# An analysis of proxy indicator data in climate change science

Tazim Chowdhury & Chin Yung Isabel Yu
Queen Mary University of London

Using techniques learnt in computational data analytics, we aimed to analyse proxy indicator data which act as indirect measurements of past climates prior to availability of instrument measurements. Amongst available datasets, we analysed tree ring, borehole, and ice core data using packages in Python such as pandas and Matplotlib, with methods such as random forest regression, linear regression, and Bayesian ridge analysis. Our results indicate that certain proxy indicators are more reliable than others, and the same applies for certain datasets for the same proxy indicator. Having determined this, in future analyses it would be beneficial to cross reference with more comprehensive and varied datasets, as well as different proxy indicators, such as corals, sediments, and pollen.

## 1. Introduction:

Climate change is, in its most general definition, a significant change in the distribution of weather patterns in the climate system lasting an extended period of time, regardless of its cause (Hughes et al. 2003). Climate change is not an unprecedented event- climate change events in recent geological past are mostly cycles of cold glacial periods (ice ages) recurring approximately every 100,000 years (Hansen & Sato 2012). Our research interest lies mostly in the field of major greenhouse phases, which provides insight to the imminent effects of global warming (Flohn 1979).

To provide evidence for climate changes, different types of proxy indicators can be recorded. In statistics, a proxy indicator is a variable which is not directly relevant but acts as an indicator of an otherwise unmeasurable variable (Dodge et al. 2003). These proxy indicators can be physical, chemical or biological, and the extent of the length of time covered by the records varied (von Storch et al. 2004). Most records date back to, at best, a few hundred years ago, but some geological data date back to much earlier. These proxies can be high-resolution, such as corals, tree rings, ice cores, and laminated sediments, and lower resolution proxies are used for supplementing the high-resolution data. These include boreholes, non-laminated sediments, glacial moraines etc. It is also worth noting that interpretations of proxy indicator data are complicated by noise and other distortions (von Storch et al. 2004).

Data from proxy indicators aims to reconstruct climate and climate change information predating historic record. The study of doing so is called paleoclimatology (Crowley & North 1991). Furthermore, models created from paleoclimate information sheds light onto the potential response of components of the Earth's systems to climate change through extrapolation. Ultimately, this also provides perspective and acts as a guide to politicians and world leaders to create policies and guidelines to adapt to and mitigate the effects of climate change (Kelly & Adger 2000).

### 1.1. Tree Rings:

Dendroclimatology, also known as tree-ring dating, is the study of past climates from annual tree rings (Schweingruber 1996). Secondary growth in tree stems result in a growth in diameter as through new growth in the tree's vascular cambrium. This results in the addition of a new visible ring in the stem (Schweingruber 1996). As tree growth is influenced by climate conditions, these climate conditions can be deciphered from the patterns such as ring

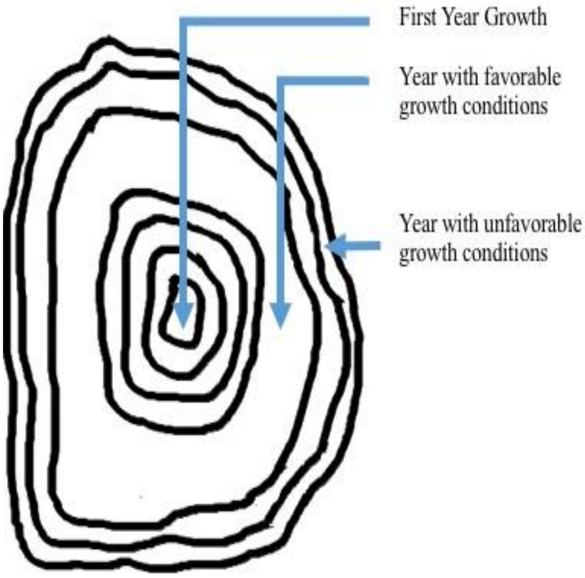


*Figure 1. An idealised horizontal section of a tree stem showing secondary growth and resulting tree rings. Rings with wider diameters represent years of more favourable growth and the opposite is true.*

width, isotopic composition, and density (Fritts 1972).

Ideally, tree rings would be clearly defined and visible, as well as forming strictly annually, reflecting a complete cycle of the season. Biological responses to climate forcing would also be stationary and not evolving. In this case, a wider ring would represent a year with more favourable growth conditions, such as volume of rainfall, atmospheric composition, soil nutrient composition, and most relevantly, temperature. The opposite case is also true – narrower tree rings reflect years with more unfavourable conditions. This is illustrated in Figure 1.

**1.2. Boreholes:**

Borehole temperature records are measured by drilling holes into the ground at different locations globally (Clauser & Mareschal 1995). With understanding of basic physics of geothermal heat flow, paleoclimatic information of surface temperature changes can be reconstructed.

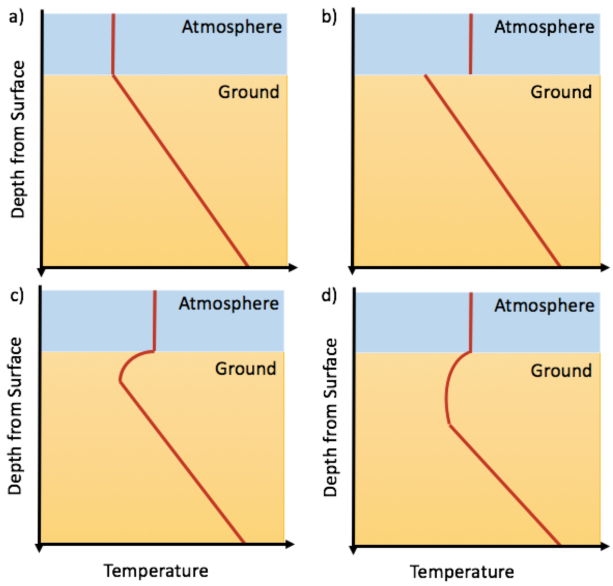


*Figure 2. An illustration of heat dispersion from the atmosphere into the ground as atmospheric temperature becomes higher than surface temperature. Figures a to d represent the passing of time.*

The hot core of the Earth of about 5000°C and the relatively cool surface of the Earth together create a geothermal gradient as we descent into the depth of the earth. This means, as heat is transferred upwards to the surface, it creates a steady rate of increase of temperature of typically 20-30°C per kilometre (Huang et al. 2000).

When atmospheric temperature increases and is warmer than the surface, heat is transferred from the surface into the ground instead. Therefore, the temperature downhole is a function of time-dependent surface and core heating (Overpeck 2000). This process of heat dispersion is illustrated in Figure 2. This process of thermal diffusion is so slow such that by measuring borehole temperatures at different depths, paleoclimatic surface temperatures can be reconstructed.

**1.3. Ice cores:**

Ice core data are collected from high mountain glaciers or polar ice sheets (Thompson et al. 2003). As layers of the ice accumulate annually, each layer of ice captures information of the climate the year it was formed. Each layer contains dust, air bubbles, as well as oxygen and hydrogen isotopes (Alley 2000). For example, the trapped air bubbles give insight to atmospheric gas composition, and isotopes reveal information about paleoclimatic temperatures (Cronin 1999). Together, these properties are used to reconstruct climate models for the past. To obtain information of the different layers of ice, different cores are drilled down with the deepest core at 3769m (Petit et al. 1999).

## 2. Methods and materials:

We obtained data for tree rings, boreholes, and ice cores from available online databases, and different types of analyses are performed on different types of data using Python and related packages and extensions. This process is described below.

### 2.1. Computational Packages

Python is used as the primary language for data analysis in this project (Rossum & Drake 2010). Extensions and packages used include NumPy (Oliphant 2006), pandas (McKinney 2011), scikit-learn (Pedregosa & Varoquaux 2011) and Matplotlib (Hunter 2007). NumPy is chosen for its ability to work quickly with large arrays, pandas for its ability to represent data as dataframes, Scikit-learn for machine learning, and Matplotlib for its visualisation. In general, the modelling process is mainly based on four steps:

1) Manipulate the input data to create arrays, this is done by using NumPy library
2) Instantiate/ call a model from the SciKit
3) Fit the data to the model
4) Make prediction using the model

### 2.2. Sources of data

Tree ring data are obtained from the International Tree-Ring Data Bank from the National Centers for Environmental Information, available online (NCEI n.d.). Tree ring data range from countries from both the northern and the southern hemisphere, ranging from USA, Sweden, Russia, Norway, and Canada. Tree ring data date back to 1600s and temperature records date back to 1850s.

Borehole data are obtained from the NCEI (NCEI n.d.). This measurement of temperature and depth descending the borehole was recorded in Australia from 1964 to 1980.

Ice core data are obtained at Dome C from the European Project for Ice Coring in Antarctica (EPICA) (Lüthi et al. 2008). This includes data about the years before present day (BP), temperature, $CO_2$ levels, and solar irradiance. BP date back to 800,000 ago.

### 2.3.Methods of analysis

Tree ring data is analysed through the correlation between the tree ring width and temperature records. This is done mostly through linear regression, which fits a line of best fit through the data and measures the Goodness-of-Fit. Extrapolation can then be used to reconstruct paleoclimatic temperature (Su et al. 2012).

Borehole data are studied by correlation of temperature and depth. The geothermal gradients are also studied and compared in the different years.

Temperature of the ice and $CO_2$ levels in air bubbles of the ice are studied using random regression analysis and Bayesian ridge regression analysis.

## 3. Results & Discussion:

### 3.3. Tree Rings:

First, exploratory analyses were performed on the available. Plots were visualized for

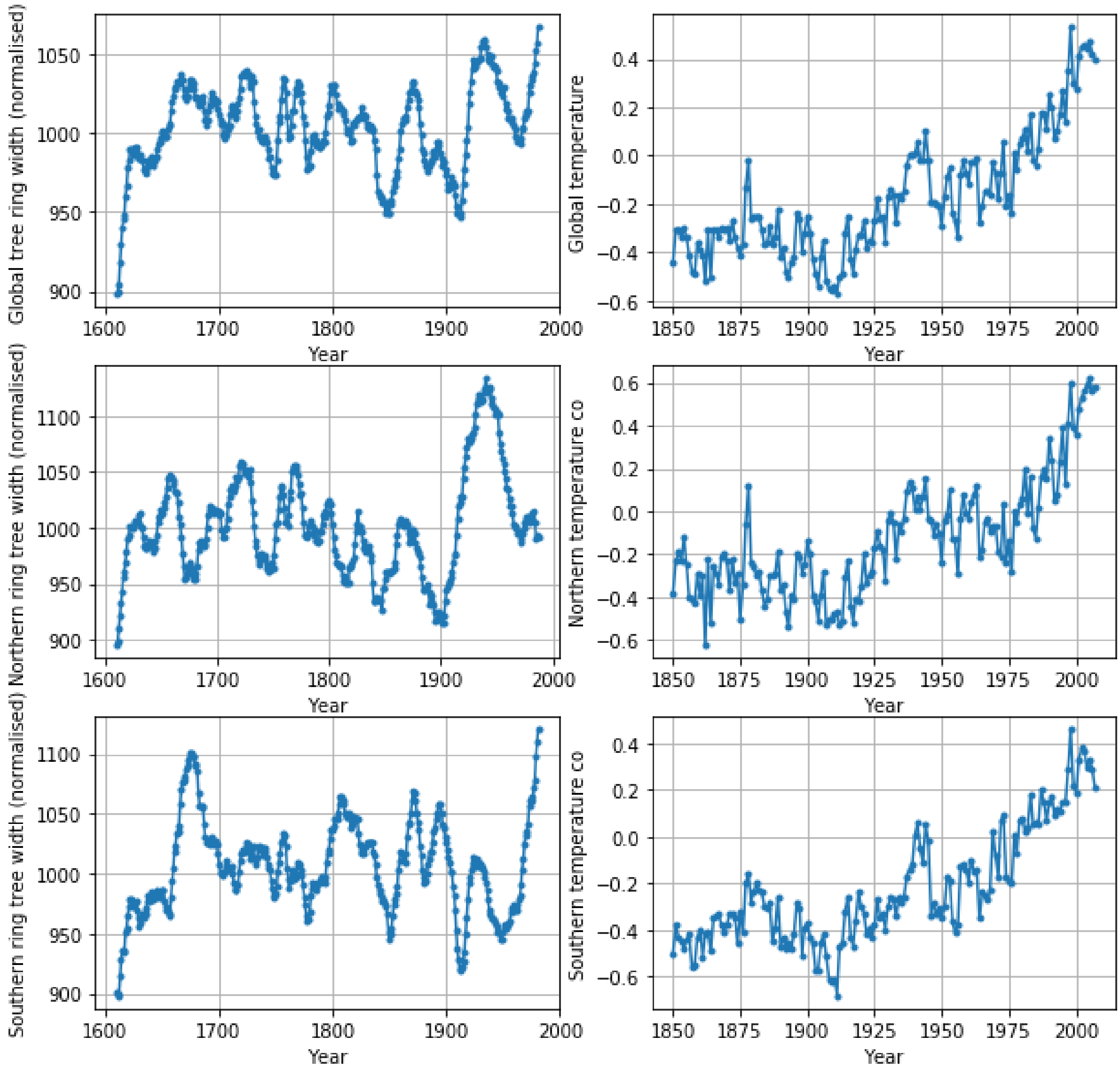


*Figure 3. Visualisation of exploratory analysis of year to tree ring width data and temperature globally, for the Northern Hemisphere, and the Southern Hemisphere.*

| *REGION* | *CORRELATION COEFFICIENT (R)* | *Y-AXIS INTERCEPT* | *REGRESSION COEFFICIENT* | *$R^2$ VALUE* |
|---|---|---|---|---|
| *GLOBAL* | 0.604897414159 | -3.54033779359 | 0.00325884 | 0.365900881657 |
| *NORTHERN HEMISPHERE* | 0.575903469915 | - 1.96624701937 | 0.00173341 | 0.199603610257 |
| *SOUTHERN HEMISPHERE* | 0.174626844877 | -1.00130450841 | 0.00069215 | -0.0865761281716 |

*Table 1. Statistical analysis of the correlation and properties of linear regression analysis between temperature and tree ring width in the 3 regions.*

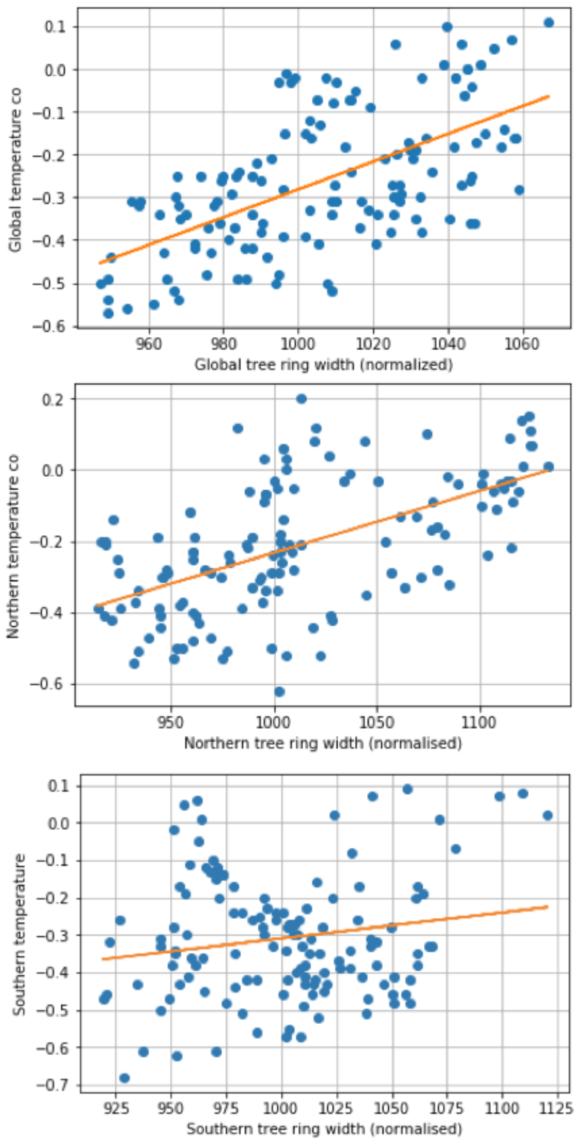


*Figure 4. Correlations between temperature and ring width for tree ring data for all three regions. The lines of best fit are also plotted.*

the three datasets used: global, northern hemisphere, and southern hemisphere. Temperature and tree ring width are plotted against year, as shown in Figure 3. The time series are seen with many fluctuations and step changes in general. At a glimpse, we can see that there are general increases in tree ring width and temperature no matter the region, though the increase is more apparent in temperature than in ring width.

With the knowledge that temperature and tree ring width are scientifically known to be coupled, we aim to analyse the Goodness-of-Fit between temperature and tree ring width globally, for the Northern hemisphere, and for the Southern hemisphere. First, the correlation coefficients are calculated, as shown in Table 1. Linear regression analyses are then applied between temperature and tree ring width for the 3 regions using Scikit-learn. The resulting regression coefficients, intercepts, and r-squared values are then calculated. The results are summarised in Table 1. Plots correlating the temperature and the tree ring width for the three regions are visualized in Figure 4, including their lines of best fit.

As stated previously, tree ring widths are used as an indicator of temperature. By correlating known temperatures and tree ring widths, the lines of best fit can be extrapolated to estimate temperatures given tree ring widths for times when temperatures were not yet recorded. This can be done by using the general formula $y=mx+c$ for the line of best fit, with m being the regression coefficient and c being the y-axis intercept. The general formulas globally, for the Northern hemisphere, and the Southern hemisphere are therefore respectively (to *3s.f.*):

$$Temp = 0.00325 * Width - 3.54 \qquad (a)$$
$$Temp = 0.00173 * Width - 1.66 \qquad (b)$$
$$Temp = 0.000692 * Width - 1.00 \qquad (c)$$

This method of climate reconstruction is not, however, without its imperfections. This is evident from the varying levels of correlation in our data, as seen through the correlation coefficient (R) and the $R^2$ value for variance. While the Northern hemisphere data and the global data have a significant level of correlation, the Southern Hemisphere data has low correlation between temperature and tree ring width. This could be due to other confounding factors affecting tree growth, such as moisture level, soil, sunlight and wind. As such, extrapolation of the Southern hemisphere should be used with caution. Moreover, tree ring data do not cover the entire Earth due to oceanic and arctic regions (Schweingruber 1996). This measurement is therefore not the most representative of proxy data indicators for temperature reconstruction.

For the complete code for the analysis of tree rings please refer to the Appendix.

### 3.2. Boreholes:

For the 1964 borehole measurement, the correlation between the depth and the temperature are tested, and the correlation is very high: 0.9993711539984198. As stated previously, linear relationships are found between temperature and borehole depth. This relationship is shown in Figure 5. Having performed a test-split, we found that there is a very high accuracy of prediction using this model (0.997 *3.s.f.*) (Ross et al. 2009). We have plotted the other data as well, as shown in Figure 7. This data set ranges from measurements of the same borehole from 1964, 1970, 1971, 1972, 1974, 1975, 1977, to 1980. Evidently, the levels of correlation are very high.

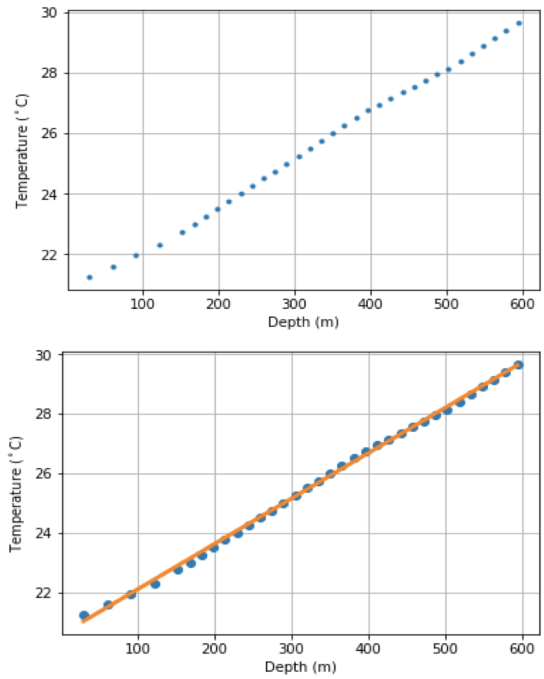


*Figure 5. Temperature along the depth of a borehole in Australia in 1964. Linear regression is applied to the data and as seen from the line-of-best-fit in the second plot.*

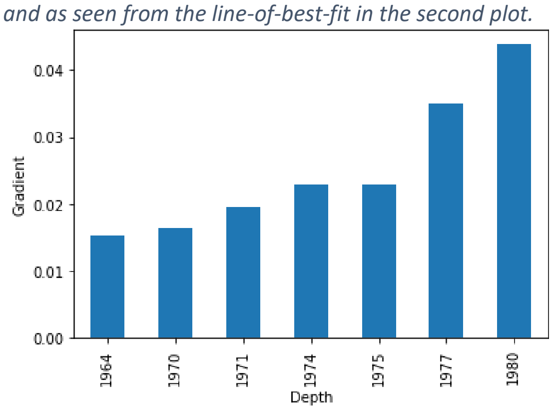


*Figure 6. Increase in gradient over time of change in temperature down the borehole.*

The gradients of each year's line-of-best-fit are as follows: 0.01526204, 0.01627011, 0.01956692, 0.0228279, 0.02293327, 0.03503116, 0.04383526. It is of interest to note these changes in gradient, as visualized in Figure 6. Climate implications of this significant increase in gradient will need to be further explored in depth to understand the meaning.

Unlike the tree ring data, however, we are not interested in good fits to the linear model. We are interested in periods of climate change which are reflected as change of the gradient in the temperature along the depth of the borehole. Therefore, we are looking for lines of best fits with less perfect correlations. This can be achieved through measuring the correlations of each year. Alternatively, in future studies, we suggest measuring the gradient along the depth of each record, and identify depths of significantly different gradients so that each the depth and therefore year of significant climate warming can be identified. Our data shows very good fits to a linear model and therefore leaves room for consideration for the degree of variance from the linear model required to identify a significant thermal transfer.

For the complete code for the analysis of boreholes please refer to the Appendix.

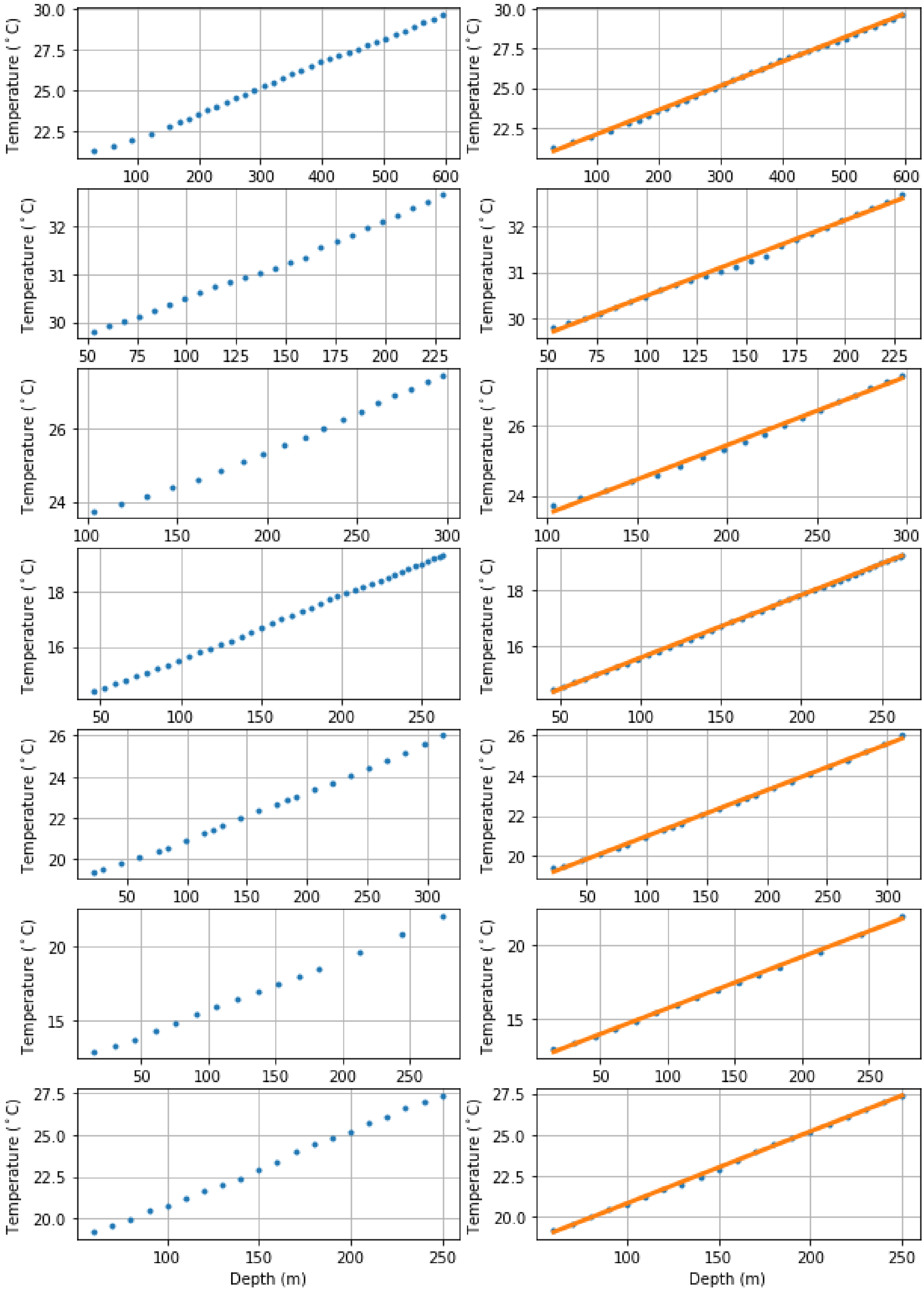


*Figure 7. Borehole temperature and depth from 1964, 1970, 1971, 1972, 1974, 1975, 1977, to 1980 from top to bottom of Figure. Lines-of-best-fit are shown on the RHS sub-figures.*

### 3.3. Ice cores:

For the ice core data, we are interested in correlating the ice core's temperature and the level of $CO^2$ in the air bubbles found in the ice. The data is shown in Figure 8. Random forest regression, linear regression and Bayesian ridge regressions are used to varying results, as shown in the Appendix.

For the complete code for the analysis of ice cores please refer to the Appendix.

## 4. Future implications:

In this project, only three types of proxy indicator data are considered. This gives us varied results on the resolution of each type of data as proxy indicators. For future studies, in terms of data used, other types of indicators should also be studied. In particular, high resolution data such as coral reefs and ice cores should be included. Moreover, sample sizes of each data set should be larger for improved representation and cross-validation. Many other types of models can also be considered for the study.

## 5. Appendix:

Please see attached Jupyter Notebooks for complete code for the above analysis, separated into the files of Tree Rings, Boreholes, and Ice cores.

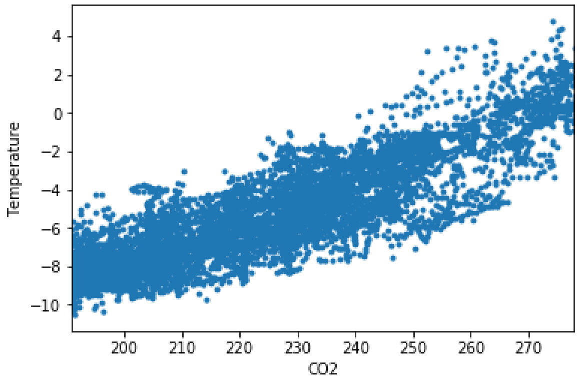

*Figure 8. Correlation of $CO_2$ levels and temperature at the ice core at Dome C from the EPICA project.*

## 6. References

Alley, R.B., 2000. Ice-core evidence of abrupt climate changes. *Proceedings of the National Academy of Sciences*, 97(4), pp.1331–1334. Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.97.4.1331.

Clauser, C. & Mareschal, J.C., 1995. Ground temperature history in Central Europe from borehole temperature data. *Geophysical Journal International*, 121(3), pp.805–817. Available at: y:%5CARCHIV_GEOWATT%5CPaper%5CPaper_archiviert%5CClauser_1995_GJI_GSTHistoryCentralEuropeFromBHT.PDF.

Cronin, T.M., 1999. *Principles of paleoclimatology*,

Crowley, T.J. & North, G.R., 1991. *Paleoclimatology*, New York, NY (United States); Oxford University Press. Available at: https://www.osti.gov/biblio/6506929 [Accessed April 3, 2018].

Dodge, Y., Marriott, F.H.C. (Francis H.C. & International Statistical Institute., 2003. *The Oxford Dictionary of Statistical Terms*, Oxford University Press. Available at: https://books.google.co.uk/books?hl=en&lr=&id=_OnjBgpuhWcC&oi=fnd&pg=PA1&dq=Oxford+Dictionary+of+Statistics+terms&ots=VtKnWTlIAL&sig=f4joHpn5oY5YcZ_OJ9VETJ_Payg#v=onepage&q=Oxford Dictionary of Statistics terms&f=false [Accessed April 3, 2018].

Flohn, H., 1979. Can Climate History Repeat Itself? Possible Climatic Warming and the Case of Paleoclimatic Warm Phases. *Developments in Atmospheric Science*, 10, pp.15–28. Available at: https://www.sciencedirect.com/science/article/pii/B9780444417664500096

[Accessed April 3, 2018].

Fritts, H.C., 1972. Tree Rings and Climate. *Scientific American*, 226(5), pp.92–100. Available at: http://www.nature.com/doifinder/10.1038/scientificamerican0572-92.

Hansen, J.E. & Sato, M., 2012. Paleoclimate Implications for Human-Made Climate Change. In *Climate Change*. Vienna: Springer Vienna, pp. 21–47. Available at: http://www.springerlink.com/index/10.1007/978-3-7091-0973-1_2 [Accessed April 3, 2018].

Huang, S., Pollack, H.N. & Shen, P.-Y., 2000. Temperature trends over the past five centuries reconstructed from borehole temperatures. *Nature*, 403(6771), pp.756–758. Available at: http://www.nature.com/articles/35001556 [Accessed April 11, 2018].

Hughes, T.P. et al., 2003. Climate change, human impacts, and the resilience of coral reefs. *Science (New York, N.Y.)*, 301(5635), pp.929–33. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12920289 [Accessed April 3, 2018].

Hunter, J.D., 2007. Matplotlib: A 2D graphics environment. *Computing in Science and Engineering*, 9(3), pp.99–104.

Kelly, P.M. & Adger, W.N., 2000. Theory and practice in assessing vulnerability to climate change and facilitating adaptation. *Climatic Change*, 47(4), pp.325–352.

Lüthi, D. et al., 2008. High-resolution carbon dioxide concentration record 650,000-800,000 years before present. *Nature*, 453(7193), pp.379–382.

McKinney, W., 2011. pandas: a Foundational Python Library for Data Analysis and Statistics. *Python for High Performance and Scientific Computing*, pp.1–9.

NCEI, Paleo Data Search | Search | National Centers for Environmental Information (NCEI). Available at: https://www.ncdc.noaa.gov/paleo-search/?dataTypeId=18 [Accessed April 12, 2018].

Oliphant, T.E., 2006. Guide to NumPy. Available at: http://web.mit.edu/dvp/Public/numpybook.pdf [Accessed April 12, 2018].

Overpeck, J.T., 2000. The hole record. *Nature*, 403(6771), pp.714–715. Available at: http://www.nature.com/articles/35001697 [Accessed April 11, 2018].

Pedregosa, F. & Varoquaux, G., 2011. *Scikit-learn: Machine learning in Python*, Available at: http://dl.acm.org/citation.cfm?id=2078195.

Petit, J.R. et al., 1999. Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature*, 399(6735), pp.429–436.

Ross, K.A. et al., 2009. Cross-Validation. In *Encyclopedia of Database Systems*. Boston, MA: Springer US, pp. 532–538. Available at: http://www.springerlink.com/index/10.1007/978-0-387-39940-9_565 [Accessed April 14, 2018].

Rossum, G. Van & Drake, F.L., 2010. Python Tutorial. *History*, 42(4), pp.1–122. Available at: http://docs.python.org/tutorial/.

Schweingruber, F.H., 1996. *Tree rings and environment: dendroecology.*, Berne: Paul Haupt AG Bern. Available at: https://www.cabdirect.org/cabdirect/abstract/19980602476 [Accessed April 3, 2018].

von Storch, H. et al., 2004. Reconstructing past climate from noisy data. *Science (New York, N.Y.)*, 306(5696), pp.679–82. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15459344 [Accessed April 3, 2018].

Su, X., Yan, X. & Tsai, C.-L., 2012. Linear regression. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(3), pp.275–294. Available at: http://doi.wiley.com/10.1002/wics.1198.

Thompson, L.G. et al., 2003. Tropical glacier and ice core evidence of climate change on annual to millennial time scales. In *Climatic Change*. pp. 137–155.