HOW ARE EMERGENT CONSTRAINTS QUANTIFYING UNCERTAINTY AND WHAT DO THEY LEAVE BEHIND?

Daniel B. Williamson and Philip G. Sansom

Emergent constraints underreport uncertainty and are based on strong, unrealistic statistical assumptions, but they need not be. We show how to weaken the assumptions and quantify important uncertainties while retaining the simplicity of the framework.

E mergent constraints have become a popular and controversial topic within the climate science community over the last number of years (Hall and Qu 2006; Wenzel et al. 2016; Cox et al. 2018). For some policy relevant quantity that we cannot observe now, for example equilibrium climate sensitivity (ECS), researchers seek to discover whether there are observations that we can make that would quantify or constrain our uncertainty in that quantity.

AFFILIATIONS: Williamson—Department of Mathematical Sciences, University of Exeter, Exeter, and Alan Turing Institute, London, United Kingdom; Sansom—Department of Mathematical Sciences, University of Exeter, Exeter, United Kingdom
CORRESPONDING AUTHOR: Daniel B. Williamson, d.williamson@exeter.ac.uk

To answer this question, the community has looked to the ensembles of the Coupled Model Intercomparison Projects CMIP3 (Meehl et al. 2007) and CMIP5 (Taylor et al. 2012), and now CMIP6 (Eyring et al. 2016). The idea is to find a (typically linear) "emergent" relationship across the models between the quantity of interest (QoI; e.g., ECS) and something that can be measured. For example, Hall and Qu (2006) found that the current seasonal cycle had a linear relationship with snow albedo feedback in CMIP models. Cox et al. (2018) relate ECS to a particular metric of climate variability. Once such a relationship is found, the models are used to estimate it via regression. Observations from the real world, coupled with the regression produce a constraint on the QoI in reality.

There are a number of reasons that this practice has caused controversy. One is the way in which the constraints are found. Some use physical reasoning to show that we would expect a linear relationship between model quantities, and then look to confirm this through the ensemble (e.g., Cox et al. 2018). Others have suggested data mining be used to find

them (e.g., Karpechko et al. 2013). Hall et al. (2019) highlight the importance of understanding the physical basis for emergent relationships. We discuss these ideas later. Another source of controversy is the simplicity of the treatment versus the complexity of the models and the quantities of interest. The argument is that the observed relationships are not emergent from the physics and hence predictive, but a result of the interaction of many different processes, well captured in the models or not, which must be better understood in order to say something about reality. A final concern is that emergent constraints actually underestimate uncertainty. Several authors have attempted to quantify the effect of uncertainty in the observations themselves without a formal statistical framework (e.g., Brient and Schneider 2016; Wenzel et al. 2016; Cox et al. 2018). Bowman et al. (2018) constructed a statistical framework for emergent constraints that properly accounts for uncertainty in the observations, but neglects other sources that we seek to address here.

In this paper we will explain the underpinning statistical assumptions and judgements that lead to the existing emergent constraints model. We will highlight the different sources of uncertainty that should be present when finding emergent constraints and show where they can enter the usual framework. We will argue for a simple generalization to existing methods that allows hitherto neglected uncertainties to be quantified and then compare results from this extended model to existing results in the literature. Our goal is to translate the existing underpinning statistical assumptions behind emergent constraints and then place them in a more general framework that allows all assumptions for any emergent constraint analysis to be transparently understood. Our framework highlights all sources of uncertainty and offers methodology for guided quantification of these additional uncertainty sources. To accompany the paper we present an open-source software tool capable of fitting the general emergent constraints model to user-inputted data that allows users to explore the effects of all sources of uncertainty on the analysis. Whether the statistical assumptions themselves are valid for any particular emergent constraint, or at all when using CMIP and observations in this way is a question for the climate community to resolve. This paper and its accompanying software can help to frame this discussion.

In the second section we present the strong statistical assumptions behind emergent constraints and generalize the framework by weakening them. We show where key uncertainties were being ignored

and show how they can be quantified going forward. In the third section we apply the generalized framework to the emergent constraint on ECS recently presented by Cox et al. (2018) to demonstrate the effect of acknowledging additional sources of uncertainty. In the fourth section we discuss quantifying these additional sources of uncertainty and present a default guided specification which is available to use through our software tool. In the fifth section we apply the new framework to a collection of constraints on ECS from the literature and discuss the interpretation of different emergent constraints analyses for the same quantity. The final section contains a discussion. The appendix contains some of the mathematical results used to derive our more general framework. The software tool and user instructions are available at https://github.com/ps344/emergent-constraints-shiny.

EXCHANGEABILITY AND EMERGENT CONSTRAINTS.
Emergent constraints are formed through relationships between climate models. Suppose we have an ensemble of climate models of size n. From each model we can obtain the value of a predictor (something we can observe) x_i, and a response (e.g., ECS) y_i, for $i = 1,\ldots, n$. The general concept is to use this ensemble data to fit a regression model:

$$y_i = \beta_0 + \beta_1 x_i, \quad i = 1,\ldots, n,$$

and use this model to "constrain" uncertainty for the response in the real world, y^*, given a value for the predictor from the real world, x^*. But what kinds of assumptions are required to underpin such an approach and in what contexts might they be valid?

Ordinary least squares and classical regression. Least squares estimates of $\boldsymbol{\beta} = (\beta_0, \beta_1)^{\mathrm{T}}$ can be obtained without any assumptions, simply by minimizing the sum of squared distances between the y_i and the $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i$ [where $\mathbf{x}_i = (1, x_i)^{\mathrm{T}}$], leading to well-known formulae for estimates $\hat{\boldsymbol{\beta}}$ (see, e.g., Draper and Smith 1998). Using such estimates to account for uncertainty in other models or reality, however, requires a statistical model to formalize the underpinning assumptions.

A classical regression assumes that for the true $\boldsymbol{\beta}$, the errors from the fit are independent and normally distributed with common variance

$$y_i = \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i + e_i, \quad e_i \sim N(0, \sigma^2).$$

The maximum likelihood estimator $\hat{\boldsymbol{\beta}}$ then coincides with the least squares estimator, and prediction

intervals can be constructed for unobserved models or even reality y^* (at x^*), if they are assumed to be independent draws from the same error distribution. But, what might fitting this model require us to assume about the climate models?

There are two ways to treat the models so that fitting this type of regression would make sense, and we will argue that, when unpacked, neither stand up to scrutiny. The first is to assume the existence of a large population of models from which we obtain independent random samples through CMIP. Reality is then another independent random draw from the same population that the models come from. Lack of independence is well documented across climate models so that, if we did believe the existence of such a population, we are sampling a narrow part of it and the regression model is simply not true. If the model is right but the sample is biased, we cannot conclude anything about the model parameters and hence the underlying population of models without modeling the bias specifically. We know there is no "random sample," the models that are in CMIP were specifically designed. That reality should be an independent draw from the population of models with the same error structure is indefensible and against everything we know about models and their relationship to reality. But what does the population of models argument mean anyway? What counts as a model from the population? Is there a resolution dependence, or a modeled process dependence? Does the population include future models at new resolutions we cannot currently run? These questions have yet to be addressed.

A second way to treat the models that does not require a large population sampled independently would be to assume that the models themselves are random. For this interpretation, uncertainty arises through the random nature of the climate model as it deviates from the line $\boldsymbol{\beta}^T\mathbf{x}$. As the models are deterministic, this randomness can only come from initial condition uncertainty leading us to view the deviation as the result of observing a random point on each model's attractor, and the line representing the mean of the attractor as it changes with x. Note this implies every model's attractor has the same "variability" (σ^2), a claim that is difficult to defend.

More natural is a Bayesian approach in which we acknowledge that, before we observe the models, we are uncertain as to what their x_i and y_i values will be, just as we are uncertain about the corresponding x^* and y^* values for reality. We do not need to view any of these values as random and coming from some distribution; they can be fixed and deterministic.

To quantify uncertainty through probabilities, the key concept here is the prior judgement of *exchangeability* between the responses given the predictors. Exchangeability is a weak assumption that amounts to indifference over labels (de Finetti 1974, 1975). Here it says that, for any i, j, we think that no information about the pairs (y_i, x_i) and (y_j, x_j) is encoded in their labels i and j. Hence, if x_i and x_j took the same value, our distribution for y_i and y_j would be the same a priori.

Here the i and j are the labels for the different climate models, so applying this assumption for an emergent constraint means that if the value of the predictor turned out to be the same for any subset of models, there is nothing else that we know about those models that would lead us to change our distribution for the response before seeing the model responses. On the other hand, a view that a particular model better represented various processes might break exchangeability if, that is, it could be articulated, for a given x how the better representation of processes would change our view of the distribution for $y|x$. For example, we might think feedbacks were captured that raised/lowered the expectation for y compared with a model with a poorer representation. The key difference here between classical independence and Bayesian exchangeability is that the former is a property of the models and the way they are chosen, and the latter is a property of the beliefs of the analyst before he/she has observed the data from the models.

Coupled with the assumption that there is a linear relationship between x_i and y_i, this type of exchangeability implies

$$E[y_i|x_i] = \boldsymbol{\beta}^T\mathbf{x}_i.$$

Assuming that $y_i|x_i$ are independent and identically distributed, as in the classical setting, trivially implies exchangeability. To make use of the weaker exchangeability assumption without assuming independence, de Finetti's representation theorem and its various generalizations (Hewitt and Savage 1955; Diaconis and Freedman 1980) imply that given exchangeability, there exists a probability model $p(y|x, \boldsymbol{\theta})$, considered to be a limit of a function of the y_i and a prior distribution on $\boldsymbol{\theta}$, $\pi(\boldsymbol{\theta})$ so that

$$p\left(y_1,\ldots,y_n \middle| x_1,\ldots,x_n\right) = \int \prod_{i=1}^{n} p\left(y_i \mid x_i, \boldsymbol{\theta}\right) \pi\left(\boldsymbol{\theta}\right) d\boldsymbol{\theta}.$$

We include this result for interest only. From a practical perspective it means that for exchangeable quantities we can behave as if a finite collection is an

independent random sample from some probability model, parameterized with $\boldsymbol{\theta}$ and with a prior distribution on $\boldsymbol{\theta}$, $\pi(\boldsymbol{\theta})$.

In the case of emergent constraints, we might view the symmetry and ubiquity of the Normal distribution as attractive for our choice of distribution and set

$$y_i \big| x_i, \boldsymbol{\theta} \sim N\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i, \sigma^2\right), \tag{1}$$

where $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \sigma^2\}$ and we choose a prior $\pi(\boldsymbol{\beta}, \sigma^2)$ to encode any prior information that we have. This is the Bayesian version of the regression problem and, though perhaps unusual at first, is, as argued above, based on much weaker assumptions than the classical version. What's more, if the usual so-called "reference" prior, $\pi(\boldsymbol{\beta}, \sigma^2) \propto 1/\sigma^2$ is used, the classical analysis and the Bayesian analysis coincide (see, e.g., Bernardo and Smith 1994; Gelman et al. 2013). So we can view the current approach used to model emergent constraints as Bayesian with the reference prior on the regression and variance parameters. We discuss physically motivated priors in the "Confidence-linked default priors for physically motivated constraints" section.

Emergent constraints and exchangeable reality. The standard procedure in the emergent constraints literature is to assume reality, $y^\star$, follows the same regression as the other models. From the statistical view we have given, this implies that $y^\star$ is assumed to be exchangeable with all of the climate models given $x^\star$. Usually $x^\star$ is taken to be the observed predictor [though Wenzel et al. (2016) and Cox et al. (2018) numerically integrate out variability in $x^\star$ and Bowman et al. (2018) provide a framework that includes modeling $x^\star$ explicitly, as we will later], and then the regression is used to predict $y^\star$ and calculate prediction intervals.

Taking the stronger classical version of this assumption first, reality is assumed to be an independent draw from the same distribution that the models were drawn from. This is the strongest possible form of assumption linking models and reality and does not seem defensible, or necessary given that it implies the weaker exchangeability assumption that we shall argue *against* below.

Rather than assume reality is an independent draw from the distribution of the models, we could assume conditional exchangeability of $y^\star$ given $x^\star$ with the y_i given x_i. This would amount to the view that there are no processes systematically missing from the models, but present in reality, that might cause us to view the behavior of the real world to be distinguishable from

that of the models. Rougier et al. (2013) dismissed this idea out of hand, yet it is the weakest form of the key assumption driving the calculations currently performed for emergent constraints. We propose a general framework to aid our discussion of the issues.

Suppose we believe the physical insight behind the linearity assumption for our emergent constraint so that

$$y^\star \big| x^\star, \boldsymbol{\beta}^\star, \sigma^{\star 2} \sim N(\boldsymbol{\beta}^{\star\mathrm{T}}\mathbf{x}^\star, \sigma^{\star 2}) \tag{2}$$

was a sensible model, but the regression coefficients and the error standard deviation, $\boldsymbol{\beta}^\star = (\beta_0^\star, \beta_1^\star)^{\mathrm{T}}$ and $\sigma^\star$ were uncertain. Suppose, further, that we believe that the relationships across the models are informative for the relationships in reality, but not necessarily the same. A natural way to express this through a statistical model is to state

$$\boldsymbol{\beta}^\star \big| \boldsymbol{\beta} \sim N(\boldsymbol{\beta}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}^\star}),$$

where

$$\boldsymbol{\Sigma}_{\boldsymbol{\beta}^\star} = \begin{pmatrix} \sigma_{\beta_0^\star}^2 & \sigma_{\beta_0^\star}\sigma_{\beta_1^\star}\rho^\star \\ \sigma_{\beta_0^\star}\sigma_{\beta_1^\star}\rho^\star & \sigma_{\beta_1^\star}^2 \end{pmatrix} \tag{3}$$

and

$$\sigma^{\star 2}\big|\sigma^2 = \sigma^2 + \sigma_R^2, \qquad \sigma_R^2 \sim \mathrm{HN}\left(0, \xi^\star\right), \tag{4}$$

with $\sigma_{\beta_0^\star}$, $\sigma_{\beta_1^\star}$, and $\rho^\star$ representing ways in which missing or incorrectly parameterized processes across models might change the emergent relationship, and $\xi^\star$ acknowledging structural uncertainty that simply makes us more uncertain about what reality might do even having observed the models. HN here indicates the Half-Normal distribution which shares the form of the PDF of the traditional Normal distribution, but with support restricted to $\sigma > 0$, (Gelman 2006). Note that one would be free to change these distributions to incorporate specific physical knowledge where available, but these assumptions are both natural (the reality coefficients are centered on the model coefficients, but uncertain, and the variance for reality is at least as big as the model uncertainty), and sufficient to illustrate a point.

The current exchangeability between models and reality assumed within the literature is recovered if the extra sources of uncertainty $\sigma_{\beta_0^\star}$, $\sigma_{\beta_1^\star}$, and $\xi^\star$ in Eqs. (3) and (4) are collapsed to zero. In the "Priors for the real world" subsection, we argue that $\rho^\star$, the correlation between the intercept $\beta_0^\star$ and slope $\beta_1^\star$, should be fixed at the value estimated from the models. Parameter $\sigma_{\beta_0^\star}$ captures our uncertainty about missing processes that might cause all models to under or over estimate the response y_i, independent of the predictor

x_i, for example, we might believe that a certain missing process will cause all models to underestimate ECS by 2 K. Parameter $\sigma_{\beta_0^\star}$ captures our uncertainty about missing processes that might alter the gradient of the constraint, for example, we might believe that there are additional feedbacks acting on $y^\star$ and depending on $x^\star$ that are missing from the models. Parameter $\xi^\star$ captures our uncertainty about how far $y^\star$ might lie from the true regression line, even if we knew the true relationship perfectly: for example, due to insufficient model resolution or missing processes not directly related to $x^\star$ or $y^\star$. If even one systematic bias in models, or one missing process known to affect the response, can be acknowledged by a researcher or the wider community, then clearly the standard emergent constraints approach is underreporting uncertainty. We demonstrate this in the "Illustration using a recently found emergent constraint" section using a recently discovered emergent constraint on ECS (Cox et al. 2018). In the "Confidence-linked default priors for physically motivated constraints" section, we propose a default approach to setting sensible values for $\sigma_{\beta_0^\star}$, $\sigma_{\beta_1^\star}$, and $\xi^\star$, in the absence of strong beliefs about specific biases or missing processes.

A complete framework for emergent constraints. We propose to use the extended emergent constraints framework described by the statistical model in Eqs. (1) and (2). We discuss general priors for the model parameters, $\pi(\boldsymbol{\beta}, \sigma^2)$, in the "Confidence-linked default priors for physically motivated constraints" section and in the illustration that follows we shall use the reference prior described above (so our regression for the models will coincide with the classical analysis). We use the model for $\boldsymbol{\beta}^\star$ given by Eq. (3) and, instead of Eq. (4), we use a Folded Normal distribution for $\sigma^\star$,

$$\sigma^\star \big| \sigma \sim \mathrm{FN}\left(\sigma, \xi^\star\right). \qquad (5)$$

The Folded Normal is a generalization of the Half-Normal distribution centered on σ instead of zero with density

$$p\left(\sigma^\star \big| \sigma\right) = \frac{1}{\sqrt{2\pi\xi^\star}} \left(\exp\left\{ -\frac{1}{2\xi^\star}\left(\sigma^\star - \sigma\right)^2 \right\} \right.$$
$$\left. + \exp\left\{ -\frac{1}{2\xi^\star}\left(\sigma^\star + \sigma\right)^2 \right\} \right), \quad \sigma^\star \geq 0.$$

We prefer this to the more natural formulation in Eq. (4) because it does not introduce extra parameters; does not bound $\sigma^\star$ below by σ, which may not be appropriate in some circumstances; is strictly positive; and tends to the normal distribution when

σ is relatively larger than $2\sqrt{\xi^\star}$. Keeping our modeling choices within the Normal family enables researchers to more easily fit their own distributions by thinking about means and standard errors for unknowns. We say more about this in the "Confidence-linked default priors for physically motivated constraints" section.

To account for the uncertainty in the observations, let $x^\star$ be the true value of the predictor in reality and z an imperfect observation of it. The simple measurement model

$$z \sim N\left(x^\star, \sigma_z^2\right), \qquad (6)$$

with given error variance σ_z^2 accounts for the observation uncertainty. Often this error might be quite large, particularly if the "observation" really comes from reanalysis. To complete the Bayesian model, a prior on the mean $x^\star$ should be given. A natural specification is

$$x^\star \sim N\left(\mu_x, \sigma_x^2\right), \qquad (7)$$

with the interpretation that, before we see the data, our best guess for real world $x^\star$ is $\mu_x \pm 2\sigma_x$. In situations where $x^\star$ must respect physical constraints (e.g., being strictly positive), other distributions can be used, without affecting the generality of the framework, or our methods of inference. Choosing a reference prior $\pi(x^\star) \propto 1$ recovers the usual emergent constraints model, and so we use this in our reference calculations throughout.

In any particular problem, we specify the rest of our prior uncertainty through the quantities $\boldsymbol{\Sigma}_{\beta^\star}$, $\xi^\star$, and σ_z^2 in Eqs. (3), (5), and (6) respectively (we shall demonstrate specification of these in our example below and more generally in the "Confidence-linked default priors for physically motivated constraints" section). Letting (Y, X) represent the ensemble, we can then use Bayesian software Stan (Carpenter et al. 2017) to generate samples from the posterior predictive distribution $p(y^\star|z, Y, X)$. We give an integral expression for this in the appendix.

The code we have provided with this paper samples from this distribution and is sufficiently flexible that any of the distributional assumptions we have made (such as the use of Normal and Half-Normal distributions) can be easily altered if required. The app we have provided allows users to add their own emergent constraint data and to experiment with the different sources of uncertainty for themselves. What follows is an illustration of these ideas through a reexamination of the Cox et al. (2018) constraint accounting for different levels of uncertainty.
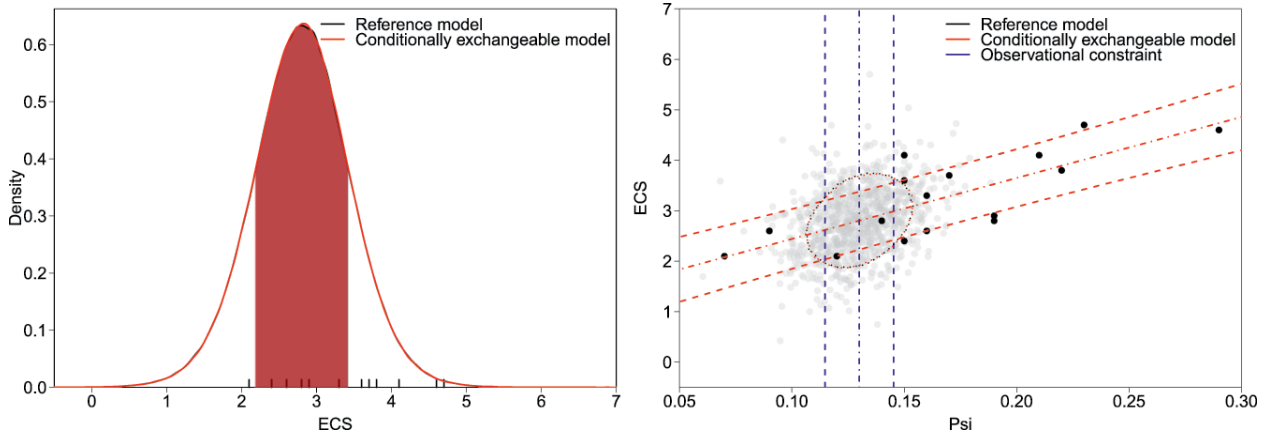
FIG. I. (left) Posterior density for ECS given the models and the observations under the reference prior and with all other uncertainties reduced to 0 K (black) and our model with $x^* \sim N(0.15, 1)$ (red). The shading represents the 66% Bayesian prediction intervals under the two models. (right) The Cox constraint vs ECS. Black dots are the CMIP5 models, the gray dots are samples from our posterior distribution for ECS. Blue vertical lines represent the uncertainty on the observation of the Cox constraint and the straight red lines are the median and prediction intervals for the regression relationship for reality. The red and black contours represent the uncertainty on ECS as it depends on the Cox constraint, with black belonging to the reference model and red, our model.

ILLUSTRATION USING A RECENTLY FOUND EMERGENT CONSTRAINT. We
start with the Ψ statistic presented by Cox et al. (2018) as an emergent constraint on climate sensitivity. The Ψ is a metric of temperature variability (standard deviation of global temperature divided by the negative root one year lag autocorrelation of temperature), with a given physical justification for why it should have a linear relationship with ECS (though some dispute that justification as part of the discussion to that paper).

We begin by introducing what we view as sensible uncertainty judgements, adding the uncertainty in layers so that the effects on the constraint can be observed. Throughout, the reference model refers to the standard emergent constraints model computed by sampling from the posterior under the reference prior. Note, throughout, that the reference prior on the regression coefficients [$\pi(\boldsymbol{\beta}, \sigma^2) \propto 1/\sigma^2$] with $\pi(x^*) \propto 1$ and with Σ_{β^*} and ξ^* in Eqs. (3) and (5) collapsed to zero recovers the usual emergent constraints model.

We use the HadCRUT4 dataset tabulated in Cox et al. (2018) to give the observations, $z = 0.13$ K, and their uncertainty $\sigma_z = 0.016$ K, in Eq. (6). For our non-reference calculations we set $\mu_x = 0.15$ K and $\sigma_x = 1$ K in Eq. (7), based on Fig. 2a of Cox et al. (2018), which shows model time series of Ψ (the data are estimated using a moving average approach) across CMIP5, that are all centered between 0.1 and 0.5 K but with an average of around 0.15 K (by eye). By setting a prior that covers all of the models with much larger uncertainty

than an expert may set, we ensure our analysis is not sensitive to the prior choice (the observation variance is orders of magnitude smaller and so this will not change the posterior very much). Figure 1 shows the posterior distribution of the emergent constraint with these prior choices and reference priors elsewhere. The shading represents the 66% Bayesian prediction interval [the probability that ECS is inside the interval is 0.66, corresponding to the IPCC's "likely" range and chosen to mirror Cox et al. (2018)], with the red curve and shading representing our model with the informed prior on x^* and the black curve representing the Bayesian reference model that coincides with the usual analysis. The reference model gives the same interval as reported in Cox et al. (2018), [2.20, 3.41 K] [black shading (left plot) and black contour (right plot)]. We overlay our model results in red with the same median estimate 2.80 K and interval of [2.20, 3.41 K].

Acknowledging additional uncertainty. Instead of assuming no uncertainty for $\boldsymbol{\beta}^*|\boldsymbol{\beta}$ and $\sigma^*|\sigma$, we look

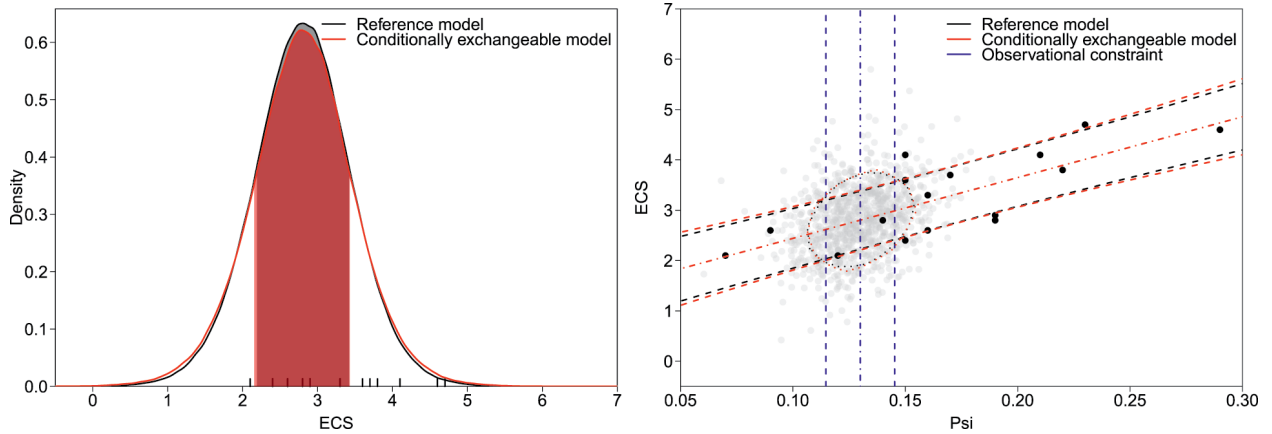| TABLE I. Posterior means and standard deviations for the model regression parameters.		
Parameter	**Mean**	**Std dev**
β_0	1.23	0.47
β_1	12.08	2.65
σ	0.59	0.12

FIG. 2. As in Fig. 1, but with the posterior uncertainties for the regression parameters adopted for the conditional variances of the reality parameters.

at the effect on the emergent constraint of adding a "reasonable" amount by specifying nonzero $\Sigma_{\beta^\star}$ and $\xi^\star$ in Eqs. (3) and (5). In the "Confidence-linked default priors for physically motivated constraints" section we offer a principled approach to setting values for these quantities, which will require a number of additional arguments and results. For illustration here, we shall define reasonable in terms of the relationship of these "reality parameter" uncertainties to the regression parameter uncertainties that come from the Bayesian model.

Having fit the Bayesian regression, we have our beliefs about the relationship between the models through samples from the posterior $\pi(\boldsymbol{\beta}, \sigma | Y)$, which can be used to calculate posterior means and standard deviations for the parameters, shown, for the Cox et al. (2018) constraint in Table 1. The posterior correlation between β_0 and β_1 is $\hat{\rho} = -0.95$.

We begin with the scenario where, given the values of $\boldsymbol{\beta}$ and σ, we would have the same uncertainty (in terms of standard deviations) for $\boldsymbol{\beta}^\star$ and $\sigma^\star$ as we currently do for $\boldsymbol{\beta}$ and σ, using the numbers in Table 1 and a correlation of $\rho^\star = \hat{\rho} = -0.95$ to construct $\Sigma_{\beta^\star}$ and ξ. This effectively doubles the marginal variance for $\boldsymbol{\beta}^\star$ and $\sigma^\star$. The emergent constraint in this scenario is shown in Fig. 2 and has a 66% interval [2.17, 3.43 K]. We can see from the interval and from the plots that, though we have acknowledged additional uncertainty at a level that may seem reasonable to some, the emergent constraint is hardly changed. Increasing all uncertainties by 10% leaves the intervals unchanged (not shown).

Note that even with the additional uncertainty specification given above, we are still virtually certain that the emergent constraint exists in reality given the expected value of the models, that is, our mean for $\beta_1^\star$ would be 12.08 K and our standard deviation

would be 3.75 K. For there to be no relationship ($\beta_1^\star$ crosses 0 K) in reality under this model would involve roughly a four standard deviation event, or a probability of 6.34×10^{-4}! Setting the standard deviation of $\beta_1^\star$ so that no relationship in reality is a two standard deviation event ($\approx$2.5% chance) and a one standard deviation event ($\approx$16.6% chance), and setting the standard deviation of $\beta_0^\star$ at 1 and 2 K for these scenarios respectively (based on an argument that says if $\beta_1^\star = 0$, then $\beta_0^\star$ should be our current best guess for ECS, which we will make more carefully in the "Confidence-linked default priors for physically motivated constraints" section), gives 66% prediction intervals of [2.10, 3.50 K] and [1.88, 3.73 K] respectively. These constraints are shown in Fig. 3 (note we added no additional uncertainty for $\sigma^\star$ for these calculations).

This example shows that not-insignificant additional uncertainty can be acknowledged for an emergent constraint, without dramatically changing the conclusions of the analysis. However, there are clearly sensible levels of additional uncertainty that could matter to an emergent constraint. In any given application, what should the additional uncertainty be? This is a fair question that might often receive the answer "that depends on the beliefs of the scientist." While it is hard to argue with this answer and, while acknowledging that any firm beliefs of the scientist that can be captured with the parameters above and openly defended should be used, we think there is a place for sensible default settings for these uncertainties that can be used and understood by any practitioner. The risk of not having such defaults is that these real additional uncertainties continue to be swept under the carpet by the community and set to zero. We present and justify our default choices below.
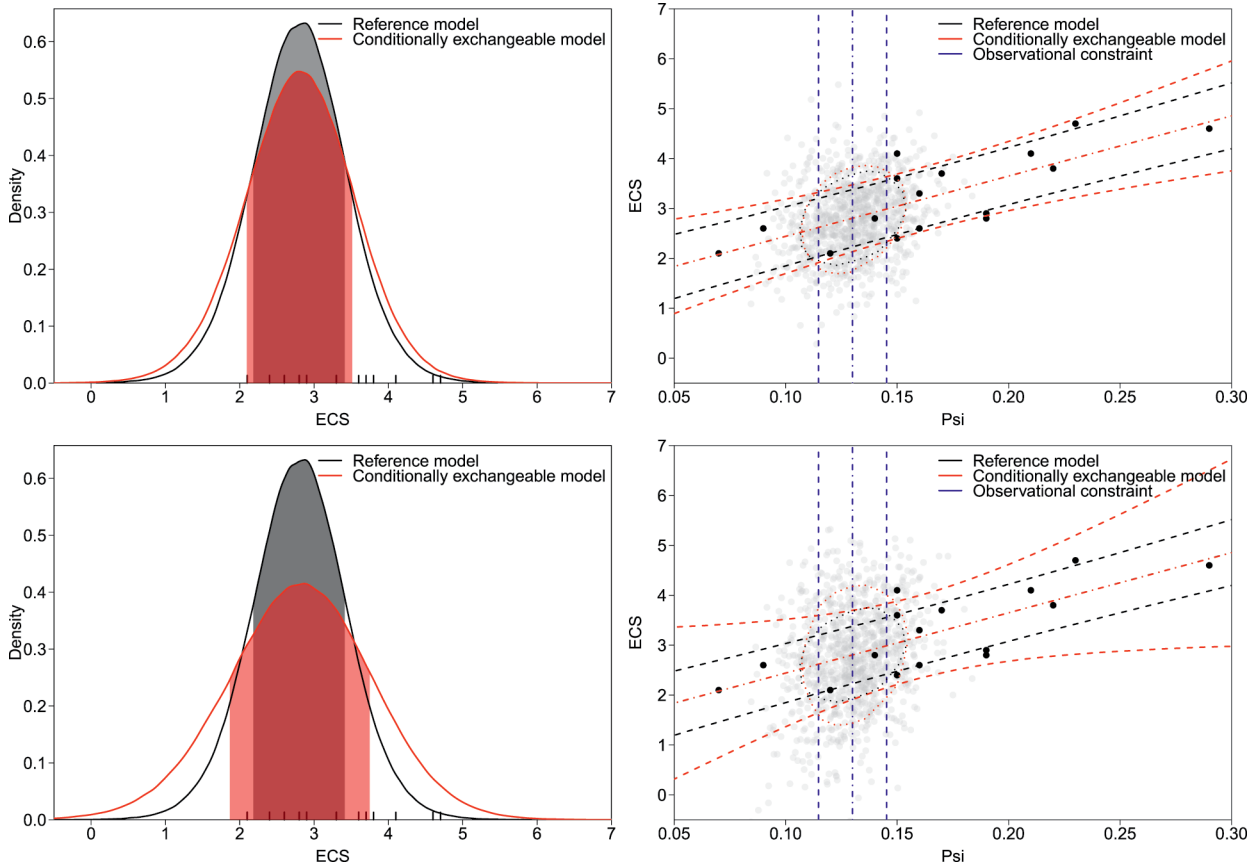
FIG. 3. (top) Emergent constraint plots given a 2.5% chance of no constraint. **(bottom)** Emergent constraint plots given a 16.6% chance of no constraint.

CONFIDENCE-LINKED DEFAULT PRIORS FOR PHYSICALLY MOTIVATED CONSTRAINTS.
The app that accompanies this paper allows the user to work with reference priors throughout and allows all of the quantities that we've introduced by hand to be set manually, giving the user ultimate control and the freedom to express their judgments. For the model regression parameters we go no further than this. In the first subsection, we describe useful subjective default priors for the regression, but we believe that in many instances ensemble sizes will be sufficient to enable the relatively safe use of the reference prior. For the reality relationships our app offers a third, *guided* specification option, based on the arguments and results from the "Priors for the real world" subsection.

Priors for the model relationships. Though the reference prior is often deemed the "objective" prior choice for regression, it actually imparts far less information than any scientist is capable of. For example, the prior states that all intervals of the same width on the real line are equally likely to contain the true intercept and slope, which is preposterous given even a rudimentary knowledge of the scale of the predictors

and responses we might see in the models. Physical knowledge of the response should at least be able to bound the prior support for β and σ^2. For example, consider finding an emergent constraint for ECS. We might view it (nearly) impossible that ECS in any model were outside of the range [0, 10 K]. So if there were no constraint at all, σ^2 should be such that the ensemble mean ECS $\pm 3\sigma$ did not cross both bounds.

A natural choice of prior is

$$\beta \sim N(\mu, \Sigma_\beta), \text{ where } \Sigma_\beta = \begin{pmatrix} \sigma_{\beta_0}^2 & \sigma_{\beta_0}\sigma_{\beta_1}\rho \\ \sigma_{\beta_0}\sigma_{\beta_1}\rho & \sigma_{\beta_1}^2 \end{pmatrix} \quad (8)$$

and $\mu = (\mu_{\beta_0}, \mu_{\beta_1})^T$. We set $\mu_{\beta_1} = 0$ to ensure that "no relationship" is the most likely outcome a priori and that the sign of the constraint will be dictated by the data. Parameter μ_{β_0} can be set to 0, with σ_{β_0} used to set limits on the prior support for the intercept, or physical arguments such as, "if the predictor were zero what would you expect the response to be" used to fix these elements of the prior. Parameter σ_{β_0} can be used to bound the prior support as discussed above. We would recommend setting the prior correlation $\rho = 0$, as negative values would indicate a linear relationship was expected and will appear in

the posterior when the constraint is estimated if that is the case.

As argued by Gelman (2006), a natural choice of prior for σ is a Half-Normal prior,

$$\sigma \sim \mathrm{HN}\left(0, \sigma_s^2\right), \qquad (9)$$

where the Half-Normal distribution shares the form of the PDF of the traditional Normal distribution, but with support restricted to $\sigma > 0$. Though this choice does not lead to analytically tractable Bayesian updating, as with, say, an inverse gamma prior, giving a limit to σ_s is far easier to do for a user, and modern inference with Stan (Carpenter et al. 2017) is extremely fast for problems of this size and type. We apply these ideas to choose a subjective prior for the Cox constraint in the models in the "Application to the Cox constraint" subsection.

Priors for the real world. Equations (2), (3), and (5) gave a model for reality $y^\star$ as a regression on some predictor $x^\star$, with "reality parameters" $\boldsymbol{\beta}^\star$ and $\sigma^\star$, that we link to the output of the models. But the interpretation, particularly for $\boldsymbol{\beta}^\star$ could be problematic. Succinctly, how can there be a regression relationship between $x^\star$ and $y^\star$ in reality when there is only one reality (one $x^\star$ and one $y^\star$)? The following construct offers us a way to think about this statistical model.

Suppose, for the generation of models in our ensemble, the values of $\boldsymbol{\beta}$ and σ could be made known to us (e.g., through many more models of the current generation being included in the sample). At some future time, an ensemble of the next generation of models will be made available to the community and we can reexamine our emergent constraint, finding β' and σ'. We expect the next generation of models to represent physical processes better. Some models will have higher resolution, others will have used the intervening years to develop new parameterizations that overcome known structural biases in their models. If β' and σ', could be made known to us, we would expect them to be different from $\boldsymbol{\beta}$ and σ, as the new physics in the models alters the relationships, even if we may not know if the improved physics would make the slope of the constraint stronger or weaker. We might consider $\boldsymbol{\beta}^\star$ and $\sigma^\star$ to be the model parameters at the limit of the process of improving all of the models and submitting large ensembles. This idea is similar to that introduced as "reification" by Goldstein and Rougier (2009) (where there is discussion of why this theoretical limit should not be reality itself). By considering how different the relationship could be from one generation of models to the next, we may be more easily able to consider the effect of missing processes on the relationship and more

comfortably able to conceptualize how/why $\boldsymbol{\beta}^\star$ might be different from $\boldsymbol{\beta}$ (and similar for $\sigma^\star$).

If limiting relationships between model processes are not a helpful thought construct for considering beliefs about $\boldsymbol{\beta}^\star$ and $\sigma^\star$, a practitioner could consider the effects of missing processes in the models on the constraint. For example, suppose we knew that a systematically missing or misrepresented process led to the response (ECS, say), being 2 too high for every model, but the slope of the line was capturing the underlying physical relationships perfectly. Then we would want to increase $\beta_0^\star$ by 2 to account for this. Similarly, if a feedback process that would strengthen (or weaken) the physical constraint were missing, we would want to adjust $\beta_1^\star$ appropriately. In this way, uncertainty on $\boldsymbol{\beta}^\star$ and $\sigma^\star$ can be considered in terms of whether the current models accurately measure the perceived constraint.

We present arguments for sensible default priors for $\boldsymbol{\beta}^\star$ and $\sigma^\star$ that depend on the level of confidence we have in the physical reasoning leading to the existence of the emergent constraint in the models transferring to reality (or the relationship between different classes of models at the conceptual limit of improvement). Our basic argument will be that, for constraints that were effectively data-mined using the current ensemble, we should have low confidence in their holding in the real world (or the next generation of models), and for those based on purely physical reasoning we might have a greater degree of confidence. To enable us to talk about our confidence in a constraint given the ensemble and to enable other researchers to make similar arguments or debate the level of confidence that should be present, we require further probabilistic arguments.

Suppose

$$\boldsymbol{\beta}^\star | \boldsymbol{\beta} \sim N\left(\boldsymbol{\beta}, \boldsymbol{\Sigma}_{\beta'}\right) \text{ and } \boldsymbol{\beta} \sim N\left(\mathbf{B}, \boldsymbol{\Sigma}_{\beta}\right),$$

then the marginal distribution for $\boldsymbol{\beta}^\star$ is

$$\boldsymbol{\beta}^\star \sim N\left(\mathbf{B}, \boldsymbol{\Sigma}_{\beta} + \boldsymbol{\Sigma}_{\beta'}\right).$$

See the appendix for a proof of this result. This result is relevant because, given the ensemble, (X, Y), we would expect

$$\boldsymbol{\beta} | (X, Y) \,\dot{\sim}\, N\left(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Sigma}}_{\beta}\right), \text{ where}$$

$$\hat{\boldsymbol{\Sigma}}_{\beta} = \begin{pmatrix} \hat{\sigma}_{\beta_0}^2 & \hat{\sigma}_{\beta_0}\hat{\sigma}_{\beta_1}\hat{\rho} \\ \hat{\sigma}_{\beta_0}\hat{\sigma}_{\beta_1}\hat{\rho} & \hat{\sigma}_{\beta_1}^2 \end{pmatrix} \text{ and}$$

$$\hat{\boldsymbol{\beta}} = \left(\hat{\beta}_0, \hat{\beta}_1\right)^{\mathrm{T}}.$$

This is a well-known limiting property of Bayesian analyses for large data and is the basis of the Laplace approximation (Gelman et al. 2013), but is particularly good for this type of quantity. Its veracity can be checked by looking at the posterior samples, a feature available in the software tool that accompanies the paper. Hence, having seen the constraint on the models, we set

$$\boldsymbol{\beta}^\star | (X, Y) \sim N\left(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Sigma}}_{\beta} + \boldsymbol{\Sigma}_{\beta^\star}\right).$$

A subtle point here is that we are assuming that the model for $\boldsymbol{\beta}^\star$ in Eq. (3) is a prior model conditioned on the ensemble (X, Y) (and it will be similar for $\sigma^\star$), rather than a prior we adopt before we see the ensemble. We believe this is the right assumption to use and reflects how emergent constraints research is done in practice. Having found a linear relationship between a predictor and a response in the ensemble (whatever physical arguments led you to look), you must then decide what this tells you about the real world. The posterior mean and variance of $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}$, and $\hat{\boldsymbol{\Sigma}}_{\beta}$, respectively, are easily computed from the posterior samples, and are provided in the data summary in the accompanying software. It remains to specify the prior covariance matrix for reality $\boldsymbol{\Sigma}_{\beta^\star}$, parameterized by $\sigma_{\beta_0^\star}$ and $\sigma_{\beta_1^\star}$ in Eq. (3).

Our "guided" expert judgements involve eliciting a researcher's confidence in the constraint holding in the real world (in the sense we made clear above). We use "confidence" here in a similar way to the IPCC, and will consider levels "virtually certain," "very likely," "likely," with these words implying the same probability levels as they do in the IPCC (99%, 90%, 66%). Suppose we have a $100(1 - \alpha)\%$ confidence level in our emergent constraint being real. We will interpret this as an interval for $\beta_1^\star$ of $[0, T]$ (for a positive constraint), so that the confidence indicates the probability of the constraint crossing zero and thus disappearing (we do not need to consider or find T). This probability is $P(\beta_1^\star < 0) = \alpha/2$. So, for example, suppose you are virtually certain that your constraint holds in reality, then $\alpha = 0.01$ and $P(\beta_1^\star < 0) = 0.005$. Given the Normal marginal distribution for $\beta_1^\star$ described above, standard calculations give

$$\sigma_{\beta_1^\star}^2 = \frac{\hat{\beta}_1^2}{\Phi\left(\alpha/2\right)^{-2}} - \hat{\sigma}_{\beta_1}^2$$

where $\Phi(\cdot)$ is the CDF of the standard Normal distribution.

The same mathematics governs the marginal distribution for $\beta_0^\star$, however, the same sign changing argument does not work for the intercept. Instead, we consider the effect on the intercept $\beta_0^\star$ if the slope $\beta_1^\star$ were to change sign. In that case, and as the slope moved through zero, the intercept should move toward our current expectation for the response. So, for ECS and with a positive constraint, $\beta_1^\star$, as that constraint reduced, the intercept, $\beta_0^\star$, should increase and cross our current mean for ECS (3 K, say) at $\beta_1^\star = 0$. Given the confidence in the constraint as above, and a response with current expectation $\mu_{y^\star}$, we set $P(\beta_0^\star \geq \mu_{y^\star}) \geq \alpha/2$, giving

$$\sigma_{\beta_0^\star}^2 = \frac{\left(\mu_{y^\star} - \hat{\beta}_0\right)^2}{\Phi\left(1 - \alpha/2\right)^{-2}} - \hat{\sigma}_{\beta_0}^2.$$

We set the correlation $\rho^\star$ between $\beta_0^\star$ and $\beta_1^\star$ to be equal to that of the models $\hat{\rho}$, which in our experience is usually large and negative, reflecting the geometry of fitting straight lines, rather than any particular judgements about the variability.

As discussed in the appendix, for $\sigma^\star | \sigma \sim \text{FN}(\sigma, \xi^\star)$ and $\sigma \sim \text{FN}(s, \xi)$, the marginal distribution for $\sigma^\star$ is

$$\sigma^\star \sim \text{FN}\left(s, \xi + \xi^\star\right).$$

This relationship is useful if we find $\sigma | (Y, X)$ to be approximately Folded Normal, which we have found to be a reasonable approximation in practice. When it is, we fit $\hat{s}$ and $\hat{\xi}$ using the regression samples numerically. Given a $100(1 - \alpha)\%$ confidence level in the constraint (as discussed above), we consider an argument based on the prior uncertainty of the response for fixing $\xi^\star$. If the constraint, $\beta_1^\star$, were really zero, our model for the response would be a mean ($\beta_0^\star = \mu_{y^\star}$), as argued previously, with uncertainty around that mean represented by $\sigma^\star$. The final judgement our guided elicitation therefore requires is a judgement for how uncertain the response is *currently*, via a standard deviation, $\sigma_{y^\star}$. Note that both $\mu_{y^\star}$ and $\sigma_{y^\star}$, because they pertain to the response (e.g., ECS), could be found via a literature review or even IPCC summaries, as we will use for the Cox constraint.

Having obtained $\sigma_{y^\star}$, we set $\xi^\star$ using the condition

$$P\left(\sigma^\star > \sigma_{y^\star}\right) \geq \alpha/2$$

so that the confidence is linked to whether the constraint actually reduces uncertainty in the response. We set this numerically as there is no analytic expression for the inverse CDF of the Folded Normal. In guided elicitation mode, the app that accompanies the paper requires only $\mu_{y^\star}$, $\sigma_{y^\star}$, and a confidence level

in the constraint (any is possible but the defaults use the IPCC levels) to complete the emergent constraints model.

Application to the Cox constraint. Applying the ideas from the "Priors for the model relationships" subsection, we use the following simple arguments to set Σ_β and σ_s. We know from previous IPCC reports that models typically have a climate sensitivity "around" 3 K and that an ECS of 10 K or a negative ECS would be hugely surprising (in a CMIP model). Under a naive assumption that each model ECS was a uniform draw from [0, 10 K] with no emergent signal at all, the regression should fit a mean of around five with no slope and the residual standard deviation, σ, should be around 2.5 K (so that two standard deviations covers the interval). This is a "worst-case" type regression where the data are far more spread than anyone familiar with ECS could possibly expect, and no signal at all. We can therefore set σ_s = 2.5 K as a weakly informative prior on σ in Eq. (9).

Parameter Ψ is on the order of 0.1 K, and ECS is on the order of 1 K. Hence, as Ψ changes, when multiplied by β_1, we should still expect a change that is on the order of 1 K. Thus, if there is a relationship, β_1 should not be more than order 10. To be cautious and only weakly informative, we set the prior standard deviation σ_{β_1} = 34 K in Eq. (8) so that a β_1 value on the order of 100 is a three standard deviation event. Note the expectation is 0 and so, in the prior, a negative relationship is as likely as a positive one. It is only the magnitude of the possible relationships that we control.

Given changes in ECS that are order 1 K at most, we would expect the intercept, β_0 to be order 1 K for ECS. To allow for the possibility of strong negative effects, we set a very cautious prior standard deviation of σ_{β_0} = 5 K, in Eq. (8) so that the event that the absolute value of the intercept is greater than 10 K is a two standard deviation event. Prior predictive checks (available in the app) show that this prior on the models offers a huge range of potential ECS and relationships, while being sensible. The posterior in this case is almost identical to the reference analysis, perhaps because the signal is clear and the ensemble is sufficiently large.

For the guided real world uncertainty specification, we interpret the IPCC likely range for ECS of [1.5, 4.5 K] as implying a central estimate of μ_{y^*} = 3 K and a standard deviation of σ_{y^*} = 1.5 K. Table 2 shows the 66%, 90%, and 95% prediction intervals under four different confidence levels. What we refer to as

TABLE 2. Bayesian prediction intervals for ECS using the Cox et al. (2018) emergent constraint with four different confidence levels in the physical arguments behind the constraint.

Confidence	66% interval	90% interval	95% interval
Virtually certain	[2.17, 3.43 K]	[1.55, 4.05 K]	[1.21, 4.39 K]
Very likely	[2.08, 3.51 K]	[1.33, 4.26 K]	[0.89, 4.70 K]
Likely	[1.80, 3.76 K]	[0.80, 4.79 K]	[0.19, 5.40 K]
Coin flip	[1.53, 4.06 K]	[0.31, 5.27 K]	[−0.41, 5.99 K]

"coin flip" is a 50% confidence level, though we use 50.1% to avoid numerical issues in our estimation procedure. We say more about this option in the discussion.

The posterior distributions for ECS under the three main levels of confidence are given in Fig. 4 and the updated intervals for the Cox et al. (2018) constraint are given in Table 2. We see in all cases that acknowledging the additional uncertainty inflates the posterior distribution and the intervals, but not so much as to remove the constraint. In all cases, having some physical confidence behind the constraint is enough to ensure that something is learned from the analysis. This is even true in the coin-flip scenario, which leads to a note of caution that we expand upon in the discussion: if constraints have been data-mined from an ensemble rather than physically motivated, we do not think this procedure should be used at all. Even fitting the model and specifying some level of confidence requires a strong scientific statement that one must be prepared to back up with physical reasoning, that is, one consequence of the emergent constraints framework, even our generalized one, is that the central estimate will be determined by the observations and will not be altered by the confidence level.

For the Cox et al. (2018) constraint in particular, we do not offer any judgements as to what the confidence in the constraint should be, as we are not physicists. If the physical reasoning is sound, however, we do insist that the reference model, with all legitimate reality uncertainties ignored, is not appropriate.

EMERGENT CONSTRAINTS IN THE LITERATURE. In this section we apply our extended framework to selected emergent constraints for equilibrium climate sensitivity published within the literature. We only select constraints published with respect to CMIP5 models and we do not include CMIP3 results within the constraints, which may lead our reference intervals to differ from those published. The constraints we choose are the sum of large- and small-scale indices for lower tropospheric mixing
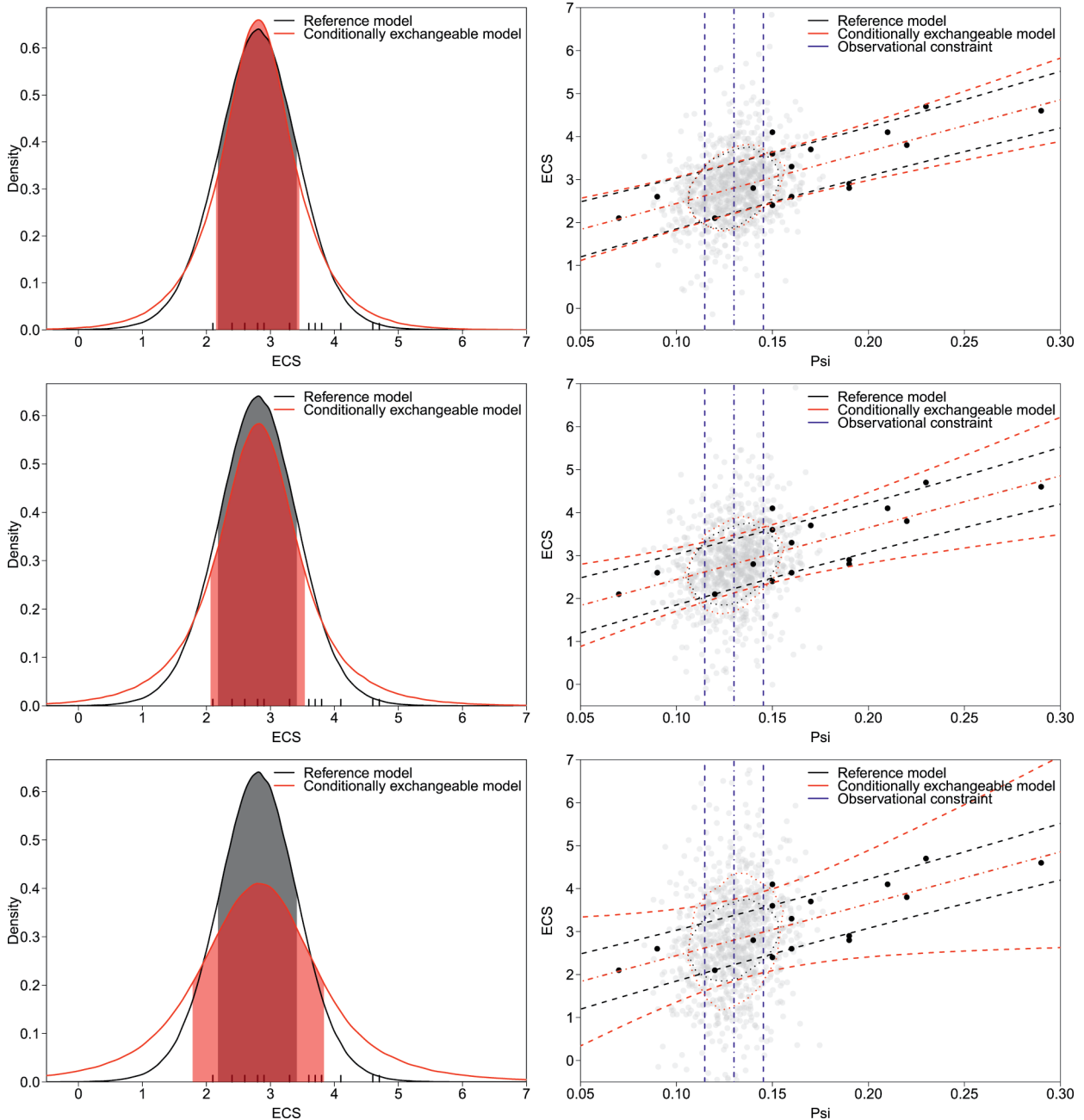
FIG. 4. (top) Emergent constraint plots for ECS given Ψ under a confidence level of *virtually certain* in the existence of the constraint. (middle) As in (top), but under a *very likely* confidence level. (bottom) As in (top), but with a confidence level of *likely*. The black lines and shading represent the reference model.

(Sherwood et al. 2014), the temporal covariance of low cloud reflection with temperature (Brient and Schneider 2016), the double intertropical convergence zone bias (Tian 2015), and the seasonal variation of marine boundary layer cloud fraction with SST (Zhai et al. 2015). The observations and their standard deviations that we used for each constraint are given in Table 3.

The results of applying our extended framework for emergent constraints to these data are given as

66% prediction intervals in Table 4, and shown as PDFs in Fig. 5, for different levels of confidence in the physical arguments behind the constraints. From the figure we see that in cases where we weaken the confidence in the constraint but where the 66% interval remains relatively unchanged, the effect of the additional uncertainty has been to inflate the tails so that our probability of extreme ECS has increased.

We have compared these analyses on alternative emergent constraints on ECS for two reasons. First,

to show that the effect of acknowledging reasonable doubt into the existence of each constraint, as discussed via the method of the "Priors for the real world" subsection, is to inflate the prediction intervals, but by a small amount rather than an amount that points to no result. We can say that emergent constraints *have* underreported uncertainty in the past, but through the given framework, in the future they need not so long as researchers are willing to state their confidence in the underlying physical argument for the linear relationship.

Our second reason is to highlight that published constraints can lead to quite different probability distributions over ECS (e.g., Sherwood predicts a higher climate sensitivity and Cox predicts a much lower climate sensitivity), and to make it clear that these distributions are not compatible in any sense. In each analysis, the authors have made (implicitly) quite different and incompatible conditional exchangeability judgements for ECS given their individual predictors, leading to different models that capture residual variability as Normal with zero mean. A meta-analysis or review of this literature for ECS that sought to give an idea of the current uncertainty in ECS itself, might stray into somehow combining these intervals or central estimates to give an objective view of the state of the science. This would be particularly troublesome if that combination put more weight on intervals that overlapped. Each interval must be thought of as the scientific judgements of the author, based on their confidence and a transparent set of statistical assumptions, as outlined in the "Exchangeability and emergent constraints" section. A form of meta-analysis might seek to take the individual judgements of a group of scientists and summarize them, but that would not lead to an objective uncertainty assessment for ECS, but rather an honest survey of the opinions of different scientists asserted with perhaps differing levels of confidence and based on transparent assumptions and beliefs.

As noted by a reviewer, each of the posterior distributions from the different emergent constraints on

TABLE 3. Observations and standard deviations used in our analyses of four emergent constraints from the literature.		
Constraint	**Obs (z)**	**Std dev (σ_z)**
Sherwood et al. (2014)	0.825 K	0.072 K
Brient and Schneider (2016)	−0.960 K	0.220 K
Tian (2015)	1.000 K	0.500 K
Zhai et al. (2015)	−1.285 K	0.565 K

ECS are symmetric about a central estimate and this may not be a realistic quantification of uncertainty for ECS. More realistic may be that the posterior should be skewed with a longer tail on higher climate sensitivities. Though our folded normal representation for σ^* breaks the usual symmetry in Normal models, the correct place to establish this type of scientific uncertainty judgement within the model is to change the Normal assumption for $y^*|x^*$ in Eq. (2) (normality across the models need not be changed). The linear mean might still be used and our arguments for uncertainty on the intercepts and slopes would be transferable, but a lognormal or shifted gamma-type structure could be used to describe reality given the observations. A benefit of our having formally provided the statistical modeling behind emergent constraints is that practitioners can clearly see which elements of the modeling can be changed in order to capture different types of assumptions.

DISCUSSION. In this paper we sought to unwrap the underpinning statistical assumptions behind the use of emergent constraints to quantify uncertainty for key unknowns in the climate system. We discussed the strong foundational assumptions underpinning the usual classical regression analysis and the interpretation of the real world as a random sample from the distribution of models. We argued that these ideas were too difficult to defend objectively.

We presented the Bayesian view of emergent constraints, and the far weaker and more reasonable a priori conditional exchangeability judgements that would lead to regression analyses that coincided

| TABLE 4. Bayesian 66% prediction intervals for ECS for different published emergent constraints using the reference model and three different confidence levels in the physical arguments behind the constraint as per our extended framework.				
Constraint	**Reference interval**	**Virtually certain**	**Very likely**	**Likely**
Sherwood et al. (2014)	[3.59, 5.02 K]	[3.42, 5.18 K]	[3.19, 5.48 K]	[2.61, 6.04 K]
Brient and Schneider (2016)	[3.10, 4.31 K]	[3.06, 4.37 K]	[2.90, 4.55 K]	[2.52, 4.88 K]
Tian (2015)	[2.66, 4.14 K]	[2.65, 4.17 K]	[2.56, 4.25 K]	[2.27, 4.54 K]
Zhai et al. (2015)	[2.52, 4.33 K]	[2.52, 4.33 K]	[2.53, 4.30 K]	[2.48, 4.35 K]
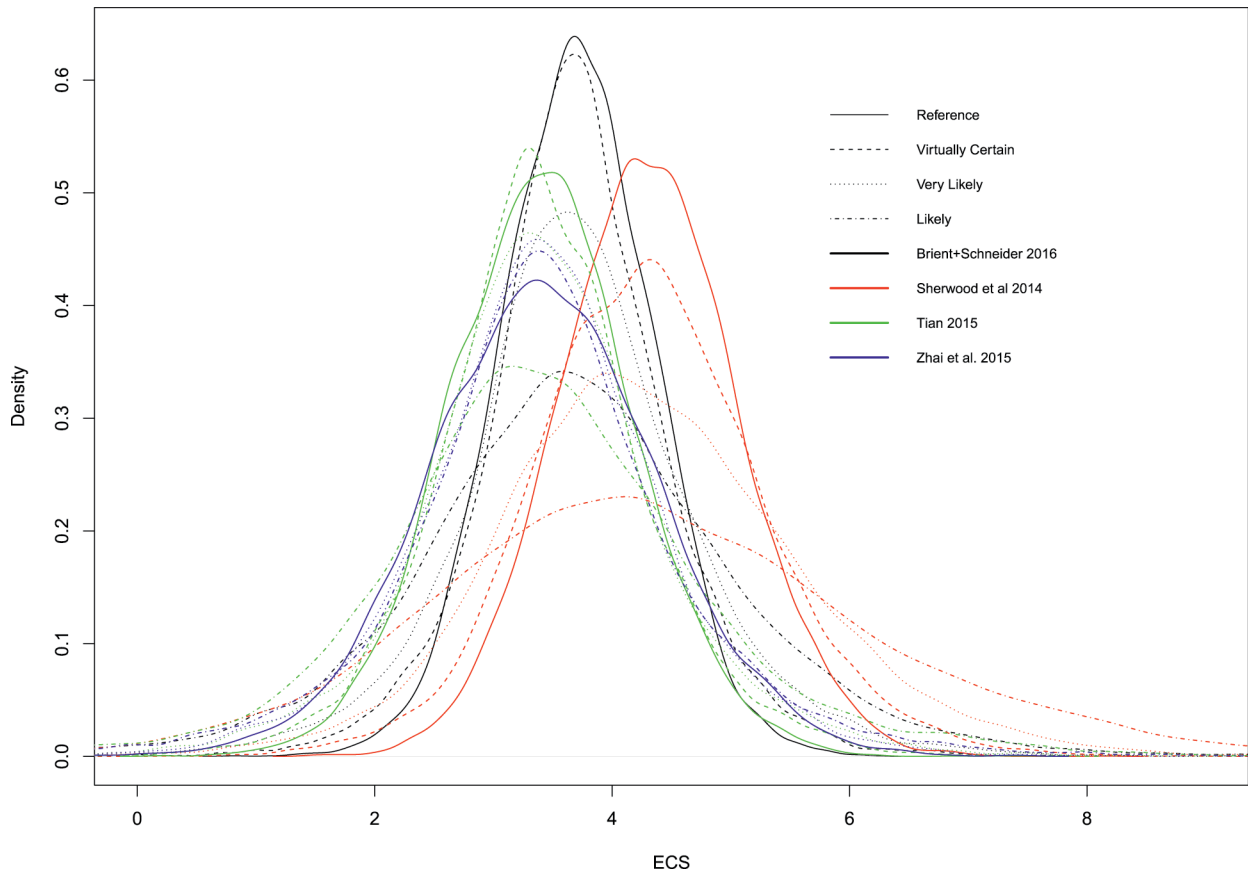
FIG. 5. Posterior probability density functions for ECS found for four different emergent constraints (colors) and four different levels of confidence in the constraint. The solid line in each case is the reference analysis.

with the classical analysis under reference priors and showed how, under this framework, standard emergent constraints analyses ignored the key uncertainties present when there are potential structural deficiencies in the current generation of models. We presented a generalized framework for emergent constraints that acknowledged these additional uncertainties, yet collapsed back to the standard model when these uncertainties were set to zero.

Our modeling looks to adopt the prior judgement that the emergent constraint is informative for reality after having observed the ensemble, to avoid incoherent models for reality beforehand and to acknowledge that these judgements should only be made sparingly. We also believe that this is how scientists think about emergent constraints. As one scientist put it to us by email, "nobody publishes an emergent constraint that doesn't correlate."

We presented a guided prior uncertainty specification that links confidence in the physical reasoning for a linear relationship between the response and the constraint to reasonable additional uncertainties through judgements about the response itself which

are either simple to specify or generally available through literature review. We have developed a software tool that allows users to do this for themselves, and have ensured that this tool also allows scientists full freedom to specify any levels of uncertainty on any of the parameters that they wish, if they do not want to follow our guided specification. Our tool is simple to use and we will maintain it for the community through GitHub. When scientists have specific judgements relating to the models and their deficiencies, we would recommend using our tool and a structured prior elicitation (Gosling 2018) to quantify these effects.

Our modeling accounts for parameter uncertainty, observation uncertainty and uncertainty about how the emergent relationship observed in the models applies to the real world. Sansom et al. (2019) also demonstrate that emergent constraints can be sensitive to uncertainty in the values of the model predictors x_i ($i = 1, \ldots, n$). Our method can be readily extended to account for these errors in variables without effecting the guided uncertainty specification, since only the model posterior distribution of the parameters given the models $\pi(\boldsymbol{\beta}, \sigma | Y, X)$ will change.

The arguments in this paper make it very clear that strong scientific judgement is implied when linking models to reality, particularly when claiming that a linear relationship between quantities across models indicates a physical relationship. Data mining for constraints may very well lead to a multiple testing problem. A simple numerical experiment can be used to illustrate the point. Generating 430,000 (normal) random numbers and stacking these into a matrix with 43 rows, generates a pseudo ensemble with 43 members and with no physical links between the 10,000 outputs. Looking at the maximum absolute correlation between outputs across the ensemble will usually return correlations between 0.70 and 0.85, well above the threshold for relationships for an emergent constraint. To base the strong beliefs required to take this relationship into the real world (in the way we have made clear) on only the discovery of a large correlation cannot be justified. For that reason, even specifying a low confidence in the constraint through our guided framework would still be inappropriate. See also Caldwell et al. (2014) for discussion of this point.

One criticism of emergent constraints is that they are overly simple, ignoring complex nonlinearities or interactions with processes that are not yet well understood or resolved by models. We do not fully agree with this criticism. When the linear relationship can be well established through mathematical and physical arguments, the conditional exchangeability judgements we have explained in this paper, amounting to indifference over labels and, though appreciating that the relationship will not be *exactly* linear, having no strong judgements as to systematic deviations from it, seem plausible in many situations.

While the models and reality themselves may well be more complex, that does not invalidate the statistical model which, rather than making strong statements about how reality/the models actually behave, captures our current knowledge and can be defended on those grounds. Of course, more complex forms of regression could be used within the framework we discuss, but the implied beliefs and the way these will be amended for transferring the constraint from models to reality will be far more complex and difficult to defend.

We hope that by making the required statistical assumptions clear and transparent, the validity of any given constraint, new or existing, can be discussed by the community in terms of the physical reasoning, the reasonableness of the exchangeability judgements, and the confidence in the current generation of models and linear relationship for a given quantity. By making software available to the community, we hope to help this debate move forward by allowing different researchers to look at the sensitivity of intervals to these judgements and to form their own views.

APPENDIX: MATHEMATICAL DETAILS. *Posterior predictive sampling.* The posterior predictive distribution for reality given the models and observations is expressed by the following integral:

$$p\left(y^{\star}\left|z,Y,X\right.\right)=\int p\left(y^{\star},\boldsymbol{\beta}^{\star},\sigma^{\star},\sigma,\boldsymbol{\beta},x^{\star}\left|z,Y,X\right.\right)d\boldsymbol{\beta}^{\star}d\sigma^{\star}d\sigma d\boldsymbol{\beta}dx^{\star}$$
$$=\int p\left(y^{\star}\left|\boldsymbol{\beta}^{\star},\sigma^{\star},x^{\star}\right.\right)p\left(\boldsymbol{\beta}^{\star},\sigma^{\star},\sigma,\boldsymbol{\beta},x^{\star}\left|z,Y,X\right.\right)d\boldsymbol{\beta}^{\star}d\sigma^{\star}d\sigma d\boldsymbol{\beta}dx^{\star}$$
$$=\int p\left(y^{\star}\left|\boldsymbol{\beta}^{\star},\sigma^{\star},x^{\star}\right.\right)p\left(\boldsymbol{\beta}^{\star}\left|\boldsymbol{\beta}\right.\right)p\left(\sigma^{\star}\left|\sigma\right.\right)p\left(x^{\star}\left|z\right.\right)p\left(\boldsymbol{\beta},\sigma\left|Y,X\right.\right)d\boldsymbol{\beta}^{\star}d\sigma^{\star}d\sigma d\boldsymbol{\beta}dx^{\star}.$$

Our software samples from each distribution within this factorization to provide posterior predictive samples.

Bayesian updates. As argued in the "Confidence-linked default priors for physically motivated constraints" section, if

$$\boldsymbol{\beta}^{\star}|\boldsymbol{\beta} \sim N\!\left(\boldsymbol{\beta}, \boldsymbol{\Sigma}_{\beta^{\star}}\right) \text{ and } \boldsymbol{\beta} \sim N\!\left(\mathbf{B}, \boldsymbol{\Sigma}_{\beta}\right)$$

then, the marginal distribution for $\boldsymbol{\beta}^{\star}$ is

$$\boldsymbol{\beta}^\star \sim N\!\left(\mathbf{B}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}} + \boldsymbol{\Sigma}_{\boldsymbol{\beta}^\star}\right).$$

To show this, we have

$$p\!\left(\boldsymbol{\beta}^\star\right) = \int_{-\infty}^{\infty} p\!\left(\boldsymbol{\beta}^\star \mid \boldsymbol{\beta}\right) p\!\left(\boldsymbol{\beta}\right) d\boldsymbol{\beta}$$

$$= \int_{-\infty}^{\infty} \left(2\pi\right)^{-1} \left|\boldsymbol{\Sigma}_{\boldsymbol{\beta}^\star}\right|^{-1/2} \exp\!\left\{-\frac{1}{2}\left(\boldsymbol{\beta}^\star - \boldsymbol{\beta}\right)^{\mathrm{T}} \boldsymbol{\Sigma}_{\boldsymbol{\beta}^\star}^{-1} \left(\boldsymbol{\beta}^\star - \boldsymbol{\beta}\right)\right\}$$

$$\left(2\pi\right)^{-1} \left|\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\right|^{-1/2} \exp\!\left\{-\frac{1}{2}\left(\boldsymbol{\beta} - \mathbf{B}\right)^{\mathrm{T}} \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1} \left(\boldsymbol{\beta} - \mathbf{B}\right)\right\} d\boldsymbol{\beta}$$

$$= A\int_{-\infty^{-}}^{\infty} \exp\!\left\{-\frac{1}{2}\left[\boldsymbol{\beta}^{\mathrm{T}}\left(\boldsymbol{\Sigma}_{\boldsymbol{\beta}^\star}^{-1} + \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\right)\boldsymbol{\beta} - 2\boldsymbol{\beta}^{\mathrm{T}}\left(\boldsymbol{\Sigma}_{\boldsymbol{\beta}^\star}^{-1}\boldsymbol{\beta}^\star + \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\mathbf{B}\right)\right]\right\} d\boldsymbol{\beta}$$

where

$$A = \left(2\pi\right)^{-2} \left|\boldsymbol{\Sigma}_{\boldsymbol{\beta}^\star}\right|^{-1/2} \left|\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\right|^{-1/2} \exp\!\left\{-\frac{1}{2}\left(\boldsymbol{\beta}^{\star\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{\beta}^\star}^{-1}\boldsymbol{\beta}^\star + \mathbf{B}^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\mathbf{B}\right)\right\}.$$

Completing the square for the integrand, it becomes proportional to a Normal distribution in $\boldsymbol{\beta}$ and so the integral becomes

$$\left(2\pi\right)\left|\boldsymbol{\Sigma}_{\boldsymbol{\beta}^\star}\right|^{1/2}\left|\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\right|^{1/2}\left|\boldsymbol{\Sigma}_{\boldsymbol{\beta}^\star} + \boldsymbol{\Sigma}_{\boldsymbol{\beta}}\right|^{-1/2} \exp\!\left\{\frac{1}{2}\left(\boldsymbol{\Sigma}_{\boldsymbol{\beta}^\star}^{-1}\boldsymbol{\beta}^\star + \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\mathbf{B}\right)^{\mathrm{T}}\left(\boldsymbol{\Sigma}_{\boldsymbol{\beta}^\star}^{-1} + \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\right)^{-1}\left(\boldsymbol{\Sigma}_{\boldsymbol{\beta}^\star}^{-1}\boldsymbol{\beta}^\star + \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\mathbf{B}\right)\right\}$$

Combining with the constant, collecting the exponential terms and simplifying gives

$$p\!\left(\boldsymbol{\beta}^\star\right) = \left(2\pi\right)^{-1}\left|\boldsymbol{\Sigma}_{\boldsymbol{\beta}^\star} + \boldsymbol{\Sigma}_{\boldsymbol{\beta}}\right|^{-1/2} \exp\!\left\{-\frac{1}{2}\left(\boldsymbol{\beta}^\star - \mathbf{B}\right)^{\mathrm{T}}\left(\boldsymbol{\Sigma}_{\boldsymbol{\beta}^\star} + \boldsymbol{\Sigma}_{\boldsymbol{\beta}}\right)^{-1}\left(\boldsymbol{\beta}^\star - \mathbf{B}\right)\right\}$$

proving the result. For the Folded Normal result of the "Priors for the real world" subsection, the technique is the same (not shown), though the integral in that case is removed by expressing the integrand as a term proportional to the PDF of a Folded Normal (rather than a Normal).

REFERENCES

Bernardo, J. M., and A. Smith, 1994: *Bayesian Theory.* Wiley, 675 pp.

Bowman, K. W., N. Cressie, X. Qu, and A. Hall, 2018: A hierarchical statistical framework for emergent constraints: Application to snow-albedo feedback. *Geophys. Res. Lett.*, **45**, 13 050–13 059, https://doi.org/10.1029/2018GL080082.

Brient, F., and T. Schneider, 2016: Constraints on climate sensitivity from space-based measurements of low-cloud reflection. *J. Climate*, **29**, 5821–5835, https://doi.org/10.1175/JCLI-D-15-0897.1.

Caldwell, P. M., C. S. Bretherton, M. D. Zelinka, S. A. Klein, B. D. Santer, and B. M. Sanderson, 2014: Statistical significance of climate sensitivity predictors obtained by data mining. *Geophys. Res. Lett.*, **41**, 1803–1808, https://doi.org/10.1002/2014GL059205.

Carpenter, B., and Coauthors, 2017: Stan: A probabilistic programming language. *J. Stat. Software*, **76** (1), https://doi.org/10.18637/jss.v076.i01.

Cox, P. M., C. Huntingford, and M. S. Williamson, 2018: Emergent constraint on equilibrium climate sensitivity from global temperature variability. *Nature*, **553**, 319–322, https://doi.org/10.1038/nature25450.

de Finetti, B., 1974: Theory of *Probability.* Vol. I, John Wiley & Sons, 300 pp.

——, 1975: *Theory of Probability.* Vol II, John Wiley & Sons, 375 pp.

Diaconis, P., and D. Freedman, 1980: Finite exchangeable sequences. *Ann. Probab.*, **8**, 745–764.

Draper, N. R., and H. Smith, 1998: *Applied Regression Analysis.* 3rd ed., John Wiley & Sons, 736 pp.

Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, 2016: Overview of the Coupled Model Intercomparison Project phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.*, **9**, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016.

Gelman, A., 2006: Prior distributions for variance parameters in hierarchical models. *Bayesian Anal.*, **1**, 515–534, https://doi.org/10.1214/06-BA117A.

——, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, 2013: *Bayesian Data Analysis.* 3rd ed. Chapman and Hall/CRC, 675 pp.

Goldstein, M., and J. C. Rougier, 2009: Reified Bayesian modelling and inference for physical systems. *J. Stat. Plann. Inference*, **139**, 1221–1239, https://doi.org/10.1016/j.jspi.2008.07.019.

Gosling, J. P., 2018: SHELF: The Sheffield Elicitation Framework. *Elicitation: The Science and Art of Structuring Judgement*, L. C. Dias, A. Morton, and J. Quigley, Eds., Springer, 61–93, https://doi.org/10.1007/978-3-319-65052-4_4.

Hall, A., and X. Qu, 2006: Using the current seasonal cycle to constrain snow albedo feedback in future climate change. *Geophys. Res. Lett.*, **33**, L03502, https://doi.org/10.1029/2005GL025127.

——, P. Cox, C. Huntingford, and S. Klein, 2019: Progressing emergent constraints on future climate change. *Nat. Climate Change*, **9**, 269–278, https://doi.org/10.1038/s41558-019-0436-6.

Hewitt, E., and L. J. Savage, 1955: Symmetric measures on Cartesian products. *Trans. Amer. Math. Soc.*, **80**, 470–501, https://doi.org/10.1090/S0002-9947-1955-0076206-8.

Karpechko, A. Y., D. Maraun, and V. Eyring, 2013: Improving Antarctic total ozone projections by a process-oriented multiple diagnostic ensemble regression. *J. Atmos. Sci.*, **70**, 3959–3976, https://doi.org/10.1175/JAS-D-13-071.1.

Meehl, G. A., C. Covey, K. E. Taylor, T. Delworth, R. J. Stouffer, M. Latif, B. McAvaney, and J. F. B. Mitchell, 2007: The WCRP CMIP3 Multimodel Dataset: A new era in climate change research. *Bull. Amer. Meteor. Soc.*, **88**, 1383–1394, https://doi.org/10.1175/BAMS-88-9-1383.

Rougier, J. C., M. Goldstein, and L. House, 2013: Second-order exchangeability analysis for multimodel ensembles. *J. Amer. Stat. Assoc.*, **108**, 852–863, https://doi.org/10.1080/01621459.2013.802963.

Sansom, P. G., D. B. Stephenson, and T. J. Bracegirdle, 2019: On constraining projections of future climate using observations and simulations from multiple climate models. *J. Amer. Stat. Soc.*, in press.

Sherwood, S. C., S. Bony, and J.-L. Dufresne, 2014: Spread in model climate sensitivity traced to atmospheric convective mixing. *Nature*, **505**, 37–42, https://doi.org/10.1038/nature12829.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, https://doi.org/10.1175/BAMS-D-11-00094.1.

Tian, B., 2015: Spread of model climate sensitivity linked to double-intertropical convergence zone bias. *Geophys. Res. Lett.*, **42**, 4133–4141, https://doi.org/10.1002/2015GL064119.

Wenzel, S., V. Eyring, E. P. Gerber, and A. Y. Karpechko, 2016: Constraining future summer austral jet stream positions in the CMIP5 ensemble by process-oriented multiple diagnostic regression. *J. Climate*, **29**, 673–687, https://doi.org/10.1175/JCLI-D-15-0412.1.

Zhai, C., J. H. Jiang, and H. Su, 2015: Long-term cloud change imprinted in seasonal cloud variation: More evidence of high climate sensitivity. *Geophys. Res. Lett.*, **42**, 8729–8737, https://doi.org/10.1002/2015GL065911.

AMS Student Leadership Circle

AMS 100th Anniversary Campaign

As the next generation of scientific leaders, students can play a role in helping lift AMS to new heights.

Donate to the AMS 100th Anniversary Campaign by joining the Student Leadership Circle. Your contribution of $20.19 will help raise funds to support the Student Conference, travel grants, and other AMS initiatives.

ametsoc.org/give