Data-Driven Probabilistic Air-Sea Flux Parameterization

Jiarong Wu, New York University

Pavel Perezhogin (NYU), David John Gagne (NCAR), Brandon Reichl (GFDL), Aneesh C. Subramanian (CU Boulder), Elizabeth Thompson (NOAA PSL), Laure Zanna (NYU)

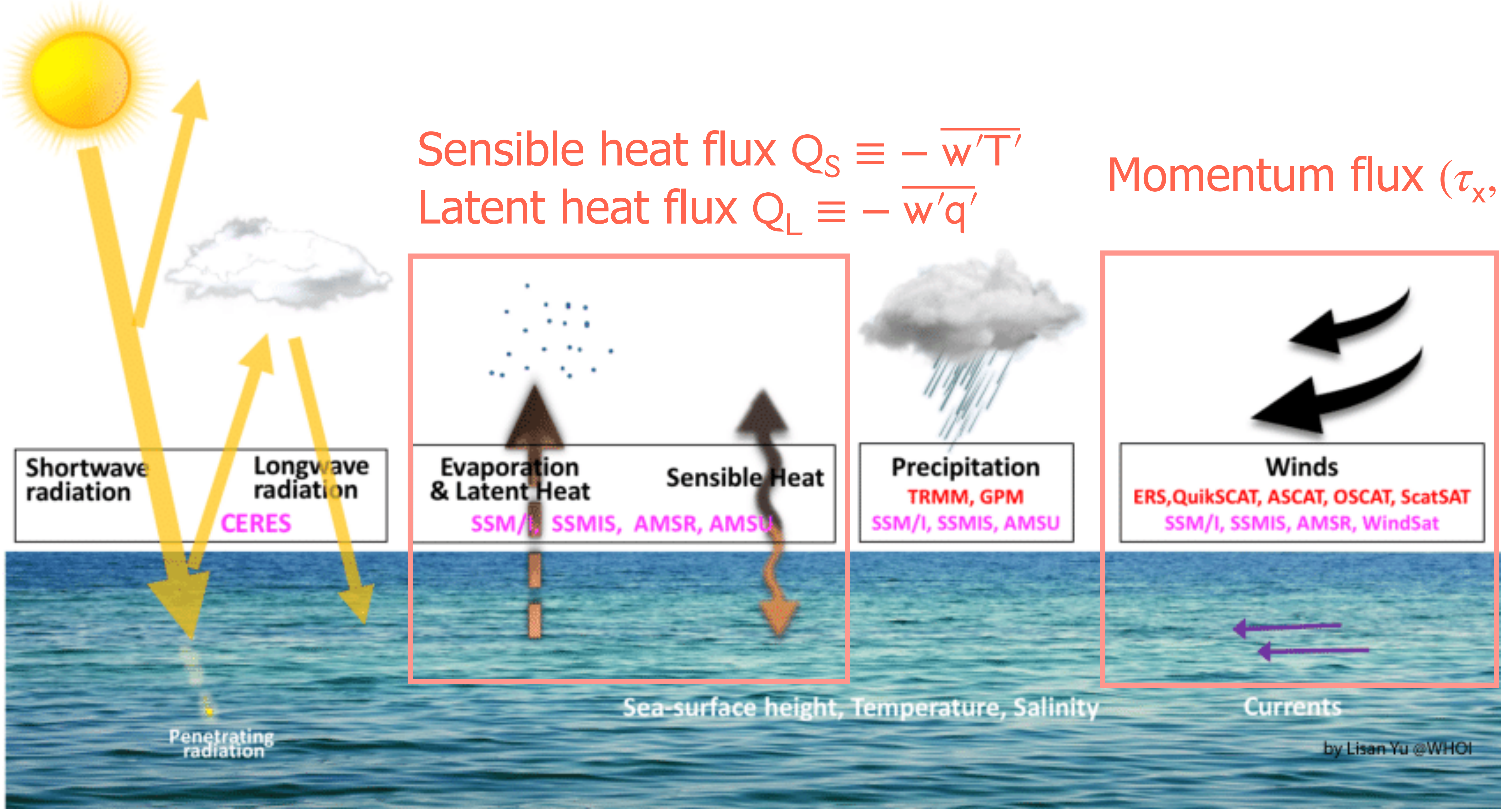
CESM Ocean Model Working Group Meeting 2025/02/27

Air-sea fluxes and their representation

Sensible heat flux $Q_S \equiv -\overline{w'T'}$

Latent heat flux $Q_L \equiv -\overline{w'q'}$

Momentum flux $(\tau_x, \tau_y) \equiv (-\overline{w'u'}, -\overline{w'v'})$



State variables:

Wind speed U_a

Air temp. T_a and humidity q_a

SST T_o

Current speed U_o

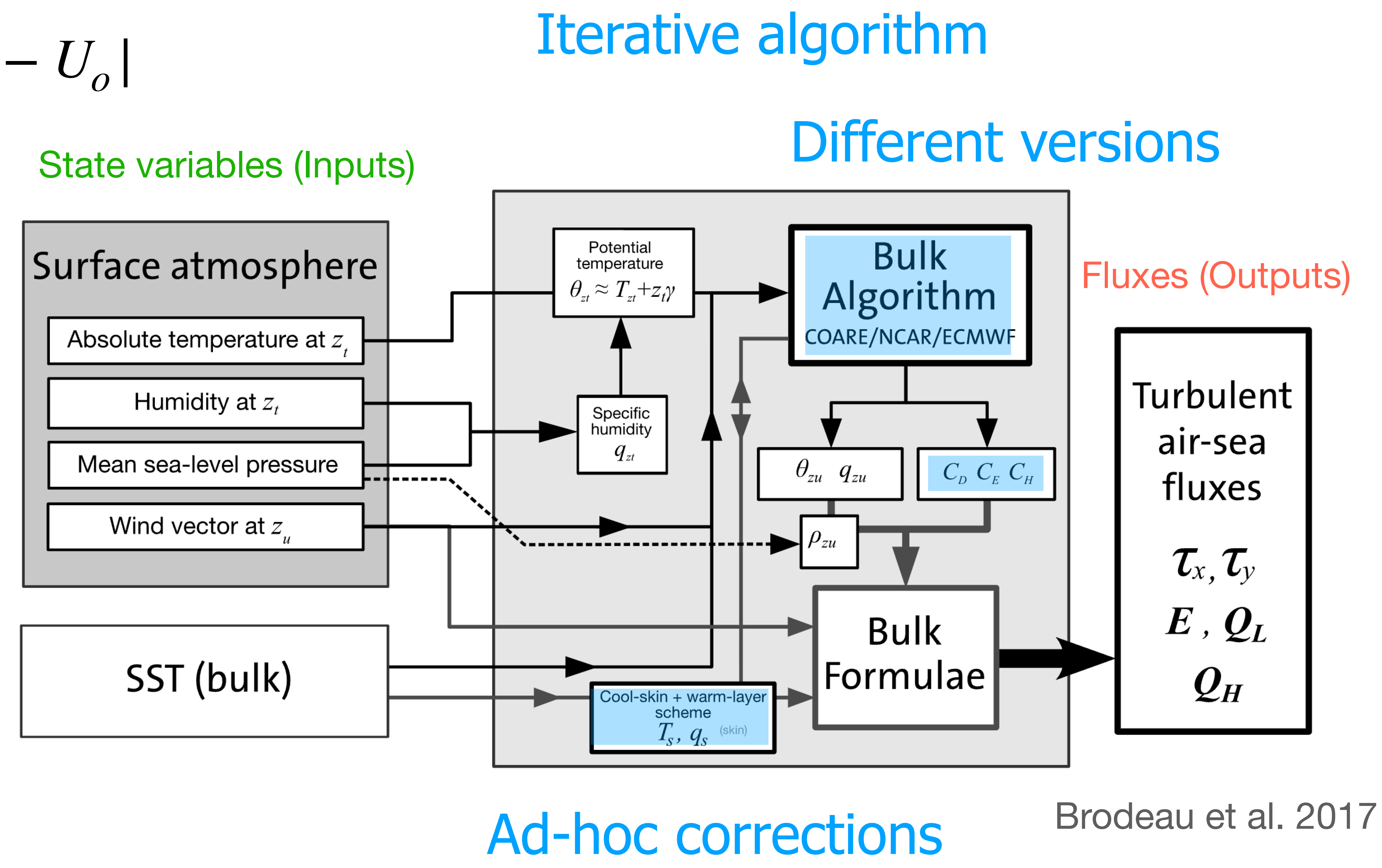
…

We need air-sea flux model in:

- Coupled GCM: prognostic variables -> fluxes as boundary conditions
- Flux products (forced GCM): observables (in-situ or satellite) -> fluxes (hard to observe)

State-of-the-art air-sea flux parameterization: bulk algorithm

- Bulk algorithms
  - Along-wind stress $\tau_x = \rho_a C_D S(U_a - U_o)$ $\quad S = |U_a - U_o|$
  - Cross-wind stress $\tau_y = 0$
  - Sensible heat flux $Q_S = \rho_a c_p C_H S(T_a - T_o)$
  - Latent heat flux $Q_L = \rho_a L_e C_E S(q_a - q_s)$
  - Use of "bulk" variables to model the surface layer
  - Physics-based (Monin-Obukhov similarity theory) + empirically fitted parameters

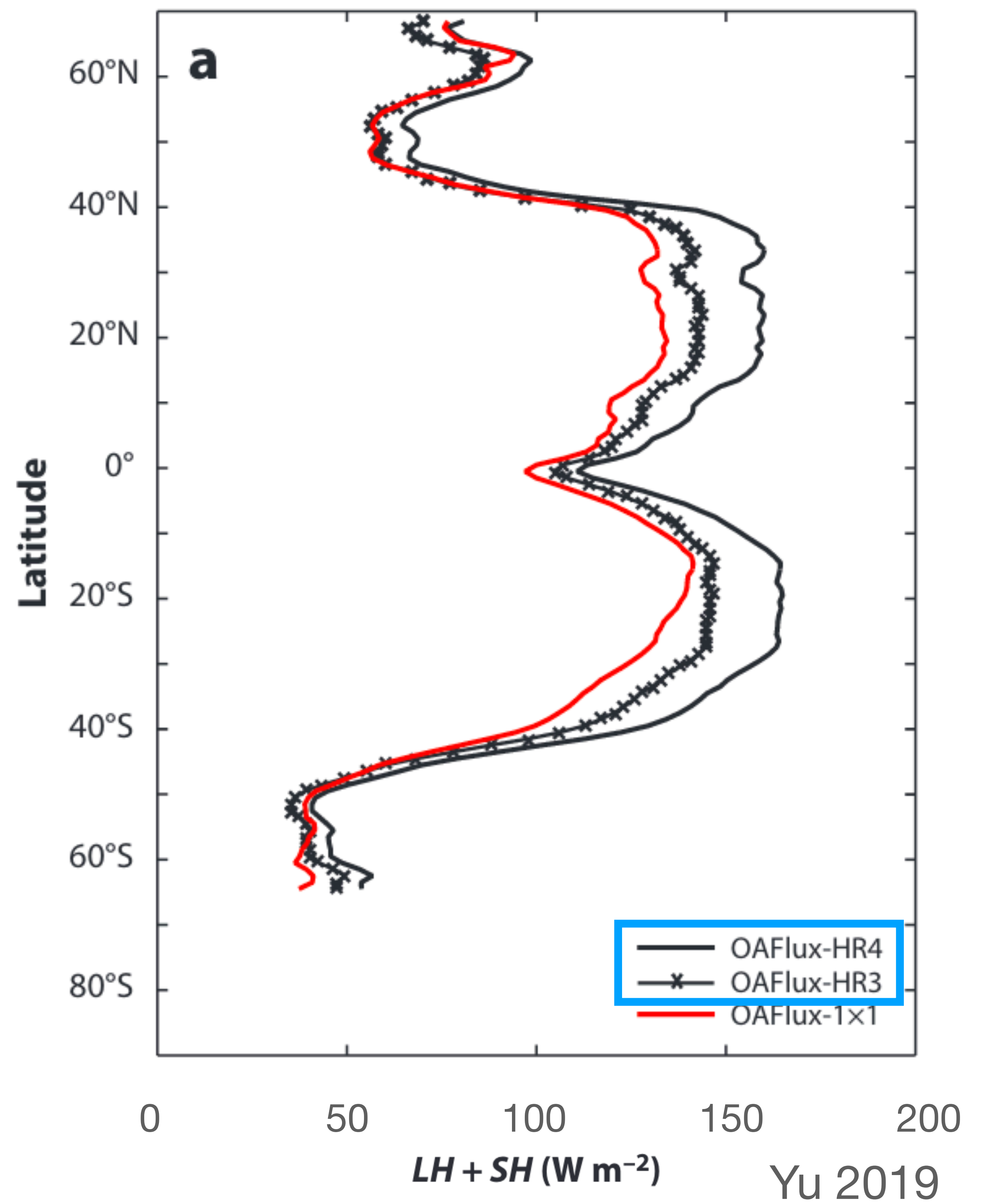


Brodeau et al. 2017

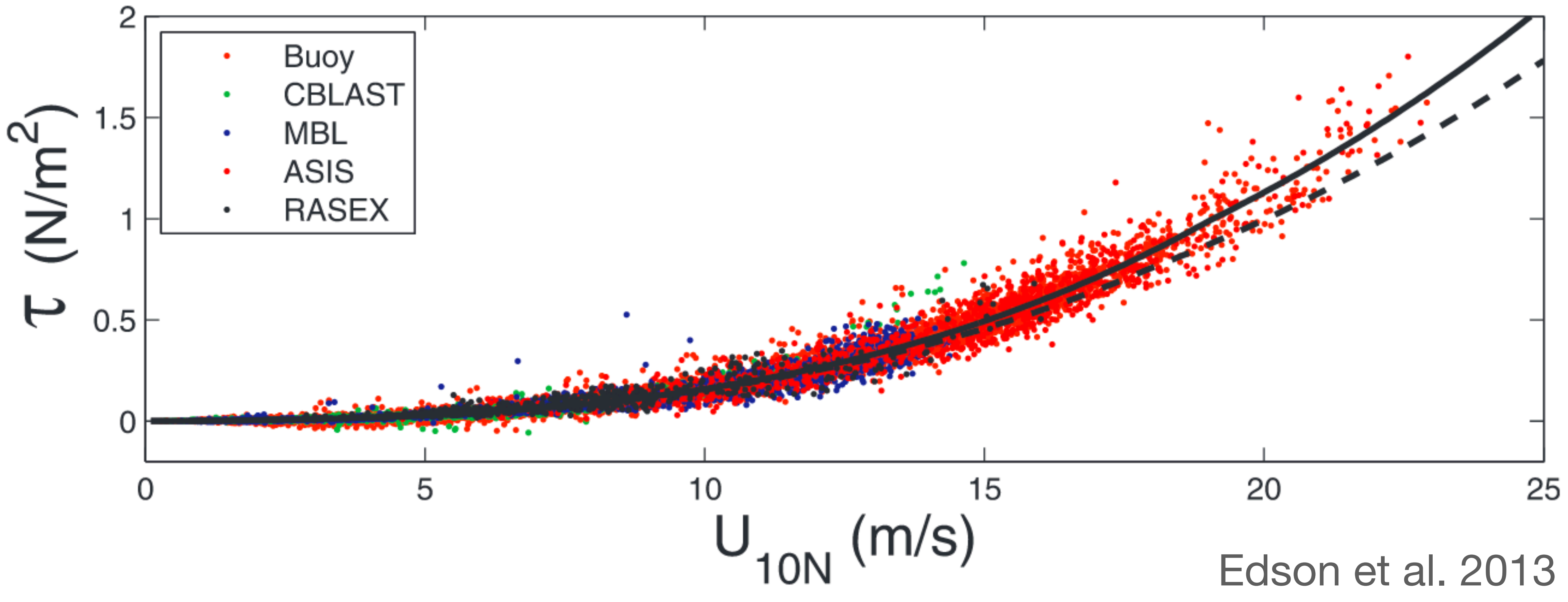
Current issues with air-sea flux modeling

- Accurate representation of air-sea fluxes across scales is challenge (for both observation and modeling).
- Using different algorithms has a considerable impact on flux estimation
  - Sensitivity studies: general circulation (Polichtchouk and Shepherd 2016), precipitation (Harrop et al. 2018), MJO (Hsu et al. 2022), SST (Bonino et al. 2022)…
- Bulk algorithms are designed to represent the mean of flux given the input state variables
  - Additional inputs: e.g. sea-state (Sauvage et al. 2023, etc.)
  - Stochastic parameterization (Williams 2012, Berner et al. 2017, etc.)

Discrepancy of Zonally averaged turbulent heat flux (annual mean)



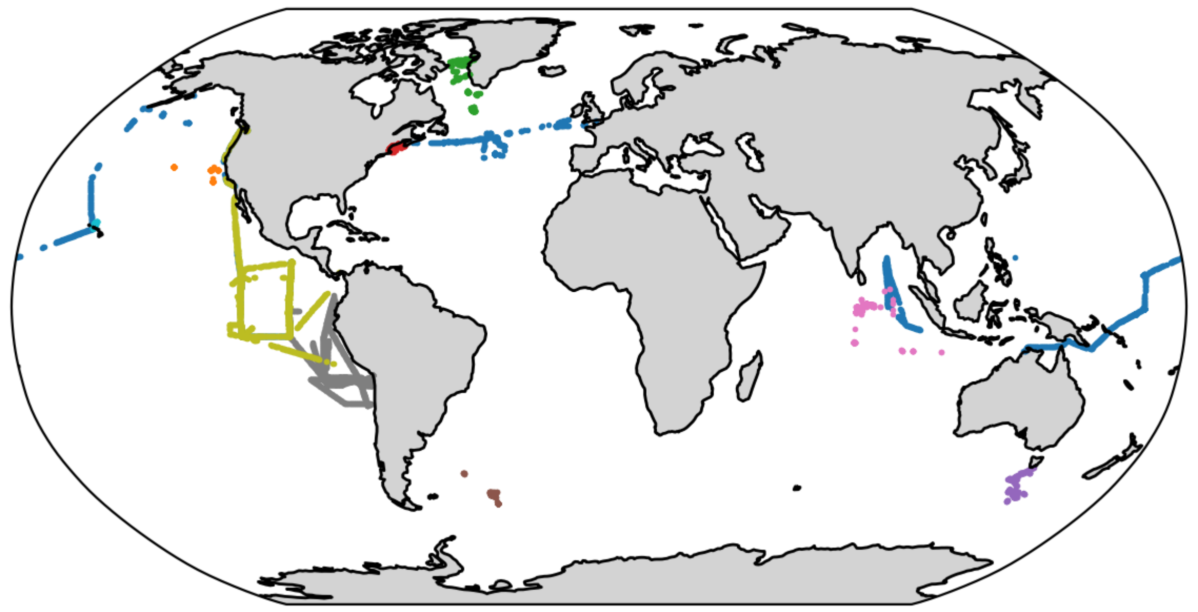
Yu 2019



Edson et al. 2013

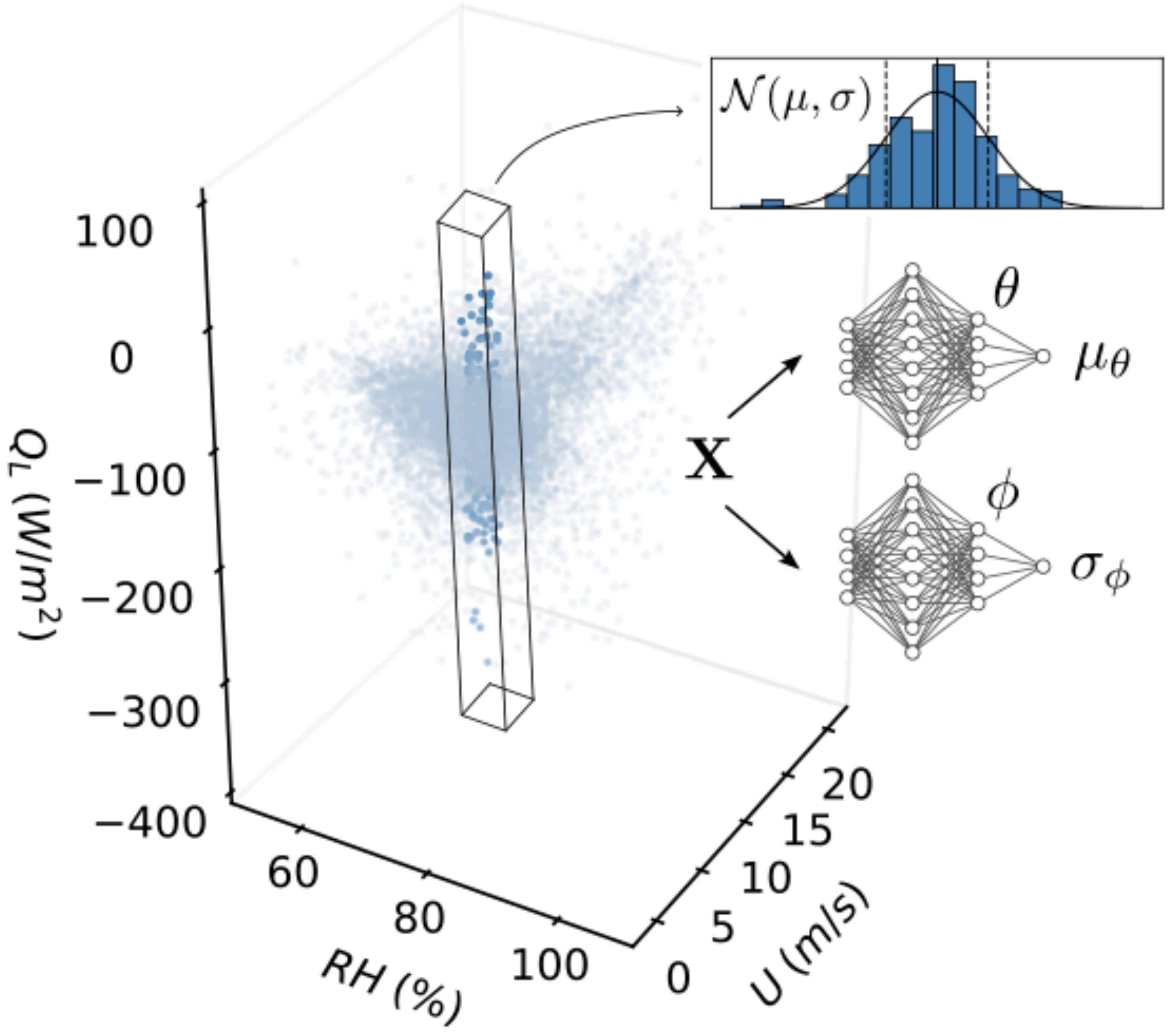
Data-driven alternative for air-sea flux modeling

- Data (NOAA PSL)
  - 10,000 samples from R/V
  - Hourly-averaged eddy-covariance $\overline{w'u'}, \overline{w'v'}, \overline{w'T'}, \overline{w'q'}$
  - No high-fidelity numerical simulations yet :(



- Method
  - Directly predict fluxes
  - Parametric distribution conditioned on inputs
  - Estimate of distribution parameters (mean and std) with neural networks

Nix and Weigend, 1994, Guillaumin and Zanna, 2021, Barnes et al., 2021, etc

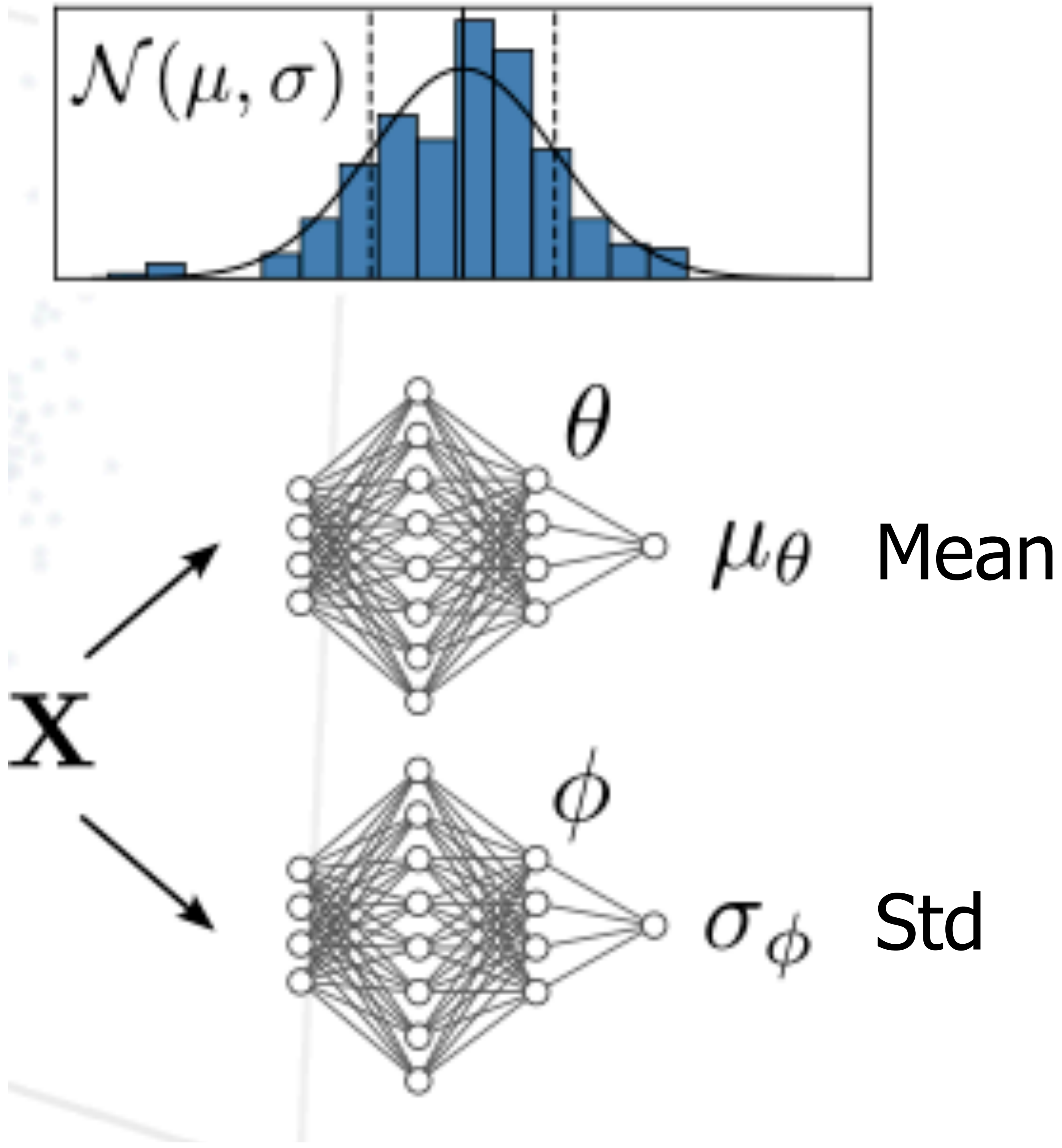


Mathematical model

- Assumption: conditional Gaussian dist.
- For each flux two ANNs
- Minimize negative log likelihood

$$L_{nll}(\theta, \phi) = \sum_{m=1}^{N} \frac{1}{2}\left[\log(\sigma_\phi^2(\mathbf{x}_m)) + \frac{(y_m - \mu_\theta(\mathbf{x}_m))^2}{\sigma_\phi^2(\mathbf{x}_m)}\right] + \text{const}\,.$$

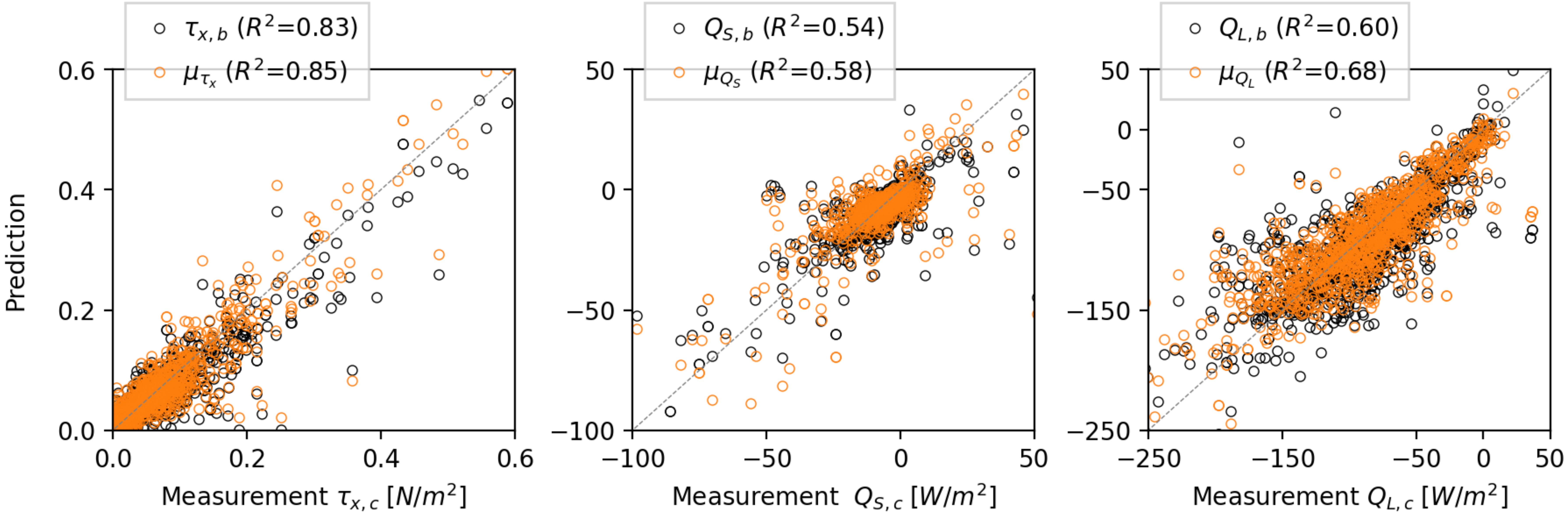
- Things that promote skills and prevent overfitting:
  - Choice of inputs $\mathbf{X} = (U_a, T_a, T_o, RH, p_a)$
  - Training on MSE before log likelihood loss
  - Early stopping



Evaluating the statistical scores

- Coefficient of determination $R^2(\hat{y}, y) = 1 - \mathbb{E}[(\hat{y} - y)^2]/\mathrm{Var}[y]$

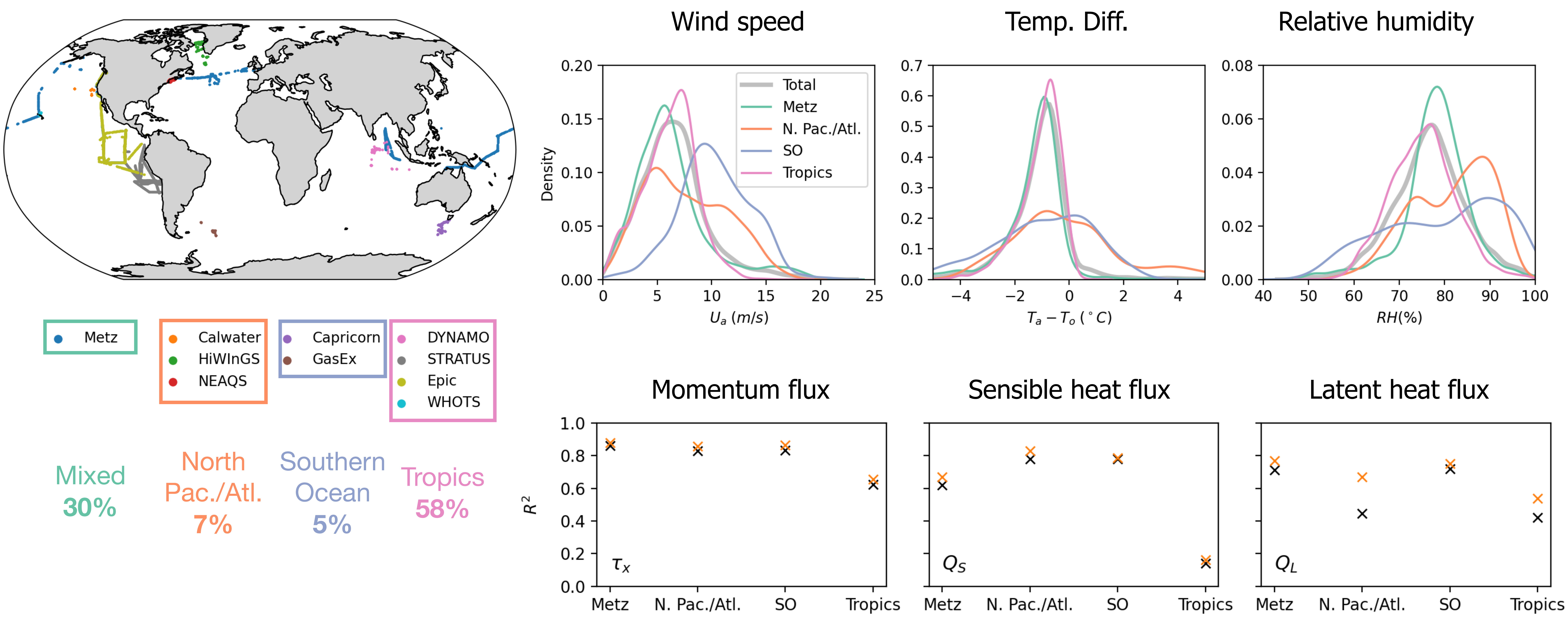
Momentum flux | Sensible heat flux | Latent heat flux



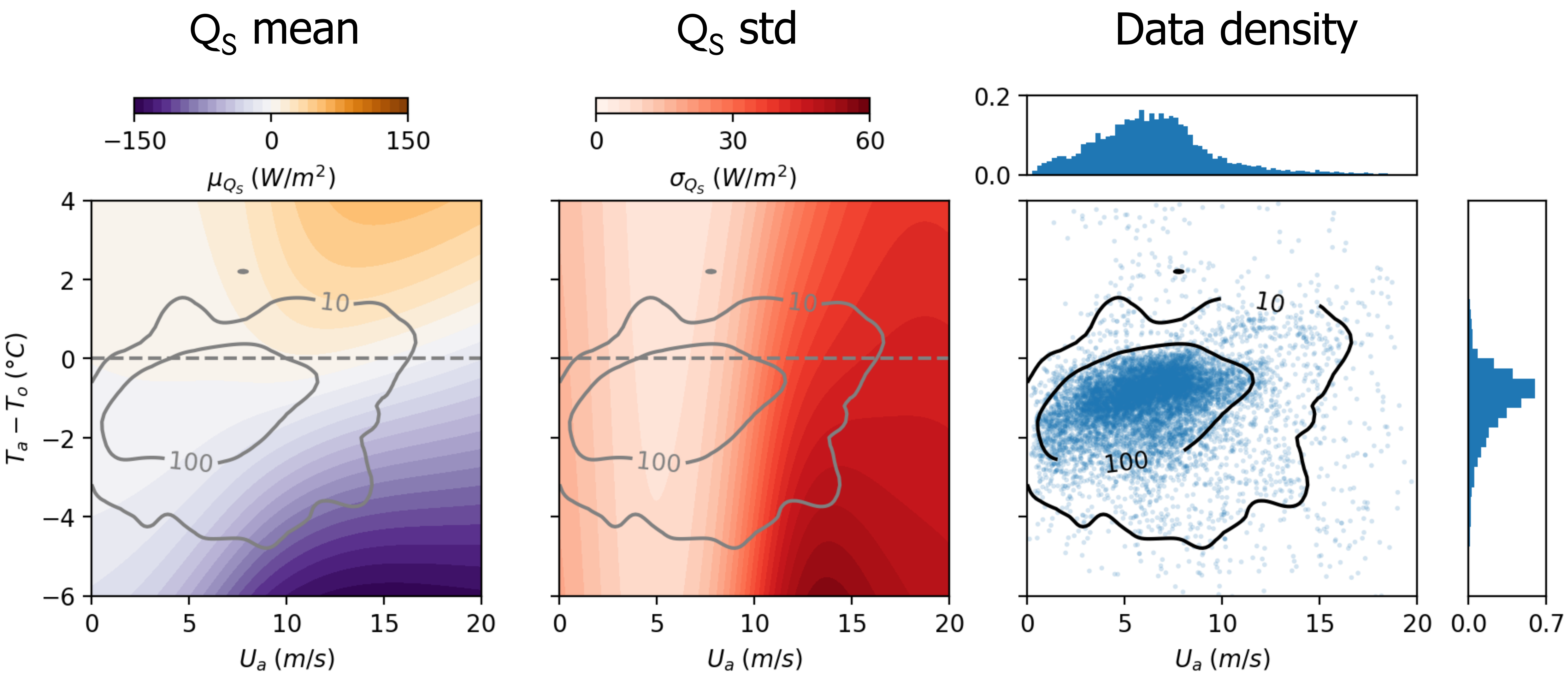
ANN

Baseline

Scores differ significantly across regions



Structure of predicted fluxes and uncertainty



- Not strictly down-gradient
- Limited by data (e.g. few high wind samples); extrapolate smoothly

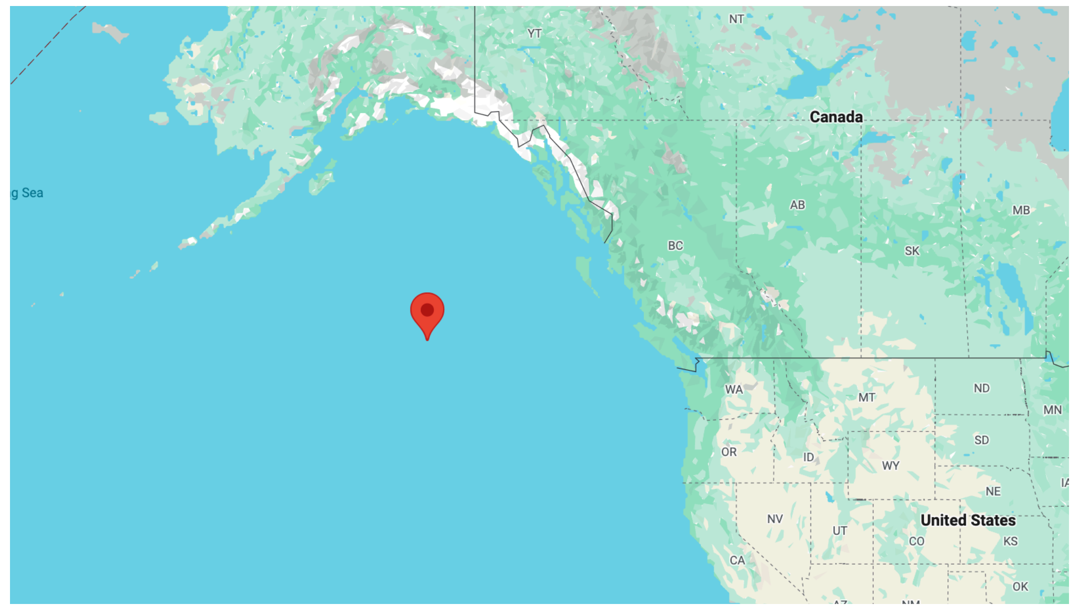
Implementation in a single-column model of upper ocean

- GOTM (General Ocean Turbulence Model)
- Surface fluxes imposed as boundary condition (also affect vertical mixing parameterization)
- Running the model in ‘forced’ way; fluxes computed offline; only modifying heat fluxes
- Limitations (ignoring horizontal advection)

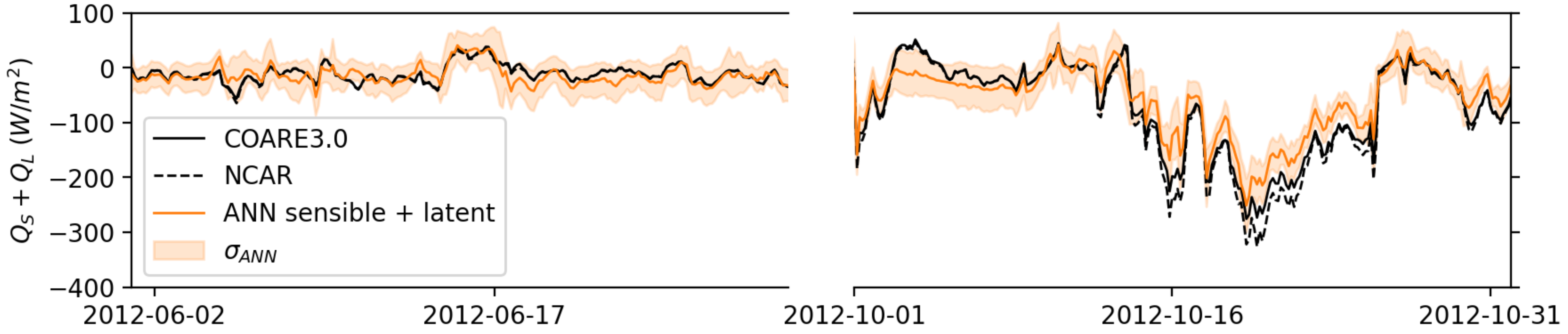
Governing equations

$$
\begin{aligned}
\partial_t U &= \partial_z[-\overline{wu}] + fV \\
\partial_t V &= \partial_z[-\overline{wv}] - fU \\
\partial_t T &= \partial_z[-\overline{w\theta}] - \partial_z Q_n - A_T \\
\partial_t S &= \partial_z[-\overline{ws}] - \partial_z F_n - A_S
\end{aligned}
$$

Vertical mixing: KPP or k-epsilon



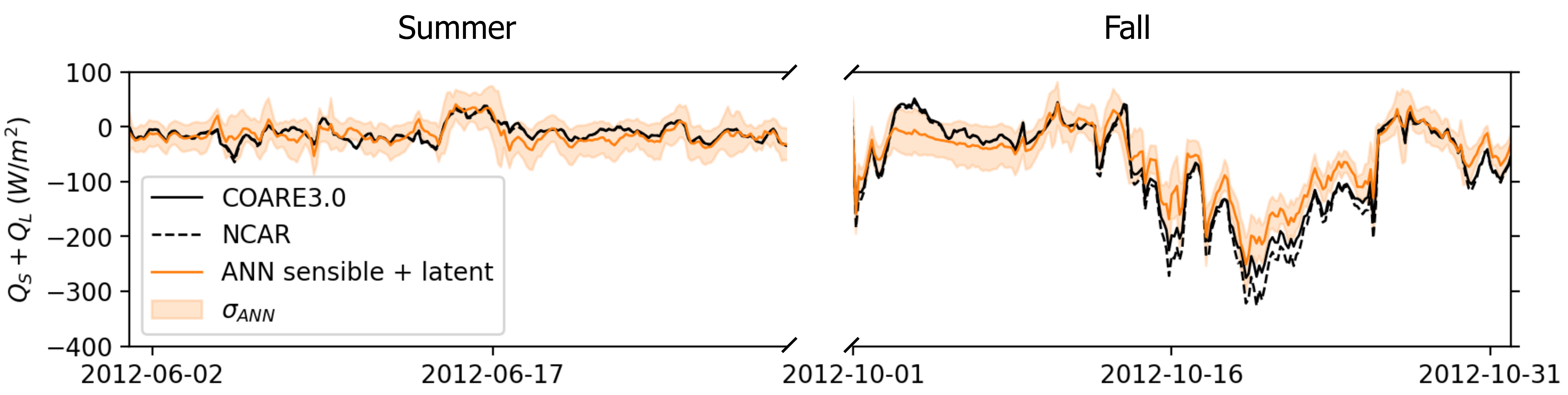
Ocean Weather Station Papa



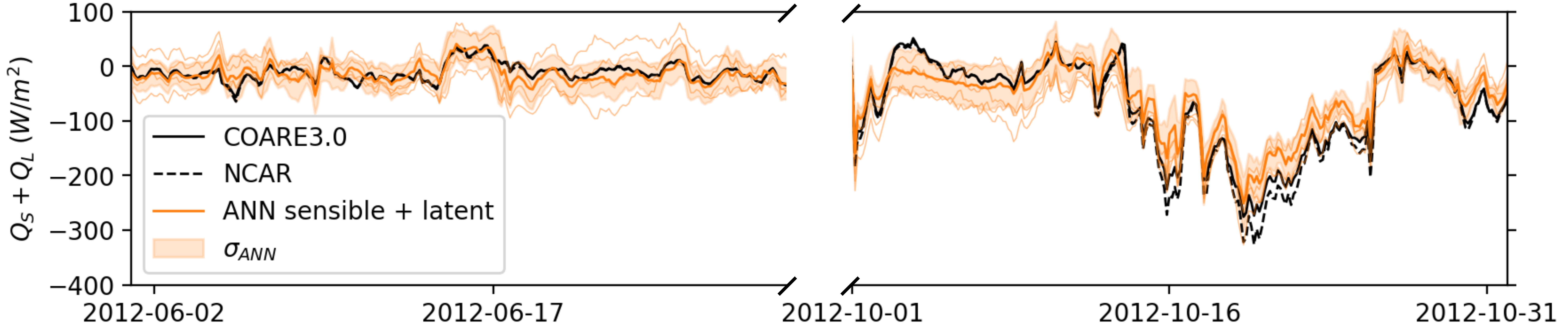
Long-term mooring records of state variables
No direct flux measurements

ANN or Baseline

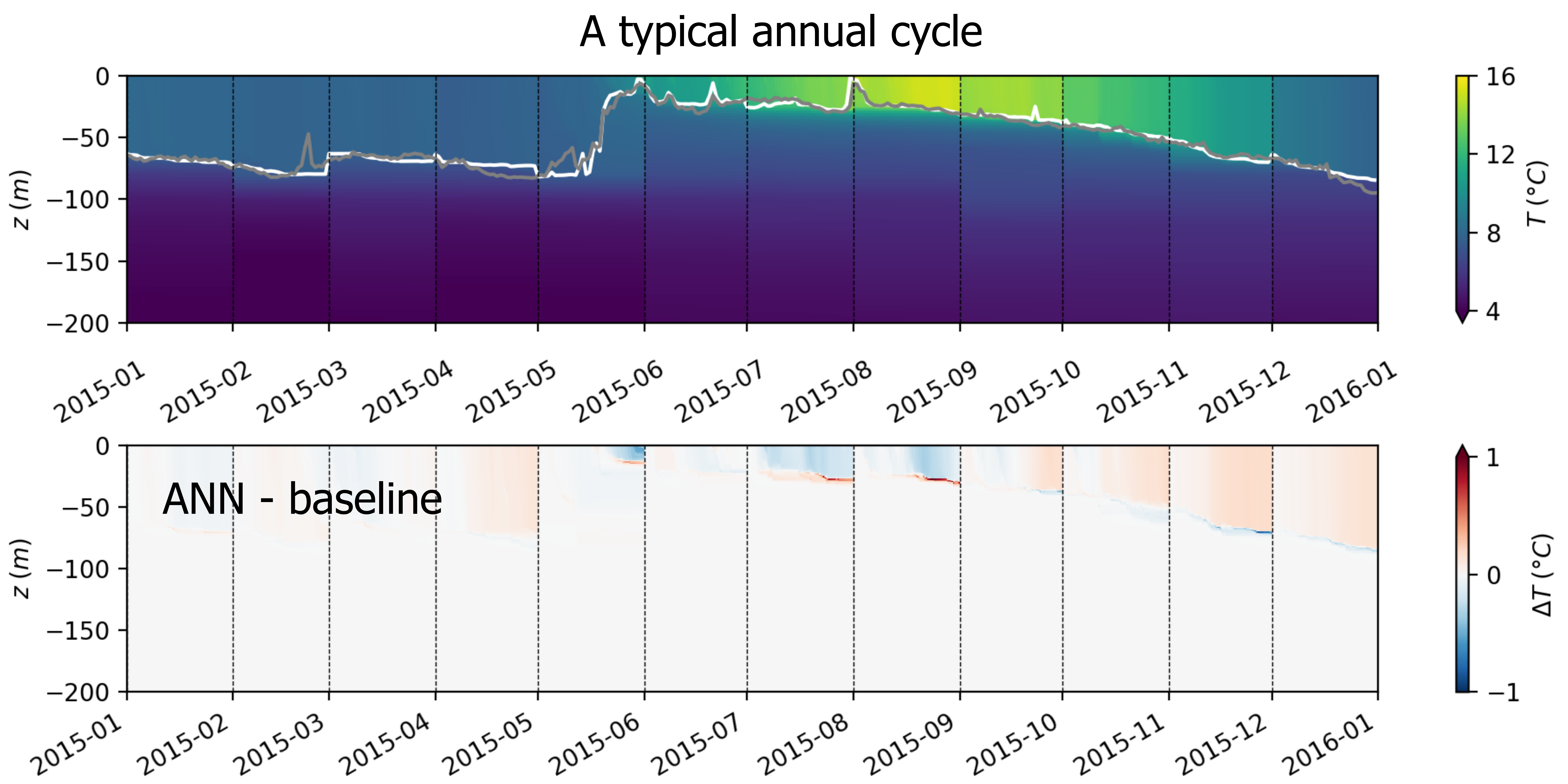
Comparing heat flux time series



Stochastically perturbed fluxes, noise generated by auto-regressive process AR(1)



Comparing state (SST and MLD)



- Sea surface temperature (SST) and Mixed layer depth (MLD) diagnosis

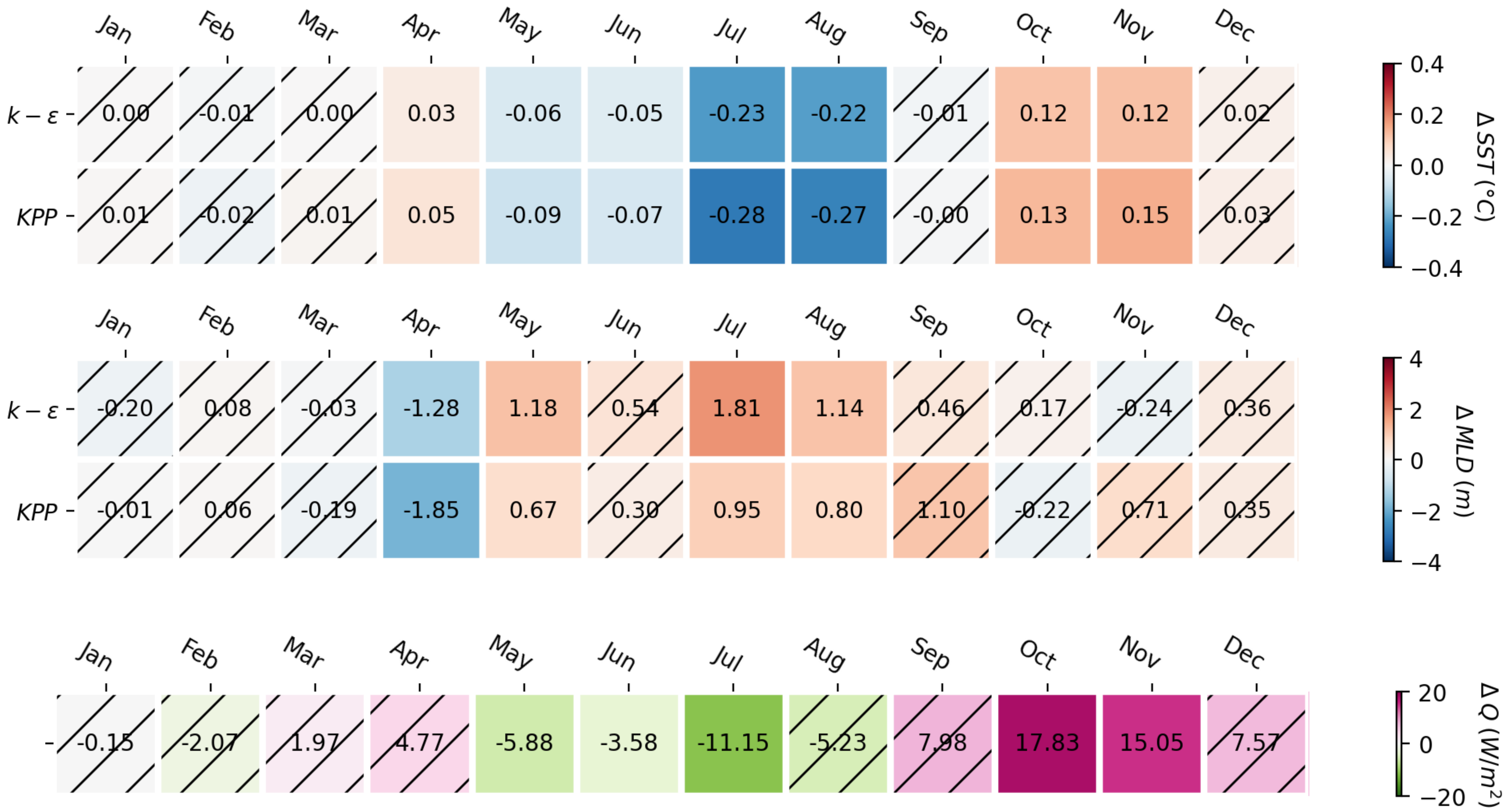
New flux model has a seasonal effect

2011, 2012, 2015, 2016

Response in SST

Response in MLD

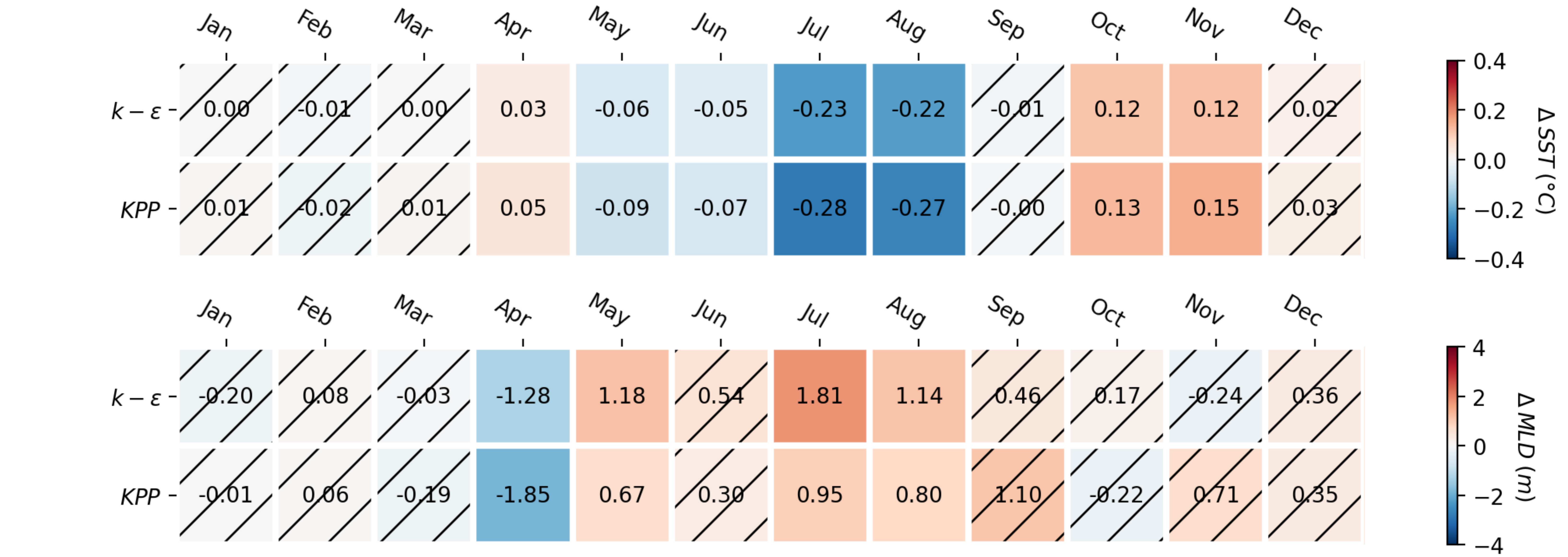
Monthly-aver. heat flux diff.



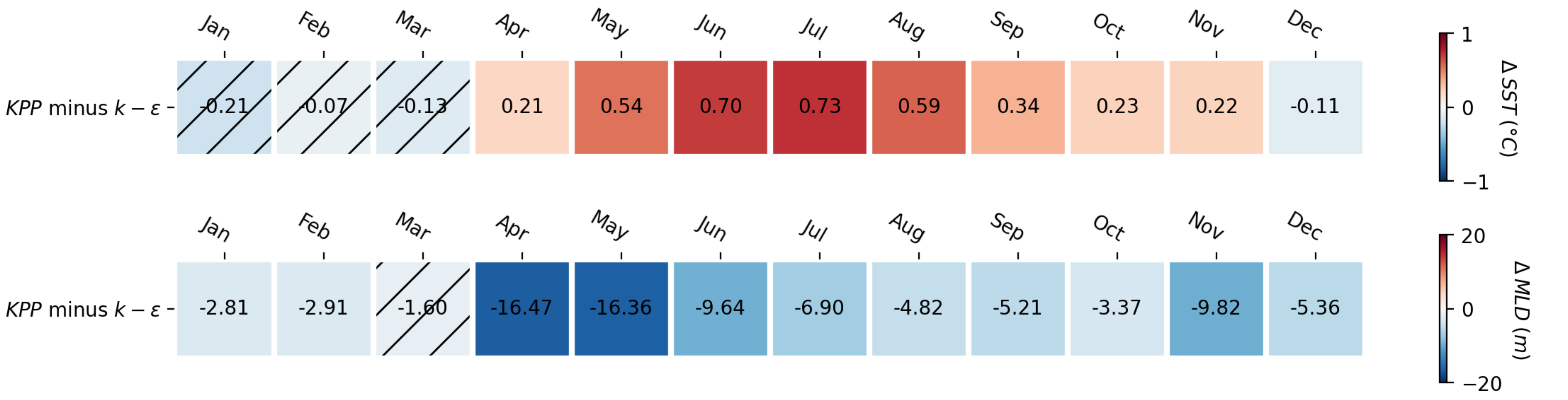
- There is a interplay with the vertical mixing. Not a simple heat budget balance.

Smaller magnitude but seasonal response

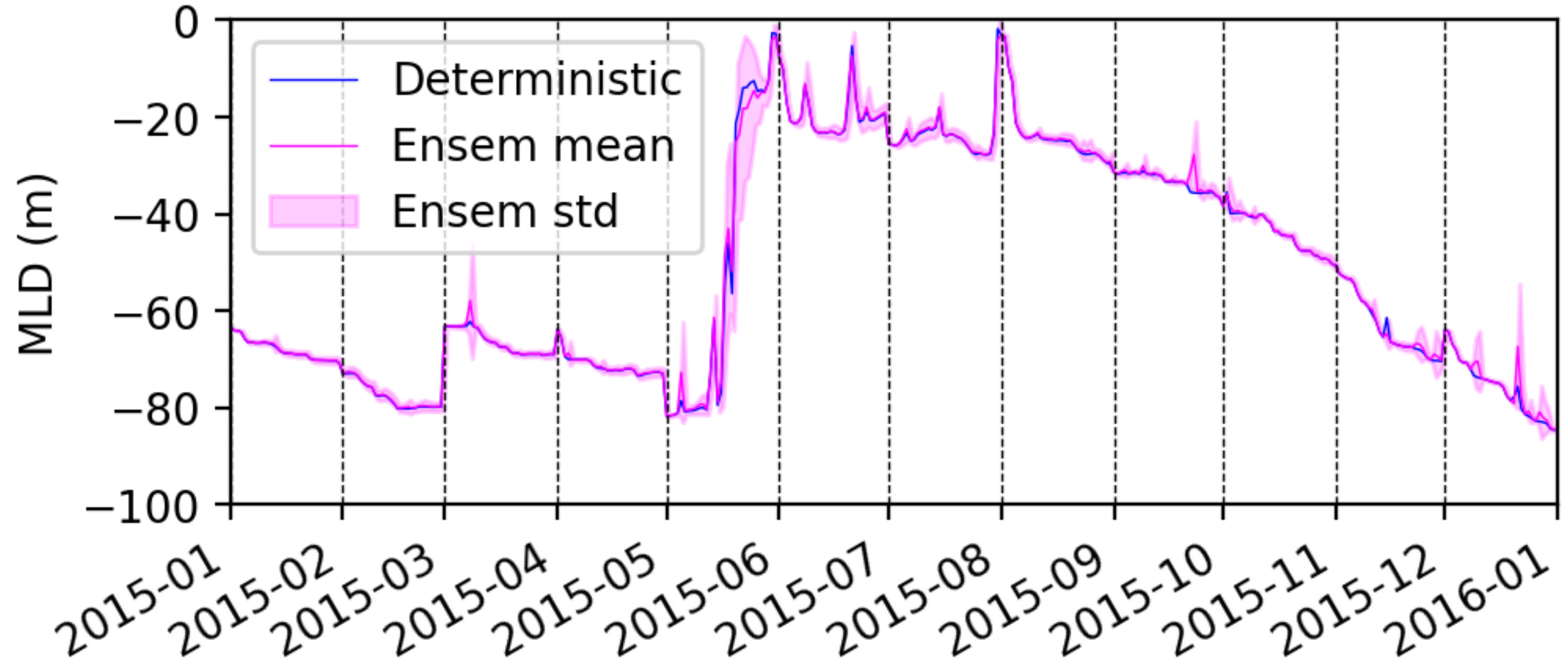
Change flux

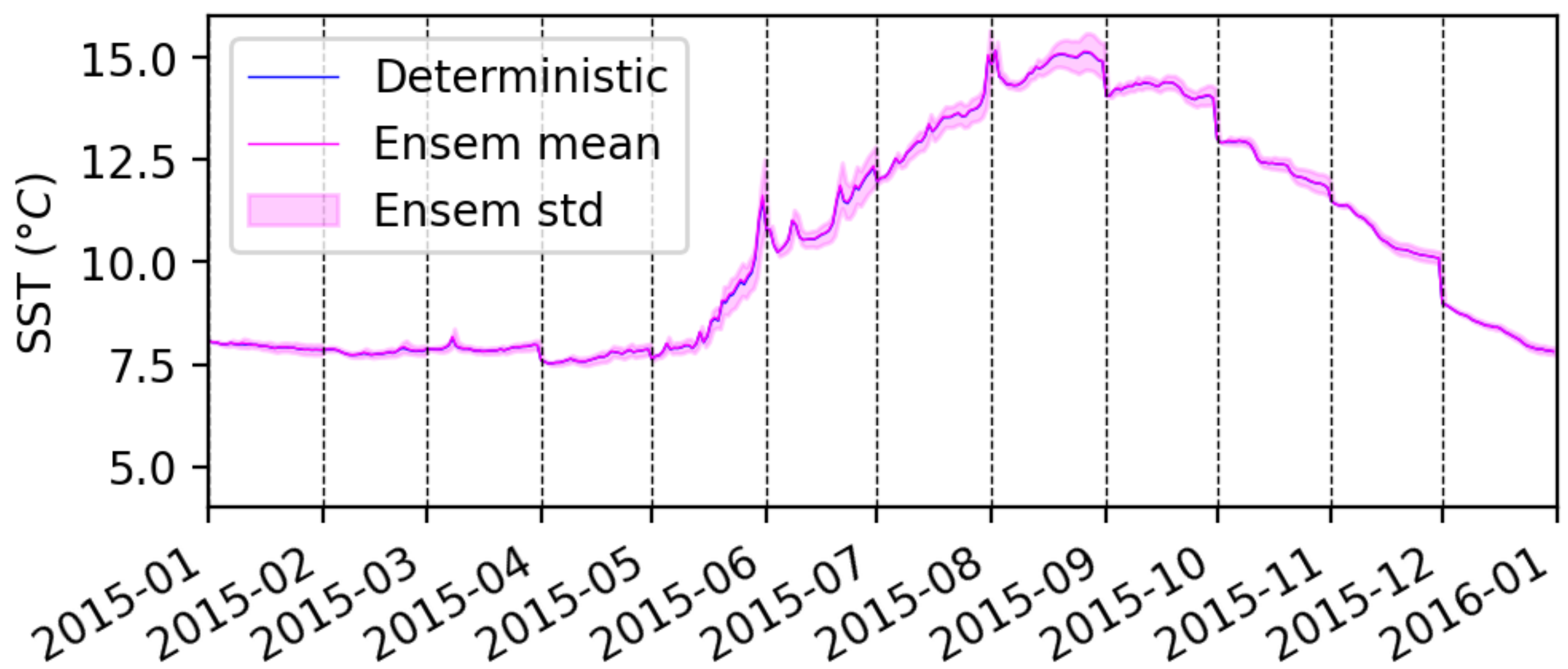


Change mixing



Spread in ensemble members





Spread in SST

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
$k-\varepsilon$	0.07	0.06	0.07	0.11	0.21	0.23	0.28	0.34	0.18	0.13	0.10	0.07
KPP	0.14	0.13	0.15	0.19	0.30	0.29	0.36	0.43	0.21	0.17	0.14	0.12

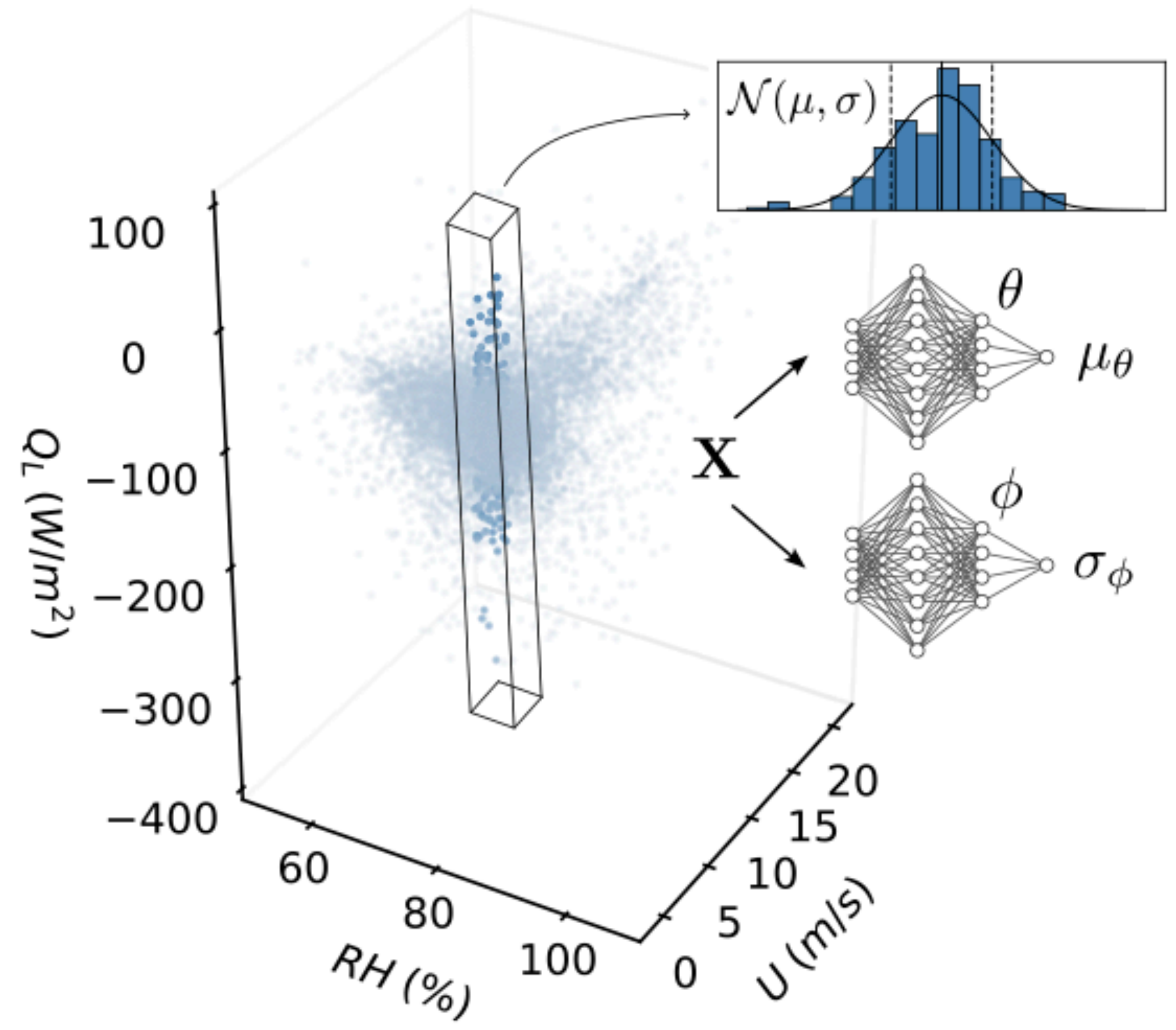
σ_{SST} (°C): 0.0 – 0.5

Spread in MLD

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
$k-\varepsilon$	0.93	0.84	0.81	4.33	7.18	3.13	2.47	1.82	1.26	1.95	1.21	0.90
KPP	0.64	1.17	1.38	8.04	7.04	2.57	2.00	1.73	2.76	2.16	1.47	0.81

σ_{MLD} (m): 0 – 5 – 10

Summary



- A **probabilistic model** for air-sea fluxes:
  - Compact ANNs and bulk inputs
  - **Mean** - similar to bulk algorithm, slightly better statistical correlation to observations
  - **Variance** - UQ and stochastic parameterization
- Implementation in single-column forced upper ocean:
  - Strong seasonality in predicted flux difference and response
  - Limitation of single column model -> coupled CESM runs
  - Large spread can have additional impact when coupled to nonlinear processes

Manuscript:

Wu, J., Perezhogin, P., Gagne., D.J., Reichl, B., Subramanian, A., Thompson, E., and Zanna, L., Data-Driven Probabilistic Air-Sea Flux Parameterization, in prep.