# Meteorológiai adatasszimiláció 1. rész

2026.11.05.

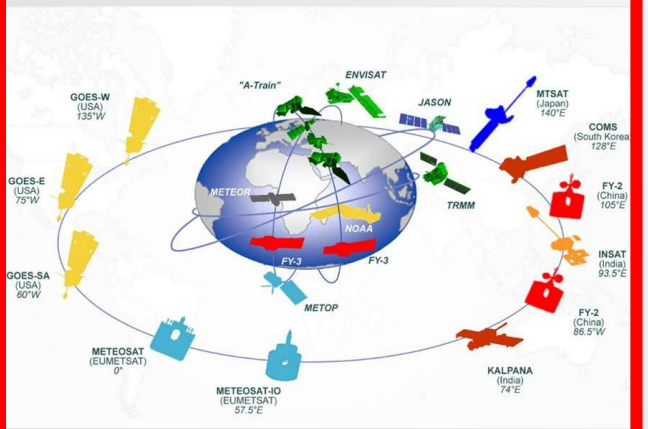
Kardos-Várkonyi Anikó

varkonyi.a@met.hu

# A számszerű előrejelzés lépései

**Adat-asszimiláció**

mérési információk gyűjtése, ellenőrzése, modell rácsra történő előállítása



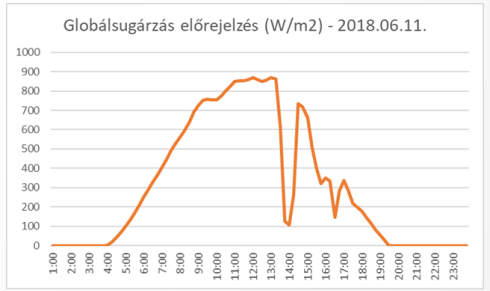

**Modell integrálás**

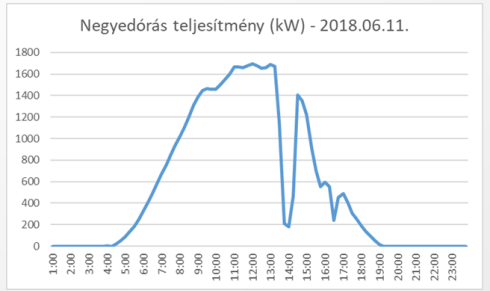
a légkör hidro-termodinamikai egyenletrendszerének közelítő megoldása




**Utófeldolgozás**

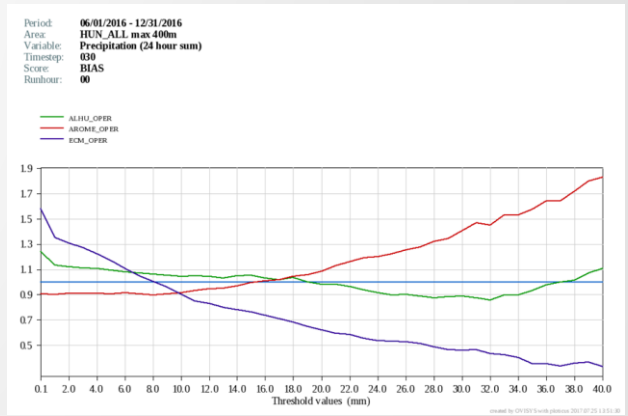
speciális paraméterek származtatása, megjelenítés




**Verifikáció**

az előrejelzések minőségének vizsgálata

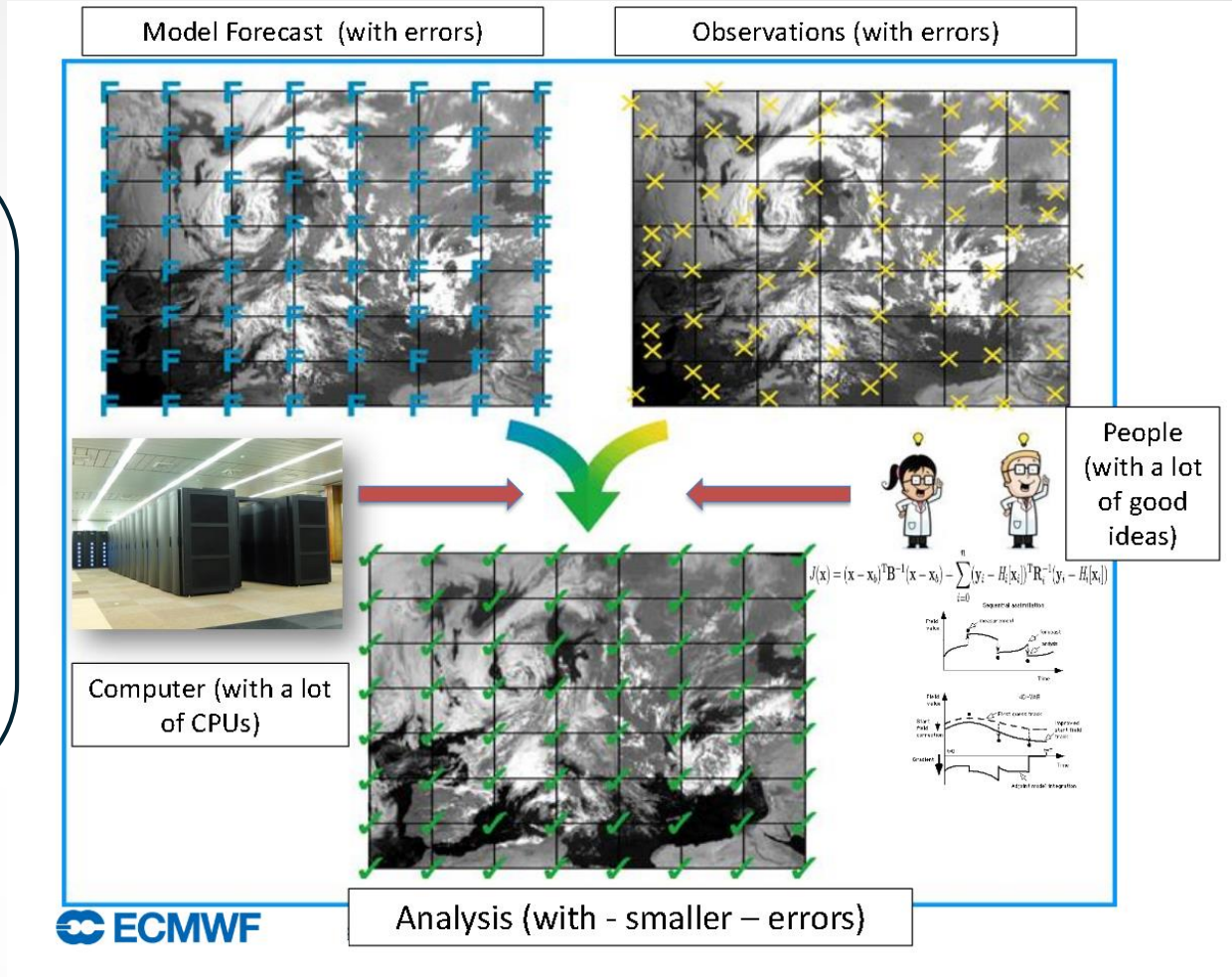


# Bevezetés

**Mi az adatasszimiláció célja?**

- Kezdeti feltételek biztosítása a numerikus modellek számára → Lényeges a pontosság a HTER megoldásának kezdeti feltételekre való érzékenysége miatt (a légkör kaotikus viselkedése)
- A bizonytalanságok meghatározása

Tehát az adatasszimiláció során analízist állítunk elő.
Analízis: Egy adott időpillanatban a légkör valós állapotáról a rendelkezésre álló információk alapján alkotott lehető legpontosabb kép




*Forrás: Massimo Bonavita, ECMWF NWP training, 2019*

**Mik szükségesek az adatasszimilációhoz?**

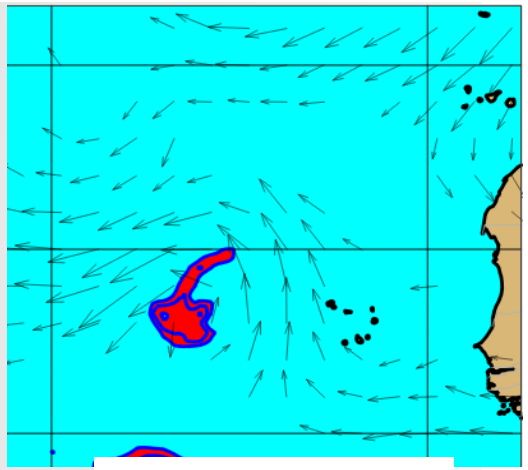
- Modell előrejelzés (background)
- Megfigyelések
- Számítási kapacitás
- Emberi erőforrás

**Mik a legnagyobb nehézségek?**

- A megfigyelések nem a modell rácspontjaiban helyezkednek el + indirekt és gyakran bonyolult változók mérése → megfigyelési operátor
- A modell előrejelzés és a megfigyelések is hibával terheltek → optimálisan kell ezeket ötvözni

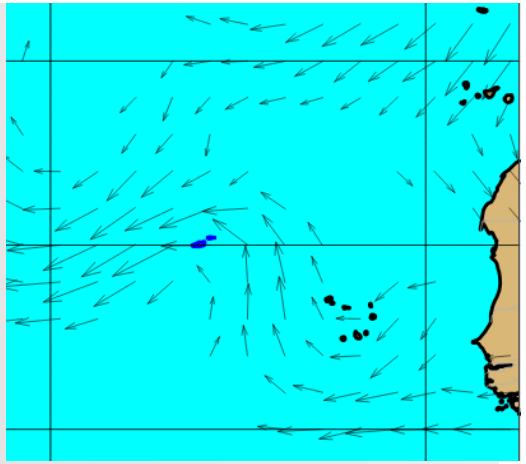
# Az adatasszimiláció jelentősége

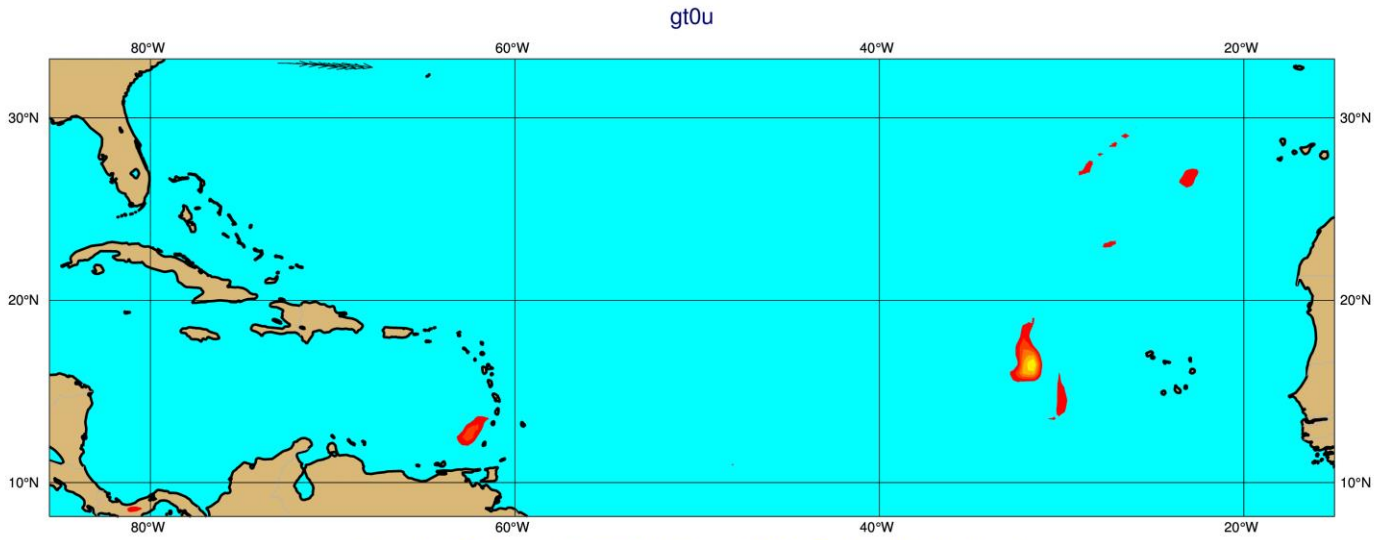
700 hPa relatív nedvesség & szél
kezdeti feltétel műholdas adatokkal



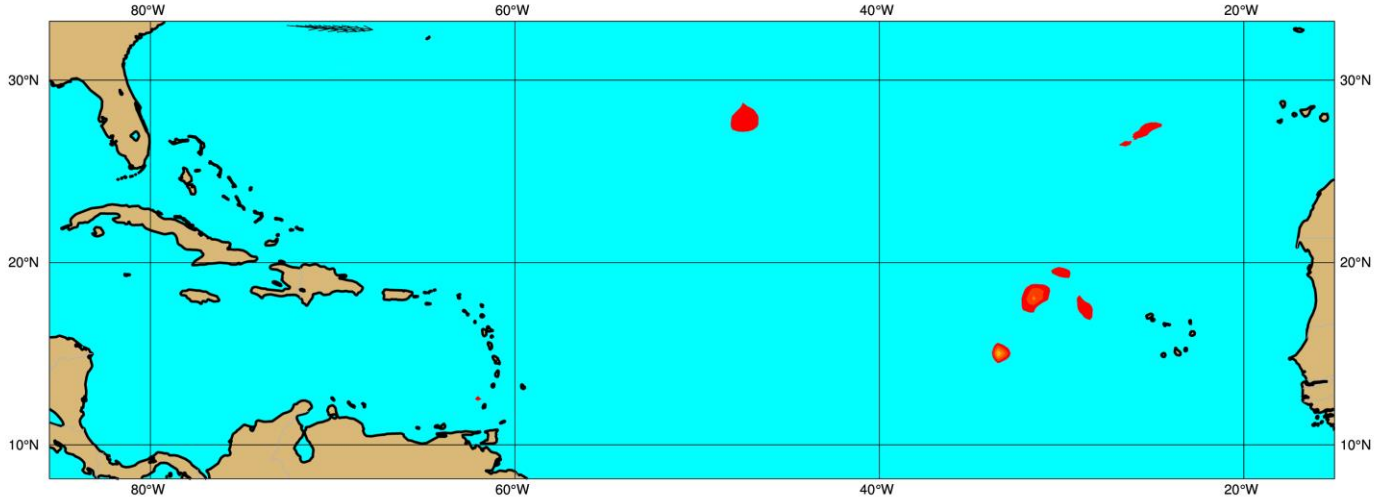
Nedvesség > 95%

700 hPa relatív nedvesség & szél
kezdeti feltétel műholdas adatok nélkül






Thursday 31 August 2017 00 UTC ecmf 500 hPa Vorticity (relative)
Thursday 31 August 2017 00 UTC ecmf 500 hPa U component of wind/V component of wind



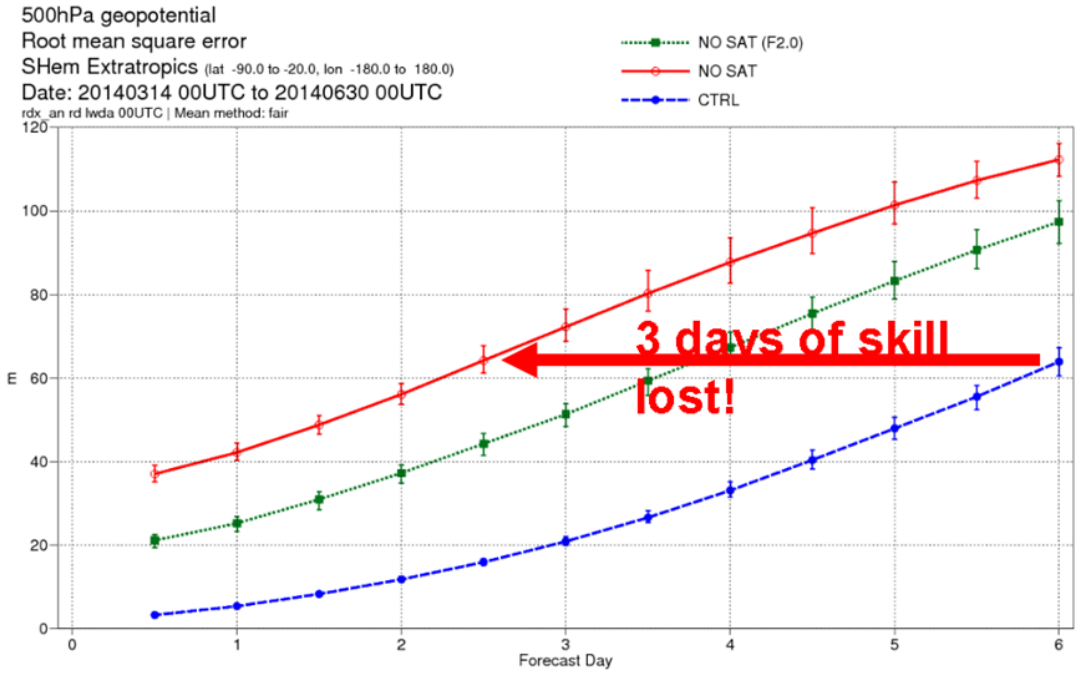
# Az adatasszimiláció jelentősége

## Műholdas mérések hatása

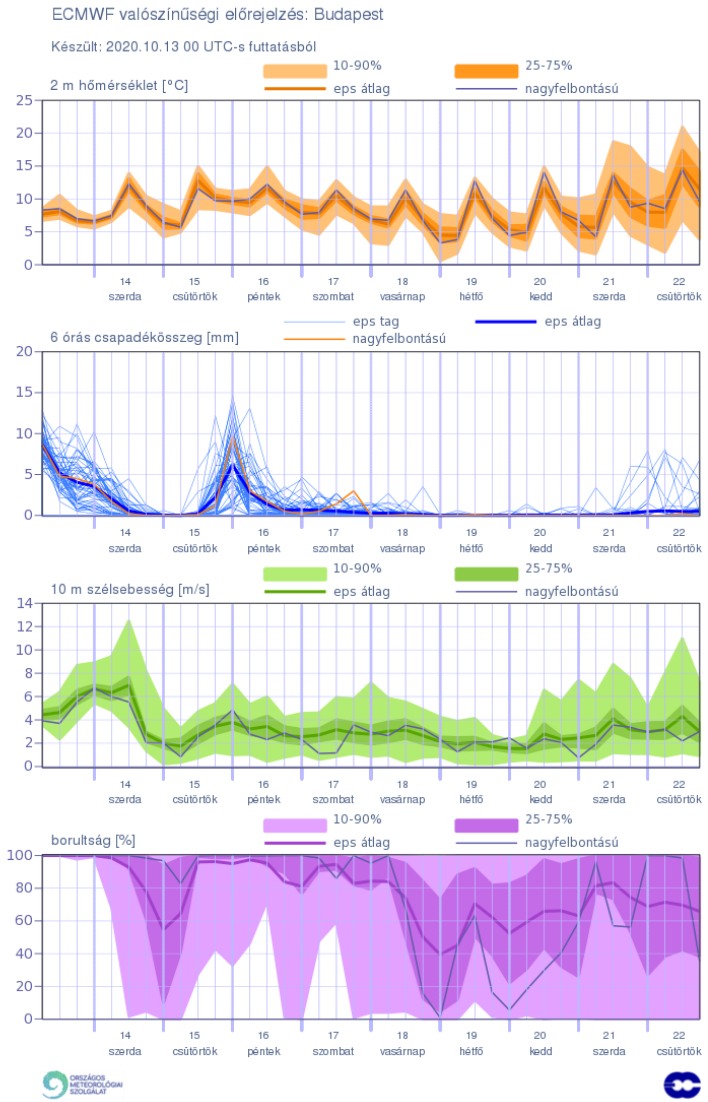
### Északi félteke

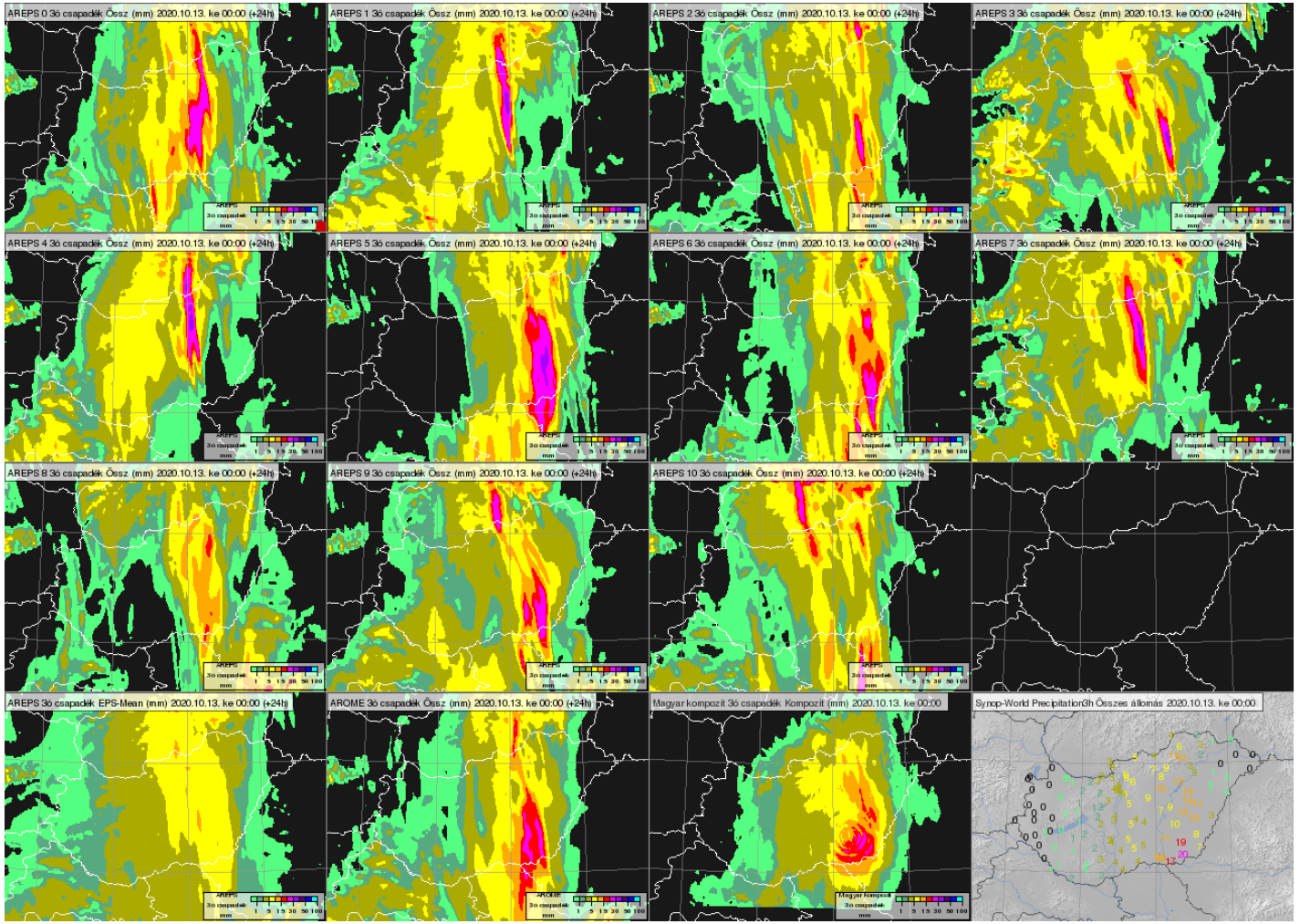



### Déli félteke



# Az adatasszimiláció jelentősége



# Milyen igényeket támasztunk?

1. A kezdeti feltétel ugyanazon a **rácson** álljon rendelkezésre, mint amin az előrejelzést elkészítjük
2. Használjunk fel **minden** rendelkezésünkre álló **információt** – **mik ezek az információk** és hogyan használjuk fel ezeket?

- Mérések és megfigyelések
- Korábbi előrejelzések
- Tudás a légkörről

**Miért nem elegendő csak a méréseket felhasználnunk?**

- A modell rácspontjaihoz képest kevesebb mérés áll rendelkezésre
- Nem elegendőek ahhoz, hogy megfelelő pontossággal az összes rácspontra meghatározzuk a modell prognosztikai változóit
- Alulhatározott lenne a feladat
- Még több információra van szükség (korábbi előrejelzések felhasználása)

# Milyen igényeket támasztunk?

1. A kezdeti feltétel ugyanazon a **rácson** álljon rendelkezésre, mint amin az előrejelzést elkészítjük
2. Használjunk fel **minden** rendelkezésünkre álló **információt** – mik ezek az információk **és hogyan használjuk fel ezeket?**

- **Optimálisan** ötvözzük a rendelkezésre álló információkat (minimális becslési hiba)
- **Vegyük figyelembe** a felhasznált **információk hibáját** (ezeket is csak becsülni tudjuk)
- A becslés legyen **összhangban a HTER-rel** (dinamikai konzisztencia)

**Milyen hibái lehetnek a felhasznált információknak?**

- Műszerek hibái
- Reprezentativitási hiba
- Átszámítási hibák
- Modell előrejelzések hibája

**Mi történik ha nem megfelelő a dinamikai konzisztencia?**

- Zajok megjelenése
- Modellinstabilitás

# Idealizált eset

- 1 térbeli pont
- 1 változó: $x$
- $x_a$ becslést adunk $x_t$-re (a valós állapotra)
- rendelkezésünkre állnak $y_1$ és $y_2$ megfigyelések $\varepsilon_1$ és $\varepsilon_2$ hibával terhelve:

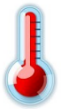
$$y_1 = x_t + \varepsilon_1$$

$$y_2 = x_t + \varepsilon_2$$

$$x_a = \hat{x}_t = f(y_1, y_2)$$

Szeretnénk a terem hőmérsékletét megbecsülni a jelen pillanatban – $T_t$

Rendelkezésünkre áll két mérés:

1. Digitális hőmérő: $T_1$ 
2. Analóg hőmérő: $T_2$ 

A rendelkezésre álló információk hibája:

$$\varepsilon_1 = T_1 - T_t \quad \text{és} \quad \varepsilon_2 = T_2 - T_t$$

# Idealizált eset

Feltesszük, hogy:

- a mérések torzítatlanok (nincs szisztematikus hiba):

$$E(\varepsilon_1) = E(\varepsilon_2) = 0$$

- a becslések (mérések) hibáinak szórásnégyzete ismert:

$$\sigma_1^2 = E(\varepsilon_1^2) \qquad \sigma_2^2 = E(\varepsilon_2^2)$$

- a becslések (mérések) hibái korrelálatlanok:

$$E(\varepsilon_1 \varepsilon_2) = 0$$

# Idealizált eset

Kérdés:

$$x_a = f(y_1, y_2) = ?$$

Több megközelítés létezik, például:

- Legkisebb négyzetek módszere
- Maximum likelihood módszer

# Legkisebb négyzetek módszere

Az **analízist** a becslések lineáris kombinációjaként állítjuk elő:

$$\left.\begin{aligned} y_1 &= x_t + \varepsilon_1 \\ y_2 &= x_t + \varepsilon_2 \\ x_a &= k_1 y_1 + k_2 y_2 \end{aligned}\right\} \quad \begin{aligned} k_1 &= ? \\ k_2 &= ? \end{aligned}$$

Szeretnénk, hogy:

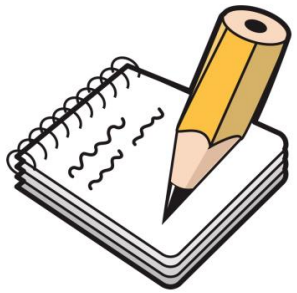
- a becslés torzítatlan legyen: $\mathrm{E}(\mathrm{x_a} - \mathrm{x_t}) = 0 \longrightarrow \mathrm{k_1} + \mathrm{k_2} = 1$
- a becslés négyzetes hibája minimális legyen: $\sigma_{\mathrm{a}}^2 = \mathrm{E}\left((\mathrm{x_a} - \mathrm{x_t})^2\right) = \min$

# 1. feladat:

$$\left.\begin{aligned} y_1 &= x_t + \varepsilon_1 \\ y_2 &= x_t + \varepsilon_2 \\ x_a &= k_1 y_1 + k_2 y_2 \end{aligned}\right\} \begin{aligned} k_1 &= ? \\ k_2 &= ? \end{aligned}$$

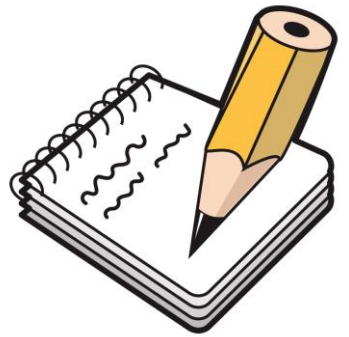
Mutassuk meg, hogy

$$E\left(x_a - x_t\right) = 0 \quad \Rightarrow \quad k_1 + k_2 = 1$$



# 1. feladat:

$$\left.\begin{aligned} y_1 &= x_t + \varepsilon_1 \\ y_2 &= x_t + \varepsilon_2 \\ x_a &= k_1 y_1 + k_2 y_2 \end{aligned}\right\}$$

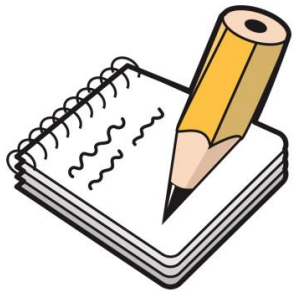


# 2. feladat:

$$\left.\begin{aligned} x_a &= k_1 y_1 + k_2 y_2 \\ k_1 + k_2 &= 1 \end{aligned}\right\} \quad x_a = k_1 y_1 + (1 - k_1) y_2$$

$$\left.\begin{aligned} y_1 &= x_t + \varepsilon_1 \\ y_2 &= x_t + \varepsilon_2 \\ x_a &= k_1 y_1 + k_2 y_2 \end{aligned}\right\}$$

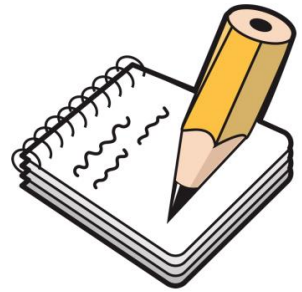
Feladat: $\min \sigma_a^2(k_1)$



A becslés: $$x_a = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} y_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} y_2$$

A becslés megbízhatósága: $$\frac{1}{\sigma_a^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} = \sum_{i=1}^{p} \frac{1}{\sigma_i^2}$$

Minden megfigyelés növeli a megbízhatóságot!

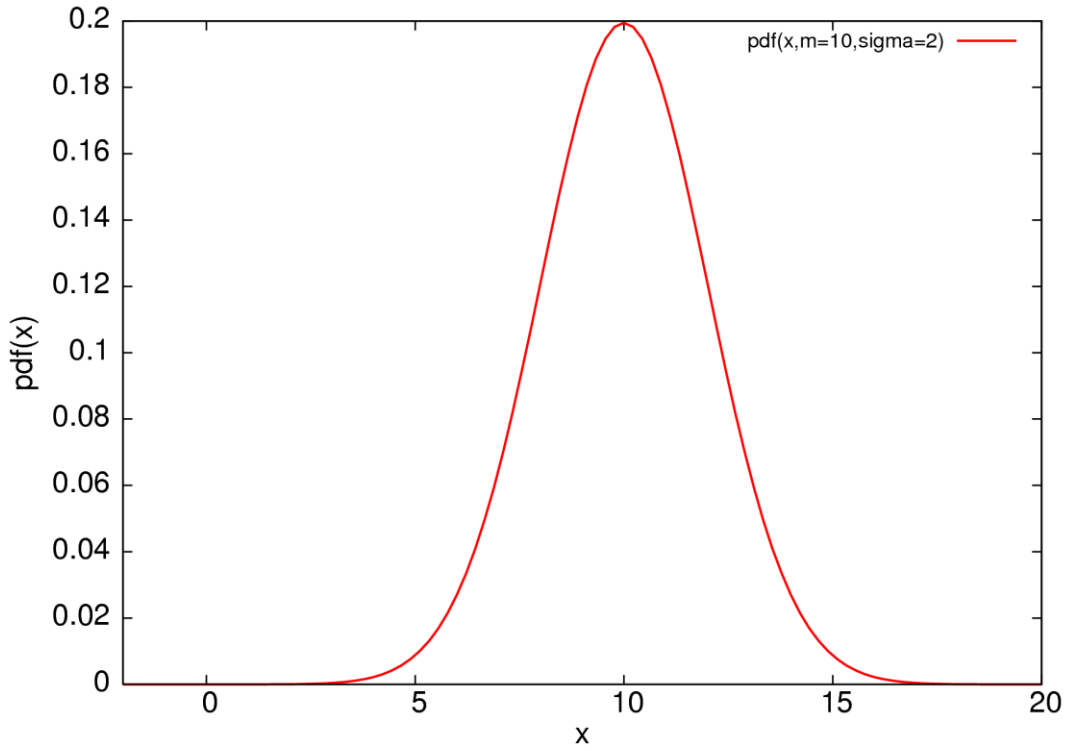


# Maximum likelihood módszer

A meteorológiai változók „jól” modellezhetők normális eloszlású valószínűségi változókként

$$x \in N\left(m, \sigma^2\right)$$

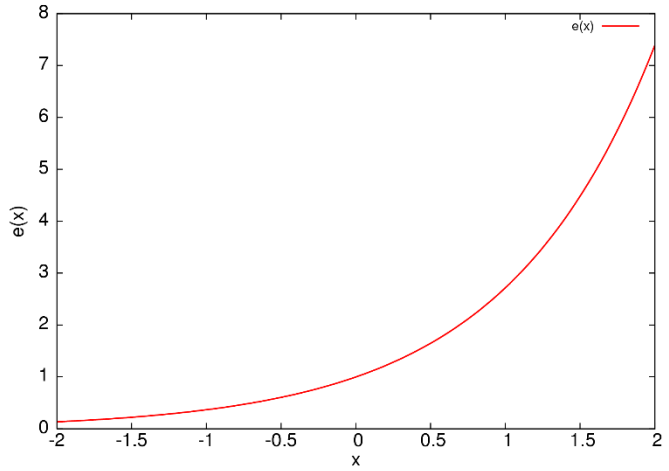
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left( -\frac{(x-m)^2}{2\sigma^2} \right)$$



Az $y_1$ és $y_2$ mérések független becslést adnak $x_t$-re $x$ várható értékkel, $\sigma_1$ és $\sigma_2$ szórással:

$$f_1(x) = \frac{1}{\sigma_1\sqrt{2\pi}} \exp\left(-\frac{(x-y_1)^2}{2\sigma^2{}_1}\right)$$

$$f_2(x) = \frac{1}{\sigma_2\sqrt{2\pi}} \exp\left(-\frac{(x-y_2)^2}{2\sigma^2{}_2}\right)$$




Becslés: $x_a$ a minta együttes sűrűségfv-ének $x$ szerinti maximuma

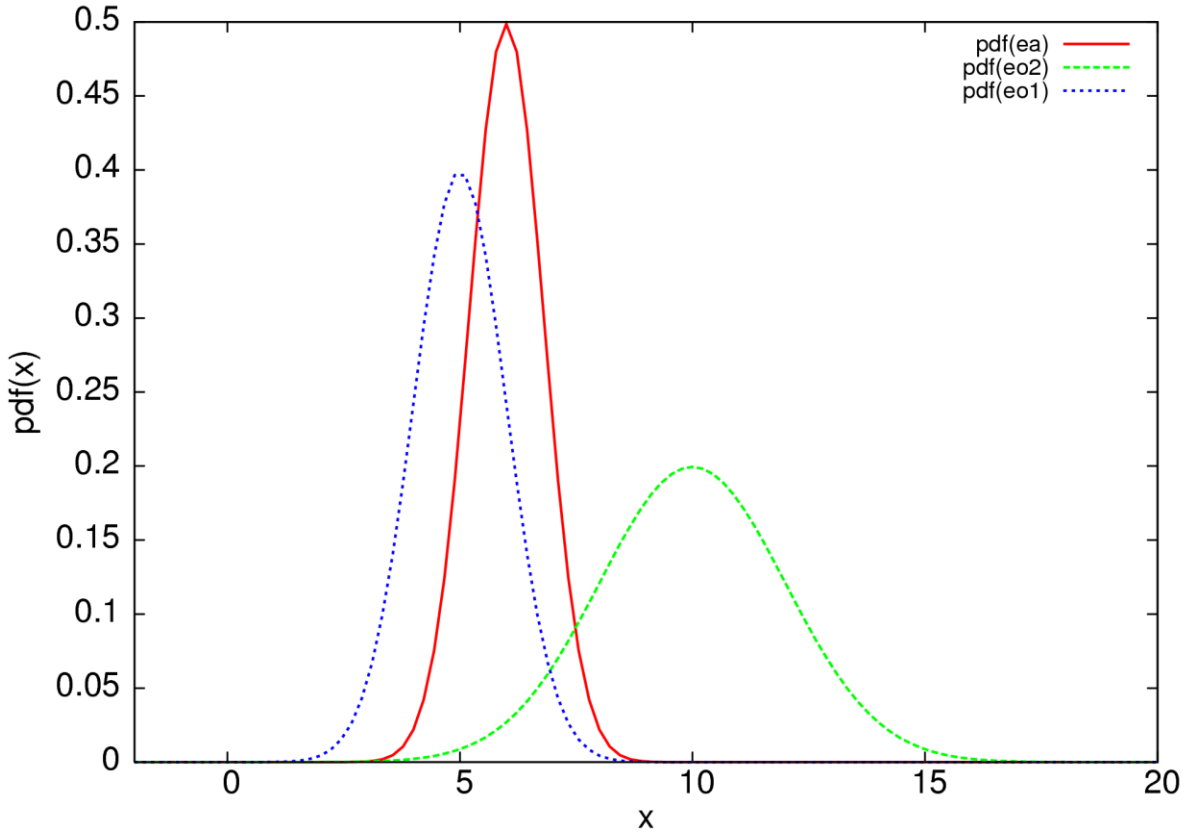
$$x_a \to \max_x f_1(x) f_2(x) = \max_x \frac{1}{\sigma_1\sigma_2 2\pi} \exp\left(-\left(\frac{1}{2}\frac{(y_1-x)^2}{\sigma_1^2} + \frac{1}{2}\frac{(y_2-x)^2}{\sigma_2^2}\right)\right) =$$

$$= \min_x \left(\frac{1}{2}\frac{(y_1-x)^2}{\sigma_1^2} + \frac{1}{2}\frac{(y_2-x)^2}{\sigma_2^2}\right)$$

$$x_a = \frac{\sigma_2^{\,2}}{\sigma_1^2+\sigma_2^2} y_1 + \frac{\sigma_1^{\,2}}{\sigma_1^2+\sigma_2^2} y_2$$

$$\frac{1}{\sigma_a^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} = \sum_i \frac{1}{\sigma_i^2} \quad \Longrightarrow$$ Minden megfigyelés növeli a megbízhatóságot!

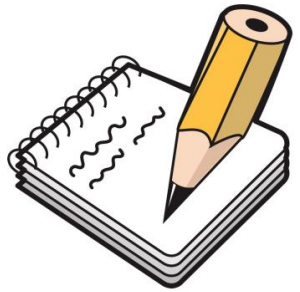
## Példa:




$y_1=5$; $\sigma_1=1$

$y_2=10$; $\sigma_2=2$

$x_a=6$; $\sigma_a=0{,}89$



# Összefoglalás

Normális eloszlású hibák esetén a legkisebb négyzetek módszere és a maximum likelihood módszerek ekvivalensek

Legkisebb négyzetek módszere
Pl. optimális interpoláció
(lineáris regresszió)

$$x_a = \sum_{i=1}^{p} k_i y_i$$

$$k_i = ?$$

Maximum likelihood módszer
Pl. variációs módszer
(veszteség függvény)

$$J(x) = \frac{1}{2} \sum_{i=1}^{p} \frac{(x - y_i)}{\sigma_i^2}$$

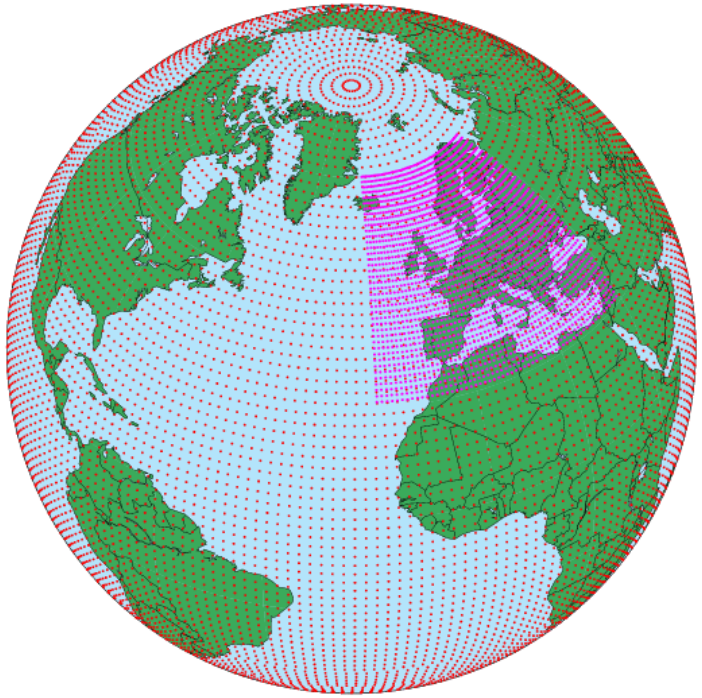
$$\operatorname{argmin} J(x) = ?$$

(iteratív minimum kereső algoritmusok)

# Adatasszimiláció a gyakorlatban

A numerikus modellek egy 3D **rácson** oldják meg a HTER-t. A rácson értelmezett meteorológiai változók együttese az **állapotvektor (továbbiakban X)** amely a mai modellek esetében **n~$10^7$** nagyságrendű.

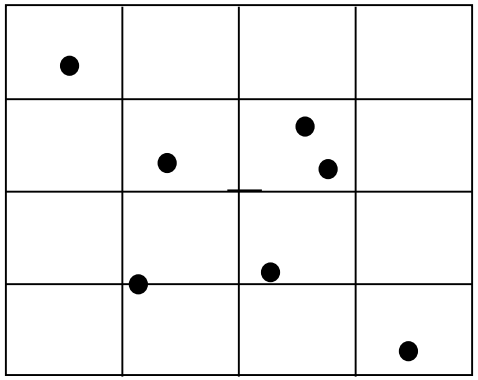
**Adatasszimiláció:** az állapotvektor megadása a kezdeti időpillanatban (**analízis**) úgy, hogy az minél közelebb legyen a valósághoz.



# Felhasználható információk

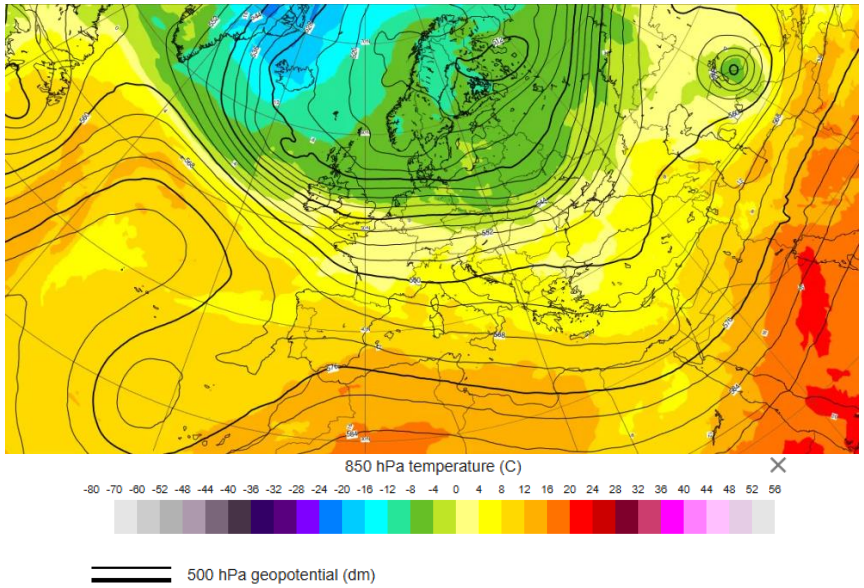
**Megfigyelések:**

- Időben rendszeres számszerű mérések
- Térben **szabálytalan rács**
- Összetett összefüggés (lehet) a mért és a modell változók között
- A felhasznált megfigyelések száma kevesebb, mint a modellrács pontjainak száma ($p \sim 10^5$)

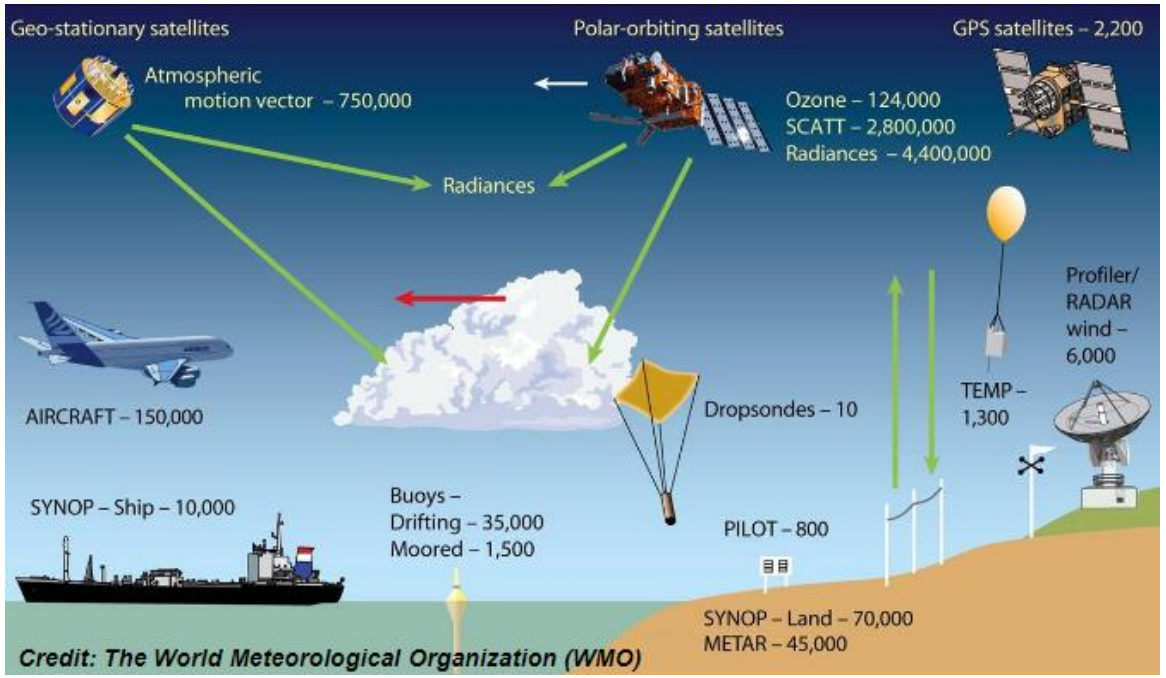


**Background/first guess:**

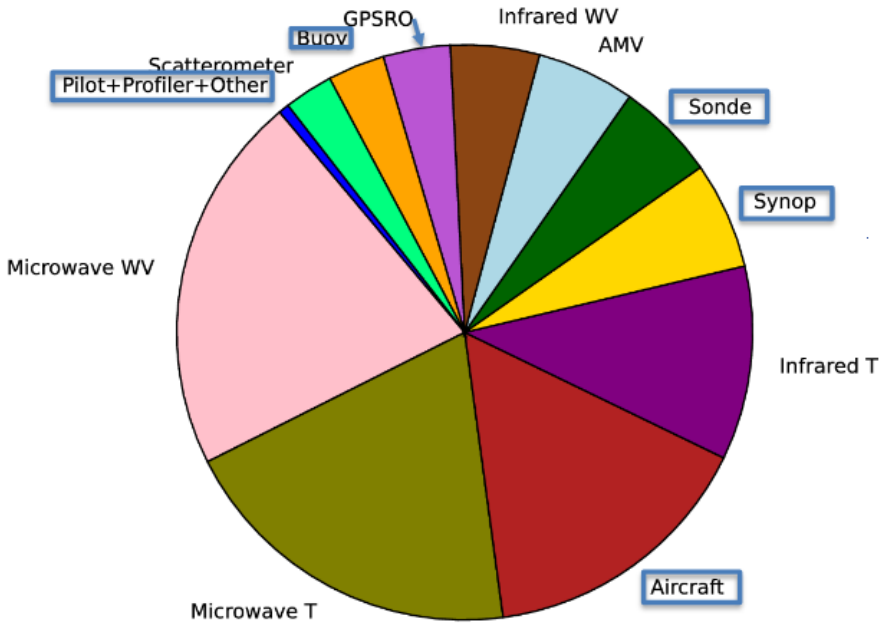
- A modell által előrejelzett állapotvektor
- Térben szabályos rács
- Közvetlenül modell változók
- Dimenzió: $n \sim 10^7$



# Megfigyelések globális rendszere




*Forrás: Tony McNally, ECMWF NWP training, 2015*

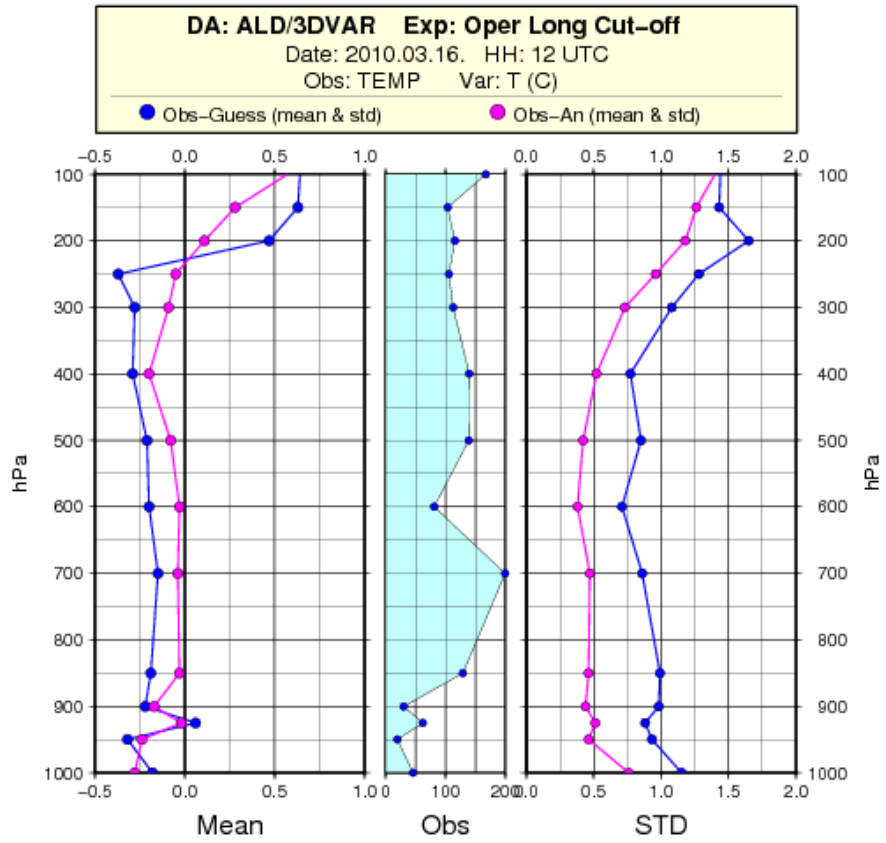



24 órás előrejelzés esetén:

- 70 % : műholdas adatok
- 30 %: in situ megfigyelések

# Konvencionális megfigyelések

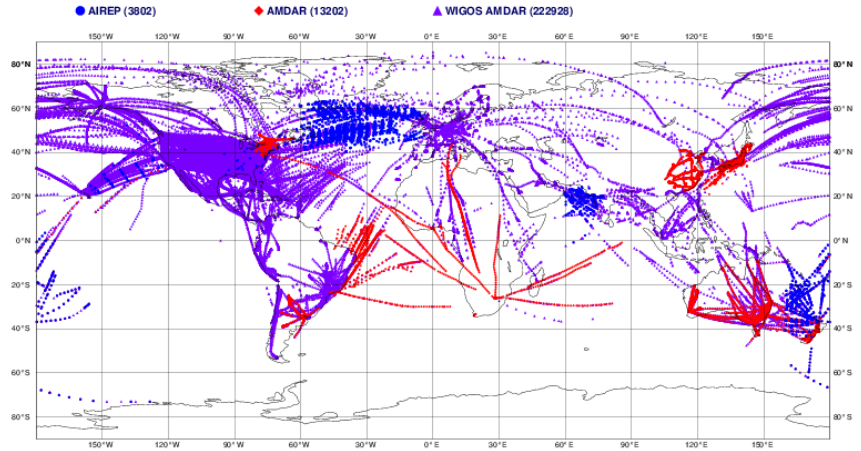
| Eszköz | Változó | Szint |
|---|---|---|
| SYNOP<br>SHIP<br>METAR | hőmérséklet, harmatpont, szél, légnyomás | Felszín, 2m, 10m, 25m |
| Bóják | hőmérséklet, légnyomás, szél | 2m |
| Rádiószondák<br>TEMPSHIP<br>Ejtőszondák | hőmérséklet, légnyomás, szél, nedvesség | Profil |
| Profilerek | szél | Profil |
| Repülőgépek | hőmérséklet, légnyomás, szél | Profil, repülési szint |



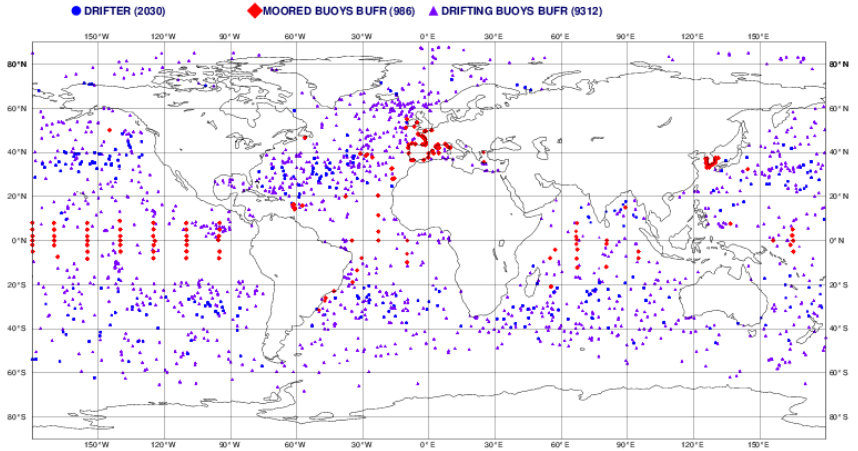

*Forrás: Tony McNally, ECMWF NWP training, 2015*

# Konvencionális/in-situ megfigyelések

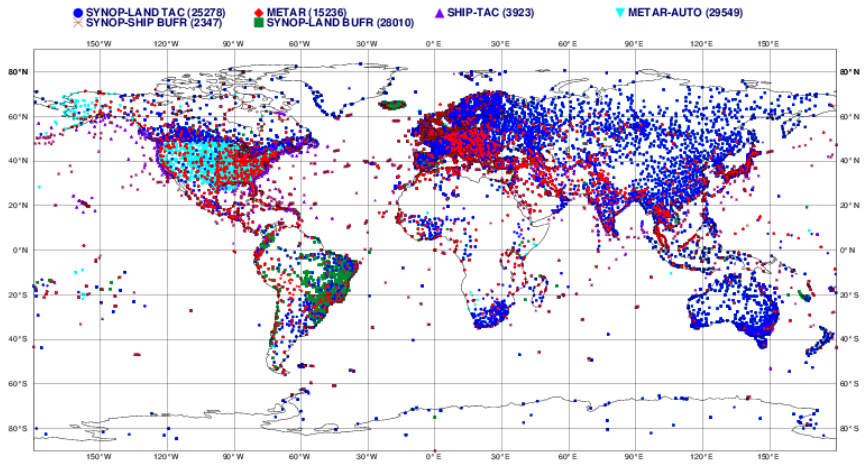
ECMWF data coverage (all observations) - AIRCRAFT
13/11/2018 00
Total number of obs = 239932



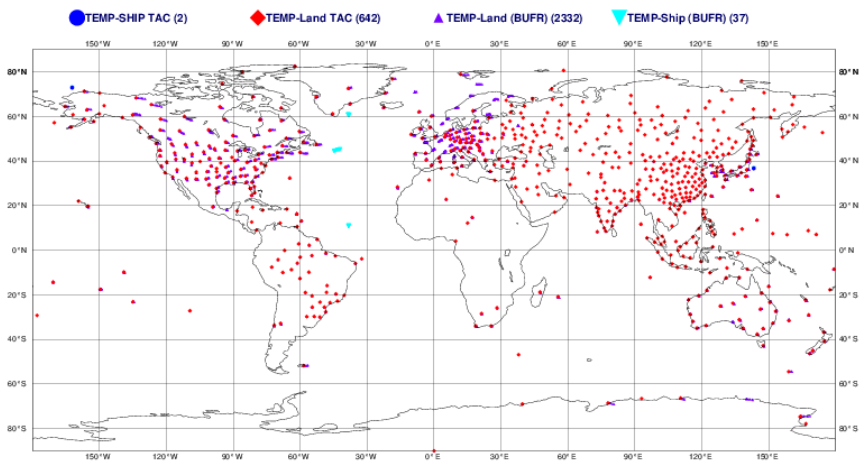

ECMWF data coverage (all observations) - BUOY
13/11/2018 00
Total number of obs = 12328




ECMWF data coverage (all observations) - SYNOP-SHIP-METAR
13/11/2018 00
Total number of obs = 104343




ECMWF data coverage (all observations) - RADIOSONDE
13/11/2018 00
Total number of obs = 3013

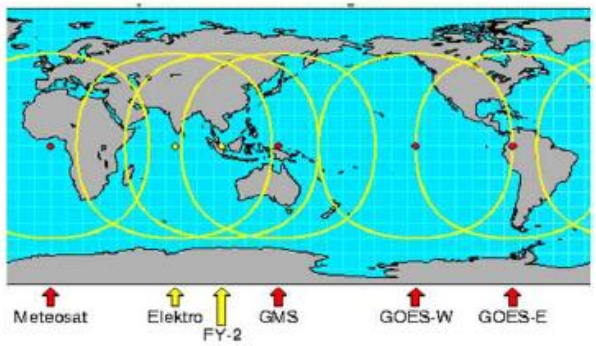


# Műholdas mérések

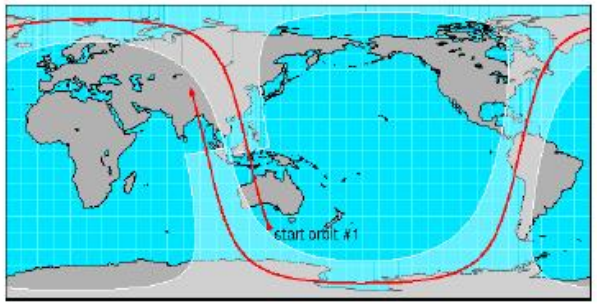
**Geostacionárius és kvázipoláris (low Earth orbit) műholdak**

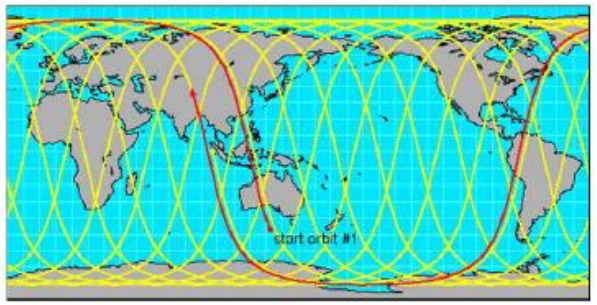
GEO



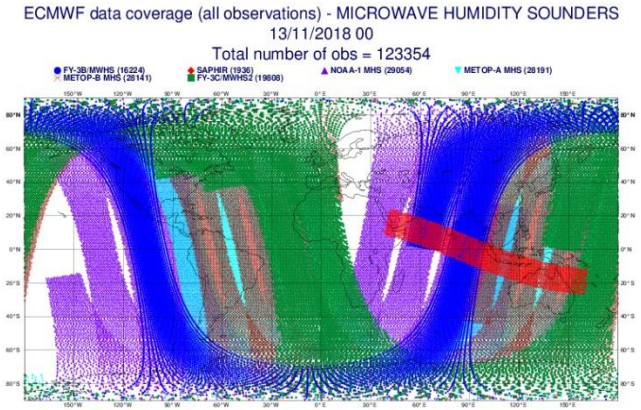

LEO

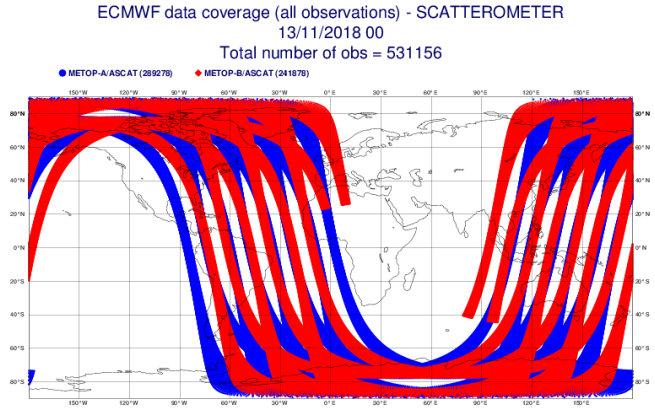


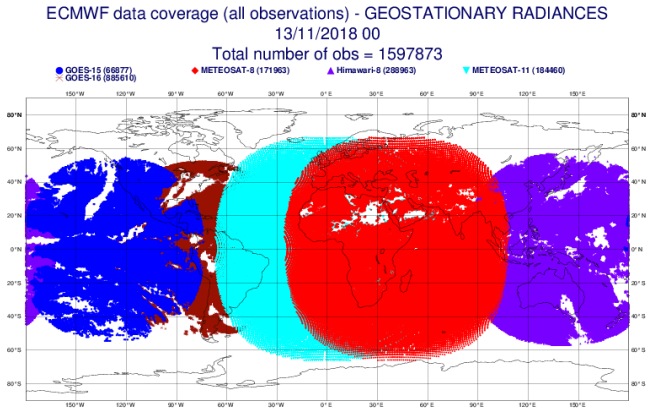



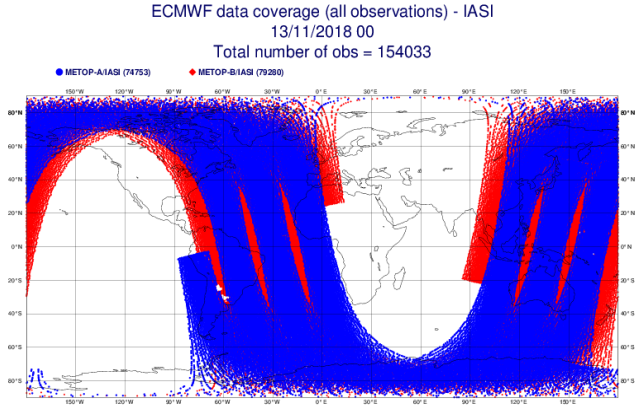

*Forrás: Tony McNally, ECMWF NWP training, 2015*

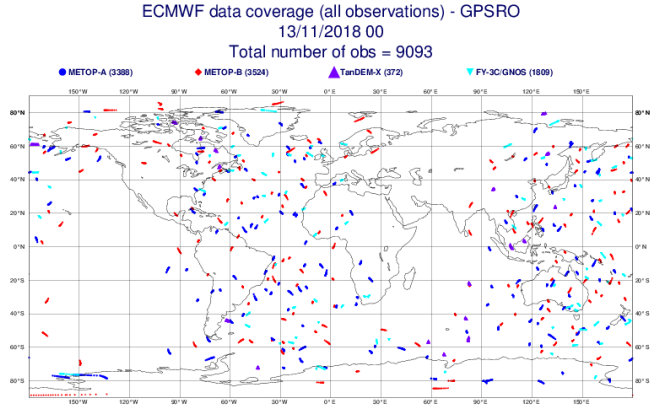
# Műholdas mérési lefedettség

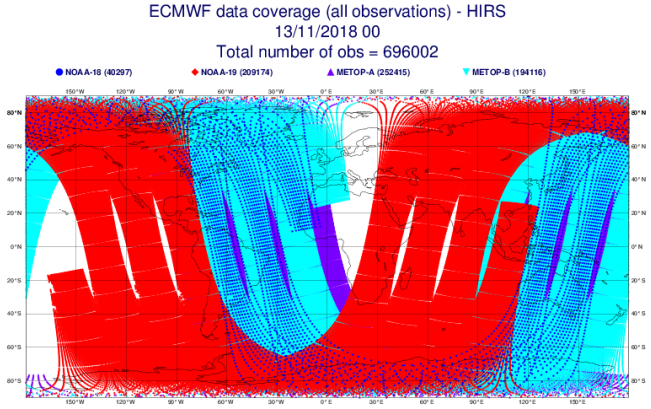


# Megfigyeléstípusok jellemzői

## Konvencionális mérések

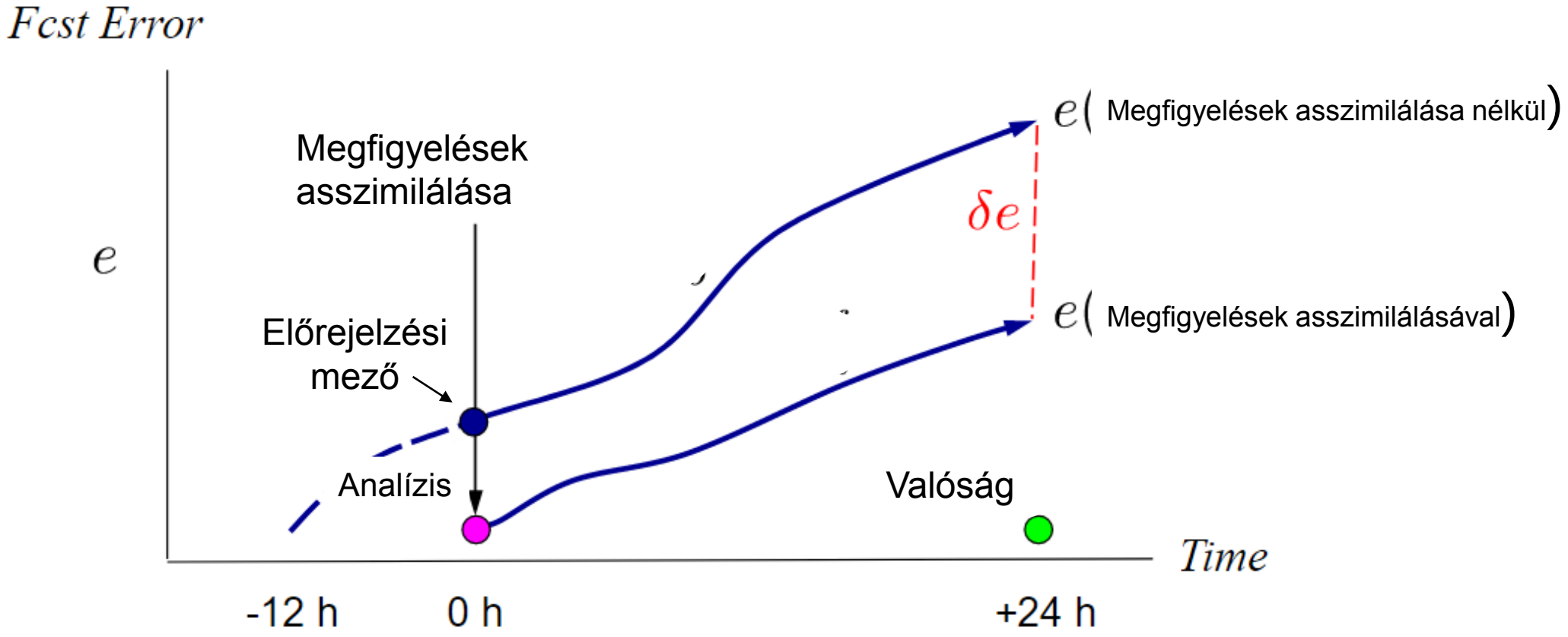
- Nem egyenletes térbeli és időbeli lefedettség. Vannak olyan területek, ahol nincsenek konvencionális mérések.
- Pontszerű mérések
- Nem egységes adatminőség
- Közvetlenül azt mérik, amit a modellben is használunk
- Értelmezésük és használatuk az adatasszimiláció során „egyszerű”

## Műholdas mérések

- Indirekt és gyakran bonyolult változók mérése, ami a modell számára NEM azonnal felhasználható → átalakítások szükségesek → **radiatív transzfer egyenlet** segítségével megoldható a probléma.
- Nehezítő körülmények (pl. felhőzet)
- Gyenge vertikális felbontás
- Bár a legtöbb műholdas megfigyelés nagyon jó minőségű, ez hirtelen megváltozhat a műszer meghibásodása vagy szoftveres probléma miatt
- Hatalmas adatmennyiség
- Egyenletes, globális térbeli lefedettség
- Nagy időbeli sűrűség

# Adatasszimiláció hatékonyságának az ellenőrzése

FSOI - Forecast Sensitivity Observation Impact



To access this activity, students go to **zzi.sh** and insert

**ake94796**

Copy class code





# Irodalom


Berre, L., 2000: Estimation of synoptic and meso scale forecast error covariances in a limited are amodel. *Mon Weather Rev* **128**, 644-667.

Bouttier, F. and Courtier, P., 1999: Data Assimilation concepts and methods. *ECMWF Lecture Series (Data assimilation and use of satellite data, http://www.ecmwf.int/newsevents/training/lecture_notes/pdf_files/ASSIM/Ass_cons.pdf)* 72.

Dévényi, D. és Gulyás, O., 1988: Matematikai statisztikai módszerek a meteorológiában, *Tankönyvkiadó*, Budapest, 443 pp.

Gandin, L.S, 1963: Objective analysis of meteorological fields, *Gidrometeorologicheskoe Izdatelstvo*, Leningrad, English translation by Israeli Program for Scientific Translations, Jerusalem, 1965

Gilbert, J. and Lemarechal, C., 1989: Some numerical experiments with variable-storage quasi-Newton algorithms, *Mathematical Programming* **45**, 407-435.

Kalnay, E., 2003: Atmospheric Modeling, Data Assimilation and Predictability, *Cambridge University Press*, Cambridge, 341 pp.

*ECMWF Data Assimilation NWP training, 2019*

*ECMWF eLearning, An introduction to Data Assimilation*
*https://learning.ecmwf.int/mod/scorm/player.php?a=179&currentorg=An_introduction_to_data_assimilation_ORG&scoid=458*