Simulating spatially resolved multi-hazards with deep learning

ALISON PEARD
JIM W. HALL
YU MO

ENVIRONMENTAL CHANGE INSTITUTE (ECI)

UNIVERSITY OF OXFORD

Context

Large-scale climate risk assessments

- Affiliations:
  - Oxford Programme for Sustainable Infrastructure Solutions (OPSIS) – PhD, Supervisor: Prof. Jim W. Hall
  - Oxford Infrastructure Analytics (OIA) – part-time
- **Focus**: **large-scale climate risk** assessments to support **long-term planning** and infrastructure **investment**
- **Projects**: Somalia, Nigeria, the Organisation of Eastern Caribbean States (OECS), Jamaica, Tanzania, Morocco, and Bangladesh
- Calculate damages and losses from **present and future** climate hazards
  - **local** and **aggregated over regions**
  - **indirect effects** on systems – e.g., transport networks, energy networks, supply chains
- Key ingredients:
  - Exposure data – location of critical assets
  - Hazard data – location, intensity, probability
  - Vulnerability models – susceptibility of exposed assets to damage/disruption

Climate hazard data: dimensions

Dimension	Subtypes	Example
Temporal	• internal dynamics • **when** events happens	• slow-moving storms • early-in-season or flood after a period of drought
Spatial	• **extent** of event • teleconnections	• widespread floods cause severe traffic disruption • droughts in multiple crop-producing regions
Multivariate	• compounding/ synergistic damages	• tropical cyclones + heatwaves
Probability	• probability of occurrence • non-stationarity	• 1-in-T-year event • climate change

Zscheischler et al. (2020). *A typology of compound weather and climate events.*
Heinrich et al. (preprint forthcoming). *Hurricanes that haven't happened, yet: Identifying unprecedented tropical cyclone scenarios.*
Matthews et al. (2019). *An emerging tropical cyclone–deadly heat compound hazard.*

Climate hazard data: hazard maps

Extreme quantiles across spatial domain

A 100-year hazard map is a grid:

$$\begin{bmatrix} \mathbf{x}_1 & \dots & \mathbf{x}_i \\ \vdots & & \\ \mathbf{x}_j & \dots & \mathbf{x}_d \end{bmatrix} \qquad x_i \;= F_i^{-1}(1-1/100)$$

$$= F_i^{-1}(0.99)$$

F_i hazard distribution at location i

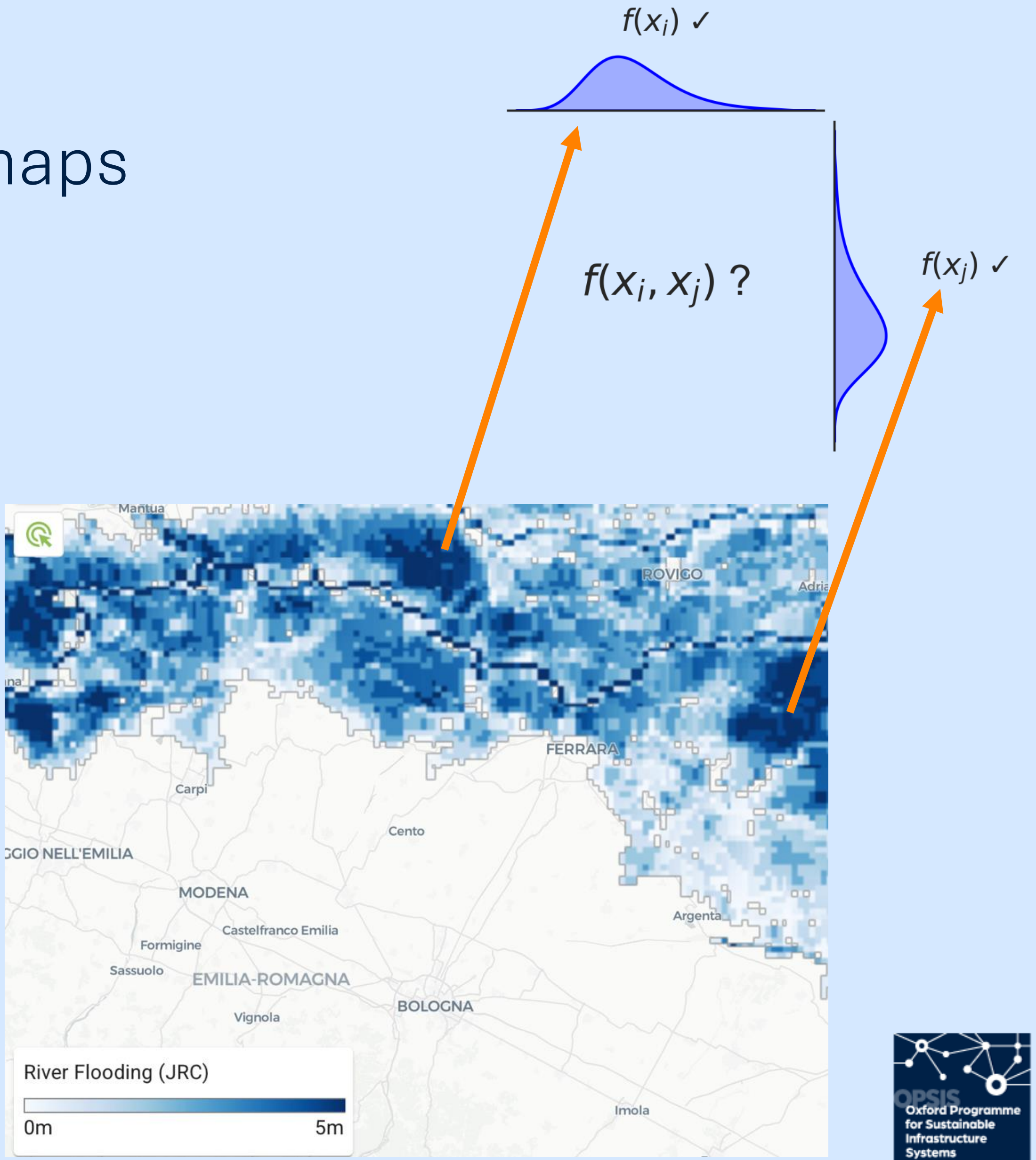
Return periods: $T = 5, 25, 50, 100, 250, 500, \dots$

Derivation: fit GEV across domain

Limitation: no joint density information

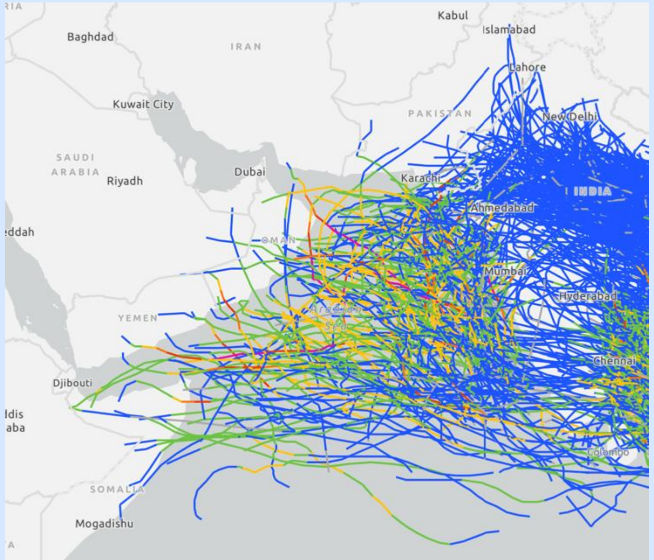
Dimensions:

- ❌ Temporal
- ❌ Spatial
- ❌ Multivariate
- ✅ Probability



100-year flood depth over Po river floodplain from [GRI Risk Viewer]

Climate hazard data: event sets and footprints



NOAA historical hurricane tracks

Event ensembles

- Large ensembles of plausible hazard events (with probabilities)

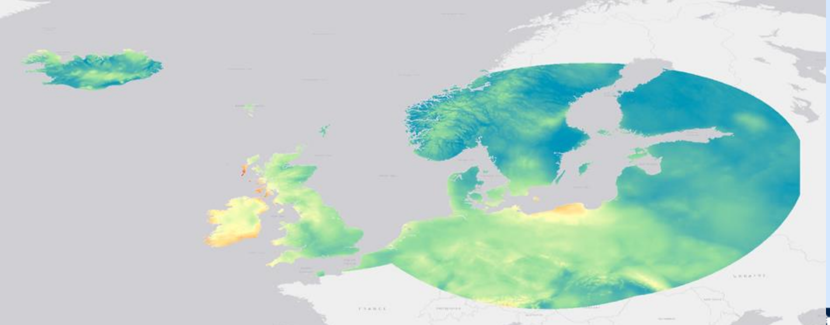
  e.g., STORM, IRIS, Fathom Global Flood Cat

- Dimensions: ✅ potentially all

Event footprints

- **Time-integrated** representation of intensity of the event over a region of interest
- Used extensively in catastrophe (Cat) modelling

Event	Metric	Spatial
Earthquake	peak ground acceleration	affected area
Flood	maximum flood depth	floodplain
Storm	peak gust wind speed	storm track


ECMWF Windstorm tracks and footprints derived from reanalysis over Europe between 1940 to present

Climate hazards: comparison

Data type	Hazard maps	Event ensembles
Pros	• Lightweight (<10 maps per hazard) • Easy-to-make • Wide variety • Full distribution at single point • Expected values over larger scales	Can capture all dimensions
Cons	• Lost dimensions: spatial, temporal, multivariate • Cannot model systemic impacts	• Difficult to create, store, and use • Limited variety, availability & transparency • Extreme event probabilities ← huge ensembles

OPSIS Oxford Programme for Sustainable Infrastructure Systems

eci

UNIVERSITY OF OXFORD

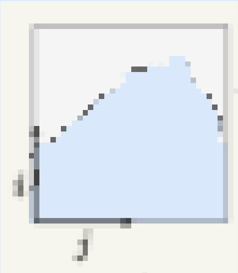
Objective: Multi-hazard event set generator

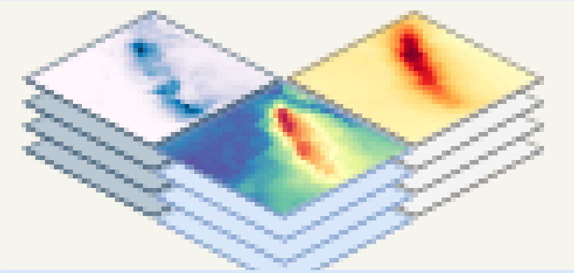
Stochastic *hazard* generator – event footprints

Essential dimensions:

- ✅ spatial
- ✅ multivariate
- ✅ probability
- ❌ temporal

Variable	Value
Dataset	ERA5 reanalysis 1940–2022
Hazards	1. Max wind speed at 10 m (ws10) 2. Total precipitation (tp) 3. Atmospheric sea-level pressure (mslp)
Total variables	64 × 64 × 3 = 12 288





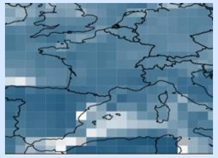
Wilks, D. S., & Wilby, R. L. (1999). *The weather generation game: A review of stochastic weather models.*

Method

Starting point: evtGAN

Boulaguiem (2022): GAN to generate **spatial extremes**

Annual max temperatures/rainfall over France/Western Europe



$$\begin{bmatrix} \mathbf{x}_1 & \dots & \mathbf{x}_i \\ \vdots & & \\ \mathbf{x}_j & \dots & \mathbf{x}_d \end{bmatrix}$$

Fit GEV to $\mathbf{x}_i \to \mu_i, \sigma_i, \xi_i$

Fit empirical CDFs $\hat{F}_i$ to $\mathbf{x}_i$

$$\Rightarrow u_i = \hat{F}_i(x_i) \sim \mathcal{U}[0,1]$$

GAN learns

$$C\left(\hat{F}_1(x_1), \dots, \hat{F}_d(x_d)\right)$$
$$\downarrow$$
$$\mathrm{GAN}(u_1, \dots, u_d)$$

Draw new sample:

$$\mathrm{GAN} \to u_1^*, \dots, u_d^*$$

Get extremes $F^{-1}_{\mu_i, \sigma_i, \xi_i}(u_i^*)$

Results

- Better representation of **extremal dependence**
- EVT-informed **extrapolation**

Modifications

- Single-hazard → multi-hazard (trivial)
- Annual maxima → event-based

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

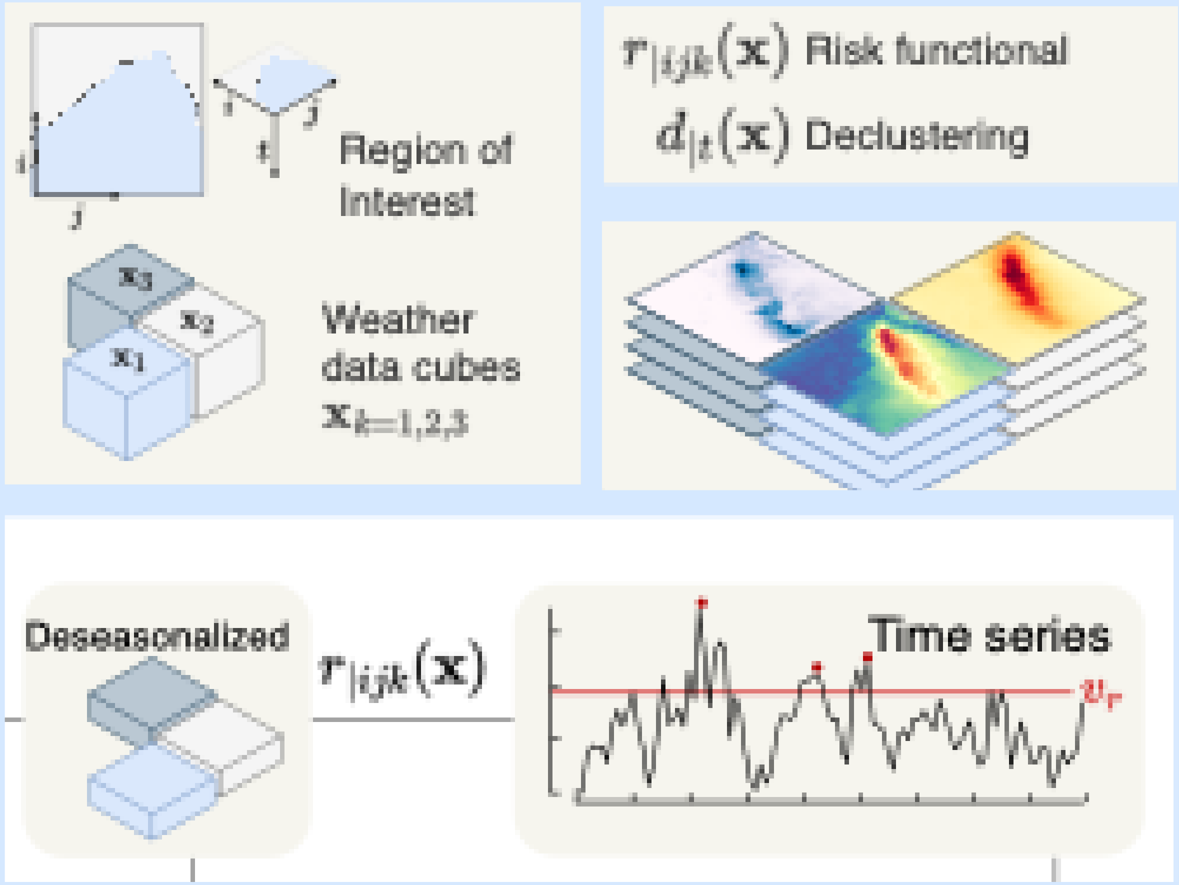
Event-based approach: POT

Univariate POT

- Peaks-over-thresholds to model events
- Declustering extracts events and handles serial correlation
- Similar to Method of Independent Storms (MIS; Cook, 1982)

Multivariate POT

- **Requirements:** Independence
- Hazard severity function $r_{|ijk}(\mathbf{x})$, e.g., max wind speed over region
- Independent events where $r_{|ijk}(\mathbf{x}) \geq v_r$ separated by run length ℓ
- Grid-search for ℓ and v_r



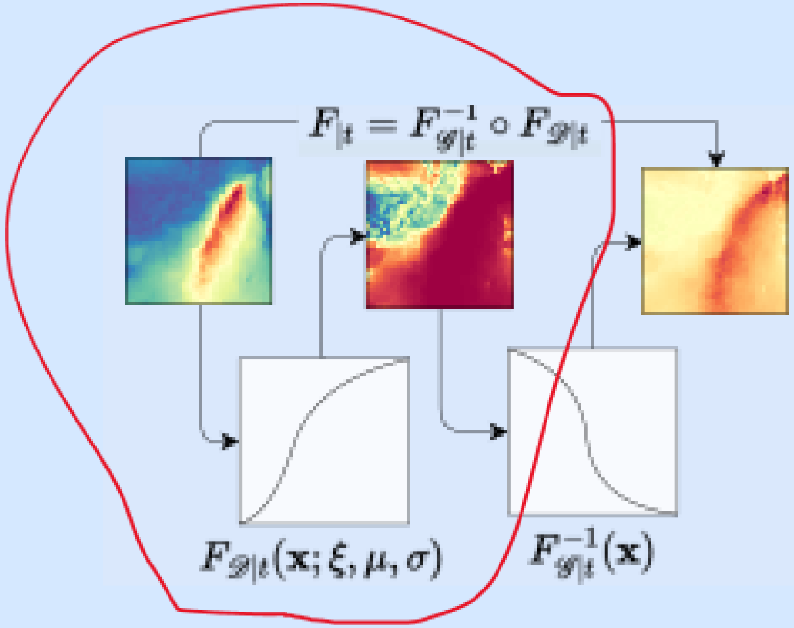
Cook . (1982). *Towards better estimation of extreme winds.*

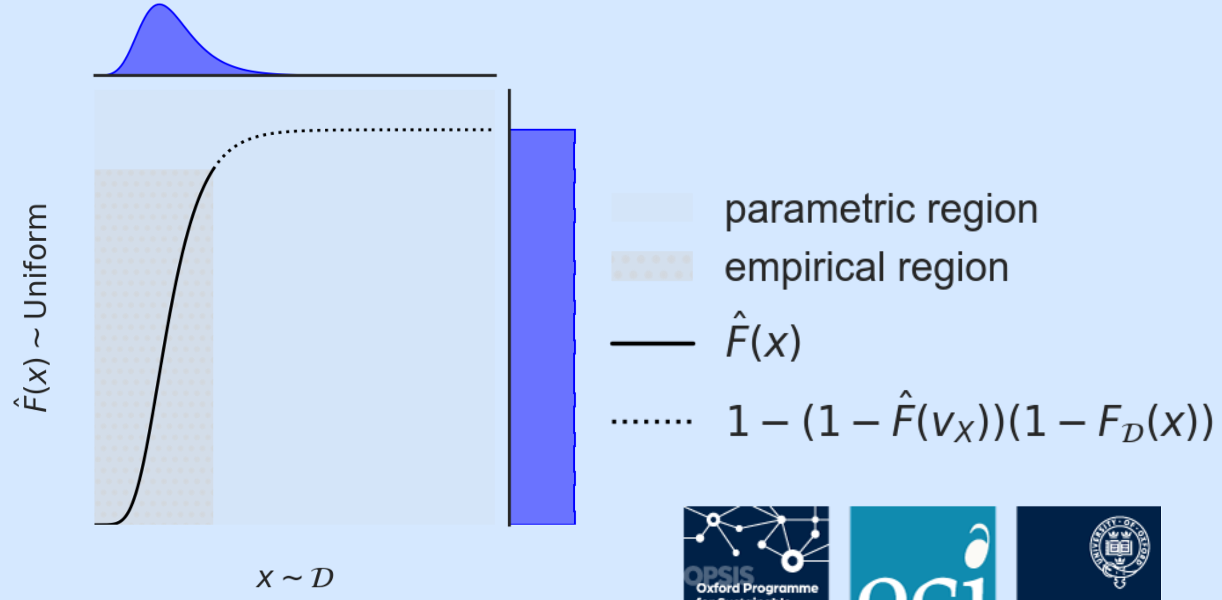
How to handle spatially localised events?

Events are spatially **localised**: only a small subset will be extreme for each event.

$$\tilde{F}(x) = \begin{cases} 1 - \left(1 - \hat{F}(v_X)\right)\left(1 - F_{\mathcal{D}}(x)\right) & \text{for } x > v_X \\ \hat{F}(x) & \text{for } x \le v_X \end{cases}$$

- $\mathcal{D}$ – generalised Pareto distribution:
  - μ – loc
  - σ – scale
  - ξ – shape

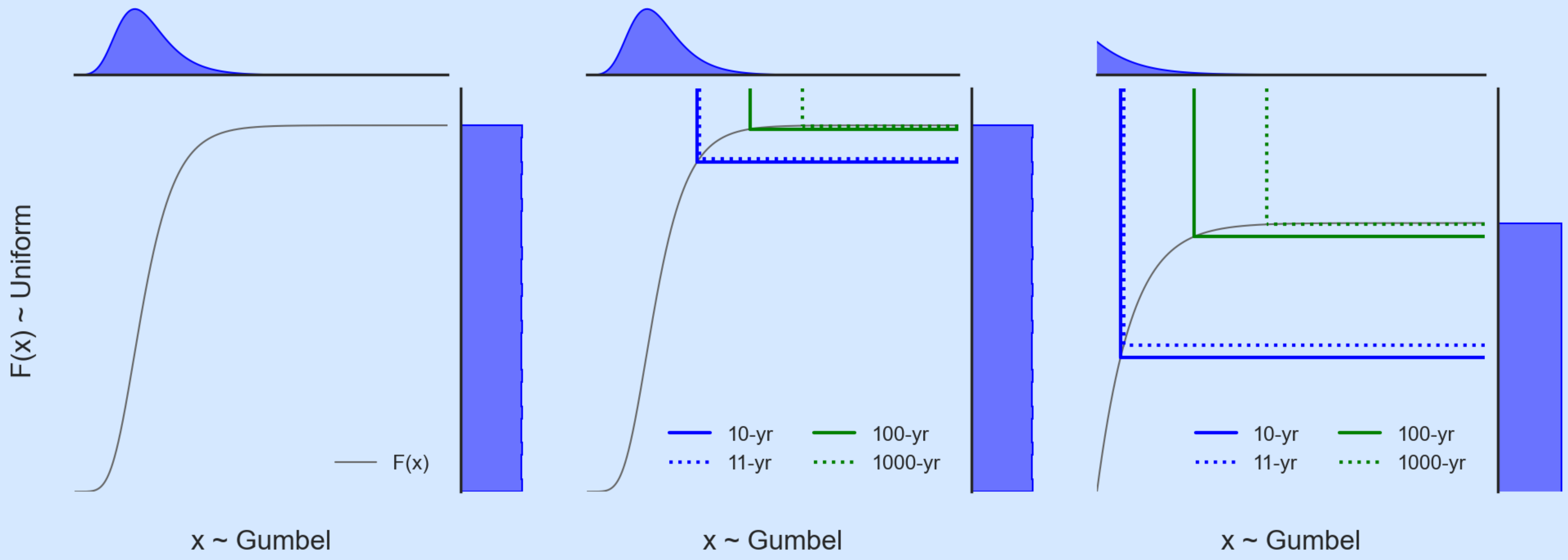




Better tail representation

Want to learn $C\left(\tilde{F}_{\mu_1,\sigma_1,\xi_1}(x_1),\ldots,\tilde{F}_{\mu_d,\sigma_d,\xi_d}(x_d)\right)$

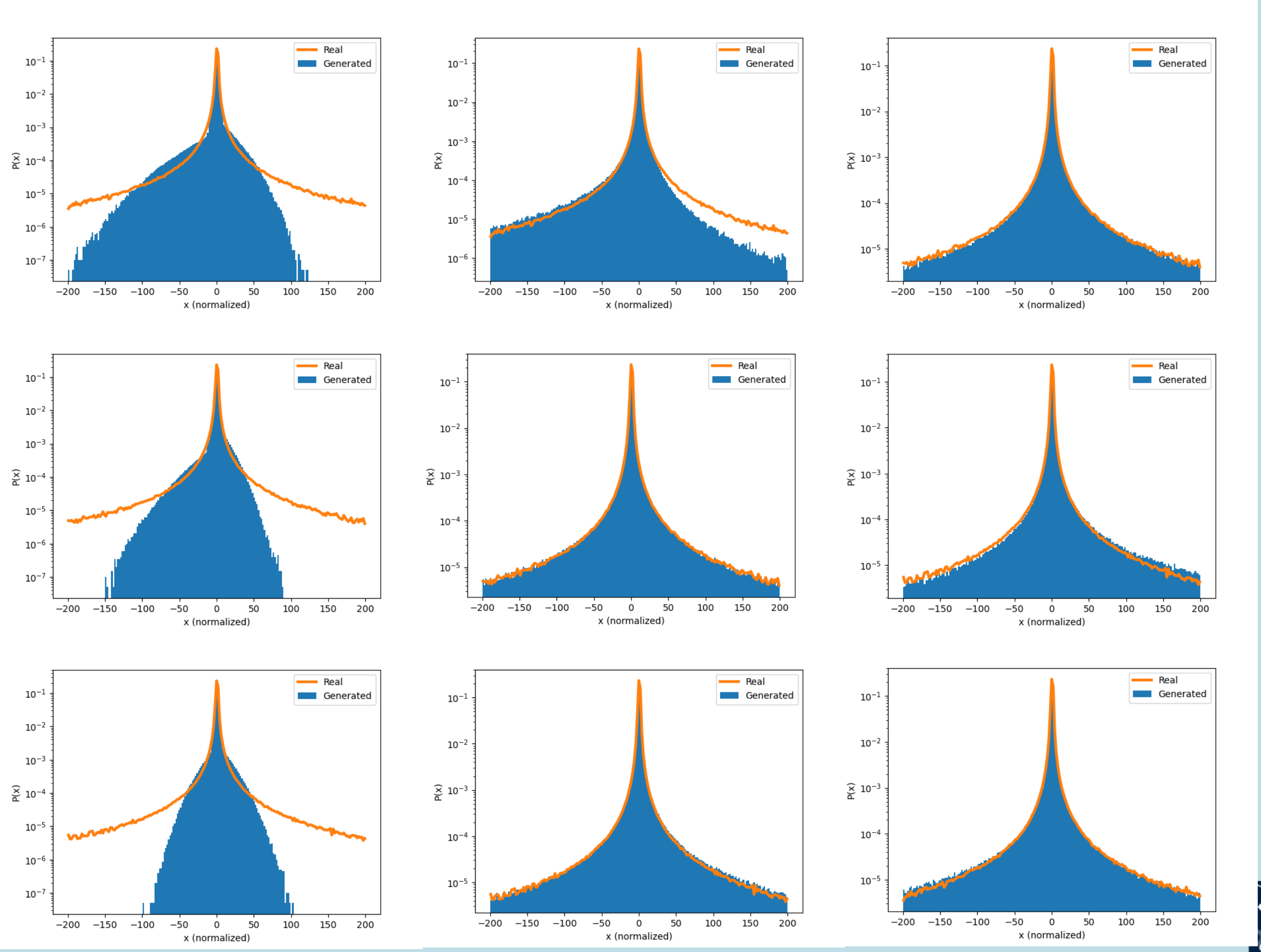
But uniform space **compresses** tails



GAN + heavy-tails

Pareto GAN

GAN + light-tails + PIT

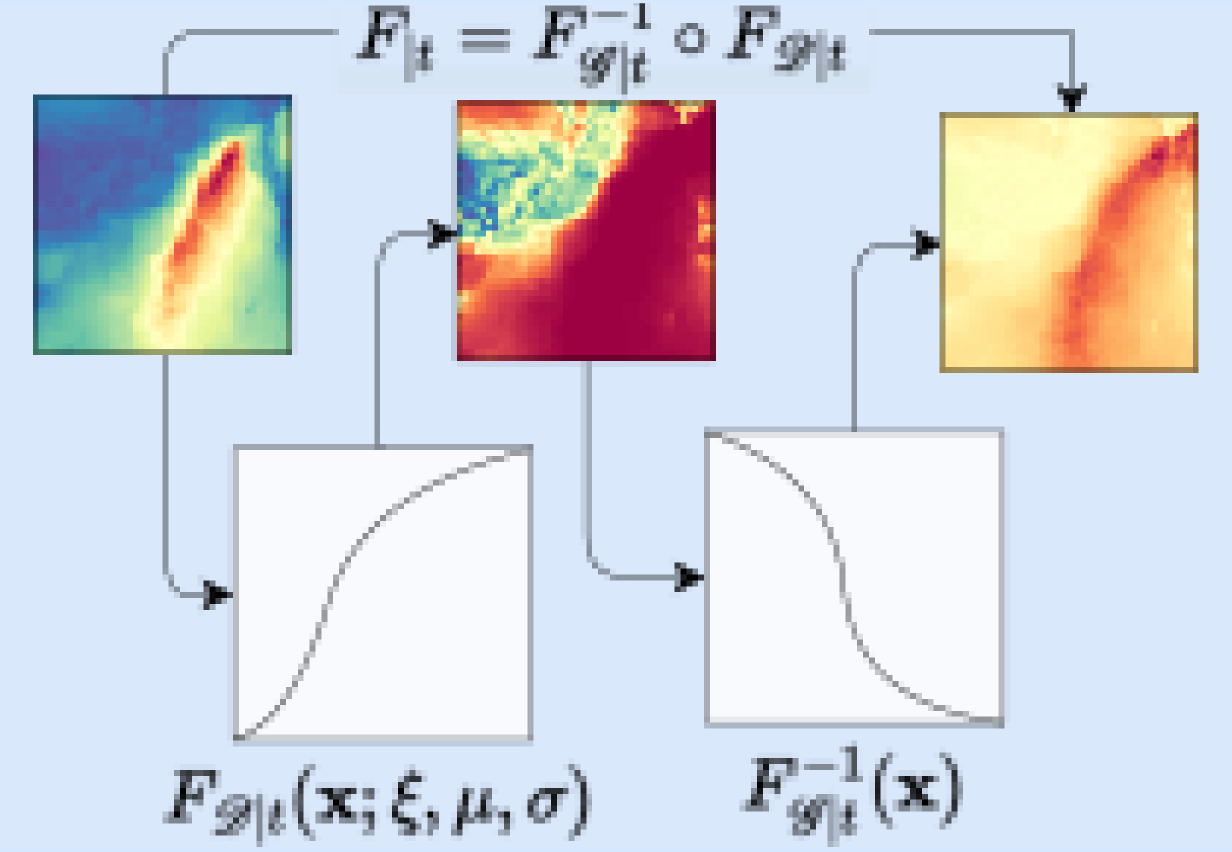


Better tail representation

Transform margins to a distribution with **light tails and** learn:

$$C\left(\mathcal{G}^{-1}\circ\tilde{F}_{\mu_1,\sigma_1,\xi_1}(x_1),\ldots,\mathcal{G}^{-1}\circ\tilde{F}_{\mu_d,\sigma_d,\xi_d}(x_d)\right) \rightarrow \text{GAN}(y_1,\ldots,y_d)$$

with $\mathcal{G}$ a Gaussian or Gumbel distribution function ($\text{MDA}_{\xi=0}$)



Training and sampling on rare events

Sampling extreme events

- ~80 years → 100–1000 samples
- StyleGAN2 with differentiable augmentation (DA)
  - Trained on 100–500 images
  - Smooth interpolation ⟹ no overfitting (Zhao et al., 2020).

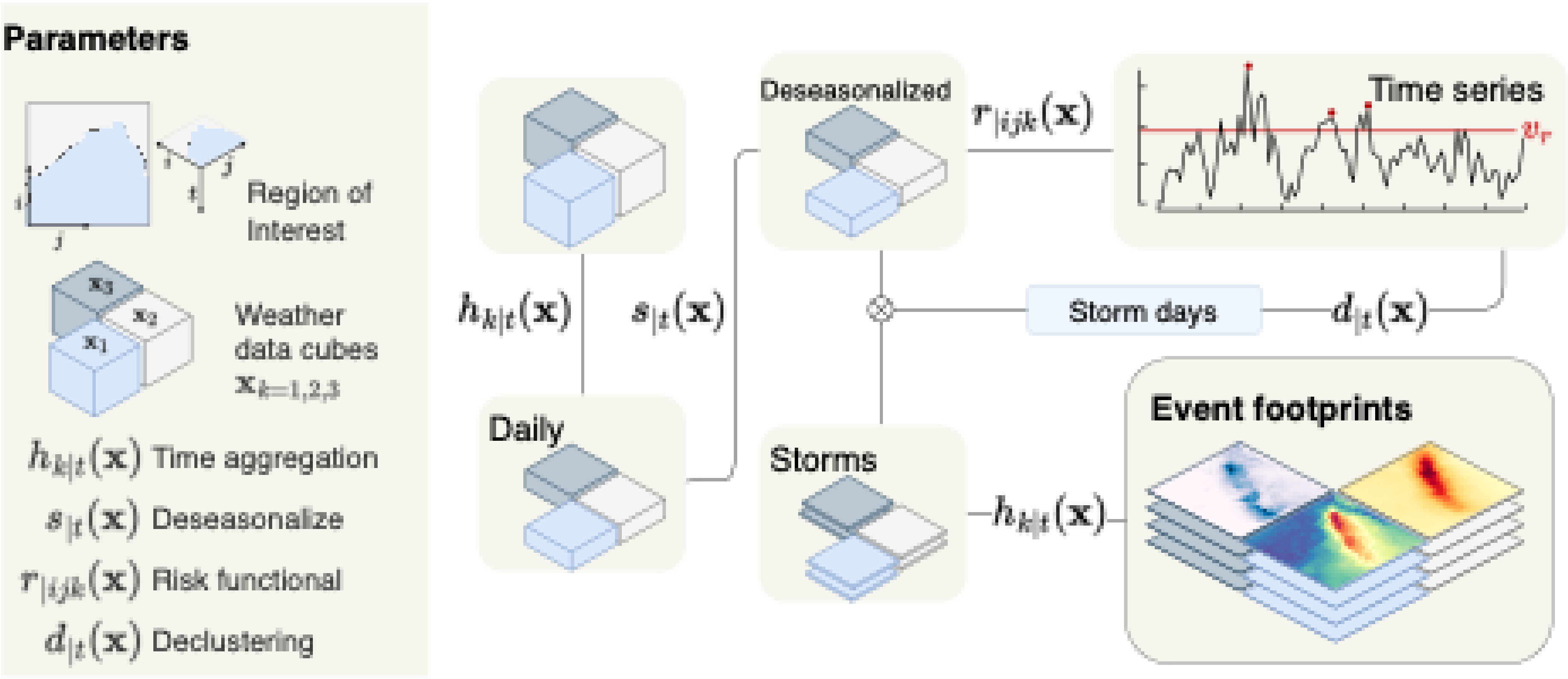
Pros	Cons
• data efficient • quick training	• old model with hardware requirements • difficult to modify architecture • flow or diffusion better

💡 Better approaches for sampling from extremes:
- progressive oversampling (collapse)
- pretraining (collapse)
- splitting algorithms

Zhao et al. (2020). *Differentiable Augmentation for Data-Efficient GAN Training.*

Recap

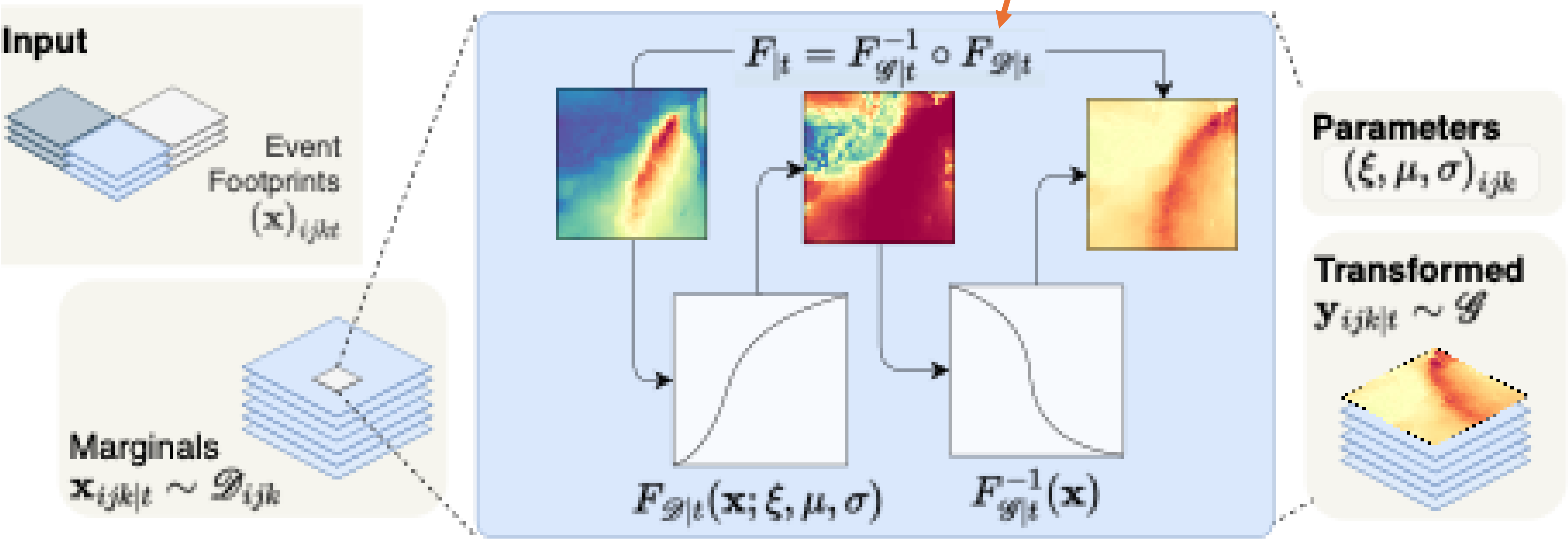
Event footprint creation



Marginal transforms

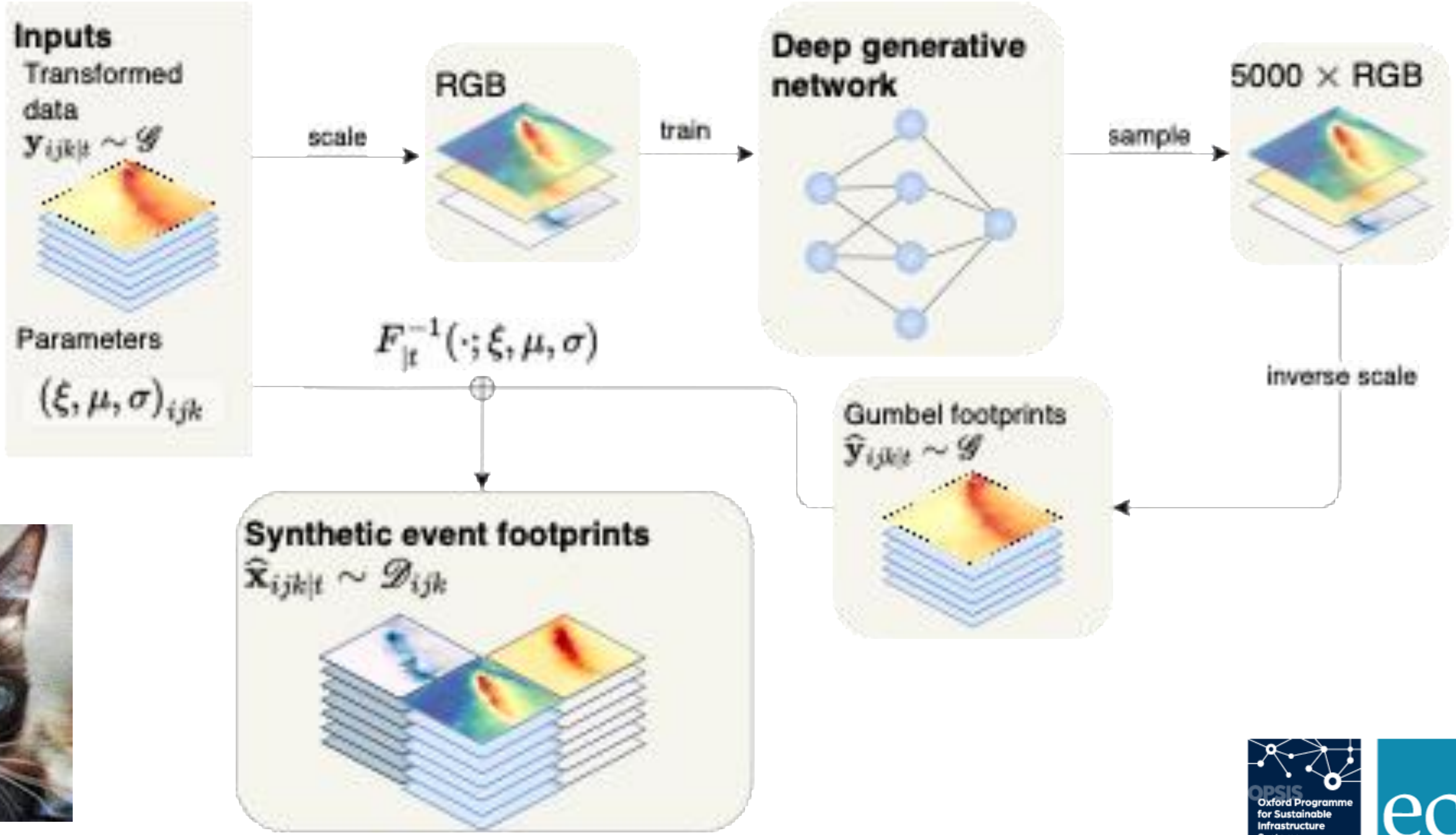
$$\tilde{F}(x) = \begin{cases} 1 - \left(1 - \hat{F}(v_X)\right)\left(1 - F_{\mathcal{D}}(x)\right) & \text{for } x > v_X \\ \hat{F}(x) & \text{for } x \leq v_X \end{cases}$$

Semiparametric GPD (Heffernan and Tawn, 2004)

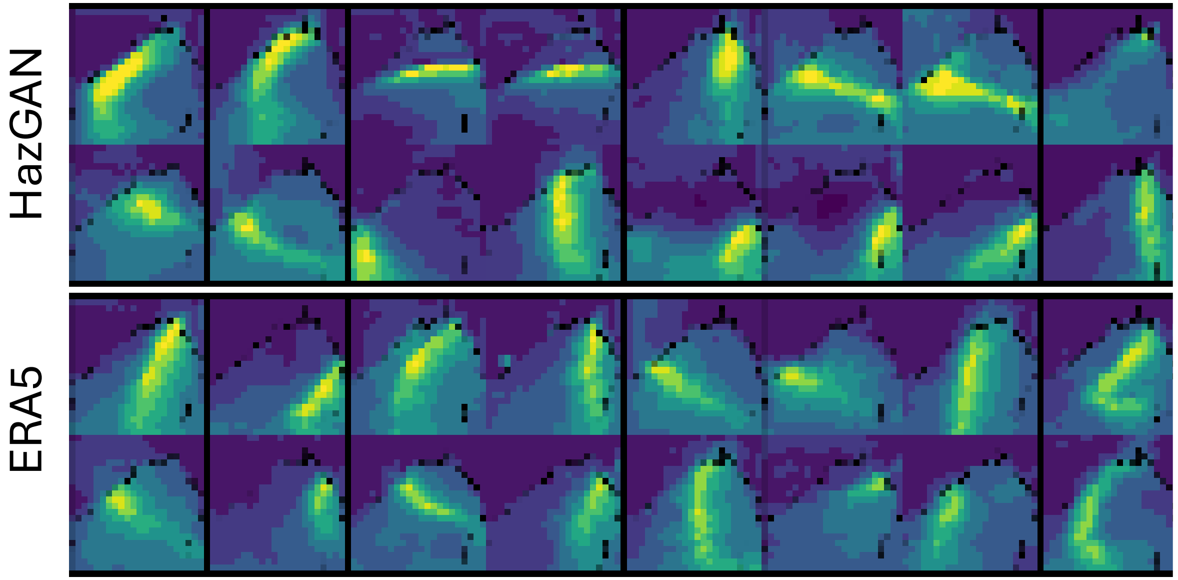


Light-tailed Gumbel or Gaussian

Training and sampling

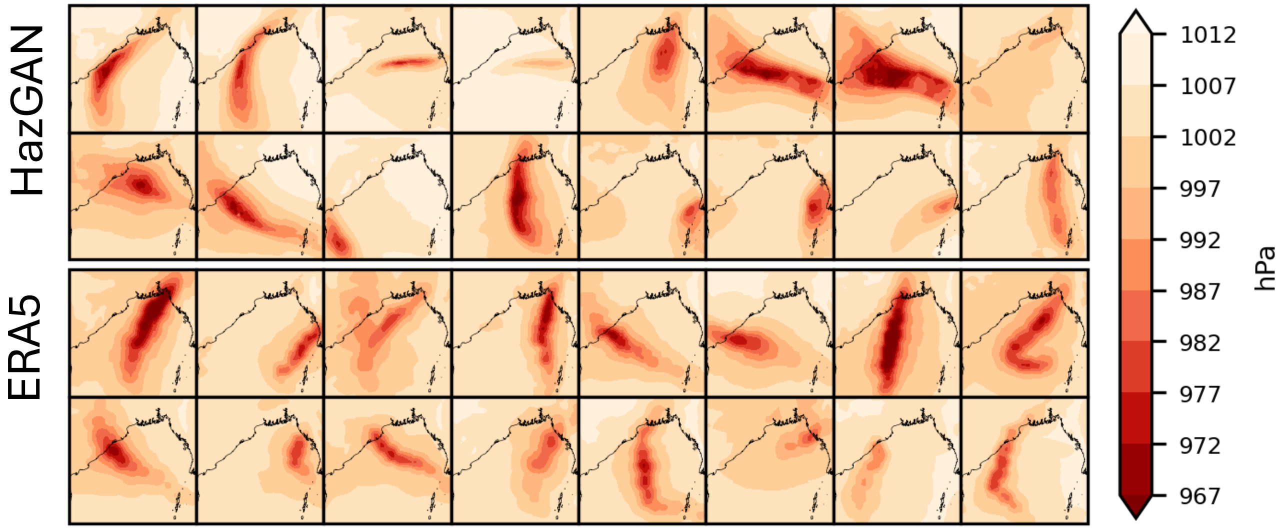


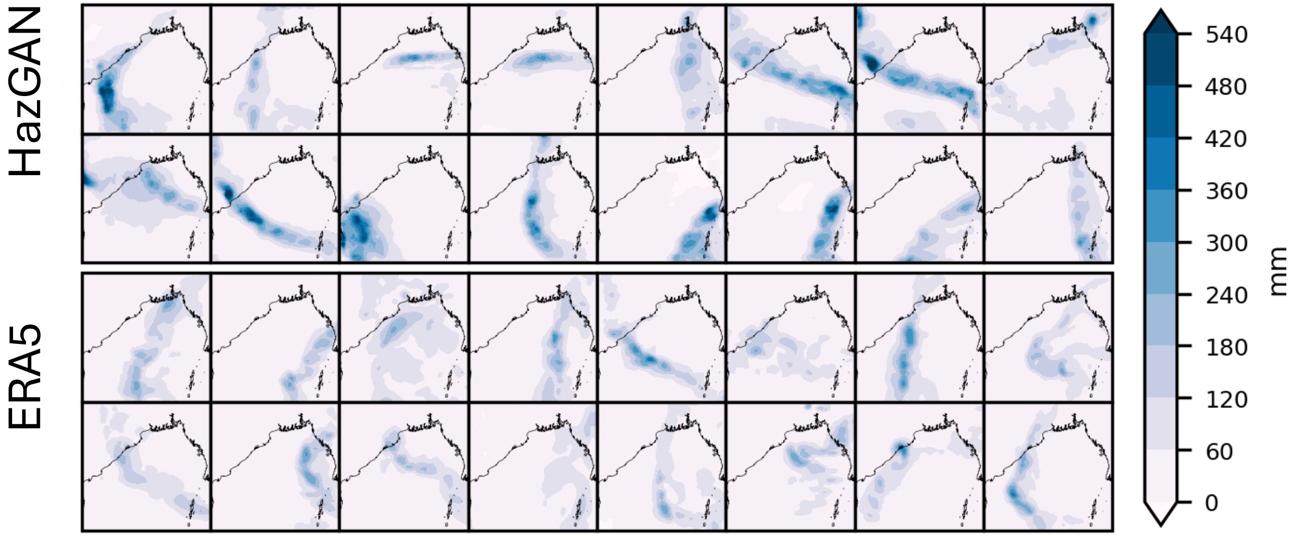
Results



Maximum wind speed

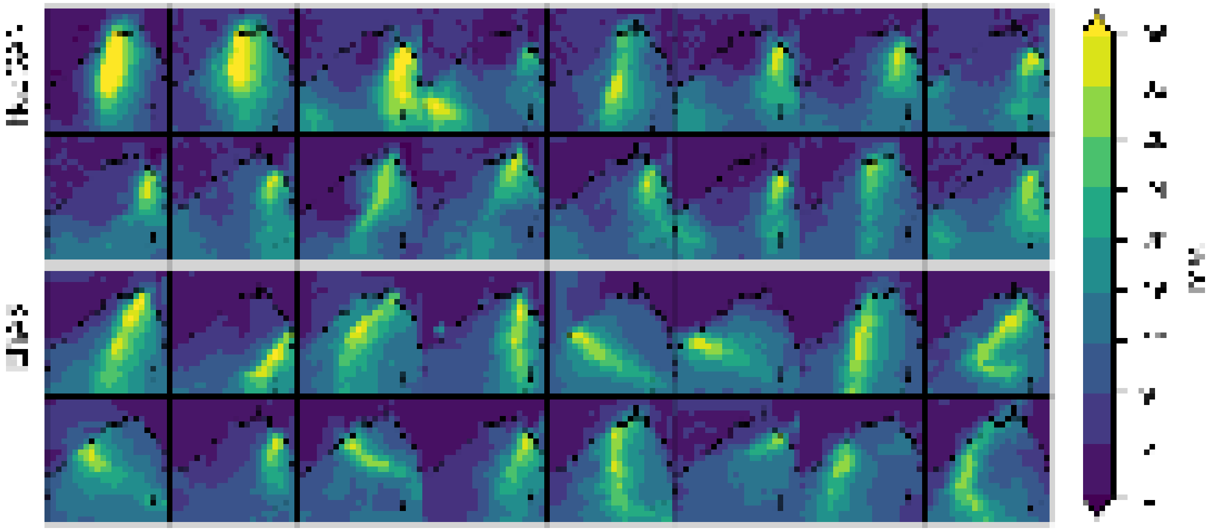
Minimum sea level pressure



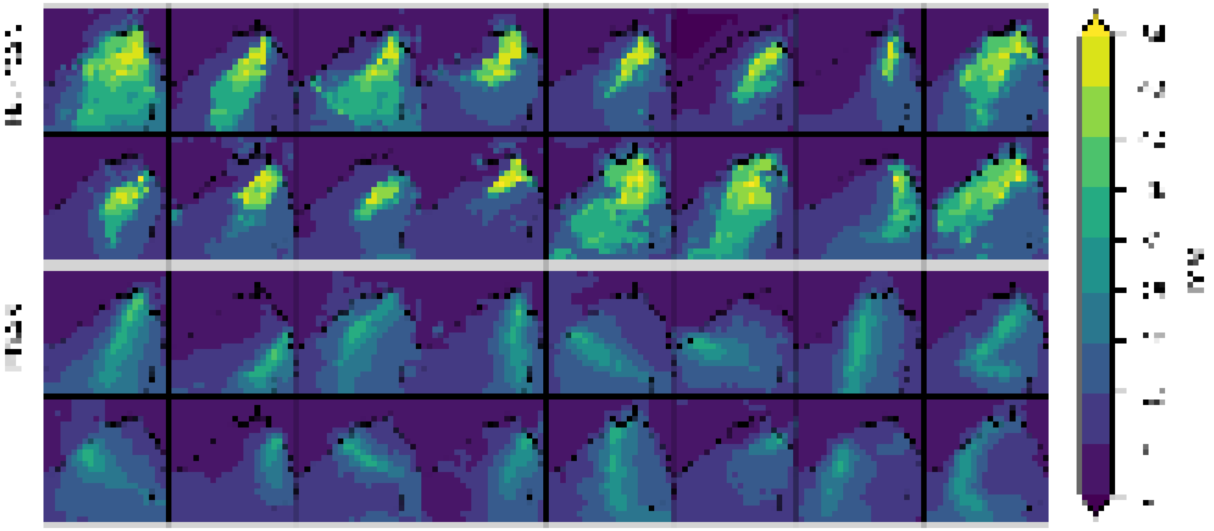


Total precipitation

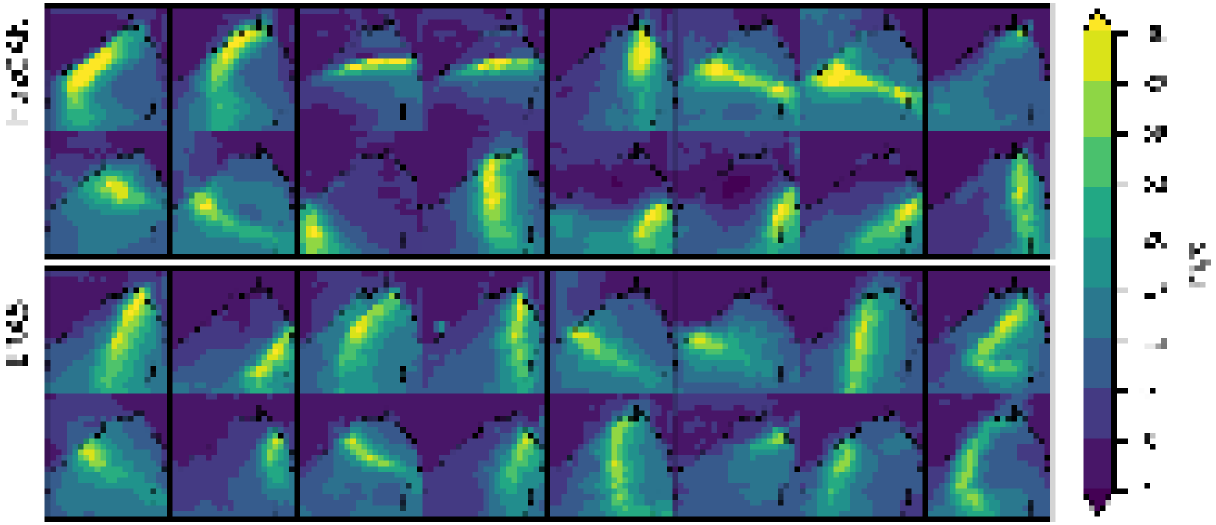
16 most extreme storms



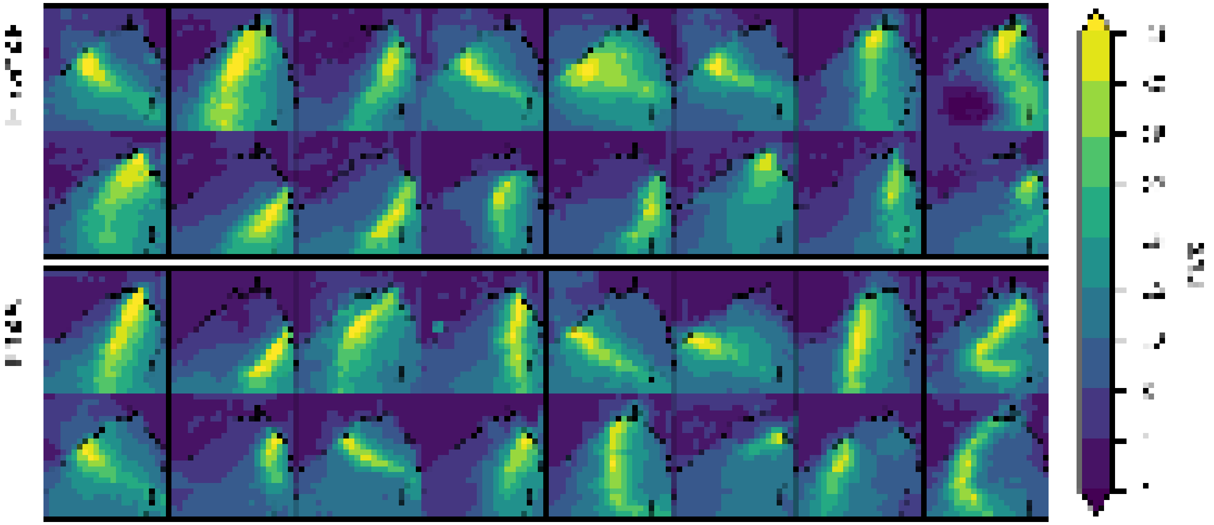
$\mathrm{GAN}(x_1, \ldots, x_d)$ (a) Rescaling only



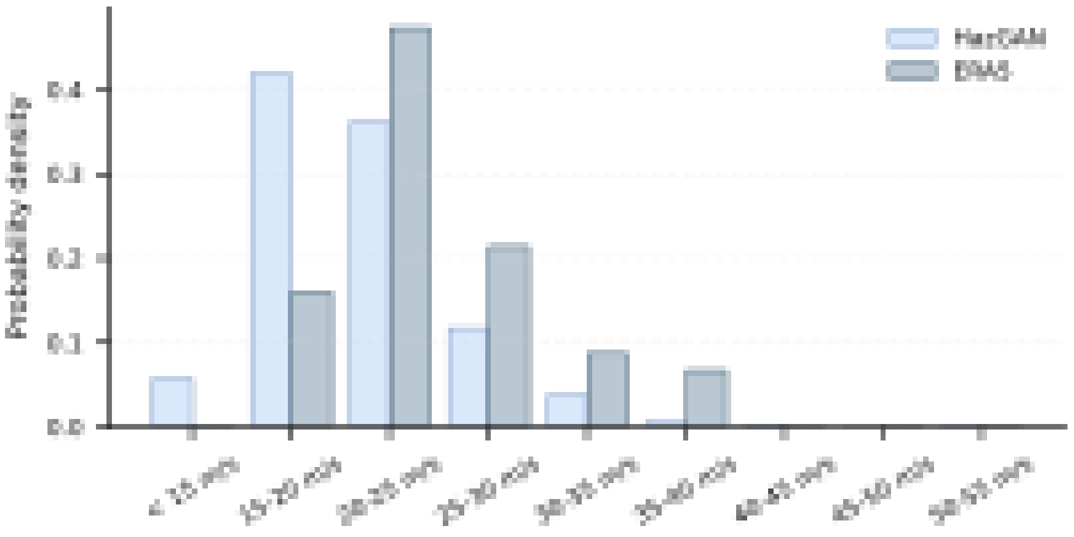
(b) Uniform $\mathrm{GAN}(u_1, \ldots, u_d)$



$\mathrm{GAN}(y_1, \ldots, y_d)$ (c) Gaussian

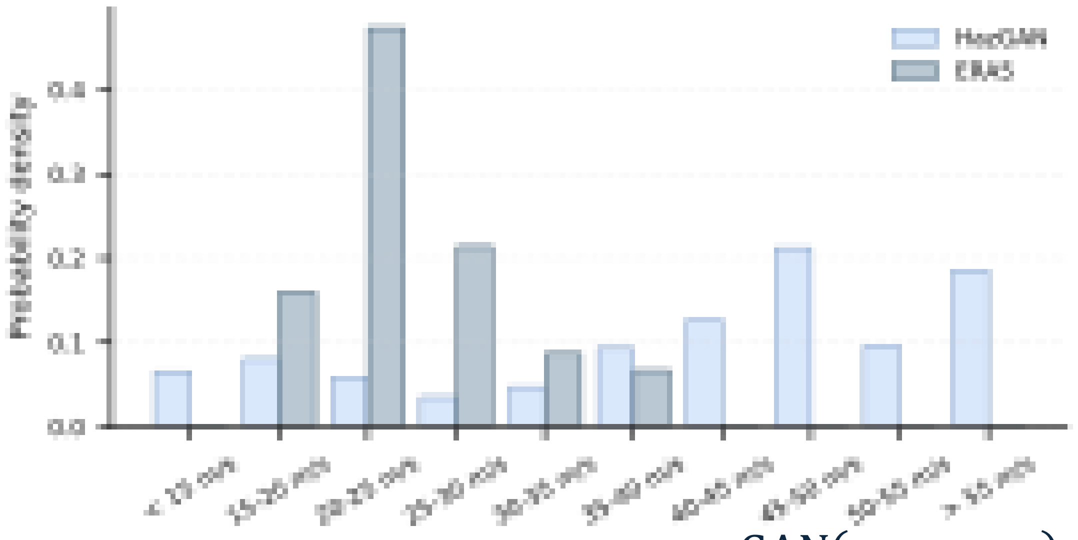


(d) Gumbel $\mathrm{GAN}(y_1, \ldots, y_d)$



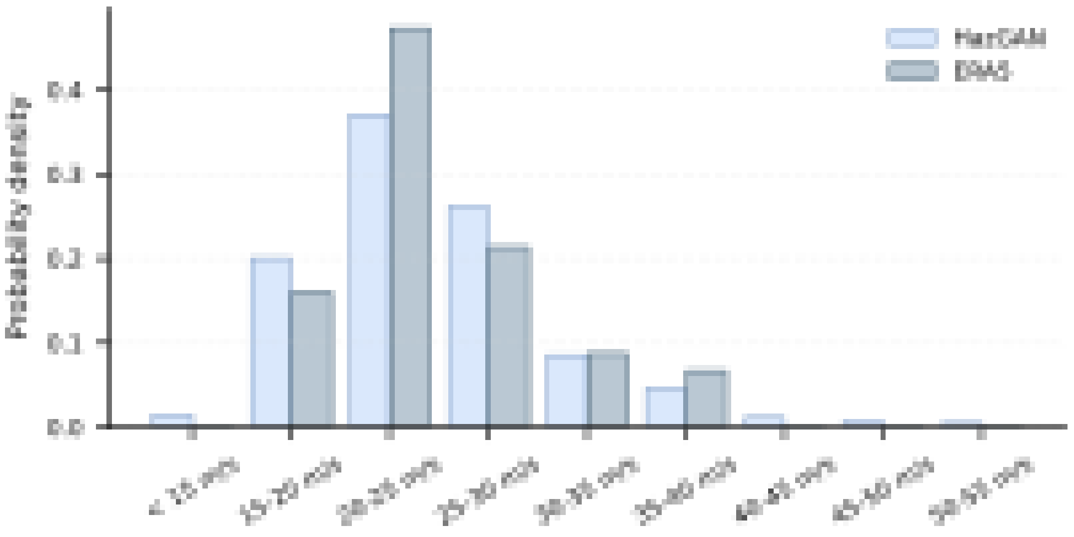
$\mathrm{GAN}(x_1, ..., x_d)$

(a) Rescaling only



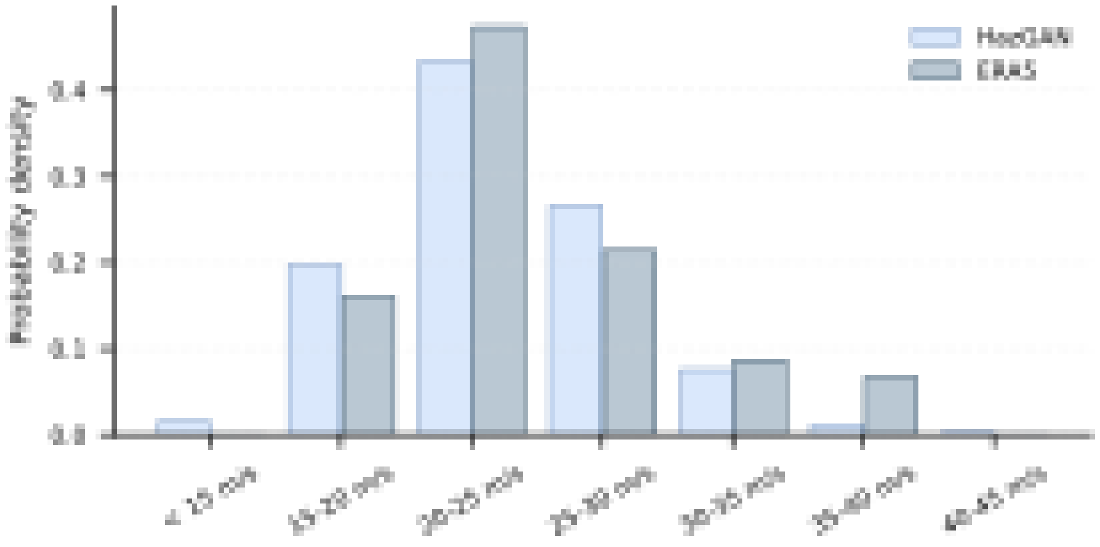
(b) Uniform

$\mathrm{GAN}(u_1, ..., u_d)$



$\mathrm{GAN}(y_1, ..., y_d)$

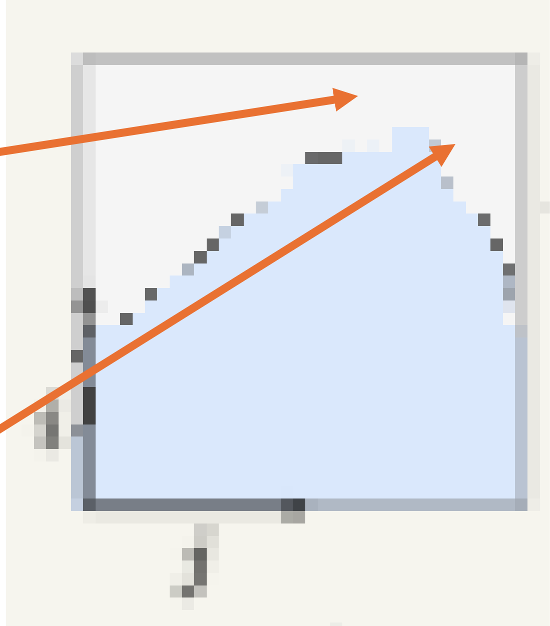
(c) Gaussian

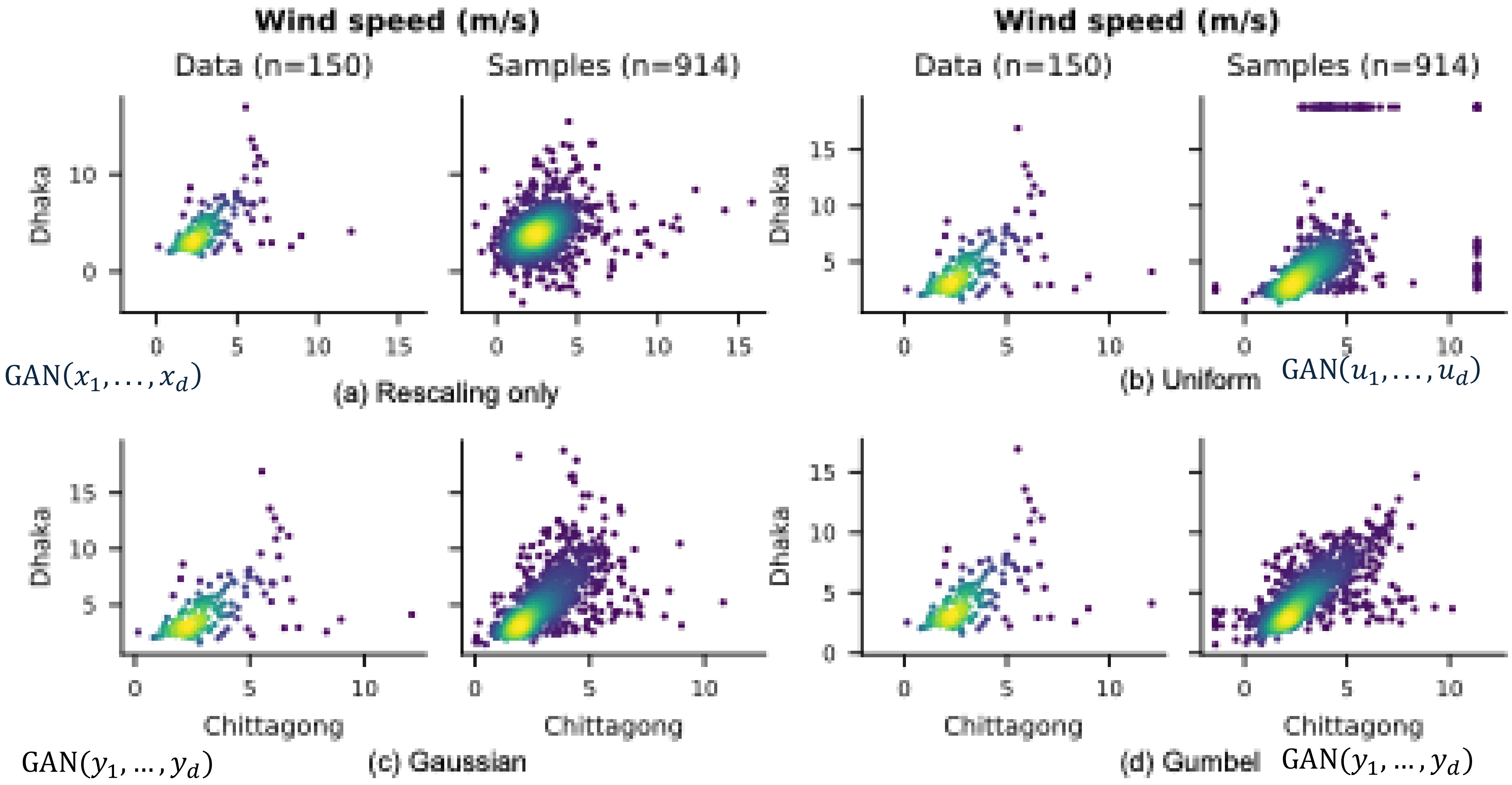


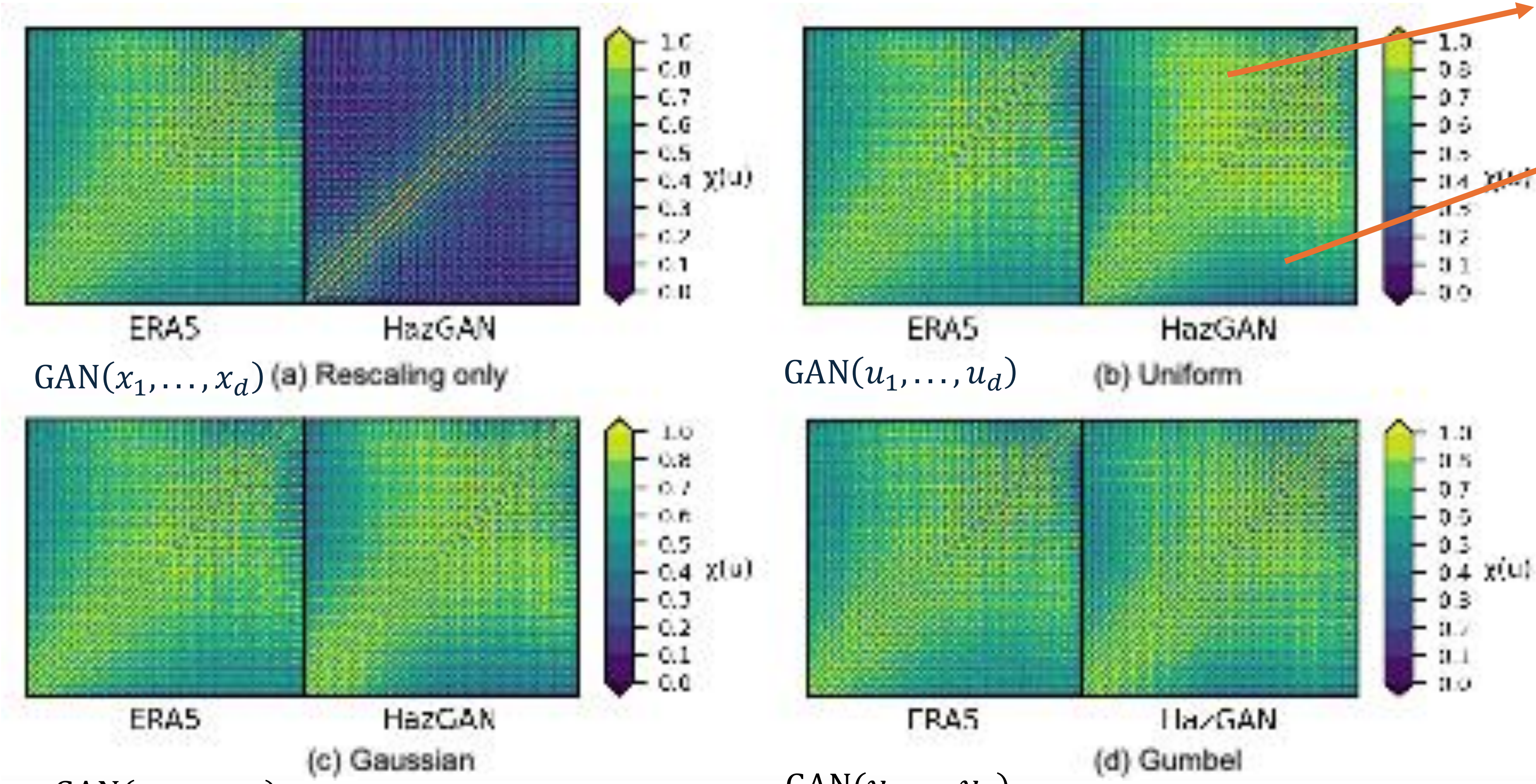
(d) Gumbel

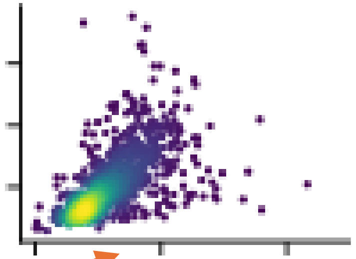
$\mathrm{GAN}(y_1, ..., y_d)$

BANGLADESH
LOW / HILLS / MOUNTAINS
Nepal
India
Dinajpur
Rangpur
Brahmaputra River
Ganges River
Jamuna
Jamalpur
Sylhat
Mymensingh
Rajshahi
Meghna
Pabna
Dhaka
Brahmanbaria
Madhumati
Narayanganj
N
Bangladesh
Comilla
Tropic of Cancer
Jessore
Padma
Kaptai Lake
India
Khulna
Barisal
Chittagong
Mongla
Mouths of the Ganges
60 mi
60 km
Bay of Bengal
Cox's Bazar
Myanmar (Burma)
Indian Ocean
Copyright©blankworldmap.net




Wind speed (m/s)
Data (n=150)
Samples (n=914)
Wind speed (m/s)
Data (n=150)
Samples (n=914)
Dhaka
Chittagong
GAN$(x_1, \ldots, x_d)$
(a) Rescaling only
(b) Uniform
GAN$(u_1, \ldots, u_d)$
GAN$(y_1, \ldots, y_d)$
(c) Gaussian
(d) Gumbel
GAN$(y_1, \ldots, y_d)$





$\mathrm{GAN}(x_1, \ldots, x_d)$

$\mathrm{GAN}(u_1, \ldots, u_d)$

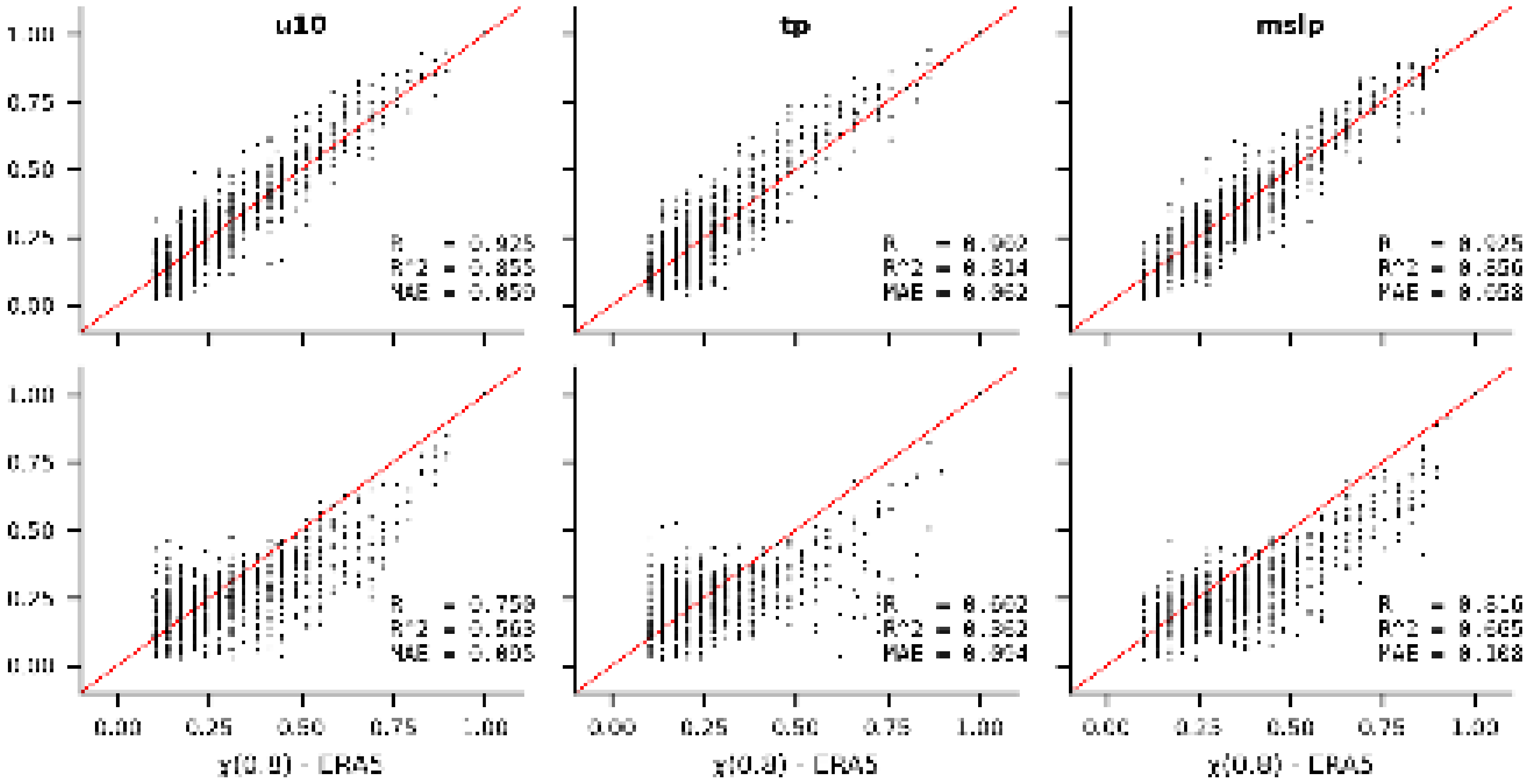
$\mathrm{GAN}(y_1, \ldots, y_d)$

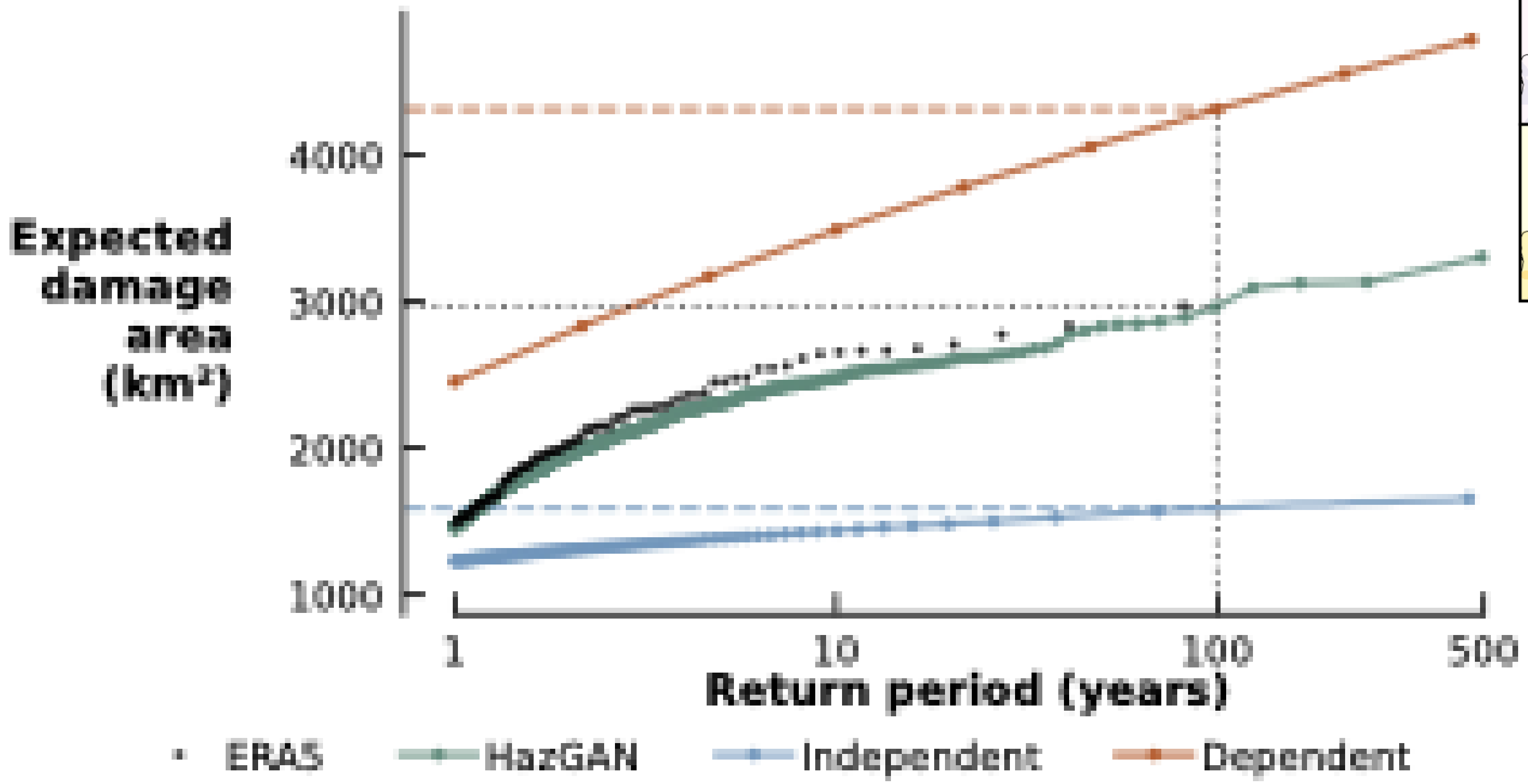
$\mathrm{GAN}(y_1, \ldots, y_d)$

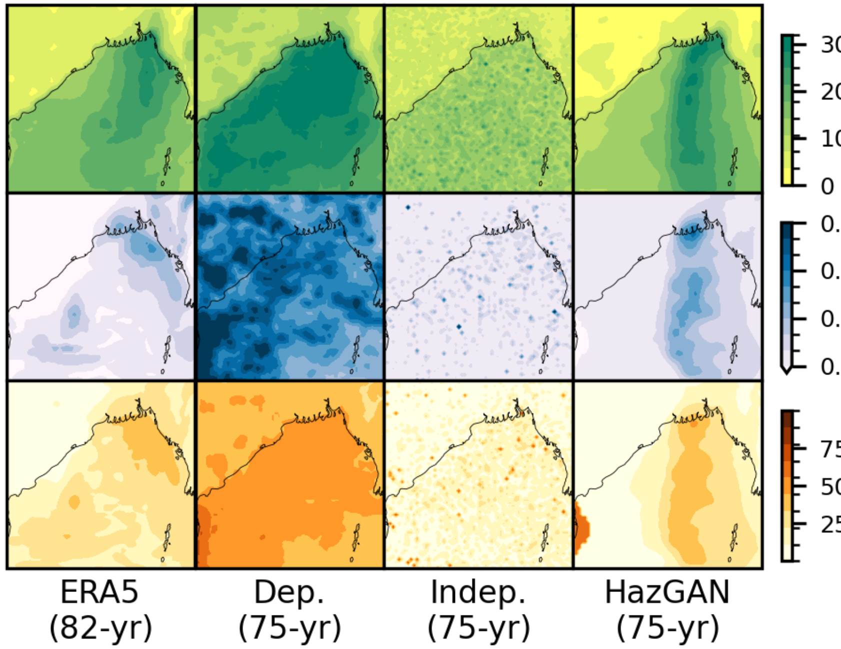
$\mathrm{GAN}(y_1, \ldots, y_d)$

HazGAN

Heffernan & Tawn (2004)

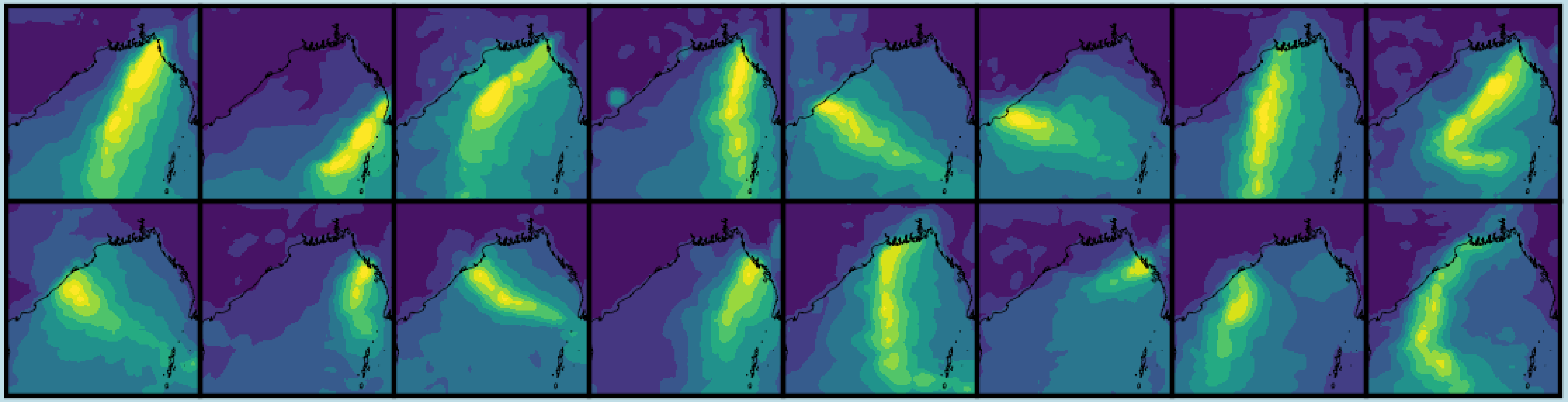



Expected damage area (km²)
4000
3000
2000
1000
1
10
100
500
Return period (years)
ERA5
HazGAN
Independent
Dependent


ERA5 (82-yr)
Dep. (75-yr)
Indep. (75-yr)
HazGAN (75-yr)

Thanks for listening

- POT approach for event selection
- Semiparametric GPD to handle localized events
- Train on data transformed to light-tailed margins
- StyleGAN2-DA for small datasets
- Good performance (so far!)

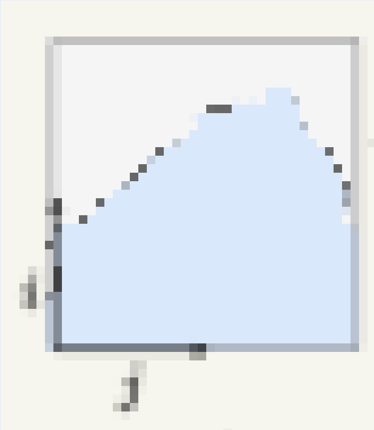


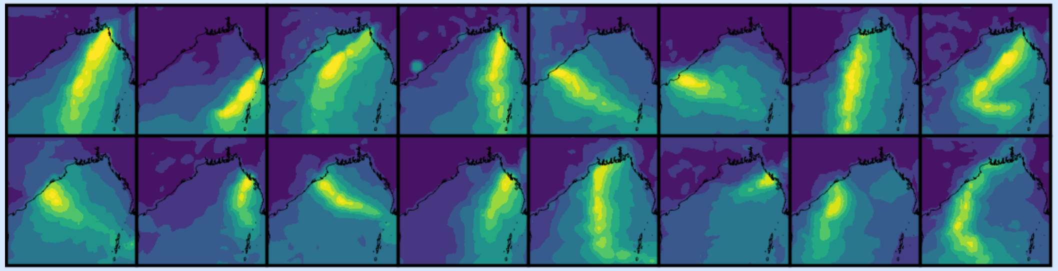
Objective: Multi-hazard event set generator

Stochastic *hazard* generator – event footprints

Essential dimensions:

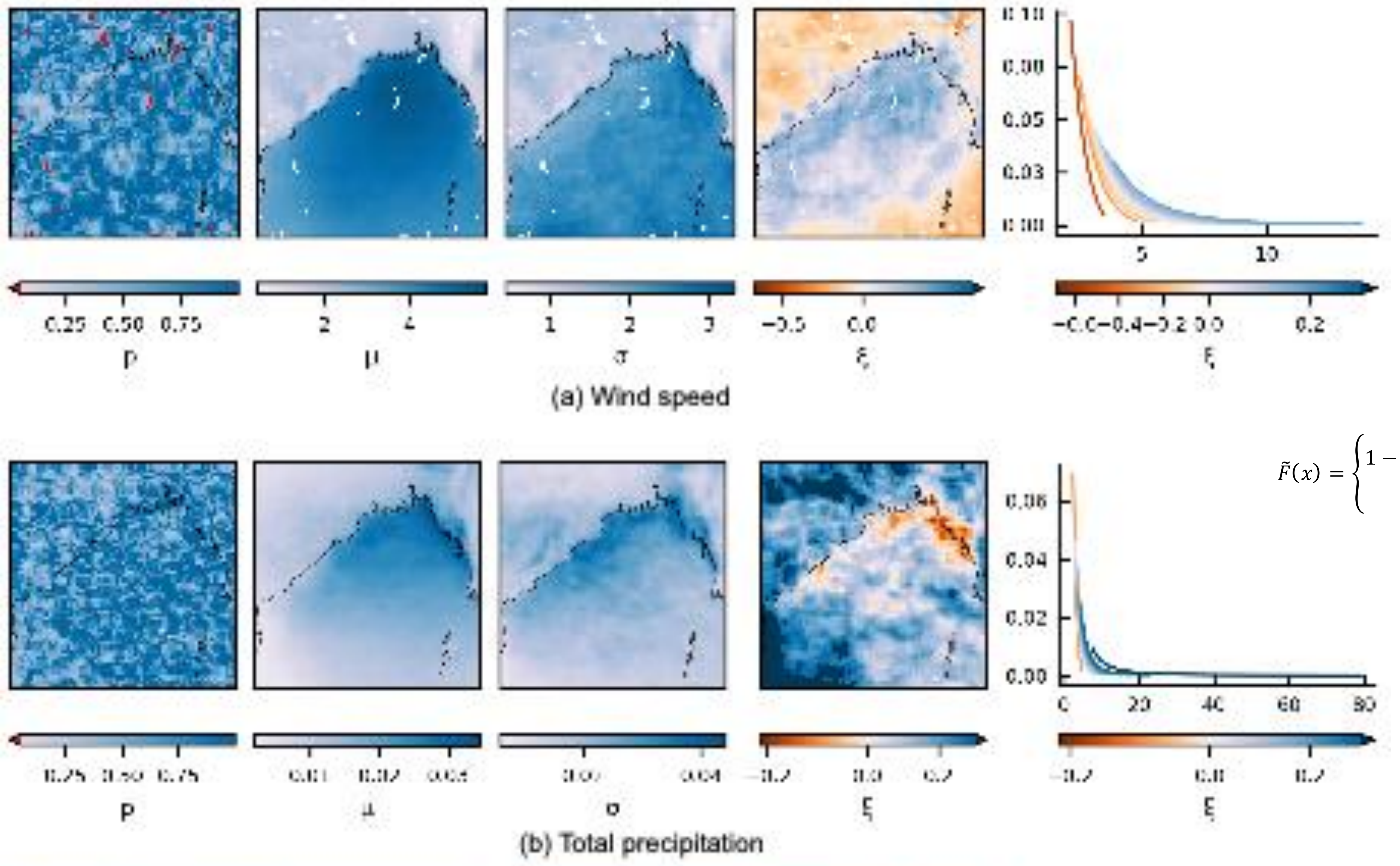
- ✅ spatial
- ✅ multivariate
- ✅ probability
- ❌ temporal



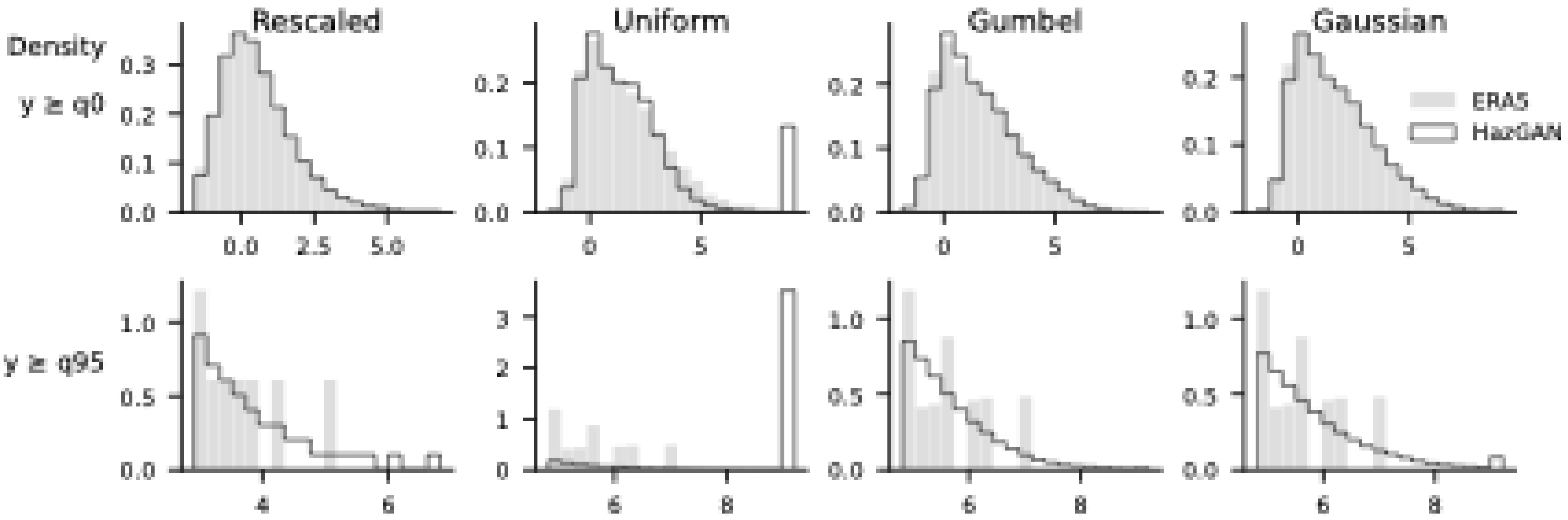


Variable	Value
Dataset	ERA5
Temporal domain	1940–2022
Spatial domain	10–25° N × 80–95° E
Spatial resolution	0.25°
Hazard variables	1. Max wind speed at 10 m (ws10) 2. Total precipitation (tp) 3. Atmospheric sea-level pressure (mslp)
Total variables	64 × 64 × 3 = 12 288

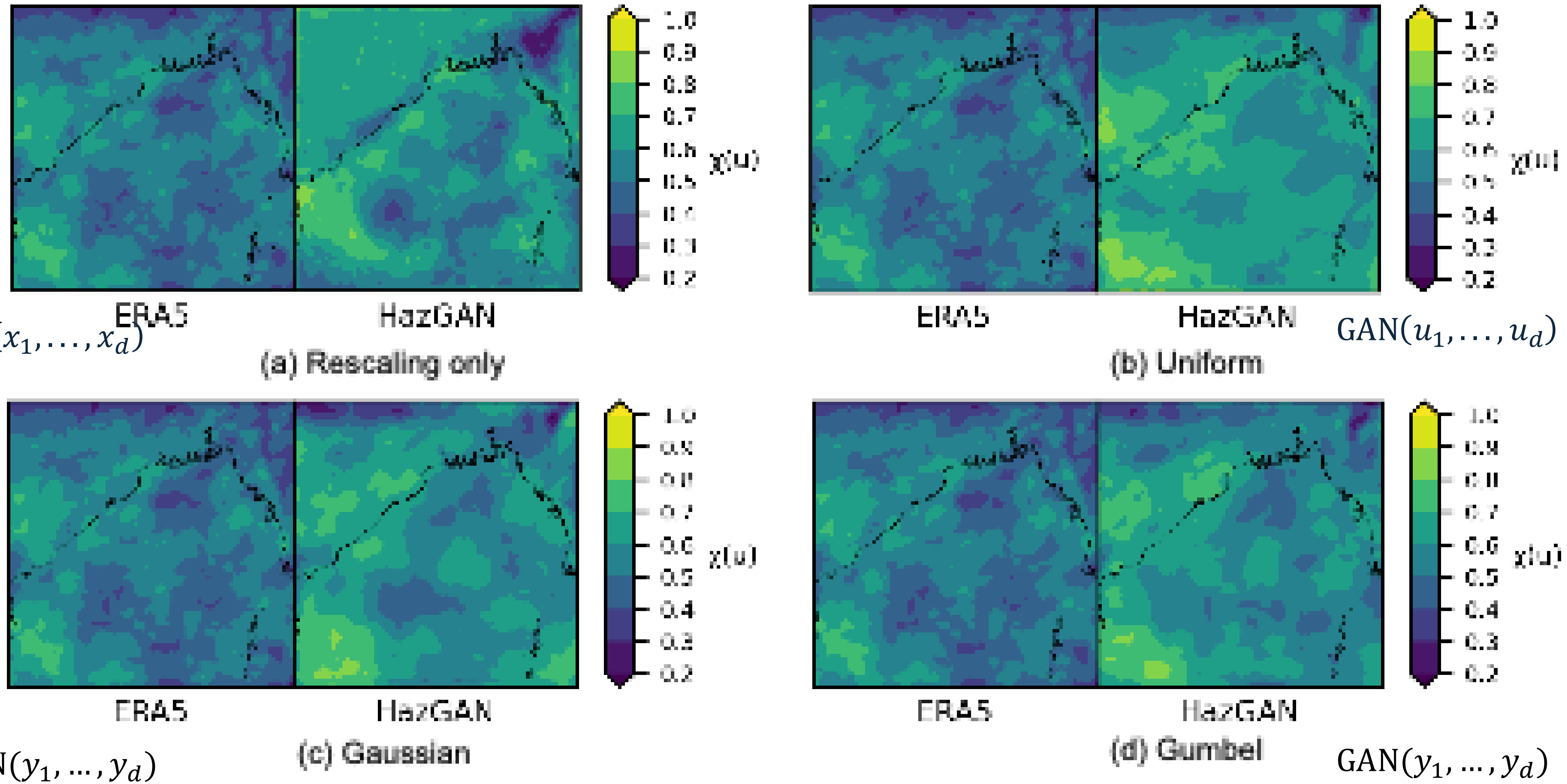
Wilks, D. S., & Wilby, R. L. (1999). *The weather generation game: A review of stochastic weather models.*



$$\tilde{F}(x) = \begin{cases} 1-\left(1-\hat{F}(v_X)\right)\left(1-F_D(x)\right) & \text{for } x > v_X \\ \hat{F}(x) & \text{for } x \leq v_X \end{cases}$$



Rescaled
Uniform
Gumbel
Gaussian
Density
y ≥ q0
y ≥ q95
ERA5
HazGAN
0.0
0.1
0.2
0.3
2.5
5.0
0
5
1.0
0.5
1
2
3
4
6
8



$\mathrm{GAN}(x_1, \ldots, x_d)$

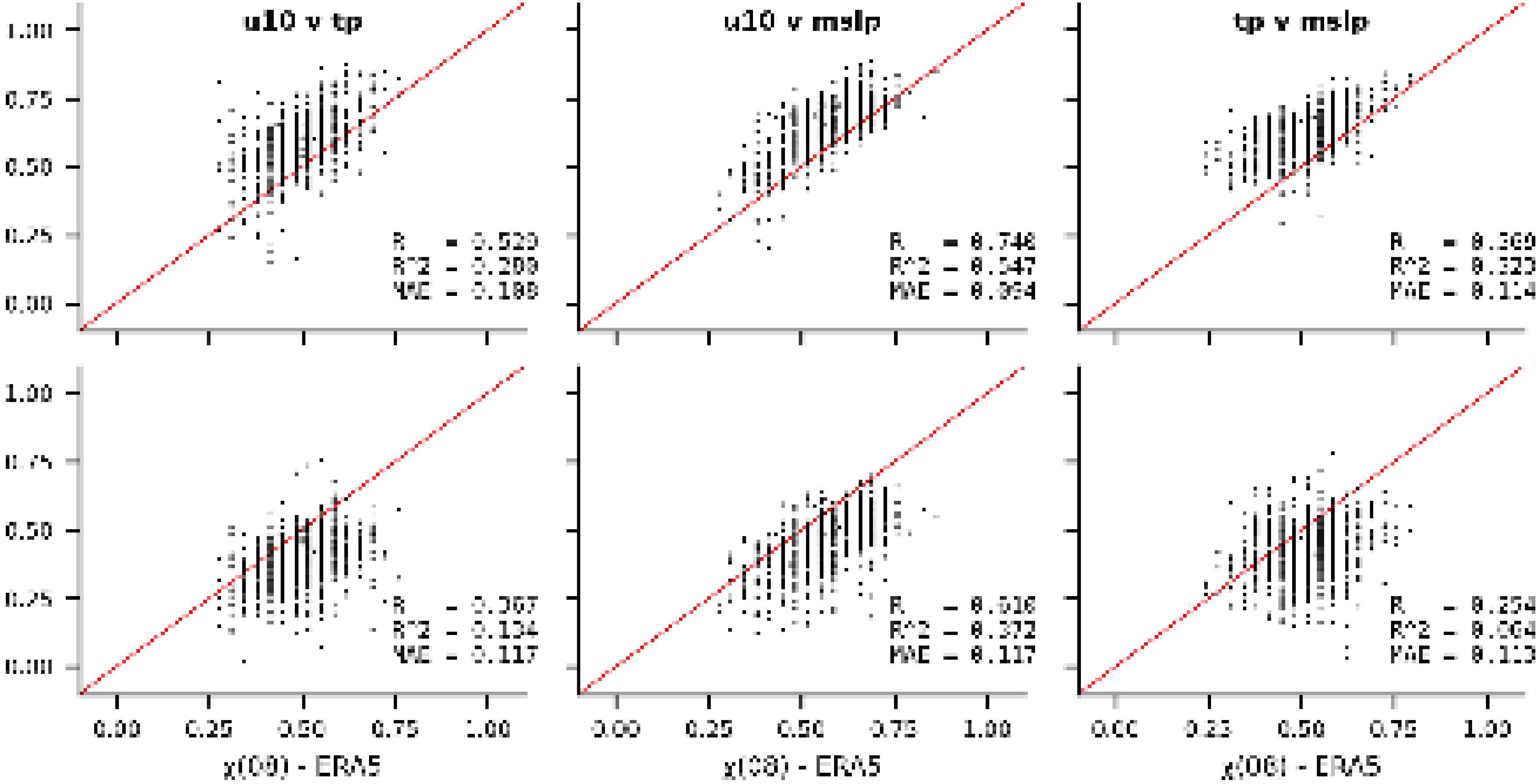
$\mathrm{GAN}(u_1, \ldots, u_d)$

$\mathrm{GAN}(y_1, \ldots, y_d)$

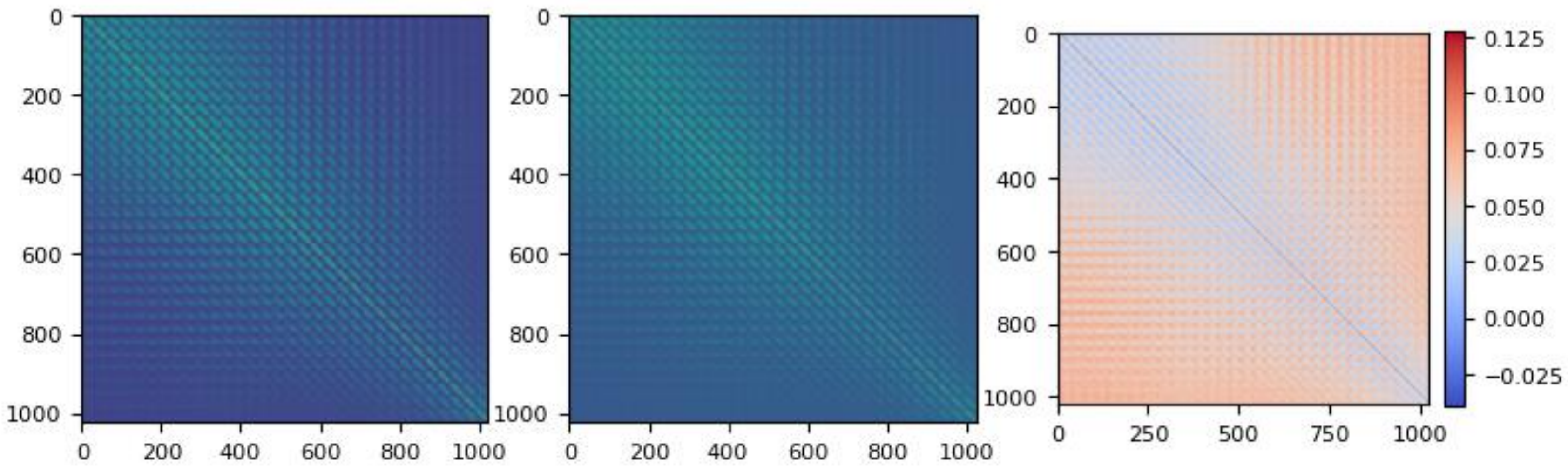
$\mathrm{GAN}(y_1, \ldots, y_d)$

HazGAN

Heffernan & Tawn (2004)



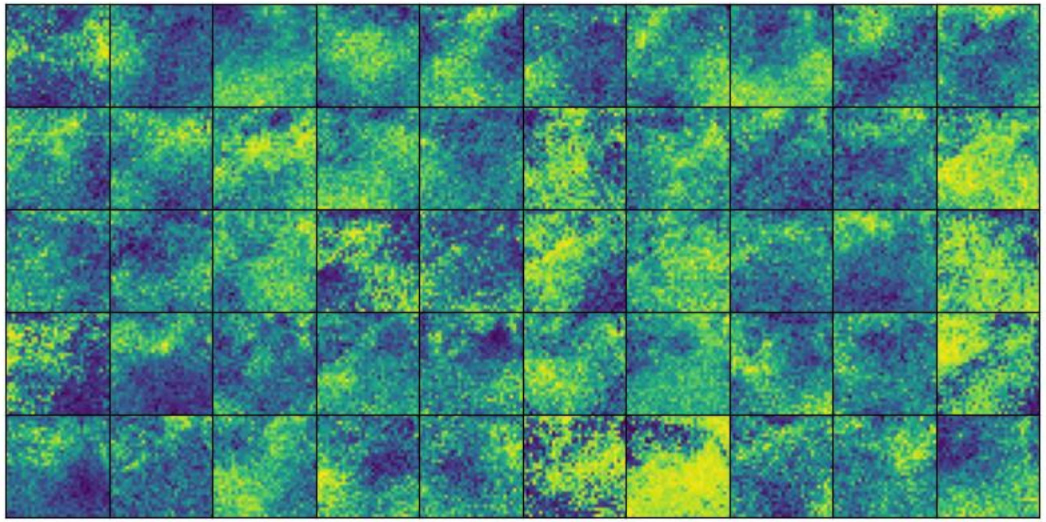
(b) Multivariate

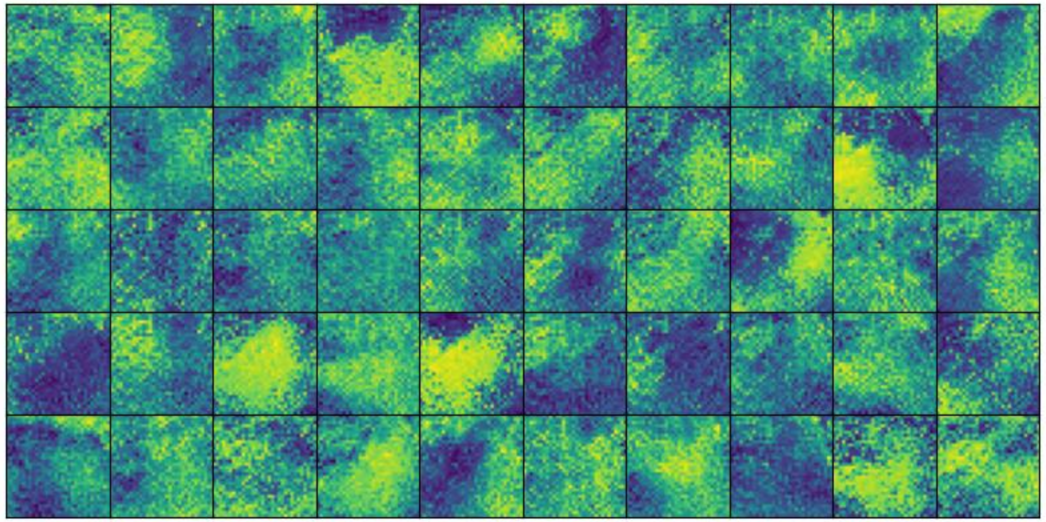


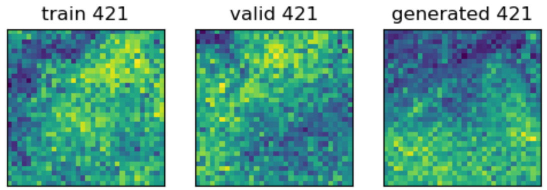
The (elliptically contoured) Gaussian scale mixtures of Huser et al. (2017), if we have a spatial process $X(s) = RW(s)$, where $W(s)$ is a standard Gaussian process with correlation function $\rho(s_i, s_j)$, and R has Pareto tail decay, $F_R(r) = 1 - Kr^{-\gamma}$, then the following relationship holds:

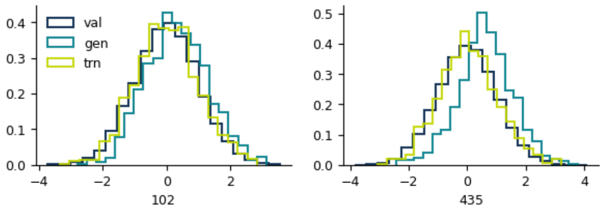
$$\chi(s_i, s_j) = 2 - 2T_{\gamma+1}\left(\sqrt{1+\gamma} \cdot \frac{1-\rho(s_i, s_j)}{\sqrt{1-\rho(s_i, s_j)^2}}\right)$$

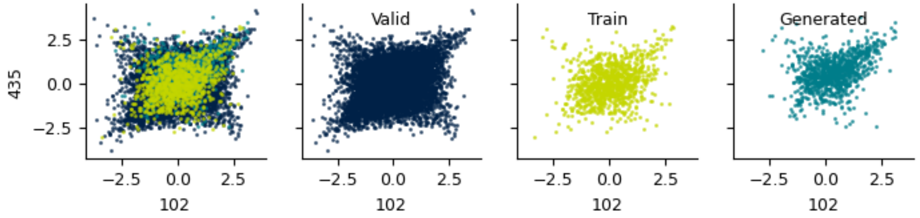
and $\eta(s_i, s_j) = 1$, where T_v denotes a student t distribution with v degrees of freedom.

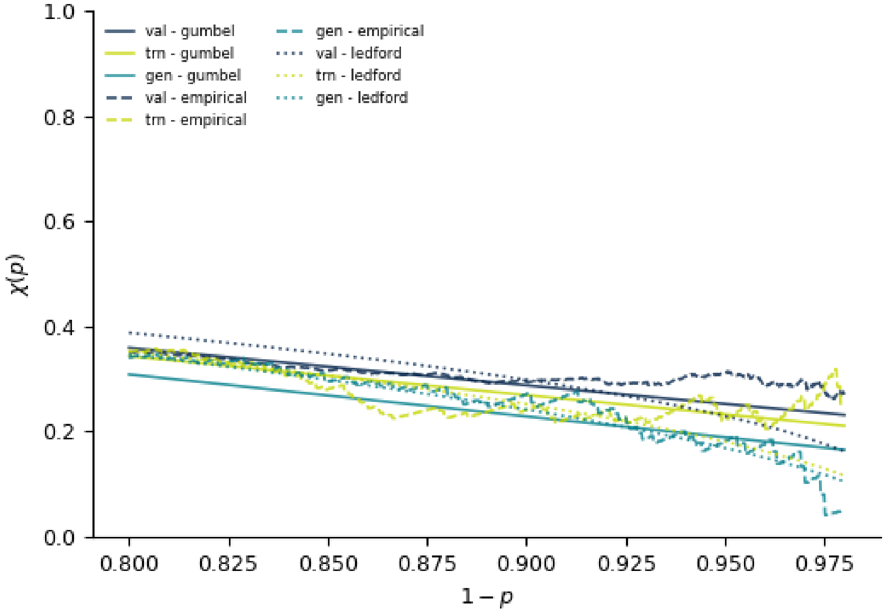


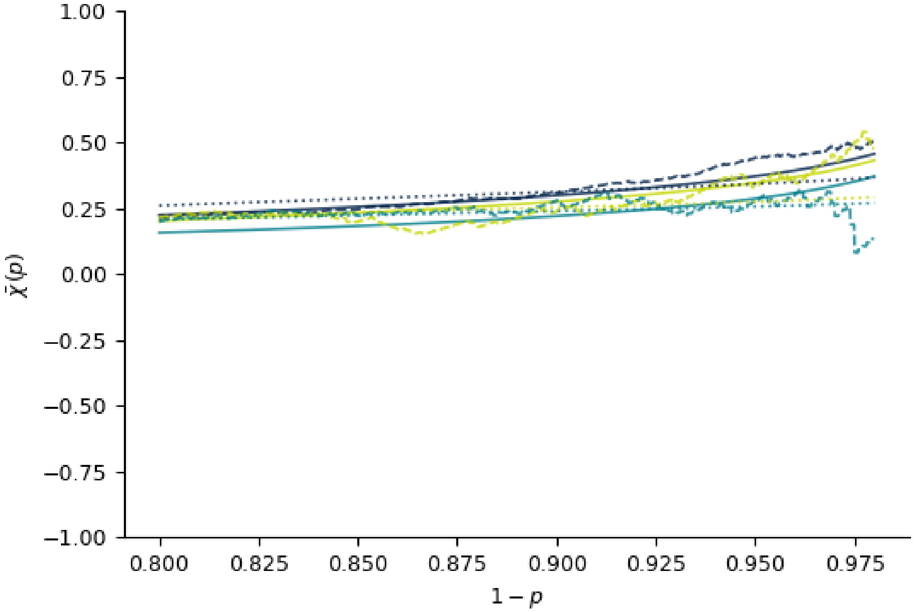




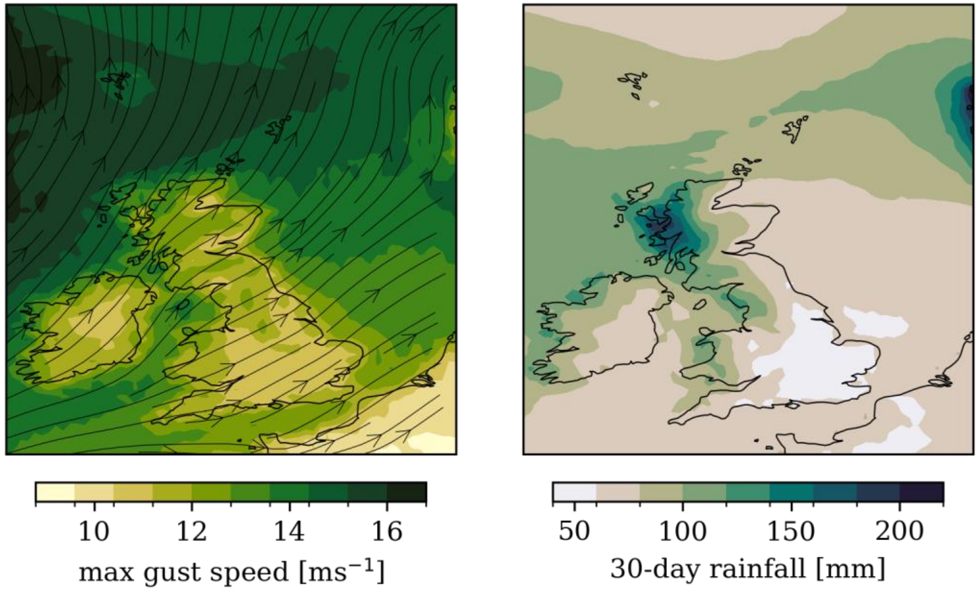


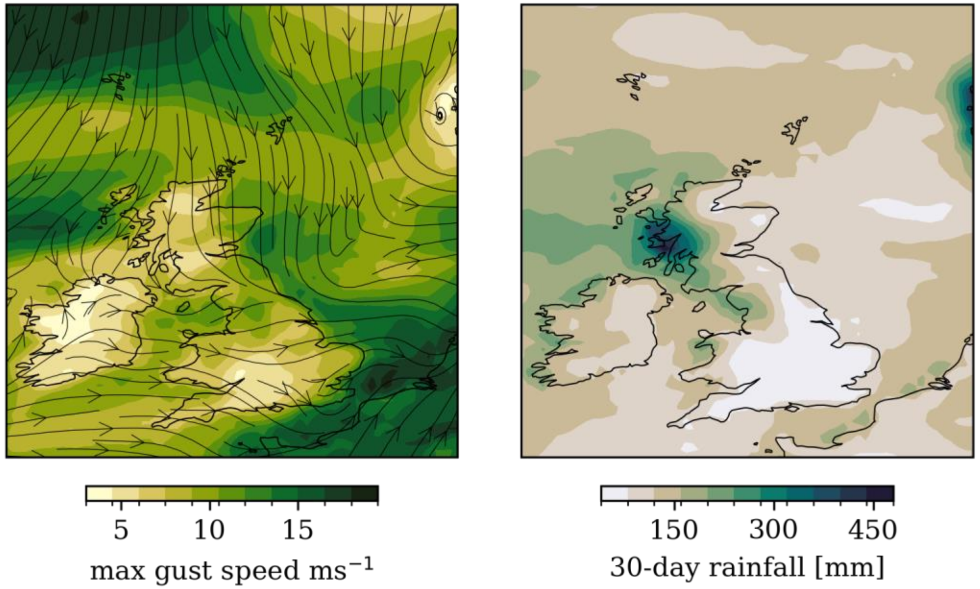


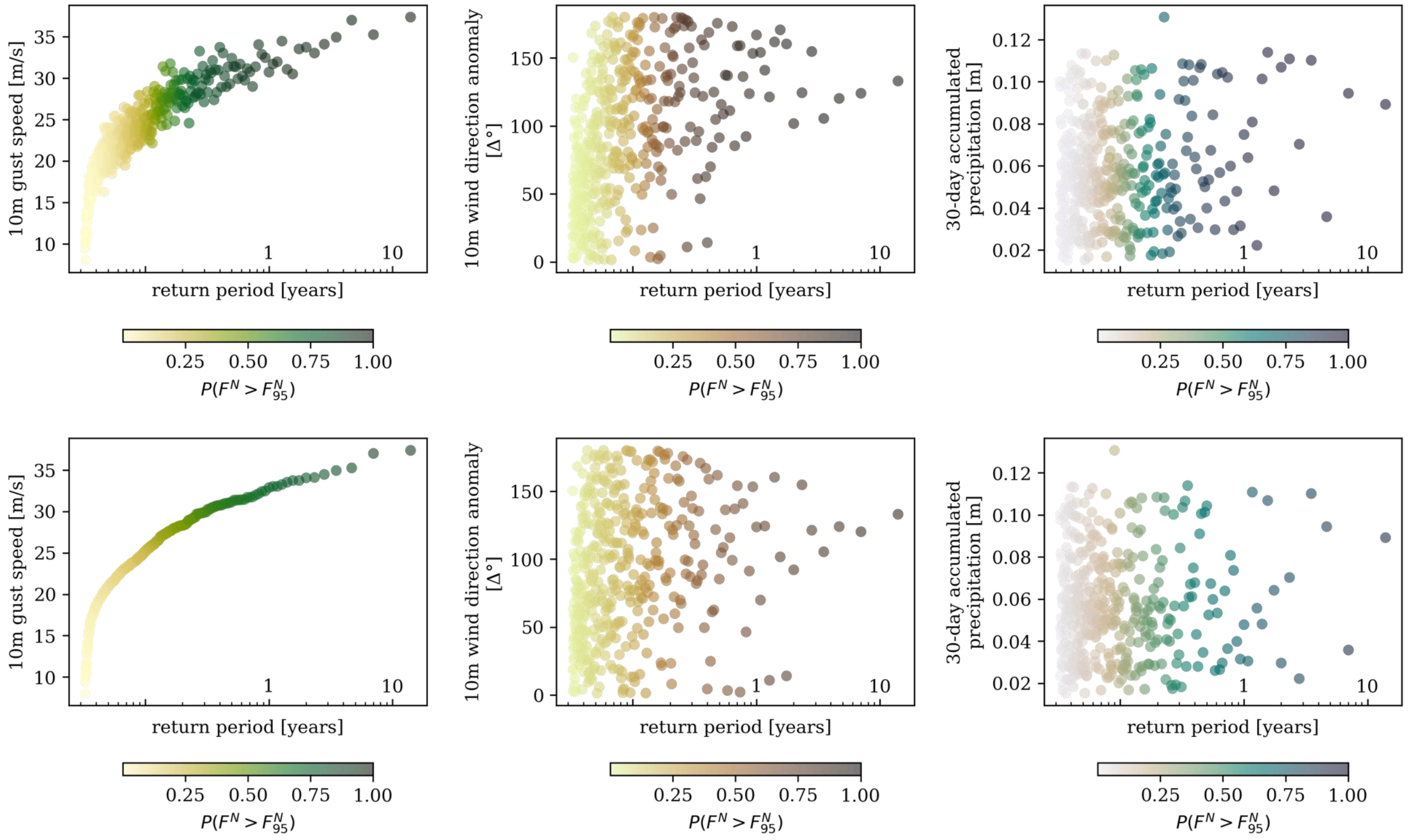


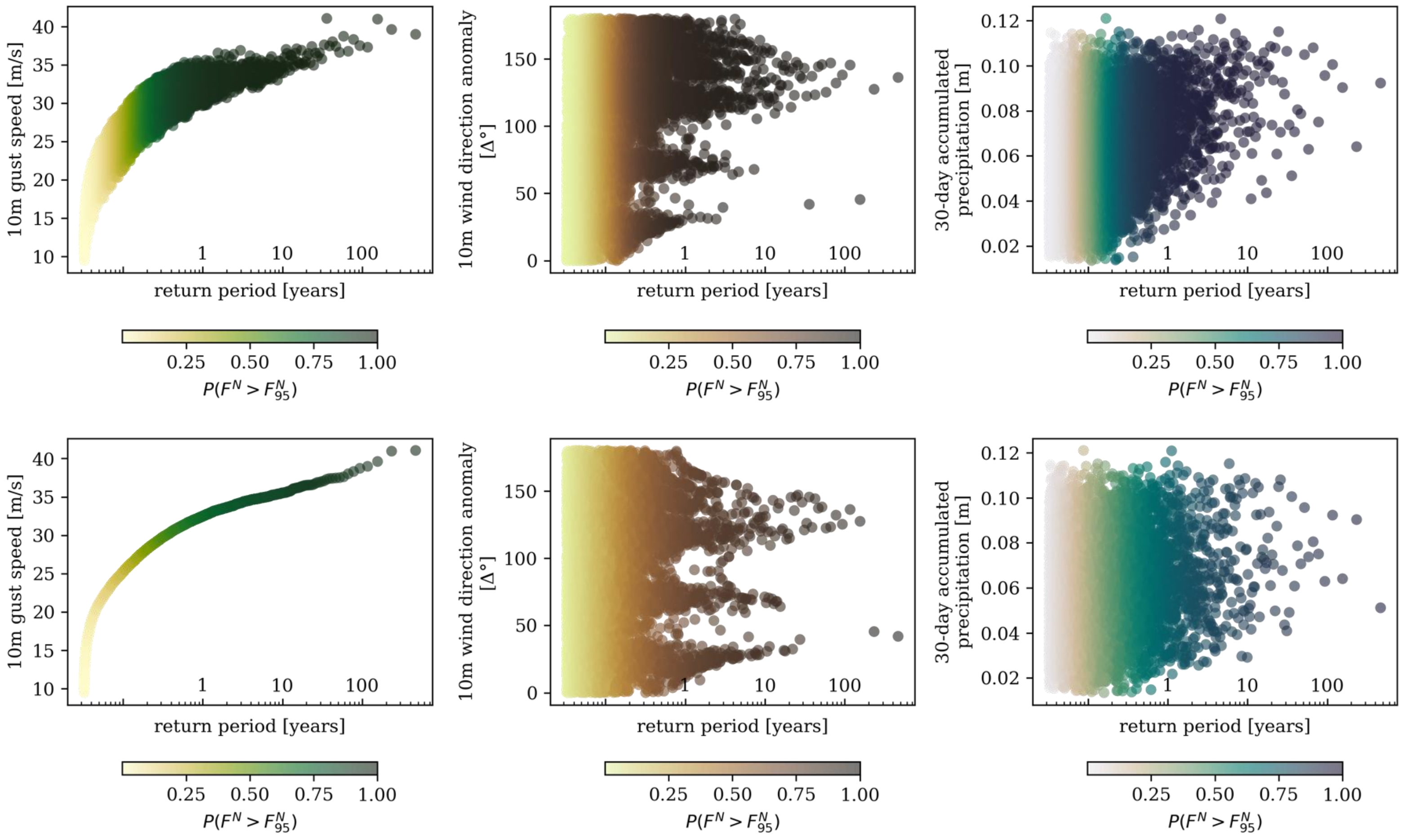


(417, 843) [seed:38]


10 12 14 16
max gust speed [ms^{-1}]
50 100 150 200
30-day rainfall [mm]


5 10 15
max gust speed ms^{-1}
150 300 450
30-day rainfall [mm]


10m gust speed [m/s]
10m wind direction anomaly [Δ°]
30-day accumulated precipitation [m]
return period [years]
$P(F^N > F^N_{95})$
0.25 0.50 0.75 1.00



10m gust speed [m/s]
return period [years]
10m wind direction anomaly [Δ°]
30-day accumulated precipitation [m]
$P(F^N > F^N_{95})$