ECMWF data analysis: methodology

The source of the data

This analysis has been conducted on the data made available by Copernicus and the ECMWF. We used the UERRA MESCAN-SURFEX[1] dataset, which comprises reanalysis of several climatological variables, and which contains the data on the temperature measured at 2 meters to the ground level.

This dataset has been chosen because of the high granularity and spatial resolution: the data is offered as a grid of 1,142,761 cells of 5,5x5,5 km2 of area . This made it possible to locate, as precisely as possible, the effects that climate change is having in every municipality. Also, that data is very rich, consisting of almost 97 billions data points: 1,142,761 cells, each having 4 temperature measurements per day for 58 years (from 01/01/1961 to 31/12/2018).

Once the data was downloaded through the Copernicus API services, we discussed thoroughly about the methodology to apply for the analysis.

How we obtained the variation values

First, the data was aggregated by month through a simple mean. Then, two 10-years averaging periods (1961-1970 and 2009-2018) have been calculated through a weighted mean, with the different number of days of the months being the weights.

When talking about temperature variations, it is less common to use 10-year averaging periods than 30-year averaging periods ("climate normal"). Below you can find how we came to our choice, and why our analysis can be considered sound.

Why we chose a 10-year averaging period

Looking around, it appears that the established way to talk about climate change has relied on the methods used by the scientific community. Namely, as researchers do, journalists tend to rely on the

1 https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-uerra-europe-single-levels?tab=overview

concept of "climate normal", which is defined as the "averages of climatological data computed for the following consecutive periods of 30 years: 1 January 1981–31 December 2010, 1 January 1991–31 December 2020, and so forth" (WMO, 2017: 2)[2].

There are several reasons why climate normals have become the standard for environmental investigations:

- 30-yr normals fit well all the climatological variables (besides temperature, for instance, precipitations, sea levels, etc.);
- 30-yr normals make up for occasional missing data;
- 30-yr normals are necessary for high order statistical techniques (such as extreme values and quantiles);
- 30-yr normals tend to smooth the influence of extreme events or anomalies.

(WMO, 2007)

Yet, none of the above is necessitated by this analysis.

On the one hand, being the data a reanalysis, there are no missing data-points. On the other, the ideal end of this analysis is purely descriptive, as it aims to make visible, at the local level, the magnitude of climate change.

Moreover, employing 30-yr normals, carries the burden of a complexity that is necessary in academia but not in journalism: the average reader could be overwhelmed by the amount of definitions that must be given, if not bewildered by the jargon[3], and by the counterintuitive choices that scientists tend/have to do (e.g., a decade starts from the year ending with 0 or ending with 1?). It is communicatively inefficient, hence a 10-yr averaging period has been chosen.

To support our methodology, there are various publications that advocate for the rethinking of concept of climate normal. There is indeed a body of literature that suggests how the 30-yr climate normal may not be the best way to evaluate temperature change, and study the change in temperatures.

There are two main reasons:

---

2 See, for instance: https://www.onedegreewarmer.eu/ or https://www.nytimes.com/interactive/2020/world/year-in-weather.html#16023

3 To explain what the deviation from the normal is, it must be defined what a deviation is, what a normal is, and what are the characteristics of a climate normal, not to mention how "normal", in common language, has a radically different meaning than in statistics. On this, for instance: https://scied.ucar.edu/blog/what%E2%80%99s-new-climate-normal

1. on a purely methodological stance, an averaging period shorter than 30 years does not affect the representativeness of the data. Some argued that for temperatures especially, a 10 year averaging period is as predictive (if not more) than a 30 years one (Huang et al., 1996), whilst others advocated for a redefinition of the standard averaging normal, underlying the scientific validity of shorter periods (Arguez & Vose, 2011), not to mention that "these values are intuitively attractive" (Kunkel & Court, 1990: 203). Going deeper into this, it can be affirmed that "whilst a fixed 30-year period may be appropriate as a reference period [...] shorter averaging periods (10 years or more for most parameters) perform as adequately as 30-year averaging periods". In addition to that, "10 years of data is adequate in most cases for the calculation of arithmetic means, more data are required for higher-order statistical properties such as quantile boundaries" (WMO, 2007: 26);
2. some made it explicit that a 30-yr climate normal may not be the proper methodological tool to capture the changes in temperature: "WMO-recommended 30-yr normals, even updated every 10 yr [...] not only have little relevance to the future climate, but are more and more often unrepresentative of the current climate", as "a direct result of rapid changes in the global climate over approximately the last 30 yr that most climate scientists agree will continue well into the future" (Livezey et al., 2007: 1771). To put it simply , 30-yr averaging periods are not representative of the major changes in temperatures that have been witnessed in the last few years and that, being a byproduct of human activity, are not going to go down anytime soon[4].

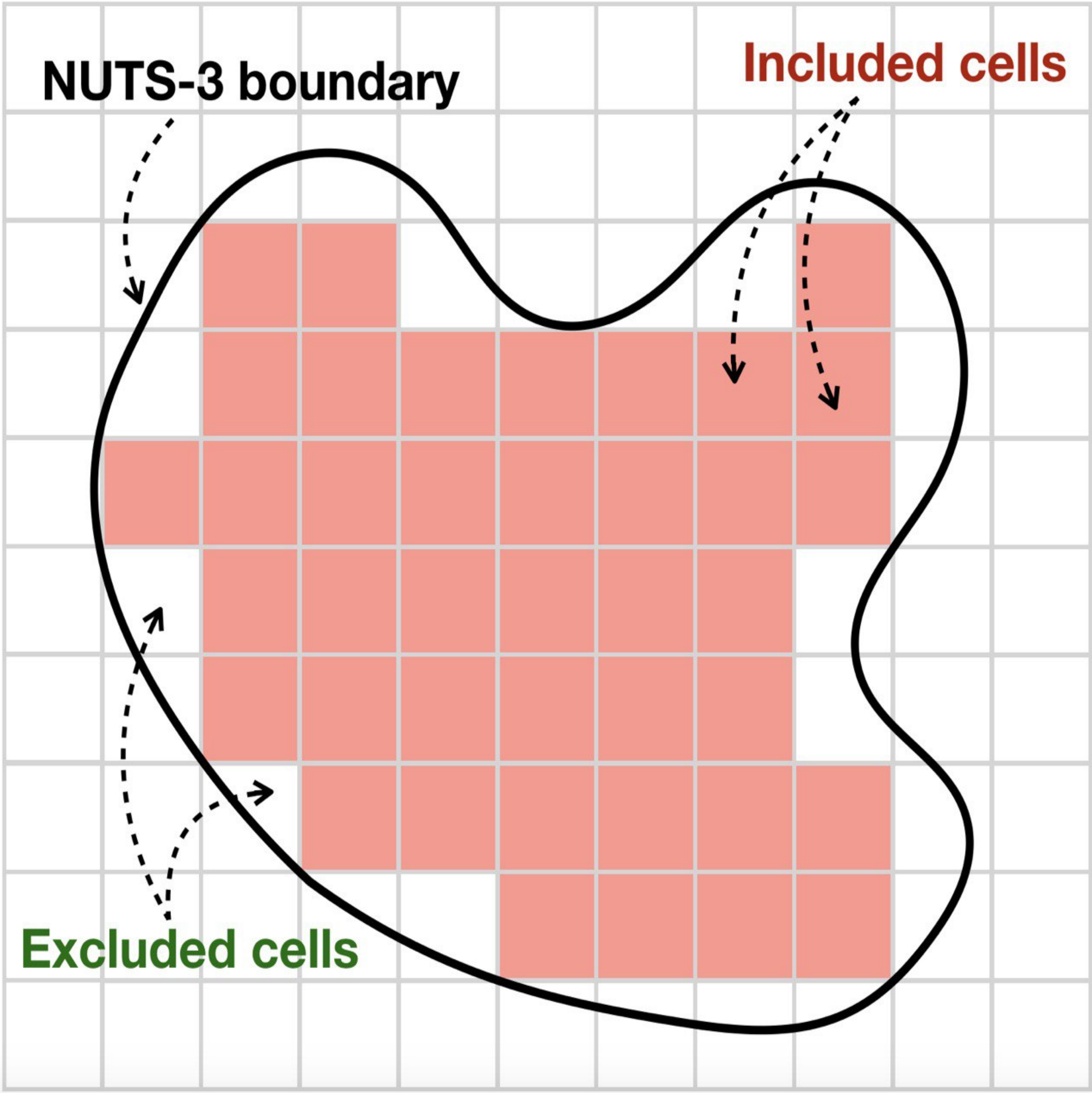
How we located temperature increases

Cells of 5,5x5,5 km2 of area do not say much to citizens. That's why we did some spatial analysis in order to connect those data with sub-regional (NUTS-3) and municipal areas: this way, we made it possible for citizens to be aware of the increase in temperature experienced in the place where they live.

Sub-regional level: Italian "province"

When it came to the NUTS-3 regions level, the methodology has been quite simple. Considering the boundaries of a given *provincia*, the mean temperature variation has been calculated as the mean of the grid cells that completely fall inside the boundaries, as exemplified by in the picture below.

4 We conducted anyway an analysis using climate normals (namely comparing the period 1961-1990 to the 1991-2018 one) in order to ascertain the variance between the two methodologies. On average, our methodology shows a variation that is 30% higher than the one obtained with wider time spans. Given that every dacade is hotter than the one before, and building on the academic corpus that argues how temperature rise is going to proceed steadily, we decided to stick with our methodology as it appears more accurate in regards to the phenomenon we wanted to analyze.

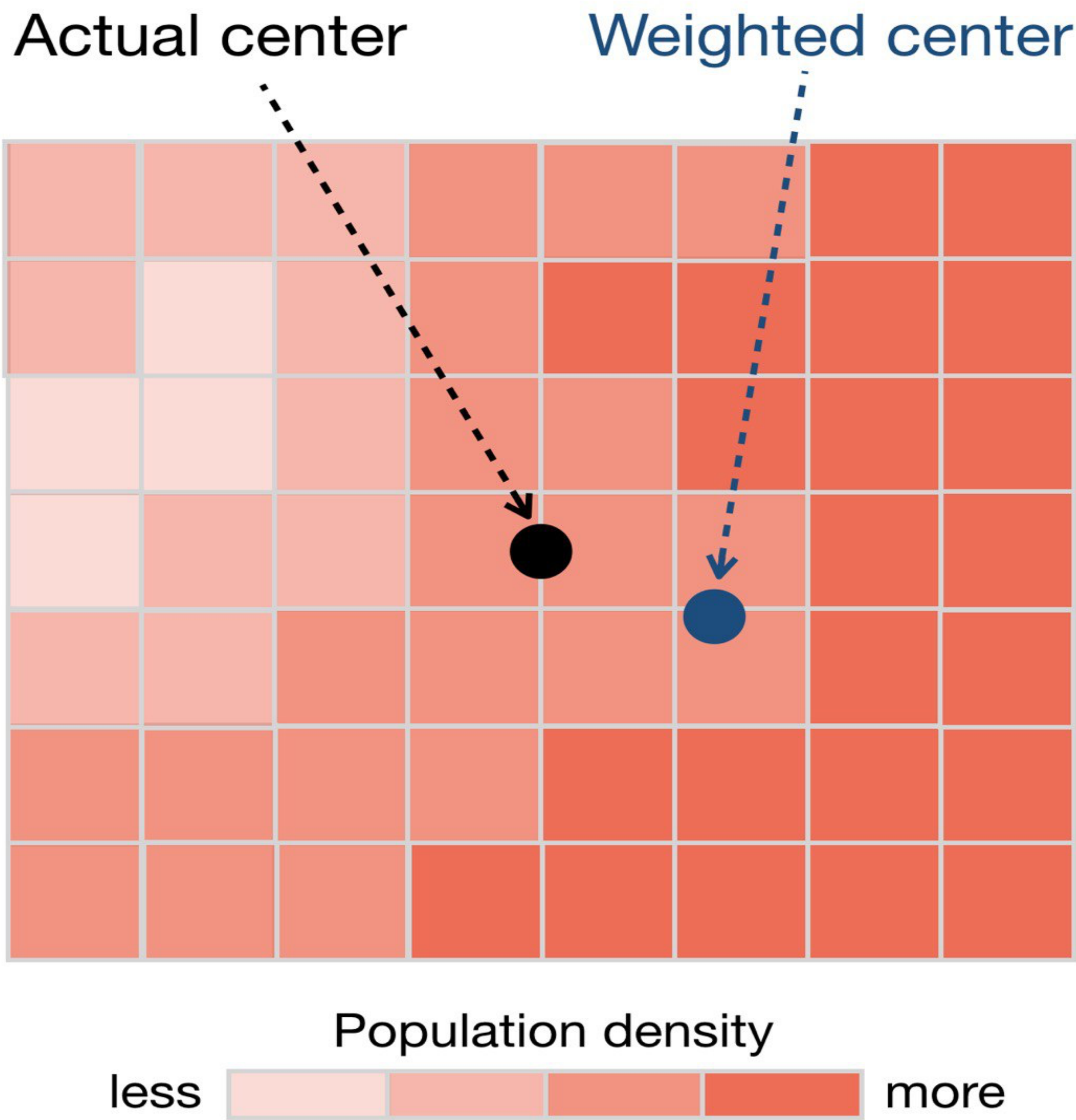


Municipalities:

For what concerned the municipalities, things became trickier for two reasons:

1. there are many municipalities that contain more than a grid cell, and many others that are too small to even contain a single one (e.g. municipality of area < 30,25 km2);
2. there is a number of municipalities that contain in their boundaries a portion of territories such as mountains, lakes, and other habitats with different inherent characteristics.

There was the need to use the same method for all the municipalities, whether they be large or small, and that had to consider the territorial diversity.
To obtain a single value, we associated the city center coordinates of every municipality to the closest grid cell of the available data (having thus a common method for any city).
Yet, to calculate the city center, instead of using the simple geometric function to calculate the exact center of the polygon, we decided to weight the center based on the population grid, so to have the center of the municipalities polygons closer to the urban areas. Indeed, when thinking about a city, the average person depicts in its mind the urban area of the city center, generally the most densely populated.
For many cities, especially smaller ones, this makes no difference, while for larger ones the difference may be significant. The figure below exemplifies this.



In the end, it has been possible to have an estimate of the impact of climate change to every italian municipality.