

# Ensemble előrejelzések II.

***„No forecast is complete without the forecast of forecast skill”***

*Tennekes, Baede, Opsteegh, 1986*

Előadó: Jávorné Radnóczy Katalin ([radnoczi.k@met.hu](mailto:radnoczi.k@met.hu))

*Korábbi előadók: Hágel Edit, Horányi András,  
Bölöni Gergely, Szűcs Mihály*

# Az előző óra tartalma

- A kaotikus rendszerekről

- Egydimenziós diszkrét példa
- A Lorenz-modellek és a vízikerek
- Kaotikus rendszerek tulajdonságai

## A légköri rendszer előrejelezhetősége

- Mik a légköri rendszer előrejelzésében a bizonytalanság forrásai?
- Mik az időjárás előrejelezhetőségének határai?

- Együttes (ensemble) előrejelzések

- Ensemble előrejelzés alapfogalmai
- Milyen valószínűségi produktumok készíthetők?
- Hogyan használhatók ezek az előrejelzések?
- Hogyan lehet a légköri bizonytalanságokat számszerűsíteni? - Perturbációs módszerek
- Mik a jó ensemble előrejelzés tulajdonságai? - Verifikáció
- Konkrét példák ensemble előrejelző rendszerekre

# Ismétlés

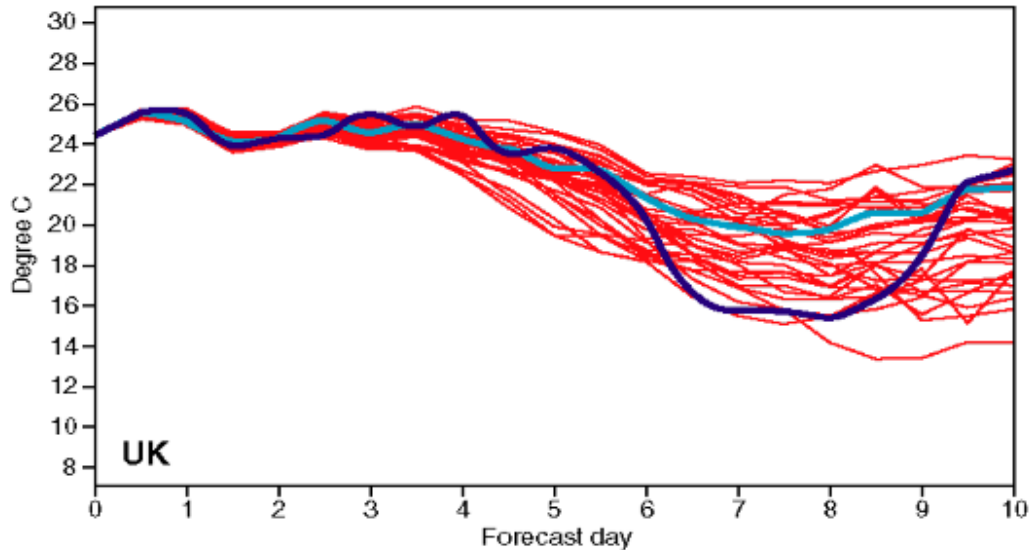
## A légkör előrejelezhetősége

- A légkör a kaotikus rendszerek jellemzőit mutatja
- Az előrejelezhetőség függ attól, hogy:
  - Hova, milyen paramétert akarunk előrejelezni? Milyen skálájú folyamatot akarunk leírni?
  - Milyen kezdeti analízis áll rendelkezésünkre? Milyen modellt használunk?
  - Milyenek a légköri feltételek? - Áramlásfüggő!

ECMWF ensemble forecast – Air temperature

Date: 26/06/1995 London Lat: 51.5 Long: 0

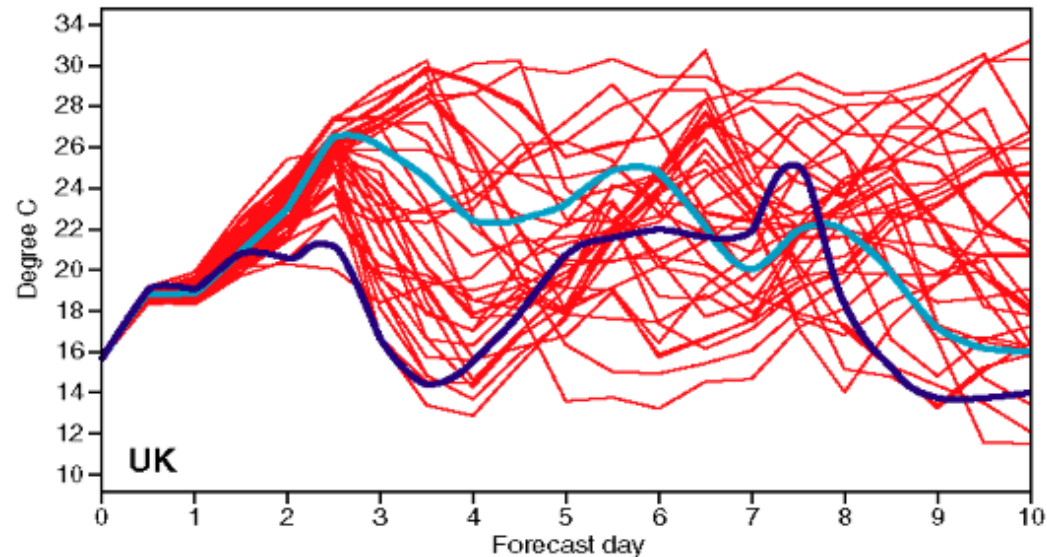
Control Analysis Ensemble



ECMWF ensemble forecast – Air temperature

Date: 26/06/1994 London Lat: 51.5 Long: 0

Control Analysis Ensemble

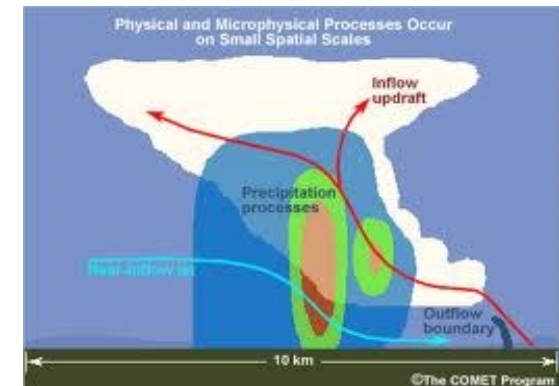
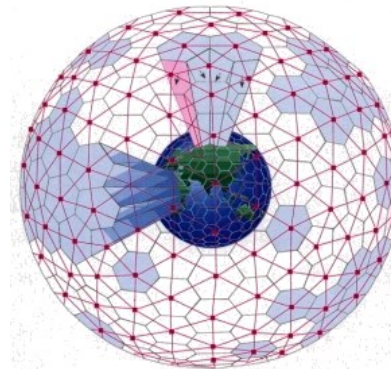
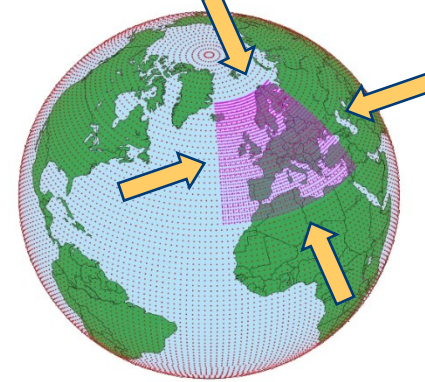
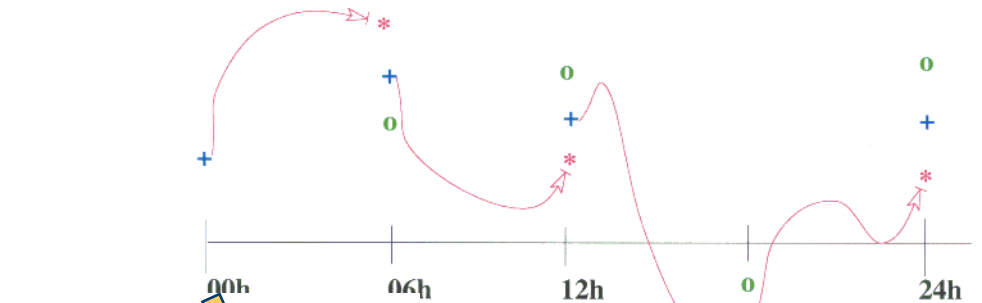


# Ismétlés

## A numerikus előrejelzés bizonytalanságai

- A légköri előrejelzések hibáinak forrása a gyakorlatban:

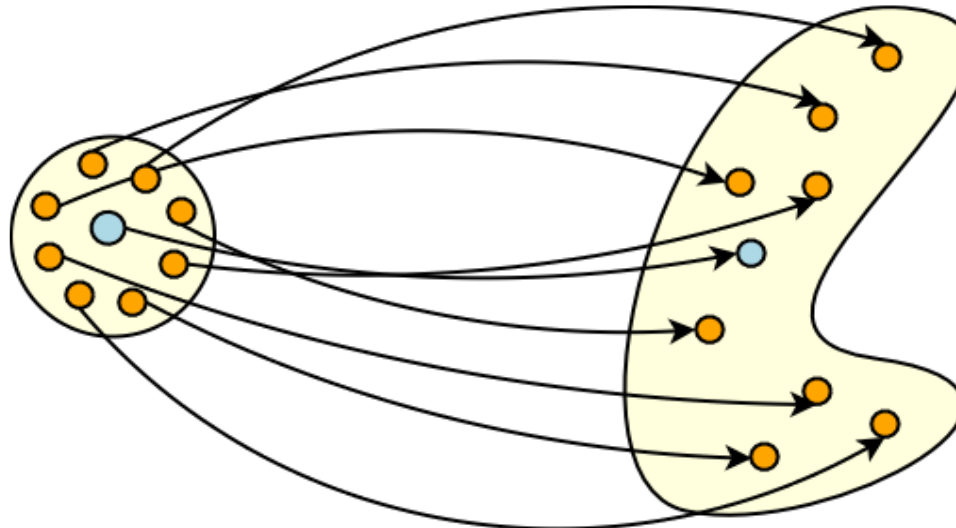
- **Kezdeti feltételek**
- **Peremfeltételek** (korlátos tartományú modelleknél oldalsó peremfeltételek)
- **Felszín** (alsó peremfeltétel, pl.: talajhőmérséklet, talajnedvesség, hóvastagság)
- A **HTDER diszkretizálása** (séma, koordináta rendszer, térképvetület, felbontás, időlépcső stb. megválasztása)
- **Fizikai parametrizáció** (a modell felbontásánál kisebb skálájú jelenségek leírása)



# Ismétlés

## EPS motiváció

- Megoldás az előrejelzések során fellépő bizonytalanságok számszerűsítésére:
    - **Nem egyetlen** előrejelzést készítünk a „legjobbnak ítélt” kezdeti feltételből kiindulva, a „legjobbnak ítélt” módszerrel.
    - **Előrejelzések együttesét** készítjük, melyek a kezdeti feltételükben különböznek csekély mértékben (a bizonytalansági határon belül), vagy a modellintegrálás során használt módszerek eltérőek.
- ⇒ valószínűségi megközelítés lehetségessé válik  
következtetni lehet az előrejelzés bizonytalanságára



# Tartalom

- A kaotikus rendszerekről
  - Egydimenziós diszkrét példa
  - A Lorenz-modellek és a vízikerék
  - Kaotikus rendszerek tulajdonságai
- A légköri rendszer előrejelezhetősége
  - Mik a légköri rendszer előrejelzésében a bizonytalanság forrásai?
  - Mik az időjárás előrejelezhetőségének határai
- Együttes (ensemble) előrejelzések
  - Ensemble előrejelzés alapfogalmai
  - Milyen valószínűségi produktumok készíthetők?
  - Hogyan használhatók ezek az előrejelzések?
  - [Hogyan lehet a légköri bizonytalanságokat számszerűsíteni? - Perturbációs módszerek](#)
  - [Mik a jó ensemble előrejelzés tulajdonságai? - Verifikáció](#)
  - [Konkrét példák ensemble előrejelző rendszerekre](#)

# Ensemble előrejelzések készítésének módszerei

**Milyen az a kis eltérés, amit az optimális kezdeti feltételhez hozzá kell adni, hogy a későbbi időlépcsőkben jól leírja a bizonytalanságokat?**

- Ensemble előrejelzések készítésére több módszer is alkalmazható (a cél, hogy az előrejelzés készítése során fellépő **összes bizonytalanságot számszerűsítsük**):
  - Kezdeti feltétel perturbációk származtatása:

”Kis eltérés a kezdeti feltételben nagy eltéréseket eredményez az előrejelzés során” elv alapján először a kezdeti feltételeket kezdték perturbálni.
  - Modellhiba reprezentációja:

A modell felépítésével összhangban lévő módszereket kerestek a modell bizonytalanságainak számszerűsítése céljából.
  - Egyéb, pragmatikus eszközök (multi-módszerek):

Trükkös és hasznos, ám kevésbé tudományos módszerek.

# Ensemble előrejelzések készítésének módszerei

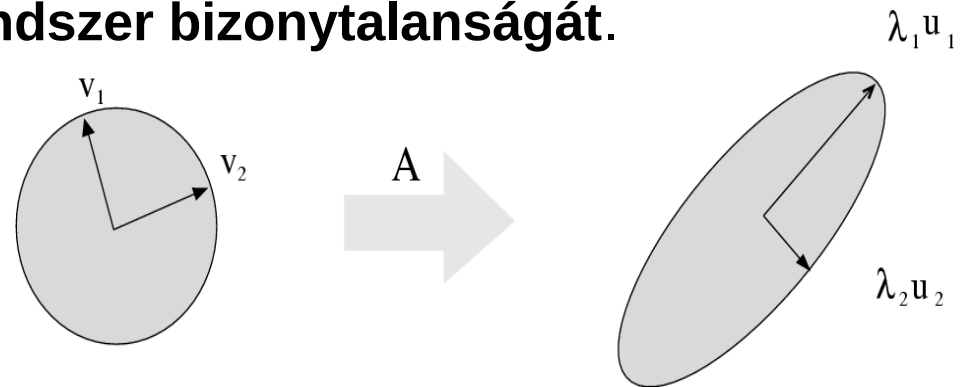
- A kezdeti feltétel perturbációk származtatása
  - Szinguláris vektorok módszere
  - Breeding (tenyésztéses) módszer
  - Ensemble adatasszimilációs módszer
- A modellhiba reprezentációja
  - Multi-fizika módszer
  - Paraméter perturbációk
  - Sztochasztikus fizika módszerek: SPPT, SPP
- Multi-módszerek
  - Multi-modell
  - Multi-analízis
  - Multi-LBC



# Szinguláris vektorok

## Motiváció

- Milyen az optimális kezdeti feltétel perturbáció?
  - A léggöri állapotváltozókat tartalmazó vektor meglehetősen nagy. Nincs kapacitásunk minden változót külön-külön perturbálni, és ilyen módon egy hatalmas tagszámú ensemble-t futtatni.
  - Helyette meg kell találni azokat az irányokat a fázistérben, melyek a legnagyobb bizonytalanságot hordozzák. Olyan **optimális perturbációkat** keresünk, melyek időben a lehető **leggyorsabban fejlődnek** és minél **nagyobb mértékben reprezentálják az egész léggöri rendszer bizonytalanságát**.



- Az egyik első módszer perturbációk előállítására (ECMWF, 1992)
- A szinguláris vektorok elve jó összhangban van a léggör kaotikus jellegével: “Kis eltérés a kezdeti feltételekben nagy eltérést eredményezhet az eredményben” - Hogy tudjuk optimálisan meghatározni a kezdeti kis eltérést?

# Szinguláris vektorok perturbációk fejlődése

- Tekintsük az  $n$ -dimenziós dinamikai rendszer nemlineáris fejlődését leíró közönséges differenciálegyenlet-rendszert:

$$\frac{dX}{dt} = F(X)$$

$X$ : a légkör állapotát leíró vektor  
 $F$ :  $X$  időbeli fejlődését leíró modell

- $x(t)$  az  $X(t_0)=x(t_0)$  kezdeti feltételhez tartozó megoldás („trajektória”).
- Módosítsuk a kezdeti feltételeket egy infinitezimális  $y(t_0)$  perturbáció hozzáadásával. Az ehhez tartozó megoldás:  $x(t) + y(t)$
- Ha a perturbáció elegendően kicsi, akkor  $F$  közelíthető a Taylor polinomjával:

$$F(x(t) + y(t)) = F(x(t)) + \frac{dF}{dX} y(t) + O(y(t)^2)$$

Megj.: a lineáris közelítés ált. +12-48 óráig érvényes

# Szinguláris vektorok perturbációk fejlődése

- Az előző két egyenlet alapján a perturbációk fejlődésére az alábbi **tangens-lineáris egyenlet** adódik:

$$\frac{dy(t)}{dt} = \frac{dF}{dX} y(t)$$

- Megjelenik  $F$  Jacobi-mátrixa.
- A tangens-lineáris egyenlet általános megoldása a következő:

$$y(t_1) = M(t_0, t_1) y(t_0)$$

- $M(t_0, t_1)$  mátrixot rezolvens-, hiba- vagy **propagátor-mátrixnak** nevezzük, ez írja le a perturbációk fejlődését.
- $M$  függ  $t_1$  megválasztásától

# Szinguláris vektorok

## a leggyorsabban fejlődő perturbációk

- A tangens-lineáris egyenlet megoldása egyszerűbben:

$$y(t_1) = My(t_0)$$

- Keressük azokat a perturbációkat, amelyeket  $M$  a lehető legnagyobb mértékben megnövel a  $(t_0, t_1)$  időszakon
- A perturbációk méretének meghatározására valamilyen  $N$  **normát** használunk:

$$\frac{\|y(t_1)\|_{N_1}}{\|y(t_0)\|_{N_2}} = \frac{\|My(t_0)\|_{N_1}}{\|y(t_0)\|_{N_2}}$$

- Ennek a hányadosnak keressük a maximumát a továbbiakban, amely függ attól, milyen normát választunk.

# Szinguláris vektorok a norma megválasztása

- A normát skalárszorozattal definiáljuk:

$$\|y\|_N = \sqrt{\langle y, y \rangle_N}$$

- Az euklideszi norma ebben az esetben nem megfelelő, mert az  $y$  perturbációs vektor különböző fizikai változókat tartalmaz (hőmérséklet, nedvesség, nyomás), így pl. a hőmérséklet értéktartománya miatt dominánsabb lenne, mint pl. a specifikus nedvesség. A normának tehát fizikai jelentést adnak:
- Az **energia normák** használata terjedt el, pl.:
  - teljes energia norma: mozgási + helyzeti + belső energia
  - CAPE norma
  - célterületre vett norma (pl. trópusi öv)

# Szinguláris vektorok

## a leggyorsabban fejlődő perturbációk

- Térjünk vissza a perturbációk vizsgálata során már látott és maximalizálni kívánt hányadoshoz:

$$\frac{\|My(t_0)\|_{N_1}}{\|y(t_0)\|_{N_2}}$$

- Válasszuk ugyanazt a normát a kezdeti- és végidőpontban egyaránt (ez a gyakorlatban nem mindig teljesül)!
- Mostantól elegendő az  $M$  propagátor  $\|y(t_0)\|=1$  egységvektoron felvett  $\|My(t_0)\|$  értékeit vizsgálni, keresni ezek maximumát

# Szinguláris vektorok

## a leggyorsabban fejlődő perturbációk

- $\|My(t_0)\|$  kifejtése a norma és skaláris szorzat kapcsolata, illetve az adjungált definíciója alapján

$$\|My(t_0)\| = \sqrt{\langle My(t_0), My(t_0) \rangle} = \sqrt{\langle M^* My(t_0), y(t_0) \rangle}$$

- $M^*M$  önadjungált, hiszen:

$$(M^* M)^* = M^* (M^*)^* = M^* M$$

- Ebből az következik, hogy  $M^*M$ -nek létezik  $v_1, \dots, v_n$ , az adott skaláris szorzatra ortonormált sajátvektor-rendszere, amihez tartozó (valós) sajátértékek:

$$\lambda_1^2 \geq \dots \geq \lambda_n^2$$

# Szinguláris vektorok a leggyorsabban fejlődő perturbációk

- A sajátvektorokra

$$\begin{aligned}\|Mv_i(t_0)\|^2 &= \langle M^* Mv_i(t_0), v_i(t_0) \rangle = \langle \lambda_i^2 v_i(t_0), v_i(t_0) \rangle = \\ &= \lambda_i^2 \|v_i(t_0)\|^2 = \lambda_i^2\end{aligned}$$

(megj.: a norma mindig pozitív szám  $\Rightarrow \lambda_i^2$  alakú felírás)

- Így  $M^*M$  sajátvektoraira a norma változása

$$\frac{\|Mv_i(t_0)\|}{\|v_i(t_0)\|} = \frac{\|v_i(t_1)\|}{\|v_i(t_0)\|} = \lambda_i$$



# Szinguláris vektorok

## jobb oldali szinguláris vektorok

- A kapott  $n$   $\lambda_1 \geq, \dots, \geq \lambda_n$  értékeket az  $M$  **szinguláris értékeinek** nevezzük, a hozzá tartozó  $v_1, \dots, v_n$  vektorokat pedig az  $M$  **jobb oldali** (előre irányuló) **szinguláris vektorainak**.
- MEGJEGYZÉSEK:
  - Az adjungált függ a normától, azaz a szinguláris értékek és vektorok is normafüggőek
  - Megmutatható, hogy  $\|My(t_0)\|$ -nak  $\lambda_1$  a maximuma („vezető szinguláris érték”), amelyet a  $v_1$  vektoron vesz fel, azaz  $v_1$  a leggyorsabban fejlődő egységnyi perturbáció („vezető szinguláris vektor”)

# Szinguláris vektorok

## jobb oldali szinguláris vektorok

### MEGJEGYZÉSEK:

A  $v_1$ -hez tartozó ortogonális kiegészítő altéren  $\lambda_2$  lesz a maximum és a  $v_2$  vektor adja meg a maximum helyét. Ha a  $v_1$  és  $v_2$  által generált altér ortogonális kiegészítő alterét tekintjük, akkor  $\lambda_3$  lesz a maximum, és a  $v_3$  vektor adja meg a maximum helyét. Tovább folytatva a sort, eljutunk egészen  $\lambda_n$ -ig, ami  $\|My(t_0)\|$  minimumát adja meg a  $v_n$  vektoron.

A jobb oldali szinguláris vektorok olyan lényegesen különböző (ortogonális) egységnyi perturbációkat adnak meg, amelyek a lehető legnagyobb mértékben növekednek (optimális perturbációk). A növekedés értéktartományát a szinguláris értékek jelölik ki.

# Szinguláris vektorok

## bal oldali szinguláris vektorok

- Vegyük a következő összefüggést:

$$MM^*(Mv_i) = M(M^*Mv_i) = M\lambda_i^2 v_i = \lambda_i^2 Mv_i$$

- Az  $Mv_i$  vektorok az  $MM^*$  mátrix  $\lambda_i^2$  sajátértékeihez tartozó sajátvektorai, azaz a  $MM^*$  és az  $M^*M$  mátrixok sajátértékei megegyeznek
- Vezessük be az  $MM^*$  mátrix normált  $(Mv_i)$  sajátvektorait:

$$u_i = \frac{Mv_i}{\|Mv_i\|}$$

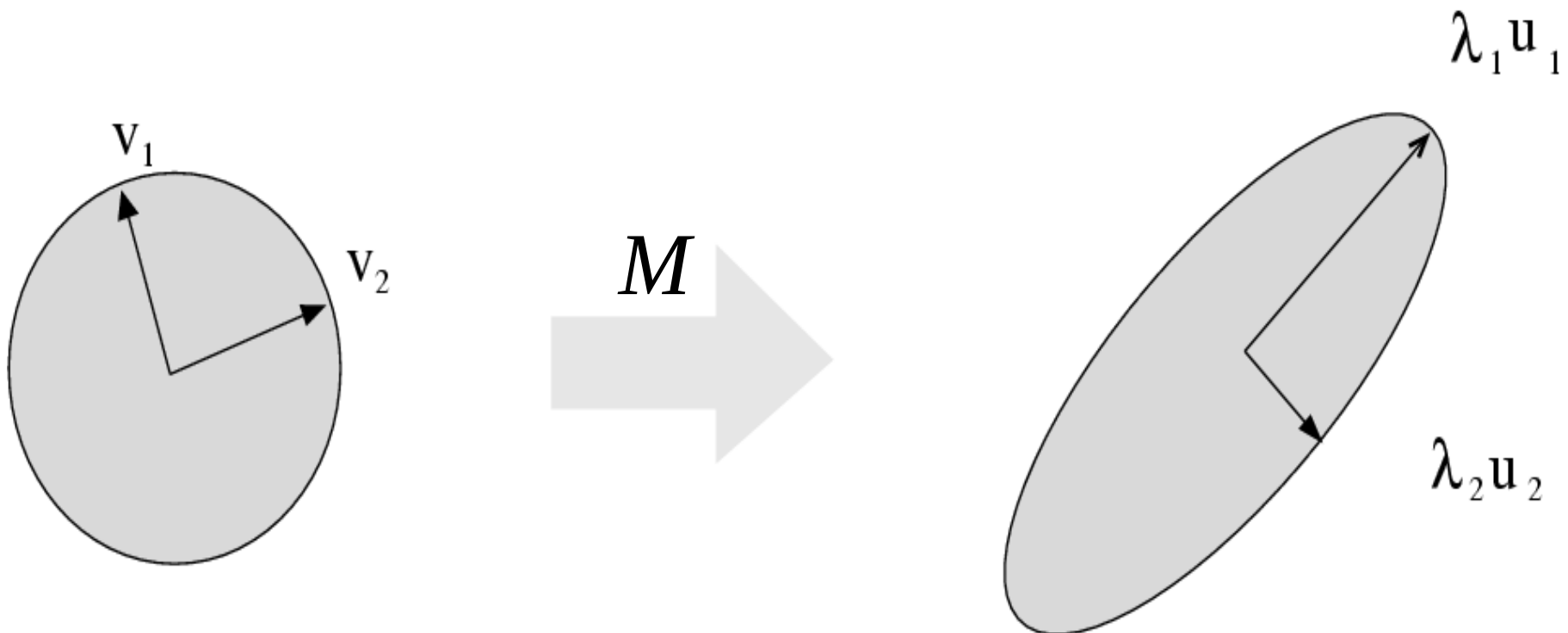
# Szinguláris vektorok

## bal oldali szinguláris vektorok

- Mivel  $\|Mv_i\|^2 = \langle Mv_i, Mv_i \rangle = \langle v_i, M^* Mv_i \rangle = \langle v_i, \lambda_i^2 v_i \rangle = \lambda_i^2$
- Az előző kettőből adódik:  $Mv_i = u_i \|Mv_i\| = \lambda_i u_i$
- Az  $M$  propagátor a jobb oldali szinguláris vektorokat az  $u_1, \dots, u_n$  vektorok irányába képezi le
- $u_1, \dots, u_n$  vektorokat a  $M$  **bal oldali** (visszafelé irányuló) **szinguláris vektorainak** nevezzük
- Azokat a lényegesen különböző perturbációkat jellemzik a vizsgált időtartam végén, amelyek a lehető legnagyobb mértékben növekedtek

# Szinguláris vektorok

Szemléltetés 2 dimenzió esetén:

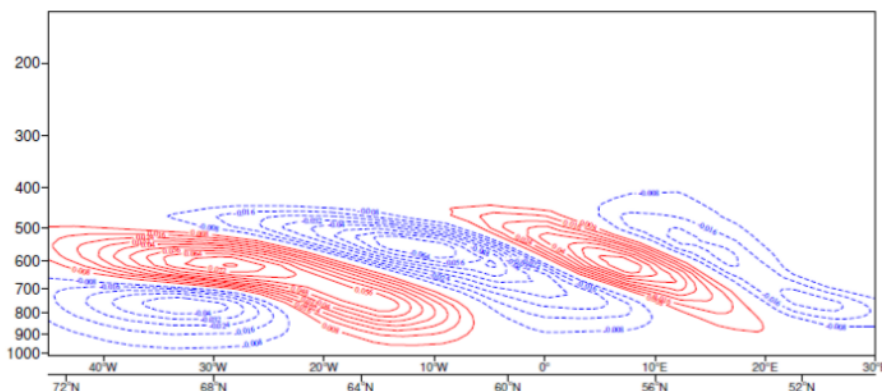
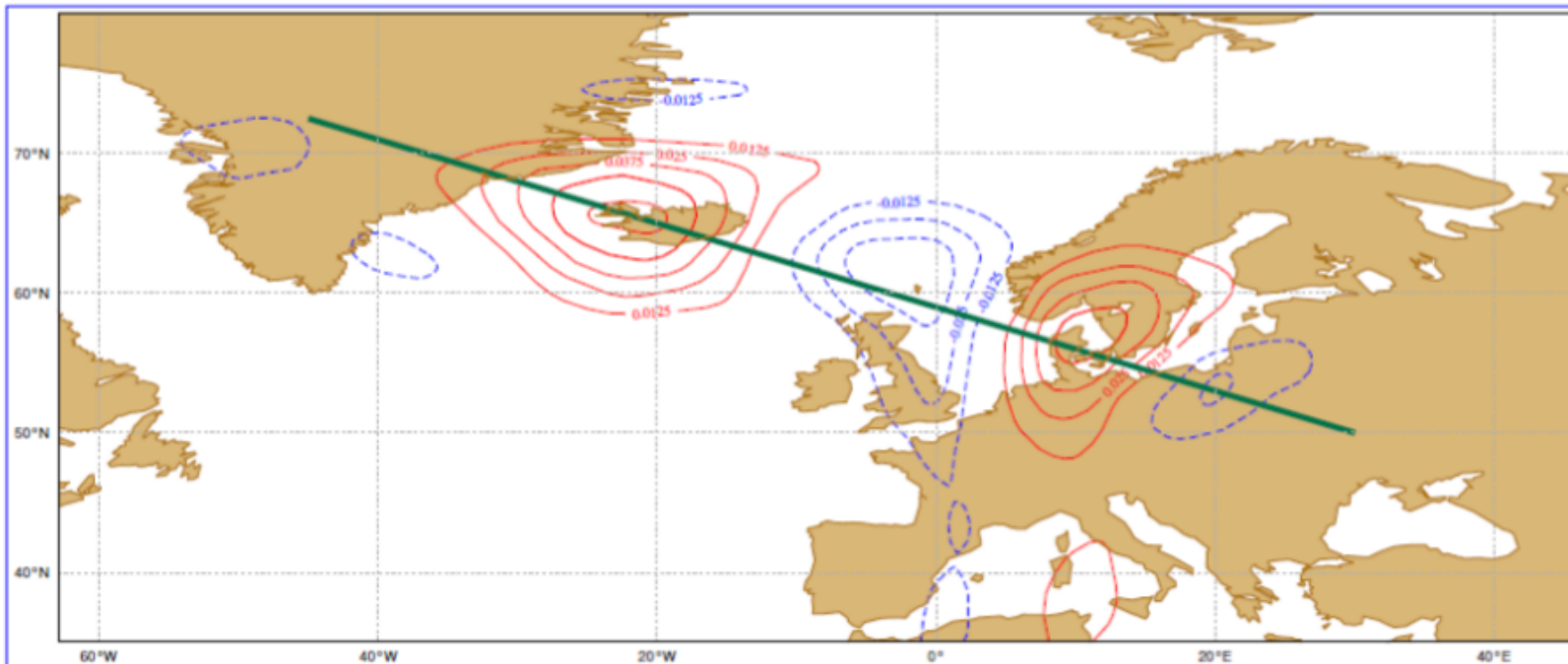


# Szinguláris vektorok

- A leggyorsabban növekvő (optimális) perturbációk megtalálásához tehát egy **sajátérték feladatot** kell megoldani
- Ez nem triviális, mivel sokdimenziós problémáról van szó. A feladat megoldásához többnyire a Lánczos-algoritmust használják
- A szinguláris vektorok megtalálása nagyon költséges
- A szinguláris vektorok függnak:
  - A norma megválasztásától
  - Az optimalizálási időtől (pl.: 12, 24, 48 óra)
  - Az optimalizálási tartománytól (lehet globális SV, vagy lehet céltartományt definiálni)

# Szinguláris vektorok

- A szinguláris vektorok általában a baroklin instabilitás területeit mutatják (időjárási frontokhoz köthető)

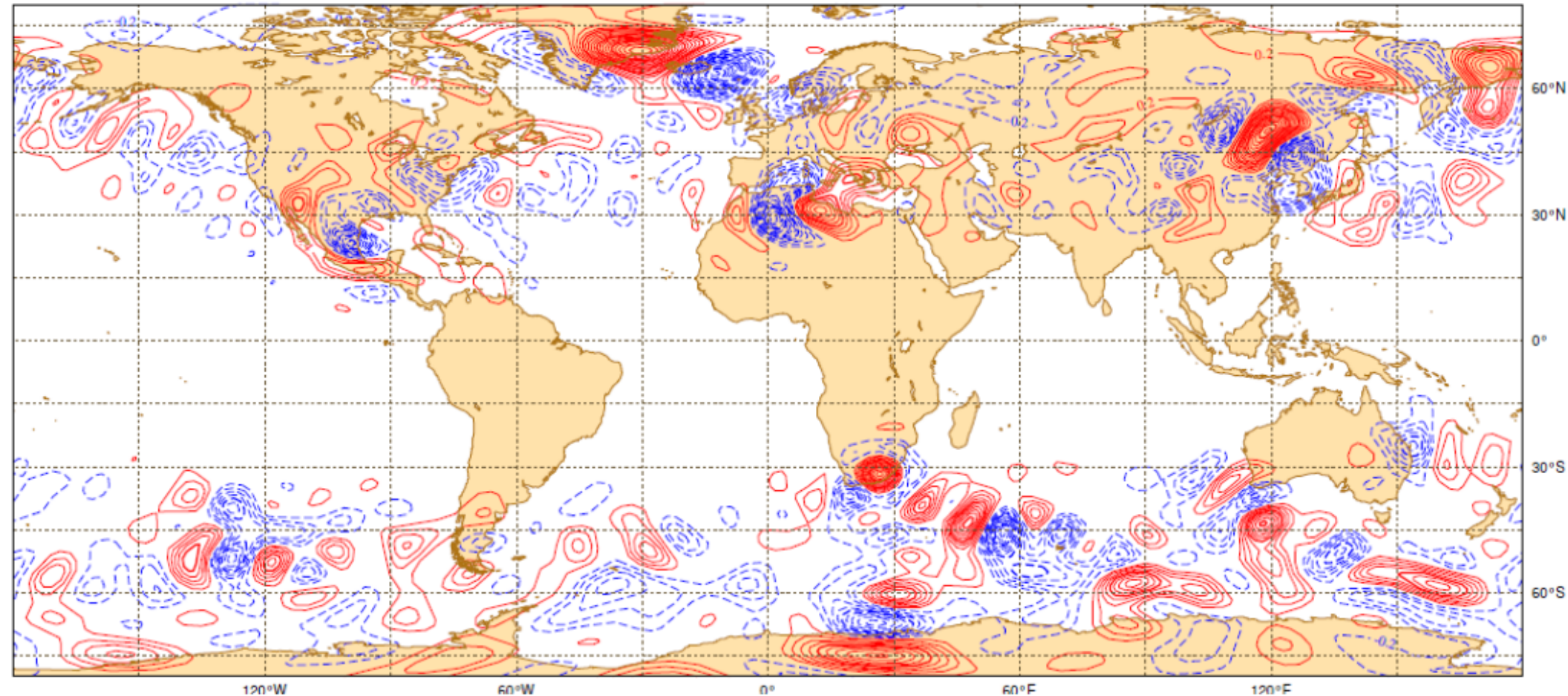


ECMWF ENS, 5. szinguláris vektor,  
2006. 03. 21. 0UTC  
fent: 700hPa hőmérséklet  
Lent: hőmérséklet vertikális metszet a zöld  
vonall mentén

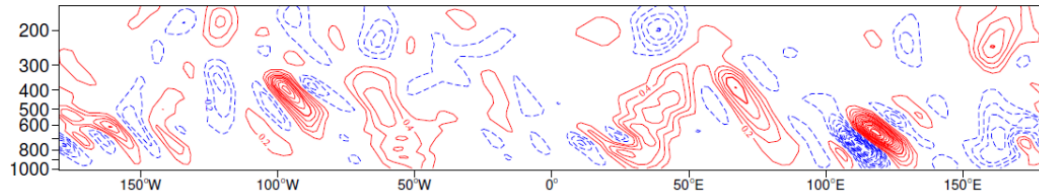
Forrás: Simon Lang előadása, ECMWF  
predictability training

# Szinguláris vektorok

- A gyakorlatban a szinguláris vektorok lineáris kombinációja adja a **kezdeti feltétel perturbációját** (analízis  $\pm$  perturbáció)



at 50°N



ECMWF ENS, SV perturb. Az 1. EPS tagban,  
2006. 03. 21. UTC

fent: 700hPa hőmérséklet

Lent: hőmérséklet vertikális metszet

Forrás: Simon Lang előadása, ECMWF predictability training

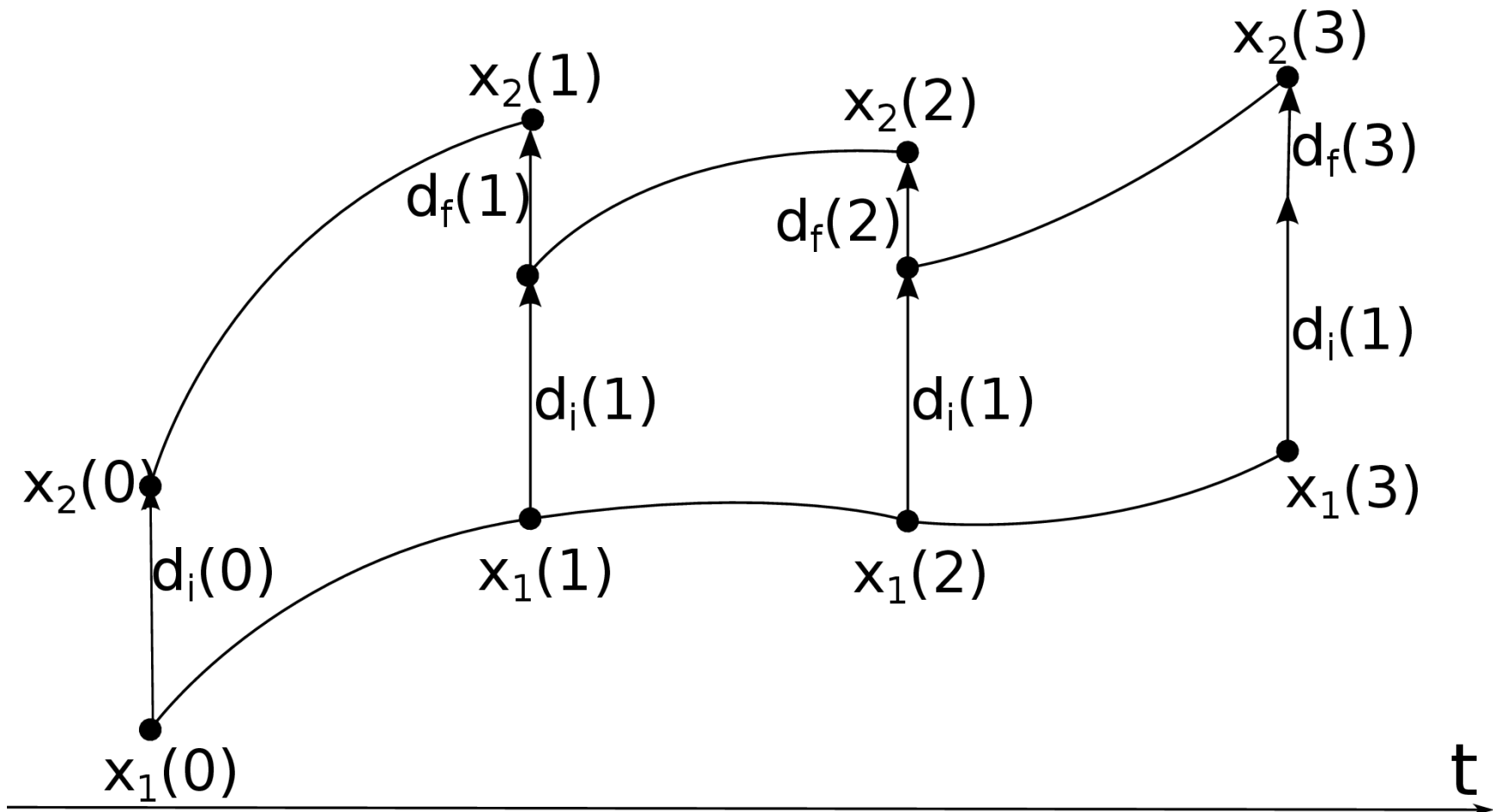


# Breeding módszer

- A szinguláris vektoroknál egyszerűbb, gyakorlatiasabb és kevésbé költséges módszer.
- Szintén az egyik első ensemble módszer (NCEP, 1992)
- Breeding (tenyésztés) módszer:
  - kezdetben véletlenszerű perturbációt adunk az analízishez
  - adott hosszúságú előrejelzés (breeding ciklus)
  - perturbáció visszaskálázása (analízis becsült hibájánál kisebbre)
  - aktuális analízis módosítása ezzel a perturbációval
  - $\Rightarrow$  4-5 nap múlva kiválasztódnak a fázistér leginkább instabilis irányai
- Véletlenszerű perturbációkból kiindulva a rendszer „kitenyészti” az optimális (leggyorsabban növekvő) perturbációkat
- Egymástól független kezdeti perturbációk kvázi-ortogonális perturbációkat eredményeznek (baroklin instabilitás területei)

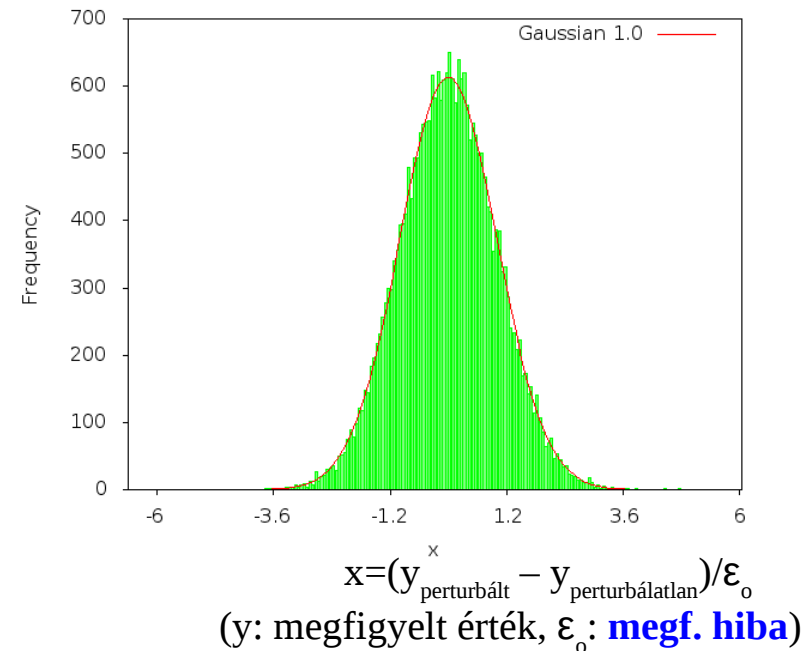
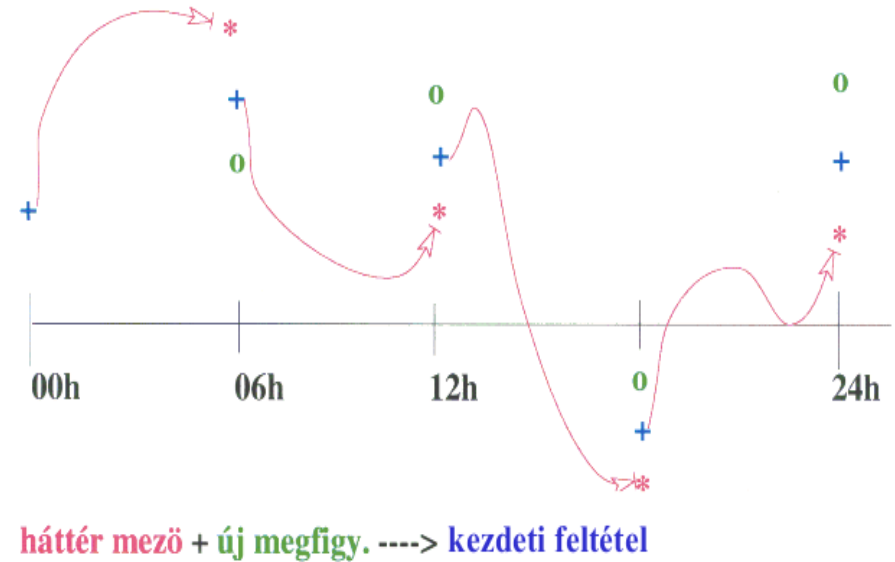
# Breeding módszer

Egyszerűsített, 1 dimenziós kép abban az esetben, ha csak egyetlen perturbációt “tenyésztünk”



# Ensemble adatasszimiláció

- EDA = Ensemble Data Assimilation
- Több adatasszimilációs ciklust futtatunk párhuzamosan
- Az adatasszimiláció során a **megfigyeléseket perturbáljuk** valamilyen Gauss-eloszlásból származó véletlen számmal
- Minden EPS-tagban más-más mérés-halmazt asszimilálunk
- A breeding módszernél a folyamatos visszaskálázás akadályozta meg, hogy a perturbációk túl nagyok legyenek. Itt az analízis „húzza vissza” folyamatosan a léggördt a valósághoz.

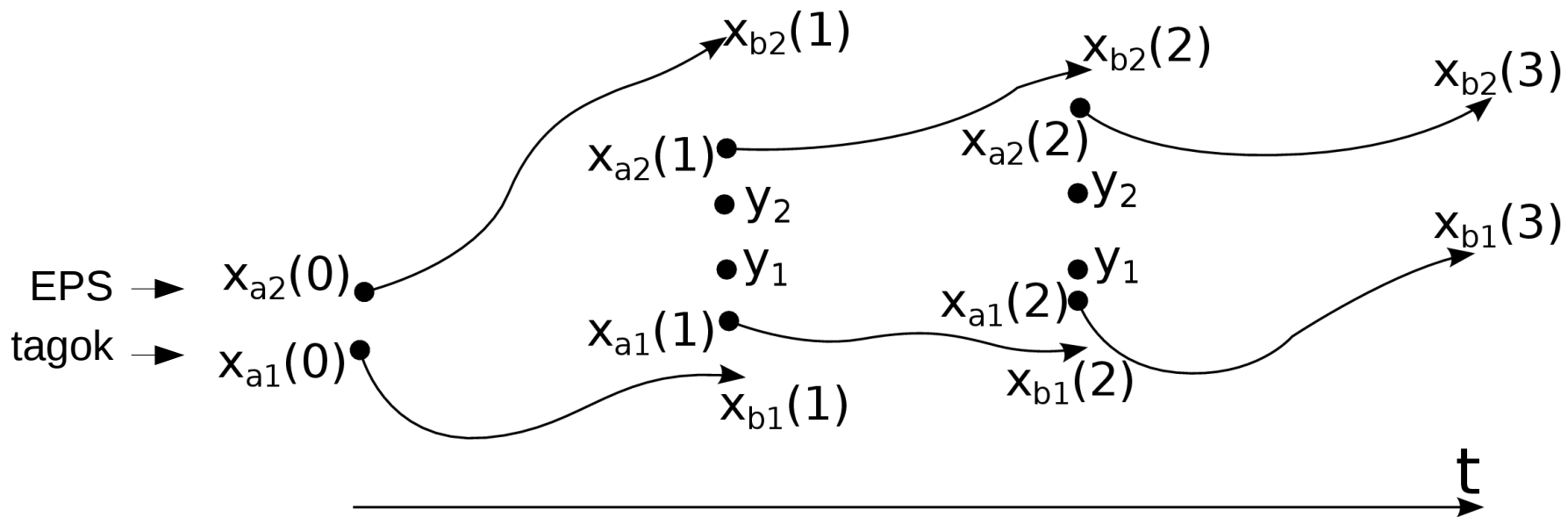


# Ensemble adatasszimiláció

Egyszerűsített, 1 dimenziós kép abban az esetben, ha csak két tagja van az asszimilációs sokaságnak. (pl. OI analízis)

$$\begin{aligned}
 x_{b1} &= Mx_{a1} \\
 x_{b2} &= Mx_{a2}
 \end{aligned}
 \Rightarrow
 \begin{aligned}
 x_{a1} &= x_{b1} + K(y_1 - H(x_{b1})) \\
 x_{a2} &= x_{b2} + K(y_2 - H(x_{b2}))
 \end{aligned}
 \Rightarrow \varepsilon_a \approx x_{a1} - x_{a2}$$

$$K = BH^T (HBH^T + R)^{-1}$$

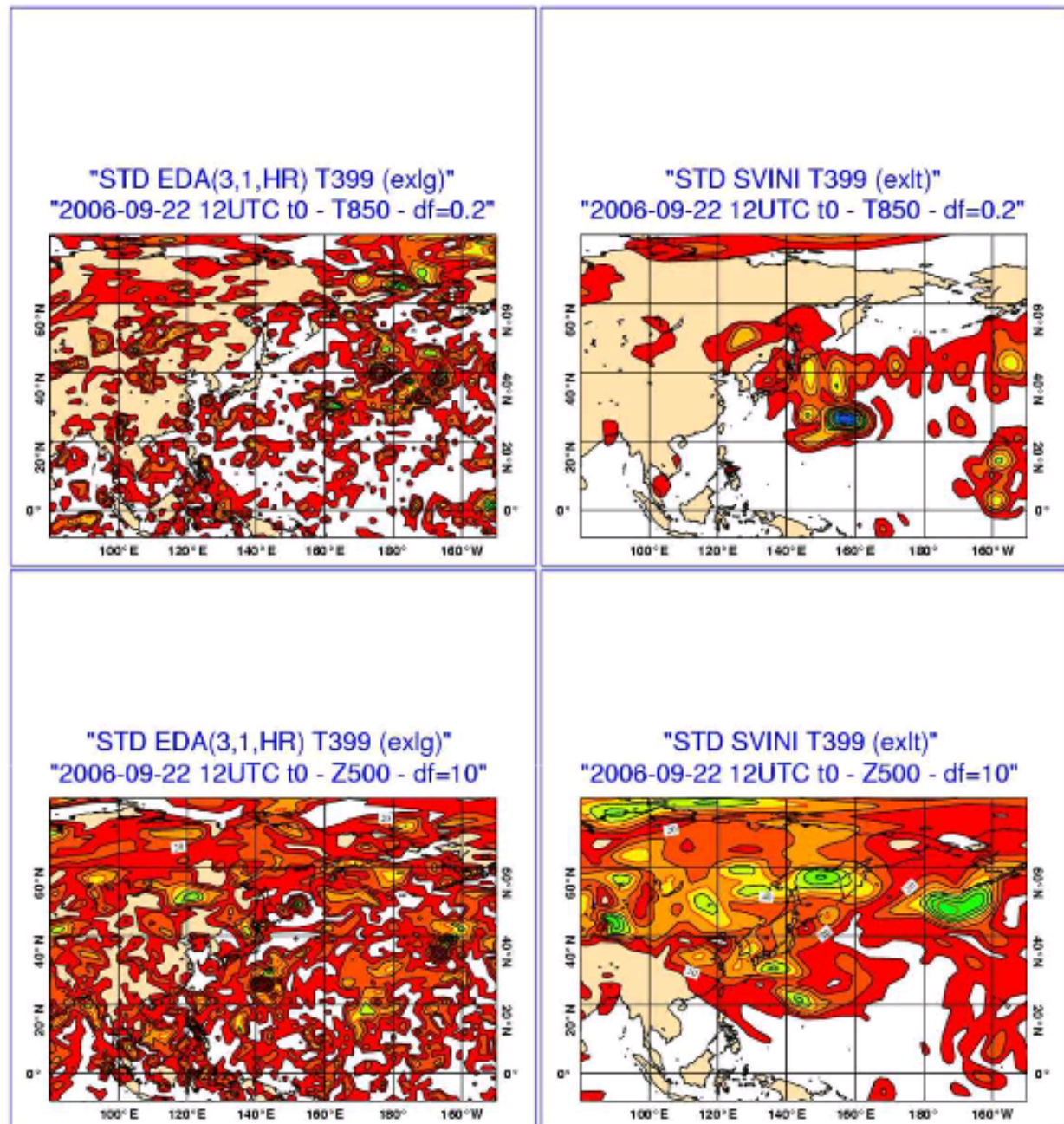


1. ill. 2. EPS tag asszimilációs ciklusa.  $x_a$  az analízis,  $x_b$  a háttér,  $y$  a megfigyelések.  $B$  a háttérhiba kovariancia mátrix,  $R$  a megf. hiba kov. mátrix,  $H$  a megf. operátor. A háttér állapotok biz. idő után eltávolodnak egymástól, így hozzájárulnak az  $\varepsilon_a$  anal. hiba repr.-hoz.

# Ensemble adatasszimiláció

Bal oldalon látható az **EDA**-ból, jobb oldalon a **szinguláris vektorokból** származtatott perturbációk térbeli struktúrája.

Az EDA perturbációk kevésbé fejlődnek optimálisan, ám már az időszak legelején is reprezentálják az analízis bizonytalanságát és kellő szórást biztosítanak a rendszer számára.



# Ismétlés - Kezdeti feltétel perturbációk

A három vizsgált módszer filozófiája kicsit különbözik:

- Szinguláris vektorok
  - Cél, hogy meghatározzuk a tér azon irányait, melyek a legtöbb bizonytalanságot hordozzák az előrejelzés elején.
  - Egy meghatározott időtávon (melyen még érvényesnek véljük a lineáris közelítést), egy meghatározott norma szerint leggyorsabban növekvő perturbációk előállítására.
- Breeding módszer
  - Cél, hogy egy ciklus során véletlen perturbációkból olyanokat tenyésszen ki a rendszer, mely az előző időszak bizonytalanságait reprezentálja.
- Ensemble adatasszimiláció
  - Mérési hiba-határon belüli perturbációkat ültetünk a mérésekre.
  - Cél, hogy párhuzamosan futtatott adatasszimilációs ciklusokon keresztül határozzuk meg a kezdeti feltétel bizonytalanságait.

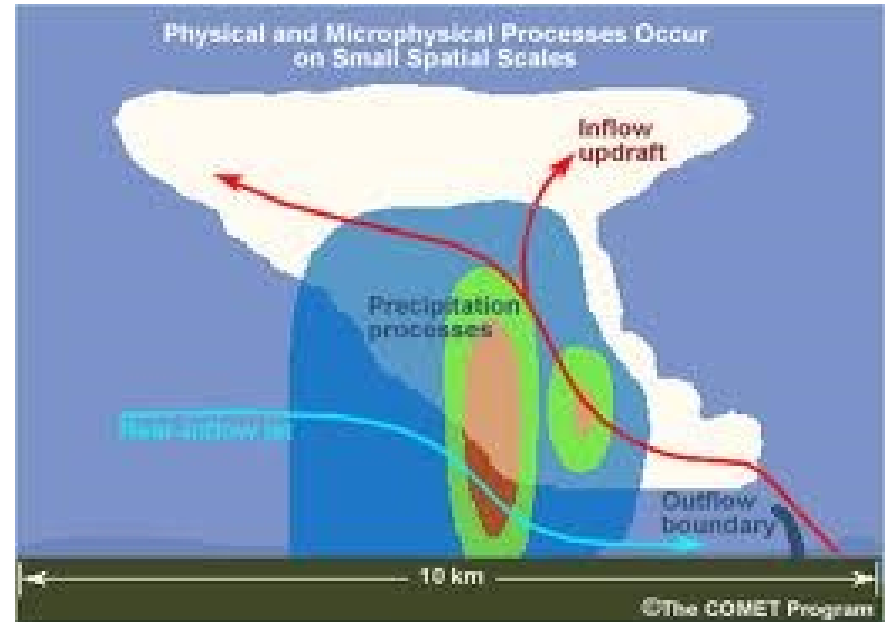
# Modellhiba reprezentációja

Tapasztalat: Nem elég a kezdeti feltétel hibáját figyelembe venni, szükség van a modellintegrálás során fellépő bizonytalanságok figyelembe vételére is!

- A modellszámításokat szokás egyszerűen két részre bontani:
  - **“Dinamika”** - a HTDER diszkretizálása és megoldása
  - **“Fizika”** - a rácstávolságnál kisebb skálájú folyamatok átlagos hatásának figyelembevétele
- Mindkét rész hordoz bizonytalanságokat:
  - **“Dinamika”** - Egy adott modellben nem szoktak többféle dinamikát használni, és nem szokták a dinamikát perturbálni
  - **“Fizika”** - Inkább a fizikai parametrizációkon keresztül próbálják figyelembe venni a modellhibát

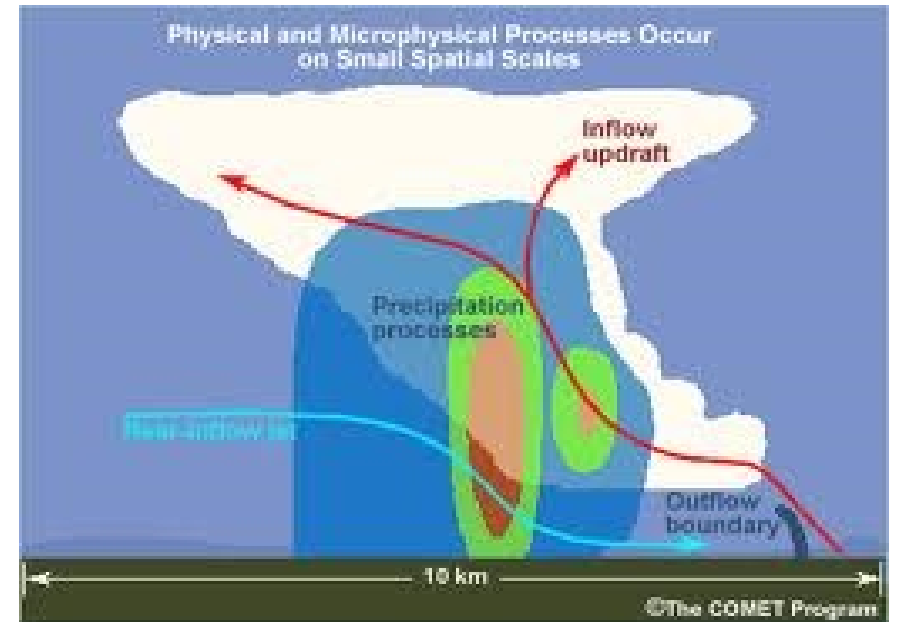
# Multi-fizika ensemble

- A rácstávolságnál kisebb skálájú, illetve túlságosan bonyolult folyamatok leírásárását **fizikai parametrizációval** végezzük
- Nem mindig egyértelmű, hogy melyik fizikai parametrizáció a legjobb (függhet időjárási helyzettől, felbontástól)
- A különböző ensemble tagok esetén különböző parametrizációkat lehet használni a bizonytalanság számszerűsítése érdekében
- Ez egy meglehetősen gyakorlatias módszer





# Paraméter perturbációk



- Bizonyos, a modellben használt paraméterek értékeit perturbálják egy-egy véletlen számmal (ésszerű határértékeken belül)
- Minden EPS tagban különböző a paraméter-készlet, de egy-egy EPS tagban egy adott modellintegrálás során ezek az értékek állandóak

# SPPT – eredeti séma

- **SPPT – Stochastically Perturbed Parameterized Tendencies**
- Eredetileg csak sztochasztikus fizikának nevezték (1999 - ECMWF)
- Az ensemble rendszer koncepciója, ha csak kezdeti feltétel perturbáció van:
  - A  $j$ -edik tag állapota ( $e_j$ )  $t_1$  időpontban a következő modellegyenlet integrálásával adható meg:

$$\frac{\partial e_j(t)}{\partial t} = A(e_j, t) + P(e_j, t)$$

$$e_j(t_1) = \int_0^{t_1} (A(e_j, t) + P(e_j, t)) dt$$

$$e_j(t=0) = e_0(t=0) + \delta e_j(t=0)$$

$A$  – a nem-parametrizált folyamatok (dinamika)

$P$  – a parametrizált folyamatok (fizika)

$\delta e$  – kezdeti feltétel perturbációk (a korábban bemutatott módszerekből származhat)

- Ötlet: 1. az egyes parametrizált folyamatok nem függetlenek egymástól  
2. a parametrizációk számítása a modellben elkülönül  
⇒ Változtassuk meg a teljes  $P$  tendenciát!

# SPPT – eredeti séma

- A rendszer koncepciója perturbált modell-egyenletek esetén:

$$e_j(t_1) = \int_0^{t_1} (A(e_j, t) + P'_j(e_j, t)) dt$$

- $P'$  – a parametrizált folyamatok perturbált tendenciája

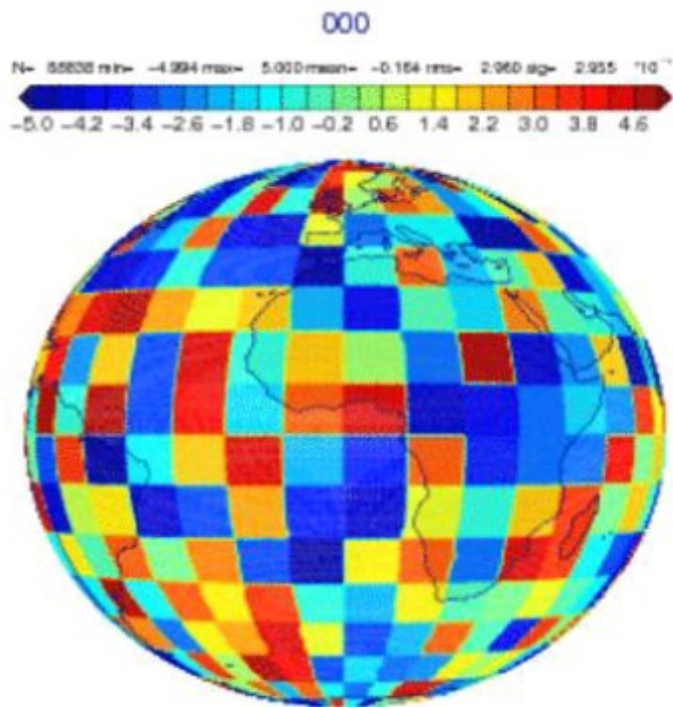
$$P'_j(e_j, t) = (1 + \langle r_j(\lambda, \varphi, t) \rangle_{D, T}) P(e_j, t)$$

- $r_j$  – Egyenletes eloszlásból vett véletlen szám, ami állandó a  $D \times D$  rácspont méretű területen  $T$  ideig. ( $\lambda, \varphi$ : koordináták)

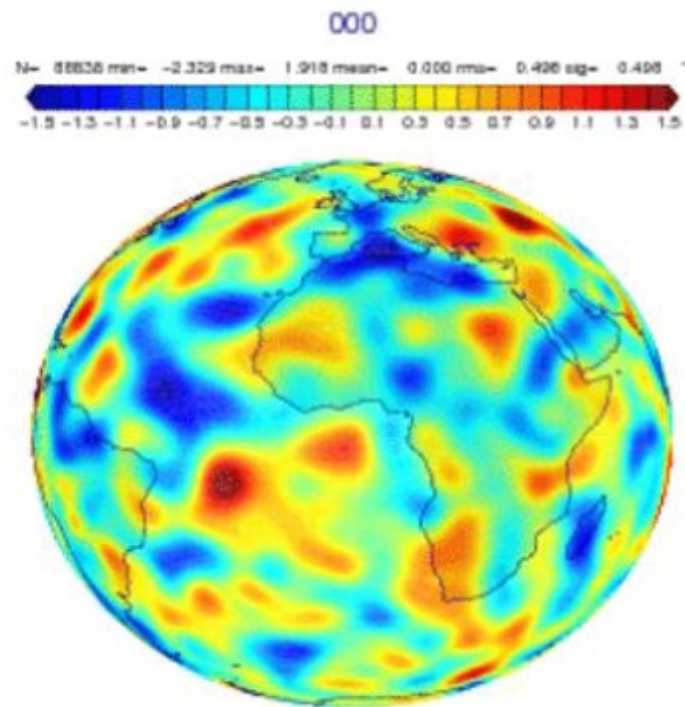
- $r$  nagyságának és tér- és időbeli korrelációjára több lehetőség van, pl.:

$r_j \in [-1; 1]$	$D = 10^\circ$	$T = 12\text{h}$
$r_j \in [-0.5; 0.5]$	$D = 5^\circ$	$T = 3\text{h}$
$r_j \in [-0.25; 0.25]$	$D = 1^\circ$	$T = 45\text{min}$

# SPPT – motiváció a séma felülvizsgálatára



Az eredeti séma perturbációi egy adott időszakon



Olyan simább mező, ami jobban összhangban van a modellhiba tér- és időbeli szerkezetével

# SPPT – felülvizsgált séma

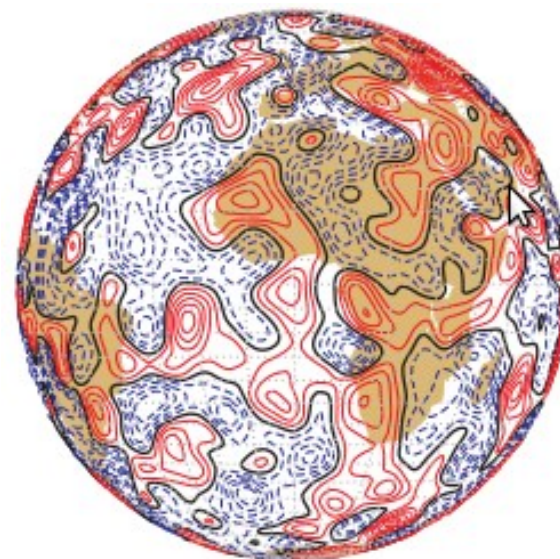
- Az új sémában  $r$  sokkal kifinomultabb módon kerül meghatározásra:

$$P'_j(e_j, t) = (1 + \alpha r_j) P(e_j, t)$$

- $r$  olyan Gauss-eloszlásból vett érték, amelynek szórását a kívánt térbeli korrelációhoz igazítják
- $\alpha$ -csak a vertikális koordinátától függő szorzó (általában az alsó és felső perem közelében csillapítják a perturbációkat)
- $r$ -t egy spektrális minta-generátorral hozzák létre:

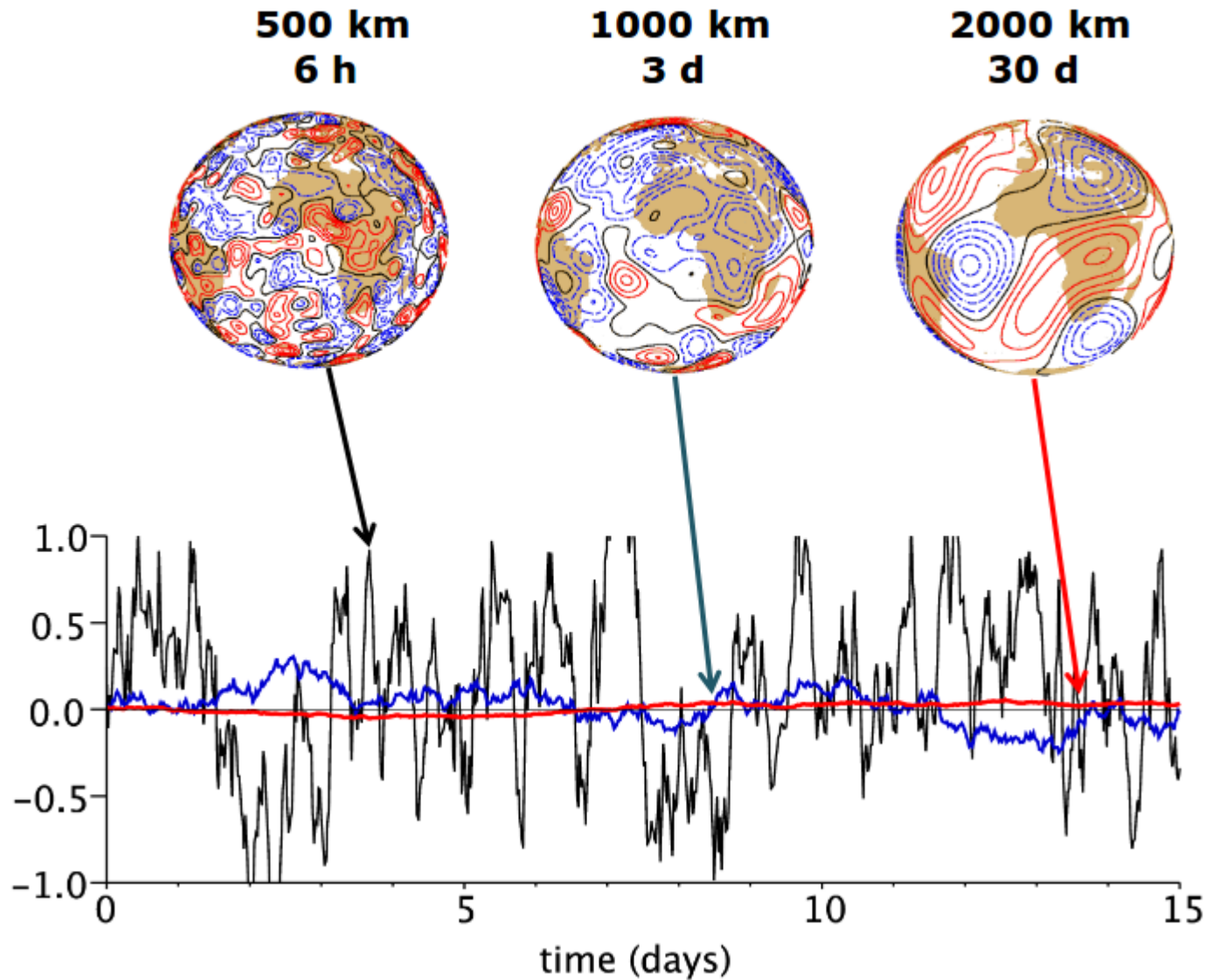
$$r = \sum_{m,n} \hat{r}_{mn} Y_{mn}$$

- $Y_{mn}$  gömbharmonikusok
- $\hat{r}_{mn}$  spektrális együtthatókat auto-regresszív folyamattal határozzák meg, amelynek lényege, hogy az egymást követő  $r$ -ek időben korrelálnak egymással



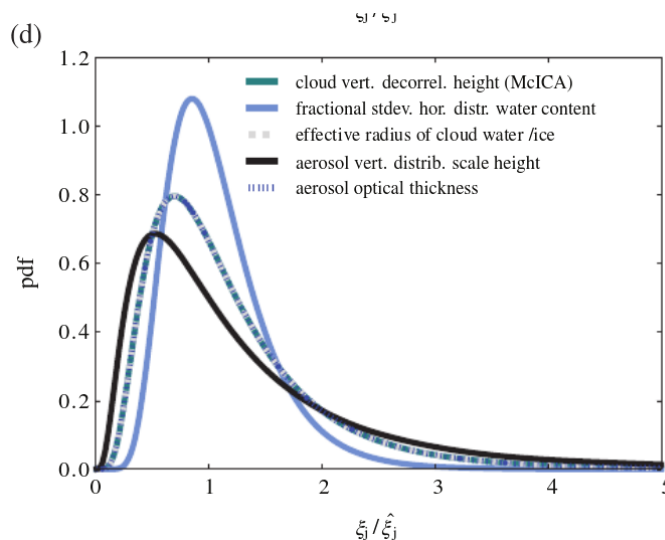
# SPPT – felülvizsgált séma

$r$  előállítható többféle mintázat kombinációjaként is (különböző skálájú folyamatok leírásának hibáját próbálják így reprezentálni), pl. (ECMWF):



# SPP – Stochastically perturbed parameters

- Tekinhető a PP és SPPT „ötvözésének”
- Paramétereket *egymástól függetlenül* perturbálják *lokálisan*, (SPPT-hez hasonló tér- és időbeli korrelációs minta)
- Meghatározzák azon legbizonytalanabb paramétereket, amelyek leginkább hozzájárulnak az előrejelzés hibájához
- Ezeket többnyire log-normál eloszlás szerint perturbálják
- Legújabb séma (2017, ECMWF)




---

## Turbulent diffusion & subgrid orography

Transfer coefficient for momentum (ocean/land)

Coefficient in turbulent orographic form drag scheme

Standard deviation of subgrid orography

Length-scale for vertical mixing in stable boundary layer

## Convection

Entrainment rate

Shallow entrainment rate

Detrainment rate for penetrative convection

Conversion coefficient cloud to rain

Zonal convective momentum transport

Meridional convective momentum transport

Adjustment time-scale in CAPE closure

## Cloud & large-scale precipitation

Relative humidity threshold for onset of stratiform condensation

Diffusion coefficient for evaporation by turbulent mixing

Critical cloud water content for autoconversion

Threshold for snow autoconversion

## Radiation

Cloud vertical decorrelation height in McICA

Fractional stdev. of horizontal distribution of water content

Effective radius of cloud water and ice

Scale height of aerosol normalized vertical distribution

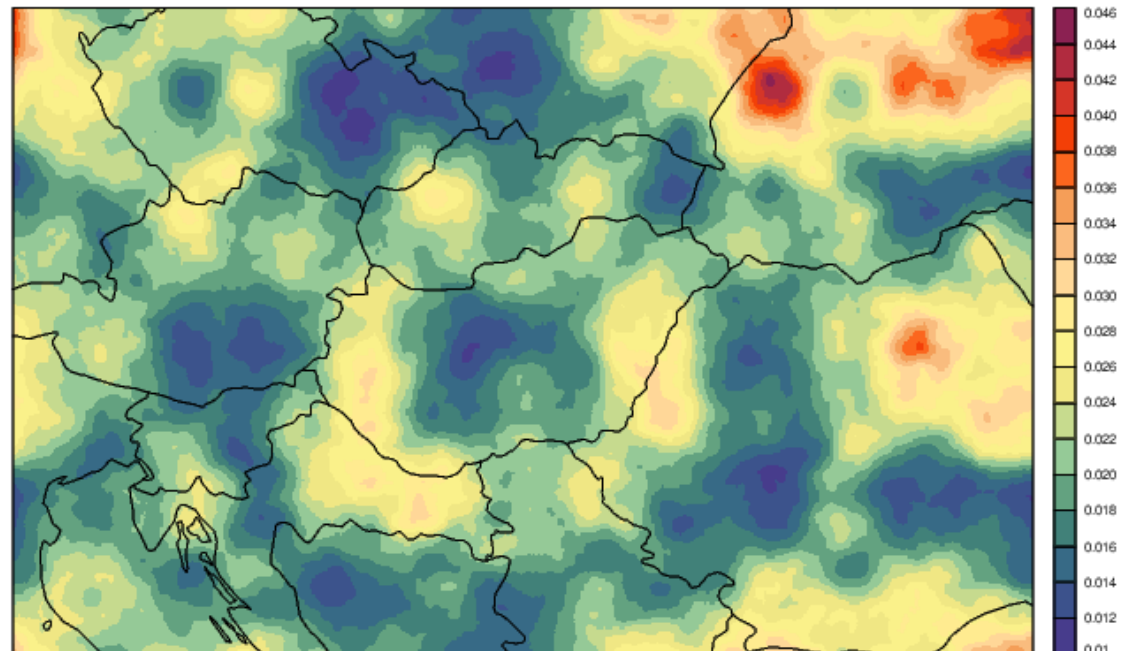
Optical thickness of aerosol

---

# SPP – Sztochasztikusan perturbált paraméterek

- Fizikai folyamatokhoz kapcsolódik: paraméterek kiválasztása, perturbációk, tér-/időbeli korreláció meghatározása
- Folyamat-szintű perturbációt tesz lehetővé (pl. sugárzás, mikrofizika)
- Finomhangolása jóval több erőforrást igényel, mint az SPPT

- LAM verzióra példa:
  - (az egyik paraméter perturbált értékei)





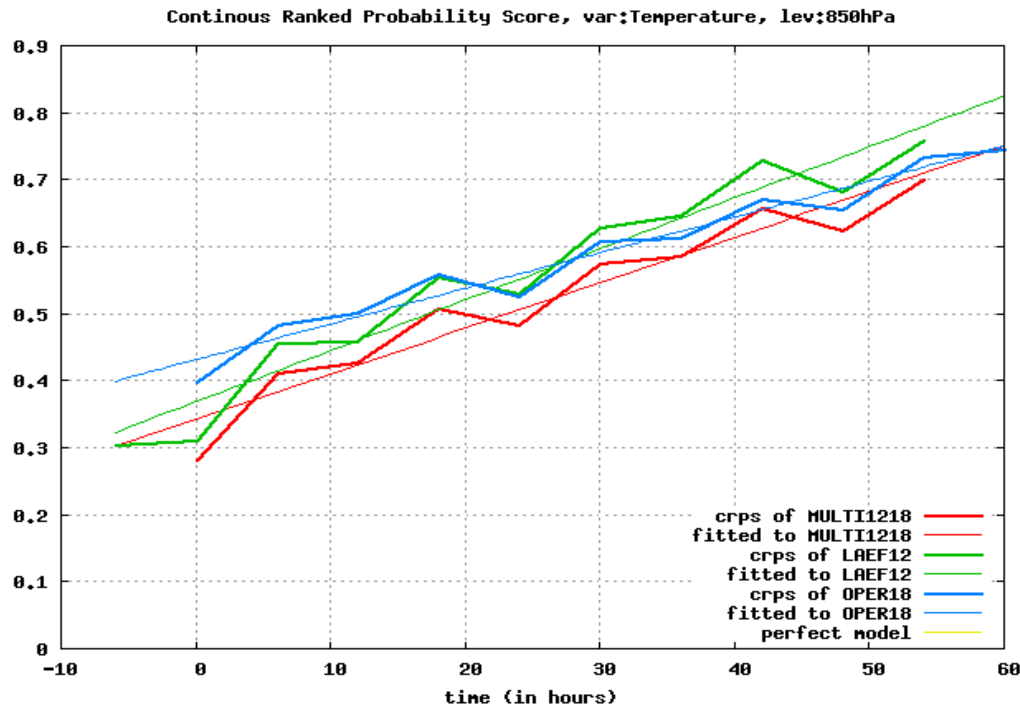
# Multi-modell ensemble

- **Több perturbálatlan modellt** rakunk össze egy ensemble rendszerbe. Ez különböző intézetek/előrejelző központok modelljeit jelenti
  - Lehetnek maguk a modellek is különbözőek, de általában minden intézet kicsit más beállításokkal futtatja a modelljeit, esetleg más méréseket vesz figyelembe az adatasszimiláció során
  - Ez egy egyszerű módszer, aminek segítségével olyan intézetek is juthatnak ensemble rendszerhez, melyeknek nem lenne erre kellő számítógépes kapacitása
- ⇒ **Poor man's ensemble prediction system**



# Multi-modell ensemble

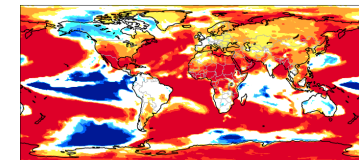
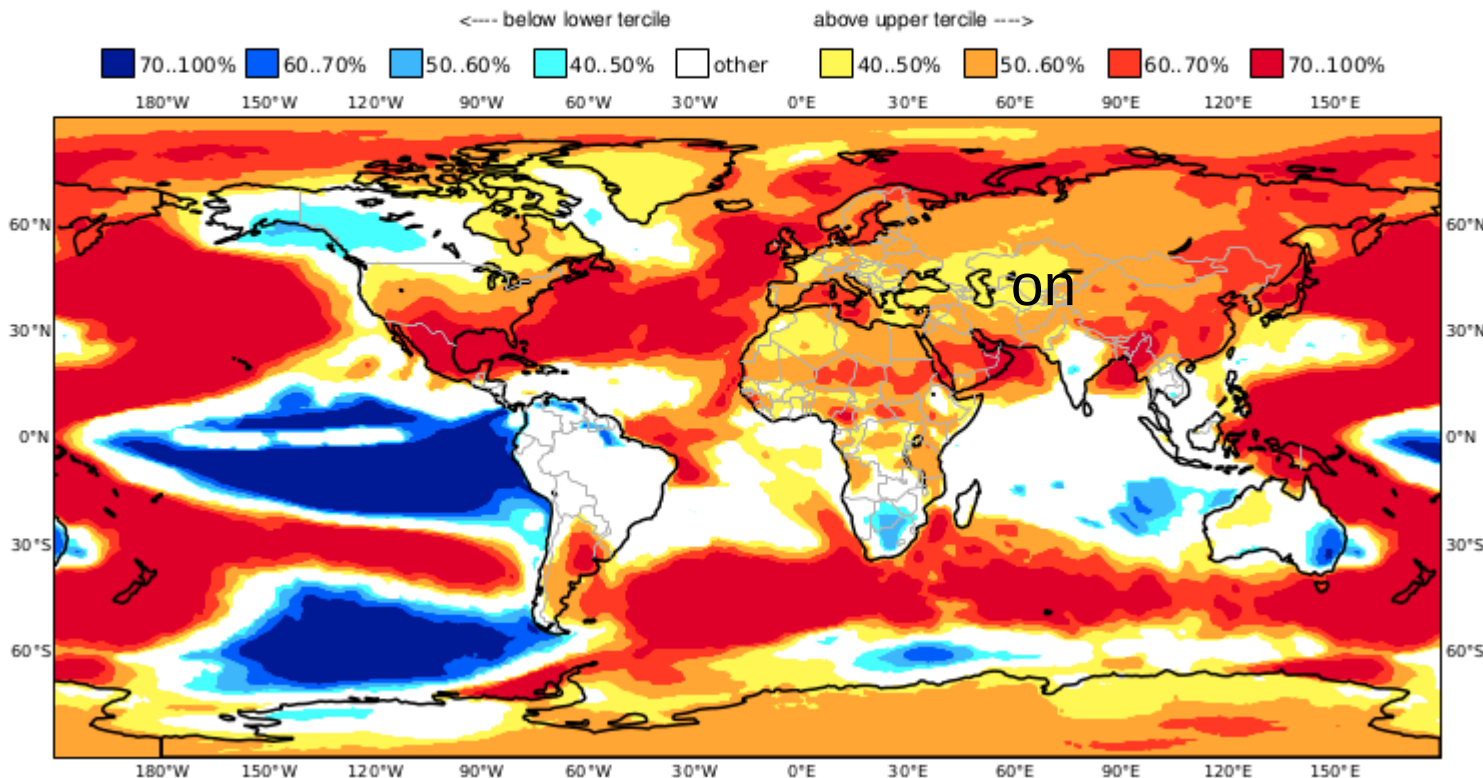
- **Rövid távú előrejelzések:** Inkább csak próbálkozások vannak nemzetközi összefogásra és multi-modell ensemble létrehozására
- Túl nagy adatátvitelre lenne szükség az operatív üzemeltetéshez
- Probléma, hogy mindenki más időpontból indíthatja, más időtávra és más területre futtathatja az előrejelzését
- Verifikációban általában megmutatható, hogy jobb a multi-modell ensemble (nemzetközi összefogások célozzák közös adatbázisok létrehozását)



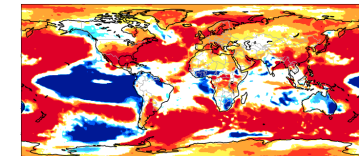
# Multi-modell ensemble

- **Szezonális- ill. klímamodellezésben** azonban használják

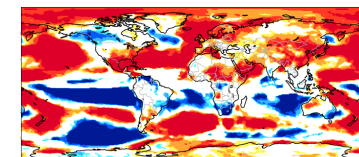
C3S multi-system seasonal forecast    ECMWF/Met Office/Météo-France/CMCC/DWD/NCEP/JMA/ECCC  
Prob(most likely category of 2m temperature)    DJF 2022/23  
Nominal forecast start: 01/11/22  
Unweighted mean



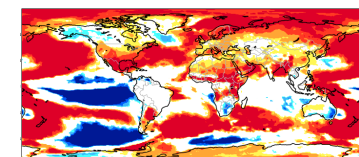
CMCC



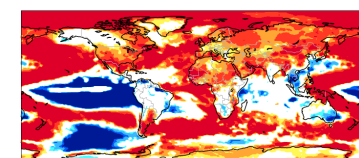
DWD



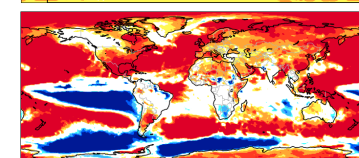
ECCC



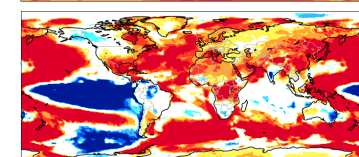
ECMWF



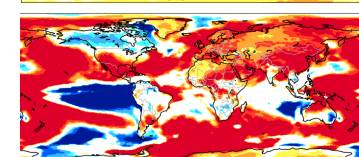
JMA



Meteo  
France



Met  
Office



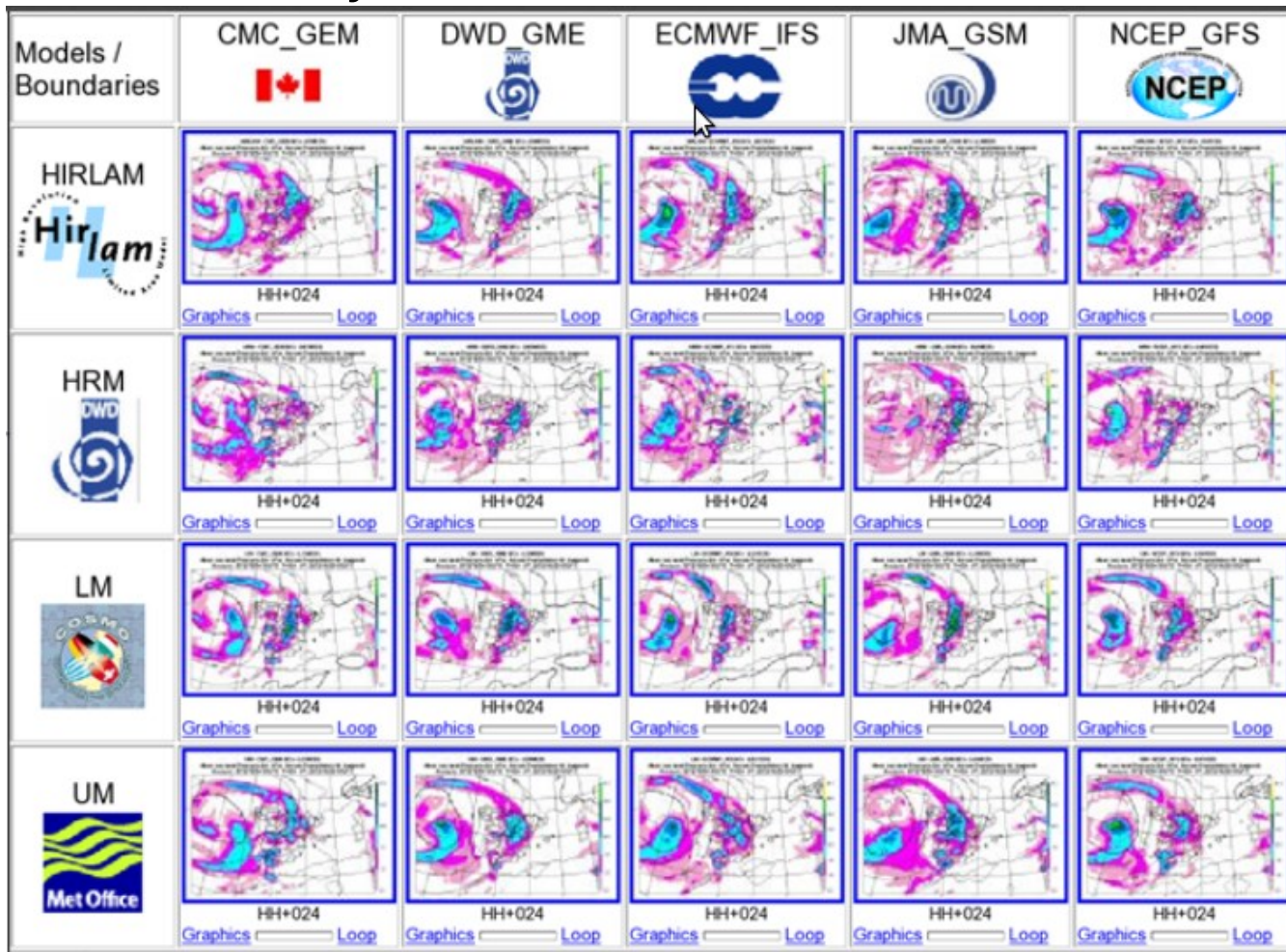
NCEP

Ábra: 2 m hőm. anomália valószínűségek, 2022/23 tél  
[https://climate.copernicus.eu/charts/c3s\\_seasonal](https://climate.copernicus.eu/charts/c3s_seasonal)



# Multi-analízis ensemble

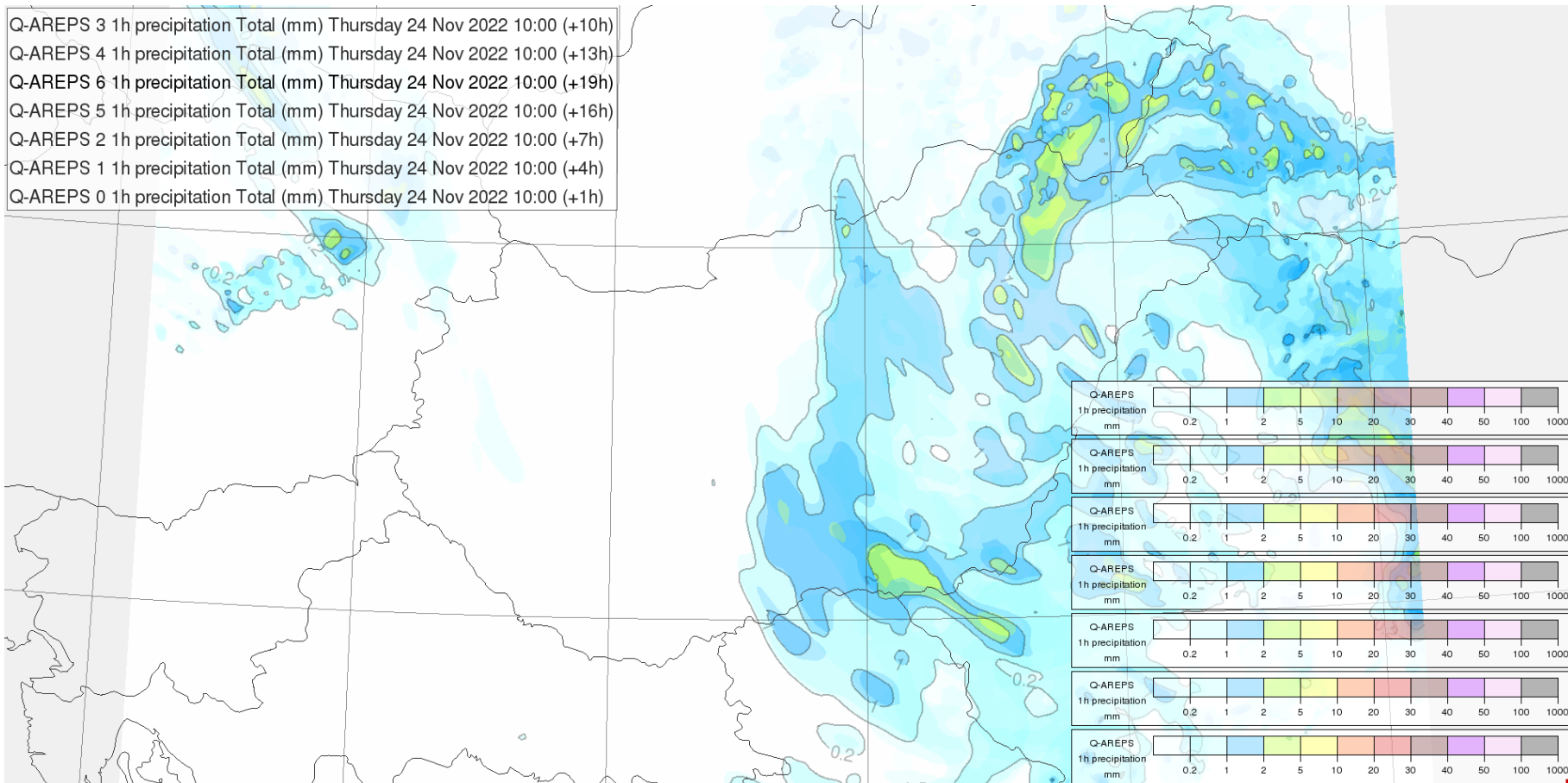
- Nem különböző modellek előrejelzéseit, hanem különböző technikákkal és/vagy modellekkel készített analíziseket használnak
- Gyakran ötvözik az előző módszerrel
  - ⇒ multi-modell/multi-analízis ensemble
- Korlátos tartományú modelleknél lehet multi-határfeltételekről beszélni



spanyol példa:  
multi-modell/  
multi-LBC

# Multi-analízis ensemble

- Nem különböző modellek előrejelzéseit, hanem különböző technikákkal és/vagy modellekkel készített analíziseket használnak.
- Ugyanazon modellnek más kezdő időpontban induló futásait együtt is nézhetjük. „Time-lagged EPS”



Kül. Kezdő időpontból induló AROME futások EPS-ként értelmezve  
(Hawk-ban elérhető)

# Ensemble előrejelzések értékelése

- A **perturbálatlan** előrejelzés esőt mondott
  - Az eső esett
  - Ez egy jó előrejelzés volt?
    - Igen
    - Nem
    - Nem tudom



# Ensemble előrejelzések értékelése

- A **perturbálatlan** előrejelzés esőt mondott
  - Az eső esett
  - Ez egy jó előrejelzés volt?
    - Igen
    - Nem
    - Nem tudom
- Az **ensemble előrejelzés** 20% esélyt adott az esőre
  - Az eső esett
  - Ez egy jó előrejelzés volt?
    - Igen
    - Nem
    - Nem tudom



# Ensemble előrejelzések értékelése

- A **perturbálatlan** előrejelzés esőt mondott
  - Az eső esett
  - Ez egy jó előrejelzés volt?
    - Igen
    - Nem
    - Nem tudom
- Az **ensemble előrejelzés** 20% esélyt adott az esőre
  - Az eső esett
  - Ez egy jó előrejelzés volt?
    - Igen
    - Nem
    - Nem tudom



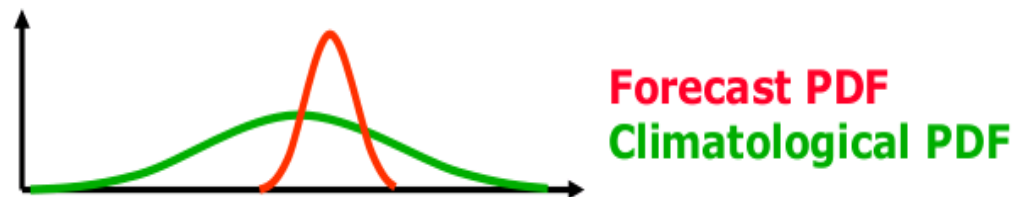
***Statisztikát kell készítenünk! (Több időpontra, állomásra)***



# Ensemble előrejelzések értékelése

Egy *valószínűségi* előrejelzés karakterisztikái:

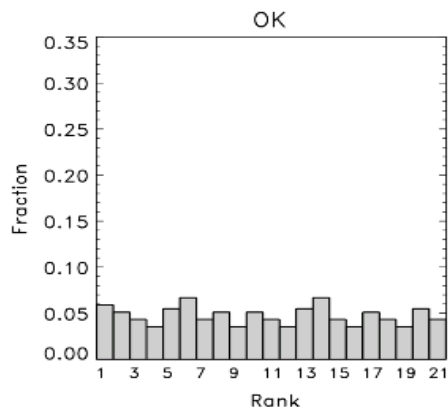
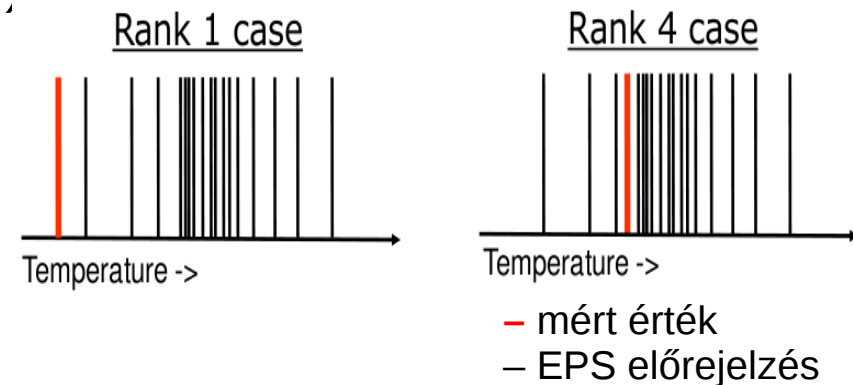
- **Megbízhatóság (reliability, consistency):**
  - Az ensemble előrejelzések statisztikailag ugyanahhoz az eloszláshoz tartoznak-e, mint a megfigyelések?
  - Megbízhatok a valószínűségekben, amiket az előrejelzésem mond? (pl.: Tényleg 20% az esély?)
  - **Pontosság (accuracy):** Várható értéket jól megfogta a modell?
- **Élesség (sharpness):** Mennyiben különbözik az előrejelzés a klímaátlag valószínűségétől (pl.: 33% eséllyel fog esni holnap)
- **Skill:** Jobb az előrejelzésem, mint a referencia? (a referencia lehet véletlen, klíma, perzisztencia)



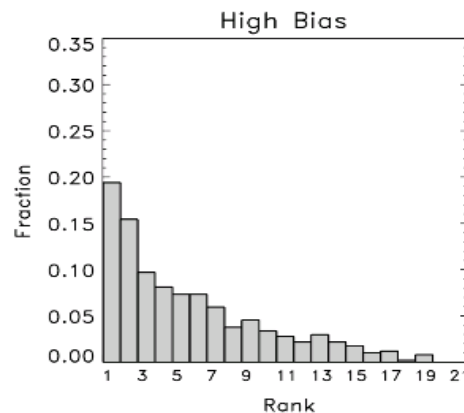
# Ensemble előrejelzések értékelése

## Talagrand-diagram

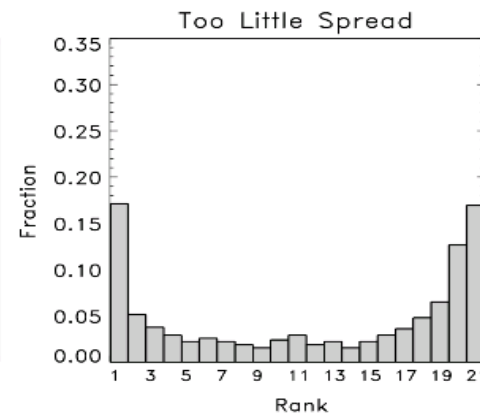
- Az  $N$  db előrejelzés  $N+1$  db részre („dobozra”) osztja az előrejelzési tartományt
- Minden megfigyelés valamelyik dobozba esik  $\rightarrow$  „rank”
- A kellően nagyszámú előrejelzés-megfigyelés pár esetén megjelenített rank-elemszám diagramot nevezzük Talagrand-diagramnak



OBS is indistinguishable from any other ensemble member



OBS is too often below the ensemble members (biased forecast)



OBS is too often outside the ensemble spread

**L** alak: EPS sz. felülbecsül

**J** alak: EPS alulbecsül

# EPS tulajdonságok

- EPS átlag:

$$\bar{x} = \frac{1}{N} \sum_{j=1}^N x_j$$

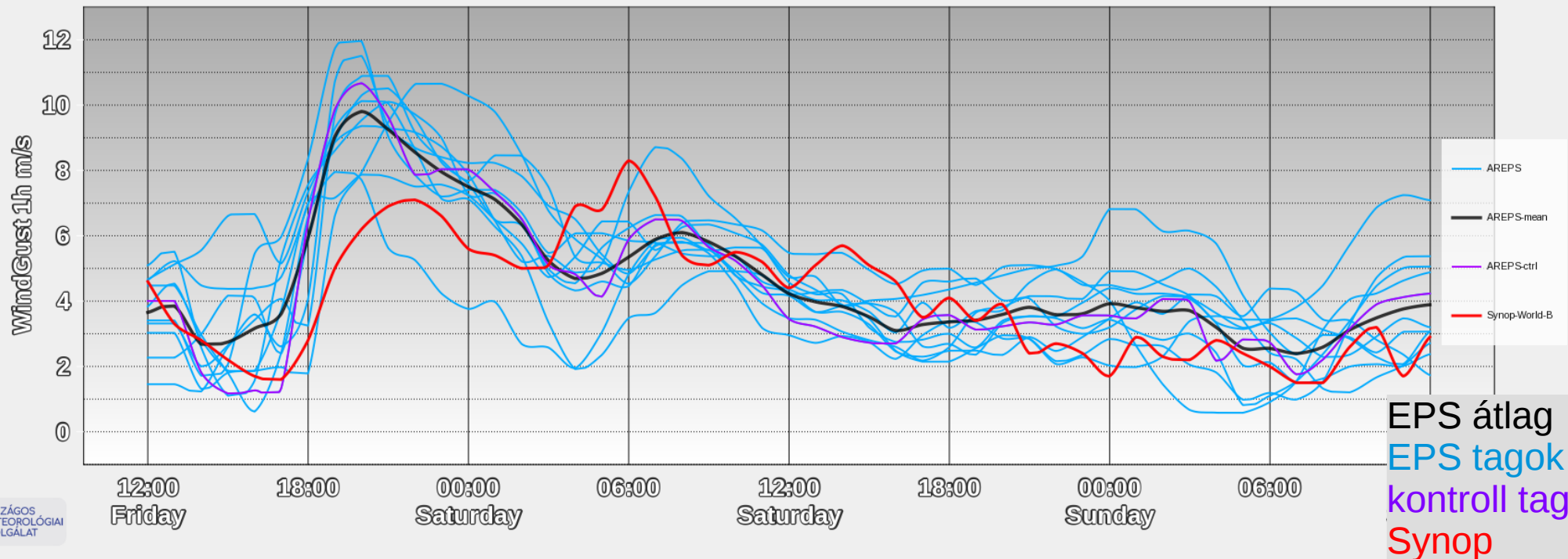
- EPS szórás:

$$s^2 = \frac{1}{N} \sum_{j=1}^N (x_j - \bar{x})^2$$

$x_j$  a  $j$ . EPS tag által előrejelzett érték,  $N$  az EPS tagok száma

AROME-EPS SZOMBATHELY

2022. November 18. - November 20.



# EPS tulajdonságok + „determinisztikus” mérőszámok

- EPS átlag:

$$\bar{x} = \frac{1}{N} \sum_{j=1}^N x_j$$

- EPS szórás:

$$s^2 = \frac{1}{N} \sum_{j=1}^N (x_j - \bar{x})^2$$

$x_j$  a  $j$ . EPS tag által előrejelzett érték,  $N$  az EPS tagok száma

egy-egy esetre  
statisztika

- bias (torzítás, szisztematikus hiba)

$$bias = \frac{1}{n} \sum_{i=1}^n ((\bar{x})_i - y_i)$$

- RMSE (root mean square error)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n ((\bar{x})_i - y_i)^2}$$

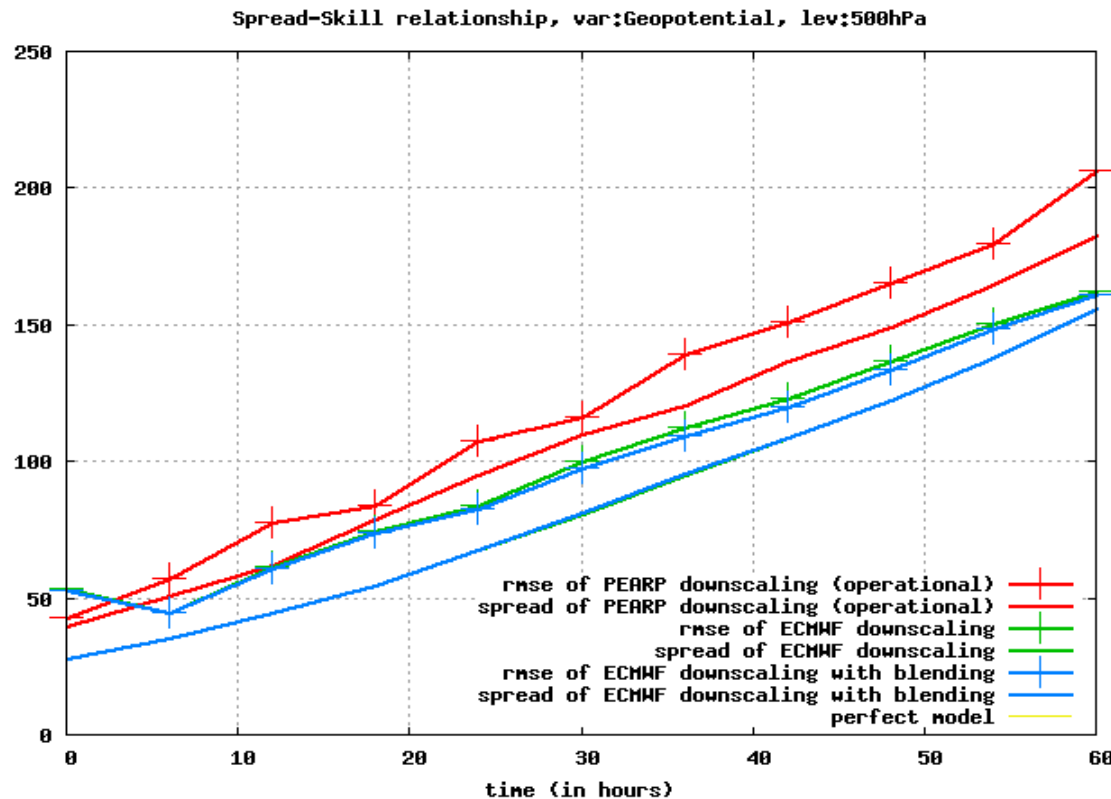
egyetlen előrejelzésre  
alkalmazható mérőszámok

- EPS átlagra, vagy külön a tagokra is számolhatók

$i$  az adott eset,  $n$  a vizsgált esetek száma,  $y_i$  a mért érték (esetleg analízis) több napi, ill. több állomásra vonatkozó előrejelzést szoktunk verifikálni

# Ensemble előrejelzések értékelése

- Spread-RMSE kapcsolat
  - Összehasonlítjuk egy-egy előrejelzett paraméterre az ensemble átlag átlagos hibáját (RMSE) a tagok átlagos szórásával
  - Azt szeretnénk, ha ez a kettő közel lenne egymáshoz (ensemble rendszer jól leírja a modell bizonytalanságát)

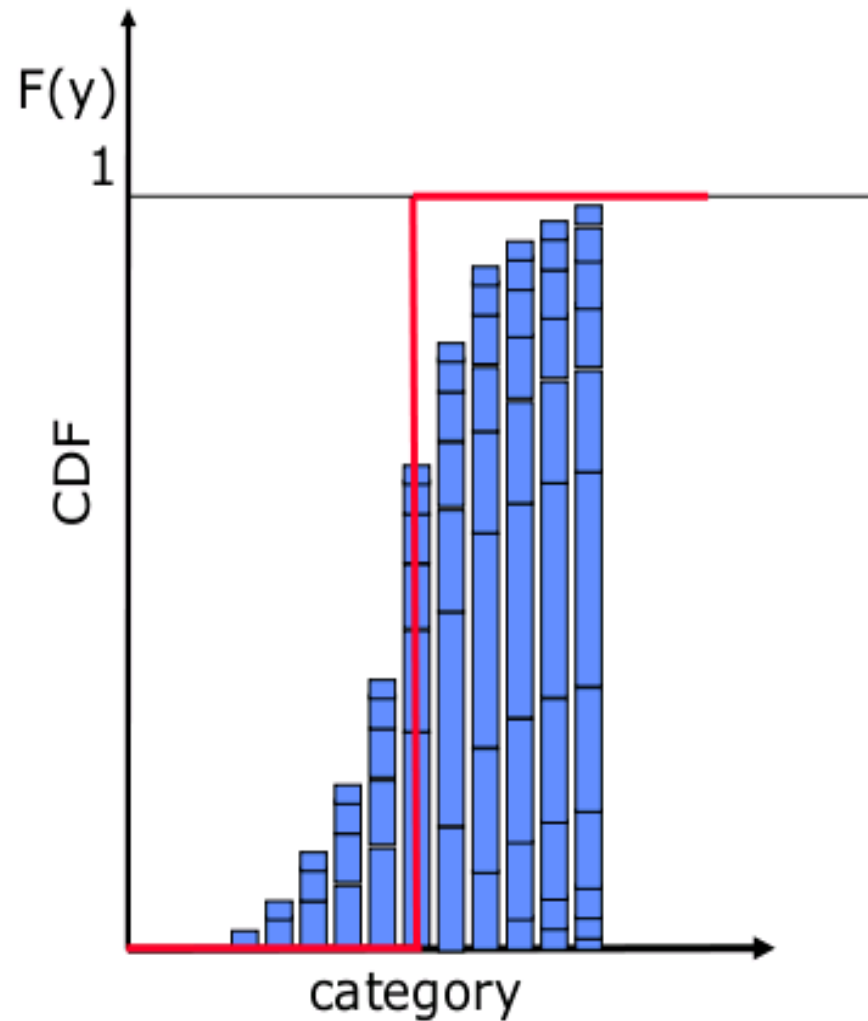
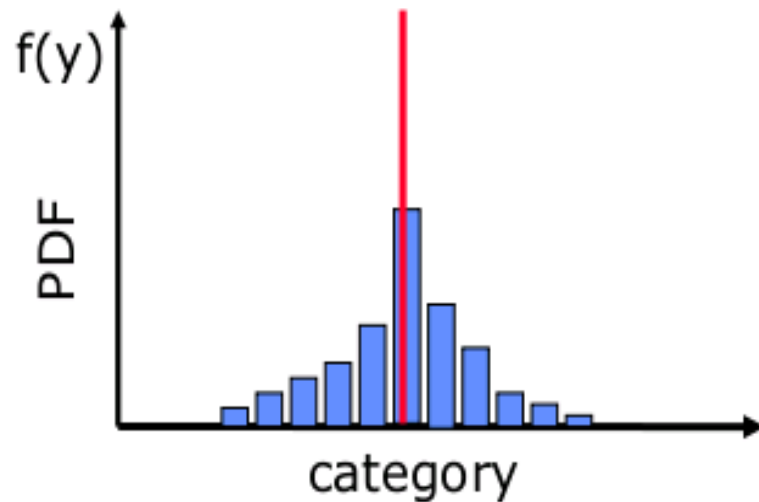


szükséges feltétele a megbízhatóságnak

Ezek közül melyik a legjobb EPS?

# Ensemble előrejelzések értékelése

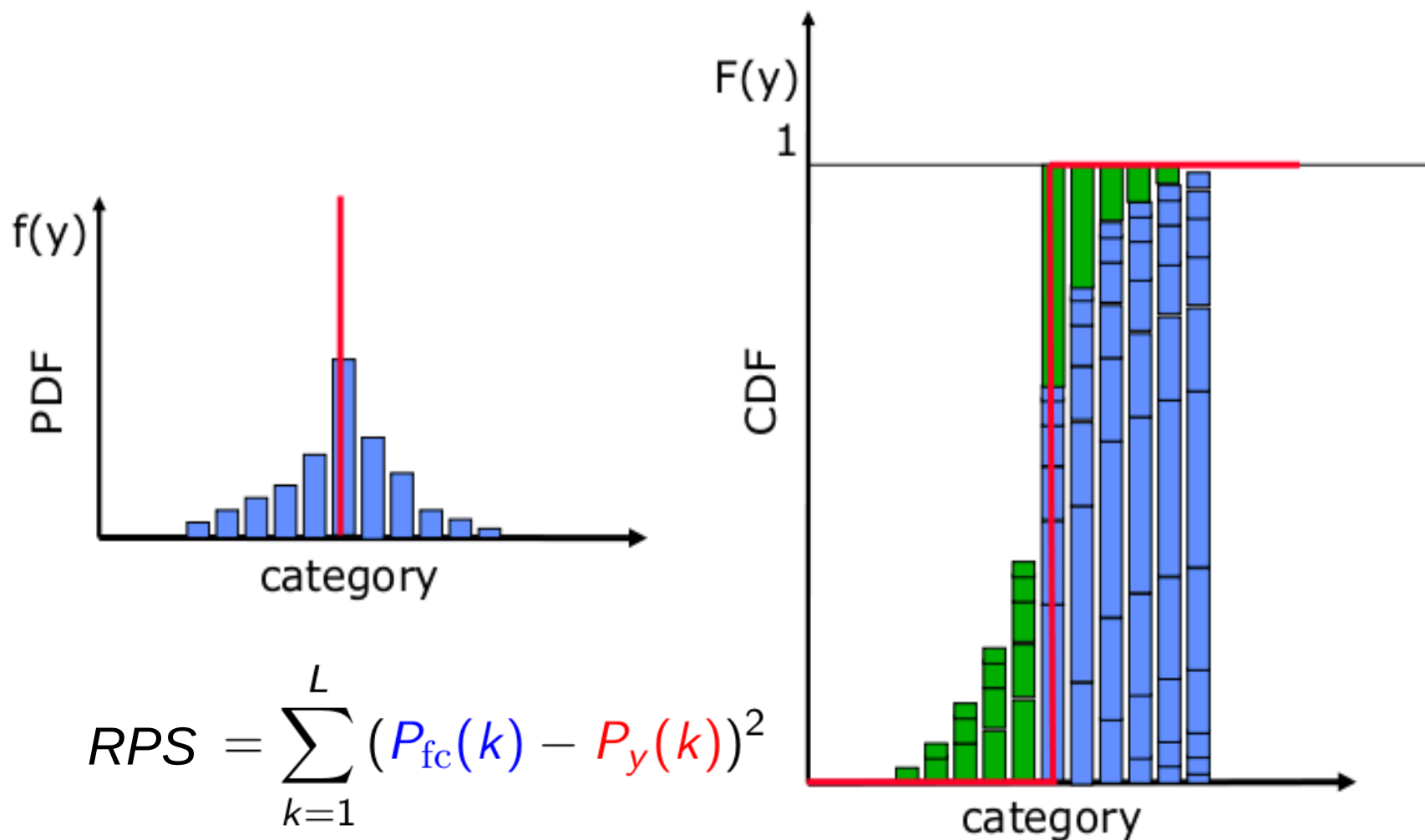
## Ranked Probability Score (RPS)



- mérés eloszlásfv.-e
  - EPS előrejelzés eloszlásfv.-e
- PDF: valószínűségi sűrűségfv.  
CDF: valószínűségi eloszlásfv.

# Ensemble előrejelzések értékelése

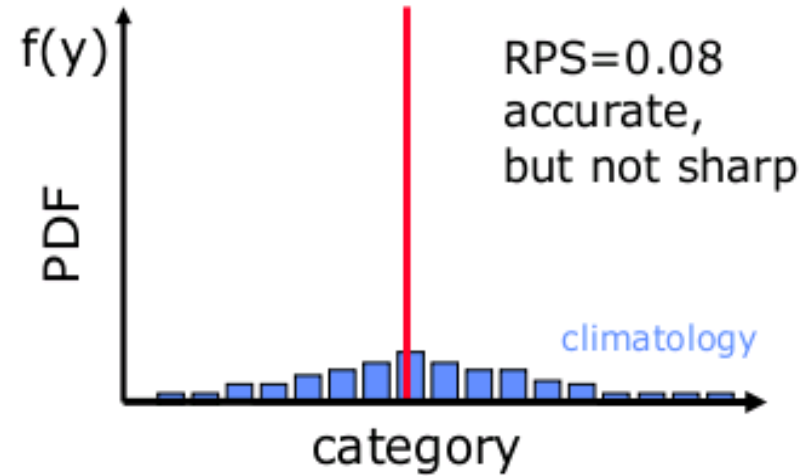
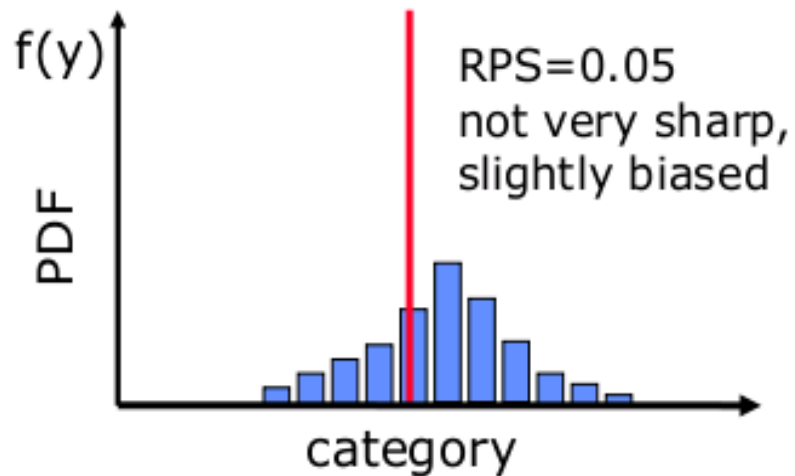
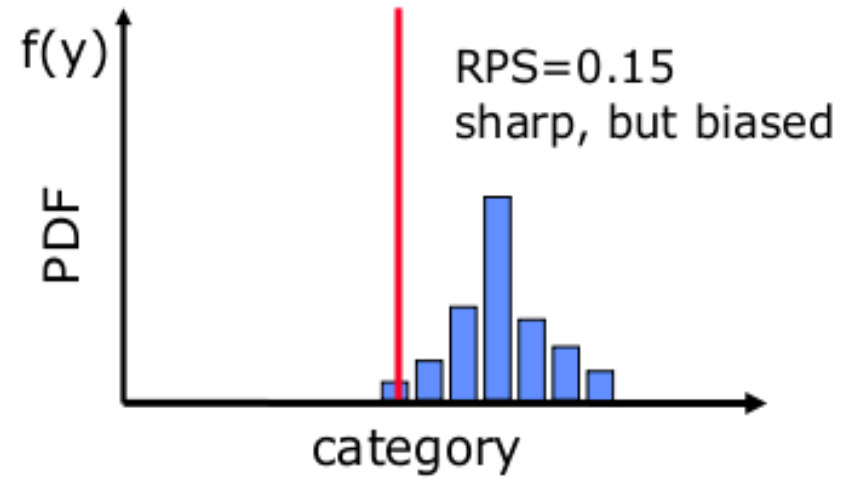
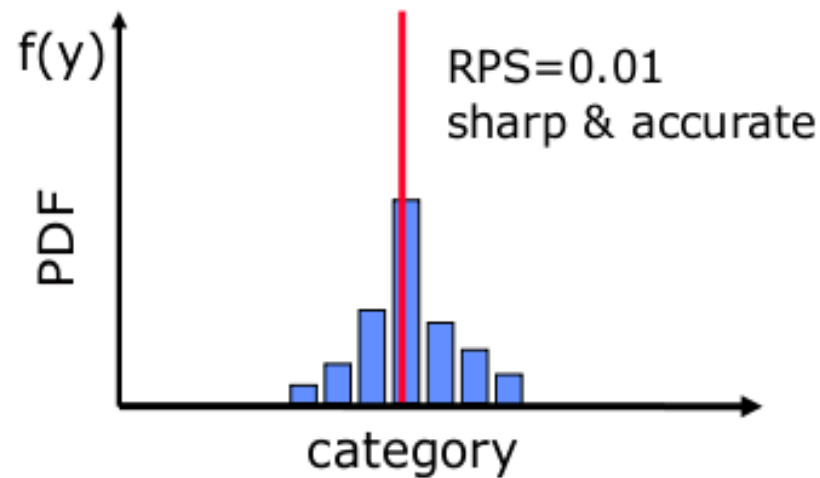
## Ranked Probability Score (RPS)



L: küszöbértékekkel definiált kategóriák  
(pl.:  $T > 0, 5, 10, 15$  °C)

# Ensemble előrejelzések értékelése

## Ranked Probability Score (RPS) - példák





# Ensemble előrejelzések értékelése

- Ranked Probability Skill Score

- A modell és a klíma előrejelzés *RPS* hányadosából képezzük
- (Hasonló képlettel származtatunk 'skill score'-t a 'score'-okból)

$$RPSS = 1 - \frac{RPS}{RPS_c}$$

- Continuous Ranked Probability Score

- Integráljuk az előrejelzett és a mért eloszlásfüggvény különbségének a négyzetét
- Az *RPS*-hez hasonló, de itt nem értéktartományokat (kategóriákat), hanem folytonos skálát használunk a vizsgált mennyiségre ( $x$ )

# Ensemble előrejelzések értékelése

- Continuous Ranked Probability Score

Minden esetben az  $N$  db előrejelzett érték egy valószínűségi eloszlásfüggvényt ad

Ennek a „távolságát” vesszük a mérés eloszlásfüggvényétől

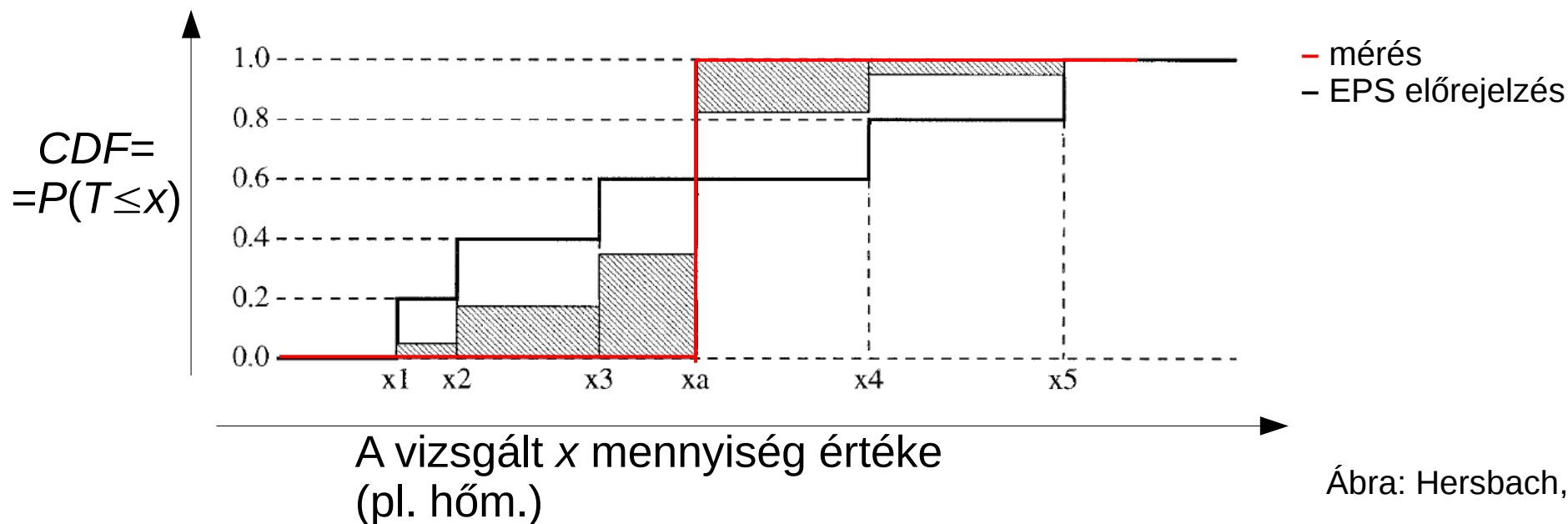
$$CRPS = \int_{-\infty}^{\infty} (CDF_{EPS}(x) - CDF_y(x))^2 dx$$

- Minél kisebb az értéke, annál jobb az előrejelzés

Az EPS egészéről ad információt

Értéke nagy, ha a bias vagy a szórás nagy

Pl.  $N=5$ -tagú EPS-re (egyetlen eset):



# Ensemble előrejelzések értékelése

- Continuous Ranked Probability Score

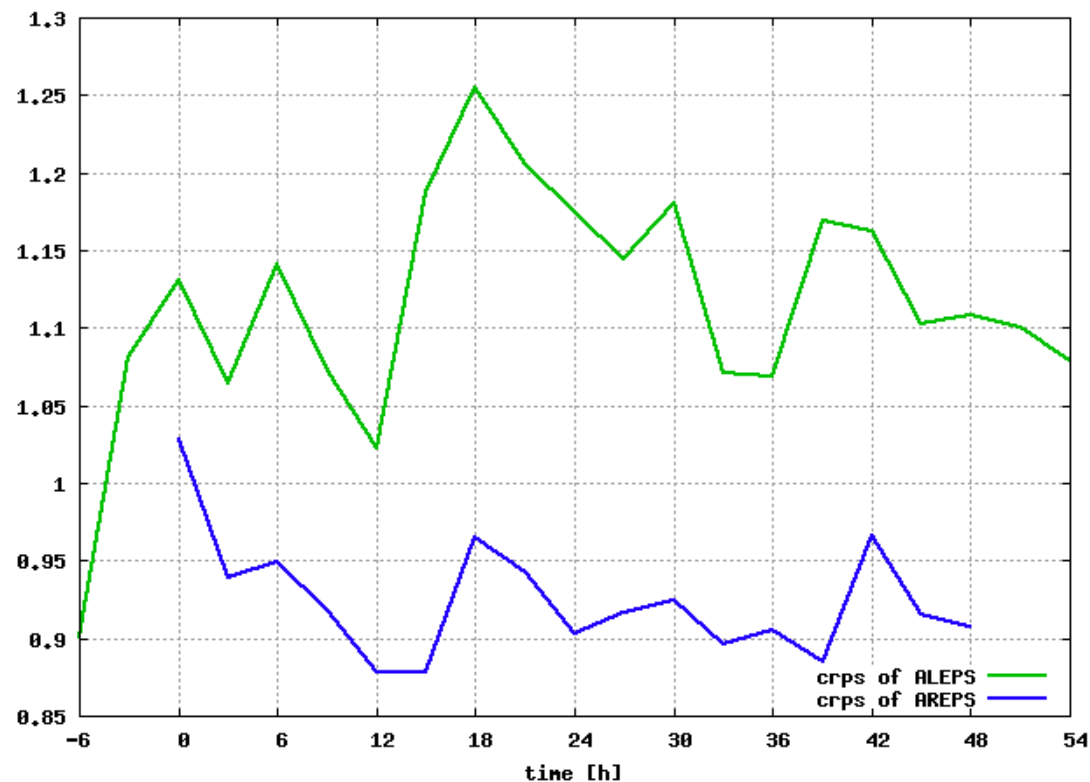
$$CRPS = \int_{-\infty}^{\infty} (CDF_{EPS}(x) - CDF_y(x))^2 dx$$

- Minél kisebb az értéke, annál jobb az előrejelzés

- Dimenziója megegyezik a vizsgált paraméterével

Szélesebesség CRPS  
2019 decemberi időszakra,  
30 magyar állomásra,  
OMSZ korábbi és jelenlegi LAM EPS

Continuous Ranked Probability Score, var:10mWindSpeed[m/s], level:Surface, location:allstat



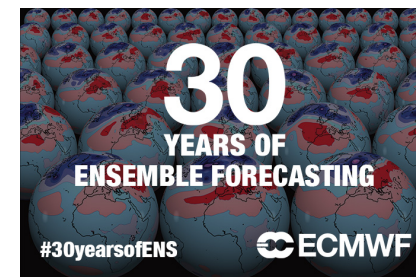
# Módszerek alkalmazása két rendszerben

- Két európai előrejelző központ globális modelljeiben:

- ECMWF (európai) - ENS

9 km

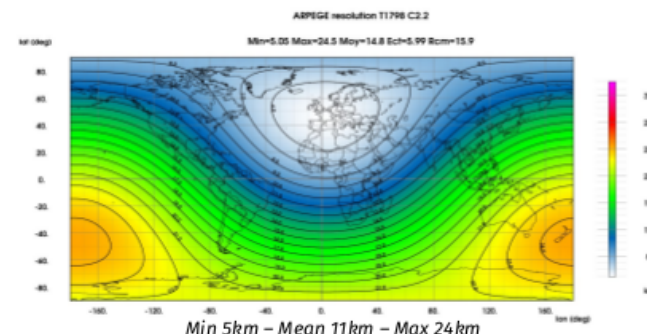
- 50+50+6 szinguláris vektor lineáris kombinációjából állítják elő SV perturbációkat
- 50 perturbált taggal fut az EDA. Az egyes EDA analízisek és az EDA EPS átlag különbségei jelentik az EDA perturbációkat
- A nagyfelbontású 4DVAR analízishez adják hozzá az SV és EDA perturbációkat => 50 tag + 1 kontroll
- SPPT-vel veszik figyelembe a modellhibát



- Météo France (francia) - PEARP

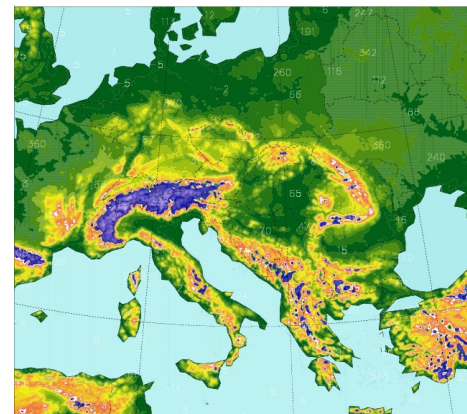
átl. 11 km

- 50 tagú 4DVAR EDA ciklus (6h)
- Ebből 35 tagot használnak, SV-okkal kombinálva => 35 tagú rövidtávú EPS
- PP + kétféle mélykonvekciós séma

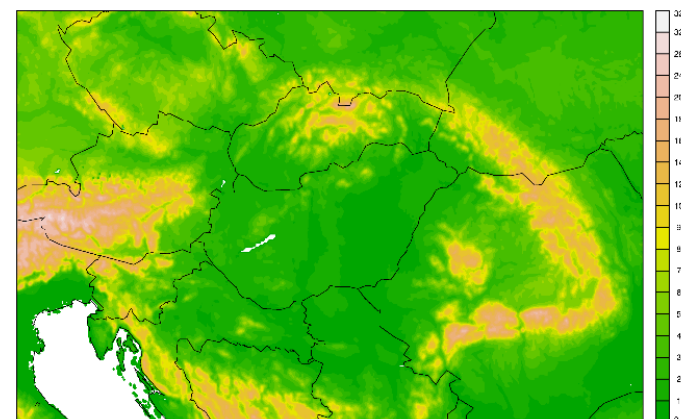


# Módszerek alkalmazása két rendszerben

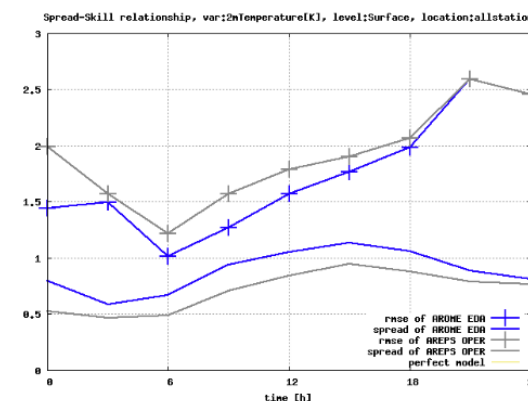
- OMSZ korlátos tartományú modelljeiben:
- ALADIN-EPS (ALARO fizikai parametrizáció)
  - 10+1 tagú ensemble rendszer
  - Korábbi operatív futás, 18UTC-ből indulva +60 órára, 8km felbontással
  - ECMWF ENS csatolás



- **AROME-EPS** („convection-permitting EPS”)
  - 10+1 tagú ensemble rendszer
  - **Operatív** futás 2020 óta. Napi 2x, 0/12 UTC +48h
  - ECMWF ENS csatolás
  - 2.5 km felbontású nem-hidrosztatikus rendszer, melyben finom struktúrák (konvektív események) is megjelennek, melyek előrejelezhetősége kisebb
  - Perturbáció: **EDA** (Lokálisan asszimilált mérések pert.)



Kísérletek SPP módszer alkalmazására



EDA  
oper.

2021  
július

# Ajánlott irodalom

- Buizza és Palmer, 1995.: The singular-vector structure of the atmospheric global circulation, J. Atmos. Sci., 52, 1434-1456
- Toth Z., Kalnay E., 1997: Ensemble Forecasting at NCEP and the Breeding Method, Mon. Weather Rew. 125, 3297-3319
- Leutbecher et al., 2017: Stochastic representations of model uncertainties at ECMWF: state of the art and future vision, Q. J. R. Meteorol. Soc. 143: 2315–2339
- Jolliffe és Stephenson: Forecast verification, Wiley, 2012
- <https://cawcr.gov.au/projects/verification/>

# Szinguláris vektorok (pótdia)

## vektor normája skalárszorzattal

- A következő, skaláris szorzat segítségével megadott hozzárendelés mindig normát ad meg:  $\|\mathbf{y}\| = \sqrt{\langle \mathbf{y}, \mathbf{y} \rangle}$
- A skaláris szorzat reprezentálható a  $D$  mátrix-szal az  $f_1, \dots, f_n$  bázison:  
$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^T D \mathbf{x} = \mathbf{x}^T D \mathbf{y}$$
- $D$  mátrix szimmetrikus és pozitív definit:  $D_{ij} = \langle f_j, f_i \rangle$
- A leírtakból következik, hogy a perturbációk nagyságára felírható:

$$\|\mathbf{y}\| = \sqrt{\langle \mathbf{y}, \mathbf{y} \rangle} = \sqrt{\mathbf{y}^T D \mathbf{y}}$$

# Szinguláris vektorok (pótdia) a norma megválasztása

Teljes energiára alapozott norma (ECMWF):

$$\|x\|^2 = \frac{1}{2} \int_0^1 \int_S (\nabla \Delta^{-1} \xi \cdot \nabla \Delta^{-1} \xi + \nabla \Delta^{-1} D \cdot \nabla \Delta^{-1} D + R_a T_r (\ln \pi)^2 + \frac{c_p}{T_r} T^2) dS \sum \left( \frac{\partial p}{\partial \eta} \right) d\eta$$

Ahol a divergenciát és az örvényességet a horizontális szélkomponensekből származtatjuk a következő módon:

$$D = \nabla_H V_H = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}$$

$$\xi = \nabla_H \times V_H = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}$$

$\pi$  a felszíni nyomás,  $\eta$  a felszínkövető vert. koord.,  $R_a$  gázáll.,  $S$  a teljes Föld-felszín,  $T_r$  refer. hőm.,  $c_p$  a fajhő áll. nyomáson



# Szinguláris vektorok (pótdia)

- A numerikus modell deriváltjára van szükség
- $M$ -et, mint mátrixot nem tudják elkészíteni, csak a neki megfelelő algoritmust
- $Mx$ -et a tangens lineáris modell integrálásával kapják
- Ebből kiszámolják az  $M^*$  adjungáltat, visszafelé integrálják az adjungált modellt

# Szinguláris vektorok (pótdia) a leggyorsabban fejlődő perturbációk

- $\|My(t_0)\|$  kifejtése a kvadratikus alakkal:

$$\|My(t_0)\| = \sqrt{\langle M^* My(t_0), y(t_0) \rangle} = \sqrt{Q_{M^* M}(y(t_0))}$$

- ahol:

$$Q_{M^* M}(y(t_0)) = \sum_{i=1}^n \lambda_i^2 |y_i(t_0)|^2$$

$$y(t_0) = \sum_{i=1}^n y_i(t_0) v_i$$

- $M^* M$  legnagyobb  $\lambda_1^2$  sajátértéke adja a legnagyobb tagot a kvadratikus alakhoz. Az ehhez tartozó  $v_1$  sajátvektor pedig mutatja az irányt, melynek perturbációja a leggyorsabban fog nőni  $(t_0, t_1)$ -en

# Szinguláris vektorok (pótdia) a leggyorsabban fejlődő perturbációk

- $(M^*M)$  sajátvektorai egyben az  $M$  ún. szinguláris vektorai, melyekre igaz, hogy:

$$M=USV^* \quad \Leftrightarrow \quad Mv_i = \lambda_i u_i$$

ahol  $U$  ill.  $V$  ortogonális mátrixok,

$S$  pedig diagonális mátrix, melynek főátlójában a  $\lambda_i$  szinguláris értékek vannak

$U$  oszlopvektorai az  $M$  bal oldali  $u_i$  („evolved”) szinguláris vektorai

$V$  oszlopvektorai az  $M$  jobb oldali (kezdeti)  $v_i$  szinguláris vektorai

Valamint  $v_i$ -k egyben  $(M^*M)$  sajátvektorai, méghozzá  $\lambda_i^2$  sajátértékkel

- $M^*M$  legnagyobb  $\lambda_i^2$  sajátértékéhez tartozó  $v_1$  sajátvektora mutatja az irányt, melynek perturbációja a leggyorsabban fog nőni  $(t_0, t_1)$ -en