SFI Centre for Research Training in Foundations of Data Science

"Investigation into different up-scaling approaches to address the double-penalty problem arising in high-resolution EPS"

Machine Learning and Dynamical techniques to be applied to ensemble-based forecasts of precipitation

Comito T. , Clancy C. , Daly C., Hally A.

October 1, 2020

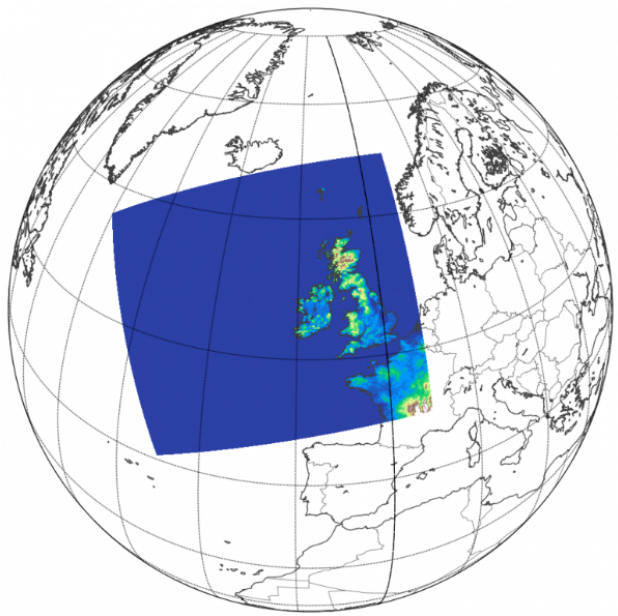
MET éireann
The Irish Meteorological Service

Intro

Presentation conceptual map

- Project justification: the Double penalty issue and the a fixed up-scaling
- Data sets overview
- Walk through algorithm steps: a Dynamical and Machine Learning approaches
- Verification and results
- Conclusions

IREPS



- Version: HARMONIE-AROME 40h1.1 (running 4 times a day, approximately 1h)
- Ensemble: $10 + 1$ control (deterministic)
- Short Spatial and temporal: regional scale and daily time-scale
- Grid: $10^3 \times 9\ 10^2$ horizontal points and 65 vertical levels; $dx = 2.5km$.
- Output: 54h forecast

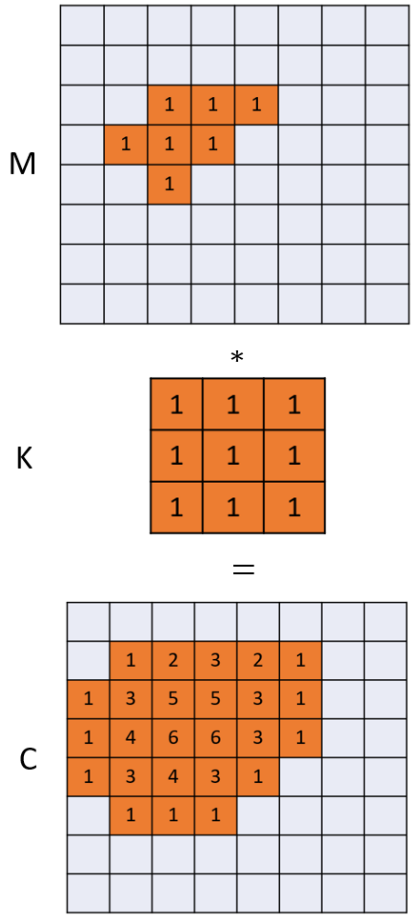
The up-scaling procedure

- It works on categorical matrix: total precipitation field is mapped onto an hit/not-hit concept object, looking at rainfall over a threshold

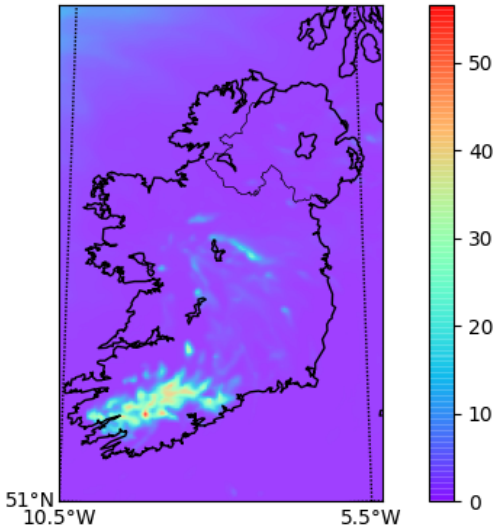
$$M_{ij} = \begin{cases} 1\ ,\ t > p_{ij} \\ 0\ ,\ t < p_{ij} \end{cases} \qquad \bar{M}_{ij} = \frac{1}{11} \sum_{m=1}^{11} M_{ij}^{m}$$

- Properties are scaled up, weighting contributions from closest neighbours through a kernel
- The discrete version of convolution function is applied

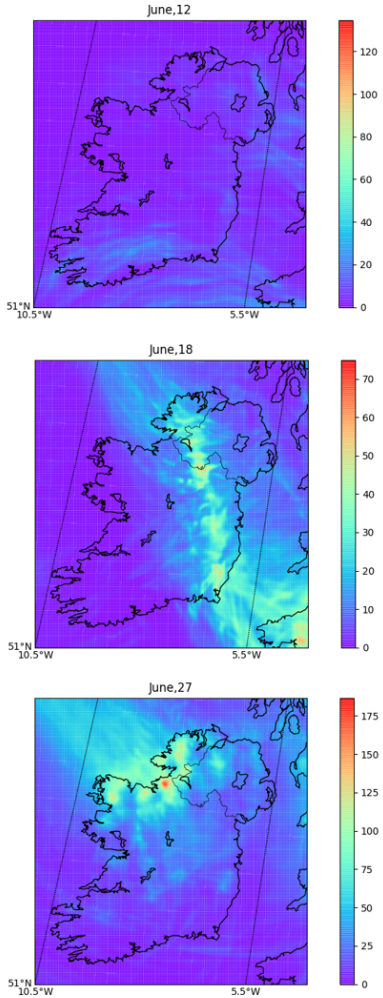
$$C_i = \sum_j M_{i+k-j} K_j, \qquad (K)_{ij} \in \mathbb{R}^{n,n}\ n = (2R+1)$$



Data sets



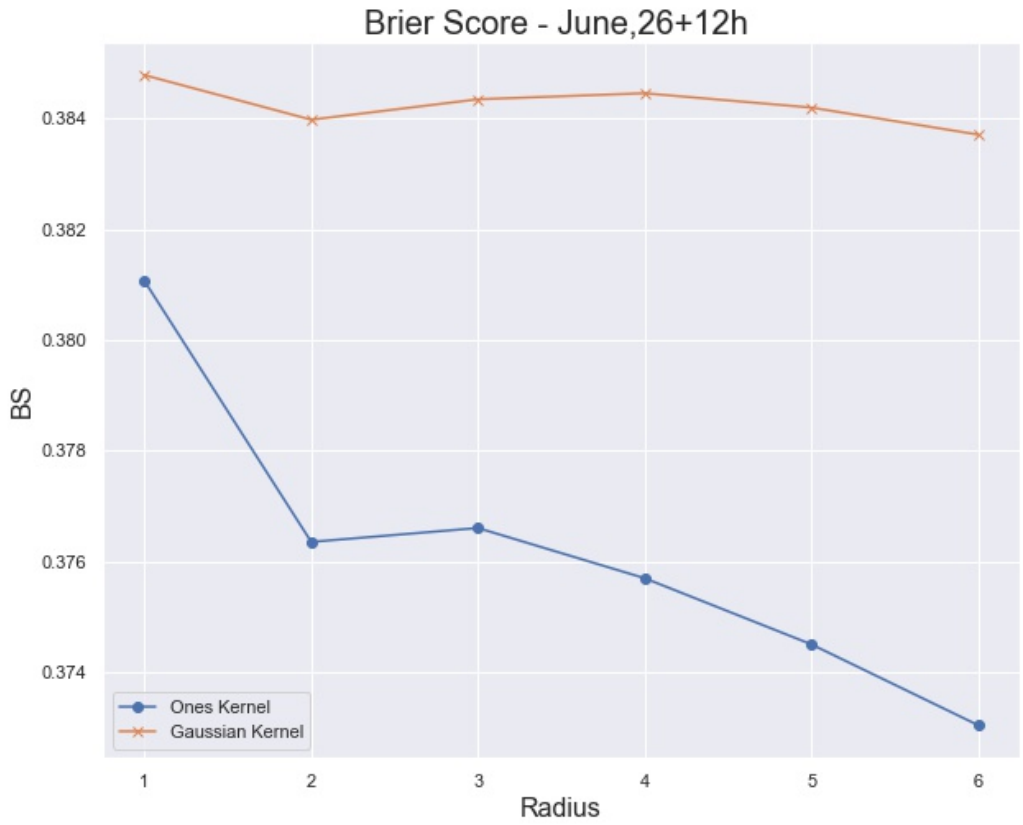
May, 9: Quick development of convective phenomena after 13 hours forecast plus. Localised rainfall in the south and sparse precipitation, with sporadic agreement over central Ireland



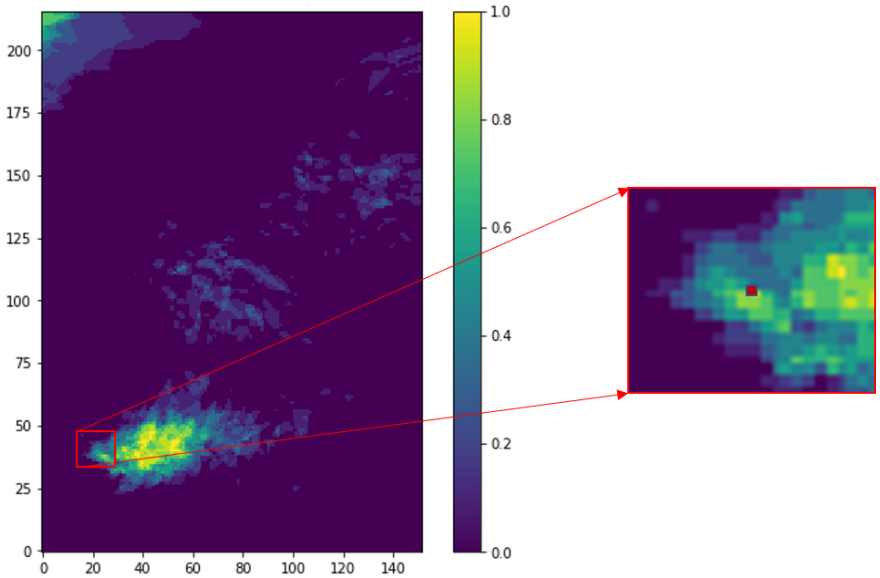
June,

- **7:** Dry and almost null precipitation;
- **9:** Non localised convective activity;
- **12:** Showers in the north. Intense squall-line manifest to SE;
- **18:** Cold front crossing the country;
- **26:** Thunderstorm followed by scattered showers;
- **27:** Severe rainfall in the north;
- **28:** Same severe precipitation, but greater concentration

Optimum Radius



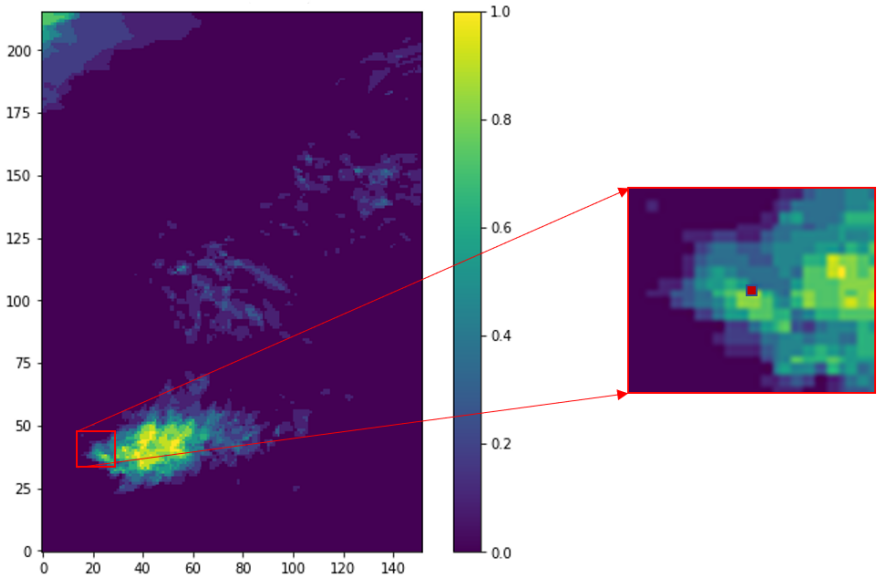
A Dynamical method



The Spread based up-scaling Algorithm

1. Evaluate the associated fraction probability matrix

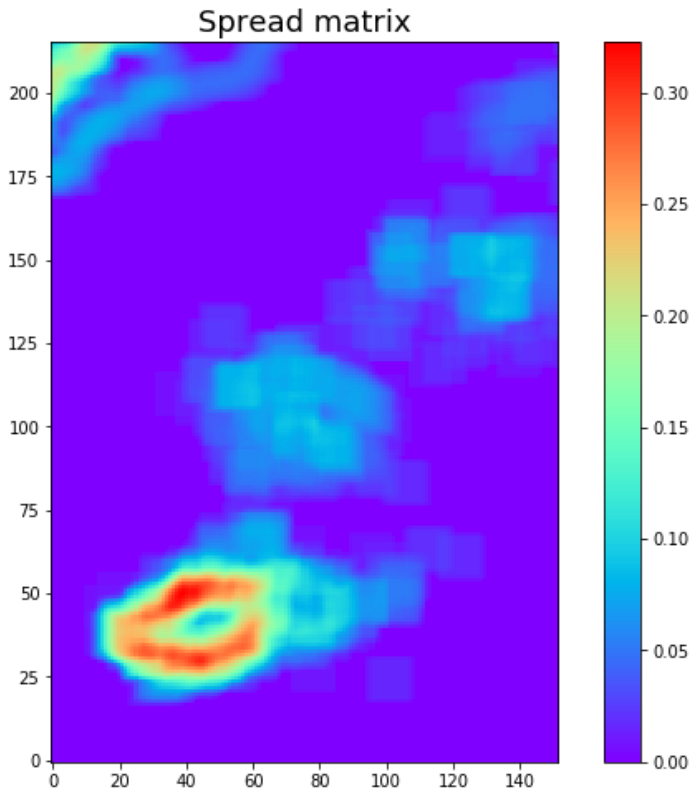
A Dynamical method



The Spread based up-scaling Algorithm

1. Evaluate the associated fraction probability matrix
2. Define a variability window

A Dynamical method

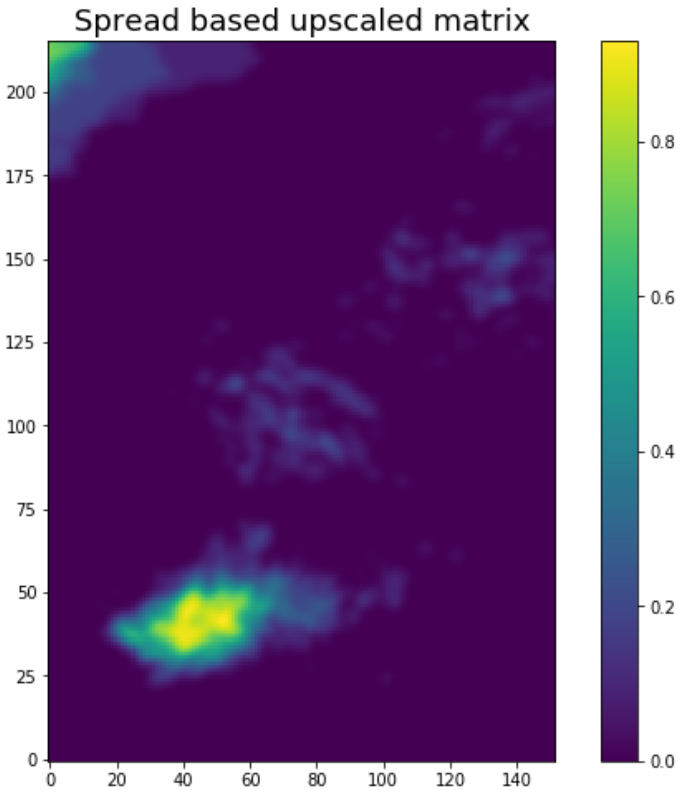


The Spread based up-scaling

Algorithm

1. Evaluate the associated fraction probability matrix
2. Define a variability window
3. Get the associated spread matrix

A Dynamical method

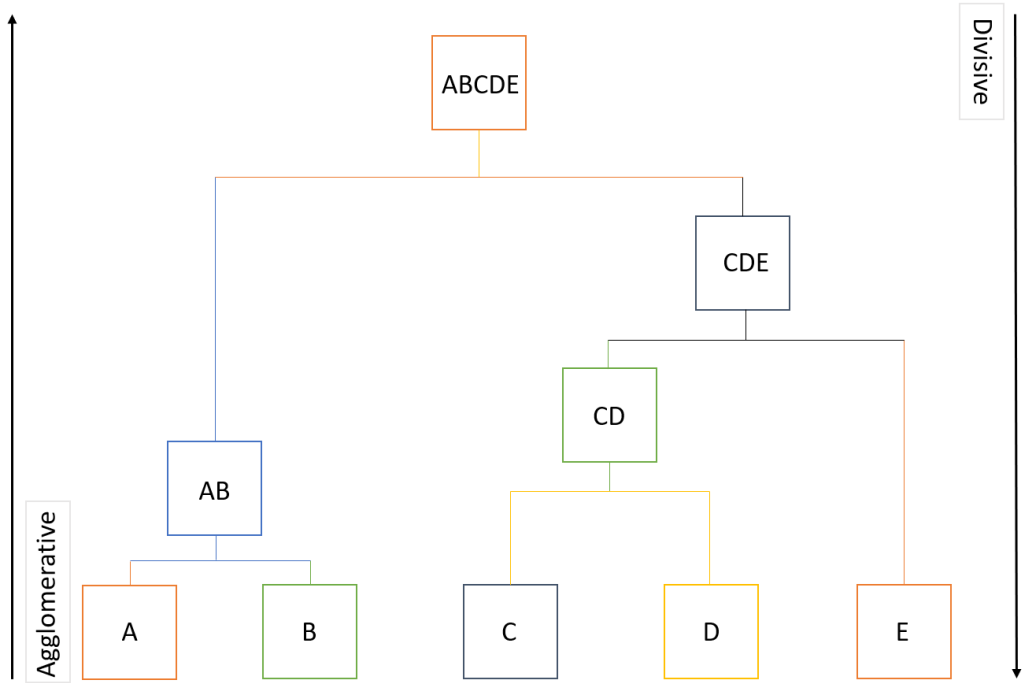


The Spread based up-scaling Algorithm

1. Evaluate the associated fraction probability matrix
2. Define a variability window
3. Get the associated spread matrix
4. Up-scaling is assigned with respect to indicator's (e.g. Standard Deviation) value

Machine Learning based approach

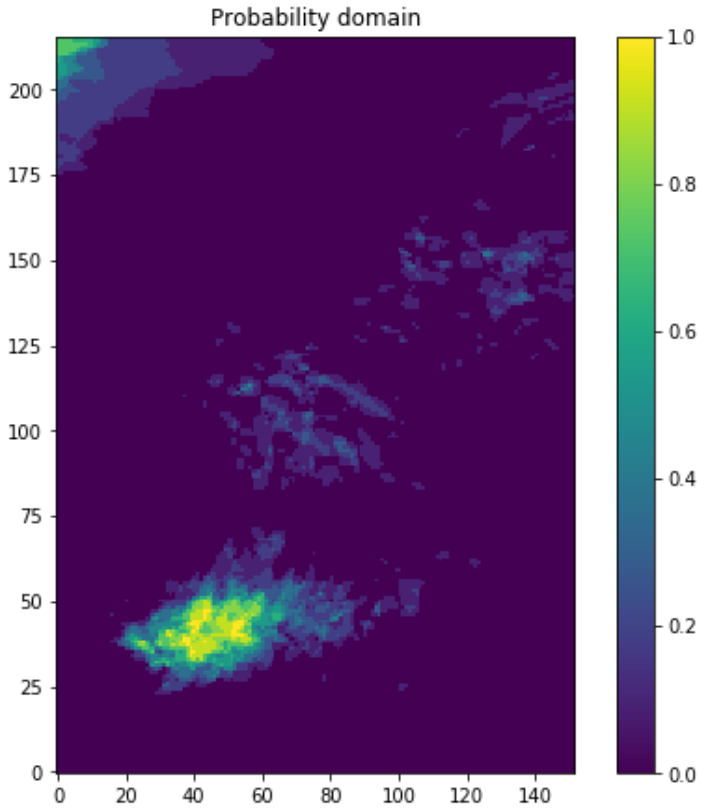
- Hierarchical clustering
- Unsupervised techniques
- No prior knowledge of the number of clusters is required
- Similarity estimation through linkage operation



Machine Learning based approach

The Clustering based up-scaling Algorithm

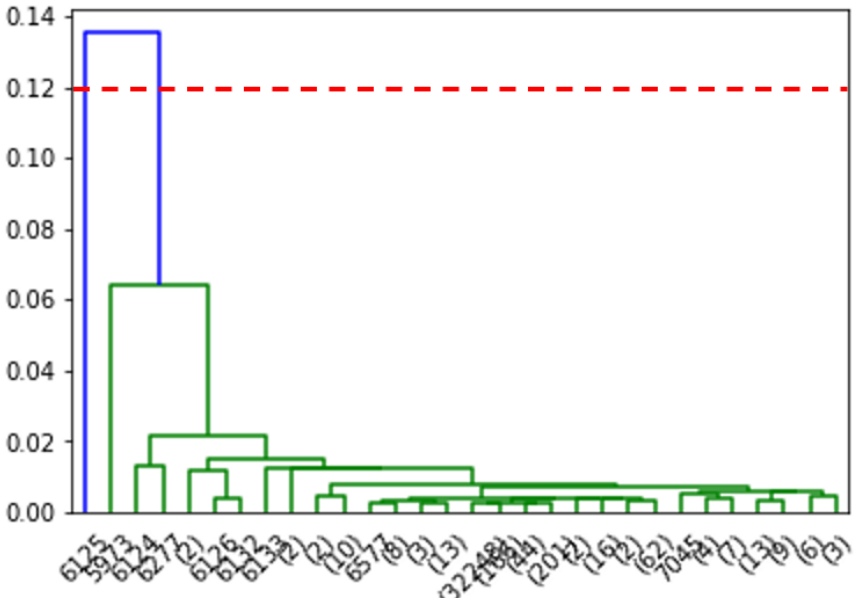
1. Perform linkage operation to get the proximity matrix



Machine Learning based approach

The Clustering based up-scaling Algorithm

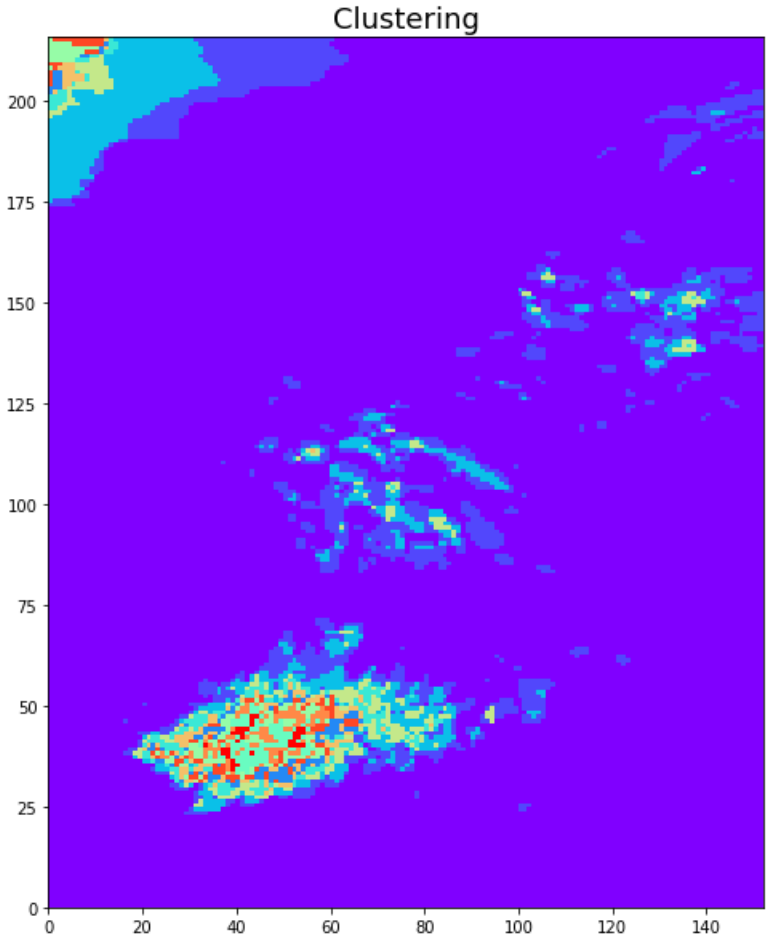
1. Perform linkage operation to get the proximity matrix
2. Find the number of clusters



Machine Learning based approach

The Clustering based up-scaling Algorithm

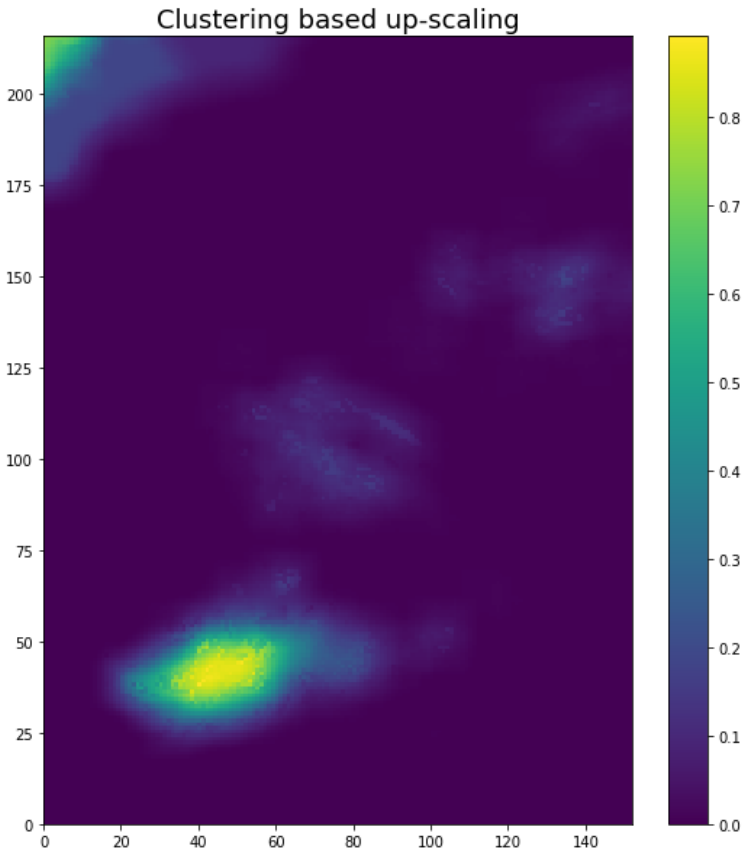
1. Perform linkage operation to get the proximity matrix
2. Find the number of clusters
3. Use Hierarchical agglomerate clustering



Machine Learning based approach

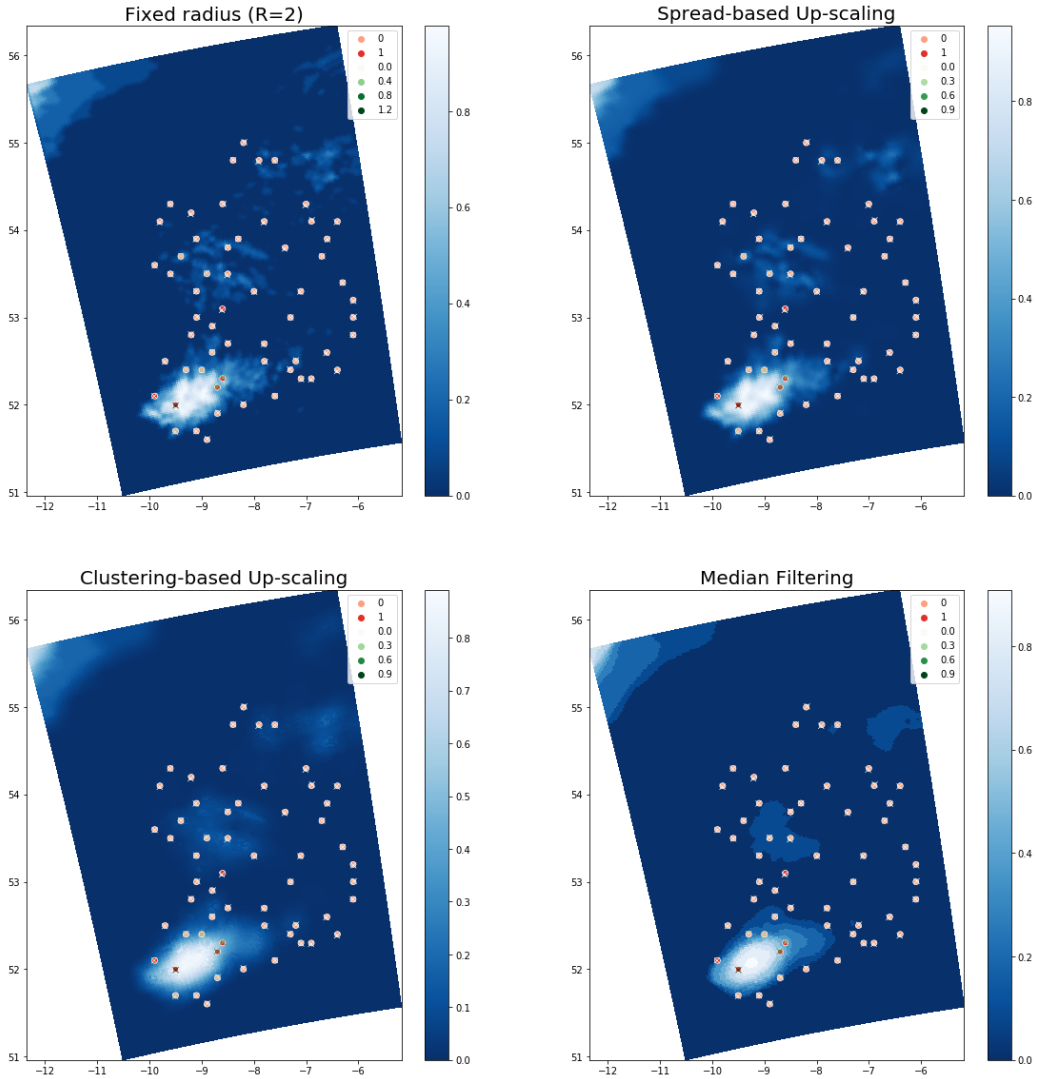
The Clustering based up-scaling Algorithm

1. Perform linkage operation to get the proximity matrix
2. Find the number of clusters
3. Use Hierarchical agglomerate clustering
4. Points within the same cluster are equally upscaled

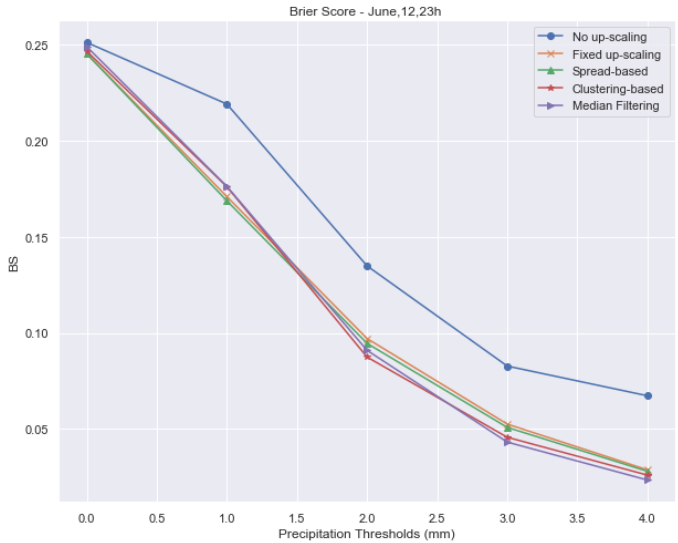


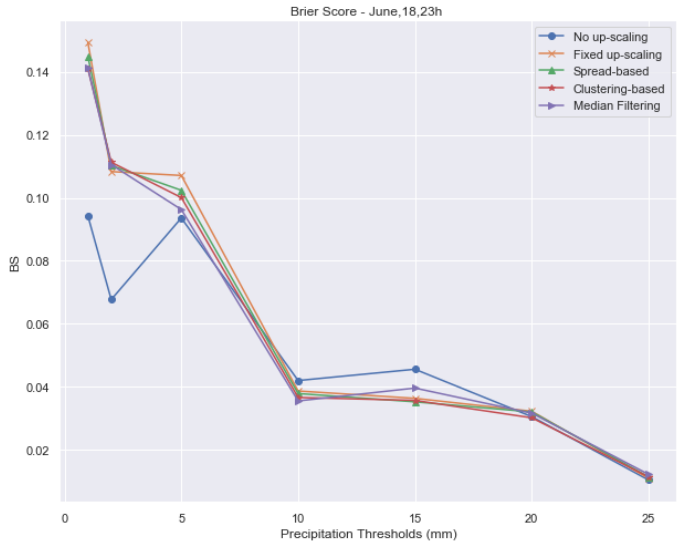
Test phase

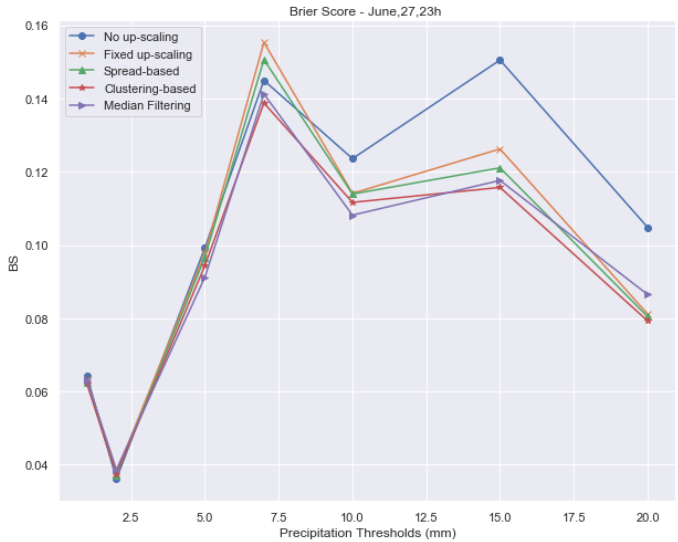
- Original fraction probability matrix
- Fixed up-scaling ($R = 2$)
- Spread based up-scaling
- Clustering based up-scaling
- Median Filtering

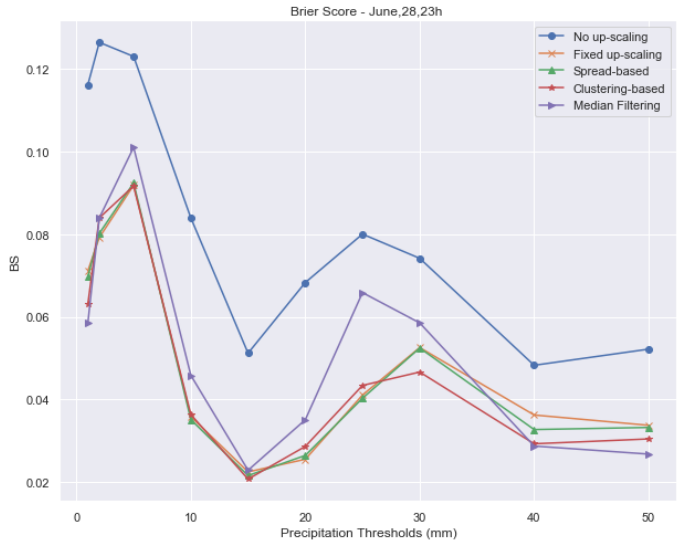


Brier Score

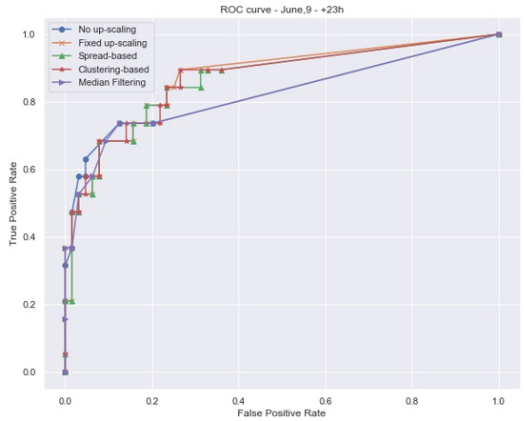

Brier Score - June,12,23h
No up-scaling
Fixed up-scaling
Spread-based
Clustering-based
Median Filtering
BS
Precipitation Thresholds (mm)

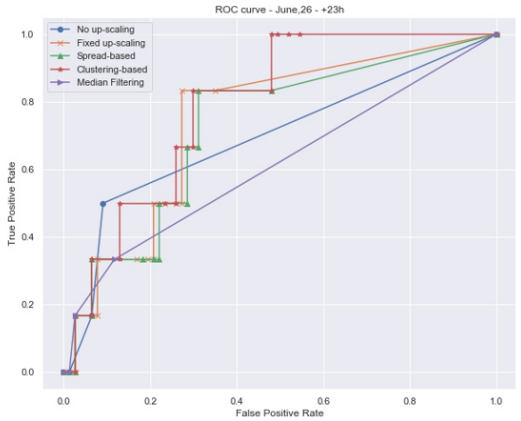

Brier Score - June,18,23h
No up-scaling
Fixed up-scaling
Spread-based
Clustering-based
Median Filtering
BS
Precipitation Thresholds (mm)

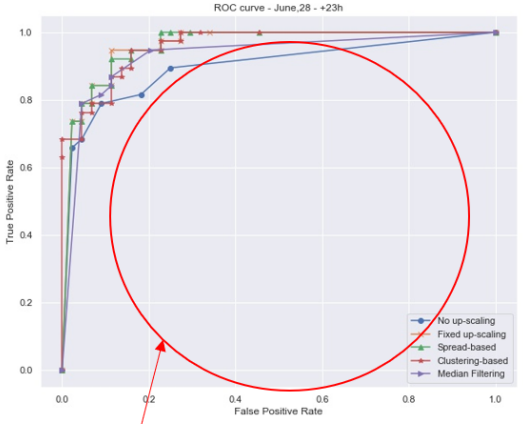

Brier Score - June,27,23h
No up-scaling
Fixed up-scaling
Spread-based
Clustering-based
Median Filtering
BS
Precipitation Thresholds (mm)

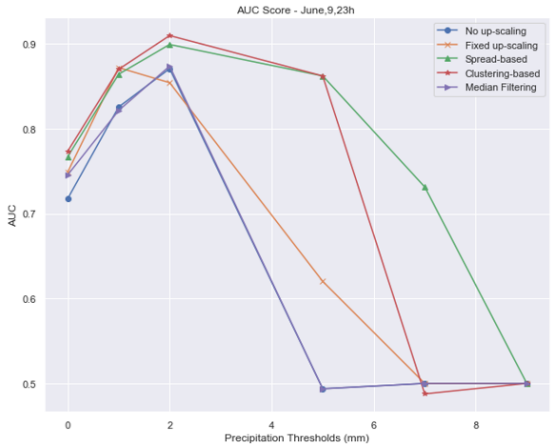

Brier Score - June,28,23h
No up-scaling
Fixed up-scaling
Spread-based
Clustering-based
Median Filtering
BS
Precipitation Thresholds (mm)

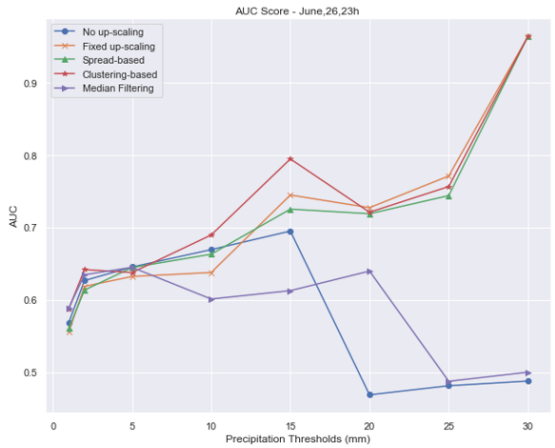
ROC & AUC

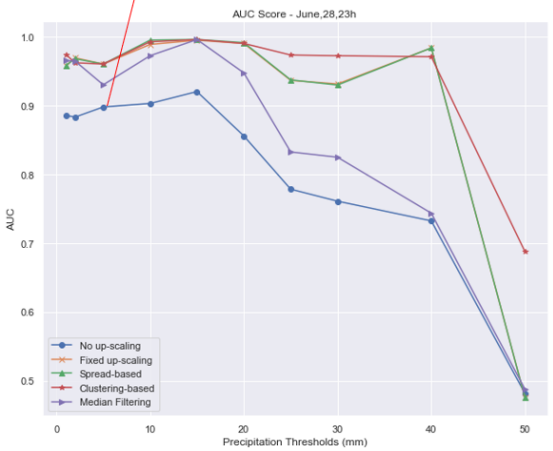

ROC curve - June,9 - +23h
No up-scaling
Fixed up-scaling
Spread-based
Clustering-based
Median Filtering
True Positive Rate
False Positive Rate


ROC curve - June,26 - +23h
No up-scaling
Fixed up-scaling
Spread-based
Clustering-based
Median Filtering
True Positive Rate
False Positive Rate


ROC curve - June,28 - +23h
No up-scaling
Fixed up-scaling
Spread-based
Clustering-based
Median Filtering
True Positive Rate
False Positive Rate


AUC Score - June,9,23h
No up-scaling
Fixed up-scaling
Spread-based
Clustering-based
Median Filtering
AUC
Precipitation Thresholds (mm)


AUC Score - June,26,23h
No up-scaling
Fixed up-scaling
Spread-based
Clustering-based
Median Filtering
AUC
Precipitation Thresholds (mm)


AUC Score - June,28,23h
No up-scaling
Fixed up-scaling
Spread-based
Clustering-based
Median Filtering
AUC
Precipitation Thresholds (mm)

Conclusion & THM

- Strong dependency on the weather scenario
- BS does not highlight any outstanding performances even though slight improvements are generally obtained
- AUC scores agree on a better ability in classifying precipitation events using the dynamical Spread and Clustering based up-scaling;
- Improved forecast skill for convective rain events

Thank You!