

Eötvös Loránd Tudományegyetem
Földrajz- és Földtudományi Intézet
Meteorológiai Tanszék

Légköri nukleációs helyzetek meteorológiai háttérének statisztikai vizsgálata

– DIPLOMAMUNKA –



Készítette:

Kelemen Tibor

Meteorológus mesterszak, előrejelző szakirány

Témavezetők:

Dr. Weidinger Tamás, ELTE TTK, Meteorológiai Tanszék

Dr. Matyasovszky István, ELTE TTK, Meteorológiai Tanszék

Dr. Salma Imre, ELTE TTK, Analitikai Kémiai Tanszék

Budapest, 2015

Tartalomjegyzék

1. Bevezetés és célkitűzések	3
2. Irodalmi háttér.....	5
2.1. A légköri nukleáció	5
2.1.1. A mérőrendszer működési elve.....	6
2.1.2. A nukleációs napok azonosítása	7
2.2. Meteorológiai adatok.....	9
2.3. Kulcsfontosságú légszennyezők.....	10
2.4. Statisztikai módszerek.....	10
2.4.1. Klaszteranalízis	11
2.4.2. Diszkriminancia analízis	14
2.4.3. Logisztikus regresszió.....	18
2.4.4. Kategóriás verifikáció	21
3. Eredmények és megvitatásuk.....	23
3.1. Időeltolás és keresztkorreláció	23
3.2. Logisztikus regresszió	32
3.3. Diszkriminancia analízis	38
3.4. Verifikáció.....	40
4. Összefoglalás	42
5. Irodalomjegyzék	44
6. Köszönetnyilvánítás.....	46

1. BEVEZETÉS ÉS CÉLKITŰZÉSEK

Az aeroszol a levegőben eloszlott szilárd és/vagy cseppfolyós részecskék kolloid rendszere. A részecskék átmérője néhány nanométertől egészen 100 μm -es nagyságig terjed; aerodinamikai átmérőjük alapján megkülönböztethetünk: durva ($d > 2,5 \mu\text{m}$) és finom részecskéket ($d < 2,5 \mu\text{m}$). Ezen kívül léteznek még ultrafinom (UF) részecskék is ($d < 100 \text{ nm}$), amelyek tömege az aeroszol részecskék össztömegének kevesebb, mint 1%-át teszi ki (Pakkanen et al., 2001), viszont szám-koncentrációjuk a legnagyobb. E részecskéket már nem aerodinamikai, hanem elektrosztatikai módszerekkel tanulmányozzuk, így méretüket sem az aerodinamikai, hanem az elektromos mozgékonyági átmérővel fejezzük ki (Salma et al., 2011). A részecskék méret-eloszlásában négy jellemző módust (csúcsot) különböztethetünk meg, az egyes módusok medián átmérője (d) alapján: durva ($d \sim 6\text{--}8 \mu\text{m}$), akkumulációs ($d \sim 0,4\text{--}0,5 \mu\text{m}$), Aitken- ($d \sim 0,1 \mu\text{m}$) és nukleációs módus ($d \sim 2\text{--}100 \text{ nm}$). A durva és finom részecskékről viszonylag sok ismerettel rendelkezünk, míg az UF részecskék vizsgálata ez idáig a háttérben maradt mérési és kísérleti nehézségek miatt. Az utóbbi évek kutatásai során kiderült, hogy az ultrafinom aeroszol részecskék szerepet játszanak az éghajlat alakulásában és egészségügyi kockázatot is jelenthetnek (IPCC, 2013; Ramanathan et al., 2001).

Az aeroszol részecskék egyik fontos éghajlati hatása a felhők keletkezésében nyilvánul meg. A felhőcseppek képződése során a nagyobb, 100–200 nm átmérőjű, főként nukleáció utáni növekedéssel kialakult (Merikanto et al., 2009) higroszkopikus részecskék felületén a vízgőz képes kondenzálódni, és felhőcsepp keletkezik. Az ilyen tulajdonságú részecskéket felhő-kondenzációs magvaknak (CCN, Cloud Condensation Nuclei) nevezzük. Az aeroszol részecskék számában, illetve higroszkopikus tulajdonságában bekövetkező lényeges változás módosíthatja a felhők mikrofizikai, illetve sugárzási jellemzőit (Chin et al., 2009). A nagyobb részecskeszám esetén a kialakuló felhőcseppeknek egyrészt nagyobb összfelületük lesz, és ezért jobban visszaverik a napsugárzást („fehérebbek”), másrészt az átlagos élettartamuk is hosszabb, tehát a sugárzás-visszaverő hatásukat hosszabb ideig fejtik ki. A kisebb cseppeket tartalmazó felhőkből ráadásul ritkábban hullik csapadék, mert a sok csepp a vízpára korlátozott mennyisége miatt nem tud kellően nagyra nőni ahhoz, hogy távozzon a levegőből. Az aeroszol részecskék tehát a felhőcseppek révén egyrészt befolyásolják a víz és a nyomanyagok körforgását, másrészt a sugárzás-visszaverés miatt „hűtik” a Földet. Ezzel

komolyan ellensúlyozzák az üvegházhatású gázok melegítő hatását, és hozzájárulnak az elérhető éghajlat fenntartásához és szabályozásához (*Borsós és Salma 2009*).

Az utóbbi évek epidemiológiai vizsgálatai kimutatták, hogy kapcsolat van a légköri ultrafinom részecskék koncentrációja és a légzőszervi megbetegedések között. Az ultrafinom részecskék többlet kockázatot jelentenek az emberi egészségre, mert kis méretük miatt keresztülhatolhatnak a légutak sejtmembránján, s bejuthatnak a vérkeringésbe, vagy a szaglóiidegen keresztül eljuthatnak a központi idegrendszerbe (*Pöschl 2005*), vagyis a szervezet bármely részébe elkerülhetnek. Az ultrafinom részecskék nagy koncentrációja ilyen módon különböző krónikus légúti megbetegedéseket okozhat, fokozhatja a tüdőgyulladás, illetve az oxidatív stressz kialakulásának az esélyét, de akár szív- és érrendszeri megbetegedéseket is előidézhet. (*Politis 2008; Oederdörster et al., 2005*).

A légköri ultrafinom aeroszol részecskék nem csak tisztán természetes és antropogén úton keletkeznek, hanem e két hatás kombinációjaként is. Fő természetes képződési folyamatuk a légköri nukleáció. Legjelentősebb antropogén forrásaik a magas-hőmérsékletű folyamatok, amelyek közvetlenül emittálják a részecskéket; így például a közúti közlekedés, a tüzelés és a háztartási fűtés, valamint a hulladékégetés (*Pöschl 2005*). Korábbi vizsgálatokból (*Németh és Salma 2014*) kiderült, hogy a légköri nukleáció nem ritka jelenség Budapesten, gyakorisága éves szinten 27%. Nukleációra viszonylag tiszta napokon, kis aeroszol koncentráció esetén számíthatunk. A folyamat során keletkező új aeroszol nagyságrendekkel megnöveli a részecskék számkoncentrációját a levegőben. A jelenség a nagyvárosokban egészségügyi és városklimatológiai problémákkal, kockázatokkal jár.

A légköri aeroszol részecskék keletkezési folyamata, és a meteorológiai állapot-határozók között kapcsolat van. Munkám fő célja, hogy megvizsgáljam a nukleációs helyzetek és a meteorológiai, illetve a légszennyezettségi állapot-határozók közötti kapcsolatrendszerét. Dolgozatom a korábbi BSc-s szakdolgozatom (*Kelemen 2013*) folytatása, melyben egy napi szintű adatbázis összeállítása után korrelációanalízis segítségével vizsgáltam a meteorológiai és levegőkémiai adatok és a nukleáció közötti kapcsolatrendszerét. Jelen kutatásban elsőként egy pontosabb, 10 perces felbontású adatbázist hoztam létre a korábbi, napi felbontású adatsorok helyett, mivel a napi szintű adatsorok nem jellemezték megfelelően a nukleáció néhány órás időtartamát. Vizsgálatom célja az volt, hogy különböző összetett statisztikai módszerek segítségével elkülönítsem azokat a paramétereket, amelyek a leginkább befolyásolják a nukleáció kialakulását.

2. IRODALMI HÁTTÉR

Elsőként bemutatom a légköri nukleáció folyamatát, a keletkező új aeroszol részecskék szám-koncentrációjának meghatározását, továbbá a nukleációs napok azonosításának elvét. Ezt követően sorra veszem az adatbázisban szereplő változókat, végül pedig leírom az adatsorok jellemzésére alkalmazott statisztikai módszereket.

2.1. A légköri nukleáció

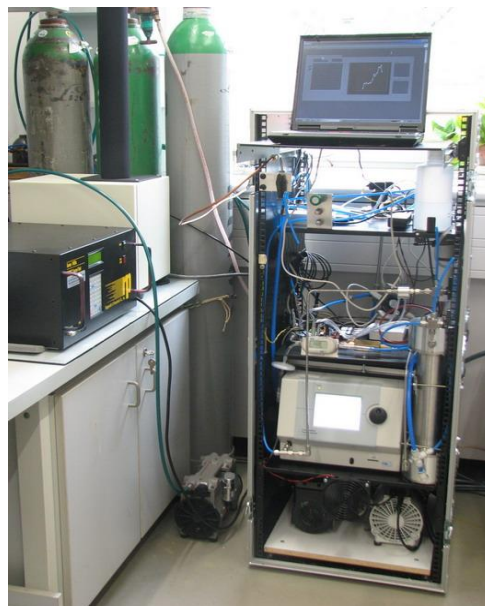
Légköri nukleációnak nevezzük azt az elsőrendű fázisátmenetet (légnemű-folyadék vagy légnemű-szilárd fázisváltás), amikor homogén összetevők (SO_2 , VOC: Volatile Organic Compounds, illékony szerves vegyületek) fotokémiai oxidációjával kisebb egyensúlyi gőznyomású, tehát kondenzációra képes gőzök (pl. H_2SO_4) keletkeznek, és ezek további légköri összetevőkkel (H_2O , NH_3 , NH_2R , NHR_1R_2 , VOC) kölcsönhatva szilárd vagy folyékony fázis megjelenését, vagyis új aeroszol részecskéket eredményeznek. A folyamat megvalósulásának kedvez, ha nagy az elővegyületek koncentrációja, erős a napsugárzás, illetve kicsi a rendelkezésre álló felület a versengő kondenzációs folyamat részére (Németh 2012). A nukleáció során, Budapesten másodpercenként 2–12 darab, új aeroszol részecske jön létre egy köbcentiméter levegőben, és a keletkezés folyamata akár egy óránál is tovább tarthat (Salma et al., 2011).

A szakirodalomban több légköri nukleációs elmélet létezik. A legegyszerűbb a homogén homomolekuláris nukleáció (egy anyagfajta esetén). Több gőz együttes nukleációja energetikailag már sokkal kedvezőbb, ezt homogén heteromolekuláris nukleációnak nevezzük. A már meglévő gócokon a heterogén nukleáció folyamata játszódik le. A nukleációt követően a részecskék kondenzációval, vagy termikus koagulációval növekednek tovább, rendszerint legfeljebb az akkumulációs módusig. Laboratóriumi kutatások alapján kimutatták, hogy a nukleáció legfőbb vegyülete a H_2SO_4 , de vannak más kémiai anyagok is, például az NH_3 , aminok, illetve különböző szerves vegyületek, amelyek szerepe a folyamatban még nem teljesen tisztázott (Benson et al., 2011). A nukleáció tehát bonyolult folyamat, mely több környezeti összetevő együttes hatására jön létre. A közöttük lévő kapcsolatrendszer korábban mért adatok statisztikai kiértékelésével próbáljuk felderíteni.

2.1.1. A mérőrendszer működési elve

A légköri aeroszol részecskék szám-koncentrációját elektromos mozgékonyyságon alapuló részecskeszeparátor (DMPS, Differential Mobility Particle Sizer) segítségével határozták meg (1. ábra). A mérőberendezés az ELTE Lágymányosi északi épületének második emeleti teraszán működött. A mérés egy éven keresztül 2008. november 4. és 2009. november 30. között folyamatosan, 10 percenkénti mintavételezéssel zajlott (Salma *et al.*, 2011).

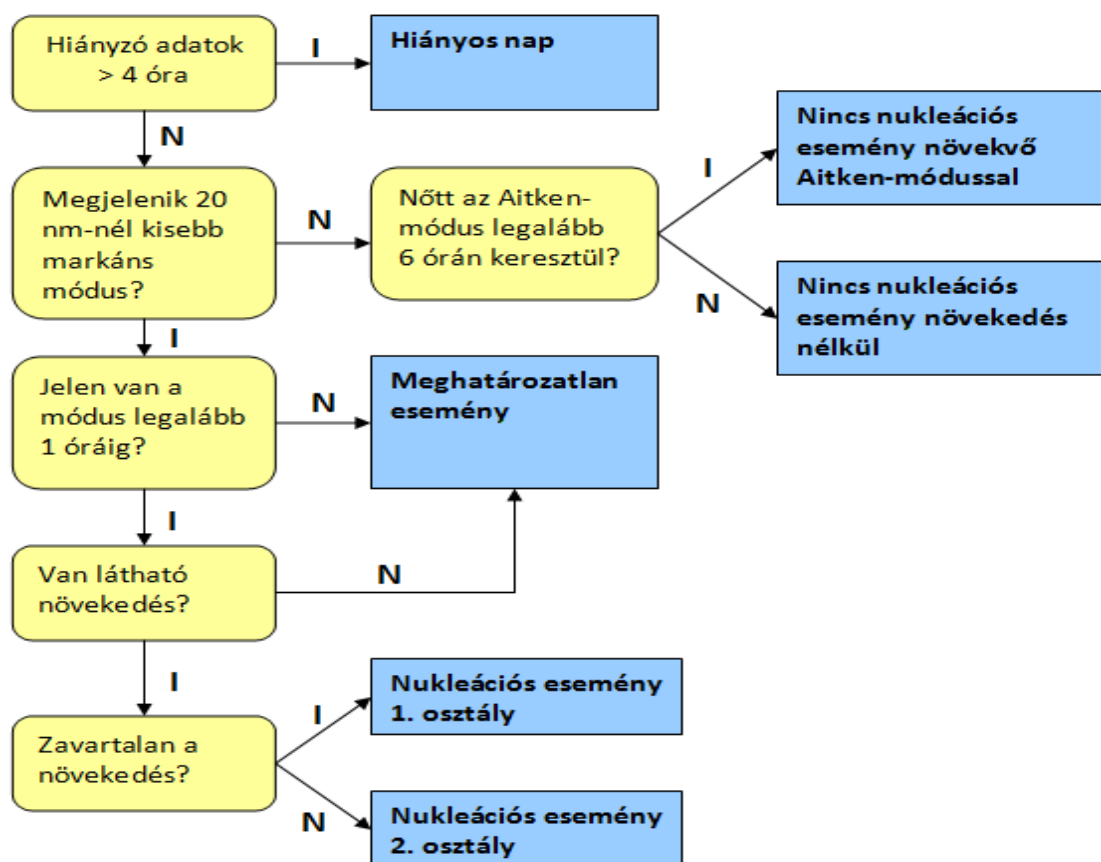
A mérőrendszer három fő egységből áll. A beáramló levegő először egy neutralizálón megy keresztül, melynek feladata, hogy az aeroszol részecskék elektromos töltését egyensúlyi töltéseloszlásra hozza. Egy radioaktív sugárforrás (^{241}Am) ionizálja a levegő molekuláit, mely az aeroszol részecskékkel ütközve egyensúlyi töltéseloszlást eredményez. A részecskék innen a differenciális mozgékonyyság analizátorba (DMA, Differential Mobility Analyser) kerülnek. A DMA elektromos mozgékonyyság alapján szétválogatja a részecskéket. A nagyfeszültséget lépésekben változtatva különböző átmérőjű részecskéket szeparálhatunk. A DMA-ból a részecskék a kondenzációs részecskeszámlálóba (CPC, Condensation Particle Counter) jutnak, melynek működése a lézerfény szórásán alapul. Annak érdekében, hogy a legkisebb részecskéket is detektálhassuk a szórás alapján, szükség van a részecskék megnövesztésére. Ezért a részecskék először egy butanollal telített, 39 °C-ra fűtött térbe jutnak, majd egy 14 °C-ra hűtött kondenzeren haladnak keresztül. Itt a butanol kondenzálódik a részecskékre, így méretük kondenzáció révén megnő, és mindegyik detektálhatóvá válik. Ezután a mintát egy kollimált lézernyalábra merőlegesen vezetjük tovább, amely szórja a lézerfényt. A szórt fény intenzitását egy fotodetektorral mérjük, ami alapján már meghatározhatjuk a részecskék számát, majd a részecskék számának méreteloszlását (Németh 2012). A DMPS műszerrel a mérések jelenleg is folynak az ELTE Lágymányosi épület mellett 2013-ban létesített BpART mérőplatformon (<http://salma.web.elte.hu/BpArt/>).



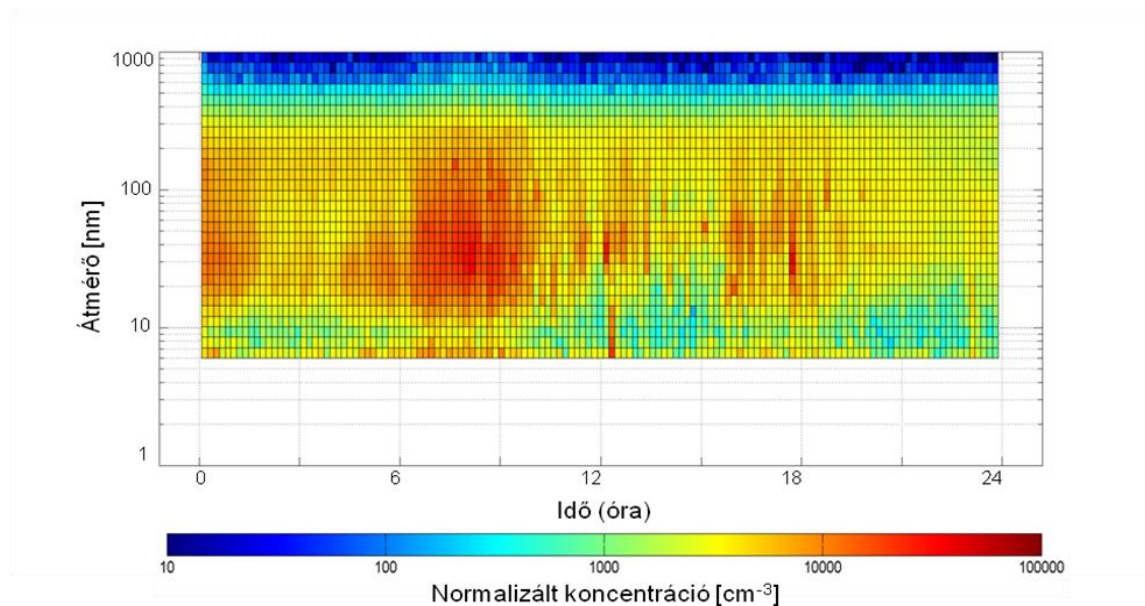
1. ábra: A mérőrendszer az egyetem egyik laborjában (forrás: Salma Imre)

2.1.2. A nukleációs napok azonosítása

A részecskeszám méreteloszlás időbeni változását kontúrgrafikonokon ábrázolják. A grafikon vízszintes tengelyén az időfelbontás szerepel 10 perces osztásközökkel, míg a függőleges tengelyén a részecskék átmérője található. Az adott időpillanatban jelen levő adott átmérőjű részecskék számkoncentrációját színek alapján lehet meghatározni, a kék a legkisebb, a piros pedig a legnagyobb koncentrációsintet jelöli (a 3. ill. a 4. ábrán példaként bemutatok egy nemnukleációs és egy nukleációs esetre vonatkozó grafikont). A kontúr-grafikon a részecskék képződésének és növekedésének folyamatát is szemlélteti. A levegőben lévő gőzökből nukleációval kialakuló, új folyékony vagy szilárd részecskék képződésekor a grafikonokon kirajzolódó, elnyúlt piros alakzat, az ún. banángörbe (Borsós és Salma 2010). A kontúrgrafikonok elemzését követően az egyes napokat a nyilvánvalóan nukleációs, a nyilvánvalóan nem nukleációs, és a nem meghatározható kategóriákba soroltuk. A mérési periódus alatt összesen 83 nukleációs és 229 nemnukleációs napot azonosítottunk. Az osztályozást a 2. ábrán látható folyamatábra alapján végeztük el.

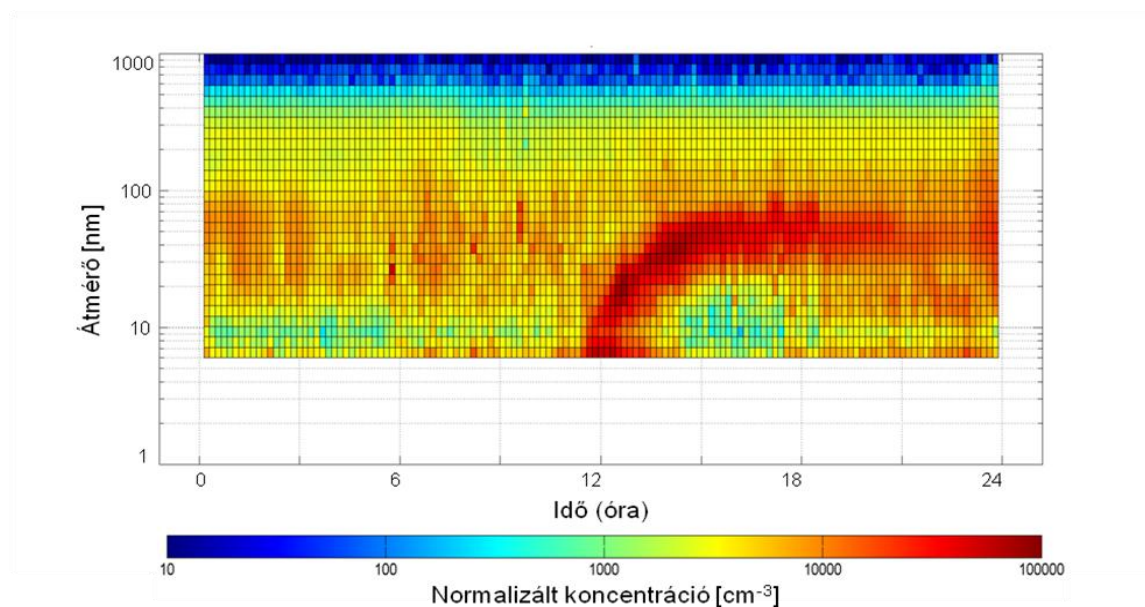


2. ábra: A nukleációs napok meghatározására szolgáló döntési eljárás folyamatábrája (forrás: Salma Imre)



3. ábra A részecskeszám méreteloszlás időváltozása egy nemnukleációs napon, 2009. 05. 04-én, hétfőn (forrás: Salma Imre)

Ezen a kontúrgrafikonon nem jelenik meg nukleációs esemény, csupán egy intenzívebb, a reggeli csúcsforgalomhoz köthető, ill. egy kevésbé intenzívebb kora esti, a közlekedés és a keveredési réteg dinamikájának együttes hatására megjelenő részecskenyelődés látható.



4. ábra A részecskeszám méreteloszlás időváltozása egy nukleációs napon, 2009. 09. 15-én, kedden (forrás: Salma Imre)

Ezen a grafikonon a közlekedés hatása mellett már megjelenik egy elnyúlt, piros alakzat az ún. banángörbe is, amely a nukleációs esemény hatására bekövetkező részecskenyelődést szemlélteti.

2.2. Meteorológiai adatok

A meteorológiai adatokat az ELTE városklíma állomása szolgáltatta, amit az Országos Meteorológiai Szolgálat üzemeltet. Az állomás műszereit a Vaisala–QLC50 típusú automata mérőrendszerhez tartozó talajhőmérséklet, fűszinti hőmérséklet, léghőmérséklet, relatív nedvesség, csapadék, globálsugárzás, szélirány, szélesség, illetve légnyomás mérésére szolgáló szenzorok alkotják (1. táblázat).

1. táblázat: Az egyes meteorológiai elemek mérőműszereinek típusai, működési elve és mérési szintjei

Meteorológiai elem	Műszer típusa	Működési elve	Mérési szint
talajhőmérséklet	Vaisala DTS12G	platina ellenállás	–20 cm, –10 cm, –5 cm
fűszinti hőmérséklet	Vaisala DTS12G	platina ellenállás	5 cm
léghőmérséklet	Vaisala HMP35D	platina ellenállás	2 m
relatív nedvesség	Vaisala HMP35D	kapacitív (HUMICAP)	2 m
csapadék	Vaisala RG13	billenő edényes	35 m
globálsugárzás	CM3, Kipp and Zonen	termoelektromos	35 m
szélirány	Vaisala WAV15A	fotoelektromos	45 m
szélesség	Vaisala WAA15A	fotoelektromos	45 m
légnyomás	Vaisala PTB210	szilikon membrános	45 m

Mind a részecskeszám koncentrációk, mind a meteorológiai paraméterek 10 perces felbontásban, egy éven keresztül lettek rögzítve. A felhasznált meteorológiai adatok: levegőhőmérséklet [°C], relatív páratartalom [%], szélesség [m/s], szélirány [°], csapadék [mm], globálsugárzás [W/m²] és légnyomás [hPa].

2.3. Kulcsfontosságú légszennyezők

A nukleáció várhatóan kapcsolatban van a levegőkémiai körülményekkel, a légköri nyomanyag koncentrációkkal. A légszennyező anyagok koncentrációját az Országos Légszennyezettségi Mérőhálózat (OLM) méri. A kiértékelés során felhasznált levegőszennyező anyagok a SO_2 [$\mu\text{g}/\text{m}^3$], PM_{10} [$\mu\text{g}/\text{m}^3$], NO_x [$\mu\text{g}/\text{m}^3$] és az O_3 [$\mu\text{g}/\text{m}^3$]. Ezek órás adatait a legközelebbi, a Kosztolányi Dezső téren elhelyezett mérőállomás szolgáltatta. Az egyes légszennyezők koncentrációját különböző módszerekkel állapítják meg, amit a 2. táblázatban foglaltam össze.

2. táblázat: Az egyes légszennyezők koncentrációjának meghatározási módszerei

Légszennyező	Módszer neve
SO_2	UV fluoreszcencia
PM_{10}	béta radiometria
NO_x	kemilumineszcencia
O_3	UV abszorpciós fotometria

2.4. Statisztikai módszerek

A nukleáció hátterében álló, a folyamat szempontjából kedvező paraméterek feltárása összetett feladat. A probléma úgy fogalmazható meg egyszerűen, hogy egy adott kétértékű függő változó (aznap volt nukleáció vagy nem volt) és az adott független változók (meteorológiai és légszennyező adatainak) esetén hogyan írhatjuk le a közöttük levő statisztikai kapcsolatot. Tehát a probléma megoldására csakis különböző összetett statisztikai módszerek jöhetnek számításba, mint például: klaszterelemzés (*Hussein et al., 2014*), diszkriminancia analízis (*Mikkonen et al., 2006*), vagy a logisztikus regresszió (*Dévényi és Gulyás 1984*). A következőkben megadom e módszerek rövid irodalmi áttekintését.

2.4.1. Klaszteranalízis

A klaszteranalízis segítségével megfigyeléseinket csoportokba (klaszterekbe) oszthatjuk, még hozzá úgy, hogy az egyes csoportokon belül a megfigyelések egymáshoz hasonlóak legyenek, de a csoportok viszont jól elkülönüljenek egymástól. Más szavakkal kifejezve az analízist arra használhatjuk, hogy megfigyeléseinket homogén, eltérő csoportokba soroljuk. A klaszterezési eljárások a megfigyelések és a klaszterek hasonlóságának vizsgálatakor a hasonlóságnak egy precízebb definícióját követelik meg. Változók csoportosítása esetén általánosan elfogadott a távolság fogalom alkalmazása a hasonlóság megállapítására. Két pont távolságának meghatározására gyakran használják az euklideszi távolság definícióját, ami a következő összefüggéssel írható le:

$$D(i, j) = \sqrt{A^2 + B^2} = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2}, (1)$$

ahol $D(i, j)$: i -edik és j -edik megfigyelések távolsága, A, B , és $(X_{1i}, X_{2i}), (X_{1j}, X_{2j})$: tetszőleges objektumok

Egy alternatív becslés a négyzetes euklideszi távolság, melynek alakja a következő:

$$D(i, j) = A^2 + B^2 = (X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2. (2)$$

Egy i megfigyelést akkor tekintünk közelebbinek (hasonlóbbnak) j -hez, egy k megfigyeléshez képest, ha $D(i, j) < D(i, k)$. A távolság becsléseket természetesen kiterjeszthetjük kettőnél több változó esetére is. Például egy $(X_{1i}, X_{2i}, \dots, X_{ki})$ és egy $(X_{1j}, X_{2j}, \dots, X_{kj})$ objektum euklideszi távolsága:

$$D(i, j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2}. (3)$$

Ezen kívül, ha ismerjük az egyes változók fontosságát kifejező $w_1, w_2, \dots, w_k$ súlyokat, akkor kiszámolhatjuk az ún. súlyozott távolság mértékét is a következőképpen:

$$D(i, j) = \sqrt{w_1(X_{1i} - X_{1j})^2 + w_2(X_{2i} - X_{2j})^2 + \dots + w_k(X_{ki} - X_{kj})^2}. (4)$$

Klaszterezési módszerek:

A klaszterezést kezdetben első megközelítésben úgy is, hogy minden egyes megfigyelést külön klaszternek tekintünk. Elsőként vonjuk össze azt a két megfigyelést, amelyek legközelebb vannak egymáshoz, így $n - 1$ klaszterünk marad a továbbiakban. A következő lépésben az egymáshoz legközelebb levő megfigyelési párt vonjuk egy klaszterbe össze, így már csak $n - 2$ klaszterünk marad. A lépéseket ezek alapján addig folytatjuk, míg végül egyetlen klasztert kapunk, amely tartalmazza az összes n db megfigyelésünket. Minden lépés során tartunk meg a távolság mértékét, amellyel a csoportokat kialakítjuk. Ahhoz, hogy meghatározzuk a csoportok számát, azokat a lépéseket kell figyelembe vennünk, amelyekben az összevonó távolság viszonylag nagy volt. A módszer problémája az, hogy hogyan tudjuk megmérni két vagy több megfigyelést tartalmazó klaszter távolságát. Minden bizonnyal a legegyszerűbb mód az, ha a klaszterekben lévő két egymáshoz legközelebb eső elem távolságát alkalmazzuk a két klaszter közötti távolság meghatározására. Ezt úgy is nevezik, mint a legközelebbi szomszéd módszer.

Ezzel szemben használhatjuk a legtávolabbi szomszéd módszerét is, amely azt mondja, hogy két klaszter távolsága a két legtávolabbi elemük távolságával adható meg. A két módszer egyes esetekben egymással megegyező eredményeket adhat, más esetekben viszont nem biztos, hogy megegyeznek. A legközelebbi szomszéd módszer valószínűleg nem a szabad szemmel észlelhető módon fogja a csoportokat kialakítani. Ez azért lehet, mert néhány közbenső lépés során összeolvasztja a legközelebbi orrpontokat egy azonos klaszterbe, a megmaradó pontokon pedig egy zsinór mentén, láncszem mintában halad végig. A legtávolabbi szomszéd módszer várhatóan azonosítja a két klasztert, mivel hajlamos arra, hogy elutasítsa azon klasztereknek az összevonását, amelyeknek az elemei lényegesen eltérnek a többi klaszterétől. Más részről a legközelebbi szomszéd módszer valószínűleg sikerebben alakítja ki a két csoportot, mint a legtávolabbi szomszéd módszer.

Egy kompromisszumos eljárás az átlagos kapcsolat alkalmazása, amely a klaszterek közötti távolságként az egyik, illetve a másik klaszterből vett elempárok közötti távolságok átlagát veszi figyelembe. A három előbbi módszer az ún. hierarchikus, összegyűjtő klaszterezési eljárások csoportjába tartozik. Hierarchikus, mivel minden klasztert az előzőleg kialakított klaszterek összevonásával kapunk. Összegyűjtő, mivel annyi

csoportból indulunk ki, ahány megfigyelésünk van, végül pedig egyetlen csoport tartalmazza az összes megfigyelésünket.

Természetesen léteznek nem-hierarchikus klaszterezési eljárások is, mint például a k-közép módszer. A k-közép eljárás a legegyszerűbb módon a következő lépésekkel írható le. Elsőként határozzuk meg önkényesen vagy tudatosan a klaszterek számát, és az egyes klaszterek tagjait! Ezután számoljuk ki a klaszterek centroidjait, és határozzuk meg az egyes megfigyelések és a centroidok közötti távolságokat! A centroid egy adott klaszterben levő elemek átlagaként adható meg, tehát a klaszterben levő megfigyelések középpontjának tekinthető. Ha egy megfigyelés közelebb van egy adott klaszter centroidjához, mint amelyhez jelenleg tartozik, akkor azt átírjuk a másik, közelebbi klaszterbe. Az előző lépést addig ismételjük, amíg minden megfigyelés a számára legközelebbi középpontú klaszterbe nem kerül. Ha nem kellő mértékben adtuk meg az első lépésben a klaszterek számát, akkor ismételjük meg az előző lépéseket egy eltérő számú klaszterre vonatkozóan, és értékeljük ki az eredményeket. A k-közép módszert a statisztikai programok gyakran használják gyors klaszterezésre (*Tryfos 1998; Füstös 2009*).

Összefoglalva tehát a klaszteranalízis meghatározza, hogy mennyi természetes csoport alkotható a megfigyeléseinkből. Az elemzés során azt is megengedi, hogy szabadon eldöntsük, hogy melyik változót melyik csoportba soroljuk be. A klaszteranalízis nem olyan, mint egy tipikus statisztikai teszt, mivel ez egy gyűjteménye a különböző algoritmusoknak, amelyek a vizsgált elemeket a megfelelő klaszterbe sorolják, jól definiált hasonlósági szabályok alapján. A módszer célja az, hogy a klasztereken belüli változékonyságot minimalizálja, a klaszterek közötti eltéréseket pedig maximalizálja (*Burns és Burns 2008*).

2.4.2. Diszkriminancia analízis

A diszkriminancia függvényelemzés egy többváltozós statisztikai módszer. A diszkriminancia analízis segítségével a folytonos független változókat két vagy több csoportba oszthatjuk. Az analízis könnyebb megérthetősége miatt célszerű először csak két csoportra vizsgálni, mivel onnan már egyszerűen áttérhetünk n csoportra. Vezessük be a következő jelölés-rendszert: X_{ji} : a j -edik független változó i -edik értéke, b_j : a j -edik változóhoz tartozó diszkriminancia együttható, Z_i : az i -edik diszkriminancia találat, z_{crit} : a diszkriminancia találathoz tartozó kritikus érték.

Legyen minden egyes Z_i diszkriminancia találat a független változók lineáris függvénye:

$$Z_i = b_0 + b_1X_{1i} + b_2X_{2i} + \dots + b_nX_{ni}. \quad (5)$$

Az osztályozási folyamat a következők szerint alakul: ha $Z_i > z_{crit}$, akkor az adott változót az 1. csoportba soroljuk, míg ha $Z_i < z_{crit}$, akkor pedig a 2. csoportba kerül. Ahol a kritikus értéket a következő összefüggés segítségével írhatjuk le:

$$Z_{crit} = b_0 + b_1X_{1i} + b_2X_{2i} + \dots + b_nX_{ni}. \quad (6)$$

Az osztályozási határ $n = 2$ osztály esetén egy egyenes vonal lesz. A vonal egyik oldalán szereplő értékek az 1. csoportba, míg a vonal másik oldalán szereplő értékek a 2. csoportba kerülnek. Ha $n = 3$, akkor az osztályozási határ egy 2D-s síkfelület lesz a 3D-s térben, tehát a besorolási határ rendszerint egy $n - 1$ dimenziós hipersík az n dimenziós térben. Egy lineáris osztályozási eljárást akkor tekinthetünk optimálisnak, ha a két csoport kovariancia mátrixai megegyeznek. A csoportok közötti távolság meghatározására általában a Mahalanobis-féle távolságot használják. A Mahalanobis-féle távolság transzformációjával az F -statisztika értékéhez jutunk, mely megmondja, hogy a két csoport statisztikailag mennyire különbözik egymástól. Az F -próbastatisztika az 1. csoporthoz tartozó x_1 átlagvektor és a 2. csoporthoz tartozó x_2 átlagvektor közötti különbséget vizsgálja. Általában azt tesszük fel, hogy a két csoport az 0,01-os szignifikancia szinten szignifikánsan különbözik egymástól. Az F -statisztika szignifikanciája, azonban nagyon gyenge indikátora annak a hatékonyságnak, amellyel az egyes változókat az 1. illetve a 2. csoportba soroljuk.

A diszkriminancia analízis során kapott eredményekkel kapcsolatban három fő kérdést tehetünk fel, melyekre a következőkben megpróbáljuk megadni a választ.

Mely független változók diszkriminálnak a legjobban?

A b_j diszkrimináló együttható nagysága és előjele meghatározza az X_j független változó hatását. Tételezzük fel, hogy a következő összefüggés fennáll:

$$|b_j| > |b_k|. \quad (7)$$

Ekkor az X_j változó jobban képes szétválasztani az 1. és a 2. csoportot, mint az X_k . Egy változó annál jobban diszkriminál, minél nagyobb hatása van a Z_i diszkriminancia találatra. A diszkriminancia együttható előjele meghatározza az X_j változó hatásának irányát. Ha a b_j pozitív, akkor X_j növekedésével Z_i értéke is nő, és minél nagyobb Z_i értéke, annál valószínűbb, hogy az i -edik jellemző paraméter az 1. csoportba tartozik.

Mennyire jól különítik el ezek a független változók az értékeket az egyes csoportok mentén?

Erre a kérdésre a klasszifikációs táblázatok és az esély kritériumok alapján adhatunk választ. Az egyik ilyen kritérium az esélyhányados kritérium:

$$c_{pro} = \alpha^2 + (1 - \alpha)^2, \quad (8)$$

a másik pedig a maximális esély kritériuma:

$$c_{max} = \max(\alpha, 1 - \alpha), \quad (9)$$

ahol c_{pro} : az esélyhányados kritérium, c_{max} : a maximális esély kritériuma, α : az 1. csoportba sorolt paraméterek aránya, $1 - \alpha$: pedig a 2. csoportba sorolt elemek aránya.

Tehát ha megvizsgáljuk a klasszifikációs táblázatainkat, és a döntési kritériumokat, akkor eldönthetjük, hogy melyik független változónak mekkora a diszkrimináló ereje.

Abban az esetben, amikor az egyik csoport sokkal nagyobb, mint a másik, akkor majdnem az összes paraméter, amit vizsgálunk a nagyobb csoportba fog kerülni. Viszont, gyakran a kisebb csoport sokkal érdekesebb számunkra ezért, ilyenkor a klasszifikációs táblázatok alapján nem tudjuk megfelelően elbírálni a független változók diszkrimináló erejét.

Milyen döntési szabályt használhatunk az egyedi osztályozások esetén?

Ha a diszkriminancia függvényt arra használjuk, hogy egyedi döntéseket osztályozzunk, akkor a téves besorolásokat is egyértelműen bele kell írunk a döntésbe. A magyarázathoz vezessük be a következő jelöléseket:

- $P(I/X_i)$: annak az utólagos valószínűsége, hogy egy paraméter az 1. csoportba tartozik, miután megvizsgáltuk az X_i független változóhoz tartozó vektorát,
- $P(II/X_i)$: ugyanez csak a 2. csoportra,
- c_{12} : a 2. csoportba való sorolás lehetséges hibája, ha az adott paraméter valójában az 1. csoportba tartozik
- c_{21} : az 1. csoportba való sorolás lehetséges hibája, ha az adott paraméter valójában a 2. csoportba tartozik.

Minden racionális esetben $c_{11} = c_{22} = 0$ kell, hogy legyen. Ha az i -edik paramétert az 1. csoportba soroljuk, akkor a lehetséges hiba a következő:

$$K_i(I) = P\left(\frac{II}{X_i}\right) c_{21}, (10)$$

hasonlóan a 2. csoportra:

$$K_i(II) = P\left(\frac{I}{X_i}\right) c_{12}. (11)$$

Ha fennáll, hogy:

$$K_i(I) < K_i(II), (12)$$

akkor az i -edik paramétert az 1. csoportba soroljuk. Természetesen ellentétes irányú reláció esetén fordítva járunk el. Ezzel megkaptuk a klasszifikációs eljárást (Morrison 1969; Füstös 2009).

Összefoglalásul, néhány megfontolás:

1. A lineáris diszkriminancia függvény akkor alkalmazható, amikor a csoportok kovariancia mátrixai egyenlők (vagy közel egyenlők).
2. Az F -próba csak a csoportok közötti eltérés statisztikai szignifikanciáját teszteli.
3. Az effektív minta méretét a kisebb csoport határozza meg.
4. A klasszifikációs döntés során vegyük figyelembe az utólagos valószínűségeket és a téves besorolások lehetséges hibáit is (*Morrison 1969*).

2.4.3. Logisztikus regresszió

A logisztikus regresszió azt határozza meg, hogy több egyidejűleg bemutatott független változó hatásának figyelembevételével hogyan tudjuk előrejelezni a függő változó értékét. A logisztikus regresszió kétféle módon használható: egyrészt, hogy előrejelezzük egy csoport tagjait, másrészt ismeretet nyújt a változók közötti kapcsolatokról és azok erősségéről.

A logisztikus regresszió előfeltételei:

- A logisztikus regresszió nem feltételez lineáris kapcsolatot a függő és a független változók között.
- A függő változó csak két értéket vehet fel (0, 1/igen, nem).
- Nem szükséges, hogy a független változók egy intervallumba essenek, és az sem, hogy normális eloszlásúak, lineárisan összefüggőek, illetve az egyes csoportokon belül egyenlő eloszlásúak legyenek.
- A csoportoknak egymást kölcsönösen kizárónak és egyértelműnek kell lennie, egy attribútum csak egy csoportban lehet, és minden attribútumnak szerepelnie kell valamelyik csoportban.
- Nagyobb minták szükségesek, mint a lineáris regresszió esetén, mert a maximum likelihood módszer együtthatói nagy minta alapján vannak megbecsülve. Magyarázó változókként minimum 50 attribútum ajánlott (*Burns és Burns 2008*).

A logisztikus regresszió egyenlete

A logisztikus regressziós analízis során a vizsgált esemény bekövetkezésének π_x valószínűségét akarjuk becsülni, ezért a $[0, 1]$ intervallumot kell transzformálni az egész számegetesre. Ezt a logisztikus függvény segítségével tudjuk megoldani:

$$\pi_x = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k}}, \quad (13)$$

$$\text{logit}(\pi_x) = \ln(\text{odds}_x) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k, \quad (14)$$

$$\text{odds}_x = \frac{\pi_x}{1 - \pi_x} = e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k}, \quad (15)$$

ahol β : a regressziós együttható, x : a vizsgált magyarázó változó, π_x : a vizsgált esemény bekövetkezésének valószínűsége, $odds_x$: a vizsgált esemény esélyhányadosa

Az *odds* (magyarul esélyhányados), azt mutatja meg, hogy egy bizonyos magyarázó-változó értékének növelése esetén mennyivel nő vagy csökken a vizsgált esemény bekövetkezésének a valószínűsége az adott esemény elmaradásának valószínűségéhez képest. Ha az értéke nagyobb, mint 1, akkor a változó növelése esetén nő a bekövetkezés valószínűsége az elmaradáséhoz képest, viszont ha kisebb, mint 1 (de legalább 0), akkor csökken. A logisztikus regresszió az X_i magyarázó-változók melletti Y kimenet valószínűségét az alábbi képlettel definiálja, ahol a β paramétereket fogjuk becsülni megfelelő b vektorral, y_i pedig az eredményváltozó i . mintaelemen megfigyelt értéke:

$$P(Y = y_i) = \frac{e^{y_i(\beta_0 + \beta_1 x_{i,1} + \dots + \beta_k x_{i,k})}}{1 + e^{\beta_0 + \beta_1 x_{i,1} + \dots + \beta_k x_{i,k}}}. \quad (16)$$

Mivel az eredményváltozó bináris, ezért célszerű maximum likelihood módszerrel becsülni b -t. Így az alábbi feladat megoldását keressük (iterációs eljárással):

$$L(b) = \prod_{i=1}^n P(Y = y_i/b) = \prod_{i=1}^n \frac{e^{y_i(b_0 + b_1 x_{i,1} + \dots + b_k x_{i,k})}}{1 + e^{b_0 + b_1 x_{i,1} + \dots + b_k x_{i,k}}} \rightarrow \max. \quad (17)$$

A modell illesztése és a likelihood függvény

A maximum likelihood módszer egy meghatározott hipotézis melletti valószínűséget is jelent. A logisztikus regresszió esetén két hipotézist is feltehetünk:

- a null hipotézis, amikor a regressziós egyenletben az együtthatók értéke nulla, és
- a váltó hipotézis, amikor a modell a magyarázó változókkal pontos, és jelentősen eltér a nullától, azaz lényegesen jobb eredményeket ad, mint a null hipotézis véletlen becslési szintje.

A likelihood módszert alkalmazva az eredmény rendszerint egy nagyon kis szám. Azért, hogy könnyen kezelhetővé tegyük, vesszük a természetes alapú logaritmusát, ezzel előállítva a log likelihood-ot (LL). Mivel a valószínűség mindig kisebb egynél, ezért az LL mindig negatív. Tehát a log likelihood módszer az alapja egy logisztikus modell tesztelésének.

Az aktuális modell mintához való közelségét, illeszkedését a likelihood arány teszt segítségével mérhetjük. Ez azt jelzi, hogy a vizsgált modell relatíve milyen távol (közel) van a szaturált modelltől, amelyet maga a minta testesít meg. Minél kisebb ez az érték, annál több információt reprodukál modellünk a mintából. A likelihood arány teszt a $-2LL$ arányon alapul, és a következő összefüggés segítségével írható le:

$$D = -2\ln \frac{L_{akt}}{L_{szat}}, (18)$$

ahol D : a deviancia, L_{akt} : az aktuális modellre, L_{szat} : pedig a szaturált (telített) modellre felírt likelihood függvény.

A kapott eredmény megközelítőleg khi-négyzet eloszlású lesz. A khi-négyzet eloszlás értékelésekor a jelentéktelen khi-négyzet értékek nagyon kevés tisztázatlan eltérést mutatnak, így a modell illeszkedése jó. Azonban, a lényeges khi-négyzet értékek azt mutatják, hogy az eltérések jelentős része tisztázatlan. A deviancia két mértéke kiemelten fontos a logisztikus regresszió esetén: a nulla deviancia, és a modell deviancia. A nulla deviancia az előrejelző nélküli, és a szaturált modell közötti különbséget mutatja. A modell deviancia pedig a legalább egy előrejelzőt tartalmazó, és a szaturált modell közötti különbséget mutatja. Ebben a vonatkozásban a nulla deviancia a bázis, amelyet összevethetünk az előrejelző modellekkel. Adván, hogy a deviancia az aktuális, illetve a szaturált modell különbségének a mértéke, a kisebb értékek jobb illeszkedést mutatnak. Tehát, hogy megbecsüljük egy előrejelzőnek, vagy az előrejelzők sorának a közreműködését, a nulla devianciából kivonhatjuk a modell devianciát, és megbecsülhetjük a különbséget egy szabadsági fokú khi-négyzet eloszlás segítségével. Ha a modell deviancia lényegesen kisebb a nulla devianciánál, akkor arra következtethetünk, hogy az előrejelző, vagy az előrejelzők sora lényegesen növeli a modell illeszkedését (*Burns és Burns 2008; Székelyi és Barna 2002*).

2.4.4. Kategóriás verifikáció

Ahhoz, hogy egy előrejelzésről megállapíthassuk, hogy jó volt-e vagy sem, valamilyen statisztikai mérőszámra van szükségünk. Ezeket a mérőszámokat az előrejelzések és a megfigyelések számszerű összehasonlítása során kapjuk. A mérőszámok segítségével tudjuk jellemezni az előrejelzésünk jóságát. Ezt az eljárást verifikációnak nevezzük.

A meteorológiában számos esetben találkozunk bináris kimenetelű, igen/nem előrejelzésekkel. Ilyenkor azt vizsgáljuk, hogy egy adott esemény bekövetkezik-e, vagy nem. Tipikus bináris előrejelzések a meteorológiában az eső, a hó, az erős szél, az éjszakai fagy, vagy a köd bekövetkezésének vizsgálata. Jelen munkában a légköri nukleáció előrejelzése során merül fel a módszer használata, amelynek az előbb felsorolt meteorológiai paraméterekhez hasonlóan minden esetben kétértékű kimenetele van. Egy bináris előrejelzés verifikálásához, első lépésként egy 2 x 2-es kontingencia táblázatot kell készítenünk, amely tartalmazza az igenek, illetve a nemek előfordulási gyakoriságát, mind az előrejelzés, mind a megfigyelés esetében (lásd 3. táblázat). Azt mondhatjuk, hogy a „helyes”, vagy a „helyes elvetés” esetén az előrejelzésünk korrekt, viszont a „hibás”, vagy „téves riasztás” esetén nem megfelelő az előrejelzés. Egy tökéletes előrejelzés esetén csak „helyes”, illetve „helyes elvetés” találatokat kellene kapnunk, a többi találat („hibás”, illetve „téves riasztás”) pedig mind nulla lenne (Nurmi 2003).

3. táblázat Egy 2x2-es kontingencia táblázat szöveges értelmezésben (Nurmi 2003)

Előrejelzés	Megfigyelés		
	Igen	Nem	Σ
Igen	„helyes”	„téves riasztás”	előrejelzés jó
Nem	„hibás”	„helyes elvetés”	előrejelzés rossz
Σ	megfigyelés jó	megfigyelés rossz	szumma

4. táblázat Egy 2x2-es kontingencia táblázat a találatokat betűkkel jelölve (Nurmi 2003)

Előrejelzés	Megfigyelés		
	Igen	Nem	Σ
Igen	a	b	$a + b$
Nem	c	d	$c + d$
Σ	$a + c$	$b + d$	$a + b + c + d = n$

A kontingencia táblázat értékeinek felhasználásával kiszámolhatunk különböző verifikációs mérőszámokat (pl. 4. táblázat). Ezen mérőszámok mindegyike lényeges információval rendelkezik, de egyetlen számként az előrejelzés jóságát nem tudják minden tekintetben jellemezni. Néhány gyakoribb mérőszámot az 5. táblázatban mutatok be.

5. táblázat Néhány fontosabb verifikációs mérőszám elnevezése és kiszámítási módja (Stephenson 2000)

Elnevezés	Kiszámítási mód
Bias	$B = \frac{a + b}{a + c}$
Hit rate	$H = \frac{a}{a + c}$
False alarm rate	$F = \frac{b}{b + d}$
Proportion correct	$PC = \frac{a + d}{n}$
Heidke skill score	$HSS = \frac{2(ad - bc)}{(a + c)(c + d) + (a + b)(b + d)}$
Gilbert skill score	$GSS = \frac{a}{a + b + c}$
Peirce skill score	$PSS = \frac{ad - bc}{(a + c)(b + d)}$

3. EREDMÉNYEK ÉS MEGVITATÁSUK

3.1. Időeltolás és keresztkorreláció

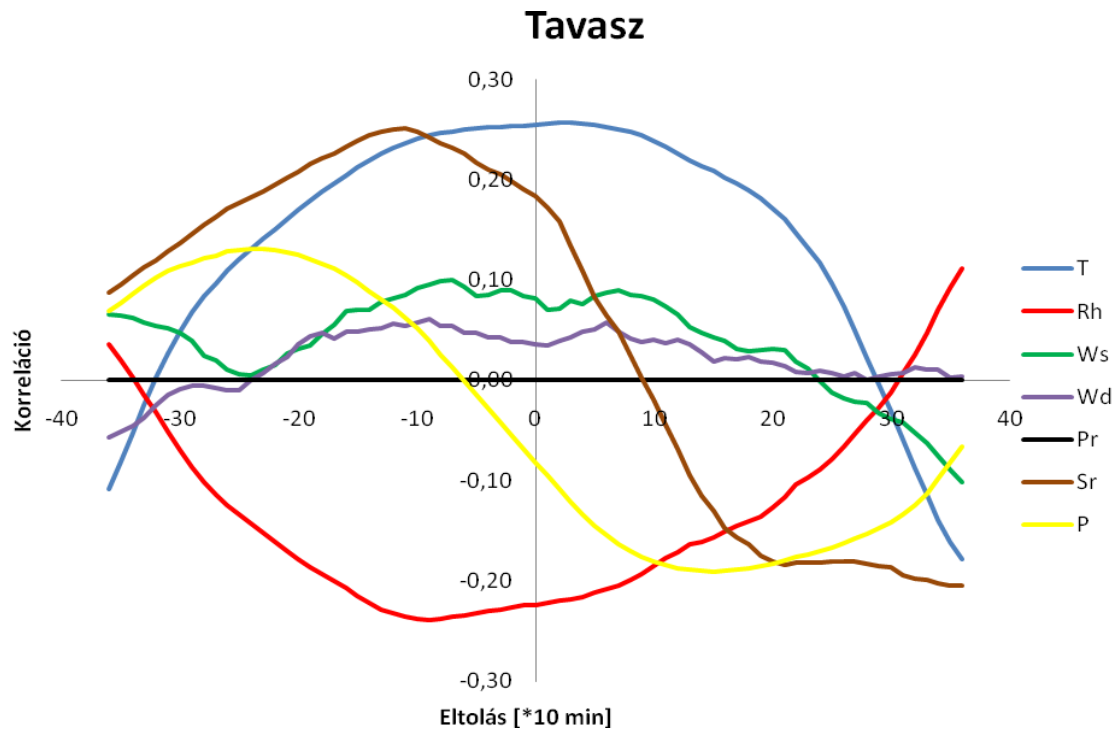
A nukleáció időtartamát 6 UTC és 12 UTC között vártam, mivel az esetek döntő többségében ezen időszak alatt ment végbe a folyamat. Mivel az egyes magyarázó változók értékei nem csak ebben a 6 órás intervallumban lehetnek érdekesek, hanem az is fontos, hogy az időszak előtt, illetve utána hogyan változott az értékük, ezért végrehajtottam egy eltolást az adatsorok és a nukleáció időszaka között. Az eltolást úgy végeztem el, hogy az adatsorokból a vizsgált 6 órás időszakra minden változóra normált-átlagokat képeztem, majd az így kapott normált-értékeket toltam el a nukleáció feltételezett 6 UTC-s kezdő időpontjához képest ± 6 órával. Az eltolást egy excel-makró segítségével készítettem el, mely kiszámolta az egyes normált értékek és a nukleáció közötti keresztkorrelációs együtthatókat is. A normálást és a korrelációs számítást az alábbi egyenletek alapján végeztem el:

$$\hat{x}_{i,j} = x_{i,j} - \bar{x}_i, (19)$$

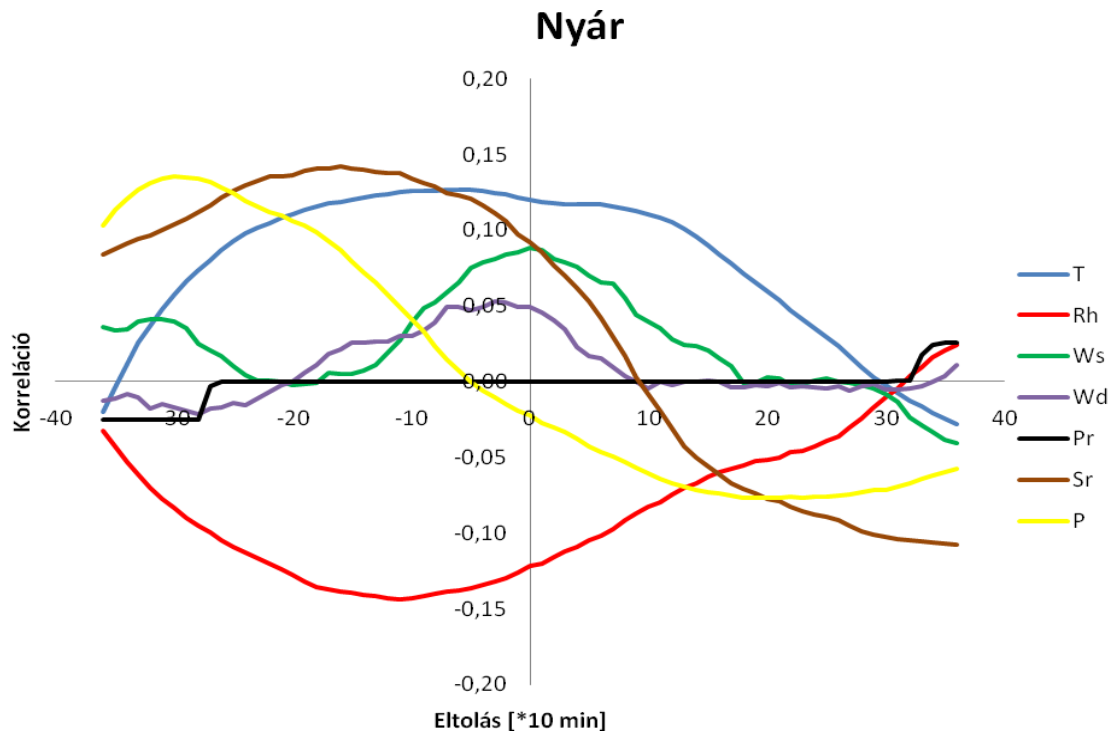
$$R(y_{i,j}, \hat{x}_{i,j+k}) = \frac{\text{cov}(y_{i,j}, \hat{x}_{i,j+k})}{\sigma_{y_{i,j}} \sigma_{\hat{x}_{i,j+k}}}, (20)$$

ahol $\hat{x}_{i,j}$: a normált érték, $x_{i,j}$: az adatsor i,j -edik eleme, $\bar{x}_i$: az adatsor vizsgált 6 órás intervallumra számított átlaga, R : korreláció, $y_{i,j}$: nukleációs adatsor i,j -edik eleme, cov : kovariancia, $\sigma_{y_{i,j}}$: nukleációs adatsor szórása, $\sigma_{\hat{x}_{i,j+k}}$: eltolt, normált-értékeket tartalmazó adatsor szórása, k : eltolás mértéke

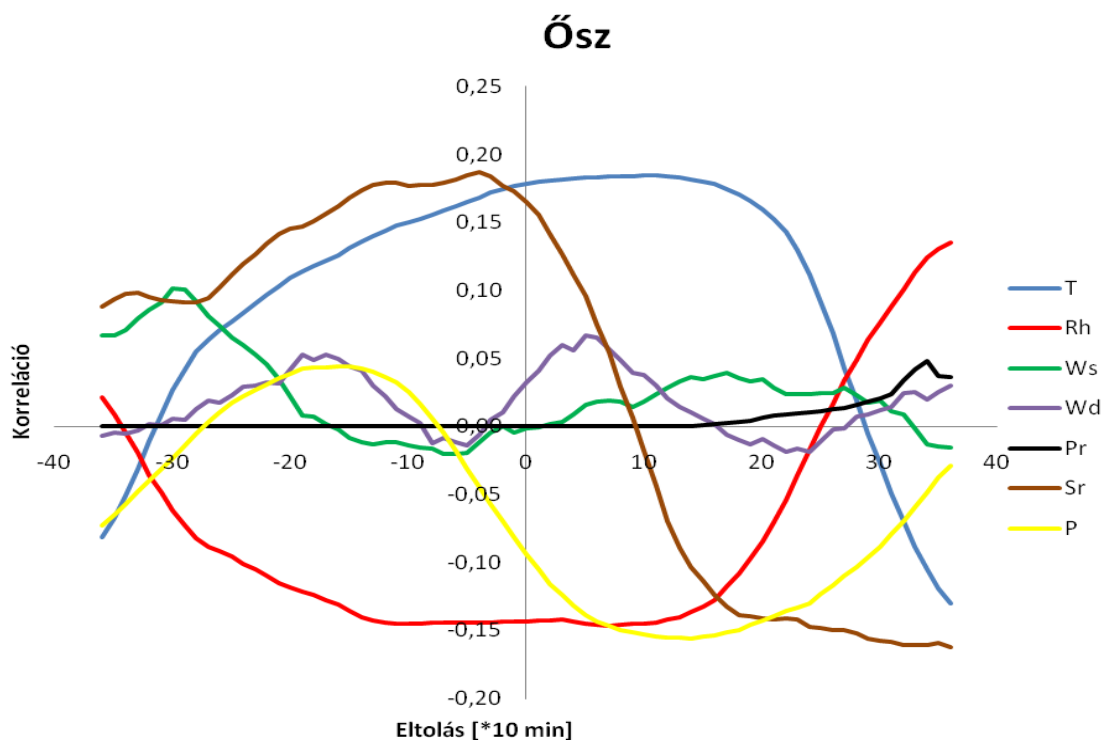
Az így előállított, új, eltolt adatsorokat évszakonként, a korrelációs együtthatóikat az eltolás mértékének függvényében ábrázolva hasonlítottam össze. A kapott évszakonkénti grafikonokból leolvasható, hogy melyik paraméternek volt nagyobb a hatása a nukleációra, és hogy ez a nukleáció előtt mennyi idővel következett be. Ezen grafikonok alapján kerültek kiválasztásra azok az adatsorok, amelyek döntően befolyásolhatják a nukleáció kialakulását.



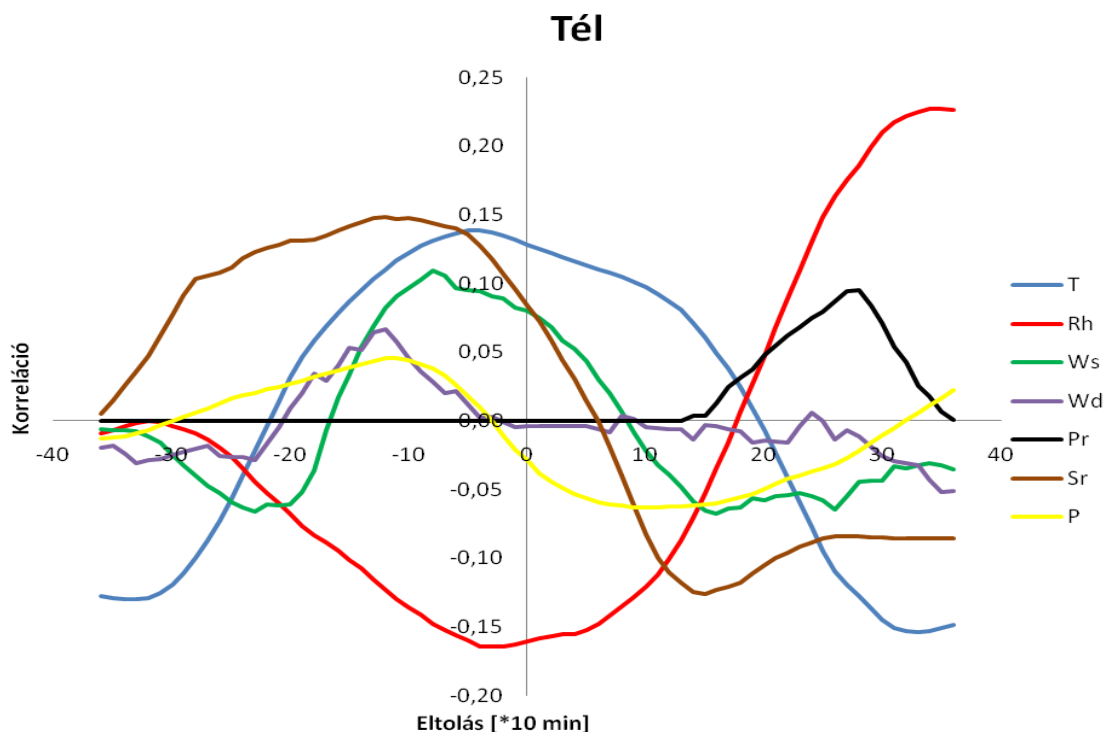
5. ábra A hőmérséklet (T), relatív nedvesség (Rh), szélsébség (Ws), szélirány (Wd), csapadék (Pr), globálsugárzás (Sr), és légnyomás (P) adatsorok nukleációval való korrelációja az időeltolás függvényében ábrázolva tavasszal.



6. ábra A hőmérséklet (T), relatív nedvesség (Rh), szélsébség (Ws), szélirány (Wd), csapadék (Pr), globálsugárzás (Sr), és légnyomás (P) adatsorok nukleációval való korrelációja az időeltolás függvényében ábrázolva nyáron.

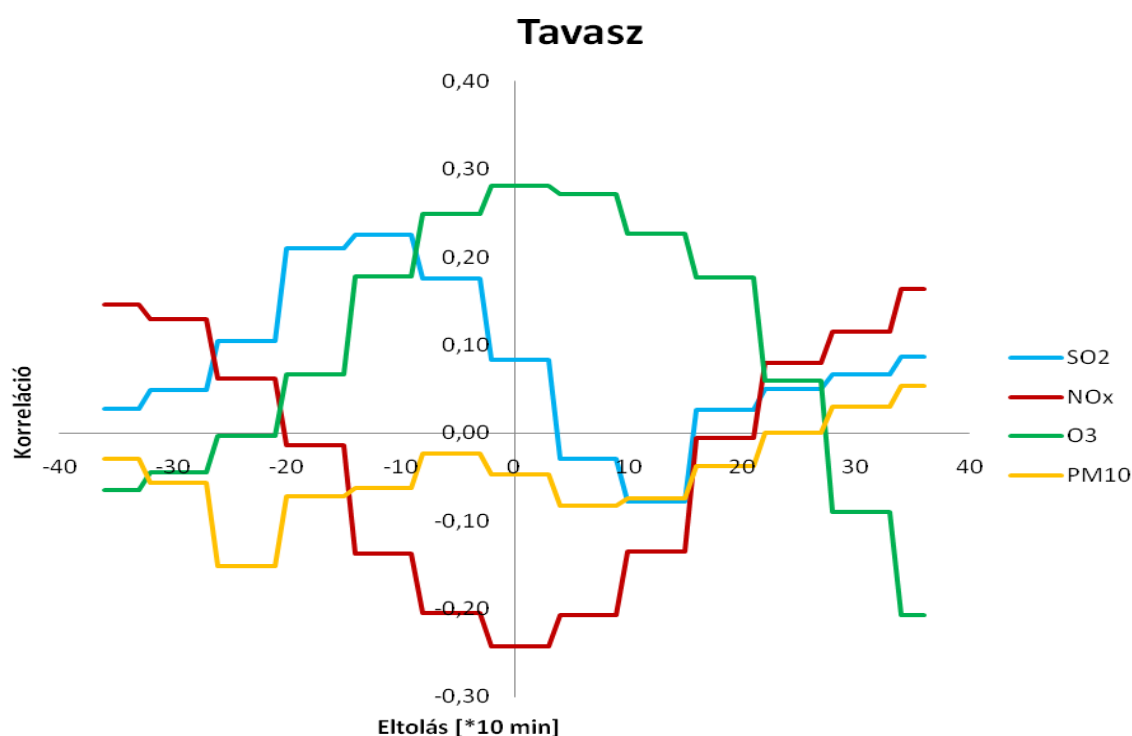


7. ábra A hőmérséklet (T), relatív nedvesség (Rh), szélesség (Ws), szélirány (Wd), csapadék (Pr), globálisugárzás (Sr), és légnyomás (P) adatsorok nukleációval való korrelációja az időeltolás függvényében ábrázolva őszel.

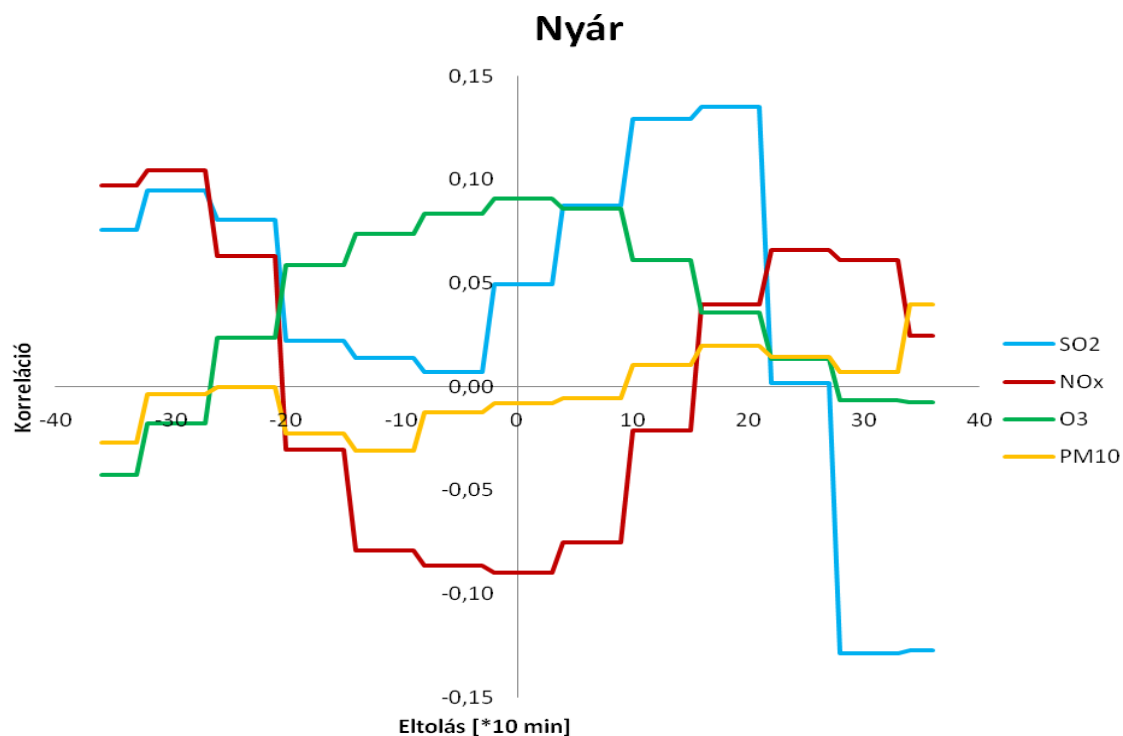


8. ábra A hőmérséklet (T), relatív nedvesség (Rh), szélesség (Ws), szélirány (Wd), csapadék (Pr), globálisugárzás (Sr), és légnyomás (P) adatsorok nukleációval való korrelációja az időeltolás függvényében ábrázolva télen.

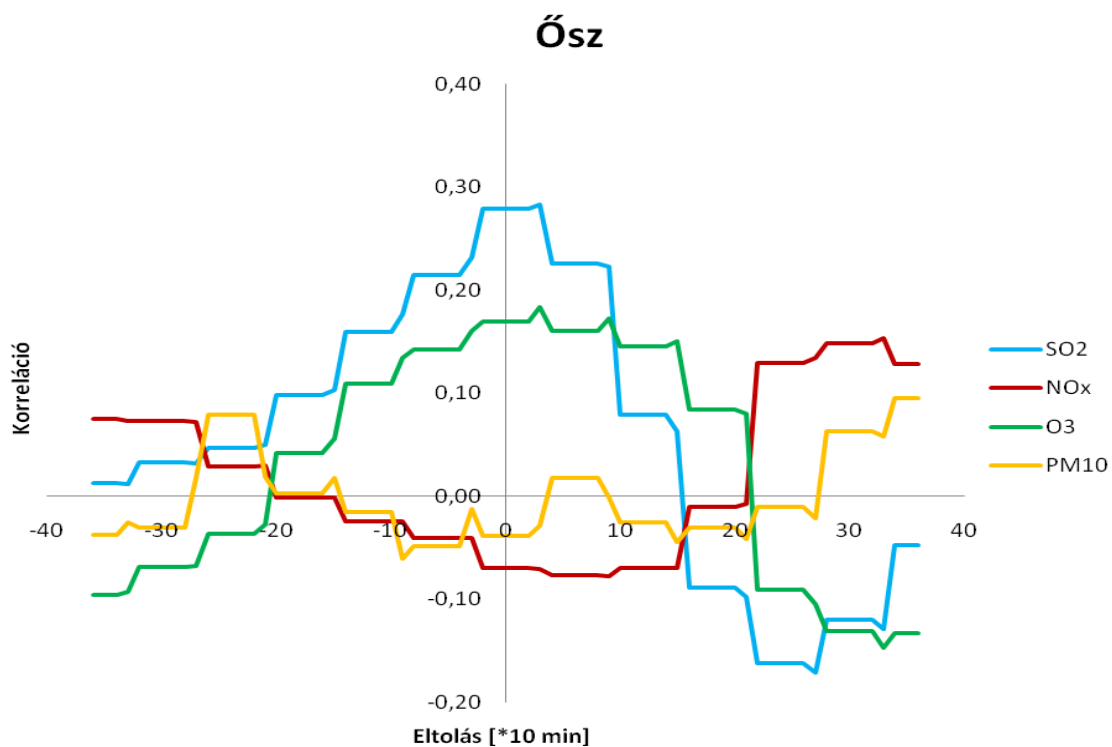
A nukleáció kialakulása szempontjából nyilván azok az adatsorok lesznek a meghatározóak, amelyek a nukleáció kezdete (0 időeltolás) előtti néhány órában magasabb, szignifikáns korrelációval rendelkeznek. A szignifikancia vizsgálatot T-próba segítségével végeztem el, amely alapján a 0,1-nél nagyobb korrelációs értékek a 0,01-os szignifikancia szint mellett szignifikánsnak adódtak. A grafikonok (5–8. ábra) alapján összességében két olyan adatsor látható, amely nagyobb korrelációval rendelkezik a folyamat előtt, mégpedig a napsugárzás és a hőmérséklet adatsora. Ezeken kívül nyáron és télen a szélsébség is nagyobb pozitív korrelációt mutat a többi adatsorhoz képest. Ami még szembetűnő az ábrák alapján, az a relatív nedvesség nukleációval szembeni negatív korrelációja, ami nem meglepő, hiszen a folyamatról való ismereteink alapján tudjuk, hogy a nagyobb nedvességtartalom erőteljes gátló tényező lehet. A szélirány, csapadék, és légnyomás értékeiből viszont nem látszik értékelhető korrelációs kapcsolat a nukleációval, így valószínűleg ezek nem, illetve csak egyedi esetekben befolyásolhatják a nukleáció kialakulását. Ilyen egyedi eset lehet, például a csapadéktevékenység, továbbá a légnyomás vagy szélirány értékekben bekövetkező hirtelen, erőteljes változás.



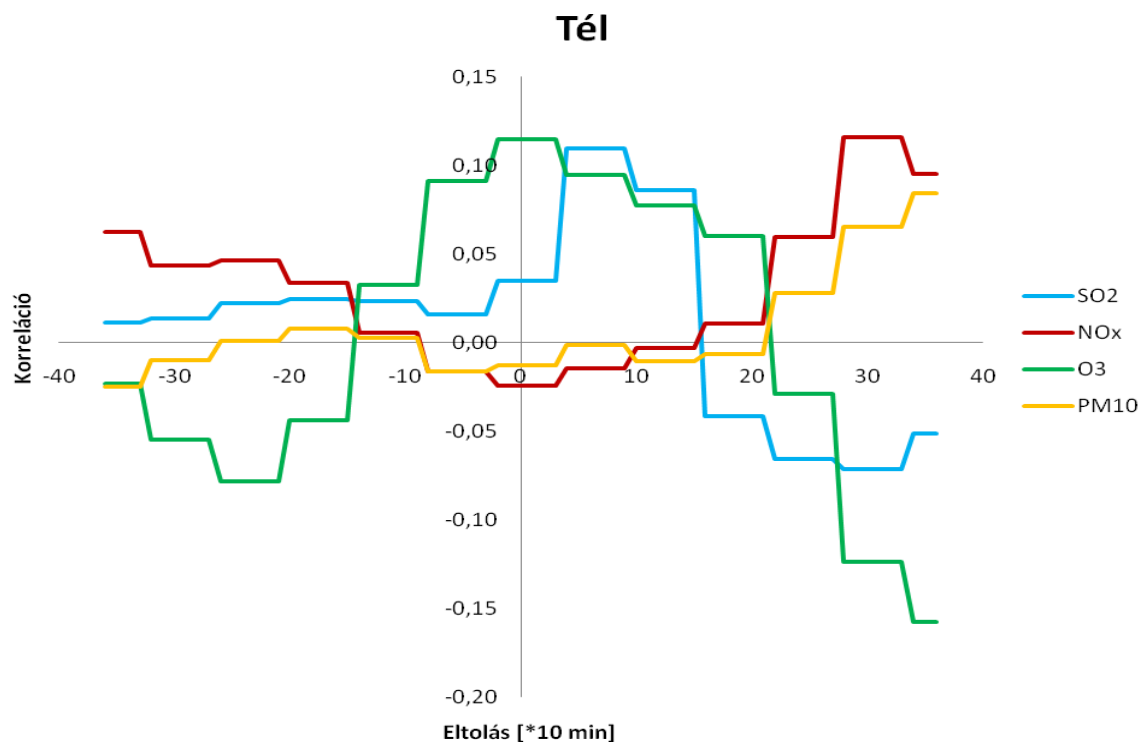
9. ábra A SO₂, NO_x, O₃ és PM₁₀ adatsorok nukleációval való korrelációja az időeltolás függvényében ábrázolva, tavasszal.



10. ábra A SO₂, NO_x, O₃ és PM₁₀ adatsorok nukleációval való korrelációja az időeltolás függvényében ábrázolva, nyáron.

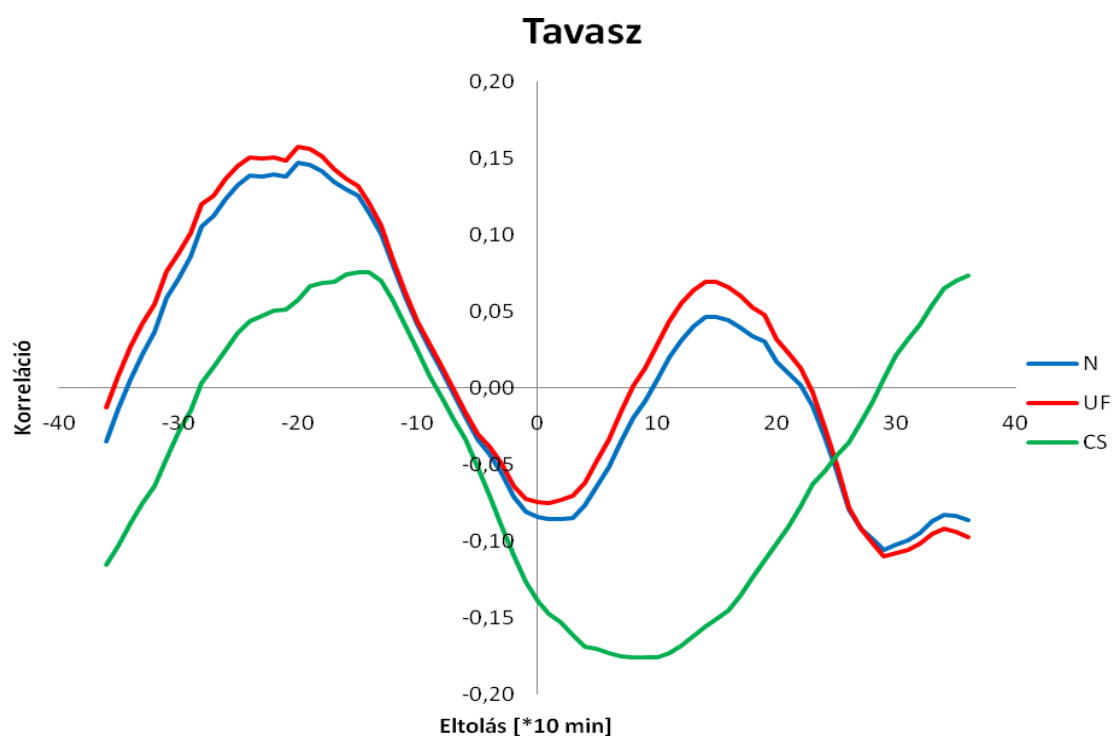


11. ábra A SO₂, NO_x, O₃ és PM₁₀ adatsorok nukleációval való korrelációja az időeltolás függvényében ábrázolva, ősszel.

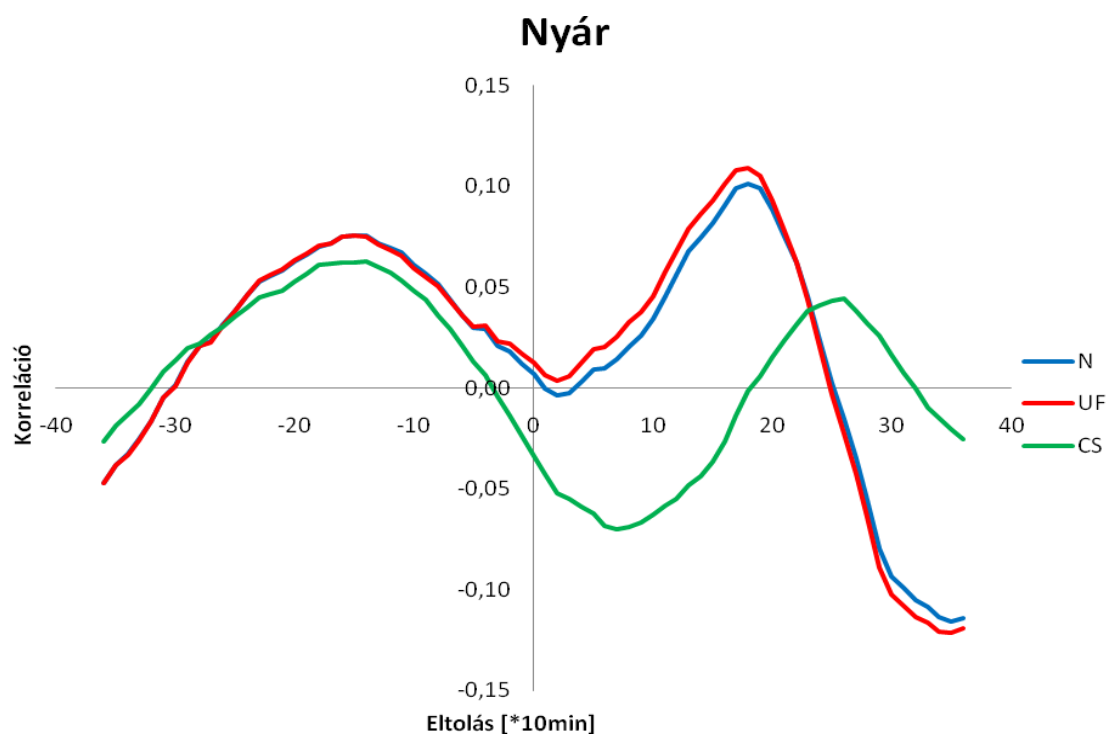


12. ábra A SO₂, NO_x, O₃ és PM₁₀ adatsorok nukleációval való korrelációja az időeltolás függvényében ábrázolva, télen.

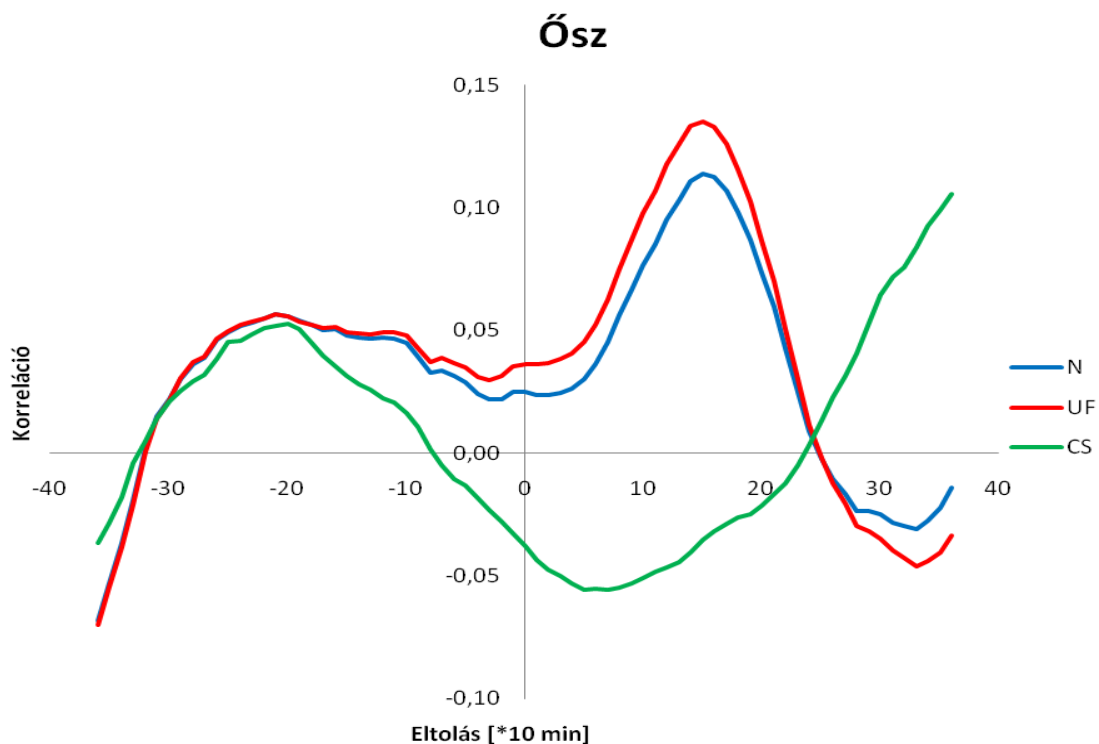
A fenti négy grafikonhoz (9–12. ábra) először is fontos megjegyezni, hogy azért rendelkeznek ilyen lépcsőzetes menettel, mert a légszennyező adatsorok esetében csak órás átlagértékek álltak rendelkezésre. Ezen kívül meg kell még említeni azt is, hogy az egyes korrelációs értékek a 0,01-os szignifikancia szinten a 0,15-os korrelációs érték felett voltak szignifikánsak. Az ábrák alapján összességében az látszik, hogy mind a négy évszakban a SO₂ és az O₃ koncentráció értékei korreláltak leginkább a nukleáció kialakulásának folyamatával. A másik két adatsort vizsgálva, elmondható, hogy a PM₁₀ koncentráció értékei egyáltalán nem mutatnak korrelációs kapcsolatot, viszont a NO_x koncentráció értékei esetében tavasszal és nyáron a többi adatsorhoz képest számottevő negatív korreláció van.



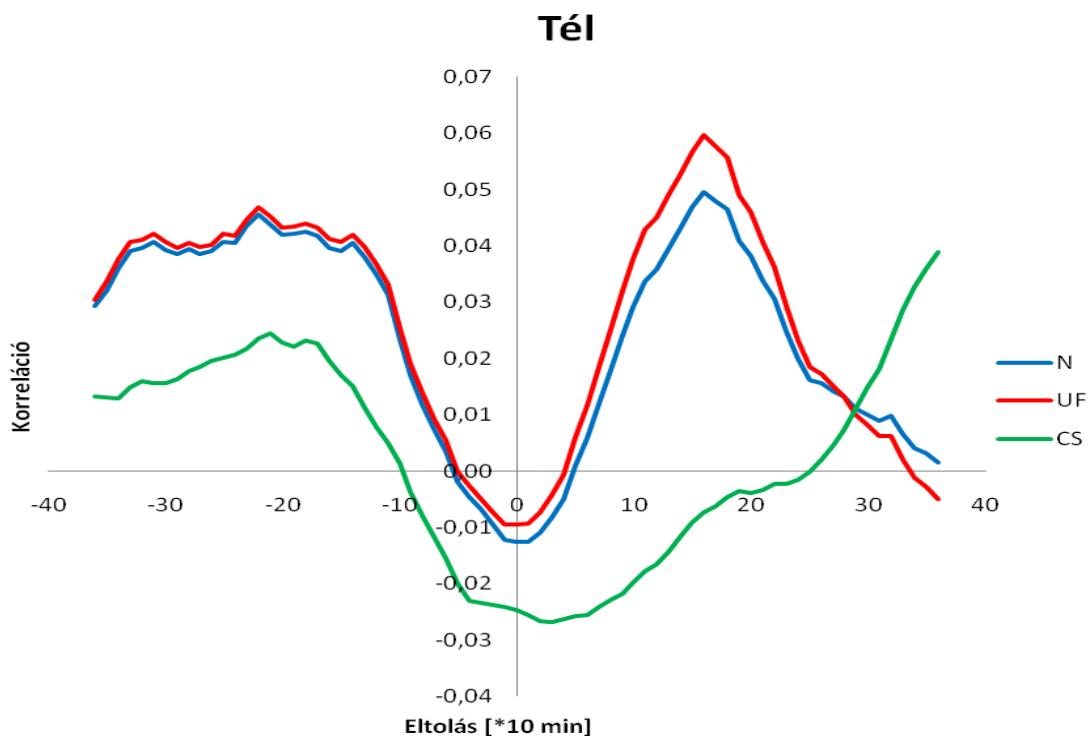
13. ábra Az ultrafinom (UF), és az összes (N) részecskére vonatkozó részecskeszám koncentráció, illetve a kondenzációs nyelő (CS) értékeinek a nukleációval való korrelációja az időeltolás mértékének függvényében, tavasszal.



14. ábra Az ultrafinom (UF), és az összes (N) részecskére vonatkozó részecskeszám koncentráció, illetve a kondenzációs nyelő (CS) értékeinek a nukleációval való korrelációja az időeltolás mértékének függvényében, nyáron.



15. ábra Az ultrafinom (UF), és az összes (N) részecskére vonatkozó részecskeszám koncentráció, illetve a kondenzációs nyelő (CS) értékeinek a nukleációval való korrelációja az időeltolás mértékének függvényében, őszel.



16. ábra Az ultrafinom (UF), és az összes (N) részecskére vonatkozó részecskeszám koncentráció, illetve a kondenzációs nyelő (CS) értékeinek a nukleációval való korrelációja az időeltolás mértékének függvényében, télen.

A kondenzációs nyelő és a részecskeszám koncentrációk korrelációs együtthatói a 0,01-os szignifikancia szinten a 0,05-os korrelációs érték felett adódtak szignifikánsnak, azonban ezen grafikonok elemzésekor nem a korrelációs kapcsolat erősségét, hanem az egyes görbék menetének alakulását hasonlítottam össze. A részecskeszám koncentráció értékek (N és UF) menetében (13–16. ábra) mind a négy évszakban jól látható, hogy a nulladik eltolási lépésnél egy mélypont, míg a negatív, illetve a pozitív eltolási tartományban egy-egy csúcs jelenik meg. Ez jól szemlélteti a nukleáció folyamatának intervallumát. A kondenzációs nyelő értékei pedig egyfajta fáziskülönbséggel jelennek meg a részecskeszám koncentrációkhoz képest, ez különösen a pozitív eltolási tartományban szembetűnő. Ugyanis, amikor a kondenzációs nyelő értéke elkezd növekedni, akkor a részecskeszám értékek elkezdenek csökkenni, ez érthető, hiszen a kondenzációs nyelő folyamata éppen a részecskék számának csökkenésével jár.

Itt jegyezném meg, hogy a kondenzációs nyelő folyamatok értéke tulajdonképpen a nukleációs esemény kezdetekor a levegőben meglévő aeroszol részecskék mennyiségére vonatkozik. A már meglévő részecskék ugyanis koagulációval képesek elvonni a növekedésnek induló, új aeroszol részecskéket, gátolva ezzel képződési folyamatukat.

3.2. Logisztikus regresszió

A fent bemutatott grafikonok alapján képet kaptunk arról, hogy mely állapothatározók fontosak a nukleáció kialakulása során. Ezek után kísérletet tettünk arra, hogy megpróbáljuk „előrejelezni”, hogy mikor számíthatunk nukleációra, azaz új aeroszol részecskék kialakulására. Ennek kiszámítására a logisztikus regresszió módszerét találtuk a legalkalmasabbnak. A folyamat kialakulása szempontjából fontosabb állapothatározókra elvégeztük a logisztikus regressziót az SPSS programcsomag segítségével. A számítást évszakonként, a nukleáció időtartamán belül óránként, külön-külön hajtottam végre, melyhez minden napra az előző 60 percet jellemző normált-átlagokat tartalmazó adatbázis állt rendelkezésemre. A program először is kiszámolta a logisztikus regresszió egyenletében vizsgált változókhoz tartozó β együtthatók értékeit, melyek egy példája a 6. táblázatban látható. Ezen együtthatókat minden órára kiszámoltam a vizsgált időtartamon belül, majd behelyettesítettem őket az alábbi egyenletbe.

$$\pi_x = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k}} \quad (21)$$

6. táblázat SPSS output a vizsgált állapothatározók β értékeire télen 10 UTC esetén

Az egyenlet változói

	B	S.E.	Wald	df	Sig.	Exp(B)
1. lépés ^a <i>T</i>	−1,487	1,682	0,782	1	0,377	0,226
<i>Rh</i>	−0,503	0,323	2,429	1	0,119	0,605
<i>Ws</i>	−0,407	0,664	0,376	1	0,540	0,666
<i>Sr</i>	−0,013	0,010	1,481	1	0,224	0,987
<i>P</i>	−1,618	1,285	1,586	1	0,208	0,198
<i>N</i>	0,006	0,005	1,569	1	0,210	1,006
<i>UF</i>	−0,007	0,005	1,712	1	0,191	0,993
<i>CS</i>	−484,848	703,543	0,475	1	0,491	0,000
<i>SO₂</i>	−0,255	0,284	0,801	1	0,371	0,775
<i>O₃</i>	0,208	0,099	4,440	1	0,035	1,231
<i>Constant</i>	−6,244	3,722	2,815	1	0,093	0,002

a: az első lépés során vizsgált változók: *T*, *Rh*, *Ws*, *Sr*, *P*, *N*, *UF*, *CS*, *SO₂*, *O₃*.

Eredményül az adott évszakra és a vizsgált órára vonatkozóan minden napra kaptam egy-egy valószínűségi értéket, amely azt mondta meg, hogy a logisztikus regresszió mekkora valószínűséget adott arra vonatkozóan, hogy az adott napon volt-e nukleáció. A számítást általában 7–11 UTC értékekre (tél esetén 8–11 UTC értékekre) hajtottam végre. Így minden napra 4 vagy 5 valószínűségi érték alapján, egy szubjektíven meghatározott kritikus valószínűségi értékhez viszonyítva hoztam döntést arról, hogy volt-e nukleáció, vagy sem. A döntési eljárás során először megvizsgáltam, hogy az adott napon a vizsgált órákra kapott valószínűségi értékek hányszor lépték túl a szubjektív kritikus értéket. Az így kapott értékeket ezután összevetettem az egyes órákra kapott valószínűségek összegével, méghozzá egy olyan kritériumrendszer segítségével, amely kizárta annak lehetőségét, hogy az esetlegesen előforduló, kritikus értéknél kisebb valószínűségek egy-egy újabb, de hibásan meghatározott esetet adjanak ki. A vizsgálatot objektív módon is elvégeztem, ekkor a valószínűségi értékekre lefutattam egy, a lineáris szeparálás módszerét alkalmazó Fortran programot, mely általam előre beállított egész számokat adott eredményül arra vonatkozóan, hogy volt-e nukleáció vagy sem. Ezeket szintén összehasonlítottam a valós megfigyelésekkel.

A módszerek előrejelzési hatékonyságát 2 x 2-es kontingencia táblázatok segítségével vizsgáltam. Ezen táblázatok értékei azt mutatják meg, hogy az előrejelzésünk hányszor találta el, illetve hányszor becsülte felül, vagy alul a nukleáció folyamatát. A következőkben bemutatom és elemzem az egyes évszakokra kapott kontingencia táblázatokat.

Tavaszi:

7. táblázat Kontingencia táblázat 0,15 kritikus valószínűségi érték esetén, tavasszal

Előrejelzés	Megfigyelés		
	Igen	Nem	Σ
Igen	12	9	21
Nem	2	51	53
Σ	14	60	74

8. táblázat Kontingencia táblázat 0,3 kritikus valószínűségi érték esetén, tavasszal

Előrejelzés	Megfigyelés		
	Igen	Nem	Σ
Igen	7	6	13
Nem	2	59	61
Σ	9	65	74

9. táblázat Kontingencia táblázat lineáris szeparálással számított értékek alapján, tavasszal

Előrejelzés	Megfigyelés		
	Igen	Nem	Σ
Igen	9	14	23
Nem	7	44	51
Σ	16	58	74

A találati arányokat (GSS) vizsgálva a legjobb eredményt a 0,15 kritikus valószínűségi érték megválasztása esetén kaptuk. Látható, hogy a számítás mind a három esetben (7–9. táblázat) inkább felülbecsülte a folyamatot. A legjobb előrejelzési értékeket a megfigyelt 16 nukleációs és 58 nemnukleációs esetre a 7. táblázat esetén kaptuk.

Nyár:

10. táblázat Kontingencia táblázat 0,15 kritikus valószínűségi érték esetén, nyáron

Előrejelzés	Megfigyelés		
	Igen	Nem	Σ
Igen	4	8	12
Nem	5	52	57
Σ	9	60	69

11. táblázat Kontingencia táblázat 0,3 kritikus valószínűségi érték esetén, nyáron

Előrejelzés	Megfigyelés		
	Igen	Nem	Σ
Igen	4	6	10
Nem	1	58	59
Σ	5	64	69

12. táblázat Kontingencia táblázat lineáris szeparálással számított értékek alapján, nyáron

Előrejelzés	Megfigyelés		
	Igen	Nem	Σ
Igen	1	8	9
Nem	9	51	60
Σ	10	59	69

Az eredmények alapján (10–12. táblázat) elmondható, hogy a 0,15 kritikus érték és a lineáris szeparálás módszere egyformán rosszul becsülte alul, illetve felül az eseteket. Ebben a nyári időszakban a vizsgált 69 esetből 10-ben volt nukleáció, ebből mind a 0,15, mind a 0,3 kritikus értékek esetén 4-4 esetet találtunk el, azonban a találati arány a 11. táblázatban volt nagyobb.

Ösz:

13. táblázat Kontingencia táblázat 0,15 kritikus valószínűségi érték esetén, összesen

Előrejelzés	Megfigyelés		
	Igen	Nem	Σ
Igen	7	7	14
Nem	0	58	58
Σ	7	65	72

14. táblázat Kontingencia táblázat 0,3 kritikus valószínűségi érték esetén, összesen

Előrejelzés	Megfigyelés		
	Igen	Nem	Σ
Igen	5	3	8
Nem	0	64	64
Σ	5	67	72

15. táblázat Kontingencia táblázat lineáris szeparálással számított értékek alapján, összesen

Előrejelzés	Megfigyelés		
	Igen	Nem	Σ
Igen	3	7	10
Nem	6	56	62
Σ	9	63	72

Az ösz esetében is a lineáris szeparálás módszere (15. táblázat) bizonyult a legrosszabbnak az alul-, illetve felülbecslések tekintetében, viszont látható, hogy a másik két táblázatban (13. és 14. táblázat) csak felülbecslések szerepeltek. A legjobb találati arány a 14. táblázat értékei alapján adódott az összesen megfigyelt 9 nukleációs és 63 nemnukleációs esetre.

Tél:

16. táblázat Kontingencia táblázat 0,15 kritikus valószínűségi érték esetén, télen

Előrejelzés	Megfigyelés		
	Igen	Nem	Σ
Igen	5	0	5
Nem	0	76	76
Σ	5	76	81

17. táblázat Kontingencia táblázat 0,3 kritikus valószínűségi érték esetén, télen

Előrejelzés	Megfigyelés		
	Igen	Nem	Σ
Igen	4	0	4
Nem	0	77	77
Σ	4	77	81

18. táblázat Kontingencia táblázat lineáris szeparálással számított értékek alapján, télen

Előrejelzés	Megfigyelés		
	Igen	Nem	Σ
Igen	5	0	5
Nem	0	76	76
Σ	5	76	81

Látható, hogy a vizsgált téli időszakban mind a három módszer (16–18. táblázat) teljes mértékben eltalálta a megfigyelt nukleációs eseteket. Ez a kiváló eredmény valószínűleg azzal magyarázható, hogy télen igen kevés, mindössze 5 nukleációs esetet regisztráltak, és ezek jól elkülönültek a többi 76 nemnukleációs típustól. A 17. táblázat esetében azért csak 4 nukleációs típust találtunk, mivel a szubjektív kritériumok alapján egy esetet a nemnukleációs típusba soroltunk.

3.3. Diszkriminancia analízis

A logisztikus regresszió analízis mellett egy ehhez nagyon hasonló többváltozós statisztikai módszer, a diszkriminancia analízis segítségével is elvégeztem az egyes típusok, a nukleációs és a nemnukleációs napok, szétválasztását. Az analízist az SPSS programcsomag segítségével hajtottam végre. A program minden évszak adott vizsgált órájára meghatározta, hogy volt-e nukleáció, vagy sem. Még hozzá, úgy, hogy minden egyes beadott paraméterhez egy-egy diszkrimináló együtthatót rendelt, amely alapján eldöntöttem, hogy mely változók a fontosabbak leginkább az elkülönítés szempontjából. Ez alapján az egyes évszakokra egy kicsit más paraméter kombinációk adták a legjobb találati arányt, de a nukleáció kialakulása szempontjából fontosnak tartott, meghatározó paraméterek (*T*, *Rh*, *Ws*, *Sr*, *N*, *SO₂*, *Proxy*) állandók voltak. Tavasszal és nyáron ez kiegészült a légnyomás és az *O₃* koncentráció, ősszel pedig a légnyomás és a *NO_x* koncentráció értékével. Télen viszont a légnyomás nem egy meghatározó paraméter, sokkal fontosabbak a légszennyező anyagok, ezért ekkor az előbbi 7 paraméterhez az *NO_x* és a *PM₁₀* koncentráció értéke társult. Összességében tehát minden évszakban 9 meghatározó állapotjelzőt találtam az analízis során.

19. táblázat Az SPSS programcsomag által tavasz 9 UTC-re generált kontingencia táblázat

		Nukleáció	Előrejelzés		Összes
			0,00	1,00	
Megfigyelés	db	0,00	46	0	46
		1,00	6	3	9
	%	0,00	100,0	0,0	100,0
		1,00	66,7	33,3	100,0

Az 19. táblázatban látható egy példa arra, hogy az SPSS programcsomag a tavasz 9 UTC-s időpontban mennyi nukleációs és nemnukleációs típust különített el, és ennek során hányszor becsülte alul, illetve felül a folyamatot. A táblázat mind darabszámban, mind pedig százalékos arányban megjeleníti az egyes találatokat. Az analízist tavasszal és nyáron 7–11 UTC, míg ősszel és télen, a később beinduló nukleációs esemény miatt, 8–11 UTC között minden egyes órára elvégeztem. A program tehát minden napra az adott vizsgált órában hozott egy döntést arra vonatkozóan, hogy volt-e nukleáció vagy sem. Ezeket az eredményeket a vizsgált intervallumokra egy Excel programban összesítettem,

még hozzá, úgy, hogy azt a napot soroltam a nukleációs típusba, amelyen a program legalább kétszer nukleációt jelzett. Végül az eredményeket kontingencia táblázatok segítségével értékeltem ki. A következőkben bemutatom az egyes évszakokra a diszkriminancia analízis alapján kapott eredményeket.

Tavaszi:

20. táblázat Kontingencia táblázat tavasszal diszkriminancia analízis esetén

Előrejelzés	Megfigyelés		
	Igen	Nem	Σ
Igen	7	2	9
Nem	7	39	46
Σ	14	41	55

Nyár:

21. táblázat Kontingencia táblázat nyáron diszkriminancia analízis esetén

Előrejelzés	Megfigyelés		
	Igen	Nem	Σ
Igen	5	1	6
Nem	5	50	55
Σ	10	51	61

Ősz:

22. táblázat Kontingencia táblázat ősszel diszkriminancia analízis esetén

Előrejelzés	Megfigyelés		
	Igen	Nem	Σ
Igen	2	0	2
Nem	3	56	59
Σ	5	56	61

Tél:

23. táblázat Kontingencia táblázat télen diszkriminancia analízis esetén

Előrejelzés	Megfigyelés		
	Igen	Nem	Σ
Igen	3	0	3
Nem	2	66	68
Σ	5	66	71

Az eredmények alapján (20–23. táblázat) megállapítható, hogy a módszer általában megközelítőleg ugyanannyiszor becsülte alul a folyamatot, mint ahányszor helyesen elkülönítette a nukleációs típusokat, a felülbecslések száma pedig elenyésző volt. Tehát a módszer elkülönítési hatékonyság szempontjából kedvezőtlenebbnek bizonyult, mint ahogy azt vártuk.

3.4. Verifikáció

24. táblázat A logisztikus regresszió és a diszkriminancia analízis módszere alapján kapott találati arányok (GSS), $(a/(a+b+c))$ az egyes évszakok esetében

	logisztikus regresszió	diszkriminancia analízis
tavas	0,52	0,44
nyár	0,44	0,45
ősz	0,40	0,40
tél	1,00	0,60

Összehasonlítva a logisztikus regresszió, illetve a diszkriminancia analízis módszere alapján kapott találati arányokat (24. táblázat), megállapítható, hogy azok nyáron és ősszel közel azonosak, míg tavasszal, és főként télen a logisztikus regresszió esetén adnak kedvezőbb értékeket. Ha csak azt vesszük figyelembe, külön-külön vizsgálva a két eljárást, hogy melyik módszer különített el helyesen több nukleációs napot, akkor is a logisztikus regresszió mondható a jobb módszernek. A két metódus lényegében hasonlóan, a fontosabb paramétereket súlyfaktorokkal ellátva próbálta a nukleációs és a nemnukleációs napokat elkülöníteni egymástól. A kapott eredmények közötti eltérés a vizsgált esetek számának különbségéből is adódik. Ugyanis a diszkriminancia analízis csak abban az esetben tudott döntést hozni, ha az adott nap adott órájában az általam kiválasztott diszkrimináló paraméterek mindegyikére vonatkozóan szerepelt érvényes érték, azaz nem volt hiányzó adat. Amennyiben valamelyik paraméter értéke mégis hiányzott, a módszer nem tudta ezt a napot egyik típusba sem besorolni. A logisztikus regresszió analízis ezzel szemben még egy-két hiányzó érték esetén is képes volt döntést hozni, az egyes típusokról, így összességében több esetet tudott elkülöníteni. Összességében tehát a logisztikus regresszió módszerének segítségével sikerült a legjobban a nukleációs, illetve a nemnukleációs napok elkülönítése.

A dolgozatomban említett, és az irodalmi áttekintésben leírt harmadik módszer a klaszterelemzés volt. Ezen eljárás több szakirodalomban is szerepelt (*Brines et al., 2014; Dall'Osto et al., 2012*), amelyekben a módszer segítségével azt vizsgálták, hogy a különböző átmérőjű aeroszol részecskék milyen meteorológiai paraméterek mellett voltak jelen a levegőben. Mivel a részecskék méreteloszlásáról nem rendelkezttem információval, az egyes típusok elkülönítésére pedig nem találtam megfelelőnek, így a klaszteranalízist végül nem alkalmaztam a vizsgálataim során. A jövőben, a további kutatások alkalmával azonban újra felmerülhet a módszer alkalmazhatóságának kérdése.

4. ÖSSZEFOGLALÁS

A légköri nukleáció, azaz az új aeroszol részecskék keletkezésének folyamata, komoly egészségügyi és éghajlati következményekkel járhat. Szükséges tehát a mechanizmus hátterének minél pontosabb megismerése. Munkám a légköri nukleációs események kutatásához kapcsolódott, melynek során azt vizsgáltam, hogy az egyes meteorológiai, illetve légszennyező paraméterek hogyan, milyen mértékben vannak hatással a nukleáció kialakulására.

A korábbi analízis során megállapítottam, hogy a napi átlagos értékek, nem jellemezték megfelelően a nukleáció időtartamát, túlzottan elsimították az adatsorokat. Ezért szükségesnek látszott az adatsorok felbontásának növelése, illetve csak azon adatok vizsgálata, amelyek beleesnek a nukleáció intervallumába. Elsőként felépítettem egy új adatbázist, mely a kulcsfontosságú légszennyezők órás átlagértékei mellett, a meteorológiai változók, a részecske-koncentrációk, illetve a kondenzációs nyelő értékeit már 10 perces felbontásban tartalmazta. A nukleáció időtartamát 6 UTC és 12 UTC között határoztam meg, mivel az esetek döntő többségében ezen időszak alatt ment végbe a folyamat. Mivel az egyes magyarázó változók értékei nem csak ebben a 6 órás intervallumban lehetnek érdekesek, hanem az is fontos, hogy az időszak előtt, illetve utána hogyan alakult az értékük, ezért végrehajtottam egy eltolást az adatsorok és a nukleáció időszaka között. Az így készített évszakonkénti grafikonokból leolvasható, hogy melyik paraméternek volt nagyobb hatása a nukleációra, s ez a nukleáció előtt mennyi idővel következett be. E grafikonok alapján kerültek kiválasztásra azok az adatsorok, amelyek a legfontosabbak voltak a nukleációs esemény kialakulása során. A grafikonok alapján tehát képet kaptam arról, hogy a vizsgált paraméterek közül melyek voltak a leginkább meghatározó környezeti változók (Sr , SO_2 , O_3 , CS , Rh) az új aeroszol részecskék keletkezése szempontjából.

A következőkben arra kerestem a választ, hogy mely paraméterek esetén várható nukleáció. Ehhez létrehoztam egy olyan adatbázist, amely a vizsgált intervallum minden egyes órájára az előző 60 percet jellemző normált-átlagokat tartalmazza. A nukleáció intervallumát így már jobban leíró adatbázis statisztikai elemzésére több módszer is felmerült; klaszteranalízis, diszkriminancia elemzés, logisztikus regresszió, de ezek közül első lépésként a logisztikus regresszió tűnt a legcélravezetőbbnek. A logisztikus regresszió által meghatározott, a nukleáció folyamatának bekövetkezését jellemző valószínűségi

értékeket összevettem a tényleges megfigyelésekkel, majd az eredményeket 2 x 2-es kontingencia táblázatok segítségével elemeztem. A kontingencia táblázatok alapján megállapítható, hogy a szubjektív módszer mind a négy évszakban jobbnak bizonyult elkülönítés szempontjából, mint a lineáris szeparálás objektív módszere. Tehát a szubjektív módszer esetében sikerült úgy megállapítani a kritériumokat, hogy jobb találati arányokat kaptunk. A vizsgált évszakok közül kiemelkedett a tél 100%-os találati aránya. Ez valószínűleg abból adódott, hogy ebben az időszakban mindössze 5 nukleációs esetet (napot) figyeltek meg a vizsgált 81-ből, és e napok, állapotathatózóik alapján, jól elkülönültek a nemnukleációs napoktól. Az egyes nukleációs és nemnukleációs típusok szétválasztását a logisztikus regresszióhoz hasonló statisztikai módszer, a diszkriminancia analízis segítségével is elvégeztem. A két módszer által kapott eredményeket összehasonlítva, azonban a logisztikus regresszió bizonyult hatékonyabbnak az elkülönítés szempontjából.

Összefoglalva, megállapítottam, hogy mely paraméterek a legfontosabbak a nukleáció kialakulásához, és ezek a folyamat kezdetéhez képest mikor jelentős hatásúak. Ezen kívül létrehoztam egy módszert az egyes nukleációs, illetve nemnukleációs napok elkülönítésére, a folyamat későbbiekben való előrejelezhetőségének vizsgálata érdekében. Fontos megjegyezni azonban, hogy ez a szétválasztási módszer még csak a korábbi mérésekre (1 éves adatsor) lett tesztelve, de a tervek között szerepel a módszer finomítása, hosszabb időszakokra történő kiterjesztése is, hiszen a mérések az ELTE Lágymányosi épület mellett 2013-ban létesített BpART mérőplatformon jelenleg is folynak.

5. IRODALOMJEGYZÉK

- Benson, D. R., Yu, J. H., Markovich, A., Lee, S. H., 2011: Ternary homogeneous nucleation of H_2SO_4 , NH_3 , and H_2O under conditions relevant to the lower troposphere. – *Atmospheric Chemistry and Physics*. 11. pp. 4755–4766.
- Borsós, T., Salma, I., 2009: Nanorészecskék a levegőben. Áldás vagy átok? – *Élet és Tudomány*. 51–52. pp. 1654–1656.
- Borsós, T., Salma, I., 2010: Budapesti banángörbe. – *Élet és Tudomány*. 21. pp. 643.
- Brines, M., Dall'Osto, M., Beddows, D. C. S., Harrison, R. M., Querol, X., 2014: Simplifying aerosol size distributions modes simultaneously detected at four monitoring sites during SAPUSS – *Atmospheric Chemistry and Physics*. 14. pp. 2973–2986.
- Burns, R., Burns, R., 2008: Business research methods and statistics using SPSS. – *SAGE Publications Ltd*, London. 560 p.
- Chin, M., Kahn, A. R., Remer, A. L., Yu, H., Rind, D., Feingold, G., Quinn, K. P., Schwartz, E. S., Streets, G. D., DeCola, P., Halthore, R., 2009: Atmospheric aerosol properties and climate impact. – *U.S. Climate Change Science Program*, Washington, DC, USA. 115 p.
- Dall'Osto, M., Beddows, D. C. S., Pey, J., Rodriguez, S., Alastuey, A., Harrison, R. M., Querol, X., 2012: Urban aerosol size distributions over the Mediterranean city of Barcelona, NE Spain. – *Atmospheric Chemistry and Physics*. 12. pp. 10693–10707.
- Dévényi, D., Gulyás, O., 1984: Matematikai statisztikai módszerek a meteorológiában – *Egyetemi tankönyv*, Tankönyvkiadó, Budapest. 443 p.
- Füstös, L., 2009: A sokváltozós adatelemzés módszerei – *Módszertani füzetek I*, MTA Szociológiai Kutatóintézete 640 p.
- Hussein, T., Molgaard, B., Hannuniemi, H., Martikainen, J., Jarvi, L., Wegner, T., Ripamonti, G., Weber, S., Vesala, T., Hameri, K., 2014: Fingerprints of the urban particle number size distribution in Helsinki, Finland: Local versus regional characteristics – *Boreal Environment Research* 19. pp. 1–20.
- IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T. F. D., Qin, G. K., Plattner, M., Tignor, S. K., Allen, J., Boschung, A., Nauels, Y., Xia, V., Bex and P. M. Midgley (eds.)]. *Cambridge University Press*, Cambridge, United Kingdom and New York, NY, USA, 1535 pp.
- Kelemen, T., 2013: Nukleációs helyzetek meteorológiai háttérének statisztikai vizsgálata budapesti mérések alapján – BSc szakdolgozat, ELTE TTK, Budapest, (témavezető: Dr. Weidinger Tamás) 29 p.
- Merikanto, J., Spracklen, D. V., Mann, G. W., Pickering, S. J., Carslaw, K. S., 2009: Impact of nucleation on global CCN. – *Atmospheric Chemistry and Physics*. 9. pp. 8601–8616.
- Mikkonen, S., Lehtinen, K. E. J., Hamed, A., Joutsensaari, J., Facchini, M. C., Laaksonen, A., 2006: Using discriminant analysis as a nucleation event classification method – *Atmospheric Chemistry and Physics* 6. pp. 5549–5557.
- Morrison, G. D., 1969: On the interpretation of discriminant analysis – *Journal of Marketing Research* 2. pp. 156–163.
- Németh, Z., 2012: Légköri nukleáció városi környezetekben. – MSc szakdolgozat, ELTE TTK, Budapest, (témavezető: Dr. Salma Imre) 40 p.
- Németh, Z., Salma, I., 2014: Spatial extension of nucleating air masses in the Carpathian Basin – *Atmospheric Chemistry and Physics* 14. pp. 8841–8848.
- Nurmi, P., 2003: Recommendations on the verification of local weather forecasts. – *ECMWF Consultancy report* 19 p.
- Oberdörster, G., Oberdörster, E., Oberdörster, J., 2005: Nanotoxicology: an emerging discipline evolving from studies of ultrafine particles. – *Environ. Health Perspect.* 113. pp. 823–839.
- Pakkanen, T., Kerminen, V., Korhonen, C., Hillamo, R., Aarino, P., Koskentalo, T., Maenhaut, W., 2001: Urban and Rural Ultrafine ($\text{PM}_{0.1}$) Particles in the Helsinki Area. – *Atmospheric Environment*. 35. pp. 4593–4607.
- Politis, M., Pilinis, C., Lekkas, T. D., 2008: Ultrafine particles (UFP) and health effects. Dangerous. Like no other PM? Review and analysis. – *Global NEST Journal*. 3. pp. 439–452.
- Pöschl, U., 2005: Atmospheric Aerosols: Composition, transformation, climate and health effects. – *Angewandte Chemie*. 46. pp. 7485–7647.
- Ramanathan, V., Crutzen, P. J., Kiehl, J. T., Rosenfeld, D., 2001: Aerosols, climate, the hydrological cycle. – *Science*. 294. pp. 2119–2124.

- Salma, I., Borsós, T., Weidinger, T., Aalto, P., Hussein, T., Dal'Maso, M., Kulmala, M., 2011: Production, growth and properties of ultrafine atmospheric aerosol particles in an urban environment. – *Atmospheric Chemistry and Physics*. 11. pp. 1339–1353.
- Stephenson, B. D., 2000: Use of the „odds ratio” for diagnosing forecast skill – *Weather and Forecasting* 15. pp. 221–232.
- Székelyi, M., Barna, I., 2002: Túlélőkészlet az SPSS-hez. – *Typotex Kiadó*, Budapest. 453 p.
- Tryfos, P., 1998: Methods for Business Analysis and Forecasting: Text & Cases, *Wiley*, 592 p.

Internetes hivatkozások:

<http://salma.web.elte.hu/BpArt/> (utolsó letöltés 2014. 11. 28.)

6. KÖSZÖNETNYILVÁNÍTÁS

Szeretném megköszönni Dr. Weidinger Tamás egyetemi docensnek és Dr. Salma Imre egyetemi tanárnak a fáradhatatlan emberi és szakmai támogatást, Dr. Matyasovszky István egyetemi docensnek pedig az egyes statisztikai módszerek kiválasztásához és alkalmazásához nyújtott szakmai segítséget, mellyel ez a dolgozat elkészülhetett.

Köszönöm Családomnak a biztató és szerető támogatást.