Environmental Modelling & Software 86 (2016) 184–203

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

Uncertainty analysis and evaluation of a complex, multi-specific weed dynamics model with diverse and incomplete data sets

Nathalie Colbach [a, *], Michel Bertrand [b], Hugues Busset [a], Floriane Colas [a], François Dugué [a], Pascal Farcy [c], Guillaume Fried [d], Sylvie Granger [a], Dominique Meunier [a], Nicolas M. Munier-Jolain [a], Camille Noilhan [a], Florence Strbik [a], Antoine Gardarin [a, 1]

[a] Agroécologie, AgroSup Dijon, INRA, Univ. Bourgogne Franche-Comté, F-21000, Dijon, France
[b] UMR Agronomie, INRA, AgroParisTech, Université Paris-Saclay, 78850, Thiverval-Grignon, France
[c] INRA, UE 115 Domaine Expérimentale d'Epoisses, F-21110, Bretenière, France
[d] Anses, Laboratoire de la Santé des Végétaux, Unité Entomologie et Plantes invasives, CBGP – Campus International de Baillarguet, CS 30016, F-34988, Montferrier-sur-Lez Cedex, France

ARTICLE INFO

Article history:
Received 4 September 2015
Received in revised form
7 April 2016
Accepted 28 September 2016

Keywords:
Weed
Population dynamics
Mechanistic model
FlorSys
Validation
Cropping system
Uncertainty analysis

ABSTRACT

Weed dynamics models are needed to test prospective cropping systems but are rarely evaluated with independent data ("validated"). Here, we evaluated the FlorSys model which quantifies the effects of cropping systems and pedoclimate on multispecific weed dynamics with a daily time step. We adapted existing validation methodologies and uncertainty analyses to account for multi-specific, multi-annual and diverse outputs, focusing on missing input data, incomplete and imprecise weed time series. Field data ranged from entirely monitored cropping system trials to annual snapshots recorded on farm fields by the French Biovigilance-Flore network. FlorSys satisfactorily predicted weed seed bank, plant densities and crop yields, at daily and multi-annual scales, at well monitored sites. It overestimated plant biomass and underestimated total flora density. Missing processes (photoperiod dependency in flowering, crop:weed competition for nitrogen) and inadequately predicted scenarios (weed dynamics in untilled fields, floras with summer-emerging species) were identified. Guidelines for model use were proposed.

1. Introduction

Weeds are the pest most harmful for crop production (Oerke, 2006). Recently, environmental and health problems led to the implementation of the ECOPHYTO plan (www.agriculture.gouv.fr/ecophyto) in France and the European directive (2009/128) which aim at reducing the risks and uses of chemical pesticides. To date, there are though no curative techniques as efficient as herbicides.[2] To reduce herbicide use, it is therefore necessary to combine several weed management techniques, aiming at both prevention and control, at medium and long-term, because weed seeds survive for several years in the soil (Gardarin et al., 2010).

We therefore need new cropping systems that reconcile reduction of herbicide use while still controlling harmful weeds. Because of the multiplicity of implicated cultural techniques and their long-term effects, models testing many and diverse cropping systems are essential for cropping system design. Among the many existing weed dynamics models (Holst et al., 2007), FlorSys (Colbach et al., 2014b; Gardarin et al., 2012; Munier-Jolain et al., 2013, 2014) is to date one of the rare models quantifying the effects of cropping systems (crop succession, cultural techniques in terms of dates and options such as tools, speed or rates) on the dynamics of multispecific weed floras with a daily time step, in interaction with pedoclimatic conditions. The model predicts the average effects of cropping practices on weed flora and also the variability of these effects as a function of weather conditions and stochastic processes (Colbach, 2010; Colbach et al., 2014a). FlorSys is based on a detailed, mechanistic representation of biophysical processes and represents a virtual field in which cropping systems can be tested and virtual measurements approximated through numerous and

* Corresponding author. INRA, UMR1347 Agroécologie, BP 86510, F-21000, Dijon, France.
E-mail address: Nathalie.Colbach@dijon.inra.fr (N. Colbach).
[1] Present address: UMR Agronomie, INRA, AgroParisTech, Université Paris-Saclay, 78850 Thiverval-Grignon, France.
[2] Except for manual weeding which is too expensive in our societies.

http://dx.doi.org/10.1016/j.envsoft.2016.09.020
1364-8152/

diverse variables describing crops, weeds and soil environment.

Before using such a model for developing practical advice, it is essential to evaluate ("validate") it. Model evaluation consists in determining "the value of a model, with respect to the proposed use of the model" (Wallach, 2006). The term "evaluation" is preferred here to "validation" as the former emphasizes "a diversity of indications about how well the model represents a process or an outcome" (Wallach, 2006), by (1) determining the level of different sources of uncertainty (and errors) in various conditions, (2) determining the domain of validity of the model and (3) questioning the relevance of the modelling choices (i.e. model structure and parameters) over the domain explored. One existing method consists in comparing model simulations to independent field observations in different agricultural and pedoclimatic contexts, using a series of statistical indicators to assess divergence between simulations and observations, and to identify error causes (Wallach, 2006). Individual parts of FlorSys (e.g. light microclimate, short-term emergence, potential crop yield) have been successfully evaluated in the past (see synthesis in Colbach et al., 2014a). The model as a whole and its ability to predict multiannual weed dynamics remain untested, mostly because of a lack of adequate data. The same is true for most existing weed dynamics models (Holst et al., 2007).

Moreover, models such as FlorSys use many input variables which are sometimes difficult to determine, either because they have no biological or technical equivalent in the field (e.g. the voxel, a 3D pixel used to discretize the canopy in light competition models) or because they are difficult to measure (e.g. weed seed bank at simulation onset). If these variables are badly estimated, prediction quality can be affected. The same problem occurs with observed variables, i.e. weed flora and seed bank dynamics are often difficult to assess, not only because they must be monitored over time but also because some species only occur rarely and almost all are aggregated in patches (Rew, 2001).

Consequently, the objective of the present paper was (1) to evaluate the FlorSys model for its ability to predict weed dynamics and their effects on crop production, i.e. to determine the conditions in which the model can be used (domain of validity), focusing on the model's ability to predict cropping system effects and to rank cropping systems as a function of weed flora and crop yield, (2) to test alternatives for key sub-models that have not yet been evaluated individually or that present deficiencies, and (3) to determine the sensitivity of the model's prediction error to a few key input variables that are difficult to determine. To achieve these goals, we adapted existing validation methodologies, developed for simpler models and easier-to-assess outputs (Wallach, 2006; Yang et al., 2014), to account for model complexity and highly variable and difficult observations, with a particular focus on incomplete data sets. Data sets ranged from cropping system trials with a complete and continuous monitoring of both cropping system history and weed flora, to the Biovigilance-Flore network collecting annual snapshots in terms of field management and weed flora from farm field surveys (Fried et al., 2008).

2. Material and methods

2.1. The weed-dynamics model FlorSys

FlorSys predicts daily weed dynamics over several years or decades as a function of cropping systems and pedoclimate (Colbach et al., 2014b; Gardarin et al., 2012; Munier-Jolain et al., 2013; Munier-Jolain et al., 2014). Only the main aspects are described here. The input variables of FlorSys consist of the daily weather, a description of the simulated field (e.g. soil texture, latitude), the initial weed seed bank (seed density and depth), the weed plant aggregation pattern (i.e. either uniformly or in patches), and the cropping system during the whole simulated period (crop sequence and management practices).

The heart of FlorSys is a generic life-cycle consisting of a succession of life-stages (e.g. dormant and non-dormant seeds in the soil, germinating seeds, emerged plant at cotyledon, seedling, vegetative stage etc) chosen for their interaction with cropping system components and light competition; it applies to annual weed species (details in appendix A.1 in supplementary material online). Pre-emergent processes are simulated for an average m^2 of the simulated field. After emergence, the crop:weed canopy of a field sub-sample (e.g. 6 m $\times$ 4 m) is simulated as an individual-based 3D canopy discretized with voxels (i.e. 3D pixels). Crop plants are placed according to their sowing pattern (e.g. row-sown or broadcasted) listed in the input variables, the weeds randomly, either with a uniform probability or aggregated in patches, depending on the user's choice. FlorSys also comprises a sub-model from STICS to predict soil climate (Brisson et al., 1998) and another from DECIBLE to predict soil structure (Chatelin et al., 2005).

The relationships between the life-stages depend on environmental variables and management techniques (appendix A.2 in supplementary material online). For many processes, probabilities are calculated deterministically for each plant from weed, cropping system and pedoclimate variables, and then the actual outcome (e.g. mortality vs. survival, plant location) is determined stochastically.

FlorSys parameters are currently available for 16 frequent and contrasting weed species: *Alopecurus myosuroides, Amaranthus retroflexus, Avena fatua*, *Capsella bursa-pastoris, Chenopodium album, Echinochloa crus-galli, Galium aparine, Geranium dissectum, Polygonum aviculare, Fallopia convolvulus, Polygonum maculosa* (previously *P. persicaria*), *Senecio vulgaris, Sonchus asper, Solanum nigrum, Stellaria media*, and *Veronica hederifolia*.

2.2. Origin of observed weed data

The preliminary step consisted in collecting data for (1) input to the model and for (2) comparison of simulated outputs with observations. Test locations consisted of two cropping system trials as well as the Biovigilance-Flore database surveying each year a variety of fields all over France (Table 1). In the cropping system trials, the cultural operations as well as the weed and crop data were collected for each field. For the Biovigilance locations, a method had to be developed to estimate cropping system history and weed dynamics from the annual snapshots of fields taken in each region (section 2.2.3).

2.2.1. The Epoisses cropping system trial with detailed weed observations

The cropping system trial at the INRA experimental station in Dijon-Epoisses (Burgundy) is described by Chikowo et al. (2009). Ten fields are managed according to five cropping system strategies, ranging from intensive herbicide-based to herbicide-free, with very diverse rotations and varying degrees of tillage and mechanical weeding (appendix C.1.1 and C.5.1 in supplementary material online). The fields were monitored for weeds from 1999 to 2012. In each field, soil seed bank was measured from 100 soil cores pooled into 10 samples, in summer 1999, and approximately every two years thereafter. *Alopecurus myosuroides* Huds. seeds were added immediately before the first crop sowing to compensate for the rarity of grass weeds. During the 13 years of the trial, weed and crop plants were identified and counted two to four times a year (every two weeks in 2006 and 2007) on 4 quadrats in the subplots where the initial seed bank was measured. The rest of the field was also sampled occasionally. Above-ground biomass was measured

Table 1
Synthetic description of the experimental sites used for FLORSYS evaluation. In each line (from 1 to 14), the best (green cells) and the worst cases (red cells) for model prediction quality (line 15) were highlighted, based on the results of section 3.

	Characteristics	Cropping system trial		Biovigilance database		
		Epoisses	La Cage	Aquitaine	Burgundy	Poitou-Charentes
	Location					
1	Latitude	47°20' N	48°48' N	~44° N	~47° N	~46° N
2	Soil texture (%clay, silt, sand)	44-50-6	17-56-27	20-40-40	48-45-7	36-58-6
	Weather: mean annual temperatures (mean monthly temperatures for Jan. and July)	10.9°C (2.5, 19.9)	11.4°C (4.1, 18.5)	12.3°C (5.0, 19.9)	10.9°C (2.5, 19.9)	12.0°C (5.7, 19.1)
	Cumulated annual precipitation	709 mm	651 mm	961 mm	709 mm	762 mm
3	Crop rotation	Very diverse	Very diverse	Maize monoculture	Oilseed rape/winter wheat/winter barley	Oilseed rape/w. wheat/ sunflower/w. wheat
	Level of detail of observations					
4	Cropping system history	Detailed	Detailed	Aggregating fields with identical combinations of Crop x plough x tillage x sowing date	Crop x plough x tillage	Crop x plough
5	Initial seed-bank measurement	Yes	No	No	No	No
6	Weed flora assessment	Quadrats	Abundance classes, then quadrats	Abundance classes	Abundance classes	Abundance classes
7	Assessments per year	>4	<1	1-2	1-2	1-2
	Observed weed flora					
8	Number of weed species	67	103	165	180	205
9	Number of FLORSYS species among the 10 most abundant	2	1	5	6	9
10	Most abundant weed species	*Alopecurus myosuroides*	*Poa annua*	*Chenopodium album*	*Alopecurus myosuroides*	*Mercurialis annua*
11	Total weed density averaged over fields and years (plants/m²)	145	148	109	76	65
12	Species density averaged over fields and years (plants/m²) of the 16 FLORSYS weed species, and [minimum non-zero, maximum values]	8.51 [10^{-4}, 27775]	1.14 [10^{-7}, 1002]	5.01 [0.05, 75]	2.19 [0.05, 75]	1.70 [0.05, 75]
	Variability in weed flora observations and simulations					
13	Variability in observations ($\sqrt{\mathbf{var}_{obs}}$)$^{\$}$	29%	76%	57%	76%	271%
14	Variability in simulations ($\sqrt{\mathbf{var}_{sim}}$)$^{\&}$	9%	32%	33%	52%	344%
	Model prediction error					
15	Discrepancy between observations and FLORSYS simulations (conclusions from section 3)	Small	High	Medium	Small	Very high

$^{\$}$Relative to ½[max-min observed values]. At Epoisses and La Cage, var_{obs} was calculated for each field, weed species and assessment date over the four quadrats; for the other three sites, $variance_{obs}$ was calculated for each cropping system, weed species and assessment date over all the fields with the crop x plough x tillage frequency belonging to the given cropping system; $^{\&}$ Relative to ½[max-min observed values]. var_{sim} was calculated for each field/cropping system weed species and assessment date over the 10 repetitions.

2–5 times a year, usually outside the seed-bank measurement subplot. Depending on years and crops, crop yield was either estimated on quadrats or corresponded to the whole field harvest.

2.2.2. *The La Cage cropping system trial with incomplete weed observations*

The cropping system trial at the INRA experimental station in Versailles-La Cage (Paris Bassin) is described by Debaeke et al. (2009). Sixteen fields are managed according to four cropping system strategies, ranging from intensive herbicide-based to organic, with diverse rotations and varying degrees of tillage and mechanical weeding (appendix C.1.2 and C.5.2 in supplementary material online). The fields were monitored generally once a year for weeds from 1998 to 2002 with scores based on weed abundance classes similar to the Barralis method (Barralis, 1976). Eight fields were also monitored in 2006–2007, every two weeks on two quadrats per field. The abundance classes were translated into density values, by using the maximum density of the class (e.g. 1 plants/m^2 for the 1–3 class). The weed seed bank was never estimated. Crop yields were combine-harvested yields from the whole fields.

2.2.3. *The Biovigilance-Flore database with cropping system and weed surveys*

The Biovigilance-Flore database was set up in 2002 to monitor agricultural practices and weed flora over France (Fried et al., 2008). Each year up to 2010, several hundred fields were assessed all over France and we chose three contrasting regions for our study (Aquitaine, Burgundy and Poitou-Charentes). Weeds were counted on unsprayed subplots, one month after crop sowing, and again 2 or 6 months later, usually after the last herbicide treatment was applied to the rest of the field. Abundance was estimated according to the Barralis method (Barralis, 1976) and translated into densities using the median density of the class (e.g., the score 3 for 3–20 individuals/m^2 was translated into 11.5). The previous crop and agricultural practices of the current year were recorded, though not necessarily *in extenso*. Fields were only monitored once or only for a few years; there were no actual weed time-courses to compare to simulated dynamics.

Consequently, the evaluation methodology had to be adapted. First, typical regional cropping systems were identified in the three study regions. For each region, the most frequent rotation was determined from the database and completed from farm surveys (Boissinot et al., 2011; Colbach et al., 2009). For each rotation, different tillage strategies were identified, combining mouldboard ploughing vs. no ploughing with low vs. high tillage frequency (appendix C.5.3 in supplementary material online). In Aquitaine, fields were further discriminated according to early vs. late sowing date. There had to be at least five fields for a given "region x cropping system x crop x cultural practices" combination in the database to keep the combination for the evaluation process. The corresponding input scenarios for the simulations were built to be as representative as possible of the fields in each database scenario; when information on agricultural practices was missing in the database, it was completed from the farm surveys, particularly in Burgundy and Poitou-Charentes. In total, 110 fields were used in Aquitaine (for 10 scenarios), 321 in Burgundy (10 scenarios), and 389 in Poitou-Charentes (11 scenarios).

2.3. Simulation scenarios

2.3.1. Features common to all sites

Simulations were run with the 20130612 version of FlorSys, initializing with a weed seed bank consisting of seeds of the 16 weed species included in FlorSys. All crops and cultural operations recorded in the monitored locations (except fungicides, insecticides and growth regulators) were used as cropping system input variables for the simulated fields. Not all grown crop species are parameterized in FlorSys; so when a species was missing, the most similar species existing in FlorSys was used instead (appendix C.3 in supplementary material online). Similarly, the FlorSys crop varieties closest to those grown in the trial were used. The simulated soil characteristics (texture etc.) were measured with soil analyses from the cropping system trials or from locations inside the Biovigilance sites (appendix C.6 in supplementary material online). Daily weather variables were recorded by INRA weather stations in the different regions (INRA Climatik platform).

Each field was simulated 10 times with identical input data, to take account of stochastic processes in FlorSys (e.g. plant location or mortality, see section 2.1). A preliminary simulation study concluded this to be the minimum number of repetitions necessary to reduce the variability between simulation runs (appendix D.1.1 in supplementary material online). The field sample (6 m by 3 m) was chosen large enough to minimize edge effects and small enough to limit simulation duration; a huge computer carrying capacity (10 000 plants m^{-2}, see section 2.5.5) was chosen to avoid underestimating plant densities; and the chosen voxel edge size (7 cm) was the one optimizing the prediction of radiation inside the canopy (Munier-Jolain et al., 2013). Simulated weed plants were aggregated in patches, with one patch per species.

Each day, two major output variables, plant density (distinguishing phenological stages) and biomass, were printed to files, for all crop and weed species. Weed seed production and crop yield were calculated at plant maturity and/or harvest.

2.3.2. Application to the cropping system trials

At Epoisses, the initial weed seed bank was initialized with the seed densities measured at the onset of the trial for the 16 FlorSys species. *A. myosuroides* seeds were added to the simulations at the same date they were added to the fields. Simulations ran for the whole monitored period, from summer 1999 to autumn 2012.

The La Cage simulations were based on the same principle, but as there were no initial seed bank measurements, the optimal one was determined during the first evaluation step (Fig. 1).

2.3.3. Biovigilance-Flore sites

The simulation plan was adapted to account for incomplete cropping system and weed data:

- The optimal initial seed bank was chosen during the first evaluation step (Fig. 1);
- Simulations lasted for 24 years, repeating the basic rotation pattern over time (e.g. 8 times the 3-year Burgundy rotation);
- For each simulated year and each of the 10 repetitions, a weather record was chosen randomly among the records available from the regional weather station, using the same 10 weather-year list for each cropping system scenario tested in the region.
- A mean soil texture was used for all fields of a given region, using data from the nearest site where measurements were available.

2.4. Step-by-step evaluation method

Fig. 1 summarizes the different evaluation steps. A preliminary step consisted in an uncertainty analysis of model output due to missing model inputs (e.g. initial weed seed bank) or inputs that are difficult to estimate (e.g. herbicide spraying conditions), using the Biovigilance data which covers three contrasting regions. The objective was (1) to rank inputs as a function of the resulting uncertainty, (2) to compare the distributions of simulated outputs and observations and check whether these data are usable for the model evaluation, and (3) to draw conclusions for choosing missing inputs for the subsequent simulation plan.

The next steps (steps 1–5) analysed the sensitivity of prediction quality to changes in inputs and choice of outputs, focusing on the well documented Epoisses site, in order (1) to estimate the part of prediction error due to missing or difficult-to-estimate inputs as well as discontinuous and approximate weed-flora monitoring, by comparing predictions obtained with estimated vs. measured data from the one well monitored site, (2) to choose the optimal inputs and transformation of output variables based on this analysis in order to discriminate prediction error due to model structure from that due to uncertainty in inputs. Impact on prediction quality was evaluated with statistical indicators for ranking situations according to their prediction error and ability to discriminate species and cropping systems.

In step 6, we looked at the effect of several unevaluated or deficient sub-models (i.e. known or suspected to produce bad predictions) by testing alternatives to the functions currently included in FlorSys. If necessary, alternative functions were chosen for individual sites. A total of four ranking indicators was used to discriminate the sub-model alternatives.

The last step consisted in analysing all model outputs for which observed data were available on at least one site in order to determine the FlorSys domain of validity and to identify additional causes of prediction error. To do so, several types of prediction error were calculated in addition to the four ranking indicators used in the previous steps.

2.5. Uncertainty analysis (step 0)

The uncertainty analysis to missing or badly known inputs was carried out using the contrasting data in terms of cropping systems and pedoclimate from the Biovigilance locations. A series of inputs was identified that are often difficult or impossible to estimate. For each region, a simulation plan of 1000 runs based on Latin Hypercube Sampling (LHS) (McKay et al., 1979) was carried out, combining 11 factors, i.e. cropping system, weather, soil and the other difficult-to-measure inputs. Because the number of runs is relatively small compared to the number of factors, a second LHS plan with additional 500 runs was simulated to check the results (section D.1 in supplementary material online). The following subsections explain the rationale and the options tested in the uncertainty analysis for each factor.

2.5.1. Cropping system

For each region and run, the simulated cropping system was chosen randomly among the 10 or 11 systems of the region.

2.5.2. Pedoclimate

The three soils corresponding to the three Biovigilance sites were used in each region. The weather series differed among regions, choosing among the 10 series of randomly chosen annual regional weather records (see section 2.3.3).

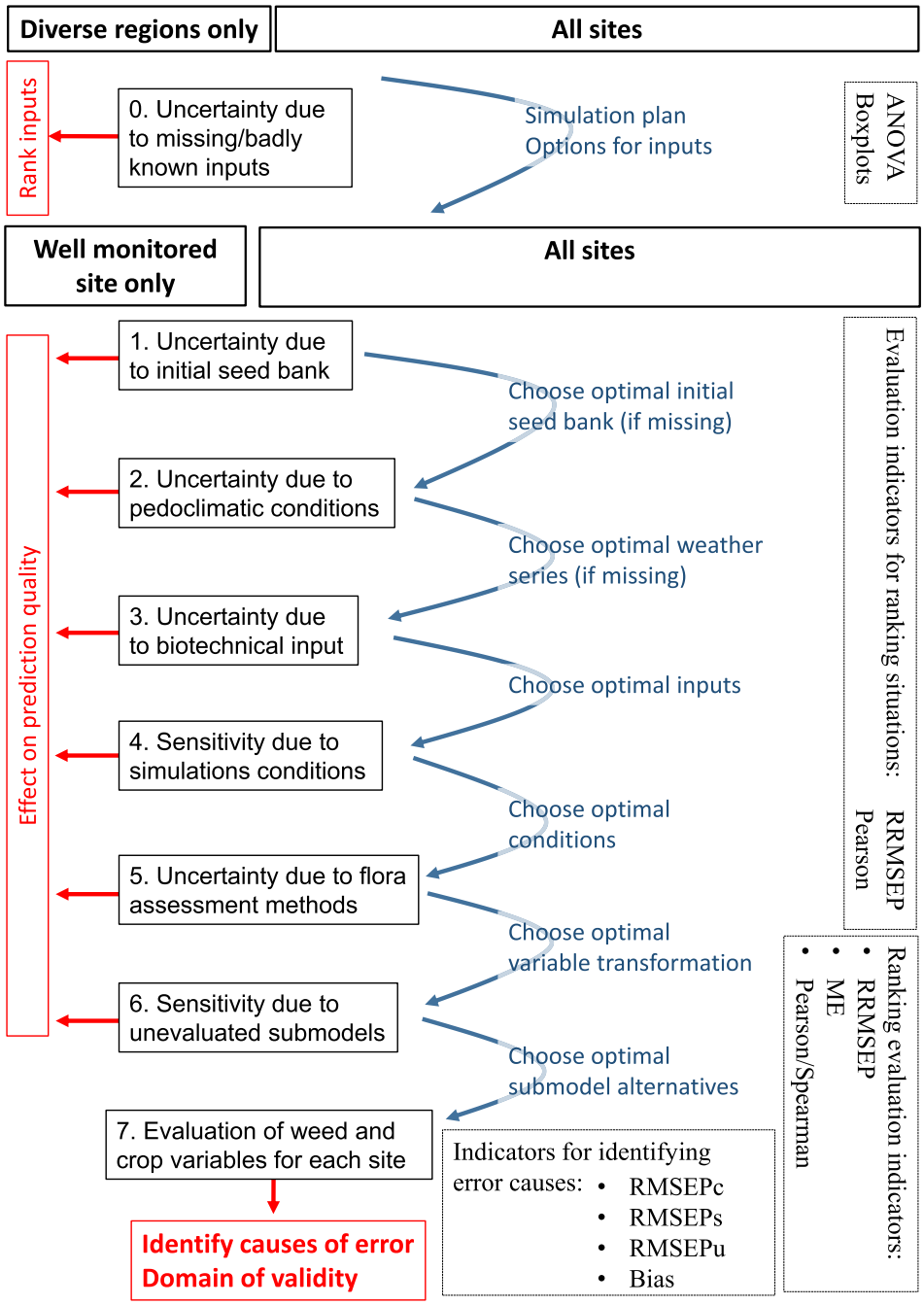

Fig. 1. Schematic representation of steps and indicators for evaluating FlorSys depending on the level of documentation and monitoring of the evaluation sites (Nathalie Colbach © 2016).

2.5.3. Initial weed seed bank

Weed seed bank in individual fields is very rarely measured because it is expensive and time-consuming. Here, two aspects were tested, the total seed density and the species pool, i.e. the proportions of the 16 FlorSys species among the seeds. The seed density at simulation onset was drawn in [2080, 25408] seeds/m^2, i.e. the minimum and maximum species seed densities measured at Epoisses and multiplied by the number of simulated weed species (i.e. 16). The seeds were distributed over the 10 top soil cm.

Ten options were used for the species pool:

- A measured pool, the average species proportions measured in the Epoisses seed bank,
- A pool based on equal species opportunity, taking account of the trade-off between seed number and seed mass, with identical seed biomasses for each species,
- Eight regional estimated pools, i.e. proportions were drawn from a normal distribution using the average and standard-error of the relative species densities of all years and plots of the location divided by the species seed mass to account for the seed number vs. seed mass trade-off. Despite large variations in species densities, the randomly drawn species pools differed

little, particularly compared to the measured pool (section D.1.1 online).

2.5.4. Other biotechnical input variables

Other biotechnical inputs were tested, focusing on those that are difficult to estimate:

- The herbicide-spraying conditions (resulting from the farmer's technical ability and temporal availability), either optimum or suboptimum. The latter results in a reduction of maximum herbicide efficiency (e.g. 0.975% grass weed mortality for a graminicide instead of 100%);
- Weed distribution pattern, either randomly over the whole field area, or aggregated into patches;
- Weed seed migration into the fields once a year, with FlorSys randomly choosing the number of seeds per species and year in a normal distribution. Here, the total weed seed immigration density was drawn in [0, 64] seeds/m^2, and the principle for choosing the species pool was the same as for initial seed bank (section 2.5.3).

2.5.5. Scale parameters

FlorSys comprises several scale parameters determining the precision of calculations which determine simulation speed and the necessary computer memory: (1) the simulated field sample area, either 6 m $\times$ 3 m, or 4 m $\times$ 2 m, which potentially misses rare species; (2) the computer carrying capacity for weed plants, drawn in [5000, 10 000] plants/m^2. In case of very large densities, a small capacity eliminates late-emerging plants that usually die a few days later, and (3) the voxel edge size, drawn in [5, 10 cm]. A larger voxel decreases the precision of plant-morphology representation and competition for light.

2.5.6. Statistics

The analysed output variable was the number of weed plants/m^2, either per weed species or summed over all species, before herbicide spraying. Outputs were averaged over the 24 simulated years and analysed with linear models as a function of inputs and all double interactions, using PROC GLM of SAS. Inputs were ranked, based on decreasing type IV sum of squares. For the major factors, boxplots were drawn to compare the distributions of simulated vs. observed outputs. Comparison of means were carried out to identify outlying options.

2.6. Modifications in the simulation plan for steps 1–7

Different scenarios were run to further assess the effects of various types of input variable uncertainty and observation errors, as well as model sensitivity to several sub-models (steps 1 to 6 of Fig. 1). In contrast to the uncertainty analysis of step 0, only one input was made to vary in each step, and outputs were compared to observations, focusing on the well monitored Epoisses site. There, most of these inputs were measured and it was thus possible to quantify the increase in prediction error when measured inputs were replaced by approximations. The 7th step analysed the prediction error remaining after missing inputs optimized to minimize their impact on prediction error. The following sub-sections detail the necessary changes in the nominal scenarios. In each step, irrespective of the tested modifications, the rest of the simulation plan remained unchanged.

2.6.1. Model sensitivity to initial seed bank

To identify the best way to estimate the initial seed bank when missing and to assess the impact on prediction quality, several initial seed banks (appendix C.2.2 – C.2.4 in supplementary material online) were tested during the first step of the evaluation process (Fig. 1). We focused on seed banks contrasting in species proportions, based on the results of the uncertainty analysis (section 3.1), testing two species pools, i.e. one based on *equal biomass*, and a *regional pool*. As there was little differences between randomly chosen pools in the uncertainty analysis, both in terms of species proportions (section 2.5.3) and effect (section 3.1), the average of the eight regional pools of section 2.5.3 were used. Total seed density was the equivalent to 125 mg seeds$\cdot$m^{-2} times the number of the simulated weed species (resulting in an average of 200 seeds per species and m^2) distributed over the top 10 cm soil. These average seed numbers and masses correspond to the median values observed at Epoisses (appendix C.1.1 in supplementary material on line). As the initial-seed-bank density was shown to have little effect (section 3.1), only one other density was tested, i.e. 16$\cdot$1250 mg seeds$\cdot$m^{-2}, and only with the regional species pool, resulting in a *ten-fold regional* seed bank.

Another solution for estimating missing seed banks was tested, which could not have been used in the uncertainty analysis. A *calibrated local* seed bank (only at La Cage) was created where the species seed biomass of 125 mg seeds$\cdot$m^{-2} was weighted, for each field, by the ratio of densities observed in the given field during the first two years vs. the densities simulated for this same field when starting simulations with the equal-mass seed bank. If necessary, densities were increased to a minimum of at least 1/18 seed/m^2 (i.e. the inverse of the simulated field area) to ensure the presence of at least one seed for each species in the simulation.

In total, 16 scenarios were run (with 10 repetitions per scenario). A 17th case resulted from including additional weed flora measurements from outside the zone where the initial seed bank was measured when analysing the Epoisses scenario starting with the measured seedbank.

2.6.2. Model sensitivity to pedoclimate

During the second evaluation step, a different soil was tested at Epoisses, using the one giving the most extreme values in the uncertainty analysis (section 3.1.2). Moreover, instead of using the weather recorded during the trial, the same principle as for Biovigilance sites was used, i.e. randomly chosen weather records for each simulated year and repetition. For the Biovigilance sites, additional scenarios were run with a different series of randomly chosen weather records. Including the nominal scenarios, a total of ten scenarios x 10 repetitions were run.

2.6.3. Model sensitivity to other biotechnical input variables

During the third evaluation step, the sensitivity to the three other biotechnical input variables was tested (see section 2.5.4): (1) the herbicide-spraying conditions was reduced from optimum to suboptimum; (2) weeds were distributed randomly over the whole field area instead of aggregated into patches; (3) weed seed immigration vs none. As the uncertainty analysis showed that only the species pool of the immigrating seeds had a slight effect (section 3.1.1), the density of the immigrating seeds was kept stable at an average of 0.625 mg/m^2 seeds per species and year. Different values were tested for the species pool (appendix D.4 in supplemental material online): (1) proportions corresponding to equal biomasses for each species (*equal-biomass* immigration), (2) the same regional pool as for initial seed bank (see section 2.6.1), depending on relative species densities in the regional flora (*regional* immigration), and (3) equal biomasses weighted by the relative mean or maximum dispersal distance (*functional* immigration) predicted according to literature (Flores-Moreno et al., 2013; Thomson et al., 2011). A total of 20 scenarios x 10

repetitions (including the nominal scenarios for the five sites) were run for this step.

2.6.4. Model sensitivity to scale parameters

During the fourth evaluation step, changes in scale parameters (see section 2.5.5) were tested which potentially make simulations faster and require less computer memory, using two extreme values from the uncertainty analysis: (1) reducing the simulated field area from 6 m × 3 m–4 m × 2 m,; (2) reducing the computer carrying capacity for weed plants from 10 000 to 5000 plants/m^2, and (3) increasing the voxel edge size from 7 to 10 cm. These changes were tested at Epoisses and with the Biovigilance sites, resulting in sixteen scenarios x 10 repetitions (including the nominal scenarios).

2.6.5. Uncertainty of observed weed data

To evaluate the impact of class-based weed flora assessment, the densities counted with quadrats at Epoisses and La Cage were transformed into equivalent abundance classes (e.g. a quadrat-based count of 9.6 plants/m^2 corresponded to the score 3 on the Barralis scale). These were back-transformed into the median class densities when calculating statistical evaluation criteria (e.g. score 3 for 3–20 individuals/m^2 was translated into 11.5), using weed variables simulated with the nominal scenarios.

Another possible source of flora assessment error are the tiny, freshly emerged seedlings just pointing their cotyledons (cotyledon stage in FlorSys), which are easy to miss in the fields, particularly when counting during the emergence flush. Model evaluation criteria were calculated once including these seedlings in the simulated output, and once without, both times using the nominal scenarios of the five locations.

2.6.6. Uncertainty related to untested sub-models

Several modifications of FlorSys sub-models were tested (appendix A.3 in supplementary material online), focusing on those developed or parameterized with insufficient data, testing: (1) a decrease in weed seedling uprooting by some tillage tools, (2) the elimination of frost damage (biomass loss, plant mortality), (3) additional phenological functions to keep weeds from flowering outside the usual observed seasons in Burgundy; (4) the elimination of herbicide interception by the canopy ("umbrella" effect). These modifications were only tested in Epoisses where weeds and crops were best characterised, except the improved phenology function. In total, 14 scenarios x 10 repetitions were run for this step.

2.6.7. Best-case scenario (step 7)

A last series of simulations consisted in combining for each location the best input and sub-model options, based on the previous simulation series, resulting in five scenarios x 10 repetitions.

2.7. Statistical analyses

2.7.1. Output variables considered

The following weed variables were used for the evaluation of model performance: the number of plants per m^2 (P), the plant biomass per m^2 (B), and the number of seeds per m^2 in the soil (S). The latter two were only available at Epoisses. These variables were measured for each species (s) on a day (d), in a quadrat in a cropping-system trial (q) or in a field of a Biovigilance site (f). These variables were compared to the corresponding simulated variables $\widehat{P}_{dsr}$, $\widehat{B}_{dsr}$ and $\widehat{S}_{dsr}$ for plants, biomass and seed bank on day d for species s and repetition r. At Epoisses and La Cage, plant and seed densities were $\log_{10}$-transformed because of the large range in variation (e.g. from 10^{-7} to 10^3 at La Cage) after adding a constant because of nil values (the constant was the highest 10^i value < the smallest observed density, i.e. 0.1 at Epoisses and 10^{-7} at La Cage).

Total number of plants (TP), biomass (TP) and seed bank densities (TS) are calculated by summing densities over all species, i.e. $\widehat{TP}_{dr} = \sum_s dP_{dsr}$, $\widehat{TB}_{dr} = \sum_s dB_{dsr}$ and $\widehat{TS}_{dr} = \sum_s dS_{dsr}$ calculated from the sixteen simulated species, $TP_{dq} = \sum_s dP_{dsq}$, $TB_{dq} = \sum_s dB_{dsq}$ and $TS_{dq} = \sum_s dS_{dsq}$ from the total number of species observed in the fields (ranging from 67 to 205, depending on locations, Table 1).

At the cropping-system trials, the crop plant densities (only at Epoisses) and the crop yield (Y) at harvest were also analysed.

2.7.2. Evaluation of predictions

Comparisons were carried out at two scales:

- The daily scale, comparing observed densities averaged over the number of quadrats (Q = 4) per field (in the case of cropping-system trials) or the number of fields (F) per cropping system (in the case of Biovigilance sites) to simulated densities averaged over the number of repetitions (R = 10), e.g. $\widehat{dP}_{ds} = \frac{1}{R}\sum_r P_{dsr}$ vs. $dP_{ds} = \frac{1}{Q}\sum_q P_{dsq}$ in the case of plant densities. These variables were used to analyse whether weed dynamics were predicted correctly.
- The multi-annual scale, comparing densities averaged over time, obtained by averaging the daily densities over the number of measurement days (D) or simulated days (D′) for each field and species, e.g. $\widehat{mP}_s = \frac{1}{D'}\sum_d dP_{ds}$ vs. $mP_s = \frac{1}{D}\sum_d dP_{ds}$. These variables were used to analyse whether the model ranked cropping systems and weed species correctly.

The daily and multi-annual variables were also compared for total weed densities. Crop plant densities and yields were only analysed at the daily scale.

2.7.3. Evaluation criteria

For each evaluation step, a series of evaluation indicators were used to compare N observed values (y_i) to simulated values ($\hat{y}_i$). During the five first steps (Fig. 1), two complementary evaluation indicators were used in order to rank the tested scenarios, one assessing overall prediction error, the other the model's ranking ability of cropping systems and weed species:

- The root square of the mean square error in predictions $RMSEP = \sqrt{\frac{\sum(y_i-\hat{y}_i)^2}{N}}$ (Wallach and Goffinet, 1987, 1989) was calculated for each location, scenario and for each weed and crop variable y (with y being dP, dPT, mP etc). RMSEP is the average prediction error and was divided here by ½[max-min observed values] to obtain the relative error RRMSEP which facilitates the comparison between variables and locations. Division by the middle of the range of variation was preferred to the usually used mean of observations because of the negative values resulting from log-transforming some y variables.
- The ranking ability, estimated here with the Pearson correlation coefficient between observed and simulated values $r = \frac{\sum(y_i-\bar{y})(\hat{y}_i-\bar{\hat{y}})}{\sqrt{\sum(y_i-\bar{y})^2}\sqrt{\sum(\hat{y}_i-\bar{\hat{y}})^2}}$. Pearson values close to 1 point to a positive correlation between observed and simulated data, but this correlation can differ from y = x (i.e. total fit between observed and simulated data), particularly in the case of a model bias.

When testing the various FlorSys sub-models, two additional ranking indicators were used:

- The modelling efficiency $EF = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$, where $\bar{y}$ is the mean of observations (Mayer and Butler, 1993), defines the ability of a model to predict the value of a variable. The closer EF is to 1, the better is the fit between observed and simulated data (Wallach, 2006). Negative EF values indicate that the mean observed value is a better predictor than the values predicted by the model; positive values are generally considered to indicate acceptable levels of model performance;
- The Spearman correlation coefficient, resulting from calculating the Pearson correlation coefficient with ranks of observed vs. simulated values instead of using the actual values. Spearman values close to 1 indicate that observed and simulated data are ranked similarly though actual values can differ considerably, both absolutely and relatively. When the Spearman coefficient exceeds the Pearson, ranks are better predicted than differences between values. When Pearson values exceed Spearman ones, ranks of similar values can be inverted.

In the last step, prediction error was decomposed into several different variables to pinpoint causes of prediction error:

- The prediction bias, i.e. the mean of predicted weed variable − observed weed variable $\frac{1}{N}\sum(\hat{y}_i - y_i)$, was compared to zero, using a *t*-test, to determine whether the model systematically over- or underestimates variables.
- The systematic error $\mathrm{RMSEP_s} = \sqrt{\frac{\sum(y_i - \overline{\overline{y}}_i)^2}{N}}$, evaluating the importance of model bias in the prediction error,
- The unsystematic error $\mathrm{RMSEP_u} = \sqrt{\frac{\sum(\overline{\overline{y}}_i - \hat{y}_i)^2}{N}}$, i.e. the remaining prediction error, with $\overline{\overline{y}}_i$ being derived from the linear regression of observed versus simulated values (Willmott, 1981).

We moreover proposed two additional calculations of RMSEP to take account of the huge variability in observations and, to a lesser degree, in simulations in our data set:

- The RMSEP was corrected for variability in observations by subtracting the mean variance of observations over possible samples $\mathrm{var_{obs}} = \frac{1}{DSQ}\sum_{dsq}(\mathrm{y_{ds}} - \mathrm{y_{dsq}})^2$ from the MSEP before applying the root-square. At Epoisses and La Cage, $\mathrm{var_{obs}}$ was calculated for each field and assessment date over the four assessment quadrats; for the other three sites, $\mathrm{var_{obs}}$ was calculated for each cropping system, weed species and assessment date over all the fields with the crop x plough x tillage frequency belonging to the given cropping system. If MSEP is small or smaller than $\mathrm{var_{obs}}$, the difference between observed and simulated values is mostly due to observation error.
- The RMSEP was corrected for variability due to stochasticity in simulations by subtracting the mean variance of simulations $\mathrm{var_{sim}} = \frac{1}{DSR}\sum_{dsr}(\mathrm{y_{ds}} - \widehat{\mathrm{y_{dsr}}})^2$ from MSEP. If MSEP is small or smaller than $\mathrm{var_{obs}}$, the difference between observed and simulated values is mostly due to stochasticity.

As a result, RMSEP corrected for measurement error and model stochasticity becomes $\sqrt{\frac{\sum(w_i - \hat{w}_i)^2}{N} - \mathrm{var}_{obs} - \mathrm{var}_{sim}}$ (Wallach, 2006). To compare the prediction quality of different FlorSys outputs and to determine the model's domain of validity in terms of variables and locations, we propose a synthetic graphical representation inspired by Coucheney et al. (2015), (1) representing RMSEP vs. $\sqrt{\mathrm{var}_{obs}}$ for each variable y and location, both standardized by the standard-deviation in observations, i.e. $\mathrm{STDEVobs} = \sqrt{\frac{\sum(y_i - \overline{y}_i)^2}{N}}$, with (2) the symbol size of each data point (variable x location) proportional to the Pearson correlation coefficient, and (3) vertical bars proportional to the prediction bias. Variable x location combinations are placed into performance classes ranging from "very good" to "bad" along the vertical axis, or into the "unclassifiable" area where observation error exceeds RMSEP.

Finally, we also propose a new indicator to analyse daily density and biomass dynamics while accounting for observed and simulated variability, i.e. the proportion of observed $\mathrm{P_{dsq}}$ and $\mathrm{B_{dsq}}$ data points that were located inside the simulated 90%-confidence interval obtained by calculating 5 and 95-percentiles from the ten $\widehat{\mathrm{P}}_{\mathrm{dsr}}$ or $\widehat{\mathrm{B}}_{\mathrm{dsr}}$ values for each date ($\widehat{\mathrm{P5}}_{\mathrm{ds}}$, $\widehat{\mathrm{P95}}_{\mathrm{ds}}$, $\widehat{\mathrm{B5}}_{\mathrm{ds}}$, $\widehat{\mathrm{B95}}_{\mathrm{ds}}$), here for instance the example for plants:

$$I = \frac{1}{DSQ}\sum_d\sum_s\sum_q i_{dsq}$$

$$\text{if } P_{dsq} \geq \widehat{\mathrm{P5}}_{\mathrm{ds}} \text{ and } P_{dsq} \leq \widehat{\mathrm{P95}}_{\mathrm{ds}} \text{ then } i_{dsq} = 1 \text{ else } i_{dsq} = 0$$

2.7.4. Analysis of residuals

To identify conditions with systematic prediction error (e.g. due to missing processes or inadequately modelled cultural techniques), analysis of residuals was carried out for each location on multi-annual weed species densities obtained with best-case simulations. Both actual residuals ($\hat{y}_i - y_i$) and absolute values of residuals ($|\hat{y}_i - y_i|$, hence error) were analysed with a Partial least squares regression (PLS regression), using cropping system descriptors and weed species as explanatory variables. PLS regression was run with PROC PLS of SAS, using the NIPALS algorithm with leave-one-out cross validation and Hotelling's T2 statistics for model comparison, and estimating missing values with 1-fold iterated imputation. Only variables with scaled and centered regression parameters (in terms of absolute value) exceeding 0.1 were considered important, and parameter estimates were only presented when variables presented an important correlation.

3. Results

3.1. Uncertainty analysis

3.1.1. Ranking inputs

At the species scale, the ranking of inputs based on variability in outputs was the same, regardless of the region (Fig. 2A). Weed species explained by far most of the variability, followed by cropping system and soil. Weed distribution pattern and weather repetition had a tiny effect. The effect of the other variables was not visible.

When looking at the total density summed over all species, cropping system was the dominant factor, followed by pedoclimate (weather repetition and soil) and species proportions among the immigrating seeds (Fig. 2B). Next came the simulation precision (field sample area, voxel edge size, and computer carrying capacity). Herbicide-spraying conditions and weed-distribution patterns were the other two factors with a visible effect.

Whatever the output scale, the density effect of initial seed bank and of seed immigration was negligible, both for total and primary effects. Primary effects were generally quite small, particularly when looking at species densities, where only the species effect was relevant (Fig. 2A). For total densities, the species pools of the

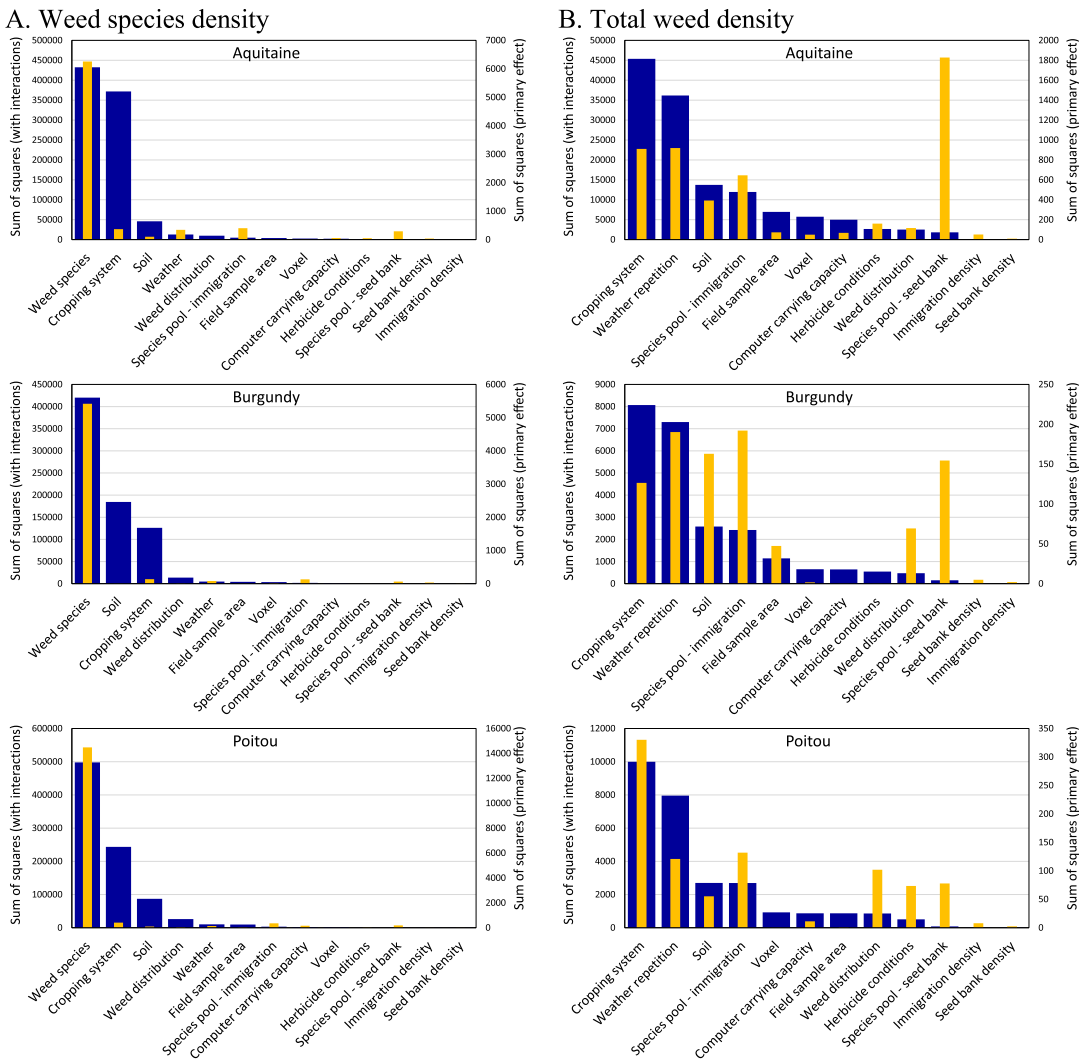

Fig. 2. Ranking of FlorSys inputs depending on explained variability in plant density averaged over the simulation. Sum of squares of linear models explaining plant density as a function of all inputs as well as all double interactions. Large blue bars show sum of squares of primary effect plus interactions (left axis), narrow yellow bars primary effects only (right axis) (Nathalie Colbach © 2016). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

initial seed bank and of the immigrating seeds were the only inputs other than cropping system and pedoclimate that notably impacted outputs (Fig. 2B).

3.1.2. Identifying problematic input values

Using different soils had a big impact in all three regions, with Poitou soil always resulting in the highest total weed density and Aquitaine soil in the lowest (37 plants/m^2 vs. 20, 63 vs 35, 61 vs 57 in Aquitaine, Burgundy and Poitou, respectively; section D.1.2 online). The analysis of the different species pools for the initial seed bank showed that the Epoisses pool resulted in significantly lower densities than the nine other pools (21 plants/m^2 vs. 26–31 in Aquitaine, 43 vs.47–49 in Burgundy). Though significant at $p < 0.05$, these variations are small compared to those observed among cropping systems (ranging in average from 4 to 66 plants/m^2, 44 to 52 and 55 to 62 in Aquitaine, Burgundy and Poitou, respectively, Fig. 3A, C, E) and, particularly, among weed species (ranging from 0.03 plants/m^2 to 16, 0.05 to 26, and 0.01 to 31 in Aquitaine, Burgundy and Poitou, respectively, Fig. 3B, D, F). The other inputs only had tiny effect: suboptimum vs. optimum herbicide spraying conditions increased total weed density (28 plants/m^2 vs. 26 in Aquitaine), uniform vs. patchy weed distribution (29 plants/m^2 vs. 26 in Aquitaine), a 4-by-2-m^2 vs 6-by-3-m^2 field sample area (29 plants/m^2 vs. 26 and 59 vs. 58 in Aquitaine and Poitou, respectively).

3.1.3. Distributions of simulated vs. observed weed densities

The distributions of simulated vs observed weed densities vary with the analysed region and output scale. In Aquitaine, the observed total weed density was usually much larger than the simulated one, except in two cropping systems (Fig. 3A). At the species scale, simulated values were always inside the observed distribution (Fig. 3B). In the other two regions, the distribution of simulated values was always inside the observed distribution (Fig. 3A). At the species scale, simulated and observed distributions always overlapped (Fig. 3B). The degree of overlapping varied among species, with simulated distributions exceeding observed distributions for three species (*A. fatua, G. aparine, F. convuvulus*). These are also the three species with the potentially tallest plants (section A.2 online).

3.2. Prediction error due to missing or badly estimated input data

Some of discrepancies between simulated and observed weed

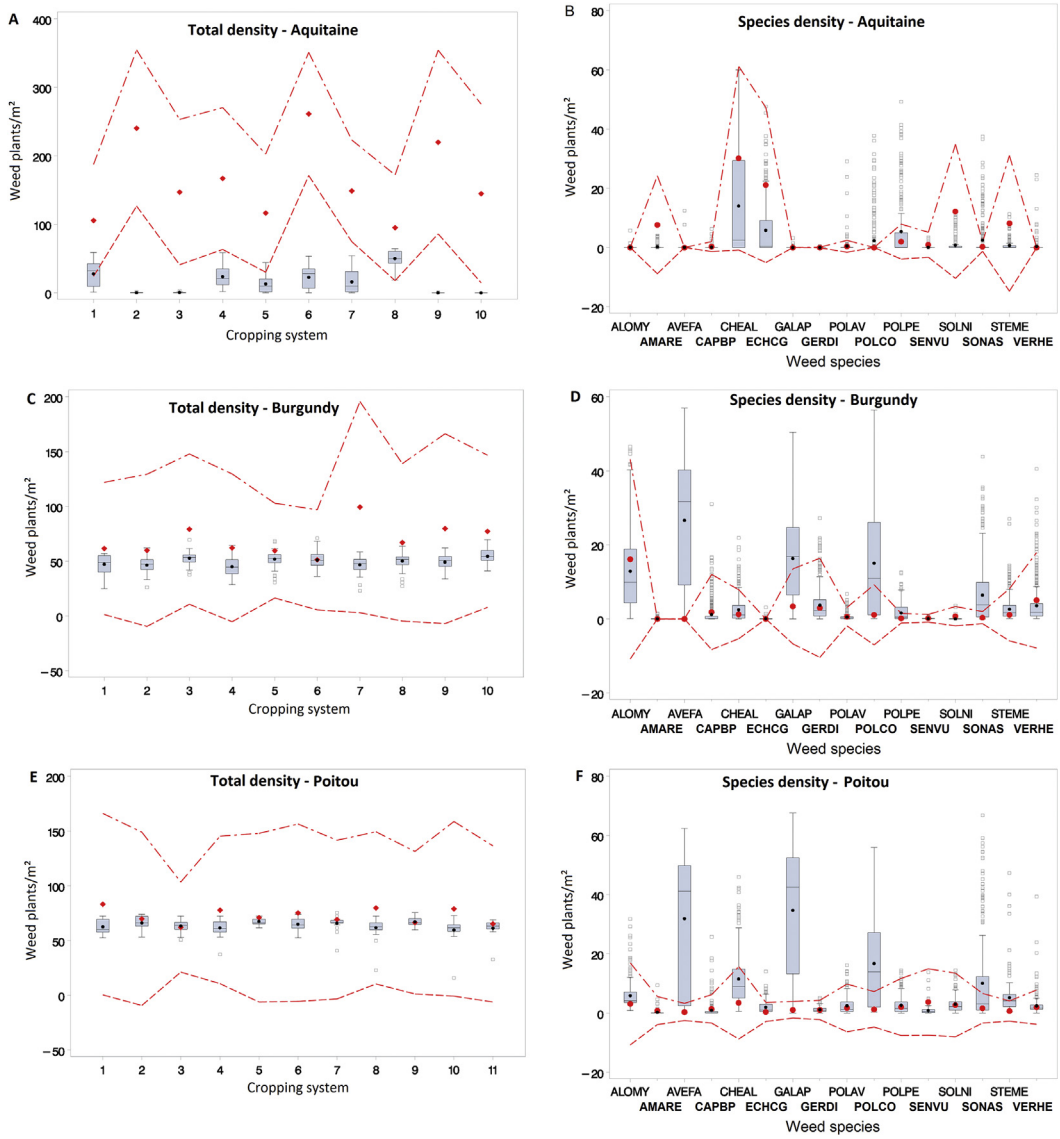

Fig. 3. Variation in simulated (boxplots) and observed (dots and lines) weed plant densities among cropping systems (total weed density) and weed species (species density). Boxplots showing minimum, 25-percentile, median, mean, 75-percentile and maximum simulated values (with outliers outside four times the interquartile range). Out of the 1000 runs per region, only those with the region's own soil were used. Dots show mean observations and lines means ± observation variability (Nathalie Colbach © 2016).

Table 2
Effect of initial weed seed bank on model ability to predict weed species densities averaged over time in a field or cropping system (step 1 of Fig. 1). Prediction quality was appreciated with relative prediction error (RMSEP divided by ½[max-min observed values] in each location) and Pearson correlation coefficient between simulations and observations (reflecting the model's ranking ability of situations and species). Bold numbers show the worst and best performances for each site.

	Initial seed bank[a]	Cropping system trial				Biovigilance database					
		Epoisses		La Cage		Aquitaine		Burgundy		Poitou-Charentes	
		RRMSEP	Pearson	RRMSEP	Pearson	RRMSEP	Pearson	RRMSEP	Pearson	RRMSEP	Pearson
1	Measured (covering all flora counting zone)	29%	0.69	No measurements of initial weed seed bank in these locations							
2	Measured (in part of flora counting zone)	**69%**	**0.40**								
3	Equal seed biomass			**104%**	**<0**	56%	0.24	27%	0.67	**201%**	**0.05**
4	Regional seed bank	**26%**	**0.74**	**76%**	**0.15**	**47%**	**0.38**	**24%**	**0.86**	**188%**	**0.12**
5	Ten-fold regional seed bank	27%	0.72	80%	0.13	49%	0.35	23%	0.84	233%	0.12
6	Local calibrated seed bank	Not tested		83%	0.09	Not tested		Not tested		Not tested	

[a] See section 2.6.1 for details.

distributions are due to missing inputs. In this section, their effect on model prediction error was quantified and the missing inputs calibrated to discriminate prediction error due to model structure from that due to uncertainty in inputs.

3.2.1. *Model sensitivity to initial seed bank*

Among the five tested virtual seed banks, the regional seed bank with the lowest seed densities resulted in the best predictions, irrespective of the location (line 4 vs. lines 3–7 in Table 2):

prediction error (RMSEP) was always lowest and ranking ability (Pearson correlation coefficient) highest. Calibrating a local seed bank from emerged weed flora in each individual field was unsuccessful at the only site where it was tested (i.e. La Cage, line 6).

The regional initial seed bank resulted in slightly better predictions than the seed bank measured at Epoisses (lines 4 vs. 1), despite starting with overestimated seed densities for some species. If the seed-bank measurement subplots only covered part of the flora assessment zones (i.e. if flora observations from outside seed-bank measurement subplots were used in the evaluation), prediction quality was even worse (lines 2 vs. 1). In both cases, the measured seed bank missed or underestimated key species observed in the emerged weed flora (e.g. *Alopecurus myosuroides* and *Capsella bursa-pastoris*, appendix C.4 in supplementary material online).

In the subsequent sections, Epoisses simulations continued to use the measured initial seed bank and weed flora observations from the seed bank measurement subplots. For the four other sites, a regional initial seed bank was used.

3.2.2. Model sensitivity to pedoclimatic conditions

If randomly chosen weather years were used at Epoisses instead of the weather measured during the experiment, prediction error nearly tripled though ranking ability only dropped by 0.11 (line 2 vs. 1 in Table 3). Similarly, if different random weather series from the same region were used for the three Biovigilance sites, prediction error could increase considerably (e.g. Poitou-Charentes) or decrease slightly (e.g. Aquitaine) compared to the reference situation. Ranking ability varied less.

Soil characteristics had less effect on prediction error, despite the large effect identified in the uncertainty analysis (section 3.1). Prediction error and ranking ability remained unchanged when the less clayey Poitou soil was used for Epoisses simulations (lines 3 vs. 1).

3.2.3. Model sensitivity to other biotechnical input variables

Deteriorated herbicide-spraying conditions did not affect weed predictions (lines 4 vs. 1 in Table 3), which is consistent with their invisible effect in the uncertainty analysis (section 3.1.1). The slight effect of weed distribution pattern in the uncertainty analysis translated here into an increase in prediction error when weeds were left to distribute randomly over the field instead of aggregated in patches, though ranking ability could slightly increase (Epoisses, La Cage, Aquitaine, lines 5 vs. 1, Table 3). The only exception was Aquitaine where ranking ability increased by more than 50% while prediction error decreased by approximately 10%.

Various types of yearly seed immigration into fields were tested. The best results were obtained with regional immigration (appendix D.4 in supplementary material online). As in the uncertainty analysis, the effect remained slight (lines 6 vs. 1) as the amount of immigrating seeds remained low. This modification though avoids empty fields in case of catastrophic events eliminating the whole weed population (e.g. excessive frost).

3.2.4. Model sensitivity to scale parameters

Decreasing the modelling precision via modified scale parameters made simulations considerably faster: decreasing the area of the simulated field sample from 18 to 8 m^2 approximately halved simulation time, reducing the computer carrying capacity from 10 000 to 5000 weed plants/m^2 reduced time by approximately 20%, and increasing the voxel edge size from 7 to 10 cm resulted in approximately 30% reduction (results not shown).

As in the uncertainty analysis, the effect of these variables was small. Reducing the field area (which increases interspecific weed competition) did not affect prediction quality (lines 7 vs. 1 in Table 3); it even increased ranking ability in Aquitaine compared to larger field areas. The reduced computer capacity reduced prediction quality in Burgundy (higher prediction error, lines 8 vs. 1). The larger voxel deteriorated predictions at Epoisses and in Burgundy but improved it in Poitou-Charentes (lines 9 vs. 1).

3.3. Uncertainty of observed weed data

When the Epoisses weed flora data was scored using the class-based method instead of quadrat-based countings, prediction error increased by approximately 10% and ranking ability decreased by 10% (lines 5 vs 1 in Table 4). The effect was similar when quadrat-based La Cage data was scored according the density classes, almost doubling the prediction error and reducing the ranking ability by one fifth (lines 5 vs. 3).

If the easy-to-miss cotyledon-stage seedlings were disregarded in the cropping system trials (where weed flora was also monitored during emergence flushes), prediction error decreased. Error decreased most at La Cage with its partial class-based monitoring

Table 3
Sensitivity of FlorSys prediction quality to pedoclimatic conditions, biotechnical inputs and scale parameters analysing weed species densities averaged over time in a field or cropping system (steps 2–4 of Fig. 1). Prediction quality was appreciated with relative prediction error (RMSEP divided by ½[max-min observed values] in each location) and Pearson correlation coefficient between simulations and observations (reflecting the model's ranking ability of situations and species). Bold numbers show deteriorated and improved prediction quality relative to the nominal scenario (if differences exceeded twice the difference between two 10-repetition runs of the nominal scenario).

Scenario	Cropping system trial				Biovigilance database					
	Epoisses		La Cage		Aquitaine		Burgundy		Poitou-Charentes	
	RRMSEP	Pearson	RRMSEP	Pearson	RRMSEP	Pearson	RRMSEP	Pearson	RRMSEP	Pearson
1 Nominal (original simulation plan)	29%	0.69	76%	0.15	47%	0.38	19%	0.88	188%	0.12
(difference between 2 nominal runs)	(1%)	(0.01)	(1%)	(0.01)	(1%)	(0.01)	(0%)	(0.01)	(12%)	(0.03)
Pedoclimatic conditions										
2 Different weather series	**80%**	**0.58**	Not tested		44%	0.42	**24%**	0.86	**231%**	**0.03**
3 Different soil[b]	30%	0.70	Not tested		Not tested		Not tested		Not tested	
Biotechnical input variables										
4 Suboptimum instead of optimum spraying conditions	28%	0.70	77%	0.14	47%	0.38	18%	0.89	169%	0.14
5 Uniform instead of patchy weed distribution	**39%**	**0.72**	**91%**	**0.18**	**52%**	**0.57**	**24%**	0.87	**439%**	**0.05**
6 Regional instead of no seed immigration[a]	28%	0.70	75%	**0.22**	49%	0.36	19%	0.87	**159%**	0.16
Scale parameters										
7 4×2 instead of $6 \times 3m^2$ simulated field sample	29%	0.71	Not tested		44%	**0.49**	19%	0.87	164%	**0.19**
8 Carrying capacity of 5000 weeds/m^2 instead of 10000	28%	0.70	Not tested		48%	0.37	22%	0.88	169%	0.15
9 Voxel edge size of 10 instead of 7 cm	**30%**	**0.65**	Not tested		47%	0.38	**21%**	0.87	**142%**	**0.20**

[a] Annual seed immigration was in average 0.625 mg seed·m^{-2}·$year^{-1}$ for each species, weighted by the relative regional weed flora.
[b] The Poitou soil was used for the Epoisses simulation.

Table 4
Effect of weed flora assessment methods on model ability to predict weed species densities averaged over time in a field or cropping system, using nominal simulations (step 5 of Fig. 1). Prediction quality was appreciated with relative prediction error (RMSEP divided by ½[max-min observed values] in each location) and Pearson correlation coefficient between simulations and observations.

	Weed flora assessment method	Cropping system trial				Biovigilance database					
		Epoisses[a]		La Cage[a]		Aquitaine		Burgundy		Poitou-Charentes	
		RRMSEP	Pearson	RRMSEP	Pearson	RRMSEP	Pearson	RRMSEP	Pearson	RRMSEP	Pearson
1	Quadrats	30%	0.69			No quadrat-based flora assessment on these sites					
2	Quadrats, disregarding cotyledon stage	29%	0.69	No continuous quadrats							
3	Mix: class-based (<2002), quadrats (>2006)	Not tested		82%	0.16						
4	Mix, disregarding cotyledon stage	Not tested		76%	0.15						
5	Class-based	33%	0.63	78%	0.13	47%	0.38	19%	0.88	188%	0.12
6	Class-based, disregarding cotyledon stage	32%	0.63	79%	0.12	47%	0.28	23%	0.86	164%	0.13

[a] Except in lines 5 and 6, weed densities were log10-transformed before averaging over time, with 0.1 and 0.000001 constants for Epoisses and La Cage, respectively, because of nil values.

(lines 4 vs. 3, and 6 vs. 5), and least at Epoisses with its continuous quadrat counting (lines 2 vs. 1, and 6 vs. 5). Conversely, disregarding the cotyledon-stage seedlings at the Biovigilance sites (where flora was assessed after the major emergence flush) usually deteriorated prediction quality (lines 6 vs. 5).

3.4. Uncertainty related to unevaluated sub-models

Two tested sub-model modifications did not change average prediction quality at Epoisses, i.e. less tillage aggressiveness (lines 2 vs.1 in Table 5A) and eliminating any frost damage in weeds and crops (lines 3 vs. 1). Eliminating herbicide interception by the canopy resulted into systematic underestimation of weed densities and generally deteriorated evaluation indicators (lines 4 vs. 1).

The phenology modification was tested at all five locations (Table 5B). Generally, prediction quality deteriorated at the Northernmost locations (increased bias and prediction error at La Cage, increased prediction error and reduced modelling efficiency and Pearson at Epoisses), improved at the Southernmost location (increased Spearman correlation in Aquitaine) and had no effect at the two intermediate ones (Burgundy and Poitou-Charentes).

3.5. Analysing the different output variables

3.5.1. Best-case simulations

The last set of simulations combined all the best options, in terms of scale parameters, biotechnical input variables and sub-model options, with choices differing among sites. In Epoisses, La Cage, Burgundy and Poitou-Charentes, the original model with yearly regional seed immigration was used because its performance was as good as the control simulation (or slightly better) and did not leave empty fields at simulation end. In Aquitaine, the model with the improved phenology function, uniform weed distribution and regional seed immigration performed best.

3.5.2. Ranking cropping systems and weed species

The model's ability to rank cropping systems and weed species was analysed via weed plant densities averaged over time for each cropping system. Cropping systems and weed species were well ranked (either Pearson or Spearman > 0.50) at all locations except Poitou-Charentes (lines 1–5 in Table 6). At three locations (Burgundy, Epoisses, Aquitaine, lines 1–3), relative differences between cropping systems and species were moreover very well predicted (Pearson ≥ 0.65) even if situations with similar density values

Table 5
Effect of FlorSys sub-models on the model's ability to predict weed species densities averaged over time in a field or cropping system (step 6 of Fig. 1). Bold numbers show deteriorated and improved prediction quality relative to the nominal scenario (if differences exceeded twice the difference between two 10-repetition runs of the nominal scenario).

	Simulation scenario	RBias[b]		Prediction error RRMSEP[b]	Modelling efficiency EF	Correlation coefficient	
						Pearson	Spearman
A. Epoisses only[c]							
1	Nominal (original model)	0%	ns[a]	29%	0.45	0.69	0.56
2	Less weed uprooting by tillage	1%	ns	29%	0.44	0.71	0.53
3	No frost damage	0%	ns	28%	0.47	0.71	0.53
4	No herbicide interception by canopy	**−2%**	ns	**30%**	**0.41**	**0.65**	0.53
B. All sites							
Nominal (original model)							
1	La Cage[c] (48°48′ N)	18%		76%	<0	0.15	0.39
2	Epoisses (47°20′ N)	0%	ns	29%	0.45	0.69	0.56
3	Burgundy (~47° N)	−9%		19%	0.68	0.88	0.48
4	Poitou-Charentes (~46° N)	45%		188%	<0	0.12	0.10
5	Aquitaine (~44° N)	−8%		47%	<0	0.38	0.40
Corrected weed phenology							
6	La Cage (48°48′ N)	**25%**		**83%**	<0	**0.18**	0.39
7	Epoisses (47°20′ N)	0%	ns	**31%**	**0.37**	**0.64**	0.53
8	Burgundy (~47° N)	−9%		19%	0.67	0.88	0.46
9	Poitou-Charentes (~46° N)	44%		191%	<0	0.08	0.05
10	Aquitaine (~44° N)	−8%		48%	<0	0.37	**0.50**

[a] ns not significantly different from zero at p = 0.05 (*t*-test).
[b] Relative ½ [max − min observed values].
[c] Weed densities were log10-transformed before averaging over time, with 0.1 and 0.000001 constants for Epoisses and La Cage, respectively, because of nil values.

Table 6
FlorSys ranking ability of cropping systems and weed species. Prediction quality for weed densities averaged over time, using best-case simulations (regional seed immigration at all sites, as well as uniform weed distribution and corrected phenology sub-model in Aquitaine) (step 7 in Fig. 1). Densities were log10-transformed for Epoisses and La Cage (with 0.1 and 0.000001 constants, respectively, because of nil values). Relative bias relative prediction error are bias and RMSEP divided by ½ [max − min observed values]. For more details on evaluation criteria, see section 2.7

Evaluated variable	Number of data points	Bias	Prediction error (RRMSEP)						Modelling efficiency EF	Correlation coefficient	
			Total	Corrected for variability in			Systematic	Unsystematic			
				Observations	Simulations	Both				Pearson	Spearman
Plants/m^2 per species											
1 Burgundy	160	−9%	19%	~0[a]	~0	~0	15%	12%	0.66	0.87	0.43
2 Epoisses	160	0% ns[b]	28%	~0	26%	~0	16%	23%	0.48	0.70	0.42
3 Aquitaine	160	−7% ns	46%	~0	30%	~0	2%	46%	<0	0.64	0.43
4 La Cage	256	22%	75%	~0	61%	~0	35%	67%	<0	0.22	0.51
5 Poitou-Charentes	176	32%	159%	~0	~0	~0	41%	153%	<0	0.16	0.11
Total plants/m^2											
6 Burgundy	10	−191%	202%	~0	164%	~0	201%	19%	<0	<0	<0
7 Epoisses	10	−132%	171%	108%	164%	97%	164%	46%	<0	<0	<0
8 Aquitaine	10	−125%	145%	77%	143%	73%	173%	18%	<0	<0	<0
9 La Cage	16	16% ns	88%	87%	84%	84%	66%	40%	<0	<0	<0
10 Poitou-Charentes	11	−438%	452%	~0	352%	~0	41%	153%	<0	<0	<0

[a] MSEP − variance < 0.
[b] ns not significantly different from zero at p = 0.05 (*t*-test).

might be inverted in terms of ranks (Pearson > Spearman). At the two sites located in the region where most of the measurements for developing FlorSys were carried out (Burgundy and Epoisses, lines 1–2), even absolute density values were well predicted (low prediction error, adequate modelling efficiency). Prediction quality of absolute density values was satisfactory in Aquitaine (line 3), though not as good as for the first two sites. Prediction quality was worst at La Cage and Poitou-Charentes (lines 4 and 5), showing FlorSys to be an acceptable semi-quantitative model.

3.5.3. Variability in observations

As seen above, missing input variables (section 3.1) as well as a less precise weed assessment method (section 3.3) explained a large part of the lower FlorSys prediction quality for the three Biovigilance sites and, to a lesser degree, the La Cage trial. Measurement error also contributed to prediction error. At the cropping-system trials, the variability between observation quadrats (line 13 in Table 1) was as large as the total model prediction error (total RMSEP for lines 2 and 4 in Table 6). At the Biovigilance sites, the between-field variability (line 13 in Table 1) considerably exceeded the model prediction error at all three sites (lines 1, 3 and 5 in Table 6). It was highest in Poitou-Charentes where field aggregation was mostly based on crops and mouldboard ploughing frequency only; it was lower in Burgundy where tillage frequency was systematically used for discriminating fields, and lowest in Aquitaine where sowing date was also used.

When correcting RMSEP for observation variability, prediction error was too small to be estimated (RMSEP corrected for variability in observations, lines 1–5 in Table 6). Most of the discrepancies between observations and simulations were thus due to measurement error, resulting either from intra-field variability (i.e. variability between sampling quadrats in cropping system trials) that was outside our scope, or from aggregating fields with dissimilar histories (Biovigilance sites).

Because of the stochastic effects in FlorSys (see section 2.1), simulated weed densities also varied between repetitions, though this variability was smaller than measurement error in the cropping-system trials (line 14 in Table 1). It increased with the length and diversity of the crop rotations for the Biovigilance sites (line 3), with the lowest variability in Aquitaine and the highest in Poitou-Charentes. Accounting for this variability did not much reduce prediction error for the cropping system trials (RMSEP corrected for variability in simulations, lines 2 and 4 in Table 6), indicating that stochasticity contributed little to prediction error. Conversely, for the Biovigilance sites, the longer and the more diverse the rotation was (line 3 in Table 1), the more stochasticity contributed to simulation vs. observation divergence (lines 1, 3 and 5 in Table 6).

3.5.4. Systematic vs. unsystematic error

When decomposing total prediction error into systematic and unsystematic error, the latter usually exceeded to former (Table 6), indicating that model errors were mainly related to the difficulty to account for dispersion rather than due to bias in model predictions. Weed species densities were though considerably overestimated at La Cage and Poitou-Charentes, i.e. the two sites where the species dominant in the fields were not among the simulated species (line 10 in Table 1). These were also the sites with the highest weed species number (line 8) and the lowest average density for FlorSys species (line 12) in their respective category (i.e. cropping system trial for La Cage, Biovigilance database for Poitou-Charentes).

Total weed densities were always badly predicted (lines 6–10 in Table 6). Most prediction error was systematic (systematic RMSEP > unsystematic RMSEP), with total densities being underestimated in all locations, except La Cage (negative bias, except in line 9). The bias was generally worse for Biovigilance sites (lines 6, 8 and 10) than for the two cropping-system trials (lines 7 and 9 in Table 6). It increased with increasing species richness (line 8 in Table 1), indicating that the weed species missing in the simulations also explain error prediction of total weed densities.

3.5.5. Weed dynamics

3.5.5.1. Daily densities are less well predicted than multi-annual densities. The model's ability to predict weed dynamics, both in terms of densities for a given day and density trends over the years, was analysed via daily weed plant densities on the two cropping-system trials. Most of the conclusions on the multi-annual weed densities remain valid: better prediction at Epoisses (lines 1 and 2 in Table 7) than at La Cage (lines 9 and 10), more unsystematic than systematic error for species densities, underestimated total densities at Epoisses and overestimated species densities at La Cage. Overall prediction quality was though lower for daily than for multi-annual densities, with negative modelling efficiencies and low correlation coefficients. In contrast to multi-annual densities,

Table 7
Evaluating FlorSys prediction quality for daily weed and crop variables, using best-case simulations (step 7 in Fig. 1). Relative bias relative prediction error are bias and RMSEP divided by ½ [max − min observed values]. For more details on evaluation criteria, see section 2.7

Evaluated variable	Initial-seed-bank measurement	Number of data points	Bias	Prediction error (RRMSEP)				Modelling efficiency EF	Correlation coefficient		Prop. Obs. in simulated interval
				Total	Corrected	Systematic	Unsystematic		Pearson	Spearman	
A. Epoisses											
Weeds											
1 Plants/m^2 per species [d]	Inside flora assessment area	7405	0% ns[a]	22%	18%	12%	19%	<0	0.41	0.29	0.88
2 Total plants/m^2 [d]		468	−17%	53%	46%	30%	43%	<0	0.20	0.17	0.35
3 Biomass/m^2 per species	Outside biomass sampling area	3808	5%	41%	35%	6%	41%	<0	0.04	0.36	0.78
4 Total biomass/m^2		241	37%	87%	67%	46%	73%	<0	0.19	0.05	0.33
5 Seed bank per species [d]	Inside seed-bank sampling area	690	9%	51%	37%	26%	44%	0.11	0.56	0.56	nc[b]
6 Total seed bank [d]		46	48%	84%	22%	66%	52%	<0	0.05	0.08	nc
Crops											
7 Plants/m^2	Inside crop-assessment area	140	−4% ns	30%	30%	29%	7%	0.57[c]	0.01	0.74	nc
8 Yield at harvest (T/ha)	Outside harvest area	64	−15%	39%	37%	24%	30%	0.56[c]	0.41	0.41	nc
B. La Cage											
Weeds											
9 Plants/m^2 per species [d]	None	2656	9%	65%	<0	38%	53%	<0	0.12	0.17	0.79
10 Total plants/m^2 [d]	None	166	1%	88%	84%	67%	56%	<0	<0	<0	0.25
Crops											
11 Yield at harvest (T/ha)	None	157	−49%	89%	nc	71%	51%	<0	<0	<0	nc

[a] ns not significantly different from zero at p = 0.05 (*t*-test).
[b] Not calculated.
[c] Observations were weighted by the inverse of standard-error over observation quadrats for each field and assessment date.
[d] Densities were log10-transformed, with 0.1 and 0.000001 constants for Epoisses and La Cage, respectively, because of nil values.

the discrepancies between observed and simulated densities were only partially due to the variability between flora-assessment quadrats (the corrected RMSEP is only slightly lower than the total RMSEP).

3.5.5.2. Prediction error is mostly due to a few days discrepancy in timing of density variations. The graphs of observed and simulated weed species densities over time (see example of Fig. 4) show that the general tendency of the species dynamics was rather well predicted (e.g. when observed species densities globally decreased with time, the simulated densities did so as well). However, simulated densities could increase or decrease a few days earlier or later than observed densities. While this did not change the overall weed dynamics and did not affect the comparison of cropping systems, it considerably deteriorated classical model-evaluation indicators such as RMSEP. The simulated confidence interval though comprised most observed data points, i.e. the proportion of observation points included in the simulated interval was excellent for weed species densities at Epoisses (line 1 in Table 7) and still rather good at La Cage (line 9). This good performance was not due to the width of the simulation interval, because weighting observations by the inverse of the interval width increased the inclusion proportion (e.g. 0.90 at Epoisses, results not shown). As observations were carried out at a much larger time step than the simulations, it was though not possible to conclude whether the model reproduced all short-term variations correctly. The analysis of the inclusion proportion confirmed that total weed densities are badly predicted at both sites.

3.5.6. Weed biomass is overestimated

Weed biomass was only measured at Epoisses and was the only weed variable measured outside the areas where the seed bank for initializing the simulation was assessed (section 3.2.1). Its prediction quality was worse than for density predictions: species biomass was systematically overestimated, with a large prediction error, though dynamics over time were mostly correctly predicted, with 78% data points included in the simulated interval (line 3 in Table 7). As before, means over time were better predicted than daily values (results not shown). Total biomass summed over all species was also systematically overestimated (line 4 in Table 7), in contrast to total plant densities, indicating that the empty space left by the missing weed species (section 3.5.4) cannot be responsible for this bias.

3.5.7. The soil seed bank dynamics is well predicted

Seed-bank densities per species were well ranked at Epoisses (line 5 in Table 7), again confirming the model's ability to simulate species dynamics over the years correctly. As in the case of weed biomass, the seed bank was slightly overestimated for individual species densities and even more so for total densities (line 6).

3.5.8. Crop yield prediction was acceptable

Crop variables were relatively well predicted at Epoisses: plant densities were ranked correctly, without a bias, but with a prediction error larger than for weeds and entirely due to model deficiencies (i.e. corrected RMSEP as high as total RMSEP, line 7 in Table 7). Crop yield was adequately predicted (modelling efficiency > 0.50), though systematically underestimated and a prediction error (3.5 T/ha) due to model deficiencies. If simulations were run without weeds, crop yield was systematically overestimated (bias = +5%) and the modelling efficiency decreased (EF = 0.52), indicating that including crop:weed competition is necessary to correctly predict crop yields.

The prediction quality considerably varied among crop species, though for many species the number of field x year combinations

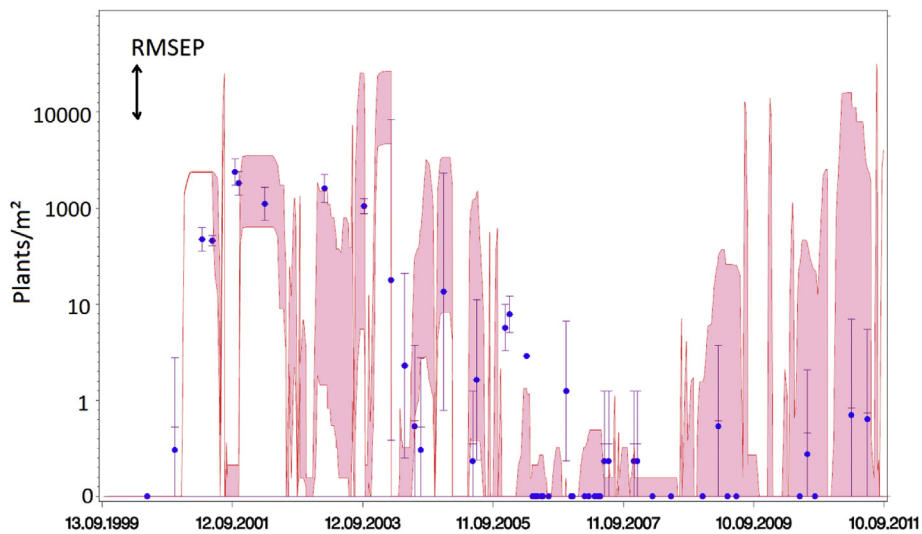

Fig. 4. Daily weed densities over time observed on the Epoisses trial (symbols and vertical bars: mean and standard-error) and simulated with FlorSys (red lines: 90%-confidence interval over 10 repetitions). Example of *Alopecurus myosuroides* In field D5 managed with a herbicide-free strategy (Nathalie Colbach © 2016). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

was too low to draw a conclusion (maize, mustard, sorghum, sunflower). Among the others, oilseed rape yield was significantly overestimated (N = 9, bias = 28%, appendix D.3.1 in supplementary material online), barley (N = 8, bias = −74%) and soybean yields (N = 6, bias = −62%) were underestimated; no bias was identified for wheat (N = 26) and triticale (N = 6). Part of the yield underestimation was due to delayed maturity in the simulations. While maturation was usually consistent with harvest dates for soybean (95%), wheat (86%) and oilseed rape (84%), it was too late than 66% for all other crops (appendix D.3.1 in supplementary material online).

Yield predictions were bad and even more underestimated at La Cage (line 11), where weed densities were generally overestimated. In contrast to Epoisses, yield prediction quality increased when disregarding weeds (biais = 0%, RMSEP = 71%, Spearman = 0.22). Nevertheless, the same kind of crop ranking was observed for the three most frequent crops, i.e. yield tended to be overestimated for oilseed rape (N = 30, bias = 37%) and underestimated for wheat (N = 99, bias = −82%) and pea (N = 26, bias = −98%).

3.6. Overall model performance

3.6.1. Graphical analysis of prediction quality

Fig. 5 shows the general performances of FlorSys for the different output variables and sites. The limits of the performance classes were fixed, based on the prediction quality of multiannual weed species densities (MWPS) identified above: the RMSEP thresholds of the "very good", "good" and "satisfactory" classes were fixed to respectively include Burgundy, Epoisses and Aquitaine. The figure summarizes the results of the previous sections: (1) prediction quality was satisfactory (yellow shaded area in Fig. 5) or good (green shaded area) for multi-annual weed plant densities per species (MWPS), (2) it was less good but still satisfactory for daily crop and weed plant and seed bank densities (DCPS, DWPS, DWSS), and (3) acceptable for crop yield (CY). (4) No definite conclusion could be drawn for several variables (e.g. MWBS, MWBT) and sites (Poitou-Charentes) because variability due to measurement exceeded total prediction error (grey shaded area). (5) Daily weed biomass (DWBS, DWBT) was badly predicted (red shaded area) and generally overestimated (positive bias), while (6) total weed plant (DWPT, MWPT) were generally underestimated (negative bias).

3.6.2. Sensitivity of residuals to cropping system variables and weed species

Analyses of residuals and error (i.e. absolute values of residuals, see section 2.7.4) were carried out for multi-annual weed species densities for each site, using cropping system descriptors and weed species as explanatory variables (appendix D.5.1 in supplementary material online). None of these variables had a significant effect on prediction bias on any of the five sites, and prediction error was only affected at La Cage. Generally, the more diverse a system was, the higher the prediction error. Diverse rotations increased prediction error (0.46), particularly with late-harvested crops (0.33) while late-sown (i.e. spring) crops decreased it (−0.68). Moreover, positive correlations with prediction error were identified for the number of mouldboard ploughing operations (centered and scaled regression parameter = 0.474), herbicides (0.59), and mowing (0.71), indicating that disturbances tend to increase simulation error by cumulating the small errors of the individual sub-models. Disturbances also increase measurement errors because they make weed densities increase or decrease, and the variance over time of a measured variable increases with the absolute value of the slope.

Tillage was the major exception: prediction error was higher in no-till (0.22) and decreased with increasing tillage frequency (−0.83) and with multiple crop sowings per year (−0.63). Error also decreased with increased observed weed density (−0.43), which is not surprising as small densities are difficult to detect in fields.

3.6.3. Identification of deficient model processes

The prediction biases of the FlorSys stages were synthesized in Fig. 6. Weed plant densities were correctly predicted but their subsequent biomass production overestimated. As the FlorSys light microclimate sub-model was satisfactorily evaluated in a previous work (Munier-Jolain et al., 2013), this points to a problem in light absorption and/or conversion. Despite biomass overestimation, both crop and weed plant densities were well predicted, indicating that the overestimation was not sufficient to result in plant mortality due to shading.

The next assessed stage after biomass was the crop yield which was often underestimated, probably because of the overestimated weed biomass and also because simulated crop maturity occurred generally too late and maturation was cut off by harvest. This confirms the deficiency in the phenology sub-model already

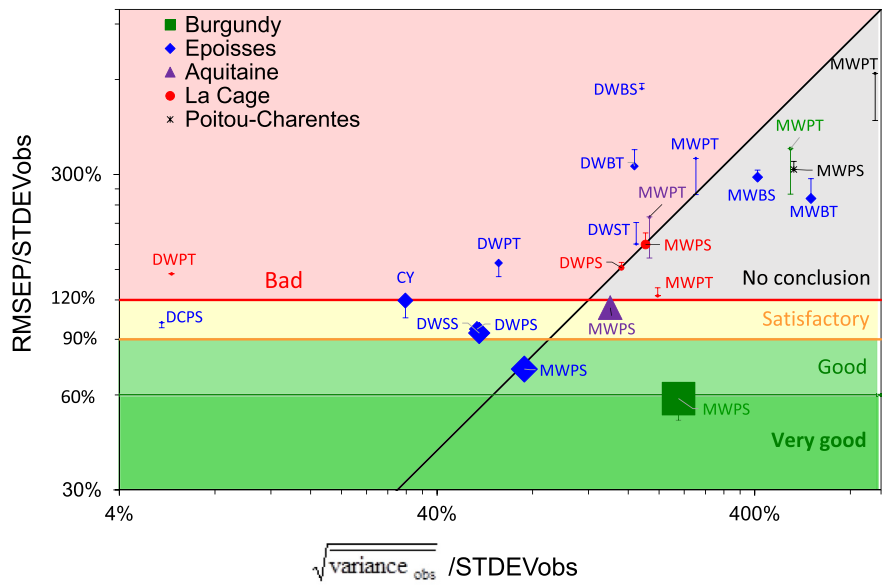

Fig. 5. Synthetic graphical representation of the overall model performances for several outputs and locations. The coordinates represent respectively the root-squared measurement variance of observations ($\sqrt{\text{variance}_{\text{obs}}}$) and the prediction error (RMSEP) normalized by the standard-deviation of observations STDEVobs over all field/cropping systems in a location. Vertical bars are proportional to the prediction bias normalized by STDEVobs. The size of the symbols is proportional to the Pearson correlation coefficient. Data points are daily weed plant, biomass and seed densities per species (DWPS, DWBS, DWSS) and daily densities summed over all species (DWPT, DWBT, DWST), multi-annual weed plant and biomass densities per species (MWPS, MWBS) and total multi-annual densities summed over all species (MWPT, MWBT), daily crop plant densities per species (DCPS) and crop yield at harvest (CY). CY could not be shown for La Cage because STDEVobs was unknown. Points to the right of the y = x line present a variability due to measurement exceeding total prediction error (Nathalie Colbach © 2016).

identified for weeds during the sensitivity analysis in step 6. It did though not result in underestimated weed seed rain, as the next assessed weed stage, the seed bank in soil, was still overestimated.

Despite the seed bank overestimation, weed plant densities were well predicted. Underestimated seed germination or pre-emergent seedling mortality are unlikely causes for cancelling out the overestimation as these would deplete the soil seed bank (and thus reduce seed-bank overestimation) and not just reduce emergence. The same applies to soil seed mortality. Overestimated seed dormancy seems more likely, reducing germination while keeping the seeds alive.

3.6.4. Domain of validity and rules for model use

Table 8.A summarizes the conclusions from the present study in terms of domain of validity. For instance, weed species densities were well predicted at the daily scale, particularly when the sixteen simulated species dominated in the regional weed flora. This output variable could though only be evaluated at two locations in the Northern half of France (i.e. Epoisses in Burgundy, and La Cage in the Paris Basin); it might therefore not be correctly predicted in more Southern locations (as was indeed the case for weed densities averaged over time). There is also a risk of densities being overestimated in continuous no-till.

Table 8.B summarizes advice on how to choose input variables and simulation plans. For instance, when weeds are rare, either in terms of densities or species, a small field sample area (e.g. 4 m × 2 m) is sufficient. The area must be increased if weeds are more frequent, to avoid overestimating intra and interspecific weed:weed competition. The field sample width perpendicular to crop rows must always be sufficient to allow several crop rows and as many rows as interrows, to avoid for instance overestimating crop:weed competition to the detriment of weed:weed competition in case of more rows than interrows. This risk is particularly high in case of large interrows.

4. Discussion

4.1. A framework for using scarce and heterogeneous datasets

The present study proposed a framework to exploit as much as possible existing but incomplete data sets for evaluating complex and long-term simulation models. One innovation was to develop methods to estimate missing key inputs such as the initial seed bank, and to compensate for the absence of multiannual surveys of field history and weed flora data by aggregating annual data from numerous similar fields. This was combined with uncertainty analyses and the use of the sole complete site (i.e. Epoisses) to evaluate the consequences of input data missing on the other sites (e.g. initial seed bank, actual weather series) to assess how much prediction error at the incomplete sites was actually due to missing input data and not to model deficiencies. It was though impossible to analyse the impact of aggregating fields with different cropping system histories into a single scenario (as we did with the Biovigilance database) as the number of fields with a similar cropping system was too low at the well documented site (i.e. Epoisses). Moreover, only a single complete site was available, and conclusions based on site-specific variables (e.g. weed flora, weather) might thus not entirely apply to different regions.

Another major aspect of the present study was the focus on variability in both observed and simulated variables, resulting in a new evaluation indicator (e.g. the proportion of observations inside the simulated confidence interval) and the adaptation of existing ones (e.g. the normalization of rMSEP for both observed and simulated variability). This is necessary as a lot of weed variability remains unpredictable (Freckleton et al., 2008), even though FlorSys already integrates many causes of weed variability with mechanistic approaches (e.g. interactions between cultural techniques and pedoclimate) (Colbach, 2010; Colbach et al., 2014a). By working with both daily and multiannual variables and by evaluating the model's ability for predicting absolute values as well as for ranking

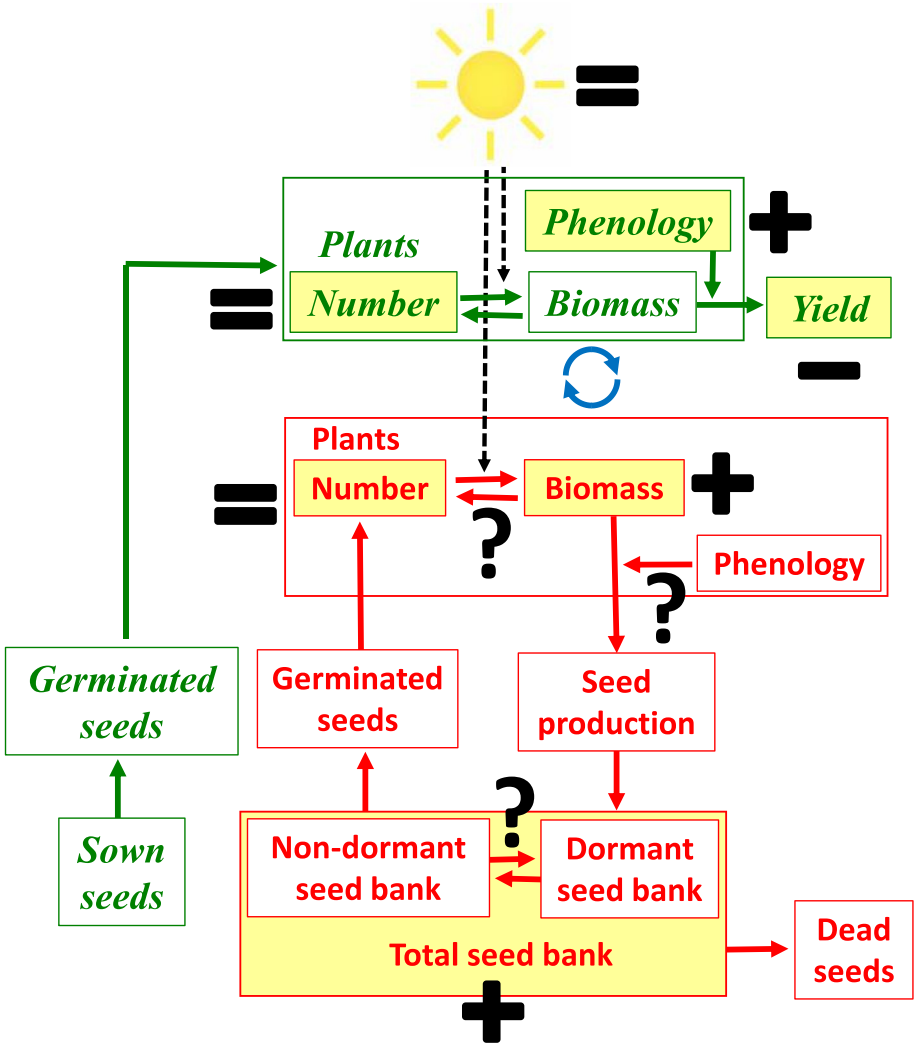

Fig. 6. Simplified representation of weed life-cycle (Plants) and crop cycle (*Plants*) in FlorSys indicating variables measured during evaluation (yellow-shaded) as well as variables well predicted (=), underestimated (−) and overestimated (+) by FlorSys for individual species. Light availability inside the 3D canopy simulated by FlorSys was evaluated in a previous work (Munier-Jolain et al., 2013). Based on this analysis, probable deficient functions were identified (**?**). (Nathalie Colbach © 2016). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

species or cropping systems correctly according to their weed infestation, we were able to assess the impact of the residual variability on weed dynamics: though residual variability influenced absolute weed densities, particularly at the daily time-scale, neither weed dynamics nor cropping system ranking were notably affected. Consequently, the major causes of weed variability interacting with cropping system seem already to have been integrated into FlorSys.

The last step consisted in proposing a graphical analysis of model outputs, inspired from Coucheney et al. (2015) but focusing on unpredictable variability. The latter was either due to processes occurring at spatial or organizational scales too fine for cropping system models (e.g. intra-field weed patch location, Rew, 2001) or at temporal scale too long for many weed monitoring plans (e.g. long-term seed survival, Gardarin et al., 2010). This led us (1) to analyse model prediction error vs. measurement error instead of plotting different types of prediction error as did Coucheney et al. (2015), (2) to distinguish situations according to the origin of measured data instead of only model output variables. Moreover, in order to focus on the main use of the model, i.e. to compare and rank cropping systems, we chose (3) to characterize variables and sites according to the model ranking ability (with Pearson correlation coefficients) instead of comparing the magnitude for observed and simulated data.

Because the model is a complex aggregation of sub-models and functions, there is a high risk of error compensation (e.g. overestimation of one sub-model is cancelled out by underestimation of another sub-model) and a great difficulty to pinpoint causes of prediction errors. This is why the present work was preceded by previous studies analysing sub-models individually and analysed several weed variables, synthesizing the results in a simplified life-cycle analysis.

4.2. How well is the functioning of the agro-ecosystem modelled?

Most evaluated weed models only concern part of the weed life-cycle, e.g. crop:weed competition (working with one crop and single, sown weed cohorts, Storkey and Cussans, 2007) or weed emergence (focusing on relative weed densities, Forcella et al., 1996). It is difficult to compare the prediction quality of FlorSys to other multi-specific weed dynamics models as very few were evaluated (Holst et al., 2007), usually for just one species (Colbach

Table 8
Summary of FlorSys evaluation results.

Variable	Prediction quality	Restriction of domain of validity
A. Domain of validity		
Daily weed variables		
Species densities (plants/m^2)	Well predicted	When the weed species most abundant in the region belongs to the 16 FlorSys species Possibly badly predicted in continuous no-till (only identified at one site) Only evaluated at two French Northern sites
Total densities (plants/m^2)	Underestimated	Only evaluated at two French Northern sites
Biomass per species (g/m^2)	Overestimated	}Only evaluated at one French Northern site where simulated species represented a major part in the regional flora
Total biomass (g/m^2)	Overestimated	
Seed bank (seeds/m^2)	Slightly overestimated	
Multi-annual weed variables (average over rotation)		
Species densities (plants·m^{-2}·year^{-1})	Well predicted	When the weed species most abundant in the region belongs to the 16 FlorSys species Possibly badly predicted in continuous no-till (only identified at one site) The further south a site is located from Burgundy, the lower prediction quality
Total densities (plants·m^{-2}·year^{-1})	Underestimated	Only evaluated at two French Northern sites
Crop variables		
Daily crop densities (plants/m^2)	Well ranked	Only evaluated at site where the dominant species was one of the 16 FlorSys species
Crop yield at harvest (T/ha)	Acceptable	When weed species density is corrected predicted Overestimated for oilseed rape and underestimated for barley and soybean Only evaluated at two French Northern sites, in non-limiting nitrogen and water conditions

Input/output variable	Advice
B. Precautions for model use	
Scale parameters	
Simulated field sample area	4 m × 2 m are sufficient in case of low densities and/or species numbers. Increase width (perpendicular to crop rows) in case of large interrow widths
Computer carrying capacity	Use 10000 plants/m^2 in case of high densities or species numbers. When reducing capacity to save simulation time, analyse simulated species and cropping system ranking rather than absolute values.
Voxel edge size	Not larger than 7 cm; always at least approximately half the smallest interrow width
Simulation plan	
Number of repetitions (for a given weather series)	Use at least ten repetitions to take account of stochastic effects in the model and obtain a realistic confidence interval for the simulated output
Number of weather repetitions	Use at least ten weather repetitions to take account of stochastic effects in the model and interactions between cropping systems and weather and obtain a realistic confidence interval for the simulated output and evaluate the robustness of cropping systems vs. weather
Simulated location	
Unknown initial seed bank	Use a "regional" seed bank, weighting weed seed biomasses (e.g. 125 mg seeds/m^2 per species) by the relative species frequency in the regional flora. Try not to overestimate competitive (tall) species. Species proportions are much more important than seed density. Analyse species and cropping system rankings rather than absolute values
Unknown soil characteristics	Use a similar soil Analyse species and cropping system rankings rather than absolute values
Southern latitudes	Use a corrective patch keeping weeds from flowering during winter in warm climates
Biotechnical input variables	
Weed distribution in field	Prefer aggregation in patches to uniform distribution (no difference in case of small simulated areas) If field areas larger than 18m^2, more than one patch per species might be necessary
Weed seed immigration	Allow annual random "regional" immigration (weighting weed seed biomass by the relative species frequency in the regional flora) or "functional" immigration (weighting weed seed biomass by their mean dispersal distance according to Thomson et al. (Flores-Moreno et al., 2013; Thomson et al., 2011)). Species proportions are much more important than seed density.
Herbicide spraying conditions	No difference between optimum and suboptimum spraying conditions (the same might not be true for bad conditions)
Simulated output variables	
Weed-impact indicators	Prefer indicators based on individuals (seeds or plants) rather than biomass Prefer indicators based on individual vs. total species variables

et al., 2007; Gonzalez-Andujar and Fernandez-Quintanilla, 1991; Munier-Jolain et al., 2002), over only few years (Munier-Jolain et al., 2002) and/or using very simple comparisons for only one output variable such as plotting simulated population growth rates calculated from final vs. initial seed banks as a function of observed final seed banks (Debaeke, 1988) or checking whether confidence intervals of predicted vs. observed seed bank densities overlap (Bohan et al., 2011b).

The present work went much further, running multiple analyses over several outputs to pinpoint possible missing processes. The results show that FlorSys already aggregates a large part of the major biophysical processes occurring in arable fields and correctly predicts weed dynamics if the major regional species are among the 16 FlorSys species, despite the numerous simplifications necessary for modelling complex systems. For instance, crop growth and development are modelled with a simplified 3D plant architecture, disregarding processes such as water stress, or nitrogen uptake and use (Munier-Jolain et al., 2013). This reduced prediction quality compared to species-specific weed dynamics model such as the *A. myosuroides* prototype of FlorSys (Colbach et al., 2007).

Despite these simplifications, crop yield prediction error (RMSEP = 3.3 T/ha, RRMSEP = 39%, EF = 0.56) was acceptable and not much worse than that of much more complex models such as STICS (RMSEP = 2.4 T/ha, RRMSEP = 30%, EF = 0.67, Coucheney et al., 2015), at least when the weed flora was correctly predicted. This was possibly because FlorSys was evaluated on a smaller range of pedoclimates and crop management strategies, focusing on trials

highlighting the major advantage of FlorSys over STICS, i.e. competition for light due to higher weed pressure. Conversely, processes disregarded by FlorSys probably presented negligible effects in these trials, particularly drought or nitrogen deficiencies, which are minor in the temperate climates of Northern France and in well-fertilized fields. Crop:weed competition for nitrogen might though have had an effect, as the yield of the most nitrophilous crop species (i.e. oilseed rape, Moreau et al., 2014) was systematically overestimated in simulations.

The main structural deficiency of FlorSys identified here resulted from the phenology sub-model solely based on temperature and seedling emergence dates, to the detriment of latitude. As many weeds are sensitive to day-length (Huang et al., 2012), weed densities were badly predicted at sites south of Burgundy because they flowered and matured too early in the simulations. Crop maturation was also affected, contributing to the lower prediction quality compared to STICS. The life-cycle analysis identified two other probable structural effects, i.e. overestimated light use efficiency in weeds and overestimated seed dormancy.

The analysis of residuals pointed to another possible structural defect. Indeed, prediction error for weed densities tended to increase in no-till fields, possibly because FlorSys disregards seed predation. This process can eliminate a large part of recent seed rains (Davis et al., 2011) and tends to be more frequent in untilled fields (Menalled et al., 2007). The evaluation of the monospecific version of FlorSys though showed that an overestimated surface seed bank does not necessarily result in an overestimated seedling density (Colbach et al., 2006), possibly because of error compensation (e.g. the possibly overestimated seed dormancy mentioned above would keep the overestimated seed bank from emerging). However, while weed seed predation has been shown to be substantial in many situations (Davis et al., 2011), it has not been demonstrated to date that it actually affects weed dynamics though a slight correlation between carabid abundance and weed seed bank decline was reported (Bohan et al., 2011a). Possibly, the eaten seeds would have disappeared from the system even in the absence of predators because of processes such as fatal summer germination which are already included in FlorSys.

Moreover, adding new sub-models to FlorSys could result in "producing an obese monster too slow and too difficult to operate" (Colbach et al., 2014a) and in increasing the prediction error resulting from badly estimated or missing input variables and parameters, as demonstrated here. We must thus assess whether adding new processes to FlorSys is necessary for the model objective (Colbach, 2010).

Another major deficiency of FlorSys is the small number of included weed species compared to the large number of species actually observed in fields (Jauzein, 1995). Despite attempting to include contrasting species types, these were not sufficient to cover all ecological niches occurring in the tested conditions and could not compensate for the missing species, thus resulting in underestimated total weed densities. The missing species were probably negligible at sites such as Epoisses where the dominant species were simulated, as neither individual weed species densities nor crop yields were overestimated. The situation was different at sites where the dominant species were not simulated, such as La Cage, resulting in overestimated weed species densities.

4.3. How to use the model for practical advice?

The present study demonstrated that the model can be used for its main objective, i.e. to rank cropping systems and weed species correctly if the major regional species are among the simulated species. Indeed, the uncertainty analysis demonstrated that many difficult-to-estimate inputs have little impact on outputs. Our study also contributed to a set of rules for safely using the model (Table 8) and proposed a method to estimate the initial seed bank, a key input that is rarely known in farmers' fields. Absolute weed values (particularly at the finer, daily scale) require accurate inputs in terms of cultural practices and pedoclimatic conditions to predict interactions and the resulting variability in effects.

Despite the model not always producing accurate predictions, it can be used to evaluate cropping systems insofar as we now know in which conditions it produces accurate predictions and how accurate these predictions are. Simulations can go beyond field experiments, for instance by testing the experimented cropping systems over longer time spans, with different climates and/or analysing output variables that cannot be easily monitored in field, e.g. daily weed dynamics, weed seeds on soil surface for feeding birds or insects (Mézière et al., 2015).

5. Conclusion

The originality of the present paper was to evaluate a complex weed dynamics model with diverse and often incomplete data sets, developing a method that can make the essential but often difficult evaluation step easier in future. As a result, we were able to define the limits of the model's domain of validity and to propose a set of guidelines for using the model for evaluating cropping systems in terms of weed dynamics. Moreover, we identified a few missing or inadequately represented processes that are pertinent for understanding and predicting weed dynamics in arable cropping systems. The latter aspect needs though to be investigated further with a global sensitivity analysis to model inputs and parameters, to understand which processes and species traits are essential for weed dynamics prediction.

Software availability

FlorSys is a free software coded in C++, available under a licence agreement at the corresponding author's and soon via the RECORD modelling platform (https://www6.inra.fr/record), together with a set of parameters for 16 major weed species and most of major French arable crops.

Acknowledgements

This project was supported by the research programme "Assessing and reducing environmental risks from plant protection products" funded by the French Ministries in charge of Ecology and Agriculture (N° 29000683), the EU project AMIGA (Assessing and Monitoring Impacts of Genetically modified plants on Agro-ecosystems, FP7-KBBE-2011-5-CP-CSA), and the French project CoSAC (ANR-14-CE18-0007). The field trials were conducted at the INRA experimental stations of Dijon-Époisses (Philippe Chamoy and Alain Berthier), and Versailles-La Cage (Patrick Saulas and UE Grandes Cultures). In addition to the authors, many people contributed to weed and crop monitoring over the years, among which Émilie Cadet, Éric Vieren, Maurice Bourlier, and Noureddine El-Mjiyad (INRA Dijon). The Biovigilance-Flore network was funded by the French Ministry of Agriculture and the database management was partially supported by ANR OGM VIGIWEED (ANR-07-POGM-003-01). The authors are grateful to the RMT Modelia (http://www.modelia.org/moodle/) and Manuel Plantegenest (INRA Rennes) for efficient statistical advice, as well as to Anne-Sophie Voisin (INRA Dijon) and several anonymous reviewers for their help in improving the present paper.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.envsoft.2016.09.020.

References

Barralis, G., 1976. Méthode d'étude des groupements adventices des cultures annuelles: application à la Cote d'Or. In: Vème Colloque International sur l'Écologie et la Biologie des Mauvaises Herbes: Dijon, France, pp. 59–68.

Bohan, D.A., Boursault, A., Brooks, D.R., Petit, S., 2011a. National-scale regulation of the weed seedbank by carabid predators. J. Appl. Ecol. 48, 888–898.

Bohan, D.A., Powers, S.J., Champion, G., Haughton, A.J., Hawes, C., Squire, G., Cussans, J., Mertens, S.K., 2011b. Modelling rotations: can crop sequences explain arable weed seedbank abundance? Weed Res. 51 (4), 422–432.

Boissinot, F., Mézière, D., Bretagnolle, V., Munier-Jolain, N.M., 2011. Réduire l'usage des herbicides en grandes cultures. Phytoma 649, 39–44.

Brisson, N., Mary, B., Ripoche, D., Jeuffroy, M.H., Ruget, F., Nicoullaud, B., Gate, P., Devienne-Barret, F., Antonioletti, R., Dürr, C., Richard, G., Beaudoin, N., Recous, S., Tayot, X., Plenet, D., Cellier, P., Machet, J.M., Meynard, J.M., Delecolle, R., 1998. STICS: a generic model for the simulation of crops and their water and nitrogen balances. I. Theory and parameterization applied to wheat and corn. Agronomie 18 (5–6), 311–346.

Chatelin, M.H., Aubry, C., Poussin, J.C., Meynard, J.M., Masse, J., Verjux, N., Gate, P., Le Bris, X., 2005. DéciBlé, a software package for wheat crop management simulation. Agric. Syst. 83 (1), 77–99.

Chikowo, R., Faloya, V., Petit, S., Munier-Jolain, N., 2009. Integrated Weed Management systems allow reduced reliance on herbicides and long term weed control. Agriculture. Ecosyst. Environ. 132, 237–242.

Colbach, N., 2010. Modelling cropping system effects on crop pest dynamics: how to compromise between process analysis and decision aid. Plant Sci. 179, 1–13.

Colbach, N., Biju-Duval, L., Gardarin, A., Granger, S., Guyot, S.H.M., Mézière, D., Munier-Jolain, N.M., Petit, S., 2014a. The role of models for multicriteria evaluation and multiobjective design of cropping systems for managing weeds. Weed Res. 54, 541–555.

Colbach, N., Busset, H., Yamada, O., Dürr, C., Caneill, J., 2006. ALOMYSYS: modelling black-grass (*Alopecurus myosuroides* Huds.) germination and emergence, in interaction with seed characteristics, tillage and soil climate. II. Evaluation. Eur. J. Agron. 24, 113–128.

Colbach, N., Chauvel, B., Gauvrit, C., Munier-Jolain, N.M., 2007. Construction and evaluation of ALOMYSYS, modelling the effects of cropping systems on the blackgrass life-cycle. From seedling to seed production. Ecol. Model. 201, 283–300.

Colbach, N., Collard, A., Guyot, S.H.M., Mézière, D., Munier-Jolain, N.M., 2014b. Assessing innovative sowing patterns for integrated weed management with a 3D crop:weed competition model. Eur. J. Agron. 53, 74–89.

Colbach, N., Granger, S., Munier-Jolain, N.M., 2009. Using weed dynamics models for evaluating and developing integrated cropping systems. In: XIIIème Colloque international sur la biologie des mauvaises herbes: Dijon, France, pp. 195–205.

Coucheney, E., Buis, S., Launay, M., Constantin, J., Mary, B., García de Cortázar-Atauri, I., Ripoche, D., Beaudoin, N., Ruget, F., Andrianarisoa, K.S., Le Bas, C., Justes, E., Léonard, J., 2015. Accuracy, robustness and behavior of the STICS soil–crop model for plant, water and nitrogen outputs: evaluation over a wide range of agro-environmental conditions in France. Environ. Model. Softw. 64 (0), 177–190.

Davis, A.S., Deadlow, D., Schutte, B.J., Westerman, P.R., 2011. Temporal scaling of episodic point estimates of seed predation to long-term predation rates. Methods Ecol. Evol. 2, 682–690.

Debaeke, P., 1988. Modelling the long-term evolution of the weed flora. II. Application to three annual broad-leaved weeds on a given site. Agronomie 8 (9), 767–777.

Debaeke, P., Munier-Jolain, N.M., Betrand, M., Guichard, L., Nolot, J.M., Faloya, V., Saulas, P., 2009. Iterative design and evaluation of rule-based cropping systems: methodology and case studies. A review. Agron. Sustain. Dev. http://dx.doi.org/10.1051/agro:2008050.

Flores-Moreno, H., Thomson, F.J., Warton, D.I., Moles, A.T., 2013. Are introduced species better dispersers than native Species? A global comparative study of seed dispersal distance. PLoS ONE 8 (6), e68541.

Forcella, F., King, R.P., Swinton, S.M., Buhler, D.D., Gunsolus, J.L., 1996. Multi-year validation of a decision aid for integrated weed management in row crops. Weed Sci. 44, 650–661.

Freckleton, R.P., Sutherland, W.J., Watkinson, A.R., Stephens, P.A., 2008. Modelling the effects of management on population dynamics: some lessons from annual weeds. J. Appl. Ecol. 45 (4), 1050–1058.

Fried, G., Norton, L.R., Reboud, X., 2008. Environmental and management factors determining weed species composition and diversity in France. Agriculture. Ecosyst. Environ. 128 (1–2), 68–76.

Gardarin, A., Dürr, C., Colbach, N., 2012. Modeling the dynamics and emergence of a multispecies weed seed bank with species traits. Ecol. Model. 240, 123–138.

Gardarin, A., Dürr, C., Mannino, M.R., Busset, H., Colbach, N., 2010. Seed mortality in the soil is related to the seed coat thickness. Seed Sci. Res. 20, 243–256.

Gonzalez-Andujar, J.L., Fernandez-Quintanilla, C., 1991. Modelling the population dynamics of Avena sterilis under dry-land cereal cropping systems. J. Appl. Ecol. 28 (1), 16–27.

Holst, N., Rasmussen, I.A., Bastiaans, L., 2007. Field weed population dynamics: a review of model approaches and applications. Weed Res. 47, 1–14.

Huang, H.-R., Yan, P.-C., Lascoux, M., Ge, X.-J., 2012. Flowering time and transcriptome variation in Capsella bursa-pastoris (Brassicaceae). New Phytol. 194 (3), 676–689.

Jauzein, P., 1995. Flore des champs cultivés. INRA-Sopra.

Mayer, D.G., Butler, D.G., 1993. Statistical validation. Ecol. Model. 68, 21–32.

McKay, M.D., Beckman, R.J., Conover, W.J., 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21, 239–245.

Menalled, F.D., Smith, R.G., Dauer, J.T., Fox, T.B., 2007. Impact of agricultural management on carabid communities and weed seed predation. Agric. Ecosyst. Environ. 118 (1–4), 49–54.

Mézière, D., Petit, S., Granger, S., Biju-Duval, L., Colbach, N., 2015. Developing a set of simulation-based indicators to assess harmfulness and contribution to biodiversity of weed communities in cropping systems. Ecol. Indic. 48, 157–170.

Moreau, D., Busset, H., Matejicek, A., Munier-Jolain, N., 2014. The ecophysiological determinants of nitrophily in annual weed species. Weed Res. 54, 335–346.

Munier-Jolain, N.M., Chauvel, B., Gasquez, J., 2002. Long-term modelling of weed control strategies: analysis of threshold-based options for weed species with contrasted competitive abilities. Weed Res. 42, 107–122.

Munier-Jolain, N.M., Collard, A., Busset, H., Guyot, S.H.M., Colbach, N., 2014. Modelling the morphological plasticity of weeds in multi-specific canopies. Field crops Res. 155, 90–98.

Munier-Jolain, N.M., Guyot, S.H.M., Colbach, N., 2013. A 3D model for light interception in heterogeneous crop:weed canopies. Model structure and evaluation. Ecol. Model. 250, 101–110.

Oerke, E., 2006. Crop losses to pests. J. Agric. Sci. 144, 31–43.

Rew, L.J.C.R.D., 2001. Spatial distribution of weeds in arable crops: are current sampling and analytical methods appropriate? Weed Res. 41, 1–18.

Storkey, J., Cussans, J.W., 2007. Reconciling the conservation of in-field biodiversity with crop production using a simulation model of weed growth and competition. Agric. Ecosyst. Environ. 122, 173–182.

Thomson, F.J., Moles, A.T., Auld, T.D., Kingsford, R.T., 2011. Seed dispersal distance is more strongly correlated with plant height than with seed mass. J. Ecol. 99 (6), 1299–1307.

Wallach, D., 2006. Evaluating crop models. In: Wallach, D., Makowski, D., Jones, J.W. (Eds.), Working with Dynamic Crop Models: Evaluating, Analyzing, Parameterizing and Using Them. Éditions INRA, Paris, pp. 11–53.

Wallach, D., Goffinet, B., 1987. Mean squared error of prediction in models for studying ecological and agronomic systems. Biometrics 43, 561–573.

Wallach, D., Goffinet, B., 1989. Mean squared error of prediction as a criterion for evaluating and comparing system models. Ecol. Model. 44, 299–306.

Willmott, C.J., 1981. On the validation of models. Phys. Geog. 2, 184–194.

Yang, J.M., Yang, J.Y., Liu, S., Hoogenboom, G., 2014. An evaluation of the statistical methods for testing the performance of crop models with observed data. Agric. Syst. 127 (0), 81–89.