Computational models for nowcasting

Czibula Istvan-Gergely

WeADL 2021 Workshop

The workshop is organized under the umbrella of WeaMyL, project funded by the EEA and Norway Grants under the RO-NO-2019-0133.
Contract: No 26/2020

Nowcasting

- “description of the current state of the weather in detail and the prediction of changes that can be expected on a timescale of a few hours.” - Keith Browning 1981
- forecasting with local detail, by any method, over a period from the present to six hours ahead, including a detailed description of the present weather - WMO Working Group on Nowcasting Research 2010
- use observations from radar, satellite, lightning networks, surface stations, wind profilers, radiosondes, etc.

Nowcasting

- Weather radar systems are the most important instruments for nowcasting, particularly for convective weather phenomena
- Radar systems directly observe precipitation particles in three dimensions over a large area with an update rate of a few minutes
- The primary goal of nowcasting is to identify and dispatch extreme weather warnings
- Severe weather warning specify the type of event (heavy rain, high winds, hail), time information, geographical location, awareness and trigger levels, severity.
- https://api.met.no/weatherapi/metalerts/1.1?show=all&lang=en

Radar products

Base radar products

- **Base Reflectivity**. A display of echo intensity measured in dBZ
- **Base Velocity**. A measure of the radial component of the wind

Derived radar products

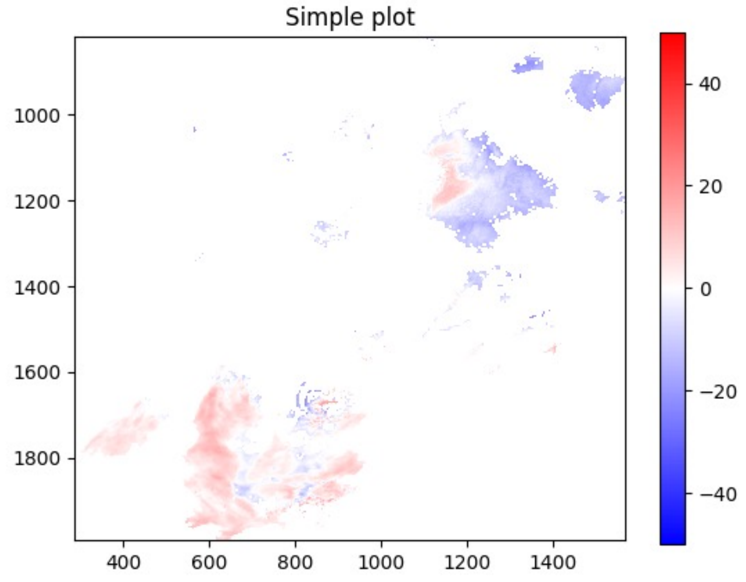
- **Composite Reflectivity**. Maximum reflectivity for the total volume within the range of the radar
- **One-Hour Precipitation**. A display of estimated one-hour precipitation accumulation

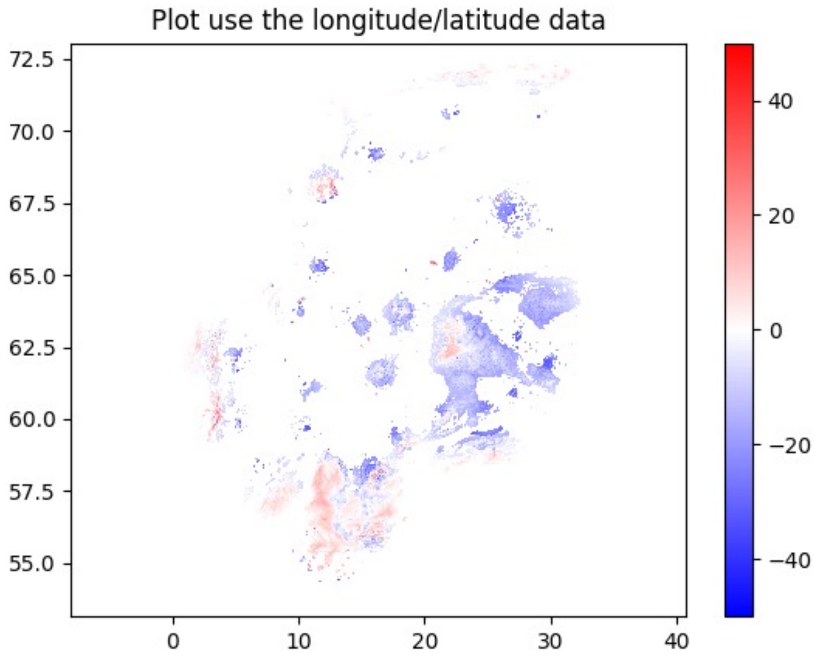
- Forecaster visualize and analyze the data using specialized tools and applications
- integrated display systems that contains observations from the various instruments and sensors on the same display with the same grid spacing for each dataset.

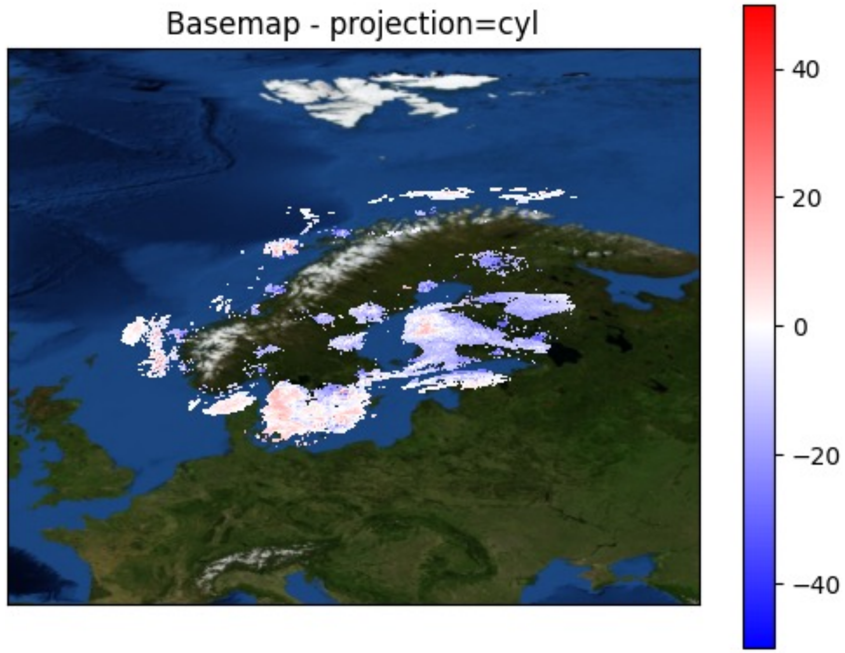
Radar products

- Data is stored in a NetCDF (Network Common Data Form) files, containing:
  - Time dimension: how often the radar is starting a new scan.
  - values: 2d matrix with the values for every acquisition
  - a geographic coordinate system (lon, lat) for each data point
  - Elevations: i.e., 9 slices: 0.5, 1, 2.6, 5.2, 8.6, 13, 18.6, 25.8 and 35 degrees.
  - spatial dimensions: x and y are projected in Azimuth Equidistant
- Other data/file formats are in use (i.e., NEXRAD Level 3, Rainbow5 XML, etc.)
  - contains similar information
- Free open-source libraries are available and can be used to facilitate reading/writing files in various standard data formats

Simple visualizations for radar products







Sample code for handling NetCDF files

Python libraries: netCDF4, pyproj, xarray, numpy, matplotlib, basemap

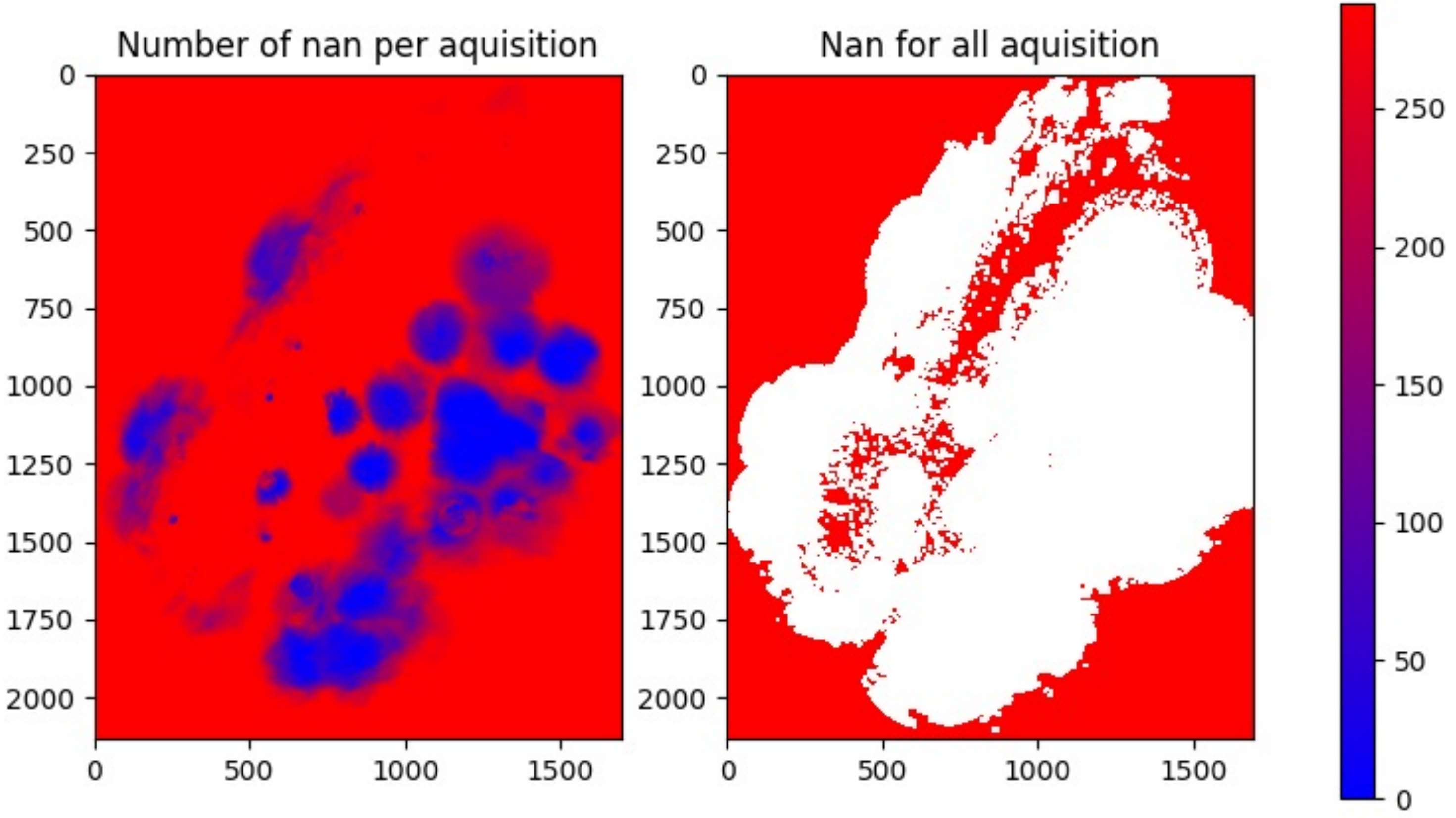
```
import netCDF4
import pyproj
import xarray as xr
def loadDataFromNetCDF(fileName):
    dataset = xr.open_dataset(fileName)
    timeStamp = 10
    #get the data (2D matrix of float values)
    reflectivity = dataset.variables["equivalent_reflectivity_factor"][timeStamp,:,:].data
    #get longitudes/latitudes
    lon, lat = dataset.variables['lon'].data,dataset.variables['lat'].data
```

Sample code for radar product visualization

```
def plotDataMatrix(reflectivityData,cmap):
    """
     Simple plot of the data, every value become a pixel in the image (same position in the image as in the data)
     reflecivityData - 2 d numpy array; cmap - (colormap,vmin, vmax)  colormap used
    """
    plt.figure()
    plt.imshow(reflectivityData,cmap=cmap[0],vmin=cmap[1],vmax=cmap[2])
    plt.colorbar()
    plt.show()

def projectToMap(reflectivityData,netCDFData,cmap):
    """
     Cylindrical projection of the data according to the longitude/latitude data
     reflecivityData - 2 d numpy array; cmap - (colormap,vmin, vmax)  colormap used
    """
    plt.figure()
    m = Basemap(llcrnrlat=43.5,urcrnrlat=82,llcrnrlon=-7,urcrnrlon=41) # centrate map to the region of interest
    X,Y = m(netCDFData.variables['lon'].data,netCDFData.variables['lat'].data) #prepare longitude/latitude data
    m.bluemarble() #pick map type
    plt.pcolormesh(X,Y,reflectivityData,cmap=cmap[0],vmin=cmap[1],vmax=cmap[2],shading='auto')
    plt.colorbar()
    plt.show()
```

Explore the data



Equivalent reflectivity factor – aggregate all day 01.07.2021

Data points per acquisition:
2134*1694=3’614’996

288 acquisition per day

NetCDF file size on disk: 948MB

58% empty(nan/red) all the time

1% nonempty all the time

Nowcasting using deep learning models

- Machine learning algorithms build a model based on training data in order to make predictions.
- Deep learning approaches are known to offer good performance if high volume of training data is available.
- Unlike traditional neural networks, deep networks are scalable, the performance of the model is likely to improve if more data and bigger models are deployed.

Nowcasting using deep learning models

- Based on the meteorologist knowledge of the local climatology and his conceptual model of the evolution for the local area, the nowcaster decide if severe weather is likely.
- The expertise and experience of the nowcaster is essential
- Realistically a deep learning approach should try to assist the human expert, not fully automatize the nowcasting process.
- The deep learning-based approach should be easily integrated into established processes and use together with existing tools
- increased precision, earlier warnings, improved operational workflow

Nowcasting as a deep learning problem

- Given existing measurements (radar, satellite, etc.) predict extreme weather events (time, location, type, severity).
- Advantages
  - Most straightforward formulation of the problem
  - Promise of “automatic nowcasting”
- Disadvantages/ Problems
  - Volume of annotated data – a deep learning approach will require large volumes of labeled training data (radar data – weather warning)

Nowcasting as a deep learning problem

- Given existing measurements (radar, satellite, etc. products) predict future measurements (product values in the future).
- Advantages
  - Plenty of labeled training data (radar product at time T – radar product at T+k)
  - Easy integration with existing tools, practices, processes - the model generate values that can be used just as we use the actual measurements.
  - If accuracy is good – earlier weather warnings, allow the use of more sophisticated and demanding analysis of the measurements
- Disadvantages
  - weather warnings not generated automatically - same process as before involving human expert

Norway grants

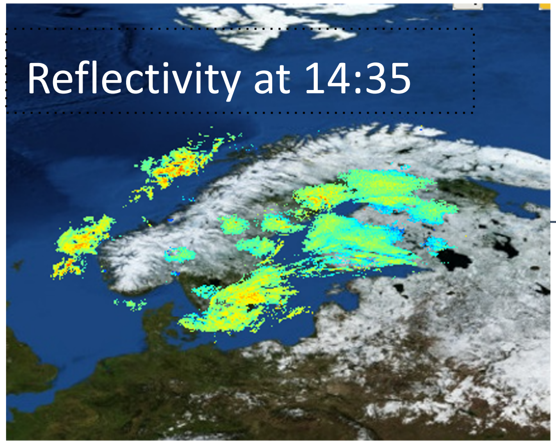
uefiscdi Unitatea Executivă pentru Finanțarea Învățământului Superior, a Cercetării, Dezvoltării și Inovării

WeaMyL

METEO ROMANIA

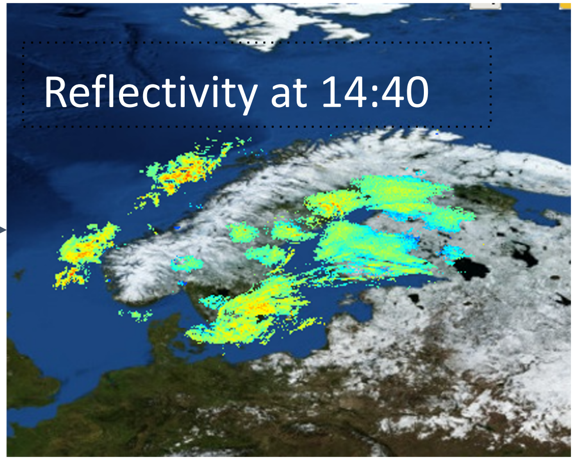
Norwegian Meteorological Institute

One-to-one approach, single time-step prediction



Machine Learning Model

Predict

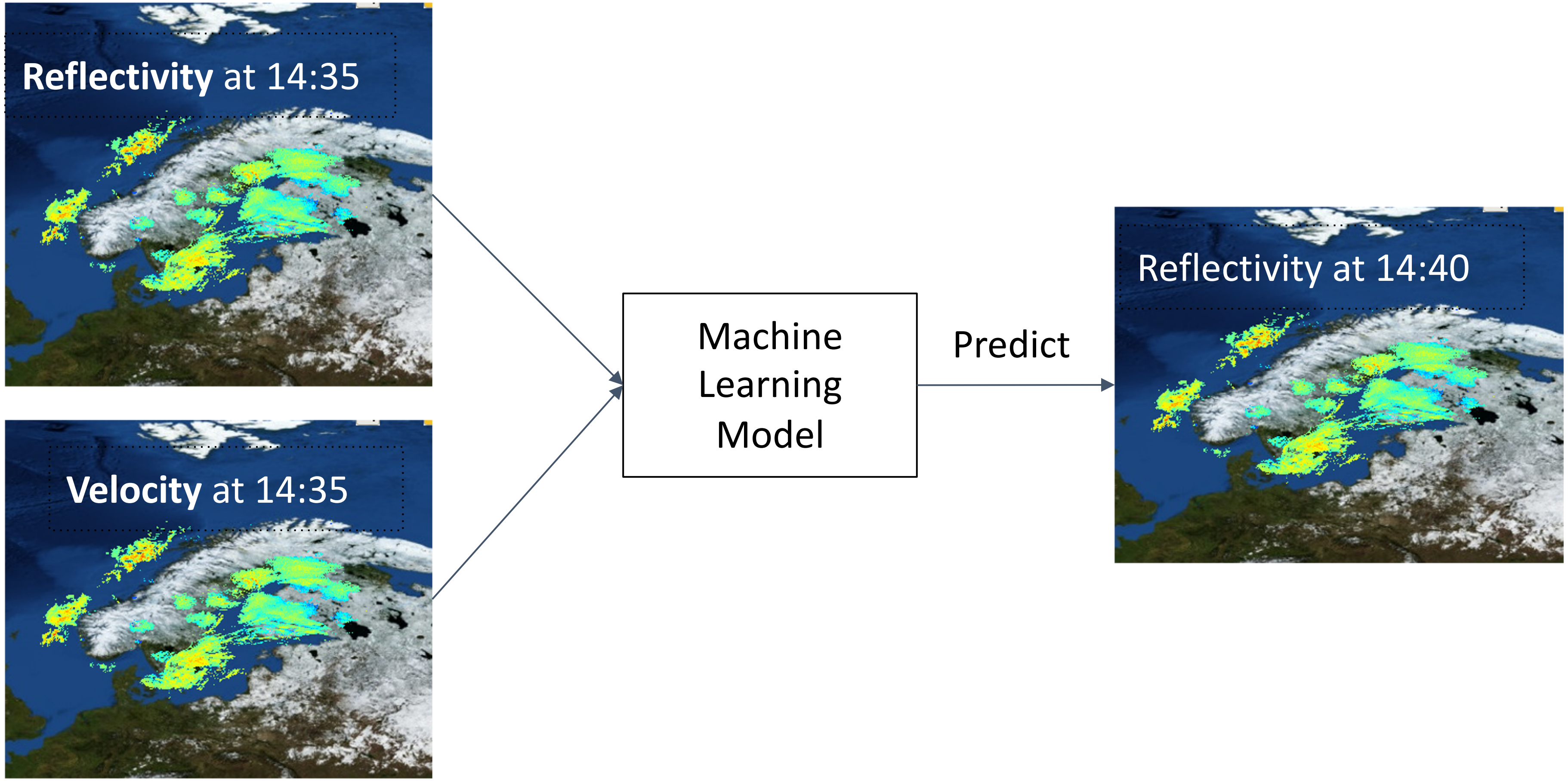


Input: **Single product** for the **entire region** at a given time **T**. Predict all values for the entire region at time **T+1** (next acquisition time)

Product1 time T → ML Model $\xrightarrow{\text{Predict}}$ Product1 time T+1

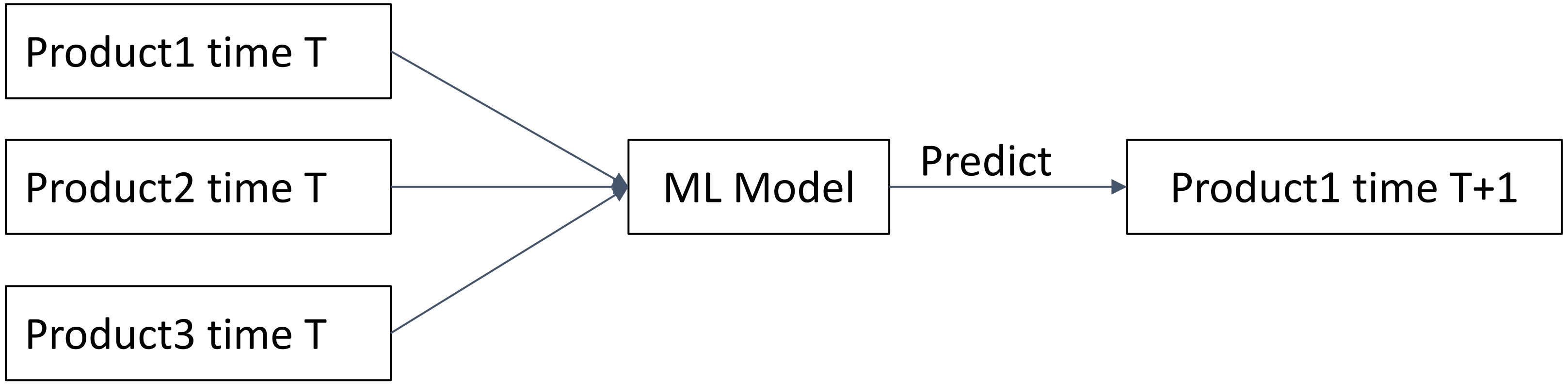
Separate model for every product we want to predict
If we need values at T+2,T+3,... need to feed values from T+1,T+2… (use predicted values for next prediction)
For prediction for 1 hour in the future we can train the model using input at T and expected output T+12

Multiple-to-one approach, single time-step prediction



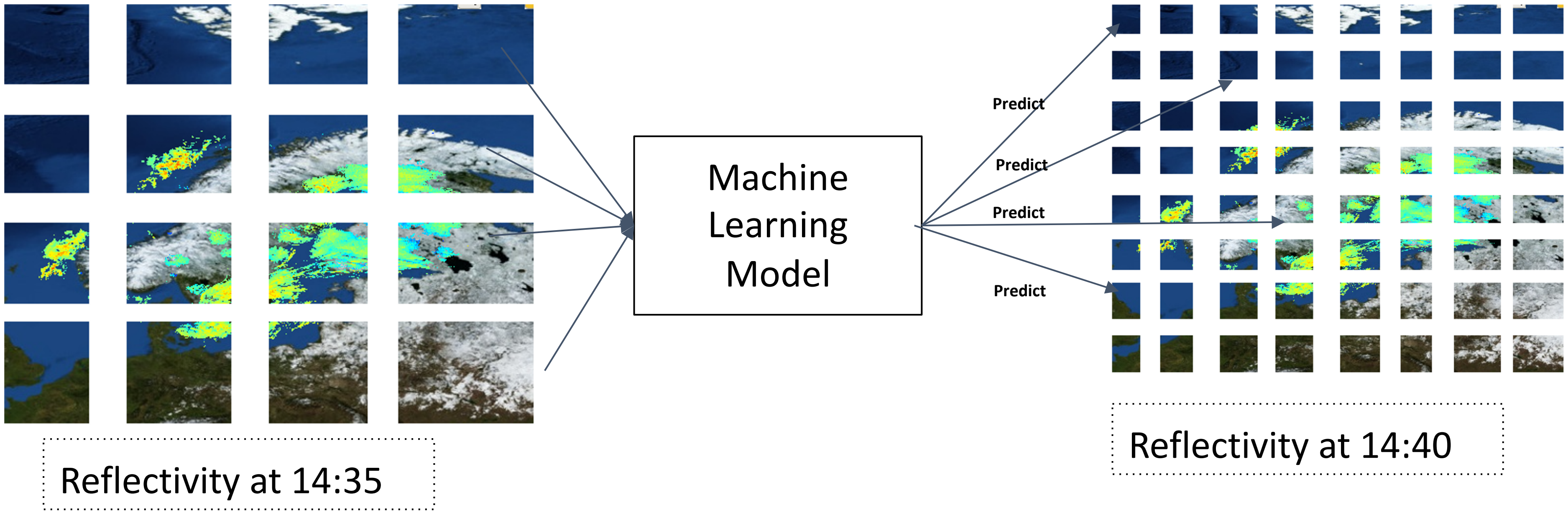
Norway grants

Input: **Multiple products** (radar, satellite, maybe combination) for **the entire region** at a given time **T**.
Predict all values for the **entire region** at time **T+1** (next acquisition time)

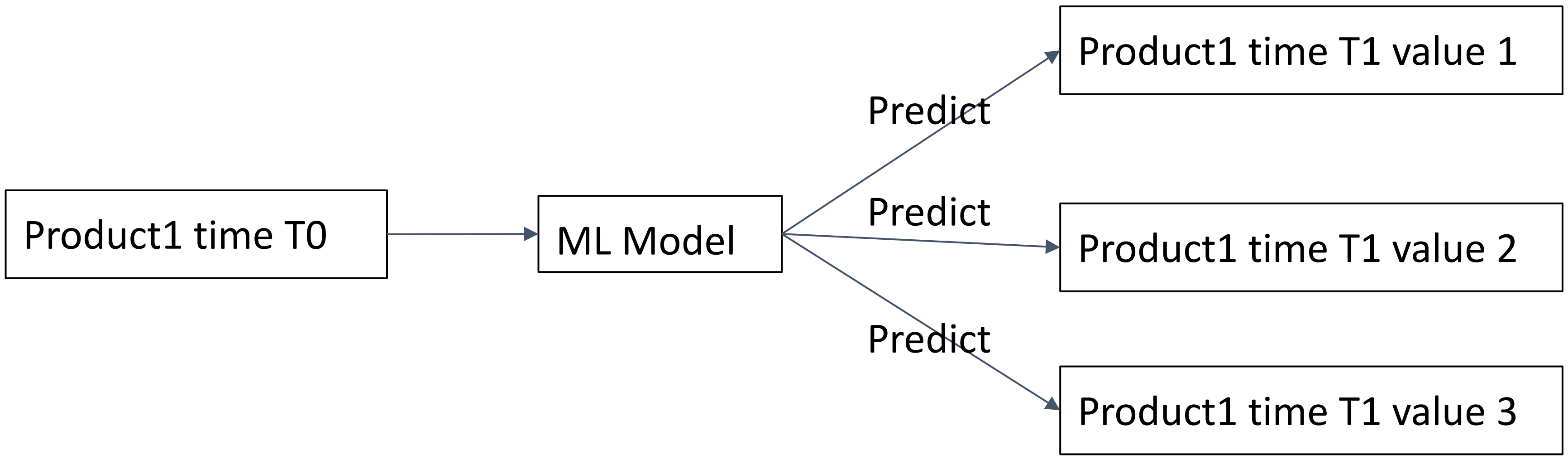


Separate model for every product we want to predict
If we need values at T+2,T+3,... need to feed values from T+1,T+2… (use predicted values for next prediction)
For prediction for 1 hour in the future we can train the model using input at T and expected output T+12
Best combination of input products need to be experimentally determined

Mosaic approach, single time-step prediction

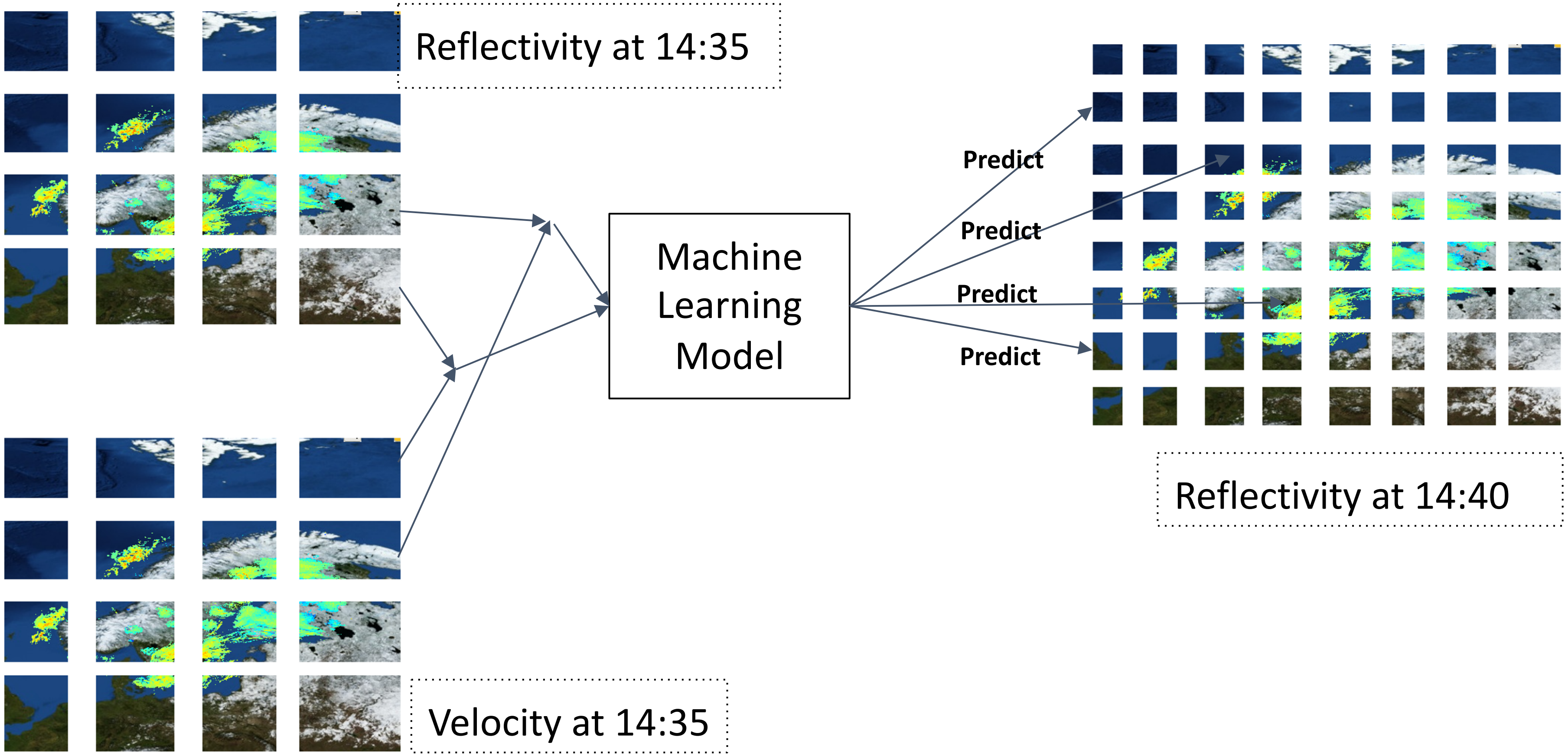


Input: **Single product** for **a small region** at a given time **T0**.
Predict a single value **(maybe a smaller region)** at time **T1** (next acquisition time)

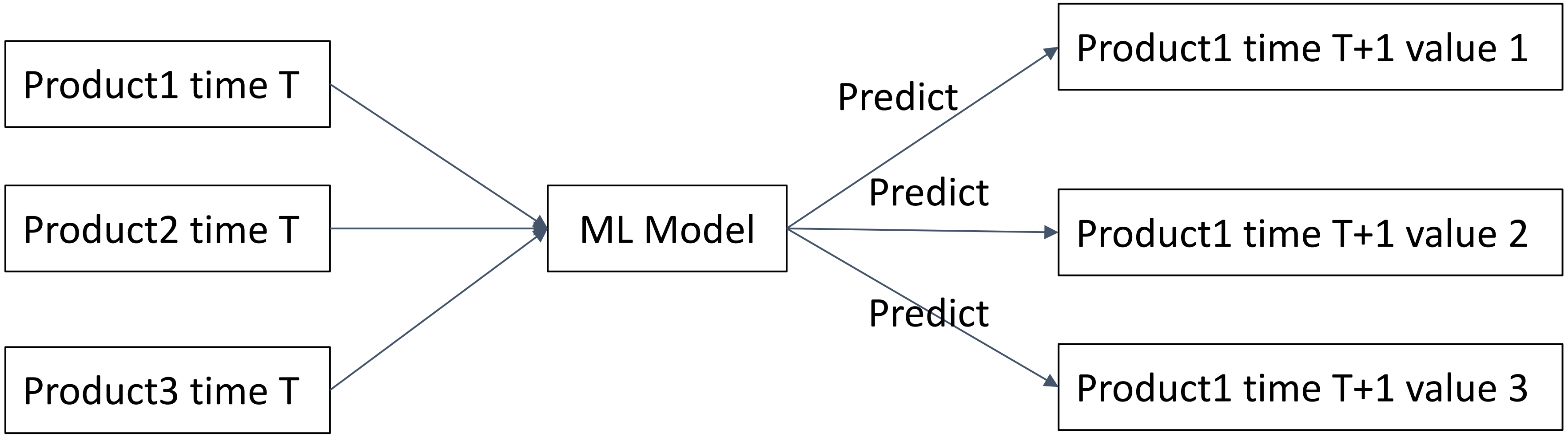


Separate model for every product we want to predict
In order to predict values for the entire region the model will be used multiple times with small regions (area separated into tiles) from the input.
If we need values at T2,T3,... need to feed values from T1,T2... (use predicted values for next prediction)
For prediction for 1 hour in the future we can train the model using input at T and expected output T+12

Mosaic multiple, single time-step prediction



Input: **Multiple products** for **a small region** at a given time **T**.
Predict a single value **(maybe a smaller region)** at time **T+1** (next acquisition time)



Separate model for every product we want to predict
In order to predict values for the entire region the model will be feeded multiple times with small regions from the input.
If we need values at T+2,T+3,... need to feed values from T+1,T+2... (use predicted values for next prediction)
For prediction for 1 hour in the future we can train the model using input at T and expected output T+12
Best combination of input products need to be experimentally determined

Other variations

Multiple variations by mixing around the following dimensions:

- Input/output variation by **products** – single, multiple, base, derived
- Input/output variation by **time steps** – single, multiple, distant time steps
- Input/output variation by **region** – entire region, tiled region, individual values
- Machine Learning **Model** - Convolutional Neural Networks (CNNs),Long Short-Term Memory Networks (LSTMs), Recurrent Neural Networks (RNNs), Autoencoders, etc.
- Handling **missing/noisy data** - ignore, replace, interpolate
- Input **data selection/transformation** - raw, cleaned, discretized, etc.
- Evaluation **measures** / model optimization - POD, FAR, CSI, Custom loss function

Norway grants

Unitatea Executivă pentru Finanțarea Învățământului Superior, a Cercetării, Dezvoltării și Inovării

WeaMyL

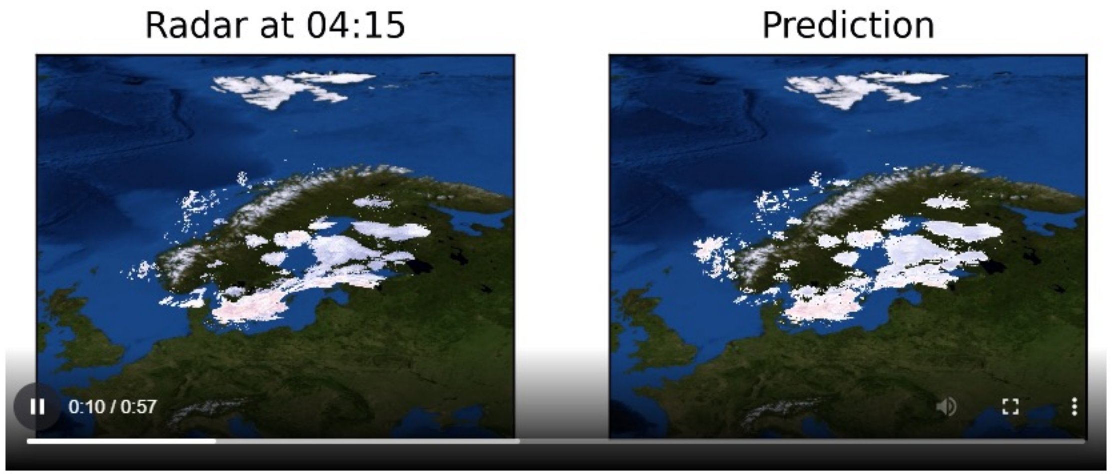
METEO ROMANIA

Norwegian Meteorological Institute

NowcastX version 0.0.1

NowcastX version 0.0.1

NowCastX version 0.0.1 - Incipient Xception-based model trained on a dataset containing 6 days with meteorological events (thredds), selected from the CAP warnings available at weatherapi. The model has been trained on a region of approximately 300km x 300km surrounding Oslo for predicting the composite reflectivity values at 5 minutes in the future using the values from the current timestamp. The day used for illustrating the predictions has not been used for training.



https://weamyl.met.no/models/