Using machine learning to parameterize moist convection: potential for modeling of climate, climate change and extreme events

Paul O'Gorman, MIT

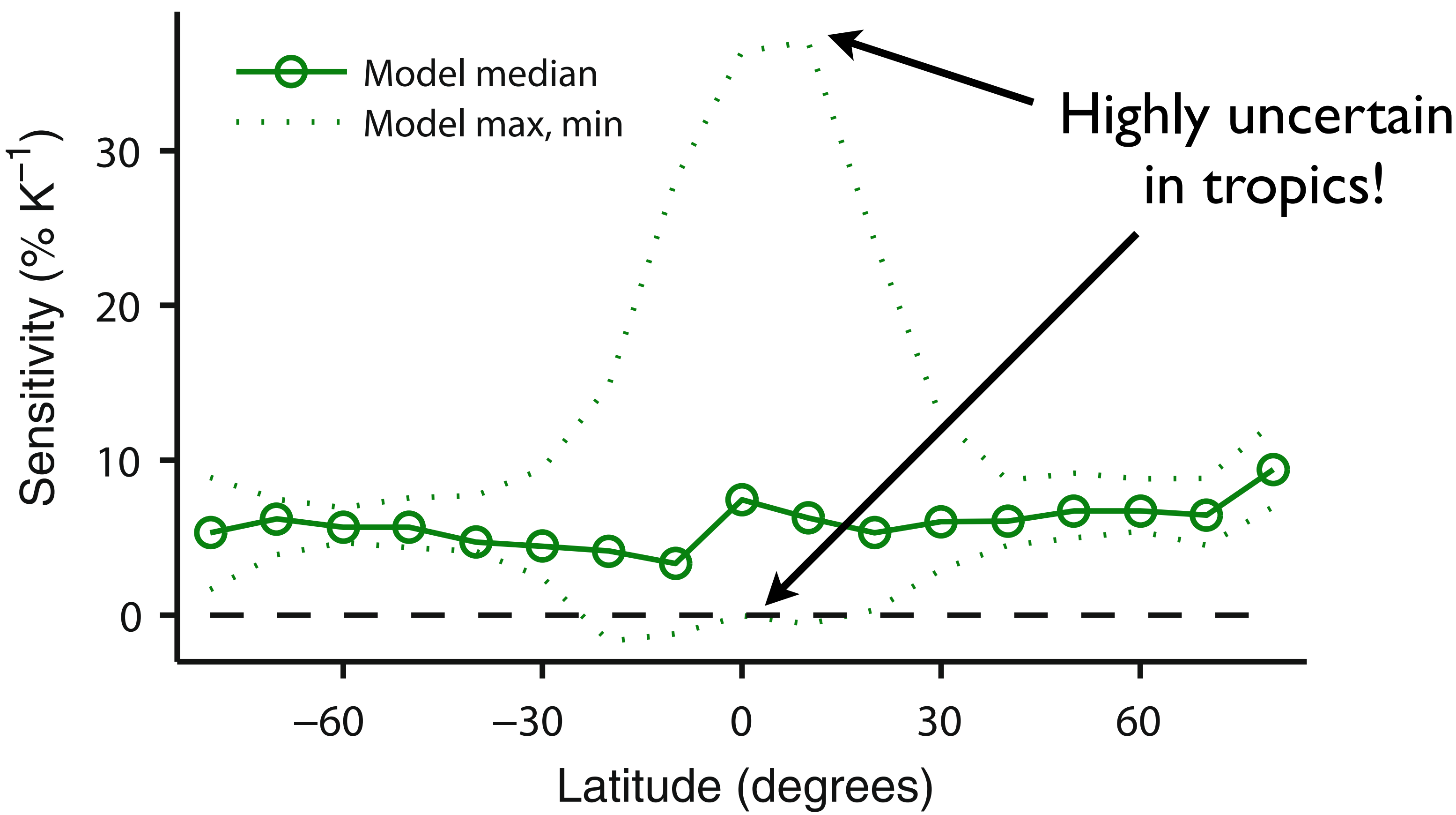
In collaboration with John Dwyer (Dia)

Funding support from the NSF and

the MIT Environmental Solutions Initiative

Clear need for improvement in parameterizations of moist convection for simulations of climate change

Response of precipitation extremes to climate warming in CMIP5 models



O'Gorman, Curr. Climate Change Reports, 2015

Four questions

Can a machine-learning-based parameterization of moist convection be used for:

1. Accurate simulations of climate?

2. Accurate simulations of extreme events (without special training)?

3. Climate change?

4. Learning about the underlying physics and dynamics?

Focus on behavior when machine-learning parameterization implemented in GCM using idealized tests

- Use "perfect-parameterization" test in which emulate a conventional parameterization
- View as best-case scenario for column-based learning
- Use idealized aquaplanet GCM to simplify analysis

O'Gorman & Dwyer, submitted to JAMES

Learn mapping from "features" (inputs) to "outputs"

$$y = f(x)$$

Nonlinear function

$$x = (T(\sigma), q(\sigma), p_s)$$

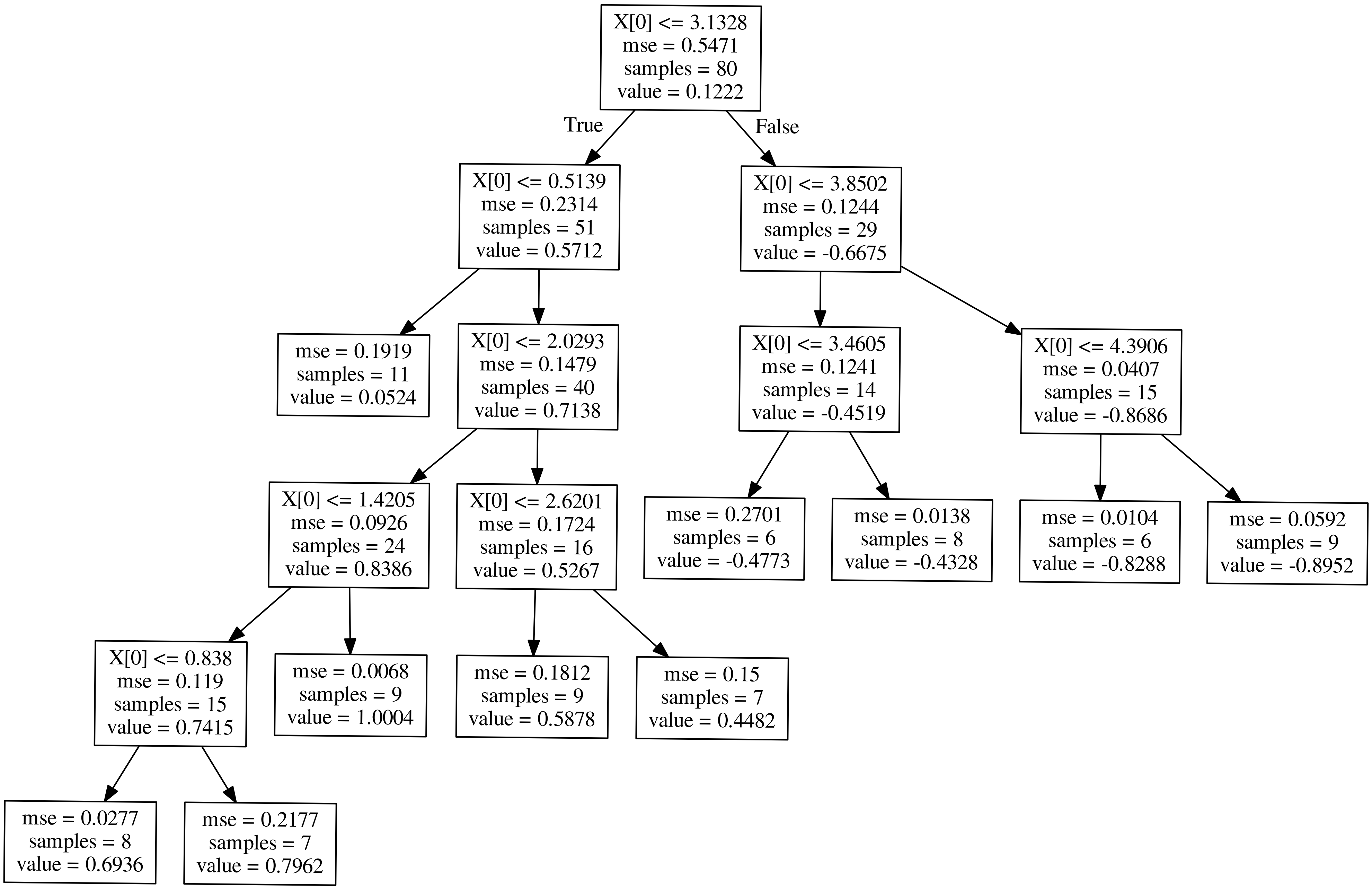
Features: vertical profiles of temperature and specific humidity

$$y = \left(\frac{\partial T}{\partial t}_{\mathrm{conv}}, \frac{\partial q}{\partial t}_{\mathrm{conv}} \right)$$

Outputs: vertical profiles of convective tendencies of temeprature and specific humidity

Include surface pressure p_s as a feature because using vertical sigma (σ) coordinate

Machine-learning algorithm:
Random Forest is an ensemble of decision trees



Why use a Random Forest for this application?

- **Found that Random Forest leads to stable and accurate simulations when implemented in GCM** (artifical neural nets lead to problems in GCM such as extremely dry or moist patches)
- **Advantage that predicted tendencies respect physical constraints (energy conservation, non-negative precipitation)**

O'Gorman & Dwyer, submitted to JAMES

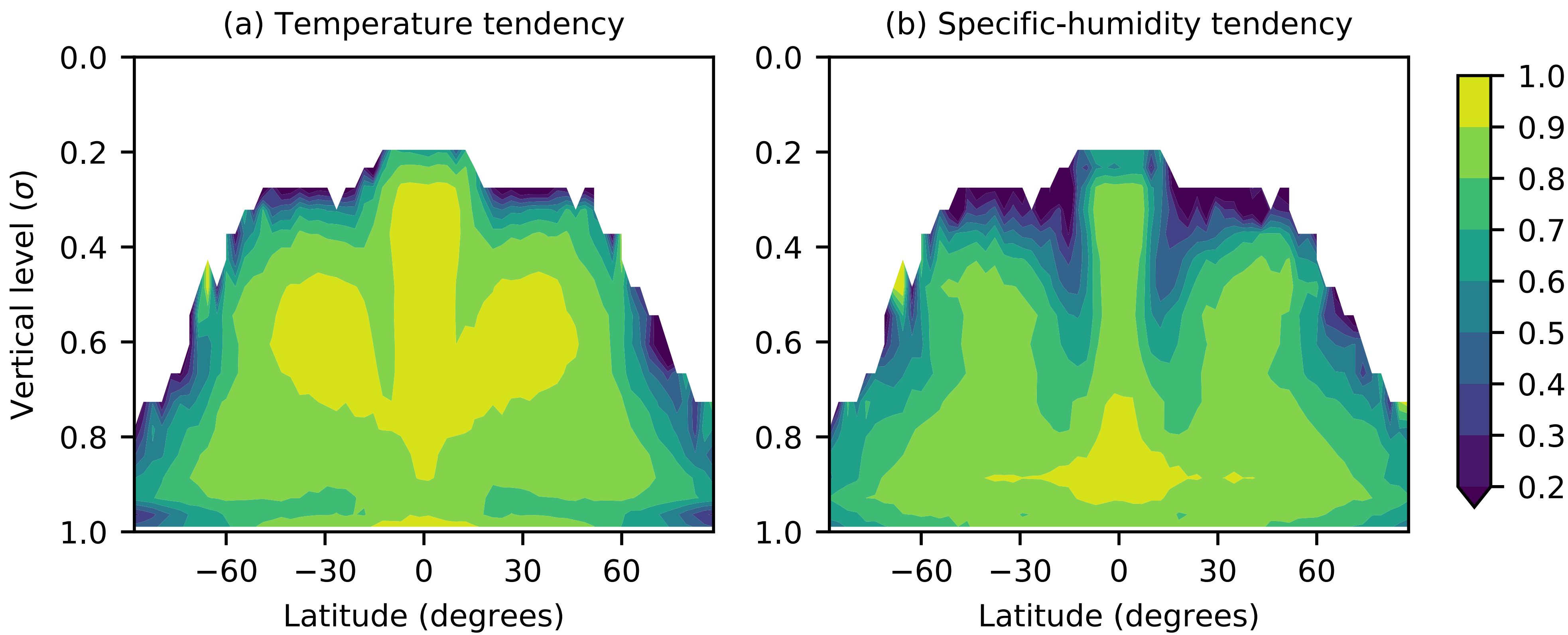
Details: idealized GCM and Random Forest

- Idealized GCM has 30 vertical sigma levels and T42 resolution, no land or ice, mixed layer ocean, gray radiation scheme for longwave *(GFDL dynamical core, as in Frierson et al 2006, O'Gorman and Schneider 2008)*

- Relaxed-Arakawa Schubert (RAS) convection scheme *(Moorthi et al 1992)*

- Trained Random Forest (10 trees) on 700,000 samples from statistical equilibrium simulation

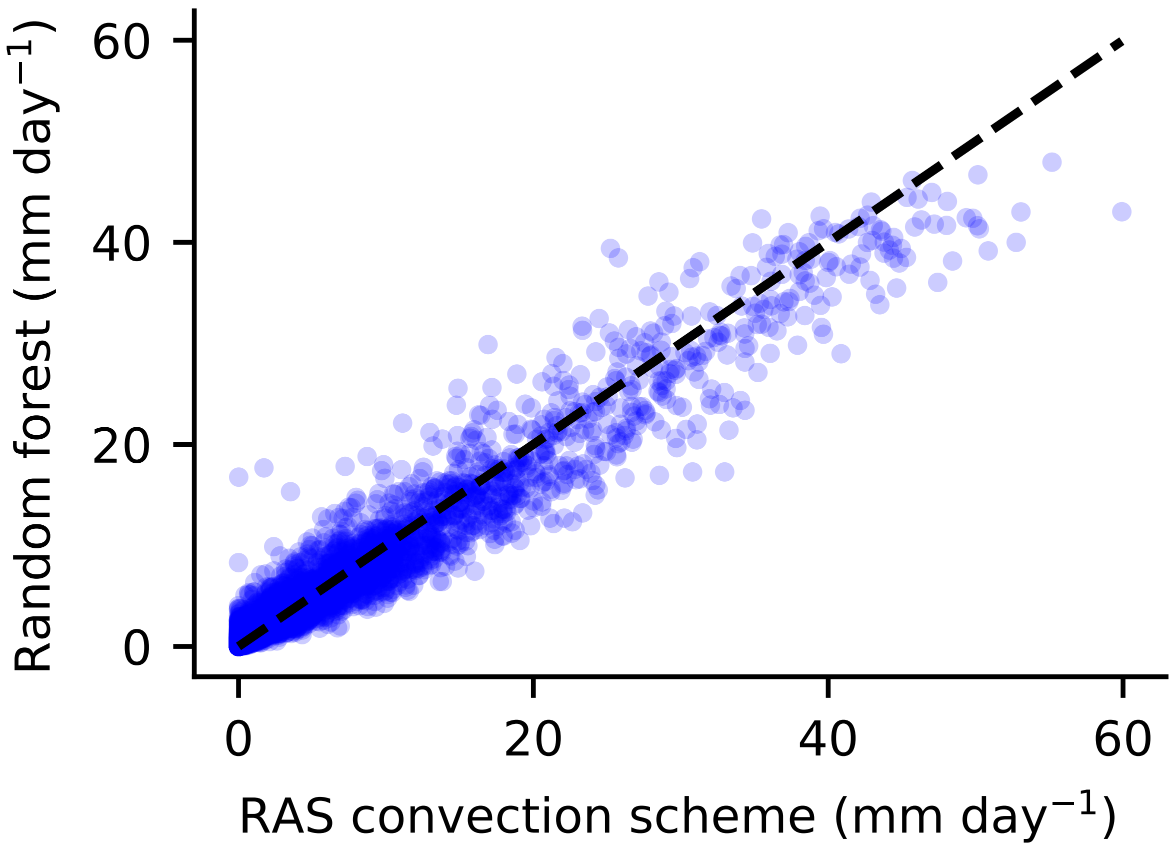
Random forest accurately predicts convective tendencies in test dataset



Correlation coefficients of convective tendencies for original convection scheme (RAS) versus Random Forest

O'Gorman & Dwyer, submitted to JAMES

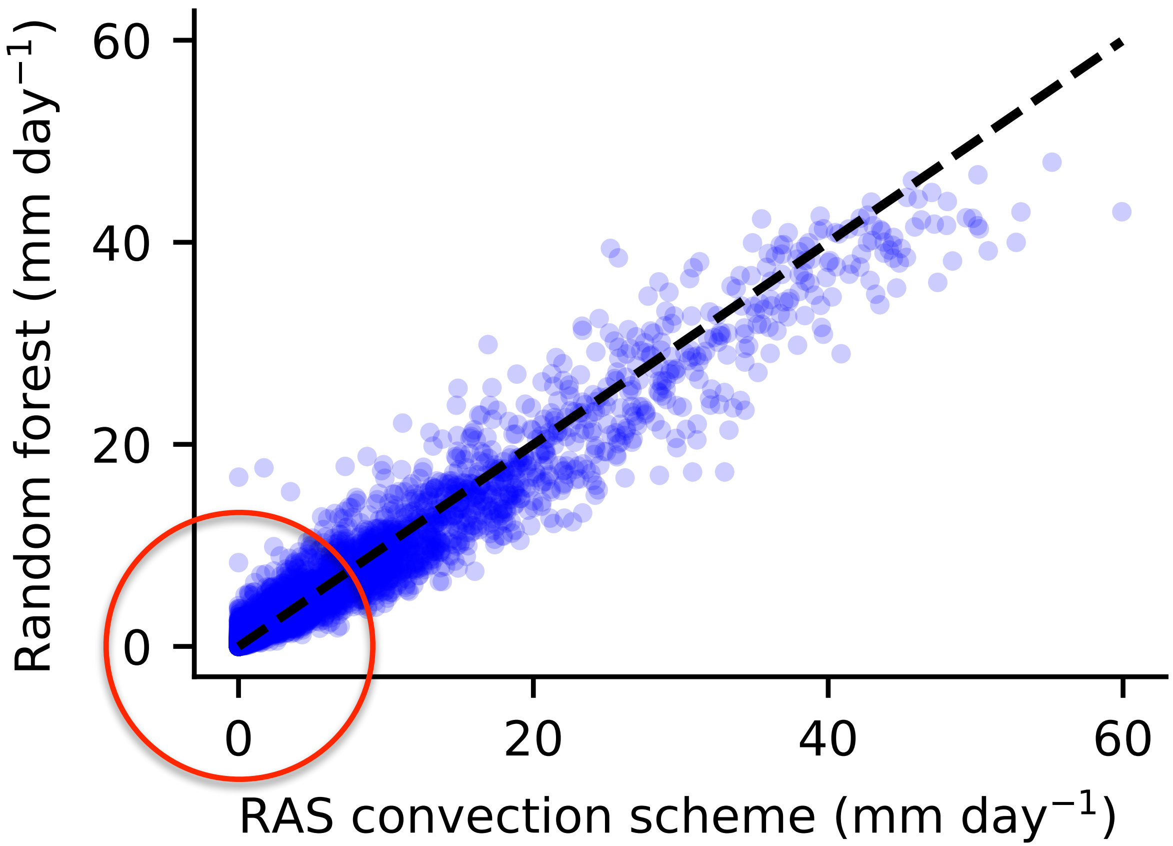
Random forest accurately predicts precipitation rate (including heavy events) in test dataset



Scatterplot of instantaneous surface precipitation rate from original convection scheme (RAS) versus Random Forest

O'Gorman & Dwyer, submitted to JAMES

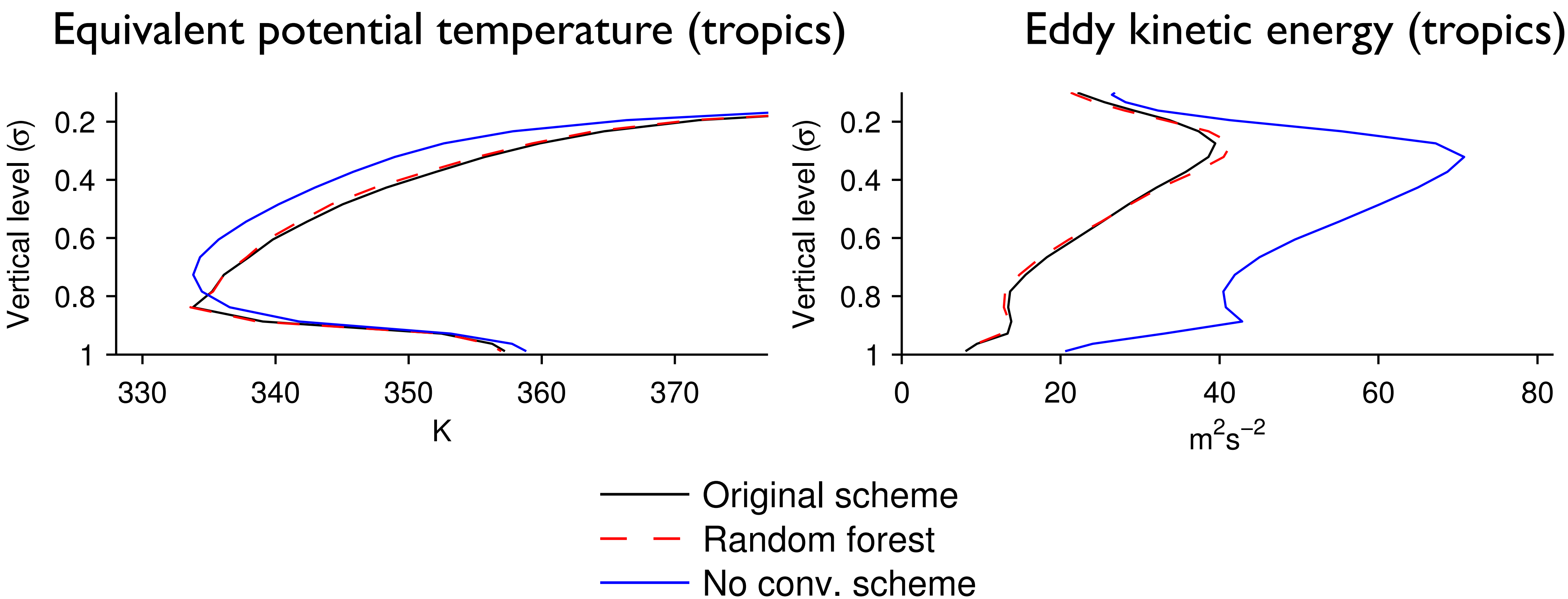
Physical constraints for precipitation and energy conservation are automatically respected



Precipitation rates are non-negative (see above)
Tendency of vertical integral of moist enthalpy is zero (not shown)

O'Gorman & Dwyer, submitted to JAMES

GCM with Random-Forest convection scheme runs stably and leads to accurate simulations of control climate

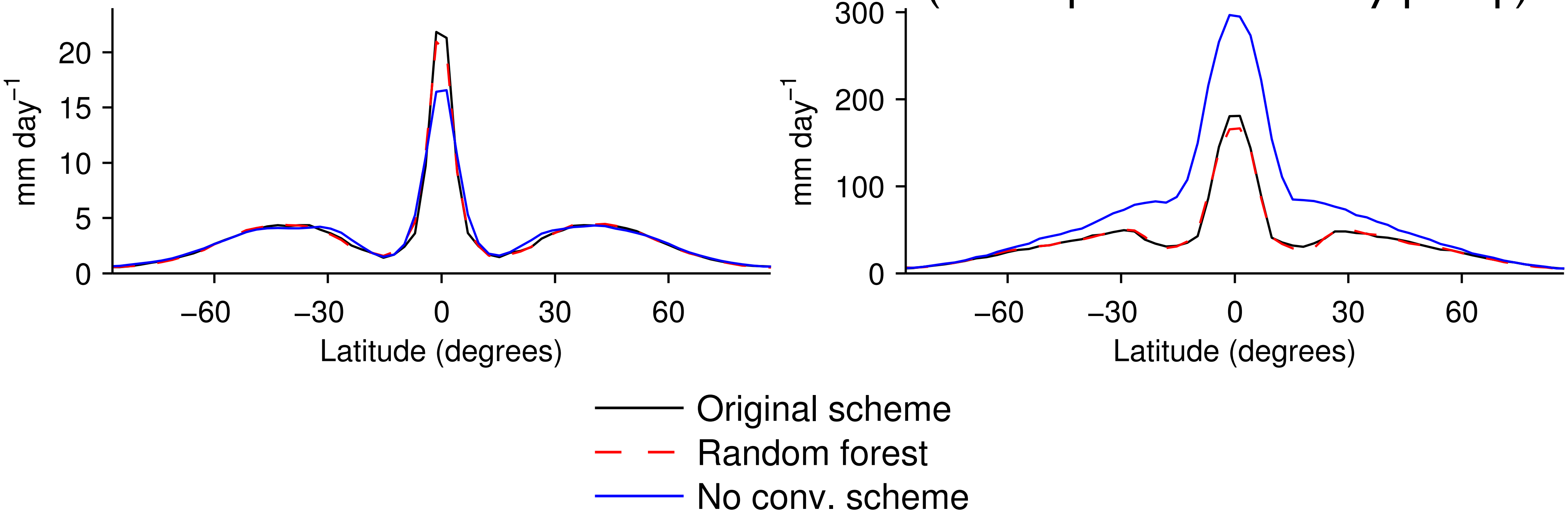


Tropical means over 20S to 20N
O'Gorman & Dwyer, submitted to JAMES

GCM with Random-Forest convection scheme does well for both mean and extreme precipitation

Mean precipitation

Extreme precipitation
(99.9th percentile of daily precip)



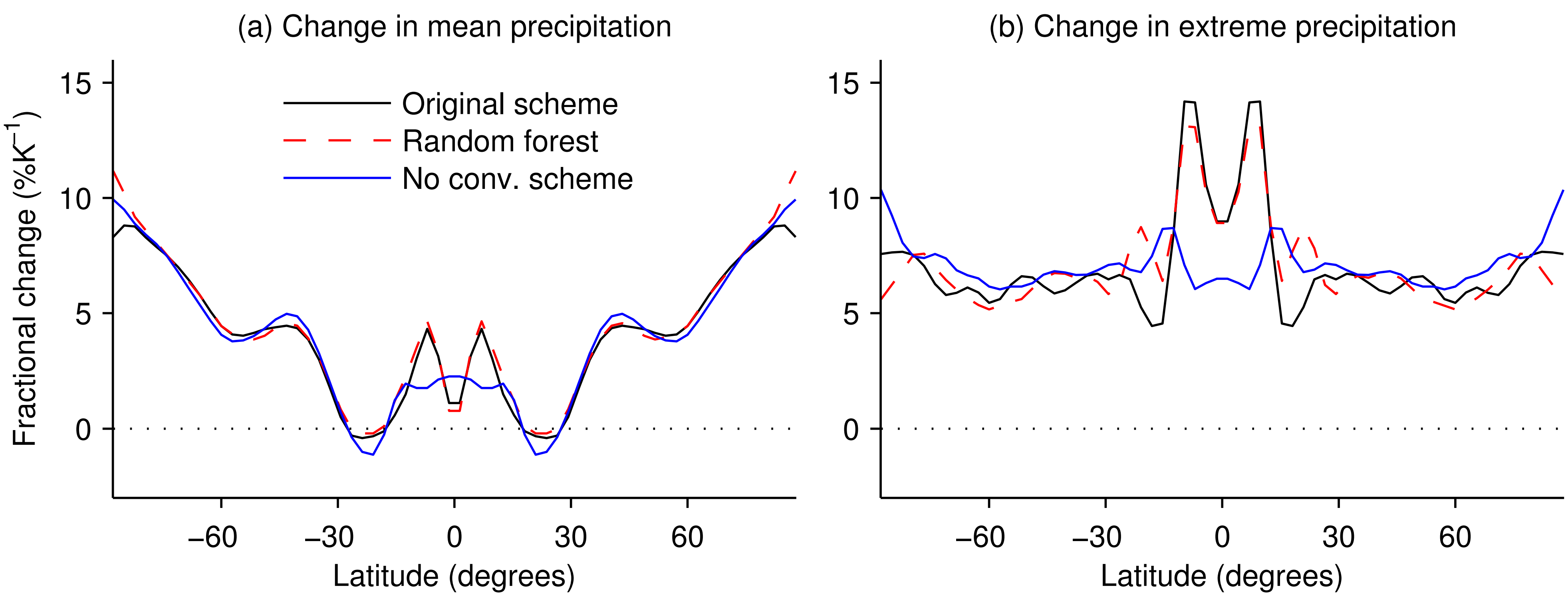
O'Gorman & Dwyer, submitted to JAMES

Climate change

- Climate change imposed in GCM by rescaling optical depth of longwave absorber in gray radiation scheme (dT = 6.5K)

- Most conservative approach is to train a separate Random Forest for each of the control and warm climates

GCM with Random-Forest scheme correctly captures climate-change response of mean and extreme precipitation



%/K relative to zonal- and time-mean surface temperature change
Extreme precipitation is 99.9th percentile of daily precipitation

O'Gorman & Dwyer, submitted to JAMES

Climate change: alternative training strategies

1. Train one Random Forest on combined samples from both control and warm climates

2. Train on control climate only

3. Train on warm climate only

Climate change: alternative training strategies

1. Train one Random Forest on combined samples from both control and warm climates

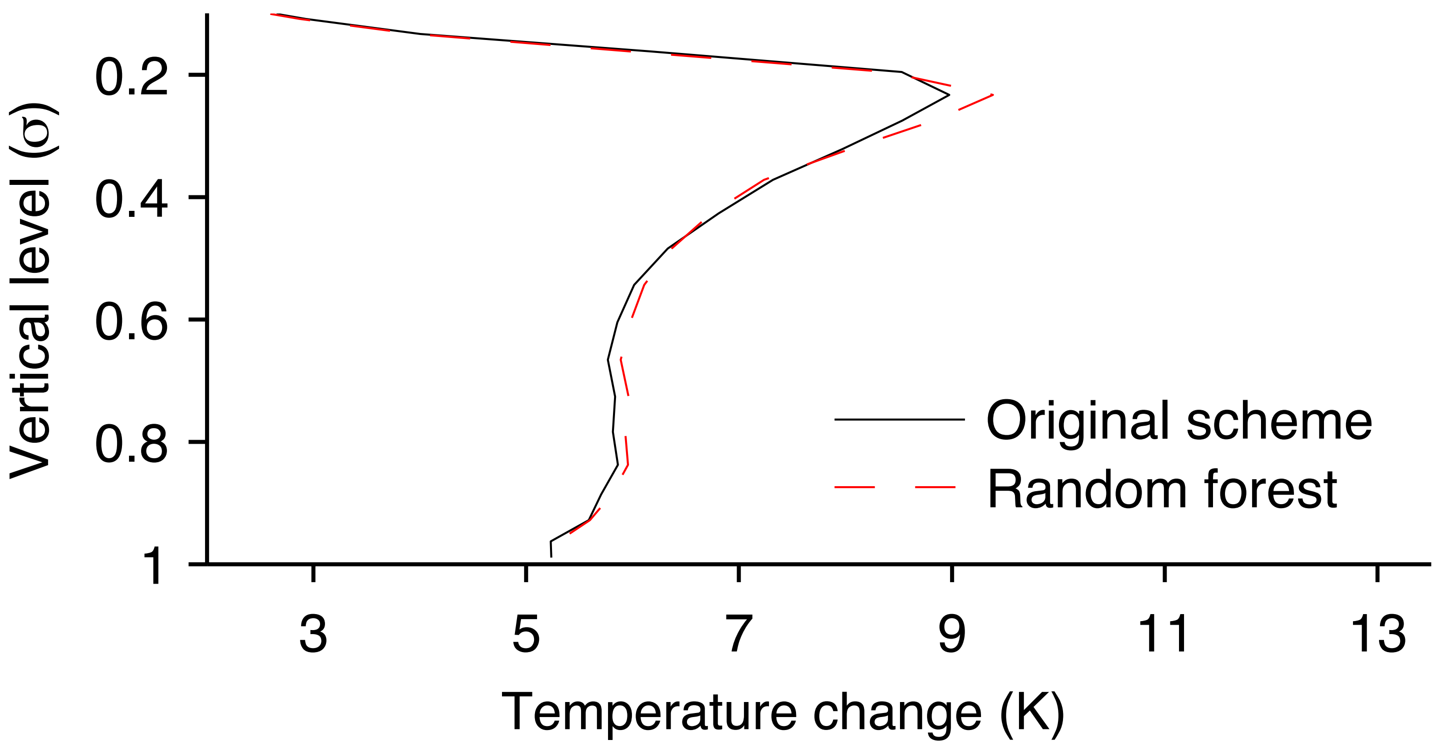
2. Train on control climate only

3. Train on warm climate only

Which approaches work or don't work?

Training a Random Forest for each climate separately: Accurate simulations of climate change

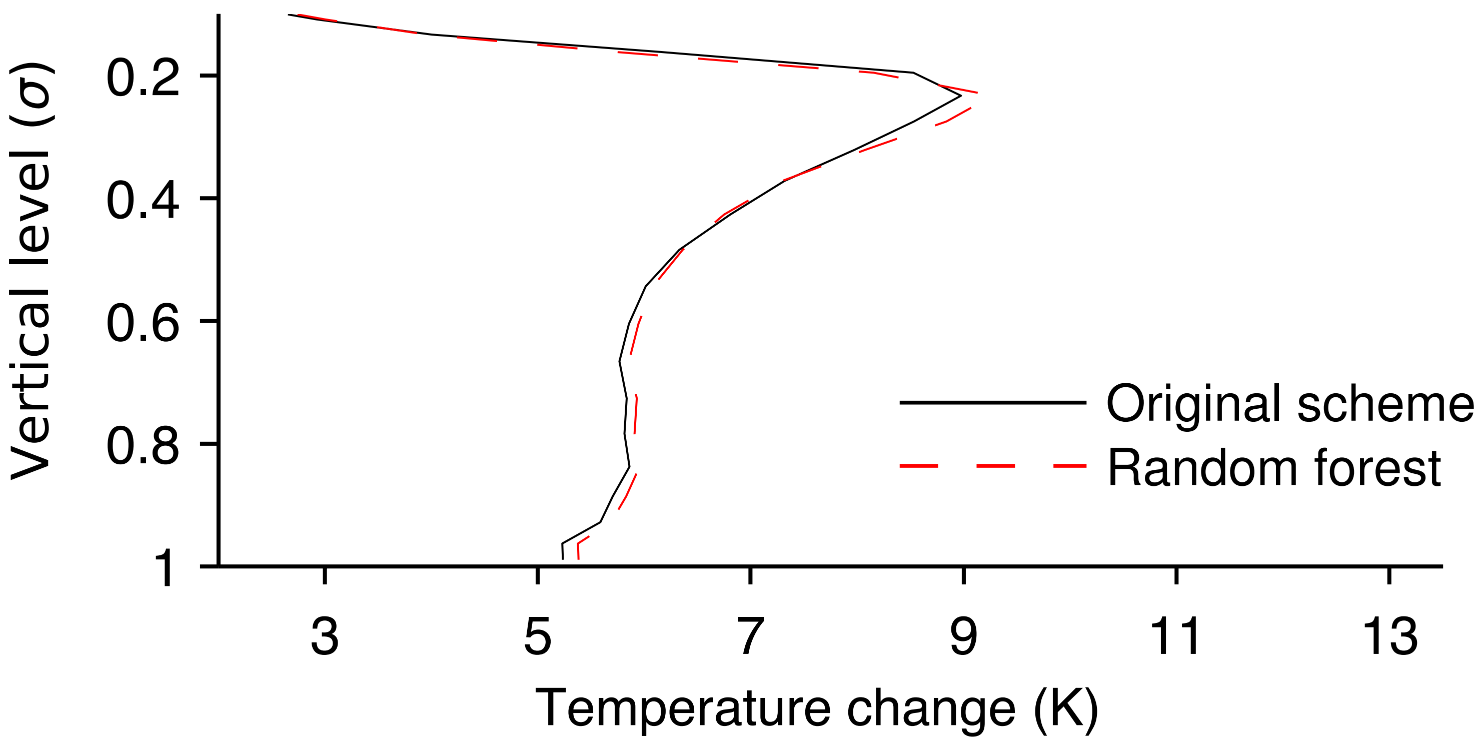
Vertical profile of warming in tropics



O'Gorman & Dwyer, submitted to JAMES

Training one Random Forest on combined samples from both climates: Accurate simulations of climate change

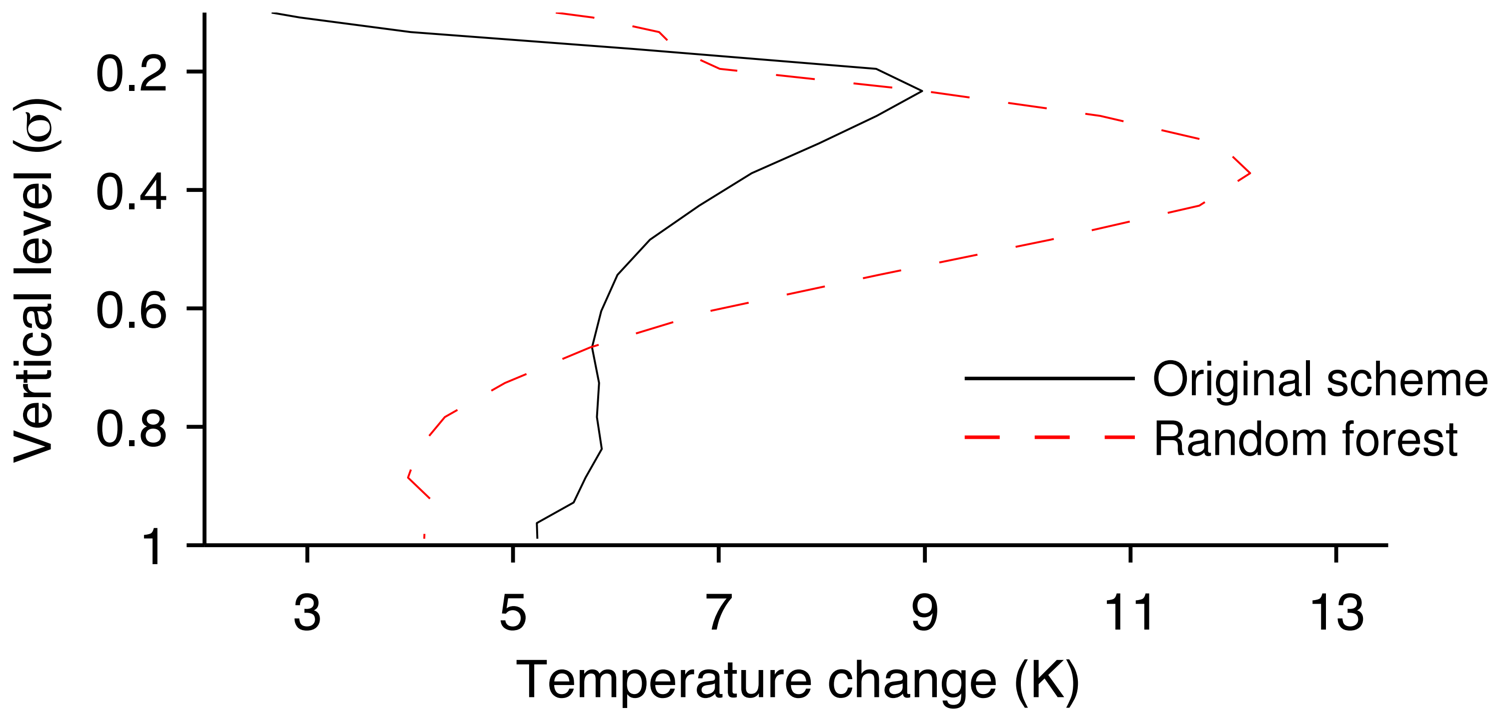
Vertical profile of warming in tropics



O'Gorman & Dwyer, submitted to JAMES

Training on control climate only:
Very inaccurate simulations of climate change

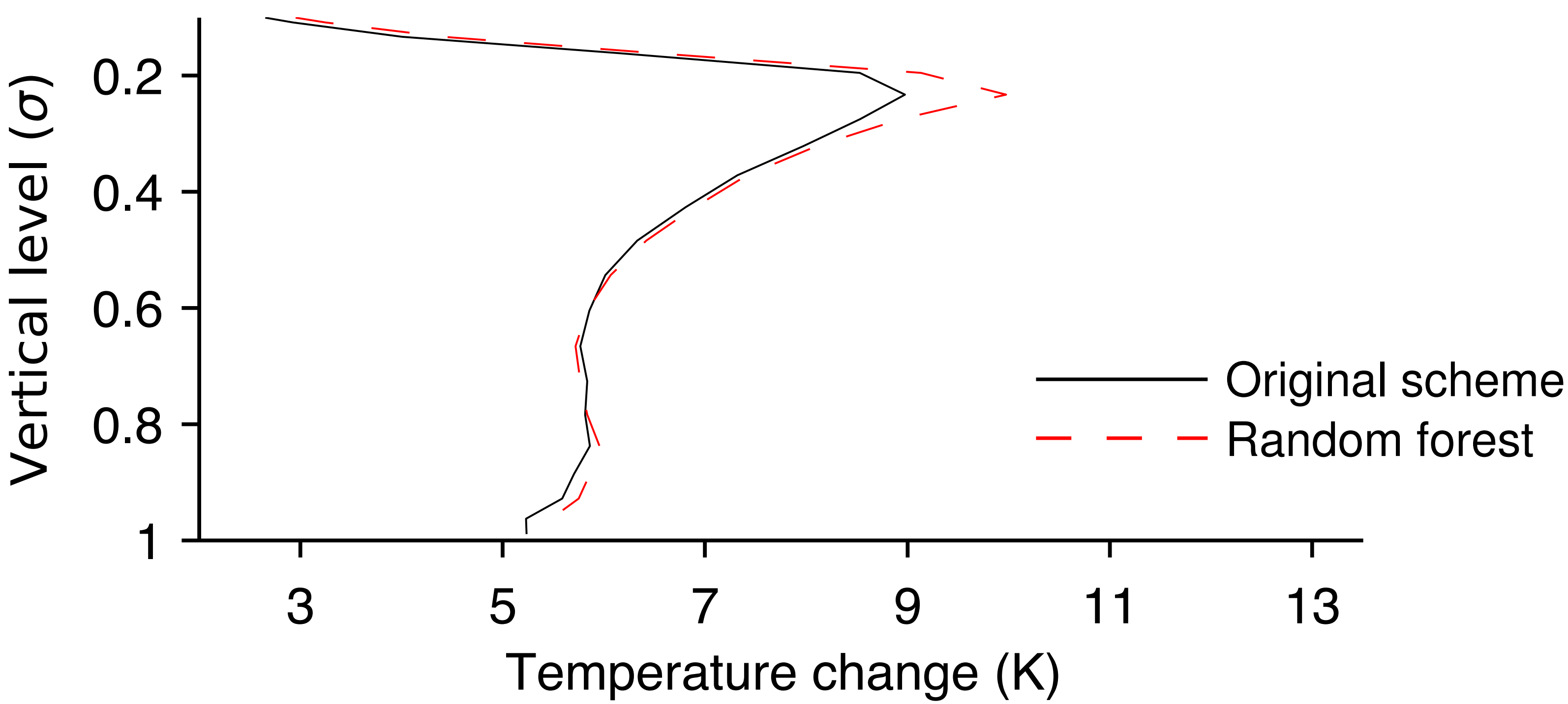
Vertical profile of warming in tropics



O'Gorman & Dwyer, submitted to JAMES

Training on warm climate only: Does surprisingly well!

Vertical profile of warming in tropics



Can find training samples in warm climate at higher latitudes that match tropical samples in control climate

Machine-learning paramterization is a nonlinear mapping that can be interrogated to learn about underlying dynamics/physics

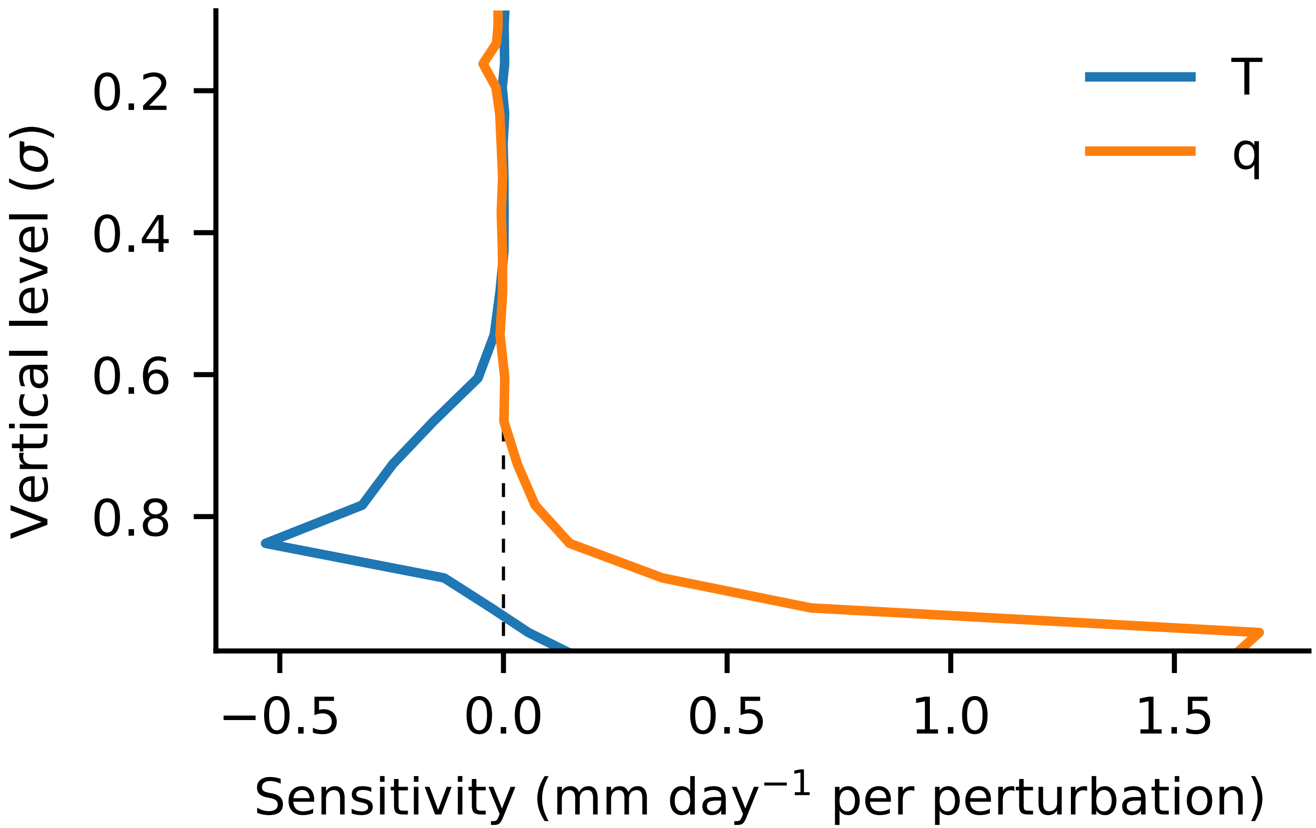
Illustrate here using two metrics:

1. Linear response function (e.g., Kuang 2010; Mapes et al 2017)

2. Feature importance

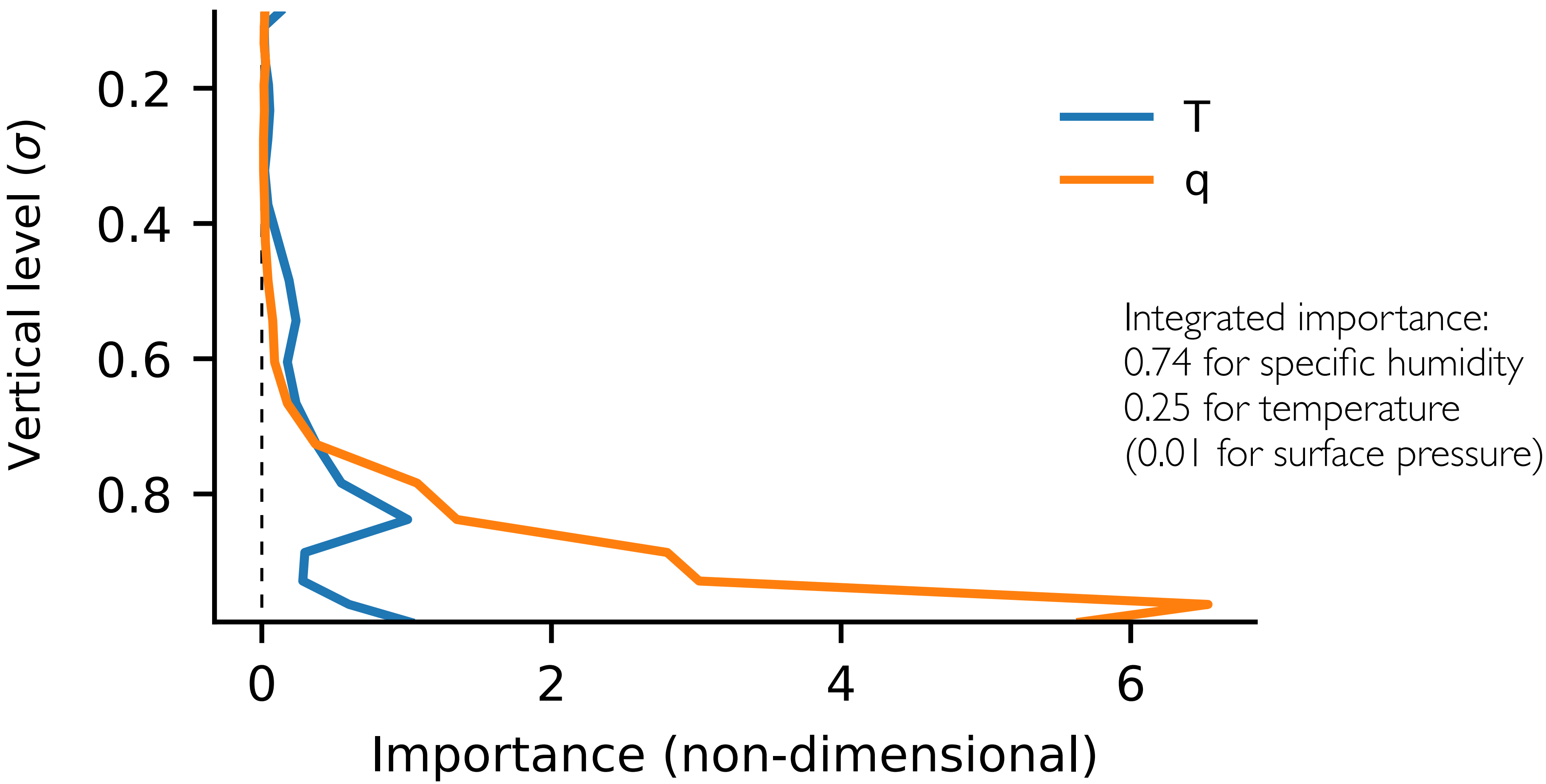
O'Gorman & Dwyer, submitted to JAMES

Linear response of surface precipitation rate to perturbations in temperature (1K) and moisture (1g/kg) at different levels



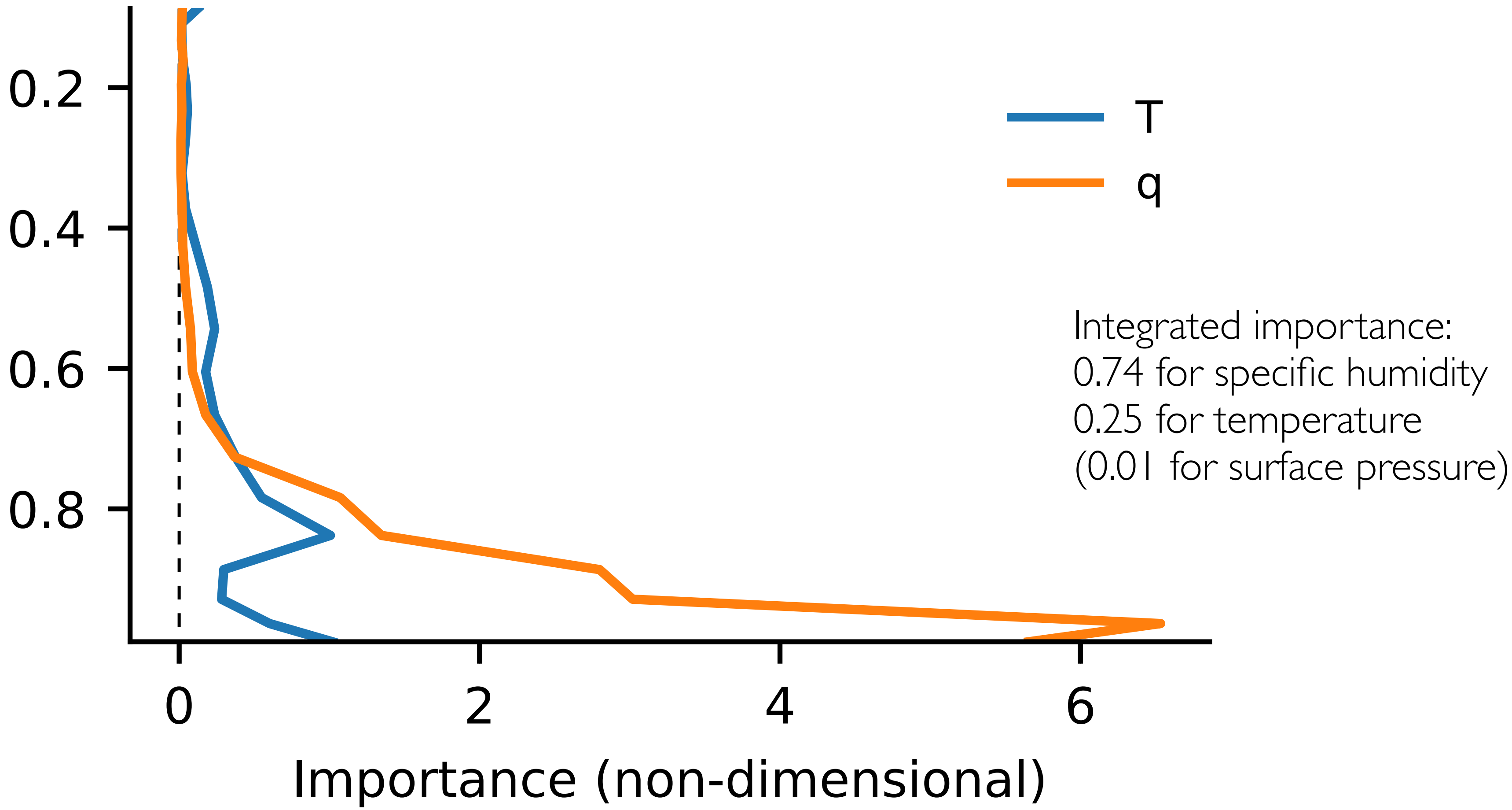
Similar to Mapes et al 2017
O'Gorman & Dwyer, submitted to JAMES

Feature importance of input temperature and moisture at different vertical levels



O'Gorman & Dwyer, submitted to JAMES

Feature importance of input temperature and moisture at different vertical levels



Doesn't require small perturbations (e.g., can apply it to question of whether convecting)

Conclusions

- Machine learning shows potential for parameterization of moist convection in "perfect-parameterization" tests

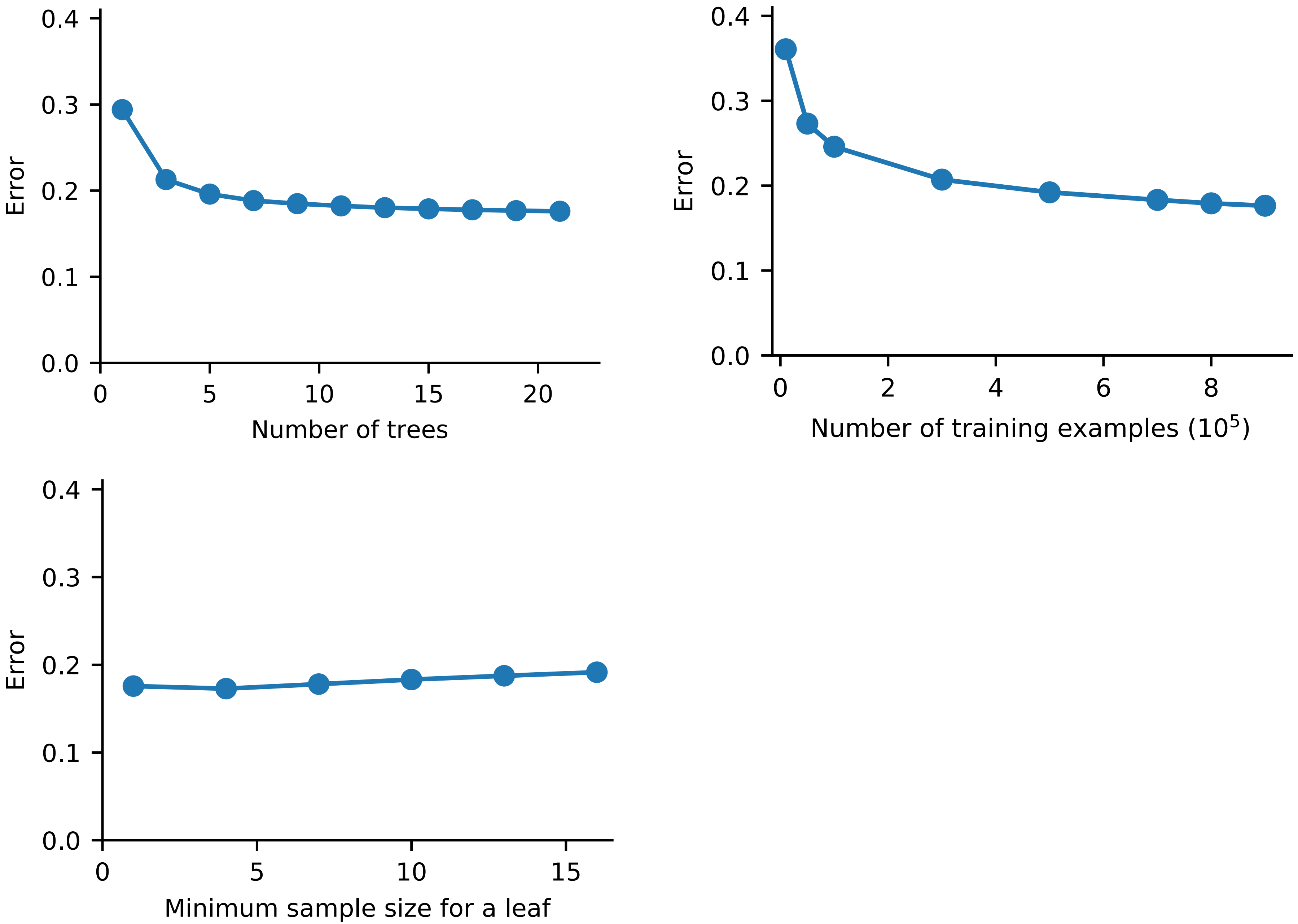
- Random forest respects physical constraints and lead to accurate climate simulations

- Need training samples from warmer climate to capture climate-change

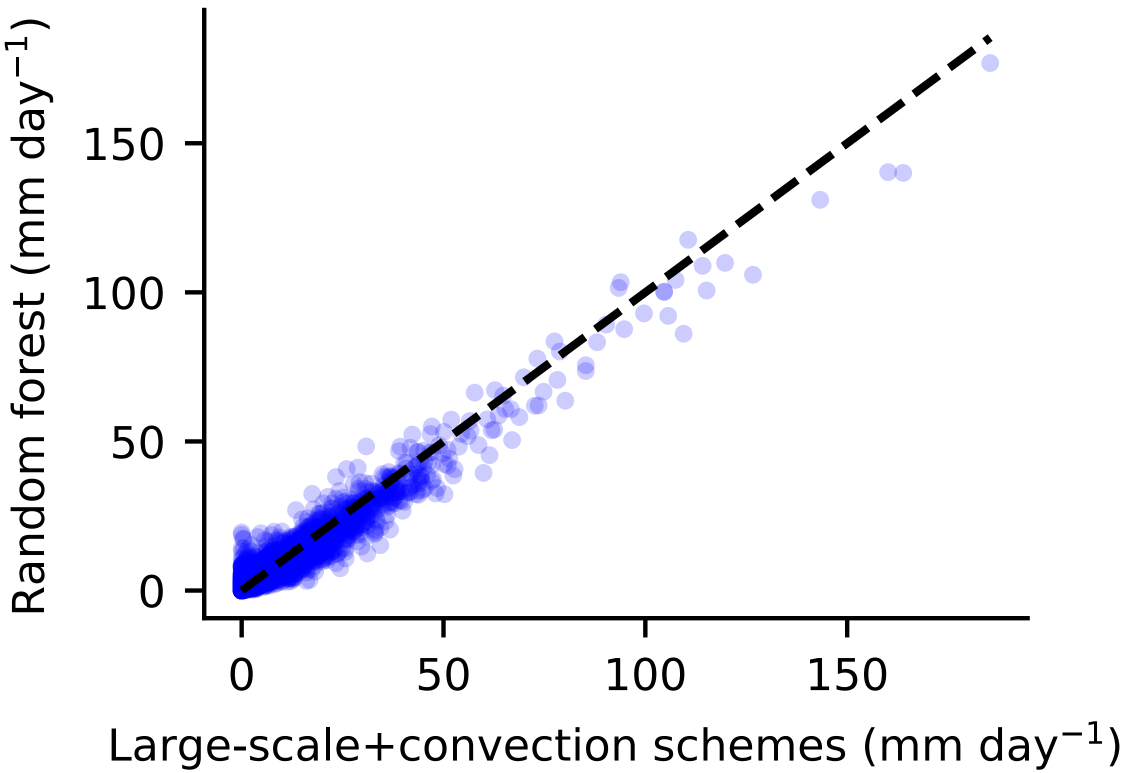
- Can use resulting nonlinear mapping to learn about interaction of convection with the large-scale environment

O'Gorman & Dwyer, submitted to JAMES

Cross-validation errors (1-R^2) versus hyperparameters



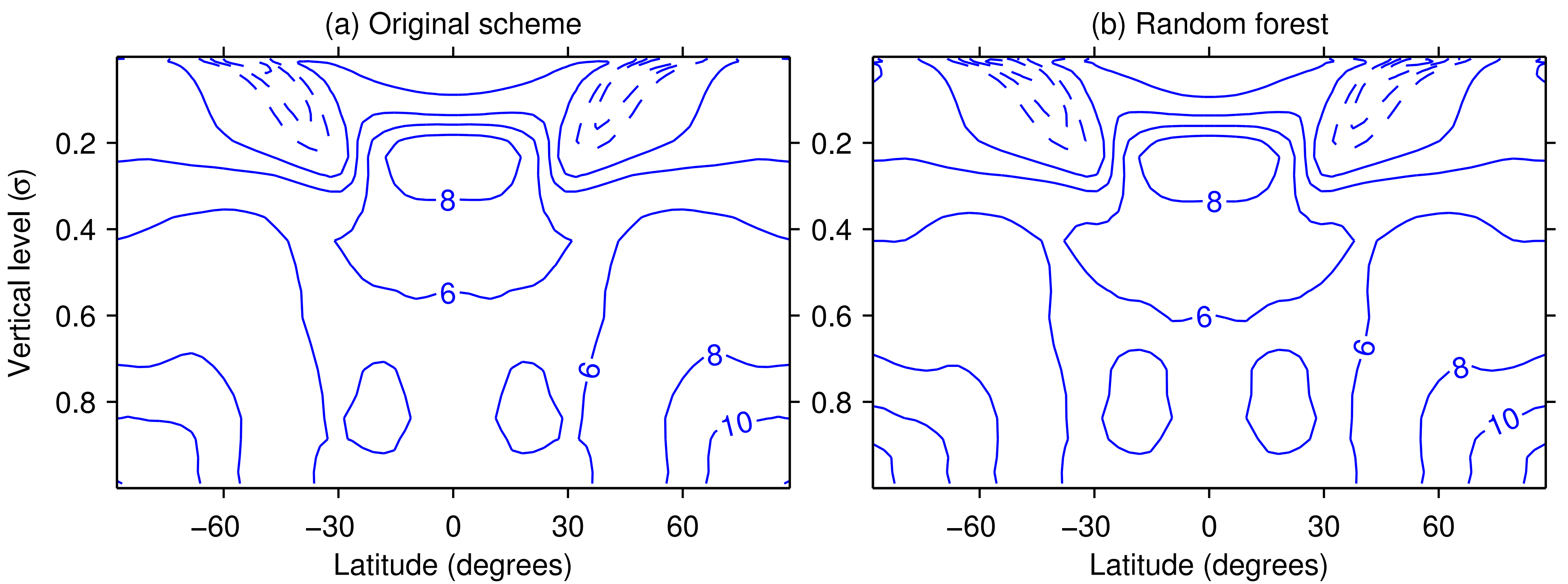
Also accurate when replace both the convection and large-scale condensation schemes



Scatterplot of instantaneous surface precipitation rate from original convection schemes (RAS+large-scale condensation) versus Random Forest

O'Gorman & Dwyer, submitted to JAMES

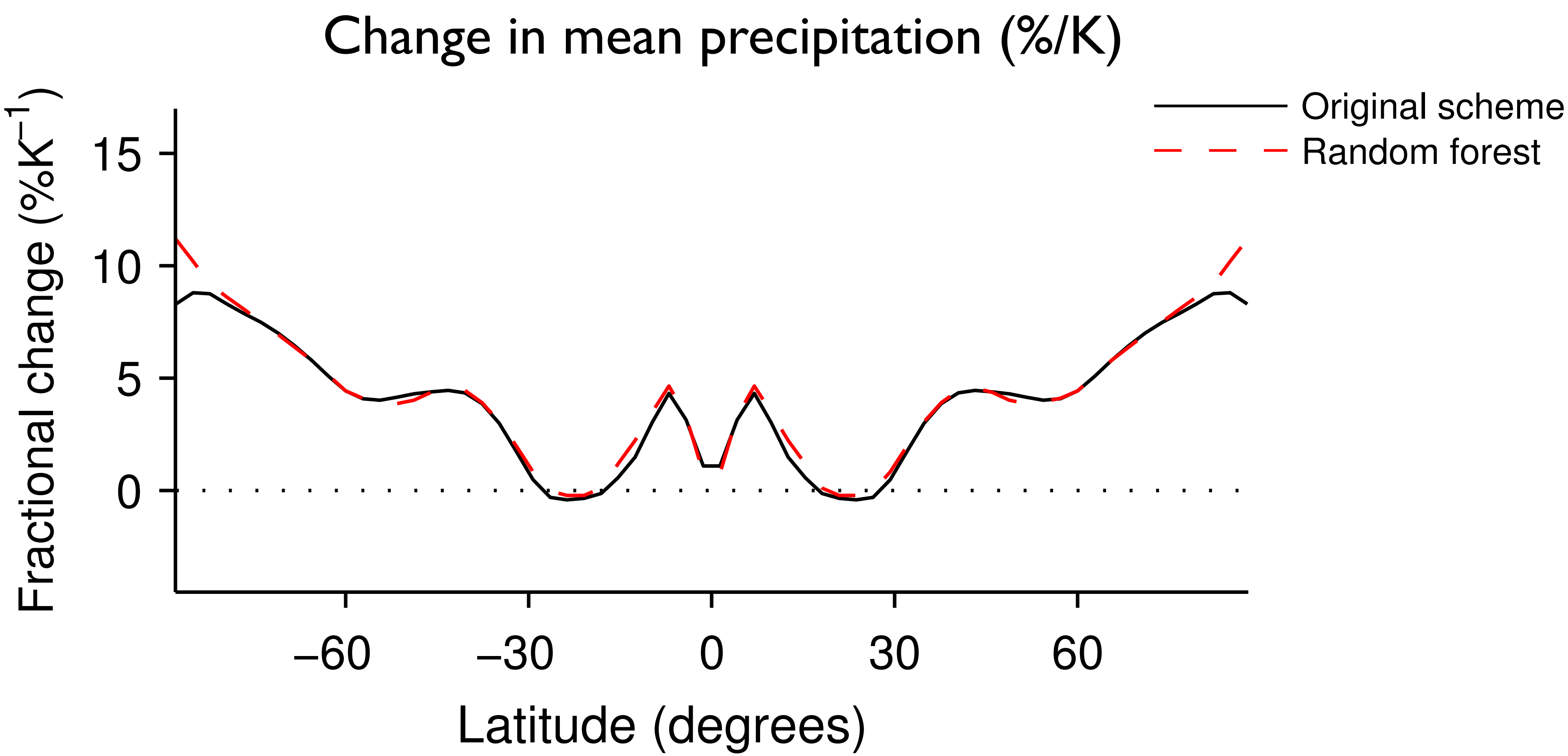
GCM with Random-Forest convection scheme correctly captures climate-change response of mean temperature



Change in mean temperature (K) with contour interval of 2K

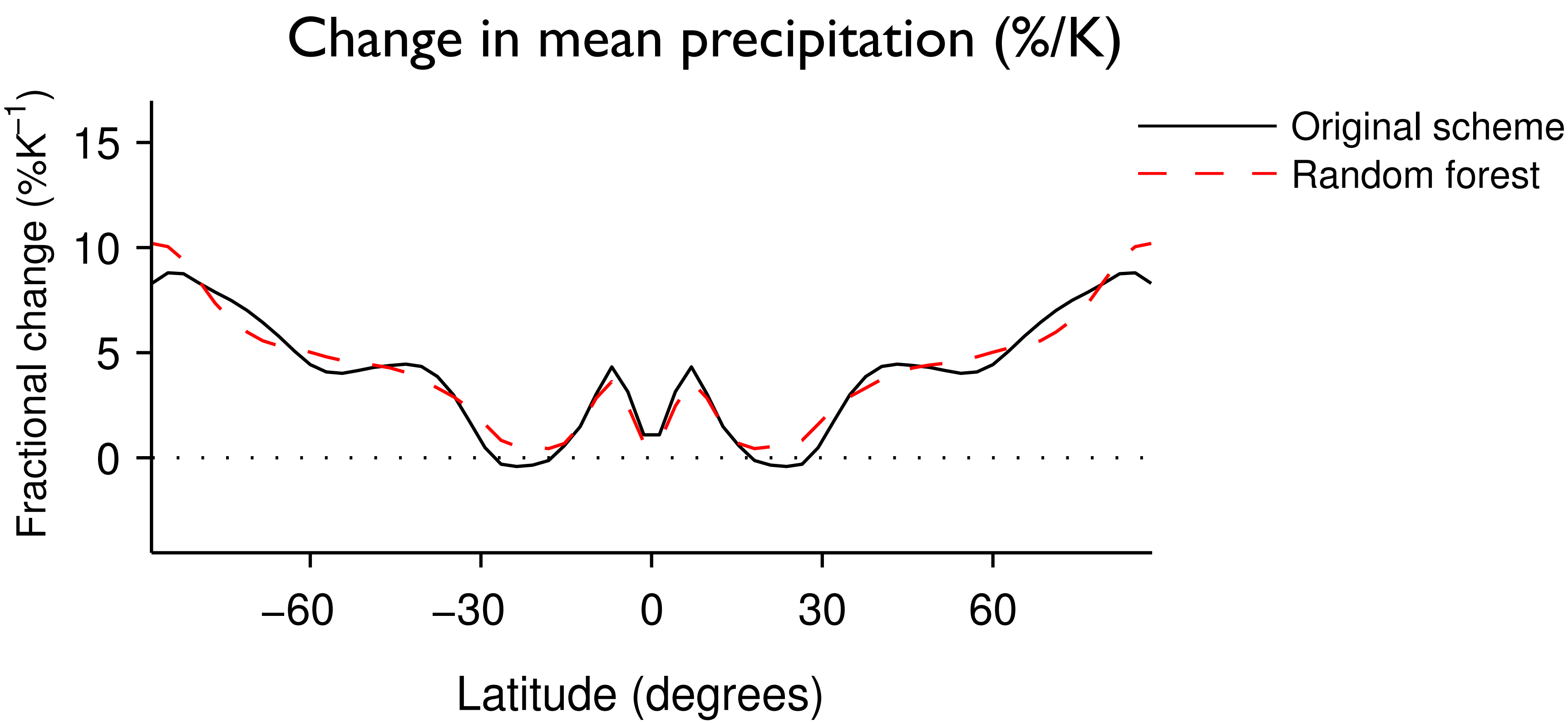
O'Gorman & Dwyer, submitted to JAMES

Training a Random Forest for each climate separately: Accurate simulations of climate change



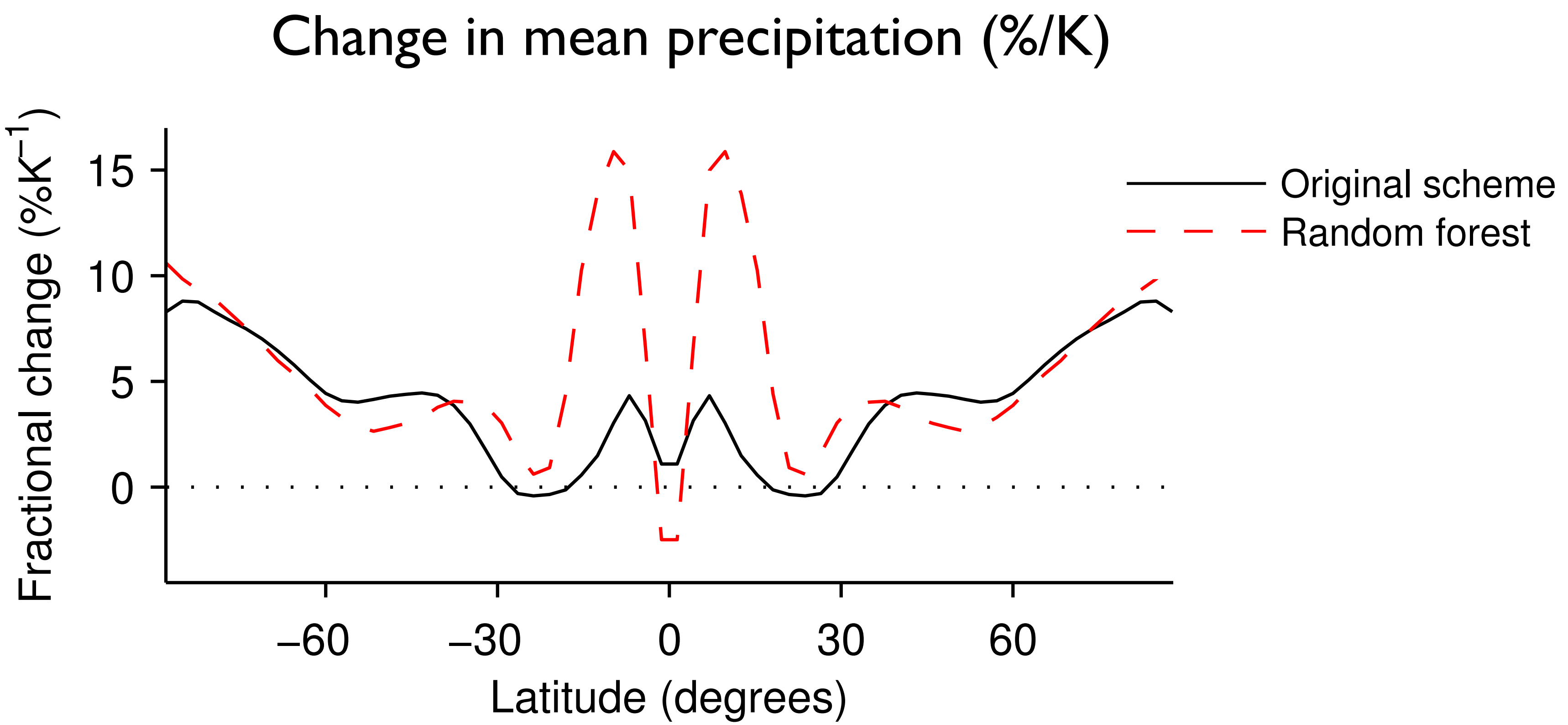
O'Gorman & Dwyer, submitted to JAMES

Training one Random Forest on combined samples from both climates: Accurate simulations of climate change



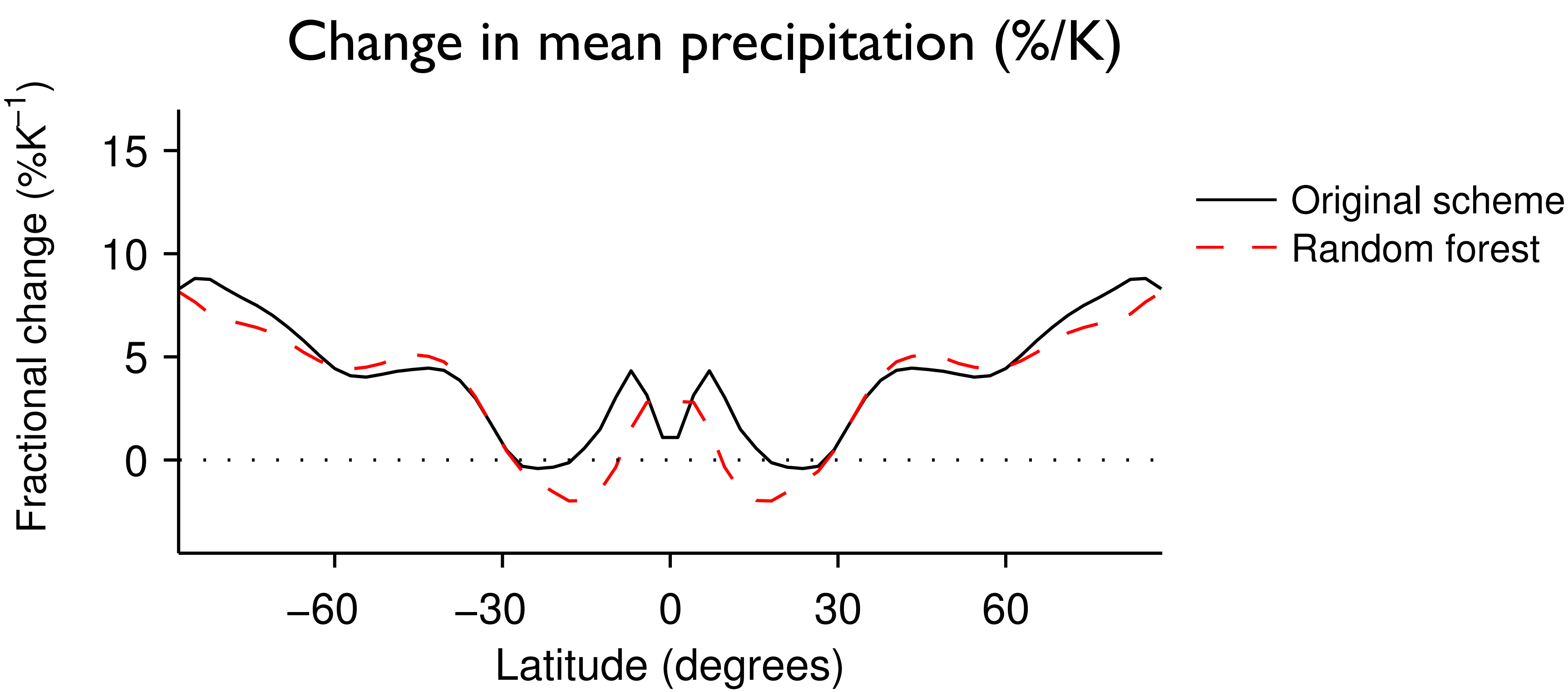
O'Gorman & Dwyer, submitted to JAMES

Training on control climate only: Very inaccurate simulations of climate change



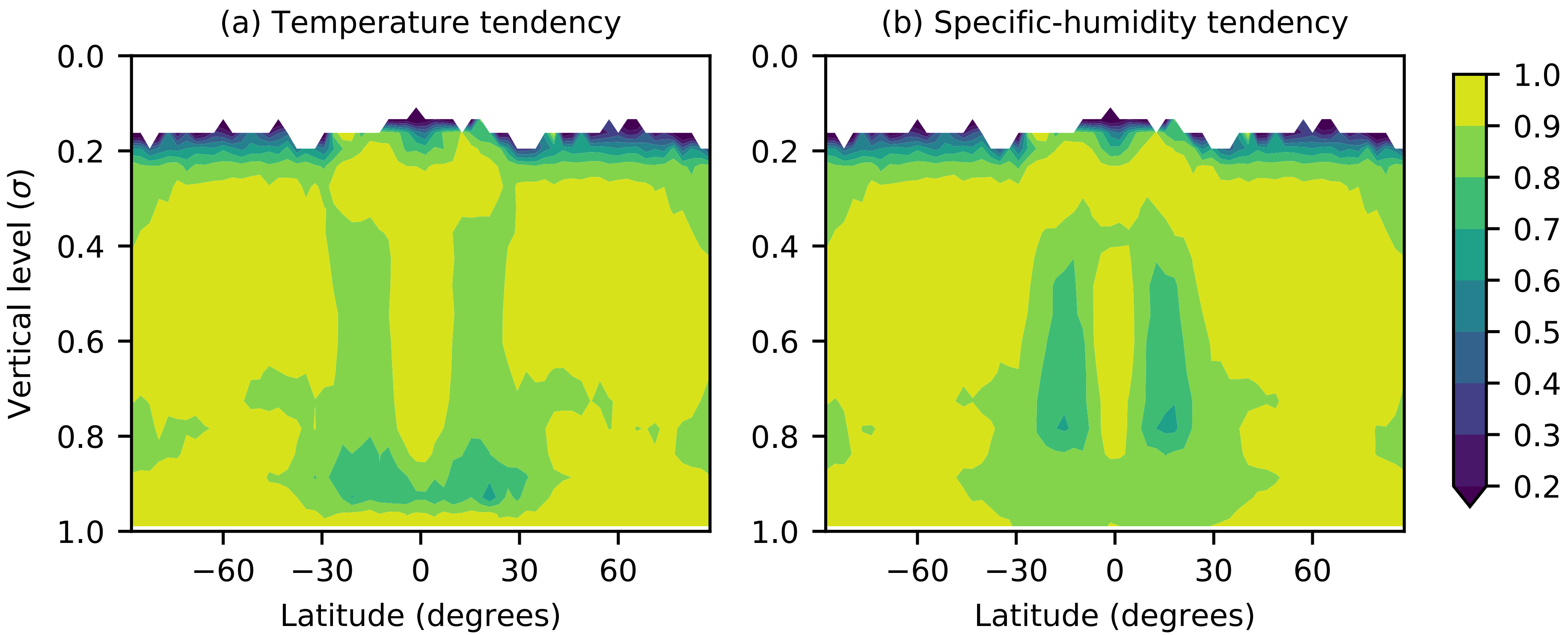
O'Gorman & Dwyer, submitted to JAMES

Training on warm climate only:
Does much better than when train on control climate only!



Asymmetry because can find training samples in warm climate at higher latitudes that match tropical samples in control climate (ultimately because of weaker temperature variability in tropics than extratropics)

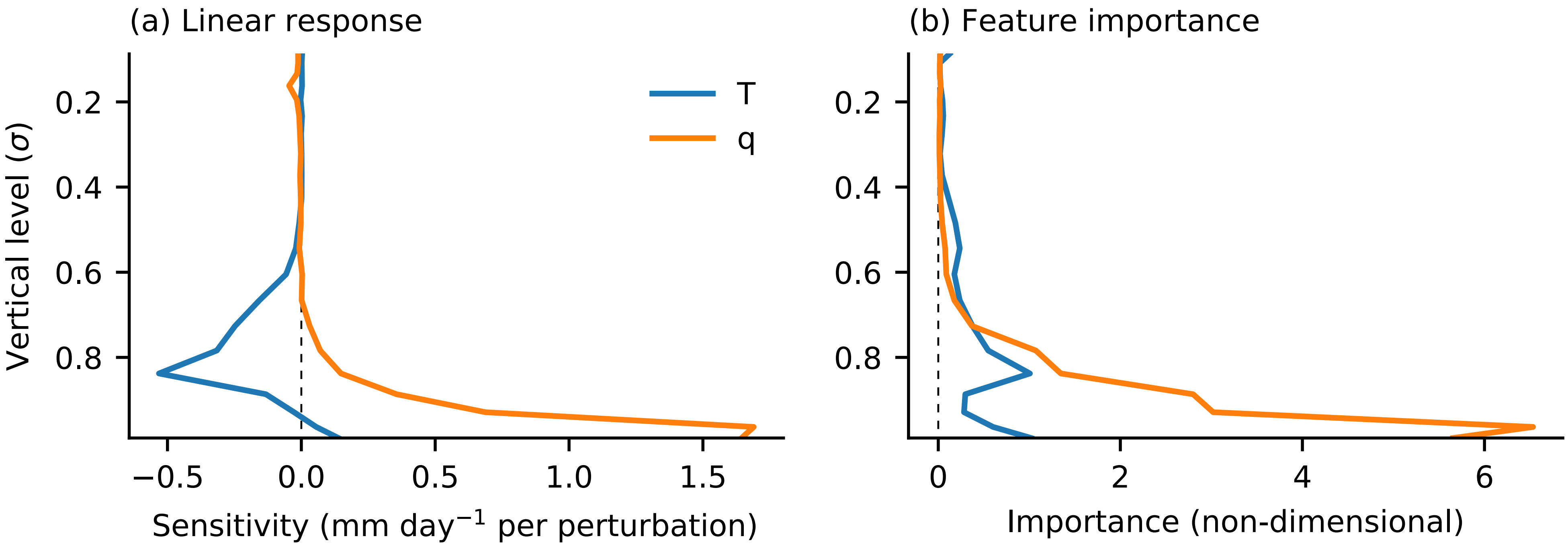
Random forest accurately predicts convection+large-scale condensation tendencies in test dataset



Correlation coefficients when train on convection+large-scale condensation tendencies using one Random Forest

O'Gorman & Dwyer, submitted to JAMES

Some similar information in linear response and feature importance, but feature importance doesn't require small perturbations



O'Gorman & Dwyer, submitted to JAMES