Hydrol. Earth Syst. Sci. Discuss., 8, C1764–C1766, 2011
www.hydrol-earth-syst-sci-discuss.net/8/C1764/2011/

Hydrology and Earth System Sciences Discussions

Interactive comment on "DREAM$_{(D)}$: an adaptive markov chain monte carlo simulation algorithm to solve discrete, noncontinuous, posterior parameter estimation problems" *by* J. A. Vrugt

B. Schaefli (Referee)

bettina.schaefli@epfl.ch

Received and published: 25 May 2011

General comment

This manuscript presents an extension of the DREAM algorithm to solve integer and non-continuous parameter estimation problems via Markov Chain Monte Carlo simulation. As the author points out, in environmental modeling, relatively little attention has been given to parameter estimation problems involving discrete variables. In other fields of optimization, the estimation of discrete parameters is a classical problem; the present manuscript represents an advance over existing solutions because the pre-

sented algorithm not only identifies globally best solutions but also provides a sample from the posterior distribution of the parameters.

While the simple extension of DREAM is very well presented and evidence is provided that it works (also from a theoretical point of view), the paper does not sufficiently discuss why an MCMC sampler for integer and non-continuous problems is actually useful. Concerning non-continuous parameter estimation, it is shown for a simple example that the results are equivalent to a continuous formulation and it is suggested (not shown) that non-continuous sampling might enhance the algorithm efficiency, but no details are given. The case of integer parameter estimation is, in exchange, not discussed from a hydrological point of view. It is suggested that it might be useful for sampling design but without any further comments.

Integer/discrete problems vs discretized continuous problems

From my point of view, the manuscript should include at least a discussion of a real example of an integer or otherwise truly discrete parameter estimation problem. The paper mentions sampling design; I can only think of one straight forward problem: deciding where or when to take a fixed amount of samples. Any more complicated question (e.g. number, location and type of measurement) would have to deal with the problem of a varying parameter vector size. The same problem occurs in the case of model design optimization, the 2nd category of integer optimization problems which I can think of. Schaefli et al., (2004) presented an example where the number of different snow types or different model structure options are estimated in the form of integer decision variables (through global optimization). Depending on the value of the integer decision variables, the total number of parameters varies, which seems difficult to handle in an MCMC scheme.

I, therefore, encourage the author to extend this manuscript beyond the current form of a technical note on his algorithm to a paper that establishes a reference for integer and/or non-continuous parameter estimation in hydrology, discussing explicitly the two

types of applications (reduction of feasible parameter space and design questions).

Likelihood function

Furthermore, I have a technical concern: the theory of sampling posterior distributions with MCMC crucially depends on the fact that the used likelihood should be reflective of the distribution of the quantity on which it is estimated. In the classical rainfall-runoff model inference, the likelihood is defined in terms of the model residuals (observed - simulated discharge values), which are assumed to have a given distribution. How can this concept be transposed to the integer Soduku case? The comment for the Soduku example ("The log-likelihood function measures the constraint violation, details of which are outside the scope of this publication") is not satisfying with this respect. I guess that the idea is that for the integer, non-ordinal Soduku case, DREAM(D) can just be used as a global optimizer.

This general problem deserves some more details, starting with a distinction between the two rather different situations: a) problems where integer or discrete decision variables /parameters influence the otherwise continuous model outcomes on which a classical likelihood function is defined or b) problems where the model outcome itself is of integer nature and where the definition of the likelihood is a priori not straight forward. To my view, the current form of the paper tends to propagate the idea that you can choose just any objective function for the MCMC sampler (which is ok to obtain "a sample" of good solutions but this sample is not a sample from the posterior distribution).

References Schaefli, B., Hingray, B., and Musy, A.: Improved calibration of hydrological models: use of a multi-objective evolutionary algorithm for parameter and model structure uncertainty estimation, Hydrology: Science and Practice for the 21st Century, London, 2004, 362-371,

---

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 8, 4025, 2011.