



Instytut Meteorologii i Gospodarki Wodnej
Państwowy Instytut Badawczy

Machine Learning in Climatological Analysis and Operational Forecast in Poland,

Bogdan Bochenek
Dział Prognoz Numerycznych ALADIN
IMGW-PIB

Zakopane, 08.03.2019



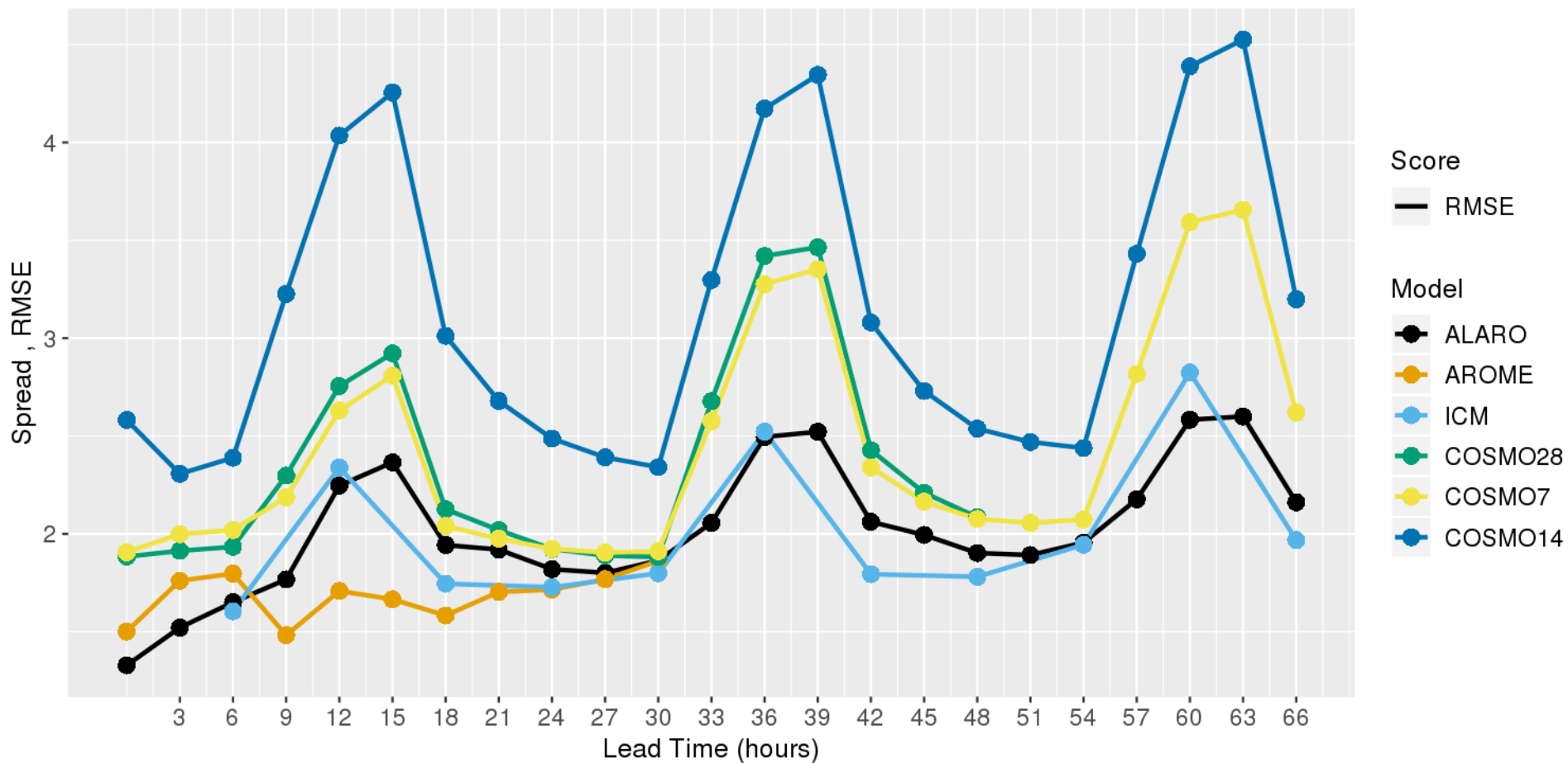
Instytut Meteorologii i Gospodarki Wodnej

Państwowy Instytut Badawczy

Spread & Skill(RMSE) : T2m

Verification Period: 2019020500-2019030700

Stations: ALL





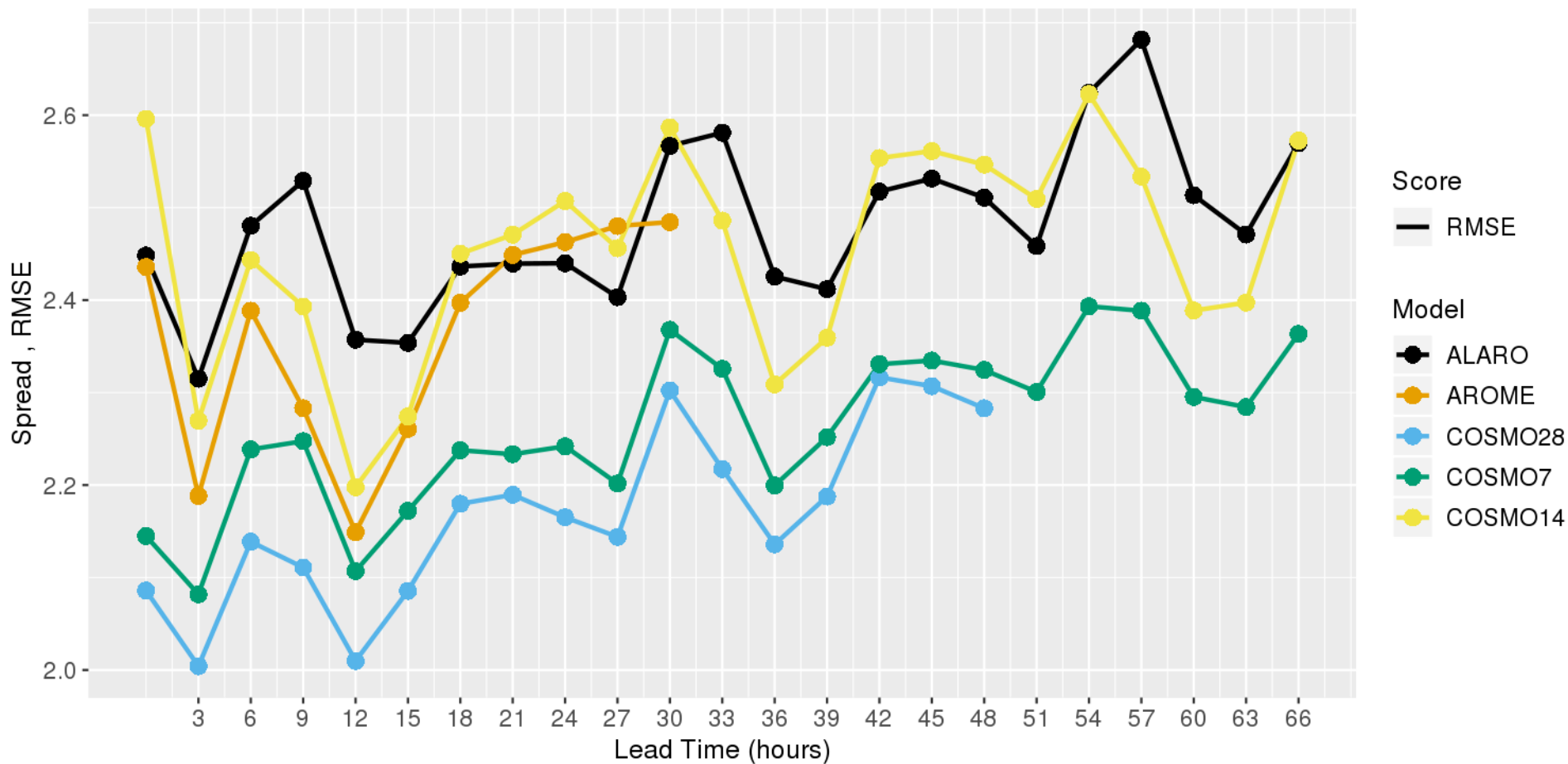
Instytut Meteorologii i Gospodarki Wodnej

Państwowy Instytut Badawczy

Spread & Skill(RMSE) : S10m

Verification Period: 2019020500-2019030700

Stations: ALL





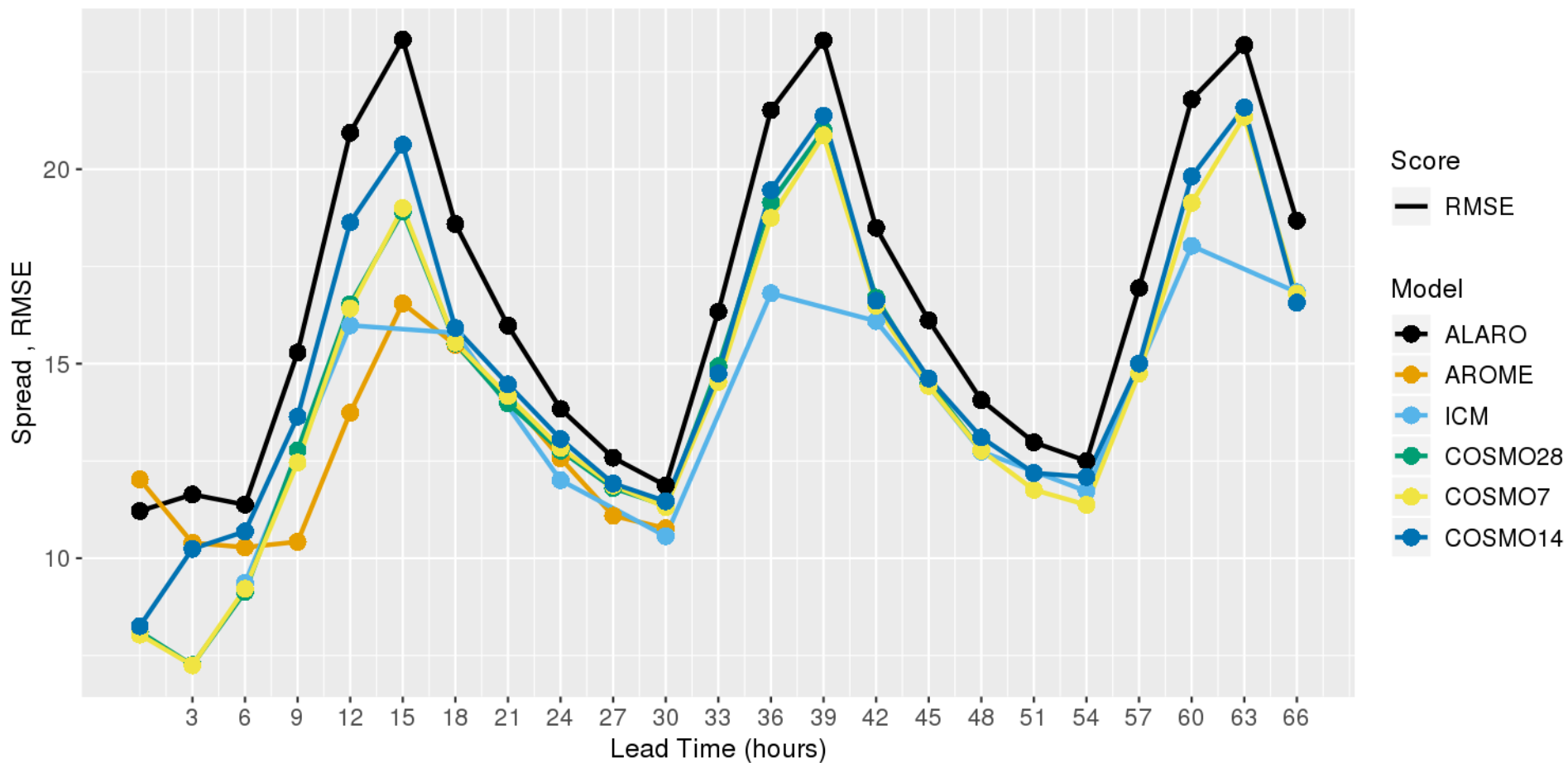
Instytut Meteorologii i Gospodarki Wodnej

Państwowy Instytut Badawczy

Spread & Skill(RMSE) : RH2m

Verification Period: 2019020500-2019030700

Stations: ALL





Instytut Meteorologii i Gospodarki Wodnej

Państwowy Instytut Badawczy

Dane:

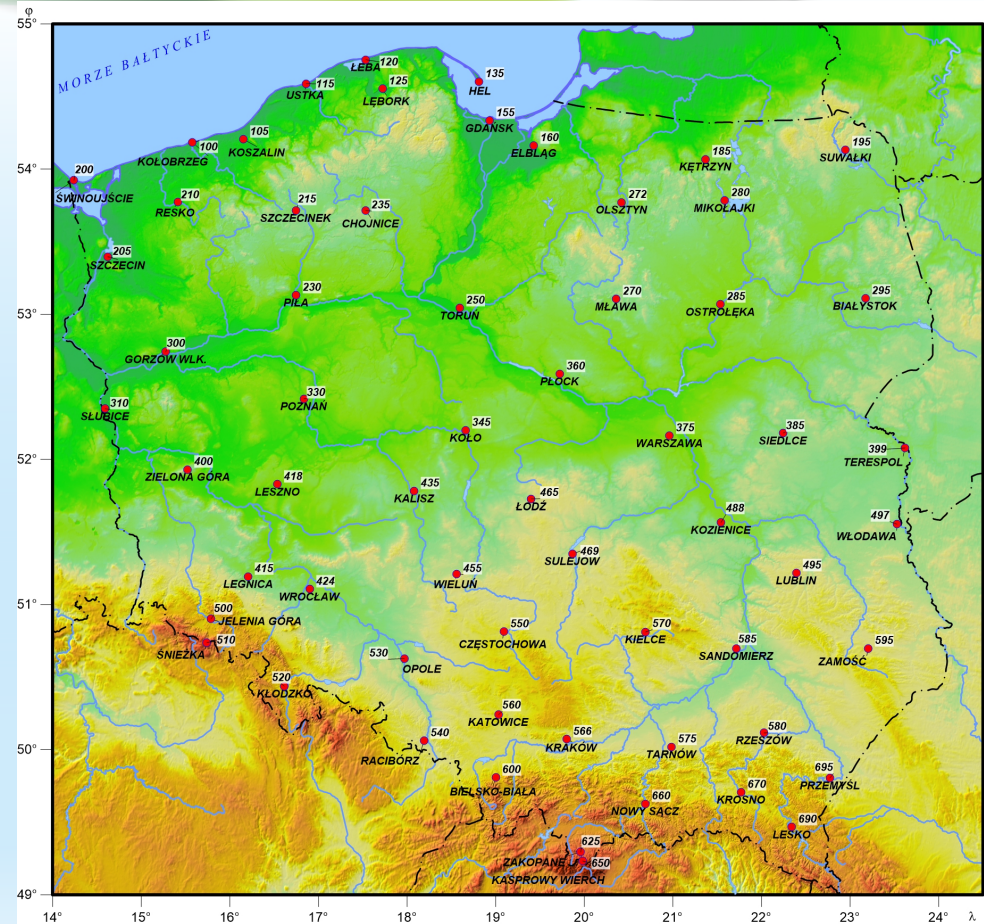
1) Dane obserwacyjne:

a) Tradycyjne (in situ), 1971 – 2015

Sieć stacji meteorologicznych
IMGW-PIB: pomiary temperatury,
wilgotności, prędkości i kierunku wiatru,
ciśnienia atmosferycznego,
zachmurzenia, opadu, zjawisk
atmosferycznych

b) Teledetekcyjne, 2006 - 2015

Dane satelitarne
Dane radarowe





Instytut Meteorologii i Gospodarki Wodnej

Państwowy Instytut Badawczy

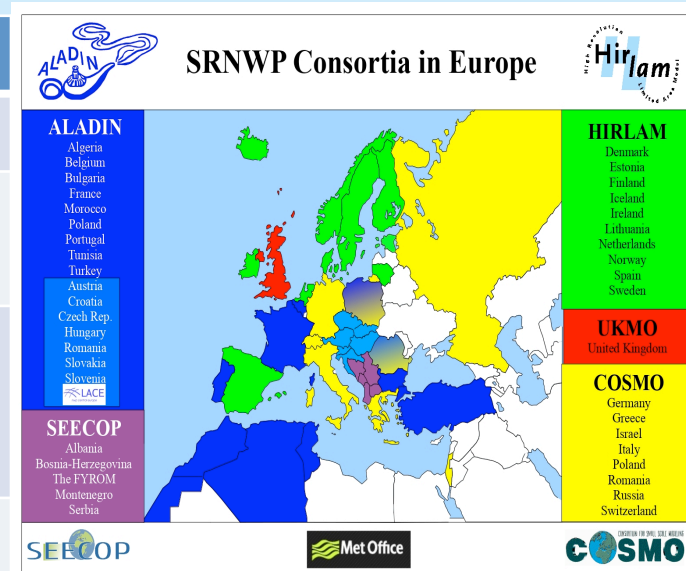
Dane:

2) Dane modelowe:

Wyniki modeli numerycznych pogody konsorcjum ALADIN – wyniki modelowanie meteorologicznego (2006-2015) oraz RCM (1971-2010) wraz ze scenariuszami klimatycznymi (2011-2030)

Reanalizy konsorcjum ECMWF – Era-Interim i ERA5 (1979-2015)

	Era-Interim	Era5
Zakres	1979 -	1950 -
Wersja modelu IFS	31 (2006 r.)	41 (2016 r.)
Rozdzielczość	79 km/ 60 poziomów (0.1 hPa)	31 km/ 137 poziomów (0.01 hPa)
Dane	Co 6 godzin	Co 1 godzinę





Dane:

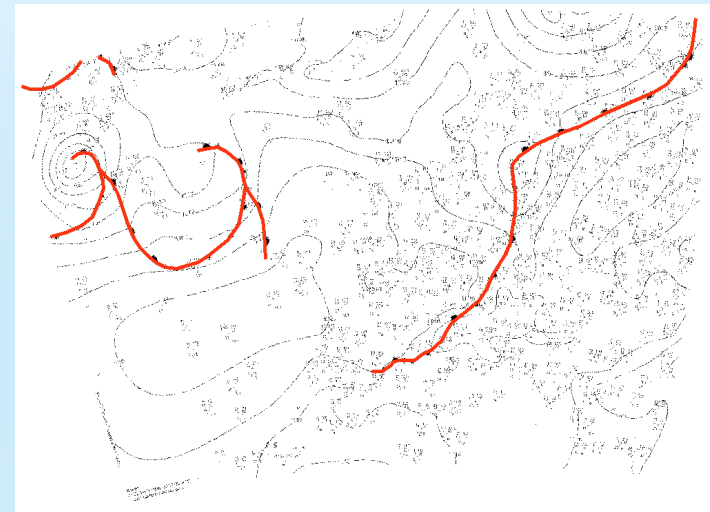
3) Pozostałe dane

Cyfrowy model terenu (DTM i SRTM)

Pokrycie terenu CORINE (CLC)

Kalendarz typów cyrkulacji atmosfery dla Polski
południowej (T. Niedźwiedź 2017)

Fronty DWD (digitalizacja)



Fronty - digitalizacja

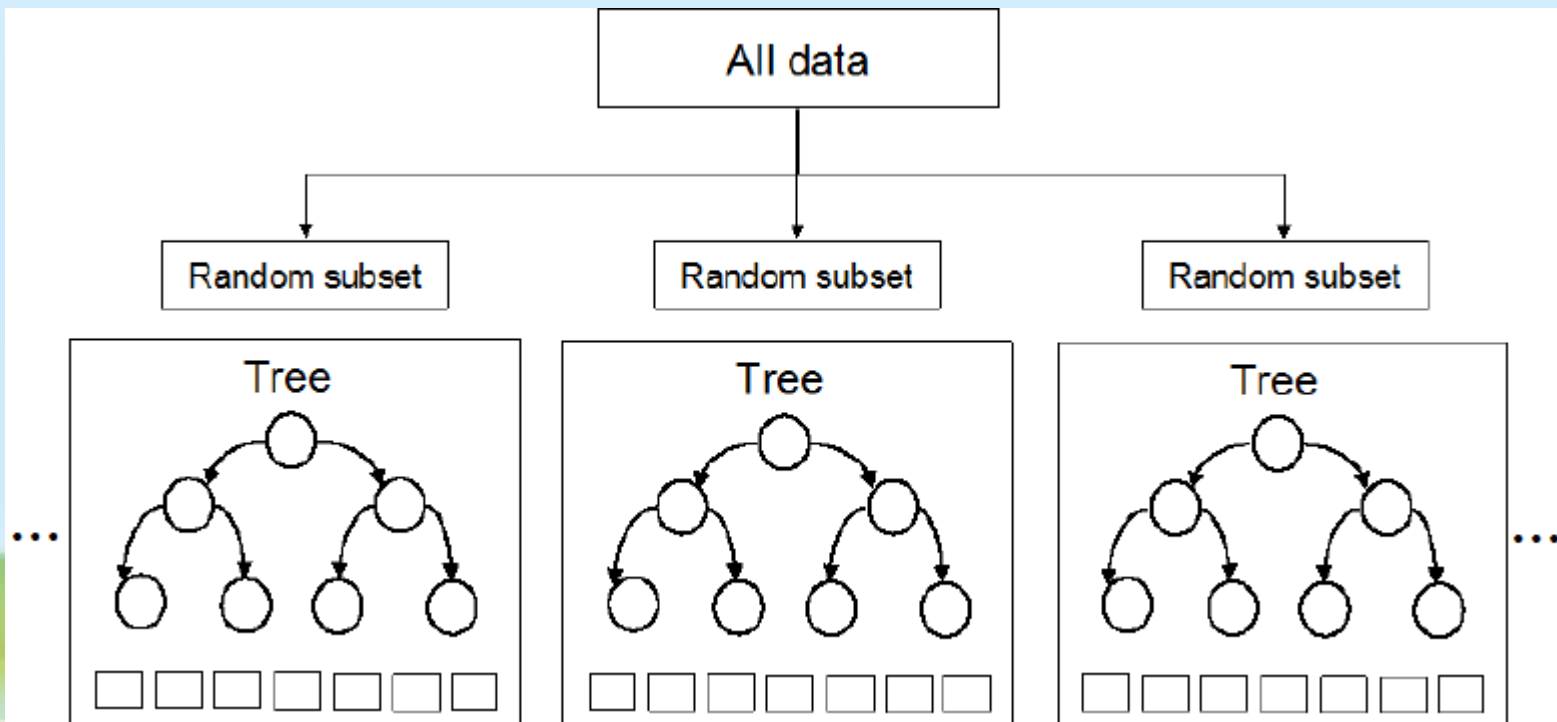


- 1) Powiązanie wyników modelowych z danymi obserwacyjnymi,
- 2) Ocena wyników z uwzględnieniem informacji o pokryciu terenu i warunkach cyrkulacyjnych,
- 3) Wybór zmiennych, które w największym stopniu wpływają na poprawę.
- 4) Badanie i ocena działania systemu – w jakim stopniu poprawia jakość danych modelowych, zastosowanie standardowych metod weryfikacyjnych: RMSE, BIAS, tablice wielodzielcze, itd.

Metody nadzorowane:

Random Forest (Losowy Las):

Budowanie zestawu losowych drzew decyzyjnych i wyłaniania na tej podstawie optymalnych zależności pomiędzy parametrami uczenia się



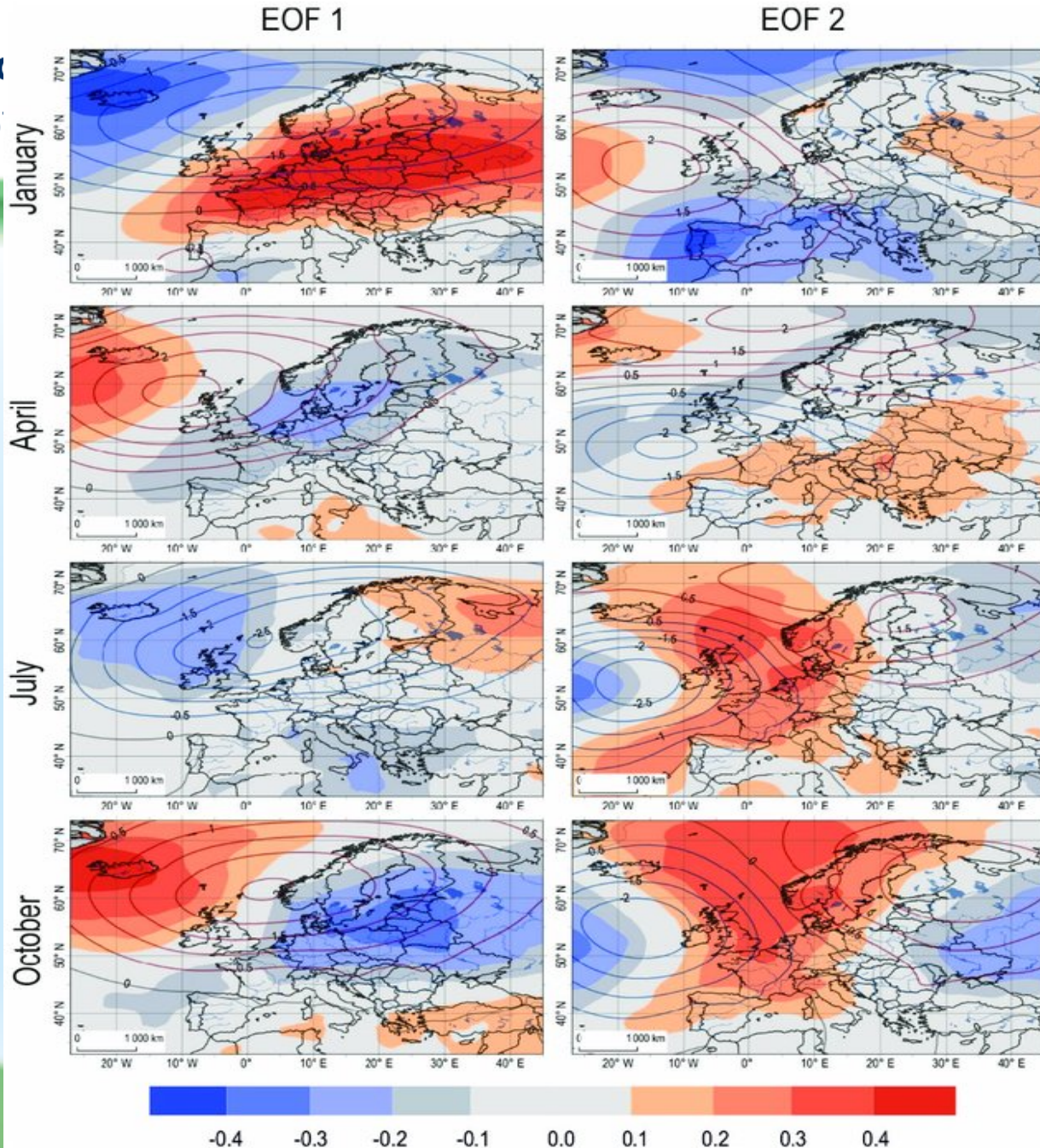


Instytut Meteorologii
Państwowy Instytut

Metody nienadzorowane:

Empirical Orthogonal
Functions (EOF):

Grupowanie danych według
wiodących wzorców



(A. Wypych et al., 2018)



Wstępne wyniki:

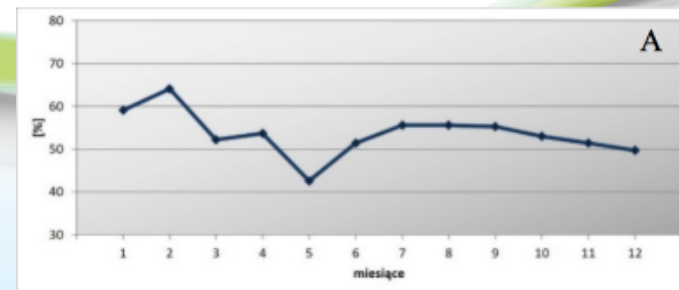
Automatyczne wyznaczanie mas powietrza

Uczenie maszynowe wykorzystujące metodę Random Forest oraz dane z reanaliz Era-Interim wraz z danymi z kalendarza mas powietrza oraz danymi obserwacyjnymi ze stacji synoptycznych IMGW-PIB.

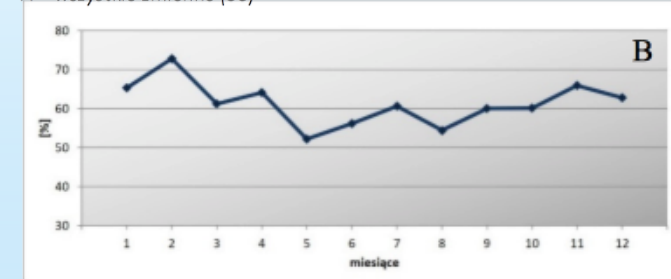
Parametry uczące:

P – (SLP)
WS – prędkość wiatru
TSR – temperatura średnia dobową
TN – temperatura minimalna
TX – temperatura maksymalna
TA – dobowa amplituda temperatury
TGR – temperatura gruntu
E – prężność pary wodnej
TD – temperatura punktu rosy
RH – wilgotność względna powietrza
DPD – niedosyt punktu rosu
RR – dobową sumą opadu
PS – grubość pokrywy śnieżnej
N – wielkość zachmurzenia
CL_CM – wysokość podstawy chmur CL_CM
SD – uśłonecznienie
VIS – widzialność

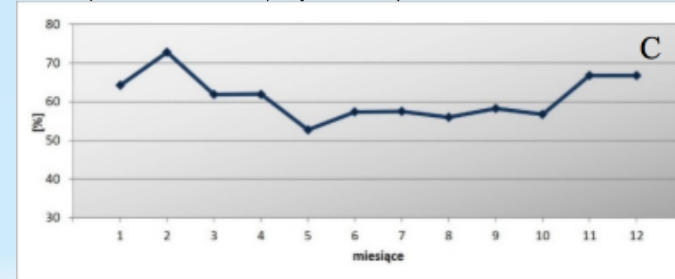
Wyniki jakości testów:



A – wszystkie zmienne (33)



B – wszystkie zmienne oraz połączone masy PPms i PPmc



C – 16 zmiennych oraz połączone masy PPms i PPmc



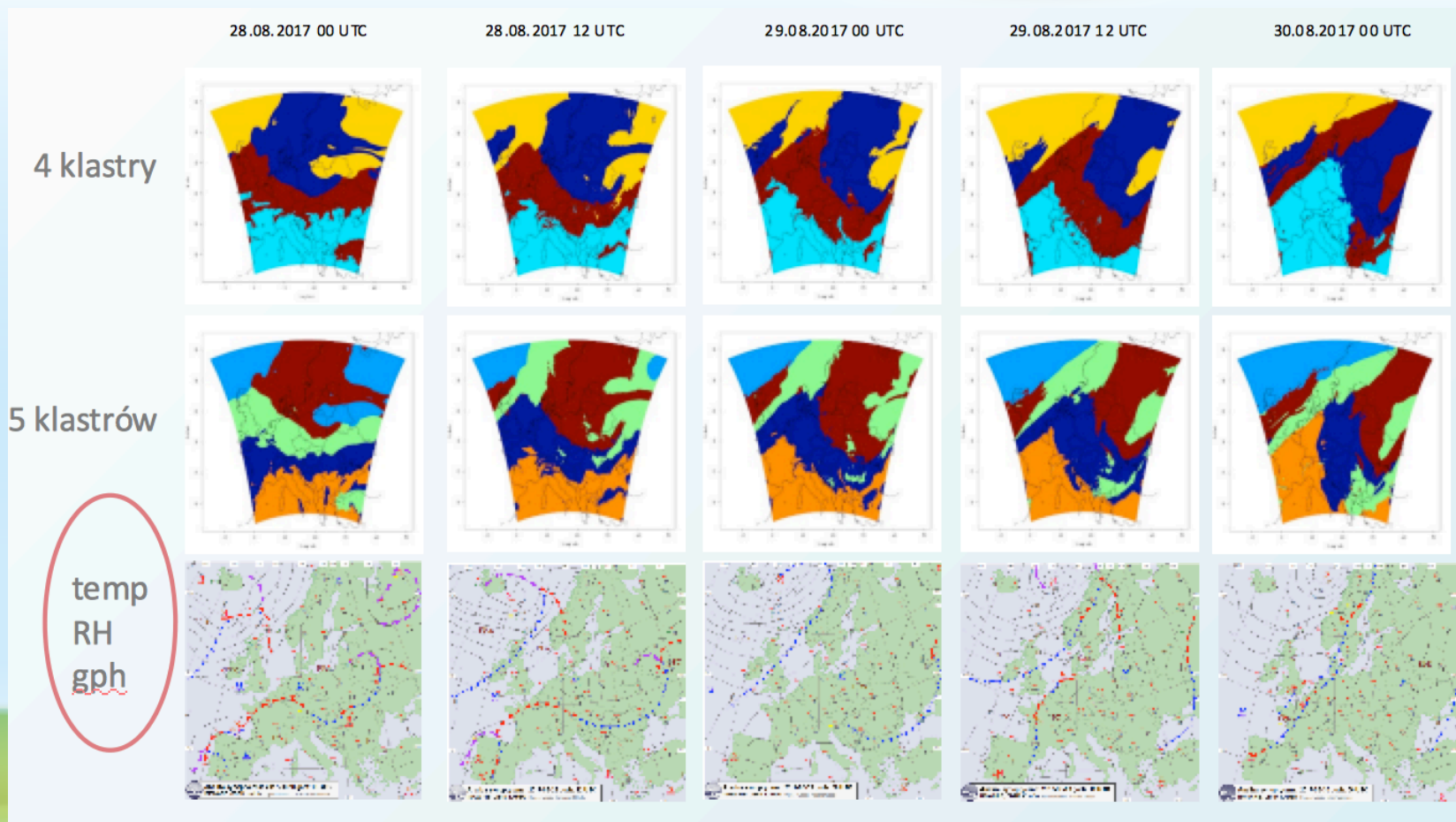
Instytut Meteorologii i Gospodarki Wodnej

Państwowy Instytut Badawczy

Wstępne wyniki:

Automatyczne wyznaczanie mas powietrza

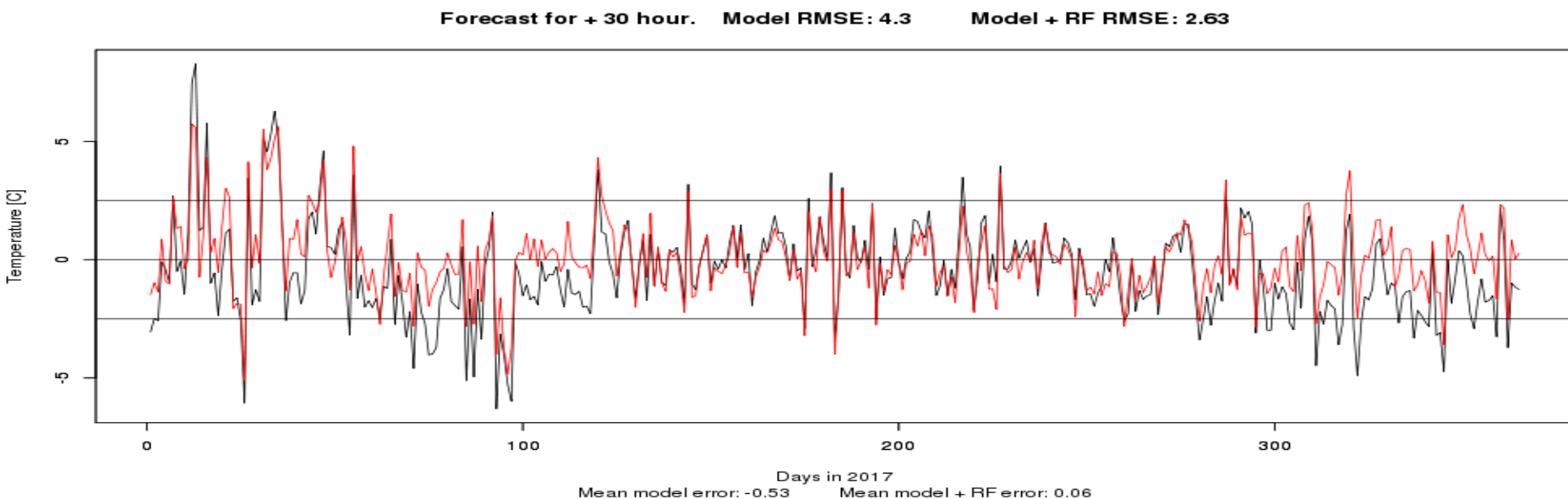
Parametry uczące
(poziomy 1000 – 850 hPa):
temp
RH
gph





Wstępne wyniki:

Poprawa wyników mezoskalowych modeli numerycznych



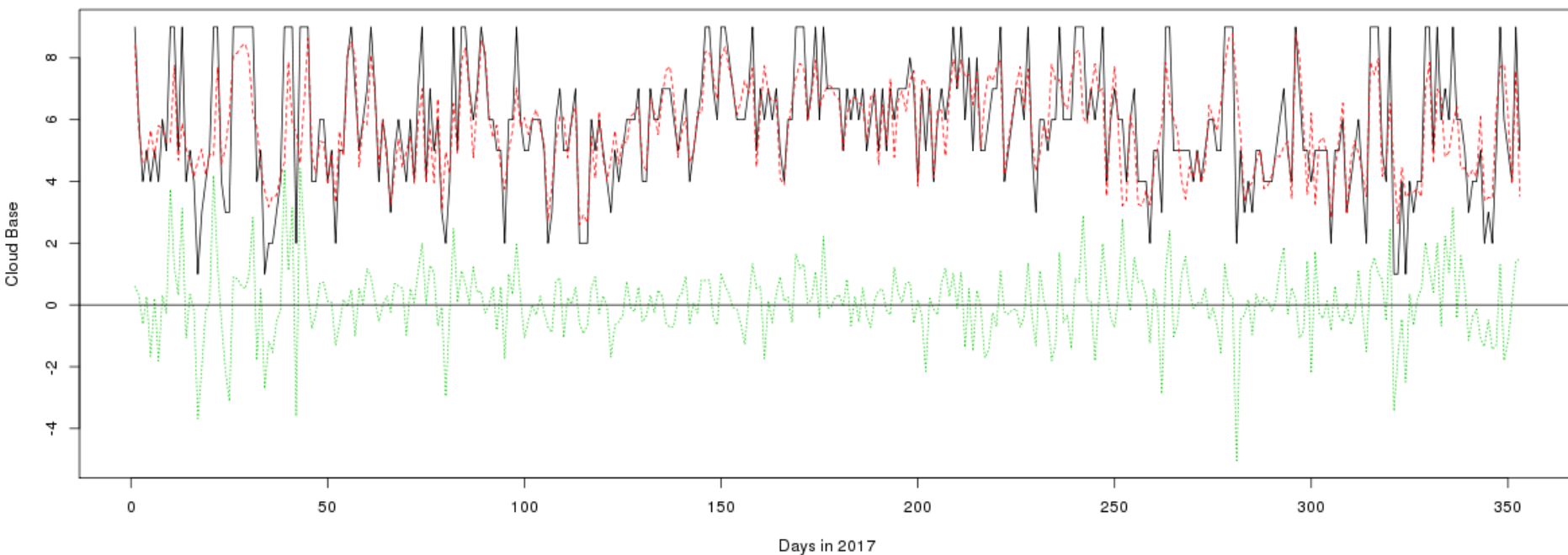
Temperatura na wysokości 2 m, Kraków 30 godzina prognozy, 2017 rok
Na czarno błąd prognozy modelu ALARO, na czerwono błąd systemu opartego
o sieci neuronowe (uczenie się 2015-2016)



Instytut Meteorologii i Gospodarki Wodnej

Państwowy Instytut Badawczy

Wstępne wyniki:



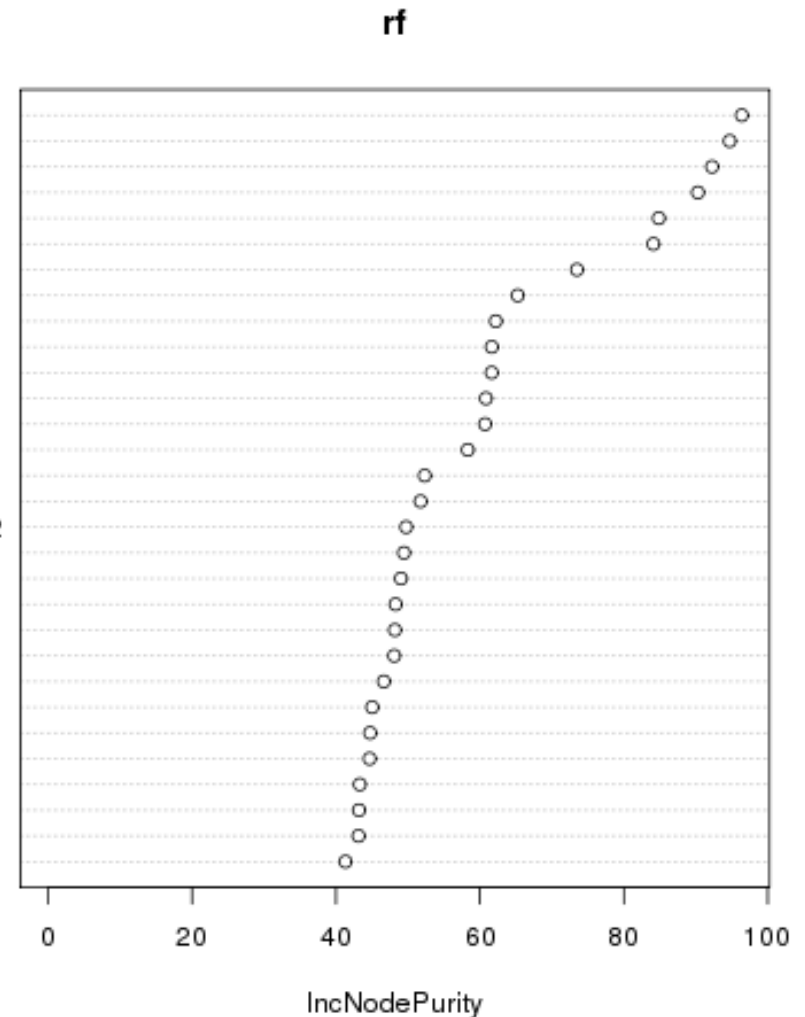
Wysokość podstawy chmur (synop 1-9), lotnisko Kraków-Balice, 2017 rok, 12 UTC
Na czarno obserwacje. Na czerwono wyniki systemu opartego o sieci neuronowe (2016 „uczenie się”) i model ALARO



Wstępne wyniki:

Wynik uczenia metodą
Random Forest
Wrocław 1980-2010
Parametry konwekcyjne
sprzyjające powstawaniu
burz

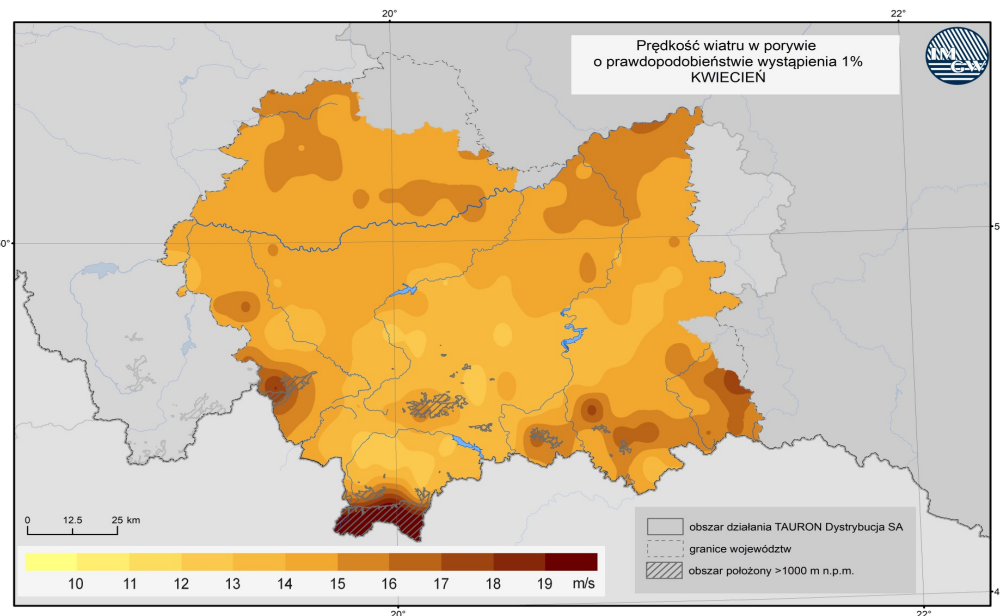
LI_12
SI_12
CAPE_12
LIFT_12
CAPEVT_12
KI_12
MLMR_12
SI
LIFT
SWEAT_12
KI
LI
TLCL_12
MLPT
CINS_12
TTI_12
BULKCAPV_12
T1000500
MLPT_12
BULK_12
MLMR
CI_12
CI
PLCL
T1000500_12
CINS
PLCL_12
VTI_12
TTI
TLCL



Wynik działania systemu opartego o metodę Random Forest
Oś y – parametr konwekcyjny i godzina pomiaru
Oś x – waga parametru

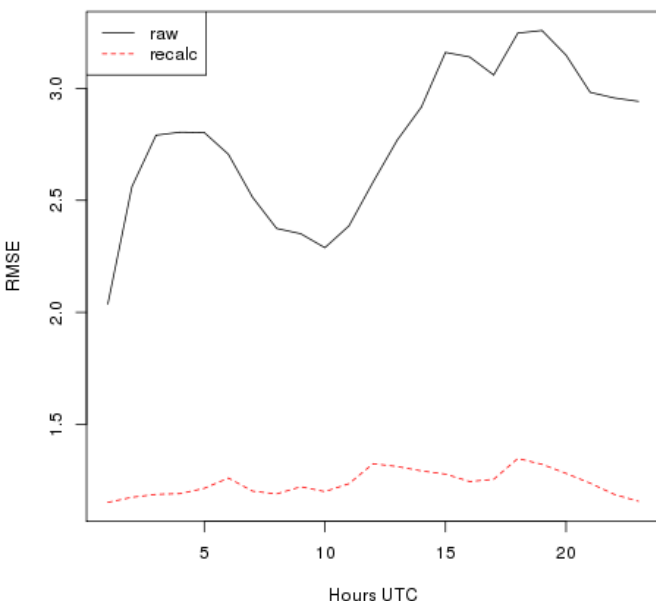
Wstępne wyniki:

Porywy wiatru – poprawa wyników modeli numerycznych (2016-2017) przy uwzględnieniu danych obserwacyjnych oraz danych DTM



Porywy wiatru - kwiecień

RMSE evol for various hours of February 2018 days



Błąd średniokwadratowy modelu AROME na czarno, błąd średniokwadratowy systemu opartego o metodę Random Forest na czerwono.



Instytut Meteorologii i Gospodarki Wodnej
Państwowy Instytut Badawczy

Dziękuję za uwagę!

Bogdan Bochenek
Dział Prognoz Numerycznych ALADIN
IMGW-PIB

Zakopane, 08.03.2019