

# Quantifying the Agreement between Observed and Simulated Extratropical Modes of Interannual Variability

*-- Routine Evaluation of Various Climate Models  
based on Big-data Sharing Ecosystem*

**Jiwoo Lee, Ph. D**

[\(jwlee@llnl.gov\)](mailto:jwlee@llnl.gov)

*Lawrence Livermore National Laboratory, California, USA*



# Contents

---

- **LLNL Introduction**
- **Part 1: Infrastructure Development for Climate Research**
  - Climate research as community efforts: CMIP & Obs4MIPs
  - Data sharing ecosystem for the community: ESGF
  - Advanced tool for climate big data analysis: CDAT
  - Help accelerating evolution of climate model: PMP
- **Part 2: Evaluation of Various Climate Models**
  - Quantifying the Agreement Between Observed and Simulated Extratropical Modes of Interannual Variability
- **Further On-going Works**

# Lawrence Livermore National Laboratory (LLNL)

## Department of Energy National Laboratories

**TOP 500**  
The List.

- 5 of Top 10 Supercomputers in DOE labs
- LLNL's Sequoia, 20PFlops, ranked at 5



Brookhaven National Laboratory

Ames Laboratory

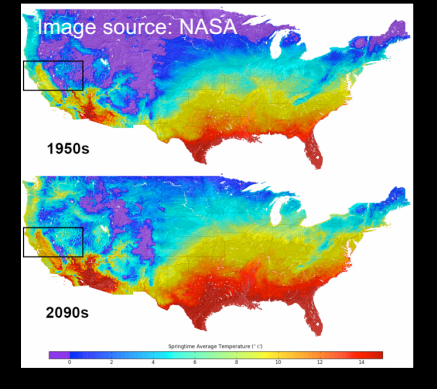
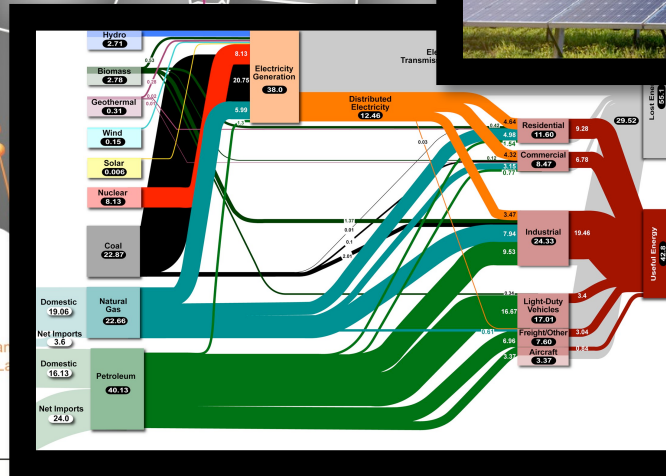
Lawrence Berkeley National Laboratory

Lawrence Berkeley National Laboratory

SLAC National Accelerator Laboratory



Lawrence Livermore National Laboratory



- Office of Science laboratory
- National Nuclear Security Administration laboratory
- Office of Fossil Energy laboratory
- Office of Energy Efficiency and Renewable Energy laboratory
- Office of Nuclear Energy, Science and Technology laboratory
- Office of Environmental Management laboratory

National Renewable Energy Laboratory



# PART 1. Infrastructure Development for Climate Research

*(next 15 slides...)*

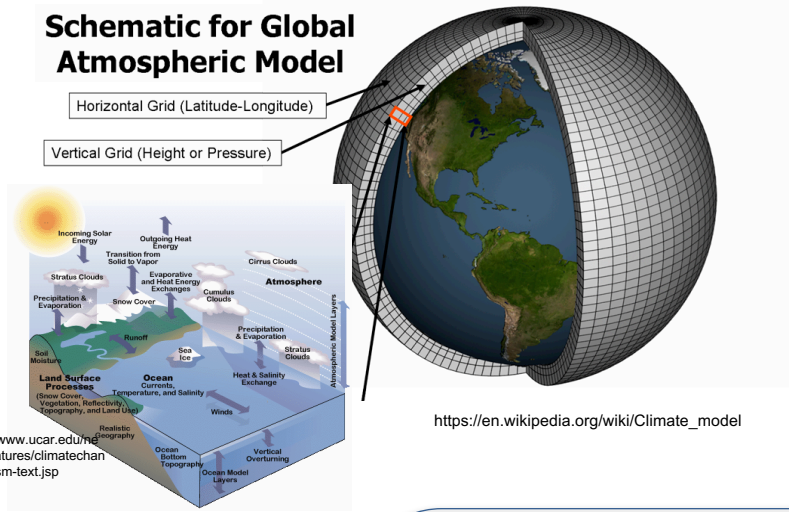
- *Climate research as community efforts: CMIP & Obs4MIPs*
- *Data sharing ecosystem for the community: ESGF*
- *Advanced tool for climate big data analysis: CDAT*
- *Help accelerating evolution of climate model: PMP*

*Acknowledgements to: Dean Williams, Charles Doutriaux, Denis Nadeau, et al. (AIMS Group);  
Peter Gleckler, Karl Taylor, Paul Durack, et al. (PCMDI)*

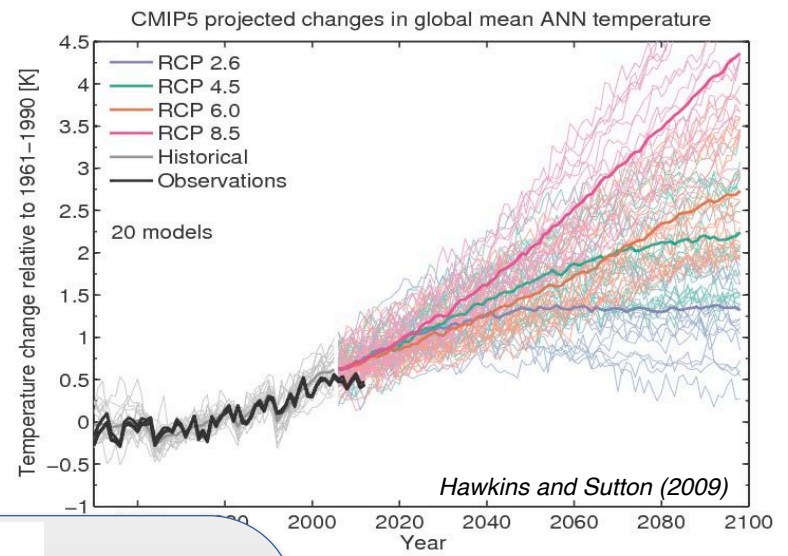


# CMIP: Climate Model Inter-comparison Project

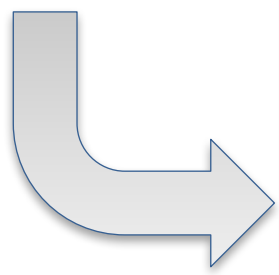
## Schematic for Global Atmospheric Model



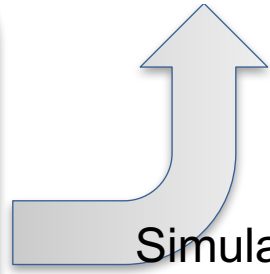
[https://en.wikipedia.org/wiki/Climate\\_model](https://en.wikipedia.org/wiki/Climate_model)



## Uncertainties...



Various climate models from all over the world!

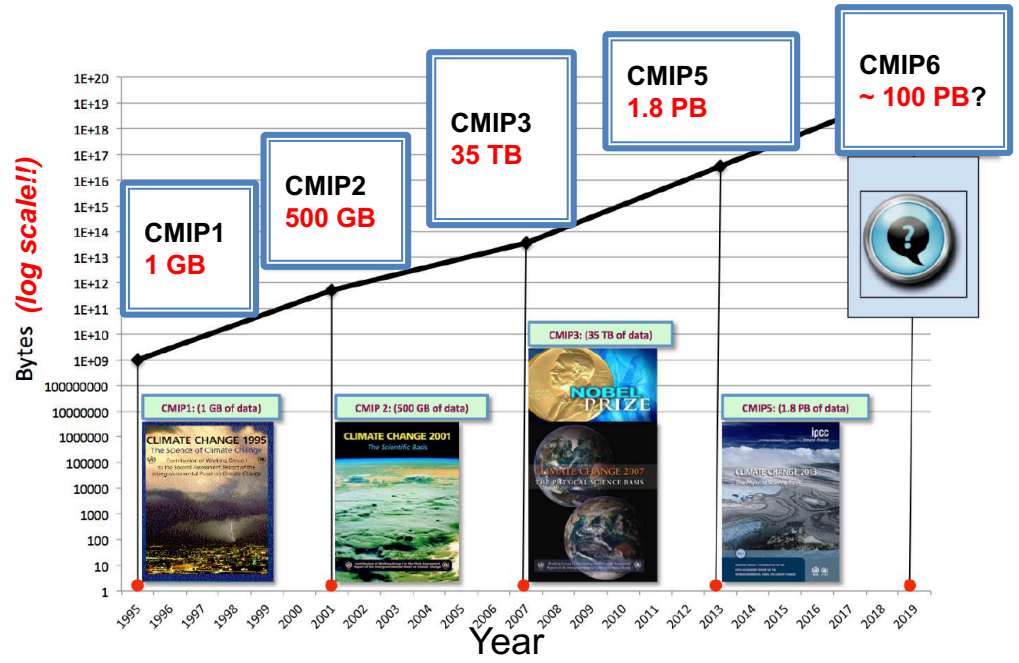


Simulate same target using different models!



# CMIP: Climate Model Inter-comparison Project

CMIP Data archived in **ESGF**  
Earth System Grid Federation



(Williams et al. 2015, BAMS)

- **CMIP** has collected a massive volume of climate model data
- CMIP research output have significantly contributed to publish **IPCC reports**



**IPCC Awards  
Nobel Prize**



PARIS2015  
UN CLIMATE CHANGE CONFERENCE  
COP21·CMP11



**Paris Agreement**

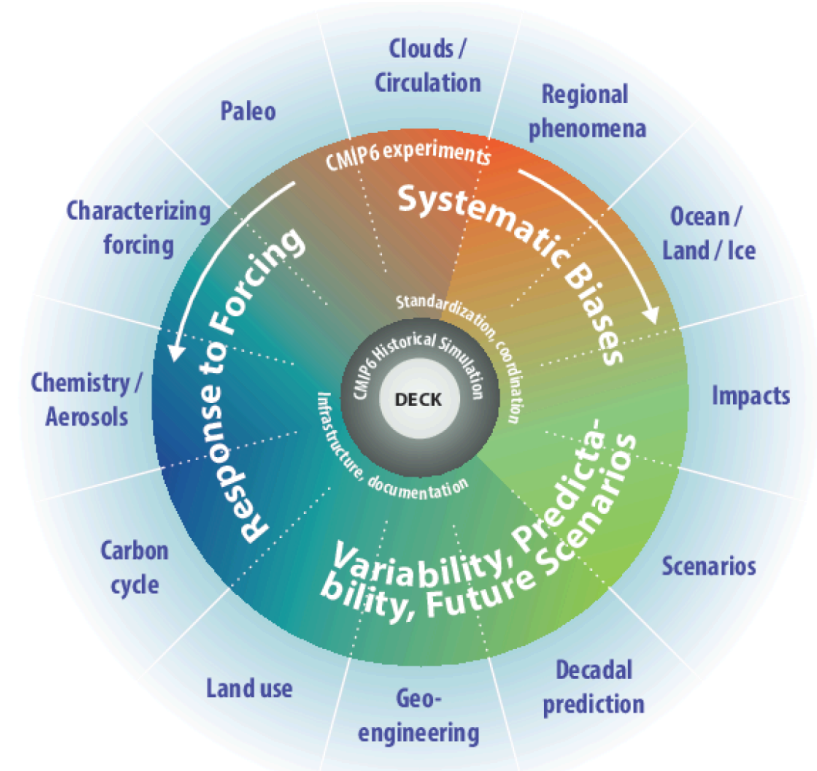


# CMIP:

## Climate Model Inter-comparison Project

### Scientific Questions of CMIP6:

- How does the Earth system respond to **forcing**?
- What are the origins and consequences of systematic **model biases**?
- How can we assess **future climate changes** given internal climate variability, predictability, and uncertainties in scenarios?



Schematic of the CMIP6 experiment design

(Eyring et al. 2016, GMD)



# Other MIPs!

<b>AerChemMIP</b>	Aerosols and Chemistry Model Intercomparison Project
<b>C4MIP</b>	Coupled Climate Carbon Cycle Model Intercomparison Project
<b>CFMIP</b>	Cloud Feedback Model Intercomparison Project
<b>DAMIP</b>	Detection and Attribution Model Intercomparison Project
<b>DCPP</b>	Decadal Climate Prediction Project
<b>FAFMIP</b>	Flux-Anomaly-Forced Model Intercomparison Project
<b>GeoMIP</b>	Geoengineering Model Intercomparison Project
<b>GMMIP</b>	Global Monsoons Model Intercomparison Project
<b>HighResMIP</b>	High-Resolution Model Intercomparison Project
<b>ISMIP6</b>	Ice Sheet Model Intercomparison Project for CMIP6
<b>LS3MIP</b>	Land Surface, Snow and Soil Moisture
<b>LUMIP</b>	Land-Use Model Intercomparison Project
<b>OMIP</b>	Ocean Model Intercomparison Project
<b>PMIP</b>	Palaeoclimate Modelling Intercomparison Project
<b>RFMIP</b>	Radiative Forcing Model Intercomparison Project
<b>ScenarioMIP</b>	Scenario Model Intercomparison Project
<b>VoMIP</b>	Volcanic Forcings Model Intercomparison Project
<b>CORDEX</b>	Coordinated Regional Climate Downscaling Experiment
<b>DynVarMIP</b>	Dynamics and Variability Model Intercomparison Project
<b>SIMIP</b>	Sea Ice Model Intercomparison Project
<b>VIACS AB</b>	Vulnerability, Impacts, Adaptation and Climate Services Advisory Board

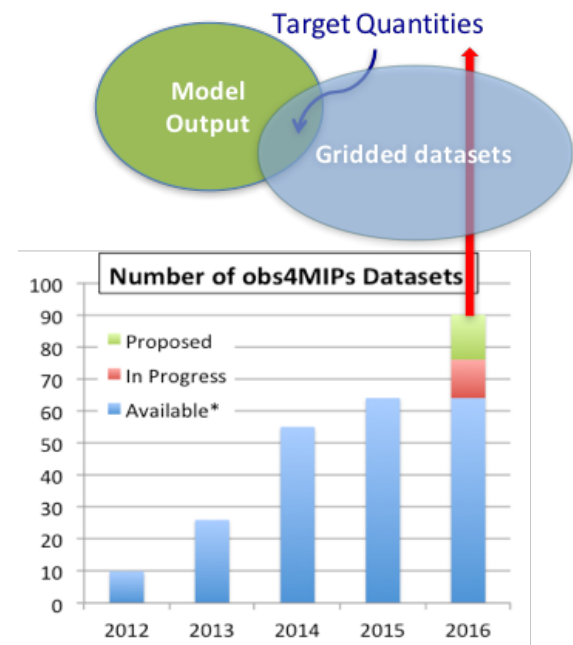
:

<https://www.wcrp-climate.org/wgcm-overview>



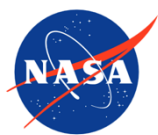
# Obs4MIPs: Observations for Model Intercomparisons Project

- A Project for identifying, documenting and disseminating observations for climate model evaluation in WCRP model intercomparisons, notably CMIP.
- Data sets accessible on the ESGF alongside CMIP model output, **adhering to the same data conventions**
- Observations to be structured in coordination with the CMIP output (e.g. NetCDF files, CF Convention, common vocabularies)



.... and growing!

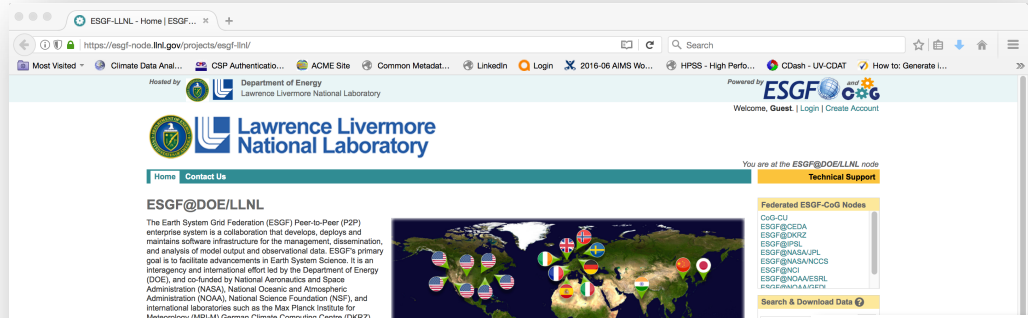
<https://www.earthsystemcog.org/projects/obs4mips/>





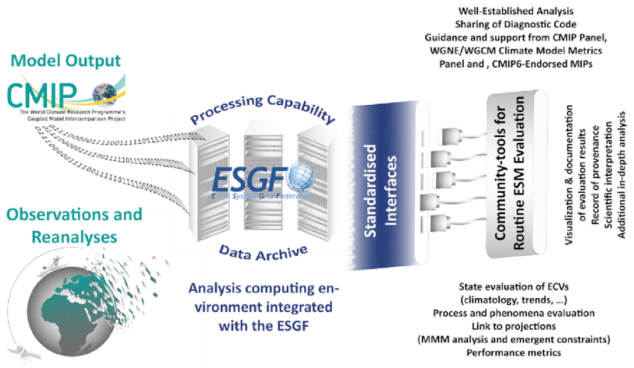
# Earth System Grid Federation

The ESGF peer-to-peer (P2P) enterprise system is one of the largest-ever collaborative data efforts in Earth system science



Major ESGF Node Sites

Institution	Gateway URL	Version	Country	Project(s)	Contact
CEDA	esgf-index1.ceda.ac.uk	2.4.0	U.K.	CMIP5, CORDEX, Obs4MIPs, SPES, ESA CCL, EUCLERA, CLIPC	alan.lwi@stfc.ac.uk
DKRZ	esgf-data.dkrz.de	2.4.0	Germany	CMIP5, CORDEX, Obs4MIPs, ISI-MIP	berger@dkrz.de
ANU NCI	esgf.nci.org.au	2.4.0	Australia	CMIP5	ben.evans@anu.edu.au
NOAA GFDL	esgfdata.gfdl.noaa.gov	2.4.0	U.S.	CMIP5, nccp2013, Obs4MIPs	hans.vahlenkamp@noaa.gov
NASA GSFC	esgf.nccs.nasa.gov	2.4.0	U.S.	CMIP5, Obs4MIPs, Ana4MIPs, NEX-GDDP, NEX-DCP30, CREATE-IP	daniel.q.duffy@nasa.gov
IPSL	esgf-node.ipsl.upmc.fr	2.4.0	France	CMIP5, CORDEX, Obs4MIPs	sebastien.denvil@ipsl.jussieu.fr
NASA JPL	esgf-node.jpl.nasa.gov	2.4.0	U.S.	Obs4MIPs, GAS5-YoTC, CMAC	luca.cinquini@jpl.nasa.gov
DOE LLNL	esgf-node.llnl.gov	2.4.0	U.S.	CMIP5, CMIP3, input4MIPs, ACME	sasha@llnl.gov
LIU	esgf-dn1.nsc.liu.se	2.4.0	Sweden	CMIP5, CORDEX, Obs4MIPs	pchengi@nsc.liu.se



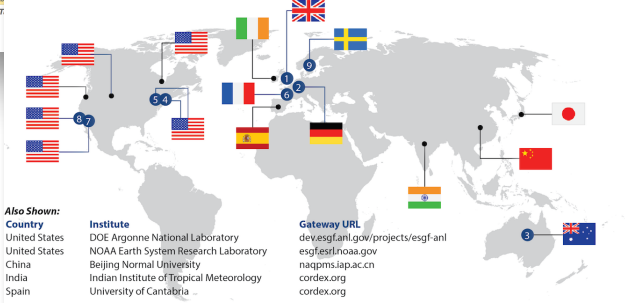
across agency and political boundaries collaborate with other software designed to meet the objectives of ESGF: e.g., software developed by NASA, NOAA, ESI, and the European ES-INES structure and tools that facilitate scientific advancements

WCRP CMIP5 World Climate Research Programme

input4MIPs

<https://esgf-node.llnl.gov/>

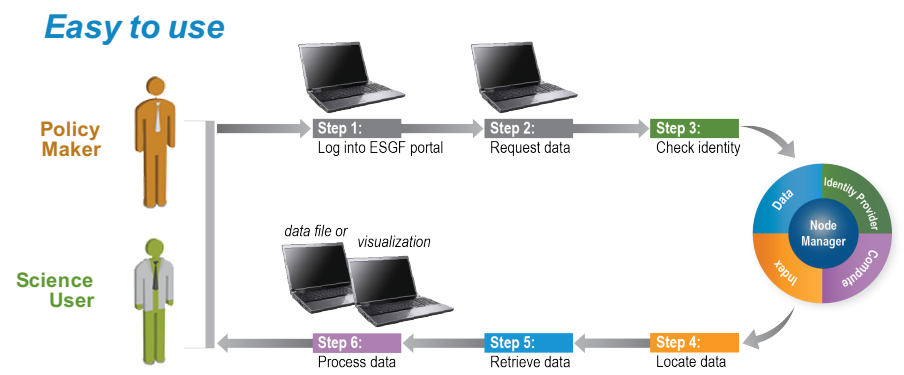
<https://esgf.llnl.gov/>



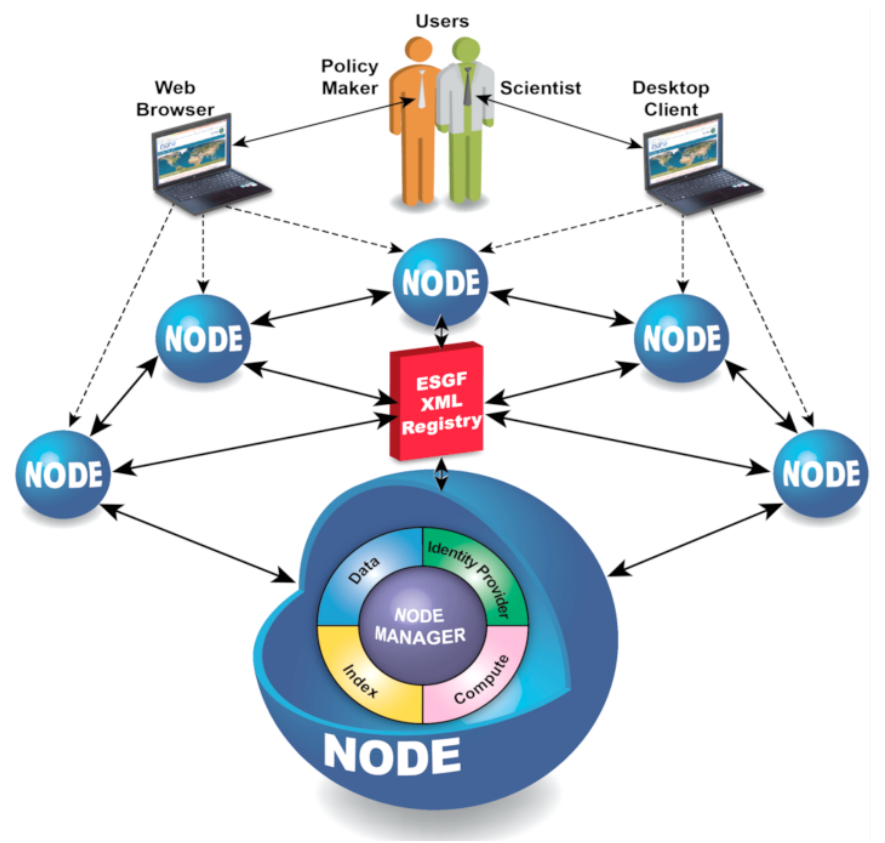


# Unified virtual data cloud

## User's View:



## Actual Structure:






# Distributed archives

## Earth system science supported projects:

- **CMIP3, CMIP5, and CMIP6**—Phases 3, 5, and 6 of the Coupled Model Intercomparison Project
- **CORDEX**—Coordinated Regional climate Downscaling Experiment
- **ACME**—Accelerated Climate Modeling for Energy
- **EUCLEIA**—European Climate and weather Events: Interpretation and Attribution
- **Obs4MIPs**—Observational for Model Intercomparison Project
- **GeoMIP**—Geo-engineering Model Intercomparison Project
- **Ana4MIPs**—Reanalysis for the coupled model intercomparison
- **LUCID**—Land-Use and Climate, Identification of robust impacts
- **ESA CCI**—European Space Agency Climate Change Initiative Earth Observation data
- **TAMIP**—Transpose-Atmospheric Model Intercomparison Project
- **PMIP**—Paleoclimate Modeling Intercomparison Project
- **HIWPP**—High Impact Weather Prediction Project
- **NEMS**—NOAA Environmental Modeling System
- **NEX-DCP30**—NASA NEX Downscaled Climate Projections
- **CREATE-IP**—Collaborative REAnalysis Technical Environment – Intercomparison Project



**4,635.828 TB**  
ESGF total data volume



**4,261.921 TB**  
CMIP5 total data volume



**59.813 TB**  
CORDEX total data volume

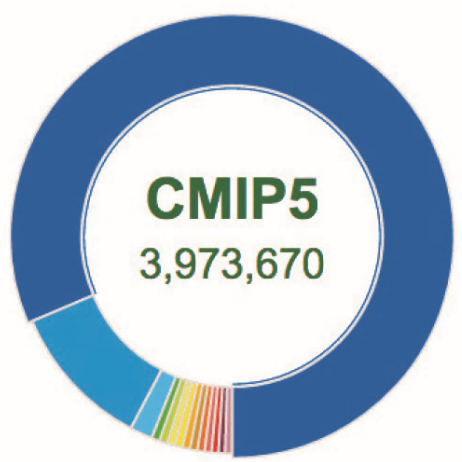
*More complete list:*

<https://esgf-node.llnl.gov/ac/list/>



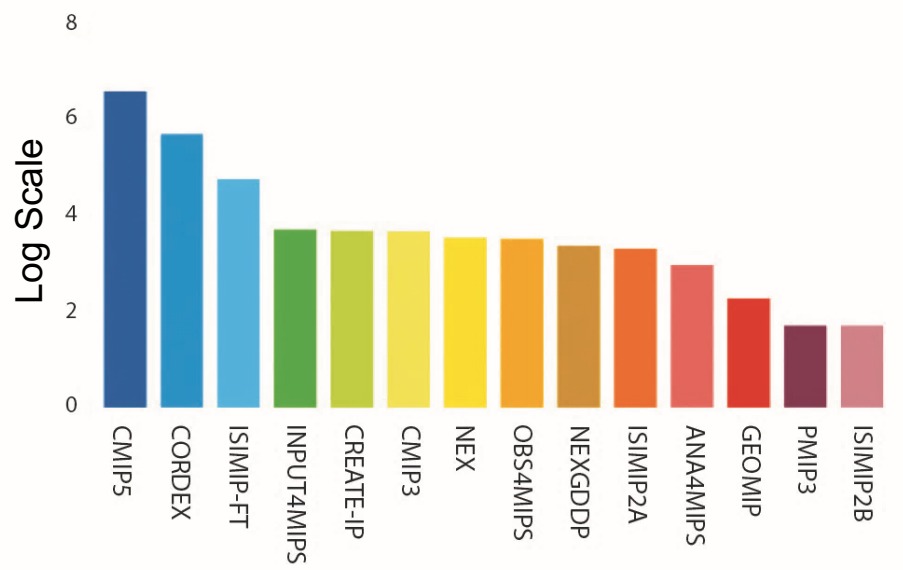
# Distributed archives

## Number of Downloads Per Project



- CMIP5 (3,973,670)
- CORDEX (510,588)
- ISIMIP-FT (58,534)
- INPUT4MIPs (5,214)
- CREATE-IP (4,890)
- CMIP3 (4,734)
- NEX (3,492)
- OBS4MIPs (3,236)
- NEXGDDP (2,376)
- ISIMIP2A (2,074)
- ANA4MIPs (942)
- GEOMIP (188)
- PMIP3 (52)
- ISIMIP2B (52)

## Number of Downloads Per Project

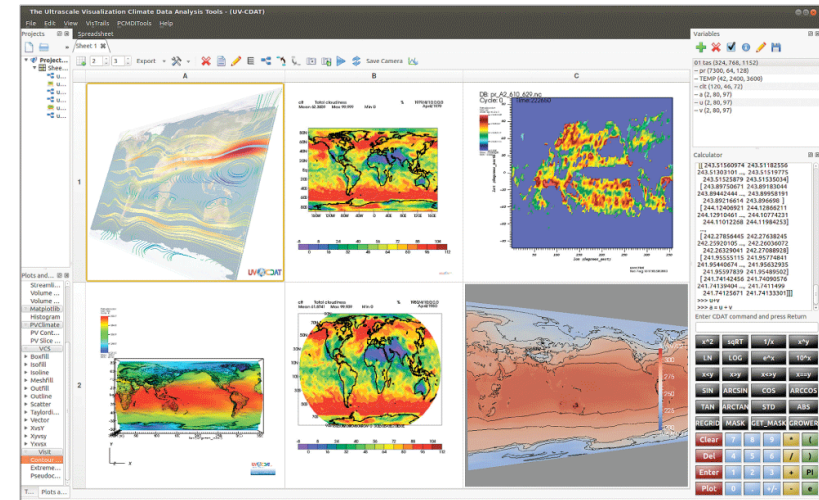
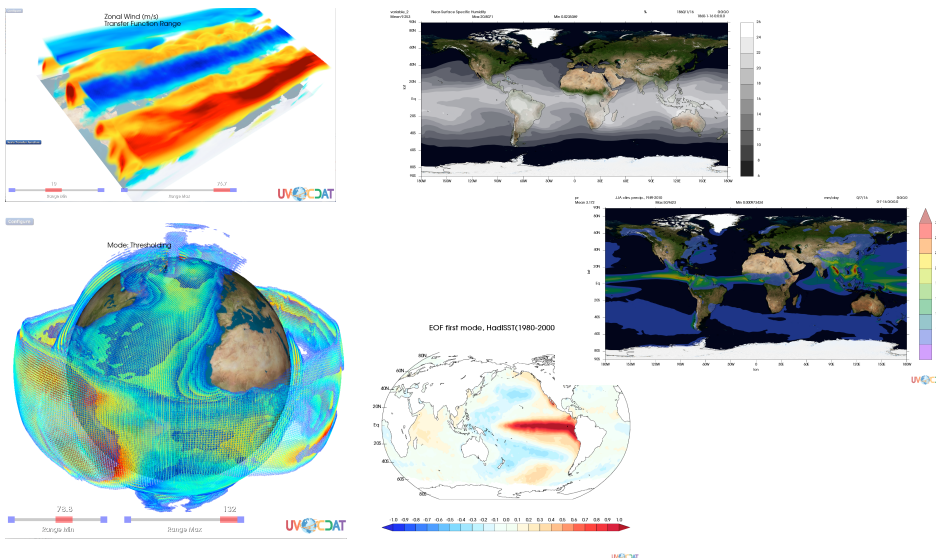


**Total number of downloads: ~ 4,500K**

- Most downloaded projects:**
1. **CMIP5**
  2. **CORDEX**
  - ⋮

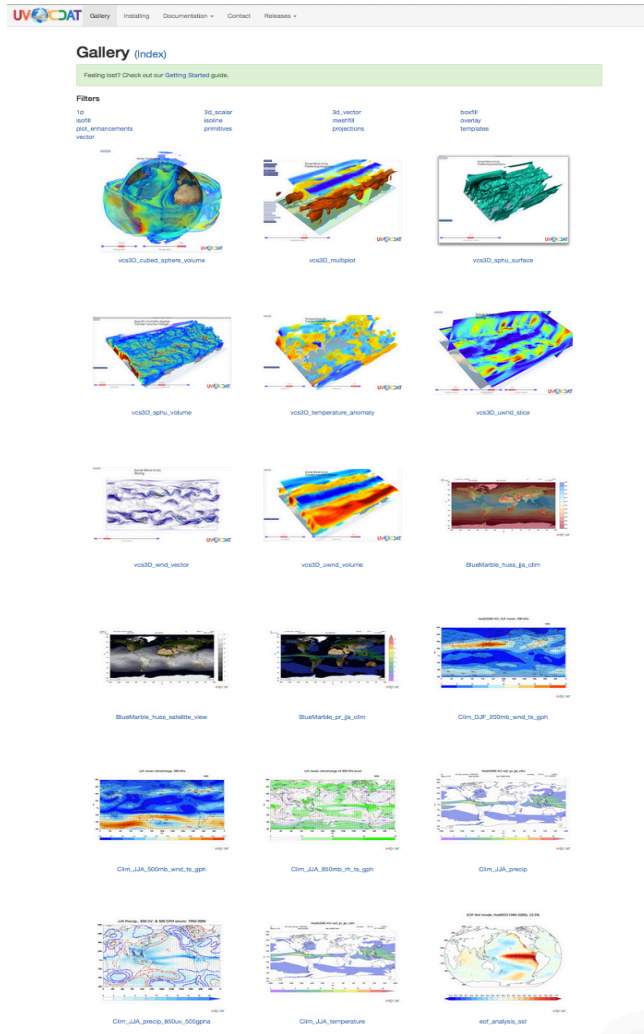
# Community Data Analysis Tools

- Python based application (GUI / Command Line)
- CDMS (Climate Data Management System)
  - Handle multi-dimensional climate variables
  - cdutil, genutil: Python library specialized for climate analysis
- VCS Graphics Module



<https://uvcdat.llnl.gov/>

# analysis examples (1)



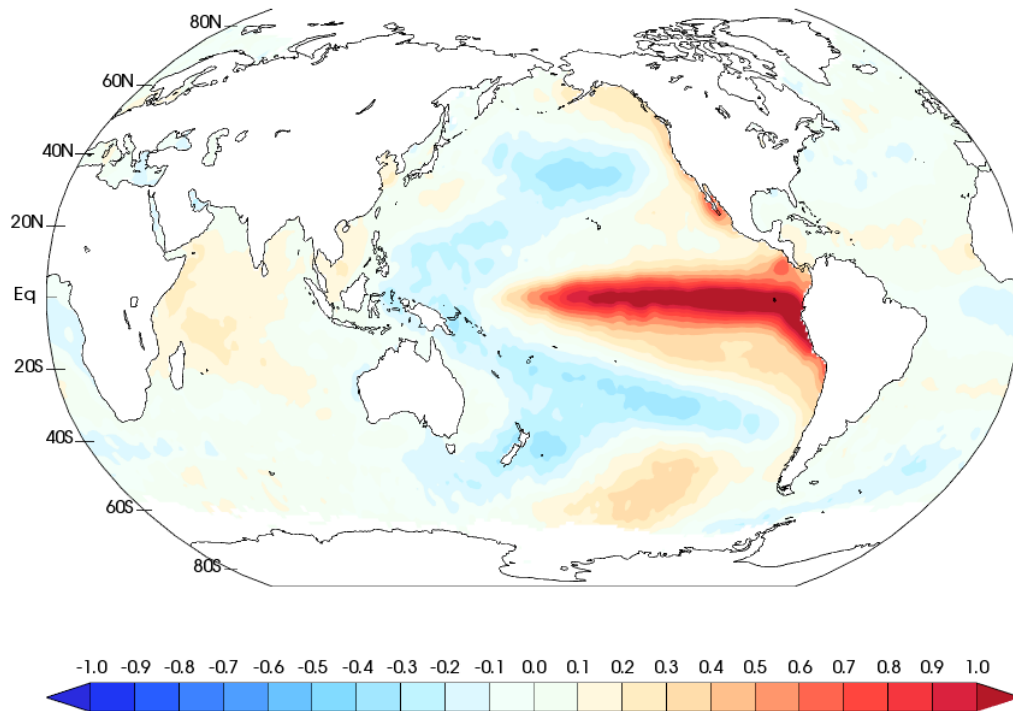
## Examples of frequently repeated analyses:

- Seasonal climatology / anomaly
- EOF analysis
- Running average
- Hovmoller & Taylor Diagram
- Multi-model ensemble
- Etc

<https://uvcdat.llnl.gov/gallery.html>

# analysis examples (2)

EOF first mode, HadISST(1980-2000), 23.5%



## Highlight of the example Python code:

```
import cdms2
from eofs.cdms import Eof

#Open file
:

# Load variable
d = ...
:

# EOF (take leading mode)
solver = Eof(d, weights='area')
eof = solver.eofsAsCovariance(neofs=1)

# Plot
:
```

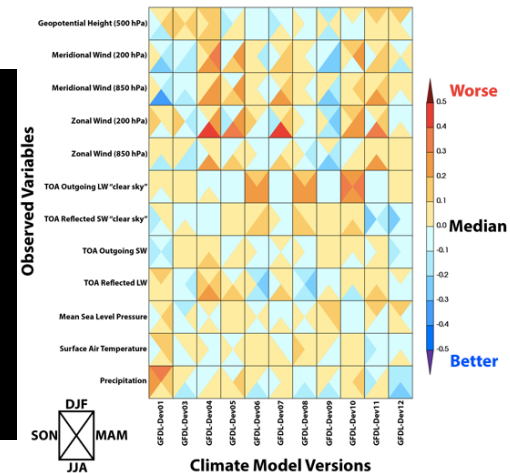
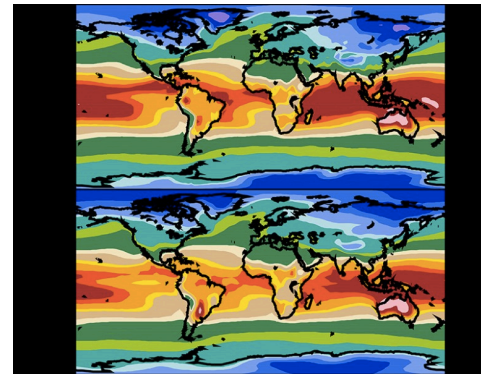
[https://uvcdat.llnl.gov/examples/eof\\_analysis\\_sst.html](https://uvcdat.llnl.gov/examples/eof_analysis_sst.html)



# PCMDI Metrics Package (PMP)

Program for Climate Model Diagnosis  
and Intercomparison

- Emphasizes a diverse suite of relatively robust high level summary statistics objectively comparing models and observations across space and time scales
- End-to-end provenance to ensure reproducibility
- **Open source Python** publicly available on **GitHub**
- Designed to enable the research community to contribute
- Currently collaborating with 5 modeling groups

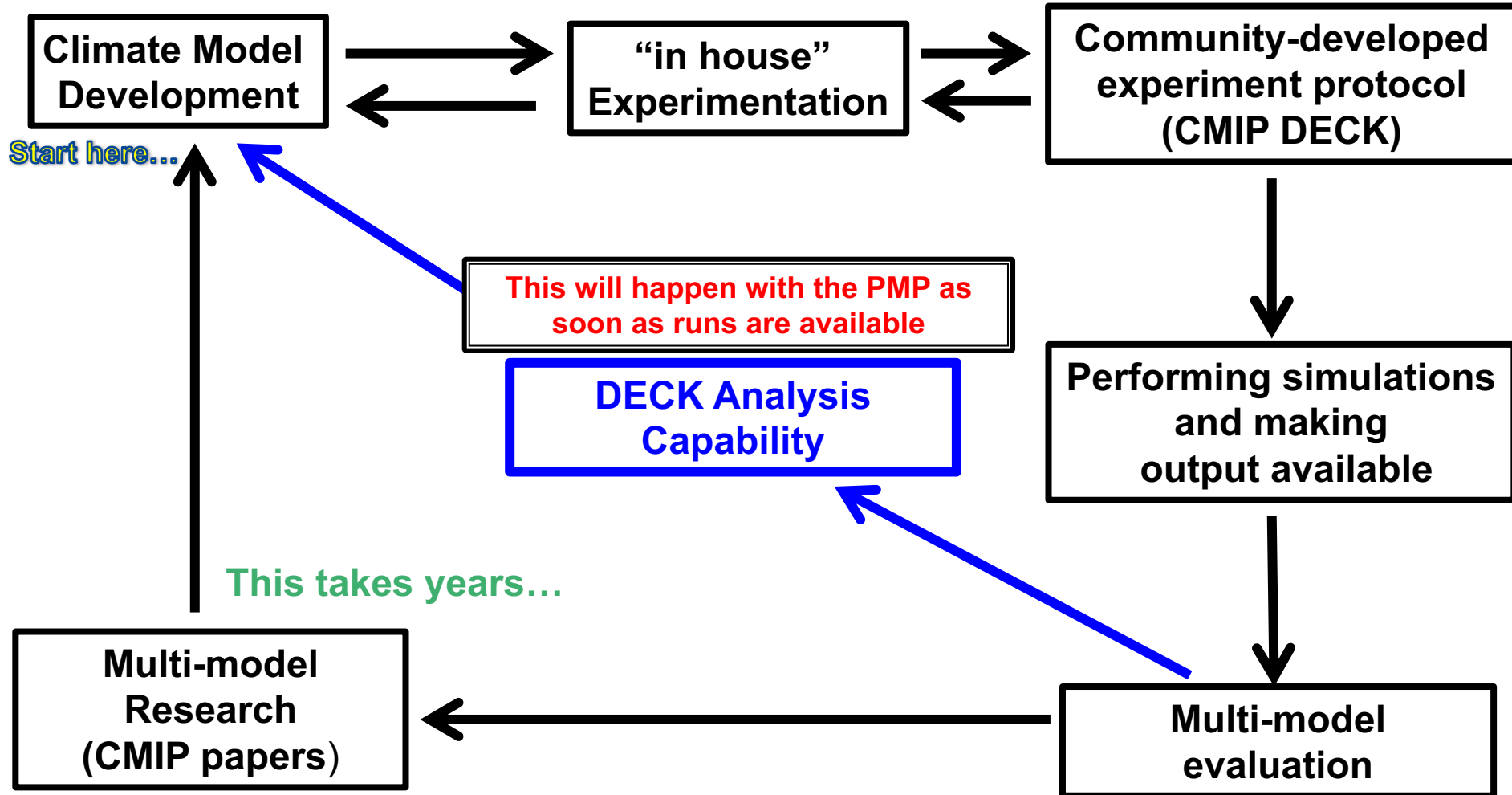


Gleckler, P. J., C. Doutriaux, P. J. Durack, K. E. Taylor, Y. Zhang, D. N. Williams, E. Mason, and J. Servonnat (2016), A more powerful reality test for climate models, *Eos*, 97, doi:10.1029/2016EO051663. Published on 3 May 2016.



## How can CMIP more directly assist model development?

**Enable modelers to leverage diverse analysis community expertise**

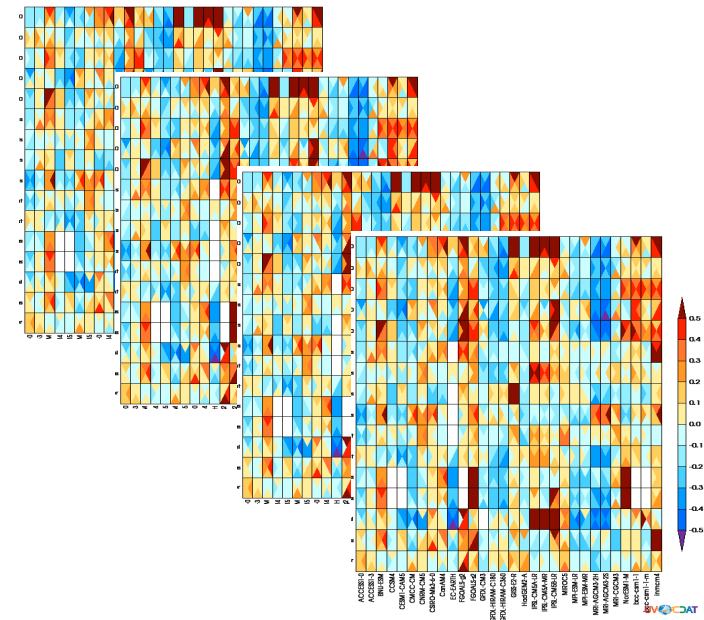
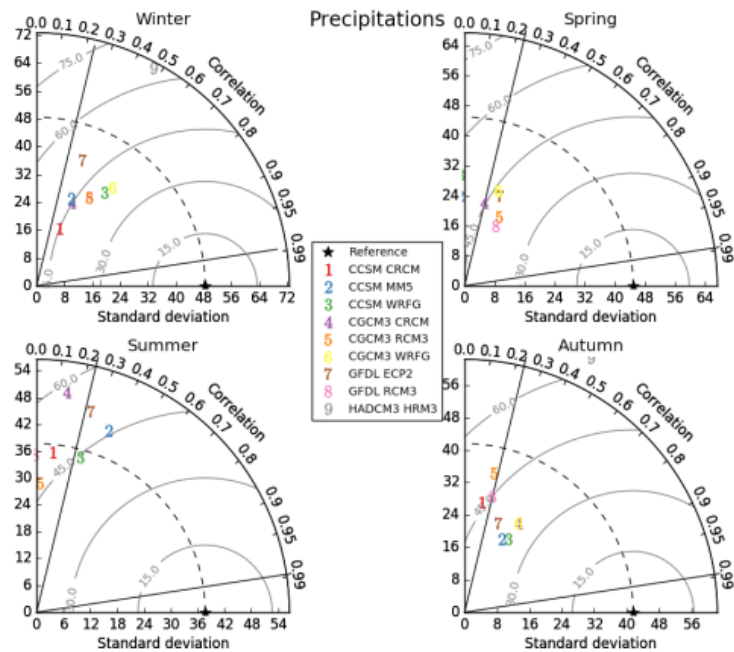




Program for Climate Model Diagnosis  
and Intercomparison

# The PCMDI Metrics Package (v1.1x) Prototyped on climatological summaries

## Taylor Diagrams and Portrait Plots: Orthogonal decompositions of large scale climatological error statistics



# PART 2. Evaluation of Various Climate Models

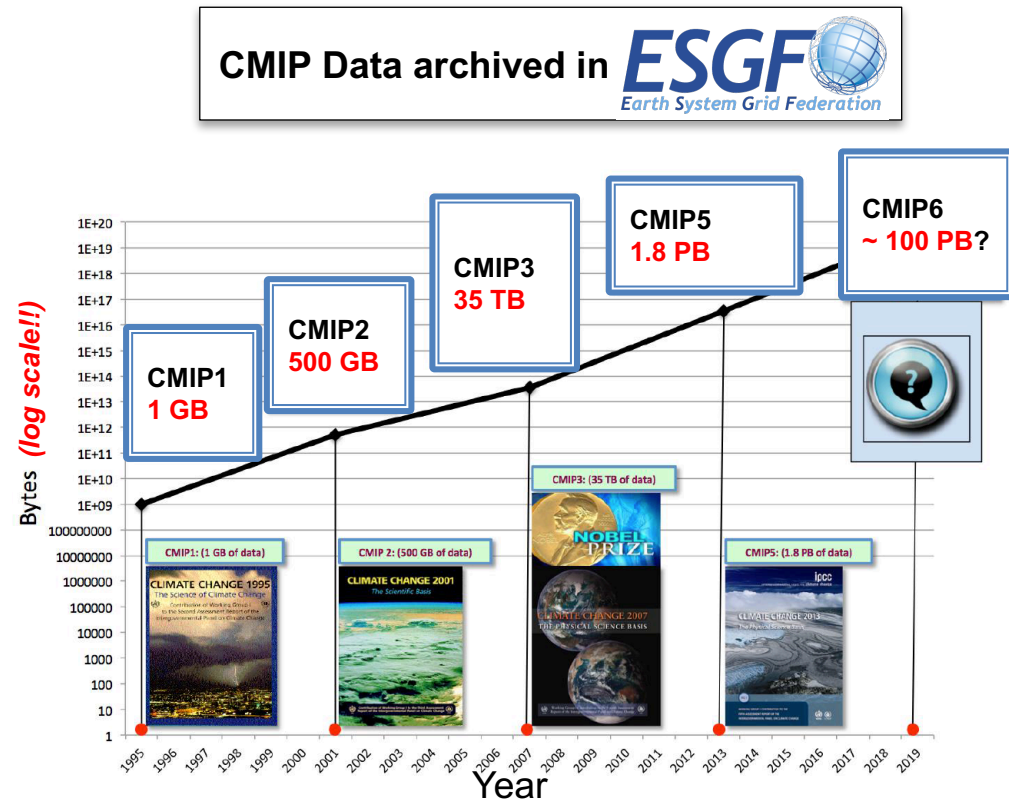
*(next 15 slides...)*

- *Scientific study based on the infrastructure, and currently being implemented to PMP*
- *Specific research title: **Quantifying the Agreement Between Observed and Simulated Extratropical Modes of Interannual Variability***

*Acknowledgements to: Kenneth Sperber, Peter Gleckler, Celine Bonfils, and Karl Taylor*

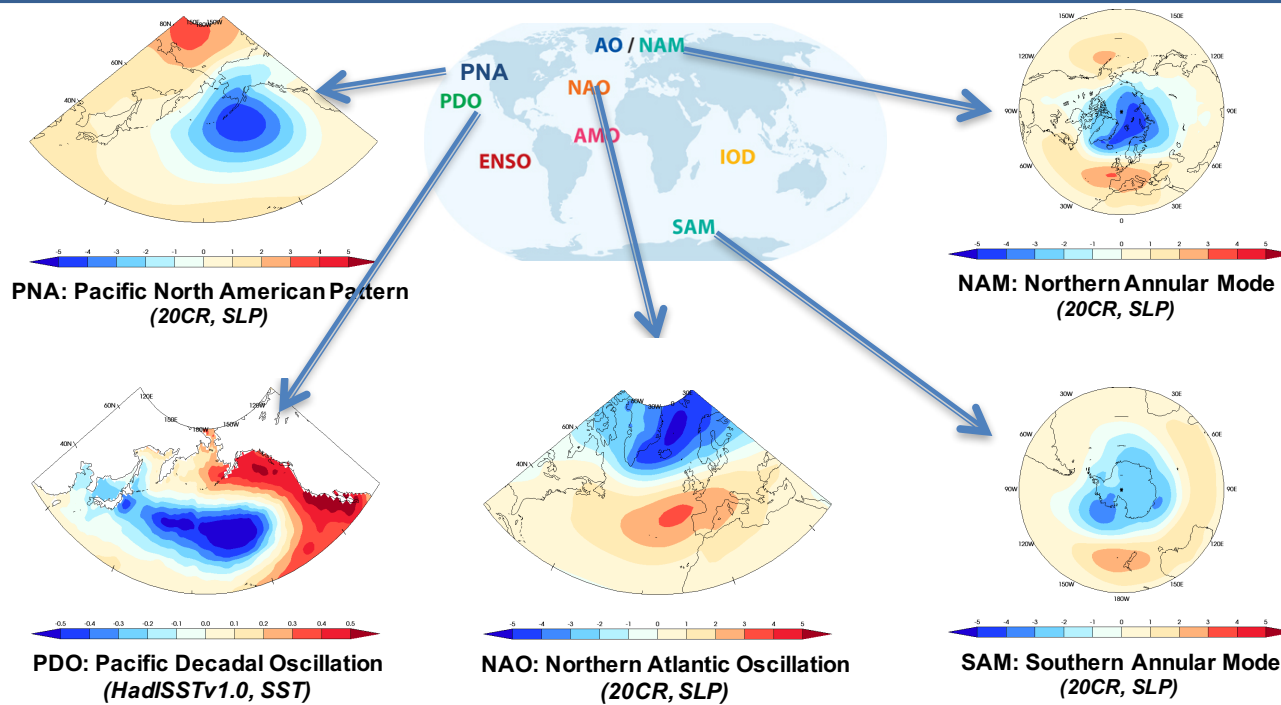
# Objective: Routine Evaluation of Models

- **CMIP** has collected a massive volume of climate model output
- To have a better understanding of model *uncertainties and performance*, it is important to *evaluate models routinely in a systematic and collective way*
- PCMDI/LLNL is developing a metrics package (the PMP) to more directly contribute to model development (via quick feedback)



(Williams et al. 2015, BAMS)

# Background: Modes of Variability



- Generally defined by the **leading EOF mode** in observations
- Represent **long-term large-scale** variance
- Important test for diagnosing model behavior, and detection & attribution

# Datasets

## ■ CMIP5 Models

- Models: **180 historical simulations (45 models with their realizations)**
- Time window: 1900-2005 (except SAM: 1956-2005)
- Variables:
  - 1) **Sea level pressure (SLP)**: Seasonal anomalies
  - 2) **Sea surface temperature (SST)**: Monthly anomalies (for PDO)
- Area-weighted average over EOF domain was removed at each time step



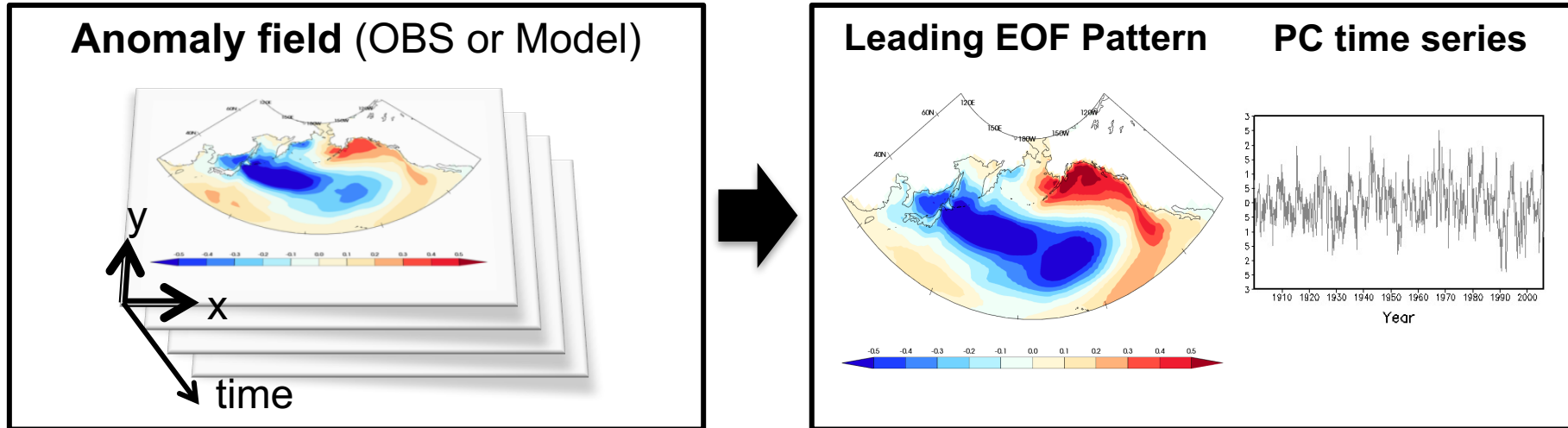
## ■ Observation based reference datasets

	SLP	SST
<b>OBS 1 (default)</b>	NOAA CIRES 20CR	HadISST v1.1
<b>OBS 2</b>	ERA 20C	HadISST v2.1
<b>OBS 3</b>	HadSLP	ERSST v3.0

## ■ Tools

- **PCMDI Metrics Package (PMP)**: Python based open-source tool (*Gleckler et al. 2016, EOS*)
- **UV-CDAT**: Python based large-scale data analysis and visualization tool (*Williams 2014, EOS*)
- **eofs**: Python library for EOF analysis (*Dawson 2016*)

# Conventional EOF Approach

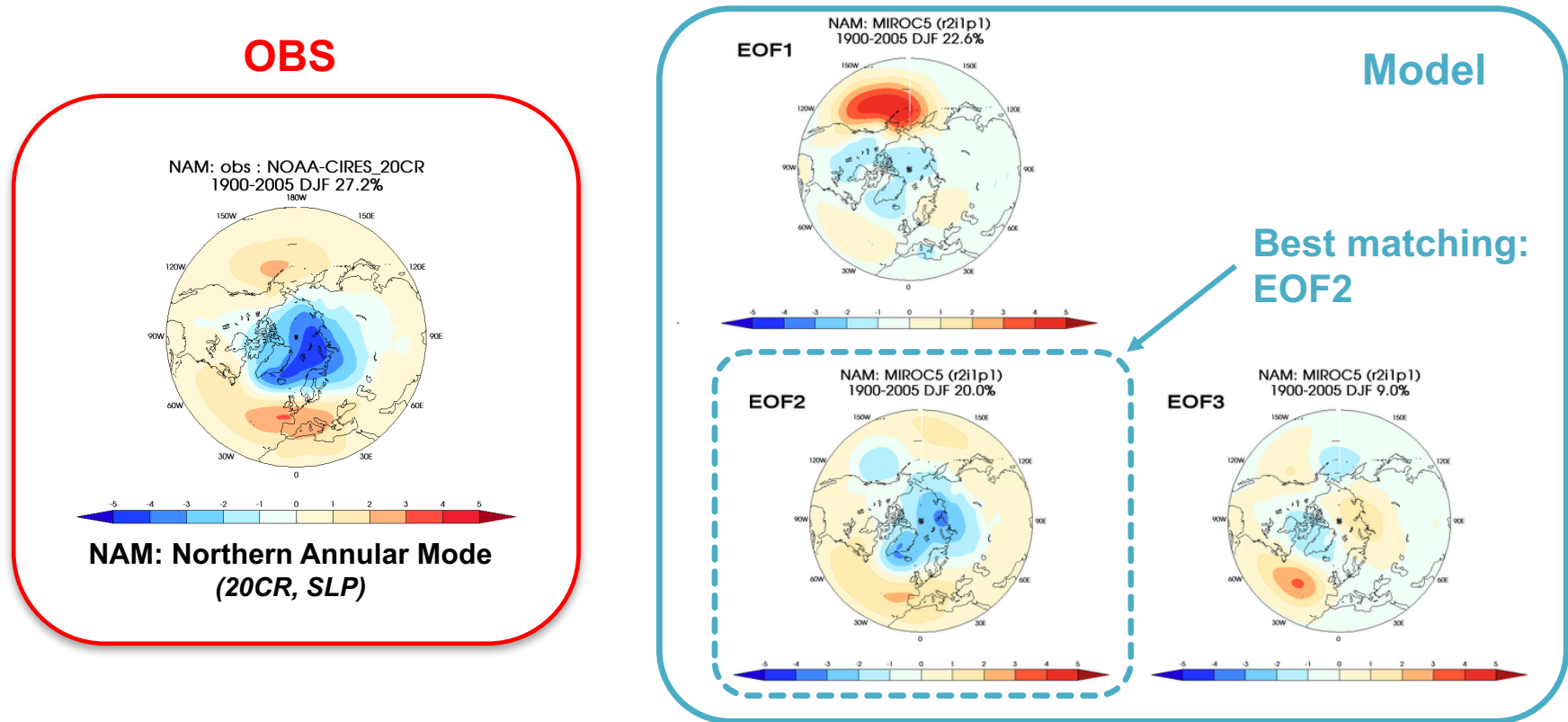


## Limitations:

1. **Sign ambiguity:** EOF sign is arbitrary, sometimes it needs to be flipped
2. **EOF mode swapping:** Cases in which leading OBS EOF better corresponds to 2<sup>nd</sup> or 3<sup>rd</sup> mode of model

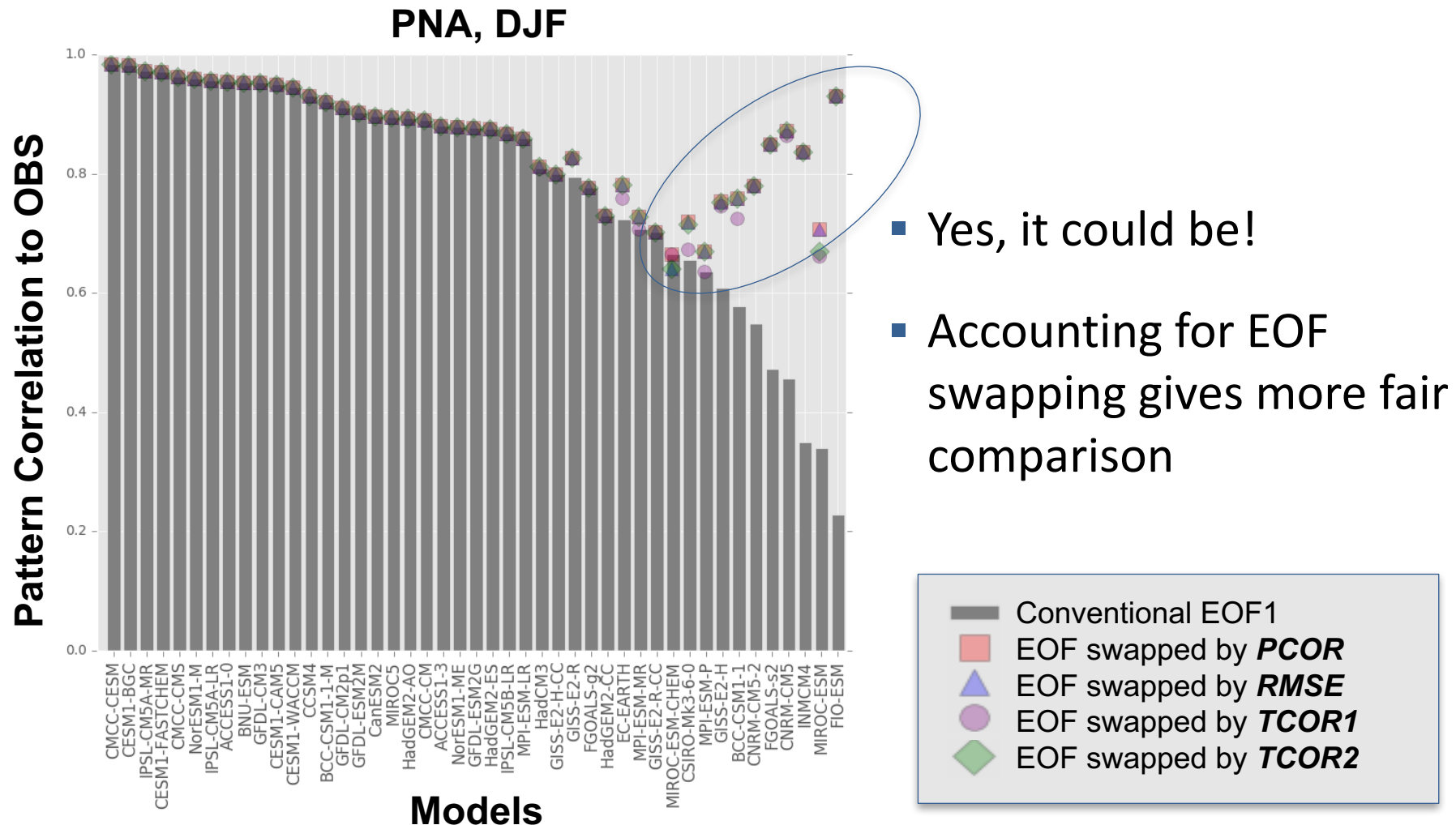
# EOF mode swapping (1): Example

- NAM (DJF) simulated by MIROC5 (r2i1p1)



The **leading OBS EOF** corresponds best to **model's 2<sup>nd</sup> EOF mode**

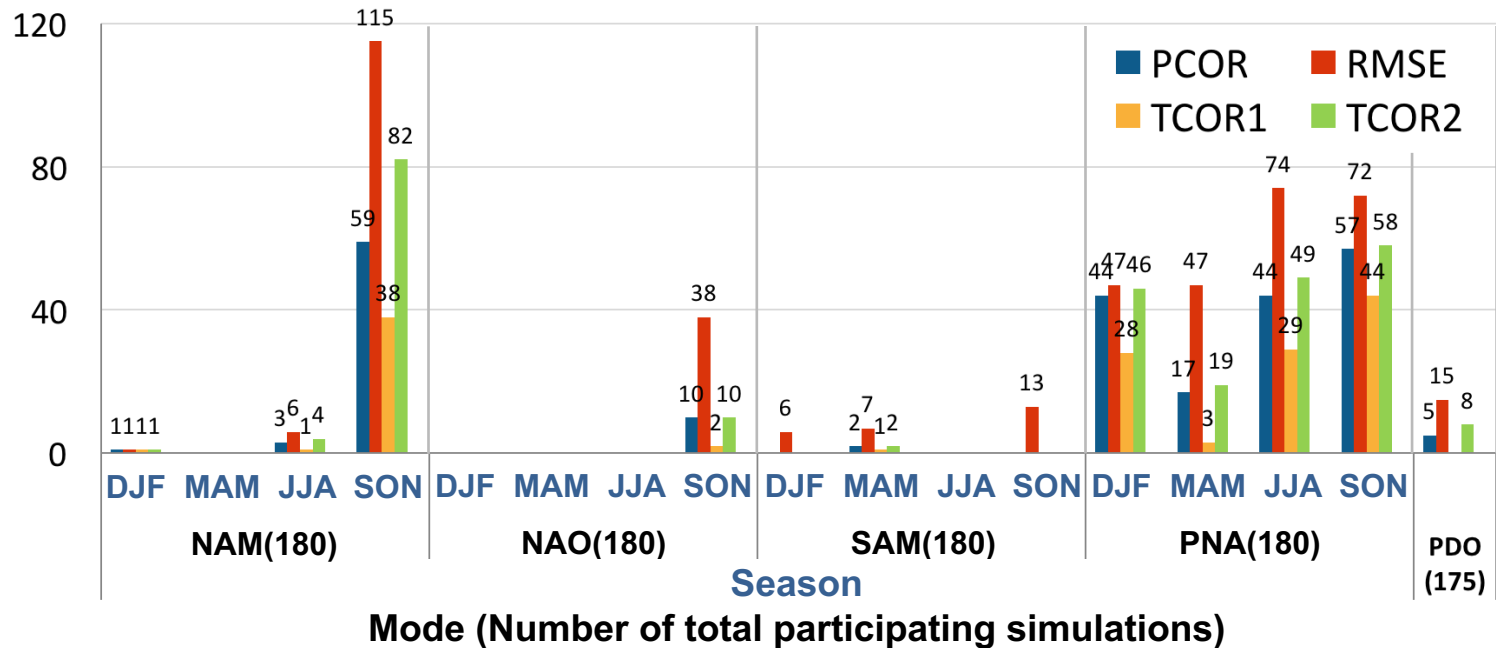
# EOF mode swapping (2): Is it significant?



- Yes, it could be!
- Accounting for EOF swapping gives more fair comparison

# EOF mode swapping (3): How often it happen?

Number of EOF swapping cases identified



Applied criteria to decide best matching EOF:

**Spatial:**

- **PCOR:** Pattern correlation
- **RMSE:** Root mean square error

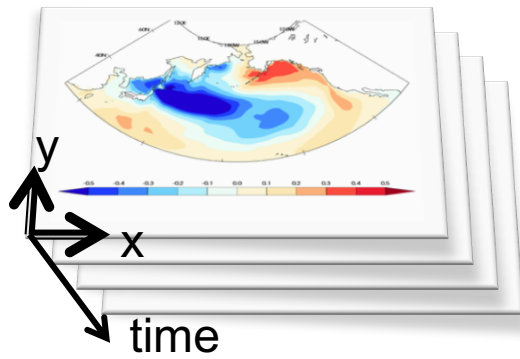
**Temporal:**

- **TCOR1:** CBF PC vs EOF PCs
- **TCOR2:** OBS PC vs tweaked CBF PCs

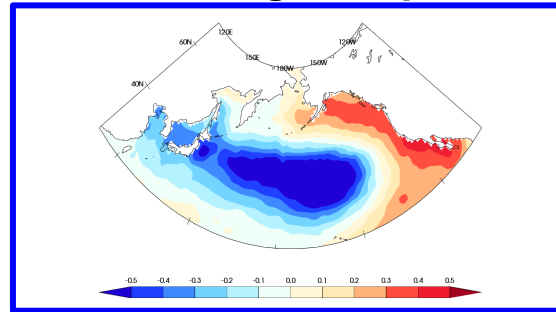
- Significant number of EOF swapping cases are identified
- There is no single best criteria

# Common Basis Function (CBF) Approach

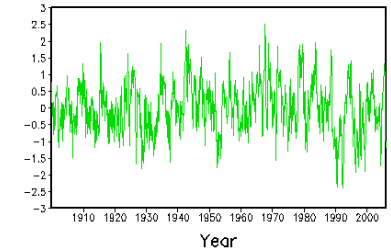
## Model's anomaly field



## OBS leading EOF pattern



## Model's CBF PC time series

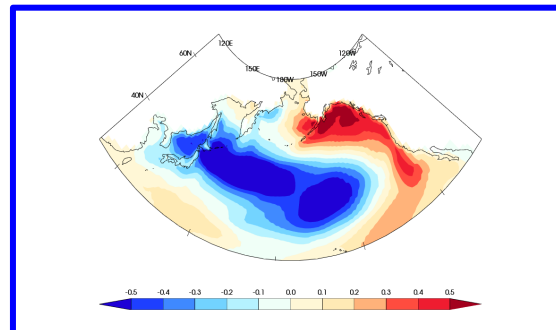


Get PC time series by projecting OBS pattern into model's anomaly space

(1) Projection

(2) Linear Regression

## Model's CBF Pattern



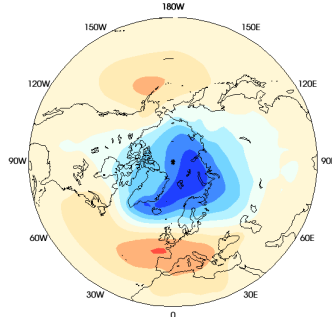
Reconstruct Pattern by linear regression between CBF PC & model field

# CBF Result: Example

## ■ NAM (DJF) simulated by MIROC5 (r2i1p1)

### OBS

NAM: obs : NOAA-CIRES\_20CR  
1900-2005 DJF 27.2%

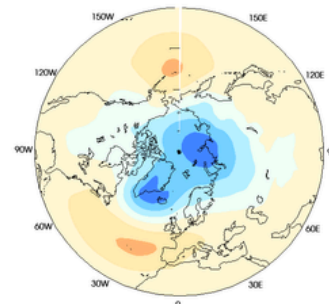


**NAM: Northern Annular Mode  
(20CR, SLP)**

### Model

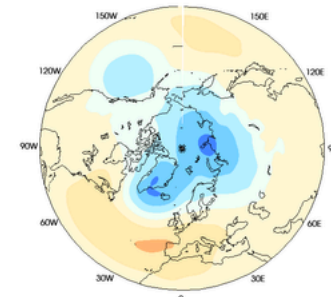
#### CBF

NAM: MIROC5 (r2i1p1) - pseudo  
1900-2005 DJF 20.1%

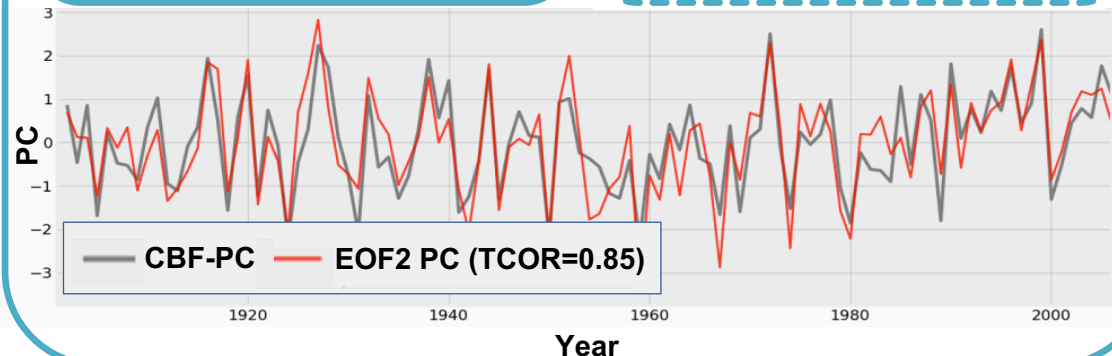


#### Best matching: EOF2

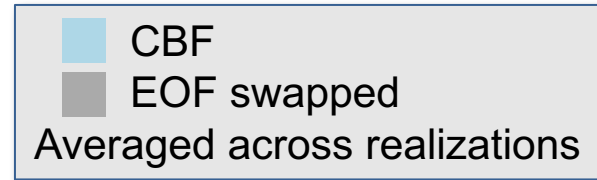
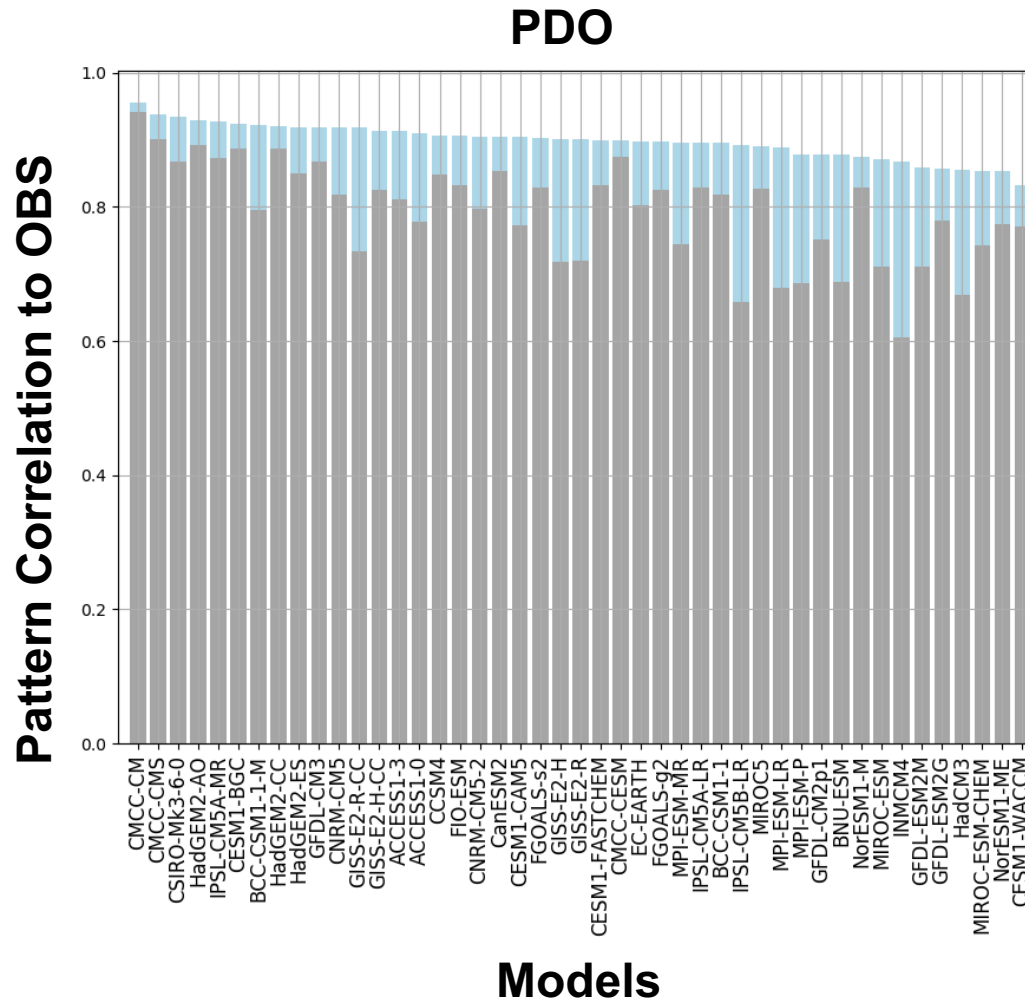
NAM: MIROC5 (r2i1p1)  
1900-2005 DJF 20.0%



- CBF pattern better corresponds to OBS than the best-matching EOF mode
- CBF PC corresponds to PC of best-matching EOF mode

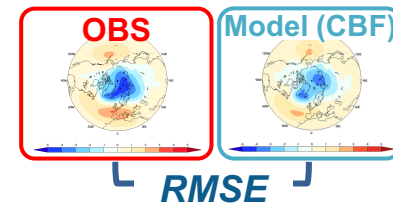


# CBF vs. EOF swapping

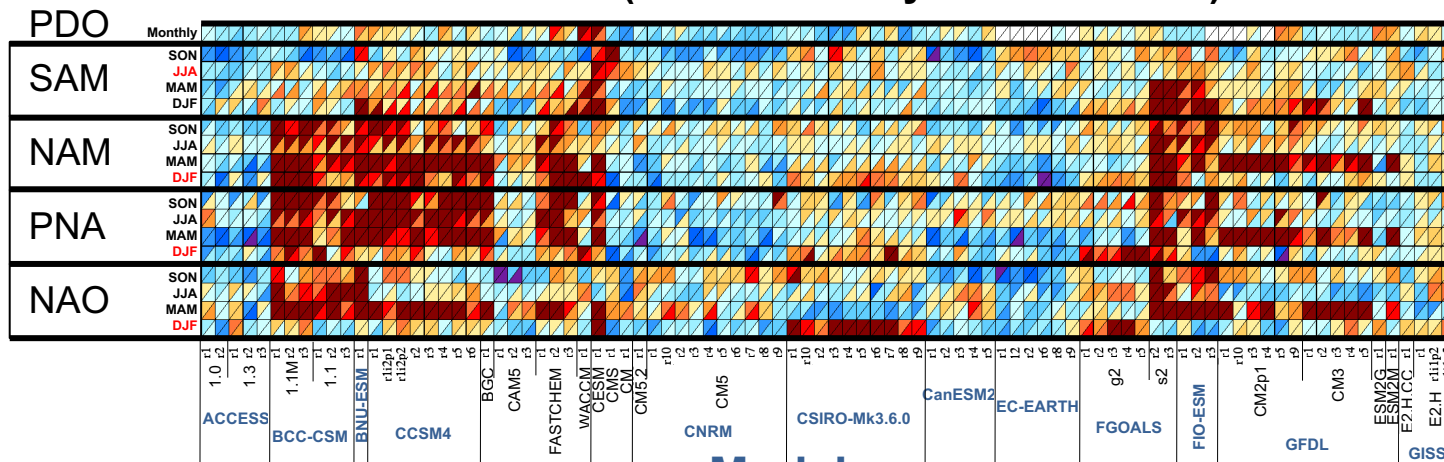


- The CBF provides a **more fair comparison with observations** than the conventional EOF
- The CBF reveals that **model skill is better** than indicated by conventional EOF analysis

# Performance Diagnostics (1): Pattern + Amplitude

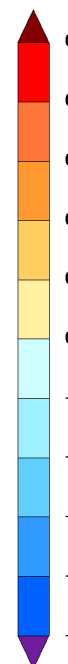


## RMSE (Normalized by median value)



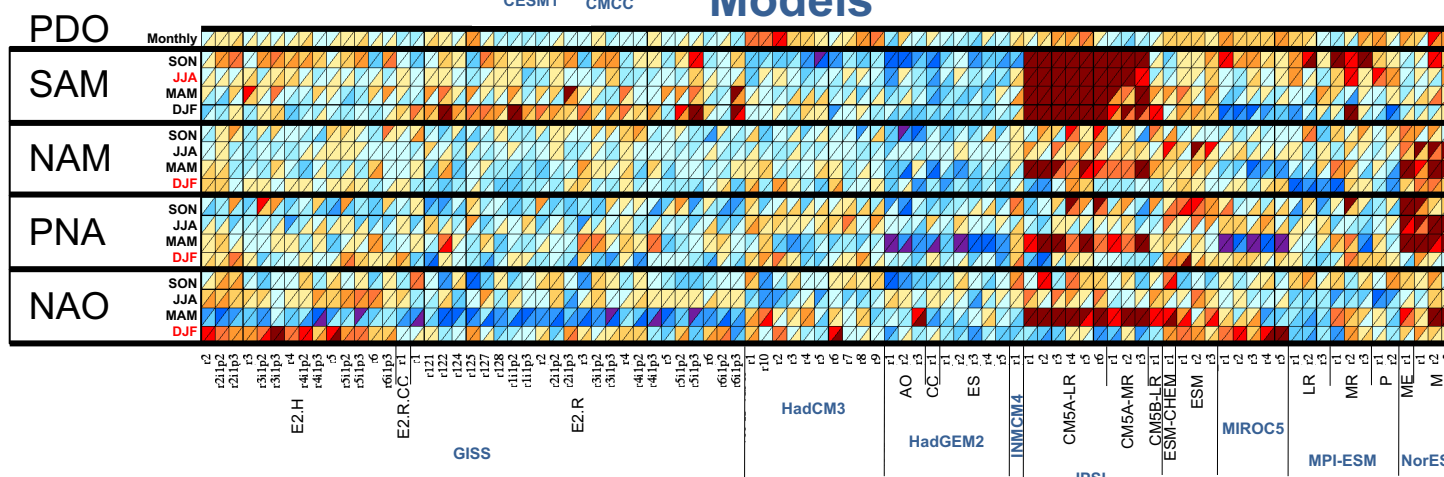
OBS1  
20CR  
OBS2  
ERA20C

Relatively  
Worse



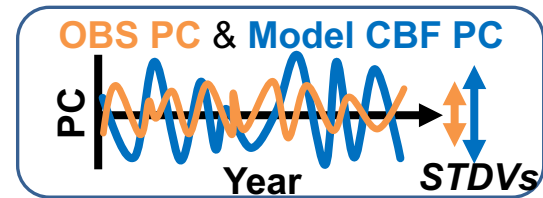
Median

Relatively  
Better



## More Models

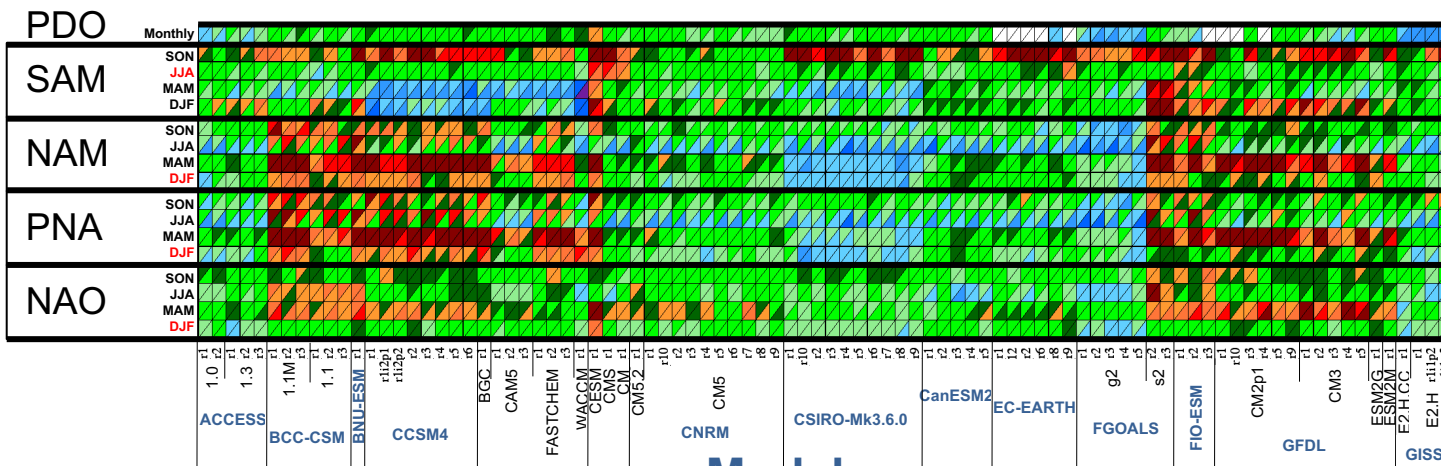
# Performance Diagnostics (2): Amplitude



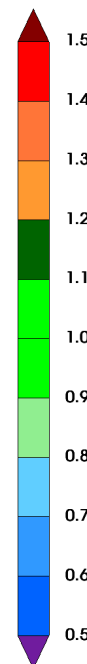
### Ratio of PC STDVs: Model/OBS

OBS1  
20CR

OBS2  
ERA20C

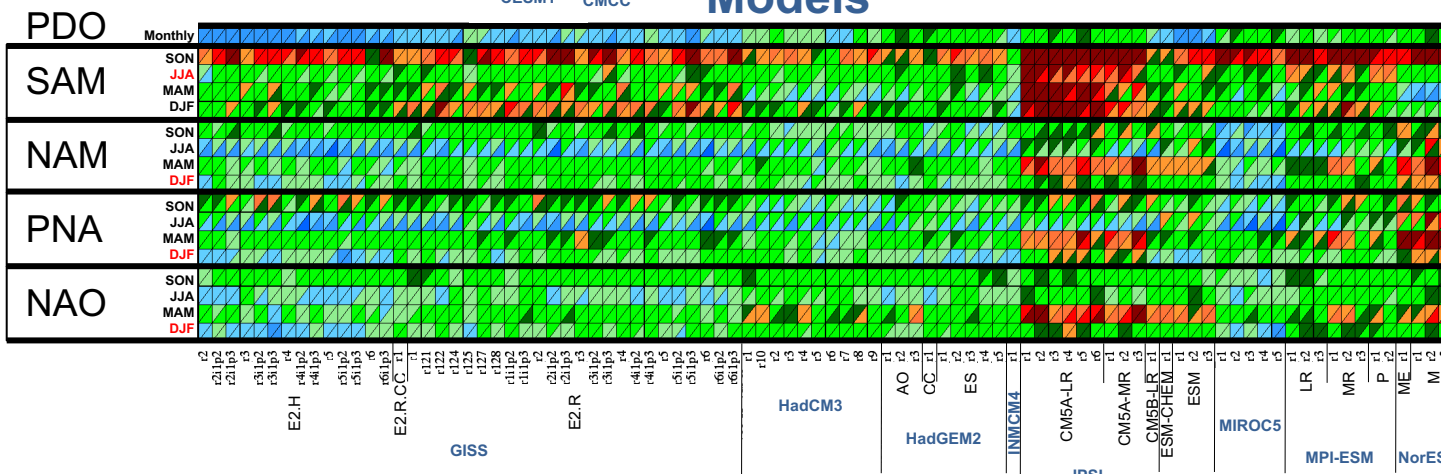


Overestimate

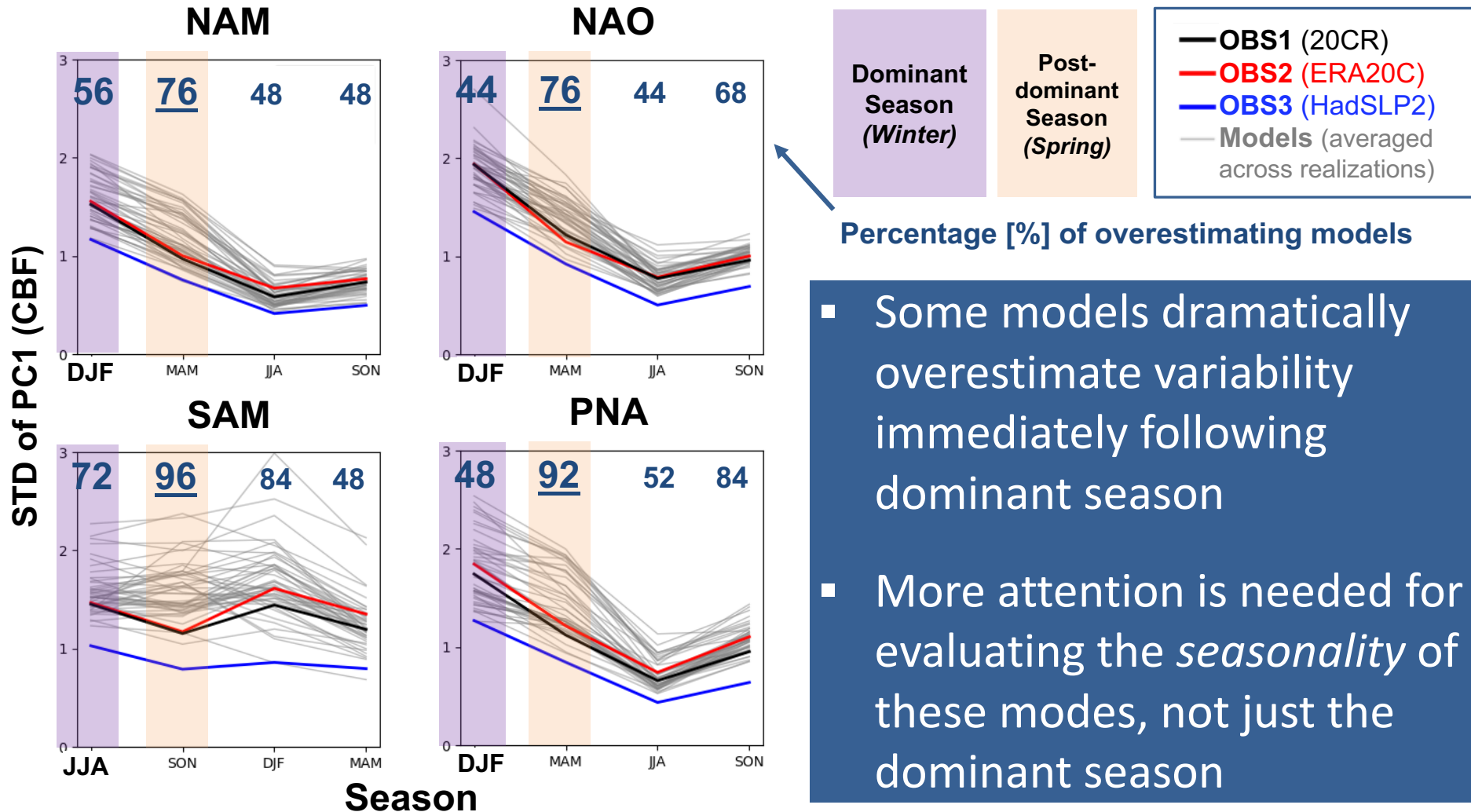


Good

Underestimate

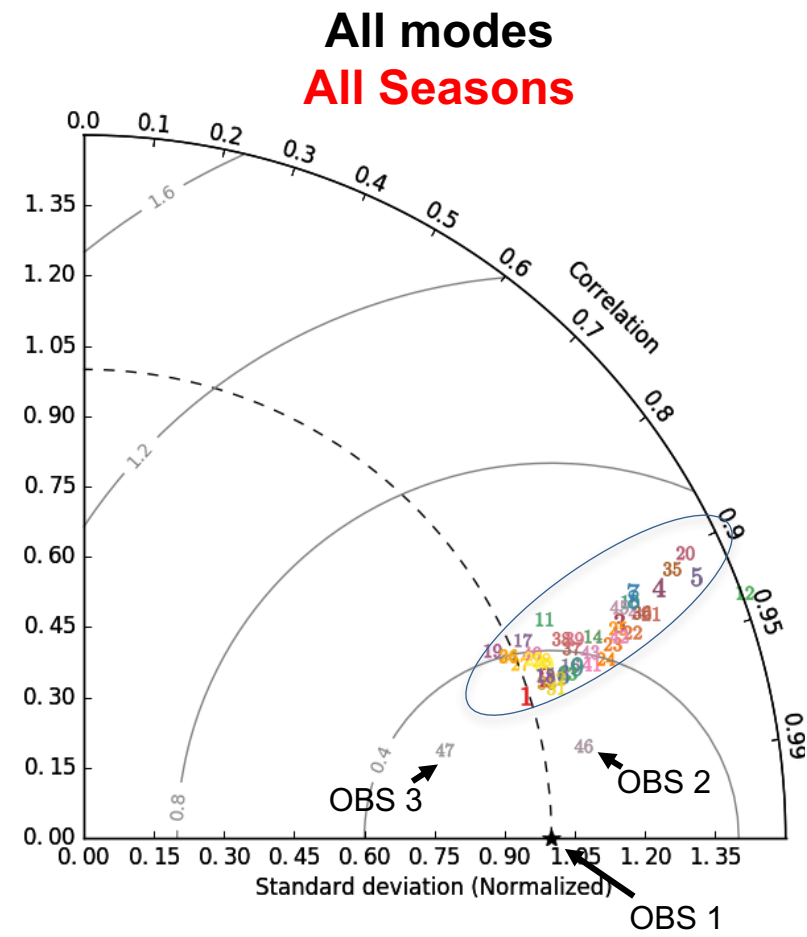
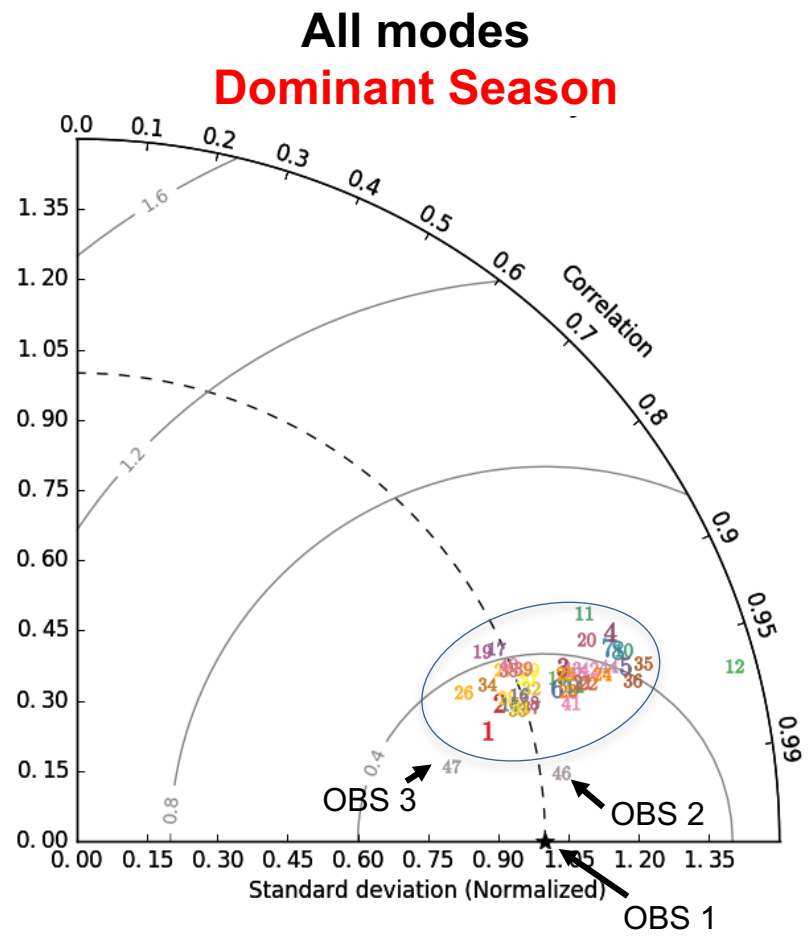


# Seasonal Variation



- Some models dramatically overestimate variability immediately following dominant season
- More attention is needed for evaluating the *seasonality* of these modes, not just the dominant season

# Taylor Diagrams



- ★ OBS 1
- 1 ACCESS1-0
- 2 ACCESS1-3
- 3 BCC-CSM1-1
- 4 BCC-CSM1-1-M
- 5 BNU-ESM
- 6 CanESM2
- 7 CCSM4
- 8 CESM1-BGC
- 9 CESM1-CAM5
- 10 CESM1-FASTCHEM
- 11 CESM1-WACCM
- 12 CMCC-CESM
- 13 CMCC-CM
- 14 CMCC-CMS
- 15 CNRM-CM5
- 16 CNRM-CM5-2
- 17 CSIRO-Mk3-6-0
- 18 EC-EARTH
- 19 FGOALS-g2
- 20 FGOALS-s2
- 21 FIO-ESM
- 22 GFDL-CM2p1
- 23 GFDL-CM3
- 24 GFDL-ESM2G
- 25 GFDL-ESM2M
- 26 GISS-E2-H
- 27 GISS-E2-H-CC
- 28 GISS-E2-R
- 29 GISS-E2-R-CC
- 30 HadCM3
- 31 HadGEM2-AO
- 32 HadGEM2-CC
- 33 HadGEM2-ES
- 34 INMCM4
- 35 IPSL-CM5A-LR
- 36 IPSL-CM5A-MR
- 37 IPSL-CM5B-LR
- 38 MIROC-ESM
- 39 MIROC-ESM-CHEM
- 40 MIROC5
- 41 MPI-ESM-LR
- 42 MPI-ESM-MR
- 43 MPI-ESM-P
- 44 NorESM1-M
- 45 NorESM1-ME
- 46 OBS 2
- 47 OBS 3

Post-dominant season (Spring) is the major contributor to the overestimation

# Summary

*Further detail:*

**J. Lee**, K. R. Sperber, P. J. Gleckler, C. W. Bonfils, and K. E. Taylor (2017)  
Quantifying the Agreement Between Observed and Simulated Extratropical Modes of Interannual Variability. *Climate Dynamics* (in review)

- We have developed **evaluation metrics** for extra-tropical **modes of variability** using **CBF approach** – *projecting model anomalies onto the observational leading EOF*
- The CBF method leads to a more **consistent** approach for evaluation, overcoming limitations of conventional EOF
- The **amplitude error** is the dominant contributor to systematic error in many models
- Models generally agree **better in the dominant season** of each mode, while many models systematically **overestimate** the variability in **post-dominant season**  
→ *More attention needs to be devoted to evaluating the seasonality of modes*

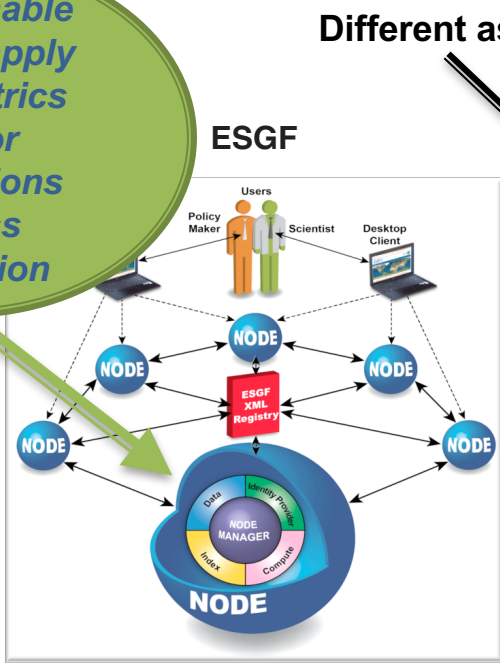
# Further Works

*(only 3 slides left!)*

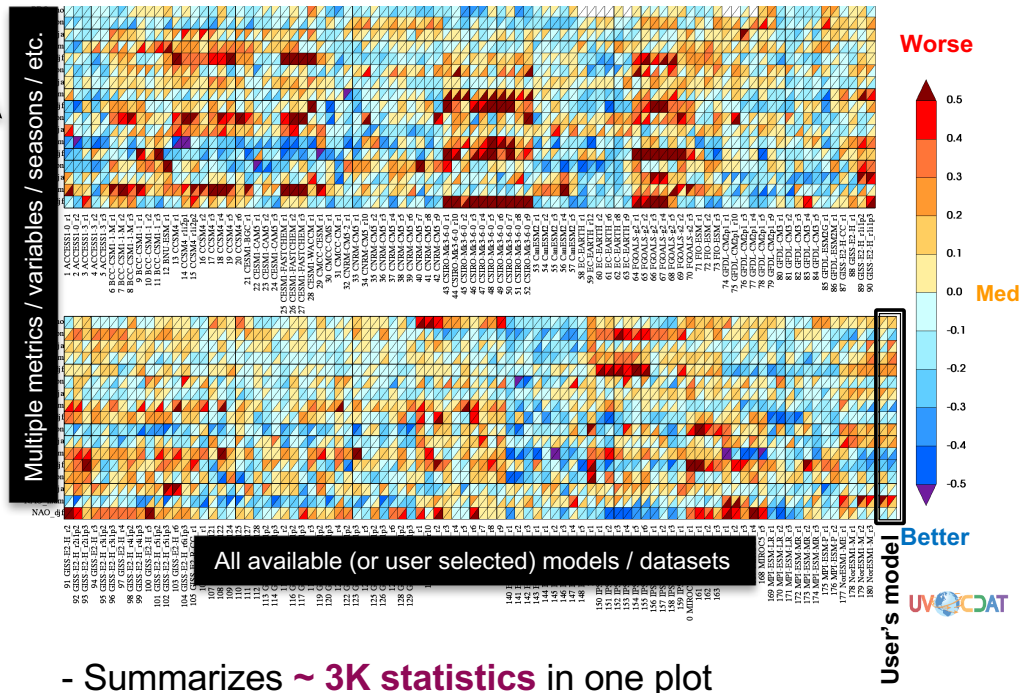
# On-going further works (1): *Server-side analysis*

- Server-side calculation of metrics

Goal: Enable user to apply new metrics and/or simulations across federation



Different aspects

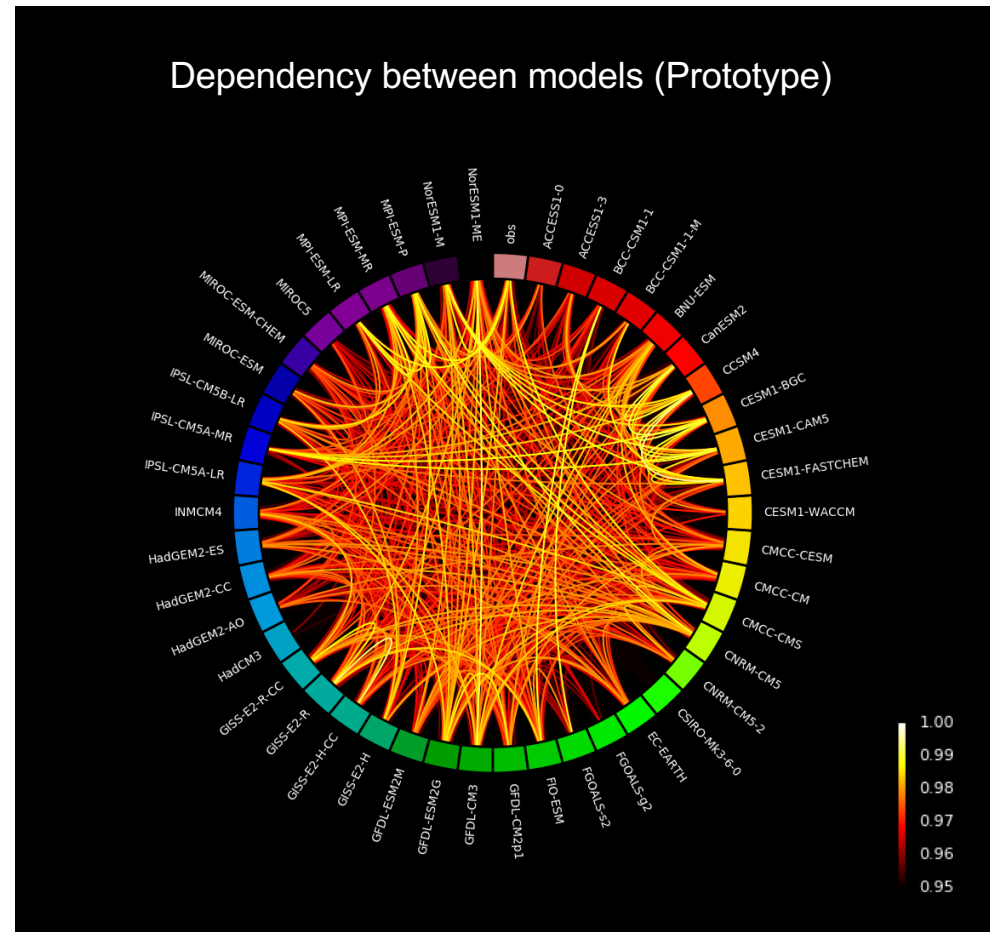


- Summarizes ~ 3K statistics in one plot

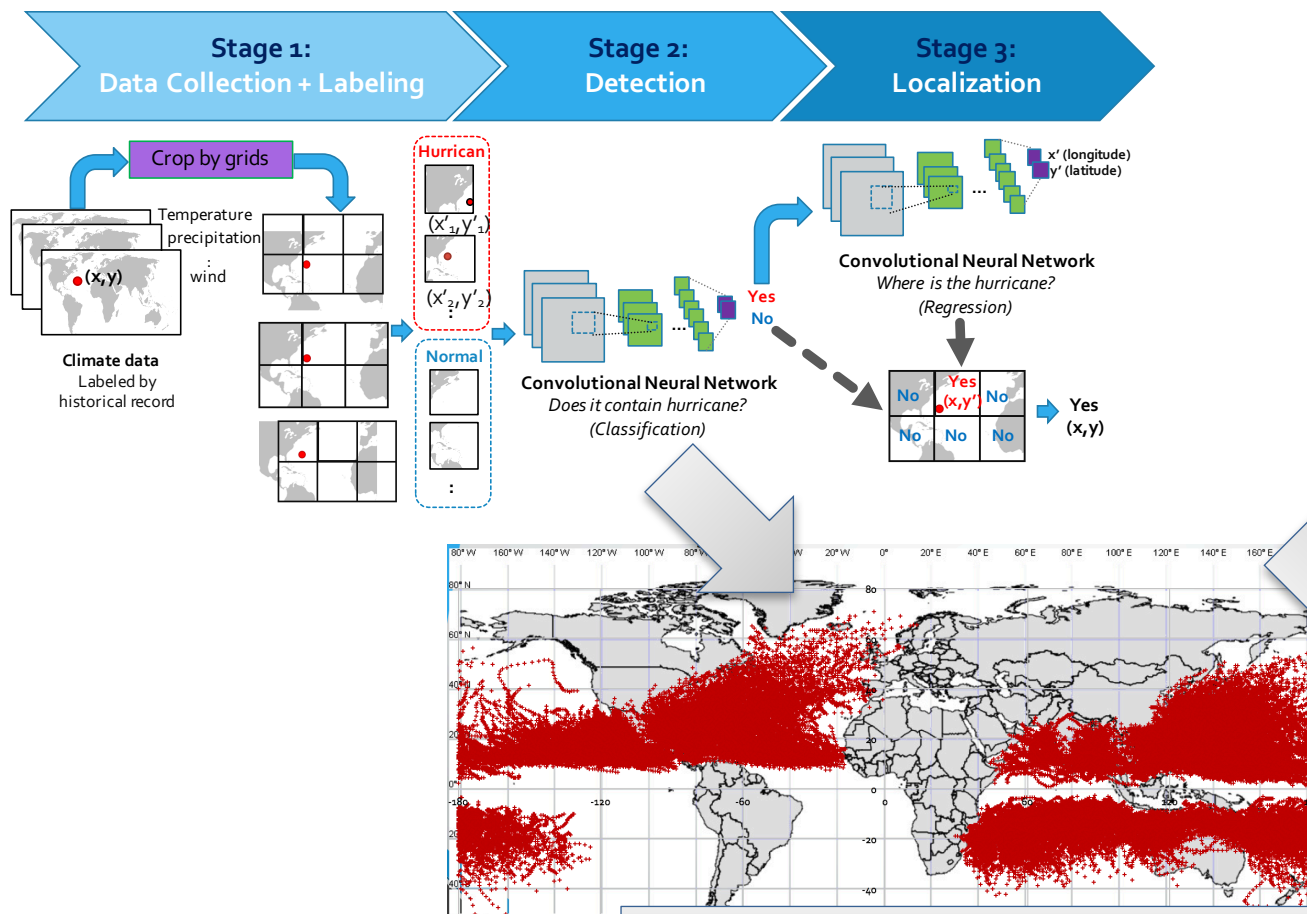
# On-going further works (2): *Big-data visualization*

Developing alternative way of big-data visualization:

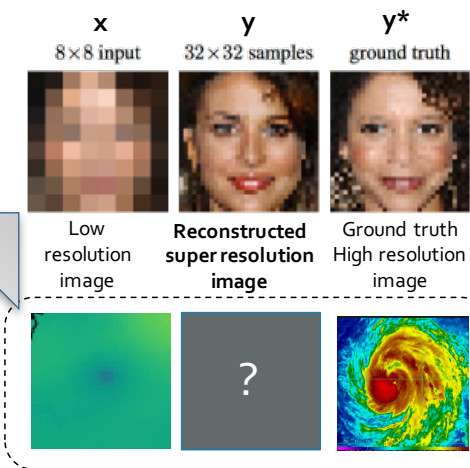
- Example plot enables identifying dependency between models
- Originally designed for bio-medical applications to identify connectivity between genes or cells
- Summarizes relationship between **more than 30K statistics** in one plot
- Open doors to *Climate model genealogy*



# On-going further works (3): *Machine Learning*



Dahl, Ryan, Mohammad Norouzi, and Jonathon Shlens. "Pixel recursive super resolution." *arXiv preprint arXiv:1702.00783* (2017)



**Courtesy of Dr. Soo Kyung Kim**

**Further detail:**  
S. K. Kim, S. Ames, J. Lee, C. Zhang, A. C. Wilson, D. Williams (2017) Massive Scale Deep Learning For Detecting Extreme Climate Events. *Climatic Informatics* (in review)

**Thank you, Questions?**

---

# Extra Back-up Slides



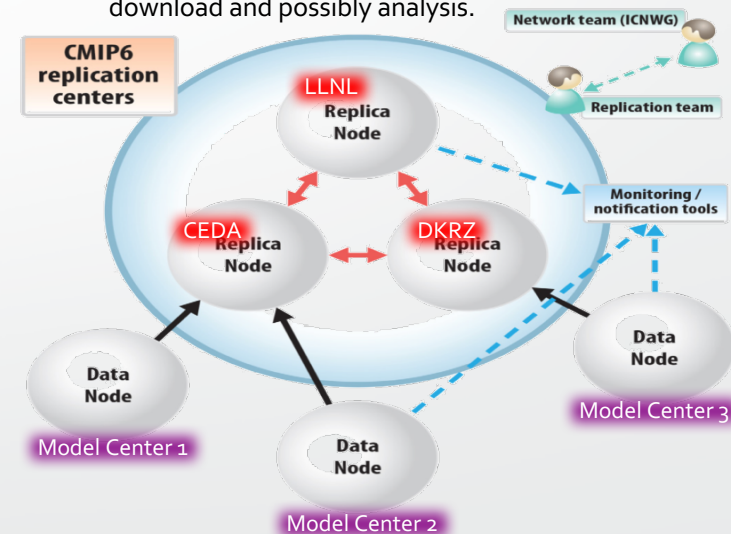
# ESGF Tier 1 and Tier 2 node site requirements for CMIP6

## • Tier 1 sites requirements

- Tier 1 sites are expected to **run the full suite of ESGF services** for data and user management, which can be used to support their own activities and those of Tier 2 sites
- Have an **uptime (>98%)**
- Have **10 petabytes of spinning disk** storage space
- Have at least a **10 gigabits per second connection** to their wide-area network provider with plans to upgrade beyond 10Gbps by 2017
- Run a 10 gigabits per second **perfSONAR host** (preferably on a physical server),
- Deploy at least four 10 gigabits per second **Data Transfer Nodes (DTNs)** in a "Science DMZ" environment with plans to run production ESGF data services on the DTNs by 2017
- Deploy sufficient high-performance storage to allow the DTNs to effectively serve CMIP6 data at high performance levels
- Publish data **using GridFTP and Globus URLs** in addition to wget URLs,
- Configure the DTNs to use Globus as well as GridFTP and wget, and
- Use **Synda for data replication** between Tier 1 sites.
- **Core monitoring services**
  - Unnoticed downtime by system administrator should not happen
  - Certificates end of validity monitoring
- Quality of Service (TBD)

## • Tier 2 sites requirements

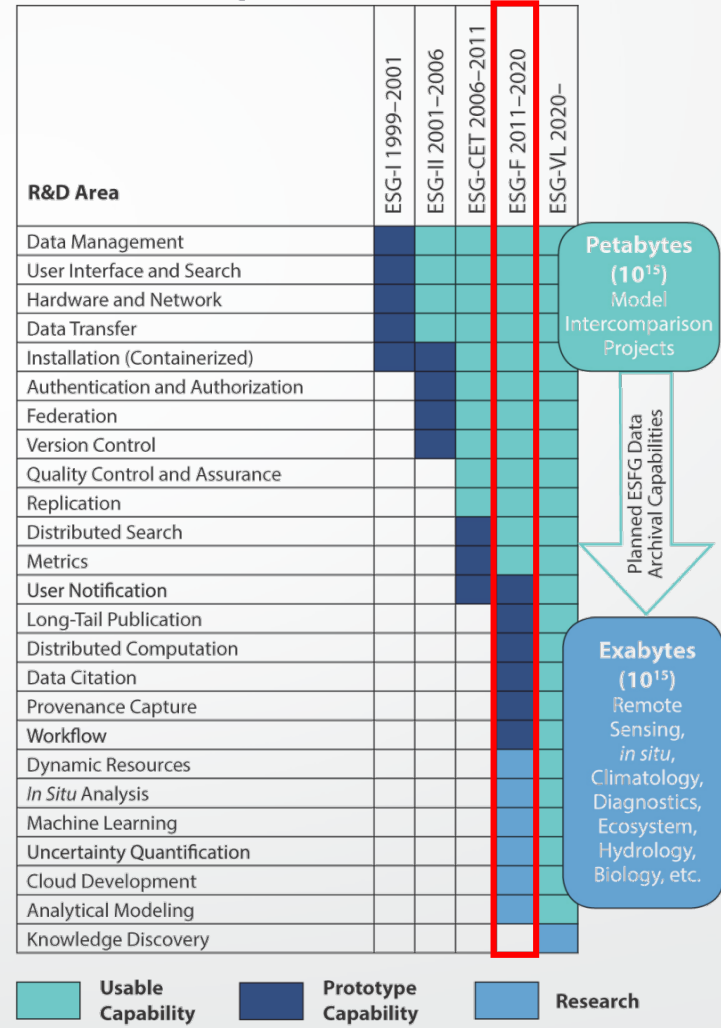
- Centers that typically have fewer physical or staff resources available for ESGF interactions but that still need to distribute a certain (possibly significant) amount of data to the scientific community.
- Tier 2 sites are **encouraged to leverage some of the services supported by Tier 1 sites**, such as a Metadata Index and Identity Provider, and focus instead on supporting local services for data download and possibly analysis.



# Future work on the horizon

- **Dynamic resources:** Manage and enhance user accessibility.
- **In situ analysis:** Enable users to obtain real-time computational results by performing calculations at the site where data resides.
- **Machine learning:** Find patterns and features in exabyte-scale data, providing users with new insights based on the vast wealth of available data.
- **Uncertainty quantification:** Connect uncertainty in underlying data with the reliability of conclusions.
- **Cloud development:** Enable easy setup of ESGF nodes and services on the cloud, either for permanent management or for transitional computational tasks.
- **Analytical Modeling:** Gauge how long calculations will take on various platforms, so that users can optimize their use of worldwide computational resources.

## ESGF Roadmap





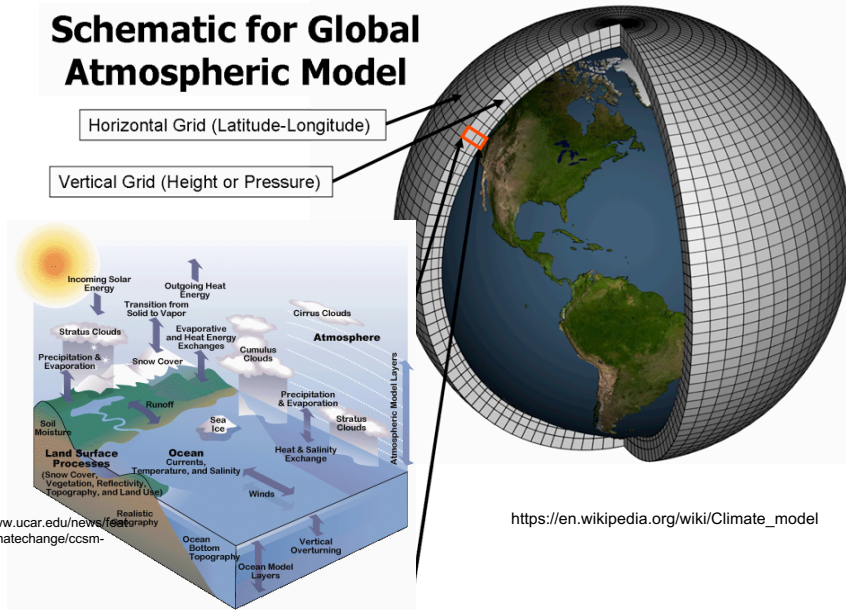
# Previous studies

- **The majority of previous studies have focused on one or two modes of variability**
- **A few studies have conducted systematic evaluation for a variety of modes, e.g.:**
  - *Stoner et al. (2009)* focused on the CMIP3 simulations
  - *Phillips et al. (2014)* has developed diagnostic package (NCAR CVDP) and released a repository for evaluating simulated modes in CMIP5
- **We expand these studies to develop metrics with an emphasis on how to:**
  - Objectively compare the models with observations, including seasonality
  - Test the skill sensitivity to:
    - 1) multiple realizations from individual models
    - 2) choice of observations
    - 3) methodological consideration
  - Ascertain the role of pattern error versus amplitude error in assessing the fidelity of the simulations using *skill metrics*

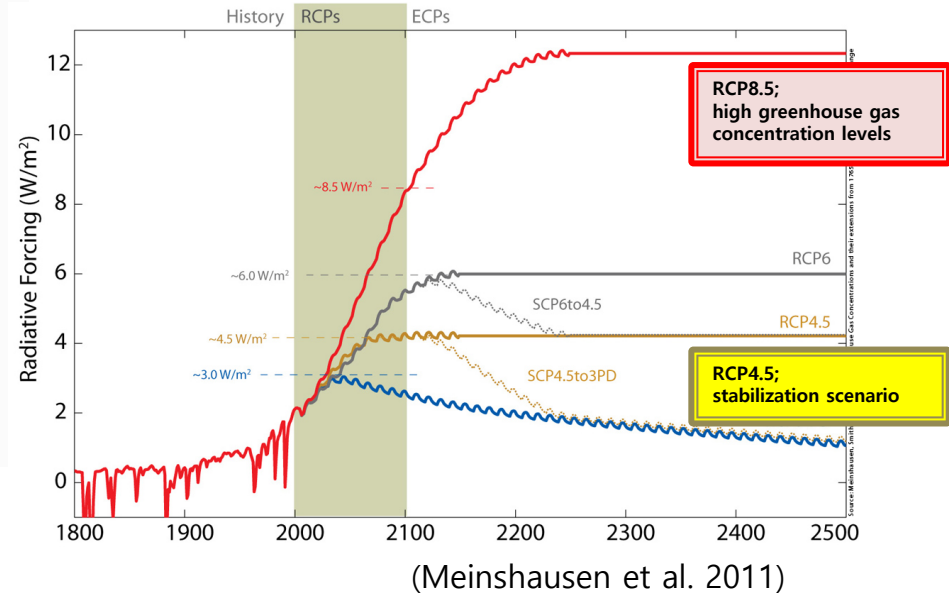


# Modeling Climate Under Future Possible Change Scenarios

## Schematic for Global Atmospheric Model



[https://en.wikipedia.org/wiki/Climate\\_model](https://en.wikipedia.org/wiki/Climate_model)

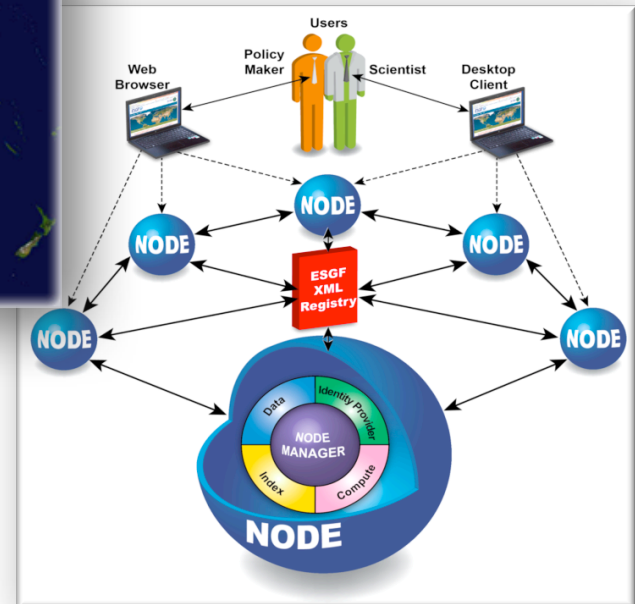


<http://www.ucar.edu/news/features/climatechange/ccsm-text.jsp>

# World-wide Climate Data Archive



+ Data standardization,  
which open door to  
***Climate Model Inter-comparison***



# Modes of variability: NAO

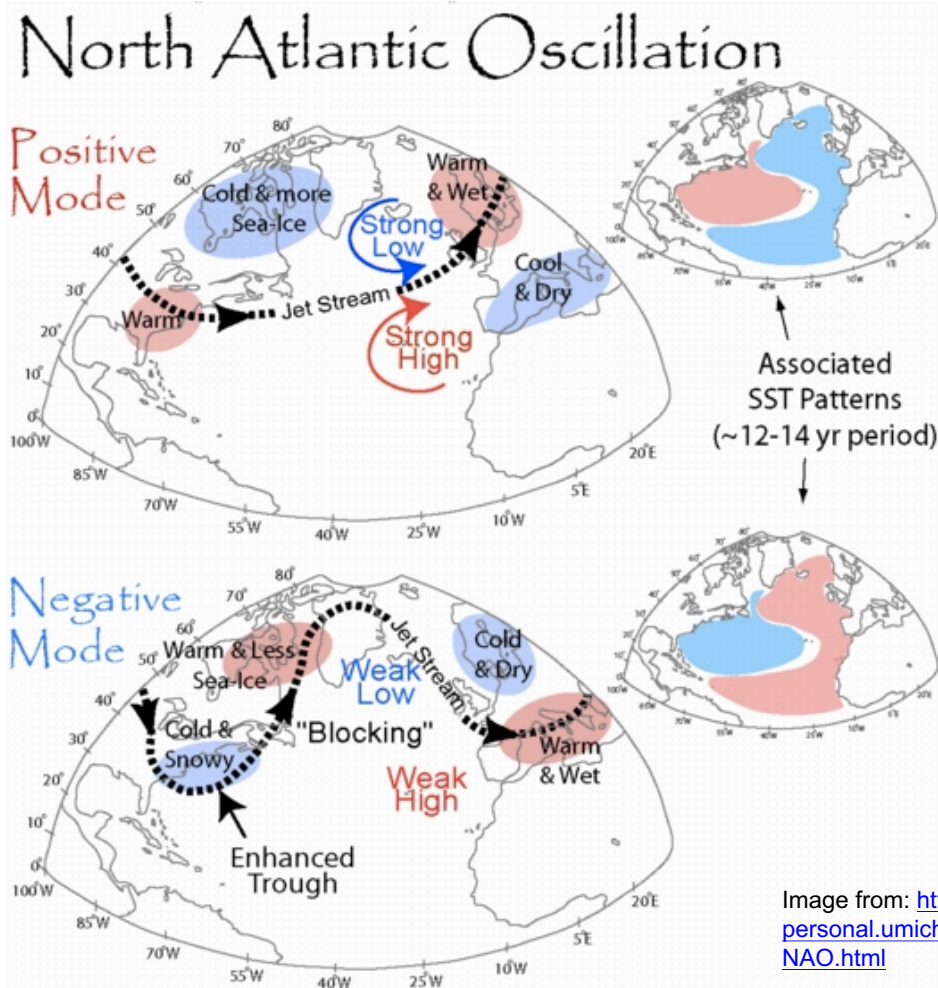
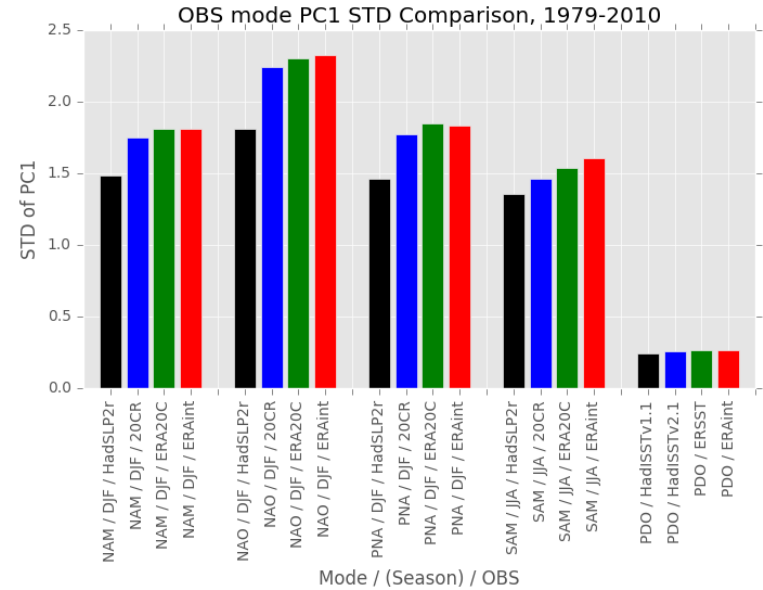
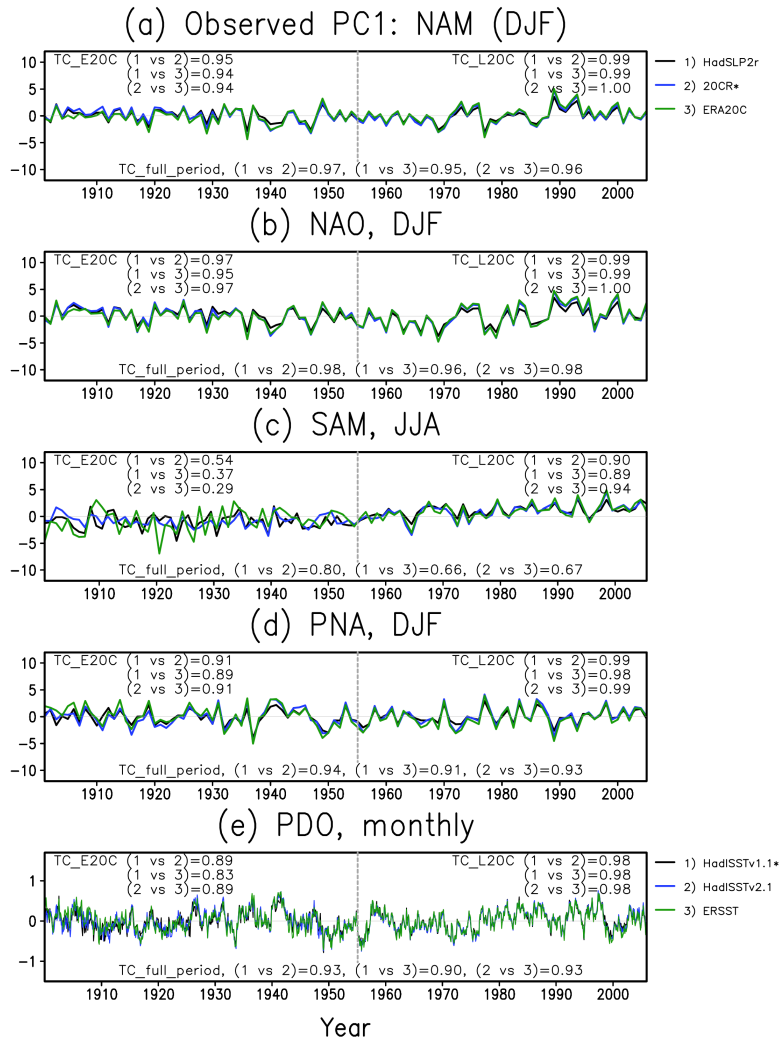


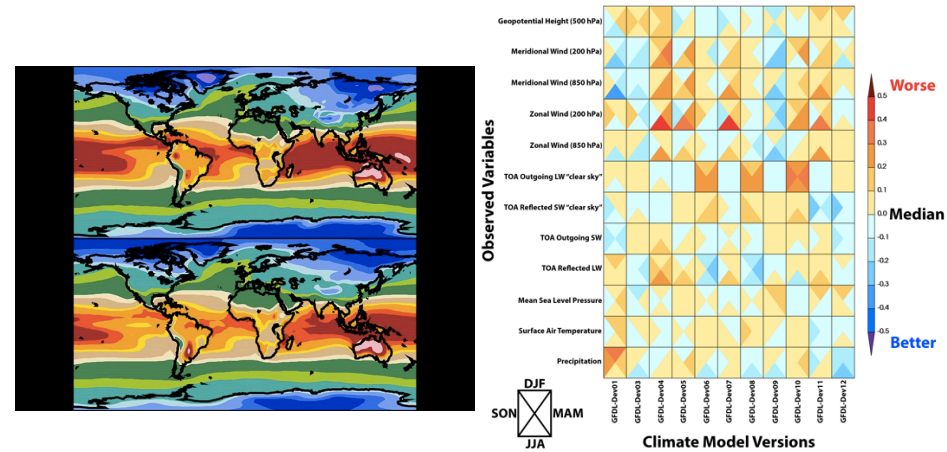
Image from: <http://www-personal.umich.edu/~auraell/precipitation/pages/NAO.html>

# Observation



# Software Development

- **Python** (free of H/W System dependency)
- **Built over:**
  - PCMDI Metrics Package (PMP)
  - UV-CDAT
  - EOFs (Dawson 2016)
- **Open source (Github)**
- **Reusable code**
  - user friendly designed
  - self-describing documentation



Gleckler, P. J., C. Doutriaux, P. J. Durack, K. E. Taylor, Y. Zhang, D. N. Williams, E. Mason, and J. Servonnat (2016), A more powerful reality test for climate models, *Eos*, 97, doi:10.1029/2016EO051663. Published on 3 May 2016.

# Python EOFs Library

- High-level interface for computing empirical orthogonal functions (EOFs) and related quantities

Dawson, A., (2016). eofs: A Library for EOF Analysis of Meteorological, Oceanographic, and Climate Data. Journal of Open Research Software. 4(1), p.e14. DOI: <http://doi.org/10.5334/jors.122>

Journal of Open Research Software

Dawson, A 2016 eofs: A Library for EOF Analysis of Meteorological, Oceanographic, and Climate Data. Journal of Open Research Software, 4, e14, DOI: <http://doi.org/10.5334/jors.122>

SOFTWARE METAPAPER

## eofs: A Library for EOF Analysis of Meteorological, Oceanographic, and Climate Data

Andrew Dawson<sup>1</sup>  
<sup>1</sup>Atmospheric, Oceanic & Planetary Physics, Department of Physics, University of Oxford, [andrew.dawson@physics.ox.ac.uk](mailto:andrew.dawson@physics.ox.ac.uk)

The eofs library provides a high-level Python interface for computing EOFs and related quantities, with a focus on correctness and ease of use. The library is designed to be modular and hierarchical, allowing computations using plain Python or through popular libraries for atmospheric and climate science.

**Keywords:** EOF analysis; Meteorology; Oceanography; Climate; Python

### (1) Overview

**Introduction**

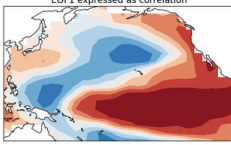
Data sets in meteorology, oceanography, and climate are typically very large, containing data covering large spatial areas, observed or modelled over long periods of time. Studying variability in these data sets can be challenging, with coherent modes of large-scale spatial and temporal variability in the atmosphere-ocean system hidden amongst the noise of smaller scale physical processes. An often used technique for examining large-scale patterns of variability in such data sets is the analysis of empirical orthogonal functions (EOFs) [1]. Decomposing a complex data set varying in time and space into a set of EOFs and associated principal component time series (PCs) can allow insight into the most dominant modes of spatial variability, for example El Niño, one of the leading modes of climate variability, is often characterised by the first EOF and PC of sea surface temperature in the tropical Pacific [2].

The EOFs and PCs of a data set describe a new basis, where instead of a series of spatial observations varying in time, the data set is represented as a set of fixed spatial patterns or modes, which represent a given amount of the total variance in the data set, and a set of time series describing how each pattern changes with time. In typical applications the first few EOFs account for a large portion of the total variance, allowing the study of one or two modes to give insight into the variability present in the data set. The method of analysis is purely mathematical and does not depend on any physical properties of the quantity being analysed.

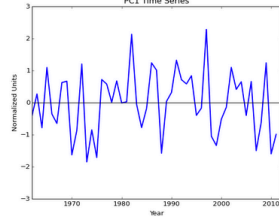
The process of computing and analysing EOFs and related structures is non-trivial, and highly error prone. For example, consider the computation of EOFs from a time-series longitude grid data to account cells due to coordinate system and care taken of an oceanographic matrix calculation as the (possibly) singular matrix. In order to undo the eigenvectors inserting any dimensions, and weight one will not just also in other d associated with onto the EOFs tion procedure and great care of each quantity. There are several computing EOF type of data using an unpubl- icated EOF analysis procedures to that cannot be kept for the analysis of the major eofs was to re- object-orient-

### El Niño

Course code



EOF1 expressed as correlation



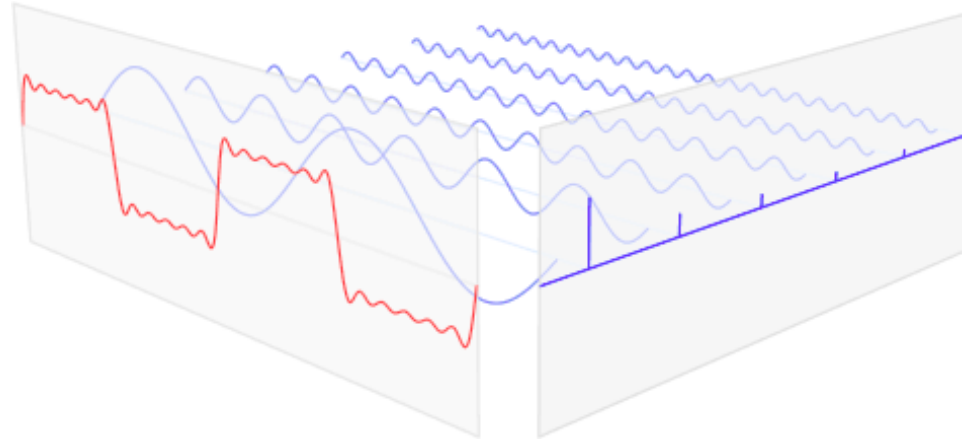
PC1 Time Series

```
from eofs.standard import EOF
from eofs.variables import example_data_path

# Read EOF anomalies using the netCDF module. The file contains
# November-March coverage of SST anomaly (in the central and western Pacific).
filename = example_data_path('el_nino_1970_2010.nc')
pc1 = Dataset(filename)['pc1']
```

# Empirical Orthogonal Function (EOF) analysis

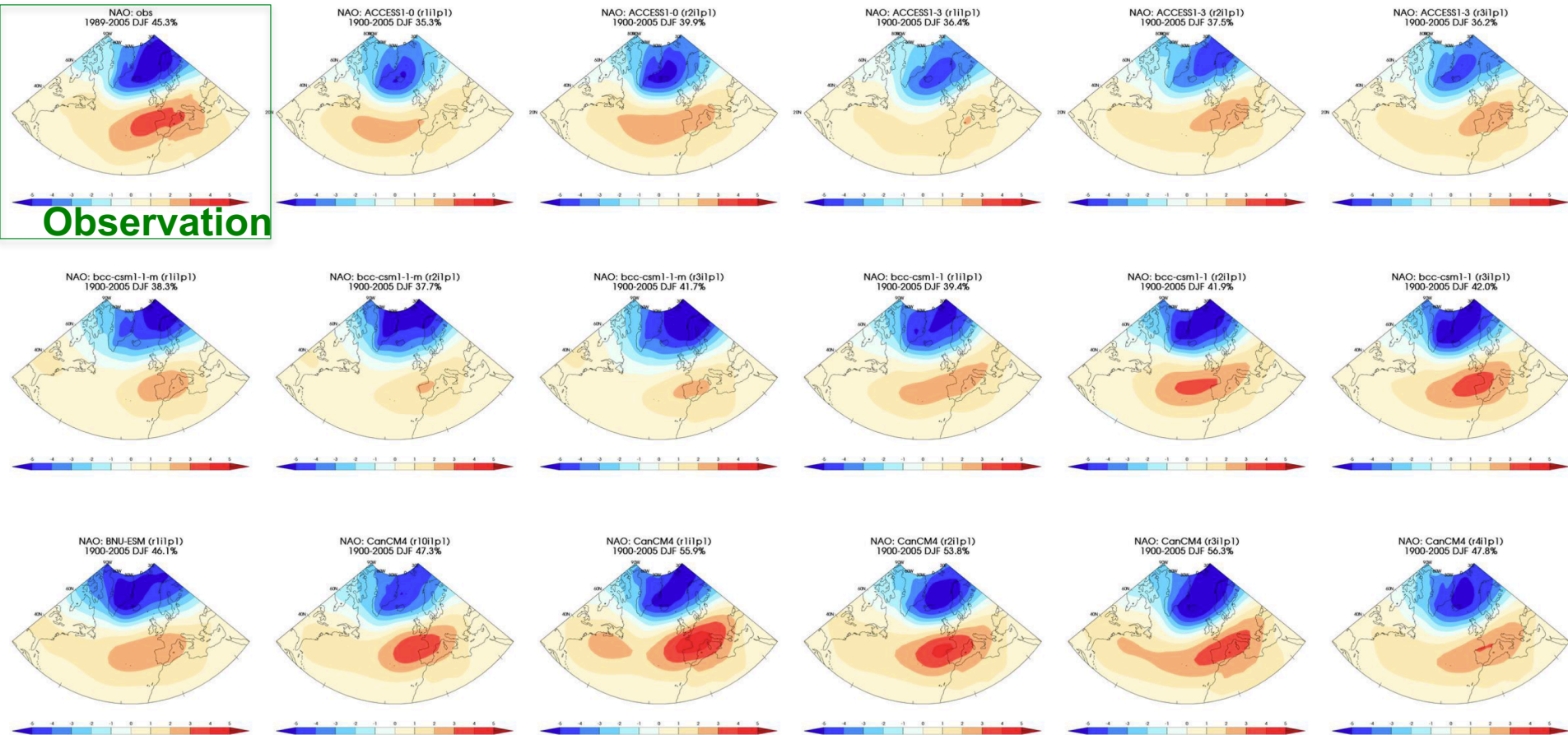
- **Fourier Transform:**



<http://tex.stackexchange.com/questions/127375/replicate-the-fourier-transform-time-frequency-domains-correspondence-illustrati>

- **EOF:**
  - **Similar concept to FT, but separates modes based on orthogonality**
  - **It is more useful when time series has “jumps” in it**

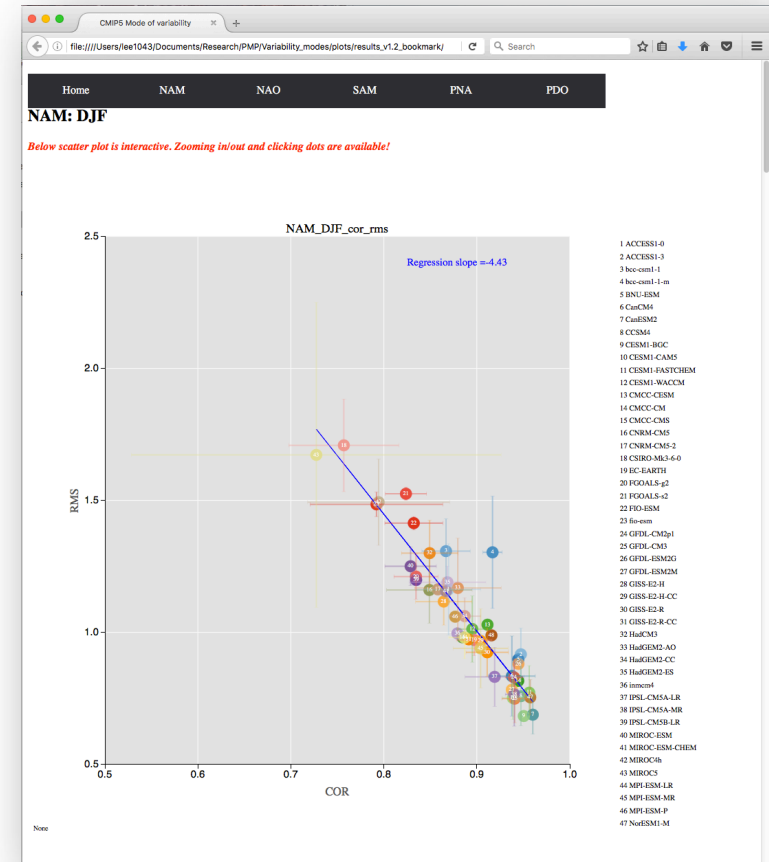
# Loop through all CMIP5 simulations.. (e.g. NAO)



5 modes, 4 seasons, about 45 models with all available realizations (1~25 per model), various type of plots;  
≈ 13,000 images at the end...

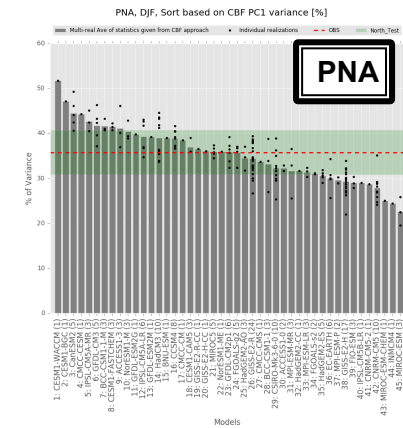
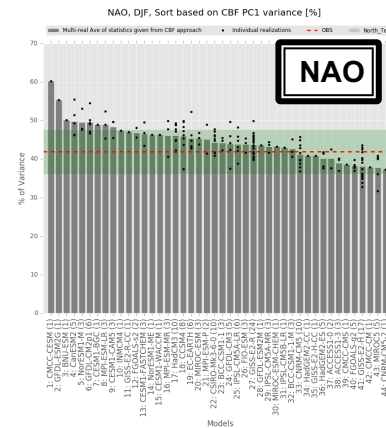
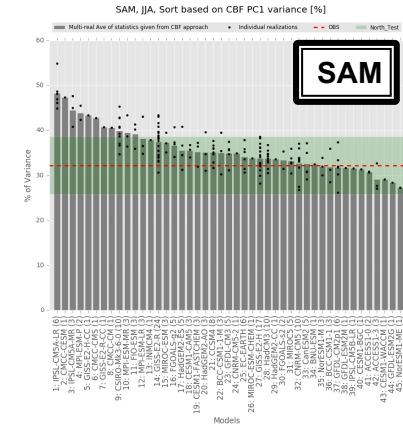
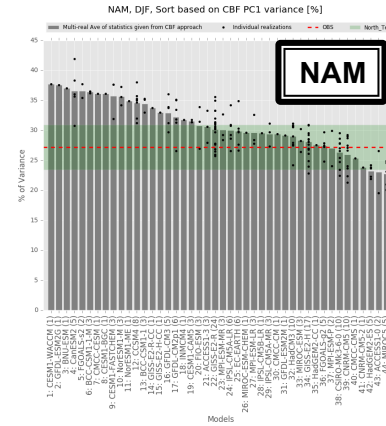
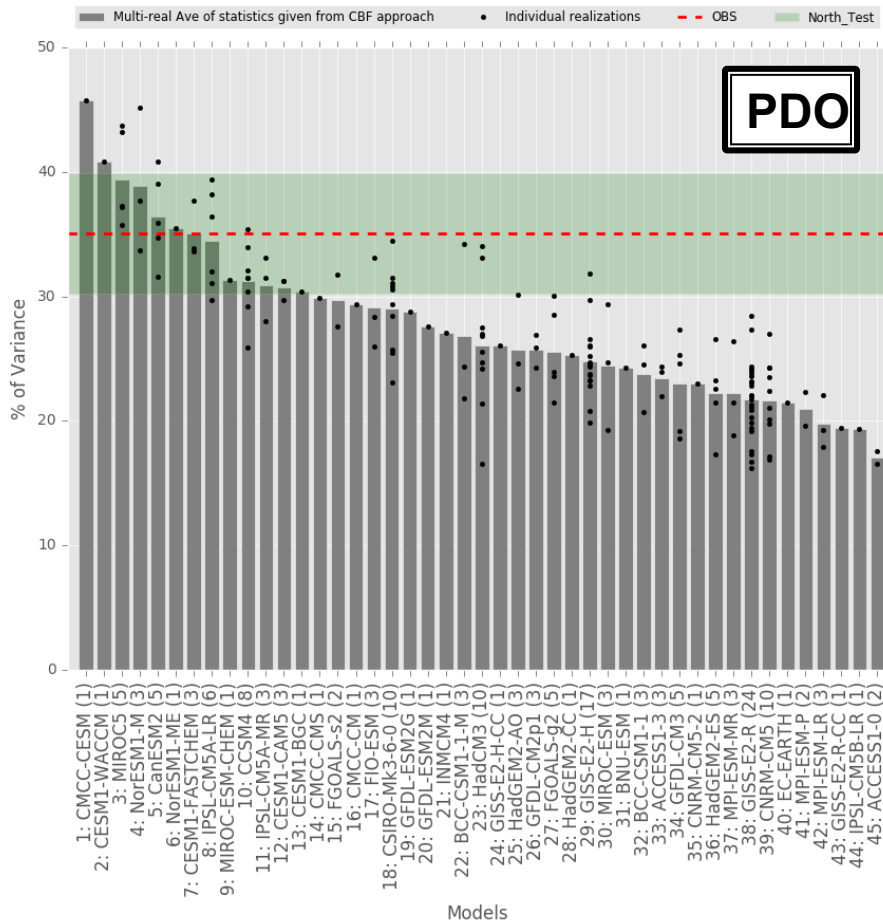
# Database Construction

- Browsing available web-like navigation
- Statistics based **Interactive plots** (**mpld3** Library)
- Starting version for **PCMDI web service**



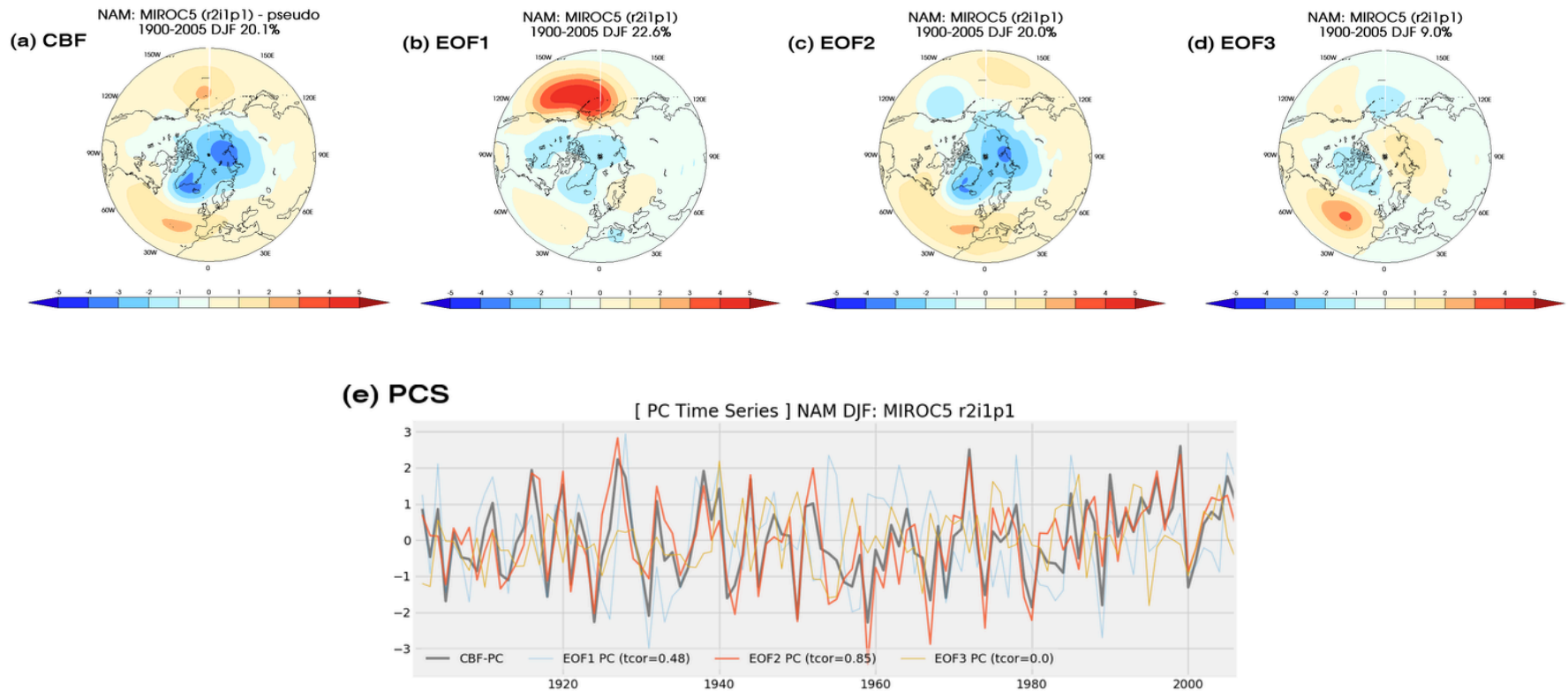
# PDO and others - Variance in time

PDO, yearly, Sort based on CBF PC1 variance [%]



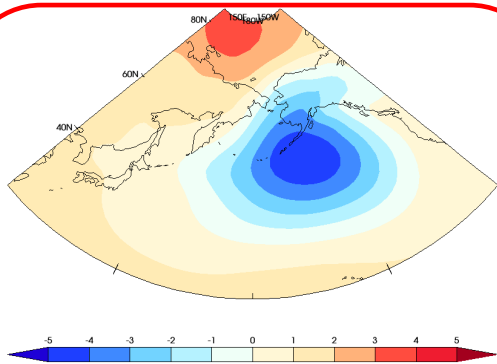
# EOF mode swapping: NAM, DJF

- NAM in MIROC5 (r2i1p1)



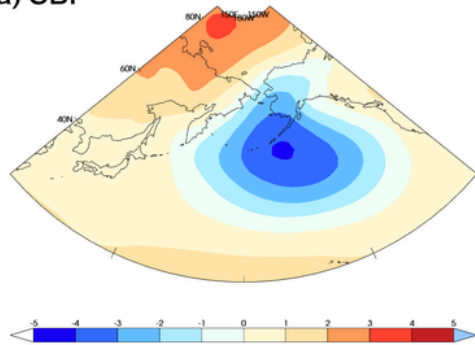
# EOF mode swapping and CBF: PNA Example

**OBS**



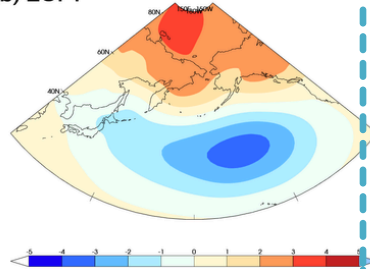
**PNA: Pacific North American Pattern (20CR, SLP)**

**(a) CBF**  
PNA: CNRM-CM5 (r1i1p1) - CBF  
1900-2005 DJF 28.9%

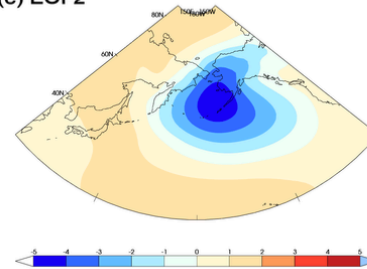


**Model**

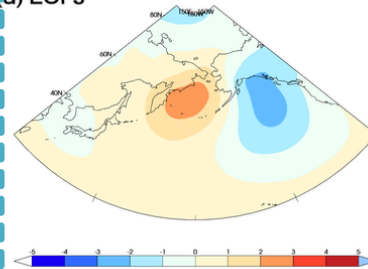
**(b) EOF1**  
PNA: CNRM-CM5 (r1i1p1)  
1900-2005 DJF 32.4%



**(c) EOF2**  
PNA: CNRM-CM5 (r1i1p1)  
1900-2005 DJF 28.3%



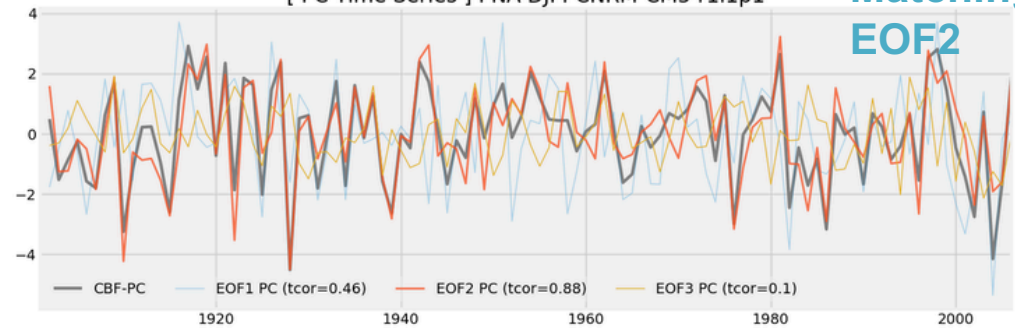
**(d) EOF3**  
PNA: CNRM-CM5 (r1i1p1)  
1900-2005 DJF 9.9%



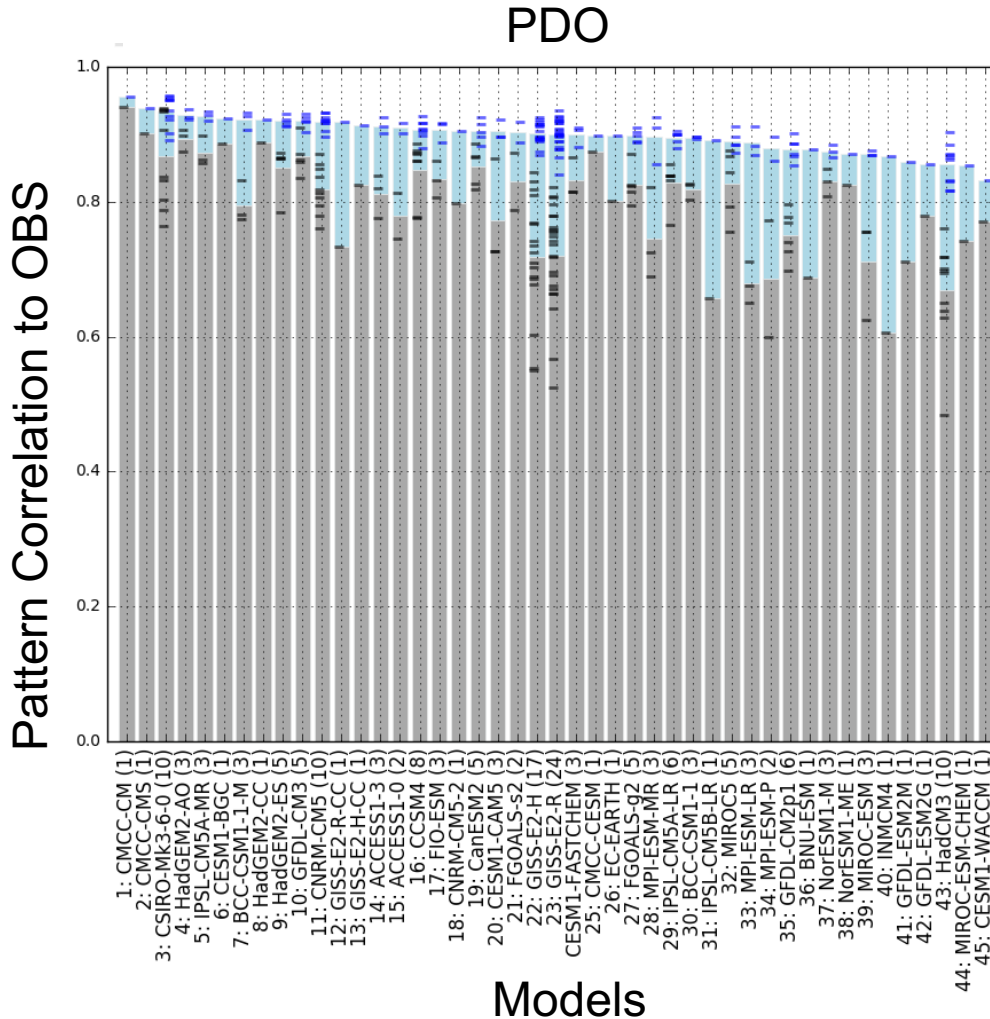
**Best Matching: EOF2**

**(e) PCS**

[ PC Time Series ] PNA DJF: CNRM-CM5 r1i1p1



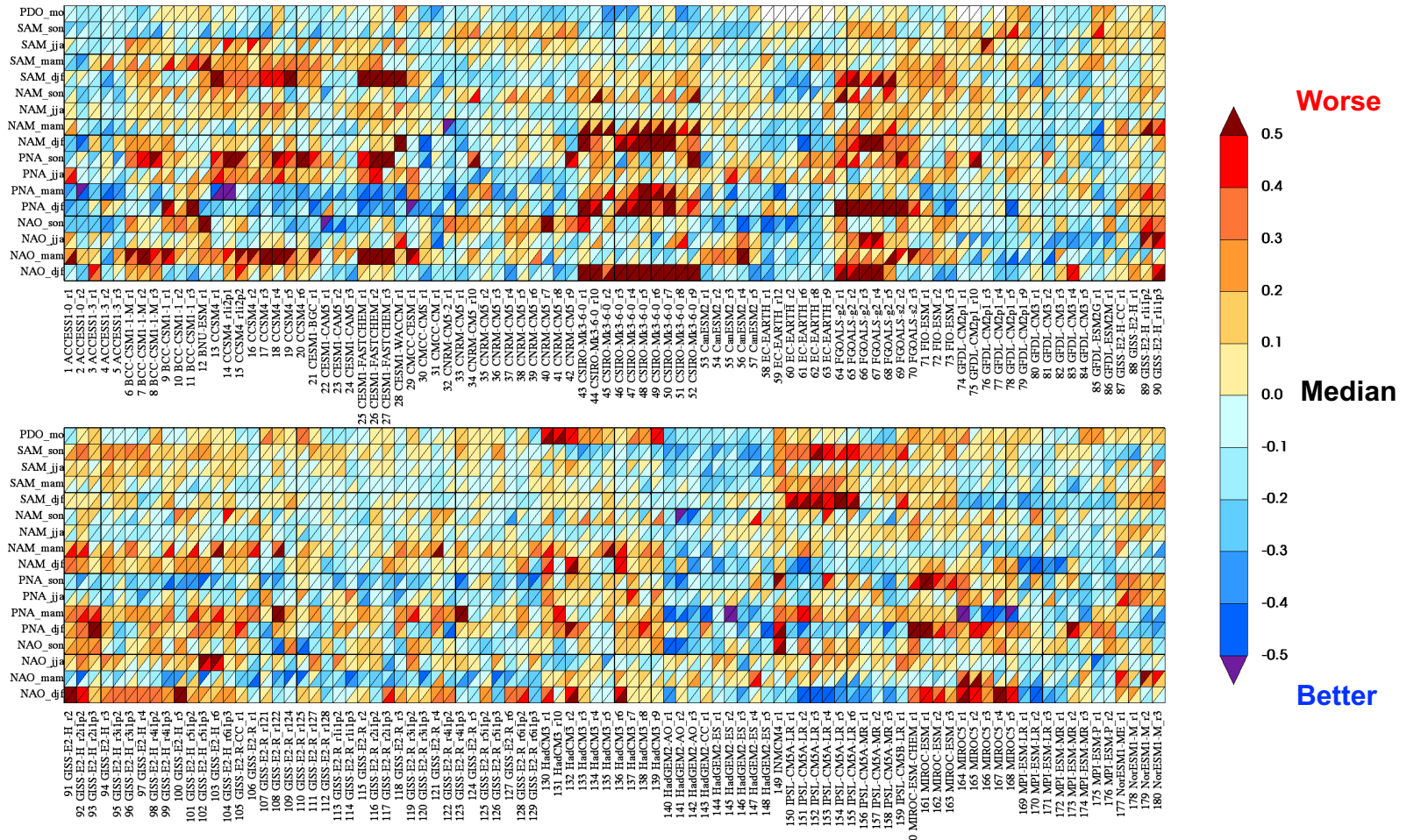
# CBF vs. EOF swapping



- CBF has better performance than EOF swapping
- CBF brings models to space of more fair comparison

# Performance Diagnostics: Pattern (RMSE) – Unit variance map

RMS using CBF approach with 20CR vs ERA20C 1900-2005 (Unit Variance Map)



# Amplitude comparison for CBF

Markers: Average of all seasons and realizations per model

