

Gépi tanulós módszerek alkalmazása a modellezésben, előrejelzésben

Szépszó Gabriella

Klimatológiai és Kutatás-fejlesztési Igazgatóság



A (fizikai) modellfejlesztés folyamata



Célkitűzés:

Fejlesztési cél kitűzése (pl. felbontás finomítása)

Előkészítés:

Modellkód és programok módosítása, előkészítése, tesztkísérletek megtervezése, beállítások áttekintése, szükséges adatok gyűjtése, futási idő felmérése

Kísérletek:

Rövidebb és hosszabb tesztstimulációk elvégzése (általában 1-hónapos időszak 2 évszaktól)

Értékelés:

Verifikáció és esettanulmányok (verifikációs mutatók, térképes áttekintés)

Felhasználói teszt:

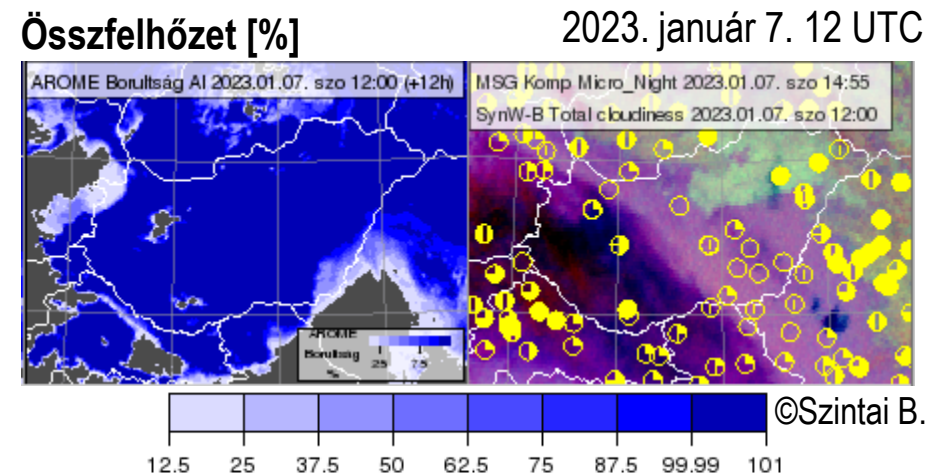
Az előrejelzők, felhasználók legalább 1 hónapig tesztelik a fejlesztést a napi feladataik elvégzésekor, „kvázi-operatív” futtatás

Operatív:

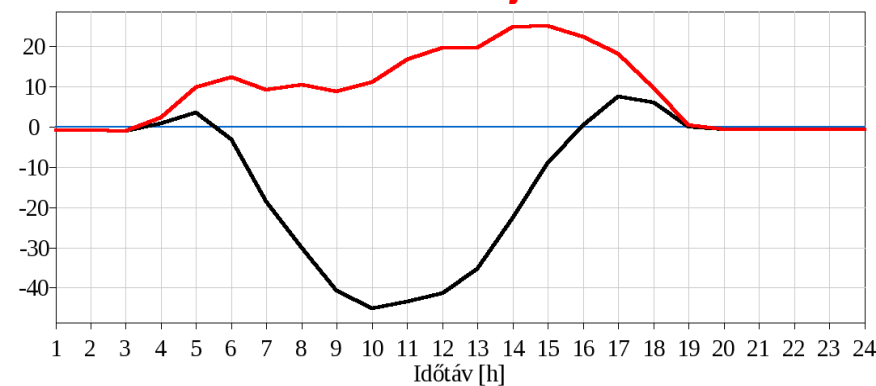
Bevezetés az operatív gyakorlatba, felügyelet

Valós fejlesztési példa 1.

- Előrejelzői visszajelzés, **2023. január**: alacsonyszintű felhőzet és nedvesség felülbecslése AROME-ban
- Első gyanú: adatasszimiláció
- **Új felhőséma**: a telítési nedvesség szórása a turbulencia séma alapján prognosztikusan (létező, bekódolt módszer kipróbálása)
- Kísérletek nyári és téli hónapokon, verifikáció, esettanulmányok
- Előrejelzői teszt visszajelzésekkel, legalább 2 hónap, évszakfüggő
- Operatív **2023 novemberétől**

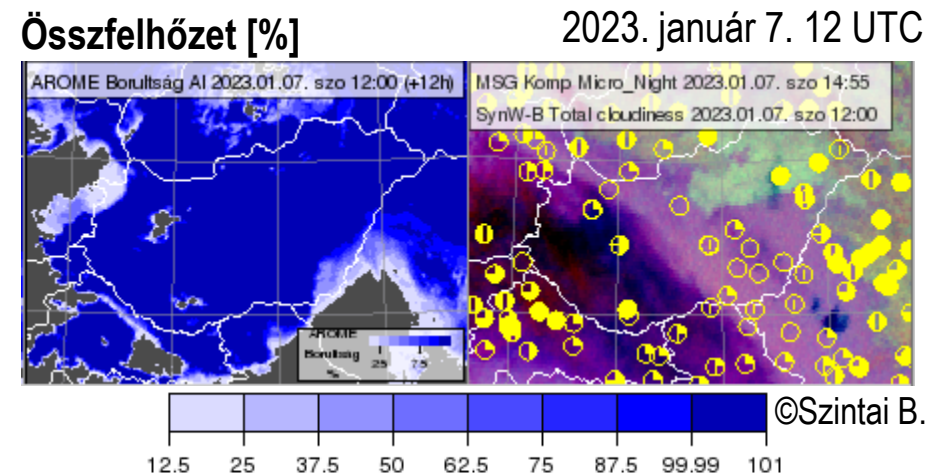


Sugárzás átlagos hibája [W/m^2]
AROME korábbi és új felhősémával

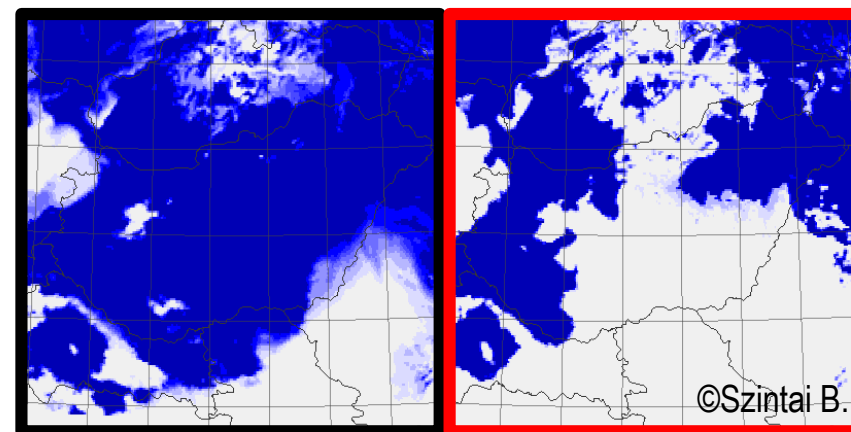


Valós fejlesztési példa 1.

- Előrejelzői visszajelzés, **2023. január**: alacsonyszintű felhőzet és nedvesség felülbecslése AROME-ban
- Első gyanú: adatasszimiláció
- **Új felhőséma**: a telítési nedvesség szórása a turbulencia séma alapján prognosztikusan (létező, bekódolt módszer kipróbálása)
- Kísérletek nyári és téli hónapokon, verifikáció, esettanulmányok
- Előrejelzői teszt visszajelzésekkel, legalább 2 hónap, évszakfüggő
- Operatív **2023 novemberétől**

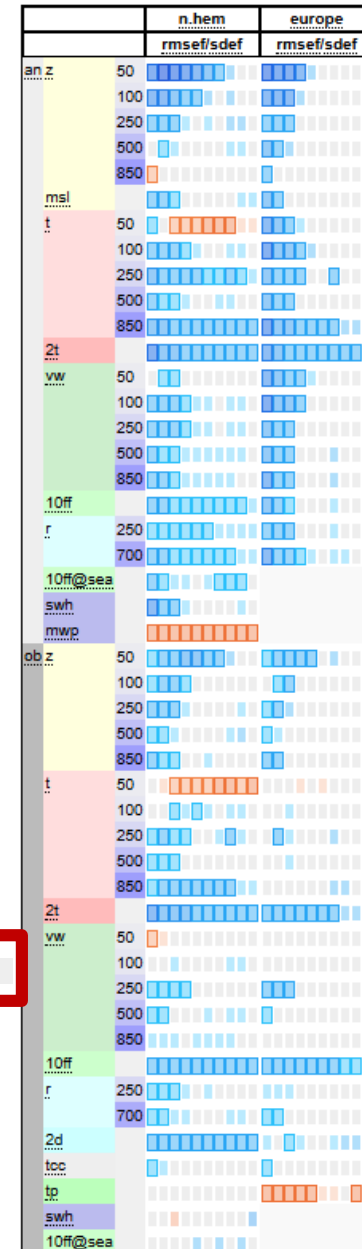


Alacsonyszintű felhőzet [%], 2023. január 7. 12 UTC



Valós fejlesztési példa 2.

- ECMWF tagországi visszajelzések: 2-méteres hőmérséklet előrejelzések hibái
- Globális középtávú modell → régióként változhat, hogy mi a hiba oka
- **Fejlesztési területek kijelölése (2020):** az asszimilációban (pl. új mérés), a felszíni folyamatok leírásában (pl. új hóséma) → a kutatók által kitalálandó, bekódolandó új modellrészek
- Sok-sok fejlesztői kísérlet
- Kísérletek legalább nyári és téli évszakon, verifikáció, esettanulmányok
- Összefésülés a többi fejlesztéssel, együttes hatás vizsgálata
- Tagországi teszt visszajelzésekkel, legalább 3 hónap
- **Operatív kb. 3-4 év múlva (2024)**

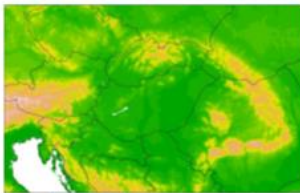


Az előrejelzések felhasználói

- Operatív előrejelző szakember (pl. repülésmeteorológus, veszélyjelző), igényei a modell előrejelzéssel szemben:
 - A légköri folyamatok 3-dimenziós leírása (pl. nyomási szinteken)
 - Fizikai konzisztencia és a fizikai folyamatok nyomon követhetősége (még az ensemble előrejelzés egyes tagjai esetében is)
 - Általánosan megbízható előrejelzés, ismert hibákkal
 - Napon belül többszöri frissítés, egyes esetekben valós idejű felhasználás
- Korlátos tartományú előrejelző modell (egyéb dinamikus modell), igények:
 - 3-dimenziós leírás a teljes modellrácsra (pl. 90 szinten)
 - Fizikai konzisztencia
- Produktum használó (pl. energia szektor), igények:
 - Nem a teljes 3-dimenziós előrejelzést használja, csak néhány meteorológiai változót, pl. globális sugárzást
 - Nagy pontosság
 - Gyakori frissítés, valós idejű felhasználás



Regionális modell (LAM)



Operatív előrejelzések ütemezése (UTC-ben)

Határfeltételek

0:00–0:20

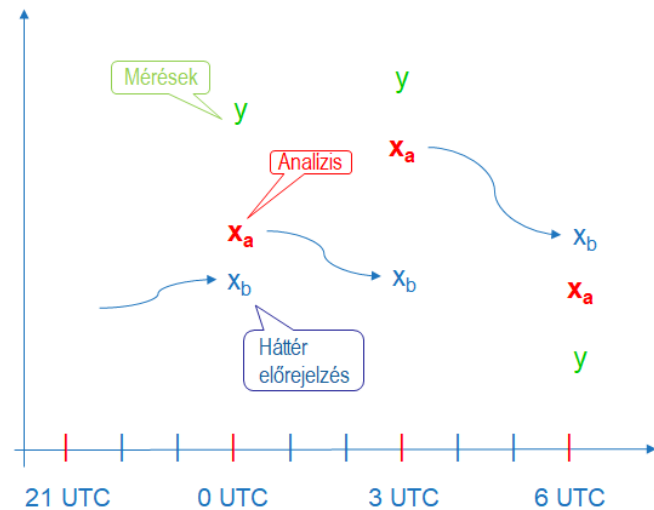
(Korábbi) globális előrejelzésből készült kivágot letöltése



Adatasszimiláció

AROME: 1:30–1:35

Mérések gyűjtése, ellenőrzése, objektív analízis → kezdeti feltétel előállítása

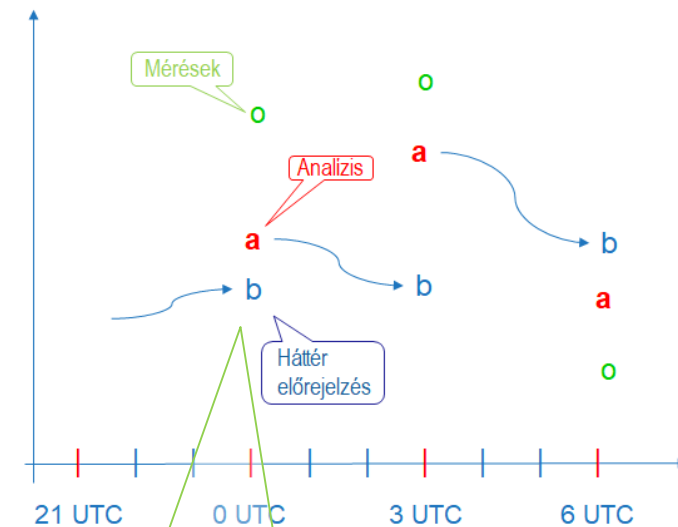


Kezdeti feltétel előállítása

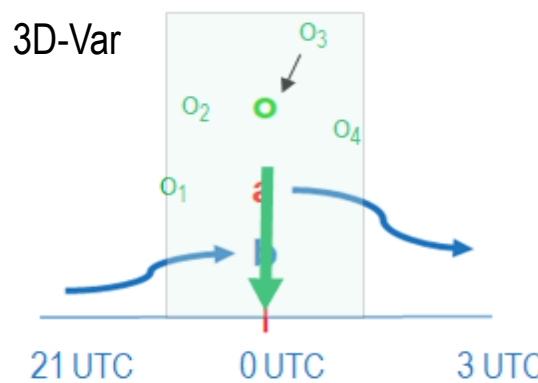
- Cél: a rendelkezésre álló információk felhasználásával a kezdeti állapot optimális becslése a modell rácshálózatán
- Felhasznált információk:
 - Mérések, megfigyelések: globális telekommunikációs rendszerből (GTS) és lokális adatcsere révén
 - Néhány-órás előrejelzések (háttér)
 - A mérési és háttér hibák becslései
- Adatasszimilációs ciklus: az AROME esetében 3 óránként, ECMWF esetében 6 óránként
- A lehető legtöbb mérés felhasználása → időablak + várakozás az adatokra → az előrejelzés késleltetve indul (AROME: +90 perc – 3D-Var, ECMWF: +3+2 óra – 4D-Var)



Adatasszimilációs ciklus



3D-Var



Operatív előrejelzések ütemezése (UTC-ben)

Határfeltételek

0:00–0:20

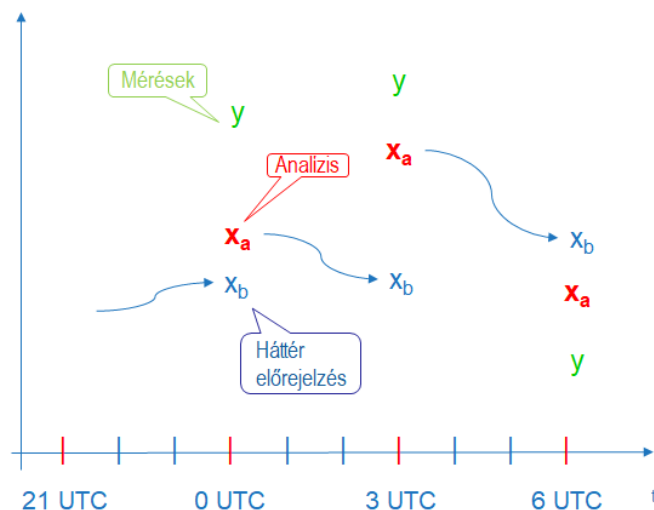
(Korábbi) globális előrejelzésből készült kivágat letöltése



Adatasszimiláció

AROME: 1:30–1:35

Mérések gyűjtése, ellenőrzése, objektív analízis → kezdeti feltétel előállítása



Modellintegrálás

AROME: 1:35–2:10

Az egyenletrendszer közelítő megoldása a modell saját vetületén és rácsán, nyers modelloutputok előállítása 15-percenként/óránként

Utó-feldolgozás

AROME: 1:44–1:57

A nyers modelloutputokból speciális paraméterek számítása a kívánt területre, szintre, felbontással

Termékek továbbítása felhasználóknak

2:00–

Gépi tanulós módszerek alkalmazása

Példák

- Felszíni jellemzők leírása
- A modellezés egyes részein:
 - Adatasszimiláció: nem-lineáris megfigyelési operátor előállítása
 - Parametrizáció: paraméter becslés, emulátor
- Nyers modelleredmények utó-feldolgozása
- Előrejelzés:
 - Nowcasting
 - Rövid- vagy középtávú előrejelzés
 - Ensemble tagok generálása

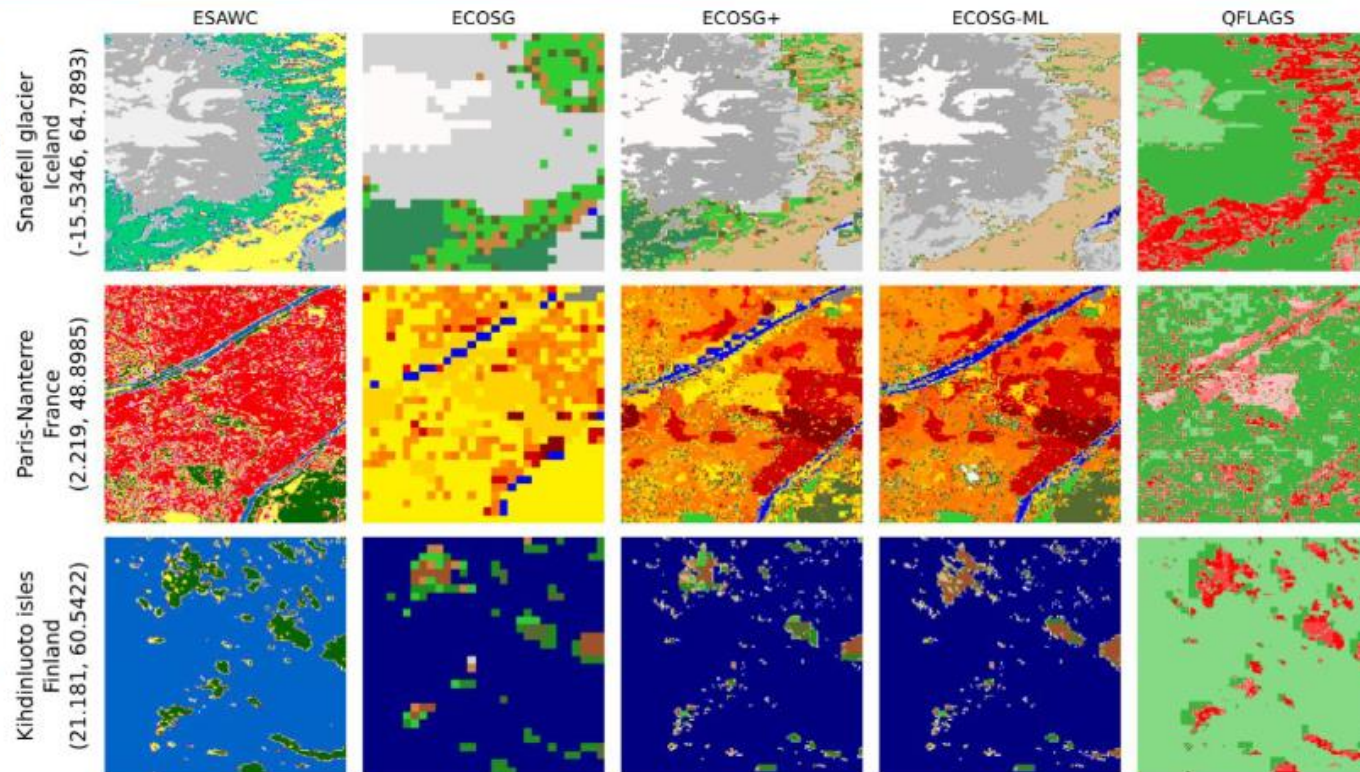
Felszíni jellemzők leírása

Next generation of physiography data is based on Machine Learning

Here, physiography generated by these two methods are compared to ESA CCI landcover Word Cover map (ESAWC) and ECOSG, accompanied by quality flags (QFLAGS).

Links to documentation:
Geoffrey Bessardon et al. (2024),
doi: [10.20944/preprints202409.0953.v1](https://doi.org/10.20944/preprints202409.0953.v1)
Thomas Rieutord et al. (2024),
doi: [10.20944/preprints202409.0942.v1](https://doi.org/10.20944/preprints202409.0942.v1)

Work now continues at FMI by
Rudolf Mård and Panu Maalampi.



Kiindulás:

- 300 m felbontású ECOCLIMAP-SG
- 10 m felbontású ESA WorldCover
- 10-100 m felbontású CORINE stb.

Módszer:
auto-encoder

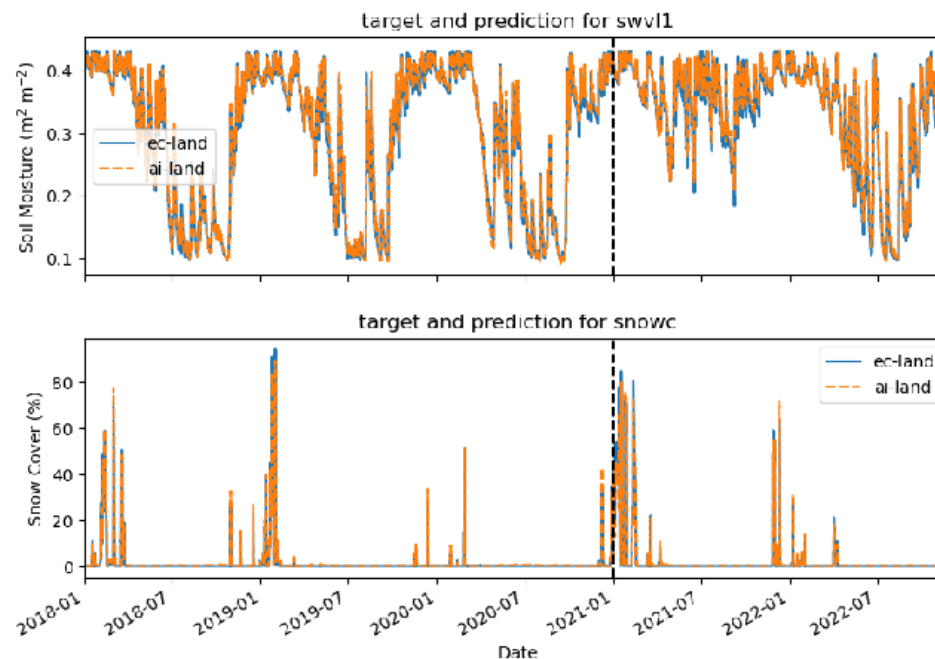
Eredmény:
60 m felbontású
ECOSG-ML

Adatasszimiláció belüli alkalmazások

Emulating the land surface for Data Assimilation



- Developing ML emulator of land surface model (ECLand), providing Jacobian and ensemble information at much reduced cost
 - 500 milliseconds vs 20 mins for 1-month Global run 🚀
- Next steps to test this in “offline” LDAS system for generation of Jacobians
- Emulator more stable than expected, no sign of divergence even after multiple years



Ewan Pinnington et al.

Analízis egyenlet:

$$\mathbf{x}_a^i = \mathbf{x}_b^i + \mathbf{K}(y^i - H(\mathbf{x}_b^i))$$

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

Felszíni asszimilációban:

$$\mathbf{H} = \begin{matrix} \frac{\partial T2M}{\partial TG1} & \frac{\partial T2M}{\partial TG2} & \frac{\partial T2M}{\partial WG1} & \frac{\partial T2M}{\partial WG2} \\ \frac{\partial HU2M}{\partial TG1} & \frac{\partial HU2M}{\partial TG2} & \frac{\partial HU2M}{\partial WG1} & \frac{\partial HU2M}{\partial WG2} \end{matrix}$$

Kiszámítása a modellezés során: a felszíni modell (pl. SURFEX) futtatása a kontroll változó kis perturbációval

Adatasszimilációon belüli alkalmazások



Funded by the European Union

Slide by Jostein Blyverket (MetNorway)

Machine learning observation operator



- Can we develop a reliable observation operator for AMSR-2 observations using machine learning?
 - Tb18 and Tb36 vertical polarization (V-pol) GHz (usually used in snow retrievals)
- Training dataset: September 2020 - May 2022, from offline SURFEX
- Three different approaches:
 - XG boost: Decision trees, no spatial context
 - **Footprint Convolutional Neural Network (CNN), learn spatial features**
 - Residual U-Net, CNN developed for the whole domain, learn spatial features at different spatial scales, domain dependent

AMSR-2: műhold adat, mikrohullámú

Származtatható: integrált vízgőz és felhővíz, csapadék, SST, SSu, **hóvastagság, tengerjég, talajnedvesség**

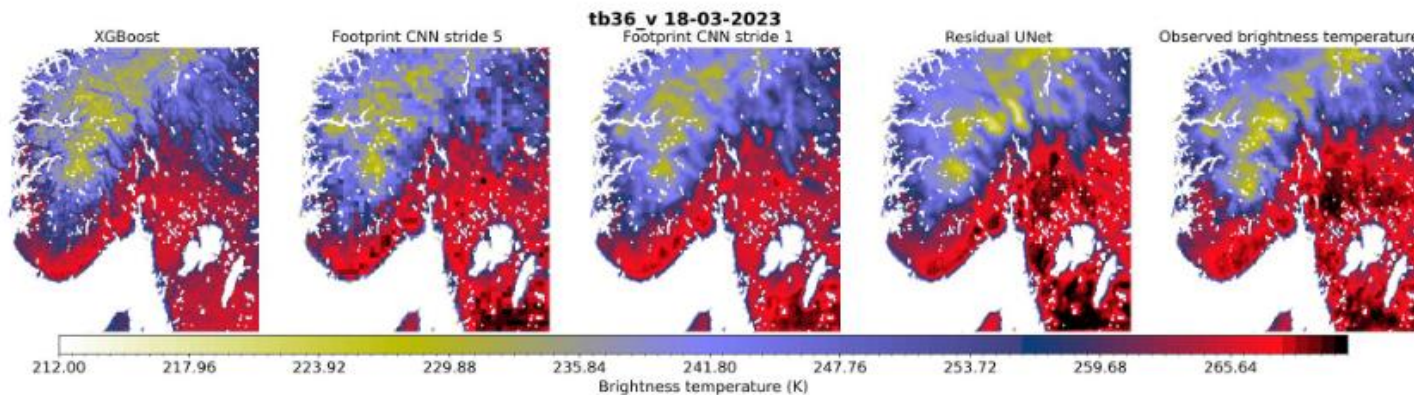


Figure xx: Example machine learning prediction of AMSR-2 Tb36 V-pol, from left, XGboost, Footprint CNN stride 5, Footprint CNN stride 1, residual U-Net, and observed brightness temperature from AMSR-2. For 18 March 2023.

Courtesy of Cyril Palermé

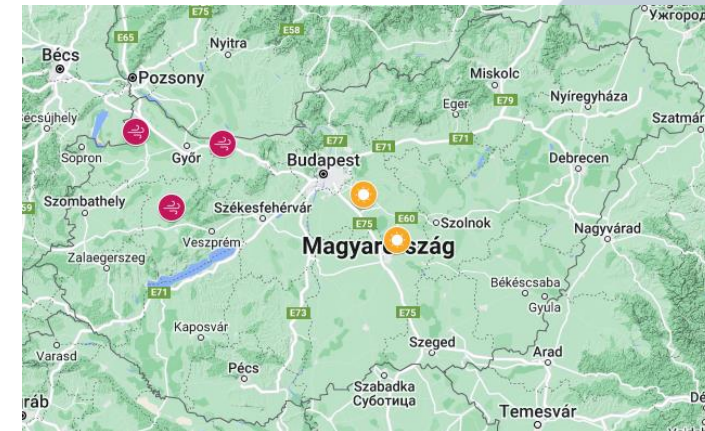
CopERNicus climate change Service Evolution -

CERISE

Nonprofit Zrt.

Nyers előrejelzések utó-feldolgozása

- Közös munka matematikusokkal: **AROME** (Korniyik Miklós) és AROME-EPS (Baran & Baran) eredményeinek utó-feldolgozása
- Célparaméter: **100-méteres szélesség, globálsugárzás**
- Felhasznált információk (szél + hőmérséklet + nyomás, sugárzás + hőmérséklet):
 - 15-percenkénti partneri mérések 3/2 pontra
 - 15-percenkénti AROME előrejelzések 36/48 órára (környezet)
- Módszer: **4/3-rétegű perceptron** (neurális háló), időtávonként
- Optimalizálás **négyzetes hibára**
- **8-** és **12-**hónapos időszak sorsolt 80/20 %-án tanítás/validáció



RMSE változás [%]			
	1. pont	2. pont	3. pont
tanítás	9	16	12
validáció	7	14	9

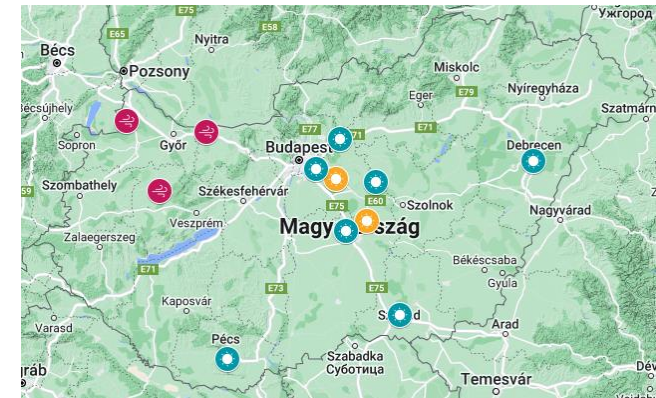
RMSE változás [%]		
	1. pont	2. pont
tanítás	21	11
validáció	21	10



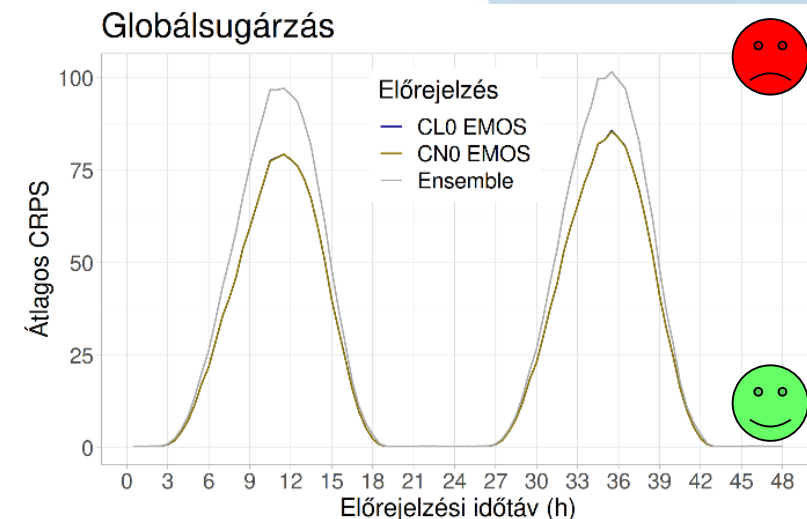
AROME-EPS utó-feldolgozás

Baran Ágnes, Baran Sándor, Debreceni Egyetem
Jávorné Radnóczy Katalin, HungaroMet

- Paraméteres **EMOS** utó-feldolgozás → eloszlás paramétereinek becslése
→ javított előrejelzés eloszlása
- Vizsgált **eloszlások**:
 - TN/LN: csonkított normális/lognormális + TN-DRN/LN-DRN (paraméterek becslése gépi tanulással)
 - CL0/CN0 – nullában alulról cenzorált logisztikus/normális
- Input adatok:
 - AROME-EPS előrejelzések 48 órára 15 percenként
 - HungaroMet sugárzás mérések** 7 pontra, **partneri szélmérések** 3 pontra
- **51/31**-napos **gördülő tanulóidőszak**
- Paraméterbecslés regionálisan és lokálisan
- Optimalizálás **CRPS-re**
- Hatás: 9-17 % CRPS javulás



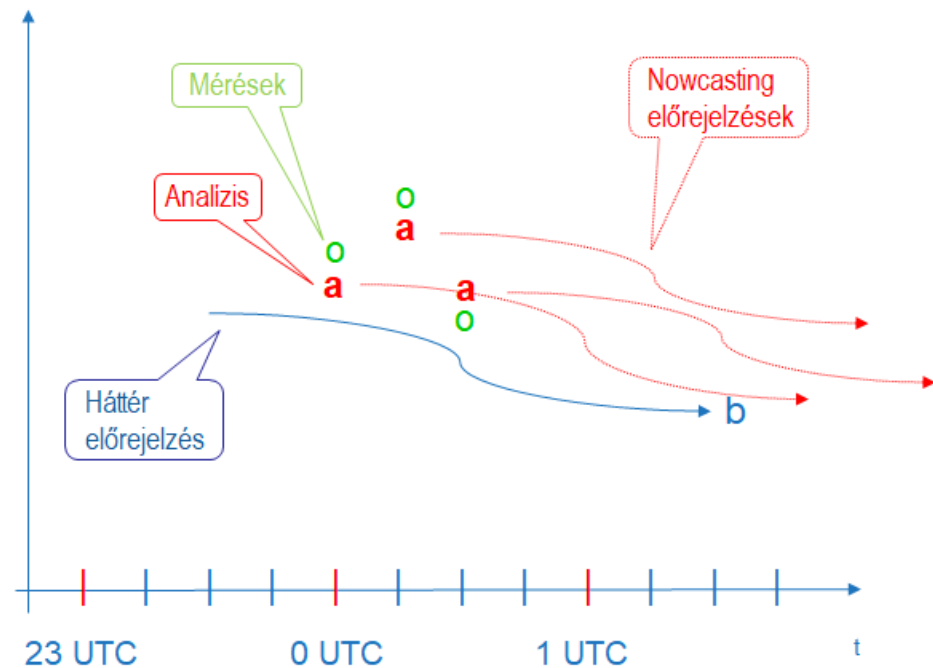
$$CRPS = \int_{-\infty}^{\infty} (F_{modell} - F_{mérés})^2 dx$$



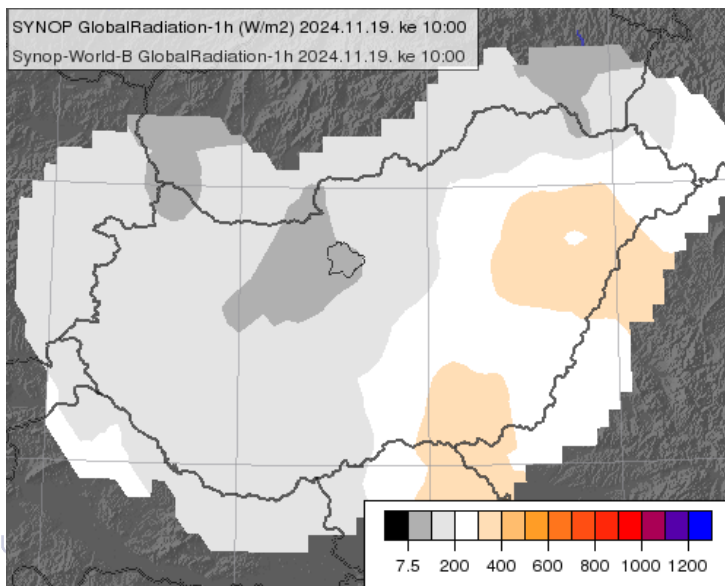
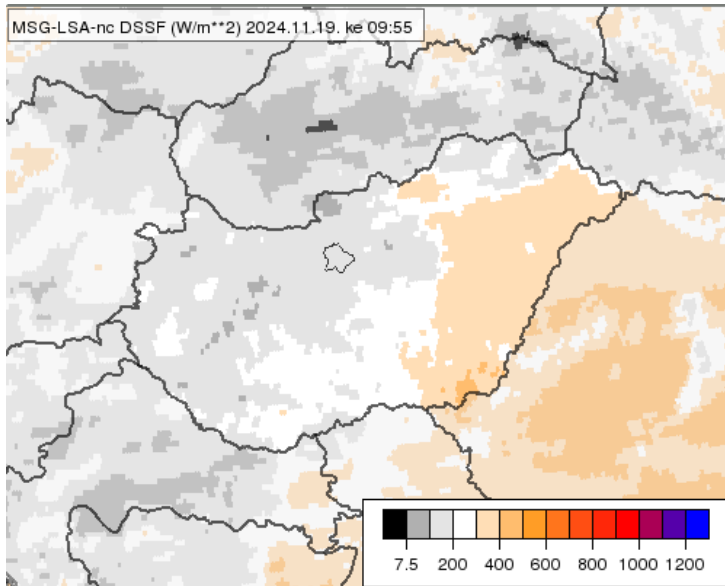
További információ:
Szépszó et al., 2023

Ultra-rövidtávú előrejelzések (nowcasting)

- Gyakori analízisből **1-3 órára szóló** ultra-rövidtávú előrejelzések
- Az analízishez a háttér (modell) előrejelzést ötvözzük mérésekkel, szempontok utóbbi kiválasztásában:
 - Nincs hosszú várakozás a mérésekre, gyors elérésű adatokat használnak
 - Az alkalmazott mérések közvetlen információt adnak a légköri eseményekről
 - A mérések is „rácson” állnak rendelkezésre
- Kézenfekvően felhasználható méréstípusok: radar (reflektivitás, szél), műhold (sugárzás, származtatott mennyiségek)
- Az analízisből egy **egyszerűsített és hatékony módszerrel** objektum „áthelyeződést” számítanak → **valós idejű felhasználás**

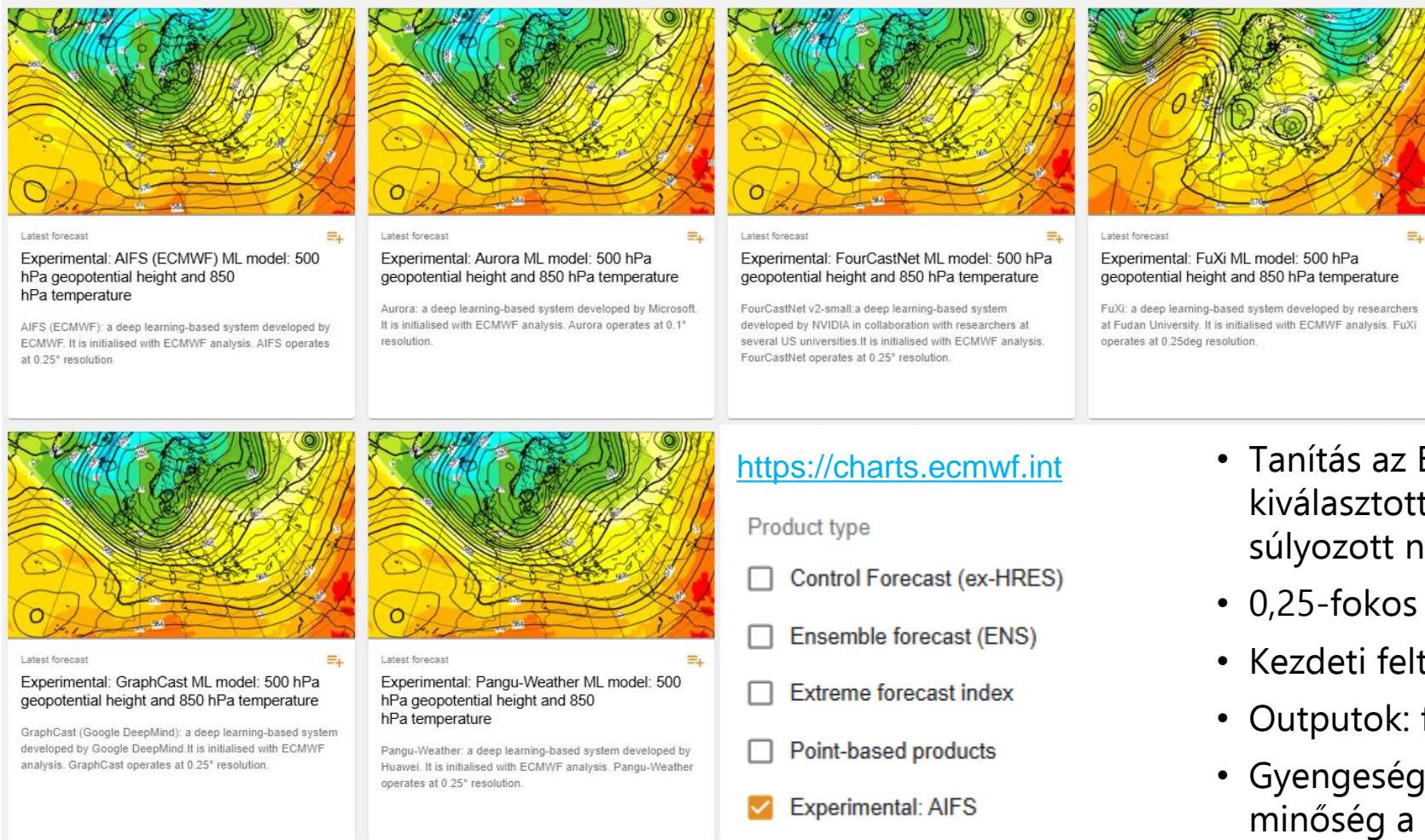


Sugárzás nowcasting – HungaroMet



- A naperőművek termelését befolyásolja az időjárási helyzet
- Az üzemeltetők számára kötelező a negyedóra bontott termelésbecslés, ami a nap során többször frissíthető (pontosítható)
- A termelés túlbecslése esetén kiegyenlítő energia → megvásárlását ráterhelik a pontatlan becslést adókra
- Intraday menetrend pontosításához közel valószerű becslések szükségesek
- Lehetséges kiindulási adat: műholdas sugárzásadatokból számított globálsugárzás produktum 15-percenként
- (Állomási mérésekkel összevetve: felülbecslés, vastagabb felhőzet esetén nagyobb hiba)
- Neurális hálón alapuló módszer, mely felismeri a térbeli mintázatokat és időbeli összefüggéseket (pl. a felhőzet esetében)
- Tanítási adatsor → 1-3 órás előrejelzés

Középtávú előrejelzés



<https://charts.ecmwf.int>

Product type

- Control Forecast (ex-HRES)
- Ensemble forecast (ENS)
- Extreme forecast index
- Point-based products
- Experimental: AIFS
- Experimental: Machine learning models

Az ECMWF-nél elérhető 6 gépi tanulós módszer operatív előrejelzése néhány meteorológiai paraméterre, de a kód is letölthető

Az **AIFS** perturbálatlan és ensemble előrejelzéseit a tagországok operatíván megkaphatják

- Tanítás az ERA5 adatbázison, optimalizáció kiválasztott mérőszámokra: AIFS – területtel súlyozott négyzetes hiba
- 0,25-fokos térbeli és 6/12-órás időbeli felbontás
- Kezdeti feltétel: ECMWF analízis
- Outputok: felszínközeli és nyomásszint változók
- Gyengeségek: sima előrejelzés, gyengébb minőség a sztratoszférában, extrém események, ensemble és hosszabb távú előrejelzések

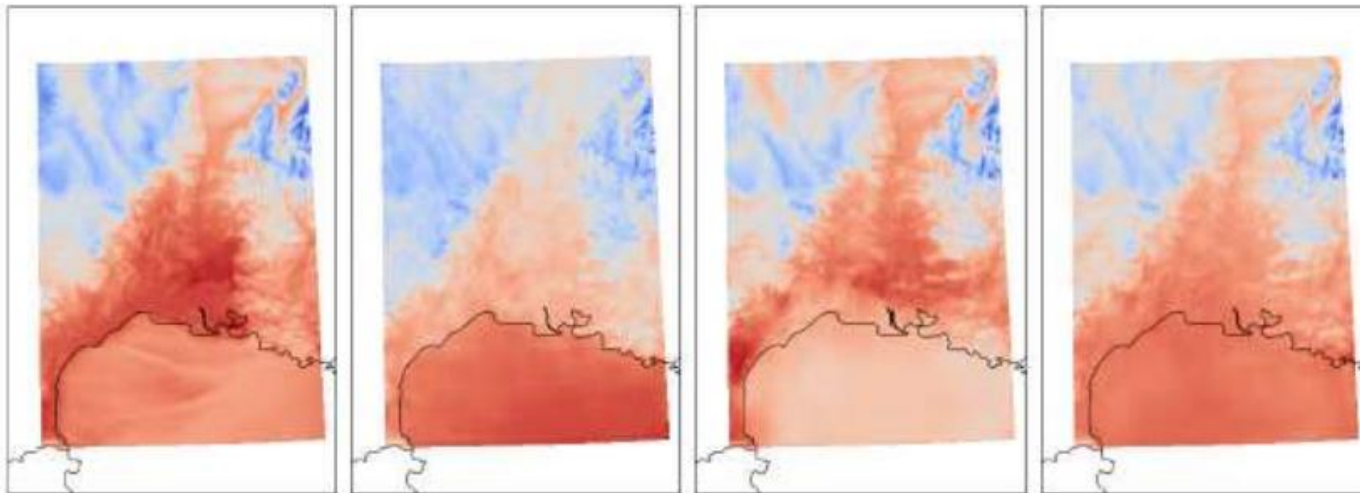
Ensemble tagok generálása

Toward next-generation Arome-EPS - Combining physics and AI

Toward big ensembles with AI-based forecast generation

- ▷ Increase the NWP sampling from $O(10)$ to $O(1000)$ thanks to complementary AI-generated forecasts
- ▷ Leverage the power of **deep generative models** such as **Generative Adversarial Networks (GANs)** to create realistic forecasts.

- Kiindulás:
a Météo-France
16-tagú, 1,3 km-es
felbontású AROME-EPS
rendszere, EDA-val,
SPPT-vel



True or fake ?

Összefoglalás

- Fizikai modellfejlesztés területei:
 - Felbontás növelése: km-es/hektometrikus skálájú globális/regionális előrejelzések
 - Módszertani fejlesztések, részfolyamatok részletesebb leírása, új folyamatok figyelembevétele
- Mesterséges intelligencia használata:
 - Az előrejelzés folyamatában (pl. asszimiláció, parametrizáció, ensemble előrejelzések)
 - Nyers modelleredmények utó-feldolgozásában
 - Nowcastingban
 - **Előrejelzés esetén alapvető hozzávaló a jó fizikai modell**
 - Kérdések: operatív használat és fejlesztés, kutatás viszonya, fizikai megértés