STA4000 READING COURSE—FINAL REPORT
LONG TIME SERIES DATA FOR ICELAND CLIMATE, VOLCANO AND POPULATION

HAO, CHUNYAN
SUPERVISOR: PROFESSOR JEFFREY S. ROSENTHAL

DECEMBER 26, 2009

1. Introduction

Climate change is one of the hot spots in the scientific talking nowadays but quantity analysis is rare. The goal of this project is looking for quantity evidence of relationship between climate change and population growth in Iceland by analyzing long time series temperature and population data.

Iceland is small island, located at 63-67° N and 18-23° W, has population over 300.000, half of which live in the capital Reykjavik and towns in the southwest. The economy of Iceland is heavily dependent on renewable natural resources, including fishing and farming. Iceland has considerably milder climate than its location just south of the Arctic Circle would imply. The average temperature of the warmest month, July, exceeds 10° C in the lowlands of southern and western Iceland, but is below that in other parts of the country. The warmest summer days around Iceland can reach 20-25° C, with the absolutely highest temperatures recorded at around +30° C. Winter in Iceland, on the other hand, is generally very mild for this northerly latitude. The coastal lowlands have mean January temperatures close to 0° C, and only in the highlands of central Iceland do the temperatures stay below -10° C. The lowest winter temperatures in northern Iceland and the highlands are generally in the range -25 to -30° C, with -39.7° C the lowest temperature ever recorded.

One advantage of choosing Iceland is that its population is stable and less effected by the war and immigrant. So the most dramatic population change is coming from natural hazard, for example, volcanoes since Iceland is volcanically and geologically active on a large scale. The other advantage is that industrialization has less impact on climate change in Iceland due to its geographical location and economic structure. In this report, we will focus on pre-industrial age (1735-1860) when we assume that population growth might be more effected by the climate than post-industrial age when technology is significantly improved to provide huge protection for people from natural hazard.

The main outcomes during the analysis are discussed in this report and the detailed data analysis is included in "Appendix".

2. Data Source

Population data一Annually census data since 1735 is available on the website of Statistical Iceland. In Table 1 and Table 2, we compared the population data between pre-industrial age (1735-1860) and post-industrial age (1860-2000), it is obvious that the population grow faster in post-industrial period (mean of growth rate is 1.0193 compared to 0.7149) and the growth rate is more stable in post-industrial era (standard deviation of 0.8228 compared to 1.3968 and smaller standard deviation value means less fluctuation). This result can be explained by the technology that is significantly improved in post-industrial era which results in more secure living space to against natural disaster.

		1735-1860	1860-2000
Population	Mean	49692.85	118285.5
	STD	6319.547	53473.13

Table 1. Population compared between pre-industrial age and post-industrial age

		1735-1860	1860-2000
Population growth rate	Mean	0.7149	1.0193
	STD	1.3968	0.8228

Table 2. Population growth rate compared between pre-industrial age and post-industrial age

In this report, we will focus in pre-industrial age (1735-1860) and we have the following plots from R. From Fig 1 and Fig 2 we know there are several points of dramatic changes in population amount which, according to the records (which is available on the website of Global Volcanism Program from Smithsonian Institute, U.S.A), is caused by several big volcano eruptions. Except those points caused by natural hazard, including volcanoes, disease and famine, we can see from Fig 2 that the growth rate is fluctuate very similar during the years without critical extreme natural events.

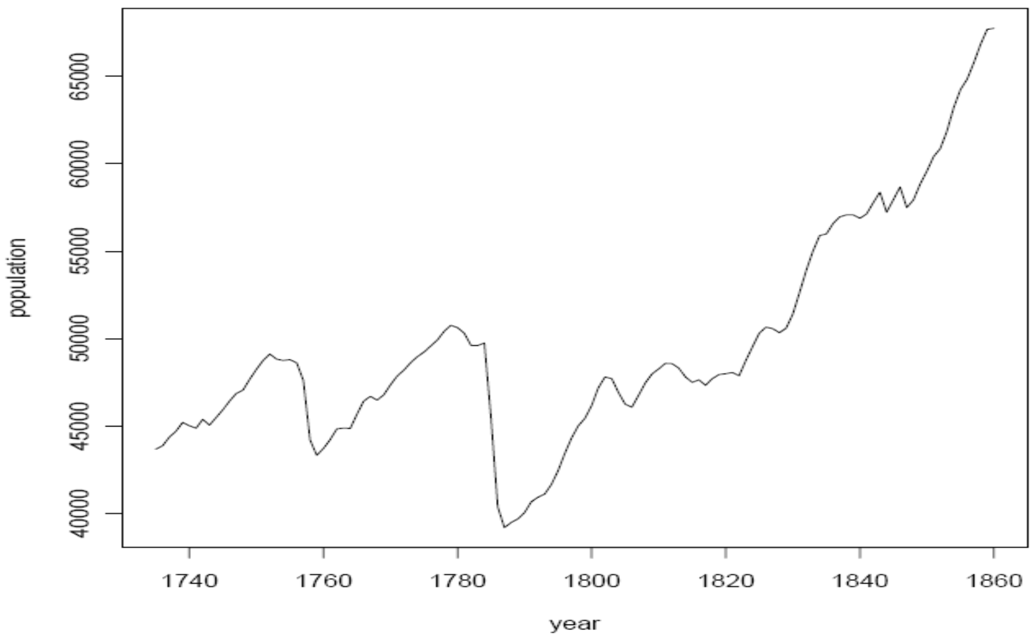

Fig 1. Population growth between 1735 and 1860

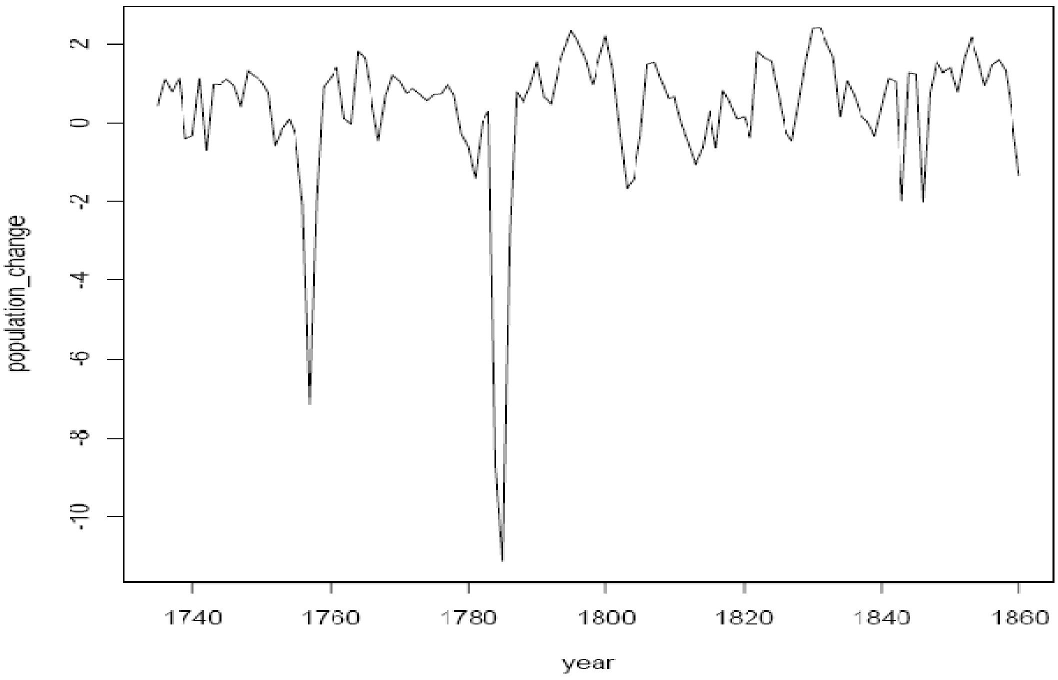

Fig 2. Population growth rate between 1735 and 1860

Instrumental temperature data—There are five stations collecting temperature data in Iceland which are Akureyri (65.7N, 18.1W), Reykjavik (64.1N, 21.9W), Stykkisholmur (65.1N, 22.7W), Teigarhorn (64.7N, 14.3W) and Vestmannaeyja (63.4N, 20.3W) and their exact locations are shown on the map below.

Map 1. Five stations in Iceland collecting instrumental temperature data

The historical instrumental data records are available on NASA GISS website and the instrumental data are dated back to about 1880 (but there are a lot of missing data from 1880 to 2008). These data are very important because we will need them to verify the correctness of reconstruct temperature data in pre-industrial age. Firstly, I extract the instrumental data from 1900 to 1980 when there are no missing data and plot it in R in Fig 3.

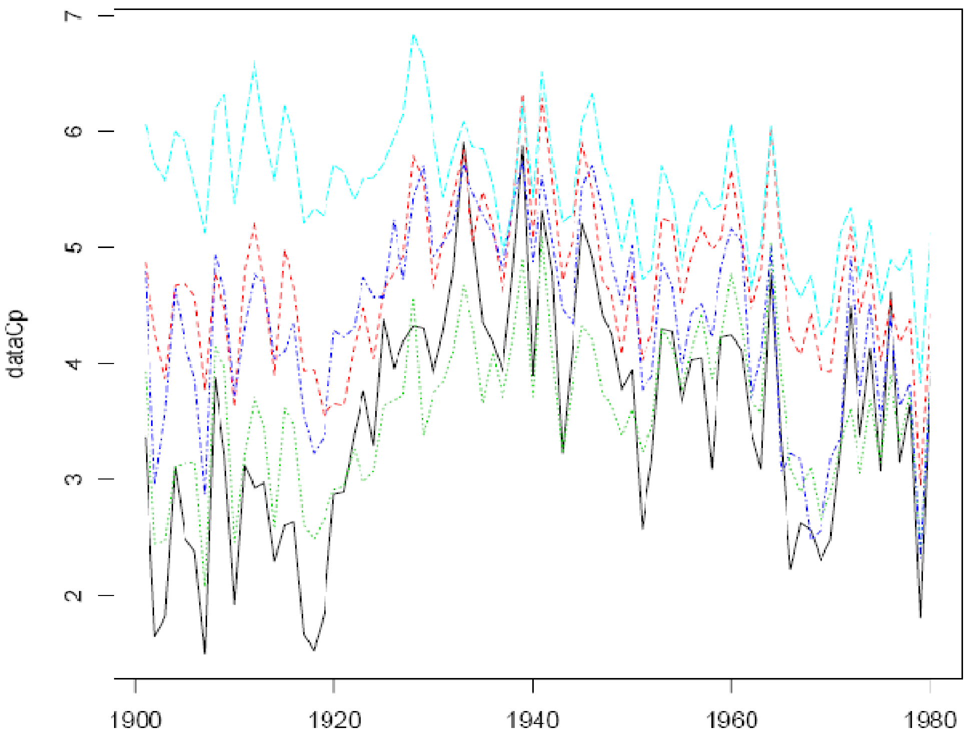

Fig 3. Iceland instrumental temperature from five stations between 1900 and 1980

It is shown in the Fig 3 that the temperature changes in Iceland are pretty consistent except data from southern Iceland Vestmannaeyja (63.4N, 20.3W).

The instrument temperature data from these five locations include the mean temperature in each month of the year (from January to December), mean temperature in each season of the year (spring (March, April, May), summer (June, July, August), autumn (September, October, November) and winter (December, January, Feburary)) and mean temperature for each year from 1901 to 1980. I compare the mean temperature from each season and mean temperature from each year respectively and find out that the temperature in Akureyri, Reykjavik, Stykkisholmur , and Teigarhorn are highly correlated with value between 0.77 to 0.87. However, the temperature of Vestmannaeyja is less correlated with other four stations (value of 0.4) because Vestmannaeyja is located in very south of Iceland effected most by the current which makes its climate very mild compared to other locations. Therefore, we decide to despite the data from Vestmannaeyja and consider that temperature data from other four locations can represent the climate change in Iceland as a whole.

3. Temperature reconstruct for pre-industrial age in Iceland

3.1 Peleoclimatology and proxy data

Because the widespread reliable instrument climate records are only available for the last 150 years or so and we certainly cannot go back in time, in order to understand how climate changes millions years ago, scientists find a way to estimate past climates based on imprints crated during past climate, known as proxies, to interpret paleoclimate and this study is called Paleoclimatology. These proxies include Microbial life, Ice core, Tree rings and sediment cores (which include diatoms, foraminifera, microbiota, pollen, and charcoal within the sediment and the sediment itself). Past climate can be reconstructed using the combination of different types of proxy records.

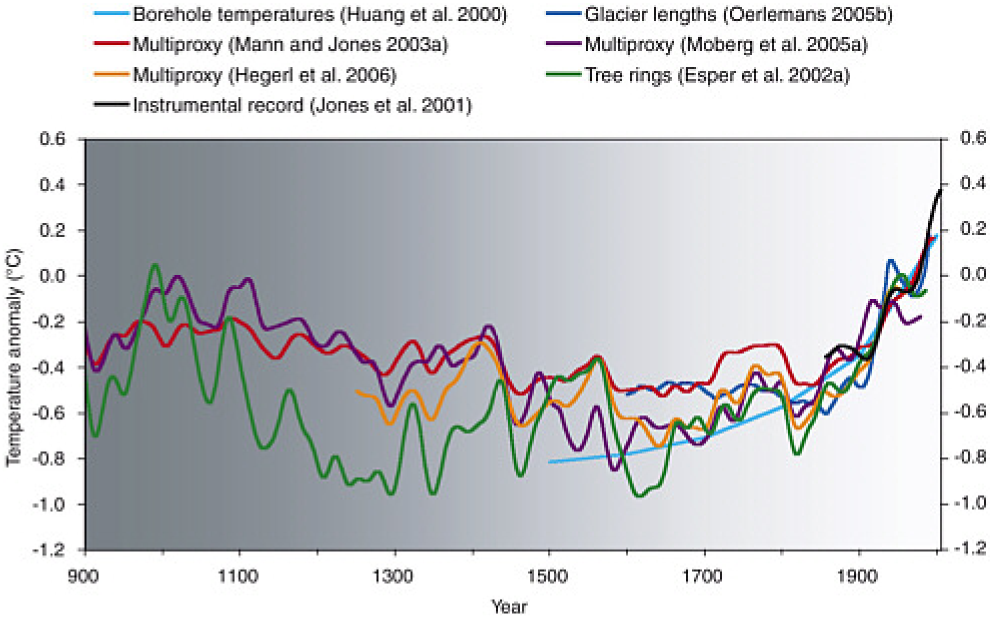

Fig 4. Temperature reconstruct by using six different proxy evidences

Fig 4 is smoothed reconstruction of large-scale (Northern Hemisphere mean or global mean) surface temperature variation from six different research proxies along with the instrumental record of global mean surface temperature. Each curve portrays a somewhat different history of temperature variations and is subject to somewhat different set of uncertainties that generally increase going back in time (as indicated by the gray shading). This set of reconstructions conveys a qualitatively consistent picture of temperature changes over the last 1,100 years and especially over the last 400 years. And it also indicates that the estimation based on proxy evidences can be very accurate.

3.2 Iceland temperature reconstruct based on sediment data

Sediment data is one of those proxy evidences that scientists use to reconstruct past climate. From NOAA website, I find temperature reconstruct data set for Iceland based on sediment data and the location, i.e. Stora Vidaruatn (66N, 15W), where the data was collected is shown in Map 2. The temperature data estimate the temperature in August in this location back to 2000 years ago. The next step is answering two questions: “How accurate is this estimate temperature” and “Can these data represent the climate of pre-industrial age of Iceland?”

Map 2. Location of Stora Vidaruatn where the sediment data was collected

In order to verify the accuracy of sediment data, I compare instrumental temperature data to sediment data between year 1900 to year 1980.

	Akureyri	Reykjavik	Stykkisholmur	Teigarhorn
Stora	0.6142	0.4110	0.3975	0.4186

Table 3. Correlation of between sediment data and annual mean of instrumental data

	Akureyri	Reykjavik	Stykkisholmur	Teigarhorn
Stora	0.4918	0.1758	0.4191	0.3465

Table 4. Correlation of between sediment data and summer season mean of instrumental data

	Akureyri	Reykjavik	Stykkisholmur	Teigarhorn
Stora	0.5271	0.1972	0.4507	0.4031

Table 5. Correlation of between sediment data and August temperature of instrumental data

From table 3-5, we know that annually the sediment estimate temperature is not accurate. Then I tried several average estimate temperature and compared it with corresponding average instrumental temperature and found out five years, eight years and ten years average of sediment data and instrumental data is highly correlated.

	Akureyri	Reykjavik	Stykkisholmur	Teigarhorn
Stora	0.7550	0.5496	0.5423	0.5208

Table 6. Correlation of 5 years average temperature data (annually)

	Akureyri	Reykjavik	Stykkisholmur	Teigarhorn
Stora	0.7911	0.5501	0.5473	0.5348

Table 7. Correlation of 8 years average temperature data (annually)

	Akureyri	Reykjavik	Stykkisholmur	Teigarhorn
Stora	0.8431	0.7010	0.6548	0.6196

Table 8. Correlation of 10 years average temperature data (annually)

From table 6-8, although all 5 years, 8 years and 10 years average temperature are highly correlated, but 5 years average provide enough information to analysis the temperature data. Also, it is not surprised that the correlation between Stora and Akureyri has the highest value since they are geographically the closest and this is also shown in Fig 5.

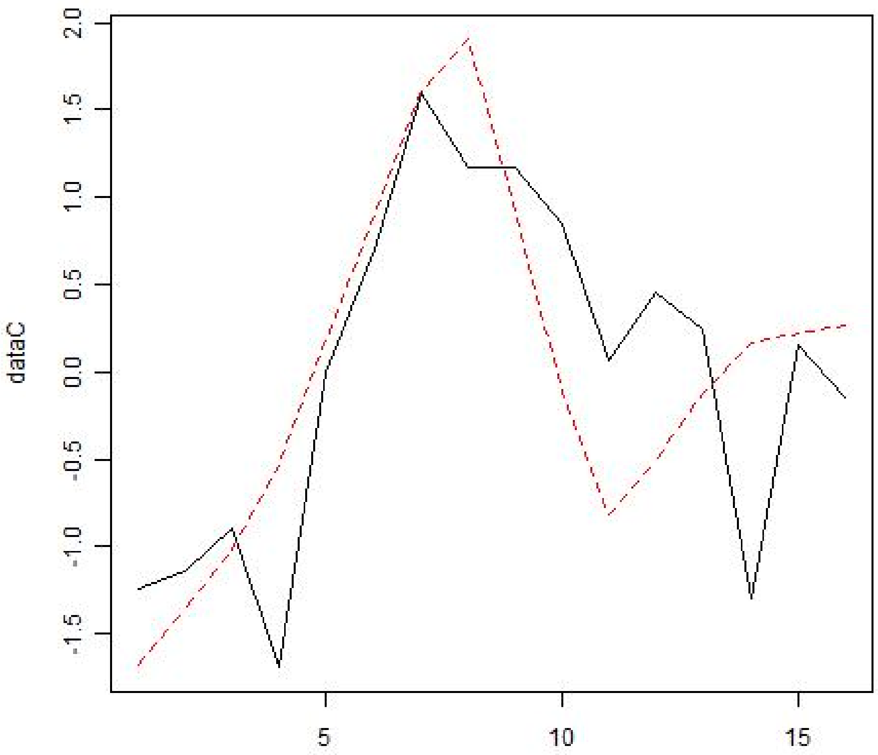

Fig 5. Plot of correlation of 5 years average temperature between Stora and Akureyri

3.3 Correlation between population and sediment data

In order to build a relationship between population and temperature, I investigate the relationship between population and sediment data and get the following table.

	Population data (1735-2000)
Sediment data(1735-2000)	0.7688151

Table 9. Correlation between population data and sediment data (1735-2000)

It's shown in the table that they are highly correlated, however, it is not concluded since this correlation might just because population number and temperature are both increasing between year 1735 and 2000. The value of correlation between population growth rate and sediment data is very small (0.2) and it doesn't improve much if we take 5 years or 10 years average temperature.

4. Model analysis

4.1 Linear model for population and temperature excluding volcano factors

From Fig 1, it seems that the population growth and time (year) have a linear relationship with quadratic terms of time, therefore, excluding the effect of natural hazard, we think about the model:

$$\text{population} = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 \text{tem}_{ave5} + \varepsilon \qquad (4.1)$$

Where: t —time of the year

tem_{ave5} —sediment estimate temperature averaged over 5 years

ε —error, following a normal distribution with constant variance.

After fitting the data, we have following result in Table 10:

Coefficients	Estimate	Std. Error	t value	Pr(>|t|)
β_0	27554	63564	0.433	0.66543
β_1	53353	4530	11.778	<2e-16
β_2	32939	10881	3.027	0.00301
β_3	2781	7984	0.348	0.72822
Residual standard error: 2770 on 122 degrees of freedom				

Table 10. Result for fitting the data to model (4.1)

From Table 10, it is obvious that the temperature is not significant for the response variable population. And when we use graphic methods to diagnose model (4.1), we found that the assumptions of ε error, assumed to be random and following a normal distribution with constant variance are violated. It suggests that the model is not fitting very well for the data.

Of course, I also check other possibilities, such as using moving average temperature, lagged temperature etc. which is listed in Appendix. It is safe to say if we didn't consider the volcano's impact on the population, temperature is not statistically significant for population growth.

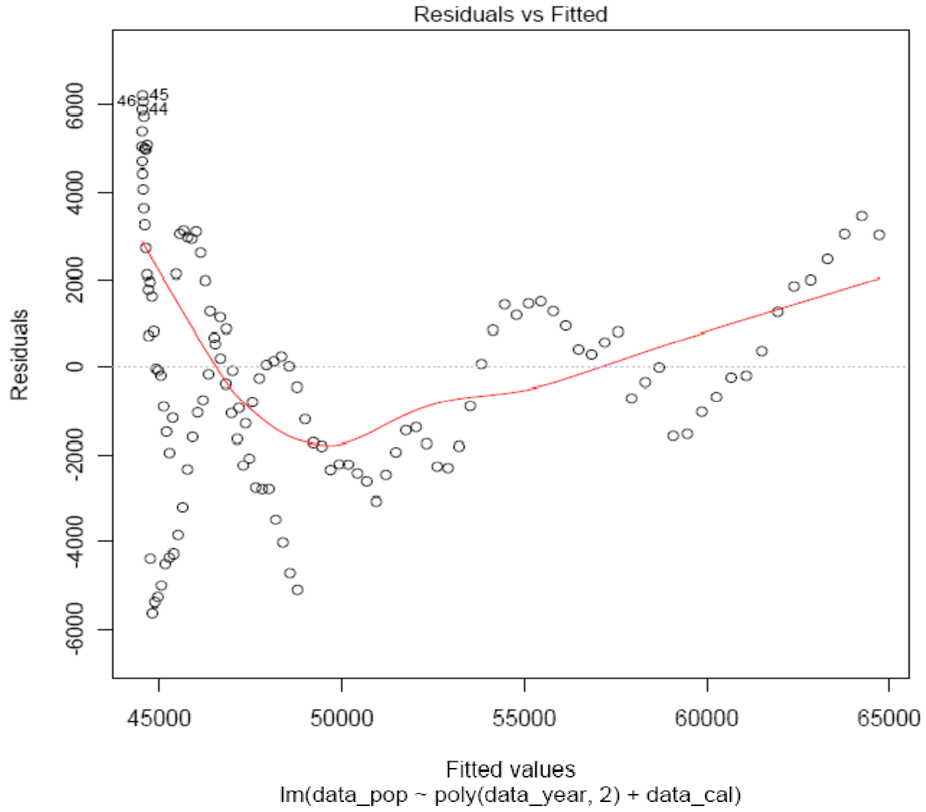

Fig 6. Plot of residual vs. fitted value for model (4.1)

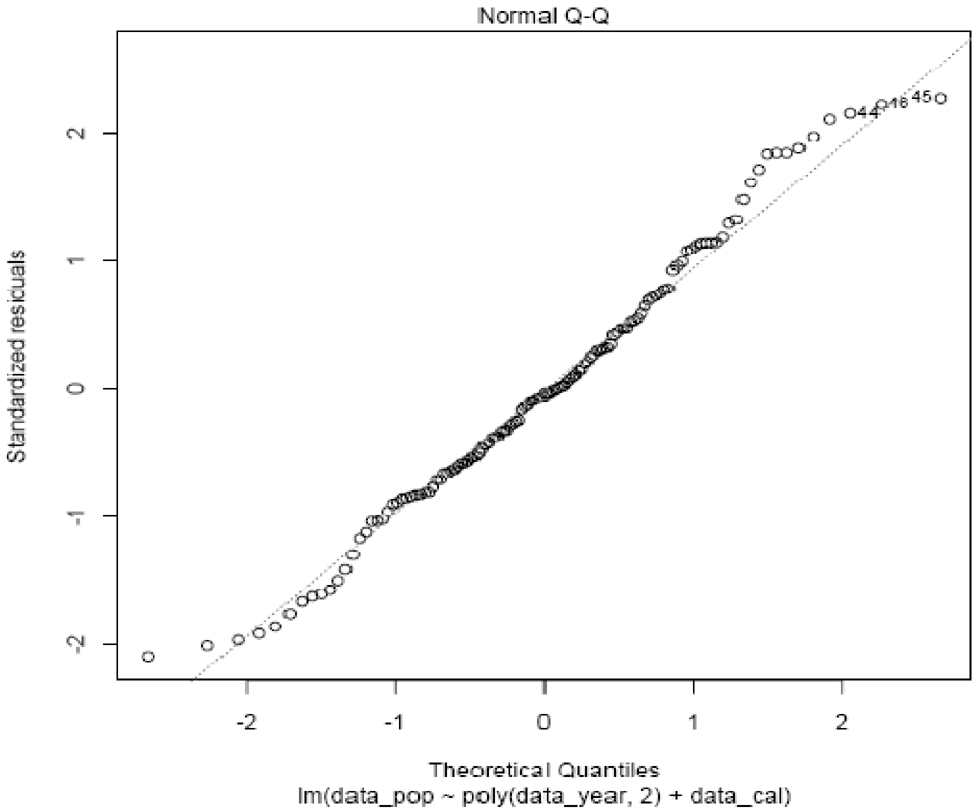

Fig 7. Q-Q norm plot for model (4.1)

4.2 Linear model for population and temperature including volcano factors

When analyzing population data that affected by volcanoes, we know two facts, first fact is every volcano has different impact on population amount and it may take several year to recover from damage of volcano. According to analysis, the average recover time is 5 years. The second fact is that the impact of volcano on population growth rate is not dramatic in Iceland. Therefore, we consider to exclude data that the year when volcanoes happened and the following 4 years after that. As to temperature, we still use sediment data with five years moving average but there are two situations we need to consider: simply using the temperature data that left after excluding volcano event OR still using 5 consecutive years moving average temperature data. So the models are:

$$\text{population growth rate} = \beta_0 + \beta_1 t + \beta_2 \text{tem}_{ave5} + \varepsilon \qquad (4.2)$$

$$\text{population growth rate} = \beta_0 + \beta_1 t + \beta_2 \text{tem}_{continue5} + \varepsilon \qquad (4.3)$$

Where: t —time of the year

tem_{ave5} —sediment estimate temperature averaged over 5 years

$\text{tem}_{continue5}$ —sediment estimate temperature averaged over 5 consecutive years

ε —error, following a normal distribution with constant variance.

After fitting the data, we have following result:

Coefficients	Estimate	Std. Error	t value	Pr(>|t|)
β_0	10.06563	4.832546	2.083	0.0417
β_1	0.001125	0.00164	0.686	0.4955
β_2	-1.366999	0.511125	-2.674	0.0097
Residual standard error: 0.5667 on 58 degrees of freedom				

Table 11. Result for fitting the data to model (4.2)

Coefficients	Estimate	Std. Error	t value	Pr(>|t|)
β_0	50.852001	15.634119	3.253	0.00191
β_1	-0.008587	0.003224	-2.663	0.01000
β_2	-4.244819	1.278677	-3.320	0.00156
Residual standard error: 0.5506 on 58 degrees of freedom				

Table 12. Result for fitting the data to model (4.3)

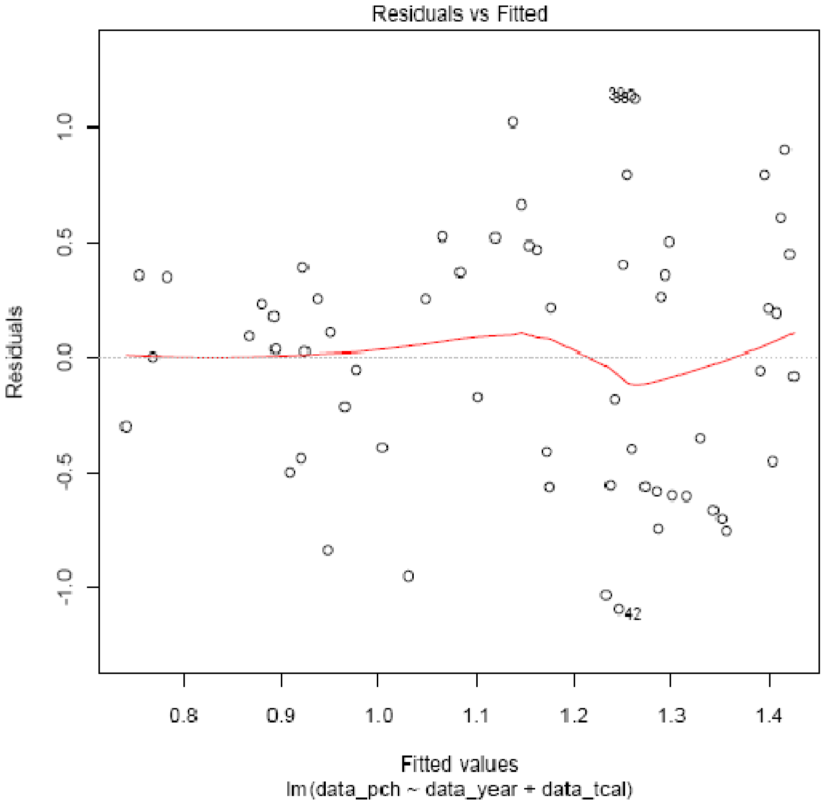

Fig 8. Plot of residual vs. fitted value for model (4.2)

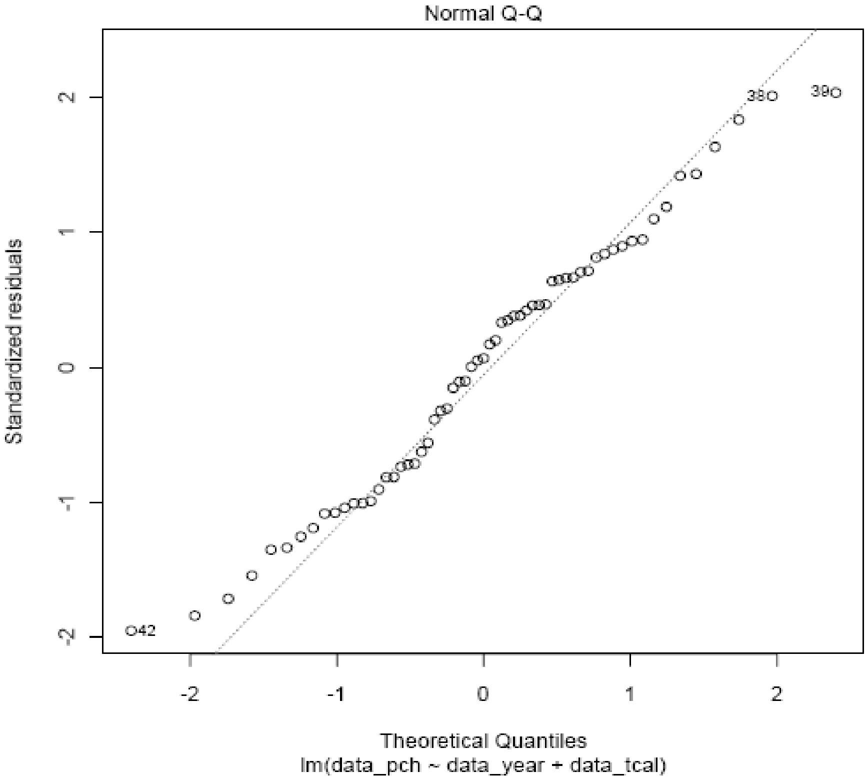

Fig 9. Plot of Q-Q norm for model (4.2)

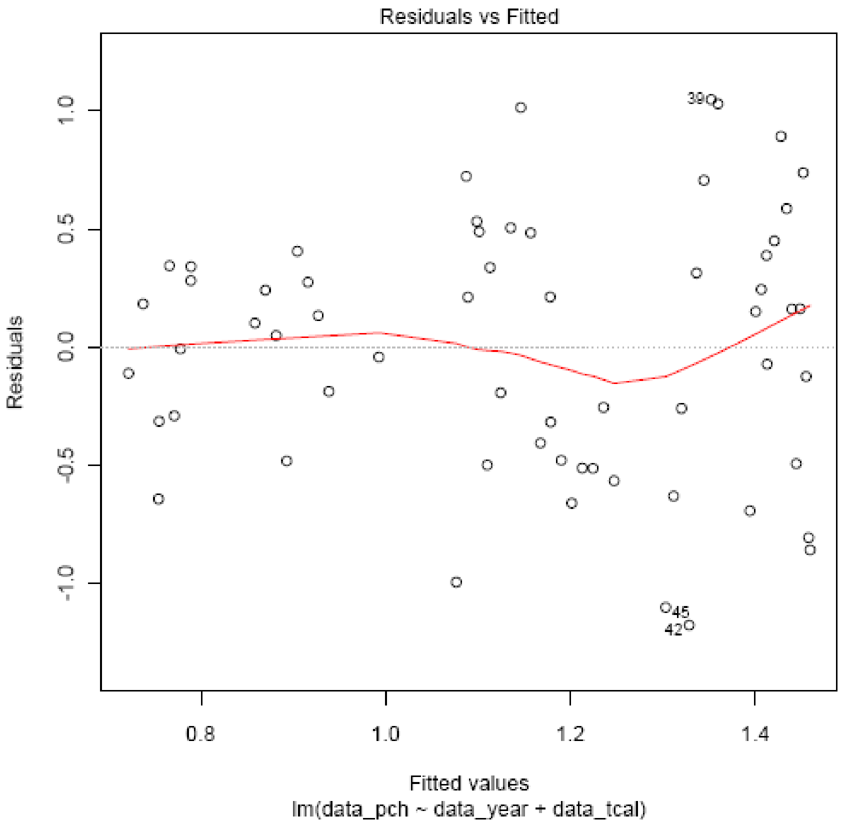

Fig 10. Plot of residual vs. fitted value for model (4.3)

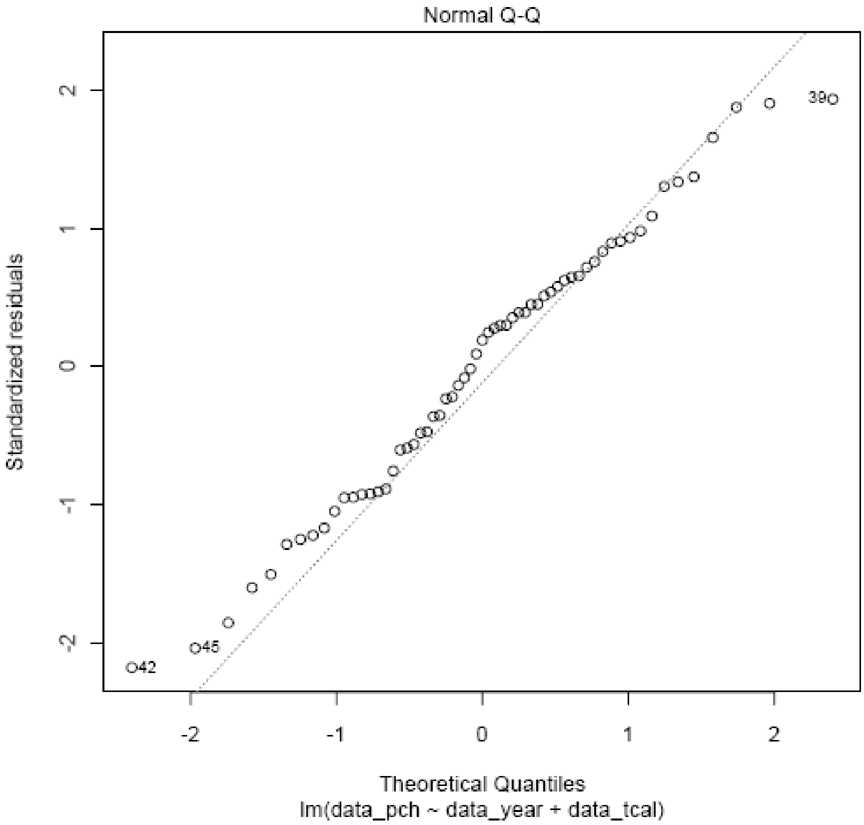

Fig 11. Plot of Q-Q norm for model (4.3)

From the result and plot we can see, when the volcanoes is taken into consideration, the average temperature, either the continuous five years average temperature or simple five years average temperature, is significant in the model (4.2) and (4.3) and the data is fitted much better than data in model (4.1). However, the assumptions of error which is random variable and following a normal distribution are not quite fit. When the response variable in model (4.2) and (4.3): population growth rate, is changed to population, we can still get the same conclusion.

5. Discussion

Estimated temperature based on sediment data from Stora is a quite good index to approximately reflect the historical climate situation in Iceland, especially five years and 10 years average temperature which shows high correlation with instrumental temperature. Of course, there are still other sources of proxy evidences to reconstruct the temperature of Iceland, such as Ice core data which is not considered in this report. However, since we don't have enough instrumental data set to compare with, it is difficult to conclude that estimate data is good enough for such a long time period.

When we simply build the model without taking natural hazard into consideration, the model doesn't fit very well with the data and temperature is statistically significant. While the natural hazard is included in the model, temperature can be statistically significant like model (4.2) and (4.3), which shows tight relationship between temperature and population growth.

Climate change and population is a very complicate problem. Even we choose to analysis the data from a country with simpler economic and population structure, it is still very difficult to draw a conclusion that whether the temperature is closely related to population growth.

6. Reference

[1] A 2000 years record of climate variations reconstructed from Haukadalsvatn, West Iceland, Journal of Paleolimnol, 2009.

[2] GISS Surface Temperature Analysis:
http://data.giss.nasa.gov/cgi-bin/gistemp/findstation.py?datatype=gistemp&data_set=1&name=&world_map.x=338&world_map.y=60

[3] Paleoclimatology: How can we infer past climates?
http://serc.carleton.edu/microbelife/topics/proxies/paleoclimate.html

[4] Statistics Iceland
http://www.statice.is

[5] Surface temperature reconstructions for the last 2,000 years, Board on Atmospheric Science and Climate, The national academies, 2006.
http://www.nap.edu/openbook.php?record_id=11676&page=R1

[6] Temperature Proxy Records Covering the Last Two Millennia: a Tabular and Visual Overview, Fredrik Charpentier Ljungqvist, Department of History, WDCA contribution series number: 2009-07
ftp://ftp.ncdc.noaa.gov/pub/data/paleo/contributions_by_author/ljungqvist2009/ljungqvist2009recons.txt

[7] The dynamic climate of Iceland
http://www3.hi.is/~oi/climate_in_iceland.htm

[8] Travel Guide to Iceland
http://www.icetourist.is/displayer.asp?cat_id=269

APPENDIX

Other models that we considered to fit the data and their results:

1. $\log(\text{population}) = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 \text{tem}_{ave5} + \varepsilon$ (P-1)

Where: t —time of the year

tem_{ave5} —sediment estimate temperature averaged over 5 years

ε —error, following a normal distribution with constant variance.

After fitting the data, we have following result in Table p1:

Coefficients	Estimate	Std. Error	t value	Pr(>|t|)
β_0	8.82971	1.31106	6.735	5.73e-10
β_1	1.10105	0.09343	11.785	<2e-16
β_2	0.35204	0.22444	1.569	0.119
β_3	0.24824	0.16468	1.507	0.134
Residual standard error: 0.05714 on 122 degrees of freedom				

Table p1. Result for fitting the data to model (P-1)

2. $\text{population change rate} = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 \text{tem}_{ave5} + \varepsilon$ (P-2)

Where: t —time of the year

tem_{ave5} —sediment estimate temperature averaged over 5 years

ε —error, following a normal distribution with constant variance.

After fitting the data, we have following result in Table p2:

Coefficients	Estimate	Std. Error	t value	Pr(>|t|)
β_0	-53.955	40.876	-1.320	0.1893
β_1	5.783	2.913	1.985	0.0494
β_2	-6.776	6.998	-0.968	0.3348
β_3	6.822	5.134	1.329	0.1864
Residual standard error: 1.782 on 122 degrees of freedom				

Table p2. Result for fitting the data to model (P-2)

3. $\text{Average population} = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 \text{tem}_{ave5} + \varepsilon$ (P-3)

Where: Average population—average population data over five years

t —time of the year

tem_{ave5} —sediment estimate temperature averaged over 5 years

ε —error, following a normal distribution with constant variance.

After fitting the data, we have following result in Table p3:

Coefficients	Estimate	Std. Error	t value	Pr(>|t|)
β_0	195713	55592	3.521	0.000613
β_1	44691	2628	17.004	<2e-16
β_2	56942	8887	6.407	3.15e-09
β_3	-18391	6992	-2.630	0.009671
Residual standard error: 2453 on 118 degrees of freedom				

Table p3. Result for fitting the data to model (P-3)

4. $\log(\text{population}) = \beta_0 + \beta_1 t + \beta_2 \text{tem}_{ave10} + \varepsilon$ (P-4)

Where: t —time of the year

tem_{ave10} —10 years LAGGED average sediment estimate temperature

ε —error, following a normal distribution with constant variance.

After fitting the data, we have following result in Table p4:

Coefficients	Estimate	Std. Error	t value	Pr(>|t|)
β_0	-1.4641	0.2706	-5.410	1.46e-07
β_1	0.0053334	0.0002049	26.024	<2e-16
β_2	0.3363	0.0401	8.392	3.41e-15
Residual standard error: 0.1683 on 253 degrees of freedom				

Table p4. Result for fitting the data to model (P-4)

Note: in this model I am using the data from 1745-2000 for both temperature and population.

Discussion of model (P-3) and (P-4)

In these two models, the results show that the temperature is significant, but when we check model validation by using graphic methods, it suggests that assumptions are violated and the data is not fitting the model very well. So, we cannot jump to conclusion that temperature is statistically significant in these two models.

5. $\text{population} = \beta_0 + \beta_1 t + \beta_2 \text{tem}_{avg5} + \varepsilon$ (P-5)

Where: t —time of the year

tem_{ave5} —sediment estimate temperature averaged over 5 years

ε —error, following a normal distribution with constant variance.

After fitting the data, we have following result in table p5:

Coefficients	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**
β_0	-1.176e+06	6.194e+04	-18.99	<2e-16
β_1	3.403e+02	1.277e+01	26.64	<2e-16
β_2	7.628e+04	5.066e+03	15.06	<2e-16
Residual standard error: 2181 on 58 degrees of freedom				

Table p5. Result for fitting the data to model (P-5)