SMAUG v1.0 – a user-friendly muon simulator for ~~transmission tomographythe imaging~~ of geological objects in 3D

Alessandro Lechmann[1], David Mair[1], Akitaka Ariga[2], Tomoko Ariga[3], Antonio Ereditato[2], Ryuichi Nishiyama[4], Ciro Pistillo[2], Paola Scampoli[2,5], Mykhailo Vladymyrov[2] and Fritz Schlunegger[1]

[1]Institute of Geological Sciences, University of Bern, Bern, CH-3012, Switzerland
[2]Albert Einstein Center for Fundamental Physics, Laboratory for High Energy Physics, University of Bern, Bern, CH-3012, Switzerland
[3]Faculty of Arts and Science, Kyushu University, Fukuoka, JP-819-0385, Japan
[4]Earthquake Research Institute, The University of Tokyo, Tokyo, JP-113-0032, Japan
[5]Physics Department "Ettore Pancini", University of Naples Federico II, Naples, IT-80126, Italy

Correspondence to: Alessandro Lechmann (alessandro.lechmann@geo.unibe.ch)

Abstract.

Knowledge about muon tomography has spread in recent years in the geoscientific community and several collaborations between geologists and physicists have been founded. As the data analysis is still mostly done by particle physicists, ~~we address the need~~much of the ~~geoscientific community to participate~~know-how is concentrated in ~~the data analysis, while not having to worry too much about the~~ particle physics ~~equations in the background. The result hereof is~~and specialised geophysics institutes. SMAUG, a toolbox consisting of several modules that cover the various aspects of data analysis in a muon tomographic experiment~~.~~, aims at providing access to a structured data analysis framework. The goal of this contribution is to make muon tomography more accessible to a broader geoscientific audience. In this study we show how a comprehensive geophysical model can be built from basic physics equations. The emerging uncertainties are dealt with by a probabilistic formulation of the inverse problem, which is finally solved by a Monte Carlo Markov Chain algorithm. Finally, we benchmark the SMAUG results against those of a recent study, which however, have been established with an approach that is not easily accessible to the geoscientific community. We show that they reach identical results with the same level of accuracy and precision.

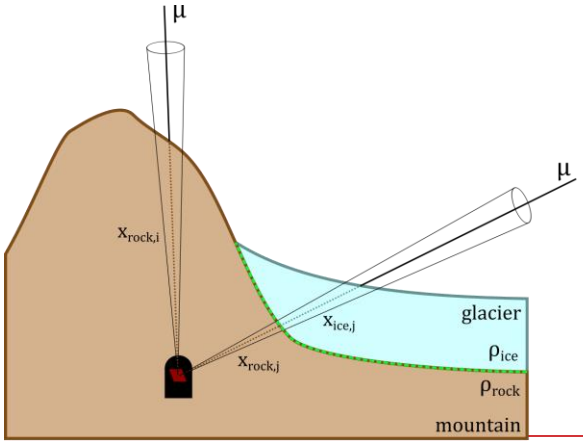
1 Introduction

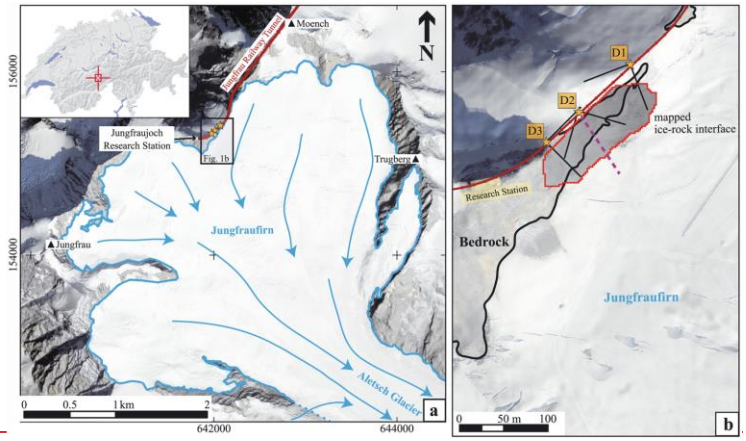
Among the manifold geophysical imaging techniques, muon tomography has increasingly gained the interest of geoscientists during the course of the past years. Before its application in Earth sciences, it was initially used for archaeological purposes. Alvarez et al. (1970) used this method to search for hidden chambers in the pyramids of Giza, in Egypt; an experiment which was recently repeated by Morishima et al. (2017), as better technologies have continuously been developed. Other civil engineering applications include the monitoring of nuclear power plant operations (Takamatsu et al., 2015) and the search

for nuclear waste repositories (Jonkmans et al., 2013) as well as the investigation of underground tunnels (e.g. Thompson et al., 2020; Guardincerri et al., 2017). A serious deployment of muon tomography in Earth sciences has only begun in the past decades. These undertakings mainly encompass the study of the interior of volcanoes in France (Ambrosino et al., 2015; Jourde et al., 2016; Noli et al., 2017; Rosas-Carbajal et al., 2017), Italy (Ambrosino et al., 2014; Lo Presti et al., 2018; Tioukov et al., 2017), and Japan (Kusagaya and Tanaka, 2015; Nishiyama et al., 2014; Oláh et al., 2018; Tanaka, 2016). Other experiments have been performed in order to explore the geometry of karst cavities in Hungary (Barnaföldi et al., 2012) and Italy (Saracino et al., 2017). Further studies (see also review article by Lechmann et al., 2021a) have been conducted by our group to recover the ice-bedrock interface of Alpine glaciers in central Switzerland (Nishiyama et al., 2017; 2019).

The core component of every geophysical exploration experiment is formed by the inversion, which might be better known to other communities as fitting or modelling. This is where the model parameters are found, which best fit the observed data. Up until now, this central part has mostly been built specifically to meet the needs of the experimental campaign at hand. On the one hand this approach has the advantage of allowing the consideration of the peculiarities of particle detectors, their data processing chain, and other models involved (e.g. the cosmic ray flux model). On the other hand, when every group develops a separate inversion algorithm, the reconstruction of the precise calculations performed in the data analysis procedure becomes a challenge. For a researcher who is not familiar with the intricacies of inversion, this might even be tougher. We thus see the need for a lightweight programme that incorporates a structured and modular approach to inversion, that also allows users with little inversion experience to familiarise themselves with this rather involved topic. This programme can be used to directly analyse experimental data in a stand-alone working environment, and the modules and theoretical foundations can be adapted, customised, and integrated into new programmes. For this reason, the code is built in the programming language Python as to facilitate the exchange between researchers and to enhance modifiability. Moreover, the source code is freely available online (Lechmann et al., 2021b).

To facilitate the further reading of our code, we introduce the reader at this point to our benchmark experiment, to which we will refer on multiple occasions throughout this work. The experimental campaign is explained in detail in Nishiyama et al. (2017) and thus we will resort to a description of the experimental design at this point. In the Nishiyama et al. (2017) study ~~we aimed at~~ we aimed at recovering the ice-bedrock interface of an Alpine glacier in central Switzerland. Figure 1 shows that we had access to the railway tunnel of the Jungfraubahnen railway company, where we installed three detectors. In our measurement we recorded muons from directions that consisted purely of rock and others where we knew that the muons must have crossed rock and ice (see the two cones in Fig. 1). From the former it was possible, together with laboratory measurements, to determine the physical parameters of the rock more precisely. ~~Conversely~~Subsequently, we utilised the ~~measurement~~directional measurements of the latter to infer the 3D structure of the interface between rock and ice underneath the glacier. Finally, we will also use the results of that experiment (Nishiyama et al., 2017) to verify our new algorithm in the present study.


μ
μ
$x_{rock,i}$
$x_{ice,j}$
$x_{rock,j}$
glacier
ρ_{ice}
ρ_{rock}
mountain


Moench
Jungfrau Railway Tunnel
Jungfraujoch
Research Station
Fig. 1b
Trugberg
N
Jungfraufirn
Jungfrau
Aletsch Glacier
0
0.5
1 km
2
a
642000
644000
156000
154000
D1
D2
D3
mapped
ice-rock interface
Research Station
Bedrock
Jungfraufirn
0
50 m
100
b

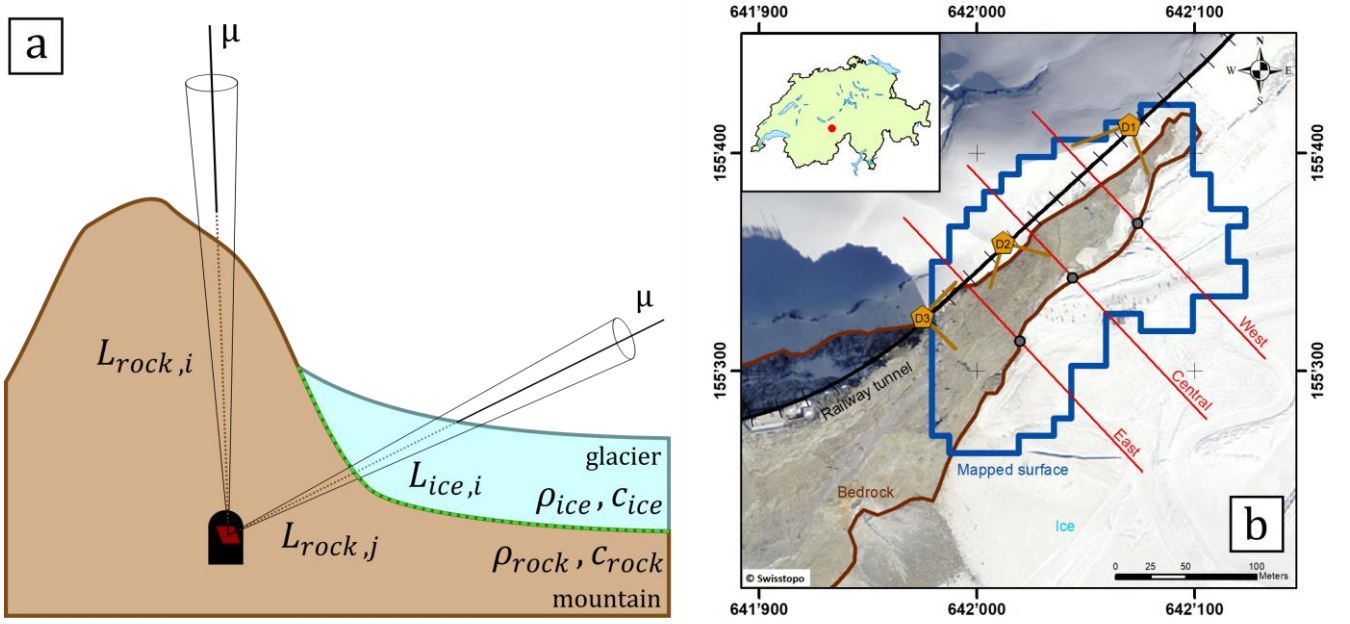


Figure 1: ~~***Left:****a)*~~ **Schematic side view of the experiment from Nishiyama et al. (2017). A muon detector (red) is placed in a tunnel or cavity and records muons from the cosmic ray flux that penetrate the material (ice & rock) from different directions. Muons that are detected along cones yield information on the amount and density of matter between the topographic surface and the detector. Based on this the interface between ice and rock (dashed green line) can be reconstructed.** ~~***Right:****b)*~~ **Modified Fig. 1 from Nishiyama et al. (2017); Overview of the study region in the central Swiss Alps.** ~~**a) shows an outline of**~~ **Detectors are indicated by yellow pentagrams (D1, D2, D3; including their viewfield) within the railway tunnel (black line). The imaged** ~~**glacier**~~**area is outlined in blue and** ~~**its localisation in Switzerland and b) indicates where**~~ **the region that we know from the digital elevation model to contain only rock is marked in brown. We verified (see Sect. 4) the reconstruction in this study with the one from Nishiyama et al. (2017) along three** ~~**detectors have been positioned in the Jungfraubahnen railway tunnel**~~**cross-sections (East, Central, West; red lines). Basemap: Orthophotomosaic Swissimage, © Federal Office of Topography swisstopo.**

1.1 Inversion – a modular view

The goal of every muon tomography study is essentially to extract information on the physical parameters (usually density and/or the thickness of a part of the material) of the radiographed object through a measurement of the cosmic ray muon flux and an assessment of its absorption as the muons cross that object. In geological applications these objects are almost always lithological underground structures such as magma chambers, cavities, or other interfaces with a high-density contrast. The reconstruction of the geometry of such structures can only be achieved if the measured muon data is compared to the results of a muon flux simulation. As stated earlier, this is the basic principle of the inversion procedure. However, the aforementioned "muon flux simulation" is not just a simple programme, but it consists of several physically independent models that act together. Taking a modular view, we will call these models "modules" from here on, as they will inevitably be part of a larger inversion code. We have visualised the components that are necessary to build an inversion and how they interact with each other in Fig. 2.

The first of the modules is the input module for the experiment results, which also considers the detectors that were used in the experiment. Typical detector setups include nuclear emulsion films (e.g. Ariga et al., ~~(~~2018), cathode chambers (e.g. Oláh et al., 2013), scintillators (e.g. Anghel et al., 2015) or other hardware solutions. Although the detailed data processing chain may be comprehensive, the related output almost always comes in the form of a measured directional (i.e. from various incident angles) muon flux or equivalently the measured directional number of muons, which will be the input to the inversion scheme. Here, we primarily work with the premise that the muon flux data and the associated errors are given. The corresponding errors can then be furnished to the code by means of an interface.

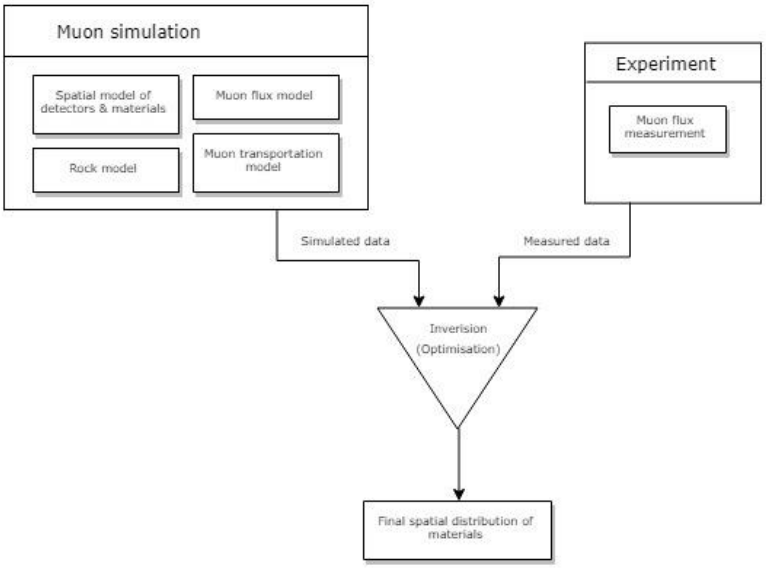
The simulation module on the other hand, consists of ~~four autonomous components~~two parts each containing two modules (see Fig. 2). ~~First, a cosmic ray muon flux~~The model parametrisation is ~~necessary, which describes~~needed in order to abstract the ~~muon abundance in the atmosphere~~ (geo)physical reality as a mathematical model. Subsequently, the forward model uses the model parametrisation and ~~which is generally dependent~~simulates an artificial dataset based on the ~~muon energy,~~ chosen parameter values.

As we want to draw inferences on the physical parameters of the involved materials, we need a "rock model" first. The latter term is used in an earlier publication (Lechmann et al., 2018) where we split a rock into its ~~incidence angle,~~mineral constituents and ~~the altitude of the detector location. Lesparre et al. (2010) list~~compute a mean composition and ~~compare various muon flux~~ a mean density needed for further calculations. Even though this is called a "rock model" (Fig. 2) the approach can be used for other materials (e.g. ice) as well. It is also possible to infuse laboratory measurements of compositions and density into this family of models ~~that may be incorporated into an extensive simulation. Second~~.

Once the materials have been described, it is necessary to model the ~~spatial distribution of the detectors as well as the initial distribution of the lithologies. Related~~experimental situation spatially. This means that materials as well as detectors have to be attributed a location in space. The central choice in the "model parametrisation" is usually to select how the material parameters are discretised spatially (this is hinted in Fig. 2 with the term "binning"). We refer the reader to Sect. 2.1 for further information on that topic. In order to conduct this parametrisation we may employ pre-existing software solutions that mainly ~~comprise~~compromise GIS- and geological 3D-modelling applications~~,~~ that excel at capturing ~~and compiling~~ geological information from various sources ~~(~~, e.g. digital elevation models (DEMs), and that also allow to compile field~~,~~ observations (maps, etc.) into a spatially organised database. ~~Third,~~Once the ~~lithologies consisting~~structure of ~~different minerals have to be translated~~the spatial model is determined, we need a physical model, that allows us to ~~a~~ calculate a synthetic dataset, based on the parameter values that we set ~~of parameters, which are a necessary input~~up. We note that the parametrisation structure remains fixed for the ~~subsequent physical~~ time of the calculations, and that changes are only performed on the parameter values.

Incident muons, on their way from the atmosphere to the detector, lose energy while traversing matter. The first step in the muon simulation~~. This can be done by a rock model (e.g. Lechmann et al., 2018), which considers the effects~~ is then to determine how much the initial energy of the ~~mass density as well as the average atomic mass and charge of the rock as a function of its mineralogical composition. Lastly, the muon fluxes at~~muon has to be, in order to be able to penetrate all the

material up to the detector~~ sites have to be simulated~~. This is done by means of a muon transportation model, which calculates all physical processes by which a muon loses kinetic energy while ~~travelling~~traveling through matter. The particle physics community has a great variety of particle simulators, the most prominent being GEANT4 (Agostinelli et al., 2003), a Monte Carlo based simulator. These have the advantage that stochastic processes resulting in energy loss are simulated according to their probabilistic occurrence - an upside that has to be traded off for longer computation times. In contrast to obtaining the full energy loss distribution, lightweight alternatives often resort to the calculation of only the mean energy loss. The solution of the resulting differential equation can even be tabulated, as has been done by Groom et al. (2001). ~~The interplay of these four submodules allows for the simulation of muon fluxes at the detector sites that are mostly located in an underground environment.~~



Lastly, based on that minimum energy, one may calculate the portion of the atmospheric muon flux that is fast enough to reach the detector. For this part a cosmic ray muon flux model is needed, which describes the muon abundance in the atmosphere, and which is generally dependent on the muon energy, its incidence angle, and the altitude of the detector location. Lesparre et al. (2010) list and compare various muon flux models that may be incorporated into an extensive simulation.

The interplay of these four modules (schematically shown in Fig. 2) allows then to simulate a dataset. It is then possible to compare the measured data with a synthetic dataset and to quantify this difference using a specific metric (usually the squared sum of residuals, which is also termed the misfit in Fig. 2). The process of changing the model parameters (here density of the materials and thicknesses of the segments in each cone), comparing the synthetic dataset with the measured dataset, and of iteratively tweaking the parameters in such a way that misfit is minimised is called “Inversion” in Fig. 2. The

solution of the inversion depends on the inversion method used, and is either the full statistical distribution of the model parameters or an estimate thereof (e.g. the Maximum Likelihood estimate).

~~**Figure 2: A schematic flowchart showing the different involved models in a muon tomographic experiment. The muon simulation consists of a model for rocks, detectors, the cosmic ray flux and a particle physical model on how muons lose energy upon travelling through rocks. These models allow for a synthetic data set to be computed, which will be compared to the actual measured data from the experiment. An optimisation problem then solves for the best set of parameters.**~~

1.2 The need for a consistent inversion environment

The ~~sole combination~~reconstruction of material parameters from muon flux data has already been performed in a variety of ways and different methods and codes have already been published. Bonecchi et al. (2015), for example, used a back-projection method such that size and location of underground objects can be determined. Jourde et al. (2015), on the ~~aforementioned four submodules does not fully justify~~other hand, describe the ~~need for a new software, as cosmic ray flux models as well as rock models can also be programmed within existing Monte Carlo simulators such as~~resolving kernel approach, where they show how muon flux data and gravimetric data can be combined to improve the resolution on the finally imaged 3D density structure. This is a useful approach especially in the planning stages of an experiment. Barnoud et al. (2019) provide a perspective on how such a joint inversion between muon flux data and gravimetric data can be combined in a Bayesian framework, whereas Lelièvre et al. (2019) investigate different methods of joining these datasets, using unstructured grids.

Existing frameworks that are especially used in physics communities are GEANT4 (Agostinelli et al., 2003) or MUSIC (Kudryavtsev, 2009). These are Monte Carlo simulators that excel at modelling how a particle (e.g. a muon) interacts with matter and propagates in space and time. The Monte Carlo aspect describes the fact that many particles are simulated to get a statistically viable distribution of different particle trajectories. This might be a very time-consuming process, as for each material distribution such a Monte Carlo simulation has to be performed, and only a fraction of all simulated muons actually hit the detector. In order to speed this calculation up Niess et al. (2018) devised a backward Monte-Carlo, which only simulates that portion of the muons that are actually observed.

In their area of use the above-mentioned sources prove very valuable. Unfortunately, ~~the application of such a Monte Carlo approach requires~~these tools and approaches have in common, that a rather good understanding of ~~programming and~~ either inversion, nuclear physics processes~~. Thus, it might prove time consuming to~~, or programming is required. Even more so if one wants to tackle our problem of interface detection, coupled with density inversion, then various parts of the above codes need to be linked together. The construction of a specialised programme ~~a specific code. Moreover, these codes are often~~is then a time-consuming process, as a) programmes might not be freely available, b) different codes might be written in ~~a~~different and specialised programming ~~language~~ languages (such as C~~++. Third,~~++), c) the compatibility between different modules ~~(e.g. cosmic ray flux and energy loss)~~ may be severely hampered, if the programme interfaces are not taken into consideration. ~~It might be even worse if the two modules are written in two different programming languages. In~~

~~addition~~Therefore, one has to carefully evaluate the benefit of such an undertaking, especially if the resulting code will most likely be tailored only to a specific problem.

We thus see the need for a versatile, user-friendly simulator, which allows users not only to quickly perform the necessary calculations without the need of additional coding, but also tailor the individual models to custom needs. A new simulator can be more useful if an inversion functionality is already included. As can be seen in Fig. 2, the inversion compares the simulated flux data with the measured ones. ~~It also attempts to reduce~~This problem is solved by finding the ~~discrepancy between measurements and simulations by optimising the~~ set of parameters ~~in the simulation, namely (~~material density and the ~~thickness distribution~~thicknesses of the overlying materials), that adhere to the constraints of the available a priori information and minimise the aforementioned discrepancy between measurement and simulation. This results in a density- or structural rock model, which best reproduces the measured data. As the energy loss equation in general is nonlinear, also the mathematical ~~optimization~~optimisation in muon tomography ~~generally is nonlinear, one has to employ~~is nonlinear. This is classically solved by either a linearisation of the physical equations or by employing nonlinear solvers~~ or even~~. A further difficulty is introduced when working in 3D. Monte Carlo techniques are, however, enough versatile to tackle these challenges, which is our main motivation for working with them. This circumstance encourages us to work with a lightweight version of a muon transport simulator, because a nonlinear inversion of Monte Carlo simulations, although mathematically preferable, is computationally prohibitive. This allows us to make use of methods from the Bayesian realm, that thrive when measurements from different sources (i.e. muon flux measurement, laboratory, geological field measurements, maps, etc.) have to be combined into a single comprehensive model. With the code presented in this paper, we aspire to make muon tomography accessible to a broader geoscientific community, as the know-how in this field is mainly concentrated in particle physics laboratories. We want to provide the tools for Earth scientists, or users that are mainly focused on the application of the method, so that they can perform their own analyses.

In this contribution we present our new code, SMAUG (Simulation for Muons and their Applications UnderGround), that allows a broader scientific community to plan and analyse muon tomographic experiments more easily, by providing them with data analysis and inversion tools. Specifically, we describe the governing equations of the physical models, and the mathematical techniques that were used. ~~Chapter~~Section 2 depicts how the muon flux simulation is conducted by its submodules and how a muon flux simulation is performed. ~~Chapter~~Section 3 then dives into the inversion module and explains how the parameters of the inferred density/rock-model can be estimated based on measured data. This ~~chapter~~section includes a description of the model and data errors and an explanation on how a subsurface material boundary can be constructed. ~~Chapter 4 provides a short overview of the program, explaining which functionality can be found in which source code. Chapter 5~~Section 4 discusses the model's performance based on the data that we collected in the framework of an earlier experimental campaign (see supplement of Nishiyama et al., 2017). ~~Chapter 6~~Section 5 then concludes this study by outlining a way of how this code can be developed further to fit the needs of the muon tomography and geology community.

~~2 The forward model: Muon flux~~In order to provide the reader with quick access to information about the vast number of variables that are used in this work, we refer to the Table 1, where all the parameters are listed and explained.

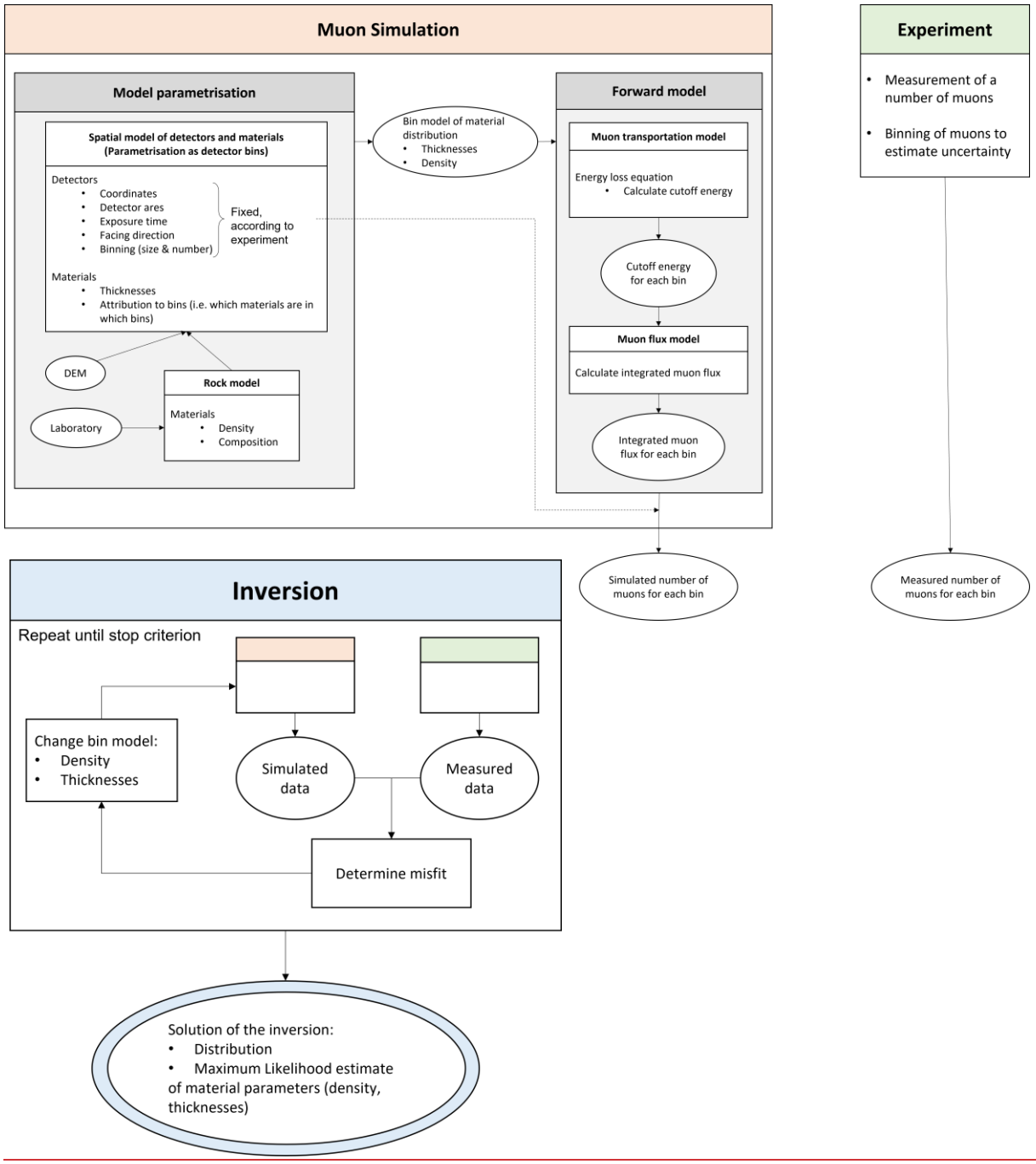


Figure 2: A schematic flowchart showing the different involved models in a muon tomographic experiment. The muon simulation consists of a model for rocks, detectors, the cosmic ray flux and a particle physical model on how muons lose energy upon travelling through rocks. These models allow for a synthetic data set to be computed. The systematic comparison between synthetic data and actual measured data and the subsequent change of the model parameters to find the set of parameters that reproduce the measured data best is an (often iterative) optimisation problem. This procedure is termed "inversion" (usually) by geophysicists.

Table 1: List of variables used in this work. Parameters are grouped into sections where they are introduced first. Parameters that are sought for muon tomography (in our case and also in general) are highlighted with an olive colour.

Variable	Unit	Description
Section 1		
m	1	Index that runs through all different materials, here $m \in \{rock, ice\}$
ρ_m	kg m^{-3}	Material density of the m-th material
c_m	1	Composition of the m-th material, given as a weight fraction of major rock forming oxides
i	1	Index that runs through all different cones/bins, here $i \in \{1, \dots, N_{cones}\}$
N_{cones}	1	Total number of cones/bins
L_{mi}	m	Thickness of the m-th material in the i-th bin
Section 2.1		
$N_{\mu,i}$	1	Number of simulated muons in the i-th bin
$E_{cut,i}$	eV	Cutoff energy in the i-th bin
$\frac{d\Phi}{dE}$	eV^{-1} s^{-1} m^{-2} sr^{-1}	Differential muon flux in the i-th bin
ΔT	s	Exposure time of the detector
$\Delta\Omega_i$	sr	Solid angle of the i-th bin
$\hat{\varphi}_i$	rad	Mean azimuth angle of the i-th bin
$\hat{\theta}_i$	rad	Mean zenith angle of the i-th bin
$\Delta A_{eff,i}(\hat{\varphi}_i, \hat{\theta}_i)$	m^2	Effective area of the detector when seen from an angle given by $\hat{\varphi}_i$, $\hat{\theta}_i$
ΔA	m^2	Area of the detector
$\Phi_i(E_{cut,i})$	s^{-1} m^{-2} sr^{-1}	Calculated muon flux in the i-th bin according to Eq. (5)
$\Delta\mathcal{E}_i$	s m^2 sr	Exposure in the i-th bin
$\vec{r}_d$	1	Normal vector to the detector surface
$\vec{r}_{\mu,in}(\hat{\varphi}_i, \hat{\theta}_i)$	1	Direction of mean muon incidence, i.e. unit vector in direction of $(\hat{\varphi}_i, \hat{\theta}_i)$. Direction vector of the i-th bin
φ_d	rad	Azimuth angle of the detector facing direction
θ_d	rad	Zenith angle of the detector facing direction
Section 2.2		
x	m	Position of a muon along a trajectory
$E(x)$	eV	Energy of the muon at position x
$-\frac{dE}{dx}$	eV m^{-1}	Energy loss of a muon traversing some matter

$\rho(x)$	$\mathrm{kg\,m^{-3}}$	Density at position x
$a(x,E)$	$\mathrm{eV\,m^2\,kg^{-1}}$	Ionisation losses
$\tilde{k}$	1	Index that runs through all three different radiative processes by which a muon can lose energy, here $\tilde{k} \in \{\mathrm{bremsstrahlung}, \mathrm{pair-production}, \mathrm{photonuclear}\}$
$b_{\tilde{k}}(x,E)$	$\mathrm{m^2\,kg^{-1}}$	Radiation losses of the $\tilde{k}$-th radiative process
N_A	$\mathrm{mol^{-1}}$	Avogadro constant
A	$\mathrm{kg\,mol^{-1}}$	Molar mass of the traversed material; written as "A" as it is numerically equivalent to the atomic mass
ν	1	Fractional energy transfer
$\frac{d\sigma_{\tilde{k}}}{dv}$	$\mathrm{m^2\,kg^{-1}}$	Differential cross-section of the k-th radiative process
E_0	eV	Energy of the muon when it reaches the detector
$\vec{L}_i$	m	Vector of the thicknesses of all m materials in the i-th bin
$\vec{\rho}$	$\mathrm{kg\,m^{-3}}$	Vector of the material densities of all m materials
$\vec{c}$	1	Vector of the compositions of all m materials
$rk(\vec{L}_i, \vec{\rho}, \vec{c})$	eV	Numerical solution of the energy loss equation, using a Runge-Kutta integration scheme

Section 3.1

d_i	1	Measured number of muons in the i-th bin
f_i	$\mathrm{s^{-1}\,m^{-2}\,sr^{-1}}$	Muon flux as a random variable in the i-th bin
μ_{f_i}	$\mathrm{s^{-1}\,m^{-2}\,sr^{-1}}$	Mean of the flux random variable in the i-th bin
$\sigma^2_{f_i}$	$\mathrm{s^{-2}\,m^{-4}\,sr^{-4}}$	Variance of the flux random variable in the i-th bin
p_k	$k = \mathrm{ion}: \mathrm{eV\,m^2\,kg^{-1}}$ else: $\mathrm{m^2\,kg^{-1}}$	Energy loss processes
ε_k	1	Relative error of the energy loss process
k	1	Index that runs through all energy loss processes, here $k \in \{\mathrm{ionisation}, \mathrm{bremsstrahlung}, \mathrm{pair-production}, \mathrm{photonuclear}\}$
$M(\vec{p}_{el}, \vec{\rho}, \vec{c}, \vec{L}_i)$	$\mathrm{s^{-1}\,m^{-2}\,sr^{-1}}$	The forward model, encompassing an energy loss calculation and a flux calculation. Results in a simulated flux.

Section 3.2

$\vec{d}$	1	Vector of observed number of muons in all bins
$\vec{m}$	different	Vector of model parameters
$\vec{m}_0$	different	Vector of initial model parameters
$\vec{m}_{new}$	different	Vector of new model parameters
J	different	Model proposal distribution
c	1	Scaling factor for model proposal variance
S	different	Prior model proposal variances
D	1	Number of model parameters

O	1	Odds ratio
p_A	1	Acceptance probability
Section 3.3		
P	1	Matrix of digital elevation model of the reconstructed interface, also called surface matrix; elements are denoted P_{rq}
r	1	Number of rows in P
c	1	Number of columns in P
H_i	m	Height of the interface in the i-th cone
Δx_i	m	Horizontal distance from interface point to the grid point P_{rq}
Δy_i	m	Vertical distance from interface point to the grid point P_{rq}
Δ_{cs}	m	Cell size of P
w_i	1	Weights of the bilinear interpolation
K	1	Smoothing matrix/kernel

2 The muon simulation

In geophysical communities this part is generally known as the forward model, i.e. a mathematical model which calculates synthetic data for given "model" parameters. In muon tomography experiments, this forward model consists of different physical models which are serially connected.

~~2.1~~2.1 A note on the parametrisation

In geophysical problems, there are many ways to parametrise a given problem. One frequently used approach is to partition the space into voxels (i.e. volume pixels of the same size) and describe them by the material parameters only. This has the advantage of imposing a fixed geometry (i.e. "Thickness" does not enter as a parameter). Unfortunately, the vast number of parameters to be determined requires a very good data coverage of the voxelised region. Another drawback might be the use of smoothing techniques (such that neighbouring voxels are forced to yield similar material parameters), that might blur any sharp interface. Because of these reasons we employed in our code another parametrisation that mimics the actual measurement process. We refer the reader again to Fig. 1 a), where two different cones (i and j) are shown. These represent two (among many) directions in which a certain number of incoming muons have been measured and (over a certain solid angle) binned together. It is now possible (as shown in cone i) to parametrise the sharp discontinuity explicitly by adding two segments (ice and rock) within that cone. As such, the thicknesses of each material within each cone L_{mi} and their material densities ρ_m are explicit parameters in the model.

2.2 Cosmic ray flux model

The nature of the data used in muon tomography generally consists of several counts within a directional bin, defined by two polar and two azimuthal angles. Additionally, the measurement is taken over a defined period of time, as well as over a given extent within the detector area. The simulated number of muons, in the i-th bin, $N_{\mu,i}$, can be calculated by this integral,

~~$N_{\mu,i}^{sim} = \int_T \iint_\Omega \iint_A \int_E \frac{dI}{dE} dE\, dA\, d\Omega\, dT$.~~ (1)

$$N_{\mu,i} = \int_T \iint_\Omega \iint_A \int_E \frac{d\Phi}{dE} dE\, dA\, d\Omega\, dT \,. \quad (1)$$

Here, ~~T~~the subscript μ indicates that muons are counted, the index $i \in \{1, \dots, N_{cones}\}$ addresses the bin/cone, T denotes the exposure time interval, ~~A~~A the detector area, ~~Ω~~Ω the solid angle of the bin and ~~E~~E the energy range of the muons that were able to be registered by the detector. There are various differential muon flux models, also referred to as the integrand, $d\Phi/dE$, in Eq. (1), that can be employed at this stage. Lesparre et al. (2010) provide a good overview on the different flux models, which can broadly be divided in two classes. On the one hand there are theoretical models, which capture the manifold production paths of muons and condense them in an analytical equation, e.g. the Tang et al., (2006) model. They contrast, on the other hand, with empirical models that were generated by fitting formulae to the results of muon flux measurements. The model of Bugaev et al. (1998) falls into this category, with later adjustments for different zenith angles (Reyna, 2006) and altitude (Nishiyama et al., 2017), which are also utilised in this study. The details of the formula are explained in Appendix A. The evaluation of Eq. (1) is rather cumbersome as strictly speaking several of the integration variables depend on each other. We may facilitate the calculation by considering that the differential muon flux model is only dependent on energy, ~~E~~E, and zenith angle, θ whereas the effective area, $\Delta A_{eff,i}$, is ~~solely~~usually (as long as the detector records muons on flat surfaces) dependent on the orientation of the bin. This is the case because muons do not necessarily hit the detector perpendicularly, such that the effective target area is usually smaller. By averaging over the zenith angle and keeping the bin size reasonably small, we may approximate Eq. (1) by

~~$N_{\mu,i}^{sim}(E_{cut,i}) = \int_{E_{cut,i}}^{\infty} \frac{dI}{dE}(E, \widehat{\theta_i})\, dE$~~ $N_{\mu,i}(E_{cut,i}) = \int_{E_{cut,i}}^{\infty} \frac{d\Phi}{dE}(E, \widehat{\theta_i})\, dE * $ ~~$\Delta T * \Delta\Delta T *$~~ $\Delta A_{eff,i}(\widehat{\varphi_i}, \widehat{\theta_i}) *$ ~~$\Delta\Omega_i \Delta\Omega_i$~~ , (2)

where ΔT is the exposure time and ~~Δ~~$\Delta A_{eff,i}$ is the effective detector area. ~~Δ~~$\Delta\Omega_i$ is the solid angle, $\widehat{\varphi_i}$ and $\widehat{\theta_i}$ are the mean azimuth and zenith angle of the i-th bin, respectively. $E_{cut,i}$, called the cutoff-energy, describes the energy needed for a muon to traverse the geological object and to enter the detector. ~~Δ~~$\Delta A_{eff,i}$ has to be scaled by the cosine of the angle between the bin direction and the detector facing direction, which can be calculated using the formula for a scalar product,

$$\Delta A_{eff,i} = \Delta A * \frac{\vec{n} \cdot \vec{d}_\mu(\widehat{\varphi_i}, \widehat{\theta_i})}{\|\vec{n}\| \|\vec{d}_\mu(\widehat{\varphi_i}, \widehat{\theta_i})\|} \frac{\vec{r}_d \cdot \vec{r}_{\mu,in}(\widehat{\varphi_i}, \widehat{\theta_i})}{\|\vec{r}_d\| \|\vec{r}_{\mu,in}(\widehat{\varphi_i}, \widehat{\theta_i})\|}, \quad (3)$$

where $\vec{n}\vec{r}_d$ is the normal vector to the detector surface and $\vec{d}_{\vec{\mu}}\vec{r}_{\mu,in}(\hat{\varphi}_i, \hat{\theta}_i)$ is the mean vector of muon incidence within the i-th bin, both of which can be chosen to feature unit length. Evaluating the scalar product in spherical coordinates, Eq. (3) yields

$$\Delta\Delta A_{eff,i} = \Delta A \Delta A * \left[\sin(\theta_d)\sin(\hat{\theta}_i)\cos(\varphi_d - \hat{\varphi}_i) + \cos(\theta_d)\cos(\hat{\theta}_i)\right]. \quad (4)$$

Here, θ_d and φ_d are the zenith and azimuth angles of the detector facing direction. It is important to note that except for $E_{cut,i}$ all variables in Eq. (2) are predetermined by the experimental setup (ΔT, $\Delta\Delta T$, ΔA) as well as by the data processing ($\hat{\varphi}_i$, $\hat{\theta}_i$), such that the number of muons $N_{\mu,i}\frac{N^{sim}_{\mu,i}}{N^{sim}_{\mu,i}}$ can be interpreted as a function of one variable, $E_{cut,i}$ only.

2.2One final tweak can be made to render Eq. (2) more accessible for future uses. We may rewrite the right-hand side of Eq. (2) in terms of a flux,

$$\Phi_i(E_{cut,i}) = \int_{E_{cut}}^{\infty} \frac{d\Phi}{dE}(E, \hat{\theta}_i)\, dE, \quad (5)$$

and an exposure

$$\Delta\mathcal{E}_i = \Delta T * \Delta A_{eff,i}(\hat{\varphi}_i, \hat{\theta}_i) * \Delta\Omega_i. \quad (6)$$

In this case, the number of number of muons in the i-th bin may be expressed by

$$N_{\mu,i}(E_{cut,i}) = \Phi_i(E_{cut,i}) * \Delta\mathcal{E}_i. \quad (7)$$

2.3 Muon transportation model

Since muons permanently lose energy when travelling through matter, they also need a certain amount of energy to enter the detector. This energy, $E_{cut,i}$, was introduced in Eq. (2) and is called the cutoff energy. If the detector is now positioned underground, the muons have to traverse more matter to reach the detector and consequently need a higher initial energy to reach the target. For this purpose, we introduce the new variable E_0, which refers to the energy needed to penetrate the detector, and we reinterpret $E_{cut,i}$ as the minimum energy that is required to traverse the matter and to be registered at the detector. For the goal of studying the interactions between particles and matter, physicists regularly use energy loss models. We base our calculations in large parts on the equations of Groom et al. (2001), where the energy loss of a muon along its path is described by an ordinary differential equation of 1st order,

$$-\frac{dE}{dx} = \rho(x) * [a(x,E) + E * b(x,E)]\,. \quad (58)$$

In Eq. (58), ρ denotes the density of the traversed material, and a and b are the ionisation loss and radiation loss parameters respectively, and x is the position along the trajectory. The radiation loss parameter groups the effects related to the bremsstrahlung, b_{brems}, the pair-production, b_{pair}, and the photonuclear interactions, b_{photo}, where

$$b(x,E) = b_{brems}(x,E) + b_{pair}(x,E) + b_{photo}(x,E). \quad (69)$$

Each of the radiative process is, in turn, calculated through

$$b_{\tilde{k}} = \frac{N_A}{A}\int_0^1 \nu \frac{d\sigma_{\tilde{k}}}{d\nu} b_{\tilde{k}} = \frac{N_A}{A}\int_0^1 \nu \frac{d\sigma_{\tilde{k}}}{d\nu} d\nu\,, \quad (710)$$

where $k \in \tilde{K}\tilde{k} \in \tilde{K} = \{$bremsstrahlung, pair-production, photonuclear$\}$ is the set of radiative processes, N_A is Avogadro's number, A is the atomic weight of the traversed material, ν is the fractional energy transfer, and $d\sigma_{\tilde{k}}/d\nu\, d\sigma_{\tilde{k}}/d\nu$ the differential cross-section of the process. Eq. (710) becomes important when modelling errors have to be included (see ChSect. 3). For a detailed discussion of the equations for a and b we refer to Groom et al. (2001). The only exception in Eq. (69) is b_{pair}, which is calculated after GEANT4 (Agostinelli et al., 2003). We selected the solution of these latter authors because it is computationally less time consuming. As the two results agree within 1-%, we deem it acceptable to exchange the two differential cross-sections.

Because Eq. (58) describes the energy loss in response to the interaction with a single-element material, certain modifications have to be made to make it also valid for rocks, which in this context represent a mixture of minerals and elements. In this case, the modified equation takes an equivalent form to Eq. (58) when replacing ρ, a, b with their mixture counterparts $\{\rho\}_{rock}, \{a\}_{rock}, \{b\}_{rock}$ (Lechmann et al., 2018), thus yielding

$$-\frac{dE}{dx} = \{\rho(x)\}_{rock} * [\{a(x,E)\}_{\text{rock}} + E * \{b(x,E)\}_{\text{rock}}]\,. \quad (811)$$

We show in Appendix B3 how the rock model (explained in ChSect. 2.34) can be used to determine these quantities.

By applying a change of variables to Eq. (811), i.e. $x' = -x$, the energy loss equation can be transformed to an energy gain equation. This has the advantage of being much easier to solve than the "final value problem" in Eq. (811). We can reorganise Eq. (811) into an initial value problem by setting the initial energy to E_0,

$$\frac{dE}{dx} = \{\rho(x)\}_{rock} * [\{a(x,E)\}_{\text{rock}} + E * \{b(x,E)\}_{\text{rock}}] \quad (912)$$

$$E(0) = E_0\,.$$

In this context E_0 is the minimal energy needed for a muon to penetrate the detector, which can be influenced by the detector design. Equation (912) is a well-investigated problem that can be solved by numerous methods. In our work we employ a standard Runge-Kutta integration scheme (see for example Stoer and Bulirsch, 2013), with a step size of 10 cmcm. As a result, it is now possible to write the cut-off energy in a functional form, where

$$E_{cut,i} = rk(\vec{x_t}\vec{L}_i, \vec{\rho}, \vec{c})\,. \quad (1013)$$

Here $rk(\cdot)$ is the function that returns the Runge-Kutta solution of Eq. (912) for defined thicknesses of materials, $\vec{x_t}\vec{L}_i$, with densities $\vec{\rho}$ and compositional parameters $\vec{c}$. Thickness and density are allowed to be vectors, as there may be more than just one material. In this case, the final energy, after the muon has passed through the first segment of materials, is the initial

energy for the second segment, etc. In order to speed up the computations – especially the calculation of the pair production cross-section, which includes two nested integrations. – we utilise customised energy loss tables. In particular, a log-log table of muon energy vs. radiation loss parameters is produced, from which the b-values, see Eq. (~~7~~10), can be interpolated. We justify this approach because the radiative losses are almost linear in a log-log plot, as can be seen in Fig. 33.1 of Tanabashi et al. (2018, p.447) for the example of copper. The general shape of the energy loss function remains the same for various materials even if the absolute values differ.

2.~~3~~4 Rock model

Equation (~~10~~13) shows that for the calculation of the cut-off energy two types of material parameters are required, which are the material density $\vec{\rho}$ and its average composition $\vec{c}$. The pre-tabulated values from Groom et al. (2001), however, include only pure elements as well as certain compounds. To extract the relevant parameters in a geological setting, a realistic rock model is needed. In an earlier work (Lechmann et al., 2018) we have shown how an integrated rock model can be constructed and how the physical parameters for a realistic rock can be retrieved. In the present work we generally use the same approach, apart from a few aspects. First, we measured the average material density directly in the laboratory, using various techniques which are explained in detail in Appendix B1. Second, in order to be able to compare the results of this study with the ones in the Nishiyama et al. (2017) publication, we consider a rock composition that corresponds to a density-modified standard rock. This is applicable, as the rock in the study of Nishiyama et al. (2017) is mostly of granitic/gneissic origin, with thicknesses rarely larger than 200 m, with the consequence that the differences are negligible. However, as the inclusion of compositional data is a planned feature for a future version of our code, we decided to include the theoretical treatment in this work. Hence, all equations are tailored to include the statistical description of such data. Compositional data for whole rock samples, which can be scaled to outcrop scale, are usually presented in one of two forms, the first being the measurement results of X-Ray Diffraction (XRD). This kind of data yields the mineral phases within a rock. Unfortunately, XRD is a rather time-consuming method. This is the reason why in muon tomographic experiments researchers often resort to a bulk chemical analysis of the rock, which is the second form of compositional data. This type of data is usually the output of dedicated X-Ray Fluorescence (XRF) measurements, describing the bulk rock composition by major oxide fractions. We note here that by the absence of information on the spatial distribution of mineral phases within a rock, we implicitly infer a homogeneous mixture of elements within the rock itself, which is thus different from our previous work (Lechmann et al., 2018). From a particle physics perspective this does not pose a real problem as the difference to a mixture of minerals is rather small. Nevertheless, we lose the power to obtain meaningful inferences that could be drawn if compositional information is being considered. As the present work aims to infer positions and uses material parameters as constraints, we can accept this drawback. Details on how compositional parameters are derived from XRF measurements, including an example, can be found in Appendix B2, and an explanation of the related influence on the energy loss equation in Appendix B3. Additionally, we assume that the density of the rock (and also the ice) is homogeneous throughout the imaged geological body. This is, of course, a simplification as there might be weathering processes that change the density of

the rock at various locations. The same is true for the ice body, which might contain crevasses. As we used this approximation already in Nishiyama et al. (2017) and since we intend to compare the calculations in this study with the earlier one, we will retain the assumption of homogeneity for the densities and compositions.

2.45 Spatial models of detectors and materials

In addition to the above explained physical models, we may also utilise available spatial data for our purposes. In this context, the use of a digital elevation model (DEM) of the surface allows the visualisation of the position of the detectors relative to the surface, as well as the spatial extent of the bins. Additionally, it allows us to determine the location where these bins intersect with the topographic surface. As a first deliverable, we can draw conclusions on which bins consist of how many parameters. For example, if we know that the detector is located underground and that there is ice at the surface, we can already infer the existence of at least 2 materials (rock and ice). For this purpose, we wrote the script "modelbuilder.py", which allows the user to attach geographic and physical information to the selected bins. This process of building a coherent geophysical model is needed for the subsequent employment of the inversion algorithm to process all the data.

3 The inverse model: A Bayesian perspective

As stated in the Introduction, we solve the inversion by using Bayesian methods. This needs an explanation as to why we chose this way and not another. First, the equations in ChSect. 2 enable us to calculate a synthetic dataset for fixed parameter values. There, one can see that the governing equations constitute a nonlinear relationship between parameter values and measured data. Despite this being of no particular interest in the forward model, the estimation of the parameters from measured data is rendered more complicated. Among muon tomographers, linearised versions have been extensively used with deterministic approaches (e.g. Nishiyama et al., 2014; Rosas-Carbajal et al., 2017), which are successfully applicable when the density or the intersection boundaries are the only variables. When deterministic approaches are viable, they efficiently produce good results. Descent algorithms or, generally speaking, locally optimising algorithms, offer a valid alternative, as they could cope with the nonlinearity of the forward model, while including all desired parameters. Even though theseOne difficulty of such algorithms suffer from possibleis that in the case of non-uniquenessunique solutions (i.e. the which occur when there are local minima that might be a solution depends heavily onto the starting model, possibly yielding multiple solutions),optimisation) the mainuser has no constraints to infer if a local or the desired global minimum has been reached. A further problem of descent methods is the calculation of the derivatives of the forward model with respect to the parameter values. The analytical calculation of the derivatives is enormously tedious because the cut-off energy results from a numerical solver of a differential equation, as can be seen in Eq. (9) & (1012) & (13). Unfortunately, numerical derivatives do not produce better results, because they might easily produce artefacts, which are hard to track down. This is especially true if the derivative has to be taken from a numerical result, which is always slightly noisy. In that

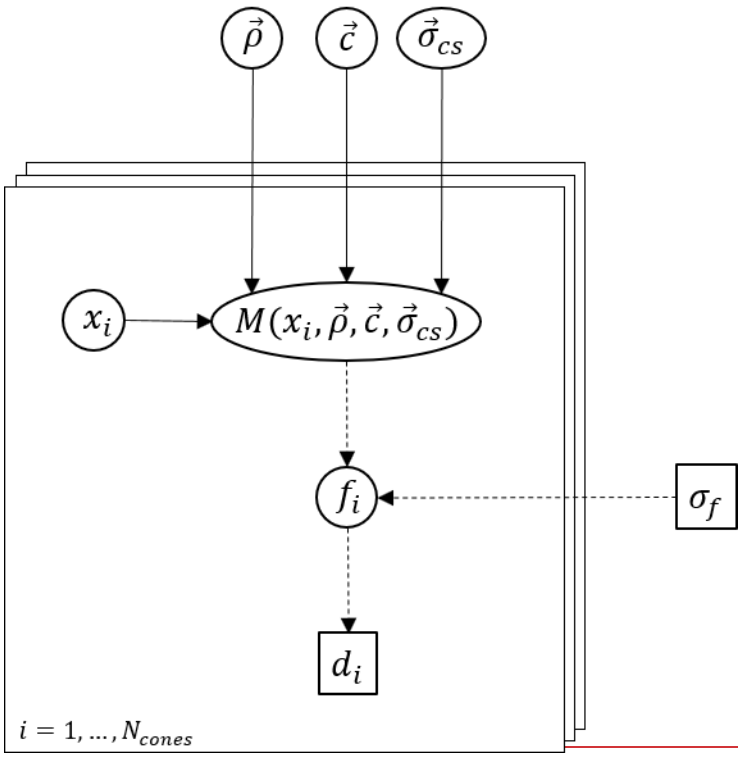
case the differentiation amplifies the "noise", resulting in unreliable gradient estimates. A good overview over deterministic inversion methods can be found in Tarantola (2005).

The reasons stated above and our goal to include as much information on the parameters as possible nudges us towards employing probabilistic methods. Those approaches are also known as Bayesian methods. The main feature that distinguishes them from the deterministic methods described above is the consistent formulation of the equations and additional information in a probabilistic way, i.e. as probability density functions (pdf). This allows us to (i) incorporate, for example, density values that were measured in the lab (including its error), (ii) set bounds on the location of the material interface, or (iii) define a plausible range for the composition of the rock. All these changes act on the pdf of the respective parameter and naturally integrate into the Bayesian inversion. We have to add that Bayesian methods do not solve the non-uniqueness problem, but they provide the user with enough information to spot these local solutions of the optimisation. Readers may find the book of Tarantola (2005) very resourceful for the explanation and illustration of probabilistic inversion. Several studies in the muon tomography community have already employed such methods with success (e.g. Lesparre et al., 2012; Barnoud et al., 2019).

The flexibility of being able to include as much information on the parameters as we consider useful comes at the price of having to solve the inversion in a probabilistic way. This can either be done using Bayes' Theorem and solving for the pdfs of the parameters of interest, or if the analytical way is not possible by employing Monte Carlo techniques. As the presence of a numerical solver renders the analytical solution impossible, we resort to the Monte Carlo approaches. In the following sections we guide the reader through the various stages of how such a probabilistic model can be set up, how probabilities may be assigned, and how the inversion can finally be solved.

3.1 Probabilistic formulation of the forward model

The starting point for a probabilistic formulation is denoted by the equations that were elaborated in ~~Ch~~Sect. 2. These deterministic equations need to be upgraded into a probabilistic framework, where their attributed model and/or parameter uncertainties are inherently described. In the following paragraphs we describe how each model component can be expressed by a pdf before the entire model is composed at the end of this ~~subchapter~~section. The model is best visualised by a directed acyclic graph (DAG), i.e. see Kjaerulff and Madsen (2008), that depicts which variables enter the calculation at what point. For our muon tomography experiment this is visualised in Fig. 3. In the following the pdfs are denoted with the bold Greek letter $\boldsymbol{\pi}$ to differentiate them from normal parameters.


$\vec{\rho}$
$\vec{c}$
$\vec{\sigma}_{cs}$
x_i
$M(x_i, \vec{\rho}, \vec{c}, \vec{\sigma}_{cs})$
f_i
σ_f
d_i
$i = 1, \dots, N_{cones}$

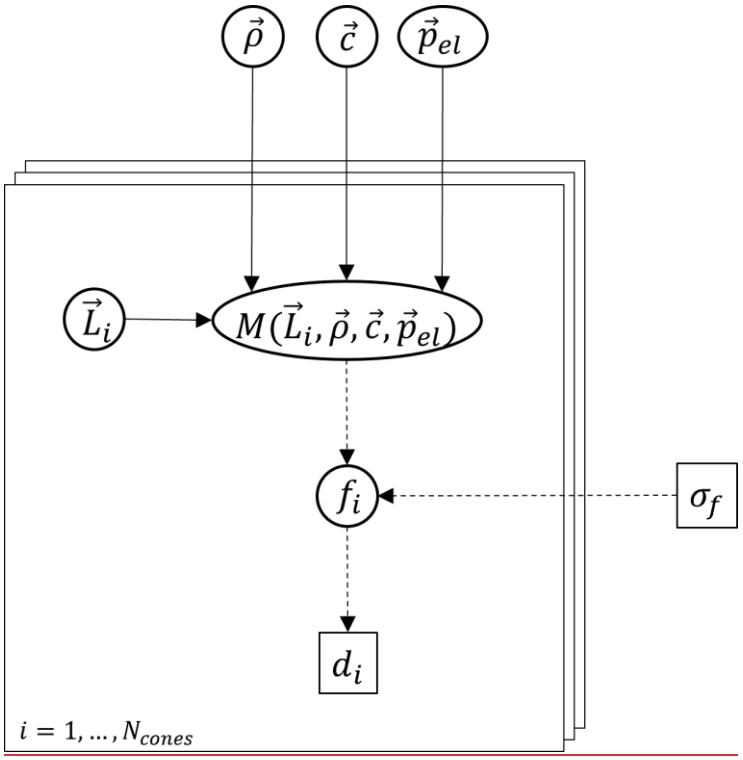

$\vec{\rho}$
$\vec{c}$
$\vec{p}_{el}$
$\vec{L}_i$
$M(\vec{L}_i, \vec{\rho}, \vec{c}, \vec{p}_{el})$
f_i
σ_f
d_i
$i = 1, \dots, N_{cones}$

Figure 3: Directed acyclic graph (DAG) for the problem of muon tomography. Variables in a square (□) denote fixed, i.e. known values and variables in a circle/ellipse (○) are generally unknown and have to be represented by a pdf. Solid arrows (→) denote a deterministic relation, i.e. within a physical model, whereas dashed arrows (⇢) indicate a probabilistic relationship, i.e. a parameter within the statistical description of the variable. $\vec{\rho}, \vec{c}$ are the density and composition for different materials, whereas ~~$\vec{\sigma}_{cs}\vec{p}_{el}$~~ contains the errors on the ~~physical cross-sections~~different processes in the energy-loss equation. σ_f describes the error on the cosmic-ray flux model. Within each cone, ~~$x_i\vec{L}_i$~~ ~~is~~denote the ~~position~~thicknesses of ~~the bedrock~~ rock and ice ~~interface~~respectively, $\boldsymbol{M}(\vec{L}_i, \vec{\rho}, \vec{c}, \vec{\sigma}_{cs})$ is the calculated flux (i.e. energy-loss model and flux model combined), f_i the actual muon flux and d_i the observed number of muon tracks.

3.1.1 Muon data

The data in muon tomography experiments are usually count data, i.e. a certain number of measured tracks within a directional bin, which has been collected over a certain exposure time and detector area. As the measured number of muons is always an integer, we may model such data by a Poisson distribution,

~~$\boldsymbol{\pi}(d_i|N_i) = \frac{N_i^{d_i} e^{-N_i}}{d_i!}$, (11~~

$$\boldsymbol{\pi}(d_i|N_{\mu,i}) = \frac{(N_{\mu,i})^{d_i} e^{-N_{\mu,i}}}{d_i!} \quad (14)$$

where d_i denotes the measured number of muons in the i-th bin and ~~N_i~~$N_{\mu,i}$ is the poisson parameter in the same bin, which can be interpreted as mean and variance of this distribution. ~~Equation (11)~~According to Eq. (7) we may ~~be rewritten~~rewrite Eq. (14) in terms of a flux, ~~f_i by~~f.

~~$\boldsymbol{\pi}(d_i|f_i) = \frac{(f_i \Delta E x_i)^{d_i} e^{-(f_i \Delta E x_i)}}{d_i!} \frac{(f_i \Delta \varepsilon_i)^{d_i} e^{-(f_i \Delta \varepsilon_i)}}{d_i!}$, (12.~~

~~(15)~~

~~where f_i is the muon flux in the i-th bin and~~

~~$\Delta E x_i = \Delta A_{eff,i} * \Delta T * \Delta \Omega_i$ (13)~~

~~is the exposure, in which $\Delta A_{eff,i}$ is the effective total detector area from Eq. (4), ΔT is the exposure time and $\Delta \Omega_i$ is the solid angle.~~

The variable here is explicitly denoted f_i to emphasise that it is not the exact calculated flux, Φ, from before but a separate variable subject to uncertainties. These two variables will be linked with each other in the next subsection.

3.1.2 Flux model

The next step is to set up a probabilistic model for the muon flux. First, we observe that "flux" is a purely positive parameter, i.e. $f_i \in [0, \infty)$. Thus, it is natural to model it by a lognormal probability distribution if estimates of mean and variance are readily available. The uncertainty on the muon flux is generally taken around 15% (e.g. Lechmann et al., 2021a) of the mean value. As it is possible, by Eq. (~~2~~5), to calculate a flux for a given cut-off energy, which we interpret as the mean of the non-logarithmic values, the parameters of the lognormal distribution (i.e. $\mu_{f_i}, \sigma_{f_i}^2$) may be expressed by

$$\sigma_{f_i}^2 = \ln\left(1 + \left(\frac{F_i(E_{cut,i})*0.15}{F_i(E_{cut,i})}\right)^2\right) \ln\left(1 + \left(\frac{\Phi_i(E_{cut,i})*0.15}{\Phi_i(E_{cut,i})}\right)^2\right) = \ln(1.0225) \quad (14\,16)$$

and

$$\mu_{f_i} = \ln\left(F_i(E_{cut,i})\right) - \frac{\sigma_{f_i}^2}{2} \ln\left(\Phi_i(E_{cut,i})\right) - \frac{\sigma_{f_i}^2}{2}, \quad (15\,17)$$

which yield the probability density function for the flux, conditional on the cut-off energy

$$\boldsymbol{\pi}\left(f_i \middle| \mu_{f_i}, \sigma_{f_i}^2\right) = \boldsymbol{\pi}\left(f_i \middle| E_{cut,i}\right) = \frac{1}{\sqrt{2\pi} * f_i * \sigma_{f_i}} \exp\left(-\frac{1}{2}\left(\frac{\ln(f_i) - \mu_{f_i}}{\sigma_{f_i}}\right)^2\right) . (16 \left(-\frac{1}{2}\left(\frac{ln(f_i) - \mu_{f_i}(E_{cut,i})}{\sigma_{f_i}}\right)^2\right) . \quad (18)$$

3.1.3 Energy loss model

The energy loss model has multiple sources of errors that have to be taken into account. Most notably, the relative errors on the different physical cross-sectionsenergy loss processes are given by Groom et al. (2001) as $\varepsilon_{ion} = 6\text{-}\%, \varepsilon_{brems} = 1\text{-}\%, \varepsilon_{pair} = 5\text{-}\%, \varepsilon_{photonucl} = 30\text{-}\%$. As it is not clearly stated as to what this error relates to, i.e. one or more standard deviations, we interpret an error like $\varepsilon_{ion} = 6\text{-}\%$ as: "within a factor of 1.06", which can be written as

$$\frac{\sigma_k}{(1+\varepsilon_k)} \frac{p_k}{(1+\varepsilon_k)} \leq \sigma_k p_k \leq \sigma_k p_k (1 + \varepsilon_k), \quad (17\,19)$$

where $k \in K =$ {ionisation, bremsstrahlung, pair-production, photonuclear}.} runs through all energy loss processes, including ionisation losses and $p_{ion} = a, p_{brems} = b_{brems}, p_{pair} = b_{pair}, p_{photo} = b_{photo}$ denote the respective energy loss terms from Eq. (12). Dividing this inequality by σ_k and taking the logarithm yields

$$-\ln(1 + \varepsilon_k) \leq 0 \leq \ln(1 + \varepsilon_k). \quad (18\,20)$$

Thus, we may attribute a Gaussian pdf in the log-space for a "log-correction factor, $l\sigma_k p_k$" by setting its mean to zero and its standard deviation to $\ln(1 + \varepsilon_k)$, i.e.

$$\boldsymbol{\pi}(l\sigma_k)(lp_k) = \frac{1}{\sqrt{2\pi} * \ln(1+\varepsilon_k)} \exp\left(-\frac{1}{2}\left(\frac{l\sigma_k}{\ln(1+\varepsilon_k)}\right)^2\right)\left(-\frac{1}{2}\left(\frac{lp_k}{\ln(1+\varepsilon_k)}\right)^2\right). \quad (19\,21)$$

With a change of variables, using the Jacobian rule as explained in Tarantola (2005), we get

$$\boldsymbol{\pi}(\sigma_k) = \frac{1}{\sqrt{2\pi} * \sigma_k * \ln(1+\varepsilon_k)} \exp\left(-\frac{1}{2}\left(\frac{\ln(\sigma_k)}{\ln(1+\varepsilon_k)}\right)^2\right), \quad (20)$$

$$\boldsymbol{\pi}(p_k) = \frac{1}{\sqrt{2\pi} * p_k * \ln(1+\varepsilon_k)} \exp\left(-\frac{1}{2}\left(\frac{\ln(p_k)}{\ln(1+\varepsilon_k)}\right)^2\right), \quad (22)$$

the lognormal pdf for the correction factor. The pdf for the ~~cross section~~energy loss uncertainty $\boldsymbol{\pi}(\vec{p}_{el})$ can now be written as a product of the four different pdfs described by Eq. (~~20~~22)

$$\boldsymbol{\pi}(\vec{p}_{el}) = \prod_{k\in K} \boldsymbol{\pi}(p_k), \quad (~~21~~23)$$

as the errors of the ~~physical cross sections~~individual energy loss processes are stochastically independent from each other. The calculated energy loss depends also on the material parameters and subsequently on their uncertainties. However, these will be explained in detail in ~~Ch~~Sect. 3.1.4. A last ~~error~~uncertainty enters by the numerical solution of the ordinary differential equation, Eq. (~~9~~12). We decided not to model this error, as its magnitude is directly controlled by the user (by setting a small enough step length in the Runge-Kutta algorithm) and thus can be made arbitrarily small. Lastly, we assume that all the errors in the energy loss model are explained by uncertainties in the ~~cross sections~~energy loss terms as well as in the material parameters. Although this assumption is rather strong, since it excludes the possibility of a wrong model, we argue that this approach works as long as the variation in these parameters can explain the variation in the calculated cut-off energy. If this requirement is met, we may model the pdf for the energy loss model as a delta function,

$$\boldsymbol{\pi}(E_{cut,i}|\vec{p}_{el},\vec{\rho},\vec{c},\vec{L}_i) = \boldsymbol{\delta}\left(E_{cut,i} - rk(\vec{p}_{el},\vec{\rho},\vec{c},\vec{L}_i)\right), \quad (~~22~~24)$$

where ~~$\vec{\sigma}_{cs} = (\sigma_k)$~~$\vec{p}_{el} = (p_k)$, $\vec{\rho}$ is the vector of all material densities, $\vec{c}$ is the vector of all compositions and ~~x_i~~$\vec{L}_i$ is the vector of thicknesses of segments used in this cone. It is now already possible to eliminate $E_{cut,i}$ as a parameter by first multiplying Eqs. (~~16) & (22~~18) & (24), which yields

$$\boldsymbol{\pi}(f_i, E_{cut,i}|\vec{\sigma}_{cs},\vec{\rho},\vec{c},~~x_i~~\vec{L}_i) = \boldsymbol{\pi}\left(f_i\middle|E_{cut,i}\right) * \boldsymbol{\pi}(E_{cut,i}|\vec{\sigma}_{cs},\vec{\rho},\vec{c},~~x_i~~)(E_{cut,i}|\vec{p}_{el},\vec{\rho},\vec{c},\vec{L}_i). \quad (~~23~~25)$$

From this expression ~~it is possible to~~one can then marginalise the parameter $E_{cut,i}$, by simply integrating over it, i.e.

$$\boldsymbol{\pi}(f_i|\vec{\sigma}_{cs},\vec{\rho},\vec{c},~~x_i~~)(f_i|\vec{p}_{el},\vec{\rho},\vec{c},\vec{L}_i) = ~~\int \boldsymbol{\pi}(f_i, E_{cut,i}|\vec{\sigma}_{cs},\vec{\rho},\vec{c},~~x_i~~)~~ \int \boldsymbol{\pi}(f_i, E_{cut,i}|\vec{p}_{el},\vec{\rho},\vec{c},\vec{L}_i)\, dE_{cut,i}\,. \quad (~~24~~26)$$

Due to the presence of the delta function in Eq. (~~22~~24), this integral is solved analytically resulting in

$$\boldsymbol{\pi}(f_i|\vec{\sigma}_{cs},\vec{\rho}_i,\vec{c}_i,x_i) = (f_i|\vec{p}_{el},\vec{\rho},\vec{c},\vec{L}_i) = \frac{1}{\sqrt{2\pi} * \mathrm{f_i} * \sigma_{f_i}} \exp\left(-\frac{1}{2}\left(\frac{\ln(f_i)-\mu_{f_i}}{\sigma_{f_i}}\right)^2\right)\left(-\frac{1}{2}\left(\frac{\ln(f_i)-\mu_{f_i}(rk(\vec{p}_{el},\vec{\rho},\vec{c},\vec{L}_i))}{\sigma_{f_i}}\right)^2\right), \quad (~~25~~27)$$

where the parameters are given by

$$\sigma_{f_i}^2 = \ln\left(1 + \left(\frac{M(\vec{\sigma}_{cs},\vec{\rho},\vec{c},x_i)*0.15}{M(\vec{\sigma}_{cs},\vec{\rho},\vec{c},x_i)}\right)^2\right) \ln\left(1 + \left(\frac{M(\vec{p}_{el},\vec{\rho},\vec{c},\vec{L}_i)*0.15}{M(\vec{p}_{el},\vec{\rho},\vec{c},\vec{L}_i)}\right)^2\right) = \ln(1.0225) \quad (~~26~~28)$$

and

$$\mu_{f_i} = \ln(M(\vec{\sigma}_{cs},\vec{\rho},\vec{c},x_i)) - \frac{\sigma_{f_i}^2}{2} \ln\left(M(\vec{p}_{el}\vec{\rho},\vec{c},\vec{L}_i)\right) - \frac{\sigma_{f_i}^2}{2}. \quad (~~27~~29)$$

Please note that ~~$M(\vec{\sigma}_{cs}, \vec{\rho}, \vec{c}, x_t)$~~$M(\vec{p}_{el}, \vec{\rho}, \vec{c}, \vec{L}_i) =$ ~~$F_t \Phi_i(rk(\vec{\sigma}_{cs}$~~$\vec{p}_{el}, \vec{\rho}, \vec{c},$ ~~x_t~~$\vec{L}_i))$ describes the combined parts of the forward model that include the energy loss and the integrated flux calculation, which is basically a composition of functions.

3.1.4 Rock model

The density model can take different forms of probability densities (see Appendix B1), such as normal, lognormal, uniform, etc. For either form, it is possible to describe it by a generic function $\boldsymbol{\pi}(\vec{\rho})$, which is a short version for a multidimensional pdf, i.e. $\boldsymbol{\pi}(\rho_{ice}, \rho_{rock})$ if the i-th cone is known to consist of two segments with two specific densities. Equivalently, the pdf for the composition (see Appendix B2) is either fixed or a multidimensional Gaussian distribution in the space of log-ratios. Thus $\boldsymbol{\pi}(\vec{c})$ can be split up to $\boldsymbol{\pi}(c_{ice}, c_{rock})$, like in the example above. Generally, we may assume that in our problem j different materials exist. We note here that the description of the composition is already probabilistic. However, the inversion in that case is not functional and works only with the mean values of the multidimensional Gaussian. The support for composition inversion is planned for a future version of the code.

The situation for the thicknesses of the segments, $\boldsymbol{\pi}$~~(x_t)~~$(\vec{L}_i)$, within the i-th cone presents itself in a similar way as for the compositions (e.g., Fig. 1). ~~As~~We know the total ~~material~~ thickness ~~is known (~~, due to our information on the detector position and the surface position from digital elevation models ~~are given), the sub-space containing the thickness parameter is endowed with the same~~. Thus we have, equal to the compositions' weight fractions (that add up to 1), a sum constraint (i.e. the sum of lengths of all segments must equal the total distance from the detector to the surface). The mathematical structure ~~as~~of the ~~one containing the composition~~ parameter ~~(i.e. one sum constraint), if~~sub-space is consequently the also the ~~cone consists of more than just one segment~~same. One can therefore safely assume that the thickness parameters ~~live~~can be presented in a log-ratio space, within which we a-priori possess no additional information ~~about the parameters~~. Thus, we attribute the thickness parameters a multidimensional uniform distribution within the log-ratio space.

3.1.5 The Joint probability density function

With the help of the DAG, introduced in Fig. 3, it is now straightforward to factorise the joint probability distribution for the whole problem, as their structure is equal. This results in

~~$\boldsymbol{\pi}(\vec{d}, \vec{f}, \vec{\sigma}_{cs}, \vec{\rho}, \vec{c}, \vec{x}) = \prod_{i=1}^{N_{cones}} \boldsymbol{\pi}(d_i|f_i)\boldsymbol{\pi}(f_i|\vec{\sigma}_{cs}, \vec{\rho}, \vec{c}, x_i)\boldsymbol{\pi}(x_i) * \prod_{j=1}^{N_{materials}} \boldsymbol{\pi}(\rho_j)\boldsymbol{\pi}(c_j) * \prod_{k \in K} \boldsymbol{\pi}(\sigma_k)$~~ ;

~~(28~~ $\boldsymbol{\pi}(\vec{d}, \vec{f}, \vec{p}_{el}, \vec{\rho}, \vec{c}, \vec{L}) = \left[\prod_{i=1}^{N_{cones}} \boldsymbol{\pi}(d_i|f_i)\boldsymbol{\pi}(f_i|\vec{\sigma}_{cs}, \vec{\rho}, \vec{c}, \vec{L}_i)\boldsymbol{\pi}(\vec{L}_i)\right] * \left[\prod_{m=1}^{N_{materials}} \boldsymbol{\pi}(\rho_m)\boldsymbol{\pi}(c_m)\right] * \left[\prod_{k \in K} \boldsymbol{\pi}(p_k)\right]$ (30)

or equivalently (and this will also be of a much better use later on) the log joint pdf

$$\boldsymbol{l\pi}\left(\vec{d},\vec{f},\vec{\sigma}_{es},\vec{\rho},\vec{c},\vec{x}\right) = \sum_{i=1}^{N_{cones}} \boldsymbol{l\pi}(d_i|f_i) + \boldsymbol{l\pi}\left(f_i|\vec{\sigma}_{es},\vec{\rho},\vec{c},x_i\right) + \boldsymbol{l\pi}(x_i) + \sum_{j=1}^{N_{materials}} \boldsymbol{l\pi}\left(\rho_j\right) + \boldsymbol{l\pi}\left(c_j\right) + \sum_{k\in K} \boldsymbol{l\pi}(\sigma_k) \quad ; \quad (29$$

$$\left(\vec{d},\vec{f},\vec{p}_{el},\vec{\rho},\vec{c},\vec{L}\right) = \left[\sum_{i=1}^{N_{cones}} \boldsymbol{l\pi}(d_i|f_i) + \boldsymbol{l\pi}\left(f_i|\vec{p}_{el},\vec{\rho},\vec{c},\vec{L}_i\right) + \boldsymbol{l\pi}\left(\vec{L}_i\right)\right] + \left[\sum_{m=1}^{N_{materials}} \boldsymbol{l\pi}(\rho_m) + \boldsymbol{l\pi}(c_m)\right] + \left[\sum_{k\in K} \boldsymbol{l\pi}(p_k)\right], \quad (31)$$

where the prefix "$\boldsymbol{l}$" denotes the logarithm of the pdf. This has the benefit of reducing the size of numbers that the code has to cope with. Moreover, many computational statistics packages already have this feature included, which renders it easy to use.

Equation (2830) depicts the full joint pdf. However, the relations between the parameters, as shown by the DAG (see Fig. 3), classify this model as a hierarchical model (Betancourt and Girolami, 2013). The key characteristic of such models is their tree-like parameter structure, i.e. the measured number of muons is related to the thickness or the density of the material by the flux parameter only, which "relays" the information. A central problem of such models is the presence of a hierarchical "funnel" (see Fig. 2 & 3 of Betancourt and Girolami, 2013), which renders it very difficult for standard MCMCMonte Carlo methods to adequately sample the model space. In high-dimensional parameter spaces this problem exacerbates even more.

Our aim to provide a simple and easy-to-use program somewhat contradicts this necessity of a sophisticated method (which inevitably requires the user to possess a strong statistical background). As the main problem is the rising number of parameters, it should be possible to mend the joint pdf by imposing thought-out simplifications.

We first get rid of the flux parameter, as for our problem it merely is a nuisance parameter. This is an official term for a parameter in the inversion which is of no particular interest but still has to be accounted for. Here specificallySpecifically, we mean that even though the calculation of the muon flux is important, we do not want to treat it as an explicit parameter that is simulated by the code. To achieve this, we integrate over all possible values of the muon flux, $\vec{f}$ within its uncertainty, so that and we can relate the results of the energy loss calculation (encoded in μ_{f_i}; see Eq. 2729) directly to the measured number of muons, d_i. This effectively reduces the number of parameters and thus the number of dimensions of the model space. This can be achieved by marginalising the flux parameter out of the joint pdf,

$$\boldsymbol{\pi}\left(\vec{d},\vec{\sigma}_{es},\vec{\rho},\vec{c},\vec{x}\right) = \int \boldsymbol{\pi}\left(\vec{d},\vec{f},\vec{\sigma}_{es},\vec{\rho},\vec{c},\vec{x}\right) d\vec{f} \quad .(30$$

$$\left(\vec{d},\vec{p}_{el},\vec{\rho},\vec{c},\vec{L}\right) = \int \boldsymbol{\pi}\left(\vec{d},\vec{f},\vec{p}_{el},\vec{\rho},\vec{c},\vec{L}_i\right) d\vec{f} \quad . \quad (32)$$

This effectively reduces to a problem where the new likelihoods have to be calculated (as d_i is given)

$$\boldsymbol{\pi}\left(d_i|\vec{\sigma}_{es},\vec{\rho},\vec{c},x_i\right)\left(d_i|\vec{p}_{el},\vec{\rho},\vec{c},\vec{L}_i\right) = \int \boldsymbol{\pi}(d_i|f_i)\boldsymbol{\pi}\left(f_i|\vec{\sigma}_{es},\vec{\rho},\vec{c},x_i\right)\left(f_i|\vec{p}_{el},\vec{\rho},\vec{c},\vec{L}_i\right) df_i, \quad (3133)$$

or fully,

$$\boldsymbol{\pi}\left(d_i|\vec{\sigma}_{es},\vec{\rho},\vec{c},x_i\right) = \int_0^\infty \frac{(f_i\Delta E x_i)^{d_i} e^{-(f_i\Delta E x_i)}}{d_i!} \frac{1}{\sqrt{2\pi}*\mathrm{f_i}*\sigma_{f_i}} \boldsymbol{\pi}\left(d_i|\vec{p}_{el},\vec{\rho},\vec{c},\vec{L}_i\right) =$$

$$\int_0^\infty \frac{(f_i\Delta\mathcal{E}_i)^{d_i} e^{-(f_i\Delta\mathcal{E}_i)}}{d_i!} \frac{1}{\sqrt{2\pi}*\mathrm{f_i}*\sigma_{f_i}} \exp\left(-\frac{1}{2}\left(\frac{\ln(f_i)-\mu_{f_i}}{\sigma_{f_i}}\right)^2\right) df_i\,, \quad (3234)$$

~~where~~which is a multiplication of Eqs. (15) and (27). In Eq. (34), μ_{f_i} and σ_{f_i} are given by Eqs. (~~26~~28) and (~~27~~29), respectively. This integral is not solvable analytically but can be evaluated by numerical integration schemes. The likelihood has a maximum when the Poisson and the log-normal pdfs fully overlap. Interestingly, this directly shows the trade-off between the flux model uncertainty and the data uncertainty. Usually, we want to measure enough muons so that the statistical counting error is smaller than the systematic uncertainty of the flux model (i.e. the width of the Poisson pdf is smaller than the width of the log-normal pdf). This can be controlled directly by the exposure of the experiment, via a larger detector area, a coarser binning, or a longer exposure time. A guide how such a calculation may be done (especially when planning a measurement campaign) can be found e.g. in Lechmann et al. (2021a).

This marginalisation roughly halves the number of parameters, but there is still another simplification, which we may use. Many muon tomography applications deal with a two-material problem, while there may also be measurement directions where only one material is present. If we conceptually split those two problems and solve them independently, it is possible to further reduce the number of simultaneously modelled parameters. In the study of Nishyiama et al. (2017), the results of which we will use later, these two cases encompass bins where we measured only rock and others where we know there is ice and rock. The joint pdf for rock bins subsequently is

$$\boldsymbol{\pi}\cancel{(\vec{d},\vec{\sigma}_{cs},\rho_{rock},c_{\mathrm{rock}}) = \prod_{i=1}^{N_{cones}^{rock}} \boldsymbol{\pi}(d_i|\vec{\sigma}_{cs},\rho_{rock},c_{\mathrm{rock}}) * \boldsymbol{\pi}(\rho_{rock})\boldsymbol{\pi}(c_{rock}) * \prod_{k\in K}\boldsymbol{\pi}(\sigma_k), (33}\ (\vec{d},\vec{p}_{el},\rho_{rock},c_{\mathrm{rock}}) = \left[\prod_{i=1}^{N_{cones}^{rock}} \boldsymbol{\pi}(d_i|\vec{p}_{el},\rho_{rock},c_{\mathrm{rock}})\right] * \boldsymbol{\pi}(\rho_{rock})\boldsymbol{\pi}(c_{rock}) * [\prod_{k\in K}\boldsymbol{\pi}(p_k)], \quad (35)$$

which leaves the problem effectively with only a handful of parameters. Solving Eq. (~~33~~35), for the rock density we retrieve $\tilde{\boldsymbol{\pi}}(\rho_{rock})$, the posterior marginal pdf for the rock density. We refer the reader to ~~Ch~~Sect. 3.2 for the details of how to solve this inverse problem. Theoretically we could also retrieve $\tilde{\boldsymbol{\pi}}(\rho_{ice})$, but this would require the detector to be positioned within the glacier ~~(see chapter 5.2 and Fig. 1 for a description of the experimental design of the Nishyiama et al. (2017) study),~~ which poses more of a practical difficulty than a mathematical one.

For the second problem, we can interpret $\tilde{\boldsymbol{\pi}}(\rho_{rock})$ as the new prior pdf for the rock density. At this point we employ one last simplification by assuming that the parameters between different cones are independent form each other. This is a rather strong presumption, which must be justified. The main mathematical problem lies in consideration of the hierarchical nature of the density parameter, which is the same for each cone and therefore not independent in different cones. We, however, argue that in cones with two materials, there are more parameters than in bins with only rock, such that we may expect the posterior pdf of the rock density of these second kind of models to be less informative than the posterior rock density pdf of Eq. (~~33~~35). This, in turn, means that the posterior rock density pdf of the two-material model largely equals the prior one if we select the posterior of the first kind of models as the prior of the second kind of models. The same is valid for the composition $\vec{c}_i$ and the ~~cross-section~~energy loss error parameters $\cancel{\vec{\sigma}_{cs}}\vec{p}_{el}$. As long as this assumption is valid, we may decompose the joint pdf into independent joint pdfs for each cone

$$\boldsymbol{\pi}\left(d_i, \vec{\sigma}_{es,i}, \vec{\rho}_i, \vec{c}_i, x_i\right) = \boldsymbol{\pi}\left(d_i \middle| \vec{\sigma}_{es,i}, \vec{\rho}_i, \vec{c}_i, x_i\right)\boldsymbol{\pi}(x_i)\prod_{j=1}^{N_{materials}} \widetilde{\boldsymbol{\pi}}(\rho_{ij})\widetilde{\boldsymbol{\pi}}(c_{ij}) * \prod_{k\in K} \widetilde{\boldsymbol{\pi}}(\sigma_{ik}). \quad (34$$

$$\left(d_i, \vec{p}_{el,i}, \vec{\rho}_i, \vec{c}_i, \vec{L}_i\right) = \boldsymbol{\pi}\left(d_i \middle| \vec{p}_{el,i}, \vec{\rho}_i, \vec{c}_i, \vec{L}_i\right) * \boldsymbol{\pi}\left(\vec{L}_i\right) * \left[\prod_{m=1}^{N_{materials}} \widetilde{\boldsymbol{\pi}}(\rho_{im})\widetilde{\boldsymbol{\pi}}(c_{im})\right] * \left[\prod_{k\in K} \widetilde{\boldsymbol{\pi}}(p_{ik})\right]. \quad (36)$$

Our inversion program enables the user to choose the type of model parametrisation, which is either the full hierarchical model given by Eqs. (~~28) & (29~~30) & (31), or the simplified single-cone-bin inversion model ("Sicobi"-model) given by Eqs. (~~33) & (34~~35) & (36).

3.2 Solution to the inverse problem

Usually in Bayesian inference, the goal is to calculate the posterior pdf, given the measured data, i.e. the quantity

$$\boldsymbol{\pi}\left(\vec{\sigma}_{es}, \vec{\rho}, \vec{c}, \vec{x} \middle| \vec{d}\right) = \frac{\boldsymbol{\pi}\left(\vec{d}, \vec{\sigma}_{es}, \vec{\rho}, \vec{c}, \vec{x}\right)}{\boldsymbol{\pi}\left(\vec{d}\right)}. \quad (35$$

$$\left(\vec{p}_{el}, \vec{\rho}, \vec{c}, \vec{L} \middle| \vec{d}\right) = \frac{\boldsymbol{\pi}\left(\vec{d}, \vec{p}_{el}, \vec{\rho}, \vec{c}, \vec{L}\right)}{\boldsymbol{\pi}\left(\vec{d}\right)} \quad (37)$$

This can be interpreted as the inferences one may draw on the parameters in a model given measured data. The denominator on the right-hand side of Eq. (~~35~~37), also called the "evidence", can be rewritten as the data marginal of the posterior, i.e.

$$\boldsymbol{\pi}\left(\vec{d}\right) = \int\int\int\int \boldsymbol{\pi}\left(\vec{d}, \vec{\sigma}_{es}, \vec{\rho}, \vec{c}, \vec{x}\right)\left(\vec{d}, \vec{p}_{el}, \vec{\rho}, \vec{c}, \vec{L}\right) d\vec{\sigma}_{es}\vec{p}_{el}\, d\vec{\rho}\, d\vec{c}\, d\vec{x}. \quad (36$$

$$\vec{L} \quad (38)$$

As Eq. (36) basically describes an integration over the whole model parameter space, this may become such an extensive computation (especially when the number of model parameters is large), that it cannot be solved in a meaningful time. However, as the evidence usually is a fixed value, the left- and right-hand side of Eq. (~~35~~37) are merely scaled by a scalar and are thus proportional to each other, i.e.

$$\boldsymbol{\pi}\left(\vec{p}_{el}, \vec{\rho}, \vec{c}, \vec{L} \middle| \vec{d}\right) \propto \boldsymbol{\pi}\left(\vec{d}, \vec{p}_{el}, \vec{\rho}, \vec{c}, \vec{L}\right). \quad (39)$$

$$\boldsymbol{\pi}\left(\vec{\sigma}_{es}, \vec{\rho}, \vec{c}, \vec{x} \middle| \vec{d}\right) \propto \boldsymbol{\pi}\left(\vec{d}, \vec{\sigma}_{es}, \vec{\rho}, \vec{c}, \vec{x}\right). \quad (37)$$

This is the starting point of Monte Carlo Markov Chain (MCMC) methods.

3.2.1 The Metropolis-Hastings algorithm

The basic MCMC algorithm, which we also use in this study, is the Metropolis-Hastings (MH) algorithm (Hastings, 1970; Metropolis et al., 1953), which allows for the sampling of the joint pdf to obtain a quantitative sample. We note, however, that many different MCMC algorithms exist for various purposes and that the MH has no special status except for being comparatively simple to use and implement. An example of another MCMC algorithm in muon tomography can be found in Lesparre et al. (2017). The authors used a simulated annealing technique on the posterior pdf in order to extract the maximum a posteriori (MAP) model. As every simulated annealing algorithm has some type of MH-algorithm at its core, we

directly use the MH-algorithm in its original form such that we not only retrieve a point estimate but a pdf for the posterior parameter distribution. The algorithm is explained in detail by Gelman (2014), such that we only provide a short pseudo-code description.

Algorithm ~~1~~ (Metropolis-Hastings):

(1) Draw a starting model, $\vec{m}_0 = (\vec{\sigma}_{es,0}\vec{p}_{el,0}, \vec{\rho}_0, \vec{c}_0, \vec{x}_0\vec{L}_0)$, by drawing $\vec{\sigma}_{es,0}\vec{p}_{el,0}, \vec{\rho}_0, \vec{c}_0, \vec{x}_0\vec{L}_0$ from their respective prior pdfs and determine the log-pdf value of this model

(2) Until convergence:

a. Propose a new model according to $\vec{m}_{new} = \vec{m}_0 + c^2 J(\vec{0}, c^2\Sigma, S)$, where ~~$\Sigma$ is the matrix of prior variances and~~ $c = 2.4/\sqrt{D}$ ~~and~~in which D is the number of parameters~~.~~, and $c^2 J(\vec{0}, S)$ is a model change drawn from a multidimensional Gaussian distribution with mean vector of $\vec{0}$ and covariance matrix of S (i.e. matrix of prior variances of the model parameters)

b. Evaluate log-pdf value of $\vec{m}_{new}$ and calculate the odds ratio:
~~r~~$O = \exp(lp(\vec{m}_{new}) - lp(\vec{m}_0))$

c. Evaluate the acceptance probability, $p_A = \min(1, $~~$r$~~$O)$ and draw a number q from the uniform distribution U(0,1).

d. If $q < p_A$: sample $\vec{m}_{new}$ & set $\vec{m}_{new} \rightarrow \vec{m}_0$,
Else: sample $\vec{m}_0$

The advantage of this algorithm, compared to a "normal" sampling, lies in its efficiency. It is often not possible, or even reasonable, to probe the whole model space, as the largest part of the model space is "empty", where the pdf-value of the posterior is uninterestingly small. The fact that regions of high probability are scarce, and this becomes worse in high dimensional model spaces, is known as the "curse of dimensionality" (Bellman, 2016). MCMC algorithms (including the here presented MH-algorithm) allows one to focus on regions of high probability, and therefore we are able to construct a reliable and representative sample of the posterior pdf. We again refer to Gelman (2014) for a discussion of why the MH-algorithm converges to the correct distribution and why we may use samples that were gained this way to estimate the posterior probability density.

3.2.2 Assessing convergence, mixing, and retrieving the samples

The above stated advantages, however, come at a price. First and foremost, we must ensure that the algorithm advances fast enough, but not too fast, through the model space. This is mainly controlled by the proposal distribution $J(0, $~~$c^2\Sigma$~~$S)$, which is taken to be a multivariate Gaussian distribution. Ideally, the covariance matrix of the proposal distribution ~~Σ~~S is equal to the covariance structure of the posterior pdf. We acknowledge that at the start of the algorithm one generally has no idea what this looks like, but we assume that a combination of the prior variances is a reasonable starting point. After a certain number

of steps, it is possible to approximate the covariance matrix of the proposal distribution with the samples taken up to at this point.

A second crucial point is the presence of a warm-up period. The starting point, which usually lies in a region of high prior probability, does not necessarily lie in a region of high posterior probability. The time it takes to move from the latter to the former is exactly this warm-up. This can usually be visualised by a trace plot, e.g. Fig. 4, in which the value of a parameter is plotted against the number of iterations. After this warm-up phase, the algorithm can be run in operational mode and "true" samples can be collected.

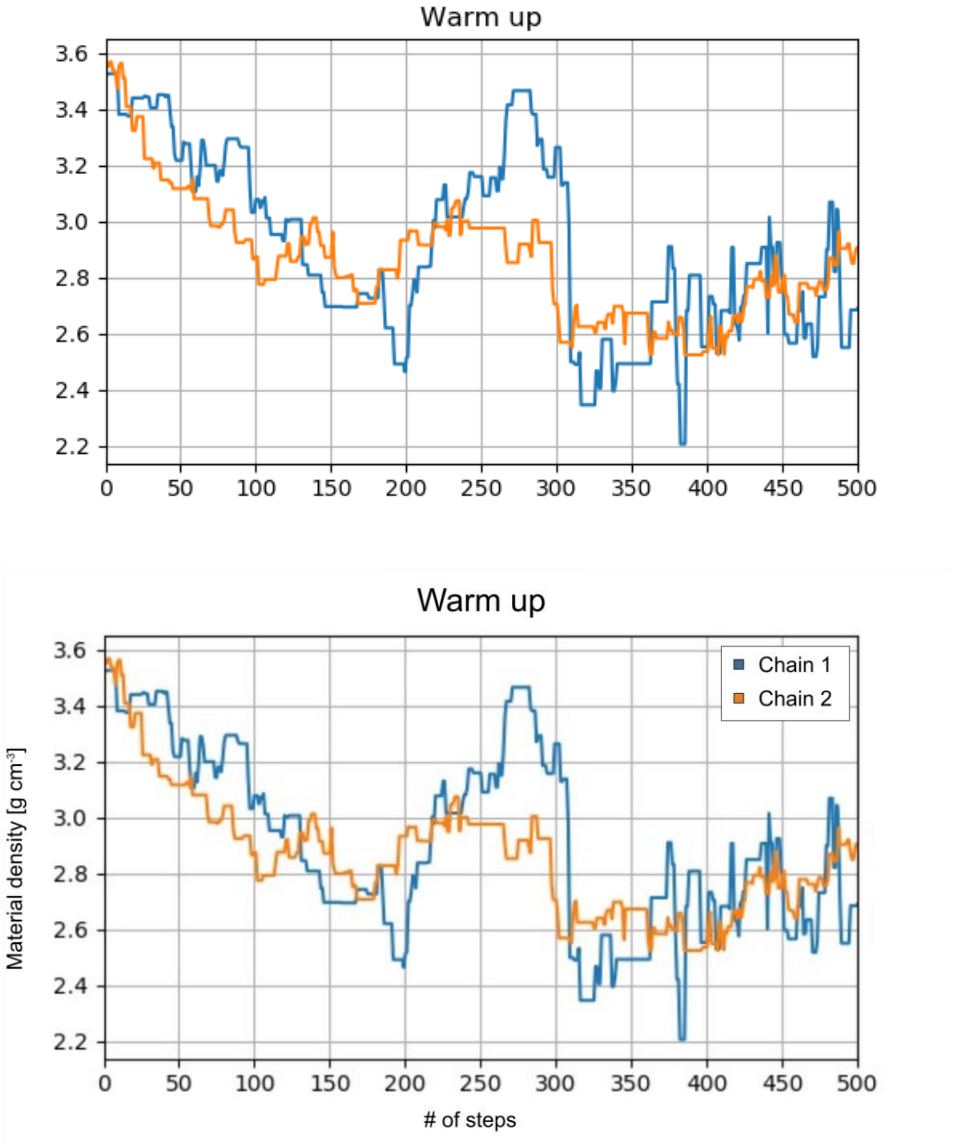


Figure 4: Example of a trace plot (2 independent chains; blue and orange) of a MH run with 500 draws. This plot shows the parameter value (~~y-axis;~~ here material density ~~in $g\ cm^{-3}$~~) vs. # of steps ~~(x-axis)~~ of a collection of cones in which we (Nishyiama et al., 2017) knew that only rock is present. This is a calculation that is included in the code base. The warm-up phase of this MH algorithm takes roughly 150 simulations indicated by the subsequent oscillation around a parameter value of $\sim 2.7 - 2.8$ ~~$g\ cm^{-3}$~~ $\mathrm{g\ cm^{-3}}$.

As in a Markov Chain the actual sample is dependent on the last one, we need a criterion to argue that the samples created in that way really represent "independent" samples. Qualitatively, we may say that if the Markov Chain forgets the past samples fast enough, then we may sooner treat them as independent from each other. Gelman (2014) suggests that in order to

assess this quantitatively, multiple MH-chains could be run in parallel and statistical quantities within and between each chain are analysed. For a detailed discussion thereof, we refer the reader to Appendix C.

Once a satisfying number of samples has been drawn from the posterior pdf, a marginalisation of the nuisance parameters can be done by looking at the parameters of interest only. These samples may then be treated like counts in a histogram, i.e. distributional estimates, or simply the interesting statistical moments, such as mean and variance, can be obtained.

3.3 Construction of the bedrock-ice interface

The main analysis programme allows us to export all parameters either as a full chain dataset, where every single draw is recorded, or as a statistical summary (i.e. mean and variance). Both are then converted to point data, i.e. (x, y, z) – data. For the subsequent construction of the interface between rock and ice (see ~~section 5.2~~Sect. 4 and Fig. 1 for the presentation of the test experiment) we only need the full-chain point-data. In the present study we restrict ourselves to a probabilistic description until the bedrock positions within a cone. It would also be possible to treat the bedrock construction within a Bayesian framework, however this would go beyond the scope of this study and is therefore left for a future adaption of the code. Nonetheless, in order to construct a surface, we rely on deterministic methods, which are explained in detail in what follows.

3.3.1 Interpolation to a grid

The "modelviewer.py" routine is able to read datasets from different detectors (which are saved as JSON-files) and computes for each cone the statistic, which the user is interested in (see "sigma" entry in program). Thus, it is possible to use the mean or, for example, the $+1\ \sigma$ position of each cone. From here onwards this point cloud is named H and contains one interface position (x, y & z coordinates) per cone. These are shown as triangles (▲) in Fig. 5.

As a second step, the programme interpolates this point cloud in a bilinear way to a rectangular grid with a user specified cell size, Δ_{cs}. This grid can be described by a matrix $P \in \mathbb{R}^{r \times c}, \mathbb{R}^{N_{row} \times N_{col}}$, where $r N_{row}$ and $c N_{col}$ are the number of rows and columns (i.e. the number of y- and x-cells, needed to cover the whole grid). The procedure is similar to the bilinear interpolation of Lagrangian markers (that carry a physical property) to a (fixed) Eulerian grid in geodynamical modelling (see Gerya, 2010, p. 116), with the difference that our physical property is the height of the ice-bedrock interface (Fig. 1).

We could also have fitted a surface through the resulting point cloud. However, by formulating this surface as a matrix we gain access to the whole machinery of linear algebra. Moreover, P can directly be interpreted as a rasterised DEM, which can be easily loaded and visualised in any GIS software. Thus, from a modular design perspective we think that the matrix formulation has more advantages than drawbacks. The bilinear interpolation is shown in more detail in Fig. 5.

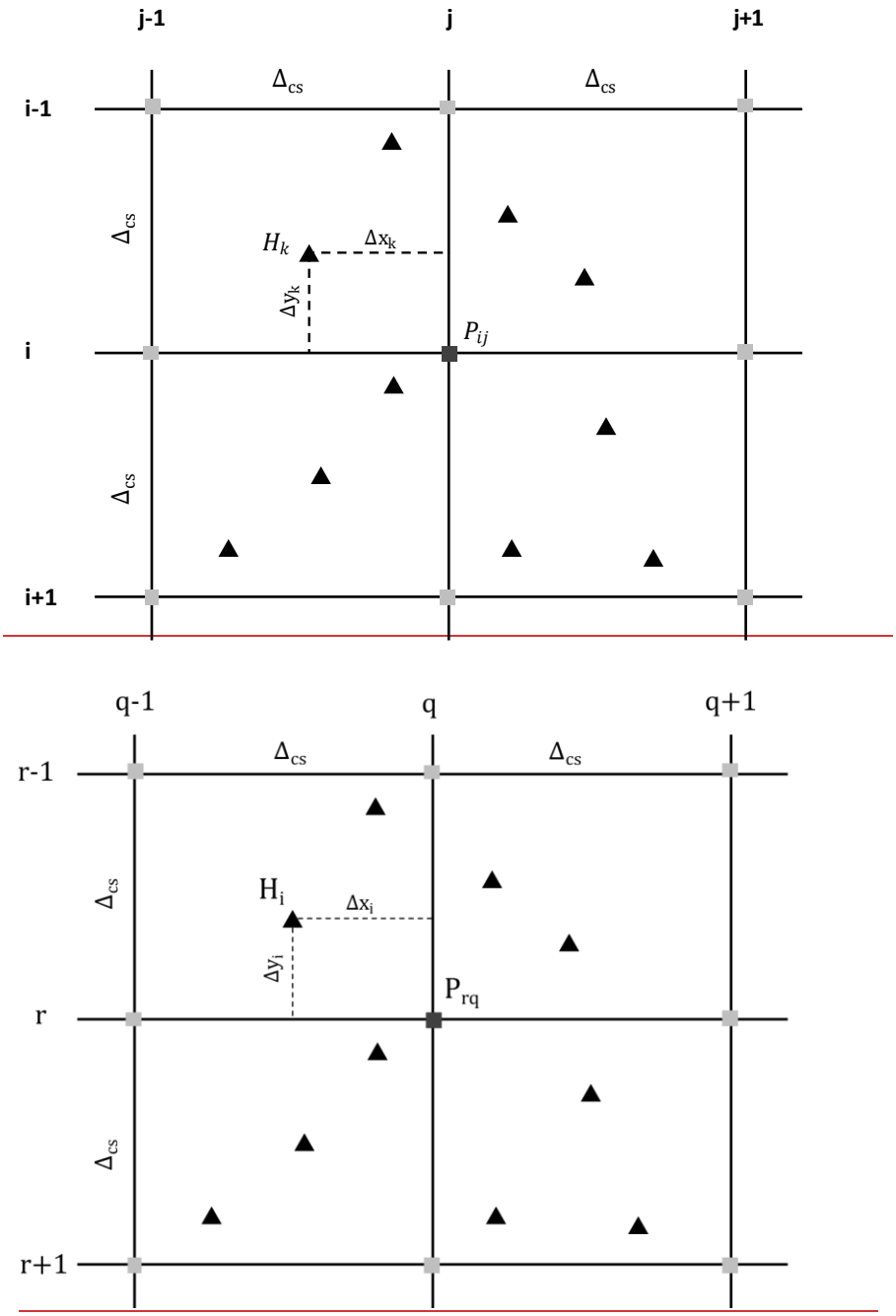


Figure 5: Two-dimensional stencil, used to summarise the bilinear interpolation of interface positions within cones (~~H_k~~H_i, ▲) to a fixed grid (~~P_{ij}~~P_{rq}, ■) with a user-defined cell size Δ_{cs}. Every interface position within a $\pm\Delta_{cs}$ interval contributes to the grid height ~~P_{ijrq}~~.

In order to calculate the height at a grid point, ~~P_{ij}~~P_{rq}, one has to form a weighted sum over the entire cone interface positions within a $\pm\Delta_{cs}$ interval, i.e.

$$\cancel{P_{ij} = \frac{\sum_k w_k H_k}{\sum_i w_i}}, \quad \cancel{(38)}$$

$$P_{rq} = \frac{\sum_i w_i H_i}{\sum_i w_i}, \quad (40)$$

where H_i is the height of the bedrock-ice interface in the i-th cone and the weights, $\cancel{w_k}w_i$, are given by

$$w_i = \left(1 - \frac{\Delta x_i}{\Delta_{cs}}\right) * \left(1 - \frac{\Delta y_i}{\Delta_{cs}}\right). \quad (41)$$

$\cancel{w_k = \left(1 - \frac{\Delta x_k}{\Delta_{cs}}\right) * \left(1 - \frac{\Delta y_k}{\Delta_{cs}}\right)}$. In Eq. (41), Δx_i and Δy_i are the horizontal and vertical distances from the interpolated grid point P_{rq}.

$\cancel{(39)}$

3.3.2 Damping & Smoothing

The concept of damping usually revolves around the idea to force parameters to a certain value (e.g. in deterministic inversion by introducing a penalty term in the misfit function for deviations from that value). From a Bayesian viewpoint this would be accomplished by setting the prior mean to a specific value. In our code we implemented this idea by allowing the user to read a DEM and a "damping weight" to the code (see "fixed length group" in code). The programme effectively computes a weighted average between the bedrock positions within the cones and a user defined DEM. The higher the chosen damping weight, the more the resulting interface will match the DEM, when pixels overlap.

The matrix formulation also enables us to use a further data processing technique without much tinkering. As geophysical data are often quite noisy, a standard procedure in nearly every geophysical inversion is a smoothing constraint. This effectively introduces a correlation between the parameters and forces them to be similar to each other. From a Bayesian perspective, we could have achieved this correlation by defining a prior covariance matrix of the thickness parameters, such that neighbouring cones should have similar thicknesses (which makes sense as we do expect the bedrock-ice interface to be relatively continuous; Fig. 1). As we work with independent cones in this study, we leave the exploration of this aspect open for a future study. Nevertheless, we offer the possibility in our code to use a smoothing on the final interpolated grid. This is achieved by a convolution of a smoothing kernel, K (see Appendix D for details), with the surface matrix P, which results in a smoothed surface matrix

$$P^S = K \cancel{*} \circledast P. \quad (\cancel{40}42)$$

Please note that the $\cancel{*}\circledast$ operator in Eq. ($\cancel{40}42$) denotes a convolution. In index notation the advantage of the linear algebra formalism becomes clear, as P^S can be expressed by

$$\cancel{P^S_{ij} = \sum_{k=-s}^{s} \sum_{l=-s}^{s} K_{k+s+1,l+s+1} P_{i+k,j+l}} \; P^S_{rq} = \sum_{k=-s}^{s} \sum_{l=-s}^{s} K_{k+s+1,l+s+1} P_{\mathrm{r}+k,\mathrm{q}+l}. \quad (43)$$

The user is free to choose the number of neighbouring pixels, *s*, across which the programme performs a smoothing. As a related matrix we use an approximation to a Gaussian kernel, which corresponds to a Gaussian blur in image processing. Whereas "smoothing" is a general term used in the geophysical community where such a process is used with an aim to force a correlation of parameters, in our case where the parameters describe a surface (Fig. 1), the convolution effectively smooths the surface, i.e. removes small-scale variations.

Finally, we added a checkbox to our code to allow it to change the order of the damping and smoothing operations. Sometimes when a strong damping is necessary, this may result in rather unsmooth features at DEM boundaries, such that it makes sense to perform a smoothing only afterwards.

4 Main modules of SMAUG

Our toolbox, SMAUG, contains several subprograms, which are executed separately. This allows the user to inspect intermediate results without any difficulty. We also tried to keep the intermediate results as portable as possible, by using JSON-files, as often as possible. Here we explain, in logical order, the rational of the submodules (a detailed user manual is separately available):

MATERIALIZER.py

This subroutine allows the user to create their own material that will be used in the subsequent model builder. The user may choose a density (either from data or directly insert mean and standard deviations) and a composition (also either from data or from the list of Groom et al., 2001).

DATA_BINNING.py

As the name suggests, this subroutine is used to spatially bin the recorded track data. The bin data (i.e. the output hereof) is then fed to the model builder.

MODEL_BUILDER.py

The model builder takes the bin data and the materials as inputs and allows the user, with help of DEMs, to allocate data and materials to certain cones. This is basically the spatial setup of the model. The resulting model file is then provided to the inversion code.

INVERSION.py

This is the main module in SMAUG, providing the functionality to perform a MCMC algorithm on the probabilistic model created with MODEL_BUILDER.py. It also includes several analysis tools to assess MCMC performance.

~~***MODEL_VIEWER.py***~~

~~The model viewer allows us to visualise the interface results, obtained and exported by INVERSION.py. It also has the functionality to dampen and smooth the resulting surfaces.~~

5 Model Verification

In this section we present examples of how the model can be employed, what it calculates and how the output is structured. We proceed by verifying, in a first step, that the physical models employed in this work yield results which are numerically consistent with the results of calculations from other studies. We will compare our results with reported values from literature in Ch. 5.1. We do this because we do not change the parameters of the flux model (except the height scaling, which has been verified by Nishiyama et al. (2017), and since the energy loss calculations is based on equations that stem from different frameworks.

In a second step, we will benchmark the results obtained by this code from real data against previously published results. For this purpose, we will re-analyse the raw data from Nishiyama et al. (2017). This is a study (see also Fig. 1) that was conducted in the Central Swiss Alps in a railway tunnel that featured a glacier (part of the Great Aletsch glacier) above. Our goal there was to estimate the thickness of the overlying glacier and thus the subsurface structure of the ice-bedrock interface. The respective calculation and discussion thereof are presented in Ch. 5.2.

5.1 Verification of energy loss calculations

The energy loss model that we use in our code generally reproduces the literature values well (below 1%) across the different energy loss processes and relevant energies. In Fig. 6 we present the energy loss calculations for each energy loss process (i.e. ionisation, bremsstrahlung, pair-production and photonuclear interactions) across energies from 10 MeV to 100 TeV for silicon.

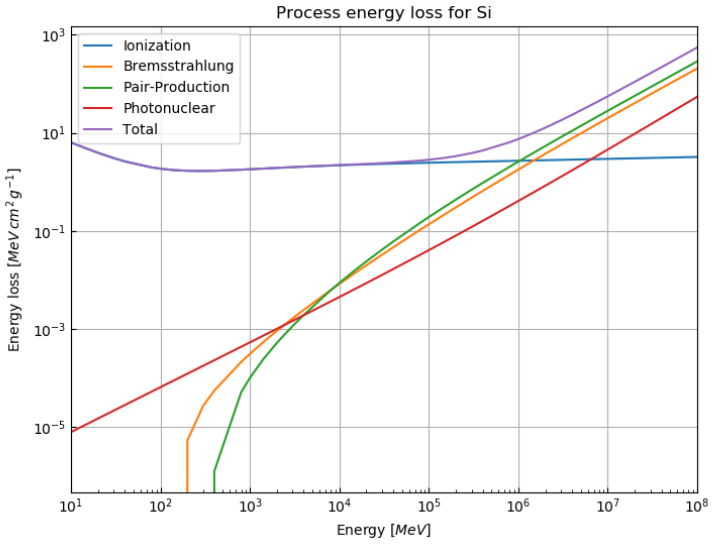


Figure 6: Log-log plot of the stopping power of the different energy loss processes for silicon. At ~ 10 GeV the radiative processes (i.e. bremsstrahlung, pair-production and photonuclear interactions) reach around 1 % of the total stopping power. At a few hundred GeV (at the so-called "critical energy") the radiative processes start to become dominant over the ionisation losses.

The overall characteristics between the different elements are the same with minor differences regarding the position of the critical energy and the 1 % - radiative point. In Fig. 7 we show the relative error of our calculations to the tabulated values from Groom et al. (2001) for the whole energy range.

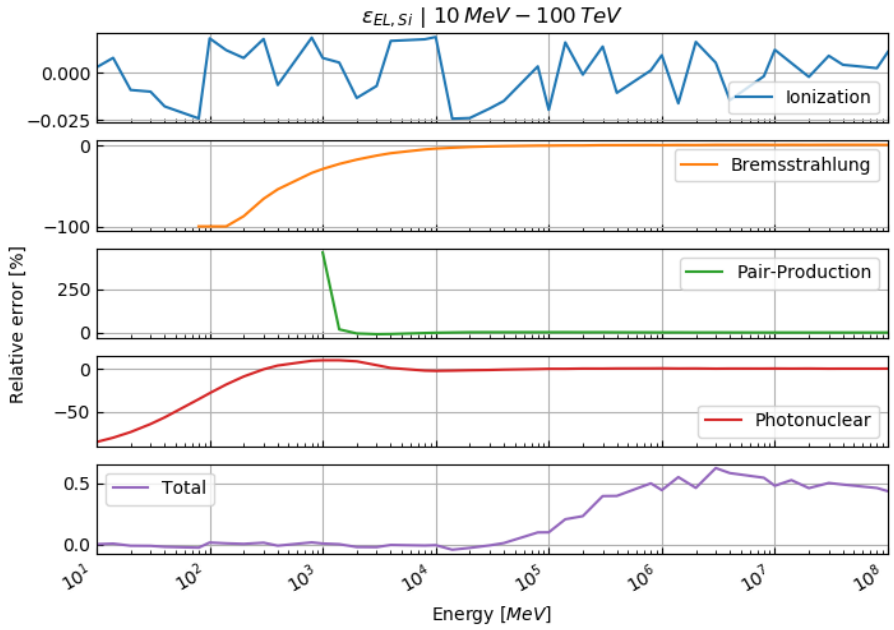


Figure 7: Relative error of our energy loss calculations compared to the tabulated values from Groom et al. (2001) for silicon. Ionisation losses agree very well with the literature values (within 0.025 %). At low energies the relative errors of the radiative processes are large and converge to a value close to 0 towards higher energies, resulting in a relative error on the total energy loss of around 0.5 % compared to literature.

We note that the energy losses by ionisation are reproduced very well over the entire energy range. We also note that the relative error on the radiative energy losses is rather large below 10 GeV. This does, however, not introduce a major bias, because below this energy, radiative energy losses are negligible compared to ionisation losses, as can be seen in Fig. 6. Furthermore, the related errors are in an acceptable range at the energy level at which radiative losses begin to become noticeable (i.e. around 100 GeV). This can be seen in Fig. 7, in the sense that the total relative error remains well bounded within 0.5 %. In the ionisation domain (i.e. below 100 GeV) the total relative error is dominated by the ionisation relative error, whereas above this energy level the relative errors on radiative losses start to prevail. A close-up of this energy range is given in Fig. 8.

There are different sources and circumstances that contribute to the error in the different energy losses processes. The scatter of the relative ionisation-loss error around 0 with a rather small deviation can be viewed as simple rounding errors. The errors on the radiative processes, however, seem to be of a more systematic nature. We explain this behaviour through a

different numerical integration scheme in Eq. (7), which tends to systematically under-/overestimate the true value, especially when the integrand comprises exponential functions. Whereas we used a Double Exponential Integration scheme (see Takahasi & Mori, 1974), the integration scheme from Groom et al. (2001) is not discernible. However, as the relative errors on the processes of energy loss remain well within the theoretical uncertainties, (see Ch. 3.1.3), we consider, that our calculation accurately reproduces the literature values for elements.

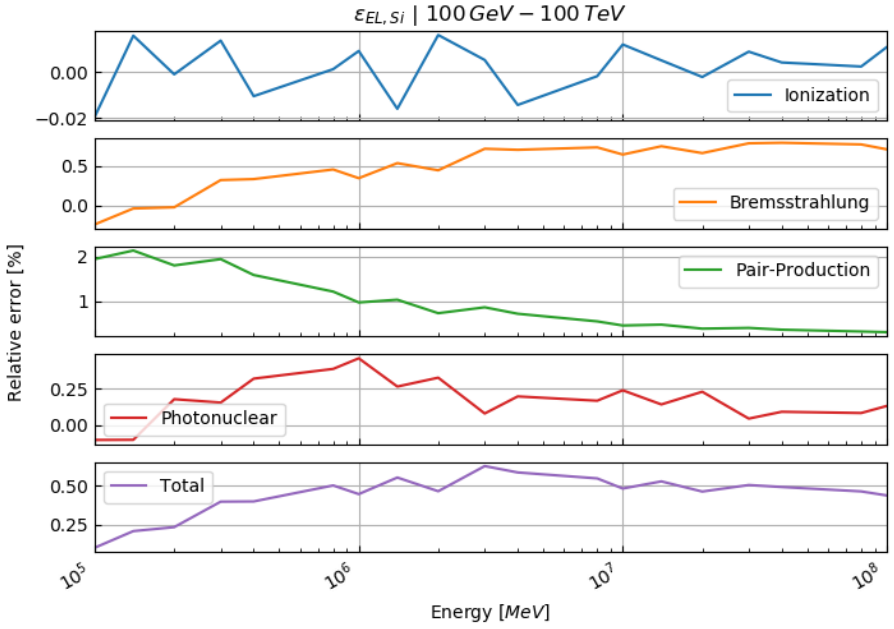
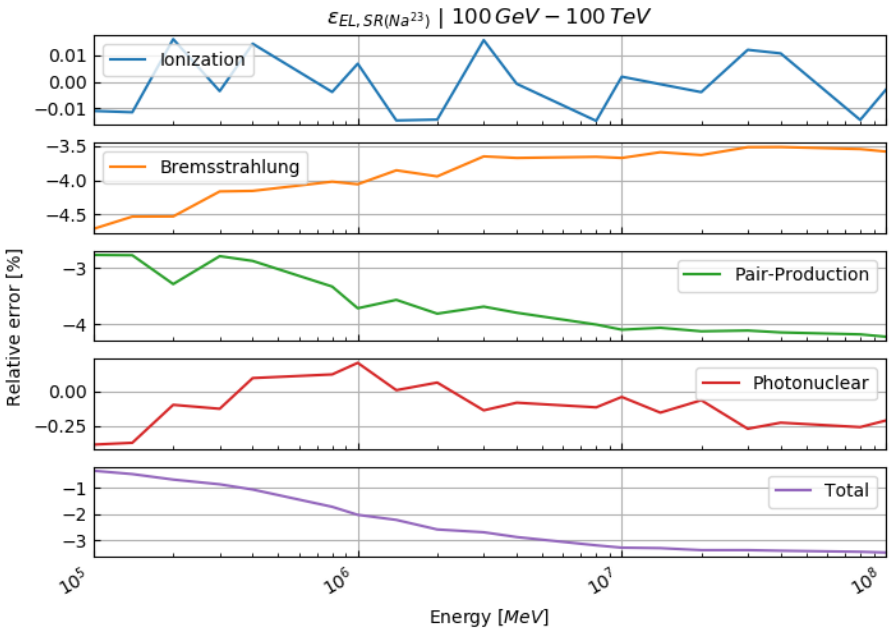


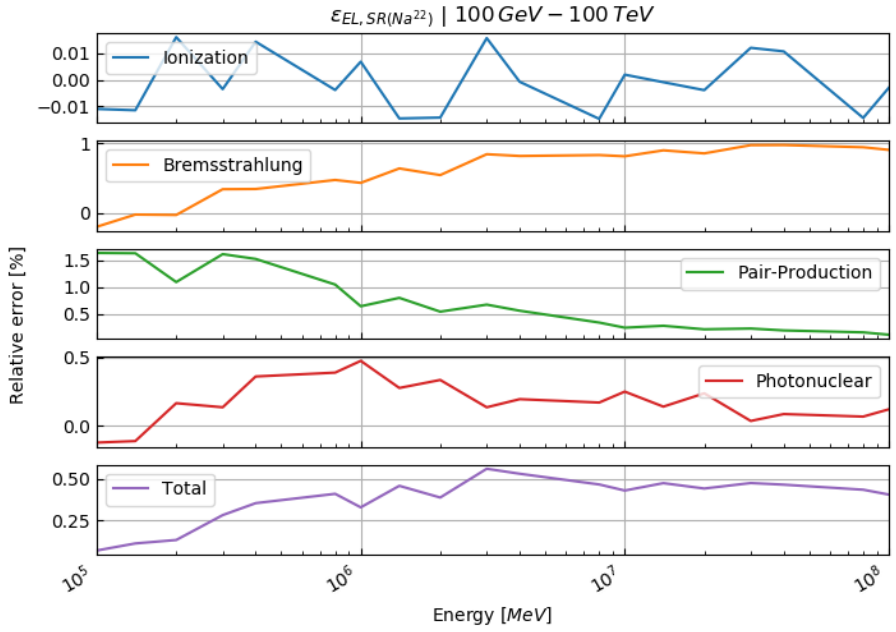
Figure 8: Relative error of our energy loss calculations for silicon compared to the tabulated values from Groom et al. (2001) at higher energies (100 GeV – 100 TeV). The relative errors remain bounded within their theoretical uncertainties (see Ch. 3.1.3).

The above calculations were performed for pure silicon. The respective figures for other four important elements in the Earth's crust (Al, Fe, Ca & O) can be found in Appendix E. Those elements are, however, not representative for any real material encountered in geological applications. For this reason, we compiled the same computations for four selected, geologically important compounds (SiO_2, $CaCO_3$, Standard Rock, ice) that are also shown in Appendix E. We summarise, that with the exception of Standard Rock, all calculations yield results that are similar to the silicon calculation above. The discrepancy for Standard Rock stems from its inconsistent definition, with respect to the different parameters. In particular, the "Standard Rock" according to Lohmann et al. (1985) has an atomic number Z of 11 (i.e. sodium) and an atomic weight A of 22, which yield the characteristic parameter values of $\langle Z/A \rangle = 0.5$ and $\langle Z^2/A \rangle = 5.5$ respectively. Note that Groom et al. (2001) list sodium as the only constituent of a standard rock. However, this material cannot be modelled by any mixture of pure elements, as common sodium consists of one neutron more and thus has a higher atomic weight (i.e. $A_{^{Na}} = 23$).

~~Consequently, the use of standard sodium would lead to different characteristic parameter values, i.e. $\langle Z/A \rangle = 0.478$ and $\langle Z^2/A \rangle = 5.263$, thus leading to an inconsistency. This is often conveyed by the phrase that standard rock "is not quite sodium" (Groom et al. 2001, p.203). In order to circumvent this problem, we advocate the exchange of $^{23}_{11}Na$ with its $^{22}_{11}Na$ isotope, that would lead to the characteristic parameter values $\langle Z/A \rangle = 0.500$ and $\langle Z^2/A \rangle = 5.501$, which are much closer to the actual definition of standard rock. For this reason, we extended the element/compound list, (which is available from http://pdg.lbl.gov/2019/AtomicNuclearProperties/expert.html) by the $^{22}_{11}Na$ isotope, assuming that all parameters are equal to the ones from $^{23}_{11}Na$. Additionally, we redefined the standard rock (i.e. material number 281 in the list) to consist only of $^{22}_{11}Na$. With this change, standard rock does not need any more special treatment and can be calculated in a way that is consistent to all other compounds. Furthermore, the relative error between the tabulated values and our modified calculation falls in line with the calculations for the other compounds and elements (Figs. 9 and 10).~~



~~Figure 9: Relative error of our energy loss calculations for a standard rock consisting of $^{23}_{11}$Na, compared to the tabulated values from Groom et al. (2001) at higher energies (100 GeV – 100 TeV).~~



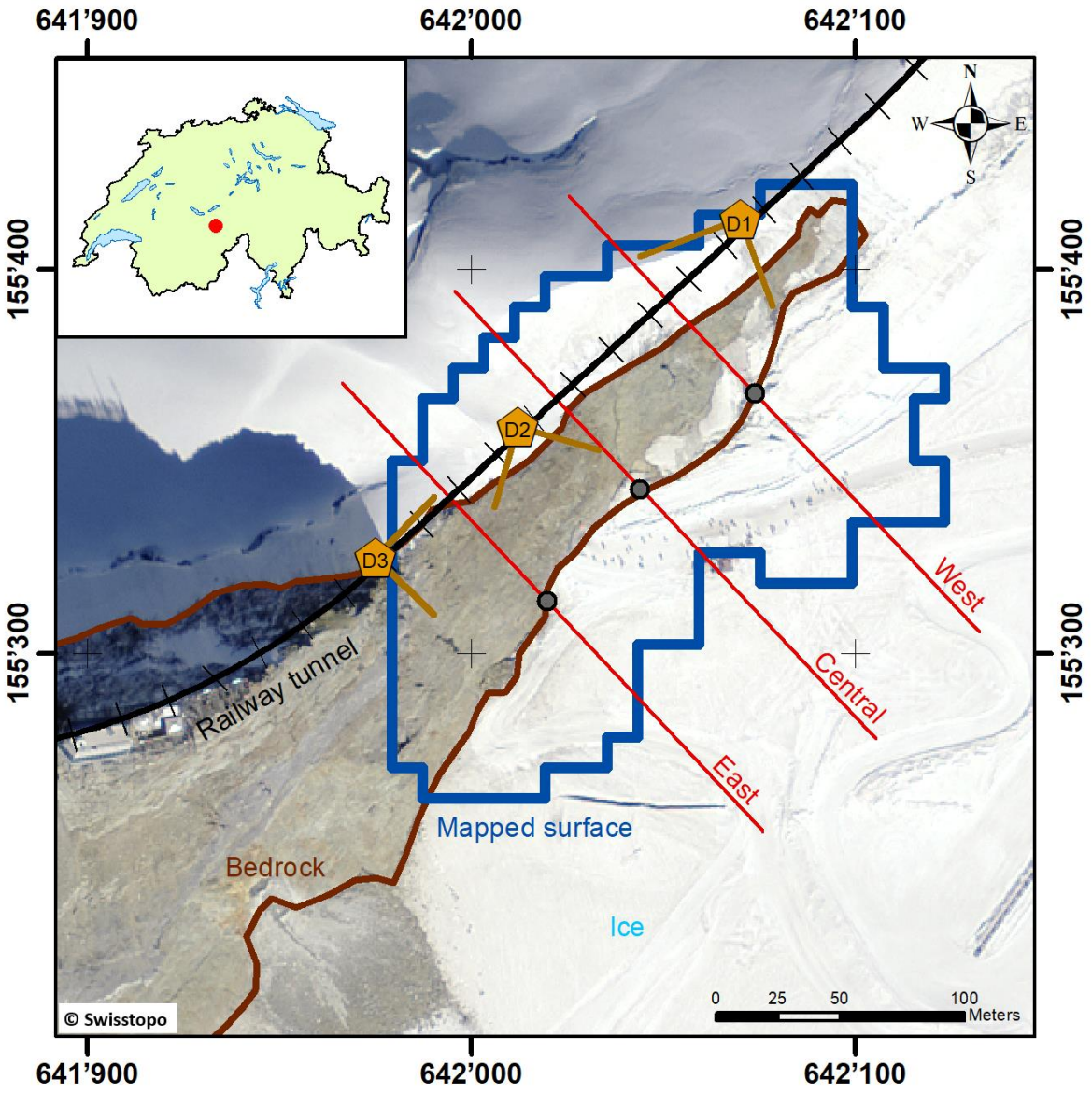
~~**Figure 10: Relative error of our energy loss calculations for a standard rock consisting of $^{22}_{11}$Na, compared to the tabulated values from Groom et al. (2001) at higher energies (100 GeV – 100 TeV).**~~

~~5.2 Verification of the bedrock-ice interface reconstruction~~

In this part we test the presented reconstruction algorithm on previously published data. For this purpose, we compare our calculations to the ones already published in the study by Nishiyama et al~~.~~, (2017), where the goal was to measure the interface between the glacier and the rock, in order to determine the spatial distribution of the rock surface (also below the glacier). ~~We could access~~ This study was conducted in the ~~Railway Tunnel to install~~Central Swiss Alps in a railway tunnel that featured a glacier (part of the ~~muon detectors beneath the ice.~~Great Aletsch glacier) above. A situation sketch is shown in Figs. ~~1 and 11~~1. For a detailed verification of the energy loss calculations we refer to Appendix E.

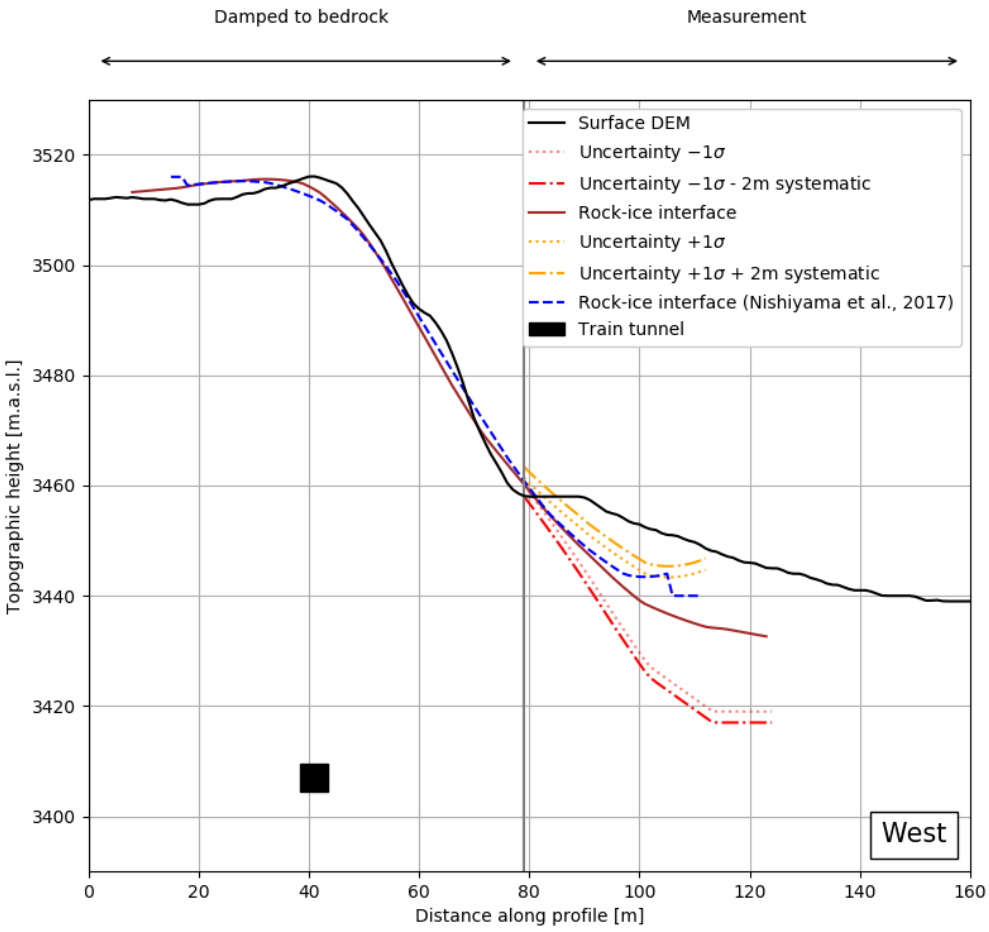
The results shown below (Figs. ~~12 – 14~~7 – 9) represent the bedrock-ice interface interpolated to an 8-metre grid, which was first damped (weight 8) and then smoothed (2 grid pixel~~).~~, i.e. $s = 2$ in Eq. 43). We assess the goodness of fit according to the three cross-sections (East, Central, West), that are shown in Fig. ~~10~~1. The crosscuts are nearly perpendicular to the train tunnel and roughly 40 ~~m~~m apart from each other. ~~Figures 12 to 14 show the three cross sections in detail.~~ In every plot, we also indicate the solution from Nishiyama et al. (2017). Please note that we added a systematic error of 2 ~~m~~m to the uncertainty planes, as the DEM we are working with has itself an uncertainty of $\pm$ 2 ~~m~~m. The dash-dotted lines mark thus

the most conservative error estimate. Moreover, we highlighted the parts of the cross-section that had either been damped to the bedrock DEM or that have been solely resolved by the measurement (see "damping marker" in ~~Fig. 11~~Figs. 6 – 8).



~~**Figure 11: Overview map of the study area at Jungfraujoch. The brown line separates the visible bedrock in the DEM from the glacier part ("Ice"). The three profiles (East, Central & West) are depicted with a red line, on which the damping marker is shown by a grey point. The extent of the reconstructed bedrock-ice interface is shown by the blue area. Additionally, the three detector positions (D1, D2 & D3) are shown by orange pentagons, including their viewfield. Basemap: Orthophotomosaic Swissimage, © Federal Office of Topography swisstopo.**~~

~~Figure 12~~Figure 6 shows the western profile, where our bedrock-ice interface and the one from the previous study agree well and both lie within the given error margins. The lack of fit in areas where the steepness changes rapidly (i.e. around 40 ~~m~~m and 80 ~~m~~m) can be explained as a smoothing artefact. Towards the end of the profile, the decreasing data coverage becomes evident as the uncertainties rise. This effect can also be seen in the jagged behaviour of the interface curves around 100 ~~m~~m to 120 ~~m~~m, hinting at the effect where the interpolation has occurred with few data.


Damped to bedrock
Measurement
Surface DEM
Uncertainty −1σ
Uncertainty −1σ - 2m systematic
Rock-ice interface
Uncertainty +1σ
Uncertainty +1σ + 2m systematic
Rock-ice interface (Nishiyama et al., 2017)
Train tunnel
West
Topographic height [m.a.s.l.]
Distance along profile [m]
3520
3500
3480
3460
3440
3420
3400
0
20
40
60
80
100
120
140
160

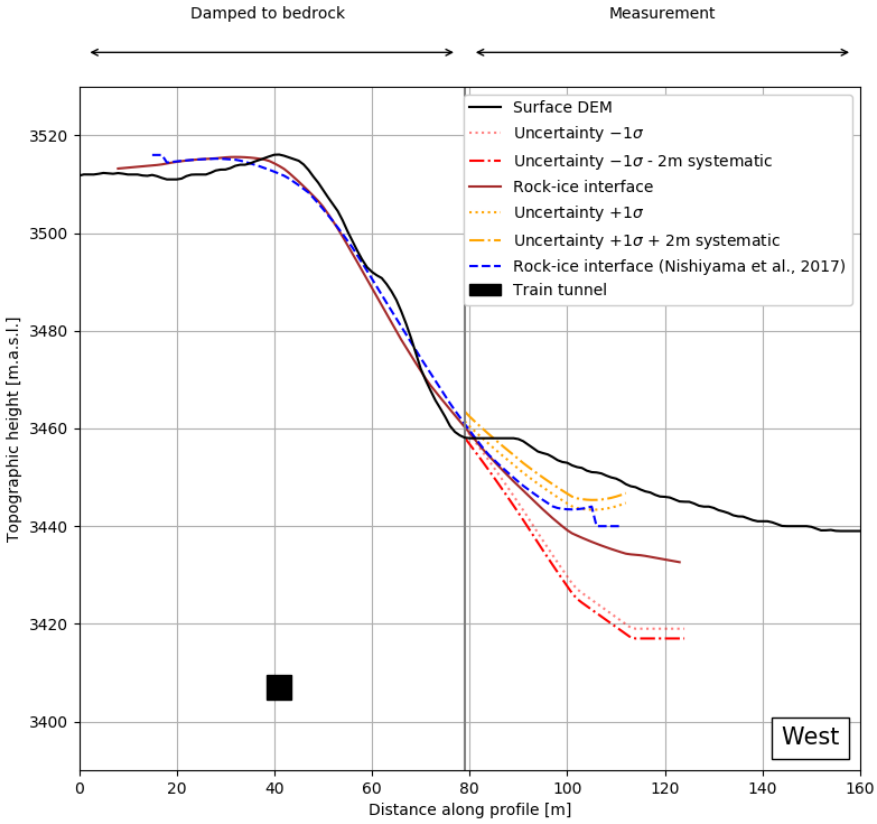
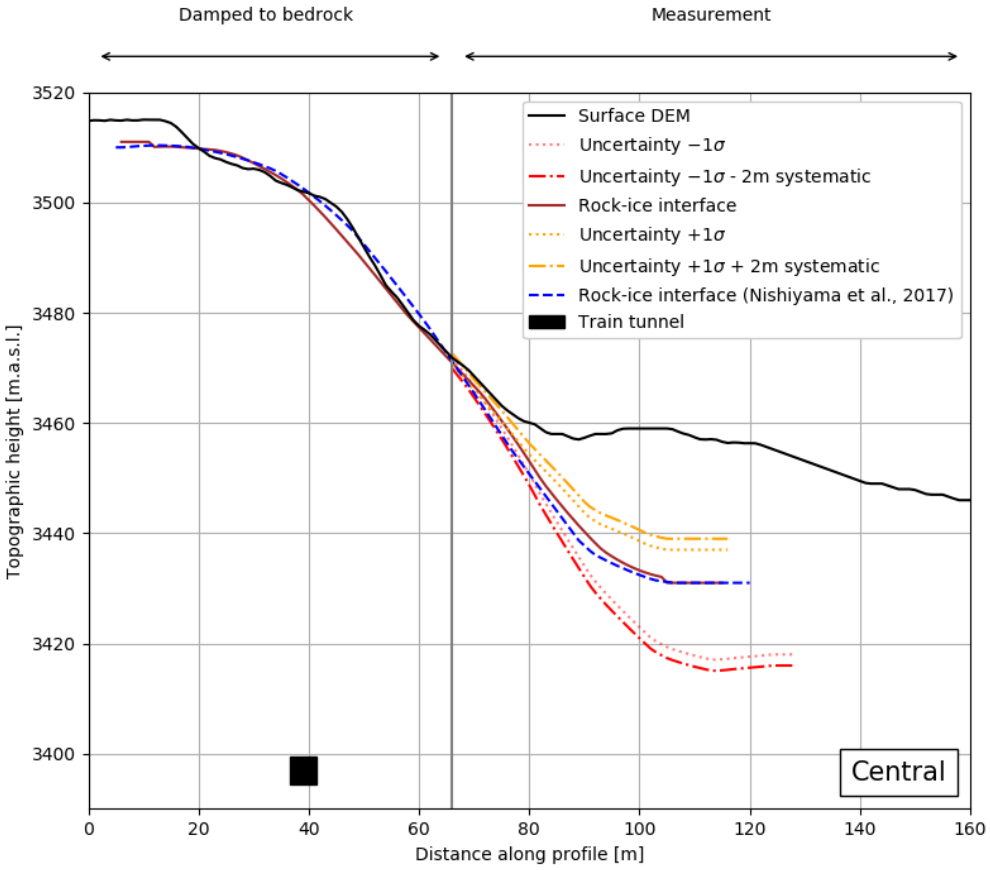


Figure 6: Western cross-section. The brown and dashed blue line indicate the ice-bedrock interface solutions of this study and the one from Nishiyama et al. (2017), respectively. 1 σ-error margins are shown in yellow (upper) and red (lower). The dotted margins encompass only the statistical variation of the interface position, whereas the dash-dotted include a ± 2 ~~m~~m systematic error which stems from the inherent DEM-uncertainty. For completeness we also show the position of the railway tunnel as a black square.

Figure ~~13~~7 presents the central profile. Similar to the western profile (Fig. ~~12~~6) the fits match quite well and are within the error margins. It may be possible that the point where the actual bedrock begins might be further down (i.e. ~80 ~~m~~m instead of 65 ~~m~~m). Here we used the same DEM and aerial photograph as in the previous study. This means that newer versions might be available that show more bedrock (due to the glacial retreat as a response to global warming).


Damped to bedrock
Measurement
Surface DEM
Uncertainty −1σ
Uncertainty −1σ - 2m systematic
Rock-ice interface
Uncertainty +1σ
Uncertainty +1σ + 2m systematic
Rock-ice interface (Nishiyama et al., 2017)
Train tunnel
Central
Topographic height [m.a.s.l.]
Distance along profile [m]
3520
3500
3480
3460
3440
3420
3400
0
20
40
60
80
100
120
140
160

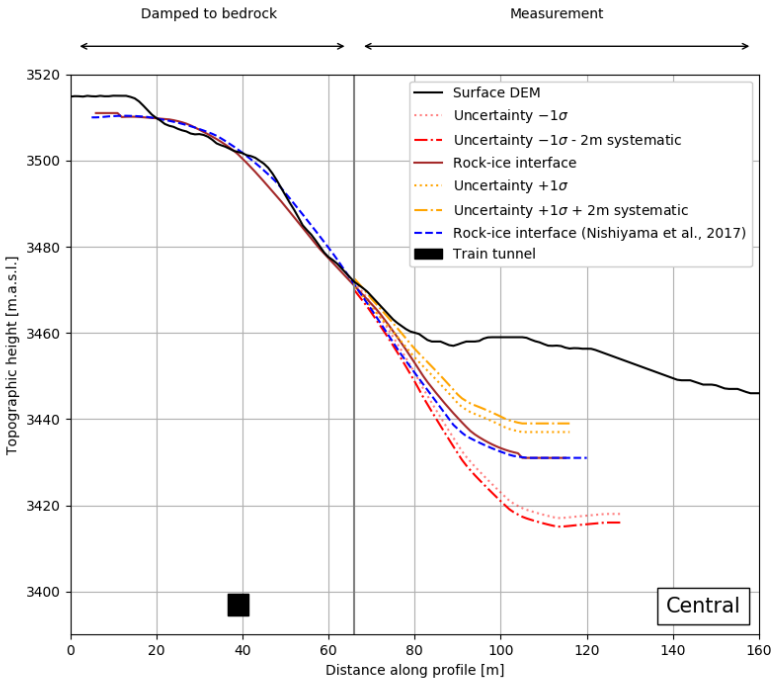
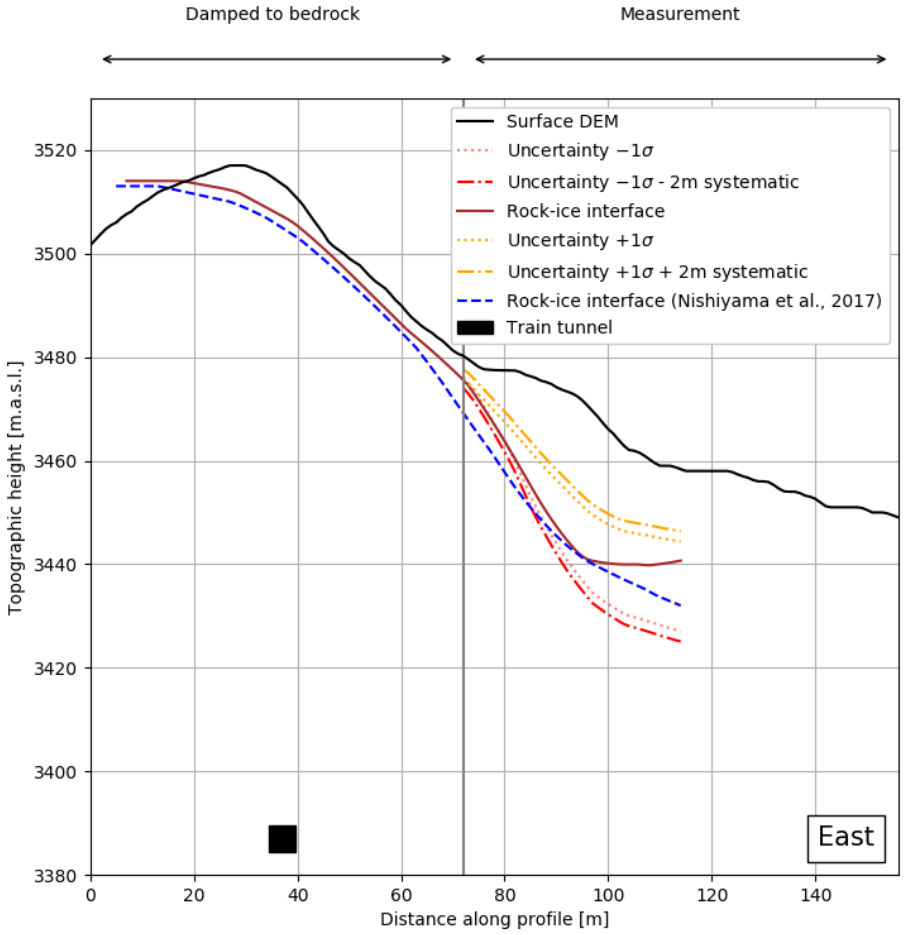


Figure 7: Central cross-section. The brown and dashed blue line indicate the ice-bedrock interface solutions of this study and the one from Nishiyama et al. (2017), respectively. 1 σ-error margins are shown in yellow (upper) and red (lower). The dotted margins encompass only the statistical variation of the interface position, whereas the dash-dotted include a ± 2 ~~m~~m systematic error, which stems from the inherent DEM-uncertainty. For completeness we also show the position of the railway tunnel as a black square.

~~**For completeness we also show the position of the railway tunnel as a black square.**~~

The eastern profile is shown in Fig. ~~14~~8. One sees that the results from this study are internally consistent. The surface from the previous study plunges down earlier with respect to the surfaces calculated here. This may in fact be a damping effect, as the bedrock-ice interface from Nishiyama et al. (2017) has not been constrained to the bedrock (via damping) and thus plunges down before the damping mark at ~72 ~~m~~m. Still, the two surfaces agree within 5 ~~m~~m, which we consider as acceptable.



In all three results (Fig. 6, 7 & 8) it can be seen that the reconstructed surfaces in the bedrock region are following the DEM within 5 m – 10 m. The reason for this deviation may be explained by the smoothing. At the beginning of this section we explained that the reconstructed interface has been smoothed by 2 grid pixels, which corresponds on an 8-metre grid to a smoothing of 16 m to each side. This is also valid for the direction perpendicular to the cross-sections shown in Figs. (6 – 8). Thus the over-/underestimation can very well be a smoothing effect. The behaviour of the reconstruction in the Western cross-section (Fig. 6) further supports this explanation, as we see at 40 m an underestimation and at 80 m an overestimation of the height. This is a typical behaviour of smoothing around “sharper” edges.

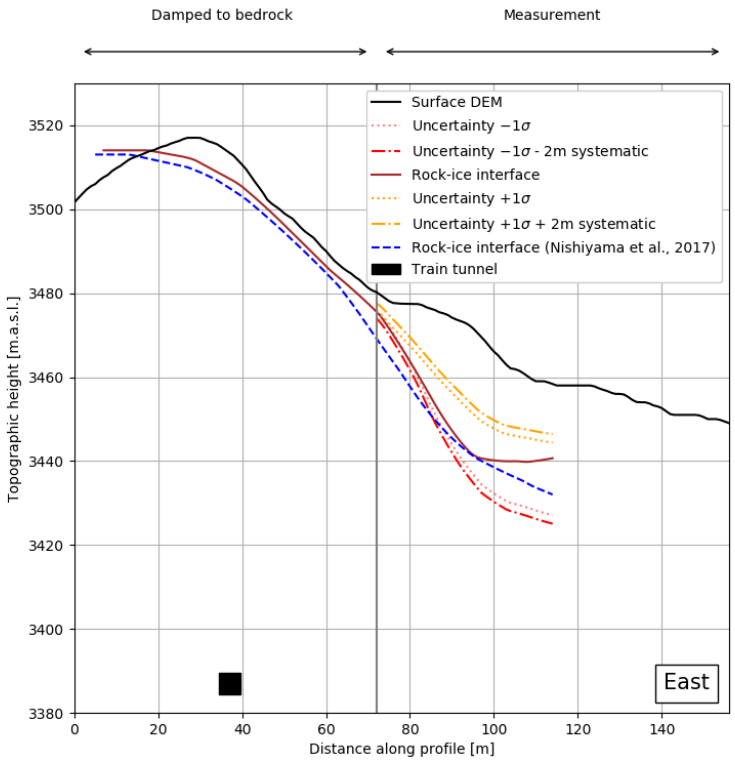


Figure 8: Eastern cross-section. The brown and dashed blue line indicate the ice-bedrock interface solutions of this study and the one from Nishiyama et al. (2017), respectively. 1 σ-error margins are shown in yellow (upper) and red (lower). The dotted margins encompass only the statistical variation of the interface position, whereas the dash-dotted include a ± 2 ~~m~~m systematic error which stems from the inherent DEM-uncertainty. For completeness we also show the position of the railway tunnel as a black square.

~~All together~~It might also be possible that the over-/underestimation might be due to heterogeneities of the rock density due to uneven fracturing and/or weathering. However, during our field-work (see Mair et al., 2018), when we also inspected the train tunnel from within, we did not see any signs of such a heterogeneous behaviour. These observations are still only superficial. For an in-depth study of this effect one would need, for example, a much longer muon flux exposure, such that the density of the rock could be better resolved. Alternatively a borehole, or another geophysical study could be performed. As we are not in possession of such information, we will not draw a definite conclusion here. Nevertheless, the performance of the whole workflow, which is shown in this study, produces results, which are similar to the ones published in the previous study (Nishiyama et al., 2017). We use the results of this comparison to validate the base of our code.

~~6~~5 Conclusion

In this study we have presented ~~a model~~an inversion scheme that allows us to integrate geological information into a muon tomography framework. The inherent problem of parameter estimation has been formulated in a probabilistic way and solved accordingly. The propagation of uncertainties thus occurs automatically within this formalism~~.~~, providing uncertainty estimates on all parameters of interest. We also considered approaches including DAGs or the simplex subspace of compositions which could be helpful to the muon tomography community while tackling their own research. We condensed these approaches in a modular toolbox. This assortment of python programmes allows the user to address the subproblems during the data analysis of a muon tomography experiment. The programmes are modular in the sense that the user can always access the intermediate results, as the files are mostly in a portable format (JSON). Thus, it is perfectly possible to only use one submodule of the toolbox while working with an own codebase. As every "tool" is embedded in a GUI, the programme is made accessible without the need to first read and consider several thousand code lines. Furthermore, we have shown that the results we obtain with our code are largely in good agreement with an earlier, already published experiment. The small deviations may be attributed to data analysis subtleties.

In its current state, SMAUG may be of help to researchers a) who plan to use muon tomography in their own research, such that the feasibility of the use of this technology can be evaluated in a virtual experiment, b) who want to use a submodule for the analysis of their own muon tomography, or c) plan to perform a subsurface interface reconstruction similar to our study. We would like to stress that this work is merely a foundation upon which many extensions can be built when it is used in other applications as well. Future content might, for example, include a realistic treatment of multiple scattering and the inclusion of compositional uncertainties in the inversion, for which we laid out the basis in this study.

930

Appendix A – Muon flux model

As many empirical muon flux models, the one that we employed consists of an energy spectrum for vertically incident muons at sea level at its core. An accepted instance is the energy spectrum of Bugaev et al. (1998), that takes the form

$$\Phi_B \frac{d\Phi_B}{dE}(p) = A_B p^{-(\alpha_3 \log_{10}^3(p) + \alpha_2 \log_{10}^2(p) + \alpha_1 \log_{10}(p) + \alpha_0)}, \quad \text{(A1)}$$

where p denotes the momentum of the incident muon in ~~$GeV * c^{-1}$~~ $\mathrm{GeV} * \mathrm{c}^{-1}$. The values of the α_i and A_B are, for example, listed in Lesparre et al. (2010). This model is an extended version of Renya (2006), to account for different incident angles,

$$\Phi_R \frac{d\Phi_R}{dE}(p, \theta) = \cos^3(\theta)\, \Phi_B(p\cos(\theta)) \frac{d\Phi_B}{dE}(p * \cos(\theta)), \quad \text{(A2)}$$

where θ is the zenith angle of the incident muon. It is important to note that the parameter values in Eq. (A1) are changed to $\alpha_0 = 0.2455$, $\alpha_1 = 1.288$, $\alpha_2 = -0.2555$, $\alpha_3 = 0.0209$ and $A_B = 0.00253$. In order to include height above sea level as an additional parameter, Hebbeker and Timmermans (2002) proposed to model the altitude dependency as an exponential decay, which modifies Eq. (A2) into

$$\Phi(p, \theta, h) \frac{d\Phi}{dE}(p, \theta, h) = \Phi_R \frac{d\Phi_R}{dE}(p, \theta) * \exp\left(-\frac{h}{h_0}\right). \quad \text{(A3)}$$

The scaling height, h_0, is usually to be taken as $h_0 = 4900m4900 + 750$ ~~$m\, c\, GeV^{-1} **$~~ p, where p, is the momentum of the incident muon in ~~$GeV * c^{-1}$~~ $\mathrm{GeV} * \mathrm{c}^{-1}$. However, as this formula is only valid up to an altitude of 1000 m above sea level, Nishiyama et al. (2017) adapted it to $h_0 = 3400$ ~~m~~ $+ 1100$ ~~$m * c * GeV^{-1} **$~~ $p * \cos(\theta)$. This was done in order to fit the energy spectrum up to 4000 m above sea level. This formula is now valid for momenta above 3 ~~$GeV * c^{-1}$~~ $\mathrm{GeV} * \mathrm{c}^{-1}$, zenith angles between 0° and 70° and an altitude below 4000 ~~m~~ m above sea level. Please note that the muon momentum, p, is related to its energy by the relativistic formula

$$p^2 c^2 = E - m_\mu^2 c^4. \quad \text{(A4)}$$

Thus, it holds that

$$\frac{d\Phi}{dE}(p, \theta, h) \equiv \frac{d\Phi}{dE}(p(E), \theta, h) \equiv \frac{d\Phi}{dE}(E, \theta, h). \quad \text{(A5)}$$

Appendix B – Rock model

B1 – Density model

The density distribution of a lithology can be determined through various methods. In our work, we ~~constructed a~~estimated the density ~~model~~of the lithology by analysing various rock samples from our study area in the laboratory. Two experimental setups were employed to gain insight into the grain, skeletal as well as the bulk density of the rocks. Grain and skeletal density were measured by means of the AccuPyc 1340 He-pycnometer, which is a standardised method that yields information on the volume. Bulk density values were then determined based on Archimedes' principle, where paraffin coated samples were suspended into water (ASTM C914-09, 2015; Blake and Hartge, 1986).

Every sample $j = 1, \dots, N$ (usually the size of a normal hand sample) has been split up into smaller subsamples $i = 1, \dots, S_j$, that were measured. The bulk density of the i-th subsample can be calculated by

$$\rho_{bulk,ij} = \frac{\rho_{H_2O} * m_{s,ij}}{(m_{s,ij} + m_{p,ij} + m_{t,ij} - m_{sus,ij}) - \left(\frac{m_{p,ij} * \rho_{H_2O}}{\rho_p}\right) - \left(\frac{m_{t,ij} * \rho_{H_2O}}{\rho_T}\right)}, \quad \text{(B1)}$$

where $\rho_{H_2O}, \rho_p, \rho_T$ denote the density of water, paraffin and the thread that was used to dip the sample into the liquid, respectively. $m_{s,ij}, m_{p,ij}, m_{t,ij}, m_{sus,ij}$ describe the mass of the sample, the paraffin coating, the thread and the apparent mass of all three components suspended in water. $m_{p,ij}, m_{t,ij}$ can then be simply obtained through

$$m_{p,ij} = m_{s,t,p,ij} - m_{s,t,ij}\,, \quad \text{(B2)}$$

as $m_{s,t,p,ij}$ denote the mass of the sample including thread and paraffin coating on one hand and $m_{s,t,ij}$ only the mass of the sample and the thread on the other hand. Further, the mass of the thread is given by

$$m_{t,ij} = m_{s,t,ij} - m_{s,ij}. \quad \text{(B3)}$$

The maximal precision of the reading is estimated at $\pm 5 * 10^{-5}$ ~~gg~~, and the commonly ignored effects regarding buoyancy in air has been estimated to introduce an error on the order of $\pm 2 * 10^{-4}$ ~~gg~~. This error has been attributed to all direct mass measurements. Moreover, because small pieces of material may detach from the sample upon attaching the thread to the sample and during the paraffin coating, we set an error of $\pm 2 * 10^{-2}$ ~~gg~~ to all measurement results. The variables in Eq. (B1) are strictly positive values. Following Tarantola (2005) we model these "Jeffreys parameters" by lognormal distributions, as they inherently satisfy the positivity constraint. Because Eq. (B1) does not simply allow a standard uncertainty propagation, the script "subsample_analysis.py" performs a Monte Carlo simulation for each subsample and attributes a final lognormal probability density function to the resulting histogram. Figure B1 illustrates such an example, where the calculation has been performed for subsample JT-20-1.

We have found 10'000 draws per subsample to be sufficient to retrieve a solid final distribution. However, this parameter can easily be changed in the script, depending on the user's preference of precision/speed. From this point onwards we may

work with a Gaussian distribution as Fig. B1 assures us that a normal pdf describes the results of the Monte Carlo simulation rather well.

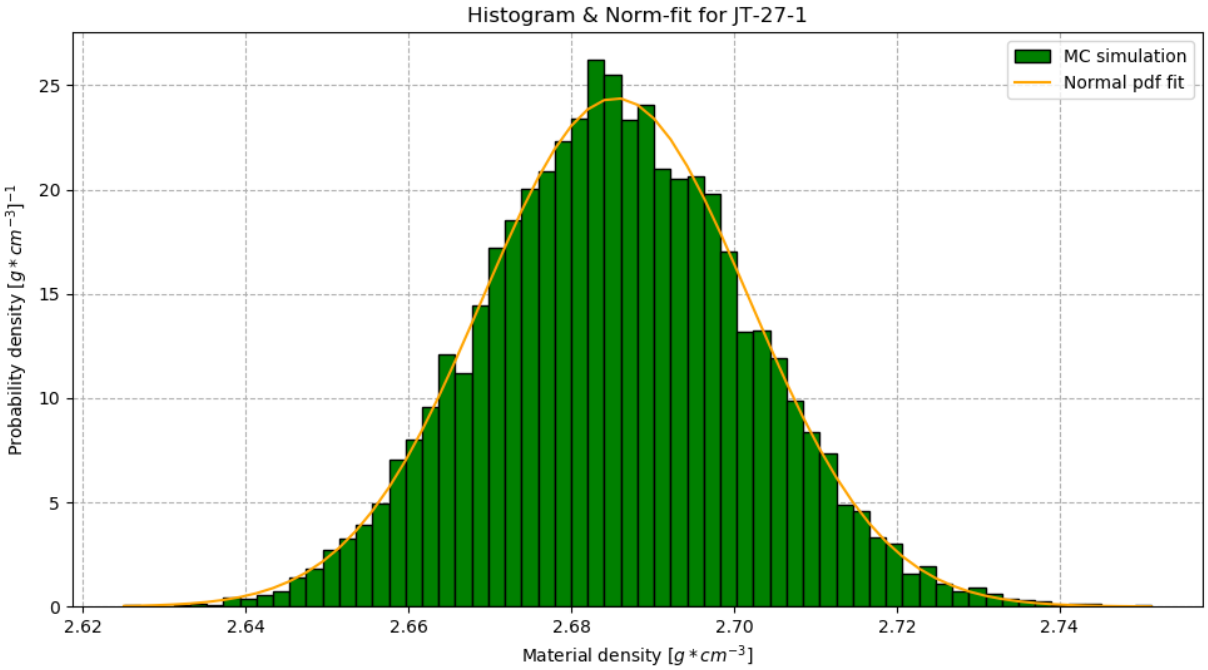


Figure B1: Example output of "subsample_analysis.py" for a bulk density measurement of subsample JT-27-1 (see supplementary information for data). Green bars represent the histogram of 10'000 Monte Carlo simulation draws. The orange curve indicates the fitted normal probability density function.

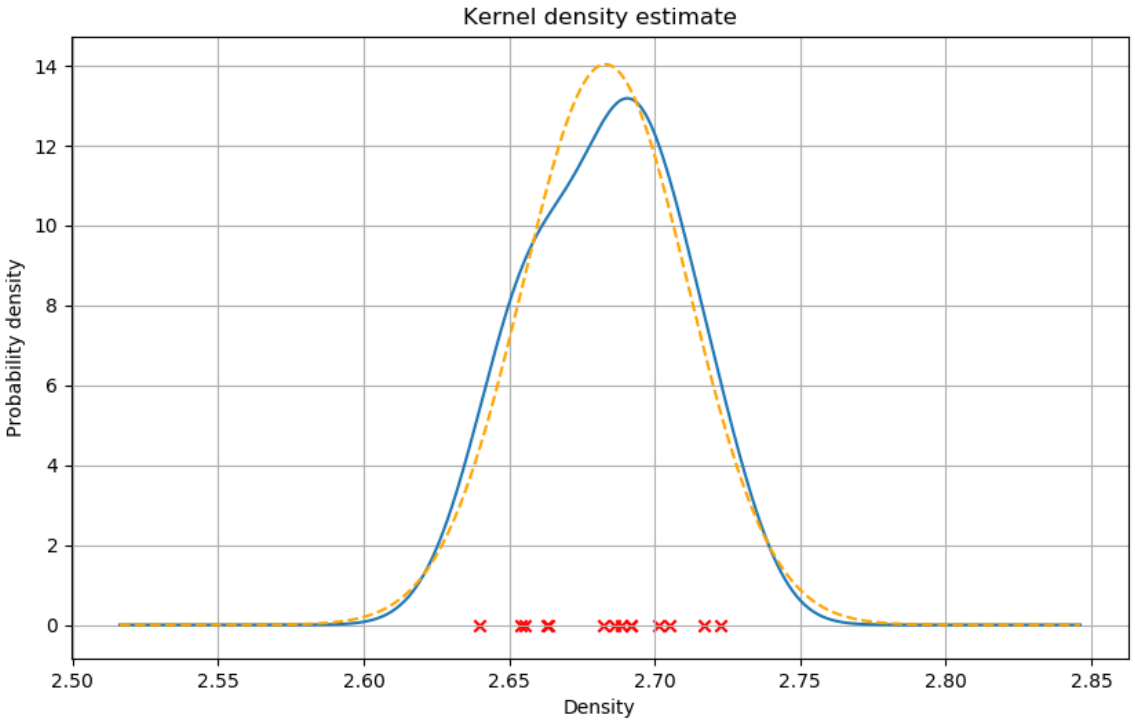
The determination of the grain and skeletal densities is simpler than the bulk density measurements because the corresponding method consists of a mass and a volume measurement, respectively. The density formula reads then simply

$$\rho_{skeletal/grain,ij} = \frac{m_{ij}}{v_{ij}}. \quad \text{(B4)}$$

The question remains as to how it is possible to construct a pdf that represents the knowledge about the whole lithology. There are two possible methods that can be readily employed at this point. The first, following largely Tarantola (2005), performs a so-called disjunction of the pdfs that corresponds to an averaging of all subsample pdfs. As Vermeesch (2012) points out, even though this might seem a "sensible strategy at first glance", there might be some problems with this method. The main problem lies in the small error on the subsamples, such that the variation between different subsamples may be larger than their attributed errors. This would not be a problem if enough subsamples could be measured, such that the resulting lithology pdf might be sampled correctly. On the other hand when one is faced with a situation where data is rather scarce, then the approach of Tarantola (2005) would result in a rather spikey pdf that would be hard to handle. For this

reason we adopted the methodology of Vermeesch (2012) where the lithology pdf is estimated by a kernel density estimation. The main difference lies in the fixed "bandwidth" of the subsample distributions. We refer to Vermeesch (2012) for more details and an in-depth discussion on this problem.

The kernel density estimation has the advantage, that only the mean values of the subsamples have to be processed as the bandwidth is determined from the spread of the subsample means. Following the methodology of Vermeesch (2012) we end up with a pdf like the one visualised in Fig. B2. We could at this point use the kernel density estimated pdf for further calculations. However, for simplicity we approximate the kde with a normal distribution and intend to add support for the kde in a later code version.



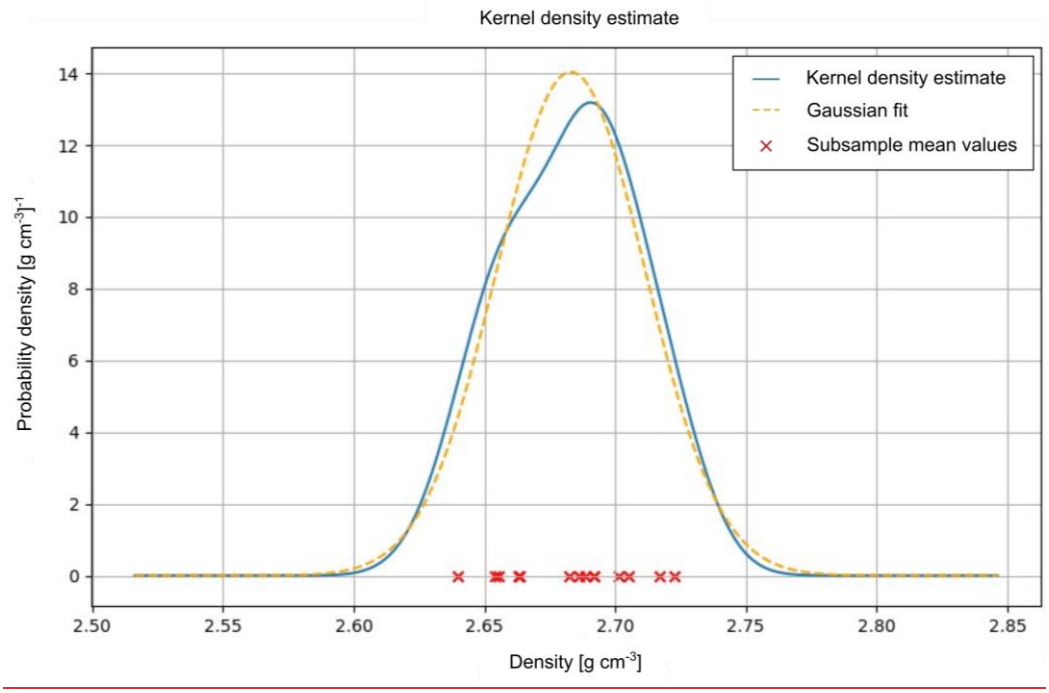


Figure B2: Example output of "materializer.py". Here a set of subsample mean values (red crosses) are processed in a kernel density estimate (solid blue line). Finally a normal distribution is fitted to the kernel density estimate (dashed yellow line).

One word of warning has to be made here. The measured densities of rock might be affected with a systematic error. Namely, the rock samples that are analysed were all gathered from surface near locations (in our case inside the tunnel or outside, i.e. where rocks are accessible). This means that they could have been subject to weathering processes that alter the density of the rock in such a way that the samples are not representative of the whole rock body anymore. Possible countermeasures would be to compare drilled samples from deeper within the rock body with the surface samples, etc.

B2 – Composition model

We have seen in Eq. (~~9~~12) that the material density parameter enters the energy loss calculations rather directly. Contrariwise, the compositional model affects the energy loss equations much more subtly through the average $\{Z/A\}_{rock}$ and $\{Z^2/A\}_{rock}$ values and mean excitation energies that need to be calculated for the entire lithology. Likewise, information on the weight percentages of the main elements within the rock is required for the quantification of the radiation loss term. Although a modal mineral analysis (e.g. the quantitative determination of mineral volumes) is preferable and can be treated according to Lechmann et al. (2018), its execution is a rather time-consuming effort. This is the reason why compositional data in muon tomography experiments predominantly consist of XRF-data, which show the abundance of major oxides within the rock. We describe here a method to incorporate such type of information in a probabilistic way thereby following Aitchison (1986). Compositional data are usually available in the form of Table B1, which presents an excerpt of four samples for illustration purposes. We refer to the excel sheet in the supplementary material of the present work for the full data.

Table B2: Excerpt of XRF data for four samples. Data in column denote weight percentages of major oxides within the rock samples.

Sample	JT01	JT02	JT19	JT20
Oxides				
SiO2	0.6131	0.5981	0.6997	0.6139
TiO2	0.0123	0.0067	0.0076	0.0094
Al2O3	0.1567	0.1873	0.1481	0.1921
Fe2O3	0.087	0.0791	0.0496	0.0686
MnO	0.001	0.0012	0.0009	0.0009
MgO	0.0359	0.0285	0.0206	0.0288
CaO	0.0202	0.0071	0.0201	0.0137
Na2O	0.0228	0.0248	0.0404	0.0323
K2O	0.0343	0.0465	0.0287	0.0469
P2O5	0.0041	0.0029	0.0021	0.0027
Sum	0.9874	0.9822	1.0178	1.0093

There are several challenges to this kind of data. First, the parameters (i.e. the oxide percentages) can take a value between 0 and 1. This means that normal as well as lognormal distributions are not suitable to describe these parameters. Second, the requirement that the sum of all parameters has to ideally equal 1 poses a constraint on this parameter space, which effectively reduces the number of independent parameters by one. Third, due to measurement uncertainties, this sum is never exactly one.

Spaces, which have this unit sum condition can be viewed as a simplex, e.g. if we had three compositional parameters, the simplex would be a 2-dimensional surface (i.e. a subspace) in this 3-dimensional parameter space. The last issue, of not summing up exactly to 1, can be remedied by projecting each sample dataset back to the simplex (Aitchison, 1986, p. 257-261). This works only if the measurement imprecisions are not too large, which works well for the examples in Table B1. With respect to the energy loss calculation, it is preferable to decompose the oxides into elements, which can be done by the following formula

$$wt_{ele,i} = \sum_{j \in \{oxides\}} wt_j * \frac{n_{ij} m_i}{m_j}, \quad \text{(B5)}$$

where m_i and m_j denote the molar mass mass of the i-th element and the j-th oxide, wt_j is the j-th datum in the column and n_{ij} is the number of atoms of the i-th element within the j-th oxide. The two transformations are visualised in Table B2.

Table B3: Element weight percent data. Transformed from oxide weight percent data with use of Eq. (B5). All data has additionally been scaled to satisfy the unit sum constraint.

Sample	JT01	JT02	JT19	JT20
Elements				
Si	0.2902	0.2846	0.3213	0.2843
Ti	0.0075	0.0041	0.0045	0.0056
Al	0.0840	0.1009	0.0770	0.1007
Fe	0.0616	0.0563	0.0341	0.0475
Mn	0.0008	0.0009	0.0007	0.0007
Mg	0.0219	0.0175	0.0122	0.0172
Ca	0.0146	0.0052	0.0141	0.0097
Na	0.0171	0.0187	0.0294	0.0237
K	0.0288	0.0393	0.0234	0.0386
P	0.0018	0.0013	0.0009	0.0012
O	0.4716	0.4711	0.4823	0.4707
Sum	1	1	1	1

In order for the data to be in a statistically convenient form, Aitchison (1986) suggests to further transform the data in Table B2 by first forming a ratio with an arbitrary element (in the list) and then taking the logarithm. For the exemplary dataset this is shown in Table B3.

Table B4: Log-ratio of element weight percentages, with respect to oxygen-wt%.

Sample	JT01	JT02	JT19	JT20
Elements				
ln(Si/O)	-0.48531565	-0.50379579	-0.40607778	-0.5042219
ln(Ti/O)	-4.14567399	-4.74687577	-4.68001075	-4.43477381
ln(Al/O)	-1.72531	-1.54064159	-1.83464118	-1.54183733
ln(Fe/O)	-2.03494223	-2.12384752	-2.64974526	-2.29276806
ln(Mn/O)	-6.39894857	-6.21033707	-6.55719484	-6.52451934
ln(Mg/O)	-3.06839321	-3.29293646	-3.67672517	-3.30896536
ln(Ca/O)	-3.47357746	-4.51287533	-3.53142599	-3.88207448
ln(Na/O)	-3.31519303	-3.22481996	-2.79600952	-2.98709658
ln(K/O)	-2.79436382	-2.48376691	-3.0254978	-2.50170175
ln(P/O)	-5.56150947	-5.90149576	-6.28344485	-5.99945492

ln(O/O)	0	0	0	0

The rationale behind this transformation is as follows. The division by an arbitrarily present element effectively transforms the space into an N-1-dimensional open space, where the parameters (i.e. ratios) may have values between 0 and ∞. The subsequent application of the logarithm further changes the space, such that the new parameters can have values between $-\infty$ and ∞. This results in so-called log-ratios, which should ideally be following a multivariate normal distribution. As a consequence, we can ~~calculate~~describe this distribution by the mean log-ratio vector across all samples as well as its corresponding covariance matrix~~, which completely describes the multivariate normal distribution~~. In addition to these statistical parameters, the script "compo_analysis.py" outputs a graph that plots for all samples an order statistic, z_r, (see Aitchison, 1986). This enables us to visualise how different the data is from a multivariate normal distribution. If equal, they should fall on the red line, shown in Fig. B3.

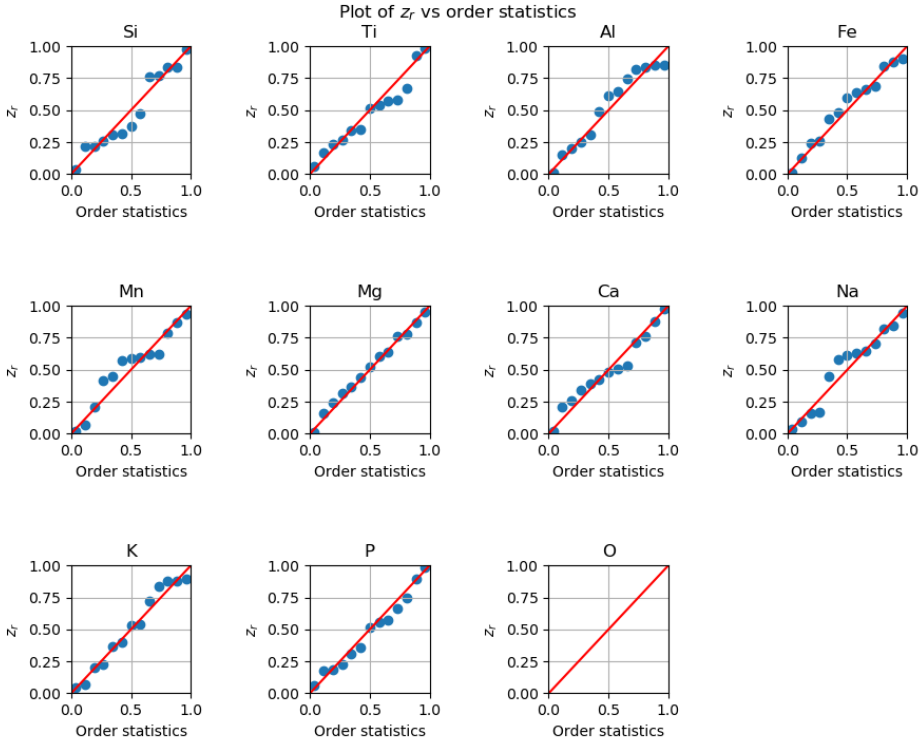


Figure B3: Visual test for multivariate normality of the log-ratio data from Table B3 (This plot shows the full dataset, of which Table B3 is only an excerpt). Each subplot checks for marginal normality. Oxygen is the denominator variable (arbitrarily chosen) and does thus not appear in the plot.

With a graph like Fig. B3 it is possible to check if the multivariate normal distribution is an appropriate model to describe the elemental composition data. For ~~our~~the example ~~that we show~~shown in Fig. B3 this looks acceptable, with only slight deviations for silicon, aluminium, manganese and sodium~~).~~. Once the normality has been verified it is possible to generate random samples from this distribution. For every drawn sample it is then possible to calculate the weight percentages of the single elements by using the inverse formula to the log-ratio transformations

$$wt_{ele,i} = \frac{\exp(r_i)}{1-\sum_{j=1}^{N_{ele}-1}\exp(r_j)}, \tag{B6}$$

for all numerator elements and

$$wt_{ele,N_{ele}} = \frac{1}{1-\sum_{j=1}^{N_{ele}-1}\exp(r_j)} \tag{B7}$$

for the denominator element (here oxygen). In Eqs. (B6) and (B7) the r_i denote the log-ratios from Table B3~~ and,~~ N_{ele} is the total number of elements (in Table B2~~).~~) and the index i runs through all elements (e.g. in Table B5) and the index j runs through all elements except the denominator variable.

B3 – Energy loss equation for rocks

As stated in Eq. (~~7~~11) the energy loss equation for rocks needs parameters that differ from the ones for pure elements. First, the expression for density can directly be exchanged according to the density model (see Appendix B1). Second, it is possible to generate an expression for the average ionisation loss within a rock by exchanging three parameters. Density values that also enter within $\{a\}_{rock}$ can again be directly changed. The average $\{Z/A\}_{rock}$ may be exchanged with the elemental Z/A by using

$$\{Z/A\}_{rock} = \sum_{i=1}^{N_{ele}} wt_{ele,i} * \frac{Z_i}{A_i}. \tag{B8}$$

$wt_{ele,i}$ are the weight fractions from Eqs. (B6) & (B7). Lastly, the mean excitation energy, $\{I\}_{rock}$, for the rock can be computed by

$$\ln\{I\}_{rock} = \frac{\sum_{i=1}^{N_{ele}} wt_{ele,i}*\frac{Z_i}{A_i}*\ln I_i}{\{Z/A\}_{rock}}. \tag{B9}$$

The radiation loss term, ~~however,~~$\{b\}_{rock}$, must be calculated as a weighted radiation energy loss over all i elements. This means that the average can be written in a rather concise form,

$$\{\mathrm{b}\}_{\mathrm{rock}} = \sum_{i=1}^{N_{ele}} wt_{ele,i} * b_{ele,i}\,. \tag{B10}$$

Appendix C – Metropolis Hastings technicalities

This appendix ~~chapter~~ is a short summary of Gelman (2014, p. 284 – 287) and we refer to these pages for a detailed discussion of the calculations. This work presents a concept of how to assess the quality of a MCMC run. In particular, the aforementioned author proposes to analyse two quantities, the potential scale reduction factor $\hat{R}$ and the effective number of simulation draws $\hat{n}_{eff}$ for every parameter of interest. For every chain of a parameter the variance between different chains and within one chain is calculated. The posterior variance of the parameter is then estimated as a weighted average of these two types of variances. Finally, $\hat{R}$ is the quadratic ratio between the posterior variance and the variance within one chain. This quantity shows if the various chains have mixed or not, i.e. if they have explored the same region of the model space. If the posterior variance is much larger than the variances of the single chains, then the chains have not sufficiently explored the same region. Gelman (2014) propose to employ a threshold of 1.1 as a rule of thumb, below which the value of $\hat{R}$ would lie.

One problem that arises in MCMC algorithms is the inherent dependence of one simulation on the one before (this is the definition of a Markov chain). One considers that such a dependency does not introduce a bias if enough samples are drawn. However, this also means, that the effective, independent sample size is much smaller than the number of simulations. Therefore, Gelman (2014) proposes to calculate the effective number of simulation draws, $\hat{n}_{eff}$ in order to assess if one has enough independent samples The underlying idea here is to evaluate the correlations within the chains. An accepted threshold value for this parameter is $5m$, where m is the number of sub-chains. For the calculation of $\hat{R}$ and $\hat{n}_{eff}$ the chains may be cut in half to generate more chains. Note, however, that $\hat{n}_{eff}$ can also be larger, which only means that the simulation standard error decreases. In our example we performed the calculations with two chains and a subdivision by 2, which means that our target quantity is around $20 (= 5 * 4)$. Most of our thickness parameters (i.e. cones with bedrock and ice; see Figs. 1 & ~~10~~6) have, in fact, a $\hat{n}_{eff} > 100$, with only a few below.

Appendix D – Construction of the smoothing kernel

As stated in the main text, the user specifies the number of neighbouring pixels s to smooth over. The main idea is to construct a roughly Gaussian smoothing kernel by approximating it with a binomial distribution. With help of the binomial coefficient we can construct a vector of weights with $L = (2 * s + 1)$ entries. The weight vector is then given by

$$w_i = \frac{1}{2^{2*s}}\binom{L-1}{i}, \tag{D1}$$

with $i \in \{0, .., L-1\}$. It is now possible to create a matrix by forming the dyadic product of $\vec{w}$ with itself, i.e.

$$K = \vec{w} \otimes \vec{w}, \tag{D2}$$

or in index notation,

$$K_{ij} = w_i * w_j . \tag{D3}$$

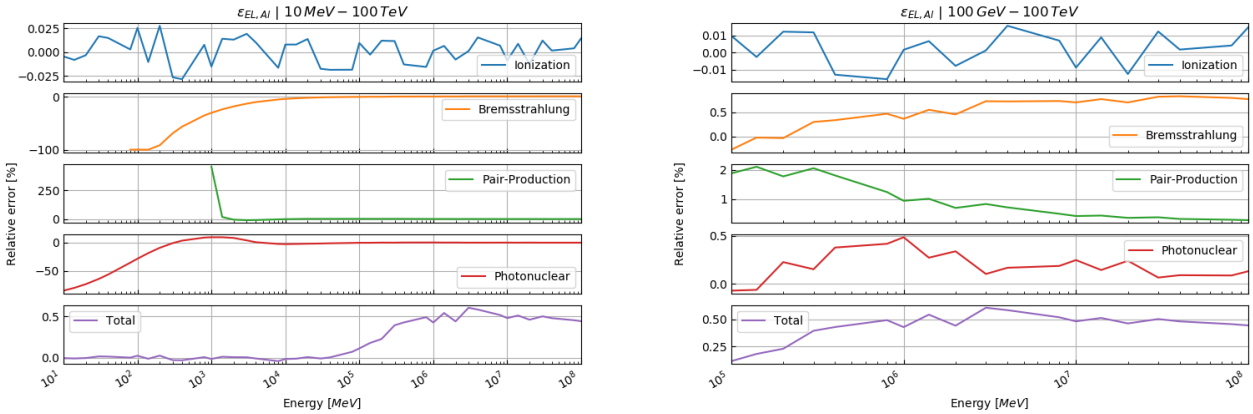
As an example, we show how a smoothing kernel that smooths over two neighbouring pixels (i.e. $s = 2$) is constructed. This is incidentally also the smoothing kernel we used to construct our ice-bedrock interface. The weight vector in this case is given by

$$\vec{w} = \frac{1}{16} * (1 \quad 4 \quad 6 \quad 4 \quad 1) . \tag{D4}$$

The weight vectors are, in fact, only the odd rows from Pascal's triangle, interpreted as vectors and normalised by a L1 norm. The smoothing matrix then takes the form

$$K = \frac{1}{256}\begin{pmatrix} 1 & 4 & 6 & 4 & 1 \\ 4 & 16 & 24 & 16 & 4 \\ 6 & 24 & 36 & 24 & 6 \\ 4 & 16 & 24 & 16 & 4 \\ 1 & 4 & 6 & 4 & 1 \end{pmatrix}. \tag{D5}$$

Appendix E – Energy loss calculations for various elements and compounds



~~Figure~~ E1~~: Relative error~~ – Verification of ~~our~~ energy loss calculations ~~compared to~~

The energy loss model that we use in our code generally reproduces the ~~tabulated~~literature values well (the relative error is generally below 1%) across the different energy loss processes and relevant energies. In Fig. E1 we present the energy loss calculations for each energy loss process (i.e. ionisation, bremsstrahlung, pair-production and photonuclear interactions) across energies from 10 MeV to 100 TeV for silicon.

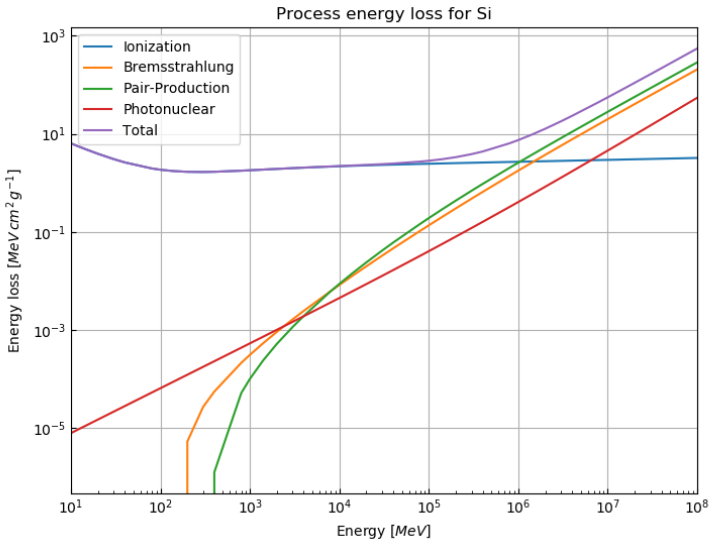
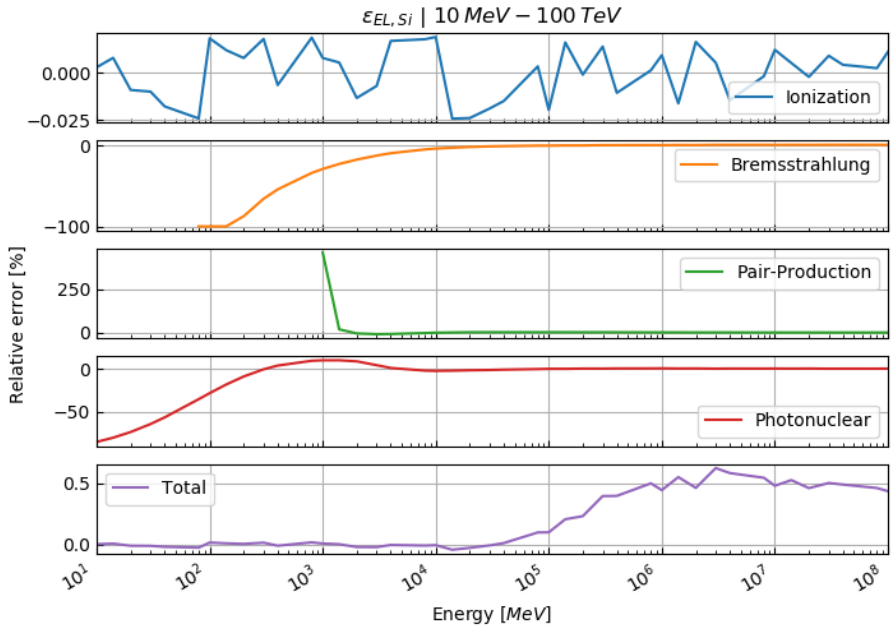


Figure E1: Log-log plot of the stopping power of the different energy loss processes for silicon. At ~ 10 GeV the radiative processes (i.e. bremsstrahlung, pair-production and photonuclear interactions) reach around 1 % of the total stopping power. At a few hundred GeV (at the so called "critical energy") the radiative processes start to become dominant over the ionisation losses.

The overall characteristics between the different elements are the same with minor differences regarding the position of the critical energy and the 1 % - radiative point. In Fig. E2 we show the relative error of our calculations to the tabulated values from Groom et al. (2001) for the whole energy range.



~~from Groom et al. (2001) for aluminium in the energy ranges: (left) 10 MeV – 100 TeV, (right) 100 GeV – 100 TeV.~~

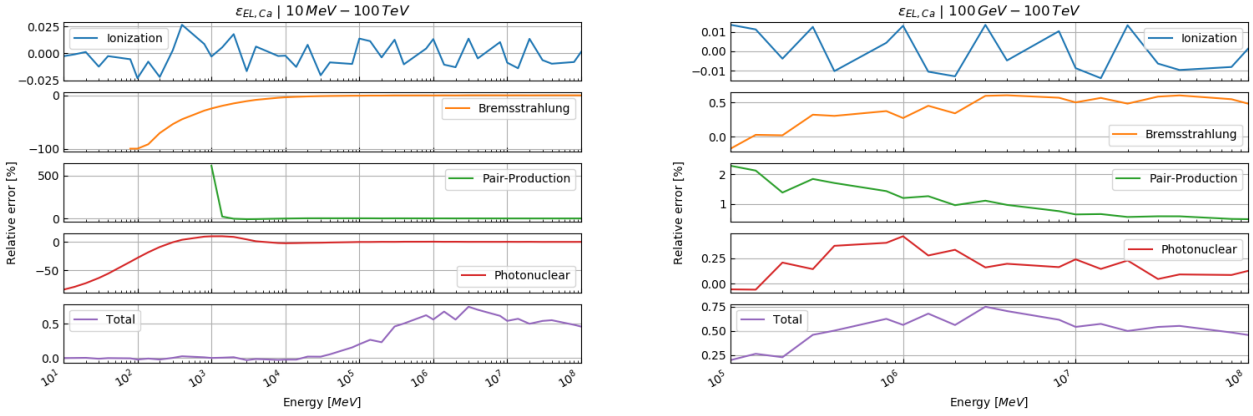


Figure E2: Relative error of our energy loss calculations compared to the tabulated values from Groom et al. (2001) for silicon. Ionisation losses agree very well with the literature values (within 0.025 %). At low energies the relative errors of the radiative processes are large and converge to a value close to 0 towards higher energies, resulting in a relative error on the total energy loss of around 0.5 % compared to literature.

We note that the energy losses by ionisation are reproduced very well over the entire energy range. We also note that the relative error on the radiative energy losses is rather large below 10 GeV. This does, however, not introduce a major bias, because below this energy, radiative energy losses are negligible compared to ionisation losses, as can be seen in Fig. E1. Furthermore, the related errors are in an acceptable range at the energy level at which radiative losses begin to become noticeable (i.e. around 100 GeV). This can be seen in Fig. E2, in the sense that the total relative error remains well bounded within 0.5 %. In the ionisation domain (i.e. below 100 GeV) the total relative error is dominated by the ionisation relative error, whereas above this energy level the relative errors on radiative losses start to prevail. A close-up of this energy range is given in Fig. E3.

There are different sources and circumstances that contribute to the error in the different energy losses processes. The scatter of the relative ionisation-loss error around 0 with a rather small deviation can be viewed as simple rounding errors. The errors on the radiative processes, however, seem to be of a more systematic nature. We explain this behaviour through a different numerical integration scheme in Eq. (10), which tends to systematically under-/overestimate the true value, especially when the integrand comprises exponential functions. Whereas we used a Double Exponential Integration scheme (see Takahasi & Mori, 1974), the integration scheme from Groom et al. (2001) is not discernible. However, as the relative errors on the processes of energy loss remain well within the theoretical uncertainties, (see Sect. 3.1.3), we consider, that our calculation accurately reproduces the literature values for elements.

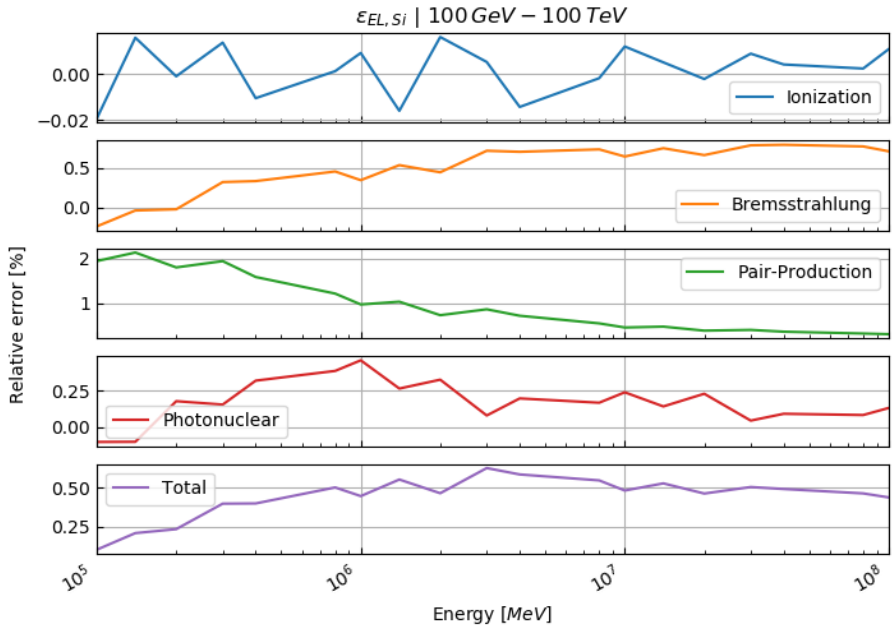


Figure E3: Relative error of our energy loss calculations for silicon compared to the tabulated values from Groom et al. (2001) at higher energies (100 GeV – 100 TeV). The relative errors remain bounded within their theoretical uncertainties (see Sect. 3.1.3).

The above calculations were performed for pure silicon. The respective figures for other four important elements in the Earth's crust (Al, Fe, Ca & O) can be found in Appendix E2. Those elements are, however, not representative for any real material encountered in geological applications. For this reason, we compiled the same computations for four selected, geologically important compounds (SiO2, CaCO3, ice) that are also listed in Appendix E2. We summarise, that with the exception of Standard Rock, all calculations yield results that are similar to the silicon calculation above. The discrepancy for Standard Rock stems from its inconsistent definition, with respect to the different parameters. In particular, the "Standard Rock" according to Lohmann et al. (1985) has an atomic number Z of 11 (i.e. sodium) and an atomic weight A of 22, which yield the characteristic parameter values of $\langle Z/A \rangle = 0.5$ and $\langle Z^2/A \rangle = 5.5$ respectively. Note that Groom et al. (2001) list sodium as the only constituent of a standard rock. However, this material cannot be modelled by any mixture of pure elements, as common sodium consists of one neutron more and thus has a higher atomic weight (i.e. $A_{Na} = 23$). Consequently, the use of standard sodium would lead to different characteristic parameter values, i.e. $\langle Z/A \rangle = 0.478$ and $\langle Z^2/A \rangle = 5.263$, thus leading to an inconsistency. This is often conveyed by the phrase that standard rock "is not-quite-sodium" (Groom et al. 2001, p.203). In order to circumvent this problem, we advocate the exchange of $^{23}_{11}Na$ with its $^{22}_{11}Na$ isotope. This would lead to the characteristic parameter values $\langle Z/A \rangle = 0.500$ and $\langle Z^2/A \rangle = 5.501$, which are much closer to the actual definition of standard rock. For this reason, we extended the element/compound-list, (which is available from http://pdg.lbl.gov/2019/AtomicNuclearProperties/expert.html) by the $^{22}_{11}Na$ isotope, assuming that all parameters are equal to the ones from $^{23}_{11}Na$. Additionally, we redefined the standard rock (i.e. material number 281 in the list) to consist only of $^{22}_{11}Na$. With this change, standard rock does not need any more special treatment and can be calculated in a way that is consistent to all other compounds. Furthermore, the relative error between the tabulated values and our modified calculation falls in line with the calculations for the other compounds and elements (Figs. ~~(2001) for calcium in the energy ranges: (left) 10 MeV – 100 TeV, (right) 100 GeV – 100 TeV.~~

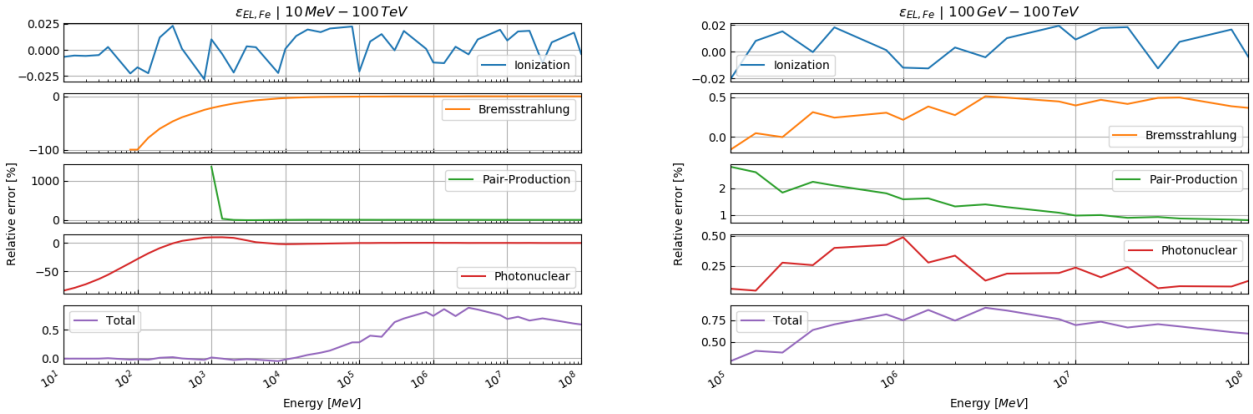
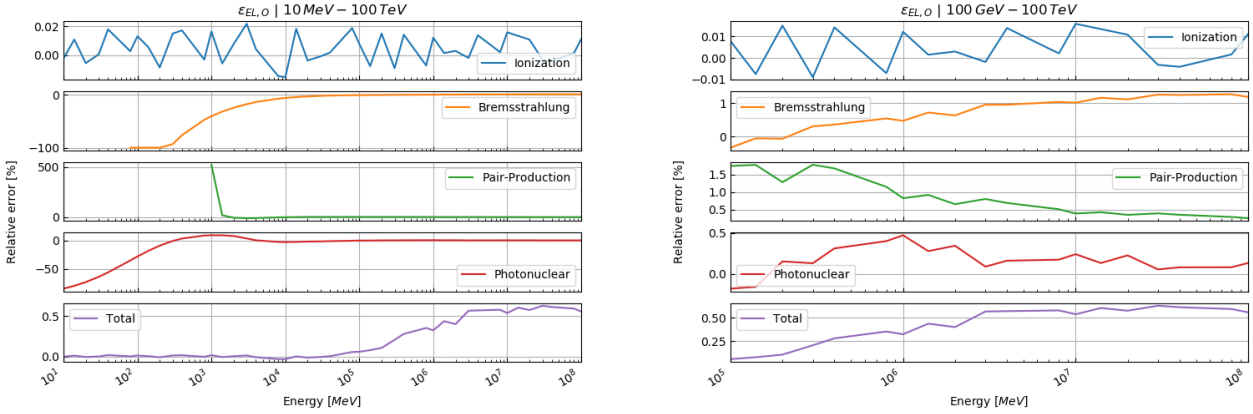


Figure E3: Relative error of our energy loss calculations compared to the tabulated values from Groom et al. (2001) for iron in the energy ranges: (left) 10 MeV – 100 TeV, (right) 100 GeV – 100 TeV.



E4 and E5).

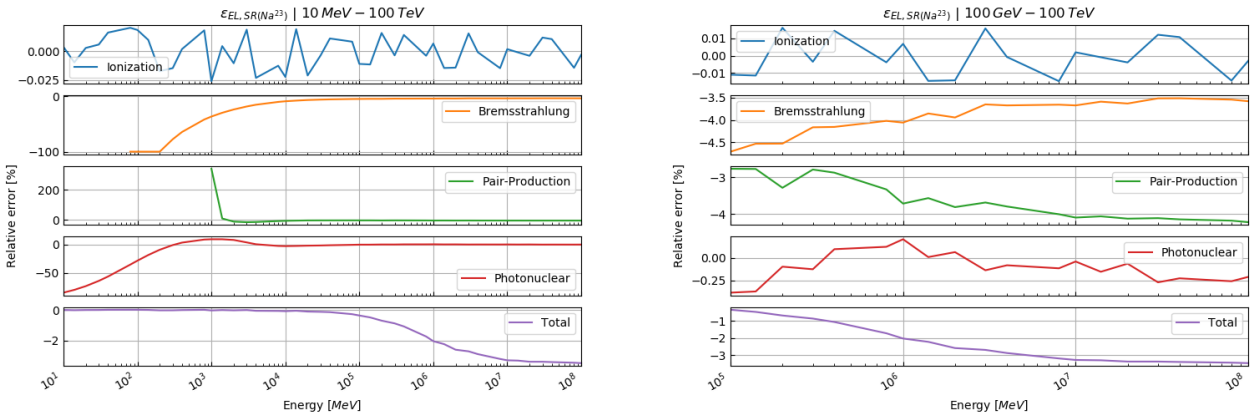
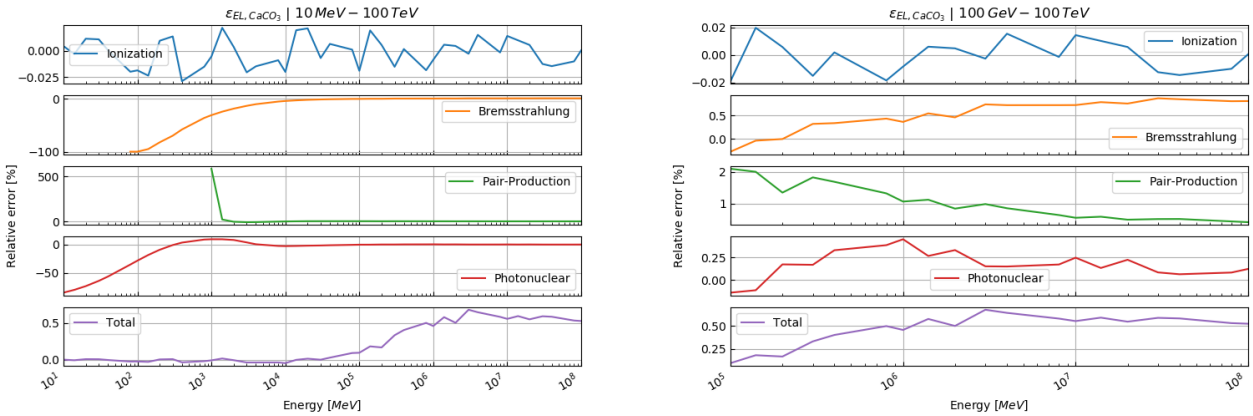
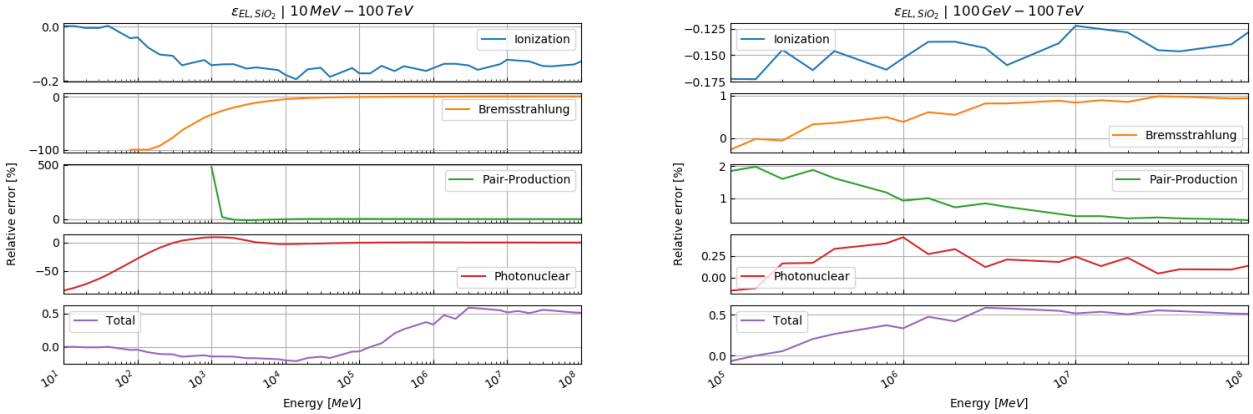


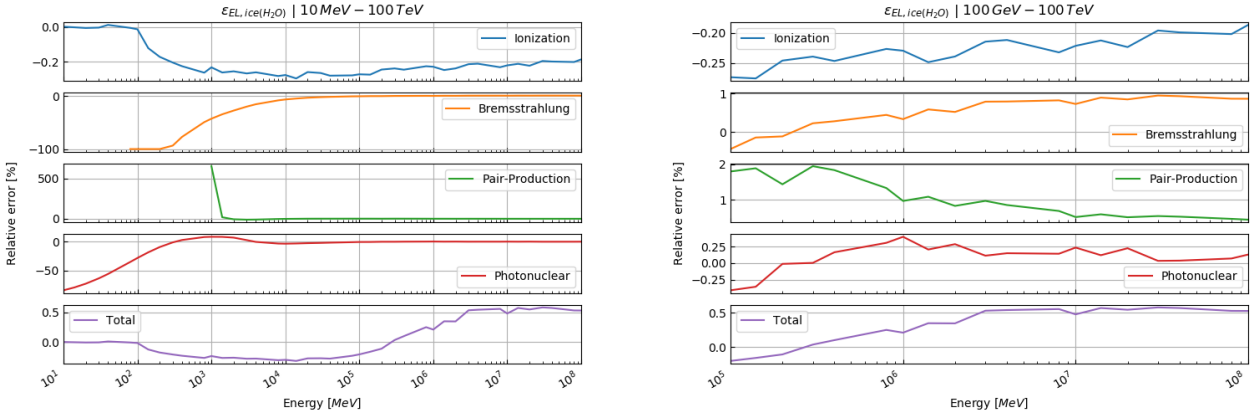
Figure E4: Relative error of our energy loss calculations compared to the tabulated values from Groom et al. ~~(2001) for oxygen in the energy ranges: (left) 10 MeV – 100 TeV, (right) 100 GeV – 100 TeV.~~



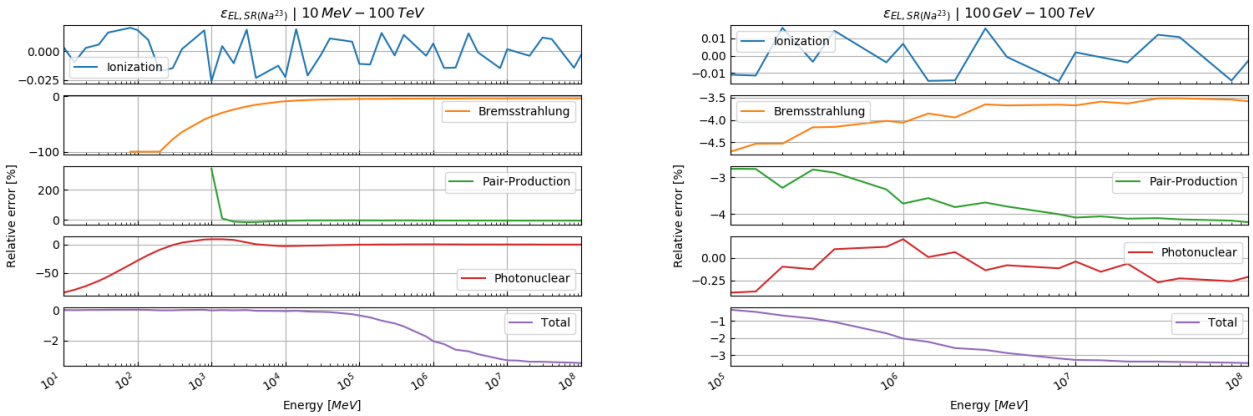
~~**Figure E6: Relative error of our energy loss calculations compared to the tabulated values from Groom et al. (2001) for calcium carbonate (calcite) in the energy ranges: (left) 10 MeV – 100 TeV, (right) 100 GeV – 100 TeV.**~~



~~**Figure E7: Relative error of our energy loss calculations compared to the tabulated values from Groom et al. (2001) for silicon dioxide (quartz) in the energy ranges: (left) 10 MeV – 100 TeV, (right) 100 GeV – 100 TeV.**~~



~~**Figure E8: Relative error of our energy loss calculations compared to the tabulated values from Groom et al. (2001) for ice in the energy ranges: (left) 10 MeV – 100 TeV, (right) 100 GeV – 100 TeV.**~~



~~Figure E9: Relative error of our energy loss calculations compared to the tabulated values from Groom et al.~~ **(2001) for standard rock ($^{23}_{11}$Na) in the energy ranges: (left) 10 MeV – 100 TeV, (right) 100 GeV – 100 TeV.**

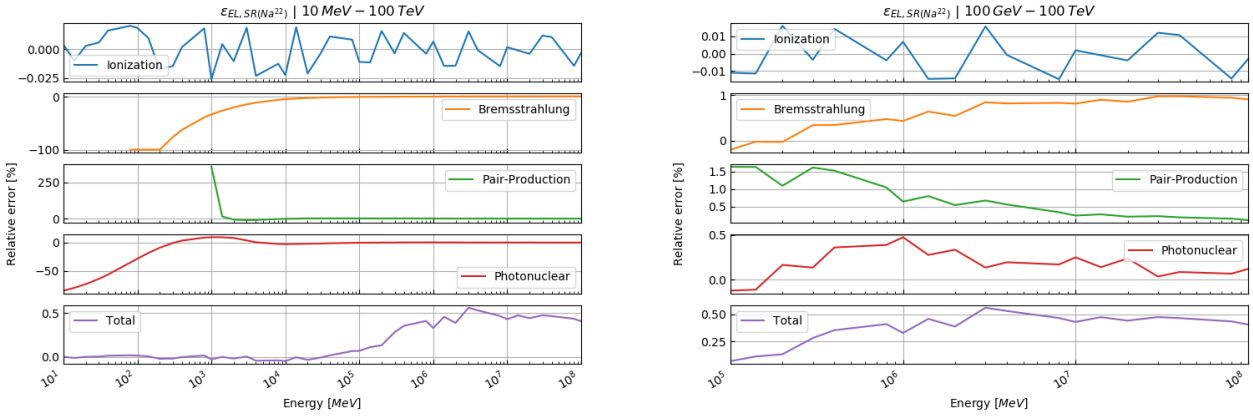


Figure E5: Relative error of our energy loss calculations compared to the tabulated values from Groom et al. ~~**Figure E10: Relative error of our energy loss calculations compared to the tabulated values from Groom et al.**~~ **(2001) for standard rock ($^{22}_{11}$Na) in the energy ranges: (left) 10 MeV – 100 TeV, (right) 100 GeV – 100 TeV.**

E2 – Additional plots of energy loss calculations

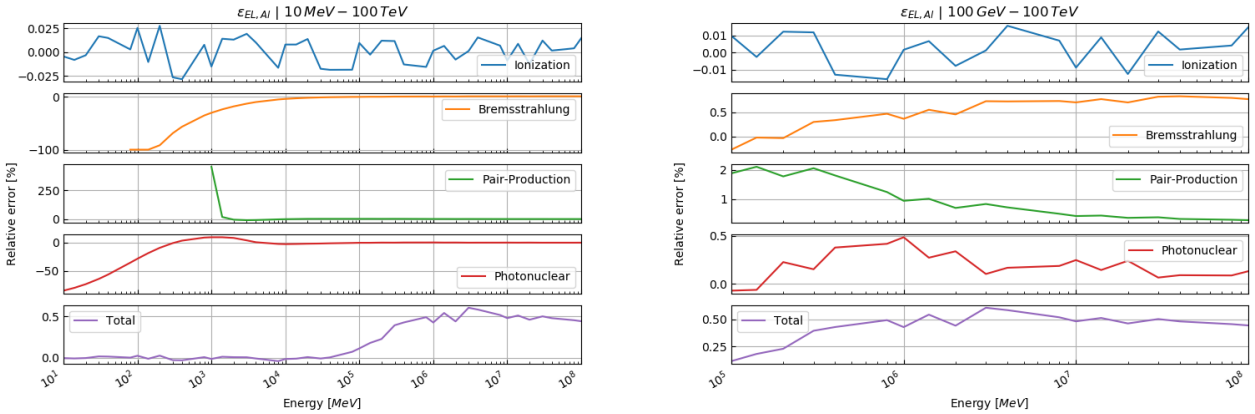


Figure E6: Relative error of our energy loss calculations compared to the tabulated values from Groom et al. (2001) for aluminium in the energy ranges: (left) 10 MeV – 100 TeV, (right) 100 GeV – 100 TeV.

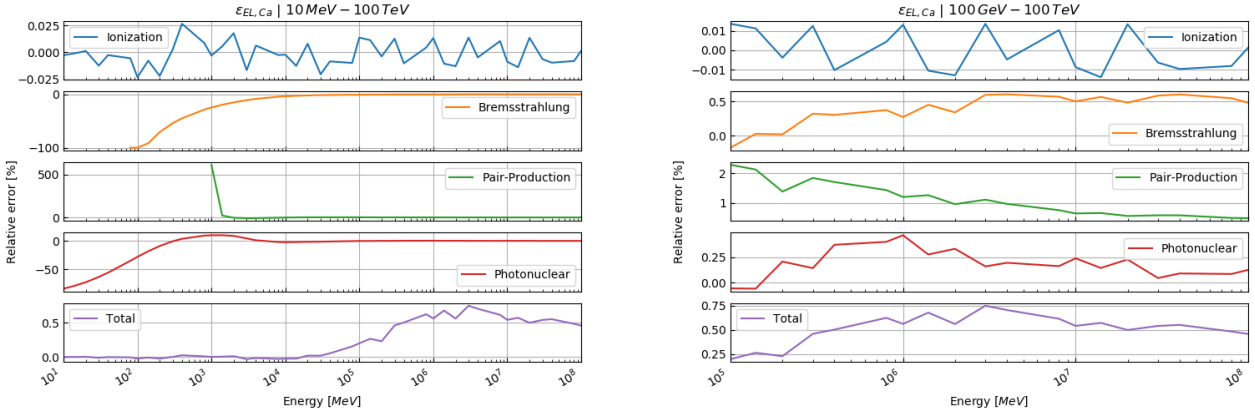


Figure E7: Relative error of our energy loss calculations compared to the tabulated values from Groom et al. (2001) for calcium in the energy ranges: (left) 10 MeV – 100 TeV, (right) 100 GeV – 100 TeV.

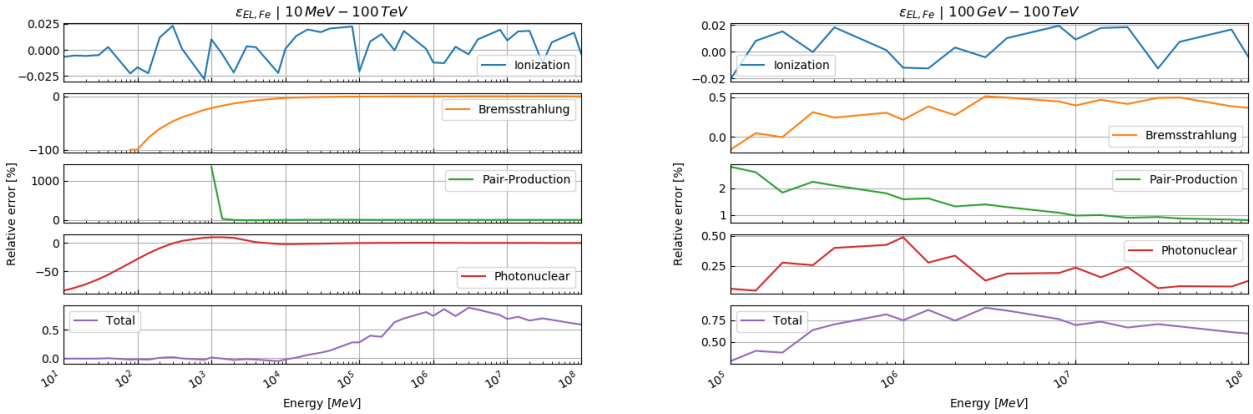


Figure E8: Relative error of our energy loss calculations compared to the tabulated values from Groom et al. (2001) for iron in the energy ranges: (left) 10 MeV – 100 TeV, (right) 100 GeV – 100 TeV.

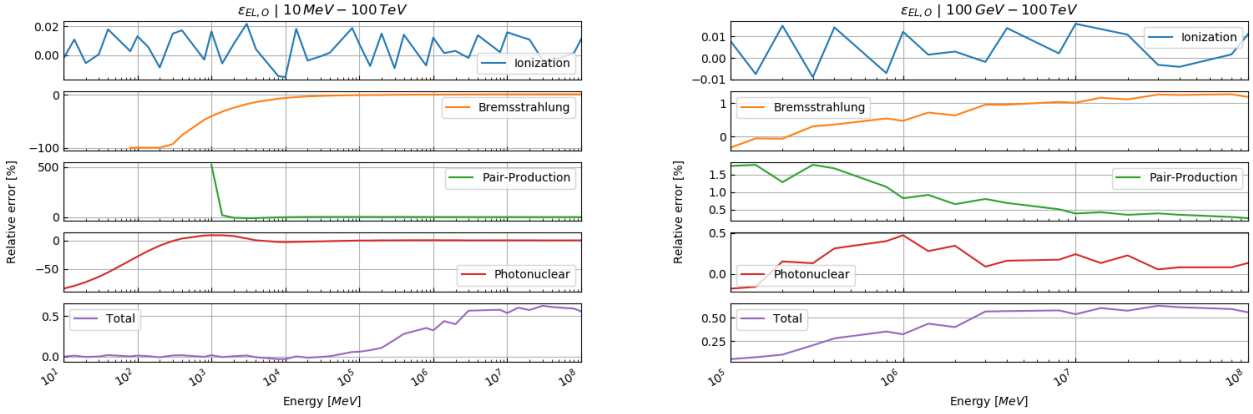


Figure E9: Relative error of our energy loss calculations compared to the tabulated values from Groom et al. (2001) for oxygen in the energy ranges: (left) 10 MeV – 100 TeV, (right) 100 GeV – 100 TeV.

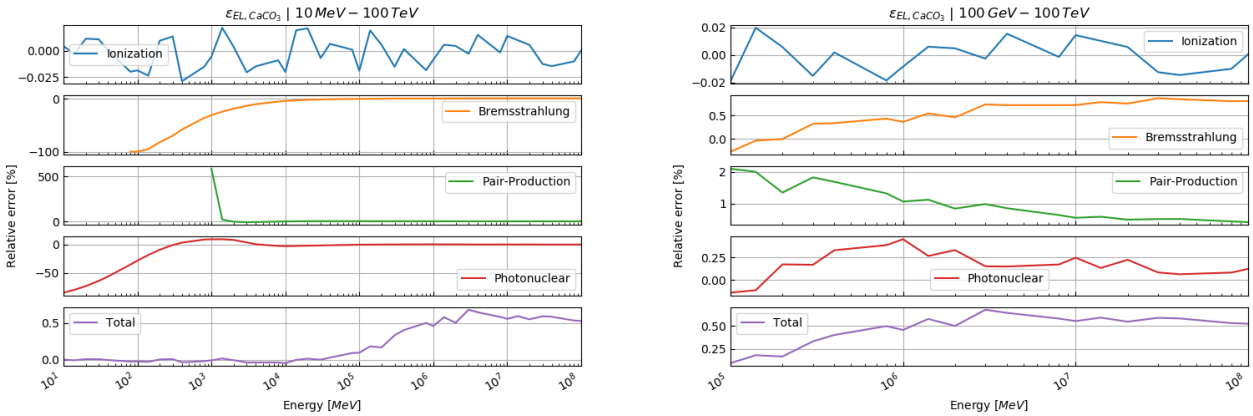


Figure E10: Relative error of our energy loss calculations compared to the tabulated values from Groom et al. (2001) for calcium carbonate (calcite) in the energy ranges: (left) 10 MeV – 100 TeV, (right) 100 GeV – 100 TeV.

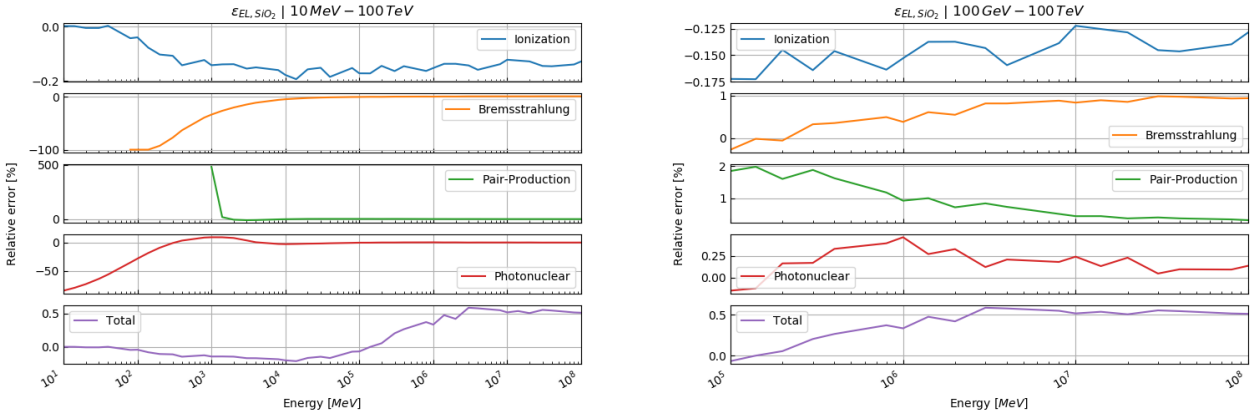


Figure E11: Relative error of our energy loss calculations compared to the tabulated values from Groom et al. (2001) for silicon dioxide (quartz) in the energy ranges: (left) 10 MeV – 100 TeV, (right) 100 GeV – 100 TeV.

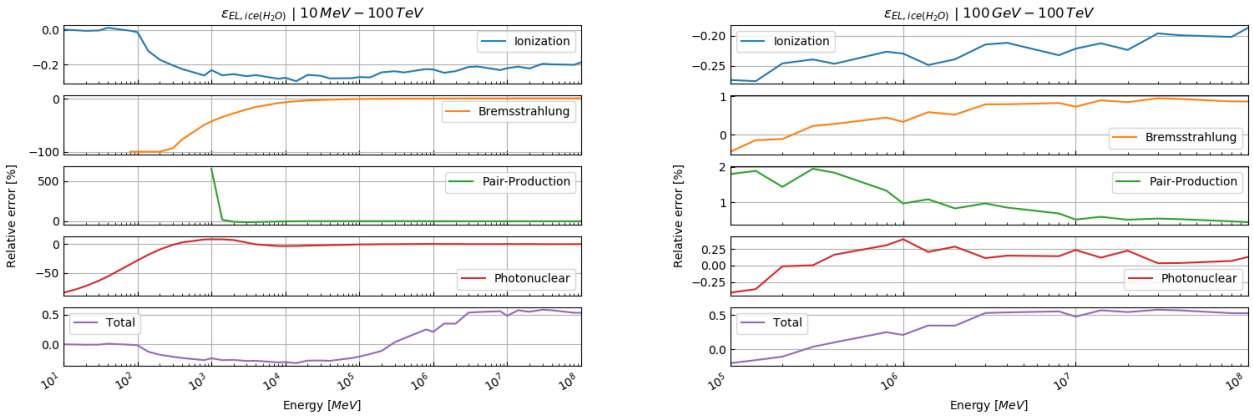


Figure E12: Relative error of our energy loss calculations compared to the tabulated values from Groom et al. (2001) for ice in the energy ranges: (left) 10 MeV – 100 TeV, (right) 100 GeV – 100 TeV.

Appendix F – Main modules of SMAUG

Our toolbox, SMAUG, contains several subprograms, which are executed separately. This allows the user to inspect intermediate results without any difficulty. We also tried to keep the intermediate results as portable as possible, by using JSON-files, as often as possible. Here we explain, in logical order, the rational of the submodules (a detailed user manual is separately available):

MATERIALIZER.py

This subroutine allows the user to create their own material that will be used in the subsequent model builder. The user may choose a density (either from measurements or directly insert mean and standard deviations) and a composition (also either from data or from the list of Groom et al., 2001).

DATA_BINNING.py

As the name suggests, this subroutine is used to spatially bin the recorded track data. The bin data (i.e. the output hereof) is then fed to the model builder.

MODEL_BUILDER.py

The model builder takes the bin data and the materials as inputs and allows the user, with help of DEMs, to allocate data and materials to certain cones. This is basically the spatial setup of the model. The resulting model file is then provided to the inversion code.

INVERSION.py

This is the main module in SMAUG, providing the functionality to perform a MCMC algorithm on the probabilistic model created with MODEL_BUILDER.py. It also includes several analysis tools to assess MCMC performance.

MODEL_VIEWER.py

The model viewer allows us to visualise the interface results, obtained and exported by INVERSION.py. It also has the functionality to dampen and smooth the resulting surfaces.

Code availability

The source code of SMAUG 1.0 is publicly and freely available on *https://doi.org/10.5281/zenodo.5547356* (Lechmann et al., 2021b). The python packages required to run SMAUG are listed in the "requirements.txt" file.

Data availability

The data of the density and XRF measurements are included (i) in the files that can be downloaded from *https://doi.org/10.5281/zenodo.5547356* as well as (ii) in the supplementary material to this publication. The raw data from the Nishiyama et al. (2017) paper is publicly and freely available from the publisher's website.

Author contributions

AL, FS and AE designed the study
AL developed the code with contributions by MV, CP and RN
AL performed the numerical experiments with support by RN
DM and AL compiled geological data
AA, TA, PS, RN and CP verified the outcome of the numerical experiments
AL wrote the text with contributions from all co-authors
AL designed the figures with contributions by DM
All co-authors contributed to the discussion and finally approved the manuscript

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

We thank the Swiss National Science Foundation (project No 159299 awarded to F. Schlunegger and A. Ereditato) for their financial support of this research project. Further, we want to thank the Jungfrau Railway Company for their continuing logistic support during our fieldwork in the central Swiss Alps. Finally, we want also to thank the High-Altitude Research Stations Jungfraujoch & Gornergrat for providing us with access to their research facilities and accommodation.

References

Agostinelli, S., et al.: Geant4—a simulation toolkit, Nucl. Instrum. Methods Phys. Res. Sect. Accel. Spectrometers Detect. Assoc. Equip., 506, 250–303, https://doi.org/10.1016/S0168-9002(03)01368-8, 2003.

Aitchison, J.: The Statistical Analysis of Compositional Data, Chapman and Hall Ltd, New York, 1986.

Alvarez, L. W., Anderson, J. A., El Bedwei, F., Burkhard, J., Fakhry, A., Girgis, A., Goneid, A., Hassan, F., Iverson, D., Lynch, G., Miligy, Z., Moussa, A. H., Mohammed-Sharkawi, and Yazolino, L.: Search for Hidden Chambers in the Pyramids, Science, 167, 832–839, 1970.

Ambrosino, F., Anastasio, A., Basta, D., Bonechi, L., Brianzi, M., Bross, A., Callier, S., Caputo, A., Ciaranfi, R., Cimmino, L., D'Alessandro, R., D'Auria, L., Taille, C. de L., Energico, S., Garufi, F., Giudicepietro, F., Lauria, A., Macedonio, G., Martini, M., Masone, V., Mattone, C., Montesi, M. C., Noli, P., Orazi, M., Passeggio, G., Peluso, R., Pla-Dalmau, A., Raux, L., Rubinov, P., Saracino, G., Scarlini, E., Scarpato, G., Sekhniaidze, G., Starodubtsev, O., Strolin, P., Taketa, A., Tanaka, H. K. M., Vanzanella, A., and Viliani, L.: The MU-RAY project: detector technology and first data from Mt. Vesuvius, J. Instrum., 9, C02029–C02029, https://doi.org/10.1088/1748-0221/9/02/C02029, 2014.

Ambrosino, F., Anastasio, A., Bross, A., Béné, S., Boivin, P., Bonechi, L., Cârloganu, C., Ciaranfi, R., Cimmino, L., Combaret, C., D'Alessandro, R., Durand, S., Fehr, F., Français, V., Garufi, F., Gailler, L., Labazuy, P., Laktineh, I., Lénat, J.-F., Masone, V., Miallier, D., Mirabito, L., Morel, L., Mori, N., Niess, V., Noli, P., Pla-Dalmau, A., Portal, A., Rubinov, P., Saracino, G., Scarlini, E., Strolin, P., and Vulpescu, B.: Joint measurement of the atmospheric muon flux through the Puy de Dôme volcano with plastic scintillators and Resistive Plate Chambers detectors, J. Geophys. Res. Solid Earth, 120, 7290–7307, https://doi.org/10.1002/2015JB011969, 2015.

Anghel, V., Armitage, J., Baig, F., Boniface, K., Boudjemline, K., Bueno, J., Charles, E., Drouin, P.-L., Erlandson, A., Gallant, G., Gazit, R., Godin, D., Golovko, V. V., Howard, C., Hydomako, R., Jewett, C., Jonkmans, G., Liu, Z., Robichaud, A., Stocki, T. J., Thompson, M., and Waller, D.: A plastic scintillator-based muon tomography system with an integrated muon spectrometer, Nucl. Instrum. Methods Phys. Res. Sect. Accel. Spectrometers Detect. Assoc. Equip., 798, 12–23, https://doi.org/10.1016/j.nima.2015.06.054, 2015.

Ariga, A., Ariga, T., Ereditato, A., Käser, S., Lechmann, A., Mair, D., Nishiyama, R., Pistillo, C., Scampoli, P., Schlunegger, F., and Vladymyrov, M.: A Nuclear Emulsion Detector for the Muon Radiography of a Glacier Structure, Instruments, 2, 7, https://doi.org/10.3390/instruments2020007, 2018.

ASTM C914-09: Standard test method for bulk density and volume of solid refractories by wax immersion, https://doi.org/10.1520/C0914-09R15, 2015.

Barnaföldi, G. G., Hamar, G., Melegh, H. G., Oláh, L., Surányi, G., and Varga, D.: Portable cosmic muon telescope for environmental applications, Nucl. Instrum. Methods Phys. Res. Sect. Accel. Spectrometers Detect. Assoc. Equip., 689, 60–69, https://doi.org/10.1016/j.nima.2012.06.015, 2012.

Barnoud, A., Cayol, V., Niess, V., Cârloganu, C., Lelièvre, P., Labazuy, P., and Le Ménédeu, E.: Bayesian joint muographic and gravimetric inversion applied to volcanoes, Geophys. J. Int., 218, 2179–2194, https://doi.org/10.1093/gji/ggz300, 2019.

Bellman, R. E.: Adaptive Control Processes A Guided Tour, Princeton University Press, Princeton, 2016.

Betancourt, M. J. and Girolami, M.: Hamiltonian Monte Carlo for Hierarchical Models, ArXiv13120906 Stat, 2013.

Blake, G. R. and Hartge, K. H.: Bulk Density, in: Methods of Soil Analysis: Part 1—Physical and Mineralogical Methods, vol. 9, edited by: Klute, A., American Society of Agronomy and Soil Science Society of America, Madison, 363–375, 1986.

Bonechi, L., D'Alessandro, R., Mori, N., and Viliani, L.: A projective reconstruction method of underground or hidden structures using atmospheric muon absorption data, JINST, 10, P02003, https://doi.org/10.1088/1748-0221/10/02/P02003, 2015.

Bugaev, E. V., Misaki, A., Naumov, V. A., Sinegovskaya, T. S., Sinegovsky, S. I., and Takahashi, N.: Atmospheric Muon Flux at Sea Level, Underground, and Underwater, Phys. Rev. D, 58, 054001, https://doi.org/10.1103/PhysRevD.58.054001, 1998.

Gelman, A.: Bayesian data analysis, Third edition., CRC Press, Boca Raton, 2014.

Gerya, T.: Introduction to numerical geodynamic modelling, Cambridge University Press, Cambridge, New York, 2010.

Groom, D. E., Mokhov, N. V., and Striganov, S. I.: MUON STOPPING POWER AND RANGE TABLES 10 MeV–100 TeV, At. Data Nucl. Data Tables, 78, 183–356, https://doi.org/10.1006/adnd.2001.0861, 2001.

Guardincerri, E., Rowe, C., Schultz-Fellenz, E., Roy, M., George, N., Morris, C., Bacon, J., Durham, M., Morley, D., Plaud-Ramos, K., Poulson, D., Baker, D., Bonneville, A., and Kouzes, R.: 3D Cosmic Ray Muon Tomography from an Underground Tunnel, Pure Appl. Geophys., 174, 2133–2141, https://doi.org/10.1007/s00024-017-1526-x, 2017.

Hastings, W. K.: Monte Carlo sampling methods using Markov chains and their applications, Biometrika, 57, 97–109, https://doi.org/10.1093/biomet/57.1.97, 1970.

Hebbeker, T. and Timmermans, C.: A Compilation of High Energy Atmospheric Muon Data at Sea Level, Astropart. Phys., 18, 107–127, https://doi.org/10.1016/S0927-6505(01)00180-3, 2002.

Jonkmans, G., Anghel, V. N. P., Jewett, C., and Thompson, M.: Nuclear waste imaging and spent fuel verification by muon tomography, Ann. Nucl. Energy, 53, 267–273, https://doi.org/10.1016/j.anucene.2012.09.011, 2013.

Jourde, K., Gibert, D., and Marteau, J.: Improvement of density models of geological structures by fusion of gravity data and cosmic muon radiographies, Geosci. Instrum. Method. Data Syst., 4, 177-188, https://doi.org/10.5194/gi-4-177-2015, 2015.

Jourde, K., Gibert, D., Marteau, J., de Bremond d'Ars, J., and Komorowski, J.-C.: Muon dynamic radiography of density changes induced by hydrothermal activity at the La Soufrière of Guadeloupe volcano, Sci. Rep., 6, 33406, https://doi.org/10.1038/srep33406, 2016.

Kjaerulff, U. B. and Madsen, A. L.: Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis, Springer, New York, 2008.

Kudryavtsev, V. A.: Muon simulation codes MUSIC and MUSUN for underground physics, Comput. Phys. Commun., 180, 339–346, https://doi.org/10.1016/j.cpc.2008.10.013, 2009.

Kusagaya, T. and Tanaka, H. K. M.: Development of the very long-range cosmic-ray muon radiographic imaging technique to explore the internal structure of an erupting volcano, Shinmoe-dake, Japan, Geosci. Instrum. Methods Data Syst., 4, 215–226, https://doi.org/10.5194/gi-4-215-2015, 2015.

Lechmann, A., Mair, D., Ariga, A., Ariga, T., Ereditato, A., Nishiyama, R., Pistillo, C., Scampoli, P., Schlunegger, F., and Vladymyrov, M.: The effect of rock composition on muon tomography measurements, Solid Earth, 9, 1517–1533, https://doi.org/10.5194/se-9-1517-2018, 2018.

Lechmann, A., Mair, D., Ariga, A., Ariga, T., Ereditato, A., Nishiyama, R., Pistillo, C., Scampoli, P., Schlunegger, F., and Vladymyrov, M.: Muon tomography in geoscientific research – a guide to best practice, ~~Manuscript submitted for publication,~~Earth-Science Reviews, 222, 103842, https://doi.org/10.1016/j.earscirev.2021.103842, 2021a.

Lechmann, A., Mair, D., Ariga, A., Ariga, T., Ereditato, A., Nishiyama, R., Pistillo, C., Scampoli, P., Schlunegger, F., and Vladymyrov, M.: SMAUG – A Simulator for Muon Applications Under Ground (Version 1.0), Zenodo, https://doi.org/10.5281/zenodo.5547356, 2021b.

Lelièvre, P. G., Barnoud, A., Niess, V., Cârloganu, C., Cayol, V., and Farquharson, C. G.: Joint inversion methods with relative density offset correction for muon tomography and gravity data, with application to volcano imaging, Geophys. J. Int., 218, 1685-1701, https://doi.org/10.1093/gji/ggz251, 2019.

Lesparre, N., Gibert, D., Marteau, J., Déclais, Y., Carbone, D., and Galichet, E.: Geophysical muon imaging: feasibility and limits, Geophys. J. Int., 183, 1348–1361, https://doi.org/10.1111/j.1365-246X.2010.04790.x, 2010.

Lesparre, N., Gibert, D., and Marteau, J.: Bayesian dual inversion of experimental telescope acceptance and integrated flux for geophysical muon tomography, Geophys. J. Int., 188, 490–497, https://doi.org/10.1111/j.1365-246X.2011.05268.x, 2012.

Lesparre, N., Cabrera, J., and Marteau, J.: 3-D density imaging with muon flux measurements from underground galleries, Geophys. J. Int., 208, 1579–1591, https://doi.org/10.1093/gji/ggw482, 2017.

Lo Presti, D., Gallo, G., Bonanno, D. L., Bonanno, G., Bongiovanni, D. G., Carbone, D., Ferlito, C., Immè, J., La Rocca, P., Longhitano, F., Messina, A., Reito, S., Riggi, F., Russo, G., and Zuccarello, L.: The MEV project: Design and testing of a new high-resolution telescope for muography of Etna Volcano, Nucl. Instrum. Methods Phys. Res. Sect. Accel. Spectrometers Detect. Assoc. Equip., 904, 195–201, https://doi.org/10.1016/j.nima.2018.07.048, 2018.

Lohmann, W., Kopp, R., and Voss, R.: Energy Loss of Muons in the Energy Range 1-10000 GeV, CERN, Geneva, 1985.

Mair, D., Lechmann, A., Herwegh, M., Nibourel, L., and Schlunegger, F.: Linking Alpine deformation in the Aar Massif basement and its cover units – the case of the Jungfrau-Eiger mountains (Central Alps, Switzerland), Solid Earth, 9, 1099-1122, https://doi.org/10.5194/se-9-1099-2018, 2018.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E.: Equation of State Calculations by Fast Computing Machines, J. Chem. Phys., 21, 1087–1092, https://doi.org/10.1063/1.1699114, 1953.

Morishima, K., Kuno, M., Nishio, A., Kitagawa, N., Manabe, Y., Moto, M., Takasaki, F., Fujii, H., Satoh, K., Kodama, H., Hayashi, K., Odaka, S., Procureur, S., Attié, D., Bouteille, S., Calvet, D., Filosa, C., Magnier, P., Mandjavidze, I., Riallot, M., Marini, B., Gable, P., Date, Y., Sugiura, M., Elshayeb, Y., Elnady, T., Ezzy, M., Guerriero, E., Steiger, V., Serikoff, N., Mouret, J.-B., Charlès, B., Helal, H., and Tayoubi, M.: Discovery of a big void in Khufu's Pyramid by observation of cosmic-ray muons, Nature, 552, 386–390, https://doi.org/10.1038/nature24647, 2017.

Niess, V., Barnoud, A., Cârloganu, C., and Le Ménédeu, E.: Backward Monte-Carlo applied to muon transport, Comput. Phys. Commun., 229, https://doi.org/10.1016/j.cpc.2018.04.001, 2018.

Nishiyama, R., Tanaka, Y., Okubo, S., Oshima, H., Tanaka, H. K. M., and Maekawa, T.: Integrated processing of muon radiography and gravity anomaly data toward the realization of high-resolution 3-D density structural analysis of volcanoes:

Case study of Showa-Shinzan lava dome, Usu, Japan, J. Geophys. Res. Solid Earth, 119, 699–710, https://doi.org/10.1002/2013JB010234, 2014.

Nishiyama, R., Ariga, A., Ariga, T., Käser, S., Lechmann, A., Mair, D., Scampoli, P., Vladymyrov, M., Ereditato, A., and Schlunegger, F.: First measurement of ice-bedrock interface of alpine glaciers by cosmic muon radiography, Geophys. Res. Lett., 44, 6244–6251, https://doi.org/10.1002/2017GL073599, 2017.

Nishiyama, R., Ariga, A., Ariga, T., Lechmann, A., Mair, D., Pistillo, C., Scampoli, P., Valla, P. G., Vladymyrov, M., Ereditato, A., and Schlunegger, F.: Bedrock sculpting under an active alpine glacier revealed from cosmic-ray muon radiography, Sci. Rep., 9, 6970, https://doi.org/10.1038/s41598-019-43527-6, 2019.

Noli, P., Ambrosino, F., Bonechi, L., Bross, A., Cimmino, L., D'Alessandro, R., Masone, V., Mori, N., Passeggio, G., Pla-Dalmau, A., Saracino, G., Scarlini, E., and Strolin, P.: Muography of the Puy de Dôme, Ann. Geophys., 60, https://doi.org/10.4401/ag-7380, 2017.

Oláh, L., Barnaföldi, G. G., Hamar, G., Melegh, H. G., Surányi, G., and Varga, D.: Cosmic Muon Detection for Geophysical Applications, Advances in High Energy Physics, https://doi.org/10.1155/2013/560192, 2013.

Oláh, L., Tanaka, H. K. M., Ohminato, T., and Varga, D.: High-definition and low-noise muography of the Sakurajima volcano with gaseous tracking detectors, Sci. Rep., 8, 3207, https://doi.org/10.1038/s41598-018-21423-9, 2018.

Reyna, D.: A Simple Parameterization of the Cosmic-Ray Muon Momentum Spectra at the Surface as a Function of Zenith Angle, arXiv Prepr. hep-ph/0604145 [online] Available from: http://arxiv.org/abs/hep-ph/0604145, 2006.

Rosas-Carbajal, M., Jourde, K., Marteau, J., Deroussi, S., Komorowski, J.-C., and Gibert, D.: Three-dimensional density structure of La Soufrière de Guadeloupe lava dome from simultaneous muon radiographies and gravity data: 3-D MUON TOMOGRAPHY OF LA SOUFRIÈRE, Geophys. Res. Lett., 44, 6743–6751, https://doi.org/10.1002/2017GL074285, 2017.

Saracino, G., Amato, L., Ambrosino, F., Antonucci, G., Bonechi, L., Cimmino, L., Consiglio, L., Alessandro, R. D. ', Luzio, E. D., Minin, G., Noli, P., Scognamiglio, L., Strolin, P., and Varriale, A.: Imaging of underground cavities with cosmic-ray muons from observations at Mt. Echia (Naples), Sci. Rep., 7, 1181, https://doi.org/10.1038/s41598-017-01277-3, 2017.

Stoer, J. and Bulirsch, R.: Introduction to Numerical Analysis, Springer Science & Business Media, New York, 2013.

Takahasi, H. and Mori, M.: Double Exponential Formulas for Numerical Integration, Publ. Res. Inst. Math. Sci., 9, 721–741, https://doi.org/10.2977/prims/1195192451, 1974.

Takamatsu, K., Takegami, H., Ito, C., Suzuki, K., Ohnuma, H., Hino, R., and Okumura, T.: Cosmic-ray muon radiography for reactor core observation, Ann. Nucl. Energy, 78, 166–175, https://doi.org/10.1016/j.anucene.2014.12.017, 2015.

Tanabashi, M., et al. (Particle Data Group): Review of Particle Physics, Phys. Rev. D, 98, 030001, https://doi.org/10.1103/PhysRevD.98.030001, 2018.

Tanaka, H. K. M.: Instant snapshot of the internal structure of Unzen lava dome, Japan with airborne muography, Sci. Rep., 6, 39741, https://doi.org/10.1038/srep39741, 2016.

Tang, A., Horton-Smith, G., Kudryavtsev, V. A., and Tonazzo, A.: Muon Simulations for Super-Kamiokande, KamLAND and CHOOZ, Phys. Rev. D, 74, 053007, https://doi.org/10.1103/PhysRevD.74.053007, 2006.

Tarantola, A.: Inverse Problem Theory and Methods for Model Parameter Estimation, 1st ed., siam, Philadelphia, 2005.

Thompson, L. F., Stowell, J. P., Fargher, S. J., Steer, C. A., Loughney, K. L., O'Sullivan, E. M., Gluyas, J. G., Blaney, S. W., and Pidcock, R. J.: Muon tomography for railway tunnel imaging, Phys. Rev. Res., 2, 023017, https://doi.org/10.1103/PhysRevResearch.2.023017, 2020.

Tioukov, V., Lellis, G. D., Strolin, P., Consiglio, L., Sheshukov, A., Orazi, M., Peluso, R., Bozza, C., Sio, C. D., Stellacci, S. M., Sirignano, C., D'Ambrosio, N., Miyamoto, S., Nishiyama, R., and Tanaka, H. K. M.: Muography with nuclear emulsions - Stromboli and other projects, Ann. Geophys., 60, 0111, https://doi.org/10.4401/ag-7386, 2017.

Vermeesch, P.: On the visualisation of detrital age distributions, Chemical Geology, 312-313, 190-194, https://doi.org/10.1016/j.chemgeo.2012.04.021, 2012.