OVERSEER®: ACCURACY, PRECISION, ERROR AND UNCERTAINTY

Mark Shepherd, David Wheeler, Diana Selbie, Laura Buckthought & Mike Freeman

AgResearch Ltd, Ruakura Campus, Hamilton

Summary

When debating the performance of models such as *Overseer*'s ability to estimate whole-farm nutrient losses, four terms are often used almost interchangeably: accuracy, precision, error and uncertainty. However, the terms are not interchangeable and it is important to consider the implications of the commonly used terminology, in the context of this farm-scale nutrient budgeting model. Given that it is not usually practicable to directly measure whole-farm nutrient losses, use of the terms accuracy or error are not directly applicable, because there is no true value to compare an estimate with. Model uncertainty is the most relevant applicable term for annual whole-farm nutrient loss estimates. Model uncertainty will be greatest for conditions where there are no, or few, data for calibration and validation. Precision in the context of *Overseer* is about precision of input information.

Introduction

Agricultural nutrient models designed for practical application aim to provide decision support tools for farm advisors, farmers and consultants. They are especially useful for complex systems or when there is limited capability for actual measurements. For example, actual quantification of nitrogen (N) leaching from soil is difficult, particularly so at a paddock scale; all measurement techniques have advantages and disadvantages (Lilburne *et al*., 2012). Consequently, it is impractical to routinely monitor nutrient losses at an individual farm scale. However, farmers need to know the consequences of their farm management decisions if nutrient use efficiency is to be improved and catchment nutrient management goals met. For this reason, farm-scale models or decision support systems/tools are developed to model nutrient flows around the farm system.

There are a number of models in use in New Zealand, or are being developed, which aim to estimate N and phosphorus (P) losses from farm systems, and cover a broad range of scale and purposes (Cichota & Snow, 2009). This range of diverse models reflects both the different level of detail and scale at which N (and P) losses can be estimated.

Models such as OVERSEER® Nutrient Budgets ('*Overseer'*) (Wheeler *et al*., 2008) must involve simplifications of complex processes, and the predictions that such models make will always involve uncertainties. In discussions relating to models, their applicability and use, a number of terms are often used: accuracy, precision, error and uncertainty. The aim of this paper is to set these terms in the context of models developed for estimating farm-scale nutrient losses, with particular reference to *Overseer*.

Background to nutrient budget modelling

Model approaches

Cichota & Snow (2009) observed that the differences in model complexity relate mainly to the numbers of pools and processes used to calculate the nutrient balance and they categorised models as ‘complex’ or ‘simple’. For the purpose of this paper we have equated these terms with mechanistic and empirical modelling approaches, respectively. They can also be equated with ‘research’ and ‘application’ models. Research type models tend to be more complex; application models tend to be simpler in their approaches, by necessity.

Empirical models can be effective in summarising data and relationships and can provide practical tools for decision making. Empirical models are statistical descriptions of observed data. The relationships underpinning the overall model are typically based on experimental data. The main approach is to gather the data, design a single equation or set of equations, and fit these to the data. Consequently, the model describes the observed data. Because of this, empirical models do not give any indication of the factors and mechanisms that produce a given response, nor the possible reasons behind a response (Thornley and Johnson, 2000). Hence, extrapolation beyond the dataset used to develop the model might be problematic.

In contrast, mechanistic modelling aims to construct mathematical representations of the behaviour of a system based on the description of processes, thereby creating a deeper level of understanding. In mechanistic modelling, the system of interest is analytically broken down into components, to which processes and properties are assigned. There are a number of ways that a system can be separated into its components. Ultimately, the set of equations that characterise the system are integrated, and the responses of the system are constructed (Thornley and Johnson, 2000). Mechanistic models typically do not fit the observed data as well as empirical models, because there are many more assumptions built into them; or the models need ‘training’ to achieve this. However, the content of mechanistic models can be more comprehensive and applies to a greater range of systems and processes with the ability to interrelate them (Thornley and Johnson, 2000). If the underpinning science is sound, then there is more scope for extrapolating beyond datasets used to validate the model.

Model scale

Scale is an important consideration in model development. Many models are designed at the plot or field/paddock scale, while policy-makers usually need models that can be used at catchment, regional and/or national scales (Addiscott, 2003). Up-scaling and down-scaling models can cause problems. For example, up-scaling reduces the accuracy of inputs by ignoring a part of natural heterogeneity. Conversely, down-scaling requires increased accuracy (Addiscott, 2003). Validation and error propagation are also potential problems that need to be considered in up or down scaling models. This then raises the question of whether validation of a model at one hierarchical level is relevant to another, as well as parameterisation of a model and the ability of parameters to be transferred across scales (Addiscott, 2003); this is perhaps more of a challenge for simpler models.

Overseer

Overseer is a whole-farm nutrient budget model that provides users with a tool to examine the impact of nutrient use and flows within a farm (product, fertiliser, effluent, supplements or transfer by animals) on nutrient use efficiency and possible environmental losses. The model also provides a means to investigate mitigation options to reduce the environmental impact of nutrients within a land use. Users range from farmers and their consultants through to policy makers and policy implementers. The main assumptions underpinning the model are that: it uses long-term annual averages, i.e. the model assumes a 'steady state'; the system is in quasi-equilibrium (inputs commensurate with production levels on the farm); users supply actual and reasonable inputs; and management practice implemented on the farm follows good practice. Version 6 was released in 2013 and marked a major upgrade (software and science) of the model as summarised by Shepherd & Wheeler (2012).

The challenge with farm systems models (and often their purpose) is to model nutrient losses at a scale that cannot be practically measured, such as farm scale N leaching. Thus, *Overseer* is building up from specific component processes, such as those involved in urine patches, to the farm level, using accepted (and often published) relationships at the paddock or sub-paddock scale to model a representation of nutrient flows at the farm level.

Where does *Overseer* fit in the classification of Cichota & Snow (2009) of 'simple' and 'complex', as described earlier? We suggest: somewhere in between, i.e. a mixture of empirical and mechanistic. *Overseer* models nutrient transfers around a farm system to initially determine, for a given nutrient, when and how much is deposited on different parts of the farm (e.g., paddock, raceway, feedpad). Sub-models, varying in complexity and approach, then model the fate of these nutrients for each of the farm locations. For example, the N leaching model uses urine patch and 'background' sub-models to determine the fate of N sources (Shepherd & Wheeler, 2013).

Assessing model performance

The following terms are often used in reference to model performance: accuracy, precision, error and uncertainty. There are international definitions for these terms (Anon., 1993).

Accuracy

The accuracy of a measurement system is defined as the degree of closeness of measurements of a quantity to that quantity's actual (true) or accepted value (where actual measurement is impractical). The concept of accuracy has limited application to the estimation of whole-farm nutrient loss because of the great technical difficulty of quantitatively measuring these losses, such as N leaching.

Error

In a modelling context, error generally refers to the difference between the modelled representation of a system, and the reality of the system (Heuvelink, 1998). The primary types of error include input, model, and output error; and models could contain combinations of these:

- *Input error* - Model parameters such as soil properties and weather and/or climatic data always contain errors. Some of these may be "human error" or mistakes, and it is important to minimise this type of error.
- *Model error* - A fault in the model itself can arise from "concept error", i.e. an error in understanding, or deliberate simplification of the system being modelled; or errors in measured data from experiments used to calibrate and validate the model. There is no specific test for these kinds of errors, but they can be exposed by sensitivity analysis and review critique. Another possibility is "error in translation", where error occurs when converting the concept or theory into a set of mathematical equations and

computer code. Translation errors are revealed during model verification (Addiscott, 2003).

- *Output error* - Output error can be a result of input error, model error or both. However, the concept of an output error clearly has limited application where actual measurement is not practicable and there is no 'accepted' value.

Precision

This is also called reproducibility or repeatability, and is the degree to which repeated measurements under unchanged conditions show the same results. This concept has some applicability to *Overseer* nutrient loss estimates.

Uncertainty

Uncertainty **(**in the context of modelling) can be defined as a potential limitation in some part of the modelling process that is a result of incomplete knowledge. The sources of uncertainty in environmental modelling can be divided into five categories (Table 1).

The concept of uncertainty is the most applicable term relating to the use of *Overseer*, i.e., given the number of assumptions and errors involved in the model, there will be a level of uncertainty attached to estimates of nutrient losses.

***Table 1**. Sources of modelling uncertainty (based on Walker et al., 2003).*

Sources of modelling uncertainty	**Brief description and comment**
Context and framing	This can include choices about the physical boundaries of the system being modelled, the range of factors to incorporate into a model, and specific prediction choices.
Inputs	Uncertainties about inputs that drive the model, e.g. fertiliser, production, supplements, soil type, climate, etc.
Model structure	Models simplify reality and may be based on an incomplete understanding of the processes and structure(s) being modelled, e.g., the *Overseer* engine and our understanding of the underpinning science.
Parameters	Parameters used in the model need to be estimated or inferred from sometimes very limited data, e.g. parameters that drive the urine N leaching, crop N leaching, etc.
Model implementation	This can include technical modelling choices and potential software bugs.

Discussion

A mathematical model such as *Overseer* can be a useful conduit for making recent research available to farmers and advisors, particularly if executed in a user-friendly computer interface. *Overseer* aims to be such a model. However, users need information on performance and clarity of required inputs, particularly as its use is widening, from farm management to support catchment nutrient management policy implementation. In any model, there is always a proportion of observed data, currently accepted theory and conjecture based on the best available information). Research models generally have a reasonable level of explicit conjecture; however, application models must give predictions based on the limited data and information available - if the uncertainty associated with an application of the model is not considered when using that model, there could be serious consequences (Thornley and Johnson, 2000).

The current dilemma for modellers is that in general, although application models for agricultural systems are related to observational data, are user friendly, and mathematically simple, they have a strong element of empiricism (Thornley and Johnson, 2000) which tends to ignore the more detailed levels of physiological and biological theory. This can results in less sensitivity of the model to changes in the environment and farm management; it also limits extrapolations. As a result, there is increasing demand for greater scope and applicability of agricultural models, requiring a more mechanistic approach; this approach requires more inputs, so is also more likely to produce predictions with larger uncertainties, however. There will therefore be an on-going tension between simplicity and complexity in approaches for modelling complex farm systems. The level of simplification chosen by the modeller will always be criticised by colleagues and other scientists who consider the model to be either too complex or not complex enough or that their particular discipline is under-represented by the model (Thornley and Johnson, 2000).

The above comments are general to all models. Questions specific to *Overseer* typically concern error, accuracy, uncertainty and precision.

Error/accuracy/uncertainty

When interpreting a model's predictive abilities, it is important to know whether the model has been calibrated. This is the process of adjusting model parameter values to maximise the agreement between a given set of data and the model outputs (Refsgaard, 2000; Trucano *et al*., 2006). The next step in the application of a model like *Overseer* is to validate the model to provide a method of assessing the confidence in the modelled outputs (i.e., testing to see how well the model outputs fit a set of independent data: Jorgensen, 2003). *Overseers*' pastoral N leaching model has had a significant amount of validation (Shepherd & Wheeler, 2013), whereas the P loss model is based on a calibration process (McDowell *et al*., 2005).

There are two major challenges in assessing the performance of *Overseer*:

- Farm-level nutrient losses are practically difficult, if not impossible, to measure accurately, so benchmarks against which to compared modelled values are rare and also carry large uncertainty.
- *Overseer* can be used for a very wide range of farm systems in many different geographical settings; validation or calibration data for all circumstances are not possible.

These challenges are additional to accounting for the complexity of farm systems and any issues around accuracy of input data (discussed later) and immediately illustrates that estimating error or accuracy of outputs is not practicably possible, given that the determination of actual values is extremely difficult (and has a large error associated with these measurements: Lilburne *et al*., 2012). We therefore contend that it is more appropriate to consider the uncertainty attached to the model's outputs. The issue then becomes that the uncertainty associated with whole-farm nutrient loss estimates will increase for situations that are well outside the calibration/validation range (Figure 1).

More data for calibration/validation data will be required to decrease this uncertainty, most notably for: cropping and beef & sheep enterprises; clay and shallow and light textured soil types; and locations with high (>1200 mm) rainfall.

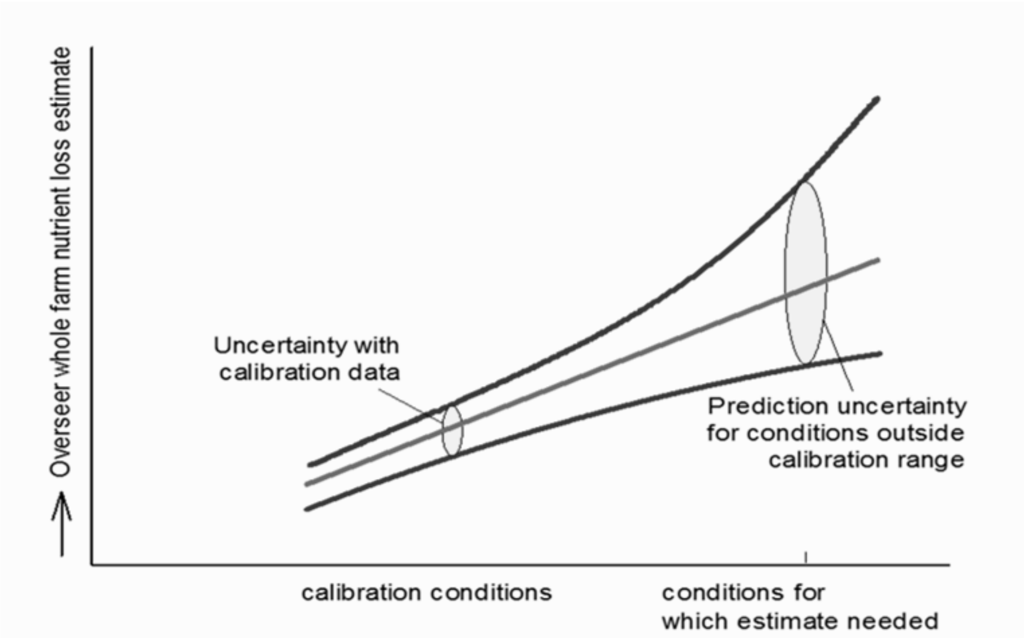

Figure 1. *Representation of Overseer uncertainty (based on Loucks et al., 2005)*

Precision

Given that precision is the measure of reproducibility or repeatability, for *Overseer*, this translates to the ability of multiple users to produce the same result. This relates to 'errors' around inputs, as discussed earlier. These 'errors' may well arise from differences in interpretation when the user is trying to translate a complex farm system into a moderate/manageable number of inputs. Improved precision of outputs will require users modelling the same (or similar) farm(s) to: (a) set up the *Overseer* file to represent the farm system in the same way, and (b) use the same input data (types and values of data). Model development has tried to address issues of data input by:

- Developing the model in such a way as to provide consistency of inputs between sectors and also between data entry methods if there is more than one way for entering data (e.g. entry of animal numbers).
- Developing the model using data, information and support structures (e.g. labels) that the farmer or consultant knows and understands.

However, setting up a farm system in *Overseer* requires a reasonable amount of interpretation and judgement by the user. The major limitation to improving precision can be potential differences in inputs entered by users. There is hence a need for guidelines for data entry and farm set-up. The development of industry-agreed 'protocols' or input guidelines will be critical for improving confidence in all applications of *Overseer.* With that, the way the farms are modelled will be consistent and hence will provide confidence in the model outputs, both in absolute and relative terms. This high level of sensitivity of whole-farm nutrient loss outputs to many input choices means that if meaningful whole-farm nutrient loss estimates are to be achieved, agreed protocols are essential.

Conclusions

Models like *Overseer* must involve simplifications of complex processes and the predictions that such models make will therefore always involve uncertainty. Given that it is not usually practicable to directly measure whole-farm nutrient losses, use of the terms accuracy or error are not directly applicable. Model uncertainty is the most relevant term for annual whole-farm nutrient loss estimates. Model uncertainty will be greatest in conditions where there are no, or few, data for calibration/validation. Precision in the context of *Overseer* is about precision of inputs. Better precision and reduction of uncertainty could be attained by developing comprehensive guidelines for entering input data into the model.

Acknowledgements

The *Overseer* development programme (2009-2013) is jointly funded by the Ministry for Primary Industries and the Fertiliser Association of New Zealand.

References

Addiscott, T.M. (2003). Modelling: Potential and Limitations. In: Benbi, D.K., Nieder, R. (Eds.), Handbook of Processes and Modelling in the Soil Plant System. Haworth Press Inc, New York.

Anon. (1993). Guide to the Expression of Uncertainty in Measurement. International Organization for Standardization (ISO) and the International Committee on Weights and Measures (CIPM): Switzerland.

Cichota, R. & Snow, V.O. (2009). Estimating nutrient loss to waterways - An overview of models of relevance to New Zealand pastoral farms. *New Zealand Journal of Agricultural Research* **52**, 239-260.

Jorgensen S.E. (2003). Environmental models and simulations, UNESCO, Encyclopedia of Life Support Systems (EOLSS).

Heuvelink, G.B.M. (1998). Error Propagation in Environmental Modelling with GIS. Taylor and Francis, London

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

Loucks D.P., van Beek E., Stedinger J.R., Dijkman, J.P.M., & Villars, M.T. (2005). Water resources systems planning and management: an introduction to methods, models and applications, Chapter 9 Model Sensitivity and Uncertainty Analysis, UNESCO.

McDowell, R.W., Monaghan, R.M. & Wheeler, D. (2005). Modelling phosphorus losses from pastoral farming systems in New Zealand. New *Zealand Journal of Agricultural Research* **48** 131-141.

Refsgaard, J.C. (2000). Towards a Formal Approach to Calibration and Validation of Models Using Spatial Data, in Grayson R. and Bloschl G, eds. Spatial Patterns in Catchment Hydrology: Observations and Modelling, Cambridge University Press.

Shepherd, M. & Wheeler, D. (2012). OVERSEER® Nutrient Budgets – the next generation. In: Advanced Nutrient Management: Gains from the Past - Goals for the Future. (Eds L.D. Currie and C L. Christensen). http://flrc.massey.ac.nz/publications.html. Occasional Report No. 25. Fertilizer and Lime Research Centre, Massey University, Palmerston North, New Zealand. 9] pages.

Shepherd, M. & Wheeler, D. (2013). How nitrogen is accounted for in OVERSEER® Nutrient Budgets. In these FLRC Workshop proceedings.

Thornley, J.H.M. & Johnson, I.R. (2000). Plant and Crop Modelling: A Mathematical Approach to Plant and Crop Physiology. The Blackburn Press, Caldwell, New Jersey.

Trucano, T.G., Swiler, L.P. Igusa, T. Oberkampf, W.L. & Pilch M., (2006). Calibration validation, and sensitivity analysis: What's what. *Reliability Engineering and System Safety* **91** 1331–1357.

Walker, W.E., Harremoës, P., Rotmans, J., van der Sluijs, J.P., van Asselt, M.B.A., Janssen, P. & Krayer von Krauss, M.P. (2003). Defining uncertainty. A conceptual basis for uncertainty management in model based decision support. *Integrated Assessment* **4,** 5-17.

Wheeler, D.M., Ledgard, S.F. & DeKlein, C.A.M. (2008). Using the OVERSEER nutrient budget model to estimate on-farm greenhouse gas emissions. Australian Journal of Experimental Agriculture 48, 99-103.