The need for sustainable and efficient practices. Proc. National Academy of Science. USA 96, 5995–6000.

Tilman, D., Cassman, K.G., Matson, P.A., Naylor, R., Polasky, S., 2002. Agricultural sustainability and intensive production practices. Nature 418.

Timsina, J., Connor, D.J., 2001. Productivity and management of rice-wheat cropping systems: issues and challenges. Field Crops Research 69, 93-132.

Trostle, R., 2008. Global Agricultural Supply and Demand: Factors Contributing to the Recent Increase in Food Commodity Prices. United States Department of Agriculture.

van Ittersum, M.K., Leffelaar, P.A., van Keulen, H., Kropff, M.J., Bastiaans, L., Goudriaan, J., 2003. On approaches and applications of the Wageningen crop models. European Journal of Agronomy 18, 201-234.

Verburg, P.H., Chen, Y., Veldkamp, T.A., 2000. Spatial explorations of land use change and grain production in China. Agriculture, Ecosystems and Environment 82, 333–354.

von Thuenen, J.H., 1966. Von Thuenen’s Isolated State. New York Pergamon Press.

Wei, X., Declan, C., Erda, L., Yinlong, X., Hui, J., Jinhe, J., Ian, H., Yan, L., 2009. Future cereal production in China: The interaction of climate change, water availability and socio-economic scenarios. Global Environmental Change 19, 34-44.

Wood, S., Sebastian, K., Scherr, S.J., 2000. Pilot analysis of global ecosystems: agroecosystems International Food Policy Research Institute and World Resources Institute, Washington, DC.

Wu, D., Yu, Q., Lu, C., Hengsdijk, H., 2006. Quantifying production potentials of winter wheat in the North China Plain. European Journal of Agronomy 24, 226-235.

Xin, L., Fan, Y.Z., Tan, M.H., Jiang, L.G., 2009. Review of Arable Land-use Problems in Present-day China. AMBIO: A Journal of the Human Environment 38, 112-115.

Tables

Table 1: Variables used in the efficiency analysis.

Variable	Definition (measure)	Source
Actual yield		
Grain	Yield of wheat, maize and rice (scale)	(Monfreda *et al.*, 2008) and SAGE (http://www.sage.wisc.edu/mapsdatamodels.html)
Frontier production function		
Temp	Deviation from optimal monthly mean temperature for grain specific growing period (scale)	Average for 1950-2000 derived from Worldclim (www.worldclim.org) with growing period information from Portmann *et al.* (2008) and LPJmL (Bondeau *et al.*, 2007)
Precip	Precipitation sum for grain specific growing period (scale)	Average for 1950-2000 derived from Worldclim (www.worldclim.org) with growing period information from Portmann and Siebert (2008) and LPJmL (Bondeau *et al.*, 2007)
Par	Photosynthetically Active Radiation (PAR) sum for grain specific growing period (scale)	Computed as described by Haxeltine and Prentice (1996)
Soil_const	Soil fertility constraints (ordinal)	Global Agro-Ecological Zones – 2000 (http://www.iiasa.ac.at/Research/LUC/GAEZ)
Inefficiency function		
Irrig	Maximum monthly growing area per irrigated grain type (scale)	MIRCA 2000 (http://www.geo.uni-frankfurt.de/ipg/ag/dl/forschung/MIRCA/index.html)
Slope	Slope (ordinal)	Global Agro-Ecological Zones – 2000 (http://www.iiasa.ac.at/Research/LUC/GAEZ)
Agr_pop	Non-urban population density as ratio of population density (below 2500 persons per km^2) and agricultural area (scale)	(Ellis and Ramankutty, 2008)
Access	Market accessibility (scale)	Derived from UNEP major urban agglomerations dataset (http://geodata.grid.unep.ch) and the Global Maritime Ports Database (http://www.fao.org/geonetwork/srv/en/main.home)
Market	Market influence (index)	Purchasing Power Parity (PPP) per country derived from CIA factbook (https://www.cia.gov/library/publications/the-world-factbook) spatially distributed through an inverse relation with variable *access*

Table 2: Coefficients for the parameters of the stochastic frontier production function at the global scale (significant at 0.05 level).

		Wheat		**Maize**		**Rice**	
Variable	*Parameter*	*coefficient**	*t-ratio*	*coefficient**	*t-ratio*	*coefficient**	*t-ratio*
Frontier production function							
Constant	$ß_0$	0.98	9.2	3.05	18.3	10.08	22.7
Ln(*temp)*	$ß_1$	-0.18	-31.8	-0.03	-19.8	-0.02	-12.4
Ln *(precip)*	$ß_2$	0.17	22.6	0.07	9.9	0.05	11.7
Ln *(par)*	$ß_3$	-0.17	-11.3	-0.24	-9.9	-0.42	-20.0
Ln *(soil_const)*	$ß_4$	0.09	14.0	-0.21	-23.3	-0.11	-10.5
Inefficiency function							
Irrig	δ_1	<-0.01	-10.1	<-0.01	-28.7	<-0.01	-20.0
Slope	δ_2	0.17	53.4	0.20	35.9	-0.05	-5.2
Agr_pop	δ_3	<-0.01	-19.7	<0.01	10.7	<0.01	7.2
Access	δ_4	0.02	14.0	0.01	6.2	0.01	5.4
Market	δ_5	<-0.01	-33.3	<-0.01	-54.8	<-0.01	-29.8
Variance parameters							
Sigma-squared	σ^2	0.26	79.0	0.82	41.7	0.80	37.4
Gamma	γ	0.47	48.1	0.91	166.3	0.91	134.4
Log-likelihood		-8411		-9350		-5356	
Likelihood ratio statistic (LR)		4307		3695		1558	
Mean efficiency		0.64		0.50		0.64	

* A positive coefficient in the frontier production function indicates that the respective variable has a positive influence on the frontier yield. A positive coefficient in the inefficiency function indicates that the respective variable has a negative influence on efficiency.

Table 3: Multiple linear regression results for efficiencies of wheat, maize, and rice production for selected world-regions.

	Unstandardized Coefficients[a]		**Standardized Coefficients**[a]
	B	*Std. Error*	*Beta*
Wheat	*USA ($r^2 = 0.25$)*		
(Constant)	-2.2×10^{-1}	2.1×10^{-3}	
Irrig	8.2×10^{-5}	6.2×10^{-6}	2.8×10^{-1}
Slope	*	*	*
Agr_pop	1.0×10^{-4}	3.6×10^{-5}	6.0×10^{-2}
Access	-5.2×10^{-3}	3.3×10^{-4}	-3.5×10^{-1}
Market	*	*	*
Wheat	*China ($r^2 = 0.38$)*		
(Constant)	-1.9×10^{-1}	4.9×10^{-3}	
Irrig	1.2×10^{-5}	1.2×10^{-6}	2.2×10^{-1}
Slope	-1.0×10^{-1}	8.6×10^{-4}	-3.6×10^{-1}
Agr_pop	3.8×10^{-5}	8.0×10^{-6}	1.1×10^{-1}
Access	*	*	*
Market	8.9×10^{-6}	1.7×10^{-6}	1.1×10^{-1}
Maize	*Mexico ($r^2 = 0.10$)*		
(Constant)	-8.1×10^{-1}	5.0×10^{-1}	
Irrig	1.1×10^{-4}	2.5×10^{-4}	1.9×10^{-1}
Slope	2.0×10^{-2}	1.0×10^{-2}	9.0×10^{-2}
Agr_pop	2.3×10^{-4}	1.0×10^{-4}	1.0×10^{-1}
Access	*	*	*
Market	2.4×10^{-5}	6.1×10^{-6}	1.7×10^{-1}
Maize	*Southern Africa[b] ($r^2 = 0.22$)*		
(Constant)	-7.7×10^{-1}	4.0×10^{-2}	
Irrig	*	*	*
Slope	*	*	*
Agr_pop	-3.7×10^{-4}	1.8×10^{-4}	-7.0×10^{-2}
Access	-2.0×10^{-2}	4.0×10^{-3}	-1.6×10^{-1}
Market	8.6×10^{-5}	1.1×10^{-5}	3.4×10^{-1}
Rice	*Thailand region[c] ($r^2 = 0.21$)*		
(Constant)	-7.5×10^{-1}	2.0×10^{-2}	
Irrig	7.0×10^{-5}	4.6×10^{-6}	4.2×10^{-1}
Slope	2.0×10^{-2}	4.5×10^{-3}	1.2×10^{-1}
Agr_pop	2.6×10^{-4}	5.0×10^{-5}	1.4×10^{-1}
Access	-2.0×10^{-3}	6.6×10^{-4}	-9.0×10^{-1}
Market	*	*	*
Rice	*Indonesia ($r^2 = 0.28$)*		
(Constant)	-4.6×10^{-1}	2.0×10^{-1}	
Irrig	1.4×10^{-5}	3.4×10^{-6}	1.6×10^{-1}
Slope	1.0×10^{-1}	3.2×10^{-3}	1.1×10^{-1}
Agr_pop	6.2×10^{-5}	1.7×10^{-5}	1.6×10^{-1}
Access	-1.6×10^{-3}	3.8×10^{-4}	-1.6×10^{-1}
Market	5.5×10^{-5}	1.1×10^{-5}	2.3×10^{-1}

* Not significant at 0.05 level

[a] A positive coefficient indicates that the respective variable has a positive influence on efficiency.

[b] Includes South Africa, Lesotho, Mozambique, Zimbabwe, Tanzania, Zambia, Malawi, Angola, Namibia, Botswana, and Swaziland

[c] Includes Vietnam, Philippines, Cambodia, Burma, Laos, and Malaysia

Figure captions

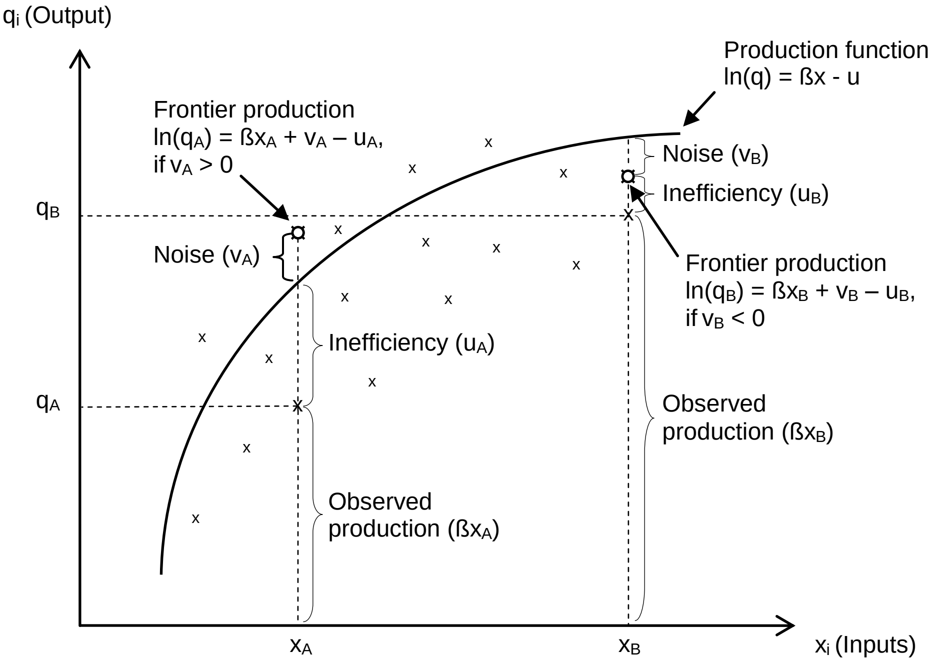
Figure 1: The stochastic Production Frontier (after Coelli et al., 2005). Observed productions are indicated with x while frontier productions are indicated with ¤. The frontier function is based on the highest observed outputs under the inputs accounting for random noise (v_i). Further deviations of the observations are due to inefficiencies (u_i). The frontier production q_i can lie above or below the frontier production function, depending on the noise effect (v_i).

Figure 2: Actual and frontier yields for wheat, maize, and rice.

Figure 3: Global yield gap for wheat (Map 1), maize (Map 2) and rice (Map 3) calculated as the difference between actual yield and estimated frontier yield.

Figure 4: Efficiencies of wheat (Map 1), maize (Map 2) and rice (Map 3) production with the most determining factors per world-region.

Figure 1



q_i (Output)
Production function
$\ln(q) = ßx - u$
Frontier production
$\ln(q_A) = ßx_A + v_A - u_A$,
if $v_A > 0$
Noise (v_B)
Inefficiency (u_B)
q_B
Noise (v_A)
Frontier production
$\ln(q_B) = ßx_B + v_B - u_B$,
if $v_B < 0$
Inefficiency (u_A)
q_A
Observed production ($ßx_B$)
Observed production ($ßx_A$)
x_A
x_B
x_i (Inputs)

Figure 2

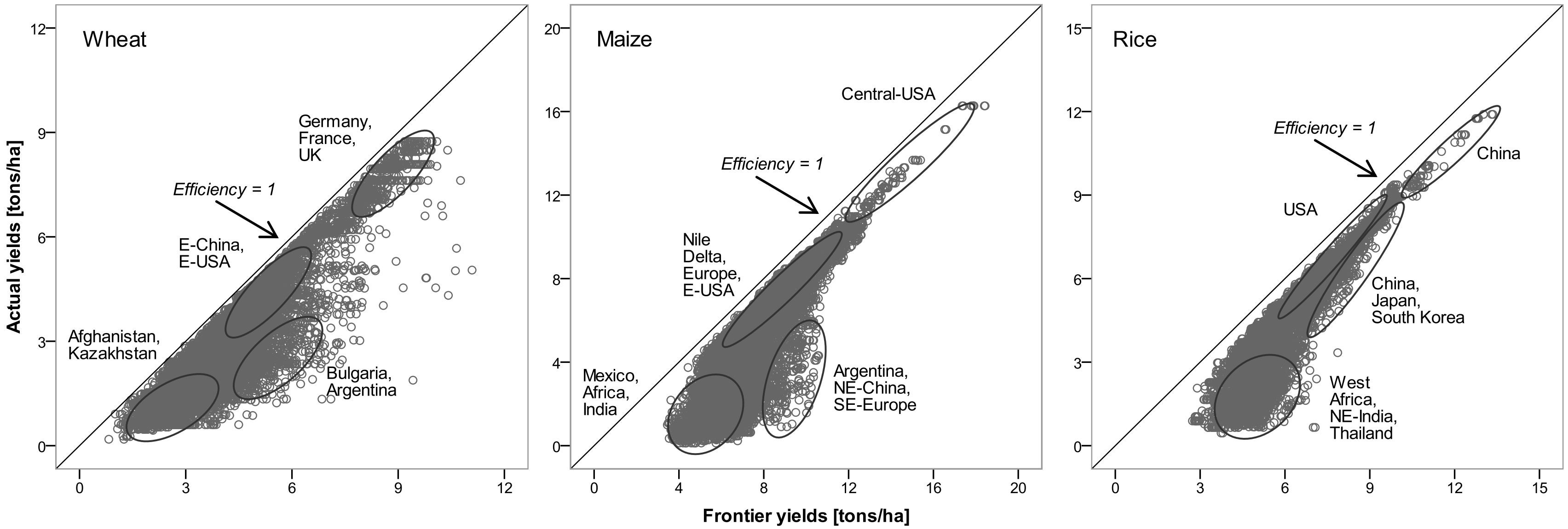

Wheat
Germany,
France,
UK
Efficiency = 1
E-China,
E-USA
Afghanistan,
Kazakhstan
Bulgaria,
Argentina
Maize
Central-USA
Efficiency = 1
Nile
Delta,
Europe,
E-USA
Mexico,
Africa,
India
Argentina,
NE-China,
SE-Europe
Rice
Efficiency = 1
China
USA
China,
Japan,
South Korea
West
Africa,
NE-India,
Thailand
Actual yields [tons/ha]
Frontier yields [tons/ha]
0
3
6
9
12
0
4
8
12
16
20
0
3
6
9
12
15

Figure 3 (Map1)

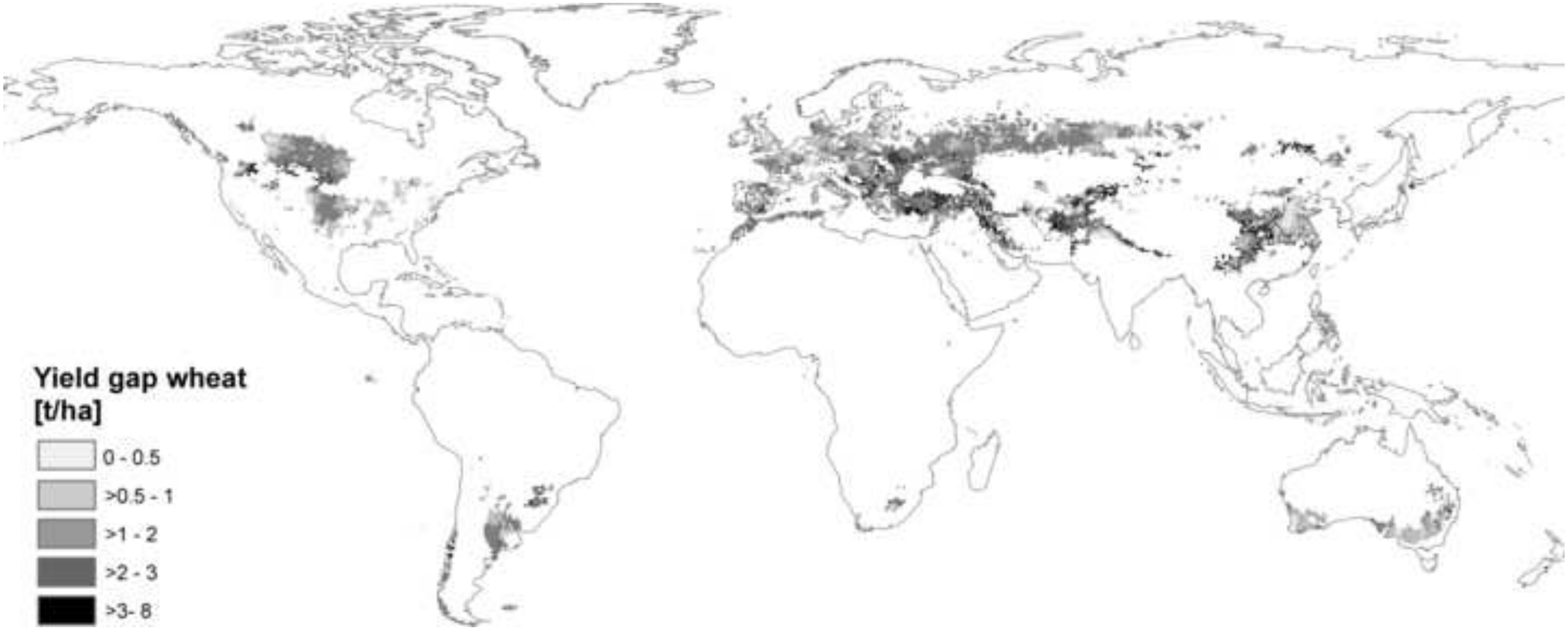


Figure 3 (Map2)

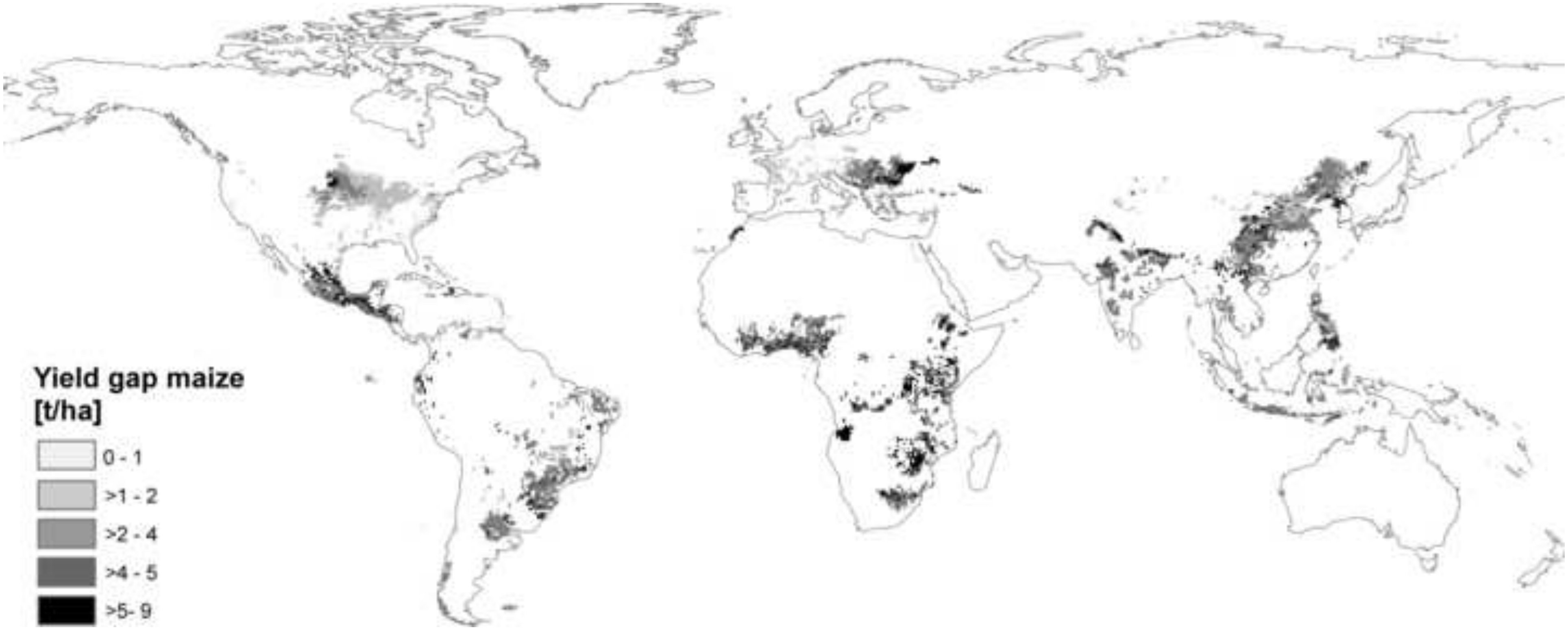


Figure 3 (Map3)

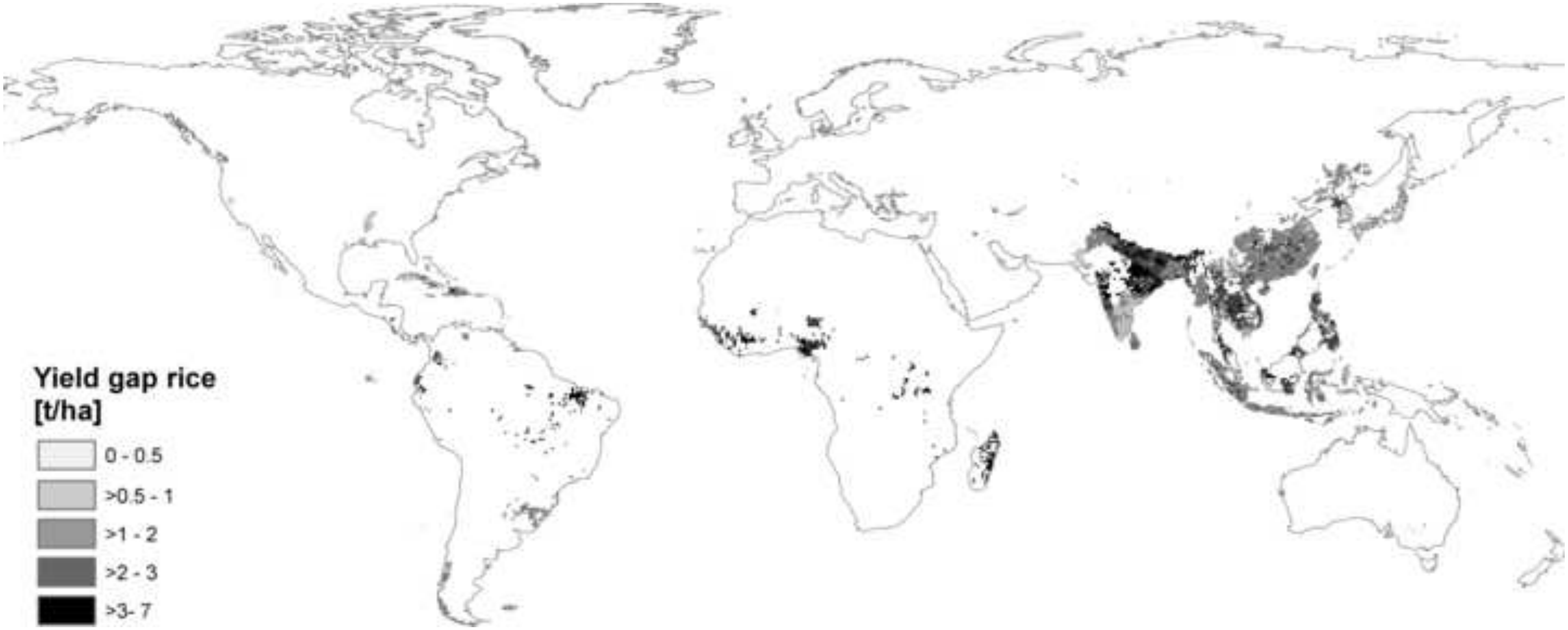


Figure 4 (Map1)

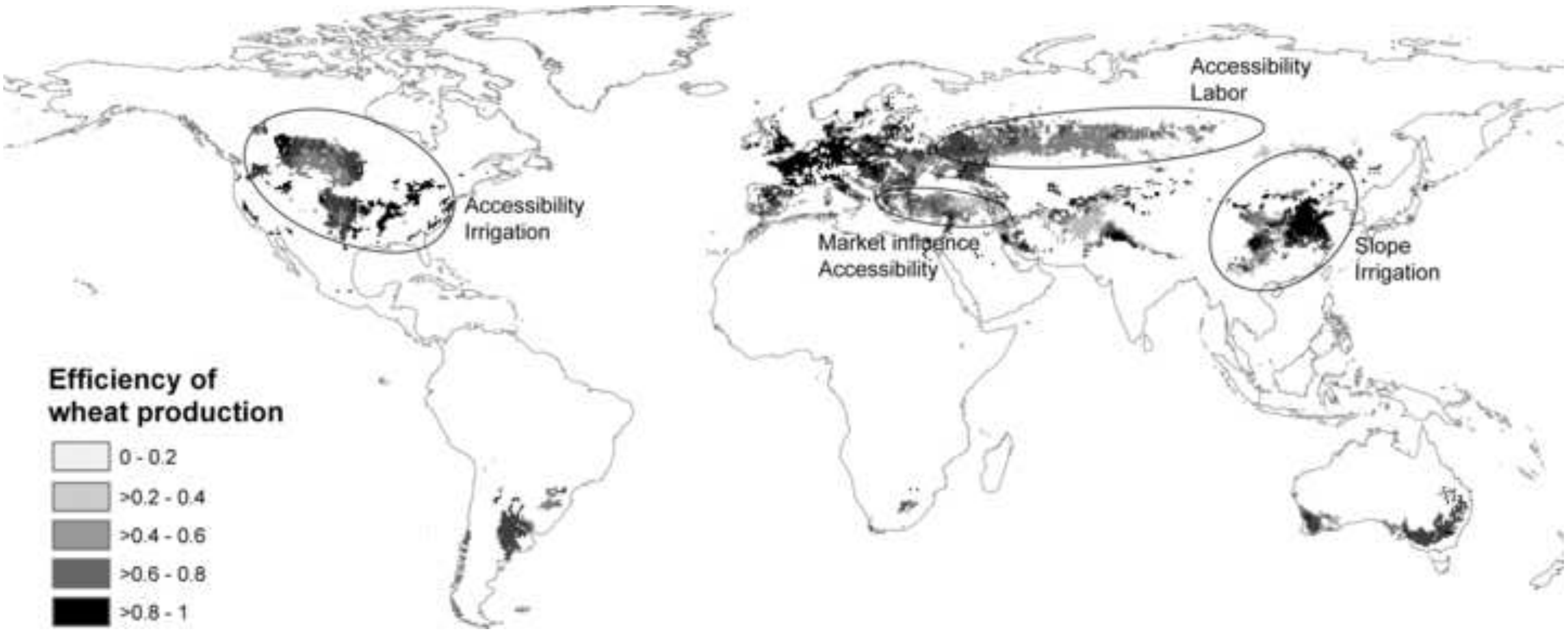


Figure 4 (Map2)

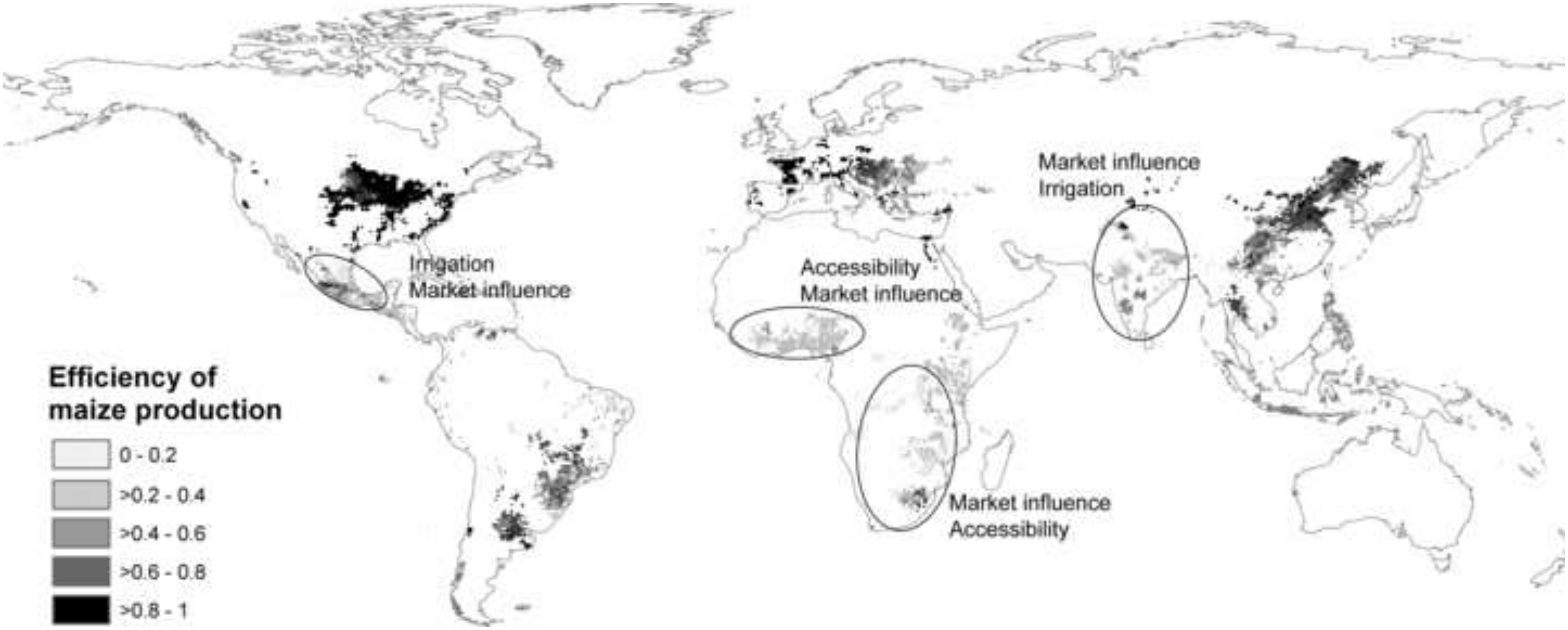


Figure 4 (Map3)