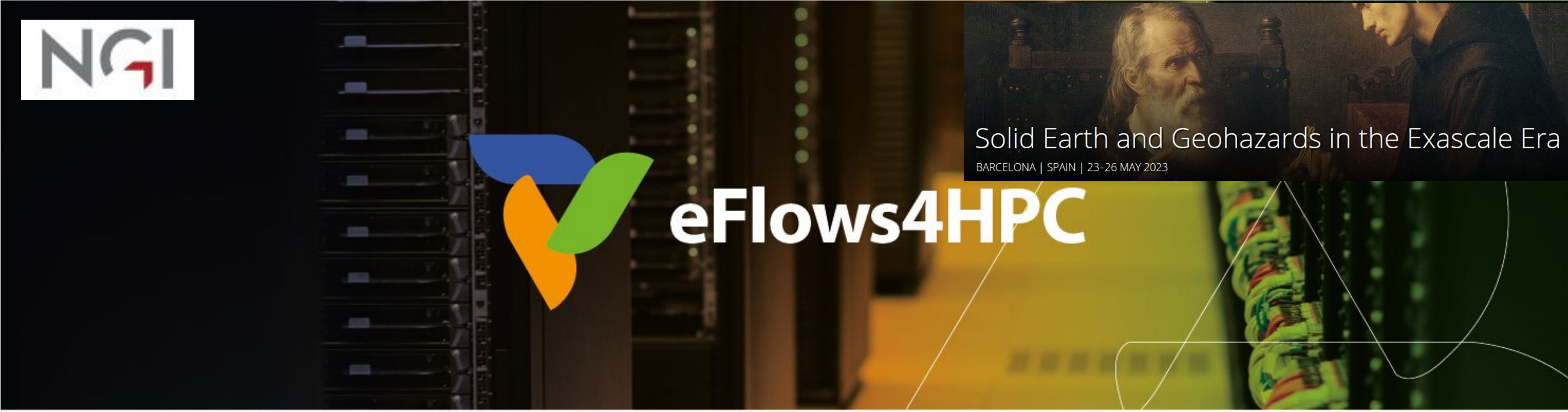


ML Emulation of High Resolution Inundation Maps

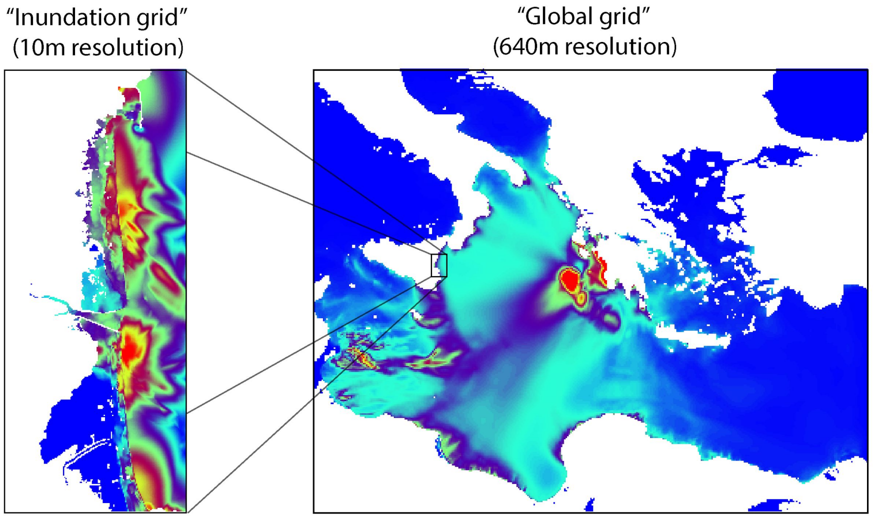
Steven J. Gibbons, Erlend Briseid Storrøsten, Finn Løvholt
Norwegian Geotechnical Institute, Norway

With grateful thanks to colleagues at INGV, Italy, the University of Málaga, Spain, the ChEESE Center of Excellence, CINECA, Rome, the Barcelona Supercomputer Center, Sigma2/UNINETT, Norway for faciliating the models and simulation results on which this work is based.

This project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 955558. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Spain, Germany, France, Italy, Poland, Switzerland, Norway.

Introduction

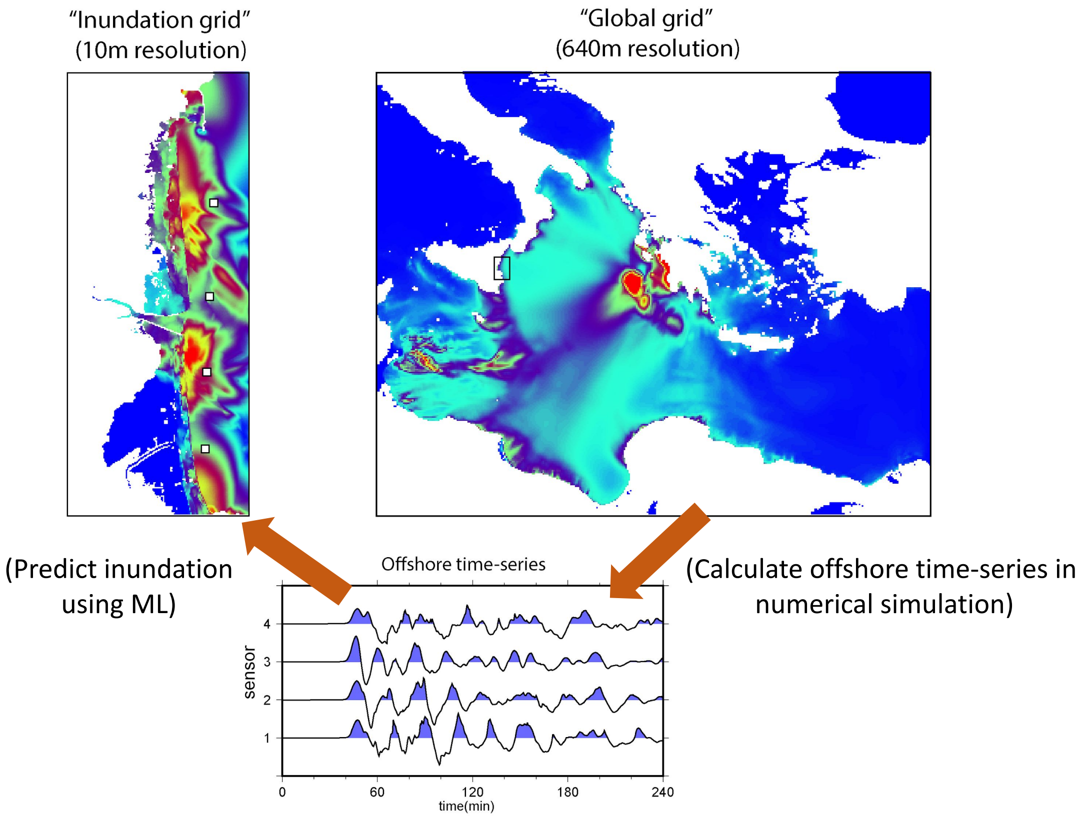
- Often need vast numbers of simulations (10000+)
- Inundation is by far the most expensive part of the calculation.
- Offshore time-series are relatively cheap to calculate.
- Is it posible to replace the simulation of the inundation by a statistical emulator?



(Full numerical simulation using nested grid)

Introduction

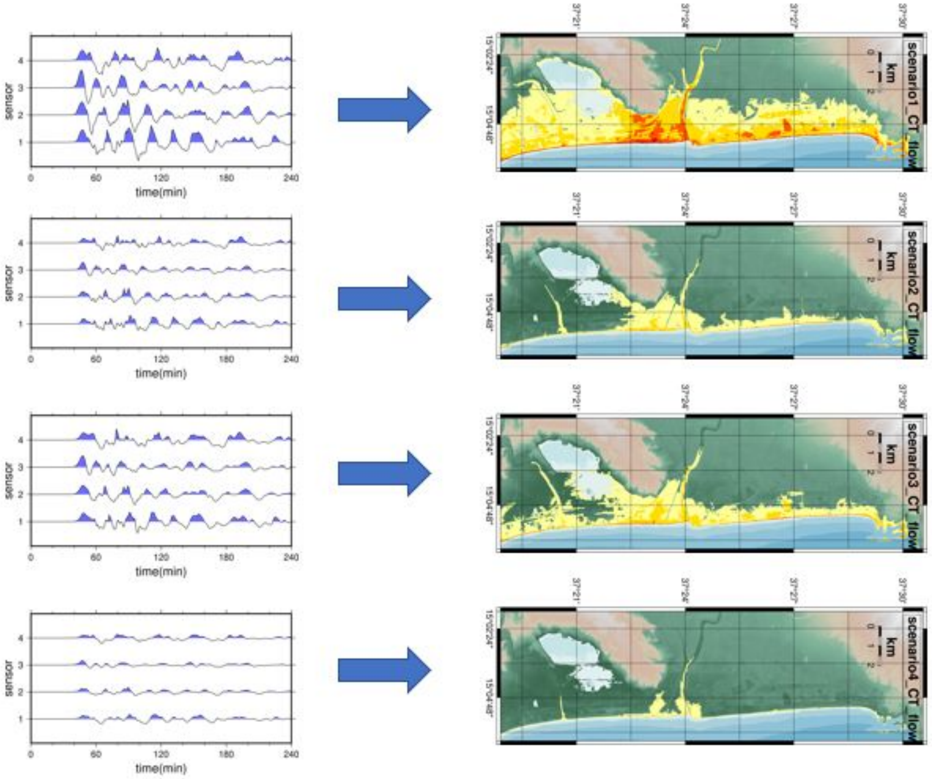
- **Often need vast numbers of simulations (10000+)**
- **Inundation is by far the most expensive part of the calculation.**
- **Offshore time-series are relatively cheap to calculate.**
- **Is it posible to replace the simulation of the inundation by a statistical emulator?**



Introduction

For such a procedure to be feasible we would require:

- Possible with a relatively small training set.
- An acceptable level of accuracy in the predictions.
- Good performance for low-probability, extreme events.
- **Significant improvement in time-to-solution relative to the full numerical simulation.**



Dataset from the ChEESE-1P Project

Probabilistic Tsunami Hazard Analysis performed for Eastern Sicily (Catania and Siracusa):

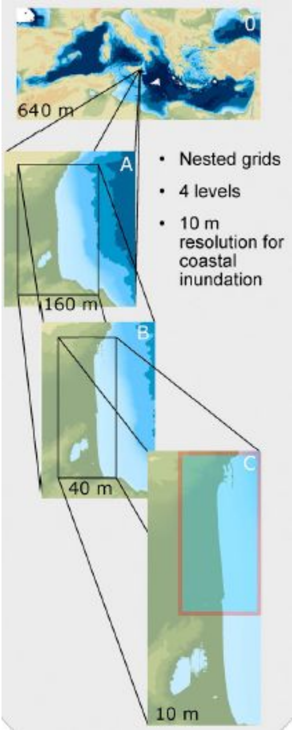
- Around 42,000 initial simulations for these sites.
- 10m resolution inundation grids for coastal regions of interest.
- Offshore time-series saved for all simulations.

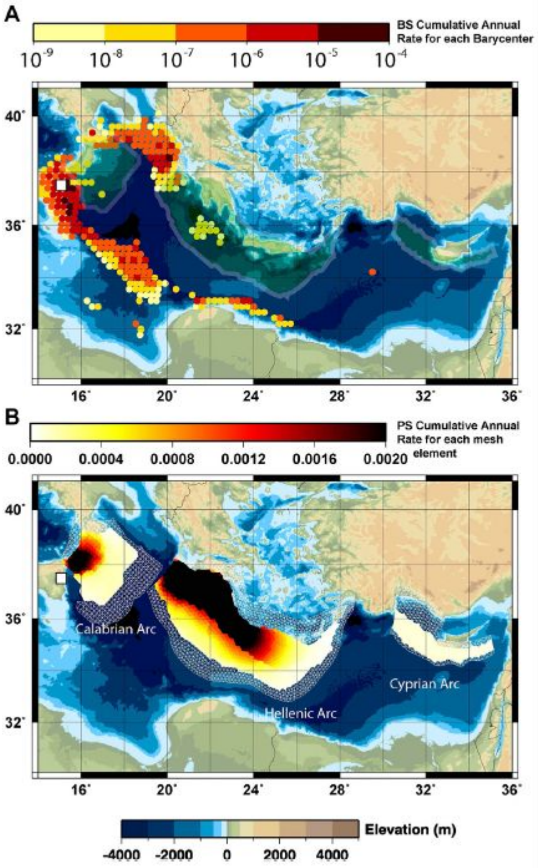
ChEESE

frontiers in Earth Science
ORIGINAL RESEARCH
published: 11 December 2020
doi: 10.3389/feart.2020.591549

Probabilistic Tsunami Hazard Analysis: High Performance Computing for Massive Scale Inundation Simulations

Steven J. Gibbons[1]*, Stefano Lorito[2], Jorge Macías[3], Finn Løvholt[1], Jacopo Selva[4], Manuela Volpe[2], Carlos Sánchez-Linares[3], Andrey Babeyko[5], Beatriz Brizuela[2], Antonella Cirella[2], Manuel J. Castro[3], Marc de la Asunción[3], Piero Lanucara[6], Sylfest Glimsdal[1], Maria Concetta Lorenzino[2], Massimo Nazaria[2], Luca Pizzimenti[2], Fabrizio Romano[2], Antonio Scala[7], Roberto Tonini[2], José Manuel González Vida[3] and Malte Vöge[1]



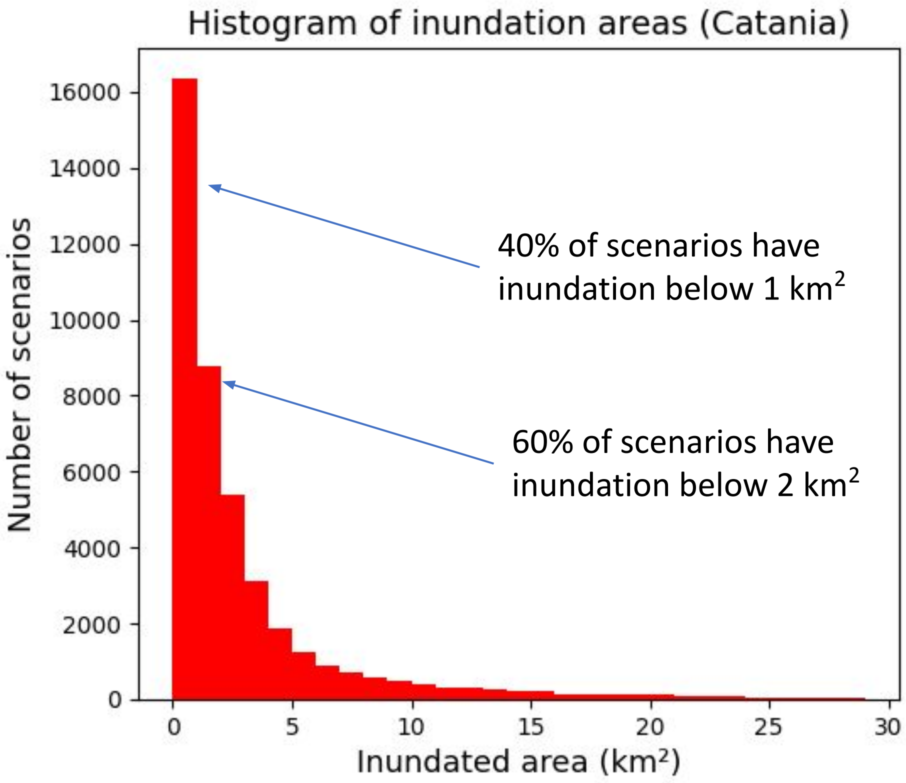


Stage 1 validation: Predicting Inundation Statistics

eFlows4HPC

Before we try to reproduce inundation maps, we test whether or not we can reproduce statistical metrics of the maps.

- **A metric of regional inundation is the total area subject to inundation in the Catania region.**
- **The vast majority of the scenarios inundate less than 3 km^2.**
- **A machine learning model able to predict more severe inundation scenarios will need an appropriate training set.**



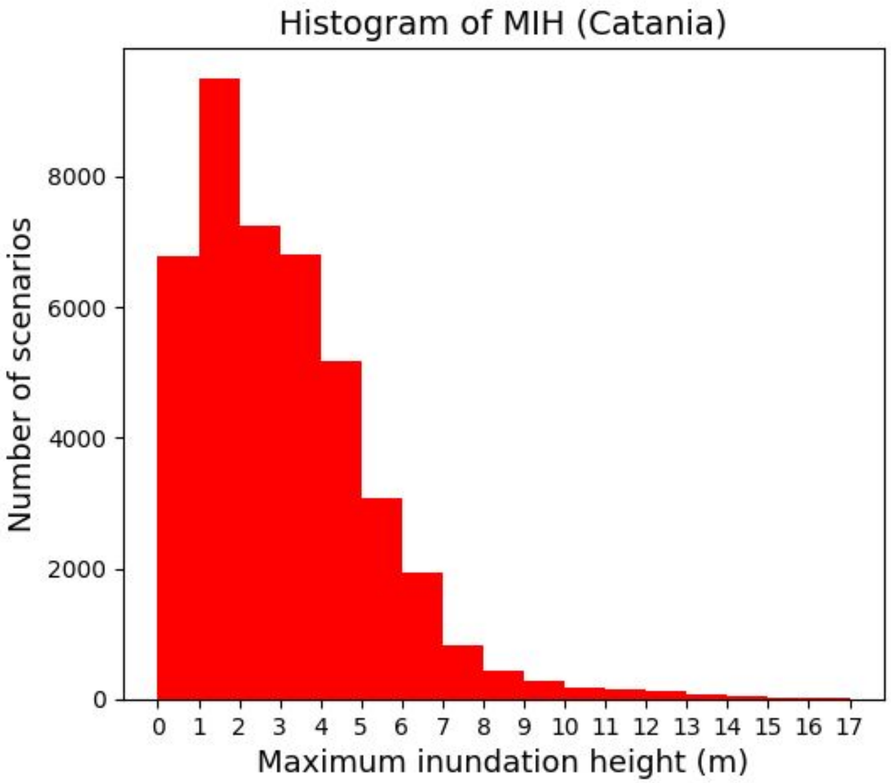
(Inundation area from numerical simulations)

Stage 1 validation: Predicting Inundation Statistics

Before we try to reproduce inundation maps, we test whether or not we can reproduce statistical metrics of the maps.

- **The Maximum Inundation Height (MIH) is the metric that is of interest – ultimately at every location on the map.**
- **Here we display the maximum MIH achieved at any location on the grid in a single simulation.**

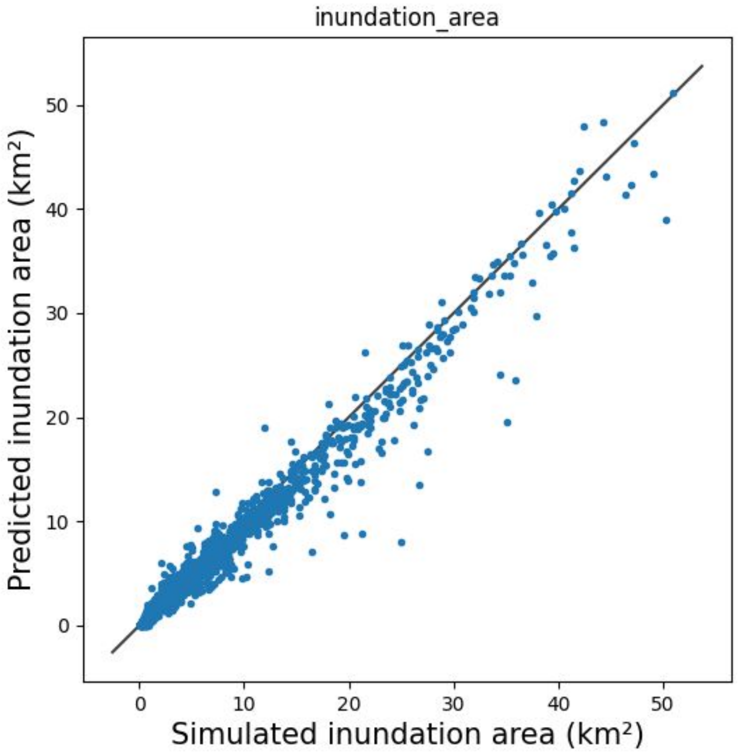
  (The highest MIH values are often very local.)



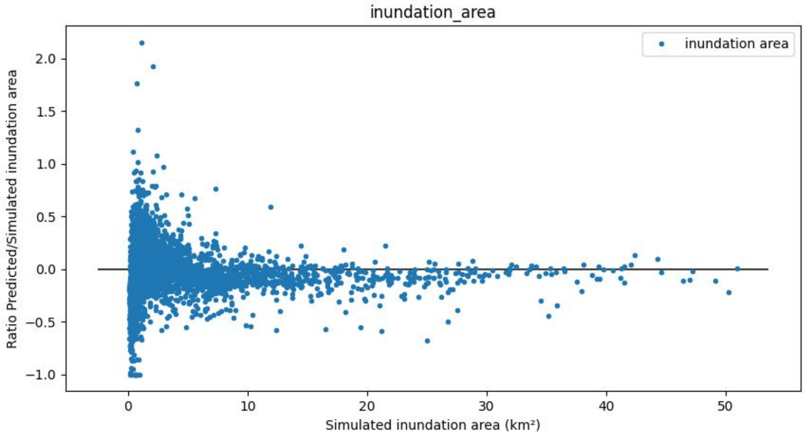
(MIH from numerical simulations)

Stage 1 validation: Predicting Inundation Statistics

(with randomly selected training set)



First test is to reproduce the total inundation area (scalar value) from offshore time-series using a CNN. Results are promising – it looks like we may underestimate the inundation in the more severe cases.



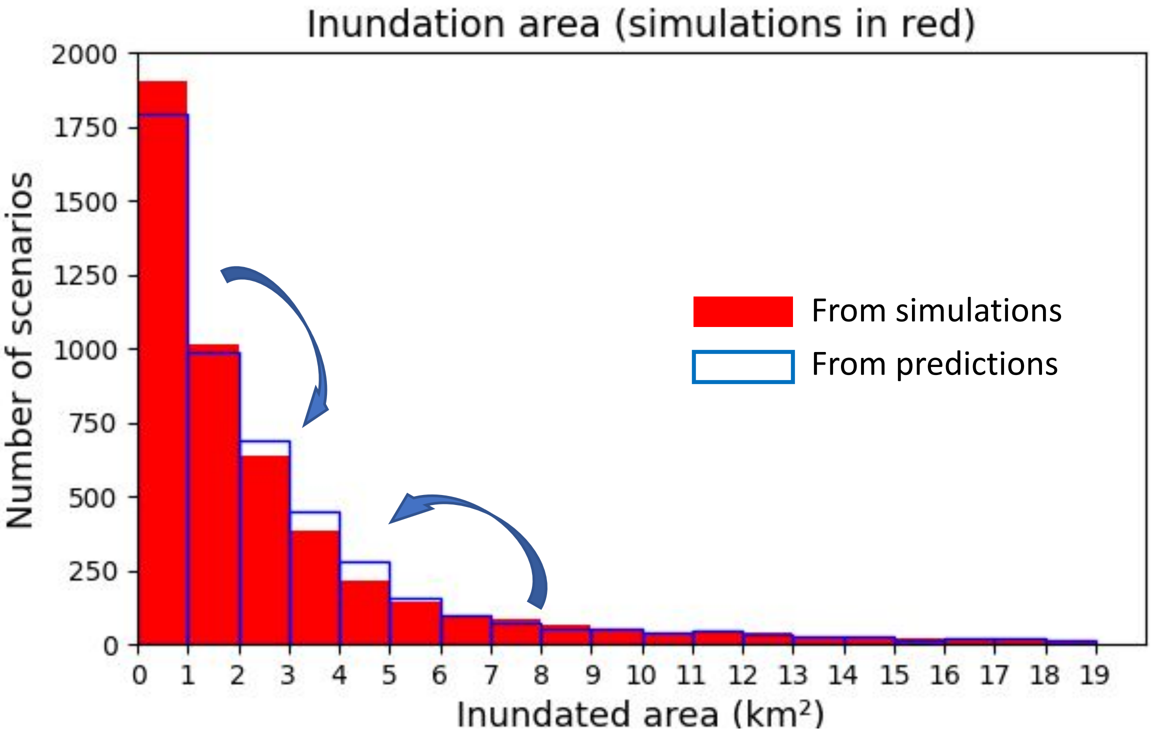
Stage 1 validation: Predicting Inundation Statistics

eFlows4HPC

(with randomly selected training set)

Looking at the statistics of the modelled and predicted inundation areas, we may be doing better than the scatterplot suggests.

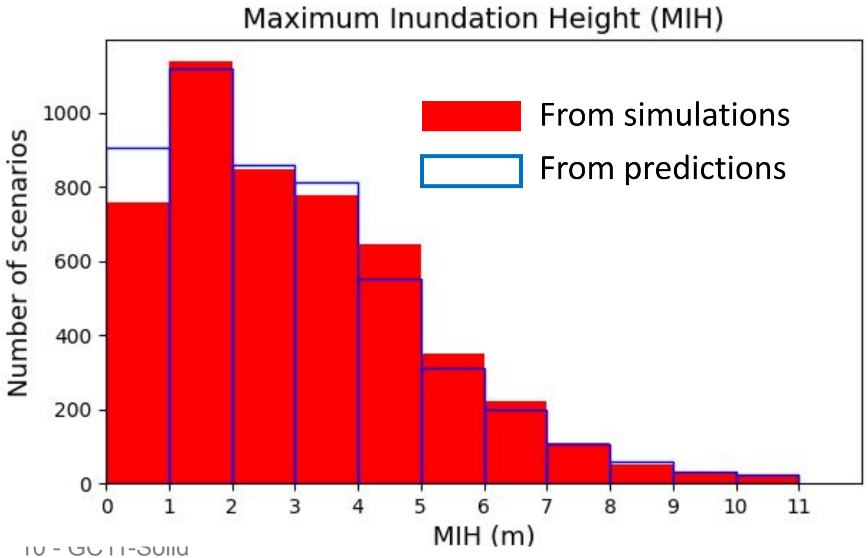
However, we clearly overestimate the inundation for some of the smallest – and underestimate for some of the largest.

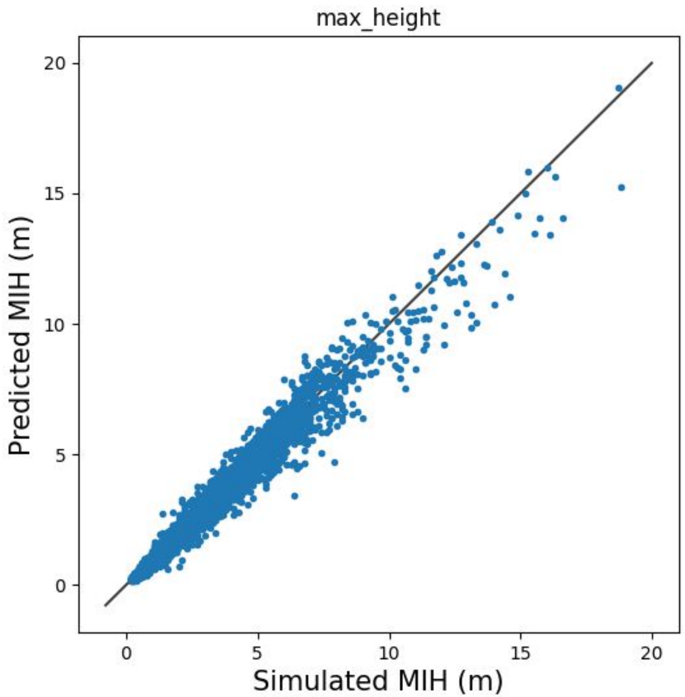


Stage 1 validation: Predicting Inundation Statistics

(with randomly selected training set)

The performance in estimating the (global) maximum inundation height (MIH) is similar. The spread appears greater than for the inundated area.

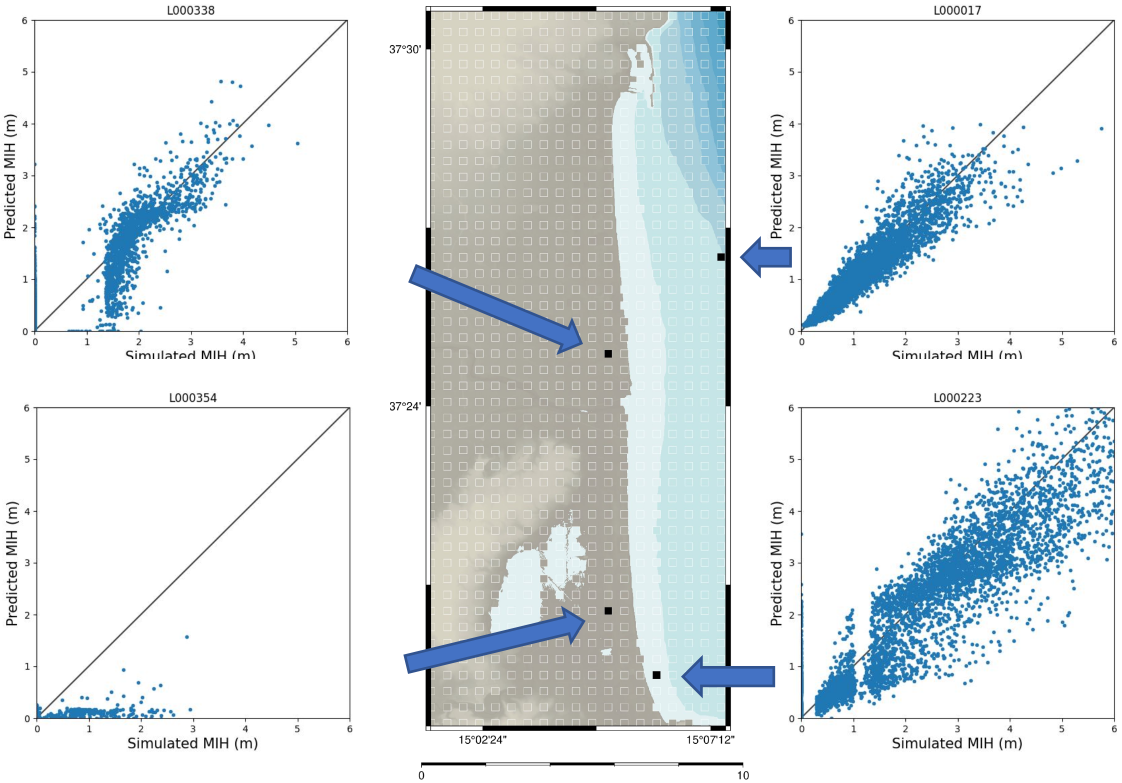




Towards predicting inundation maps

Our simple model (which worked quite well for the global scalar quantities struggles to model the maximum inundation heights at individual sites.

- Performance is especially poor for inland points that are rarely inundated.
- We do not exploit the joint distribution of the targets at the different locations.



So – we need to rethink this!

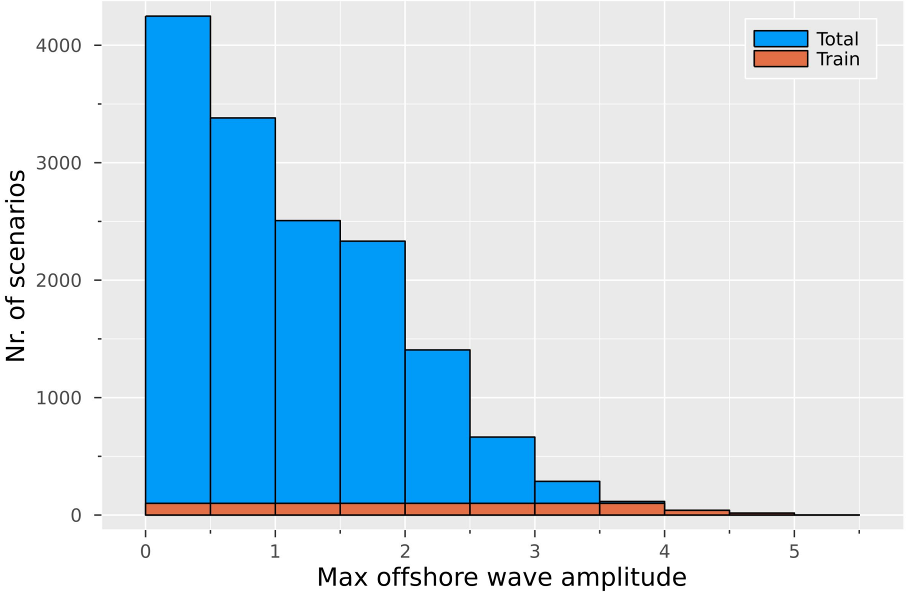
- If we pick training sets randomly we will likely need an unacceptably high number of training scenarios.
- Our inundation parameter space is very large (several million pixels) – or is it?
  (It is always the lowest-lying locations, or those closest to
  the shore that are inundated the most.)

  - There is likely significant correlation between one inundation map and the next. Can we exploit this?
  - Can we cover our parameter space with far fewer parameters?

(1) Choosing an appropriate training set

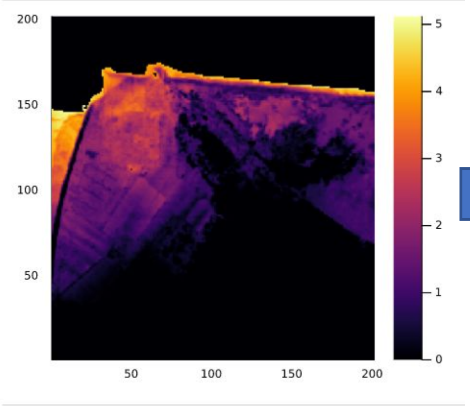
However, it is the statistics of inundation we want to predict – we do not have access to these values when choosing our training set.

- **To the right is the distribution of maximum offshore wave amplitudes (at a single tide-gauge offshore Catania)**
- **Displayed is a possible way of selecting a training set which covers both the more typical (low to moderate inundation) and the extreme cases (high inundation)**

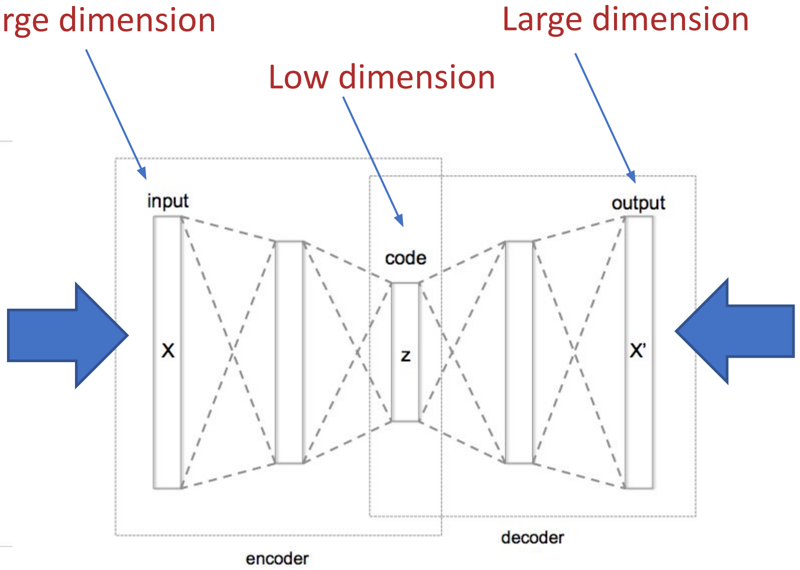


(2) Dimensionality of our inundation maps

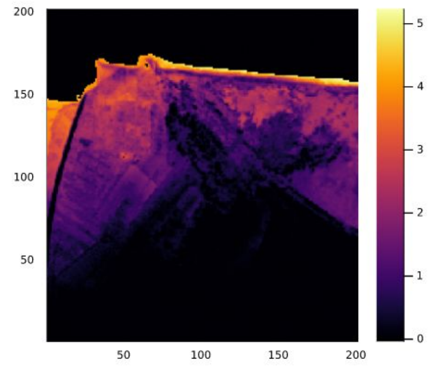
input



Simulated inundation map



output



Reconstructed inundation map

Autoencoders

https://en.wikipedia.org/wiki/Autoencoder

(2) Dimensionality of our inundation maps

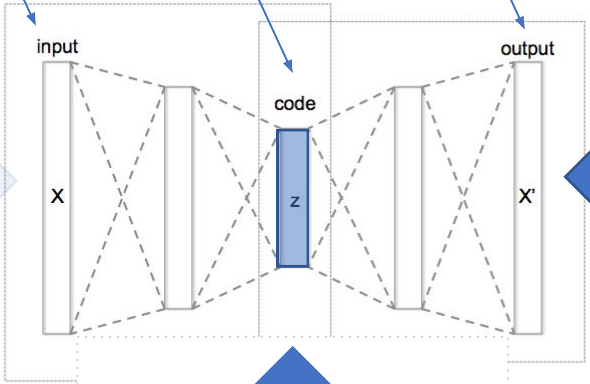
input

Simulated inundation map

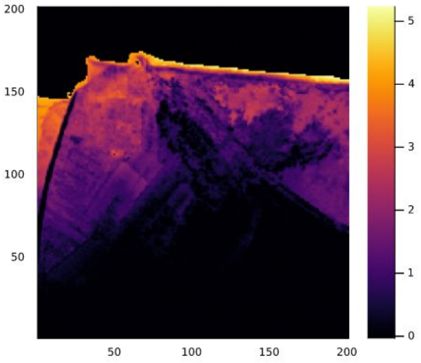
Large dimension

Low dimension

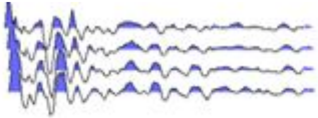
Large dimension



output



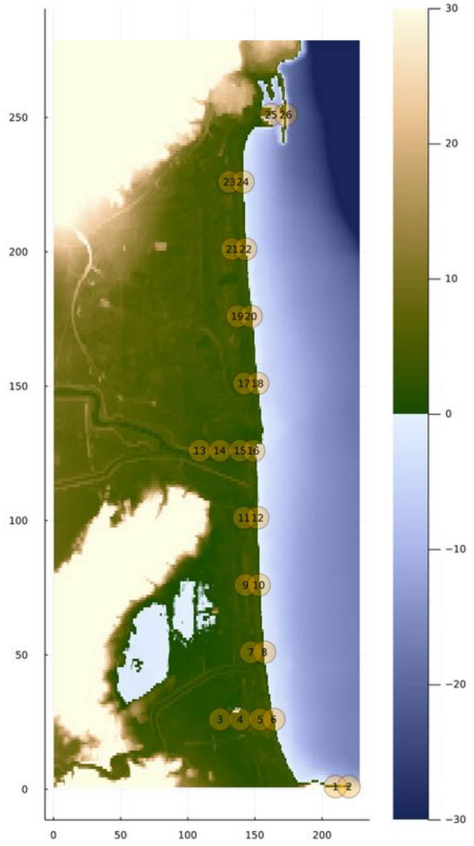
Reconstructed inundation map



Convolutional encoder

https://en.wikipedia.org/wiki/Autoencoder

Prediction of inundation maps



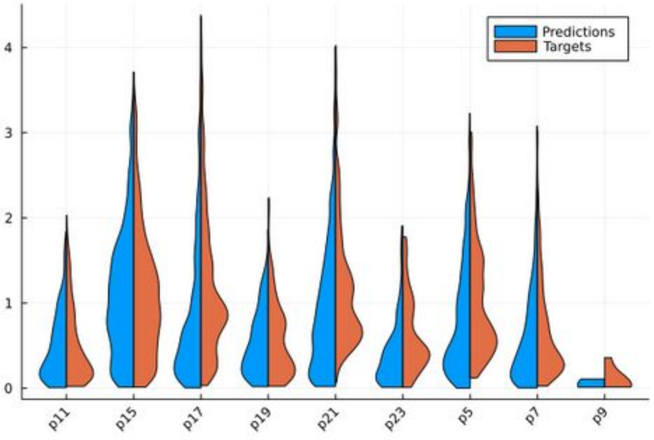
- We test the viability of predicting the inundation using an auto-encoder.
- We want to investigate the size of the latent space needed to be able to accurately reconstruct inundation patterns.
- We want to investigate the size and selection of the training sets on the ability to predict an inundation pattern from offshore time-series.

We be considering the entire grid of points but for ease of display will consider a distinct set of points along the coast at Catania.

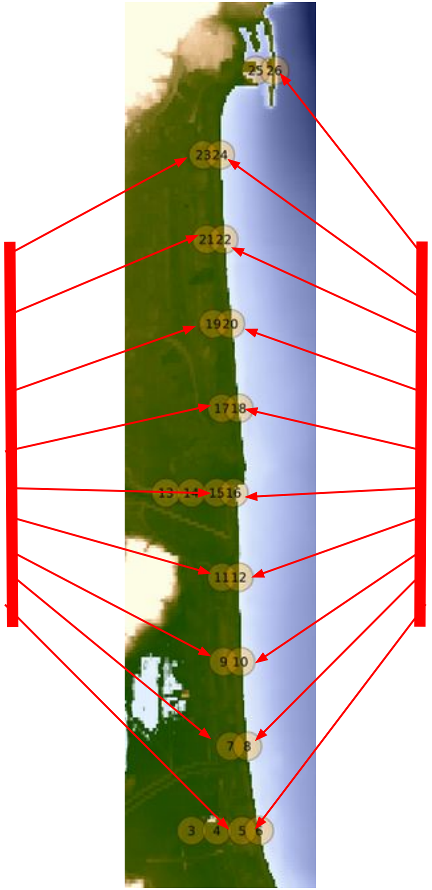
Some locations are right at the shoreline and are essentially always inundated.

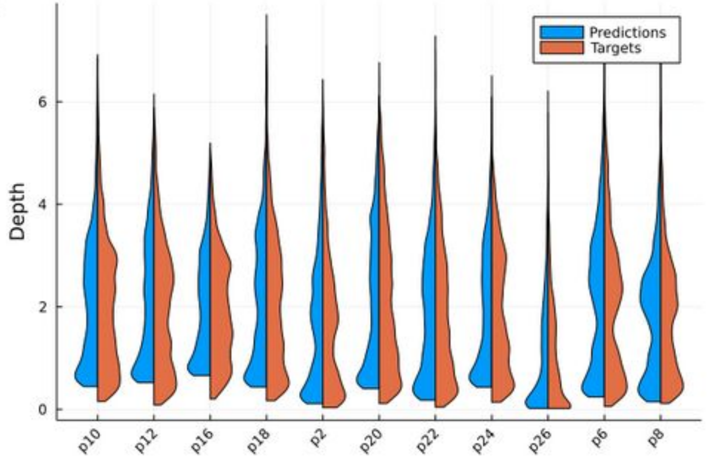
Some locations are further inland and at some height. They are more seldom inundated.

Prediction of inundation maps



(Locations further inland)





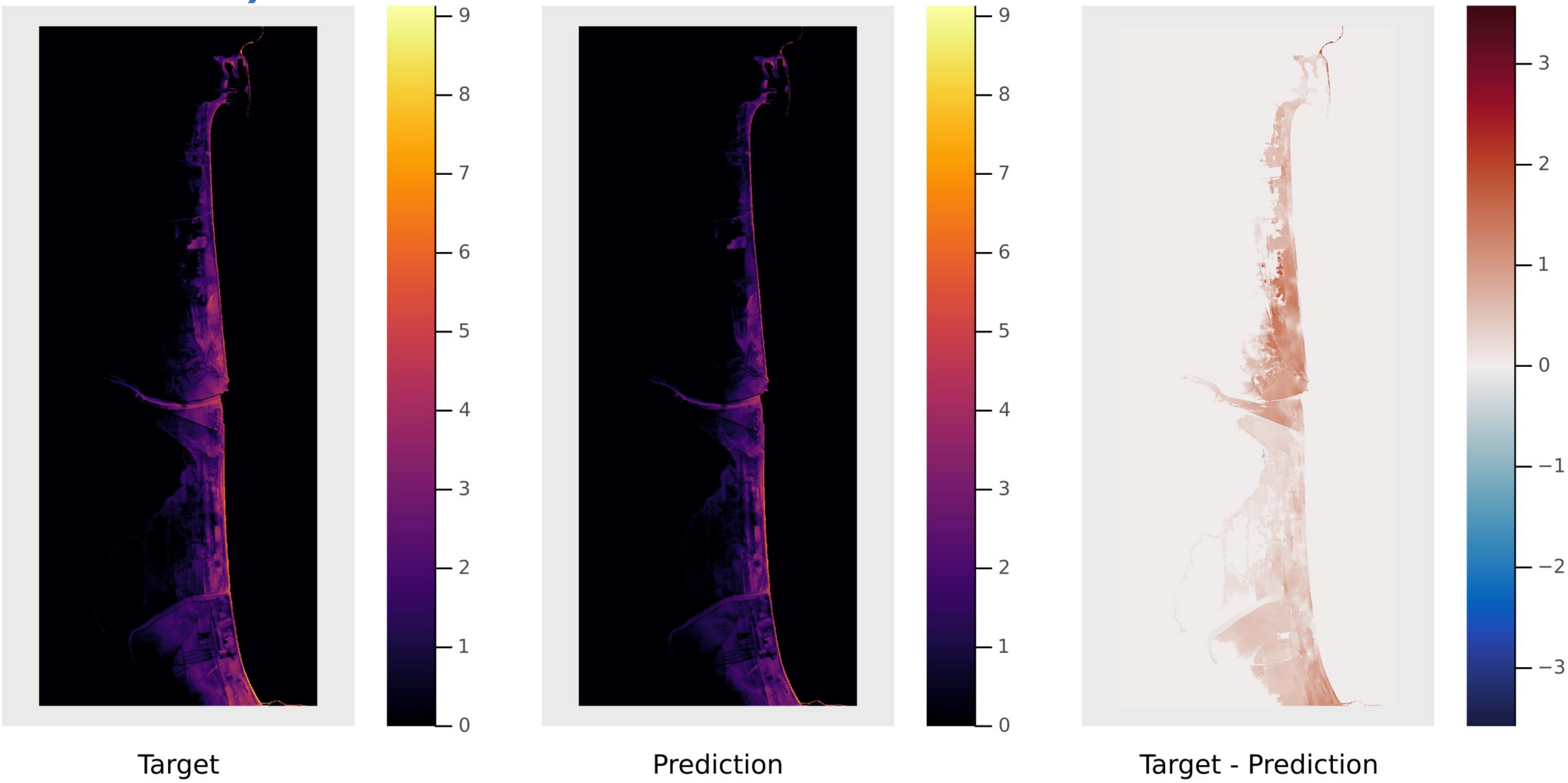
(Locations near the waterfront)

From simulations

From predictions

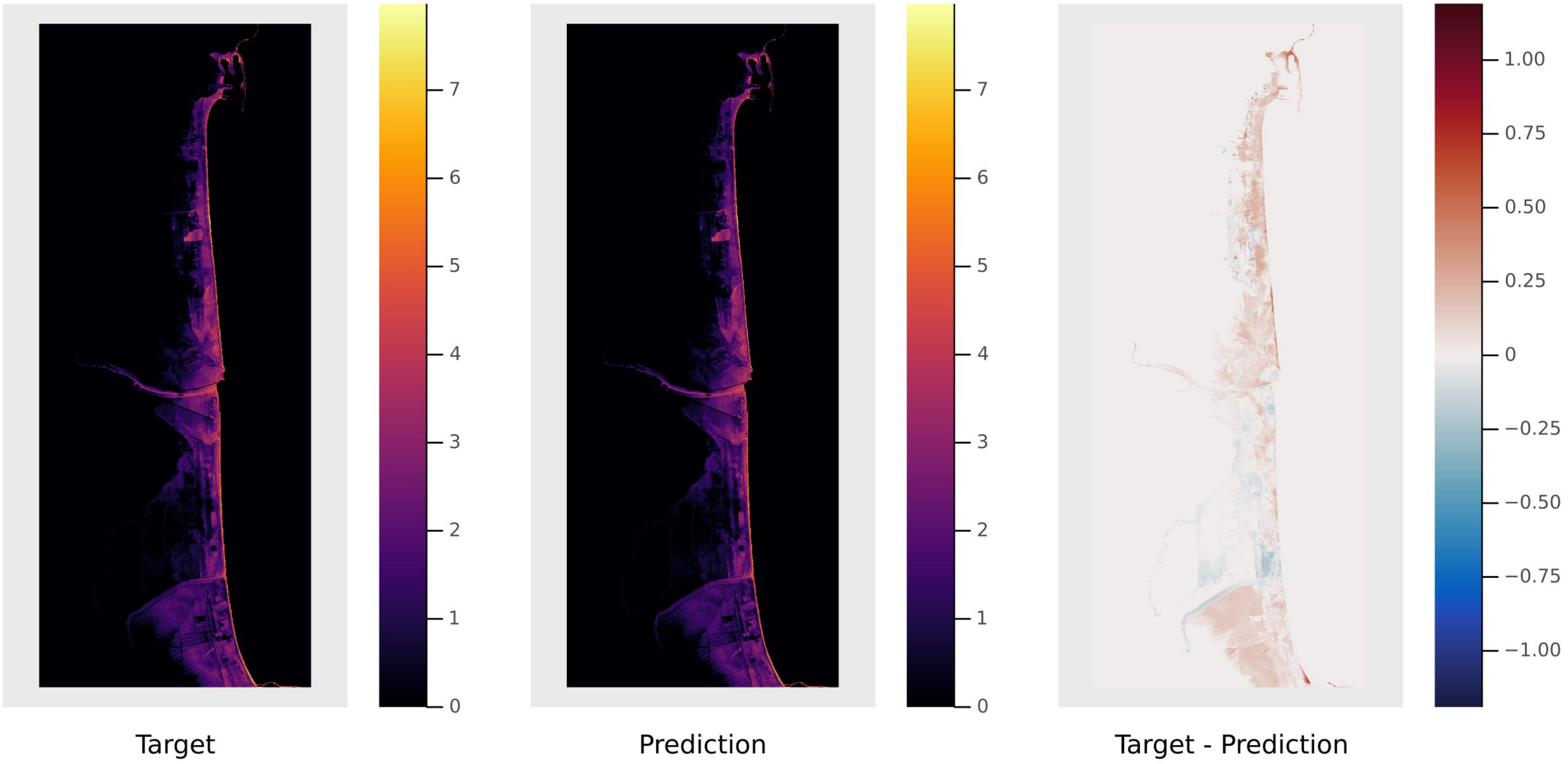
The inundation at the locations further inland is underestimated more often than the inundation at the near-shore locations.

Results: Example 1 – Large inundation (highest error)



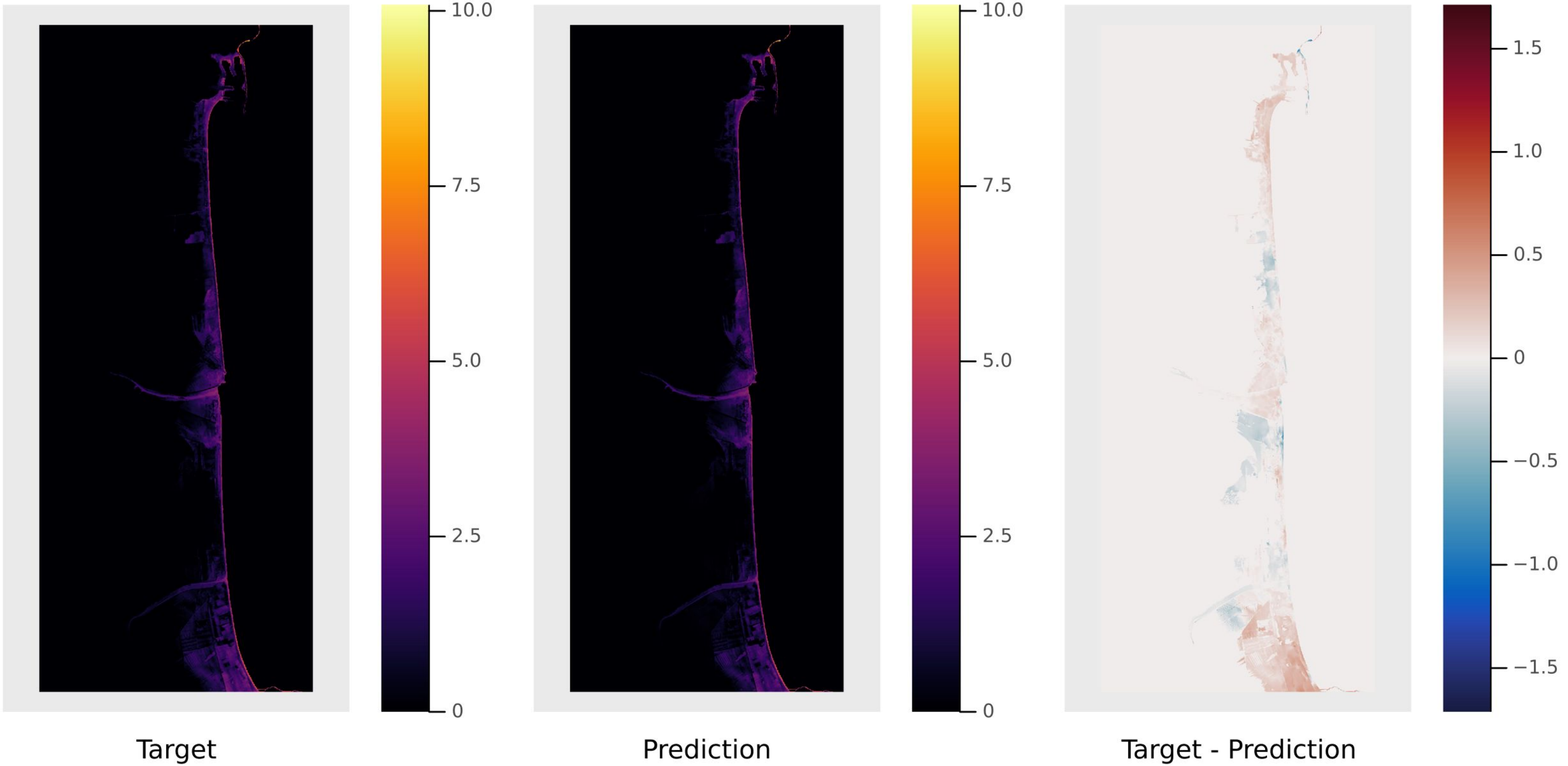
Largest inundation (l^1) on test set (also largest error). 286 training samples.

Results: Example 2 – Large inundation



Second largest inundation (l^1) on test set. 286 training samples.

Results: Example 3 – Error low but spatial pattern not very accurately reproduced.



95% quantile (l^1) on test set. 286 training samples.

Conclusions

- A convolutional network acting on offshore time-series is able to predict tsunami inundation statistics and maps.
- Predictions for the coast of Eastern Sicily (Bay of Catania) were satisfactory with training sets with several hundred scenarios. (As opposed to many thousands needed for PTHA or PTF)
- An Autoencoder allows us to represent the high-dimensional inundation maps with a far lower dimensional latent space. (Significant work remaining on determining optimal dimensions)
- Training set must have sufficient size (to avoid overfitting) and be sufficiently diverse (to cover higher inundations).

Thank you!

eFlows4HPC

Enabling dynamic and Intelligent workflows in the future EuroHPC ecosystem

www.eFlows4HPC.eu

@eFlows4HPC eFlows4HPC Project

This project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 955558. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Spain, Germany, France, Italy, Poland, Switzerland, Norway.