Field Crops Research 177 (2015) 49–63

Contents lists available at ScienceDirect

Field Crops Research

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

How good is good enough? Data requirements for reliable crop yield simulations and yield-gap analysis

Patricio Grassini[a,*], Lenny G.J. van Bussel[b], Justin Van Wart[a], Joost Wolf[b], Lieven Claessens[c,d], Haishun Yang[a], Hendrik Boogaard[e], Hugo de Groot[e], Martin K. van Ittersum[b], Kenneth G. Cassman[a]

[a] University of Nebraska-Lincoln, PO Box 830915, Lincoln, NE 68583-0915, USA
[b] Plant Production Systems Group, Wageningen University, PO Box 430, 6700 AK Wageningen, The Netherlands
[c] International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), PO Box 39063, 00623 Nairobi, Kenya
[d] Soil Geography and Landscape Group, Wageningen University, PO Box 47, 6700 AA Wageningen, The Netherlands
[e] Alterra, Wageningen University and Research Centre, PO Box 47, 6700 AA Wageningen, The Netherlands

ARTICLE INFO

Article history:
Received 19 December 2014
Received in revised form 7 March 2015
Accepted 8 March 2015
Available online 30 March 2015

Keywords:
Crop simulation
Yield gap
Yield potential
Weather data
Cropping system

ABSTRACT

Numerous studies have been published during the past two decades that use simulation models to assess crop yield gaps (quantified as the difference between potential and actual farm yields), impact of climate change on future crop yields, and land-use change. However, there is a wide range in quality and spatial and temporal scale and resolution of climate and soil data underpinning these studies, as well as widely differing assumptions about cropping-system context and crop model calibration. Here we present an explicit rationale and methodology for selecting data sources for simulating crop yields and estimating yield gaps at specific locations that can be applied across widely different levels of data availability and quality. The method consists of a tiered approach that identifies the most scientifically robust requirements for data availability and quality, as well as other, less rigorous options when data are not available or are of poor quality. Examples are given using this approach to estimate maize yield gaps in the state of Nebraska (USA), and at a national scale for Argentina and Kenya. These examples were selected to represent contrasting scenarios of data availability and quality for the variables used to estimate yield gaps. The goal of the proposed methods is to provide transparent, reproducible, and scientifically robust guidelines for estimating yield gaps; guidelines which are also relevant for simulating the impact of climate change and land-use change at local to global spatial scales. Likewise, the improved understanding of data requirements and alternatives for simulating crop yields and estimating yield gaps as described here can help identify the most critical "data gaps" and focus global efforts to fill them. A related paper (Van Bussel et al., 2015) examines issues of site selection to minimize data requirements and up-scaling from location-specific estimates to regional and national spatial scales.

1. Introduction

Yield potential (Yp) is defined as the yield of an adapted crop cultivar as determined by solar radiation, temperature, carbon dioxide, and genetic traits that govern length of growing period, light interception by the crop canopy and its conversion to biomass, and partition of biomass to the harvestable organs (Evans, 1993; van Ittersum and Rabbinge, 1997). Water-limited yield potential (Yw) is determined by these previous factors and also by water supply amount and distribution during the crop growth period and field and soil properties that affect soil water availability such as slope, plant-available soil water holding capacity, and depth of the root zone (Lobell et al., 2009; van Ittersum and Rabbinge, 1997; Van Ittersum et al., 2013). For a specific location and year, the crop yield gap (Yg) is defined as the difference between Yp (irrigated systems) or Yw (rainfed) and average actual farm yield (Ya). The magnitude of Yg provides a benchmark of current land productivity in relation to the biophysical yield ceiling, and an estimate of the additional crop production that could potentially be achieved, on existing cropland area, through improved management that alleviates all limiting factors other than weather factors. Estimates of Yp,

* Corresponding author. Tel.: +1 402 472 5554; fax: +1 402 472 7904.
E-mail addresses: pgrassini2@unl.edu (P. Grassini), Lenny.vanBussel@wur.nl (L.G.J. van Bussel).

http://dx.doi.org/10.1016/j.fcr.2015.03.004
0378-4290/

Yw, and Yg also provide the foundation for more detailed studies to identify underpinning causes of the observed Yg, and for *ex-ante* evaluation of impact from adoption of new technologies, changing climate, and land-use change.

Accuracy in Yg estimation depends on the error associated with estimates of Yp (or Yw) and Ya[1]. Amongst methods to estimate Yp or Yw, crop simulation models provide the most robust approach because they account for the interactive effects of genotype, weather, and management (GxExM) on yields across agro-ecological zones and years (Van Ittersum et al., 2013). To minimize errors in estimating Yp and Yw, crop simulation models require high-quality input-data on weather, soil, and crop management (Aggarwal, 1995; Rivington et al., 2005; Bert et al., 2007). These models need also to be rigorously evaluated for their ability to reproduce major GxExM interactions (Passioura, 1996; Kersebaum et al., 2007; Van Ittersum et al., 2013). Likewise, reliable simulation of Yp and Yw requires specification of the cropping system and water regime in which a crop is grown as determined by crop sequence, dates of sowing and physiological maturity for the most widely used cultivars, and whether the crop is fully irrigated, partially irrigated, or rainfed (Folberth et al., 2012; Van Wart et al., 2013c). Finally, the error associated with the estimate of average annual Ya will also determine the accuracy of the Yg estimate.

Crop yield simulation is an important component of yield-gap analysis, hence, the above-mentioned sources of uncertainty related with estimates of Yp (or Yw) also affect other kinds of studies that rely on crop yield simulations and the required data therein. For example, studies on climate change, and land use change involving crop simulation models applied at global or regional spatial scales are abundant in recent literature (*e.g.*, Challinor et al., 2014a; Rosenzweig et al., 2014). However, several recent publications have identified a number of substantive concerns associated with data sources and methods used in such studies (Van Ittersum et al., 2013; Van Wart et al., 2013a). These concerns include: (i) poor quality of weather and soil data, (ii) unrealistic assumptions about the cropping-system context, (iii) poorly calibrated crop simulation models, and (iv) lack of transparency about underpinning assumptions and methods. For example, Nelson et al. (2010) used 50-y monthly average gridded (5′ resolution) weather data and coarse assumptions about the cropping system (*e.g.*, a single crop variety was simulated for the entire world) to produce a global assessment of climate change impact on crop yields and land-use change. A similar approach was followed by Bagley et al. (2012) to simulate changes in water availability and potential crop yields in the world's breadbaskets. In both studies, no information was provided about how models were calibrated to simulate yield potential. Similarly, Rosenzweig et al. (2014) used an ensemble of models to simulate crop yields based on gridded daily weather data, coarse assumptions about cropping systems, and crop model parameters that were forced to reproduce current regional or national Ya averages. Another pitfall of these three studies is failure to account for multiple-crop systems (*i.e.*, fields planted with more than one crop in the same year, such as the rice-wheat system that is widely practiced in Asia) or cropping systems where irrigated and rainfed systems co-exist within the same geographic area.

In most cases, use of poor quality or coarse-scale weather, soil, and cropping-system data for yield-gap analysis, as well as for other studies on climate change, food security, and land-use change that rely on crop yield simulations, is due to the fact that high quality data at finer spatial resolution do not exist, so pragmatic short-cuts are required to achieve the full terrestrial coverage. These short-cuts, however, are rarely evaluated for their ability to reproduce Yp, Yw and Yg values estimated using high-quality, measured data. Without such validation, Yp, Yw, and Yg estimates with coarse-scale data sources can seriously distort results, decreasing their usefulness to inform regional or national policies and effective prioritization of research and development investments for agriculture (Rivington et al., 2004; Van Wart et al., 2013a,c). In contrast, one can find studies on yield-gap analysis for specific locations with data that are only available for few and specific site-years, which are not representative of larger spatial areas and do not allow upscaling to regional or global levels (*e.g.*, Fermont et al., 2009; Grassini et al., 2011). Surprisingly, despite wide use of crop simulation models for yield-gap analysis (263 results in the Web of Science by Nov 15th, 2014), there are no published guidelines about standard sources and quality of data input for weather, soil, actual yields, and cropping-system context, or requirements for calibration of crop models used in such studies.

In summary, a robust approach to simulate accurate crop yield potential and estimate Yg requires: (i) input data that meet minimum quality standards at the appropriate spatial scale, (ii) agronomic relevance with regard to cropping-system context, (iii) proper calibration of crop models used, and (iv) flexibility and transparency to account for different scenarios of data availability and quality. Here we address the current lack of guidelines on data and methods for yield gap analysis, by developing a systematic approach for selection of data inputs based on the lessons learned from establishing the Global Yield Gap Atlas (www.yieldgap.org). The paper focusses on yield-gap analysis at specific 'point' locations, and their surrounding inference zone, based on application of crop simulation models to estimate Yp or Yw (hereafter called 'targeted areas'). An inference zone is defined as an area with similar climate such that there is relatively little variation in crop management practices. This paper has implications not only for yield-gap analysis but also for other studies related with climate change, food security, and land-use change because these studies typically rely on crop yield simulations and the required data therein. A separate paper describes the methodology for site selection, spatial delimitation of the inference zone around a location, and up-scaling local estimates of Yg to regional and national scales (Van Bussel et al., 2015).

2. Data requirements for yield-gap analysis

2.1. Overview

Yield-gap analyses at large spatial scale require enormous amounts of input data, because simulated and actual crop yields are strongly determined by the spatial and temporal variation in environmental conditions and cropping system context. Based on the concept that it is better to use primary data for crop growth simulations than to use aggregated or interpolated average input data (De Wit and Van Keulen, 1987; Rabbinge and van Ittersum, 1994; Penning De Vries et al., 1997), the Global Yield Gap Atlas (www.yieldgap.org) utilizes a 'bottom-up' approach for yield-gap analysis. A limited number of locations are selected such that these account for the greatest proportion of total national production of the crop being evaluated. For these locations, 'point-based' estimates of Yp, Yw, Ya, and Yg are derived, which are subsequently up-scaled to climate zones and national spatial scales (Van Wart et al., 2013b; Van Bussel et al., 2015). This site selection and up-scaling process helps to limit the number of locations for which site-specific data on weather, soils, and cropping system are required, which in turn facilitates the focus on quality of the underpinning data and helps ensure local to global relevance of the analysis. Principles that underpin the data selection approach

[1] Accuracy is the closeness of a measurement (or simulation) to the true value.