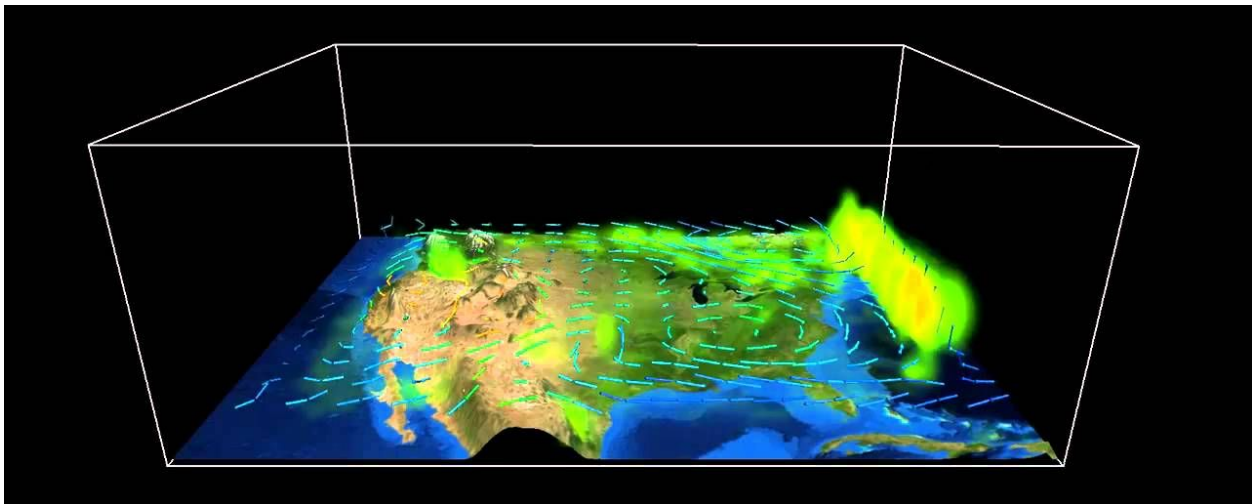


Matematyczne modelowanie procesów meteorologicznych i klimatycznych -> prognozowanie i weryfikacja modelu atmosferycznych Global Forecasting System (GFS) oraz modelowanie regresyjne

# Projekt zaliczeniowy:

## weryfikacja krótkoterminowych prognoz pogody oraz modelowanie regresyjne



---

„Remember that all models are wrong;  
the practical question is how wrong  
do they have to be to not be useful”

[George E.P. Box, 1987]

---

## 1. Wprowadzenie.

### **A: Weryfikacja archiwalnych prognoz modelu numerycznego GFS dla obszaru Polski:**

Większość znanych portali internetowych a także w dużym stopniu prognozy pogody znane z mediów bazują na publicznie dostępnych wynikach obliczeń amerykańskiego modelu GFS ([https://en.wikipedia.org/wiki/Global\\_Forecast\\_System](https://en.wikipedia.org/wiki/Global_Forecast_System)).

Oceń skuteczność prognoz pogody dla tego modelu analizując serie modelowe oraz oficjalne obserwacje meteorologiczne z wybranej stacji meteorologicznej dla temperatury powietrza i opadów atmosferycznych

1. Wybierz stację meteorologiczną dla której będzie wykonywany projekt końcowy oraz dopisz skład grupy w arkuszu na stronie:

<https://docs.google.com/spreadsheets/d/19oQkKqwZzYdt9Pd4GCgbaHXjejEvdEPaCVOJMeAa3U8/edit?usp=sharing>

Jeśli któraś ze stacji jest już przypisana do innej grupy wybierz inną lokalizację

*Uwaga: Projekt może być wykonany w dowolnym środowisku obliczeniowym (w tym także np. w arkuszu kalkulacyjnym) i powinien być oddany w formie spójnego i odpowiednio opisanego raportu. Poniższe podpowiedzi dedykowane są dla środowiska RStudio.*

### **Dane obserwacyjne:**

Oficjalne dane meteorologiczne można nabyć odpłatnie z IMGW-PIB w kwocie ok. 1-2 PLN za 1 daną (np. 24 obserwacje temperatury / dzień \* ~30 dni w miesiącu =~ 1000PLN). Spora część danych wchodząca w tzw. zakres podstawowy pomiarów meteorologicznych dostępna jest za darmo na serwerach WMO w postaci depesz SYNOP. Depesze te są rozkodowywane przez kilka serwisów internetowych, wśród których do najpopularniejszych należy zaliczyć portal [ogimet.com](http://ogimet.com).

---

Dane w serwisie ogimet.com można wygenerować w postaci zestawienia generując odpowiednie zapytanie w adresie internetowym. Przykładowo w linku:

*<http://ogimet.com/cgi-bin/gsynres?lang=en&ind=12330&decoded=yes&ndays=40&ano=2016&mes=12&day=15&hora=06>* kolejne pogrubione wartości oznaczają:

**12330** - międzynarodowy kod Światowej Organizacji Meteorologicznej (WMO) stacji analogicznie do informacji z arkusza kalkulacyjnego, w którym wpisywałeś(aś) skład grupy

**40** - liczba dni dla których wygenerowano zestawienie

**2016** - rok

**12** - miesiąc

**15** - dzień

**06** - godzina

**Zmodyfikuj link w taki sposób aby pobrać dane dla wybranej stacji meteorologicznej i zapisz je np. w arkuszu kalkulacyjnym lub wykorzystaj przygotowane dane dla kilku stacji.**

Na potrzeby realizacji projektu niezbędne będzie wykorzystanie godzinowych wartości temperatury powietrza na wysokości 2 m n.p.t. [T(C)] oraz sum opadów atmosferycznych w interwałach 24-godzinnych [Prec(mm)]. Zakres czasowy dla opadów w depeszach SYNOP jest zmienny, stąd też konieczne będzie ręczne przeliczenie wartości w tej kolumnie aby uzyskać sumy opadów za 24h do godz. 6 UTC.

Dane obserwacyjne powinny zawierać wartości dla zakresu dat od **"2016-11-14 06:00:00 UTC"** do **"2016-12-11 06:00:00 UTC"**

**Dane w formie wstępnie przetworzonej dostępne są na stronie**

**<http://openmeteo.pl/projekt/> i nazwane w formacie "kod\_WMO.txt"**

Odpowiednio przygotowany plik wczytaj do środowiska obliczeniowego (np. RStudio i nazwij go 'obserwacje').

---

A1. Wykonaj wykres zmienności temperatury powietrza (1 wykres dla wartości godzinowych). Przygotuj oddzielny obiekt z wartościami minimalnymi, maksymalnymi i średnimi temperaturami powietrza (2-gi wykres) oraz dobowej sumy opadów atmosferycznych. (3-ci wykres). Pamiętaj o odpowiednim podpisaniu osi wykresów oraz wyjaśnieniu użytych symboli/kolorów (w podpisie lub w formie legendy do wykresu). W przypadku wartości dobowych najlepiej stworzyć oddzielny obiekt z wyliczonymi wartościami w oparciu o dane godzinowe. Dopuszczalne jest agregowanie danych dla temperatury powietrza w dobach kalendarzowych (umieść wówczas tą informację w raporcie).

A2. Korzystając z technik modelowania regresyjnego oceń w jakim tempie rosła/spadała (1) średnia dobowa temperatura powietrza, (2) minimalna dobowa temperatura powietrza, (3) maksymalna dobowa temperatura powietrza w analizowanym okresie czasu w przeliczeniu na 1 lub 10 dni.

## **Prognoza pogody modelu GFS**

Archiwalne dane prognostyczne modelu GFS dla obszaru Polski dostępne są na stronie <http://openmeteo.pl/projekt/>; dane dostępne są w formatach CSV lub RDS (oba pliki zawierają dokładnie te same informacje).

Są to symulacje modelu GFS przycięte do obszaru Polski. Start symulacji następował codziennie o godz. 6 UTC od 14. listopada 2016 r. do 8. grudnia 2016 r. w horyzoncie wyprzedzenia prognozy od 0 do 72 godzin. Poprawnie wczytany zbiór danych powinien zawierać 2,053,125 rzędów i 7 kolumn. Kolejne wartości w kolumnach oznaczają:

date1 -> czas początkowy symulacji / czas wydania prognozy

date2 -> czas obowiązywania prognozy

lon -> długość geograficzna oczka siatki obliczeniowej

lat -> szerokość geograficzna oczka siatki obliczeniowej

TMP -> temperatura powietrza na wysokości 2 m n.p.t. w Kelwinach ( $T(^{\circ}\text{C}) = \text{TMP} - 273.15$ );

pamiętaj o przeliczeniu wartości!

---

PRATE -> opad atmosferyczny docierający do powierzchni terenu w ciągu godziny [mm/h]  
hh -> wyprzedzenie prognozy (date2-date1) w godzinach

## **Wczyt danych GFS w środowisku R**

Pobierz dane na dysk oraz ustaw katalog roboczy w RStudio na folder zawierający pobrany plik. Jeśli jest to plik gfs.rds można go wczytać do nowotworzonego obiektu 'dane' za pomocą komendy:

```
dane <- readRDS("gfs.rds")
```

## **Wybór najbliższego oczka siatki względem lokalizacji punktu pomiarowego**

Znajdź współrzędne najbliższego oczka siatki obliczeniowej modelu względem stacji pomiarowej wiedząc, że siatkę modelu GFS tworzą regularne gridy w odległościach co 0.25 stopnia.

Przykładowo, jeśli najbliższe oczko siatki obliczeniowej modelu względem stacji pomiarowej ma współrzędne lon=17.00 oraz lat=52.50 to jednym ze sposobów pozbycia się wszystkich niepotrzebnych danych jest użycie instrukcji 'filter' z pakietu 'dplyr' (może zaistnieć potrzeba jego zainstalowania), czyli:

```
library(dplyr)

dane <- filter(dane, lon==17, lat==52.5)
```

W ten sposób zamiast ponad 2 miliona rzędów powinno pozostać jedynie 1825 poprawnych wartości obserwacyjnych dla każdego z analizowanych elementów meteorologicznych.

## **Połączenie obu baz danych**

Dobrym zwyczajem jest połączenie serii danych obserwacyjnych i modelowych do postaci jednej ramki danych, tak aby w kolejnych kolumnach były wartości obserwowane i

---

odpowiadające im wartości prognoz. Połączenie dwóch zbiorów jest możliwe np. w oparciu o funkcję 'join' z pakietu 'plyr'. Jeśli w obu zbiorach danych będą takie same nazwy kolumn i takie same typy obiektów, wówczas R postara się je odpowiednio połączyć. Ważne aby format i klasa obiektów były takie same.

W tym celu warto utworzyć nową kolumnę w ramce danych "obserwacje" nazwaną "date1", w której można utworzyć obiekt analogicznej klasy jak w obiekcie 'gfs' przechowującej czas. Jeśli obiekt 'obserwacje' wygląda następująco:

```
> head(obserwacje)
  czas.   hh. t2m  prec   date2
1 2016-11-14 06:00:00 -3.9 ---- 2016-11-14 06:00:00
2 2016-11-14 07:00:00 -4.5 ---- 2016-11-14 07:00:00
```

To kolumnę date2 można stworzyć wykorzystując funkcje 'strptime' oraz 'paste' :

```
obserwacje$date2 <- strptime(paste(obserwacje$czas., obserwacje$hh.),
"%Y-%m-%d %H:%M:%S", tz='UTC')
```

Następnie połączony zbiór danych (np. 'baza') można utworzyć za pomocą polecenia:

```
baza <- join(obserwacje, dane)
```

## 2. Weryfikacja prognoz pogody

**A1:** Oblicz najczęściej stosowane miary jakości prognoz pogody dla elementów ciągłych (Jolliffe, Stephenson 2011). Wykonaj obliczenia statystyk weryfikacyjnych dla godzinowych wartości temperatury powietrza biorąc pod uwagę następujące miary:

- błąd systematyczny lub średni (ME - Mean Error) (bias)
  - $ME = (1/n) \sum (f_i - o_i)$ ; gdzie: f - prognoza, o - obserwacja
- średni błąd bezwzględny (MAE - Mean Absolute Error)
  - $MAE = (1/n) \sum |f_i - o_i|$
- pierwiastek ze średniego błędu kwadratowego (RMSE - Root Mean Squared Error)

- $RMSE = \sqrt{(1/n) \sum (f_i - o_i)^2}$
- Współczynnik korelacji Pearsona
- oblicz maksymalny błąd prognozy dla każdego z analizowanych modeli jaki wystąpił w rozpatrywanym okresie
- Jaki jest udział (procentowo) trafnych prognoz pogody przyjmując za dopuszczalny błąd +/-2°C?

W celu zachowania przejrzystości uzyskanych rezultatów wyniki zapisz w postaci tabelarycznej dla czasów wyprzedzenia prognozy z interwałem co 6 godzin (tj. 0, 6, 12... 72).

Opcjonalnie wyniki obliczeń można wzbogacić np. wykresem pudełkowym (tzw. boxplot) lub diagramami Taylora.

**A2:** Wykonaj wykresy rozrzutu (ang. *Scatter plots*) lub wykresy kwantylowe (w R komenda `qqplot`) pozwalające na określenie typowych błędów modelu dla horyzontu czasowego prognozy 0-23h. Opisz w jakich zakresach wartości temperatur model przejawia tendencję do przeszacowywania wartości temperatury, a w jakich zakresach wartości ta tendencja jest odwrotna?

**B: Wskaźniki weryfikacyjne dla opadu** oblicz w oparciu o tzw. tabele wielodzzielcze (lub tablice kontyngencyjne). Stwórz dla prognoz pogody z wyprzedzeniem 24, 48 i 72h odpowiednie tablice kontyngencyjne.

Tabela 1. Schemat tworzenia tablic kontyngencyjnych.

Wystąpienie opadu		obserwacja		
		Tak	Nie	Suma
prognoza	Tak	a	b	a+b
	Nie	c	d	c+d
	Suma	a+c	b+d	a+b+c+d



---

Na podstawie stworzonych tablic kontyngencyjnych oblicz następujące wskaźniki (dla każdej z analizowanych prognoz i wyprzedzenia czasowego) oraz zapisz je w postaci tabelarycznej:

- **FBI (Frequency Bias Index)**
  - $FBI = (a+b)/(a+c)$
- **POD (Probability of detection)**
  - $POD = a/(a+c)$
- **FAR (False Alarm Ratio)**
  - $FAR = b/(a+b)$

Przeprowadź dyskusję uzyskanych wyników oraz skrótkowo podsumuj całość wykonanej analizy w zadaniach 1-2. Opisz czy model przejawiał tendencję do systematycznego występowania błędów, które można łatwo wyeliminować (np. stosując techniki statystyczne) czy też miały one charakter losowy? Jak oceniasz jakość prognozy modelu GFS w analizowanym okresie czasu? Czy długość prognozy wpływa na jej dokładność? Itp.

#### LITERATURA:

Ian T. Jolliffe, David B. Stephenson (2011): Forecast Verification: A Practitioner's Guide in Atmospheric Science, 2nd Edition, ISBN: 978-0-470-66071-3; 292 pages