

Interactive comment on “Filling the white space on maps of European runoff trends: estimates from a multi-model ensemble” by K. Stahl et al.

K. Stahl et al.

kerstin.stahl@hydrology.uni-freiburg.de

Received and published: 18 April 2012

We appreciate the encouraging review. The question of identifying individual models has been an ongoing debate within the WATCH project community, and the topic of several other papers that are based on a similar set of model simulations. The conclusion from these studies is that it is indeed difficult to explain why some models performed better or worse for the different performance metrics by trying to relate the results to various properties of the model structure or process representation/processes (Haddeland et al., 2011, Gudmundsson et al., 2012; Gudmundsson et al., in revision; Prudhomme et al., 2011, and various Technical Reports on www.eu-watch.org). In light of this experience we chose not to identify individual models for the following reasons. First, discussing individual models would indeed make a very long paper, which,

C969

likely will not conclude differently than the aforementioned studies. Second, this would take away attention from the primary focus on the trends and on the general effect that the same meteorological input can produce rather different hydrological trends in modelled runoff from an ensemble of models. Third, choosing a subsample of models to make a better ensemble is against current interpretation that ensemble means work because of random model errors caused by the variety of model structures, none of which is perfect and each having its own specific strengths or weaknesses. We feel that the decision not to name the individual models is justified in this case, but suggest to add more information on other studies addressing this topic and to improve the arguments for not identifying the individual models.

Response to Specific comments:

Within WATCH, the modellers had some constraints to adjust to a common protocol, but still had some freedom to apply the model in their custom way, which included calibration for WaterGAP as stated in Haddeland et al. 2011. The common protocol was defined as part of the (broader) WaterMIP and is described in Haddeland et al. 2011. WaterGAP does rank overall high in all applications, though not necessarily best. We do not think, however, that this difference in modelling setup affects the results of our study significantly, as the WaterGAP model is calibrated against discharge from large basins whereas in our study, runoff is compared at the grid cell scale.

We thank the referee for pointing out a number of details that concern readability, preciseness, etc. We agree that all of these require our attention and will address them in the revision.

References:

Gudmundsson, L. et al. (2012) Comparing Large-Scale Hydrological Model Simulations to Observed Runoff Percentiles in Europe, *J Hydrometeor.* 13(2): 604-620. DOI: 10.1175/JHM-D-11-083.1.

C970

Gudmundsson L., Wagener T., Tallaksen L. M., Engeland K. Evaluation of nine Large-Scale Hydrological Models With Respect to Seasonal Runoff Climatology in Europe. *Wat. Res. Res.* (In revision)

Haddeland, I. et al. (2011) Multimodel estimate of the global terrestrial water balance: Setup and first results. *J. Hydrometeor.*, 12(5): 869–884.

Prudhomme, C. et al. (2011) How well do large-scale models reproduce regional hydrological extremes in Europe? *J. Hydrometeor.*, 12(6): 1181–1204.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 9, 2005, 2012.