Review: Meta-analysis of nitrous oxide emission factors for excreta deposited onto pasture

Daniel Gerhard

26/09/2019

Objective

A meta analysis, combining data from several experiments at different locations in New Zealand, is used to obtain estimates of emission factors for various farm animals on various types of land. It should be investigated if some of the emission factors for different animal or land categories can be combined to reduce the number of factors, but still making an adequate prediction.

The report closely follows Kelliher et al. 2014, who presented the statistical analysis of emission factors for 185 field trials between the years 2000 and 2013.

Database

Each experiment in the database is designed in a similar way:
At a location several treatments, dung or urine from different animals, are placed in separate chambers at different N loads. Each experiment includes at least one control chamber with no addition of dung or urine. Most of the experiments are arranged in a block structure, with a replication for each treatment level and a control in each block.

There are some inconsistencies about reporting the blocking structure in the database. For some experiments no block identifiers are specified, although each control is assigned to a certain treatment level. For other experiments, the blocks might have been reported as separate trials.

The cumulative loss for each chamber is reported, which is a summary over multiple measurements from each chamber. Each of these measurements is again an estimate of flux at time zero, based on extrapolating repeated measurements over time. Each of these measurements are reported without a measure of uncertainty about the estimation.

The response used for the analysis is the emission factor, which is the difference between cumulative losses for a treatment and control divided by the N load times 100. This response is treated as an observation instead of an estimate, ignoring the corresponding estimation uncertainty. This will lead to an underestimation of the variability of the emission factors and also might result in biased estimates; but without any additional estimates of uncertainty in the database, there is no way of incorporating these in the data analysis.

Statistical Model

Logarithmic transformation

Based on visual inspection of the observed EF_3, increasing variability with an increase in mean values can be observed, which leads to a skewed distribution with some extreme values of high EF_3. As the cumulative losses are estimates based on extrapolated values for a flux at time zero, negative values can occur; and with the EF_3 being calculated as differences of cumulative losses, any additional sources of variability in these estimates can cause additional negative values for EF_3.

The proposed logarithmic transformation of the EF_3 is a reasonable solution to represent the variance-mean relationship by switching from an additive to a multiplicative model. But there is a problem with negative values, for which the logarithm is not defined. This problem is solved with a small shift of the full sample

by adding a small difference of the smallest (negative) observation to 0 to every observation. This shift is dependent on the sample, that is, with a new sample a different adjustment would be made, and it affects the scale of the full sample. The estimates of the expected emission factors are obtained from the original scale of the EF_3 values without the logarithmic transformation.
One possible alternative would be the assumption of a detection limit at a small positive value of EF_3. Either any value below the threshold can be set to the detection limit, then assuming a truncated log-normal distribution, or the values below the threshold can be treated as censored observations, estimating emission factors conditional on the proportion of observations below the detection limit. Any of these alternatives will be more difficult to implement and it would be questionable if the estimation will be much improved compared to the logarithmic transformation of the shifted sample.

Instead of applying the logarithmic transformation to the EF_3, the cumulative losses can be transformed after adjusting for negative estimates. An advantage would be the reasoning that negative cumulative losses are unrealistic, motivating the rescaling to only positive values by an additive shift. Adding the same shift to the control and treatment observations of cumulative losses, might not affect the EF_3 calculation in a major way. But a logarithmic transformation of the cumulative losses will change the nominator in the definition of EF_3

$$EF_3 = \frac{\text{Treatment } N_2O - \text{Control } N_2O}{\text{N load}} \times 100$$

to

$$\log\left(\text{Treatment } N_2O\right) - \log\left(\text{Control } N_2O\right) = \log\left(\frac{\text{Treatment } N_2O}{\text{Control } N_2O}\right),$$

which can be interpreted as the logarithm of a relative change compared to a control. Hence, any transformation of individual components of the emission factor will change its interpretation.

Distributional assumptions

After the logarithmic transformation, the EF_3 response is assumed to follow a normal distribution. This log-normal assumption seems to be a reasonable approximation that captures the increasing variance with an increasing mean.

I do not see a reason why additionally an assumption of a Poisson distribution is used as an alternative. Technically, the Poisson likelihood can be evaluated at non-integer values larger or equal to zero; but theoretically, it is based on a discrete distribution function for modelling the number of events in a certain time interval. As a main difference to the log-normal model, the Poisson distribution assumes that the mean is equal to the variance, which is quite a strict assumption, especially for the EF_3 measurements.

Model parameters

The main predictors are dummy-coded variables for estimating the effects of different animals, steepness, dung/urine, or their interaction. But the model also needs to capture the sampling structure, which is reflected by a single random trial effect, separating the between trial variability from the within-trial variability.

There are additional sources of variability that could be captured by random effects. E.g. spatial effects can be modelled by the variance between regions, which would also address the fact that the number of trials are varying between regions. Within each trial, almost all studies are planned as a block design, combining a control with a specific set of treatment chambers. Ignoring the blocking structure would lead to the assumption that each EF_3 observation is independent, but in reality several EF_3 measurements are based on the same control observation.

There is some controversy about determining model complexity in hierarchical models and choosing an appropriate number of random effects. Barr et al. 2013 suggest to use the maximal number of random effects making sure that the uncertainty about parameter estimates is not underestimated. As a response, Bates et al. 2015 showed that some complex random effect structures might not be supported by the information in the sample and can lead to convergence problems. Hence, with only small samples within a single trial, a more complex mixed-effect model that takes the blocking structure into account, might not be supported by the data.

Model selection

One of the main aspects of the study is the aim to reduce the number of emission factors by combining animal and steepness categories.

In order to find an optimal number of emission factor categories, a model selection procedure is used, finding a model with a compromise between prediction accuracy and complexity.

Separating the dung and urine datasets based on visual inspection is a reasonable first step. Any attempt in finding a combined model might lead to an unnecessary complex modelling procedure.

Based on a step-down procedure, a single sample-based path is followed through the complete set of possible sub-models, where two models are compared by testing the statistical significance of a term in the model with a p-value threshold of 0.05. Based on the rather arbitrary threshold of a type-I-error rate of 0.05, the selected model might not be able to select an optimal model with respect to the prediction accuracy; but, the selection procedure will result in a model that includes terms that can explain a certain amount of variability in the response.

Only the effects of terms remaining in the selected model can be interpreted as being on importance; nothing can be said about the terms that have been removed from the model. Further, any inference and probability statements will be conditional on the selected final model; it follows that hypothesis tests for statistical significance and comparisons between treatments do not take the model selection uncertainty into account, which will result in an underestimation of standard errors and p-values for confirmatory testing.

Multiple comparison procedures (Tukey-type tests) are used to find differences between treatment means of EF_3 values. But this procedure is not suitable to decide about combining emission factor categories, as two categories are combined when no statistical significant difference is found. A higher number of comparisons would result in a larger penalty for multiple testing, which will consequently lead to a larger p-value for a single comparison; hence, with a large enough number of treatments no statistical significant effect will be found, collapsing everything into a single emission factor category.

Reporting emission factors

The reported emission factors are calculated as arithmetic means for certain livestock class/excreta type/slope class categories. Averaging the EF_3 on the original scale assumes a different model compared to averaging on the logarithmically transformed scale, which would lead to geometric means instead.
Hence, there is a disagreement between the model, which is used to decide about collapsing categories, versus the model, that is used to obtain estimates for the expected EF_3. It can be assumed that the arithmetic means are generally higher compared to the geometric means, being more influenced by the longer tails of an assumed log-normal distribution.
Reporting arithmetic means has the advantage of comparability with all previous literature and guidelines, which are also based on arithmetic means.

A problem with the use of arithmetic means is reporting adequate measures of uncertainty about the estimates. When reporting 95% confidence intervals for the arithmetic mean of EF_3, a nonparametric bootstrap method is used to estimate the standard deviation of a bootstrap distribution of the mean. This methodology only includes information about the sampling structure by conditioning on the observed sample; therefore, any correlation between measurements and unbalanced sample allocation in trials or locations is not taken into account. This probably will lead to the underestimation of the uncertainty about estimating the arithmetic means.

Comparison of Method 1 vs Method 2

Viewing the comparison between

- Method 1: Individual EF_3 means for each of the full factorial combinations of livestock class/excreta type/slope class categories.

- Method 2: Pooling some categories based on model selection methodology

as a model selection problem, then, under the assumption that the full model is a good representation of the true data-generating process, the decision is mainly about a bias-variance trade-off for the EF_3 estimates. Using the most complex model (Method 1) will result in less biased estimates with larger uncertainty, whereas Method 2 will increase the prediction accuracy by introducing some bias to the estimation process.

From an applied statistical perspective, we can assume that the data-generating process is much more complex than the full model with arithmetic means for the 16 combinations of categories, with a highly unbalanced amount of information from the sample. Without enough information from the sample a more simplified model might miss important aspects of a signal to predict EF_3 levels. Then the most complex model (Method 1) should ensure that there will be a reasonable prediction for an emission total based on the available sample data with the cost of a higher variability.

But the reduction of the number of EF_3 values, collapsing the medium and steep slope categories for urine, and only using a single EF_3 estimate for dung has the practical advantage of easier calculation and reporting. The predicted EF_3 is assumed to have a higher precision, where the additional bias might be negligible compared to the uncertainty caused by experimental errors in measurements and calculating the cumulative losses.
The benefit of simplicity for reporting EF_3 can easily outweigh some additional bias of pooling the EF_3 categories.

References

Barr, D. J., Levy, R., Scheepers, C. and Tily, H. J. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of Memory and Language, 68, 255-278.

Bates, D., Kliegl, R., Vasishth, S. 2015. Parsimonious mixed models. arXiv, 1506.04967.

Kelliher, F.M., Cox, N., Van Der Weerden, T.J., De Klein, C.A.M., Luo, J., Cameron, K.C., Di, H.J., Giltrap, D., Rys, G. 2014. Statistical analysis of nitrous oxide emission factors from pastoral agriculture field trials conducted in New Zealand. Environmental Pollution 186, 63-66.

Saggar, S., Giltrap, D.L., Davison, R., Gibson, R., DeKlein, C., Rollo, M., Ettema, P., Rys, G. 2015. Estimating direct N2O emissions from sheep, beef, and deer grazed pastures in New Zealand hill country: accounting for the effect of land slope on the N2O emission factors from urine and dung. Agriculture Ecosystems & Environment 205, 70–78.