QuadTune: A poor man's tuner for expensive global models

Vincent Larson, Zhun Guo, Ben Stephens,
Nils Wittwer, Luis Hasenauer

E3SM Webinar
9 Oct 2025

When we tune a model, we have two goals:

- Tune away parametric error; and
- Get hints about what part of the model structure to change next, i.e. which model equations are still wrong.

The big gains in accuracy will come from improvements to model structure. So we want to tune as quickly as possible, so that we can resume working on the model structure.

Our problem: How can we tune an expensive model?

Both SCREAM simulations and coupled ocean-atmosphere simulations are expensive.

Unless we run short-duration simulations, running a perturbed parameter ensemble (PPE) with hundreds of members is prohibitive.

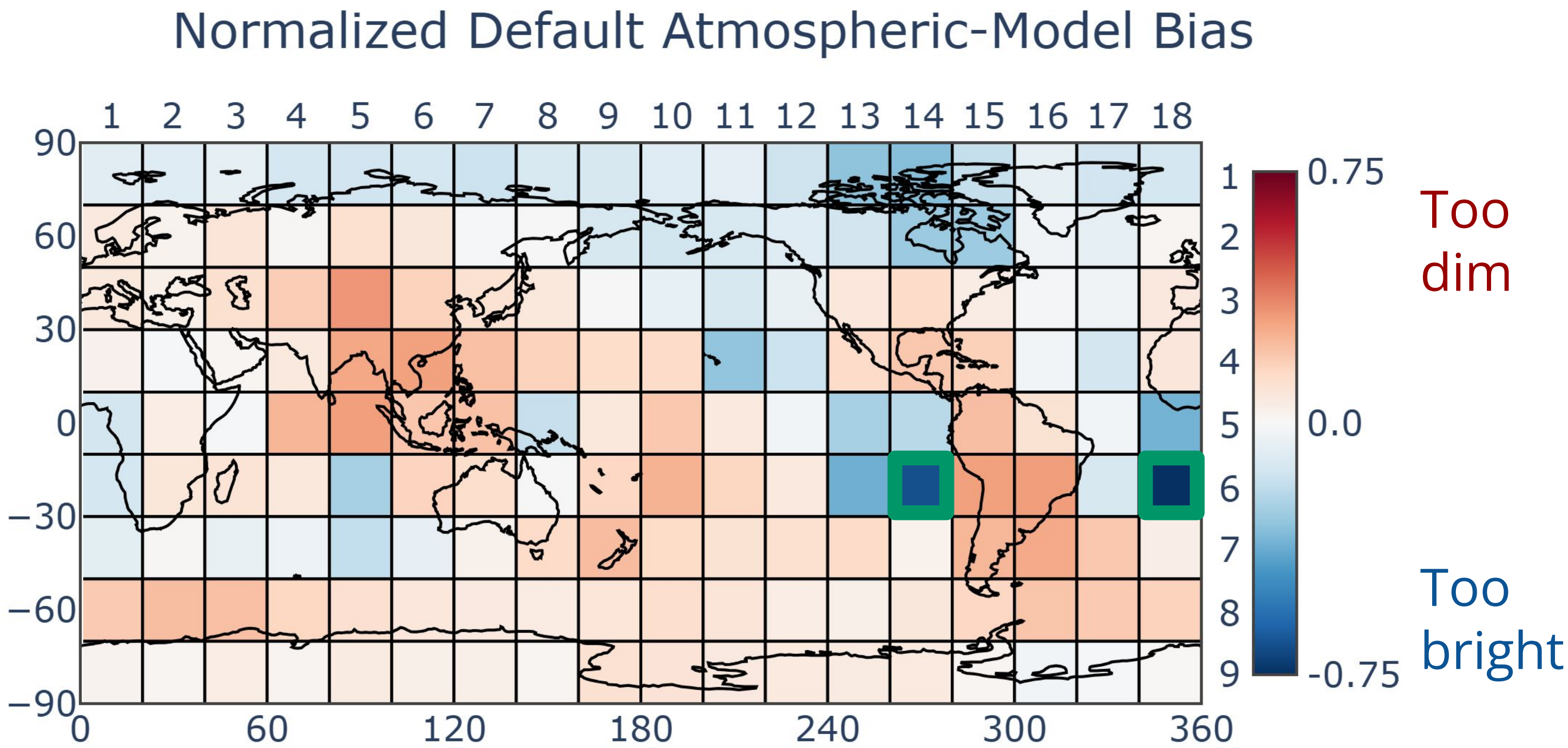
Outline of talk

- We attempt to reduce the cost of tuning by use of a tuner (“QuadTune”) that uses a simple quadratic emulator

- Example tuning results from a global atmospheric model (EAM)

- Two archetypal model errors: Tuning trade-offs and stubborn biases

QuadTune adjusts *P* parameter values, p_j, in order to best match *N* regional metrics m_i (e.g., SWCF in Sc regions, box 6_14 or box 6_18)



The QuadTune tuning recipe:

1. Choose regional metrics (e.g., SWCF in 20°×20° regions, but alternatively, we can choose only a few custom regions)

2. Choose *P* tuning parameters

3. Run *2P+1* global simulations, varying parameters one at a time, perturbing each high and then low (expensive)

4. Minimize difference between model and obs, and create diagnostic plots (cheap)

A linear-regression view of tuning: The goal of tuning is to find a single *dp* that dots into each row (region) and yields the corresponding rhs bias

$$
\begin{array}{r}
\text{stratocumulus region} \\[2.5em]
\text{cumulus region} \\[2.5em]
\text{warm pool region}
\end{array}
\begin{bmatrix}
\dfrac{\partial m_{Sc}}{\partial p_1} & \dfrac{\partial m_{Sc}}{\partial p_2} \\[1.5em]
\dfrac{\partial m_{Cu}}{\partial p_1} & \dfrac{\partial m_{Cu}}{\partial p_2} \\[1.5em]
\dfrac{\partial m_{WP}}{\partial p_1} & \dfrac{\partial m_{WP}}{\partial p_2}
\end{array}
\end{bmatrix}
\begin{bmatrix}
\delta p_1 \\[1.5em]
\delta p_2
\end{bmatrix}
\approx -
\begin{bmatrix}
\delta b_{Sc} \\[1.5em]
\delta b_{Cu} \\[1.5em]
\delta b_{WP}
\end{bmatrix}
$$

$$
\begin{array}{cc}
\text{sensitivity to parameter 1} & \text{sensitivity to parameter 2}
\end{array}
$$

The number and location of regions can be chosen flexibly. They are ``samples."

Tuning might fail if a row is small (insensitive to all parameters) or if two rows are equal but the corresponding biases are different.

QuadTune emulates the parameter dependence as a linear term plus a diagonal quadratic term

We expand the emulator in a Taylor series and set it equal to the obs (Neelin et al. 2010, Bellprat et al. 2012):

$$\text{Forward Model} = \underbrace{m_{i,def}}_{\substack{\text{model}\\\text{default}}} + \sum_j \underbrace{\frac{\partial m_i}{\partial p_j}}_{\substack{\text{sensitivity}\\\text{matrix}}} \delta p_j + \underbrace{\frac{1}{2} \sum_j \frac{\partial^2 m_i}{\partial p_j^2}}_{\substack{\text{quadratic}\\\text{diagonals}}} (\delta p_j)^2$$

built by 2P+1 global simulations

$$\text{Forward Model} \approx \text{Obs} \qquad \underbrace{\delta b_i}_{\text{bias}} \equiv \underbrace{m_{i,def}}_{\substack{\text{model}\\\text{default}}} - \underbrace{m_{i,obs}}_{\substack{\text{observed}\\\text{value}}}$$

$$-\delta b_i \approx \sum_j \frac{\partial m_i}{\partial p_j} \delta p_j + \frac{1}{2} \sum_j \frac{\partial^2 m_i}{\partial p_j^2} (\delta p_j)^2$$

Nota bene: There's no intercept in this problem.

Alternatively, if an existing PPE is available, then it can be used to fit a quadratic emulator

$$-\delta b_i \approx \sum_j \frac{\partial m_i}{\partial p_j} \delta p_j + \frac{1}{2} \sum_j \frac{\partial^2 m_i}{\partial p_j^2} \left(\delta p_j\right)^2$$

Then the quadratic emulator aids interpretability of the parameter dependence. That is, it helps us understand what PPE-based tuning is doing.

The emulator serves as the basis for plots that are shown later.

Outline of talk

- We attempt to reduce the cost of tuning by use of a tuner (“QuadTune”) that uses a simple quadratic emulator

- Example tuning results from a global atmospheric model (EAM)

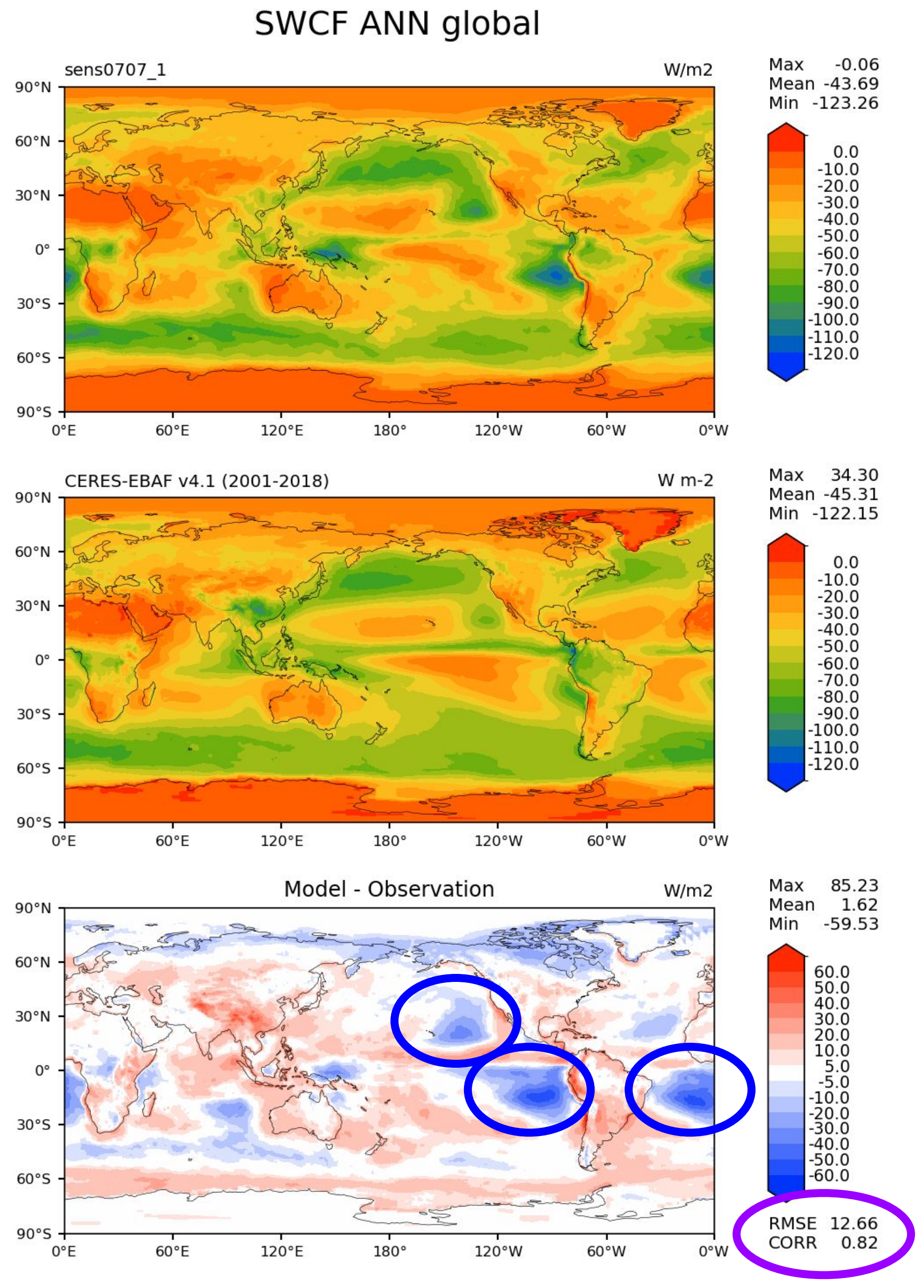
- Archetypal model errors: Tuning trade-offs and stubborn biases

Now we present an example tuning run of a global atmospheric model, EAMv~3.

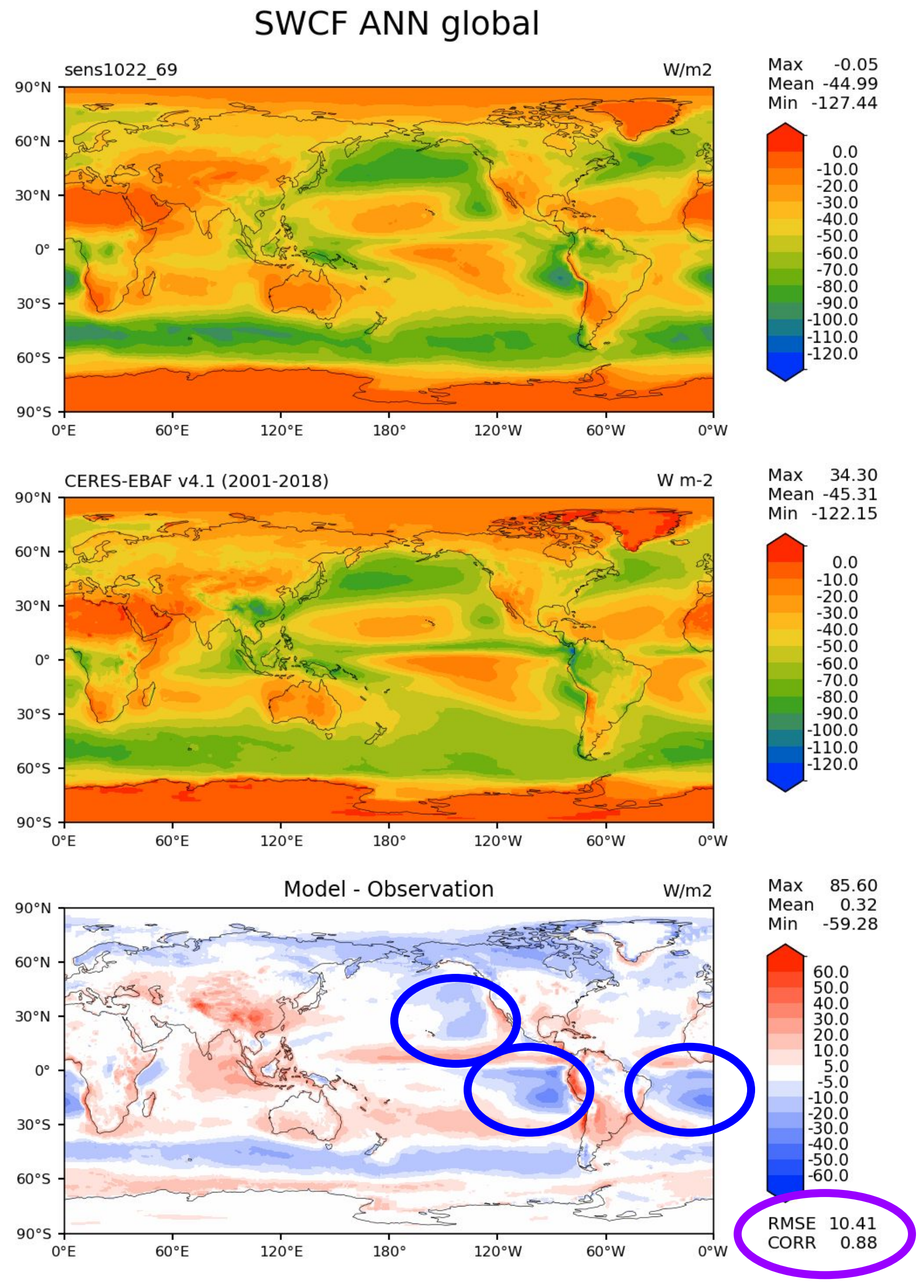
Zhun and I needed to re-tune because we introduced a new version of CLUBB ("CLUBB-taus", Guo et al. 2021).

We tune for *P=5* CLUBB parameters. Each of the *2P+1=11* runs lasts 14 months. In this example, we attempt to match SWCF in all our 20°×20° boxed regions.

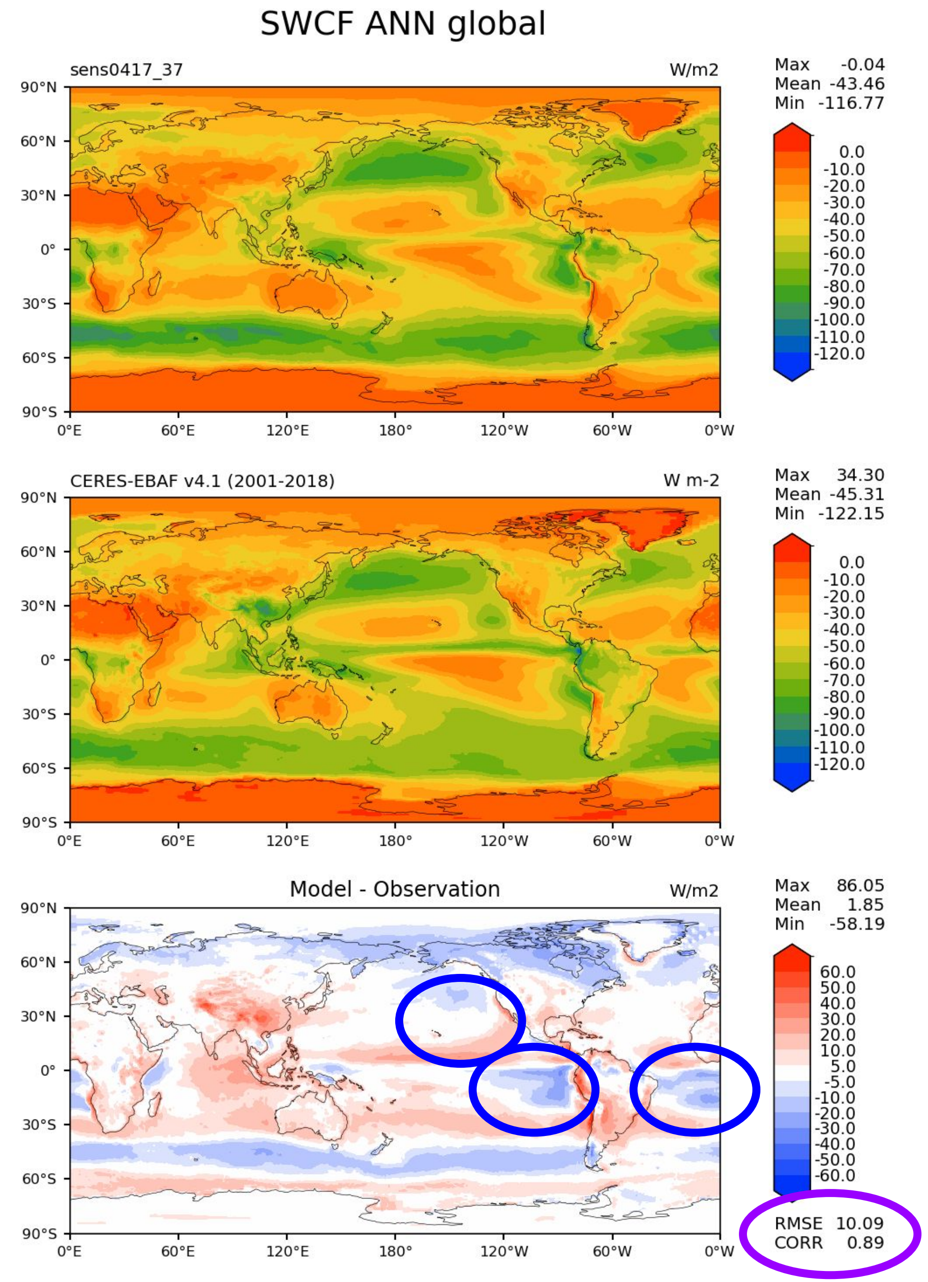
When we started, the far-coastal Sc were too bright.



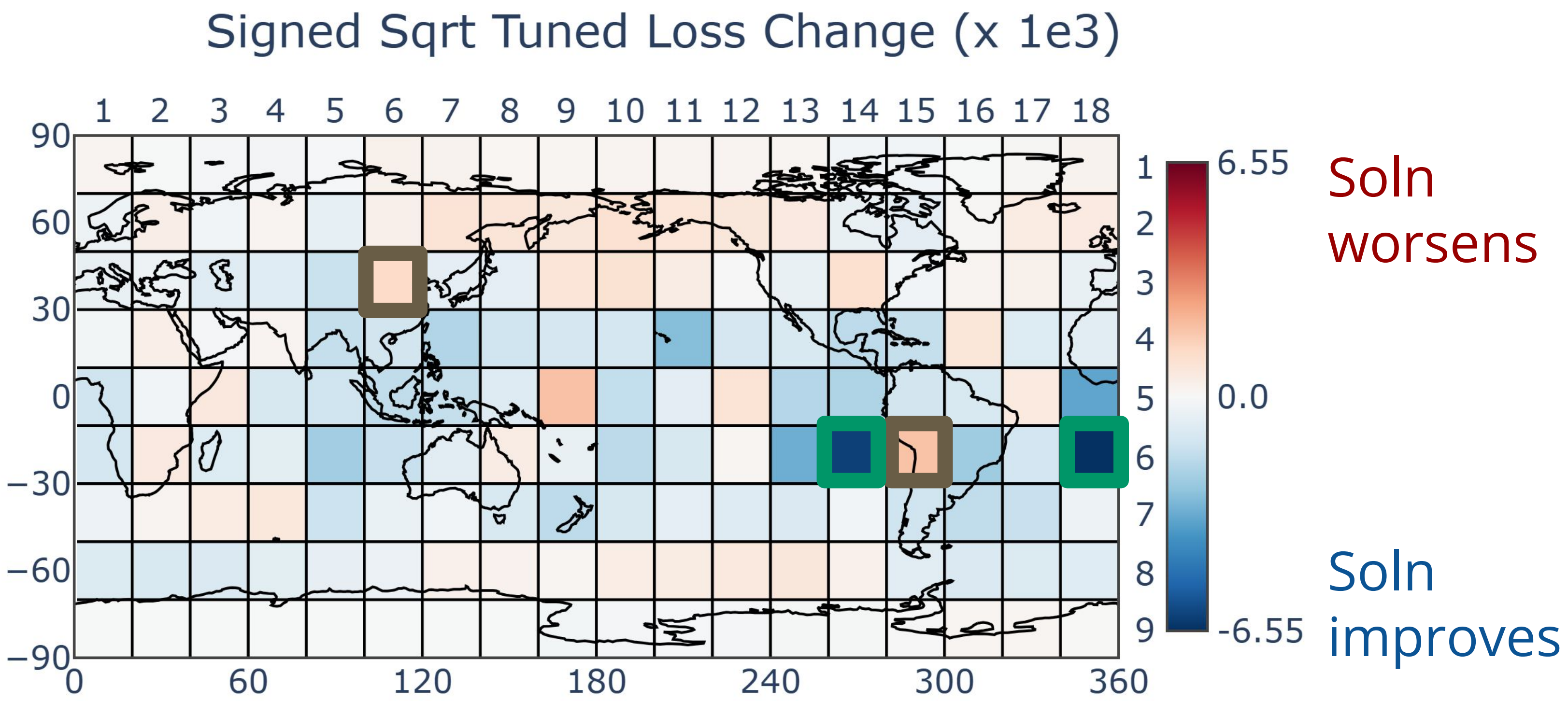
QuadTune dims the far-coastal Sc…



…but doesn't reduce the RMSE as much as Zhun's hand tuning:

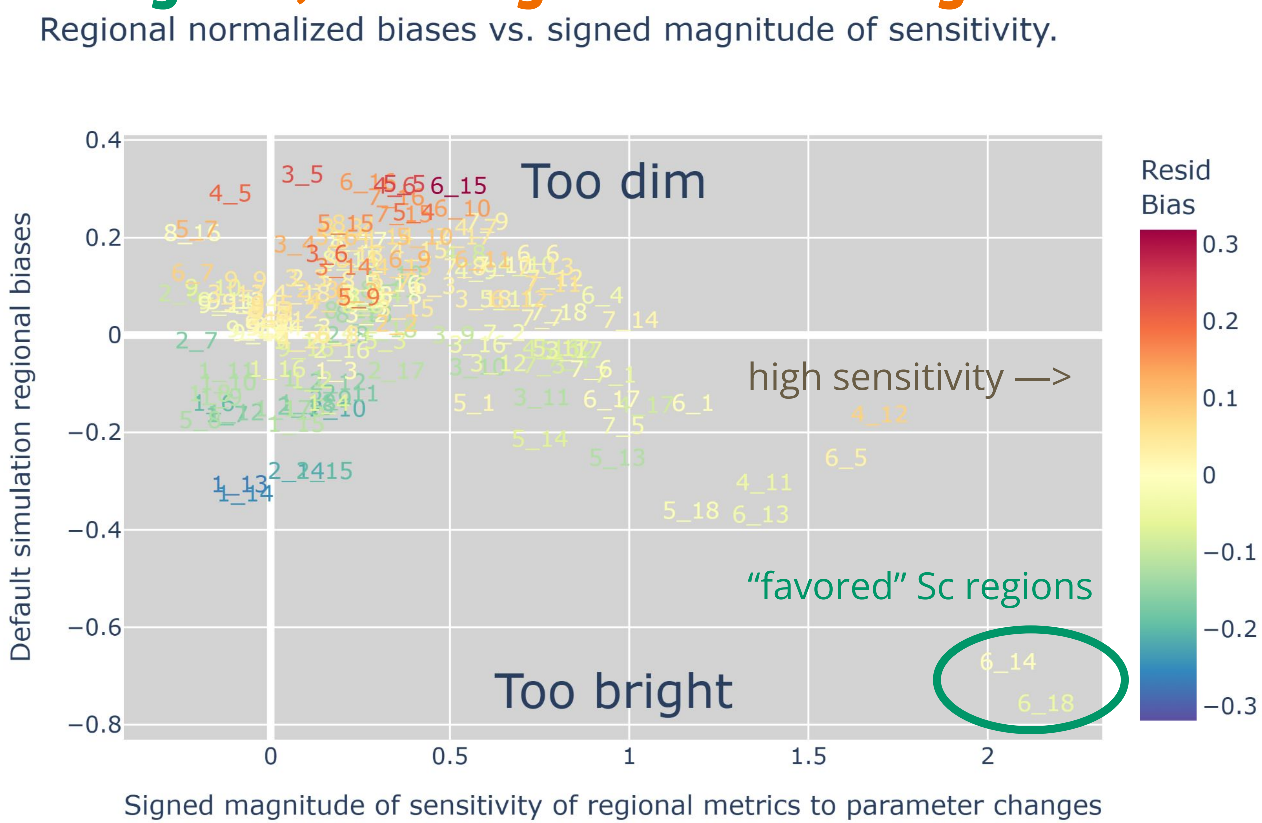


QuadTune worsens biases in the red regions. Why? QuadTune tries to address this with some diagnostics.



We'll take a look later at the brown-boxed degraded regions in China (3_6) and Bolivia (6_15) and the green-boxed improved Sc regions (6_14 and 6_18).

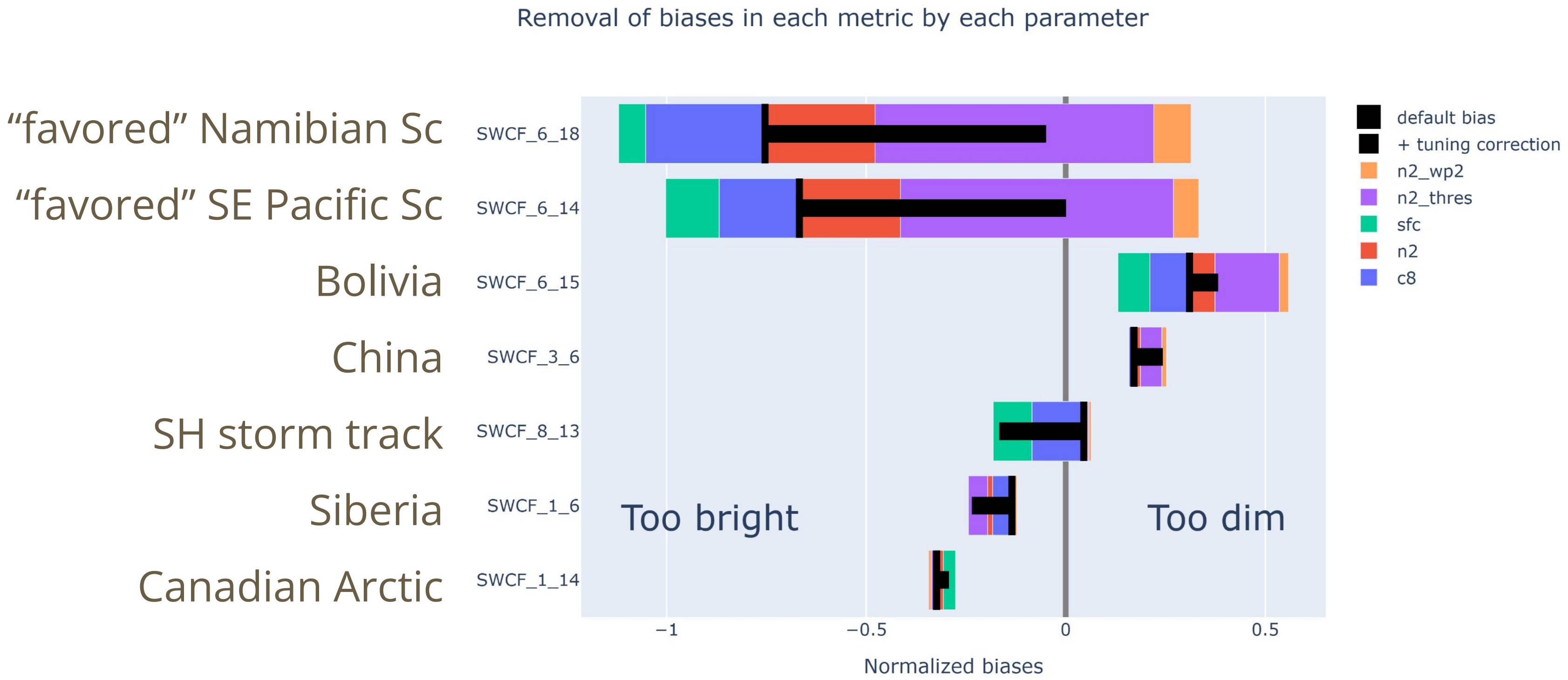
What is QuadTune doing? It removes strong biases in the sensitive Sc regions, and it ignores other regions



Yellow points have little residual bias. Red and blue points have large-magnitude residual bias.

$$\begin{bmatrix} \frac{\partial m_{Sc}}{\partial p_1} & \frac{\partial m_{Sc}}{\partial p_2} \\ \frac{\partial m_{Cu}}{\partial p_1} & \frac{\partial m_{Cu}}{\partial p_2} \\ \frac{\partial m_{WP}}{\partial p_1} & \frac{\partial m_{WP}}{\partial p_2} \end{bmatrix} \begin{bmatrix} \delta p_1 \\ \delta p_2 \end{bmatrix} \approx - \begin{bmatrix} \delta b_{Sc} \\ \delta b_{Cu} \\ \delta b_{WP} \end{bmatrix}$$

QuadTune also includes a graphical representation of the tuning matrix equation:



$$\begin{bmatrix} \frac{\partial m_{Sc}}{\partial p_1} & \frac{\partial m_{Sc}}{\partial p_2} \\ \frac{\partial m_{Cu}}{\partial p_1} & \frac{\partial m_{Cu}}{\partial p_2} \\ \frac{\partial m_{WP}}{\partial p_1} & \frac{\partial m_{WP}}{\partial p_2} \end{bmatrix} \begin{bmatrix} \delta p_1 \\ \delta p_2 \end{bmatrix} = \begin{bmatrix} \frac{\partial m_{Sc}}{\partial p_1}\delta p_1 & + & \frac{\partial m_{Sc}}{\partial p_2}\delta p_2 \\ \frac{\partial m_{Cu}}{\partial p_1}\delta p_1 & + & \frac{\partial m_{Cu}}{\partial p_2}\delta p_2 \\ \frac{\partial m_{WP}}{\partial p_1}\delta p_1 & + & \frac{\partial m_{WP}}{\partial p_2}\delta p_2 \end{bmatrix} \approx - \begin{bmatrix} \delta b_{Sc} \\ \delta b_{Cu} \\ \delta b_{WP} \end{bmatrix}$$

Outline of talk

- We attempt to reduce the cost of tuning by use of a tuner (“QuadTune”) that uses a simple quadratic emulator
- Example tuning results from a global atmospheric model (EAM)
- Archetypal model errors: Tuning trade-offs and stubborn biases

Tuning trade-off: Bolivia (6_15) has a positively correlated sensitivity with favored Sc regions, but "wrong" bias

Bolivia (6_15) has a similar sensitivity to all the parameters as do the stratocumulus regions (6_14 or 6_18). However, whereas the Sc are too bright, Bolivia is too dim.

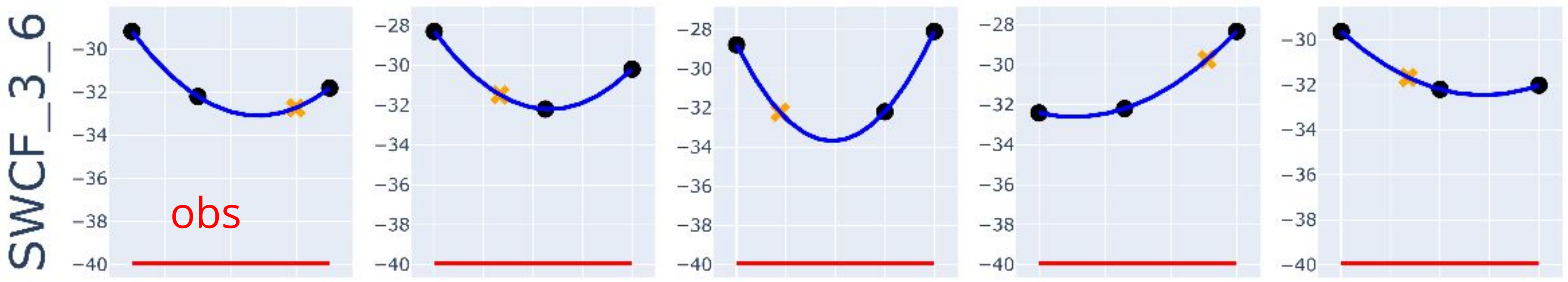
Therefore, in whatever way QuadTune adjusts the parameter values, improving Sc will necessarily worsen Bolivia.

Some regional biases are not the result of tuning trade-offs. They're just local biases.

They can't be improved regardless of how we treat other regions.

Nonlinear Zugzwang: (China, 3_6)

For region 3_6, the dependence of SWCF on each parameter is parabolic, and each parabola curves away from the observed value of SWCF. Hence the default parameter value is the best possible value.

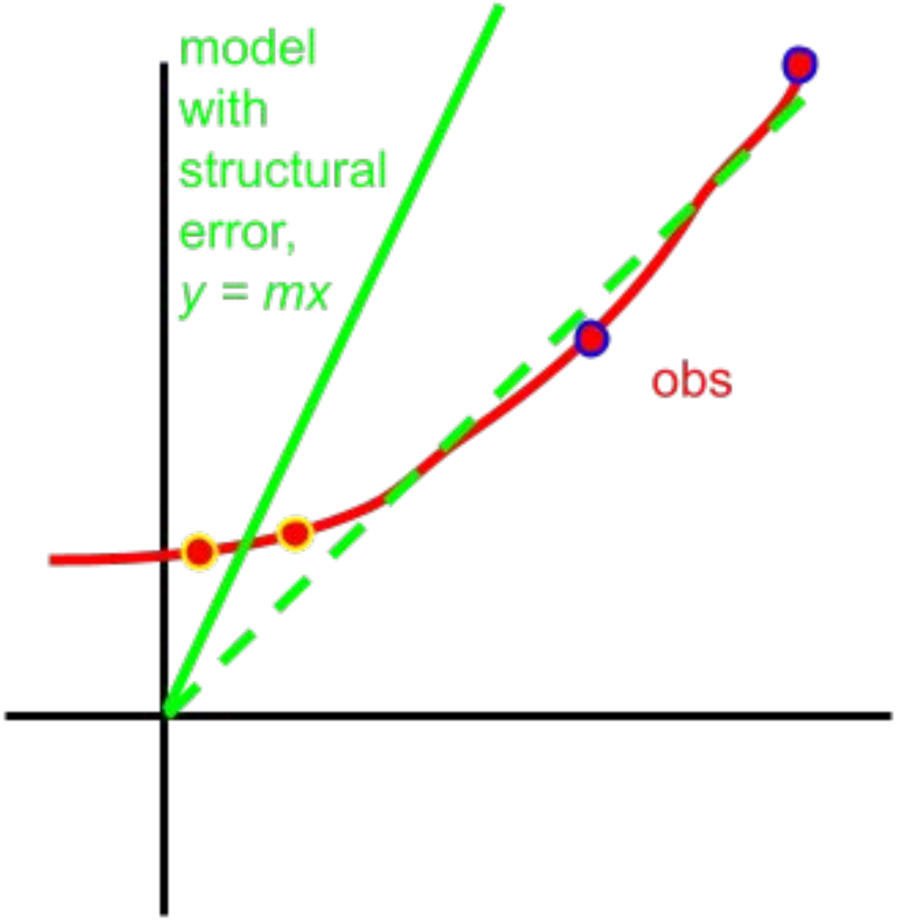


parameter values →

QuadTune can also generate an ensemble of parameter sets

We want the spread in parameter sets to be related to model structural error.

To do that, QuadTune bootstrap-samples the regions. That is, it samples the regions with replacement, so that some regions are double counted and others are omitted.

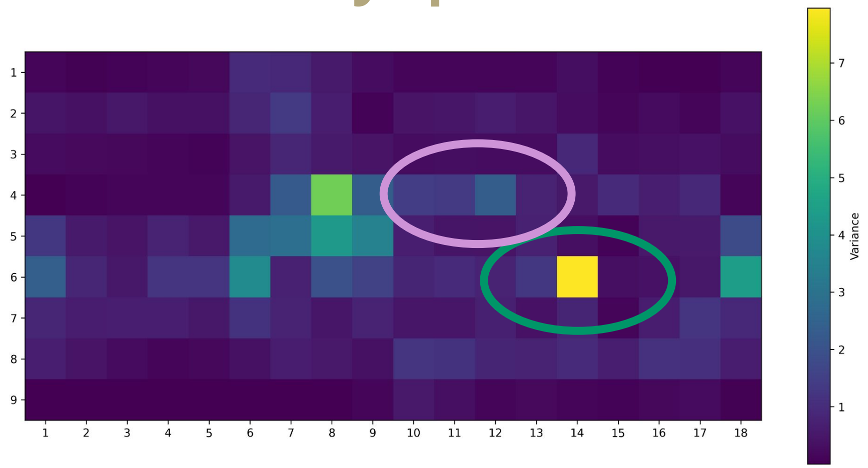


The bootstrap parameter ensemble can be used to estimate uncertainty in, e.g., climate sensitivity

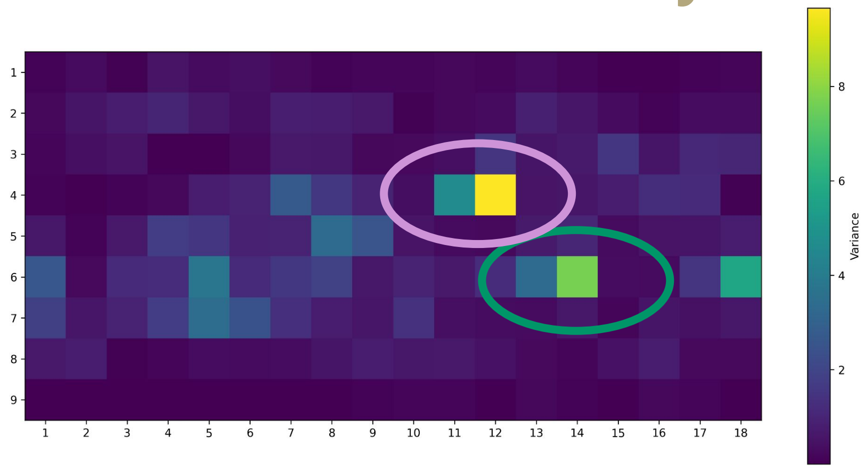
The parameter ensemble can be combined with future climate projections (actually, just SST+4K) to yield an uncertainty estimate.

In this example, the uncertainty increases in the DYCOMS (California) Sc and decreases in the VOCALS (SE Pac) Sc.

Present-day spread



Future uncertainty



What can we learn from QuadTune?

- We learn which regional biases involve trade-offs with other regions, and which regions have stubborn biases.

- We learn when to give up! If the tuner doesn't yield acceptable results, then we should either 1) find new parameters or 2) re-formulate the model structure.

- We can estimate uncertainty with a bootstrap ensemble.

Thanks for your time!