5 | Reproducing experiments of ice nucleation in atmospheric chemistry

Fabian Herrmann[1], Evelyn Jantsch[2], Philipp Cimiano[1], Thomas Koop[2]

1 – Semantic Computing Group, Faculty of Technology & Cognitive Interaction Technology Excellence Center (CITEC), Bielefeld University
2 – Faculty of Chemistry, Bielefeld University

Abstract

This chapter describes a case study in reproducing results in the area of atmospheric chemistry. The specific result reproduced is described in the paper *'BINARY: an optical freezing array for assessing temperature and time dependence of heterogeneous ice nucleation'* by Budke and Koop [1]. The study investigated the conditions under which ice nucleation occurs using Snomax®, a commercial ice inducer containing freeze-dried nonviable bacterial cells from *Pseudomonas syringae*, as a test substance for the investigation of heterogeneous ice nucleation processes. The ice inducing bacterial cell agents are known to be active at high temperature and are used in snow cannons. The study considered a temperature range between 0°C and -12°C. The main result was the finding that two classes of nucleations occur at a number ratio of about 1 to 1000 in the chemical samples, based on the difference of 3 orders of magnitude of the temperature plateau values. As a result of the Conquaire project, we reimplemented the original workflow relying on OriginPro in Python and could reproduce the central figure of the above mentioned paper by Budke and Koop using free and open software. This thus counts as a case of full analytical reproducibility. The data and scripts for the paper by Budke and Koop are available at `https://gitlab.ub.uni-bielefeld.de/conquaire/atmospheric_chemistry`.

Keywords

Atmospheric Chemistry, Ice Nucleation, Pseudomonas syringae, Snomax

5.1 Introduction

The study of ice formation is an active research area in the atmospheric sciences [2]. For example, ice crystals occur in high altitude clouds and they are responsible for initiating most precipitation above continents [2, 3, 4]. From a thermodynamic point of view, crystalline ice is the stable phase of water below the melting temperature T_m, which is 0°C at ambient pressure, see Figure 5.1. In many cases, the formation of ice crystals is kinetically inhibited and can occur at lower temperature either via homogeneous or via heterogeneous nucleation, see Figure 5.1.
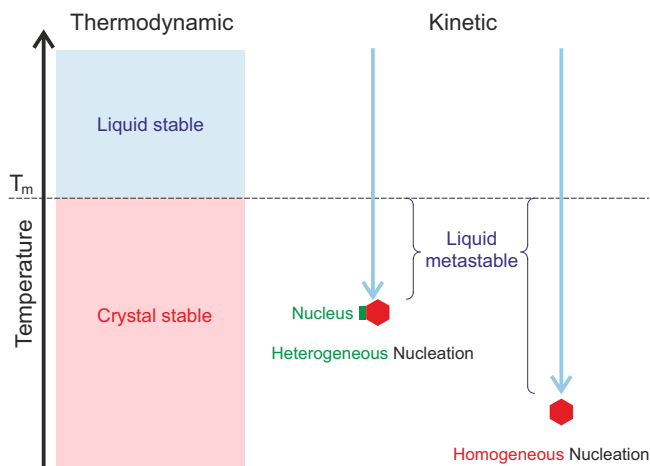

Figure 5.1: Schematic depiction of different nucleation mechanisms for the freezing of water. T_m is the melting temperature of the crystalline phase ice; adapted with changes from Koop, 2004 [5].

For homogeneous freezing, a number of water molecules have to arrange themselves into an ice-like cluster, termed critical ice embryo, in order to trigger ice formation. The size of this critical embryo is temperature dependent and decreases with decreasing temperature, thus making ice nucleation more likely at lower temperature. For example, micrometer-sized pure water droplets freeze homogeneously at approximately -38°C (homogeneous nucleation temperature) [5]. In contrast, heterogeneous ice nucleation can occur at higher temperatures – even close to the melting temperature of ice – depending upon the presence and activity of so-called ice nuclei (IN), see Figure 5.1 [5, 3, 6, 7]. Laboratory experiments can be employed to help understanding the processes that lead to ice nucleation in the atmosphere. By investigating ice nucleation temperatures of different IN, we can quantify different IN activities, which can be used for parametrizations of ice formation in atmospheric cloud models [8].

The activity of an IN material suspended in a water droplet can be obtained from the measured number of active sites per dry mass $n_m(T)$ as a function of

temperature. Equation 5.1 presents a definition for $n_m(T)$, where T is temperature, $K(T)$ is the experimentally observed cumulative number of active sites per volume of water, and C_m is the mass concentration of IN in the water.

$$n_m(T) = \frac{K(T)}{C_m} \tag{5.1}$$

$K(T)$ can be obtained from equation 5.2.

$$f_{ice}(T) = \frac{n_{ice}(T)}{n_{tot}} = 1 - e^{-K(T) \cdot V_{drop}} \tag{5.2}$$

Here, $f_{ice}(T)$ represents the experimentally observed cumulative ice fraction, which is defined by the ratio of the number of droplets frozen at temperature T, $n_{ice}(T)$, and the total number of investigated droplets, n_{tot}. V_{drop} is the droplet volume.

Established methods for the determination of heterogeneous ice nucleation temperatures have different advantages and disadvantages. For instance, larger droplet volumes encounter a higher probability of impurities. In contrast, smaller volumes are often realized through emulsions and, therefore, an oil phase is in contact with the water droplet, which may influence results for those IN (e.g. pollen and fungal spores) which have an affinity to hydrophobic phases, i.e. the concentration of suspended IN would be overestimated in such cases [9]. Many experimental techniques are droplet arrays based on a method originally developed by Vali and Stansbury, where small volume droplets are placed on a cooling stage [10, 11]. However, since no oil phase is used to enclose the droplets, frozen droplets grow by water vapor transport from the remaining supercooled liquid droplets, according to the Wegener-Bergeron-Findeisen process. Moreover, sometimes frost halos form around frozen droplets. These ice rings tend to grow on the surface below the droplets and can cause ice nucleation in adjacent supercooled droplets. Budke and Koop [1] took these potential shortcomings into account when they developed a new device to investigate ice nucleation termed BINARY (short for Bielefeld Ice Nucleation ARraY), which was used in the present study. The different droplets in BINARY are separated in individual compartments thus preventing water vapor transfer between neighboring droplets.

Snomax® is a well-studied IN material and, therefore, a good reference substance for testing new methods [12, 13, 14]. Snomax® is a commercial product containing freeze-dried cells from *Pseudomonas syringae*, a rod-shaped bacterium living on a variety of plants. *P. syringae* bacteria are known to induce heterogeneous ice nucleation at very high temperatures of approximately -2°C (class A) and also in a temperature range of about -7 to -10°C (class C) [15]. The latter study was a multi-group intercomparison project and also included data from the BINARY setup. Using this setup Budke and Koop determined $n_m(T)$ in a temperature range between 0°C and -12°C [1].

5.2 Methods

In this section, we describe the experimental settings and methods as well as the main results described in the paper by Budke and Koop [1].

5.2.1 Experiment settings and Data acquisition pipeline

In the study under investigation [1], the BINARY technique was used to determine heterogeneous ice nucleation temperatures of Snomax®. Therefore a certain dry mass of Snomax® (m) was suspended in freshly double-distilled water of volume V_{H_2O} to obtain the desired mass concentration $C_m = m/V_{H_2 0}$ of Snomax® in water. 36 droplets (each $V_{drop} = 1$ µL) of such a suspension were pipetted into the compartments of a polydimethylsiloxane (PDMS) lattice placed on a hydrophobic glass surface, resulting in a 6 x 6 droplet array as shown in Figure 5.2a. The droplet compartments are sealed with another glass slide on top of the PDMS lattice to prevent droplet evaporation (see Figure 5.2b).
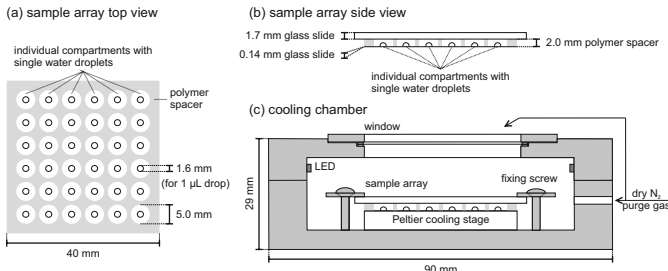

Figure 5.2: Schematic picture of the Bielefeld Ice Nucleation ARraY (BINARY) setup. **(a)** Top view of the 6 x 6 droplet array. The droplets are separated from each other by a polymer lattice creating individual compartments. **(b)** Side view showing the sealing of the compartments by top and bottom glass slides. **(c)** Position of the sample array on the Peltier cooling stage inside the cooling chamber. Figure is taken from Budke and Koop, 2015 [1].

This sample array is positioned on a Peltier stage within a cooling chamber (Linkam LTS120) as shown in Figure 5.2c. A metal frame presses the whole array onto the Peltier cooling stage with the help of fixing screws to assure a homogeneous and efficient heat transfer. The Peltier stage is connected to a heat sink bath at 5°C and its top side can be cooled to -40°C at cooling rates between 0.1 and 10°C min⁻¹. All experiments described below were measured at a cooling rate of 1°C min⁻¹. Small cold-light white LED stripes are fixed at the top edges inside the cooling chamber aiding the visualization of phase changes through light scattering (liquid droplets appear darker whereas ice crystals appear brighter due to the backscattered light). A CCD camera (QImaging MicroPublisher 5.0 RTV) is mounted above the whole setup to observe the

droplets through a 40 x 40 mm window in the top ceiling of the cooling chamber. Both the interior of the cooling chamber and the surface of the top window are purged with dry nitrogen during the experiment to prevent water condensation. A LabVIEW™ virtual instrument is used to detect ice nucleation and melting events from the digital images obtained by the CCD camera. In detail, for each compartment the average gray value of all pixels within a predefined area is obtained. These gray values range between 0 for black pixels and 255 for white pixels.

Figure 5.3b and c show a representative behavior of the gray values and their changes for the compartment marked by a yellow box in panel (a). Starting with the red curve at 4°C, the average gray value in Figure 5.3b is almost constant until the droplet freezes at -3.9°C, as indicated by a sharp jump. This steep increase is also shown as the derivative in Figure 5.3c. After a temperature of -10°C is reached, heating is started (green curves) and ice melting begins at 0°C, again indicated by a gray value change. The thresholds for defining the occurrence of nucleation and melting events are gray value changes larger than 1 and -1, respectively (dotted red and green lines in Figure 5.3c).

5.2.2 Methods applied to analyze the experimental data

For each individual droplet, the uncalibrated heterogeneous ice nucleation temperatures T_{nuc} are obtained and saved in a text file for offline calibration and further analysis. The temperature calibration function and how it was developed from experiments is discussed in detail in the paper [1]. Briefly, the calibrated nucleation temperature T_{cal} can be obtained using equation 5.3, where r is the cooling rate of 1°C min^{-1}.

$$T_{cal} = -((-6.03165) + 0.02113 \cdot (273.15 + T_{nuc}) - (3.59774 + (-0.02956)$$
$$(273.15 + T_{nuc}) + 6.10156 \cdot 10^{-5} \cdot (273.15 + T_{nuc})^2) \cdot (-r) + T_{nuc} \quad (5.3)$$

Each T_{cal} value is then binned into temperature intervals of 0.1°C width, i.e. all T_{cal} values within the interval $X_{low} \leq T_{cal} < X_{up}$ get sorted into the bin X_{low}. Thereafter, T_{cal} will only be used as the binned value T. Now the number of individual T_{cal} data are counted to gain $n_{ice}(T)$ and n_{tot} for determining f_{ice} using equation 5.2. This counting is done for all droplets with the same Snomax® concentration, so each measured concentration has one cumulative ice fraction ranging from 0 to 1. Using equation 5.4 (derived from equation 5.1 and 5.2) $n_m(T)$ is obtained for each concentration and can be plotted for all investigated temperatures.

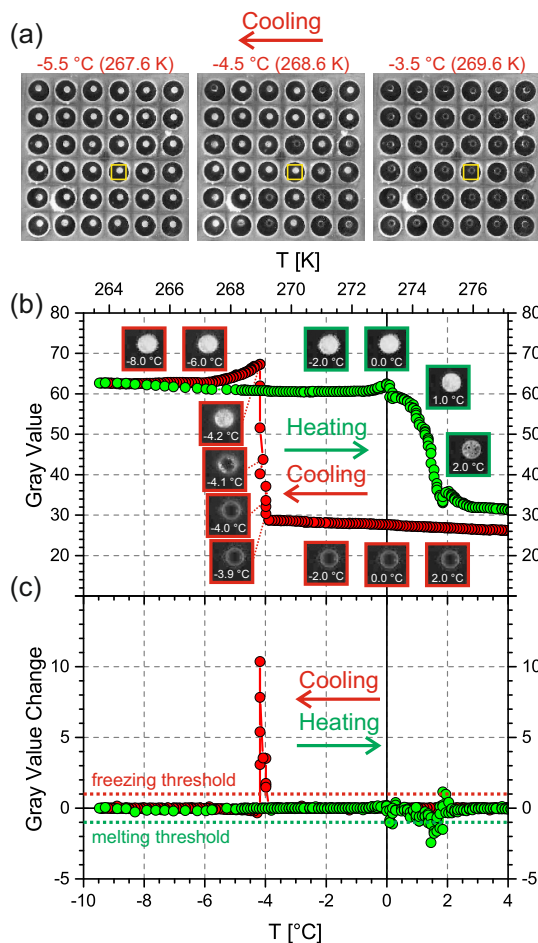$$n_m(T) = \frac{-\ln(1 - f_{ice}(T))}{C_m \cdot V_{drop}} \quad (5.4)$$

Figure 5.3: Typical experiment with Snomax®-containing droplets (0.1 $\mu g\,\mu L^{-1}$) showing the automatic detection of ice nucleation events by the change in brightness during freezing. **(a)** Image series of the 6 x 6 droplet array during cooling. **(b)** Measured gray value of the droplet compartment indicated by the yellow box in panel (a) during cooling (red) and heating (green). Freezing and melting start at -3.9°C and 0.0°C, respectively. **(c)** Plot of the change in gray value between successive images showing peaks at the phase transition temperatures. Threshold values of ±1 for the automatic attribution of freezing and melting are indicated by the dashed lines. Figure is taken from Budke and Koop, 2015 [1].

5.2.3 Main Results

Figure 5.4 presents the main result of the paper in form of a combined curve of $n_m(T)$ values from several Snomax® suspensions with different concentrations (see color code).

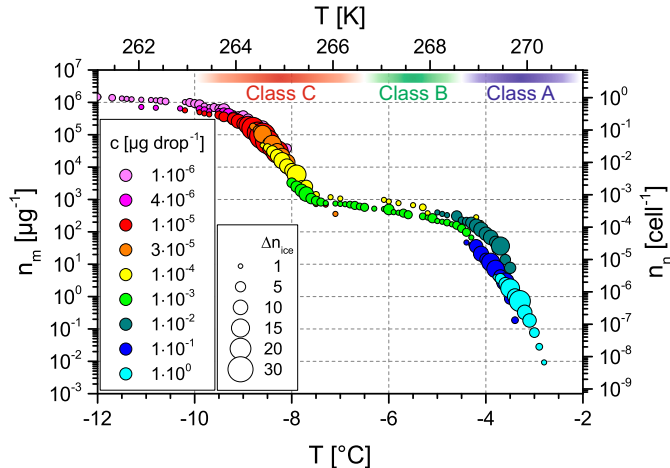

Figure 5.4: Experimentally determined active site density per unit mass of Snomax® $n_m(T)$ versus temperature. Symbol colors indicate data from droplets with different Snomax® concentrations; symbol size indicates the number of nucleating droplets per temperature interval (0.1°C). The temperature range for different classes of IN is also indicated by the colored bars. Figure is taken from Budke and Koop, 2015 [1].

Two steep increases can be seen, which represent two different types of IN activity at different temperature regimes. Plateaus in a $n_m(T)$ plot, e.g. between -4.5°C and -7.5°C, can be interpreted as temperatures where no IN is active. It should be noted that data points below -12°C down to -35°C were obtained, but are not shown since they did not reveal any other IN (purple symbols). Also indicated in Figure 5.4 are the temperature ranges for different IN classes as defined in the literature [16]. Two different classes of IN in Snomax® were identified, inducing ice nucleation at about -3.5°C (class A) and at about -8.5°C (class C). For class A IN in Snomax® $n_m(T)$ ranges from about 10^{-2} μg^{-1} up to almost 10^3 μg^{-1}. However, the number of active sites is much larger for class C IN as the increase starts at 10^3 μg^{-1} rising to almost 10^6 μg^{-1}, indicating that class C IN are about a factor of 10^3 more abundant than class A IN. The number of active sites can also be expressed as a number of active sites per cell (i.e., $n_n(T)$ on the right axis in Figure 5.4). Hence, there is about one class C active site per *P. syringae* cell.

5.3 Analytical Reproducibility

As a main objective of this study, we defined the goal of being able to independently reproduce the plot shown here in Figure 5.4 as main result of the work described by Budke and Koop [1]. The validation was done by calculating the calibrated temperatures from the temperature for a given cooling rate and Snomax® concentration; for each concentration bin, the $f_{ice}(T)$ was calculated. The calculations had been done originally using OriginPro for the original paper, while we reproduced these calculations using a custom Python program.

5.3.1 Research Data - Primary

The data was read off the BINARY experiment setup, then processed entirely by OriginPro, a proprietary computer software from OriginLab Corporation, that is mainly used for interactive scientific graphing and data analysis on the Microsoft Windows platform only. It is a GUI software with a spreadsheet-like front end which uses a column-oriented data processing approach for calculations. It has its own file format, **.OPJ**, for project files which are directly processed by the system for statistics, data analysis and visualization.

The group uses OriginPro along with a scripting language known as **LabTalk** that allows finer control, by writing small macros that run over the data analysis process for the experiment data. With LabTalk the group programs routine operations, including batch operations, with customizable graph templates and analysis dialog box themes. Various features exist to save a collection of operations within the workbook, viz., saving a suite of operations, auto recalculation on changes to data or analysis parameters, and different analysis templates.

5.3.2 Research Data - Analyzed and Processed

The Snomax® data file contains data from the OPJ data file that is read into the Origin software system. The data was exported into tab-separated files with OriginPro as *.txt files with six TAB delimited columns. The calibration data numbers start from line four with the headers confined to the first three lines; viz. the first line has the column names, while line 2 contains the data description or unit, and the third line contained information about the substance.

For the computational reproducibility experiment, we used Python to process these text files for data analysis and visualization based on the validated raw data. After calibrating the temperature, the python script binned the data, then grouped the data for all columns by concentration (decreasing) into different bins and then within each concentration bin the data is sorted by (decreasing) calibrated temperature T_{cal}. Afterwards, $f_{ice}(T)$ was determined for each temperature value in each bin. In the last step the mass concentration of Snomax® and the volume of the droplets are converted into the active site density per unit mass, $n_m(T)$.

After tabulating $f_{ice}(T)$ and $n_m(T)$ for each concentration bin, the resulting data is stored in a csv file that became the input data to reproduce the plot from the original paper shown in Figure 5.4.

5.3.3 Data Workflow Lifecycle

In order to reproduce the mentioned plot, the functionality implemented originally in the OriginPro frameworks was reproduced using a Python program. The resulting workflow implemented in Python reproduces the workflow implemented in OriginPro and schematically represented in Figure 5.5.
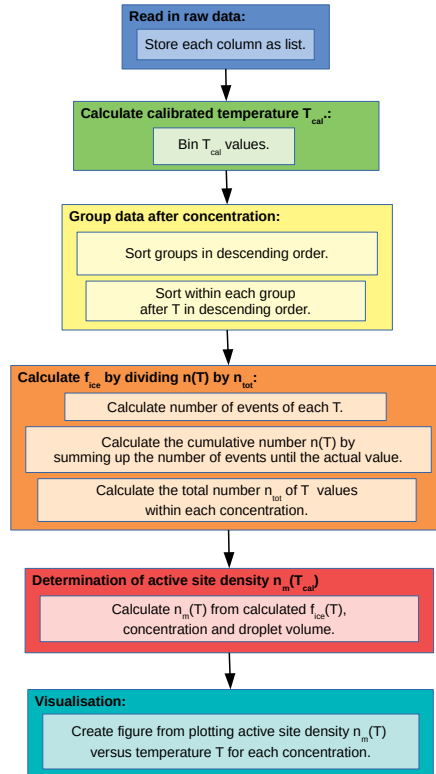

Figure 5.5: Schematic representation of analytical workflow as implemented in Python program.

First, the given raw data is read in and each column is saved as a list. In the second step, the calibrated temperature T_{cal} is calculated from the measured temperature T_{nuc} and the cooling rate r with formula 5.3. In the third step, the data is grouped by the concentration C_m into different bins and is sorted in

descending order. Within each bin the data is sorted by the temperature T in descending order. In the fourth step, for each bin a new table is generated. The bin is grouped by the temperature T and a second row is introduced showing the number of occurrences of each different binned temperature. A third row is used to summarize the occurrences including the current temperature. It shows the number of droplets up to the current temperature. This value and the total number of all droplets in this bin are used to calculate the frozen fraction $f_{ice}(T)$ with the given formula 5.2. Then it is appended to the table. In the fifth step, the active site density per unit mass, $n_m(T)$ is calculated from $f_{ice}(T)$, the concentration C_m and the droplets volume V_{drop} with the formula 5.4 and is appended as fifth column to the new table. Thereafter, this table is saved as a **.CSV** file, a common data format used by researchers with many tools for file input-output operations. As a second result, the generated table is used to reproduce and plot the graph in Figure 5.6 which displays our graph and the graph from the original paper for comparison. With the given raw data the results from the original experiment could be successfully reproduced using Python, an open source programming language.

5.3.4 Summary of Reproducibility Experiment

We reproduced the results from the analyses from the original paper as shown in the visualization graph Figure 5.6 by plotting $n_m(T)$, the cumulative number of IN per dry mass of Snomax® as a function of calibrated temperature. Origin software is a proprietary analysis toolbox with no equivalent libre software alternative. Hence, the original OPJ data file format can only read data into the Origin software system. The system allows data to be exported into tab-separated files with delimited columns. Due to the complexity and time associated with learning to use a new system like Origin, we opted to use Python to code the formulae and run the data files to be analyzed. In addition, Python is open source and is supported by many platforms.

Two particularly strong increases in $n_m(T)$ are observed, one at about $-3.5\,°\text{C}$ $(269.6K) \pm 0.5K$ and one at $-8.5\,°\text{C}(264.6K) \pm 0.5K$, indicating the presence of two distinct classes of ice nucleators with different activation temperatures.

The two plateaus at temperatures just below each increase of $n_m(T)$ in Figure 5.6 arise when no INs are active at these temperatures in the investigated suspensions. The $n_m(T)$ values of the plateaus differ by about 3 orders of magnitude, from which it can be inferred that the two classes of Snomax® ice nulceations occur at a number ratio of about 1 to 1000 in the samples. The active site densities per cell $n_n(T)$, shown in Figure 5.6 on the right axis, were calculated using the specific particle number of cells in Snomax®.
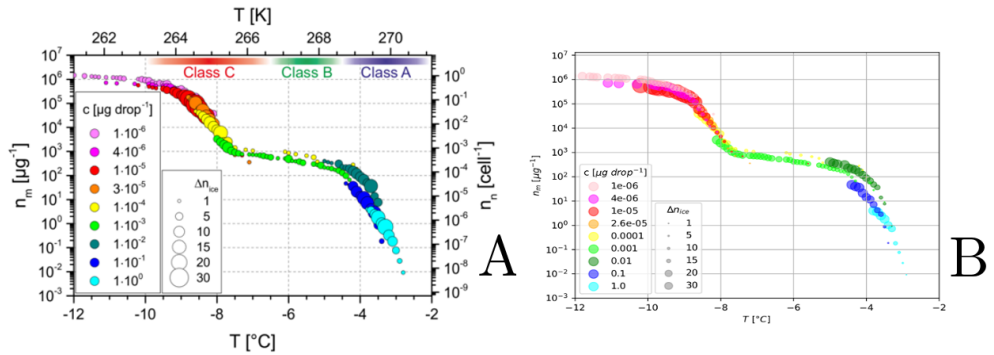
Figure 5.6: Experimentally determined active site density per unit mass of Snomax® $nm(T)$ versus temperature. A: Original version of diagram as published by Budke and Koop [1]; B: diagram resulting from reproducing the computational workflows of Budke and Koop as described in this paper. Symbol colors indicate data from droplets with different Snomax® concentrations; symbol size indicates the number of nucleating droplets per temperature interval (0.1°C). The temperature range for different classes of IN is also indicated by the colored bars.

5.4 Conclusion

In this work we could successfully reproduce the main results of the paper by Budke and Koop [1], reproducing the original analytical workflow using Origin-Pro by using free and open software, in this case a Python program implemented as part of the Conquaire project. Here, we thus have a case of limited reproducibility as the direct reproduction would have required obtaining a commercial license for OriginPro and re-creating the GUI interactions used in the original work. Instead, we have opted for a re-implementation of the original analysis in Python. We have thus not directly reproduced the original workflow, but developed a workflow that can be regarded as functionally equivalent. As we did not reproduce the original workflow exactly, we have a case of limited analytical reproducibility as defined in chapter 1 of this book. The data has been uploaded to the DFG FOR1525 project website (https://www.ice-nuclei.de/), where it is available upon request. Moreover, the data has been verified by an intercomparison paper by Wex et al. [15]. As a result of Conquaire, both the data and the script are available in a Git repository for further re-use and verification. While there is not yet a DOI for the dataset, the dataset and script are referenceable via a GIT repository, even down to a particular version.

Acknowledgments

References

[1] C. Budke and T. Koop. BINARY: an optical freezing array for assessing temperature and time dependence of heterogeneous ice nucleation. *Atmospheric Measurement Techniques*, 8(2):689–703, 2015.

[2] Hans R Pruppacher and James D Klett. *Microphysics of Clouds and Precipitation.* Kluwer Academic Publishers, New York, 2 edition, 1997.

[3] Will Cantrell and Andrew Heymsfield. Production of ice in tropospheric clouds: A review. *Bulletin of the American Meteorological Society*, 86(6):795–807, 2005.

[4] Dennis Lamb and Johannes Verlinde. *Physics and Chemistry of Clouds.* Cambridge University Press, Cambridge, 2011.

[5] T Koop. Homogeneous Ice Nucleation in Water and Aqueous Solutions. *Z. Phys. Chem*, 218:1231–1258, 2004.

[6] P J DeMott, A J Prenni, X Liu, S M Kreidenweis, M D Petters, C H Twohy, M S Richardson, T Eidhammer, and D C Rogers. Predicting global atmospheric ice nuclei distributions and their impacts on climate. *Proc. Natl. Acad. Sci. U. S. A.*, 107(25):11217–11222, 2010.

[7] B J Murray, D O'Sullivan, J D Atkinson, and M E Webb. Ice nucleation by particles immersed in supercooled cloud droplets. *Chem. Soc. Rev.*, 41(19):6519–6554, 2012.

[8] C Hoose and O Möhler. Heterogeneous ice nucleation on atmospheric aerosols: a review of results from laboratory experiments. *Atmos. Chem. Phys.*, 12(20):9817–9854, 2012.

[9] B G Pummer, H Bauer, J Bernardi, S Bleicher, and H Grothe. Suspendable macromolecules are responsible for ice nucleation activity of birch and conifer pollen. *Atmos. Chem. Phys.*, 12(5):2541–2550, 2012.

[10] Gabor Vali and E J Stansbury. Time-dependent Characterstics of the Heterogeneous Nucleation of Ice. *Canadian Journal of Physics*, 44(3):477–502, mar 1966.

[11] Gabor Vali. Supercooling of Water and Nucleation of Ice (Drop Freezer). *Am. J. Phys.*, 39(10):1125, 1971.

[12] L R Maki, E L Galyan, M M Chang-Chien, and D R Caldwell. Ice nucleation induced by pseudomonas syringae. *Appl. Microbiol.*, 28(3):456–459, sep 1974.

[13] G Vali, M Christensen, R W Fresh, E L Galyan, L R Maki, and R C Schnell. Biogenic Ice Nuclei. Part II: Bacterial Sources. *J. Atmos. Sci.*, 33(8):1565–1570, aug 1976.

[14] O Möhler, D G Georgakopoulos, C E Morris, S Benz, V Ebert, S Hunsmann, H Saathoff, M Schnaiter, and R Wagner. Heterogeneous ice nucleation activity of bacteria: new laboratory experiments at simulated cloud conditions. *Biogeosciences*, 5(5):1425–1435, 2008.

[15] Heike Wex, S Augustin-Bauditz, Yvonne Boose, Carsten Budke, Joachim Curtius, Karoline Diehl, Axel Dreyer, Fabian Frank, Susan Hartmann, Naruki Hiranuma, Evelyn Jantsch, Zamin a. Kanji, Alexei Kiselev, Thomas Koop, Ottmar Möhler, Dennis Niedermeier, Björn Nillius, Michael Rösch, Diana Rose, C Schmidt, Isabelle Steinke, and Frank Stratmann. Intercomparing different devices for the investigation of ice nucleating particles using Snomax as test substance. *Atmospheric Chemistry and Physics*, 15(3):1463–1485, feb 2015.

[16] M A Turner, F Arellano, and L M Kozloff. Three separate classes of bacterial ice nucleation structures. *J. Bacteriol.*, 172(5):2521–2526, 1990.