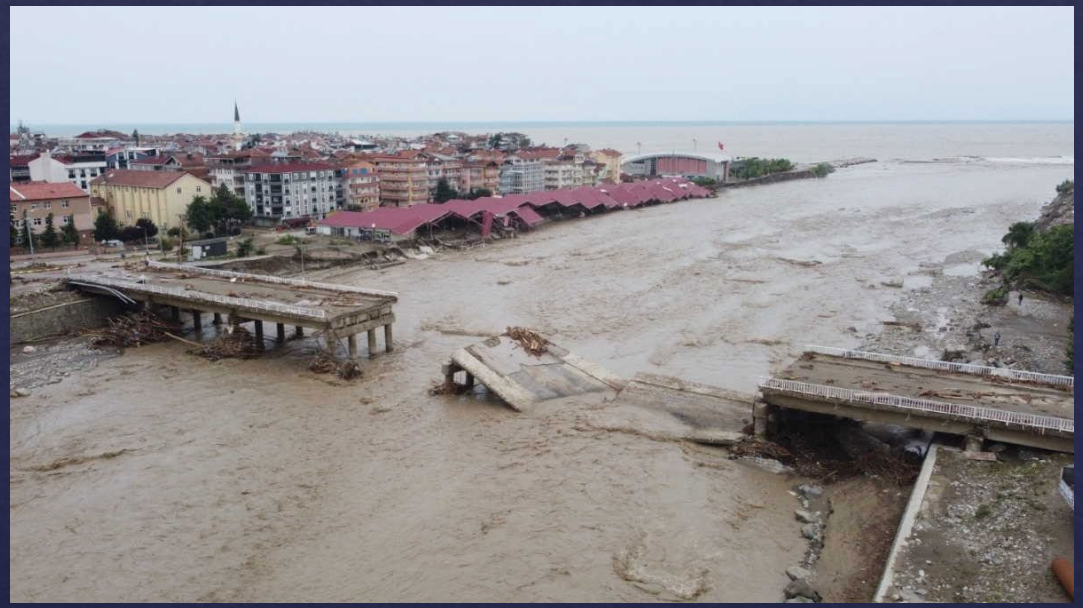
SELF-ORGANIZING MAPS AS A DIMENSION REDUCTION APPROACH FOR SPATIAL GLOBAL SENSITIVITY ANALYSIS VISUALIZATION

GIScience and Big Data for Hazard Research Session

University of Massachusetts Amherst
BE REVOLUTIONARY™

Seda Şalap-Ayça, PhD
Department of Geosciences
@ Mount Ida Campus
ssalapayca@umass.edu

How do we approach hazard?



What are the commonalities ?

How can we understand and try to find a solution for hazard?

ssalapayca@umass.edu

What is a model?

Represent a real thing.

Represent **some characteristics of** the real thing but **ignores other.**

ssalapayca@umass.edu

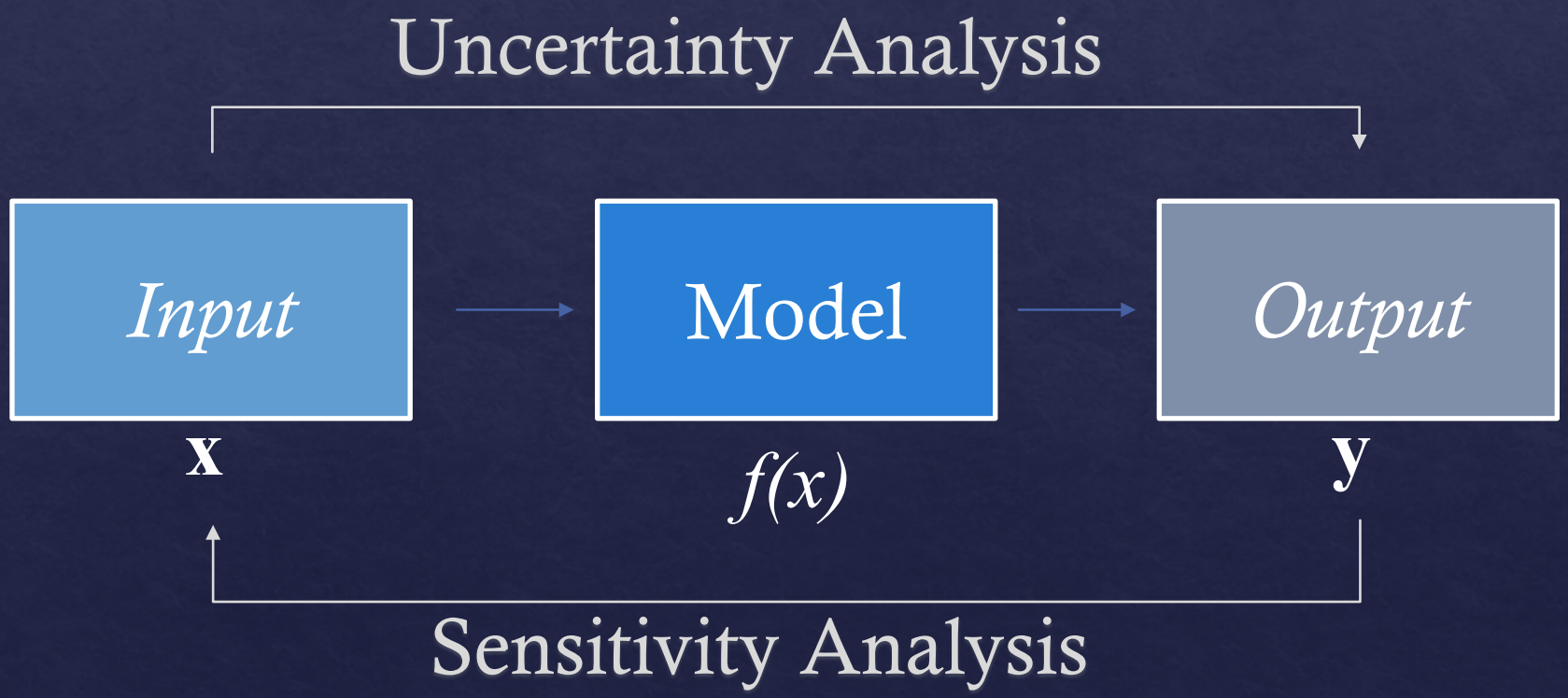
What if all models are wrong?

"All models are wrong, but some are useful."

George E. P.Box ,1987

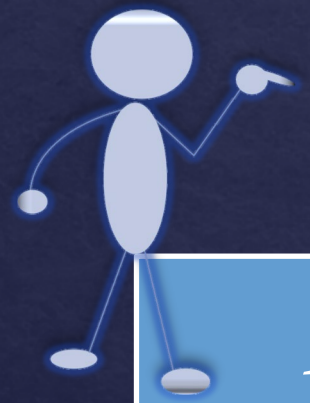
ssalapayca@umass.edu

Spatially Explicit Uncertainty and Sensitivity Analysis



ssalapayca@umass.edu

Spatially Explicit Uncertainty and Sensitivity Analysis



Input → Model → *Output*

x — *f(x)* — **y**

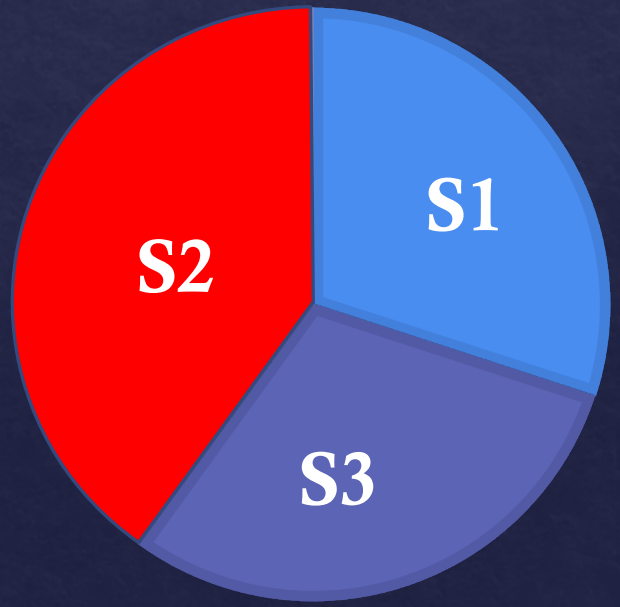
ssalapayca@umass.edu

Global Sensitivity Analysis

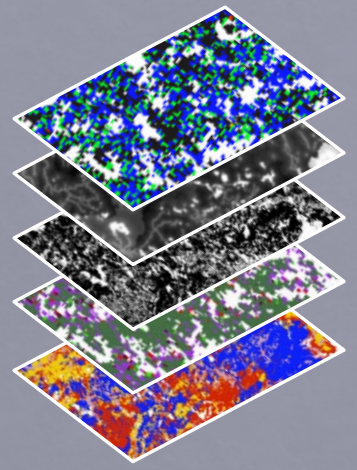
- Variance Decomposition (Sobol) – decomposition of output variance into parts

FIRST ORDER - INDIVIDUAL EFFECTS

HIGHER ORDER INTERACTIONS

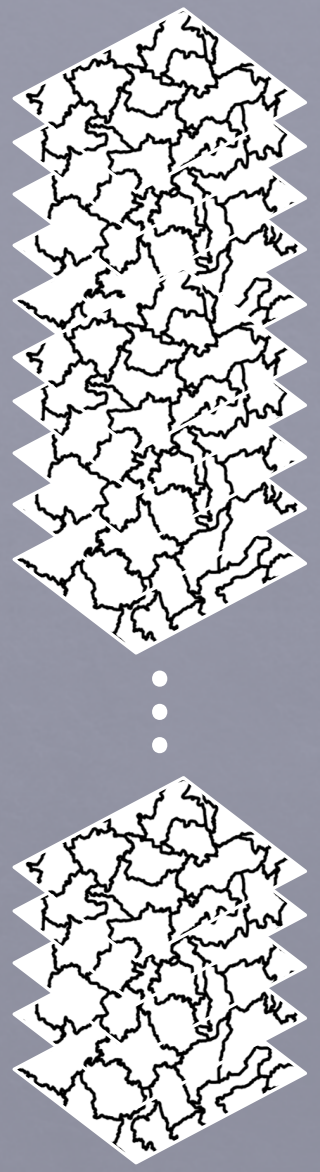


ssalapayca@umass.edu



N1 –Wildlife
N2 – Water Quality
N3-Erosion
N4 –Enduring Benefits
N5-Air Quality

Criteria Value Maps



Uncertainty Analysis

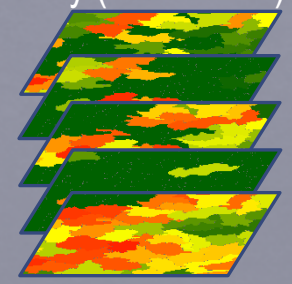
Mean Suitability Map

Uncertainty Map

Sensitivity Analysis

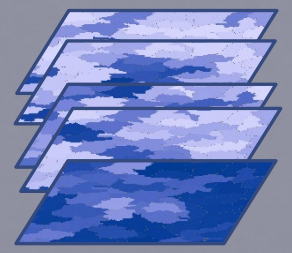
Output Variability (First Order) Maps

S_1
S_2
S_3
S_4
S_5



Interactions(Total-Order) Maps

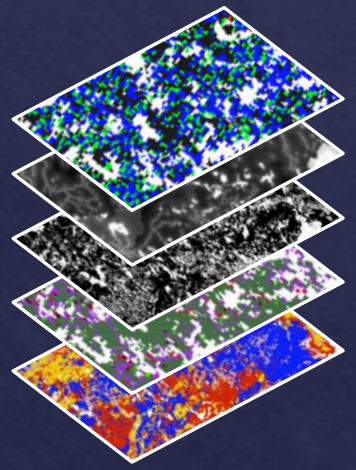
S_{T1}
S_{T2}
S_{T3}
S_{T4}
S_{T5}



ssalapayca@umass.edu

Temporal Sensitivity

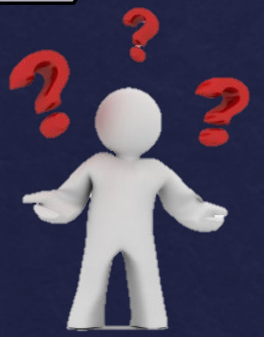
10 outputs



5 inputs

t_1 t_2 t_3 t_4

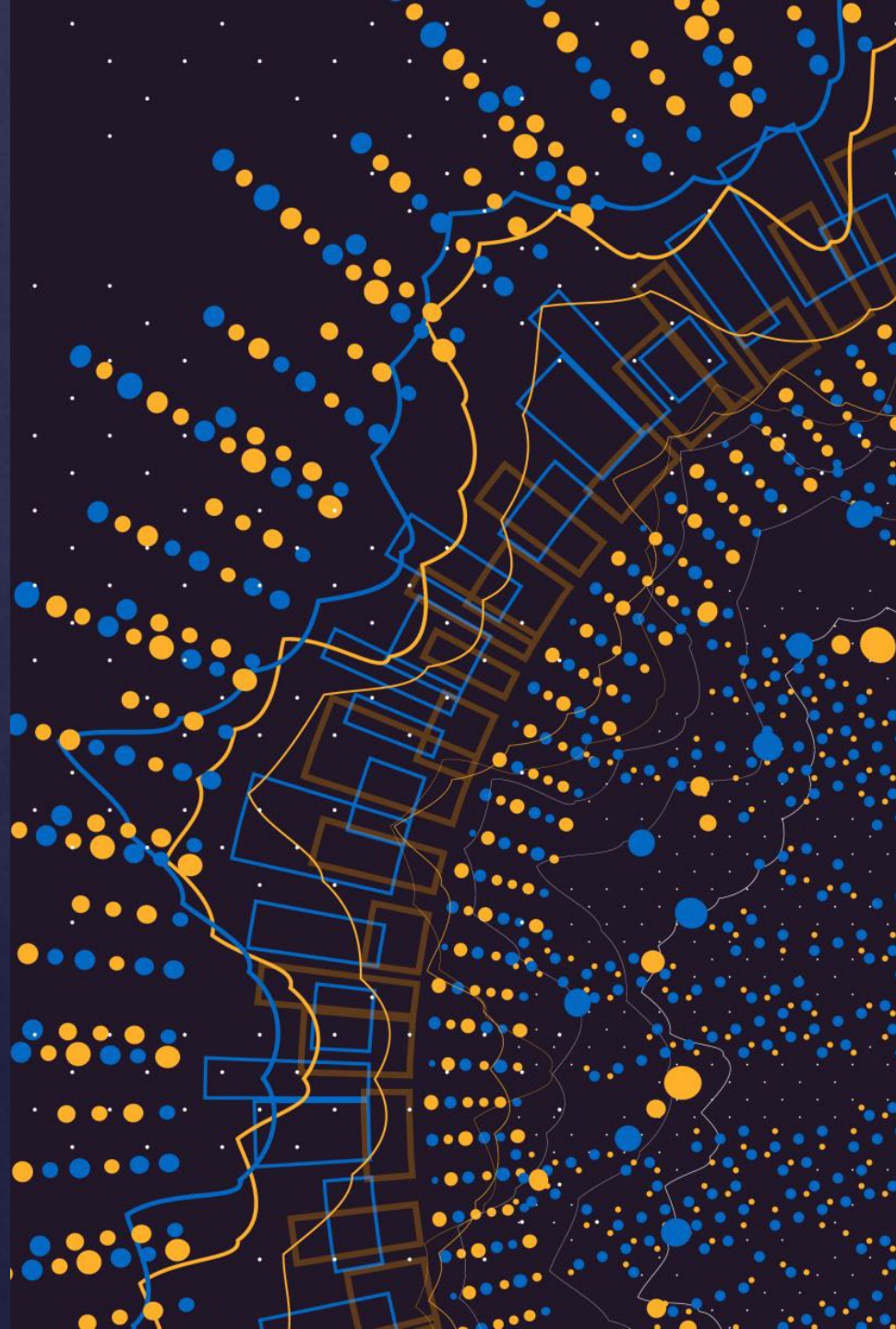
Global Sensitivity Analysis Output



ssalapayca@umass.edu

Self-Organizing Maps – Kohonen Networks

- SOM is used to explore and visualize patterns in high-dimensional datasets
- Introduced by Teuvo Kalevi Kohonen in 1982
- Clustering technique that identifies groups in a dataset without having to use traditional statistical techniques



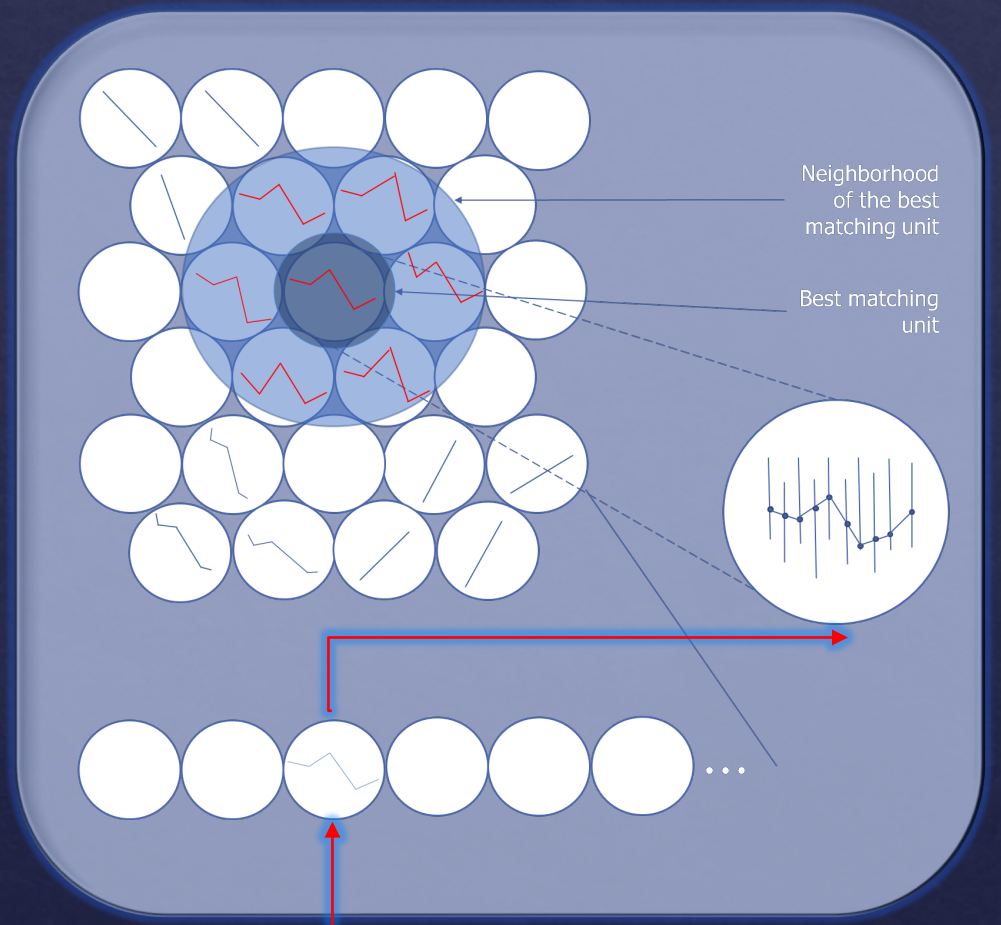
Self-Organizing Maps

- *TOPOLOGY*

  Each neuron in the SOM grid is closely related to each other and each of the inputs are connected to each of the output nodes by means of a connection weight.

- OBJECTIVE

  Find the set of centroids (neurons) to represent the cluster, but with topological restrictions.



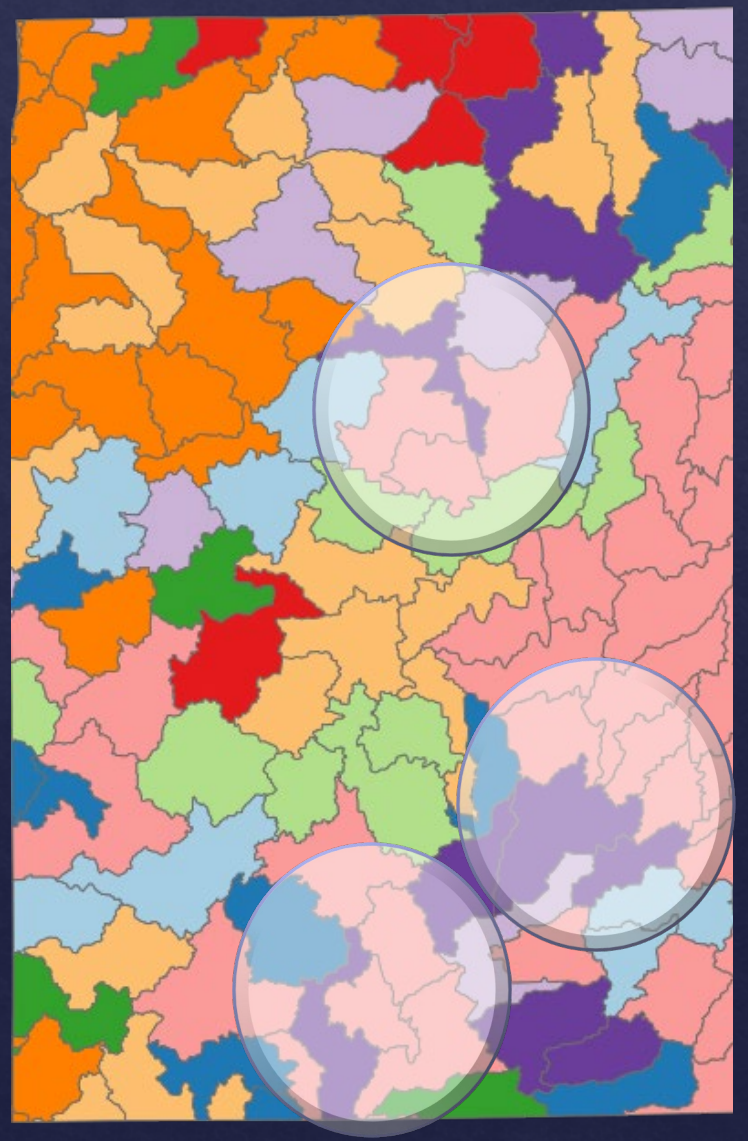
$$\{feature_i\} : (attribute_1, attribute_2, \cdots, attribute_n)$$

ssalapayca@umass.edu

SOM Weight Lattice

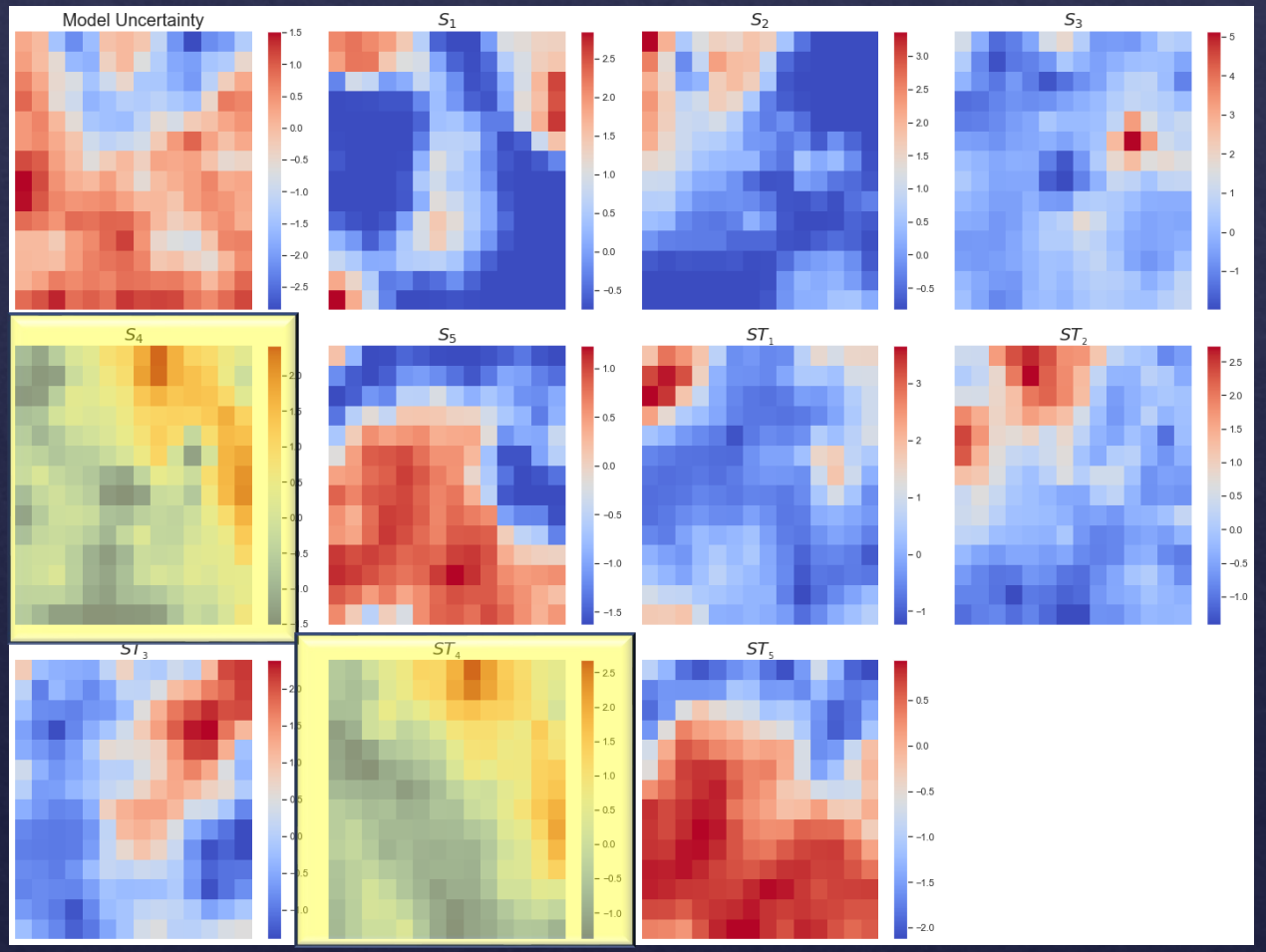
Data: SOM Cluster Array and GSA Indices
Result: Thematic Map for GSA
N = Number of Features;
(x,y) = Unified Matrix Location;
m = Number of Clusters;
GSA dictionary = (key=range(0,len(m); values=GSA label);
for $N \neq 0$ **do**
    get the (x,y) of N^{th} feature;
    get the m at (x,y);
    join m, N, in the data frame;
    update cluster ID as with GSA dictionary;
end

GSA Influence Map



Most Relevant Feature Plane for GSA

Data: SGSA output tabular data (i.e txt, csv)
Result: Pandas dataframe of GSA results for MiniSom
$N \leftarrow 2n$;
while $N \neq 0$ **do**
    read current S_i index map;
    normalize the current S_i index map;
    concatenate S_i index map in i^th column in the data frame;
end



ssalapayca@umass.edu

Conclusion and Charting a Course Forward

- As a powerful **confidence mechanism**, **SGSA supports modeling** in spatial science
- **SOM-based exploratory analysis** helps to **reduce the high-dimensional SGSA output**
- The analyst can utilize the ***most relevant feature plane comparison*** to ***reduce the number of SGSA indices***
- Next steps are
  - Exploring the dependencies on SOM parameters
  - Improving performance and reducing the computational Cost

ssalapayca@umass.edu

Thanks for your attention

ssalapayca@umass.edu

@sedasalap