# Performance Assessment of OverseerFM Version 6.6.0

November 2026

## Summary

OverseerFM version 6.6.0 changes how crops take up nitrogen (N). Before release, the new version was tested against N leaching measured in field trials, and compared with the previous version, 6.5.13. The results are:

- **Pastoral systems: unchanged.** Modelled N leaching agrees with measurements as well as before, and is rated Very Good on all three performance statistics.
- **Cropping systems: improved.** On the same 40 trial points, agreement with measurements moves from Good to Very Good. Version 6.5.13 over-estimated crop leaching by 11% on average; version 6.6.0 is within about 3%.
- **New crop.** Italian ryegrass grown as a forage crop can now be modelled. With its three trial points included, performance remains Very Good.

## Introduction

Every change to the Overseer science model is tested against measured data before it is released, following Overseer's standard validation workflow (Tavernet, 2025). This report presents that test for version 6.6.0.

Version 6.6.0 makes three changes that affect N leaching from crops:

- **Updated crop coefficients.** Growth curves, harvest timing and the N concentration of the harvested product and crop residues were revised for most crops. Each crop now also has its own root mass, root N concentration and maximum luxury N uptake. Five new crops were added. The new coefficients come from joint work by Overseer Ltd and the Bioeconomy Science Institute (Khaembah et al., 2025, 2026), and was independently peer reviewed by a domain expert before release.
- **Dynamic crop N uptake.** A crop can now take up more N when more mineral N is available in the soil (luxury uptake), up to a limit set for each crop.
- **Root biomass correction.** Root biomass is now calculated as a share of total plant biomass, as intended, rather than of above-ground biomass. This slightly increases root N uptake and the N returned to the soil when roots break down.

Four other corrections in this release do not affect modelled N losses and are not covered here. All changes are described in the release notes (Overseer Ltd, 2026).

The assessment answers two questions:

- Is performance for pastoral systems unchanged? The changes are not intended to affect pasture.
- Does version 6.6.0 predict N leaching from crops closer to what was measured in field trials?

# How the model was tested

## Measured datasets

Two groups of field trials with measured N leaching were used (Table 1). In this report, a point is one trial treatment: its measured N leaching compared against the value modelled by OverseerFM.

- **Grazed pasture (farmlet) dataset.** Paddock-scale grazed pasture trials in New Zealand, including Scott Farm, Edendale and Tussock Creek. This dataset has been used in every previous Overseer performance assessment. The trials were selected and set up as described by Watkins and Shepherd (2014), Selbie et al. (2020) and Smith et al. (2020), and the assessment method is described by Tavernet (2023).
- **Cropping dataset.** The Lincoln crop rotation trial, a two-year Plant & Food Research trial of potato-based rotations under different fertiliser and irrigation (Khaembah et al., 2015), and five crop experiments added for this release (Revill et al., in prep).

| Dataset | System | Location | Points |
|---|---|---|---|
| Grazed pasture (farmlet) | Paddock-scale grazed pasture | New Zealand | 29 |
| Lincoln A Block, rotations 1 and 2 | Potato-based arable rotations | Lincoln, Canterbury | 12 |
| Added crop experiments: Carey et al. (2016, 2017, 2018), Malcolm et al. (in prep), Macdonald et al. (2005) | Oats and Annual/Italian ryegrass (forage crop system) sown after urine (field lysimeters and a growth chamber); cover crops before spring barley | Canterbury and Southland, New Zealand; Woburn and Rothamsted, United Kingdom | 31 |

*Table 1. Measured datasets used in this assessment.*

Until now, cropping performance was tested against the Lincoln trial alone: one site and one soil. The five added experiments cover more crops, climates and soils, including catch crops sown after winter forage grazing. They will remain part of the standard validation set for future releases.

The wheat crop in Rotation 2 of the Lincoln trial was affected by disease, which was identified from the trial records after the version 6.5.13 assessment. Its six analyses (reporting year ending 2006) were removed for both versions, so the comparison between versions is not affected. The Lincoln figures in this report are therefore not directly comparable with those in the version 6.5.13 assessment.

## How the two versions were compared

Each trial treatment was set up as an analysis in OverseerSci, following Smith et al. (2020) for the farmlets and Revill et al. (in prep) for the crop experiments, and run with both version 6.5.13 and version 6.6.0. Farm inputs, climate, soils, management and settings were identical, so any difference in results comes only from the science model changes between versions 6.5.13 and 6.6.0.

As in previous assessments, each trial was run with the climate recorded in each trial year, and where a treatment was measured over several years the values were averaged over those years (Tavernet, 2023).

The added experiments were brought together for this release as a preliminary set to assess the new crop model. They are of different types and scales: single-season field lysimeters with urine patches, one growth-chamber run (Carey et al., 2018), and cover-crop trials, rather than the multi-year averages of the farmlets. Only Malcolm et al. (under review) spans two seasons, 2021 and 2022, which are kept as separate points because the urine rates and sowing dates differed between years. These points therefore carry more year-to-year variation than the farmlet averages, and for some of them drainage was taken from the measurements. The results for these experiments should be read as a first indication of how the new crop model performs. A detailed and finalised assessment, including how each experiment was represented in OverseerSci and the results by type of experiment, will be given in Revill et al. (in prep).

Three points, the Italian ryegrass treatments of Carey et al. (2017), can only be run in version 6.6.0, because Italian ryegrass as a forage crop is one of the crops added in this release. The main comparison between versions therefore uses the 40 points that both versions can run. Results for all 43 points are also reported.

## How performance is measured

Agreement between modelled and measured N leaching is summarised with three statistics recommended by Moriasi et al. (2007) and used in all Overseer performance assessments:

- **Nash–Sutcliffe efficiency (NSE):** how well the model reproduces the differences between treatments. A value of 1 is a perfect match; 0 means the model does no better than the average of the measurements.
- **RMSE–observations standard deviation ratio (RSR):** the typical size of the model error, relative to how much the measurements vary. A value of 0 is a perfect match; lower is better.
- **Percent bias (PBIAS):** whether the model over- or under-estimates on average, calculated as 100 × Σ(modelled − measured) / Σ(measured). A positive value means the model over-estimates; 0 means no bias.

Each statistic is rated with the thresholds for N given by Moriasi et al. (2007) (Table 2). Results are also shown as 1:1 plots, with version 6.5.13 on the left and version 6.6.0 on the right. Each dot is one point; a dot on the dashed line means the modelled and measured values are the same.

| Rating | NSE | RSR | PBIAS |
|---|---|---|---|
| Very Good | > 0.75 | ≤ 0.50 | < 25% |
| Good | 0.65 to 0.75 | 0.50 to 0.60 | 25 to 40% |
| Satisfactory | 0.50 to 0.65 | 0.60 to 0.70 | 40 to 70% |
| Unsatisfactory | ≤ 0.50 | > 0.70 | ≥ 70% |

*Table 2. Performance ratings for N, after Moriasi et al. (2007).*

# Results: pastoral systems

Performance on the farmlet dataset is unchanged (Table 3, Figure 1). Both versions are rated Very Good on all three statistics, and the only difference is 0.1 percentage points of bias.

| Version | NSE | RSR | PBIAS | Number of data Points |
|---|---|---|---|---|
| 6.5.13 | 0.80 (Very Good) | 0.44 (Very Good) | +1.5% (Very Good) | 29 |
| 6.6.0 | 0.80 (Very Good) | 0.44 (Very Good) | +1.6% (Very Good) | 29 |

*Table 3. Performance statistics for the grazed pasture (farmlet) dataset.*

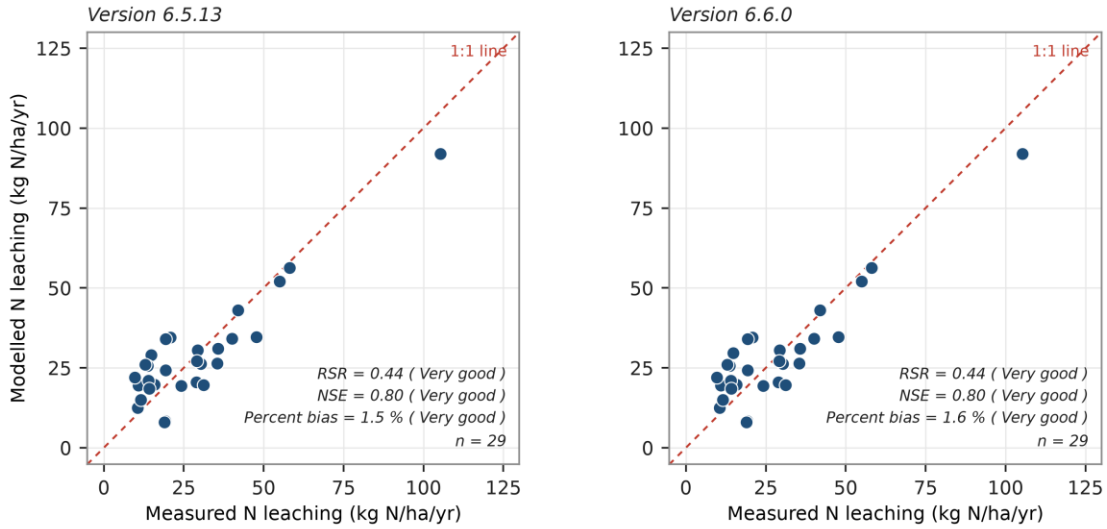



*Figure 1. Measured and modelled N leaching for the grazed pasture (farmlet) dataset, OverseerFM versions 6.5.13 (left) and 6.6.0 (right). The dashed red line is the 1:1 line.*

To check that the changes were applied as intended, every block in the farmlet dataset was also compared. Of 381 blocks, 21 gave a different result in version 6.6.0:

- 11 crop and fodder crop blocks, where the changes apply directly, with differences of up to 56 kg N/ha;
- 10 pasture blocks on farms that also have a crop block. All but two changed by 1 kg N/ha, through nutrients moved between blocks on the same farm.

No pasture block on a farm without a crop block changed. The farmlet points are calculated from the measured pasture blocks only, so crop and fodder crop blocks are not included directly. Their effect on the points is indirect and small: in one trial treatment, where part of the pasture block was rotated through turnips, two measured pasture blocks each changed by 1 kg N/ha. This moved one point slightly and accounts for the 0.1 percentage point difference in bias.

# Results: cropping systems

Version 6.6.0 predicts N leaching under crops more accurately than version 6.5.13 (Table 4, Figure 2). On the 40 points that both versions can run:

- Predictions follow the measurements more closely. Version 6.6.0 better reproduces which treatments leach more and which leach less, and its typical error is smaller, particularly at low and moderate leaching. In statistical terms, NSE rises from 0.73 to 0.84 and RSR falls from 0.51 to 0.39, which moves both from Good to Very Good.

- The average over-estimate has been significantly reduced. Version 6.5.13 predicted about 11% more leaching than was measured, on average. Version 6.6.0 is within about 3% of the measurements, slightly under rather than over (PBIAS changes from +11.3% to −3.1%).

| Version | NSE | RSR | PBIAS | Number of data Points |
|---|---|---|---|---|
| 6.5.13 | 0.73 (Good) | 0.51 (Good) | +11.3% (Very Good) | 40 |
| 6.6.0, same points | 0.84 (Very Good) | 0.39 (Very Good) | −3.1% (Very Good) | 40 |
| 6.6.0, all points | 0.87 (Very Good) | 0.36 (Very Good) | −1.4% (Very Good) | 43 |

*Table 4. Performance statistics for the cropping dataset. The first two rows compare the versions on the same 40 points. The third row adds the three Italian ryegrass points that only version 6.6.0 can run.*

With the three Italian ryegrass points included, version 6.6.0 remains Very Good on all three statistics. These points include the highest measured leaching in the dataset, around 350 kg N/ha/yr. A point with a large measured value tends to raise NSE, which is why the comparison on the same 40 points is the fairer measure of the improvement.

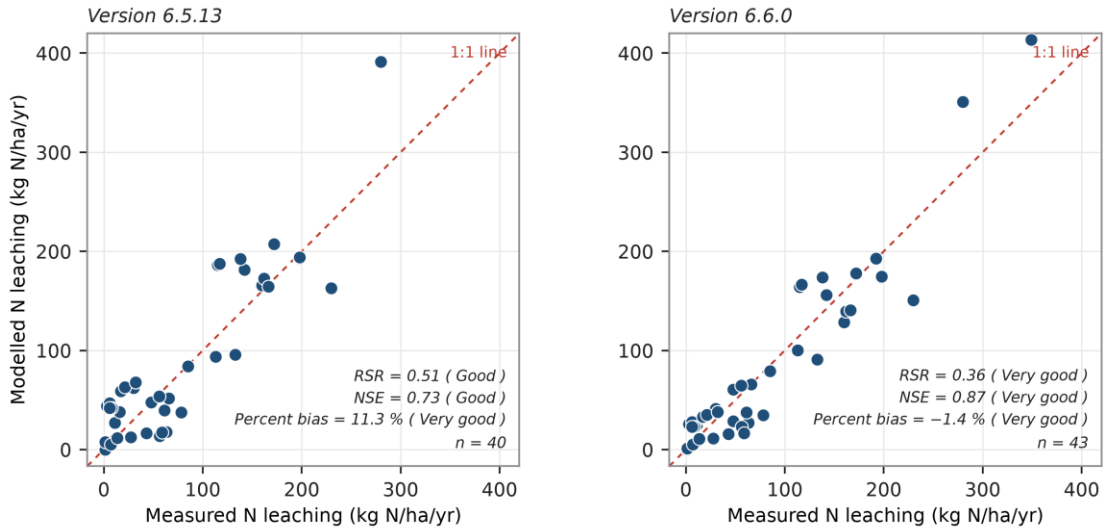



*Figure 2. Measured and modelled N leaching for the cropping dataset (Lincoln rotations and added crop experiments), OverseerFM versions 6.5.13 (left, 40 points) and 6.6.0 (right, 43 points). The dashed red line is the 1:1 line.*

In both versions, the largest differences from measurements are in treatments with high leaching, above about 100 kg N/ha/yr (Figure 2). Because most cropping points are single seasons, part of this scatter reflects the weather of individual years rather than the long-term behaviour of the crop system.

## What this means for Overseer users

- **Pastoral farms.** In the farmlet dataset, no pasture block on a farm without crops changed. Farms without crop blocks can expect the same N leaching results as with version 6.5.13.
- **Farms with crops.** N leaching from crop and fodder crop blocks will change. The size and direction of the change depend on the crop, its yield and the soil N available to it. Across the cropping trials, modelled leaching moved closer to the measured values and was lower on average. Pasture blocks on the same farm may change slightly, through nutrients moved between blocks.
- **A wider evidence base.** Crop leaching is now tested against 43 data points from six experiments in New Zealand and the United Kingdom, instead of a single trial at one site.
- **Reading individual results.** In everyday use, OverseerFM runs on long-term climate and estimates long-term average N leaching for a farm system, not the leaching of a particular year. In this assessment it was run with the actual weather of each trial year, so that its results could be compared with what was measured. As with any model, individual results can differ from measurements, particularly where leaching is high. The cropping dataset still covers a limited range of soils and climates, and more field trial data will be added as it becomes available.

## Conclusions

- Performance for pastoral systems is unchanged: NSE 0.80, RSR 0.44 and PBIAS +1.6%, all Very Good.
- Performance for cropping systems improves from Good to Very Good: on the same 40 points, NSE rises from 0.73 to 0.84, RSR falls from 0.51 to 0.39, and PBIAS moves from +11.3% to −3.1%.
- The changes act on crop and fodder crop blocks, as intended.

Version 6.6.0 improves how OverseerFM represents crop N uptake, without affecting its performance for pastoral systems.

## References

Carey, P.L., Cameron, K.C., Di, H.J., Edwards, G.R., & Chapman, D.F. (2016). Sowing a winter catch crop can reduce nitrate leaching losses from winter-applied urine under simulated forage grazing: a lysimeter study. Soil Use and Management, 32(3), 329–337. https://doi.org/10.1111/sum.12276

Carey, P.L., Cameron, K.C., Di, H.J., & Edwards, G.R. (2017). Comparison of nitrate leaching from oats and Italian ryegrass catch crops following simulated winter forage grazing: a field lysimeter study. New Zealand Journal of Agricultural Research, 60(3), 298–318. https://doi.org/10.1080/00288233.2017.1336103

Carey, P.L., Cameron, K.C., Di, H.J., & Edwards, G.R. (2018). Does sowing an oats catch crop reduce nitrate leaching from urine deposition following simulated winter forage grazing? A growth chamber experiment. Plant and Soil, 431, 37–52. https://doi.org/10.1007/s11104-018-3742-2

Khaembah, E.N., Brown, H.E., Sharp, J.M., & Zyskowski, R. (2015). Soil nitrogen and soil water dynamics in crop rotations: Estimation with the multiple crop single purpose model. In: Currie, L.D., & Burkitt, L.L. (Eds.), Moving farm systems to improved nutrient attenuation (Occasional Report No. 28). Fertilizer and Lime Research Centre, Massey University, Palmerston North, New Zealand.

Khaembah, E.N., Gee, M., Jenkins, H., Brown, H., & Sharp, J. (2025). Review of root biomass and root nitrogen concentration of the OVERSEER crop model. Bioeconomy Science Institute report prepared for Overseer Limited. SPTS No. 27932.

Khaembah, E.N., Revill, A., Malcolm, B., Brown, H., Jenkins, H., Tavernet, J.-P., Gellatly, S., Bruwer, M.-J., Kirikiri, K., & Cheetham, W. (2026). Improvement of nitrogen uptake in OverseerFM. Bioeconomy Science Institute report prepared for the Ministry for Primary Industries. SPTS No. 28480.

Macdonald, A.J., Poulton, P.R., Howe, M.T., Goulding, K.W.T., & Powlson, D.S. (2005). The use of cover crops in cereal-based cropping systems to control nitrate leaching in SE England. Plant and Soil, 273, 355–373. https://doi.org/10.1007/s11104-005-0193-3

Malcolm, B.J., Hu, W., Maley, S.C., Dellow, S., Taylor, A., Beare, M.H., & Jenkins, H. (under review). Catch crops reduce nitrogen leaching from forage crop grazing and mitigate tillage-induced soil mineralisation. Agriculture, Ecosystems and Environment. Preprint: https://doi.org/10.2139/ssrn.7049780

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., 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(3), 885–900. https://doi.org/10.13031/2013.23153

Overseer Ltd. (2026). Release notes for version 6.6.0. https://support.overseer.org.nz/hc/en-us/articles/63985503895321

Revill, A., Tavernet, J.-P., & Bruwer, M.-J. (in prep). Extended crop validation datasets for OverseerFM model performance assessment: selection of experiments, representation in OverseerSci, and derivation of performance indicators. Overseer Ltd.

Selbie, D., Smith, C., Mercer, G., & 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.

Smith, C., Selbie, D., Mercer, G., & Shepherd, M. (2020). Evaluation of the OVERSEER® pastoral block N model (v.6.3.1): 2. Setting up files & systems. Report for Overseer Limited. Client report number RE450/2020/053.

Tavernet, J.-P. (2023). Assessment of Overseer model performance with experimental data from grazed pastures. Overseer Ltd. https://support.overseer.org.nz/hc/en-us/article_attachments/49053437510809

Tavernet, J.-P. (2025). Validation process: standard workflow for updates to the science model. Overseer Ltd.

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