Review of “Toetsing van de KringloopWijzer”

By Rogier Schulte, Wageningen Economic Research

Preamble

The report “Toetsing van de KringloopWijzer” evaluates the potential use of the KringloopWijzer as a policy instrument for the calculation of nitrogen and phosphorus excretion and production on dairy farms.

The report and is clearly structured and written, the objectives are clear and the methodology is well laid out (subject to some recommendations below).

Overall assessment

- The Kringloopwijzer (KLW) has been tested against what appears to be an impressive empirical database (“meetweken”).
- Section 2 would benefit from two flow diagrams that show: 1) data inputs; 2) estimated parameters and 3) the order of calculations used for A) the “observations” and B) the “predictions”, respectively.
- I support the overall conclusion by the authors that the KLW is capable of estimating the variation in manure and crop production across a range of dairy farms, *subject to the observations and reservations below*.
- Specifically, the authors have convincingly shown that – in its totality – the KLW provides a better estimate of N and P excretion than the statutory (“forfortaire”) estimations.
- I support the recommendations for further research, with the exception that sensitivity analyses that do *not* require further empirical data should possibly have been part of this current study.

Answers to specific questions

Is de KringloopWijzer, als instrument voor de berekening van de stikstof- en fosfaatproductie in mest en de ruwvoerproductie op een melkveebedrijf, bruikbaar en betrouwbaar voor beleidsdoeleinden?

Kan de KringLoopWijzer worden gebruikt ter vervanging van de diergebonden normen (RVO-Tabellen 4 en 6) op een melkveebedrijf?

Where appropriate, and within the range of farming systems assessed in this report, the KLW may be used as a tool to compute N and P production in manure and crops, for use in policy purposes. The report shows that – in its totality – the KLW provides a better estimate of N and P excretion than the statutory (“forfortaire”) estimations. This superior performance must be evaluated in the context of:

- The additional work associated with data input by the user;
- The risk of users “picking and choosing” the method that is most favourable to farm management, which may not necessarily align with the methods most favourable to nutrient efficiency and environmental integrity;
- The risk of using an advisory tool as a policy tool, which may blur the line between advice and enforcement, with potential long-term consequences for the relationship between farmers, advisors and researchers.

There may be merit in discussing such wider considerations in the report.

Voor welke typen melkveebedrijven is de KringLoopWijzer geschikt; ook voor melkveebedrijven met een neventak, bedrijven die vee uitscharen, bedrijven die zelf melk verwerken/verkopen en bedrijven die meer dan één vestiging hebben?

I would be concerned about extending the use of KLW as a policy tool outside the range of farming systems assessed in this report, especially given my reservations about some of the methodological aspects; see below.

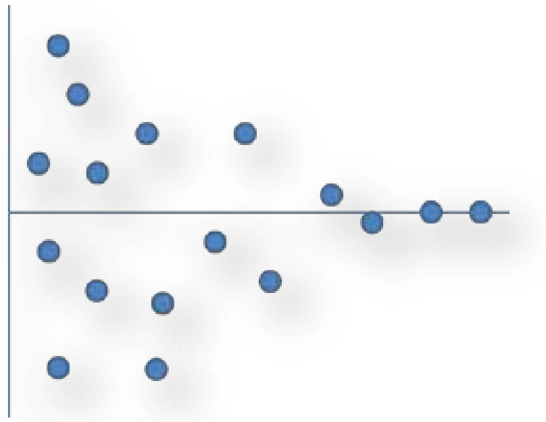
Is de toets zoals beschreven in het rapport betrouwbaar uitgevoerd; is de werkwijze van de meetweken een voldoende betrouwbare toets?

No, the report does not use the correct procedures for model evaluation; this is the area where I have significant concerns:

- Tables 3.2 to 3.8: differences should be tested for significance. If they are not significant, then technically there is no difference between observations and predictions.

- Figure 3.1 – 3.8: The authors use correlations to assess the alignment between observations and predications. However, correlations say very little about this alignment. Observations and predictions can be perfectly correlated, and yet be significantly different from each other. The appropriate way to assess differences between observations and predictions is to test for both absolute differences (is the intercept significantly different from 0?) and relative differences (is the slope significantly different from 1?).

- Figures 3.9 – 4.3: Here, incorrect indicators and assessments are used.
    - "Afwijking" should be given in absolute, rather than relative values. Use of relative values results in dampening of overestimations are dampened and amplification of underestimations.
    - In additionally, the test should be performed on |error| (i.e. the absolute values – turning negative into positives), or an alternative test should be used to assess the impact of the X-variable on the error of the prediction. In its current format, the X-variable may impact on the error, even when the correlation is computed as zero. Consider e.g. the following, hypothetical scenario:



      In this example, the error (plotted on the Y-axis) is reduced at higher values of X. However, this is not captured by the correlation between X and Y, which will be close to 0.
    - Single correlations are used to elucidate which X-variables impact on the error in the predictions. However, we may expect a considerable degree of correlation between the X-variables themselves – which gives rise that observed correlations may in fact be pseudo-correlations. An alternative assessment (e.g. backward or stepwise multiple regression? Monte Carlo analysis?) should be considered that assesses the impact of all X-variables simultaneously.

  - The data in these figures (particularly on X-axis) does not appear to be normally distributed, which gives rise to outliers that may have excessive leverage in the correlations --> consider log transforming?
- Section 4.2 focusses on the “average error” between observations and predictions. Of more relevance are the outliers of the absolute errors – how large are these, and what is the cause of their discrepancy?

Welke mogelijke verbeteringen van de KringLoopWijzer kunnen leiden tot een hogere betrouwbaarheid?

In general, I support the authors’ recommendations for further research (subject to comments above). If use of the KLW is to be extended to include farms characterised by grazing (“weidegang”), then there may be merit in direct prediction of the N and P uptake by grass, and N and P uptake by the animal. In the current KLW, this is the “sluitpost” (Section 2.2); therefore, predictions on N and P in grass currently include all accumulated error terms of the preceding predictions. The expertise for modelling N and P uptake in grass is available in the Netherlands.