Meteorológiai adatasszimiláció

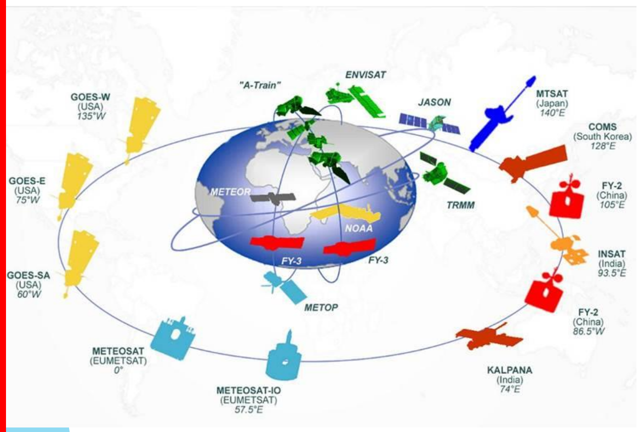
2023.11.09. és 2023.11.23.

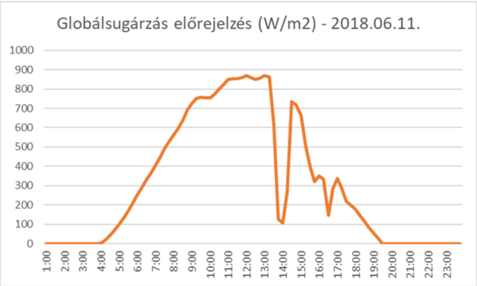
Kardos-Várkonyi Anikó

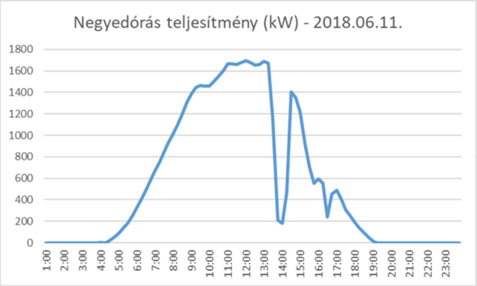
varkonyi.a@met.hu

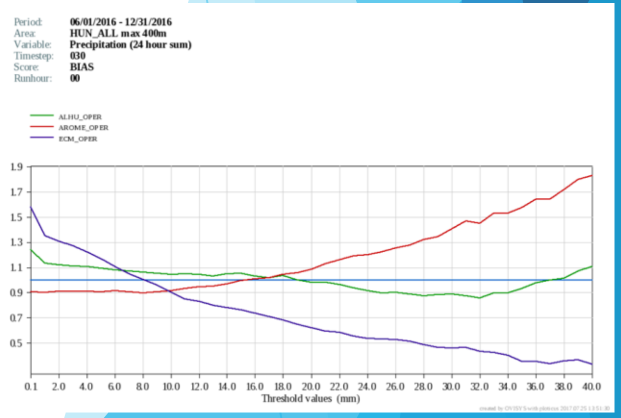
A számszerű előrejelzés lépései

Adat-asszimiláció	Modell integrálás	Utófeldolgozás	Verifikáció
mérési információk gyűjtése, ellenőrzése, modell rácsra történő előállítása	a légkör hidro-termodinamikai egyenletrendszerének közelítő megoldása	speciális paraméterek származtatása, megjelenítés	az előrejelzések minőségének vizsgálata









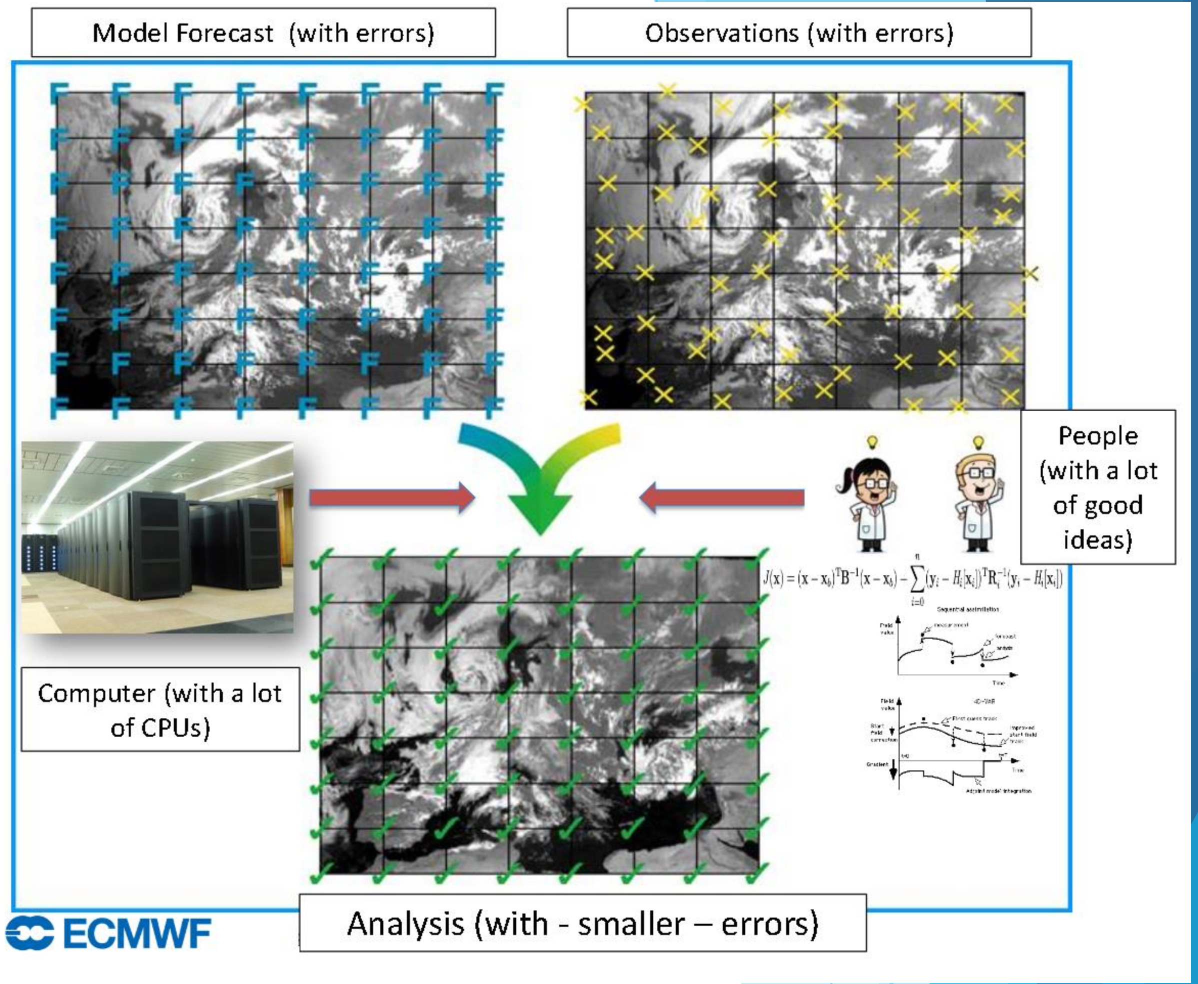
Bevezetés

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

Mik szükségesek az adatasszimilációhoz?

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



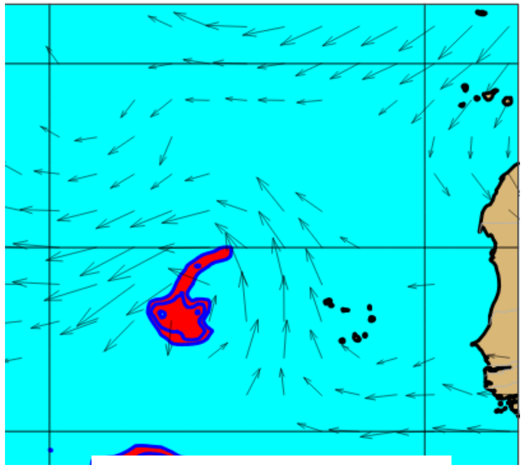
Forrás:Massimo Bonavita, ECMWF NWP training, 2019

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

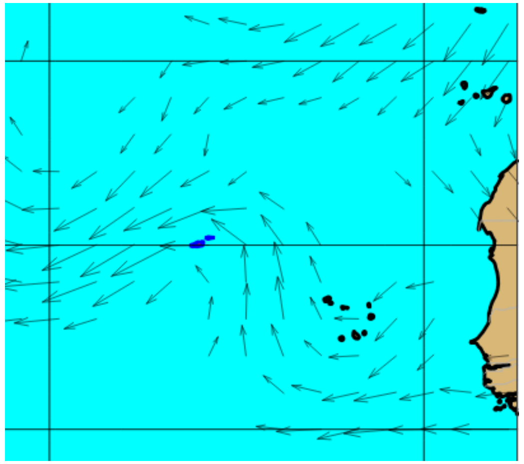
Az adatasszimiláció jelentősége

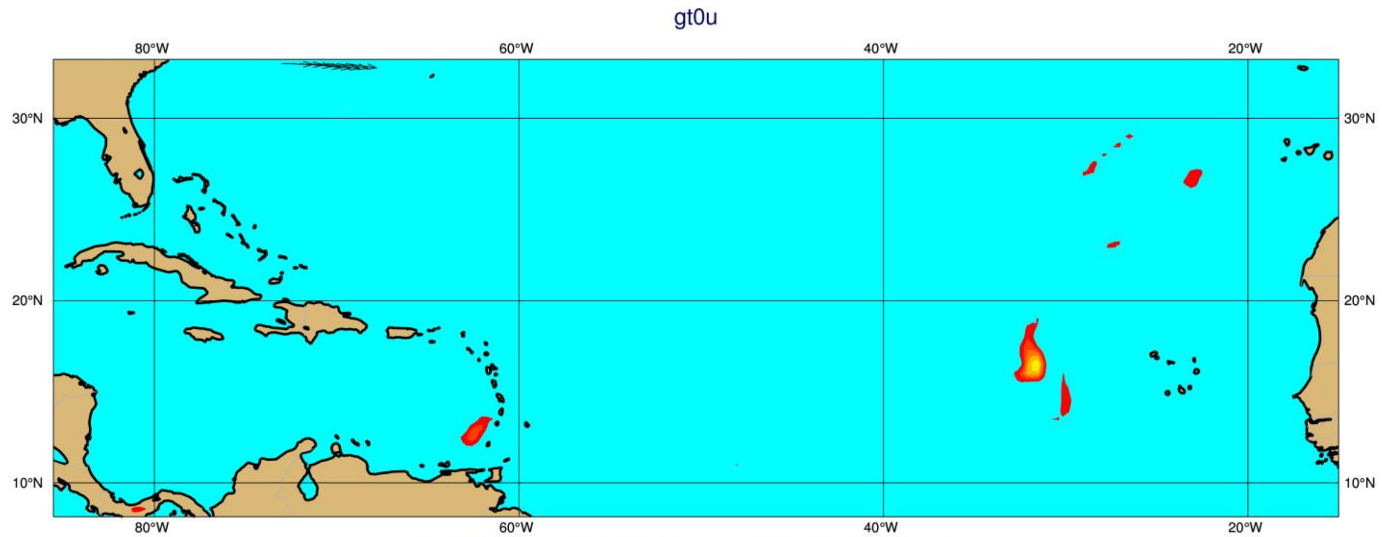
700 hPa relatív nedvesség & szél
kezdeti feltétel műholdas adatokkal



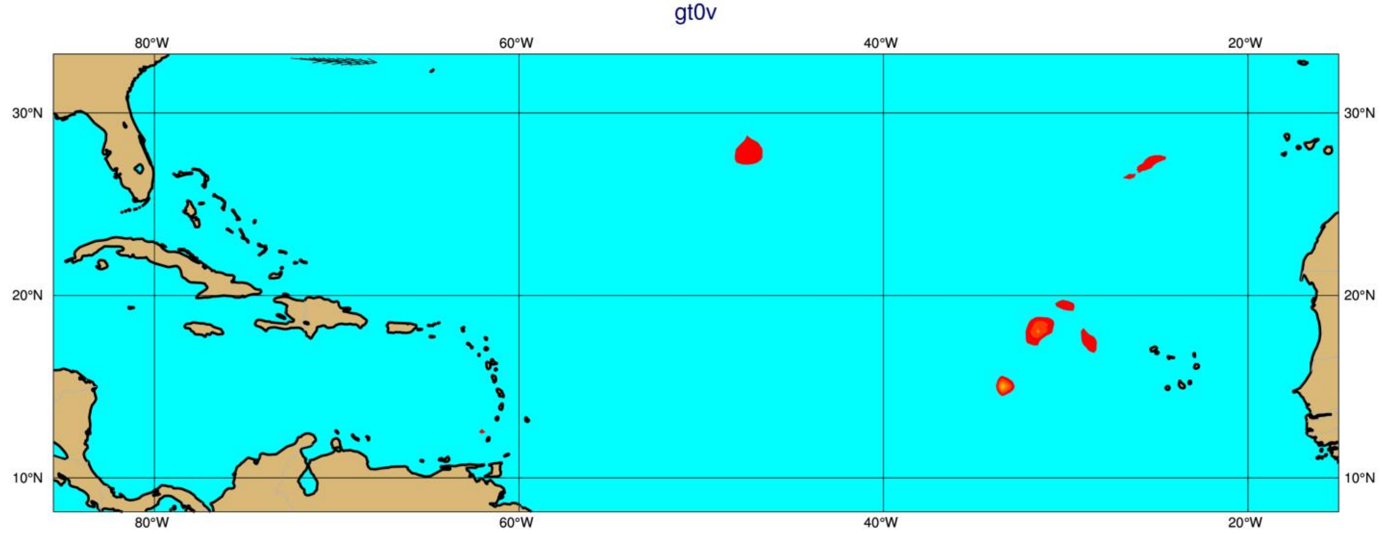
Nedvesség > 95%

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

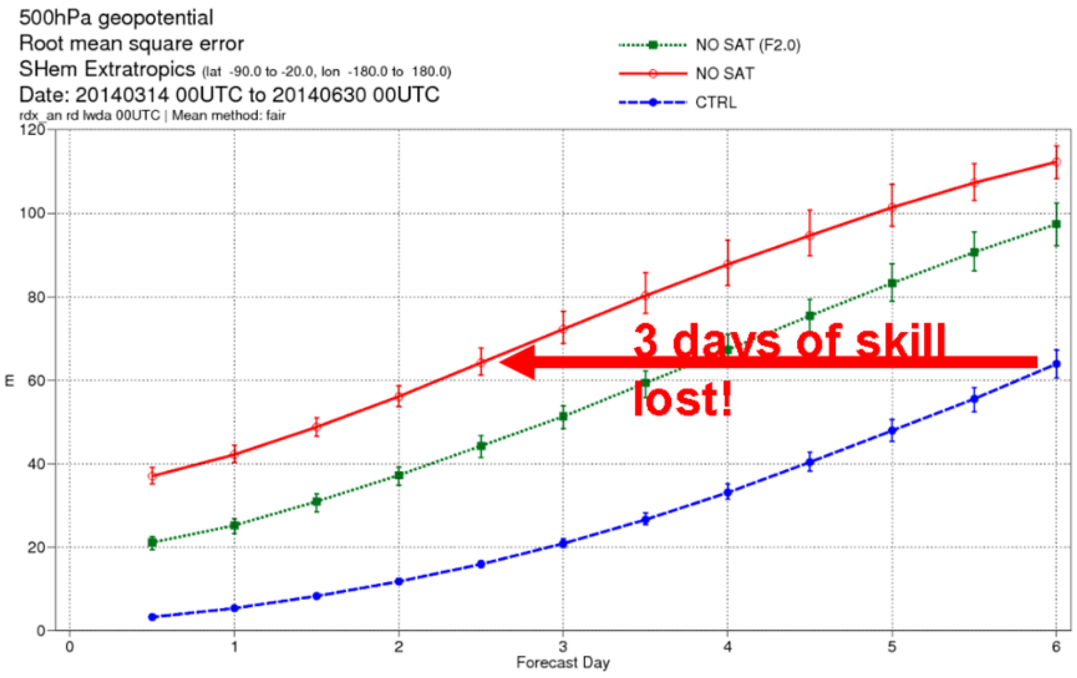


Az adatasszimiláció jelentősége

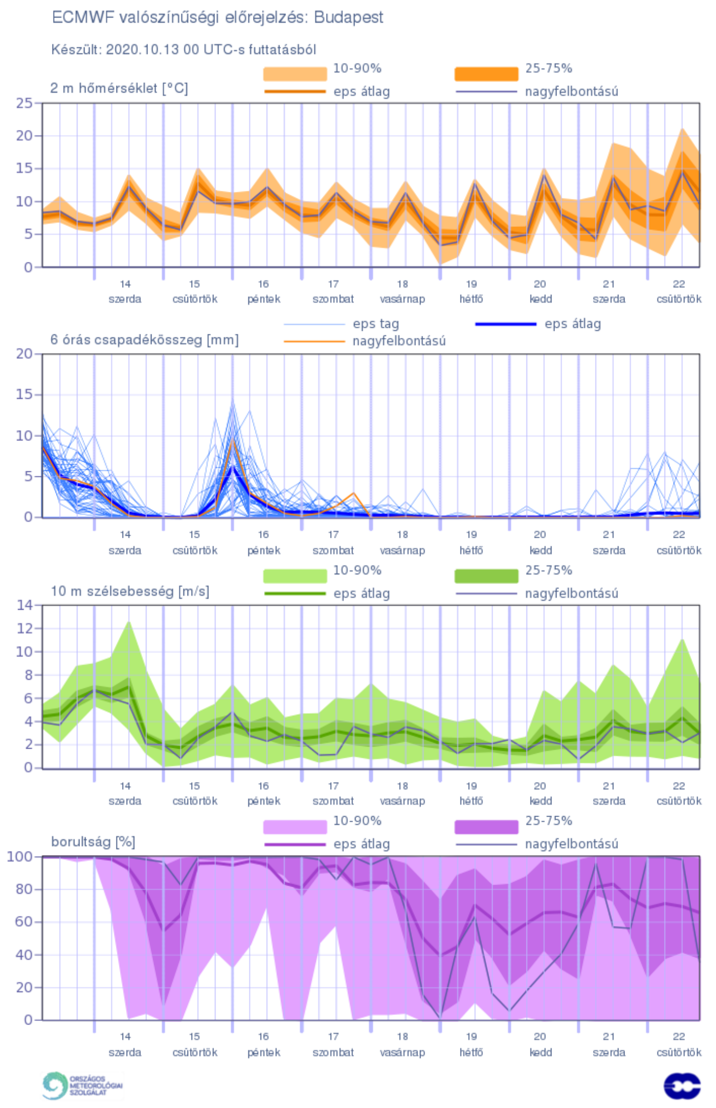
Műholdas mérések hatása

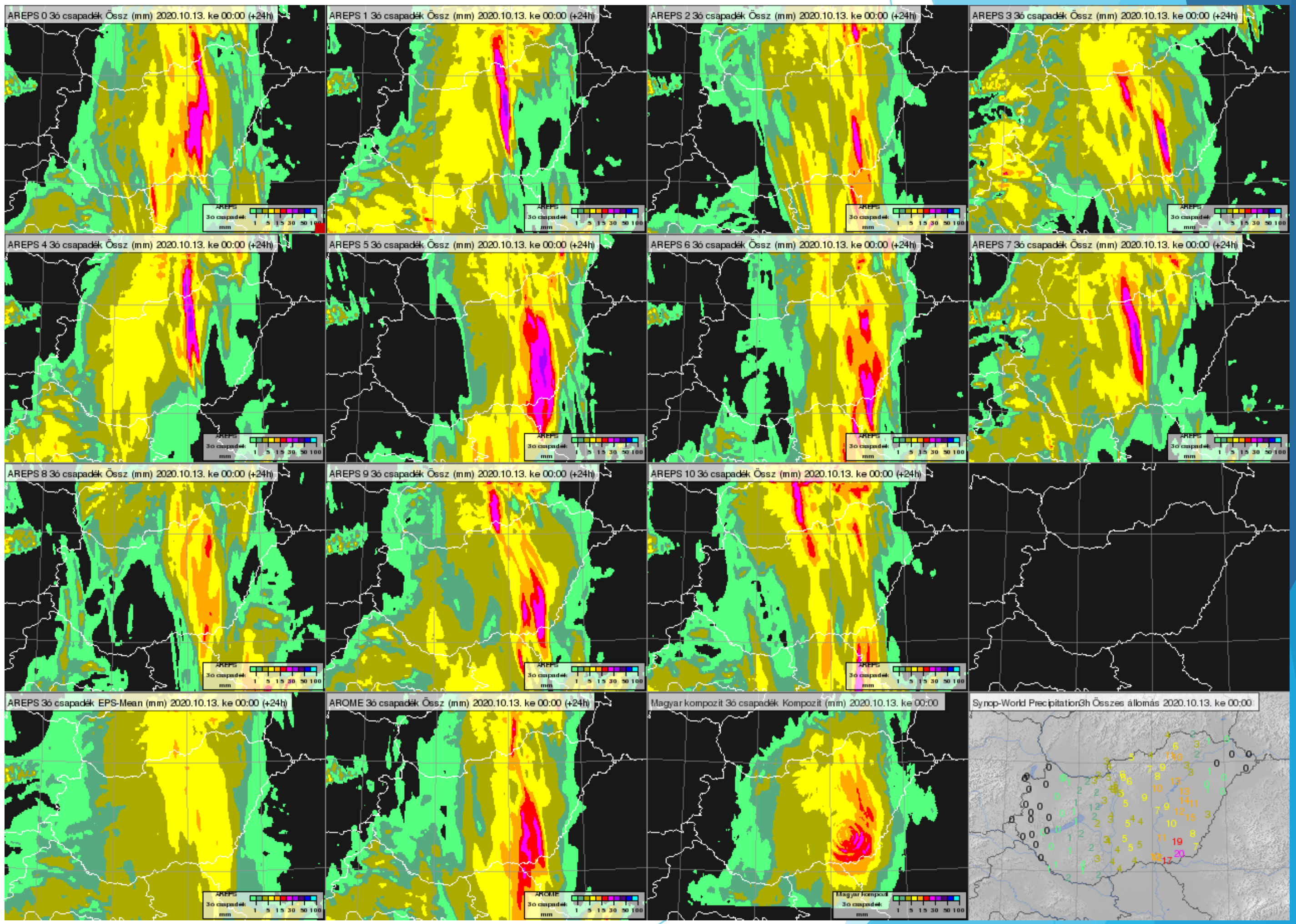
Északi félteke

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?

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)

- Mérések és megfigyelések
- Korábbi előrejelzések
- Tudás a légkörről

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 ε_1 és ε_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:

$\varepsilon_1 = T_1 - T_t$ és $\varepsilon_2 = T_2 - T_t$

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:

$$x_a = f\left(y_1, y_2\right) = ?$$

Több megközelítés létezik, például:

- Legkisebb négyzetek módszere
- Maximum likelihood módszer

Legkisebb négyzetek módszere

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

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

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:

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

Mutassuk meg, hogy

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

1. feladat:

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

$$E(k_1 y_1 + k_2 y_2 - x_t) = 0$$

$$E(k_1(x_t + \varepsilon_1) + k_2(x_t + \varepsilon_2) - x_t) = 0$$

$$E(k_1 x_t + k_1 \varepsilon_1 + k_2 x_t + k_2 \varepsilon_2 - x_t) = 0$$

$$E(k_1 x_t) + E(k_1 \varepsilon_1) + E(k_2 x_t) + E(k_2 \varepsilon_2) - E(x_t) = 0$$

$$k_1 E(x_t) + \cancel{k_1 E(\varepsilon_1)} + k_2 E(x_t) + \cancel{k_2 E(\varepsilon_2)} - E(x_t) = 0$$

Torzítatlanság miatt

$$k_1 E(x_t) + k_2 E(x_t) - E(x_t) = 0$$

$$E(x_t)(k_1 + k_2 - 1) = 0$$

Ez csak akkor 0, ha: $k_1 + k_2 = 1$

2. feladat:

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

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

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

$$\sigma_a^2(k_1) = E\left((x_a - x_t)^2\right) = E\Big(\big(\overbrace{k_1(\underbrace{x_t + \varepsilon_1}_{y_1}) + (1-k_1)(\underbrace{x_t + \varepsilon_2}_{y_2})}^{x_a} - x_t\big)^2\Big)$$

$$= E(\cancel{k_1 x_t} + k_1 \varepsilon_1 \cancel{+ x_t} - \cancel{k_1 x_t} + (1-k_1)\varepsilon_2 \cancel{- x_t})^2)$$

$$= E\left(\left(k_1 \varepsilon_1 + (1-k_1)\varepsilon_2\right)^2\right)$$

$$= k_1^2 E\left((\varepsilon_1)^2\right) + 2k_1(1-k_1)\underbrace{E(\varepsilon_1 \varepsilon_2)}_{0 \text{ (korrelálatlanság miatt)}} + (1-k_1)^2 E\left((\varepsilon_2)^2\right)$$

$$= \boxed{k_1^2 \sigma_1^2 + (1-k_1)^2 \sigma_2^2}$$

min → k_1 szerinti derivált 0

k szerinti derivált: $2k_1\sigma_1^2 - 2(1-k_1)\sigma_2^2 = 0$

$$k_1\sigma_1^2 = (1-k_1)\sigma_2^2$$

$$k_1(\sigma_1^2+\sigma_2^2) = \sigma_2^2$$

$$k_1 = \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2} \Longrightarrow k_2 = 1-k_1 = \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}$$

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

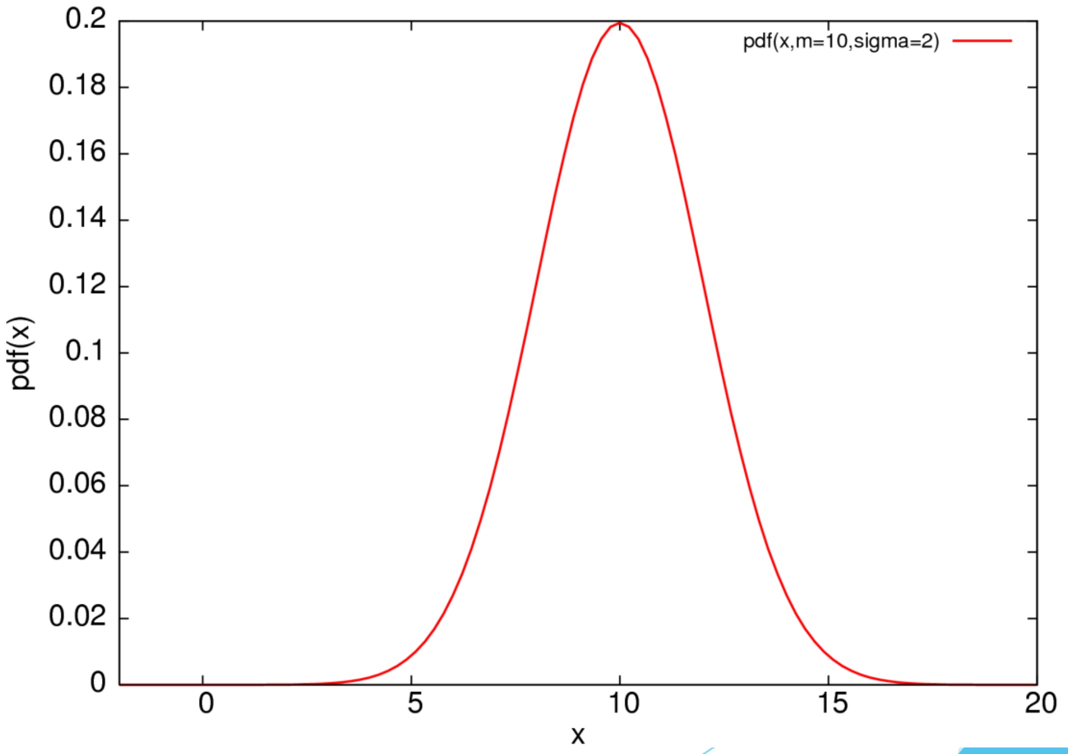
Minden megfigyelés növeli a megbízhatóságot!

Maximum likelihood módszer

A meteorológiai változók „jól” modellezhetők normális eloszlású valószínűségi változókként

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

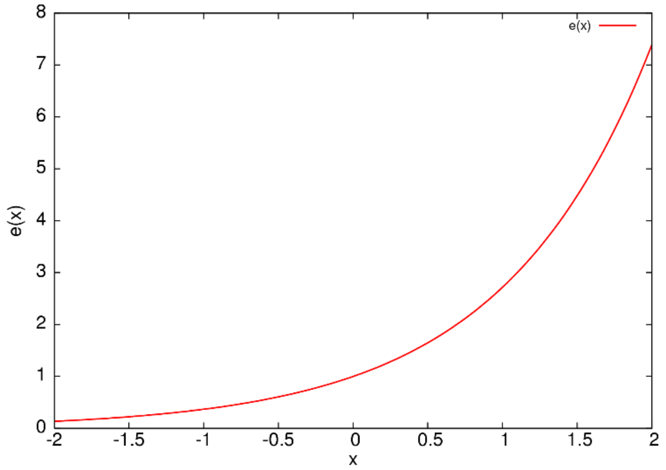
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$$



Az y_1 és y_2 mérések független becslést adnak x_t-re x várható értékkel, σ_1 és σ_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

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

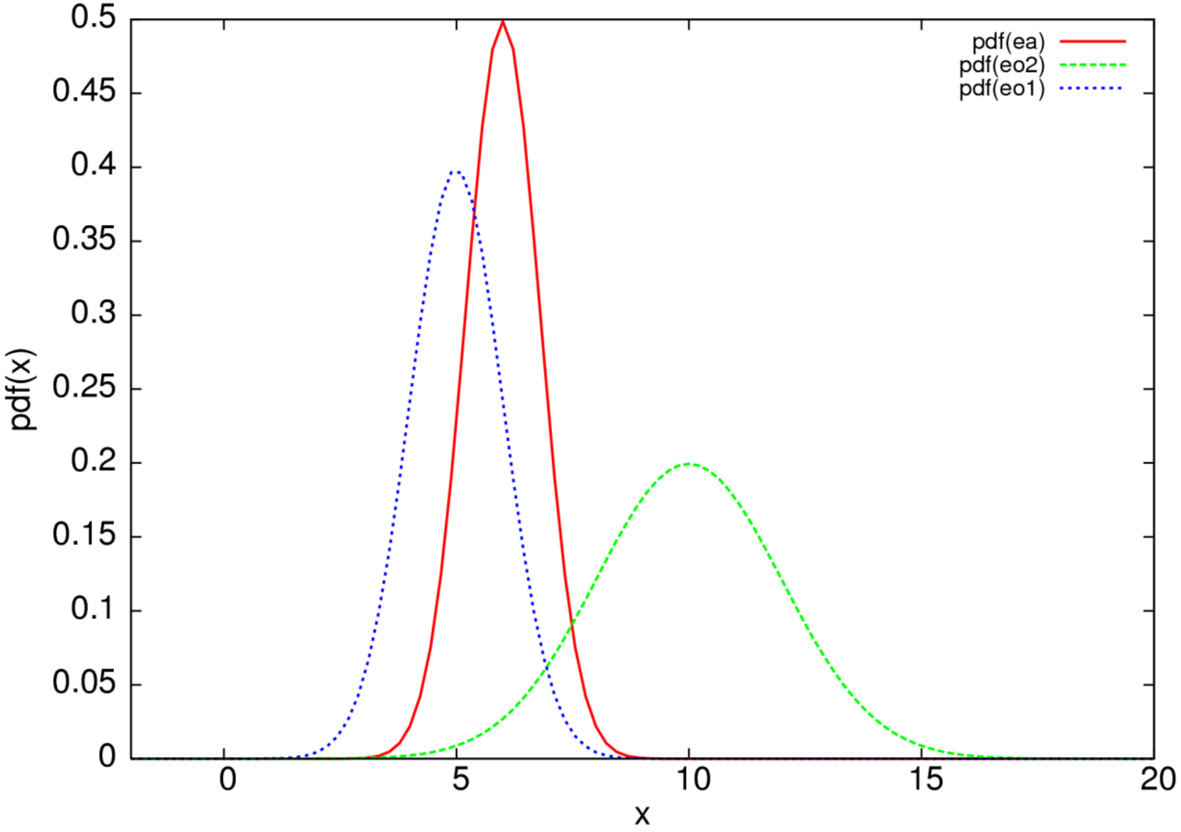
$$= \min_x \left(\frac{1}{2}\frac{(y_1-x)^2}{\sigma_1^2} + \frac{1}{2}\frac{(y_2-x)^2}{\sigma_2^2} \right)$$

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

$$\frac{1}{\sigma_a^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} = \sum_i \frac{1}{\sigma_i^2}$$

$\Longrightarrow$ Minden megfigyelés növeli a megbízhatóságot!

Példa:



y_1=5; σ_1=1

y_2=10; σ_2=2

x_a=6; σ_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)

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

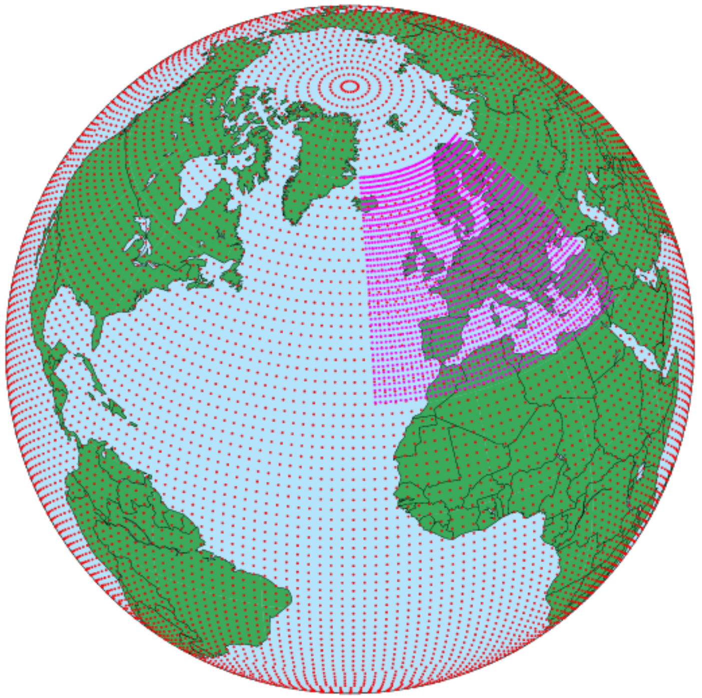
$$\operatorname{argmin} J(x) = ?$$

(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 $\mathbf{x}$)** amely a mai modellek esetében **n~10**^{7} nagyságrendű.

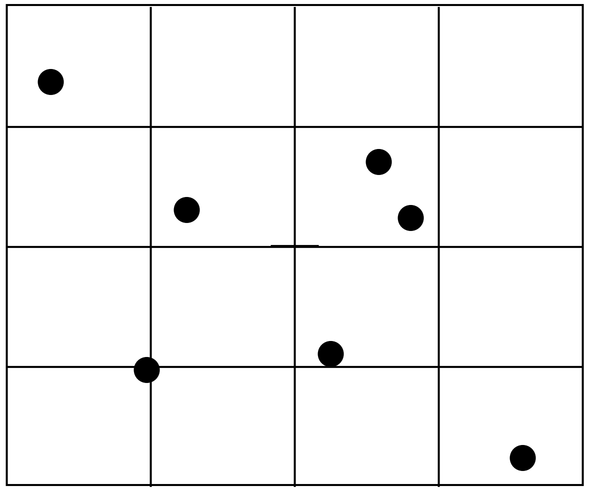
Adatasszimiláció: az állapotvektor megadása a kezdeti időpillanatban (**analízis**) úgy, hogy az minél közelebb legyen a valósághoz.



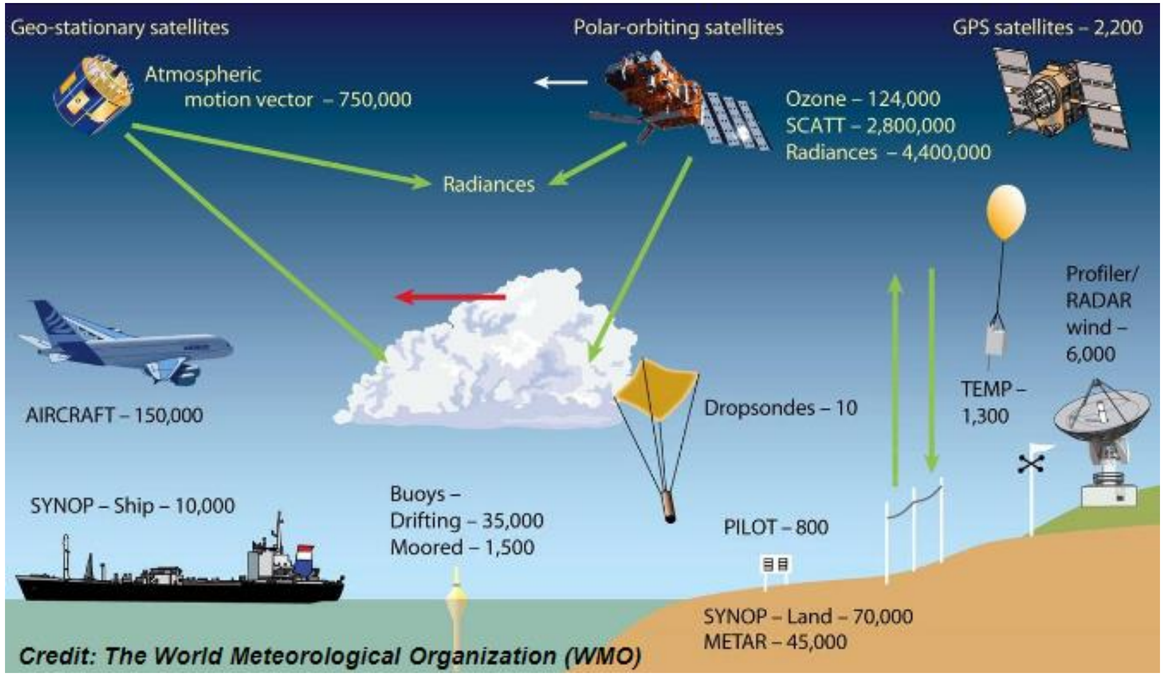
Felhasználható becslések

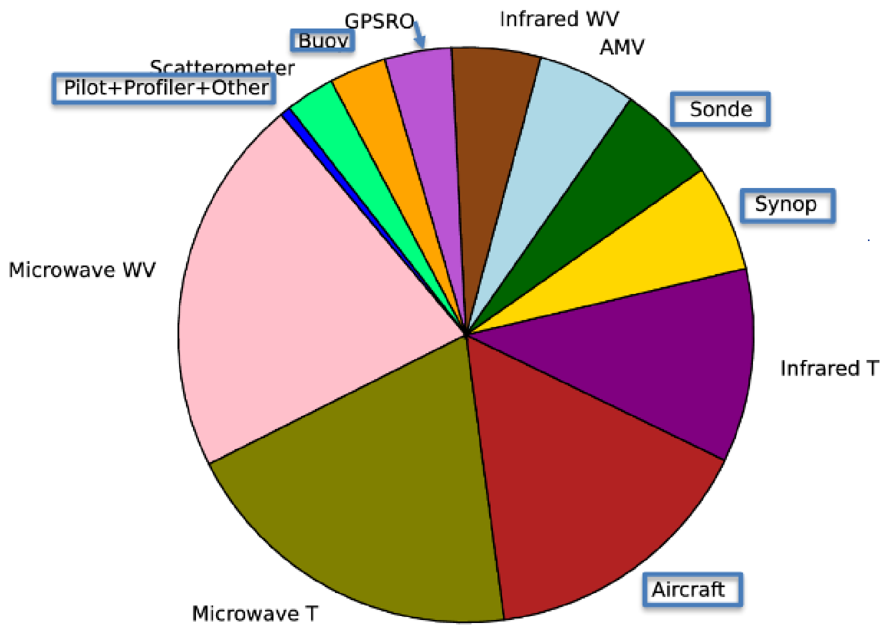
Megfigyelések:

- Időben rendszeres számszerű mérések
- Térben **szabálytalan rács**
- Összetett összefüggés (lehet) a mért és a modell változók között
- Analízisenként $\mathbf{p \sim 10^5}$ nagyságrendű megfigyelést használunk



Megfigyelések globális rendszere





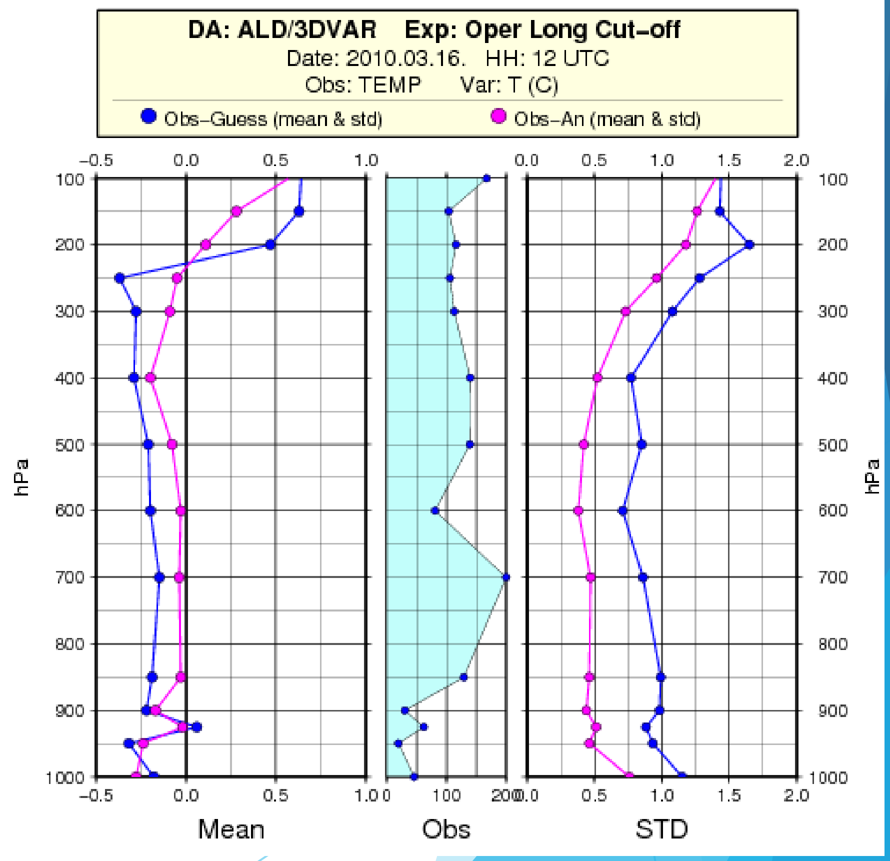
Forrás: Tony McNally, ECMWF NWP training, 2015

24 órás előrejelzés esetén:

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

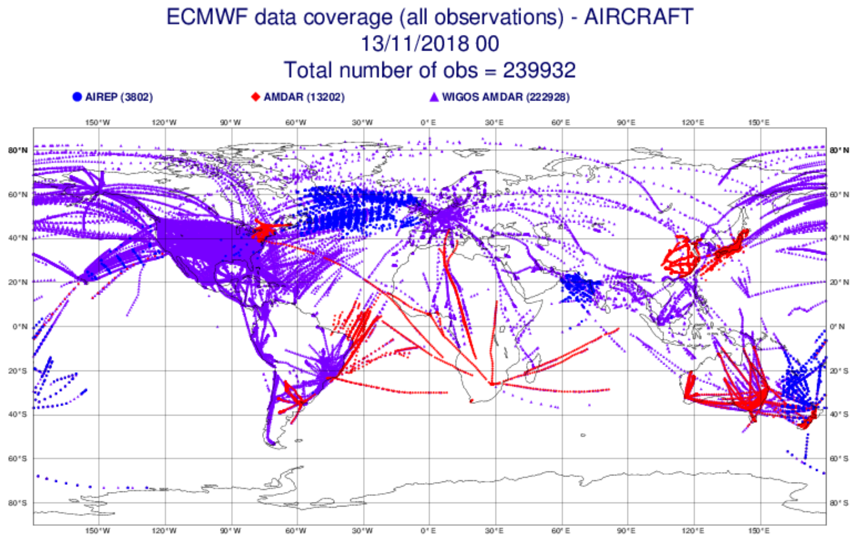
Konvencionális megfigyelések

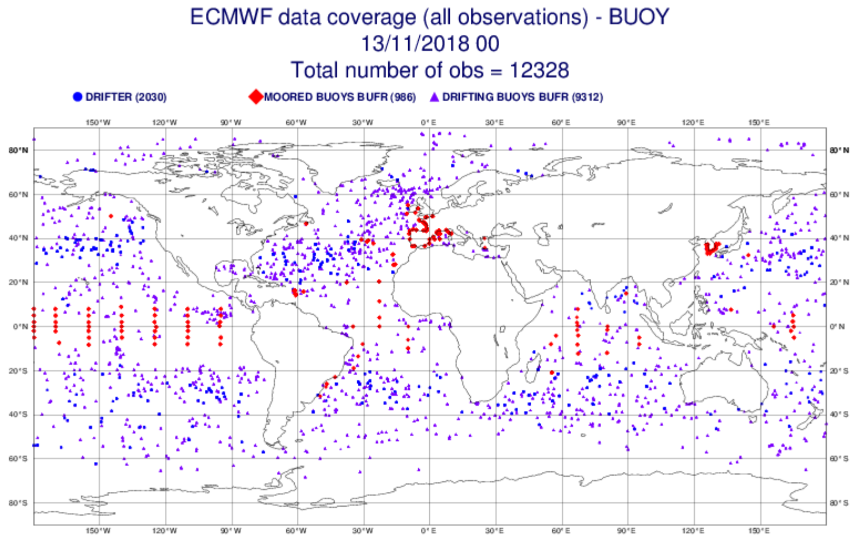
Eszköz	Változó	Szint
SYNOP SHIP METAR	hőmérséklet, harmatpont, szél, légnyomás	Felszín, 2m, 10m, 25m
Bóják	hőmérséklet, légnyomás, szél	2m
Rádiószondák TEMPSHIP Ejtőszondák	hőmérséklet, légnyomás, szél, nedvesség	Profil
Profilerek	szél	Profil
Repülőgépek	hőmérséklet, légnyomás, szél	Profil, repülési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

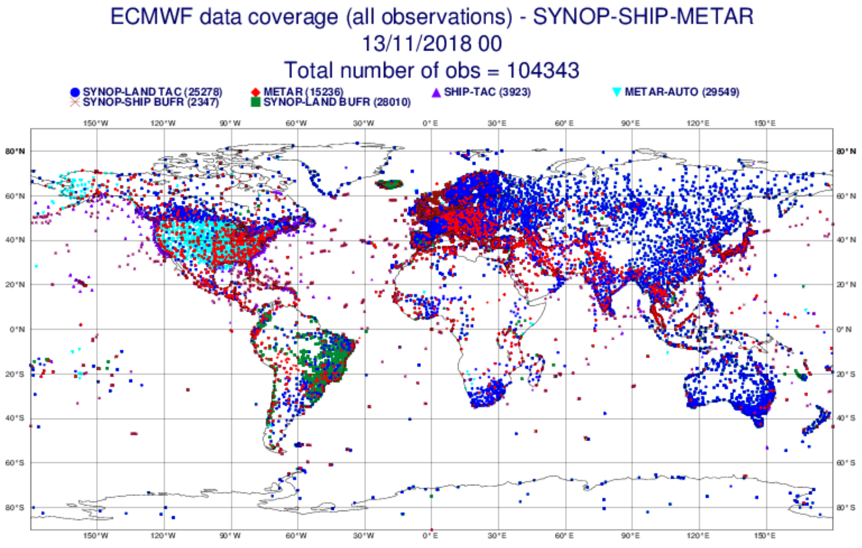


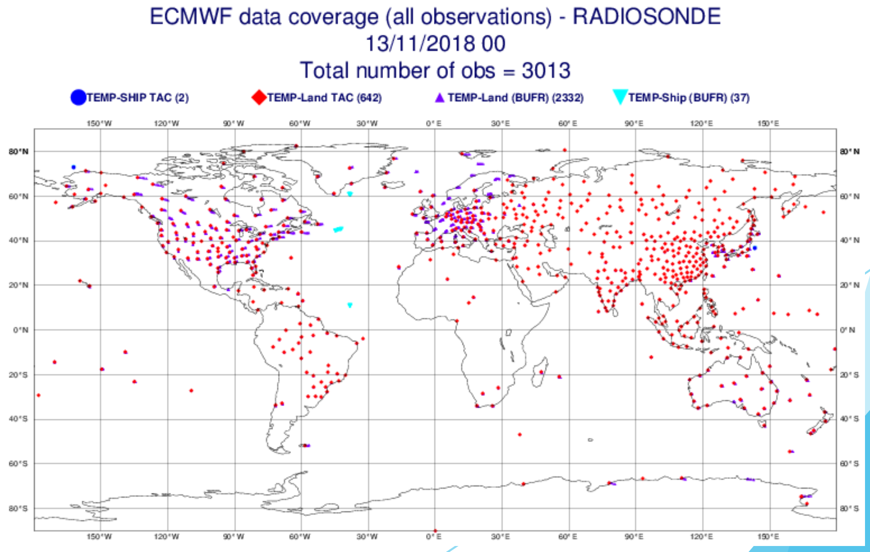
Forrás: Tony McNally, ECMWF NWP training, 2015

Konvencionális/in-situ megfigyelések









In-situ megfigyelések jellemzői

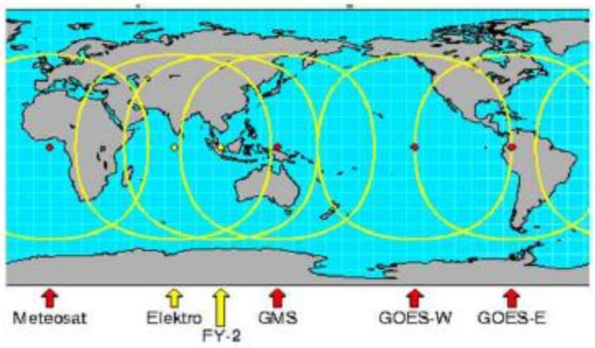
- Nem egyenletes térbeli és időbeli lefedettség
- Pontszerű mérések
- Nem egységes adatminőség

- Közvetlenül azt mérik, amit a modellben is használunk
- Értelmezésük és használatuk az adatasszimiláció során „egyszerű”

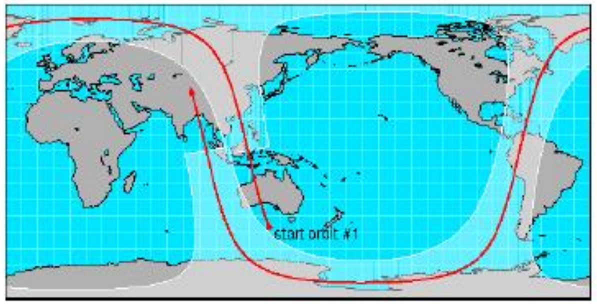
Műholdas mérések

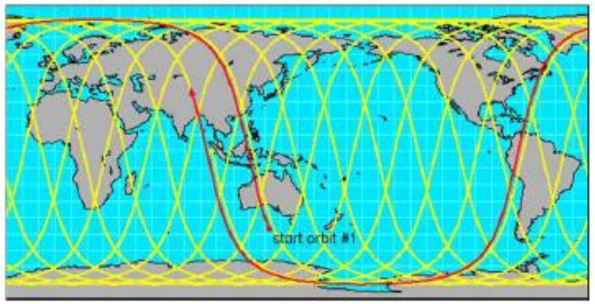
Geostacionárius és kvázipoláris (low Earth orbit) műholdak

GEO



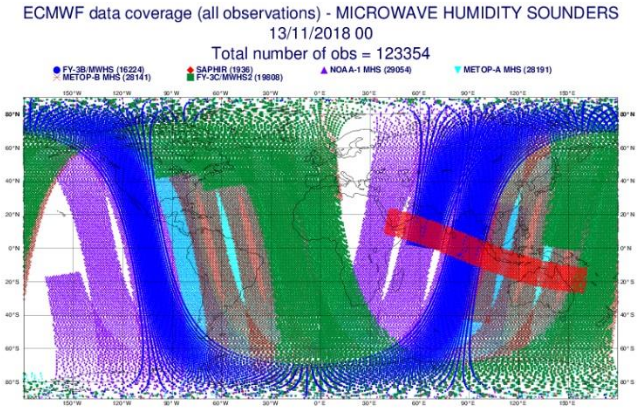
LEO

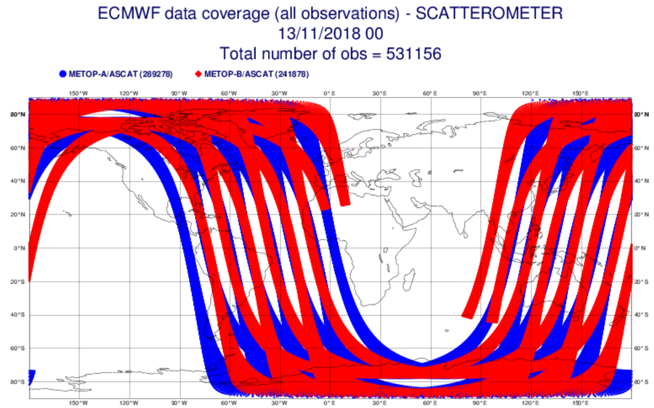


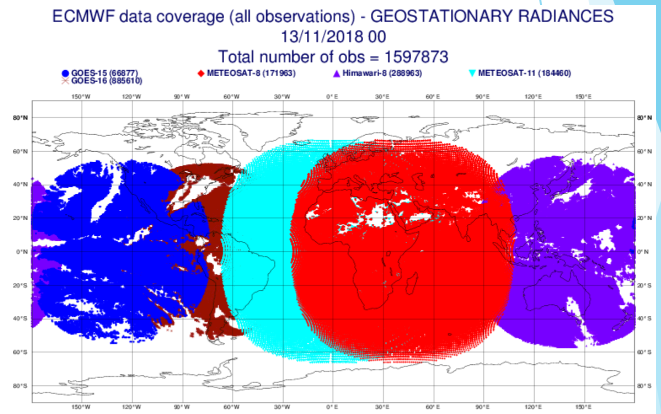


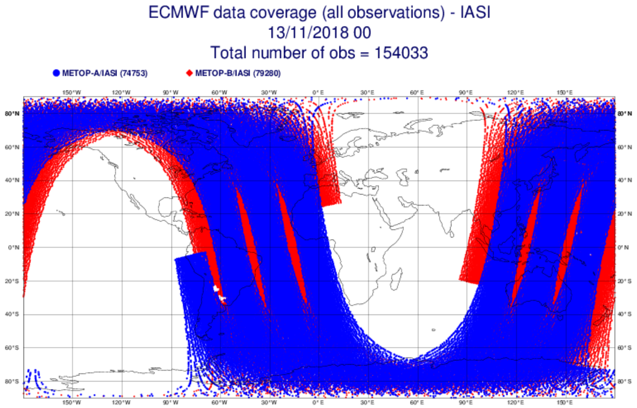
Forrás: Tony McNally, ECMWF NWP training, 2015

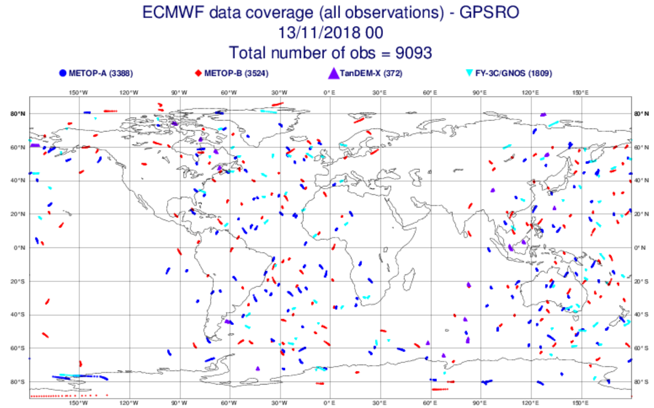
Távérzékeléses mérési lefedettség

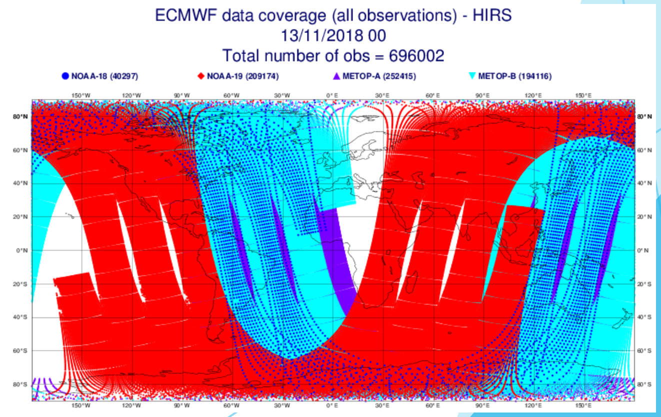












Távérzékelési megfigyelések jellemzői az adatasszimiláció szempontjából

- Nehézség a műholdas adatokkal: 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)
- Hatalmas adatmennyiség

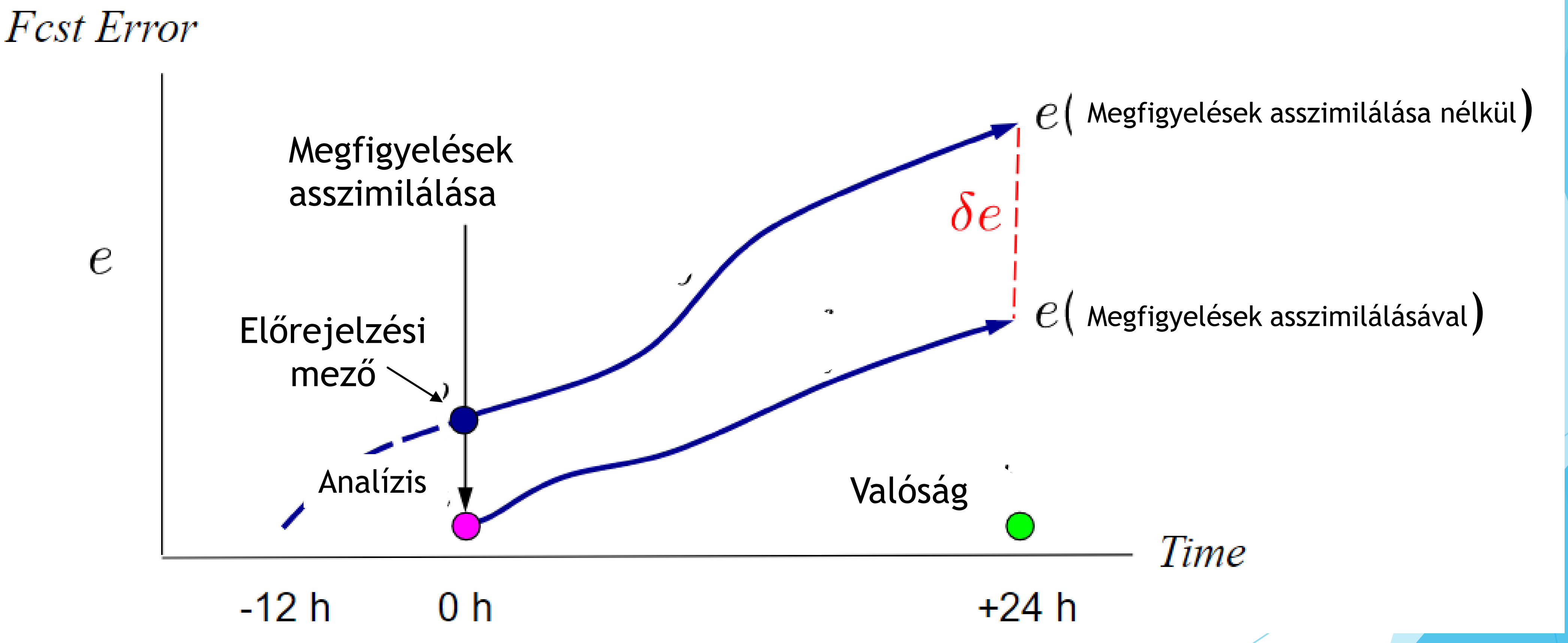
Hátrányok

- Egyenletes, globális térbeli lefedettség
- Nagy időbeli sűrűség
- A műholdas „pixelek” közelebb állnak a modellrácshoz

Előnyök

Adatasszimiláció hatékonyságának az ellenőrzése

FSOI - Forecast Sensitivity Observation Impact



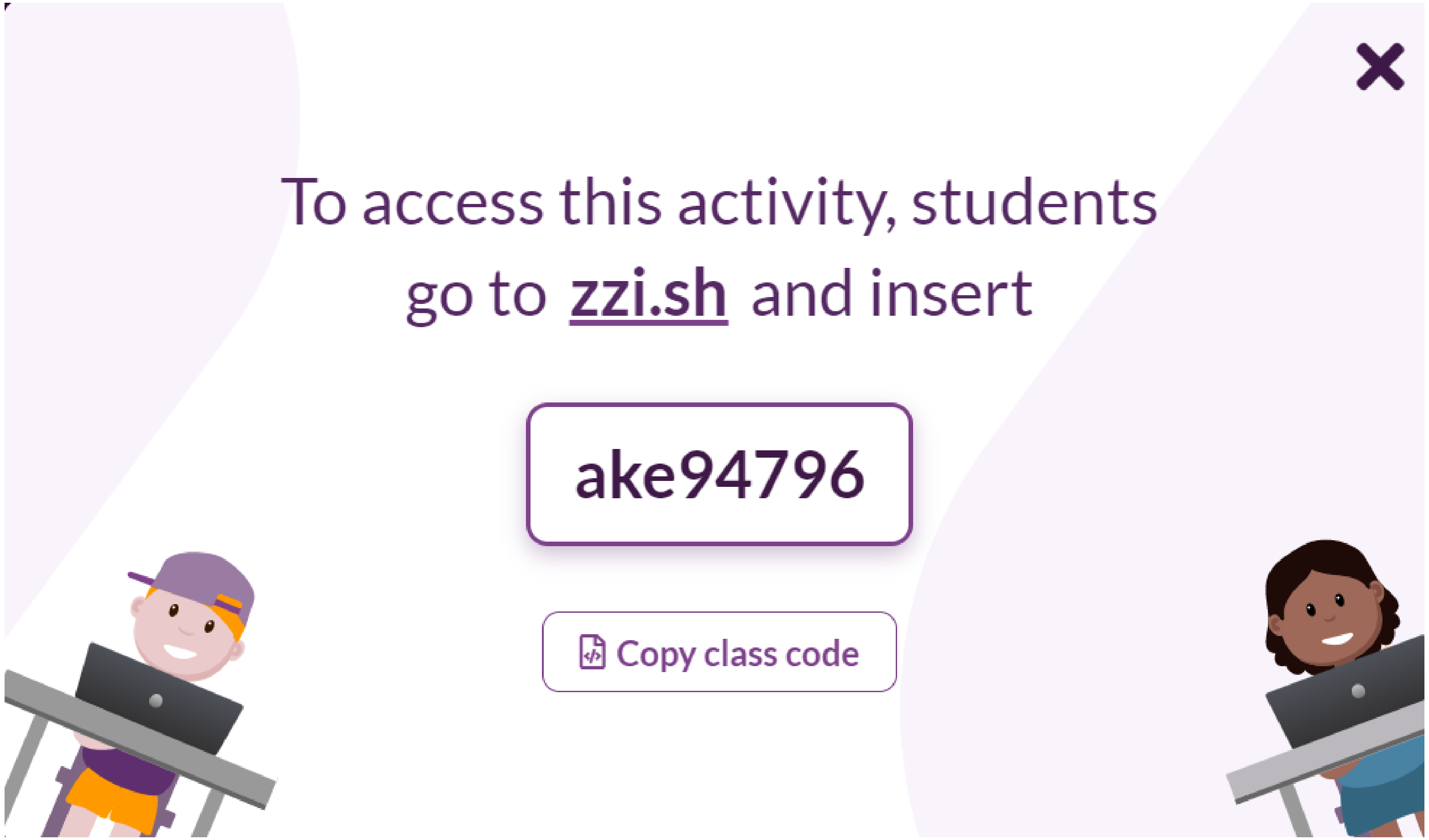
Felhasználható becslések-folytatás

Megfigyelések:

- A megfigyelések nem feltétlenül esnek a modellrácspontokba
- Nem közvetlenül a modellezett mennyiségeket mérik
- A megfigyelések száma ma is kevesebb, mint a modellrács pontjainak száma

Background/first guess:

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


To access this activity, students go to zzi.sh and insert
ake94796
Copy class code