The Cryosphere, 17, 539–565, 2023
https://doi.org/10.5194/tc-17-539-2023
© Author(s) 2023. This work is distributed under
the Creative Commons Attribution 4.0 License.

The Cryosphere
Open Access

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[8,9], Amalie R. F. Mygind[10], Troels Petersen[10], Andrea Spolaor[2], Sebastiano Vascon[1], Marcello Pelillo[1], Patrizia Ferretti[1], Rafael S. dos Reis[11], Jefferson C. Simões[11,12], Yuval Ronen[4], Barbara Delmonte[9], Marco Viccaro[13,14], Jørgen Peder Steffensen[3], Dorthe Dahl-Jensen[3], Kerim H. Nisancioglu[4,5], and Carlo Barbante[1,2]

[1]Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice, 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]GeoSphere Austria, Neulinggasse 38, 1030 Vienna, Austria
[7]Department of Biological Sciences, University of Bergen, Thormøhlensgate 53A, 5006 Bergen, Norway
[8]Laboratory of Environmental Chemistry (LUC), Paul Scherrer Institute, Forschungsstrasse 111, 5232 Villigen, Switzerland
[9]Department of Environmental Sciences, University of Milano-Bicocca, Piazza della Scienza 1, 20126 Milan, Italy
[10]Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, 2100 Copenhagen, Denmark
[11]Centro Polar e Climático, Universidade Federal do Rio Grande do Sul, Porto Alegre 91501-970, Brazil
[12]Climate Change Institute, University of Maine, Orono, ME 04469, USA
[13]Dipartimento di Scienze Biologiche Geologiche e Ambientali, Università degli Studi di Catania, Corso Italia 57, 95129 Catania, Italy
[14]Istituto Nazionale di Geofisica e Vulcanologia – Sezione di Catania, Osservatorio Etneo, Piazza Roma 2, 95125 Catania, Italy

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

Received: 15 July 2022 – Discussion started: 26 August 2022
Revised: 2 January 2023 – Accepted: 18 January 2023 – Published: 7 February 2023

Abstract. Insoluble particles in ice cores record signatures of past climate parameters like vegetation dynamics, volcanic activity, and aridity. For some of them, the analytical detection relies 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 seven commonly found classes, namely mineral dust, felsic and mafic (basaltic) volcanic ash grains (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 potential and its 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 procedures. 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.

Published by Copernicus Publications on behalf of the European Geosciences Union.

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, to 800 000 and 125 000 years ago and register variability in 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 contains 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 the production of past climate records at various spatial and temporal 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, the glass component of volcanic ash, pollen grains, and other biological matter, as well as microfossils sourced from oceans 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, and parameters controlling atmospheric transport and deposition.

By far the most abundant particle type in ice cores is mineral dust particles that are sourced from continental surfaces and are transported and dry- or wet-deposited onto ice sheets and glaciers (Legrand and Mayewski, 1997). The detection of dust is fundamental for investigating 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 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 actually being 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), which is an electrical-based analyzer that operates in discrete mode and cannot 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 in ice cores can contain volcanic minerals, rock fragments, and volcanic glass shards. Across the spectrum of volcanic material, in this work we target cryptotephras and glass shards from individual eruptions that can form stratigraphically distinct deposits in ice cores and in marine and terrestrial sediments that are invisible to the naked eye (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 either by using different methods or in combination. Potential volcanic layers can be identified by electrical conductivity or sulfate concentration measurements during CFA analyses (e.g., Wolff et al., 1995) at high resolution. Not all tephra layers, particularly cryptotephras, 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). For example, in the glacial period, tephra-rich deposits consistently lack coeval chemostratigraphic peaks, partially due to signal neutralization by dust (Bourne et al., 2015). Manual discrete subsampling of such selected intervals of interest is also carried out to maximize tephra layer identification. During this method, ice samples are individually processed and manually inspected using optical bench microscopy (e.g., Bourne et al., 2015; Cook et al., 2018). If the presence of cryptotephra is confirmed, then the glass shards are individually counted. 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.

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 are manually evaluated by palynologists via optical bench microscopy. In summary, the extraction and detection of climate-relevant ice core particles is extremely laborious.

Over the last 10 years, neural networks and in particular convolutional neural networks (CNNs) 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 a 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). Despite the advances of such techniques, their application to environmental studies has lagged behind to very few and recent examples (Kerr et al., 2020; Viertel and König, 2022).

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. In our setup, to generate the ice particle imagery, we rely on a flow imaging microscopy instrument (the FlowCam; Fig. A1) 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 six classes of particles, namely mineral dust, tephra (basaltic and felsic), and three pollen grains potentially present in alpine ice records, i.e., *Corylus avellana*, *Quercus robur*, and *Quercus suber*. An additional seventh 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 The FlowCam: settings and image and feature extraction

The FlowCam instrument (Yokogawa/Fluid Imaging Technologies; VS-IV 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, connection tubing can allow sampling from discrete sample vials or from a continuous flow system (Appendix A). The flow cell used in our setup (depth $= 80\,\mu m$; width $= 570\,\mu m$) allows the flow of particles up to $80\,\mu m$ in diameter in the maximum dimension. A 1.0 mL volume syringe pump is set to operate at a flow rate of $0.02\,mL\,min^{-1}$. While passing through the flow cell, the sample is imaged by a camera equipped with a $20\times$ magnification objective. The camera flash duration is set to $65\,\mu s$ and is operated at the maximum 22 frames per second. 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 determine a calibration factor of $0.2752\,\mu m$ per pixel in the resulting monochrome 1280×960 pixels 8 bit TIFF images.

The mechanics of particle image creation is performed by the native FlowCam software (VisualSpreadsheet v3.4). All camera 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 to be 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, then they are considered to be background and set to 0. Once the signal–background binary image is created, the single particle images are extracted by segmenting out the pixels flagged as signal (Fig. A1). Each created image thus represents one particle. The threshold value, set to 18, and the camera focus are optimized by acquiring images of spherical polystyrene $25\,\mu m$ beads and by minimizing the standard deviation of the resulting size distribution (Fig. S1 in the Supplement). For each acquired particle image, the FlowCam software calculates a number of numerical features, hereafter also referred to as "metadata", mostly reflecting the geometrical properties of the particles. These numerical features 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 seven classes of particles, i.e., mineral dust, tephra (basaltic and felsic), three pollen species of *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–C7). Each item in 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 one-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 in the training dataset is visually inspected and validated by the human eye.

- The training dataset of dust particles is created by measuring the water solutions of FD066 (Linsinger et al., 2019; Table 1; Fig. C1), an aluminum oxide powder containing particles with a mean size distribution of $2.5\,\mu m$ and rarely exceeding $6\,\mu m$ (Table 2). Such a dust training set is therefore suited to mimic dust found in inland Antarctic and Greenland ice cores, typically below $4\,\mu m$ (e.g., Delmonte et al., 2002; Ruth et al., 2003).

Table 1. Training dataset.

Class	Sample type	Sample origin	Approximate size range	No. of training items
(1) Dust	Conundrum powder (Al_2O_3)	Standard Reference Material (ERM-FD066)	< 10 µm	8000
(2) Felsic tephra	Campanian Ignimbrite	Southeastern Romania (43–44° N, 23–24° E)	(8, 80) µm	7125
(3) Basaltic tephra	Grímsvötn (Iceland)	Kirkjubæjarklaustur (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	47 223
(5) Pollen *Q. robur*	Pollen	Portugal (41.476–41.155° N, 8.701–8.563° W)	(10, 40) µm	35 276
(6) Pollen *Q. suber*	Pollen	Portugal (41.155–41.151° N, 8.565–8.660° W)	(10, 40) µm	31 745
(7) Contamination/blurry	Outer-core ice samples	GRIP ice core external layer	(5, 80) µm	11 439

- 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 basaltic ash is darker. The felsic tephra training dataset consists of Campanian Ignimbrite volcanic ash from the 39.3 ± 0.1 ka Phlegraean Fields eruption (Fedele et al., 2003; Table 1; Fig. C2). The phonolitic–trachytic (~ 60 wt % SiO_2) ash was sampled ~ 1000 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). Ash samples were collected on 22 May 2011 in the town of Kirkjubæjarklaustur, about 70 km southwest of the Grímsvötn caldera. After collection, samples were dried and stored in plastic beakers. Ashes of both types were dry-sieved at 63 µm to limit the maximum dimension and fit the max. 80 µm size constraint (min. ~ 8 µm) of the flow cell. 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 blurry 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, i.e., *C. avellana*, *Q. robur*, and *Q. suber* (Table 1; Figs. C4, C5, and 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 were prepared by mixing together pollen from different trees. Both *Quercus* species were collected in Portugal and treated similarly. Occasionally, if pollen grains flow at the boundary of the camera field of view, they result in being partially captured. We decided to keep fractional pollen images 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). The ice core surface 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 particle 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. While blurry images are an intrinsic limitation of this methodology, the contamination/blurry class serves the purpose of an important controlled channel for the model to be able to identify particles that do not carry climate significance.

2.3 Model

2.3.1 Hybrid deep neural 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 multilayer perceptron (MLP), fed, respectively, by particle images and the corresponding 34-dimensional (34-d) numerical feature vectors (metadata). The CNN consists of 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 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.

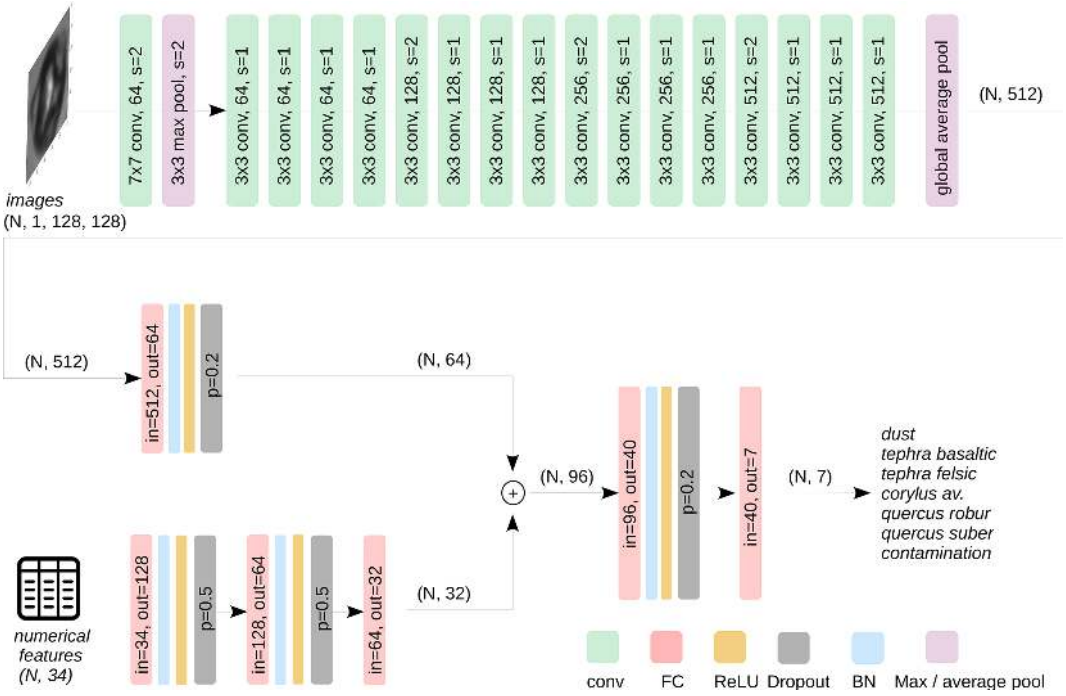

Figure 1. Model architecture. The top branch of the network is a ResNet-18 CNN (He et al., 2016). BN, ReLU layers, and skip connections are omitted for clarity. The bottom branch operates on the numerical features and consists of three-layer multilayer perceptron. The separate outputs of the two branches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

2.3.2 Data preprocessing and augmentation

All images are reshaped by linear interpolation to 128×128 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

The data are split into three separate datasets, namely training, validation, and test. Both the validation and test datasets consist of a random subset of 500 items per class, 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 training dataset is unbalanced, a weighted loss is implemented by enforcing a different weight w for each class c as follows (Eq. 1):

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

Underfitting and overfitting is checked after every epoch (a cycle of training the network) by monitoring the loss and the accuracy of 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 and 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 %, with the remaining fraction being misclassified mostly as the

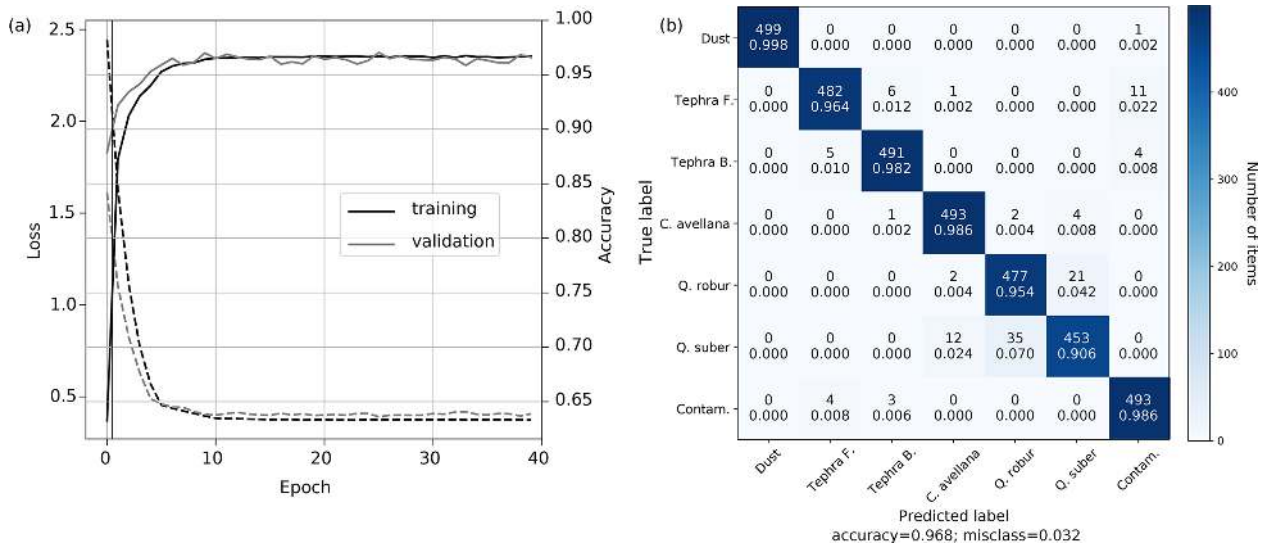

Figure 2. (a) Model loss (dashed) and accuracy (solid) evaluated during training (black) and validation (gray). **(b)** Confusion matrix of the best model evaluated on the test dataset. The accuracy across all classes is 96.8 %. Most misclassifications occur within the two *Quercus* classes.

wrong *Quercus*. No misclassification is found between the three pollen types and all other classes.

3 Results and discussion

The following discussion is divided into three sections. In Sect. 3.1, we investigate the FlowCam's 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 Sect.3.2, we discuss pollen and the representativeness of their training datasets. In Sect. 3.3, the model is deployed on Greenland ice core samples containing known volcanic ash horizons.

3.1 Dust

3.1.1 Standard Reference Material: size reconstruction, limits of detection (LOD), 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 area-based diameter distribution (ABD; Appendix B). 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 imaged volume of the sample, and multiplying by the density. The aluminum oxide density is $3.96\,\mathrm{g\,cm^{-3}}$. An alternative metric to the ABD is the equivalent spherical diameters (ESDs; 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 tends 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 therefore be considered appropriate for reconstructing the size of dust particles with a distribution similar to that of the FD066 material and of spheres (Fig. S1).

Given the low dust concentrations in ice core records, it is crucial to investigate the blank levels of the analytical system and the impurity content of the used glassware. We define "system blank" as the instrumental response to ultrapure water (UPW; $18.2\,\mathrm{M\Omega\,cm^{-1}}$) directly injected into the system. The system blanks can be thought of as the blank level of a CFA system, 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 five 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 LOD of the system blanks are, respectively, 6 ppb (parts per billion) and

Table 2. Comparison between the FD066 ABD size distributions certified by scanning electron microscopy (SEM; Linsinger et al., 2019) and calculated using the FlowCam (this study).

FD066 ABD size distribution percentiles	ABD (µm, $\pm 1\sigma$) 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

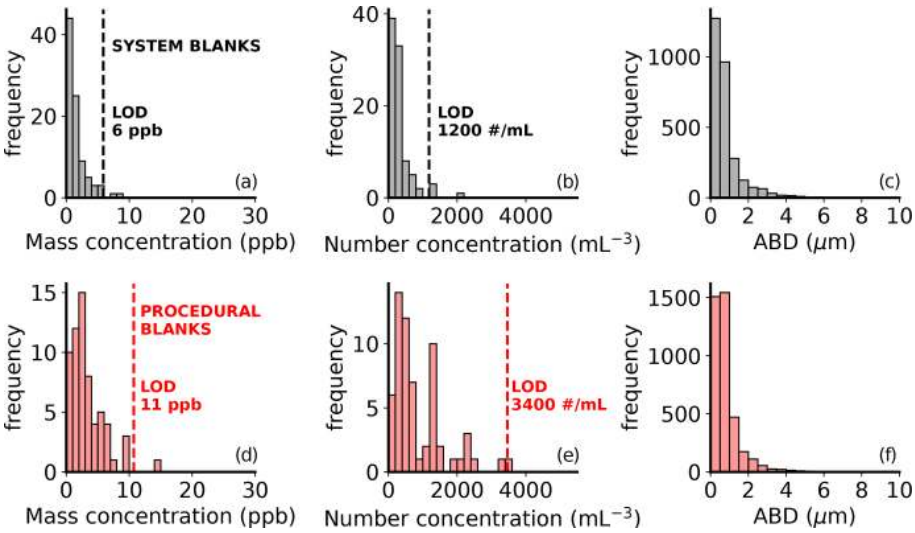

Figure 3. Analysis of $n = 91$ system blanks (black; **a**, **b**, and **c**) and $n = 63$ procedural blanks (red; **d**, **e**, and **f**). Panels **(a)** and **(d)** show the median mass concentrations of 1.0 ± 1.6 ppb (parts per billion; 1σ) and 2.4 ± 2.8 ppb (1σ) result in a 3σ LOD of 6 ppb for system blanks and 11 ppb for procedural blanks, respectively. Panels **(b)** and **(e)** show the number concentration distributions and respective LOD. Panels **(c)** and **(f)** indicate the size distributions of blank particles in system (**c**; $N = 2864$) and procedural (**f**; $N = 3945$) 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}.

1200 no. mL^{-1} (no. mL^{-1} is the number or particles per unit of sample volume). The mass concentration and number concentration LOD of the procedural blanks are, respectively, 11 ppb and 3400 no. mL^{-1}. In comparison, the LOD of the CC is 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 LOD thus allows quantification of dust in all sites globally, except for coastal Antarctic interglacial records. It is likely possible to further lower the instrument LOD 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 (parts per million; 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 of 3.96 g cm^{-3}. Similarly prepared samples are measured by a Coulter counter (CC), 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 from 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 the low Antarctic to high mid-latitude glacier dust levels. The residual distribution (mean

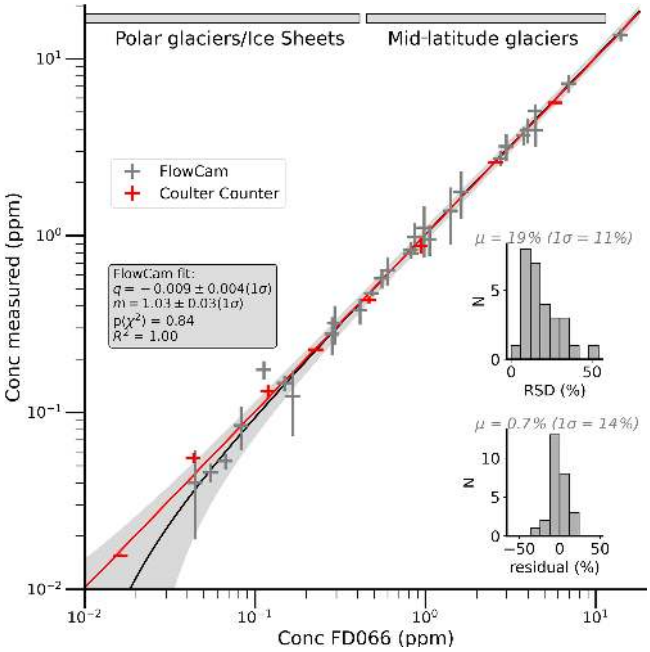

Figure 4. Comparison between nominal and measured concentrations of FD066 dust samples using the FlowCam and the Coulter counter. An orthogonal distance regression on the FlowCam data (black line with 3σ confidence interval in gray) shows good linearity over 3 orders of magnitude. The red line refers 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, and 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.

of 0.7 %; $1\sigma = 14$ %) 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 relative standard deviation (RSD) distribution) and averages 19 % ($1\sigma = 11$ %). The CC measurements also show good linearity ($q = 0.001 \pm 0.002$; $m = 1.05 \pm 0.05$). This experiment shows that both instruments yield accurate size and volume reconstructions for the irregularly shaped FD066 particles.

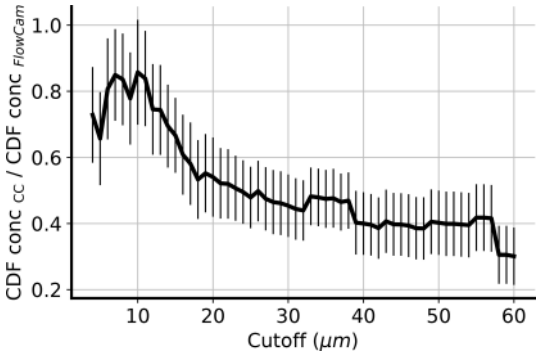

Figure 5. Mass concentration cumulative distribution function (CDF) ratios between the CC and the FlowCam as a function of a size cutoff. The best agreement is found at a cutoff value of 10 µm. If larger particles are included in the quantification of the concentration, then the FlowCam concentrations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

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 g cm^{-3} is assumed.

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 have 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, $\mathrm{CDF}(x) = \int_0^x \mathrm{conc}(z)\mathrm{d}z$, 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 2-fold. 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^{-1}, and maximizing the camera shutter to 22 frames per second). For example, for a 3 min analysis, only 0.025 mL of the sample is imaged, compared to 0.5 mL on the CC. The low statistics have 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 (≥ 50 µm) particle concentra-

tions 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 detected throughout the analysis time. We argue that the fast gravitational separation of particles of different sizes leads to underestimated concentrations, especially for a 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 increasing from 3 to 60 µm. The comparison is quantified by evaluating the slope of an orthogonal distance linear regression between the concentration cumulative distribution functions (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 ~ 3. This analysis is consistent with the good match previously found when using the small-sized FD066 material (Fig. 4).

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 20× 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) and the FlowCam mass (and number) 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 and (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. Five different *C. avellana* types were made available for this experiment, labeled A, B, C, D, and E, and they reflect samples collected from different trees within the same sampling region. We built three training datasets, i.e., type A, type B, and type "mix", with the last one prepared by mixing all five types together. We then train the model four 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 *Corylus* B, the percentage of *Corylus* classification in the *Corylus* B sample increases to 96 %. If a *Corylus* mix training set is used, then the correct accuracies are 97 %. If the model is trained with all datasets joined together (A + B + mix), then 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 each time it is used 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 *Q. 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 it is 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.

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 %, similar 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 increas-

Table 3. Pollen experiment results. The accuracies are indicated as the average of the N replicates (*C. avellana* $N = 5$, *Q. robur* $N = 3$, and *Q. suber* $N = 10$). In parentheses, the standard deviation of the replicates are given. Bold font refers to the cases in which the most general training datasets are used.

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 ($N = 35\,276$) *Q. suber* A + B ($N = 31\,745$)	0.48 (0.03)	0.07 (0.01)	0.45 (0.03)
C. avellana B ($N = 13\,713$) *Q. robur* A + B ($N = 35\,276$) *Q. suber* A + B ($N = 31\,745$)	0.961 (0.007)	0.005 (0.004)	0.034 (0.003)
C. avellana mix ($N = 25\,186$) *Q. robur* A + B ($N = 35\,276$) *Q. suber* A + B ($N = 31\,745$)	0.970 (0.007)	0.006 (0.001)	0.023 (0.006)
C. avellana A + B + mix ($N = 47\,723$) *Q. robur* A + B ($N = 35\,276$) *Q. suber* A + B ($N = 31\,745$)	**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 ($N = 47\,723$) *Q. robur* A ($N = 10\,239$) *Q. suber* A + B ($N = 31\,745$)	0.11 (0.01)	0.023 (0.005)	0.87 (0.01)
C. avellana A + B + mix ($N = 47\,723$) *Q. robur* B ($N = 24\,537$) *Q. suber* A + B ($N = 31\,745$)	0.038 (0.007)	0.910 (0.01)	0.051 (0.003)
C. avellana A + B + mix ($N = 47\,723$) *Q. robur* A + B ($N = 35\,276$) *Q. suber* A + B ($N = 31\,745$)	**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 ($N = 47\,723$) *Q. robur* A + B ($N = 35\,276$) *Q. suber* A ($N = 10\,663$)	0.47 (0.10)	0.49 (0.09)	0.038 (0.008)
C. avellana A + B + mix ($N = 47\,723$) *Q. robur* A + B ($N = 35\,276$) *Q. suber* B ($N = 20\,582$)	0.07 (0.03)	0.03 (0.01)	0.90 (0.03)
C. avellana A + B + mix ($N = 47\,723$) *Q. robur* A + B ($N = 35\,276$) *Q. suber* A + B ($N = 31\,745$)	**0.07 (0.03)**	**0.03 (0.01)**	**0.90 (0.03)**

ing 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*, $98 \pm 2\,\%$, 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.

We finally train a model five times using the largest datasets, namely *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 sam-

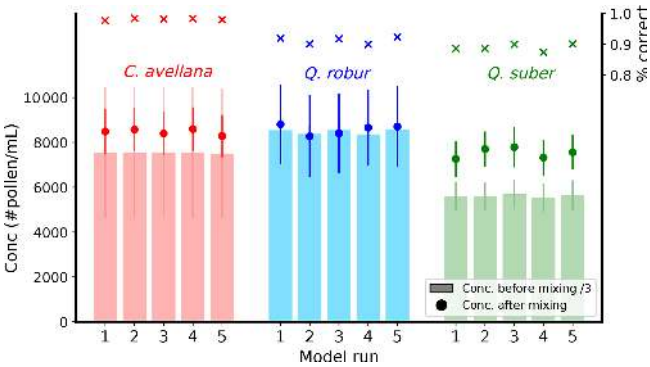

Figure 6. Quantification of pollen concentrations in single-type samples and in a mixed sample. The model was deployed to classify and quantify pollen concentrations in three samples containing purely *C. avellana* (red), *Q. robur* (blue), and *Q. suber* (green) pollen. The percentages of correct predictions among the three pollen classes are indicated in the top right axis as a function of five 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 aliquots).

ple. 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 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 is indicated by 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, leading 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 the case of low concentrations, a similar underestimation of the absolute number of pollen is to be expected, by a factor ~ 2 (Fig. 5). Intensive analysis of alpine ice core records (where pollen is expected) is the next logical step.

3.3 Tephra

We deployed the model to investigate the content of 12 samples from the Greenland Ice Core Project (GRIP) ice core (Table 4). Specifically, seven of these contain known tephra deposits, selected from the tephrochronology framework of Cook et al. (2022), while the remaining five samples are known to be devoid of tephra grains (i.e., tephra grains were not observed by bench microscopy).

The seven tephra deposits were resampled 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 et al. (2022), to ensure the same deposits could be found and thus producing replicate tephra-containing ice core samples. The five 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 to those selected for our training dataset, i.e., felsic (rhyolitic), mafic (basaltic), 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 et al., 2022), and 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 ice 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.

3.3.1 Optical microscopy for tephra analysis

The seven samples chosen for replicate tephra analysis in this study were originally identified using optical microscopy, following the sampling methodology outlined in Cook et al. (2022). Specifically, the samples were melted, centrifuged, and 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 per milliliter (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

Table 4. GRIP sample details and tephra counts by manual optical microscopy. The sample ages are derived from the GICC05 chronology. Note that b2k means before the year 2000. Bølling–Allerød/Greenland Interstadial 1 is GI-1 and Glacial/Greenland Stadial 2 is GS-2.

GRIP sample (age)	Sample depth interval (cm)	Depth (m)	Microscopy ice meltwater (mL)	Microscopy tephra counts ($\pm 1\sigma$)[b]	Microscopy tephra concentrations (no. mL^{-1})
3046 (GI-1b, 13 186 yr b2k)	0–20	1674.75–1674.95	33	0	0
	20–40[a]	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, 14 191 yr b2k)	0–20[a]	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, 17 238 yr b2k)	0–20[a]	1816.1–1816.3	34	18 ± 1	0.53 ± 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, 17 326 yr b2k)	0–20	1817.75–1817.95	33	0	0
	20–40	1817.95–1818.15	34	0	0
	40–55[a]	1818.15–1818.3	23	431 ± 100	19 ± 4

[a] Sample that corresponds to the specified age. [b] The uncertainties are estimated.

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, i.e., $\sim 8\,\mu m$. 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 (outlined in Sect. 3.3.1) 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 in volume. 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, which is 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 no. mL^{-1} (Table 5; column J). Although of 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 by a factor of about 2–3 from the real concentrations (Sect. 3.1.2; Fig. 5) because the fluidics/loss of material as gravitational settling preferentially affects large particles.

3.3.3 Human assessment of modeled tephra predictions

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

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

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
GRIP sample (age)	Depth interval (cm)	Volume (mL)	Imaged volume (mL)	Acquired particles	Dust	Tephra felsic, basaltic (Human1 yes, maybe, no), (Human2 yes, maybe, no)	*Corylus*	*Quercus robur*, *Quercus suber*	Cont.	AI tephra conc. (no. mL^{-1})	AI tephra conc. |Human1 Y+M (no. mL^{-1})	AI tephra conc. |Human1 Y (no. mL^{-1})	AI tephra conc. |Human2 Y+M (no. mL^{-1})	AI tephra conc. |Human2 Y (no. mL^{-1})
3046	0–20	56	0.23	198 961	188 053	43, 43 (10, 30, 46), (9, 33, 44)	16	28, 33	10 745	3.3	1.6	0.4	1.6	0.4
3046	20–40	56	0.33	257 002	241 826	114, 184 (83, 115, 100), (15, 152, 131)	31	50, 49	14 748	15	9.7	4.1	8.2	0.7
3046	40–55	45	0.42	58 217	54 481	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	46 571	42 512	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	86 352	81 280	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	38 055	35 203	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	170 457	164 457	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	349 773	338 911	87, 89 (9, 56, 111), (0, 111, 65)	16	23, 32	10 615	14	5.3	0.7	9.1	0
3303	40–55	28	0.33	334 060	322 070	41, 72 (17, 55, 41), (0, 72, 41)	4	8, 11	11 854	12	7.7	1.8	7.7	0
3306	0–20	36	0.33	608 856	588 023	99, 118 (8, 72, 137), (0, 117, 100)	36	84, 55	20 441	18	6.7	0.7	9.9	0
3306	20–40	37	0.31	437 339	425 309	67, 82 (5, 46, 98), (0, 91, 58)	26	57, 28	11 770	12	4.1	0.4	7.4	0
3306	40–55	30	0.36	499 420	489 356	69, 64 (26, 59, 48), (1, 55, 77)	9	14, 17	9891	11	7.1	2.1	4.7	0.08

tephra predictions in the 12 GRIP samples (irrespective of whether they are predicted as felsic or basaltic) into three classes of "yes", "maybe", and "no" (Table 5). According to Human1 (Human2), of all 1671 images, 16 % (2 %) are positively validated as tephra, 37 % (56 %) are dubious, and 47 % (41 %) are not considered tephra (Fig. S4). Of all the AI-predicted tephras, Human1 therefore considers 53 % of them are possible tephras (yes + maybe), while Human2 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 (Figs. 7, S5). It is worth noting that some tephra shards are positively validated even in those samples for which no tephra was previously found using optical microscopy. This is possibly related to the fact that the network detection accuracy does not vary with concentration, whereas the human eye is probably more trained to recognize particles if their number exceeds a certain threshold. Further analyses would be needed to quantitatively support this hypothesis. However, according to both analysts, the tephra modeled predictions include a number of minerals, such as feldspar and quartz, and a few contamination particles. Minerals, closely resembling tephra grains, 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 for differentiating tephra from minerals.

The source of minerals inside the FlowCam-measured samples can be 2-fold; 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 contamination in the GRIP samples ($n = 89\,329$), ~ 900 false positives (out of the n=1671 tephra predictions) could be misclassified as tephras. This further advocates the need for measuring clean samples in future studies.

Meltwater from the 12 samples run through the FlowCam was subsequently collected and then mounted in epoxy for tephra identification using optical microscopy and the methodology outlined in Sect. 3.3.1. This was required to verify that replicate samples were consistent with those of Cook et al. (2022). Despite some potential sample loss through the syringe pump, we found that samples were consistent, and tephra grains, consistent with either basaltic or rhyolitic grains, were present in seven samples and absent in five others.

3.3.4 Investigating the network dynamics

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 128×128 image becomes compressed into a 64-d vector representation. We inspect such a 64-d 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 64-d representations and let UMAP learn a 2-d embedding space of the data (Fig. 8). In such a representation the embedded data appear clustered according to their respective classes, with a few items being misplaced (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 in distinguishing these types of pollen. Overall, the high degree of separation between the training items is well reflected in the 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 64-d vectors are extracted and reprojected onto the learned 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 occasionally be 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 become 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

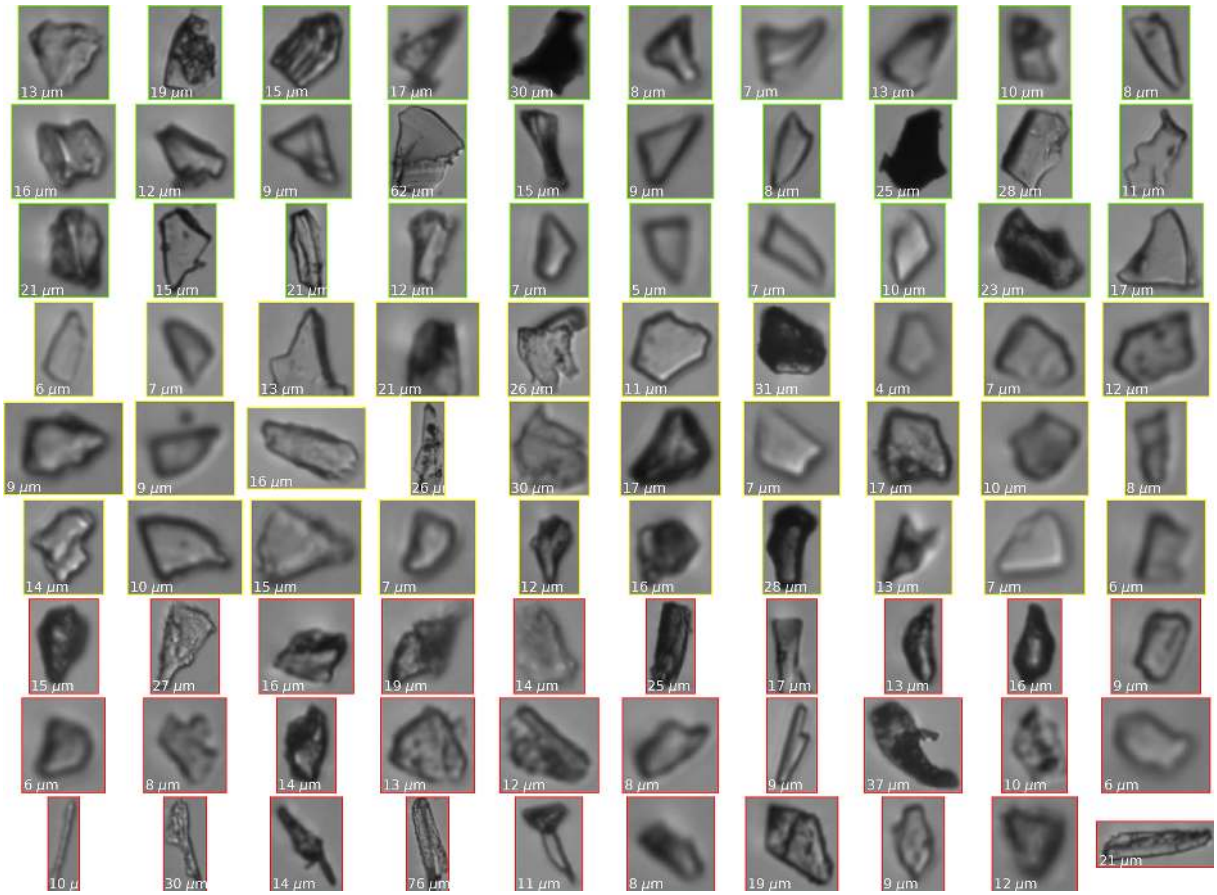

Figure 7. A random subset of the AI-predicted tephras in the 3136 0–20 cm GRIP sample assessed by Human1, color-coded according to the given validation of yes (green), maybe (yellow), or no (red). The particle diameters (ABD) are shown in the bottom-left corners.

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-like particles likely from glove fabrics. Blurry images are also present in this class (as the model was trained to do so), and 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 $\geq 10\,\mu m$. These two features are consistent with the pollen training images, probably leading to such a 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 seven classes of particles, including 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 $\sim 300\,000$ images per second 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.

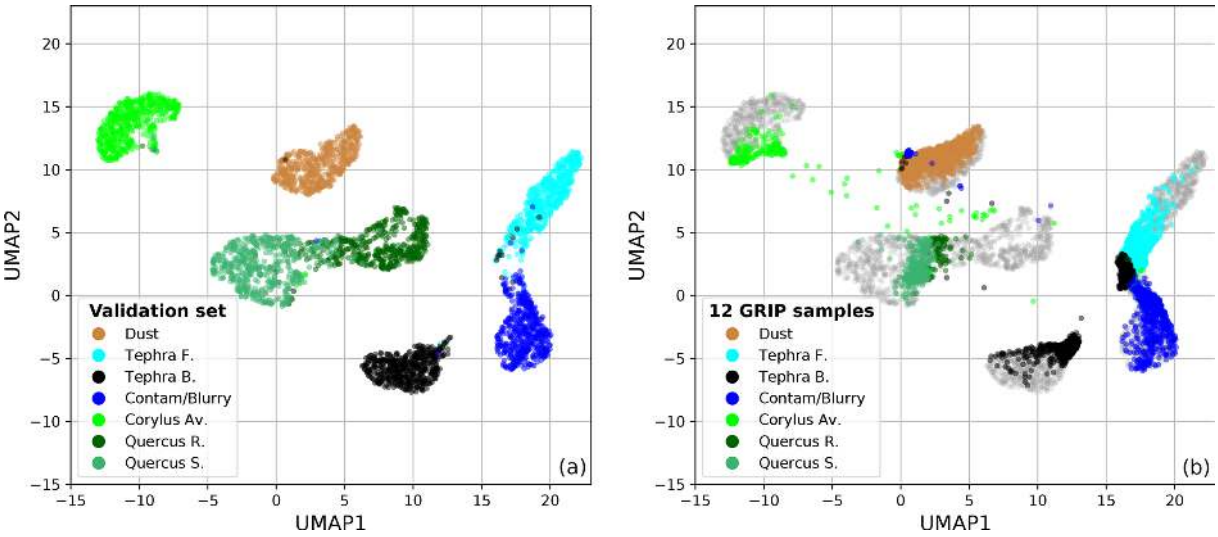

Figure 8. UMAP 2-d visualization of the network 64-d layer of the CNN branch. In panel **(a)**, UMAP is run on the validation dataset. In panel **(b)**, the learned UMAP space is used to project all images of the $n = 12$ GRIP samples. The items are color-coded according to their predicted class. Gray items represent the validation items.

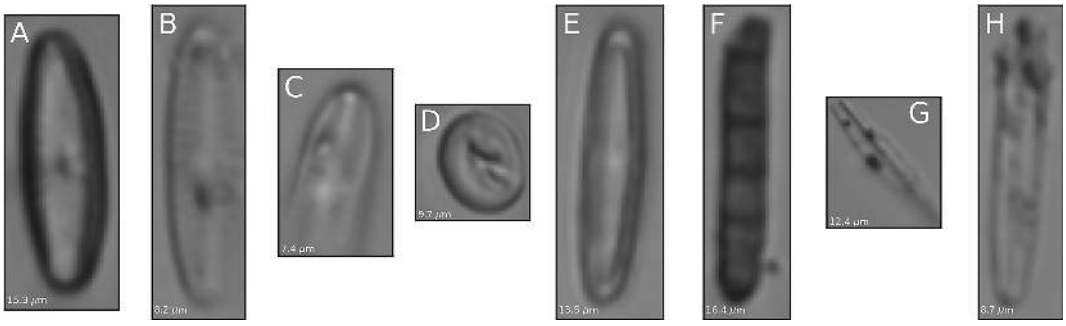

Figure 9. Diatoms identified in the Quelccaya Ice Core from the acquired FlowCam images. Particle D can be a Centrales diatom (possibly *Cyclotella* genus) or an algae. Particle F can possibly be a fungus. All other particles are Pennales diatoms. The particle diameters (ABD) are indicated in the bottom left corners. The presence of diatoms in this ice record has been previously reported, using SEM, 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 ($< 10\,\mu m$) dust particles within 1σ of the certified values. The mass concentrations can be replicated within 1 % over a range from a 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 $= 0.86 \pm 0.16$) only for particles smaller than $\sim 10\,\mu m$. The FlowCam exhibits a drop in efficiency in detecting larger particles that can lead to an underestimated mass concentration of up to a factor of 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 it 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, then *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 core samples, containing known tephra deposits, for a total of over 3×10^6 images. Almost the entirety of the images is classified as either dust or contamination/blurry particles, with the latter from the external core surface. A total of 1671 particles are classified as tephra (either felsic or basaltic). Inspection of such images by two tephra experts suggests that

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

Particle class	Advantages	Disadvantages	Suggested upgrades
Polar dust ($< 10\,\mu m$)	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 $< 10\,\mu m$.	Underestimation of the $> 10\,\mu m$ fraction.	Higher-volume cell.
Tephra (volcanic glass)	Can support human-operated bench microscopy. Autonomous, no sample preprocessing, and CFA ready.	Limited to fraction $> 8\,\mu m$. Underestimation of particles $> 10\,\mu m$. Low statistics. Image quality.	Higher-volume cell. Polarizer.
Pollen	Can support human-operated bench microscopy. Autonomous, no sample preprocessing, and CFA ready.	Underestimation of particles $> 10\,\mu m$. Low statistics. Image quality.	Required training datasets tailored to the ice core site. Higher-volume cell.
New particles (e.g., diatoms)	Autonomous and no sample preprocessing. Easy to implement by adding training datasets.	Underestimation of particles $> 10\,\mu m$. Low statistics.	Requires specific training datasets. Higher-volume cell.

only up to $\sim 50\,\%$ are possible tephra, with the remaining $\sim 50\,\%$ 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 in Cook et al. (2022).

Building on this work, we envision promising avenues for further research and upgrades in two main fields, namely 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 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 in 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 camera with a faster shutter rate). Importantly, a polarizer would be key to separating tephra from anisotropic minerals. An improved system should be ideally tested and deployed within a CFA workflow, targeting continuous particle records from ice cores.

Appendix A: Segmentation of particle images and outflow recovery

The instrument is equipped with a syringe pump (in our case with a 1.0 mL volume) placed downstream of the flow cell. The syringe pump draws sample fluid until its volume is filled, and then discharges it through an outlet tubing. In such a configuration, the sample outflow can be collected via the outflow tubing while the pump is being discharged. Such a collection, however, would integrate 1.0 mL of sample volume, which is not ideal if a fraction of the sample is needed. Additionally, there is no instrument-continuous outflow while the 1.0 mL pump volume is being filled, which is not compatible with continuous flow analysis setups.

We therefore suggest replacing the default syringe pump with a peristaltic pump, which ensures (i) a continuous flow and (ii) full control of the sample outflow collection via, e.g., a valve switch connected to the outflow tubing. If used within a CFA setup, the instrument inflow would simply require replacing the default discrete-mode pipette tip with a tubing connecting the instrument placed upstream with the FlowCam flow cell inlet.

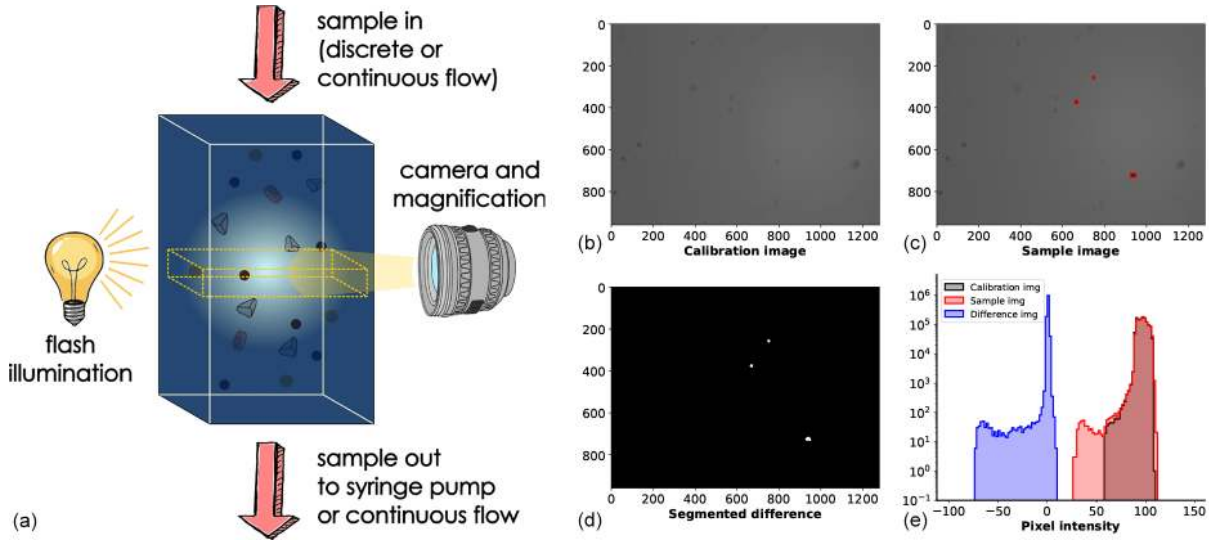

Figure A1. The experimental setup **(a)** is presented alongside the segmentation of particle images **(b–e)**. A calibration image of the camera view is obtained prior to the analysis when no sample is pumped into the system **(b)**. During the analysis, each frame **(c)** is compared to the calibration image, and a pixel-by-pixel difference is calculated **(e)** and thresholded to extract the single particle images. This procedure is performed by the FlowCam software.

Appendix B: Metadata

Table B1. Metadata.

Feature	Explanation
(1) Area	Number of pixels in the thresholded (binary) grayscale 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) = area. However, if parts of the particle are transparent, and therefore do not threshold as particle, then area (filled) > area.
(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) = $(\pi/4) \times$ geodesic thickness$^2 \times$ geodesic length.
(5) Biovolume (p. spheroid)	Biovolume (spheroid) = $(\pi/6) \times$ Legendre minor$^2 \times$ 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 is $(4 \times \pi \times \text{area})/\text{perimeter}^2$. (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]; 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 is $\text{perimeter}^2/(4\pi \times \text{area})$; (real ≥ 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, area-based diameter)	The diameter based on a circle with an area that is equal to the area (feature no. 1; real > 0).
(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-detecting convolution filter has been applied to the raw camera image (real [0, 255]).
(15) Elongation	The inverse of geodesic aspect ratio (real ≥ 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 is (geodesic length / length)−1; (real ≥ 0).

Table B1. Continued.

Feature	Explanation
(19) Fiber straightness	A shape parameter computed from Geodesic length and length. Formula is length / geodesic length (real ≥ 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 the length and thickness by solving the equations of area = geodesic length × geodesic thickness perimeter = 2 × (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 to the ratio of perimeter to convex perimeter (real ≥ 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 ≥ 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]).
(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).

Appendix C: Training dataset images

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

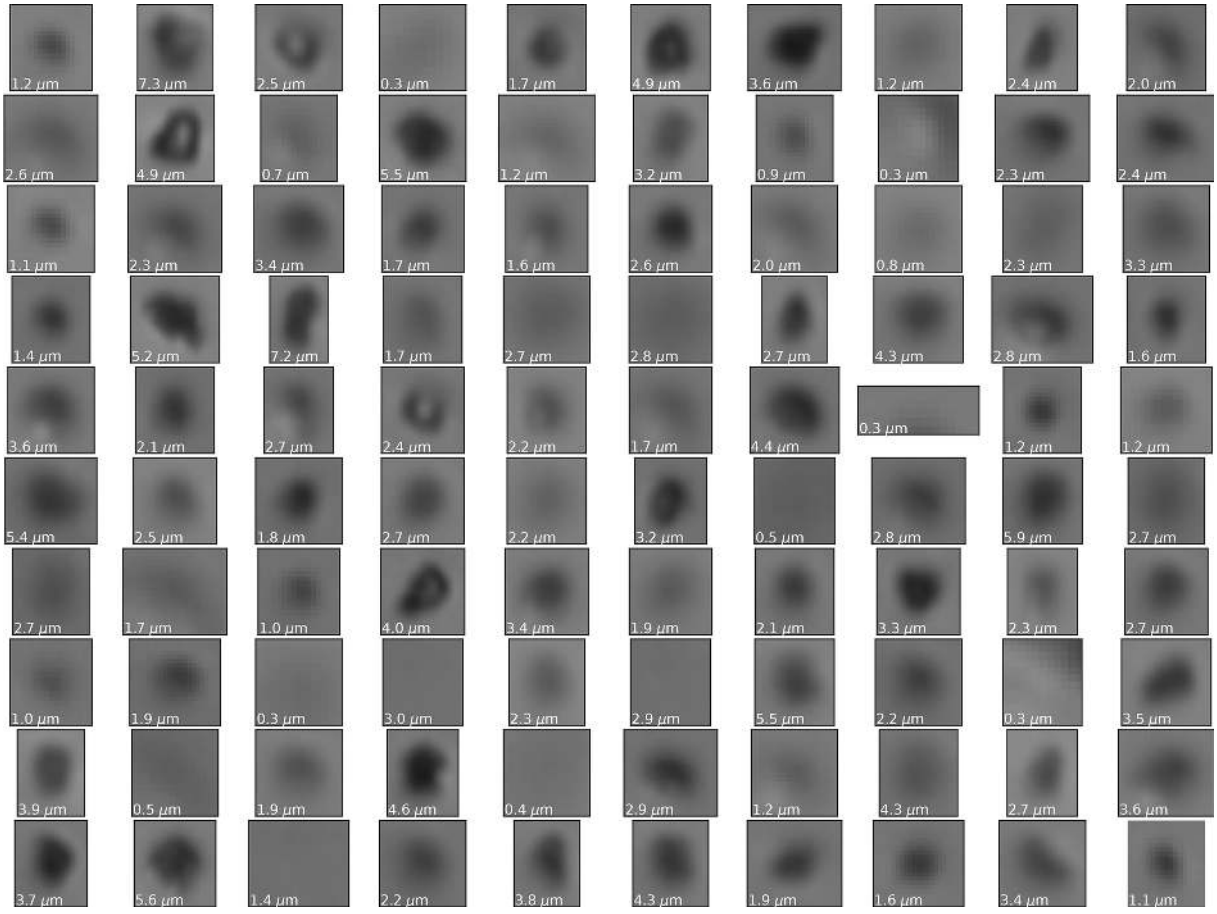

Figure C1. Dust.

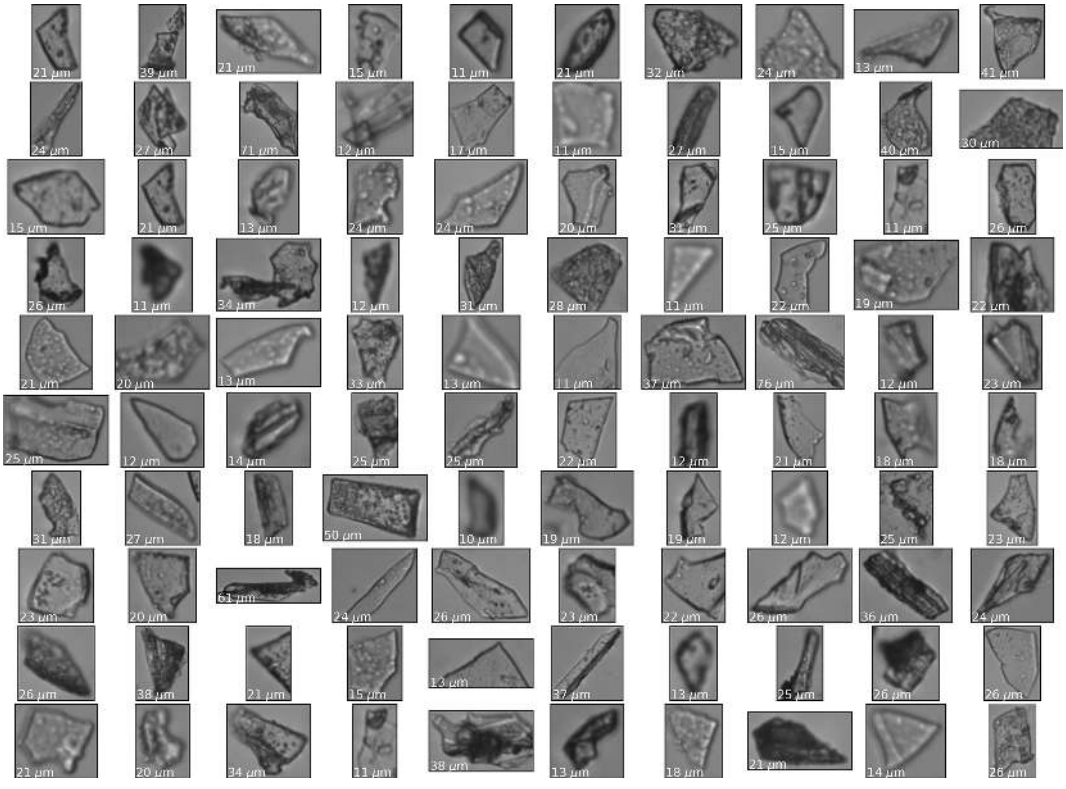

Figure C2. Felsic tephra.

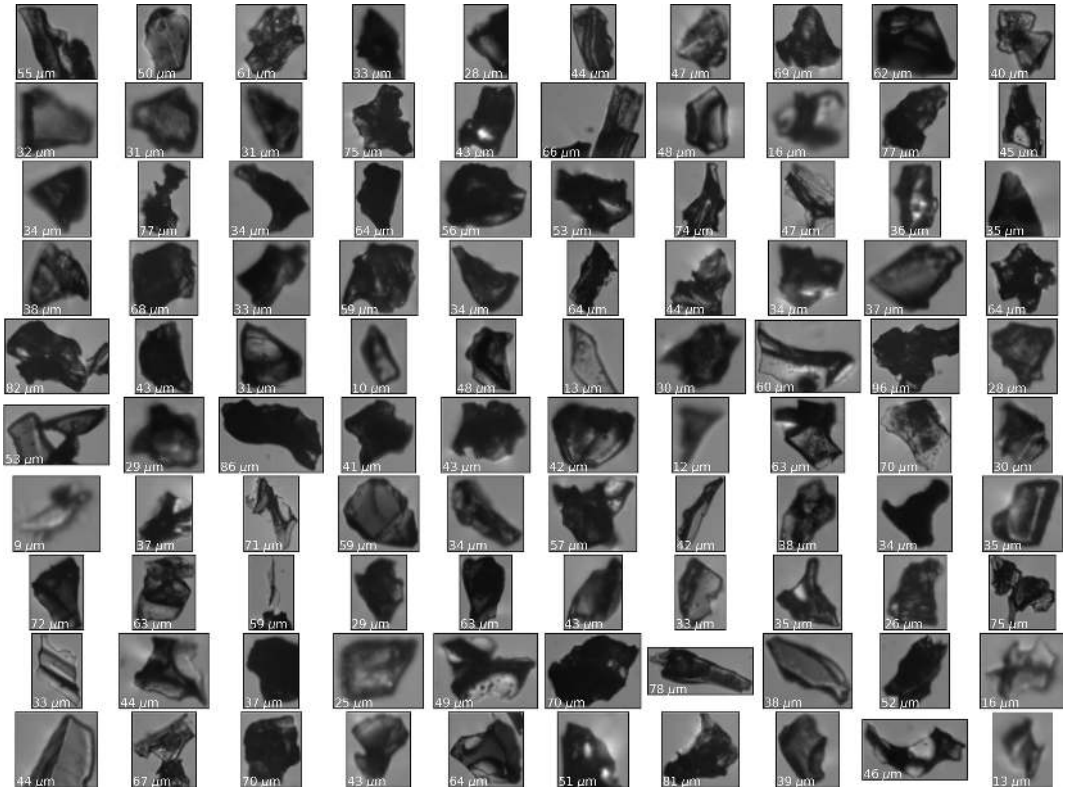

Figure C3. Basaltic tephra.

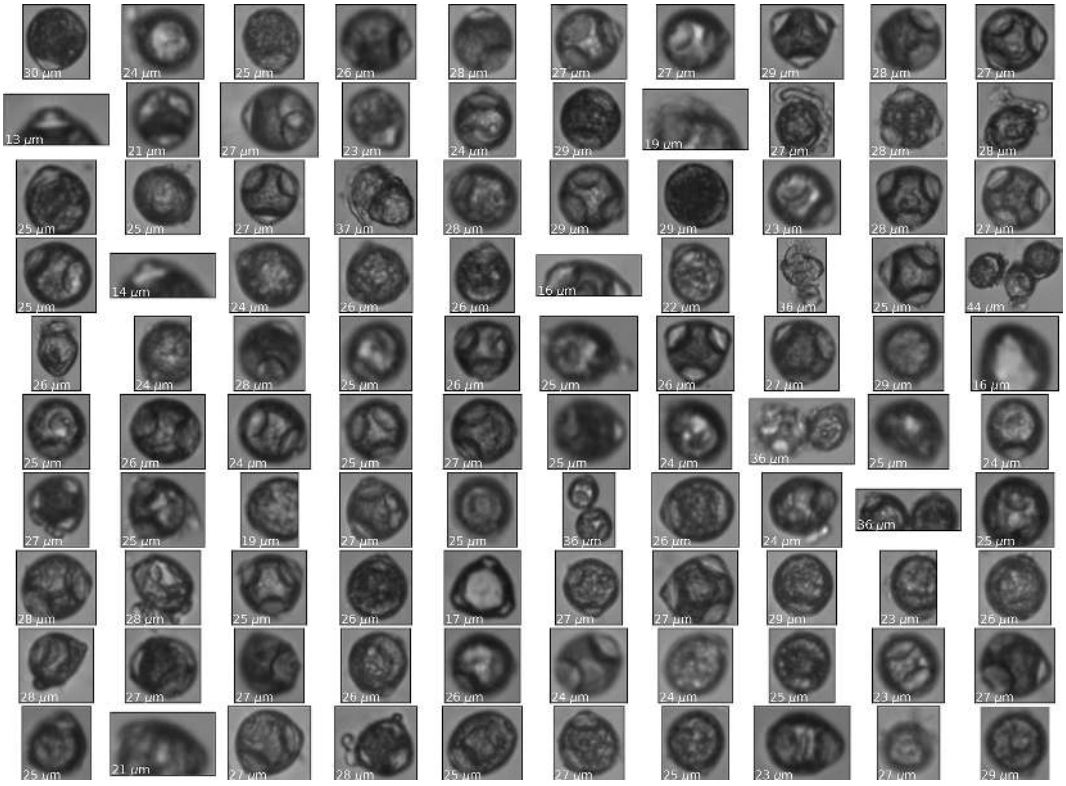

Figure C4. *Corylus avellana* pollen.

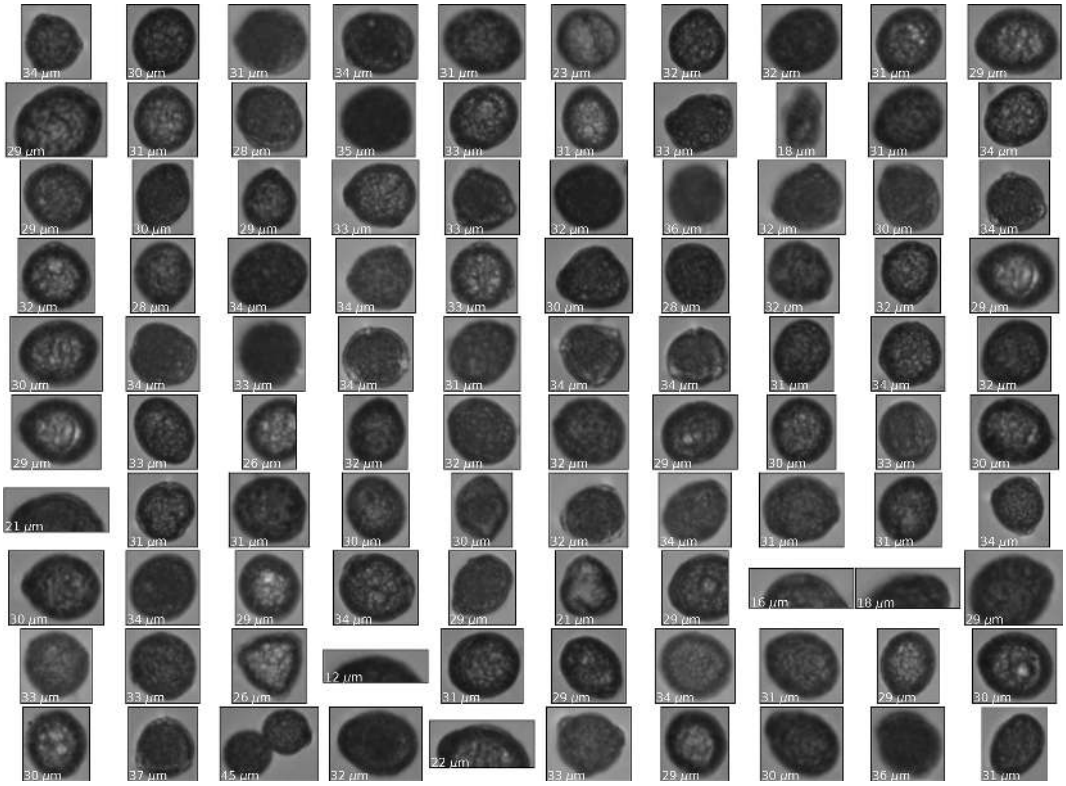

Figure C5. *Quercus robur* pollen.

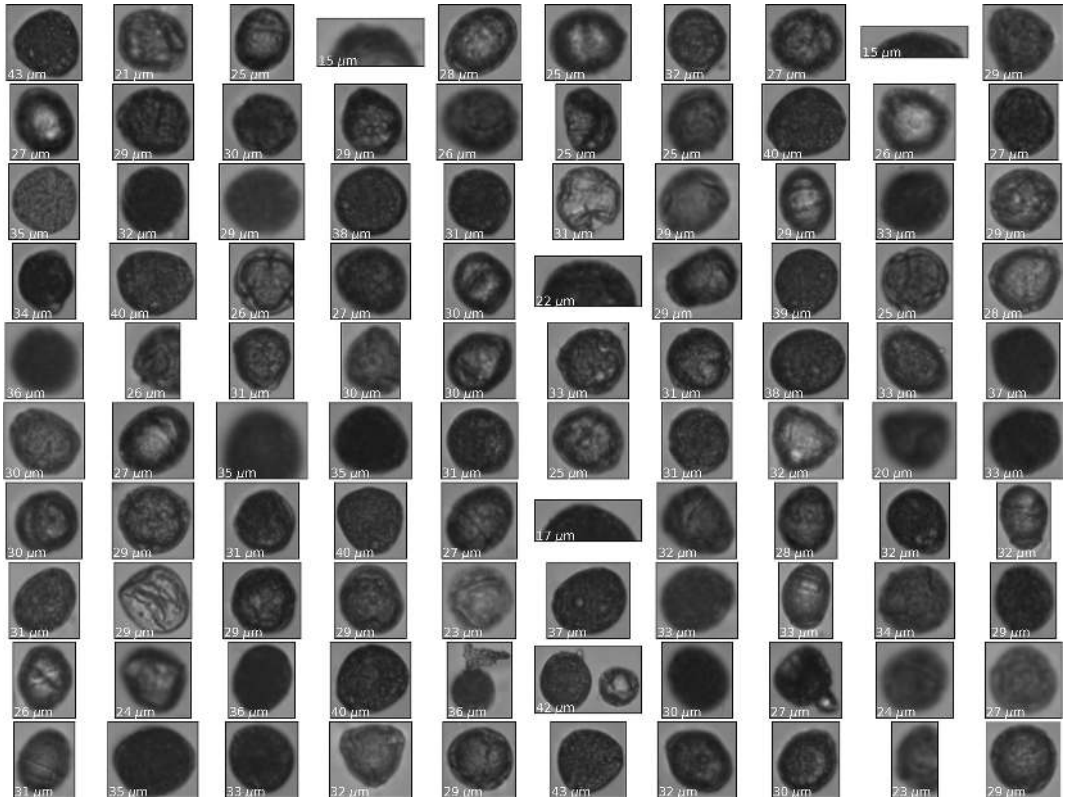

Figure C6. *Quercus suber* pollen.

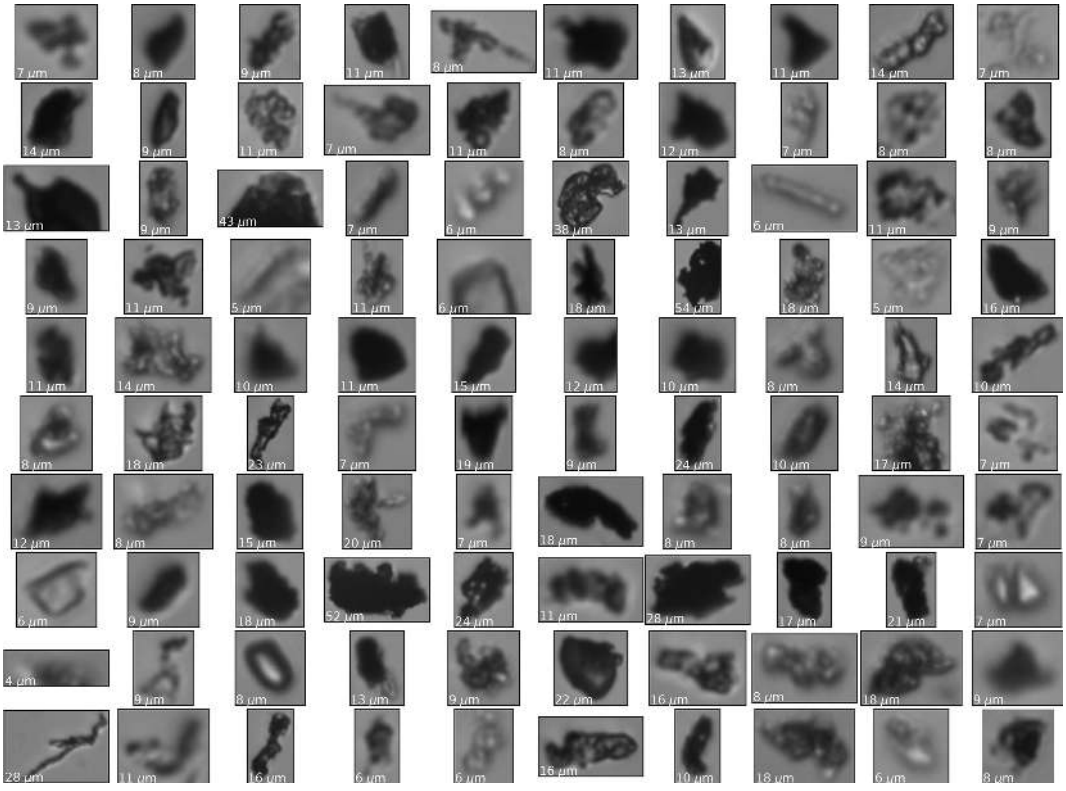

Figure C7. Contamination/blurry particles.

Code and data availability. The training and GRIP datasets will be deposited on Zenodo (https://doi.org/10.5281/zenodo.7591282, Maffezzoli, 2023a). The code will be made publicly available at https://github.com/nmaffe/icelearning (last access: 1 February 2023) and referenced on Zenodo (https://doi.org/10.5281/zenodo.7591227, Maffezzoli, 2023b).

Supplement. The supplement related to this article is available online at: https://doi.org/10.5194/tc-17-539-2023-supplement.

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, and YR. The samples were provided by WvdB, EC, DF, FM, AR, GB, AS, FdR, JS, BD, MV, JPS, and 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, and EC wrote the paper, using feedback from all other co-authors.

Competing interests. At least one of the (co-)authors is a member of the editorial board of *The Cryosphere*. The peer-review process was guided by an independent editor, and the authors also have no other competing interests to declare.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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

Financial support. This research has been carried out during the ICELEARNING project, supported by the European Union's Horizon 2020 Marie Skłodowska-Curie Actions (grant no. 845115).

Review statement. This paper was edited by Benjamin Smith and reviewed by two anonymous referees.

References

Abbott, P. M. and Davies, S. M.: Volcanism and the Greenland ice-cores: the tephra record, Earth-Sci. Rev., 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, https://doi.org/10.5194/tc-15-4807-2021, 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, Environ. Sci. Technol., 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, Clim. Past, 14, 21–37, https://doi.org/10.5194/cp-14-21-2018, 2018.

Bourgeois, J. C.: Seasonal and interannual pollen variability in snow layers of arctic ice caps, Revi. Palaeobo. Palyno., 108, 17–36, 2000.

Bourne, A. J., Cook, E., Abbott, P. M., Seierstad, I. K., Steffensen, J. P., Svensson, A., Fischer, H., Schüpbach, S., and Davies, S. M.: A tephra lattice for Greenland and a reconstruction of volcanic events spanning 25–45 ka b2k, Quaternary Sci. Rev., 118, 122–141, 2015.

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, e2021PA004293, https://doi.org/10.1029/2021PA004293, 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 Sci. Rev., 181, 200–206, 2018.

Cook, E., Abbott, P. M., Pearce, N. J., Mojtabavi, S., Svensson, A., Bourne, A. J., Rasmussen, S. O., Seierstad, I. K., Vinther, B. M., Harrison, J., Street, E., Steffensen, J. P., Wilhelms, F., and Davies, S. M.: Volcanism and the Greenland ice cores: A new tephrochronological framework for the last glacial-interglacial transition (LGIT) based on cryptotephra deposits in three ice cores, Quaternary Sci. Rev., 292, 107596, https://doi.org/10.1016/j.quascirev.2022.107596, 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 Planet. Sc. Lett., 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, Clim. Dynam., 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, 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Miami, 20–25 June 2009, 8, https://doi.org/10.1109/CVPR.2009.5206848, 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, Sci. Adv., 3, e1600446, 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), J. Glaciol., 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, Geophys. Monogr., 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, J. Glaciol., 61, 815–824, 2015.

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, https://doi.org/10.5194/tc-15-4135-2021, 2021.

Fritz, S. C., Brinson, B. E., Billups, W., and Thompson, L. G.: Diatoms at> 5000 meters in the Quelccaya Summit Dome Glacier, Peru, Arct. Antarct., Alp. Res., 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, J. Glaciol., 53, 585–596, 2007.

He, K., Zhang, X., Ren, S., and Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification, Proceedings of the IEEE International Conference on Computer Vision, 1026–1034, https://doi.org/10.1109/ICCV.2015.123, 2015.

He, K., Zhang, X., Ren, S., and Sun, J.: Deep residual learning for image recognition, Proc. CVPR IEEE, 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 J. Geo-Mar. Sci., 46, 1261–1264, 2017.

Kerr, T., Clark, J. R., Fileman, E. S., Widdicombe, C. E., and Pugeault, N.: Collaborative deep learning models to handle class imbalance in flowcam plankton imagery, IEEE Access, 8, 170013–170032, 2020.

Krizhevsky, A., Sutskever, I., and Hinton, G. E.: Imagenet classification with deep convolutional neural networks, Adv. Neur. In., 25, https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf (last access: 1 February 2023), 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, J. Sea Res., 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, Rev. Geophys., 35, 219–243, 1997.

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 Technol., 343, 652–661, 2019.

Loshchilov, I. and Hutter, F.: Decoupled weight decay regularization, arXiv [preprint], https://doi.org/10.48550/arXiv.1711.05101, 2017.

Lowe, D. J.: Tephrochronology and its application: a review, Quat. Geochronol., 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.

Maffezzoli, N.: ICELEARNING – Datasets, Zenodo [data set], https://doi.org/10.5281/zenodo.7591282, 2023a.

Maffezzoli, N.: nmaffe/icelearning: v0.1.0 pre-release (v0.1.0), Zenodo [code], https://doi.org/10.5281/zenodo.7591227, 2023b.

McInnes, L., Healy, J., and Melville, J.: Umap: Uniform manifold approximation and projection for dimension reduction, arXiv [preprint], https://doi.org/10.48550/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, J. Biogeogr., 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, J. Geophys. Res.-Earth, 109, F04001, https://doi.org/10.1029/2004JF000125, 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 Sci. Rev., 49, 52–63, 2012.

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., Delmotte, M. Kotlyakov, V. M., Legrand, M., Lipenkov, V. Y., Lorius, C., PÉpin, L., Ritz, C., Saltzman, E., and Stievenard, M.: Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica, Nature, 399, 429–436, https://doi.org/10.1038/20859, 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 Res., 39, https://doi.org/10.33265/polar.v39.3511, 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, Front. Environ. Sci., 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, J. Geophys. Res.-Atmos., 108, 4098, https://doi.org/10.1029/2002JD002376, 2003.

Ruth, U., Barbante, C., Bigler, M., Delmonte, B., Fischer, H., Gabrielli, P., Gaspari, V., Kaufmann, P., Lambert, F., Maggi, V., Marino, F., Petit, J. R., Udisti, R., Wagenbach, D., Wegner, A., and Wolff, E. W.: Proxies and measurement techniques for min-

eral dust in Antarctic ice cores, Environ. Sci. Technol., 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, Springer, 95–119, https://doi.org/10.1007/1-4020-2146-1_6, 2004.

Sigl, M., Winstrup, M., McConnell, J. R., Welten, K. C., Plunkett, G., Ludlow, F., Büntgen, U., Caffee, M., Chellman, N., Dahl-Jensen, D., Fischer, H., Kipfstuhl, S., Kostick, C., Maselli, O. J., Mekhaldi, F., Mulvaney, R., Muscheler, R., Pasteris, D. R., Pilcher, J. R., Salzer, M., Schüpbach, S., Steffensen, J. P., Vinther, B. M., and Woodruff, T. E.: 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, Clim. Past, 14, 601–608, https://doi.org/10.5194/cp-14-601-2018, 2018.

Simonyan, K. and Zisserman, A.: Very deep convolutional networks for large-scale image recognition, arXiv [preprint], https://doi.org/10.48550/arXiv.1409.1556, 2014.

Turney, C. S., Harkness, D. D., and Lowe, J. J.: The use of microtephra horizons to correlate Late-glacial lake sediment successions in Scotland, J. Quaternary Sci., 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, Ann. Glaciol., 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, Front. Earth Sci., 8, https://doi.org/10.3389/feart.2020.622386, 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, Quatern. Int., 293, 22–33, 2013.

Viertel, P. and König, M.: Pattern recognition methodologies for pollen grain image classification: a survey, Mach. Vision Appl., 33, 1–19, 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, J. Geophys. Res.-Atmos., 100, 16249–16263, 1995.

Zeiler, M. D. and Fergus, R.: Visualizing and understanding convolutional networks, in: European conference on computer vision, ECCV 2014: 13th European Conference, Zurich, Switzerland, 6–12 September 2014, Springer, 818–833, https://doi.org/10.1007/978-3-319-10590-1_53, 2014.

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