Clim. Past Discuss.,
https://doi.org/10.5194/cp-2019-162-SC2, 2020

Climate of the Past
Discussions
Open Access
EGU

Interactive comment on "A Bayesian framework for emergent constraints: case studies of climate sensitivity with PMIP" *by* Martin Renoult et al.

Philip Sansom

p.g.sansom@exeter.ac.uk

Received and published: 19 March 2020

Printer-friendly version

Discussion paper

Comment on Renoult et al.

Philip Sansom and Daniel Williamson

March 19, 2020

Summary

The authors propose an alternative statistical framework for exploiting emergent relationships between multiple climate models in order to constrain future climate. Several advantages are claimed for the proposed method over existing methods, in particular the ability to specify a prior for the quantity of interest directly and independently of the predictive model, and consequently the ability to combine multiple data sources or multiple emergent constraints in a straightforward way. Our view is that the statistical reasoning is difficult to justify for the example of Equilibrium Climate Sensitivity, and impossible for emergent constraints in general. That this is the case is difficult to discern from the current presentation as the model is not written nor justified specifically. This review will attempt to write down the new assumptions and their implications for emergent constraints. The authors make claims regarding the ability of the method to account for model inadequacy, that we argue are not correct. Finally, we attempt to reproduce the results given in the paper and fail for reasons we shall discuss.

Printer-friendly version

Discussion paper

Statistical reasoning for emergent constraints

When analysing emergent constraints the aim is to obtain a prediction of some quantity of interest $Y^\star$ in the real world, given computer simulations $\mathbf{Y} = (Y_1, \ldots, Y_M)'$ of that quantity, simulations $\mathbf{X} = (X_1, \ldots, X_M)'$ of some other quantity that we can observe in the real world, and an observation $X^\star$ of the real world. Mathematically we aim to form the posterior predictive distribution

$$p(Y^\star \mid X^\star, \mathbf{X}, \mathbf{Y}) = \int p(Y^\star, \boldsymbol{\theta} \mid X^\star, \mathbf{X}, \mathbf{Y}) d\boldsymbol{\theta}.$$

where $\boldsymbol{\theta}$ is a vector of unknown parameters. We omit observation uncertainty for brevity.

Like any mathematical model, a statistical model should be the result of transparent chain of logical reasoning. Unfortunately, the authors do not explicitly specify a probability model, and do not supply the reasoning they use to construct the implied model. This makes it extremely difficult for a non-statistician to judge the claims made for the proposed approach, and difficult even for an experienced analyst to reconstruct the underlying reasoning. We attempt to do this here, as the modelling and assumptions are absolutely critical to the emergent constraints debate and because both this approach and the standard approach cannot be correct at the same time. If there is validity in one of them, users need to have as much access to the assumptions and theory underpinning each in order to work out which.

The standard model specified by Bowman et al. (2018), Williamson & Sansom (2019) and used implicitly in every other emergent constraints study is

$$p(Y^\star, \boldsymbol{\theta} \mid X^\star, \mathbf{X}, \mathbf{Y}) \propto p(Y^\star \mid X^\star, \boldsymbol{\theta}) p(\mathbf{Y} \mid \mathbf{X}, \boldsymbol{\theta}) \pi(\boldsymbol{\theta}).$$

In practice, if $X^\star$ is not known precisely, then a prior $\pi(X^\star)$ must also be specified.

Printer-friendly version

Discussion paper

From the manuscript, it is possible to deduce that the implied alternative model is

$$p(Y^\star\boldsymbol{\theta} \mid X^\star, \mathbf{X}, \mathbf{Y}) \propto p(X^\star \mid Y^\star, \boldsymbol{\theta}')\pi(Y^\star)p(\mathbf{X} \mid \mathbf{Y}, \boldsymbol{\theta}')\pi(\boldsymbol{\theta}').$$

The central feature of the proposed approach is to specify $p(X^\star \mid Y^\star)$ and $p(X \mid Y)$ rather than $p(Y^\star \mid X^\star)$ and $p(Y \mid X)$. It is important to realise that these are two fundamentally different statistical models, they cannot both be valid at the same time, and will inevitably result in different inferences for $Y^\star$. Though both equations hold mathematically (as they are valid factorisations of the joint distribution), they imply different conditional independencies in the modelling that need to be physically interpreted and are certainly not interchangable. Therefore, one must consider carefully the underlying reasoning before adopting one or the other. For emergent constraints, $Y^\star$ is usually a measurable property of the future climate and $X^\star$ an observable property of the current or historical climate. Therefore it makes immediate sense to adopt the standard model for emergent constraints, i.e., the future depends upon the past via $p(Y^\star \mid X^\star)$ and $p(Y \mid X)$. In those cases the proposed approach would make no sense because it explicitly states that the past depends on the future via $p(X^\star \mid Y^\star)$ and $p(X \mid Y)$. Equilibrium Climate Sensitivity (ECS) is operationally defined as "the temperature anomaly reached at equilibrium following a instantaneous doubling of CO_2". To us, it seems natural to view this as a future climate quantity, and so it makes sense to adopt the standard model for ECS. We also point out that generally statistical reasoning demands that we predict quantities of interest using observables as predictors, though tradition is not as concrete an argument as the causal one.

It might be possible to justify the proposed model for certain quantities of interest, but for an operationally defined future quantity, we would have to be willing to accept the reversal of time's arrow, i.e., past depends on future. It may be that the authors have in mind some alternative definition or interpretation of ECS that renders time's arrow an illusion for this quantity in particular. We are statisticians and take no view on the validity of such a physical argument, but we must strongly insist that the physical argument behind the statistical model be made explicit, well defended, and open to

the scrutiny of other researchers within the field. The models are different and the inferences and conclusions will be different. We feel that transparency is ultimately key to resolving differences.

We note that the given approach would render any claims of causality for emergent constraints impossible, undermining efforts elsewhere in the community to put emergent constraints on a firmer theoretical basis, e.g., Hall et al. (2019).

Results and reproducibility

The authors supply all the data used in the study in Table 1, and the observations and priors are given in the text, making it possible to check some of the claims made. We were interested to reproduce these results and compare with our approach. However the statistical model is not stated explicitly in its entirety, nor are the models it is compared to. For the sake of transparency, we interpret the model used by the authors to be

$$\begin{aligned} X_m &\sim \text{Normal}(\beta_0 + \beta_1 Y_m, \sigma^2) \quad \text{for } m = 1, \ldots, M \\ X^\star &\sim \text{Normal}(\beta_0 + \beta_1 Y^\star, \sigma^2) \\ Z &\sim \text{Normal}(X^\star, \sigma_Z^2) \end{aligned}$$

where Z is the observed value of $X^\star$ and σ_Z^2 is the observation uncertainty. The default priors are

$$\begin{aligned} \beta_0 &\sim \text{Normal}(0, 1^2) \\ \beta_1 &\sim \text{Normal}(-1, 1^2) \\ \sigma &\sim \text{Half-Cauchy}(5) \\ Y^\star &\sim \text{Cauchy}(2.5, 3) \quad \text{truncated at } 0. \end{aligned}$$

For the Last Glacial Maximum $z = -2.2$ and $\sigma_Z = 0.7/\Phi^{-1}(0.95) = 0.43$, and for the mid-Pliocene Warm Period $z = 0.8$ and $\sigma_Z = 1.6/\Phi^{-1}(0.95) = 0.97$, where Φ^{-1} is the quantile function of the normal distribution. This is the model we implement and it is trivial to do so using the Stan probabilistic programming language, however we are unable reproduce many of the reported results. Our comparisons were based on 30 000 samples from four independent chains after discarding the first 5 000 samples as warm-up for a total of 100 000 samples. The chains were checked for convergence both visually and using Gelman-Rubin diagnostics. The medians are usually within believable sampling error, but there are often large discrepancies in the credible intervals.

In particular, for mPWP PlioMIP1 we obtain posterior median 2.5K $(90\%$ CI $0.4 - 6.3$K$)$ compared to 2.4K $(90\%$ CI $0.5 - 5.0$K$)$, and for mPWP PlioMIP1+PlioMIP2 we obtain median 2.4K $(90\%$ CI $0.4 - 5.6$K$)$ compared to 2.4K $(90\%$ CI $0.4 - 5.0$K$)$. We are, of course, open to the possibility that we have failed to interpret the manuscript text and to implement the correct model, however the Python scripts that accompany the manuscript suggest that we have, and our source code is included for transparency. Examination of those same Python scripts reveals that four chains of only $2\,000$ samples each with no warm-up were used in the production of the reported results (a total of only $8\,000$ samples). Both STAN and PyMC3 (used by the authors) implement the No U-Turn Sampler (NUTS) variant of Hamiltonian Monte Carlo for efficient mixing and fast convergence, so our results should be comparable The lack of warm-up / burn-in period used by the authors is likely to lead to skewed estimates, even using NUTS. Further, the inefficient use of importance sampling to account for the observation uncertainty and the small total number of samples given notorious the difficulty of efficiently sampling from the Cauchy distribution are all likely to contribute to the differences we see when implementing their framework. Unless we have misunderstood their modelling (and by itself this would be an argument for making it explicit in the manuscript), we are not sure that the numbers given in the text actually represent the posteriors of their alternative model faithfully.

Bowman et al. (2018) include explicit expressions for projection using the Kalman filter model (Equations 17 & 23). However using these expressions we are also unable to reproduce the results for the Kalman filter quoted in Table 2 of the manuscript. Examining the Python script that accompanies this manuscript reveals an obvious error in the expression for the posterior variance on Line 80. The authors should therefore revisit the calculation of these credible intervals.

We were able to reproduce the Ordinary Least Squares (OLS) estimates and intervals. However, in Section 2.1 the authors equate OLS with frequentist linear regression. This is incorrect. As discussed in detail in Williamson & Sansom (2019), OLS is a purely algorithmic method of parameter estimation in a mathematical model. The OLS estimates of the mean parameters in a linear regression model are equal to those obtained by frequentist maximum likelihood estimation, *but OLS provides no estimate of uncertainty in either the parameters or the prediction*. Frequentist regression is difficult to justify in a climate change context, but Bracegirdle & Stephenson (2011) presented emergent constraints within a frequentist linear regression framework and this approach has been adopted in many subsequent studies. However, the authors present heuristic uncertainty estimates based on mean-squared errors and on lines 287-288 claim that "OLS" underestimates uncertainty compared to their method. In fact, when standard frequentist regression is used, the inference for ECS in the LGM PMIP2 experiment is very similar to the proposed model with median 2.8K and 90% confidence interval 1.0–4.5K. As Williamson & Sansom (2019) point out, this is equal to an equal tailed 90% Bayesian credible interval under reference priors. Credible intervals from the reference model under the other experiments are less similar to the proposed model, but wider than either the "OLS" or Kalman filter estimates.

Treatment of model inadequacy

In the statistical reasoning above we omitted discussion of model inadequacy for brevity. In lines 130–134 and lines 376–378 the authors claim that in their approach model inadequacy can be entirely accounted for by specifying a larger residual variance for reality than the models and it is not necessary to consider differences in the regression parameters. With a minor modification to the proposed model, the intercept can be made independent of the slope, and therefore any additional uncertainty about the intercept in the real world can safely be pushed into the residual since both sources of uncertainty are independent of $Y^\star$. However, any uncertainty in the slope leads to uncertainty about $X^\star$ that is dependent on $Y^\star$, i.e, the width of the predictive interval for $X^\star$ increases with the distance of $Y^\star$ from the prior mean. Therefore any additional uncertainty in the slope in the real world is also conditional on the value of $Y^\star$ and not accounted for by the residual variance. This is a simple matter of geometry and is therefore unavoidable. Williamson and Sansom (2019) developed a coherent elicitation of the regression parameters and structural error that was designed to account for these geometric considerations, and any emergent constraints framework that wishes to account for structural error, whether using the authors approach or the standard one, must grapple with the geometry.

References

Bowman, K. W., Cressie, N., Qu, X., & Hall, A. D. (2018). A hierarchical statistical framework for emergent constraints: application to snow-albedo feedback. Geophysical Research Letters, 45(23), 13050–13059. https://doi.org/10.1029/2018GL080082

Bracegirdle, T. J., & Stephenson, D. B. (2012). Higher precision estimates of regional polar warming by ensemble regression of climate model projections. Climate Dynamics, 39(12), 2805–2821. https://doi.org/10.1007/s00382-012-1330-3

all, A., Cox, P., Huntingford, C., & Klein, S. (2019). Progressing emergent constraints on future

climate change. Nature Climate Change, 9(4), 269–278. https://doi.org/10.1038/s41558-019-0436-6
Williamson, D. B., & Sansom, P. G. (2019). How are emergent constraints quantifying uncertainty and what do they leave behind? Bulletin of the American Meteorological Society, 100, 2571–2588. https://doi.org/10.1175/bams-d-19-0131.1

```
## Load libraries
library(rstan)

## Model fitting
fit.blm <- function(X, z, sigmaz,
                    chains = 4, iter = 3e4, warmup = 5e3, cores = 4) {

  ## Fit ``Bayesian Linear Regression'' with truncated Cauchy prior
  data <- list(M = nrow(X), x = X$x, y = X$y, z = z, sigmaz = sigmaz)
  blm  <- sampling(object = blm.model, data = data, chains = chains,
                   iter = iter, warmup = warmup, cores = cores)
  blm  <- as.data.frame(blm)

  ## Return results
  round(quantile(blm$ystar, c(0.50,0.05,0.95)), 1)

}

## STAN code for ``Bayesian Linear Regression''
blm.code <- "
  data {
    int<lower=0> M;        // number of models
    real x[M];             // predictand
```

```
    real y[M];              // predictor
    real z;                 // observation
    real sigmaz;            // observation uncertainty
  }
  parameters {
    real beta0;             // intercept
    real beta1;             // slope
    real<lower=0> sigma;    // standard deviation
    real xstar;             // latent predictand state
    real<lower=0> ystar;    // latent predictor state
  }
  model {
    // Priors
    beta0 ~ normal( 0.0, 1.0);
    beta1 ~ normal(-1.0, 1.0);
    sigma ~ cauchy( 0.0, 5.0) T[0,];
    ystar ~ cauchy( 2.5, 3.0) T[0,];
    // Models
    for (m in 1:M)
      x[m] ~ normal(beta0 + beta1*y[m], sigma);
    // Observations
    z ~ normal(xstar, sigmaz);
    // Reality
    xstar ~ normal(beta0 + beta1*ystar, sigma);
  }
"

## Compile STAN model
blm.model = stan_model(model_name = "blm", model_code = blm.code)
```

```
##############################
## The Last Glacial Maximum ##
##############################

## Data
pmip2 <- data.frame(x = c(-2.70,-2.73,-2.16,-3.18,-2.42,-2.73,-1.37),
                    y = c(+4.00,+4.40,+2.70,+3.40,+2.30,+3.30,+1.80))
pmip3 <- data.frame(x = c(-2.56,-3.46,-2.41,-2.58,-1.68,-2.80,-3.15),
                    y = c(+3.20,+4.13,+4.67,+3.45,+3.25,+2.60,+3.37))
pmip4 <- data.frame(x = c(-2.06,-2.23),
                    y = c(+3.01,+2.66))

## Observations
z      <- -2.2
sigmaz <-  0.7/qnorm(0.95)

## Results
pmip2.fit   <- fit.blm(pmip2, z, sigmaz)
pmip3.fit   <- fit.blm(pmip3, z, sigmaz)
pmip23.fit  <- fit.blm(rbind(pmip2,pmip3), z, sigmaz)
pmip234.fit <- fit.blm(rbind(pmip2,pmip3,pmip4), z, sigmaz)


##################################
## The mid-Pliocene Warm Period ##
##################################
```

```
## Data
pliomip1 <- data.frame(x = c(1.03,1.33,1.99,1.16,2.18,1.15,1.93,1.45,2.14),
                       y = c(3.20,3.40,4.05,2.80,4.10,3.20,3.30,2.10,3.37))
pliomip2 <- data.frame(x = c(0.92,2.12,1.37),
                       y = c(2.60,4.50,2.29))

## Observations
z      <- 0.8
sigmaz <- 1.6/qnorm(0.95)

## Results
pliomip1.fit  <- fit.blm(pliomip1, z, sigmaz)
pliomip12.fit <- fit.blm(rbind(pliomip1,pliomip2), z, sigmaz)
```

Interactive comment on Clim. Past Discuss., https://doi.org/10.5194/cp-2019-162, 2020.

Printer-friendly version

Discussion paper