Tavernet, J.P. 2024. Overseer assessment of modelling nitrogen leaching under pasture. In: Opportunities for improved farm and catchment outcomes (Eds. C.L. Christensen, D.J. Horne and R. Singh). http://flrc.massey.ac.nz/publications.html. Occasional Report No. 36. Farmed Landscapes Research Centre, Massey University, Palmerston North, New Zealand. 7 pages

OVERSEER ASSESSMENT OF MODELLING NITROGEN LEACHING UNDER PASTURE

Jean-Paul Tavernet

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

Abstract

The Overseer model – derived from over 30 years of research into New Zealand (NZ) farming environments and conditions – is an agricultural management tool based on scientific principles to estimate nutrient flows in productive agricultural systems. These estimates identify potential environmental risks related to GHG emissions and nutrient loss through runoff and leaching. Model evaluations are essential to determine the robustness of Overseer and give confidence in its estimates. This study aimed to perform a specific type of evaluation that compared the Overseer model paddock-scale N leaching estimates with experimental data from paddock-scale grazed pastures and showed the challenges and limitations of such an exercise. This work was primarily based on research by Selbie *et al.* (2020) that reviews the published literature on N leaching from grazed pastures (dairy, sheep, and beef farms) in combination with Smith *et al.* (2020), which describes the method to configure these experimental measurements into the Overseer model framework. The adopted approach selects the appropriate experimental data according to the conditions of use of the model.

With the proposed approach, the comparison of Overseer N leaching estimates with experimental data from NZ paddock-scale grazing systems shows a linear correlation and produces a ‘very good’ performance rating using the metrics of Moriasi et al. (2007). Similarly, the comparison of the N-model leaching estimates with experimental data produces a ‘satisfactory’ to ‘good’ performance rating for a limited number of cropping systems. This work also shows the importance of the selection of experimental data, the challenges associated with representing these measurements in the model, and the limitations of such comparisons when applied to a small number of experimental sites.

The complementary use of alternative evaluation methods, such as scenario analysis or comparison with process-based models, seems reasonable due to the methodological limits of comparisons with experimental data. Using different types of evaluations will constitute a body of evidence, further confirming the performance of the Overseer model.

Introduction

Scientific models play a decisive role in various fields, including environmental regulation. It is essential to recognise that models are not infallible predictors of reality but rather indispensable tools for guiding decision-making processes. This is because models are inherently limited by computing power, underlying assumptions, and incomplete knowledge. As the National Research Council (NRC) (NRC, 2007) states, models should not be seen as generators of absolute truth but rather as information tools. Perfect models encompassing all processes of reality are and will remain inaccessible. These intrinsic limitations of models suggest that their evaluation/assessment should be considered an essential and ongoing part of the model life cycle.

OverseerFM®, cited as Overseer in this article, encompasses key elements relevant to the study of nutrient flows and greenhouse gas (GHG) emissions in most New Zealand agricultural systems. Given the difficulty of directly measuring these flows or emissions at the farm level, models such as Overseer are used. Overseer considers the complex interaction between agricultural management, soils, plants, animals, and climate to estimate nutrient losses and GHG emissions at the farm scale. However, like any mathematical model, Overseer simplifies complex processes through a set of equations and therefore has inherent assumptions and limitations. When applied beyond these limits, Overseer's results, like those of other models, should be interpreted with caution due to potential inaccuracies. The main assumptions and limitations, as well as the implications of their use, are discussed in Freeman et al. (2016).

Regular assessment/evaluation, calibration, and validation of the model is essential for continuous development and improvement. Nitrogen (N) leaching estimates from the Overseer model are evaluated by comparing them with experimental data at different sub-model levels (e.g., animal ME, hydrology, etc.) despite the inherent challenges. This comparison can be applied to each sub-model contributing to the overall estimate of N leaching. However, due to the inability to directly measure N leaching at the farm scale, these comparisons are limited to the block level.

Separate reports summarise the evaluation of specific sub-models against block-scale experimental data for drainage (Shepherd, 2019), cropping systems (Brown et al., 2020), inter-urine patch areas (Shepherd & Selbie, 2019), dry matter and N intake (Shepherd et al., 2020), urine patch (Shepherd & Selbie, 2020) and pastoral block N model (Shepherd et al., 2020). These reports generally indicate reasonable consistency between the sub-model results and the available experimental data, except for the pastoral block model, which showed a weaker correlation.

This report aims to evaluate the performance of the Overseer N model against experimental data from grazed pastures at the paddock scale. Initially, the focus will be on the challenges and associated limitations of comparing Overseer model estimates with experimental data. Subsequently, the performance assessment will be presented. Finally, potential prospects for long-term evaluation will be discussed.

Data sourcing

This study uses data on N leaching loss measurements in New Zealand pastoral blocks, compiled comprehensively in Selbie *et al.* (2020). The data comes from field experiments conducted at the scale of a paddock, commonly referred to as a farmlet.

A brief overview of these experimental setups can be found in the appendices of Welten *et al.* (2021), with various insightful observations and comments highlighted in Shepherd *et al.* (2015). For a more condensed summary of the experimental sites carried out until 2015, we can refer to Watkins and Shepherd (2014).

The methodology for preparing experimental data and configuring Overseer files to reflect field site measurements and treatments is detailed in Smith *et al.* (2020). Specifically, individual Overseer files were created in Overseer for each unique combination of site, treatment, soil type and year. This ensures a granular representation of the experimental design. Subsequently, the results from each year within a treatment, soil type and site combination were aggregated to provide an average value for each treatment per site.

Challenges and limitations

The main challenges and limitations encountered when comparing the Overseer model results with experimental N leaching data in New Zealand highlight the complexities involved with such a task. The main challenges (Selbie *et al.*, 2020) are:

- Disparity in experimental purposes: existing experimental data originating from research conducted at paddock-scale, focuses on specific management practices and their impacts on productivity, profitability, and environmental sustainability rather than model assessment work, resulting in a heterogeneous collection of measurements (not consistent across all experiments).
- Short experiment duration: full comparison with the Overseer would require long-term averages (10+ years) of N leaching at various research sites, but most experiments typically last 1 to 3 years and are potentially biased by extreme leaching events or atypical weather conditions.
- Limited geographical scope and measurement uncertainty: the available experimental data come from only a few sites in a limited number of geographical areas, with varying numbers of measurements per site. This limits the data's ability to fully reflect the diversity of soil types, climates, and agricultural systems across New Zealand, compounded by the uncertainties inherent in N leaching measurements.
- Uncertainty in Overseer file configuration: evaluating the Overseer model's N leaching requires setting up individual Overseer files for each experimental site and treatment, but missing information from published experiments can lead to uncertainties in the modelled results.
- Outlier management: despite careful data selection, outliers within the dataset were identified. These outliers likely arise from specific situations like pre-experimental land-use practices or unusually wet years. Statistical analyses and expert judgment were used to identify and address such outliers, enabling a more trustworthy data set for comparison between experimental and modelled data.

Despite these limitations, careful data selection, including appropriate handling of outliers, and adequacy of the experimental design with Overseer (e.g. measuring nitrogen leaching at 60 cm depth) helps ensure there is a more robust evaluation of the Overseer model performance.

Comparison with experimental data from grazed pastures

To ensure the validity of the comparison between N leaching estimates generated by the Overseer model and measurements obtained from various experimental sites across New Zealand (farmlets), outliers and measurements falling outside the model's intended application domain have been excluded from our analysis. This process is further outlined in Table 4 of Appendix C in Tavernet et al. (2023). The resulting dataset is presented in Table 1.

Region	**Site**	**Total number of measurements**
Manawatu	Massey DCG	6
	Massey P21	2
	Grasslands Ruz-Jerez	3
Otago	Kelso	3
	Telford P21	9
Southland	Edendale	12
	Tussock creek (2009)	20

	Tussock creek (2016)	11
Waikato	Ruakura N	9
	Scott farm (RED)	30
	Scott farm (prototype)	27
	Scott farm P21	40
	Lake Taupo Hoog	6

Table 1: Summary of paddock-scale experiments (farmlets) used to assess Overseer performance against experimental data from grazed pastures.

The remaining measurements, representing various agricultural systems and environmental conditions, form the basis for the comparison between experimental N leaching and the Overseer model estimates.

This study uses three established performance indicators to assess the agreement between Overseer estimates and measured N leaching data. These indicators, detailed in Moriasi *et al.* (2007), are:

i. Nash-Sutcliffe Efficiency (NSE): this metric measures the normalised root mean square difference between the estimated and observed values.
ii. Percentage bias (PBIAS): this indicator expresses the average tendency of the model to underestimate or overestimate the measured values.
iii. Ratio of root mean square error to standard deviation of measured data (RSR): this metric compares the root mean square error of the model to the variability of the measured data.

Consistent with the criteria described in Table 2, lower RSR values and higher NSE values indicate better model performance. Table 2 provides a detailed breakdown of the different performance scores associated with these indicator values.

Performance rating	RSR	NSE	(N, P) PBIAS (%)
Very good	$0.00 \leq RSR \leq 0.50$	$0.75 < NSE \leq 1.00$	$PBIAS < \pm 25$
Good	$0.50 < RSR \leq 0.60$	$0.65 < NSE \leq 0.75$	$\pm 25 \leq PBIAS < \pm 40$
Satisfactory	$0.60 < RSR \leq 0.70$	$0.50 < NSE \leq 0.65$	$\pm 40 \leq PBIAS < \pm 70$
Unsatisfactory	$RSR > 0.70$	$NSE \leq 0.50$	$PBIAS \geq \pm 70$

Table 2: General performance ratings for recommended statistics for a monthly time step model, extracted from Moriasi et al. (2007).

To evaluate the performance of the model, two comparison approaches were undertaken.

The first approach, shown in Figure 1a, involves a direct comparison of modelled N leaching averages over years (experiment duration) by site, treatment and soil type with experimental data obtained from selected pastures.

The second approach, shown in Figure 1b, focuses on a comparison that mitigates potential regional biases. Modelled N leaching averages are compared to experimental data averaged over years by site and treatment, encompassing measurements across all soil types within each site-treatment combination. This approach reduces the influence of the Waikato region, which has more measurements than other regions. Therefore, this methodology facilitates a more representative model performance assessment at the national level, ensuring that regional variations in data availability do not influence the overall assessment.

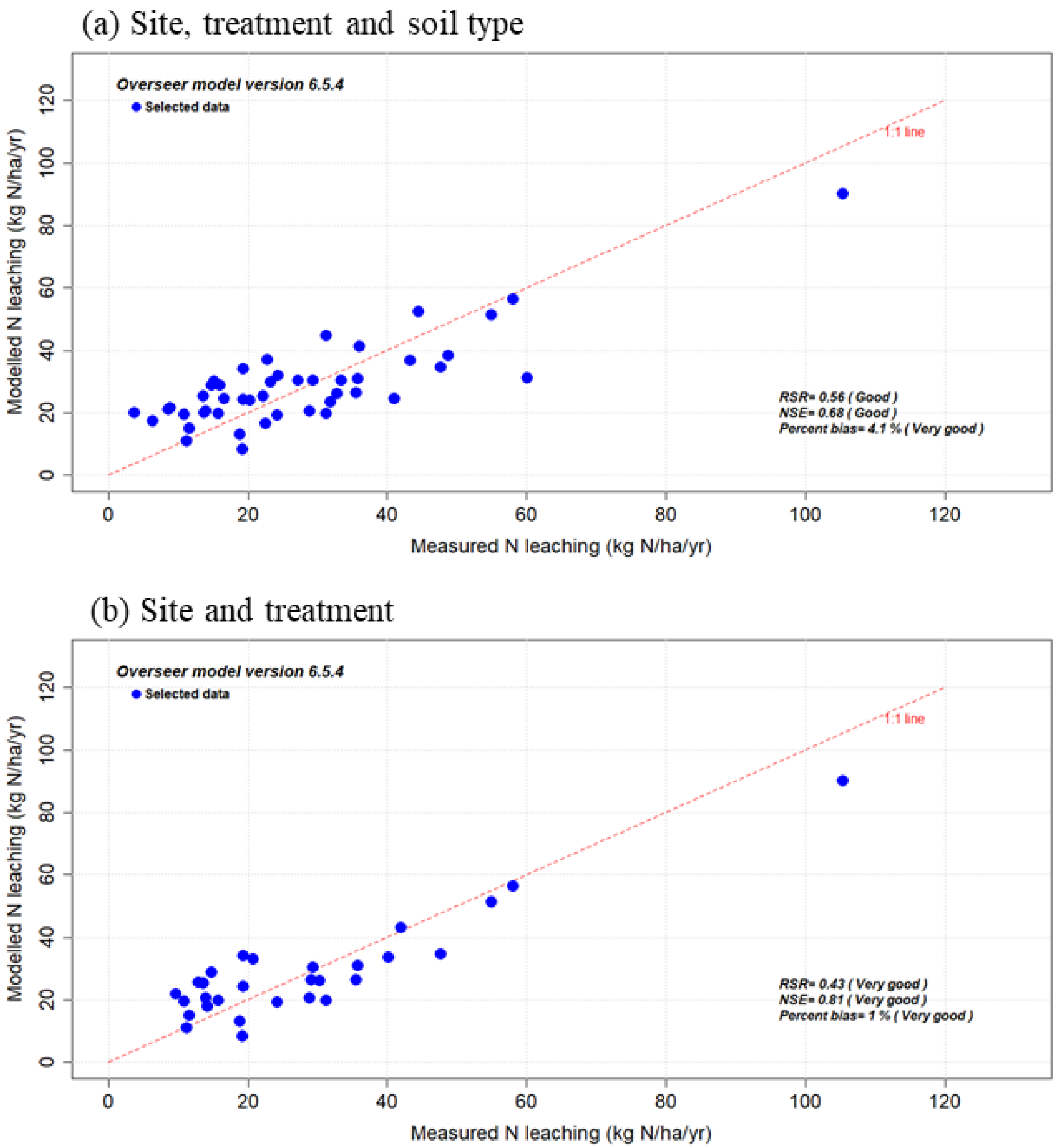

Figure 1: Comparison between experimental and modelled annual N leaching loss (kg N/ha/year) based on measured data from New Zealand paddock-scale grazing system studies and modelled using the Overseer model. Data are the mean of measurements (at least 2) over the years (a) per site, treatment, and soil types, (b) per site and treatment. The values of the three indications and the performance metrics described in Moriasi et al. (2007) are also shown.

The performance of the Overseer model for estimating N leaching from grazed pastures, is rated from 'good' to 'very good' based on Moriasi *et al.* (2007). The RSR, NSE, and PBIAS values collectively suggest that the Overseer model provides reliable estimates of N leaching from grazed pastures.

However, it is crucial to acknowledge a limitation in Figure 1, where average values are represented by single points that are not equivalent due to the significant variation in the number of measurements used to calculate them. This discrepancy in measurement numbers should be considered when interpreting the results and comparing the performance of the Overseer model across different sites and treatments.

Perspectives on long-term evaluation of model performance

Evaluating the performance of the Overseer model in estimating N leaching requires a long-term perspective and the implementation of complementary qualitative and quantitative methods to address the limitations of traditional experimental data comparison approaches.

A promising approach is scenario (or sensibility) analysis. Scenarios would be developed in collaboration with scientists and would likely encompass variations in farm management practices (e.g., stocking rates, fertiliser application) and farm characteristics (e.g., soil types, climate). The Overseer science team will then work with relevant experts to evaluate these scenarios using experimental data, results from other models, or expert knowledge. This project is currently in the scoping phase, with the aim of evaluating the first scenarios over the coming years.

Another evaluation option is to compare Overseer's N leaching estimates with results from process-based models like Agricultural Production Systems sIMulator (APSIM, Keating *et al.*, 2003). Previous work compared Overseer's N leaching estimates with APSIM's 20-year averages for specific cropping systems. Although the results showed consistency, the study highlighted the need to calibrate certain input variables due to the increased flexibility (or customizability) of APSIM compared to the Overseer model. This type of comparison could be extended in the long-term to encompass a wider range of New Zealand agricultural systems.

The integration of scenario analysis and process-based model comparisons into the Overseer model evaluation strategy will provide insight into its performance in estimating N leaching and improve model confidence in various agricultural contexts.

Conclusions

The assessment of the performance of the Overseer model for New Zealand grazing systems at paddock scale reveals a 'very good' performance rating when we compare model estimates with experimental data, using the approach described in this study. This result is consistent with previous assessments (Welten *et al.*, 2021; Anon, 2012), demonstrating the effectiveness of the model for grazed pastoral systems, which represent approximately three-quarters of New Zealand's agricultural land.

However, it is crucial to broaden the comparison of the Overseer estimates with experimental data from other agricultural systems.

Finally, to address the methodological limitations of comparisons with experimental data, the implementation of complementary methods, such as scenario analyses and comparisons with process-based models, is recommended to further justify the performance and limitations of the model.

References

Anon, B. 2012. Changes to the pastoral N model. Overseer FM Technical note 5, August 2012.

Brown, H., Khaembah, E.N. 2020. Evaluation of the overseer nitrogen model for cropping systems. Plant&Food research, PFR SPTS No. 20143.

Freeman, M., Robson, M., Lilburne, L., McCallum-Clark, M., Cooke, A., & McNae, D. 2016. Using OVERSEER in regulation – technical resources and guidance for the appropriate and consistent use of OVERSEER by regional councils, August 2016. Report prepared by Freeman Environmental Ltd for the OVERSEER Guidance Project Board.

Keating, B. A., Carberry, P. S., Hammer, G. L., Probert, M. E., Robertson, M. J., Holzworth, D., … & Smith, C. J. (2003). An overview of APSIM, a model designed for farming systems simulation. European journal of agronomy, 18(3-4), 267-288.

Moriasi, D.N., Arnold, J.G., Liew, M., Bingner, R.L., Harmel, R.D., & Veith, T.L. (2007). Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. Transactions of the ASABE, 50, 885-900.

National Research Council (2007). Models in Environmental Regulatory Decision Making. Committee on Models in the Regulatory Decision Process, Board on Environmental Studies and Toxicology, Division on Earth, and Life Studies. Washington, D.C.: National Academies Press.

Selbie, D., Smith, C., Mercer, G., and Shepherd, M. (2020). Evaluation of the OVERSEER® pastoral block N model (v.6.3.1): 1. Selection of farmlet experiments. Report for Overseer Limited. Client report number: re450/2020/051

Shepherd, M., Wheeler, D., Freeman, M., and Selbie, D. (2015). Rationale for Overseer® Nutrient Budgets evaluation and recalibration. AgResearch report no. RE500/2015/034.

Shepherd, M. (2019). Evaluation and validation of the OVERSEER® drainage model (v.6.3.1). Report for Strategic Science Investment Fund. AgResearch Report no. RE450/2019/052.

Shepherd, M., Selbie, D. (2019). Evaluation and validation of the OVERSEER® pastoral background N leaching sub-model (v.6.3.1). Report for Strategic Science Investment Fund. AgResearch Report no. RE450/2019/040.

Shepherd, M., Smith, L.C., Selbie, D. (2020). Comparison of Overseer-estimated dry matter and nitrogen intake with measured data. Report for AgResearch Ltd. AgResearch Report no. RE450/2020/029.

Shepherd, M., Selbie, D. (2020). An evaluation of the OVERSEER® urine patch N leaching sub-model. Report prepared for Strategic Science Investment Fund. AgResearch Report no. RE450/2020/022.

Shepherd, M., Selbie, D., Smith, L.C., Mercer, G. (2020). Evaluation of the OVERSEER® pastoral block N model (v.6.3.1) 3. Results. Report for Overseer Limited. AgResearch report no. RE450/2020/053.

Smith, C., Selbie, D., Mercer, G., and Shepherd, M. (2020). Evaluation of the OVERSEER® pastoral block N model (v.6.3.1): 2. Setting up files & systems. AgResearch report no. RE450/2020/053.

Tavernet, J-P. (2023). Assessment of Overseer model performance with experimental data from grazed pastures. Report prepared for MPI.

Watkins, N. L., and Shepherd, M. A. (2014). A compendium of New Zealand pasture farmlet experiments measuring nitrogen leaching. Nutrient management for the farm, catchment, and community. L. D. Currie and C. L. Christensen. Palmerston North, New Zealand, Fertilizer and Lime Research Centre, Massey University. 27.

Welten, B., Mercer, G., Smith, C., Sprosen, M., and Ledgard, S. (AgResearch); with contribution from: Khaembah, E., Fraser, T., and Thomas, S. (Plant and Food Research) 2021. Refining estimates of nitrogen leaching for the New Zealand agricultural greenhouse gas inventory. Report for Ministry for Primary Industries, client report number: RE450/2020/077.