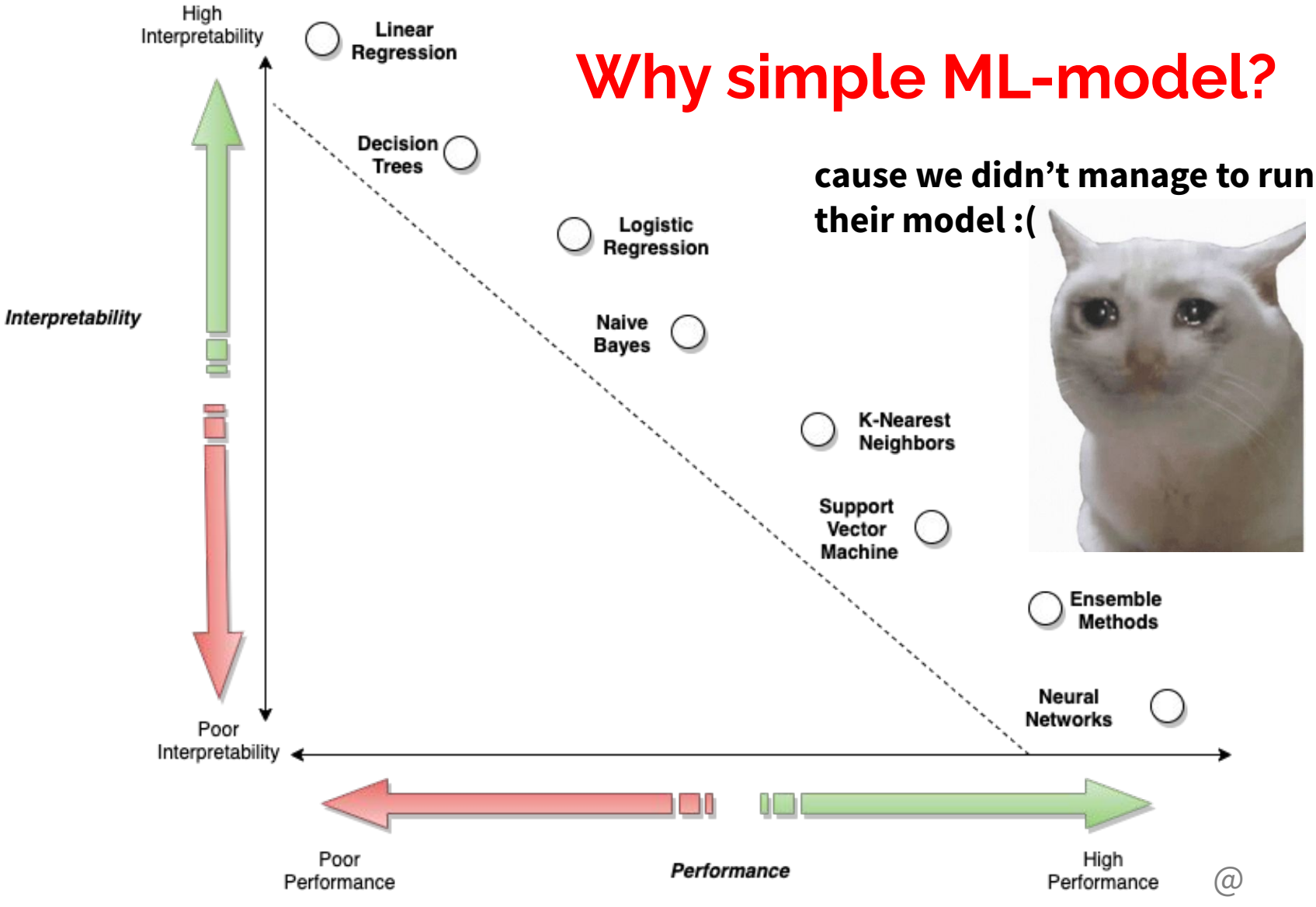
ML-based Convective Parameterization

Jawairia Ahmad, Collin Victor, Lloyd Villanueva, Siddharth

Rain or no rain?: A simple ML-model to predict rain.

Why simple ML-model?



cause we didn't manage to run their model :(

@

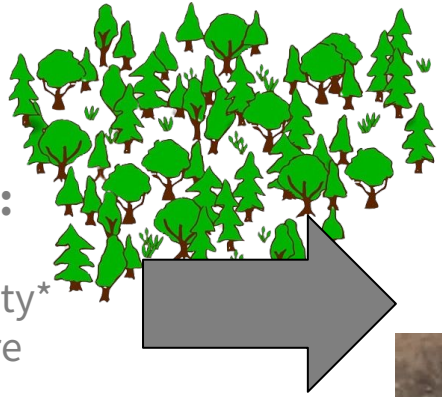
ML model pipeline

Input:

- Temperature*
- Specific humidity*
- Surface Pressure
- Insolation
- Surface latent heat flux
- Surface sensible heat flux

*with 60 vertical levels
60th -> surface

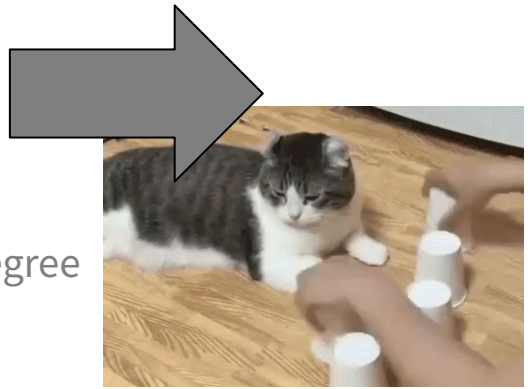
124 total features
resolution : 1.5 by 1.5 degree



Features selected:

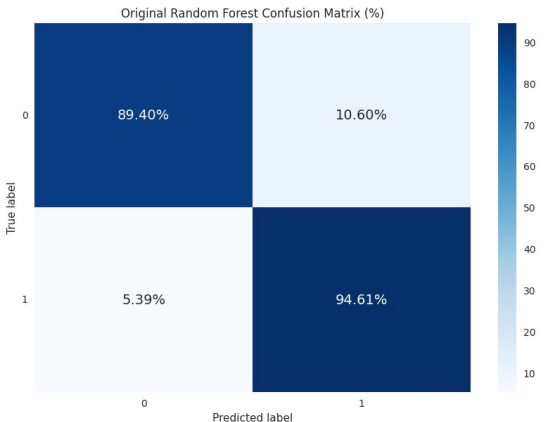
- Temperature*
- Specific humidity*
- Surface Pressure
- Insolation
- Surface latent heat flux
- Surface sensible heat flux

25 total features
*reduced levels

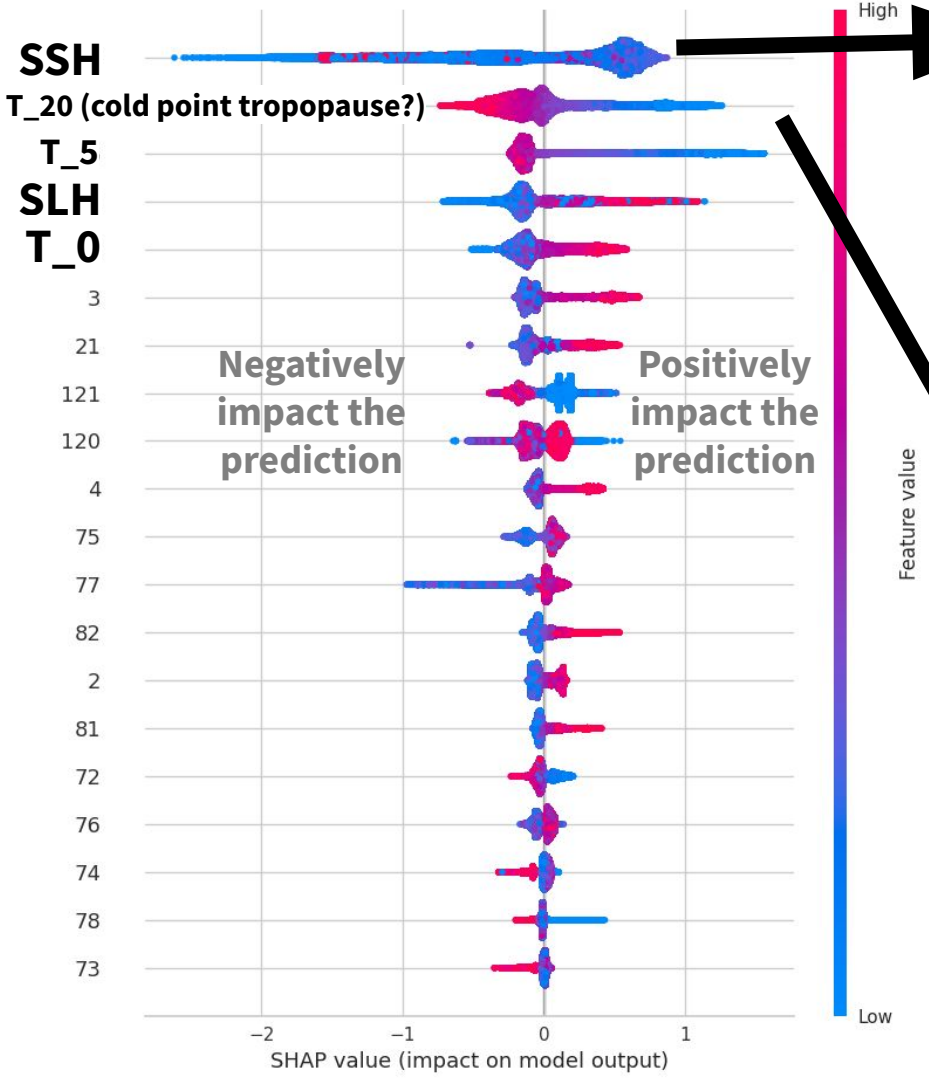


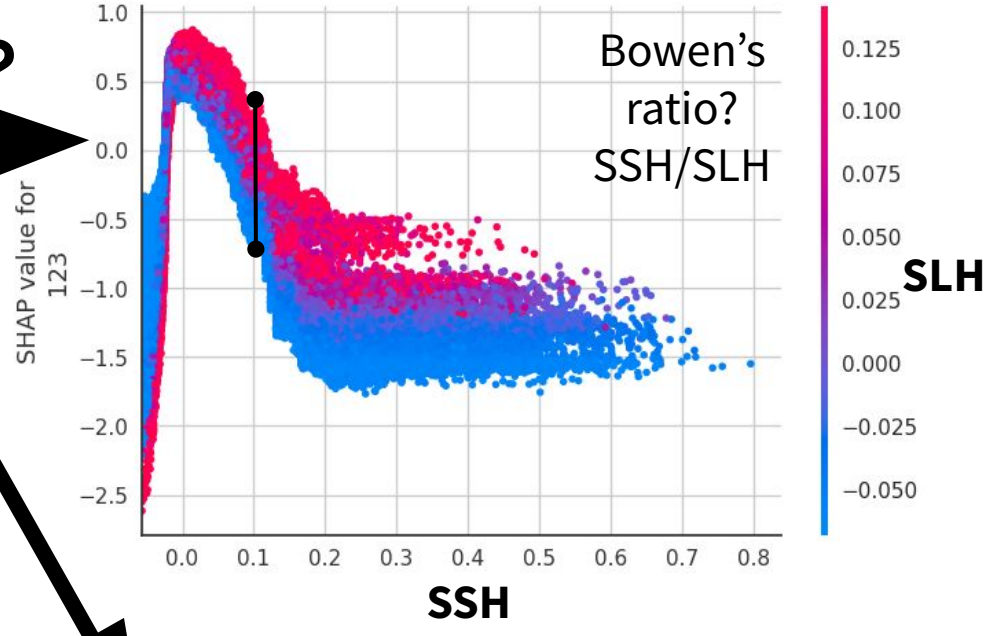
No rain

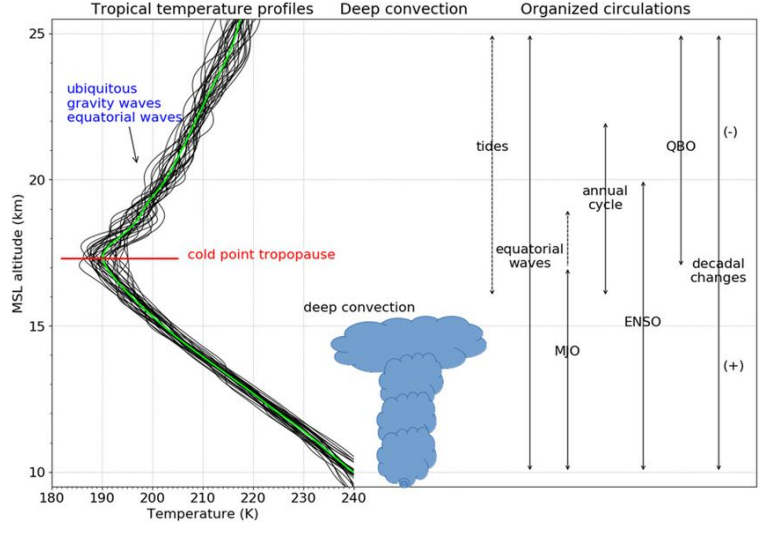
Rain



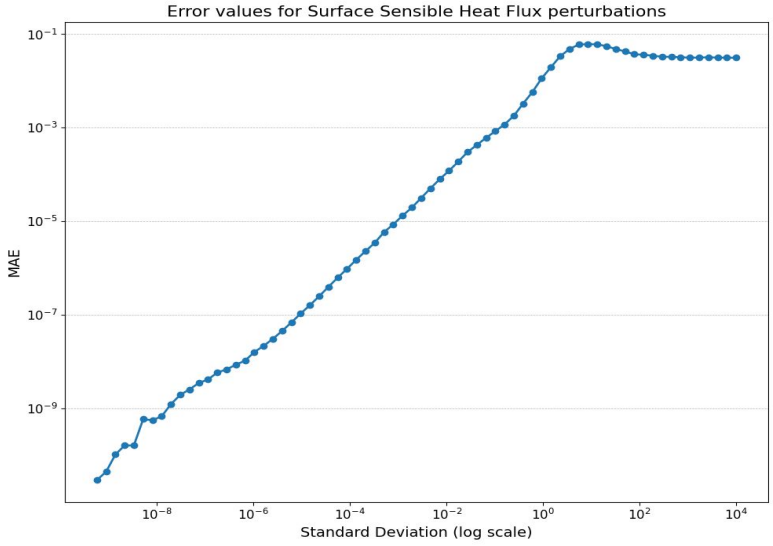
How the model decide?

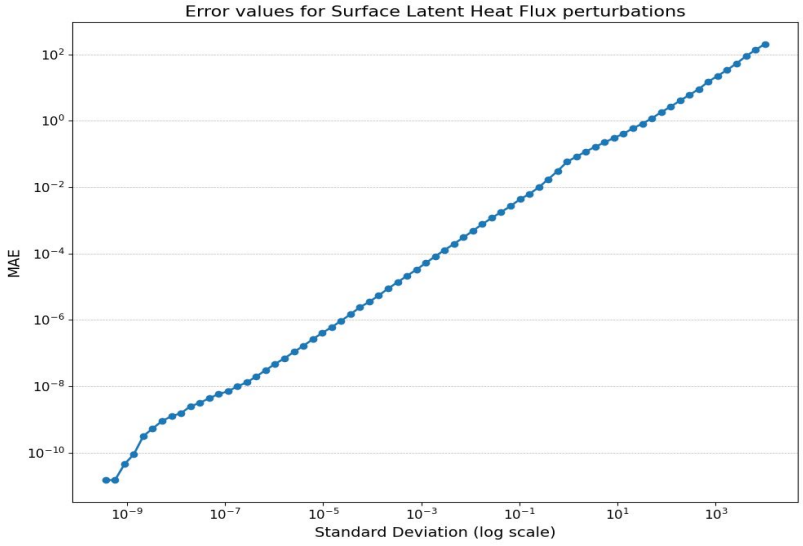


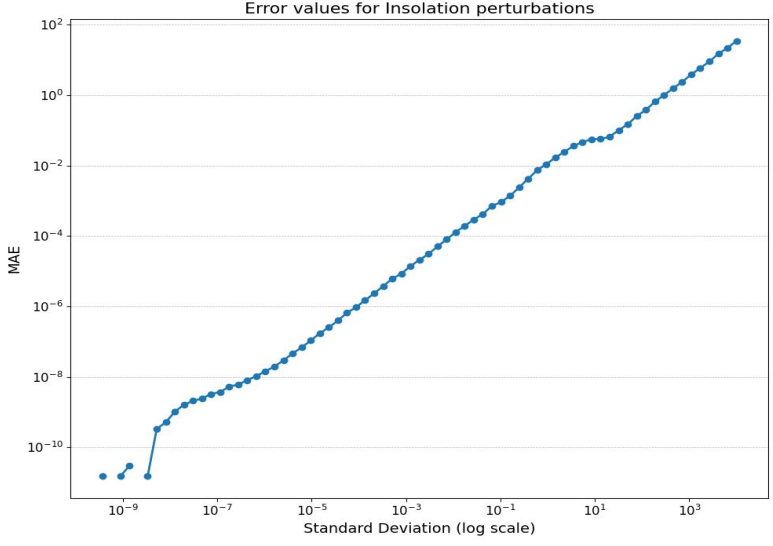


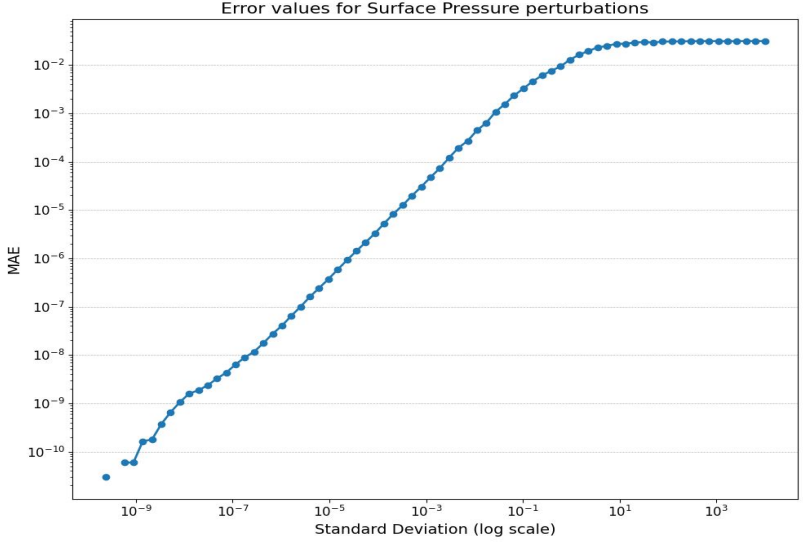


Perturbation Analysis

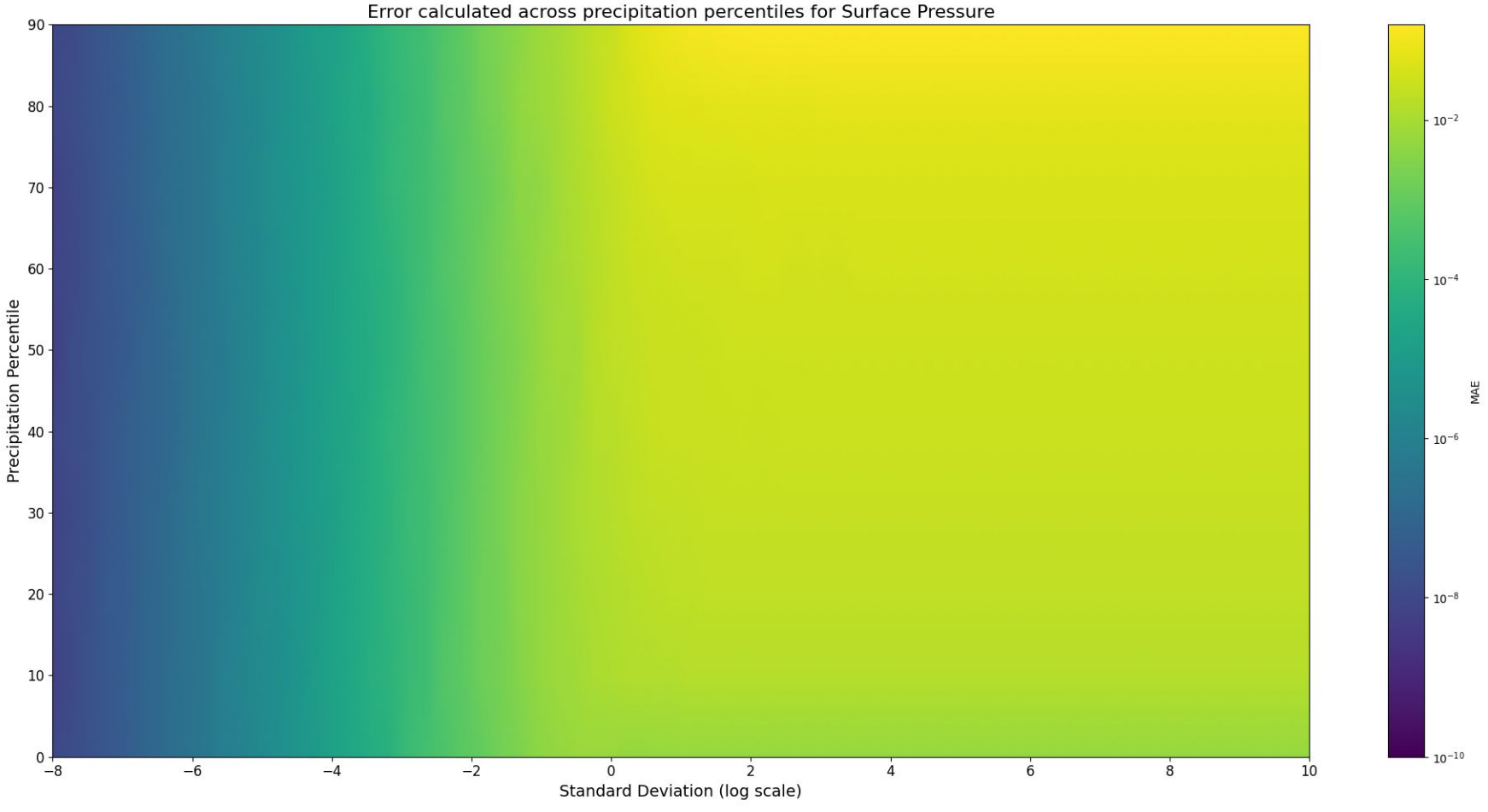

Error values for Surface Sensible Heat Flux perturbations
MAE
Standard Deviation (log scale)

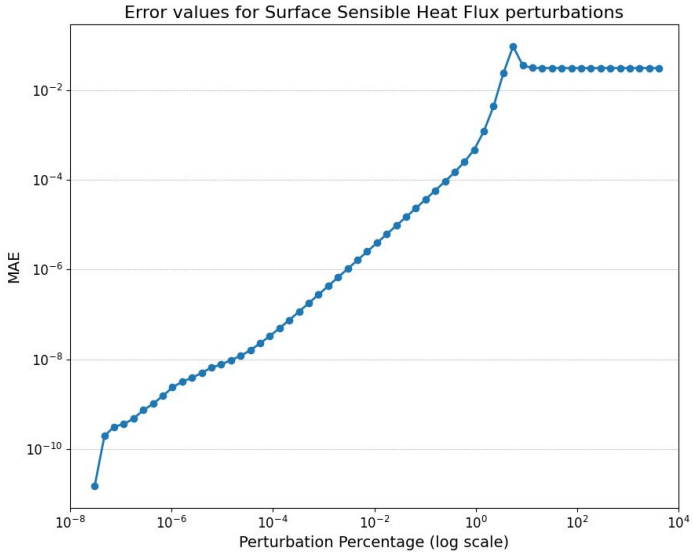

Error values for Surface Latent Heat Flux perturbations
MAE
Standard Deviation (log scale)

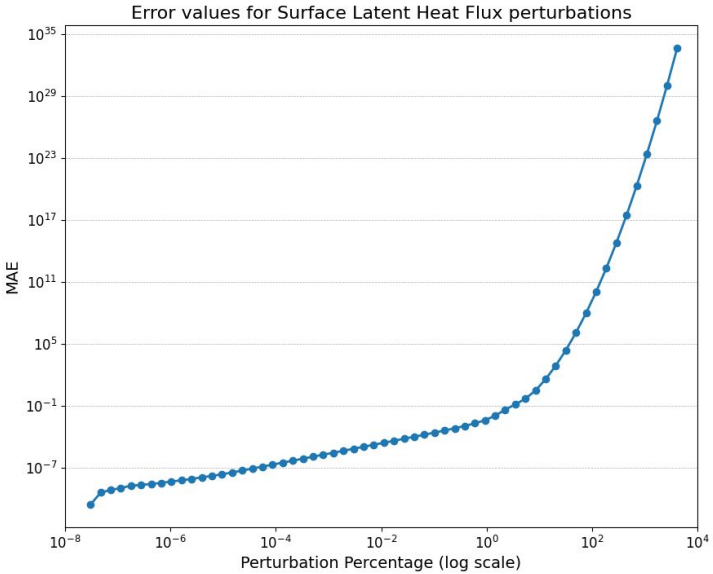

Error values for Insolation perturbations
MAE
Standard Deviation (log scale)

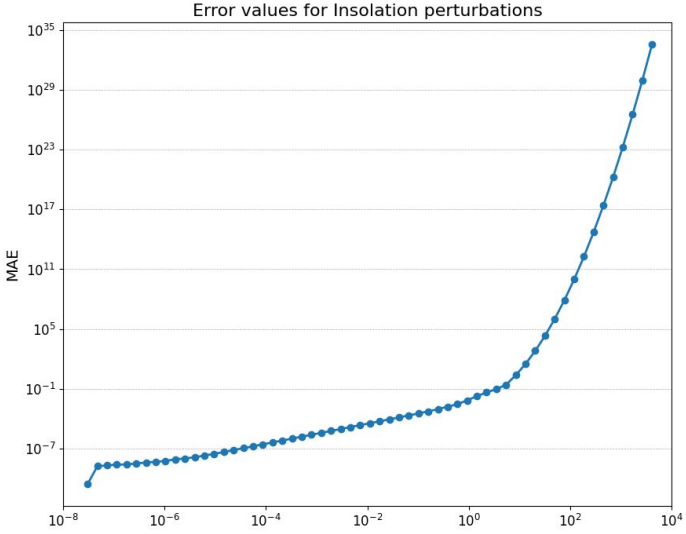

Error values for Surface Pressure perturbations
MAE
Standard Deviation (log scale)

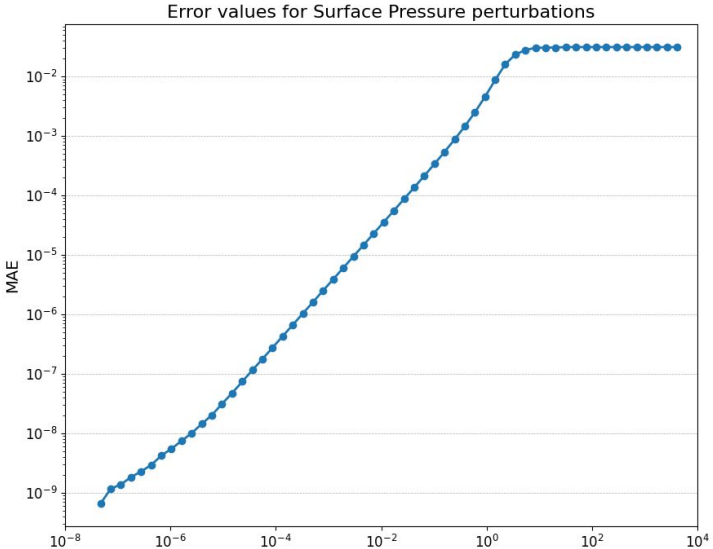
Perturbation Across Precipitation Percentiles



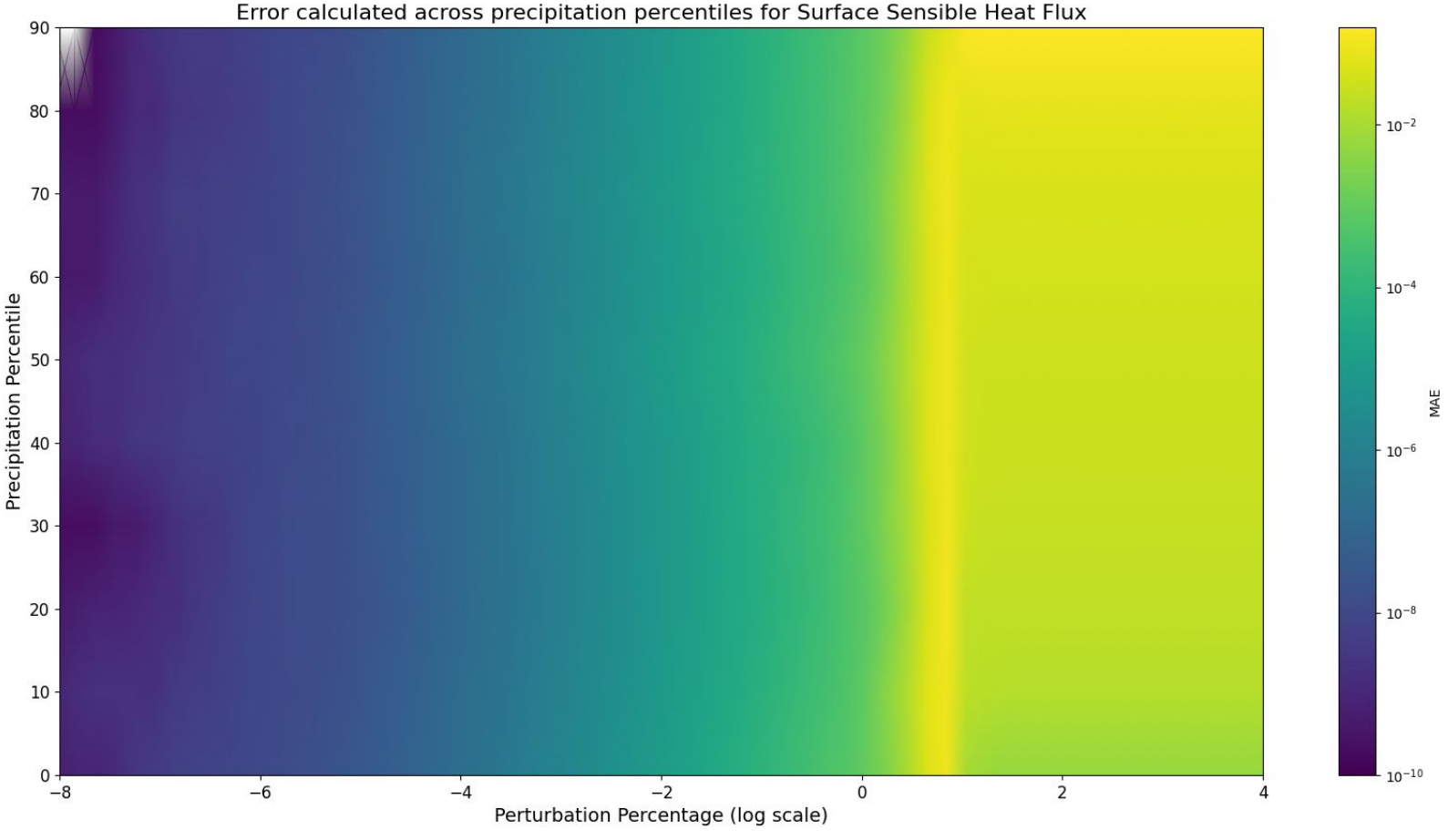

Error values for Surface Sensible Heat Flux perturbations
MAE
Perturbation Percentage (log scale)


Error values for Surface Latent Heat Flux perturbations
MAE
Perturbation Percentage (log scale)


Error values for Insolation perturbations
MAE


Error values for Surface Pressure perturbations
MAE

Perturbation Across Precipitation Percentiles



Rainfall prediction sensitivity to Latent and Sensible Heat Flux

Simple CNN model- ClimSim dataset

Layer (type)	Output Shape	Param #
conv1d_2 (Conv1D)	(None, 9, 32)	96
flatten_2 (Flatten)	(None, 288)	0
dense_4 (Dense)	(None, 10)	2,890
dense_5 (Dense)	(None, 1)	11

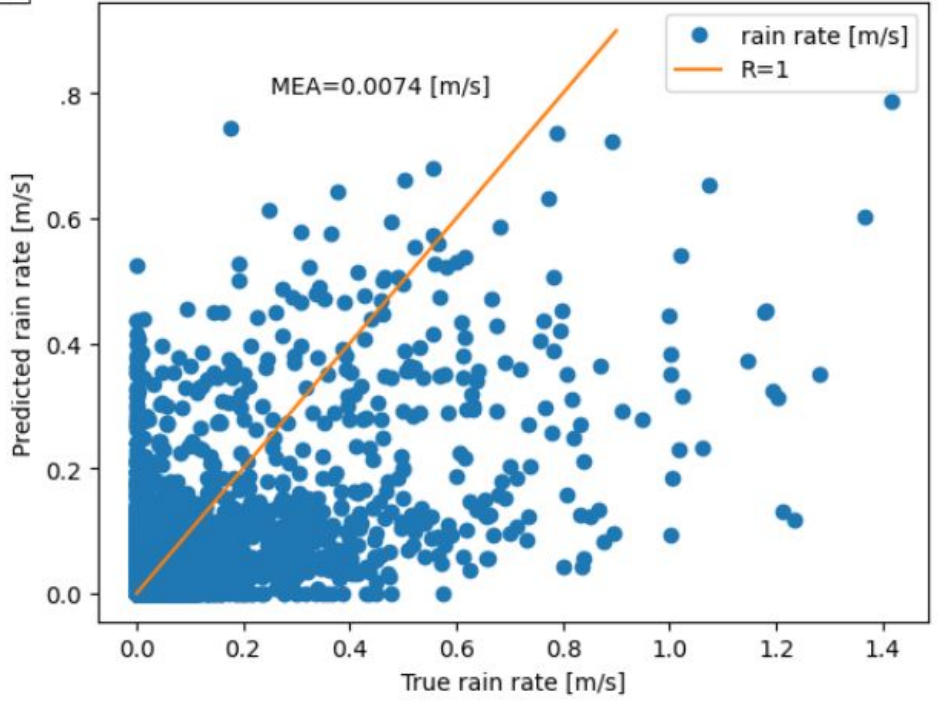
Total params: 2,997 (11.71 KB)
Trainable params: 2,997 (11.71 KB)
Non-trainable params: 0 (0.00 B)

1) Training with <u>constant</u> learning rate on >1million samples
2) Training with <u>changing</u> learning rate on >70K samples

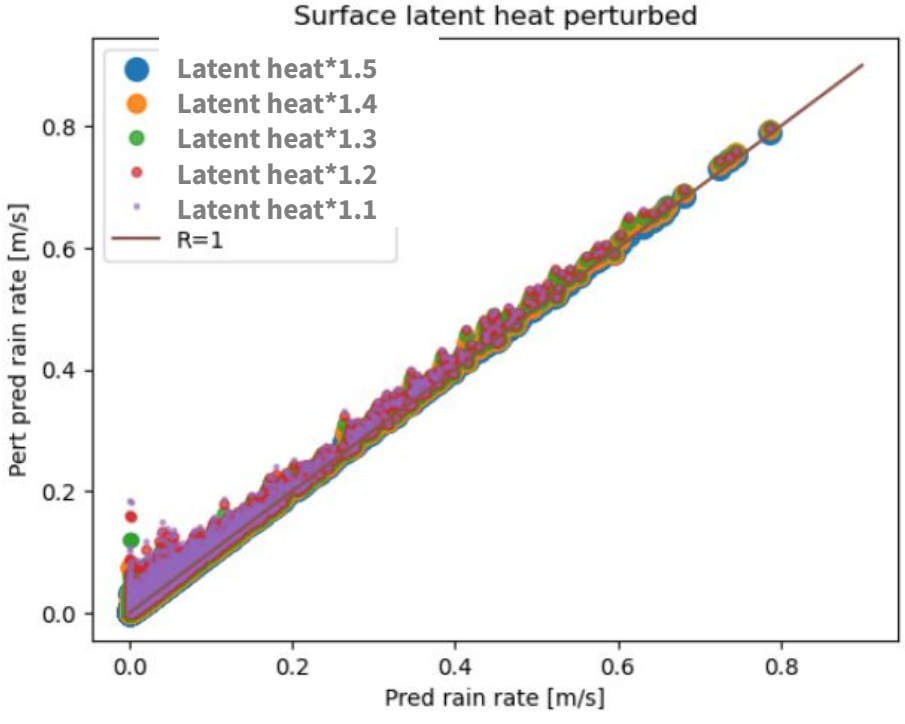
Inputs

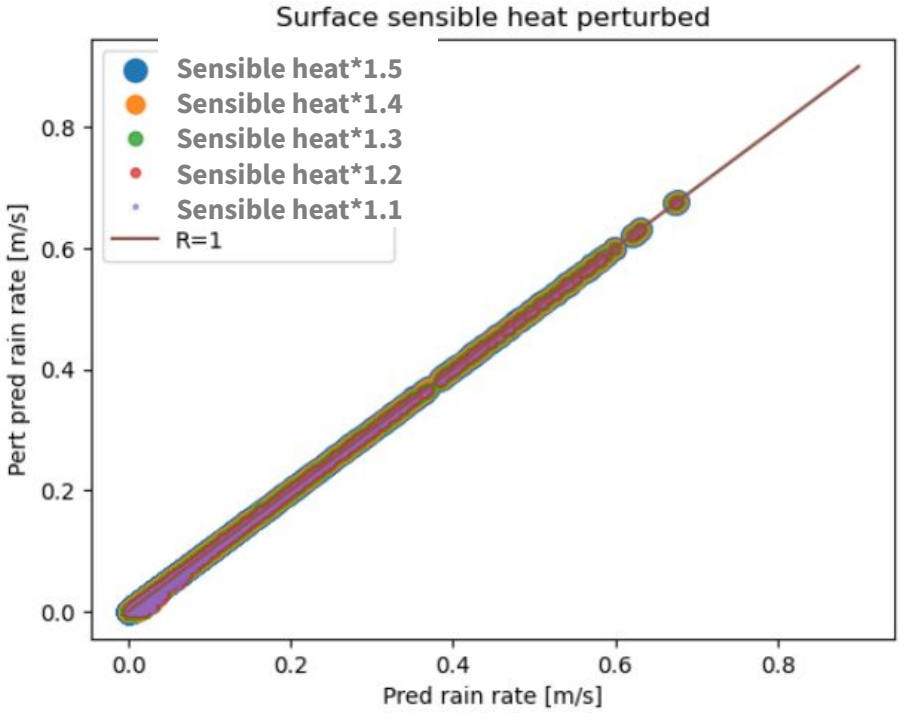
1. Temp surface
2. Temp mid
3. Temp top
4. Sp humidity surface
5. Sp humidity mid
6. Sp humidity top
7. Surface pressure
8. Insolation
9. Surface latent heat
10. Surface sensible heat

Target
Rain rate



Sensitivity to latent and sensible heat flux





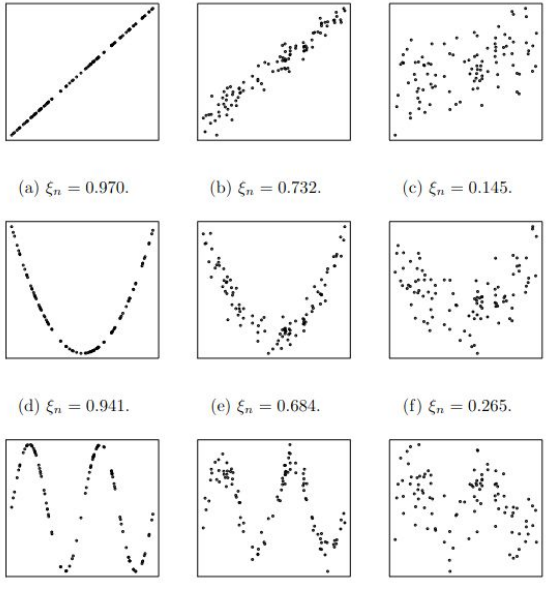
Predicted rain rate is more sensitive to latent heat!

Feature Selection for Precipitation Parametrization- ERA5

Feature Selection for Precipitation Parametrization

Input: Temperature, Specific humidity, Winds (low, mid and upper troposphere), Solar insolation, Sensible heat, Latent heat, Surface Pressure

Method: Mutual Information and Non-linear Correlation (Chatterjee, 2020)



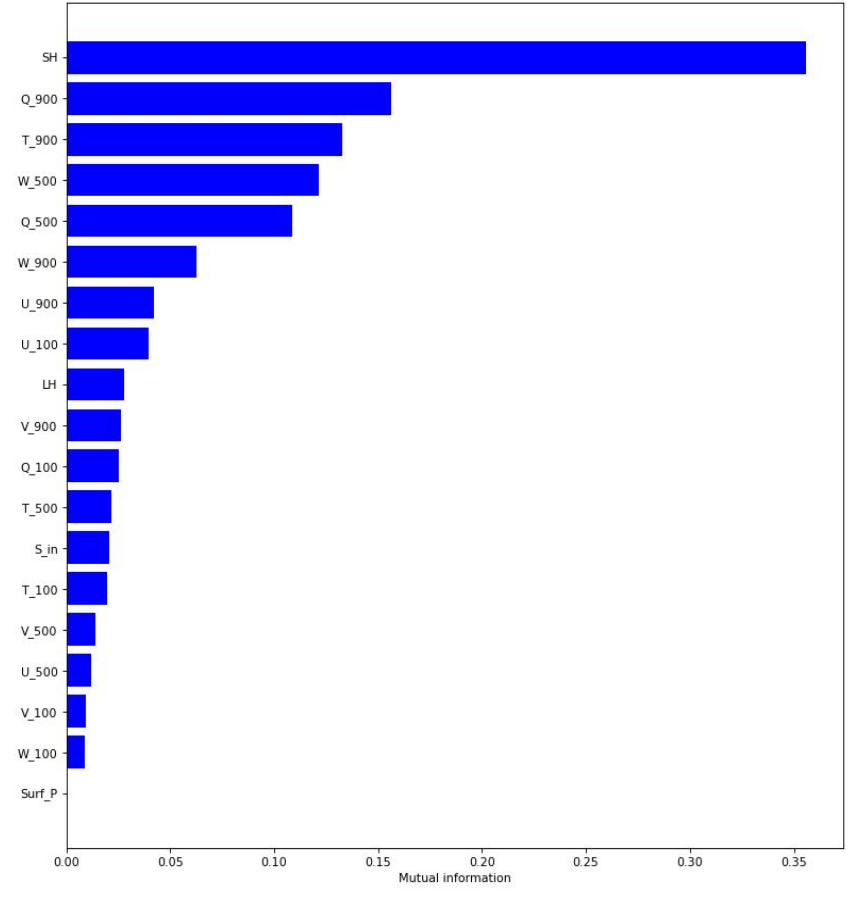
(a) $\xi_n = 0.970$. (b) $\xi_n = 0.732$. (c) $\xi_n = 0.145$.

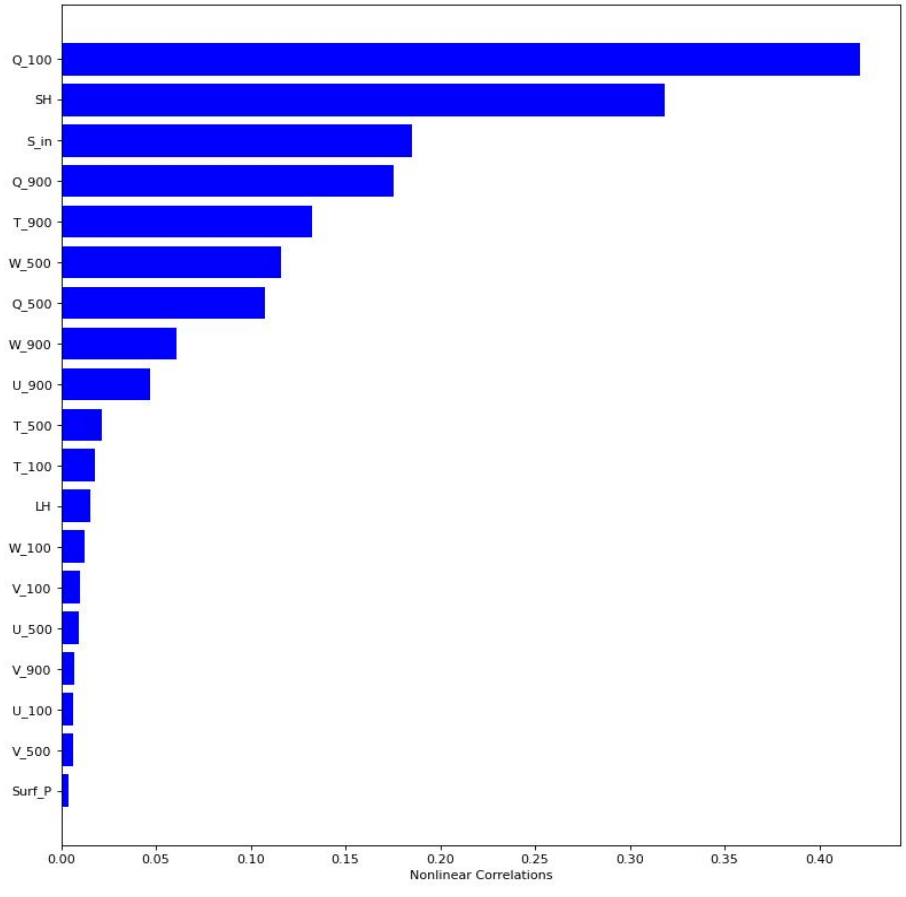
(d) $\xi_n = 0.941$. (e) $\xi_n = 0.684$. (f) $\xi_n = 0.265$.

$$I(X;Y) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X)$$
$$= \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)\,p(y)}$$

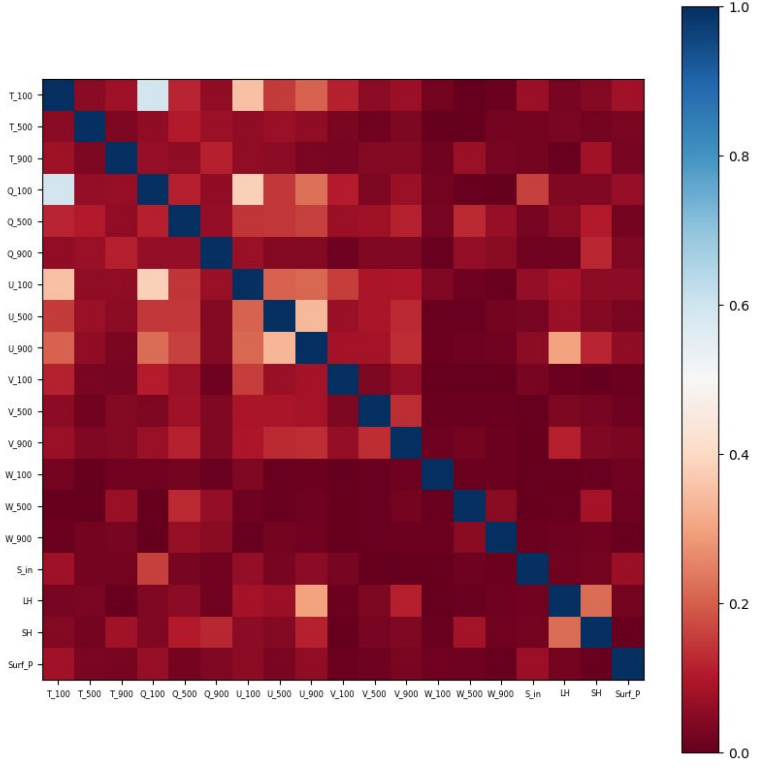
$$\xi_n(X,Y) := 1 - \frac{3 \sum_{i=1}^{n-1} |r_{i+1} - r_i|}{n^2 - 1}$$

Feature Importance based on Mutual Information and Nonlinear Correlations





Mutual Information among input variables



Thanks to Sara Shamekh!