Défi IA : challenge kick-off

Gwennaëlle LARVOR
07/10/2021

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

METEO FRANCE

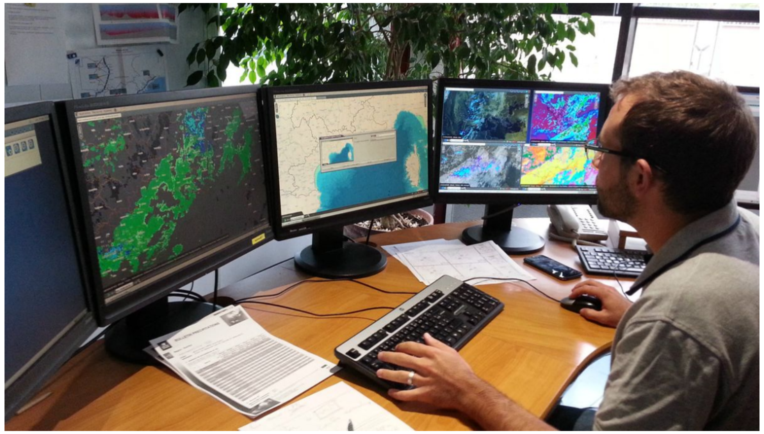
Table of contents

1. Context

- Forecast weather is a really hard task!

A lot of extreme events

A lot of information



Weather forecaster at METEO FRANCE

Can AI help to:

- perform a summary?
- make decisions?

1. Context

- The answer is yes!
  - Today, AI is used at METEO FRANCE to improve weather forecasts
  - So, help us to do even better!

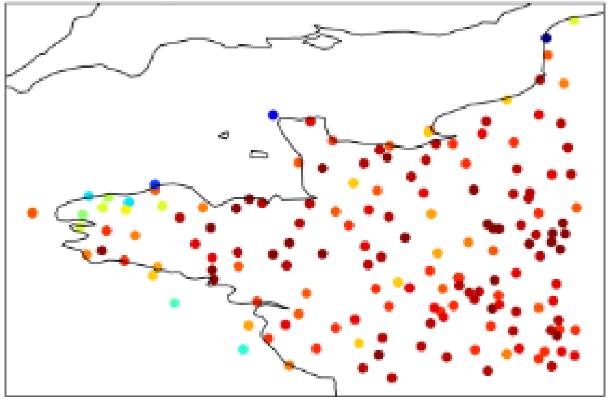
1. Context

- Existing weather forecast challenge
    - made by the french national meteorology school (ENM-> École Nationale de la Météorologie)
    - open to everyone!
    - weather forecasts with ‘classic’ methods (human summary of different data sources)
    - let’s challenge them with our AI algorithms!
    - for more information : http://concours-previ.enm-toulouse.fr/

2. Subject

- Objective : predict the accumulated **daily** rainfall **(on 24h)** on ground stations

Ground station = measurement stations

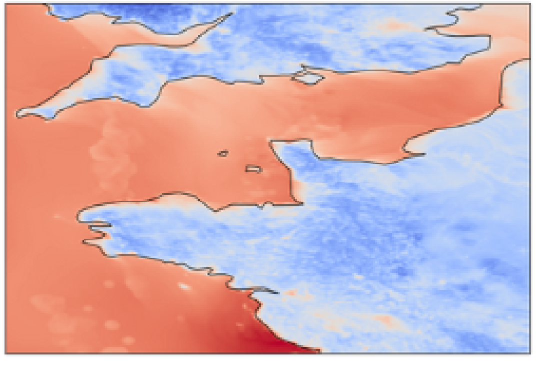


Built throughout the French territory

3. Data sources (1)

- 2 data types
  - **measurement** stations (X_station)
    - rainfall, temperature, wind...etc
    - measurements **at a date t**

  - weather **forecasts** (X_forecast)
    - made by weather forecast systems from METEO FRANCE
    - predict **in the future** the weather



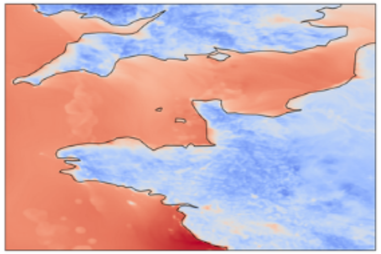
RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

METEO FRANCE

3. Data sources (2)

- Use the **available** data at a time t to predict the rainfall
  - measurements (the day before) : the day D-1
  - weather forecasts : the day D

X_forecast : weather forecasts



X_station : weather measurements

Prediction 0

Prediction 1

... D-1 D D+1 D+2 ...

Time

...

X_station_0

X_forecast_0

Y_0

X_station_1

X_forecast_1

Y_1

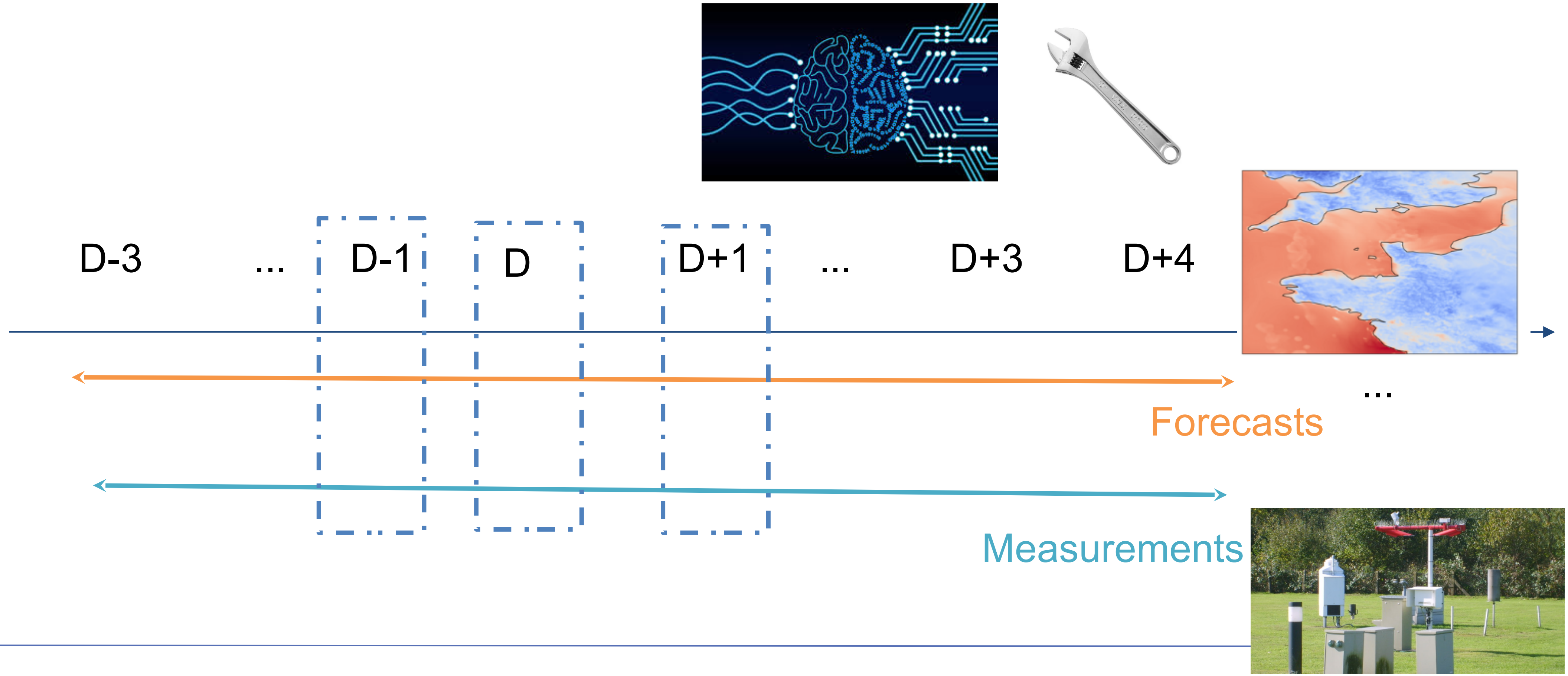
...

Y : measured rainfall

4. Approaches

N°1 : learn from past errors

On the train part -> compute the past errors (forecasts VS measurements) -> algorithms coefficients
ex : overestimate the rainfall or underestimate in such weather scenario



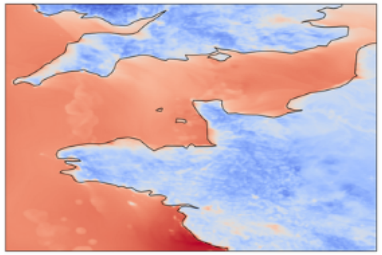
RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

METEO FRANCE

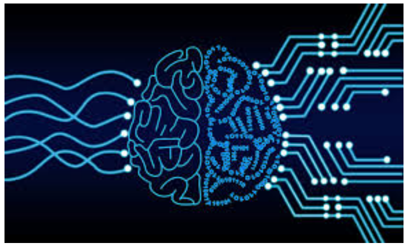
4. Approaches

N°1 : learn from past errors

On the test part -> apply our algorithm to improve the day D forecast



+



->

Prediction 0

... D-1 D D+1 D+2 ...

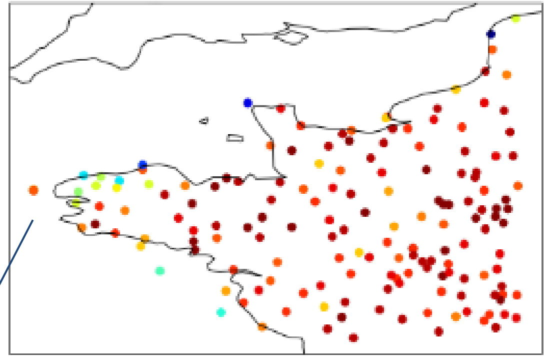
Time

... X_forecast_0 ...

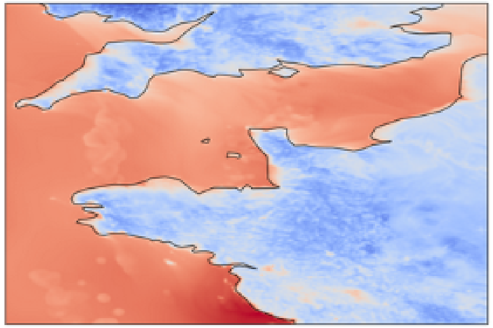
Y_0

4. Approaches

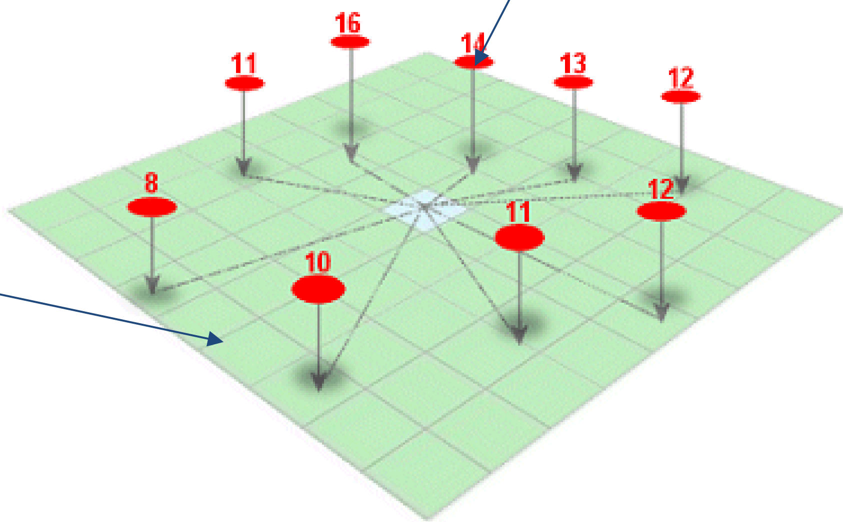
Information repartition in space



Measurements
on punctual locations



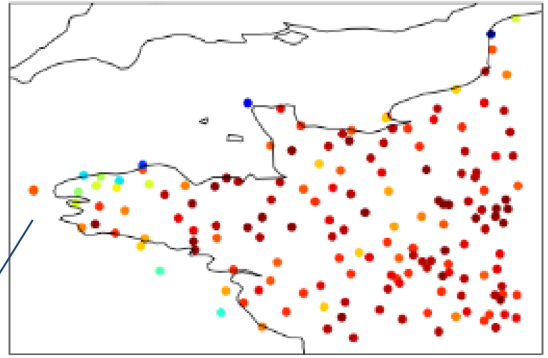
Forecasts
“Regular grid” in latitude/longitude



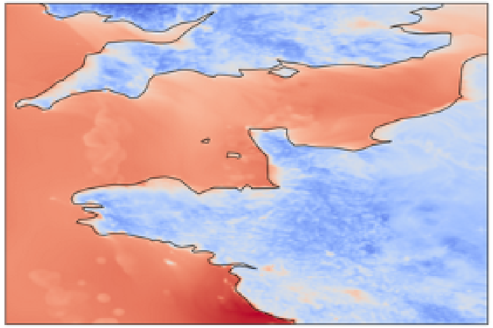
How to compare data with different coordinates (latitude/longitude)? -> **spatial interpolation**

4. Approaches

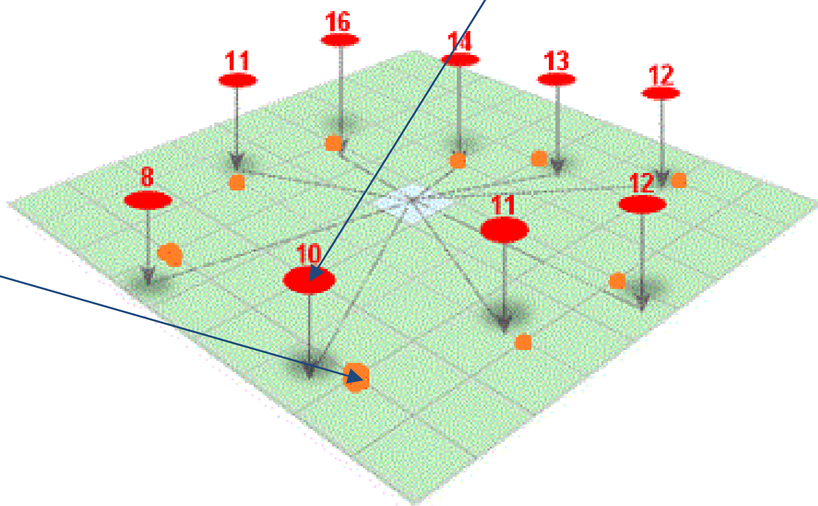
Simple method : nearest neighbor



Measurements
on punctual locations



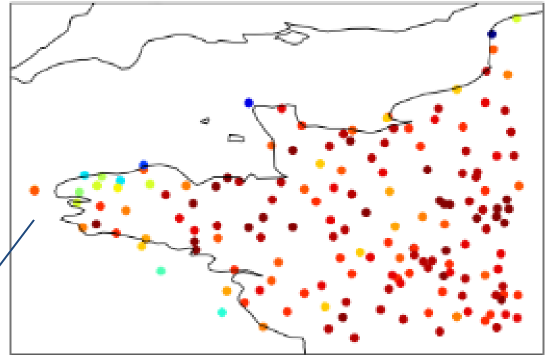
Forecasts
“Regular grid” in latitude/longitude



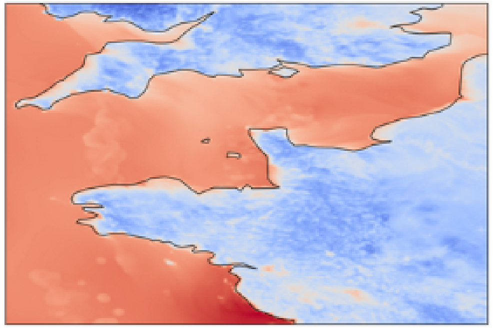
For each station point, choose the nearest forecast point

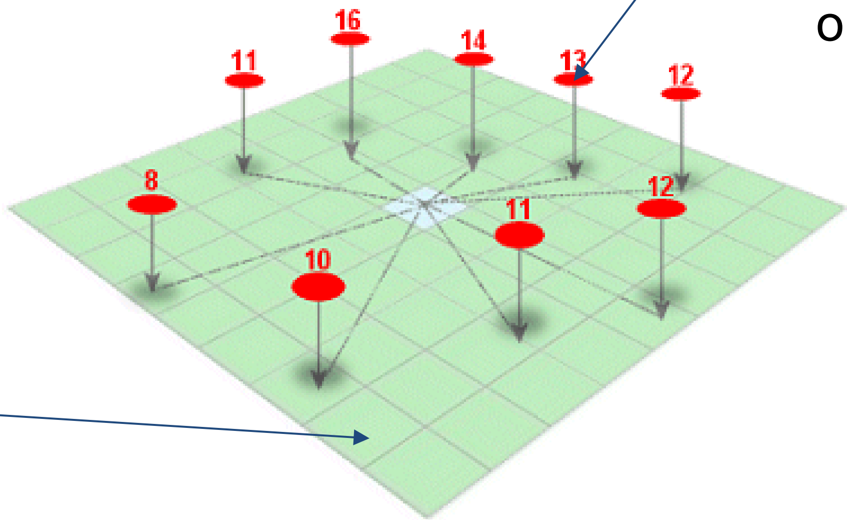
4. Approaches

Other method : use several points (linear, quadratic method...etc)



Measurements
on punctual locations





Forecasts
“Regular grid” in latitude/longitude

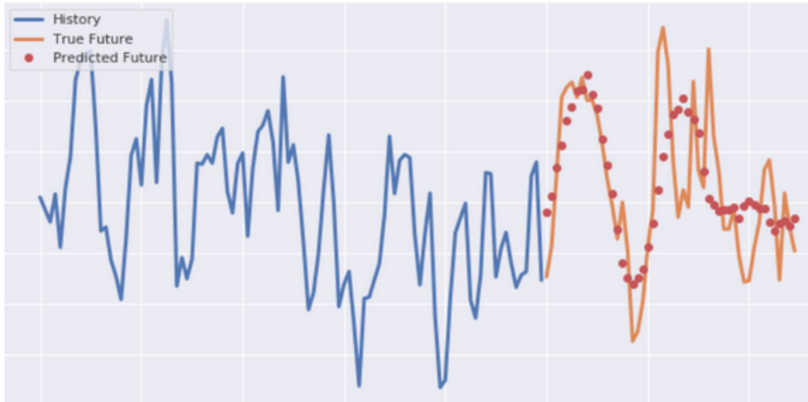
For each station point, use all the points with **weights according to their distances**:

- a lot of importance about nearest points
- less weights about furthest points

4. Approaches

N°2 : predict the future from the past measurements (time series prediction)

For a given station, from the measurements of the day D-1, predict the rainfall on the day D
Existing methods : ARIMA, LSTM...etc



X_station : weather measurements

Prediction 0

... D-1 D D+1 D+2 ...

Time

...

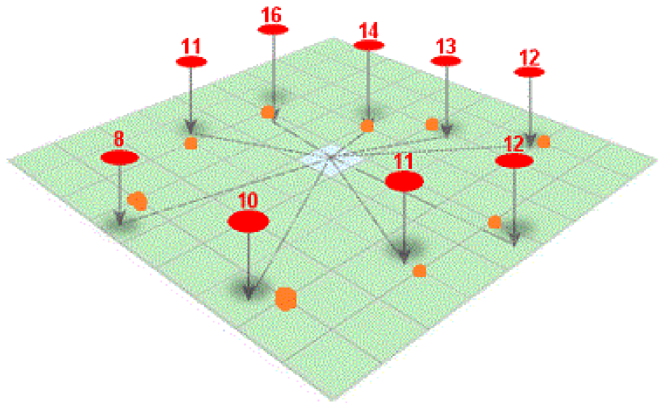
X_station_0

Y_0

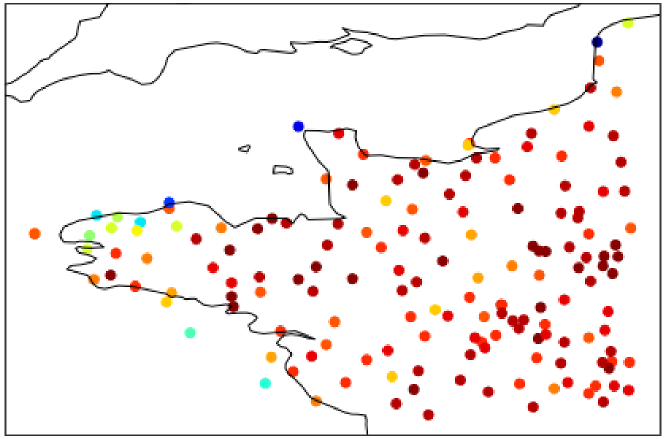
...

5. Baselines

Baselines : references to beat

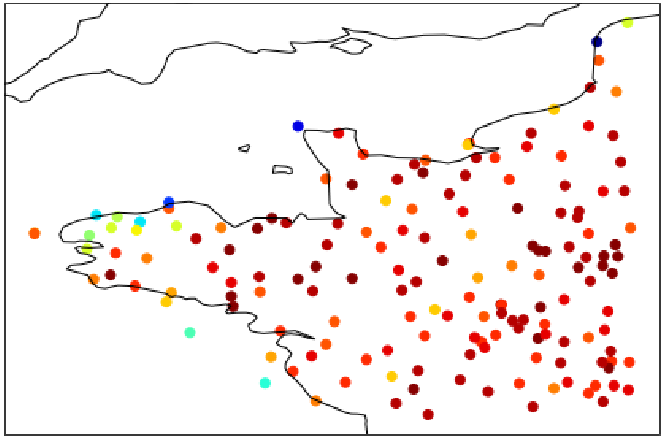


Baseline_observation



The **measured** rainfall the day **D-1**

Baseline_forecast



The **forecasted** rainfall the day **D** by the **nearest** point (for each station)

6. Scoring

- MAPE score (Mean Absolute Percentage Error):
  - n : number of predictions (every date and every station)
  - yi : actual rainfall
  - ŷi : predicted rainfall

$$MAPE = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

- Rules
  - to avoid 0 on denominator
    - add +1 to yi and ŷi before submitting
  - Remove NaNs

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

METEO FRANCE

7. Data structures

Y and baselines

- CSV files
  - date : one per day
  - number_sta : station number
  - Ground_truth (or Prediction) : accumulated rainfall value (in mm)
  - Id : number_sta + index day -> need Xi to predict Yi

date	number_sta	Ground_truth	Id
2Jan16 – 31Dec17	14.1m – 95.7m	0 – 93.5	183747 unique values
2016-01-02	14066001	3.4	14066001_0
2016-01-02	14126001	0.5	14126001_0
2016-01-02	14137001	3.4000000000000004	14137001_0
2016-01-02	14216001	4.0	14216001_0
2016-01-02	14296001	13.3	14296001_0
2016-01-02	14357002	4.800000000000001	14357002_0

7. Data structures

X_station

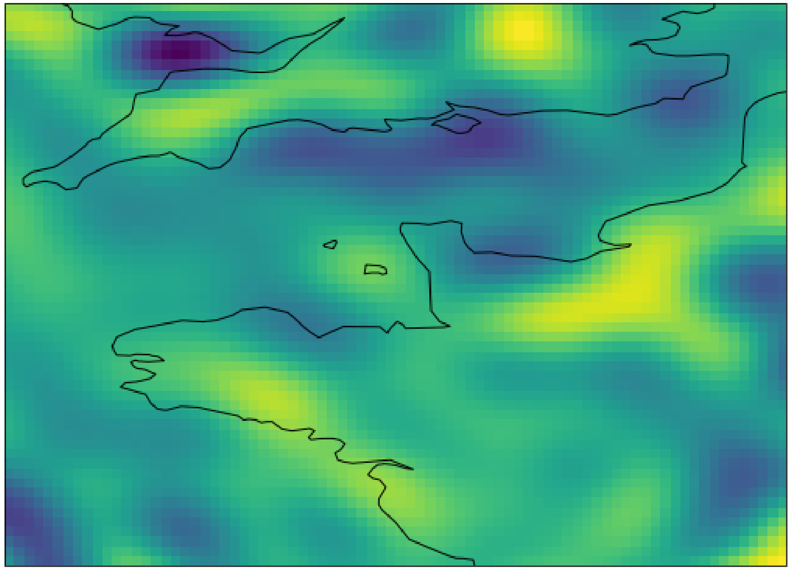
- CSV files
  - date : one per hour
  - number_sta : station number
  - several weather parameters (wind, temperature, humidity, rainfall)
  - Id : number_sta + index day **+ hour** -> need Xi to predict Yi

# number_sta	date	# ff	# precip	Id
14.1m – 95.7m	1Jan16 – 31Dec17	0 – 33.1	0 – 61	4409474 unique values
14066001	2016-01-01 00:00:00	3.05	0.0	14066001_0_0
14066001	2016-01-01 01:00:00	2.5699999999999994	0.0	14066001_0_1
14066001	2016-01-01 02:00:00	2.2600000000000007	0.0	14066001_0_2
14066001	2016-01-01 03:00:00	2.62	0.0	14066001_0_3
14066001	2016-01-01 04:00:00	2.99	0.0	14066001_0_4
14066001	2016-01-01 05:00:00	2.5	0.0	14066001_0_5
14066001	2016-01-01 06:00:00	2.32	0.0	14066001_0_6
14066001	2016-01-01 07:00:00	2.2899999999999996	0.0	14066001_0_7

7. Data structures

X_forecast (1)

- netCDF files, several variables
    - 2D files : latitude, longitude, forecast time
        - 2 forecast systems from METEO FRANCE
            - AROME (finer resolution)
            - ARPEGE
    - 3D files : latitude, longitude, **height**, forecast time
        - ARPEGE only



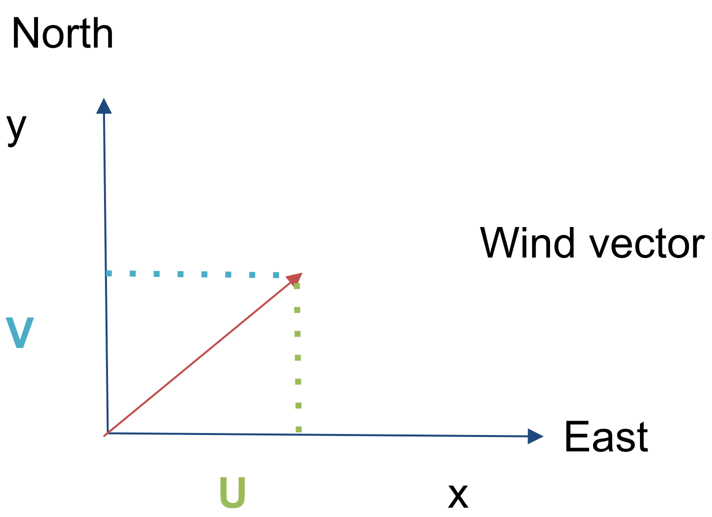
ARPEGE model
14/02/2017 12:00
vertical velocity
vertical level 850hPa

7. Data structures

X_forecast : 2D files

- 2D files : latitude, longitude, forecast time
- weather parameters
  - at 2m above ground : temperature, humidity and **dew point***
  - at 10m above ground : wind speed and direction, **U and V wind components***
  - Mean sea level pressure
  - Precipitation

If temperature=dew point, then there is fog

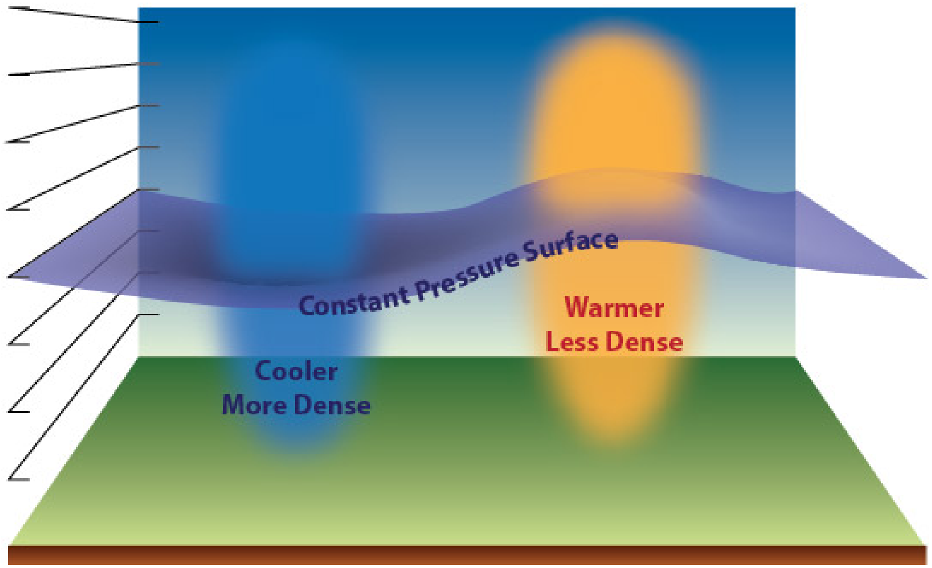


7. Data structures

X_forecast : vertical levels

- netCDF files, several variables
  - 3D files : latitude, longitude, **height**, forecast time
    - ARPEGE only, 2 different height levels
      - height levels (in m) - 20,100,500...3000m
      - isobaric levels (in hPa) - 1000, 950, 925...500 hPa

Air pressure ↘ when the height ↗ (air density ↘)
Pressure levels move according to the weather situation
-> interesting to have information on a given pressure level (and not height level)
-> pressure levels : the most used in meteorology



7. Data structures

X_forecast : file structure

example of one file

```
data
```

xarray.Dataset

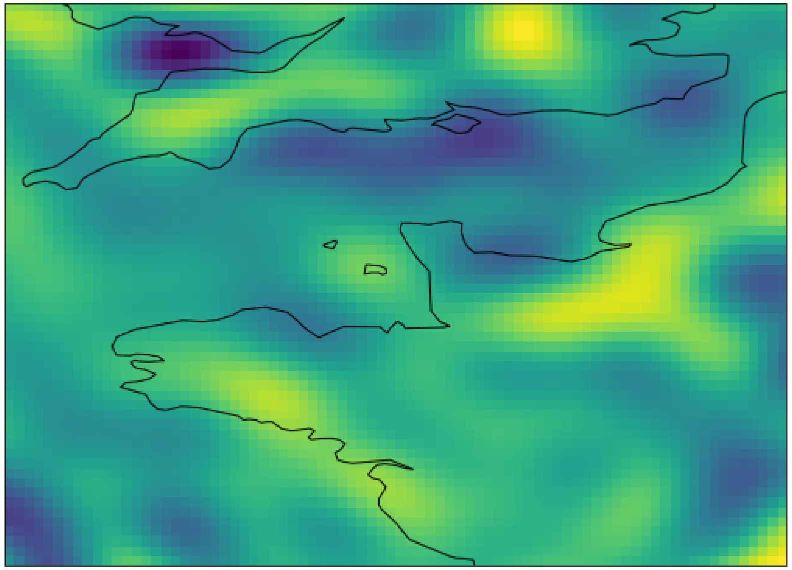
▸ Dimensions: (**isobaricInhPa**: 7, **latitude**: 58, **longitude**: 80, **valid_time**: 17, **Id**: 17)

▾ Coordinates:

isobaricInhPa	(isobaricInhPa)	int16	1000 950...
latitude	(latitude)	float32	51.9 51.8 ...
longitude	(longitude)	float32	-5.842 -5....
valid_time	(valid_time)	datetime64[ns]	2017-02-...
Id	(Id)	object	'409_0' '4...

▾ Data variables:

time	()	datetime64[ns]	...
t	(valid_time, isobaricInhPa, latitude, longitude)	float32	...
p3014	(valid_time, isobaricInhPa, latitude, longitude)	float32	...
r	(valid_time, isobaricInhPa, latitude, longitude)	float32	...
ws	(valid_time, isobaricInhPa, latitude, longitude)	float32	...
p3031	(valid_time, isobaricInhPa, latitude, longitude)	float32	...
u	(valid_time, isobaricInhPa, latitude, longitude)	float32	...

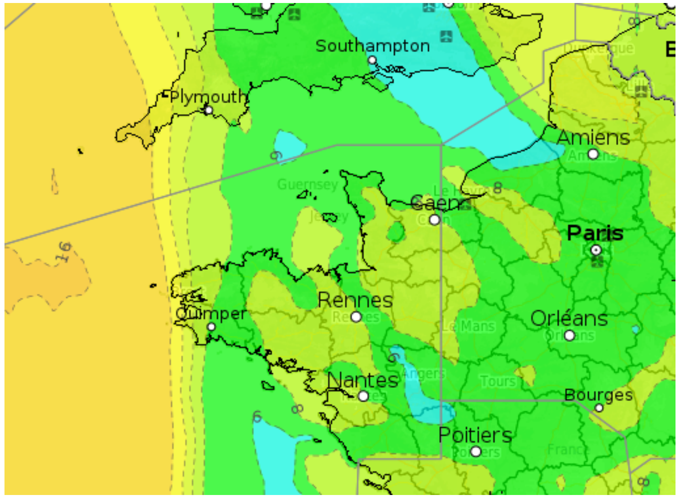


ARPEGE model
14/02/2017 12:00
vertical velocity
vertical level 850hPa

7. Data structures

X_forecast : 3D files (1)

- 3D files : latitude, longitude, forecast time, height
- weather parameters
  - height (in m)
    - pressure
  - isobar levels (in hPa)
    - temperature, **theta’ w* (temperature)**
    - humidity
    - wind speed and direction, U and V wind components, vertical velocity (in Pa.s-1)
    - **geopotential***

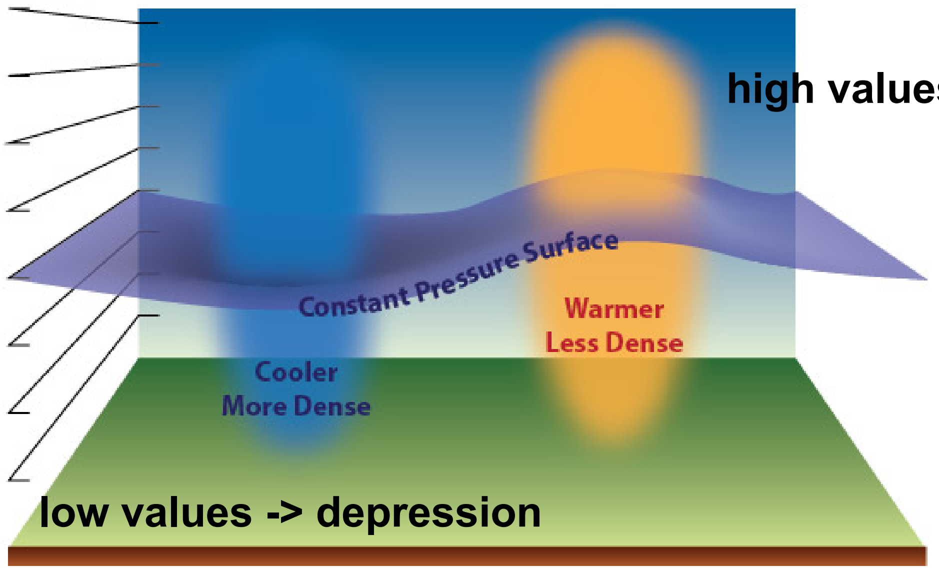


theta’w -> compare temperatures of air samples from different altitudes.
commonly used to track visually the dynamic areas of depressions (especially at 850hPa -> close to ground).

7. Data structures

X_forecast : 3D files (2)

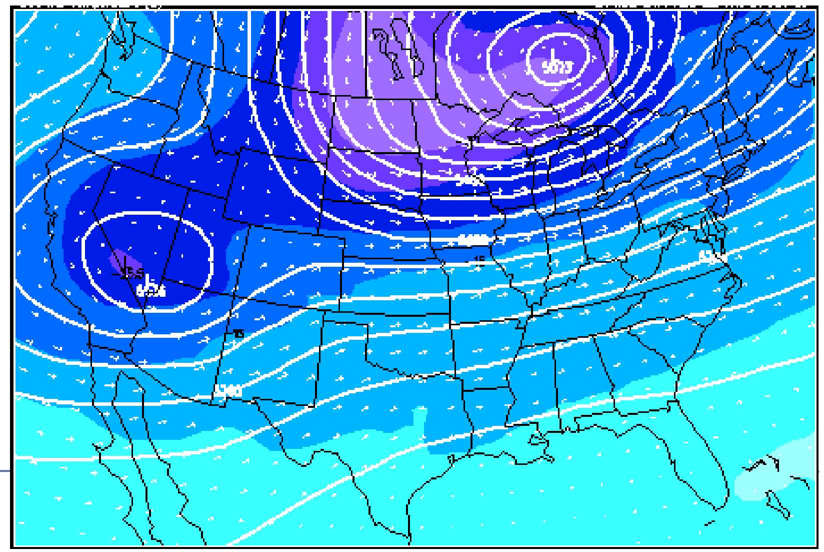
- **geopotential** ~ height of pressure surface above mean sea-level



high values -> anticyclone

low values -> depression

Commonly used map : geopotential and temperature at 500 hPa (track altitude dynamics)



More documentation

More documentation about the data : **https://meteofrance.github.io/meteonet/**

Thank you for your attention!