

Steady-state soil organic matter approximation model: application to the Pasture Simulation Model

Romain Lardy^{a,b,c}, Anne-Isabelle Graux^a, Bruno Bachelet^{b,c}, David R.C. Hill^{b,c}, Gianni Bellocchi^a

^a*French National Institute for Agricultural Researches, Grassland Ecosystems Research Unit, 234 Avenue du Brézat, 63100 Clermont-Ferrand, France*

^b*Clermont University, Blaise Pascal University, LIMOS - UMR 6158CNRS, 63173 Aubière, France*

^c*ISIMA, Computer Science & Modelling Institute, BP 10125, F-63177*

Email addresses: rlardy@clermont.inra.fr; graux@clermont.inra.fr; drch@isima.fr; bachelet@isima.fr; gianni.bellocchi@clermont.inra.fr

Abstract: Spin-up runs usually used to initialize mechanistic biogeochemical models highly increase the time required to make simulations. The aim of this paper is to evaluate the use of linear and quadratic regression models, as an alternative way to initialize such models. This option is illustrated with the grassland ecosystem Pasture Simulation model (PaSim) under a range of climate, soil and management conditions. Coupled to the CENTURY model for the soil processes, PaSim simulates fluxes of C, N, water and energy at the soil-plant-animal-atmosphere interface for managed grasslands at the plot scale. This study demonstrates the feasibility of approximating steady state SOM (Soil Organic Matter) by a quadratic regression. For instance, PaSim initialization using a quadratic regression with P-ET₀ (Climatic Water Balance indicator) is about 500 times faster than using spin-up runs. However, quadratic SOM regression provides a 10-15% gap, due to the existing variability in SOM response to climate (e.g. ~7% standard deviation for one P-ET₀ value of the climate year). Anyway, these quadratic regressions could be used in future vulnerability assessments that require a prohibitive number of simulations for complex models.

Keywords: Grassland; metamodel; soil organic matter; steady-state

1. INTRODUCTION

Soil organic matter (SOM) is the primary reservoir in grasslands of organic carbon (C) and nitrogen (N) and plays a key role in mitigating GHG (Green House Gas) emissions. Classical biogeochemical models incorporate a mechanistic view of the SOM dynamics and provide a sound basis for development of generalised response signals of SOM pools (Guo and Giffort [2002]). There is a concern that the different SOM pools are vulnerable to future warming (IPCC [2001]), and this emphasizes the need for understanding SOM dynamics.

There is a great number of relevant parameters influencing climate change vulnerability, and a high level of uncertainties in climate change impact studies (e.g. emission scenarios, climate modelling, downscaling and initialization, and modelling of the impacts on a target system). So assessing the vulnerability of soil stocks needs many simulation runs to gain an accurate understanding of the influence of environmental variables. This suggests that a pertinent Design of

Experiment (DOE) should be employed to reduce the time required for simulations (Lardy et al. [2011a]). Such an amount of simulations also needs high performance computing resources to scatter simulations in distributed environments.

However, due to the often lack of experimental data and/or the difficulty to link up measurements with SOM input values, a common way to initialize models is to put them at equilibrium with climate and management. For that initialization, spin-up runs are usually performed to bring the soil C and N pools to steady-state (e.g. millennia, Wutzler and Reichstein [2007]). To avoid the prohibitive computational time required by this conventional approach, we can build response surfaces (i.e. metamodels) of SOM, i.e. approximations of the relationship between inputs and outputs in much simpler terms than the full simulation model (Kleijnen et al. [2005]). In this study, we document the creation of such response surfaces simulated under a range of climate, soil and management conditions in France by the Pasture Simulation Model (PaSim, Riedo et al. [1998]). This study is also a first step in SOM vulnerability assessment, taking into account different sources of uncertainties (Lardy et al. [2011a]), such as climate, management, soil and plant species.

The next section presents the grassland model used in this study. The third section describes the DOE and the steps taken to prepare and conduct the experiment. Then we analyze the results and discuss them. The concluding section identifies key results and explores future research needs.

2. MODEL DESCRIPTION

The Pasture Simulation model (PaSim, Riedo et al. [1998]) is a multi-year, plot-scale, biogeochemical model to simulate water, C and N cycles in grassland systems on a daily to sub-daily time step. Soil processes are based on the CENTURY model of Parton et al. [1988]. Photosynthetic-assimilated C is either respired or allocated dynamically to one root and to three shoot compartments. Accumulated aboveground biomass is used by either cutting or grazing, or enters a litter pool. Soil organic Carbon (SOC) is represented in three pools (active, slow and passive) with different potential decomposition rates, while above and belowground plant residues and organic excreta are partitioned into structural and metabolic pools. The N cycle considers three types of N inputs to the soil via atmospheric N deposition, fertilizer N addition, and symbiotic N fixation by legumes. The inorganic soil N is available for root uptake and may be lost through leaching, ammonia volatilization and nitrification/denitrification, the latter processes leading to nitrous oxide (N₂O) gas emissions to the atmosphere. Management includes N fertilization, mowing and grazing and can either be set by the user or optimized by the model (Vuichard et al., [2007]; Graux [2011]). The vegetation is simulated at the community scale without accounting for species interactions. In this PaSim version, nitrogen fixation is simulated by assuming a constant legume fraction. The animal module [Graux et al., 2011] simulates the performance of grazing ruminants (suckler cows with calves, dairy cows and heifers) in response to climate and management. This version 5.3 of PaSim was used in this study, with the algebraic method for equilibrium search method (Lardy et al. [2011b]) already developed to reduce computation time.

3. DESIGN OF EXPERIMENT

Design of experiments (DOE) has a rich history, with many theoretical developments and practical applications in a variety of fields. Since the beginning of computer simulation, DOE has been an active research field (Kempthorne [1952]; Amblard et al. [2003]). In the modelling field, DOE is a needed tool for

efficiently testing and analysing the behaviour of a model (Kleijnen [1987]). Most of model simulations aim at exploring and/or testing the behaviour of the model. A parameter or a model input is called a factor in the DOE terminology, and it could either be qualitative or quantitative (Kleijnen et al. [2005]). Each factor can take two or more values, called levels. An experimental design is a combination of factor levels.

In environmental dynamics modelling, models became increasingly more complex at the pace of the growth of computational power. Due to the high number of model parameters and the computation time required for a single run, the needed time by a sequential machine is usually too expensive for a full factorial DOE. This implies that, first of all smart but less complete DOE are used and, on top of that, distributed computing is required. The use of a proper DOE will help to get, firstly, all the information we are looking for. The second point is to have the smallest number of simulation runs for a desired accuracy, which implies the optimization of the total computation time. Computation time is then considerably reduced thanks to the distribution of processes on parallel architectures.

3.1. Factor choice and simulation domain

The size of the simulation domain was reduced by considering exemplary climate and management conditions in France. Similarly, the number of potential agricultural practices was restrained to mown grasslands. Default vegetation parameters were used (see Riedo [1998]). An experiment was run based on the combination of three kinds of factors:

- Soil. We used 102 dominant grassland soils of France (through the French ANR 'VALIDATE' project <http://www1.clermont.inra.fr/validate>), characterized by texture (silt, sand and clay), depth, bulk density and pH. Other soil characteristics were not considered in the design as they are highly linked to these four properties. The data do not contain the whole combinations of depth x texture x density x pH, thus a qualitative value was assigned to each soil (i.e. one number per soil) for the purpose of DOE's conception. On the other hand, soil properties were used for response surface estimation.
- Climate. For each spin-up run, we repeated a cycle of three years until the equilibrium is reached. One can show that three years is a good compromise between speed and performance (data not shown). Climate data are from 1970 (or later) to 2006, at 12 sites, representative of the climate in France (408 years of hourly weather data), provided by the French ANR 'CLIMATOR' project (http://w3.avignon.inra.fr/projet_climator). Due to the complexity of choosing an array of variables representative of the climate, and in order to have a consistent coverage of the domain, climate was considered as a qualitative trait for the conception of the DOE. We also added atmospheric CO₂ concentration and site elevation to the design.
- Management. Constant over the three years of the spin-up runs, different agricultural practices depend on the number of cuts and nitrogen fertilization rates. Systems were analyzed with one to four cutting events per year, with about one month interval between each one. Fertilization rates (in the form of ammonitrate) varied from 0 to 120 kg N ha⁻¹ per year, applied 30 days before the first cutting event and two days after other cuts (except after the last cut). The same amount of N was supplied at each fertilization event under condition that a minimum of 40 kg N ha⁻¹ is provided at each event. The presence of legumes in the sward is also linked to practices, and was added to the design (in the form of fraction in the sward).

3.2. Choice of the experimental design

Due to the number of factors and their levels in the study (Table 1), most classical designs (e.g. factorial design) are not directly applicable. We decided to use a Latin

Hypercube Design (LHD) for its good space-filling properties with a relatively low number of sample points. The size of LHD is the least common multiple of the levels (i.e. 157080). A simple transformation was applied on each factor to get the correct level. We refined it by optimizing the ‘maximin’ criterion (Johnson et al. [1990]) over 100 designs.

Table 1. DOE Factors and levels

Name of the factor	Number of levels	Information
Climate year 1	408	Three levels of input are considered in the response surface regression
Climate year 2	408	
Climate year 3	408	
Atmospheric CO ₂ concentration	8	320–390 ppm
Soil	102	Depth, bulk density, pH and texture are used in response surface regression
Number of cuts	4	1; 2; 3; 4
1 st cutting date	7	1 st April to 1 st May
2 nd cutting date	7	16 th May to 15 th June
3 rd cutting date	7	1 st July to 31 st July
4 th cutting date	7	16 th August to 15 th September
Legume fraction	5	0; 10; 20; 30; 40%
Nitrogen fertilization rate	6	0; 40; 60; 80; 100; 120 kg N ha ⁻¹ yr ⁻¹
Elevation	20	50–1000 m a.s.l.

3.3. Regression methods

The design previously described, allows us to build a response of the SOM to climate, soil and management. In our case, we restricted our study to the simplest approaches, i.e. linear and quadratic regressions, with different climate entries. Indeed, we used three alternative models according to temporal grain:

- a single variable per year: $P-ET_0$, where P is the annual precipitation sum, and ET_0 the annual sum of the reference potential evapotranspiration by Allen et al. [1998], adjusted of the CO₂ effect (Olioso et al. [2010])
- five variables per season: average air temperature, average global radiation, average air humidity, average wind speed and precipitation sum
- five variables per month: average air temperature, average global radiation, average air humidity, average wind speed and precipitation sum

As the order of the years is negligible on the total SOM (<1%, data not shown), years were sorted by aridity conditions for the regression (i.e. by $P-ET_0$ in the first case, and by the Martonne-Gottman index (De Martonne [1942]) in the two other cases. All the regressions and statistical tests were done using “biglm” package by Lumley [2011] in the software R [2011].

4. RESULTS AND DISCUSSION

We launched simulations on the Biomed VO grid, thanks to the OpenMOLE software by Reuillon et al [2010], in order to reduce the computing time required (about four years on a single modern CPU) by the number of runs involved.

The goodness of SOM regressions was analyzed by three performance indices, the Relative Root Mean Squared Error (RRMSE, Table 2), the Root Mean Squared Relative Error (RMSRE, Table 3), and the modelling efficiency (EF, Table 4). All these indices use the comparison between values predicted by regression (P_i) and values given by PaSim (O_i) for each i^{th} simulation in the design. For RRMSE and RMSRE, the nearer to zero they are, the better the prediction is, whereas best predictions are at one for EF. All the indices calculated for soil organic nitrogen

showed the same pattern as those calculated for SOC (data not shown). The results for SOC (Table 2, Table 3, Table 4) clearly show that climate information contained in the annual variable P-ET₀ is sufficient for a linear regression of SOC pools at equilibrium. Indeed, the RRMSE decreases when moving from annual (P-ET₀) to seasonal and monthly variables (e.g. radiation, temperature), because the addition of explanatory variables reduces the squared errors (and improve the efficiency) of least-squares regressions. However, the RMSRE, which gives more weight to the relative error and less to the total amount of the error, is only improved in five of 22 cases.

Table 2. RRMSE ("Relative Root Mean-Squared Error") values of linear and quadratic regressions of five carbon (C) pools and their totals for three climatic input levels (P-ET₀, Season or Month) on design points.

$$RRMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}} \times 100 ; \text{ where } P_i \text{ is the } i^{\text{th}} \text{ prediction by the regression model and } O_i \text{ the } i^{\text{th}} \text{ corresponding value by PaSim}$$

	Linear			Quadratic		
	P-ET ₀	Season	Month	P-ET ₀	Season	Month
Metabolic C	38.87	36.21	36.09	34.17	29.14	27.78
Structural C	30.25	28.81	28.75	21.56	16.60	15.69
Active C	27.54	26.59	26.49	18.34	14.99	14.08
Slow C	21.03	19.85	19.75	14.99	11.18	10.38
Passive C	20.89	19.71	19.62	14.85	11.04	10.25
Total C	21.39	20.12	20.03	15.26	11.20	10.42

Table 3. RMSRE ("Root Mean-Squared Relative Error") values of linear and quadratic regressions of five carbon (C) pools and their totals, for three climatic input levels (P-ET₀, Season or Month) on design points.

$$RMSRE = \sqrt{\frac{\sum_{i=1}^n \left(\frac{P_i - O_i}{O_i}\right)^2}{n}} \times 100 ; \text{ where } P_i \text{ is the } i^{\text{th}} \text{ prediction by the regression model and } O_i \text{ the } i^{\text{th}} \text{ corresponding value by PaSim}$$

	Linear			Quadratic		
	P-ET ₀	Season	Month	P-ET ₀	Season	Month
Metabolic C	83.05	81.61	81.44	60.95	56.70	54.59
Structural C	38.99	39.95	39.94	21.32	19.98	19.55
Active C	35.77	36.06	36.05	21.94	20.67	19.71
Slow C	28.47	28.55	28.49	17.33	15.39	14.42
Passive C	28.38	28.44	28.37	17.10	15.11	14.15
Total C	28.62	28.75	28.69	17.23	15.20	14.27

Table 4. Modelling Efficiency values of linear and quadratic regressions of five carbon (C) pools and their totals, for three climatic input levels (P-ET₀, Season or

Month) on design points. $Efficiency = 1 - \frac{\sum_{i=1}^n (P_i - O_i)^2}{\sum_{i=1}^n (\bar{O} - O_i)^2}$; where P_i is the ith prediction by the regression model and O_i the ith corresponding value by PaSim

	Linear			Quadratic		
	P-ET ₀	Season	Month	P-ET ₀	Season	Month
Metabolic C	0.183	0.360	0.367	0.468	0.662	0.702
Structural C	0.662	0.703	0.704	0.853	0.918	0.927
Active C	0.609	0.645	0.649	0.858	0.909	0.921
Slow C	0.688	0.731	0.735	0.863	0.928	0.939
Passive C	0.687	0.731	0.734	0.863	0.929	0.939
Total C	0.695	0.739	0.742	0.865	0.931	0.941

The transition from a linear to a quadratic regression can considerably improve performance, regardless of the three statistical criteria considered (e.g. efficiency increases from ~ 0.7 to ~ 0.9 for the total soil C). This improvement is not solely due to the increased number of explanatory variables. Indeed, the quadratic regression with P-ET₀ contains 172 against 196 variables in the linear regression

with monthly climatic input. Unlike linear regression, moving from an annual climate variable to five seasonal variables causes an improvement of the quadratic regression (e.g., RRMSE decreased from 15% to 11% for total soil C).

The poor quality of the regression of metabolic compartment is partly due to the rapid turnover (0.5 years) (Parton et al. [1988]) of this component, and its strong relationship with the state of plant biomass. Given its rapid turnover, this compartment is more influenced by last year climate than by the full climatic cycle. Indeed, the regressions were improved by the use of unsorted years [data not shown], with an efficiency of 0.629 against 0.367 for a linear regression using monthly climate data. However, the prediction gain is only apparent for the metabolic compartment, which represents on average only 0.89% of the total soil biomass. The lack of sorting led to a slight drop in the prediction accuracy (e.g. 0.71 vs. 0.69 for the efficiency of the total organic matter with P-ET₀).

PaSim showed a failure rate of 4.0‰, only on alkaline soil ($8 \leq \text{pH} \leq 8.5$), where numerical instability in computing the ammonia concentration was exacerbated by dry climate and intensive management. It is also interesting to note that 0.95‰ of the equilibria do not exist (even after a 100 cycles with the matrix equilibrium search (Lardy et al. [2011b])). These situations are characterized by more cutting events than on average (ca. 2.8 vs. 2.5), lower fraction of legumes (on average, 10% vs. 20% for the overall design) and lower fertilization rates (46 kg N ha⁻¹ vs. an average 66 kg N ha⁻¹). These conditions are unsustainable for grasslands, turning out into excessive exploitation (mowing) in relation to resources (resource-poor nitrogen), and causing lack of balance if climate and management were extended indefinitely. Indeed, in the absence of any equilibrium the model slowly moves to herbage zero biomass.

Although SOM decomposition rates are temperature dependent in the model, elevation does not significantly change SOM values in this study, as the effects of elevation and climate were tested independently (the elevation effect was tested without changing input climate data and conversely). As expected, atmospheric CO₂ concentration, legume fraction and the amount of fertilizer applied have a positive effect on the organic matter, because they promote plant growth and thus increase the inflow of SOM. Similarly, mowing has a negative effect due to the export of material. As the model simulates SOM throughout the soil profile, it makes sense to find a positive effect of soil depth, as for the bulk density which has an overall positive effect. The increase in each of the texture fractions has a negative effect, however, which is lower for silt content. It is interesting to note that the effect of the agrometeorological indicator P-ET₀ has a positive effect on organic matter (in fact, an arid climate results in decreased productivity of the grassland associated with reduced flow). Moreover the effect of the two driest years is almost twice as much as the wettest year (standardized regression coefficients, for the total carbon, with linear regression with P-ET₀: 6.0, 5.2, 3.0 kg C m⁻² mm⁻¹). These results show that the behaviour of the PaSim model is coherent with the state-of-the-art of plant-soil interactions.

We can observe (Figure 1) that the response of SOC at equilibrium to climate is noisy. Indeed, we can easily cover a quite big range of SOC values for each value of P-ET₀ (of a single meteorological year). We can evaluate on this two examples, a standard deviation of 5-7%. This can be explained by the fact that *in situ* intra-annual vegetation dynamics contain threshold effects, for example phenology, and that there are non-linearities in the equations (e.g. energy budget) and inputs (e.g. precipitations). Somehow, it proves the interest of biogeochemical models use (compared to approximation models) to simulate SOM dynamics. It shows that a model like PaSim is able to produce information that cannot be fully reproduced by regressions. Moreover, with regression models, it is harder to detect the lack of equilibrium than with usual equilibrium simulations. Indeed, regressions can

provide SOM equilibrium values even if there is almost no more herbage biomass (unrealistic situations).

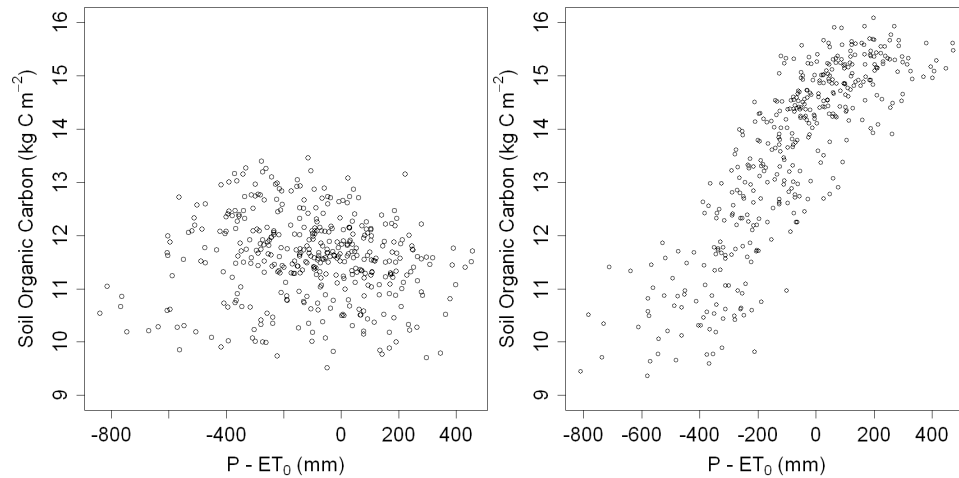


Figure 1. Influence of the change of one climatic year on SOC at equilibrium. Left: extensive, arid grassland on shallow soil; right: humid, intensive grassland on deep soil.

5. CONCLUSION

This study demonstrates the feasibility of approximating SOM at equilibrium by a quadratic regression in order to faster initialize biogeochemical mechanistic models. Using $P-ET_0$ may be sufficient to capture the climate input information. For example, initialization by quadratic regression with $P-ET_0$ appears to be about 500 times faster than PaSim spin-up run, with a 10-15% gap due to the variability in SOM response to climate (e.g. ~7% standard deviation for one of the three climate year). Vulnerability analysis is a multi-step process, which includes model-based sensitivity and uncertainty analyses at different levels (e.g. Lardy et al. [2011a]), and thus requires a huge amount of simulation runs. The present quadratic regressions could be also used in sensitivity analyses that require prohibitive number of simulations for complex models. Moreover, the current study may be extended to projections of future climate conditions, thereby allowing calculating vulnerability indices on the potential stock of organic matter (i.e. at steady-state) using current climate as baseline. Similarly, developments in the field of non-existence of equilibrium may be of interest. So, for further investigations, it may be necessary to prove that this approximation model contains specific trends and global maxima and minima. The authors also intend to test alternative approximation models, for instance based on the Kriging method.

ACKNOWLEDGMENTS

This work was supported by the EC FP7 'GHG-Europe IP' project. Climate data was provided by the French ANR 'CLIMATOR' project and soil data by the French ANR 'VALIDATE' project.

REFERENCES

- Allen, R.G., Pereira, L.S., Raes, D., Smith, M., Crop evapotranspiration: Guidelines for computing crop water requirements. *Irrigation and Drainage Papers* 56. UN-FAO, Rome, Italy, 1–15, 1998.
- Amblard, F., Hill, D.R.C., Bernard, S., Truffot, J., Deffuant, G., MDA Compliant Design of SimExplorer A Software Tool to Handle Simulation Experimental

- Frameworks. Summer Computer Simulation Conference, Montreal, Canada, 279-284, 2003.
- De Martonne, E., Nouvelle carte mondiale de l'indice d'aridité. *Annales de Géographie* 51, 242–250, 1942.
- Guo, L.B., Gifford, R.M., Soil carbon stocks and land-use change: a meta-analysis. *Global Change Biology*, 8, 345–360, 2002.
- Graux, A.-I., Modeling climate change impacts on grassland ecosystems. Adaptation strategies for forage production systems, Blaise Pascal University, Ph.D. thesis, 535 p., 2011.
- Graux, A.-I., Gaurut, M., Agabriel, J., Baumont, R., Delagarde, R., Delaby, L., Soussana, J.-F., Development of the Pasture Simulation Model for assessing livestock production under climate change, *Agriculture, Ecosystems & Environment*, 144(1), 69–91, 2011.
- IPCC (Intergovernmental Panel on Climate Change). *Impacts, adaptation, and vulnerability climate change 2001. Third Assessment Report of the IPCC*. University Press, Cambridge, UK. 1032 p., 2001
- Johnson, M.E., Moore, L.M., Ylvisaker, D., Minimax and maximin distance designs, *Journal of Statistical Planning and Inference*. 26, 131–148, 1990.
- Kempthorne, O., *The Design and Analysis of Experiments*. Wiley, Publications in Statistics, 631 p., 1952.
- Kleijnen, J.P.C., *Statistical tools for simulation practitioners*. New York: Marcel Dekker, 429 p., 1987.
- Kleijnen, J.P.C., Sanchez, S.M., Lucas, T.W., and Cioppa, T.M., State-of-the-Art Review: A user's guide to the brave new world of designing simulation experiments. *INFORMS Journal on Computing*, 17, No.3, pp. 263–289, 2005.
- Lardy, R., Bellocchi, G., Bachelet B., Hill, D.R.C., Climate change vulnerability assessment with constrained design of experiments, using a model driven engineering approach. ESM'2011, The European Simulation and Modelling Conference 2011, Oct. 24-26, Guimarães, Portugal, 354–362, 2011a
- Lardy, R., Bellocchi, G., Soussana, J.-F., A new method to determine soil organic carbon equilibrium, *Environmental Modelling & Software*, 26, 1759–1763, 2011a.
- Lumley, T., biglm: bounded memory linear and generalized linear models. R package version 0.8., <http://CRAN.R-project.org/package=biglm>, 2011
- Olioso, A., Huard, F., Guillioni, L., Prise en compte du CO₂ dans le calcul de l'évapotranspiration de référence. Présentation des méthodes et des résultats du projet CLIMATOR, Versailles, France, June 17-18, 66–67, 2010.
- Parton, W.J., Stewart, J.B.W., Cole, C.V., Dynamics of C, N, P and S in grassland soils: a model. *Biogeochemistry*, 5, 109–131, 1988.
- Reuillon, R., Chuffart, V., Leclaire, M., Faure, T., Dumoulin, N., Hill, D.R.C., Declarative task delegation in OpenMole. In proceedings of *International Conference on High Performance Computing and Simulation*, Caen, June 28–July 2, pp. 55–62, 2010
- Riedo, M., Grub, A., Fuhrer, J., A pasture simulation model for dry matter production, and fluxes of carbon, nitrogen, water and energy. *Ecological Modelling* 105, 141–183, 1998.
- R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2011
- Vuichard, N., Ciais, P., Viovy, N., Calanca, P., Soussana, J.-F., Estimating the greenhouse gas fluxes of European grasslands with a process-based model: 2. Simulations at the continental level. *Global Biogeochemical Cycles*, 21, 1005, 13p., 2007
- Wutzler, T., Reichstein, M., Soils apart from equilibrium – consequences for soil carbon balance modelling. *Biogeosciences* 4, 125–136, 2007.