

Doktori beszámoló

3.félév

Márkuszt István

(istvan.markusz@gmail.com)

Hálózatok magas dimenziós modellezése

Témavezető: Dr. Pollner Péter

ELTE TTK Fizika Doktori Iskola

Statisztikus fizika, biológiai fizika
és kvantumrendszerek fizikája program

1. Bevezetés

A félév jelentős részében az Afrikai Sertéspestis (ASP) modell paramétereinek illesztésével voltam elfoglalva. A modell komplexitásából eredően nagy számítási kapacitást igényel. A jelenlegi eredményeket a februári NetSciX konferencián fogom bemutatni, illetve a cikk írása is folyamatban van.

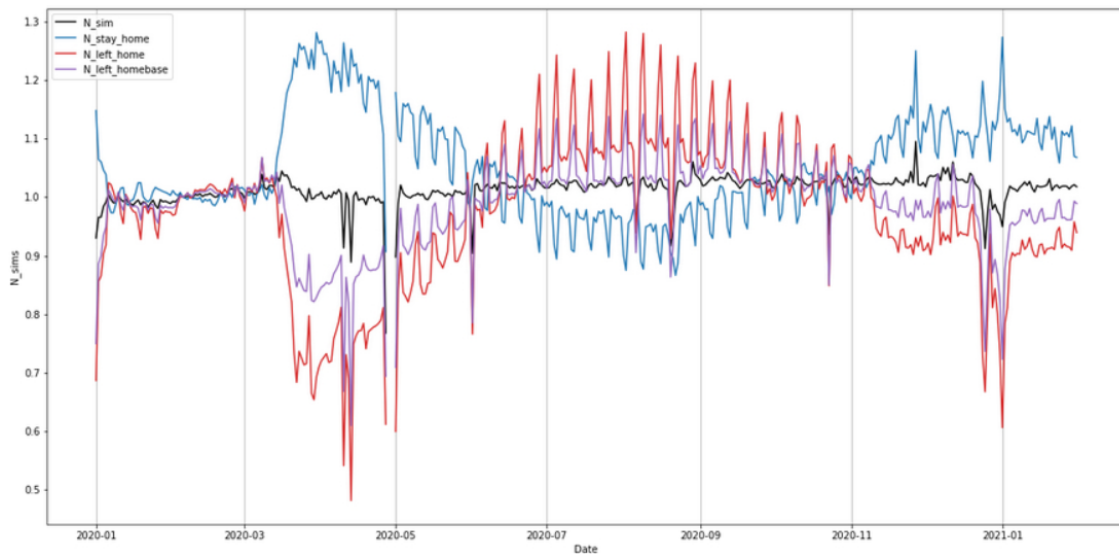
Ezen felül az előző beszámolóban említett, Nemzeti Közszolgálati Egyetem szociológusaival folytatott exploratív analízisből cikket írtunk.

A COVID-19 Magyarországi terjedésének reprodukálásához hozzáférést kaptam szennyvíztelepekben mért víruskoncentráció adatokhoz. Az adatok a mérés módszertanából eredően meglehetősen jelentős szórással rendelkeznek, de mégis sikerült aggregált formában korrelációt találni a szennyvíz koncentrációk és a járvány hivatalosan rögzített számadatai között. Az adatok felhasználhatósági köre további analízist igényel.

2. Kutatás

2.1. Afrikai Sertéspestis Modell paramétereinek illesztése

A terjedési modell komplexitása, illetve a stochasztikus szimulációk miatt a valós adatokhoz való egzakt illesztés (például maximum likelihood módszerrel) matematikailag lehetetlen. Ennek áthidalására egy Markov-lánc Monte-Carlo módszert választottam, mely ráadásul adaptív lépésközzel dolgozik, és képes multimodális paraméter eloszláso-



1. ábra. Az otthon maradók, valamint az otthonukat, illetve otthon településüket elhagyók aránya egy átlagos 2019 őszi héthez képest. A KSH ingázási adatait ezen számokkal skálázva felhasználható a COVID-19 magyarországi terjedésének modellezésére.

kat is feltérképezni [1]. Ehhez sok párhuzamos szimuláció futtatása szükséges, így az Atlasz szuperszámítógép kézenfekvőnek tűnt erre a célra.

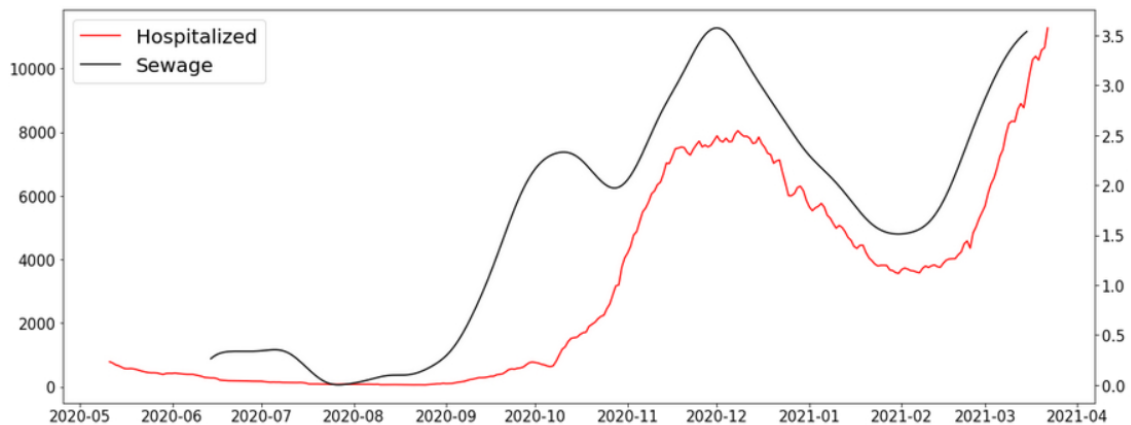
A szuperszámítógépen való futtatás több, nem várt technikai nehézségbe ütközött. Mivel a megoldások megtalálása körülményes, lévén hogy egy beküldött futtatásnak napokba telik elindulnia, és csak parancssoros felületen lehet hibát keresni.

Bár a paraméterek optimalizálása továbbra is hátra van, sokat tanultam a slurm illetve kubernetes rendszerekről, illetve az ezeken való futtatásról.

2.2. COVID-19 magyarországi terjedésének reprodukálása

A KSH korábbi ingázási adatait skáláztuk városonként az otthonukat elhagyók számával (1. ábra), és ez alapján szimuláljuk a járvány terjedését.

A COVID-19 lejelentett statisztikái helyett használjuk 24 magyar szennyvíztelepen kvantitatív PCR-rel mért víruskoncentrációs adatokat. Az adatpontokat tisztítva, a mérési eljárás technológiai részleteinek ismeretében először egy országos referencia-trendet állapítottunk meg. Az összes telep adatát átlagolva és 6 hetes csúszóátlagot nézve sikerült kinyerni egy görbét, amely viszonylag jó korrelációt mutat a lejelentett kórházi ápoltak számával (lásd 2. ábra), ezzel igazolva az adatok összefüggését a valósággal. A továbbiakban célkitűzésünk, hogy a telepkehez tartozó egyes területegységek átfertőzöttségét megbízhatóan próbáljuk becsülni ezekből a számokból.



2. ábra. A szennyvíztelepeken mért kvantitatív PCR adatok összevetve a kórházban ápoltak számával. A görbék viszonylag jó egyezést mutatnak, azonban ehhez az összes telephez tartozó város populációjával súlyozott átlagának további 6 hetes csúszóátlagolása volt szükséges.

2.3. Egyetemek kollaborációs hálójának és a QS rangsor kapcsolatának exploratív elemzése

A Nemzeti Közszerződési Egyetem szociológusai legfőképp az "államilag támogatott szervezetek" (ilyen például a CERN is) szerepére kíváncsiak az egyetemek kollaborációs hálójában. Bár az látszik, hogy ezekből jó pár kiterjedt kollaborációs körrel rendelkezik, a rangsorra gyakorolt befolyás nem volt kimutatható. Ezen eredmény után a kutatás inkább exploratív fázisba került, jelenleg a hálózat leíró statisztikáit vizsgálják.

3. Publikációk

- Comparison of imaging modalities used in follow up for left ventricular noncompaction patients: which is optimal? (elfogadásra vár, 3.szerző)
- Path of excellence: a co-authorship network analysis of European Research Council grant winners in social sciences (elfogadásra vár, 2. szerző)

4. Tanulmányi tevékenység

A félév során a PhD programomban meghirdetett alábbi három tárgyat végeztem el:

1. Extrémek, rekordok és sorrend statisztikák a természetben (FIZ/3/075E)
2. A gépi tanulás új eredményei szeminárium (FIZ/3/092)
3. Fejlődésbiológiai mechanizmusok kvantitatív modelljei (FIZ/3/056E)

5. Konferenciák, iskolák

A félév során rendszeresen részt vettem a *Deep learning szemináriumon* az ELTE-n, melyen többször elő is adtam.

Elfogadták a jelentkezésemet a NetSciX februári konferenciájára, melyen negyed órás prezentációt tartok majd a sertéspestis terjedésének modellezéséről.

6. Oktatási tevékenység

Tanársegédként részt vettem az *Adatbázisok* tárgyban, melyen egy konzultációt és két félév végi projektfeladatot vezettem.

Tanársegédként két téma vezetését vállaltam a *Data Science laboratory* tárgyban.

Továbbá segédkeztem a *Valószínűségszámítás* tárgy beadandóinak javításában

Hivatkozások

- [1] C. J. Ter Braak. A markov chain monte carlo version of the genetic algorithm differential evolution: easy bayesian computing for real parameter spaces. *Statistics and Computing*, 16(3):239–249, 2006.