CGC: an open-source Python module for geospatial data clustering

Francesco Nattino

ITC – University of Twente

17 November 2022

netherlands eScience center

Contributors

Francesco Nattino

Ou Ku

Meiert Grootes

Raul Zurita-Milla

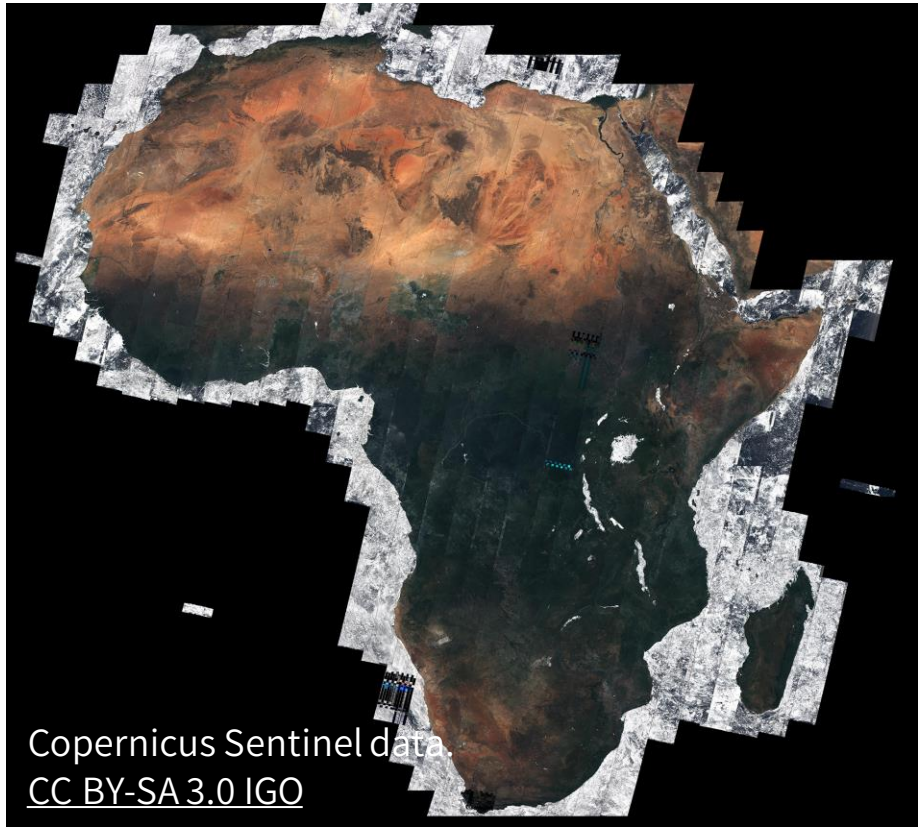
Serkan Girgin

Emma Izquierdo-Verdiguier

netherlands eScience center

Motivation and challenges

Space \ Time	T1	T2	T3	…	TN
S1	1	15	3	…	1
S2	2	65	4	…	21
S3	5	70	3	…	34
S4	7	115	4	…	2
…	…	…	…	…	…
SM	3	67	1	…	40



Clustering spatio-temporal data

netherlands eScience center

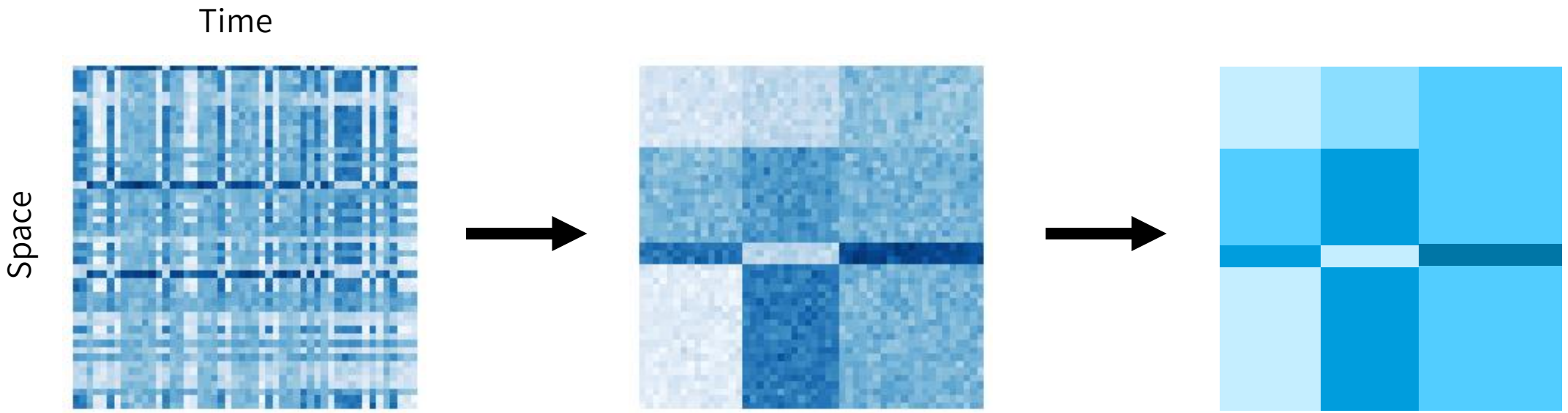
Clustering Geo-data Cubes (CGC)

- Iterative **co-clustering** (2D) and **tri-clustering** (3D) algorithms.
- Optional cluster refinement step.
- Implementations for **single machine** and **computer clusters** (Dask).
- Generic partitional clustering algorithms, targeting geospatial data.

phenology/cgc

netherlands eScience center

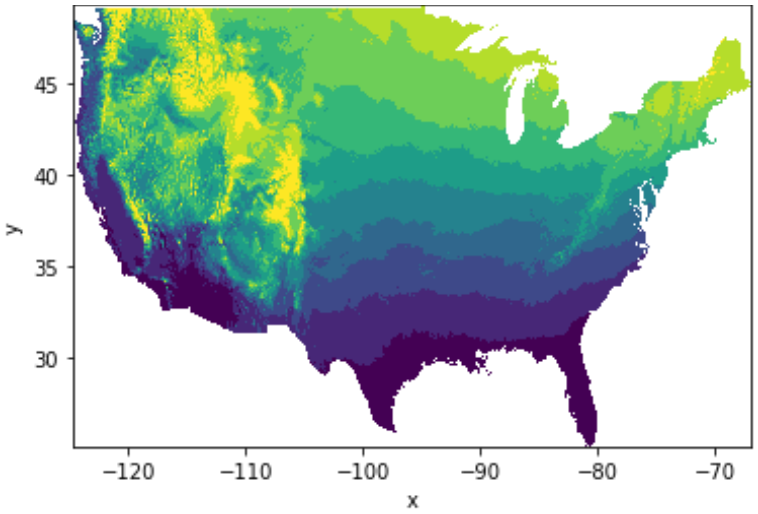
Clustering spatio-temporal data



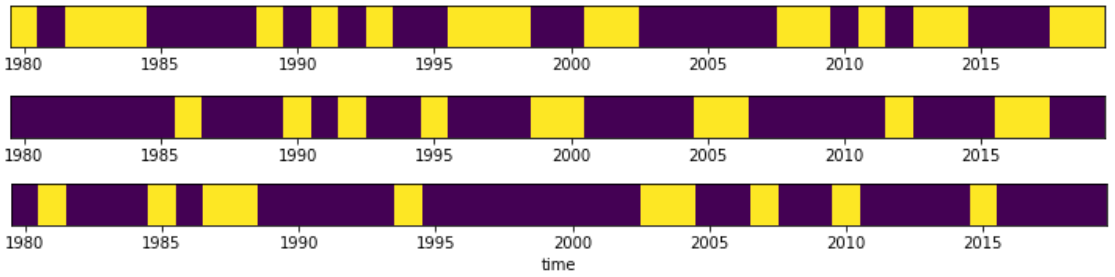
netherlands eScience center

Use case: Identifying phenoregions

- Gridded spring indices (day of first leaf and bloom) at **continental scale** and **high resolution** (1 km) for three plant species.
- Extract phenological patterns (**co-clustering**) and correlation between indices (**tri-clustering**).



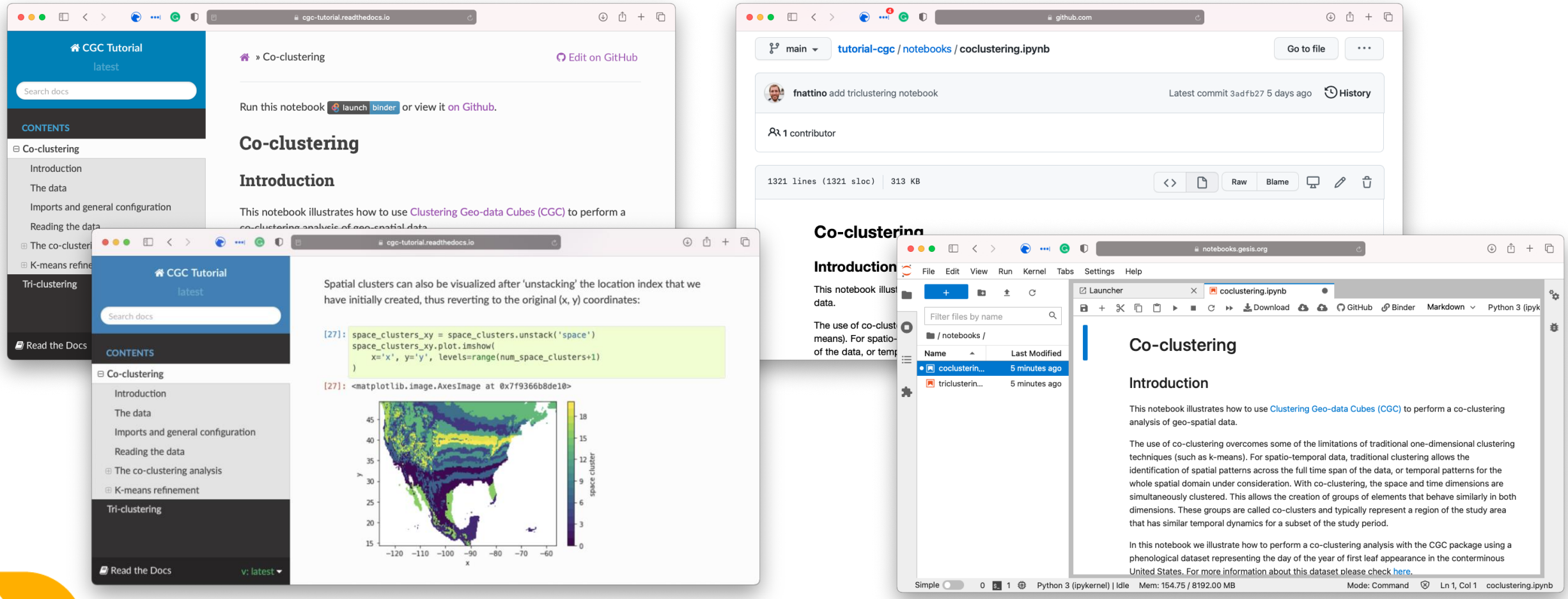
Space clusters



Time clusters

netherlands eScience center

Tutorials



https://cgc-tutorial.readthedocs.io

netherlands eScience center

Let’s stay in touch

www.eScienceCenter.nl

f.nattino@esciencecenter.nl

+31 (0)20 460 4770

phenology/cgc