LETTER

Reply to Robert et al.: Model criticism informs model choice and model comparison

Robert, Mengersen, and Chan (RMC) represent our approach to model criticism in situations when the likelihood cannot be computed (1) as a way to "contrast several models with each other" (2, 3). Moreover, RMC argue that model assessment with Approximate Bayesian Computation under model uncertainty (ABCμ) is unduly challenging and question its Bayesian foundations. We disagree, and clarify that ABCμ is a probabilistically sound and powerful tool for criticizing a model against aspects of the observed data x_0.

Coherence and Power of ABCμ in Model Assessment

ABCμ (1) probes a sampling model M by summarizing the data similar to ABC and introducing error terms, each associated to one summary, to quantify the frequency of observed error magnitude under θ,

$$\varepsilon_{1:K} \to \xi_{\theta,x_0}(\varepsilon_{1:K}) = \lim_{h\to 0} \int \delta_h \Big((\rho_k(S_k(x), S_k(x_0)) - \varepsilon_k)_{1:K} \Big) f(x|\theta, M) dx.$$

Because we shift only the observed summaries (without further adjustments, as in ref. 2),

$$\theta, \varepsilon_{1:K} \to f_{\rho,\tau}(x_0|\theta, \varepsilon_{1:K}, M) = \xi_{\theta,x_0}(\varepsilon_{1:K})$$

is proportional to a density in x_0. Furthermore, transformations of $\xi_{x_0,\theta}(\varepsilon_{1:K})\pi_{\varepsilon_{1:K}}(\varepsilon_{1:K})$ must also change the scale τ of $\pi_{\varepsilon_{1:K}}$ when the Jacobian is not constant (see ref. 4 for details). It is well known that ABC and ABCμ are not invariant to changes in $x \to \rho_k(S_k(x), S_k(x_0))$ and τ. This leaves ABCμ probabilistically sound, but calls for sensitivity analyses.

In contrast to ABC, the error $\varepsilon_{1:K}$ is no longer understood as a latent random variable, and we termed $f_{\rho,\tau}(x_0|\theta, \varepsilon_{1:K}, M)$ "augmented likelihood" (1) simply to indicate that the state space was extended. Labeling $f_{\rho,\tau}(x_0|\theta, \varepsilon_{1:K}, M)$ a "shifted likelihood" seems more appropriate (4).

Using multiple error terms $\varepsilon_{1:K}$ is integral to our method. ABCμ is powerful in revealing model mismatch whenever all discrepancies are simultaneously away from zero for any θ. To escape unidentifiability, the crux is to use errors associated to codependent summaries that may reveal model inconsistencies (see ref. 4 and Figs. 1 and 2). These summaries need not be sufficient. Finally, the smaller τ can be chosen, the more we are able to criticize a fitted model, and the influence of π_θ is attenuated (4).

Because ABCμ makes only fuller use of the data already generated in ABC [e.g., Std-ABCμ (1), mcmcABCμ (4)], no extra computational cost is incurred (see Fig. 2). The second algorithm in ref. 1 may help when the data-generating process is volatile.

Model Criticism and Model Comparison

Model criticism and model comparison are complementary; methods for model comparison attempt to choose between candidate models, even if all of them do not match the data.

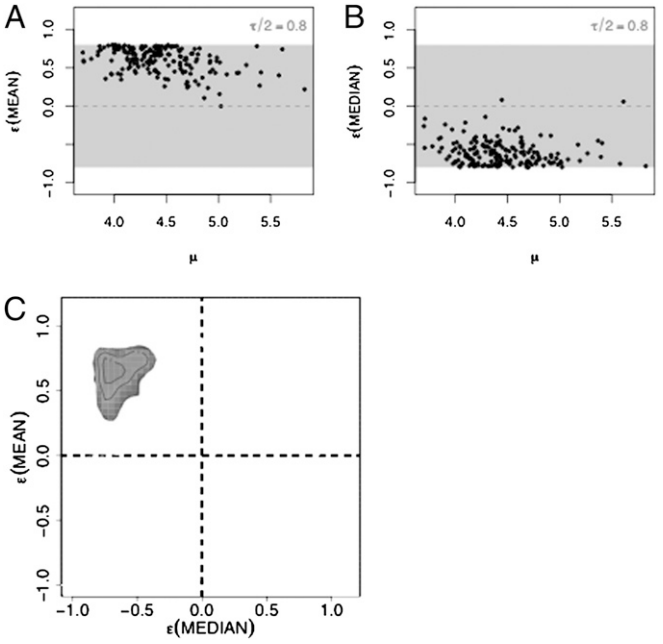

Fig. 1. To illustrate the power of ABCμ in revealing existing model mismatch, consider 100 independent samples that are exponentially distributed with rate 0.2, suppose that each sample is generated according to a Gaussian likelihood model with unknown mean $\mu \in \mathbb{R}$ and $\sigma^2 \geq 0$ (denoted by M_2) and summarize the data with the sample mean and median. We consider $\rho_k(S_k(x), S_k(x_0)) = S_k(x) - S_k(x_0)$, a normal-inverse gamma prior $\pi_0(\theta|M_2) = \pi(\mu|\sigma^2, M_2)\pi(\sigma^2|M_2) = \mathcal{N}(\mu; \mu_0, \sigma^2)\mathrm{IG}(\sigma^2; \alpha_0, \beta_0)$ with $\mu_0 = 5$, $\alpha_0 = 4$, and $\beta_0 = 75$. The hyperparameters are chosen such that it is possible to employ a simple ABCμ algorithm, and results are insensitive to the choice of μ_0, α_0, β_0, but may require more advanced numerical algorithms. Crucially, we set $\pi_{\varepsilon_{1:K}}(\varepsilon_{1:K}|M_2) = \prod_{k=1}^{2} 1/\tau_k \mathbf{1}\{|\varepsilon_k| \leq \tau_k/2\}$ with $\tau_k = 1.6$ sufficiently small, as is standard practice in ABC. We ran Std-ABCμ (1) and show in *A* and *B* the accepted samples (black, only μ) together with the computed discrepancies as well as the accepting region of Std-ABCμ (gray). The heat plot in *C* reconstructs the density $f_{\rho,\tau}(\varepsilon_{\bar{x}}, \varepsilon_{\text{median}}|x_0, M_2)$. ABCμ clearly suggests model mismatch even though the two summaries are not sufficient for $\theta = (\mu, \sigma^2)$ under M_2. With larger τ and/or unbalanced choices of τ, it is more difficult to identify model mismatch, indicating the importance of choosing τ in such a way that the conflict between both summaries ($K > 1$) propagates into $f_{\rho,\tau}(\varepsilon_{\bar{x}}, \varepsilon_{\text{median}}|x_0, M_2)$.

Approximate Bayes' factors can address both model comparison and model criticism when the likelihood cannot be readily evaluated, and may provide different answers (4). Credibility intervals (1) are informal diagnostics to indicate directions into which a faulty model could be modified.

In biology, we now often face quantities of data that are hard to analyze under current computer resources (e.g., molecular genetic data) and/or are intricate (e.g., interaction networks), or we cross boundaries of biological organization (e.g., systems biology). In such challenging circumstances, ABCμ offers statistical rigor during the initial stages of model design to identify one or a suite of models that are in agreement with many aspects of the data at the same time. This clearly informs model choice.

Author contributions: O.R. designed research; O.R. performed research; O.R., C.A., C.W., and S.R. analyzed data; and O.R., C.A., C.W., and S.R. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed: E-mail: oliver.ratmann@duke.edu.

www.pnas.org/cgi/doi/10.1073/pnas.0912887107

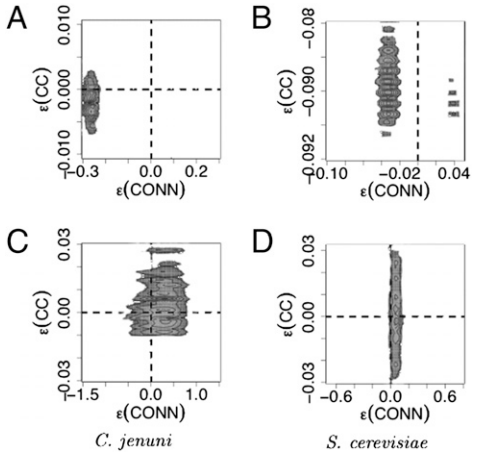

Fig. 2. To illustrate the computational tractability of ABCμ in real-world applications, we analyzed two large protein–protein interaction data sets [*Campylobacter* jejuni (5) and *Saccharomyces cerevisiae* (6)]. As before, interaction networks were summarized with a set of seven codependent topological statistics to quantify the adequacy of the duplication-divergence model DDA+PA and the preferential attachment model PAH under the sampling scheme RS2 (1). The prior density π_θ(θ|DDA + PA) was uniform, π_θ(θ|PAH) was very broad, $\pi_{\varepsilon_{1:K}}$ was set in both cases to $\prod_{k=1}^{K} 1/\tau_k \exp(-2|\varepsilon_k|/\tau_k)$, and τ_k were chosen small. We used the Metropolis–Hastings algorithm mcmcABCμ to obtain samples from $f_{\rho,\tau}(\theta, \varepsilon_{1:K}|x_0, M)$ (4). Because we record only the discrepancies that are computed in any case, ABCμ has exactly the same computational complexity than ABC with the approximation kernel $\pi_{\varepsilon_{1:K}}$. The heat plots show a two-dimensional slice of our seven-dimensional posterior error density for PAH (A, B) and DDA+PA (C, D) across both data sets, indicating that preferential attachment fails to account for the degree correlations and the neighborhood structure in the topology of the data sets. On average, it took 2 min to evaluate all summaries for one *S. cerevisiae* simulation, and hence it would require an extra $2 \times 500/60 \geq 16$ h to obtain 500 samples from the approximate posterior predictive error density proposed by RMC (4). Assuming a generously large acceptance probability of 10%, the extra time required to obtain 500 samples from the weighted posterior predictive error (see ref. 4) is more than 6 days.

ACKNOWLEDGMENTS. OR acknowledges support from NSF grant NSF-EF-08-27416, CA is supported by an EPSRC Advance Research Fellowship, CW by the Danish Research Council, and SR by the Centre for Integrative Systems Biology at Imperial College as well as grant G-0600-609 from the Medical Research Council.

***Oliver Ratmann*[a,b,1], *Christophe Andrieu*[c], *Carsten Wiuf*[d], *and Sylvia Richardson*[e]**

[a]*Department of Biology, Duke University, NC 27708;* [b]*Statistical and Applied Mathematical Sciences Institute, Research Triangle Park, NC 27709;* [c]*Department of Mathematics, University of Bristol, Bristol BS8 1TW, United Kingdom;* [d]*Bioinformatics Research Center, University of Aarhus, 8000 Aarhus C, Denmark; and* [e]*Centre for Biostatistics, Imperial College London, London W2 1PG, United Kingdom*

1. Ratmann O, Andrieu C, Wiuf C, Richardson S (2009) Model criticism based on likelihood-free inference, with an application to protein network evolution. *Proc Natl Acad Sci USA* 106:10576–10581.
2. Robert CP, Mengersen KL, Chen C (2009) Letter: Model choice versus model criticism *Proc Natl Acad Sci USA*, [doi:10.1073/pnas.0911260107].
3. Robert CP, Mengersen KL, Chen C (2009) Model choice versus model criticism. arXiv: 0909.5673v2.
4. Ratmann O, Andrieu C, Wiuf C, Richardson S (2009) Notes: Model criticism informs model choice and model comparison. arXiv: arXiv:0912.3182v1.
5. Parrish JR, et al. (2007) A proteome-wide protein interaction map for Campylobacter jejuni. *Genome Biol* 8:R130.
6. Reguly T, et al. (2006) Comprehensive curation and analysis of global interaction networks in Saccharomyces cerevisiae. *J Biol* 5:11.