POLITECNICO MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE

Non-Parametric Statistics Project

Master of Science in Mathematical Engineering.
Course : Non-Parametric Statistics.

Author: **Simon Trottier & Noé Debrois**

Student ID: 246573 & 242751
Advisor: Prof. Simone Vantini
Co-advisors: Prof. Francesca Ieva, Dr. Andrea Cappozzo, Dr. Alfredo Gimenez Zapiola
Academic Year: 2023-2024
Github : `https://github.com/NoeDebrois/NPS-Project-2023-2024.git`

Abstract

This project presents a non-parametric statistical analysis conducted in collaboration with an NGO focused on climate education. Utilizing meteorological data spanning from 1978 to 2022, the study employs bootstrap hypothesis testing to examine trends in average monthly temperatures worldwide, aiming to reject the null hypothesis of a zero slope coefficient and thereby confirm global warming. By extending the analysis to encompass the entire planet through regression techniques, significant statistical zones are delineated using p-value cartography. Additionally, predictions for future temperature trends are made using conformal prediction methods, offering insights into potential climate scenarios in various regions over the coming decades. While acknowledging the inherent limitations of temperature-based analyses, particularly in capturing the complexity of climate dynamics, the study aims to present results in a comprehensible manner accessible to the general population. Beyond assessing global temperature rise, the project explores the diverse manifestations of climate change, including heightened frequency of extreme weather events such as precipitation and temperature extremes. The findings underscore the multifaceted nature of climate change and the urgent need for informed action.

Keywords: Climate Change, Non-parametric Statistics, Bootstrap Hypothesis Testing, Regression Analysis, Extreme Weather Events

Contents

Abstract **i**

Contents **iii**

Introduction & Goals of the project **1**

1 Assessment of the global warming **3**

1.1 Collection, study and pre-processing of the data 3

1.1.1 What data did we collect and for which purpose ? 3

1.1.2 How did we process the data to build our own database ? 4

1.2 Is there a shift in temperature from 1978 to 2022 ? 5

1.2.1 At the station scale 5

1.2.2 A "world map of statistical significance" 7

1.2.3 How to expand the shift results to the whole world despite the lack of stations ? 14

1.3 Forecasting of the global warming 18

2 Climate change is also about extreme events **21**

2.1 Data Preparation 21

2.2 Methodology 22

2.3 Detection of extreme events linked with temperature 23

2.4 Detection of extreme events linked with precipitation 24

3 Conclusion **27**

Bibliography **29**

Introduction & Goals of the project

Climate change stands as one of the most pressing challenges of our time, with far-reaching implications for ecosystems, economies, and human well-being. As the scientific community strives to comprehend the complexities of this phenomenon, there is a growing imperative to communicate its impacts in a manner accessible to all. In response to this need, our project endeavors to employ non-parametric statistical methodologies to elucidate the trends and manifestations of climate change, bridging the gap between scientific research and public understanding. Indeed, our work can be seen as a mission for a climate advocacy NGO, informing the public about the impacts of climate change and promoting beneficial actions.

Focusing on meteorological data spanning from 1978 to 2022, our study delves into two interconnected aspects : the intricate interplay between temperature shifts (analyzed by comparing monthly data points) and the growing frequency of extreme weather events (examined through daily data). Leveraging non-parametric statistical techniques, particularly bootstrap hypothesis testing, we scrutinize the trajectory of average monthly temperatures against time, seeking evidence to support the hypothesis of a non-zero slope coefficient — a hallmark of global warming.

Expanding beyond the data from 1 315 weather stations, we extrapolate our findings to a world scale through regression, thereby offering insights into temperature trends across diverse geographical regions. Complementing this analysis, we deploy p-values computation techniques [2] to highlight statistically significant zones, providing a clear visualisation of climate change's spatial heterogeneity.

Furthermore, our study transcends the singular focus on global temperature rise, embracing the multifaceted nature of climate change. By investigating the incidence of extreme weather events, such as heightened precipitation or heatwaves, we underscore the diverse manifestations of climate variability. Notably, regions where the statistical significance of global warming is pronounced often exhibit an increased frequency of extreme weather events, amplifying the urgency for localized adaptation and mitigation strategies [1], [3].

In addition to retrospective analyses, our project adopts a forward-looking approach, employing conformal prediction methods to forecast temperature trends in the coming decades. By integrating these predictive insights with our spatial mapping framework, we offer stakeholders, policymakers, and the general public (even those without a scientific background) a glimpse into the future climate scenarios that may unfold in their respective regions.

While acknowledging the inherent limitations of simplistic temperature-based analyses in capturing the full complexity of climate dynamics, our endeavor underscores the power of statistical tools in translating complex scientific insights into digestible formats. Through visually engaging graphs and accessible presentations, our aim is to empower individuals, irrespective of their scientific background, to grasp the evolving climate realities and take informed action towards a sustainable future.

1 | Assessment of the global warming

1.1. Collection, study and pre-processing of the data

1.1.1. What data did we collect and for which purpose ?

For the first part of our project, focused on estimating global climate warming, we collected data from the Meteostats database (`https://meteostat.net/it/`) spanning from January 1st, 1978, to December 31st, 2022. These data include monthly averages of minimum, maximum, and mean temperatures, as well as precipitation levels.

In the second part of our analysis, where our objective was to investigate the frequency of extreme weather events such as heavy precipitation (potentially leading to flooding [1]) or extreme heat spikes (indicative of heatwaves whose characteristic time is of a few days, and potentially leading to fires [3]), we gathered daily data from the same database. Given the characteristic time scales of these events, daily frequency data were deemed appropriate. Again, we collected information on minimum, maximum, and mean temperatures, as well as precipitation.

It's worth noting that the data were sourced from 1315 weather stations distributed across the globe (we generated a grid and chose the closest station). While this constitutes a significant number, it's important to acknowledge the geographical concentration of these stations, as illustrated in the map in Figure 1.1.

Stations are primarily clustered in regions including North America, Western Europe, Sub-Saharan West Africa, and the Southeast Asian archipelago. Notably, there is limited data available for Russia, despite its vast land area, as well as Greenland. Moreover, data from Asia, particularly around China, are scarce.

All temperature data are expressed in degrees Celsius, while precipitation levels are recorded in millimeters.

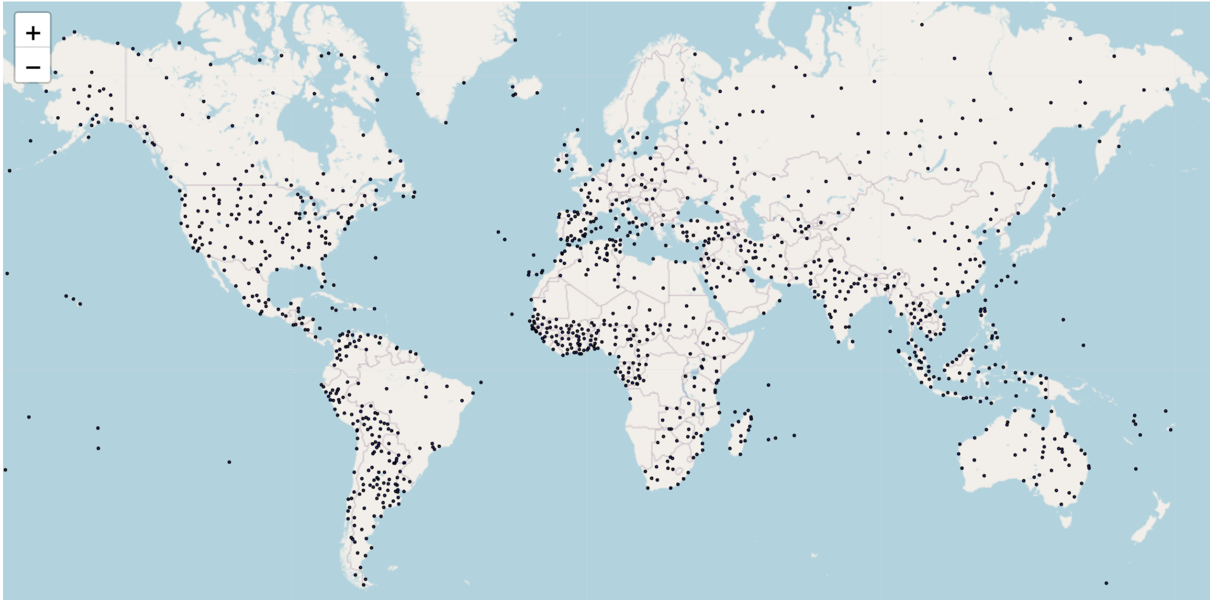


Figure 1.1: Distribution of the 1315 selected weather stations across the globe

1.1.2. How did we process the data to build our own database ?

For the following statistical study, we decided to keep the monthly average temperature and the monthly average precipitation data from the 1315 stations across the world. More specifically, our Python script created for each station a csv file which contains these data. Here is an image showing how those csv files look like :

Time	Tavg	Tmin	Tmax	Prcp
1995-08-31	22.76	17.52	27.87	3.63
1995-09-30	17.45	12.27	22.73	3.16
1995-10-31	13.35	8.01	19.45	0.56
1995-11-30	6.92	2.88	11.2	1.41
1995-12-31	3.62	1.01	6.61	5.2
1996-01-31	3.53	0.77	6.39	3.86
1996-02-29	2.82	-1.07	6.36	2.18
1996-03-31	6.05	2.01	10.19	0.71
1996-04-30	13.14	8.13	18.37	2.47
1996-05-31	17.62	11.54	23.06	2.5
1996-06-30	21.97	15.77	27.37	2.16
1996-07-31	22.35	16.57	27.16	2.49
1996-08-31	21.77	16.62	27.16	3.81
1996-09-30	16.01	11.31	21.2	1.32
1996-10-31	12.71	8.9	16.94	3.81

Figure 1.2: Our Database for Verona ; One point per month ; We'll mainly work with Tavg and Prcp in this part of the study

1.2. Is there a shift in temperature from 1978 to 2022 ?

1.2.1. At the station scale

First, we conducted an analysis to quantify the effects of climate change on monthly average temperature at the 1315 weather stations. The bootstrap analysis focused on monthly average temperature data obtained from the 1315 stations, which were prepared as explained before, and was designed to test whether there was a shift in monthly average temperature, for each twelve months, from 1978 to 2022 (e.g for each station, we compared each January average temperature, over the 45 years of the study).

Using linear regression analysis, we explored the relationship between monthly average temperature and time. To ensure the reliability of our findings, we employed the Bootstrap method, a robust resampling (with replacement) technique. This method allowed us to obtain a pointwise estimation of the regression coefficients, as well as reverse percentile confidence interval, providing a measure of uncertainty around the estimated coefficient.

These confidence intervals were crucial for assessing the significance and variability of the observed trends. If a confidence interval did not contain zero for the estimated coefficient, it indicated that we could reject the null hypothesis stating that this coefficient is equal to zero. So in our context, if the confidence interval does not contain zero, it suggests that the slope of the month-by-month temperature shift is likely different from zero.

Here is the detailed methodology we followed to build such a study :

- *Linear Regression Analysis:* The code utilizes the ordinary least squares method to estimate the relationship between the year and the average temperature for each month. In particular, we estimated the slope.
- *Bootstrap Method:* To quantify the uncertainty on the estimated slope, the code employs the Bootstrap method. It generates Bootstrap samples by resampling with replacement the residuals from the initial model, then estimates the slope for each Bootstrap sample. These Bootstrap slope coefficients are used to construct confidence intervals.
- *Reverse Percentile Confidence Intervals:* Reverse Percentile confidence intervals are calculated using Bootstrap quantiles for each slope coefficient. These intervals capture the uncertainty surrounding the slope estimates.
- *Exporting Results:* The results are stored in a three-dimensional array (RP) and

exported to a CSV file for further analysis or visualization.

Using the exported CSV, we were able to see on a world map the stations that recorded a statistically significant (using the previously explained test) monthly average temperature shift from 1978 to 2022, for each of the twelve months. For the stations with no significant slope, we considered there was no monthly average temperature shift. For the other significant stations, and for each of the twelve months of the year, we were even able to get a numeric estimation (as well as a confidence interval) of this shift slope.

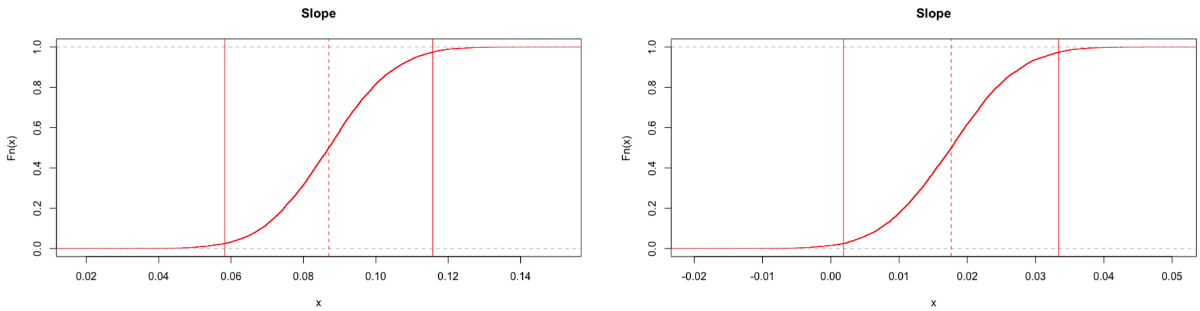


Figure 1.3: Pointwise estimation and Bootstrap Reverse Percentile Interval for the slope in the linear model of monthly average temperature VS year, for June, in Verona (left), in Cape Town (right)

So there is a shift in temperature in both Verona (for example in Verona it is about +0.08 degree per year if we look at the pointwise estimate) and Cape Town in June, but this shift seems to be way stronger in Verona than in Cape Town. We will discuss more about this later.

Above we showed an example for two particular stations at one particular month. But what is great is that we can compute these estimations (pointwise and uncertainty quantification) for every station on earth and at each and every month. Take a look at the maps we are able to produce, with the dot size corresponding to the value of the shift (bigger the dot, bigger the shift) and the dot color corresponding to the statistical significance (we will talk more about this later).

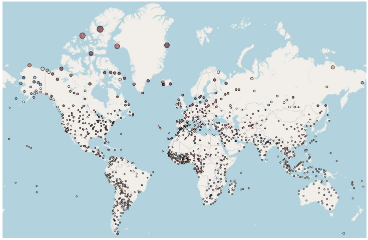
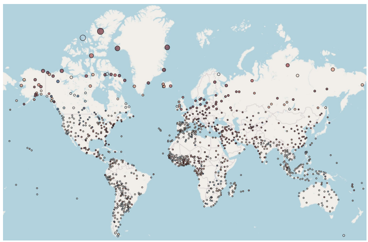
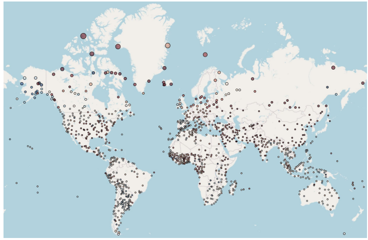
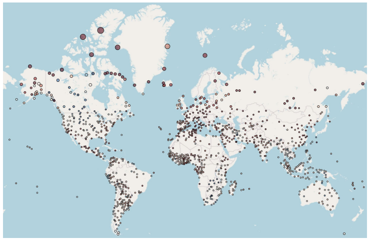
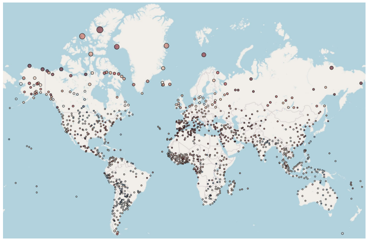
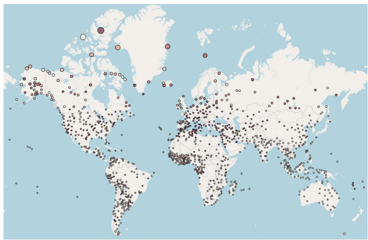
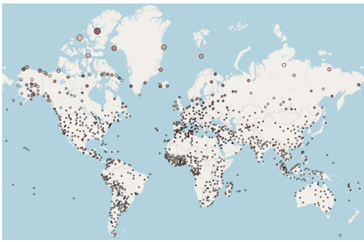
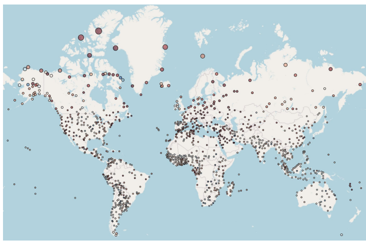
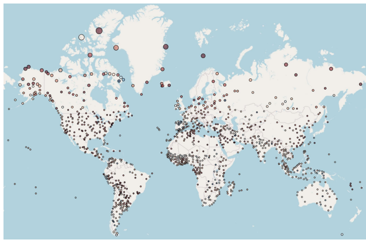
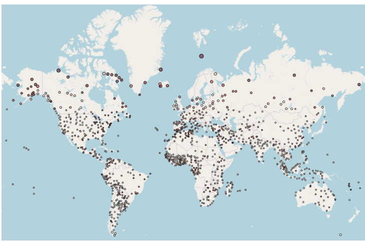
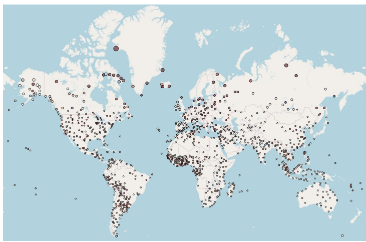
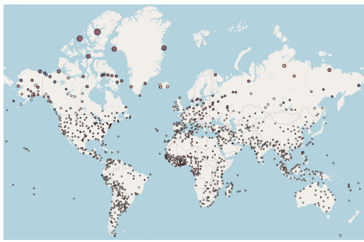

January

February

March

April

May

June

July

August

September

October

November

December

Figure 1.4: Estimated shift per station

What is clearly visible is the difference in the intensity of this shift (see the dot size) when we look at different places on earth. This is a well known fact about global warming : the warming is not the same when we look at an equator point or next to the poles.

1.2.2. A "world map of statistical significance"

In order to detect the places on earth were the climate change is significant statistically speaking, we first decided to use the same method as in [2]. This consists in three main steps :

- The first one consists in assigning weights to the data points in order to be able to approximate the integral of a function on an area of the earth using only the values of the data points.
- The second step is to choose the statistic T we will use and to be able to determine its integral value T^I for any "circle" I on the sphere that is the earth. Using this, we are able to determine the p-value associated p^I using a permutation test.
- Finally, we must compute the adjusted p-value $\widetilde{p}(x) = \underset{I \ s.t. \ x \in I}{sup} \ p^I$.

However, in our case it is not so clear how to apply it since we don't have regularly spaced data points on the earth. In fact, our data stations are really irregular and there are also huge voids. So, the first step on the proposed method in [2], which consists of making a triangulation with the data points, was not really consistent in our case. However, when we looked at what this triangulation is created for, we discovered that it was only to put weights on the data points : each data point weights a third of the total area of the triangles it belongs to. It is to say, the more a point is isolated, the more its weight is high. So, we decided to follow the same idea to put weights on our stations : each station weights $\frac{1}{N_{neigbours}}$ where $N_{neigbours}$ corresponds to the number of stations in a circle of radius R around the station. We tried different values for R (from 100km to 1000km) and we decided to use only 320km in the report since the results were quite stable.

Once our data points were associated with a weight, we were now able to go on with the next step of the method : choosing the statistic. First, when we deal with temperatures, our framework is the same as in part 6.1 of [2], so we decided to use the same idea. We assume that for each year $y \in 1978, ..., 2022$, month $m \in 1, ..., 12$ and location x, $Temp_y(x, m) = a(x, m) + y * b(x, m) + \epsilon(x, s)$. The hypothesis test we want to compute in this case is the following one :

H_0 : There is no augmentation -> corresponding to $H_0(x, m) : b(x, m) = 0$

H_1 : There is an augmentation -> corresponding to $H_1(x, m) : b(x, m) > 0$

To match this hypothesis test, we use a permutation test computing the t statistic with cutoff : $T(x, m) = max\left(0, \frac{\hat{b}(x,m)}{SE(b(x,m))}\right)$

Now we work month by month and we must estimate the value T^I for any "circle" I on the earth. However, strictly speaking is it not computationally possible to do so. However, we decided to do it for 6 different sizes of circles with radius going from $\sim 320km$ to $\sim 3200km$ (this corresponds to limit angle between 0.05 rad and 0.5 rad) and with central points on a grid G where spaces correspond to 1° in latitude and 1° in longitude. In order to estimate T^I for each of these circle, we compute : $T^I = \sum_{station \in I} T(station) * weight(station)$.

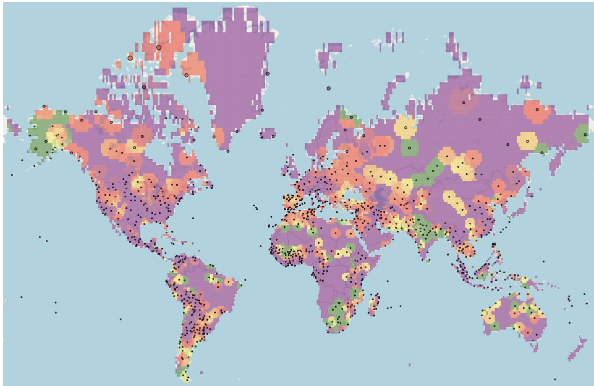
At this point, for each month m, we are now able to determine the p-value $p(m)^I$ corresponding to each of these circles with a permutation test. In our case, we used 500 permutations per station. We then created 6 maps (one for each radius R_i) and in each of these maps, a point with location x has a value corresponding to $v_i(x,m) = p(m)^I$ with $I = B(x, R_i)$ (See figure 1.6). We define $I_i \triangleq \{B(x, R_i) : x \in G\}$.

From these intermediate maps, we computed the partially adjusted p-value, it is to say for each radius size R_i, we created the same maps as before but this time, each value corresponds to : $v_i'(x,m) = \underset{I \in I_i : x \in I}{max} p(m)^I$.(See figure 1.7)

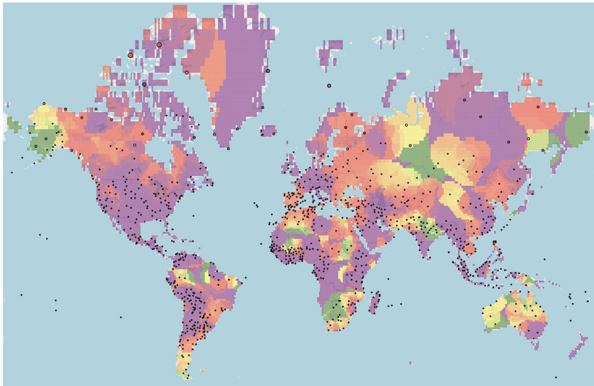
Finally, for each month we were able to create an approximated map of the adjusted p-value by taking for each points of the grid x, $\widetilde{v}(x,m) = \underset{i \in \{1,\ldots,6\}}{max} v_i'(x,m)$. (See figure 1.8)

We are aware that this method has faults, the main ones being that our data stations are not well spaced on the earth and that we could reasonably ran only 6 sizes of "circles" for computational reasons. But we wanted to try and deal with it and the results seem coherent with the map of the p-value determined for each station.

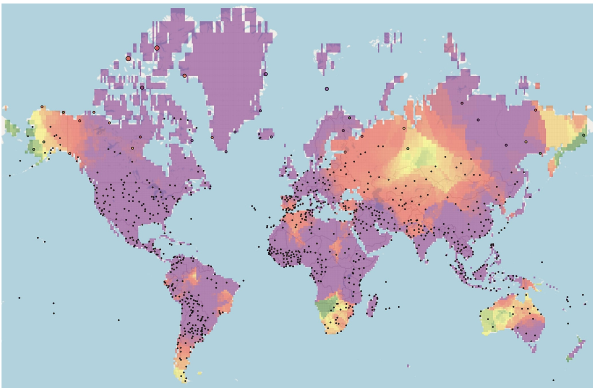
Figure 1.5: p-value scale



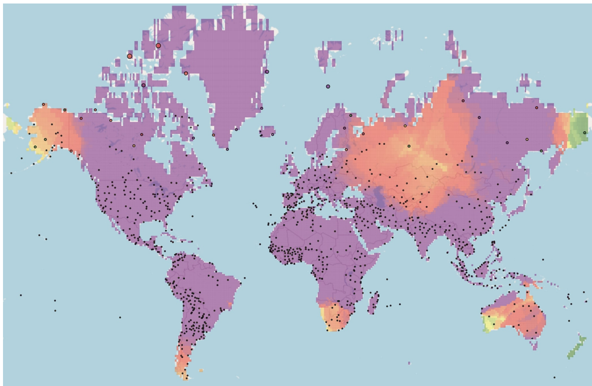
R $\sim 320km$



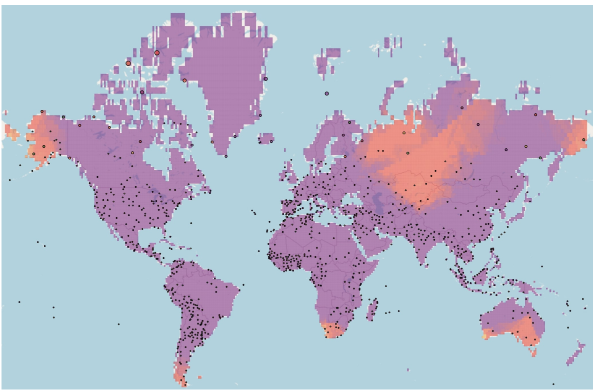
R $\sim 640km$



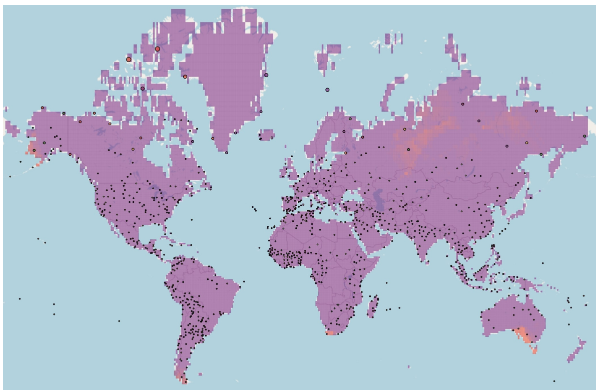
R $\sim 1280km$



R $\sim 1920km$

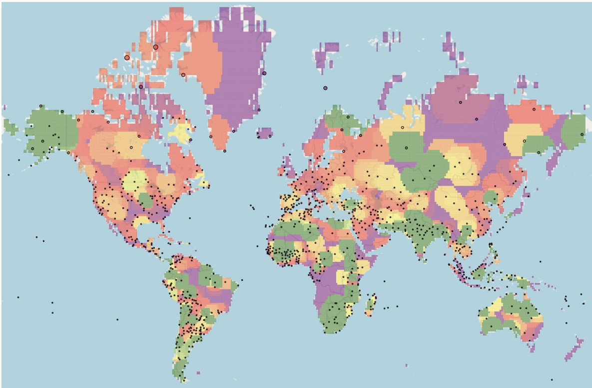


R $\sim 2560km$

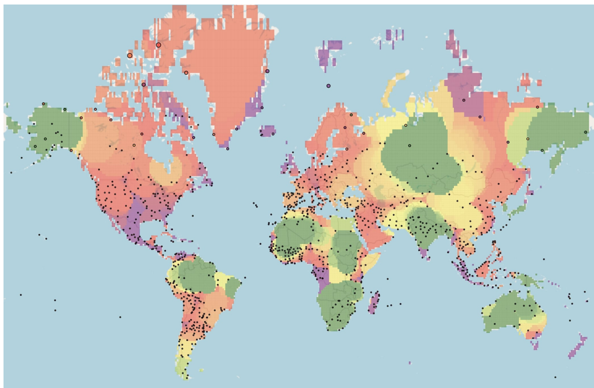


R $\sim 3200km$

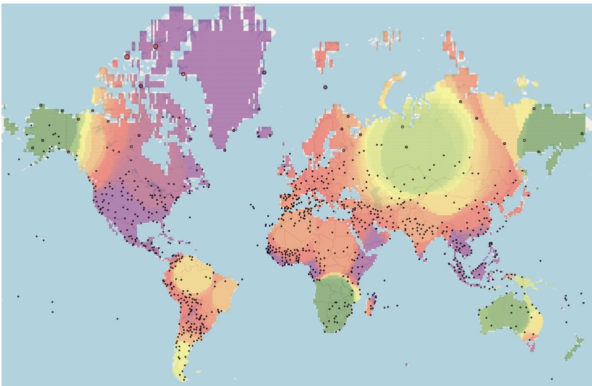
Figure 1.6: p^I per radius R in January



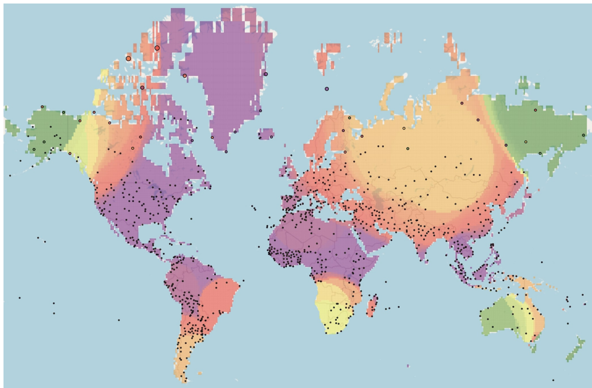
R $\sim 320km$



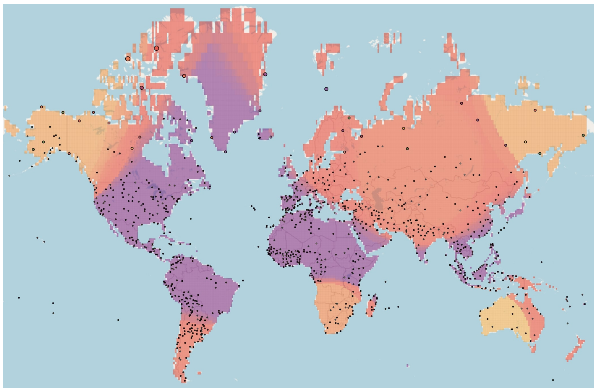
R $\sim 640km$



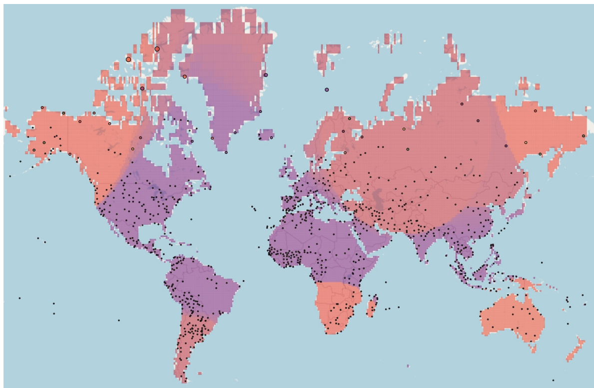
R $\sim 1280km$



R $\sim 1920km$



R $\sim 2560km$



R $\sim 3200km$

Figure 1.7: p-value partially adjusted per radius R in January

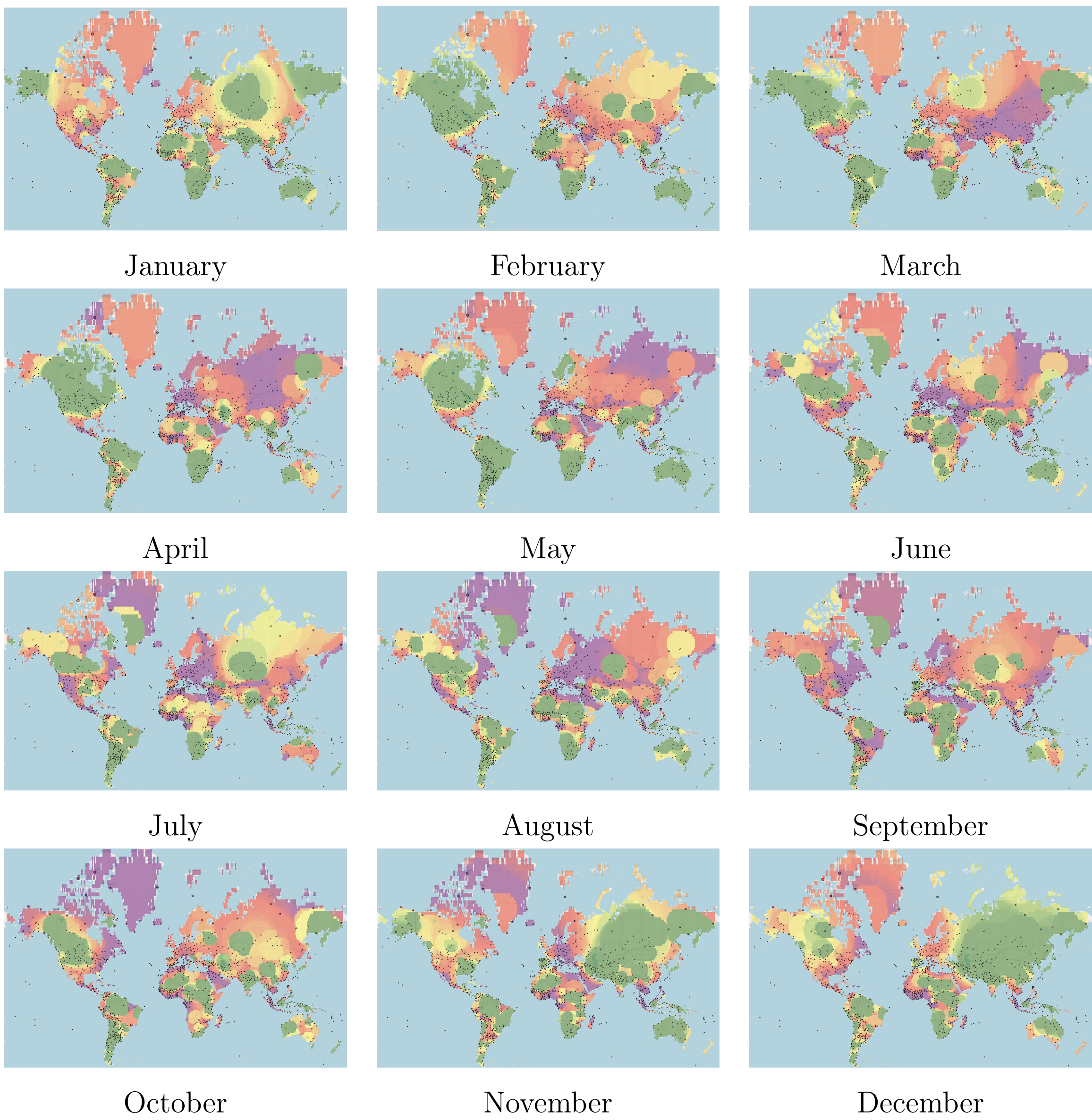


Figure 1.8: Adjusted p-value for each month (temperature)

From these different figures, there are many conclusions we can draw. The first one is clearly that all parts of the world are not affected in the same way by climate change. Indeed, the parts of the world that seem to be the most affected are western Europe, the west and east coasts of the United States, especially in the summer, Indonesia, the North of Africa and the Greenland region.

We applied the exact same method but with the precipitations this time in order to see if some places in the world where undergoing clear changes concerning the rain. The only thing we modified is the statistic of the hypothesis test since in this case we have no a priori on the modification : it could be a augmentation as well as a diminution. So the new test we perform was the square trend instead of the t test with cutoff. The results correspond to the figure 1.9.

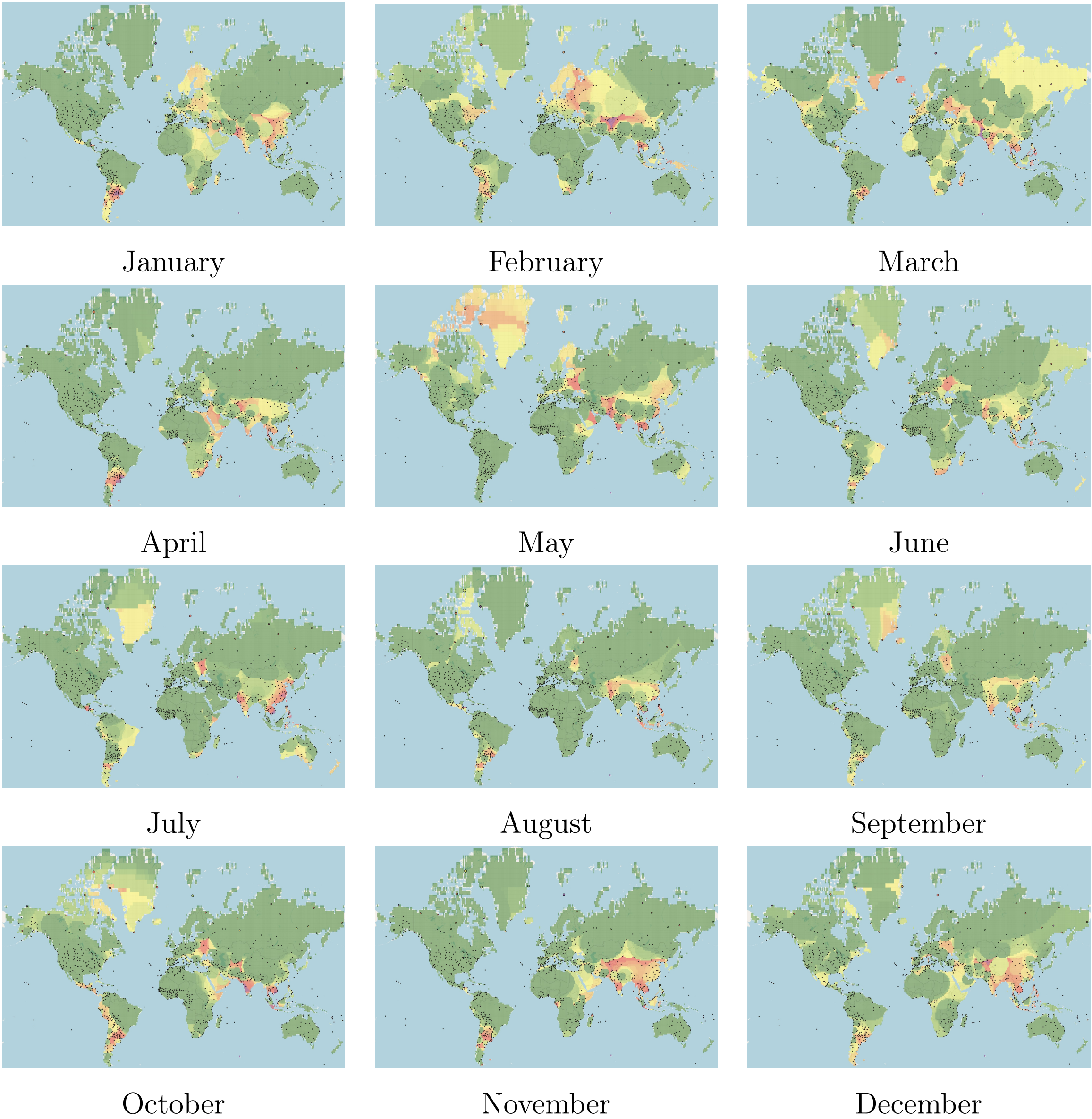


Figure 1.9: Adjusted p-value for each month (precipitations)

Concerning the precipitations, it is difficult to see any clear area that undergo modifications since most of the world looks green and so not significant at all.

1.2.3. How to expand the shift results to the whole world despite the lack of stations ?

In order to expand the results found at the station scale concerning the shift in temperature, we decided to use 2 different methods :

- The first one is the k closest neighbours which is quite classical in the regression framework. In the presented result we used k=5.
- The second method we decided to apply was inspired by the classical one dimensional linear regression between two points a and b : $f(x) = f(a)\frac{b-x}{b-a} + f(b)\frac{x-a}{b-a}$. In order to adapt this idea, we defined a radius of influence ($R \sim 2000km$ in the results) and the value of a point p corresponds to : $v(p) = \frac{\sum_{p_i \in B(p,R)} \frac{v(p_i)}{geo(p,p_i)}}{\sum_{p_i \in B(p,R)} \frac{1}{geo(p,p_i)}}$ where $geo(p_a, p_b)$ is the geodesic distance between p_a and p_b.

However, since the data stations are not properly spaced on the earth, we chose to apply a weight to each station (as in the significance case) in order to take into account the isolation of data points. Otherwise many points at the same location would have too much weight compared to a single point. Hence, we defined the same radius of 320km to put weights on the stations. Finally, we have $v(p) = \frac{\sum_{p_i \in B(p,R)} v(p_i) \frac{w(p_i)}{geo(p,p_i)}}{\sum_{p_i \in B(p,R)} \frac{w(p_i)}{geo(p,p_i)}}$. Where $w(p_i)$ is the weight of the station at location p_i.

In practice, we put $max(0.0001, geo(p, p_i))$ instead of $geo(p, p_i)$ in order to avoid zero-division errors. We could also have used the global formula which doesn't involve zero-division but it would be less convenient from a programming point of view : $v(p) = \frac{\sum_{p_i \in B(p,R)} v(p_i) w(p_i) \prod_{j!=i} geo(p,p_j)}{\sum_{p_i \in B(p,R)} w(p_i) \prod_{j!=i} geo(p,p_j)}$.

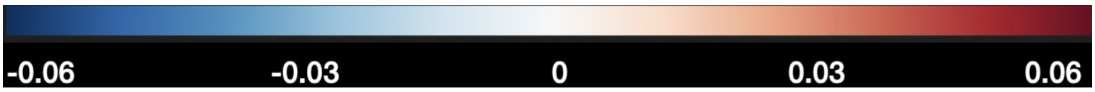
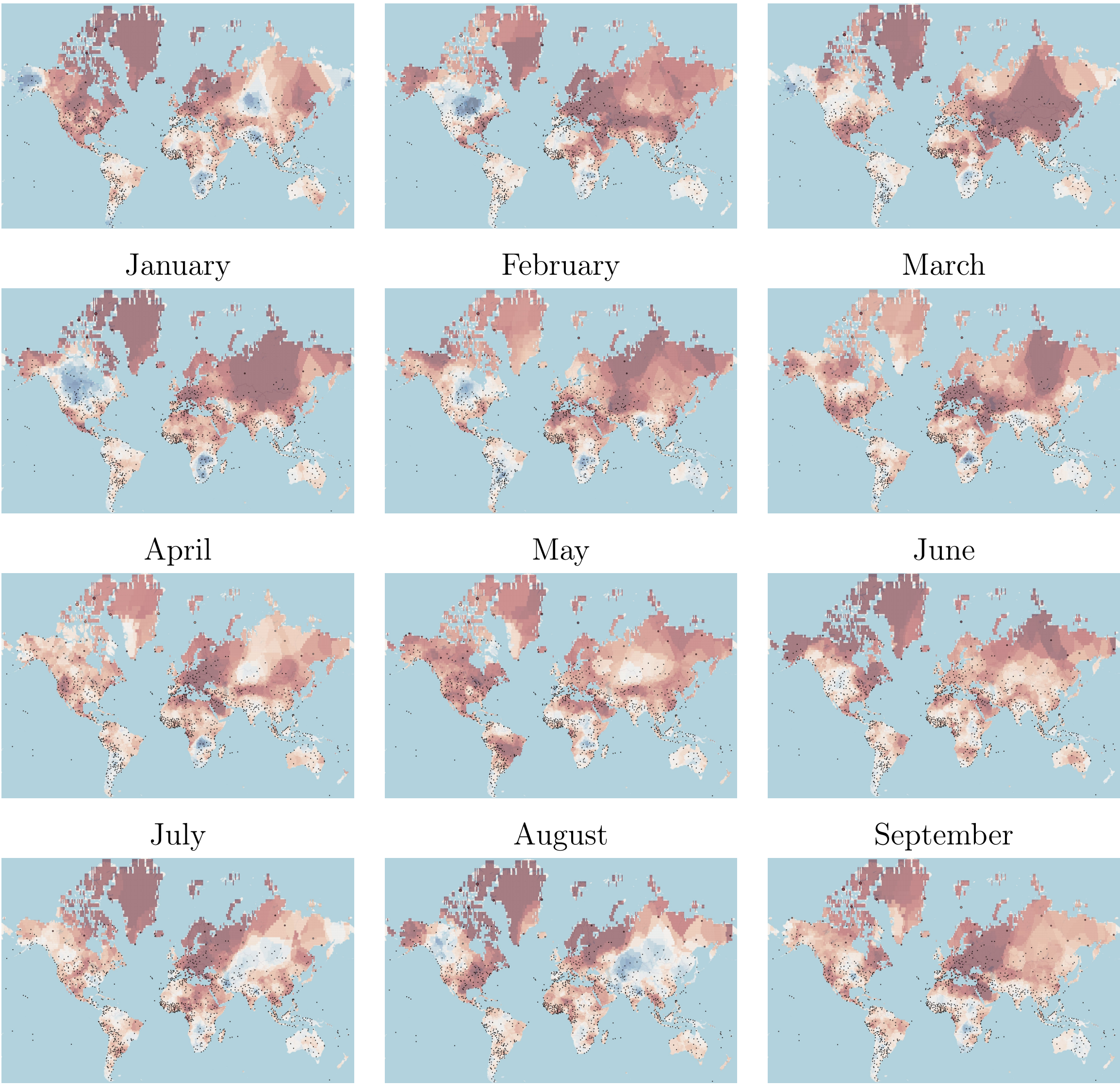
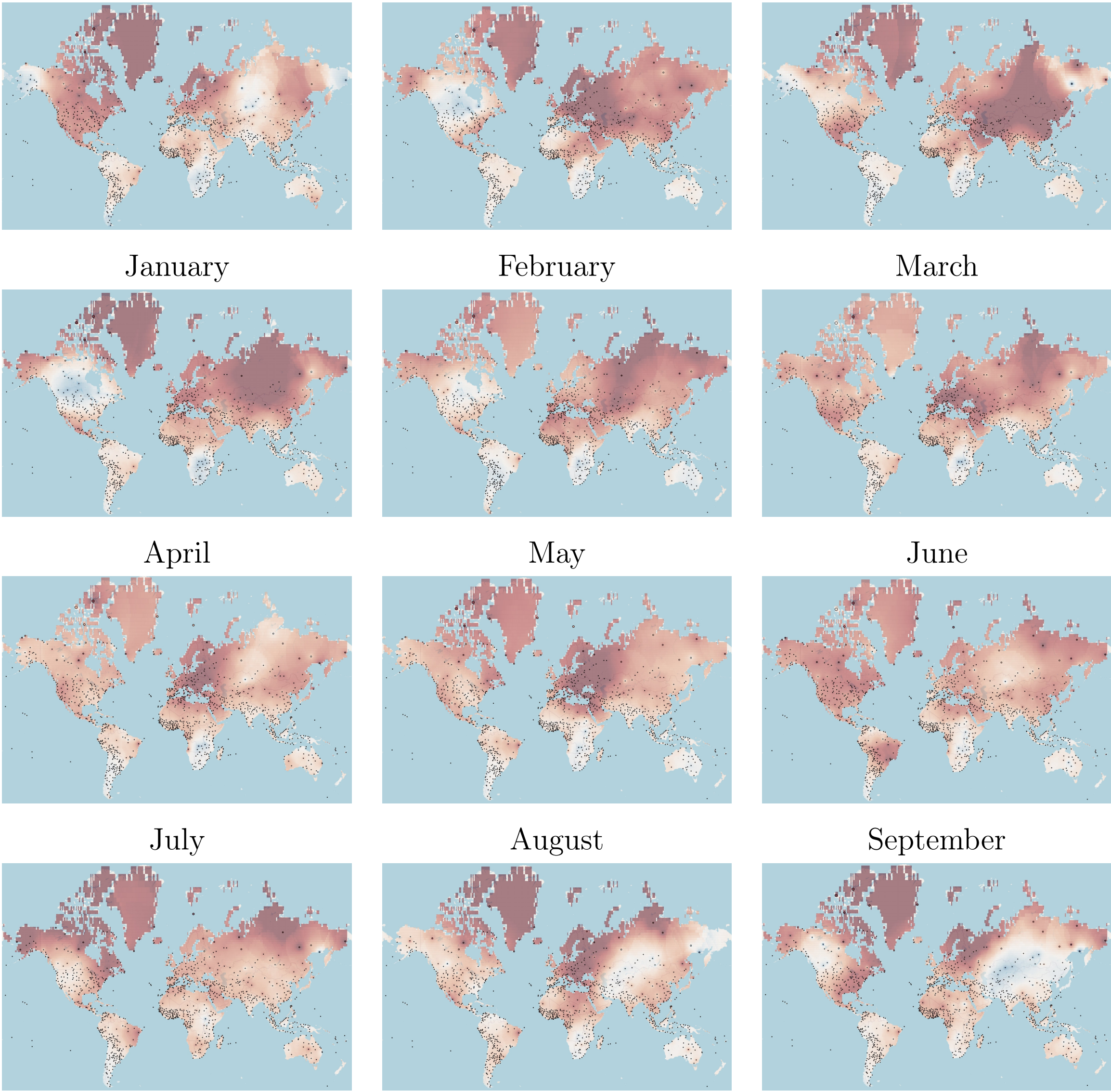


Figure 1.10: shift scale (degree per year)



January February March
April May June
July August September
October November December

Figure 1.11: Estimated shift using k-closest neighbours for each month



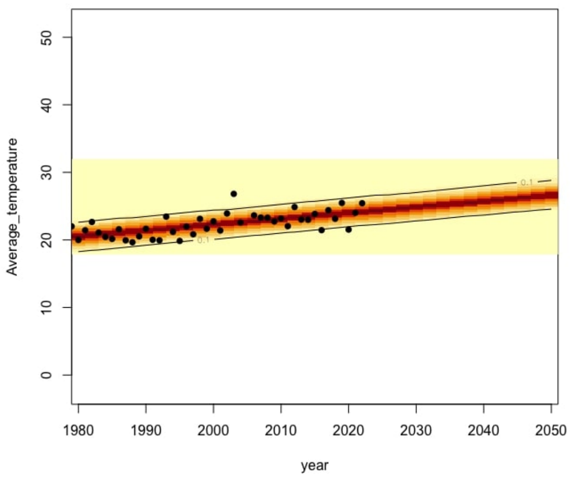
January February March
April May June
July August September
October November December

Figure 1.12: Estimated shift using 2D linear-like method

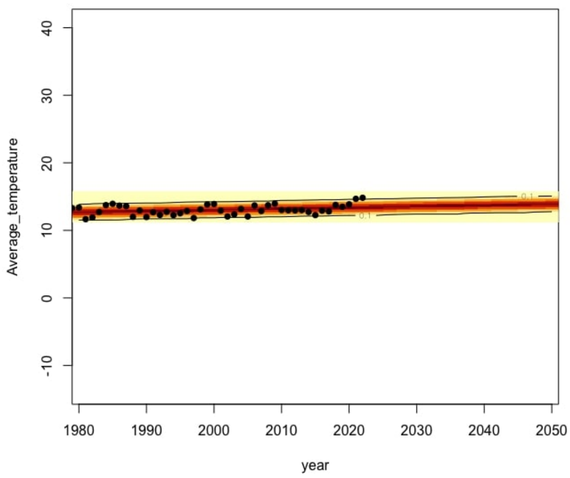
When we compare both methods, they seem to give close results, however, the second one produces a much smoother map due to the linearisation. Globally, we find the same regions as in the significance test corresponding to an increase in temperature. However, it is also important to say that our results should be consider as really relevant in regions where we have more data stations only.

1.3. Forecasting of the global warming

Another thing we wanted to deal with are temperature previsions. In fact, we wanted to see whether the temperature in the future could be really different from the actual ones and the previous ones. In order to do so, we decided to use the conformal prediction framework to establish prediction intervals for the average temperature. We hesitated between full conformal and split conformal but since we have "only" 45 years of data, it seemed better to use full conformal method. The discrepancy measure we use is the one corresponding to the linear regression to be coherent with our precedent frameworks. We decided to show the results only for one month (June) and 2 different locations : Verona and Cape Town. We decided so in order to show the two main cases we faced, one with a clear augmentation and another one which seems quite stable.



Verona



Cape Town

Figure 1.13: Conformal prediction bands for the average temperature in June

The 90% prediction intervals we deduced from this method are the following ones for 2030, 2040 and 2050 at Verona : [22.9,26.9],[23.7,27.9],[24.7,28.7]. If we also compute the bands for 1980, we see that the temperature will be really different from what it has been in the past : [18.3,22.5].

However, in Cape Town, we have another behaviour in June since the 90% prediction intervals for 2030, 2040 and 2050 are the following : [12.5,14.7],[12.7,14.9],[12.9,14.9]. So in Cape town, the temperature seems much more stable from an increasing point of view but also if we consider the variance at a fixed year.

Finally, we see that the different places on earth are really undergoing different effects due to climate change and some places seem not to be impacted while for other ones, there is almost no doubt that the temperatures are rising.

2 | Climate change is also about extreme events

In this section, we aimed to detect extreme meteorological events, such as heatwaves or heavy precipitation, at a daily scale, to investigate potential coherence with the findings of the first part of the study, which focused on global warming trends. On one hand, we anticipate an increase in the frequency of 'extreme events' in terms of temperatures in regions experiencing significant global warming. On the other hand, we believe that a more detailed examination of precipitation at a daily scale may yield more insightful results. This is because we know that extreme precipitation events, such as those leading to flooding [1], typically occur within a day and are among the observed consequences of climate change.

2.1. Data Preparation

To detect extreme weather events, we utilized historical daily weather data spanning from 01/01/1978 to 31/12/2022. This dataset included variables such as temperature, precipitation, sourced from the reliable meteostat database. This time, in analyzing temperature data, we primarily focused on maximum daily temperature readings. In other words, for each day, we utilized a single value representing the highest temperature recorded. As previously, we obtained this data from the same 1315 stations. For one particular station, e.g Poitiers (France), the data look like these :

Time	Tavg	Tmin	Tmax	Prcp
1978-01-01	7.1		11	0
1978-01-02	1.7	0	10	0
1978-01-03	3.8	0	10	0
1978-01-04	7	4	11	0
1978-01-05	0.6	-1	10	0
1978-01-06	0.2	-2	6	0
1978-01-07	0.9	-4	5	0
1978-01-08	3.6	0	10	0
1978-01-09	3.8	1	7	0
1978-01-10	6.3	4	11	0
1978-01-11	5		10	14
1978-01-12	2.4	0	3	9.9
1978-01-13	2.6	2	10	0
1978-01-14	2.8		5	0
1978-01-15	3.4		7	0
1978-01-16	2.6	-1	10	0
1978-01-17	2.1	0	4	0.5
1978-01-18	0.8		10	1
1978-01-19	1.3	-3	10	3
1978-01-20	2.2	0		1
1978-01-21	1.3	-1	10	0
1978-01-22	3.4	1	10	5.1
1978-01-23	2.4	-1	10	5.1

Figure 2.1: Daily Data for Poitiers, January 1978. We work with Tmax & Prcp.

2.2. Methodology

Daily maximum temperature and precipitation data were collected from the selected 1315 weather stations covering the entire study period. We computed the 95th quantile for both maximum daily temperature and maximum daily precipitation, in order to identify the top 5% of values. These represent the highest temperature and precipitation events observed during the study period. The years corresponding to these extreme events were determined. This involved identifying which specific years had extreme maximum daily temperature and precipitation values falling within the top 5% of the distribution. The identified years containing extreme events were then plotted on a new histogram. This visualization provided a clear representation of the distribution of extreme events across the study period, highlighting periods characterized by unusually high maximum daily temperature or precipitation. Please note that this methodology is only "observational". We do not perform another non-parametric statistical study here, we just check at hand if the extreme events have a tendency to be observed in the recent years (which would suggests an increase in their frequency) or on the whole timescale ; and if we can observe more frequent extreme events in the areas that were previously identified as having a statistically significant global warming.

2.3. Detection of extreme events linked with temperature

Here, we showcase the outcomes of the earlier methodology applied to Cape Town (South Africa) and Verona (Italy) for the month of June. This implied applying the methodology to each day of June spanning from 1978 to 2022 for these two regions.

Firstly, we observe that there is no discernible pattern indicating an increase in the frequency of heatwaves at a daily scale in Cape Town. Conversely, it appears that Verona has experienced a growing number of heatwaves at a daily scale in recent years (since 2000), as depicted in the second plot.

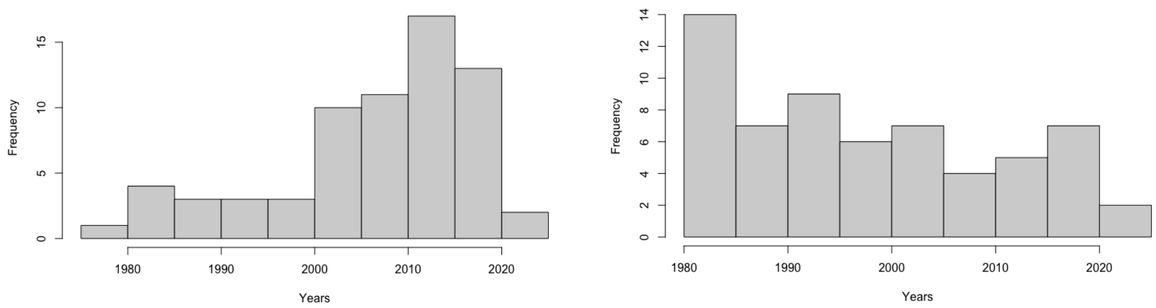


Figure 2.2: Daily Tmax in Verona (left), Cape Town (right) for June days, top 5%

Let's now delve into the insights provided by our previous study regarding the month of June, focusing on the statistical significance world map (p-values) when examining a potential shift in maximum temperature for the month of June. We observe that Cape Town is depicted in green and Verona in purple on the map. This coloration suggests that there is no statistical evidence of a temperature shift in Cape Town, while such evidence is apparent for Verona. In this regard, the observations we make in this section align with the findings of our initial study.

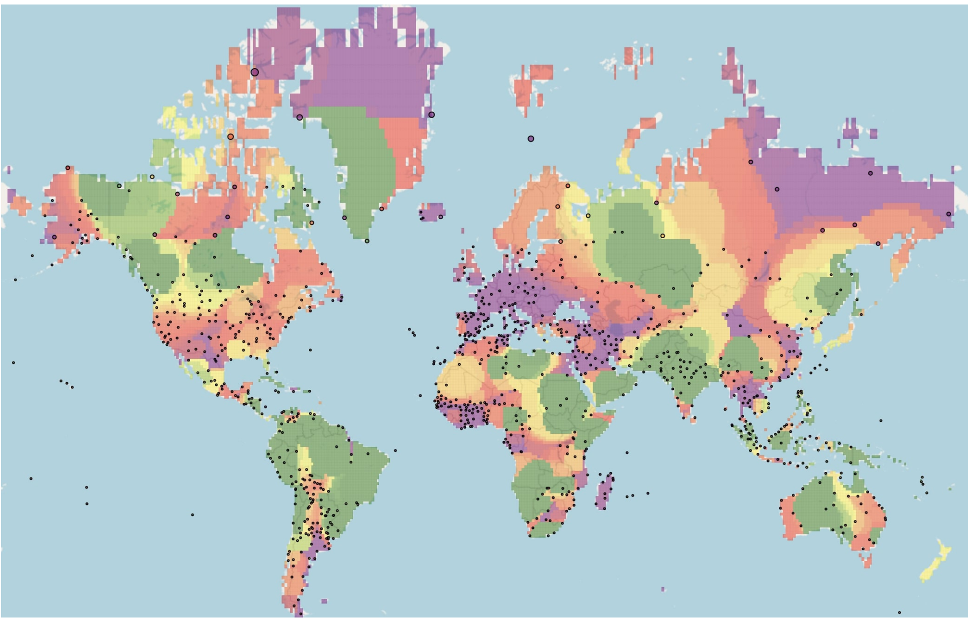


Figure 2.3: P-values world map for Tmax, June

2.4. Detection of extreme events linked with precipitation

Once again, regarding precipitation, the findings are not readily discernible. Upon examination of the following two plots, it becomes evident that the top 5% of extreme daily precipitation events in both Cape Town (South Africa) and Verona (Italy) during the month of June span from 1978 to 2022, with no discernible "accelerating pattern."

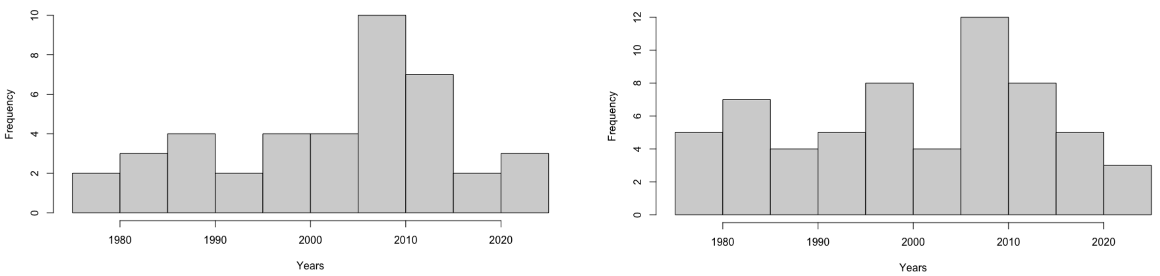


Figure 2.4: Daily Prcp in Cape Town (left) Verona (right) for June days, top 5%

Let's review the findings from our prior study. In this analysis, we once again generated a world map depicting p-values to assess alterations in precipitation patterns, focusing specifically on the month of June. When examining the regions of Cape Town and Verona, we observe no significant evidence indicating a shift in precipitation patterns, as indicated by the absence of purple shading representing low p-values.

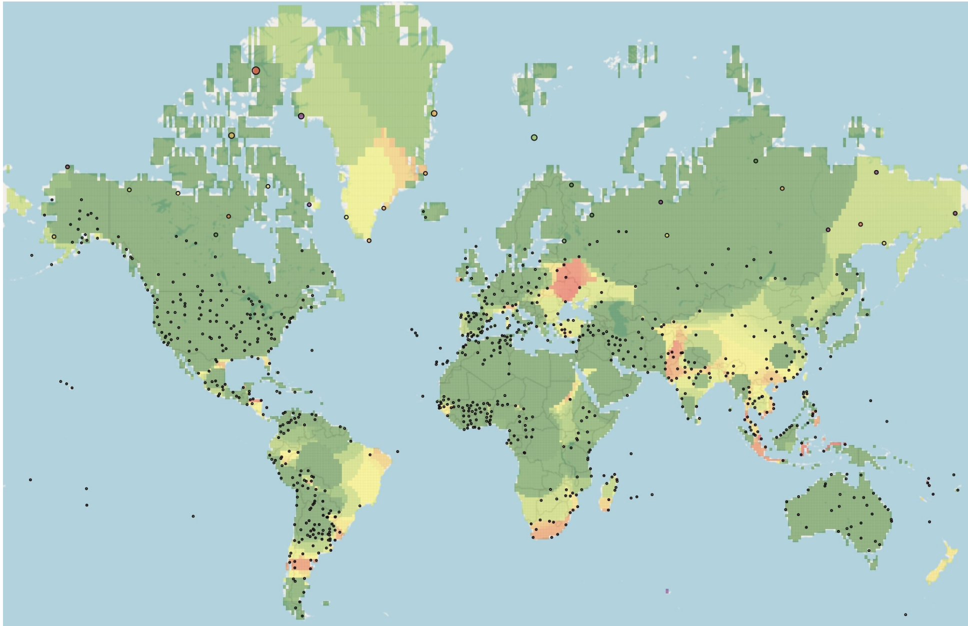


Figure 2.5: P-values world map for Prcp, June

The observed alignment between extreme events and the patterns identified in the analysis of global warming underscores the fact that climate change manifests in various consequences, encompassing not only global warming but also highly intense meteorological events. As a future endeavor, we could conduct a non-parametric statistical study similar to the one conducted in the initial part of this report, but with a focus on daily data. This expanded analysis would deepen our understanding of extreme events and their implications.

3 | Conclusion

In conclusion, our study employed an innovative method for calculating p-values globally, inspired by Olsen, Pini and Vantini's article from 2023 [2], to assess the quantitative effects of climate change. Through these calculations, we were able to reaffirm known facts regarding the pace of climate change in specific regions worldwide. For instance, our findings confirm that regions nearer to the poles, such as Greenland, exhibit more pronounced warming trends compared to equatorial areas. Additionally, our analysis revealed that temperature plays a more significant role than precipitations.

Furthermore, the disparity in statistical significance between precipitation and temperature data can be attributed to the greater variance observed in precipitation data. Moreover, our analysis of monthly average temperatures led to conclusions consistent with those derived from a secondary analysis focusing on the frequency of extreme events, such as short-term heatwaves.

Finally, our objective to present results in a clear and accessible manner, even to non-scientific audiences, has been successfully achieved. The maps we have generated are intuitive and readily comprehensible, fulfilling our goal of communicating our findings effectively.

Bibliography

[1] A. Laurin. Les inondations dans le nord de la france pourraient coûter 255 à 400 millions d'euros [The floods in northern France could cost 255 to 400 million euros]. *Les Echos*. URL `https://www.lesechos.fr/finance-marches/banque-assurances/les-inondations-dans-le-nord-de-la-france-pourraient-couter-255-a-400-millions-`

[2] N. L. Olsen, A. Pini, and S. Vantini. Local inference for functional data on manifold domains using permutation tests, 2023.

[3] S. Sengupta. The fingerprints on chile's fires and california floods: El niño and warming. *New York Times*. URL `https://www.nytimes.com/2024/02/05/climate/california-floods-chile-wildfires-global-warming.html?searchResultPosition=2`.