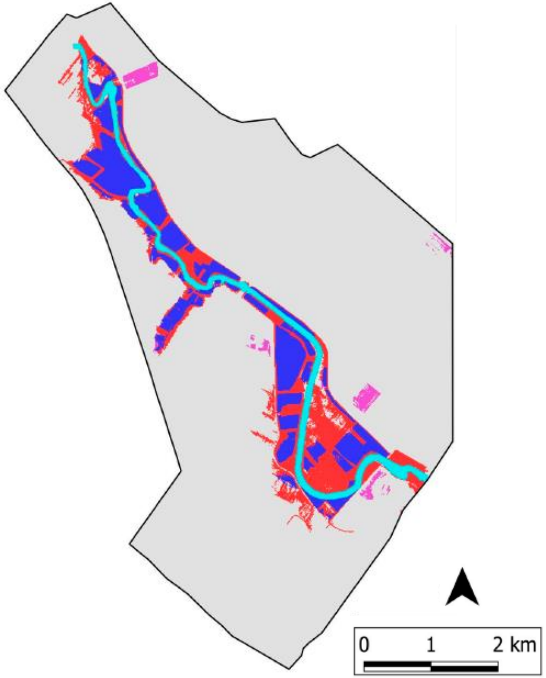
Can performance metrics accounting for the flood extent shape improve inundation model calibration?

Thaine H. Assumpção, Ioana Popescu, Andreja Jonoski, Dimitri P. Solomatine

EGU General Assembly 2020

Hydrodynamic flood models

current metrics



	Observed flood	No observed flood
Simulated flood	A	B
No simulated flood	C	D

A confusion matrix (contingency table) is created

Image source: Scarpino, S.; Albano, R.; Cantisani, A.; Mancusi, L.; Sole, A.; Milillo, G. (2018). Multitemporal SAR Data and 2D Hydrodynamic Model Flood Scenario Dynamics Assessment. *ISPRS Int. J. Geo-Inf.*, *7*, 105.

Hydrodynamic flood models
current metrics

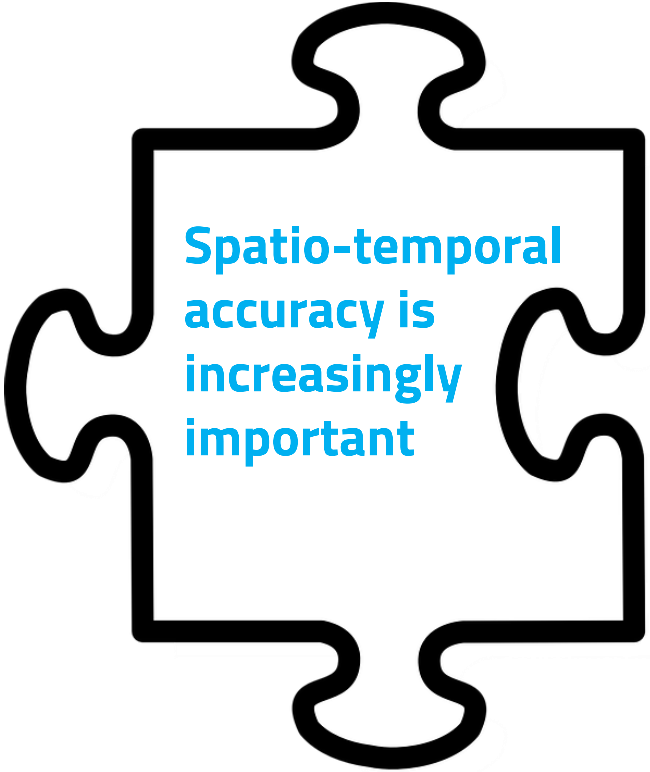
Name	Equation
Bias	$\frac{A+B}{A+C}$
Proportion Correct (PC)	$\frac{A+D}{A+B+C+D}$
Critical Success Index (CSI) or Threat Score ($F^{<2>}$)	$\frac{A}{A+B+C}$
$F^{<3>}$	$\frac{A-C}{A+B+C}$
$F^{<4>}$	$\frac{A-B}{A+B+C}$
Hit rate (H)	$\frac{A}{A+C}$
False alarm rate (F)	$\frac{B}{B+D}$
Pierce Skill Score (PSS)	$H-F$

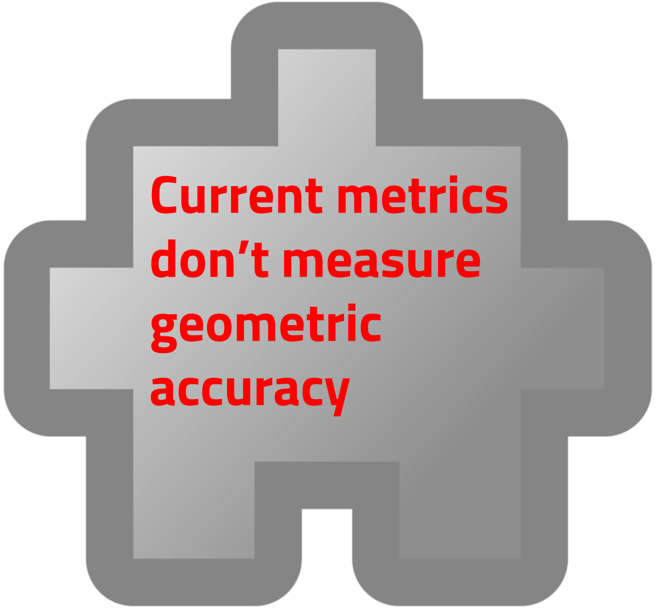
	Observed flood	No observed flood
Simulated flood	A	B
No simulated flood	C	D

Metrics are calculated over the confusion matrix

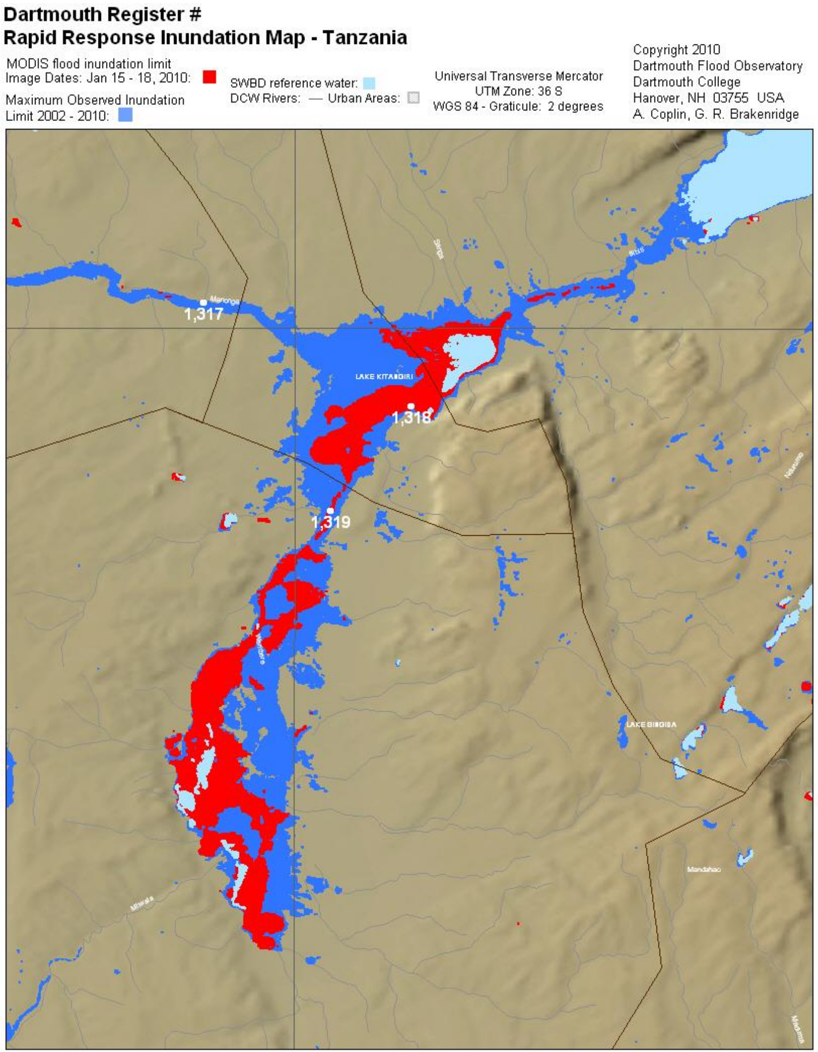
Metrics measure overlap accuracy

What is the problem?

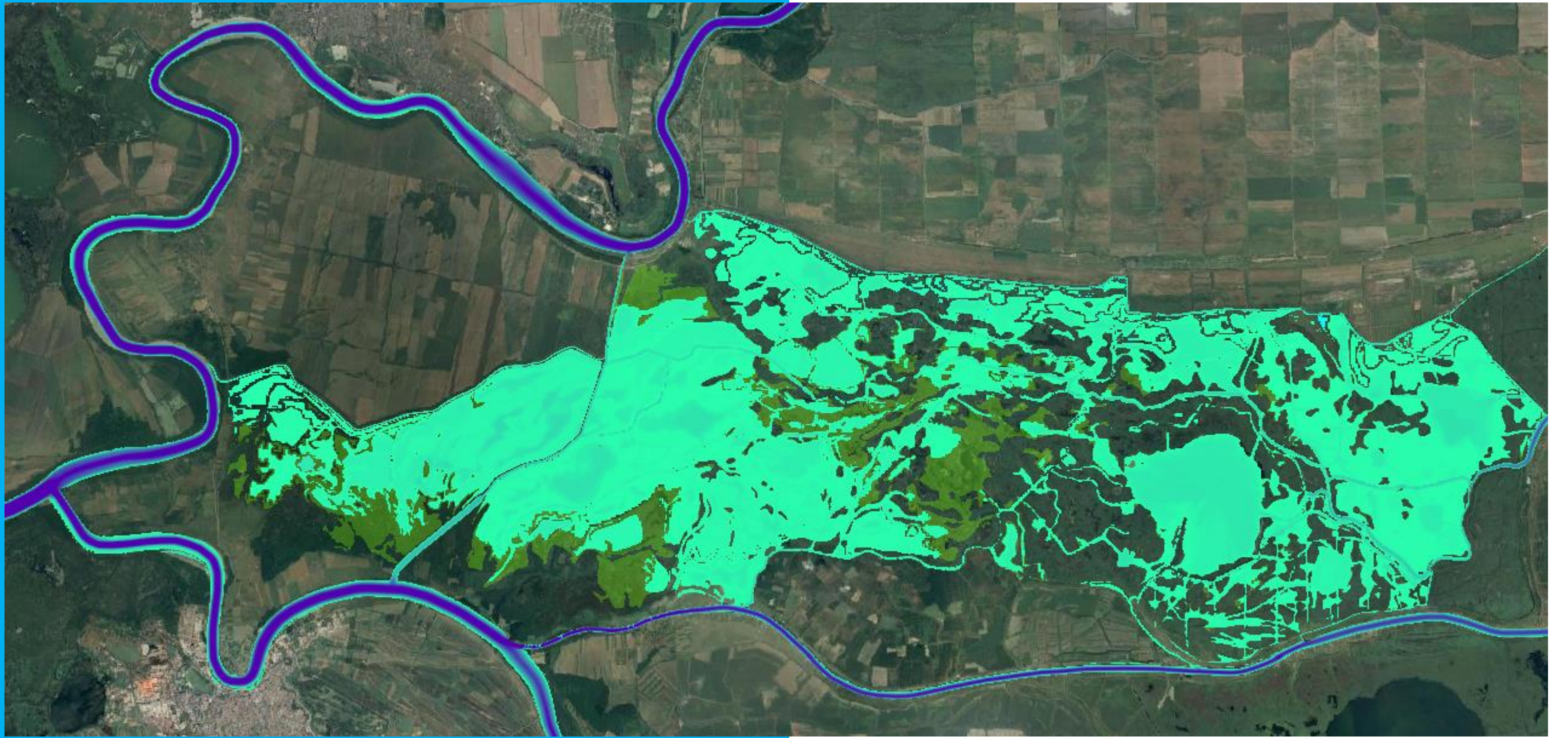




Comparing shapes visually is simple …

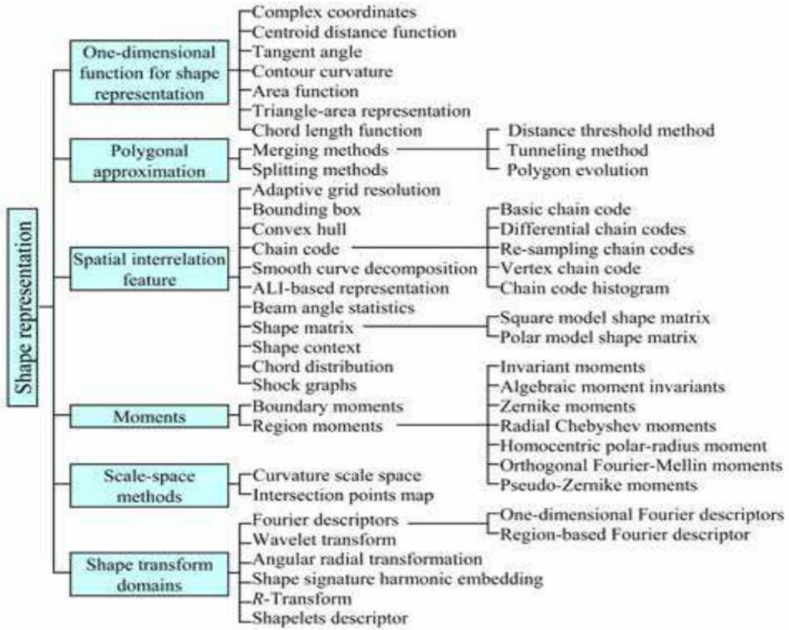


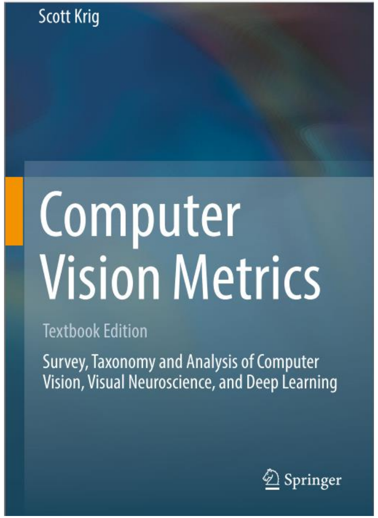
... until it isn't



Sontea-Fortuna, Danube Delta, Romania

Possible solution
look at the flood extent as a shape





Mingqiang, Y., Kidiyo, K., & Joseph, R. (2008). *A Survey of Shape Feature Extraction Techniques. Pattern Recognition Techniques, Technology and Applications*. https://doi.org/10.5772/6237

MAIN OBJECTIVES

To test if traditional flood extent performance metrics are able to capture differences in shape; if shape-based metrics are an alternative

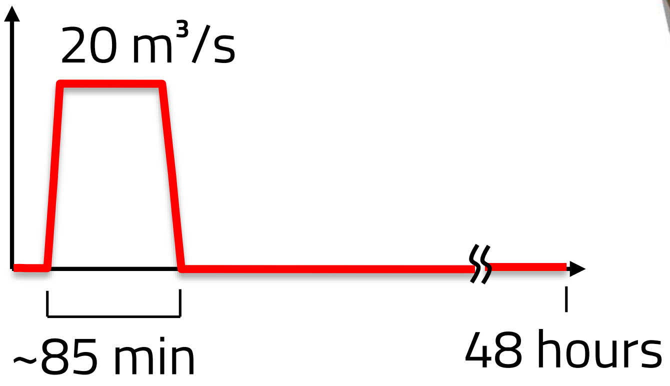
Some tested metrics

Traditional metrics
Bias
Proportion Correct (PC)
Critical Success Index (CSI) or Threat Score ($F^{<2>}$)
$F^{<3>}$
$F^{<4>}$
Hit rate (H)
False alarm rate (F)
Pierce Skill Score (PSS)

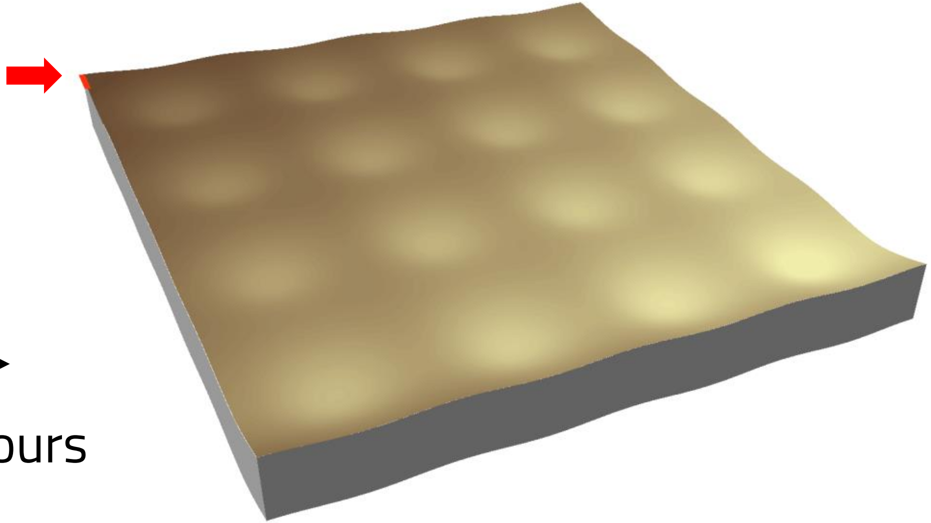
Shape-based metrics	
Centroid (difference)	Shape descriptors
Eccentricity (difference)	Shape descriptors
Solidity (difference)	Shape descriptors
Hausdorff Distance	Shape dissimilarity
Modified Hausdorff Distance	Shape dissimilarity

Experiment #1a - evaluation

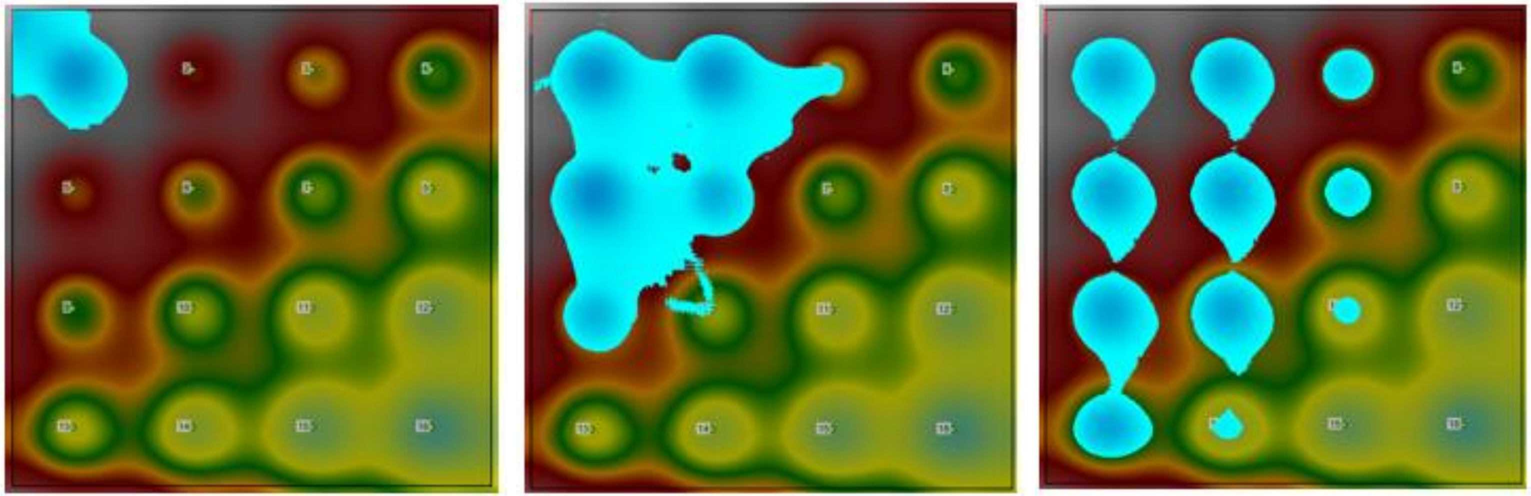
benchmark case study (EPA-UK experiment)



Néelz, S., & Pender, G. (2013). Benchmarking the Latest Generation of 2D Hydraulic Flood Modelling Packages. Bristol: Environment Agency Bristol.



t =30 min | t = 2 hours | t =12 hours



n=0.03 ➡ observed

Evolution of inundation extent and depth with time, where the Manning value of 0.03 is considered the observed value

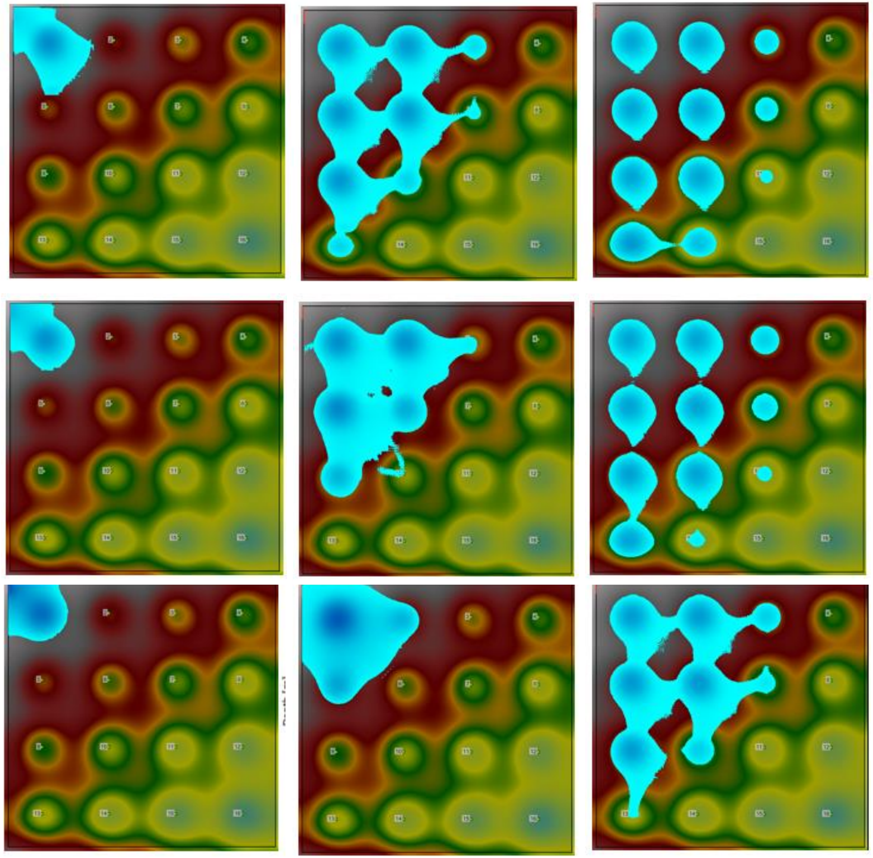
To get different flood extents, the Manning coefficient was varied from n=0.01 to n=0.16

t =30 min | t = 2 hours | t =12 hours

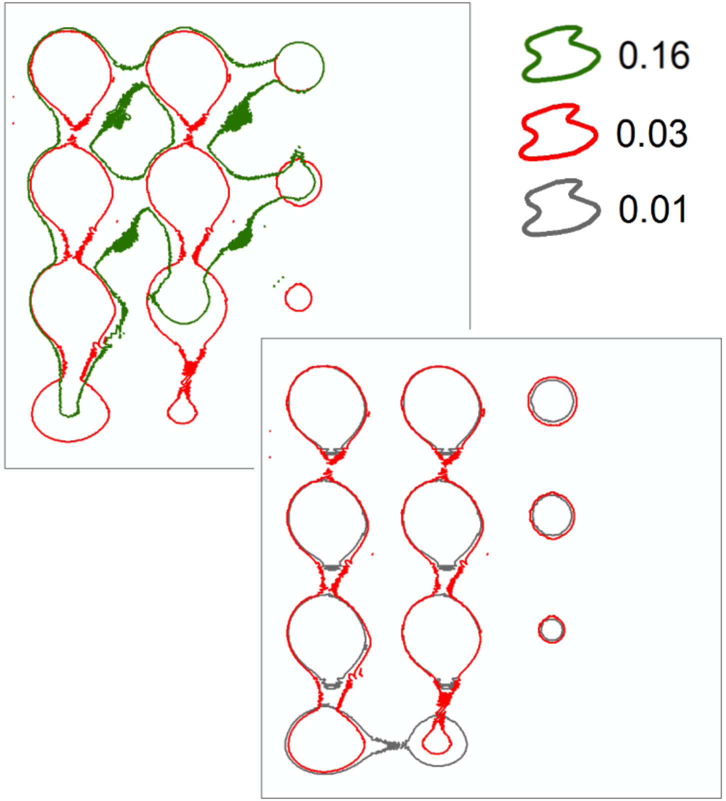
(simulated) n =0.01

(observed) n = 0.03

(simulated) n =0.16



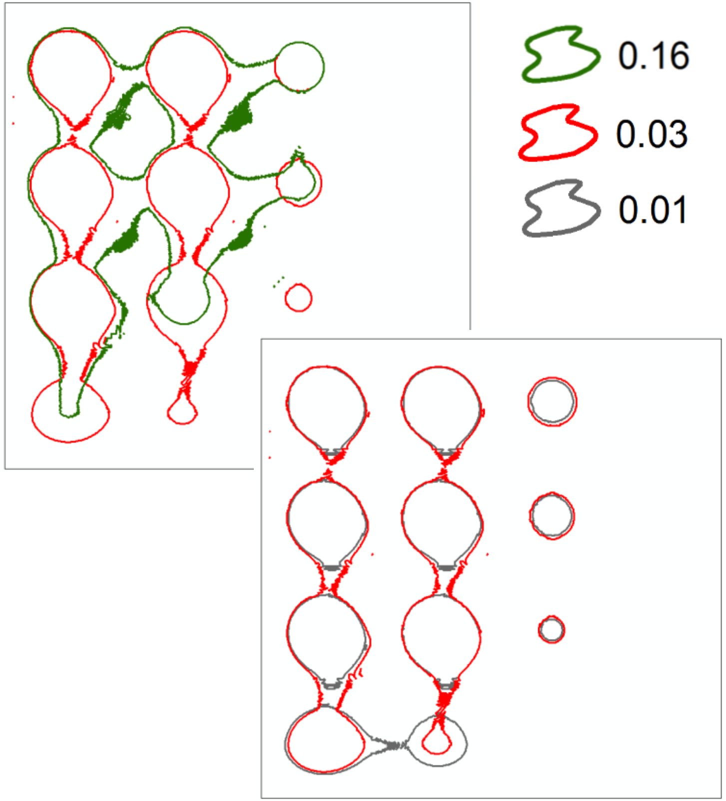
t = 12 hours



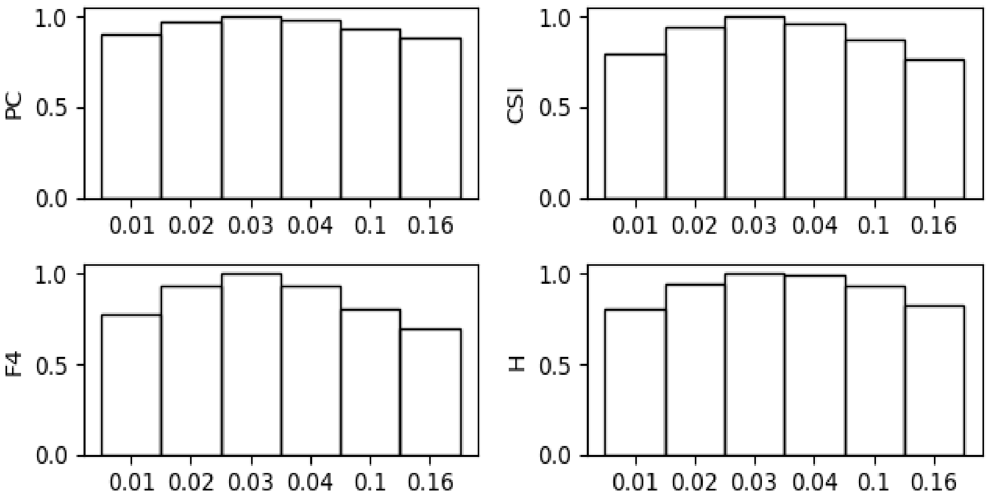
Visual inspection was used as the benchmark to assess if metrics capture or not similarity in shape

In this case:
The shape generated with Manning **0.01** is **more similar to** the observed (**0.03**) than the shape generated with Manning **0.16**

0.01 > 0.16



t = 12 hours

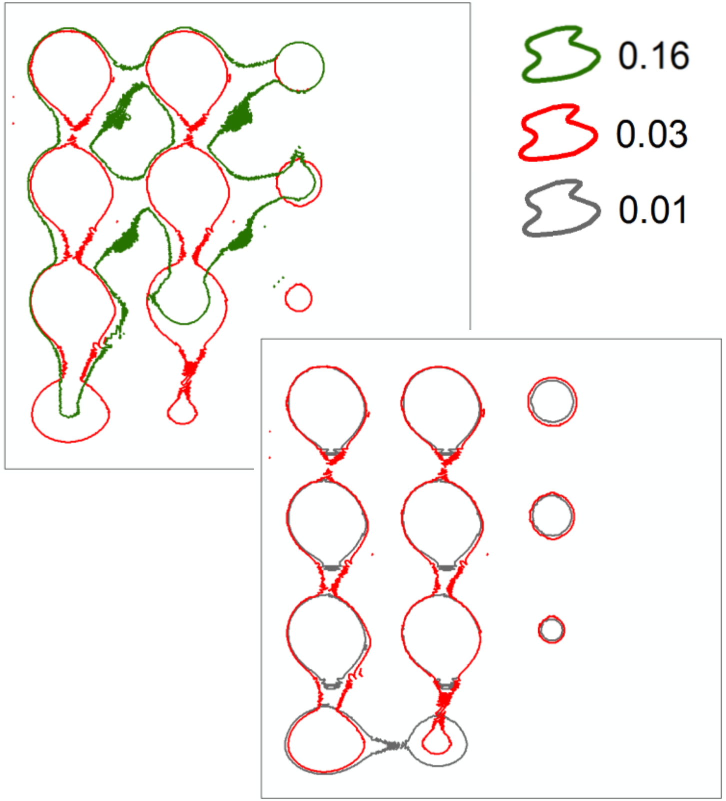


Bias	PC	CSI
F3	**F4**	H
F	**PSS**	

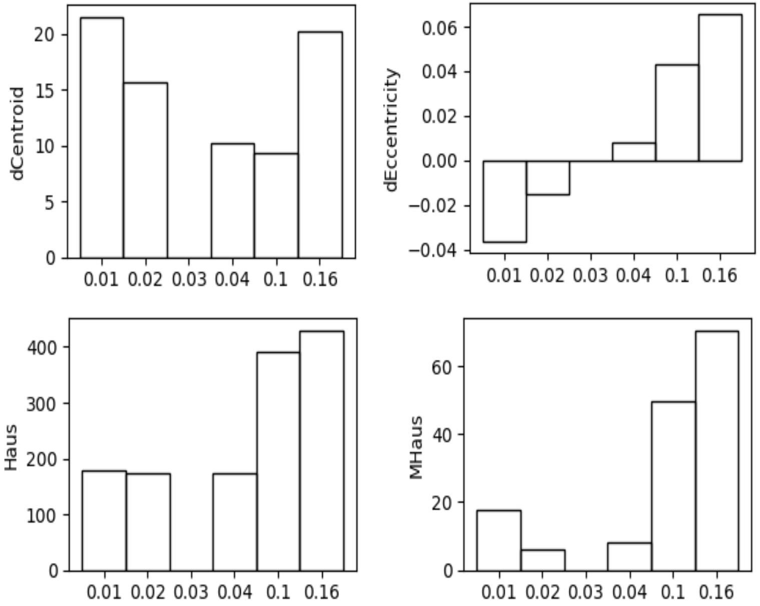
captured
not captured

0.01 > 0.16

For many traditional metrics, similarity is not captured



t = 12 hours



dCentroid	dEccentricity	dSolidity
Haus	MHaus	

captured
not captured

0.01 > 0.16

For many shape-based metrics, similarity is captured

Experiment #1b – calibration

benchmark case study (EPA-UK experiment)

Experiment #1b - calibration
setup

Algorithm

- Differential Evolution
- 15 members initial population
- Maximum of 5 generations

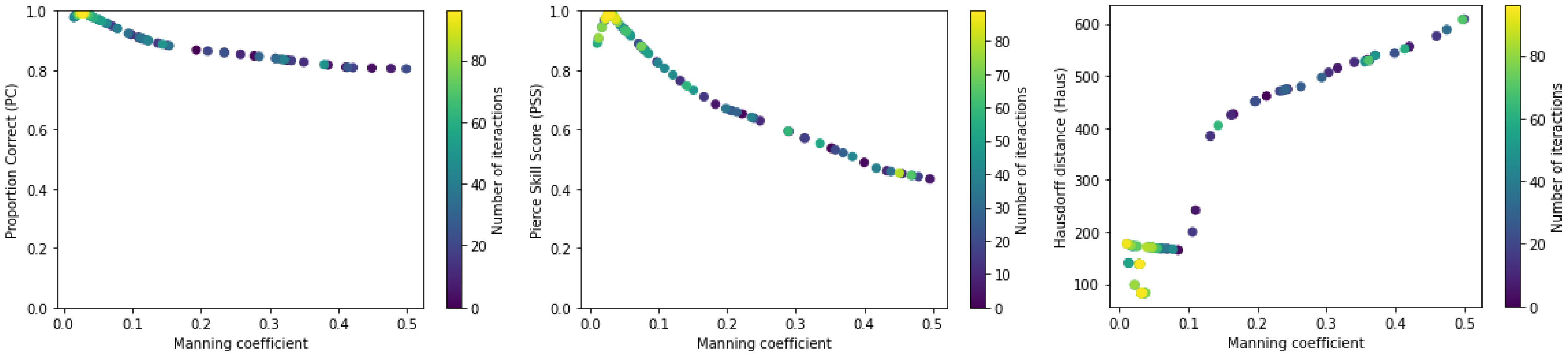
Bounds

- Manning coefficient from 0.01 to 0.5

Tested objective functions

- Proportion Correct (PC): Widely used in the literature
- Pierce Skill Score (PSS): Among best performing traditional metric in previous tests and used in the literature
- Hausdorff distance (Haus): Among best performing shape metrics in previous tests and computationally efficient

Experiment #1b - calibration
results

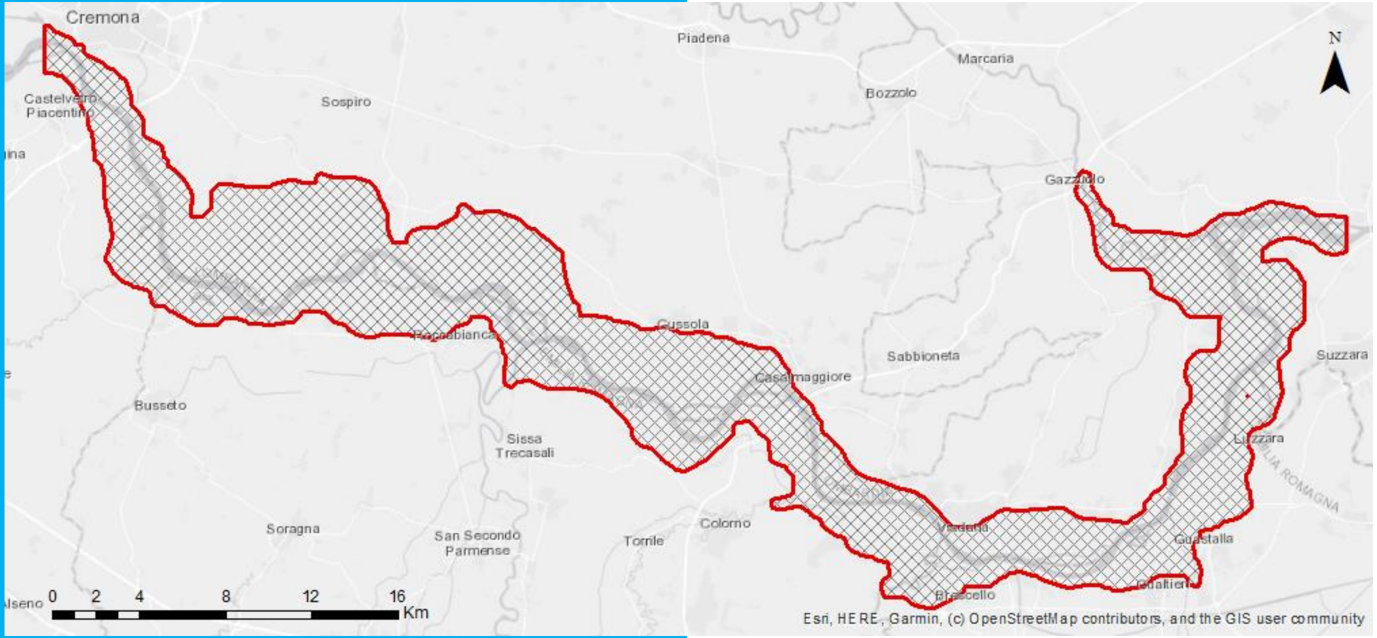


	PC [-]	PSS [-]	Haus [m]
Optimal Manning	0.0276	0.0281	0.0310
Objective function	0.992	0.990	82.92

Not conclusive results: none has reached the a solution;
PSS seems to have the most favorable landscape

Experiment #2 - calibration
Po River

- Italy (north)
- Between the stream gauges of Cremona and Borgoforte
- 98 km



Experiment #2 Po River
original setup

Developed by Tarek Hamouda (2018)

Model
- HEC-RAS (5.0.3)
- Fully 2D model with breaklines
- 2m LIDAR DEM (Po river Basin Authority)
- 90m computational grid

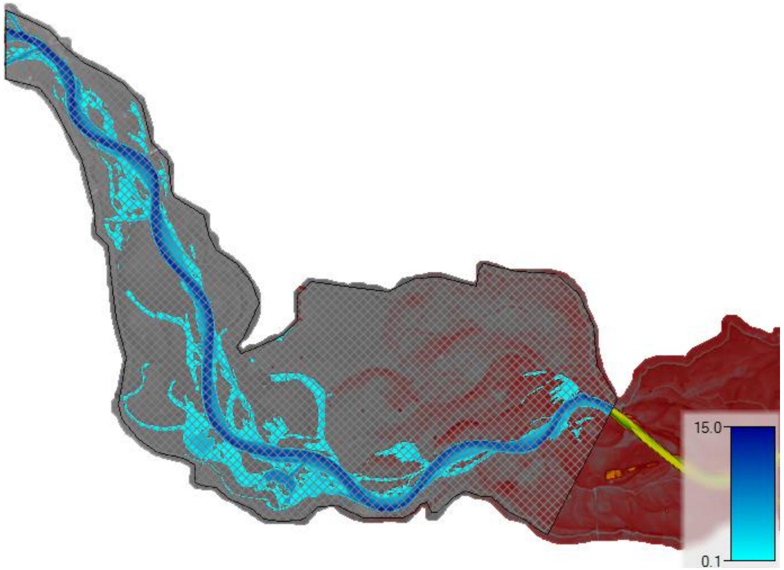
Calibration
- Water levels for a 60-year flood event in 2000

Hamouda, T. (2018). Impact of micro-topography and bathymetry modification on inundation modelling with different magnitudes based on SRTM data. Master Thesis Dissertation. UNESCO-IHE. Delft. The Netherlands.

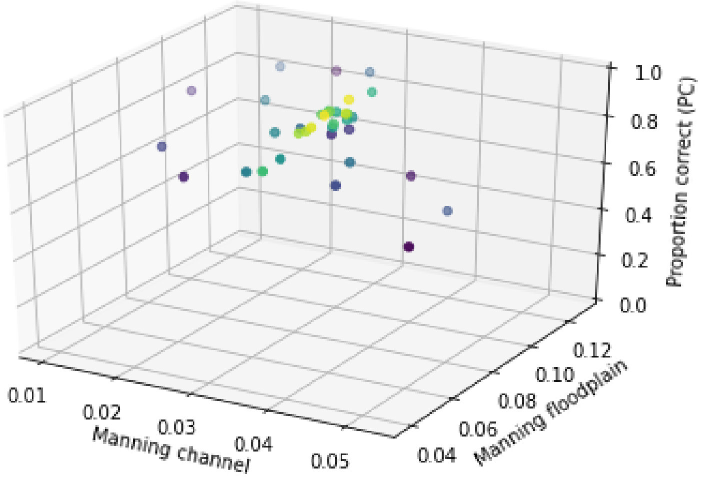
Experiment #2 Po River
this work setup

Model
- Shorted version (HEC-RAS 5.0.7)
- 2-year return period (peak at ~5400 m^3/s)
- Observed value: calibrated flood extent
  - Manning channel: 0.032
  - Manning floodplain: 0.08

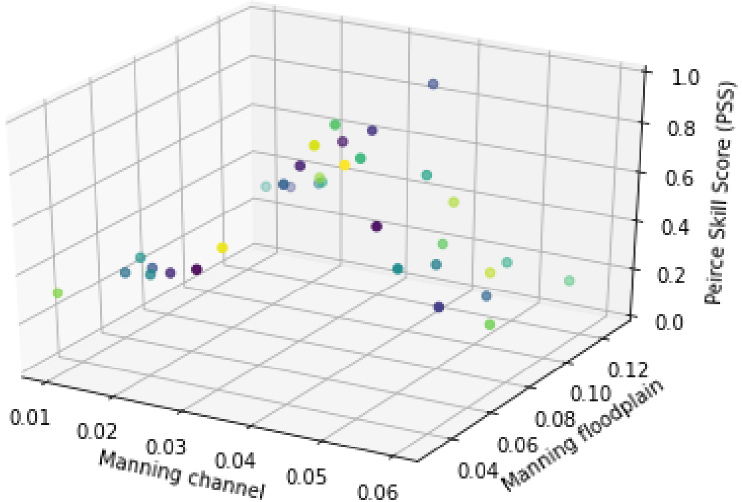
Optimization
- Objective function: PC, PSS and Haus
- Ranges
  - Manning channel: 0.01 - 0.06
  - Manning floodplain: 0.03 – 0.13



Experiment #2 Po River
results

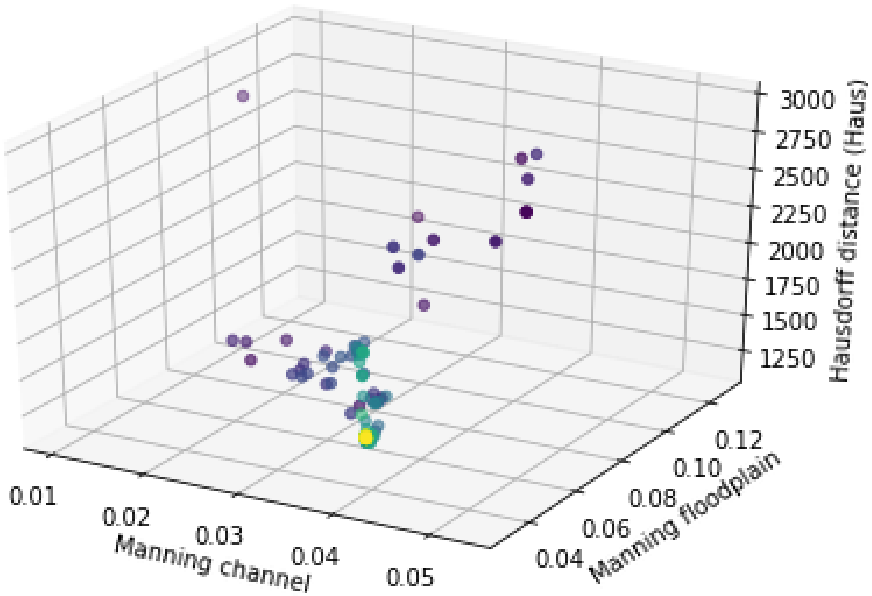


PC has optimal parameters very close to the observed ones, although the landscape of PSS still seems more favorable



	PC [-]	PSS [-]
Optimal n_{ch}	0.0349	0.0360
Optimal n_{fp}	0.0883	0.0778
Objective function	0.962	0.905

Experiment #2 Po River
results



Results were worse for this metric, mainly in terms of the Manning floodplain coefficient

	Haus [m]
Optimal Manning channel	0.0353
Optimal Manning floodplain	0.0557
Objective function	1154.61

Conclusions

- Traditional metrics can fail to distinguish shapes
- Shape-based metrics can be used for that
  - **To predict spatio-temporal variations, it is a good idea to evaluate spatio-temporal variations (add to our diagnostics toolbox)**

- Can we better calibrate inundation models? Maybe not yet

⇨ This research is in progress, we need to:
  - Test more metrics
  - Improve optimization
  - Test with real shape data (flood extent from remote sensing)

References

Hamouda, T. (2018). Impact of micro-topography and bathymetry modification on inundation modelling with different magnitudes based on SRTM data. Master Thesis Dissertation. UNESCO-IHE. Delft. The Netherlands.

Mingqiang, Y., Kidiyo, K., & Joseph, R. (2008). A Survey of Shape Feature Extraction Techniques. Pattern Recognition Techniques, Technology and Applications. https://doi.org/10.5772/6237

Néelz, S., & Pender, G. (2013) Benchmarking the Latest Generation of 2D Hydraulic Flood Modelling Packages. Bristol: Environment Agency Bristol.

Scarpino, S.; Albano, R.; Cantisani, A.; Mancusi, L.; Sole, A.; Milillo, G. (2018) Multitemporal SAR Data and 2D Hydrodynamic Model Flood Scenario Dynamics Assessment. *ISPRS Int. J. Geo-Inf.*, *7*, 105.

Credits

Thanks!

Any questions?

You can find me at
t.hermanassumpcao@un-ihe.org