INTRODUCING THE US CLIVAR DATA SCIENCES WORKING GROUP

For scientifically focused discussion on emerging tools

Mike Pritchard
Associate Professor
University of California, Irvine

Road map: The Data Science WG

I. Motivation.

Who we are, context, the goal.

II. Three themes.

1. How should we change modeling practices?
2. What is potential for data-driven discovery (patterns, predictability)?
3. Learn & talk - which methods are achieving breakthrough potential?

III. Objectives, timeline, upcoming events.

Webinars, collecting tools & data, how to get involved.

Road map: The Data Science WG

→ I. Motivation.

Who we are, context, the goal.

II. Three themes.

1. How should we change modeling practices?
2. What is potential for data-driven discovery (patterns, predictability)?
3. Learn & talk - which methods are achieving breakthrough potential?

III. Objectives, timeline, upcoming events.

Webinars, collecting tools & data, how to get involved.

Road map: The Data Science WG

I. Motivation.

→ Who we are, context, the goal.

II. Three themes.

1. How should we change modeling practices?
2. What is potential for data-driven discovery (patterns, predictability)?
3. Learn & talk - which methods are achieving breakthrough potential?

III. Objectives, timeline, upcoming events.

Webinars, collecting tools & data, how to get involved.

Co-conspirators.

Amy Braverman

Principal Statistician
NASA / JPL

Elizabeth Barnes

Associate Professor
Colorado State U.

Pierre Gentine

Associate Professor
Columbia U.

CLIVAR RESEARCH

How does climate vary and
change on multiple timescales?

Analysis of observations

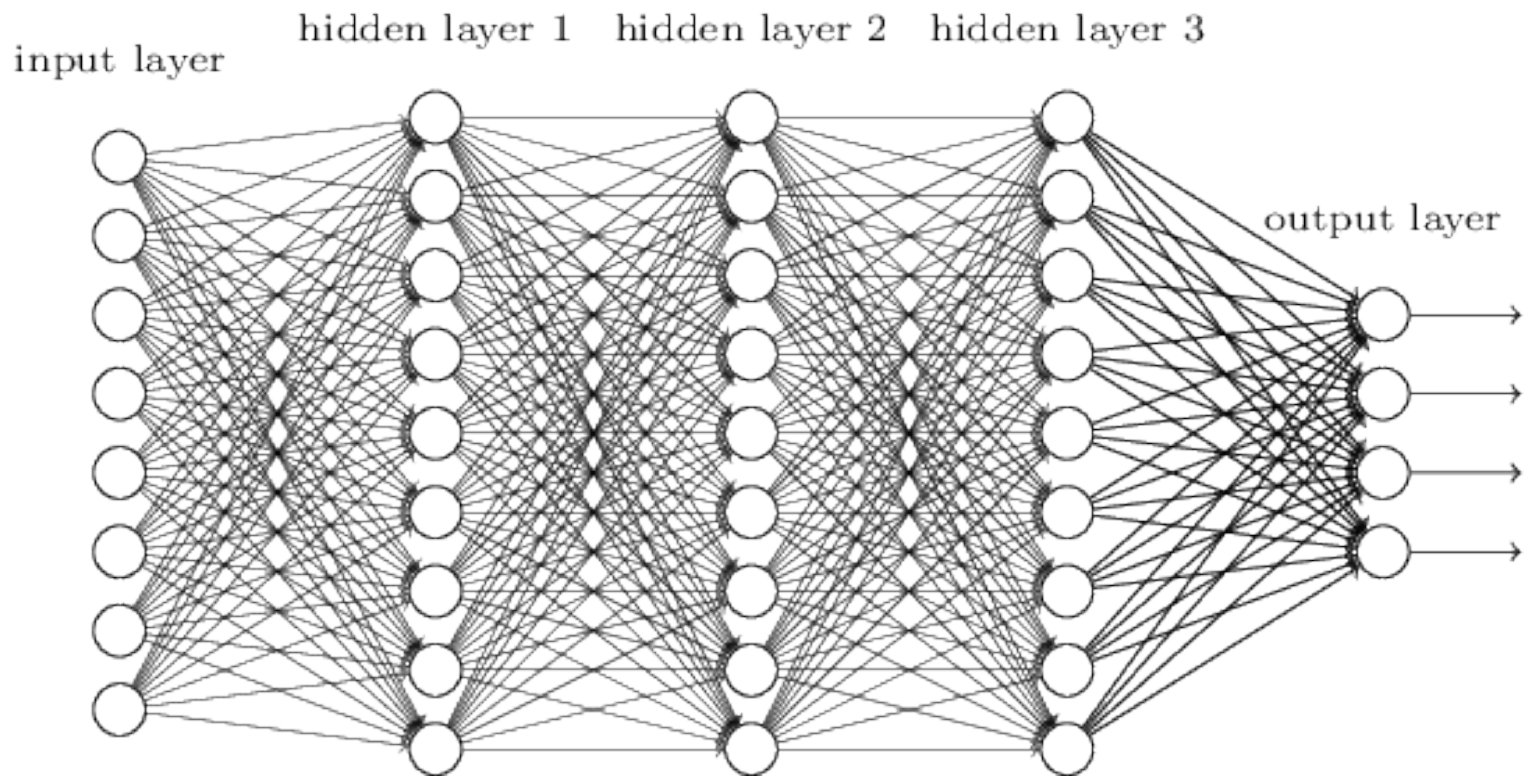
Modeling of observations

Exploding information.

Analysis of observations

Modeling of observations

New computational tools to meet the challenge.



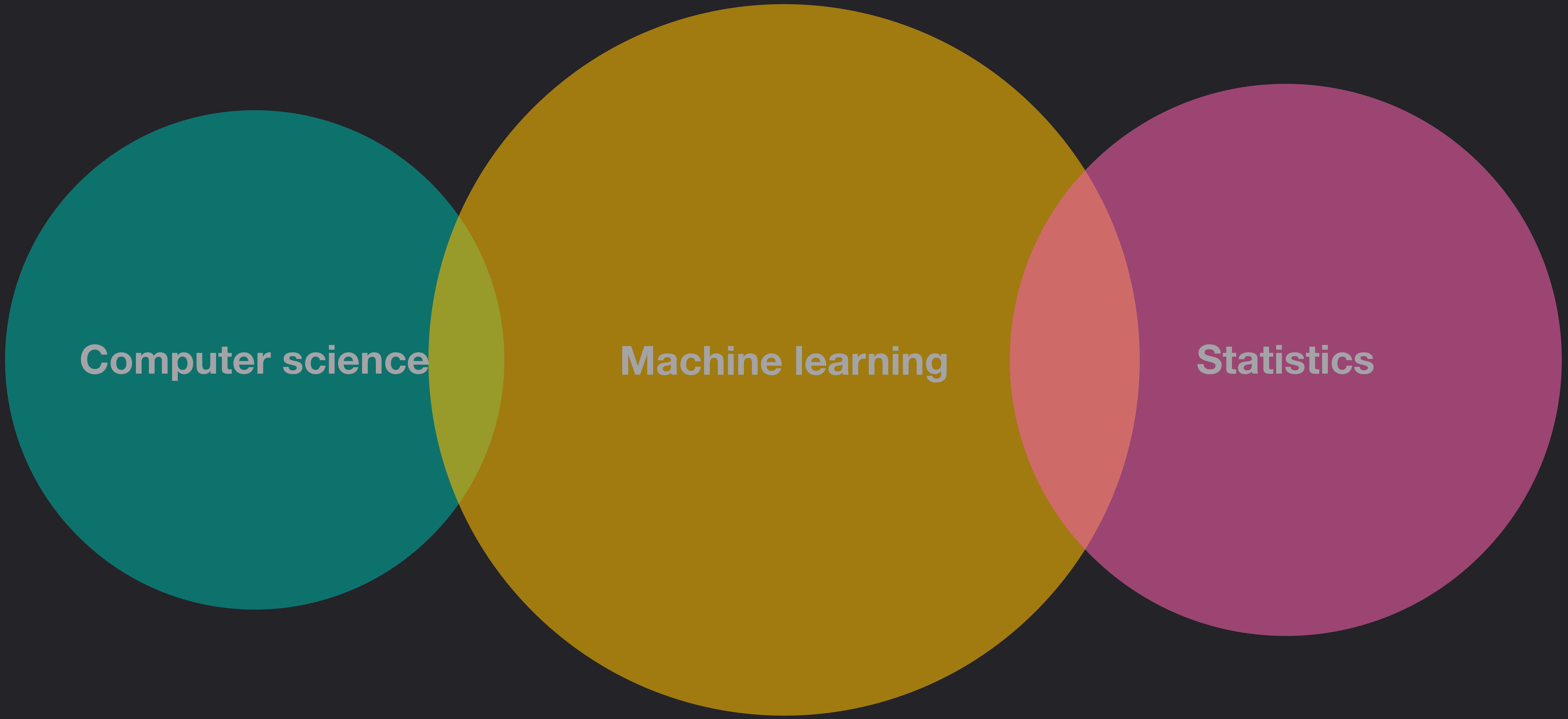
Example: “Deep Neural Networks” — powerful emulators of high-dimensional nonlinear functions disrupting industry and science.

Evidence that data sciences are transforming engineering, science & the economy.

- Near-human-level image classification
- Near-human-level speech recognition
- Near-human-level handwriting transcription
- Improved machine translation
- Improved text-to-speech conversion
- Digital assistants such as Google Now and Amazon Alexa
- Near-human-level autonomous driving
- Improved ad targeting, as used by Google, Baidu, and Bing
- Improved search results on the web
- Ability to answer natural-language questions
- Superhuman Go playing

Example: Deep Neural Networks have driven “breakthroughs .. in historically difficult areas of machine learning”

Behind the tools are new methodologies & algorithms.



This is what our team means by "data science"

NEW TOOLS,
NEW SCIENCE?

Discovery of relationships and processes in large datasets that may have gone unnoticed?

Computationally efficient emulation of physical models?

MAIN GOAL

To foster...

Understanding

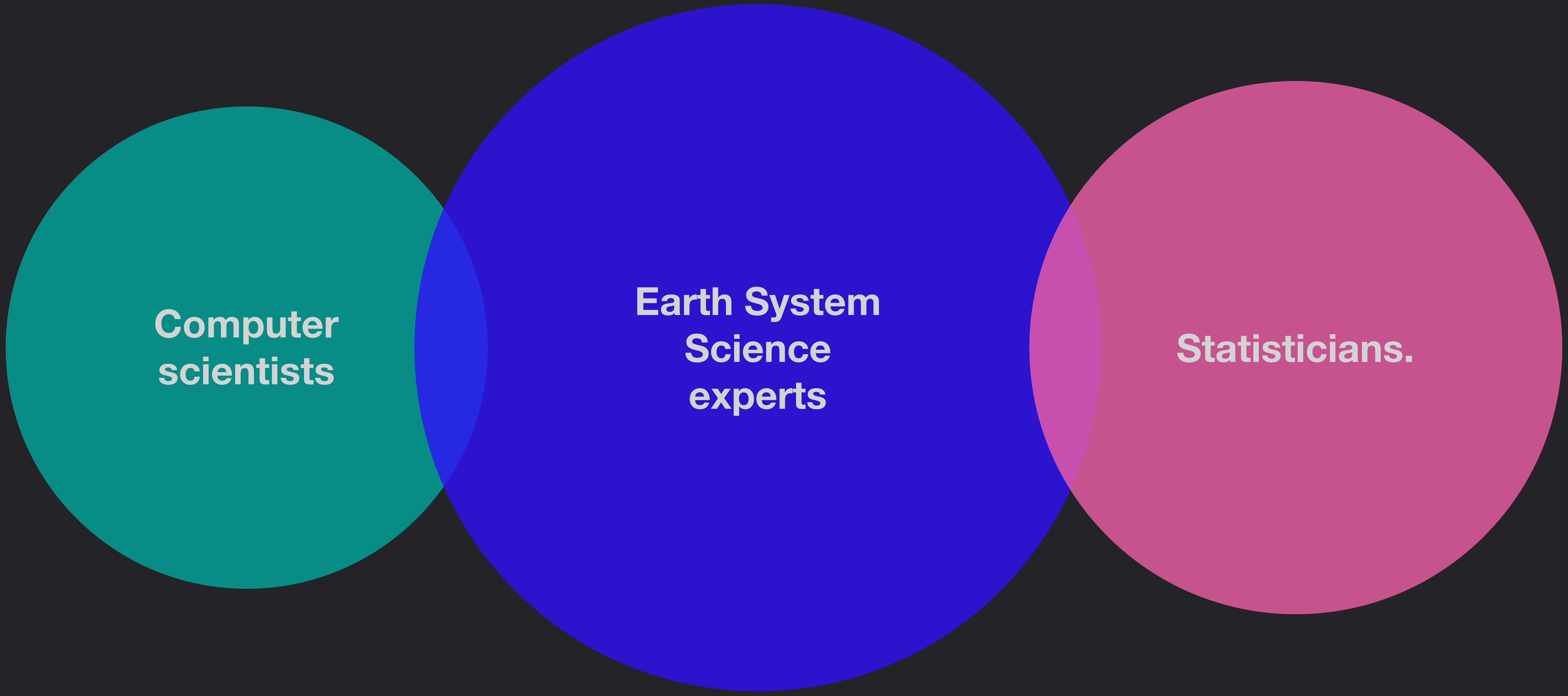
Adoption

Development

Of modern data science tools

in ways that advance CLIVAR science.

The WG will try to unite:



Road map: The Data Science WG

I. Motivation.

→ Who we are, context, the goal.

II. Three themes.

1. How should we change modeling practices?
2. What is potential for data-driven discovery (patterns, predictability)?
3. Learn & talk - which methods are achieving breakthrough potential?

III. Objectives, timeline, upcoming events.

Webinars, collecting tools & data, how to get involved.

Road map: The Data Science WG

I. Motivation.

Who we are, context, the goal.

→ II. Three themes.

1. How should we change modeling practices?
2. What is potential for data-driven discovery (patterns, predictability)?
3. Learn & talk - which methods are achieving breakthrough potential?

III. Objectives, timeline, upcoming events.

Webinars, collecting tools & data, how to get involved.

Road map: The Data Science WG

I. Motivation.

Who we are, context, the goal.

II. Three themes.

→ 1. How should we change modeling practices?

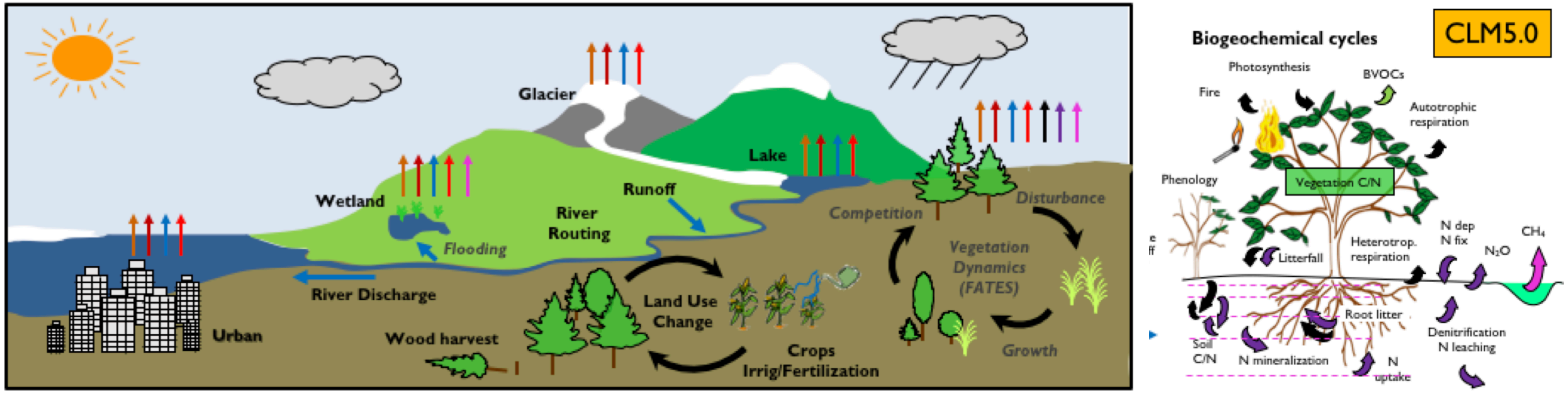
2. What is potential for data-driven discovery (patterns, predictability)?

3. Learn & talk - which methods are achieving breakthrough potential?

III. Objectives, timeline, upcoming events.

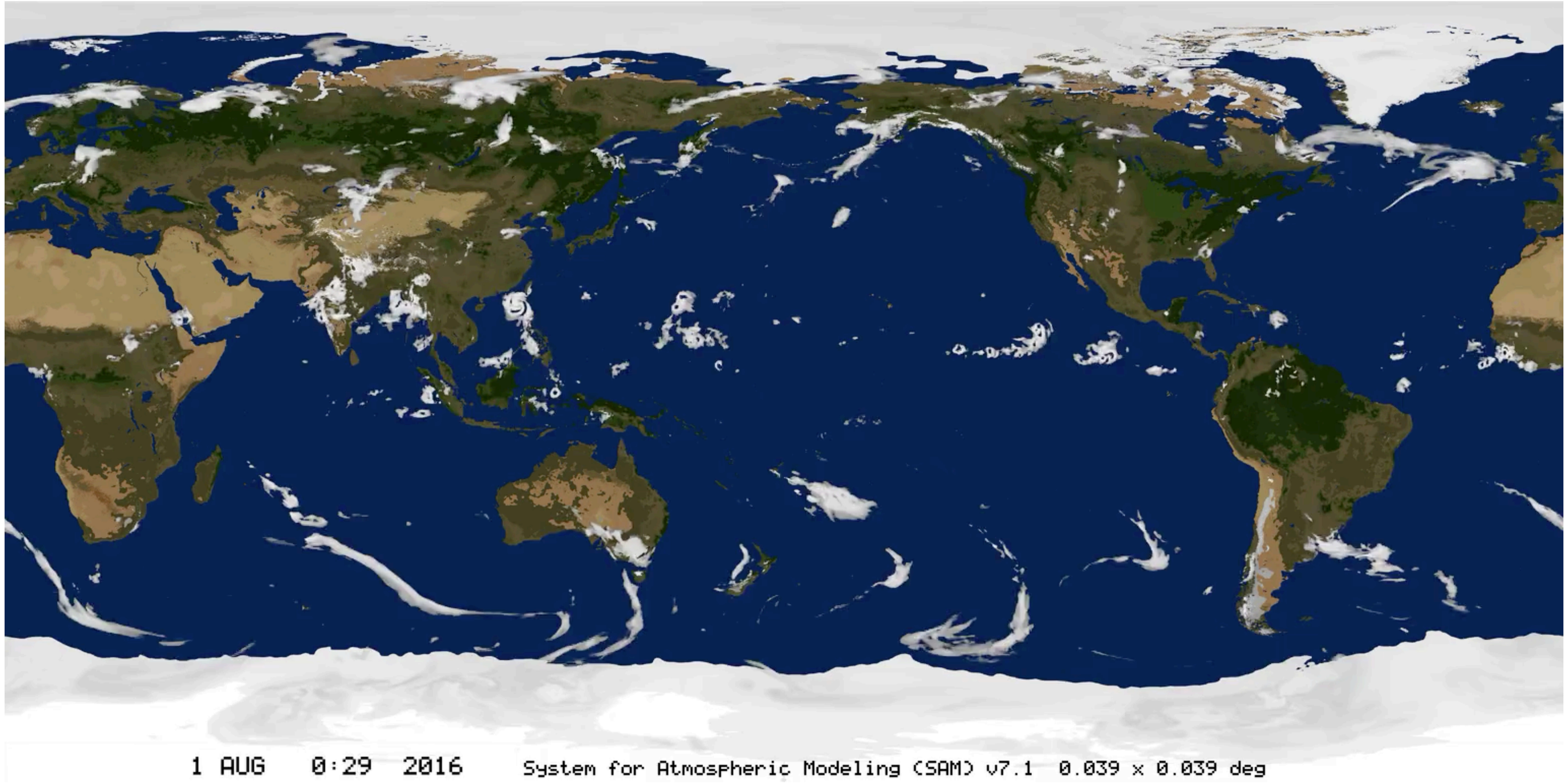
Webinars, collecting tools & data, how to get involved.

Some areas of the climate system have weak physical constraints but rich data constraints.



Example: Terrestrial or oceanic biosphere modeling

Other areas are becoming rich in quality synthetic data.



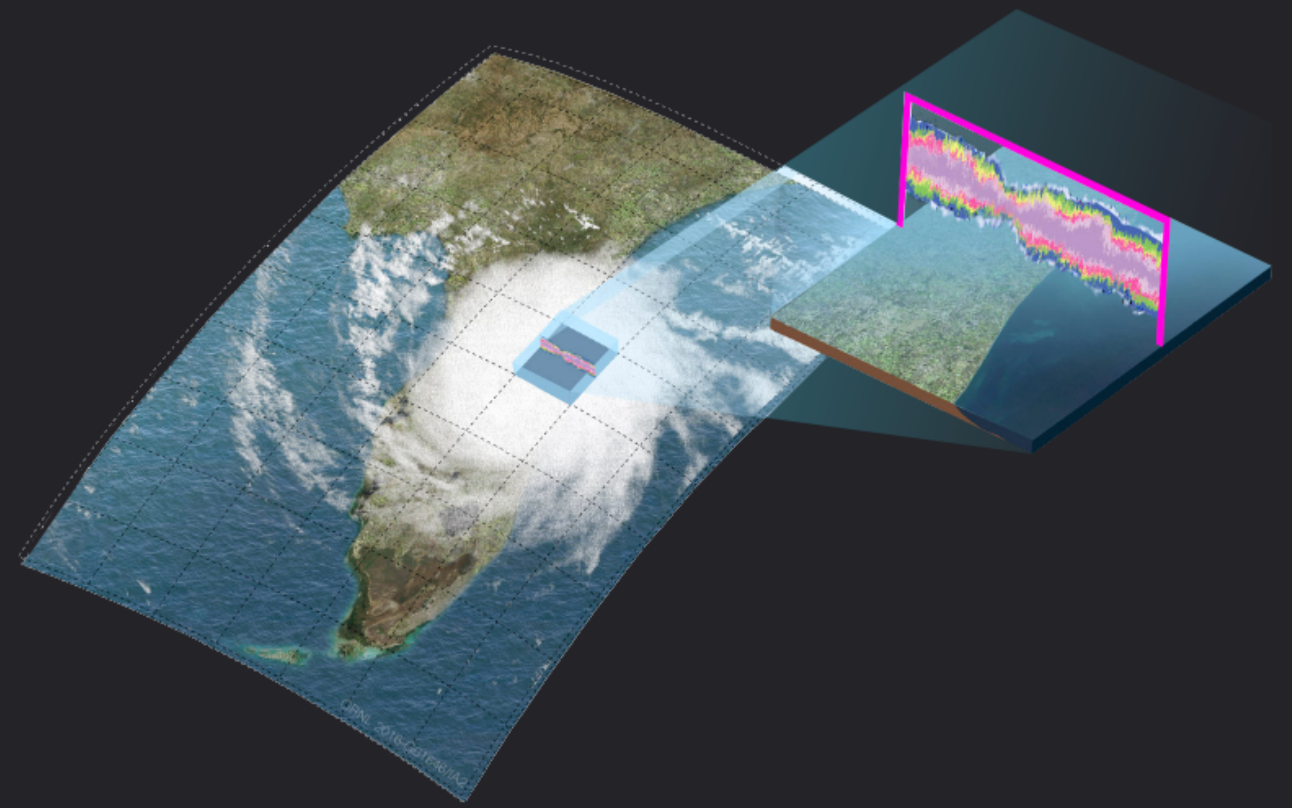
Example: Global Cloud-Resolving model output.

Animation by Marat Khairoutdinov, Stony Brook U. (System for Atmospheric Modeling v7.1)
1 sim-day integrates in 4 hours using ~ 4,000 NCAR processors.

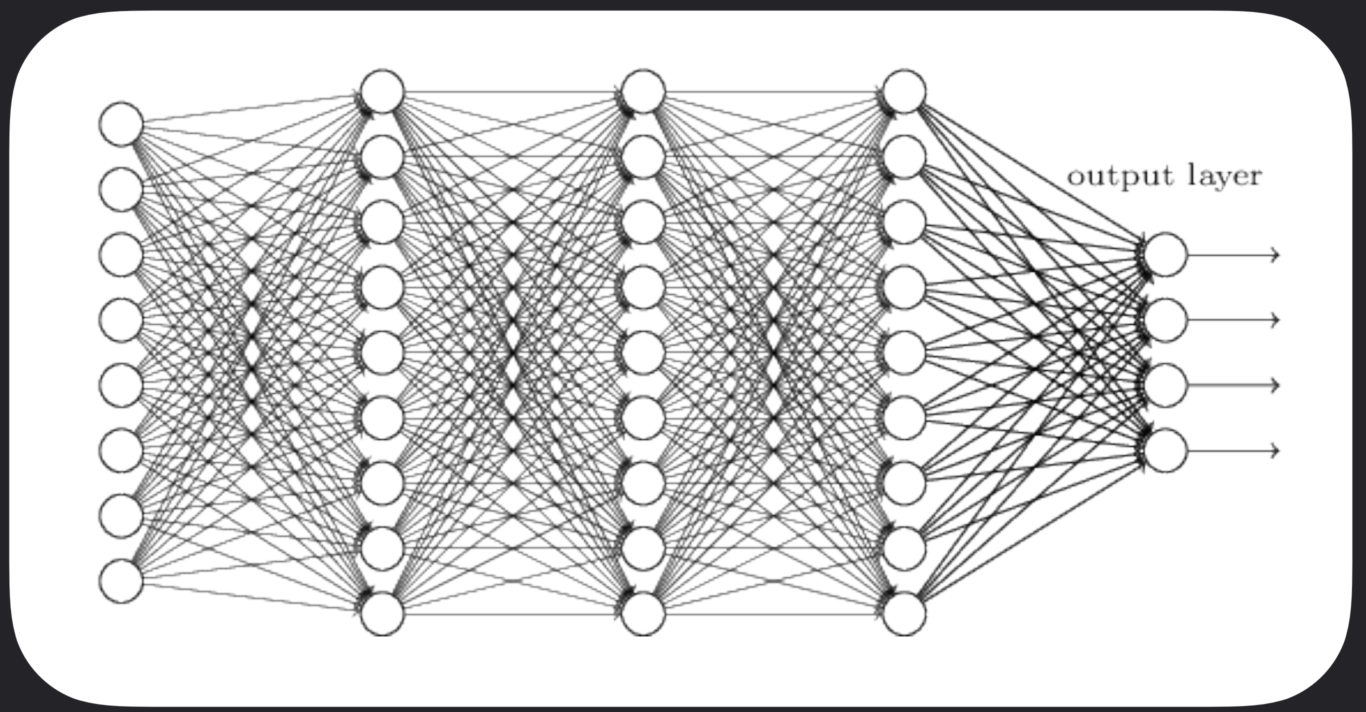
Example: Neural networks for emulating superparameterization?

Global aquaplanet testbed

Can 140,000,000 outputs from 1 year of ~ 10,000 cloud-resolving models...



Be fit by a deep neural network?



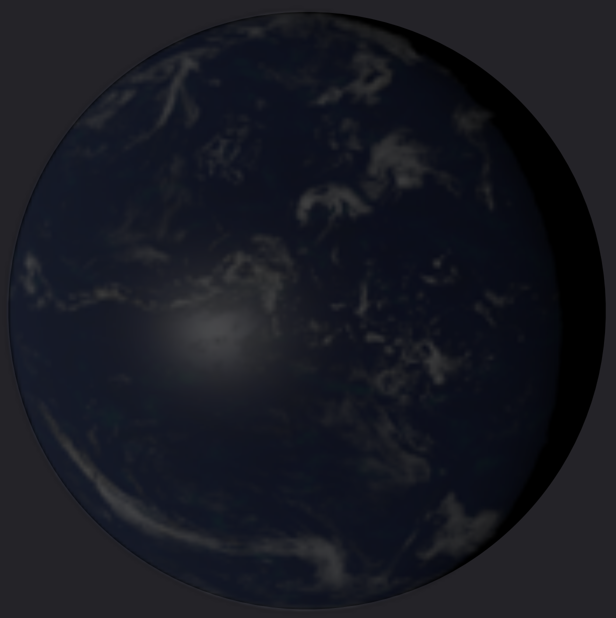
Gentine, Pritchard, Rasp et al., *GRL*, 2019.

Is deep learning viable for emulating superparameterization?

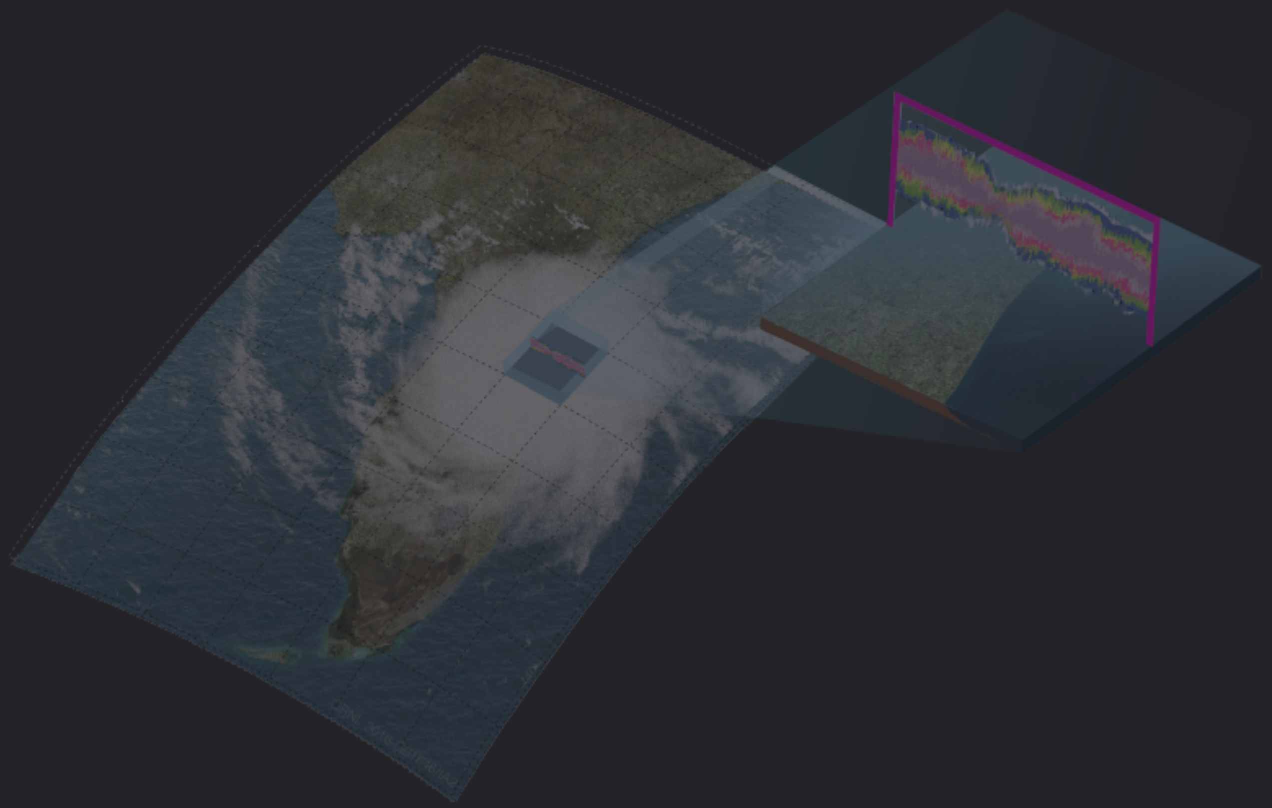
Quite possibly!

The “Cloud Brain”

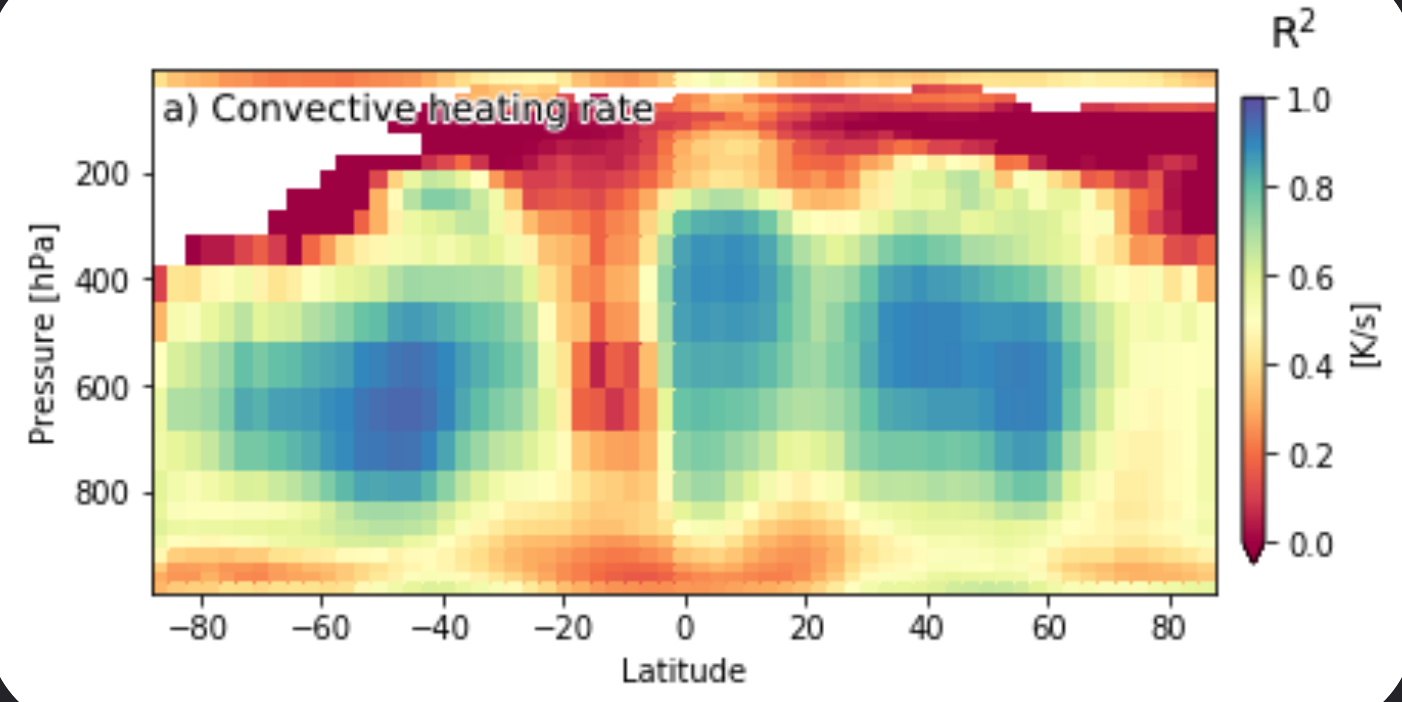
Global aquaplanet testbed



Can 140,000,000 outputs from 1 year of ~ 10,000 cloud-resolving models...



Be fit by a deep neural network?



tropospheric
heating by convection and
radiation.

WHAT IS THE OUTLOOK?

For replacing process parameterization with data-driven machine learning emulators?

Glimmer of recent success in cloud physics

But many outstanding issues challenges:

Interpretability?

Generalizability?

Stability?

Physical constraints?

How should uncertainties be incorporated?

What are the philosophical trade-offs?

Our community has only scratched the surface.

Road map: The Data Science WG

I. Motivation.

Who we are, context, the goal.

II. Three themes.

→ 1. How should we change modeling practices?

2. What is potential for data-driven discovery (patterns, predictability)?

3. Learn & talk - which methods are achieving breakthrough potential?

III. Objectives, timeline, upcoming events.

Webinars, collecting tools & data, how to get involved.

Road map: The Data Science WG

I. Motivation.

Who we are, context, the goal.

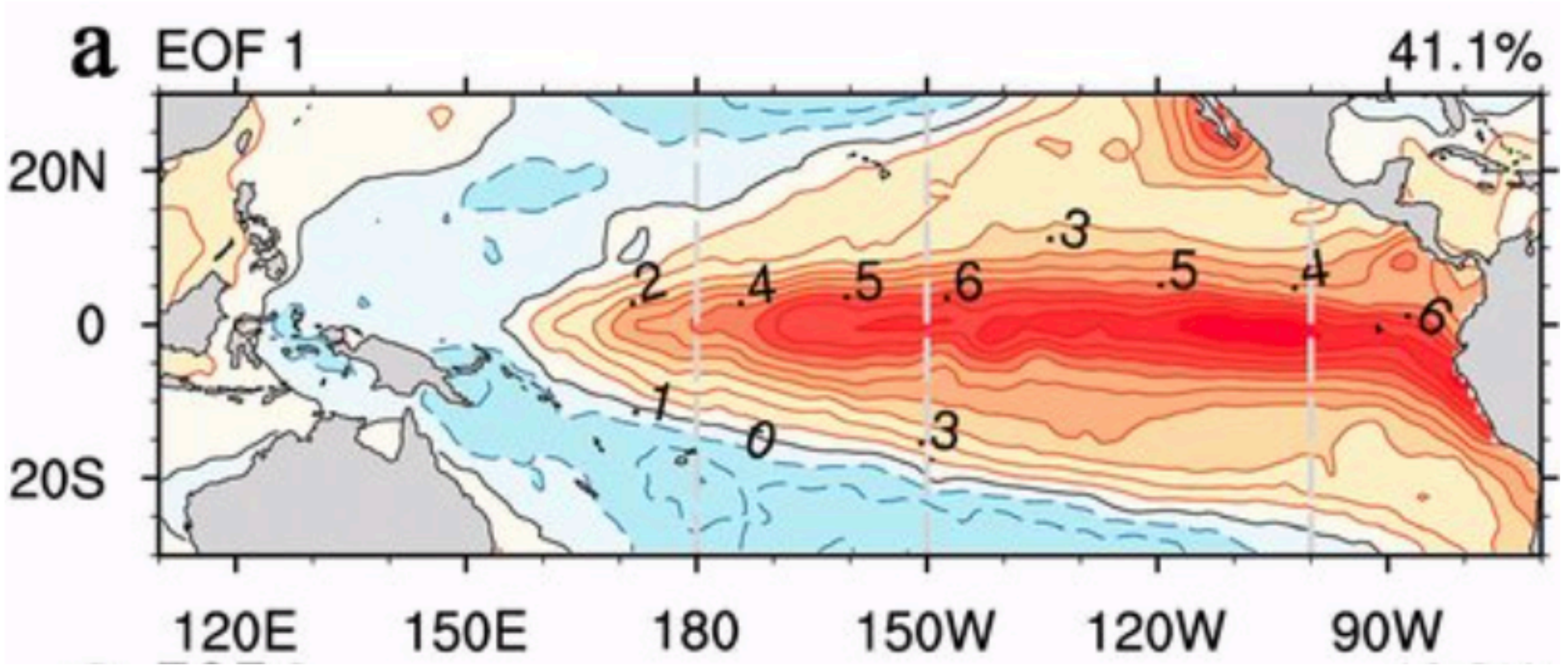
II. Three themes.

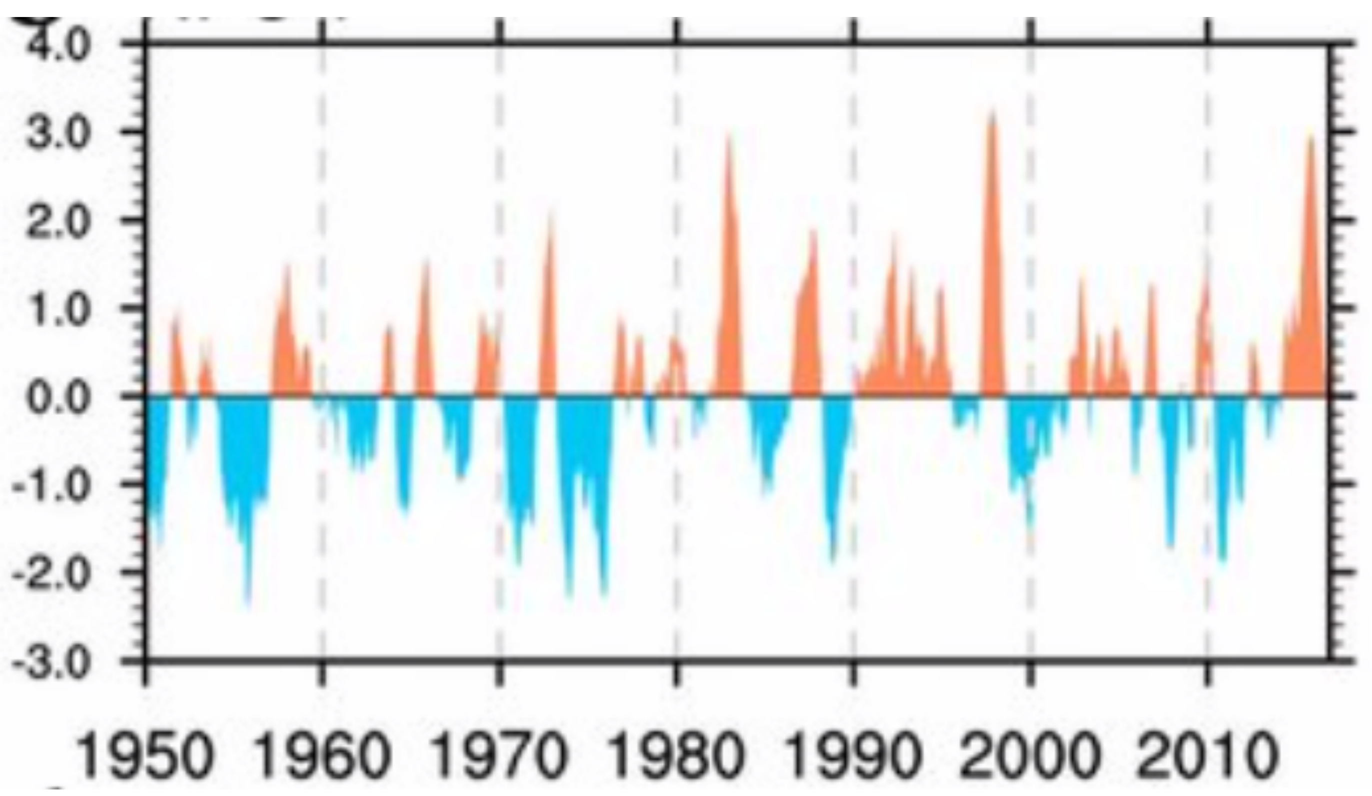
1. How should we change modeling practices?
2. What is potential for data-driven discovery (patterns, predictability)?
3. Learn & talk - which methods are achieving breakthrough potential?

III. Objectives, timeline, upcoming events.

Webinars, collecting tools & data, how to get involved.

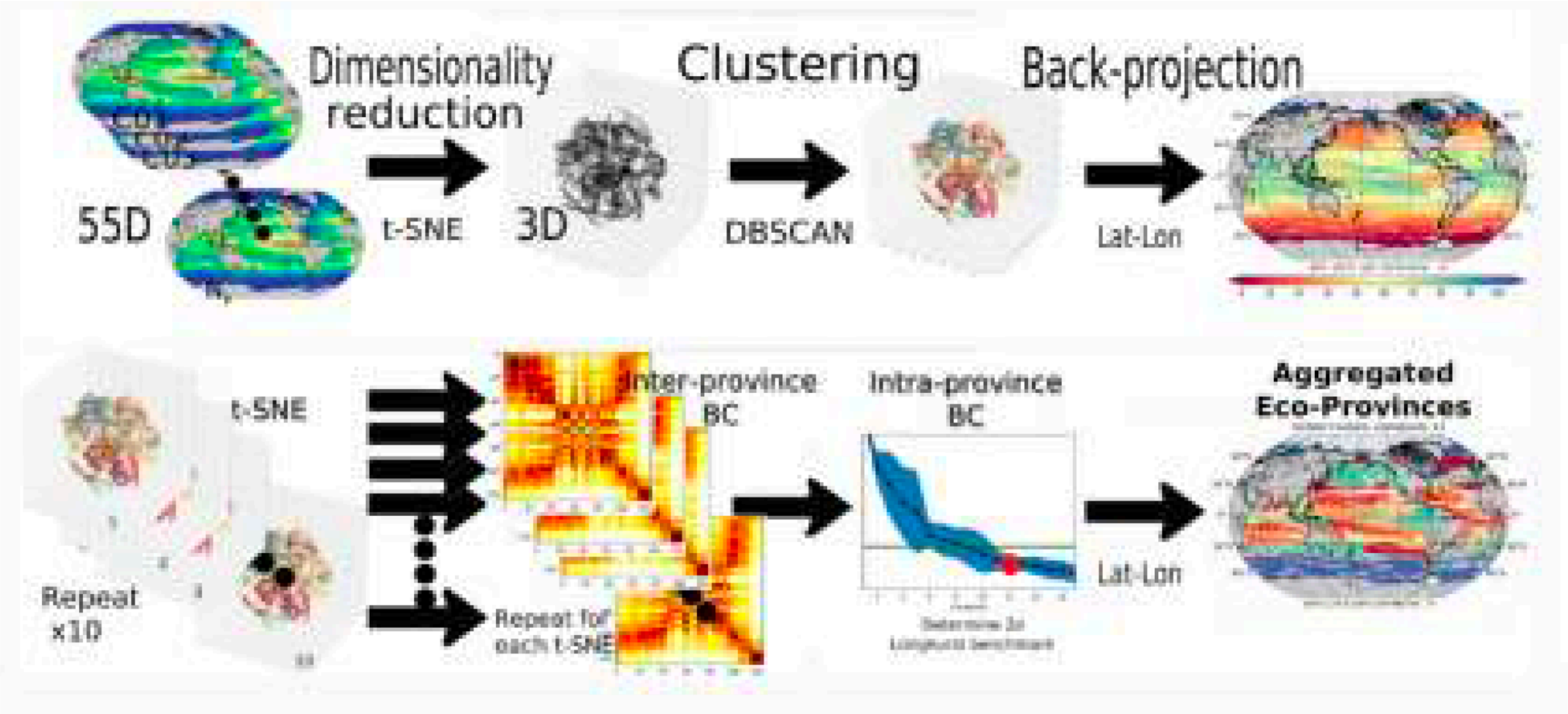
Data-driven ways to identify correlations and relevant patterns.

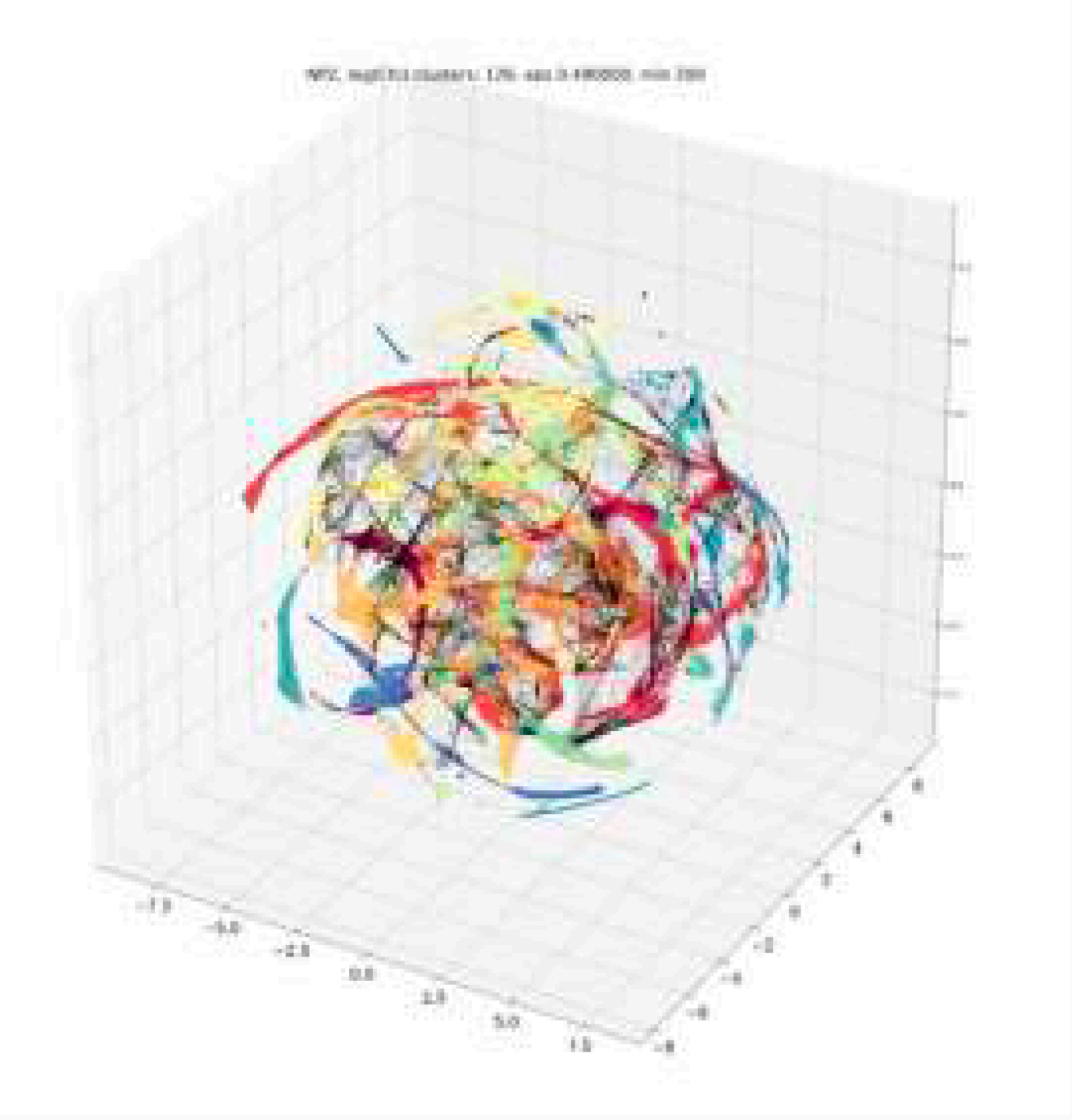




EOFS: We've been making good use of them for decades

New methods from data sciences have potential to help.

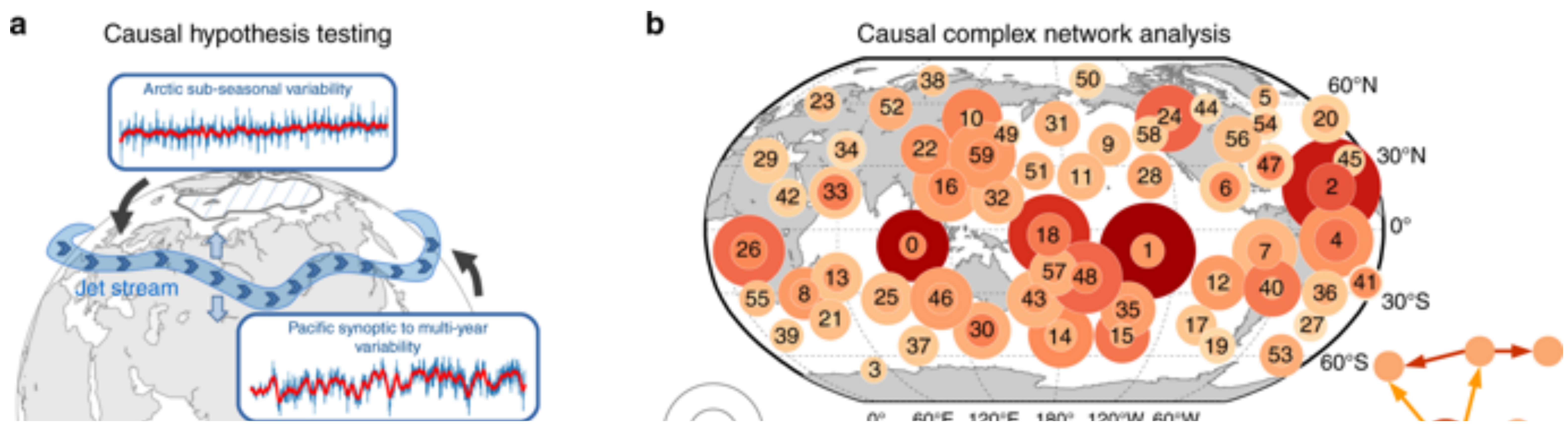




More “tools for the toolbox”

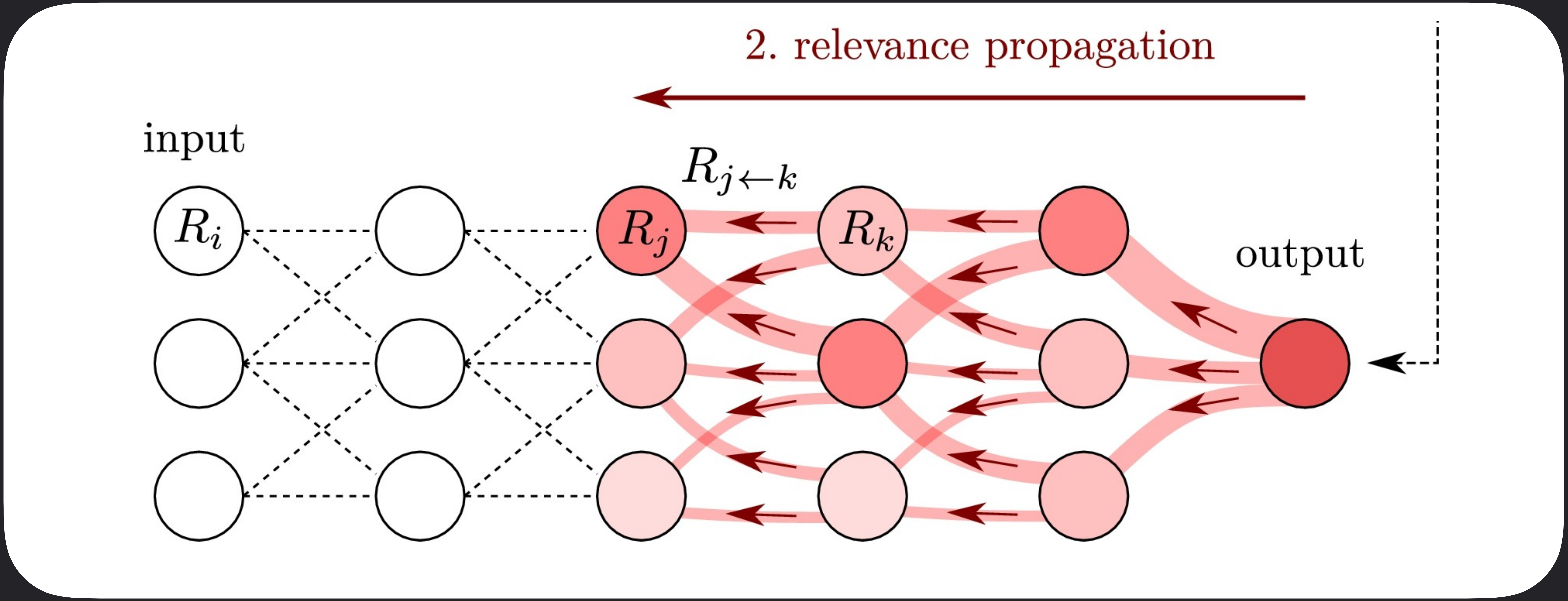
Figures courtesy of Maike Sonnewald.

Example: Causal inference theory for studying teleconnections



Runge et al., *Nature Comm.*, 2019.

Example: Machine learning is beginning to be interpretable.



Montevan et al., *Digital Signals Processing*, 2018.

OPPORTUNITIES.

Developing machine learning interpretability methods for climate applications.

Promising avenues

Optimal input analysis?

Layer-wise relevance propagation?

How to leverage data-hungry methods when samples infrequent (e.g. extreme events)

How should uncertainties be incorporated?

Road map: The Data Science WG

I. Motivation.

Who we are, context, the goal.

II. Three themes.

1. How should we change modeling practices?
2. What is potential for data-driven discovery (patterns, predictability)?
3. Learn & talk - which methods are achieving breakthrough potential?

III. Objectives, timeline, upcoming events.

Webinars, collecting tools & data, how to get involved.

Road map: The Data Science WG

I. Motivation.

Who we are, context, the goal.

II. Three themes.

1. How should we change modeling practices?
2. What is potential for data-driven discovery (patterns, predictability)?
3. Learn & talk - which methods are achieving breakthrough potential?

III. Objectives, timeline, upcoming events.

Webinars, collecting tools & data, how to get involved.

We are in the infancy of adopting emerging data science tools.

Some are tinkering but scattered across disciplines

Which specific tools are:

reproducibly helpful?

statistically novel?

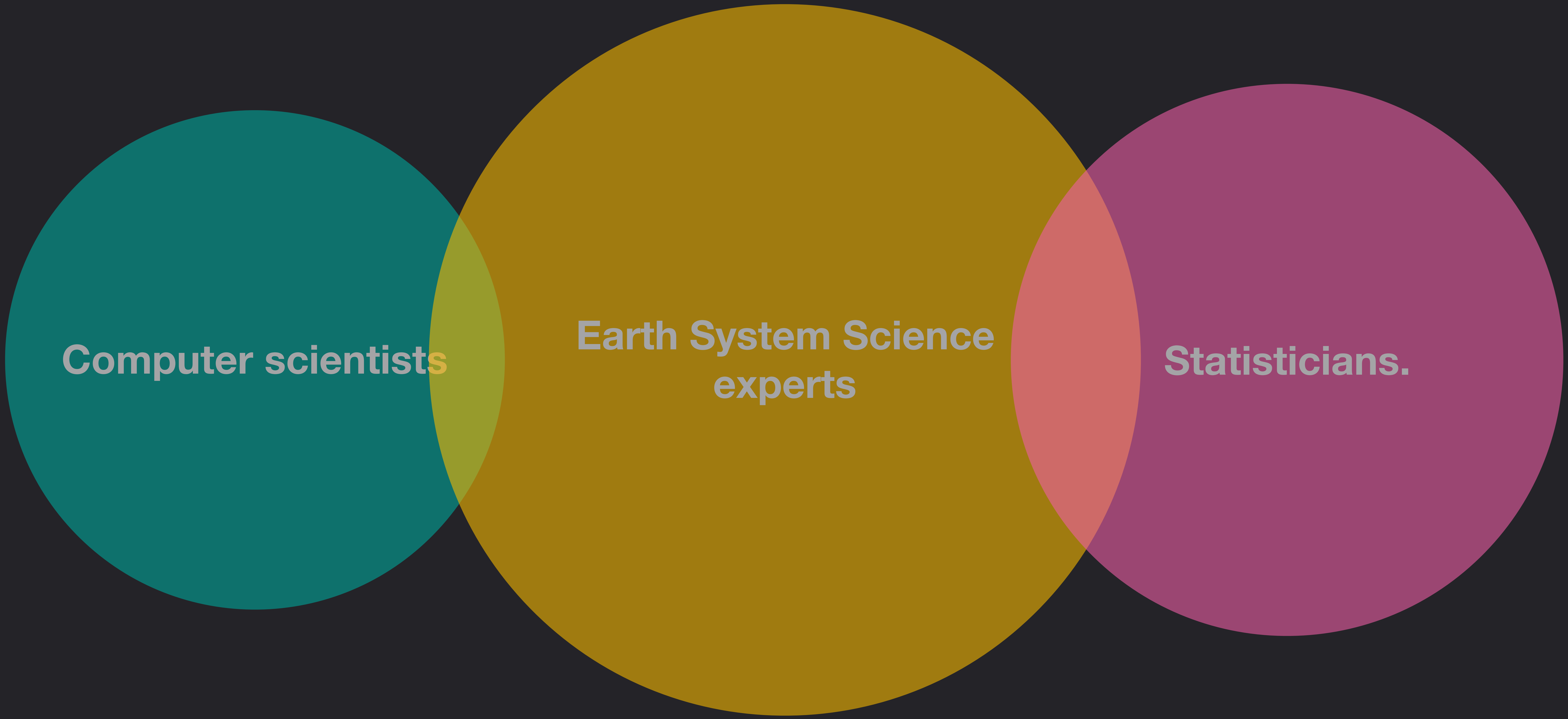
technically approachable?

gaining consensus?

How should the community organize itself?

(including this working group)

CONVERSATIONS NEEDED.



Road map: The Data Science WG

I. Motivation.

Who we are, context, the goal.

II. Three themes.

1. How should we change modeling practices?
2. What is potential for data-driven discovery (patterns, predictability)?
3. Learn & talk - which methods are achieving breakthrough potential?

III. Objectives, timeline, upcoming events.

Webinars, collecting tools & data, how to get involved.

Road map: The Data Science WG

I. Motivation.

Who we are, context, the goal.

II. Three themes.

1. How should we change modeling practices?
2. What is potential for data-driven discovery (patterns, predictability)?
3. Learn & talk - which methods are achieving breakthrough potential?

→ III. Objectives, timeline, upcoming events.

Webinars, collecting tools & data, how to get involved.

Road map: The Data Science WG

I. Motivation.

Who we are, context, the goal.

II. Three themes.

1. How should we change modeling practices?
2. What is potential for data-driven discovery (patterns, predictability)?
3. Learn & talk - which methods are achieving breakthrough potential?

III. Objectives, timeline, upcoming events.

→ Webinars, collecting tools & data, how to get involved.

WHAT'S NEXT?

WHAT'S NEXT?

Working group invitations are out (yesterday)

Bi-monthly webinars begin this fall.

WG discussions with community

Curated experiences

Curated tools (& how-to guides)

Working group discussions with each other (years 2 & 3)

Online presence

Publications: Years 1 & 3.

THANKS

@CLIVAR_DataSci

mspritch@uci.edu