Environmental Modelling & Software 74 (2015) 104–113

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

Using site-specific data to estimate energy crop yield

A. Laurent [a, b], C. Loyce [b, a], D. Makowski [a, b], E. Pelzer [a, b, *]

[a] INRA, UMR 211 Agronomie, F-78850, Thiverval-Grignon, France
[b] AgroParisTech, UMR 211 Agronomie, F-78850, Thiverval-Grignon, France

ARTICLE INFO

Article history:
Received 23 April 2015
Received in revised form
4 September 2015
Accepted 16 September 2015
Available online 27 September 2015

Keywords:
Bayesian model
Energy crops
Site-specific data
Yield estimation
Miscanthus × *giganteus*

ABSTRACT

The estimation of energy crop yields is important, to help the firms responsible for collecting them to estimate biomass production in a given area, for example. A Bayesian modelling framework for site-specific yield estimation is presented in this paper. The proposed approach is based on a hierarchical model describing between-site and within-site yield variability. Probability distributions are used to describe the uncertainty of model estimations. The model can be fitted to site-specific yield data, to obtain both average and site-specific yield estimates. Site-specific yield data may be obtained from measurements for crop species other than those for which estimations are required, or from past measurements on perennial crop species grown over a period of several years at a given site. These two options were illustrated in two case studies, in which our model was used to estimate the yields of several energy crops. In most situations, site-specific yield estimations were more accurate than average estimations.

1. Introduction

Biomass (agricultural crops, wood, green or organic waste), as a source of renewable energy, could help to ensure security of the energy supply while reducing net greenhouse emissions and increasing agroecosystem diversity (Heaton, 2004; Kerckhoffs and Renquist, 2012). Biomass can be converted into several types of energy, such as heat, electricity, and biofuel. Energy crops compete for land with food and feed crops, and are therefore a source of controversy. The growth of energy crops on surplus cropland and degraded land unsuis for arable production appears to be a promising alternative (Metzger and Hüttermann, 2009; Rahman et al., 2014) that could reduce the competitive pressure for land. Another way of reducing this competition for land would be to select energy crops with high yields.

Yield estimations can address different types of questions. Consider, for example, *Miscanthus* × *giganteus* (hereafter referred to as *M. giganteus*), a perennial energy crop with a high yield potential (Heaton, 2004). This crop species is typically grown during 15–20 years. During the cultivation period, yield of *M. giganteus* varies from year to year. The yield tends to increase during the first 3–5 years and then reaches a maximum value (Lesur et al., 2013; Miguez et al., 2008). Temporal predictions of yield for this crop would thus be useful, as they would help farmers and collecting firms to anticipate the future yields of recently established crops in a given area, thereby making it possible to estimate more accurately the overall profitability of the crop, or the storage capacity required.

Yield estimations can also help bioenergy firms and farmers' advisers to select the most appropriate energy crop from a list of candidate species. It is generally possible to cultivate several types of energy crop in any given area (Cadoux et al., 2014). As crop yields vary considerably between sites and between years (Miguez et al., 2012), it is not easy to identify the species likely to be the most productive species. In the absence of yield data for a given energy crop at a site of interest, available yield values for other energy crops grown at the same site could be used to estimate yield of the missing crop species. The development of models of this kind could help bioenergy firms to diversify the energy feedstock.

Several types of process-based model have been proposed for the simulation of energy crop yields, particularly for *M. giganteus*. Clifton-Brown et al. (2000) developed a mechanistic model for predicting *M. giganteus* yield in Southern Ireland. Another mechanistic model, MISCANFOR, was developed by Hastings et al. (2009) for the prediction of *M. giganteus* yields as a function of climatic and soil conditions. Miguez et al. (2009) developed a semi-mechanistic model for estimating *M. giganteus* yield as a function of thermal time. Recently, Strullu et al. (2014) developed a process-based

* Corresponding author. INRA, UMR 211 Agronomie, F-78850, Thiverval-Grignon, France.
E-mail address: Elise.Pelzer@grignon.inra.fr (E. Pelzer).

http://dx.doi.org/10.1016/j.envsoft.2015.09.008
1364-8152/

model for simulating biomass dynamics in *M. giganteus* shoots. As this model runs for one crop species only, it cannot be used to compare yields of different crop species in a given growing area. More generally, the calibration of parameters of process-based models is difficult and requires a lot of experimental data (Wallach, 2011). Some input variables of these models may be difficult to measure in farmers' fields (e.g., mineral N in the soil at the beginning of the crop cycle), limiting their potential applications.

Lesur et al. (2013) and Miguez et al. (2008) developed statistical models for estimating yield trends over time for *M. giganteus*. Mola-Yudego and Aronsson (2008) also proposed a statistical model for estimating yields of *Salix* (another perennial crop) over a period of three years. These statistical models include only a limited number of input variables, facilitating their large-scale use, but they may not estimate yield values accurately, due to the variability of energy crop yields across sites and years.

A Bayesian modelling framework made it possible to combine statistical models with site-specific data for the estimation of energy crop yields. The principle is to adjust a statistical model to site-specific yield data, and then to use the fitted model to estimate unobserved yield values. With this approach, site-year effects are estimated through the single or small number of site-specific yield measurements used to adjust the model. These yield measurements may be collected for the species of interest or for other species cultivated at the same site. Thus, no information about soil and climate data is required. Another important advantage of the proposed Bayesian framework is that it provides a quantitative assessment of uncertainty about yield estimation, in the form of a probability distribution (Aguilera et al., 2011; Chen and Pollino, 2012).

The two datasets used in this study are described below. Then our Bayesian framework is presented and its use is illustrated in two case studies in which *i)* the yield of *M. giganteus* is estimated, using past yield data for this species collected at a specific site (case study 1), *ii)* the yields of 36 energy crop species are estimated from yield data collected for alternative crop species (case study 2). The accuracy of yield estimations is assessed in both case studies.

2. Materials and methods

2.1. Datasets used in the two case studies

The main characteristics of the two datasets are presented in Table 1 and described in detail below.

2.1.1. Case study 1: dataset used to estimate yield of M. giganteus for one extra year

This dataset was used to estimate *M. giganteus* yields in future years from past yield data. Yield data were collected from 19 farmers' fields in eastern central France (Burgundy). This region has a semi-continental climate with a mean annual rainfall of 723 mm and a mean annual temperature of 10.9 °C (averaged over 2001–2014, measured locally at Ouges, 47°15′46.3″N, 5°4′26.1″E). *M. giganteus* crops were established on nine fields in 2009 and 10 fields in 2010.

From the second growing season onwards, *M. giganteus* yields were measured in February, as described by Bazot et al. (2014). Yield was not measured during the first growing season (2009 or 2010), because biomass production levels were too low (*M. giganteus* was crushed at the end of December). The last yield measurements were made in February 2014. Three to four sets of yield data were thus available per site, depending on the year of establishment. These data are presented in Fig. 1.

2.1.2. Case study 2: dataset used to estimate yields of different energy crop species

In this case study, the yield of an energy crop was estimated in a given area from yield data collected in the same area, but for a different crop species. The objective was to estimate the yields of an energy crop species, hereafter denoted as species 2, for site-years for which yields were measured for a reference species, hereafter denoted as Ref, different from species 2. The idea was to estimate the yield of species 2 in a given site-year from the yield of Ref measured for the same site-year, based on the correlation between the yields of the reference species Ref and species 2 in other site-years.

A dataset of 856 observations of yield, expressed in tons of dry matter per ha and per year (Laurent et al., 2015) was used. These yield data were collected in 93 experimental site-years in 12 countries and were extracted from 28 published scientific papers. Yields of at least two species were measured for each of the experimental site-years included in the dataset. A separate subset of data was defined for each of 31 pairs of species (Table 2, Fig. 2), and was used to estimate yields of one species from yield data collected for the other species in the same site-year.

2.2. Statistical model for yield estimations

2.2.1. General framework

A hierarchical Bayesian statistical model for crop yield estimation was defined. A within-group level, a between-group level, and a level defining prior distributions were included. Groups corresponded either to sites (case study 1) or site-years (case study 2) (Table 1).

2.2.1.1. Within-group level. This level describes the probability distribution of the yield data within a given group (i.e., within a given site or site-year). Let Y_{ij} be the jth yield data collected in the ith group. Y_{ij} is related to a set of explanatory variables X_{ij} (e.g., time, crop species; Table 1) as follows:

$$Y_{ij} = f\left(X_{ij}, \theta_i\right) + \varepsilon_{ij} \tag{1}$$

$$\varepsilon_{ij} \sim N\left(0, \sigma_\varepsilon^2\right)$$

where f is a function relating Y_{ij} to X_{ij} and to a set of group-specific parameters θ_i, and ε_{ij} is a residual term. Here, all residuals are assumed independent and normally distributed with variance σ_ε^2, but Eq. (1) can be modified to deal with other types of distributions.

Table 1
Main characteristics of the datasets used in case studies 1 and 2.

Case study	Group (index i)		Explanatory variables (X_{ij})		Model function (f)
	Definition	Number	Definition	Number	Type
1	Site	9 or 10	Time	3 or 4	Logistic
2	Site-year	From 2 to 25	Species	2	Linear