Reply to Referee #1

We thank you Referee#1 for her/his constructive comments on our manuscript. Please find below our replies in blue.

This manuscript provides a novel and necessary technique to improve the finding and targeting of tephra layers (and other particle types) in ice core samples. The authors did a rigorous analysis by imaging synthetic and natural particles and used machine learning to identify the different particle groups. The size of the particles that they can image is greater than most other grain size analysis techniques being used today in ice core research. Marine particles (diatoms and sponge spicules) are a little underrepresented but can sometimes be larger than the tubbing diameter (80 microns). Complex rock fragments (i.e. lava flows) are also not really identified in this study. Many particulate layers in Antarctica are wind-blow rock fragments and would have a distinctive shape when compared to mineral fragments and glass shards. This along with adding more types of pollen to their image dataset could be an area for future improvement.

We agree with the Referee - the training datasets can be certainly extended as a future improvement, with additional classes. In this work, among volcanic material, we have targeted only glass shards. We point out that the 80 µm flow cell is not necessarily a hard constraint. A bigger flow cell can be installed, but the trade-off is that typically in the standard FlowCam configuration the magnification would be lower (e.g. 10X). This would unlock the possibility to detect bigger particles but the resulting images would be smaller.

There are a number of comments in the attached pdf that need some clarification or elaboration. Many of my comments deal with the FlowCam setup. It would be really helpful if there was a picture or diagram of the FlowCam setup, even if in the appendix. I was confused by the orientation of the tubbing and the gravitational settling of larger particles. A cross-section diagram of the tube would help explain the imaging volume (41.8%) and the problem with large and blurry particles.

The imaged volume % is solely explained by the combination of pump rate, camera shutter rate and flow cell/optics geometry. The flow cell is placed in a vertical orientation, with the sample flowing in the same direction as gravity. The imaged volume % (41.8%) is automatically calculated by the FlowCam software based on the settings above. We have now added a picture in Appendix A to clarify the experimental setup.

Being able to image particles assess their grain size and give them a particle type (e.g. tephra, dust, etc.) before a tephra specialist gets the samples is extremely helpful and will improve the number of tephra found in ice cores and will decrease the time needed to find said tephra. However, the authors do not discuss how to physically capture the particles after FlowCam analysis. Capturing these particles so that they can be analyzed by SEM or EMPA is the most important part of this type of work. It is great to know the particle type and grain-size distribution but this method falls short if geochemistry on the particles is not obtained. It would be great if the authors would elaborate on capturing these particles. Their goal is to help both the CFA and the tephra communities. The CFA community doesn't like to run particles through their MS and the tephra community wants those particles. This method can be extremely helpful in spotting these interesting intervals.

We have now added our recommendations on how to collect the instrument outflow in Appendix A.

Overall this is an excellent paper that addresses a need in the ice core community. With some minor corrections and a few elaborations, this manuscript is ready to publish. I hope to see this type of FlowCam analysis being used in more labs and on more cores.

Thanks - we hope this work will be beneficial to various communities.

Please also note the supplement to this comment:
https://tc.copernicus.org/preprints/tc-2022-148/tc-2022-148-RC1-supplement.pdf

Please find our replies to your detailed comments in the pdf attached to this file, with the exception of the following one which is reported here given its importance:

L443-444: Limit of detection can be two things. Is this the lowest limit of detection or the maximum limit of detection? What is the maximum number of particles you can image? This would be important for a thick tephra deposit. Or what are the fewest particles you can see. This is important for cryptotephra where 10's of particles may be present.

We refer to LOD as the lowest concentration of an analyte in a sample that can be detected [1]. We calculated it for dust since we could perform a controlled experiment for this particle type by changing the dust standard concentrations.
For other particles such controlled conditions could not allow a LOD assessment - however, the following back-on-the-envelope calculation offers some insights for tephra:

If considering all concentration and dilution steps, we calculate that we concentrated the analyzed GRIP tephra samples by 20-50 times. With the level of tephra concentration of these samples, reported in the 1-100 shards/mL range from optical microscopy assessments (Table 4), our model detects tephra shards in the number of few tens to few hundreds, with roughly 0.3 mL of sample analyzed (Table 5). To get the same statistics without any concentration step, the amount of sample to be analyzed should be increased to 6-15 mL, clearly not compatible with continuous/online flow analysis requirements. Of course, such volumes/statistics can change according to the original concentration of tephra shards in the samples. At the moment the 42% image volume is the biggest challenge for the detection of low concentration particles. To our knowledge the newer FlowCam models feature a higher volume capacity (with the same 20X magnification).

On the other hand, we don’t see major obstacles for the system to detect ‘too many’ particles, other than the flow cell becoming clogged. Such a ‘clogging threshold’, however, has not been established within this work.

[1] IUPAC. Compendium of Chemical Terminology, 2nd ed. (the "Gold Book"). Compiled by A. D. McNaught and A. Wilkinson. Blackwell Scientific Publications, Oxford (1997).

https://doi.org/10.5194/tc-2022-148
Preprint. Discussion started: 26 August 2022

The Cryosphere Discussions

Detection of ice core particles via deep neural networks

Niccolò Maffezzoli[1,2], Eliza Cook[3], Willem G.M. van der Bilt[4,5], Eivind N. Støren[4], Daniela Festi[6], Florian Muthreich[5,7], Alistair W. R. Seddon[5,7], François Burgay[8], Giovanni Baccolo[9,10], Amalie R. F. Mygind[11], Troels Petersen[11], Andrea Spolaor[2], Sebastiano Vascon[1], Marcello Pelillo[1], Patrizia Ferretti[1], Rafael S. dos Reis[12], Jefferson C. Simões[12,13], Yuval Ronen[4], Barbara Delmonte[9], Marco Viccaro[14], Jørgen Peder Steffensen[3], Dorthe Dahl-Jensen[3], Kerim H. Nisancioglu[4,5], and Carlo Barbante[1,2]

[1]Ca' Foscari University of Venice, Department of Environmental Sciences, Informatics and Statistics, Via Torino 155, 30172 Venice, Italy
[2]Institute of Polar Sciences, ISP-CNR, Via Torino 155, 30172 Venice, Italy
[3]Physics of Ice, Climate and Earth, Niels Bohr Institute, University of Copenhagen, 2200, Copenhagen, Denmark
[4]Department of Earth Science, University of Bergen, Allégaten 41, 5020 Bergen, Norway
[5]Bjerknes Centre for Climate Research, Jahnebakken 5, 5020 Bergen, Norway
[6]Institute for Interdisciplinary Mountain Research, Austrian Academy of Sciences, Innrain 25, 6020 Innsbruck, Austria
[7]Department of Biological Sciences, University of Bergen, Thormøhlensgate 53A, 5006 Bergen, Norway
[8]Paul Scherrer Institut, Laboratory of Environmental Chemistry (LUC), Forschungsstrasse 111, 5232 Villigen PSI, Switzerland
[9]Department of Environmental Sciences, University of Milano-Bicocca, P.zza della Scienza 1, 20126 Milan, Italy
[10]Milano-Bicocca Section, Istituto Nazionale di Fisica Nucleare INFN, Milan, Italy
[11]Niels Bohr Institute, University of Copenhagen, 2100, Copenhagen, Denmark
[12]Centro Polar e Climático, Universidade Federal do Rio Grande do Sul, Porto Alegre, 91501-970, Brazil
[13]Climate Change Institute, University of Maine, Orono, ME 04469, USA
[14]Università degli Studi di Catania, Dipartimento di Scienze Biologiche, Geologiche e Ambientali, 95129 Catania, Italy

Correspondence: Niccolò Maffezzoli (niccolo.maffezzoli@unive.it)

Abstract. Insoluble particles in ice cores record signatures of past climate parameters like vegetation, volcanic activity or aridity. Their analytical detection depends on intensive bench microscopy investigation and requires dedicated sample preparation steps. Both are laborious, require in-depth knowledge and often restrict sampling strategies. To help overcome these limitations, we present a framework based on Flow Imaging Microscopy coupled to a deep neural network for autonomous image classification of ice core particles. We train the network to classify 7 commonly found classes: mineral dust, felsic and basaltic volcanic ash (tephra), three species of pollen (*Corylus avellana*, *Quercus robur*, *Quercus suber*) and contamination particles that may be introduced onto the ice core surface during core handling operations. The trained network achieves 96.8% classification accuracy at test time. We present the system's potentials and limitations with respect to the detection of mineral dust, pollen grains and tephra shards, using both controlled materials and real ice core samples. The methodology requires little sample material, is non destructive, fully reproducible and does not require any sample preparation step. The presented framework can bolster research in the field, by cutting down processing time, supporting human-operated microscopy and further unlocking the paleoclimate potential of ice core records by providing the opportunity to identify an array of ice core particles. Suggestions for an improved system to be deployed within a continuous flow analysis workflow are also presented.

https://doi.org/10.5194/tc-2022-148
Preprint. Discussion started: 26 August 2022

The Cryosphere Discussions
Open Access
EGU

1 Introduction

Ice cores provide some of the most valuable continuous records of the Earth's past climate. While the oldest Antarctic and Greenland cores date back respectively 800,000 and 125,000 years ago and register variability of climate parameters at hemispheric scales (North Greenland Ice Core Project members, 2004; EPICA community members, 2004), ice stored in glaciers and small ice caps located at lower latitudes typically contain fingerprints of local to regional climate changes on centennial to millennial timescales (Schwikowski, 2004). The analytical detection of impurities contained in the ice matrix allows to produce records of past climate at various spatial and time scales. Alongside gas bubbles and soluble chemical compounds, the ice matrix stores insoluble particulate matter, hereafter referred to as 'particles'. Among the types of particles is mineral dust, volcanic ash, pollen grains and other biological matter, as well as fossils sourced from the ocean or lakes such as diatoms and foraminifera. Each particle type carries its own climate significance and its concentration depends on factors such as the source strength and emission mechanisms, the relative distance between core site and source region, as well as parameters controlling atmospheric transport and deposition.

By far the most abundant particle type in ice, are mineral dust particles, that are sourced from continental surfaces and transported and dry or wet deposited onto ice sheets and glaciers (Legrand and Mayewski, 1997). The detection of dust is fundamental to investigate the extent of arid areas in the past, the paleo-atmospheric circulation and to assess the role of mineral dust aerosol in Quaternary climate changes (Petit et al., 1999; Lambert et al., 2008). Thanks to its preservation, dust records can be used to synchronize deep ice cores in the absence of other proxies, thus supporting ice core dating (e.g. Bohleber et al., 2018; Eichler et al., 2000; Dome Fuji Ice Core Project Members, 2017). Dust measurements are routinely carried out in continuous while melting ice cores in Continuous Flow Analysis setups (CFA, Bigler et al., 2011) using optical systems such as the laser-based Klotz Abakus sensor. As the abundance of dust particles is orders of magnitude higher than other insoluble particles, Abakus measurements are commonly associated with dust, despite the instrument being actually insensitive to the type of particle entering the detector. Additionally, Abakus values require an accurate calibration with an independent technique, typically the Coulter Counter (CC), an electrical-based analyzer that operates in discrete mode and can't be run on CFA setups (Petit et al., 1981). The mismatch and calibration between the Abakus and the CC impurity detection is an active research topic within the ice core community (Simonsen et al., 2018). The higher accuracy of the CC comes at the expense of its discrete mode of use; moreover it is also particle insensitive.

Volcanic ash deposits are also found in ice cores and can contain volcanic minerals, rock fragments as well as volcanic glass shards. Referred to as 'cryptotephras', these invisible and low concentrations layers of glass shards can form stratigraphically distinct deposits in ice cores, as well as in marine and terrestrial sediments (e.g. Lowe and Hunt, 2001; Turney et al., 1997). The identification of volcanic glass (hereafter referred to as tephra) provides direct evidence of past volcanic activity (Abbott and Davies, 2012; Sigl et al., 2015) and provides a crucial tool to date and synchronize paleorecords (ice, marine and lake) and therefore to establish absolute and synchronized chronologies (Lowe, 2011). The analytical detection of tephra layers in ice cores is typically carried out in a two-stage routine. First, potential volcanic layers are identified by spikes of electrical conductivity or sulfate concentrations during CFA analyses (e.g. Wolff et al., 1995). Afterwards, manual discrete sub-sampling

https://doi.org/10.5194/tc-2022-148
Preprint. Discussion started: 26 August 2022

of such selected horizons is carried out from a dedicated ice core section and the ice samples are individually processed and manually inspected using optical bench microscopy (e.g. Cook et al., 2018A). If the presence of tephra is confirmed, the grains are individually counted. Not all tephra layers however correspond to acidity or sulfate peaks and vice versa, given the different emission, transport and deposition of gaseous species and particulate volcanic material (Legrand and Mayewski, 1997; Davies et al., 2010). This makes the identification of tephra extremely time consuming and in some cases serendipitous. While attempts have been made to automate particle detection (e.g. Van der Bilt et al., 2021, in sediment records), the methodology for investigating tephra in ice cores typically requires a huge time commitment by tephra experts that leads to the identification of few major eruptions (Svensson et al., 2008; Lin et al., 2022).

Pollen analyses from snow and ice records provide information on past vegetation and atmospheric circulation changes (Bourgeois, 2000). Over relatively short timescales, pollen records with springtime maxima associated with vegetation blooms can also be used as a dating method (Nakazawa et al., 2004; Festi et al., 2021). Like tephra, pollen analyses need several and laborious preprocessing steps in which discrete ice samples are cut, melted, and pre-concentrated (e.g. Festi et al., 2015, and references therein). Finally, the presence and the number of pollen grains is manually evaluated by palynologists via optical bench microscopy. In summary, extraction and detection of climate-relevant ice core particles is extremely laborious.

In this work we investigate the extent to which autonomous and simultaneous detection and classification of ice core particles can be achieved with deep Neural Networks. Over the last ten years Neural Networks and in particular Convolutional Neural Networks (CNN) have become the state-of-the-art methods in digital image classification tasks. Since the proposed architecture of Krizhevsky et al. (2012), the field experienced rapid growth that spawned major breakthrough and optimization of a number of aspects including increasing model depth (Simonyan and Zisserman, 2014), understanding the dynamics of internal layers (Zeiler and Fergus, 2014) and facilitating the gradient flow (He et al., 2016). In the ImageNet classification challenge (Deng et al., 2009), CNN-based architectures have surpassed human accuracy (He et al., 2015). In our setup, to generate the ice particle imagery, we rely on a flow imaging microscopy instrument (the FlowCam), able to produce images of particles captured within a liquid stream continuously pumped through the instrument. We develop a mixed Convolutional and Fully Connected neural network to classify the imagery into 6 classes of particles: mineral dust, tephra (basaltic and felsic), and three pollen grains potentially present in alpine ice records: *Corylus avellana*, *Quercus robur* and *Quercus suber*. An additional 7-th class of Contamination/Blurry particles is included as a control channel for the model to be able to identify those particles that do not provide climate information.

2 Methods

2.1 FlowCam settings and optimization

The FlowCam instrument (Yokogawa Fluid Imaging Technologies, VS-I-B-model) located at the Earth Surface Sediment Laboratory (EARTHLAB, University of Bergen, Norway) is used to capture images of particles in ultrapure water or ice meltwater samples. The FlowCam is a benchtop flow imaging cytometer equipped with a visible range optical camera. The liquid sample is injected into the system by manual pipetting and it is drawn by a syringe pump to a quartz flow cell. Alternatively, a connec-

https://doi.org/10.5194/tc-2022-148
Preprint. Discussion started: 26 August 2022

tion tubing can allow sampling from discrete sample vials or from a continuous flow system. The flow cell used in our setup (depth=80 µm; width=570 µm) allows the flow of particles of up to 80 µm in diameter in the maximum dimension. A 1.0 mL volume syringe pump is set to operate at a flow rate of 0.02 mL/min. While passing through the flow cell the sample is imaged by a camera equipped with a 20x magnification objective. The camera flash duration is set to 65 μs and is operated at the maximum 22 frames per seconds. With the aforementioned settings the imaged sample volume, i.e. the percentage of volume imaged by the camera, is 41.8 %. This parameter is determined by the combination of camera frame rate, pump speed and flow cell geometry. The system optics determines a calibration factor of 0.2752 µm/pixel in the resulting monochrome 1280x960 pixels 8-bit TIFF images.

The mechanics of particle image creation is performed by the native FlowCam software (VisualSpreadsheet v3.4). All image frames captured during analyses are compared to a calibration image acquired prior to the analysis (Fig. A1). In every image, the pixels are considered 'signal' (i.e. set to 1) if their intensities are higher or lower than their intensities in the calibration image by a threshold value. If the pixel intensity differences do not exceed the threshold, they are considered 'background' and set to 0. Once the signal-background binary image is created, the particles are obtained by segmenting out the pixels flagged as signal (Fig. A1). The threshold value, set to 18, and the camera focus are optimized by acquiring images of spherical polystyrene 25 µm beads and by minimizing the standard deviation of the resulting size distribution (Fig. S1). For each acquired particle image, the FlowCam software calculates a number of numerical features, hereafter also referred to as metadata, mostly reflecting the particles' geometrical properties and are calculated by classic computer vision algorithms. In this work we use n=34 metadata (Appendix B).

2.2 Training dataset

The classification model is based on a supervised learning approach. The training dataset consists of images and related metadata for 7 classes of particles: mineral dust, tephra (basaltic and felsic), three pollen species: *Corylus avellana*, *Quercus robur* and *Quercus suber* and an additional class that consists of contamination particles that are found on the external surface of ice cores (Table 1, Figs. C1 to C7). Each item of the training dataset consists of a particle image and the corresponding array of 34 numerical metadata. The training dataset of each class (except for the contamination class) is created by preparing and measuring samples that contain only one type of particle so that each acquisition yields a purely 1-class batch. The samples are created by preparing solutions in ultrapure water and multiple acquisitions are repeated until several thousand images are collected. Every image of the training dataset is visually inspected and validated by the human eye.

- The training dataset of dust particles is created by measuring water solutions of FD066 (Linsinger et al., 2019, Table 1, Fig. C1), an aluminum oxide powder containing particles with a mean size distribution of 2.5 µm and rarely exceeding 6 µm (Table 2). Such a dust training set is therefore suited to mimic Antarctic and Greenland ice core dust, which is typically centered on 2–4 µm (e.g. Delmonte et al., 2002; Ruth et al., 2003).

- Two tephra classes, felsic and basaltic, are included in the training dataset, primarily because of their detectable color differences that result from a different geochemistry. Felsic (silica-rich) tephras are typically lighter in color, while

https://doi.org/10.5194/tc-2022-148
Preprint. Discussion started: 26 August 2022

The Cryosphere Discussions
Open Access
EGU

Table 1. [illegible]ng dataset.

Class	Sample type	Sample origin	Approximate size range	# training items
1. Dust	Conundrum powder (Al_2O_3)	Standard Reference Material (ERM-FD066)	< 10 µm	8000
2. Felsic tephra	Campanian Ignimbrite	Southeast Romania (43-44° N, 23-24° E)	(8, 80) µm	7125
3. Basaltic tephra	Grímsvötn (Iceland)	Kirkjubaejarklaustur, Iceland (63.78° N, 18.09° W)	(8, 80) µm	6271
4. Pollen *C. avellana*	pollen	Austria (47° 16' 14.31" N, 11° 22' 29.22" W)	(10, 40) µm	47223
5. Pollen *Q. robur*	pollen	Portugal (41.476-41.155° N, 8.701-8.563° W)	(10, 40) µm	35276
6. Pollen *Q. suber*	pollen	Portugal (41.155-41.151° N, 8.565-8.660° W)	(10, 40) µm	31745
7. Contamination/Blurry	Outer core ice samples	GRIP ice core external layer	(5, 80) µm	11439

basaltic ash is darker. The felsic tephra training dataset consists of Campanian Ignimbrite volcanic ash from the 39.3±0.1 ka BP Phlegrean Fields eruption (Fedele et al., 2003, Table 1, Fig. C2). The phonolitic-trachytic (∼60 wt.% SiO_2) ash was sampled ∼1,000 km from its source (Veres et al., 2013). Our basaltic tephra consists of volcanic ash from the Icelandic Grímsvötn 2011 eruption (Table 1, Fig. C3). Samples ash were collected on May 22, 2011 in the town of Kirkjubæjarklaustur, about 70 km southwest of the Grímsvötn caldera. After collection, samples have been dried and stored in plastic beakers. Ashes of both types were dry sieved at 63 µm to limit the maximum dimension and fit the flow cell's max. 80 µm size constraint (min. ∼8 µm). This range (8-80 µm) is consistent with the size that is typically considered during cryptotephra manual counting by bench microscopy (Gow and Meese, 2007; Narcisi et al., 2012; Abbott and Davies, 2012; Plunkett et al., 2020). It is important to note that, for both tephra classes, only those tephra images that could be clearly validated by an experienced tephra analyst were included in the training dataset. This resulted in discarding a very large fraction of [illegible]y imagery. This decision was adopted to drive the model to yield clearer tephra predictions and reduce ambiguous predictions (i.e. for tephra, purity is prioritized over efficiency).

- Three pollen species are included in the training dataset: *C. avellana*, *Q. robur* and *Q. suber* (Table 1, Figs. C4, C5, C6). *C. avellana* branches were collected near Innsbruck (Austria) in February 2019 from multiple trees within a radius of 500 m. The inflorescence was matured in the lab and the samples prepared by mixing together pollen from different trees. Both *Quercus* species were collected in Portugal and treated similarly. Given the diameter of the pollen grains, (*Corylus*: 25 µm; *Quercus*: 30 µm) and the optics not allowing to capture the whole cell width, in a number of images the pollen grain appears only visible for a fraction of its entirety. We decided to keep fractional pollen images above 10 µm to increase the sensitivity of the model to correctly classify pollen even when grains are only partially visible.

- The seventh class (Contamination/Blurry) consists of two types of particles. The first includes 'contamination' particles from the GRIP ice core external surface (Table 1, Fig. C7). Such a layer typically contains particles from the core drilling, cutting and handling operations such as paper wrap, glove clothing fibers and graphite from the pencil used to mark the core sections. The second type of particles added to this class includes relatively large and poor quality images, i.e. out of focus. The particles collected for this class are obtained from GRIP sample measurements followed by offline manual validation and labeling. The 7th class serves the purpose of a controlled channel to reduce false positives in all

https://doi.org/10.5194/tc-2022-148
Preprint. Discussion started: 26 August 2022

other classes that bring a climate significance. While blurry images is an intrinsic limitation of this methodology, the Contamination/Blurry class allows to flag the particles that carry climate significance from those that do not.

2.3 Model

2.3.1 Hybrid deep neural network

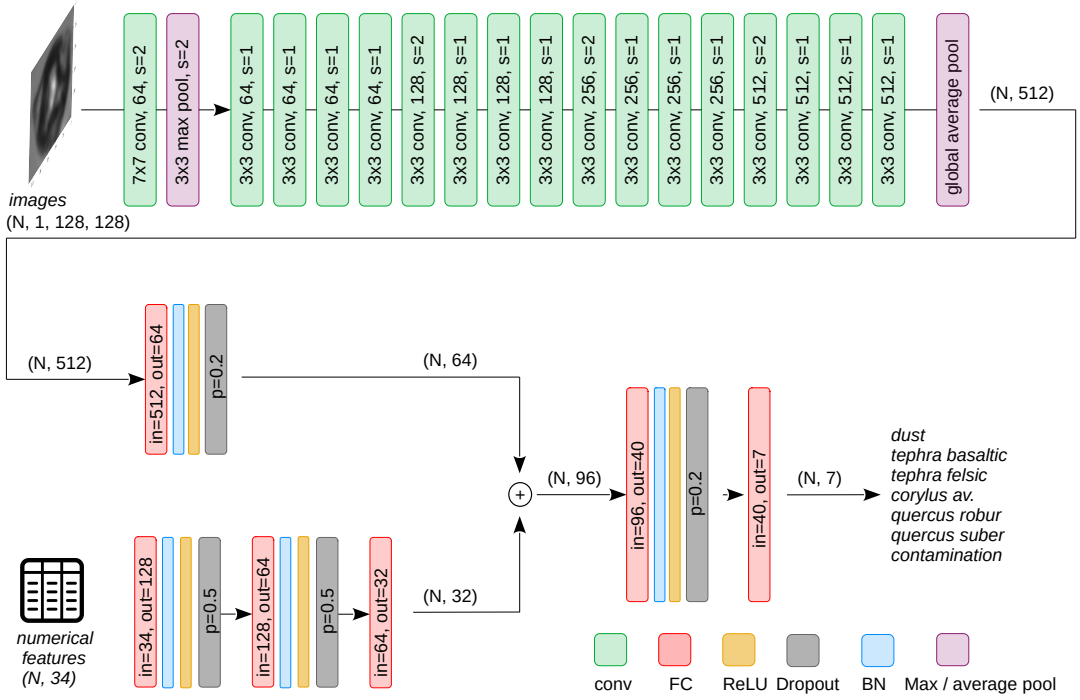

Figure 1. Model architecture. The top branch of the network is a resnet-18 CNN (He et al., 2016). BN and ReLU layers are omitted for clarity, as well as skip connections. The bottom branch operates on the numerical features and consists of 3-layer Multilayer Perceptron. The separate outputs of the two branches are concatenated into a final classification branch. Indicated in brackets are the input and output shapes of some layers along the network.

The developed model is a hybrid network that supports mixed data inputs (Fig. 1). It is composed of two branches, a Convolutional Neural Network (CNN) and a Multi Layer Perceptron (MLP), fed respectively by particle images and the corresponding 34-dimensional numerical feature vectors (metadata). The CNN consists of a resnet-18 architecture (He et al., 2016). This network is composed of multiple convolution layers that progressively increase the number of filters while decreasing the feature map size. Batch Normalization (BN) layers are placed right after each convolution layer and before ReLU (Rectified Linear Unit) activations. The network ends with an average pooling layer and a final FC (Fully Connected) layer that compresses the image into a 64-d embedding. This vector is concatenated to the output of the MLP, formed by two series of

https://doi.org/10.5194/tc-2022-148
Preprint. Discussion started: 26 August 2022

FC-BN-ReLU-Dropout layers followed by a final FC layer that produces a 32-d representation. Following the concatenation of the two network branches, a first FC-BN-ReLU stack is placed before the final FC layer that precedes a sigmoid activation.

2.3.2 Data preprocessing and augmentation

All images are reshaped by linear interpolation to 128x128 pixels. The downside of reshaping compared to zero-padding (i.e. increasing the image size by adding zeros to the borders) is that warping effects are introduced in images with large height to width differences, and the fact that the size information is lost. However, zero-padding to the largest image size would largely increase the computational complexity. We also argue that the size information is retained by the model in the metadata branch, that includes multiple features related to the geometry and the size of the particles. A per-image normalization to zero mean and unit variance is used to preprocess the images. Data augmentation during training consists of random rotations (p=0.5), as either horizontal, vertical or both horizontal and vertical flips. All metadata are also normalized by scaling to zero mean and unit variance.

2.3.3 Model training, validation and test

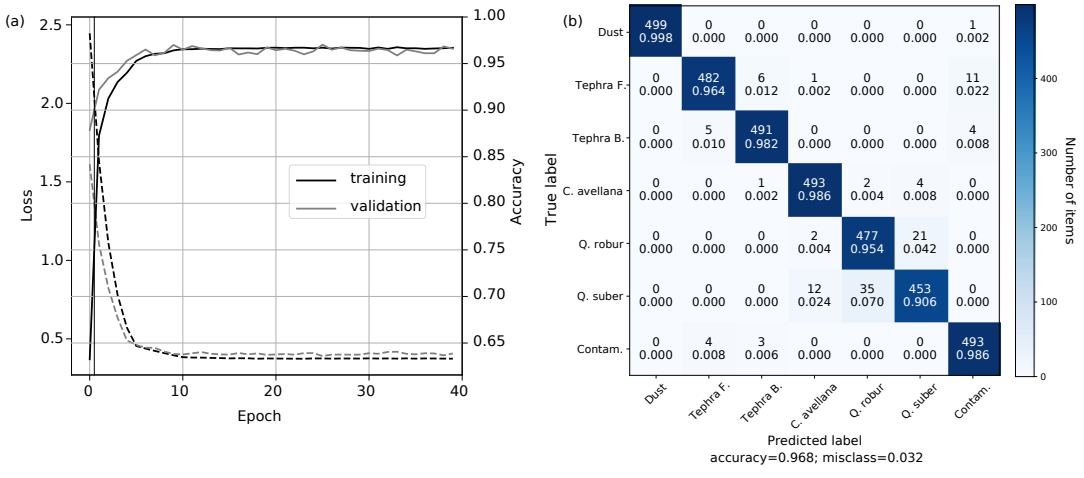

Figure 2. (a) Model loss (dashed) and accuracy (solid) evaluated during training (black) and validation (gray). (b) Confusion matrix of the best model evaluated on the test dataset. The accuracy across all classes is 96.8%. Most misclassifications occur within the two *Quercus* classes.

The data are split into three separate datasets: training, validation and test. Both the validation and test datasets consist of a random 500 item/class subset, for a total of 3500 items. A transfer learning approach is adopted for the convolutional branch of the network as the CNN pretrained on the Imagenet dataset is found to train faster. The whole network is trained on mini batches of 512 items using a binary cross entropy loss. The training dataset size of each class is indicated in Table 1. Since the

https://doi.org/10.5194/tc-2022-148
Preprint. Discussion started: 26 August 2022

training dataset is unbalanced, a weighted loss is implemented by enforcing a different weight w for each class c (Eq. 1):

$$w_c = \frac{\max_c size(c)}{size(c)}, c \in classes. \quad (1)$$

Underfitting and overfitting is checked after every epoch (a cycle of training the network) by monitoring the loss and the accuracy on the validation dataset. Adaptive AdamW is used as optimizer (Loshchilov and Hutter, 2017), with a learning rate of 10^{-4}, betas=(0.9, 0.999), a weight decay of 0.01 and a dedicated scheduler that imposes a learning rate decay of 0.1 every 5 epochs. The best hyperparameters (dropout probabilities, number and dimensionality of FC layers) are found by random search by maximizing the accuracy on the validation dataset. The final best model is evaluated on the test dataset.

The model converges to an average 96.8% accuracy across all classes in 15 epochs (Fig. 2). Dust and *C. avellana* images are classified with very high accuracy. Slightly lower accuracy is found among the two tephra classes, with on average 1% particles classified as the wrong tephra class and some 1-2% misclassified as 'Contamination'. The *Quercus* species are identified with an accuracy of ∼90-95%, the remaining fraction being misclassified mostly as the wrong *Quercus*. No misclassification is found between the three pollen and all other classes.

3 Results and discussion

The following discussion is divided into three sections. In Section 3.1 we investigate the FlowCam ability to correctly detect dust, with particular focus on the reconstruction of the size distribution and the mass concentrations, followed by the comparison with the Coulter Counter on a number of alpine ice core samples. In section 3.2 we discuss pollen and the representativeness of their training datasets. In Section 3.3 the model is deployed on Greenland ice core samples containing volcanic ash horizons.

3.1 Dust

3.1.1 Standard Reference Material: size reconstruction, LODs and mass concentrations

The certified reference material ERM FD066 Aluminum Oxide powder is used to evaluate the performance of the system as a dust detector. We measure a solution containing FD066 powder, run the model on the acquired images and metadata and evaluate the [illegible] distribution. All particles are classified as 'dust' by the model. The number-weighted ABD distribution percentiles are consistent within 1σ to the certified values (Table 2, Fig. S2).

The mass concentration of a sample can be calculated by summing the particles ABD-based volumes, dividing by the sample imaged volume and multiplying by the density. The Aluminium Oxide density is 3.96 $\mathrm{g/cm^3}$. An alternative metric to the ABD is the Equivalent Spherical Diameters (ESD, Appendix B), a measure of an object size based on its orientation. However, we find that ESD volume quantifications are not consistent with the expected volume distribution of FD066 samples (not shown), in agreement with previous studies that found that ESD leads to overestimate volumes of particles with extended parts and appendages (Karnan et al., 2017; Kydd et al., 2018). Our results show that the ABD metric can be therefore considered appropriate for reconstructing the size of dust particles with a distribution similar to that of the FD066 material, as well as of spheres (Fig. S1).

https://doi.org/10.5194/tc-2022-148
Preprint. Discussion started: 26 August 2022

The Cryosphere Discussions
Open Access
EGU

Table 2. Comparison between the FD066 ABD size distributions certified by Scanning Electron Microscopy (Linsinger et al., 2019) and calculated using the FlowCam (this study).

FD066 ABD size distribution percentiles	ABD (µm, ±1σ) certified by SEM (Linsinger et al., 2019)	ABD (µm) reconstructed by FlowCam (this study)
$x_{5,0}$	1.07 ± 0.23	1.14
$x_{10,0}$	1.28 ± 0.24	1.06
$x_{25,0}$	1.71 ± 0.28	2.07
$x_{50,0}$	2.4 ± 0.4	2.5
$x_{75,0}$	3.3 ± 0.4	3.3
$x_{90,0}$	4.4 ± 0.4	4.5
$x_{95,0}$	5.1 ± 0.4	5.1

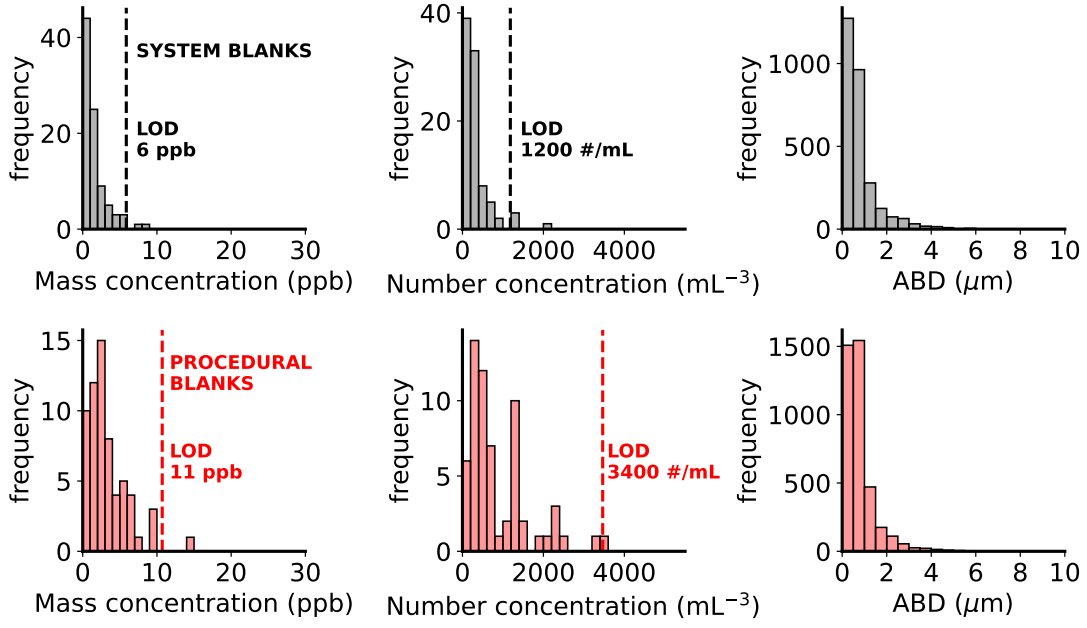

Figure 3. Analysis of n=91 system blanks (black, top row) and n=63 procedural blanks (red, bottom row). Median mass concentrations (left distributions) of 1.0±1.6 ppb(1σ) and 2.4±2.8 ppb(1σ) result in a 3σ LOD of 6 ppb for system blanks and 11 ppb for procedural blanks respectively. The middle plots show the number concentration distributions and respective LODs. The right plots indicate the size distributions of blank particles in system (top, N=2864) and procedural (bottom, N=3945) blanks, rarely exceeding 3 μm. All particles are classified as dust by the model. The mass concentrations are calculated assuming a density of 2.5 g/cm^3.

Given the low dust concentrations in ice core records, it is crucial to investigate the blank levels of the analytical system as well as the impurity content of the used glassware. We define *system blank* as the instrumental response to ultrapure water (UPW, 18.2MΩ/cm) directly injected into the system. The system blanks can be thought of as the blank level of a CFA system,

https://doi.org/10.5194/tc-2022-148
Preprint. Discussion started: 26 August 2022

in which no discrete vials are used and the sample stream directly feeds the FlowCam from the melt head (although a tubing connection would be needed). We define *procedural blanks* as the instrumental response to UPW stored in sterile ultra-clear polypropylene VWR centrifuge tubes (model 21008-216) prewashed 5 times with UPW. No acids are used. A set of n=91 system blanks and n=63 procedural blanks are investigated (Fig. 3). The model classifies the totality of particles in both the system and procedural blanks as dust, with diameters rarely exceeding 3 µm (Fig. 3). The limits of detection (LOD) are calculated as the median plus 3 standard deviations. The mass concentration and number concentration LODs of the system blanks are respectively 6 ppb and [illegible] #/mL. The mass concentration and number concentration LODs of the procedural blanks are respectively 11 ppb and 3400 #/mL. In comparison, the LOD of the [illegible]s reported as 2 ppb (Ruth et al., 2008). The lowest dust concentrations in ice records are found in Antarctica during interglacial periods, with levels of about 10 ppb over the plateau (Lambert et al., 2008) and a few ppb towards high accumulation coastal sites (Vallelonga et al., 2004). The FlowCam LODs thus allow quantification of dust in all sites globally except for coastal Antarctic interglacial records. It is likely possible to further lower the instrument LODs by operating the FlowCam inside a clean room.

We next evaluate the quantification of dust mass concentrations, by comparing the FlowCam to the Coulter Counter. Discrete dust samples for FlowCam analyses are prepared by diluting a known mass of FD066 material (weighted on a 10^{-6} g accuracy scale) in ultrapure water, and subsequent dilutions using VWR centrifuge tubes. The concentration of the final samples ranged from 44 ppb to 14 ppm (Fig. 4). All acquired particle images are classified as dust by the model. The ABD-based volumes are converted to mass using the FD066 density, 3.96 $\mathrm{g/cm^3}$. Similarly prepared samples are measured by Coulter Counter at the University of Milano-Bicocca, by adopting the same analytical steps as described in (Baccolo et al., 2021). The LOD of the CC, calculated as 3 standard deviations above the average of n=7 UPW samples, is 10 ppb. For both the FlowCam and CC experiments the blank levels are subtracted to the concentration values of the samples. The FlowCam mass concentrations are consistent with the expected values and a good linear agreement is found across the investigated range (Fig. 4), spanning from low Antarctic to high mid-latitude glacier dust levels. The residual distribution (mean of 0.7%, 1σ=14%) suggests an accurate combination of camera focus and particle volume estimation and no systematic uncertainty in the volume quantification. The precision is evaluated by multiple repetitions of the same samples (typically 3 to 5, shown as the error bars on the points and in the RSD distribution) and averages 19% (1σ=11%). The CC measurements also show good linearity (q=0.001±0.002, m=1.05±0.05). This experiment shows that both instruments yield accurate size and volume reconstructions for the irregularly shaped FD066 particles.

3.1.2 Ice core dust mass concentrations

The FlowCam and the CC mass concentration reconstructions are compared by analyzing n=24 ice samples from the Quelccaya ice cap (Peru, Reis et al., 2022). Since the CC is particle insensitive for this comparison the classifier coupled to the FlowCam is switched off. Two aliquots for each sample are measured by CC (Milan, Italy) and by FlowCam (Bergen, Norway). The CC is operated with a 2-60 µm capillary to accommodate the large particles common in alpine records. Each sample quantification results from the average of three measurements. In the calculations of the mass concentrations a density of 2.5 $\mathrm{g/cm^3}$ is assumed.

https://doi.org/10.5194/tc-2022-148
Preprint. Discussion started: 26 August 2022

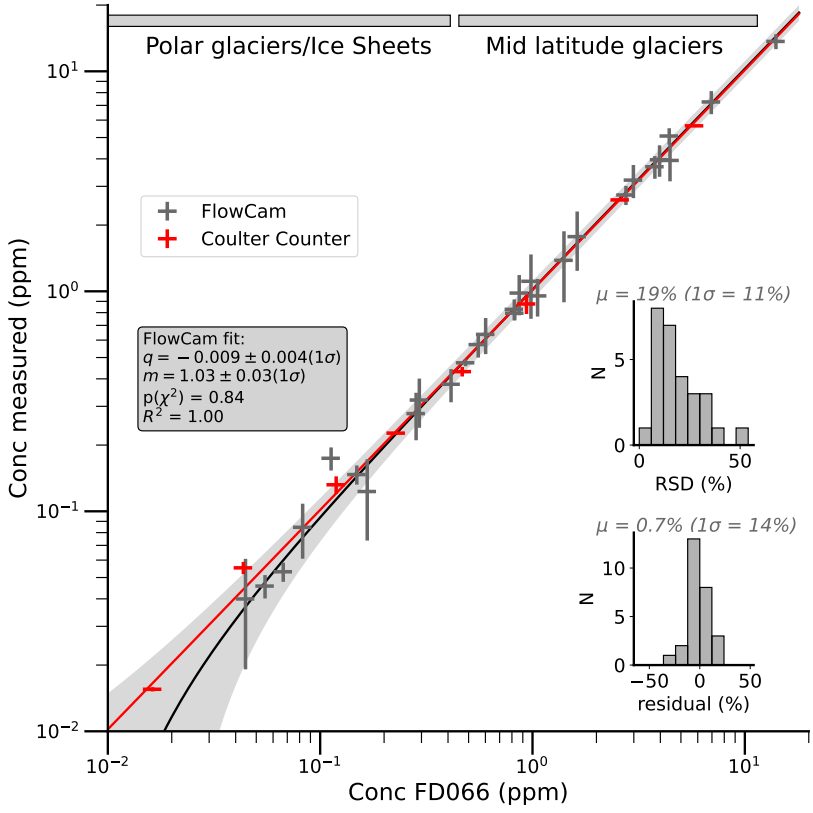

Figure 4. Comparison between nominal and measured concentrations of FD066 dust samples using the FlowCam and the Coulter Counter. An orthogonal distance regression on the FlowCam data (black line with 3σ confidence interval in gray) shows good linearity over three orders of magnitude. The red line refers to the linear fit on the CC data. The y-error bars reflect 1 standard deviation of multiple repetitions of the same sample. All x-errors are estimated as 10% of the FD066 prepared concentrations and account for the uncertainties in the dilutions and plastic adsorption effects. Both insets refer to the FlowCam measurements: the top inset shows the relative standard deviation (RSD) distribution; the bottom inset shows the distribution of the residuals, defined as the difference between the expected and measured concentrations. The top bars indicate the approximate ranges of dust concentration in polar and mid-latitude records.

The samples exhibit a very large size distribution, with particle sizes extending to 60 µm and a volume weighted size distribution centered between 10 and 20 µm. The dust concentrations (we here refer to the total insoluble content as dust for simplicity) range from 1 to 15 ppm and a median of 2 ppm. The comparison of the two instruments across the batch of 24 samples reveals that the FlowCam mass concentrations are systematically lower than those measured by CC. In particular, the cumulative distribution function of the mass concentration, $CDF(x) = \int_0^x conc(z)dz$, reveals that fewer big particles are captured by the FlowCam compared to the CC, explaining the lower values of the FlowCam (Fig. S3).

We argue that the causes are twofold. First, the FlowCam images a very low amount of volume (the highest efficiency achievable in our setup, 41.8%, is reached by minimizing the pump rate, 0.02 mL/min, and maximizing the camera shutter

https://doi.org/10.5194/tc-2022-148
Preprint. Discussion started: 26 August 2022

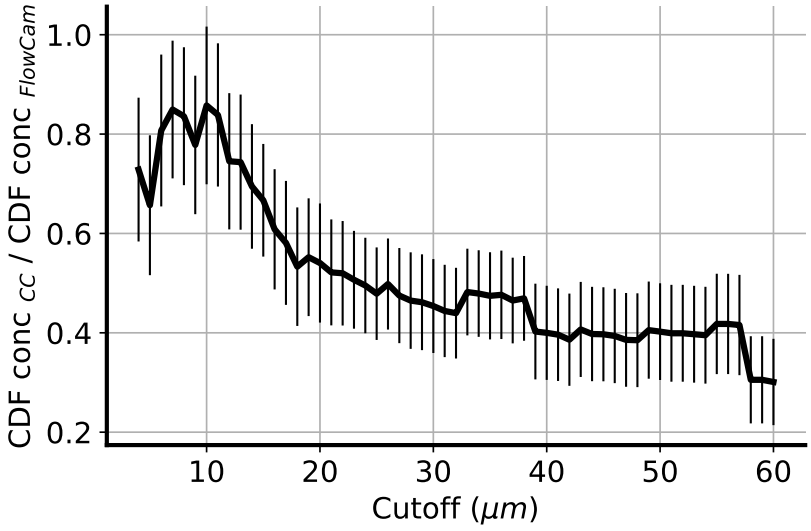

Figure 5. Mass concentration CDF ratios between the CC and the FlowCam as a function of a size cutoff. The best agreement is found at a cutoff value of 10 µm. If larger particles are included in the quantification of the concentration, the FlowCam concentrations are consistently lower than the CC.

to 22 FPS). For example, for a 3 min analysis only 0.025 mL of sample is imaged, compared to 0.5 mL on the CC. The low statistics has a notable effect in the estimation of the mass concentration, since big particles are rare and provide a large contribution to the volume. The underestimation of large ($\geq$ 50 µm) particle concentrations using the FlowCam compared to manual microscopy has been previously reported (Kydd et al., 2018). A possible second cause for the FlowCam undershoot is the discrete mode of analysis. During manual sample injection into the FlowCam, big particles quickly flow through the instrument by gravitational settling, while smaller particles remain more easily suspended in the solution and are continuously injected throughout the analysis time. We argue that the fast gravitational separation of particles of different sizes leads to underestimated concentrations especially for high analysis time. It may be possible to reduce the gravitational separation by using a continuous-flow injection system with the tubing placed horizontally, as in an ice core CFA melting system, or by operating in discrete mode using sample agitation equipment.

We investigate to which extent (in terms of particle size) the concentrations of the CC and the FlowCam can be compared. For each method we calculate the concentration of all 24 samples by only considering particles smaller than a certain value, progressively increased from 3 to 60 µm. The comparison is quantified by evaluating the slope of an orthogonal distance linear regression between the concentration CDFs with respect to the size cutoff (Fig. 5). The best agreement is found if only particles up to ca. 10 µm are accounted for (m=0.86$\pm$0.16). For bigger particle sizes, the FlowCam underestimates the CC concentrations by up to $\sim$3. This analysis is consistent with the good match previously found using the small-sized FD066 material (Fig. 4).

https://doi.org/10.5194/tc-2022-148
Preprint. Discussion started: 26 August 2022

From the two FlowCam-CC comparisons carried out on the small-sized FD066 dust and on alpine samples we conclude that, in the experimental conditions of our setup (discrete mode of operation, 80 µm flow cell and 20x magnification), the FlowCam is to be used for evaluating mass concentrations of particles up to only ∼10 µm. For samples containing larger particles the mass (and number), the FlowCam concentrations will be underestimated by up to ∼3. To improve the accuracy, the statistics at high particle sizes can be increased by i) increasing the efficiency of the instrument using larger volume cells, ii) increasing the measurement time alongside sample agitation equipment.

3.2 Pollen

Given the similarity of pollen grains, we investigate the representativeness of the three training datasets used to classify these types of particles. The analysis is carried out by training the model using slightly different training datasets and by evaluating the classification accuracy on controlled samples of specific types. 5 different *C. avellana* types were made available for this experiment, labeled A, B, C, D, E. They reflect samples collected from different trees within the same sampling region. We build three training datasets: type A, type B, and the last one prepared by mixing all five types together (referred to as Mix). We then train the model 4 times separately using type A, type B, type Mix and all of them together (A+B+Mix), and each time we evaluate the pollen predictions of a pure type B dataset. The training datasets of all the other classes are kept fixed. In particular, the *Q. robur* and *Q. suber* datasets consist of two types for each one mixed together. After each training session, a validation stage on 500 images of each type is evaluated for performance and hyperparameter tuning. No substantial change in any hyperparameter is found to be affecting the accuracy on the validation set, which is consistently 0.97-0.99 for *C. avellana* and between 0.90-0.96 for the two *Quercus* species. The model trained with the Corylus A dataset yields only 48% correct Corylus predictions, when deployed on a Corylus B sample (N=5 replicates, Table 3). It appears that the Corylus A training dataset is not fully representative of the Corylus B sample. If the model is trained with a Corylus B, the percentage of Corylus classification in the Corylus B sample increases to 96%. If a Corylus Mix training set is used, the correct accuracies are 97%. If the model is trained with all datasets joined together (A+B+Mix), the correct Corylus predictions are 98%. The best result can therefore be achieved if the model is trained with the widest dataset in terms of particle variability.

A similar test is carried out for the *Q. robur* class. The model is trained separately using a *Q. robur A*, a *Q. robur* B and a joined *Q. robur* A+B dataset and used each time to classify a pure *Q. robur* B sample (N=3 replicate measurements). The model is trained by keeping a fixed *Corylus* A+B+Mix and *Q. suber* A+B datasets. The results show that only 2.3% of the images in the *Q. robur* B sample are correctly classified as *Q. robur* if the model is trained with the *Q. robur* A dataset (Table 3). The correct predictions are 91% if the *Q. robur* B training dataset is used instead. By training using a joint *Q. robur* A+B dataset, the percentage of correct robur predictions remains similar. Unlike the *Corylus* test, no *Q. robur* Mix is available.

The test on the *Q. suber* type is analogous to the *Q. robur* test. The model is trained three times separately on a *Q. suber* A, a *Q. suber* B and on a *Q. suber* A+B dataset and each time used to classify a pure *Q. suber* B sample (N=10 measurements). When the model is trained with the *Q. suber* A dataset, only 4% of the pollen in the *Q. suber* B sample are correctly classified as *Q. suber* (Table 3). The correct classifications rise to 90% if the model is trained with either the *Q. suber* B or with the *Q. suber* A+B dataset.

https://doi.org/10.5194/tc-2022-148
Preprint. Discussion started: 26 August 2022

The Cryosphere Discussions
Open Access
EGU

From these tests we conclude that the representativeness of the training dataset is crucial to achieve the highest pollen classification accuracy. For all three pollen types, the best classification is achieved with the largest training datasets. Under this condition the classification accuracy of *C. avellana*, *Q. robur* and *Q. suber* is respectively 98±1%, 91±1% and 90±3%, similarly to what was previously found (Fig. 2b). We argue that a further increase in accuracy and a more general model may be achieved by further increasing the training datasets in both variability and in size. As a sense of the model predicting power, it should be noted that expert palynologists cannot efficiently classify the *Q. robur* and *Q. suber* species by looking at the FlowCam images. The state-of-the-art classification accuracy between *Q. robur* and *Q. suber*, 98±2%, can be achieved by analyzing the different pollen chemical signatures by using Fourier-transform infrared spectroscopy (Muthreich et al., 2020). We also find that the absolute number of images classified as pollen varies by just, on average, 0.4% , suggesting that pollen detection (irrespective of the pollen class) is largely independent of the choice of the training dataset.

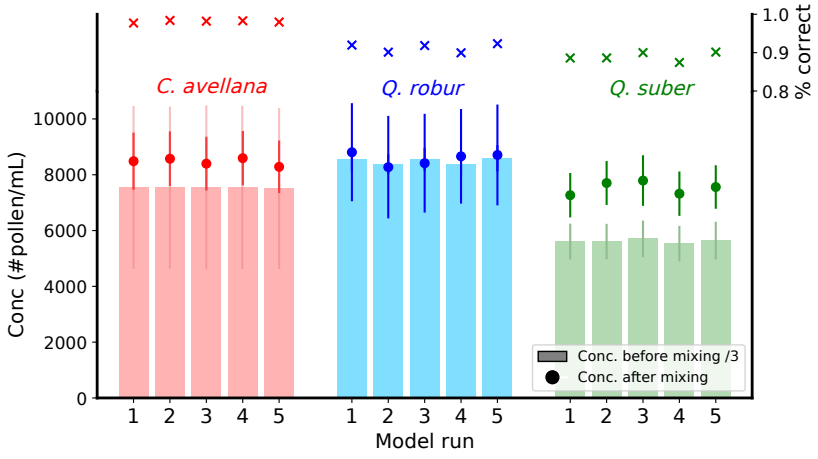

Figure 6. Quantification of pollen concentrations in single types samples and in a mixed sample. The model was deployed to classify and quantify pollen concentrations in three samples containing purely *C. Avellana* (red), *Q. robur* (blue) and *Q. suber* (green) pollen. The percentages of correct predictions among the three pollen classes are indicated in the top right axis as a function of 5 independent model runs. One third of the sample pollen concentrations (as averages of 5, 3 and 10 aliquots respectively for *C. Avellana*, *Q. robur* and *Q. suber*) are indicated as histogram bars along with 1σ error bars displayed with light colors. The model was also used to classify and quantify pollen in a 1:1:1 mix of the original samples (dots and solid colored error bars reflect the average values and 1σ of 10 aliquot).

We finally train a model 5 times using the largest datasets: *C. avellana* A+B+Mix, *Q. robur* A+B and *Q. suber* A+B. The model is then used to classify particles in three samples containing only one type of pollen. The analysis is performed on 5 aliquots of the *C. avellana* sample, 3 aliquots of the *Q. robur* sample and 10 aliquots of the *Q. suber* sample. Afterwards, the three samples are mixed together in a 1:1:1 volume ratio and the model is used to classify particles in 10 aliquots of the mixed sample. As previously found the model behaves well in classifying the *C. avellana* pollen, with 98% of all particles classified correctly (Fig. 6, red). The classification accuracy for the *Q. robur* and *Q. suber* averages 90% accuracy (Fig. 6, blue and green). The concentrations of the pollen species before mixing (bars) and the concentration of the species as classified by the

https://doi.org/10.5194/tc-2022-148
Preprint. Discussion started: 26 August 2022

model after the mixing agrees (dots) is reasonably consistent for the *C. avellana* and *Q. robur* pollen, while some departure from the expected concentration is found for the *Q. suber* class. The results do not show significant differences with respect to the model runs, suggesting that the model converges to similar parameters. However, in all separate runs a significant spread is found between the aliquots, particularly with respect to *C. avellana* classification (1σ are indicated as the error bars in Fig. 6), which suggests that robust quantification of pollen concentrations should be achieved by multiple measurements. The *Q. suber* concentration mismatch is tentatively attributed to the cell being partially clogged that led to an underestimated concentration before mixing.

The pollen experiments suggest that the developed framework is promising for pollen autonomous classification under the condition that the most representative datasets are used for training. Additionally, the representativeness of fresh pollen as a training dataset for microfossil ice core pollen should be investigated. We also stress that, in case of low concentrations, a similar underestimation of the absolute number of pollen is to be expected, by a factor $\sim$2 (Fig. 5). Intensive analysis of alpine ice core records (where pollen is expected) is the next logical step.

3.3 Tephra

We now deploy the model to investigate the content of 12 samples from the Greenland Ice Core Project (GRIP) ice core (Table 4). In particular, 7 of these contain known tephra deposits, selected from the tephrochronology framework of Cook (2022), while the remaining 5 samples are known to be devoid of tephra grains (i.e. tephra grains were not observed by bench microscopy).

The 7 tephra deposits were re-sampled, by removing a strip of 55 cm of ice (referred to as a 'bag') using a band saw. Each bag strip of ice was then cut into three sections, at resolutions of 20 or 15 cm, using the same depth intervals as Cook (2022) to ensure the same deposits could be found, and thus producing replicate tephra-containing ice core samples. The 5 tephra-free samples were derived from ice adjoining each of the tephra layers, i.e. the remaining ice per bag. The deposits chosen for this experiment date back to the Bølling–Allerød/Greenland Interstadial 1 (GI-1) and glacial/Greenland Stadial 2 (GS-2) periods and comprise tephra of a similar geochemical composition as those selected for our training dataset: felsic (basaltic), mafic (rhyolitic), or a mix thereof. For each selected depth interval, two replicate samples are obtained: the first was analyzed for tephra by optical bench microscopy (Cook, 2022); the second one is analyzed by Flow Microscopy followed by our particle classification model. It is important to note that, although extracted from the same horizon, the samples dedicated to the two analyses are different and non-homogeneity can affect the lateral distribution of insoluble matter at the same depth interval (Cederstrøm et al., 2021). Additionally, we note that the samples contain contamination particles, as the outer surface of the core collects impurities from drilling and processing activities. The samples dedicated to tephra investigation have typically been extracted from these external sections, as the analyst is able to distinguish tephra from other types of matter using the bench microscopy.

https://doi.org/10.5194/tc-2022-148
Preprint. Discussion started: 26 August 2022

3.3.1 Optical microscopy for tephra analysis

The samples dedicated to tephra analysis by optical microscopy were prepared as in Cook et al. (2018B). Specifically, the samples were melted, centrifuged, evaporated and the remaining material was embedded in epoxy resin. Optical microscopy tephra counts range between 0 to 5000 shards per sample, corresponding to concentrations from 0 to 111 shards/mL (Table 4). The counting errors, estimated in Table 4, also incorporate the uncertainties related to the loss of material during the centrifuge and due to adhesion onto the used plastic tubes. It is worth noting that microscopy counts of tephra are typically only performed above a size threshold for which the human operator is confident to differentiate tephra grains from mineral dust: ∼8 µm. Replicate counting on the same samples would be needed to more rigorously quantify the manual counting errors.

3.3.2 Flow Imaging Microscopy and particle classification

The samples dedicated to FlowCam analyses, whose original volumes were between 28 and 56 mL (Table 5), were concentrated by centrifuge down to less than 0.5 mL, following the same sample processing adopted for optical microscopy for the sake of consistency, except for the embedding in epoxy resin. As an additional step, given the very high particle concentration that would obstruct the flow cell, the samples were diluted by adding ultrapure water, between 0.5-1.0 mL. The imaged volume of each sample was 0.2-0.3 mL. In total, up to hundreds of thousands of images were collected per sample, for a total of 3,085,063 images (Table 5). As expected, most particles (91-98% of the total content) are classified as dust by the model. The remaining fraction is almost fully explained by Contamination/Blurry particles (2-9%). Their presence derives from the nature of the analyzed samples, extracted from the core surface and thus loaded with external impurities. It is possible that the Contamination/Blurry predictions contain some particles of climate significance but we expect this number to be very small. A total of n=921 particles are classified as pollen (209 *C. avellana*, 375 *Q. robur* and 337 *Q. suber*, Table 5). By visually inspecting these particles it is clear that, due to their blurriness, only few of them can be confidently identified as pollen (or spores), but the large majority of these predictions remain dubious (Fig. S5). We note that the three species of pollen used to train the model do not fit with the spectrum of pollen species that may be found in Greenland. A better choice for polar records would be a training dataset of *Betula* pollen - ubiquitous in Arctic paleoclimate records. We also argue that very likely a high number of contamination particles are falsely predicted as pollen. The reason for such classification outcome by the model is the round shape of such particles, and their similar size to that of the three pollen species (Sect. 3.3.4).

A total of n=1671 particles are classified as tephra (949 basaltic and 722 felsic, Table 5). The tephra concentrations in the samples, irrespective of the two types, range from 3.3 to 18 #/mL (Table 5, col J). Although in the same order of magnitude, there are significant sample-to-sample differences compared to concentrations determined by manual counting (Table 4). It should be noted that the samples measured using the two techniques are different and some non-homogeneities with regard to tephra deposition can be expected (Pyne-O'Donnell, 2011). We also argue that, while the model accuracy does not depend on the tephra concentration, human-operated microscopy is probably more effective for higher concentrations. This could explain why the modeled concentrations are always above zero. We also note that the modeled values are expected to be underestimated

https://doi.org/10.5194/tc-2022-148
Preprint. Discussion started: 26 August 2022

by a factor of about 2-3 from the real concentrations (Sect. 3.1.2, Fig. 5), because of fluidics/loss of material as gravitational settling preferentially affects large particles.

3.3.3 Human assessment of modeled tephra predictions

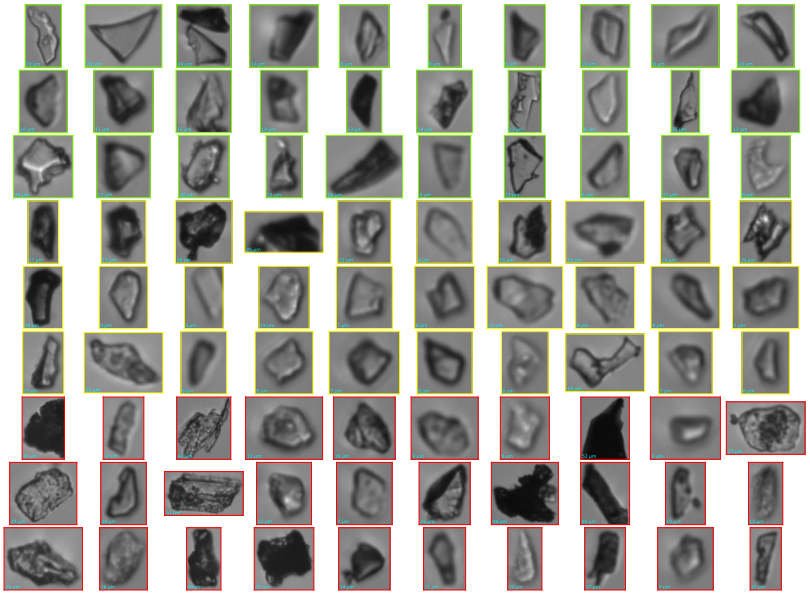

Figure 7. A random subset of the AI-predicted tephras in the 3136 0-20cm GRIP sample assessed by Human#1, color coded according to the given validation: yes (green), maybe (yellow), no (red). [illegible] diameters are shown in the bottom left corners.

To further explore the model predictions and investigate the mismatch, two tephra experts were asked to assess and classify, based on the FlowCam images, all (n=1671) modeled tephra predictions in the 12 GRIP samples (irrespective of whether they are predicted as felsic or basaltic) into 3 classes: 'yes', 'maybe' and 'no' (Table 5). According to Human#1 (Human#2), of all 1671 images 16% (2%) are positively validated as tephra, 37% (56%) are dubious and 47% (41%) are considered not tephra (Fig. S4). Of all the AI-predicted tephras, Human#1 therefore considers 53% of them are possible tephras ('yes'+'maybe'), while for Human#2 considers them to be 58%. It should be noted however that the agreement between the two operators is weak (Fig. S4): the quality of the FlowCam images often precludes a confident optical assessment of the particles (Fig. 7, Fig. S5). Interestingly, some tephra shards are positively validated even in those samples for which no tephra was previously found during optical microscopy. According to both analysts, the tephra modeled predictions include a number of minerals such as feldspar and quartz, and a few contamination particles. Minerals are routinely found during manual microscopy assessments but can be confidently recognized using cross-polarized light (Lowe, 2011), which allows the analyst to easily distinguish isotropic non-crystalline tephra from anisotropic minerals. In our current setup this key function is not available, but a circular polarizer should be implemented on the FlowCam for future studies and will be key to differentiate tephra from minerals.

https://doi.org/10.5194/tc-2022-148
Preprint. Discussion started: 26 August 2022

The source of minerals inside the FlowCam-measured samples can be twofold: they can derive from active dust sources proximal to the core site such as ice-free Iceland or Greenland (Simonsen et al., 2018), or be introduced artificially onto the core surface during the laboratory handling procedure, similar to the source of the contamination particles). At this stage, it is not possible to further speculate on the relative importance of these two sources of minerals, and additional measurements of replicate clean ice samples would be needed. With respect to the presence of minerals within the set of tephra predictions in the GRIP samples, the consulted experts point out that some images of minerals are also found within the two tephra training datasets. Hence, the tendency to classify minerals as tephra is to some extent embedded in the model. Measurements of clean ice are also needed to minimize the rate of tephra false positives from the contamination class (∼1%, Fig. 2). Given the large prior of contaminations in the GRIP samples (n=89329), 900 false positives (out of the n=1671 tephra predictions) could be misclassified as tephras. This further advocates the need of measuring clean samples in future studies.

3.3.4 Investigating the network dynamics

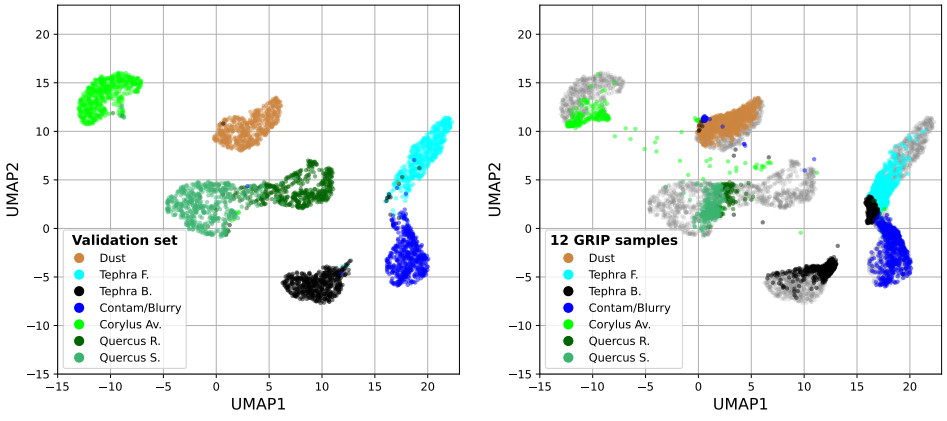

Figure 8. UMAP 2D visualization of the network 64d layer of the CNN branch. On the left umap is run on the validation dataset. On the right the learnt UMAP space is used to project all images of the n=12 GRIP samples. The items are color coded according to their predicted class. Gray items represent the validation items.

To better understand the network dynamics and how the images are classified into the different classes, we probe the output of the last FC layer of the convolutional branch of the architecture (Fig. 1). At this network depth, each original 128x128 image becomes compressed into a 64-dimensional vector representation. We inspect such a 64d space using UMAP, an unsupervised manifold learning and dimension reduction algorithm (McInnes et al., 2018). We first inject the trained network with a random dataset of 500 items/class from the validation dataset, for a total of 3500 items. We extract the 64d representations and let UMAP learn a 2D embedding space of the data (Fig. 8). In such representation the embedded data appear clustered according to their respective classes, with few items mispositioned in the wrong classes (basaltic, felsic tephra and contamination/blurry) and with some degree of overlap between the two *Quercus* classes that evidences the higher difficulty of the network to distinguish these types of pollen. Overall the high degree of separation between the training items is well reflected in the

https://doi.org/10.5194/tc-2022-148
Preprint. Discussion started: 26 August 2022

confusion matrix (Fig. 2). The parametric UMAP model generated using the training data is then applied to the combined dataset of n=12 GRIP samples comprising all 3,085,063 images. The images are injected into the network, and the 64d vectors are extracted and reprojected onto the learnt UMAP space (Fig. 8). Overall, the GRIP items are projected on top of the training clusters, with the exception of a secondary smaller cluster of Tephra B. found encompassed within the Contamination and Tephra F. clusters which evidences that some Tephra B. images incorporate some features that are common to all three classes. The *Quercus* predictions are located at the intersection of the two respective training clusters. Some *C. avellana* predictions are found scattered outside its training cluster, thus not fully representing the features of the training images. Figure S5 shows the same plot with the dots replaced by images. Such a representation also allows us to inspect a number of features. For example, different light conditions characterize images located in different areas within the dust cluster (both the validation and GRIP data). The light from the camera flash can be occasionally redirected to the camera shutter if the dust particle is oriented in such a way that the light becomes significantly backscattered. In such a condition the dust particles becomes white on a darker background. Different colors are also found within the training Tephra B. cluster, mostly consisting of dark particles and fewer brighter particles located at the margins on the cluster. The Tephra B. GRIP cluster contains a higher proportion of bright particles compared to its training counterpart. Bright tephra classifications are more frequently predicted as Tephra F., although a secondary cluster of bright Tephra B. images is found positioned at the interface between the Tephra F. and contamination clusters. The contamination cluster contains a number of particles that have been introduced during handling operations, such as long and rod looking particles likely from glove fabrics. Blurry images are also present in this class (as the model was trained to do so). They may or may not be legitimate ice core particles. Particles classified as pollen in the GRIP samples are blurrier than those in the training sets. However, they generally show round shapes and significant size $\geq$10 µm. These two features are consistent with the pollen training images, probably leading to such classification outcome. Similar to tephra, the investigation of pollen particles should be carried out on clean samples to avoid the presence of contamination particles being falsely classified as pollen.

4 Conclusions and perspectives

We developed a framework for the detection, autonomous classification and quantification of climate-relevant insoluble particles in ice core samples that can provide support and complement human-operated optical microscopy. Our approach is fully reproducible, non-destructive and does not require any sample preparation, thus saving time and material. It couples Flow Imaging Microscopy to a deep neural network for image classification. The network is trained on 7 classes of particles: mineral dust, volcanic ash or tephra (basaltic and felsic), three species of pollen grains (*C. avellana*, *Q. robur* and *Q. suber*) and a class consisting of Contamination/Blurry particles. The architecture, comprising a Convolutional and a Fully Connected network, achieves 96.8% accuracy on the test set. Training 40 epochs requires ~30 min on a GeForce RTX 3090. The model operates at ~300,000 img/s at test time and allows online deployment. Some key advantages, disadvantages and suggested upgrades to the system developed in this work are outlined in Table 6.

https://doi.org/10.5194/tc-2022-148
Preprint. Discussion started: 26 August 2022

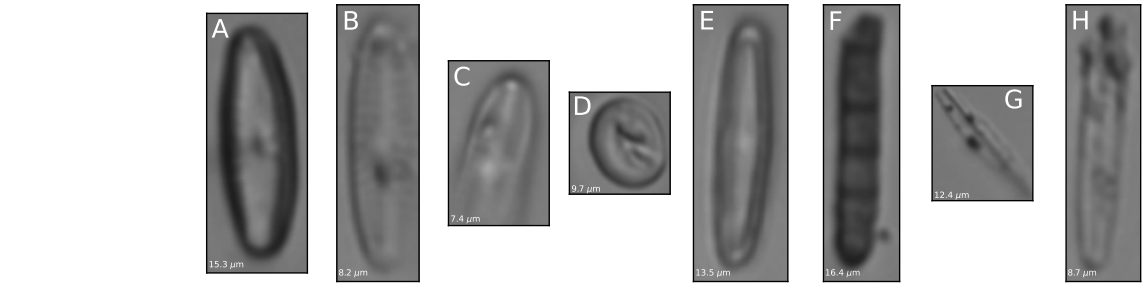

Figure 9. Diatoms identified in the Quelccaya ice core from the acquired FlowCam images. Particle D can be a Centrales diatom (possibly Cyclotella genus) or an algae. Particle F can possibly be a fungus. All other particles are Pennales diatoms. The diameters are indicated in the bottom left corners. The presence of diatoms in this ice record has been previously reported using SEM microscopy by Fritz et al. (2015). A promising future application will be to naturally extend the model by incorporating additional training classes, including diatoms. At this stage this has not been possible.

The system was investigated as a dust detector. The FlowCam can reconstruct the size distribution of Standard Reference Material fine-grained (<10 µm) dust particles within 1σ of the certified values. The mass concentrations can be replicated within 1% over a range from few ppb to 10 ppm, with an average precision of 19%. The Limit of Detection for dust ranges from 6 to 11 ppb. The comparison of mass concentrations with the Coulter Counter reveals a good agreement (ratio=0.86±0.16) only for particles smaller than ∼10 µm. The FlowCam exhibits a drop in efficiency in detecting larger particles that can lead to an underestimated mass concentration of up to a factor 3. This drawback affects all types of particles and should be carefully considered. In the presented setup, the FlowCam offers a valid alternative to the Coulter Counter and to the Abakus as a dust detector for polar ice cores, with the advantage of being sensitive to the particle type.

We tested the classification of freshly-collected pollen grains and found - perhaps unsurprisingly - that the representativeness of the training datasets is of exceptional importance. If the model is trained using the most general pollen datasets, *Corylus avellana* can be classified at ∼98% accuracy, while *Quercus robur* and *Quercus suber* can be classified at ∼90% accuracy.

We applied the model to 12 GI-1 and GS-2 Greenland ice samples, containing known tephra deposits, for a total of 3+ million images. Almost the entirety of the images is classified as either dust or Contamination/Blurry particles, the latter from the external core surface. 1671 particles are classified as tephra (either felsic or basaltic). Inspection of such images by two tephra experts suggests that only up to ∼50% are possible tephra, the remaining ∼50% consisting of either Contaminations or minerals such as quartz and feldspar. At this stage, our framework can support tephra analyses by providing first-order information on the occurrence of volcanic layers, but we could not quantitatively replicate the tephra concentrations obtained by optical microscopy.

Building on this work, we envision promising avenues for further research and upgrades in two main fields: data and hardware.

- The existing training datasets should be extended by including other relevant particles that may be found in ice core records (e.g. diatom frustules, Fig. 9, or *Betula* pollen). The noise baseline introduced by contamination/blurry particles

https://doi.org/10.5194/tc-2022-148
Preprint. Discussion started: 26 August 2022

should be better established by measuring clean samples. Meaningful integrations between the data that result from our method and from human-operated optical microscopy should be outlined.

- Improvements of the hardware should target both the quality of the imagery (by using the more resolved color camera featured by the FlowCam 8100 model) and the statistics (by installing a higher volume cell alongside a faster shutter rate camera). Importantly, a polarizer would be key to separate tephra from anisotropic minerals. An improved system should be ideally tested and deployed within a CFA workflow, targeting continuous particle records from ice cores.

Code and data availability. The training and GRIP datasets will be deposited on Zenodo. The code will be made publicly available at https://github.com/nmaffe/Icelearning-software.

https://doi.org/10.5194/tc-2022-148
Preprint. Discussion started: 26 August 2022

Table 3. Pollen experiment results. The accuracies are indicated as the average of N replicates (*C. avellana*: N=5, *Q. robur*: N=3, *Q. suber*: N=10). In brackets the standard deviation of the replicates.

Model training dataset	Model inference on a pure *C. avellana* B sample		
	N_{cor}/N_{pollen} (1σ)	N_{rob}/N_{pollen} (1σ)	N_{sub}/N_{pollen} (1σ)
C. avellana A (N=7824) *Q. robur* A+B (N=35276) *Q. suber* A+B (N=31745)	0.48 (0.03)	0.07 (0.01)	0.45 (0.03)
C. avellana B (N=13713) *Q. robur* A+B (N=35276) *Q. suber* A+B (N=31745)	0.961 (0.007)	0.005 (0.004)	0.034 (0.003)
C. avellana Mix (N=25186) *Q. robur* A+B (N=35276) *Q. suber* A+B (N=31745)	0.970 (0.007)	0.006 (0.001)	0.023 (0.006)
C. avellana A+B+Mix (N=47723) *Q. robur* A+B (N=35276) *Q. suber* A+B (N=31745)	**0.984 (0.004)**	**0.005 (0.002)**	**0.012 (0.003)**
Model training dataset	Model inference on a pure *Q. robur* B sample		
C. avellana A+B+Mix (N=47723) *Q. robur* A (N=10239) *Q. suber* A+B (N=31745)	0.11 (0.01)	0.023 (0.005)	0.87 (0.01)
C. avellana A+B+Mix (N=47723) *Q. robur* B (N=24537) *Q. suber* A+B (N=31745)	0.038 (0.007)	0.910 (0.01)	0.051 (0.003)
C. avellana A+B+Mix (N=47723) *Q. robur* A+B (N=35276) *Q. suber* A+B (N=31745)	**0.036 (0.007)**	**0.914 (0.009)**	**0.050 (0.005)**
Model training dataset	Model inference on a pure *Q. suber B* sample		
C. avellana A+B+Mix (N=47723) *Q. robur* A+B (N=35276) *Q. suber* A (N=10663)	0.47 (0.10)	0.49 (0.09)	0.038 (0.008)
C. avellana A+B+Mix (N=47723) *Q. robur* A+B (N=35276) *Q. suber* B (N=20582)	0.07 (0.03)	0.03 (0.01)	0.90 (0.03)
C. avellana A+B+Mix (N=47723) *Q. robur* A+B (N=35276) *Q. suber* A+B (N=31745)	**0.07 (0.03)**	**0.03 (0.01)**	**0.90 (0.03)**

https://doi.org/10.5194/tc-2022-148
Preprint. Discussion started: 26 August 2022

Table 4. GRIP sample details and tephra counts by manual optical microscopy. The sample ages are derived from the GICC05 chronology. * Sample that corresponds to the specified age. ** The uncertainties are estimated.

GRIP sample (age)	Sample depth interval (cm)	Depth (m)	Microscopy ice meltwater (mL)	Microscopy tephra counts (±1σ)**	Microscopy tephra concentrations (#/mL)
3046 (GI-1b, 13186 yr b2k)	0-20	1674.75 - 1674.95	33	0	0
	20-40*	1674.95 - 1675.15	36	1062±50	30±1
	40-55	1675.15 - 1675.3	33	16±1	0.48±0.03
3136 (GI-1e, 14191 yr b2k)	0-20*	1724.25 - 1724.45	45	5000±3000	111±67
	20-40	1724.45 - 1724.65	38	57±5	1.5±0.1
	40-55	1724.65 - 1724.8	28	0	0
3303 (GS-2.1a, 17238 yr b2k)	0-20*	1816.1 - 1816.3	34	20±1	0.59±0.03
	20-40	1816.3 - 1816.5	33	0	0
	40-55	1816.5 - 1816.65	28	365±20	13.0±0.7
3306 (GS-2.1a, 17326 yr b2k)	0-20	1817.75 - 1817.95	33	0	0
	20-40	1817.95 - 1818.15	34	0	0
	40-55*	1818.15 - 1818.3	23	500±100	22±4

https://doi.org/10.5194/tc-2022-148
Preprint. Discussion started: 26 August 2022

Table 5. GRIP sample modeled predictions obtained from the FlowCam measurements. Col F: number of tephra predictions. In parenthesis are indicated the number of tephra (irrespective of the two tephra classes) validated as Yes, Maybe or No by the Human1 by evaluating the FlowCam images. For example, the 3046 0-20 sample would contain 88 tephras, of which 10/88 are positively validated by the operator, 30/88 are uncertain and 46/88 are considered not tephra. Col J: tephra concentration calculated by considering the number of all AI-predicted tephras, e.g. 43+43 for the 3046 0-20 sample. Col K: tephra concentration calculated by considering the number of AI-predicted tephra, constrained to Human1 'Yes' + 'Maybe'. Col L: tephra concentration calculated by considering the number of AI-predicted tephra, constrained to Human1 'Yes'. Col M: same as column K, but constrained to Human2 counts. Col N: same as column L, but constrained to Human2 counts.

GRIP sample (Age)	A Depth interval (cm)	B Volume (mL)	C Imaged volume (mL)	D Acquired particles	E Dust	F Tephra felsic, basaltic (Human1 Yes, Maybe, No), (Human2 Yes, Maybe, No)	G Corylus	H Quercus robur, suber	I Cont.	J AI Tephra conc (#/mL)	K AI Tephra conclhuman1 Y+M (#/mL)	L AI Tephra conclhuman1 Y (#/mL)	M AI Tephra conclhuman2 Y+M (#/mL)	N AI Tephra conclhuman2 Y (#/mL)
3046	0-20	56	0.23	198961	188053	43, 43, (10, 30, 46), (9, 33, 44)	16	28, 33	10745	3.3	1.6	0.4	1.6	0.4
3046	20-40	56	0.33	257002	241826	114, 184, (83, 115, 100), (15, 152, 131)	31	50, 49	14748	15	9.7	4.1	8.2	0.7
3046	40-55	45	0.42	58217	54481	51, 32, (9, 25,49), (2, 51, 30)	13	30, 28	3582	4.0	1.6	0.4	2.5	0.1
3136	0-20	54	0.31	46571	42512	30, 149, (58, 69, 52), (9, 107, 63)	11	33, 20	3816	9.5	6.7	3.1	6.1	0.5
3136	20-40	52	0.33	86352	81280	39, 53, (12, 38, 42), (0, 59, 33)	14	12, 18	4936	5.2	2.8	0.7	3.3	0
3136	40-55	42	0.31	38055	35203	30, 29, (8, 21, 30), (1, 33, 25)	22	15, 22	2734	3.9	1.9	0.5	2.3	0.07
3303	0-20	33	0.30	170457	164457	52, 34, (28, 30, 28), (2, 59, 25)	11	21, 24	5858	7.9	5.3	2.6	5.6	0.2
3303	20-40	32	0.32	349773	338911	87, 89, (9, 56, 111), (0, 111, 65)	16	23, 32	10615	14	5.3	0.7	9.1	0
3303	40-55	28	0.33	334060	322070	41, 72, (17, 55, 41), (0, 72, 41)	4	8, 11	11854	12	7.7	1.8	7.7	0
3306	0-20	36	0.33	608856	588023	99, 118, (8, 72, 137), (0, 117, 100)	36	84, 55	20441	18	6.7	0.7	9.9	0
3306	20-40	37	0.31	437339	425309	67, 82, (5, 46, 98), (0, 91, 58)	26	57, 28	11770	12	4.1	0.4	7.4	0
3306	40-55	30	0.36	499420	489356	69, 64, (26, 59, 48), (1, 55, 77)	9	14, 17	9891	11	7.1	2.1	4.7	0.08

The Cryosphere Discussions
Open Access
EGU

https://doi.org/10.5194/tc-2022-148
Preprint. Discussion started: 26 August 2022


The Cryosphere Discussions
Open Access
EGU

Table 6. Advantages, disadvantages and suggested upgrades to the system presented in this work.

Particle class	Advantages	Disadvantages	Suggested upgrades
Polar dust (<10 µm)	CFA ready. Accurate mass concentration reconstruction.	Detection Limit close to Antarctic interglacial values.	Deployment in a clean room.
Alpine dust	Accurate mass concentration reconstruction for dust <10 µm.	Underestimation of the >10 µm fraction.	Higher volume cell.
Tephra (volcanic glass)	Can support human-operated bench microscopy. Autonomous, no sample preprocessing, CFA ready.	Limited to fraction >8 µm. Underestimation of particles >10 µm. Low statistics. Image quality.	Higher volume cell. Polarizer.
Pollen	Can support human-operated bench microscopy. Autonomous, no sample preprocessing, CFA ready.	Underestimation of particles >10 µm. Low statistics. Image quality.	Required training datasets tailored to the ice core site. Higher volume cell.
New particles (e.g. diatoms)	Autonomous, no sample preprocessing. Easy to implement by adding training datasets.	Underestimation of particles >10 µm. Low statistics.	Requires specific training datasets. Higher volume cell.

https://doi.org/10.5194/tc-2022-148
Preprint. Discussion started: 26 August 2022

The Cryosphere
Discussions
Open Access
EGU

Appendix A: Segmentation of particle images

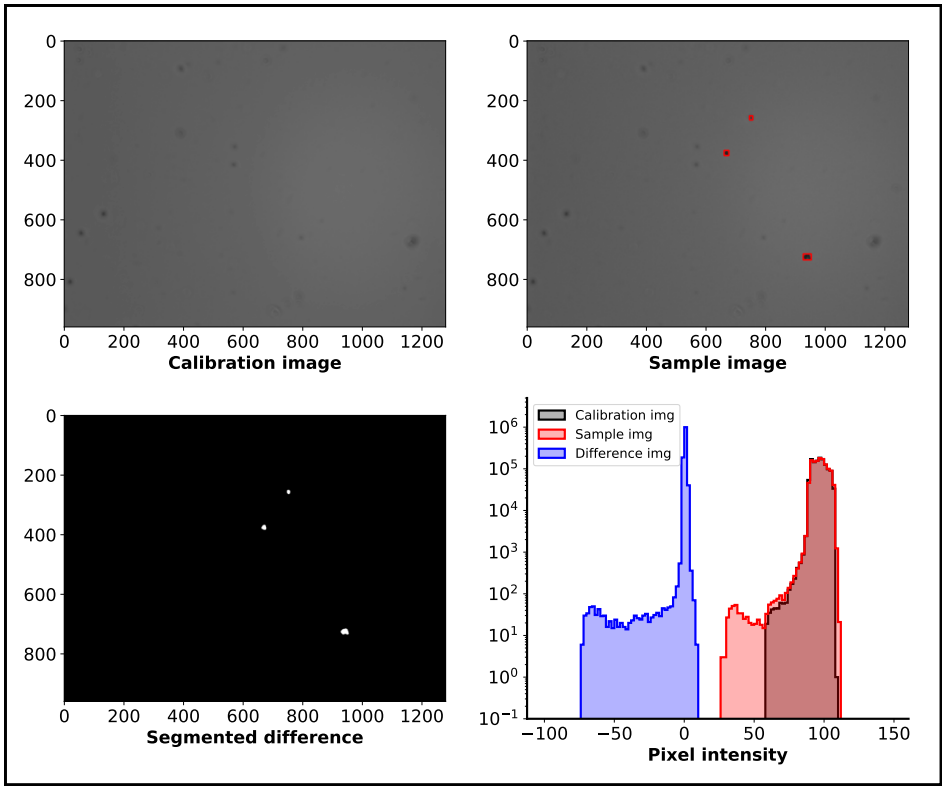

Figure A1. Segmentation of particle images. A calibration image is obtained prior to the analysis when no sample is pumped into the system (top left). During the analysis each frame (top right) is compared to the calibration image (bottom right) and a pixel-by-pixel difference is calculated (bottom right) and thresholded to yield the single particle images. The procedure is performed by the FlowCam software.

https://doi.org/10.5194/tc-2022-148
Preprint. Discussion started: 26 August 2022


The Cryosphere Discussions

Appendix B: Metadata

Feature	Explanation
1. Area (ABD)	Number of pixels in the thresholded (binary) greyscale image converted to a measure of area by use of the calibration factor. (real > 0)
2. Area (Filled)	The area represented by the particle edge and all the pixels inside the edge. (real >0) In the case of an opaque particle Area (Filled) = ABD. However, if parts of the particle are transparent, and therefore do not threshold as "particle", then Area (Filled) > ABD.
3. Aspect Ratio	The ratio of the lengths of the axes of the Legendre ellipse of inertia of the particle. The Legendre ellipse of inertia is an ellipse with its center at the particle's centroid, and with the same geometrical moments, up to second order, as the original particle area. A circle has the value 1.0 as does a square. Values near zero are for particles that are long and thin. (real [0, 1])
4. Biovolume (Cylinder)	Biovolume (Cylinder) = (π/4) * Geodesic Thickness2 * Geodesic Length
5. Biovolume (P. Spheroid)	Biovolume (Spheroid) = (π/6) * Legendre Minor2 * Legendre Major
6. Circle Fit	Deviation of the particle edge from a best-fit circle, normalized to the range [0,1] where a perfect fit has a value of 1. (real [0, 1]; 1 is the value for a perfect circle; values near zero are for particles that are not at all circular)
7. Circularity	A shape parameter computed from the perimeter and the (filled) area. A circle has a value of 1.0. Circularity is the inverse of Compactness. Formula: (4 x π x Area) / Perimeter^2. (real [0,1]) = Pixel Grid = Perimeter = Best-Fit Circle = Area (Filled)
8. Circularity (Hu)	An alternative measure of circularity that often provides a better indication of the circular shape of a particle than does Circularity, especially if the particle is very small or its edge has defects. A circle has a value of 1.0. (real [0, 1]). Ref: Žunić et al. (2010).
9. Compactness	A shape parameter derived from the perimeter and the (filled) area. The more convoluted the shape, the greater the value. A circle has a value of 1.0. Compactness is the inverse of Circularity. Formula: Perimeter^2 / (4π * Area). (real $\geq$ 1)
10. Convex Perimeter	An approximation of the perimeter of the convex hull of a particle. Derived from feret measurements.
11. Convexity	A shape parameter that is computed as the ratio of filled area to the area of the convex hull of the particle. This property is sometimes called Solidity. A circle has a value of 1.0. (real [0, 1]) (A simple way of thinking of the convex hull is to imagine taking a rubber band and stretching it around the filled area)
12. Diameter (ABD, Based Diameter)	The diameter based on a circle with an area that is equal to the ABD Area. (real > 0)

https://doi.org/10.5194/tc-2022-148
Preprint. Discussion started: 26 August 2022

13. Diameter (ESD, Equivalent Spherical Diameter)	The mean value of 36 Feret measurements. (real > 0)
14. Edge Gradient	Average intensity of the pixels making up the outside border of a particle after a Sobel Edge Detect convolution filter has been applied to the raw camera image. (real [0, 255])
15. Elongation	The inverse of Geodesic Aspect Ratio. (real $\geq$ 1; 1 is the value for a circle or square; larger values are for elongated particles)
16. Feret Angle Max	Angle of the largest Feret measurement. (real [-90, +90])
17. Feret Angle Min	Angle of the smallest Feret measurement. (real [-90, +90])
18. Fiber Curl	A shape parameter computed from Geodesic Length and Length. Also known as Curl Index. Formula: (Geodesic Length / Length) – 1. (real $\geq$ 0)
19. Fiber Straightness	A shape parameter computed from Geodesic Length and Length. Formula: Length / Geodesic Length. (real $\geq$ 0)
20. Geodesic Aspect Ratio	The ratio of Geodesic Thickness to Geodesic Length. Elongation is the inverse of this ratio. (real [0, 1])
21. Geodesic Length	Values obtained by modeling the particle as a rectangle and computing length and thickness by solving the equations: Area = Geodesic Length x Geodesic Thickness Perimeter = 2 x (Geodesic Length + Geodesic Thickness) where Area is filled area and Perimeter is the length of the particle edge not including the lengths of edges of holes in the particle. (real > 0)
22. Geodesic Thickness	See Geodesic Length
23. Intensity	The average grayscale value of the pixels making up a particle (grayscale sum / number of pixels making up the particle). (real [0, 255]; 255 is most intense)
24. Length	The maximum value of 36 feret measurements. (real > 0)
25. Particles Per Chain	The number of particles that were grouped into one particle based on the nearest neighbor distance. (integer > 1; almost always 1 if nearest neighbor distance is 0)
26. Perimeter	The length of the particle edge not including the lengths of edges of holes in the particle. (real > 0)
27. Roughness	A measure of the unevenness or irregularity of a particle's surface-the ratio of perimeter to convex perimeter. (real $\geq$ 1; 1 is the value for a filled shape with convex perimeter; larger values are for particles that have interior holes and/or a non-convex perimeter)
28. Sigma Intensity	Standard deviation of grayscale values. (real $\geq$ 0)
29. Sum Intensity	Sum of grayscale pixel values. (real > 0)
30. Symmetry	A measure of the symmetry of the particle about its center. If a particle is symmetric about the center then the value of Symmetry is 1.0. Typically used to locate 'broken' or partial particles. (real [0, 1])

https://doi.org/10.5194/tc-2022-148
Preprint. Discussion started: 26 August 2022

31. Transparency	1 – (ABD Diameter / ESD Diameter). (real [0, 1]; 0 is the value for a filled circle; values near 1 are for an elongated or irregular shape or a shape that has many interior holes)
32. Volume (ABD)	Sphere volume calculated from ABD Diameter. (real > 0)
33. Volume (ESD)	Sphere volume calculated from ESD Diameter. (real > 0)
34. Width	The minimum value of 36 feret measurements. (real > 0)

https://doi.org/10.5194/tc-2022-148
Preprint. Discussion started: 26 August 2022

The Cryosphere Discussions
Open Access
EGU

Appendix C: Training dataset images

Random batches of n=100 training images of each class. The images have been reshaped for better visualization. The diameters are indicated in the bottom left corners. Zoom in for best view.

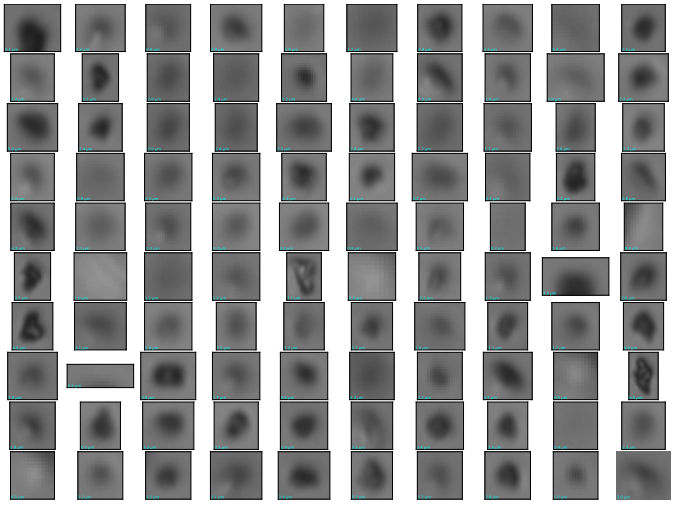

Figure C1. Dust.

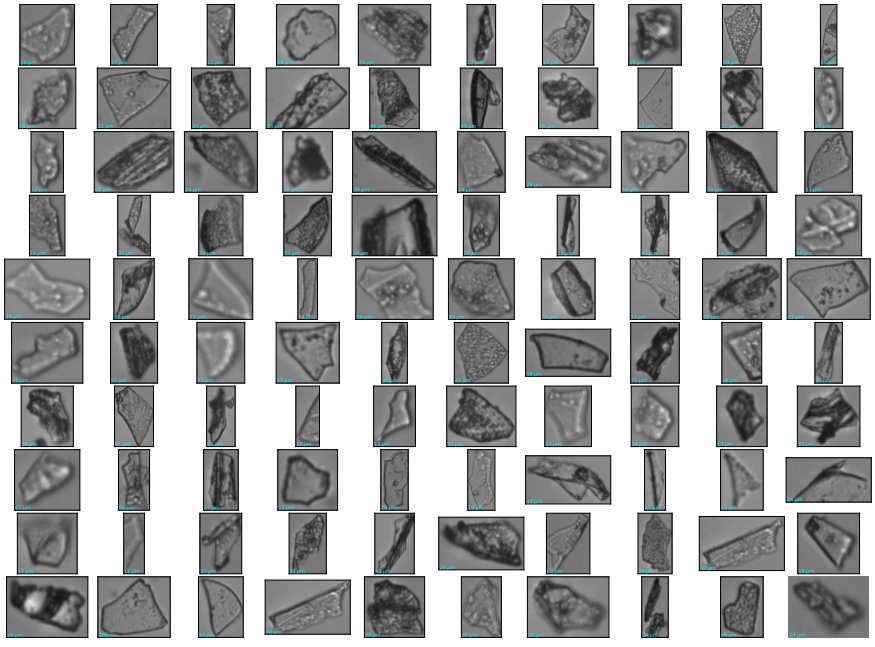

Figure C2. Felsic tephra.

https://doi.org/10.5194/tc-2022-148
Preprint. Discussion started: 26 August 2022

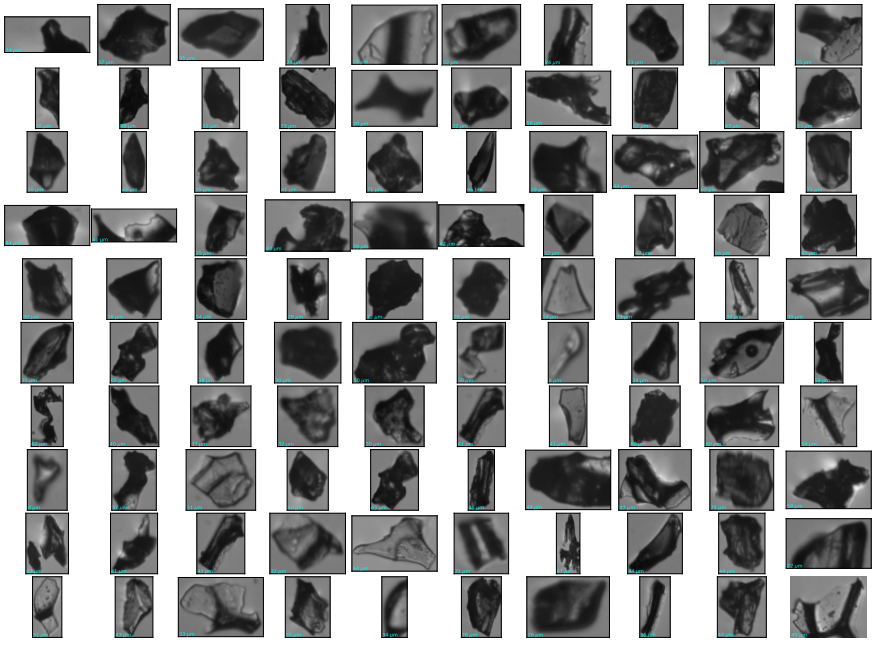

Figure C3. Basaltic tephra.

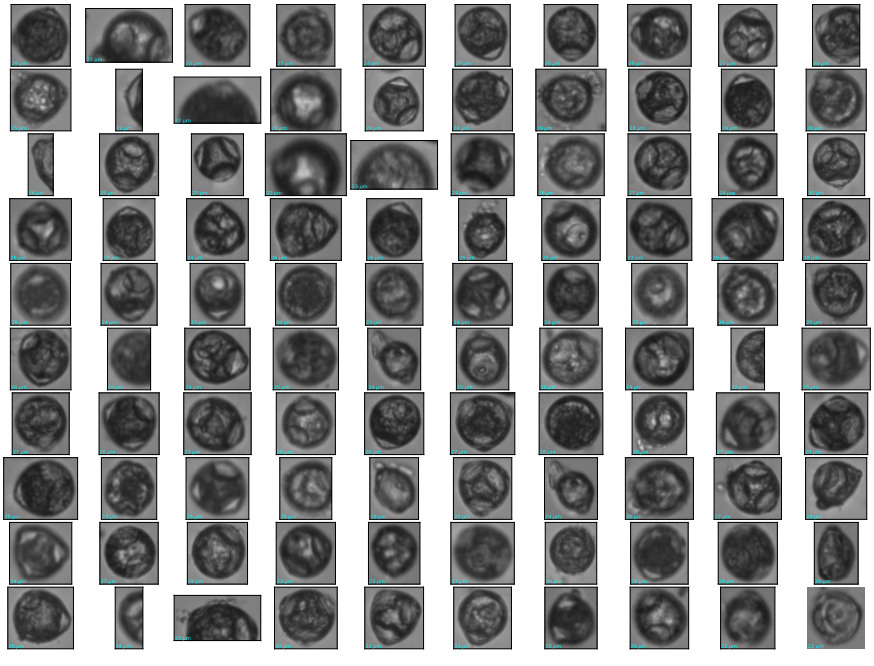

Figure C4. *Corylus avellana* pollen.

https://doi.org/10.5194/tc-2022-148
Preprint. Discussion started: 26 August 2022

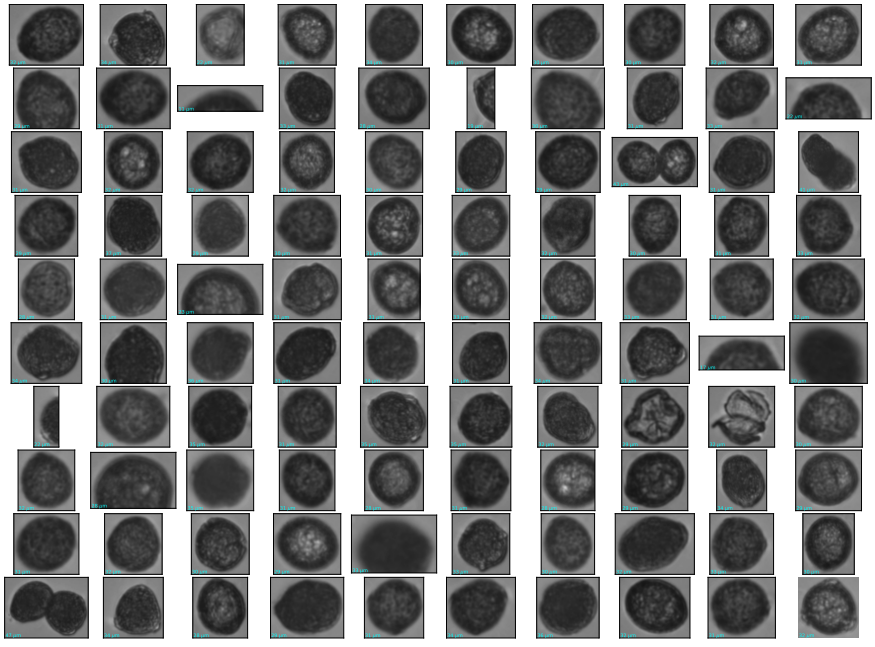

Figure C5. *Quercus robur* pollen.

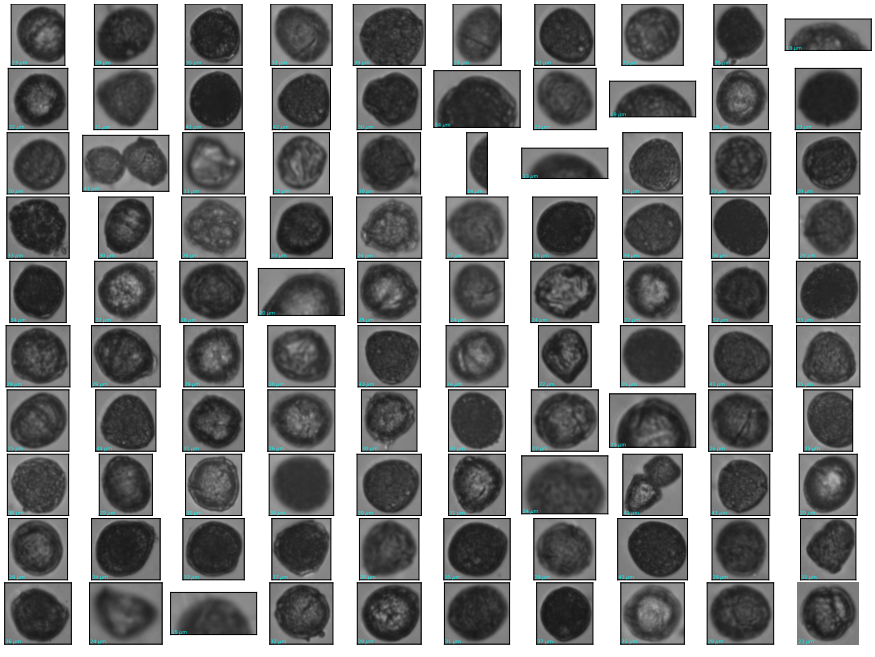

Figure C6. *Quercus suber* pollen.

https://doi.org/10.5194/tc-2022-148
Preprint. Discussion started: 26 August 2022

The Cryosphere
Discussions
Open Access
EGU

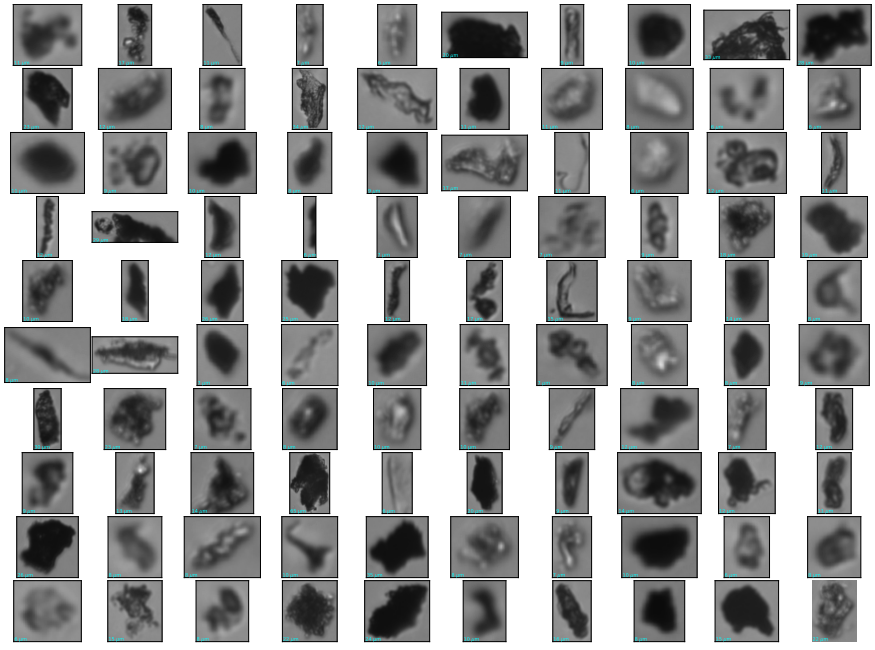

Figure C7. Contamination/Blurry particles.

Author contributions. NM conceived and conceptualized the idea. NM coded the model with support from AM, TP, SV and MP. The measurements were carried out by NM, ES, WvdB, EC, GB, FdR, JS, YR. The samples were provided by WvdB, EC, DF, FM, AR, GB, AS, FdR, JS, BD, MV, JPS, DJ. All authors contributed to the data analysis and interpretation of the results. EC and WvdB carried out the human validation of all tephra predictions. NM, WvdB, EC wrote the manuscript using feedback from all other co-authors.

Competing interests. Kerim Nisancioglu is member of the editorial board of the journal.

Acknowledgements. We would like to thank Harry Nelson for his support in setting up the FlowCam for ice core measurements.

https://doi.org/10.5194/tc-2022-148
Preprint. Discussion started: 26 August 2022

References

Abbott, P. M. and Davies, S. M.: Volcanism and the Greenland ice-cores: the tephra record, Earth-Science Reviews, 115, 173–191, 2012.

Baccolo, G., Delmonte, B., Di Stefano, E., Cibin, G., Crotti, I., Frezzotti, M., Hampai, D., Iizuka, Y., Marcelli, A., and Maggi, V.: Deep ice as a geochemical reactor: insights from iron speciation and mineralogy of dust in the Talos Dome ice core (East Antarctica), The Cryosphere, 15, 4807–4822, 2021.

Bigler, M., Svensson, A., Kettner, E., Vallelonga, P., Nielsen, M. E., and Steffensen, J. P.: Optimization of high-resolution continuous flow analysis for transient climate signals in ice cores, Environmental science & technology, 45, 4483–4489, 2011.

Bohleber, P., Erhardt, T., Spaulding, N., Hoffmann, H., Fischer, H., and Mayewski, P.: Temperature and mineral dust variability recorded in two low-accumulation Alpine ice cores over the last millennium, Climate of the Past, 14, 21–37, 2018.

Bourgeois, J. C.: Seasonal and interannual pollen variability in snow layers of arctic ice caps, Review of Palaeobotany and Palynology, 108, 17–36, 2000.

Cederstrøm, J. M., Van der Bilt, W. G., Støren, E. W., and Rutledal, S.: Semi-Automatic Ice-Rafted Debris Quantification With Computed Tomography, Paleoceanography and Paleoclimatology, 36, e2021PA004 293, 2021.

Cook, E., Portnyagin, M., Ponomareva, V., Bazanova, L., Svensson, A., and Garbe-Schönberg, D.: First identification of cryptotephra from the Kamchatka Peninsula in a Greenland ice core: Implications of a widespread marker deposit that links Greenland to the Pacific northwest, Quaternary Science Reviews, 181, 200–206, 2018A.

Cook, E., Davies, S. M., Guðmundsdóttir, E. R., Abbott, P. M., and Pearce, N. J.: First identification and characterization of Borrobol-type tephra in the Greenland ice cores: new deposits and improved age estimates, Journal of quaternary science, 33, 212–224, 2018B.

Cook, E. e. a.: Volcanism and the Greenland ice-cores: a new tephrochronological framework for the last glacial-interglacial transition based on cryptotephra deposits in three ice cores, in review, 2022.

Davies, S. M., Wastegård, S., Abbott, P., Barbante, C., Bigler, M., Johnsen, S., Rasmussen, T. L., Steffensen, J., and Svensson, A.: Tracing volcanic events in the NGRIP ice-core and synchronising North Atlantic marine records during the last glacial period, Earth and Planetary Science Letters, 294, 69–79, 2010.

Delmonte, B., Petit, J., and Maggi, V.: Glacial to Holocene implications of the new 27000-year dust record from the EPICA Dome C (East Antarctica) ice core, Climate Dynamics, 18, 647–660, 2002.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L.: Imagenet: A large-scale hierarchical image database, in: 2009 IEEE conference on computer vision and pattern recognition, pp. 248–255, Ieee, 2009.

Dome Fuji Ice Core Project Members: State dependence of climatic instability over the past 720,000 years from Antarctic ice cores and climate modeling, Science Advances, 3, https://doi.org/10.1126/sciadv.1600446, 2017.

Eichler, A., Schwikowski, M., Gäggeler, H. W., Furrer, V., Synal, H.-A., Beer, J., Saurer, M., and Funk, M.: Glaciochemical dating of an ice core from upper Grenzgletscher (4200 m asl), Journal of Glaciology, 46, 507–515, 2000.

EPICA community members: Eight glacial cycles from an Antarctic ice core, Nature, 429, 623 – 628, https://doi.org/10.1038/nature02599, 2004.

Fedele, F. G., Giaccio, B., Isaia, R., and Orsi, G.: The Campanian Ignimbrite eruption, Heinrich Event 4, and Paleolithic change in Europe: A high-resolution investigation, Geophysical Monograph-American Geophysical Union, 139, 301–328, 2003.

Festi, D., Kofler, W., Bucher, E., Carturan, L., Mair, V., Gabrielli, P., and Oeggl, K.: A novel pollen-based method to detect seasonality in ice cores: a case study from the Ortles glacier, South Tyrol, Italy, Journal of Glaciology, 61, 815–824, 2015.

https://doi.org/10.5194/tc-2022-148
Preprint. Discussion started: 26 August 2022

Festi, D., Schwikowski, M., Maggi, V., Oeggl, K., and Jenk, T. M.: Significant mass loss in the accumulation area of the Adamello glacier indicated by the chronology of a 46 m ice core, The Cryosphere, 15, 4135–4143, 2021.

Fritz, S. C., Brinson, B. E., Billups, W., and Thompson, L. G.: Diatoms at> 5000 meters in the Quelccaya Summit Dome Glacier, Peru, Arctic, Antarctic, and Alpine Research, 47, 369–374, 2015.

Gow, A. J. and Meese, D. A.: The distribution and timing of tephra deposition at Siple Dome, Antarctica: possible climatic and rheologic implications, Journal of Glaciology, 53, 585–596, 2007.

He, K., Zhang, X., Ren, S., and Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification, in: Proceedings of the IEEE international conference on computer vision, pp. 1026–1034, 2015.

He, K., Zhang, X., Ren, S., and Sun, J.: Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778, 2016.

Karnan, C., Jyothibabu, R., Manoj Kumar, T., Jagadeesan, L., and Arunpandi, N.: On the accuracy of assessing copepod size and biovolume using flowCAM and traditional microscopy, Indian Journal of Geo-Marine Sciences, 46, 1261 – 1264, 2017.

Krizhevsky, A., Sutskever, I., and Hinton, G. E.: Imagenet classification with deep convolutional neural networks, Advances in neural information processing systems, 25, 2012.

Kydd, J., Rajakaruna, H., Briski, E., and Bailey, S.: Examination of a high resolution laser optical plankton counter and FlowCAM for measuring plankton concentration and size, Journal of Sea Research, 133, 2–10, 2018.

Lambert, F., Delmonte, B., Petit, J.-R., Bigler, M., Kaufmann, P. R., Hutterli, M. A., Stocker, T. F., Ruth, U., Steffensen, J. P., and Maggi, V.: Dust-climate couplings over the past 800,000 years from the EPICA Dome C ice core, Nature, 452, 616–619, 2008.

Legrand, M. and Mayewski, P.: Glaciochemistry of polar ice cores: A review, Reviews of geophysics, 35, 219–243, 1997.

Lin, J., Svensson, A., Hvidberg, C. S., Lohmann, J., Kristiansen, S., Dahl-Jensen, D., Steffensen, J. P., Rasmussen, S. O., Cook, E., Kjær, H. A., et al.: Magnitude, frequency and climate forcing of global volcanism during the last glacial period as seen in Greenland and Antarctic ice cores (60–9 ka), Climate of the Past, 18, 485–506, 2022.

Linsinger, T. P., Gerganova, T., Kestens, V., and Charoud-Got, J.: Preparation and characterisation of two polydisperse, non-spherical materials as certified reference materials for particle size distribution by static image analysis and laser diffraction, Powder Technology, 343, 652–661, 2019.

Loshchilov, I. and Hutter, F.: Decoupled weight decay regularization, arXiv preprint arXiv:1711.05101, 2017.

Lowe, D. J.: Tephrochronology and its application: a review, Quaternary Geochronology, 6, 107–153, 2011.

Lowe, D. J. and Hunt, J. B.: A summary of terminology used in tephra-related studies, Tephra: Les Dossiers de l'Archeo-Logis, 1, 17–22, 2001.

McInnes, L., Healy, J., and Melville, J.: Umap: Uniform manifold approximation and projection for dimension reduction, arXiv preprint arXiv:1802.03426, 2018.

Muthreich, F., Zimmermann, B., Birks, H. J. B., Vila-Viçosa, C. M., and Seddon, A. W.: Chemical variations in Quercus pollen as a tool for taxonomic identification: Implications for long-term ecological and biogeographical research, Journal of Biogeography, 47, 1298–1309, 2020.

Nakazawa, F., Fujita, K., Uetake, J., Kohno, M., Fujiki, T., Arkhipov, S. M., Kameda, T., Suzuki, K., and Fujii, Y.: Application of pollen analysis to dating of ice cores from lower-latitude glaciers, Journal of Geophysical Research: Earth Surface, 109, 2004.

Narcisi, B., Petit, J. R., Delmonte, B., Scarchilli, C., and Stenni, B.: A 16,000-yr tephra framework for the Antarctic ice sheet: a contribution from the new Talos Dome core, Quaternary Science Reviews, 49, 52–63, 2012.

https://doi.org/10.5194/tc-2022-148
Preprint. Discussion started: 26 August 2022

North Greenland Ice Core Project members: High-resolution record of Northern Hemisphere climate extending into the last interglacial period, Nature, 431, 147–151, 2004.

Petit, J.-R., Briat, M., and Royer, A.: Ice age aerosol content from East Antarctic ice core samples and past wind strength, Nature, 293, 391–394, 1981.

Petit, J.-R., Jouzel, J., Raynaud, D., Barkov, N. I., Barnola, J.-M., Basile, I., Bender, M., Chappellaz, J., Davis, M., Delaygue, G., et al.: Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica, Nature, 399, 429–436, 1999.

Plunkett, G., Sigl, M., Pilcher, J. R., McConnell, J. R., Chellman, N., Steffensen, J., and Büntgen, U.: Smoking guns and volcanic ash: the importance of sparse tephras in Greenland ice cores, Polar research, 39, 2020.

Pyne-O'Donnell, S.: The taphonomy of Last Glacial–Interglacial Transition (LGIT) distal volcanic ash in small Scottish lakes, Boreas, 40, 131–145, 2011.

Reis, R. S. d., da Rocha Ribeiro, R., Delmonte, B., Ramirez, E., Dani, N., Mayewski, P. A., and Simões, J. C.: The Recent Relationships Between Andean Ice-Core Dust Record and Madeira River Suspended Sediments on the Wet Season, Frontiers in Environmental Science, 10, https://doi.org/10.3389/fenvs.2022.840884, 2022.

Ruth, U., Wagenbach, D., Steffensen, J. P., and Bigler, M.: Continuous record of microparticle concentration and size distribution in the central Greenland NGRIP ice core during the last glacial period, Journal of Geophysical Research: Atmospheres, 108, 2003.

Ruth, U., Barbante, C., Bigler, M., Delmonte, B., Fischer, H., Gabrielli, P., Gaspari, V., Kaufmann, P., Lambert, F., Maggi, V., et al.: Proxies and measurement techniques for mineral dust in Antarctic ice cores, Environmental science & technology, 42, 5675–5681, 2008.

Schwikowski, M.: Reconstruction of European air pollution from Alpine ice cores, in: Earth Paleoenvironments: records preserved in mid-and low-latitude glaciers, pp. 95–119, Springer, 2004.

Sigl, M., Winstrup, M., McConnell, J. R., Welten, K. C., Plunkett, G., Ludlow, F., Büntgen, U., Caffee, M., Chellman, N., Dahl-Jensen, D., et al.: Timing and climate forcing of volcanic eruptions for the past 2,500 years, Nature, 523, 543–549, 2015.

Simonsen, M. F., Cremonesi, L., Baccolo, G., Bosch, S., Delmonte, B., Erhardt, T., Kjær, H. A., Potenza, M., Svensson, A., and Vallelonga, P.: Particle shape accounts for instrumental discrepancy in ice core dust size distributions, Climate of the Past, 14, 601–608, 2018.

Simonyan, K. and Zisserman, A.: Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556, 2014.

Svensson, A., Andersen, K. K., Bigler, M., Clausen, H. B., Dahl-Jensen, D., Davies, S., Johnsen, S. J., Muscheler, R., Parrenin, F., Rasmussen, S. O., et al.: A 60 000 year Greenland stratigraphic ice core chronology, Climate of the Past, 4, 47–57, 2008.

Turney, C. S., Harkness, D. D., and Lowe, J. J.: The use of microtephra horizons to correlate Late-glacial lake sediment successions in Scotland, Journal of Quaternary Science: Published for the Quaternary Research Association, 12, 525–531, 1997.

Vallelonga, P., Barbante, C., Cozzi, G., Gaspari, V., Candelone, J.-P., Van De Velde, K., Morgan, V. I., Rosman, K. J., Boutron, C. F., and Cescon, P.: Elemental indicators of natural and anthropogenic aerosol inputs to Law Dome, Antarctica, Annals of Glaciology, 39, 169–174, 2004.

Van der Bilt, W. G., Cederstrøm, J. M., Støren, E. W., Berben, S. M., and Rutledal, S.: Rapid tephra identification in geological archives with computed tomography: experimental results and natural applications, Frontiers in Earth Science, p. 710, 2021.

Veres, D., Lane, C. S., Timar-Gabor, A., Hambach, U., Constantin, D., Szakács, A., Fülling, A., and Onac, B. P.: The Campanian Ignimbrite/Y5 tephra layer–A regional stratigraphic marker for Isotope Stage 3 deposits in the Lower Danube region, Romania, Quaternary International, 293, 22–33, 2013.

https://doi.org/10.5194/tc-2022-148
Preprint. Discussion started: 26 August 2022

Wolff, E. W., Moore, J. C., Clausen, H. B., Hammer, C. U., Kipfstuhl, J., and Fuhrer, K.: Long-term changes in the acid and salt concentrations of the Greenland Ice Core Project ice core from electrical stratigraphy, Journal of Geophysical Research: Atmospheres, 100, 16 249–16 263, 1995.

Zeiler, M. D. and Fergus, R.: Visualizing and understanding convolutional networks, in: European conference on computer vision, pp. 818–833, Springer, 2014.

Žunić, J., Hirota, K., and Rosin, P. L.: A Hu moment invariant as a shape circularity measure, Pattern Recognition, 43, 47–57, 2010.