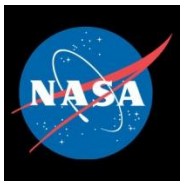
NSF Expeditions in Computing

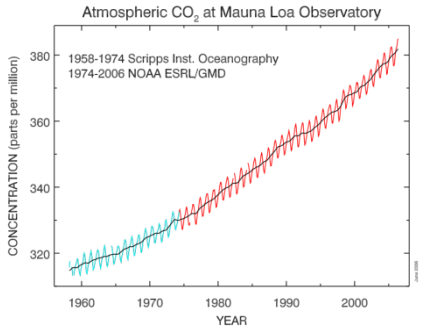
Understanding Climate Change: A Data Driven Approach

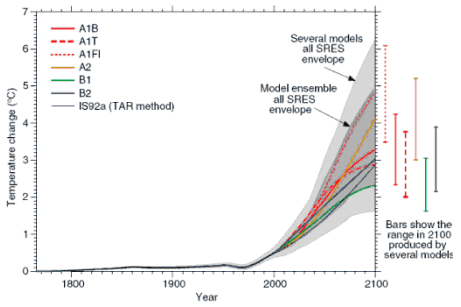
Vipin Kumar
University of Minnesota
kumar@cs.umn.edu
www.cs.umn.edu/~kumar



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



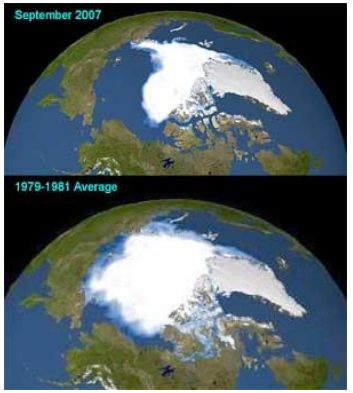


- **Consequences can be dire**
  - Extreme weather events, regional climate and ecosystem shifts, abrupt climate change, stress on key resources and critical infrastructures

- **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



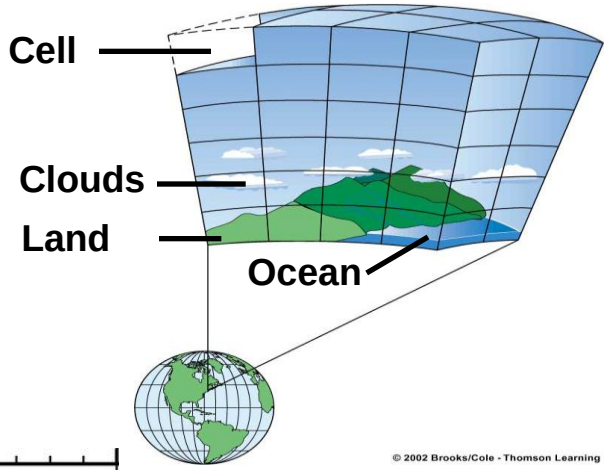
The Vanishing of the Arctic Ice cap
ecology.com, 2008

Key outstanding science challenge:
Actionable predictive insights to credibly inform policy

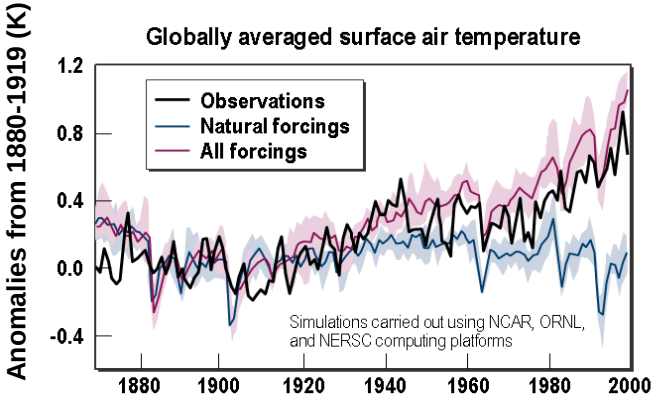
Understanding Climate Change - Physics based Approach

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

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

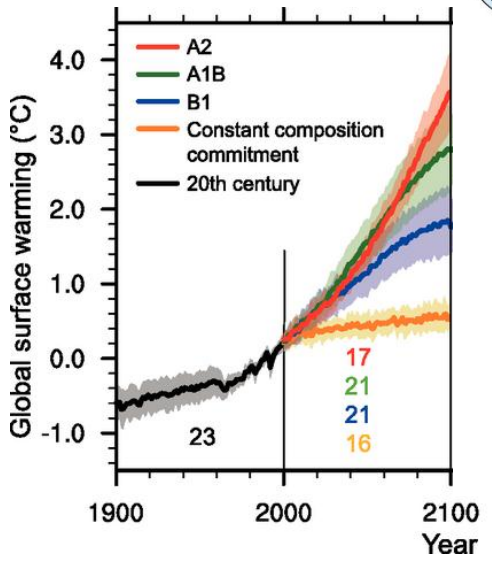


Temperature increases are human-induced
The anthropogenic climate change "fingerprint"



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

Figure Courtesy: ORNL



IPCC (2007)

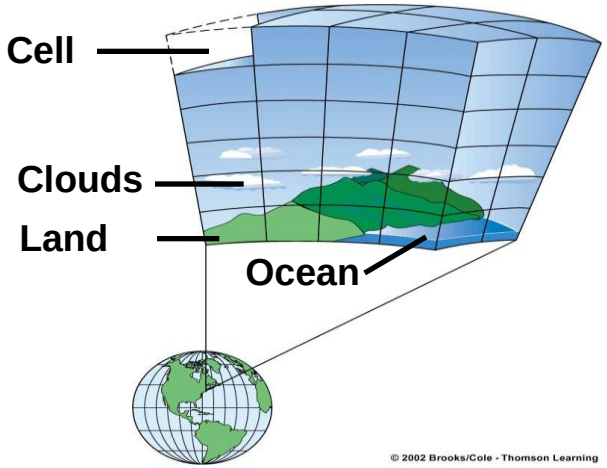
Projection of temperature increase under different **Special Report on Emissions Scenarios** (SRES) by 24 different GCM configurations from 16 research centers used in the **Intergovernmental Panel on Climate Change** (IPCC) 4th Assessment Report.

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

Understanding Climate Change - Physics based Approach

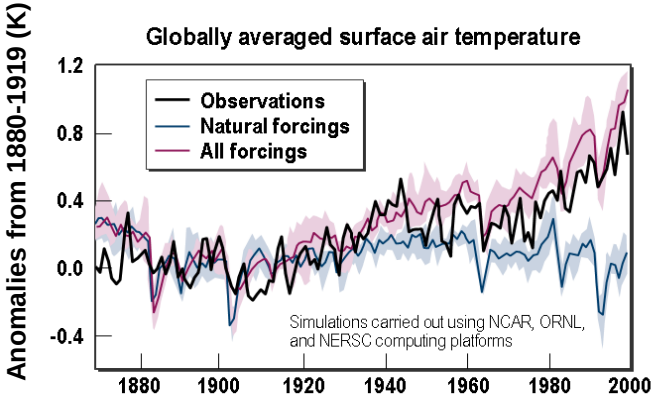
General Circulation Models: Mathematical models with physical equations based on fluid dynamics

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

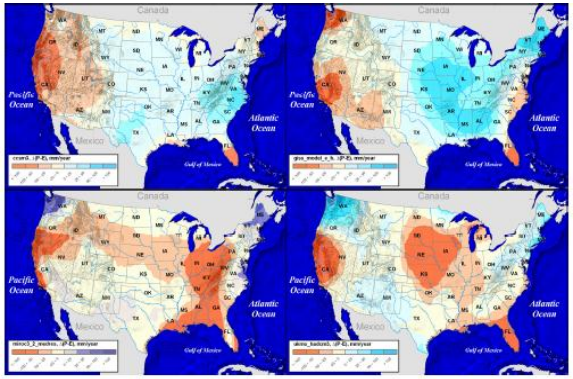
Figure Courtesy: ORNL

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

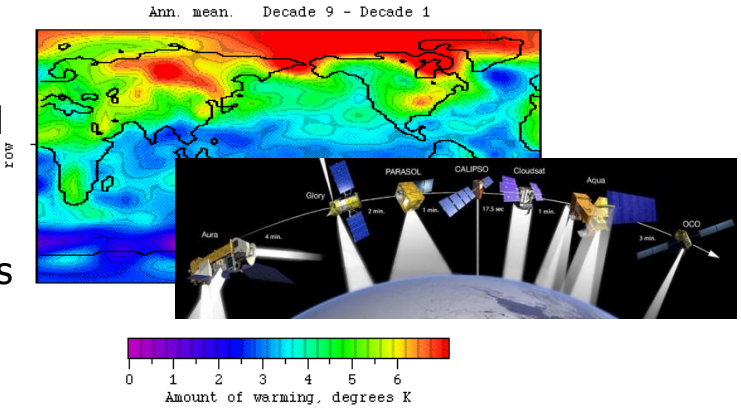


Regional hydrology exhibits large variations among major IPCC model projections

Data-Driven Knowledge Discovery in Climate Science

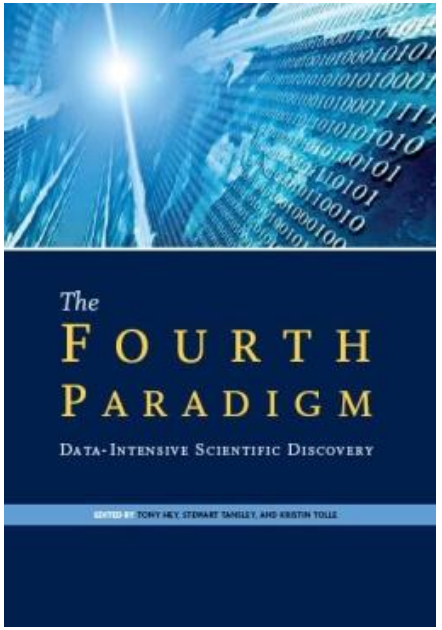
From data-poor to data-rich transformation

- **Sensor Observations**: Remote sensors like satellites and weather radars as well as in-situ sensors and sensor networks like weather station and radiation measurements
- **Model Simulations**: IPCC climate or earth system models as well as regional models of climate and hydrology, along with observed data based model reconstructions



- Data guided processes can complement hypothesis guided data analysis to develop predictive insights for use by climate scientists, policy makers and community at large.

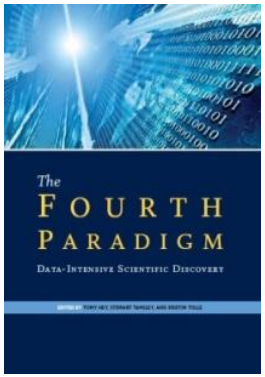
"***The world of science has changed … data-intensive science [is] so different that it is worth distinguishing [it] … as a new, fourth paradigm for scientific exploration***." - **Jim Gray**



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)

Transformative Computer Science Research

Enabling large-scale data-driven science for complex, multivariate, spatio-temporal, non-linear, and dynamic systems:

The Expedition is an end-to-end demonstration of this major paradigm for future knowledge discovery process.

- **Fusion plasma**
- **Combustion**
- **Astrophysics**
- **....**

Complex Networks
Enable studying of collective behavior of interacting eco-climate systems

Nonlinear, space-time lag, geographical, multi-scale

Relationship Mining
Enable discovery of complex dependence structures such as non linear associations or long range spatial dependencies

Nonlinear, spatio-temporal, multi-variate, persistence, long memory

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

Nonlinear, spatio-temporal, multivariate

relationships

Community structure-function-dynamics

kernels , features, dependencies

High Performance Computing
Enable efficient large-scale spatio-temporal analytics on future generation exascale HPC platforms with complex memory hierarchies

Large scale, spatio-temporal, unstructured, dynamic

CS and Climate Science Synergies

- ***Physics based models inform data mining***
  - More credible variables (e.g., SST) → more crucial variables (e.g., precipitations/hurricanes)
  - Better modeled processes (e.g., atmospheric physics) → more crucial processes (e.g., land surface hydrology)
  - Global and century scale → regional and decadal scale

- ***Data mining informs physics based models***
  - Better understanding of climate processes
  - Improved insights for parameterization schemes

Measures of Success in CS Research

Does the proposed research enable significantly better data analysis than before or enable new kinds of analysis?

- **Predictive Modeling**
  - Improved capabilities for multivariate spatio-temporal regression that can simultaneously capture the dependencies between spatial-temporal objects (e.g., temperature, pressure) and help find novel climate patterns not found by standard approaches
  - New capabilities for detecting extreme events using quantiles and quantile regression for multivariate data

- **Relationship Mining**
  - Creation of entirely new capabilities in association mining from approaches that extend / modify traditional approaches to handle spatio-temporal data
  - Creation of a completely novel approach for association analysis that reduces the number of patterns and the time required to find them
  - Improvement in the performance of complex networks and predictive models by using nonlinear relationships that more faithfully represent reality

- **Complex Networks**
  - Detection of known climate patterns and tracking of the evolution of climate patterns in time using complex networks that capture non-linear relationships in multivariate and multiscale data

- **High Performance Computing**
  - Scalable analytics code for spatio-temporal data
  - Enabling of large scale data driven science that serves as a demonstration of the value of the data driven paradigm

Climate-based Measures of Success

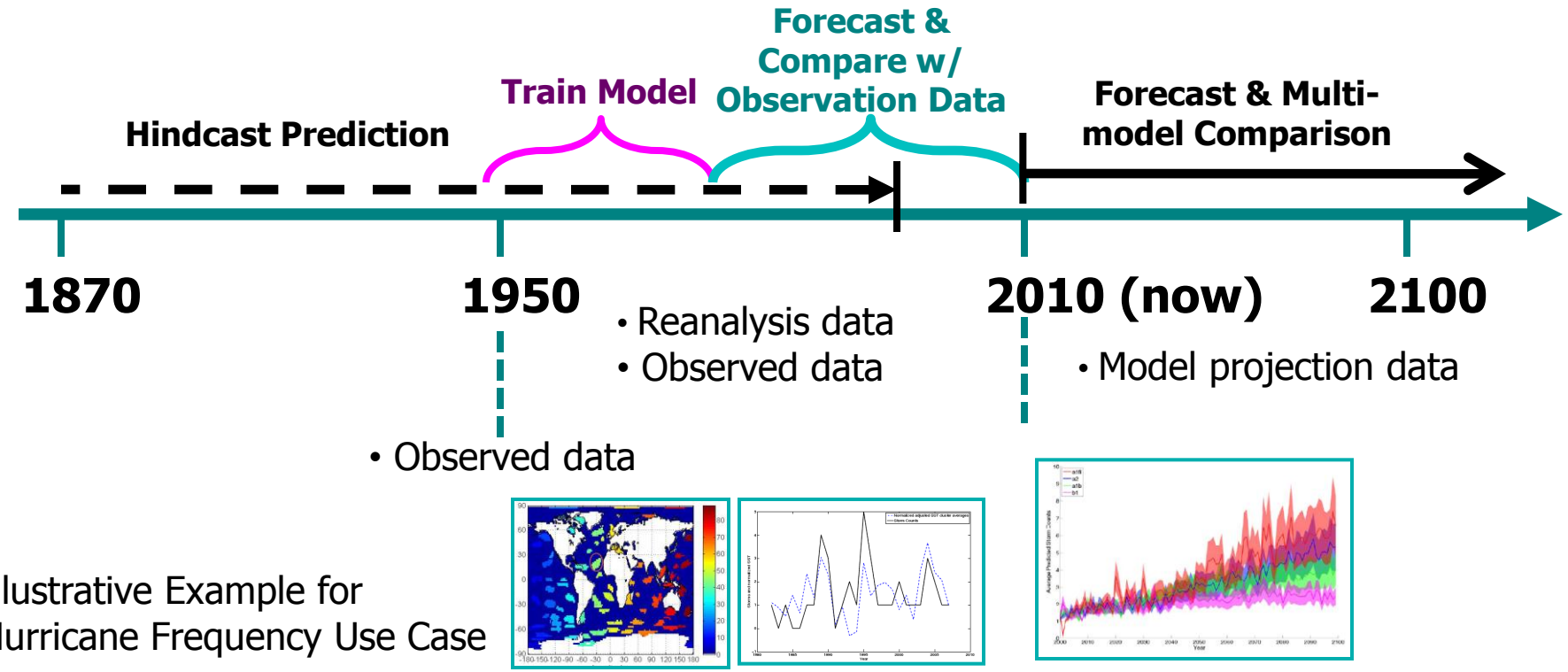
- Ultimate success is to have data-driven analysis included as a standard part of climate projections and impact assessment (e.g., for IPCC).

- Achieving this will require measuring and demonstrating success in two key areas
  - **Filling critical gaps in climate science** (*Nature*, 2010)
    - Reduction of uncertainty in climate predictions at regional and decadal scales
    - Improved predictive insights for key precipitation processes
    - Providing credible assessments of extreme hydro-meteorological events
    - New knowledge about climate processes (e.g., teleconnection patterns)
  - **Improve climate change impact assessments and inform policy**
    - Distinguishing between natural and anthropogenic causes
    - Improved assessments of climate change risks across multiple sectors

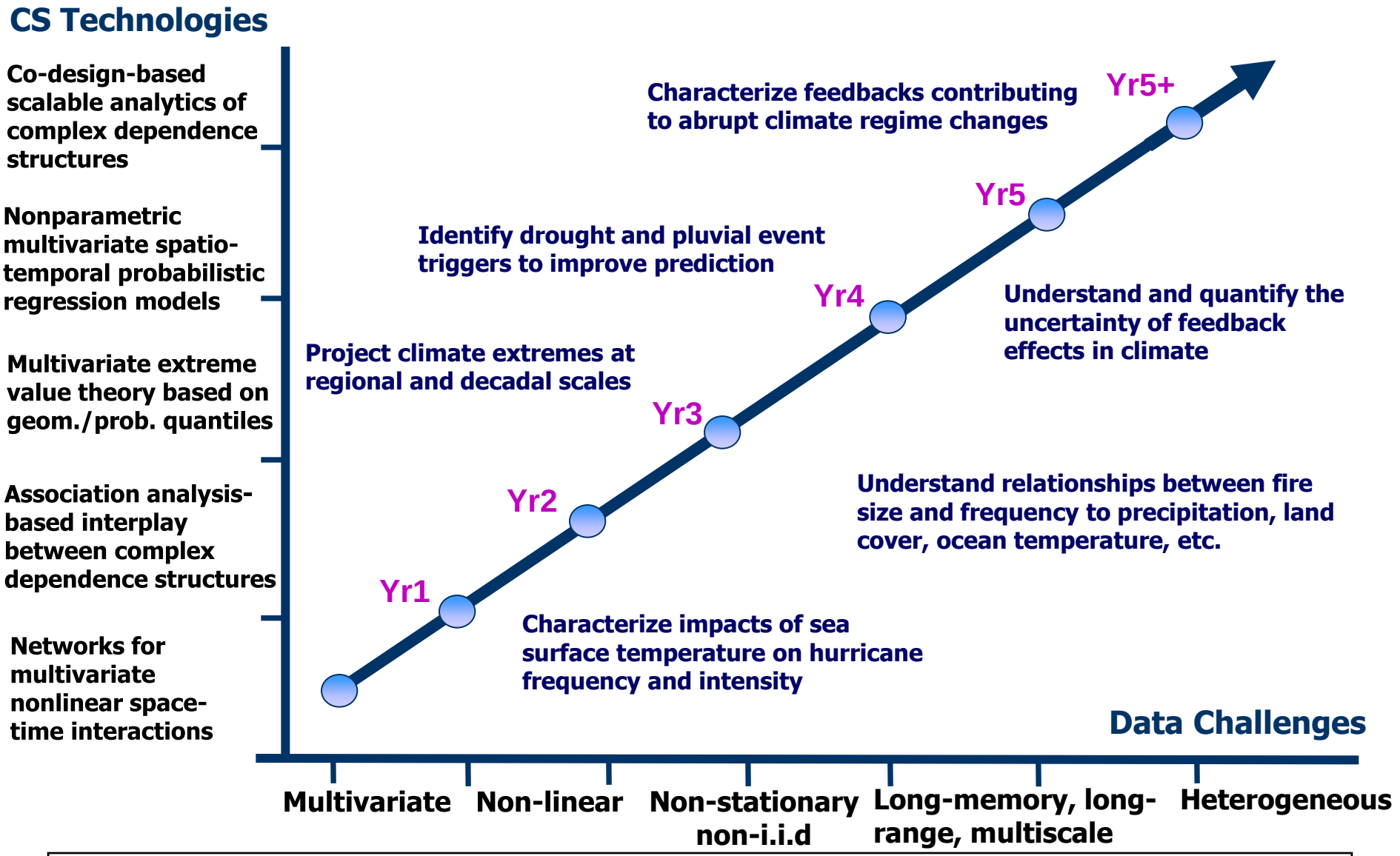
- Specific use cases will provide the context for these evaluations

Science Impact View: Evaluation Methodology

- Can we improve the projection (e.g., hurricane frequency, intensity)?
- Can the data-driven model identify:
  - which regional climate variables are informative/causal (e.g., SST, wind speed)?
  - what is the relationship (+/- feedback) between these variables?
- To what extent does the data-driven model inform climate scientists?

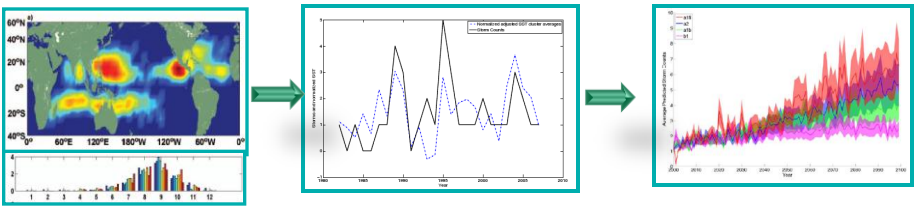


Conceptual View: Driving Use-Cases



Example Driving Use Cases

Impact of Global Warming on Hurricane Frequency



Find non-linear relationships

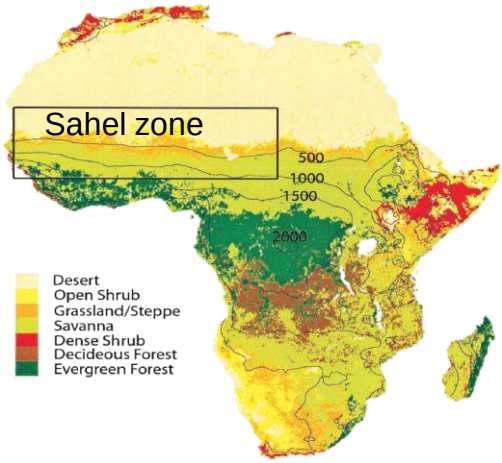
Validate w/ hindcasts

Build hurricane models

Regime Shift in Sahel

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

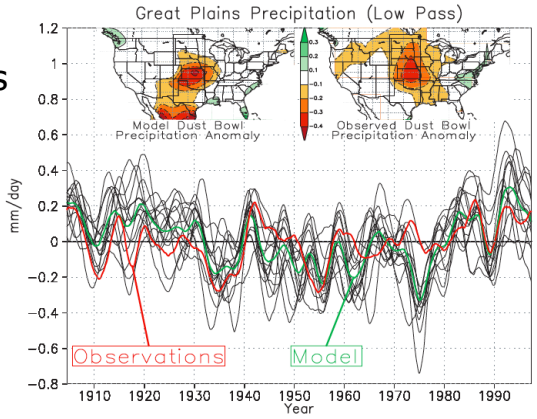
Regime shift can occur without any advanced warning and may be triggered by isolated events such as storms, drought



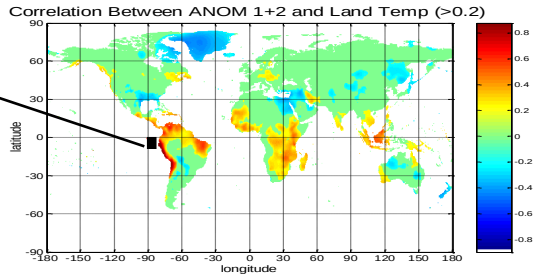
1930s Dust Bowl

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



Goals for the Annual Workshop

- Engage wider community in dialogue
  - Share research results and learn about related efforts
  - Identify important / relevant problems in climate science that need to be addressed using data driven methods
  - Build acceptance for data analytic methods in climate change science
- Build Collaborations
- Encourage students and the next generation of researchers to “think big” about problems relevant to society