

Ensemble előrejelzések II.

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

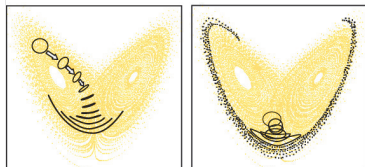
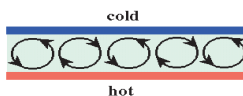
Előadó: Jávorné Radnóczy Katalin (radnoczy.k@met.hu)

Korábbi előadók: Hágel Edit, Horányi András,
Böloni Gergely, Szűcs Mihály

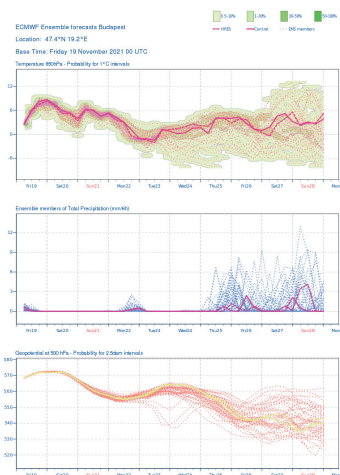
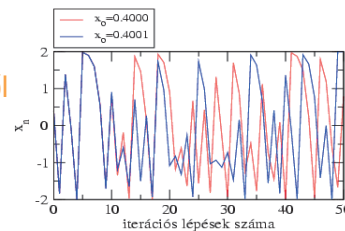


Ismétlés Kaotikus rendszerekről

- Láttuk a káosz megjelenését egyszerű, determinisztikus, nemlineáris rendszerekben.
- Ezekhez hasonlóan a légkört leíró HTDER is nagyon érzékeny a kezdeti feltételekre: **kis különbség a kezdeti mezőkben nagy eltérést jelenthet az előrejelzésben.**



Az előrejelezhetőség függ a légkör aktuális állapotától

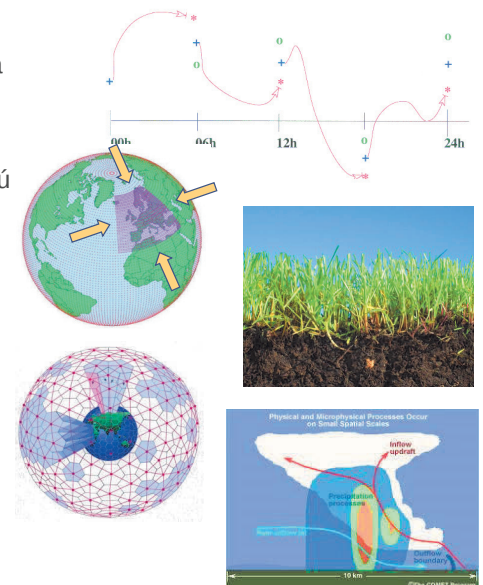


Az előző óra tartalma

- 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 légköri előrejelzés bizonytalanságai

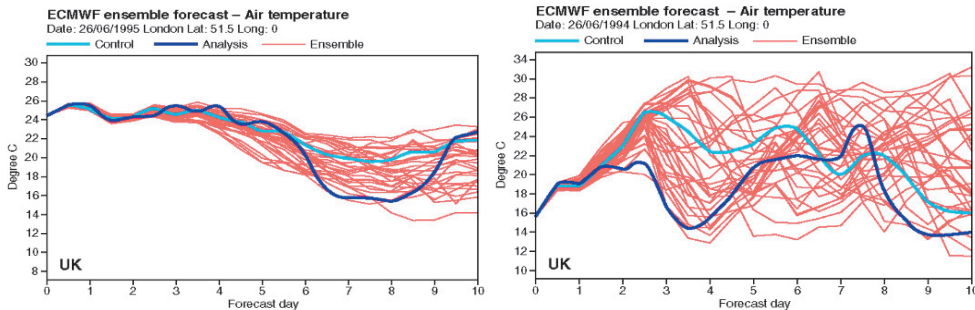
- A légköri előrejelzések hibáinak forrása a gyakorlatban:
 - Kezdeti feltételek**
 - Peremfeltételek** (korlátozott tartományú modelleknél oldalsó peremfeltételek)
 - Felszín** (alsó peremfeltétel, pl.: talajhőmérséklet, talajnedvesség, hóvastagság)
 - A HTDER diszkretizálása (séma, koordináta rendszer,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



Ismétlés

A légkör előrejelezhetősége

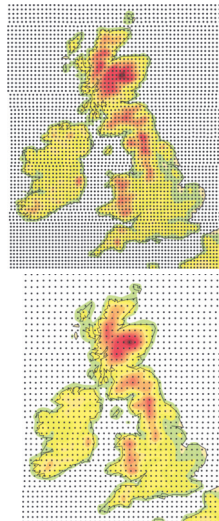
- Az előrejelezhetőség függ attól, hogy:
 - Hova, milyen paramétert akarunk előrejelezni? Milyen skálájú folyamatot akarunk leírni?
500hPa geopotenciál vagy 10 perces csapadékösszeg előrejelzése a könnyebb?
 - Milyen kezdeti analízis áll rendelkezésünkre? Milyen modellt használunk?
 - Milyenek a légköri feltételek? - Áramlásfüggő!



Ismétlés

EPS alapfogalmak

- Mint láhattuk, egy-egy előrejelzés elkészítése rendkívül költséges.
- Annak érdekében, hogy több előrejelzést futtathassunk, kompromisszumokat kell kötnünk
- **Rosszabb felbontáson**, nagyobb időlépcsővel szoktak ensemble előrejelzések készítése során futtatni.
- A jobb felbontású, 'hagyományos' módszerrel készült előrejelzést szokás '**determinisztikus**' előrejelzésnek is nevezni, de egyre inkább a '**nagy-felbontású**' vagy '**HRES**' név jön divatba.
- Természetesen az **ensemble** előrejelzések egyes tagjai is determinisztikusak.
- EPS=Ensemble Prediction System**
- Kontroll tag**=Perturbálatlan, rosszabb felbontáson futtatott tagja a sokaságnak

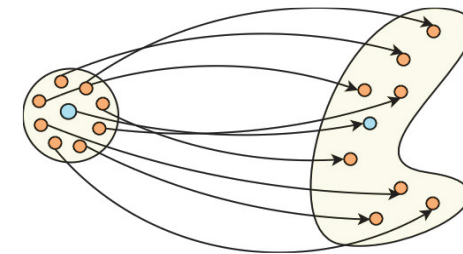


Ismétlés

EPS motiváció

- Megoldás az előrejelzés során eluralkodó bizonytalanságok számszerűsítésére:
 - Nem egyetlen** előrejelzést készítünk a „legjobbnak ítélt” kezdeti feltételtől 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.

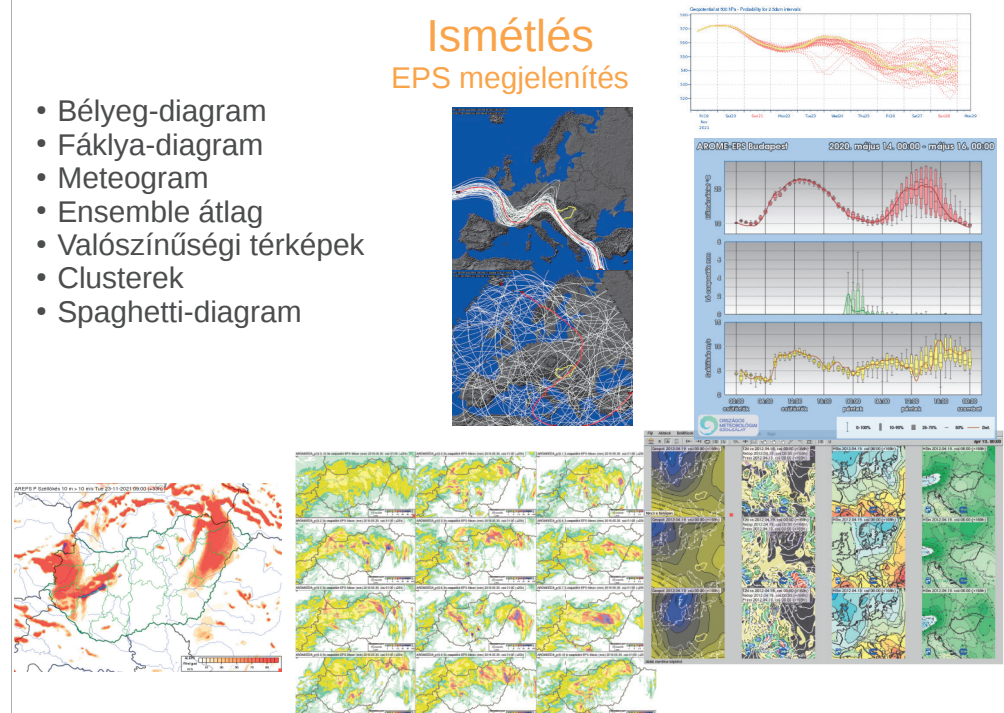
⇒ valószínűségi megközelítés lehetségessé válik
következtetni lehet az előrejelzés bizonytalanságára



Ismétlés

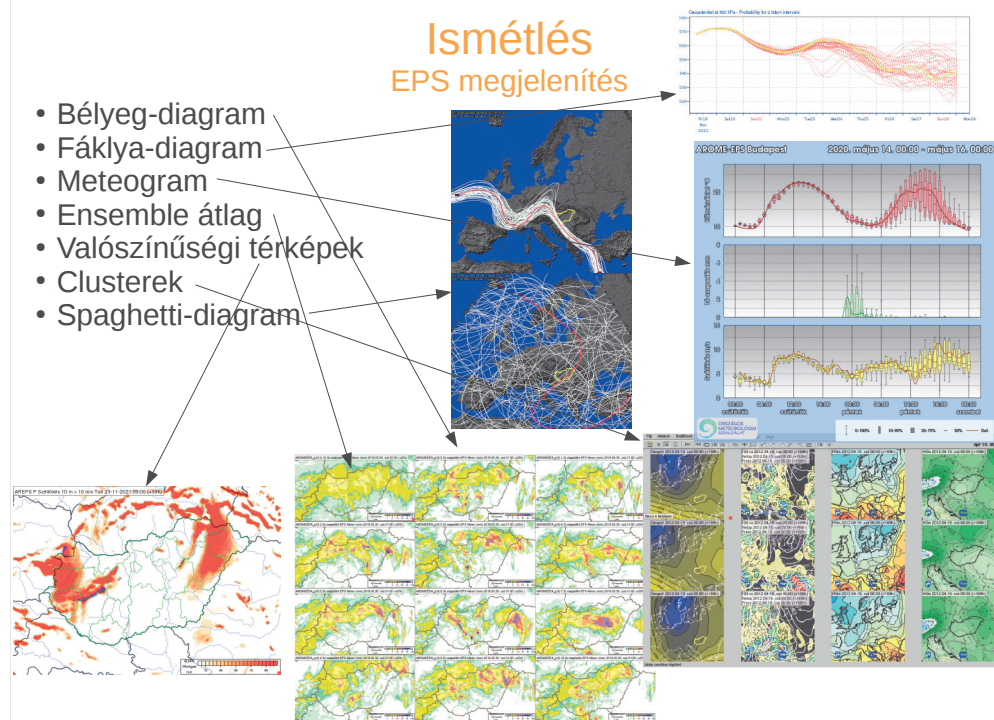
EPS megjelenítés

- Bélyeg-diagram
- Fáklya-diagram
- Meteogram
- Ensemble átlag
- Valószínűségi térképek
- Clusterek
- Spaghetti-diagram



Ismétlés EPS megjelenítés

- Bélyeg-diagram
- Fáklya-diagram
- Meteogram
- Ensemble átlag
- Valószínűségi térképek
- Clusterek
- Spaghetti-diagram



Ensemble előrejelzések felhasználása

• Üzleti döntések meghozatala:

- Egy utcára nyíló kávézó rendelkezik terasszal és azon 20 szabadtéri hellyel is.
- A tulajdonosnak két nappal korábban el kell döntenie, hogy kipakol-e a teraszra és annak üzemeltetéséhez alkalmaz-e extra munkaerőt.
- Ezt a döntést az eső valószínűségének függvényében hozza meg.



Ensemble előrejelzések felhasználása

- Tegyük fel, hogy az extra munkaerő bére 10.000 Ft, az elérhető plusz bevétel 50.000 Ft
- Ha a tulajdonos nem alkalmaz extra munkaerőt, a plusz bevétel 0 Ft, de költsége sincsen
- Ha alkalmaz extra munkaerőt
 - és eső lesz, akkor költsége 10 000 Ft
 - és nem lesz eső, akkor a költsége 10 000 Ft, bevétele pedig 50 000 Ft
- Tegyük fel, hogy annak a valószínűsége, hogy a kérdéses napon eső lesz = P

Milyen P -re éri meg kinyitni a teraszt?



Ensemble előrejelzések felhasználása

$$\text{BEVÉTEL} = P \cdot (0 - 10.000) + (1 - P) \cdot (50.000 - 10.000) = 40.000 - 50.000P > 0$$

- Ha tehát $P < 0.8$ megéri extra munkaerőt alkalmazni, míg $P > 0.8$ esetében nem
- 'Determinisztikus' előrejelzés esetén a tulajdonos mindig alkalmazna extra munkaerőt ha nem jelzünk előre esőt, és soha, amikor szerepel a prognózisban csapadék
- 'Determinisztikus' esetben az igen/nem döntést maga a modell igen/nem kimenetele határozná meg. Ensemble előrejelzés esetén az igen/nem döntést az esemény valószínűsége és az árak együttesen határozzák meg.

Ensemble előrejelzések felhasználása

- Egy építési vállalkozónak sürgősen le kell aszfaltoznia egy útfelületet. Viszont ebben a P valószínűséggel érkező eső megakadályozhatja.



- Ha elvégzi rendben a munkát, akkor 1millió forintot kap
- A munkaerő és az aszfalt költsége 400 ezer forint, amit mindenképp ki kell fizetnie, ha kiszáll a helyszínre; még akkor is, ha eső miatt nem tudnak dolgozni
- Ha nem végzi el a munkát, akkor mindenképp 100 ezer forint kötbért kell fizetnie.
- Mi az a P valószínűségi érték, ami alatt a vállalkozónak érdemes kimenni aszfaltozni?

Ensemble előrejelzések felhasználása

- Egy építési vállalkozónak sürgősen le kell aszfaltoznia egy útfelületet. Viszont ebben a P valószínűséggel érkező eső megakadályozhatja.
 - Ha elvégzi rendben a munkát, akkor 1millió forintot kap
 - A munkaerő és az aszfalt költsége 400 ezer forint, amit mindenképp ki kell fizetnie, ha kiszáll a helyszínre; még akkor is, ha eső miatt nem tudnak dolgozni
 - Ha nem végzi el a munkát, akkor mindenképp 100 ezer forint kötbért kell fizetnie.
- Mi az a P valószínűségi érték, ami alatt a vállalkozónak érdemes kimenni aszfaltozni?

BEVÉTEL=

>

Ensemble előrejelzések felhasználása

- Egy építési vállalkozónak sürgősen le kell aszfaltoznia egy útfelületet. Viszont ebben a P valószínűséggel érkező eső megakadályozhatja.
 - Ha elvégzi rendben a munkát, akkor 1millió forintot kap
 - A munkaerő és az aszfalt költsége 400 ezer forint, amit mindenképp ki kell fizetnie, ha kiszáll a helyszínre; még akkor is, ha eső miatt nem tudnak dolgozni
 - Ha nem végzi el a munkát, akkor mindenképp 100 ezer forint kötbért kell fizetnie.
- Mi az a P valószínűségi érték, ami alatt a vállalkozónak érdemes kimenni aszfaltozni?

$$\text{BEVÉTEL} = P(0 - 400000 - 100000) + (1 - P)(1000000 - 400000) > -100000$$

$$-500000P + 600000 - 600000P > -100000$$

$$-1100000P > -700000$$

$$P < 7/11 \sim 64\%$$

Tartalom

- 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

Ensemble előrejelzések készítésének módszerei

- 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.

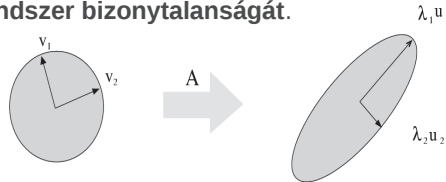
Ensemble előrejelzések készítésének módszerei

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

Szinguláris vektorok Motiváció

- Milyen az optimális kezdeti feltétel perturbáció?

- A légköri 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 légköri rendszer bizonytalanságát**.



- Történelmi háttér

- Az egyik első módszer perturbációk előállítására. (1992 - ECMWF)
- A szinguláris vektorok elve jó összhangban van a légkör 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:

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

- $x(t)$ 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:

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

Szinguláris vektorok perturbációk fejlődése

- Az előző két egyenlet alapján 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:

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

- Arra vagyunk kíváncsiak, hogy M milyen perturbációkat növel a legnagyobb mértékben 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

Szinguláris vektorok a leggyorsabban fejlődő perturbációk

- A következő, skaláris szorzat segítségével megadott hozzárendelés mindig normát ad meg: $\|y\| = \sqrt{\langle y, y \rangle}$

- A skaláris szorzat reprezentálható a D mátrix-szal az $f_1, \dots, f_n$ bázison:

$$\langle x, y \rangle = y^T D x = x^T D y$$

- D mátrix szimmetrikus és pozitív definit: $D_{ij} = \langle f_j, f_i \rangle$
- A leírtakból következik, hogy a perturbációk nagyságára felírható:

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

Szinguláris vektorok a norma megválasztása

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

- Logikusnak tűnhet az euklideszi skaláris szorzatból származtatni a normát. Azonban y perturbációs vektor különböző fizikai változókat tartalmaz (hőmérséklet, nedvesség, nyomás), így pl. a hőmérséklet értéktartománya miatt dominánsabb lenne, mint pl. a specifikus nedvesség.
- Ezért inkább az **energia normák** használata terjedt el:
- A légkör teljes energiája három komponensből tevődik össze:
 - Mozgási energia
 - Helyzeti energia
 - Belső energia

Szinguláris vektorok a norma megválasztása

Teljes energiára alapozott norma (ECMWF):

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

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

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

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

π a felszíni nyomás, η a felszínkövető vert. koord., R_a gázáll., S a teljes Föld-felszín, T_r refer. hőm., c_p a fajhő áll. nyomáson

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

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

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

- Válasszuk ugyanazt (pl. a teljes energia) normát a kezdeti és végidőpontban egyaránt (ez a gyakorlatban nem mindig teljesül):

$$\frac{\|My(t_0)\|}{\|y(t_0)\|} = \left\| M \frac{y(t_0)}{\|y(t_0)\|} \right\|$$

- Mostantól elegendő az M propagátor $\|y(t_0)\|=1$ egységvektoron felvett $\|My(t_0)\|$ értékeit vizsgálni

Szinguláris vektorok 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 λ_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 jobb oldali szinguláris vektorok

- A kapott $\lambda_1 \geq \dots \geq \lambda_n$ 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 λ_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 λ_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 λ_3 lesz a maximum és a v_3 vektor adja meg a maximum helyét. Tovább folytatva a sort eljutunk egészen λ_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 jobb oldali szinguláris vektorok

Tetszőleges szinguláris vektorra igaz tehát:

$$M^* M v_i = \lambda_i^2 v_i$$

$$\|v_i(t)\| = \sqrt{\langle M^* M v_i(t_0), v_i(t_0) \rangle} = \sqrt{\langle \lambda_i^2 v_i(t_0), v_i(t_0) \rangle} = \lambda_i \|v_i(t_0)\|$$

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

Szinguláris vektorok bal oldali szinguláris vektorok

- Vegyük a következő összefüggést:

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

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

$$u_i = \frac{M v_i}{\|M v_i\|}$$

Szinguláris vektorok

bal oldali szinguláris vektorok

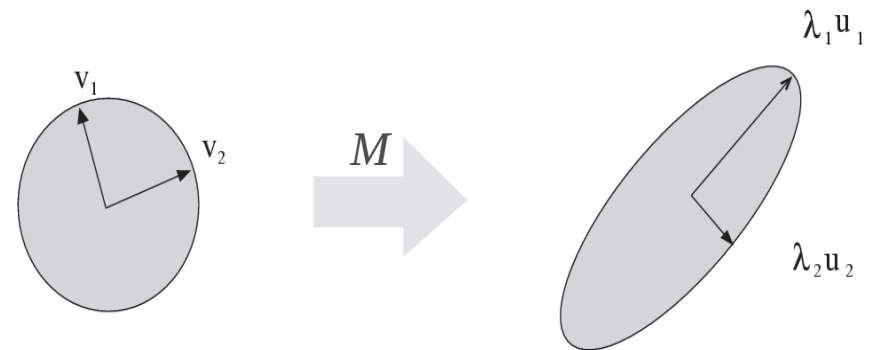
- Mivel: $\|Mv_i\|^2 = \langle Mv_i, Mv_i \rangle = \langle v_i, M^* M v_i \rangle = \langle v_i, \lambda_i^2 v_i \rangle = \lambda_i^2$
- Az előző kettőből adódik: $Mv_i = u_i \|Mv_i\| = \lambda_i u_i$
- Az M propagátor a jobb oldali szinguláris vektorokat az $u_1, \dots, u_n$ vektorok irányába képezi 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

- 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 függenek:
 - A norma megválasztásától
 - Az optimalizálási időtől (pl.: 12, 24, 48 óra)
 - Az optimalizálási tartománytól (lehet globális SV, vagy lehet céltartományt definiálni)

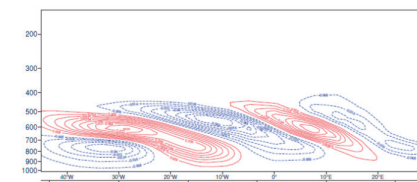
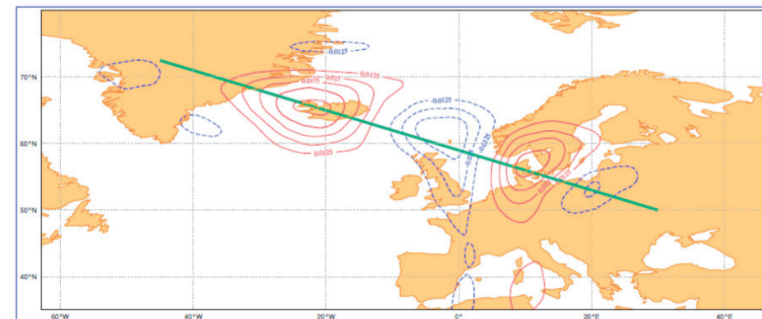
Szinguláris vektorok

Szemléltetés 2 dimenzió esetén:



Szinguláris vektorok

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

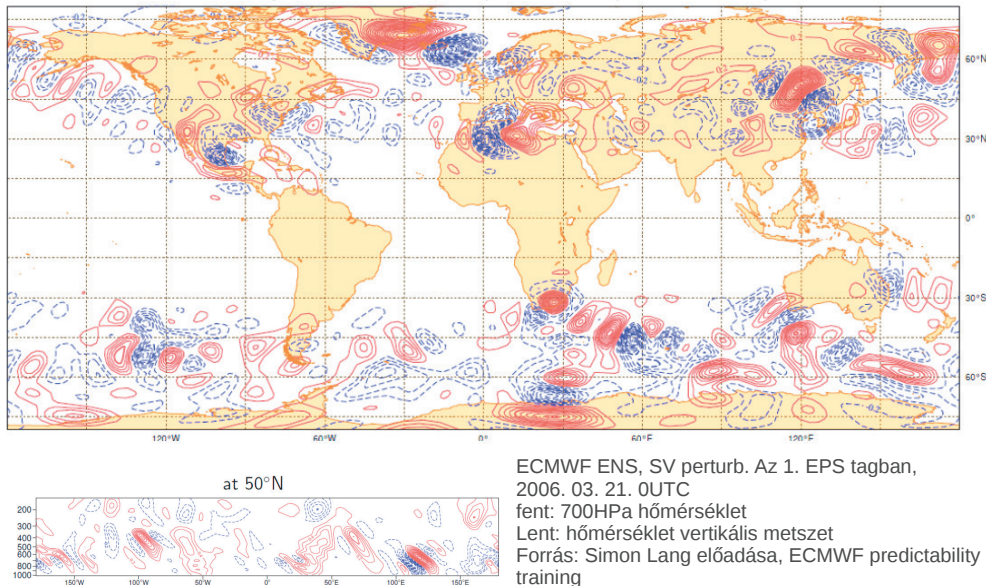


ECMWF ENS, 5. szinguláris vektor,
2006. 03. 21. 0UTC
fent: 700hPa hőmérséklet
Lent: hőmérséklet vertikális metszet a zöld
vonal mentén

Forrás: Simon Lang előadása, ECMWF
predictability training

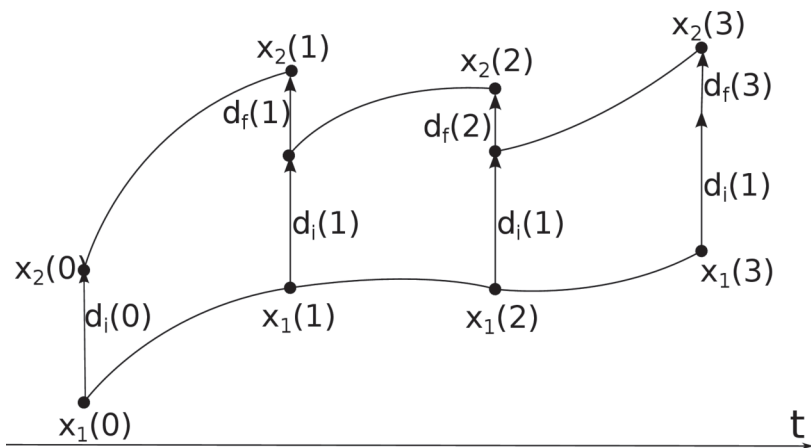
Szinguláris vektorok

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



Breeding módszer

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

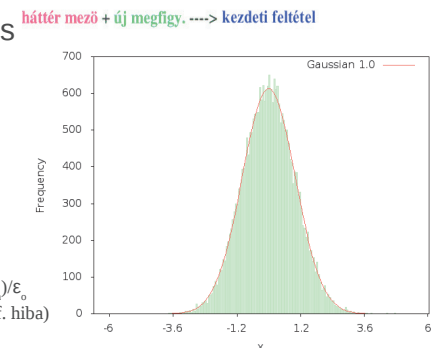
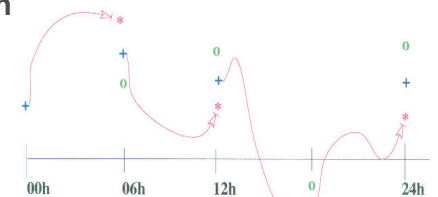


Breeding módszer

- A szinguláris vektoroknál egyszerűbb, gyakorlatiasabb és kevésbé költséges módszer.
- Szintén az egyik első ensemble módszer (1992 - NCEP)
- 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 perturbációkat

Ensemble adatasszimiláció

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



$$x = (y_{\text{perturbált}} - y_{\text{perturbálatlan}}) / \epsilon_0$$

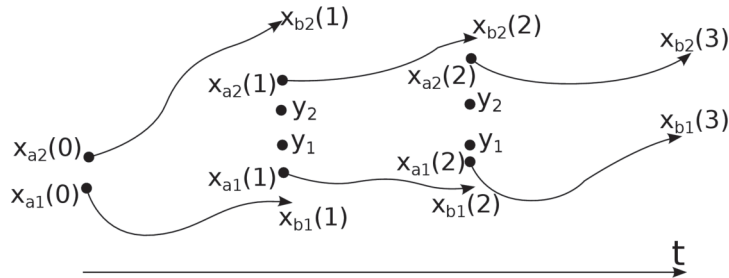
(y: megfigyelt érték, ϵ_0 : megf. hiba)

Ensemble adatasszimiláció

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

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

$$K = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}$$



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

Ismétlés - Kezdeti feltétel perturbációk

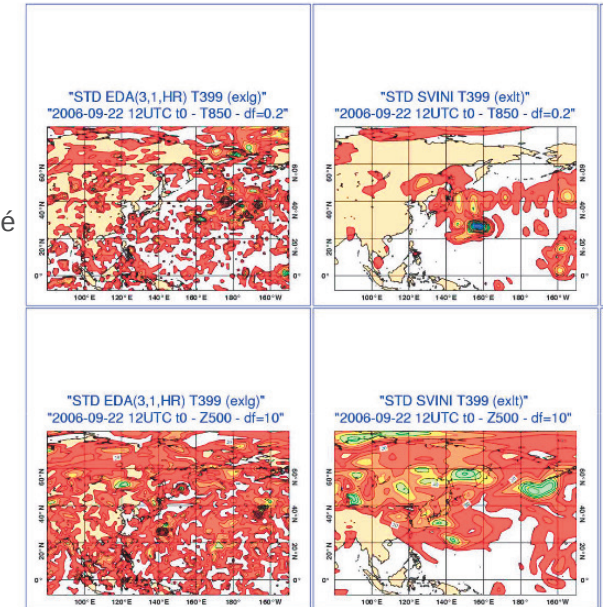
A három vizsgált módszer filozófiája kicsit különbözik:

- **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 előállítás.
- **Breeding módszer**
 - Cél, hogy egy ciklus során véletlen perturbációkból olyanokat tenyesszen ki a rendszer, mely az előző időszak bizonytalanságait reprezentálja.
- **Ensemble adatasszimiláció**
 - Mérési hibahatáron belüli perturbációkat ültetünk a mérésekre.
 - Cél, hogy párhuzamosan futtatott adatasszimilációs ciklusokon keresztül határozzuk meg a kezdeti feltétel bizonytalanságait.

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.

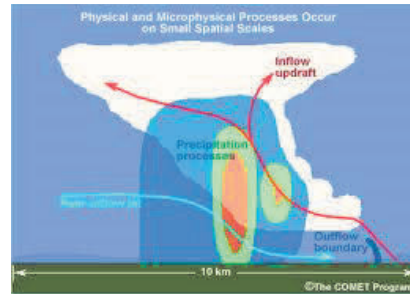


Modellhiba reprezentációja

- A modellszámításokat szokás egyszerűen két részre bontani:
 - **“Dinamika”** - a HTDER diszkrétizálása és megoldása
 - **“Fizika”** - a rács távolságnál 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.

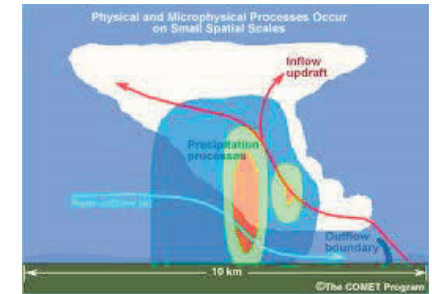
Multi-fizika ensemble

- A ráctávolságnál 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

- A ráctávolságnál kisebb skálájú, illetve túlságosan bonyolult folyamatok leírásárását **fizikai parametrizációval** végezzük.
- Az alkalmazott paraméterek értékeit perturbálják egy-egy véletlen számmal (ésszerű határértékeken belül).



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:

$$e_j(T) = \int \{A(e_j; t) + P(e_j; t)\} dt$$

- A j-edik tag állapota **T** időpontban a következő modellegyenlet integrálásával adható meg:

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

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

- **A** – a nem-parametrizált folyamatok (dinamika)
- **P** – a parametrizált folyamatok (fizika)
- **δe** – kezdeti feltétel perturbációk (a korábban bemutatott módszerekből származhat)

SPPT – eredeti séma

- A rendszer koncepciója perturbált modell egyenletek esetén:

$$e_j(T) = \int \{A(e_j; t) + P'(e_j; t)\} dt$$

$$\frac{\partial e_j}{\partial t} = A(e_j; t) + P'(e_j; t)$$

- **P'** – a parametrizált folyamatok perturbált tendenciája

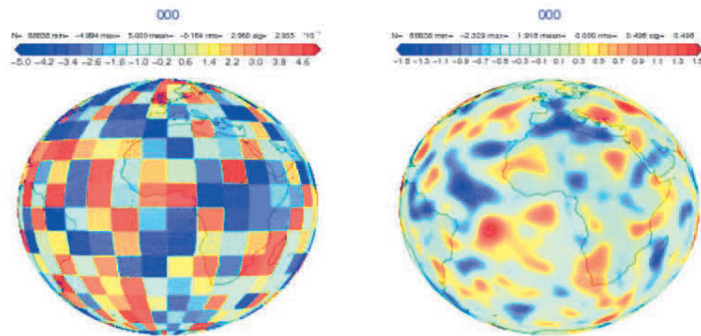
$$P'_j(e_j; t) = (1 + \langle r_j(\lambda; \varphi; t) \rangle_{D,T}) P_j(e_j; t)$$

- **r_j** – Egyenletes eloszlásból vett véletlen szám, ami állandó a **$D \times D$** rácpont méretű területen **T** ideig.

- **r** mérete és tér- és időbeli korrelációjára többféle lehetőség van, pl.:

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

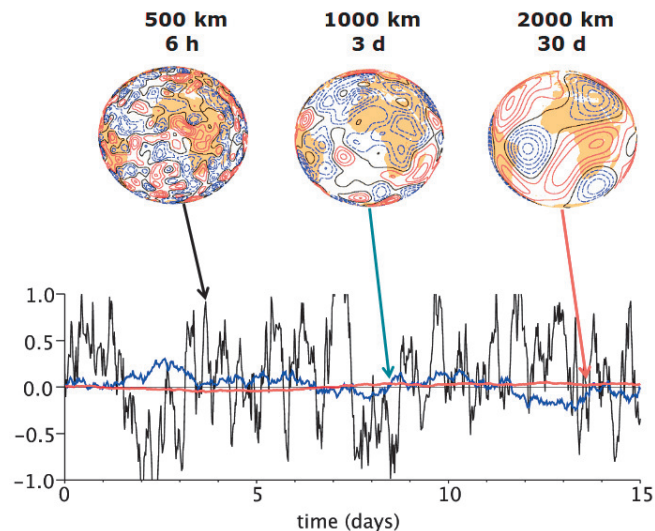
SPPT – motiváció a séma felülvizsgálatára



Bal oldal – az eredeti séma perturbációi egy adott időszakon
 Jobb oldal – olyan simább mező, ami jobban összhangban van a modellhiba tér- és időbeli szerkezetével

SPPT – felülvizsgált séma

r előállítható többféle mintázat kombinációjaként is (különböző skálájú folyamatok leírásának hibáját próbálják így reprezentálni):



SPPT – felülvizsgált séma

- Az új sémában r sokkal szofisztikáltabb módon kerül meghatározásra:

$$P'(e_j; t) = (1 + \alpha r_j) P_j(e_j; t)$$

- r olyan Gauss eloszlásból vett érték, melynek 0 a várható értéke és σ a szórása, valamint az $[-n\sigma; n\sigma]$ intervallumba van szorítva, ahol $n = 2$ vagy 3 .
- α -val magasságfüggővé lehet tenni (általában az alsó és felső perem közelében csillapítják a perturbációkat)
- r -t egy spektrális minta generátorral hozzuk létre:

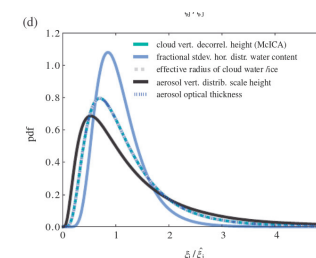
$$r = \sum_{mn} (\hat{r})_{mn} Y_{mn}$$

- r -t Y_{mn} gömbharmonikus és $(r)_{mn}$ spektrális együttható határozza meg
- $(r)_{mn}$ -et egy autoregresszív folyamat határozza meg, ahol:
 L a horizontális korrelációs hossz
 τ az időbeli korrelációt határozza meg



SPP – Stochastically perturbed parameters

- Paramétereket perturbálják *lokálisan*, térben és időben korrelált módon
- Meghatározzák azon legbizonytalanabb paramétereket, amelyek leginkább hozzájárulnak az előrejelzés hibájához
- Ezeket többnyire log-normál eloszlás szerint perturbálják egymástól függetlenül
- Fizikai folyamatokhoz kapcsolódik (paraméterek, perturbációk)
 $\Rightarrow$ Folyamat-szintű
- 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

Forrás: Leutbecher et al. 2017, ECMWF ENS

Multi-modell ensemble

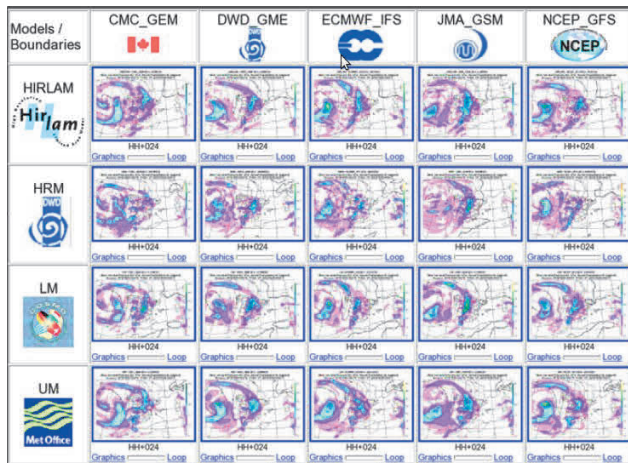
- Több 'determinisztikus' modellt rakunk össze egy ensemble rendszerbe. Ez különböző intézetek/előrejelző központok modelljeit jelenti.
- Lehetnek maguk a modellek is különbözőek, de általában minden intézet kicsit más beállításokkal futtatja a modelljeit, esetleg más méréseket vesz figyelembe az adatasszimiláció 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

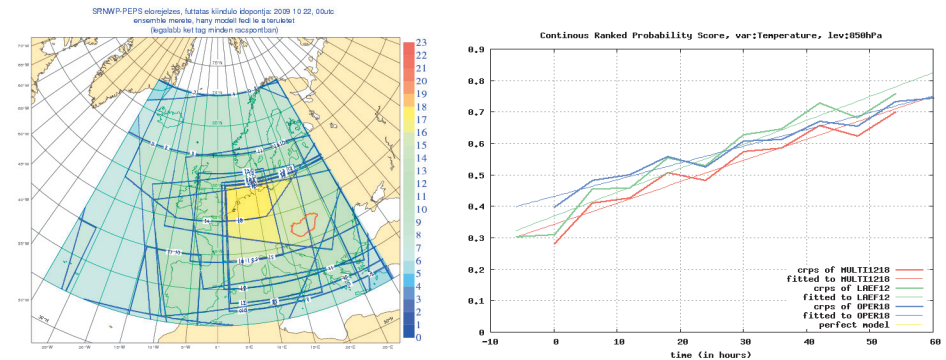
Multi-analízis ensemble

- 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.



Multi-modell ensemble

- 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 bemutattható, hogy jobb a multi-modell ensemble (nemzetközi összefogások célozzák közös adatbázisok létrehozását)



Ensemble előrejelzések értékelése

- A 'determinisztikus' előrejelzés esőt mondott
 - Az eső esett
 - Ez egy jó előrejelzés volt?
 - ☐ Igen
 - ☐ Nem
 - ☐ Nem tudom



Ensemble előrejelzések értékelése

- A '**determinisztikus**' 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 '**determinisztikus**' 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)

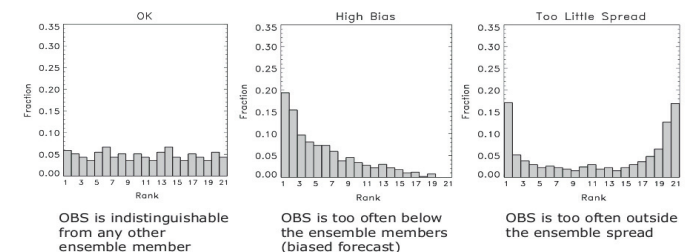
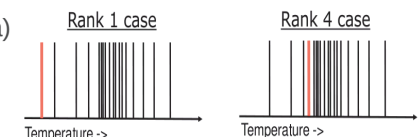
Ensemble előrejelzések értékelése

- Egy valószínűségi előrejelzés karakterisztikái:
 - **Konzisztencia:** A megfigyelések statisztikailag ugyanahhoz az eloszláshoz tartoznak, mint az előrejelzett ensemble.
 - **Megbízhatóság:** Megbízhatok a valószínűségekben, amiket az előrejelzésem mond? (pl.: Tényleg 20% az esély?)
 - **Élesség:** Mennyiben különbözik az előrejelzés a klímaátlag valószínűségétől (pl.: 33% eséllyel fog esni holnap)
 - **Pontosság:** Mennyiben különbözik az előrejelzés a klímaátlag valószínűségétől? Jól megfogta a modell?
 - **Skill:** Jobb az előrejelzésem, mint a referencia? (a referencia lehet véletlen, klíma, perzisztencia)



Ensemble előrejelzések értékelése

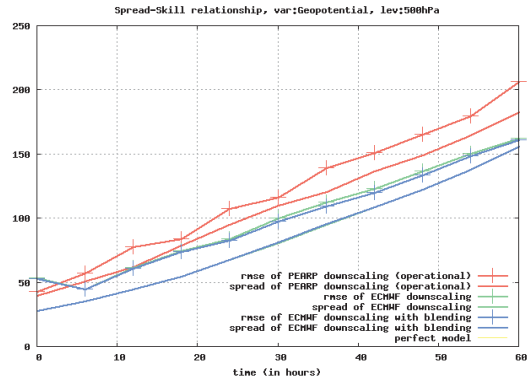
- Talagrand diagram
 - Az n db előrejelzés n+1 db részre (dobozra) osztja az előrejelzési tartományt.
 - Minden megfigyelés valamelyik dobozba esik.
 - A kellően nagyszámú előrejelzés-megfigyelés pár esetén megjelenített doboz elemszám diagramot nevezzük Talagrand diagramnak.



Ensemble előrejelzések értékelése

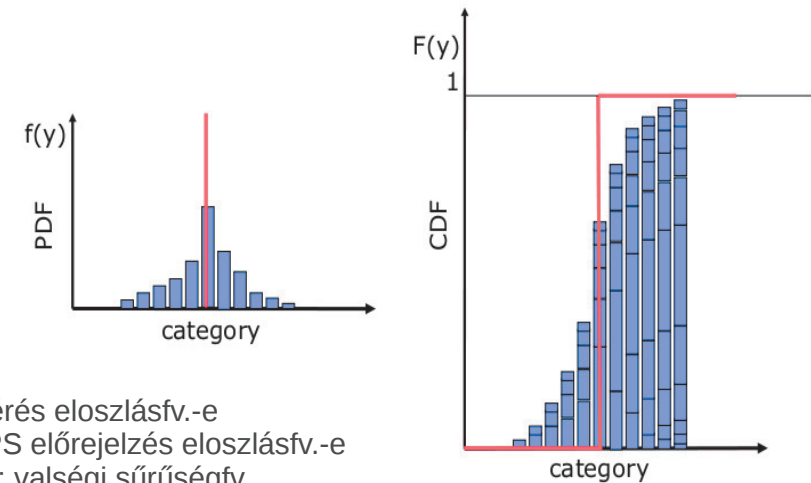
• Spread-Skill kapcsolat

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



Ensemble előrejelzések értékelése

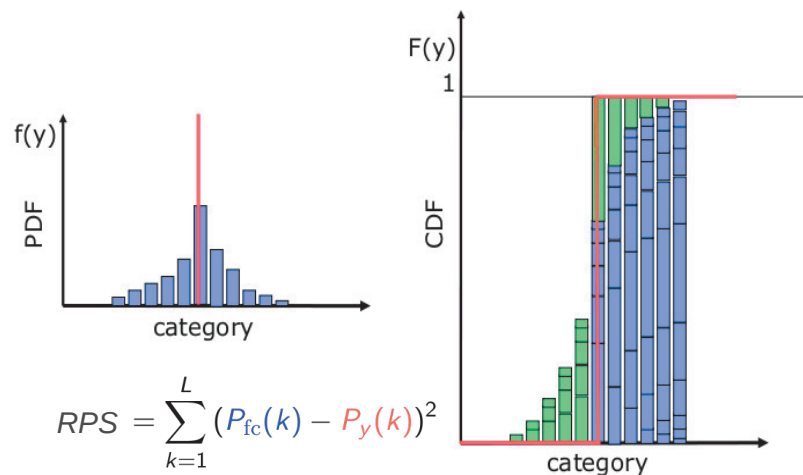
Ranked Probability Score (RPS)



- mérés eloszlásfv.-e
- EPS előrejelzés eloszlásfv.-e
- PDF: valószínűségi sűrűségfüggvény
- CDF: valószínűségi eloszlásfüggvény

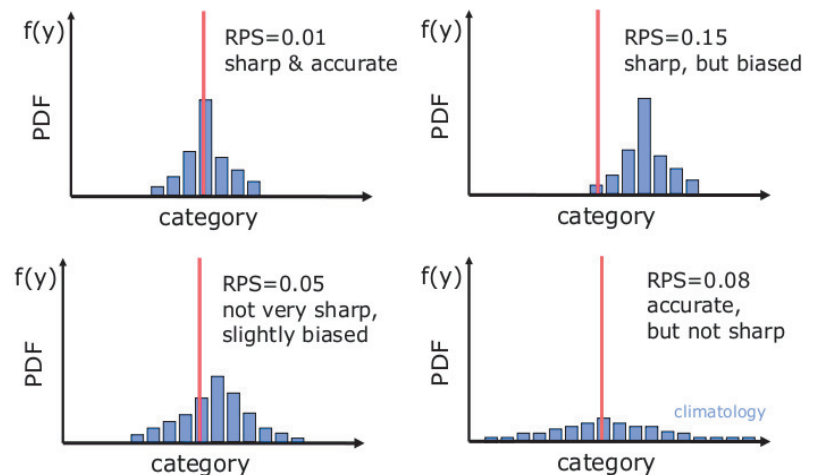
Ensemble előrejelzések értékelése

Ranked Probability Score (RPS)



L: küszöbértékekkel definiált kategóriák

Ensemble előrejelzések értékelése



Ensemble előrejelzések értékelése

• Ranked Probability Skill Score

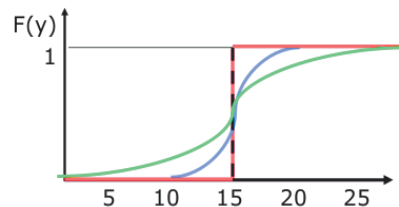
- A modell és a klíma előrejelzés RPS hányadosából képezzük
- Mindig ezzel a képlettel származtatunk 'skill score'-t a 'score'-okból.

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

• Continuous Ranked Probability Score

- Integráljuk az előrejelzett és a mért eloszlás függvény különbségének a négyzetét.
- Minél kisebb az értéke, annál jobb

$$CRPS = \int_{-\infty}^{+\infty} (F_{FC}(x) - F_{OBS}(x))^2 dx$$



Módszerek alkalmazása két rendszerben

• Két európai előrejelző központ globális modelljeiben:

• ECMWF (európai) - ENS

- 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 (amit nem perturbáltak)
- SPPT-vel veszik figyelembe a modellhibát

• Meteo France (francia) - PEARP

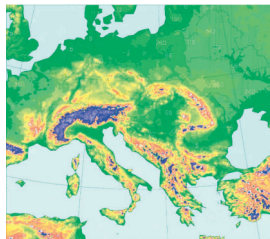
- 50 tagú 4DVAR EDA ciklus (6h)
- Ebből 35 tagot használnak, SV-okkal kombinálva => 35 tagú rövidtávú EPS
- Kontrollt mindig a legjobbnak ítélt analízissel futtatják, a többi 10 különböző fizikai csomaggal.

Módszerek alkalmazása két rendszerben

• OMSZ korlátos tartományú modelljeiben:

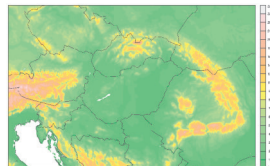
• ALADIN-EPS (ALARO fizikai parametrizáció)

- 10+1 tagú ensemble rendszer
- Korábbi operatív futás, 18UTC-ből indulva +60 órára, 8km felbontással
- Kísérletek EDA és SPPT módszerek alkalmazására
- ECMWF ENS csatolás



• AROME-EPS („convection-permitting EPS”)

- 10+1 tagú ensemble rendszer
- Operatív futás 2020 óta. Napi 1x, 0 UTC +48h
- ECMWF ENS csatolás
- 2.5 km felbontású nem-hidrosztatikus rendszer, melyben finom struktúrák (konvektív események) is megjelennek, melyek előrejelezhetősége kisebb
- Kísérletek EDA és SPPT módszerek alkalmazására



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.: Ensemble Forecasting at NCEP and the Breeding Method 1997, Mon. Weather Rev. 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.

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

- (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 λ_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

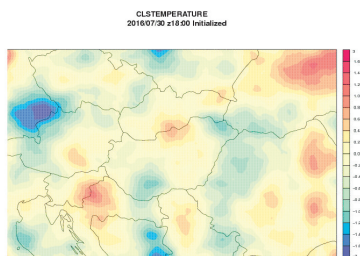
Valamint v_i -k egyben (M^*M) sajátvektorai, méghozzá λ_i^2 sajátértékkel

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

SPPT – Korlátos tartományú modellben

r mintázat optimális hangolása kihívásokkal teli korlátos tartományon. Fontos, hogy a pontbeli értékek szórása, tér- és időbeli korrelációja megfelelően működjön.

r AROME tartományon történő fejlődése
3 órán keresztül



r mező térbeli metszetének fejlődése

