Tavernet, J.P., 2023. IMPACT OF USING DIFFERENT CLIMATE DATASETS IN 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. 9 pages

IMPACT OF USING DIFFERENT CLIMATE DATASETS IN THE OVERSEER MODEL

Jean-Paul Tavernet

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

Abstract

OverseerFM is a support tool for farm management that uses a combination of inputs, including climate datasets provided by NIWA (National Institute of Water & Atmospheric), to determine N-loss estimates at a block and farm scale. This study investigates the impact of using climate datasets with different temporal resolutions on the N-loss estimates generated by the Overseer science model, without increasing data input requirements for the end-user.

The study focuses on the impact of using daily resolution data within the hydrology sub-model in the first instance. The paper will show the comparisons of anonymised Overseer N-loss estimates using the following:

- the existing 30-year long-term average monthly scale climate dataset used from version 6.4.3 of the model,
- the average of 30 individual climate years of N-loss estimates using monthly scale climate data,
- the average of 30 individual climate years of N-loss estimates using daily scale climate data.

The overall findings on the impact of using different climate datasets will be discussed in the context of recently calculated estimates of model output uncertainty relating to climate inputs.

Introduction

Overseer is a long-term decision support tool designed to estimate long-term average farm-scale nutrient loss and greenhouse gas (GHG) emissions from a farm system. The tool assesses the impact of long-term farm management practice changes on these estimates and supports farm management decision-making to support environmental planning.

The Overseer model is based on the hypotheses that inputs are in equilibrium with production and that management and site characteristics (soil and climate) are relatively consistent over multiple years. This approach limits the number of inputs requested from users, as described in the introduction chapter of the Overseer Technical Manual (Wheeler et al., 2022). Overseer uses location-specific climate conditions from its long-term climate database, in line with the internationally recognised timeframe for defining average climate (Tait, 2022).

The recent review of the N-loss component of the Overseer model (MPI (Ministry for Primary Industries), 2021) increased interest in assessing the impact of using higher temporal resolution climate data on Overseer's N-loss estimate.

This study assesses the impact of using climate datasets of different resolutions in the Overseer model on the N-loss estimate for different farm systems.

Methodology

Climate data

The Overseer model uses a long-term monthly average climate dataset generated by NIWA and is based on 30 years of climate observations (1991 to 2020) collected at climate stations around New Zealand. The climate data is produced by interpolating these observations on a 500m spatial resolution grid across New Zealand (Wratt, 2006). This provides long-term monthly average climate (L-Av) data for rainfall, potential evapotranspiration (PET) and temperature. Since farms are geolocated in OverseerFM, the monthly climate data used for a given farm is that of the nearest point on the L-Av climate dataset grid.

As most Overseer sub-models operate on monthly time steps, monthly data is directly used. Indeed, all sub-models using temperature progress in monthly timesteps. However, the hydrological sub-model that calculates the soil water balance uses daily time steps to model, among other things, drainage, and surface runoff.

Rainfall and PET daily data used by the hydrology sub-model are obtained by distributing monthly values into daily values according to a "daily patterns" mechanism based on pseudo climatic regions defined in Overseer. Rutherford et al. (2008) defined 15 pseudo-regions, which are not necessarily contiguous in geographic space, based on the amount of precipitation and the strength of seasonality (Wheeler, 2022). This allows monthly precipitation and PET values to be broken down into a pattern of typical daily values based on the pseudo region in which a farm is located. Water budgets from the hydrology sub-model are then aggregated up to monthly values for further calculations to align with the monthly time step in the other sub-models.

Climate datasets

This study investigates the impact of using three different NIWA climate datasets (1991-2020) on the Overseer N-loss estimate:

- Current 30-year average monthly dataset (L-Av)
    - Observations made at climate stations around the country are interpolated to generate estimated climate data overlaying a spatial 500m resolution grid across NZ (Tait, 2022).
    - Long-term, 30-year average of monthly estimates of rainfall, PET, and temperature.
- Thirty individual years of monthly climate data (M-Val)
    - Monthly data overlaying a spatial 500m resolution grid across NZ generated by interpolation, as outlined for the L-Av dataset above.
    - Total monthly estimates of rainfall, PET, and average monthly temperature.
- Thirty years of daily climate data (D-Val)
    - NIWA's daily VCSN (Virtual Climate Station Network) dataset, generated on a spatial 5 km resolution grid across NZ as described in Tait et al. (2006).
    - Daily estimates of rainfall, PET, and the average temperature.

Farm dataset

The impact on N-loss estimates using different climate data sets is assessed by comparing the different estimates obtained for a defined set of analyses. In the Overseer farm database, a farm is represented by several description files depending on the analysis requested or the year of the request. Selecting farms means selecting one description file for each farm. To limit bias when comparing results, 6178 anonymised farms with similar sources of inputs were selected using the following criteria:

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

This approach is designed to ensure that the analyses selected represent real rather than hypothetical farm systems.

Method

For a given farm, N-loss is estimated for each climate dataset. For the different estimates, only the input climate data is modified, e.g., year-to-year management decisions (fertiliser application, irrigation, animal distribution) remain unchanged regardless of the climatic conditions of the year studied. Any interpretation of results generated by this approach needs to account for the temporal disconnect between climate and management data. For example: monthly irrigation management does not change with the daily climate dataset.

The process for inputting the different datasets and calculating the different N-loss estimates is summarised below (see also *Figure 1*):

- L-Av: Average long-term monthly climate data (1991-2020):
  The daily rainfall and PET are computed using the "daily pattern" process. The model is run once, and a single N-loss estimate is obtained per farm, denoted $Nloss_{longterm}$.
- MVal: 30 years of actual monthly climate data (1991-2020):
  The same as the L-Av, except that each year of monthly climate data is run through the model. There are 30 N-loss estimates calculated per farm, one per year, denoted $Nloss_{monthly_i}$, where i is a year from 1991 to 2020. The 30-year average of these estimates is denoted $\overline{Nloss_{monthly}}$.
- DVal: 30 years of daily climate data (1991-2020):
  The model is run 30 times, once for each year of daily climate data. As a result, there are 30 N-loss estimates per farm, one per year, denoted $Nloss_{daily_i}$ where i is a year from 1991 to 2020. The 30-year average of these estimates is denoted $\overline{Nloss_{daily}}$.
  The hydrology sub-model uses the daily climate data directly; the monthly values required by the other sub-models are obtained by aggregating the daily values. The monthly rainfall and PET are the monthly sum of the daily rainfall and PET, respectively. The monthly temperature is the monthly average of the mean daily temperature.

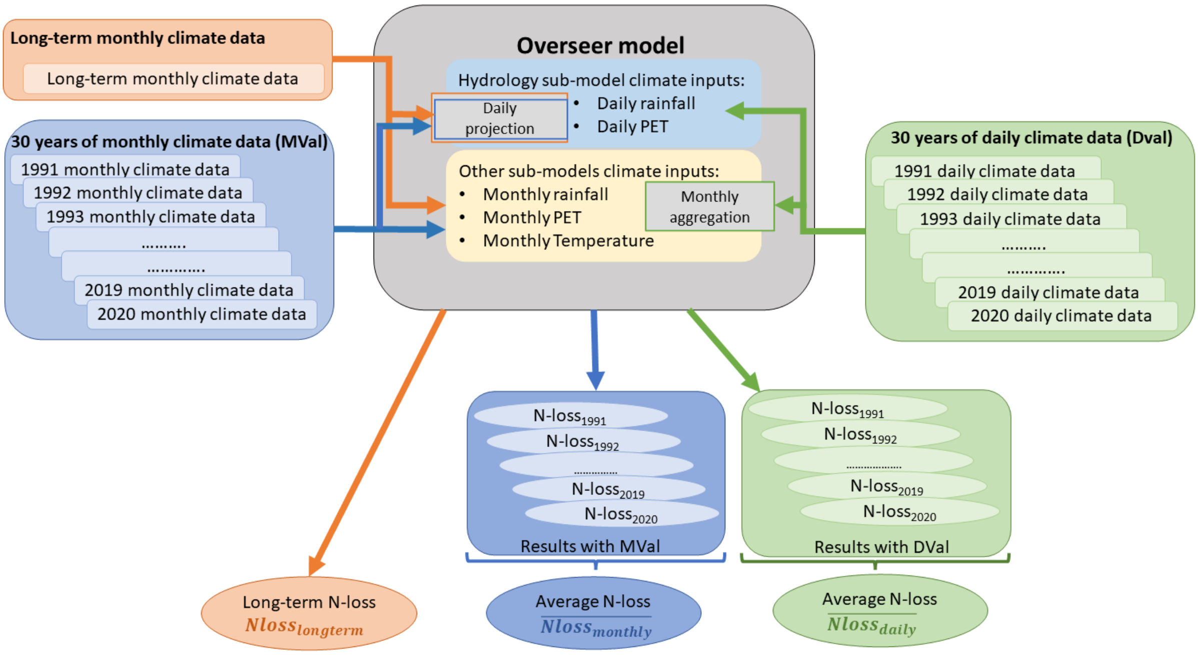

Figure 1: Description of the process for calculating the different N-loss estimates obtained from the different climate datasets.

The different results obtained for a given farm are summarised in Figure 2 as an example.

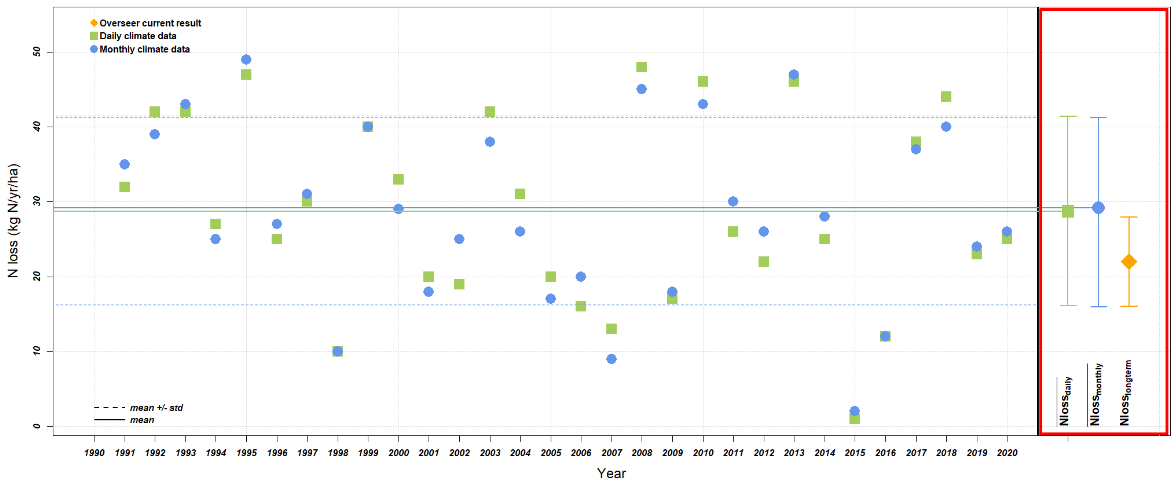

Figure 2: 30 years of annual N-loss estimates from DVal (blue) and MVal (green). Means (solid lines) and standard deviations (dotted lines) are represented by the coloured lines and projected into error bars in the red rectangle. The long-term N-loss estimate with uncertainty (Overseer, 2022a) is indicated in orange.

Results

Year-to-year N-loss variability

The variability of the estimated N-loss from year to year ($Nloss_{monthly_i}$ or $Nloss_{daily_i}$) for each farm (6178) is measured by the Relative Standard Deviation (RSD or coefficient of variation), defined as the ratio of the standard deviation by the mean of the modelled N loss distribution.

The distribution of RSD values for all selected farms when using the MVal and DVal datasets is characterised by an average and a standard deviation of 23±7% and 24±8%, respectively (Overseer, 2022b). The origin of the year-to-year variability illustrated in Figure 3 was hypothesised to be due to annual rainfall, as our previous sensitivity analysis work identified this variable significantly influences the N-loss estimate (Overseer, 2022a). This hypothesis was investigated by calculating the correlation coefficient between the annual rainfalls and the N-loss estimates using MVal for all the selected farms. With an average correlation coefficient of 0.8±0.1, the year-to-year variability of the N-loss estimates can be primarily explained by the year-to-year variation of the annual rainfall (Overseer, 2022c).

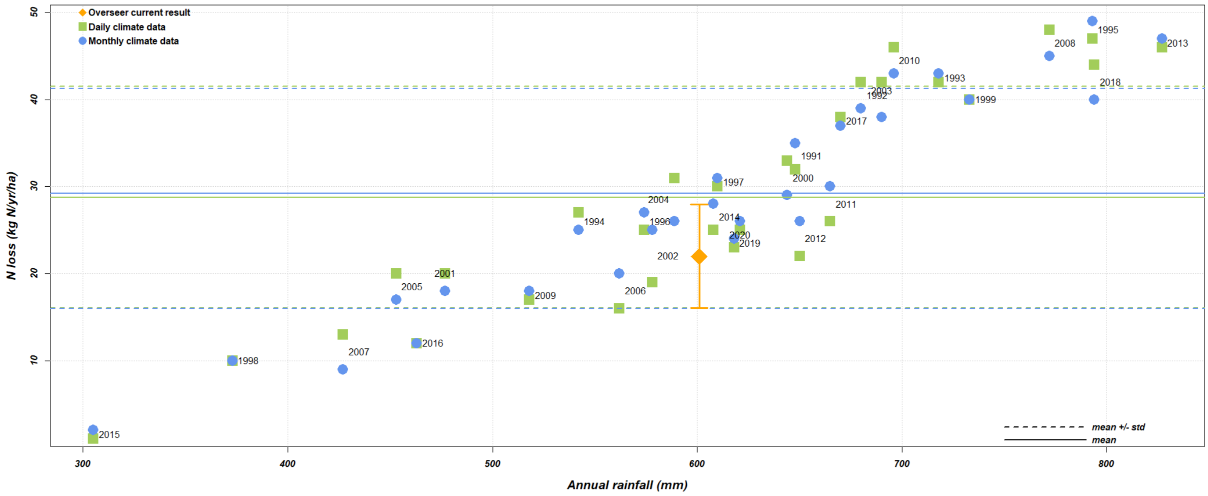

Figure 3: Example of comparison of the N-loss estimates from DVal and MVal datasets versus the annual rainfalls. The year of the climate data used is indicated.

N-loss comparison of the MVal and DVal datasets

Figure 4 shows the results for the 0.1th and 99.9th percentile of the distributions, representing the comparison of the averaged N-loss estimates for selected farms. The averaged N-loss per farm is produced by averaging the annual N-loss estimates from DVal and MVal datasets.

The percentage difference in the distribution of the N-loss estimates (*Figure 4*, inset) is defined as:

$$\Delta_{ratio}\,(\%) = \frac{\overline{Nloss_{daily}} - \overline{Nloss_{monthly}}}{\overline{Nloss_{monthly}}} * 100$$

$\overline{Nloss_{daily}}$ and $\overline{Nloss_{monthly}}$ are the N-loss averages obtained using the DVal and MVal datasets, respectively.

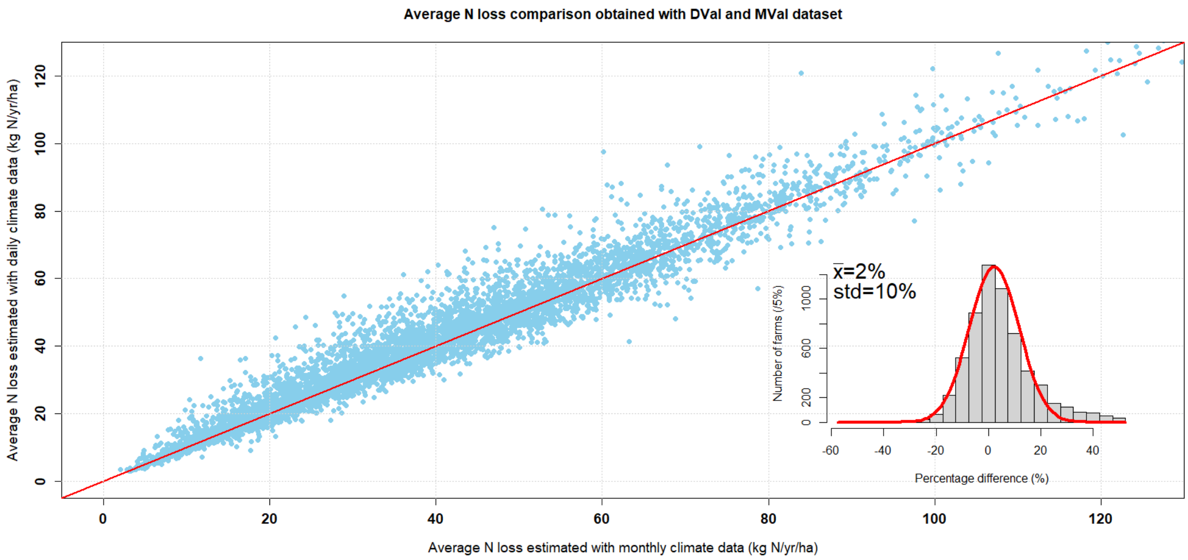

Figure 4: Comparison of the averaged N-loss estimates for 6178 farms using the DVal and MVal datasets. Inset: Percentage difference of the distribution.

The results show a narrow distribution with a standard deviation of 10% and a bias of 2%. With a p-value of 10^{-4} obtained using the Kolmogorov-Smirnov test (Williams, 2001), the averaged N-loss estimates obtained with the two climate data sets are statistically comparable (Overseer, 2022c).

A positive distribution tail was observed where the N-loss estimates were ≥20% higher using the DVal dataset compared to the MVal dataset. The distribution's tail has a couple of reasons:

- Possible extreme episodic daily climatic events or atypical weekly climatic patterns in the DVal dataset. These events are absent in the MVal dataset because daily climatic data are determined using typical values ("daily patterns" mechanism).
- The difference in the spatial resolution of the Dval dataset (5 km) compared to the MVal dataset (500 m). The impact of the difference in resolution is difficult to quantify because the landscape profile must be considered farm by farm.

N-loss comparisons with L-Av dataset

For each farm, the long-term N-loss is compared to the average of N-loss results obtained with the MVal dataset in Figure 5.

The percentage difference in the distribution of the N-loss estimates (Figure 5, inset) is defined as:

$$\Delta_{ratio}\ (\%) = \frac{\overline{Nloss_{monthly}} - Nloss_{longterm}}{Nloss_{longterm}} * 100$$

where $Nloss_{longterm}$ is the N-loss estimate based on the L-Av dataset, and $\overline{Nloss_{monthly}}$ is the average of the N-loss estimates based on the MVal dataset.

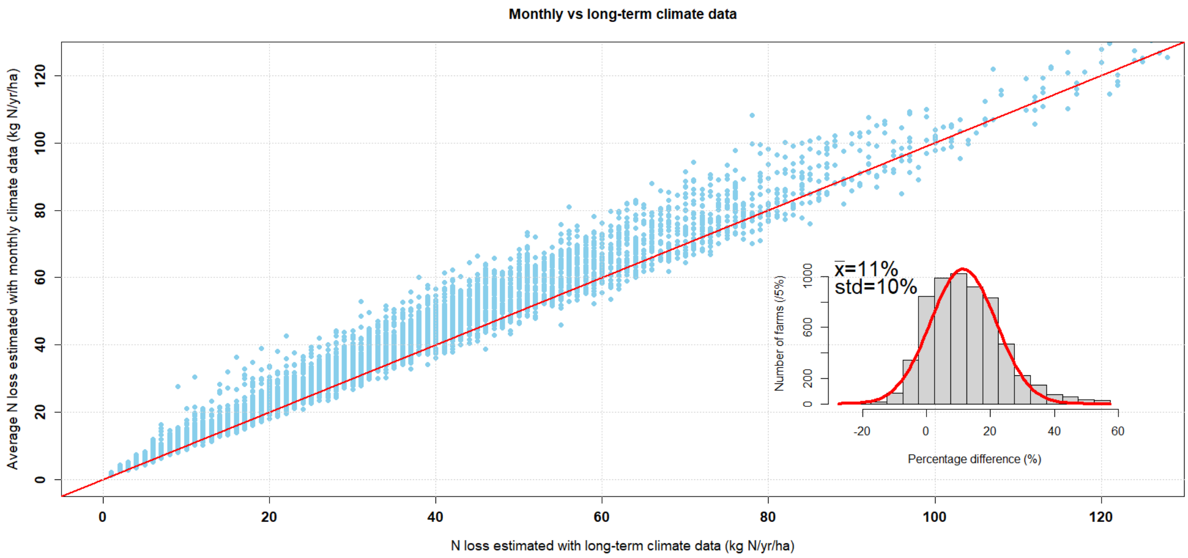

Figure 5: Comparison of the long-term annual N-loss estimate (L-Av) for 6178 farms with the averaged estimates using the MVal dataset. Inset: Percentage difference of the distribution.

The comparison with MVal shows a mean positive bias of 11% with a standard deviation of 10%. Thus, the MVal dataset shows N-loss estimates that are 11% higher than those obtained using the current climate dataset (L-Av). This difference can be explained as follows:

(1) Overseer, like other N-loss modelling tools, is a nonlinear model. This translates mathematically to Jensen's inequality (Jensen, 1906) which states that if F(x) is a nonlinear function, the mean of the function must differ from the value of the nonlinear function to the mean of the variables, ($\overline{F(x_i)} \neq F(\bar{x})$) where $\bar{x}$ is the average of x_i values. Any significant positive deviation from the average value is not offset by a negative balance; therefore, it is expected that the average of 30 annual N-loss estimates will be greater than the single long-term result. This naturally leads to lower estimates when climate event values are averaged (long-term climate dataset).

(2) To avoid the requirement for additional user input, the long-term average management practices are retained for the analysis with the MVal dataset, which could contribute to bias. For example, the long-term timing of fertiliser applications is used even if a specific month in a given year experiences above-average rainfall; this artificially increases the nitrogen loss estimate. Thus, to improve the accuracy of N loss, it is recommended to use the individual 30 years of monthly and/or daily management practices.

Conclusions

The climate parameters have a crucial influence on the N leaching estimated by Overseer. They are currently defined at the month level at the location of a farm from a long-term monthly climate dataset (L-Av) provided by NIWA. We studied the impact of different temporal-resolution input climate data on the N loss estimated by Overseer. The N-loss estimates generated using 30 years of interpolated monthly climate data (MVal), 30 years of interpolated daily climate data (DVal), and the L-Av climate data were compared, with the following main findings:

- The MVal and DVal N-loss estimates (annual and averaged) are statistically comparable. This provides confidence in the use of "daily climate patterns" for the generation of daily climate data used in the hydrology model and indicates that the use of interpolated daily climate data is unlikely to deliver material benefit over monthly data to the model output.
- Although not statistically significant, the trend towards higher N-loss estimates using the MVal and DVal datasets is likely to be an artefact due to (i) N-loss being a threshold process (Jensen) i.e., a stochastic process with only positive results and (ii) the difference in the temporal resolution between the long-term average management practices and annual climate datasets. An adjustment for year-to-year management practices could be studied to determine the impact of these on these average annual N-loss estimates.

The average N loss distributions obtained with the actual climate data (DVal and MVal) are consistent with the long-term N loss distribution if the uncertainties are considered. This means that the N loss comparisons give no benefits to one of the options over the two others. However, switching from the current option to using one of the two other climate data sets examined would imply the need for 30 years of management practices from users, or at least a revised definition of management practices coupled with additional costs and a loss of user-friendliness.

Acknowledgement

The authors thank Andrew Tait from NIWA (National Institute of Water & Atmospheric) for providing all the climate datasets and his valuable comments.

References

Jensen, J. L. W. V. (1906). "Sur les fonctions convexes et les inégalités entre les valeurs moyennes". Acta Mathematica. 30 (1): 175–193. doi:10.1007/BF02418571.

MPI, Ministry for Primary Industries (2021). Overseer whole-model review, Assessment of the model approach. Government of New Zealand. Available from https://apo.org.au/node/313610

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

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

Overseer Limites (2022,c) Assessment of the impact of using different temporal climate data on N-loss estimated by OVERSEER. Available from http://www.overseer.co.nz.

Rutherford K., McKergow L., and Rupp D. (2008) Nutrient attenuation and hydrology modules for Overseer NIWA Client Report: HAM2008-088. 75 pages.

Tait A., Henderson R., Turner R., and Zheng X. (2006). Thin plate smoothing spline interpolation of daily rainfall for New Zealand using a climatological rainfall surface. International Journal of Climatology, 26, pp 2097-2115.

Tait A. (2022) Personal communication.

Williams, D., Weighing the Odds: A Course in Probability and Statistics, Cambridge University Press, 2001, 548 p.

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

Wratt D.S., Tait A., Griffiths G. et al. (2006) Climate for crops: Integrating climate data with information about soils and crop requirements to reduce risks in agricultural decision-making. Journal of Applied Meteorology (13):305-315.