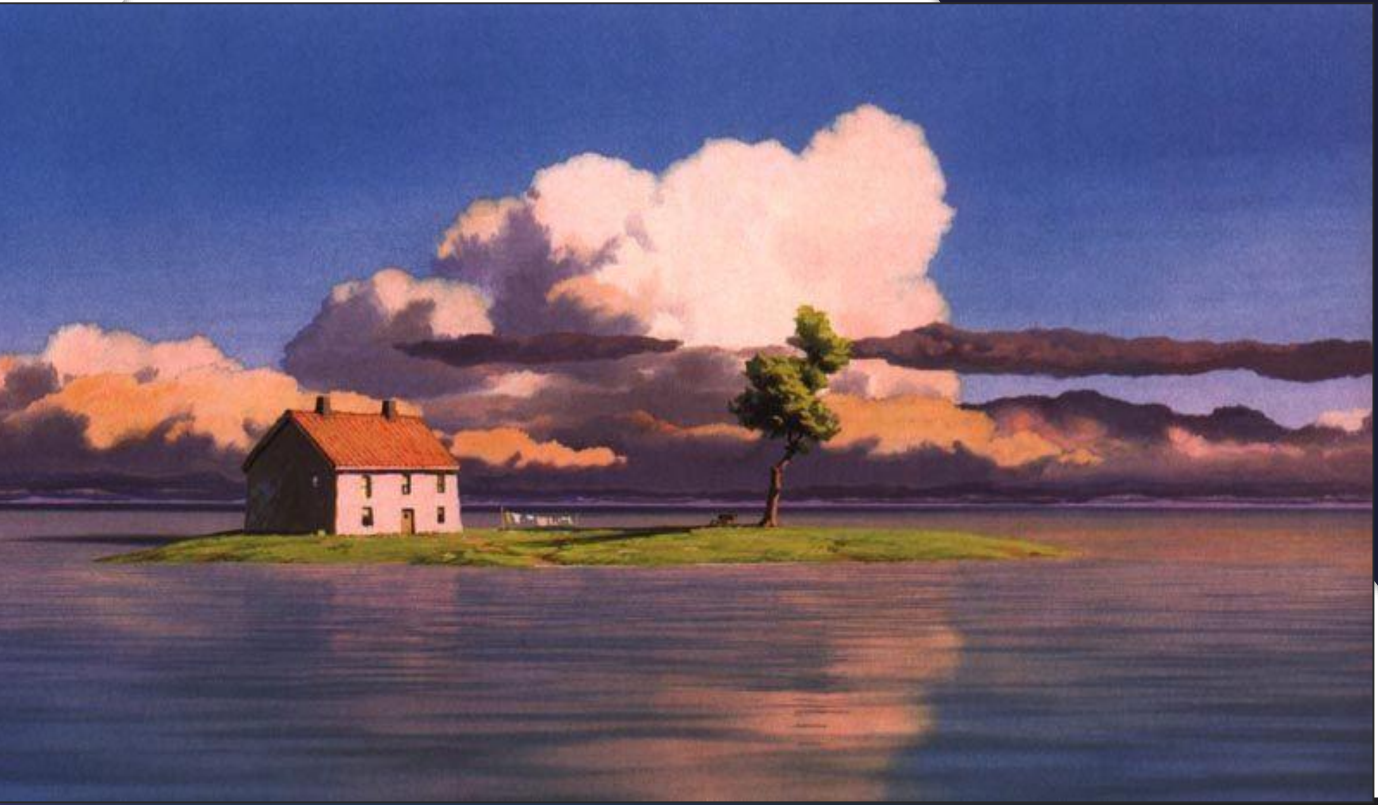
Potentialities of machine learning in climate change research: A few examples

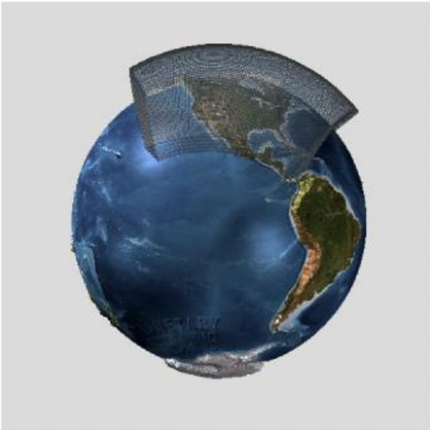


Alexis Hannart
OURANOS

21 mai 2019

Ouranos in a few words

- Basé à Montréal, créé par les membres en 2002
- Masse critique d'experts pour assurer le développement et la coordination de R&D interdisciplinaire, appliquée et orientée vers les usagers de l'adaptation
- Innovation par une recherche collaborative connectée avec praticiens/décideurs (opérations, politique, planification, stratégique)



1. Un programme en Science du climat dédié à la production de scénarios climatiques et à la modélisation climatique aux échelles régionales
2. Un programme multidisciplinaire et multi-institutionnel en Vulnérabilités, Impacts et Adaptation

Environnement Canada / Environment Canada · Hydro Québec · Québec · Montréal · Ontario Power Generation · Université du Québec École de technologie supérieure

McGill · UQÀM Université du Québec à Montréal · INRS Université de recherche · Université Laval · Communauté métropolitaine de Québec · RioTinto · UQAR · Manitoba Hydro

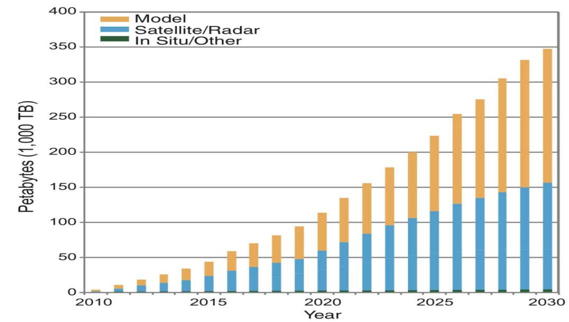
Gestion de l'eau **R. Turcotte et A. Blondlot**	Ressources forestières **D. Houle**	Environnement Maritime	Énergie **I. Chartier et J. Clavet-Gaumont**	Environnement nordique **S. Bleau et R. Siron**	Environnement bati **N. Bleau**	Écosystème & biodiversité **R. Siron**	Agriculture, pêches et aquaculture commerciales **A. Blondlot**	Santé **P. Gosselin et C. Campagna**	Tourisme **S. Bleau**

Agenda

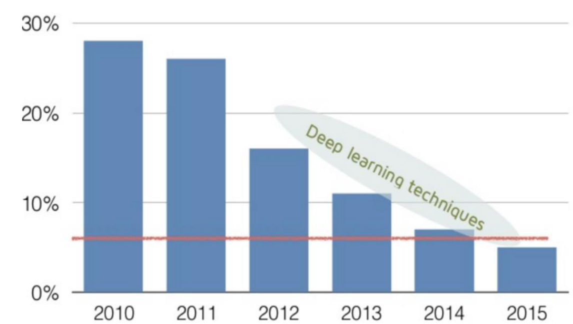
- Context
- A few case studies
- A few challenges

A simple story

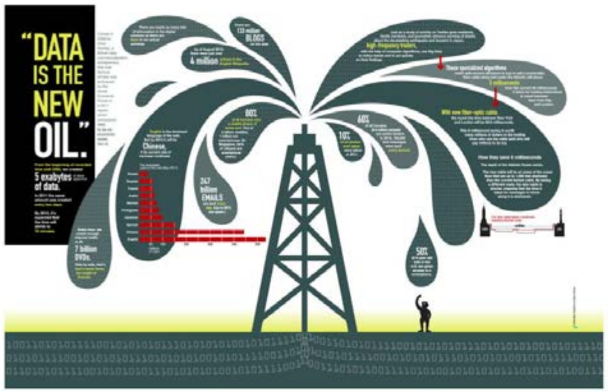
- Exponential trend on data generation and storage,



- Matched by smart algorithms and large computional power,

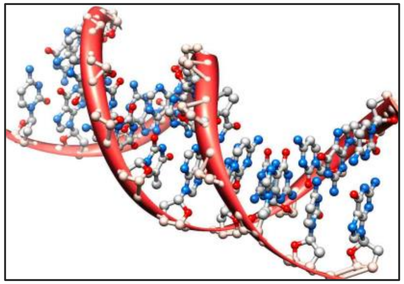


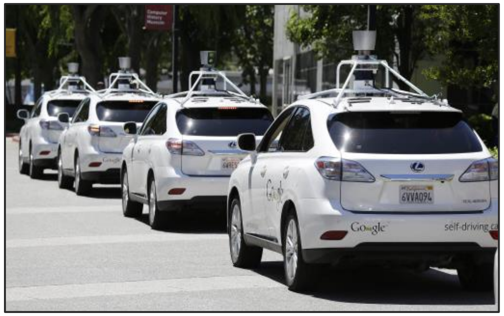
- New applications, products, services, and tools for science.

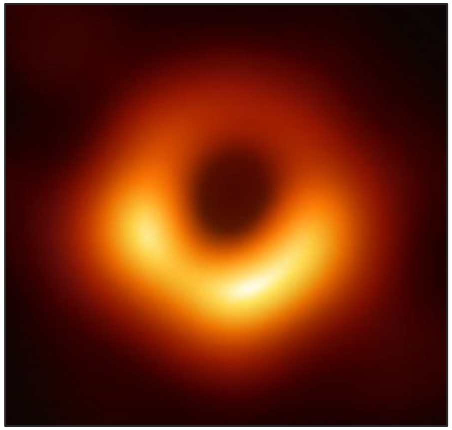


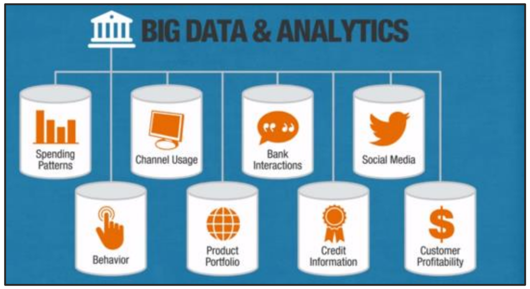
The ‘fourth revolution’

- Search Engines & Internet
- Health & Genomics
- Astrophysics
- Banking & Finance
- Transport & Logistics
- Marketing & Media
- Energy & Distribution
- Agriculture & Forestry
- Urbanism

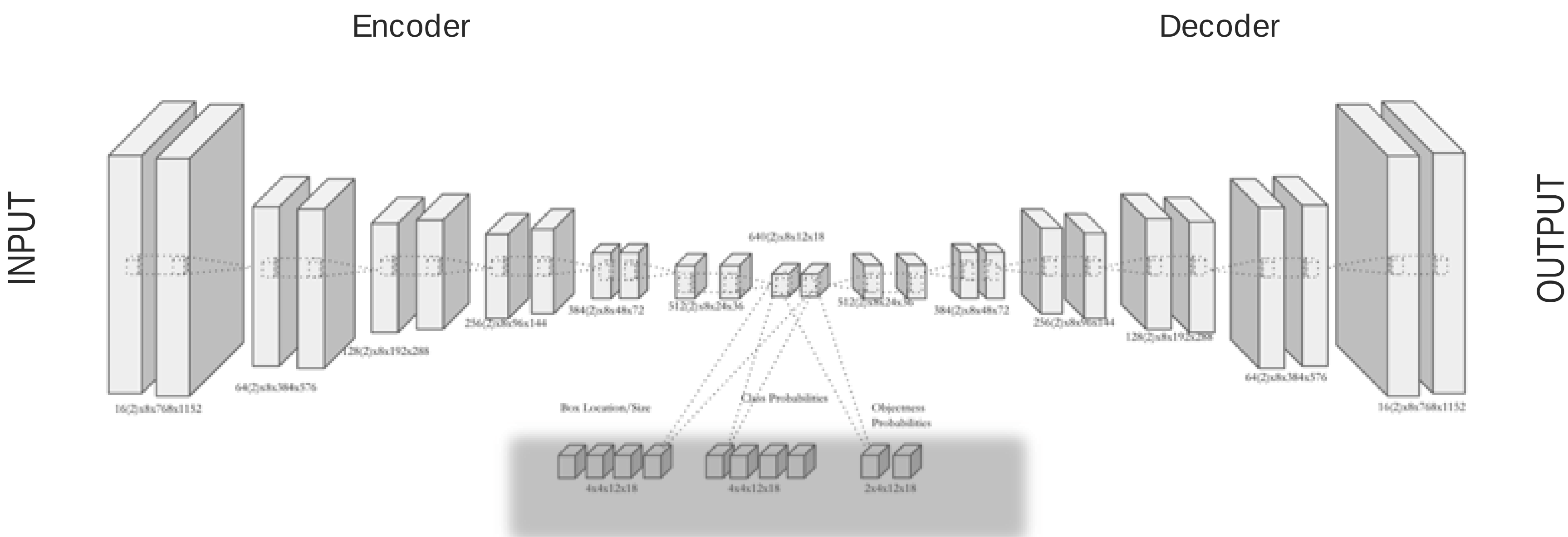






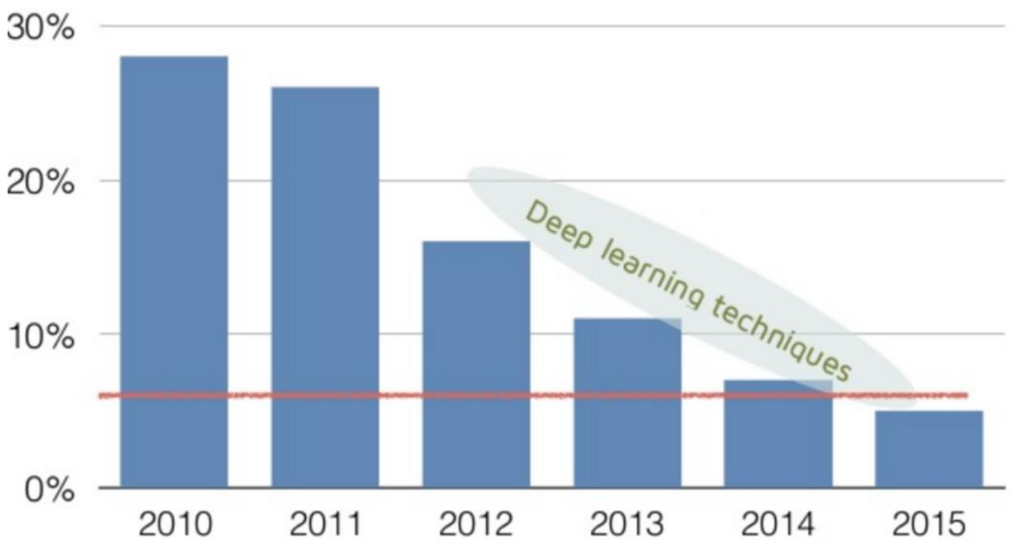


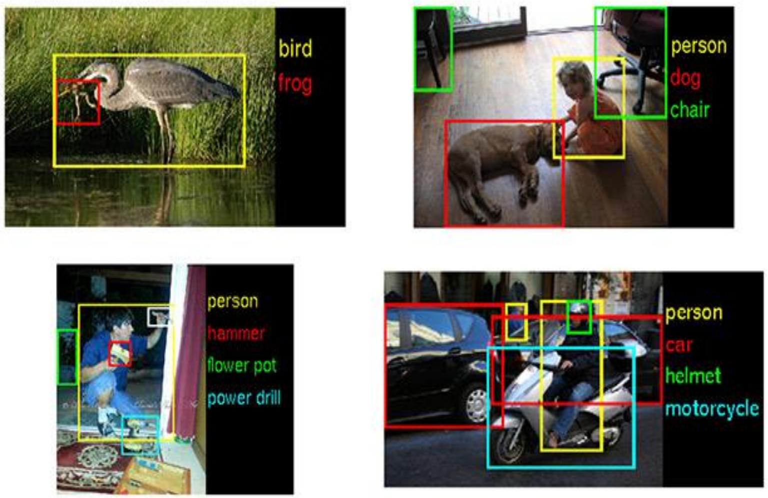
Deep learning – example of a network architecture



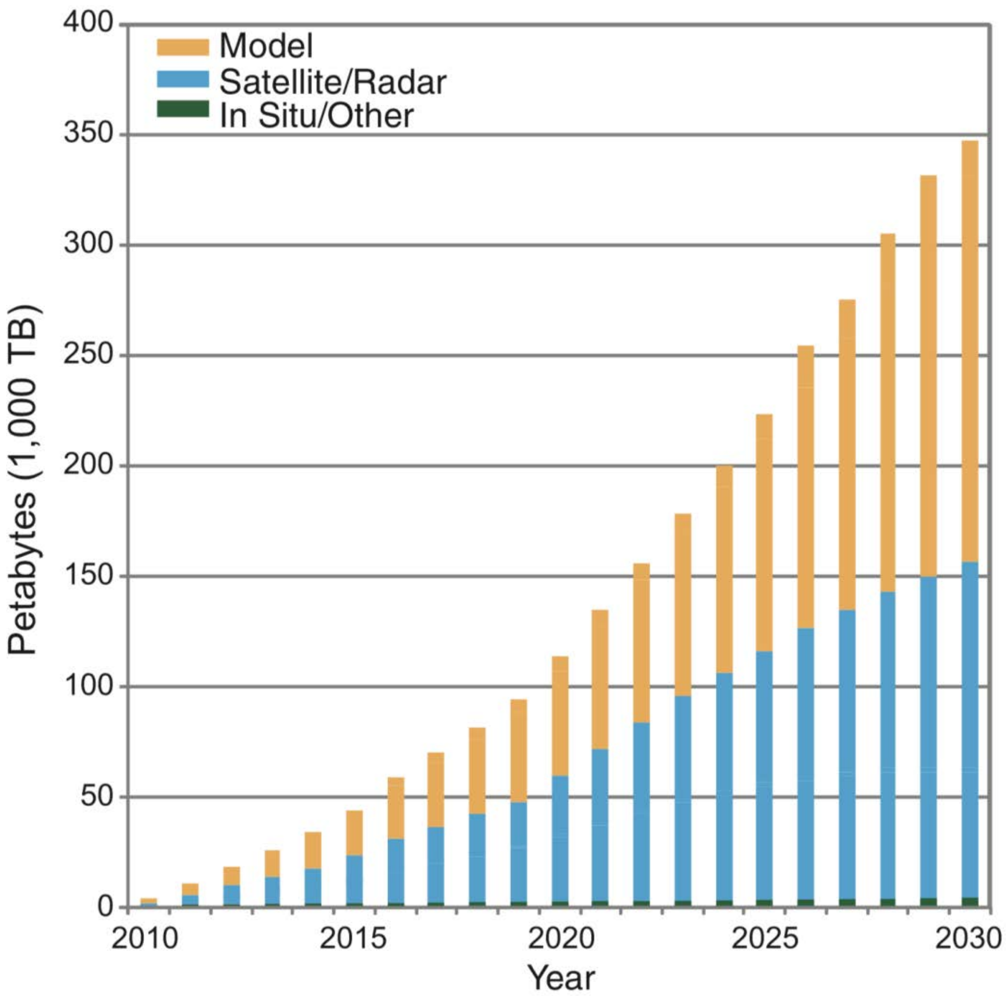
Skill trend in image recognition

ImageNet challenge



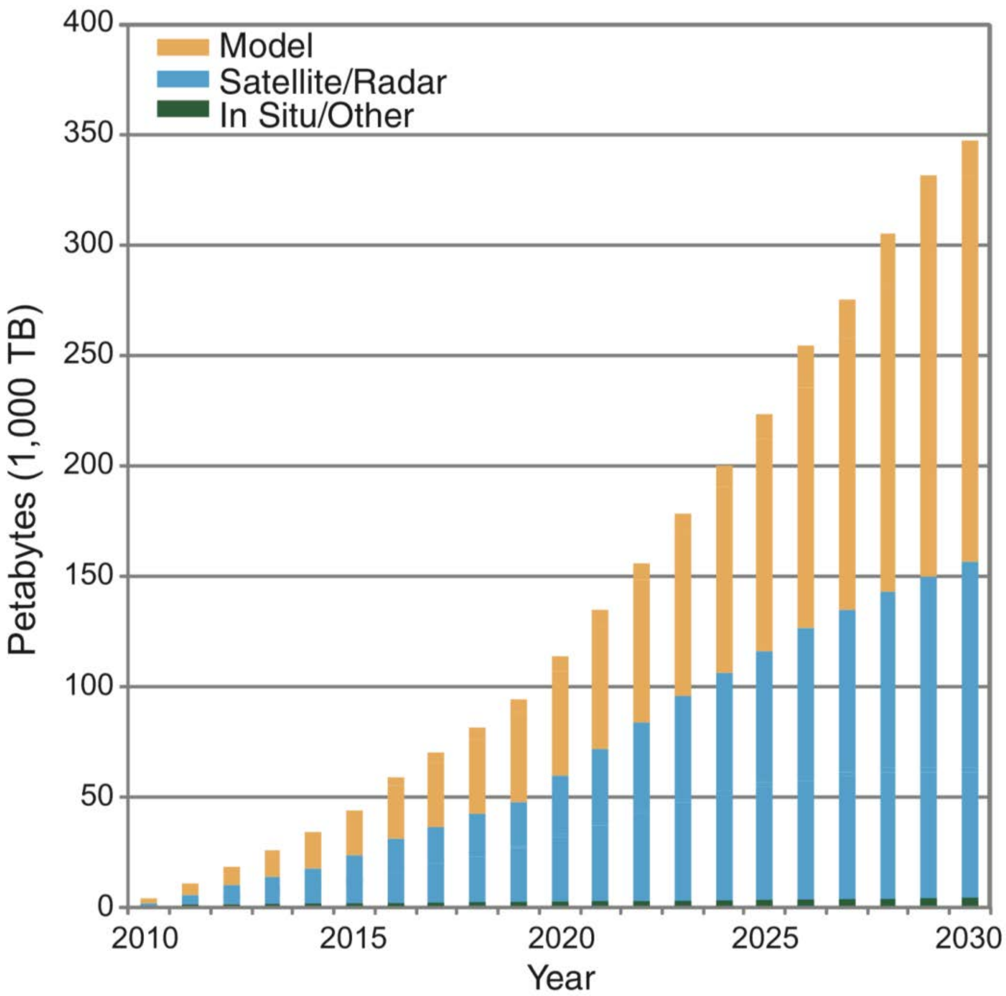


Data volume trend in climate science



Overpeck et al. 2011

Data volume trend in climate science



Should climate science and climate services join the party ?

Overpeck et al. 2011

World Economic Forum report 2018

WORLD ECONOMIC FORUM

COMMITTED TO IMPROVING THE STATE OF THE WORLD

Fourth Industrial Revolution for the Earth Series

Harnessing Artificial Intelligence for the Earth

Climate change

- Clean power
- Smart transport options
- Sustainable production and consumption
- Sustainable land-use
- Smart cities and homes

Biodiversity and conservation

- Habitat protection and restoration
- Sustainable trade
- Pollution control
- Invasive species and disease control
- Realizing natural capital

Healthy Oceans

- Fishing sustainably
- Preventing pollution
- Protecting habitats
- Protecting species
- Impacts from climate change (including acidification)

Water security

- Water supply
- Catchment control
- Water efficiency
- Adequate sanitation
- Drought planning

Clean air

- Filtering and capture
- Monitoring and prevention
- Early warning
- Clean fuels
- Real-time, integrated, adaptive urban management

Weather and disaster resilience

- Prediction and forecasting
- Early warning systems
- Resilient infrastructure
- Financial instruments
- Resilience planning

Climate Informatics, NCAR, from 2011 to present

NCAR UCAR CISL Computational & Information Systems Lab

Software | Data Portals | User Support | Resources | About Us

Daily Bulletin | Resource Status | Newsroom | Events

CLIMATE INFORMATICS

- Home
- Important Dates
- Registration
- Application for Travel Support
- Paper Submission Guidelines
- Poster Guidelines
- Hackathon

HOME

CI 2017

7th International Workshop on Climate Informatics
September 20-22, 2017

Hosted by the National Center for Atmospheric Research in Boulder, CO

Climate Informatics Workshop

About Climate Informatics

We have greatly increased the volume and diversity of climate data from satellites, environmental sensors and climate models in order to improve our understanding of the climate system. However, this very increase in volume and diversity can make the use of traditional analysis tools impractical and necessitate the need to carry out knowledge discovery from data. Machine learning has made significant impacts in fields ranging from web search to bioinformatics, and the impact of machine learning on climate science could be as profound. However, because the goal of machine learning in climate science is to improve our understanding of the climate system, it is necessary to employ techniques that go beyond simply taking advantage of co-occurence, and, instead, enable increased

Big Data & Environment, Buenos Aires, November 2015

- 8 sponsors: CNRS-INSU, Fondation Science Mathématique de Paris, Dept. of Computation of UBA, ANR, MINCyT, CONICET, …
- 12 nationalities: Arg., Fr., USA, Canada, Chili, Brazil, Ur., Peru, ...
- 6 thematical fields: Climate, Maths, Machine learning, Computer Science, Biology, Ecology.

Accueil > dse2017

dse2017

- dse2017
- Program
- Talks
- Posters
- Venue
- Accommodations
- Committees

Abstract

Environmental sciences have experienced a data deluge with the explosion in the amount of data produced by sensors and models that monitor, measure and forecast the Earth system. This exponential trend in data availability is expected to continue in the future thereby creating many new opportunities, needs and challenges. On the other hand, data science has emerged as a wide multidisciplinary dynamic which addresses challenges associated to large and complex data and encompasses diverse fields in applied mathematics and computer science.

Aim

The conference will gather researchers that have an expertise in one of the two areas (data science, environmental data) and some interest for the other. Its main goal is to explore the fruitful interplay between the two areas, and ultimately to help create new connections and collaborations between the scientific communities involved. Another objective is to propose some high level courses and practices at the interaction of these two areas.

Climate change and AI, California, June 2019

Climate change is widely agreed to be one of the greatest challenges facing humanity. We already observe increased incidence and severity of storms, droughts, fires, and flooding, as well as significant changes to global ecosystems, including the natural resources and agriculture on which humanity depends. The 2018 UN report on climate change estimates that the world has only thirty years to eliminate greenhouse emissions completely if we are to avoid catastrophic consequences.

ICML 2019 Workshop

Many in the ML community wish to take action on climate change, yet feel their skills are inapplicable. This workshop will showcase the many settings in which machine learning can be applied to reducing greenhouse emissions and helping society adapt to the effects of climate change. Climate change is a complex problem requiring simultaneous action from many directions. While

Organizers

David Rolnick (UPenn)
Alexandre Lacoste (ElementAI)
Tegan Maharaj (MILA)
Jennifer Chayes (Microsoft)
Yoshua Bengio (MILA)
———
Karthik Mukkavilli (MILA)
Narmada Balasooriya (ConscientAI)
Di Wu (MILA)
Priya Donti (CMU)
Lynn Kaack (ETH Zürich)
Manvitha Ponnapati (MIT)

AI in weather and climate, Montreal, July 2019

JM07 - Artificial Intelligence and Big data in Weather and Climate Science (IAMAS, IAHS)

Convener: Philippe Roy (Canada, IAMAS)

Co-Conveners: Alexis Hannart (Canada, IAMAS), David Hall (USA, IAMAS), Allen Huang (USA, IAMAS), Ashish Sharma (Australia, IAHS)

Description

Rapid advances in artificial intelligence, combined with the availability of enormous amount of data (termed Big Data) is opening new avenues for climate analysis and climate scenarios. The long awaited promises of AI is now common in many disciplines. Applying AI methods, combined with physical knowledge, can improve climate analysis and provide better climate simulations and climate products, notably for high-impact events, such as floods, wildfires and winds.

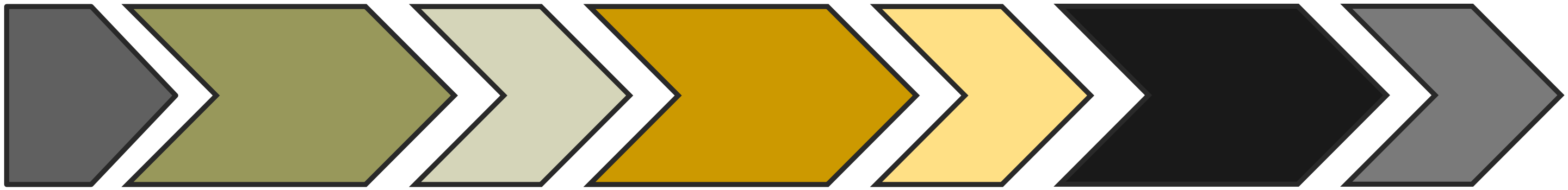
Contributions are welcome in the following areas, but not limited to:

- Decision-making tools for climate and weather related hazards;
- Data mining and explorations approaches
- Pattern recognition and classification
- Climate and weather emulators
- Smart-grid and smart cities applications combining AI and weather and climate data
- Novel approaches in the domain of natural hazards using AI methods

Climate change: from causes to consequences

Causes
(GHGs emissions)

Consequences
(misc.)



Climate
model

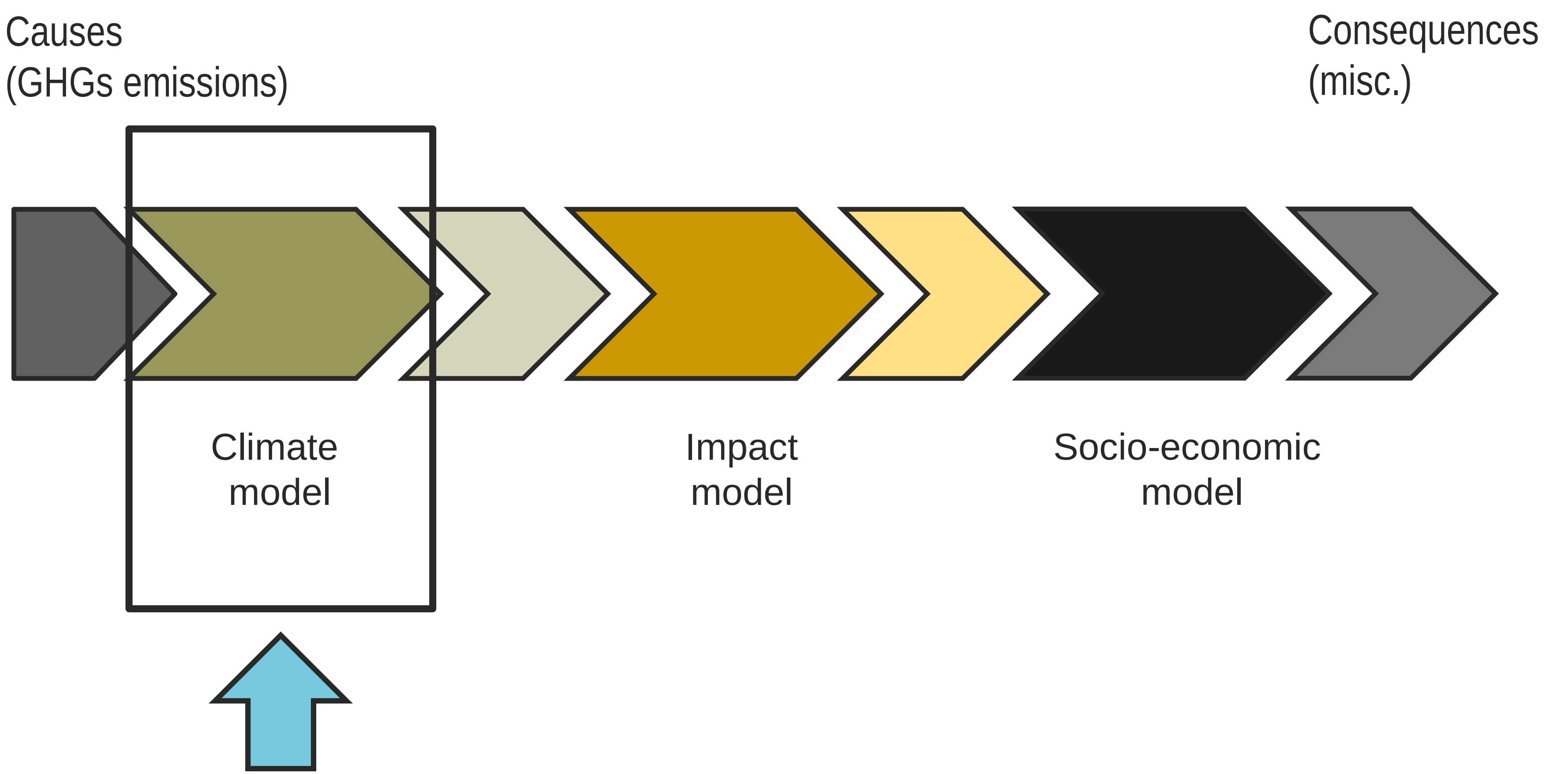
Impact
model

Socio-economic
model

Agenda

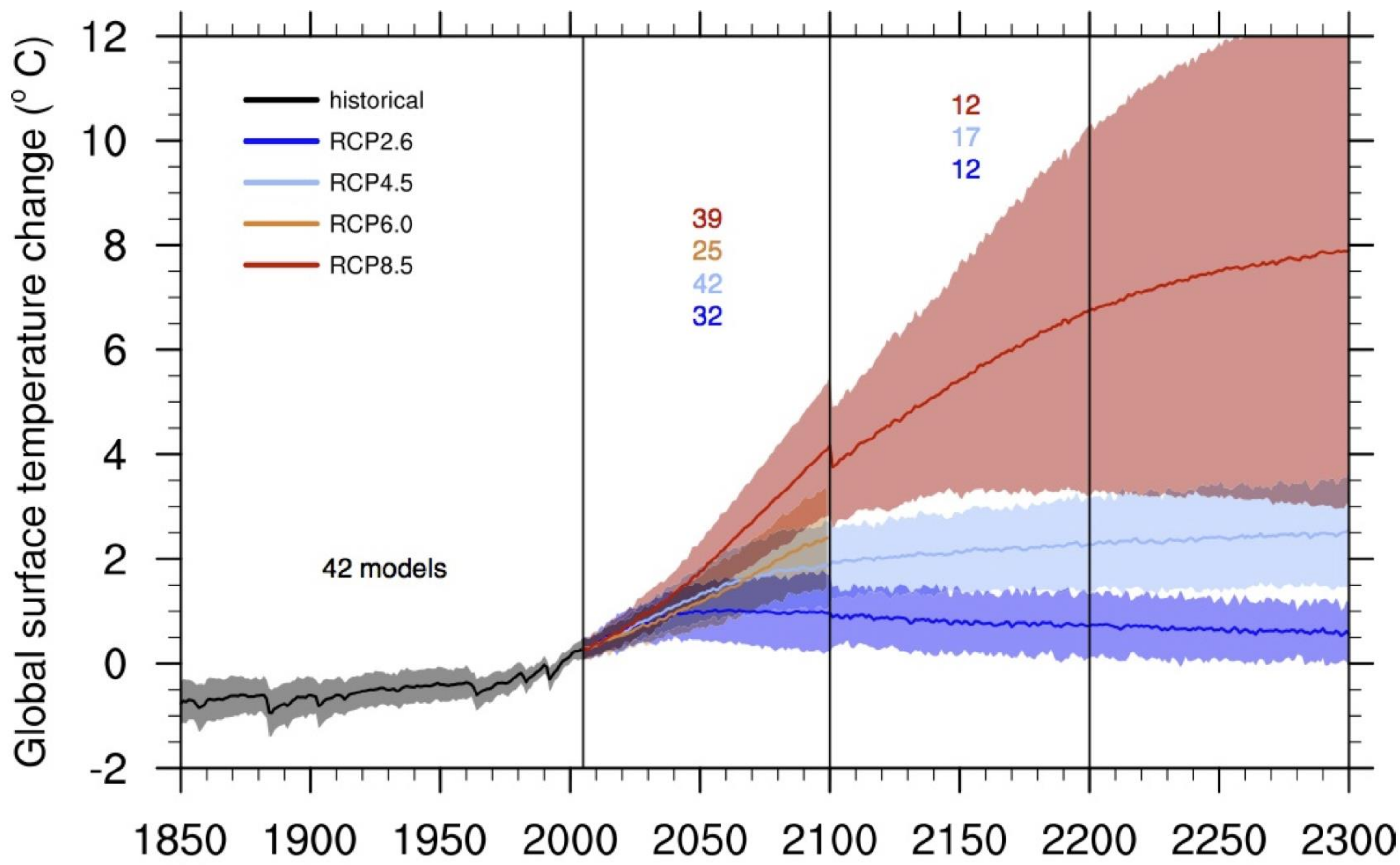
- Context
- A few case studies: subgrid parameterization
- A few challenges

Context

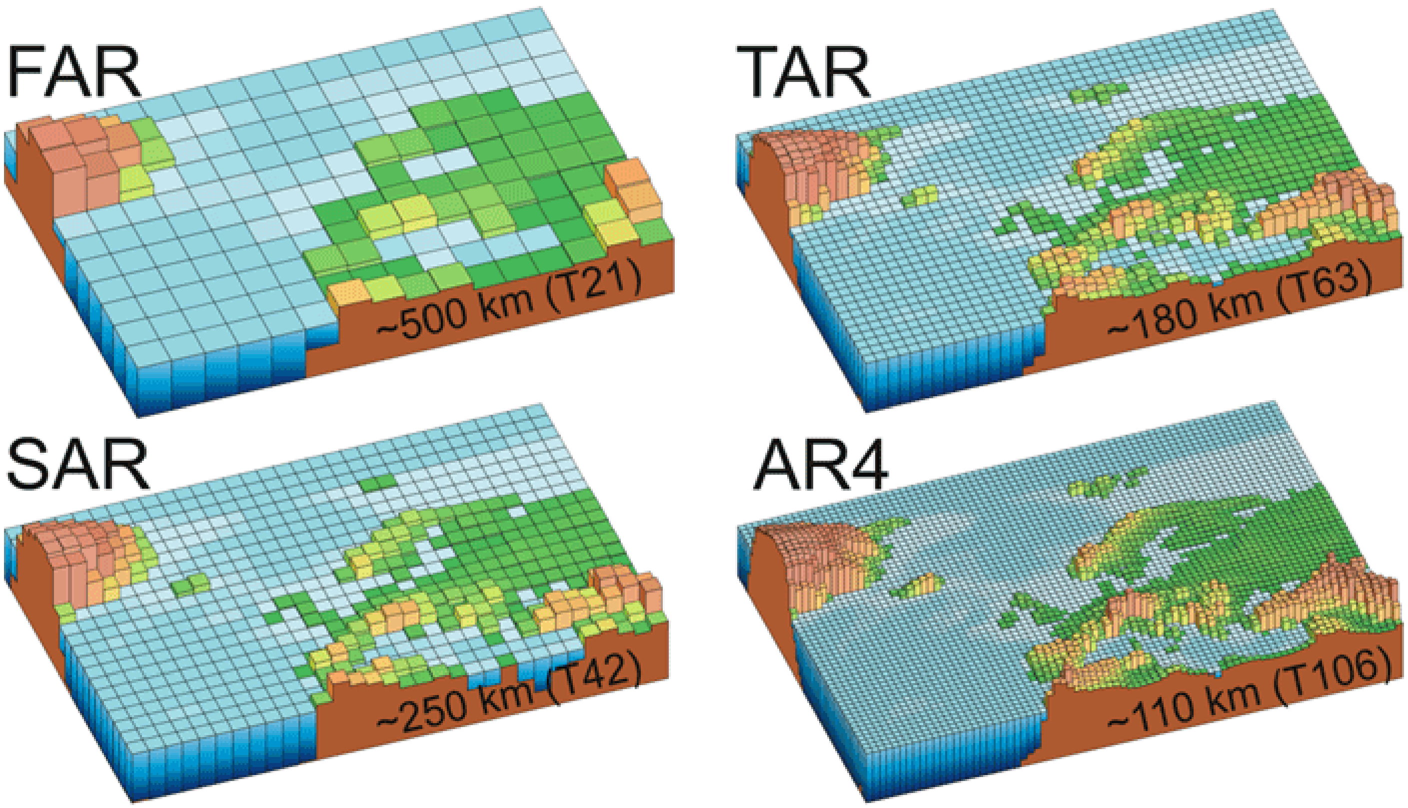


Causes
(GHGs emissions)
Consequences
(misc.)
Climate
model
Impact
model
Socio-economic
model

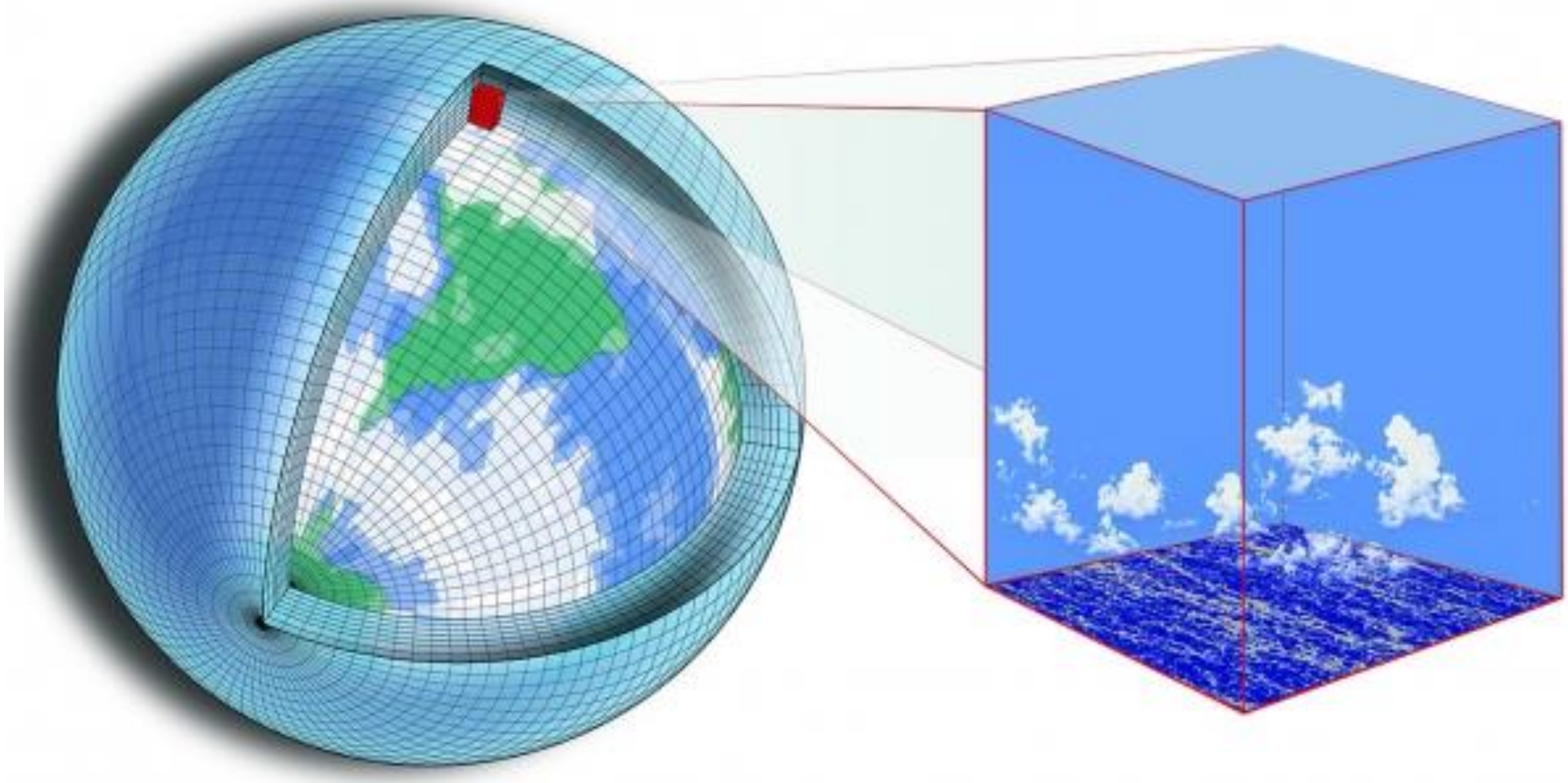
Future projections of climate change



Climate models



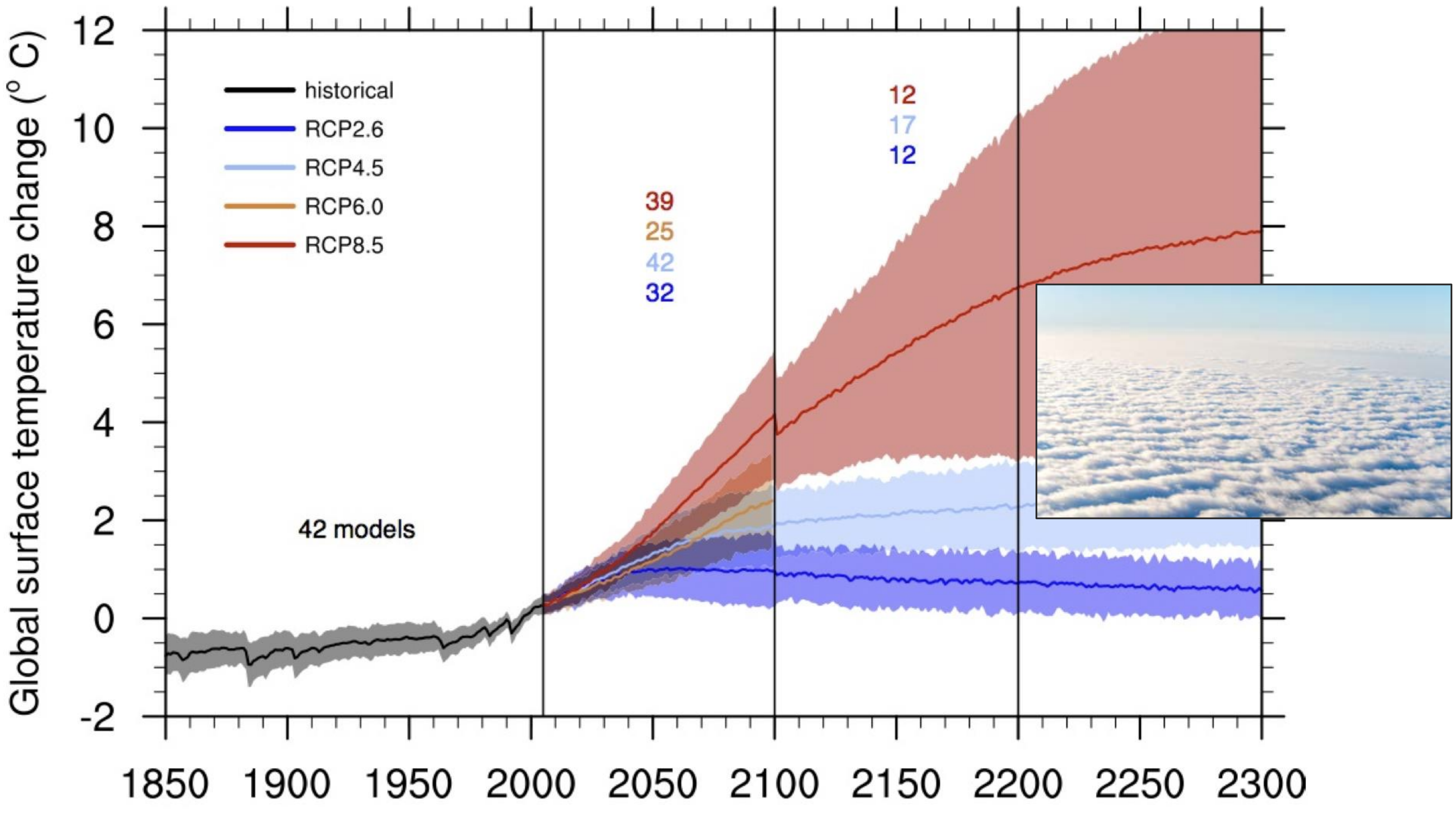
Climate models: subgrid processes



Clouds

Low level clouds: stratocumulus

Stratocumulus response is a major part of uncertainty



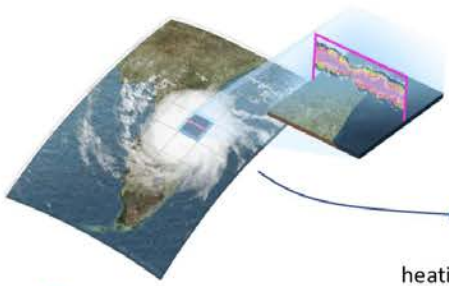
Getting around the computational wall: the AI trick

Deep learning can skillfully approximate sub-grid climate model physics harvested from cloud-resolving simulations.

Is deep learning viable for sub-grid parameterization?

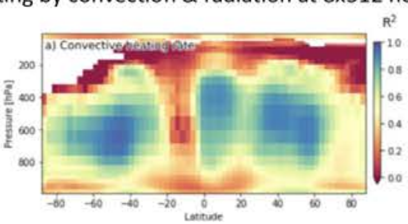
Aquaplanet SPCAM testbed
1 year for training, 1 for validation
Globally diverse meteorological regimes

Can the 140M outputs from
1 year of 9k Cloud Resolving Models...
(solutions of accurate radiative transfer & explicit CRM equations)



SuperParameterization

....Be fit by a deep, fully connected network?
Yes, e.g. $R^2 > 0.7$ for mid-tropospheric heating by convection & radiation at 8x512 nodes.



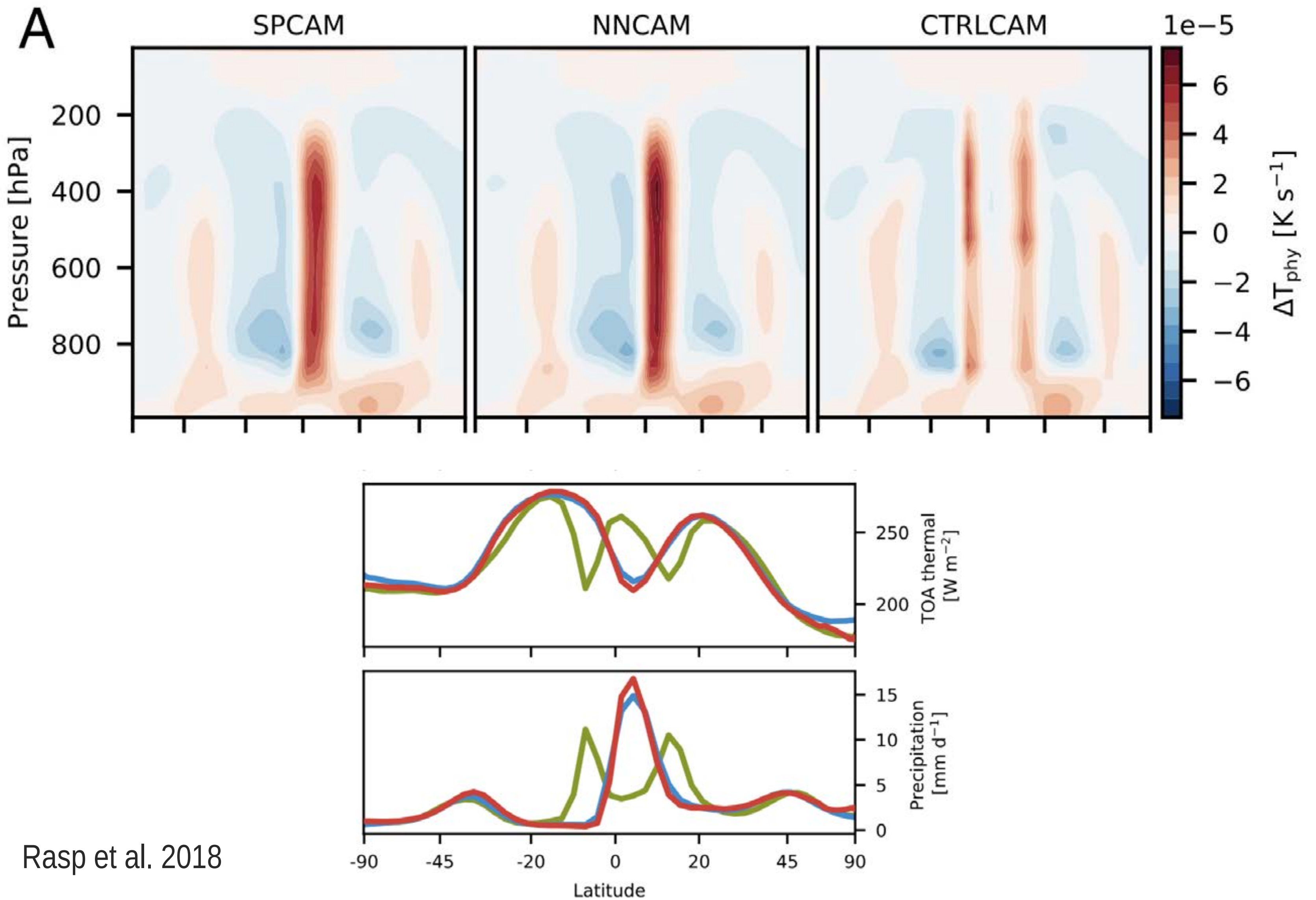
Possibly!
Just 3 months' hi-res sim data is enough for a good fit!

The "Cloud Brain"

Geophysical Research Letters
Could machine learning break the convection parameterization deadlock?
P. Gentine, M. Pritchard, S. Rasp, G. Reinaudi & G. Yacalis. May 2018.

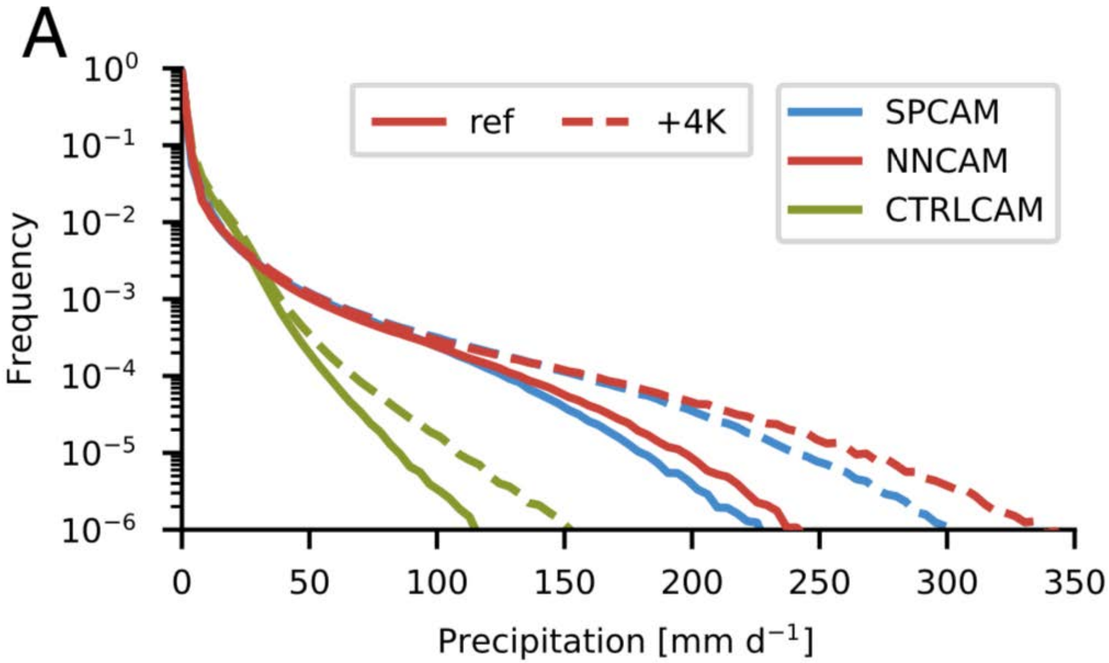
Image credits: Illustration by Tony Gold for IBM systems magazine, ORNL visualization lab.

Some encouraging early results



Rasp et al. 2018

Some encouraging early results



A promising way forward

CLIMA
CLIMATE MODELING ALLIANCE

A NEW APPROACH TO CLIMATE MODELING

CLIMATE MACHINE

We are developing the first Earth system model that automatically learns from diverse data sources. Our model will exploit advances in machine learning and data assimilation to learn from observations and from data generated on demand in targeted high-resolution simulations, for example, of clouds or ocean turbulence. This will allow us to reduce and quantify uncertainties in climate predictions.

SCALABLE PLATFORM

We are engineering a modeling platform that is scalable and built for growth. For processing data and for simulating the Earth system, it will exploit state-of-the-art algorithms to run on the world's fastest supercomputers and on the cloud. It will be scalable to ever finer resolution globally, and its targeted high-resolution simulations will provide detailed local climate information where needed.

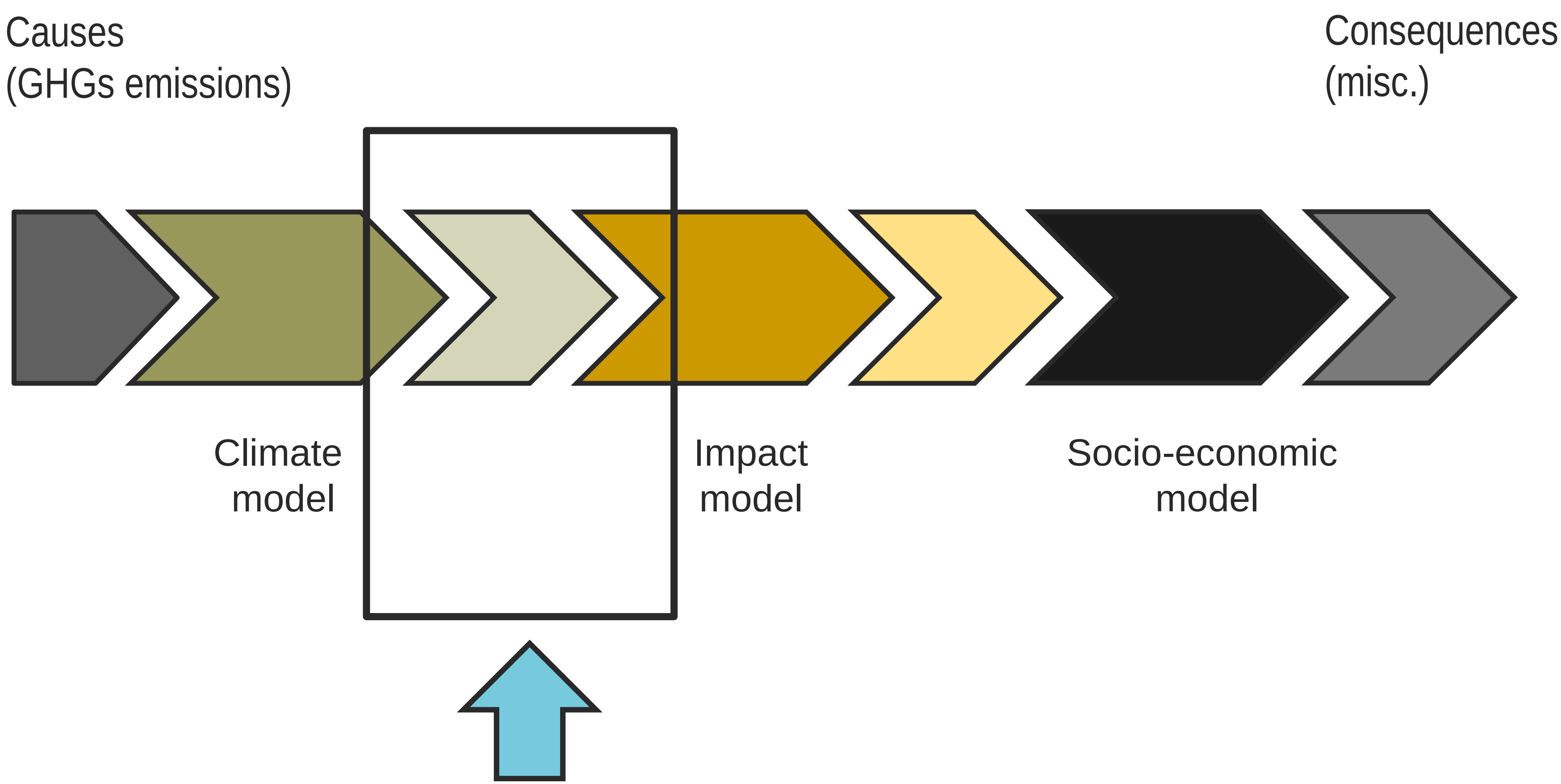
OPEN HUB

We are committed to transparency and open science principles. Our modeling platform is open source, and our results are available to the public. We will provide interfaces to our modeling platform so that it can become the anchor of an ecosystem of front-end apps. These apps may provide detailed models, for example, of flood risks, risks of extreme heat, crop yields, and other climate impacts.

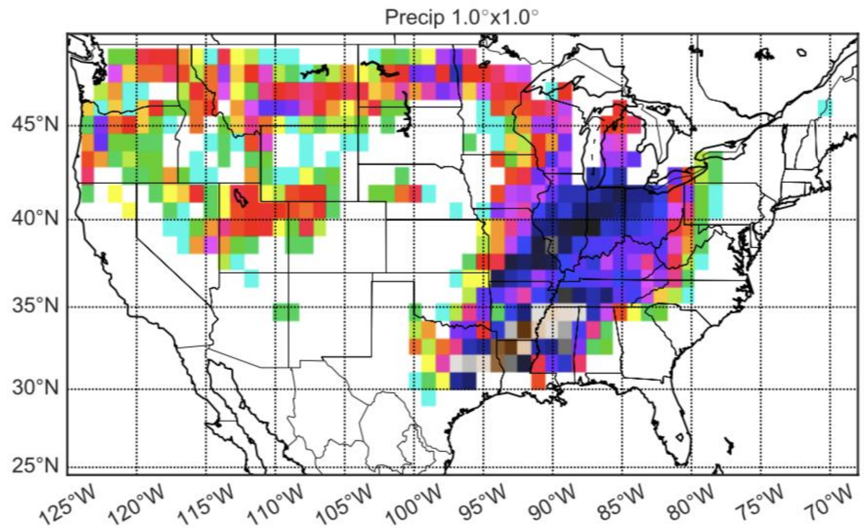
Agenda

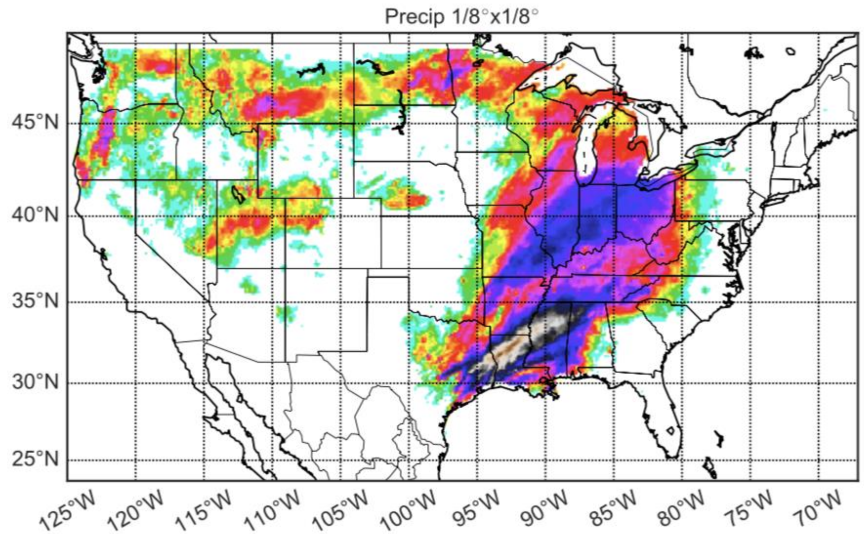
- Context
- A few case studies: statistical downscaling
- A few challenges

Context

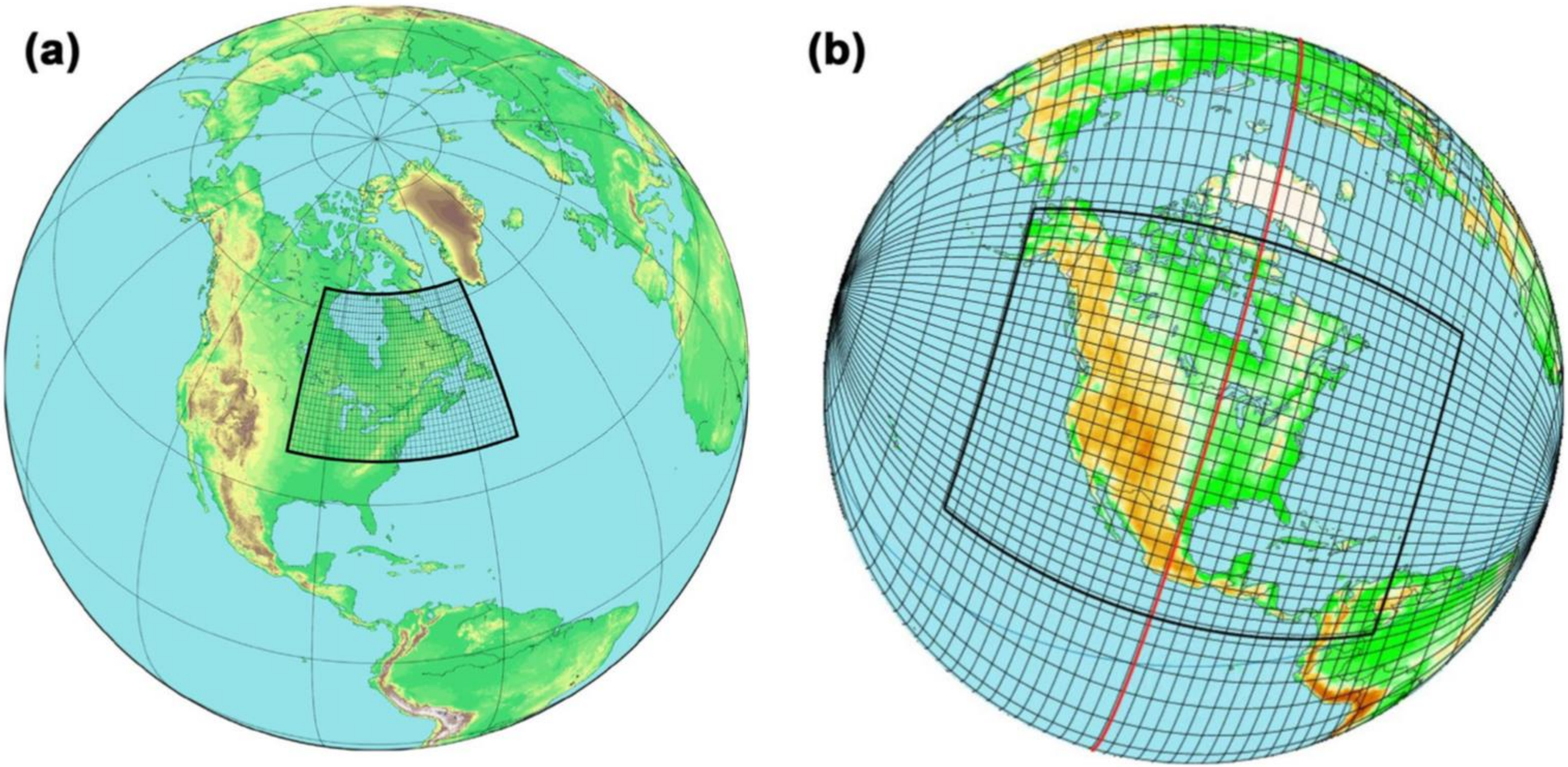

Causes
(GHGs emissions)
Consequences
(misc.)
Climate
model
Impact
model
Socio-economic
model

Context

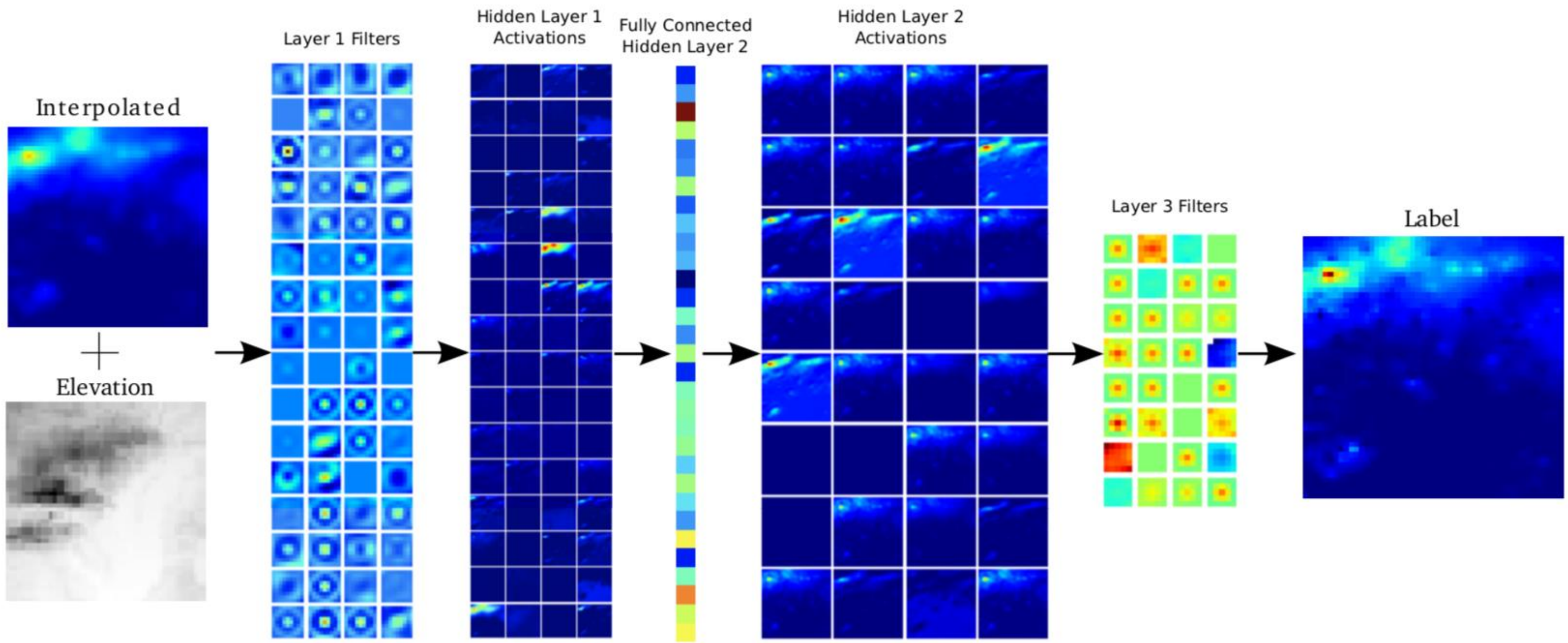

Precip 1.0°x1.0°
45°N
40°N
35°N
30°N
25°N
125°W
120°W
115°W
110°W
105°W
100°W
95°W
90°W
85°W
80°W
75°W
70°W


Precip 1/8°x1/8°
45°N
40°N
35°N
30°N
25°N
125°W
120°W
115°W
110°W
105°W
100°W
95°W
90°W
85°W
80°W
75°W
70°W

Climate models

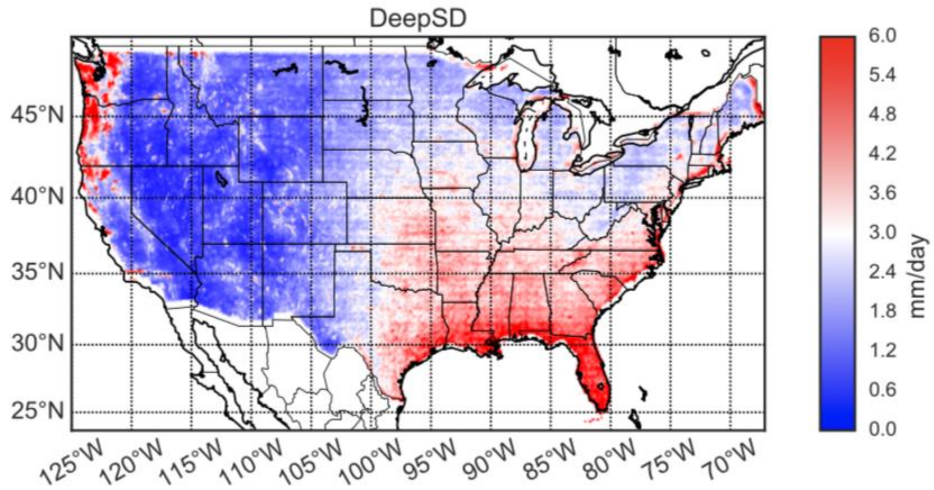


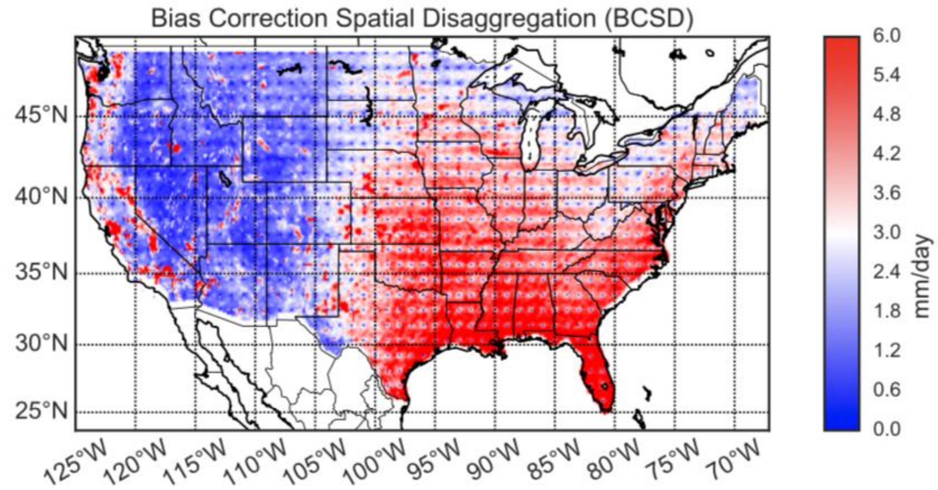
A deep convolutional net for downscaling

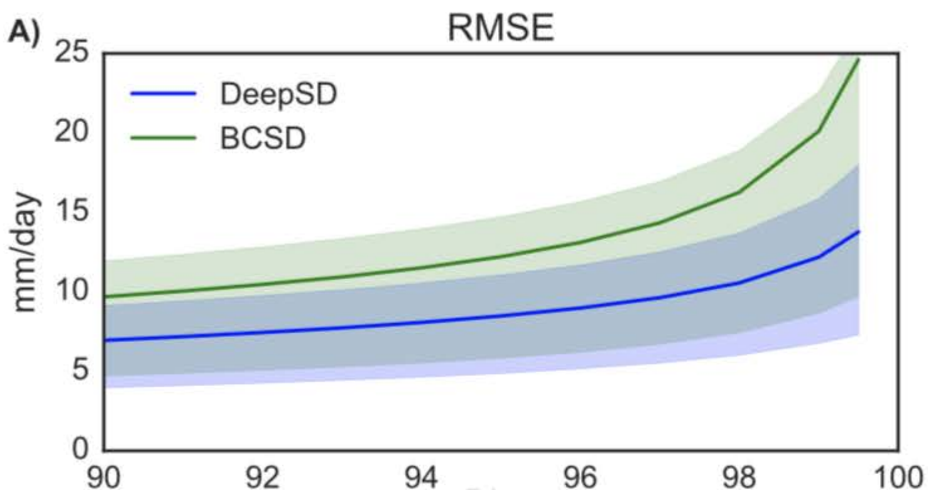


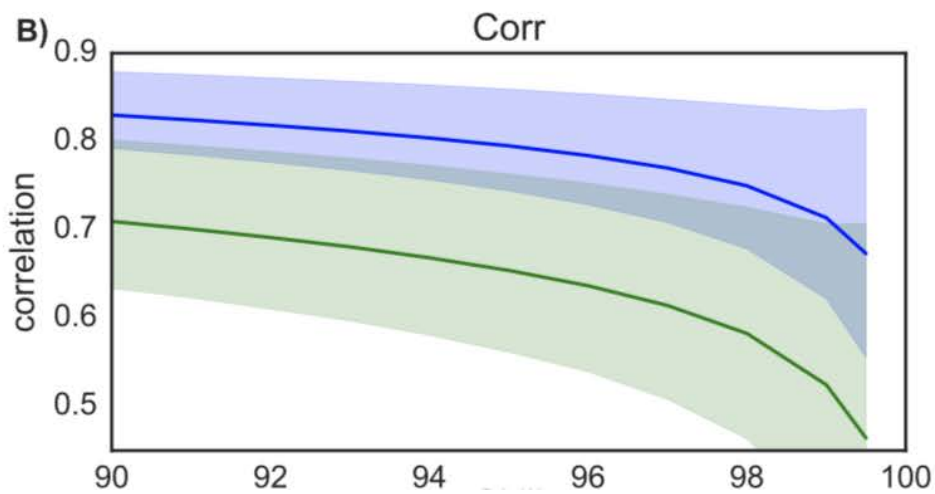
Vandal et al. 2017

Results


DeepSD
mm/day


Bias Correction Spatial Disaggregation (BCSD)
mm/day

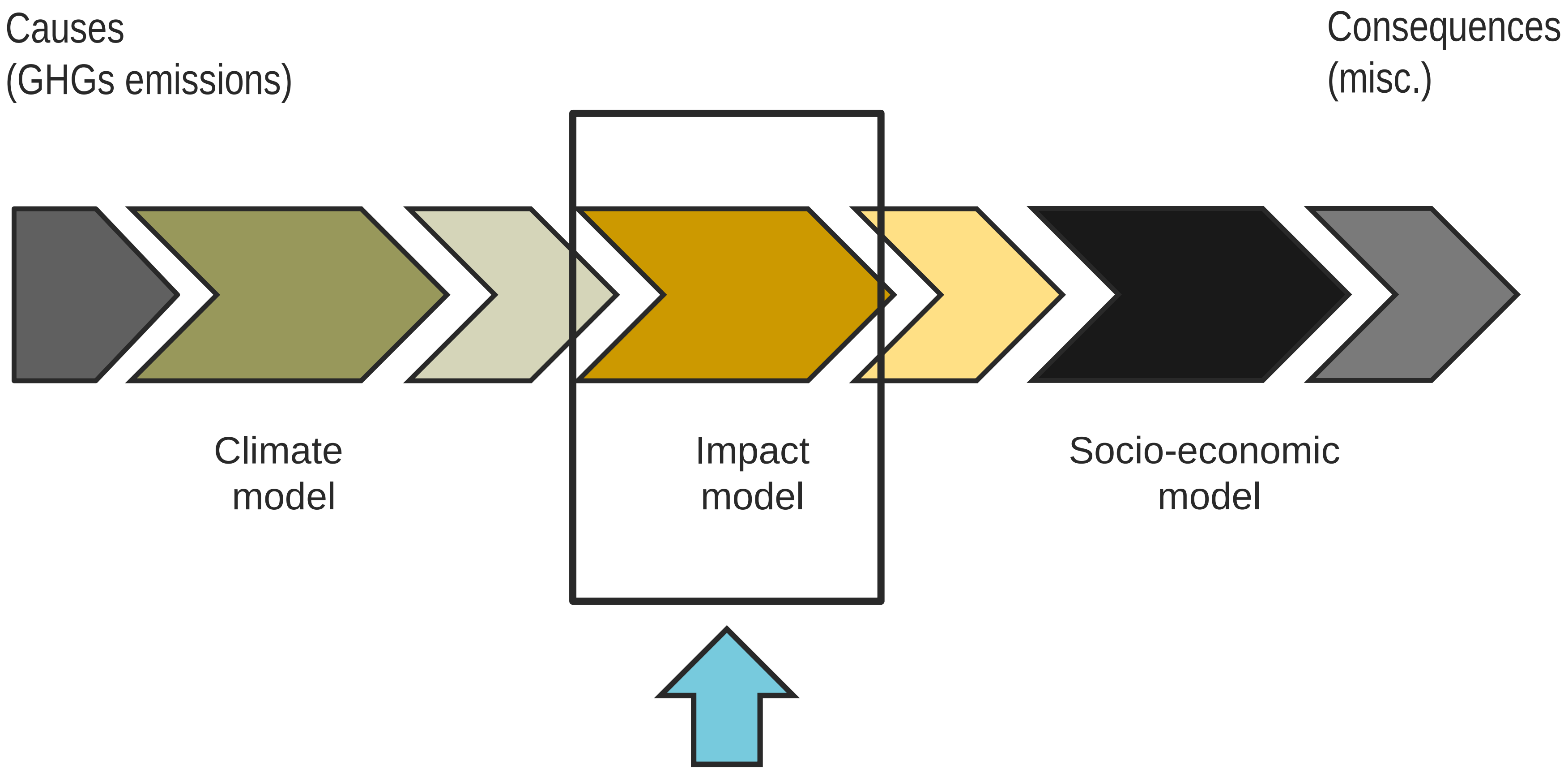

A)
RMSE
DeepSD
BCSD
mm/day


B)
Corr
correlation

Agenda

- Context
- A few case studies: flood mapping
- A few challenges

Context

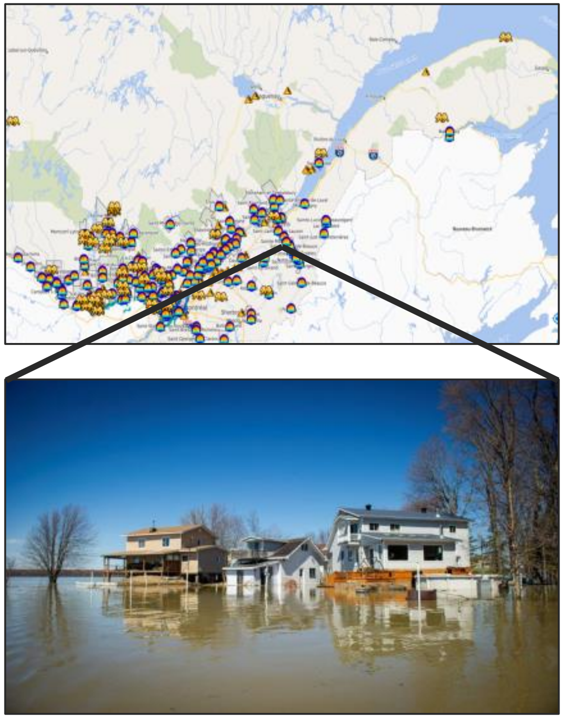


Causes
(GHGs emissions)
Consequences
(misc.)
Climate
model
Impact
model
Socio-economic
model

Spring floods in Quebec

April 2017

April 2019

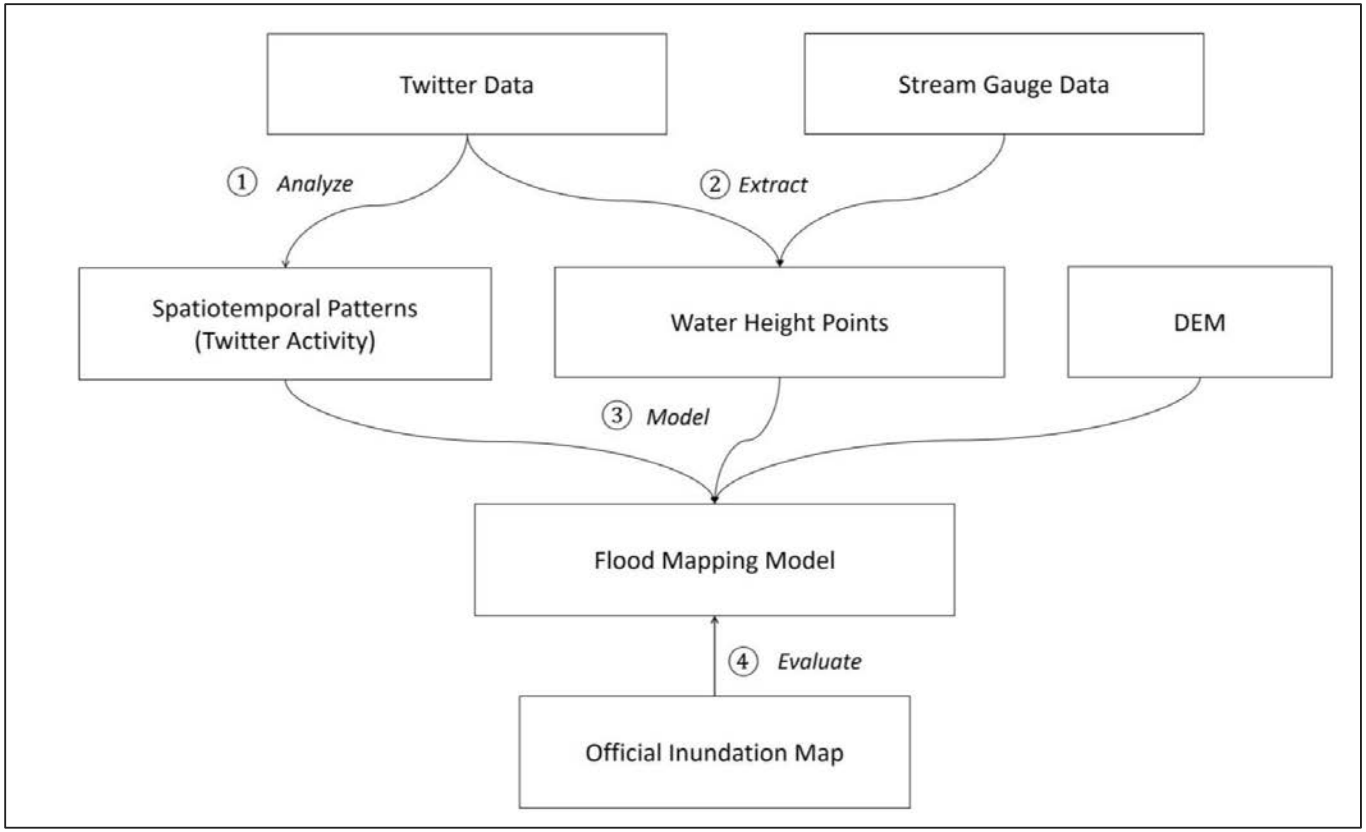


Source: GéoMSP, Québec

Challenges in observing floods

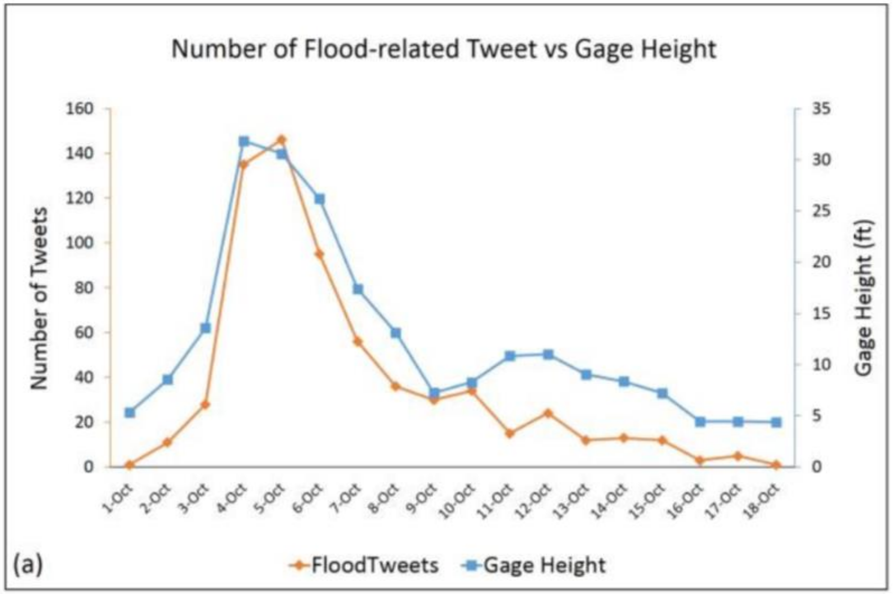
Remote Sensors (i.e. spaceborne or airborne)	**Citizens Sensors** (i.e. local disaster participants)
• Heavily used before, during, and after disaster • Systems and experts (with limited local area knowledge) in places to interpret data quickly from outside AOI • Contextual limitations: no cultural, economic, or social information – only visual/spectral analysis • Temporal limitations: satellite paths do not allow capture between passes (sometimes days) • Spatially limited: resolution may be too coarse. May need to stich images to capture impacts • Environmental limitations: imagery may be obscured by clouds, smoke, dust, or debris	• Applications in emergency management in infancy • Data contributed by people more intimate knowledge about local area • Contextually specific: Locals hear, see, feel, synthesize, and report impacts • Near-real time reporting enables changes to be tracked as they occur • Spatially proximal, very fine scale can be aggregated up and used for detailed spatial analysis • Environmental limitations only controlled by conditions on the ground, ability to see impacts • Spatial limitations: skewed in space, more populated areas tend to have more data • Data quality limitations: unstructured, noisy, and high uncertainty. Need validation

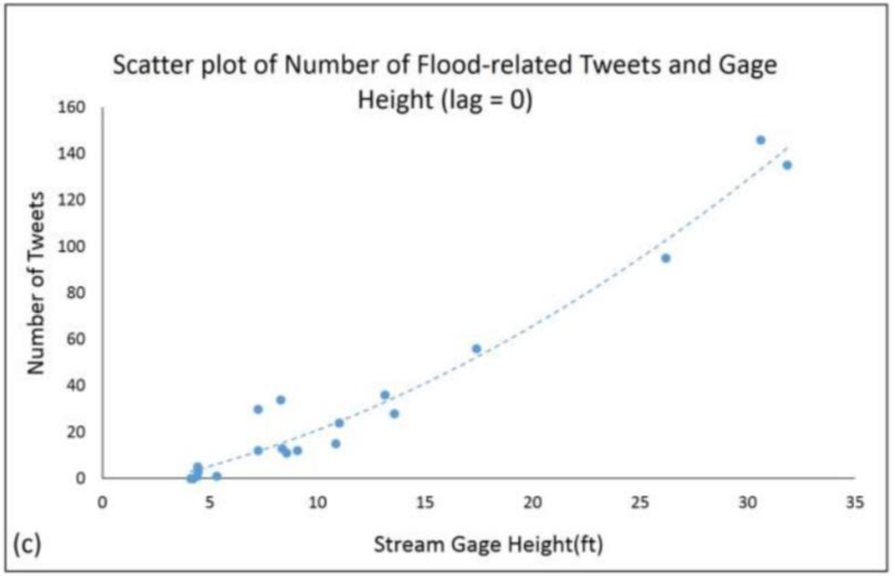
Proposed approach leveraging Twitter



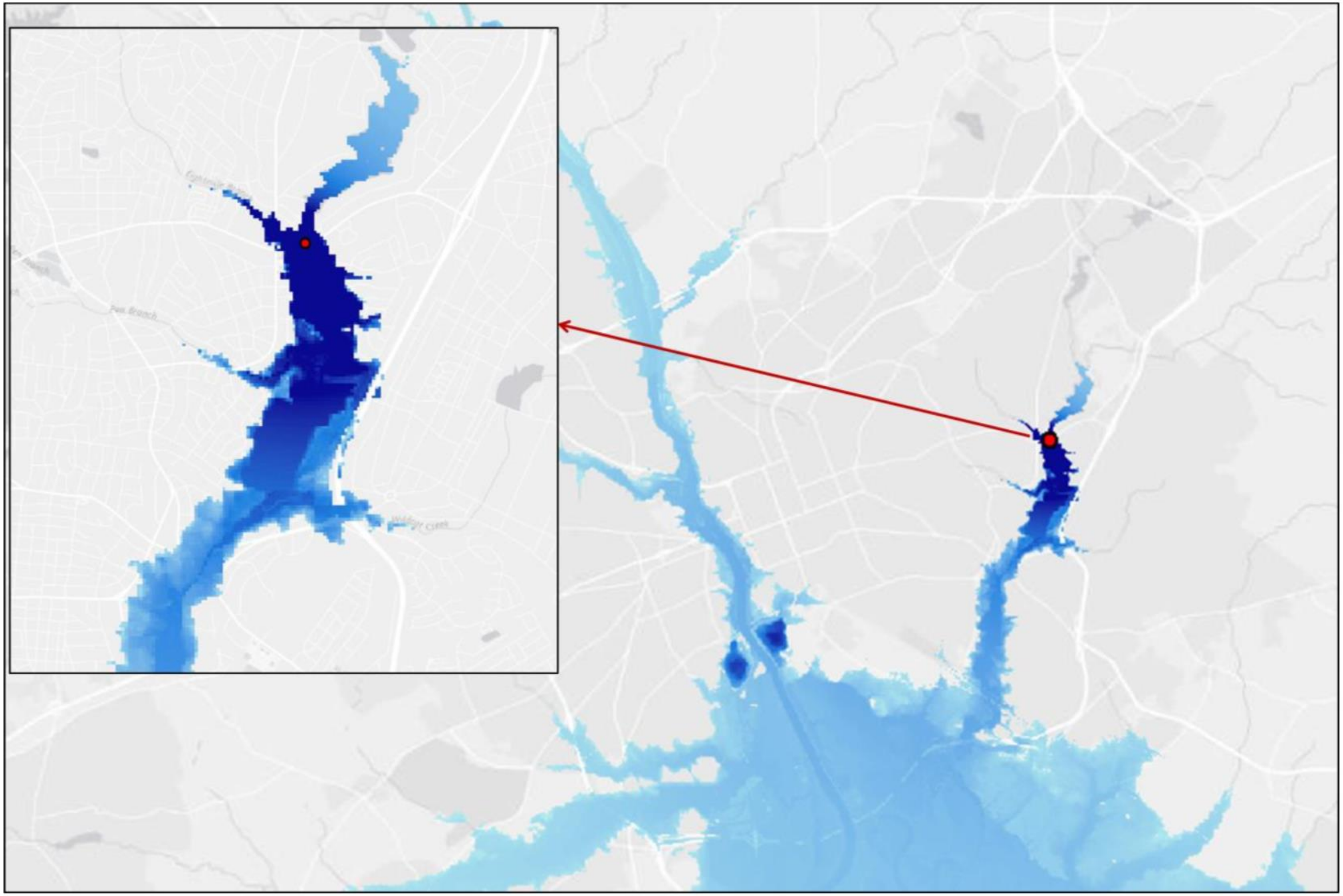
Li et al. 2017

Results

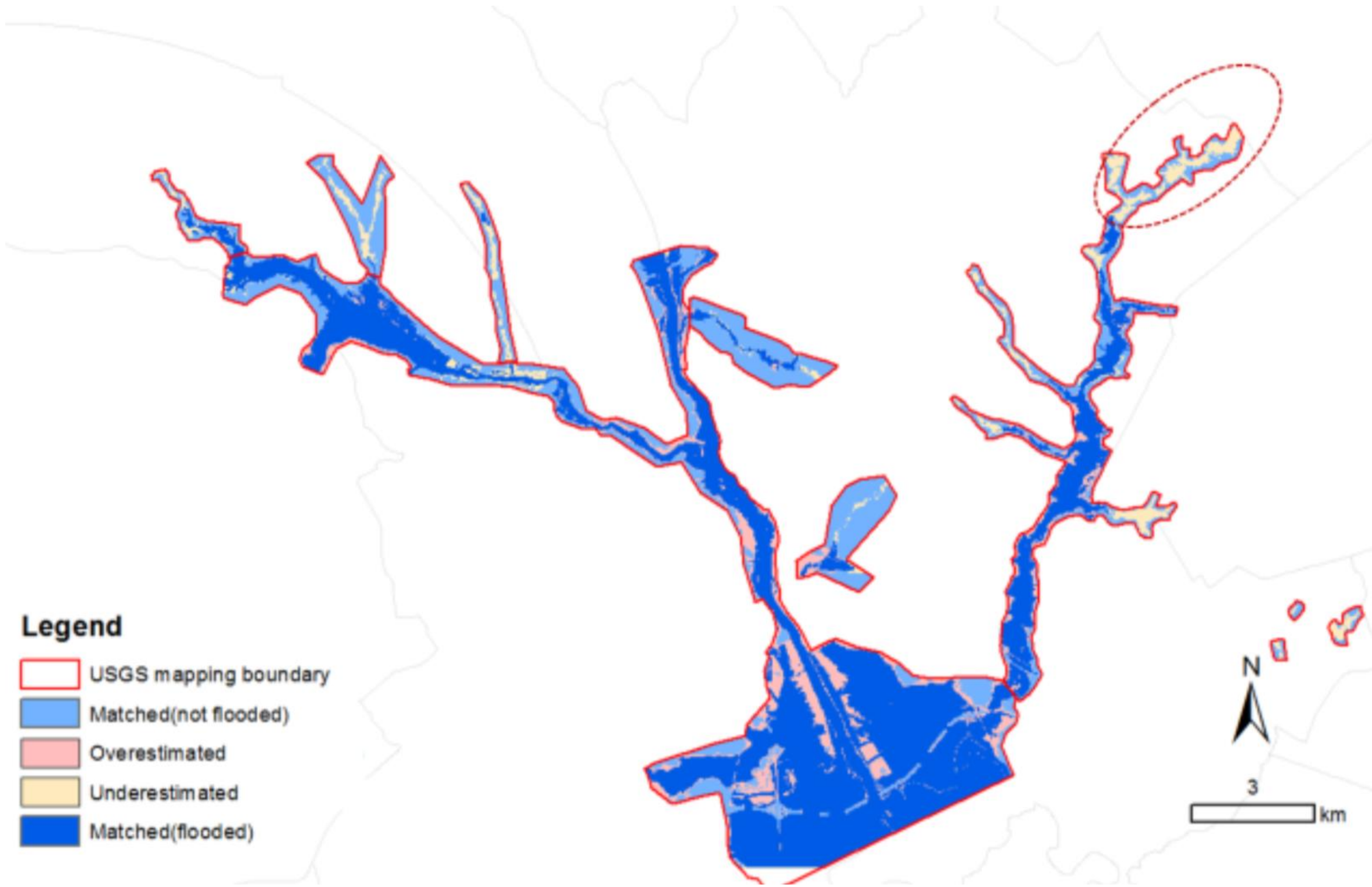

Number of Flood-related Tweet vs Gage Height
Number of Tweets
Gage Height (ft)
160
140
120
100
80
60
40
20
0
35
30
25
20
15
10
5
0
1-Oct
2-Oct
3-Oct
4-Oct
5-Oct
6-Oct
7-Oct
8-Oct
9-Oct
10-Oct
11-Oct
12-Oct
13-Oct
14-Oct
15-Oct
16-Oct
17-Oct
18-Oct
(a)
FloodTweets
Gage Height


Scatter plot of Number of Flood-related Tweets and Gage Height (lag = 0)
Number of Tweets
160
140
120
100
80
60
40
20
0
0
5
10
15
20
25
30
35
(c)
Stream Gage Height(ft)

Results



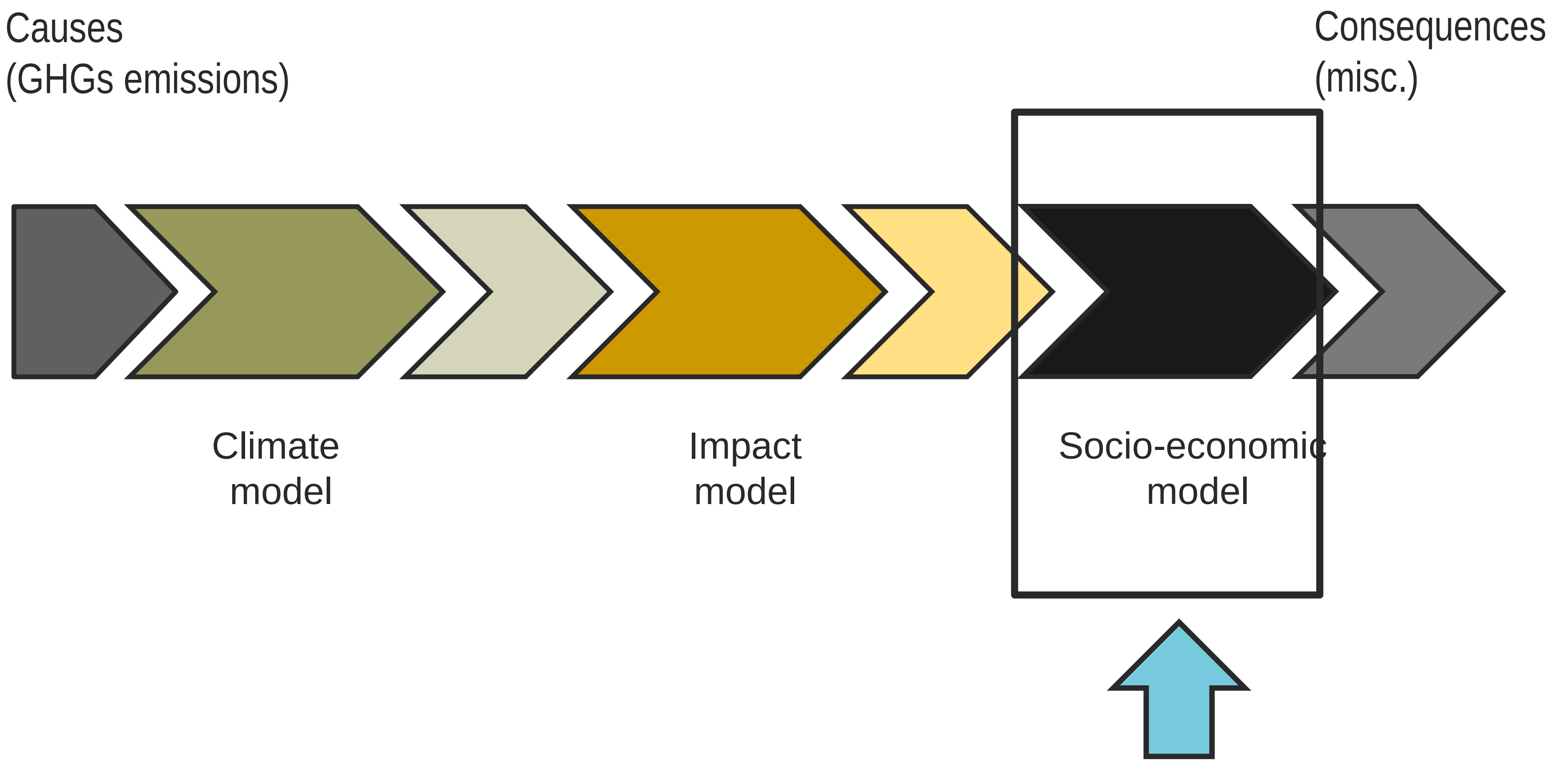
Skill


Legend
USGS mapping boundary
Matched(not flooded)
Overestimated
Underestimated
Matched(flooded)
N
3
km

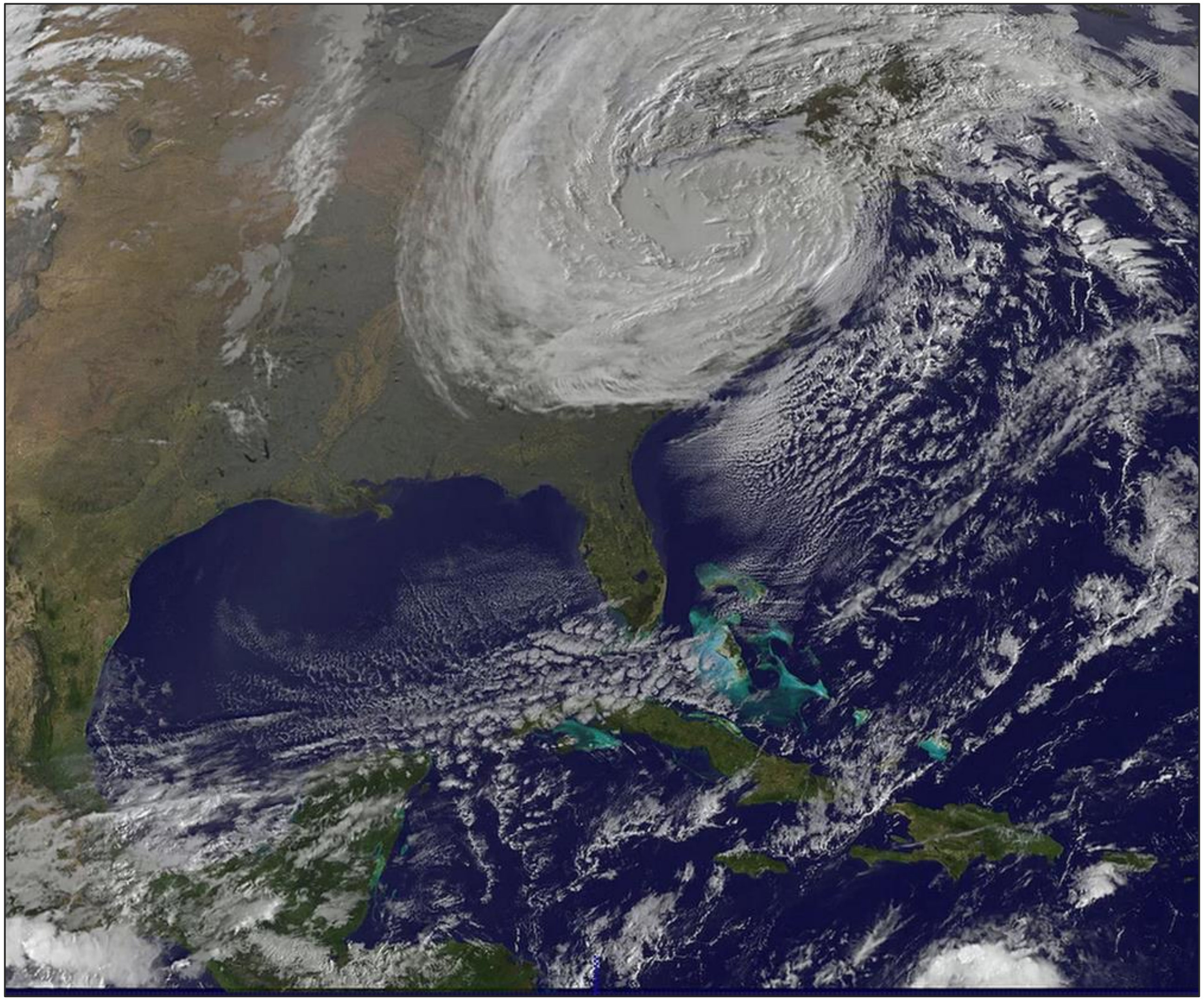
Agenda

- Context
- A few case studies: damage assessment
- A few challenges

Context


Causes
(GHGs emissions)
Consequences
(misc.)
Climate
model
Impact
model
Socio-economic
model

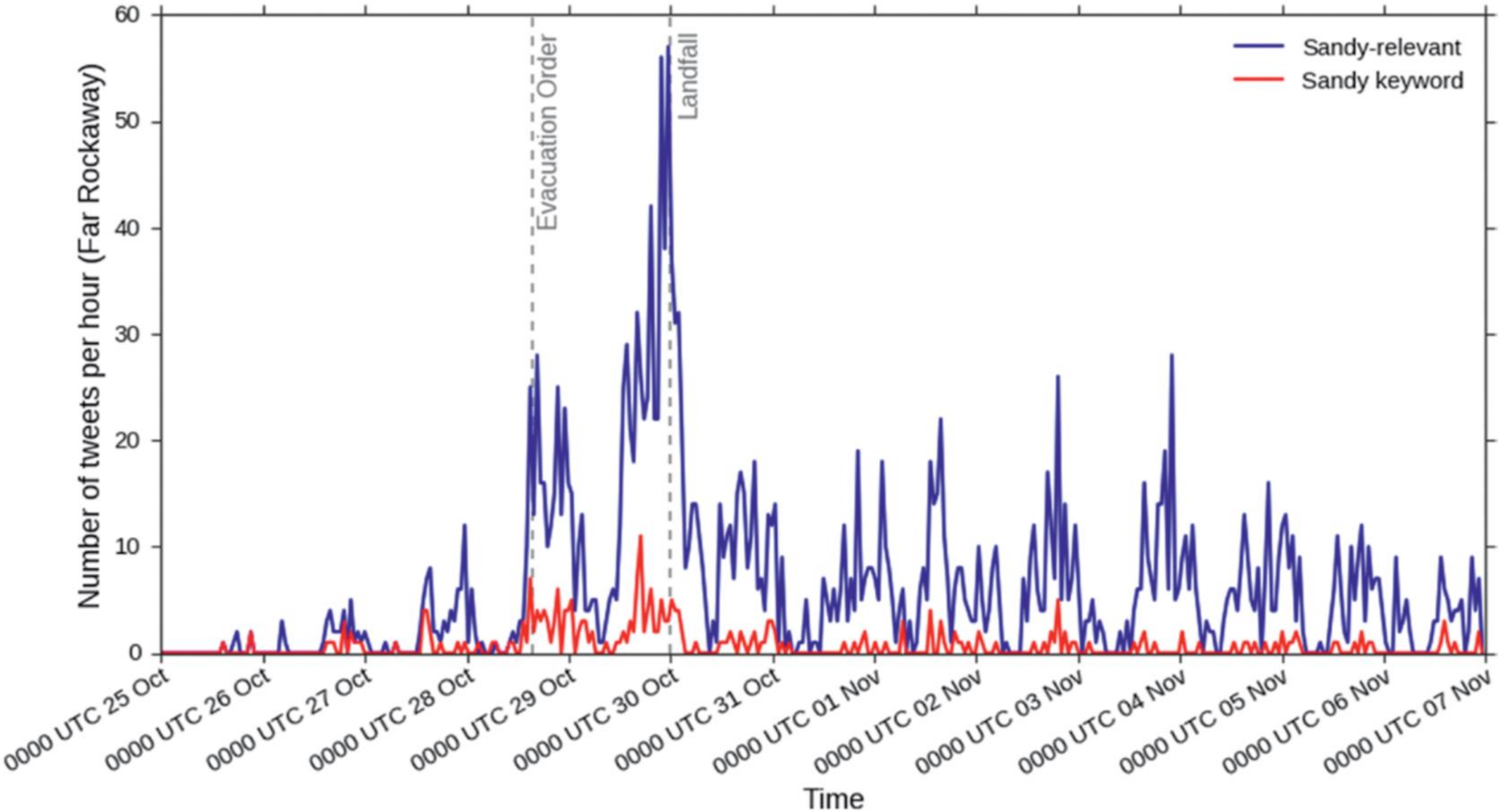
Hurricane Sandy



Keywords appearing in tweets

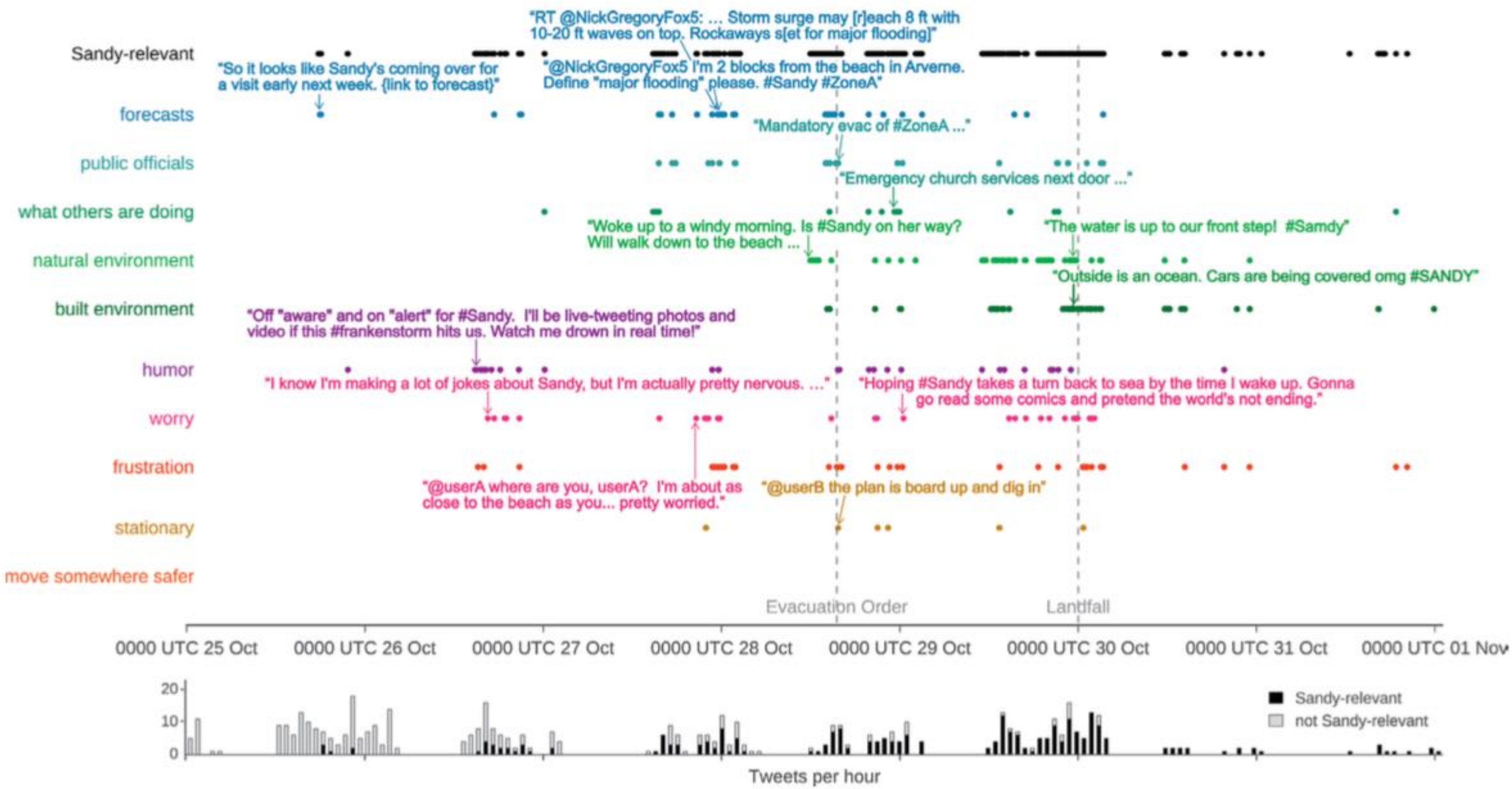
Keyword	Count
power	4 825 717
sandy	4 745 099
hurricane	4 680 290
weather	3 333 025
storm	2 555 196
gas	1 991 524
Governor	498 135
stay safe	484 732
recovery	431 591
climate	420 217
FEMA	329 789
flooding	264 132
no power	261 998
climate change	236 009
wall st	233 411
blackout	213 520
mta	206 504
frankenstorm	205 467
Cuomo	92 014
prayforusa	91 293

Tweets during the event

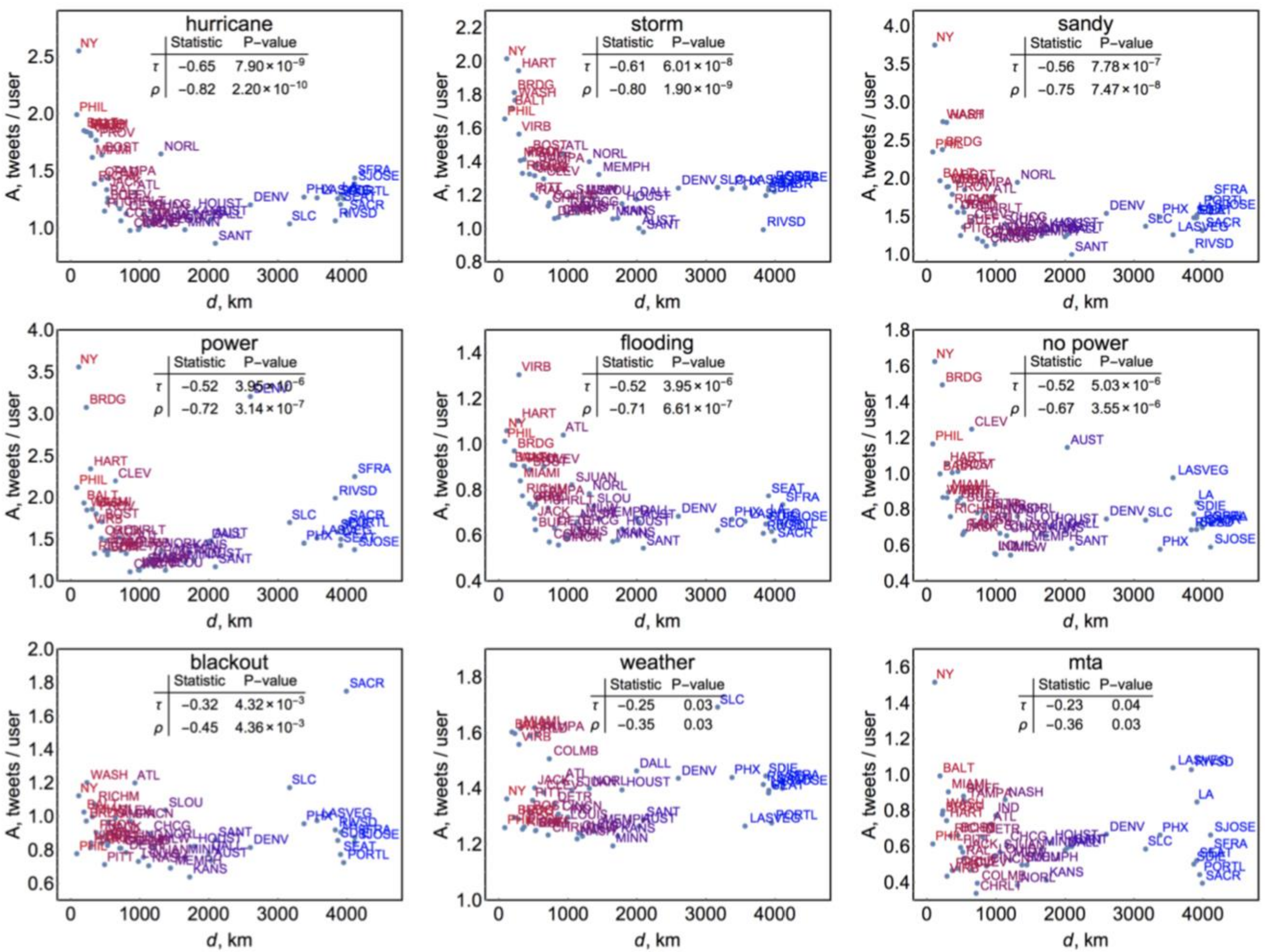


Morss et al. 2017

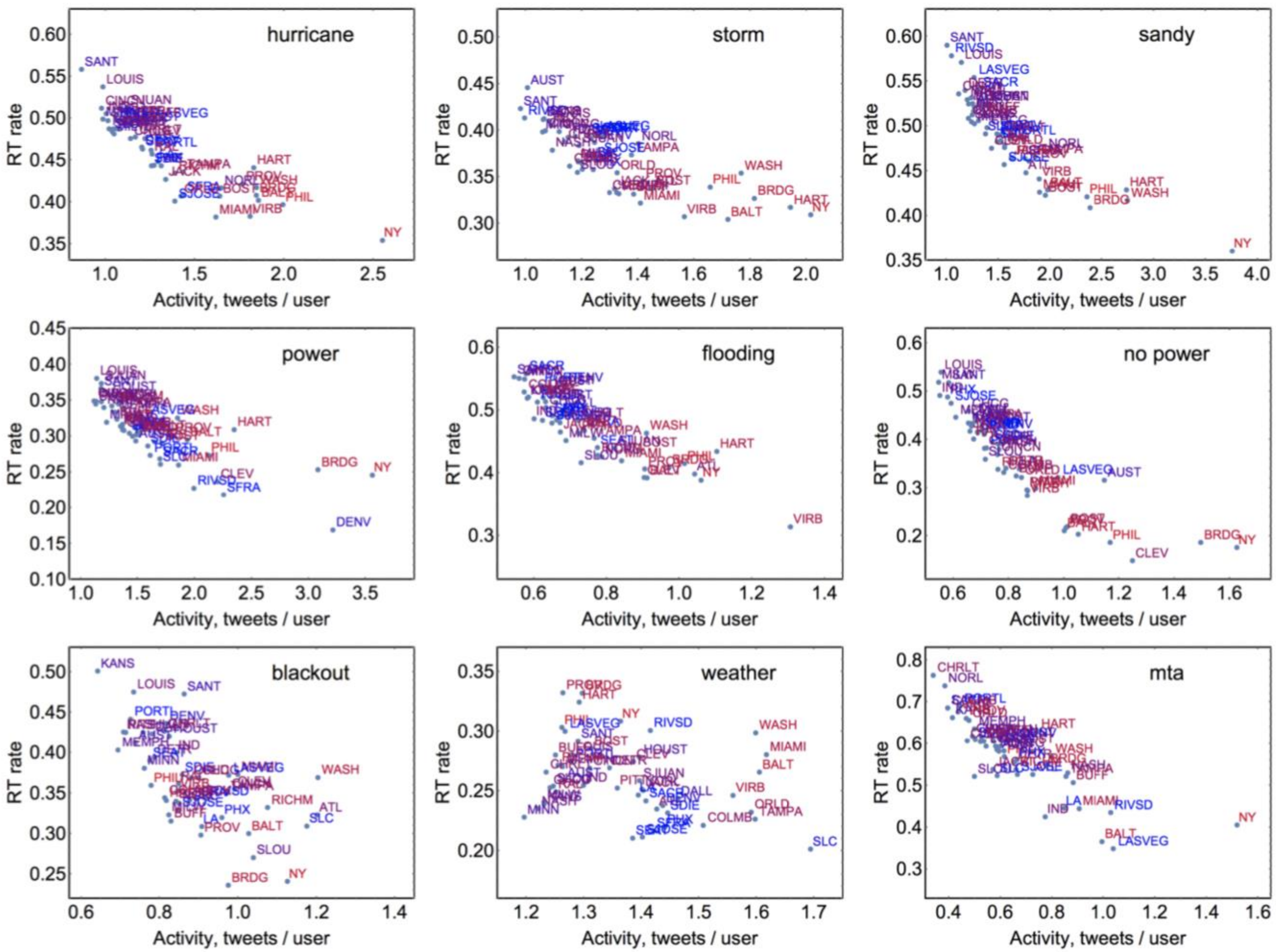
Classifying tweets



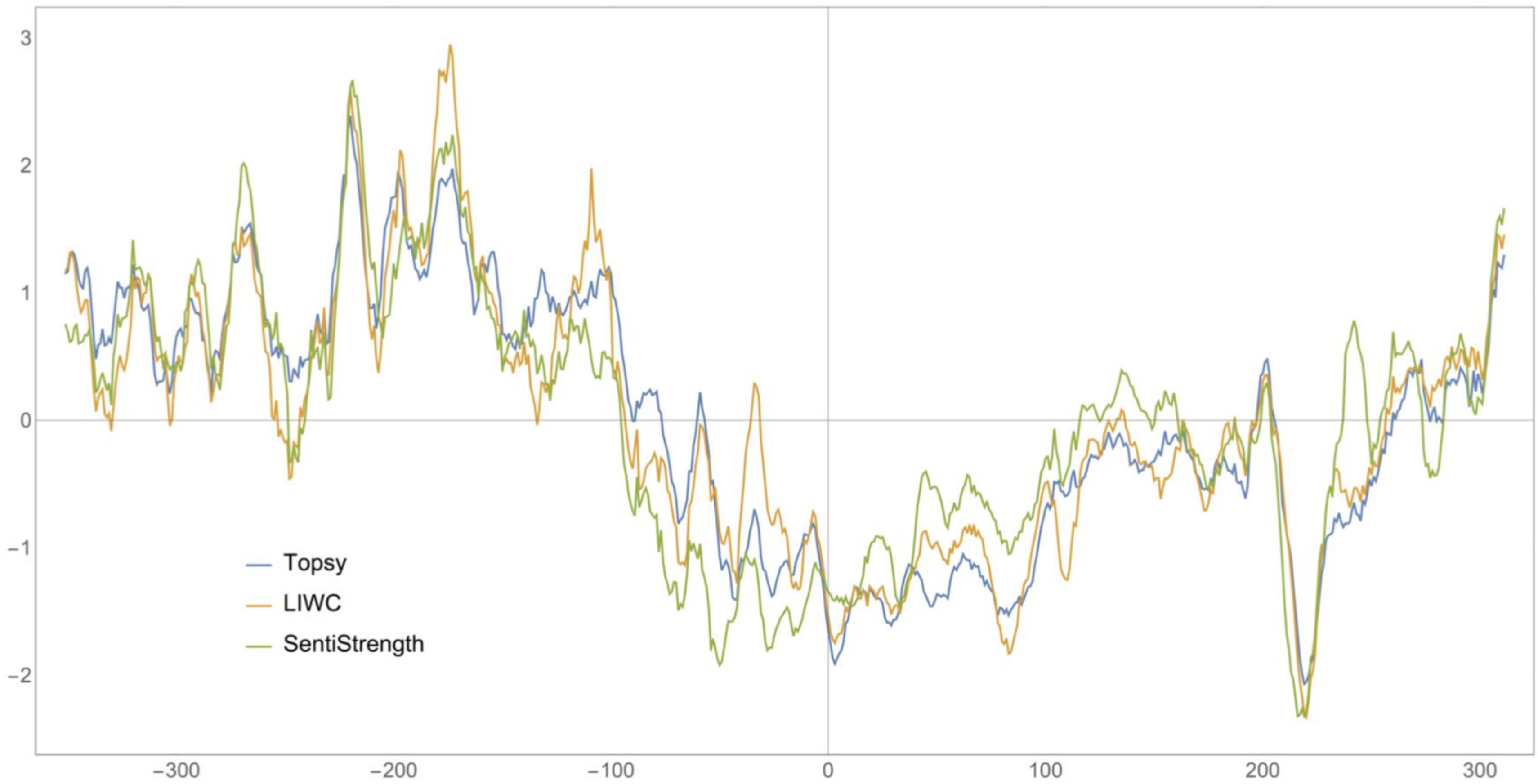
Quantifying the relevance of keywords



Quantifying the relevance of keywords

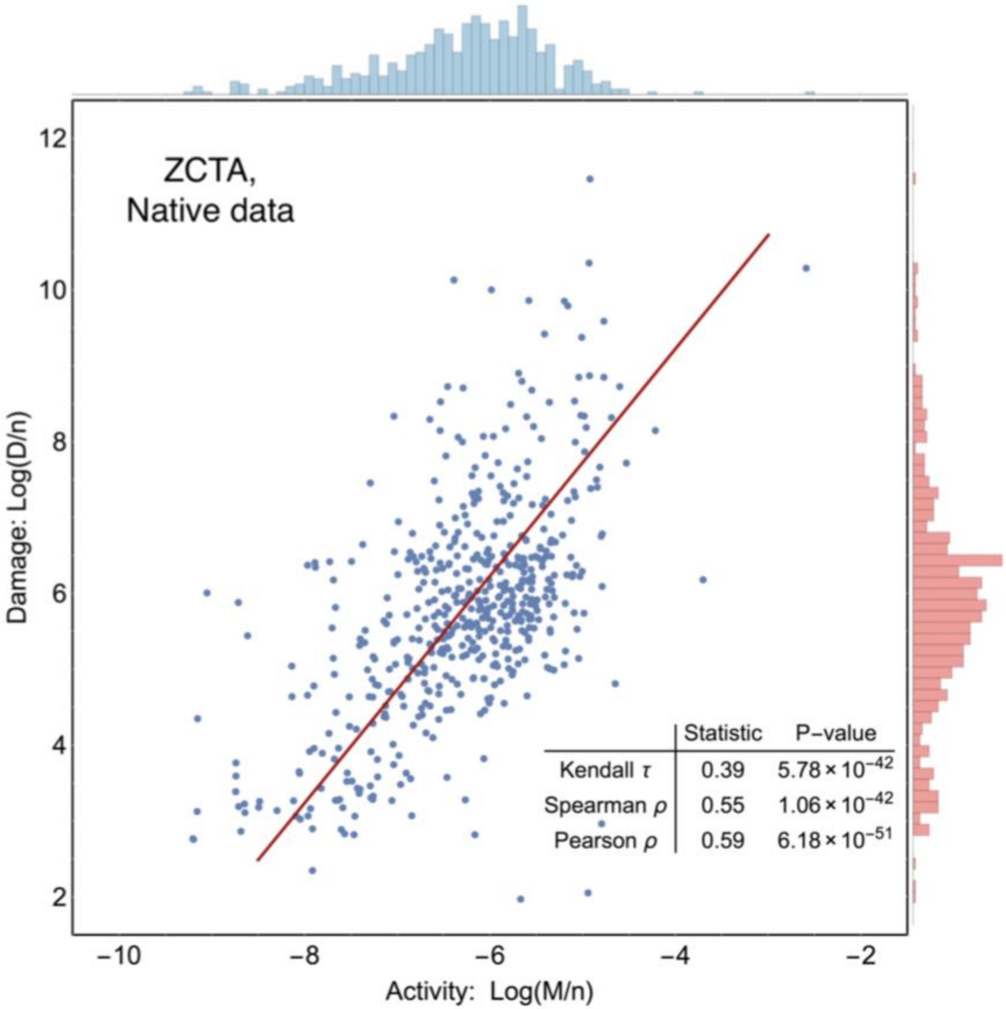


Sentiment indexes



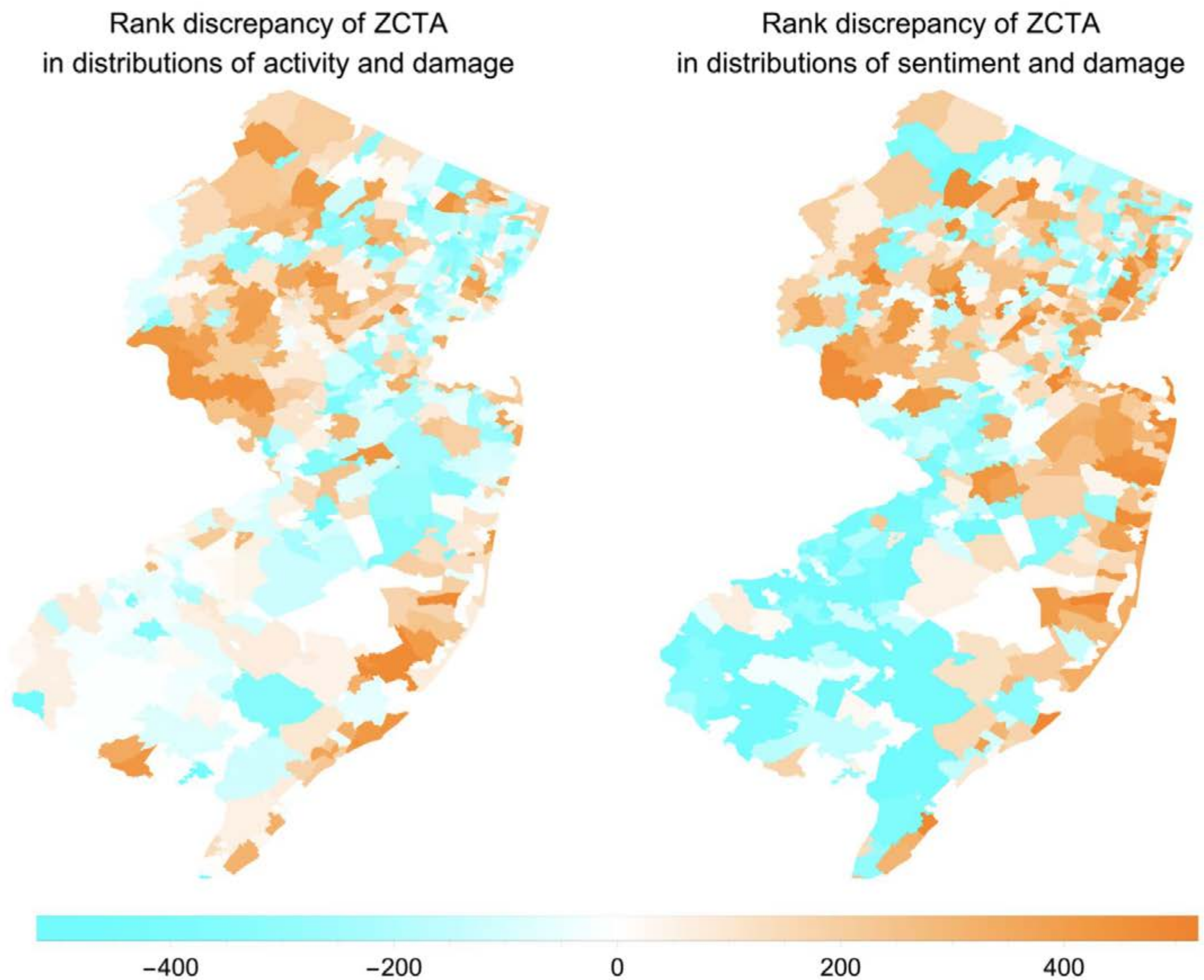
Hours

Impact indexes



	Statistic	P−value
Kendall τ	0.39	5.78×10^{-42}
Spearman ρ	0.55	1.06×10^{-42}
Pearson ρ	0.59	6.18×10^{-51}

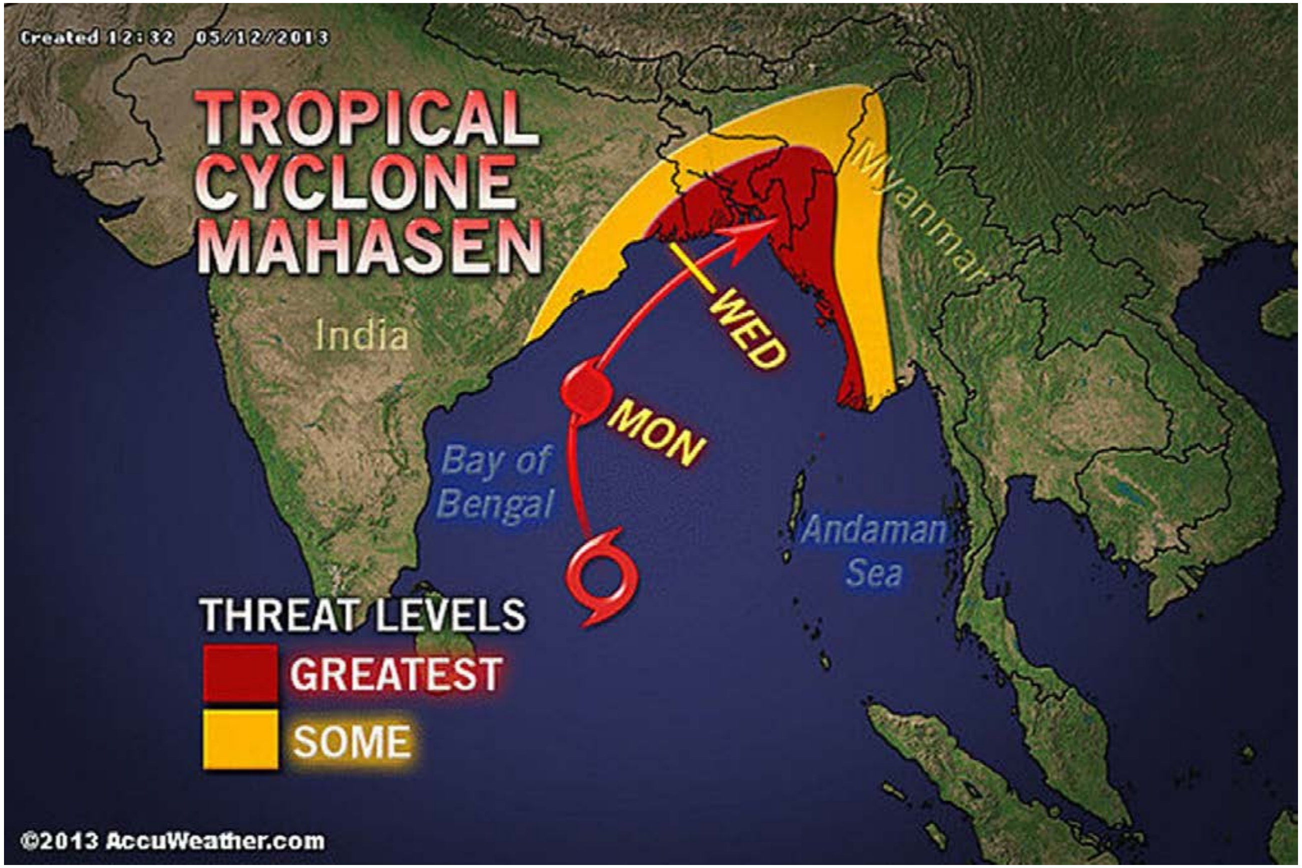
Skill


Rank discrepancy of ZCTA
in distributions of activity and damage
Rank discrepancy of ZCTA
in distributions of sentiment and damage
−400
−200
0
200
400

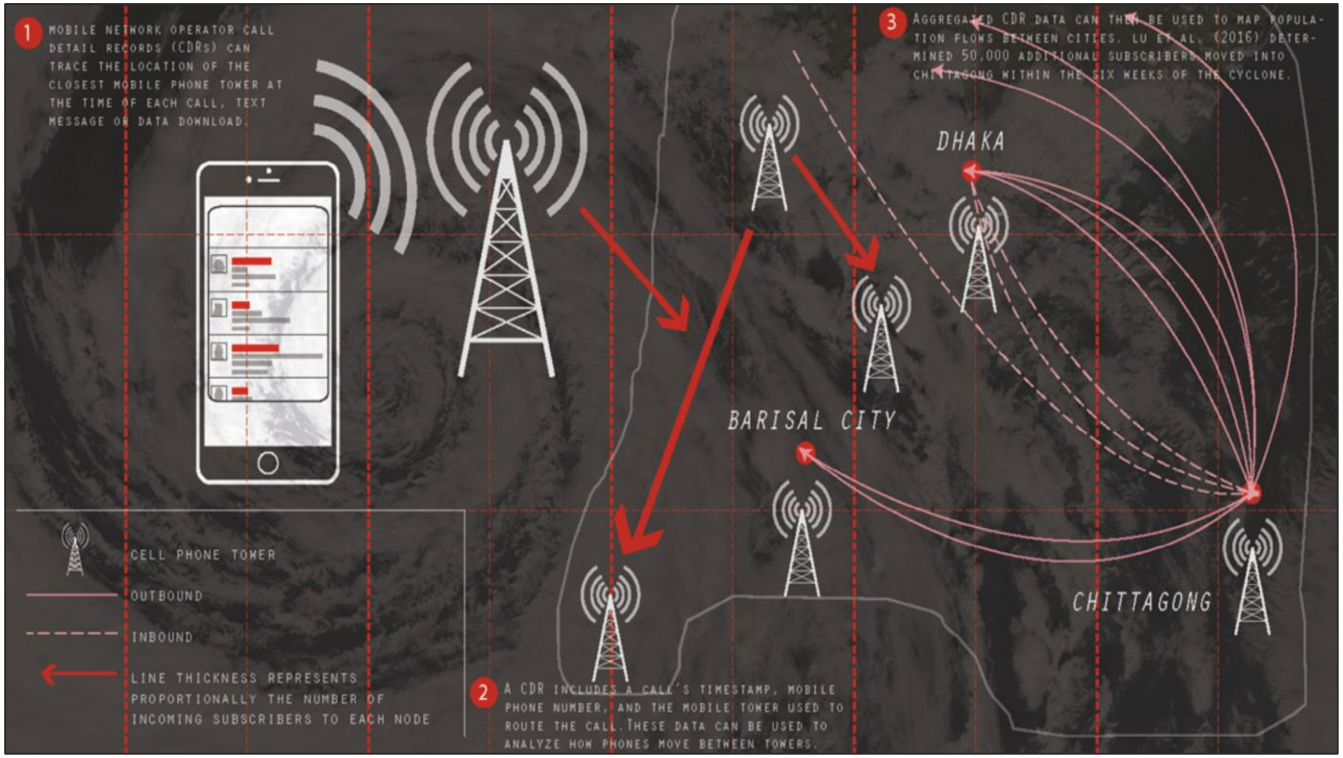
Agenda

- Context
- A few case studies: vulnerability assessment
- A few challenges

Tropical cyclone Mahasen, 2013

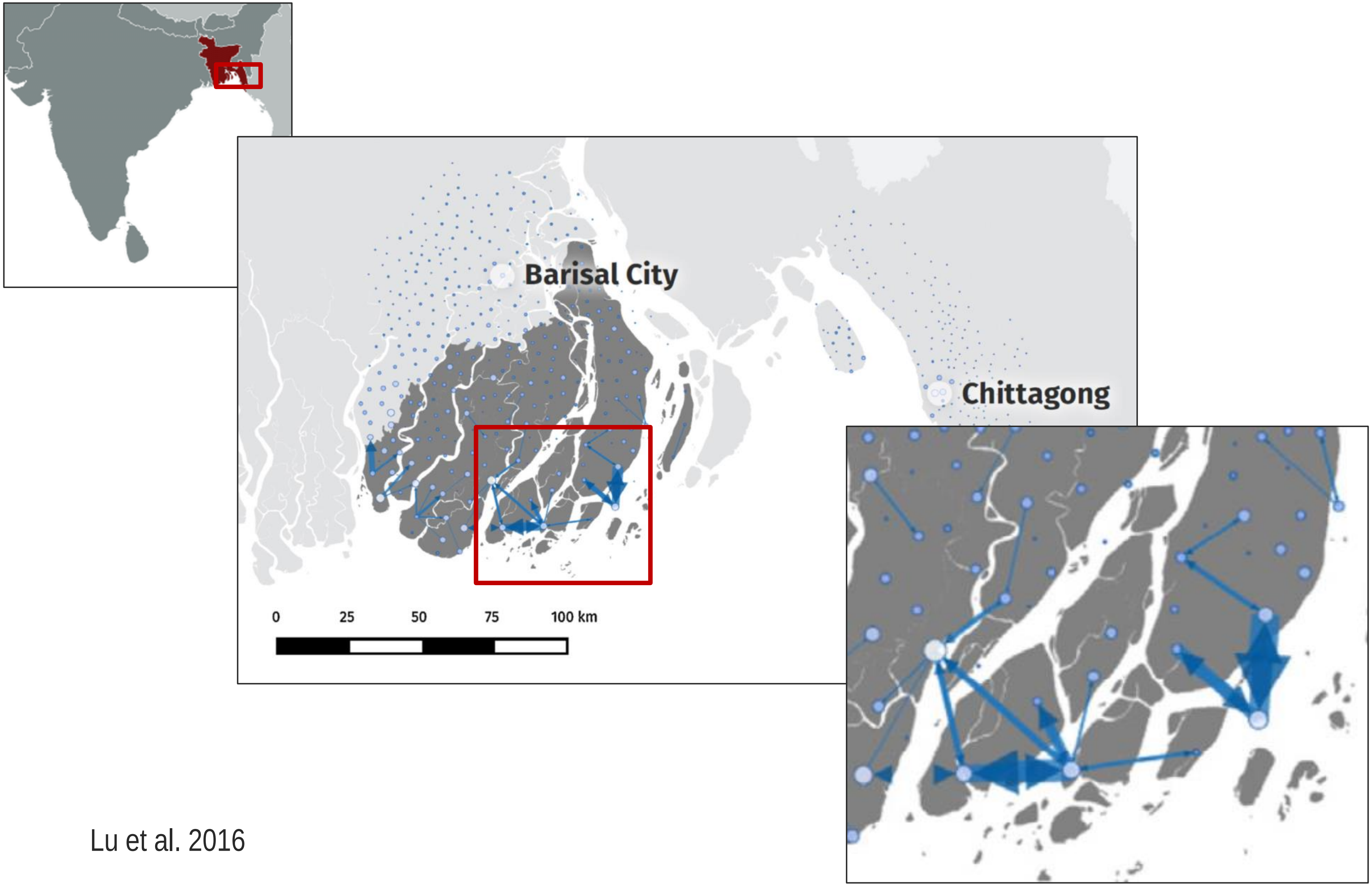


Using mobile network data to track mobility patterns



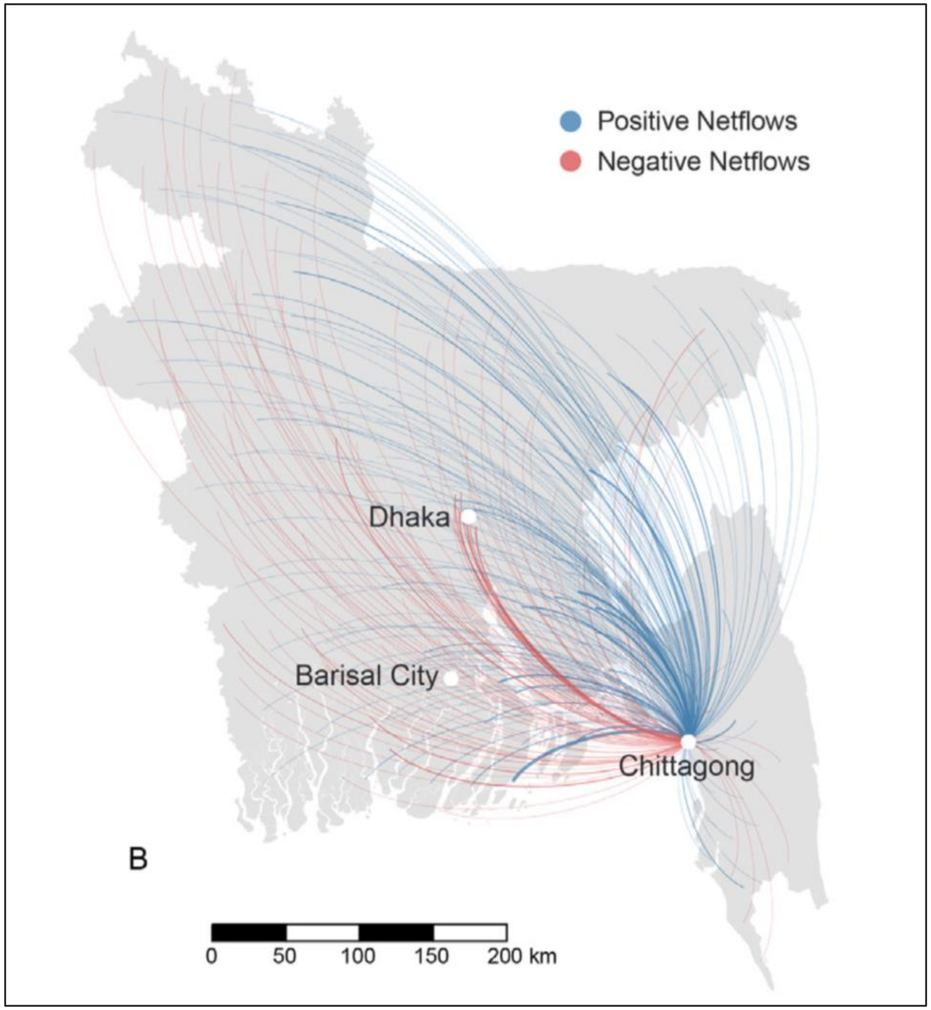
Ford et al. 2016

Population mobility during the storm



Lu et al. 2016

Population mobility after the storm

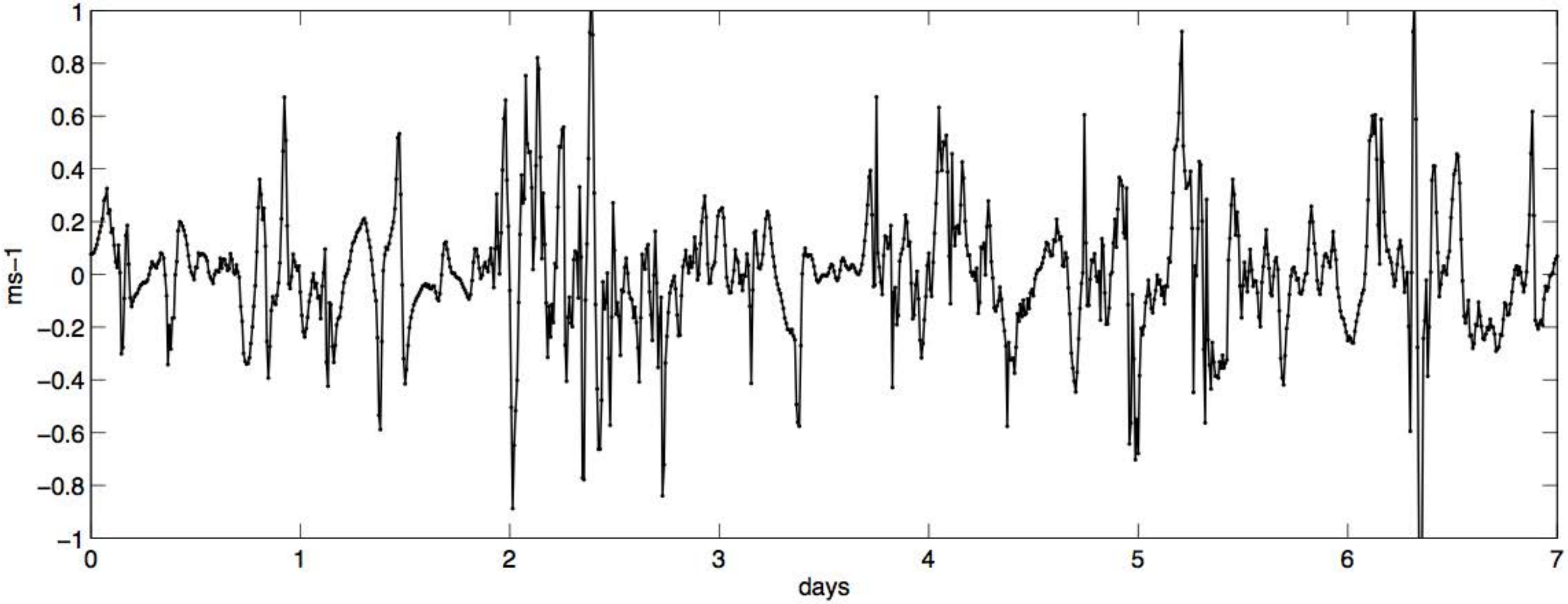


Agenda

- Context
- A few case studies: short term prediction
- A few challenges

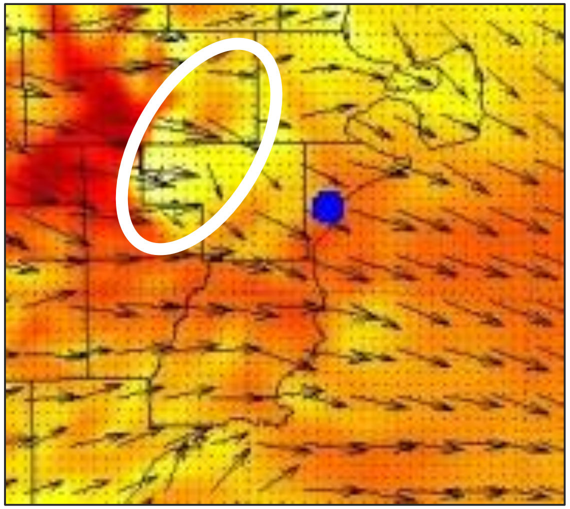
Wind power generation

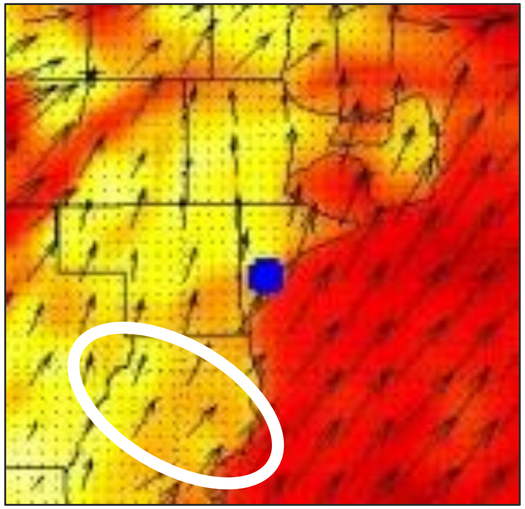
Windspeed variations



Use spatio-temporal patterns to predict wind

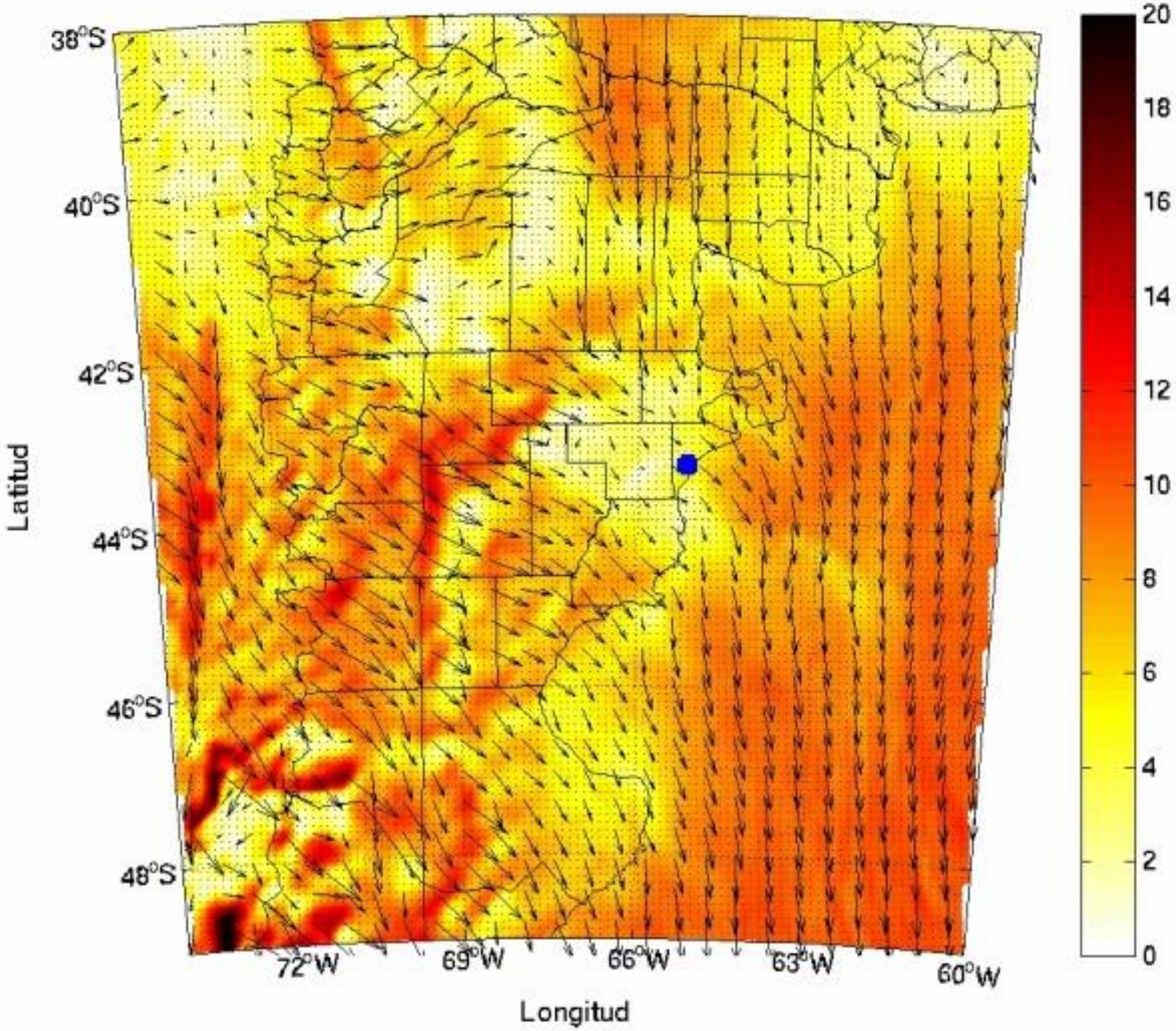
- Idea of “upstream prediction”



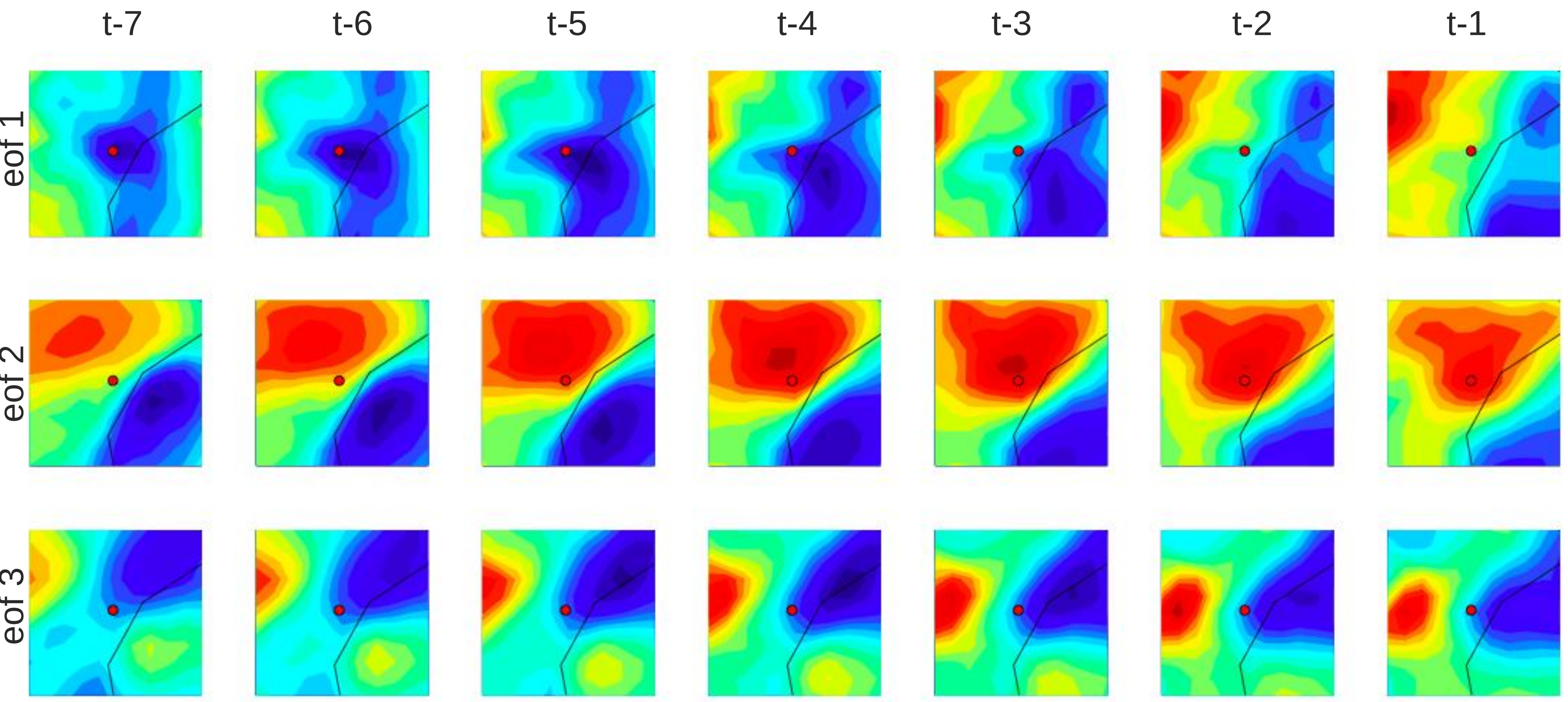


Hannart et al. (in prep.)

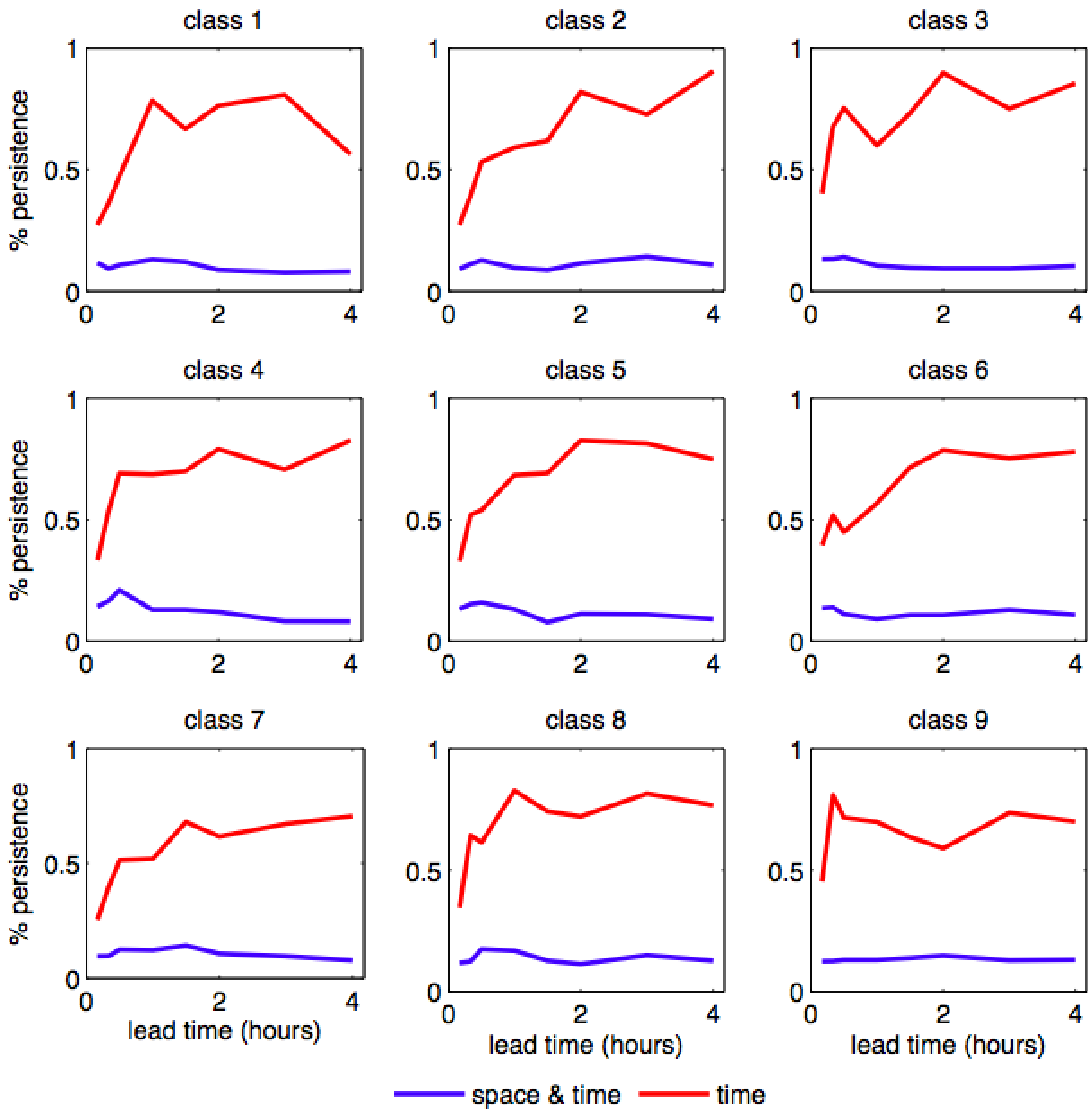
Data: long simulation **at 10’ time** step



Tool: spatio-temporal Gaussian process



Skill



class 1
class 2
class 3
class 4
class 5
class 6
class 7
class 8
class 9
% persistence
lead time (hours)
space & time
time

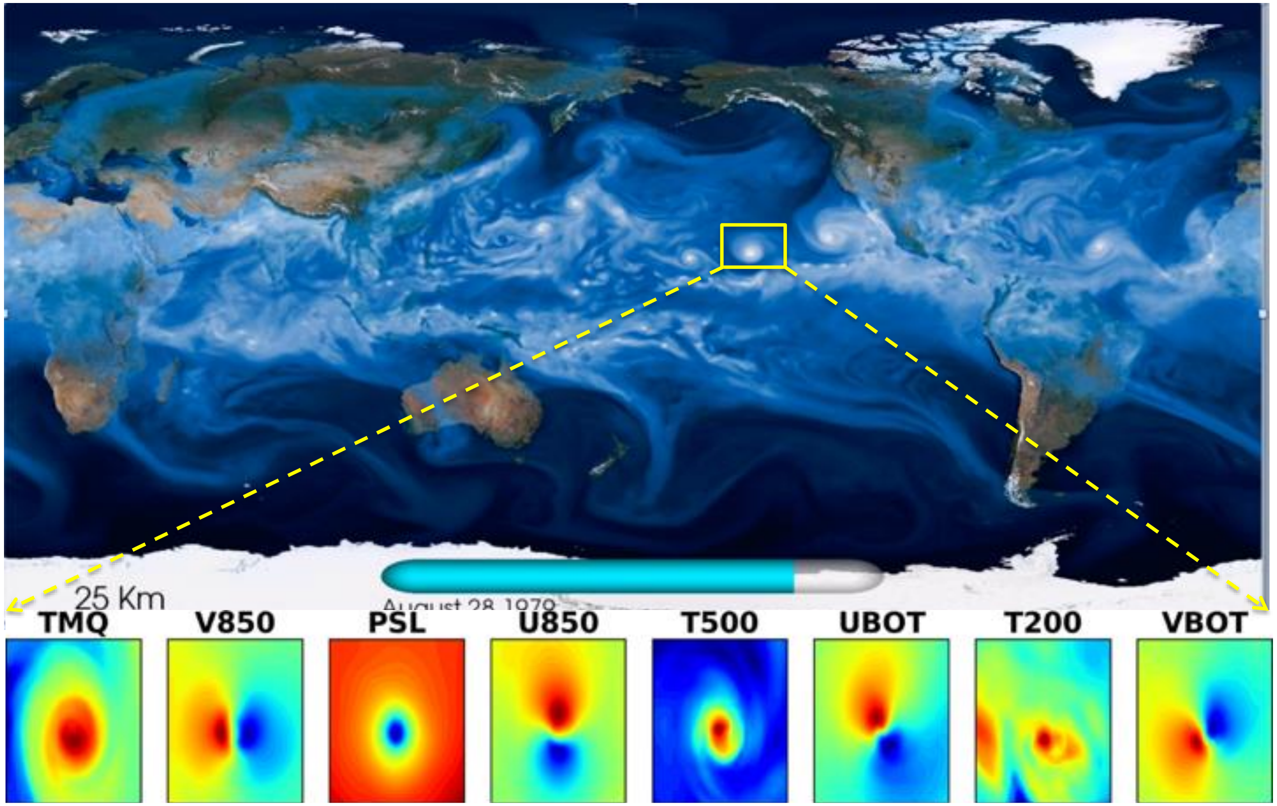
Agenda

- Context
- A few case studies: tracking weather systems
- A few challenges

Hurricane season 2017

8 September 2017 06.00pm GMT

Multi-channel image classification

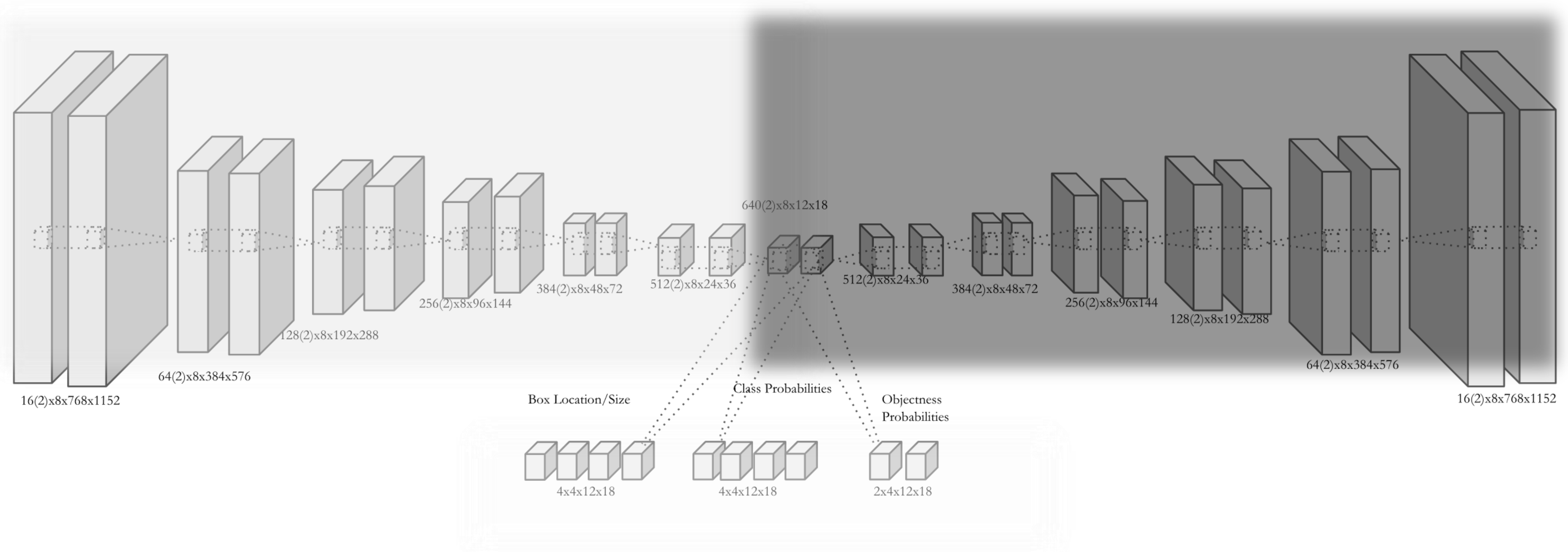


Prahbat et al. 2016

Semi-Supervised Convolutional Architecture

Encoder

Decoder



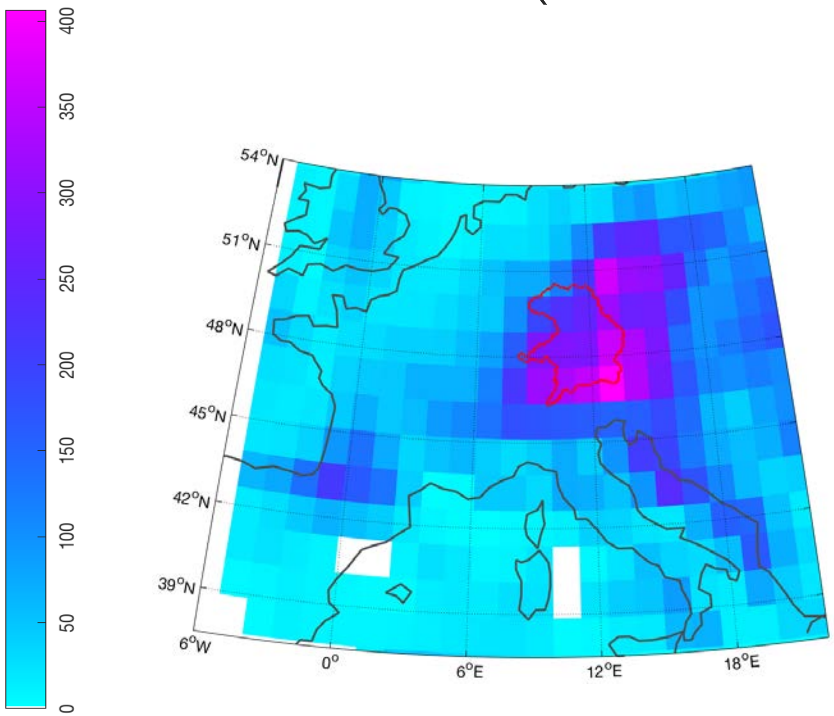
Classification + YOLO Bounding Box Regression

Contributors: Evan Racah (LBL), Chris Pal, Chris Beckham, Tegan Maharaj (U. Montreal)

Mediterranean cyclones (Vb)

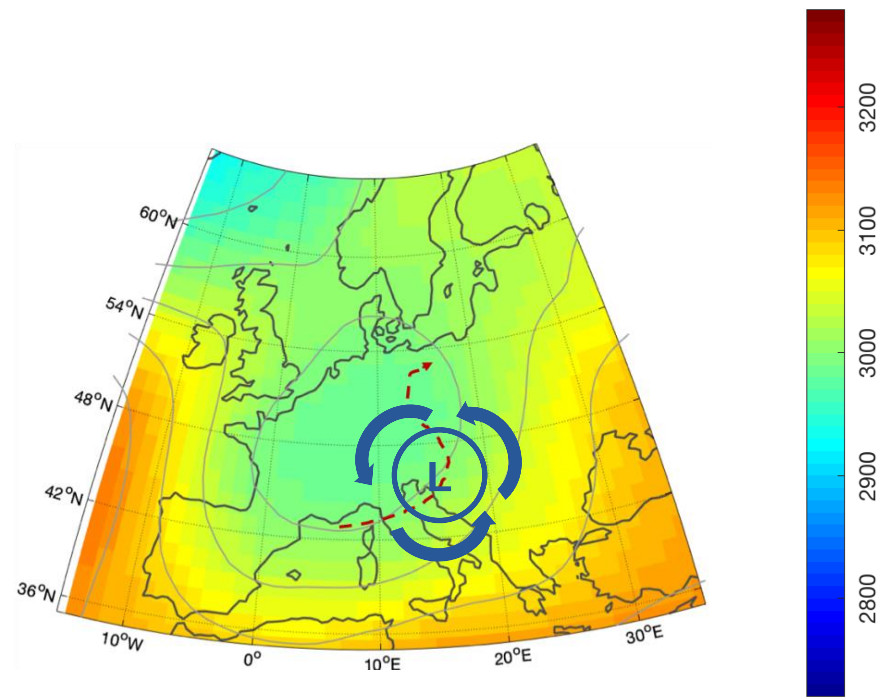
Precipitation sum during the
2013 flood event (29.5. - 2.6.2013)

mm



Mean synoptic pattern at 700 hPa
during the 2013 flood event

gpm

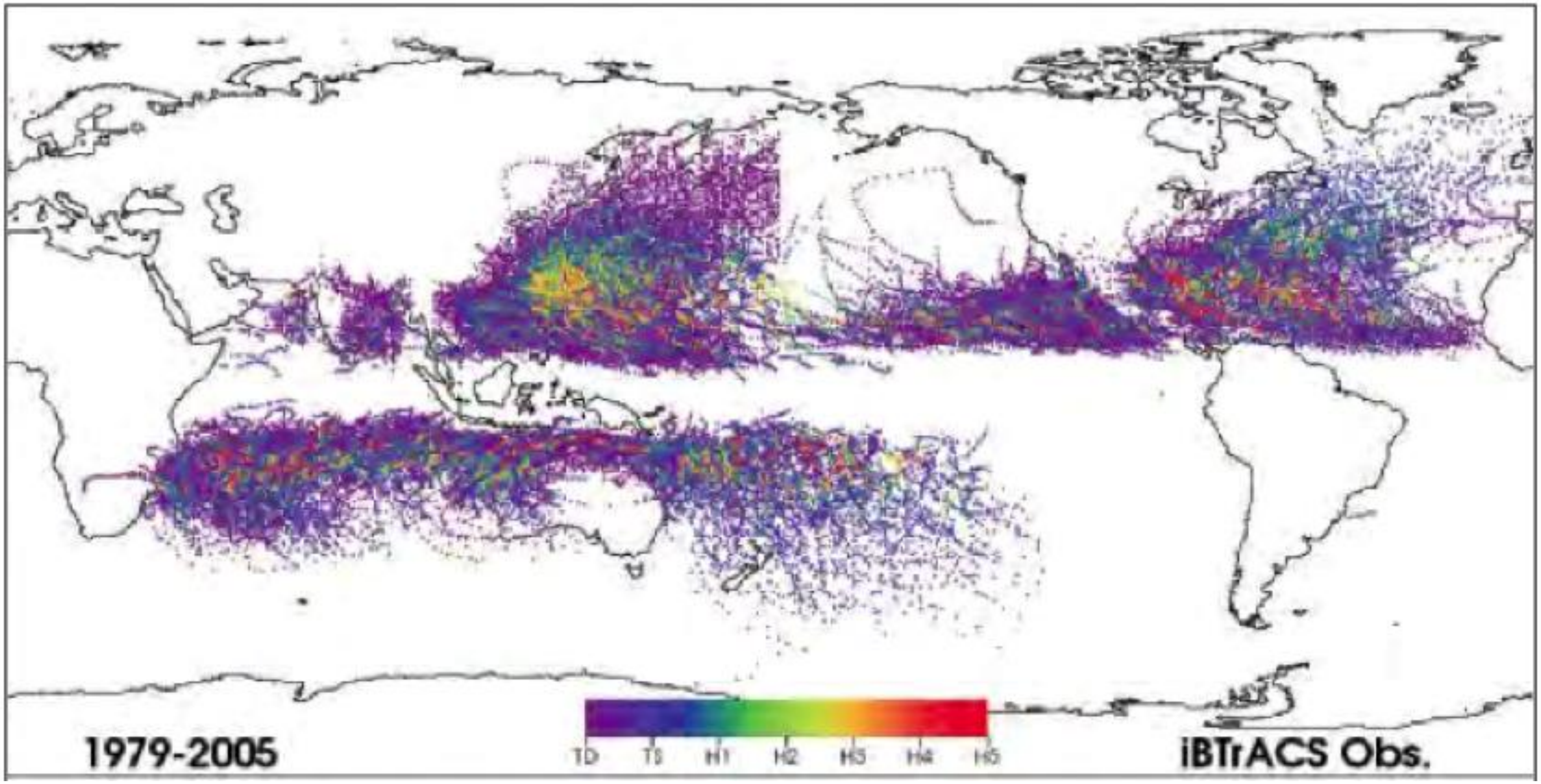


CLIMEX project

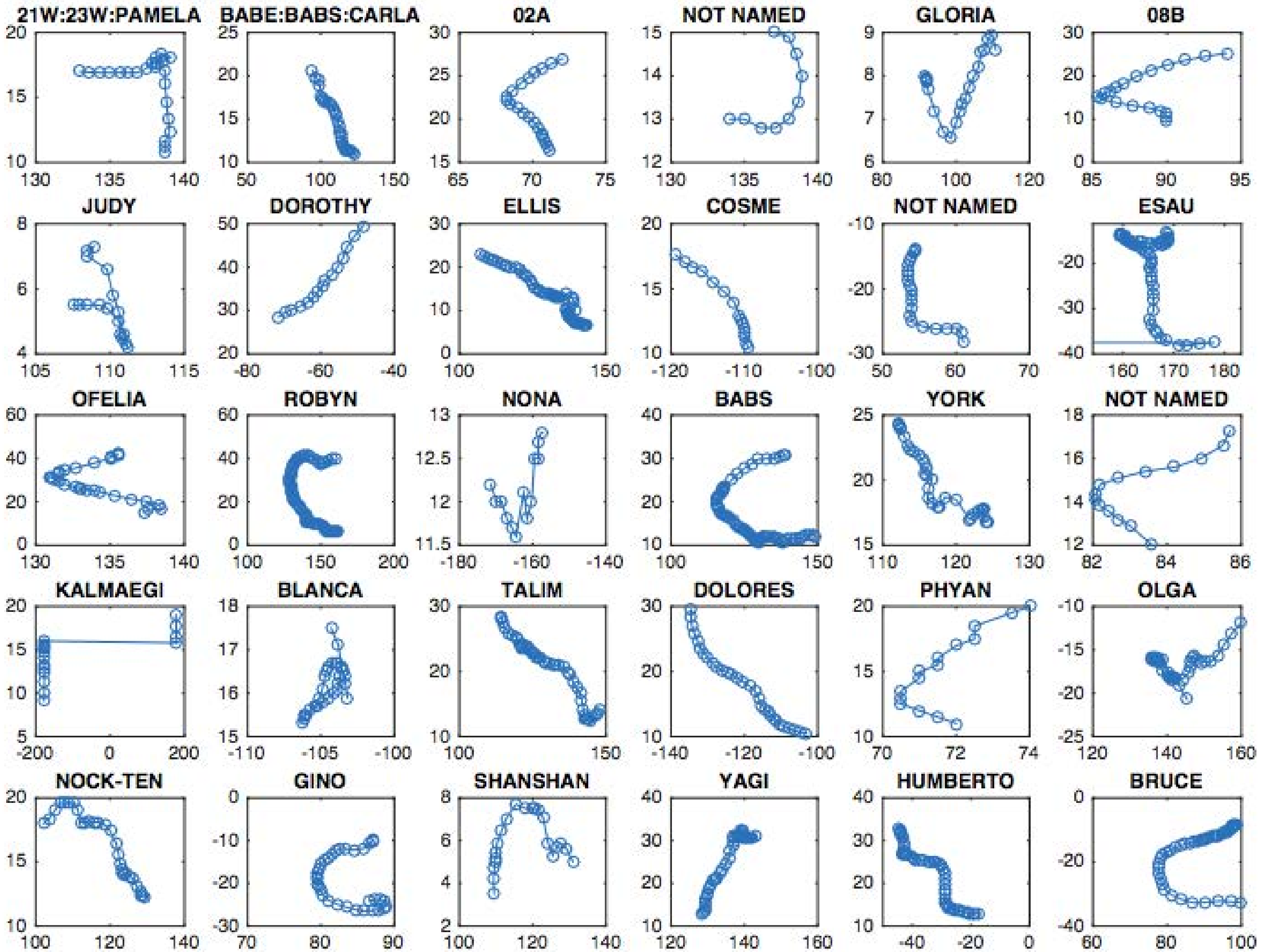
Agenda

- Context
- A few case studies: understanding changes in extremes
- A few challenges

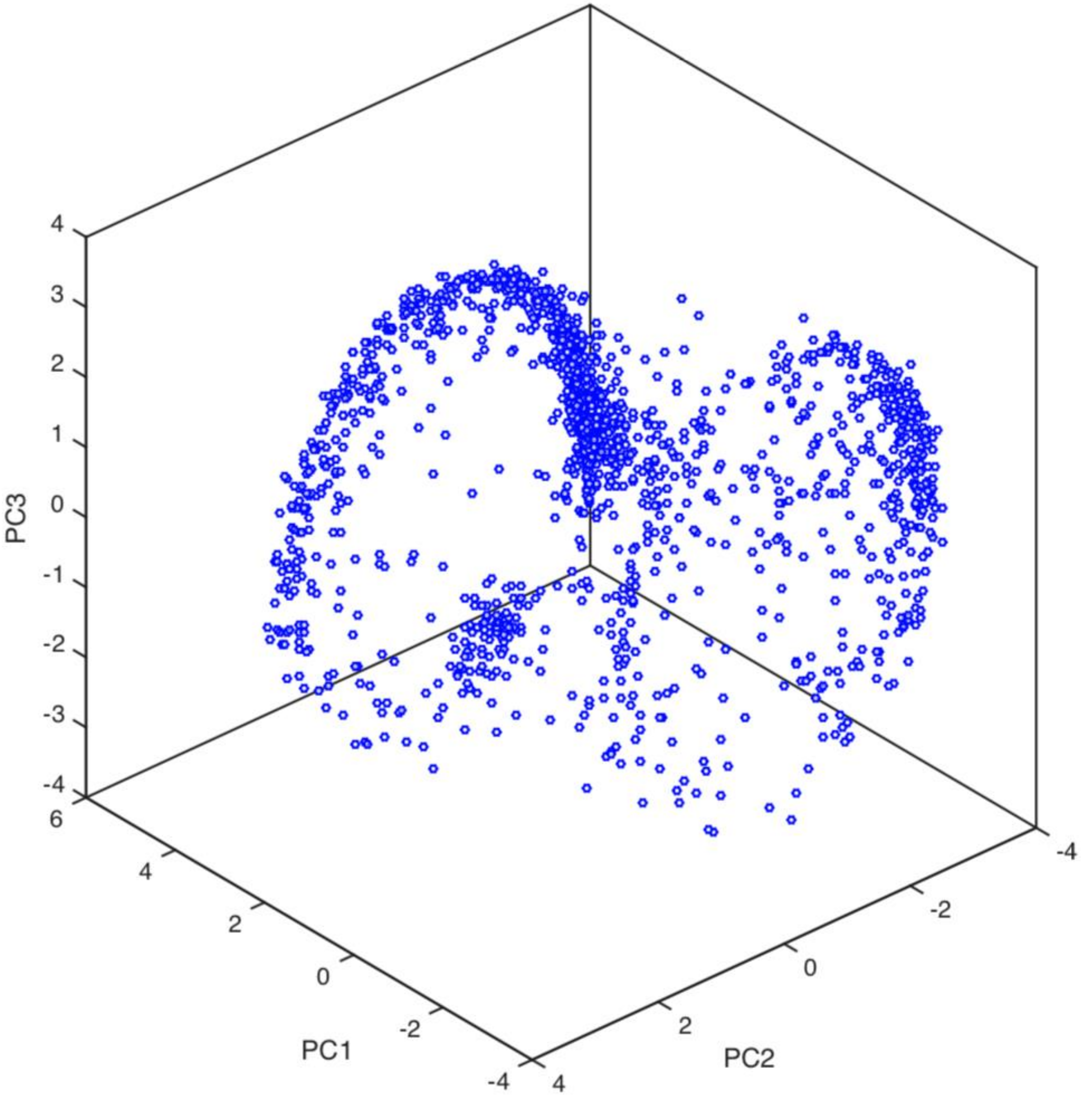
Spatial plot of tropical cyclones tracks



Individual trajectories

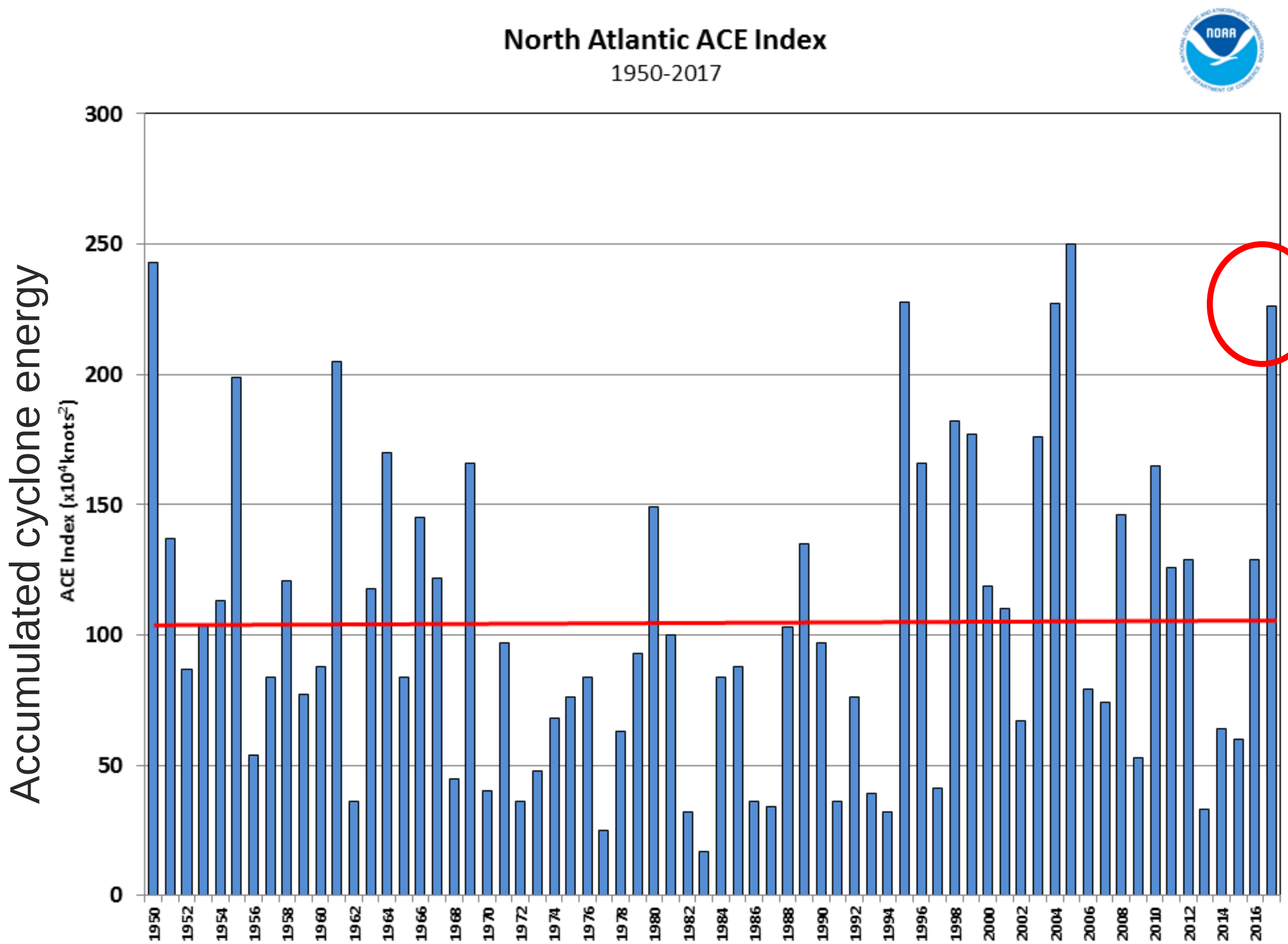


Individual trajectories: dimension reduction



Hannart et al. (in prep.)

Temporal plot of tropical cyclones occurrences



compatible with natural variations

NOAA National Centers for Environmental Information, State of the Climate: Hurricanes and Tropical Storms for Annual 2017, published online January 2018, retrieved on July 27, 2018 from https://www.ncdc.noaa.gov/sotc/tropical-cyclones/201713.

Old versus recent storms: supervised classification

Creating features that describe storms

Input

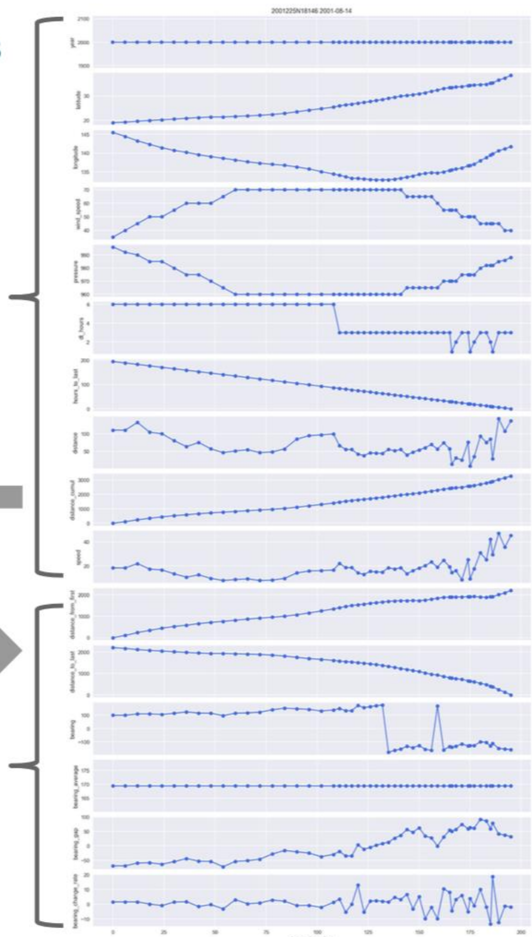
Each storm is described by **several timeseries** (several measurements every 6 hours)

- lat, lon
- wind speed
- pressure

Extra geo features

- storm speed
- bearing
- bearing volatility

1



2

Output

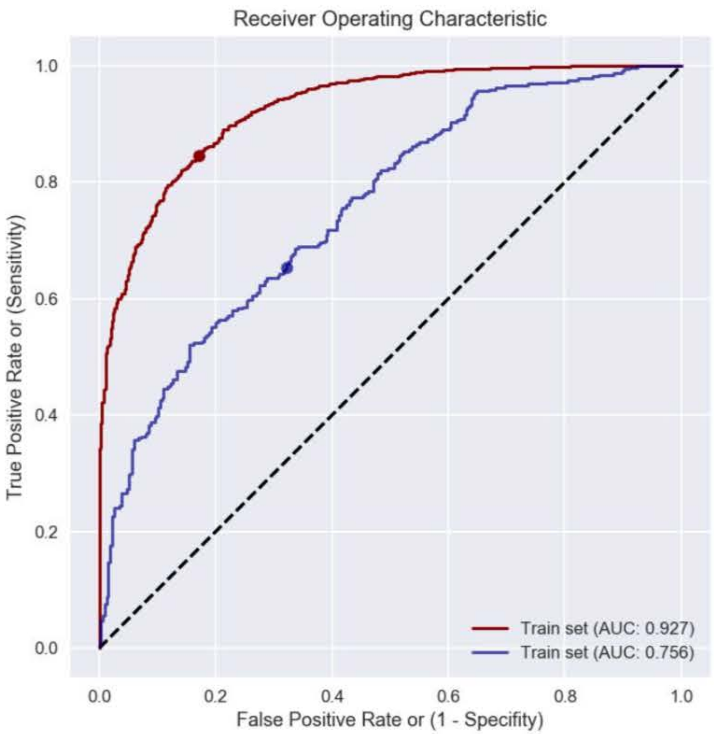
Each timeseries is **aggregated** using standard metrics (min, max, mean, std).

The **time dimension is thus removed**

id	1979343S08093
sample	train
target	False
predict	False
predict_proba	0.0010587
basin	SI
current_basin	SI
name	02S:CLAUDETTE:VIOLA:VIOLA/CLAUDE
nature	NR
sub_basin	WA
time	1979-12-09 00:00:00
year	1980
bearing_average_first	73.5765
bearing_change_rate_max	10.4237
bearing_change_rate_mean	-0.00929555
bearing_change_rate_min	-14.3821
bearing_change_rate_std	4.41162
bearing_gap_max	55.0722
bearing_gap_mean	-5.9513
bearing_gap_min	-81.1751
bearing_gap_std	38.698
bearing_max	128.649
bearing_mean	67.6252
bearing_min	-7.59855
bearing_std	38.698

Results – work in progress

Classification performance



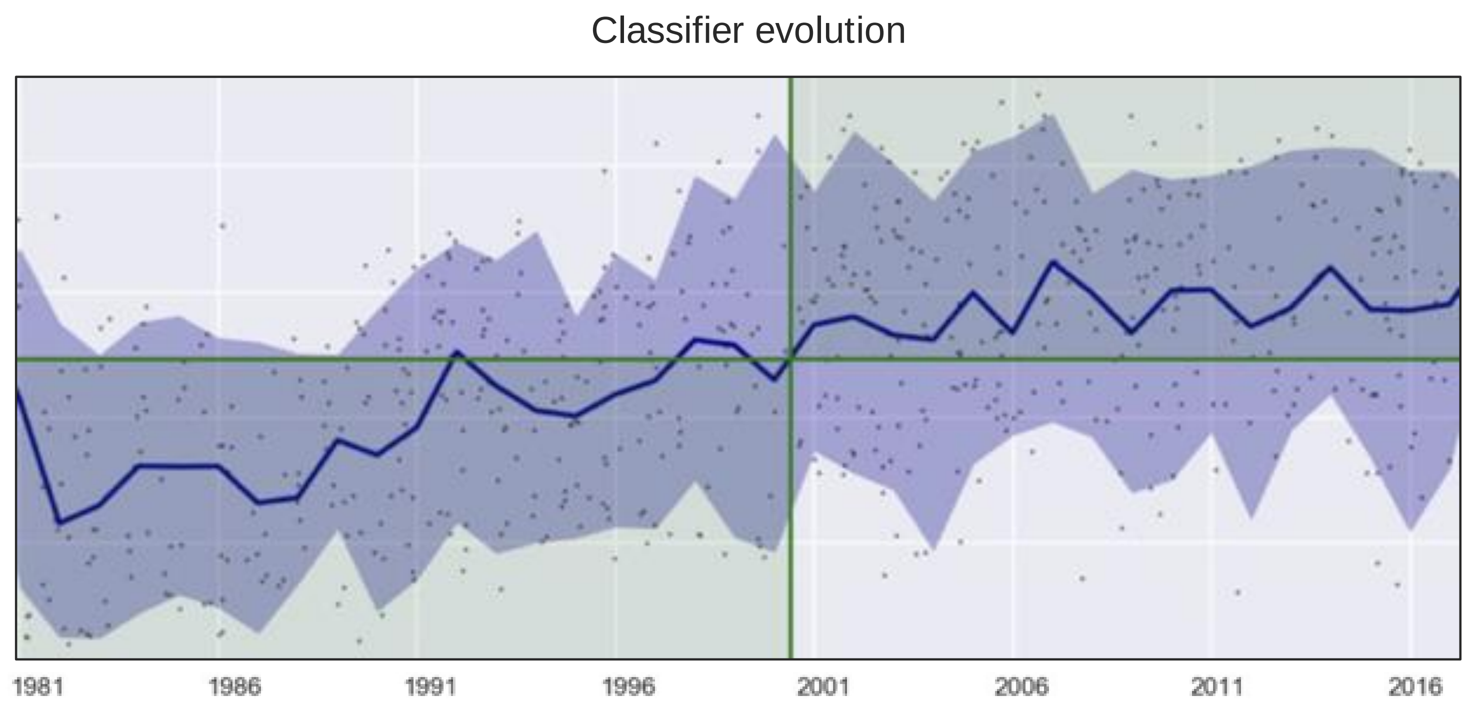
Model

- Gradient Boosting (sklearn) without complex hyper-optimisation
- target: is this storm after year 2000?
- train: 1900 storms
- test: 600 storms

Performances (test set)

- AUC: 75%
- Precision = Recall = 67%

Results – work in progress

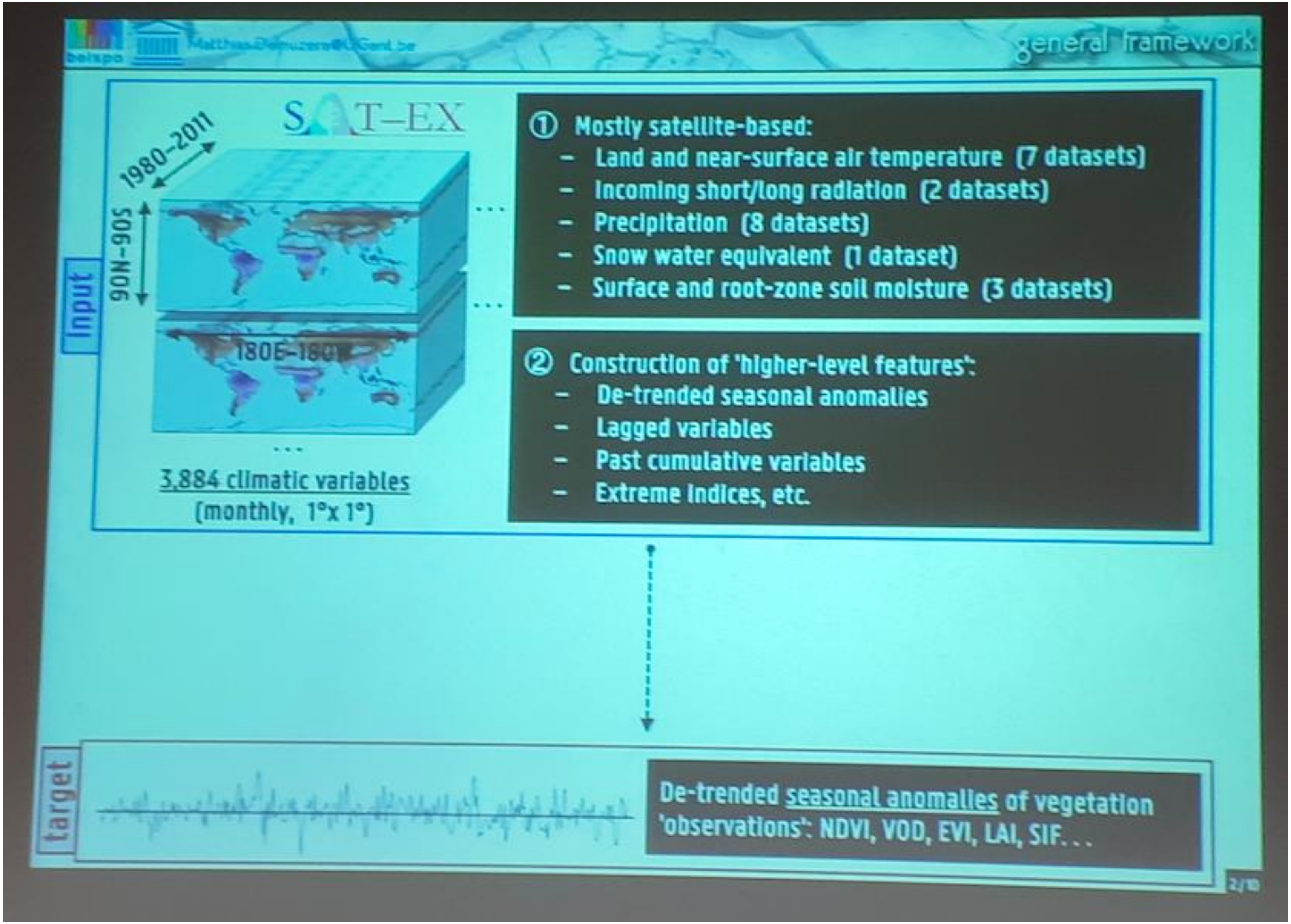


- The probability of hurricanes with z>0.5 has increased by a factor 6.
- Something has changed.
- Work in progress:
  - robustness check & verification on simulations
  - physical interpretation of the classifier

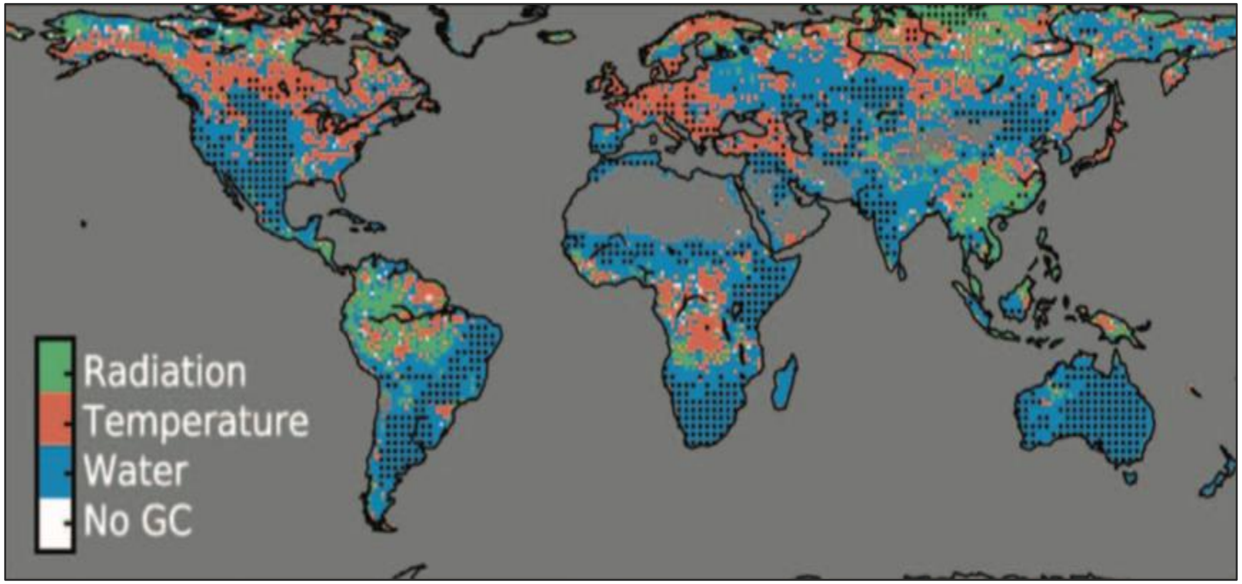
Agenda

- Context
- A few case studies: tracking causality
- A few challenges

Climatic drivers of vegetation growth

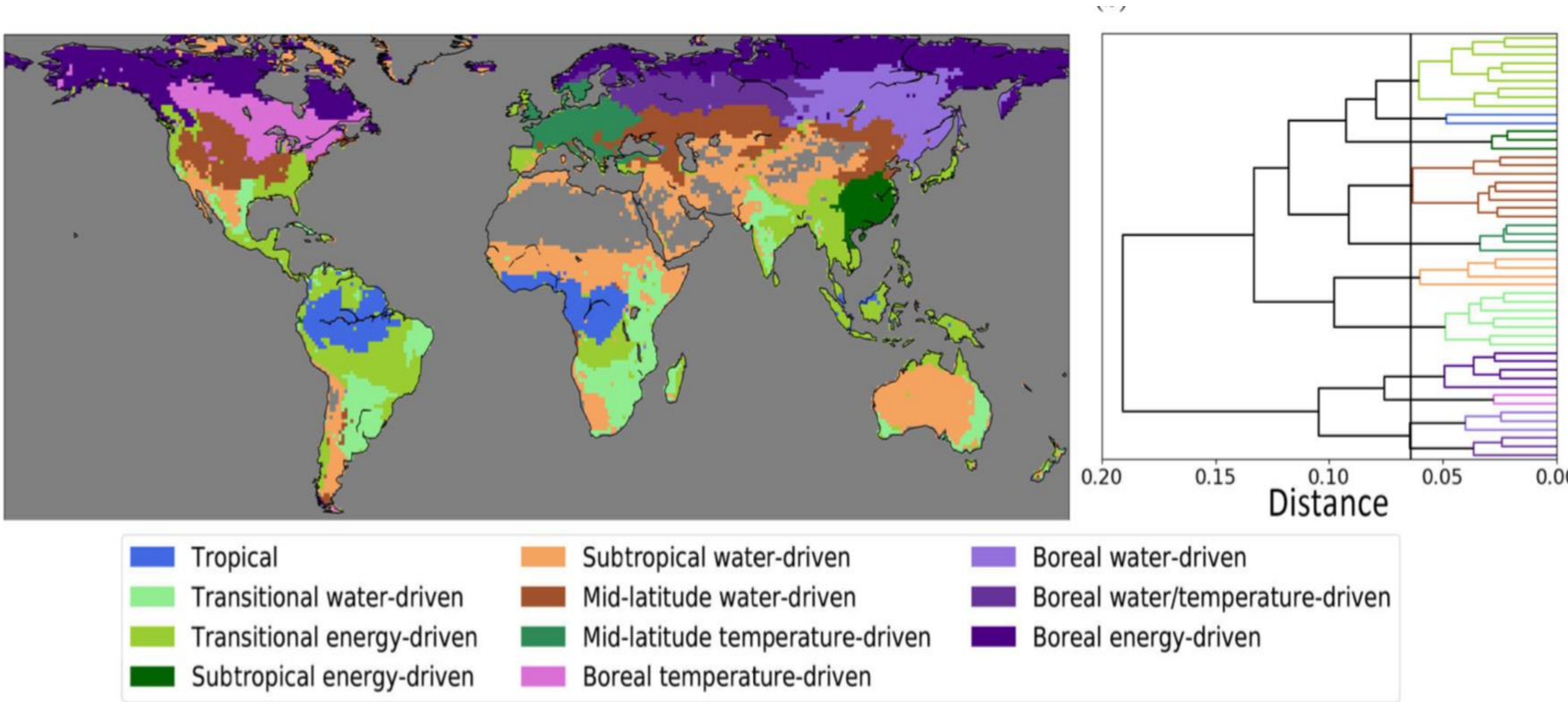


Nonlinear Granger causal classification



Pappagianopoulou et al. 2017

Multi-task learning causal classification



Pappagianopoulou et al. 2018

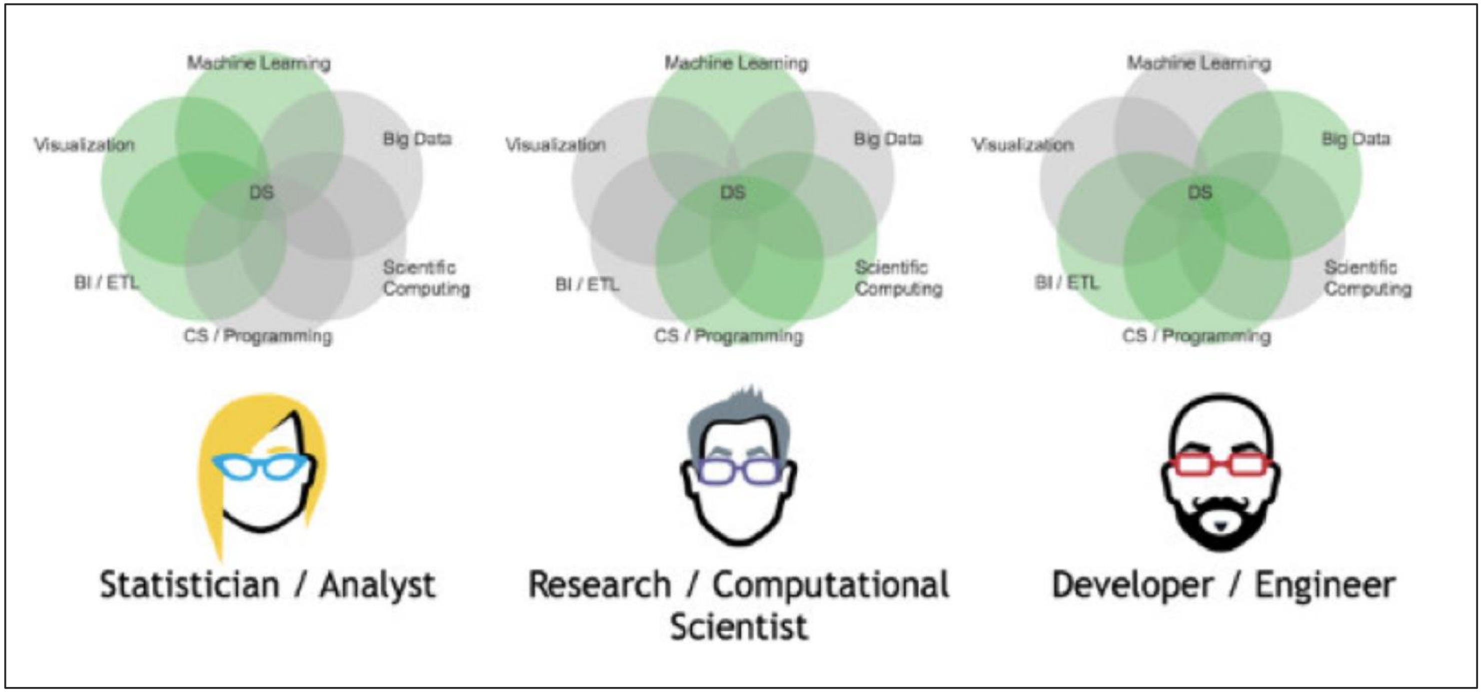
Agenda

- Context
- A few case studies
- A few challenges

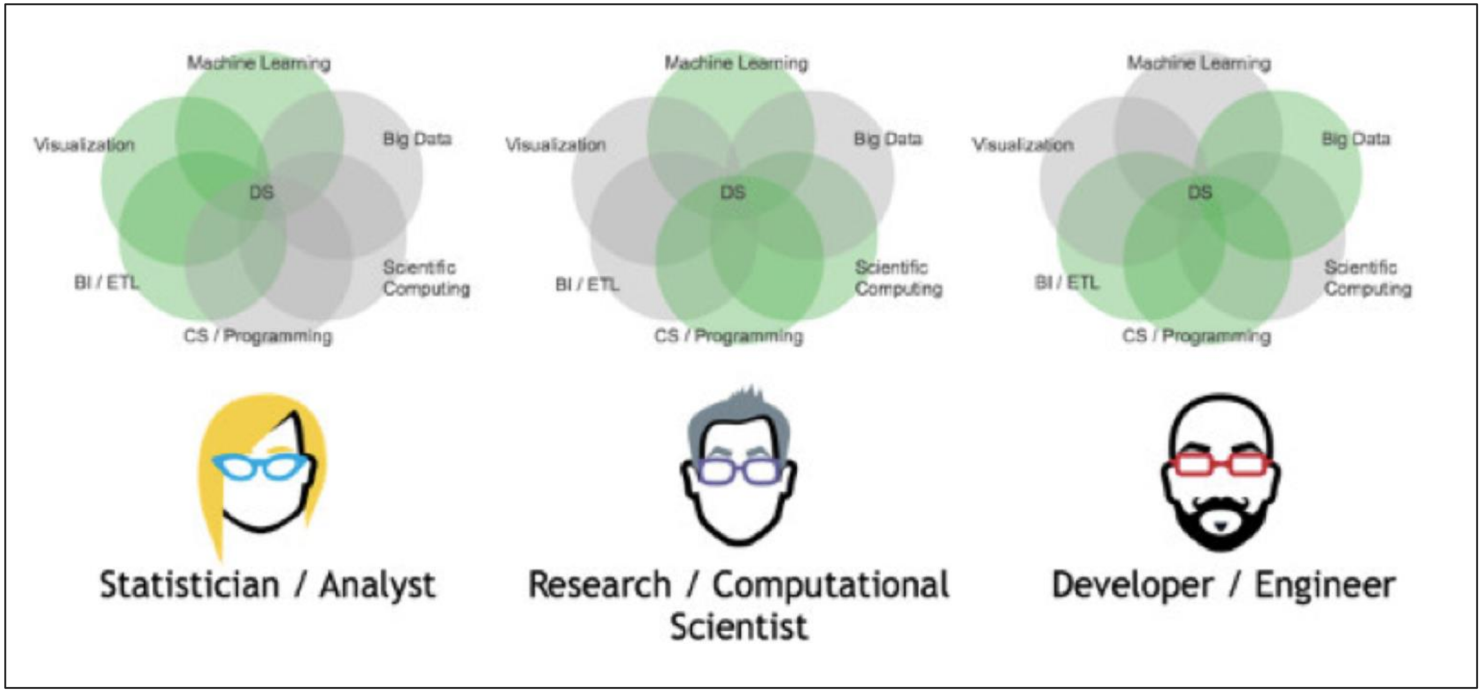
OURANOS

Technological complexity challenge

HR challenge



Organizational challenge

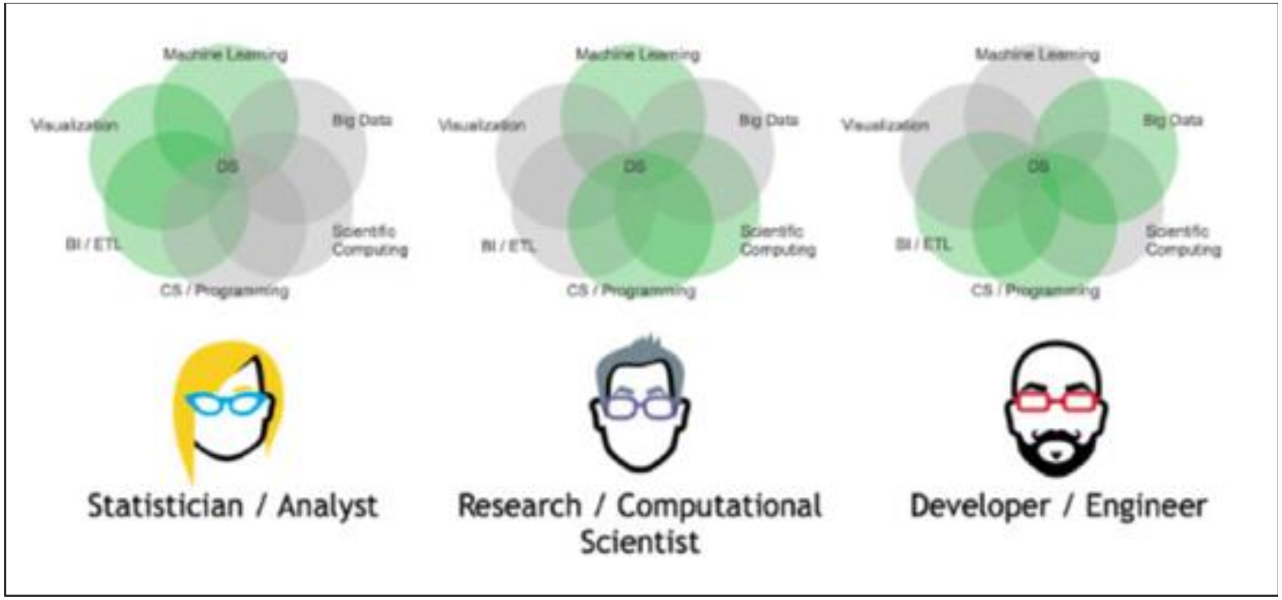


+

Climate Scientists / Impact Scientists

Epistemological challenge

Empirical models



+

Physical models

Climate Scientists / Impact Scientists

Data access challenge

Research in the 21st Century Context

Increasingly complex data-driven decision-making

Data volumes getting too large for standard local analysis

Multidisciplinary research hindered by absence of common language and formats

Moving from deductive to inductive science

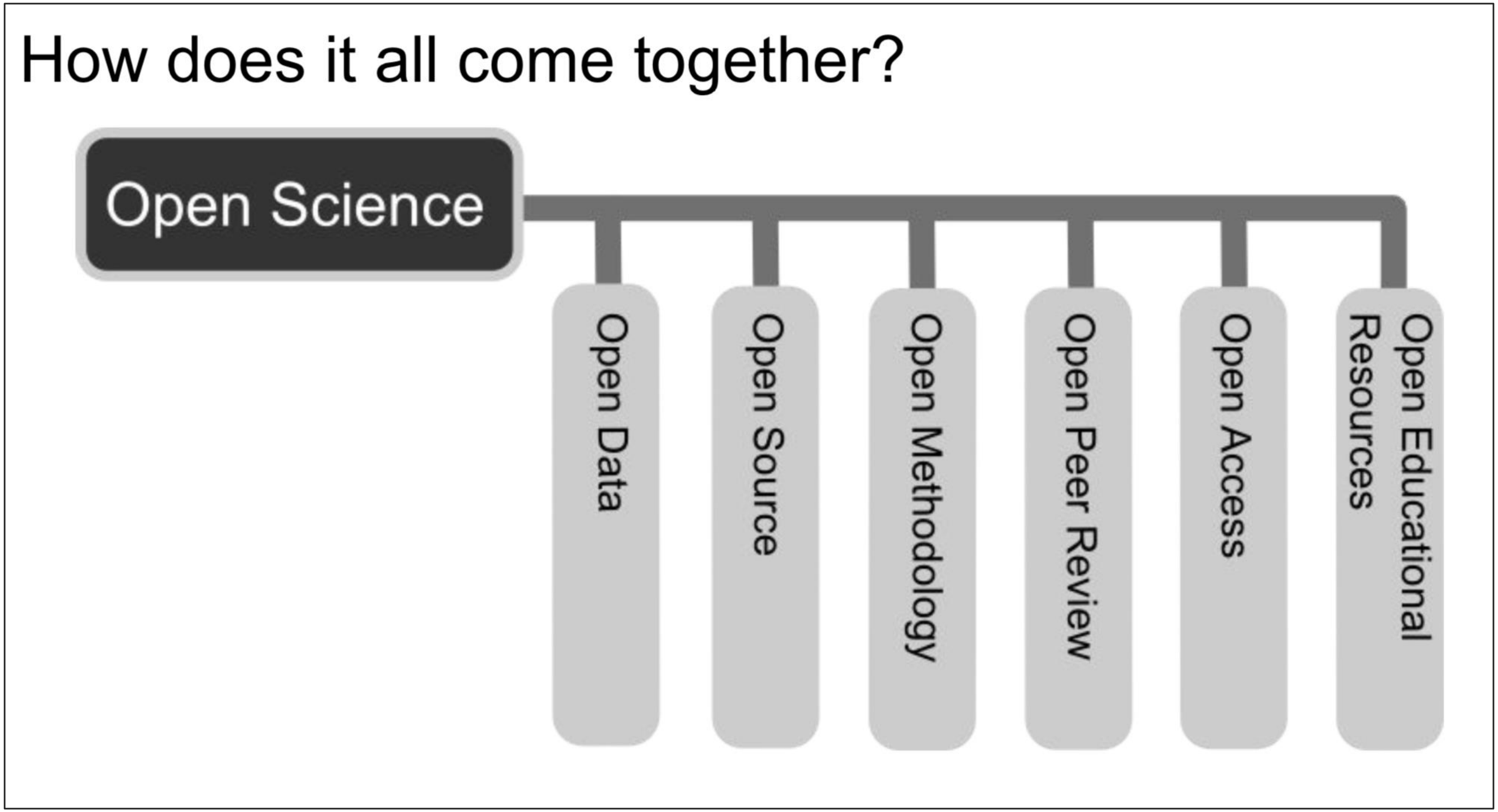
Competitive scientific environment at the national and individual level

Pool resources / Abstract low-level details

Collaborate to solve BIG problems

Courtesy of Trevor Smith and David Huard

Data access challenge



Courtesy of Trevor Smith and David Huard

Data access challenge

Canadian GSDI

- GeoSpatial Data Infrastructure
  - Research Project Financing
  - Interoperable Data Standards
  - Data Creation/Sharing Policies
  - Geospatial Network Research
  - **"Open Government"**
  - **OGC Web Services**
    - **WMS/WFS/WCS**
    - **CanVec/GeoGratis**

https://www.nrcan.gc.ca/earth-sciences/geomatics/canadas-spatial-data-infrastructure

https://open.canada.ca/en/open-maps

Science Gateways

Managed networked environments including community-developed set of tools, applications and digital data collections that are integrated through a tailored web-based environment that support the whole research cycle.

Support data-intensive and multidisciplinary science

A lot of activity on science gateways and virtual research environment

- Europe: Copernicus
- Australia: nectarlabs
- Canada: Enviro-Net

https://catalog.sciencegateways.org

ICSG International Coalition on Science Gateways

RDA RESEARCH DATA ALLIANCE

Courtesy of Trevor Smith and David Huard

Data access challenge

What is PAVICS?

Platform for the **A**nalysis and **V**isualization of **C**limate **S**cience

PAVICS

Data exploration and processing platform that allows users to perform climate analyses on large data sets over the Internet

Supported by **CANARIE**, developed alongside researchers at **CRIM**

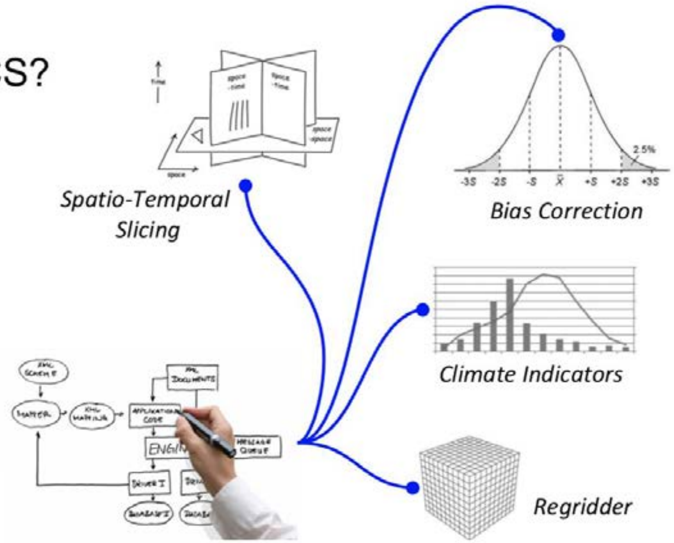
Integration of "Open" technologies with Climate Science
    Supporting transparency and availability → **Collaboration**

Research Software supporting Open Climate Services

What is PAVICS?

PAVICS

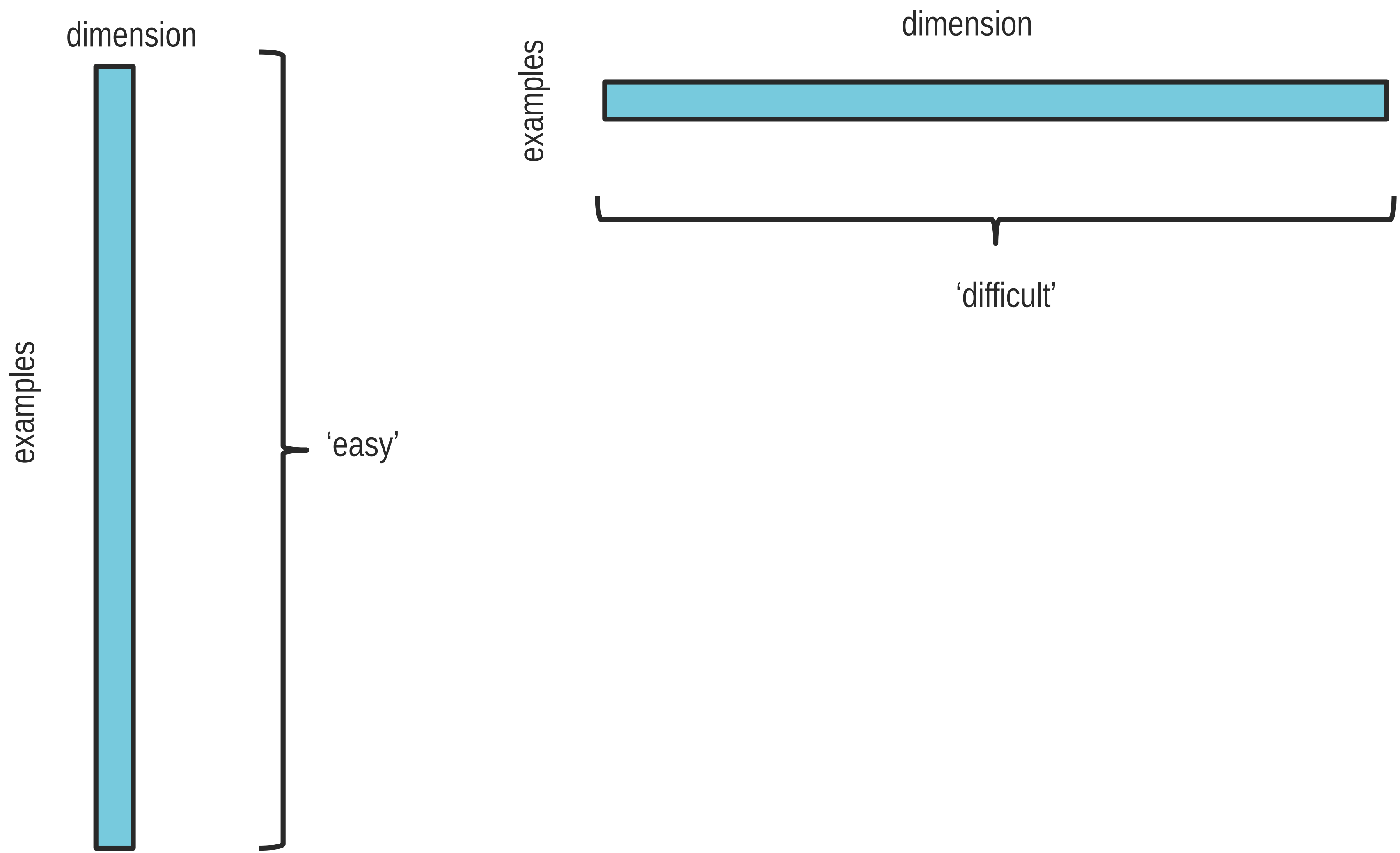
Software platform



Courtesy of Trevor Smith and David Huard

Theoretical challenges

- ‘Small data’ algorithms



Theoretical challenges

- ‘Small data’ algorithms
- Quantification of uncertainty

→ Gaussian processes

→ Bayesian updating

Data assimilation,
Ensemble Kalman filter

Bocquet et al. 2015

Hannart et al. 2015

Big Data Trophy, Paris, 2015

Theoretical challenges

- ‘Small data’ algorithms
- Quantification of uncertainty
- ‘Ilities’: interpretability, reliability and causality

→ Gaussian processes

→ Bayesian updating

→ Bayesian networks

Hannart et al. 2016a

Hannart et al. 2016b

Hannart et al. 2017

Hannart et al. 2018

Conclusion

- An outlook was given on a short and non-exhaustive list of examples of applications of AI in climate science.
- Promising early results, and room for more.
- Exciting organizational, technological and theoretical challenges need to be addressed to foster this research.