Machine learning and snowflakes

A. Berne

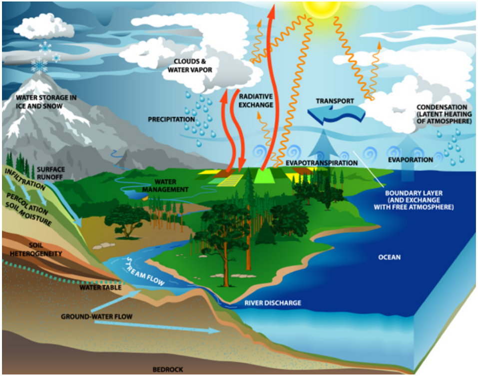
Environmental Remote Sensing Laboratory
EPFL, Lausanne, Switzerland

Applied Machine Learning Days 2019 - Lausanne, 29 Jan. 2019

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Importance of precipitation

Precipitation = flux of (solid/liquid) water from atmosphere to Earth surface.



- Crucial for the water cycle, the climate system and ecosystems.
- At the interface of many environmental disciplines.

Societal impacts

Floods

Agriculture

Transportation

Power supply

Importance of snowfall

- Significant if not dominant in total precip at high altitudes/latitudes.
- How much will it snow? Down to which altitude? What kind of snow?

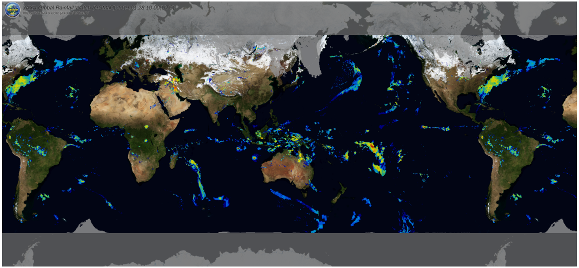
Water storage

Traffic...

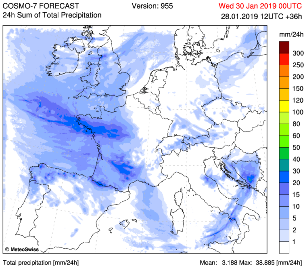
Mountain tourism

- Modelling and forecasting of (solid) precipitation is difficult.
- Cold microphysics is not as well understood as warm microphysics.

Machine Learning?



Global map of precipitation - 28/01/2019 (source: GSMaP)

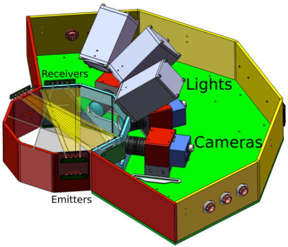


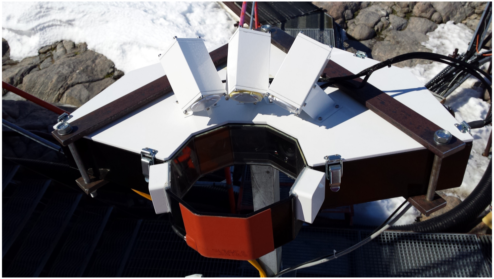
Forecasted precip (source: MeteoSwiss)

- Development of remote sensing capabilities (satellite + ground-based).
- Development of NWP and climate models (more processes + higher res.)
- Large multidimensional data sets with large (subgrid) variability
  $\Rightarrow$ potential for ML approaches to extract relevant information...

Outline

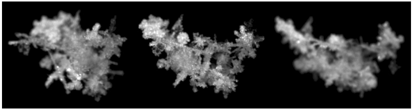
The Multi-Angle Snowflake Camera





Measurements

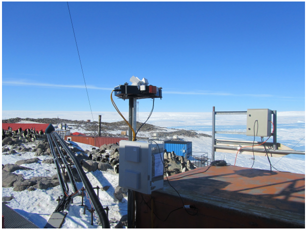
- 3 images from 3 different coplanar angles.
- Resolution $\sim$ 33 μm, sampling area $\sim$ 8.3 cm^2.
- Falling velocity [m.s^{-1}].



Data sets

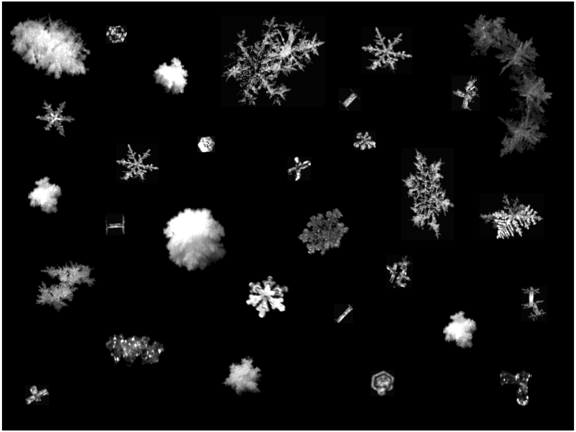
Dumont d'Urville (2015-16,2017)

Davos (2015-16)



About 500k images in each data set...

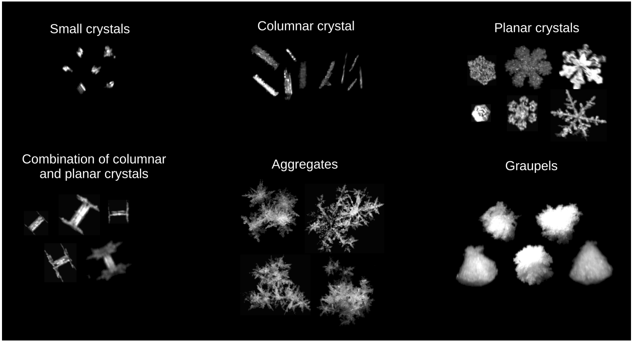
Examples of snowflakes pictured by the MASC



- Variability of shape, size, etc.
- Human eye/brain can distinguish and group pictures.
- Huge number of pictures → we need an automatic classification.

Output of classification scheme (1)

- Started with 10 main categories taken from Magono and Lee (1966).
- Removed classes rarely observed (germ of snow, comb. of planar crystals).
- Added aggregates and small crystals.
- Merged similar classes (col. - needles ; plates / sectored plates / dend.).



Output of classification scheme (2)

Complementary information : riming degree and melting or not.

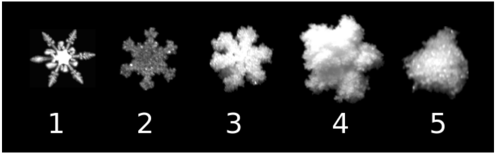
Riming degree

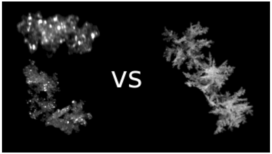
- Continuous value in general...
- Trained as ordinal value from 1 to 5.

(adapted from *Mosimann et al. 1994*)

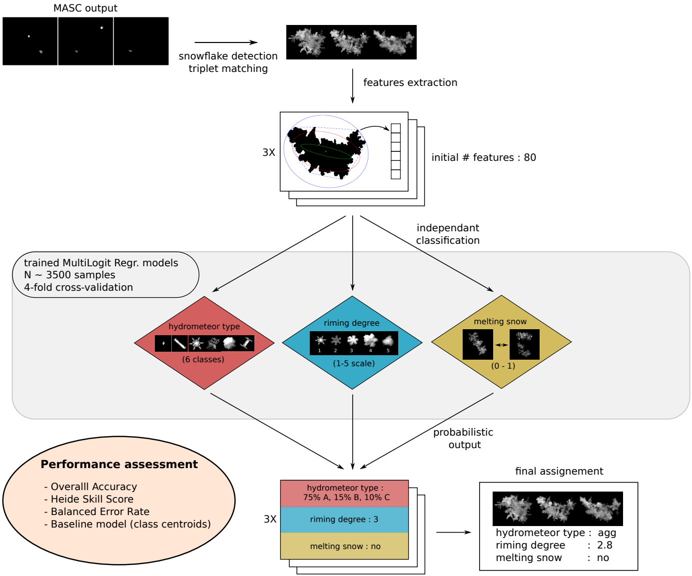
Melting snow

- Cont. value in general...
- Here: boolean (dry / melting).



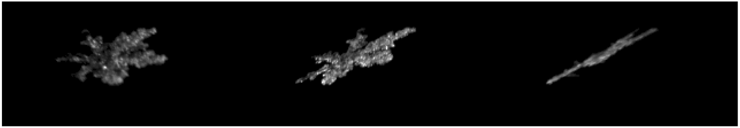


Summarizing flowchart



Application to MASC image triplets

Example:



65% planar crystal	71% planar crystal	~97% columnar crystal
30% aggregate	26% aggregate	3% planar crystal
5% others	3% others	~0% others
Riming = 2.2	Riming = 2.5	Riming = 2.0
Dry	Dry	Dry

$\Rightarrow$ Dry, moderately rimed planar crystal.

Blowing snow vs snowfall

- Blowing snow = ground snow particles suspended by wind.
- Snowfall = snowflakes formed in atm. and falling to the surface.

$\Rightarrow$ Crucial to distinguish the two for water mass balance.

Visible over ridges

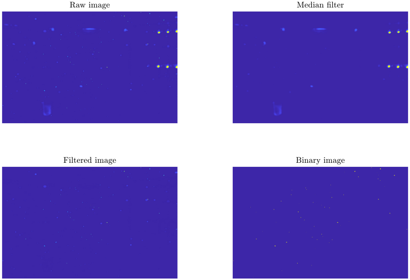
http://la-haut.over-blog.fr/2014/01/2-jours-en-foret-dans-la-neige.html

Significant over polar ice sheets

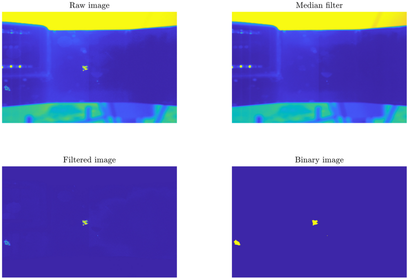
Blowing snow from MASC

Thanks to campaigns in the Swiss Alps (Davos) and in Antarctica (DDU)
$\rightarrow$ MASC data $\pm$ contaminated by blowing snow.

MASC image with BS

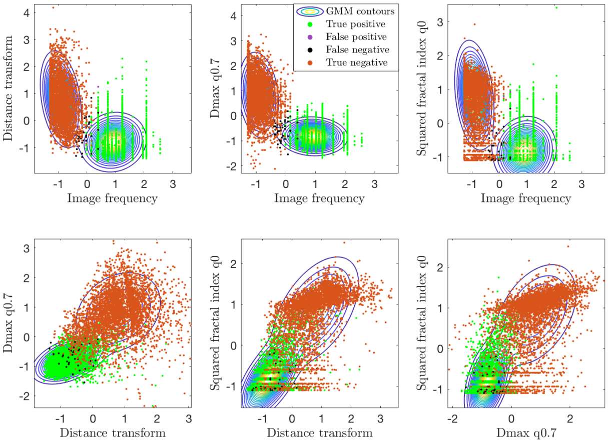


MASC image without BS



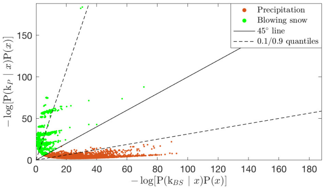
Classification using GMM

- 4 features related to image frequency; particle number + size + geometry.
- Training set manually built (4623 images of precipitation, BS and mixture).

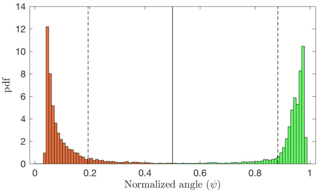


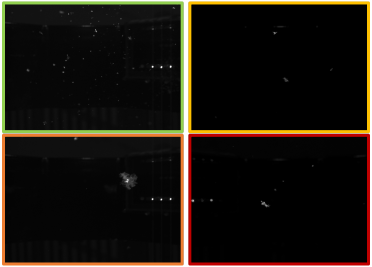
Mixed images

- Challenge: when BS and precip at the same time...
- GMM: probability for each GM → angle $\psi = \frac{2}{\pi} \arctan\left\{ \frac{-\log[P(k_P|\mathbf{x}_i)P(\mathbf{x}_i)]}{-\log[P(k_{BS}|\mathbf{x}_i)P(\mathbf{x}_i)]} \right\}$

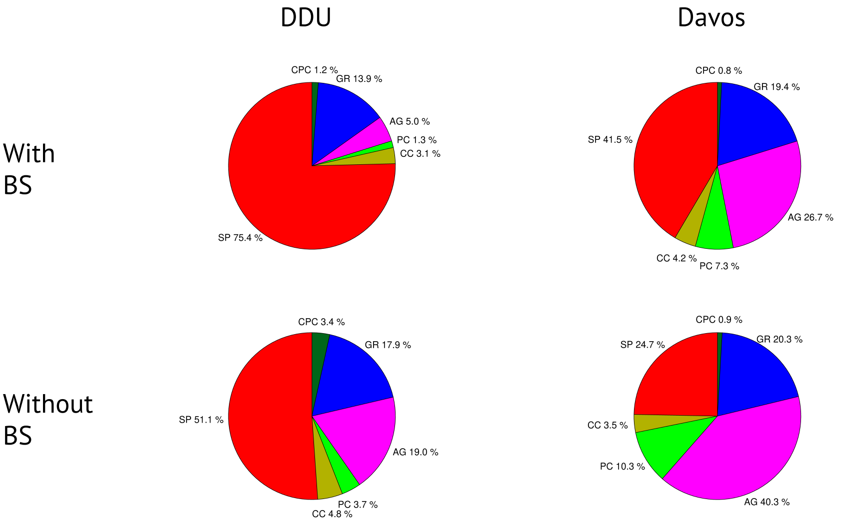


date_vec_unique	1 ID	2 Label	3 Normalized_Angle	4 Flag_mixed
25/03/2016 16:56:35	7383	1	0.7422	0.8178
25/03/2016 16:56:39	7384	0	0.1590	NaN
25/03/2016 16:56:39	7385	1	0.5659	0.5865
25/03/2016 16:56:40	7386	0	0.1437	NaN
25/03/2016 16:56:42	7387	0	0.1920	NaN
25/03/2016 16:56:42	7388	0	0.2260	0.0537





Outcome for snowflake type statistics



Summary

Snowflake classification

- MASC provides high-resolution photographs of snowflakes.
- MLR used to automatize the classification of large MASC data sets.
- Subsequent analysis of snowflake microphysical properties useful for process understanding and model parameterization improvement.

Blowing snow vs snowfall

- Ground-level measurements of snowfall potentially contaminated by blowing snow.
- GMM were employed to distinguish and quantify mixing of blowing snow and snowfall particles.

ML and precipitation

A variety of potential applications

- Parameterization of subgrid processes in climate/meteo models.
- Downscaling: use large-scale information to constrain downscaled fields at higher resolutions.
- Precipitation estimation (e.g. by data fusion) and nowcasting (extrapolation of spatial field).
- Early warning for precipitation-related natural hazards (e.g. floods, avalanches, landslides, debris flows).

Challenges for / expectations from ML

- How to deal with non-stationary uncertainty in observations / references?
- Include/combine a priori knowledge and physics, possibly at different scales, in the learning step.
- Provide information about important processes at work.
- Provide probabilistic ouput suitable for many env./climate variables.

Thank you