Tavernet, J.P., 2023. SENSITIVITY AND UNCERTAINTY ANALYSES FOR N-LOSS ESTIMATES BY THE OVERSEER MODEL. In: Diverse Solutions for Efficient Land, Water and Nutrient use. (Eds. C.L. Christensen, D.J. Horne and R. Singh). http://flrc.massey.ac.nz/publications.html. Occasional Report No. 35. Farmed Landscapes Research Centre, Massey University, Palmerston North, New Zealand. 8 pages

SENSITIVITY AND UNCERTAINTY ANALYSES FOR N-LOSS ESTIMATES BY THE OVERSEER MODEL

Jean-Paul Tavernet

Overseer Limited
Level 4, 2 Woodward Street, Wellington, New Zealand
Email: jeanpaul@overseer.org.nz

Abstract

Sensitivity and uncertainty analyses provide information that helps to understand a science model's performance and outputs. As the first step to better understanding N-loss estimates generated by the Overseer science model, sensitivity and uncertainty analyses were undertaken to identify key input parameters and interactions, that significantly alter model estimates for dairy, beef and sheep and cropping farm systems.

This work utilised real-world anonymised farm setups to carry out local and global sensitivity analyses on a minimum of 30 parameters for different farm systems in Overseer. It confirmed that Overseer's N-loss estimate was most sensitive to changes in key climate and soil parameters with minimal parameter interactions. Uncertainties for these inputs were propagated through the model, and the combined output uncertainties in the N-loss estimate relating to soil and climate, averaging 27±9% across all the farm systems, is consistent with other models in the field. The predicted uncertainty of other parameters identified in the sensitivity analyses, due to independence from parameter interactions, will also be discussed.

Introduction

Overseer is a farm management tool that uses a long-term, quasi-equilibrium agricultural science model (Wheeler et al., 2022) to estimate nutrient losses, including nitrogen (N) losses, from a farm system.

The Overseer model, like most agricultural simulation models, utilises a large amount of data and input parameters to calculate its outputs. That said, it is common for a model of this nature that a subset of key parameters strongly influence the variability in certain model outputs.

To provide a better understanding of the N-loss estimates from the Overseer model, it is important to understand the model's sensitivity and uncertainty in the model output relative to the uncertainty in the key model inputs. The scope of this work was to first undertake a sensitivity analyses for the three main farming systems referenced in Overseer to identify which of the analysed inputs have the most significant impact on the modelled farm level N-loss. Next, this work aimed to study the uncertainty of this modelled farm level N-loss in relation to the uncertainty of the key parameters identified in the sentivity analysis.

Sensitivity analyses (SA)

Approach

The following approach, based on previous pilot studies (e.g., Wheeler et al. (2020)), is employed for the sensitivity analyses:

- Inputs that significantly impact the modelled N loss at the total farm level are identified by a one-at-a-time (OAT) local SA (LSA). This method changes each input parameter around their nominal values one at a time and then quantifies the effect (sensitivity index) on the N-loss estimate.
- The influences and the effects of the interactions between the most significant parameters identified in the local SA (OAT analysis) are then quantified by a global SA (GSA). The sensitivity of the most influential parameters and their interactions are assessed with a two-level full factorial analysis using a variance-based method for each farm system.

Farm dataset

All analyses are performed at the farm level using anonymised farm setups in the Overseer database. A specific advantage of this approach is that the analyses are carried out in an operational, real-world context capturing, for example, both climate information and the diversity of farm management. The three different types of agricultural systems defined for this study are:

- Dairy farms: only dairy enterprise.
- Beef & Sheep farms: beef and/or sheep and/or deer enterprises.
- Cropping farms: only crops with no animal enterprise.

To limit bias in comparisons of results and ensure analyses represent real rather than hypothetical farm systems, comparable farm setups were also selected using the following criteria:

- Geolocated: to ensure the accuracy of climate data selection.
- Overseer 'Year-end' run type, describes the current farm system: to ensure the farm setup represents the farm system, not a scenario.
- Most recent 'Year-end' farm description file Overseer: to ensure the most up-to-date farm setup.

Local sensitivity analysis – OAT method

The most effective approach to a SA for a model with a significant number of input parameters is to vary each parameter of interest one at a time, while keeping the remaining parameters fixed at their nominal values. This type of analysis is called a local sensitivity analysis (LSA). The variation in a model output (in this case N-loss) due to the change in an input parameter quantifies the model's sensitivity to that variable. The sensitivity index (S_i) of an input parameter quantifies the rate of increase or decrease in a model output (i.e. N-loss). The method of calculation is explained in Overseer (2022a) and is based on Helton (1993) and Borgonova (2008).

The OAT SA is carried out at the farm level, which means that the key parameters are identified at the farm level.

Parameter selection

With more than 100 inputs or parameters, the most likely parameters influencing the N-loss estimate are selected to better manage the analyses. The selection of parameters is informed by previous reports on this topic e.g., Wheeler et al. (2020), experts' advice and users' feedback.

This process identifies 46 (input and internal) parameters as potentially influencing the N loss for the farms with animals, and 36 parameters are identified for the crop farms. The full description of these parameters can be found in Overseer (2022a).

LSA results

For each selected parameter and each selected farm, the sensitivity index (S_i) is determined by varying the value of each parameter around its nominal value, while the other parameters remain constant. The S_i values are then classified by type of farms. The distributions obtained thus classify the influence of the parameters on the N-loss at the farm level. The influence of a parameter is characterised by the interquartile range (IQR) of the distribution. IQR is the width of an interval that contains the middle 50% of the values in the distribution. The interquartile ranges of the sensitivity index distributions of each selected input parameter for dairy farms are presented in Figure 1.

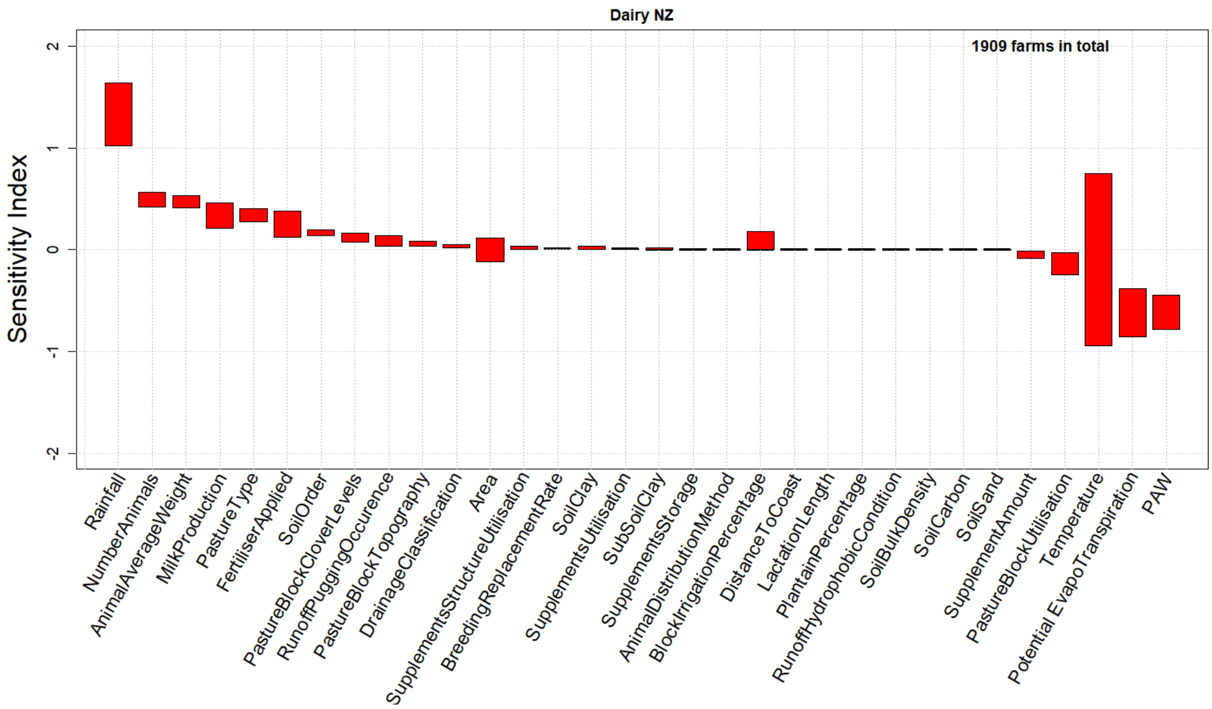

Figure 1: Interquartile range of the sensitivity index for the selected dairy farms.

The main conclusions of the analysis are:

- The most influential input parameters are rainfall, potential evapotranspiration (PET), temperature, soil profile available water (PAW), applied fertiliser, number of animals, milk production, average animal weight, distributed supplements, and pasture utilisation.
- The large IRQ for temperature is due to the fact that certain processes modelled in Overseer have a nonlinear relationship with the temperature. For example, the pasture growth rate increases with the temperature above a certain threshold and can be limited by low soil moisture, which is also partly a function of temperature, thus giving a

complex relationship between pasture growth and temperature. Therefore, the width of the S_i distribution interquartile range across New Zealand is larger than the others because of the different temperature and soil moisture ranges across New Zealand.

All the results of the LSA concerning the other enterprises as well as for the different regions and for the different types of farms can be found in Overseer(2022a). For all types of farms, the climate parameters (rainfall, temperature, and PET) and the soil water holding capacity (from which PAW is calculated) are the most influential input parameters.

Global sensitivity analysis (GSA)

GSA, based on a variance method, estimates individual parameter variables' statistical influence and interactions with other parameters (Sobol & Kuchereko, 1993). The sensitivity of a given input parameter measures its contribution to the model output variance. The two standard sensitivity indices are (1) the first-order index (or 'main effects'), which measures the direct contribution to the model output variance, and (2) the total-order index (or 'total effects') which measures the overall contribution (direct and indirect through interactions with the other parameters of interest) to the model output variance (Homma & Saltelli, 1996). The first-order index used in this analysis is also called global sensitivity index (GSI) and ranks the parameters' influence. The computation of the GSI in this study used a full factorial design and a classical analysis of variance decomposition (Lamboni et al.,2011).

Parameter selection

The main parameters identified as the most influential with the OAT analysis are the inputs for the GSA. The GSA follows a two-level full factorial design. The parameters are discretised in two levels, 'low' and 'high'. For each farm, the low and the high level are set as the parameter's nominal value minus 25% and plus 25%, respectively. This design requires 2M model evaluations per farm, where M is the number of parameters.

The GSA is carried out with a subset of farms selected randomly across New Zealand. It is anticipated that future analyses will include expanding the number of farms and parameters to better understand the model's sensitivities.

GSA result for dairy farms

The GSIs (Global Sensitivity Index) for a subset of 109 randomly distributed dairy farms across New Zealand are presented in Figure 2. The first observation is that most of the variance in the model output can be explained through the input parameters in the analysis, such that only 1% ,on average, of the output variance cannot be explained (Figure 2, in *pale yellow*).

The second observation is that the direct effect of key input parameters explains 92±3% of the N loss variance. Interactions between key input parameters (Figure 2, in grey) can be attributed to only 7±2% of the variance in N loss. The relative importance of these interactions is also tested with a linear regression model using the 11 most influential parameters as inputs. The relative weakness of interactions for each studied farm is confirmed by an average adjusted R-squared value of 0.92±0.03 from a linear regression between N loss and key input parameters. Consequently, the sensitivity indexes (S_i), which quantify the impact of input parameters, may also be used as indicators of the direction of change of N-loss when an input parameter varies.

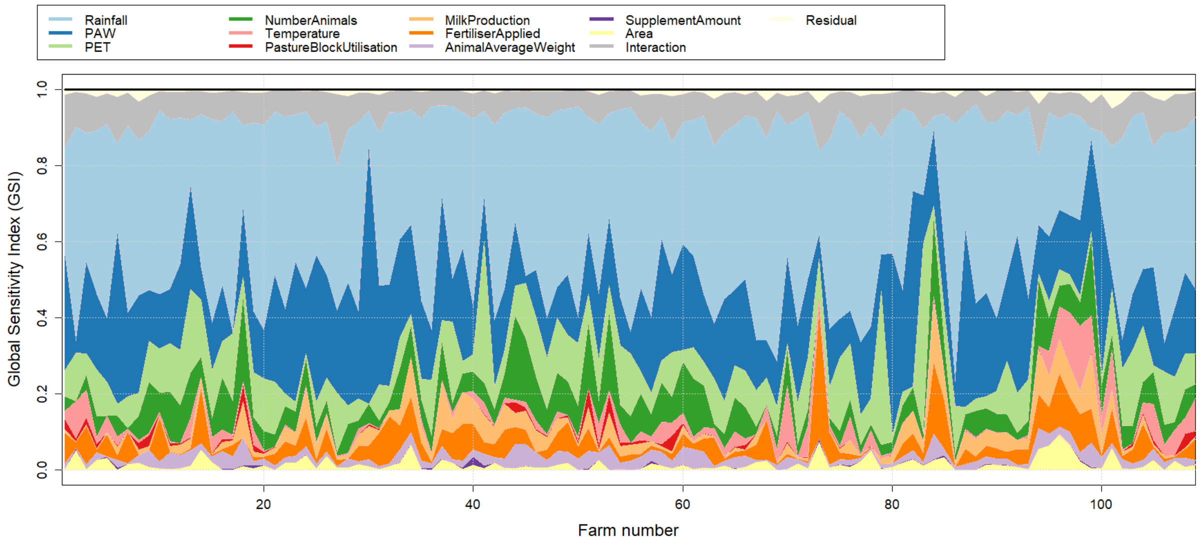

Figure 2: GSI (proportion of variance explained) for each selected dairy farm.

The GSA results for the different types of farms can be found in Overseer (2022a). Overall, the GSA confirmed the results from the OAT analysis, whereby rainfall and profile available water (PAW) are identified as the most influential parameters. This finding was expected as N-loss is governed by drainage, which is mainly a function of rainfall and the capacity of the soil to contain the water.

Uncertainty analysis

The aim of this analysis is to determine the uncertainty relating to the most influential parameters (rainfall, temperature, PET, and soil water holding capacity) common across three main farm systems in Overseer (dairy, beef & sheep, cropping) and to quantify the uncertainty in the N-loss estimate due to these key parameters. In this report, the uncertainty of a parameter defines the variability or dispersion of the values assigned to this parameter. It is quantified by the standard deviation of the distribution of the assigned values.

In line with the sensitivity analyses, the uncertainty analysis is based on real farm systems and conditions referenced in the Overseer database. The Monte Carlo technique is used to propagate the uncertainty of the rainfall, temperature, PET, and soil water holding capacity (from which PAW is calculated) parameters within the model by sampling the distributions of these input parameters to quantify the uncertainty of the N-loss estimate.

Method

Climate uncertainties

Knowing the averages and standard deviations of the climatic input data makes it possible to view each climate input parameter as a random variable normally distributed as a first approximation. The central limit theorem (30 monthly measurements over the period 1991-2020) allows us to create reasonable statistical models of sample averages, hypothesising a normal distribution (Stirzaker, 2003). The model uses a monthly scale for input climate data. Accordingly, each month is treated independently and represented by a normal distribution, a mean, and a standard deviation estimated at the farm's location. For a given location, each normal monthly distribution is independently randomly sampled (Monte Carlo method) and used as an input parameter to calculate the estimate of N-loss. The process is repeated 100 times.

Soil water holding capacity uncertainties

Overseer uses S-map soil data, where available, to provide users with soil property information (Lilburne et al., 2012). Soil information is provided as a map of areas containing one or more soils (siblings). A sibling is a member of a soil family. Further information about S-map and soil properties is available on the Landcare (Manaaki Whenua) website (https://smap.landcareresearch.co.nz/). Each sibling has a defined set of soil parameters. As the output parameters of the S-map model are strongly correlated, Landcare provides 100 parameter sets ('realisations') per S-map sibling, giving a reasonable and statistically valid representation of the distribution in soil water holding capacity parameters (Lilburne et al., 2016). Therefore, the model is run 100 times with different soil water holding capacity values.

N-loss estimate uncertainty

The resulting replications are used to assess the variance of the N-loss estimate distribution. The mean and the standard deviation of the N-loss estimate distribution are subsequently used to assess the average uncertainty in the N-loss estimate due to the input uncertainty. The relative standard deviation (RSD), defined as the ratio of the standard deviation to the mean of the N-loss estimate distribution, is used to compare the uncertainties between parameters whose values are variable.

The uncertainty analysis is carried out on the same farm datasets as the sensitivity analyses, further requiring soil characteristics to be provided by S-map.

N-loss estimate uncertainty due to soil and climate data

The uncertainties of the N-loss estimate due to the combined effect of the climate and soil uncertainties were determined by sampling both the climate and soil information distributions for the 2175 farms with S-map data comprised of 1166 dairy, 960 beef & sheep, and 49 crop farms.

The standard deviation of the N-loss distribution for soil and climate was estimated by sampling different climate data distributions and soil realisations in combination with 100 replications per farm. Figure 3 shows the RSD versus the N-loss mean for each studied farm. The mean and the standard deviation of the N-loss RSD values when farms are stratified into 5 groups with an equal number of farms are also represented by red crosses.

The combined uncertainties of the soil and climate data result in an average uncertainty of N-loss of 27±9% for all the farms studied. Overall, the uncertainty results for N-loss are equivalent across the different types of farming (Overseer, 2022b).

For N-loss values less than 10 kg N/ha/yr, the average relative uncertainty reaches 35% with a significative variability (±15%), which is explained by the fact that Overseer is a threshold model. This type of model predicts no effect below a critical value which depends on a set of thresholds, while an effect of some magnitude exists above this value. This effect leads to significant variability near threshold values, i.e., at low N-loss values in the case of Overseer.

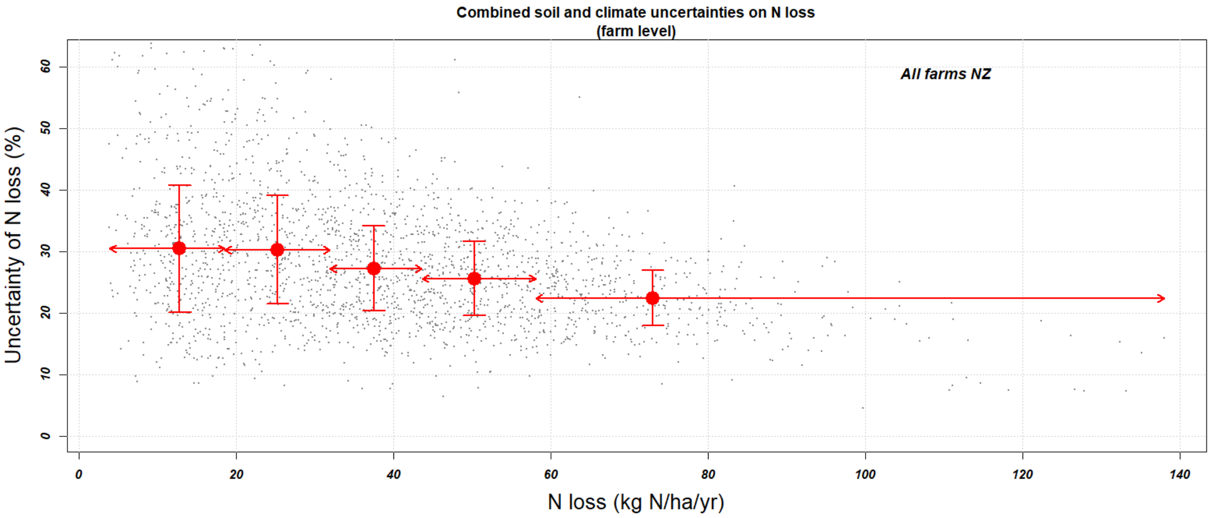

Figure 3: Relative standard deviation (RSD) of the N-loss distribution versus the N-loss mean for each studied farm for the different types of farming referenced in Overseer when climate and soil parameters uncertainty distributions are sampled. Each point stands for a farm. Red crosses represent the mean and standard deviation of the RSD when farms are stratified into 5 groups with an equal number of farms.

Conversely, beyond 100 kg N/ha/yr, the average uncertainty on the N-loss estimated by Overseer is 15±5%. The general trend is a decrease in N-loss uncertainty with increasing N-loss estimates. This observation is similar for the variability (±5%), allowing to narrow the confidence interval with the high values of N-loss. The results for the different types of farms referenced in the Overseer can be found in Overseer (2022b).

Conclusions

The LSA quantifies the influence of input parameters on the N-loss estimates at the farm level. The most influential input parameters are identified for farm types referenced in Overseer. Across all types of farms, climate data (rainfall, PET, temperature) and soil water holding capacity are the most crucial input parameters for the N-loss estimate at the farm level.

The GSA results indicate that the impact of the interactions between the studied input parameters on farm level N-loss estimates is considered to be weak compared to the direct effects. Under these conditions, the value of sensitivity indexes (S_i) can be used directly to predict the direction of change of N-loss due to a variation in the input parameters, i.e., a variation of x% of an input parameter leads to a variation of S_i*x% in the N-loss estimate at the farm level.

For all types of farms, the climate parameters (rainfall, temperature, and PET) and the soil water holding capacity are the most influential input parameters.

The combination of the soil and climate data uncertainties lead to an average uncertainty on the N-loss estimated by Overseer of 27±9% for the studied farms (2175 in total). This value increases to 35±10% at low values of N-loss (<10 kg N/ha/yr) and decreases to 25±7% when the value of the N-loss is greater than 40 kg N/ha/yr. The average uncertainty is less than 20% when the N-loss value exceeds 70 kg N/ha/yr.

In the absence of significant interactions between influential parameters, it is possible to use the sensitivity index to predict the contribution of a parameter's uncertainty to the uncertainty in the N-loss estimate. In this context, the parameters defining the climate and the soil contribute most of the uncertainty on the losses of nitrogen. The contribution of the other parameters is minimal.

Future improvements of the sensitivity analyses should consider widening the range of inputs e.g., monthly animal numbers or the different uses of farm structures. For the uncertainty analysis, the other sources of uncertainty (e.g., internal parameters and calibration data) should be characterised.

Acknowledgement

The authors would like to thank Linda Lilburne (Manaaki Whenua Landcare) and Andrew Tait (National Institute of Water & Atmospheric) who provided input parameter uncertainties. This data was crucial for this work.

References

Borgonova, E. (2008) Sensitivity analysis of model output with input constraints: a generalized rationale for local methods. Risk Analysis, 28 (2008), pp. 667-680

Homma, T. and Saltelli, A. (1996) Importance measures in global sensitivity analysis of nonlinear models. Reliability Engineering and System Safety, vol. 52, pp. 1–17, 1996.

Lamboni, M., Makowski, D., Monod, H., 2011. Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models. Reliability Engineering & System Safety, volume 96, pp. 450-459.

Lilburne L., Carrick S., Webb T. and Moir J. (2012) Computer-based evaluation of methods to sample nitrate leached from grazed pasture. Soil Use and Management, 28 (1) (2012), pp. 19-26.

Lilburne L., McNeill S., Cuthill T., and Roudier P. (2016) Exploring the uncertainty of soil water holding capacity information. Spatial Accuracy 2016, Montpellier, France, 5-8th July 2016.

Overseer Limited (2022a) Sensitivity Analyses on Nitrate Loss at the farm level estimated by OVERSEER. Available from http://www.overseer.co.nz.

Overseer Limited (2022b) Uncertainty Analysis – Propagation of Input Climate Data and Soil Uncertainties on Nitrate-loss estimated by OVERSEER. Available from http://www.overseer.co.nz.

Sobol, I. and Kuchereko, S. (1993) Sensitivity estimates for nonlinear mathematical models. Mathematical Modelling and Computational Experiments, vol. 1, pp. 407–414, 1993.

Wheeler, D. M., Meenken, E.D., Green, P and Mercer, G.J.K (2020) Sensitivity Analysis of Overseer: Preliminary analysis of input parameters affecting intake. Report for Overseer Limited RE450/2020/024

Wheeler (2022), Overseer Technical Manuals. Available from http://www.overseer.co.nz.