**Geoscientific
Model Development
Discussions**

Interactive comment on "Using multi-model averaging to improve the reliability of catchment scale nitrogen predictions" *by* J.-F. Exbrayat et al.

F. Anctil (Referee)

francois.anctil@gci.ulaval.ca

This paper, which promotes multi-model averaging, provides an interesting perspective of the pros and cons of simple ensemble averaging (SEA) and of reliability ensemble averaging (REA) for catchment scale nitrogen predictions based on four models of various complexities. The paper is well written, easy to read, and informative. Suggestions for improvement follow.

Principal comments

The pool of proposed models is rather limited for an ensemble study. I understand that some catchment scale nitrogen prediction systems are quite time consuming to implement and that the number of plausible options are limited, but I do not think that

C778

we can produce generalizable conclusions from four model and a series of scenarios reducing the fertiliser application on a single watershed. Still, the authors advocate, from such a set-up, that REA is superior to SEA. But is it true for all possible models and watersheds? Authors need to be more modest in their discussion and conclusion.

I do not understand why the authors limited their ensemble experiment to the fertiliser application scenarios. They could also have used some of the metrics in Table 2 to compare REA and SEA. For instance, I would have liked to know if REA and/or SEA, when applied to the four models at hand, produce better RMSE values than the LAS-CAM model.

Page 2292, line 13: we can read "...in spite of the demonstrated improvement in prediction reliability." Multimodel averaging is not magic; it does not always work. The authors need to nuance their statement.

A somewhat more nuanced statement is found page 2296, line 22: "Previous studies on multi-model averaging techniques set in a variety of environmental modelling contexts have demonstrated that the simple mean of a MME usually outperforms its members taken separately in terms of goodness-of-fit metrics." The problem is that when people find a situation for which ensemble averaging does not work; they tend to not publishing their findings.

Other comment

Page 2293, line 7: "Because of the sandy nature of the soils, evaporation is high (2000 mm yr-1)". The authors are certainly discussing of the potential evapotranspiration. Evaporation cannot surpass precipitation.