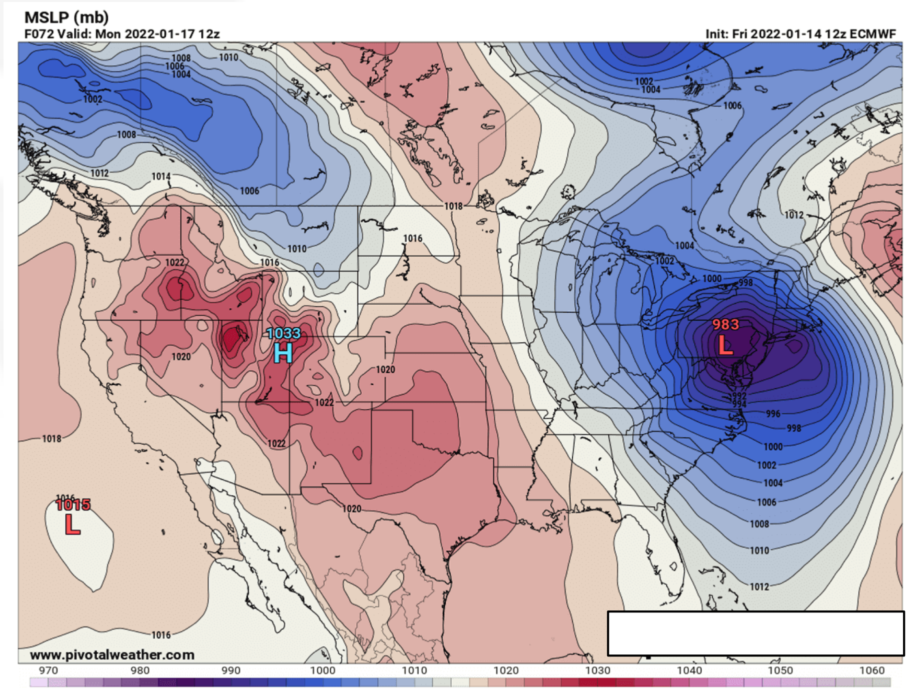
When machine learning deciphers the 'language' of atmospheric air masses

Davide Faranda, Lucas Fery, Flavio Pons
Berengere Podvin, Berengere Dubrulle

Lucas.Fery@lsce.ipsl.fr
Davide.faranda@cea.fr

An artificial intelligence for Meteorology?

A **meteorologist** interpret everyday weather by looking at the motion of **cyclones (low pressure system)** and **anticyclones (high pressure systems)**. Thei bring stable and dry or wet rainy weather



To understand **climate**, we need, to read the weather maps for several decades, and categorize cyclones and anticyclones, their position & their frequency.

This task is beyond human capabilities, so we use machine learning techniques as an artificial intelligent meteorology

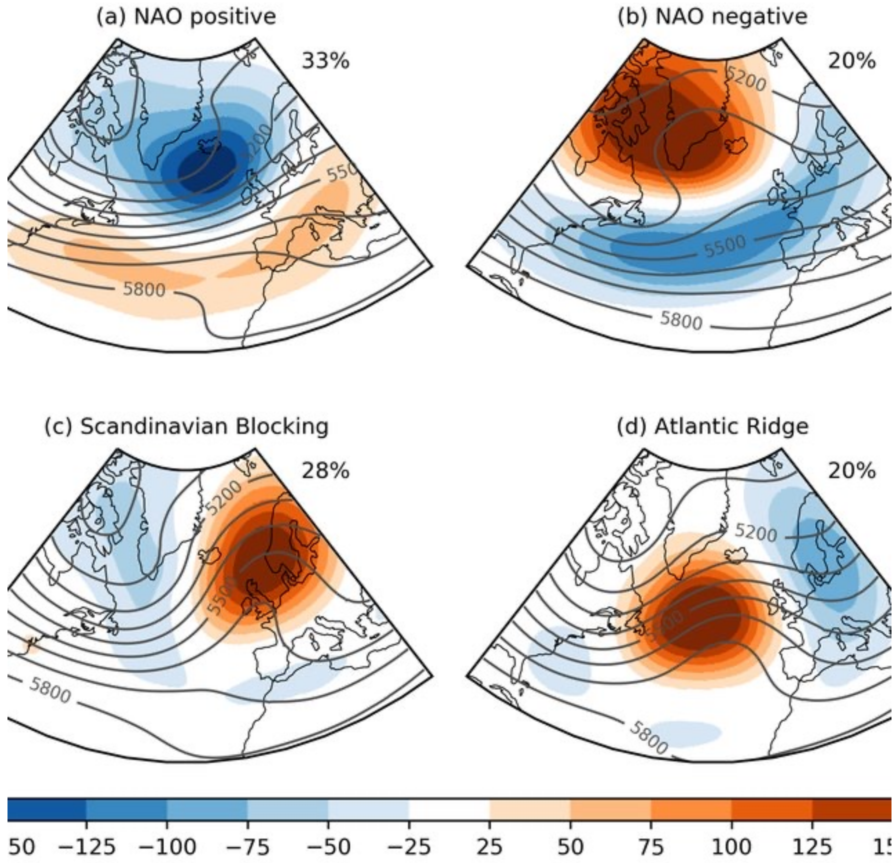
Why a new machine learning technique?

Tradition statistical **techniques to decompose automatically the weather dynamics** into simple elements work differently than human brain

They associate to each map, another map that contains a mixture of cyclones and anticyclones

⇒ **Lack of interpretability**

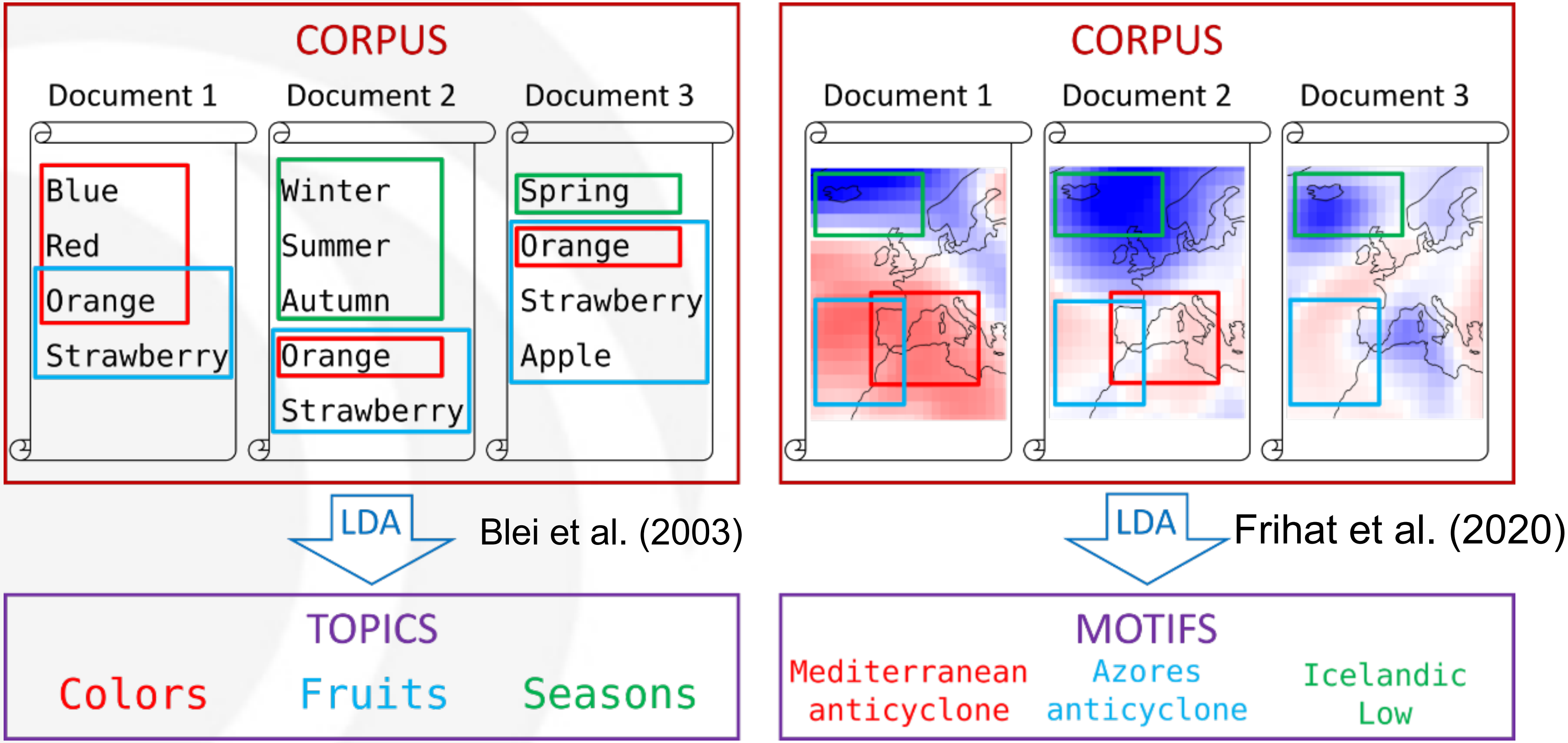
⇒ **We use the Latent Dirichlet Allocation that works as a human meteorologist**



Van der Wiel et al. 2019

Latent Dirichlet Allocation (LDA)

Generative probabilistic model used in linguistics as a topic model



Latent Dirichlet Allocation (LDA)

Blei et al. (2003)

Generative probabilistic model used in linguistics as a **topic model**

CORPUS

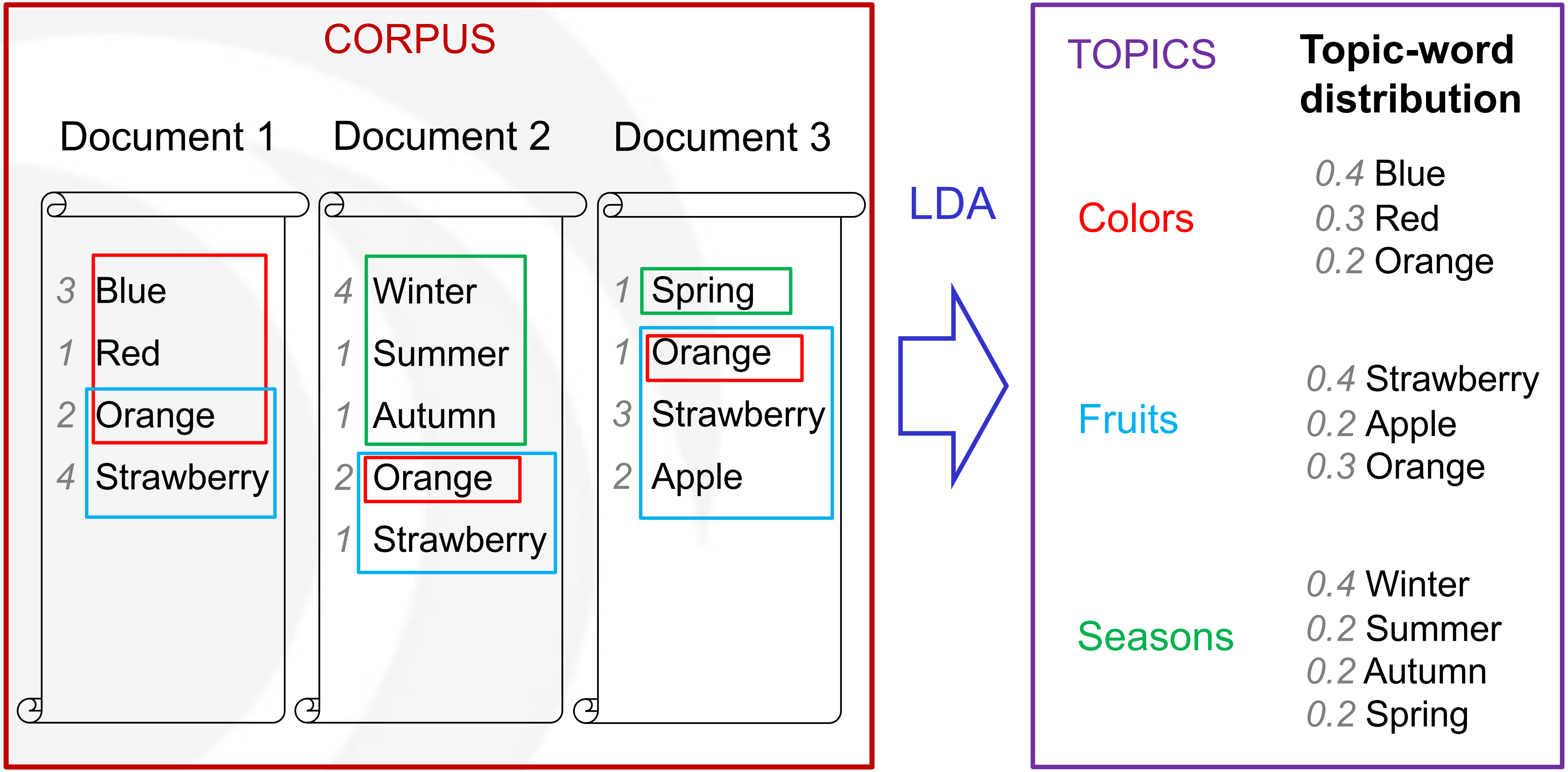
Document 1	Document 2	Document 3
3 Blue	4 Winter	1 Spring
1 Red	1 Summer	1 Orange
2 Orange	1 Autumn	3 Strawberry
4 Strawberry	2 Orange	2 Apple
	1 Strawberry	

- Documents = bags of words
- Words belong to a **finite vocabulary** $V = \{\mathrm{w_i}\ , i \in [1, n]\}$
- Documents represented by vectors d^l
- $d_i^l \in \mathbb{N}$: number of occurrences of w_i in document l

Latent Dirichlet Allocation (LDA)

Blei et al. (2003)

Generative probabilistic model used in linguistics as a **topic model**



Latent Dirichlet Allocation (LDA) Blei et al. (2003)

Moreover, documents are represented as **mixtures of topics** :

Document 1

3 Blue
1 Red
2 Orange
4 Strawberry

LDA

0.7	Colors
0.3	Fruits

Document-topic distribution

Latent Dirichlet Allocation (LDA)

Blei et al. (2003)

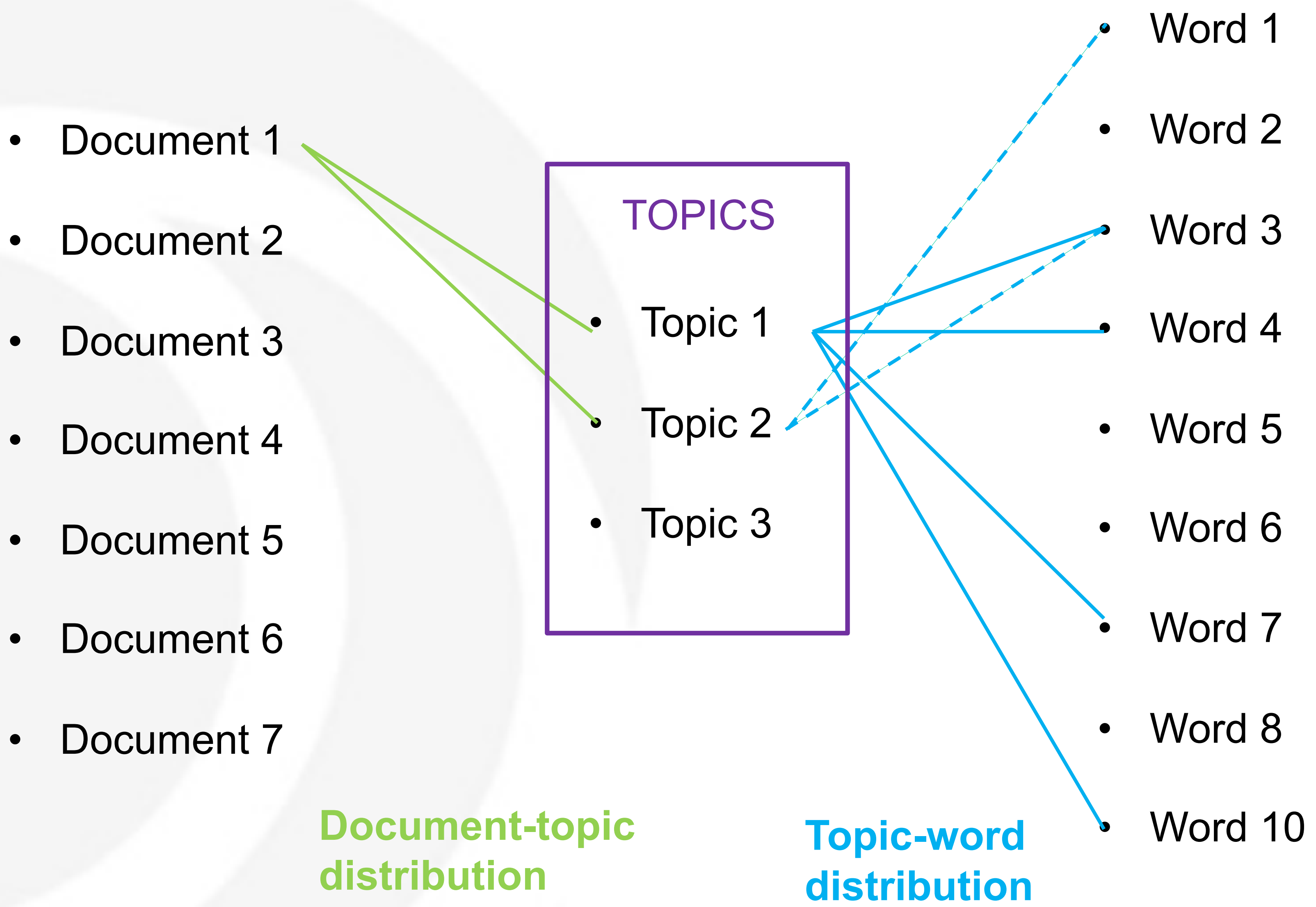
- Document 1
- Document 2
- Document 3
- Document 4
- Document 5
- Document 6
- Document 7

Words occurrences

- Word 1
- Word 2
- Word 3
- Word 4
- Word 5
- Word 6
- Word 7
- Word 8
- Word 10

Latent Dirichlet Allocation (LDA)

Blei et al. (2003)



Application of LDA to gridded observables

Analogy proposed by *Frihat et al. (2020)* :

Linguistics	Physics
Documents	Snapshots (gridded maps)
Words	Grid-points
Number of occurrences	Grid-points values
Topics	Spatial patterns (named *motifs*)

Grid-point values are converted to positive integers via rescaling, digitization and thresholding

Frihat, M., Podvin, B., Mathelin, L., Fraigneau, Y., & Yvon, F. (2021). Coherent structure identification in turbulent channel flow using latent Dirichlet allocation. Journal of Fluid Mechanics, 920, A27. doi:10.1017/jfm.2021.444

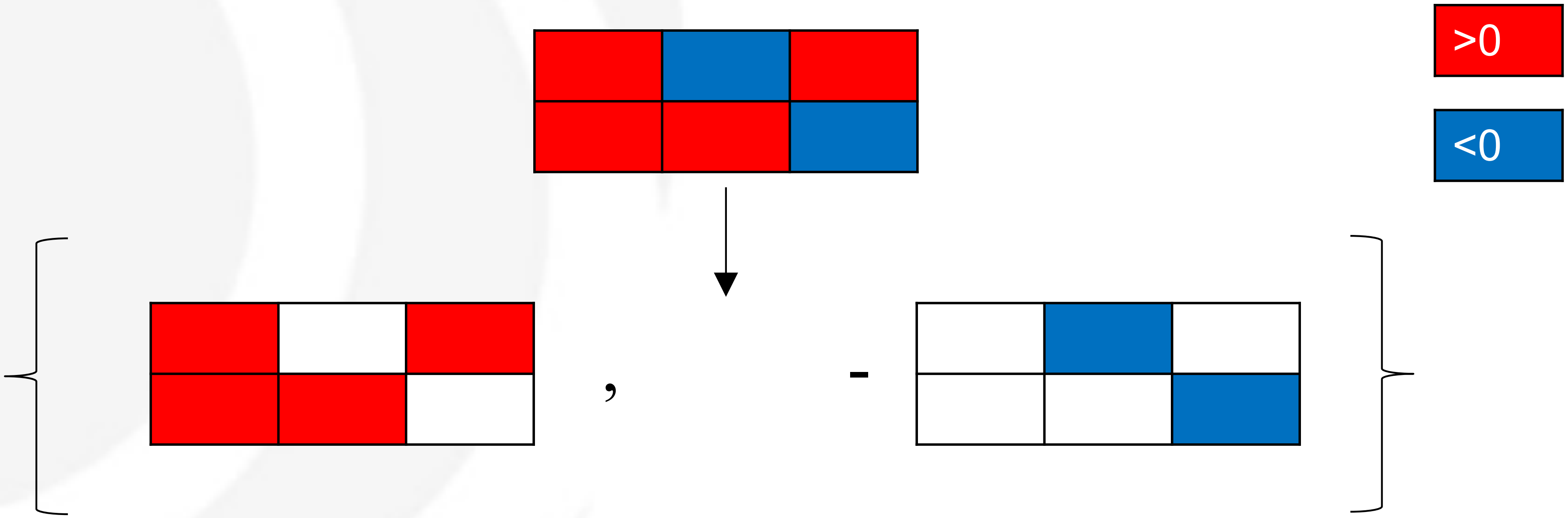
Application of LDA to gridded observables

Here applied to daily sea level pressure anomaly (NCEP/NCAR reanalysis) over North Atlantic

Need to convert our **observable values** to **positive integers** :

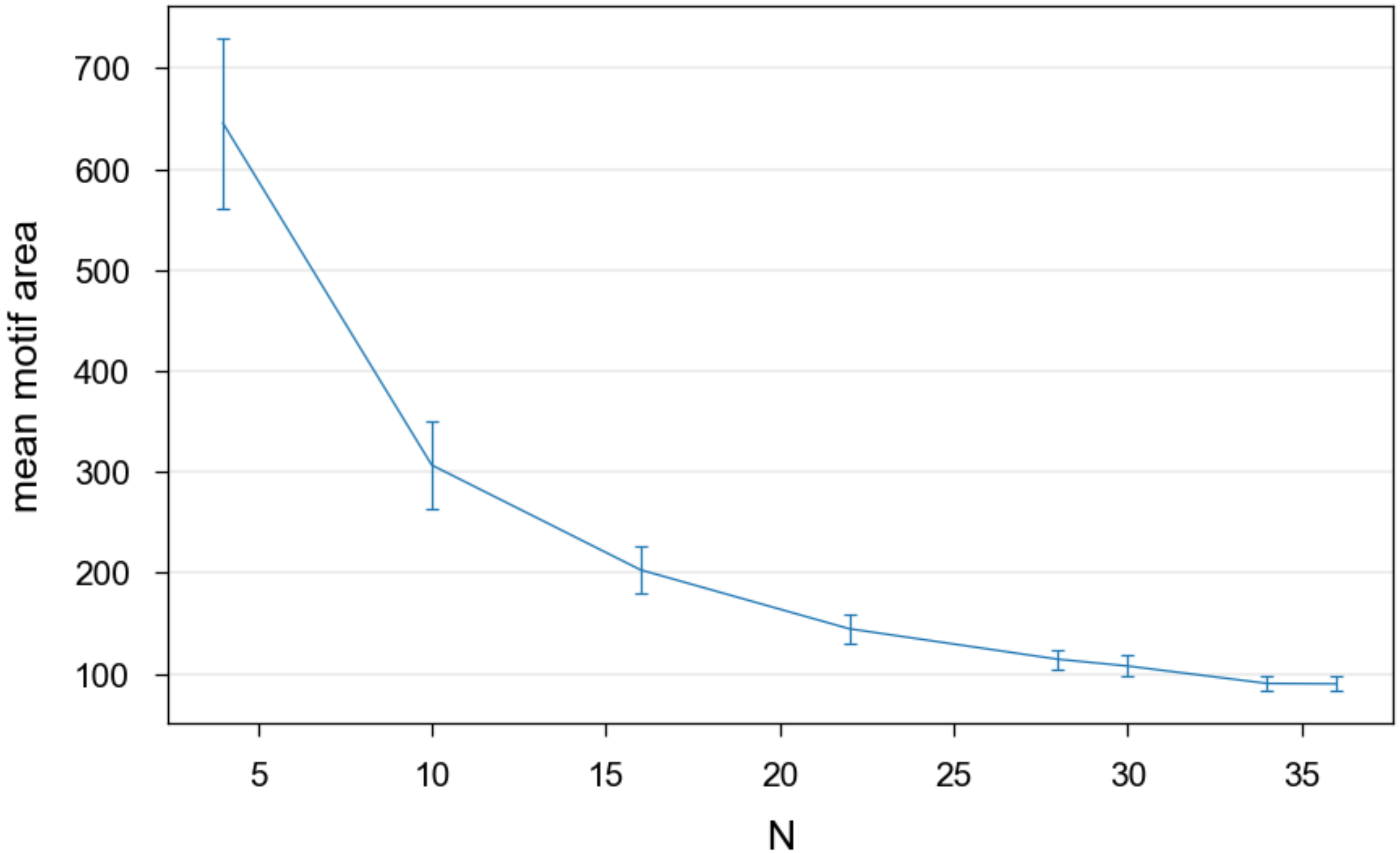
Problem : anomaly can be >0 or <0

Solution : **Split our maps in two maps** (double the ‘vocabulary’ size)



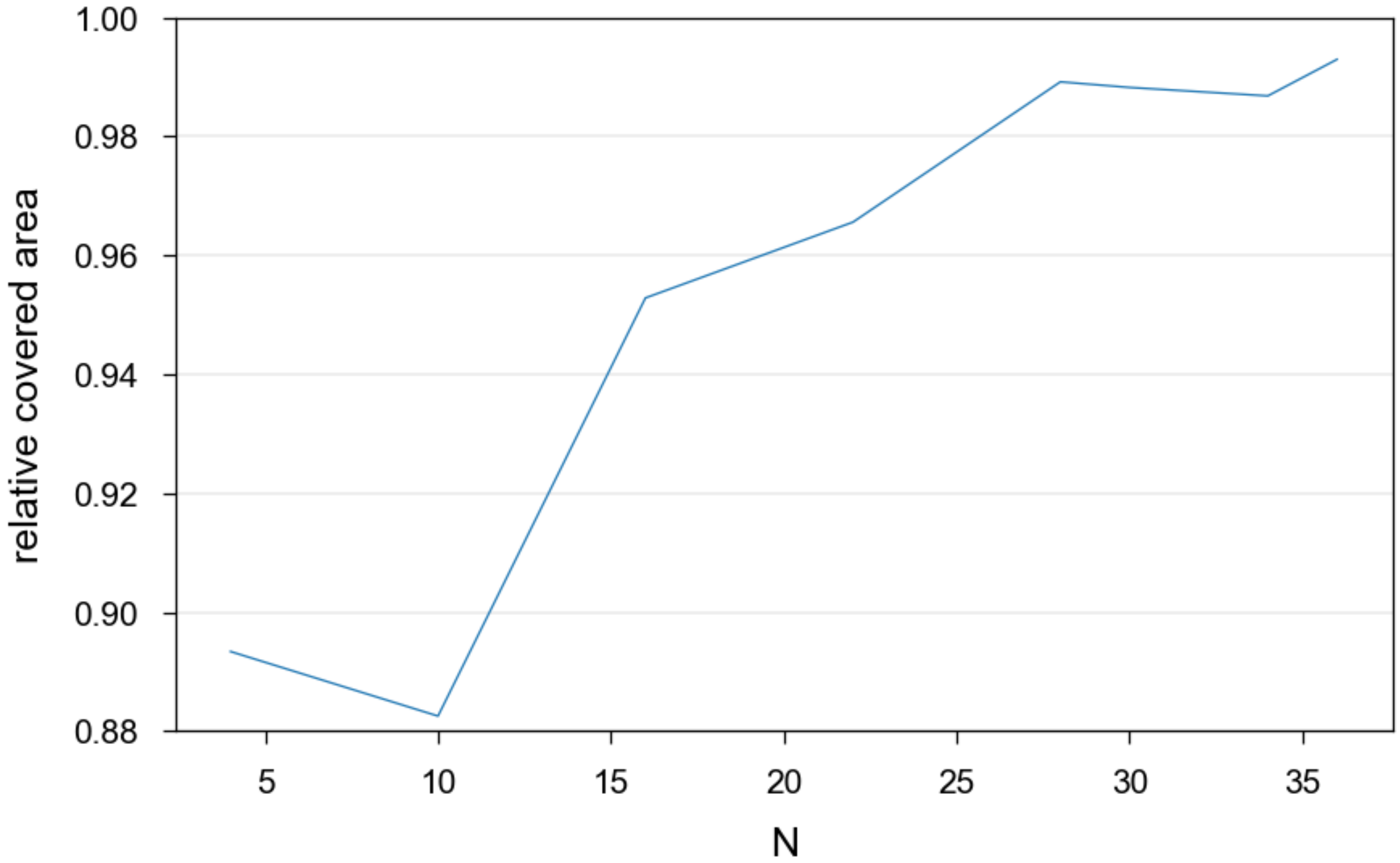
Selection of the number of motifs

The area of the motifs converges to the **typical size of cyclones and anticyclones** ($R \sim 1000 - 1500\ km$)



Selection of the number of motifs

Relative covered area **reaches 99% at 28 motifs**

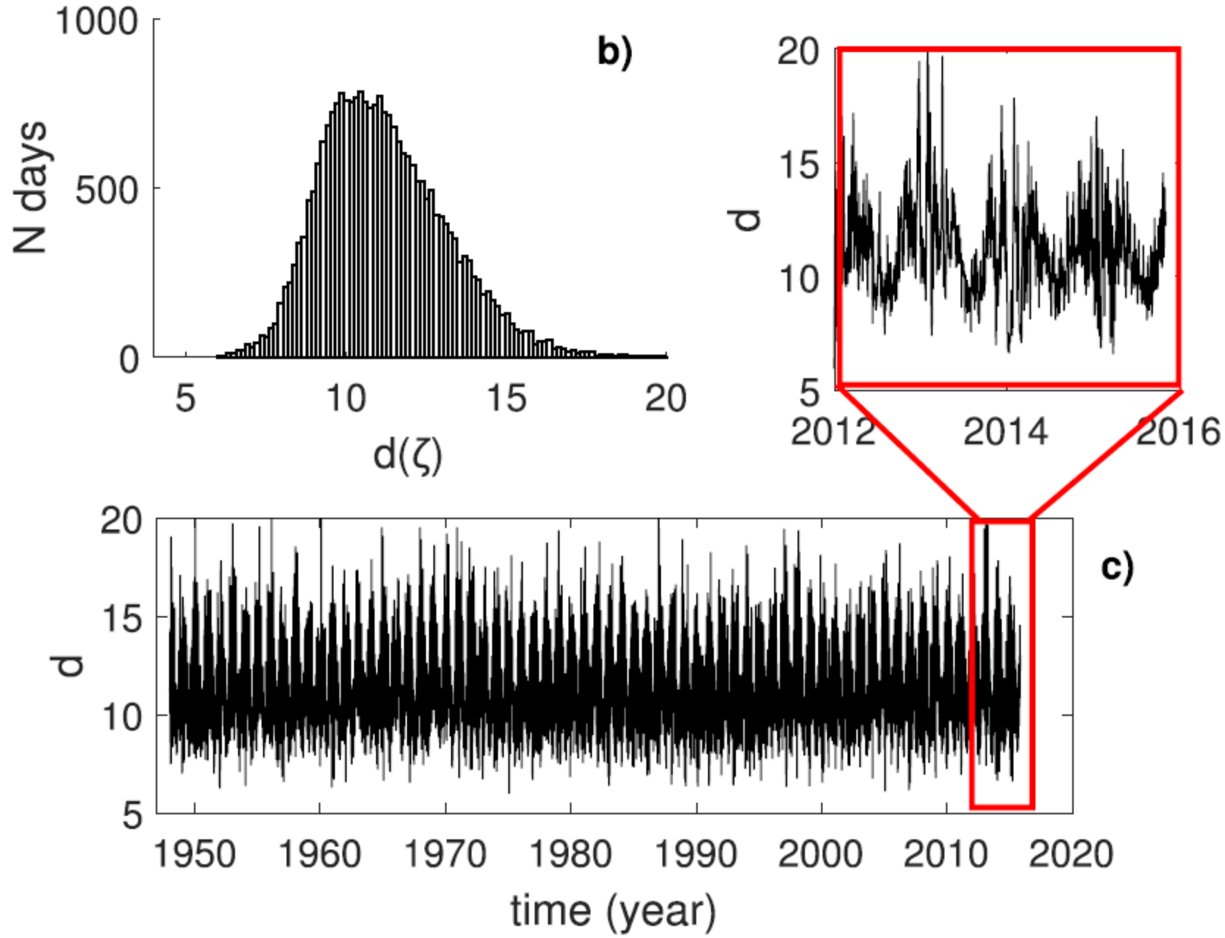


Selection of the number of motifs

We choose to **set N as 28**

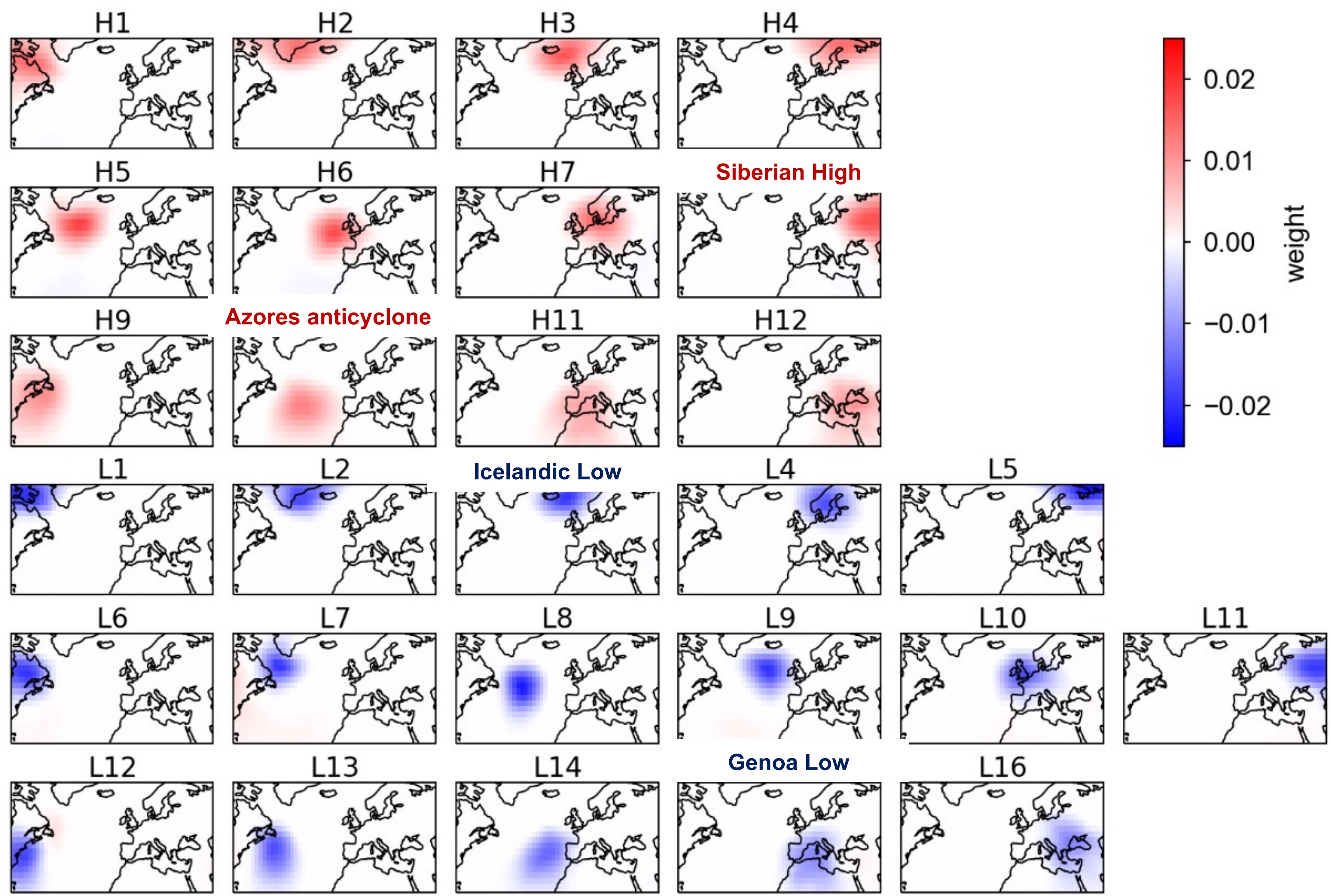
Further motivated by independent analysis on the same dataset :

28 is the **upper bound of the number of degrees of freedom** obtained computing the local attractor dimensions



Faranda, Davide, Gabriele Messori, and Pascal Yiou. "Dynamical proxies of North Atlantic predictability and extremes." *Scientific reports* 7.1 (2017): 1-10.

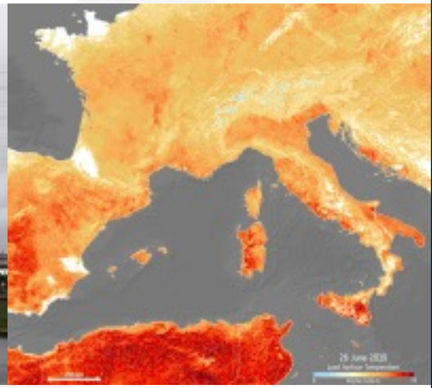
Motifs in daily pressure maps NCEP 1948-2020



Extreme events

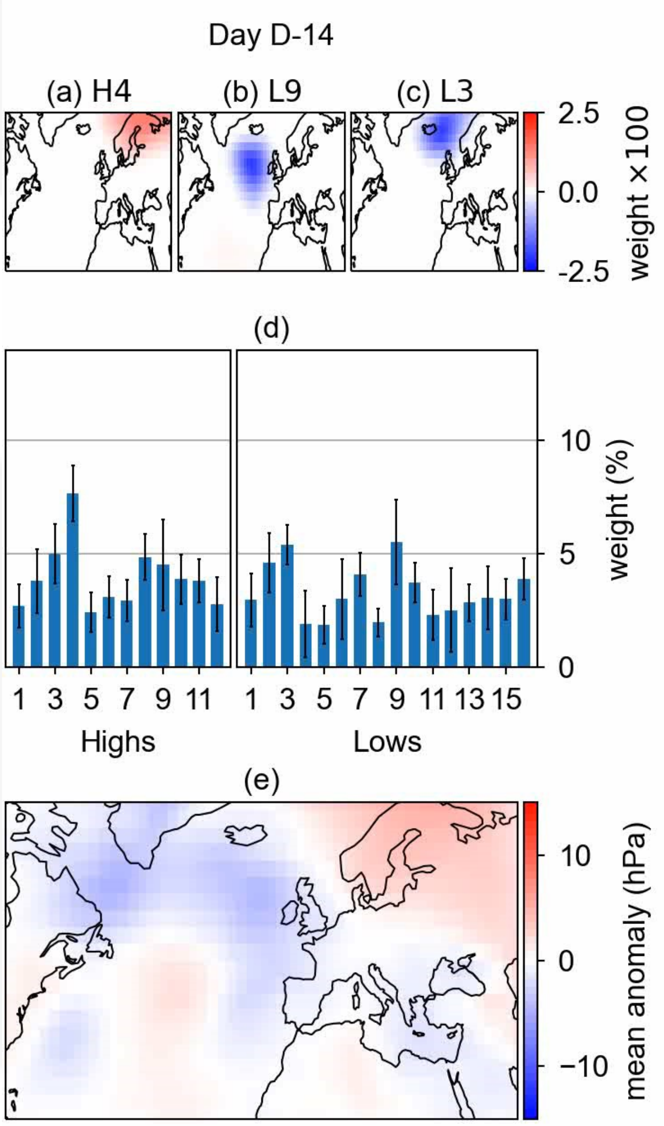
We use EM-DAT where European Extreme events are referenced **based on their impacts** (death toll, victims, state of emergency…) and **not on statistical analysis**

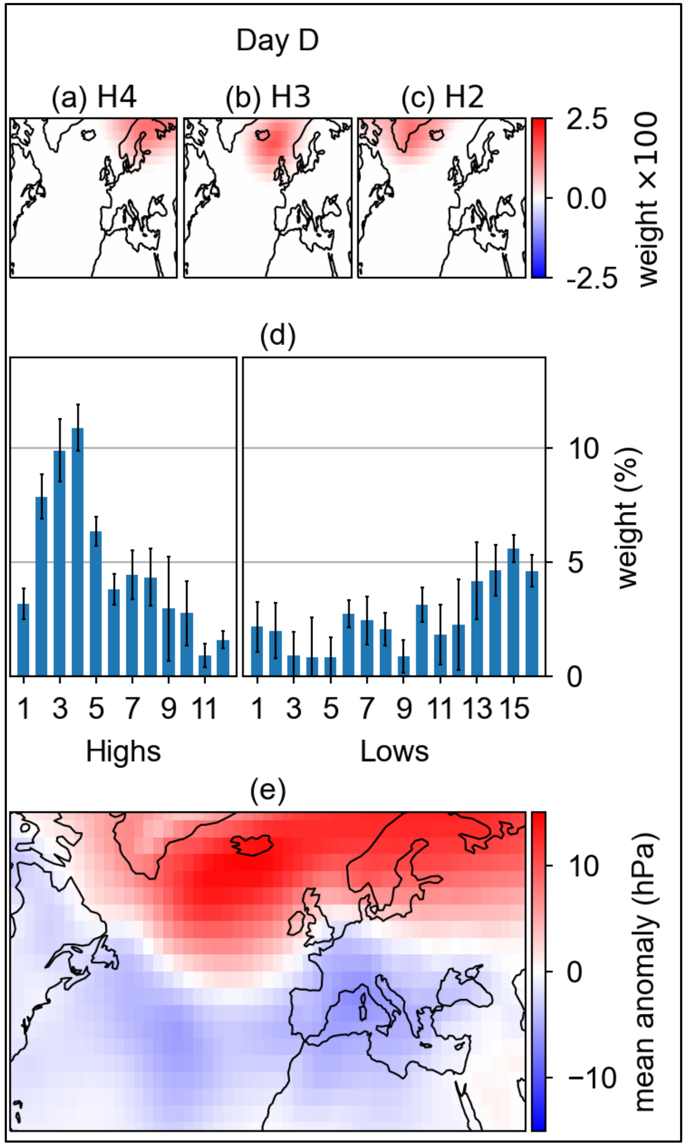
- Can these events be related to the motifs defined with LDA?
- Are there Motifs that are precursors of extreme events?



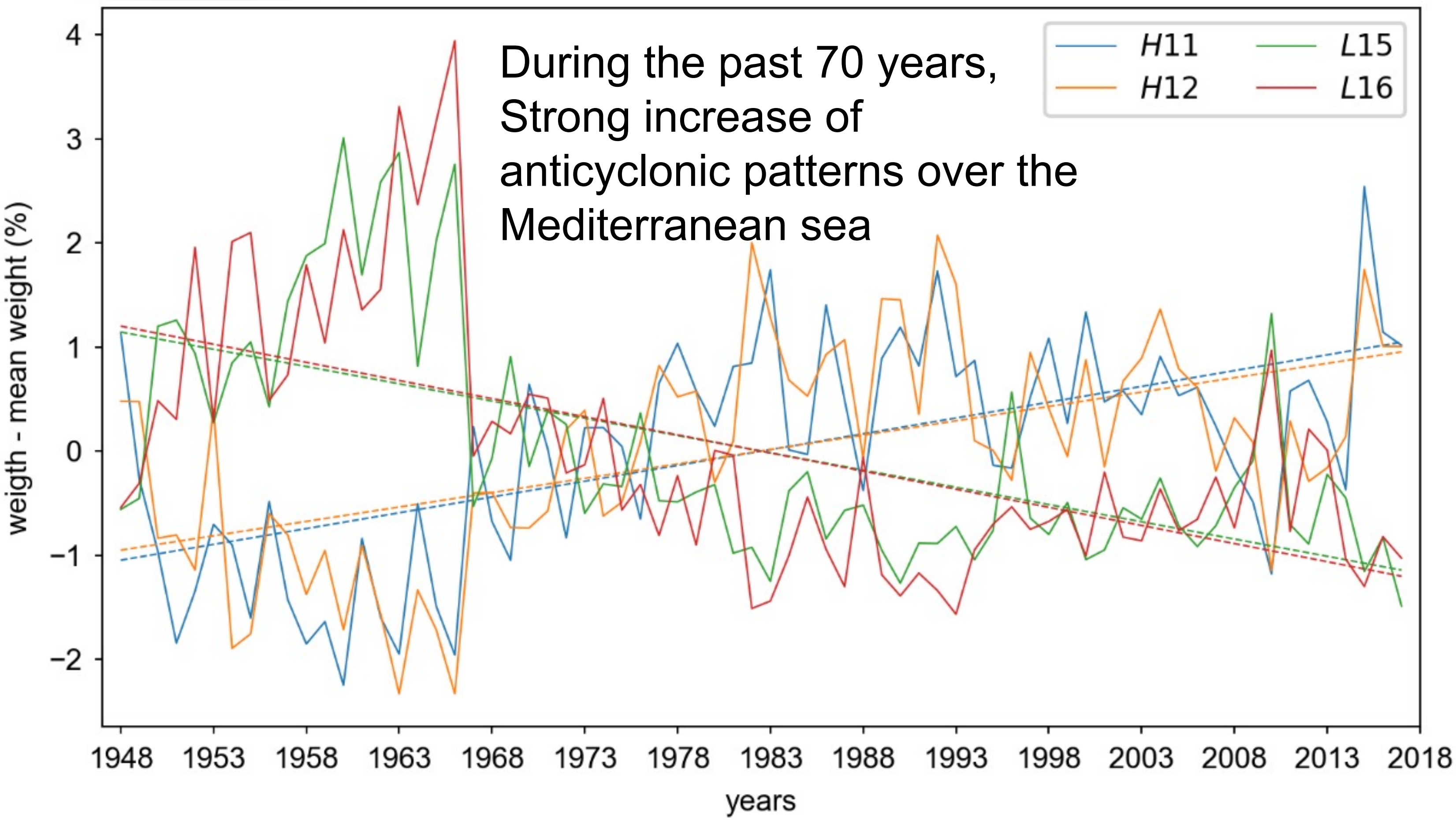
Focus on Cold Spells over Europe

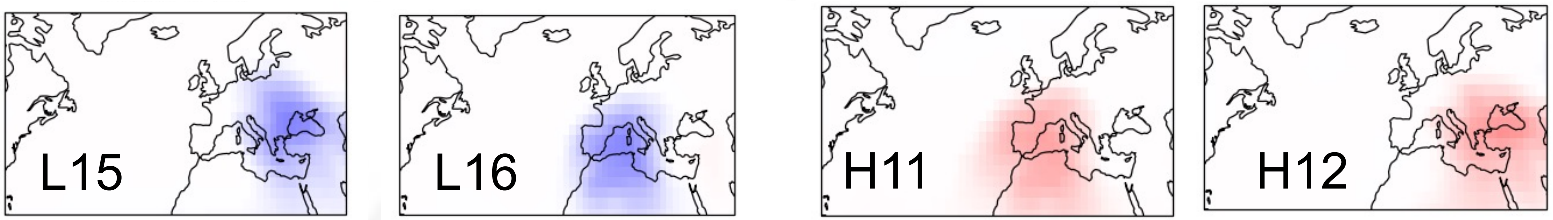
Extreme events : cold waves





LDA highlights Climate change





Conclusions

- LDA adapted and applied to **gridded atmospheric observable** (sea level pressure anomaly)
- Identification of **relevant localized patterns : cyclones and anticyclones**
- Identification of few relevant motifs and **precursors for extreme events**
- Allow to detect **trends in patterns recurrence and intensity**

Lucas Fery, Berengere Dubrulle, Berengere Podvin, Flavio Pons, Davide Faranda. Learning a weather dictionary of atmospheric patterns using Latent Dirichlet Allocation. 2022. *Geophysical Resarch Letters (in press)* https://hal-enpc.archives-ouvertes.fr/X-DEP-MECA/hal-03258523v1

Applications open: Summer school in Trieste

Artificial Intelligence for Detection and Attribution of Climate Extremes

ICTP

20 June - 2 July 2022
An ICTP Hybrid Meeting
Trieste, Italy

Further information:
http://indico.ictp.it/event/9802/
smr3717@ictp.it

During the last 5-10 years, a large number of extreme weather and climate events in Europe and worldwide have occurred, causing damage to infrastructure and casualties especially in developing countries. This has raised the question about the role of climate change in altering the odds or the magnitude of a number of such events and the new "science of attribution" has began with several attribution published all around the world. The aim of the school is to define techniques to tackle the problem of attributing meteorological extreme events to climate change by mean of machine learning technologies. Lectures will also focus on determining causal links of extreme events with the underlying climate dynamics as the atmospheric circulation. The school will also discuss and provide the bases for communicating attribution results to the general public, stakeholders and other scientists in an exact although non specialist language.

Directors:

E. BEVACQUA, Helmholtz Centre for Environmental Research - UFZ, Germany,
E. COPPOLA, ICPT, Italy
D. COUMOU, Vrije Universiteit Amsterdam, Netherland
D. FARANDA, LSCE -IPSL, CNRS, France
A. JEZEQUEL, LMD, ENS Paris, France
R. VAUTARD, LSCE -IPSL, CNRS, France
M. VRAC, LSCE -IPSL, CNRS, France
P. YIOU, LSCE -IPSL, CEA, CNRS, France

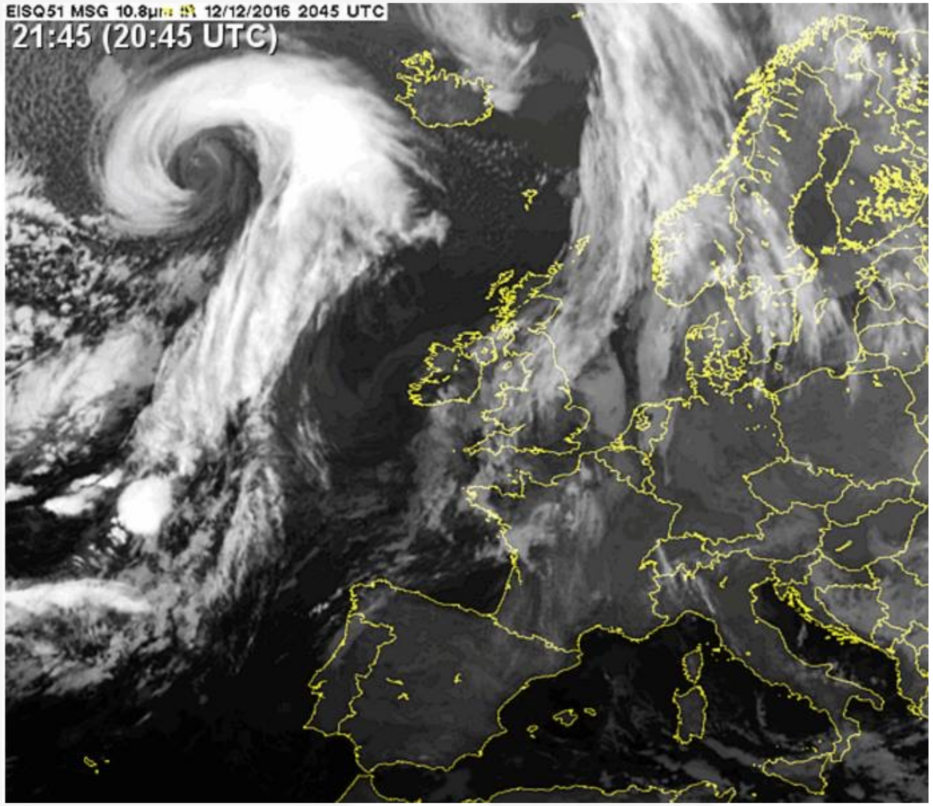
Local Organiser:

E. COPPOLA, ICTP, Italy

Speakers:

E. BARNES, Colorado State U, USA
E. BEVACQUA, UFZ Leipzig, Germany
G. CAMPS-VALLS, ISP-UVEG, Spain
E. COPPOLA, IPCC, Italy

Thank for your Attention



To keep in touch

davide.faranda@lsce.ipsl.fr

@DaviFaranda

Any Questions?

References

- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *the Journal of machine Learning research*, *3*, 993-1022.

- Frihat, M., Podvin, B., Mathelin, L., Fraigneau, Y., & Yvon, F. (2020). Coherent structure identification in turbulent channel flow using Latent Dirichlet Allocation. *arXiv preprint arXiv:2005.10010*.

- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., ... & Joseph, D. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American meteorological Society*, *77*(3), 437-472.

- EM-DAT database https://www.emdat.be/

- Preprint at https://www.researchsquare.com/article/rs-608588/v1