Project 1: Snow particles

MATH-516 Applied Statistics

Linda Mhalla

2026-02-16

1 Data

- data from a (former) PhD student at the Laboratory of Cryospheric Sciences at EPFL, essentially snow-flake diameters
  - shared with the permission of the authors of this paper
- the total number of particles measured (variable `particles.detected`) and the fraction (variable `retained [%]`) of particles belonging to each diameter bin (given by `startpoint` and `endpoint`)
  - only binned data are available (and the grid is not equidistant)

```
  X startpoint endpoint retained.... particles.detected
1 1      0.000    0.060          3.3             705044
2 2      0.060    0.065          0.8             705044
3 3      0.065    0.070          0.9             705044
4 4      0.070    0.076          1.1             705044
5 5      0.076    0.082          1.3             705044
6 6      0.082    0.089          1.3             705044
```

2 The Goal

Melo et al. (2022) show that grain size distribution of surface snow significantly modifies dynamics of wind-driven snow transport (saltation) $\Rightarrow$ realistic snow transport modelling requires an accurate probabilistic description of particle diameters

The goal is to simulate diameters from a distribution, which is as close as possible to the observed data, in order to study aeolian transport of snow using certain numerical models

- i.e., the goal is to do Monte Carlo: **how to simulate snow-flake diameters that are compatible with the data?**

Expert knowledge: a mixture of two log-normal distributions is a good model for the diameters

3 Tasks for You

1. Is the assumption viable, i.e. is bi-log-normal distribution a reasonable model for the data?
   - simple exploration of the data
2. Write down the likelihood of the binned data AND the likelihood of the jittered data
3. Fit the bi-log-normal distribution in order to be able to simulate the data easily using
   - jittering and EM algorithm OR direct optimisation (e.g., local search starting from the jittered EM result), AND
   - a Bayesian approach
4. Test whether the diameters come from a bi-log-normal distribution
   - parametric bootstrap and goodness of fit

4 MATH-517 Content

- Week 1.1: Introduction & Software & Data Considerations
- Week 1.2: Graphics & Visualization
- **Week 2: Kernel Density Estimation**
- Week 3: Non-parametric Regression
- **Week 4: Cross-validation**
- **Week 5: EM Algorithm**
- **Week 6: EM Algorithm**
- **Week 7: Monte Carlo**
- **Week 8: Bootstrap**
- **Week 9: Bootstrap**
- **Week 10: Bayesian Computations**
- **Week 11: Bayesian Computations**
- Week 12: Decision Trees
- Week 13: Conformal Prediction
- Week 14: $\emptyset$
    - Weeks in bold are pertinent to Project 1
    - Weeks 1.1-1.2 established the workflow needed for all the projects