Methods and Tools for the Application of the UF-ECT to New Climate Models

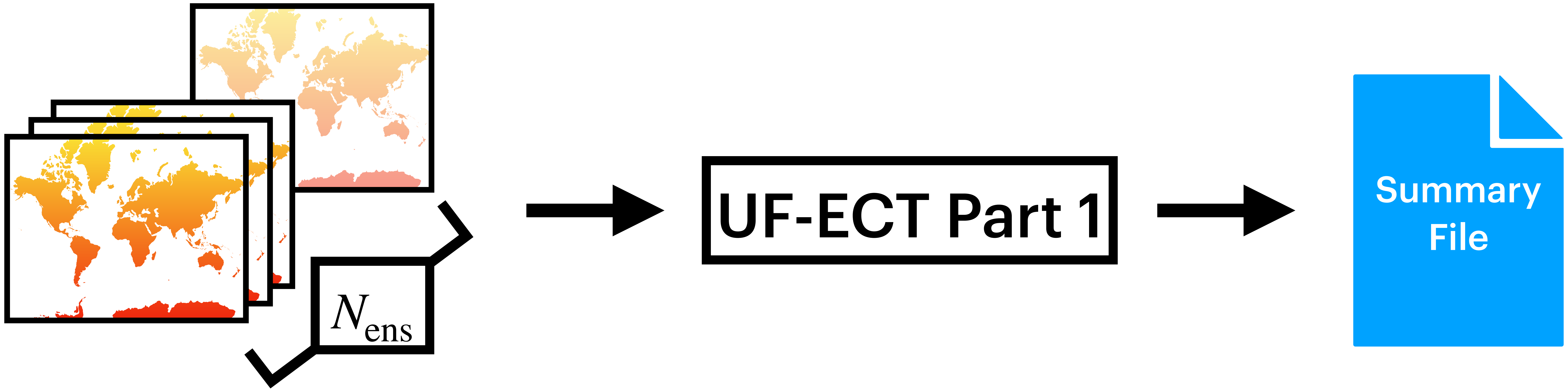
Teo Price-Broncucia, Allison Baker, and Dorit Hammerling

Purpose of UF-ECT

Identify inconsistencies between large scientific model outputs run in different configurations.

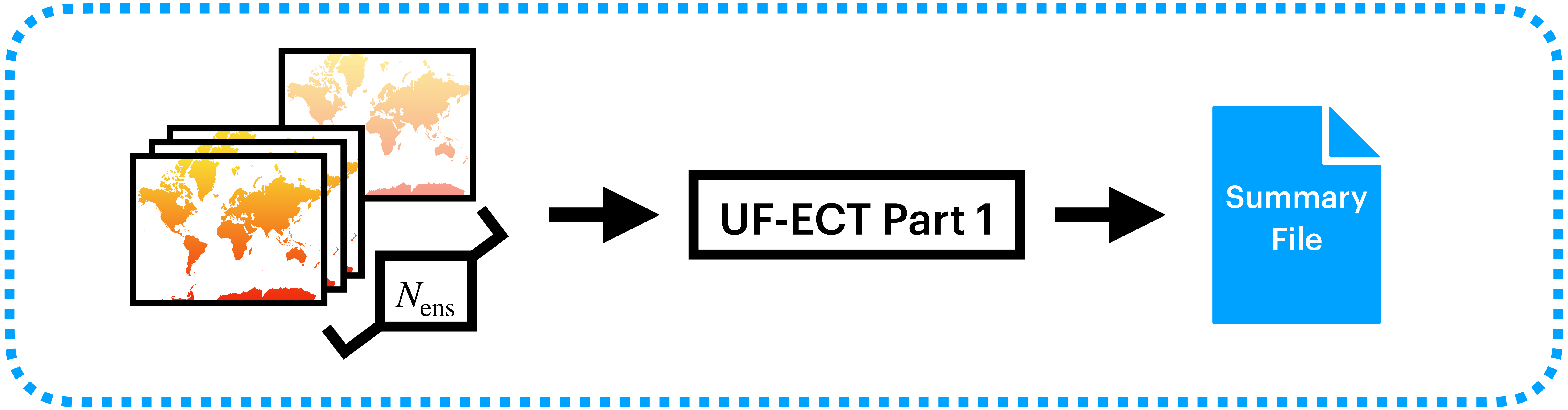
The UF-ECT: Overview

1) Characterize model variability
using large ensemble of model runs.



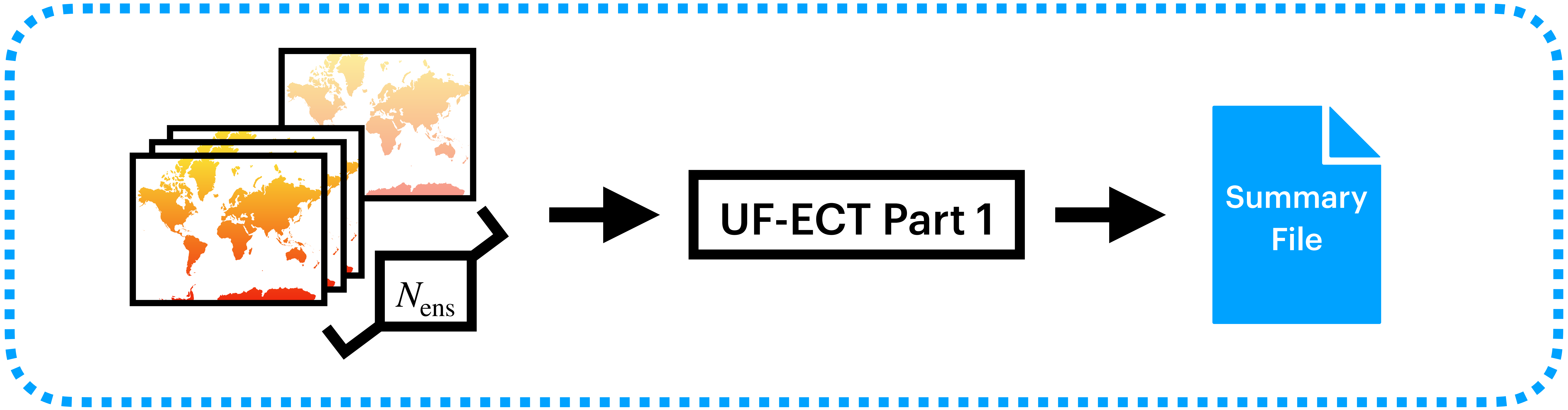
The UF-ECT: Overview

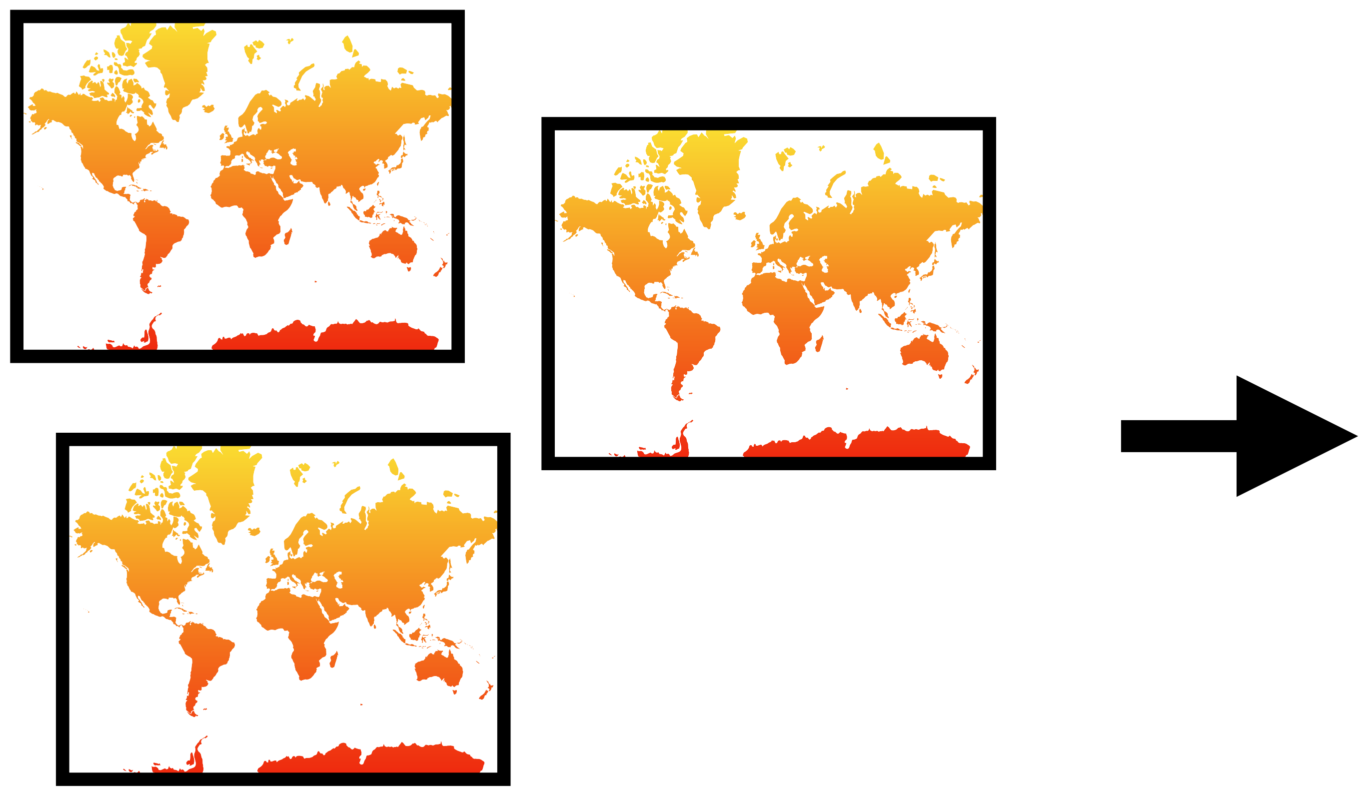
1) Characterize model variability using large ensemble of model runs.



The UF-ECT: Overview

1) Characterize model variability using large ensemble of model runs.

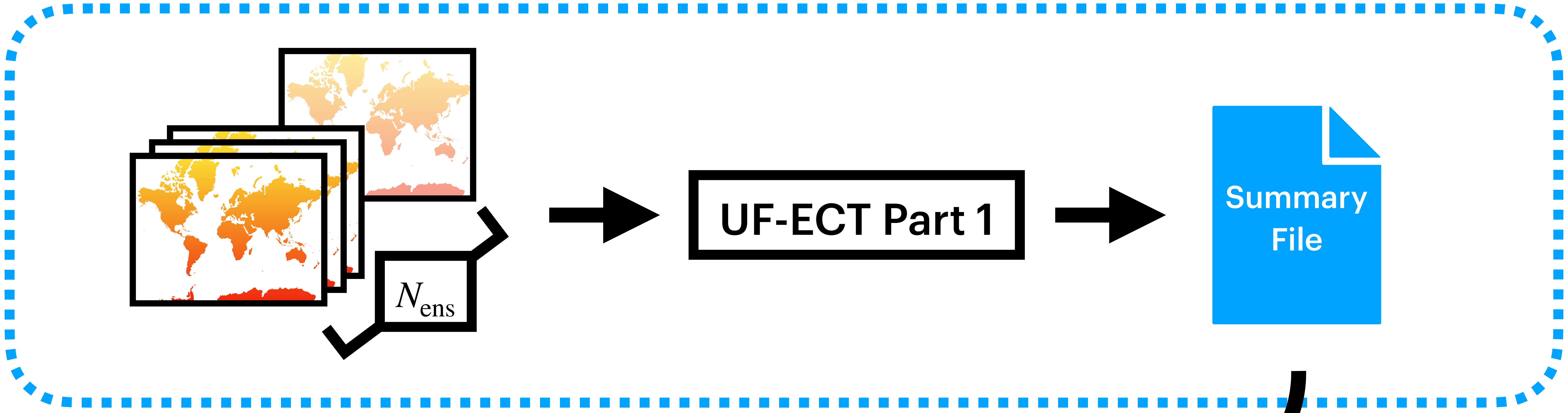


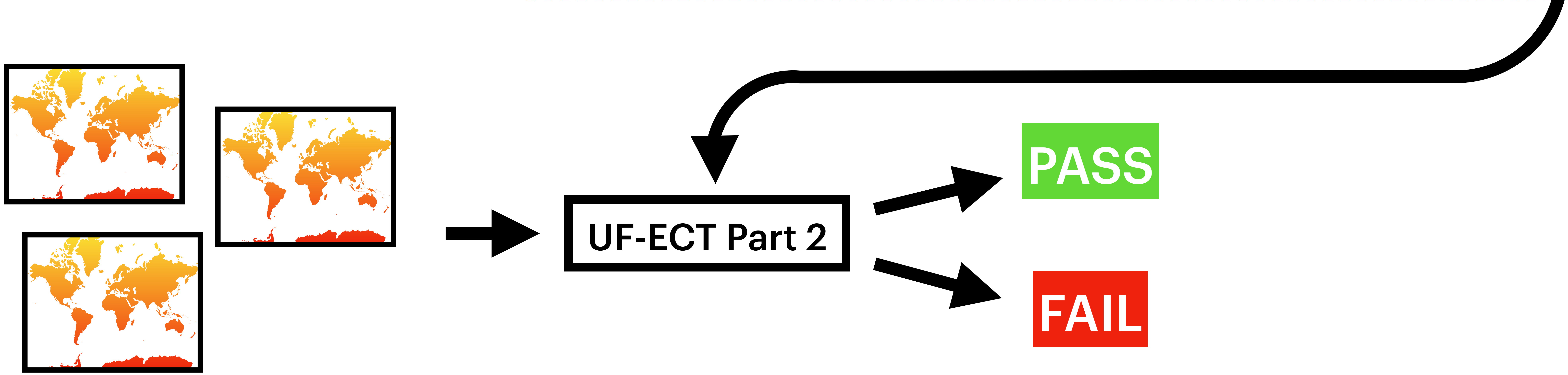


2) Test new configurations using a small set of model runs.

The UF-ECT: Overview

1) Characterize model variability using large ensemble of model runs.

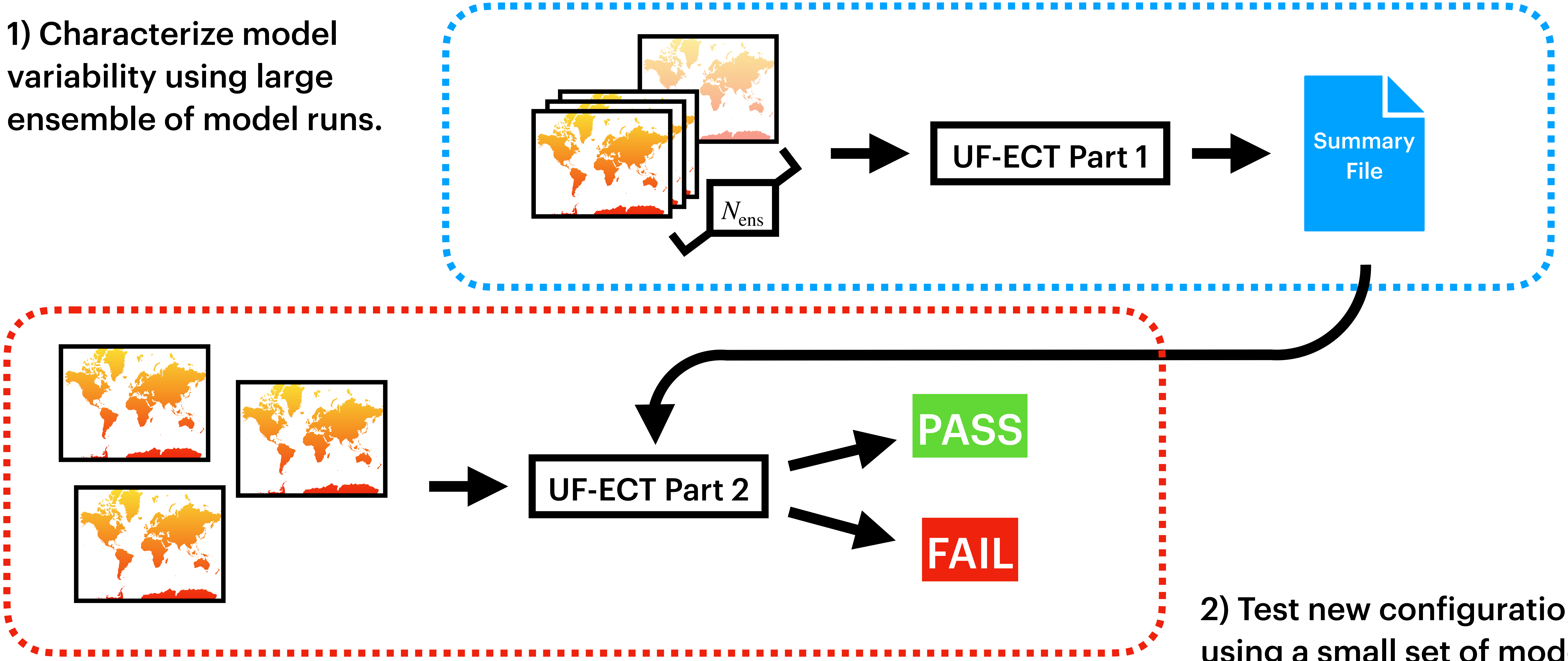




2) Test new configurations using a small set of model runs.

The UF-ECT: Overview

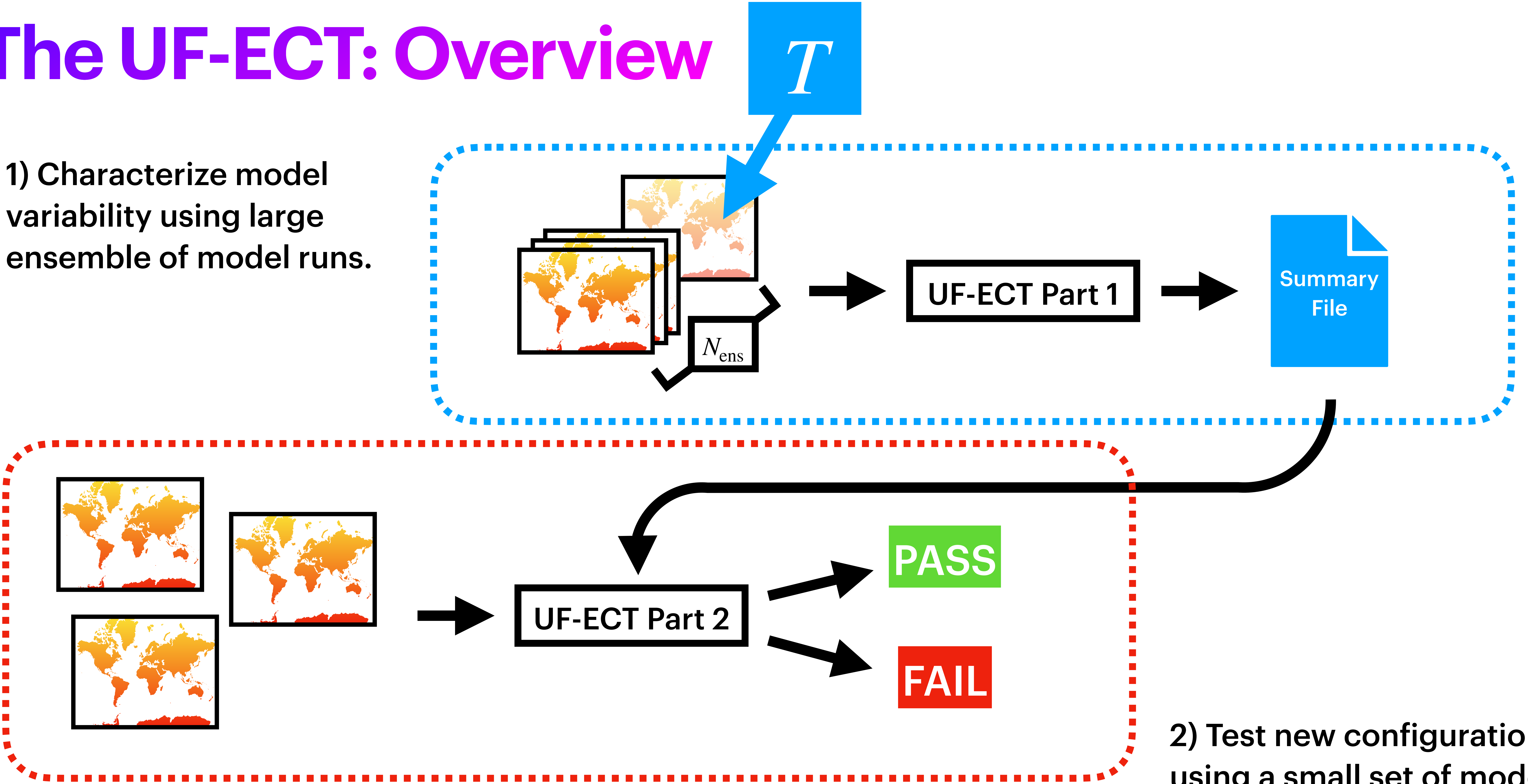
1) Characterize model variability using large ensemble of model runs.



2) Test new configurations using a small set of model runs.

The UF-ECT: Overview

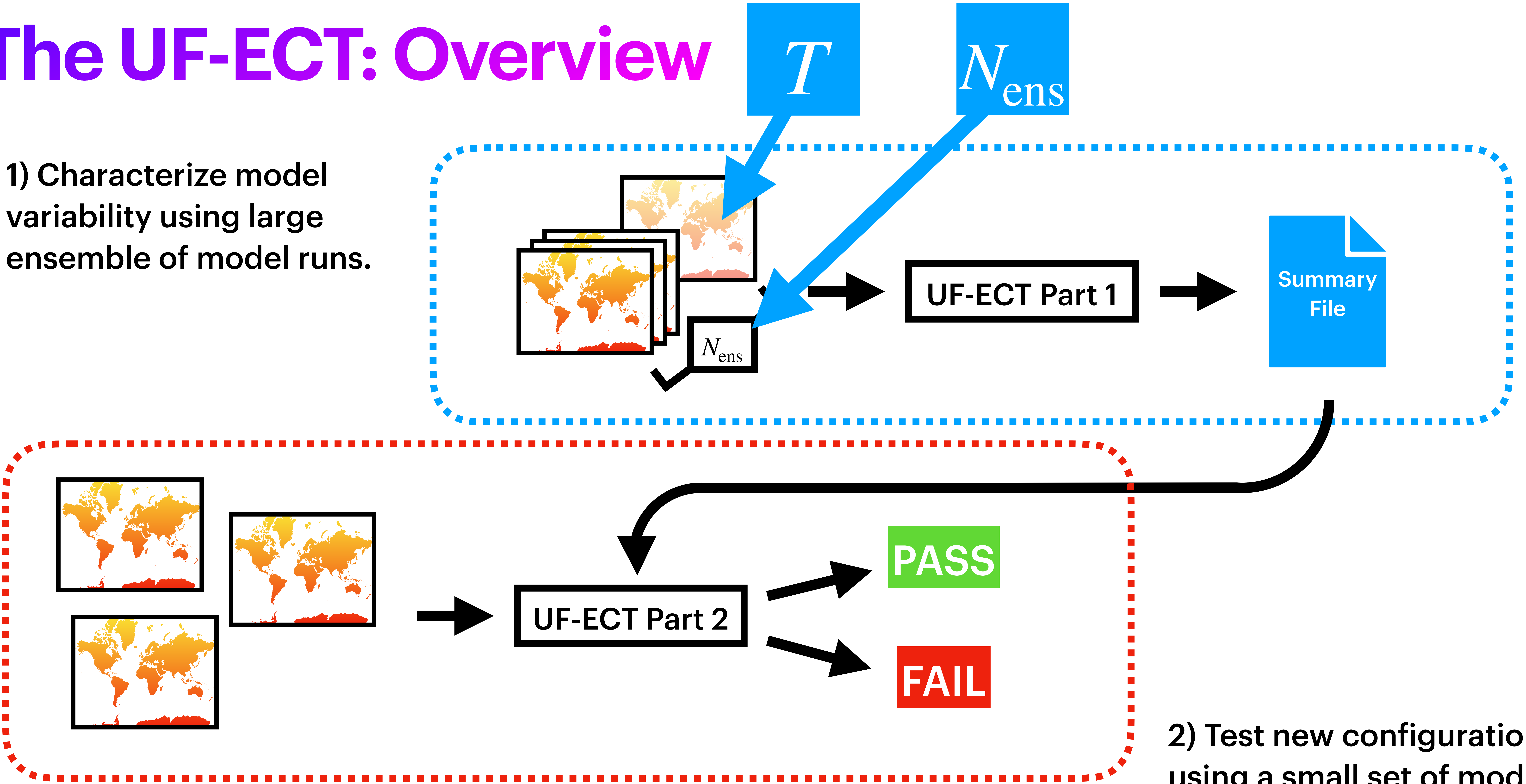
1) Characterize model variability using large ensemble of model runs.



2) Test new configurations using a small set of model runs.

The UF-ECT: Overview

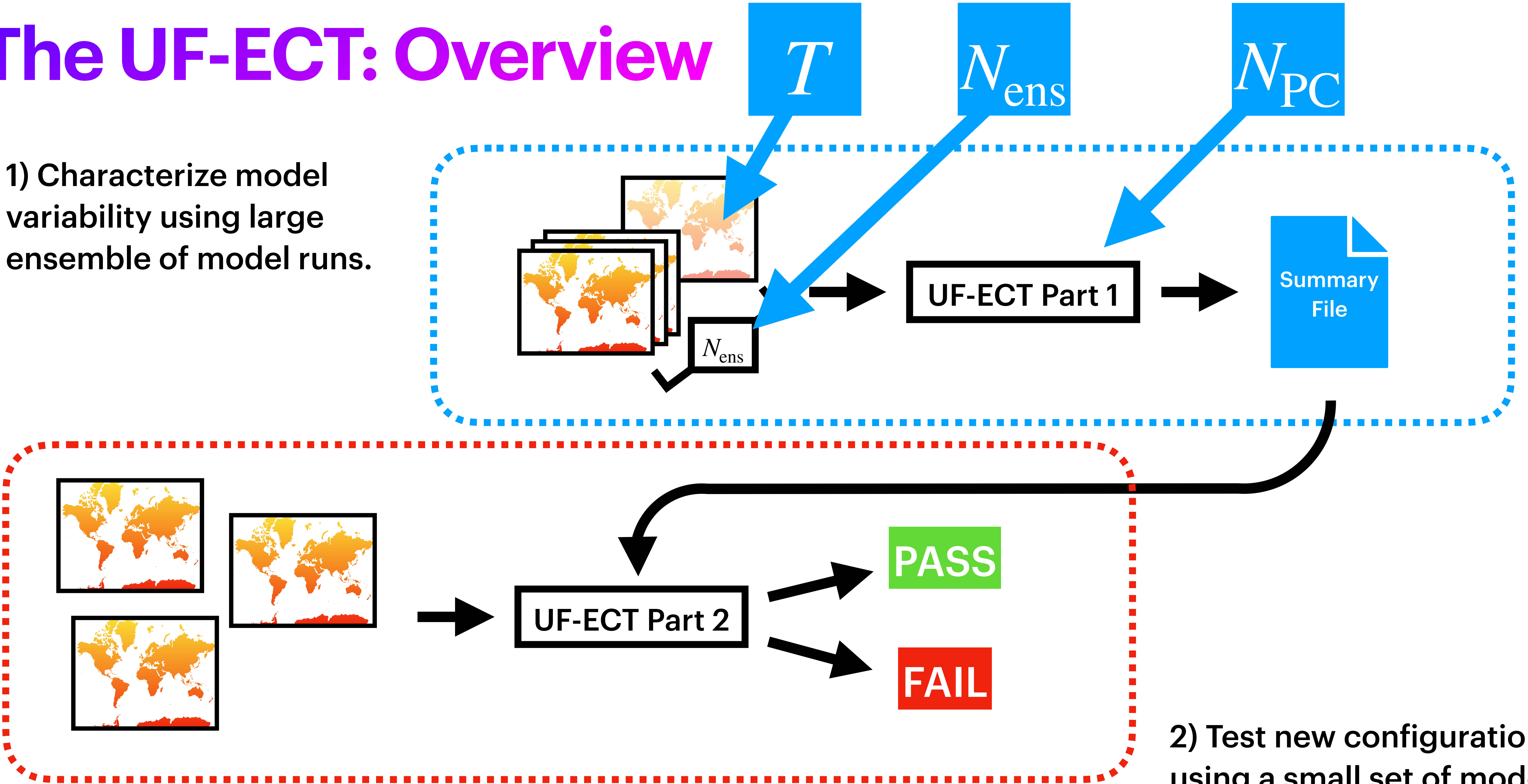
1) Characterize model variability using large ensemble of model runs.



2) Test new configurations using a small set of model runs.

The UF-ECT: Overview

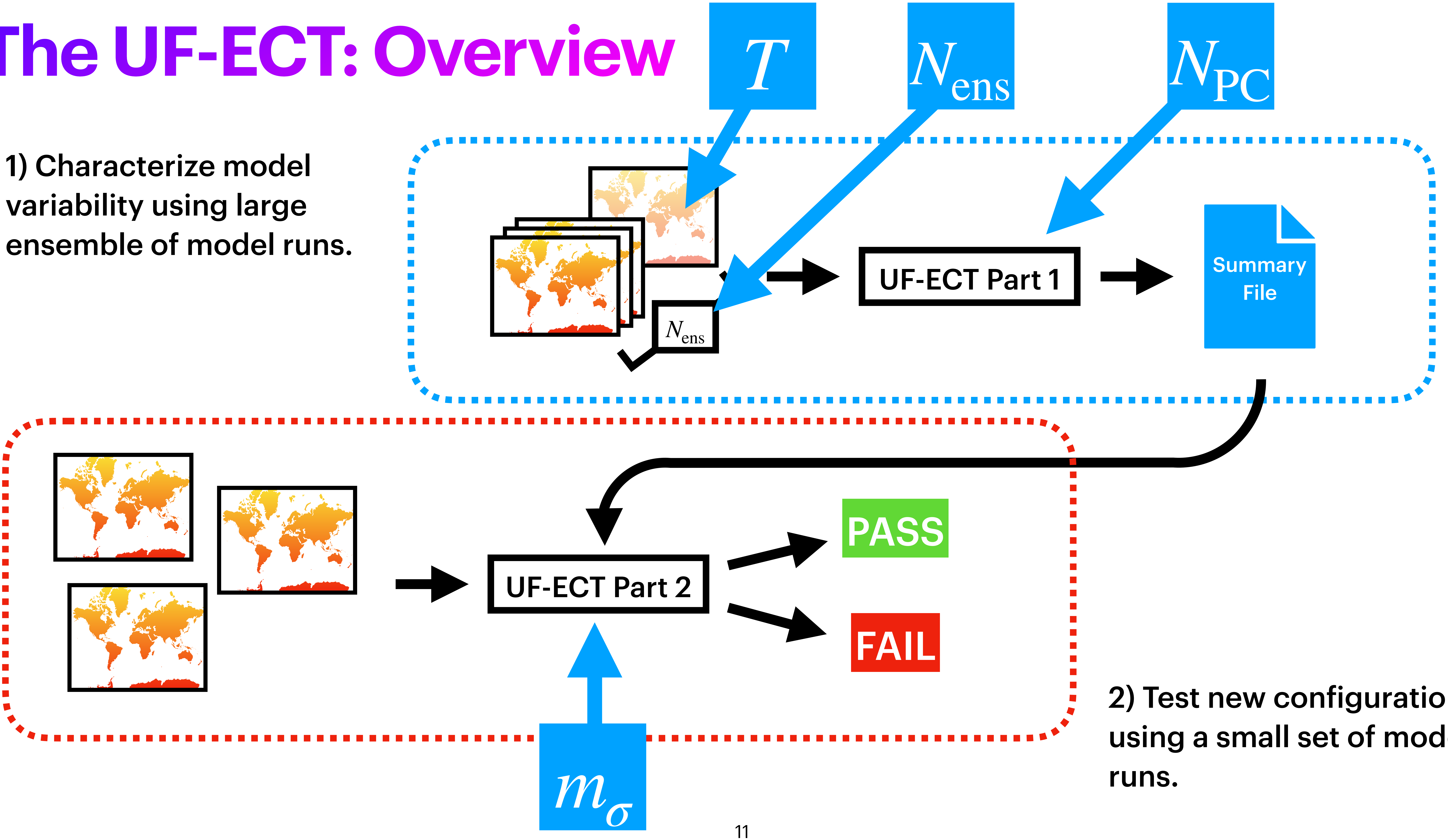
1) Characterize model variability using large ensemble of model runs.



2) Test new configurations using a small set of model runs.

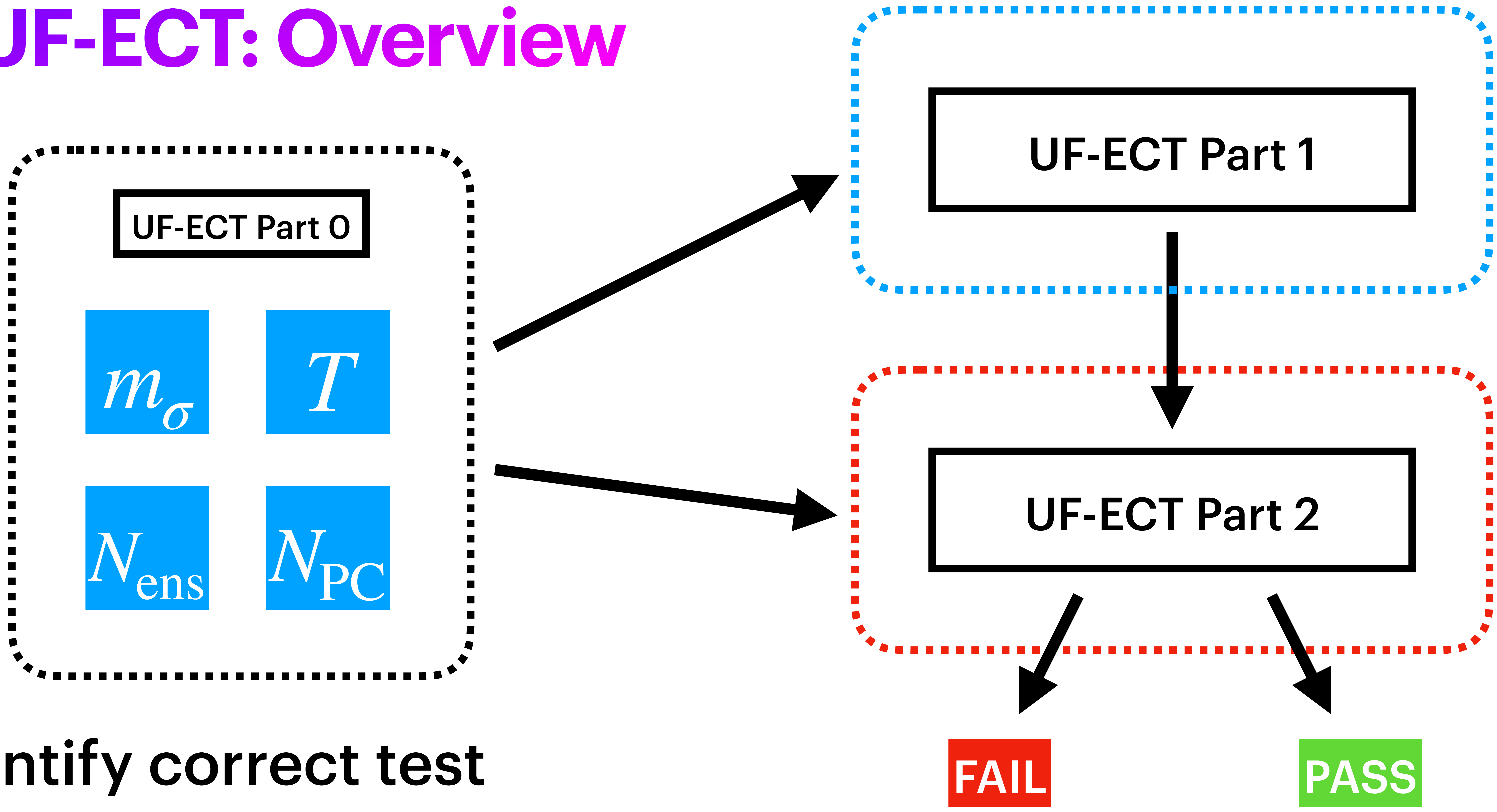
The UF-ECT: Overview

1) Characterize model variability using large ensemble of model runs.



2) Test new configurations using a small set of model runs.

The UF-ECT: Overview



0) Identify correct test parameters for given model.

How to Apply the UF-ECT to New Models?

T	N_{PC}
N_{ens}	m_σ

How to Apply the UF-ECT to New Models?

1. Develop a recipe for applying the UF-ECT to new models.

T

N_{PC}

N_{ens}

m_{σ}

How to Apply the UF-ECT to New Models?

1. Develop a recipe for applying the UF-ECT to new models.
   - Can we choose test parameters in a cohesive way to:

T | N_{PC}

N_{ens} | m_{σ}

How to Apply the UF-ECT to New Models?

1. Develop a recipe for applying the UF-ECT to new models.
   - Can we choose test parameters in a cohesive way to:
     - Detect changes across our model.

T N_{PC}

N_{ens} m_{σ}

How to Apply the UF-ECT to New Models?

1. Develop a recipe for applying the UF-ECT to new models.
    - Can we choose test parameters in a cohesive way to:
        - Detect changes across our model.
        - Ensure usability for model developers.

T N_{PC}

N_{ens} m_σ

How to Apply the UF-ECT to New Models?

1. Develop a recipe for applying the UF-ECT to new models.
   - Can we choose test parameters in a cohesive way to:
     - Detect changes across our model.
     - Ensure usability for model developers.
2. Demonstrate our approach on a different earth system model (MPAS-A)

T

N_{PC}

N_{ens}

m_{σ}

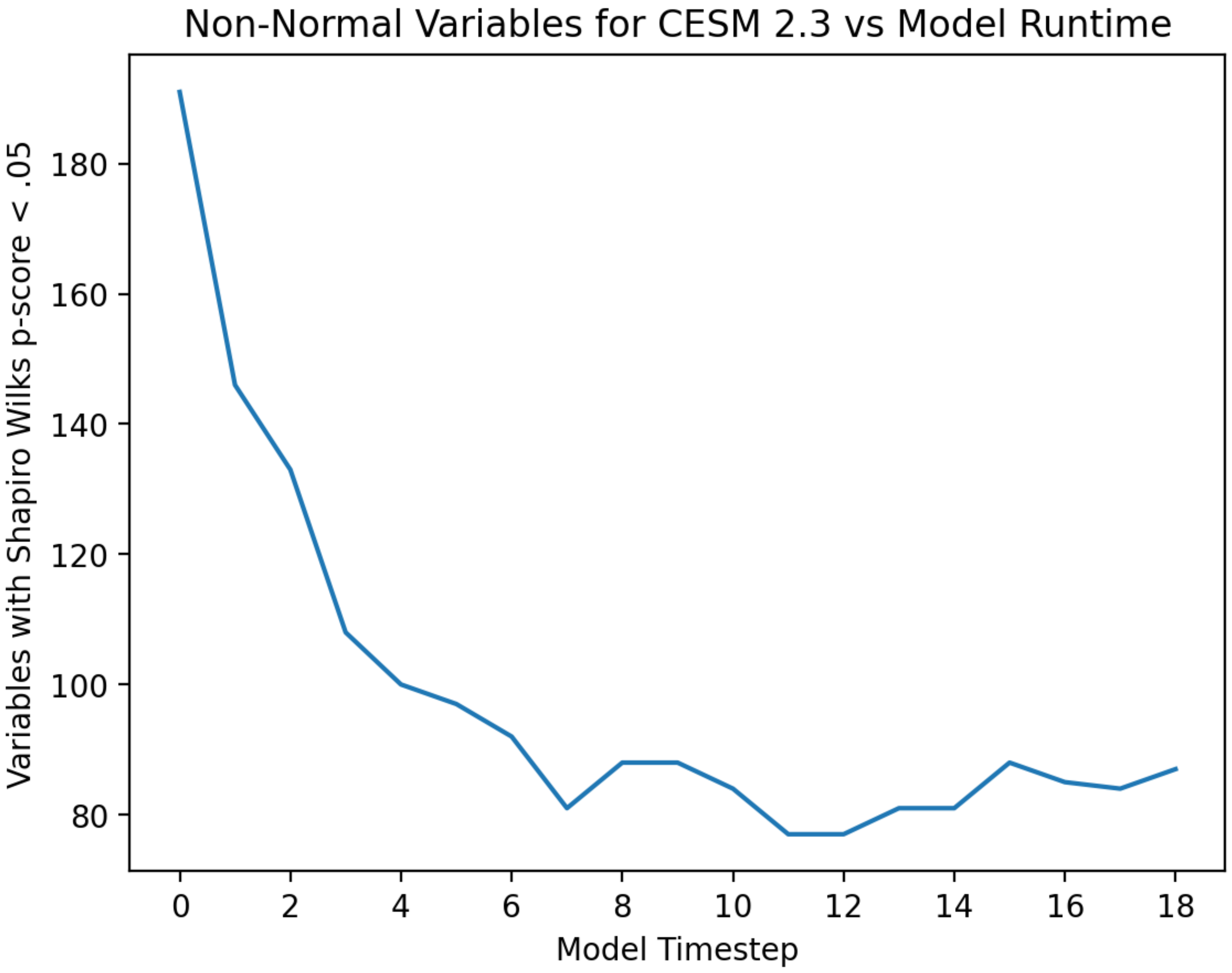
How to Apply the UF-ECT to New Models?

1. Develop a recipe for applying the UF-ECT to new models.
   - Can we choose test parameters in a cohesive way to:
     - Detect changes across our model.
     - Ensure usability for model developers.
2. Demonstrate our approach on a different earth system model (MPAS-A)
3. Identify if previous results are still appropriate for updated CESM - CAM.

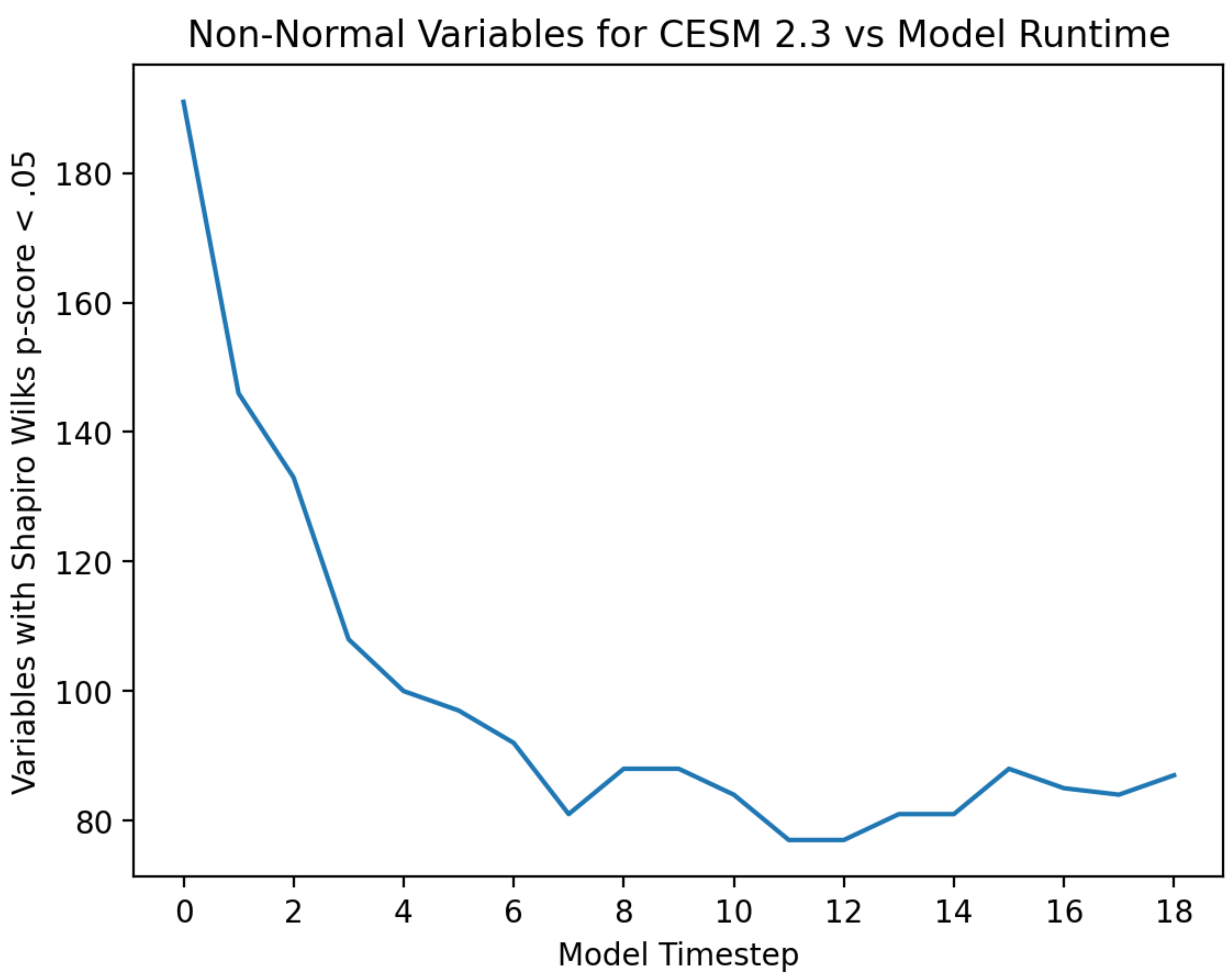
T N_{PC}

N_{ens} m_{σ}

T: How Long To Run Model?

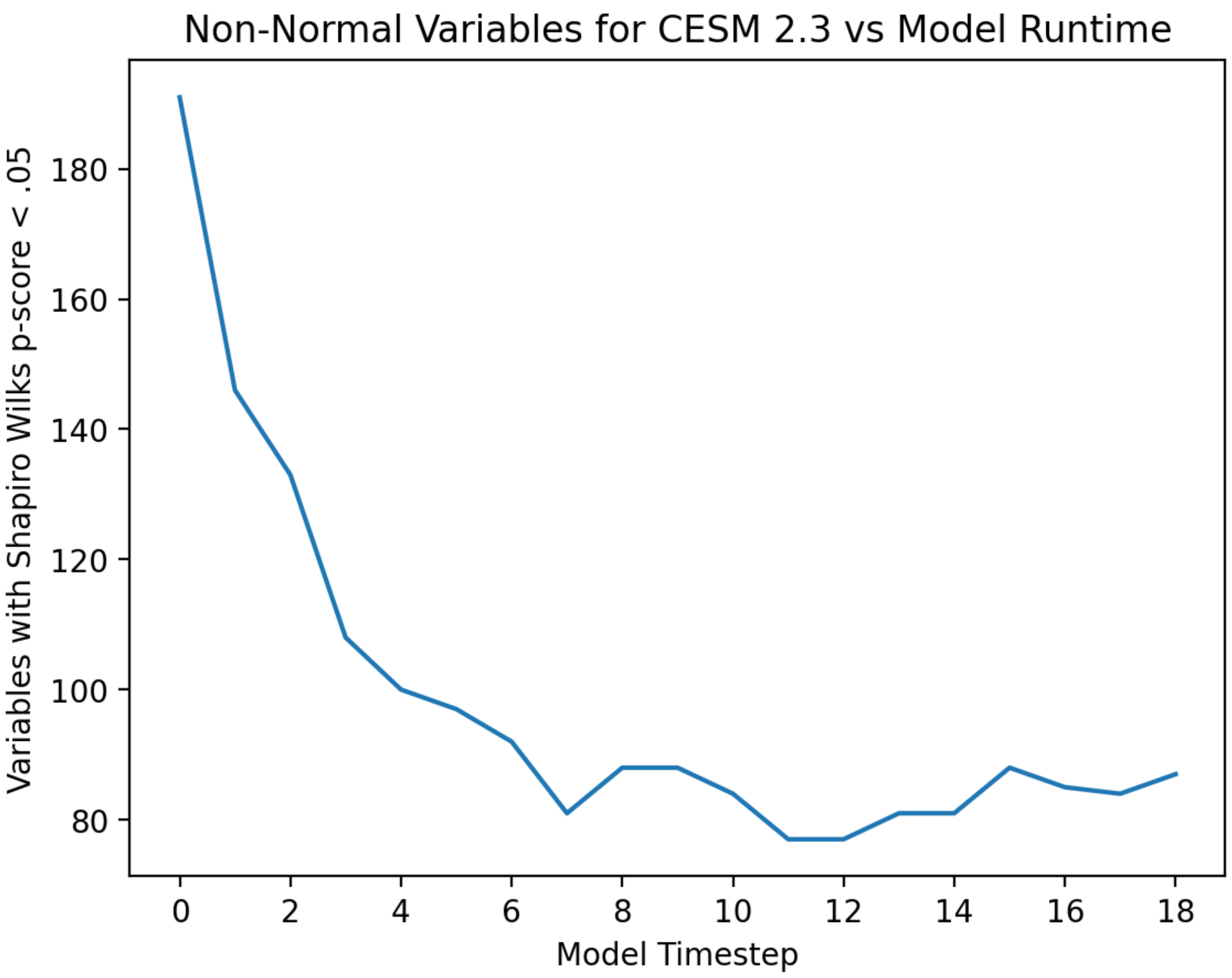


T: How Long To Run Model?



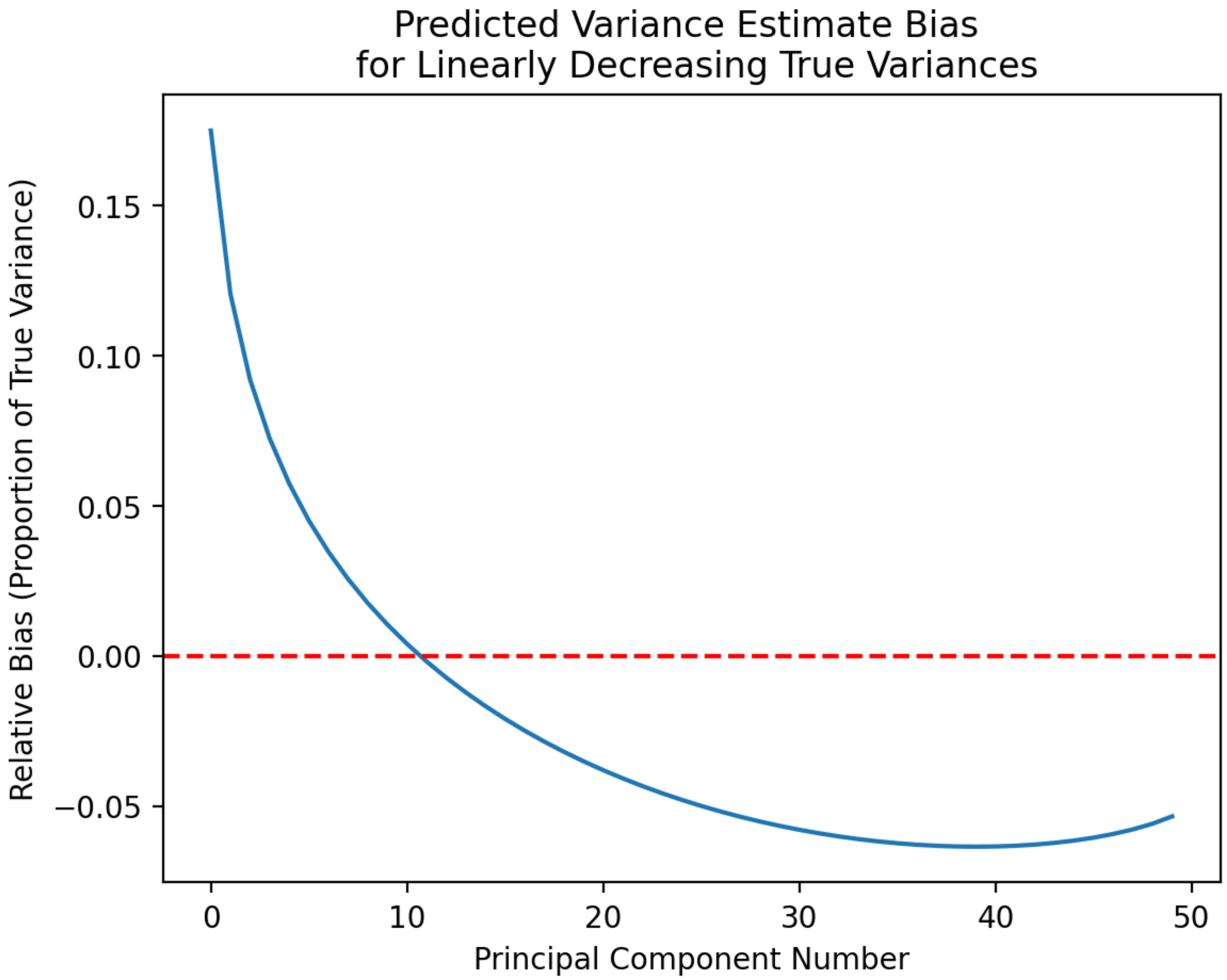
- Using a common test for normality (Shapiro-Wilks) and waiting for the number of non-normal variables to stabilize was an easy way to determine when the model had been run long enough.

T: How Long To Run Model?

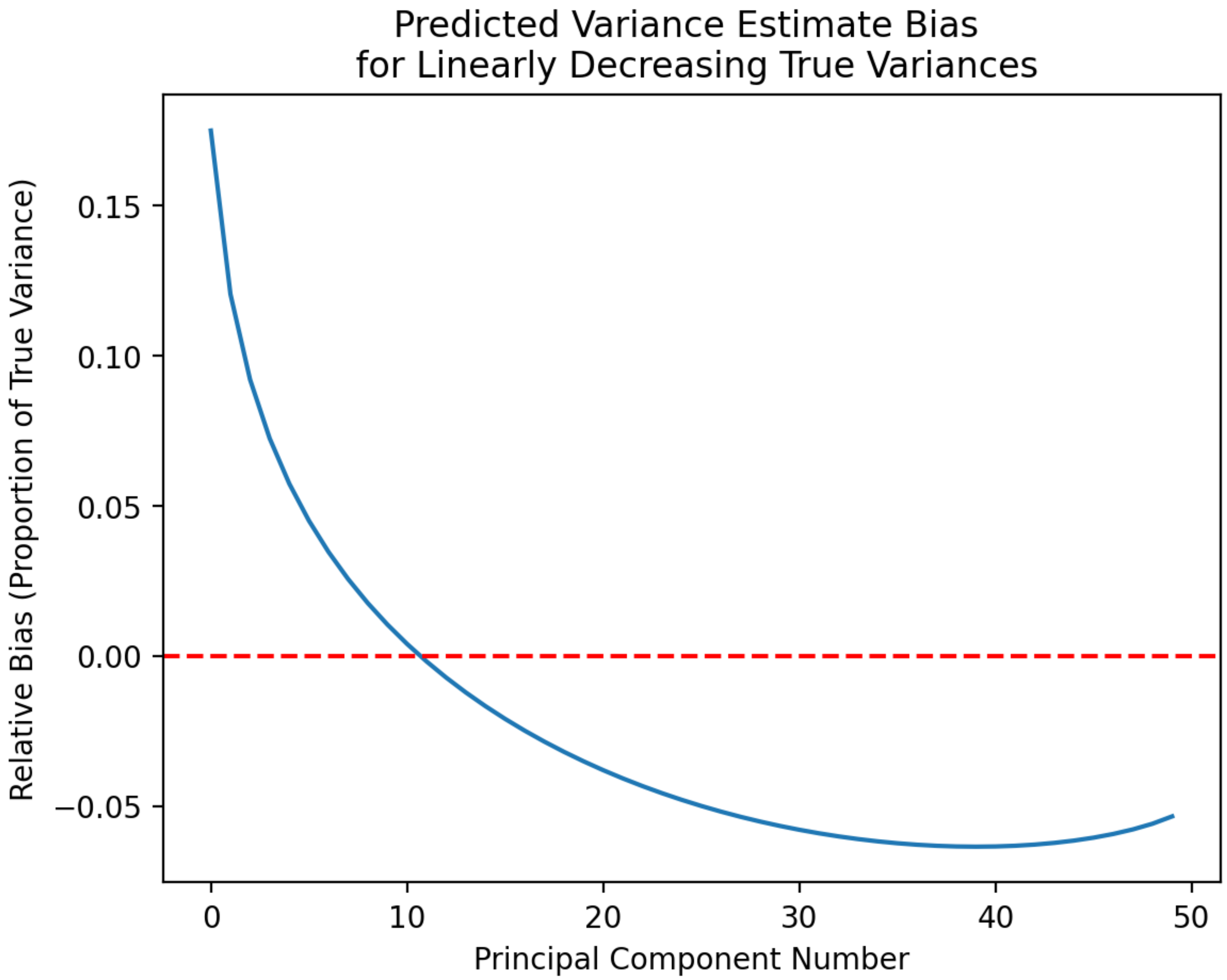


- Using a common test for normality (Shapiro-Wilks) and waiting for the number of non-normal variables to stabilize was an easy way to determine when the model had been run long enough.
- This indicates that our initial perturbations in one field have propagated through the model.

Small Detour:
PCA Variance Estimation Bias

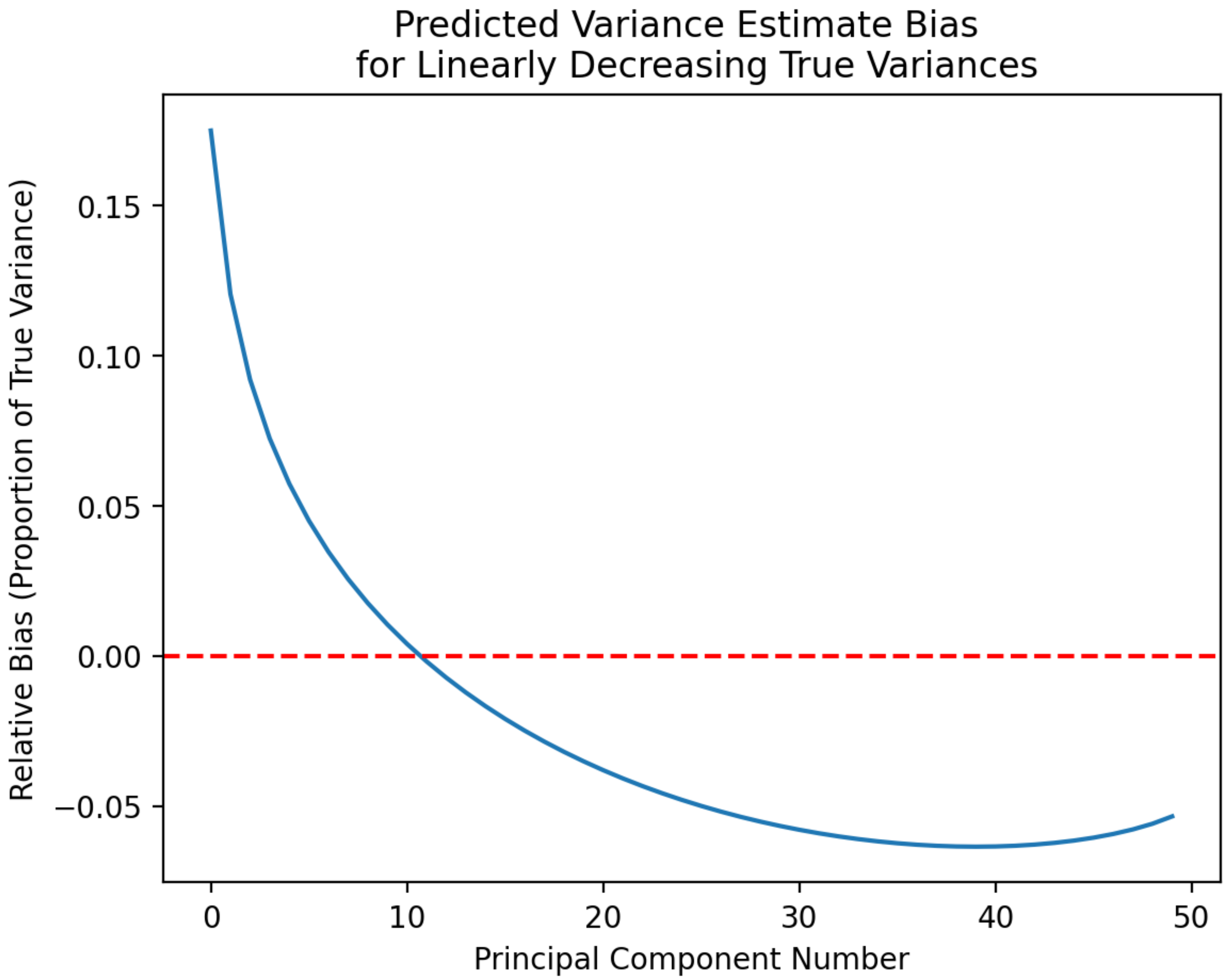


Small Detour:
PCA Variance Estimation Bias



- Estimates of PC variance are biased according to:

Small Detour: PCA Variance Estimation Bias

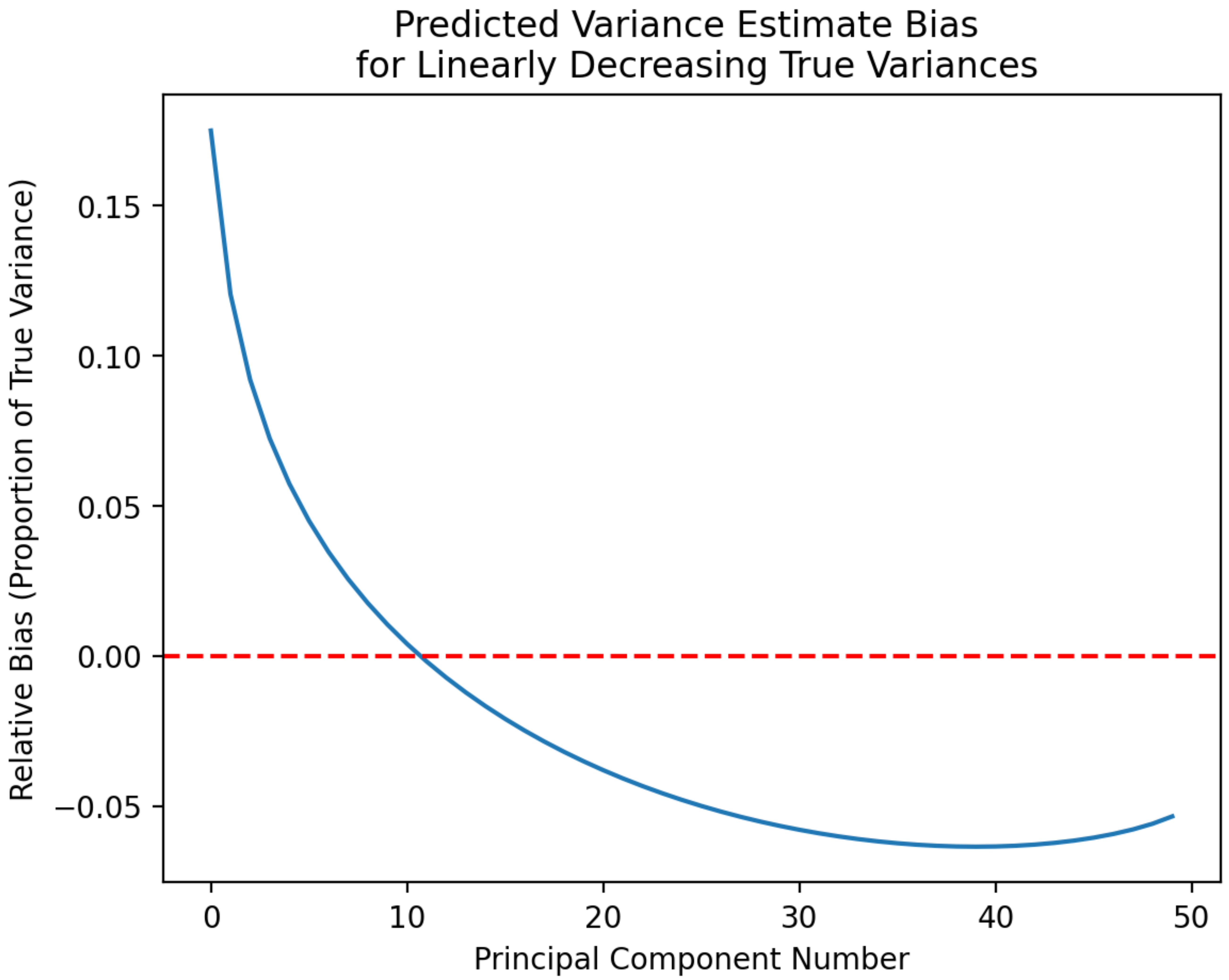


- Estimates of PC variance are biased according to:

$$E(l_i) = \lambda_i \left[1 + \frac{1}{n} \sum_{j \neq i}^{p} \left(\frac{\lambda_j}{\lambda_i - \lambda_j} \right) \right] + O(1/n^2)$$

[Lawley, 1956]

Small Detour: PCA Variance Estimation Bias



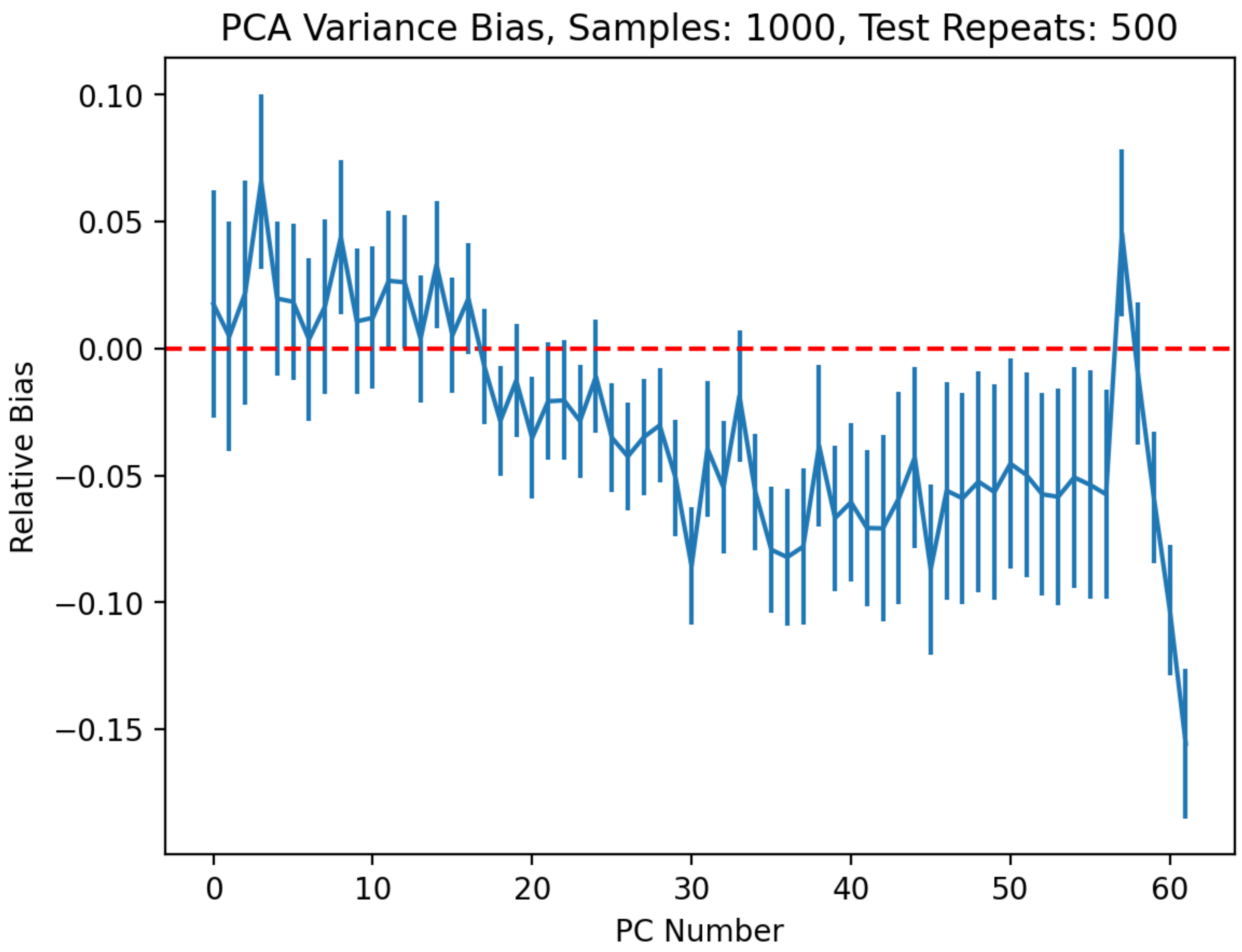
- Estimates of PC variance are biased according to:

$$E(l_i) = \lambda_i \left[1 + \frac{1}{n} \sum_{j \neq i}^{p} \left(\frac{\lambda_j}{\lambda_i - \lambda_j} \right) \right] + O(1/n^2)$$

[Lawley, 1956]

- Effect is large variances are overestimated, small variances are underestimated.

Small Detour: PCA Variance Estimation Bias



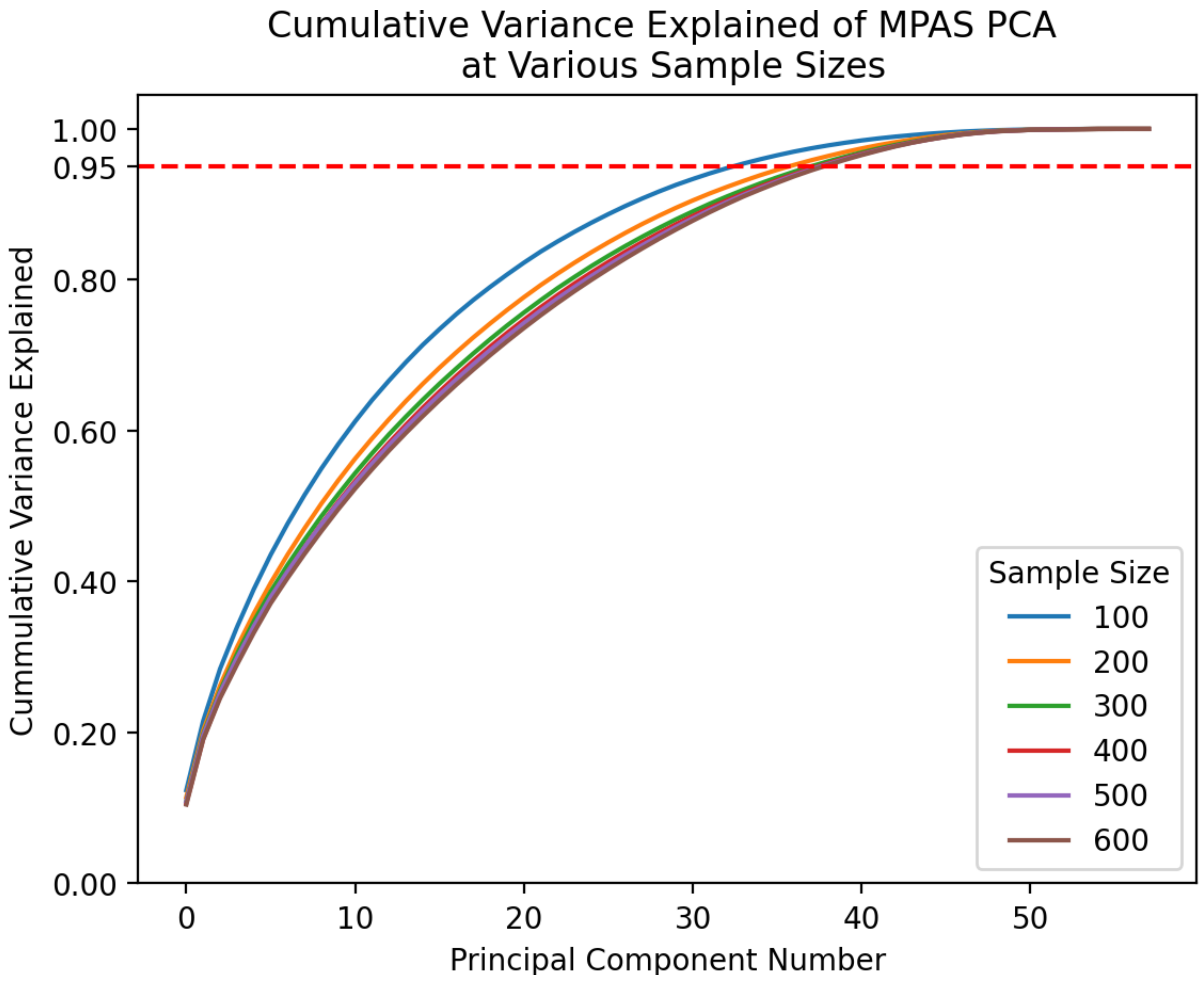
- Estimates of PC variance are biased according to:

$$E(l_i) = \lambda_i \left[1 + \frac{1}{n} \sum_{j \neq i}^{p} \left(\frac{\lambda_j}{\lambda_i - \lambda_j} \right) \right] + O(1/n^2)$$

[Lawley, 1956]

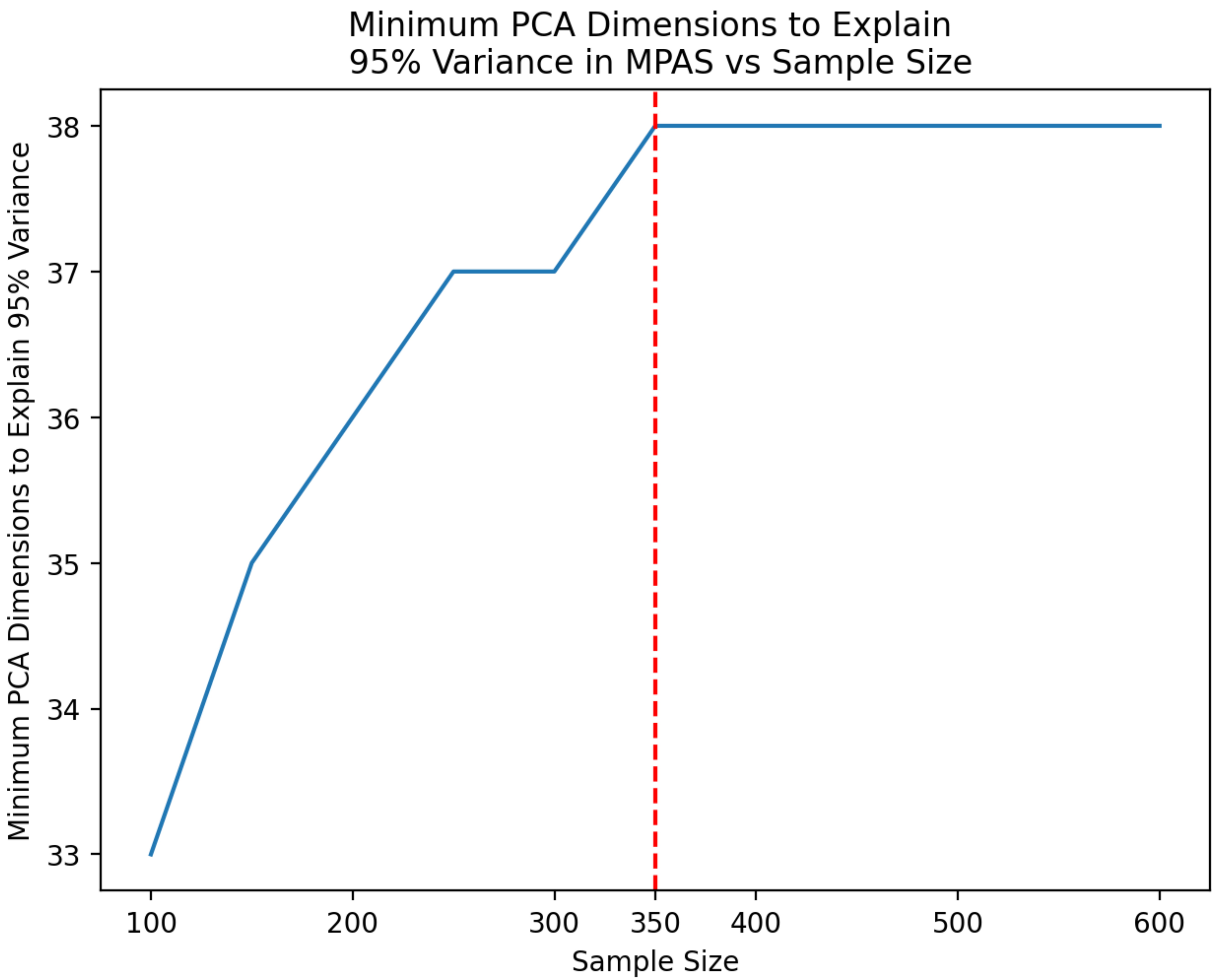
- Effect is large variances are overestimated, small variances are underestimated.

N_{PC}: How Many PCA Dimensions to Use?



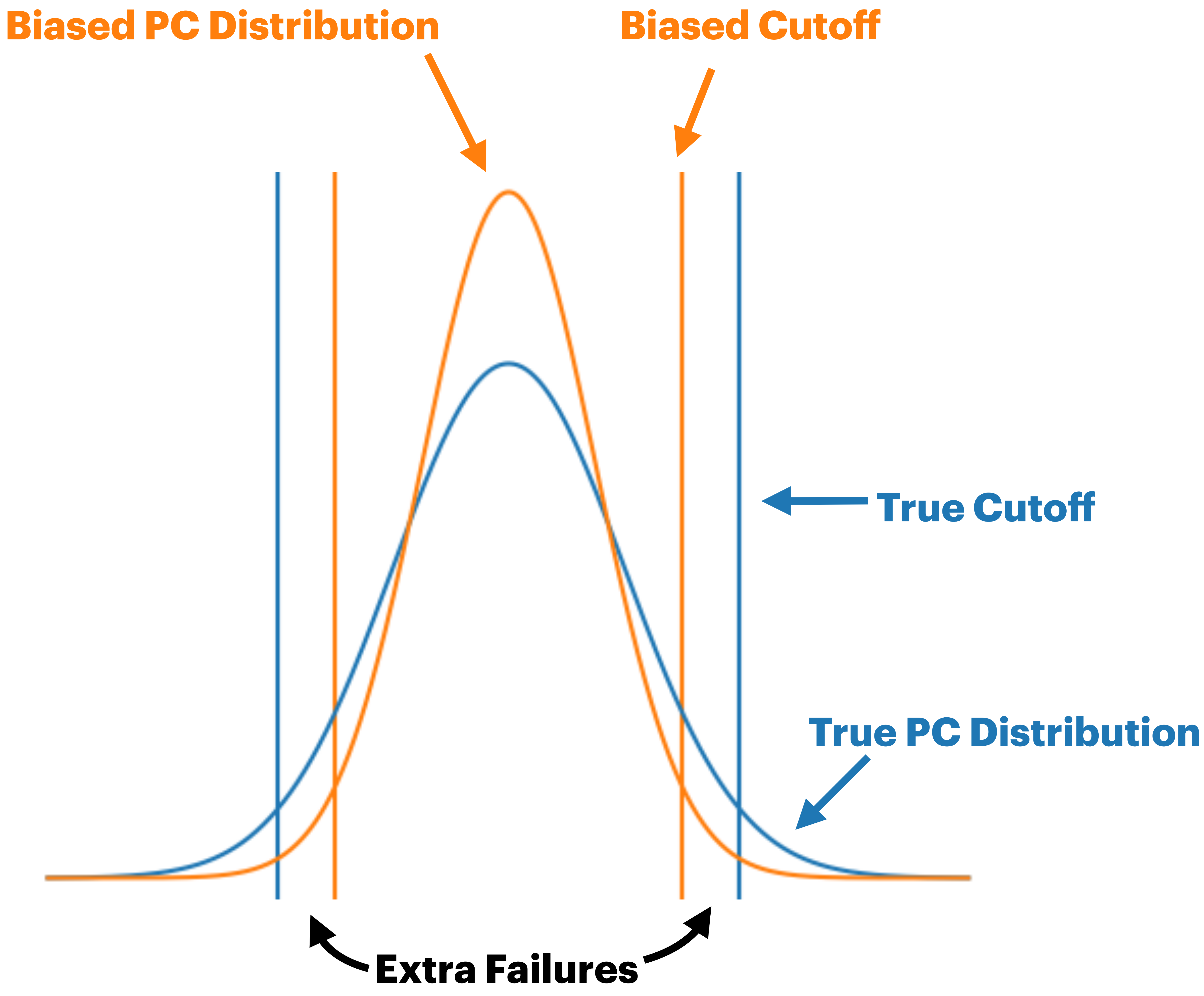
- Bias changes our estimate of how many PC’s we need to adequately capture our model.

N_{PC}: How Many PCA Dimensions to Use?

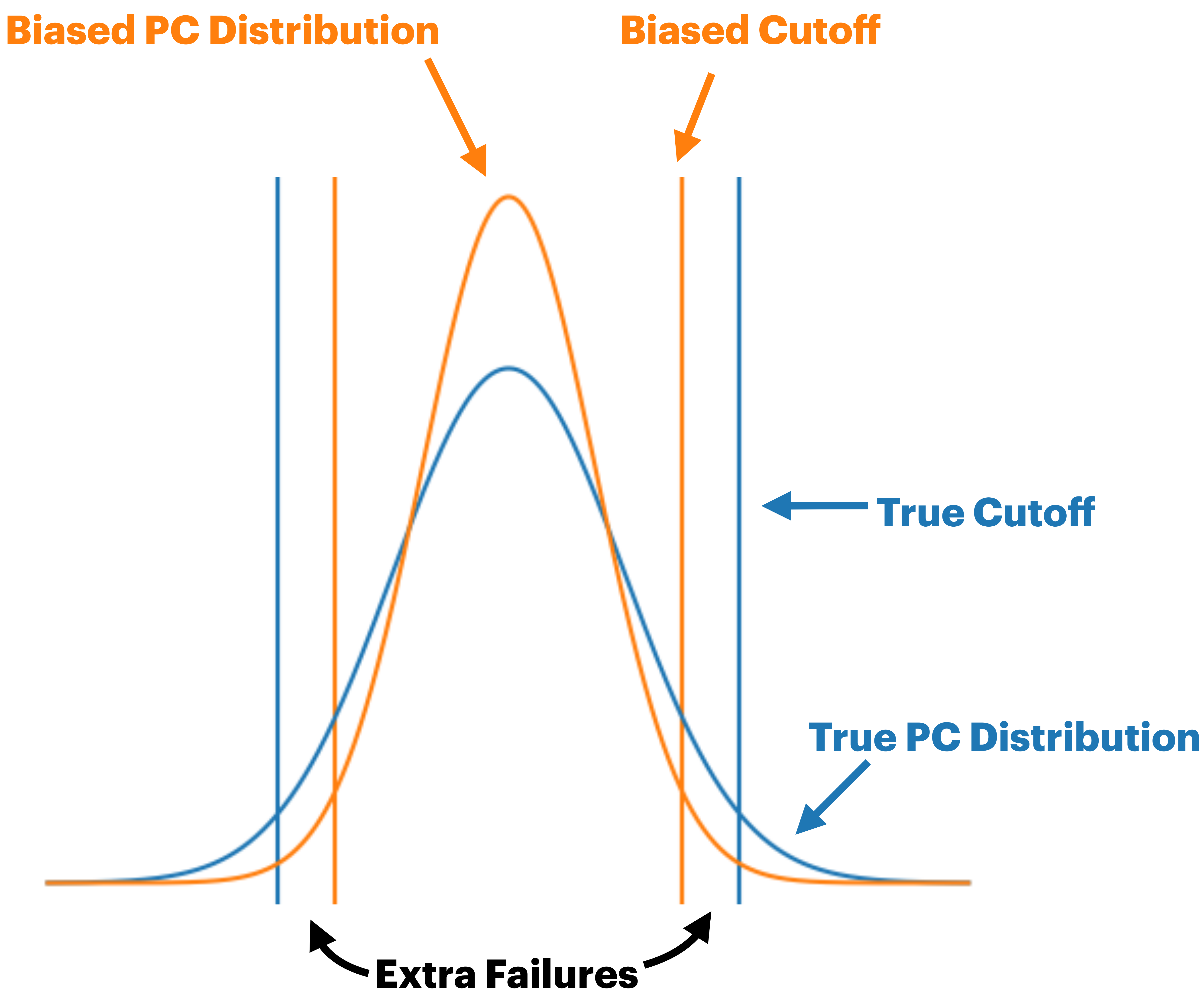


- This changes our estimate of how many PC's we need to adequately capture our model.
  - We address this by increasing our ensemble size until the number of PC dimensions required to describe 95% of our model variance stabilizes.

N_{ens}: How Large of an Ensemble?

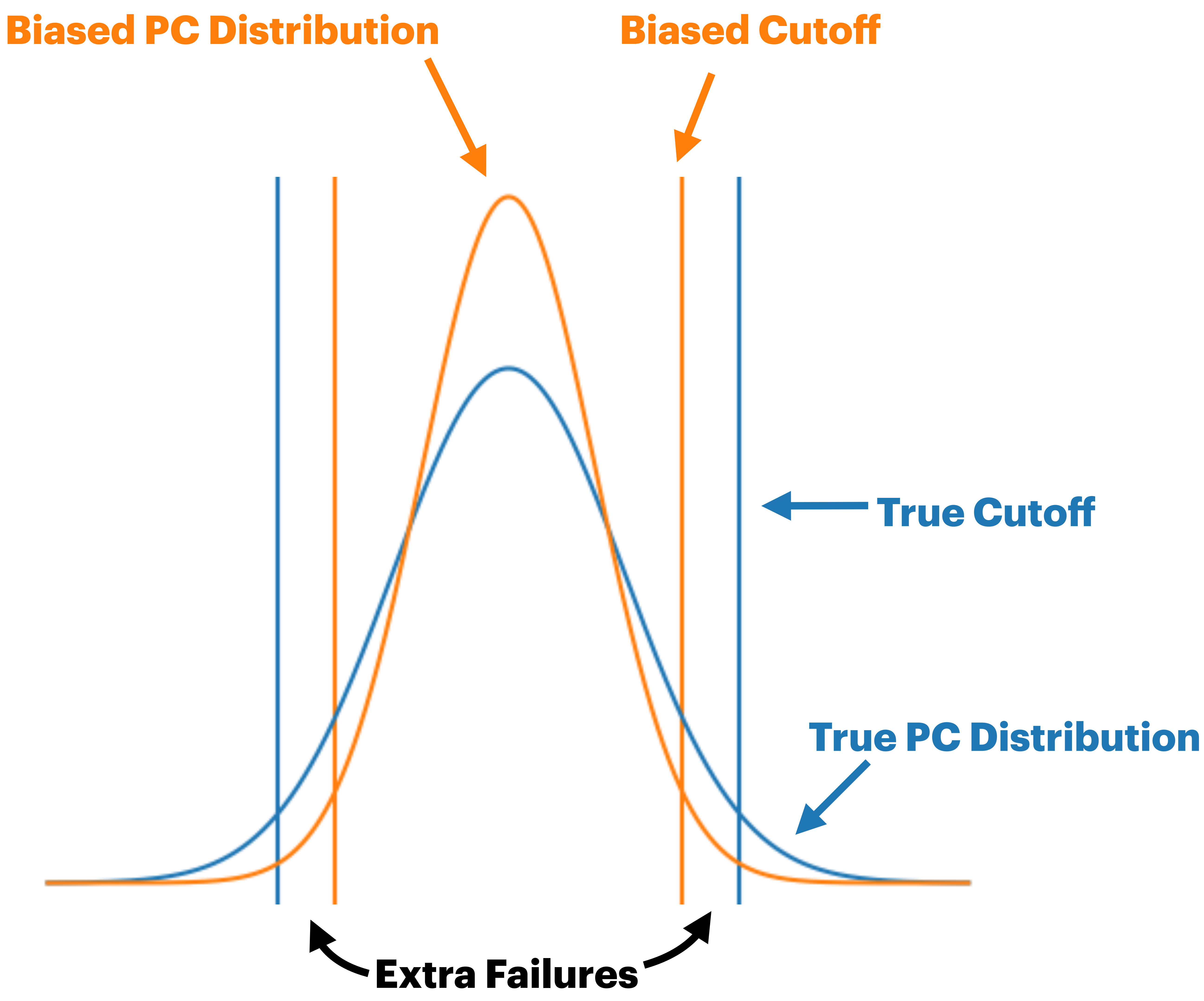


N_{ens}: How Large of an Ensemble?



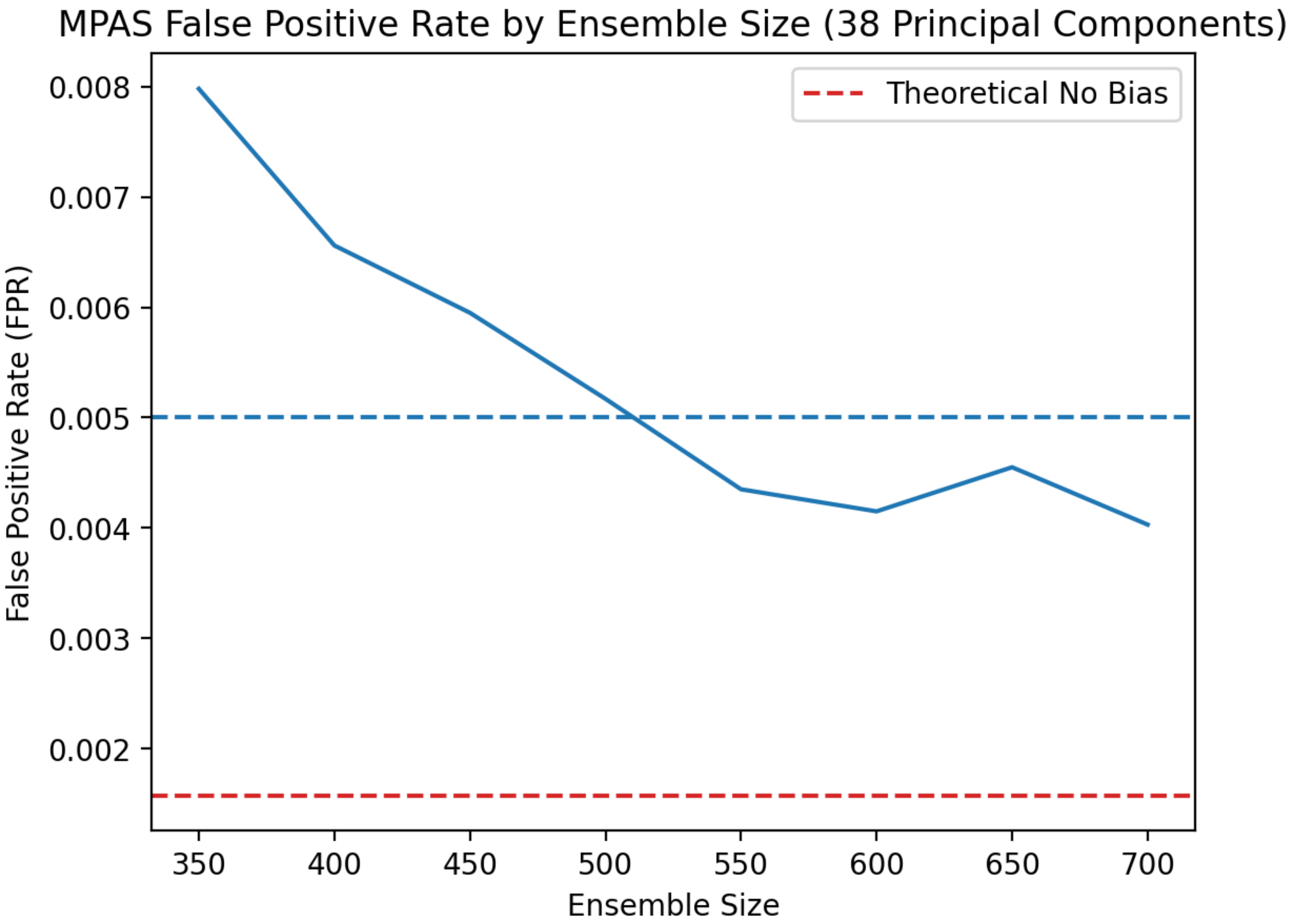
- Bias can cause erroneous failures due to underestimating the variance of specific PC's.

N_{ens}: How Large of an Ensemble?



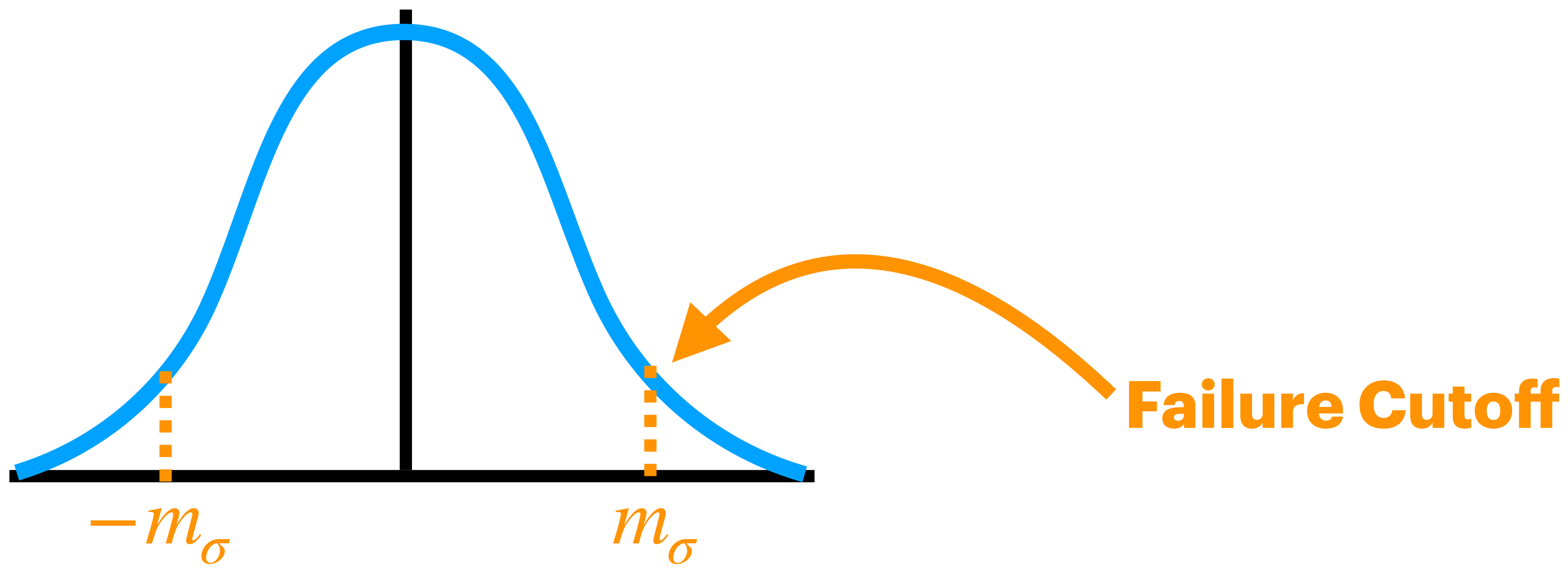
- Bias can cause erroneous failures due to underestimating the variance of specific PC's.
  - Overestimates don't average out in the UF-ECT design.

N_{ens}: How Large of an Ensemble?



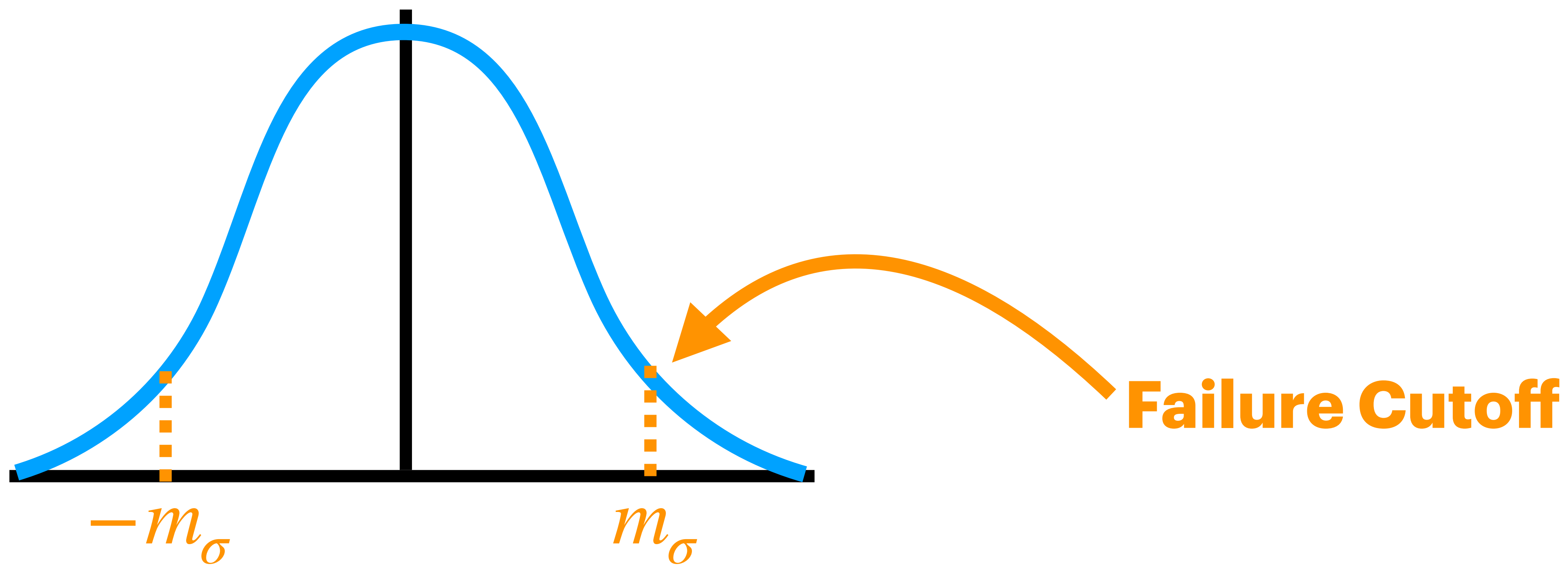
- Bias can cause erroneous failures due to underestimating the variance of specific PC's.
  - Overestimates don't average out in the UF-ECT design.
  - In order to limit bias in individual PC's we increase ensemble size until false positive rate (FPR) is acceptable.

m_σ: How To Set Failure Cutoff?



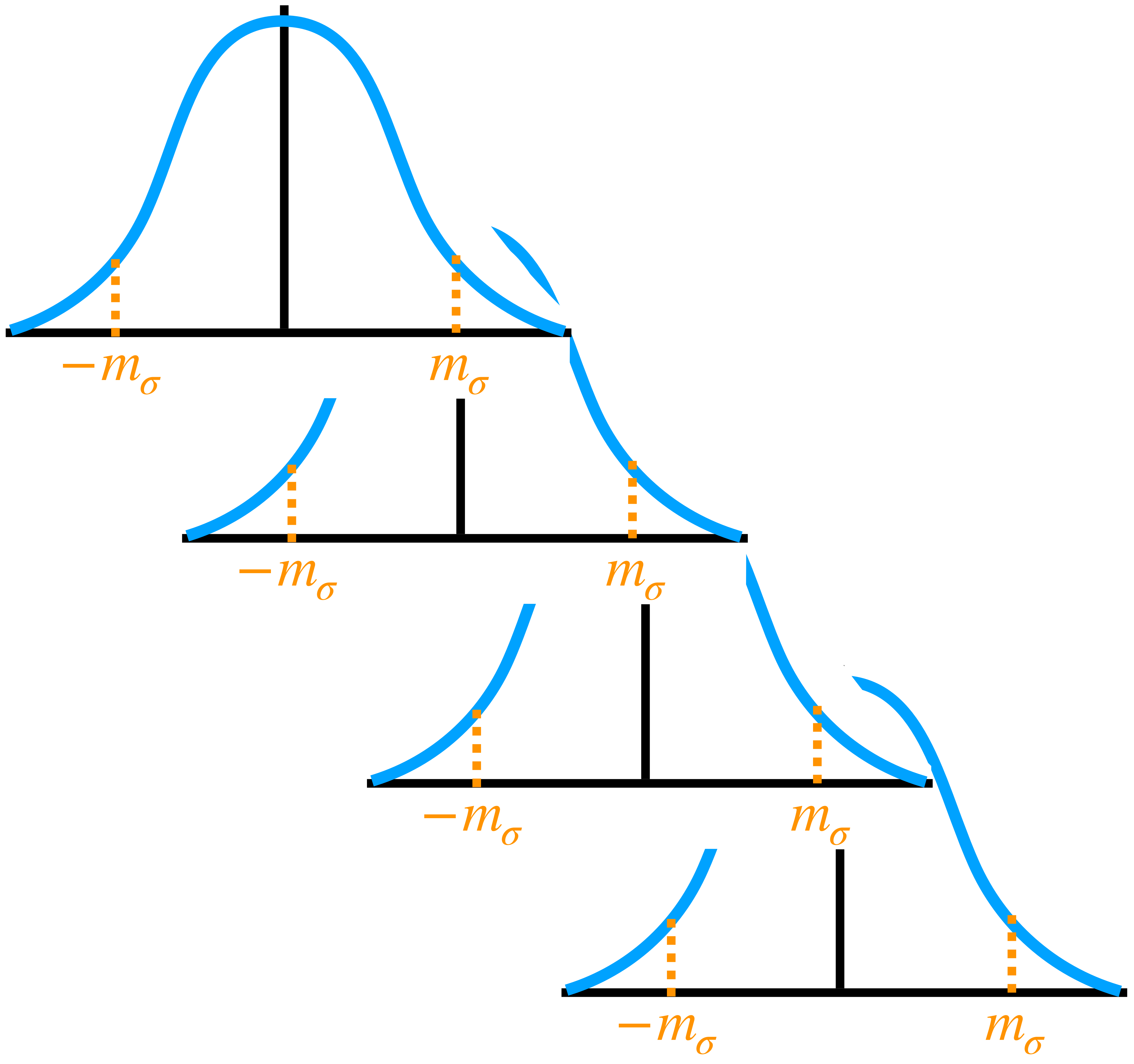
m_σ: How To Set Failure Cutoff?

- Originally set to $m_\sigma = 2$.



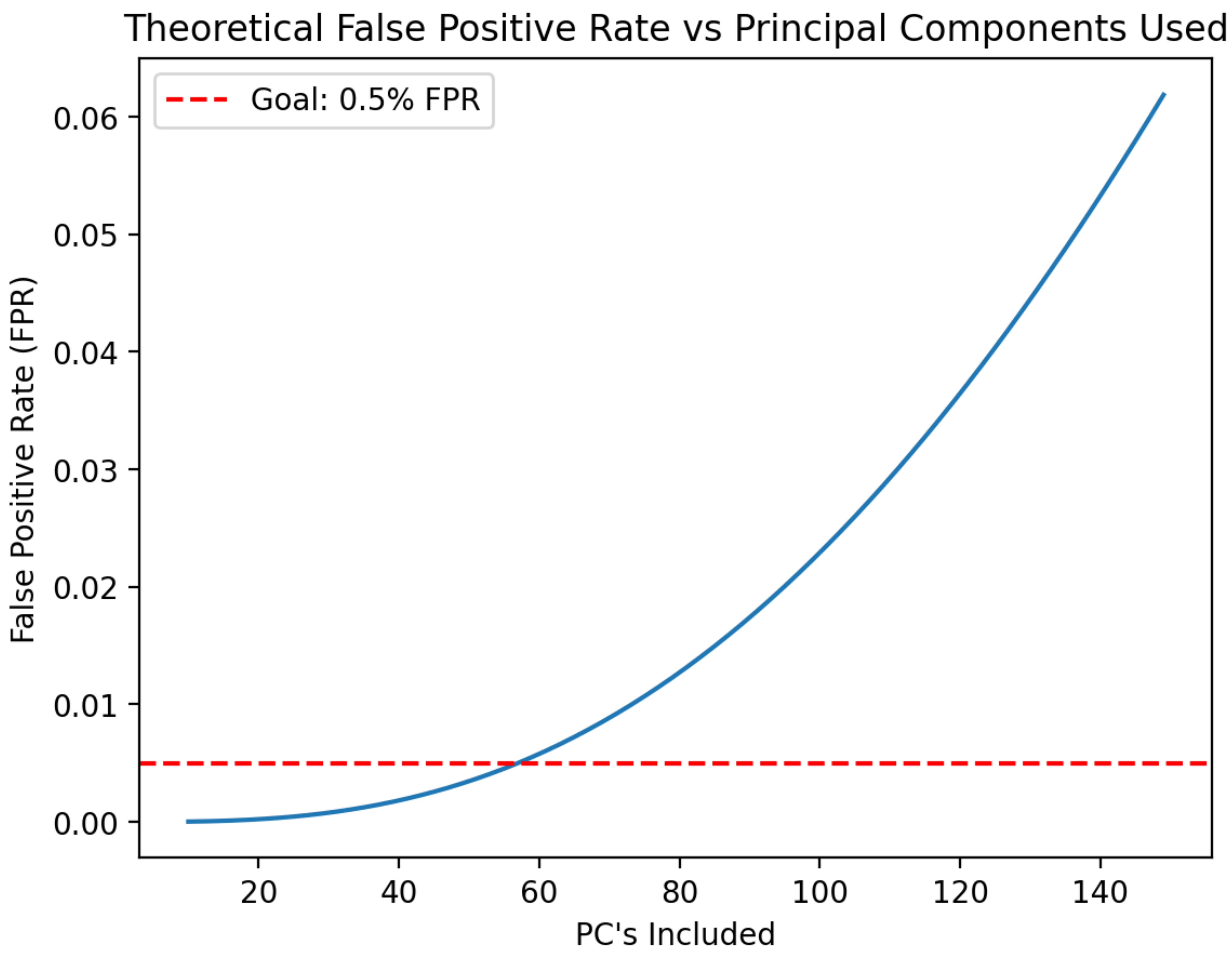
m_σ: How To Set Failure Cutoff?

- Originally set to $m_\sigma = 2$.
- Bigger models -> more PC's -> more chances to fail -> higher FPR.



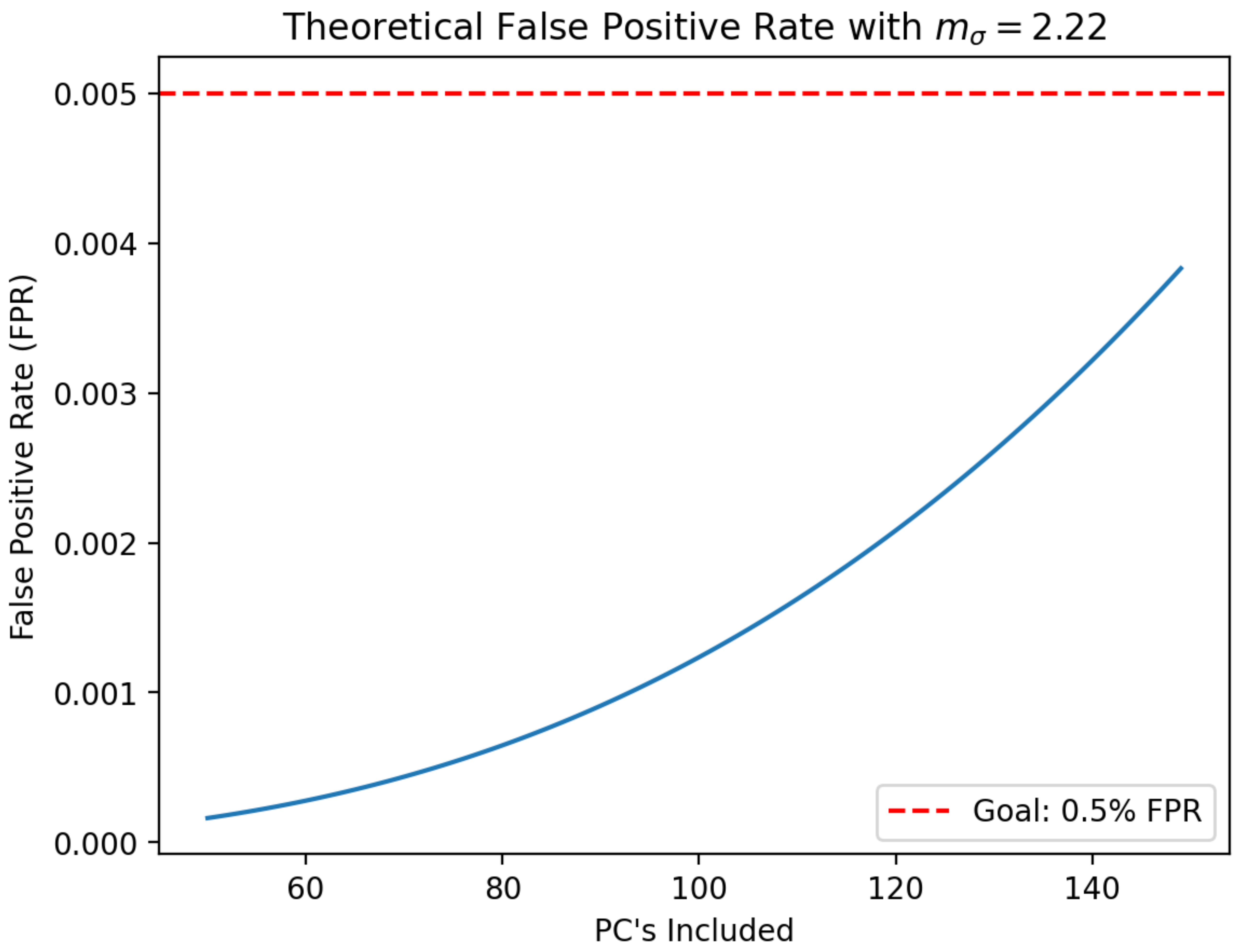
m_σ: How To Set Failure Cutoff?

- Originally set to $m_\sigma = 2$.
- Bigger models -> more PC's -> more chances to fail -> higher FPR.



m_σ: How To Set Failure Cutoff?

- Originally set to $m_\sigma = 2$.
- Bigger models -> more PC's -> more chances to fail -> higher FPR.
- We introduce a step to numerically solve for a reasonable m_σ at our chosen N_{PC}



Overall Approach

Overall Approach

1. Determine an appropriate length to run a model.

Overall Approach

1. Determine an appropriate length to run a model.
    - Want to make sure perturbations have propagated through the model.

Overall Approach

1. Determine an appropriate length to run a model.
    - Want to make sure perturbations have propagated through the model.
2. Determine an appropriate number of PC's to use.

Overall Approach

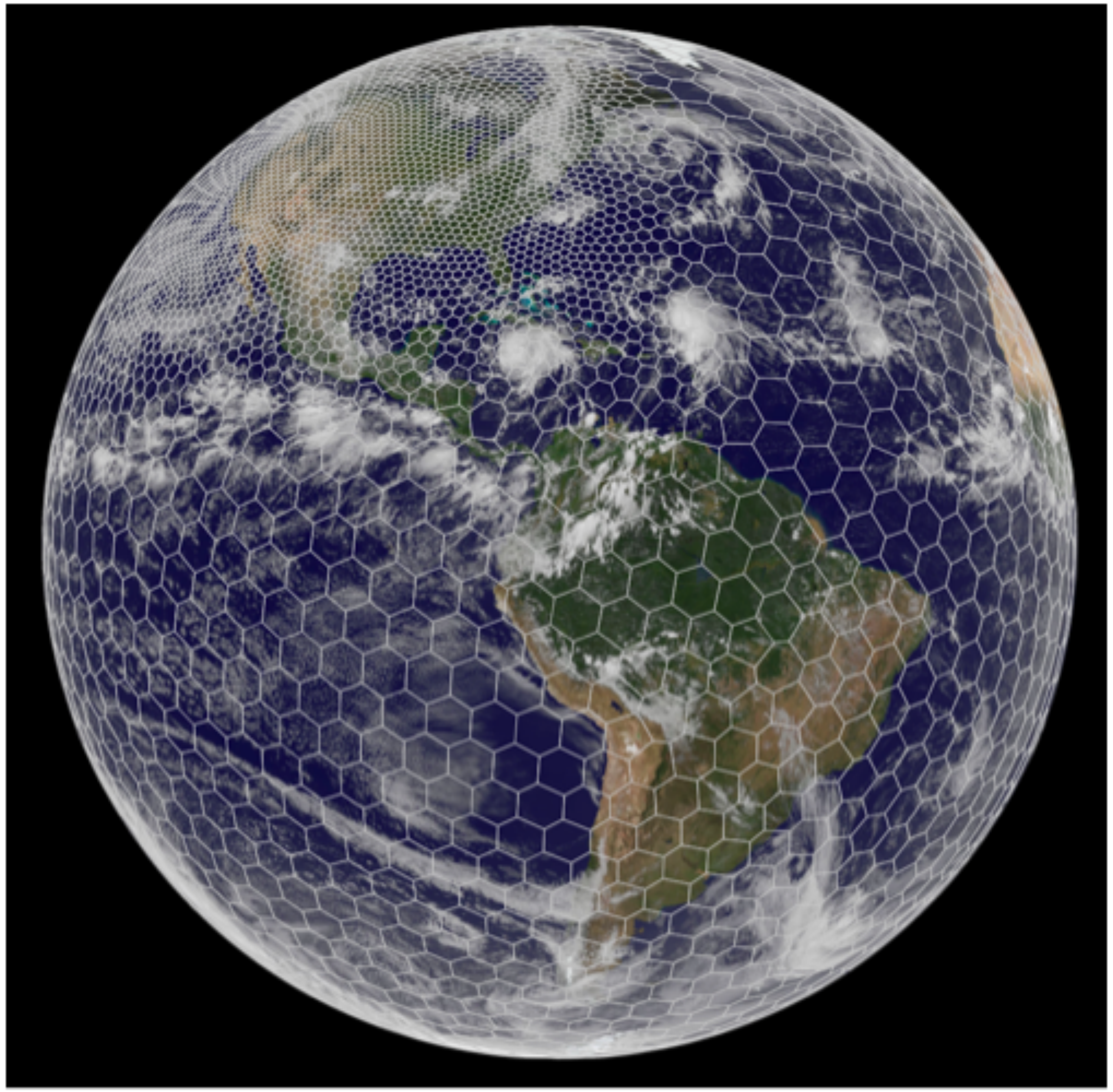
1. Determine an appropriate length to run a model.
    - Want to make sure perturbations have propagated through the model.
2. Determine an appropriate number of PC's to use.
    - Make sure we capture most of the variance of the model.

Overall Approach

1. Determine an appropriate length to run a model.
    - Want to make sure perturbations have propagated through the model.
2. Determine an appropriate number of PC's to use.
    - Make sure we capture most of the variance of the model.
    - Also sets a minimum ensemble size.

Overall Approach

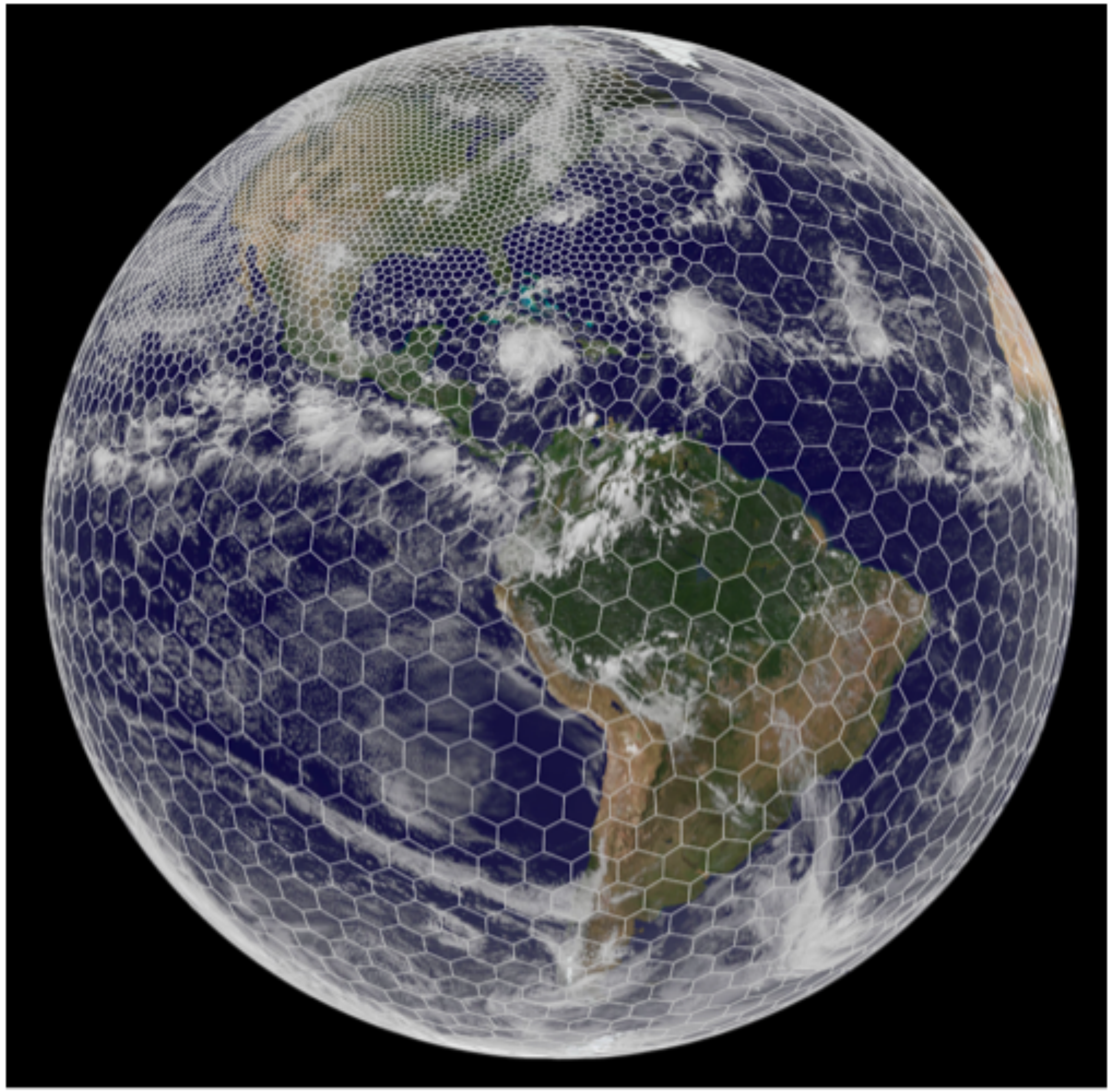
1. Determine an appropriate length to run a model.
    - Want to make sure perturbations have propagated through the model.
2. Determine an appropriate number of PC’s to use.
    - Make sure we capture most of the variance of the model.
    - Also sets a minimum ensemble size.
3. Determine appropriate failure cutoff and ensemble size to prevent too many erroneous failures.



Testing with MPAS

Model Across Prediction Scales

- In order to test our procedure for determining UF-ECT parameters we applied them to a new model, MPAS-A.



Testing with MPAS

Model Across Prediction Scales

- In order to test our procedure for determining UF-ECT parameters we applied them to a new model, MPAS-A.
- MPAS is a climate model based around unstructured Voronoi meshes.

UF-ECT Test Types

False Positives

Do additional runs from the same configuration fail?

Non-Climate Changing Modifications

Changes that might lead to BFB changes, but aren't expected to affect scientific conclusions.

Climate Changing Parameters

Does the test detect the change of scientific parameters?

Testing Non-Climate Changing Modifications

Test Title	Test Description	Test Result (EET Failure Rate)

EET = Run test with many different ensembles / test runs

Testing Non-Climate Changing Modifications

Test Title	Test Description	Test Result (EET Failure Rate)
Compiler	Change from Intel's Fortran Compiler to GNU	0.12%

Testing Non-Climate Changing Modifications

Test Title	Test Description	Test Result (EET Failure Rate)
Compiler	Change from Intel's Fortran Compiler to GNU	0.12%
Core Count	Change from 36 cores to 96 cores	0.2%

Testing Non-Climate Changing Modifications

Test Title	Test Description	Test Result (EET Failure Rate)
Compiler	Change from Intel's Fortran Compiler to GNU	0.12%
Core Count	Change from 36 cores to 96 cores	0.2%
Compiler Optimizations	Change from Intel -03 compiler optimizations to -01	0.12%

Testing Non-Climate Changing Modifications

Test Title	Test Description	Test Result (EET Failure Rate)
Compiler	Change from Intel's Fortran Compiler to GNU	0.12%
Core Count	Change from 36 cores to 96 cores	0.2%
Compiler Optimizations	Change from Intel -03 compiler optimizations to -01	0.12%
Order of Operations	Change part of MPAS convection code to do a set of operations in a different, but mathematically equivalent, order.	1.67%

Testing Non-Climate Changing Modifications

Test Title	Test Description	Test Result (EET Failure Rate)
Compiler	Change from Intel's Fortran Compiler to GNU	0.12%
Core Count	Change from 36 cores to 96 cores	0.2%
Compiler Optimizations	Change from Intel -03 compiler optimizations to -01	0.12%
Order of Operations	Change part of MPAS convection code to do a set of operations in a different, but mathematically equivalent, order.	1.67%
Precision	Change from double to single precision	100%

Testing Non-Climate Changing Modifications

Test Title	Test Description	Test Result (EET Failure Rate)
Compiler	Change from Intel's Fortran Compiler to GNU	0.12%
Core Count	Change from 36 cores to 96 cores	0.2%
Compiler Optimizations	Change from Intel -03 compiler optimizations to -01	0.12%
Order of Operations	Change part of MPAS convection code to do a set of operations in a different, but mathematically equivalent, order.	1.67%
Precision	Change from double to single precision	100%
New Cluster	Run on default Derecho configuration (Intel compiler)	37.91%

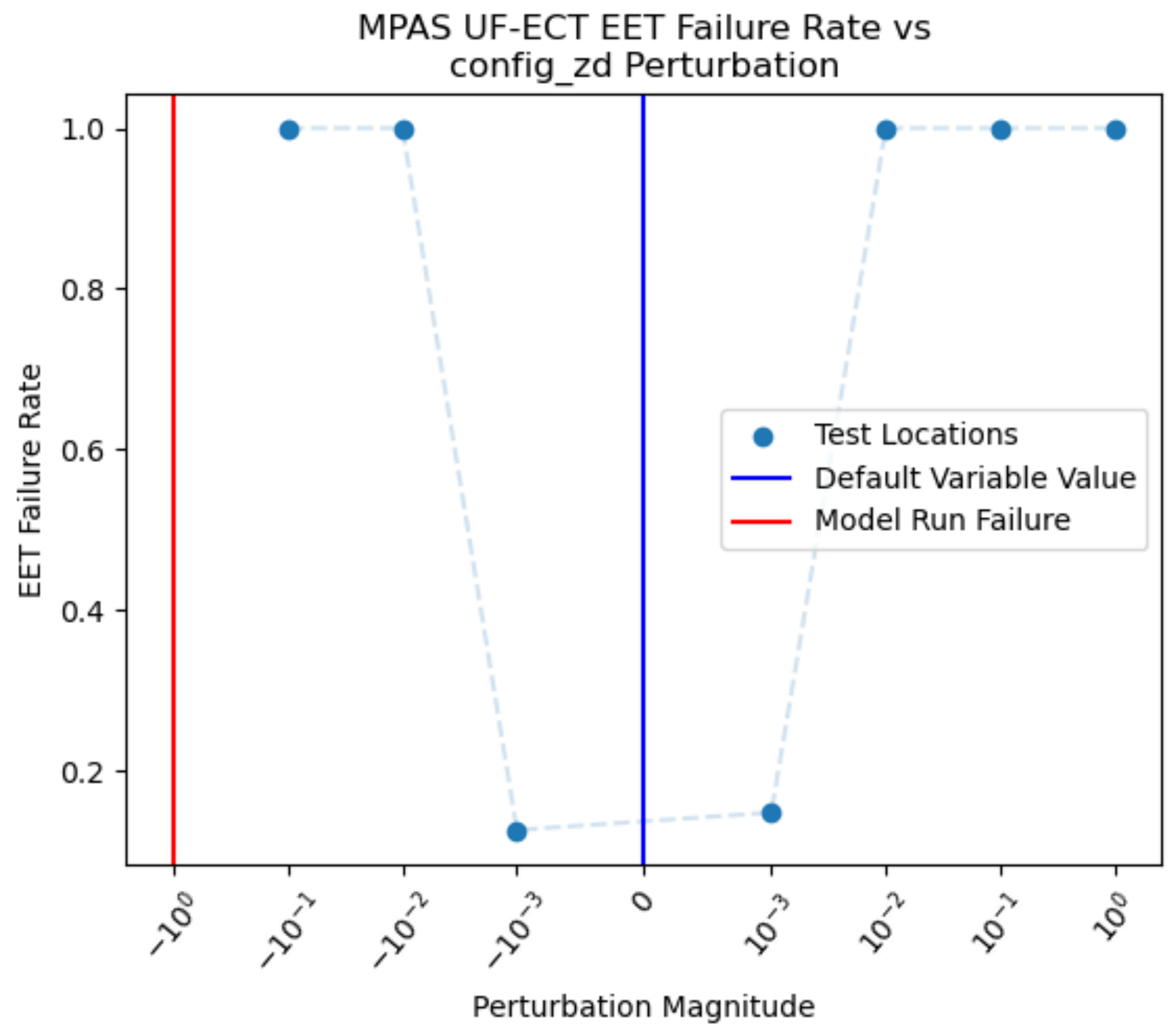
Testing Non-Climate Changing Modifications

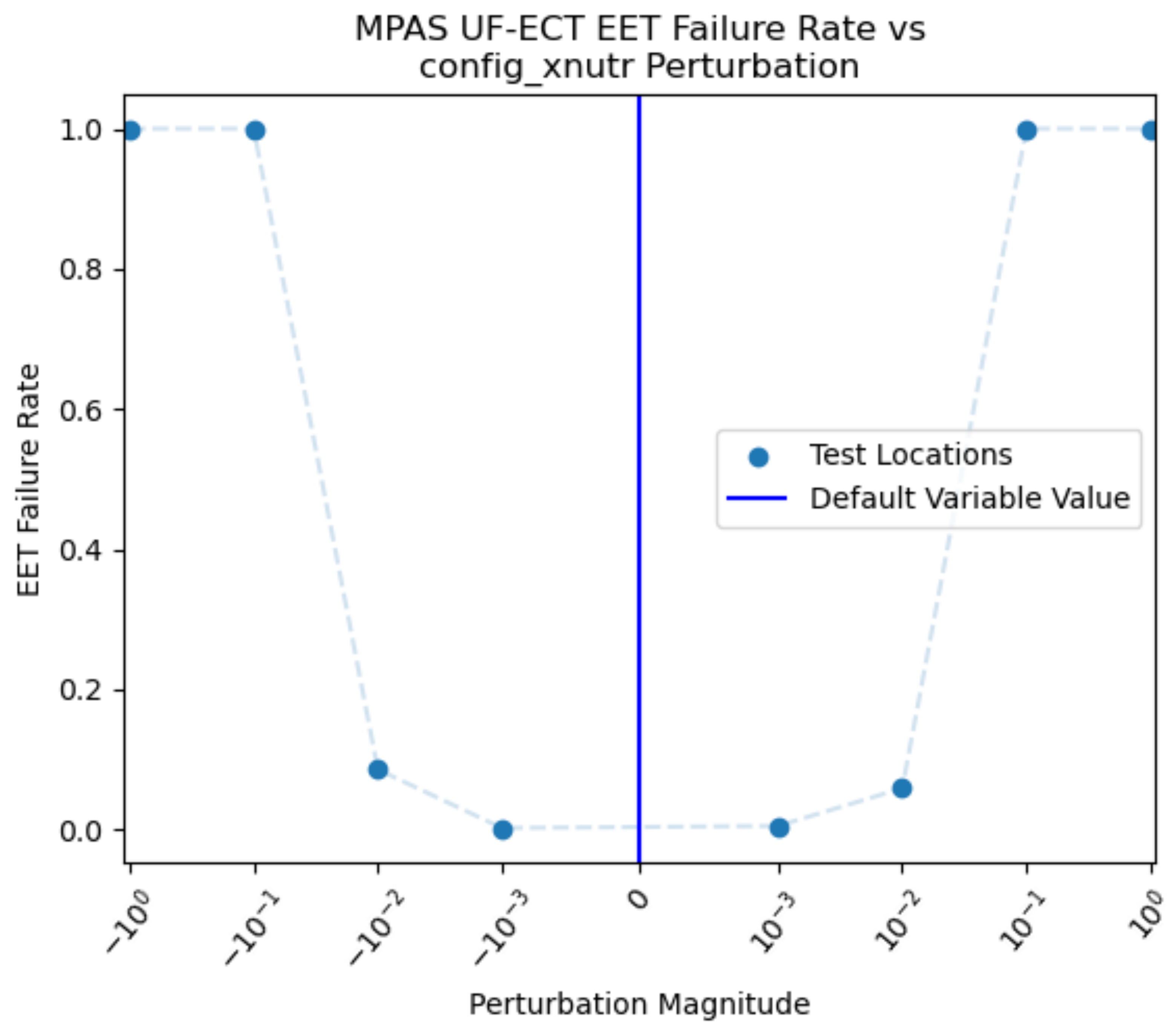
Test Title	Test Description	Test Result (EET Failure Rate)
Compiler	Change from Intel's Fortran Compiler to GNU	0.12%
Core Count	Change from 36 cores to 96 cores	0.2%
Compiler Optimizations	Change from Intel -03 compiler optimizations to -01	0.12%
Order of Operations	Change part of MPAS convection code to do a set of operations in a different, but mathematically equivalent, order.	1.67%
Precision	Change from double to single precision	100%
New Cluster	Run on default Derecho configuration (Intel compiler)	37.91%
New Cluster (No FMA)	Run on default Derecho configuration (Intel compiler) but without FMA.	0.15%

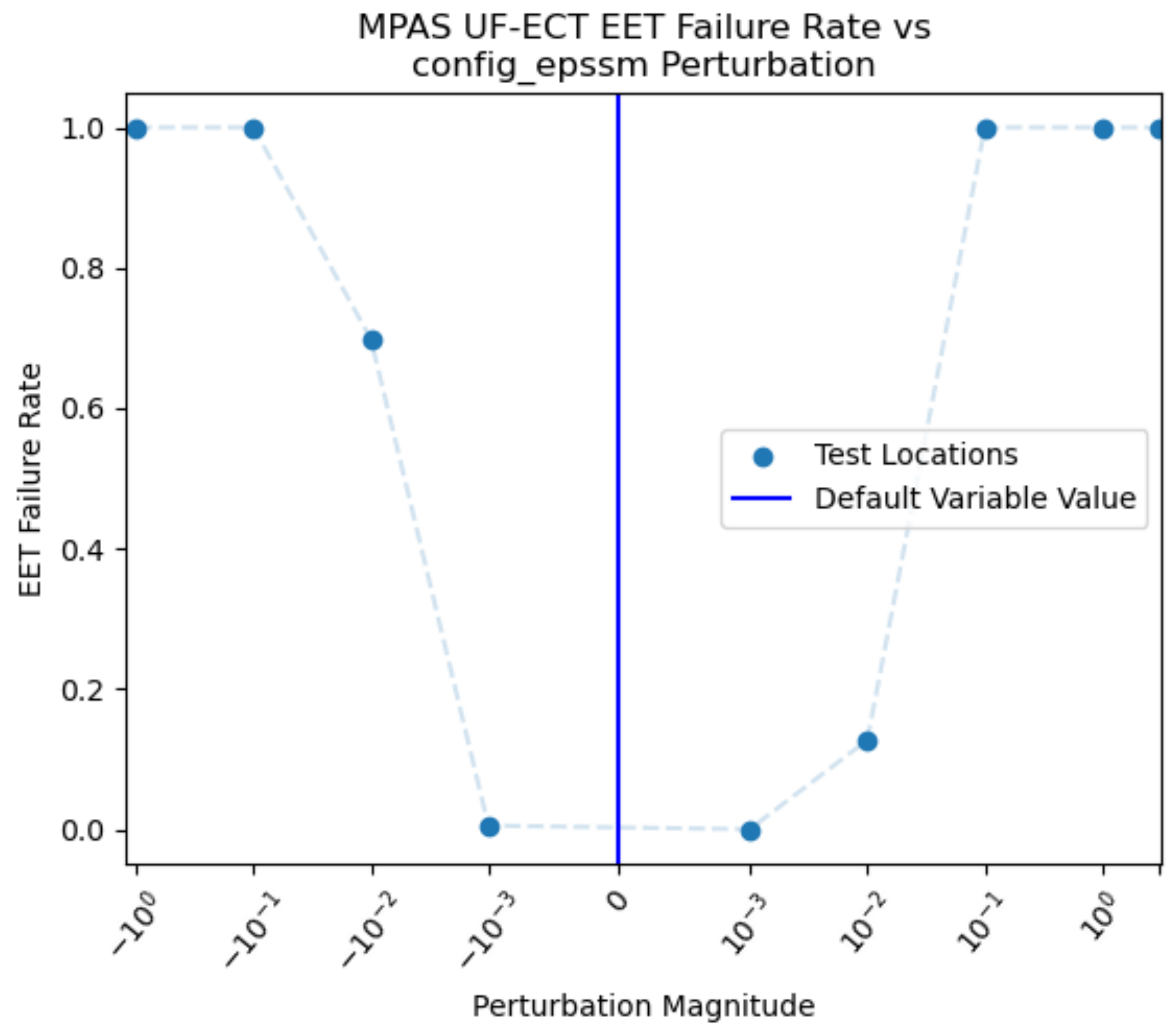
Climate Changing Parameters

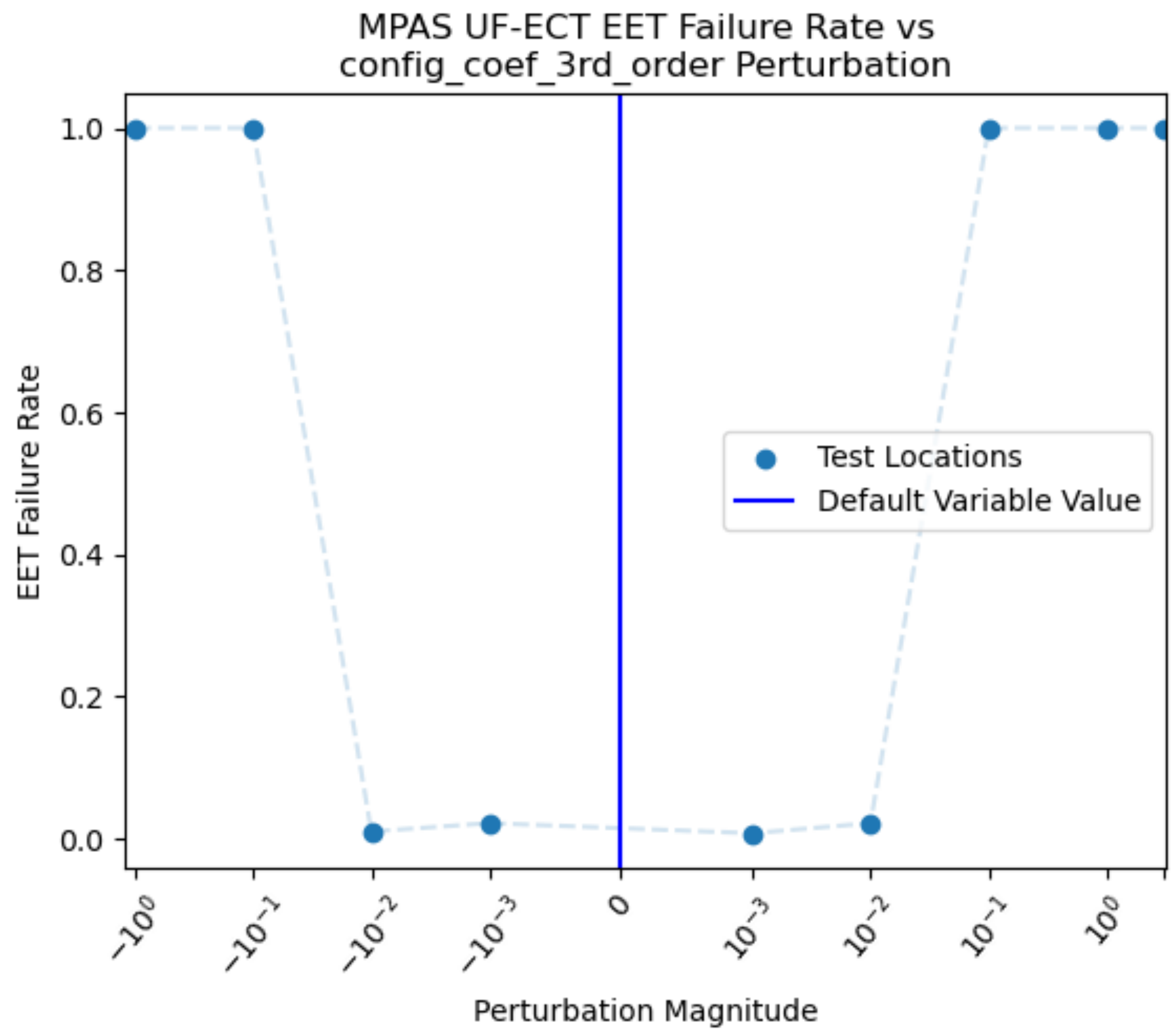
Parameter	Units	Description	Default Value
config_zd	m	Height MSL to begin w-damping profile.	22,000
config_xnutr	-	Maximum w-damping coefficient at model top	0.2
config_epssm	-	Off-centering parameter for the vertically implicit acoustic integration	0.1
config_coef_3rd_order	-	Upwinding coefficient in the 3rd order advection scheme	0.25

Climate Changing Parameters

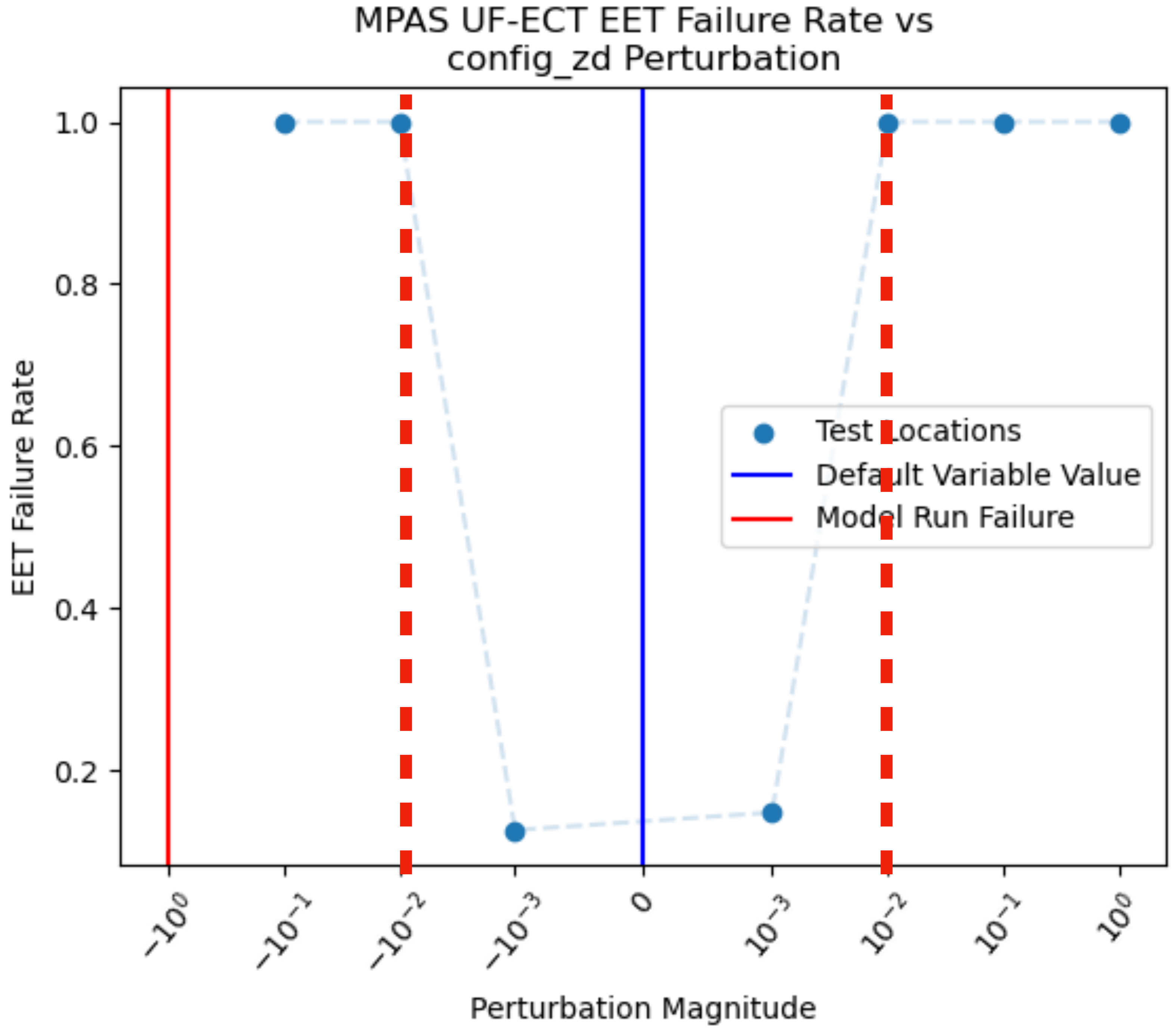


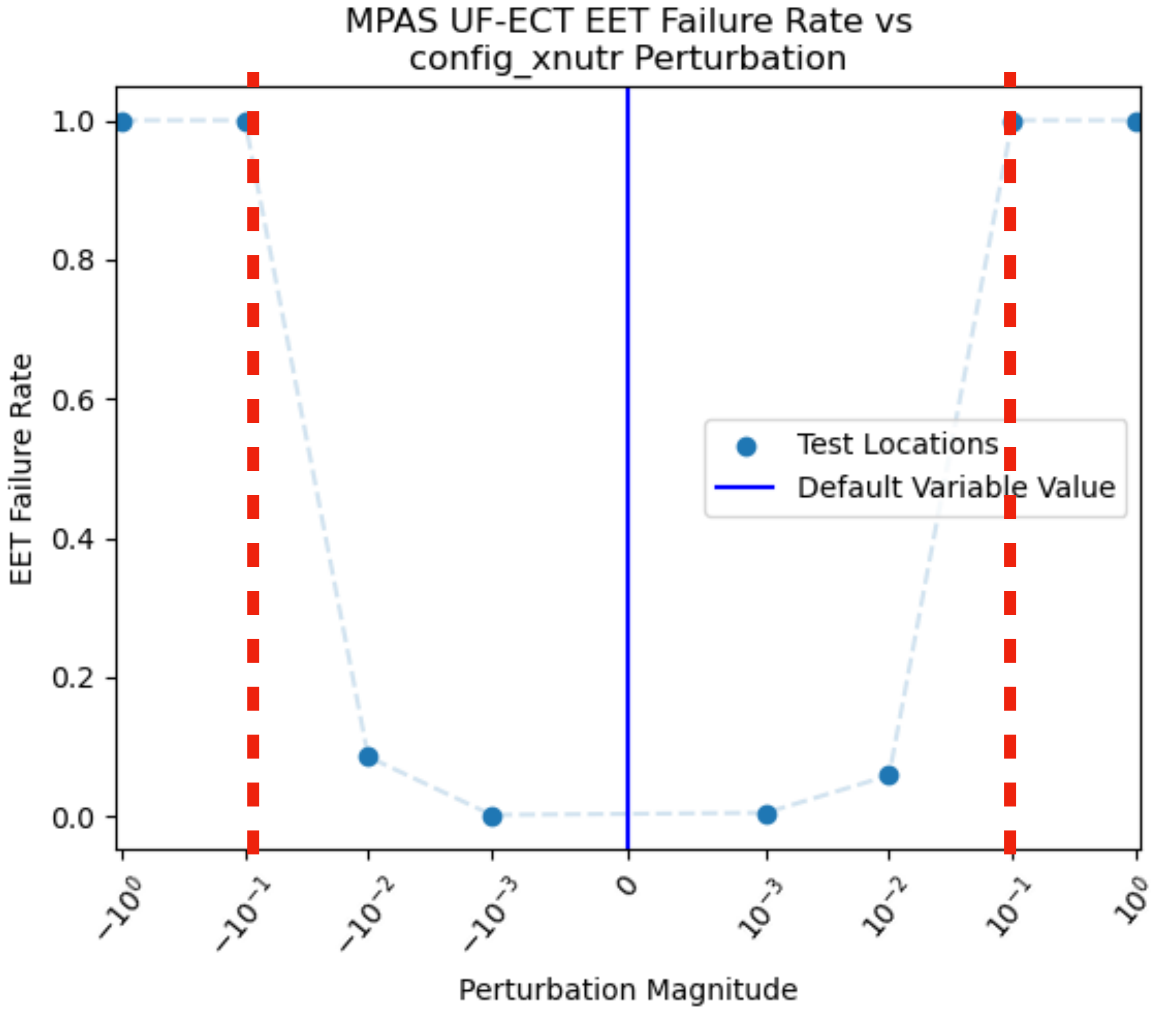


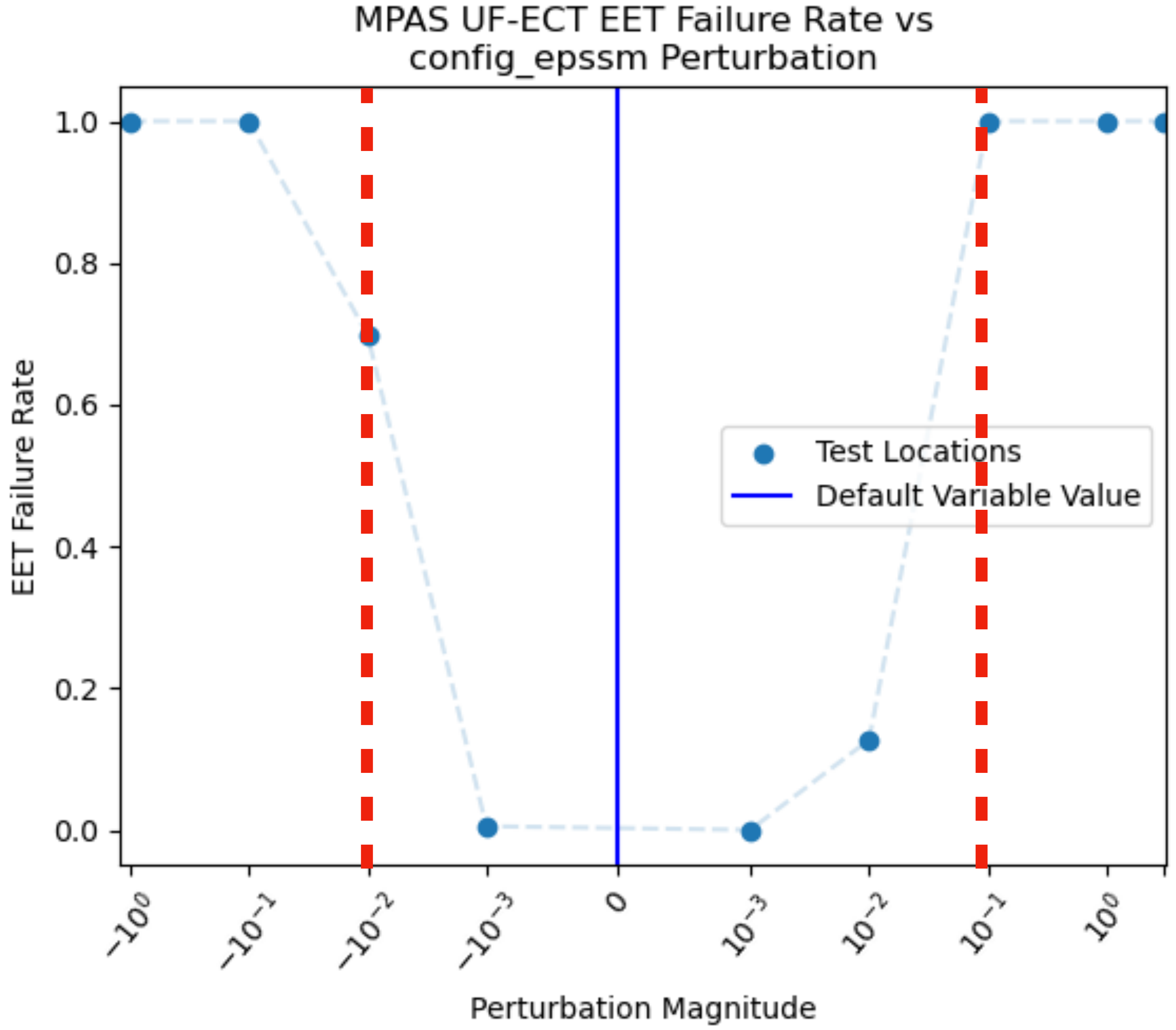


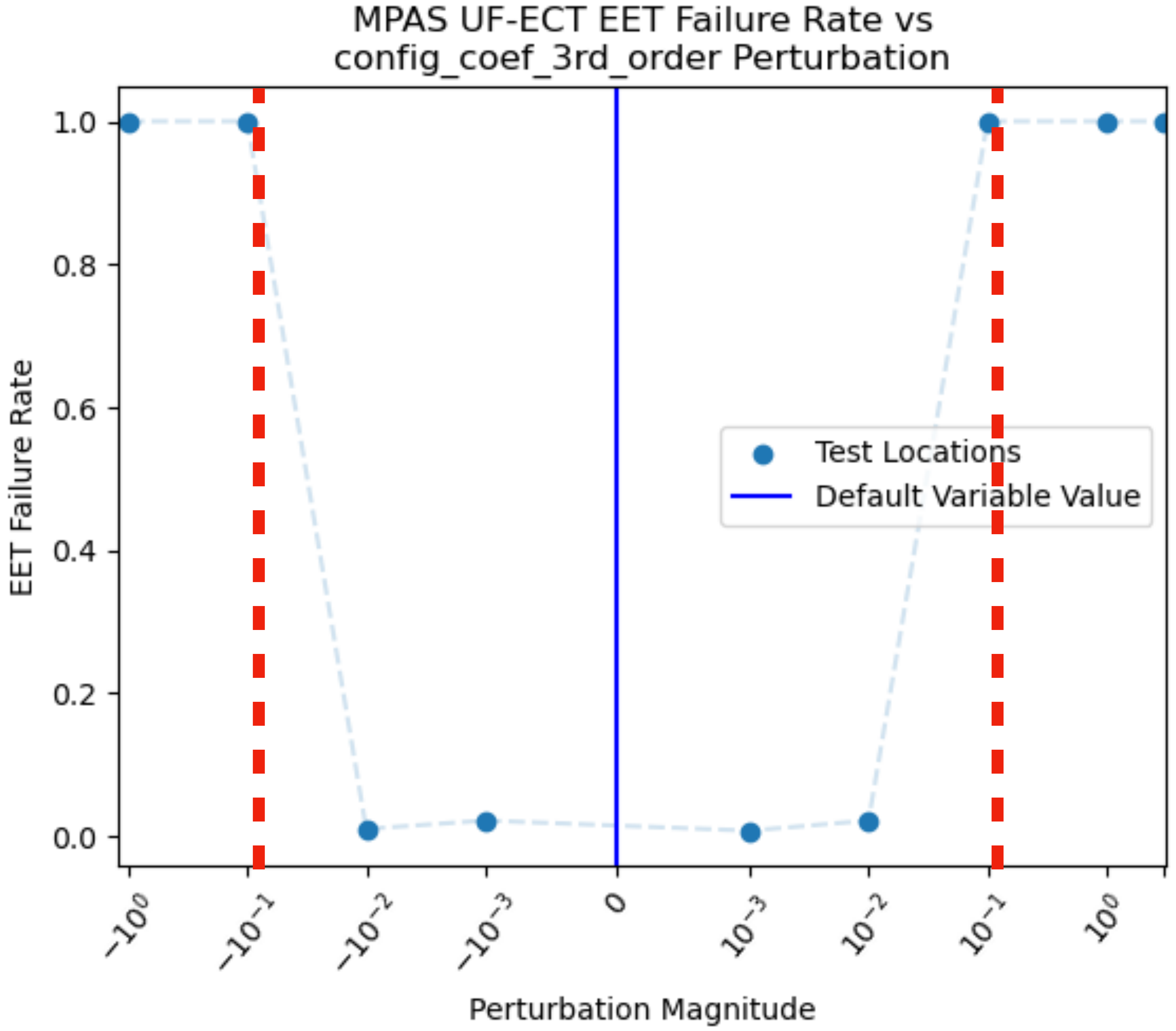


Climate Changing Parameters

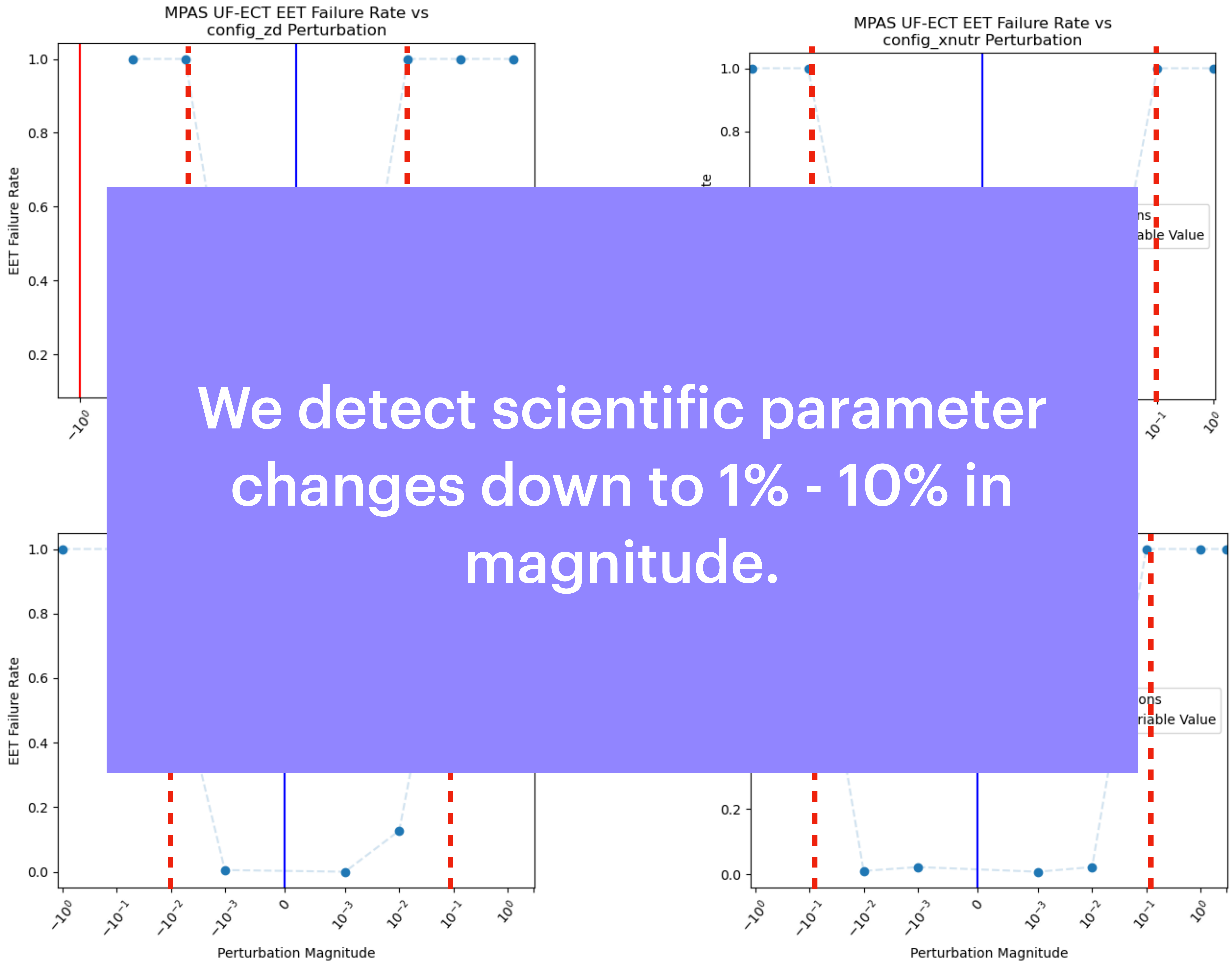








Climate Changing Parameters



We detect scientific parameter changes down to 1% - 10% in magnitude.

CESM-CAM: Do we need to update our test parameters?

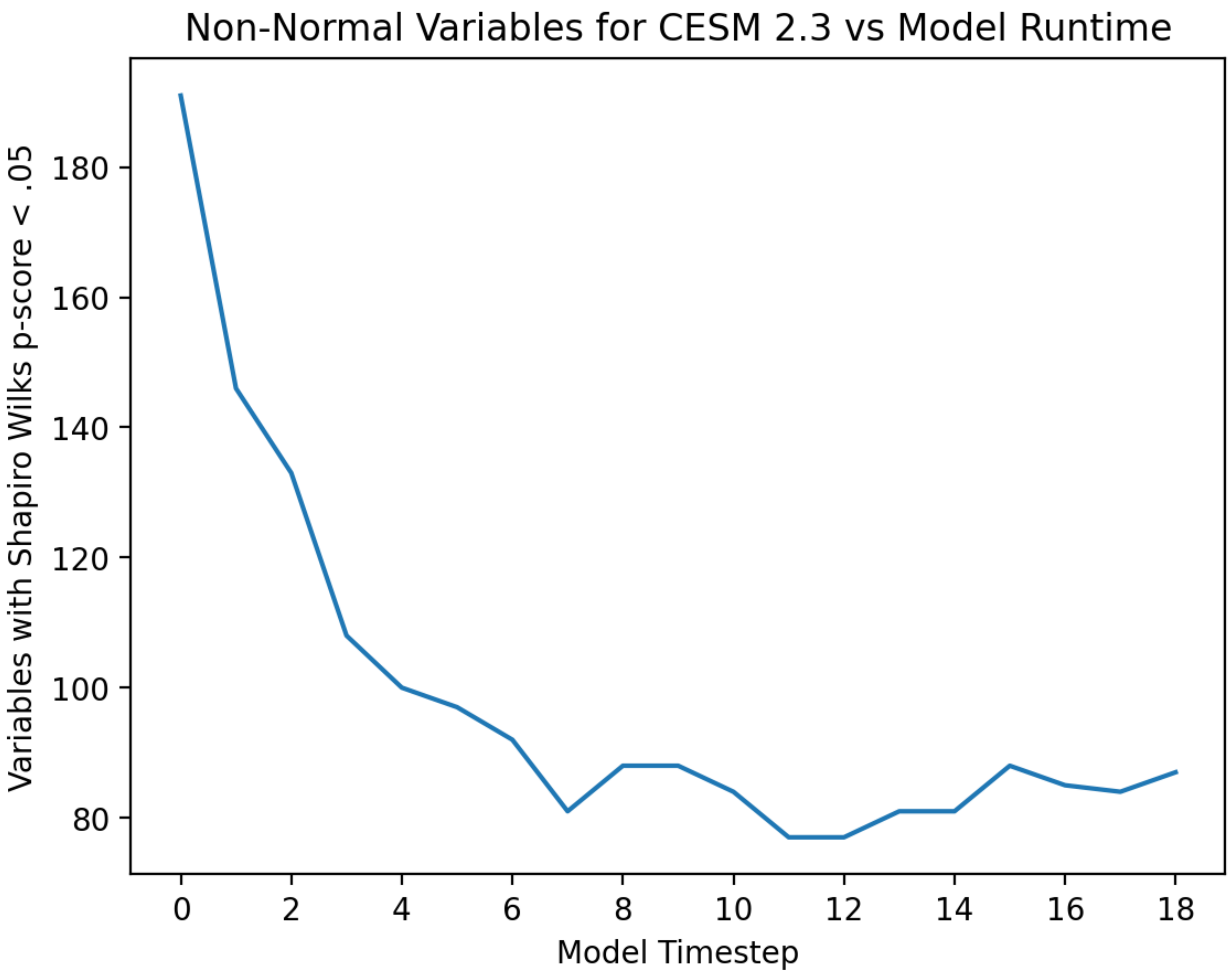
CAM 5.3

- 108 default variable outputs.
  - After exclusions of variables that introduce numerical issues in the test.
- T = 9 timesteps (4.5 hours)
- N_{PC} = 50
- m_{σ} = 2
- N_{ens} = 350

CAM 6.3

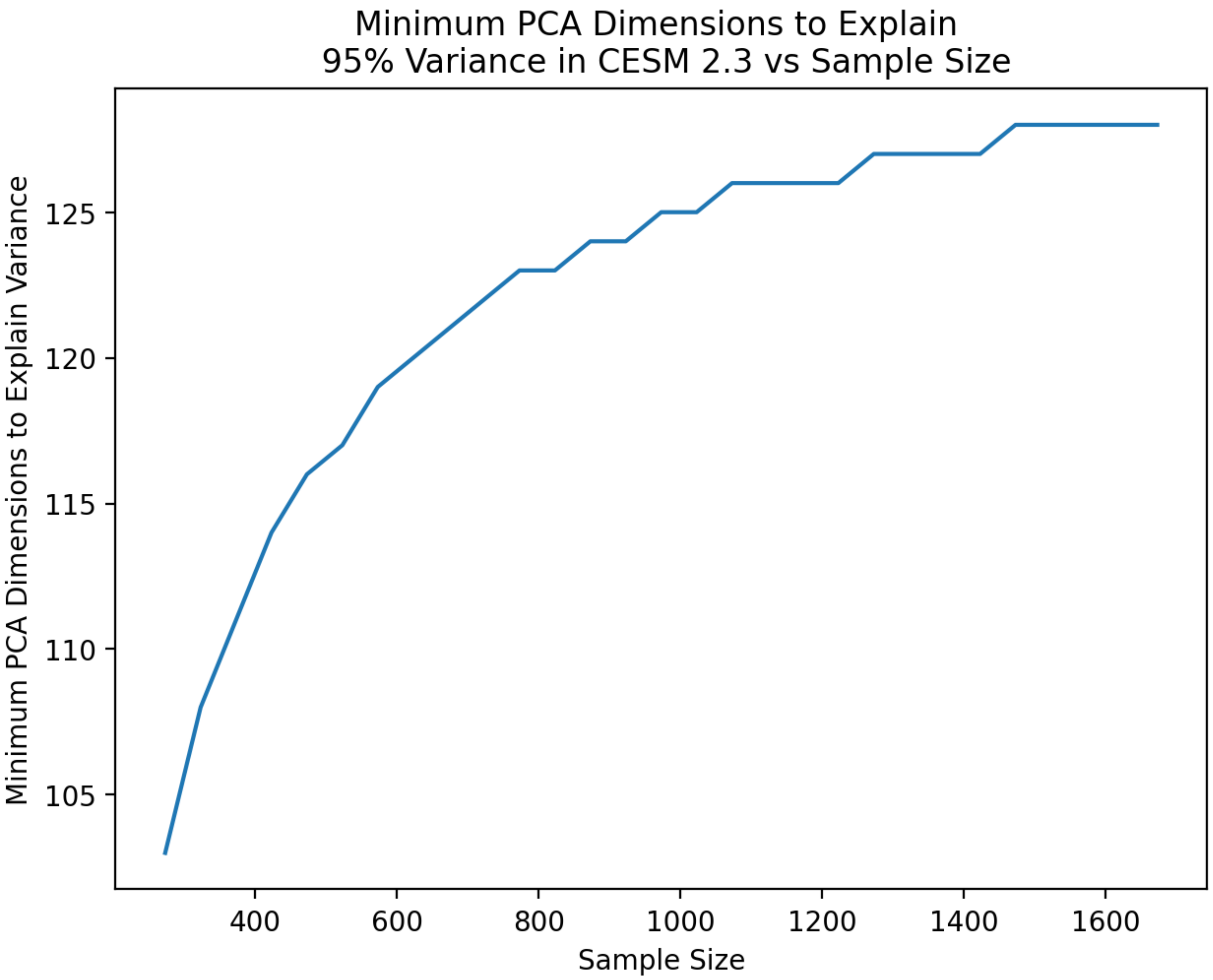
- 275 default output variables
- New physics!

CAM 6.3



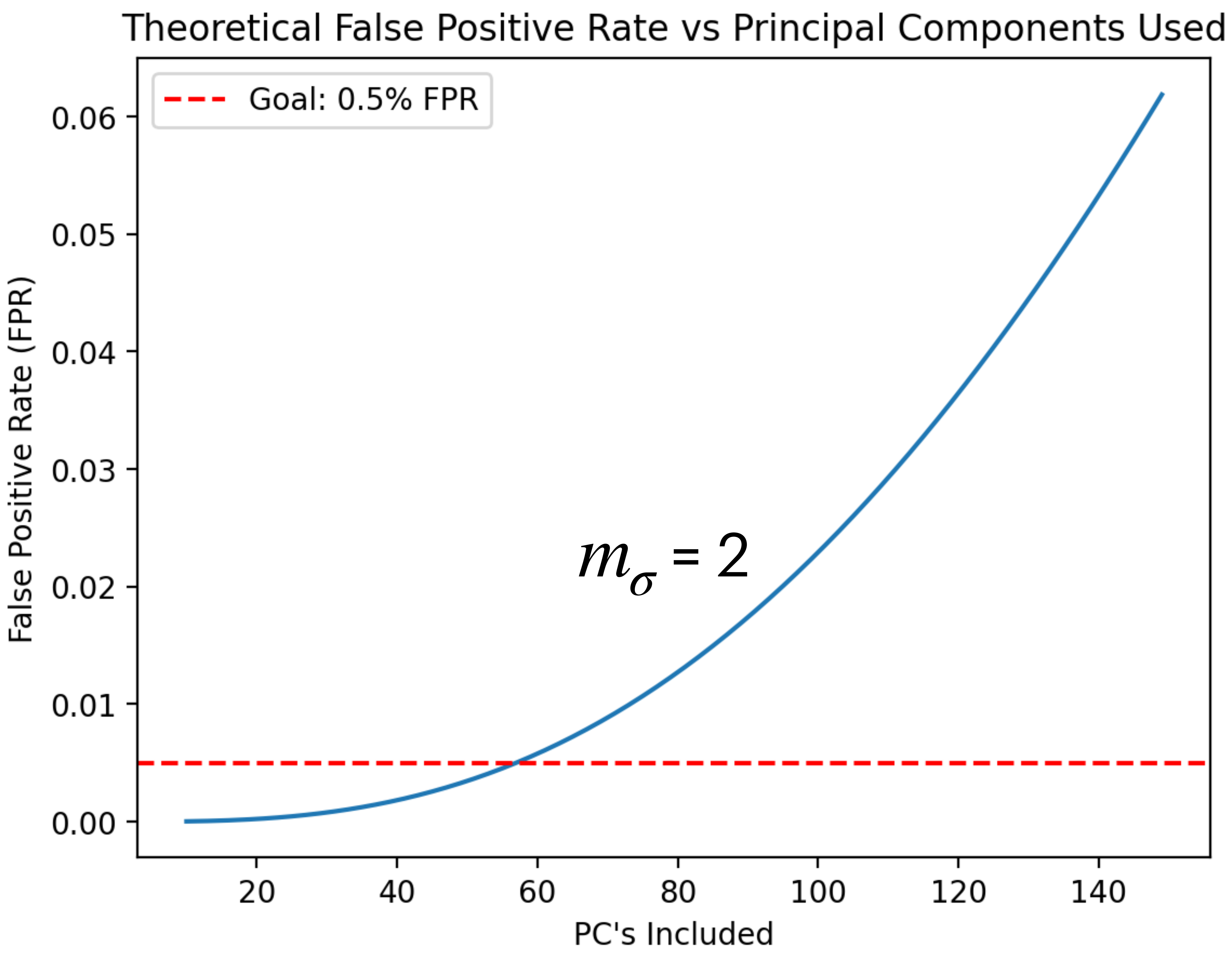
- 275 default output variables
- New physics!
- T > 7 timesteps

CAM 6.3



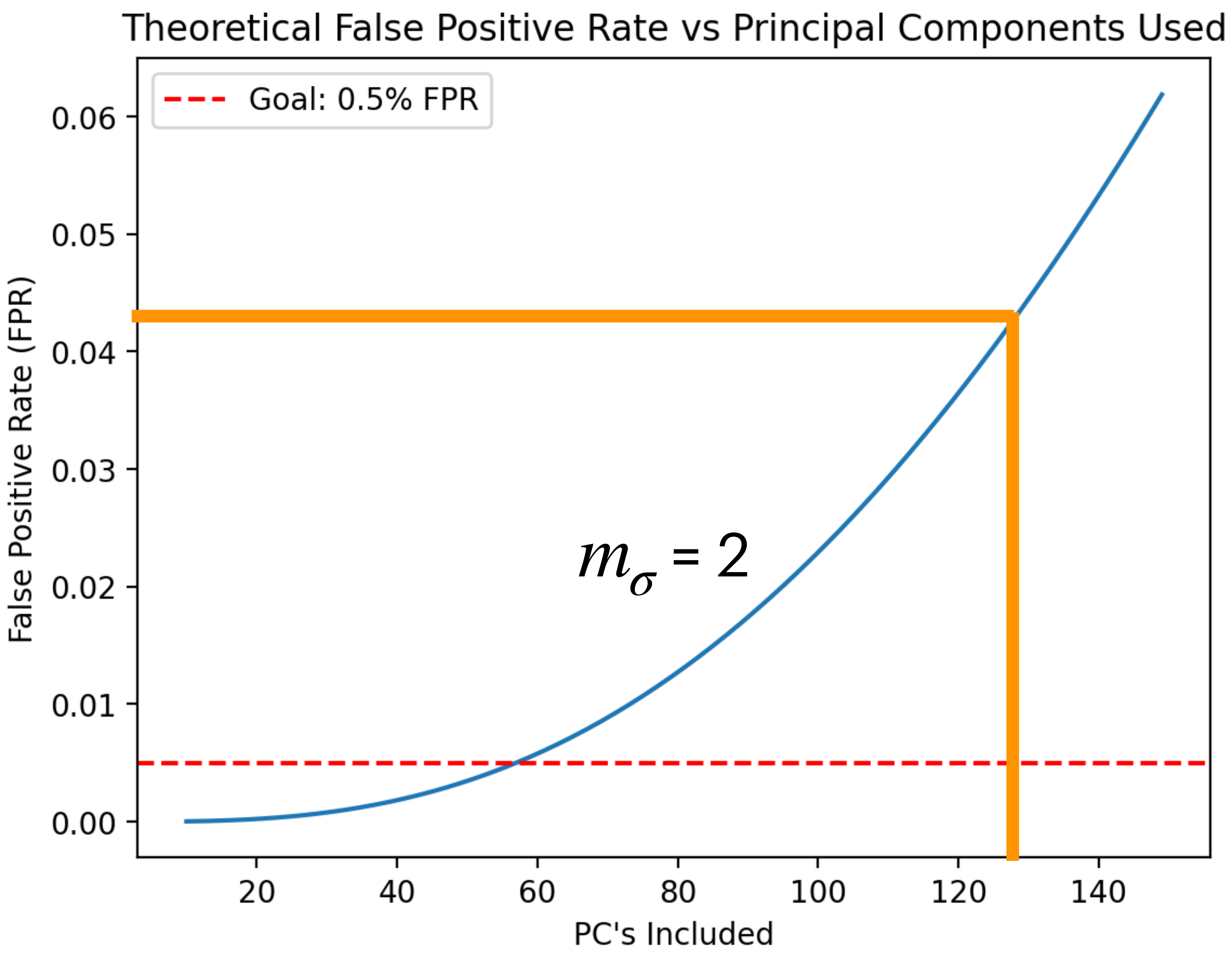
- 275 default output variables
- New physics!
- T > 7 timesteps
- N_{PC} = 128

CAM 6.3



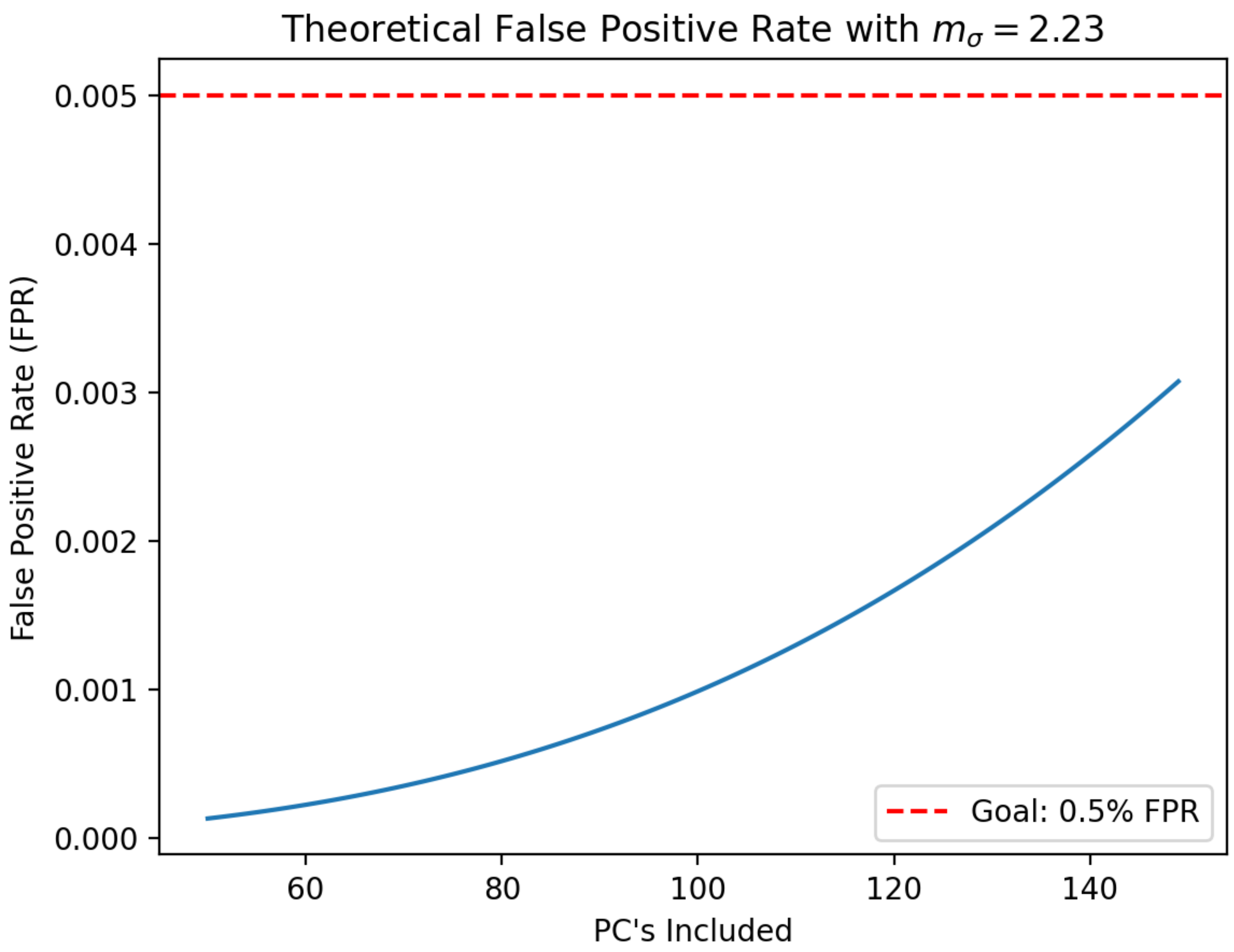
- 275 default output variables
- New physics!
- $T > 7$ timesteps
- $N_{\mathrm{PC}} = 128$
- $m_\sigma = ?$

CAM 6.3



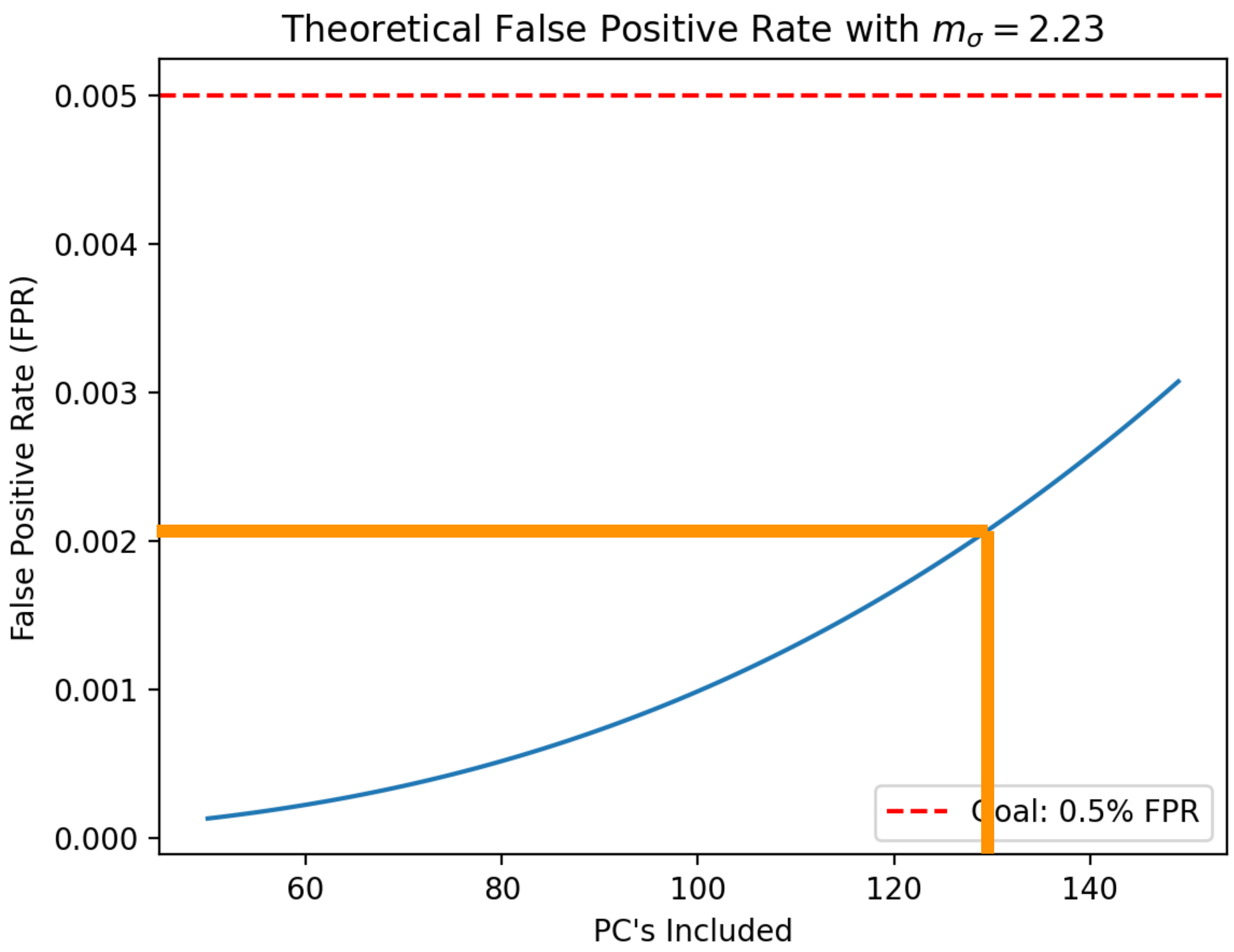
- 275 default output variables
- New physics!
- $T > 7$ timesteps
- $N_{\mathrm{PC}} = 128$
- $m_\sigma = ?$

CAM 6.3



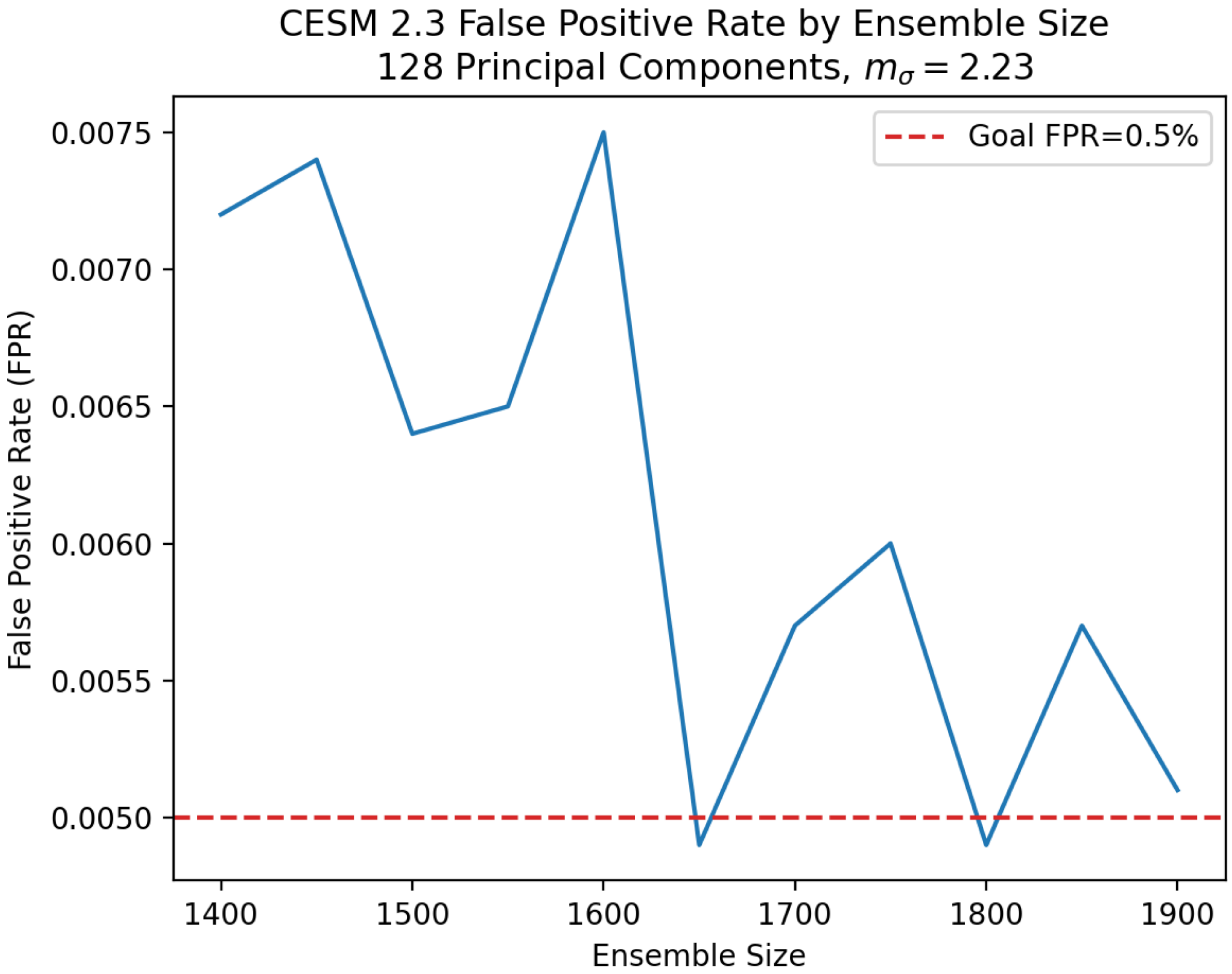
- 275 default output variables
- New physics!
- T > 7 timesteps
- N_{PC} = 128
- m_σ = 2.23

CAM 6.3



- 275 default output variables
- New physics!
- T > 7 timesteps
- N_{PC} = 128
- m_σ = 2.23

CAM 6.3



- 275 default output variables
- New physics!
- T > 7 timesteps
- N_{PC} = 128
- m_σ = 2.23
- N_{ens} = 1650

CAM 5.3

- 108 default variable outputs.
- T = 9 timesteps (4.5 hours)
- N_{PC} = 50
- m_σ = 2
- N_{ens} = 350

CAM 6.3

- 275 default output variables
- New physics!
- T > 7 timesteps
- N_{PC} = 128
- m_σ = 2.23
- N_{ens} = 1650

CAM 5.3

- 108 default
- T = 9 times
- N_{PC} = 50
- m_σ = 2
- N_{ens} = 350

CAM 6.3

- put variables

Big change in test parameters many be required when model changes!

Thanks!

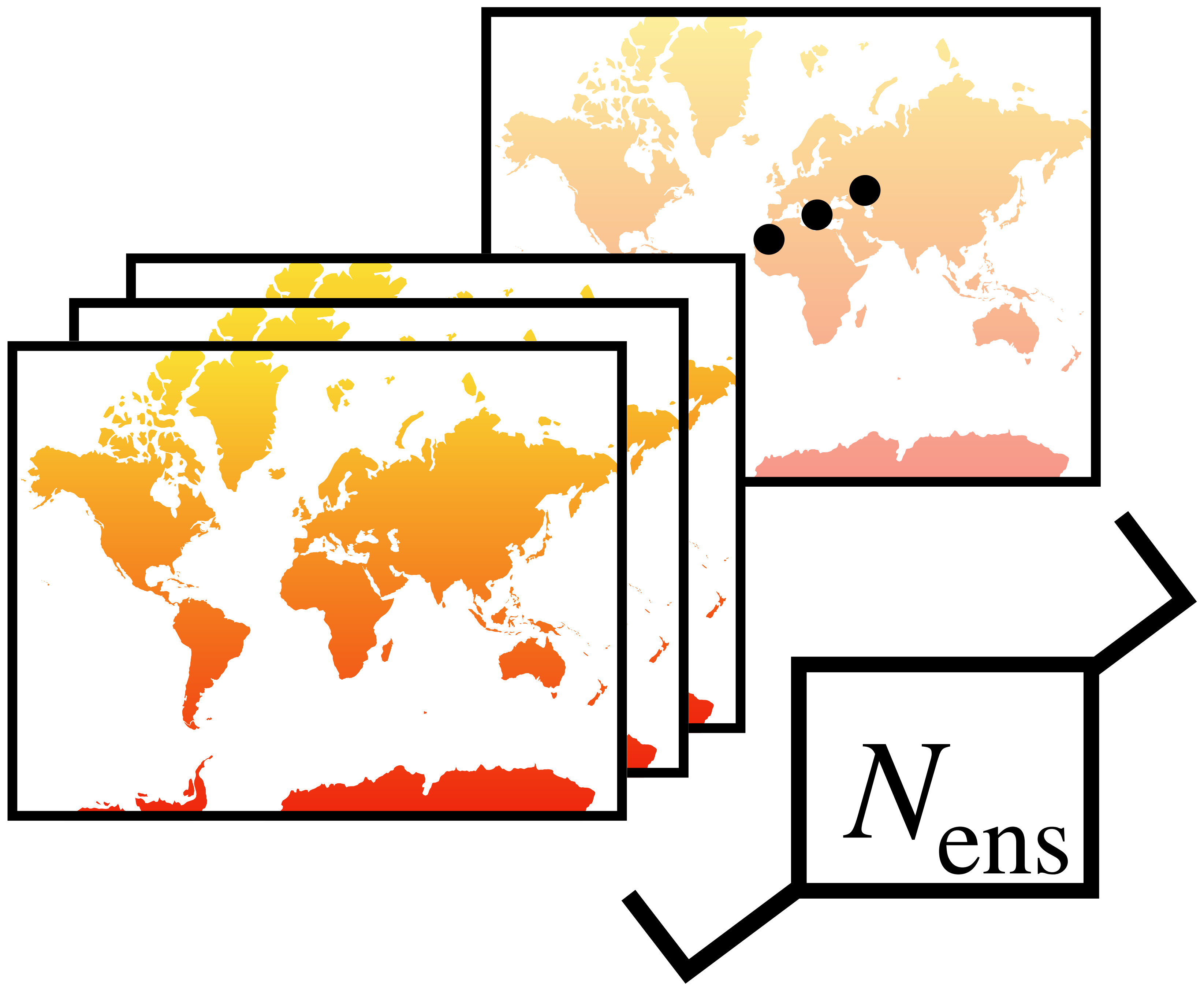
Back Slides

The UF-ECT: Part 1

1.1) Start with set of initial conditions (IC's).

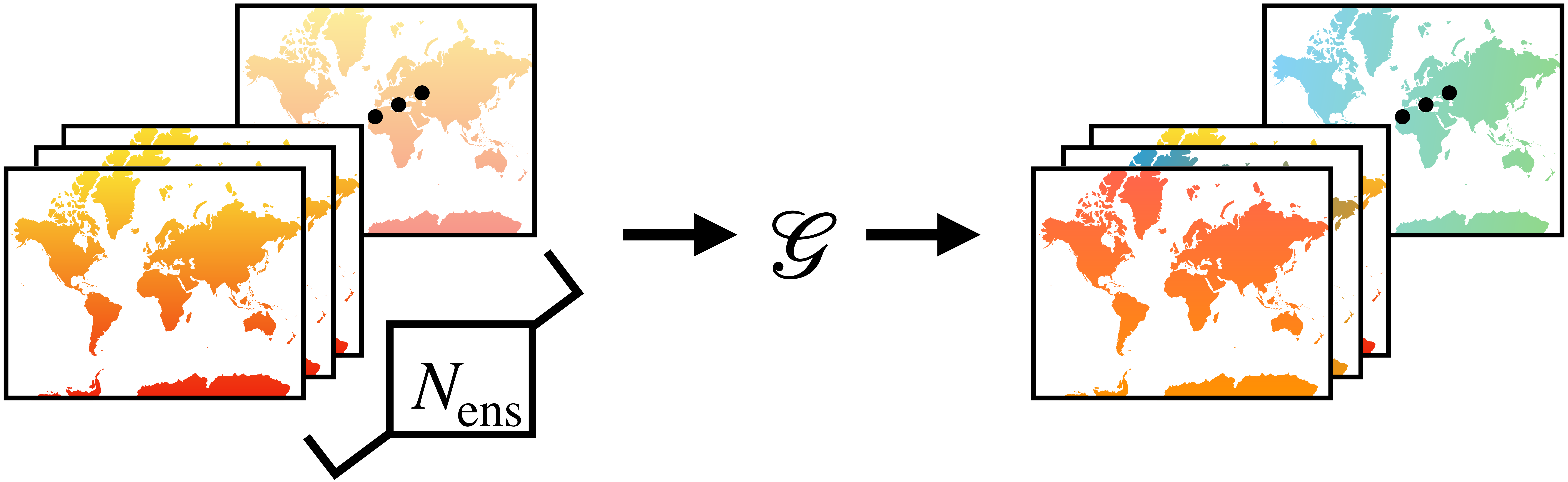
The UF-ECT: Part 1

1.2) Perturb IC's to create ensemble.



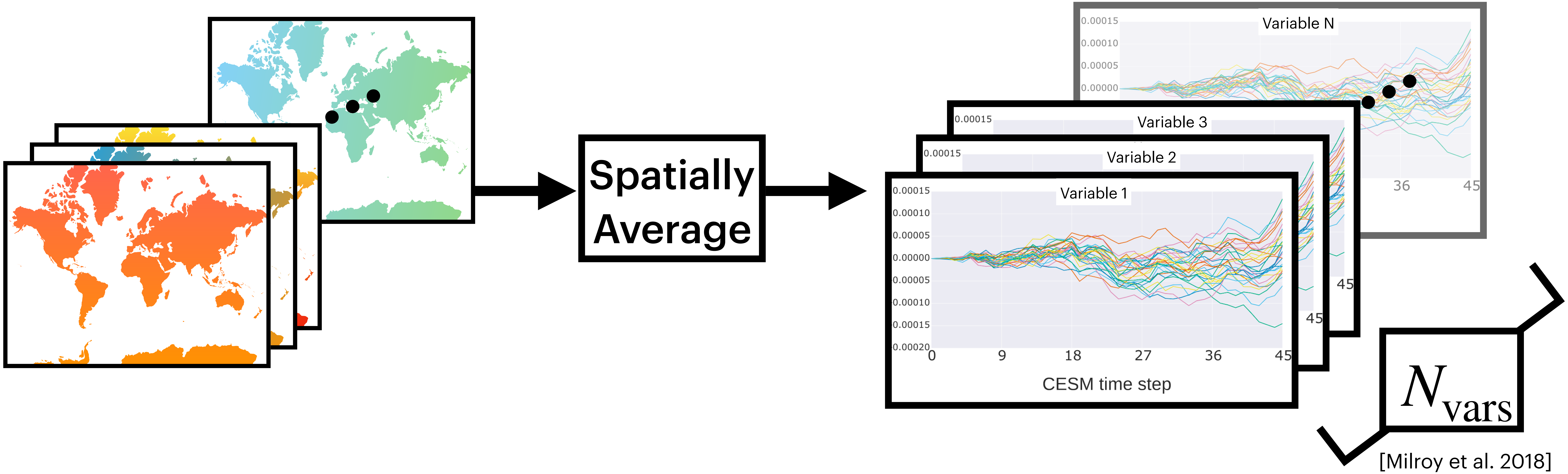
The UF-ECT: Part 1

1.3) Run model for simulation length T using ensemble of IC's to create ensemble of outputs.



The UF-ECT: Part 1

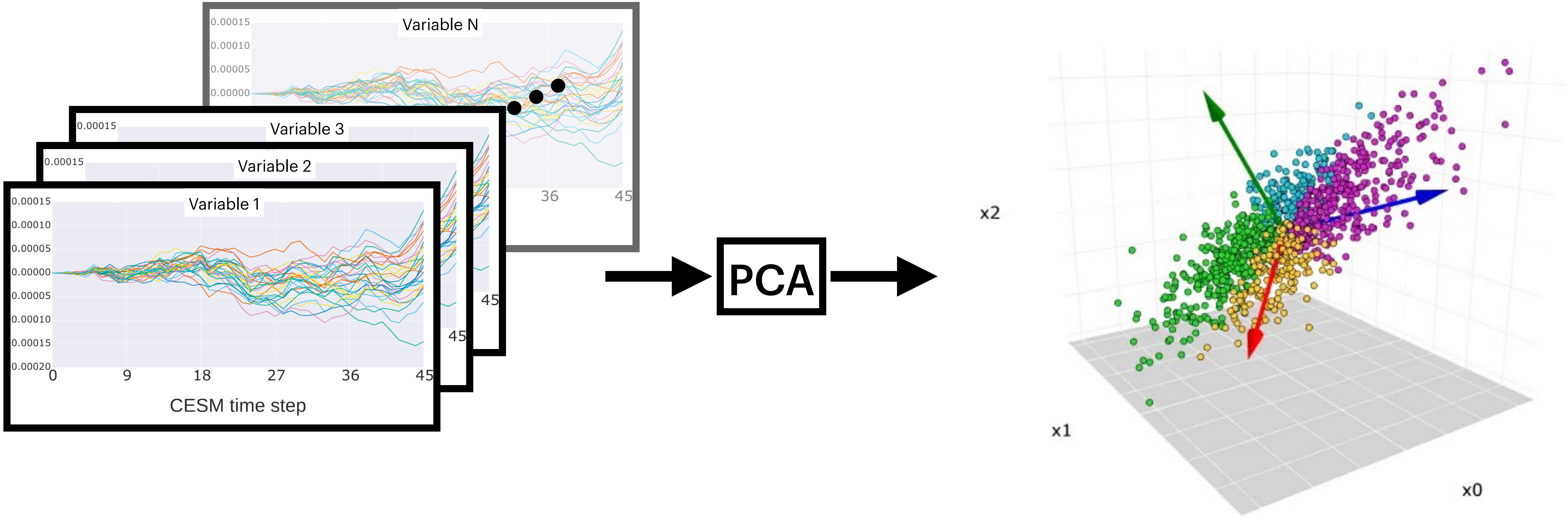
1.4) Spatially average model outputs.



[Milroy et al. 2018]

The UF-ECT: Part 1

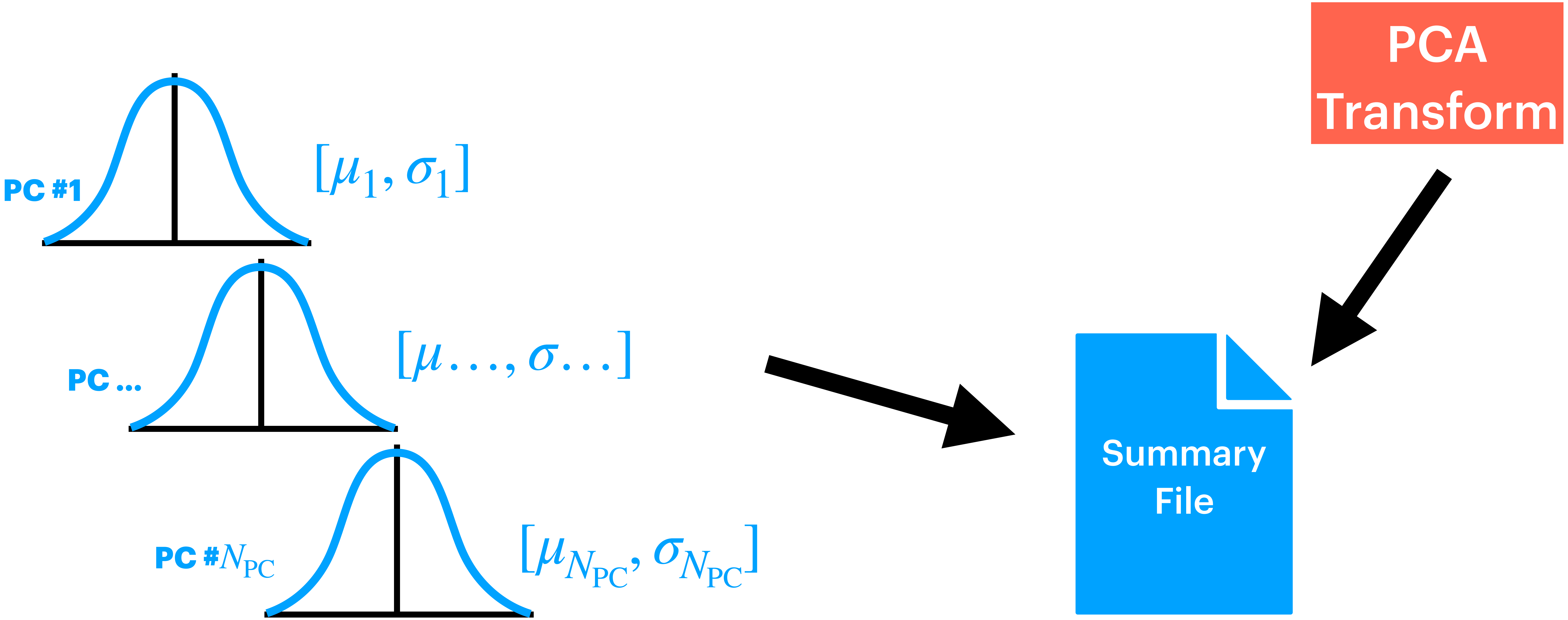
1.5) Use Principal Component Analysis to find orthogonal basis that explains most of the variance in the ensemble.



[Milroy et al. 2018]

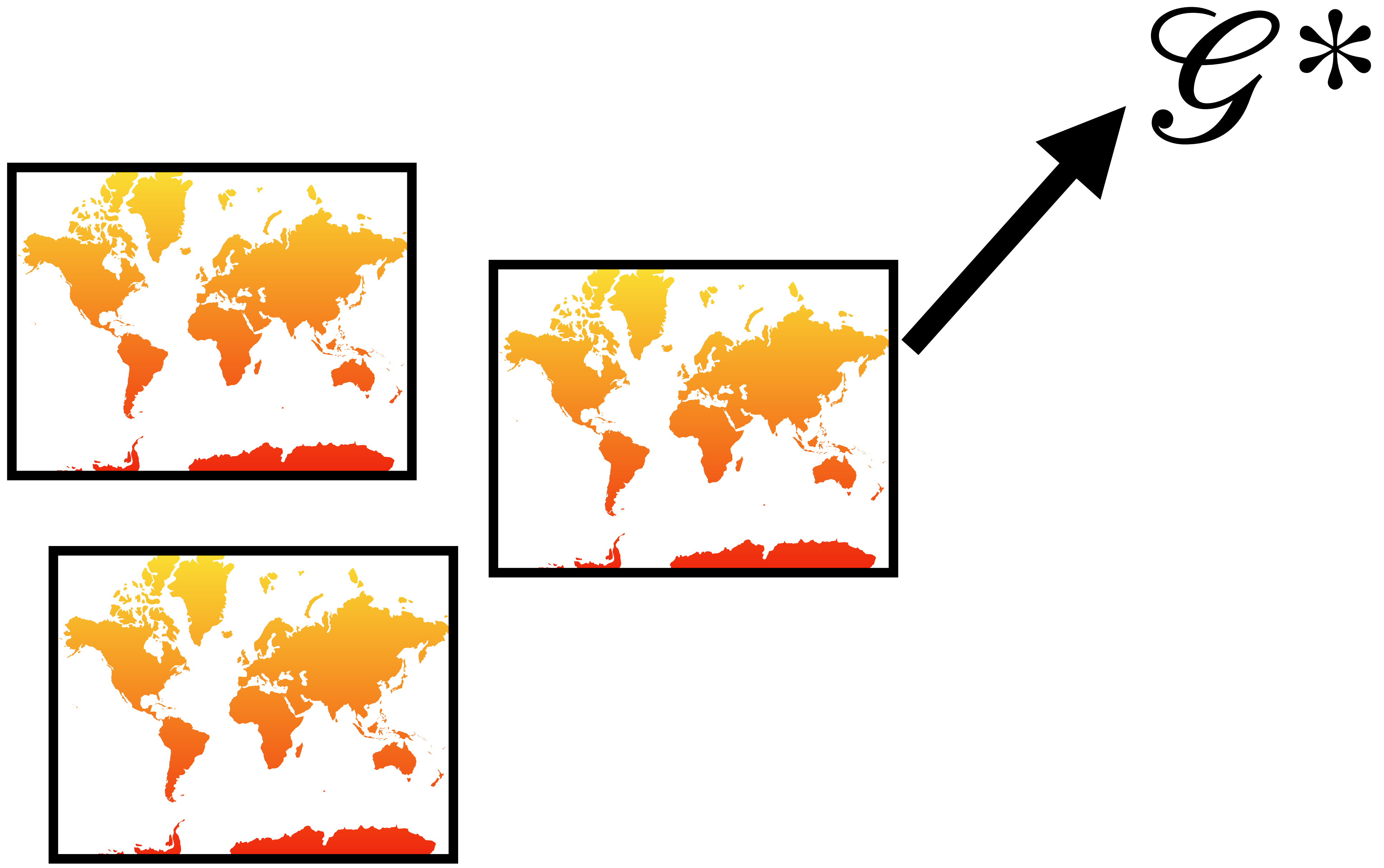
The UF-ECT: Part 1

1.6) Along with PCA transform, save PC distributions of ensemble to summary.



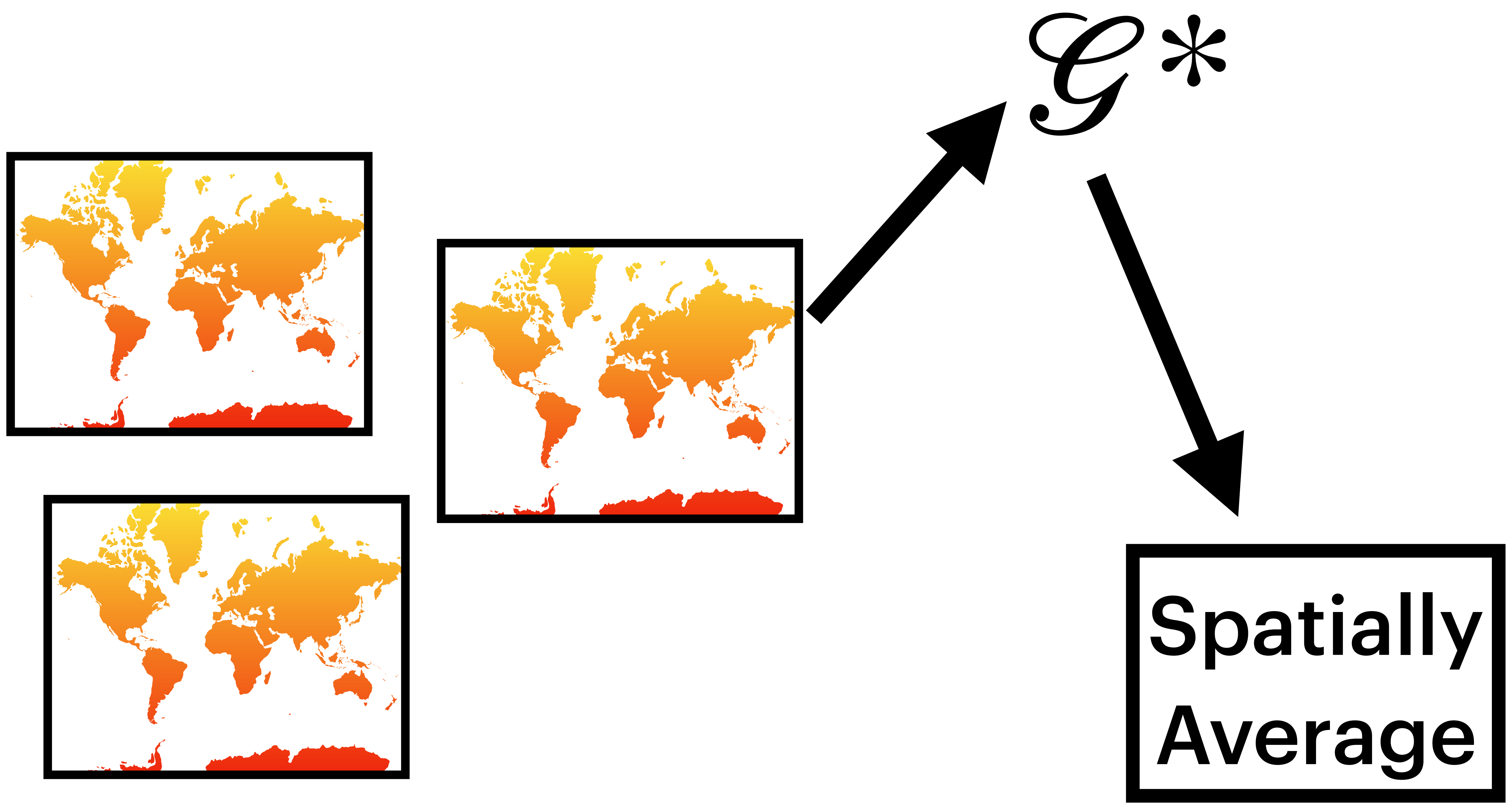
The UF-ECT: Part 2

2.1) Run small set of perturbed runs using new model configuration.



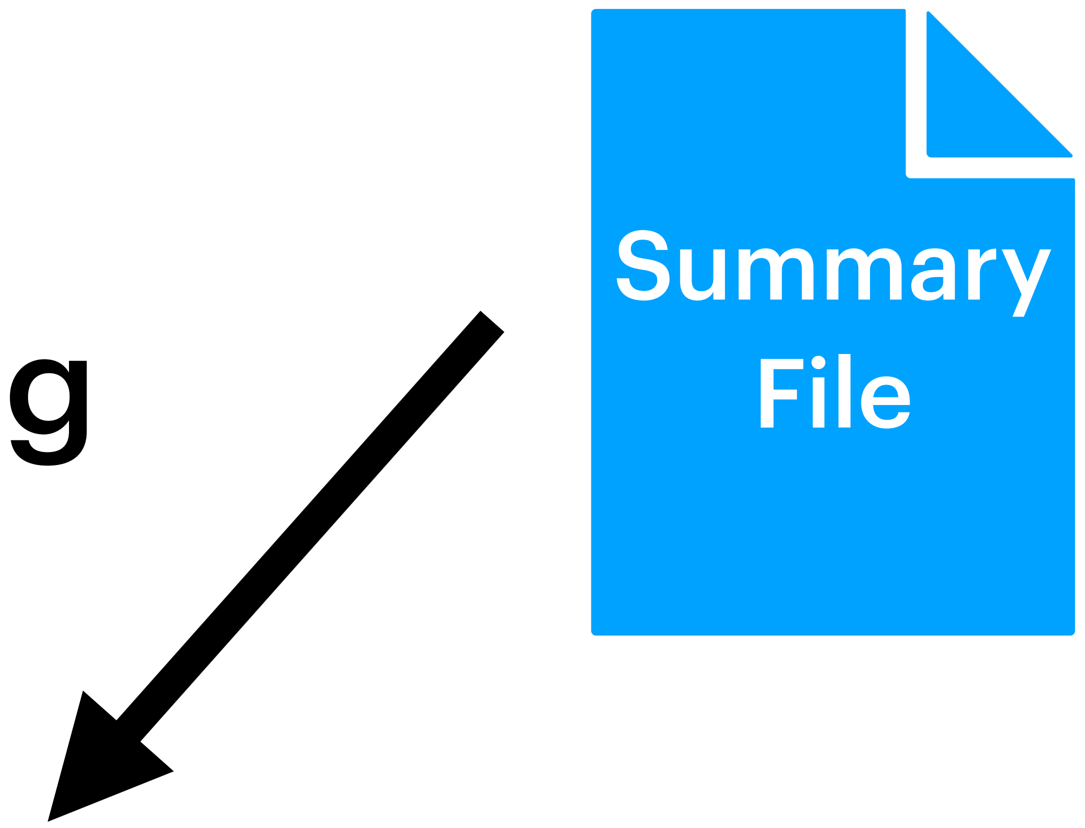
The UF-ECT: Part 2

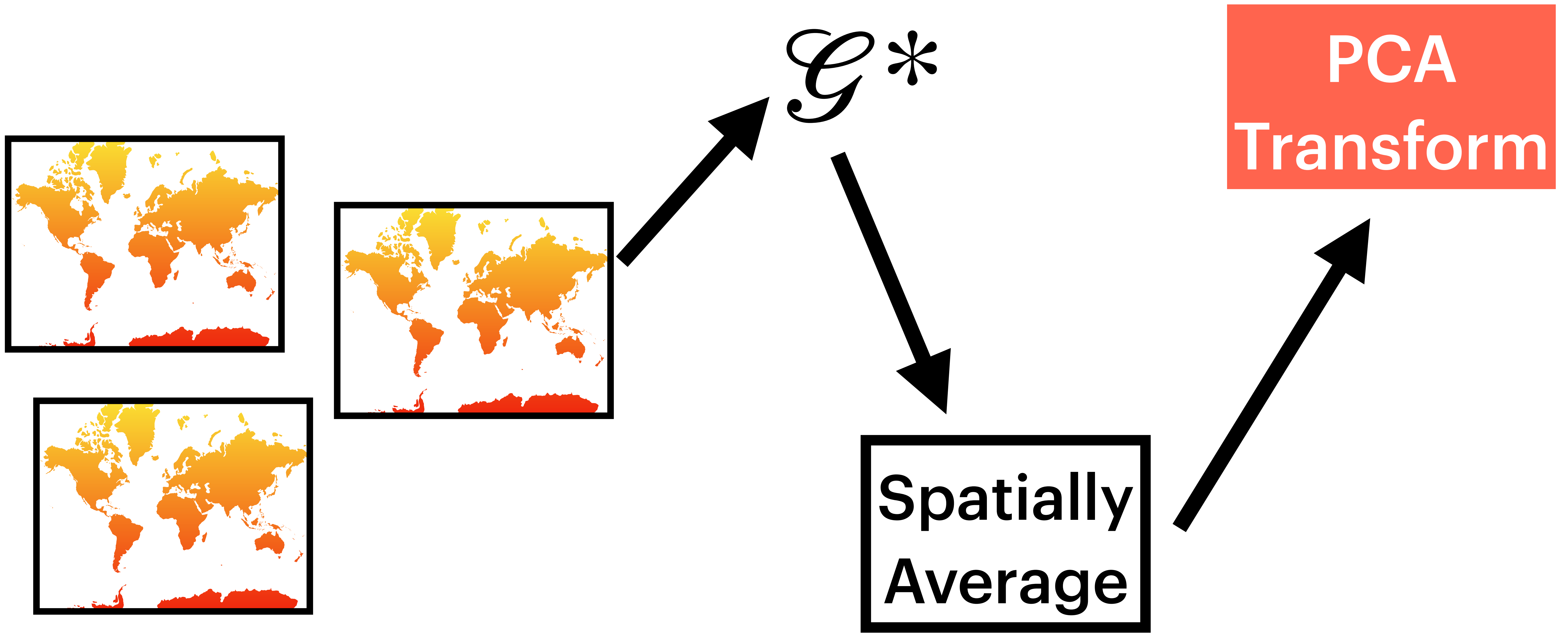
2.2) Spatially average new model outputs.



The UF-ECT: Part 2

2.3) Transform new average outputs using saved PCA transform.





The UF-ECT: Part 2

2.4) Compare transformed outputs to original distributions. Issue pass/fail based on result.