Válasz Prof. Dr. Szántai Tamás opponensi véleményére

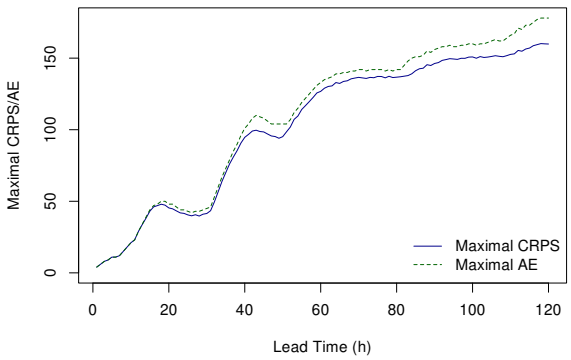
Először is szeretném megköszönni, hogy Szántai Tamás Professzor Úr elvállalta a dolgozatom véleményezését, azt tüzetesen végigolvasta és bőven a hivatalos határidő lejárta előtt elkészítette bírálatát. Nagyon nagy örömet okoz, hogy munkámat pozitívan értékelte, eredményeimet elegendőnek tartotta az MTA doktori cím megszerzéséhez és javasolta a nyilvános védés kitűzését.

Az alábbiakban a Professzor Úr által feltett kérdésekre szeretnék válaszolni.

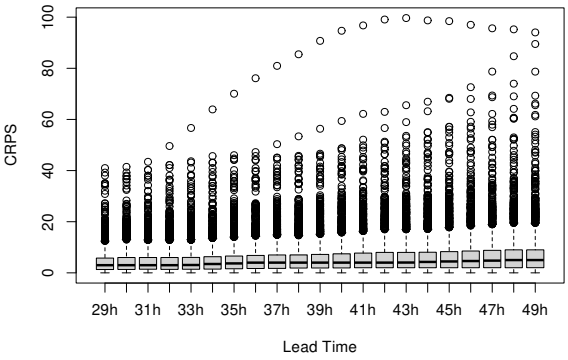
1. *Tud-e valamilyen indokot mondani arra, hogy a Box-Cox transzformációban az egyes időlépésekben használt λ paraméterek szépen csökkenő értékében a 42, és a 110 időlépések környékén miért található egy-egy ettől lényegesen eltérő ,,kipúposodás" (lásd 2.1. ábra)?*

   Annak érdekében, hogy az általunk javasolt BMA modellt minél korrektebbül össze tudjuk hasonlítani a Hemri and Klein (2017) által leírt EMOS modellel, mindkét módszert ugyanazon adathalmazon kellett vizsgálni. Emiatt az egyes időlépcsőkhöz tartozó λ Box-Cox transzformációs paramétereket egyszerűen átvettük a Hemri and Klein (2017) cikkből, azok külön vizsgálatával eddig nem foglalkoztunk. A szerzők a paramétereket úgy határozták meg, hogy a transzformált adatokra illesztett EMOS modelleknek az eredeti adatsorra vonatkozó előrejelző képességét optimalizálták.

   A Professzor Úr kérdése kapcsán azonban mi is megvizsgáltuk, van-e valami különös a vízállás ensemble előrejelzésekben, ami indokolná a 42 h és 110 h környéki kiugró viselkedést. Véleményünk szerint 42 h környékén van néhány kiemelkedően rossz előrejelzés. Ezt támasztja alá az alábbi ábra (a) része, ami a nyers előrejelzésekhez és a kapcsolódó megfigyelésekhez tartozó CRPS értékek, illetve az ensemble medián abszolút hibája időlépcsőnkénti maximumát mutatja, és erre enged következtetni a CRPS értékek dobozábráit mutató (b) ábra is. Sajnos 110 h környékén még ennyi kapcsolatot sem találtunk a nyers előrejelzések illeszkedése és a Box-Cox paraméterek viselkedése között.

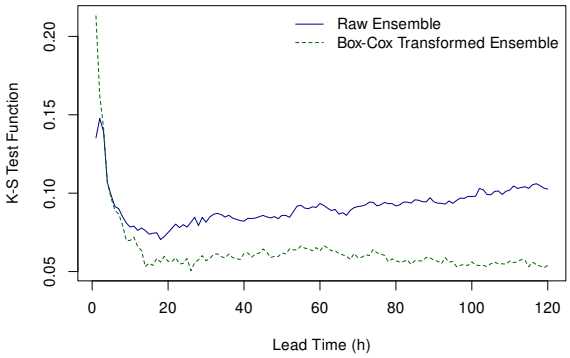


(a)



(b)

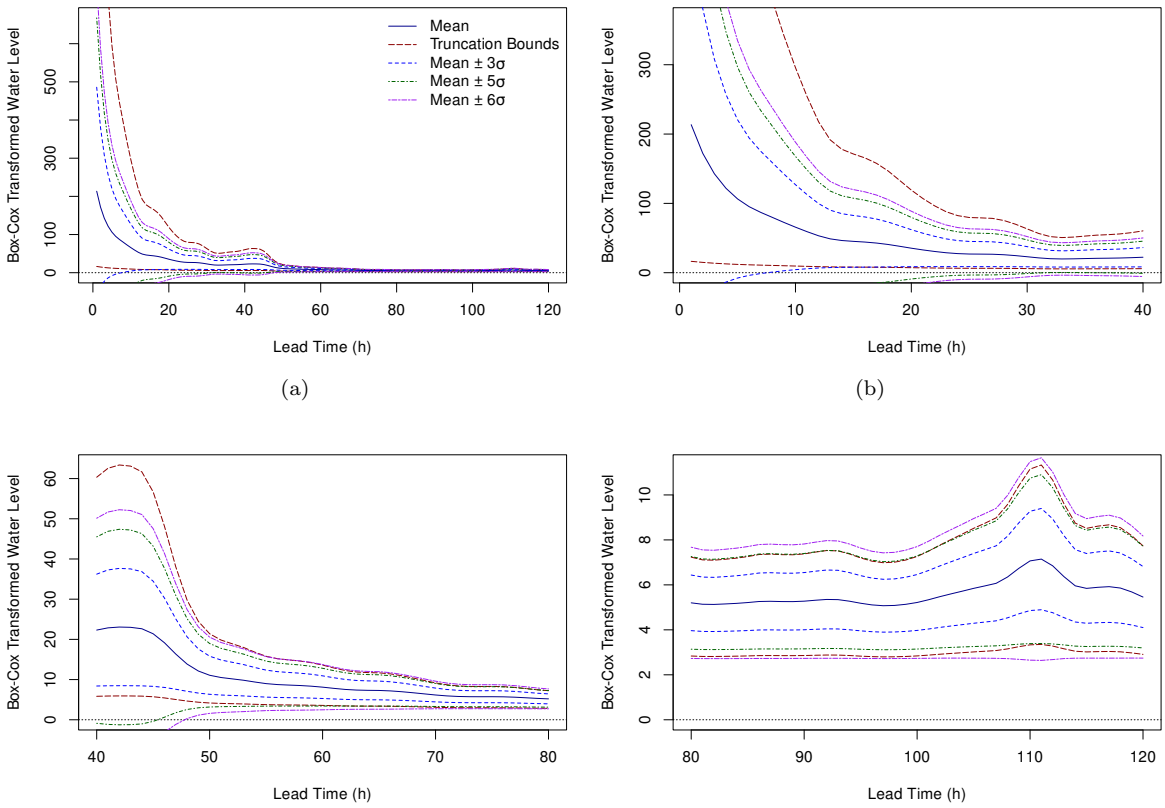
Megvizsgáltuk ezek után azt is, hogy a Box-Cox transzformáció önmagában mennyit javít az ensemble átlag megfigyeléstől való eltérésének a normális eloszláshoz való illeszkedésén. Az alábbi ábra a megfelelő Kolmogorov-Szmirnov próbák próbafüggvényét ábrázolja az előrejelzési horizont függvényében (minél kisebb, annál jobb). Látszik, hogy a javasolt Box-Cox paraméterek 20 h után egy meglehetősen sima görbét eredményeznek, tehát a 42 h és 110 h körüli anomáliák nem jelennek meg. Megjegyeznék még, hogy azért döntöttünk ezen egyszerű próba mellett, mert korábban ugyancsak hidrológiai előrejelzéseknél Hemri *et al.* (2015) is a Kolmogorov-Szmirnov teszt próbastatisztikáját minimalizálva kereste meg a Box-Cox transzformáció paramétereit.



Tisztában vagyunk vele, hogy a fenti indoklás messze nem tökéletes, de talán mentségünkre szolgál, hogy a Box-Cox paramétereket eredetileg meghatározó hidrológus szakemberrel (Stephan Hemri, MeteoSwiss) való konzultáció sem adott ettől meggyőzőbb eredményt.

2. *A 2.1. szakaszban mindkét oldalon csonkolt, normális eloszlással közelíti a vízállás adatok Box-Cox transzformáltjait, ahol az alsó csonkolás értéke az addig mért legkisebb vízállás felének; a felső csonkolás pedig az addig mért legnagyobb vízállás duplájának a transzformált értéke. Az esettanulmányokban ellenőrizte-e, hogy a csonkolási intervallum lefedi-e teljesen a várható érték plusz-mínusz mondjuk 3, 4, vagy esetleg 5-szörös szórás intervallumot? Mert ha igen, akkor a csonkolás gyakorlatilag hatástalan, egyszerűbb lenne magával a csonkolás nélküli eloszlással számolni.*

Teljesen jogos a Professzor Úr észrevétele. Igen, megvizsgáltuk a kérdést, bár erre külön nem tértünk ki sem az eredményeket bemutató Baran *et al.* (2019) cikkben, sem pedig az értekezésben. Az alábbi ábrán a Box-Cox transzformált megfigyelések átlaga, az csonkolási intervallum határainak transzformáltjai, valamint az átlag körüli 3, 5, illetve 6 szórás szélességű intervallumok láthatóak az előrejelzési horizont függvényében (az (a) ábra mutatja a teljes 1–120 h horizontot, a (b), (c) és (d) ábrák pedig rendre az 1–40 h, 40–80 h és 80-120 h szakaszokat). Látható, hogy az 5, illetve 6 szórás szélességű intervallumok már gyakran kilógnak csonkolási intervallumból, ami indokolja az csonkolt normális eloszlás használatát.



3. *A 3.1 szakasz elején azt írja, hogy a szélsebesség modellezésére tipikusan valamilyen nemnegatív értéket felvevő ferde eloszlást célszerű használni, mint amilyen például a Weibull, vagy a gamma eloszlás. Mi indokolja, hogy a 3.1.2 pontban mégis a BMA modellben a* $[0, \infty)$ *félegyenesre csonkolt normális eloszlás alkalmazását javasolja?*

A szélsebesség ensemble előrejelzések utófeldolgozására elsőként kifejlesztett BMA modellben az előrejelző eloszlás gamma eloszlások keveréke (Sloughter *et al.*, 2010). Itt azonban problémát jelenthet, hogy a hőmérséklet, illetve légnyomás modellezésére használható normális BMA modellel ellentétben a paraméterek becslésére szolgáló EM algoritmus minden egyes M (maximization) lépése egy-egy numerikus optimalizálást igényel, ami jelentősen lassítja az algoritmust. A csonkolt normális (TN) eloszláson alapuló BMA modellhez a motivációt egyrészt az adta, hogy ugyanez az eloszlás jól működött Thorarinsdottir and Gneiting (2010) szélsebesség utófeldolgozására javasolt ensemble model output statistics eljárásánál. A másik érv az EM algoritmus Lee and Scott (2012) által csonkolt, illetve cenzorált normális eloszlások keverékének paraméterbecslésére javasolt verziója volt, ami a normális keverékhez hasonlóan ugyancsak zárt formulákat használ mind az E (expectation), mind pedig az M lépésben. Ez nem mondható el más, ugyancsak szélsebesség modellezésre használt eloszlásokra, mint a Weibull, a log-normális, vagy az általánosított extrémérték, illetve annak frissen tesztelt (Baran *et al.*, 2021) nullában alulról csonkolt változata. Amint azt az értekezés 3.7 ábrája mutatja, amellett, hogy az eset-

tanulmányainkban a TN BMA modell előrejelző képessége szignifikánsan felülmúlta a gamma BMA modellét, az M lépés numerikus optimalizálásának kiváltása jelentős sebességnövekedést is eredményezett.

Tisztelettel,

.....

Dr. Baran Sándor
egyetemi docens

Debrecen, 2021. április 3.

Irodalom

- Baran, S., Hemri, S. and El Ayari, M. (2019) Statistical post-processing of water level forecasts using Bayesian model averaging with doubly-truncated normal components. *Water Resour. Res.* **55**, 3997–4013.
- Baran, S., Szokol, P. and Szabó, M. (2021) Truncated generalized extreme value distribution based EMOS model for calibration of wind speed ensemble forecasts. *Environmetrics*, doi:10.1002/env.2678.
- Hemri, S. and Klein, B. (2017) Analog based post-processing of navigation-related hydrological ensemble forecasts. *Water Resour. Res.* **53**, 9059–9077.
- Hemri, S., Lisniak, D. and Klein, B. (2015) Multivariate postprocessing techniques for probabilistic hydrological forecasting. *Water Resour. Res.* **51**, 7436–7451.
- Lee, G and Scott, C. (2012) EM algorithms for multivariate Gaussian mixture models with truncated and censored data. *Comput. Stat. Data Anal.* **56**, 2816–2829.
- Sloughter, J. M., Gneiting, T. and Raftery, A. E. (2010) Probabilistic wind speed forecasting using ensembles and Bayesian model averaging. *J. Am. Stat. Assoc.* **105**, 25–37.
- Thorarinsdottir, T. L. and Gneiting, T. (2010) Probabilistic forecasts of wind speed: ensemble model output statistics by using heteroscedastic censored regression. *J. R. Stat. Soc. Ser. A Stat. Soc.* **173**, 371–388.