

Idősor elemzés a rendszerfelügyeletben

Önálló laboratórium feladat összefoglalója (1. félév)

Gruber Richárd (V1WG4B)

Konzulens: Égel Zoltán

**BME Méréstechnika és Információs Rendszerek Tanszék
Informatikai infrastruktúra tervezése szakirány, 2008/2009. I. félév**

A félév elején megismerkedtünk néhány a témához kapcsolódó cikkel, mely alapján egy átfogó képet kaptam arról, hogy milyen távlati céljaink vannak a témával kapcsolatban.

Ezt követően megállapodtunk abban, hogy az idősorok elemzéséhez a Mathworks Matlab programjának R2008b verzióját fogjuk használni.

Megfelelő gyakorló idősorokat kerestünk a feladatunkhoz, így választásunk az angol folyók vízjárását tartalmazó idősorokra esett. Kiválasztottuk azon folyókat, amelyek valamilyen szempontból érdekesek lehetnek számunkra, így a vizsgált folyók a Temze(Thames), Severn, Kennet, Mersey, Mole, Coln.

Első lépésként meg kellett oldanom, hogy az idősorokat tartalmazó .csv fájlok kezelhetőek legyenek a Matlab alatt, így készítettem egy beolvasó rutint, amely egy-egy mátrixba olvasta be az idősorok tartalmát.

Következő lépésként elemzést kellett készítenem a folyókról, mely tartalmazza a legkisebb, legnagyobb vízhozamot, illetőleg az átlagos vízhozamot az egyes években. Ehhez rutinokat hoztam létre, melyek reményeim szerint a későbbi munkák során is felhasználhatóak lesznek. Ekkor szembesültem a problémával, hogy az ún. NaN (Not a Number) értékek egyrészt módosíthatják az eredményeket, másrészt lehetetlenné teszik az átlagolást. Ezt kétféleképpen oldhattuk volna meg, egyrészt egyszerűen figyelmen kívül hagyjuk ezeket az időszakokat, mikor NaN értékeink vannak, másrészt javító rutint írok rá, és megpróbálom megbecsülni, hogy mikor milyen értéket vesz fel az adott folyó idősora. Az utóbbi mellett döntöttünk, mivel ez a későbbiek során kicsit átalakítva felhasználható lesz az előrejelzésnél.

Öt alapvető rutint hoztam létre, melyek más és más módszeren alapultak. Az első a hiányzó érték környezeti átlagát vette alapul és helyettesítette be, a második hasonlóképpen következtette ki a hiányzó értéket, csak a közelebbi értékek nagyobb súllyal számítottak be a becsült értékbe. A harmadik módszer vette az adott napot minden egyes évben, és ezeknek vette az átlagát. A negyedik módszer a meredekség alapján következtette ki a behelyettesíthető értéket, míg az ötödik szintén a meredekség alapján, csak ez négyzetes átlagolással. Miután megnéztem, hogy az egyes módszerek milyen eredménnyel becsülik meg az adott hiányzó értékeket, arra jutottam, hogy a két legeredményesebb módszer egyértelműen a finomított környezeti átlagú (kettes) és a meredekséget alapul vevő (négyes) módszer. Megvizsgáltam, hogy amikor egyikük nagyobbat téved, abban az esetben a másik milyen eredményt ad, és arra jutottam, hogy kiegészíthetik egymást. A hirtelen emelkedések és esések esetében ugyanis a négyes módszer volt a pontosabb, míg az esetek nagyobb részében a finomított környezeti átlagot felhasználó. Így létrehoztam egy Hybrid nevű rutint a két módszer kombinálására, mely figyel, hogy mikor van hirtelen esés vagy emelkedés a vízhozamban, és ettől függően használja az egyes módszereket. Az eredmény egy sikeresebb becslő függvény lett.

A NaN javító rutinok létrehozása után következett az alapvető trendek megfigyelése a folyók viselkedésében, melyeket felhasználhatunk majd az előrejelzésnél, hiszen akkor értelemeszerűen nem állnak rendelkezésünkre a becsülendő adat időpontja utáni napok adatai.

Végül elkezdtem az előrejelző rutin megvalósítását, de ez még nem került befejezésre.