Geoderma 318 (2018) 167–181

Contents lists available at ScienceDirect

Geoderma

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

Key functional soil types explain data aggregation effects on simulated yield, soil carbon, drainage and nitrogen leaching at a regional scale

Elsa Coucheney[a,*], Henrik Eckersten[b], Holger Hoffmann[c], Per-Erik Jansson[d], Thomas Gaiser[c], Franck Ewert[c,e], Elisabet Lewan[a]

[a] Department of Soil and Environment, Swedish University of Agricultural Sciences, Lennart Hjelms väg 9, Uppsala, SE, Sweden
[b] Department of Crop Production Ecology, Swedish University of Agricultural Sciences, Uppsala, Sweden, Germany
[c] Crop Science Group, INRES, University of Bonn, Katzenburgweg 5, 53115 Bonn, Germany
[d] Department of Land and Water Resources Engineering, KTH-Royal Institute of Technology, Stockholm, SE, Sweden
[e] Leibniz Centre for Agricultural Landscape Research (ZALF), Eberswalder Str. 84, 15374 Müncheberg, Germany

A R T I C L E I N F O

Editor: A.B. McBratney

Keywords:
Soil properties
Soil-crop models
Input data aggregation
Sensitivity analysis
Soil spatial pattern
Upscaling

A B S T R A C T

The effects of aggregating soil data (DAE) by areal majority of soil mapping units was explored for regional simulations with the soil-vegetation model CoupModel for a region in Germany (North Rhine-Westphalia). DAE were analysed for wheat yield, drainage, soil carbon mineralisation and nitrogen leaching below the root zone. DAE were higher for soil C mineralization and N leaching than for yield and drainage and were strongly related to the presence of specific soils within the study region. These soil types were associated to extreme simulated output variables compared to the mean variable in the region. The spatial aggregation of these key functional soils within sub-regions additionally influenced the DAE. A spatial analysis of their spatial pattern (i.e. their presence/absence, coverage and aggregation) can help in defining the appropriate grid resolution that would minimize the error caused by aggregating soil input data in regional simulations.

1. Introduction

Modelling agricultural production and adaptation to the environment at regional or global scales is receiving much interest in the context of a growing food demand (Tilman et al., 2011) and climate change (Ewert et al., 2015). Two important issues are to identify areas with high yield potential (van Wart et al., 2013) and sustainable management practices that would minimize environmental impacts (e.g. soil degradation, GHGs emissions, nutrient leaching). Process-based soil-crop models describe the flows of mass and energy in the soil-plant-atmosphere system and have been applied and tested in many different contexts, e.g. for CoupModel (Jansson, 2012), STICS (Coucheney et al., 2015), APSIM (Zhang et al., 2012), DNDC (Giltrap et al., 2010), DANUBIA (Lenz-Wiedemann et al., 2010) or CERES, WOFOST, CropSyst, WARM, and SWAP (Confalonieri et al., 2009). As such, they represent valuable tools for predicting agricultural production in diverse agro-environmental contexts (e.g. Jeuffroy et al., 2014) as well as for assessing impacts on the environment; e.g. leaching of nitrates (Conrad and Fohrer, 2009a), changes in soil carbon (Gervois et al., 2008) and GHGs emissions (De Gryze et al., 2011). They are also used to make predictions in response to climate change (e.g. Tubiello et al., 2000) and management changes (e.g. Ng et al., 2000) at the small plot or field scales where input data are considered to be spatially homogeneous. In this context, they are also increasingly applied at regional (e.g. Gaiser et al., 2009) and global scales (Rosenzweig et al., 2014). This raises new challenges related to model input data, calibration and evaluation and the use of different methods of upscaling and downscaling adds new sources of modelling uncertainties (Ewert et al., 2011).

In regional-scale modelling, one major concern is the need to take into account the spatial variability of the environmental conditions (e.g. climate, soils, management practices) used as model inputs. Previous studies showed the effects of input data quantity and quality on model predictions (Grassini et al., 2015) or evaluated model predictions when applied in diverse agro-environmental conditions (Balkovič et al., 2013; Coucheney et al., 2015). Other recent studies have assessed the errors caused by upscaling methods for a range of agro-environmental contexts and models (Hoffmann et al., 2016a; Kuhnert et al., 2016; Van Bussel et al., 2011; Zhao et al., 2015a). These studies are a step further towards the identification and development of scaling methods that minimize these errors for particular climate, soil and management conditions.

* Corresponding author.
E-mail address: elsa.coucheney@slu.se (E. Coucheney).

https://doi.org/10.1016/j.geoderma.2017.11.025
Received 13 July 2017; Received in revised form 17 November 2017; Accepted 20 November 2017
Available online 13 January 2018

Ewert et al. (2011) present different scaling methods from which two main groups can be distinguished: those based on aggregating model outputs (Zhao et al., 2016) and those based on aggregating inputs. In the latter case, the model is usually run for equal-sized grid cells covering the region at a pre-defined resolution (Angulo et al., 2013, 2014; de Wit et al., 2005; Folberth et al., 2012; Hoffmann et al., 2015; Hoffmann et al., 2016a, 2016b; Jégo et al., 2015; Kuhnert et al., 2016; Zhao et al., 2015a). The challenge is then to generate input data for each grid cell (see Zhao et al., 2015b) from data sources available at finer ('upscaling') or coarser ('downscaling') resolutions and which result in the smallest errors in comparison to simulations carried out for each single cell of the grid. Studies of this type have focused mainly on the effects of upscaling climate and soil input data on the predictions for a range of models.

While uncertainties in radiation and precipitation are recognised as a major source of uncertainty for crop yield prediction, de Wit et al. (2005) concluded that average unbiased estimates of weather data are sufficient for predicting yield at the regional scale. Similarly, Angulo et al. (2013, 2014) found that aggregating climate and soil input data from 10 to 100 km had a small effect on barley yields in the south of Finland simulated with four different crop models. Limited data aggregation effects were also obtained when aggregating climate input for simulations of yield and net primary productivity of wheat and maize under water-limited conditions in western Germany using several crop models (Hoffmann et al., 2015; Kuhnert et al., 2016; Zhao et al., 2015a). The DAE related to soil input data were larger in the same region (Hoffmann et al., 2016a). Olesen et al. (2000) and Jégo et al. (2015) found that fine resolution data for both precipitation and soil properties was important for predictions of representative wheat and maize yields in Denmark and southern Quebec respectively. In the latter region, one reason for this was a high pedo-diversity and a strong spatial correlation between soil type and the rainfall data used in the model simulations. DAE are therefore highly dependent on the region. The small effect of weather data aggregation in the German study regions might be due to limited spatial variation in climate combined with a moderate response of modelled yield to those variations (Angulo et al., 2013; Zhao et al., 2015a). Zhao et al. (2015a) further showed that the DAE due to climate aggregation differed among output variables, being positively correlated with the spatial heterogeneity (variation) of the variables concerned. Zhao et al. (2015b) concluded that a high spatial resolution of climate data is desired for regions with high environmental heterogeneity. In line with this, Kuhnert et al. (2016) found a higher impact of climate aggregation on simulated NPP for single years when extreme events such as drought occurred, because model outputs varied spatially more in these years compared with long-term average values. In addition, Hoffmann et al. (2015) showed that the DAE differed among models and identified the climate variables that had the most influence. This study highlighted the importance of also considering the sensitivity of model outputs to the input data (e.g. Hoffmann et al., 2016a). The question arises as to whether the DAE for a specific model and given output variables can be predicted given the spatial distribution of input data and the model sensitivity to the input. For example, variation in altitude was found to be a good proxy of climate DAE for the LINTUL-SIMPLACE model applied in Germany (Zhao et al., 2015b).

Soil properties such as texture, bulk density, porosity and organic matter content strongly influence the soil hydraulic properties (Schaap and Leij, 1998) and biogeochemical processes (Riffaldi et al., 1996). These impact crop growth and its sensitivity to climate under limiting water and nutrient conditions significantly (Kravchenko and Bullock, 2000). Soil variability is therefore one of the most important factors underlying spatial variability in crop yields (Wassenaar et al., 1999). Soil data aggregation may also have a strong impact on simulations of other variables apart from crop yield, such as soil organic carbon stocks (Zhang et al., 2014) or water and N dynamics (Kersebaum and Wenkel, 1998). Furthermore, the importance of soil properties for yield predictions depends not only on the climate (e.g. Timlin et al., 1998) but also on management practices such as fertilization. For example, Folberth et al. (2016) showed that the variability in yield related to soil type may exceed weather-related variability in scenarios characterized by low fertilization and irrigation amounts. DAE may therefore be larger in regions with a high pedodiversity (i.e. high spatial variation in soil types; Jégo et al., 2015) and in climates or under certain management practices that lead to high water and nitrogen stress.

The present study explores soil DAE on simulated yields of winter wheat, drainage, N leaching and C mineralisation in the region of North Rhine-Westphalia (NRW) in western Germany (Hoffmann et al., 2016a). Simulations were run with a process oriented soil-vegetation model (the CoupModel; Jansson, 2012) for gridded soil data and a spatially uniform climate to ensure that the spatial variability in outputs is related only to variation in soil properties. The soil data was aggregated by selecting the dominant soil at each coarser resolution (i.e. the soil mapping unit covering the areal majority is selected at each coarser scale). The objective of this study was to investigate the contribution of specific soils and their spatial distribution to the DAE for the selected model outputs. We hypothesize that specific combinations of soil properties ('*key soils*') generate extreme model outputs and that the spatial distribution (e.g. coverage, spatial aggregation) of these critical soils within the region strongly influences the spatial variation of the model outputs and therefore the DAE. To test this hypothesis we propose and apply an approximation of the DAE as a function of these *key soils* and their spatial coverage in the region. In addition, the influence of the degree of spatial aggregation of these key soils within an area was investigated with respect to four sub-areas of NRW.

2. Material and methods

2.1. Study area

The study area, the state of North Rhine-Westphalia (NRW, 6 E–9.5 E, 50 N–52.5 N, Fig. 1), is located in west-central Germany with a temperate humid climate. Half of the region (34,098 km^2) is covered by flat plains and the topography rises from the northeast towards the southeast with a maximal elevation of 843 m. Agricultural land represents > 60% of the area, with winter wheat and silage maize as the main crops.

The whole climate and soil data used in the study can be obtained from Hoffmann et al. (2016b), a brief description is given below.

2.1.1. Climate data

Gridded 1 km × 1 km daily weather data on maximum, minimum and mean temperature, daily rainfall, solar radiation and wind speed over the 29-year period from 1982 to 2011 were obtained by combining daily data from > 200 local weather stations with gridded (1 km) monthly data from the German Meteorological Service (DWD, 2014; see Siebert and Ewert, 2012; Zhao et al., 2015a for precise descriptions). The regional average annual temperature was 9.1 °C and mean annual precipitation was 802 mm with a spatial coefficient of variation of 9.3% and 18% respectively.

2.1.2. Soil data

Gridded 300 m × 300 m soil data (texture, soil layers, depth, C content) were obtained by aggregating mapping units by areal majority using a soil map at a scale of 1:50,000 obtained from the Geological Service North-Rhine Westphalia (Geological Service NRW, 2004), see Angulo et al. (2014) for a more detailed description. The soil data were complemented by soil physical parameters (e.g. water holding capacity) estimated from the texture class by applying pedo-transfer functions developed for German soils (Eckelmann et al., 2005). Topsoil organic carbon and pH were taken from the database FIS StoBo (LANUV, 2014), while the organic carbon content and C:N-ratio of subsoil layers was approximated using pedotransfer functions (Angulo et al., 2014; Eckelmann et al., 2005). Each soil is characterized by a unique set of