

Egész értékű idősorok alkalmazásai

Applications of integer-valued time series

ISPÁNY Márton, Ph.D.

Debreceni Egyetem, Informatikai Kar
4028 Debrecen, Kassai út 26., Magyarország
ispany.marton@inf.unideb.hu; <https://inf.unideb.hu/>

Abstract

Periodically correlated and seasonal non-negative integer-valued autoregressive processes are investigated. A spectral criterion involving the model parameters is given for the existence and uniqueness. The structure of the mean and covariance functions of the periodically stationary distribution of the model is derived. Two infinite series representations for the process, the moving average, and the immigrant generation, are established. A real data application is presented for Pickup data.

Keywords: integer-valued autoregression; binomial thinning; periodicity; seasonality; Pickup data

Kivonat

Periodikusan korrelált és szezonális, nem negatív egészértékű autoregressziós folyamatokat vizsgálunk. A folyamat létezésére és egyértelműségére vonatkozóan megadunk egy, a modell paramétereit figyelembe vevő spektrális kritériumot. Levezetjük a modell periodikusan stacionárius eloszlása átlag- és kovarianciafüggvényeinek szerkezetét. Meghatározzuk a folyamat két végtelen sorral való előállítását: a mozgóátlagot és a bevándorlók generációin alapulót. Bemutatunk egy valós adatalkalmazást az átvételi (Pickup) adatokra vonatkozóan.

Kulcsszavak: egész értékű autoregresszió; binomiális ritkítás; periodikusság; szezonális; Pickup adatok

1. BEVEZETÉS

Egy diszkrét idejű sztochasztikus folyamatot $\{Y_t\}$, ahol t egész számot jelöl, periodikusan korrelálnak (PC) vagy tág értelemben vett periodikusan stacionáriusnak nevezünk, ha az átlagfüggvénye $\mu(t) := E(Y_t)$ és az autokovariancia-függvénye $R(s, t) := \text{cov}(Y_s, Y_t)$ minden s, t esetén létezik, és periodikus függvények $s \geq I$ egész periódussal. Ha $s = I$, akkor a folyamat gyengén stacionárius. A PC-folyamatok számos területen előfordulnak, például az orvostudományban, a hidrológiában, a klimatológiában és a légszennyezés területén. A legújabb eredményeket illetően lásd pl. Gardner és tsai [8] és Hurd és Míamee [10]. A PC-folyamatok modellezésének természetes módja az, ha megengedjük, hogy a stacionárius modellek paramétereit idővel periodikusan változzanak. Például az autoregressziós és mozgóátlag (ARMA) modellt kiterjesztették periodikus ARMA (PARMA) modellre a PC-folyamatok esetében. A PARMA-folyamatok létezését Pagano [17] és Vecchia [19] tanulmányozta. A PARMA modellekre vonatkozó statisztikai következtetés megtalálható Lund és Basawa [14], valamint Basawa és Lund [3] tanulmányaiban.

Egy gyengén stacionárius folyamatot szezonálisan korrelálnak nevezünk, ha a γ autokovariancia függvénye, amelyet a következőképpen definiálunk: $\gamma(h) := R(t + h, t)$ minden $h, t \in \mathbb{Z}$ esetén, szezonális mintázatot mutat. A gyengén stacionárius szezonális idősorok esetében a Box–Jenkins-módszertan részeként széles körben használt paraméteres modell a szezonális ARMA (SARMA) folyamatok osztálya, legyenek azok tisztán additív vagy multiplikatív folyamatok – lásd pl. Shumway és Stoffer [18] 3.9. szakaszát és az ott található példákat. A SARMA-modellekben történő következtetéssel többek között Chatfield és Prothero [3] foglalkozott. A gyakorlatban az idősorok mind szezonális, mind periodikus viselkedést mutathatnak, mivel periodikus autokovariancia szerkezetükben szezonális mintázatok is megjelenhetnek. A periodikusan változó paraméterekkel rendelkező szezonális ARMA-folyamatokat (SPARMA) Basawa és társai [4] vezették be. Ilyen idősorok gyakran fordulnak elő az ökonometriában, lásd a 4.2. szakaszt Koopman és társai [12]-ban, valamint

Hindrayanto és társai [9] 3.3. szakaszát, ahol SPARIMA-modelleket alkalmaztak az amerikai havi munkanélküliség idősorára. A valós értékű idősorok szezonálisának és periodicitásának modellezéséről további részletek a Lund [13] és Baek és tsai [3] munkákban találhatóak.

Az utóbbi időben egyre nagyobb az érdeklődés az olyan $\{Y_t\}$ számláló folyamatok iránt, ahol Y_t egyedek, tárgyak, vagy események t időpontban történő előfordulásainak számát jelöli; lásd pl. Davis és tsai [6] és Weiss [20]. A számláló folyamatok esetében a standard AR-modell analógja az egész értékű AR (INAR) modell. Ez a ritkító operátoron alapuló modell versenyképes alternatívának bizonyult más megközelítésekkel szemben, pl. a paraméter-vezérelt modellek, mint a látens Gauss-transzformációk, a rejtett Markov- és GLARMA-modellek; a megfigyelés-vezérelt modellek, mint a lineáris és log-lineáris Poisson-autoregresszió, valamint a kópula-alapú modellek. Az elsőrendű INAR (INAR(1)) modellt McKenzie [15] és Al-Osh és Alzaid [1] vezették be, egymástól függetlenül.

Filho és tsai [7]-ban bevezetésre került a PINAR(1,1_S) nevű, nem negatív egész értékű idősor-modell. A modellt később tovább vizsgáltuk [11]-ban. Tudomásom szerint ez volt az első olyan egyszerű egész-értékű idősor-modell, amely egyszerre mutat periodikus és szezonális autokorrelációt. A modellt alkalmaztuk antibiotikumot kapó légúti betegségben szenvedő emberek napi számának modellezésére a sürgősségi ellátásban a közegészségügyi rendszer számára. A PINAR(1,1_S) modell számos korábbi idősor modellnek az általánosítása. A cikkekben bizonyítottuk a PINAR(1,1_S) folyamat létezését és egyértelműségét amennyiben teljesül egy, a modellparamétereket figyelembe vevő, spektrális feltétel. Levezettük néhány tulajdonságát és megadtuk két végtelen sor által való előállítását: a bevándorló generációkon alapuló sorfejtést, amely egymástól független komponensekből áll; valamint a végtelen mozgóátlag kifejtést, amely korrelálatlan, de egymástól már függő komponenseket tartalmaz. Az eredmények azt mutatják, hogy a másodrendű struktúráig a PINAR(1,1_S) folyamat nem különbözik egy megfelelő PAR(1,1_S) folyamattól, azonban az innovációs folyamat függőségi szerkezete már sokkal bonyolultabb az egész-értékű esetben. A vizsgálatok fő eszköze az idősor implicit stationárius vektor állapotterez leírása.

2. A MODELL

Az elsőrendű, periodikus és szezonális, egész-értékű $\{Y_t\}$ autoregressziós folyamatot, amelyet PINAR(1,1_S) folyamatnak nevezünk S szezonális periódussal, ahol $S \in \{2, 3, \dots\}$, a következő periodikus sztochasztikus differenciaegyenlet megoldásaként tudjuk bevezetni:

$$Y_{kS+s} = \sum_{j=1}^{Y_{kS+s-1}} \xi_{k,s,j} + \sum_{j=1}^{Y_{kS+s-S}} \eta_{k,s,j} + \varepsilon_{kS+s}$$

ahol $k \in \mathbb{Z}$, $s = 1, \dots, S$. Az egyenletben a Y_{kS+s} jelölés a k -adik periódus s -edik szezonjára vonatkozó sorozatot jelöli. Például havi adatok és éves szezonális esetén $S = 12$, s az év hónapja, k pedig az év indexe.

A könnyebbség kedvéért a nem periodikus Y_t jelölést és a periodikus Y_{kS+s} jelölést felcserélhetően fogjuk használni. Az egyszerűség kedvéért feltételezzük, hogy a $\{\xi_{k,s,j}\}$ és $\{\eta_{k,s,j}\}$ véletlen változók függetlenek és Bernoulli-eloszlásúak eloszlásúak, ahol a várható érték $E(\xi_{k,s,j}) = \alpha_s$, illetve $E(\eta_{k,s,j}) = \beta_s$ minden $s = 1, \dots, S$ és $k \in \mathbb{Z}$, $j \in \mathbb{N}$ esetén. Megjegyezzük, hogy az eredmények érvényesek általánosabb ritkító operátorokra is, amelyek más, korlátlanul osztható számláló eloszlásokon alapulnak. Az $\alpha_s, \beta_s \in [0, 1]$ paramétereket, ahol $s = 1, \dots, S$, a modell autoregressziós együtthatóinak nevezzük, ahol az α_s -ek az elsőrendű időbeli függőségért, az β_s -ek pedig a szezonális függőségért felelősek. A bevándorlási folyamat $\{\varepsilon_t\}$ nem-negatív egész értékű véletlen változók egy sorozata, ahol az $\varepsilon_k := (\varepsilon_{kS+1}, \dots, \varepsilon_{kS+S})$, $k \in \mathbb{Z}$, véletlen vektorok függetlenek és azonos eloszlásúak, véges λ várható érték vektorral. A bevándorlási véletlen változók egy adott időszakon belül azonban korrelálhatnak egymással. Megjegyezzük, hogy ez a tulajdonság a modellt általánosabbá teszi az eredeti modellnél, ld. Filho et al. [7]. Ezenkívül a $\{\xi_{k,s,j}\}$, $\{\eta_{k,s,j}\}$ és $\{\varepsilon_k\}$ halmazok egymástól függetlenek. Az üres összeg értékét 0-ként definiáljuk a modell képleteiben. Az összes véletlen változó egy közös valószínűségi téren van definiálva.

A PINAR(1,1_S) modellnek három véletlenszerű komponense van, mivel a folyamat elemeinek halmaza az s -edik időszakban a következő három, egymástól független új elemekből álló részhalmazból áll. Az előző generációból fennmaradt elemek, amelyek túlélési valószínűsége α_s . Az egy korábbi időszak elemeiből származó, β_s születési valószínűséggel rendelkező, szezonálisan újonnan született elemek. Valamint a rendszerbe λ_s intenzitással belépő új elemek, amelyek a bevándorlásnak felelnek meg. Ezenkívül az α_s, β_s autoregressziós

együtthatók és a λ_s bevándorlási intenzitások, $s = 1, \dots, S$, az S periódusnak megfelelően időszakosan változnak. Ebben az összefüggésben a PINAR(1,1 $_S$) modell mind a szezonalitást, mind a periodicitást figyelembe veszi az autoregressziós együtthatókban, vagyis egyfajta szezonálisan korrelált periodikus (ciklo-stacionárius) modellnek tekinthető, hasonlóan a standard lineáris idősorokra vonatkozó periodikus SARIMA-modellhez.

A PAR-modellre ismert állapotterres megközelítést követve a PINAR(1,1 $_S$) modell a mátrixos binomiális ritkító operátor segítségével algebrai formában átírható az alábbi módon. A mátrixos binomiális ritkító operátort, amelyet mátrixos Steuel- és van Harn-operátornak is neveznek, M° jelöléssel jelöljük, ahol az $S \times S$ dimenziójú $M^\circ = (m_{i,j}^\circ)$ mátrix egymástól független binomiális ritkító operátorokból áll. Definiáljuk az $\mathbf{Y}_k$ állapot vektorokat úgy mint $\mathbf{Y}_k := (Y_{kS+1}, \dots, Y_{kS+S})$, $k \in \mathbb{Z}$, és tekintsük az $\{\mathbf{Y}_k\}$ és $\{\boldsymbol{\varepsilon}_k\}$ nem-negatív egész értékű sztochasztikus folyamatokat. Ekkor belátható, hogy a fenti modell egyenlet az alábbi stacionárius vektor egyenletre írható át:

$$\mathbf{Y}_k = A_k \circ \mathbf{Y}_k + B_k \circ \mathbf{Y}_{k-1} + \boldsymbol{\varepsilon}_k$$

ahol az A_k és B_k mátrixos binomiális ritkító operátorokat a PINAR(1,1 $_S$) modellbeli megfelelő véletlen összegekkel definiáljuk. Ezt az egyenletet a PINAR(1,1 $_S$) modell implicit állapotterres leírásának nevezzük. Ebben a leírásban $\{A_k, B_k, \boldsymbol{\varepsilon}_k\}$ teljesen független sorozatot alkot. Az $S \times S$ dimenziós, az A_k és B_k mátrixos binomiális ritkító operátorokhoz tartozó, A és B várható érték mátrixok az alábbi formában írhatóak fel:

$$A = \begin{bmatrix} 0 & \dots & 0 \\ \alpha_2 & \ddots & \vdots \\ 0 & \alpha_S & 0 \end{bmatrix} \quad B = \begin{bmatrix} \beta_1 & \dots & \alpha_1 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \beta_S \end{bmatrix}$$

Ílymódon, A és B nem-negatív mátrixok.

3. ELMÉLETI EREDMÉNYEK

Először bevezetjük a következő spektrális feltételt. Az $A + B$ mátrix spektrális sugara kisebb mint 1, azaz ez a mátrix egy nem-negatív stabil vagy konvergens mátrix. Ez a feltétel jól ismert többek között a periodikus ARMA folyamatok elméletéből. Első eredményként megadjuk az állapotterres leírás stacionárius megoldásának $\boldsymbol{\mu}$ várható érték vektorát és $\boldsymbol{\Sigma}$ variancia mátrixát. Először is látható, hogy ezek a mennyiségek az alábbi egyenleteket elégítik ki:

$$\boldsymbol{\mu} = (I - A)^{-1}(B\boldsymbol{\mu} + \boldsymbol{\lambda})$$

és

$$\boldsymbol{\Sigma} = (I - A)^{-1}(B\boldsymbol{\Sigma}B^T + \text{diag}((V_A + V_B)\boldsymbol{\mu}) + \boldsymbol{\Lambda})((I - A)^{-1})^T$$

ahol $\boldsymbol{\Lambda}$ a bevándorlás vektor variancia mátrixa. Amennyiben teljesül a spektrális feltétel, úgy ezek az egyenletek egyértelműen megoldhatóak és a $\boldsymbol{\mu}$ valamint $\boldsymbol{\Sigma}$ mennyiségekre explicit kifejezés adható.

Második eredményként a következő jellemzés adható az $\{Y_t\}$ PINAR(1,1 $_S$) folyamatra. Az alábbi állítások ekvivalensek egymással: (i) $\{Y_t\}$ egy periodikusan korrelált folyamat S periódus hosszal; (ii) A nem-negatív egész értékű véletlen vektor $(Y_1, \dots, Y_S)$ várható érték vektora ($\boldsymbol{\mu}$) és kovariancia mátrixa ($\boldsymbol{\Sigma}$) létezik és véges valamint kielégíti a fenti egyenleteket; (iii) $\{\mathbf{Y}_k\}$ egy gyengén stacionárius folyamat, amelynek $\boldsymbol{\mu}$ várható érték vektora és $\boldsymbol{\Sigma}$ variancia mátrixa kielégíti a fenti egyenleteket.

A harmadik eredmény azt mondja ki, hogy az $\{\mathbf{Y}_k\}$ gyengén stacionárius folyamatnak létezik egyoldali mozgóátlag előállítása az alábbi formában:

$$\mathbf{Y}_k = \boldsymbol{\mu} + \sum_{j=0}^{\infty} ((I - A)^{-1}B)^j \mathbf{W}_{k-j}$$

ahol a $\{\mathbf{W}_k\}$ folyamat egy fehér zaj.

A negyedik eredmény a PINAR(1,1 $_S$) folyamat S -dimenziós stacionárius állapotter $\{\mathbf{Y}_k\}$ reprezentációjának létezését és egyértelműségét biztosító tétel. A spektrális feltétel mellett létezik egy majdnem biztosan egyértelmű, a jövőtől nem függő, és gyengén stacionárius nem-negatív egész-vektor értékű $\{\mathbf{Y}_k\}$ folyamat, amely kielégíti a stacionárius állapot vektor egyenletet. Ez a $\{\mathbf{Y}_k\}$ folyamat előállítható egy majdnem biztosan konvergens végtelen sorral:

$$\mathbf{Y}_k = \sum_{j=1}^{\infty} \mathbf{Z}_k^{(j)}$$

ahol $\mathbf{Z}_k^{(j)}$ a bevándorlók j -edik generációs utódainak számát jelöli a k -adik időpontban lévő. Az $\{\mathbf{Y}_k\}$ folyamat erősen stacionárius és ergodik, várható érték vektora $\boldsymbol{\mu}$, valamint varianciamátrixa $\boldsymbol{\Sigma}$. Ezentúl, az $\{\mathbf{Y}_k\}$ folyamat autokorreláció mátrix függvénye megegyezik egy első rendű vektor autoregressziós (VAR) folyamat autokorreláció mátrix függvényével.

4. ALKALMAZÁS ÁTVÉTELI (PICK-UP) ADATOKRA

Ebben a fejezetben azt a számláló idősort (Pickup data) vizsgáljuk meg, ami egy átvételi pontokat kezelő vállalat egyik átvételi pontjáról (PUP – pick up point) naponta felvett csomagok számát adja meg. Ez egy csv fájlban tárolt, egy átvételi pontnál felvett összes csomag napi összesítéséből áll. Az adatállomány elérhető a <https://github.com/cabani/forecastingparcels> oldalon; lásd Nguyen et al. [16]. Ez a valós adatsor a 2017. július 3. és 2019. december 29. közötti időszakra vonatkozik, ami $t = 910$ napi ($n = 130$ hét) megfigyelést eredményez. Az 1. ábrán látható sorozat erős periodikus ingadozást mutat abban az értelemben, hogy az átlag és a variancia periodikusan változik. Ezt alátámasztják az idősor autokorreláció függvény és periodogramja is, ld. a 2. ábra. Mindezen empirikus eredmények azt támasztják alá, hogy a PINAR(1,1₅) alkalmas lehet az idősor dinamikájának leírására. Az idősor kérésre elérhető a szerzőtől, az azt elemző R- és Python-kódokkal együtt. Az adatokat az R szabványos idősor könyvtáraival dolgoztuk fel, a perarma függvény segítségével a periodikus idősorok elemzéséhez, valamint olyan Python-könyvtárakkal, mint a csv, numpy, pandas és scipy. Az eredményeket a két szoftver kimeneteinek összehasonlításával is ellenőriztük.

Az adatokra egy PINAR(1,1₇) modellt illesztettünk az Yule-Walker módszer segítségével. Ez a becslési módszer nem igényel semmilyen feltételezést a ritkító operátor és a bevándorlás eloszlásáról. A PINAR(1,1₇) modell paramétereinek becsléseit az 1. táblázat tartalmazza. Ezek a paraméterértékek mélyebb betekintést nyújtanak az idősor dinamikájának megértéséhez. Azt mutatják, hogy az elsőrendű időbeli függőség vasárnap kivételével minden nap szignifikáns, és erősebb, mint a szezonális függőség. A legnagyobb elsőrendű autoregressziós együttható csütörtökön figyelhető meg, ami erős függőséget jelez az előző szerdai értéktől, míg a legnagyobb szezonális autoregressziós együtthatók pénteken és szombaton figyelhetők meg, ami erősebb szezonális függőségre utal az egymást követő hétvégék között. Ezzel szemben a bevándorlás hatása a hét elején a legnagyobb (hétfőn és kedden). Végül, mivel az autoregressziós paraméterek vasárnap nem szignifikánsak, az idősor itt fehér zajként viselkedik.

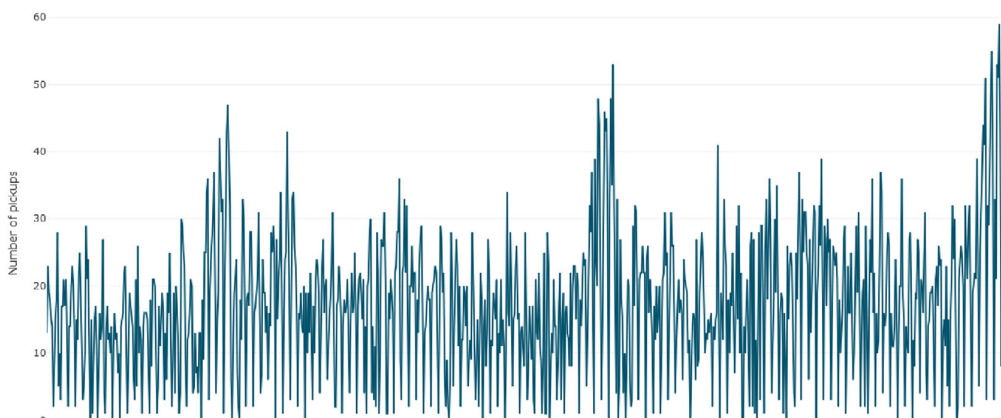
Táblázatok:

PINAR(1,1₇) modell paramétereinek becslése a Pickup adatokon

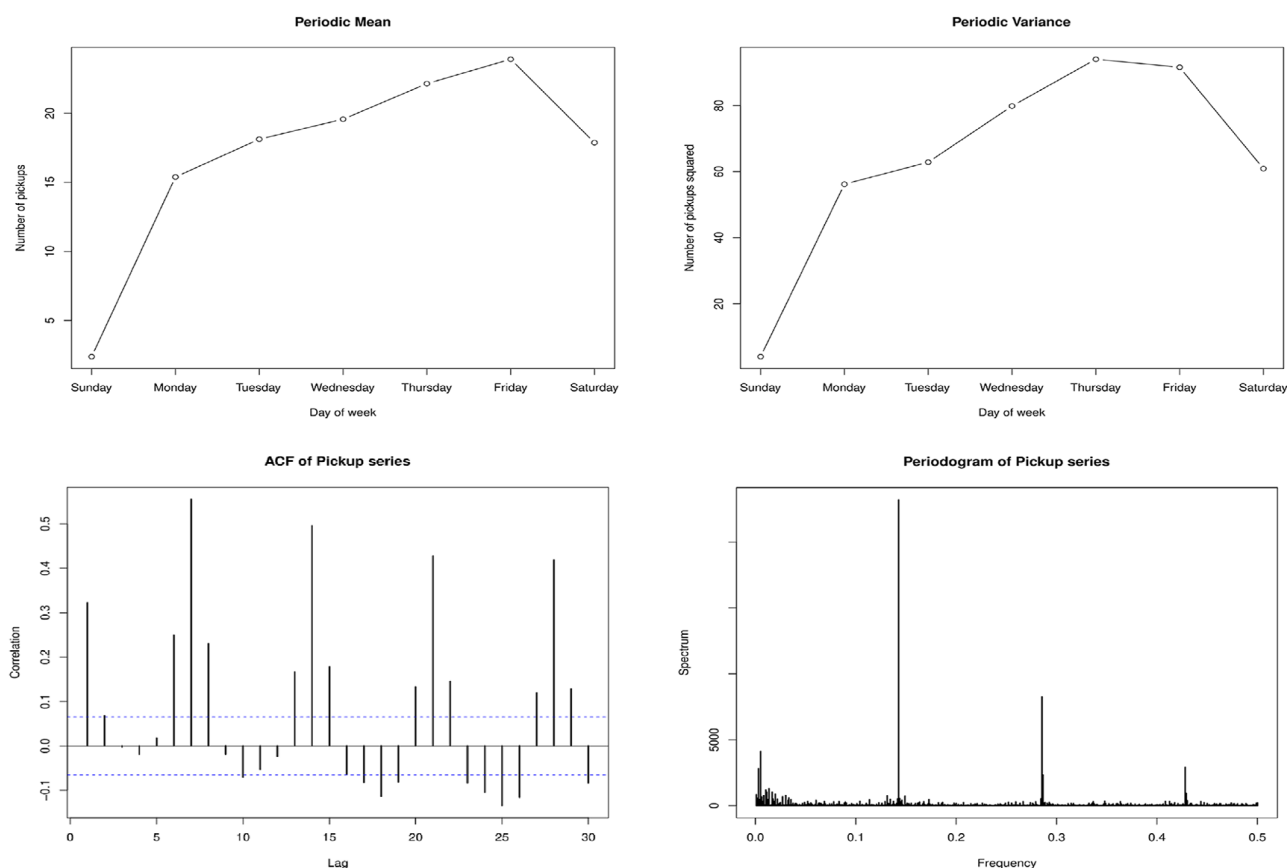
1. táblázat

	Vasárnap	Hétfő	Kedd	Szerda	Csütörtök	Péntek	Szombat
α	0.065	0.224	0.280	0.337	0.547	0.398	0.346
β	-0.072	0.165	-0.014	0.171	0.196	0.207	0.218
λ	1.393	12.321	14.072	10.122	7.092	10.137	5.698

Ábrák:



1. ábra. Az átvételi (Pickup) adatok idősora



2. ábra. Az átvételi (Pickup) adatok periodikus átlaga, varianciája, autokorreláció függvénye és periodogramja

KÖSZÖNETNYILVÁNÍTÁS

Ezúton szeretnék köszönetet mondani Pascal Bondon (CNRS, CentraleSupélec, Laboratoire des Signaux et Systèmes, Université Paris-Saclay, Gif-sur-Yvette, France), Valdério Anselmo Reisen (Graduate Program in Environmental Engineer, Universidade Federal do Espírito Santo, Vitoria, Brazil) és Paulo Roberto Prezotti Filho (Instituto Federal do Espírito Santo (IFES), Vitoria, Brazil) kutatóknak a közös kutatómunkáért és a gyümölcsöző beszélgetésekért, amely jelentősen hozzájárult jelen munka elkészítéséhez.

IRODALMI HIVATKOZÁSOK

- [1] Al-Osh MA, Alzaid AA. *First-order integer-valued autoregressive (INAR(1)) process*. Journal of Time Series Analysis, 1987 8(3):261–275.
- [2] Baek C, Davis RA, Pipiras V. *Sparse seasonal and periodic vector autoregressive modeling*. Computational Statistics and Data Analysis, 2017 106:103–126.
- [3] Basawa IV, Lund R. *Large sample properties of parameter estimates for periodic ARMA models*. Journal of Time Series Analysis, 2001 22(6):651–663.
- [4] Basawa IV, Lund R, Shao Q. *First-order seasonal autoregressive processes with periodically varying parameters*. Statistics & Probability Letters, 2004 67:299–306.
- [5] Chatfield C, Prothero DL. *Box-Jenkins seasonal forecasting: problems in a case-study*. Journal of the Royal Statistical Society: Series A, 1973 136(3):295–315.
- [6] Davis RA, Fokianos K, Holan SH, Joe H, Livsey J, Lund R. 2021. *Count time series: a methodological review*. Journal of the American Statistical Association 116(535):1533–1547.
- [7] Filho PRP, Reisen VA, Bondon P, Ispány M, Melo MM, Serpa FS. *A periodic and seasonal statistical model for non-negative integer-valued time series with an application to dispensed medications in respiratory diseases*. Applied Mathematical Modelling, 2021 96:545–558.
- [8] Gardner WA, Napolitano A, Paura L. *Cyclostationarity: half a century of research*. Signal Processing, 2006. 86(4):639–697.
- [9] Hindrayanto I, Koopman SJ, Ooms M. *Exact maximum likelihood estimation for non-stationary periodic time series models*. Computational Statistics and Data Analysis, 2010 54:2641–2654.

- [10] Hurd HL, Miamee A. *Periodically Correlated Random Sequences: Spectral Theory and Practice*, Vol. 355 John Wiley & Sons, Hoboken, NJ, 2007.
- [11] Ispány M, Bondon P, Reisen VA, Filho PRP. *Existence of a periodic and seasonal INAR process*. Journal of Time Series Analysis, 2024 45:980–1005.
- [12] Koopman SJ, Ooms M, Hindrayanto I. *Periodic unobserved cycles in seasonal time series with an application to US unemployment*. Discussion Paper No. 06-101/4. Amsterdam and Rotterdam: Tinbergen Institute, 2006. <https://www.econstor.eu/handle/10419/86347>.
- [13] Lund R. *Choosing seasonal autocovariance structures: PARMA or SARMA*. In Economic Time Series: Modelling and Seasonality, Bell WR, Holan SH, McElroy TS (eds.). CRC Press, Boca Raton, 2012 63–80.
- [14] Lund R, Basawa IV. *Recursive prediction and likelihood evaluation for periodic ARMA models*. Journal of Time Series Analysis, 2000 21(1):75–93.
- [15] McKenzie E. Some simple models for discrete variate time series. Journal of the American Water Resources Association, 1985 21(4):645–650.
- [16] Nguyen T-T-T, Cabani A, Cabani I, De Turck K, Kieffer M. Load prediction of parcel pick-up points: model-driven vs data-driven approaches. International Journal of Production Research, 2023 62(11):4046–4075.
- [17] Pagano M. *On periodic and multiple autoregressions*. The Annals of Statistics, 1978 6(6):1310–1317.
- [18] Shumway RH, Stoffer DS. *Time Series Analysis and Its Applications With R Examples*. Springer Texts in Statistics, 4th ed. Springer International Publishing, Cham, Switzerland, 2017.
- [19] Vecchia AV. *Periodic autoregressive-moving average (PARMA) modeling with applications to water resources*. Journal of the American Water Resources Association, 1985 21(5):721–730.
- [20] Weiss CH. *An Introduction to Discrete-Valued Time Series*. Wiley, Hoboken, NJ, 2018.