



UNIVERSITY of OULU
OULUN YLIOPISTO

COMBINING VARIOUS DATA SOURCES

Case: The weather and net blotch

INTRODUCTION

- Combining the open and closed data
 - Challenges
- Case from agriculture
 - Correlation between the weather conditions and existing of net blotch, one plant disease (barley)
 - Data collection and analysis
 - Results
 - Potential usage of the results



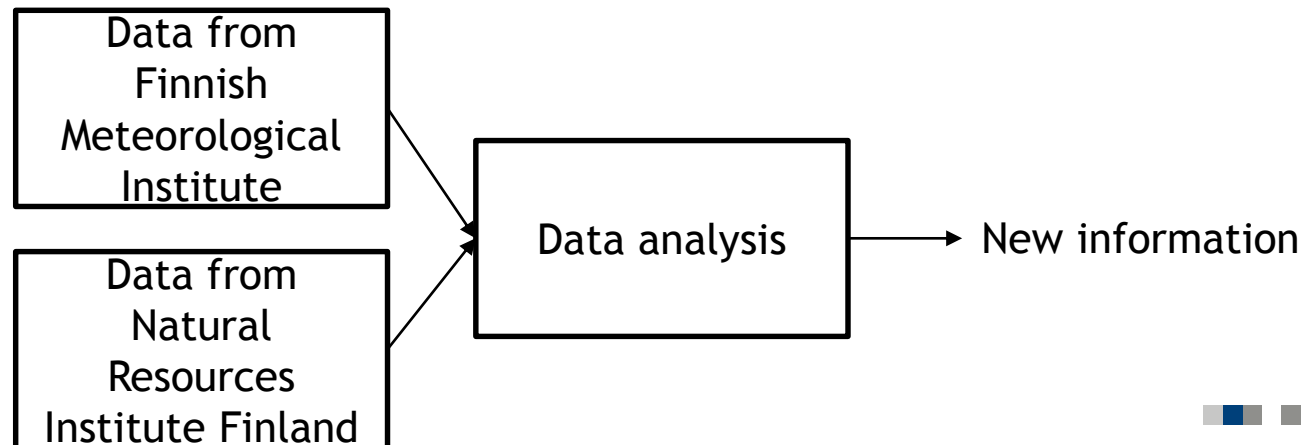
DATA, WHERE?

- A lot of data collected every day
 - Weather conditions
 - Traffic
 - Consumer habits
 - Accidents
 - ...
 - Measurements from industry, agriculture, health care and so on
- Availability of data differs
 - Open and close data



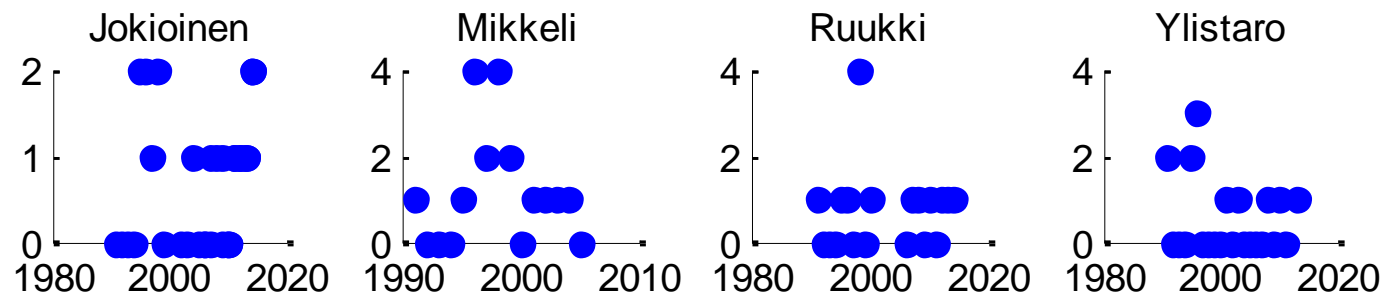
CASE; NET BLOTCH AND WEATHER

- Aim to find out weather conditions which predicts the appearance of the net blotch
- Finnish Meteorological Institute, open data
 - Data collection
- Natural Resources Institute Finland, measurements, close data
 - Data ready to use in xls-format



NATURAL RESOURCES INSTITUTE FINLAND, NET BLOTCH DATA

- Measurements from different places in different years
- Information:
 - No net blotch = 0
 - A lot of net blotch = 5
 - The rest = something between



FMI, OPEN DATA

- Data available for everyone
- Requires registration
- User can select the measurement station and listed variables

- Some challenges with usability
 - Queries in the internet, data in xml-format
 - Data includes mixed up letters and numbers
 - Time limit: Data can be loaded for only one week maximum
 - Vs. the earliest net blotch data is from 1991

- Maybe easier to use with R-code?



EXAMPLE, FMI DATA, ONE QUERY

```

- <target:LocationCollection gml:id="sampled-target-1-1-42m">
  - <target:member>
    - <target:Location gml:id="obsloc-fmiid-101104-pos-42m">
      <gml:identifier codeSpace="http://urn:fmf:fi:namespace:stationcode:fmiid">101104</gml:identifier>
      <gml:name codeSpace="http://urn:fmf:fi:namespace:locationcode:name">Jokioinen Jokioisten observatorio</gml:name>
      <gml:name codeSpace="http://urn:fmf:fi:namespace:locationcode:geoid">-16000162</gml:name>
      <gml:name codeSpace="http://urn:fmf:fi:namespace:locationcode:wmo">2963</gml:name>
      <target:representativePoint xlink:href="#point-101104"/>
      <target:region codeSpace="http://urn:fmf:fi:namespace:location:region">Jokioinen</target:region>
    </target:Location>
  </target:member>
</target:LocationCollection>
</sams:sampledFeature>
- <sams:skape>
  - <gml:Point gml:id="point-1-1-42m" srsName="http://www.opengis.net/def/crs/EPSG/0/4258" srsDimension="2">
    <gml:name>Jokioinen Jokioisten observatorio</gml:name>
    <gml:pos>60.81401 23.49761</gml:pos>
  </gml:Point>
</sams:skape>
</sams:SF_SpatialSamplingFeature>
</om:featureOfInterest>
- <om:result>
  - <wml2:MeasurementTimeseries gml:id="obs-obs-1-1-42m">
    - <wml2:point>
      - <wml2:MeasurementTVP>
        <wml2:time>2014-01-01T00:00:00Z</wml2:time>
        <wml2:value>4.4</wml2:value>
      </wml2:MeasurementTVP>
    </wml2:point>
  </wml2:point>
  - <wml2:point>
    - <wml2:MeasurementTVP>
      <wml2:time>2014-01-01T12:00:00Z</wml2:time>

```



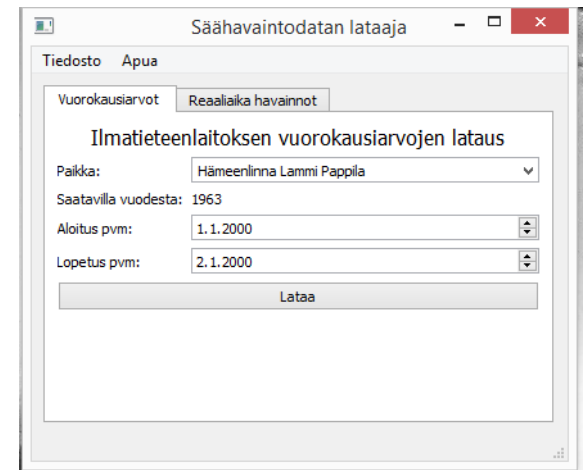
EXAMPLE, FMI DATA OPENED IN EXCEL

	A	B	C	D	E	F	G	H
1	timeStamp	numberMatched	numberReturned	ns8:id	ns3:href	ns8:id2	ns8:beginPosition	ns8:endPosition
2	2015-08-26T11:56:02Z	13	13	WFS-g5TtDp2VzpSlid.mUifWLTN.dv_SJTowqYWbbpdOt.Lnl5dsPTTv3c3Trvlw9NGXk6ddNO3L2w7OuXhh08oV	time1-1-1		2013-07-11T00:00:00Z	2013-07-17T00:00:00Z
3	2015-08-26T11:56:02Z	13	13	WFS-g5TtDp2VzpSlid.mUifWLTN.dv_SJTowqYWbbpdOt.Lnl5dsPTTv3c3Trvlw9NGXk6ddNO3L2w7OuXhh08oV	time1-1-1		2013-07-11T00:00:00Z	2013-07-17T00:00:00Z
4	2015-08-26T11:56:02Z	13	13	WFS-g5TtDp2VzpSlid.mUifWLTN.dv_SJTowqYWbbpdOt.Lnl5dsPTTv3c3Trvlw9NGXk6ddNO3L2w7OuXhh08oV	time1-1-1		2013-07-11T00:00:00Z	2013-07-17T00:00:00Z
5	2015-08-26T11:56:02Z	13	13	WFS-g5TtDp2VzpSlid.mUifWLTN.dv_SJTowqYWbbpdOt.Lnl5dsPTTv3c3Trvlw9NGXk6ddNO3L2w7OuXhh08oV	time1-1-1		2013-07-11T00:00:00Z	2013-07-17T00:00:00Z
6	2015-08-26T11:56:02Z	13	13	WFS-g5TtDp2VzpSlid.mUifWLTN.dv_SJTowqYWbbpdOt.Lnl5dsPTTv3c3Trvlw9NGXk6ddNO3L2w7OuXhh08oV	time1-1-1		2013-07-11T00:00:00Z	2013-07-17T00:00:00Z
7	2015-08-26T11:56:02Z	13	13	WFS-g5TtDp2VzpSlid.mUifWLTN.dv_SJTowqYWbbpdOt.Lnl5dsPTTv3c3Trvlw9NGXk6ddNO3L2w7OuXhh08oV	time1-1-1		2013-07-11T00:00:00Z	2013-07-17T00:00:00Z
8	2015-08-26T11:56:02Z	13	13	WFS-g5TtDp2VzpSlid.mUifWLTN.dv_SJTowqYWbbpdOt.Lnl5dsPTTv3c3Trvlw9NGXk6ddNO3L2w7OuXhh08oV	time1-1-1		2013-07-11T00:00:00Z	2013-07-17T00:00:00Z
9	2015-08-26T11:56:02Z	13	13	WFS-g5TtDp2VzpSlid.mUifWLTN.dv_SJTowqYWbbpdOt.Lnl5dsPTTv3c3Trvlw9NGXk6ddNO3L2w7OuXhh08oV	time1-1-1		2013-07-11T00:00:00Z	2013-07-17T00:00:00Z
10	2015-08-26T11:56:02Z	13	13	WFS-g5TtDp2VzpSlid.mUifWLTN.dv_SJTowqYWbbpdOt.Lnl5dsPTTv3c3Trvlw9NGXk6ddNO3L2w7OuXhh08oV	time1-1-1		2013-07-11T00:00:00Z	2013-07-17T00:00:00Z
11	2015-08-26T11:56:02Z	13	13	WFS-g5TtDp2VzpSlid.mUifWLTN.dv_SJTowqYWbbpdOt.Lnl5dsPTTv3c3Trvlw9NGXk6ddNO3L2w7OuXhh08oV	time1-1-1		2013-07-11T00:00:00Z	2013-07-17T00:00:00Z
12	2015-08-26T11:56:02Z	13	13	WFS-g5TtDp2VzpSlid.mUifWLTN.dv_SJTowqYWbbpdOt.Lnl5dsPTTv3c3Trvlw9NGXk6ddNO3L2w7OuXhh08oV	time1-1-1		2013-07-11T00:00:00Z	2013-07-17T00:00:00Z
13	2015-08-26T11:56:02Z	13	13	WFS-g5TtDp2VzpSlid.mUifWLTN.dv_SJTowqYWbbpdOt.Lnl5dsPTTv3c3Trvlw9NGXk6ddNO3L2w7OuXhh08oV	time1-1-1		2013-07-11T00:00:00Z	2013-07-17T00:00:00Z
14	2015-08-26T11:56:02Z	13	13	WFS-g5TtDp2VzpSlid.mUifWLTN.dv_SJTowqYWbbpdOt.Lnl5dsPTTv3c3Trvlw9NGXk6ddNO3L2w7OuXhh08oV	time1-1-1		2013-07-11T00:00:00Z	2013-07-17T00:00:00Z
15	2015-08-26T11:56:02Z	13	13	WFS-g5TtDp2VzpSlid.mUifWLTN.dv_SJTowqYWbbpdOt.Lnl5dsPTTv3c3Trvlw9NGXk6ddNO3L2w7OuXhh08oV	time1-1-1		2013-07-11T00:00:00Z	2013-07-17T00:00:00Z
16	2015-08-26T11:56:02Z	13	13	WFS-g5TtDp2VzpSlid.mUifWLTN.dv_SJTowqYWbbpdOt.Lnl5dsPTTv3c3Trvlw9NGXk6ddNO3L2w7OuXhh08oV	time1-1-1		2013-07-11T00:00:00Z	2013-07-17T00:00:00Z
17	2015-08-26T11:56:02Z	13	13	WFS-g5TtDp2VzpSlid.mUifWLTN.dv_SJTowqYWbbpdOt.Lnl5dsPTTv3c3Trvlw9NGXk6ddNO3L2w7OuXhh08oV	time1-1-1		2013-07-11T00:00:00Z	2013-07-17T00:00:00Z
18	2015-08-26T11:56:02Z	13	13	WFS-OL_a2SFMPjBQy8ziGD2GN.dBNwSJTowqYWbbpdOt.Lnl5dsPTTv3c3Trvlw9NGXk6ddNO3L2#time1-1-1				
19	2015-08-26T11:56:02Z	13	13	WFS-OL_a2SFMPjBQy8ziGD2GN.dBNwSJTowqYWbbpdOt.Lnl5dsPTTv3c3Trvlw9NGXk6ddNO3L2#time1-1-1				



FMI DOWNLOADER

- Simple user interface for downloading FMI data
- Available from internet:
 - <http://tumetsu.github.io/FMI-weather-downloader>
- Developed by Tuomas Salmi
 - Salmi.tuomas@gmail.com
- Fast and easy to use
- Data in csv-format => usable
- Can't load every variables
 - A place and date, rainfall per day, temperature; minimum, maximum and average per day



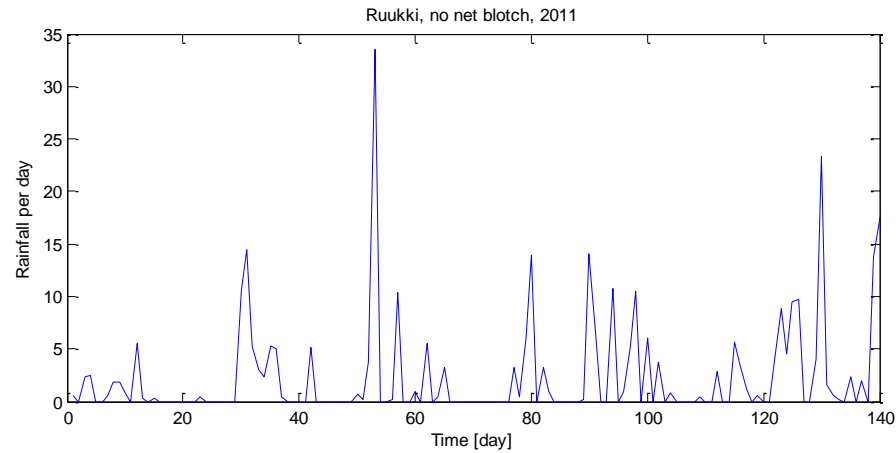
DATA ANALYSIS

- The weather data were grouped according the place and net blotch
 - Ruukki:
 - First group - no net blotch
 - Second group - a small amount of net blotch
 - Third group - a big amount of net blotch
- Data timing according to growing season => The first point of the weather data is the beginning of the growing season - in every data set
- Analysis of the variables from the weather data: What is different between the years grouped according to net blotch

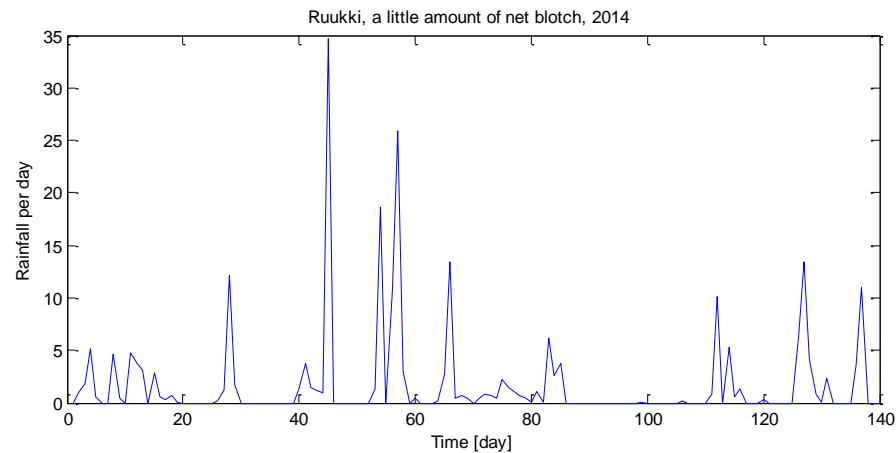


DATA ANALYSIS

- Example, Ruukki
- Rainfall per day
- No net blotch

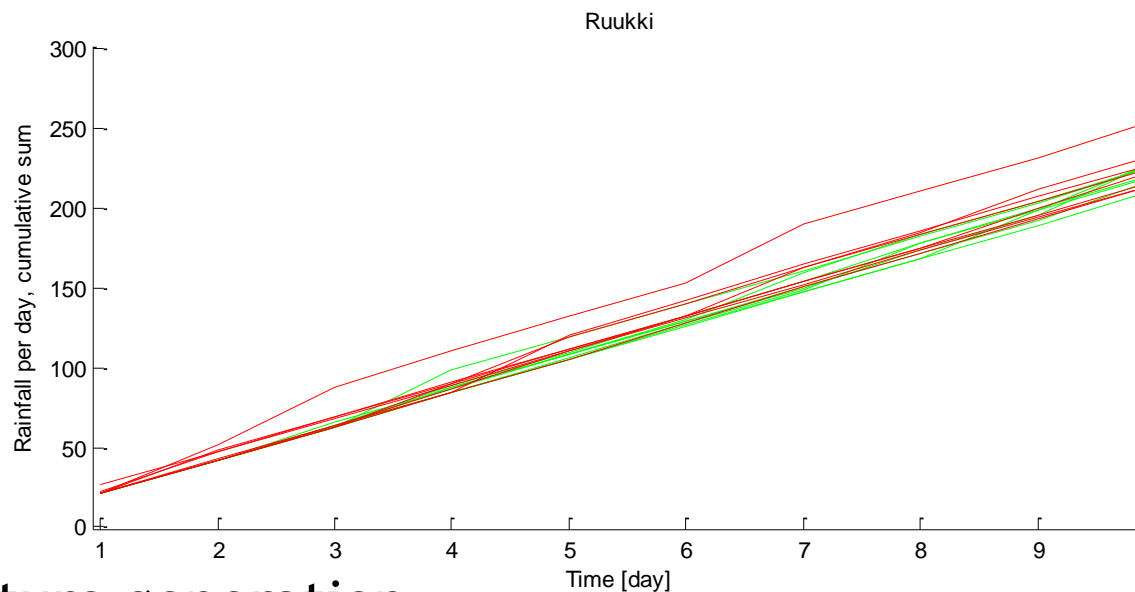


A little amount of
net blotch



DATA ANALYSIS

- Feature; cumulative sum of the rainfall per day
 - Green = years with no net blotch, red = years with net blotch

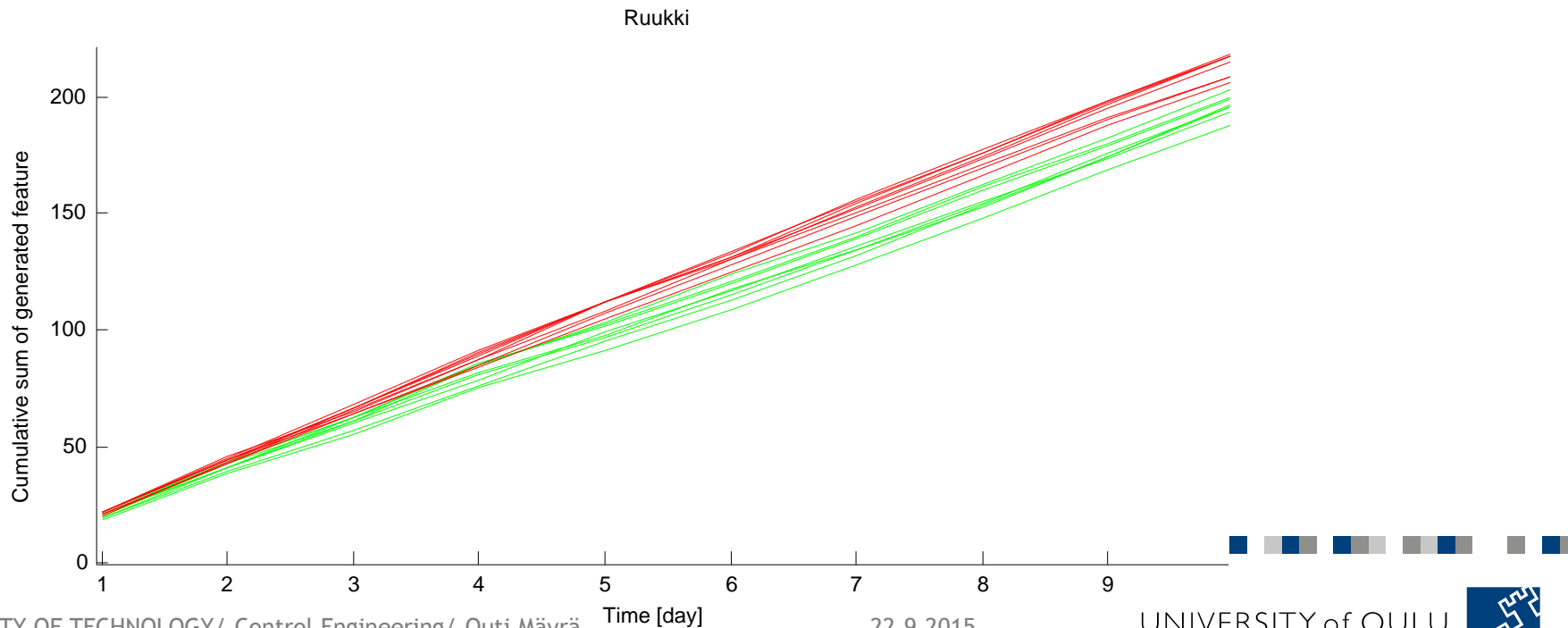


- Feature generation
 - Combining variables with different mathematical operations
 - Ranking the generated features and selecting the suitable one



DATA ANALYSIS

- Feature; $(\ln(z)+\ln(x))*\ln(z)$;
 - z = minimum daily temperature
 - x = rainfall per day
 - Used algorithm from Ruusunen M.: Signal Correlations in Biomass Combustion - an Information Theoretic Analysis. Acta Univ Oulu C 459, 2013. PhD Thesis



THE USE OF RESULTS

- Predict the appearance of the net blotch
 - Optimize the use of chemical plant protectants
 - Avoid the unnecessary spraying of the fields
- Positive effects to crop
- Less chemical => save money and nature



THANK YOU!



- Picture from Bachelor's thesis of Ville Ruuhonen (2015) Disease resistance of the new spring cereal varieties

