Considering unknown uncertainty in imperfect models: nitrogen mineralization as a case study

Maria P. Vilas[a] [ID], **Frederick R. Bennett[b]** [ID], **Kirsten Verburg[c]** [ID] **and Matthew P. Adams[d,e]** [ID]

[a] Department of Resources, Queensland Government, Brisbane QLD Australia
[b]Department of Environmental Science, Queensland Government, Brisbane QLD Australia
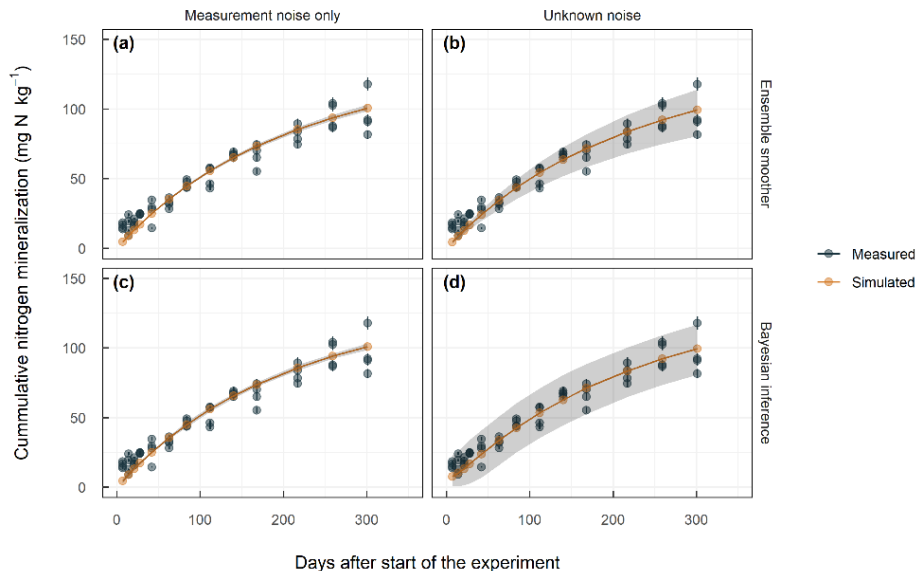*[c]CSIRO Agriculture and Food, Canberra ACT Australia*
[d]School of Mathematical Sciences, ARC Centre of Excellence for Mathematical and Statistical Frontiers, and the Centre for Data Science, Queensland University of Technology, Brisbane QLD Australia
[e]School of Chemical Engineering, The University of Queensland, St Lucia QLD Australia
Email: maria.delapazvilas@resources.qld.gov.au

Abstract: All models are imperfect, so it is important to consider uncertainty in their predictions. When calibrating models to measured data, uncertainty in the model can be simultaneously estimated, although there are multiple methods available for doing this. In this study, we applied ensemble smoothers and Bayesian inference to calibrate a model of nitrogen mineralization in soils. We obtained mineralization measurements from a previously published study that measured changes in inorganic nitrogen over long-term laboratory incubations in a soil located in the Mackay Whitsundays region (North Queensland). Simulations were performed using the Agricultural Production Systems Simulator (APSIM). We inferred two parameters that characterize the size of the simulated soil organic carbon pools (fbiom and finert) because it is difficult to estimate these parameters from measurements only. For the calibration, we considered two different sources of uncertainty: measurement noise and noise of unknown origin, the latter of which includes all non-measurement related errors. We found that ignoring noise of unknown origin can result in an overly optimistic representation of the model error (Figure 1a,c). On the contrary, incorporating noise of unknown origin can lead to a more accurate representation of the uncertainty in the predictions, with model predictions providing adequate coverage of measurements (Figure 1b,d). We show that parameterizing fbiom and finert is difficult because these parameters are correlated, hence different combinations of parameters can equally well simulate the measured data. We suggest that future work needs to provide a means of parameterizing at least one of these fractions independently to facilitate parameter identifiability.

Figure 1. Measured (grey dots) and simulated (orange lines) nitrogen mineralization (mg N kg^{-1}) for the ensemble smoother with multiple data assimilation (ES-MDA) (a), flexible iterative ES-MDA (b), Bayesian inference with measurement noise only (c), Bayesian inference with measurement noise and additional noise (d). The error bars in the measured data represent the error of the laboratory method. Shaded areas represent the predicted 95% credible intervals.

Keywords: *Soil nitrogen, mineralization, Sequential Monte Carlo, ensemble smoothers, APSIM*

1.   INTRODUCTION

Models are an imperfect representation of the true state of a system. When models are used to inform decision-making, it is important to incorporate uncertainty in the predictions (Ascough et al., 2008; Maier et al., 2016). Uncertainty arises from various sources: inherent randomness, measurement error, systematic error, systematic variation, model uncertainty, and subjective judgment (Uusitalo et al., 2015). Calibrating models to measured data provides a means of estimating some of this uncertainty.

Uncertainty due to measurement error can be estimated and dealt with in probabilistic frameworks. If the available data includes estimates of measurement error, model-data calibration can be used to investigate other uncertainty sources. However, quantifying multiple sources of uncertainty is difficult because they can be hard to distinguish from each other (Certain et al., 2018). Treating all non-measurement errors together as one term or distribution that accounts for all "errors of unknown origin" is a first pass approach to dealing with this problem. In this paper, we demonstrate that neglecting errors of unknown origin can lead to inaccurate estimates of model parameters and compromise the accuracy of the predictions.

To ensure that our conclusions do not depend on the model-data calibration methods used, we present results from two different calibration approaches: ensemble smoothers (Emerick & Reynolds, 2013) and Bayesian inference (Girolami, 2008). A recent study presented a flexible iterative ensemble smoother with multiple data assimilation (flexible ES-MDA) that accounts for errors of unknown origin such as model structural errors (Rammay et al., 2021). The authors demonstrated that only accounting for measurement error can give rise to biased and over-confident estimates of model parameters. Errors of unknown origin can also be considered within a Bayesian inference framework (e.g. Xu and Valocchi, 2015), although the statistical treatment of these errors in Bayesian inference is markedly different from ES-MDA. This paper's objective is to examine how errors of unknown origin affect model parameters and predictions. This is timely as there is increasing interest in parameter identifiability issues within mechanistic soil biogeochemical models (Marschmann et al., 2019). To achieve the objective of this study, we applied ensemble smoothers and Bayesian inference methods to fit a mechanistic model of nitrogen mineralization to measured data. We used the Agricultural Production Systems Simulator (APSIM; http://www.apsim. info/; Holzworth et al., 2014) to simulate a nitrogen mineralization experiment undertaken by Allen et al. (2019). Calibration of APSIM to the data from Allen et al. (2019) was used to estimate two parameters in the APSIM model: "finert", which represents the proportion of initial organic carbon assumed to be inert in the soil, and "fbiom", which represents the proportion of non-inert carbon in the microbial biomass within the soil. These two parameters are used in APSIM to calculate the initial size of two conceptual carbon pools: biom (microbial biomass) and hum (the rest of soil organic matter), where mineralization of organic carbon occurs (Probert et al., 1998, 2005). Mineralization dynamics are highly sensitive to the fractions defined by the "finert" and "fbiom" parameters, so these parameters are crucial to estimate it correctly, but there are currently no measurements that accurately represent these fractions (Luo et al., 2014). Hence, the accurate estimation of these two critical parameters in APSIM not only provides an ideal test case for our analysis, but also provides useful information for improving future simulations of nitrogen mineralization in soils.

2.   METHODS

2.1.   Measured data

We obtained mineralization data from Allen et al. (2019) who measured changes in inorganic nitrogen over long-term (301 days) laboratory incubations across 15 soils obtained from sugarcane fields in North Queensland. Incubations were performed at constant temperature (35ºC) and soil water (field capacity). Soils were sieved to remove any fresh organic matter. Four replicates were collected for determination of soil inorganic nitrogen content on days 0, 7, 14, 21, 28, 42, 63, 84, 112, 140, 168, 217, 259 and 301. The interested reader is referred to Allen et al. (2019) for details on the experimental design.

2.2.   Model set up

We used the APSIM model (version 7.10) configured with soil water and nitrogen modules (Probert et al., 1998). APSIM parameters were either default or measured values. We simulated the laboratory incubation for one soil from the Mackay Whitsundays (-21º 11' 3''S, 148º 58' 7''E) using its verified soil parameterization from Vilas et al., (2021) and measured soil characteristics from Allen et al. (2019) (Table 1). Default values were used for soil carbon dynamics, including turnover rates of 0.0081 and 0.00015 day^{-1} for the biom and hum pools, respectively (Probert et al., 1998) and nitrification rates (Meier et al., 2006; Probert et al., 1998).

Table 1. Soil parameters used in APSIM for the Mackay Whitsundays soil.

Parameter	Value	Units	Source
Bulk density (BD)	1.200	g cm^{-3}	Vilas et al. (2021)
Soil lower limit (LL15)	0.250	mm mm^{-1}	Vilas et al. (2021)
Soil drained upper limit (DUL)	0.460	mm mm^{-1}	Vilas et al. (2021)
Soil saturation (SAT)	0.520	mm mm^{-1}	Vilas et al. (2021)
Soil organic carbon (OC)	2.295	%	Allen et al. (2019)
Soil pH	5.200	Units of pH	Allen et al. (2019)
Soil C:N ratio	16	-	Allen et al. (2019)

Only one soil layer (0-10 cm) was simulated to represent the incubation experiment. Simulations were run for 301 days at a constant soil water (field capacity) and temperature (35ºC) to match the experimental conditions.

2.3. Calibration

The posterior distribution of the parameters fbiom and finert was produced using: (a) an ensemble smoother with multiple data assimilation (ES-MDA, Emerick & Reynolds, 2013), (b) a flexible iterative ensemble smoother with multiple data assimilation (flexible ES-MDA, Rammay et al., 2021), (c) Bayesian inference (Girolami, 2008) using a likelihood function that assumes the only discrepancy between the model and data is measurement noise, and (d) Bayesian inference using a likelihood function that assumes the discrepancy between the model and data is measurement noise plus error of unknown origin. Hence (a) and (c) represent different calibration procedures that ignore error of unknown origin, and (b) and (d) represent calibration procedures that explicitly account for error of unknown origin.

All four measurement replicates were used in the calibration. For each replicate a measurement error with a standard deviation of 4% was assumed as it represents the error of the laboratory method used to measure inorganic nitrogen in the soil (APHA et al., 2012). All approaches (a-d) require definition of prior distributions for fbiom and finert; these distributions represent our initial beliefs about the probability distributions for the model parameters. The prior distributions for fbiom and finert were truncated normal distributions with means of 0.093 and 0.580, respectively (Biggs et al., 2021; Vilas et al., 2021), standard deviations equal to one-quarter of the difference between these parameters' stated upper and lower bounds and truncated within these upper and lower bounds. The lower bounds of fbiom and finert were 0.05 and 0.4, respectively, to match the default APSIM parameterization (Dalgliesh et al., 2016). The upper bounds were then set to 0.15 for fbiom and 0.8 for finert, which is within the range of previous studies (Luo et al., 2019). For the ensemble smoothers, an ensemble of 500 members was used, with 10 iterations. Analogously, our computational implementation of Bayesian inference used an ensemble of 500 members.

Ensemble smoothers

The ES-MDA (Emerick & Reynolds, 2013) was originally developed to address some inefficiencies identified with the ensemble Kalman filter (EnKF) (Evensen, 1994). A significant advantage of the ES-MDA over the EnKF is that an ensemble of data realizations is assimilated simultaneously rather than sequentially. The ES-MDA is theoretically grounded in Bayesian Statistics and armed with a prior parameter distribution, $\mathbf{M}$, a vector of observed data, d_{obs}, and a deterministic forward solver, g, data is conditioned using the procedure outlined in Algorithm 1.

Algorithm 1 Standard ES-MDA

Choose the number of ES-MDA steps, N_a, and inflation factors such that $\sum_{i-1}^{N_a} \frac{1}{\alpha_i} = 1$

for $i = 1,\ldots, N_a$ **do**

$D_{uc} = d_{obs} + \sqrt{\alpha_i}\, \boldsymbol{C}_D^{1/2} z_d$, perturb the observation vector for each ensemble member where $z_d \sim \mathcal{N}\left(0, \mathrm{I}_{N_d}\right)$

$\mathbf{D} = g(\mathbf{M})$, build ensemble of model outputs

$\mathbf{M} \leftarrow \mathbf{M} + \mathbf{C}_{MD}(\mathbf{C}_{DD} + \alpha \mathbf{C}_D)^{-1}(\mathbf{D}_{uc} - \mathbf{D})$, update the ensemble of model parameters

end for

$\mathbf{C}_D$ is the covariance matrix of the measurement noise, $\mathbf{C}_{DD}$ is the auto-covariance matrix of the predicted data at the current iteration and $\mathbf{C}_{MD}$ is the cross-covariance between the predicted data and the model parameter ensemble given by the standard formulas:

$$C_{MD} = \frac{1}{N_a - 1} \sum_{i=1}^{N_a} (m_j - \bar{m})(d_j - \bar{d})^T$$

and (1)

$$C_{DD} = \frac{1}{N_a - 1} \sum_{i=1}^{N_a} (d_j - \bar{d})(d_j - \bar{d})^T$$

The ES-MDA assumes that the forward model accurately represents the physical phenomena giving rise to the measurement data, in such a way that the mismatch between the assimilated model output and the measured data is dominated by measurement error. Rammay et al. (2021) demonstrated that when the model is deficient, model structural error can give rise to biased and over-confident estimates of model parameters. They proposed an adjustment to the standard ES-MDA to account for model errors of unknown origin to improve the quality of estimated parameters. Their flexible ES-MDA uses model residuals derived directly from the ensembles and therefore has no increase computational overhead compared to the standard ES-MDA. The reader is referred to Rammay et al. (2021) for a discussion on how this adjustment is implemented.

Bayesian inference

Bayesian inference was implemented using a Sequential Monte Carlo sampling algorithm described in Adams et al. (2020a, 2020b), except with one additional resampling step at the end of algorithm to make all posterior samples possess the same weight for ease of subsequent computations. Although many other implementations of Bayesian inference are available, we used Sequential Monte Carlo sampling primarily because of its ability to exploit parallel computing architectures (Dai et al., 2020). For the fit of the APSIM model to the data, assuming that the only noise is measurement noise, the likelihood function for the model-data difference was Gaussian in all data points, with a standard deviation equal to 4% of each measured data point. For the fit of the APSIM model to the data, assuming that noise is measurement noise plus error of unknown origin, the likelihood function for the model-data difference was again Gaussian in all data points. In this latter fit, the standard deviation was instead set to $\sqrt{(\sigma_{d,j}^2 + \sigma_m^2)}$, where $\sigma_{d,j}$ is equal to 4% of the jth measured data point and σ_m is the standard deviation of the normal distribution that characterizes additional error of unknown origin. This square root expression is correct if the imperfect model's assumed noise structure is a summation of measurement noise and additional noise. Such a noise structure agrees with that assumed by flexible ES-MDA (Rammay et al. 2021) if both the measurement noise and additional noise follow Gaussian distributions. Under this assumed noise structure, σ_m was treated as a third parameter to be estimated. A uniform prior for σ_m, bounded between 0 and 20 mg N kg^{-1}, was assumed. In both calibrations undertaken via Bayesian inference (measurement noise only, and measurement noise plus error of unknown origin), the likelihood functions were truncated so that measurements of negative values of cumulative nitrogen mineralization were assigned zero probability.
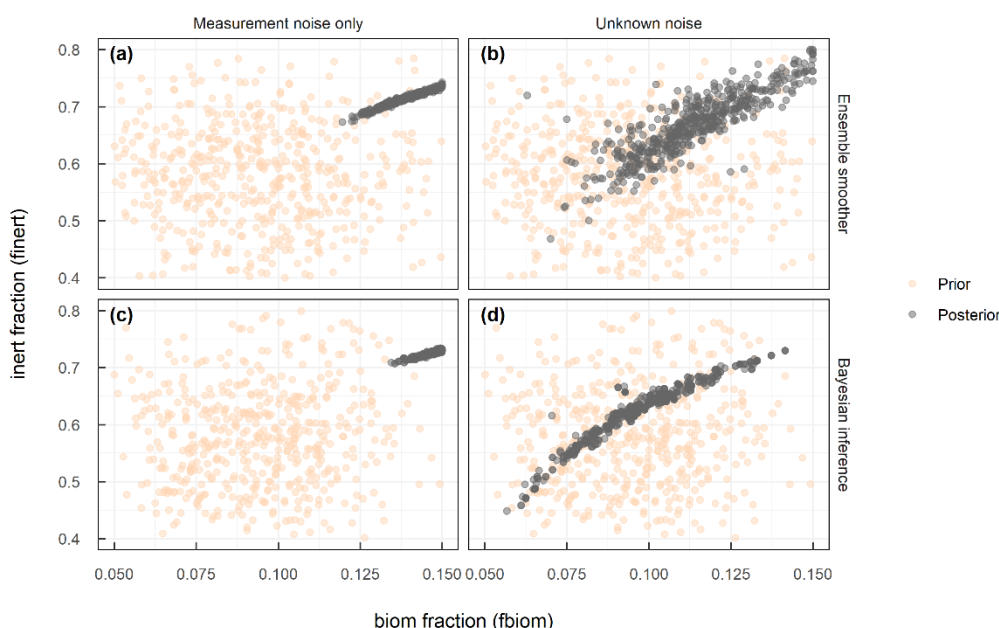
3.   RESULTS

Predictions of models fitted assuming that the only noise is measurement noise provided inadequate coverage of the measurement data, regardless of the calibration technique (Figures 1a,c: orange lines and grey shaded areas). When additional noise was also assumed to be present Figures 1b,d), the coverage was far more reasonable. Thus, ignoring noise sources other than measurement noise can compromise the prediction accuracy of models fitted to data if probabilistic forecasts are sought. It is worth noting that Bayesian inference with the additional error term (Figure 1d) provided better coverage for the measured data points than the flexible ES-MDA (Figure 1b). For fitting the APSIM model to data using Bayesian inference with both noise types included, the posterior estimate for the standard deviation parameter σ_m characterizing the noise of unknown origin was 8.7±1.5 mg N kg^{-1}. Given that the measurement noise is assumed to possess a standard deviation of 4% of measured values, and the largest measured values were of the order of 100 mg N kg^{-1} (Figure 1), we find that the noise of unknown origin appears to be at least twice as large as the error in the laboratory measurement.

Parameter values including their uncertainty (mean ± sd) for the different approaches are shown in Table 2. In general, the standard deviation of the calibrated model parameters was much higher for the calibration techniques with additional noise (flexible ES-MDA and Bayesian inference) than with measurement noise only.

Table 2. Parameters fbiom and finert (mean ± sd) estimated using the four different calibration methods (BI = Bayesian inference). Notice, however, that in Figure 2 the distribution of parameter values is more complex than a bell-shaped curve.

Parameter	Standard ES-MDA	Flexible ES-MDA	BI with measurement noise	BI with both noise types
fbiom	0.139 ± 0.006	0.114 ± 0.017	0.147 ± 0.003	0.097 ± 0.017
finert	0.714 ± 0.012	0.668 ± 0.056	0.725 ± 0.005	0.622 ± 0.054

When additional noise was considered in the calibration procedures, the range of plausible values for fbiom and finert based on the data, was substantially larger (Figure 2b,d) than when additional noise was excluded from the calibration procedures (Figure 2a,c). This suggests that the exclusion of additional noise from calibration procedures leads to overfitting of model parameter values. This conclusion held regardless of whether ES-MDA or Bayesian inference was used, although there were also subtler differences between these methods' estimates of plausible parameter values based on the data. What is also clear is that there is positive correlation between the two parameters' values, regardless of which of the four calibration procedures was used (Figure 2).

Figure 2. Samples of plausible paired values of fbiom and finert parameters, obtained from the 500 ensemble members approximating the prior (light orange) and posterior (grey) distributions obtained from standard ES-MDA (a), flexible ES-MDA (b), Bayesian inference with measurement noise only (c), and Bayesian inference with both noise types (d).

4. DISCUSSION

Models are often calibrated to observations to estimate parameters that are difficult to measure directly. In our case study, we inferred two parameters of APSIMs' soil module (fbiom and finert) from measurements of nitrogen mineralization using Bayesian and ensemble smoother methods that can both account for different sources of noise. Overall, both methods showed that ignoring unknown noise leads to a collapse in the diversity of the posterior parameter distributions, giving rise to overly optimistic estimates of the parameters as illustrated in the Results section. Our findings suggest that accounting for unknown sources of uncertainty will lead to more defensible predictions, enabling a broader applicability of model outputs.

Both the ensemble smoothers and the Bayesian inference approach resulted in similar parameter estimations. Bayesian inference has been applied to calibrate parameters that describe the carbon dynamics in APSIM (Luo et al., 2015, 2019). However, Bayesian inference has high computational costs compared with ensemble smoothers (Rammay et al., 2021). Thus, ensemble smoothers may be a good alternative for fast calibration of agricultural models.

All four calibration techniques showed a correlation between the posterior fbiom and finert parameter values. APSIM's mineralization stems from the turnover of the biom and hum conceptual carbon pools and is determined by the size of the pools and their decay rates. A higher value of finert will reduce the amount of organic carbon (hum + biom) available to be mineralized. This effect can be off-set by a higher fbiom, as the biom pool has a higher turn-over rate than the hum pool (0.0081 vs 0.00015 day^{-1}). As fbiom is expressed relative to the non-inert organic carbon fraction, the increase in fbiom needs to be larger for an increase in finert at the higher end of the scale, explaining the curvilinear nature of the relationship. The correlation between APSIM fbiom and finert presents a challenge for the parameterization of these fractions. Ideally, additional measurements should independently inform one of these fractions. Unfortunately, there are currently no measurements that accurately represent each of these fractions (Luo et al., 2014). The issue of parameter identifiability has been widely discussed in the literature, with studies suggesting to change, reduce complexity or re-parameterize models (e.g. Browning et al., 2020). Thus, future work needs to evaluate strategies to deal with parameter identifiability issues in agricultural models.

ACKNOWLEDGEMENTS

REFERENCES

Adams, M. P., Koh, E. J. Y., Vilas, M. P., Collier, C. J., Lambert, V. M., Sisson, S. A., Quiroz, M., McDonald-Madden, E., McKenzie, L. J., & O'Brien, K. R. (2020). Predicting seagrass decline due to cumulative stressors. *Environmental Modelling & Software*, *130*, 104717. doi: 10.1016/j.envsoft.2020.104717

Adams, M. P., Sisson, S. A., Helmstedt, K. J., Baker, C. M., Holden, M. H., Plein, M., Holloway, J., Mengersen, K. L., & McDonald-Madden, E. (2020). Informing management decisions for ecological networks, using dynamic models calibrated to noisy time-series data. *Ecology Letters*, *23*(4), 607–619. doi: 10.1111/ELE.13465

Allen, D. E., Bloesch, P. M., Orton, T. G., Schroeder, B. L., Skocaj, D. M., Wang, W., Masters, B., & Moody, P. M. (2019). Nitrogen mineralisation in sugarcane soils in Queensland, Australia: I. evaluation of soil tests for predicting nitrogen mineralisation. *Soil Research*, *57*(7), 738. doi: 10.1071/SR19031

APHA, & AWWA. (2012). *Standard methods for the examination of water and wastewater* (E. W. Rice, R. B. Baird, A. D. Eaton, & L. S. Clesceri (eds.); 22nd editi, Vol. 49, Issue 12). doi: 10.5860/choice.49-6910

Ascough, J. C., Maier, H. R., Ravalico, J. K., & Strudley, M. W. (2008). Future research challenges for incorporation of uncertainty in environmental and ecological decision-making. *Ecological Modelling*, *219*(3–4), 383–399. doi: 10.1016/J.ECOLMODEL.2008.07.015

Biggs, J. S., Everingham, Y., Skocaj, D. M., Schroeder, B. L., Sexton, J., & Thorburn, P. J. (2021). The potential for refining nitrogen fertiliser management through accounting for climate impacts: An exploratory study for the Tully region. *Marine Pollution Bulletin*, *170*, 112664. doi: 10.1016/j.marpolbul.2021.112664

Browning, A. P., Warne, D. J., Burrage, K., Baker, R. E., & Simpson, M. J. (2020). Identifiability analysis for stochastic differential equation models in systems biology. *Journal of the Royal Society, Interface*, *17*(173), 20200652. doi: 10.1098/rsif.2020.0652

Certain, G., Barraquand, F., & Gårdmark, A. (2018). How do MAR(1) models cope with hidden nonlinearities in ecological dynamics? *Methods Ecol Evol*, *9*, 1975–1995. doi: 10.1111/2041-210X.13021

Dai, C., Heng, J., Jacob, P. E., & Whiteley, N. (2020). *An invitation to sequential Monte Carlo samplers*. Retrieved from http://arxiv.org/abs/2007.11936

Dalgliesh, N., Hochman, Z., Huth, N., & Holzworth, D. (2016). CSIRO agriculture and food: A protocol for the developmentof APSOIL parametervalues for use in APSIM. *CSIRO Agriculture and Food*, *20*(September), 1–24. Retrieved from https://www.apsim.info/wp-content/uploads/2019/10/Parameters-for-soil-water-Ver24.pdf

Emerick, A. A., & Reynolds, A. C. (2013). Ensemble smoother with multiple data assimilation. *Computers & Geosciences*, *55*, 3–15. doi: 10.1016/J.CAGEO.2012.03.011

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, *99*(C5), 10143–10162. doi: 10.1029/94JC00572

Girolami, M. (2008). Bayesian inference for differential equations. *Theoretical Computer Science*, *408*, 4–16.

doi: 10.1016/j.tcs.2008.07.005

Holzworth, D. P., Huth, N. I., deVoil, P. G., Zurcher, E. J., Herrmann, N. I., McLean, G., Chenu, K., van Oosterom, E. J., Snow, V., Murphy, C., Moore, A. D., Brown, H., Whish, J. P. M., Verrall, S., Fainges, J., Bell, L. W., Peake, A. S., Poulton, P. L., Hochman, Z., … Keating, B. A. (2014). APSIM – Evolution towards a new generation of agricultural systems simulation. *Environmental Modelling & Software*, *62*, 327–350. doi: 10.1016/J.ENVSOFT.2014.07.009

Luo, Z., Eady, S., Sharma, B., Grant, T., Liu, D. L., Cowie, A., Farquharson, R., Simmons, A., Crawford, D., Searle, R., & Moore, A. (2019). Mapping future soil carbon change and its uncertainty in croplands using simple surrogates of a complex farming system model. *Geoderma*, *337*, 311–321. doi: 10.1016/j.geoderma.2018.09.041

Luo, Z., Wang, E., Fillery, I. R. P., Macdonald, L. M., Huth, N., & Baldock, J. (2014). Modelling soil carbon and nitrogen dynamics using measurable and conceptual soil organic matter pools in APSIM. *Agriculture, Ecosystems and Environment*, *186*, 94–104. doi: 10.1016/j.agee.2014.01.019

Luo, Z., Wang, E., Zheng, H., Baldock, J. A., Sun, O. J., & Shao, Q. (2015). Convergent modelling of past soil organic carbon stocks but divergent projections. *Biogeosciences*, *12*(14), 4373–4383. doi: 10.5194/bg-12-4373-2015

Maier, H. R., Guillaume, J. H. A., van Delden, H., Riddell, G. A., Haasnoot, M., & Kwakkel, J. H. (2016). An uncertain future, deep uncertainty, scenarios, robustness and adaptation: How do they fit together? *Environmental Modelling and Software*, *81*, 154–164. doi: 10.1016/j.envsoft.2016.03.014

Marschmann, G. L., Pagel, H., Kügler, P., & Streck, T. (2019). Equifinality, sloppiness, and emergent structures of mechanistic soil biogeochemical models. *Environmental Modelling & Software*, *122*, 104518. doi: 10.1016/J.ENVSOFT.2019.104518

Meier, E. A., Thorburn, P. J., & Probert, M. E. (2006). Occurrence and simulation of nitrification in two contrasting sugarcane soils from the Australian wet tropics. *Australian Journal of Soil Research*, *44*(1), 1–9. doi: 10.1071/SR05004

Probert, M. E., Delve, R. J., Kimani, S. K., & Dimes, J. P. (2005). Modelling nitrogen mineralization from manures: Representing quality aspects by varying C:N ratio of sub-pools. *Soil Biology and Biochemistry*, *37*(2), 279–287. doi: 10.1016/j.soilbio.2004.07.040

Probert, M. E., Dimes, J. P., Keating, B. A., Dalal, R. C., & Strong, W. M. (1998). APSIM's water and nitrogen modules and simulation of the dynamics of water and nitrogen in fallow systems. *Agricultural Systems*, *56*(1), 1–28. doi: 10.1016/S0308-521X(97)00028-0

Rammay, M. H., Elsheikh, A. H., & Chen, Y. (2021). Flexible iterative ensemble smoother for calibration of perfect and imperfect models. *Computational Geosciences*, *25*(1), 373–394. doi: 10.1007/s10596-020-10008-z

Uusitalo, L., Lehikoinen, A., Helle, I., & Myrberg, K. (2015). An overview of methods to evaluate uncertainty of deterministic models in decision support. In Environmental Modelling and Software (Vol. 63, pp. 24–31). Elsevier Ltd. doi: 10.1016/j.envsoft.2014.09.017

Vilas, M. P., Shaw, M., Rohde, K., Power, B., Donaldson, S., Foley, J., & Silburn, M. (2021). Ten years of monitoring dissolved inorganic nitrogen in runoff from sugarcane informs development of a modelling algorithm to prioritise organic and inorganic nutrient management. *Science of The Total Environment*, 150019. doi: 10.1016/J.SCITOTENV.2021.150019

Xu, T., & Valocchi, A. J. (2015). A Bayesian approach to improved calibration and prediction of groundwater models with structural error. *Water Resources Research*, *51*(11), 9290–9311. doi: 10.1002/2015WR017912