TOWARD A DATA-DRIVEN APPROACH FOR LARGE-SCALE STORM SURGE PREDICTION FROM TROPICAL CYCLONES

Itxaso Odériz[1]
MSCA individual fellow
itxaso.oderiz@unican.es

Iñigo J. Losada[1]

Sanne Muis[2,3]

[1]IHCantabria - Instituto de Hidráulica Ambiental de la Universidad de Cantabria, Santander, Spain
[2]Environmental Hydrodynamics and Forecasting, Deltares, Delft, Netherlands
[3]Institute for Environmental Studies, Vrije Universiteit, Amsterdam, Netherlands

TRANSCLIMA
MARIE SKŁODOWSKA-CURIE ACTIONS
Funded by the European Union
Barcelona Supercomputing Center
Centro Nacional de Supercomputación
UC
FUNDACION INSTITUTO DE HIDRÁULICA AMBIENTAL DE CANTABRIA
IHcantabria INSTITUTO DE HIDRÁULICA AMBIENTAL UNIVERSIDAD DE CANTABRIA

TRANSCLIMA

0 MOTIVATION AND CONTEXT
Adapting to unprecedented tropical cyclones

In recent years, areas have been impacted by unprecedented tropical cyclones (TCs).

1) TC Kenneth struck northern Mozambique — a region previously unaffected by TCs — causing 45 casualties and $1 billion in damages.
2) The 2024 Atlantic hurricane season featured major hurricanes in the Gulf of Mexico, with Helene and Milton prompting U.S. coastal warnings due to rapid intensification.
3) In 2024, the Philippines was hit by six consecutive typhoons, three of which were classified as super-typhoons.

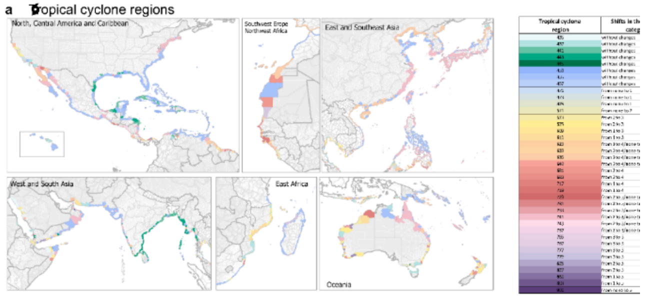
Adapting to unprecedented TCs is a major concern, presenting significant challenges for implementing responses.

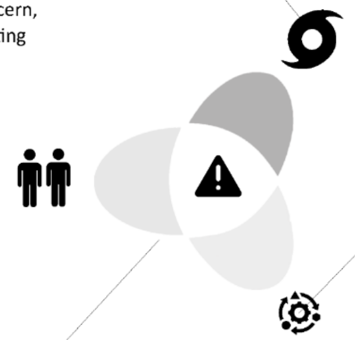
1 TROPICAL CLIMATE REGIONS
Where are areas that are going to be face unprecedented TCs?

Data from STORM (Bloemendaal et al., 2022). Historical period (1980-2017) and SSP5-8.5 (2015-2050)

Return periods (RP), weighted by category based on Emanuel (2011). Coastal areas were classified based on:

1) Highest TC category indicator,
2) Mayor TCs (> category 2) frequency indicator and
3) Minor TCs (< category 3) frequency indicator





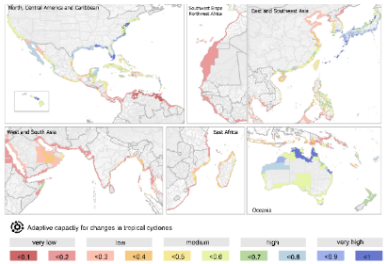
3 ADAPTIVE CAPACITY AS VULNERABILITY UNDER CLIMATE CHANGE
Providing information at subnational level

Local-Level Experience Indicator: IBTrACS dataset (1980 to 2023).

National-Level Insurance Coverage Indicator: Ratio of insured damage to total recorded damage caused by TC).

National-Level readiness readiness Indicator: ND-GAIN readiness indicators of a country's preparedness.

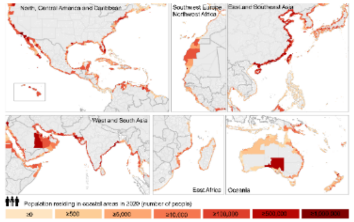
$$AC_{TC,i}=LC_{TC,i}+IP_{TC,i}+R_i$$



Following the IPCC AR6 risk framework (Reisinger et al. 2020)

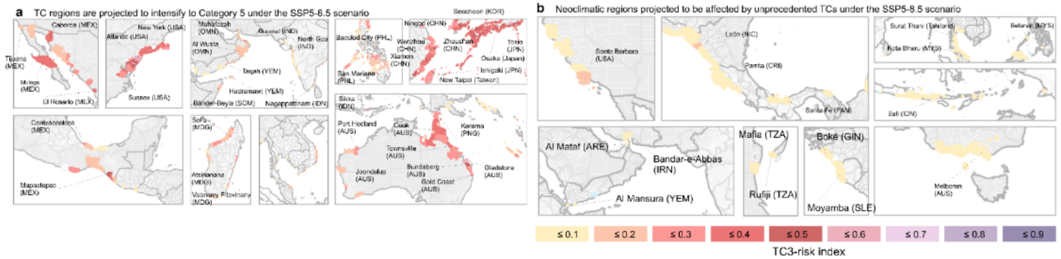
2 EXPOSED POPULATION
People centric

Number of people residing at coastal locality in 2020 from Worldpop and population growth using SSP5 (fossil-fuelled development), calculated as the difference between the population in 2050 and 2020, based on CIESIN projections.



4 RISK HOTSPOTS
For tropical cyclone regions

Developed the Tropical Cyclone under Climate Change risk index (TC3-risk index) to assess risk across different climatic regions.

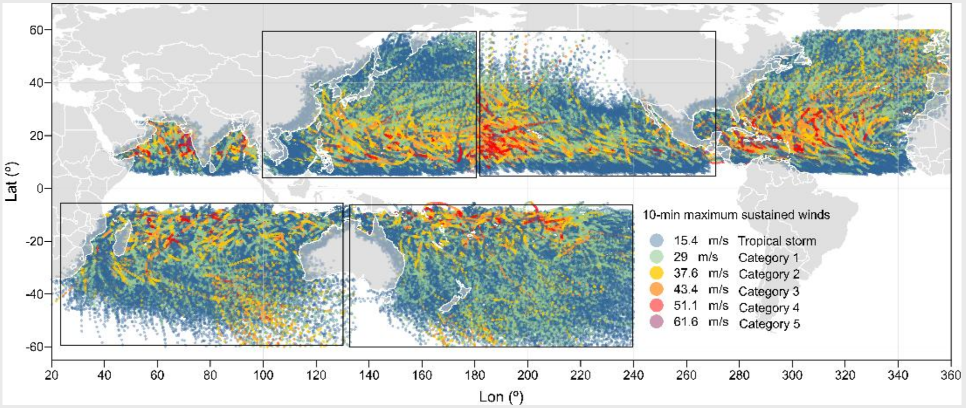


Odériz et al. (under review) Global hostspots of unprecedented tropical cyclone risk to tailored adaptation.

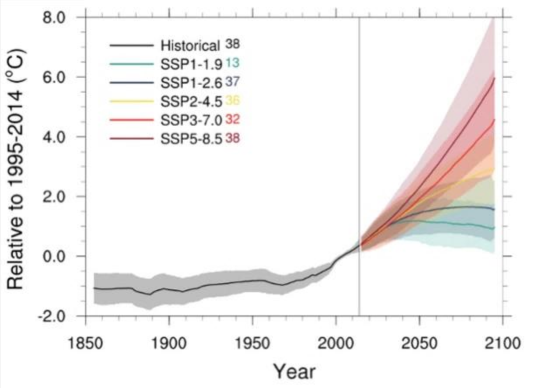
MARIE SKŁODOWSKA-CURIE ACTIONS
Funded by the European Union
UC | FUNDACIÓN INSTITUTO DE HIDRÁULICA AMBIENTAL DE CANTABRIA | IHcantabria INSTITUTO DE HIDRÁULICA AMBIENTAL UNIVERSIDAD DE CANTABRIA

MOTIVATION

- Storm surges are among the deadliest hazards associated with tropical cyclones (TCs) .
- Uncertainty in TCs under climate change is critical for long-term risk assessments.
- Computational limitations restrict the use of such tools, especially in countries with limited technical infrastructure.
- We need better focus TWL for support decision making and climate change adaptation with a global coverage.



CMIP6, CMIP7, etc.



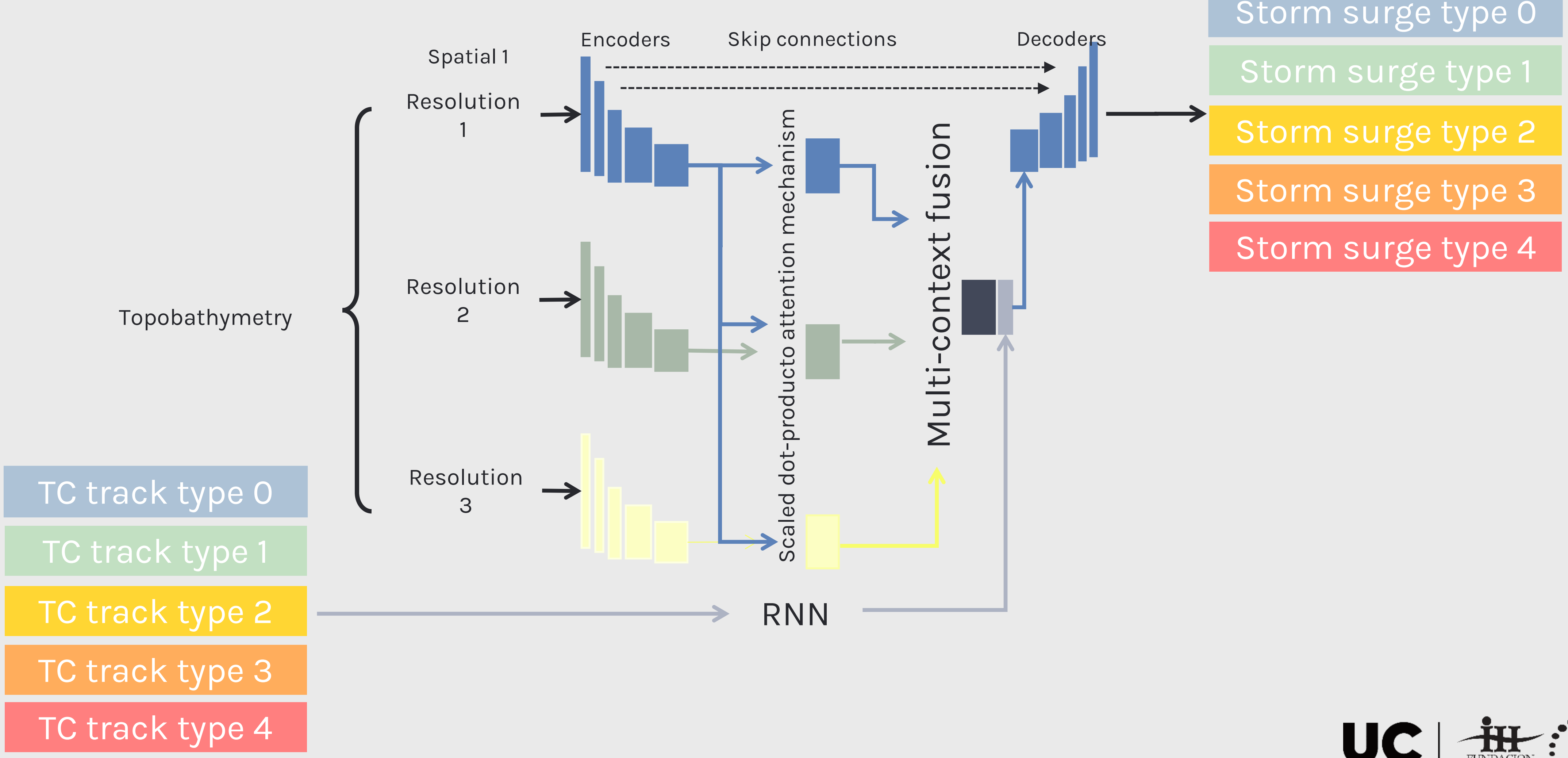
GCMs

ACCESS1.0
BCC-CSM1.1
CCSM4
CNRM-CM5
CNRM-CM5-2
CSIRO-Mk3.6.0
CanESM2
FGOALS-g2
GFDL-CM3
GFDL-ESM2G

Rapid, large-scale tropical cyclones (TCs) hazard assessment tools

MOTIVATION

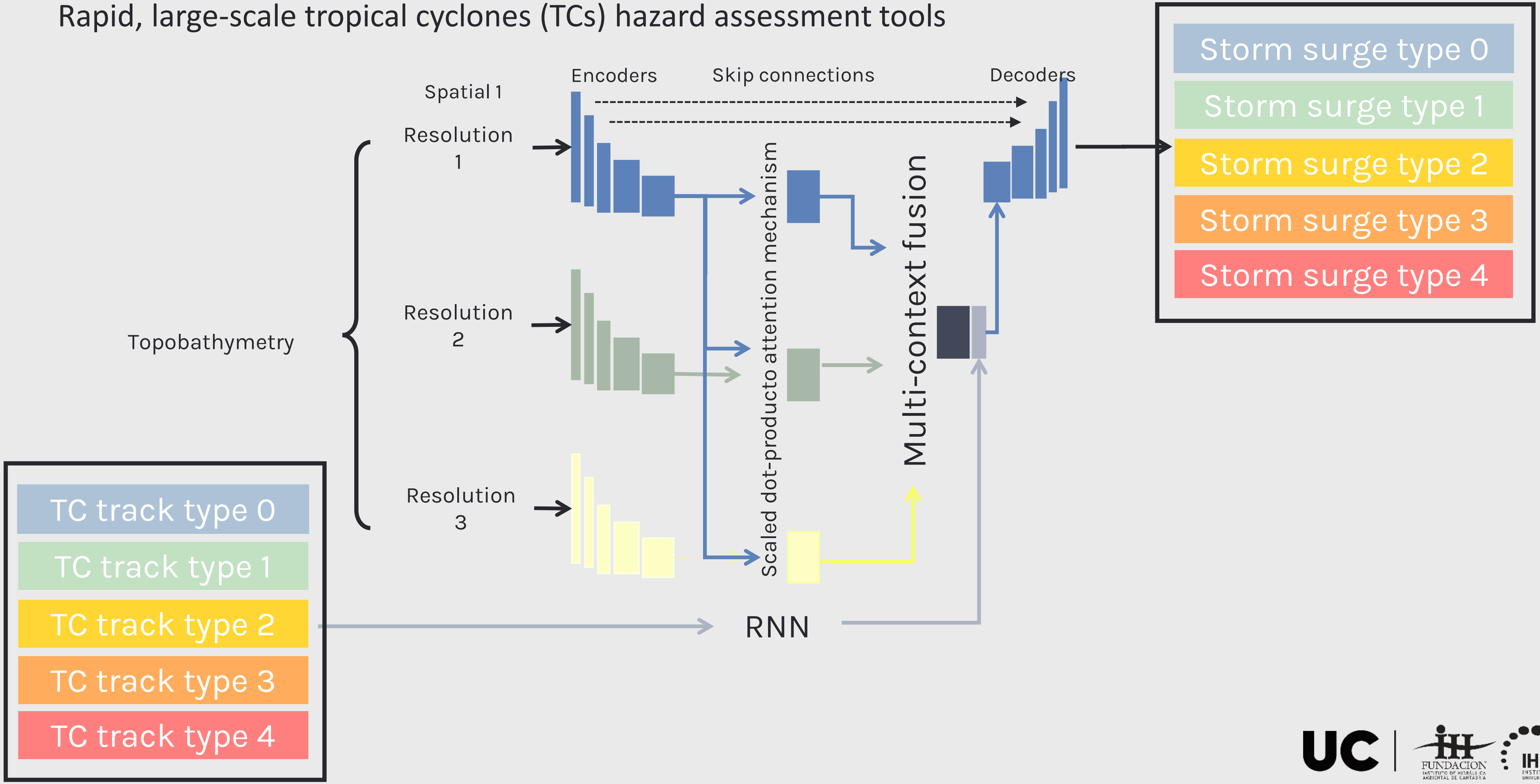
Rapid, large-scale tropical cyclones (TCs) hazard assessment tools



UC | FUNDACION IH | IH cantabria

MOTIVATION

Rapid, large-scale tropical cyclones (TCs) hazard assessment tools



OBJECTIVE

Explore the data

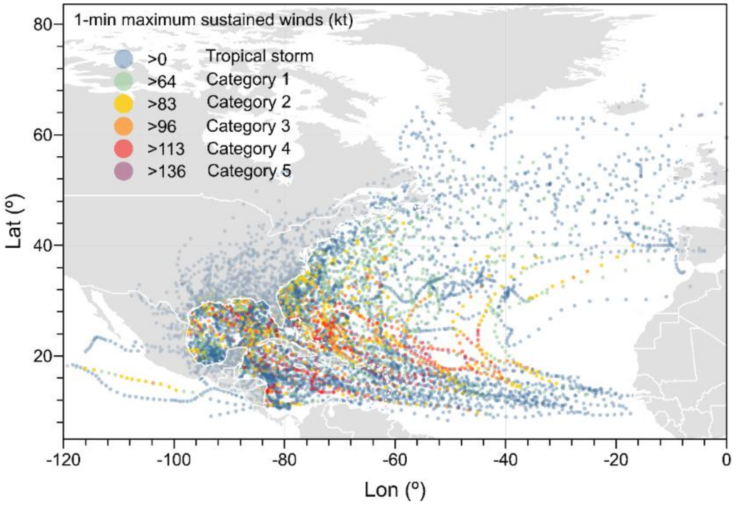
01	To understand it and to control the training process and
02	To guide training and ensure the model learns from all types of storm surge–related tropical cyclone events

00 DATA

Residual storm surge (Muis et al., 2016) GSTM reanalysis with 25 km on coastal aras.

We focused only on TCs from IBTrACS v4 (Knapp et al.,2010) for the period 1980-2019:

- which are well defined in ERA5 (Bourdin et al., 2022): Subtropical storms, short-lived alternate tracks, and storms lasting less than one day are not considered. Tracks that do not reach tropical storm are also removed(≥16 m/s 10-min sustained winds).
- Landfalling cyclones are included.
- IBTrACS 3-hourly data are interpolated to 1-hourly resolution.
- Radius of maximum winds are calculated for each time step with (Knaff et al 2015).
- The TC lifetime is extended by 10 days after dissipation to account for storm residual effects.

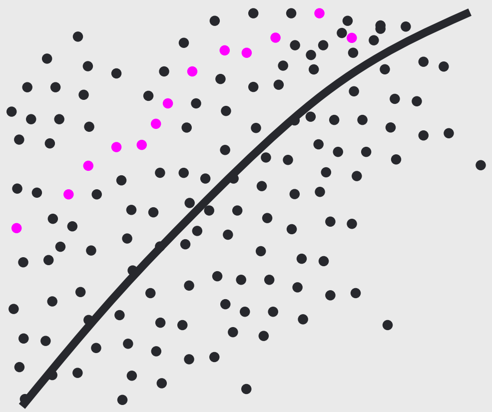


Total in IBTrACS	Well represented by ERA 5 (Bourdin et al., 2024)
604	191

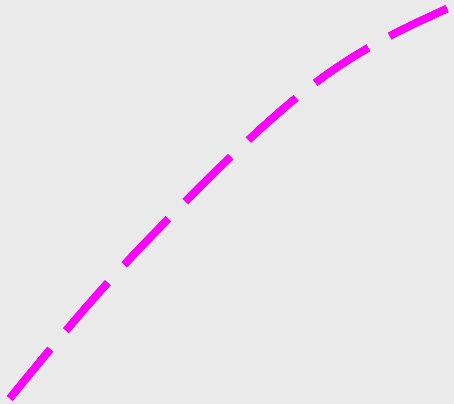
Bourdin, S., Fromang, S., Dulac, W., Cattiaux, J., and Chauvin, F.: Intercomparison of four algorithms for detecting tropical cyclones using ERA5, Geosci. Model Dev., 15, 6759–6786, https://doi.org/10.5194/gmd-15-6759-2022, 2022.
Knaff, J. A., S. P. Longmore, R. T. DeMaria, y D. A. Molenar, 2015: Improved tropical cyclone flight-level wind estimates using routine infrared satellite reconnaissance. J. Appl. Meteor. Climatol., 54, 463–478, doi:10.1175/JAMC-D-14-0112.1
Knapp, K. R., M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann, 2010: The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying tropical cyclone best track data. Bulletin of the American Meteorological Society, 91, 363-376. doi:10.1175/2009BAMS2755.1
Sanne Muis, Maialen Irazoqui Apecechea, José Antonio Álvarez, Martin Verlaan, Kun Yan, Job Dullaart, Jeroen Aerts, Trang Duong, Rosh Ranasinghe, Dewi le Bars, Rein Haarsma, Malcolm Roberts, (2022): Global sea level change time series from 1950 to 2050 derived from reanalysis and high resolution CMIP6 climate projections.
Copernicus Climate Change Service (C3S) Climate Data Store (CDS). DOI: 10.24381/cds.a6d42d60 (Accessed on 01-08-2025)

METHODS

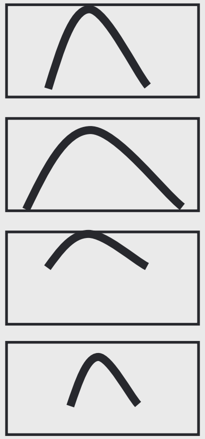
01 KM-DBSCAN
Filtering storm surge and drivers



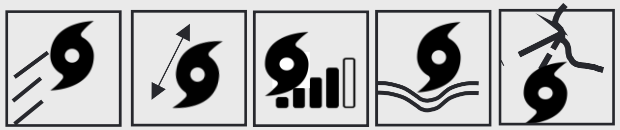
02 Event augmentation
Splitting a TC track into multiple TC events



03 K-mean: storm surge types
Clustering ss-hydrographs and TC characteristics

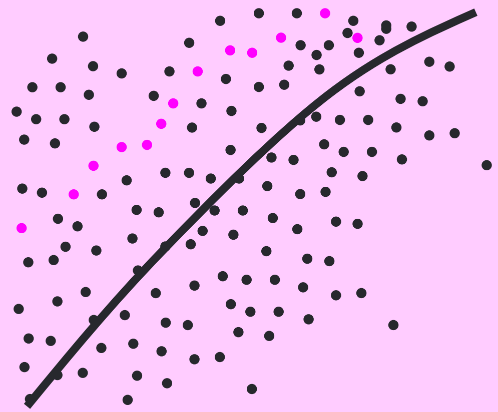


04 Random Forest
Feature importance of TC variables by surge types

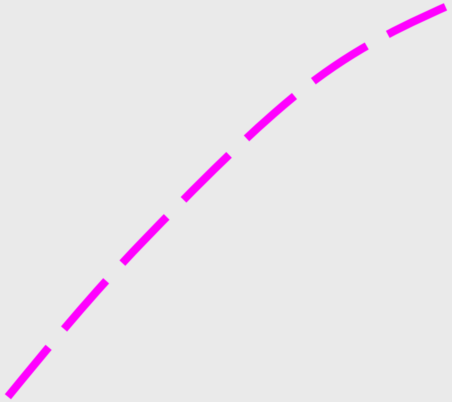


RESULTS

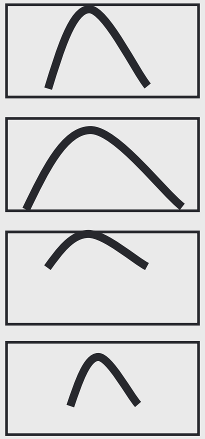
01 **KM-DBSCAN**
Filtering storm surge and drivers



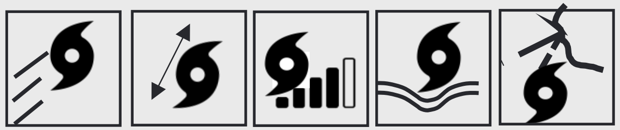
02 **Event augmentation**
Splitting a TC track into multiple TC events



03 **K-mean: storm surge types**
Clustering ss-hydrographs and TC characteristics



04 **Random Forest**
Feature importance of TC variables by surge types



01 TRACKING SYSTEM: KM-DBSCAN

KM-DBSCAN (Dynamic clustering with **K-M**eans and an static clustering **DBSCAN**)

01 Dynamic K-Means Clustering (Odériz et al 2021) Group values that are similar

The maximum area influence is 10 degrees (~ 1000 km)

$X = \{x_1, x_2, \dots, x_n\}$

$C = \{c_1, c_2, \dots, c_k\}$ *where K is number of clusters*

$S_j = \left\{x_i \mid \|x_i - c_j\|^2 \le \|x_i - c_k\|^2, \forall k \in \{1, \dots, K\}\right\}$ *where K is number of clusters*

Where Sj is the set of points assigned to cluster j and applying the Euclidean distance

The algorithm recalculates the centroids of each cluster $c_j = \frac{1}{|S_j|}\sum_{x_i \in S_j} x_i$

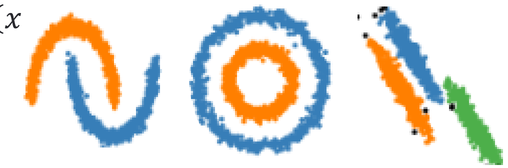
Where $|S_j|$ is the number of points in cluster j with Eucledean distance $J = \sum_{j=1}^{k}\sum_{x_i \in S_j}\|x_i - c_j\|^2$

ss >0.1 m

02 DBSCAN: Group values that are spatially proximate

$N_\varepsilon(x_i) = \{x_i \in X \mid \|x_j - c_i\| \le \varepsilon\}$

$x_j \in N_\varepsilon(x$

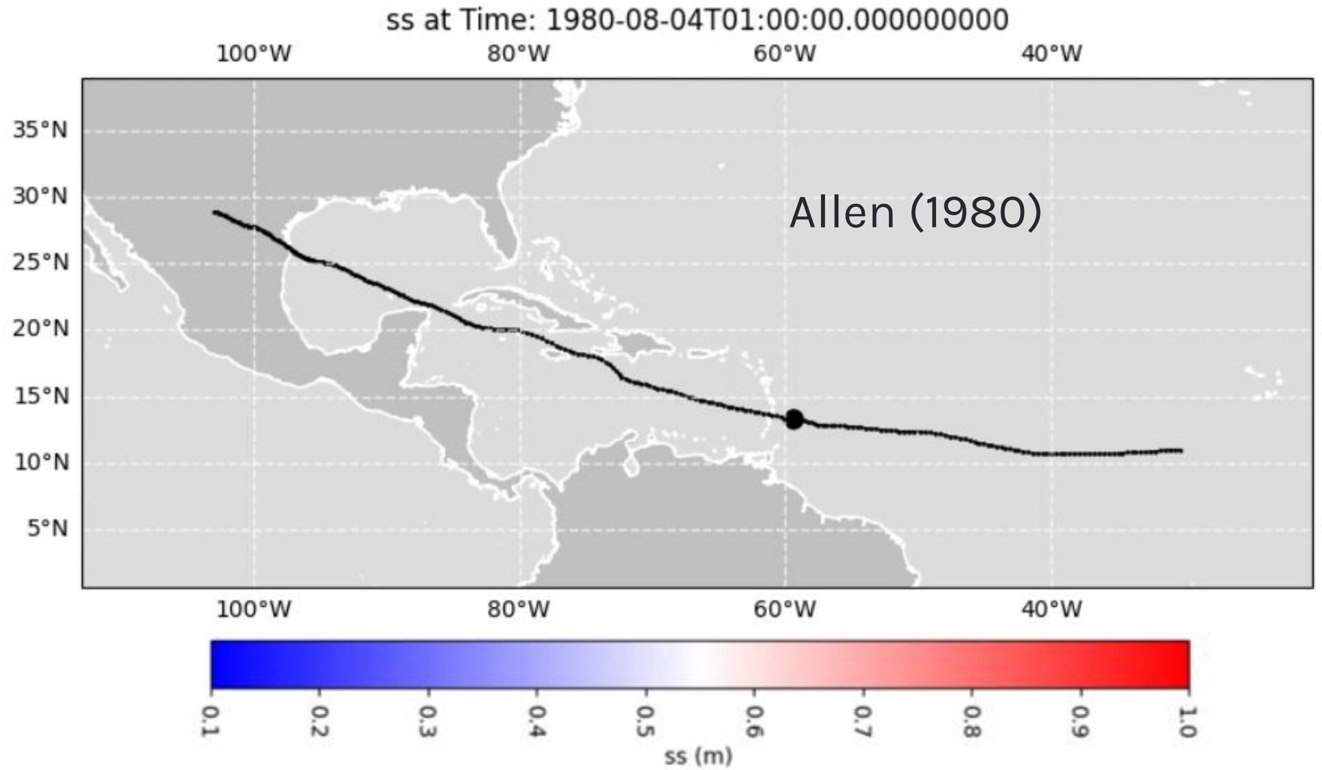


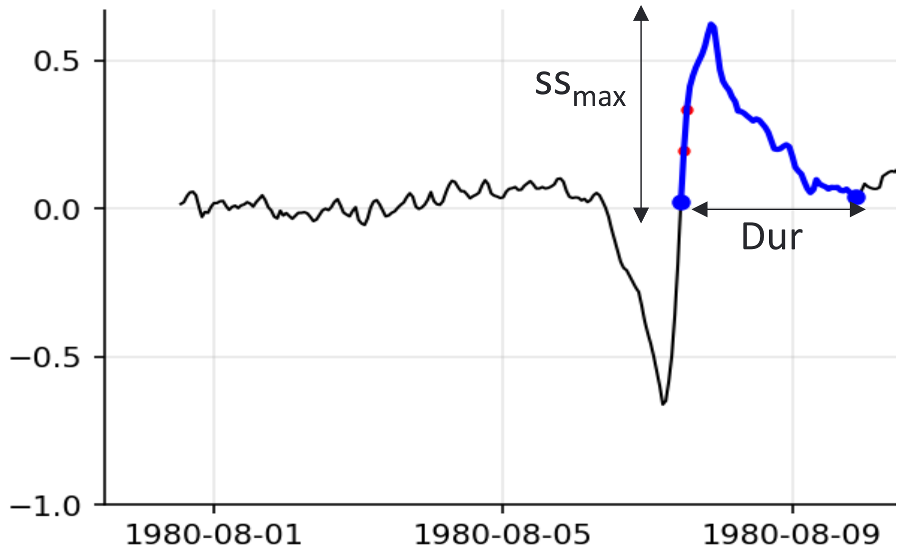
- radius of the neighborhood around a point(eps)=1
- Minum samples are 1, every point is automatically a core point.

Odériz et al. (in preparation)

UC | FUNDACION IH | IHcantabria

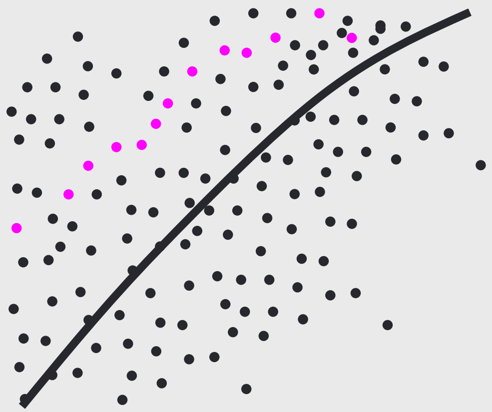
01 TRACKING SYSTEM: KM-DBSCAN



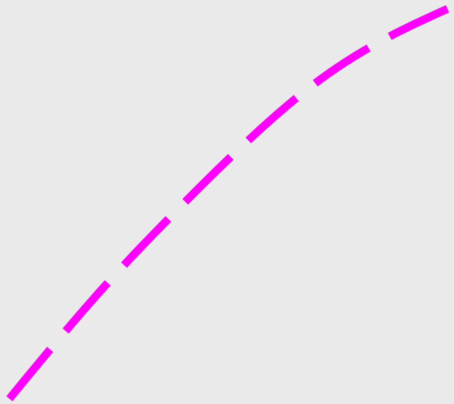


RESULTS

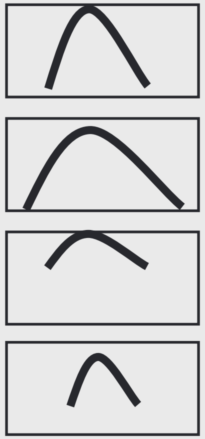
01 KM-DBSCAN
Filtering storm surge and drivers



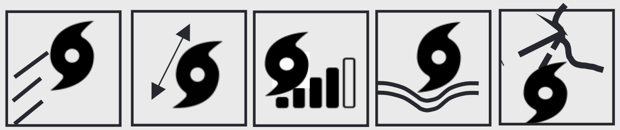
02 Event augmentation
Splitting a TC track into multiple TC events



03 K-mean: storm surge types
Clustering ss-hydrographs and TC characteristics

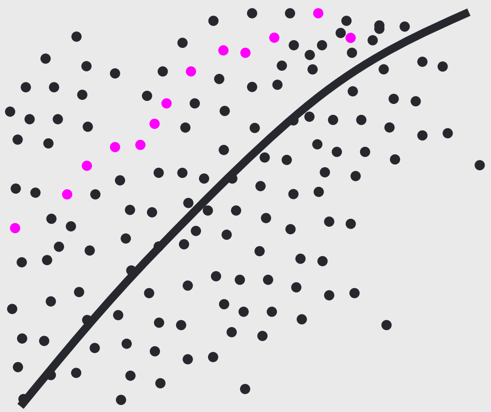


04 Random Forest
Feature importance of TC variables by surge types

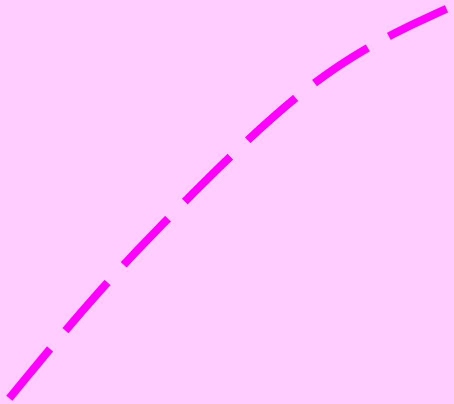


RESULTS

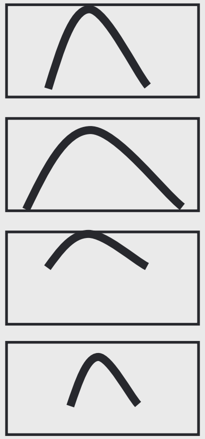
01 KM-DBSCAN
Filtering storm surge and drivers



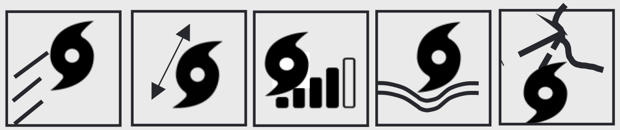
02 Event augmentation
Splitting a TC track into multiple TC events



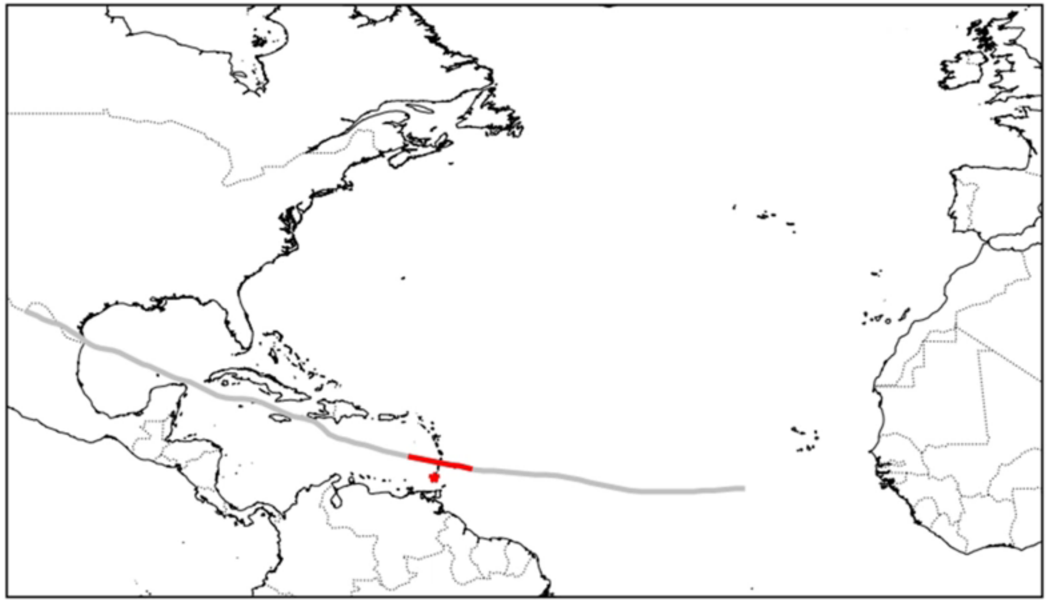
03 K-mean: storm surge types
Clustering ss-hydrographs and TC characteristics

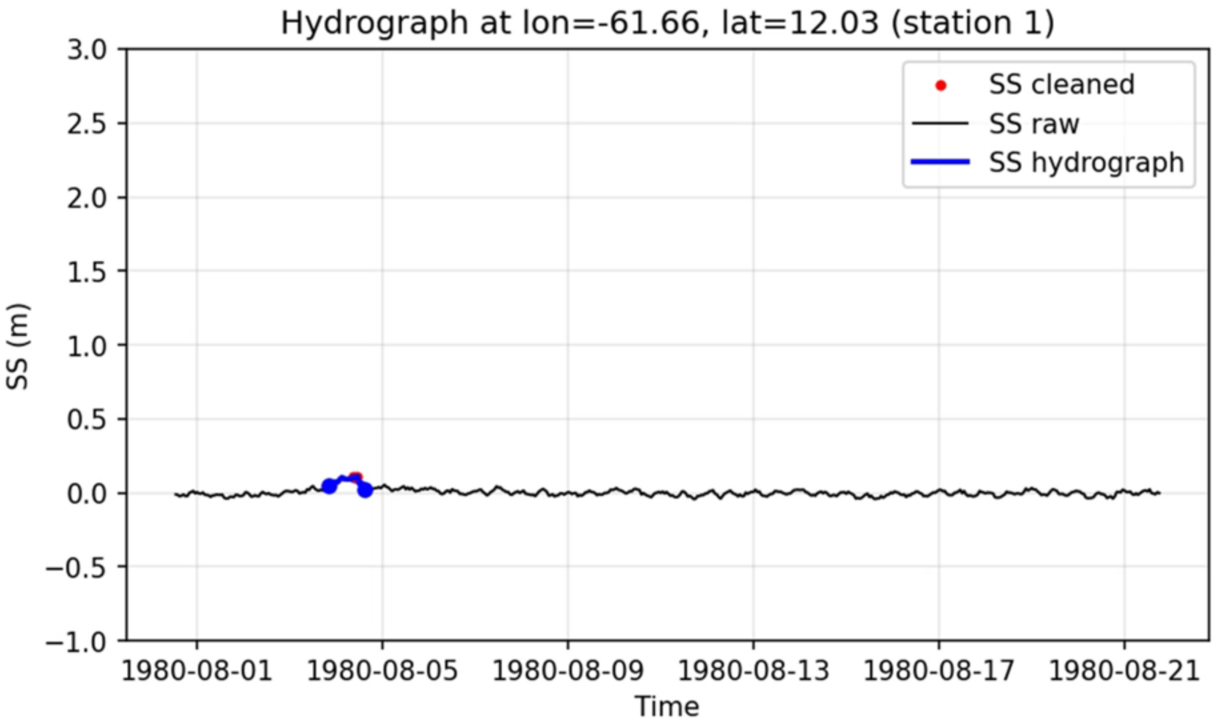


04 Random Forest
Feature importance of TC variables by surge types



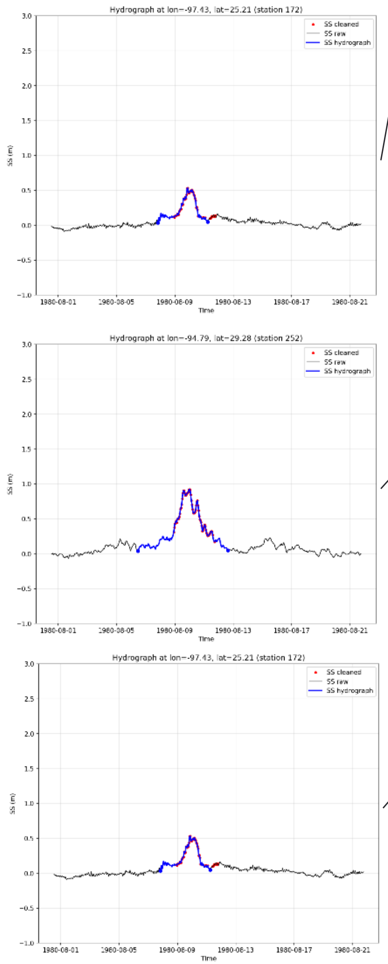
02 EVENT AUGMENTATION: SPLITTING A TC INTO MULTIPLE TCS

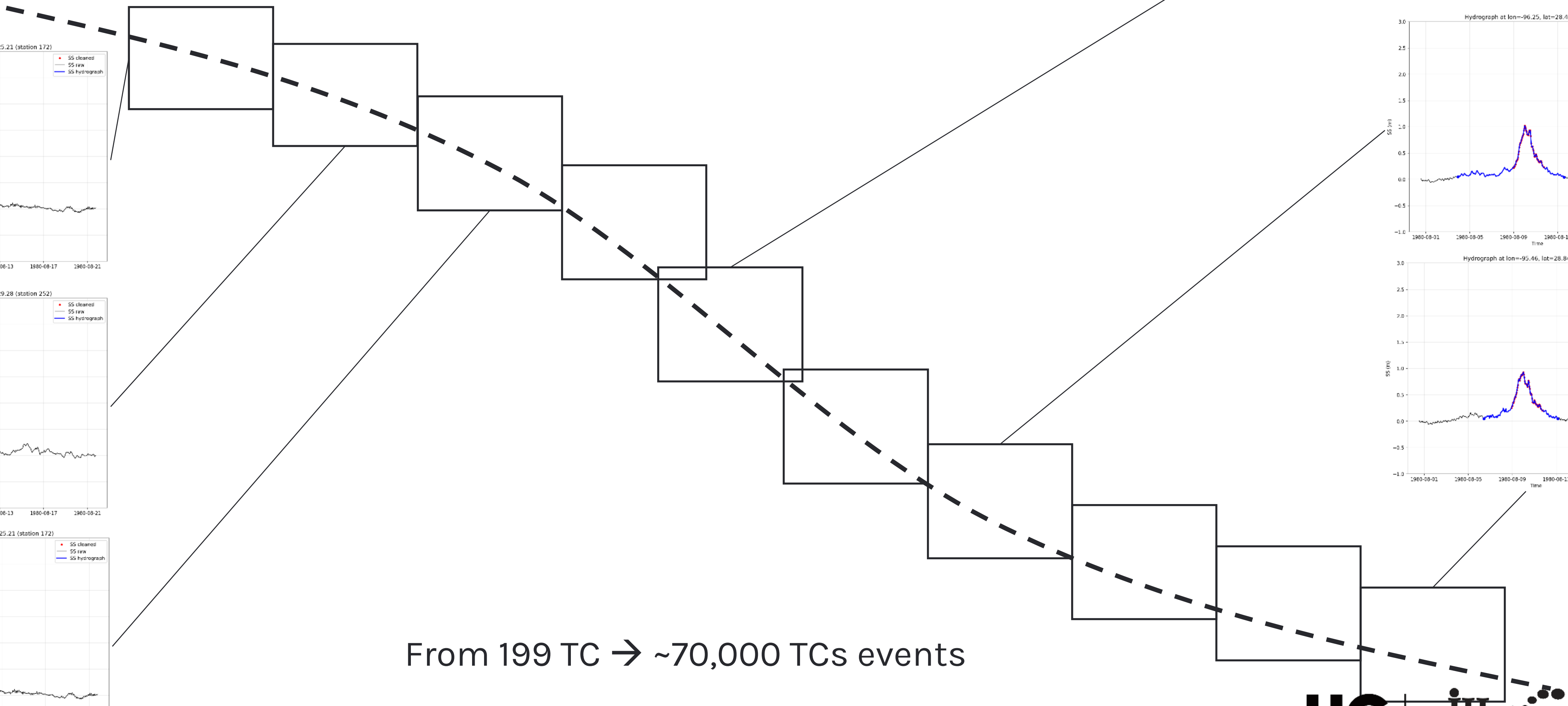


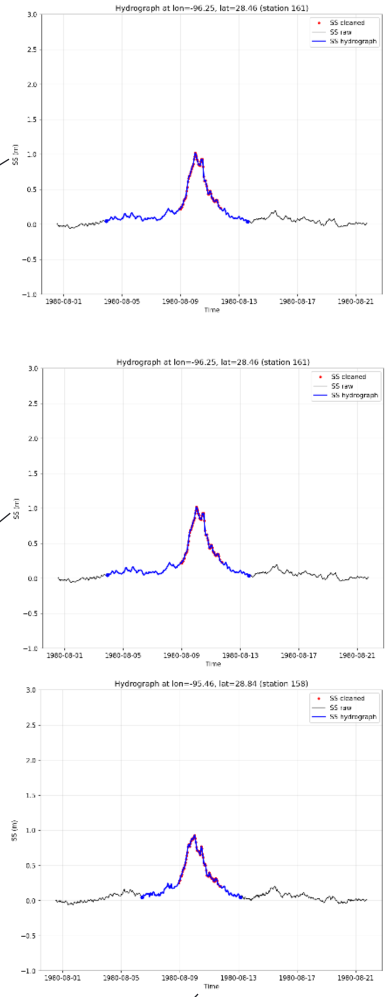


02 EVENT AUGMENTATION: SPLITTING A TC INTO MULTIPLE TCS

02 EVENT AUGMENTATION: SPLITTING A TC INTO MULTIPLE TCS



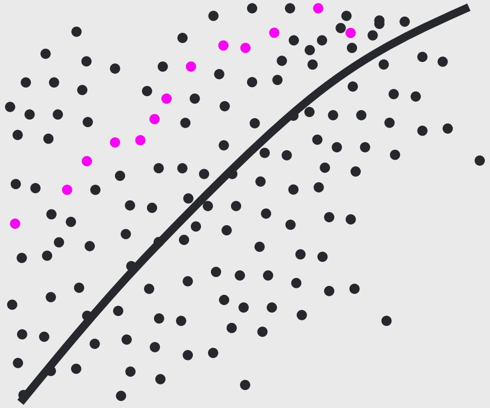




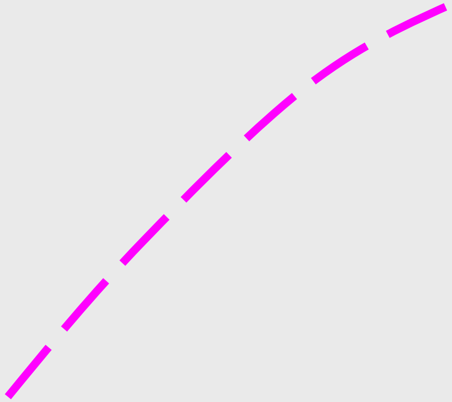
From 199 TC → ~70,000 TCs events

RESULTS

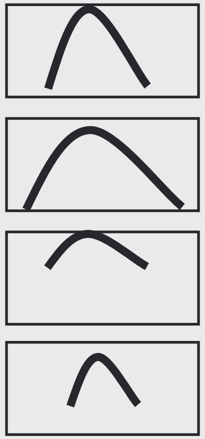
01 KM-DBSCAN
Filtering storm surge and drivers



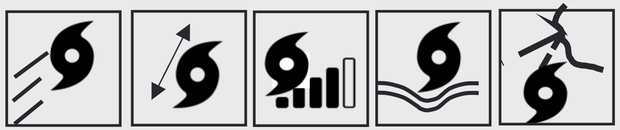
02 Event augmentation
Splitting a TC track into multiple TC events



03 K-mean: storm surge types
Clustering ss-hydrographs and TC characteristics

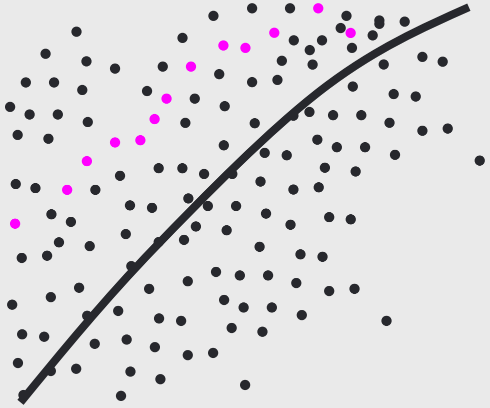


04 Random Forest
Feature importance of TC variables by surge types

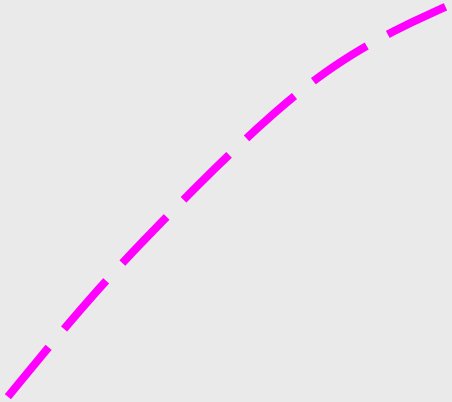


RESULTS

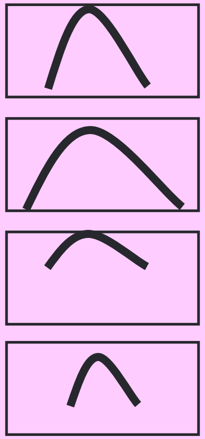
01 KM-DBSCAN
Filtering storm surge and drivers



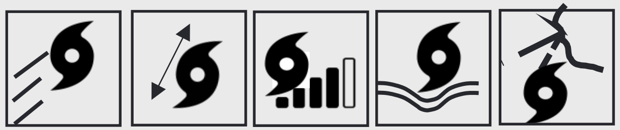
02 Event augmentation
Splitting a TC track into multiple TC events



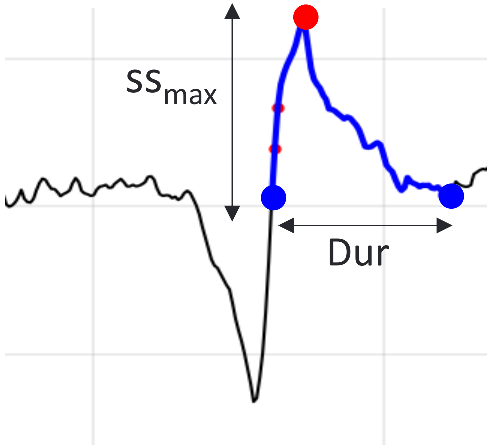
03 K-mean: storm surge types
Clustering ss-hydrographs and TC characteristics

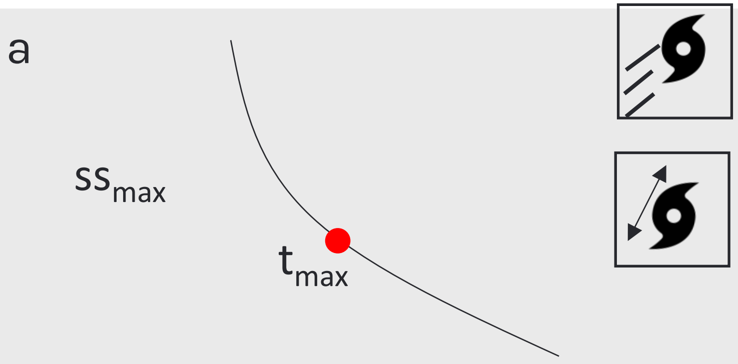


04 Random Forest
Feature importance of TC variables by surge types

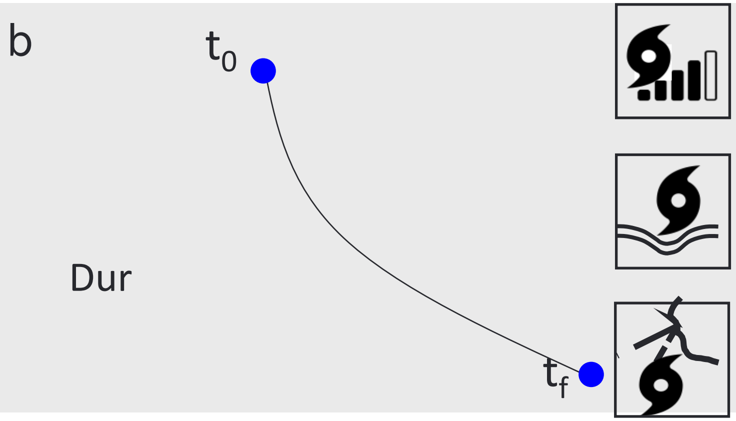


K-MEAN: STORM SURGE TYPES





TC_{speed} (m/s): Translation speed of the TC.
t_{asc} (hr): Time during which the surge rises.
W_{max} (knots): Maximum wind speed.
P_{min} (mb): Minimum central pressure of the TC.
Dir_{TC} (º): Direction of the TC in degrees.
TC_{dist} (km): Distance between the storm surge location and the TC.
RMW (nmiles): Radius of maximum winds.

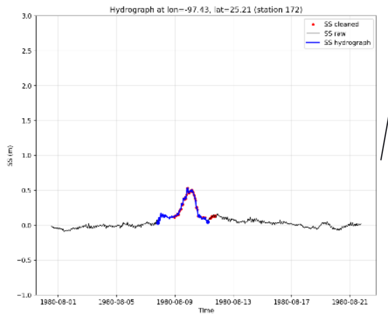


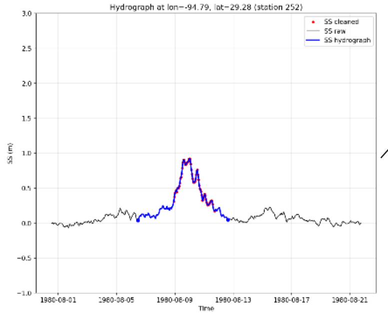
$\overline{W_{max}}$ (knots): average of the wind maximum speed during the storm surge event
$\overline{Dir_{Tc}}$ (º): average of the storm direction during the storm surge event
$\overline{RMW}$ (km): average radius of maximum winds during the storm surge event
$\overline{P_{min}}$ (mb): average minimum central pressure of the TC during the storm surge event
$\overline{TC_{speed}}$(m/s): average translation speed of the TC during the storm surge event

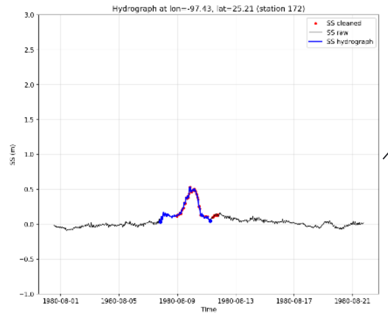
Elbow method to estimate the optimal number of clusters for the dataset data1 (k values from 2 to 10).

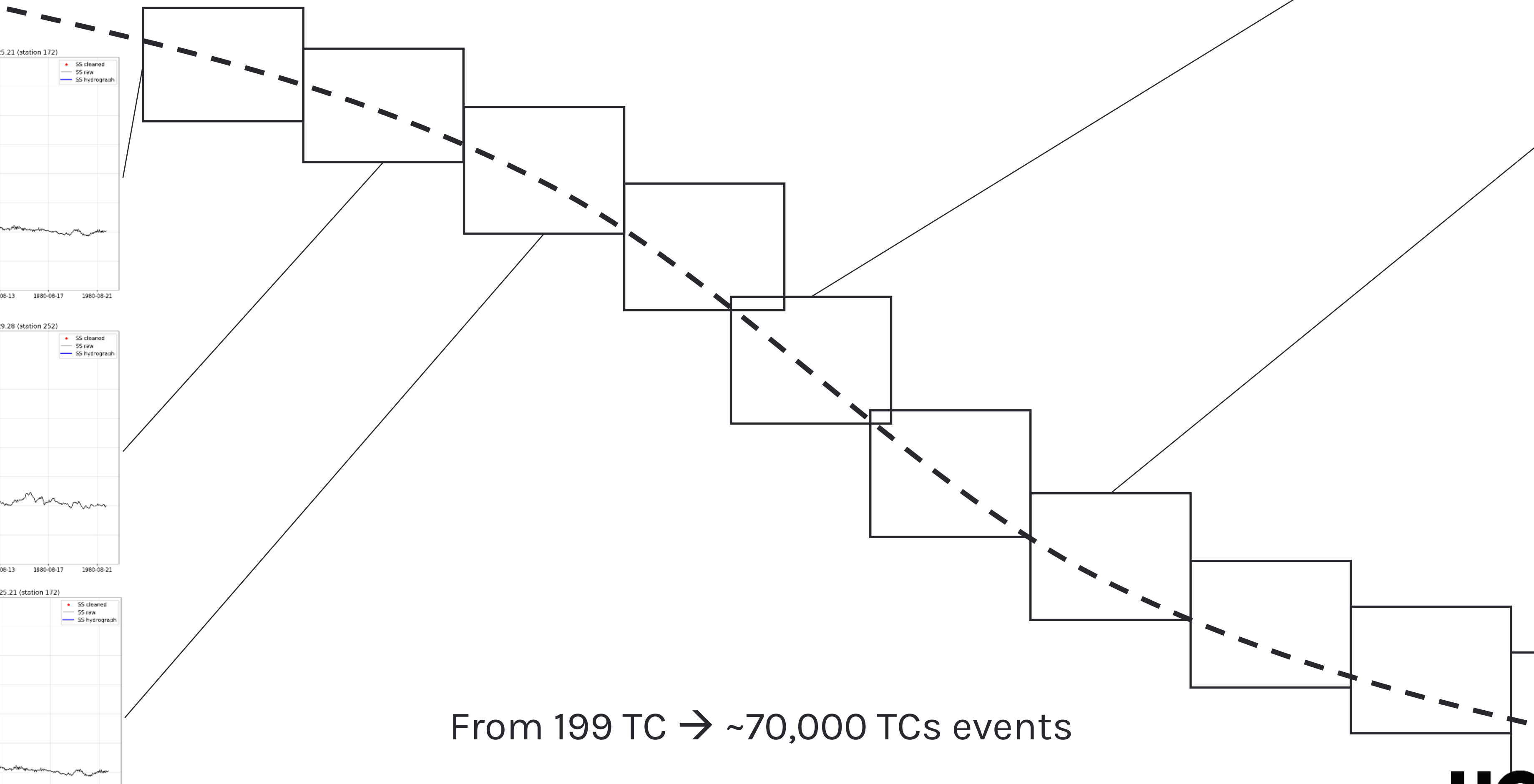
Normalize data scaling each variable to the range [0, 1].

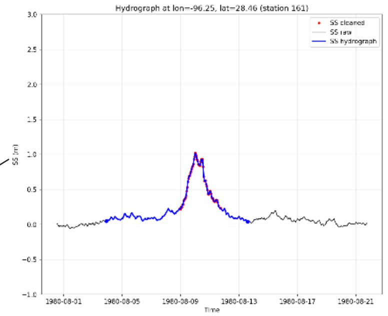
03 K-MEAN: STORM SURGE TYPES

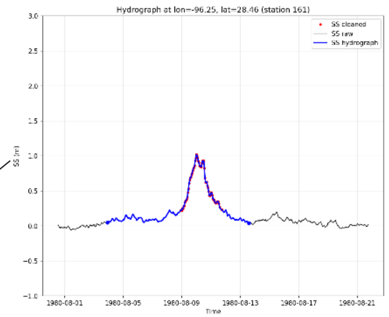


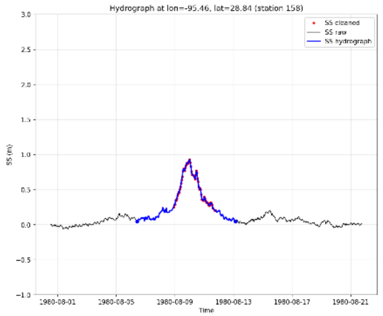






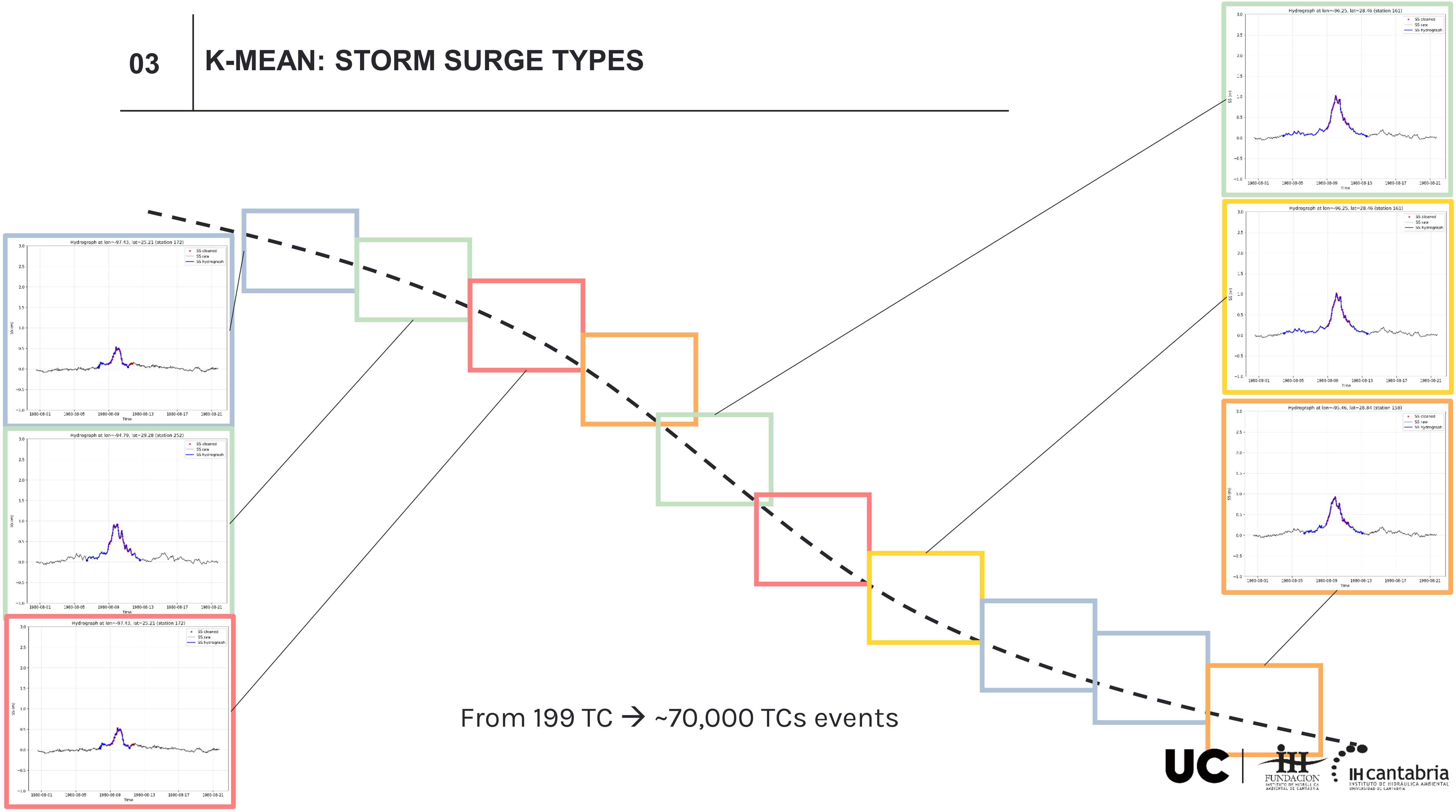






From 199 TC → ~70,000 TCs events

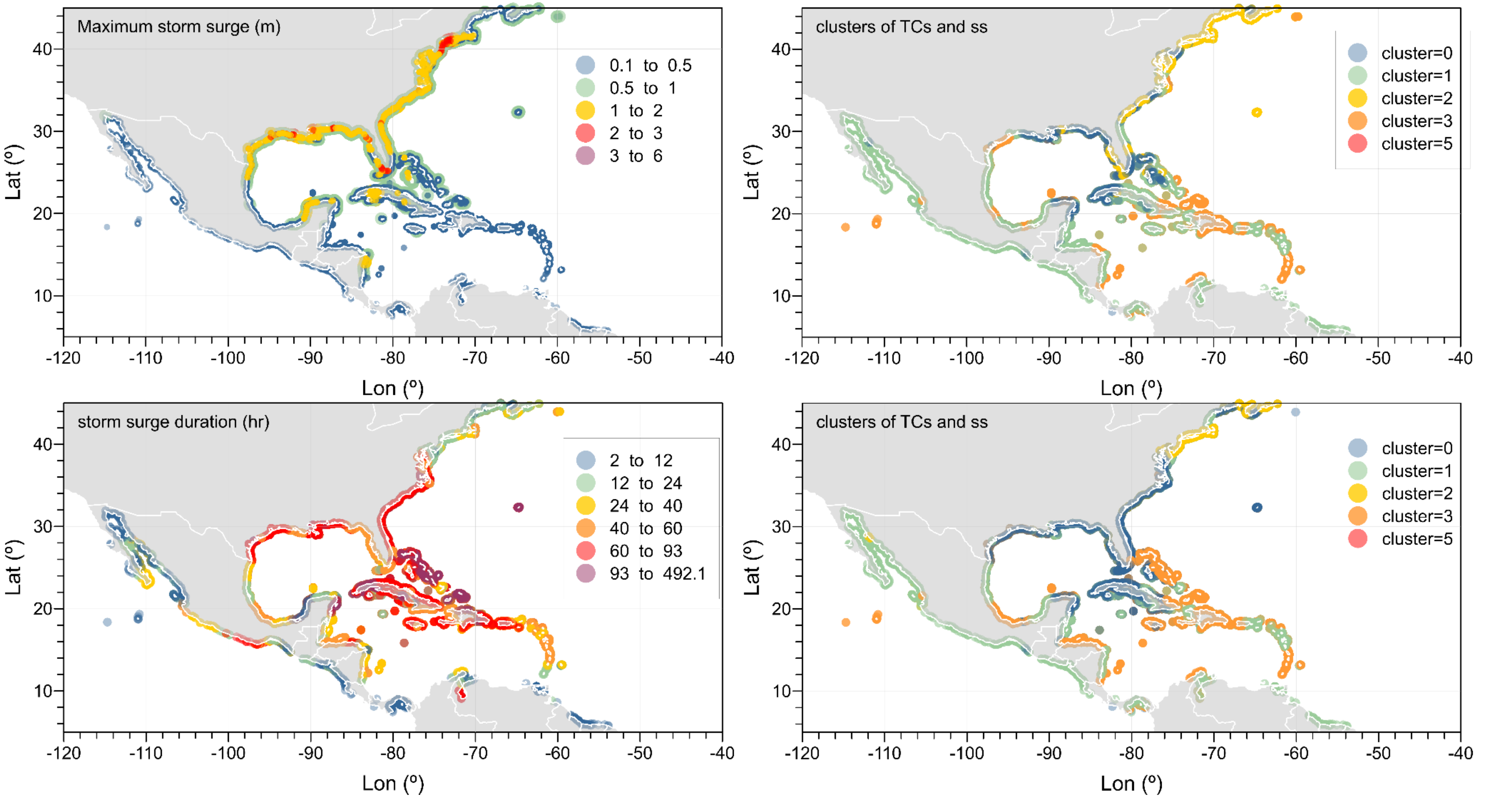
03 K-MEAN: STORM SURGE TYPES



From 199 TC → ~70,000 TCs events

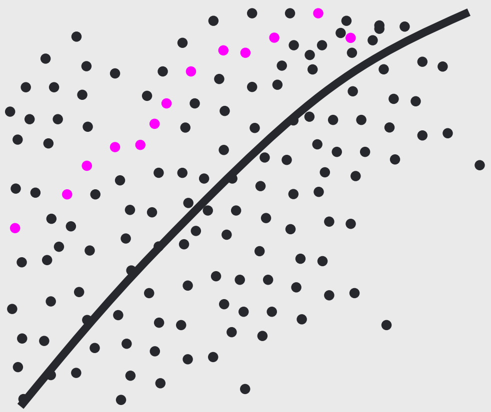
03

K-MEAN: STORM SURGE TYPES

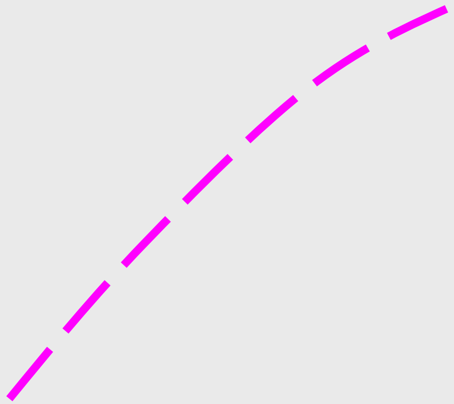


RESULTS

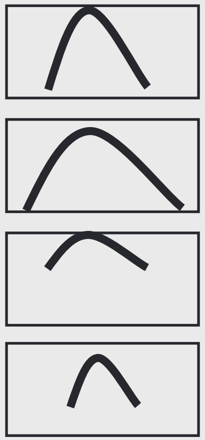
01 KM-DBSCAN
Filtering storm surge and drivers



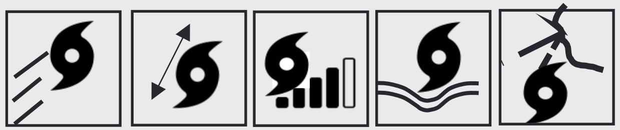
02 Event augmentation
Splitting a TC track into
multiple TC events



03 K-mean: storm surge types
Clustering ss-hydrographs
and TC characteristics

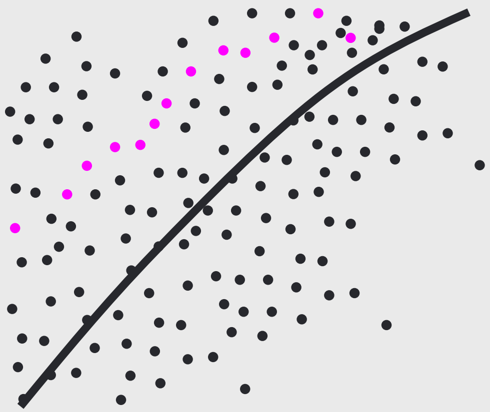


04 Random Forest
Feature importance
of TC variables by surge types

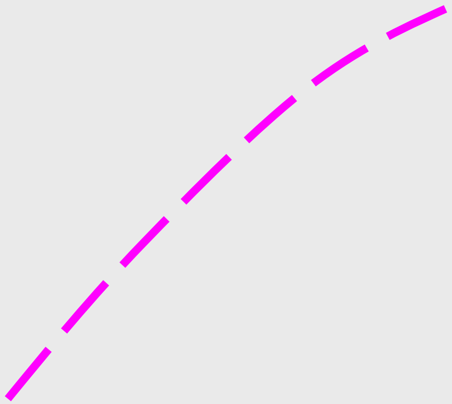


RESULTS

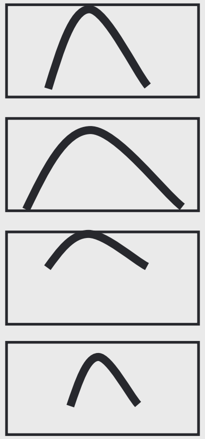
01 KM-DBSCAN
Filtering storm surge and drivers



02 Event augmentation
Splitting a TC track into multiple TC events



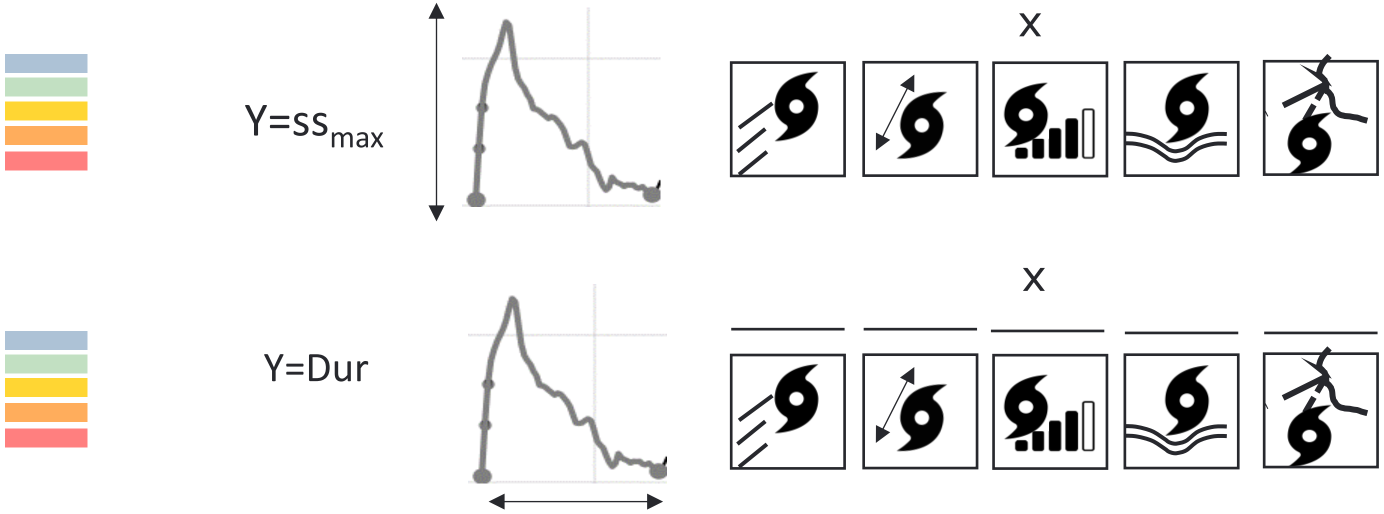
03 K-mean: storm surge types
Clustering ss-hydrographs and TC characteristics



04 Random Forest
Feature importance of TC variables by surge types

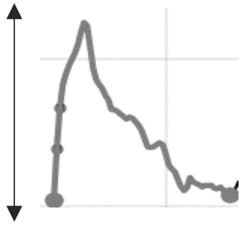
04 RANDOM FOREST: FEATURE IMPORTANCE

- Extract predictor variables (X) and target variable (y) for the cluster.
- A Random Forest Regressor is an ensemble learning method for regression, (300 trees, 42 running times).
- KFold cross-validation: 80% training, 20% validation with 5 folds, randomizes the data before splitting.
- For each fold, Compute R^2 scores, the model's predictions are compared to the true values of y.



04

RANDOM FOREST: FEATURE IMPORTANCE



Type of storm surge	R2 (μ)	R2 (std)	TC_{dist}:	RMW	TC_{speed}	W_{max}	Dir_{Tc}	P_{min}
3	0.75	0.05	1.80	1.14	0.37	0.34	0.23	0.20
1	0.85	0.02	2.57	0.71	0.62	0.51	0.41	0.21
0	0.86	0.01	0.88	0.76	0.65	0.40	0.18	0.13
4	0.79	0.03	0.94	0.64	0.52	0.38	0.24	0.24
2	0.88	0.01	1.05	0.80	0.48	0.25	0.22	0.14

SS type 0: Storm surge peaks are strongly influenced by cyclone proximity and storm size. Cyclone speed also matters here.

SS type 1: storm size is the most important factor, but proximity and wind speed.

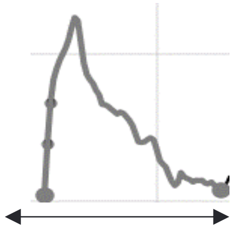
SS type 2: cyclone size and distance drive the surge, but storm direction also plays a significant role — suggesting coastal orientation effects.

SS type 3: Predictability is lower, indicating more complex or heterogeneous cases. Surge is influenced by cyclone size and distance, but also direct wind forcing and pressure play a role.

SS type 4: This regime emphasizes proximity and size, with storm direction again emerging as important, reflecting track–coastline alignment.

04 | RANDOM FOREST: FEATURE IMPORTANCE

Dur



Type of storm surge	R2 (μ)	R2 (std)	$\overline{Dir_{Tc}}$	$\overline{TC_{speed}}$	$\overline{RMW}$	$\overline{P_{min}}$	$\overline{W_{max}}$:
3	0.90	0.03	0.86	0.74	0.69	0.60	0.35
1	0.89	0.02	0.93	0.85	0.47	0.42	0.36
0	0.93	0.01	0.93	0.62	0.55	0.53	0.43
4	0.97	0.00	1.12	0.49	0.44	0.29	0.15
2	0.97	0.01	0.83	0.65	0.48	0.40	0.32

Cluster 0 Duration is primarily controlled by storm direction, with pressure and translation speed as secondary influences.

Cluste 1 central pressure and track direction dominate surge duration, highlighting the role of cyclone intensity and orientation.

Cluster 2 Duration is shaped by a combination of low pressure and direction, with storm size and speed also contributing.

Cluster 3 Duration is linked mainly to cyclone direction and forward speed, together with storm size.

Cluster 4 This regime is almost completely determined by cyclone direction, suggesting a strong link between track alignment and surge persistence.

FUTURE WORK

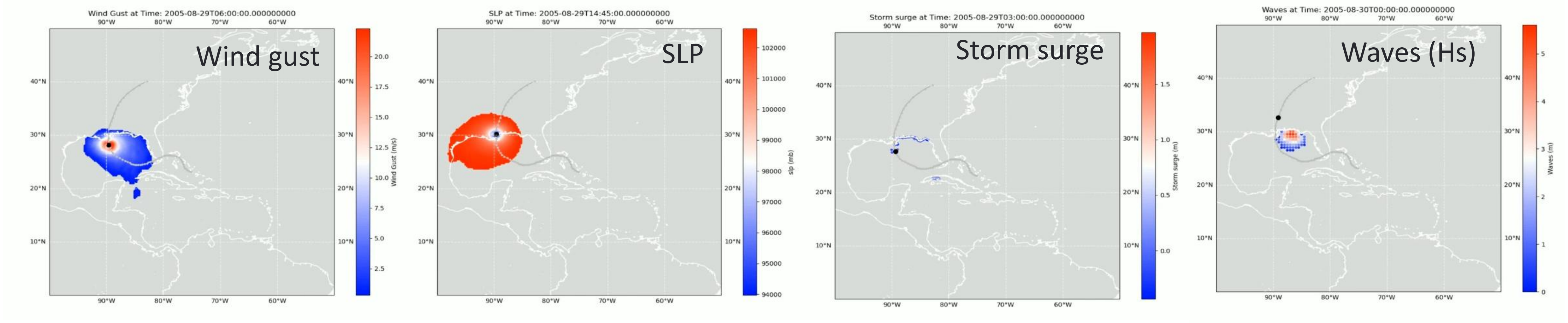
01	Analyze globally the data
02	Integrate coastal configuration as a parameter
03	Considering analysis 6000 storm TC-surge events
04	Training and validate the model
05	Incorporate waves and tides to estimate TWL

THANK YOU!

Itxaso Odériz

MSCA individual fellow, itxaso.oderiz@unican.es

IHCantabria - Instituto de Hidráulica Ambiental de la Universidad de Cantabria, Santander, Spain



Detection and Global Risk Assessment of Coastal Transitional Climate and Neo-Climate Regions in a Warming World 01111262 – TransClima – HORIZON-MSCA-2022-PF-01 funded by the European Union

TRANSCLIMA

MARIE SKŁODOWSKA-CURIE ACTIONS

Funded by the European Union

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

UC | FUNDACIÓN IH cantabria