Understanding Climate Change: Opportunities and Challenges for Data Intensive Computational Science

Vipin Kumar

University of Minnesota

kumar@cs.umn.edu
www.cs.umn.edu/~kumar

Collaborators and Group Members:

Shyam Boriah, Jaya Kawale
Karsten Steinhaeuser, Stefan Leiss
Arjun Kumar , Snigdhansu Chatterjee
Peter Sneider, Joe Knight,
Michael Steinbach
University of Minnesota

Alok Choudhary
Northwestern university

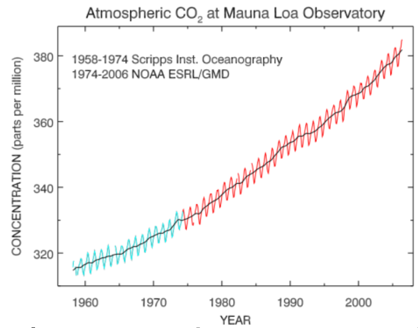
Auroop Ganguly
ORNL

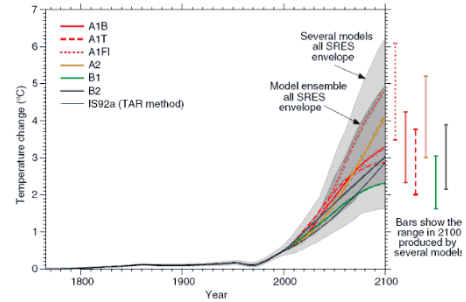
Fred Semazzi,
Nagiza Samatova
NC State

Research funded by NSF, NASA, Planetary Skin Institute, ISET-NOAA, MN Futures program

Climate Change: The defining issue of our era

- **The planet is warming**
  - Multiple lines of evidence
  - Credible link to human GHG (green house gas) emissions



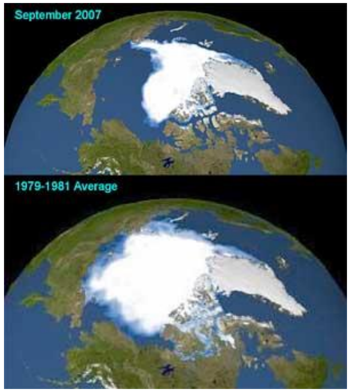


- **Consequences can be dire**
  - Extreme weather events, regional climate and ecosystem shifts, abrupt climate change, stress on key resources and critical infrastructures

- **There is an urgency to act**
  - Adaptation: "Manage the unavoidable"
  - Mitigation: "Avoid the unmanageable"

- **The societal cost of both action and inaction is large**

Russia Burns, Moscow Chokes
NATIONAL GEOGRAPHIC, 2010



The Vanishing of the Arctic Ice cap
ecology.com, 2008

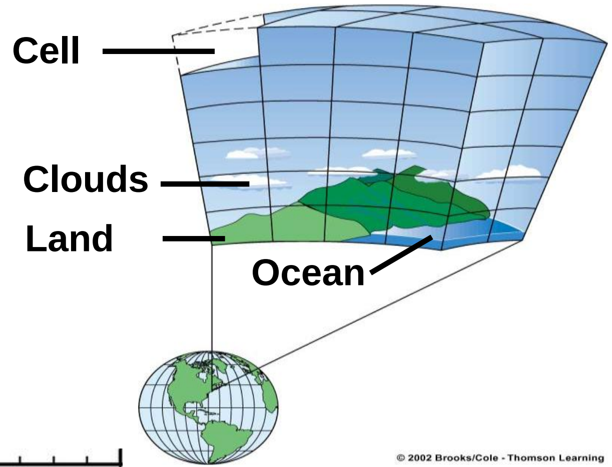
Key outstanding science challenge:

Actionable predictive insights to credibly inform policy

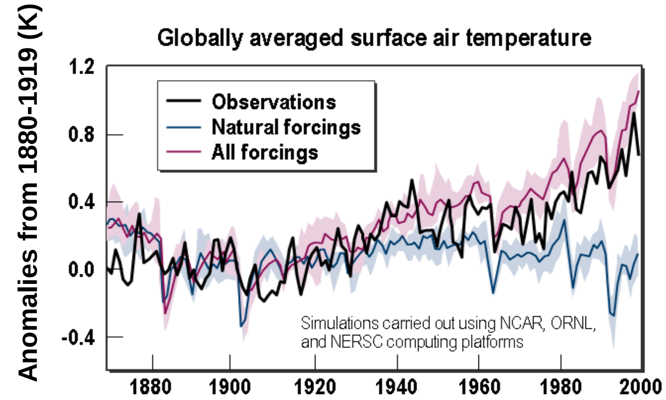
Understanding Climate Change - Physics based Approach

General Circulation Models: Mathematical models with physical equations based on fluid dynamics

Parameterization and non-linearity of differential equations are sources for uncertainty!

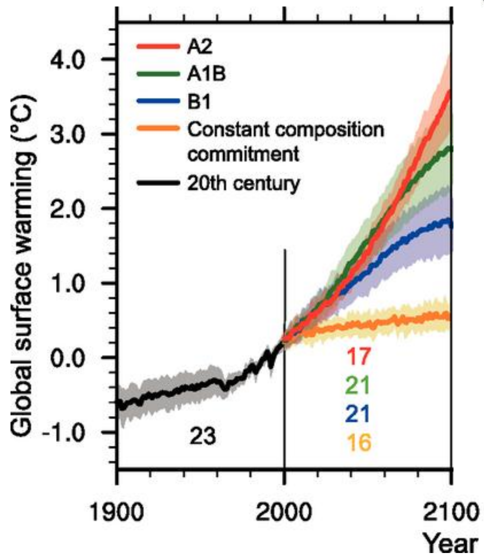


Temperature increases are human-induced
The anthropogenic climate change "fingerprint"



In the absence of human-induced changes to the atmosphere, the earth would be in a cooling trend

Figure Courtesy: ORNL



IPCC (2007)

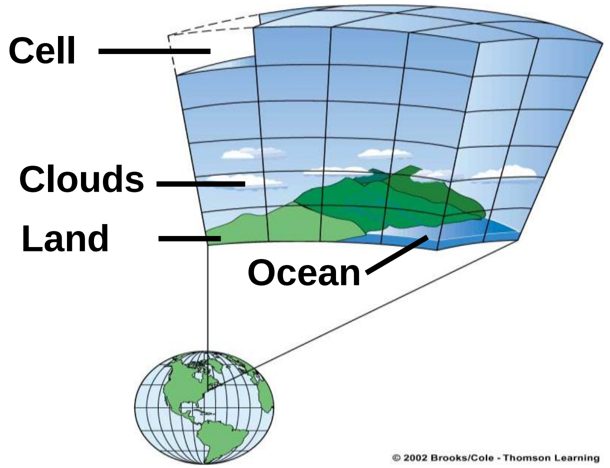
Projection of temperature increase under different **Special Report on Emissions Scenarios** (SRES) by 24 different GCM configurations from 16 research centers used in the **Intergovernmental Panel on Climate Change** (IPCC) 4th Assessment Report.

A1B: "integrated world" balance of fuels
A2: "divided world" local fuels
B1: "integrated world" environmentally conscious

Understanding Climate Change - Physics based Approach

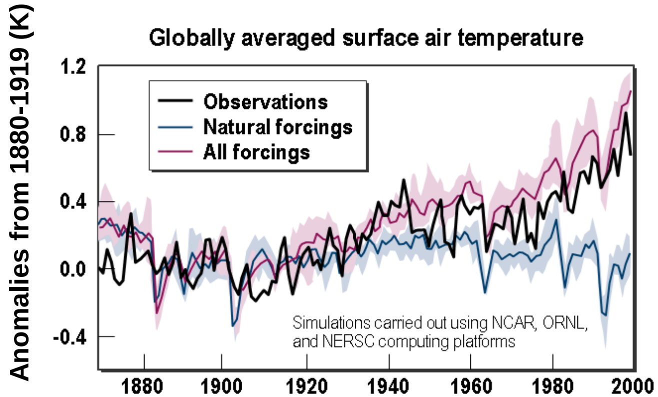
General Circulation Models: Mathematical models with physical equations based on fluid dynamics

Parameterization and non-linearity of differential equations are sources for uncertainty!



Temperature increases are human-induced

The anthropogenic climate change "fingerprint"



In the absence of human-induced changes to the atmosphere, the earth would be in a cooling trend

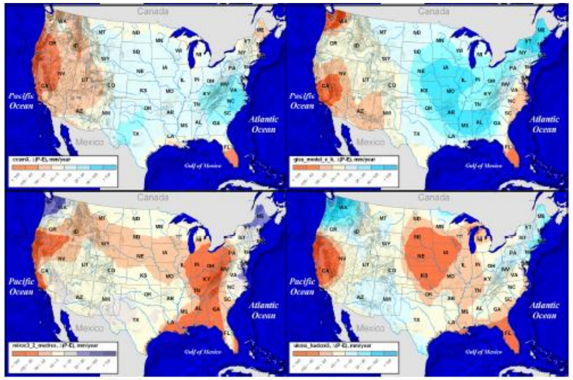
Figure Courtesy: ORNL

Physics-based models are essential but not adequate

- Relatively reliable predictions at global scale for ancillary variables such as temperature
- Least reliable predictions for variables that are crucial for impact assessment such as regional precipitation

"The sad truth of climate science is that the most crucial information is the least reliable"
(Nature, 2010)

Disagreement between IPCC models

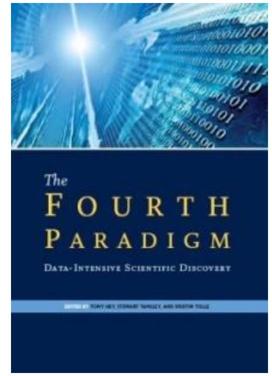


Regional hydrology exhibits large variations among major IPCC model projections

NSF Expedition: Understanding Climate Change - *A Data-Driven Approach*

Project aim:

A new and transformative data-driven approach that complements physics-based models and improves prediction of the potential impacts of climate change



"... data-intensive science [is] ...a new, fourth paradigm for scientific exploration." - Jim Gray

Transformative Computer Science Research

Predictive Modeling

Enable predictive modeling of typical and extreme behavior from multivariate spatio-temporal data

Relationship Mining

Enable discovery of complex dependence structures: non-linear associations or long range spatial dependencies

Complex Networks

Enable studying of collective behavior of interacting eco-climate systems

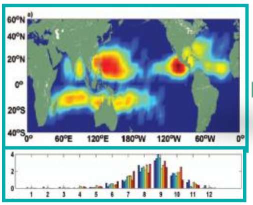
High Performance Computing

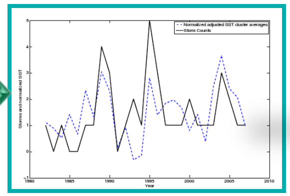
Enable efficient large-scale spatio-temporal analytics on exascale HPC platforms with complex memory hierarchies

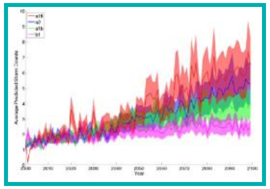
- *Science Contributions*
  - Data-guided uncertainty reduction by blending physics models and data analytics
  - A new understanding of the complex nature of the Earth system and mechanisms contributing to adverse consequences of climate change
- *Success Metric*
  - Inclusion of data-driven analysis as a standard part of climate projections and impact assessment (e.g., for IPCC)

Some Driving Use Cases

Impact of Global Warming on Hurricane Frequency







Find non-linear relationships

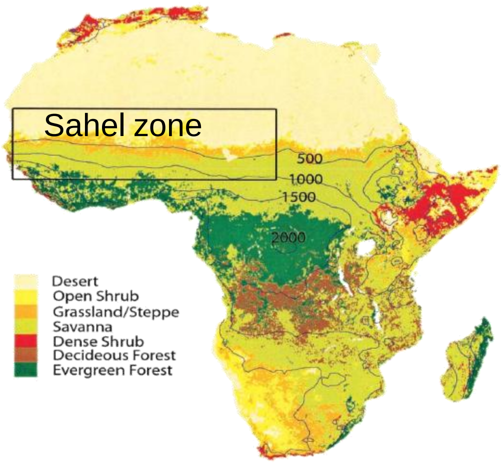
Validate w/ hindcasts

Build hurricane models

Regime Shift in Sahel

Onset of major 30-year drought over the Sahel region in 1969

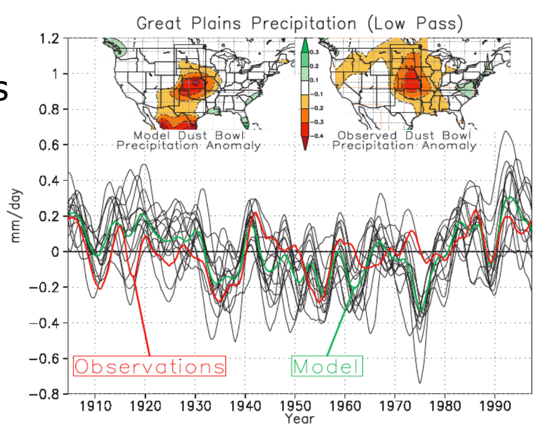
Regime shift can occur without any advanced warning and may be triggered by isolated events such as storms, drought



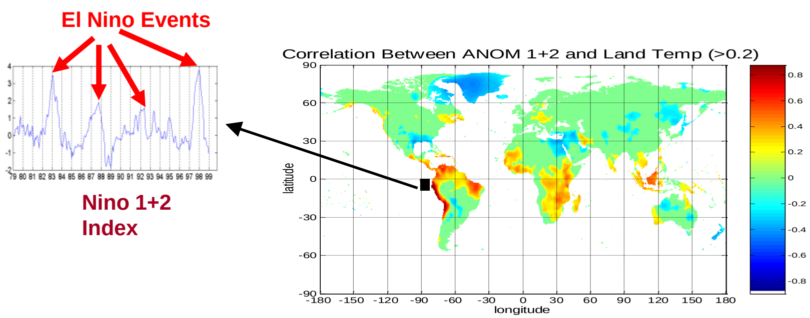
1930s Dust Bowl

Affected almost two-thirds of the U.S. Centered over the agriculturally productive Great Plains

Drought initiated by anomalous tropical SSTs (Teleconnections)

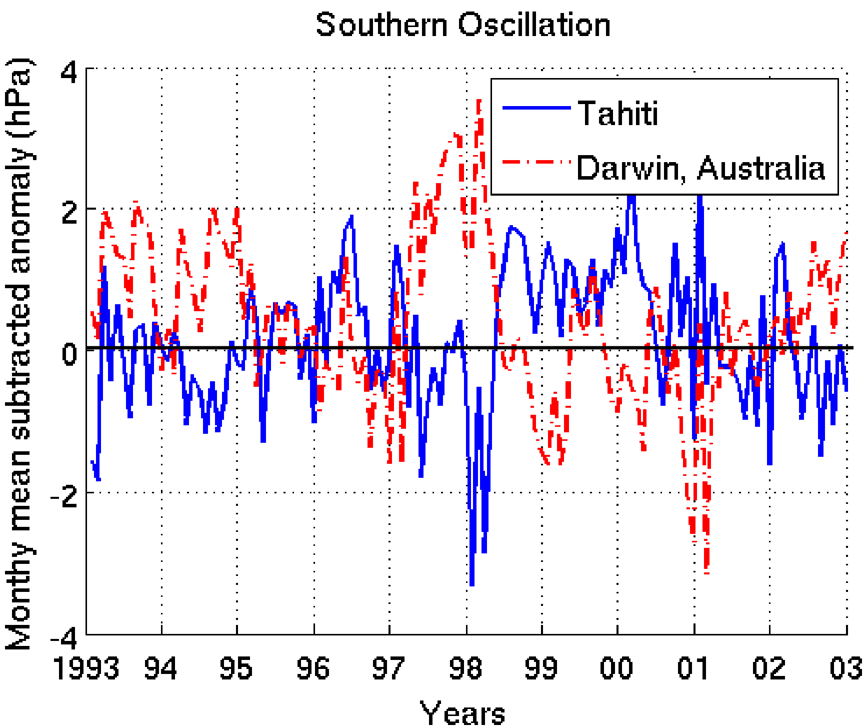


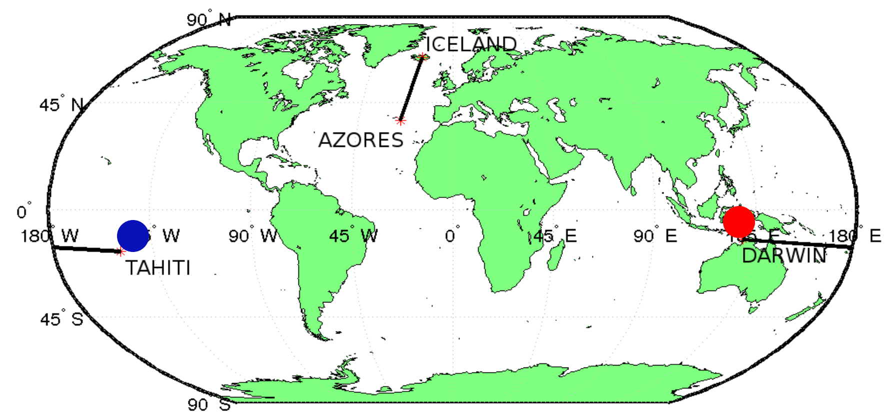
Discovering Climate Teleconnections



Understanding climate variability using Dipole Analysis

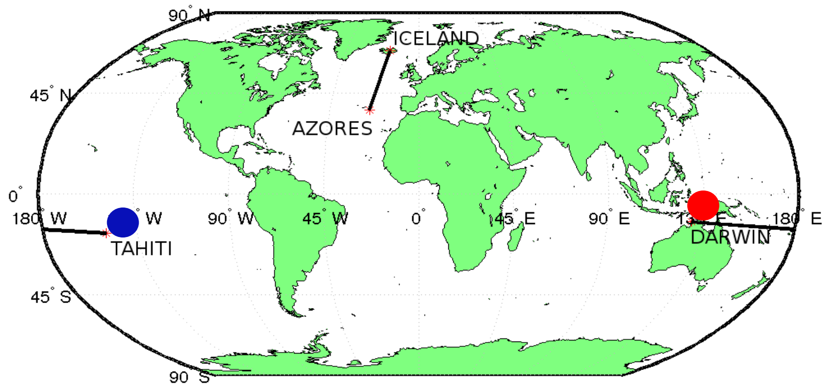
Dipoles represent a class of teleconnections characterized by anomalies of opposite polarity at two locations at the same time.

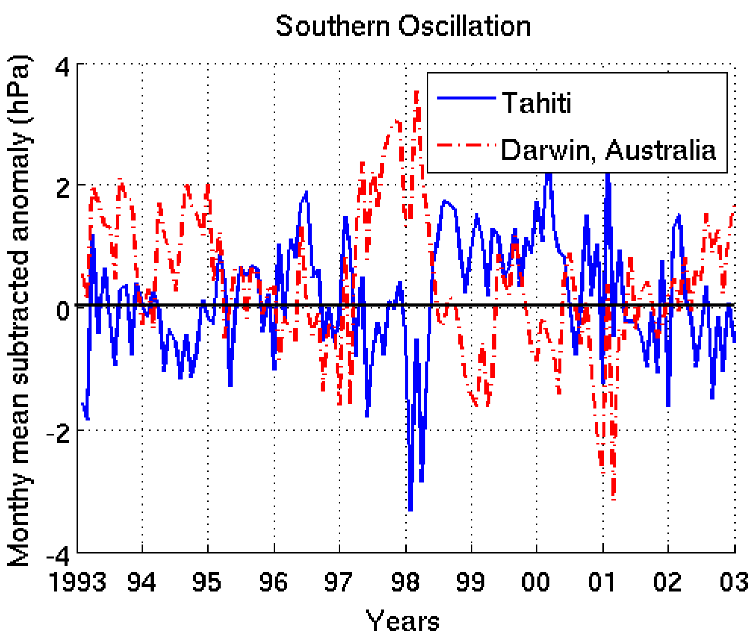


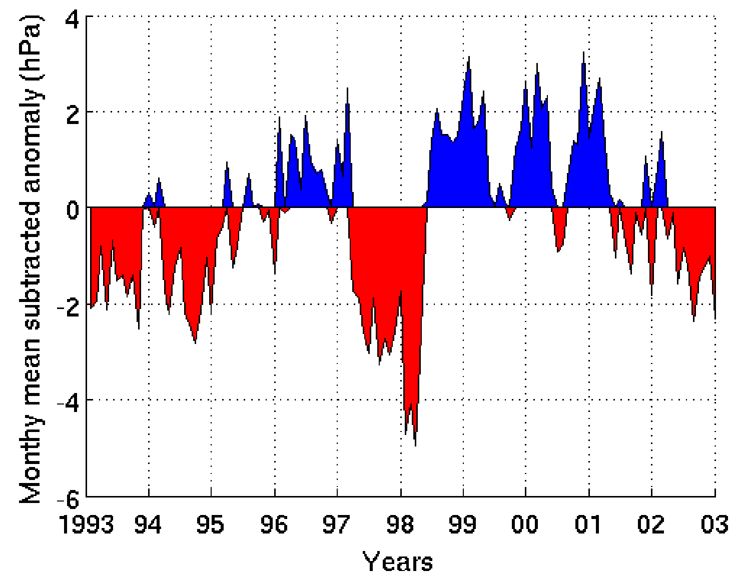


Understanding climate variability using Dipole Analysis

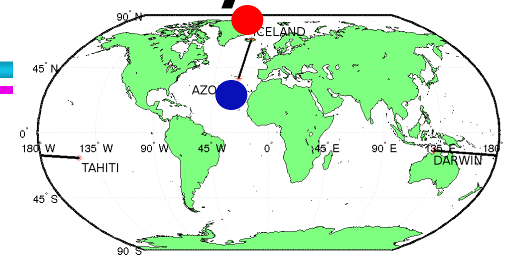
Dipoles represent a class of teleconnections characterized by anomalies of opposite polarity at two locations at the same time.





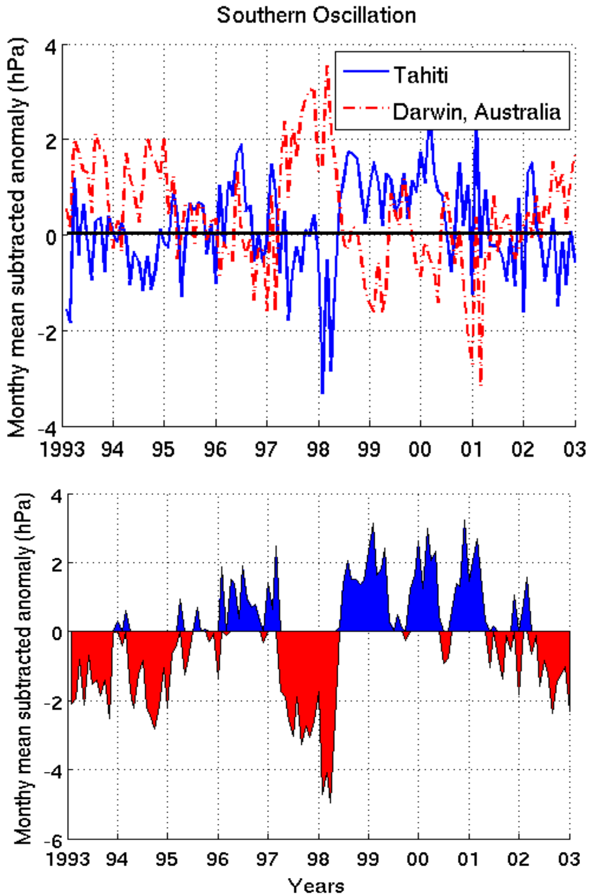


Understanding climate variability using Dipole Analysis

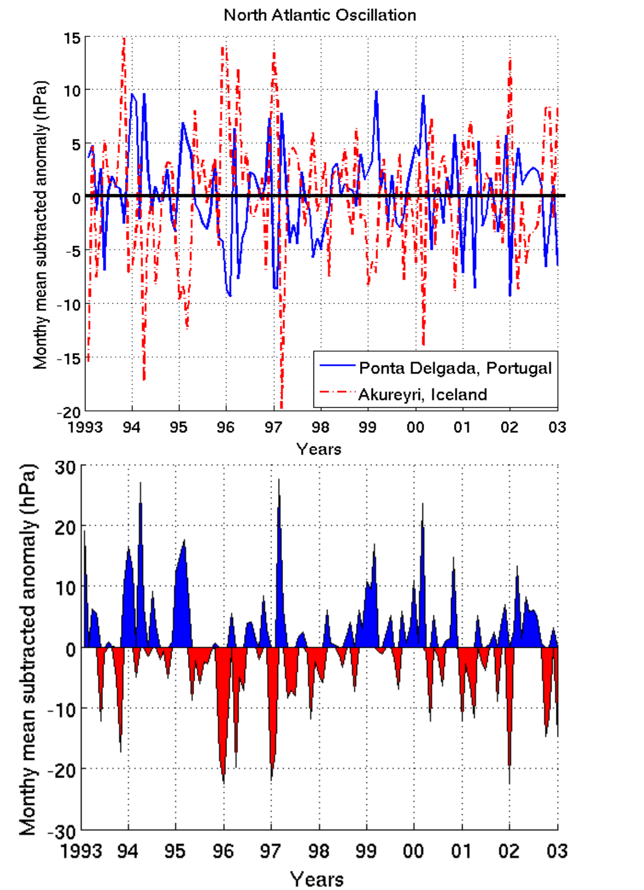


Dipoles represent a class of teleconnections characterized by anomalies of opposite polarity at two locations at the same time.

Southern Oscillation: Tahiti and Darwin



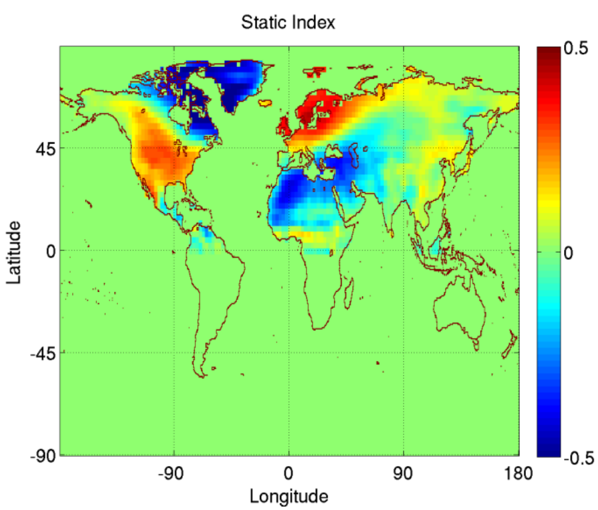
North Atlantic Oscillation: Iceland and Azores

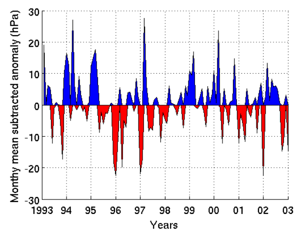


Importance of dipoles

Crucial for understanding the climate system, especially for weather and climate forecast simulations within the context of global climate change.

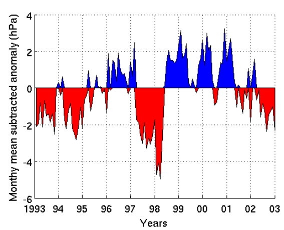
NAO influences sea level pressure (SLP) over most of the Northern Hemisphere. Strong positive NAO phase (strong Islandic Low and strong Azores High) are associated with above-average temperatures in the eastern US.

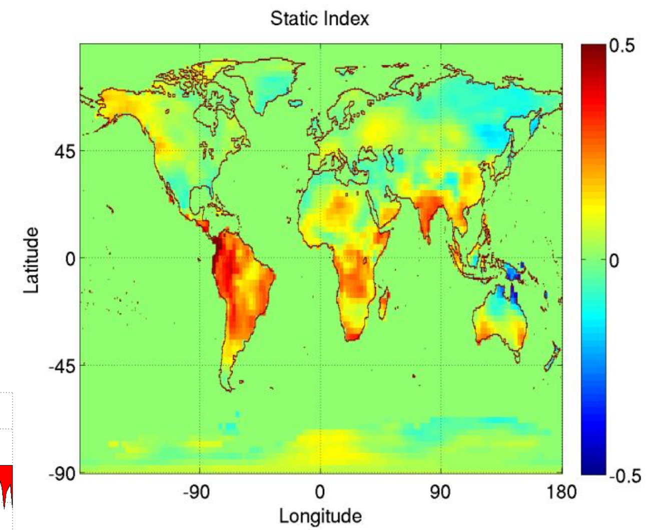




Correlation of Land temperature anomalies with NAO

SOI dominates tropical climate with floodings over East Asia and Australia, and droughts over America. Also has influence on global climate.





Correlation of Land temperature anomalies with SOI

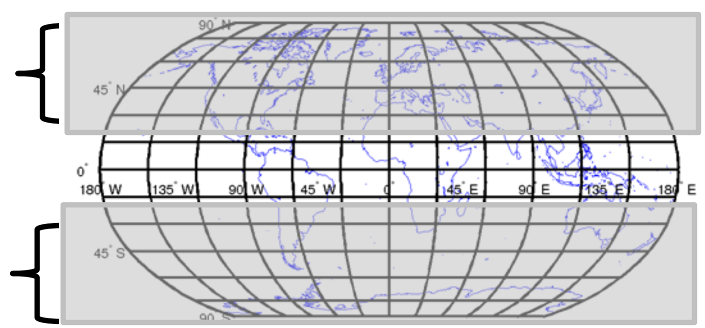
List of Well Known Climate Indices

Index	Description
SOI	**Southern Oscillation Index:** Measures the SLP anomalies between Darwin and Tahiti
NAO	**North Atlantic Oscillation:** Normalized SLP differences between Ponta Delgada, Azores and Stykkisholmur, Iceland
AO	**Arctic Oscillation:** Defined as the first principal component of SLP northward of 20° N
PDO	**Pacific Decadel Oscillation:** Derived as the leading principal component of monthly SST anomalies in the North Pacific Ocean, poleward of 20° N
WP	**Western Pacific:** Represents a low-frequency temporal function of the ‘zonal dipole' SLP spatial pattern involving the Kamchatka Peninsula, southeastern Asia and far western tropical and subtropical North Pacific
PNA	**Pacific North American:** SLP Anomalies over the North Pacific Ocean and the North America
AAO	**Antarctic Oscillation:** Defined as the first principal component of SLP southward of 20° S
NINO1+2	Sea surface temperature anomalies in the region bounded by 80° W-90° W and 0° -10° S
NINO3	Sea surface temperature anomalies in the region bounded by 90° W-150° W and 5° S-5° N
NINO3.4	Sea surface temperature anomalies in the region bounded by 120° W-170° W and 5° S-5° N
NINO4	Sea surface temperature anomalies in the region bounded by 150° W-160° W and 5° S-5° N

Discovered primarily by human observation and EOF Analysis

AO: EOF Analysis of 20N-90N Latitude

AAO: EOF Analysis of 20S-90S Latitude

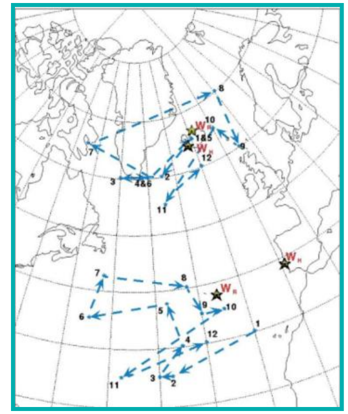


Motivation for Automatic Discovery of Dipoles

- The known dipoles are defined by static locations but the underlying phenomenon is dynamic

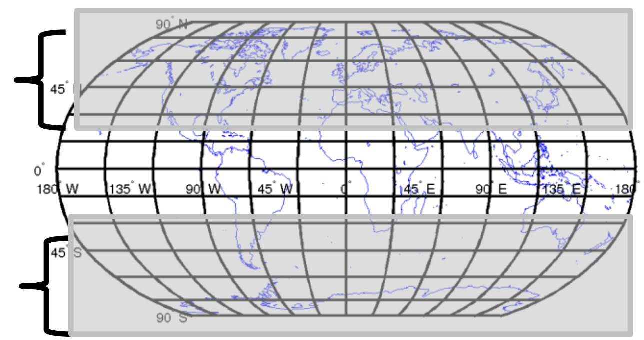
- Manual discovery can miss many dipoles

- EOF and other types of eigenvector analysis finds the strongest signals and the physical interpretation of those can be difficult.



Dynamic behavior of the high and low pressure fields corresponding to NOA climate index (Portis et al, 2001)

AO: EOF Analysis of 20N-90N Latitude

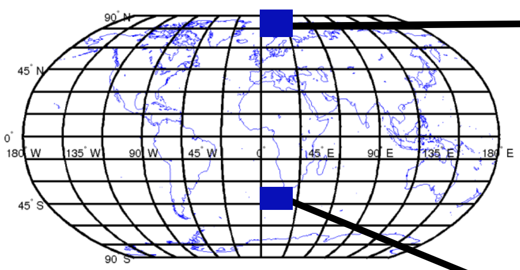


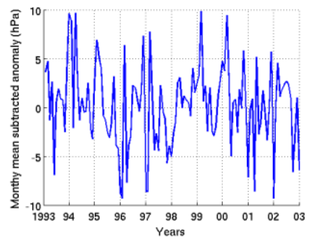
AAO: EOF Analysis of 20S-90S Latitude

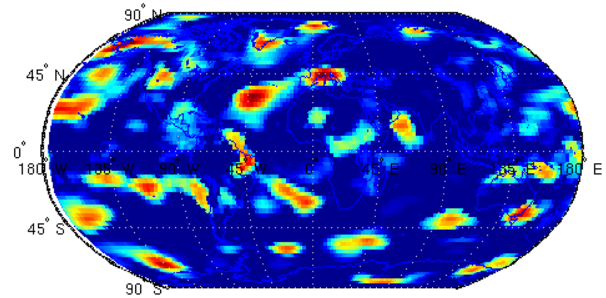
Discovering Climate Teleconnections using Network Representation

Shared Reciprocal Nearest Neighbors (SRNN) Density

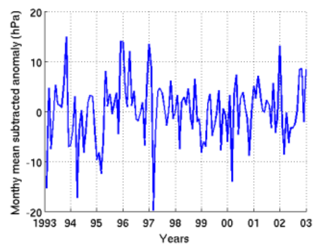
Climate Network



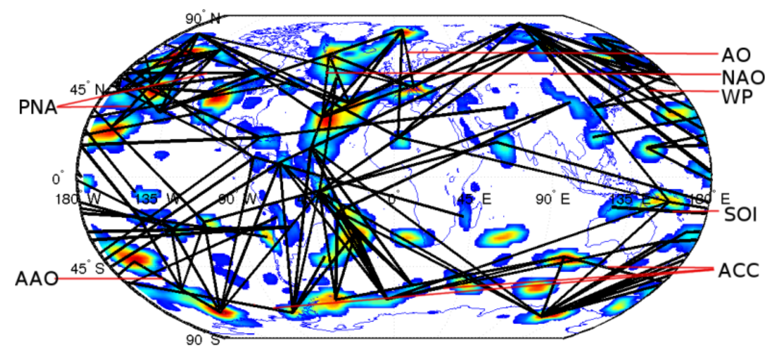




Nodes in the Graph correspond to grid points on the globe.



NCEP



Edge weight corresponds to correlation between the two anomaly timeseries

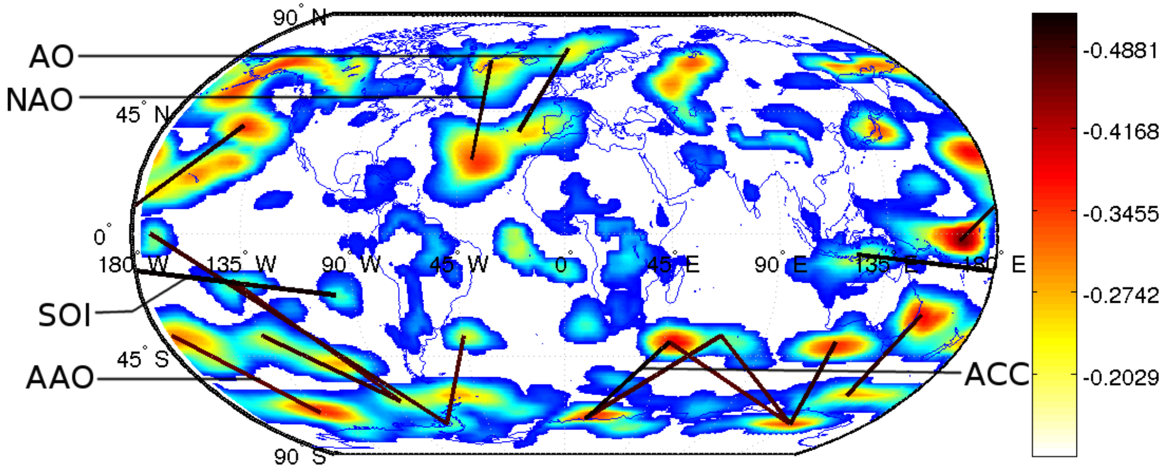
Dipoles from SRNN density

Benefits of Automatic Dipole Discovery

- Detection of Global Dipole Structure
  - Most known dipoles discovered
  - New dipoles may represent previously unknown phenomenon.
  - Enables analysis of relationships between different dipoles
- Location based definition possible for some known indices that are defined using EOF analysis.
- Dynamic versions are often better than static
- Dipole structure provides an alternate method to analyze GCM performance

Detection of Global Dipole Structure

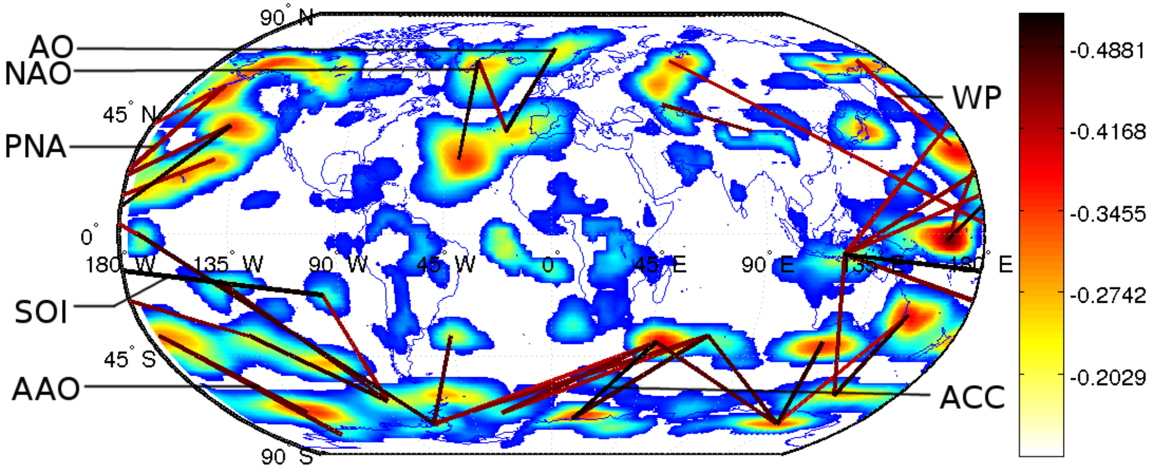
NCEP (National Centers for Environmental Prediction) Reanalysis Data



- Most known dipoles discovered
- Location based definition possible for some known indices that are defined using EOF analysis.
- New dipoles may represent previously unknown phenomenon.

Detection of Global Dipole Structure

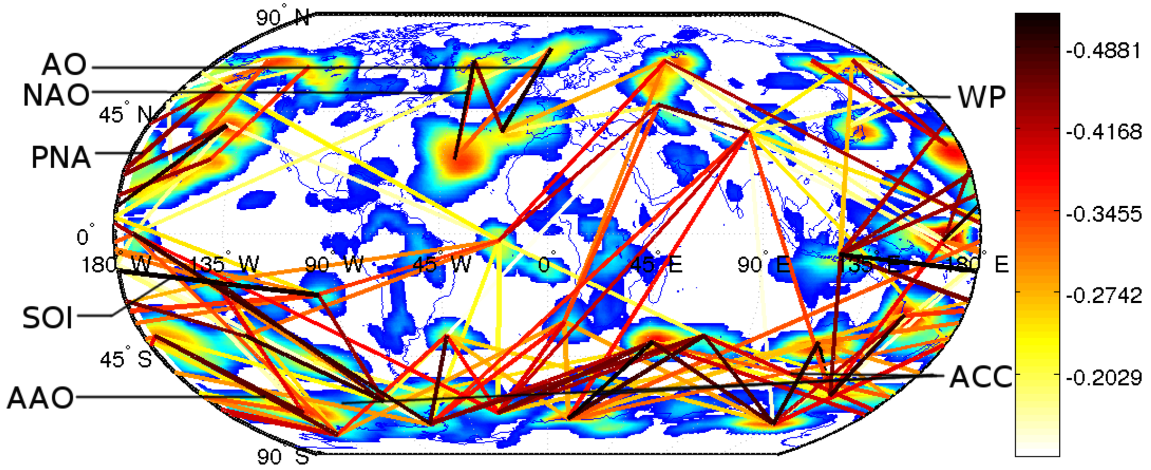
NCEP (National Centers for Environmental Prediction) Reanalysis Data



- Most known dipoles discovered
- Location based definition possible for some known indices that are defined using EOF analysis.
- New dipoles may represent previously unknown phenomenon.

Detection of Global Dipole Structure

NCEP (National Centers for Environmental Prediction) Reanalysis Data



- Most known dipoles discovered
- Location based definition possible for some known indices that are defined using EOF analysis.
- New dipoles may represent previously unknown phenomenon.

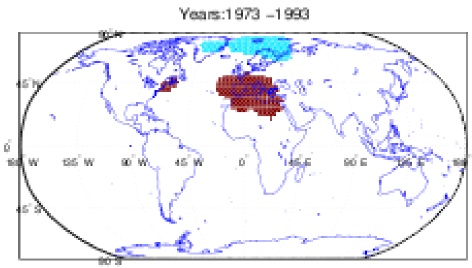
Results: Most known Dipoles discovered

- Correlation with Static Indices: This shows that the dynamic dipoles are a good representative of the static ones.

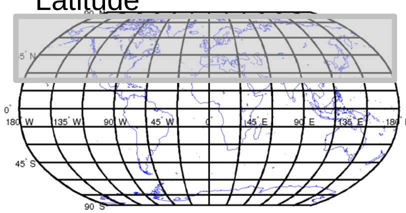
Table 3: Correlation of our dynamic indices with known climate indices (K=300)

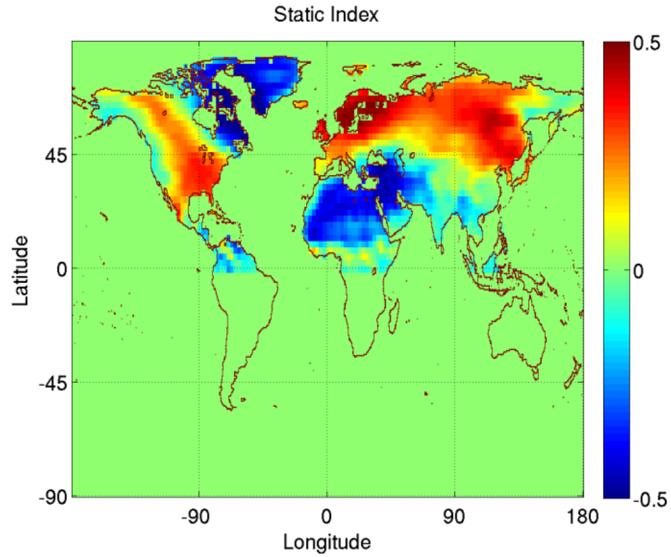
Network	A1					A1 + community				
	SOI	NAO	AO	AAO	WP	SOI	NAO	AO	AAO	WP
1	0.8885	0.7686	0.8665	-	0.7166	0.8761	0.7686	0.8665	-	0.7163
2	0.8696	0.7729	0.8506	-	0.7231	0.7378	0.7711	0.8529	-	0.7232
3	0.9012	0.7312	0.8560	-	0.7400	0.8952	0.7317	0.8580	-	0.7399
4	0.8895	0.8044	0.8353	-	0.7306	0.8828	0.8043	0.8353	-	0.7298
5	0.8983	0.7279	0.8037	-	0.7523	0.8540	0.7283	0.8037	-	0.5017
6	0.9214	0.7498	0.8648	-	0.7602	0.9195	0.7488	0.8702	-	0.7408
7	0.8387	0.7769	0.8137	-	0.7604	0.8318	0.7819	0.8137	-	0.7318
8	0.8946	0.7581	0.8407	0.8763	0.7240	0.8933	0.7582	0.8407	0.8797	0.4369
9	0.8746	0.7609	0.8597	0.8835	0.7103	0.8737	0.7621	0.8597	0.8809	0.7095
Mean	0.8863	0.7612	0.8434	0.8799	0.7353	0.8627	0.7608	0.8445	0.8803	0.6642

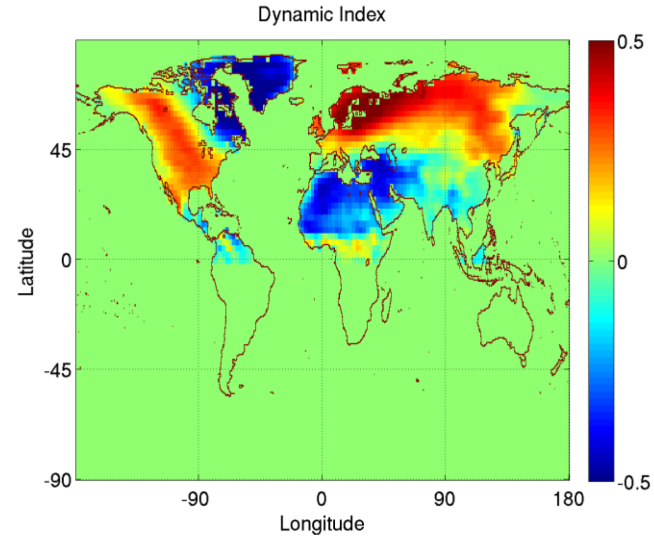
Results: Location Based definition AO



Static AO: EOF Analysis of 20N-90N Latitude



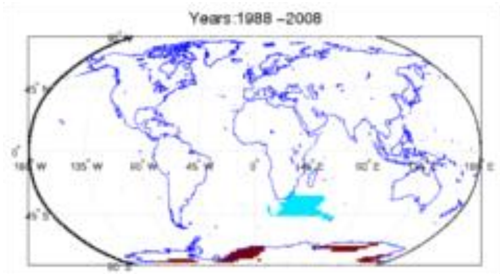




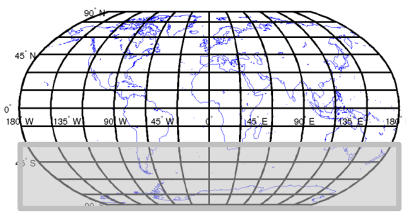
Impact on Land temperature Anomalies using Static and Dynamic AO

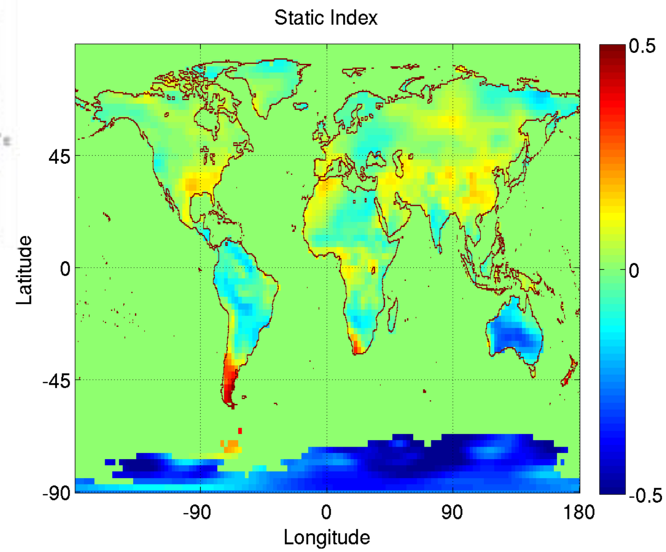
- Mean Correlation between static and dynamic index: 0.84
- Impact on land temperature anomalies comparatively same using static and dynamic index

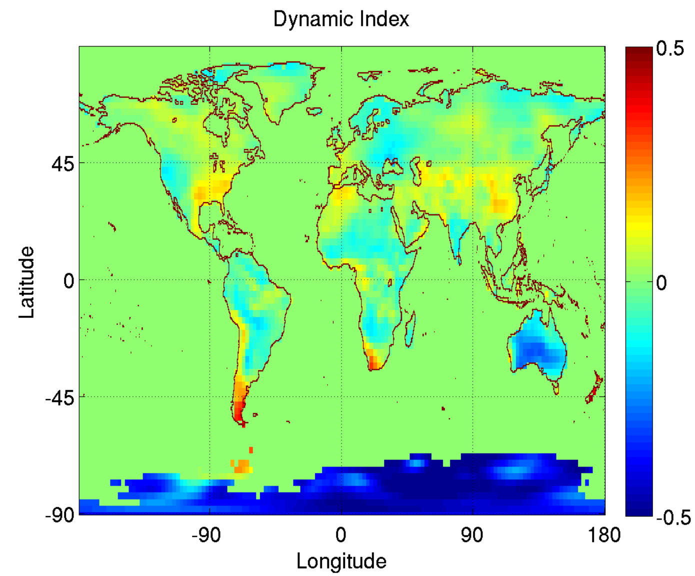
Results: Location Based definition AAO



Static AAO: EOF Analysis of 20S-90S Latitude



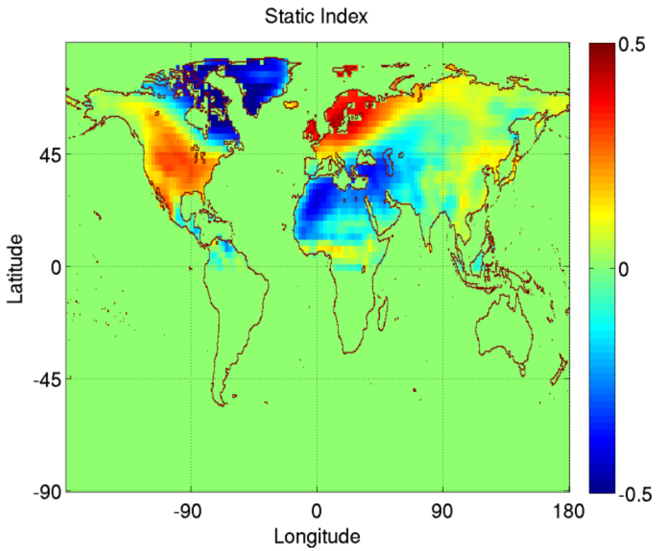


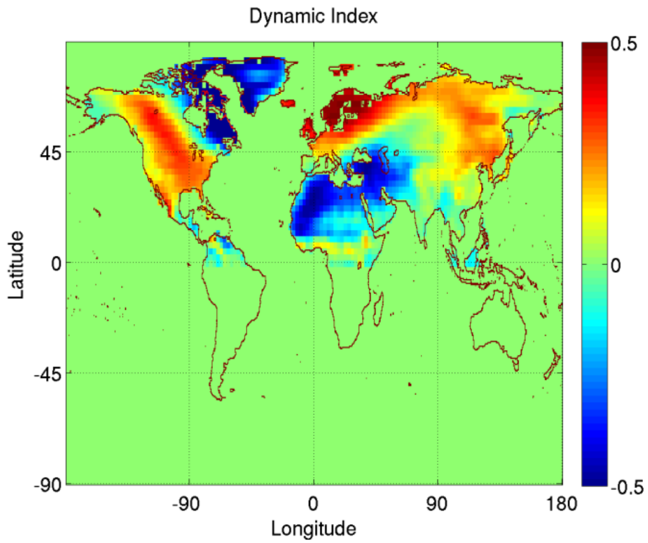


Impact on Land temperature Anomalies using Static and Dynamic AAO

- Mean Correlation between Static and Dynamic index = 0.88
- Impact on land temperature anomalies comparatively same using static and dynamic index

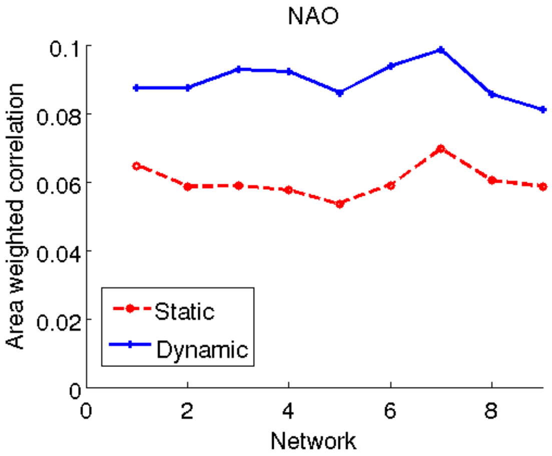
Static vs Dynamic NAO Index: Impact on land temperature



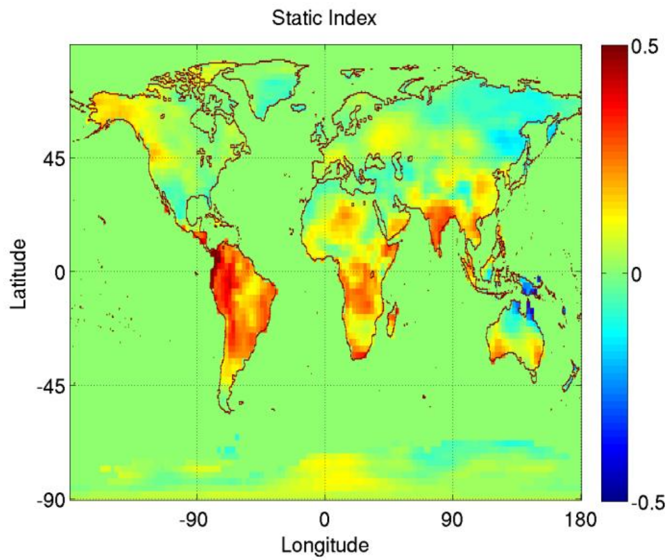


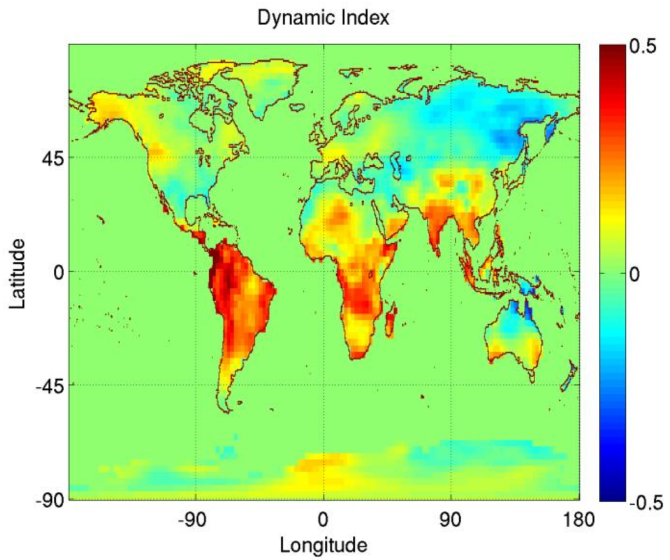
The dynamic index generates a stronger impact on land temperature anomalies as compared to the static index.

Figure to the right shows the aggregate area weighted correlation for networks computed for different 20 year periods during 1948-2008.



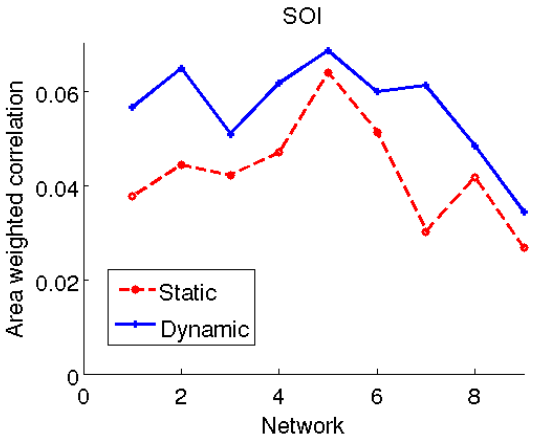
Static vs Dynamic SO Index: Impact on land temperature



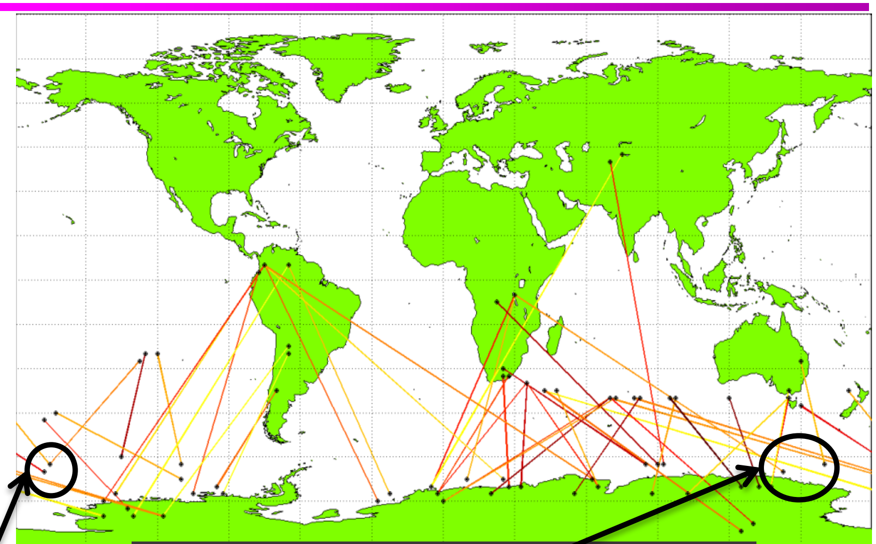


The dynamic index generates a stronger impact on land temperature anomalies as compared to the static index.

Figure to the right shows the aggregate area weighted correlation for networks computed for different 20 year periods during 1948-2008.

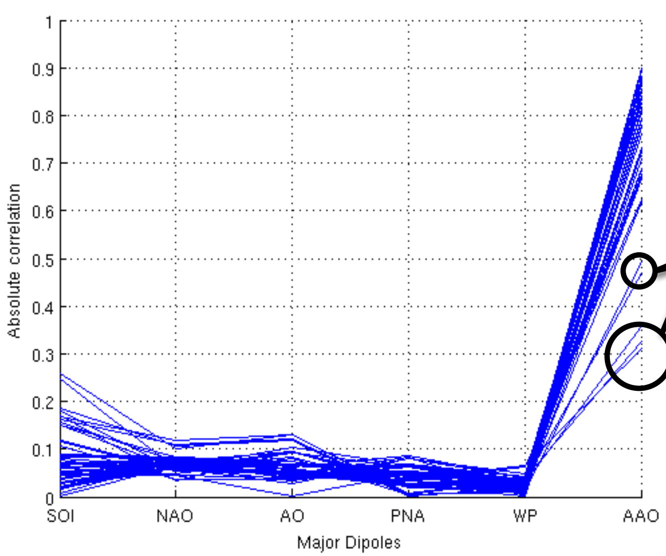


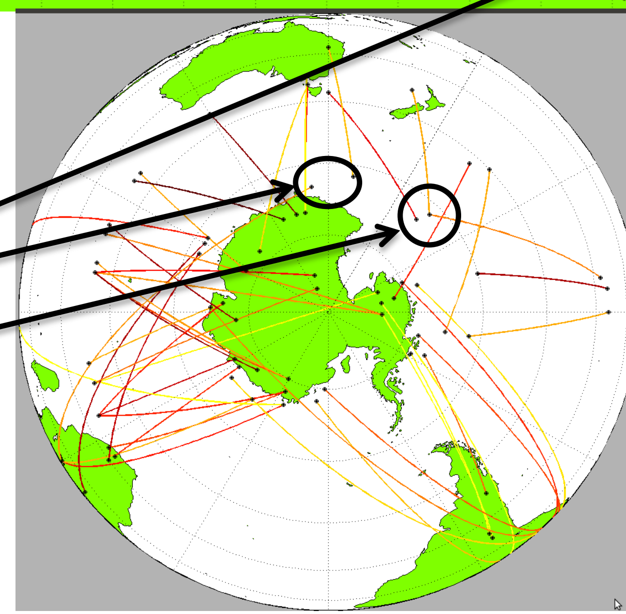
A New Dipole Around Antarctica?



- 3 major dipole structures can be seen.
- The AAO and two others shown in figure
- A newer phenomenon which is not captured by the EOF analysis?

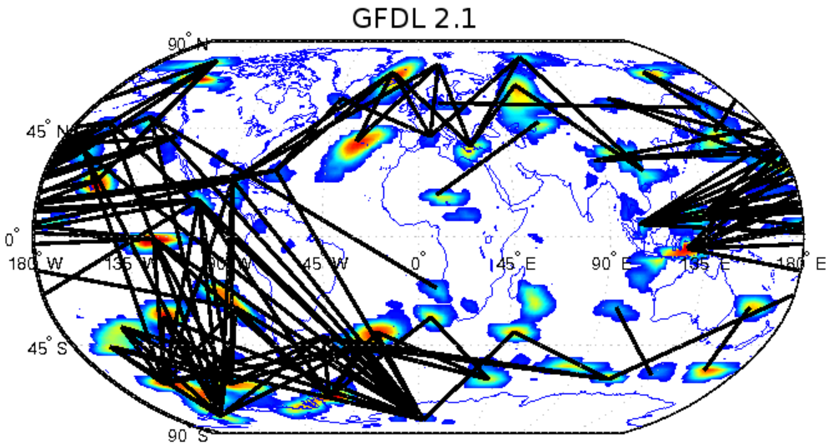
Correlation plot with major dipoles

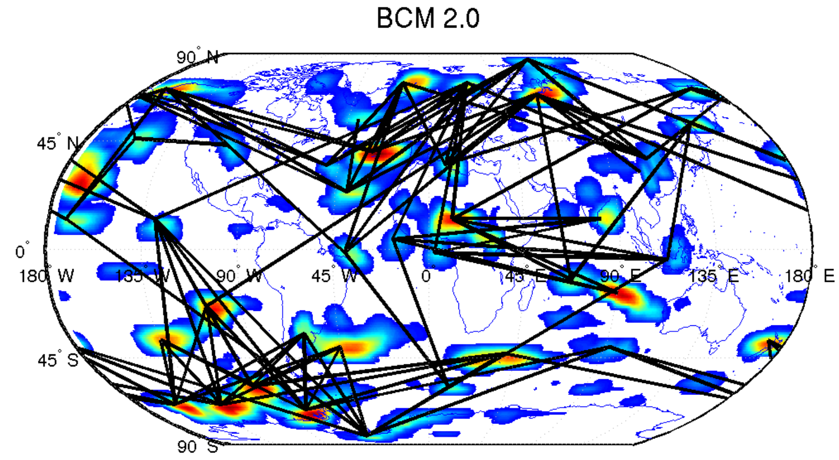




Comparison of Climate Models using Dipole Structure

Hindcast data



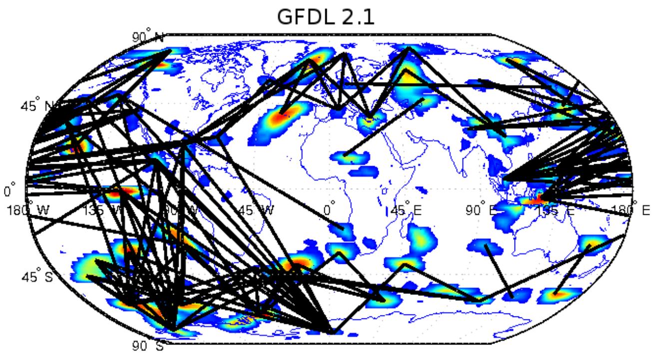


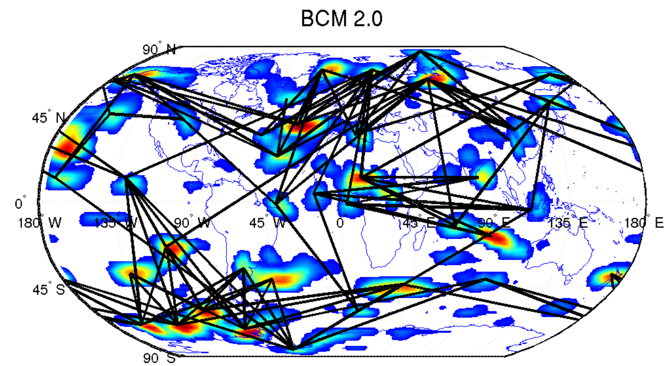
- Differences in dipole structure can offer valuable insights to climate scientists on model performance

- Strength of the dipoles varies in different climate models
  - SOI is only simulated by GFDL 2.1 and not by BCM 2.0.

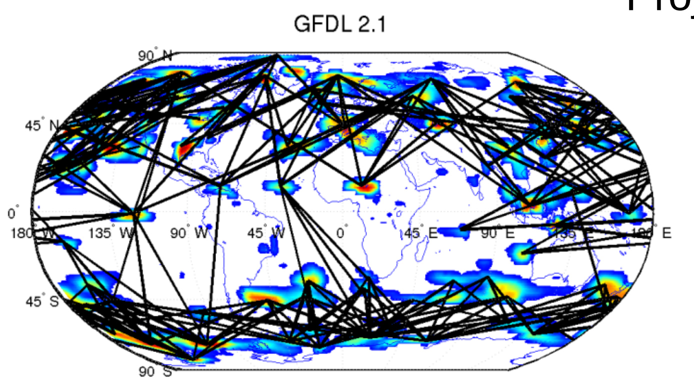
Comparison of Climate Models using Dipole Structure

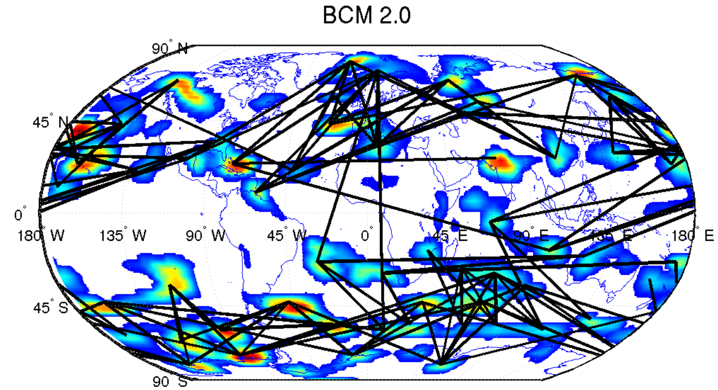
Hindcast data





Projection data





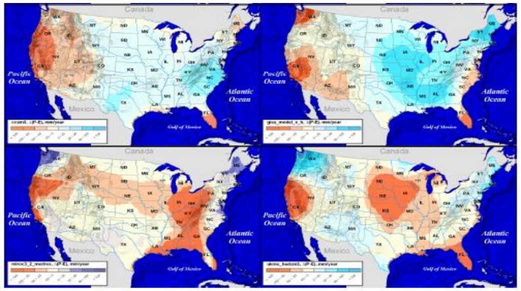
- Dipole connections in forecast data provide insights about dipole activity in future.

- For e.g. both forecasts for 2080-2100 show continuing dipole activity in the extratropics but decreased activity in the tropics. SOI activity is reduced in GFDL2.1 and activity over Africa is reduced in BCM 2.0. This is consistent with archaeological data from 3 mil. years ago, when climate was 2-3°C warmer (Shukla, et. al).

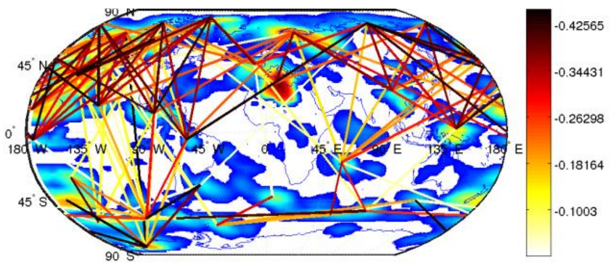
Relating Dipole Structure to Model Prediction

- The dipole structure of the top 2 models are different from the bottom two models
  - NCAR-CCSM and NASA-GISS miss SOI and other dipoles near the Equator

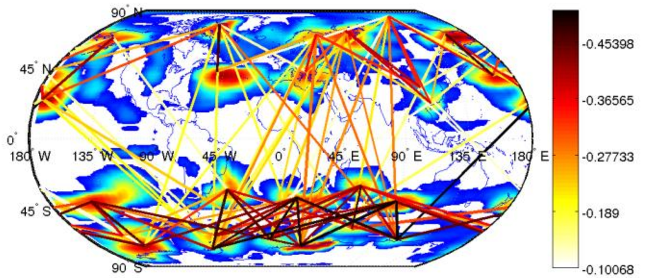
Disagreement between IPCC models



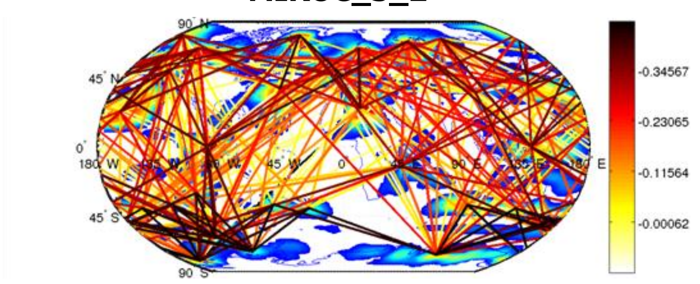
NCAR-CCSM



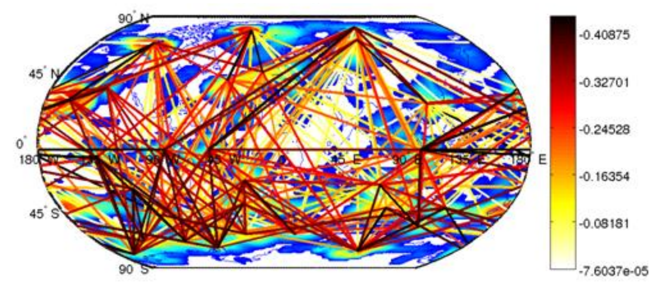
NASA-GISS



MIROC_3_2

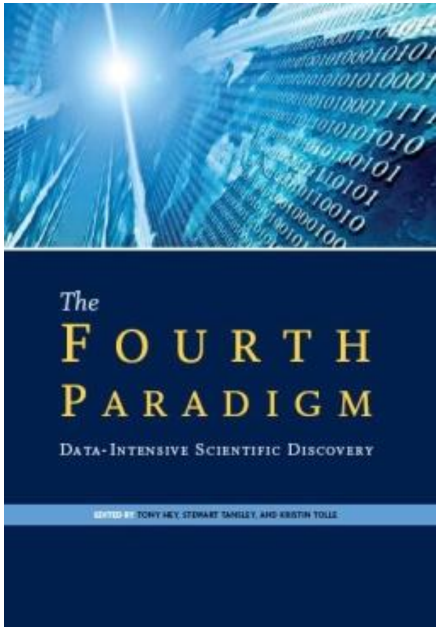


UKMO-HadCM3



Summary

- Data driven discovery methods hold great promise for advancement in the mining of climate and eco-system data.
- Scale and nature of the data offer numerous challenges and opportunities for research in mining large datasets.



"*The world of science has changed … data-intensive science [is] so different that it is worth distinguishing [it] … as a new, fourth paradigm for scientific exploration.*" - **Jim Gray**

References

- Kawale, J, Steinbach M, Kumar V, 2011. Discovering dynamic dipoles in climate data, SDM 2011

- Steinbach, M., Tan, P., Kumar, V., Klooster, S., and Potter, C. 2003. Discovery of climate indices using clustering. In SIGKDD international Conference on Knowledge Discovery and Data Mining. KDD, pg. 446-455, 2003

- Tsonis, A. A., Swanson, K. L., Roebber, P. J. What Do Networks Have to Do with Climate. In Bulletin of the American Meteorological Society, vol. 87, no. 5, pg. 585-595, 2006

- Donges, J. F., Zou, Y., Marwan, N. Complex networks in climate dynamics. In European Physical Journal Special Topics, 174 (1), pp. 157-179, 2006

- Steinhaeuser, K., Chawla, N. V., Ganguly, A. R. An exploration of climate data using complex networks. KDD Workshop on Knowledge Discovery from Sensor Data, pp. 23-31, 2009

- Portis, D. H., Walsh, J. E., El Hamly, Mostafa and Lamb, Peter J., Seasonality of the North Atlantic Oscillation, Journal of Climate, vol. 14, pg. 2069-2078, 2001.