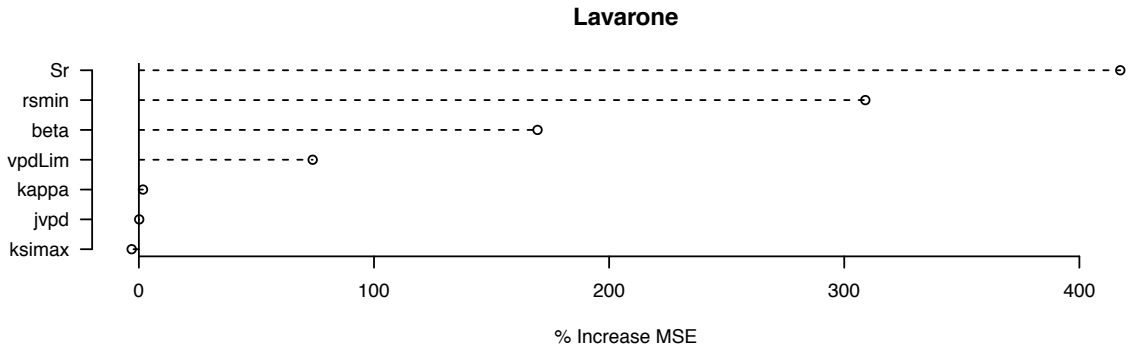
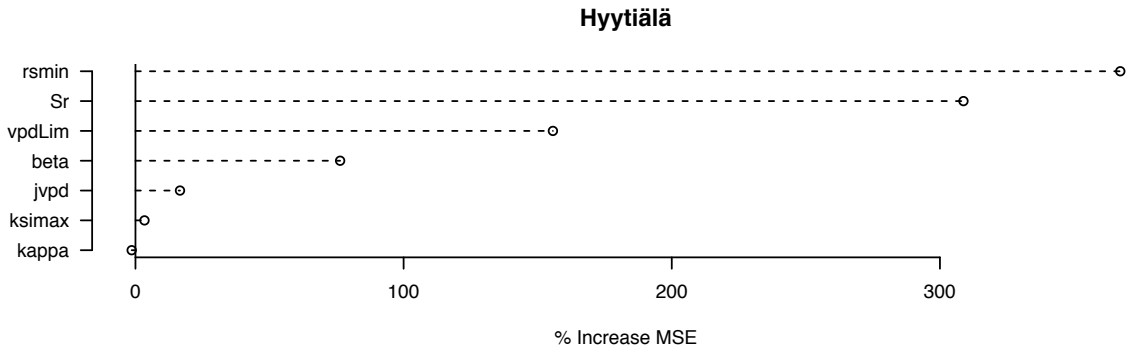
We would like to thank reviewer #2 for their helpful and constructive comments. We take this opportunity to discuss the main points raised in this review and propose ways to address these concerns in a revised version.

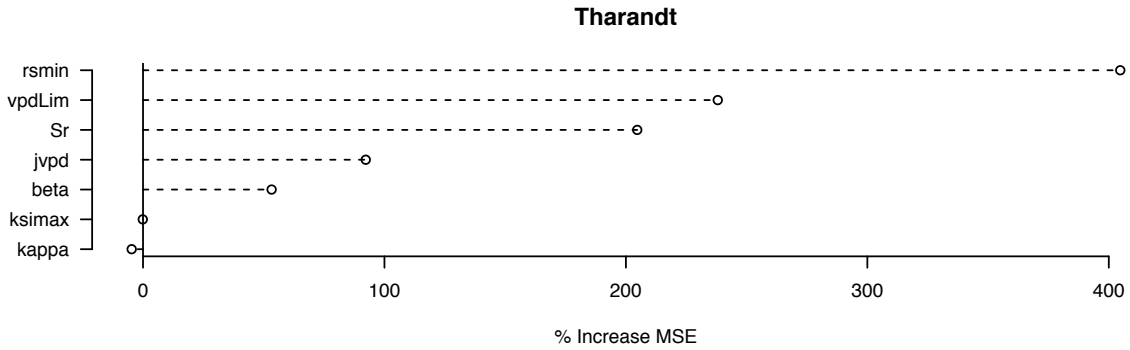
Some of the points have been raised by both reviewers 1 and 2. We address the issues regarding the article structure, thematic focus and research goals/questions mainly in our response to review #1, and concerns regarding sensitivity and uncertainty mainly in our response to review #2.

Uncertainty of the FORHYTM parameters and the role of Sr

- Reviewer #2 notes that the sensitivity of FORHYTM to Sr and other parameters is not assessed. We agree that this makes it difficult to interpret the calibrated values and to decide how reliable the calibration procedure is to estimate Sr. Another manuscript of ours (referred to as "in prep.") has been accepted for publication in Environmental Modelling and Software and will be published in volume 102. This article reports on a sensitivity analysis of FORHYTM, assessing the effects of changes in parameters (including soil and canopy properties, as well as the Jarvis parameters) on model outputs (long-term total evaporation, as well as a drought index based on the ratio of actual to potential transpiration). This analysis was conducted at ten sites with contrasting hydro-climatic conditions, including dry inneralpine, temperate lowland and cold subalpine sites. In this study, Sr was among the most influential parameters at all sites, for both output variables.
- To assess whether this also applies to the KGE scores in the present study, we conducted a sensitivity analysis on the calibration runs. At each of the eddy covariance sites, a Random Forest model was fitted to the outputs of the calibration runs, with the seven parameters as predictors, and the KGEavg score as the dependent variable. The Random Forest algorithm provides a measure of variable importance, which allows a ranking of the predictors by their influence on the output. The importance score is based on the increase in model prediction error when the values of a predictor are permutated (see e.g. Liaw and Wiener (2002)) Of the 7 parameters, Sr was the most influential variable at 5 sites, the second most influential at 6 sites and 3rd most influential at 4 sites. While the absolute values of the importance scores vary over several iterations of the RF algorithm, the ranking of the parameters remains stable. Other variables of importance are RSmin and l_vpd. The following figure shows the variable importance scores at three stations, with Sr in the first, second and third position:







Sensitivity and uncertainty of the G10 model

- We also agree that it is important to assess the sensitivity of the G10 model to its inputs (climate statistics, physiological parameters, LAI, soil WHC). We have started to assess the effect of variations in physiological parameters in Sections 2.1.4 and 3.3. However, we now feel that a more formal sensitivity and uncertainty analysis would be much more informative.
- We have generated 2000 estimates of Sr at each station, with perturbations of all parameters by up to 20%. The parameters include (1) plant physiological parameters for trees and grass, (2) climate statistics, (3) site characteristics such as LAI and soil WHC. In addition, the start and end of the growing season were also shifted back or forward by up to 10 days (which, in turn, also affects the climate statistics calculated over the growing season). Sampling was again done with the Latin Hypercube method.
- The resulting standard deviations of Sr ranged between 18 and 59 mm across sites. The spread, however, is much larger. Distributions of Sr tend to have long

tails, with a few extreme values. It will be useful to determine which parameter combinations lead to these extremes.

- We also applied a Random Forest model to determine variable importance, in a similar manner as for the FORHYTM calibration runs. Preliminary results indicate for example that the parameter values for grass have little importance at all sites. The sensitivity rankings further suggest geographical differences, with e.g. LAI being more sensitive at the Mediterranean sites than elsewhere. While these preliminary results are promising, it might be necessary to define more specific uncertainty bounds than 20% for each parameter. This will be based on a brief discussion of the various sources of uncertainty associated with each parameter.

Uncertainty of the eddy covariance and soil moisture data

- Although we briefly mention some of the issues of eddy covariance and soil moisture data, we agree that it is worthwhile to discuss these in more detail, and to put the FORHYTM results and calibrated values in this context. We plan to include a paragraph or subsection in the discussion where we briefly summarize the recent literature published by the EC community on these issues (regarding e.g. random and systematic measurement errors, as well as spatial heterogeneity), and relate them to the analysis in this article.

- The review points out several statements where it is not clear to what extent they are supported by the analyses. We will carefully review each of the statements and either clarify the link to the analysis or reformulate them to avoid making unsubstantiated claims.

Article structure and readability

- We also appreciate the feedback on the structure and readability of the article. In our response to reviewer #1, we have outlined a possible reformulation of the research goals and questions. We also propose to rearrange the structure, so that only the parts that directly relate to the research questions are kept in the main text, and accessory parts are shifted to the appendix. For example, the validation of FORHYTM is useful to estimate the reliability of model results and of the calibrated values (in addition to the sensitivity analysis outlined above), but does not directly relate to the research questions and can be moved to the appendix. Also, we suggest to move the section on numerical approximation to the Appendix. For the description of FORHYTM, we can now refer to the Environ. Modell. Softw. article. We believe that this will further clarify the goals of the article and enhance readability.
- We will also carefully consider the points on cross-referencing, and presentation of the results. We will add color to the figures where appropriate.

Liaw, A. and M. Wiener (2002): *Classification and Regression by randomForest*. R News 2(3), 18-22.