Impediments to model tuning: Stubborn biases, tuning trade-offs, and nonlinear parameter dependence

Vincent Larson, Zhun Guo, Ben Stephens

AMWG Winter Meeting

5 Feb 2025

Our problem: When a new (better?) parameterization is introduced into a global atmospheric model, typically the results get worse (!)

Why? Because there are compensating errors among the former suite of parameterizations. Those compensations are disrupted when a former parameterization is removed.

When a new parameterization worsens results, we want to…

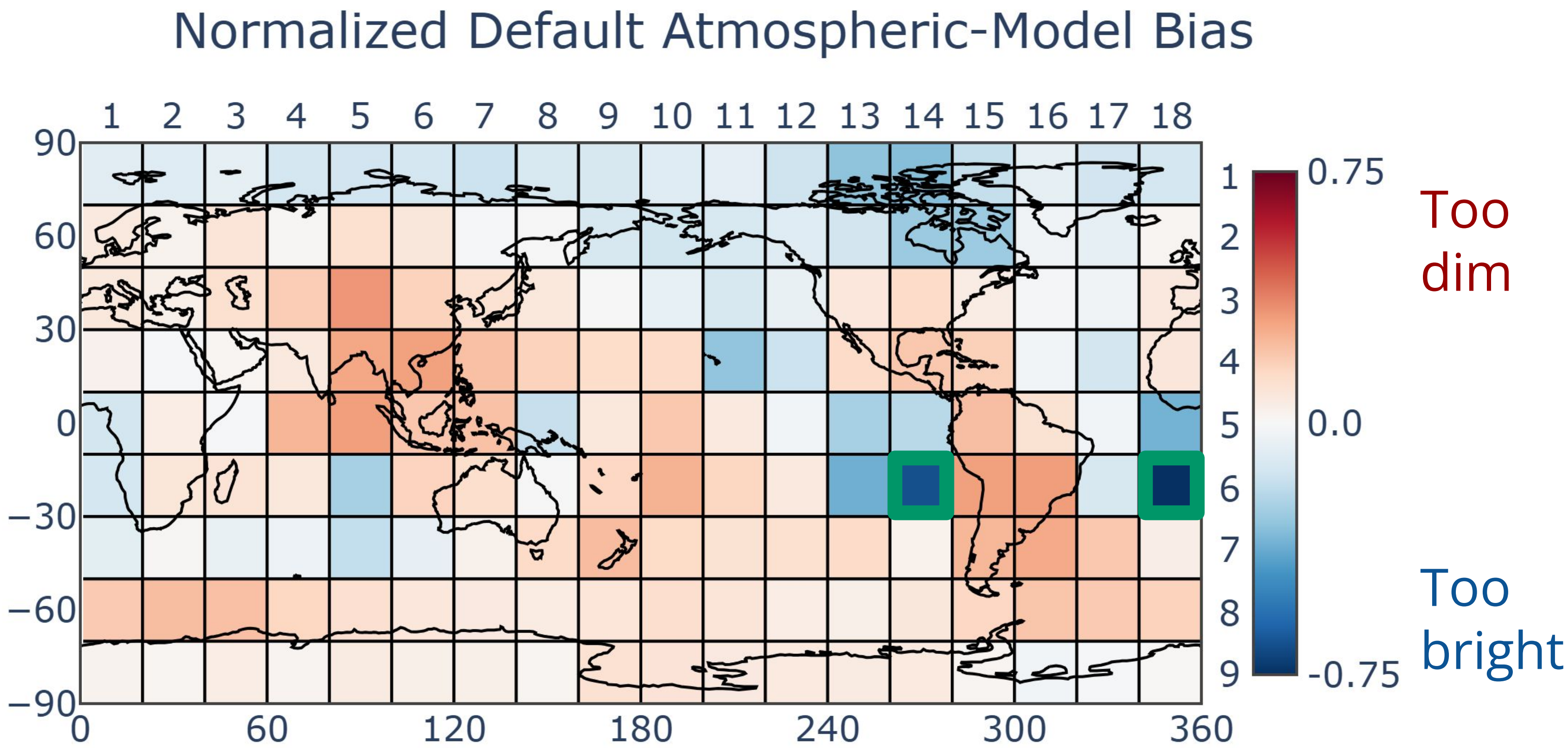
- Retune, in order to see if the parameterization looks promising; and
- Get hints about what part of the model structure to change next, i.e. which model equations are still wrong.

We want to re-tune as quickly as possible, so that we can resume working on the model structure. That is where we'll realize the big gains in accuracy.

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)

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)

Ideally, we would like to run all the *2P+1* global simulations overnight. (QuadTune is a poor man's tuner!)

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

stratocumulus region

cumulus region

warm pool region

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

s

sensitivity to parameter 1

sensitivity to parameter 2

Tuning 2 parameters can't remove the bias in all 3 regions unless the spatial pattern of sensitivity happens to be consistent with the spatial pattern of bias.

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

$$\text{Forward Model} \approx \text{Obs}$$

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

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

To reduce the number of global model simulations needed, QuadTune neglects parameter interactions

Neglecting parameter interactions, we need to run only *2P+1* global simulations, where *P* is the number of tunable parameters.

Neglecting parameter interactions is expected to lead to ~15% error.

Outline of talk

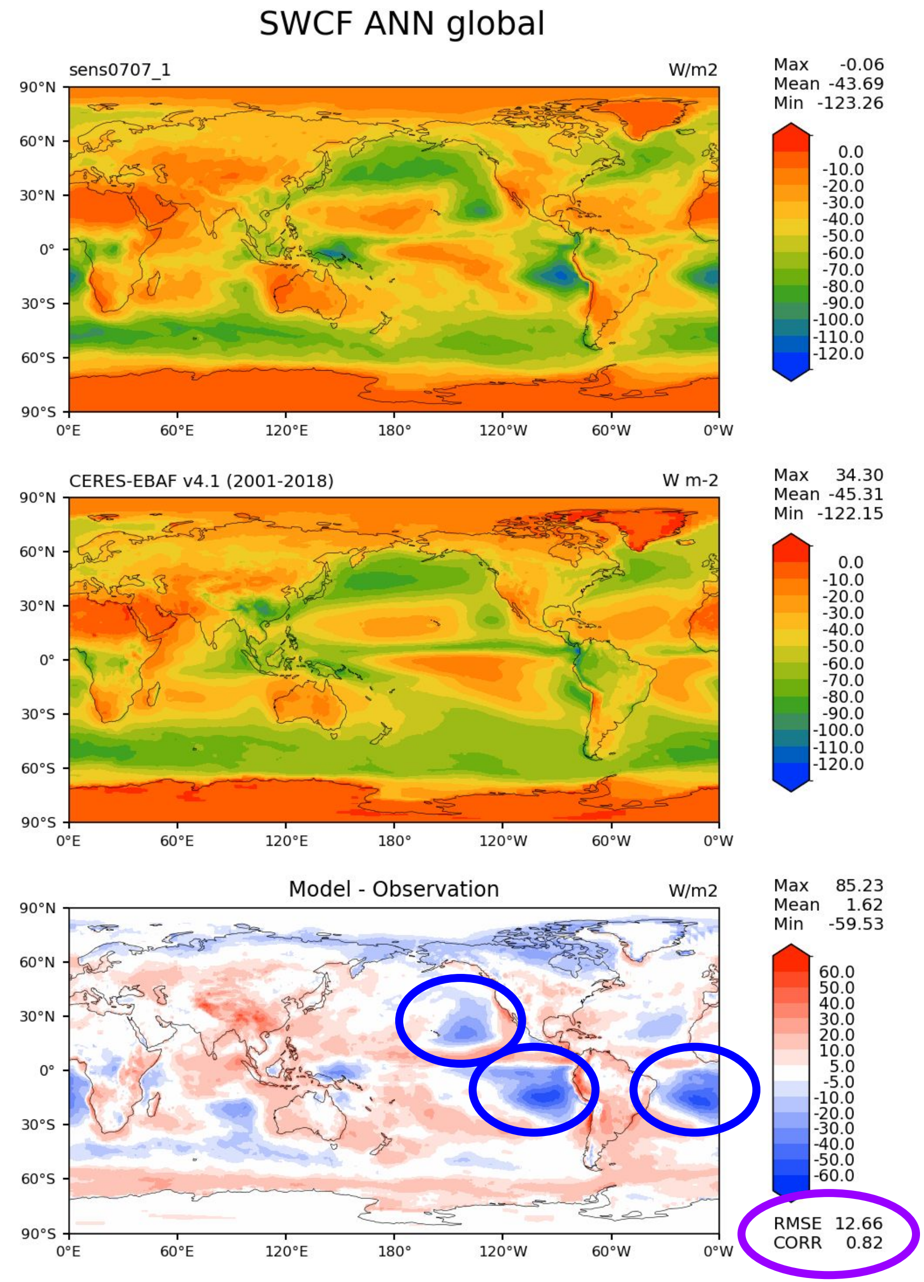
- We attempt to reduce the cost of tuning by use of a tuner (“QuadTune”) that uses a simple quadratic emulator
- Example tuning results from a global atmospheric model (EAM)
- Archetypal model errors: Tuning trade-offs and stubborn biases

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

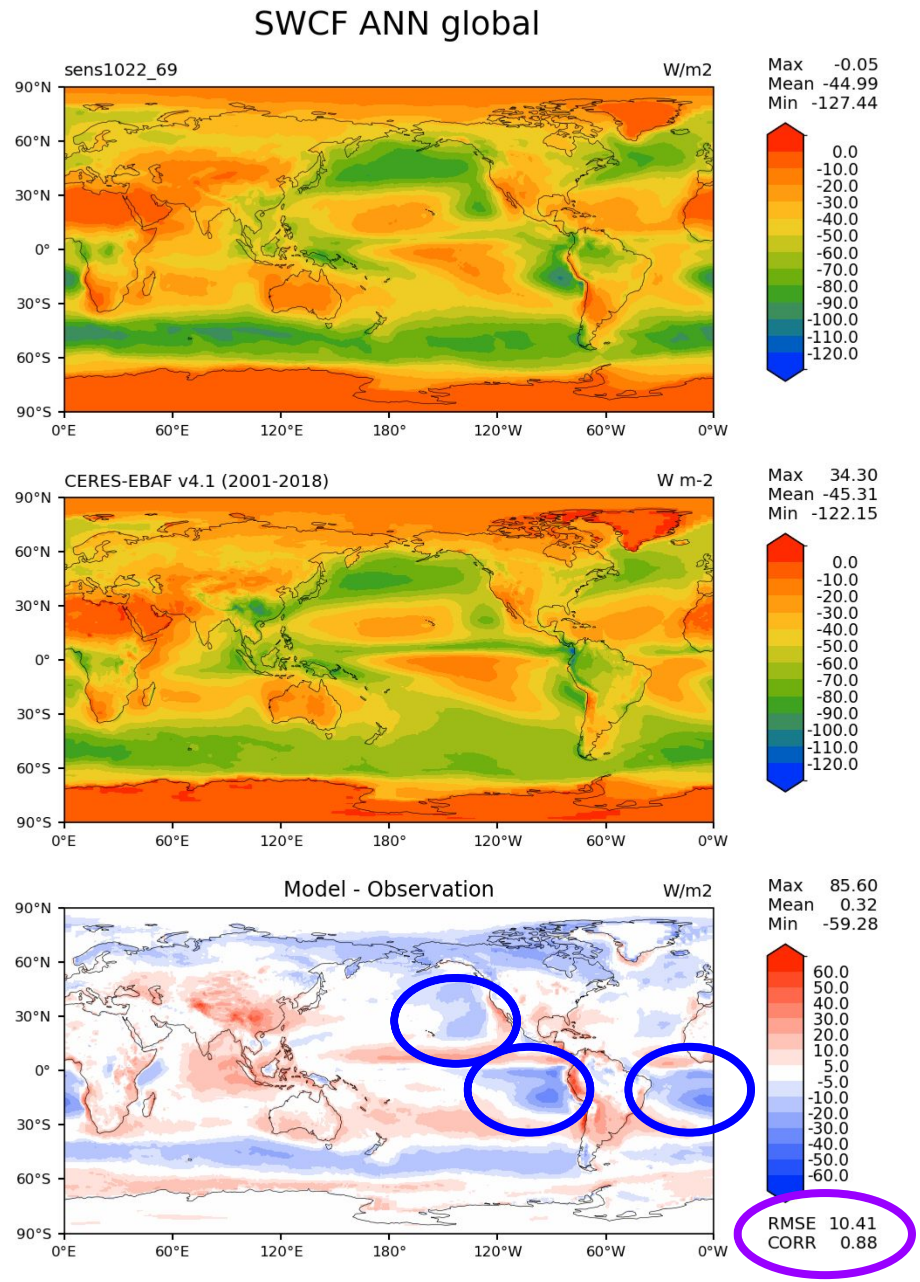
Zhun and I needed to re-tune because we introduced a new version of CLUBB (“CLUBB-taus”, Guo et al. 2021).

We tune for 5 CLUBB parameters. Each of the 11 runs lasts 14 months. In this example, we attempt to match SWCF in all our boxed regions.

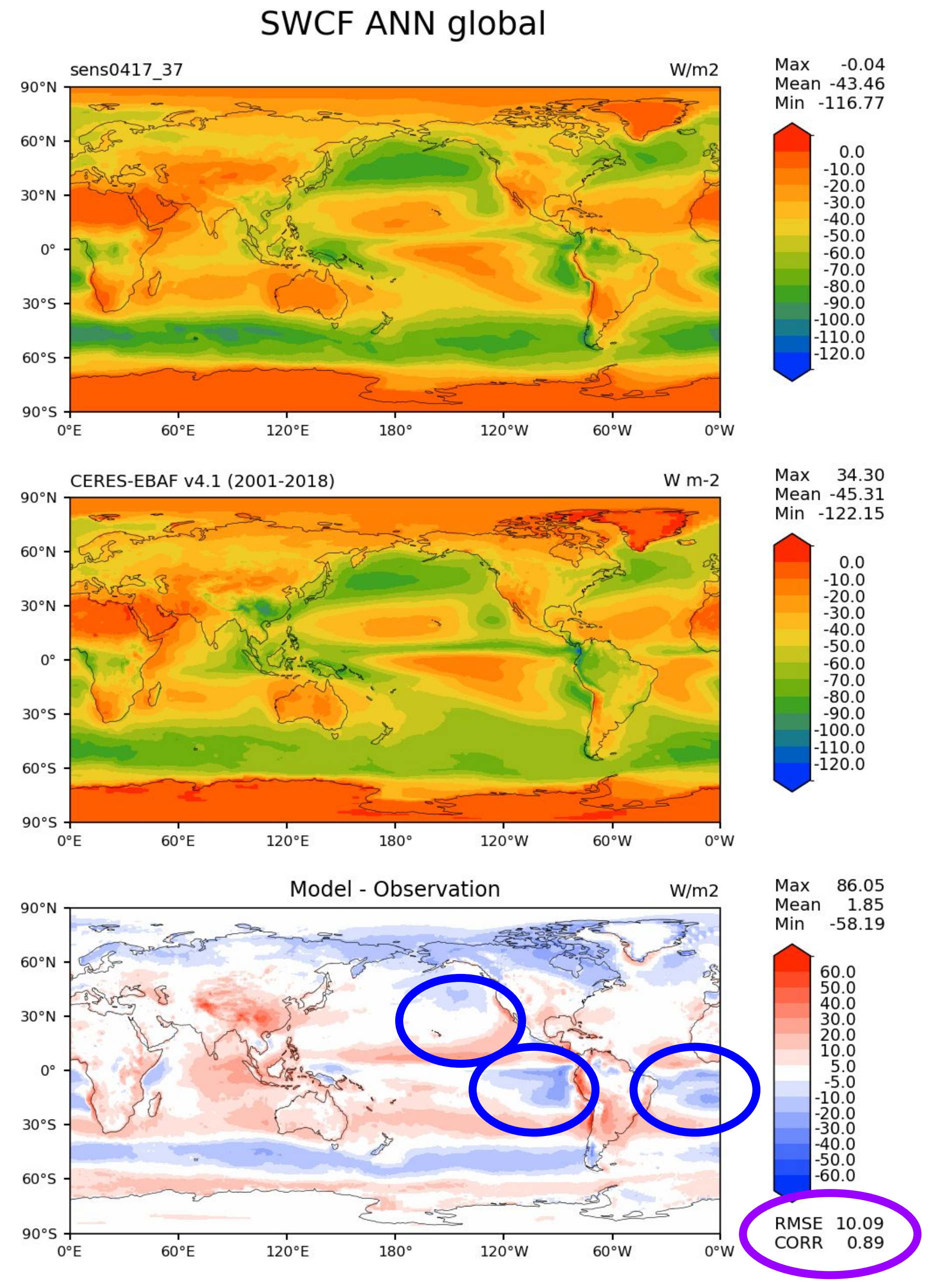
When we started, the far-coastal Sc were too bright.



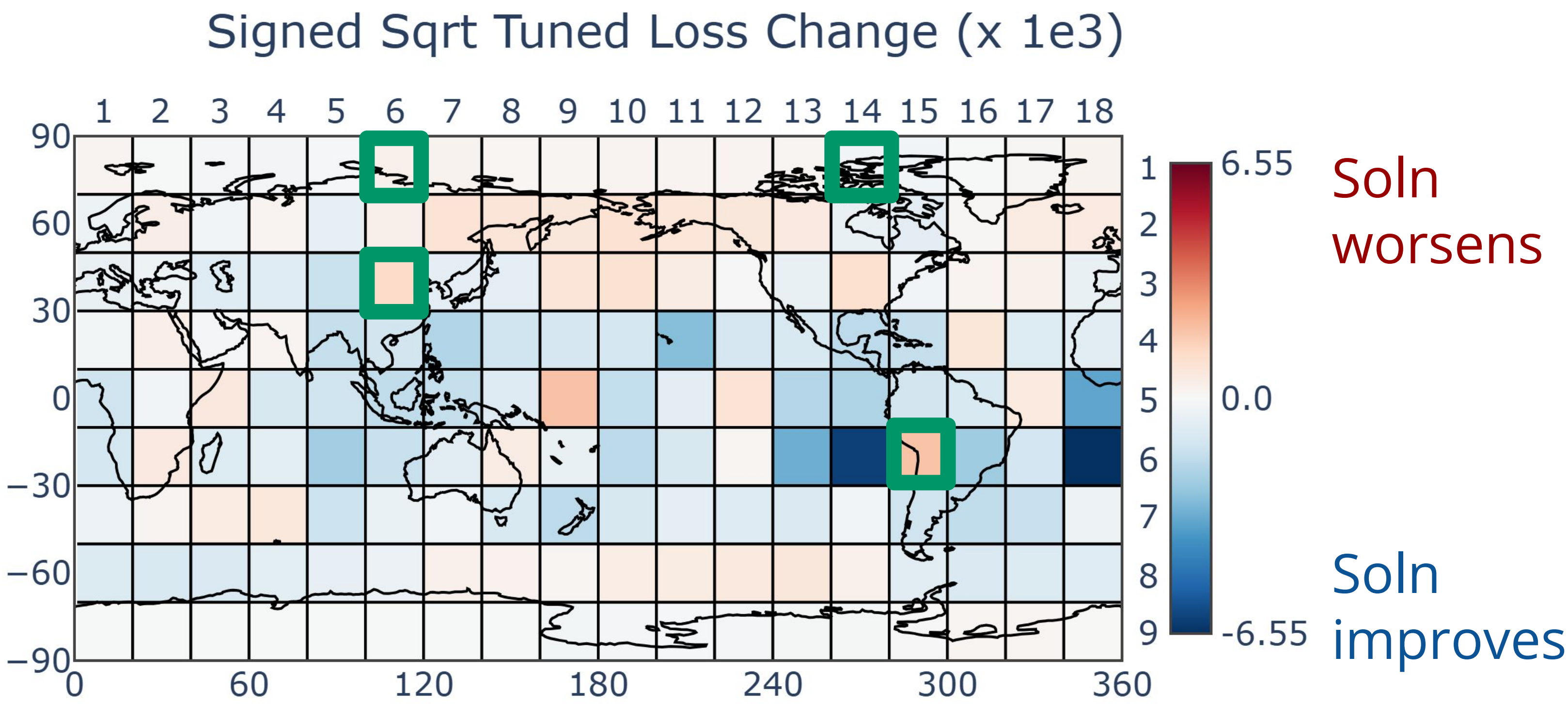
QuadTune dims the far-coastal Sc…



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

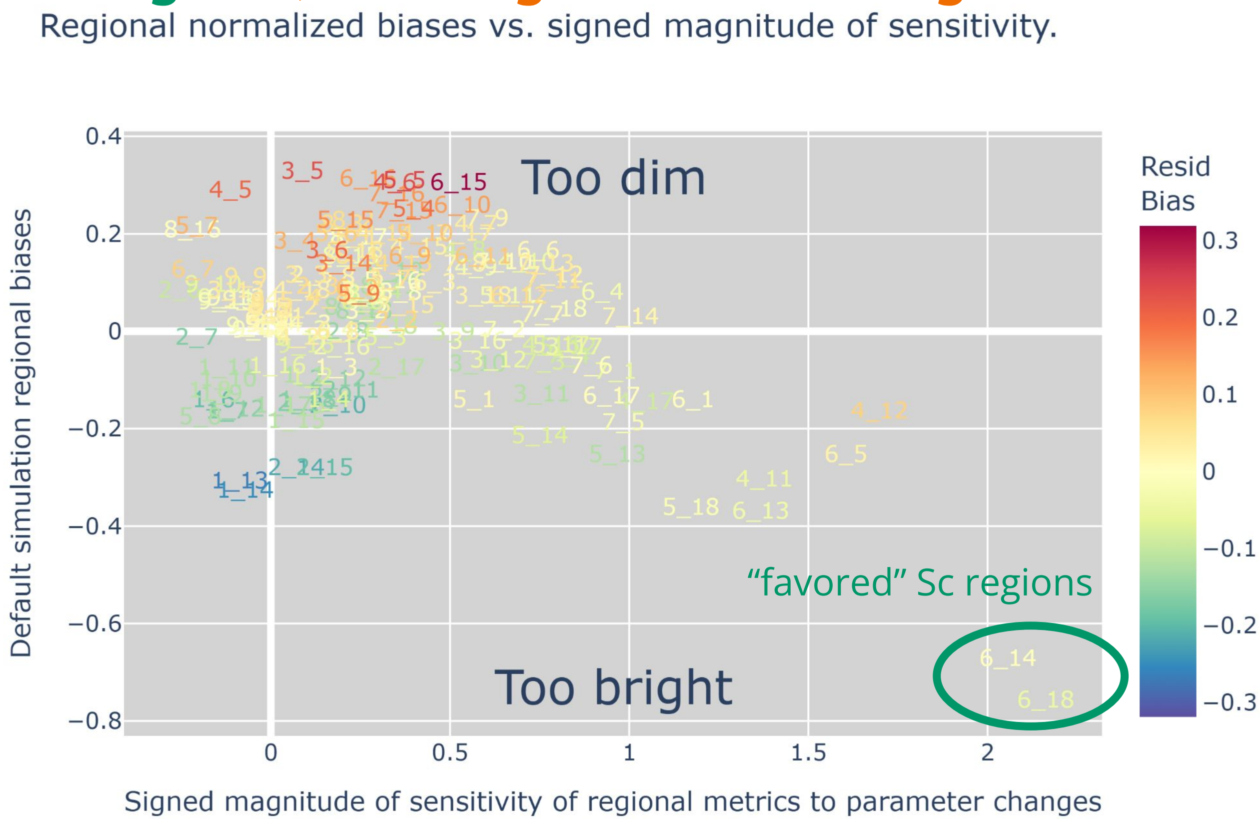


QuadTune worsens biases in the red regions. Why? QuadTune tries to address this with some diagnostics.



We'll take a look later at the green-boxed regions.

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's diagnostics include a "3-dot" plot showing how each regional metric is influenced by parameter perturbations

Siberia

Canadian Arctic

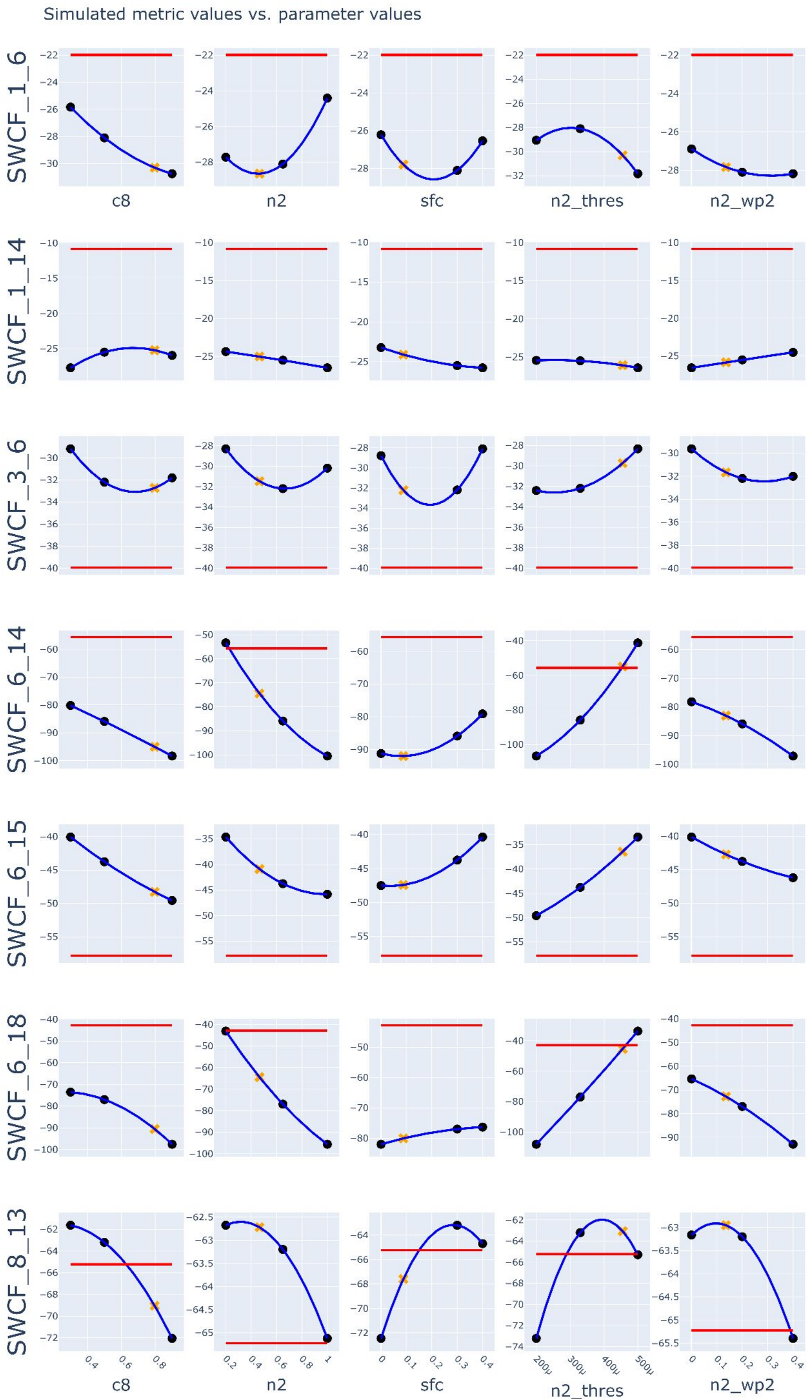
China

"favored" SE Pacific Sc

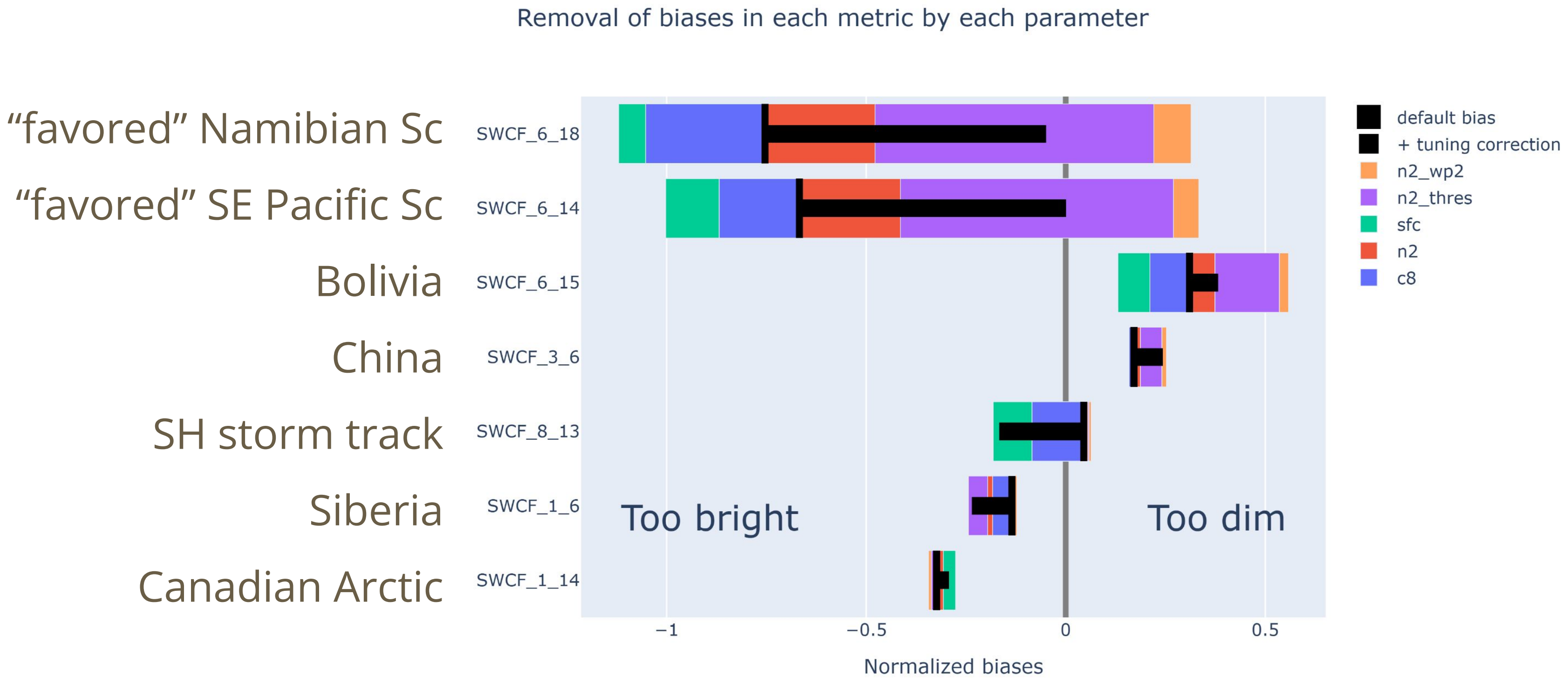
Bolivia

"favored" Namibian Sc

SH storm track



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 1: 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.

Tuning trade-off 2: Siberia (1_6) has a “correct-sign" bias but the “wrong” sensitivity

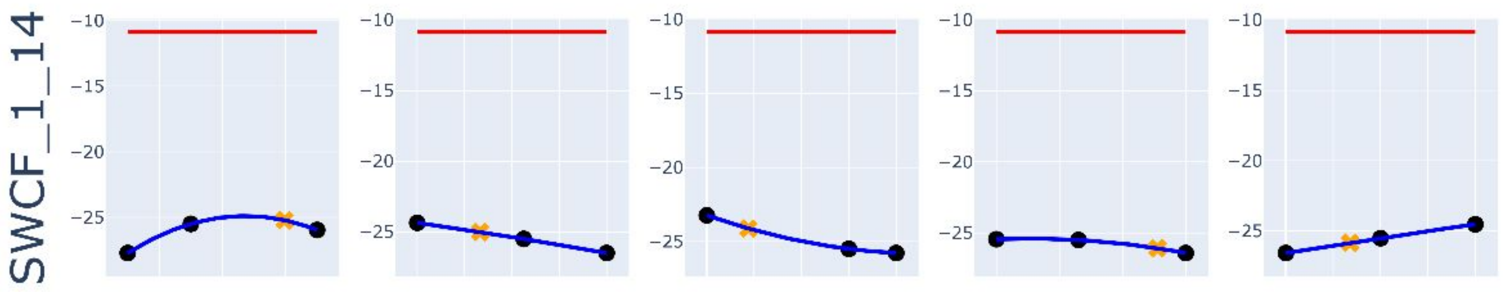
Siberia (1_6) has the same-sign bias as the Sc regions (too bright), but its response to, for instance, parameter n2_thresh has the opposite sign. The reason is that 1_6’s response to n2_thresh is strongly nonlinear.

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.

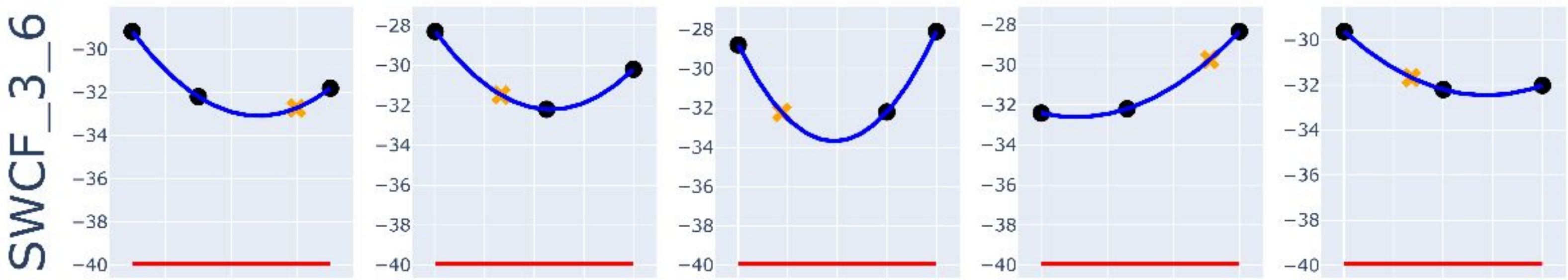
Stubborn bias: (Canadian Arctic, 1_14)

Region 1_14 has a non-negligible bias but has little sensitivity to any parameter. The large bias and small sensitivity means that 1_14 resides on or near the y-axis of the bias-sensitivity scatterplot, and far from the x-axis, which has zero bias.



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.



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.

Thanks for your time!