Hydrol. Earth Syst. Sci. Discuss., 8, C1069–C1071, 2011
www.hydrol-earth-syst-sci-discuss.net/8/C1069/2011/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Köppen versus the computer: an objective comparison between the Köppen-Geiger climate classification and a multivariate regression tree" *by* A. J. Cannon

B. Orlowsky (Referee)

boris.orlowsky@env.ethz.ch

Received and published: 20 April 2011

The author introduces a partitioning algorithm (the Multivariate Regression Tree, MRT) to come up with a classification of the climates of the world. He compares them to the traditional classification based on the Köppen-Geiger (KG) rules. When evaluating the two classifications with respect to the monthly climatologies of precipitation and temperature, the MRT classification shows a better reproduction of these climatologies than the KG classification, in particular for precipitation.

This is an interesting and well written paper, which I would certainly recommend for

publication after addressing some minor comments.

1. My main comment concerns the choice of the predictands for the MRT and the subsequent comparison between MRT and KG. The author argues that the main goal of such a classification is to obtain regions of high intra-annual climatological homogeneity, where the climatological characteristics are represented by the monthly climatologies of temperature and precipitation. The MRT, based on the same predictors as the KG classification, is consequently grown such that these monthly climatologies are optimally predicted. While this is certainly a reasonable target for such a classification, it is not the one at which KG aimed (their classification relied much on vegetation types, for example). A better reproduction of these predictands by the MRT (compared to KG) is therefore natural in my view. Could the discussion of the comparison be adapted in this sense? For example, I would not describe it as "objective" (as done in the abstract), which in my understanding would only apply to the comparison of two classification approaches aiming at the same target.
2. P2348L3: Not being an expert in this field: Is it true that the binary data clustering is the most common one?
3. P2348L9f: Could you give a reference for "association analysis"?
4. P2348L17ff: Is this auto-associative MRT relevant for your study? If so: Does the non-distinction between X and Y mean that $X = Y$ (was a bit unclear to me)? If not: Could it be removed from the text?
5. Section 3: From what I understand, MRT is like a multivariate extension of CARTs. If this understanding is correct, could it be emphasized in this section?
6. P2350L15: Does "exhaustive" mean "not greedy"? I am a bit confused, later on P2356L3 you talk of a greedy optimization.

7. P2351L25f: Such predictive aspects are also shared by CARTs.

8. P2353L25ff: How do you calculate the CIs for EV?

9. P2355L26ff: Many of the following sentences would in my view better fit in Section 3.

10. A very minor comment: Although I am not at all skeptical about the application of MRT in this context, I would personally express its prospects as an essential means of education for future students rather as a hope, not as something at eye level with KG (concerning the last sentences of the abstract and Section 6) ;-)

11. Fig. 6: Would it make sense to arrange either the MRT or the KG diagrams such that the pairs in each row correspond to each other in terms of the “winners” from Table 2?

12. Fig. 10: Why don’t you show the CIs here? I didn’t quite understand how this figure is computed – is it averages over the single month EVs?