On the use of circulation classifications by self-organizing maps toward studying extreme weather events

Jan Stryhal, Eva Plavcová

Institute of Atmospheric Physics, Czech Academy of Sciences, Prague

The goals of the study are as follows:

1. To assess the potential of classifications by self-organizing maps (SOMs) in synoptic-climatological studies of extreme weather events

2. To assess the potential of Sammon mapping, routinely used to validate SOMs, for visualizations and classifications of the circulation continuum

Synoptic climatology
EGU 2020

What are self-organizing maps?

Self-organizing maps are a neural-network based method of circulation pattern typing.

Unlike other classification methods, a SOM is able to **sort circulation types (CTs)** into an array (“map”) that **preserves the topological structure** of the data space (therefore, similar CTs are close to each other).

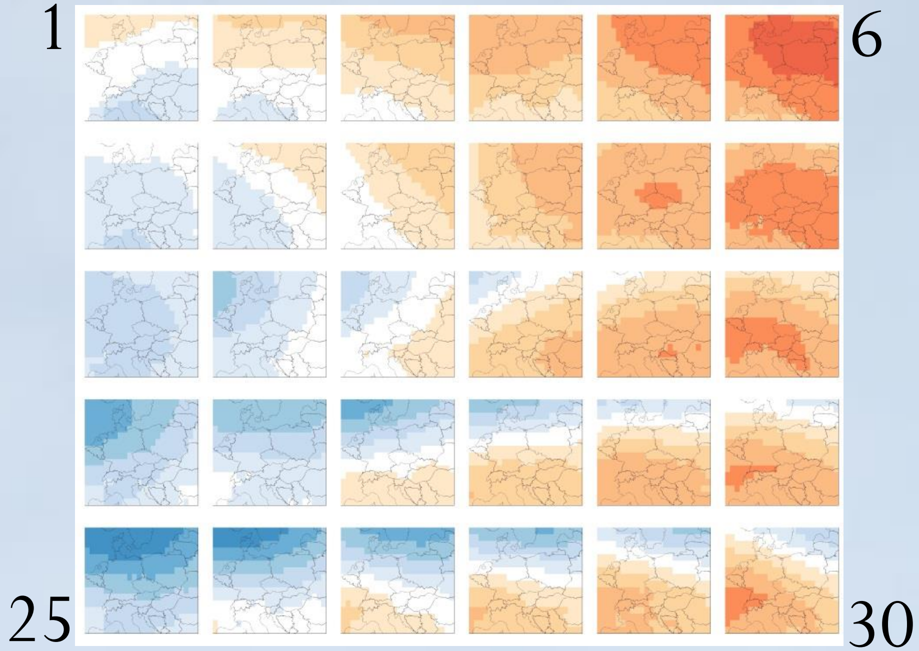


Fig. 1. A rectangular 2D 5×6 SOM of ERA5 extended winter daily SLP patterns *(central Europe, XI– III 1979–2017, SLP intervals: 5 hPa)*

One way to assess the degree of topological structure preservation is via **Sammon mapping** (SM; Sammon 1969).

SM projects all objects (in this case 30 multidimensional CT centroids shown in Fig. 1) onto (typically) a 2D plane, such that the difference between the Euclidean distances over all pairs of CTs in the original (SLP) space and in the new Sammon space is minimized.

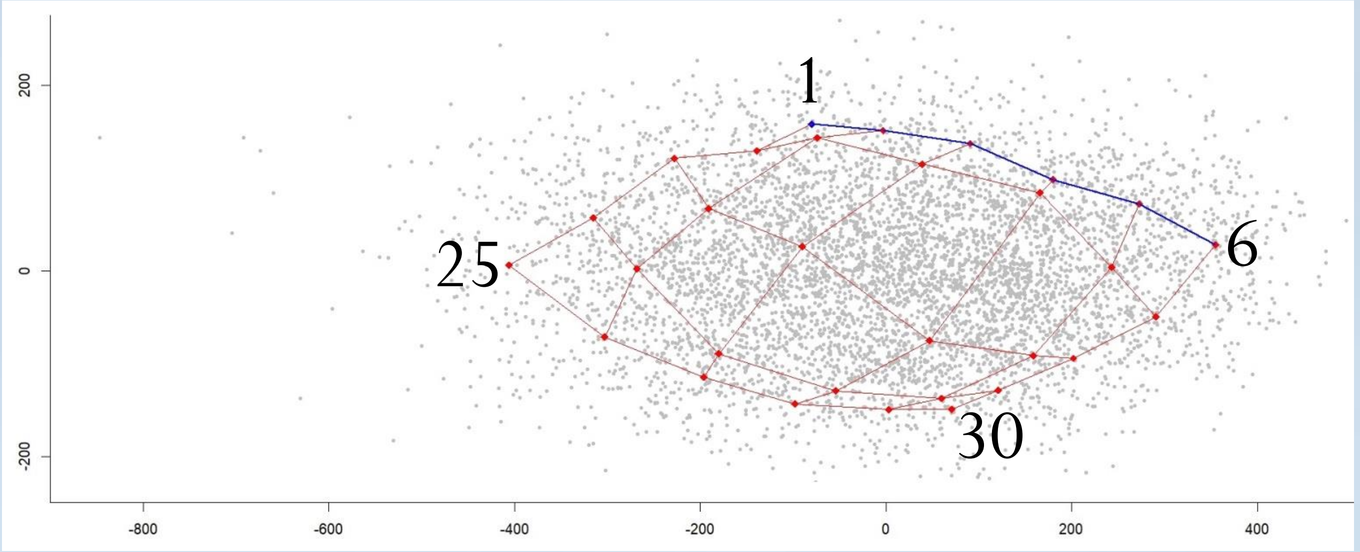


Fig. 2. Sammon map of the SOM in Fig. 1 (red dots) among all daily SLP patterns (grey dots) used to train the SOM. *The first (top) row of CTs is highlighted in **blue***

Could SOMs be used to study links between atmospheric circulation and extreme weather?

SOMs, or circulation classifications in general, represent a **powerful tool for studying links to circulation**.

However, **classifications typically provide too generalized a view** on the circulation continuum to be effective in studying weather extremes, which are often linked to rather rare circulation fields.

Here, we study whether **larger SOMs** could be trained that would still be topologically stable and would allow for studying and comparing various extreme winter weather events in various datasets (reanalyses, RCMs).

The R 'kohonen' package by Wehrens and Kruisselbrink is used to train SOMs (https://cran.r-project.org/web/packages/kohonen/kohonen.pdf) and the MASS R library is used for Sammon mapping.

Multiple parameters are necessary to predefine when creating a SOM, including:

- **The number of nodes** (or, CTs) and **shape of the SOM array** (number of rows and columns, rectangular vs. hexagonal array)
- **Number of iterations** (times the complete dataset is presented to the network
- **Learning rate** (indicating the amount of change in each iteration)
- **Neighbourhood function** (bubble vs. Gaussian) and **the radius of neighbourhood** (influencing how many and how distant SOM nodes are changed in each iteration

Can one train large SOMs and use them to study extremes?

The effect of all abovementioned parameters on the resulting classifications were tested, **the neighbourhood radius (NR) appearing to be the crucial one.**

With the NR reduced to one, the classification becomes identical to k-means. This leads to far better classifications in terms of separation of clusters and explained variation of data; however, the topological structure dissipates (*compare Figs. 3 and 4*).

On the other hand, **too large values of NR lead to only very poor and useless classifications** that span only the centre of the data space.

Therefore, it does not seem to be possible to train a large SOM that would span the whole data space (and would thus be representative of outlying circulation patterns); **larger SOMs are only finer representations of the same (limited) data subspace,** rather than extending toward outlying data points.

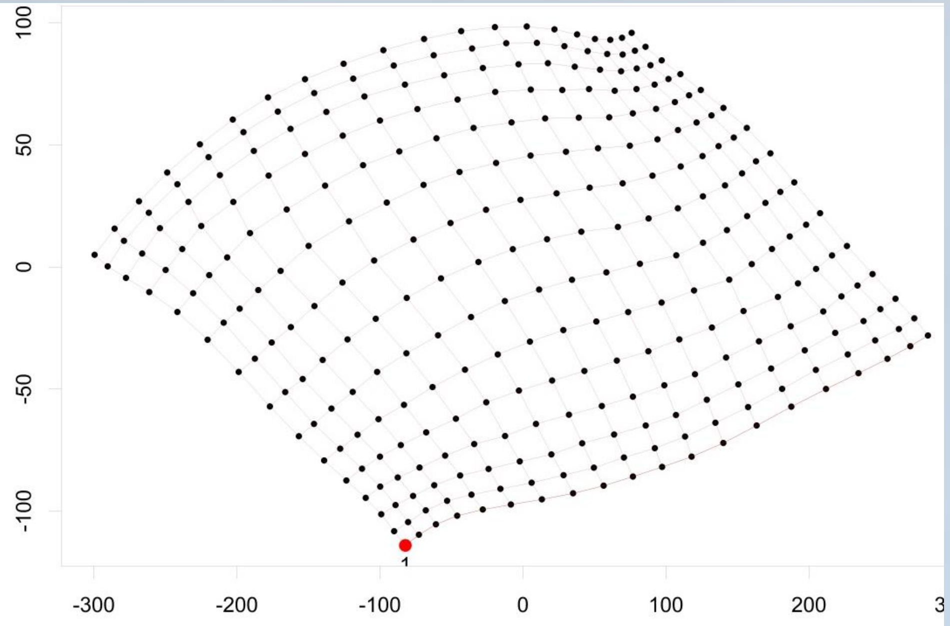


Fig. 3. *Sammon map of a 20×15 SOM trained with the radius parameter linearly decreasing from 10 to 3 units.*

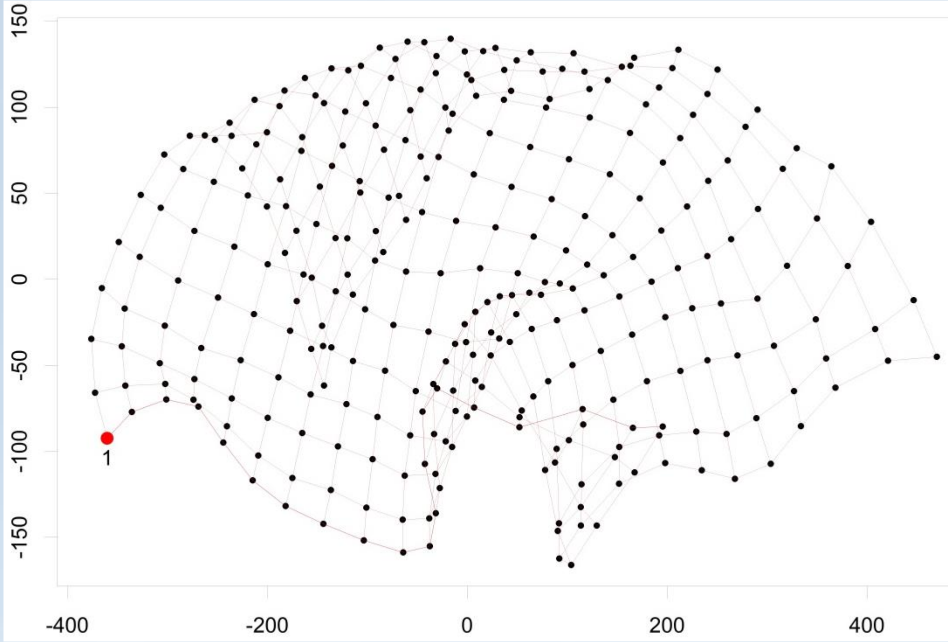


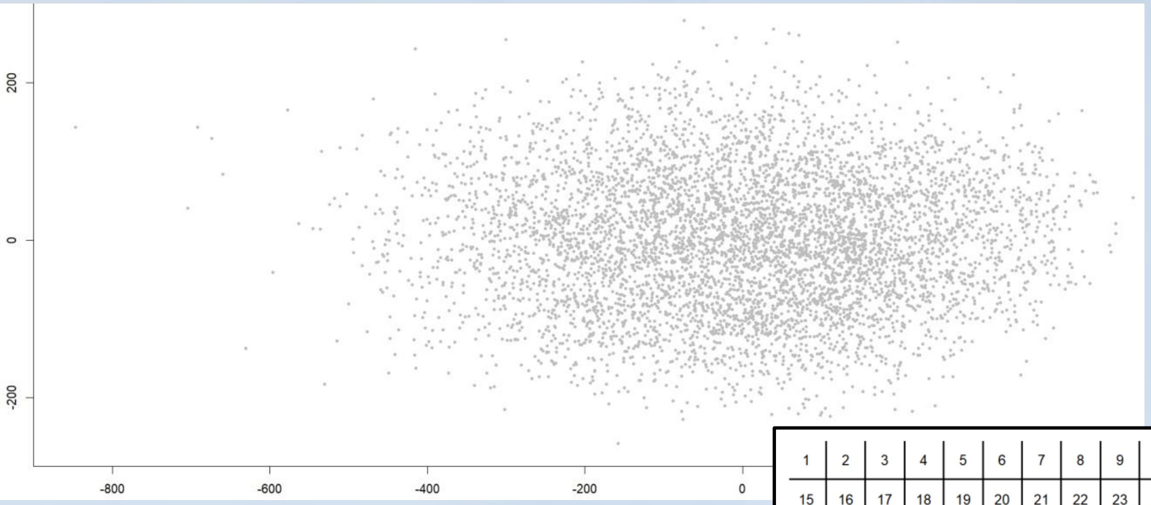
Fig. 4. *As in Fig. 3, except for radius between 10 and 1 unit.*

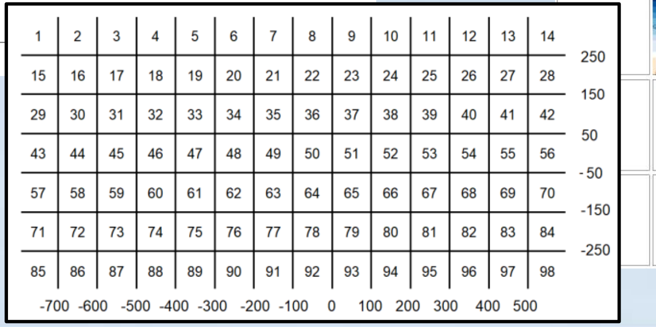
Could we use Sammon mapping instead of SOMs to create classifications representative of both the topology of data and extreme circulation patterns?

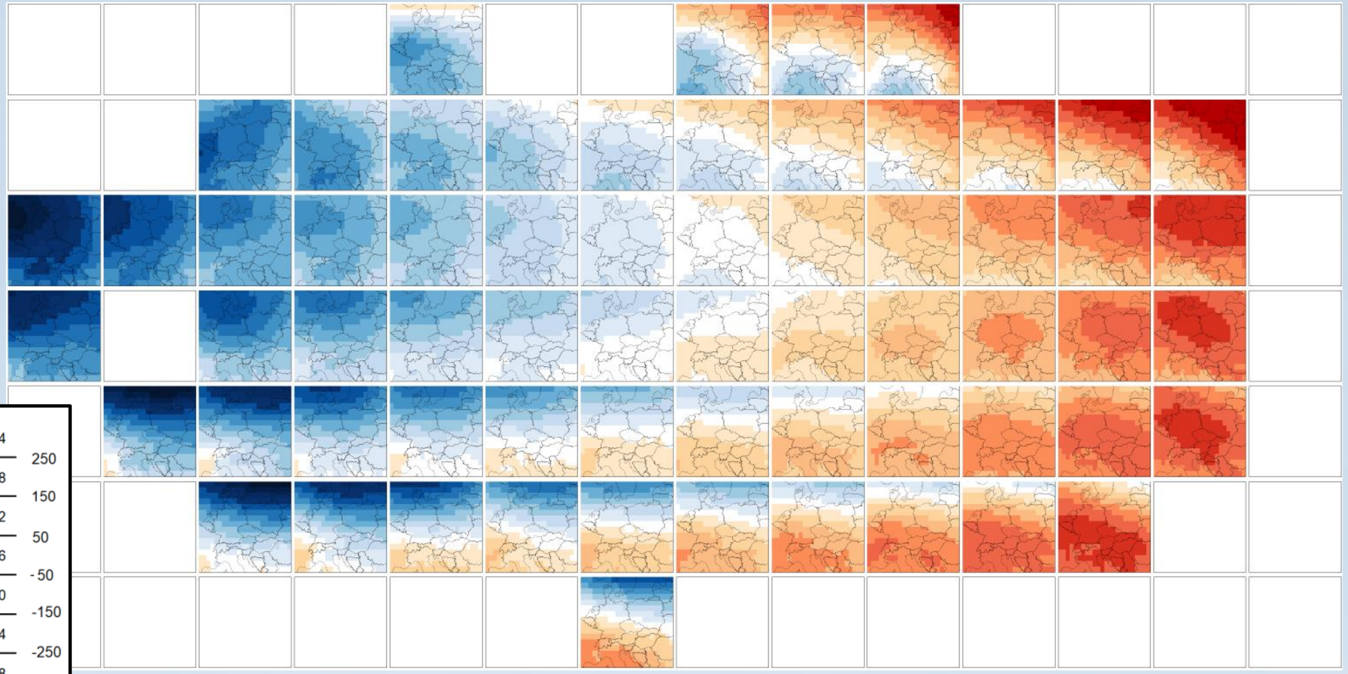
SOMs seem to have many limitations if one wants to analyse extreme atmospheric states. In cases in which we want to emphasize the topological structure of the atmospheric circulation continuum (see examples below), **Sammon mapping (SM) itself could represent a viable alternative.**

SM can also be understood as **a non-linear alternative to projection methods such as those based on PCA** (e.g., projection onto a component plane) or planes (graphs) defined via circulation indices (e.g., strength vs. direction of flow), as in the case of the Jenkinson-Collison method).

Fig. 5. *Sammon map of ERA5 daily SLP fields (left), classes covering the whole data space at regular intervals (middle) and their centroid patterns (right)*







On the use of circulation classifications by self-organizing maps toward studying extreme weather

Examples and Summary

Naturally, this method cannot replace but rather **complement traditional classification approaches**.

It makes it possible to **visualize extreme circulation patterns leading to various extreme weather events as part of the continuum of geophysical fields.** Unlike other projection approaches (PCA, arbitrarily defined indices, etc.), it does not discard significant parts of data variance in the process.

The boundaries of the Sammon space are given by the training data, but **can be extended** by projecting additional data onto the SM, including additional periods and/or models.

Results can be visualized either for classifications (as in Fig. 6), but since each (daily) pattern has its own exact coordinates, the degree of generalization can be chosen accordingly to the need of a researcher.

Fig. 6. *ERA5 CT centroids (top), CT frequencies and CT-conditioned probabilities of extreme weather days (bottom). The definition of extreme weather days is based on an evaluation of extremity of temperature, precipitation, and wind gusts across 48–52°N and 10–20°E; approximately 3 % of days are selected for each variable.*