1 General comments

1.1 Topic and modelling approach

This is an interesting paper about the difficult issue of explaining spatio-temporal variation in N2O-emission from soils. The authors mostly focus on temporal variation, in particular regarding emission-events following fertilization of grassland (a grass-clover mixture). Some attention is given to variation in the vertical direction, between different soil layers. Horizontal heterogeneity is not investigated.

The authors use the model ecosys as their analysis tool. The model is highly complex. It has a detailed representation of plant physiology, soil composition, microbial populations dynamics, diffusive and mass-flow transport processes, and chemical transformations - and all that with vertical spatial variation represented in the form of multiple canopy and soil layers.

Modelling N2O-flux variability is a notoriously difficult problem, and there are arguably no great success stories yet, so it is good to see another attempt. Some of the modelling results shown here look promising indeed: the model is able to account for some of the differences in N2O emission factors between years and fertilization-events, as observed at the Oensingen site in Switzerland over a period of 8 years. But the use of the highly complex ecosys model is the stand-out feature of this study, and much of the following discussion is about whether or not the complexity was shown to be essential.

*I have elaborated on the principles underlying the development of ecosys on l. 122 – 129 of the Introduction:*

(1) *all algorithms in the model must represent physical, biochemical and biological processes studied in basic research programs (e.g. convective-diffusive transport, oxidation-reduction reactions) so that these algorithms can be parameterized independently of the model*
(2) *this parameterization must be conducted at spatial and temporal scales smaller than those of prediction (in this case seasonal $N_2O$ fluxes) so that site-specific effects on predicted values are not incorporated into the algorithms.*

*These principles are designed to avoid as much as possible the use of site- and model-specific algorithms that may lack application in sites and models other than those for which they were developed.*

*In fact, although such models may appear complex, because they are parameterized from independent experiments, they can be better constrained than simpler models derived from site-specific data using model-specific formulations. The advantage of using these principles is that complex ecosystem behavior (e.g. N2O emissions) arising from interactions among different site conditions (e.g. soil water and temperature) can be simulated without additional parameterization, as noted several times in the Discussion (e.g. ll. 682 – 685, ll. 692 – 695, ll. 700 – 708, ll. 721 – 723, ll. 792 – 794). I therefore prefer the term ‘comprehensive’ rather than ‘complex’ to describe models such as ecosys.*

1.2. Is the model complexity needed?

Is all the complexity built into the ecosys model needed for simulation of N2Oemissions?

The authors claim so. They claim that process-based models for N2Oemission "must" follow several prescripts (lines 77, 79, 95), one of which reads: "These models must also explicitly represent the effects of mineral N, Ts and theta on the demand for vs. supply of O2 and alternative e- acceptors NO3-, NO2- and N2O, and the oxidation-reduction reactions by which these e- acceptors are reduced. However earlier process models have : : :".
This is a bold claim about the required modelling detail. But such statements on what models "must" look like are subjective and never fully defensible. Even complex models are (over)simplifications and other model structural choices could have been made. Despite the complexity of ecosys, it is still simplified compared to reality (e.g. plant hormones are ignored, as are within-plant variation in N-C ratio and chlorophyll fraction, the soil is assumed to be horizontally homogeneous etc. etc.).
So the authors have
made their own choice as to what fraction of known system complexity to represent in their model, and others might justifiably make different choices. For example, the modelling results reported here - and real emissions likely too - depend heavily on how fast the sward can regrow after harvesting. Fast plant regrowth requires much nitrogen, thereby leaving little substrate for N2O-emission. But we know that postharvest regrowth in grass swards suffers delays because of the weak photosynthetic machinery in uncut leaves at the bottom of the sward (a grass field after cutting is more yellow than green). However, this is not represented in ecosys - it assumes constant nitrogen and chlorophyll concentrations. Should we now say that ecosys "must" be changed to incorporate that? Many more such examples could be given.
The fact that the model ignores horizontal spatial variation is a huge simplification given what we know about small-scale spatial variability for N2O-emission.
The spatial simplification
seems at odds with the enormous detail in process representation as those processes operate in 3D not 1D. Indeed, large variation was observed between chambers regarding their annual total fluxes (Table 3), but this information was not used in the model-data comparison (chamber-data were averaged). The highest measured annual fluxes per chamber were nearly four times larger than the smallest values in several years, with presumably orders of magnitude larger relative differences at shorter time scales. Averaging environmental conditions is justified in a linear model, not in a nonlinear model like ecosys. So, given that the horizontal spatial heterogeneity was averaged away, can there be any justification for using a complex nonlinear model?

*Provision is made in ecosys for partial remobilization of structural C, N and P during senescenence of older, lower leaves, the nonstructural products of which are available for synthesis of structural C, N and P during growth of newer, upper leaves, thereby reinvesting nutrients from the lower to the upper parts of the canopy. However this part of the modelling is getting away from the stated purpose of the paper, to model ecological controls on N2O emissions, although it does indicate the need for comprehensive modelling of these controls.*

*I had not addressed the issue of spatial variability in the model because I had no data by which to characterized spatial variability in the field. However I have now addressed this issue by adding material to the manuscript that shows the response of modelled N2O emissions to hypothesized small increases in surface soil bulk density as might arise from natural variation or from compaction by field traffic (ll. 380 – 384, ll. 607 – 618, ll. 774 – 782, Fig. 8, Table 4). This large response would account for much of the spatial variation measured in the field experiment. The ability of ecosys to model this response without further parameterization arose from the* *detail and comprehensiveness with which processes are represented in the model, as pointed out in the Discussion.*

1.3 Performance of ecosys

We can also ask: Does the model teach us anything about N2O-fluxes from grasslands that simpler models can't? That question is not directly answered in this paper (because no other models - or model versions - are considered) but the authors do compare their model results to observations.

*The model provides a detailed explanation of the sensitivity of N2O fluxes to conditions in surface litter and near-surface soil. Such sensitivity has been inferred from other more basic research studies as cited in the paper, but nonetheless these conditions are rarely, if ever, characterized in field experiments. This sensitivity would appear to be a lesson offered from this modelling study that might guide future field methodology.*

One strength of the analysis here was that little site-specific calibration was carried out. Parameters were - apparently - kept at generic values, except for site-specific soil properties. It would be good though if more detail were provided about the parameterisation procedure: which parameters exactly were informed by knowledge of the local system and which were not? An advantage of working with little or no site-specific calibration, was that the data could be considered as independent from the modelling and thus used to assess the quality of the modelling.

*The key site inputs for weather, plant and soil properties and land management were described in the Model Experiment and either taken from earlier studies or listed in Tables 1 and 2. One key land management practice that had to be modelled for this study was harvest removal fraction because it affected subsequent plant regrowth and hence N2O emissions. The fractions selected for the harvests in this study were constrained by measurements of post-harvest LAI and CO2 exchange, as shown in the Results.*

Some of the ecosys results are very good, but not all of them. One good result, not even pointed out by the authors, was the fact that ecosys ranked the 8 years of annual fluxes in approximately the correct order (Table 3). The four years with lowest measured annual fluxes (2003-2006) were also the ones with the lowest simulated fluxes, and likewise for the two years with the highest fluxes (2008-2009) even though those two years had low fertilization (Table 3). Given the universal problems with modelling N2Ofluxes, this is a good result. So years are in the right order, but interannual variation was underestimated. The simulated magnitude of interannual flux-variability (standard deviation across the 8 years) was about three times too low compared to the data, but even that is a typical modelling result: models often underestimate ecosystem variability.
Very good results are also mentioned on p. 13, with the model correctly predicting differing emission factors for selected fertilization events (with higher emissions in some events despite lower fertilization levels). That is good, but why was the analysis not carried our for all emission events in the data? Could you expand the analysis to include all events, perhaps using the concept of the event-specific emission factor (see Flechard et al. 2007)? Why not for every observed emission event show the measured and modelled start/end dates, peak flux, cumulative flux? That would give a far more comprehensive picture of the quality of the simulations.

*I have added total measured vs modelled N2O-N emitted during each emission event to Fig. 3 (formerly Fig. 2). Start/end dates and peak fluxes are already represented in Fig. 3. I have done some further work to improve the readability of this Fig.*

The timing of emission-events seems to be reasonably well predictable using a simple empirical model (Smith & Massheder 2014 Nutr Cycl Agroecosyst 98:309-326): does the ecosys model improve on that?

*Predictions by such models are limited to the site conditions under which the model was derived. We wish to develop a more robust model capable of widespread application.*

Not all results are equally impressive. Spring emission events (up to DOY 180) are generally missed by the model (Fig. 2). Why is that?
And peaks of emission events after DOY 180 are often overestimated (Fig. 2), again why? Exceptionally accurate simulations are shown for two emission events that were singled out for closer study in Figs 3 and 4 - it would be better to have a more representative (or comprehensive) choice of events.

*All emission events modelled and measured from 2004 through 2009 were represented in Fig. 3. The events in Figs 4 and 5 were the largest ones that seemed to be the most important to simulate accurately as stated on ll. 437 - 438, and so were studied in more detail. The possible reasons for underestimating the spring emissions in 2008 were discussed on p. 20. Otherwise spring emissions were underestimated in 2004 and 2005, but overestimated in 2006 and 2009 (Fig. 3), which does not suggest a model bias for these emissions. In all cases, emissions modelled and measured in spring were significantly smaller than those in summer.*

It would also be good if at least one more site were included besides Oensingen, to test the modelling capacity for different soils and climates.

*The test data in this study were recorded under diverse conditions (winter, spring, summer and autumn) over 6 years and so present a good test of the model. Other sites have been used in earlier studies, as cited in the paper, and will be used in the future.*

1.4 How much of model performance is due to its complexity?

*As noted in the Introduction, the key advantage of more complex models is that they enable better constrained tests of model output against more comprehensive and diverse site data (e.g. Figs 4 and 5) than are possible with simpler models.*

In any case, it is not at all clear to what extent the model results shown are due to the complexity of the model being used. There was vastly more detail in the modelling than in the data, so no testing of underlying simulated processes could be carried out. And perhaps more importantly, no comparison of the performance of ecosys with other models, or simplified versions of ecosys itself, were carried out. All that leaves us unclear about which of the many processes and mechanisms represented in ecosys are essential for predicting N2O-emission rates.

*Given the complexity of ecological controls on N2O emissions, I don't believe it is productive to try to determine which processes are essential and which are not under any specific set of site conditions, as all processes in the model (e.g.those that contribute to GPP, Ra, Rh, energy* *exchange, gas transfer and many more) may to some degree affect these emissions. Attempting to subtract any of these processes can only reduce model robustness.*

1.5 A complex model as hypothesis generator

Of course, there is another line of reasoning to justify model complexity: complex models may help explore possible mechanistic explanations for observations.
In lines 399-451, the authors give us story lines, derived from the modelling, as to what happened in the soil leading up to various emission events. These are highly interesting and show the value of the model that was being used. However, in the final analysis the key mechanisms seem to be fairly simple. The authors provide a nice summary in lines 449-451: "model findings indicated the importance to N2O emissions of surface and near-surface theta after precipitation, and of plant management (intensity and timing of defoliation in relation to N application) and its effect on subsequent CO2 fixation". Another nice short summary is given in lines 686-697. Given these simple explanations, could the same results not have been achieved with a much simpler model? One that models these summary mechanisms without superfluous detail? For example, the authors do not mention microbial activity (or population dynamics) anywhere in lines 399-451, so do they need to be modelled at all if the aim is forecasting N2O-emissions?

*Given the complexity of ecological controls on N2O emissions, how do we know what is 'superfluous'? Therefore emissions need to be simulated with comprehensive ecosystem models in which all the key processes known to affect emissions are represented, rather than with models that represent a limited subset of processes in which the effects of unrepresented processes 'hardwired' into the code. Such models may simulate emissions under site-specific conditions, but may not do so under conditions that differ markedly from those for which they were hardwired. That is why these simpler models are continuously being reformulated.*

Moreover, some of the results, e.g. those about the importance of the top soil layer, are already known from the experimental literature as cited by the authors (Neftel et al. 2000; van der Weerden et al. 2013; Pal et al. 2013). Also the model results concerning sensitivity to intensity of foliage removal at harvests (Ruzjerez et al. 1994; Imer et al. 2013). So what does the authors' modelling study add to that? Arguably, the model analysis did help formulate and evaluate hypotheses on the mechanisms underlying the phenomena. And the paper usefully builds on that with their interesting sensitivity analysis showing the possible importance of harvest intensity and harvest timing for emission rates.

*These effects may be known qualitatively from experiments, but models that represent these effects from basic processes may provide a robust method to predict these effects quantitatively under diverse conditions of soil, weather and land management needed for larger-scale studies.*

1.6 Overall

Overall, this is interesting work, somewhat marred by the enormous complexity of the model without the reader being able to judge the necessity of that complexity. The absence of any uncertainty analysis, and the focus on just a few of the many emission events observed, and the use of data from just one site, also make it hard to evaluate the work.
I look forward to future modelling work from the authors in which they show how much

of their model's complexity is essential, and how much can be stripped away without affecting model performance.

*We have added a study of surface bulk density effects on N2O emissions, as noted earlier, to demonstrate how the model's comprehensiveness allows it to respond realistically to specified changes in a measureable input.*

2 Specific comments

Some claims about the accuracy of the simulations are difficult to judge, as Figures 2 to 4 are much too small, and the latter two are much too crowded. The first line of the discussion states that "Most N2O emission events measured from 2004 to 2009 were simulated within the range of measurement uncertainty, estimated to be about 30% of mean values (Fig. 2)". Apart from the fact that no basis for the uncertainty level is given, the degree of model-data correspondence cannot be judged from the figures. Perhaps use fewer but bigger graphs and move some to the Supplementary Material where they also need not be so small?

*I would like to keep Figs 3 and 4 (now Figs 4 and 5) as is, because they demonstrate the diversity of tests of which a more detailed model is capable, and hence the better constraints on model performance that such tests enable.*

Lines 250-251: "The soil [at the Oensingen field site : : :], key properties of which are given in Table 1". No, those are clearly not measured soil properties but assumed model inputs. It is impossible to measure exactly the same values, to three significant digits, of bulk density, field capacity etc. at five different depths. So how were the soil properties quantified in reality? Were they set to values that gave proper model performance?

*As clearly indicated in the footnotes for this table, the soil hydraulic properties were derived from a pedotransfer function cited in the footnotes. Further details about the sources of the soil properties used in this study have been added to the footnotes along with some updated references for methodologies.*

The eddy covariance system: can you state the magnitude and location of its footprint relative to the positions of the chambers: do the N2O- and CO2-fluxes refer to the same part of the field?

*Further details about the spatial relations of the EC tower and chambers have been added in ll. 335 - 340.*

Line 376: What is that uncertainty assessment of 30% based on?

*More information about uncertainty in the measured fluxes has been added in ll. 427 - 433.*

The last footnote to Table 1 is incorrect. Soil composition (sand, silt, clay) cannot have been recalculated as a function of SOC and CF, because those latter variables varied with depth in the range 0.28-1.50 m, whereas your values for soil composition were constant over that same depth range.

*These variables were recalculated from their input values which sum to 1000 to account for the presence of the non-mineral soil fraction in each layer. I have removed this footnote because the point is not important enough to dwell on.*

3 Technical corrections

The list of References is not fully alphabetical: see Conant et al. (2005), and also the many Grant et al. references.

*done*

The paper is very well-written and I noticed only few minor language errors (lines 40, 108, 295 and the second caption lines of Figs 3 and 4).

*done*

There are unit errors on line 518.

*done*

Figures 3-4 require a microscope.

*I would like to keep these figures (now 4 and 5) as is, because they indicated the well-constrained testing of which a comprehensive model such as ecosys is capable. I have done some further work to simplify Figs 7 and 8 (formerly 6 and 7). I have also enlarged fonts and simplified the axes in Fig. 3 to make it more readable.*

The Supplementary Material is very long but unnecessarily so: much of the text and many of the equations are irrelevant for the grassland modelling carried out here. There is information on modelling woody plants ["branches", "twigs", "coarse woody litter"], C4 plants, methane emissions, horizontal heterogeneity i.e. x- and y-variables in equations etc. etc. - none of which is used in the current study. Also, the within-text referencing is often wrong. For example, there is no section called "Energy Exchange" either "above" or anywhere else. Likewise for "Autotrophic Respiration and Growth" etc. etc. Either make - and carefully check - a document that is specific to this text, or refer to a proper general ecosys-document not linked to this paper. It would also be good if the model description began with some basic overall facts: How many state variables does the model have, how many input variables, how many parameters (and how many are deemed universal constants, functional type constants, site-specific parameters). Some diagrams of model structure would also be helpful.

*The Supplementary Material includes a general coverage of ecosys. The reader is referred in the Model Development to specific appendices in the Supplementary Material that are directly relevant to modelling N2O emissions, and may thus ignore those appendices that are not, such as those with the sections.  Model structure, input variables and key parameters are summarized in Grant(2001) as cited in the paper. I am committed to publishing all key algorithms for use by other modelling projects, rather than leaving the model as a ‘black box’.*

*However to facilitate reader understanding of model structure, I have added a flow diagram showing key processes governing N2O emissions as Fig. 1 to the paper.*

Header of Table 2: replace "2004" with "2001".

*Done*

Table 3: the NEP-values are redundant (NEP = GPP + Re).

*Field values of NEP are directly derived from EC measurements and so provide the better test of modelled values. A direct comparison of modelled and measured NEP  is therefore informative.*

Fig. 4e: "1-3 m" should be "1-3 cm".

*done*