Nitrogen flows on organic and conventional dairy farms

Supporting Information to an article in

Nutrient Cycling in Agroecosystems

Rasmus Einarsson, Christel Cederberg, and Jonatan Kallus

March 2017

1 Calculation of confidence intervals

If X and Y are normally distributed, and Y is unlikely to be zero, then the so-called Geary-Hinkley transformation can be used to show that the quotient $Z = X/Y$ is also roughly normally distributed (Hayya et al. 1975). Approximate parameter values of the distribution of Z of the approximate normal distribution of Z are given by

$$\mu_z \approx \mu_x/\mu_y + \sigma_y^2 \mu_x/\mu_y^3 - \rho\sigma_x\sigma_y/\mu_y^2,$$
$$\sigma_z^2 \approx \sigma_y^2 \mu_x^2/\mu_y^4 + \sigma_x^2/\mu_y^2 - 2\rho\sigma_x\sigma_y\mu_x/\mu_y^3,$$

where μ are the means, σ^2 the variances, and ρ is the correlation coefficient between X and Y (Hayya et al. 1975).

As in the main text, we consider the example of chain N surplus indicator. On average, the chain N surplus is $S_c = S_f + kF$, where S_f is the farm-gate surplus, F is the amount of purchased feed N, and k is a constant representing the average amount of off-farm N surplus per unit of purchased feed. Given a group of n sampled farms, we estimate the indicator as

$$\widehat{\mathrm{CS}}_{\mathrm{M}} = \frac{\widehat{S}_c}{\widehat{P}} = \frac{\sum_{i=1}^{n}\left(S_{f,i} + kF_i\right)}{\sum_{i=1}^{n} P_i}.$$

where $S_{f,i}$, F_i and P_i are the individual farms' measured farm-gate surpluses, feed N purchases, and product N sales. When n is sufficiently large, both $\widehat{S}_c$ and $\widehat{P}$ are approximately normally distributed, so that $\widehat{\mathrm{CS}}_{\mathrm{M}}$ is approximately distributed as a ratio of two normal distributions. Thus, the sample mean, sample variances and sample correlation coefficients of $X = (S_f + kF)$ and $Y = P$ can be used to estimate the normal distribution approximating the sampling distribution $\widehat{\mathrm{CS}}_{\mathrm{M}}$, and thus to calculate confidence intervals. A similar procedure works for all three indicators.

To check that the calculations were correct, we also estimated confidence intervals using bootstrap resampling. We resampled the organic and conventional sets of farms 10,000 times and thus obtained 10,000 average organic and conventional farms, for which the three indicator values could be calculated. The resulting distributions of indicator values was clearly normally distributed, and the 95% confidence intervals agreed very well with the approximated parametric confidence intervals.

2 Bayesian uncertainty analysis

2.1 Motivation and method description

The calculation of confidence intervals described in the paper assumes that the indicator estimates are correct in expectation, or in other words, that the estimates converge to the true values as the sample size grows. However, this is not necessarily true since some parameters in the calculation are uncertain. For example, if the estimates of BNF are biased, so are the estimated indicator values. This type of parameter uncertainty is the focus of the this uncertainty analysis.

We used a Monte Carlo sampling method to assess the importance of uncertainties in some of the parameters of the calculation, comprised of four steps: (1) estimating a joint distribution of the uncertain parameters, (2) drawing a number of random parameter sets (we drew 10,000) from the distribution, (3) calculating all results for each parameter set, and (4) analyzing the distribution of the results.

We find it natural to interpret this method in a Bayesian sense. The parameter values are unknown, but we can assign a joint probability distribution to them that represents our best knowledge. The corresponding results, being a deduction from the parameters, should then be interpreted in the same way. Hence, we interpret the distribution of results in a Bayesian sense, and it is then appropriate to calculate credible intervals for them.

The following sources of uncertainty were considered:

- Even if sampled farms are in principle representative of conventional and organic dairy farms, the sample mean is only an estimate of the population mean. To quantify the uncertainty about the population mean, we used Bayesian bootstrap resampling (Rubin 1981) to obtain a distribution of average farms. This is the only type of uncertainty that is reflected in the confidence interval in the article.

- The N budgets for the off-farm crop cultivation were based on statistical data. For each crop yield and average N application rate, we assumed a normal distribution with coefficient of variation equal to the reported sample coefficient of variation (Statistics Sweden 2014). For soybeans, where comparable statistical data was not available, we assumed a coefficient of variation of 10% for the yield.

- As mentioned in the paper, the contributions of BNF to N budgets are known to be rather uncertain. We have no reason to believe that the estimates are biased one way or another, but it seems plausible that future research will change our best estimates. Therefore, we also added a normally distributed 10% relative error to the BNF estimates. In other words, we multiplied the estimates by a normally distributed error factor with mean 1 and standard deviation 0.1.

We assumed that all these uncertainties were independent of each other. Thus, a sample from the joint parameter distribution was simply drawn as one Bayesian bootstrap sample of average organic and conventional farms, together with independent samples of all other parameters. Similarly as in the frequentist approach, we calculated 95% credible intervals for the indicator values and their differences between the systems to test whether they were significantly different.

2.2 Results

Not surprisingly, the 95% credible intervals obtained using this approach are substantially wider than the 95% confidence intervals in the paper. Figure 1 shows a summary of the results analogous to one of the figures in the main text.

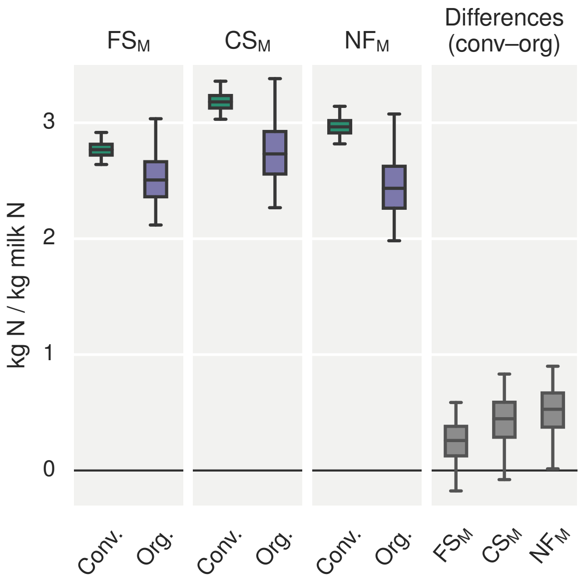

Figure 1 – Comparison of farm-gate N surplus (FS_M), chain N surplus (CS_M), and N footprint (NF_M) between average conventional and organic milk. The boxes and whiskers give information about the uncertainties, based on 10,000 random parameter sets. Boxes show the interquartile range (i.e. the central 50% of the distribution) and whiskers show a 95% credible interval.

Figure 2 gives more information about the relationship between conventional and organic N footprints. Similar illustrations for the N surplus indicators would look similar, so the following paragraphs to some extent applies to all three indicators. All three panels of Figure 2 show the distributions of N footprint values obtained in the Bayesian analysis. In the middle and bottom panel, each point corresponds to one sample from the parameter distribution.

One crucial thing to note in Figure 2 is that, although the unconditional distributions (top panel) overlap substantially, the organic N footprint is higher than the conventional for only about 2% of the sampled parameter sets (the points right of the dashed line in the middle panel). This is partially explained by the correlation between the conventional and organic N footprints (see middle panel). This correlation means that at least one parameter affects both values in the same direction. The only parameter that can cause correlation between organic and conventional N footprints is the bias term we added to BNF estimates, since it is the only random parameter that is common to both calculations. In the bottom panel of Figure 2, the N footprints and their differences are plotted against the sampled values of the relative BNF bias parameter. This possible bias is the largest contributor to uncertainty in the indicator values: If the BNF has been overestimated by 20%, the conventional N footprint may be almost 45% higher than the organic, but if the BNF has been underestimated by 20%, the conventional and organic N footprints are likely to be equal.

The uncertainties in crop yields and N application rates in off-farm feed production are not an important contributor to uncertainty in the indicator estimates. Although the crop yields and N application rates may vary substantially in some cases, it is very likely that a negative variation in one crop is cancelled by a positive variation in another crop; and thus it is unlikely that the joint variation in yields and N application rates gives rise to any large change in the indicator values.

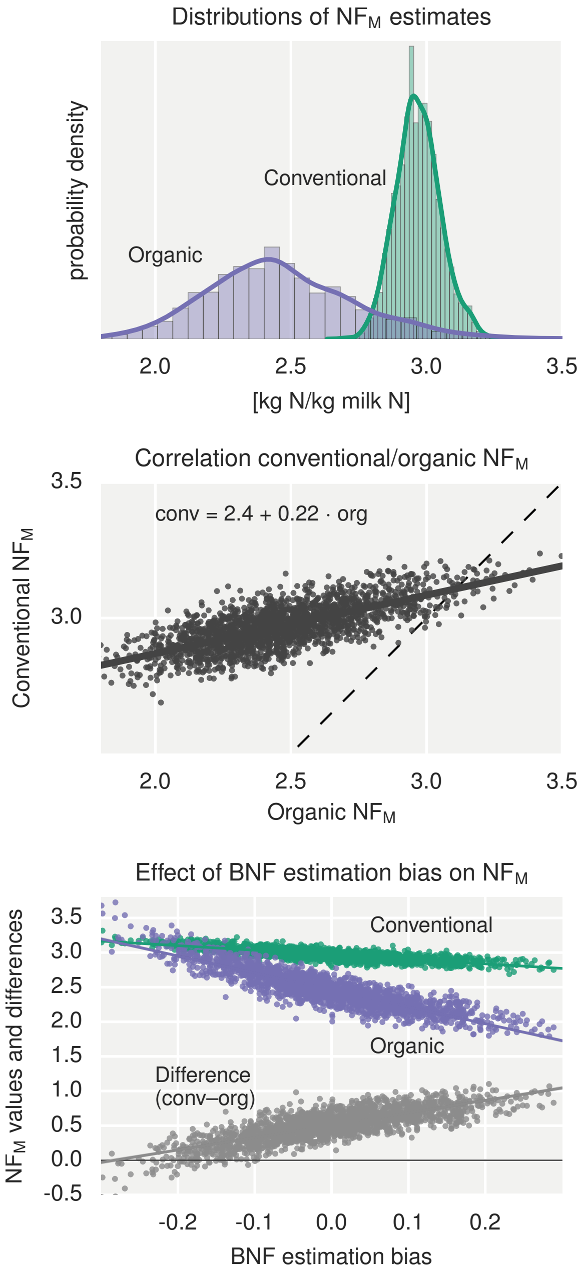

Figure 2 – Monte Carlo uncertainty analysis of the milk N footprint (NF_M), based on 10,000 samples from the parameter distribution. Each point in the middle and bottom panels correspond to one parameter set. Top panel: Distributions of footprint values. Middle panel: Joint distribution of conventional and organic N footprints. The dashed line is the 1:1 line: To the left of it, the conventional footprint is larger than the organic, and vice versa. Bottom panel: Regressions of the N footprints and their differences against sampled values of BNF estimation bias parameter.

References

Hayya, Jack, Donald Armstrong, and Nicolas Gressis (1975) A Note on the Ratio of Two Normally Distributed Variables. *Management Science* **21** (11), 1338–1341.
Rubin, Donald B. (1981) The Bayesian Bootstrap. *The Annals of Statistics* **9** (1), 130–134.
Statistics Sweden (2014) *PM Lantbruksstatistik 2013:2*. Statistics Sweden, p. 85.