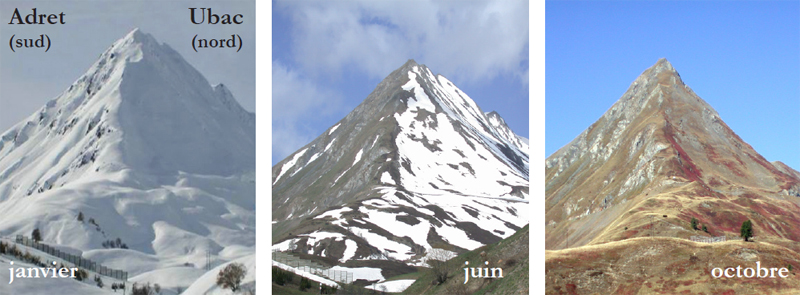
When I was hiking, skiing or cycling in the French Alps, I was always wondering how the snow falling and melting processes work. It is quite complicated and depends on many features. Snow falling depends mostly of the fresh snow, of course. Snow melting is more complicated: it is due to :

* the air temperature
* the solar insolation and energy, due to energy
  + the mountain slope combined to the Sun ray. Snow on a south-facing slope melts faster and on a north slope. You can especially see that difference in April, May and June
* the rain, especially if that rainwater is warm
* some wind can also snatch the upper snow levels
* the mass of the snow also : a big mass of snow will, due to calorific features, tend to need more time to melt. That is what we see with Greenland of some mountain glaciers which are still in place while the temperature is above 20°C
* some other local features that are difficult to find and difficult to measure physically. You can sometimes see covered by a one meter layer of snow next to a snow-free area, while the physical values I talked about previously were the same.

So, instead of trying to use plenty of thermodynamical formulas (and I forgot most of them, haha, or don’t know them), why not trying to use some Deep Learning methods ? If I can find enough data combining all those features in the French mountais on an opendata website, the Recurrent networks seem to be an interesting method.

*Extreme example of what we can see in the Northern hemisphere mountains from April to June: on the south-facing slope(on the left) snow has completely melted due to a longer Sun exposure rather than air temperature. On the north-facing slope (on the right), snow remains*

Luckily Météo-France, the French national meteorological service, updated on their opendata branch some interesting data of the snow height and other weather features in the mountains. The website link is :

[Données Publiques de Météo-France - Observations du réseau nivo-météorologique (meteofrance.fr)](https://donneespubliques.meteofrance.fr/?fond=produit&id_produit=94&id_rubrique=32)

Data is in .csv format. There is one .csv file for each month. What we can notice :

* Measures go as far as January 1996. For the current data, data are updated daily.
* Measure points are mostly situated in winter sport resorts, very close to the villages. A minority is located higher in the mountains, mostly near cable car stations
* Mostly of point of measurement per winter resort area. Sometimes 2 or 3. Chamonix has for example 4 points.
* Available features are explained here :

<https://donneespubliques.meteofrance.fr/client/document/doc_parametres_nivo_197.pdf>

Good news : among those features, those I talked about at the beginning are here : snow height of course, wind, rain, temperatures, clouds.. And other important geographical features : latitude, longitude, altitude, station name..

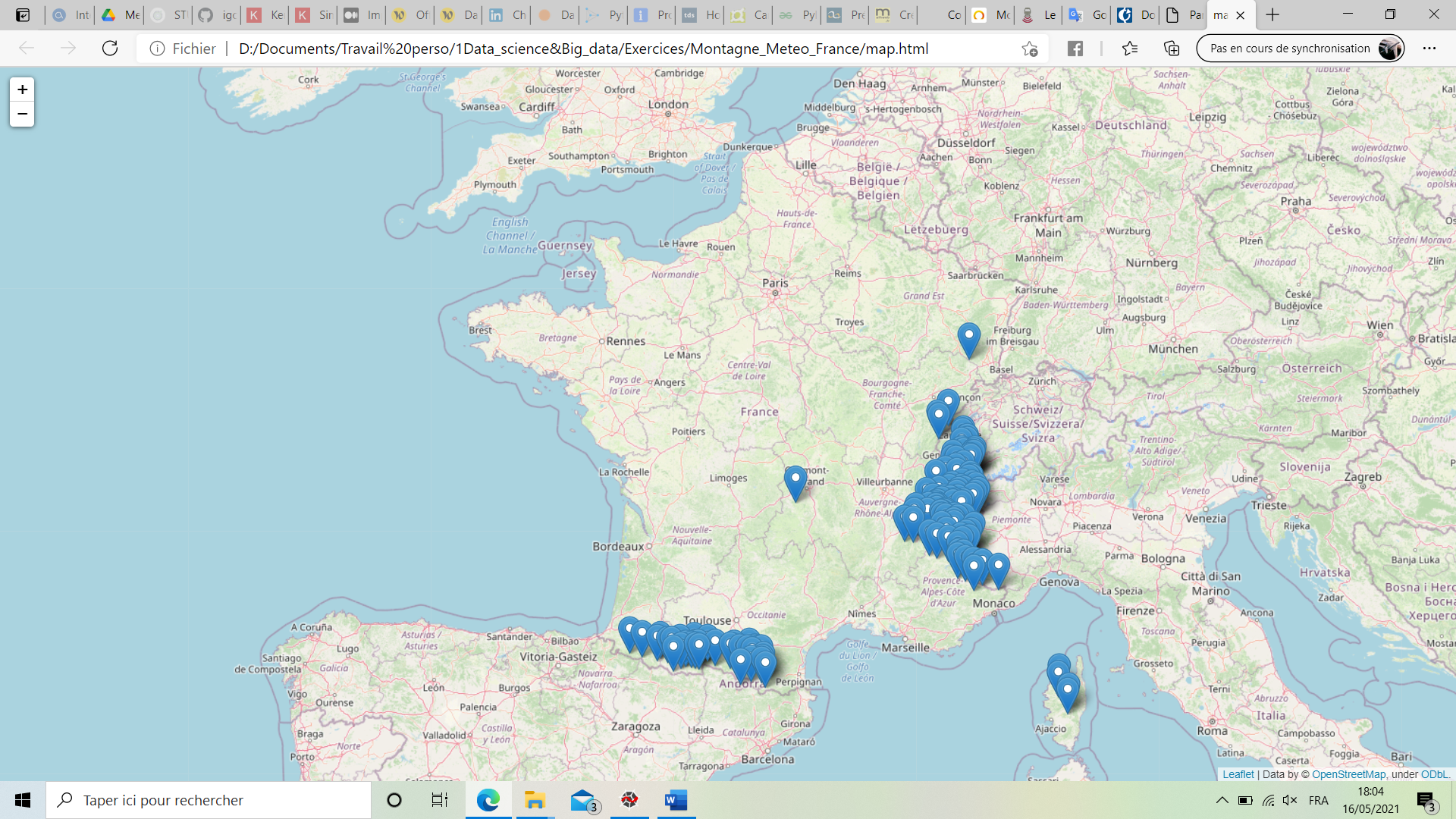
Which results in more than 647 000 measures in all french mountains from January 1996 to May 2021

1. STATIONS

Inspired by this article :

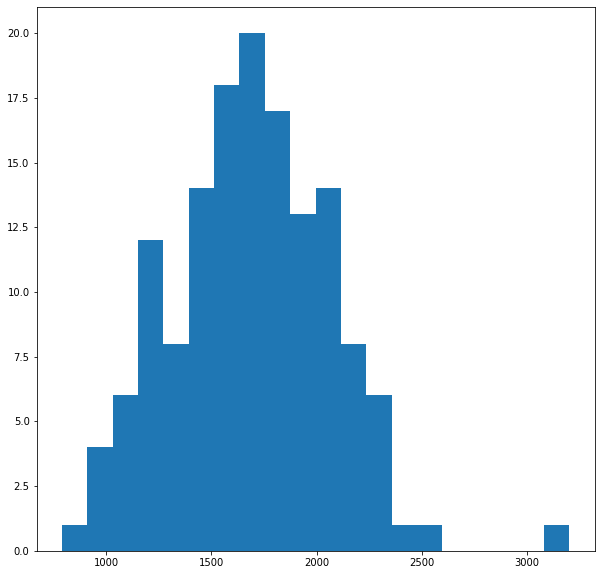
[Cartographier le manteau neigeux avec Python — Makina Corpus (makina-corpus.com)](https://makina-corpus.com/blog/metier/2019/python-carto)

[[[[[[ PYTHON PLOT OPENSTREETMAP CODE ]]]]]]



That projection on an OpenstreetMap data, using the Python folium package, shows that the points are mostly located in the French Alps, but also an important part in the Pyrenees, and some on Corsican mountains, a few in the Jura, one in the “Massif Central” and one in the Vosges.

If I look at the altitude distribution of the stations :



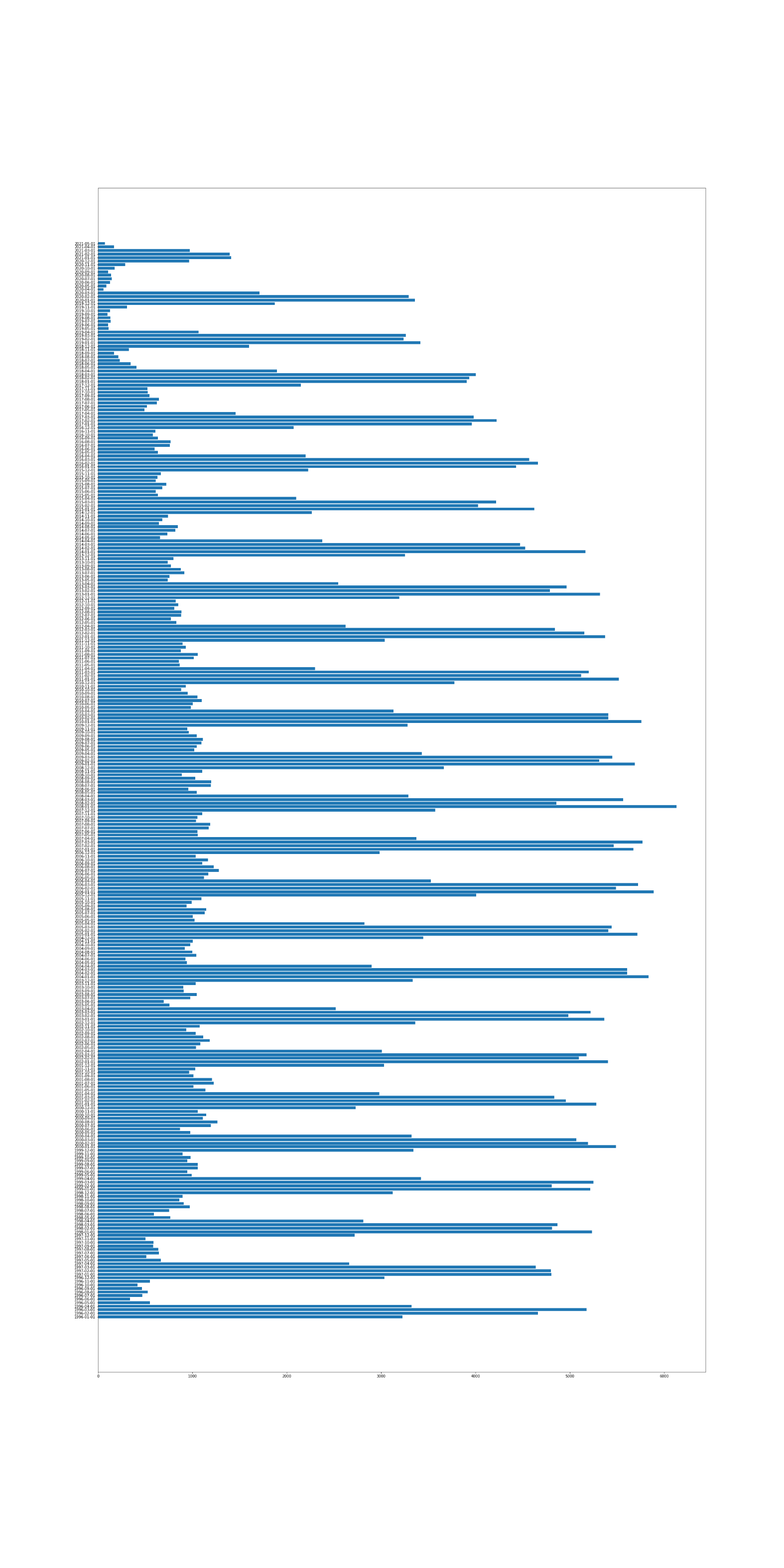
The distribution is centered on ~1700 meters. The mean is 1693 meters

1. FEATURES
2. FEATURES

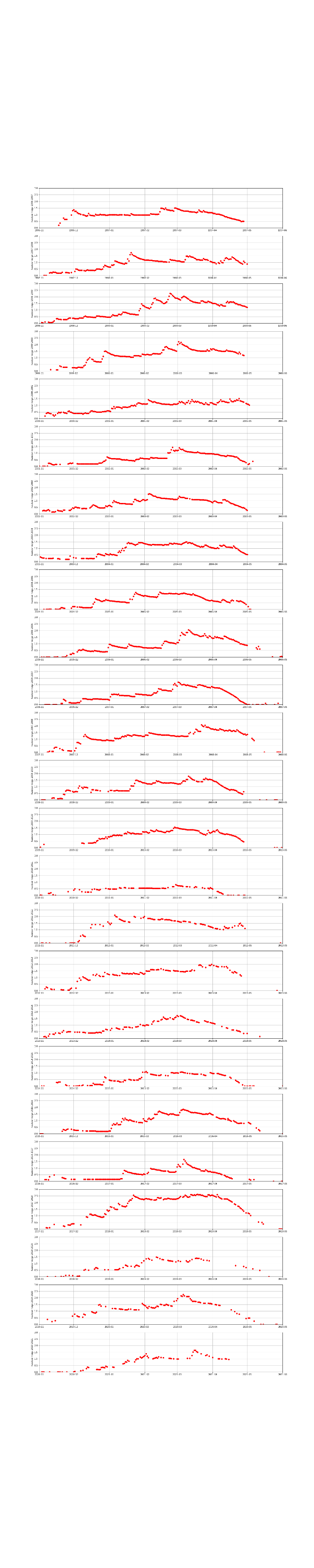
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| French description | English description | Name in the dataset | type | Unity | Completeness  1996-May 2021 |
| Indicatif OMM station |  | Numer\_sta | Char |  |  |
| Date (UTC) | Date UTC | date | Char | YMDHmS |  |
| Altitude de la station haut\_sta réel m | Station altitude | Haut\_sta | Float | Meter |  |
| Direction du vent moyen 10 mn | Wind average direction – last 10 min | dd | Int | degree |  |
| Vitesse du vent moyen 10 mn ff réel m/s | Wind average speed -last 10 min | ff | Float | Meters/sec |  |
| Température | Current temperature | t | Float | Kelvin |  |
| Point de rosée | Dew point | td | Float | Kelvin |  |
| Humidité | Humidity | u | nt | % |  |
| Temps présent ww int code | Present time |  |  |  |  |
| Temps passé 1 w1 int code |  |  |  |  |  |
| Temps passé 2 w2 int code |  |  |  |  |  |
| Nebulosité totale n réel % |  |  |  |  |  |
| Nébulosité des nuages de l'étage |  |  |  |  |  |
| inférieur |  |  |  |  |  |
| nbas int octa |  |  |  |  |  |
| Hauteur de la base des nuages de |  |  |  |  |  |
| l'étage inférieur |  |  |  |  |  |
| bbas int octa |  |  |  |  |  |
| Type des nuages de l'étage inférieur cl int code |  |  |  |  |  |
| Type des nuages de l'étage moyen cm int code |  |  |  |  |  |
| Type des nuages de l'étage supérieur ch int code |  |  |  |  |  |
| Précipitations dans les N dernières |  |  |  |  |  |
| heures |  |  |  |  |  |
| rrN réel mm |  |  |  |  |  |
| Température minimale sur N heures tnN réel K |  |  |  |  |  |
| Température maximale sur N heures txN réel K |  |  |  |  |  |
| Hauteur totale de neige ht\_neige réel m |  |  |  |  |  |
|  |  |  |  |  |  |

1. Temps passé 1 w1 int code
2. Temps passé 2 w2 int code
3. Nebulosité totale n réel %
4. Nébulosité des nuages de l'étage
5. inférieur
6. nbas int octa
7. Hauteur de la base des nuages de
8. l'étage inférieur
9. bbas int octa
10. Type des nuages de l'étage inférieur cl int code
11. Type des nuages de l'étage moyen cm int code
12. Type des nuages de l'étage supérieur ch int code
13. Précipitations dans les N dernières
14. heures
15. rrN réel mm
16. Température minimale sur N heures tnN réel K
17. Température maximale sur N heures txN réel K
18. Hauteur totale de neige ht\_neige réel m
19. Hauteur de la neige fraîche ssfrai réel m
20. Periode de mesure de la neige fraiche perssfrai réel m
21. Phénomène spécial phenspeN réel code
22. Etendue cche nuageuse 1 nnuage1 int code
23. Température de surface de la neige t\_neige réel K
24. Etat de la neige etat\_neige int code
25. Profondeur d'enfoncement de la sonde prof\_sonde int m
26. Nuages dans la vallée nuage\_val int code
27. Chasseneige
28. en altitude chasse\_neige int code
29. Description de l'avalanche observée aval\_descr int code
30. Genre d'avalanche aval\_genre int code
31. Altitude de départ de l'avalanche aval\_depart int code
32. Exposition de l'avalanche aval\_expo int code
33. Estimation du risque d'avalanche aval\_risque int code
34. Direction vent
35. TIME DISTRIBUTION

As we



Number of measures per month from 1996 (bottom) to 2020(top), ordered chronologically. As we can see, every year the most filled months are January, February, and March. December and April are about half filled. January 2008 is the most filled (>6000 measures)



Example of measures of snow height from November, 1st to June, 30th every year, from 2011 to 2020, for quite filled station : La Plagne.

1. DATA STRUCTURE

Of course, among the 647 000 measures and considering all the features, a lot of data is missing. I don’t know how the weather stations work. The stations seem to be composed of thermometers, manometers, other weather instruments, probably computers to get the list of features. Many features are probably not computed because of computing errors, of missing measures. Out of the winter period, some stations are probably disactivated. I first noticed that :

* For
* Measures are mostly concentrated on the winter season. It begins mostly in November or December, and ends in March or April. Some stations keep sending data out of the winter range.

1. DATA PROCESSING AND CLEANING

STEPPING INTO THE MODEL

If we look at the available features :

1. FEATURES TO KEEP

* Snow height : this will be the feature to predict
* Fresh snow height
* Temperature min on the last 24h
* Temperature max on the last 24h
* Altitude of the station
* Mean wind speed

1. FEATURES TO ADD

Some features were not present here, but I tried to add them by myself if it is not too difficult to build :

* Sun ray : using the latitude of the station and the day in the year, I computed a value proportional to the energy received by the station when the Sun is at the highest point in the sky in the day. And considering the value 1 when it is at the zenith, and zero when it is on the horizon.
* Slope of the station : probably useful mostly for March and April. I simply looked at the location of the station on geoportail.gouv website and tried to locate the slope.. 0 for a fully north slope, and 1 for a fully south slope. This is of course approximate

1. STATIONS AND TIME RANGES TO KEEP

As I said previously, the measures are mostly concentrated between December and April. One of my main thoughts was to try to model the period between April and June, the most difficult period to model, when the snow melting depends of the Sun, the rain, the slope, the density.. Unfortunately after mid-April the data is often missing.

The idea was to build shapes of equal number of days, one for each winter season and for each station.

I eventually decided to build a batch that goes from the 15th of December, to the 15th of April. For the bissextile years, it ends on the 14th of April.

1. FILLING HOLES

Difficulty of filling holes depends

* Snow height : probably the most critical feature. That feature was quite good filled. I decided to do an interpolation for the holes
* Fresh snow height : quite a critical information also, as it directly influences the snow heigh. All the snow height rises are only due to the fresh snow.

Snow height[d} = Snow h

* Temperature min on the last 24h
* Temperature max on the last 24h
* Altitude of the station
* Mean wind speed