#11 eRka meetup -> Cracow R User Group Meeting

PM_{10} concentrations in Kraków
- climatology of the air pollution and possibilities of short-term forecasting

Bartosz Czernecki

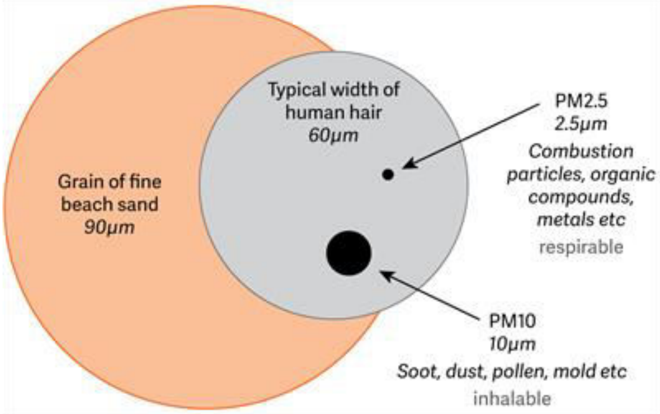
Adam Mickiewicz University in Poznań

Department of Climatology
Faculty of Geosciences
Adam Mickiewicz University in Poznań, Poland
nwp@amu.edu.pl

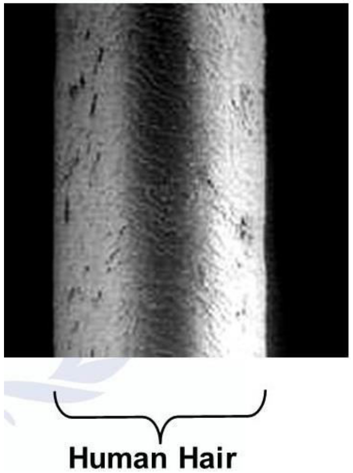
Overview (1)
– What is particulate matter (PM)?

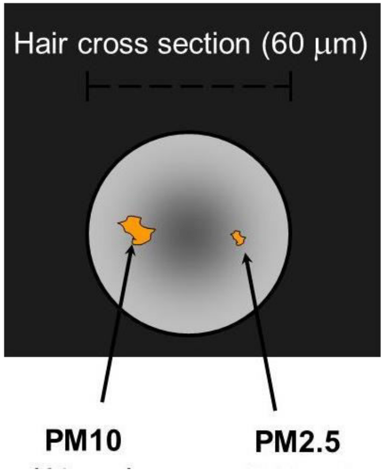
- **Particulate air pollution** is defined as an **air-suspended mixture of both solid and liquid particles.**

- They are often separated into three classes:
  - coarse (10µm and 2.5µm) -> settle quickly
  - fine (0.1 to 2.5µm in diameter)
  - ultrafine particles (<0.1µm in diameter) -> remain in suspension for longer



- To put things into **perspective**:
  - a human hair is about 50-70µm
  - a grain of sand has a diameter of 90µm





- **Synonyms**: dust, particulate matter, inhalable particles, respirable particles, smoke, mist

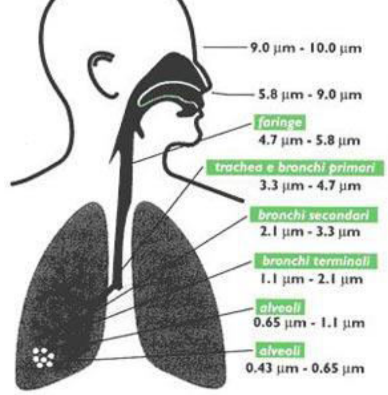
Overview (2)
– Where does it come from?

Sources:

- Particulate matter can come from both human and natural sources

- **Natural:**
  - Sea salt, forest fires, pollen, Sahara desert (etc..)
- **Human related (main sources):**
  - Road dust, construction activities, vehicle emissions (PM_{10})
  - Fuel burning, industrial combustion processes, vehicle emissions ($PM_{2.5}$)

- The chemical properties of PM_x vary depending on sources of particles. It is important to note that particulates are not one particular chemical substance.



Overview (3) – Why should we care?

The growing awareness is associated with the potential damaging effects to our health:

1. PM_{10} (or 2.5) make up a large proportion of dust that can be drawn deep into the lungs
   (larger particles tend to be trapped in the nose, mouth or throat)

2. Primary health effects include damage to the respiratory and cardiovascular systems.

3. They can penetrate the deepest parts of the lungs as well as access the gas exchange regions of the lung via diffusion.

4. The WHO believes particles are **affecting more people worldwide than any other pollutant.** (reason for **44,000 fatalities per year in Poland, and 440,000 in the EU)**

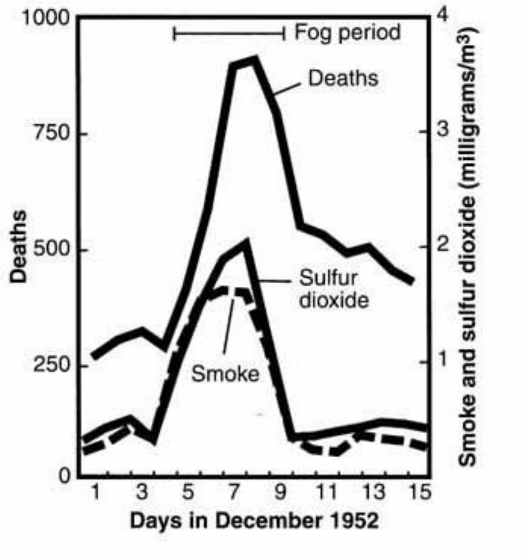
Acute effects	Chronic effects
Lung inflammatory reactions	Increase in lower respiratory symptoms
Respiratory symptoms	Reduction in lung function in children
Adverse effects on the cardiovascular system	Inrease in chronic obstructive pulmonary disease
Increase in hospital admissions	Reduction in lung function in adults
Increase in mortality	Reduction in life expectancy
Increase in medication usage	

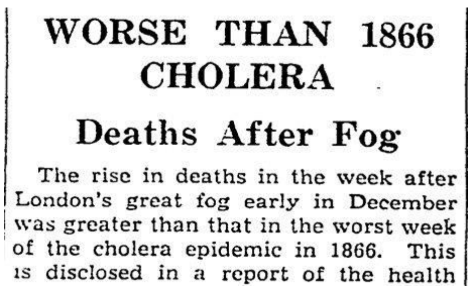
Health effects attributed to exposure to suspended PMx
Source: Health aspects of air pollution (WHO, 2004)

Overview (4)

Example – The great smoke of London (1952)

(aka „killer fog”, „big smoke”)





WORSE THAN 1866 CHOLERA

Deaths After Fog

The rise in deaths in the week after London's great fog early in December was greater than that in the worst week of the cholera epidemic in 1866. This is disclosed in a report of the health

Background:

- 5 to 9 December 1952
- cold weather in an anticyclone and near-windless conditions
- airborne pollutants – mostly arising from the use of coal

Government medical reports estimated ~ over **1,000 deaths per day** as a direct result of the smog and **100,000 got ill** by the smog's effects on the **human respiratory tract.**

The **total number of fatalities 12,000**

Rethinking of air pollution. New regulations implemented

Overview (6)
– Regulations

The European Union has developed an extensive body of legislation which establishes health based standards:

Pollutant	Concentration	Averaging period	Legal nature (Date of limit value enters into force)	Permitted exceedances each year
$PM_{2.5}$	25 uq/m3	1 year	1.1.2015	-
PM_{10}	50 uq/m3	24 hours	1.1.2005	35
PM_{10}	40 uq/m3	1 year	1.1.2005	-

http://ec.europa.eu/environment/air/quality/standards.htm

Alert level in Poland for PM10 – 300uq/m3 for 24h hour mean
(used to 200 uq/m3, but this value is now assigned as informing level)

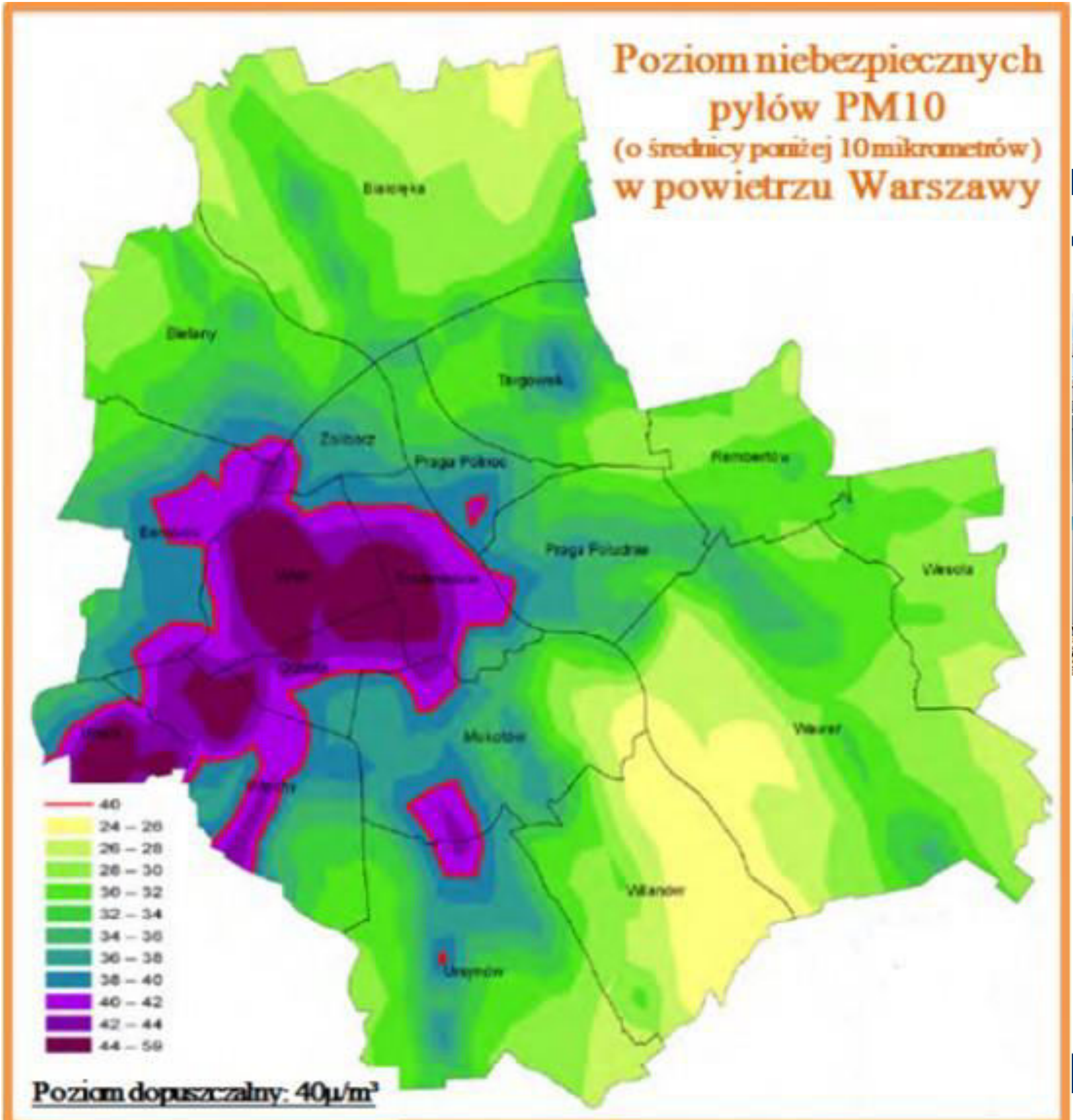
Rozporządzenie Ministra Środowiska z dnia 24 sierpnia 2012 r. w sprawie poziomów niektórych substancji w powietrzu (Dz. U z 2012 r., poz. 1031)

Overview (7)
– Need for operational monitoring!

Differences between countries:

- Many countries have legislation that require ambient air quality monitoring:
  - Responsibility prescribed for governmental or local authorities
  - Additional regulations for particular industries (petrochemical, waste, etc...)
  - No international agreements = no homogenized data series = no valuable comparison between some countries

- **Poland:**
  - **Operational monitoring**: Chief Inspectorate for Environmental Protection and related
  - **81 stations** in Poland (mostly in big cities)
  - Standards: (www.gios.gov.pl/artykuly/126/Badania-tla-zanieczyszczenia-atmosfery-wg-programow-miedzynarodowych)

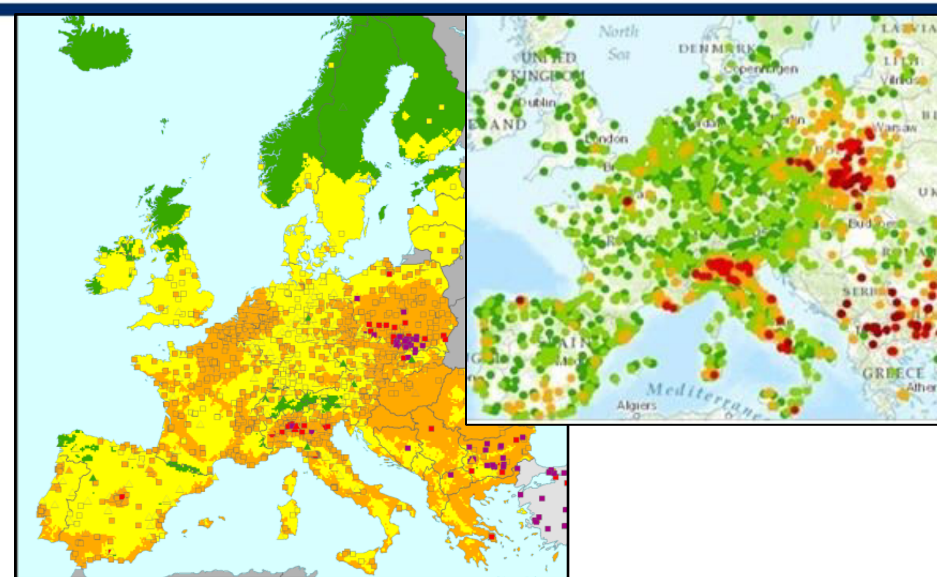

Poziom niebezpiecznych
pyłów PM10
(o średnicy poniżej 10 mikrometrów)
w powietrzu Warszawy
Białołęka
Bielany
Targówek
Żoliborz
Praga Północ
Rembertów
Praga Południe
Wesoła
Mokotów
Wawer
Wilanów
Ursynów
40
24 – 26
26 – 28
28 – 30
30 – 32
32 – 34
34 – 36
36 – 38
38 – 40
40 – 42
42 – 44
44 – 50
Poziom dopuszczalny: 40μ/m³

nitoring!

Part 2

Climatology

Data availibility for the city of Cracow for the needs of climatological analysis



Background:

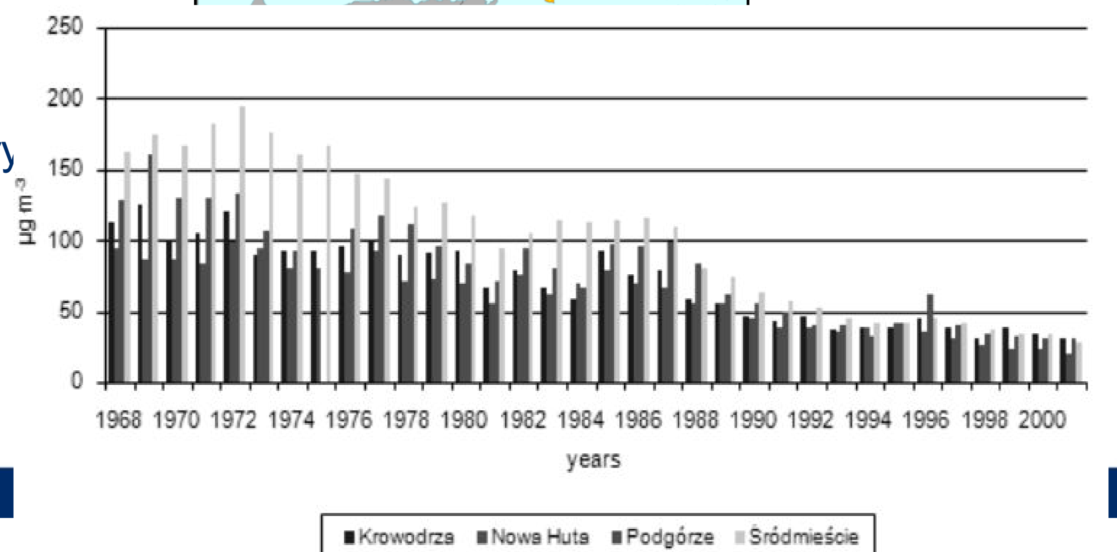
Cracow needs to deal with the problem of a large PMx concentration and year-by-year occupies top positions in the ranking of most polluted air

(Bokwa 2008, Juda-Rezler and Manczarski 2010; Krynicka and Drzeniecka-Osiadacz 2013; Rogula-Kozłowska et al. 2014; Rozbicka and Michalak 2015, European Environment Agency 2015).

Historical data:

- 1968-2001 -> the Voivodship Sanitary Epidemiological Station

- 1992-onwards -> Inspectorate for Environmental Protection in Kraków



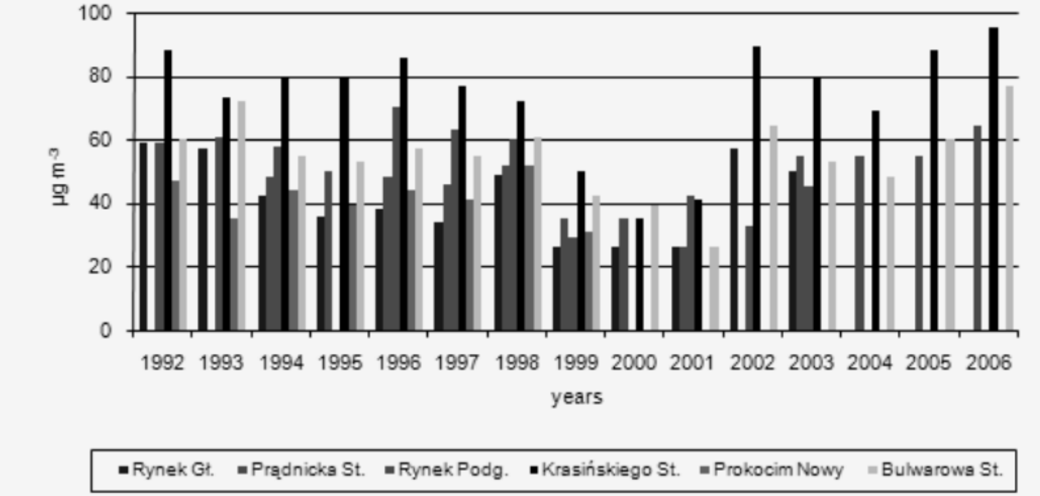
[A. Bokwa 2008]

Data:
- Some details and possible problems

Problems:

- Different stations' locations in the last ~25 years and different measurement techniques influences:
  - Accuracy/Range of errors
  - Problems with designing long-term trends
  - Problems with creating spatial patterns

- Hourly data not available for free (but not expensive) and available only from a few locations after 200x



Mean annual concentrations of PM10 (in μg·m^{-3}) in measurement points in Kraków in 1993-2006.

[A. Bokwa 2008]

Decision require compromise:

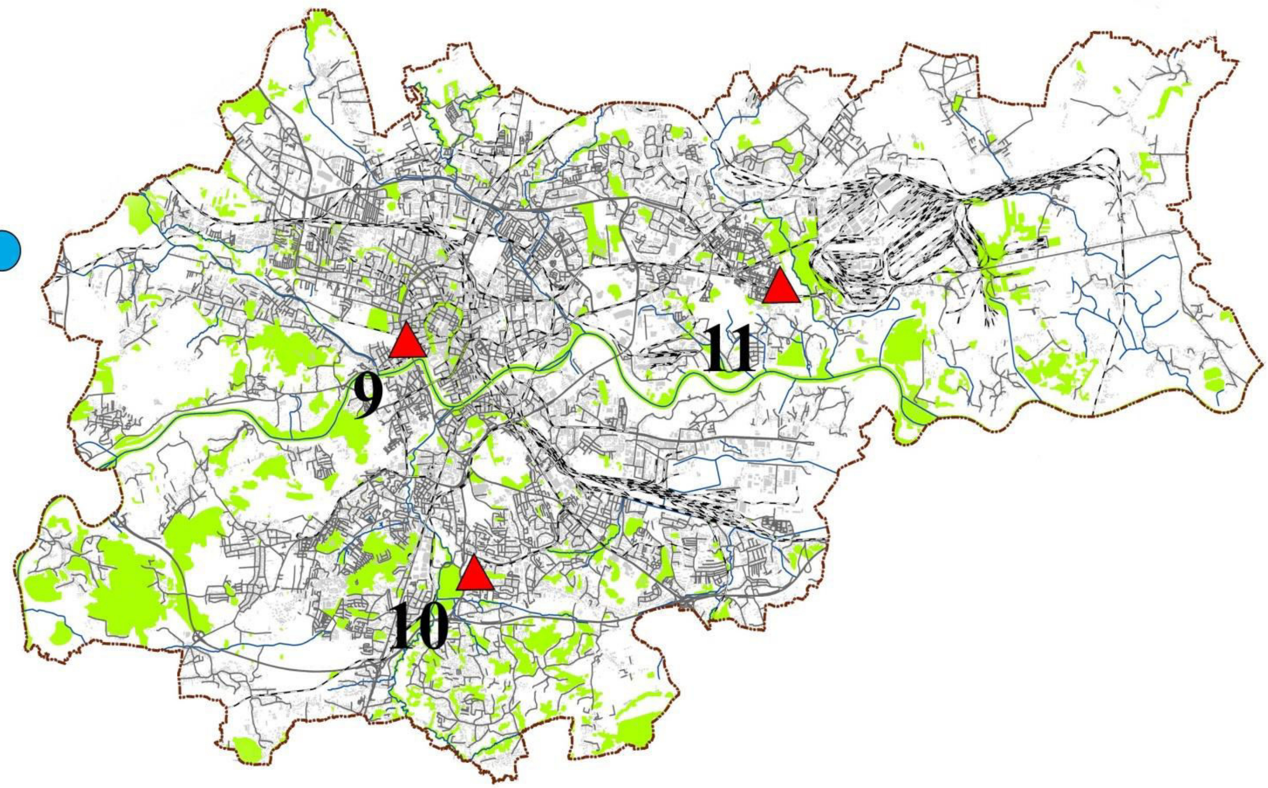
- Usage of shorter but more reliable dataset
  (10 years, 2006-2015)

- **9 - Krasińskiego**
- **10 – Kurdwanów**
- **11 – Nowa Huta**

Data:
- Some details and possible problems

Kraków Balice (meteorological station)

air pollution monitoring stations



0 5 10 km

Decision require compromise:

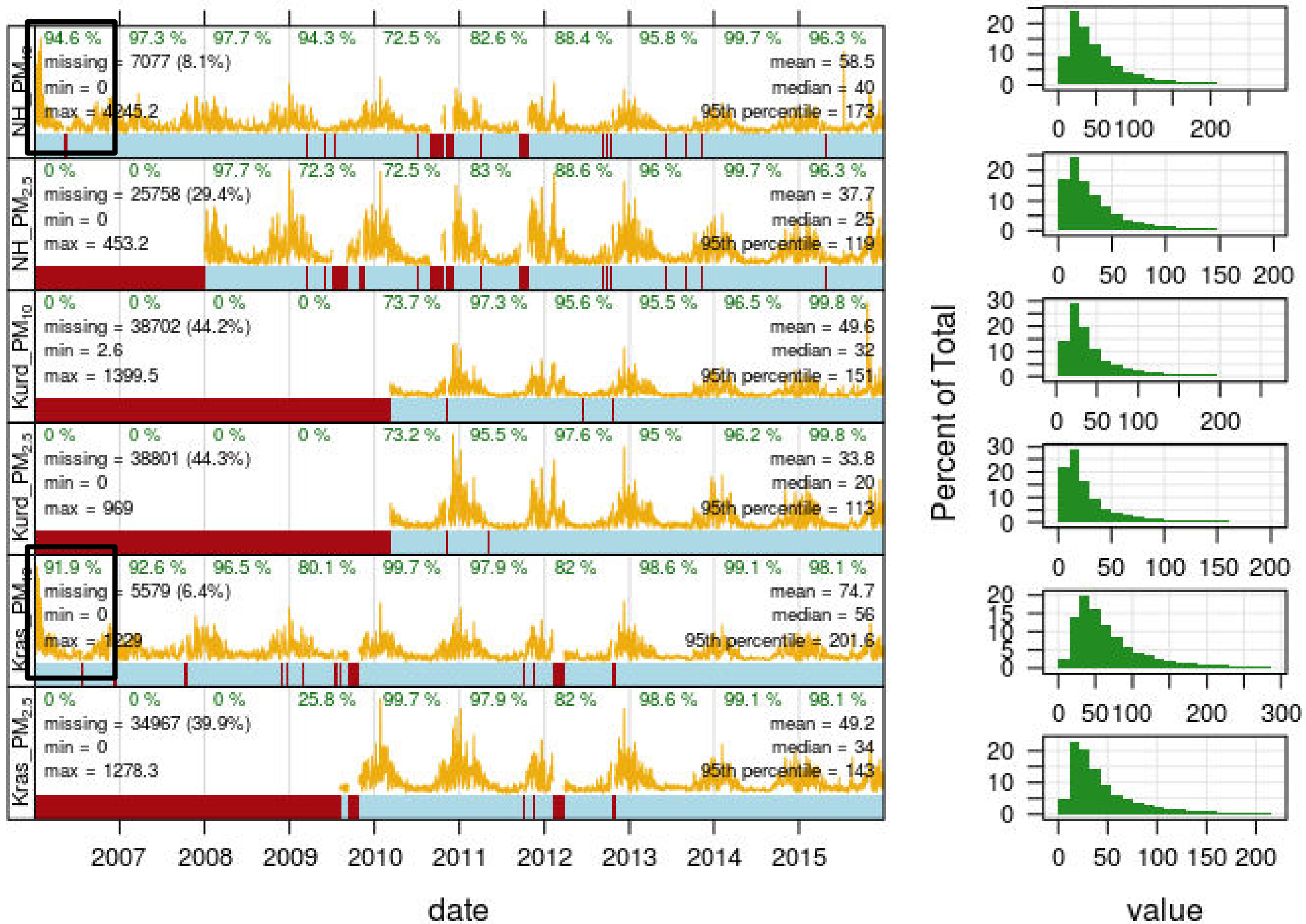
- Usage of shorter but more reliable dataset **(10 years, 2006-2015)**

- **9 - Krasińskiego**
- **10 – Kurdwanów**
- **11 – Nowa Huta**

Data:
- Air quality data mining in R

- **Manual investigation**

(~90k rows, 3x PM_{10}, 3xPM_{10})
- cleaning and finding possible „buggy” data



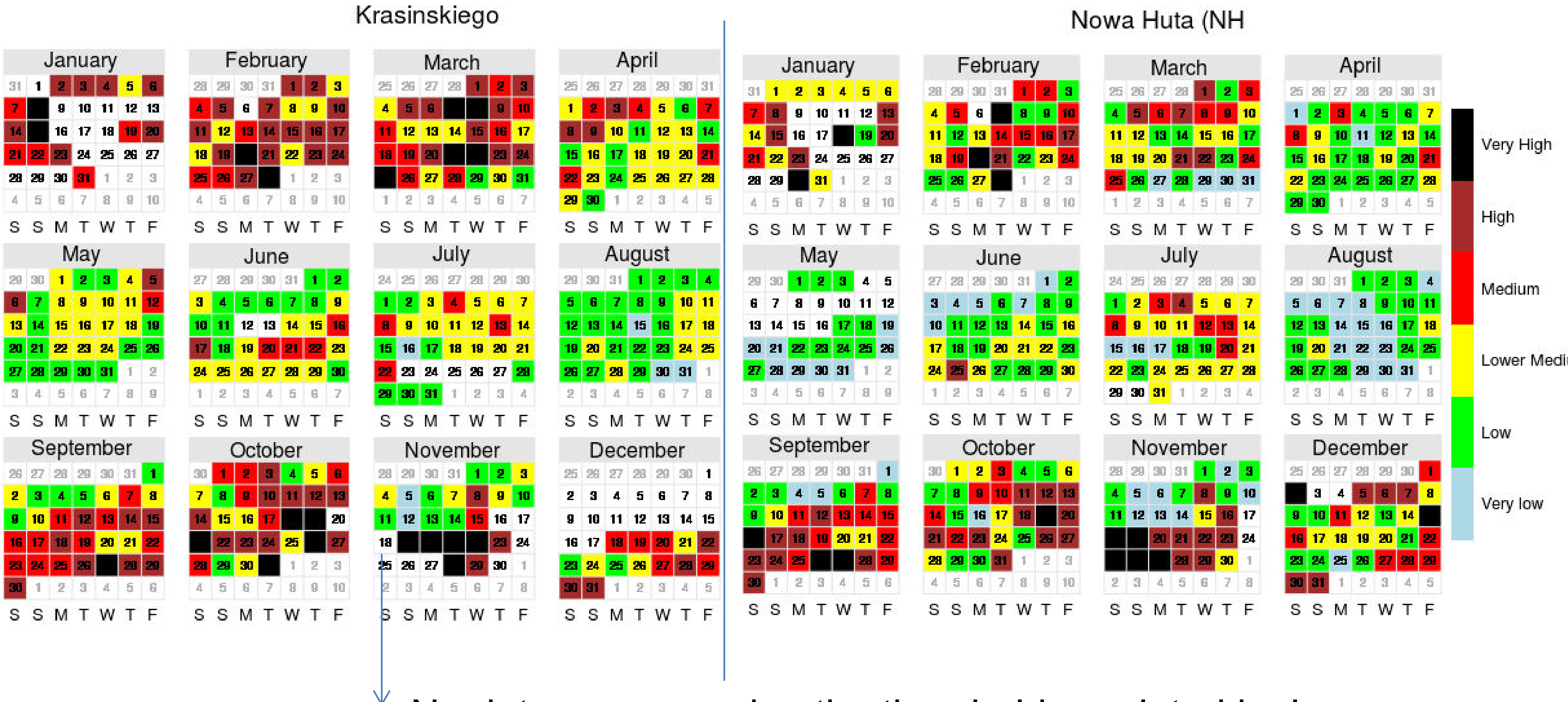
„openair” R package

(Carslaw, Ropkins 2012)

First conclusions:

- Only 2 stations with almost complete dataset after 2006 (PM_{10})
- Usually around 2-4% of observations missing

Data:
- Zoom to particular year with calendarPlot

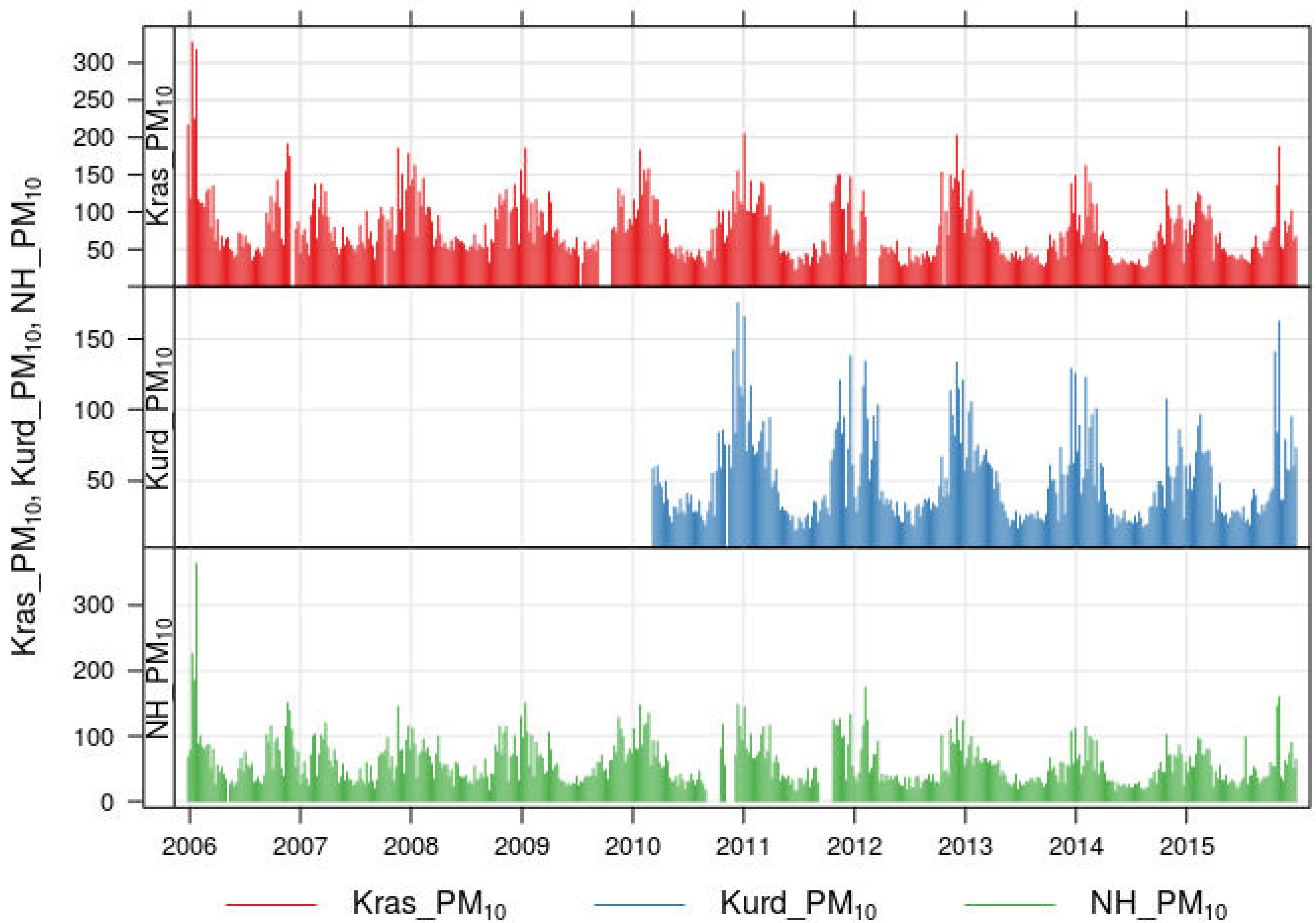


No data = no crossing the threshold regulated by law

calendarPlot(tt, pollutant = "Kras_pm10", breaks = c(0, 25, 50, 75, 100, 150, 200),labels = c("Very low", "Low", "Lower Medium", "Medium", "High", "Very High"),cols = c("lightblue", "green", "yellow", "red", "brown", "black"), statistic = "mean", main='Krasinskiego’, year=2006)

Data:

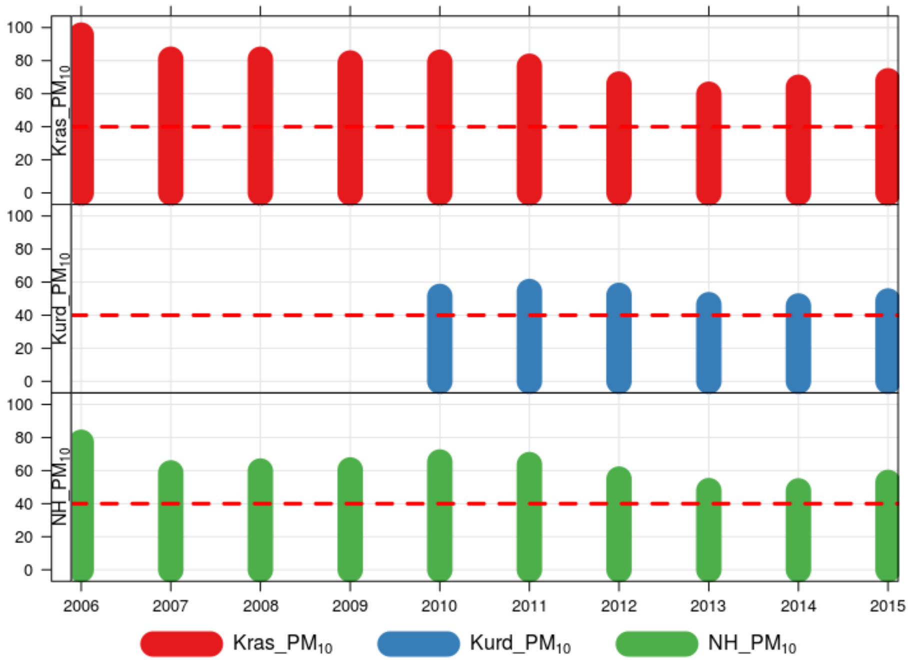
- Detecting seasonality, removing noise and making synchronized time-series plots



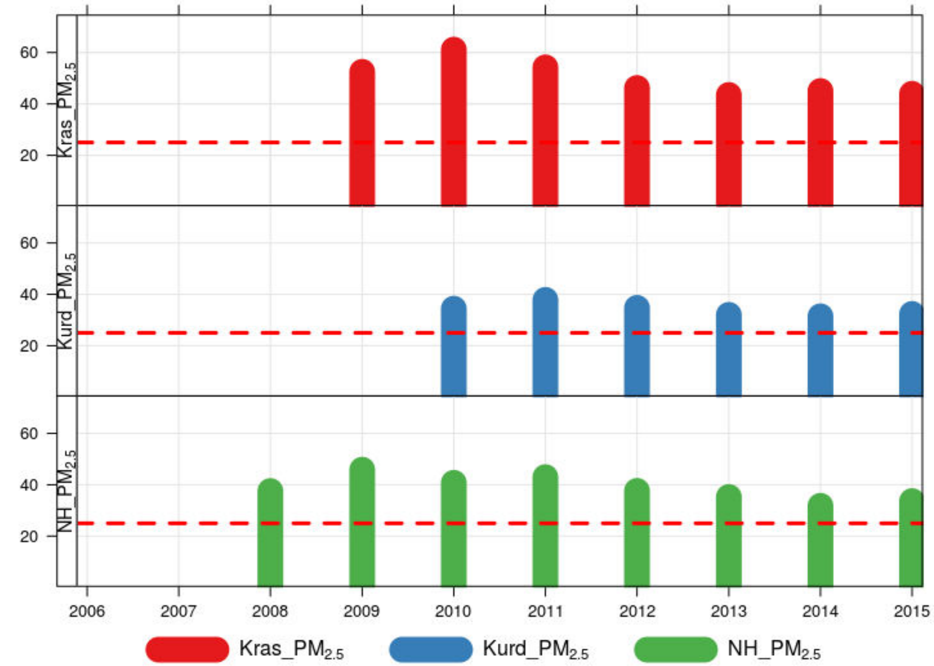
timePlot(tt, pollutant = c("Kras_pm10", "Kurd_pm10", "NH_pm10"), avg.time = "week",stack = F, plot.type = "h")

Data:
- Yearly standards

PM10
(mean annual concentration set by law = 40uq/m3)



PM2.5
(mean annual concentration set by law = 25uq/m3)



timePlot(tt, pollutant = c("Kras_pm25", "Kurd_pm25", "NH_pm25"), avg.time = "year",stack = F, plot.type = "h", statistic = "mean", group =F, key = T, ci = T, ref.y = list(25, lty=2, col="red", lwd=3), ylim=c(5,70), ylab="", lwd=20)

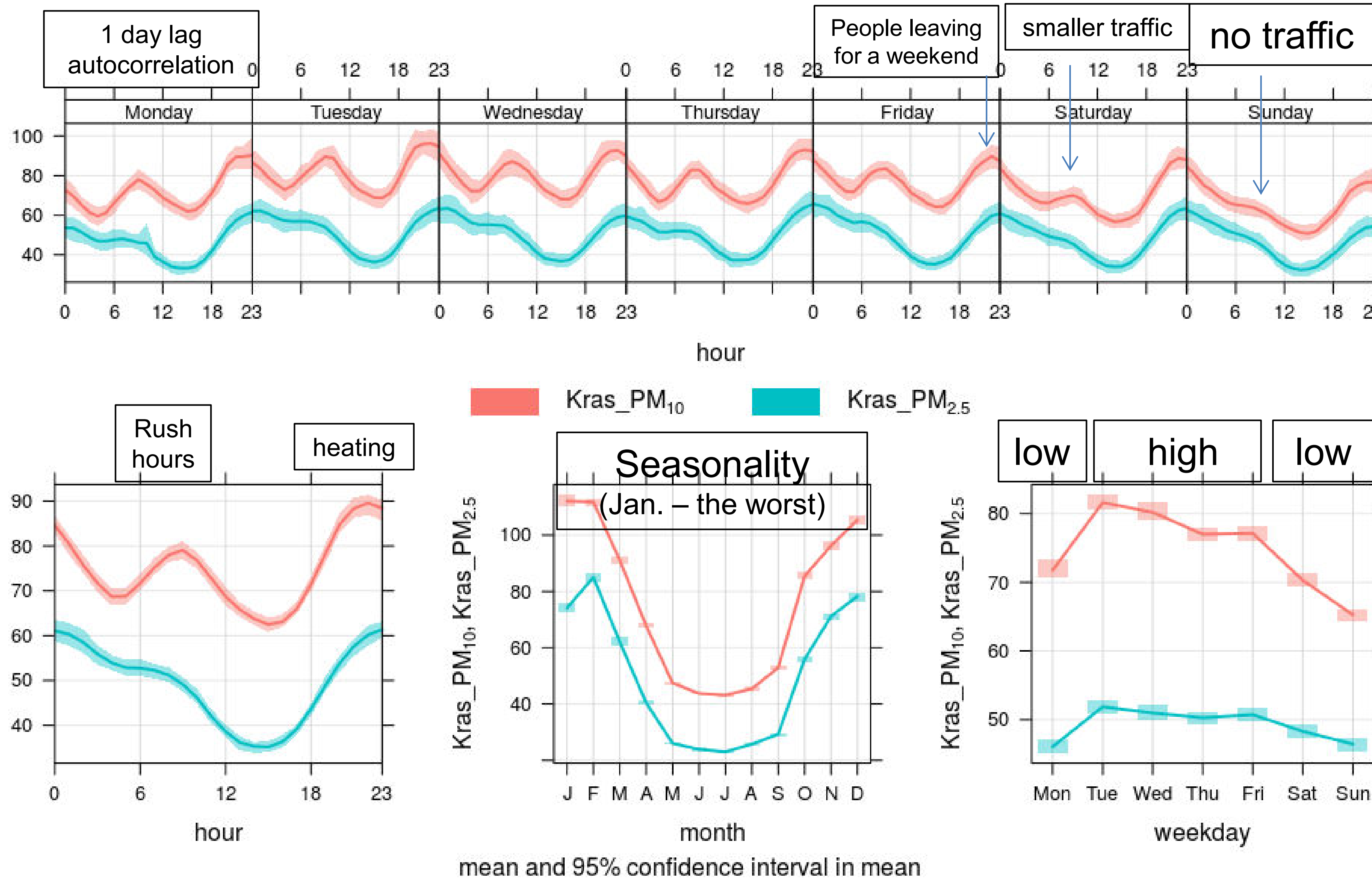
Data:
- Yearly standards

Number of days with mean PM10 concentration above 50 uq/m3
(standards set by law 35 days)

Year	Krasińskiego	Kurdwanów	Nowa Huta	Temp.	HDD
2006	255	-	198	8.4	91224
2007	249	-	165	9.3	82867
2008	270	-	179	9.4	81315
2009	226	-	169	8.7	87494
2010	226	82	128	7.5	97710
2011	203	137	155	8.8	86247
2012	139	122	129	8.6	90329
2013	163	116	128	8.7	88215
2014	190	108	127	9.8	77475
2015	206	111	127	10.0	78069
mean	**212,7**	**112,7**	**150,5**	**8.9**	**86094**

```
hdd <- function (x, x0 = 18, na.rm = TRUE) {  cold <- x < x0;   hdd <- sum; (x0 - x[cold], na.rm = na.rm) ;   return(hdd) }
przekroczenia <- function(x) sum((x>50), na.rm = T)
tt %>% select(-Kras_pm25,-Kurd_pm25, -NH_pm25, -date) %>% group_by(yy,mm,dd) %>% summarise_all( .funs="mean",
na.rm=T)  %>%  group_by(yy) %>% summarise_all(., .funs="przekroczenia")
```

Data:
- Time variation



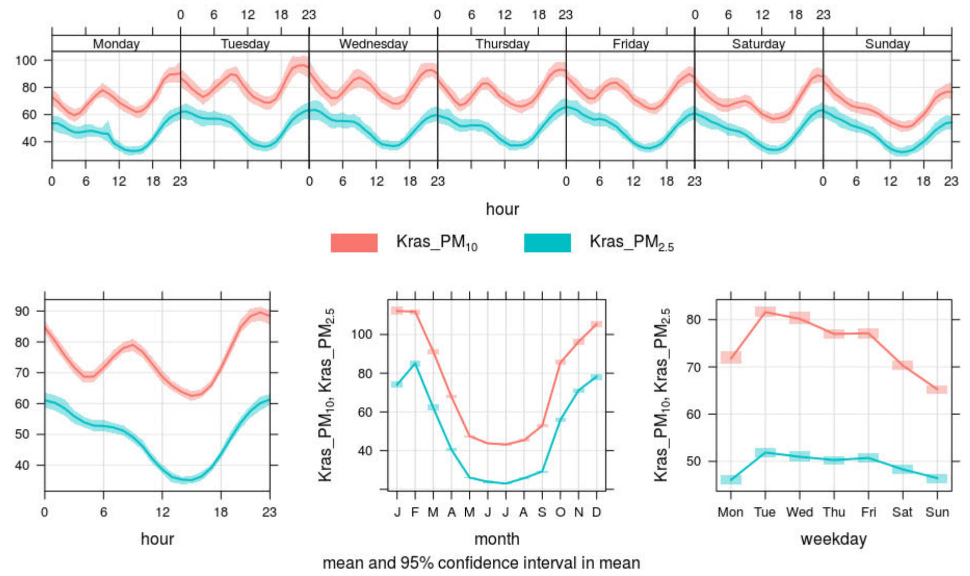
Data:
- Time variation

Brief summary:

- Clear reaction of PMx levels to different types of human activity

- traffic rush hours -> increase of PM10 concentr. on working days ~ 20uq/m3
- „best” air quality conditions around 14 UTC (15 or 16 local time, i.e. 1–2 h after the daily maximum air temperature)

- the daily maxima clear response to heating needs (ca. 40 µg m−3), usually observed midnight;
- more influencial than transport!

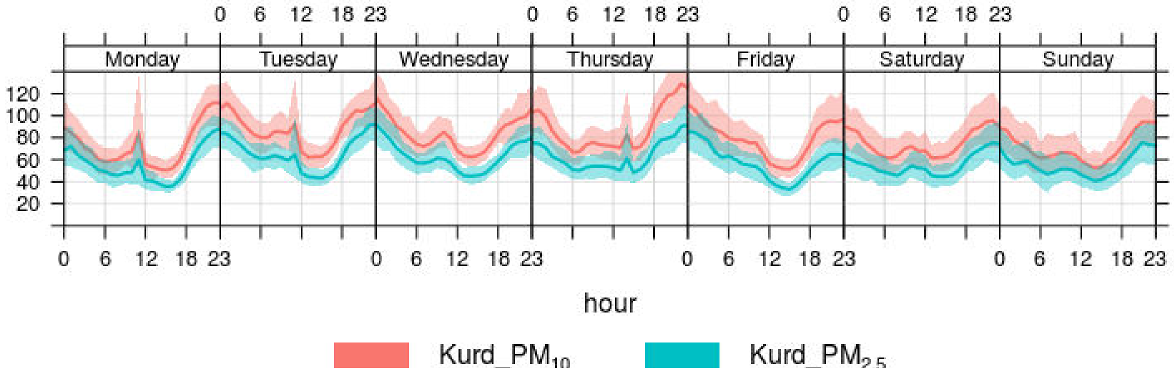


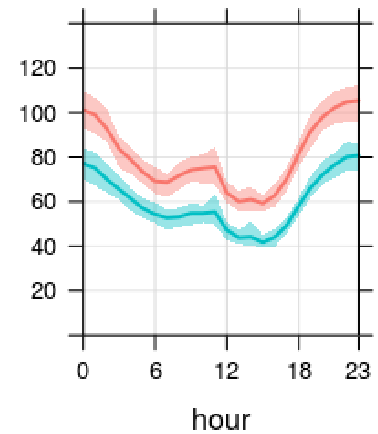
Outdoor activity:

- Air quality almost twice worse around midnight than at 15 local time!
- 2 best times during a day: before morning and afternoon rush hours

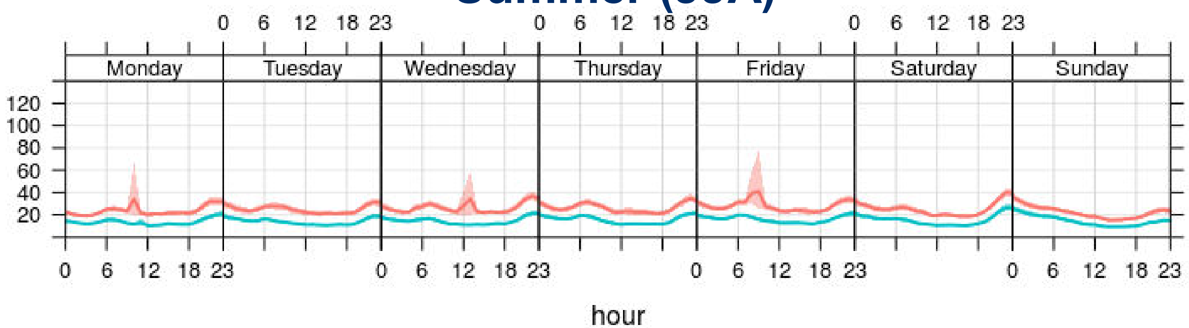
Data:
- Time variation - seasonal aspects

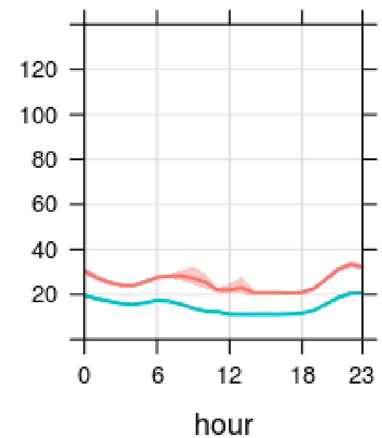
Winter (DJF)





Summer (JJA)

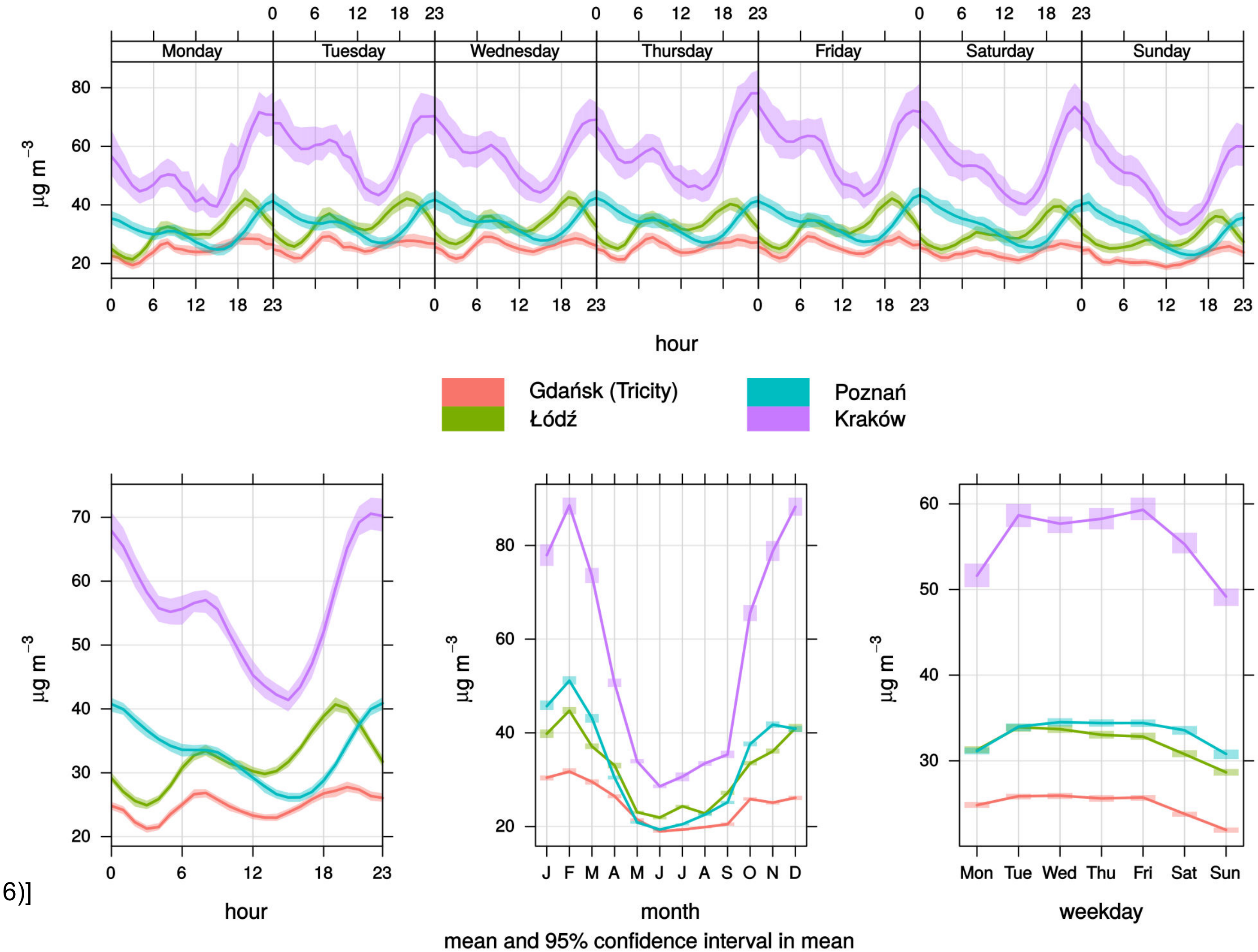




Conclusions:

- In winter season heating needs (on average) doubles PM10 concentration

Data:
- Time variation – short comparison



Is the air quality in Cracow really that bad on the Polish background?

- **Yes**

Are those regularities similar?

- **Not always...**

[J.Jędruszkiewicz, B. Czernecki, M.Marosz (2016)]

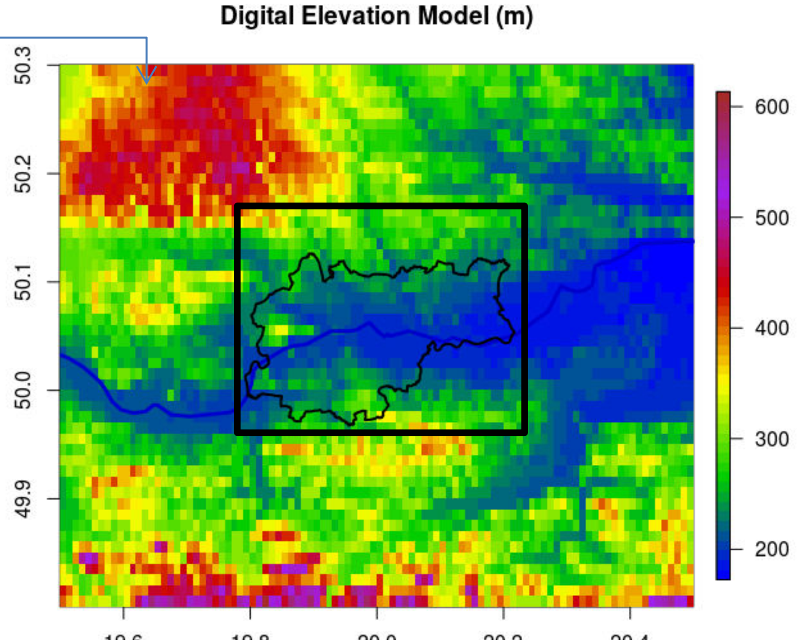
Data:
- Looking for a reason – „literature” mining

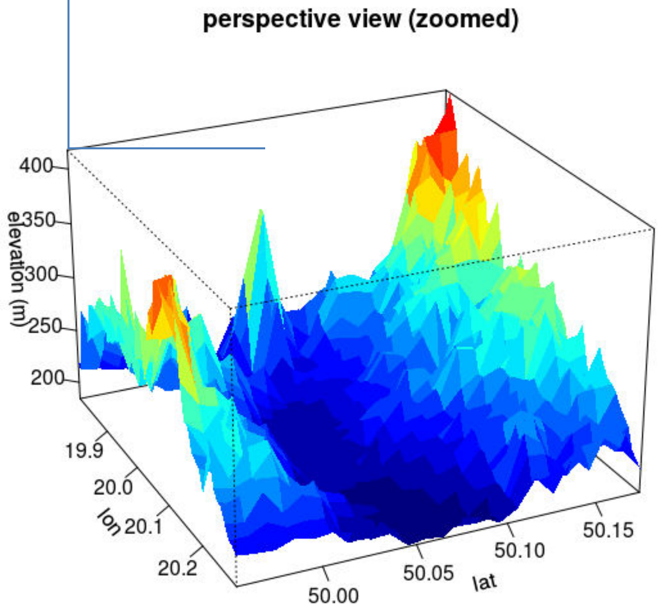
On average 70% impact of environmental conditions:

- Meteorological conditions
  (esp. wind spd., PBL height)
- Geographical features (location)
- Both categories strongly inter-related

On average 20-30% human activity:

- Heating systems
- Transport
- Industry





library(GA); persp3D(lon,lat,slope2, theta = 70, phi = 20, zlab = "elevation (m)", main="perspective view")

Data:

- Looking for a reason – meteorological data

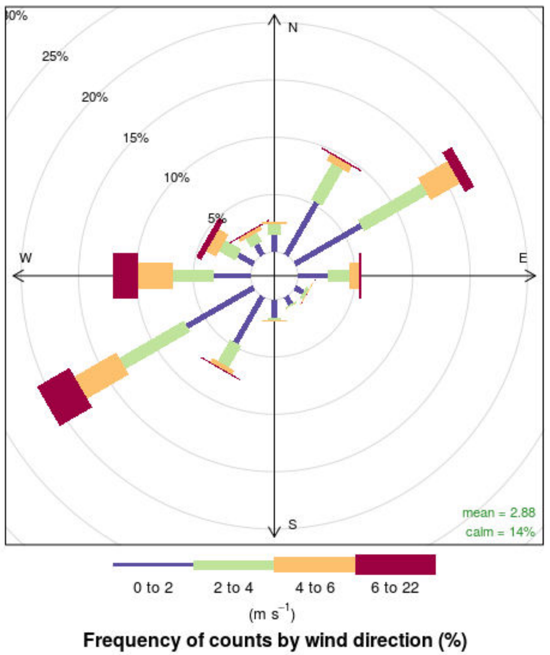
Credits:

- Official (**observational**) 1-hourly **meteorological dataset** for the city of Cracow provided by Institute of Meteorology and Water Managament – National Research Institute
    - Free for scienfitic purposes, otherwise ~1.5PLN per record

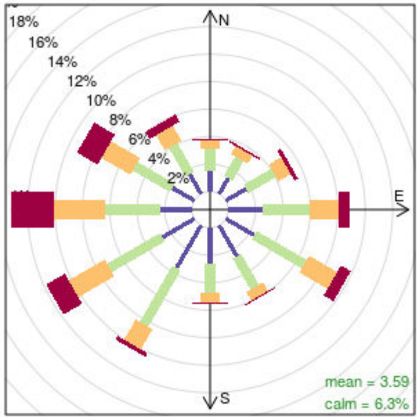
- ERA-Interim reanalysis project (3-hourly gridded dataset) written in NetCDF format

Ventillation conditions:

- **One of the lowest mean wind speed in Poland (~30% smaller than average)**
- **The highest frequency of wind calms (wind speed < 1m/s)**
- **Clearly valley-driven frequency of wind directions**

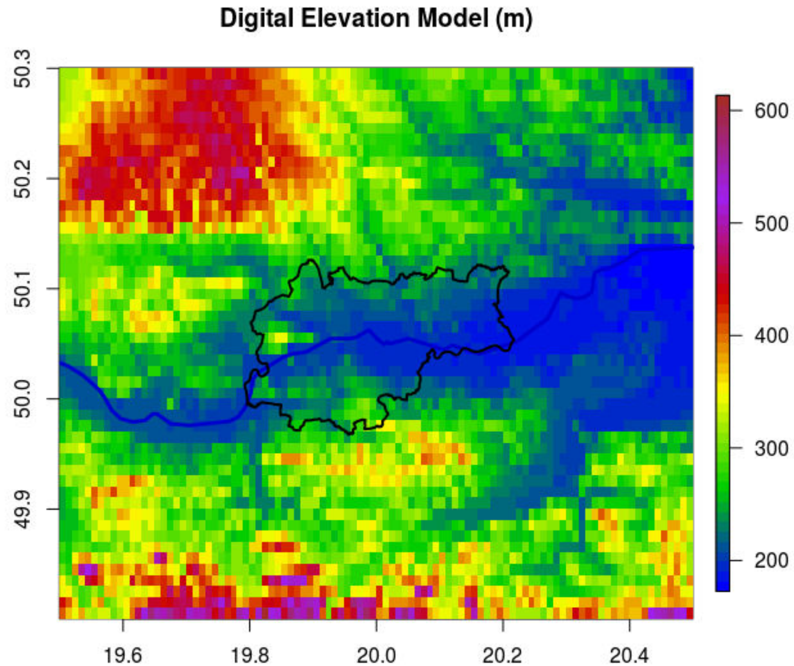


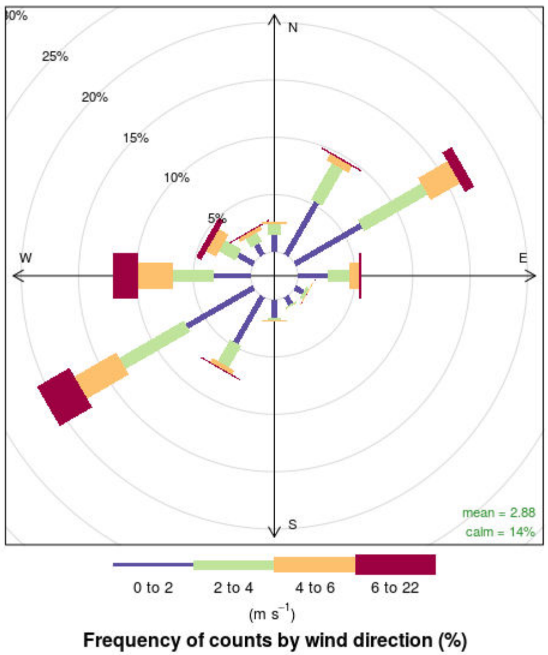
Kraków



Poland

Data:
- Looking for a reason – meteorological data



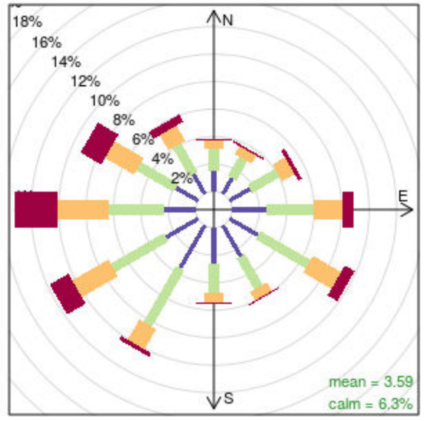


Kraków

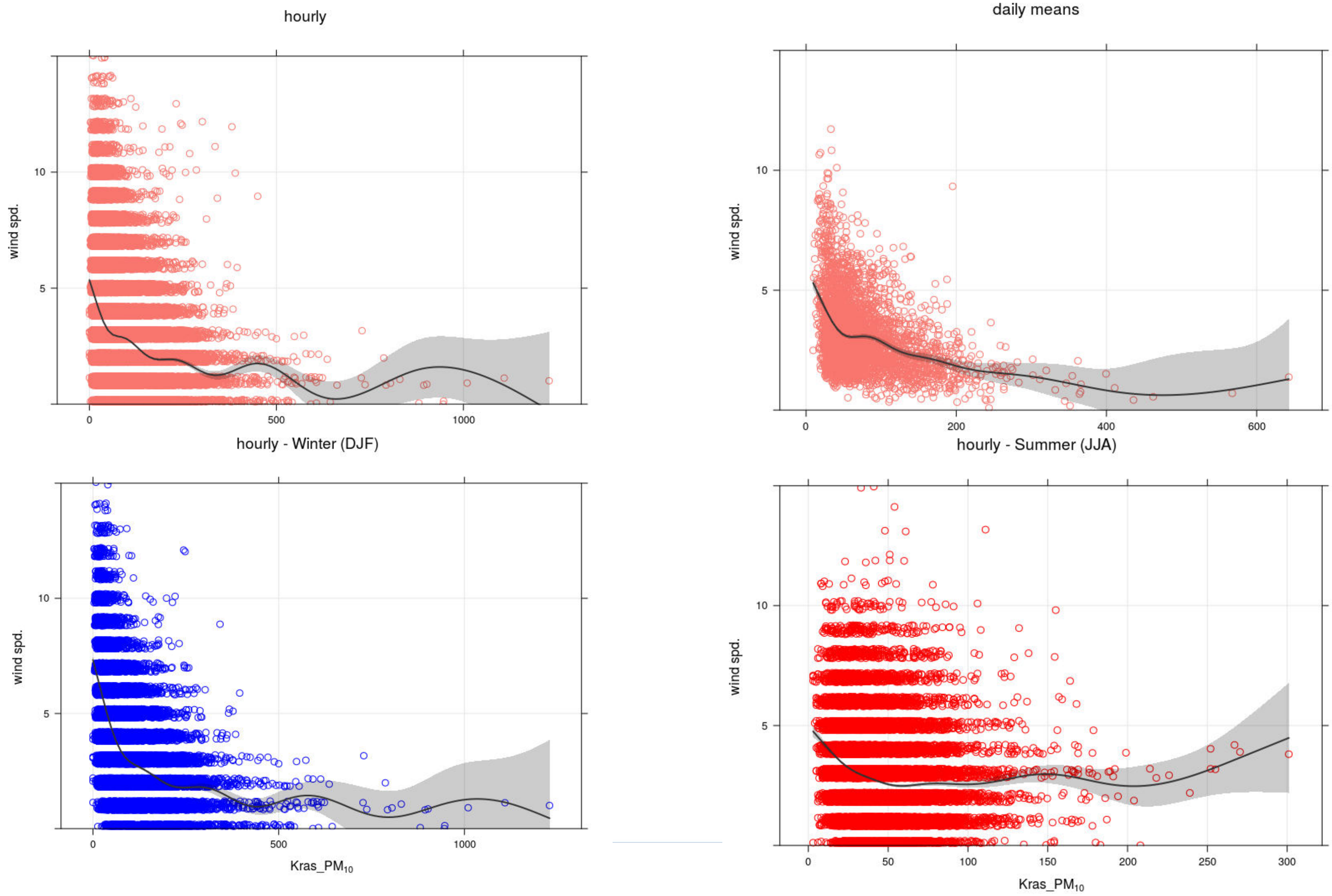
Ventillation conditions:

- One of the lowest mean wind speed in Poland (~30% smaller than average)
- The highest frequency of wind calms (wind speed < 1m/s)
- Clearly valley-driven frequency of wind directions

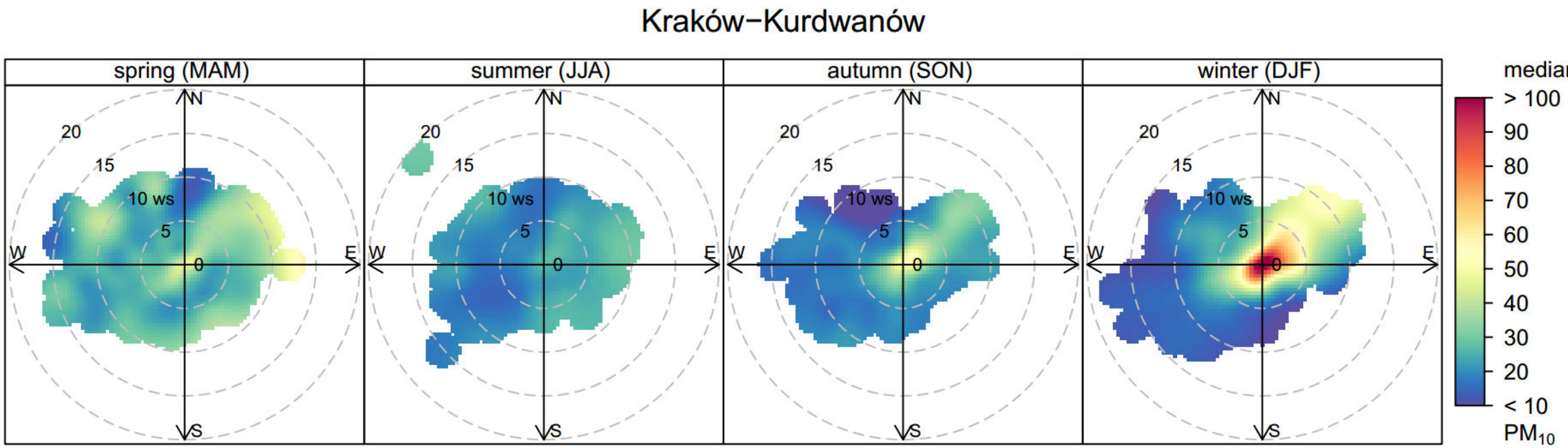
Poland



Data:
- Scatter plots: Wind speed

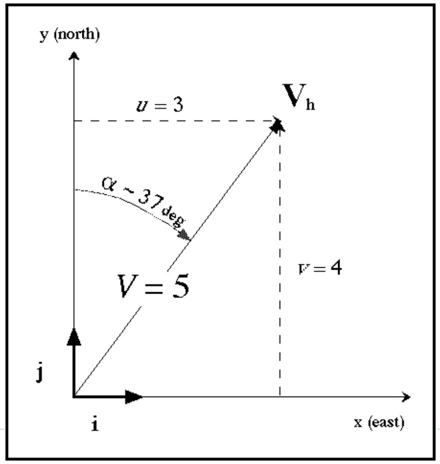


Data:
- Scatter plots: Wind direction + wind speed



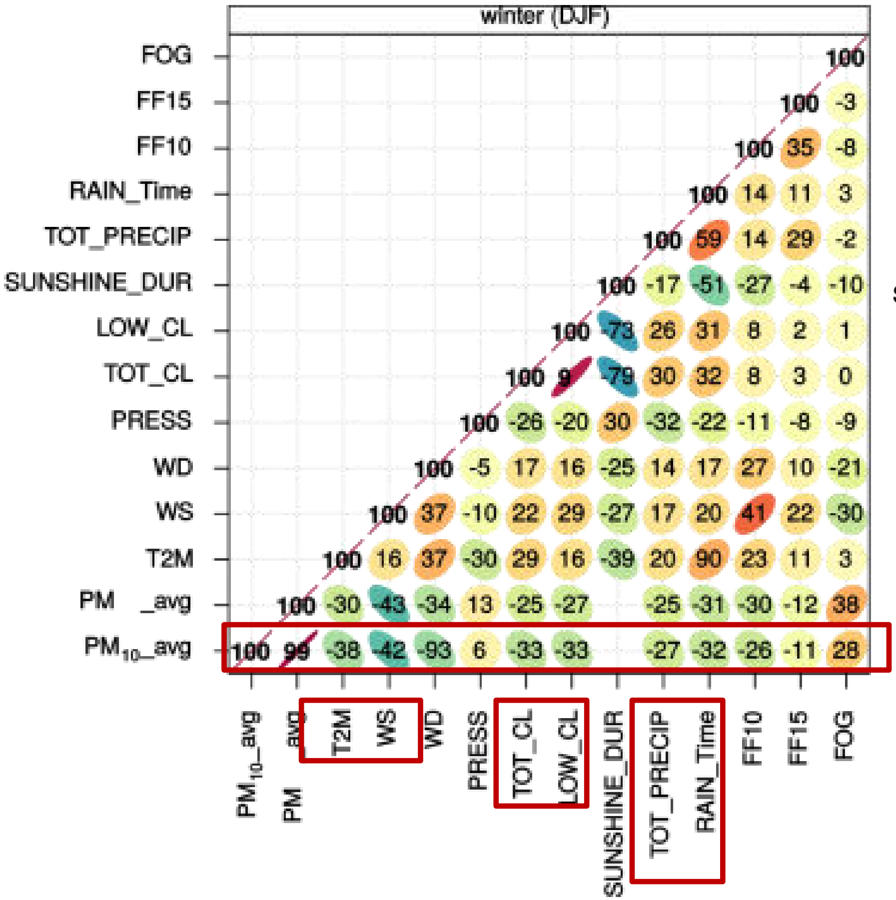
How to deal with averaging of wind direction?
(brief details)

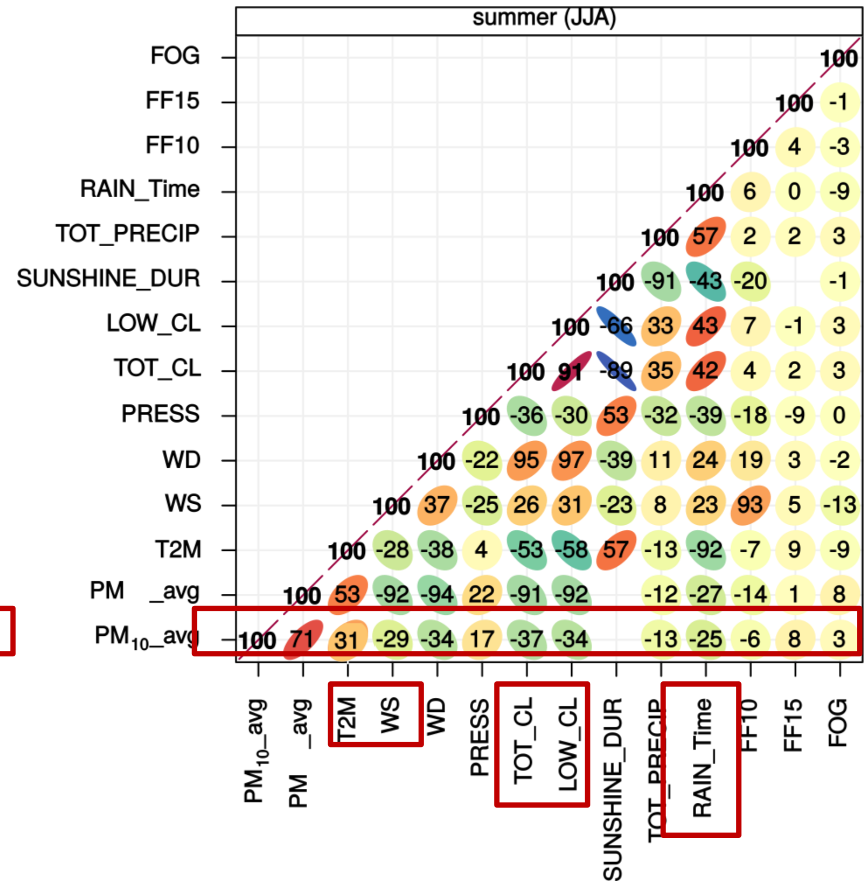
- **Adding angle 10 and 350 doesn't give north! (0)**
- **Applying vector calculations using**
  i = wind speed * cos(wind direction)
  j = wind speed * sin(wind direction)



Data:

Are there any other variables influencing PMx levels? -> example of daily means






```
summary(lm(Kurd_pm10~T2M+WS+TOT_CL, data=winter))$r.squared                >0.44
```

Air pollution dispersion inside Atmospheric boundary layer (PBL)

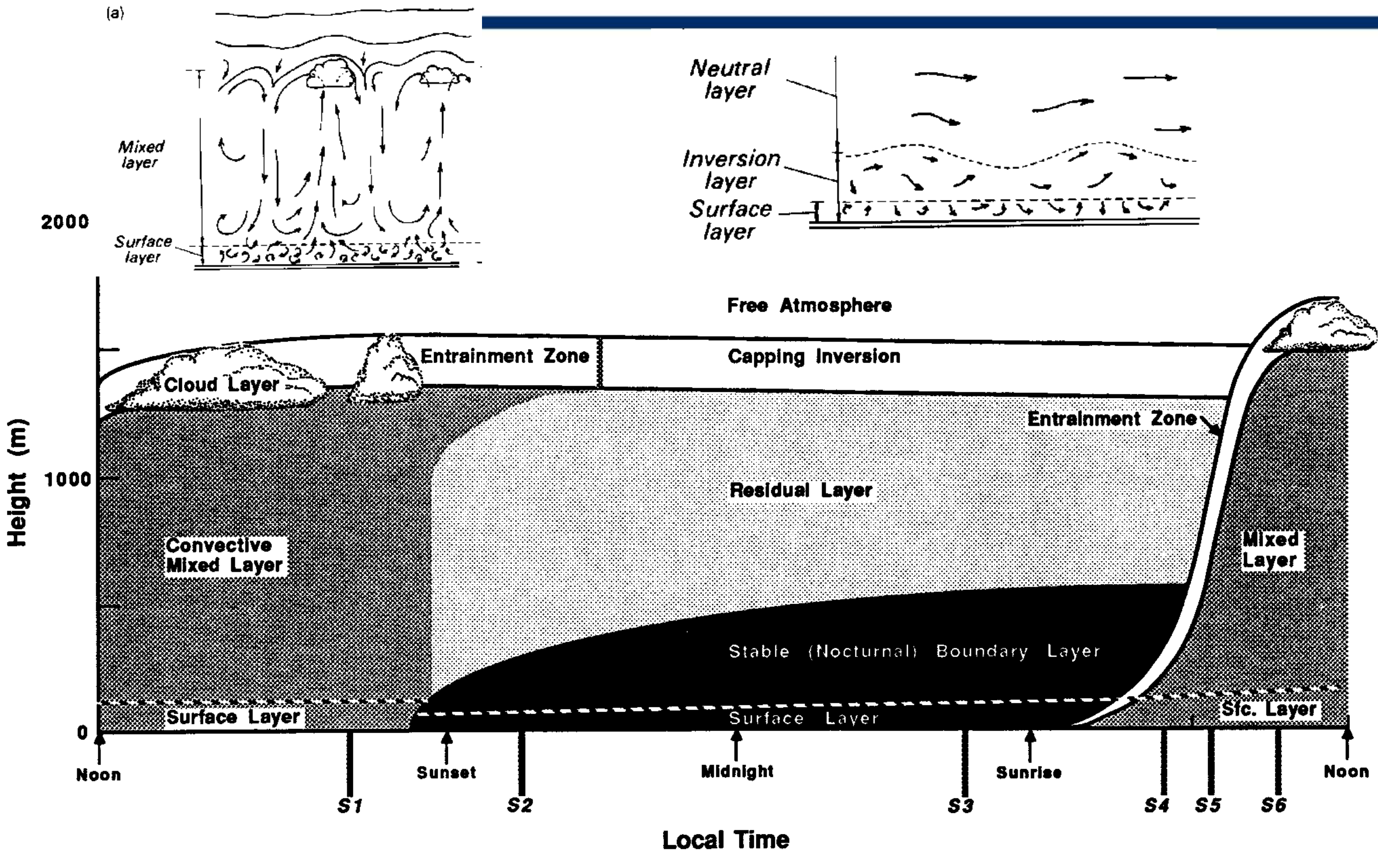
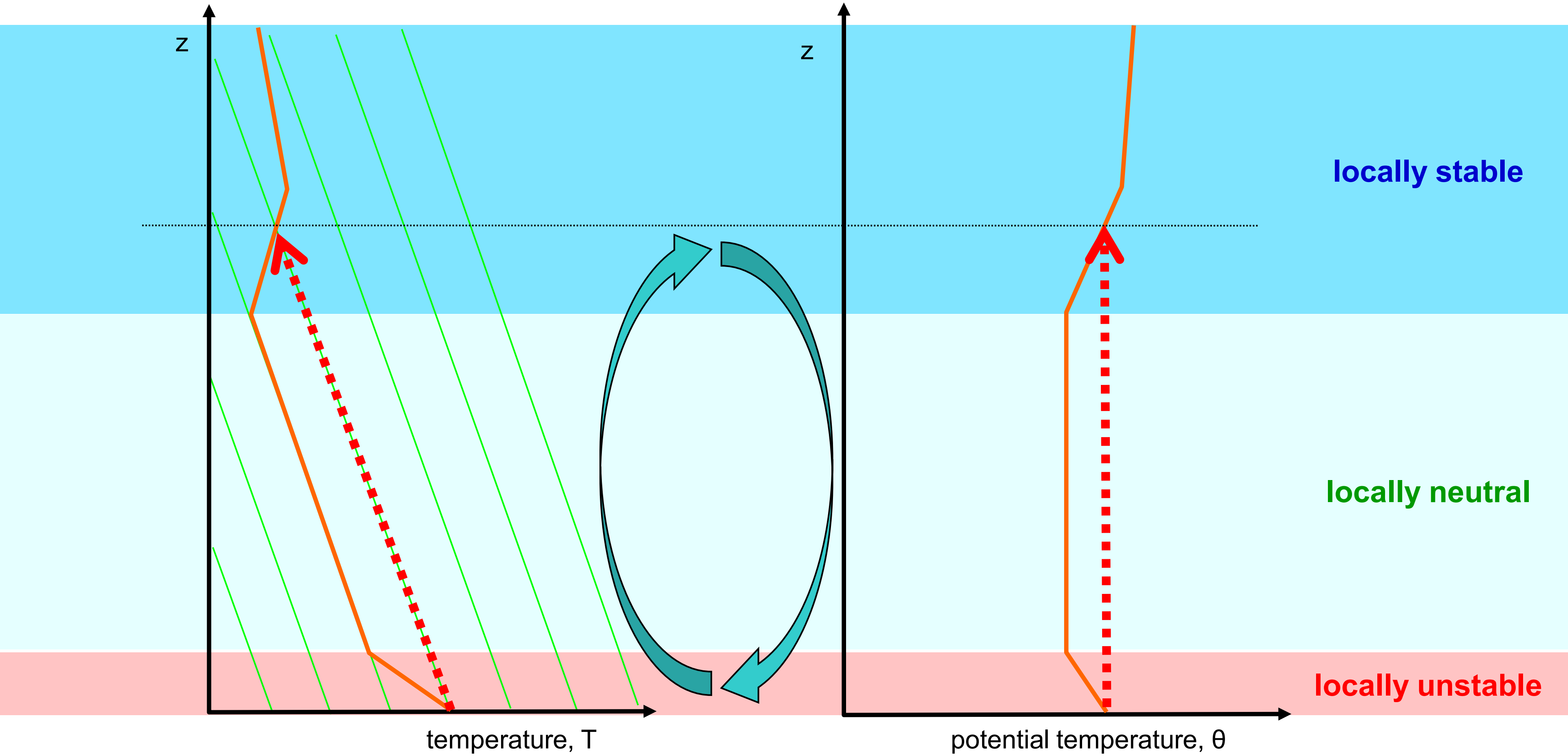


Fig. 1.7 The boundary layer in high pressure regions over land consists of three major parts: a very turbulent mixed layer; a less-turbulent residual layer containing former mixed-layer air; and a nocturnal stable boundary layer of sporadic turbulence. The mixed layer can be subdivided into a cloud layer and a subcloud layer. Time markers indicated by S1-S6 will be used in Fig. 1.12.

Physics of mixing conditions in PBL profile and atmospheric stability



PBL evalution in a daily cycle

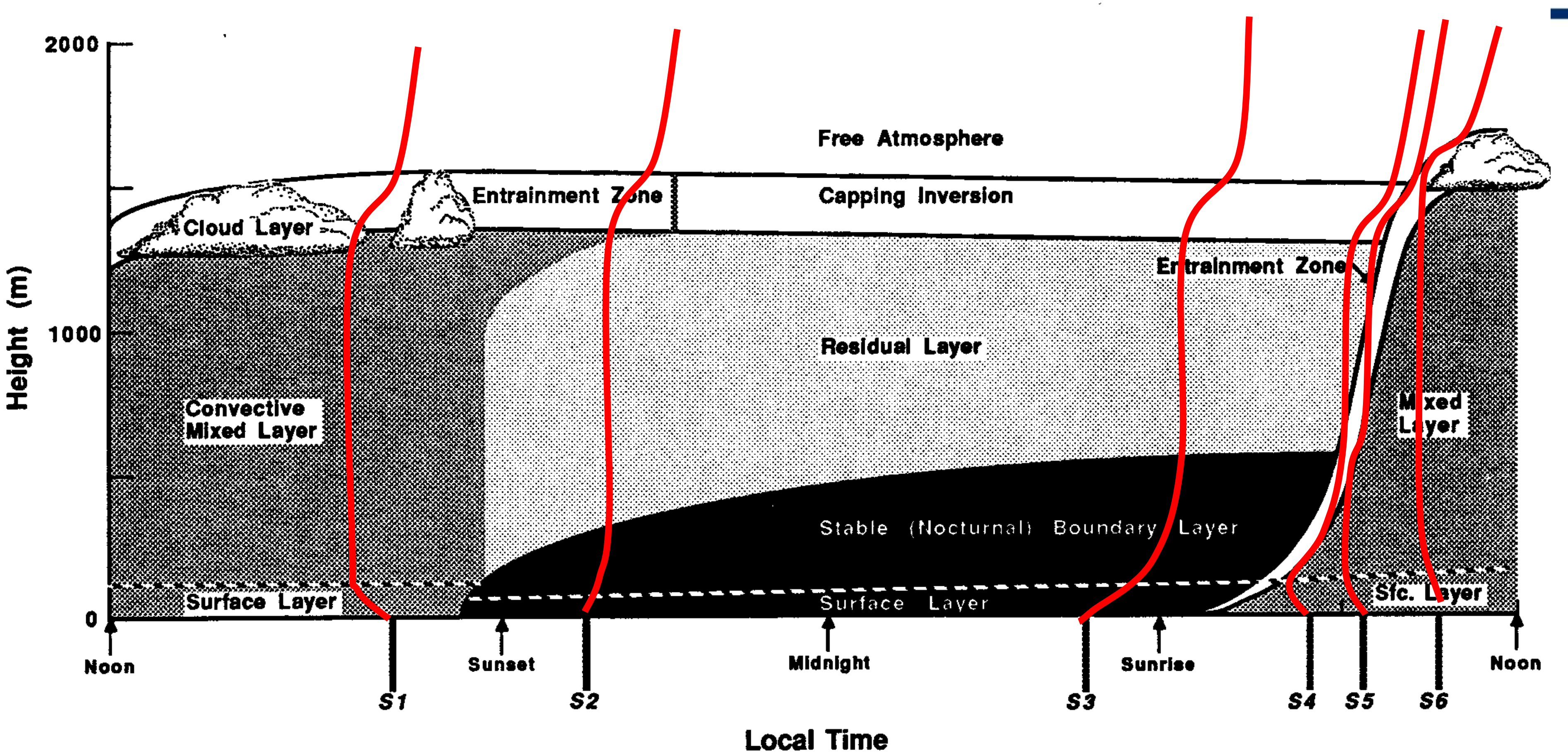


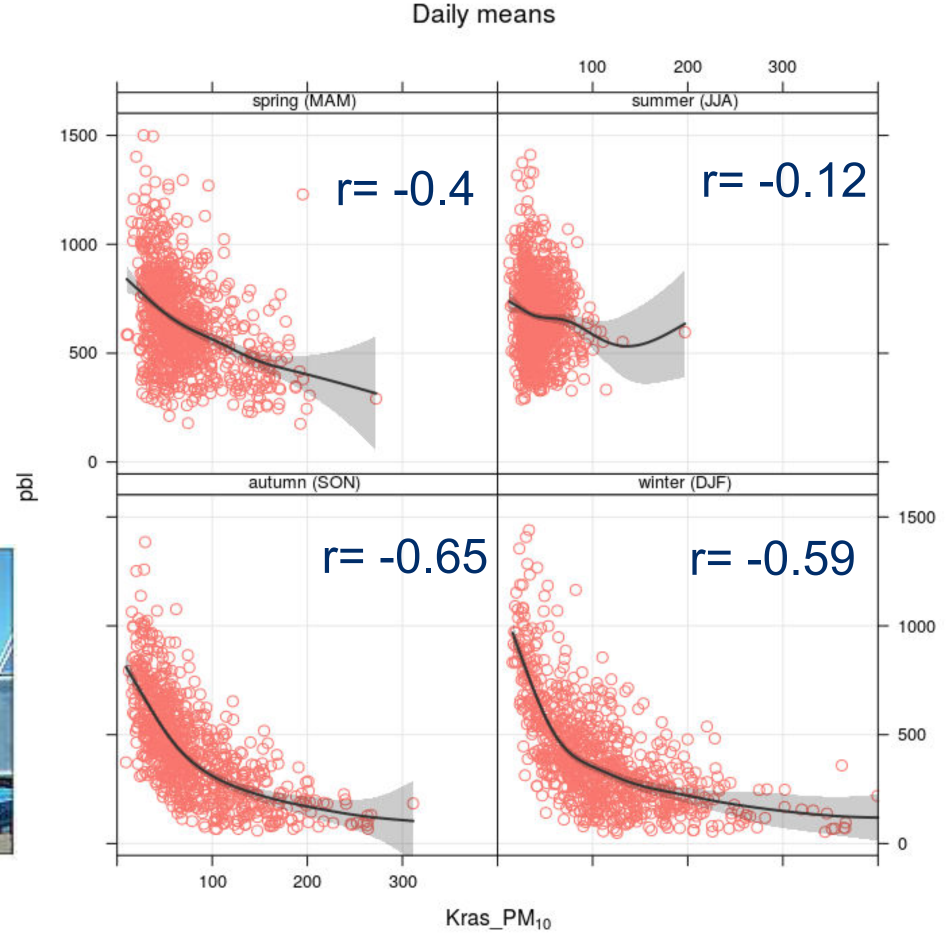
Fig. 1.7 The boundary layer in high pressure regions over land consists of three major parts: a very turbulent mixed layer; a less-turbulent residual layer containing former mixed-layer air; and a nocturnal stable boundary layer of sporadic turbulence. The mixed layer can be subdivided into a cloud layer and a subcloud layer. Time markers indicated by S1-S6 will be used in Fig. 1.12.

How to estimate PBL height?

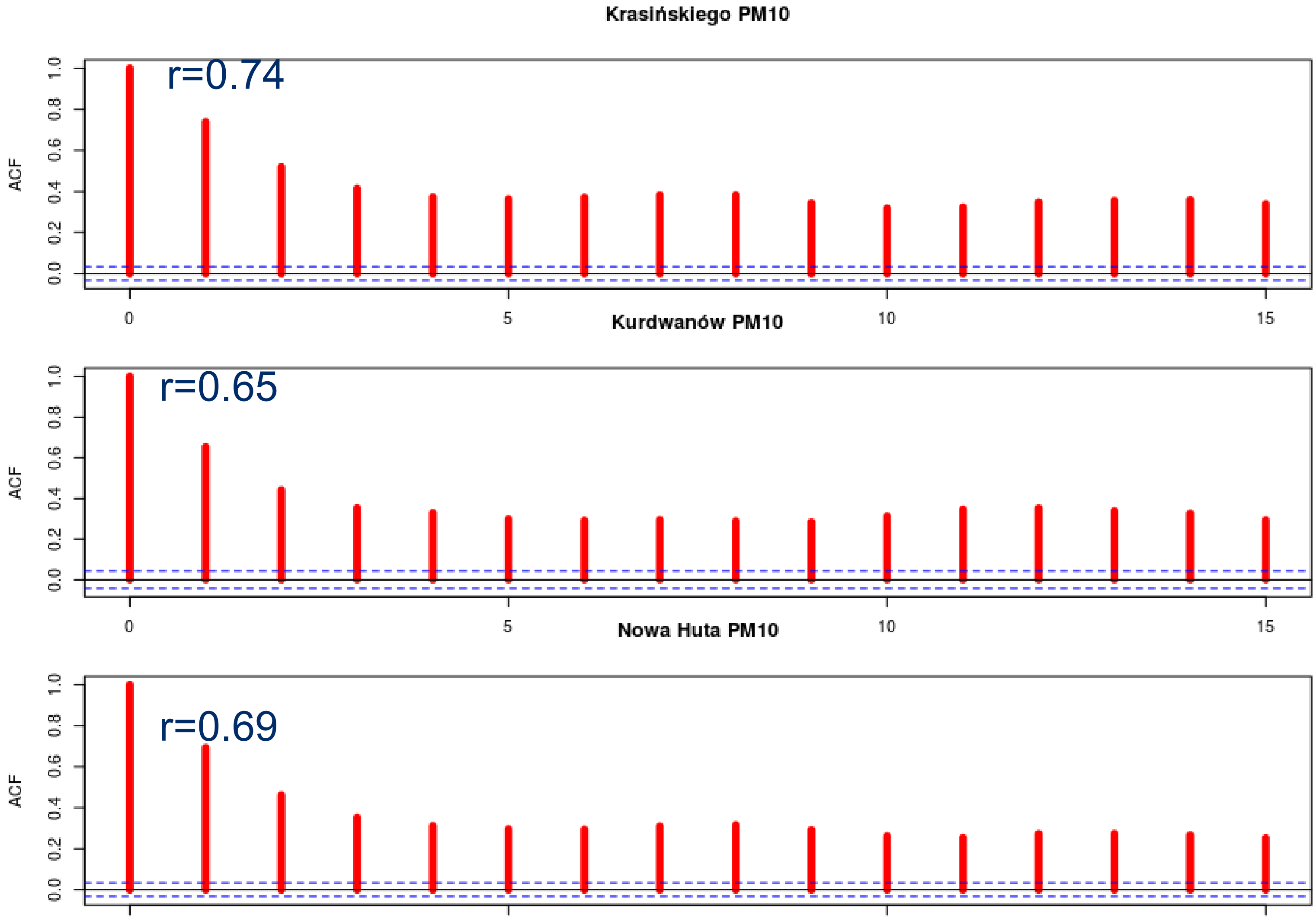
- **Atmospheric radiosonde data:** 00 and 12 UTC, 3 points in PL
- **... or sodar soundings**

Nearest grid of ERA-Interim reanalysis data [ncdf4 package]

> annual correlation r= -0.59



PMx autocorrelation



Part 3

modeling / forecasting

Air quality forecasting methods

Dynamical (deterministic, numerical)

- Use law of physics to solve transformation of atmospheric features in time
- Based on emission data inventory and meteorological forecasting simulations

- **Pros:**
  - Forecast continous in space
  - Unrelated to historical air quality data

- **Cons:**
  - Extremely expensive to run even in comparison to classical NWP
  - Strongly related to parametrization features
  - Inherit errors from NWP models

- e.g.: WRF-Chem, GEM, Calpuff/Calmet, Gaussian plume models

Statistical

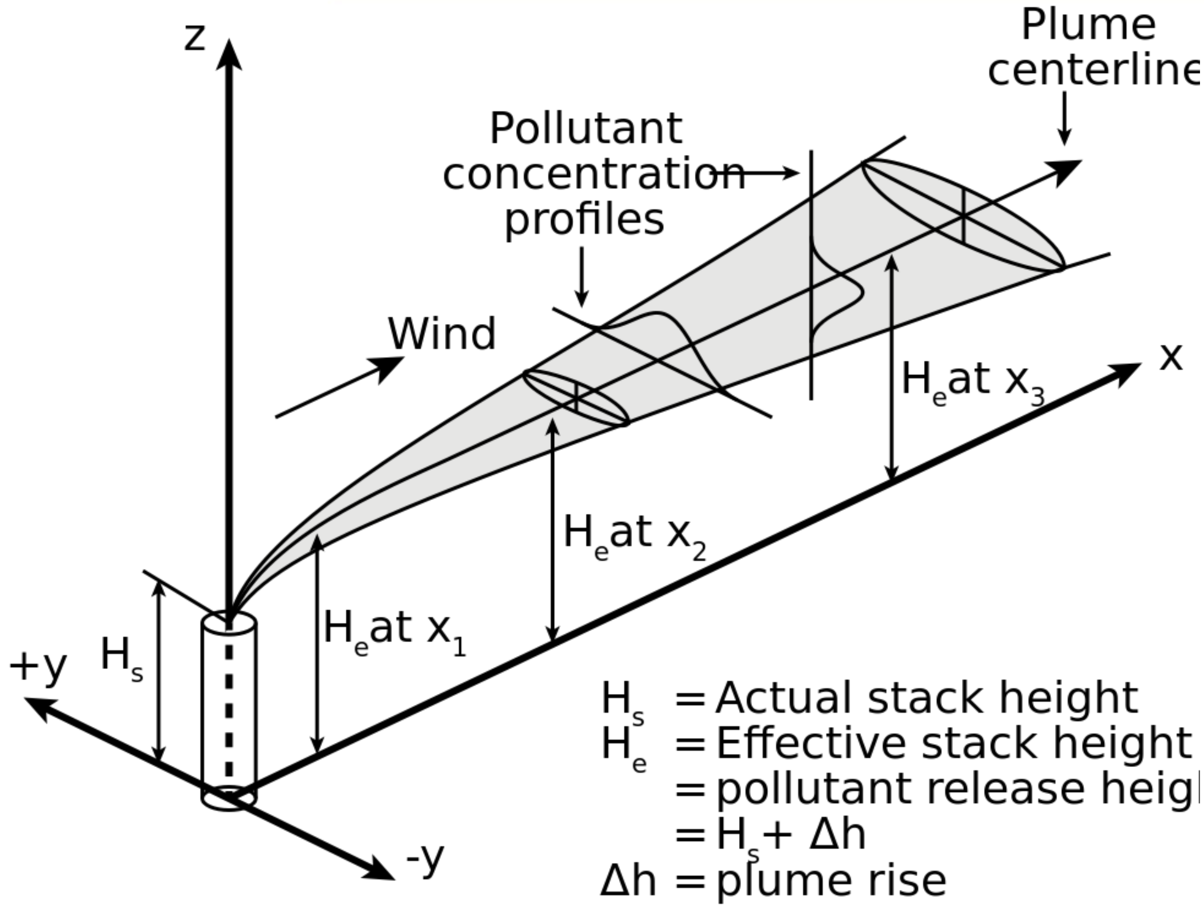
- Base on historical relationships between air pollution measurements and emissions

- **Pros:**
  - Computationally very cheap
  - Might be well fitted to end-user needs

- **Cons:**
  - Only for points with historical measurement
  - Variance inflation
  - Does not assume changes in emission from sources

- e.g: ARMAX, ARIMA, Neural Nets, RF

Gaussian plume model in R

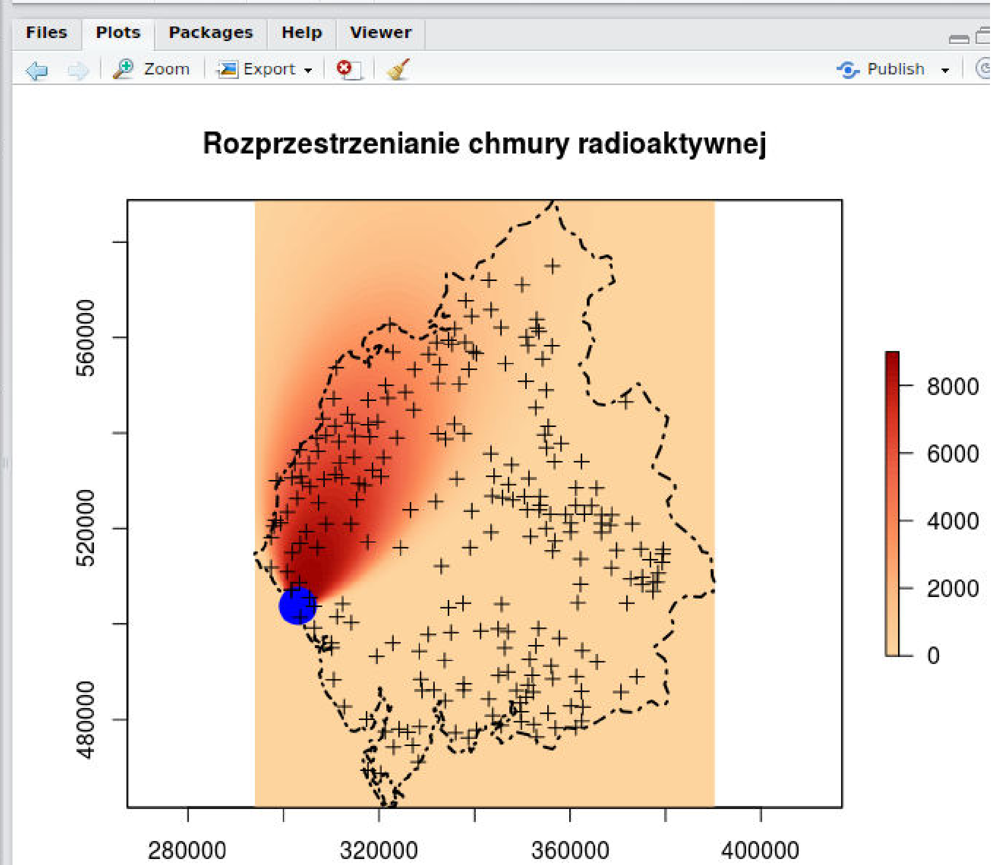


$$C(x,y,z)=\frac{E}{2\pi\sigma_y\sigma_z u}\exp\left(-\frac{y^2}{2\sigma_y^2}\right)\left[\exp\left(-\frac{(z-H_e)^2}{2\sigma_z^2}\right)+\exp\left(-\frac{(z+H_e)^2}{2\sigma_z^2}\right)\right]$$

Gaussian plume model in R

-base R coupled with GIS packages

```r
sp.theme(set = TRUE, regions = list(col = kolory))

# Zastosujemy w obliczeniach prosty model dyspersji zanieczyszczen
# Calkowita koncentracja zanieczyszczeń w dystansie 'd' i dla kąta θ z
punktowego źródła zanieczyszczeń
# dana jest poniższym wzorem:

#f=α*exp(-(d/β)2e-^(κ*cos(θ-ϕ)^2))

# gdzie:
# α - poziom zanieczyszczeń w punkcie emisji
# β - współczynnik odległości
# ϕ - dominujący kierunek smugi zanieczyszczeń
# κ - ekscentryczność (mimośrodowość)


# zapiszmy teraz powyższy wzór w postaci funkcji programistycznej:

pfunc <- function(d2, ang, a, b, k, phi)
{  a * exp(-(d2*exp(-k*cos(ang-phi))^2/b^2))  }

# a następnie zastosujmy ją do obliczeń w regularnej siatce punktów
plume <- function(src,dst,a,b,k,phi)
{  src = coordinates(src)
dst = coordinates(dst)
d2 = (dst[,1]-src[,1])^2 + (dst[,2]-src[,2])^2
ang = atan2(dst[,2]-src[,2],dst[,1]-src[,1])
pfunc(d2,ang,a,b,k,phi) }


# modyfikacja ukladu wspolrzednych do postaci kompatybilnej z naszymi
wzorami fizycznymi (zapisanymi w układzie SI)
toOSGB <- function(s){spTransform(s,CRS("+init=epsg:27700"))}
cumbria <- toOSGB(readOGR(dsn="cumbria.shp",layer="cumbria",verbose=FALSE))
# warstwa GIS: granice administracyjne
```

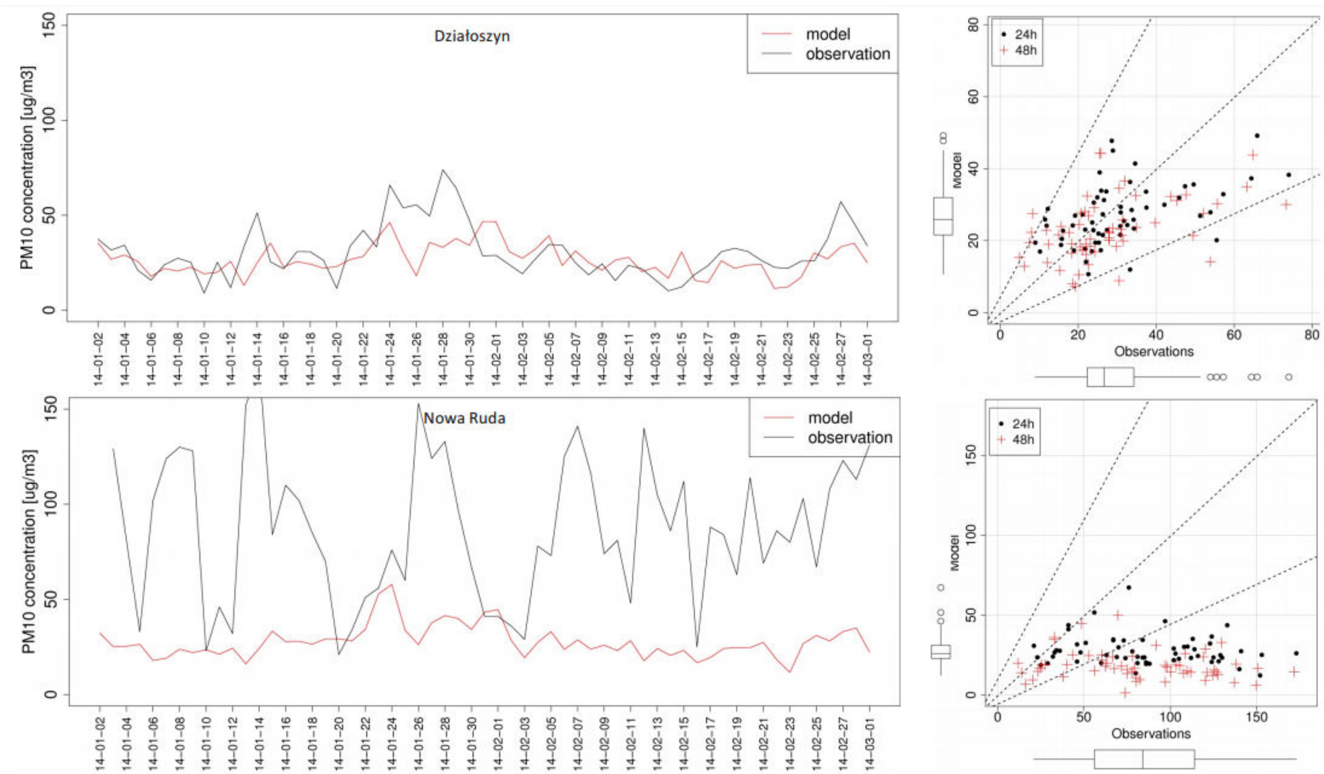


Looking for forecasting solution...

Gaussian plume model is too simplistic...

but...

Dynamical models are also sometimes far from being perfect...



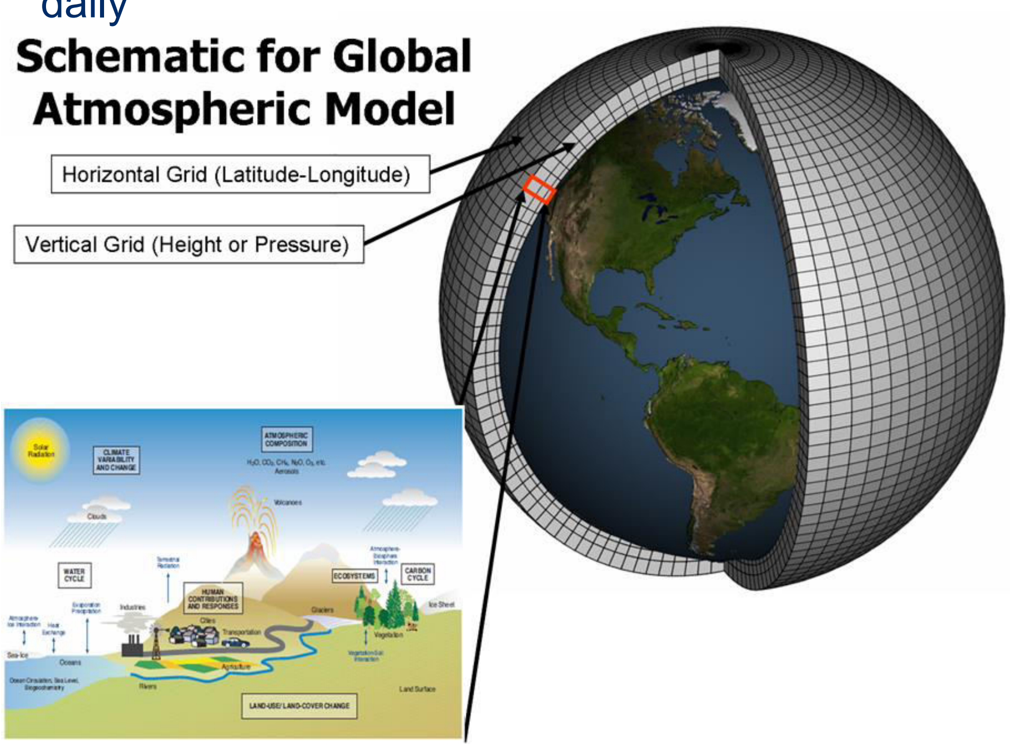
Werner M., Kryza M., Ojrzyńska H., Skjoth C., Wałaszek K., Dore A. (2016)
http://eprints.worc.ac.uk/3356/1/PM10_forecast_HARMO.pdf

Idea:

- Create hybride type of model based on:
(1) NWP forecasting systems
(2) machine learning algorithms

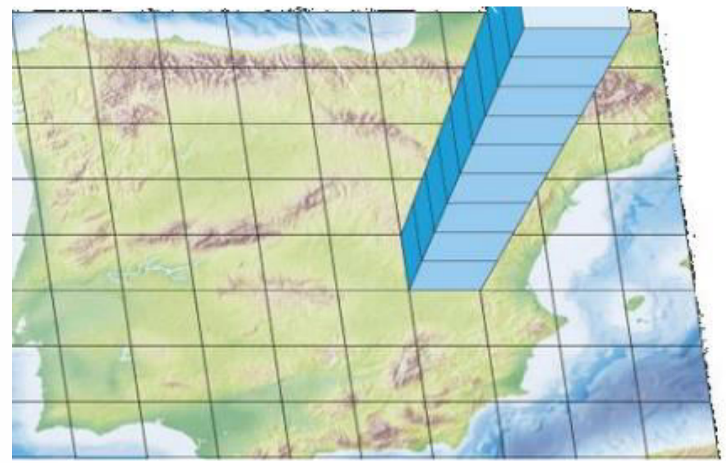
Dynamical downscaling of meteorological forecast

Global Forecasting System (**GFS**) [0.25x0.25 deg. grid, 64 vertical levels] x 4 times daily



no. of grids = 1440*1440*64
> 6 PetaFlops required

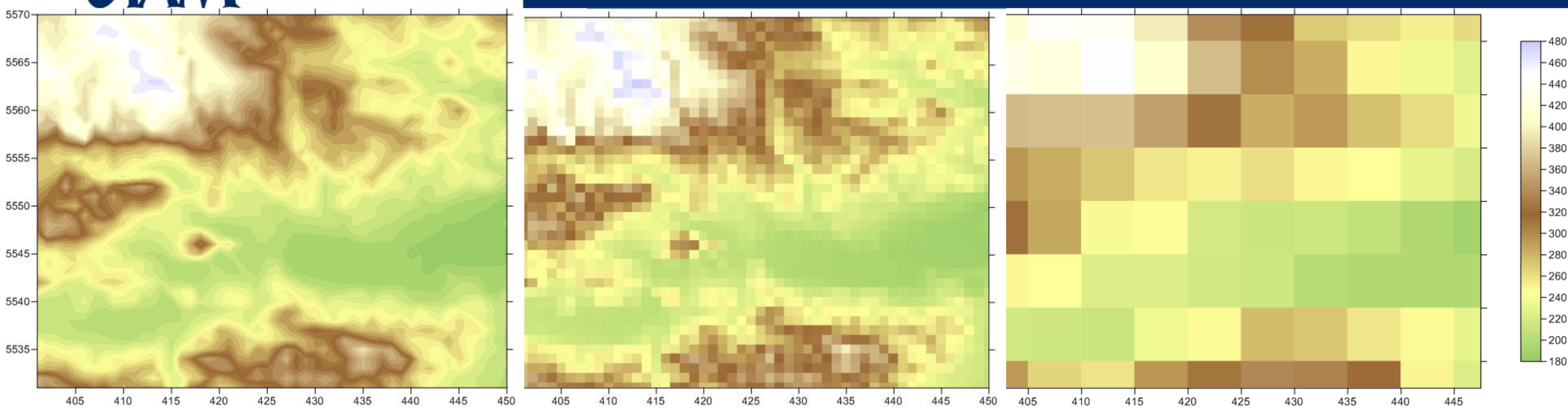
Limited Area model [LAM] – Weather Research and Forecasting Model (**WRF**)



5 km horizontal grid resolution
45 vertical levels

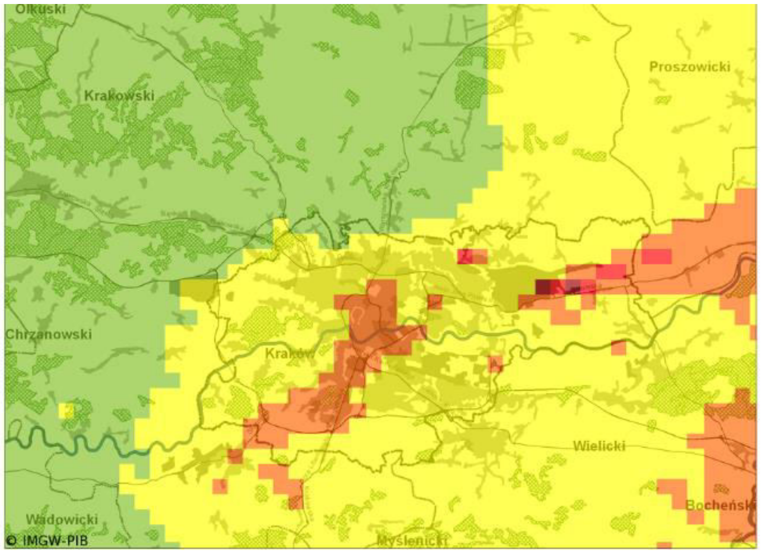
Intel Skylake Xeons CPU based cluster (28 cores per node + IB)
- available in PCSS

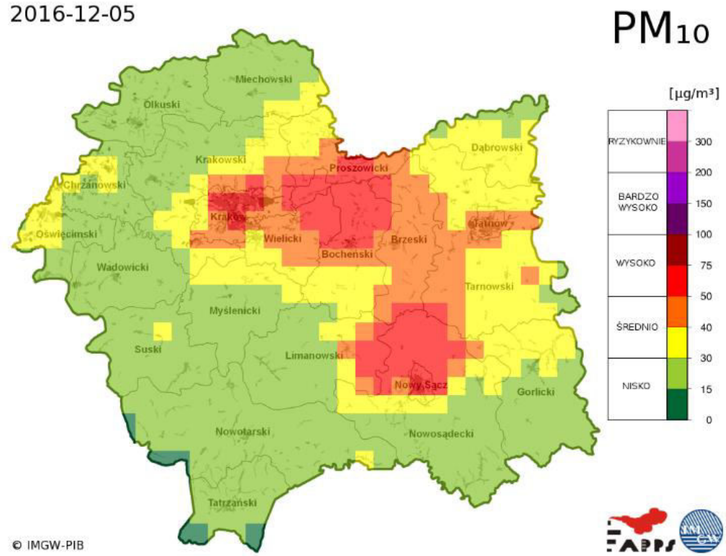
The importance of dynamical model's resolution



90 m (SRTM) 1 km (forecast for Kraków) 5 km (forecast for Małopolska)

Courtesy of IMGW-PIB
(www.smog.imgw.pl)





Machine learning algorithms

Three types of data sources used as predictors (for daily means):

1) Meteorological predictors derived from WRF model:

- Air temperature
- PBL height
- Wind speed and direction
- Cloud cover

2) Human activity based on „calendar"

- Day of week
- Working day/day off (R package: ‚wolnedni')

3) Autocorrelation

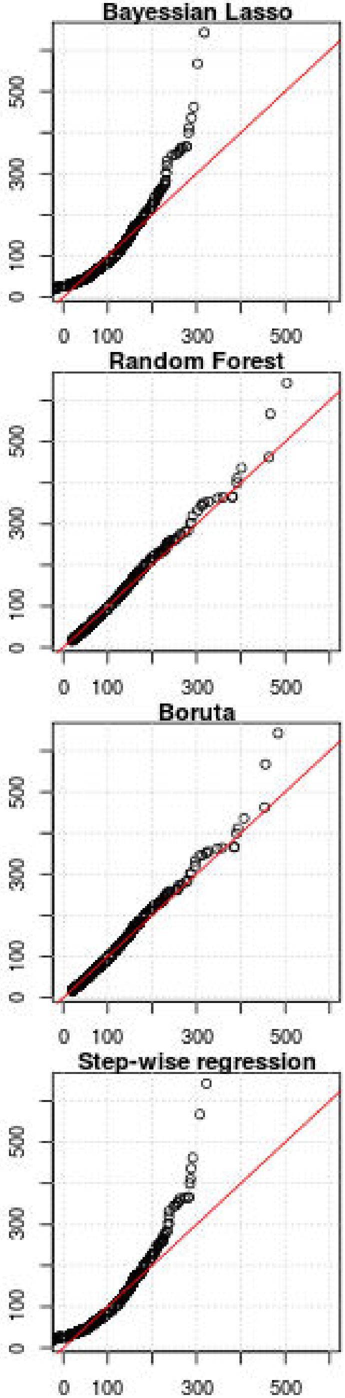
- 1 day lag

Machine learning algorithms
(caret package)

1) Models:
- Stepwise regression with AIC
- Random Forest
- Boruta
- Bayessian Lasso

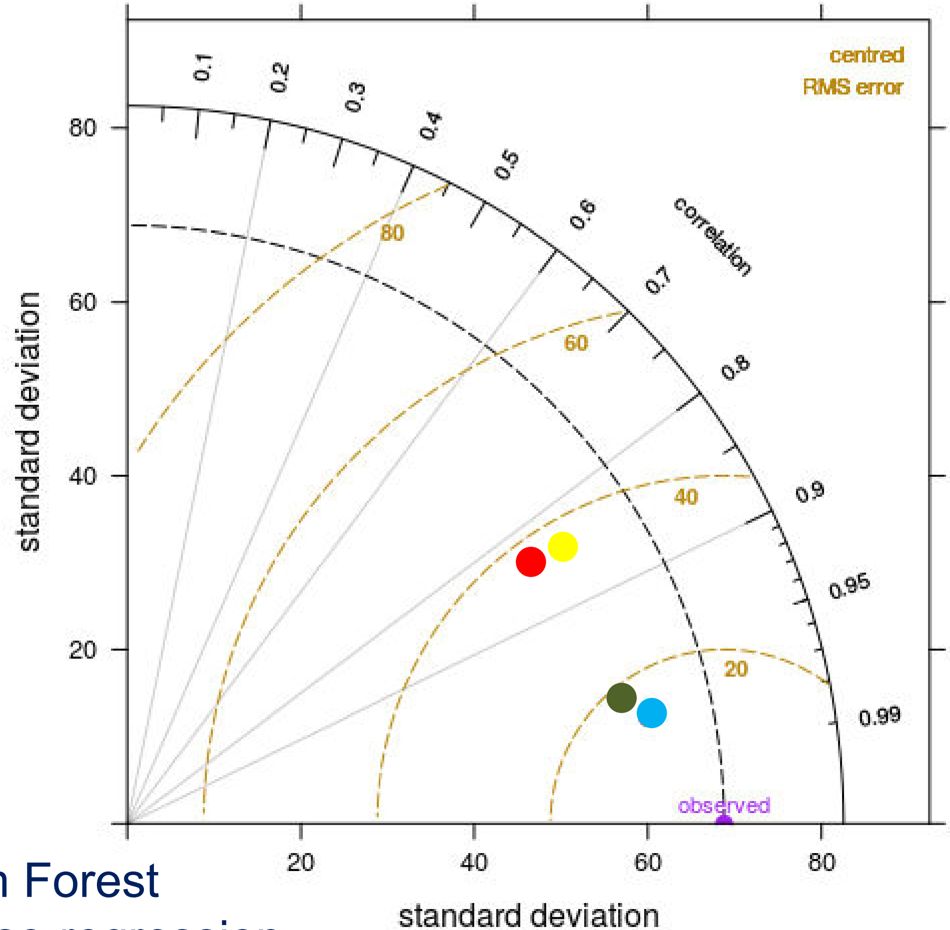
2) Preprocessing
- scale
- center
- YeoJohnson

3) Cross-validation
- K-fold (10-folds, repeated 10-times)



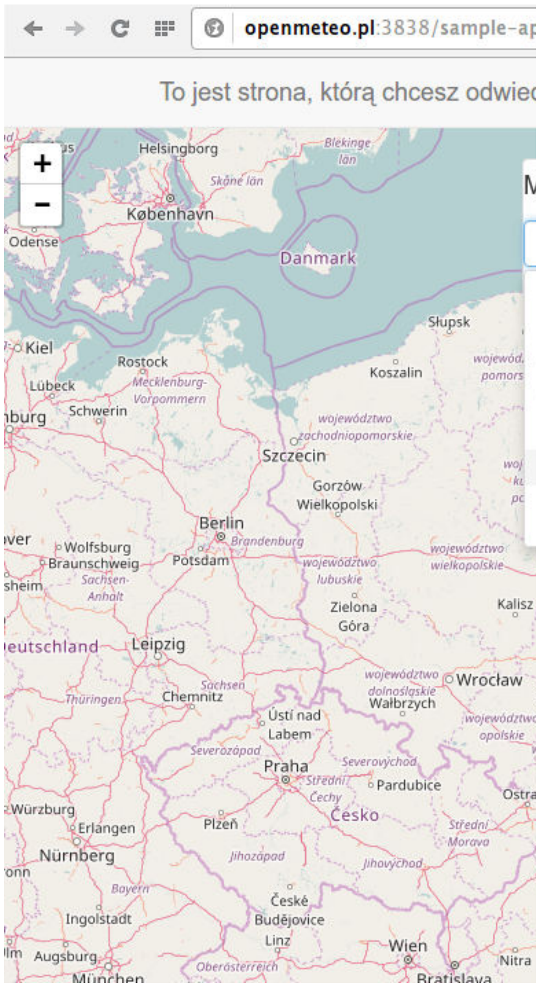
Verification

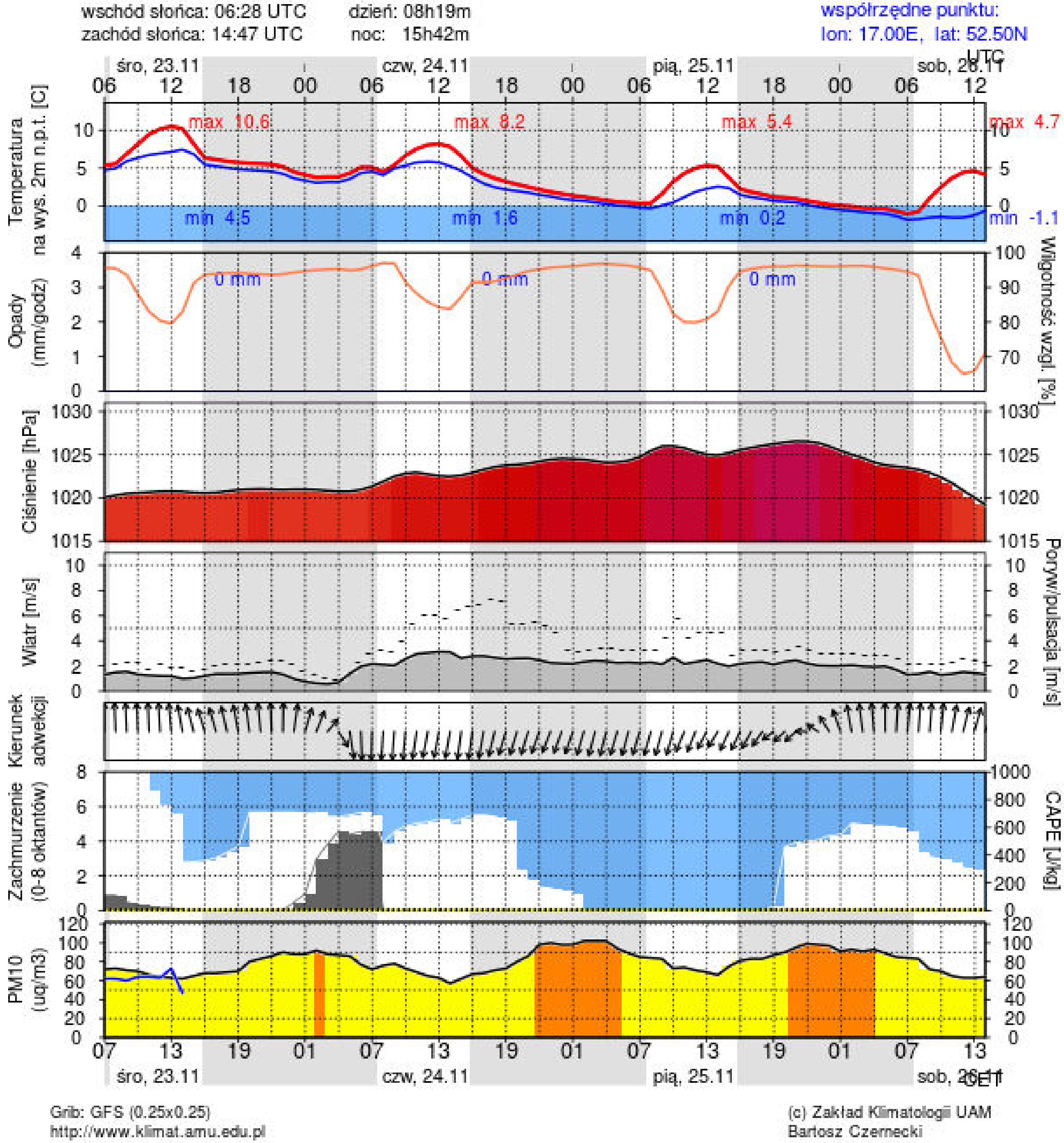
(„verification” and „plotrix” packages)



Legend:

- Blue: Boruta
- Olive: Random Forest
- Yellow:step-wise regression
- Red: Bayessian Lasso


openmeteo.pl:3838/sample-ap
To jest strona, którą chcesz odwie
Helsingborg
København
Danmark
Odense
Kiel
Rostock
Lübeck
Schwerin
Słupsk
Koszalin
Szczecin
Gorzów Wielkopolski
Berlin
Potsdam
Wolfsburg
Braunschweig
Zielona Góra
Kalisz
Leipzig
Wrocław
Wałbrzych
Chemnitz
Praha
Pardubice
Česko
Würzburg
Erlangen
Nürnberg
Plzeň
Ingolstadt
České Budějovice
Augsburg
Linz
Wien
Bratislava
München


wschód słońca: 06:28 UTC
zachód słońca: 14:47 UTC
dzień: 08h19m
noc: 15h42m
współrzędne punktu:
lon: 17.00E, lat: 52.50N
UTC
śro, 23.11
czw, 24.11
pią, 25.11
sob, 26.11
Temperatura na wys. 2m n.p.t. [C]
max 10.6
max 8.2
max 5.4
max 4.7
min 4.5
min 1.6
min 0.2
min -1.1
Opady (mm/godz)
0 mm
0 mm
0 mm
Wilgotność wzgl. [%]
Ciśnienie [hPa]
Wiatr [m/s]
Porywy/pulsacja [m/s]
Kierunek adwekcji
Zachmurzenie (0-8 oktantów)
CAPE [J/kg]
PM10 (ug/m3)
CET
Grib: GFS (0.25x0.25)
http://www.klimat.amu.edu.pl
(c) Zakład Klimatologii UAM
Bartosz Czernecki

Plans for a nearest future:

1) Models:

- Account for a wet deposition -> historical precipitation available as 12 hours totals (hindcast simulation required)
- Bias correction of NWP models based on hindcast simulation
- Testing for neural networks (outside R :( )
- Confidence intervals based on ensemble models

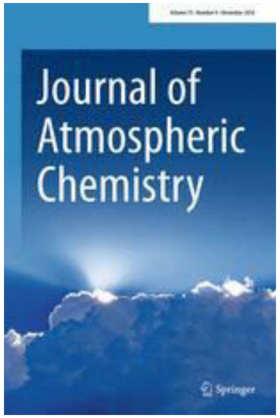
2) Web interface

- Finishing shiny + leaflet layout
- Integrating with other modules (convection [thunderstorm/tornadoes] and pollen)

3) Mobile app

- ...

4) Cooperation



Czernecki B., Półrolniczak M., Kolendowicz, L. et al.
(2016) *Influence of the atmospheric conditions on PM10 concentrations in Poznań, Poland*
doi:10.1007/s10874-016-9345-5

Thank you for your attention

Part of computations were carried out in the Poznań Centre for Networking and Supercomputing
(http://hpc.man.poznan.pl),
Grant No. 295.