Nonlin. Processes Geophys. Discuss., 1, C522–C526, 2014
www.nonlin-processes-geophys-discuss.net/1/C522/2014/

Nonlinear Processes in Geophysics
Discussions
Open Access

Interactive comment on "Bayesian optimization for tuning chaotic systems" *by* M. Abbas et al.

C. Snyder (Referee)

chriss@ucar.edu

Received and published: 25 September 2014

General comments:

This manuscript tests a Bayesian optimization (BO) algorithm in two simple problems. The BO algorithm is interesting and, in the geophysical literature I am familiar with, novel. Overall, the study is largely preliminary and the results are limited; even the simple test problems are only briefly analyzed.

Specific comments:

1. The writing should be revised for clarity. In addition, the descriptions of the experiments are incomplete and do not include sufficient information to reproduce the results. See detailed comments that follow.

2. The manuscript mainly presents very preliminary steps toward understanding how

Full Screen / Esc
Printer-friendly Version
Interactive Discussion
Discussion Paper

BO might be applied to actual prediction systems and whether it might yield significant or minimal improvements to those systems. Improvements could be made in several aspects:

a) There is little motivation for the experiments chosen and little explanation of what has been learned from the experiments.

b) The manuscript provides no sense of the magnitude of improvements achieved by tuning, since there is no analysis of how predictions are altered by the tuning. See section 4 of Wilks (2005), for examples of diagnostics that might be considered. Without such analysis, the manuscript shows only that BO finds a minimum.

c) There is no exploration (or even much discussion) of how the method might scale to larger problems (both with more parameters and with higher-dimensional states) and what difficulties might be encountered. An especially important issue is the calculation of the likelihood – see my comment 3.

d) There is no exploration of how the method compares with other possible approaches. For example, would the tuning accomplished here with BO fail with other methods, such as an ensemble Kalman filter with the state augmented by the parameters θ? As another example, how does BO compare with the approach of Wilks (2005) to tune representations of the fast degrees of freedom in the same Lorenz (1995) system?

3. A key issue is how to calculate or approximate the likelihood $p(y_{1:k}|\theta)$, especially when there are many observations, many parameters and the state dimension is large. Typically, we might expect that each individual observation carries little information about the parameters θ, and thus it is crucial to integrate the information in many observations (across space and time) to tune the parameters. Errors in calculating the likelihood may easily swamp that accumulated information Indeed, this may well be an issue even in the experiments presented in the manuscript; the authors use approximate methods but do not demonstrate their accuracy.

NPGD
1, C522–C526, 2014

4. In general, we would like $p(\theta|y_{1:K})$ rather than a single value of θ that maximizes $p(\theta|y_{1:K})$. Can the method presented here be extended to estimate a pdf for θ?

Detailed comments and typos:

section 1: Neelin et al. (2010 PNAS) deserve a citation both in the introductory paragraph and where response-surface techniques are discussed.

p 1284, l 25: forecast skills –> forecast skill

p 1286, l 15-16: awkward phrase

p 1288, l 9-10: Change "is computed using covariance function $k(\theta_i, \theta_j|\eta)$ for the corresponding inputs θ_i, θ_j " to "is computed using a covariance function $k(\theta_i, \theta_j|\eta)$ that depends on θ_i, θ_j". If that changes the meaning you intend, then I don't understand how you're specifying the covariance.

p 1289, l 6: i) Is this σ the same as that used in (2)? If so, why? ii) σ is used again, with a different meaning (I think), in (20).

p 1290, the EI acquisition function: I'm confused what EI is doing and how (8-9) describe that. Equation (8) (and the explanatory text) says that the acquisition function $g(\theta)$ is equal to the expectation of $f(\theta)$ minus μ^+, the current maximum value found for the mean of $f(\theta)$ at some θ_i. This seems to me to imply $g(\theta) = E(f(\theta)) - \mu^+ = \mu(\theta) - \mu^+$, which is not what (9) says. Please clarify.

p 1292, following (12): Usually, the EnKF is described as sampling correctly from the predictive distribution [i.e. the r.h.s. of (12)], but approximating the update (13).

p 1293, l 11-12: I suggest saying " ... C^{est}_{k-1} is the estimated covariance of $p(s_{k-1}|y_{1:k-1})$."

p 1293, l 23-25: I can't parse this sentence.

p 1294, l 4-5: "... which results in noiseless likelihood evaluations." The approximate

NPGD
1, C522–C526, 2014

likelihood (15) is a deterministic function of θ, but it is only an approximation and therefore contains error. Shouldn't the likelihood be consider "noisy" here too, in the sense that we shouldn't require the GP estimate of $f(\theta)$ to match the likelihood exactly? Is your assumption that (15) has errors that are too small to matter?

p 1294, l 7-8: "fluid motion dynamics on a rotating cylinder" This is a confusing way to think of the QG model (and it leads the authors to compute distances in a odd way later, p 1296, l 12-13). The periodic-channel geometry of the model is a computational artifice and not a physically realizable property – note that the planetary rotation is normal to the horizontal plane of the model. Calling this a "periodic channel" or a "zonally periodic channel" will make more sense to readers familiar with idealized QG simulations and will not lead astray those who are not.

p 1295, l 1: physic –> physical

section 4.1-2: What are boundary conditions at the channel walls? Does the model include any dissipation?

section 4.2: Judging by the definitions of the Froude numbers F_1 and F_2 and the Rossby number R_s, it seems that the parameters D_1, D_2 should be dimensional. My guess is the "units" are in fact meters, so that the layer depths are 6 km and 4 km, and the grid spacing is 100 km.

section 4.2: Please specify $S(x, y)$.

p 1296, l 17: if i and j are in the same layer –> if i and j are in different layers

p 1296, l 17: It is not obvious what the distance between layers should be (i.e. numerical value is assigned to h_{ij} when points are in different layers).

p 1296, l 19: The "nugget" term and its motivation are mysterious.

p 1297, l 4: It seems that a covariance function based on the 2D (Euclidean, allowing for periodicity in the zonal direction) distance would be fine. The model definitely does

NPGD
1, C522–C526, 2014

not represent fluid on the surface of a cylinder.

section 4.3: Plese give other BO parameters, in addition to $\zeta = 0$.

p 1298, l 5: . . . iterations to 200. While –> . . . iterations to 200, though

p 1298, l 13: bad samples –> good samples (I think??)

section 5: An obvious and very helpful step would be to explain how these experiments relate to those of Wilks (2005) with the same model.

p 1298, l 19-20: I don't see that a "noisy" likelihood (i.e. one that is stochastic function of the observations and θ) is a crucial distinction Even if you used a deterministic EnKF, the likelihood would only be approximate, even though it was a deterministic function of the inputs ($y_{1:k}$, θ). This error is equally important and is present in your first example (the QG model) as well.

p 1300, l 12: Eq. (23) –> $g(x_k, \theta)$

p 1300, l 13-16: Please give details of how the likelihood is calculated.

p 1300, l 19: LHS not defined

section 5.3: Again, please give more details of set-up for BO. How were the parameters η selected and adjusted? What was the prior on θ?

Interactive comment on Nonlin. Processes Geophys. Discuss., 1, 1283, 2014.