A deep learning approach for simulating compound coastal hazards over large domains

Alison Peard

Waves and Flows Seminar, 27th March 2024

OPSIS
Oxford Programme for Sustainable Infrastructure Systems

UNIVERSITY OF OXFORD

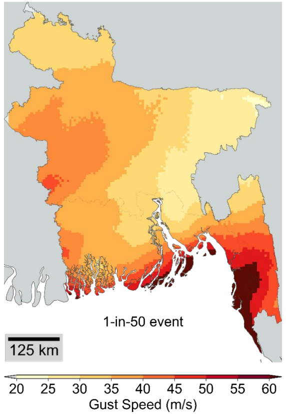
UKRI
Engineering and Physical Sciences Research Council

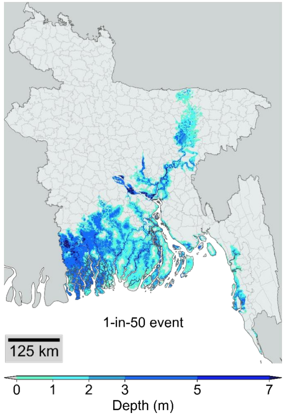
Compound coastal hazards over large domains

- Why we want to simulate …
- How it’s usually done and limitations
- Why deep learning a good candidate
- Similar work
- Current approach
- (Preliminary) results
- Conclusions and next steps

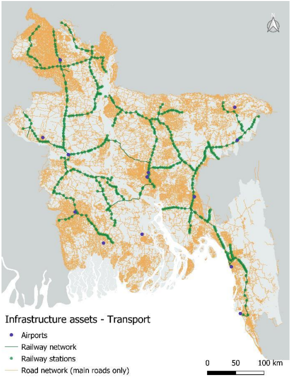
Motivation: large-scale spatial risk analyses

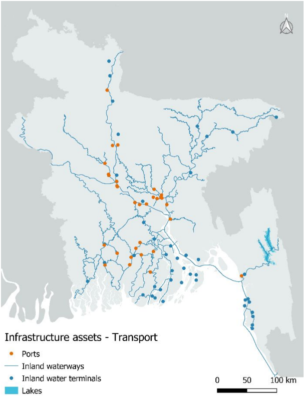
Hazards

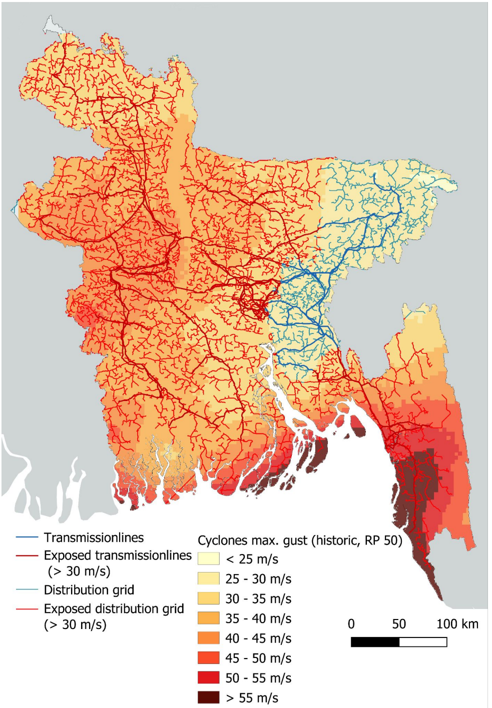




Critical infrastructure





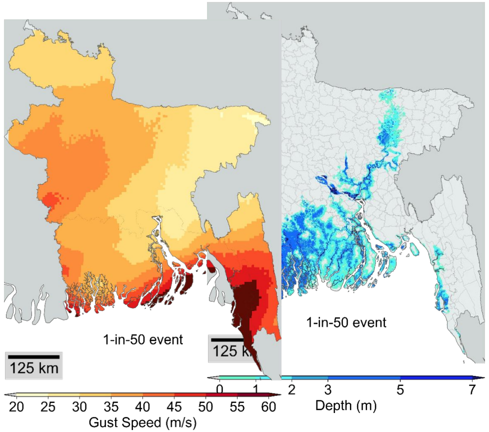


Exposure

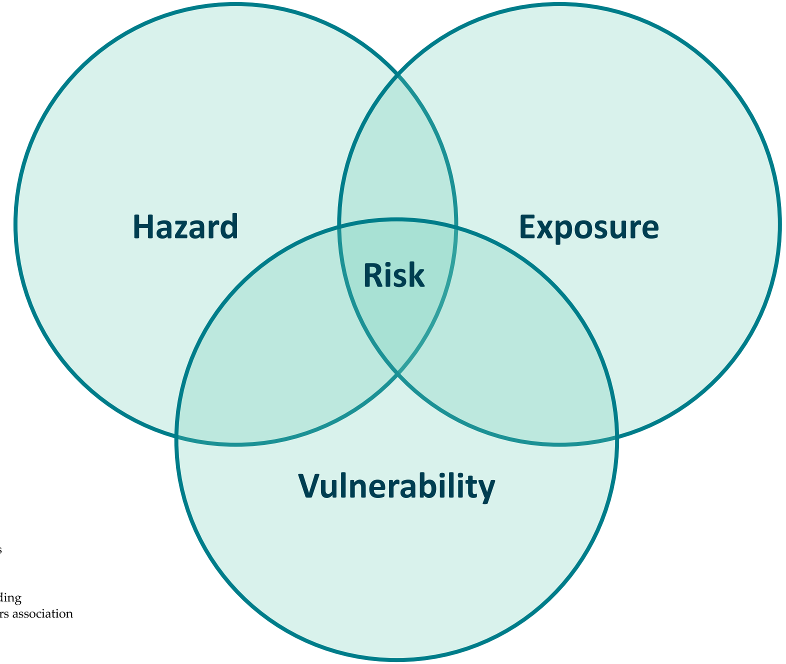
- National/continental scales
- Critical infrastructure
- Climate hazards
- Risk:
  - Exposure and vulnerability of different sectors
  - Vulnerability hotspots
  - Cascading risks
    - Societal
    - Economic

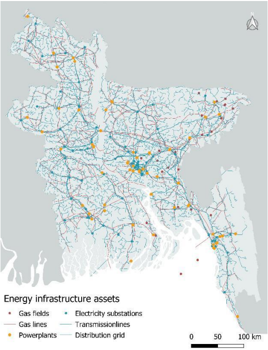
https://doi.org/10.1038/s41558-024-01974-8

Motivation: data for large-scale spatial risk analyses

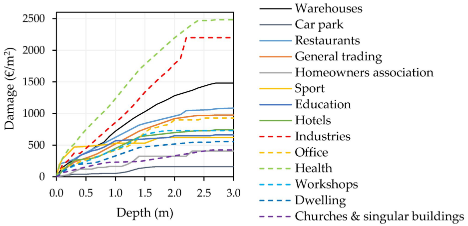


Hazard data



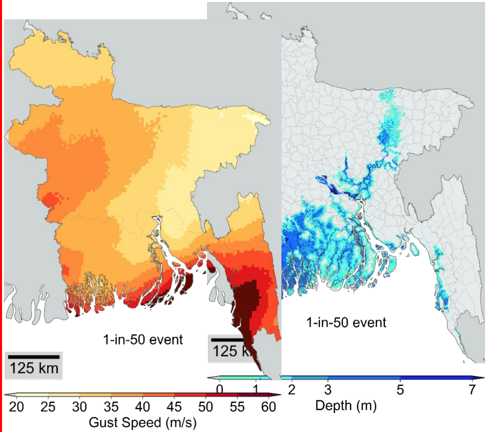


Geospatial data

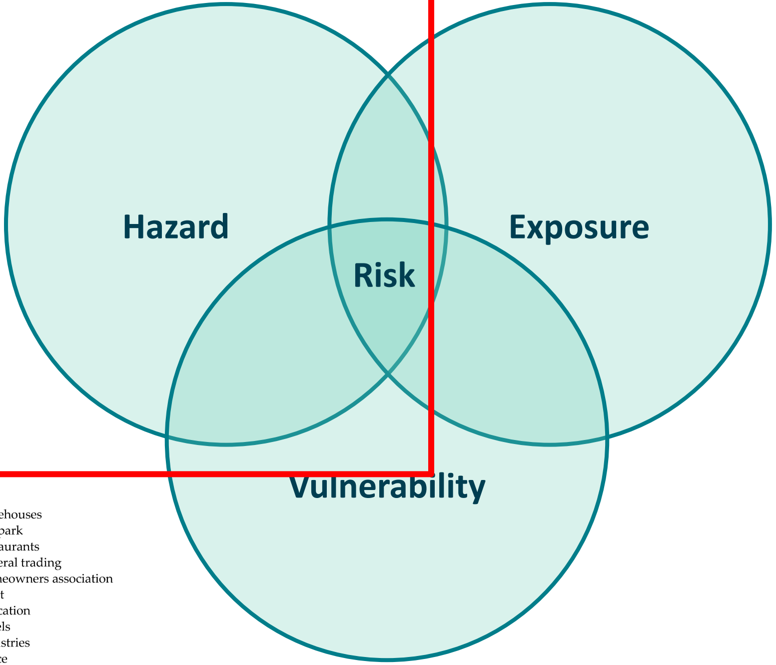


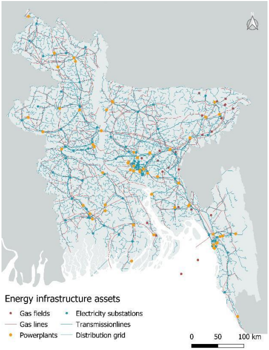
[Flood Depth–Damage Curves for Spanish Urban Areas]

Motivation: data for large-scale spatial risk analyses

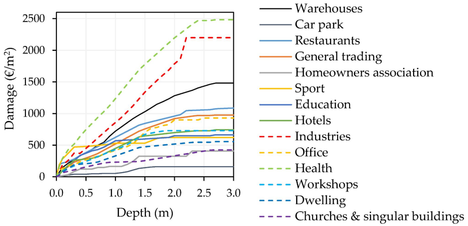


Hazard data



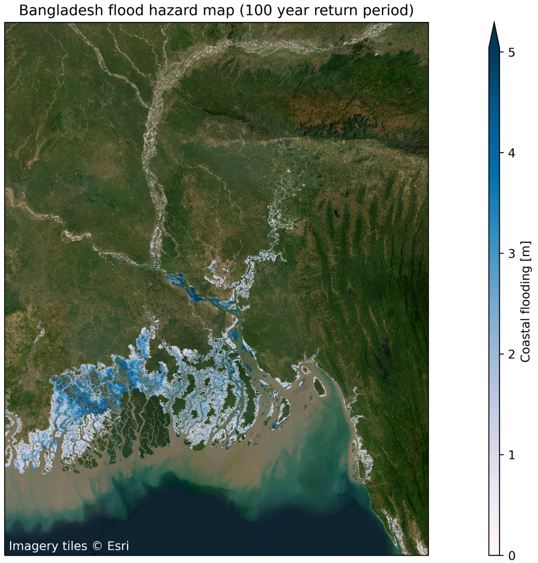


Geospatial data



[Flood Depth–Damage Curves for Spanish Urban Areas]

Large scale hazards: homogeneous return periods



Hazard maps

- One-in-T year hazard => 1/T probability
- No uncertainty estimates
- Single map per time frame/RP/scenario
- From numerical simulations

- Univariate - single hazard

- Assume complete dependence between all locations

Time frame	Hazard	Climate Scenario	Return period
2010, 2030, 2050	Coastal flooding	RCP 4.5, RCP 8.5	2, 5, 10, 25, 50, 100
…			

Engineering and Physical Sciences Research Council
OPSIS Oxford Programme for Sustainable Infrastructure Systems
eci
UNIVERSITY OF OXFORD

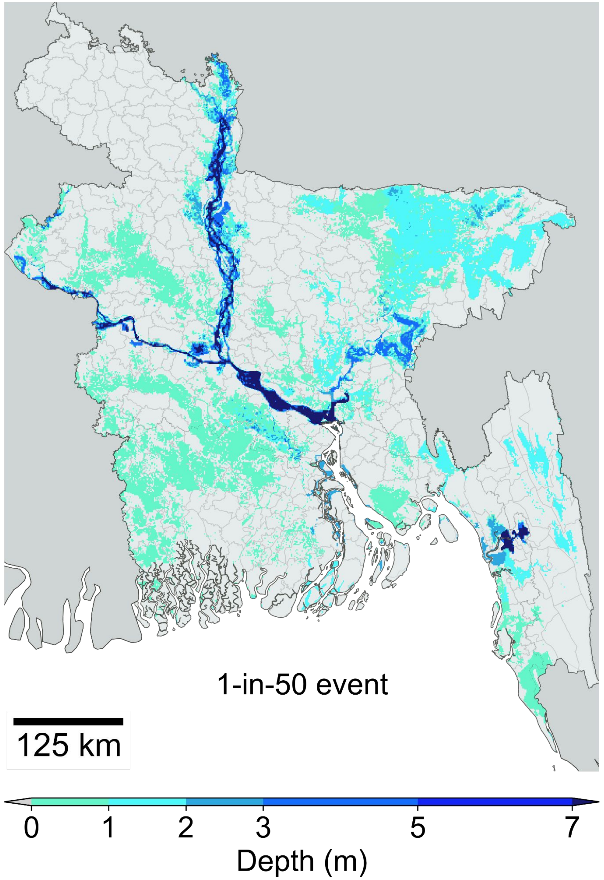
Rough work: impact/applications

Vorogushyn

- National investments in flood risk mitigation
- Insurance portfolio management
- Hazard and Exposure ↑ vulnerability↓
- The solidarity principle set out in the EU Flood Dir requires that communities should not be adversely affected by risk reduction measures implemented elsewhere.
- FRMPs flood risk management plans
- Lack of *spatially consistent* large-scale flood risk assessments
- Prevailing approach: assembling local-scale flood hazard and risk assessments
- Cost-benefit ratios for interventions to be quantified properly
- “These interactions vastly modify flood risk and may lead to profoundly different mitigation/adaptation measures and policies.“
- Spatially-homogenous return period of floods
- :For instance, the probability of a single storm event resulting in a flood with a 100-year return period discharge peaks at all gauges in a large-scale basin is far below 0.01
- does not provide insight into possible flood extents or consequences related to single extreme events.
- Probability of it happening *somewhere* is much higher, but of it happening *everywhere* is far lower!
- What does consistent mean here?
- Flood event sets

Get some numbers on compound

Large scale hazards: homogeneous return periods



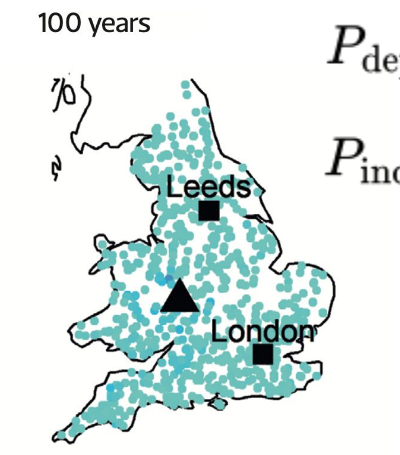
Questions you might want to answer

- What is the probability of getting an event of a certain magnitude (RP) somewhere in the region?
- What is the probability of getting a large event in more than one location at the same time?
- What are expected annual damages from hazard events?

Engineering and Physical Sciences Research Council
Oxford Programme for Sustainable Infrastructure Systems
UNIVERSITY OF OXFORD

Large scale hazards: example I

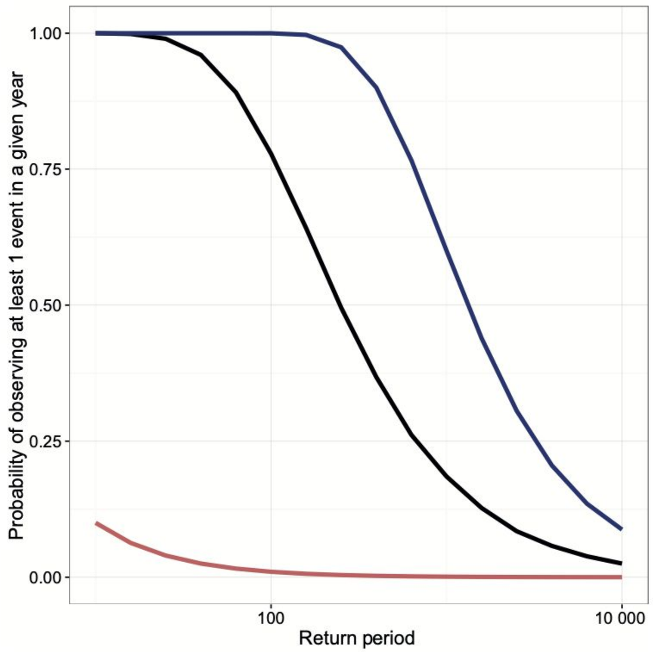
"What is the probability, *P*, that a one-in-100-year event occurs somewhere in England and Wales this year?" - Towe, Tawn, and Lamb (2018)



River flow gauges over England and Wales

$$P_{\text{dependence}} = \frac{1}{T}$$

$$P_{\text{independence}} = 1 - \left(1 - \frac{1}{T}\right)^N$$



Towe, Tawn, and Lamb (2018) *Why extreme floods are more common than you might think*

Large-scale hazards: example II

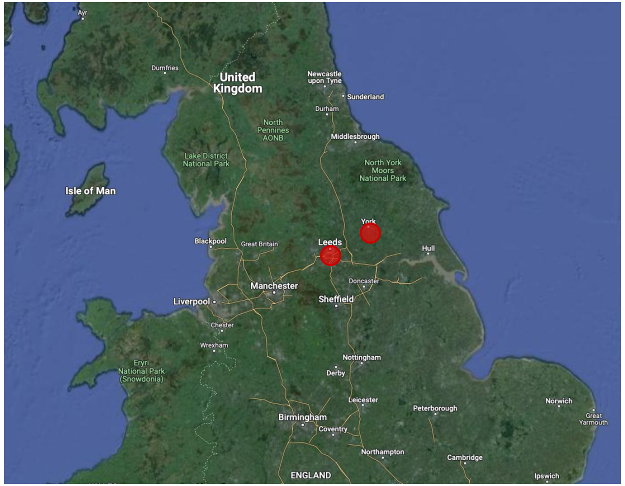
Leeds (River Aire) and York (River Ouse)

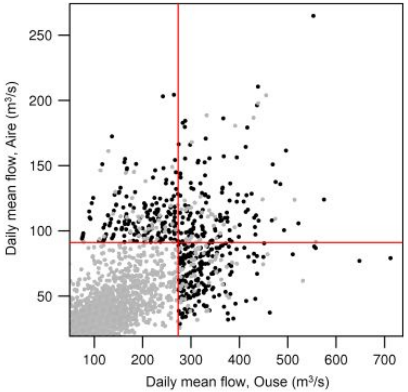
Bivariate conditional exceedance model

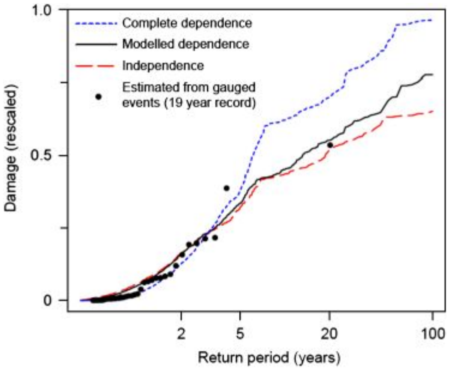
5,000 simulated events where at least one site exceeds threshold (red line)

Losses due to flooding

- Complete dependence overestimates losses
- Complete independence underestimates losses







Lamb (2010) *A new method to assess the risk of local and widespread flooding on rivers and coasts*

Large-scale hazards: compound hazards

Table 1 | Non-exhaustive list of documented climate-related hazards for which drivers are dependent as well as combinations of dependent hazards with potentially large impacts

Hazard(s)	Climatic drivers	Reference(s)
Drought	Precipitation, evapotranspiration, historic evolution of soil moisture, temperature	[35,77,78]
Physiological heat stress	Temperature, atmospheric humidity, strongly dependent on diurnal cycle	[56]
Fire risk	Temperature, precipitation, relative humidity, wind, lightning	[55,79]
Storm risk	Wind speed, humidity, large scale atmospheric circulation	[94,95]
Coastal flood	River flow, precipitation, coastal water level, surge, wind speed	[11,12,30]
Flood risk at river confluences	Precipitation, water levels of contributing rivers, large-scale atmospheric circulation	[31]
Concurrent drought and heat	Temperature, precipitation, evapotranspiration, atmospheric humidity	[7,35]
Concurrent wind and precipitation extremes	Wind speed, precipitation, orography, large-scale atmospheric circulation	[34]
Concurrent heat and air pollution	Temperature, sulfur dioxide, NO_x, particulate matter (PM_{10})	[6,76]

- Co-occurring extremes of different climate processes can lead to greater losses
- Climate hazards are often related, likely to co-occur, e.g.
  - Wind and rain
  - Heatwave and drought
- Univariate approach neglects this risk

Zscheischler *et al.* (2020) *Future climate risk from compound events*

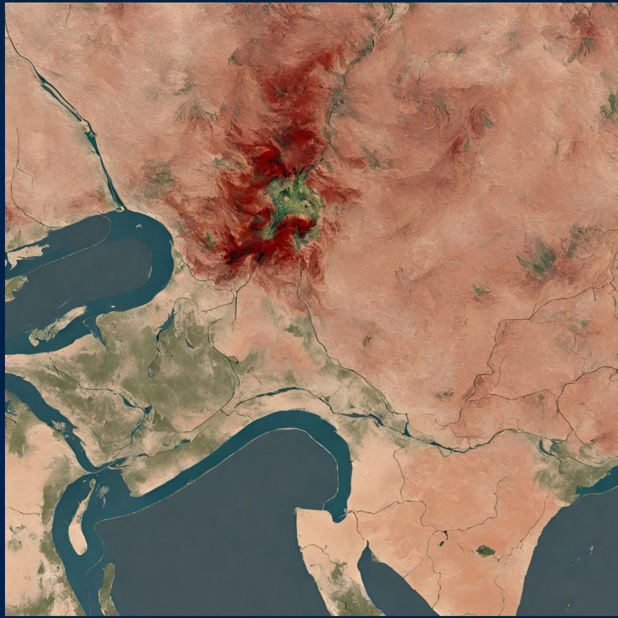
Large-scale hazard maps: implications

- Inaccurate expected damage
- Underestimate yearly probability of an extreme event *somewhere*

Spatially-compounding
- Systems may be affected by hazards in multiple locations, e.g., food systems by multiple crop failures
- Don’t account for where emergency services might be stretched between locations

Multivariate

- Co-occurring extremes can lead to much worse damages, e.g., rain, high tides, and strong winds causing compound coastal flooding



Ideal solution

- Large portfolio of plausible multivariate events accounting for dependence
    - Across space
    - Between climate processes

Engineering and Physical Sciences Research Council
OPSIS Oxford Programme for Sustainable Infrastructure Systems
eci
UNIVERSITY OF OXFORD

Summary

Hazard map approach has flaws

- Doesn’t account for varying dependence between
  - Locations
  - Hazards
- No uncertainty estimates
- Leads to under/overestimation of damages
- Affects climate adaptation planning
  - Wrong priorities

Engineering and Physical Sciences Research Council
Oxford Programme for Sustainable Infrastructure Systems
eci
UNIVERSITY OF OXFORD

Statistical models: example for bivariate case

- E.g., for single location
  - Variable 1: wind speed
  - Variable 2: precipitation

- Want to understand:
  1. Their marginal distributions $p_X(x) = \int_y p_{X|Y}(x \mid y)\, p_Y(y)\, \mathrm{d}y$
  2. Dependence between them
  3. Dependence in tails

Usual approach is to *split this in two*…

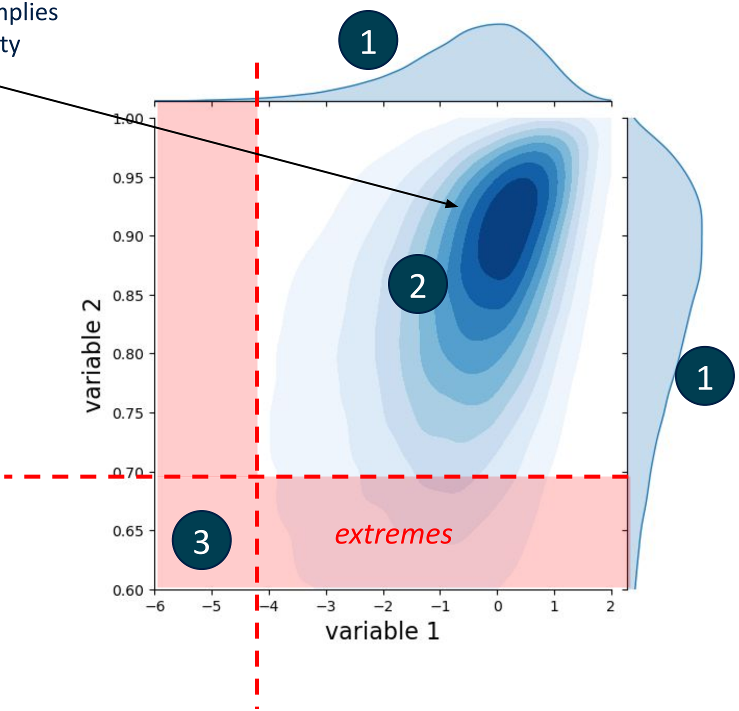
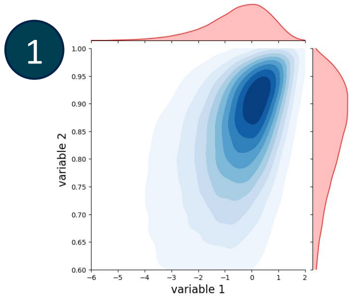


Image: [3 dead as Hurricane Irma batters Florida with high winds, heavy rain | Arab News]

Traditional approaches: statistical models for bivariate case

Usual approach is to *split this in two*…

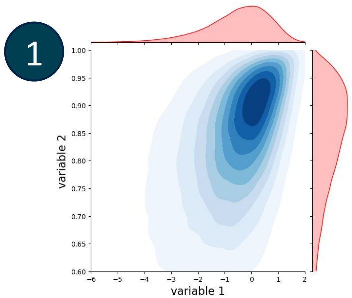
1. Learn model to describe marginals

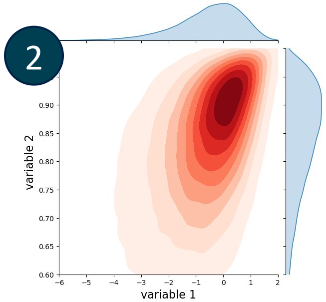


Traditional approaches: statistical models for bivariate case

Usual approach is to *split this in two*…

1. Learn model to describe marginals
2. Learn model to describe dependency

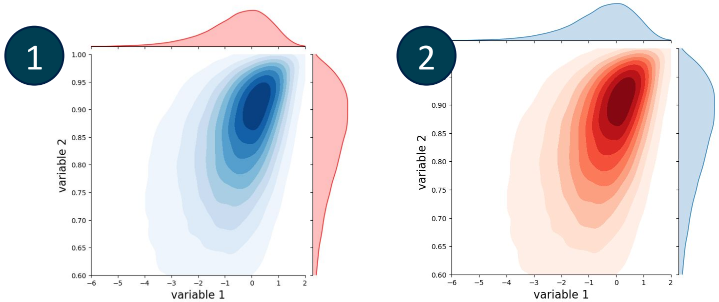




Traditional approaches: statistical models for bivariate case

Usual approach is to *split this in two*…

1. Learn model to describe marginals
2. Learn model to describe dependence



Two most popular approaches:

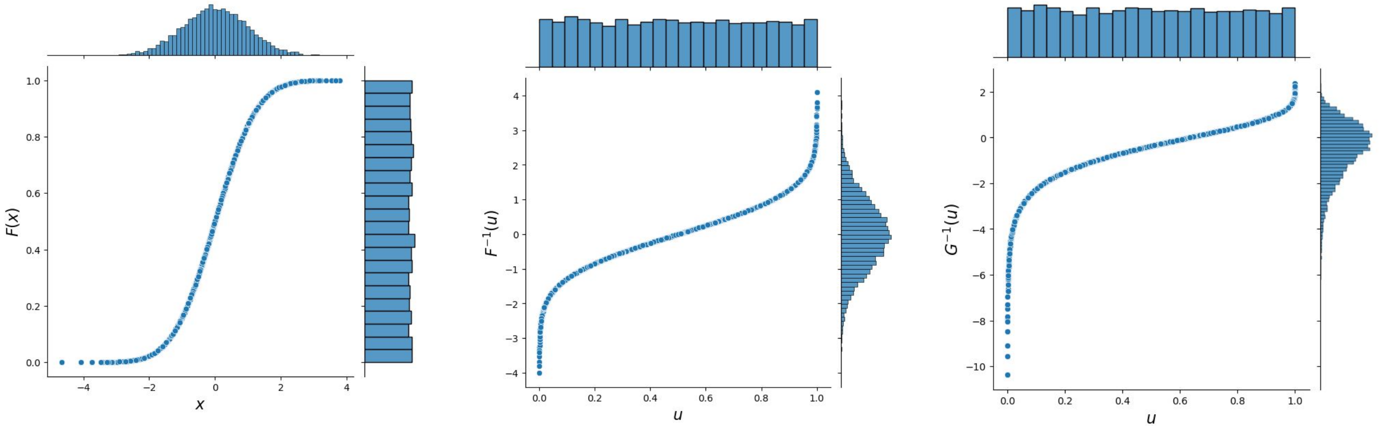
- Copulas[1]
- Conditional exceedance model[2]

1. Nelsen, R.B.: An Introduction to Copulas. Lecture Notes in Statistic vol. 139, Springer, Berlin Heidelberg New York (1999)
2. Heffernan J.E. & Tawn J.A. A conditional approach for multivariate extreme values (with discussion). J R Stat Soc B 2004, 66, 497–546.

Engineering and Physical Sciences Research Council
Oxford Programme for Sustainable Infrastructure Systems
eci
UNIVERSITY OF OXFORD

Copulas: probability integral transform

Probability integral transform (PIT)



X ⟶ Uniform   Uniform ⟶ X   Uniform ⟶ Gumbel

Copula

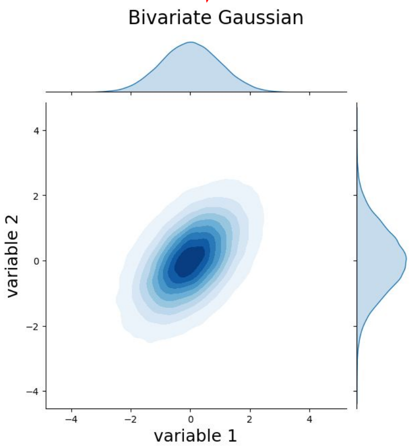
Multivariate distribution defined by $C(U_1, U_2, \ldots, U_n)$ where $U_i \sim \mathrm{Uniform}(0,1)$

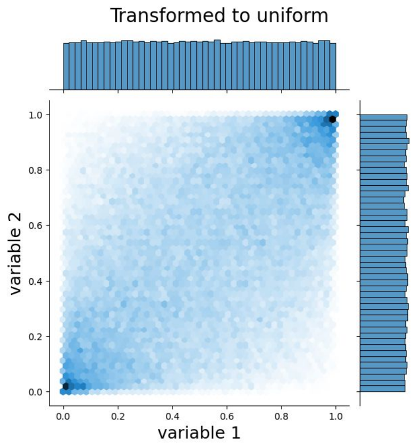
Copulas: fitting and sampling

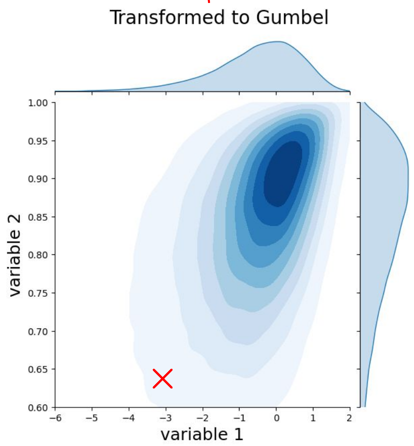
1. Fit
2. Simulate

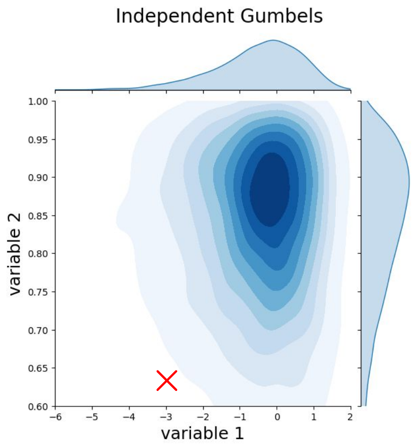
Transform
F(x)

1









$$X \sim \mathcal{N}_2(\mu, \Sigma) \qquad X \in \mathbb{R}^2$$

$$\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

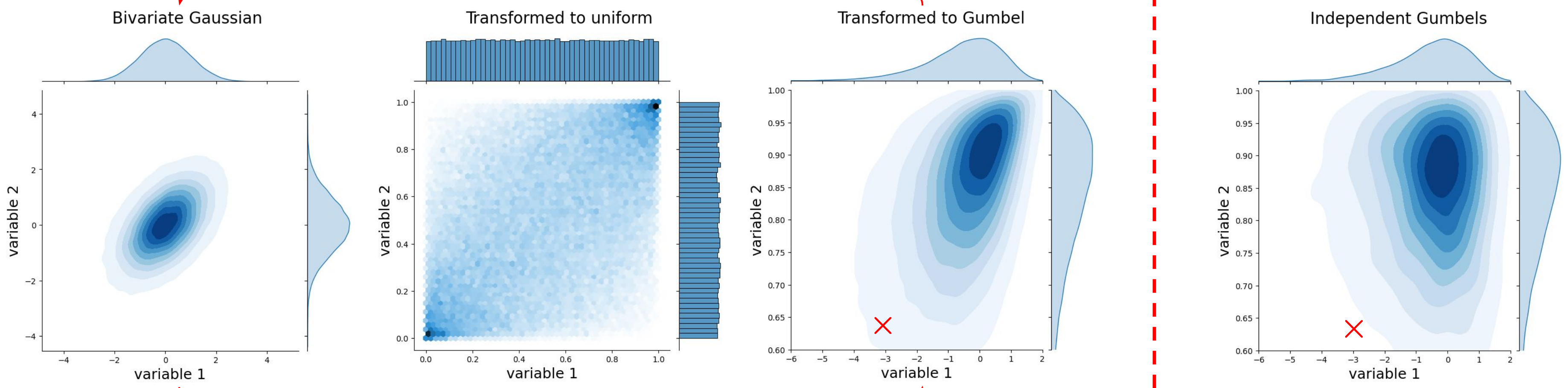
2

Transform
$F^{-1}(x)$

Copulas: fitting and sampling

Transform F(x)

1



$$X \sim \mathcal{N}_2(\mu, \Sigma) \quad X \in \mathbb{R}^2$$

$$\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

2

Transform $F^{-1}(x)$

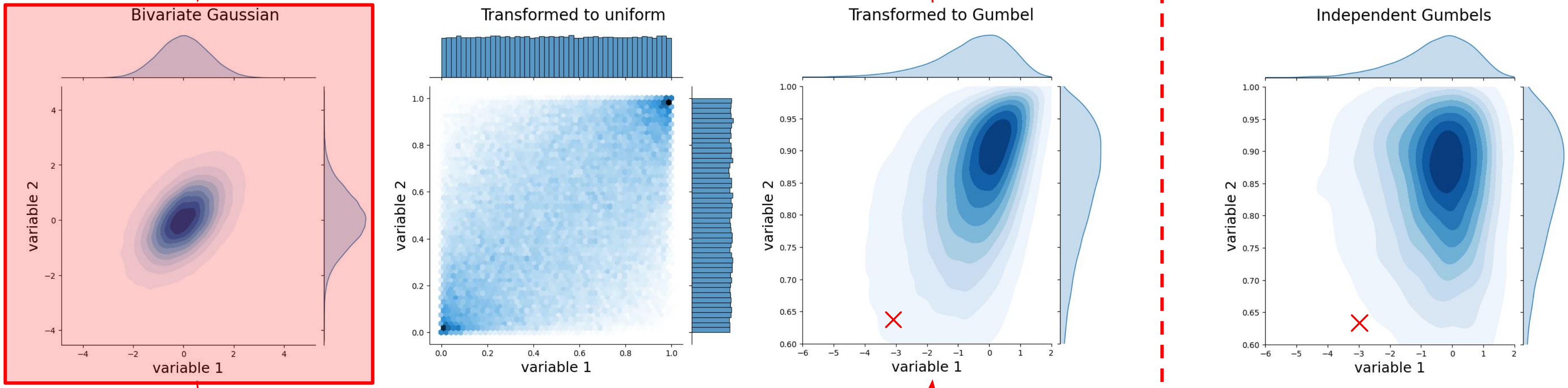
an *m*-dim elliptical copula has at least *m*(*m*−1)/2 parameters[1]

1. Schölzel, C. & Friederichs, P. Multivariate non-normally distributed random variables in climate research – introduction to the copula approach. Nonlinear Processes in Geophysics (2007)

Copulas: fitting and sampling

Transform
F(x)

1



$X \sim \mathcal{N}_2(\mu, \Sigma) \quad X \in \mathbb{R}^2$

$$\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

2

Transform
$F^{-1}(x)$

while my_mcmc: gently(samples)

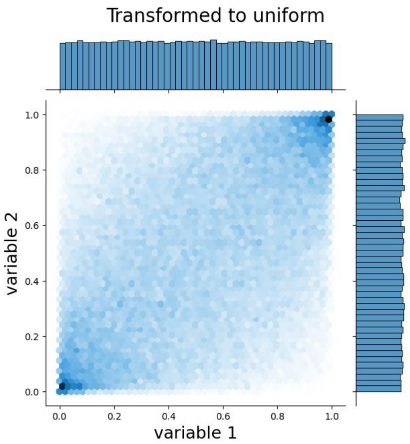
Copulas: fitting and sampling

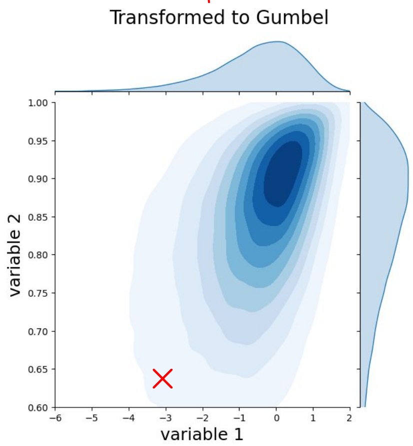
Transform
F(x)

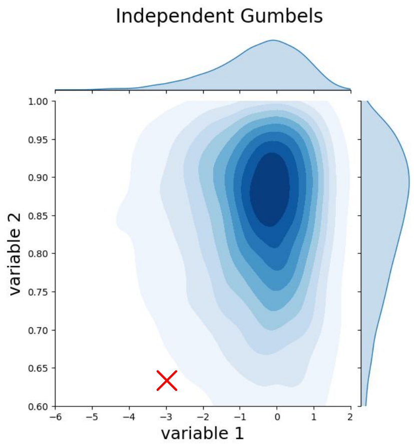
1

Deep generative model

- High dimensions
- Nonlinear relationships
- Few assumptions
- Implicitly learns distribution







$X \sim \mathcal{N}_2(\mu, \Sigma) \quad X \in \mathbb{R}^2$

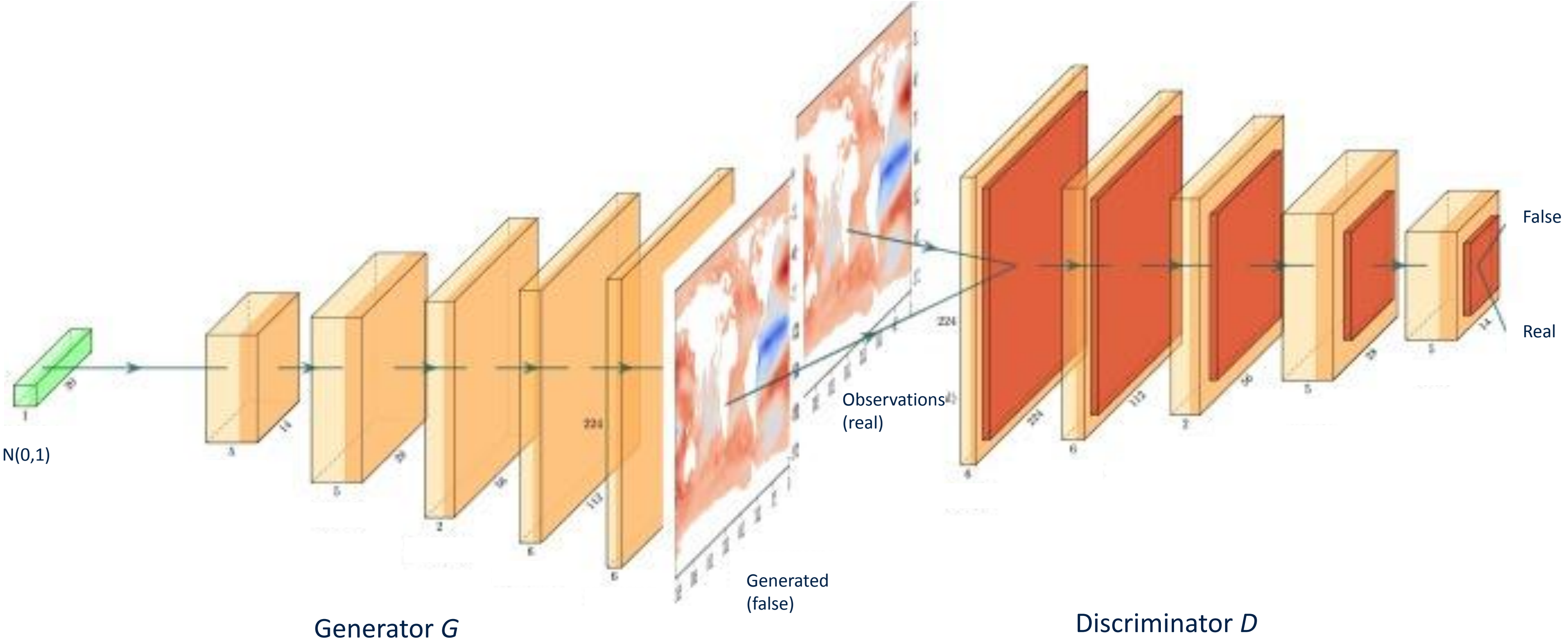
$$\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

2

Transform
$F^{-1}(x)$

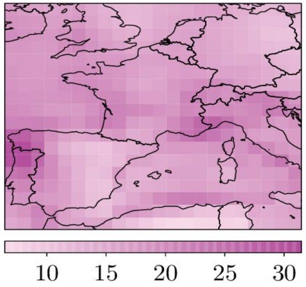
Deep generative model: generative adversarial networks (GAN)

$$\min_G \max_D V(D,G) = \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))].$$



1. Reimers & Requena-Mesa *Deep Learning - an Opportunity and a Challenge for Geo- and Astrophysics, Chapter 13* (2020)

Boulaguiem *et al.* (2022) evtGAN model

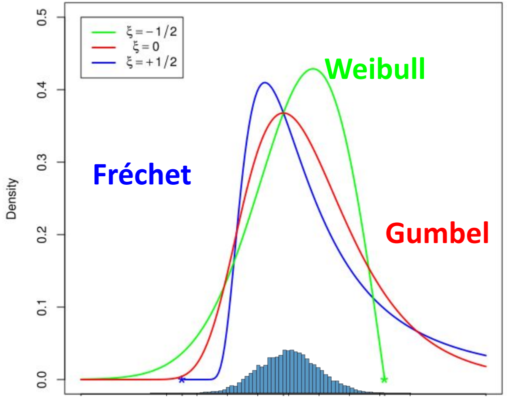


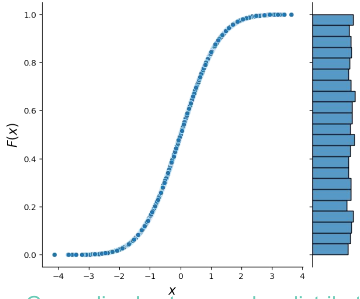
Annual maximum temperature over Western Europe

Data

- Annual maximum temperature OR precipitation
- Western Europe

Generalised extreme value (GEV) densities





[Generalized extreme value distribution - Wikipedia]

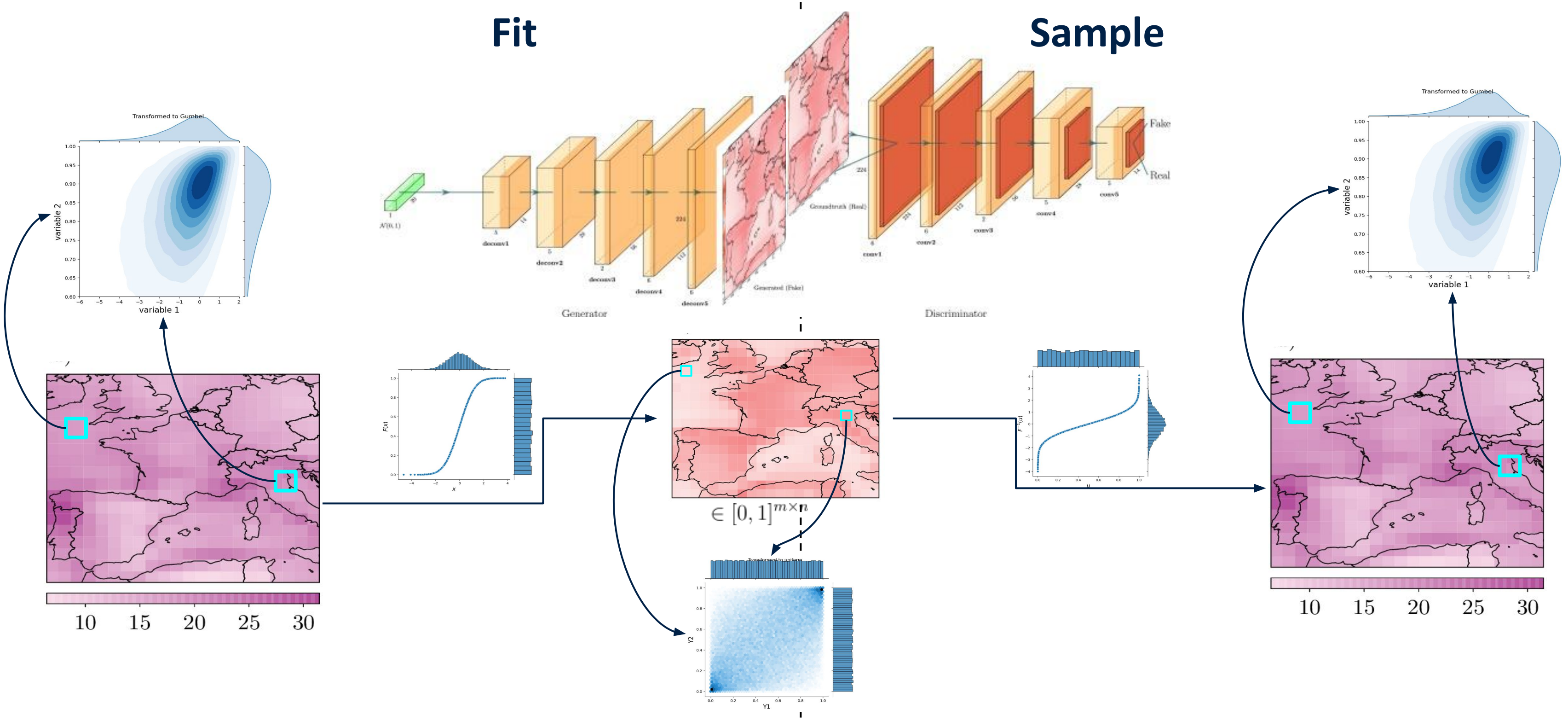
Generalised extreme value (GEV) distribution

$$F(z) = \begin{cases} \exp\left(-\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\xi}\right), & \xi \neq 0 \\ \exp\left(-\exp\left(\frac{x-\mu}{\sigma}\right)\right), & \xi = 0 \end{cases}$$

Boulaguiem *et al.* (2022) *Modeling and simulating spatial extremes by combining extreme value theory with generative adversarial networks*

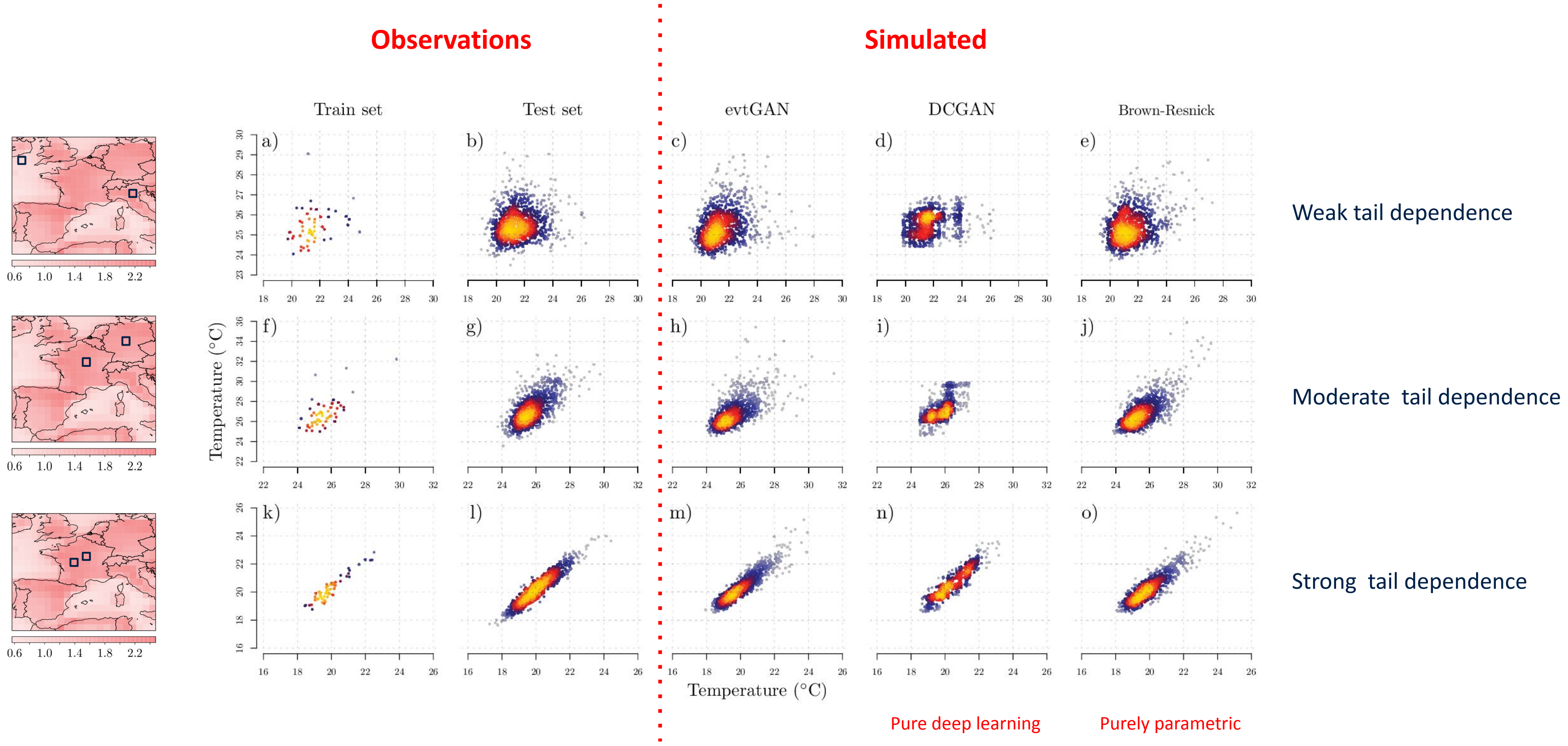
Engineering and Physical Sciences Research Council
Oxford Programme for Sustainable Infrastructure Systems
eci
UNIVERSITY OF OXFORD

Boulaguiem *et al.* (2022) evtGAN model



Boulaguiem *et al.* (2022) *Modeling and simulating spatial extremes by combining extreme value theory with generative adversarial networks*

Boulaguiem *et al.* (2022) evtGAN results



Boulaguiem *et al.* (2022) *Modeling and simulating spatial extremes by combining extreme value theory with generative adversarial networks*

Boulaguiem *et al.* (2022) evtGAN: what's left to do?

- Demonstrates appropriateness of deep generative models for learning dependence
- Not yet useful for our analyses

Potential improvements

- One year time window: too large for co-occurring events
- Univariate
- Single snapshot in time

Engineering and Physical Sciences Research Council
OPSIS Oxford Programme for Sustainable Infrastructure Systems
eci
UNIVERSITY OF OXFORD

Our approach: hazardGAN

- Train on multivariate data at same time ✅
- Smaller (daily/weekly) time windows ✅

Remaining to do:

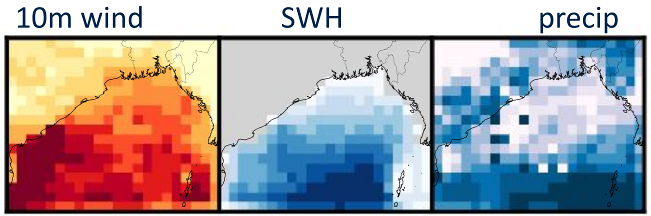
- Event-based approach ✅
- Different deep generative model architectures ✅☐☐
- Temporal extension ☐

Engineering and Physical Sciences Research Council
Oxford Programme for Sustainable Infrastructure Systems
University of Oxford

hazardGAN: data

- ERA5 reanalysis data 2013-2022
    - 3285 samples
- Daily maximum
    - 10m wind speed
    - Significant wave height
    - Total precipitation
- Bay of Bengal
    - 10-25º East
    - 80-95º North
- Resolution
    - Original: 0.25º x 0.25º
    - Resampled to 0.5º x 0.8º
    - 18 x 22 = 396 pixels per hazard
- Split into training/valid/test sets

$$\binom{396}{2} = 78,210 \text{ unordered pixel pairs for a wind only.}$$

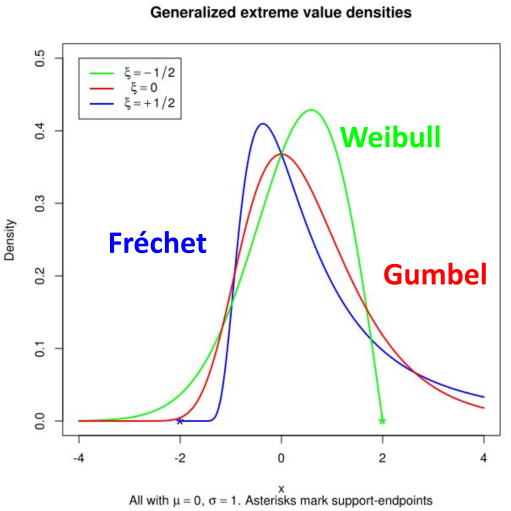


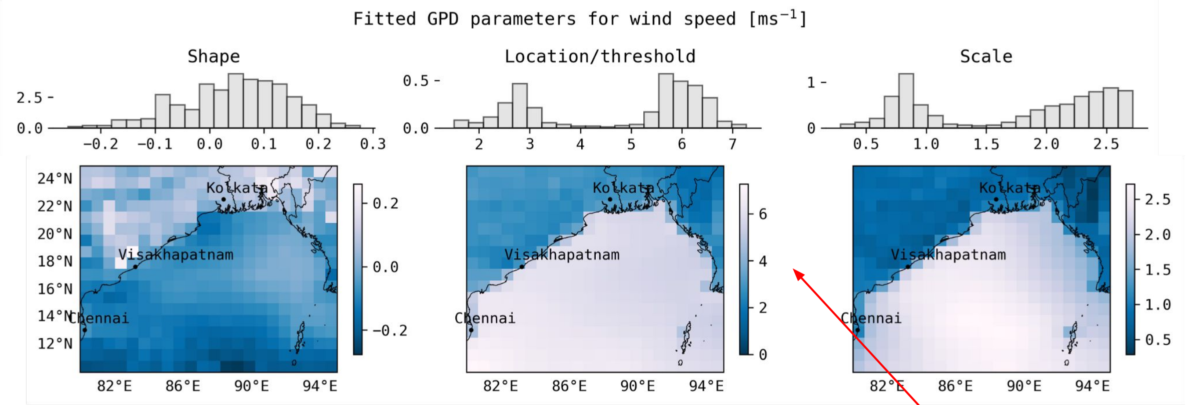
Training sample wind speed (left), SWH (center), precipitation (right) for single day

Boulaguiem *et al.* (2022) *Modeling and simulating spatial extremes by combining extreme value theory with generative adversarial networks*

Engineering and Physical Sciences Research Council
Oxford Programme for Sustainable Infrastructure Systems
eci
UNIVERSITY OF OXFORD

hazardGAN: parametric model





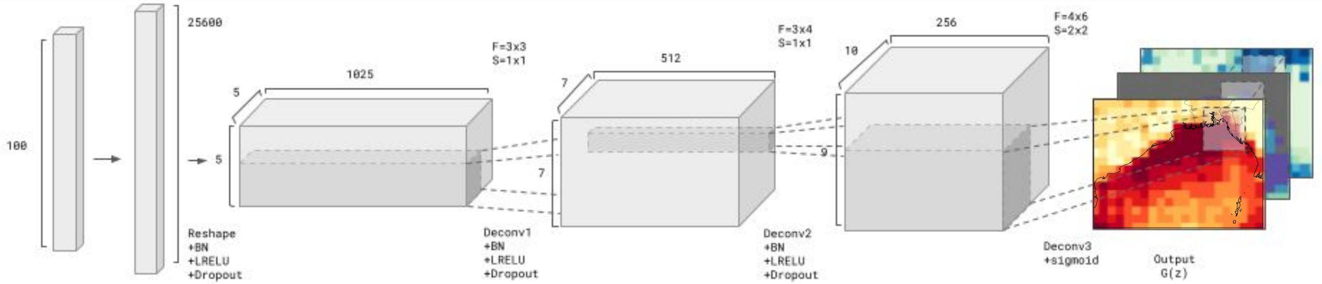
colour bar

<u>GEV fit to daily maxima:</u>

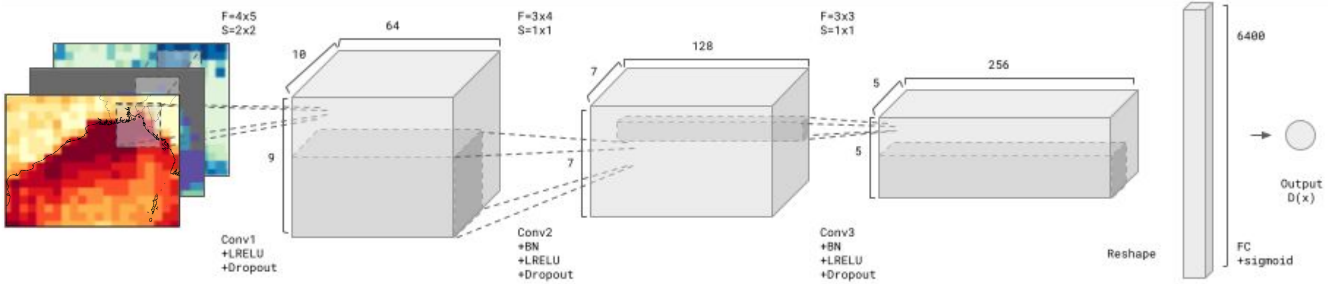
- Exhibits autocorrelation ⇒ violates GEV ML assumption
    - Event-based approach in future work
- Fit OK (for now)
- Weibull offshore winds ☑️

hazardGAN: generative model

Generator



Discriminator



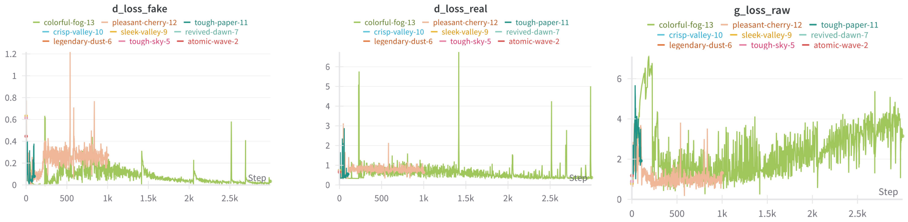
Engineering and Physical Sciences Research Council

Oxford Programme for Sustainable Infrastructure Systems

UNIVERSITY OF OXFORD

hazardGAN: generative model

Parameter	Values	Selected Value	Description
seed	[0, 1, 2, 6, 7, 42]	7	Included in sweep for reproducibility.
learning_rate	0.0001 - 0.0003	0.00013367626823798716	Controls the optimizer step size and rate of convergence.
beta_1	0.1 - 0.5	0.22693882275467836	Controls the exponential decay rate for the first moment estimates of the gradient for the Adam optimizer.
lrelu	0.1 - 0.4	0.2991161912395133	The gradient to assign to negative values in the Leaky ReLU function.
dropout	0.3 - 0.6	0.44053850596844424	Frequency the dropout function sets inputs to zero in training to prevent overfitting.
training_balance	[1, 2]	2	Ratio of training loops for discriminator vs. generator.



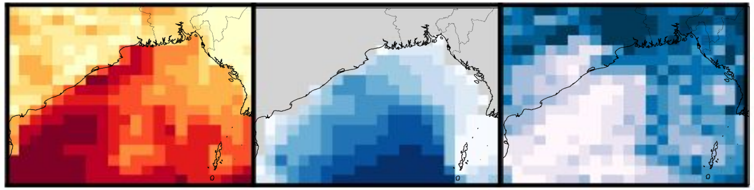
- 1000 epochs
- Training size N=1000
- Batch size 50
- Adam optimizer
- WandB Bayesian hyperparameter tuning

Final tuned parameters

hazardGAN: results

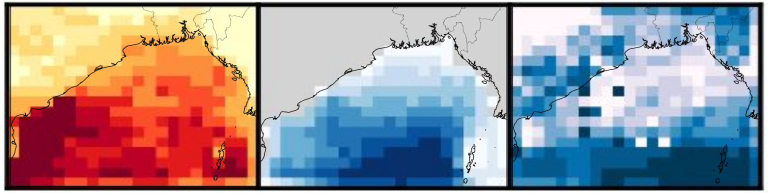
Training sample

10m wind | SWH | Total precip



Generated sample

10m wind | SWH | Total precip

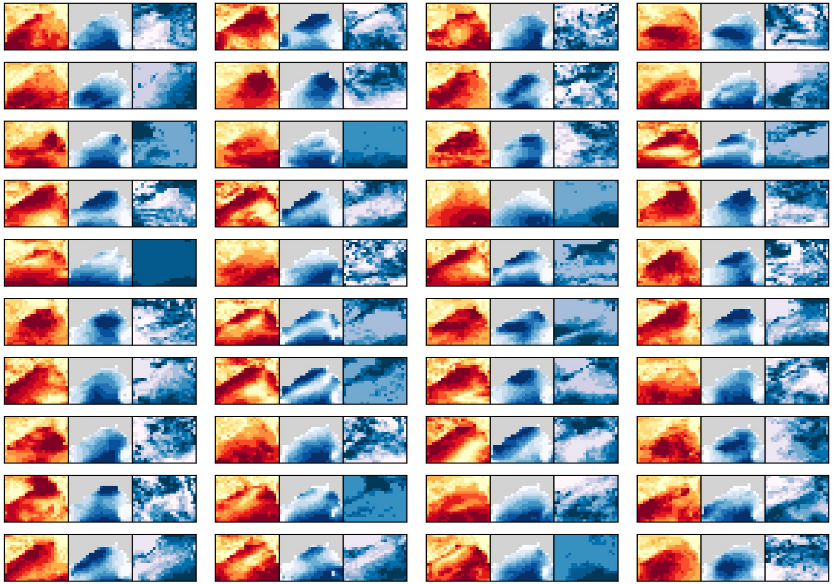


Samples, where each three images are wind speed (left, orange), SWH (center, blue), precipitation (right, blue) for single day

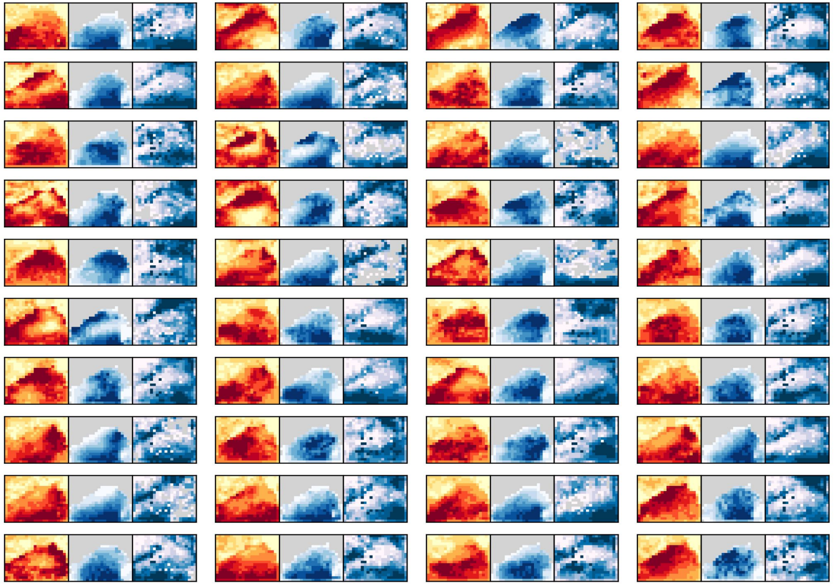
(Colour scales are the same)

hazardGAN: results

Training sample



Generated sample



Samples, where each three images are wind speed (left, orange), SWH (center, blue), precipitation (right, blue) for single day

Engineering and Physical Sciences Research Council

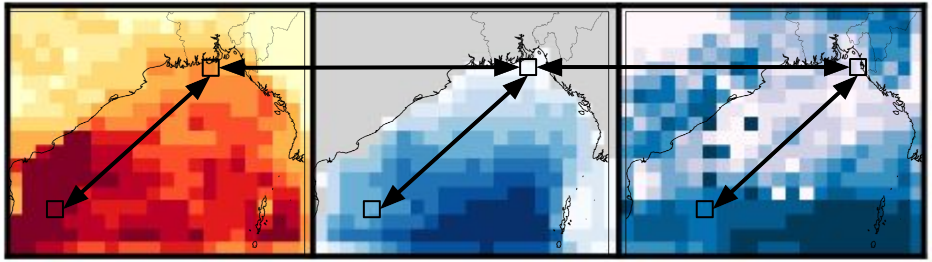
Oxford Programme for Sustainable Infrastructure Systems

UNIVERSITY OF OXFORD

Does hazardGAN learn spatial/multivariate relationships?

Evaluation:
Is it capturing relationships
- Across space?
- Between variables?

Is it capturing *tail* relationships
- Across space?
- Between variables?



Metrics
1. Pearson (regular) correlation
2. Extremal coefficient θ in [1, D], where D is # dimensions
3. Extremal correlation χ=(D-θ)/(D-1) (analogous to Pearson)

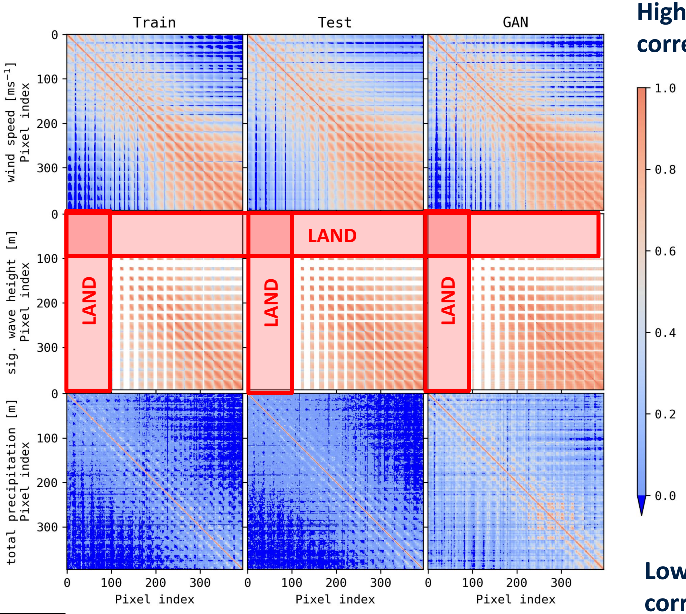
$$\hat{\theta}_{123} = \frac{N}{\sum_{n=1}^{N} \min\left(\frac{1}{Y_{n1}}, \frac{1}{Y_{n}2}, \frac{1}{Y_{n3}}\right)} \qquad Y_{ni} = -\log(U_{ni})^{-1}$$

θ↑: tail independence
θ↓: tail dependence

hazardGAN: dependence across space

Correlation

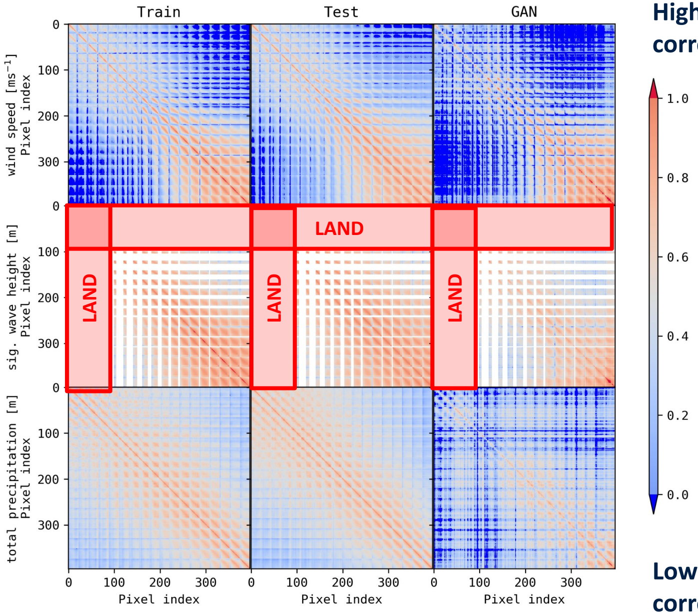
High correlation



Low correlation

Extremal correlation χ

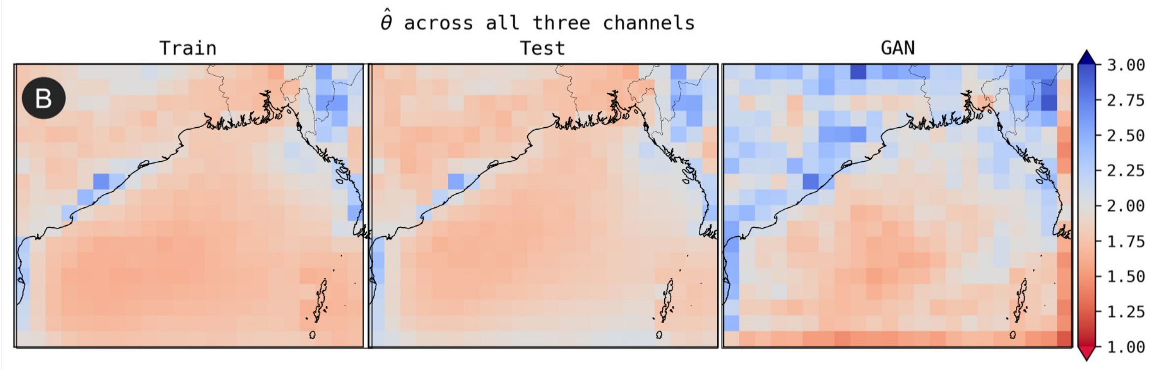
High tail correlation



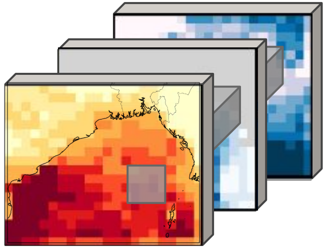
Low tail correlation

LAND - LAND	LAND - WATER
WATER - LAND	WATER- WATER

hazardGAN: dependence between variables



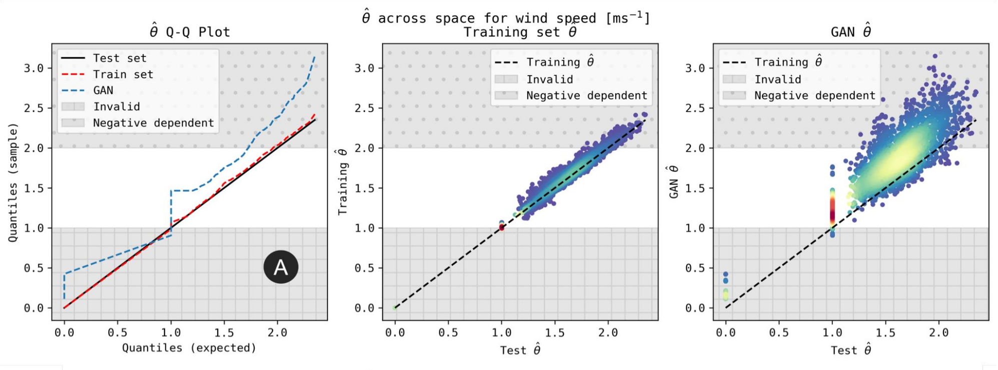
$$\hat{\theta}_{123} = \frac{N}{\sum_{n=1}^{N} \min\left(\frac{1}{Y_{n1}}, \frac{1}{Y_{n}2}, \frac{1}{Y_{n3}}\right)} \qquad Y_{ni} = -\log(U_{ni})^{-1}$$



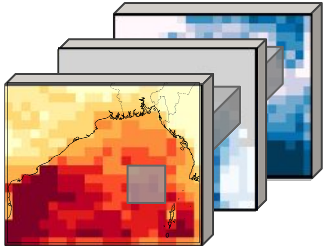
θ↑: tail independence

θ↓: tail dependence

hazardGAN: dependence between variables



$$\hat{\theta}_{123} = \frac{N}{\sum_{n=1}^{N} \min\left(\frac{1}{Y_{n1}}, \frac{1}{Y_{n}2}, \frac{1}{Y_{n3}}\right)} \qquad Y_{ni} = -\log(U_{ni})^{-1}$$



θ↑: tail independence

θ↓: tail dependence

Summary

Cons

- Underestimating tail dependence
- Black box dependence
- Inherits training data errors

Pros

- Learning overall dependence structure
- Much higher dimensions
- Nonlinear relationships
- Flexible tail dependence
- Parametric extremes extrapolation
- Flexible for regions/hazards
- Fast once trained

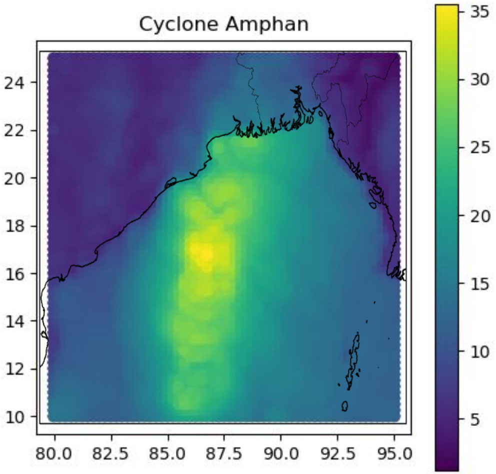
Next

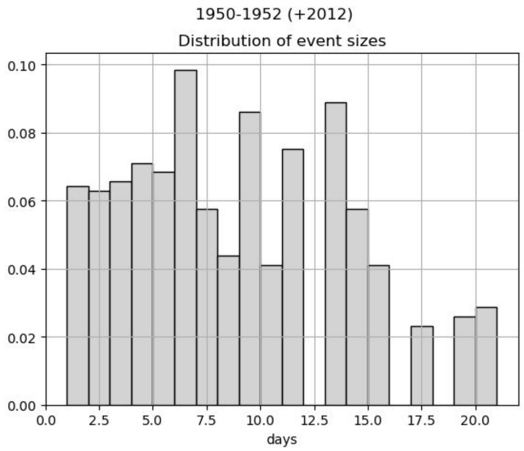
- Daily maxima → event maxima ✅
- Deep learning architectures
  - Wasserstein GAN-GP ✅
  - Heavy-tailed latent space ☐
  - Flow-matching ☐
- Temporal info ☐

UKRI Engineering and Physical Sciences Research Council
OPSIS Oxford Programme for Sustainable Infrastructure Systems
eci
UNIVERSITY OF OXFORD

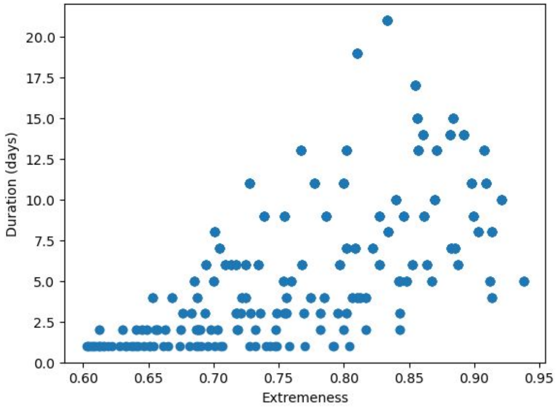
Next steps: temporal events

- Define event by relative *extremeness*
- Daily maxima → eventwise maxima
  - No more autocorrelation



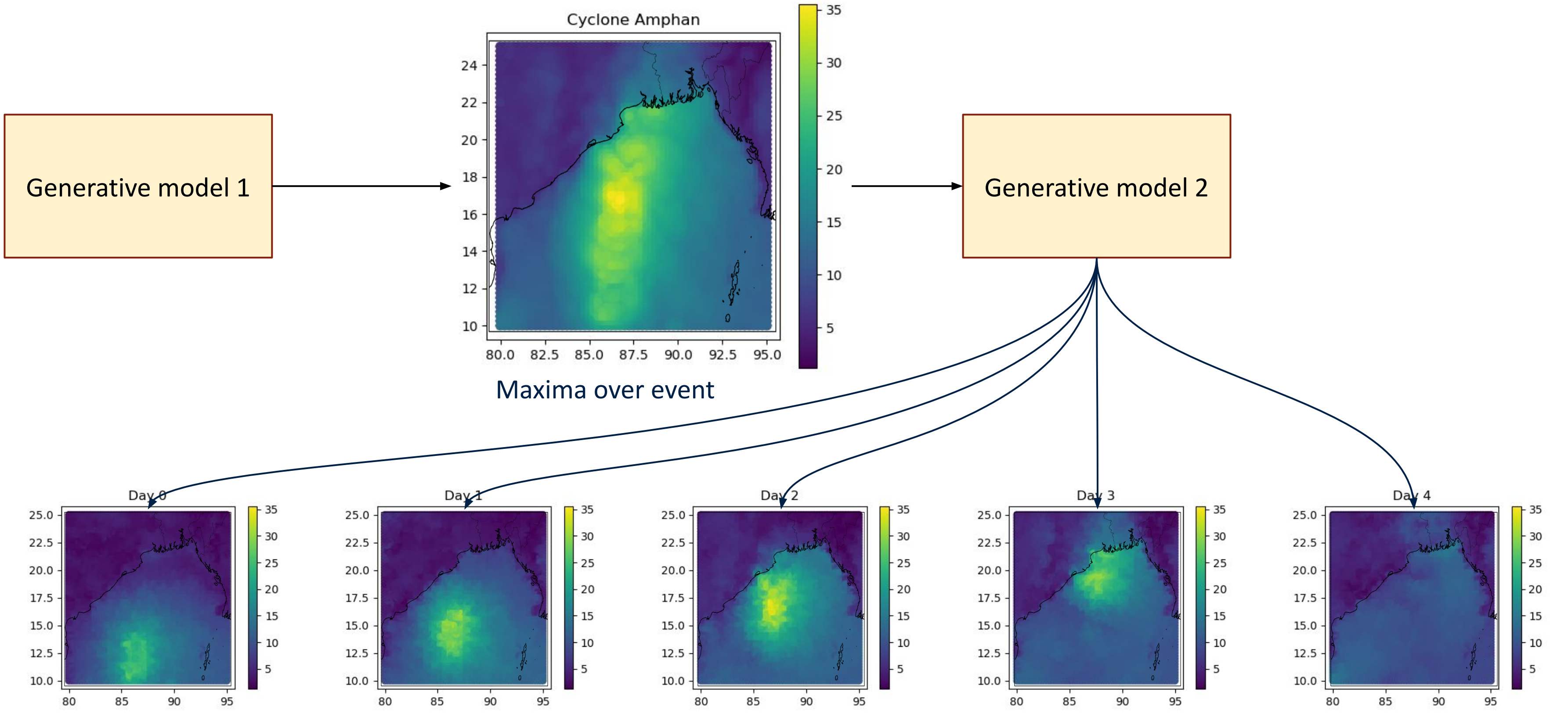


Max event lasts 20 days

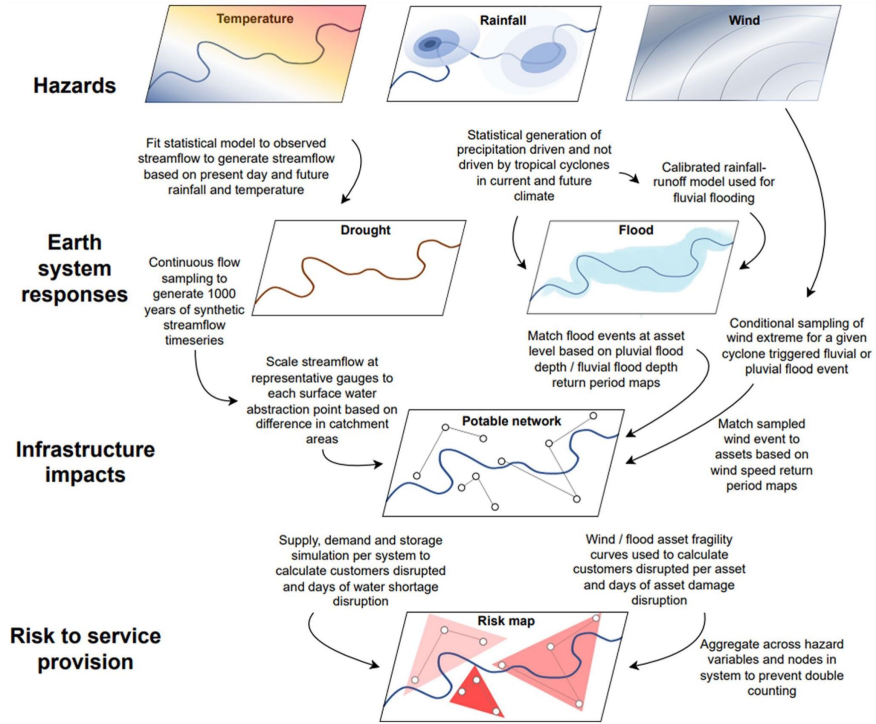


Extremeness and duration of event are correlated

Next steps: temporal events



Applications



- Flexible:
  - Variables
  - Regions
- Scenario modelling
  - Water resources:
    - Temperature
    - Precipitation
  - Biodiversity:
    - Precipitation
    - Seasons
  - Climate risk:
    - Compound hazards

Becher *et al.* (2023) *A multi-hazard risk framework to stress-test water supply systems to climate-related disruptions*

Thanks for listening!

alison.peard@ouce.ox.ac.uk

OPSIS Oxford Programme for Sustainable Infrastructure Systems

UNIVERSITY OF OXFORD

Engineering ar
Physical Scienc
Research Cour

Applications

- Framework can be applied to other regions/climate variables
- Generates 1000s of large-scale plausible and coherent events in seconds
- Scenario modelling
- Quantify risk from compound hazards over large scales
- Input for numerical models

High level impact slide, what

What will my methods change IRL?

Currently

-

With temporal extension

- Input for numerical simulations, e.g., compound flood simulations

Engineering and Physical Sciences Research Council

Oxford Programme for Sustainable Infrastructure Systems

eci

UNIVERSITY OF OXFORD

Large-scale hazard maps: implications

- Food systems vulnerable to multiple co-occurring droughts and heatwaves in crop-producing regions (Anderson et al., 2019; Mehrabi and Ramankutty, 2019, Boulaguiem 2022).

- Spatially compounding events: Zscheischler et al . 2020
  - Multiple connected locations affected by same/different hazards withing a time window
  - Infrastructure systems affected in more than once location at once

- Widespread flooding/cyclones that happen in close succession stretch emergency services

- Different hazards in same location and time can cause worse event than if single drivers

- Rain and strong winds co-occur often and can lead to compound flooding in coastal areas: pluvial and coastal