نشریه حفاظت منابع آب و خاک، سال دهم، شماره دو، زمستان ۱۳۹۹

ISSN 2251-7480

کاربرد مدل هیبرید ARIMA و رگرسیون بردار پشتیبان جهت بهبود پیش‌بینی سری زمانی

لاله پرویز[1*]، بهاره سعید آبادی[2]

۱*) دانشیار دانشکده کشاورزی، دانشگاه شهید مدنی آذربایجان، تبریز، ایران

ایمیل نویسنده مسئول مکاتبات: laleh_parviz @yahoo.com

۲) دانش آموخته کارشناسی دانشکده کشاورزی، دانشگاه شهید مدنی آذربایجان، تبریز، ایران

تاریخ دریافت: ۱۳۹۸/۰۵/۱۱ تاریخ پذیرش: ۱۳۹۹/۰۸/۲۰

چکیده:

بررسی دقیق ساختار اصلی سری زمانی نقش مهمی در افزایش دقت پیش‌بینی مدل ARIMA دارد. هدف این تحقیق بررسی تاثیر جداسازی مدلسازی بخش خطی و غیرخطی سری زمانی در نتایج مدل ARIMA است. تفکیک مدلسازی سری‌های عملکرد محصول گندم و ذرت دانه‌ای (استان‌های کرمانشاه و اصفهان) در بخش خطی مربوط به مدل ARIMA بود و در بخش غیرخطی با رگرسیون بردار پشتیبان انجام گرفت(مدل هیبرید). نتایج مدلسازی می‌تواند تحت تاثیر نوع ترکیب مورد استفاده بخش غیر خطی در مدل هیبرید تغییر یابد، به‌عنوان نمونه در سری زمانی ذرت دانه‌ای در استان کرمانشاه مقدار RMSE در ترکیبی فقط با باقی‌مانده‌ها ۱/۵۲ و در ترکیبی با سری زمانی ۱۵/۰۳ برآورد شد. در سری زمانی گندم در استان اصفهان با مدل هیبرید میزان کاهش آماره‌های RMSE، MAE و UII به‌ترتیب برابر با ۴۵/۹۴، ۵۲/۲۹ و ۴۶ درصد بود که بیانگر بهبود نتایج با مدل هیبرید و تفکیک مدلسازی بخش خطی و غیرخطی سری زمانی است. مقادیر GMER در هر چهار سری زمانی بزرگتر از یک بودند که حاکی از بیش‌برآورد مقادیر پیش‌بینی شده مدل هیبرید می‌باشد. مقایسه متوسط مقادیر آماره‌ها در دو استان حاکی از تاثیر نوع اقلیم در مبحث مدلسازی است چرا که متوسط مقادیر هر آماره در هر دو مدل (ARIMA و هیبرید) و در هر دو محصول در استان اصفهان نسبت به کرمانشاه کاهش داشت (میزان کاهش RMSE و UII به ترتیب ۲۴/۷۲ و ۱۲/۲۴ درصد). بنابراین تفکیک مدلسازی بخش خطی و غیرخطی می‌تواند دقت نتایج مدل ARIMA را افزایش دهد.

کلید واژه‌ها: ARIMA، بخش غیر خطی، رگرسیون بردار پشتیبان، هیبرید

مقدمه

با توجه به اهمیت پیش‌بینی عملکرد محصول در تصمیم‌گیری‌های مربوط به بازاریابی محصولات کشاورزی، برنامه‌ریزی کشاورزی، تامین مواد غذایی، تجارت بین المللی مواد غذایی، پایداری اکوسیستم،... در سال‌های اخیر از ساختارهای مختلف جهت برآورد عملکرد محصول استفاده شده است (Kumar Paul and Sinha 2016) که نمونه‌هایی از آن شامل مدل‌های شبیه-سازی رشد محصول (Lecerf *et al.*, 2019)، تاثیرات اقلیمی بر عملکرد محصول (زارع ابیانه، ۱۳۹۲) و استفاده از مفاهیم سری زمانی (Verma *et al.*, 2012; Sharma *et al.*, 2014; Biswas *et al.*, 2018) می‌باشد. مدل‌های عملکرد محصول بیان خلاصه‌ای از تقابل محصول با محیط آن است و می‌توانند از همبستگی ساده عملکرد با تعداد محدودی از متغیرها تا مدل‌های آماری پیچیده را شامل گردند. استفاده از مدل‌های آماری پیچیده در بیان شرایط حاکم دچار پیچیدگی می‌شوند (Suman and Urmil Verma, 2016). در مدلسازی عملکرد محصول با

رویکرد اقلیمی سعی در افزایش کارایی است چرا که اینوع مدلسازی تنها متغیرهای اقلیمی را در فرآیند شبیه-سازی در نظر می‌گیرد. در تحقیق پرویز و پیمایی (۱۳۹۷) در پیش‌بینی عملکرد چهار گیاه گندم، جو، سیب‌زمینی و ذرت دانه‌ای در استان‌های آذربایجان شرقی و غربی، رویکرد ترکیبی (اقلیمی- استوکستیکی) نسبت به رویکرد اقلیمی کارایی مدل را افزایش داد. استفاده از مفاهیم سری زمانی در پیش‌بینی عملکرد محصول با توجه به در نظر گرفتن تاثیر تغییرات عملکرد محصول در سال‌های مختلف گزینه دیگری است. بررسی تغییرات زمانی عملکرد محصول با استفاده از ساختار درونی آنها می‌تواند در این زمینه کمک‌کننده باشد. در سال‌های اخیر مدل ARIMA در این زمینه کاربرد چشمگیری داشته است، به عنوان نمونه از مدل ARIMA (2,1,0) در پیش‌بینی تولید ذرت استفاده کردند که ملاک انتخاب مدل ترسیم همبستگی‌نگار و همبستگی‌نگار جزئی بود. مدل منتخب حاکی از ۱۳/۷۶ درصد افزایش در تولید ذرت در پنج سال مدنظر تحقیق بود (Sharma *et al.*, 2018). در سال های اخیر مدل ARIMA با استفاده از تفکیک مدلسازی بخش خطی و غیر خطی در قالب مدل‌های هیبرید ارتقاء چشمگیری در پیش‌بینی سری زمانی بوجود آورده است. مدل‌های هیبرید در واقع ترکیب مدل ARIMA با سایر مدل‌ها از جمله شبکه عصبی مصنوعی، رگرسیون بردار پشتیبان،... می‌باشند. با در نظر گرفتن نقاط قوت مدل ARIMA و شبکه عصبی مصنوعی به‌ترتیب در مدلسازی بخش خطی و غیر خطی، از مدل هیبرید در پیش‌بینی سری زمانی استفاده شد. نتایج تحقیق Peter Zheng (2003) حاکی از عملکرد بهتر مدل هیبرید نسبت به مدل ARIMA و شبکه عصبی مصنوعی بود، به‌طوری‌که میزان MAD در مدل هیبرید نسبت به مدل ARIMA و ANN به‌ترتیب ۷/۹۷٪ و ۷/۸۳٪ کاهش داشت. تجربه دیگر از مدل هیبرید شامل ترکیب مدل SARIMA و شبکه عصبی مصنوعی با استفاده از الگوریتم ژنتیک در پیش‌بینی مقدار تولید صنعت ماشین در تایوان بود. دلیل استفاده از الگوریتم ژنتیک مربوط به بهینه‌سازی تعداد نرون‌ها در لایه پنهان و تعداد پارامترهای یادگیری ساختار شبکه عصبی می‌باشد. مدل پیشنهاد شده در پیش‌بینی سری زمانی فصلی به‌عنوان مرجع قابل قبولی بود (Liang, 2009). مدل هیبرید در تحقیق Ruiz-Aguilar و همکاران (۲۰۱٤) جهت پیش‌بینی حجم بازرسی، ترکیبی از مدل SARIMA و رگرسیون بردار پشتیبان بود. مدل هیبرید عملکرد مدل‌های تکی از جمله SARIMA و رگرسیون بردار پشتیبان را بهبود بخشید. شاخص سازش از مدل SARIMA به مدل هیبرید از ۰/۸۷به ۰/۸۹ رسید. در تحقیق دیگر از مدل هیبرید (ترکیب SARIMA و ماشین بردار پشتیبان) در پیش‌بینی مقدار تولیدات صنعت ماشین آلات در تایوان استفاده شد. در مقایسه عملکرد سه مدل SARIMA، SVM و مدل هیبرید، خطای ریشه متوسط مربعات نرمال شده (NMSE[1]) به‌ترتیب ۰/۲۸، ۰/٤٦ و ۰/۱۱ گزارش شد. کمینه مقدار خطا مربوط به عملکرد مدل هیبرید بود. همچنین توانایی مدل هیبرید در پیش‌بینی نقاط عطف مربوط به دوره صحت‌سنجی قابل توجه بود (Chen and Wang, 2007). در تحقیق Zaynoddin و همکاران (۲۰۱۸) مدل هیبرید در پیش‌بینی بارندگی ماهانه در مناطق گرمسیری از کارایی بالایی برخوردار بود.

هدف این مقاله پیش‌بینی عملکرد محصولات گندم و ذرت دانه‌ای در دو استان اصفهان و کرمانشاه با دو رویکرد برپایه ساختار زمانی است. رویکرد اول مربوط به مدل ARIMA و رویکرد دوم مربوط به ترکیب مدل ARIMA و رگرسیون بردار پشتیبان (مدل هیبرید) بود. از رهیافت باکس- جنکینز (۱۹۷۰) در ساخت مدل ARIMA و از تابع کرنل و پارامتر تنظیم کننده در ساخت رگرسیون بردار پشتیبان استفاده شد. مقایسه عملکرد دو

[1] Normalized mean square error

مدل براساس برخی از آماره‌های ارزیابی عملکرد مدل انجام گرفت.

مواد و روش‌ها

منطقه مورد مطالعه

در این مطالعه به بررسی عملکرد دو رویکرد مدلسازی در پیش‌بینی سری زمانی عملکرد محصول گندم و ذرت دانه‌ای در استان‌های کرمانشاه و اصفهان پرداخته شده است که موقعیت آنها در شکل ۱ نشان داده شده است. طول دوره آماری مورد استفاده مربوط به سال‌های ۱۳۶۱ تا ۱۳۹۵ بوده است.

آمار مربوط به سری زمانی عملکرد محصول از آمارنامه وزارت جهاد کشاورزی اخذ شد.

رگرسیون بردار پشتیبان

ماشین بردار پشتیبان در مسایل مربوط به طبقه‌بندی و رگرسیون استفاده می‌شود. ماشین بردار پشتیبان با طبقه-بندی از تئوری یادگیری آماری براساس کمینه‌سازی ریسک ساختاری عمل می‌کند. از نظر تئوریکی، خطای مورد انتظار ماشین یادگیری و مسایل بیش‌برازش کمینه می‌شود. در نظر بگیرید که (x,y) مجموعه‌ای از N نمونه باشند که بردار ورودی شامل m جزء است و y مقادیر خروجی مرتبط است. تخمینگر ماشین بردار پشتیبان (f) در رگرسیون به‌صورت ریاضی براساس رابطه ۱ بیان می‌شود.

$$f(x) = w.\phi(x) + b \qquad (۱)$$

w: بردار وزن، b: اریبی، $\phi(x)$ تابع کرنل.

مساله بهینه‌سازی را می‌توان به‌صورت مساله بهینه-سازی محدب نوشت تا جوابی برای معادله ۱ بدست آید. بنابراین تابع هدف و قیود به‌صورت رابطه ۲ و ۳ نوشته می‌شود.

$$\frac{1}{2} w^T w + C\sum_{i=1}^{N} \xi_i + C\sum_{i=1}^{N} \xi_i^* \qquad (۲)$$

$$\begin{cases} w^T \phi(x_i) + b - y_i \le \varepsilon + \xi_i \\ y_i - w^T \phi(x_i) - b \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0, i = 1,\ldots, N \end{cases} \qquad (۳)$$

شکل ۱. موقعیت مکانی استان‌های مورد مطالعه

C: پارامتر تنظیم کننده که میزان خطای تجربی را در مساله بهینه‌سازی تعیین می‌کند، ξ_i^*و ξ_iمتغیرهای کمبود.

این مساله معمولا در فرم دوگان با تشکیل لاگرانژین حل می‌شود. این عملیات در فضای ورودی نسبت به فضای با بعد بالا با تابع کرنل انجام می‌گیرد. توابع معمول کرنل عبارتند از چند جمله‌ای، خطی، پایه شعاعی و سیگموئید (Hamidi *et al.*, 2014; Byvatov *et al.*, 2003).

مدل ARIMA

در ساخت مدل ARIMA از روند تکراری باکس- جنکینز (۱۹۷۰) استفاده می‌شود که شامل سه مرحله تکراری شامل شناسایی مدل، تخمین پارامترهای مدل و آزمون نکویی برازش است. ایده اصلی شناسایی مدل این است که اگر یک سری زمانی از مدل ARIMA حاصل شود، باید برخی از خصوصیات خودهمبستگی تئوریکی را داشته باشد. باکس- جنکینز (۱۹۷۰) تابع خودهمبستگی و تابع خودهمبستگی جزئی مربوط به داده‌ها را جهت شناسایی مرتبه‌های مدل ARIMA پیشنهاد دادند. در مرحله شناسایی مدل تبدیل داده جهت ایجاد سری زمانی ایستا لازم است. ایستایی مرحله مهم و ضروری در ساخت مدل ARIMA می‌باشد که در پیش‌بینی سری زمانی مفید است. در صورت وجود روند و ناهمگونی در سری زمانی، تفاضل‌گیری و تبدیلات توانی جهت حذف روند و ایستایی واریانس لازم است. تخمین پارامترهای مدل از روند بهینه‌سازی غیرخطی با هدف کمینه‌سازی خطا پیروی می‌کند. مرحله آزمون نکویی برازش مربوط به کنترل فرضیه مدل در ارتباط با خطا (ε_t) است. جهت انجام این مرحله چندین آماره نکویی برازش و نمودار باقی‌مانده‌ها پیشنهاد شده است که در صورت اثبات عدم کفایت مدل، یک مدل جدید باید مورد آزمایش قرار گیرد (Narasimha Murthy *et al.*, 2018; Peter Zhang, 2003).

ترکیب مدل ARIMA و رگرسیون بردار پشتیبان (مدل هیبرید)

مدل ARIMA و رگرسیون بردار پشتیبان موفقیت‌هایی در حوضه‌های خطی و غیرخطی داشته‌اند، با این حال هیچ یک از آنها الگوی کلی و مناسبی برای هر شرایطی نمی‌باشند. عملکرد مدل ARIMA در مسایل پیچیده غیر-خطی ممکن است مناسب نباشد، از طرف دیگر مدل‌هایی مانند شبکه عصبی مصنوعی در مدلسازی مسایل خطی دارای نتایج قابل تاملی هستند. مارخام و راکس (۱۹۹۸) به این نتیجه رسیدند که عملکرد مدل شبکه عصبی مصنوعی در مسایل رگرسیون خطی وابسته به اندازه نمونه و سطح نویز است. بنابراین از آنجایی که شناسایی کامل مشخصه‌های داده‌ها در مسایل واقعی سخت است رویکرد هیبرید استراتژی مناسبی در مورد مسایل واقعی است.

اگر یک سری زمانی از دو بخش خطی و غیر خطی تشکیل شده باشد، می‌توان سری زمانی را بصورت رابطه ٤ نشان داد.

$$y_t = L_t + N_t \qquad (۴)$$

L_t: بیانگر جزء خطی، N_t: نمایانگر جزء غیر خطی.

این دو جزء باید از روی داده‌ها تخمین زده شوند، به این‌صورت که در ابتدا از مدل ARIMA برای مدلسازی جزء خطی استفاده می‌شود و سپس باقی‌مانده‌های مدل خطی شامل بخش غیرخطی خواهند بود. اگر e_t نمایانگر باقی‌مانده در زمان t از مدل خطی باشد (رابطه ۵)

$$e_t = y_t - \hat{L}_t \qquad (۵)$$

L_t: مقادیر پیش‌بینی شده در زمان t

مدلسازی باقی‌مانده‌ها می‌تواند با استفاده از روش‌های هوش مصنوعی مانند رگرسیون بردار پشتیبان، ... با ترکیب‌های مختلف انجام گیرد که یکی از ترکیب‌های مورد نظر در رابطه ٦ بیان شده است.

$$e_t = f(e_{t-1}, e_{t-2}, \ldots, e_{t-n}) + \varepsilon_t \qquad (٦)$$

f: تابع غیر خطی، ε_t: خطای تصادفی.

ترکیب بخش خطی و غیرخطی ($\hat{y}_t = \hat{L}_t + \hat{N}_t$) امکان پیش‌بینی سری زمانی را به‌صورت هیبرید می‌دهد (Chen and Wang 2007; Peter Zhang, 2003).

مقایسه عملکرد مدل‌های مورد بررسی

جهت بررسی عملکرد مدل ARIMA و هیبرید از برخی آماره‌های خطا استفاده شده است که عبارتند از خطای ریشه متوسط مربعات (RMSE[1])، میانگین خطای مطلق(MAE[2]) و آماره‌ای جهت ارزیابی کیفیت پیش‌بینی(UII که معادله آماره براساس اختلاف بین مقادیر مشاهداتی و پیش‌بینی بر مقادیر مشاهداتی است). در مقایسه عملکرد چندین مدل، مدلی با عملکرد بهتر دارای کمینه مقدار RMSE، MAE، UII است (Zeynoddin *et al.*, 2018)

نتایج و بحث

پیش‌زمینه مدلسازی سری زمانی، تفکیک سری زمانی به دو بخش واسنجی و صحت‌سنجی است که در این تحقیق از سال‌های ۱۳۶۱ تا ۱۳۸۵ جهت واسنجی مدل و از سال‌های ۱۳۸۶ تا ۱۳۹۵ جهت صحت‌سنجی مدل‌ها استفاده شد. بیشترین و کمترین میزان عملکرد محصول در دو استان به ترتیب مربوط به ذرت دانه‌ای در استان اصفهان (۶/۴۴ تن بر هکتار) و گندم در استان کرمانشاه (۳/۲۲ تن بر هکتار) بوده است. جهت پیش‌بینی سری زمانی عملکرد محصول از دو رویکرد ARIMA و هیبرید استفاده شد.

مدل ARIMA

اولین گام در مدلسازی سری زمانی بررسی نرمال بودن سری‌های زمانی و وجود روند در سری زمانی است که در این تحقیق از آزمون کولموگروف- اسمیرنوف جهت بررسی نرمال بودن داده‌ها و از آزمون من- کندال جهت بررسی وجود روند استفاده شد که نتایج در جدول ۱ آورده شده است.

جدول ۱. نتایج بررسی نرمال بودن سری زمانی و وجود روند

استان	نام محصول	آماره آزمون کولموگروف- اسمیرنوف	آماره آزمون من- کندال
اصفهان	گندم	۰/۸۵	۲/۸۵
	ذرت دانه‌ای	۰/۴	۱/۵۸
کرمانشاه	گندم	۰/۸	۵/۹
	ذرت دانه‌ای	۰/۲۴	۵/۰۸

با توجه به جدول ۱، سطح معنی‌داری آزمون کولموگروف- اسمیرنوف در هر چهار سری زمانی بزرگتر از ۰/۰۵ بودند، بنابراین سری‌ها نرمال می‌باشند و نیازی به تبدیل نرمالسازی نیست. با توجه به آماره من- کندال سری‌های زمانی عملکرد محصول گندم در استان اصفهان، گندم و ذرت دانه‌ای در استان کرمانشاه دارای روند افزایشی معنی‌داری است. در مرحله بعد برای تعیین مرتبه‌های مدل از معیار آکائیک استفاده شد که مرتبه‌های مورد ارزیابی از ۰ تا ۵ بودند (p=0,1,2,3,4,5- q=0,1,2,3,4,5). مدل‌های منتخب براساس کمینه مقدار معیار آکائیک و گذراندن آزمون نکویی برازش در جدول ۲ آورده شده است.

جدول ۲. مدل منتخب سری‌های زمانی عملکرد محصول در دو استان

اصفهان		کرمانشاه	
گندم	ذرت دانه‌ای	گندم	ذرت دانه‌ای
ARIMA (1,1,1)	ARIMA (1,0,5)	ARIMA (2,1,1)	ARIMA (0,1,2)

مدل هیبرید

همان طور که در بخش مواد و روش‌ها اشاره شد مدل هیبرید از دو قسمت تشکیل شده است ۱- قسمت خطی ۲- قسمت غیرخطی. در قسمت خطی از مدل‌های منتخب

[1] Root Mean Square Error
[2] Mean Absolute Error

جدول ۲ استفاده شد و سپس رگرسیون بردار پشتیبان جهت مدلسازی بخش غیرخطی یا باقیمانده‌ها بکار گرفته شد تا در نهایت ترکیب بخش خطی و غیرخطی انجام گیرد. ترکیبات مختلفی که در مدلسازی باقی‌مانده‌های مدل مورد استفاده قرار گرفت شامل ترکیب شماره ۱- $e_t = f(e_{t-1}, e_{t-2}) + \varepsilon_t$ ترکیب شماره ۲- $e_t = f(e_{t-1}, e_{t-2}, e_{t-3}) + \varepsilon_t$ ترکیب شماره ۳- $y_t = f(y_{t-1}, y_{t-2}) + \varepsilon_t$ ترکیب شماره ٤- $y_t = f(y_{t-1}, y_{t-2}, \hat{L}_t) + \varepsilon_t$ بودند. یکی از مراحل استفاده از رگرسیون بردار پشتیبان، تحلیل حساسیت مدل است که در این تحقیق تحلیل حساسیت مربوط به توابع کرنل و C به‌عنوان پارامتر تنظیم کننده مدل بود. توابع کرنل مورد استفاده شامل خطی، پایه شعاعی و سیگموئید بودند. نتایج تحلیل حساسیت مربوط به عملکرد گندم در جدول ۳ آورده شده است.

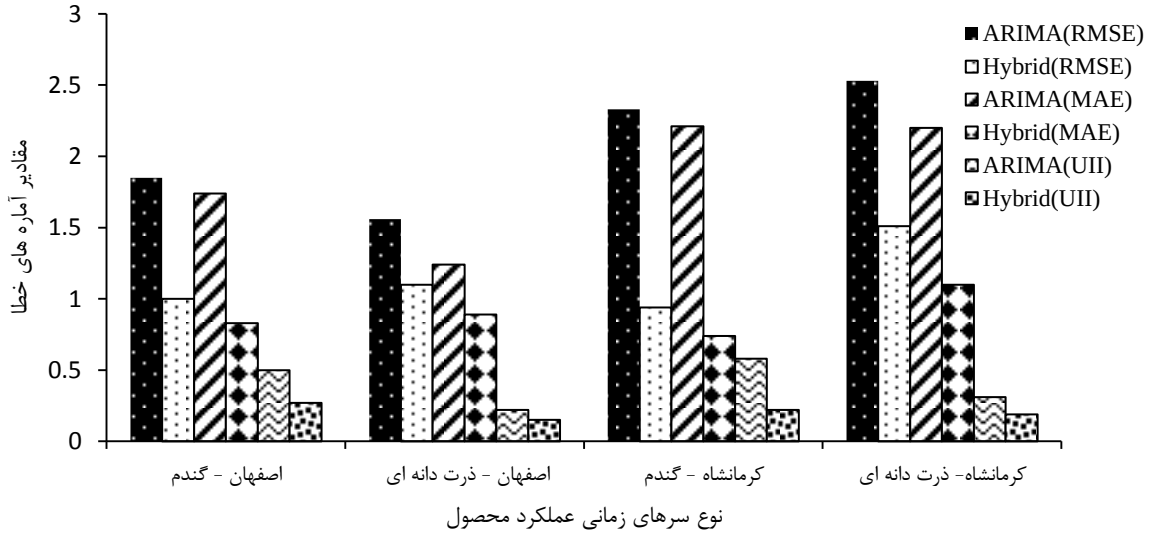
با توجه به جدول۳، کمترین میزان خطا مربوط به ترکیب شماره ۱ (تابع سیگموئید C=0.03) است. کمترین میزان خطا در سری زمانی ذرت دانه‌ای (استان اصفهان)، گندم (استان اصفهان) و ذرت دانه‌ای (استان کرمانشاه) به‌ترتیب مربوط به ترکیب شماره ۱ (تابع شعاعی C=0.03)، ترکیب شماره ۱ (تابع شعاعی C=2) و ترکیب شماره ۱ (تابع شعاعی C=0.03) بوده است. مساله قابل توجه نوع ترکیب مورد استفاده در مدلسازی باقی‌مانده‌ها می‌باشد به طوری که از ترکیب ۱و۲ به ترکیب ۳ و٤ میزان خطا به شدت افزایش می‌یابد به‌عنوان نمونه در سری زمانی ذرت دانه‌ای کرمانشاه با تابع خطی و C=0.03 میزان خطا از مقدار ۱/۵۲ در ترکیب ۱ به مقدار ۱٦/۲ در ترکیب ٤ رسیده است، در واقع تغییرات خطا با تغییرات ترکیب مورد استفاده در مدلسازی باقی‌مانده‌های مدل به مراتب بیشتر از تغییرات خطا با تغییرات تابع کرنل یا تغییرات خطا با تغییرات C می‌باشد. در تحقیق Chen و Wang (۲۰۰۷) و رویز- اگیلار و همکاران (۲۰۱٤) نیز ترکیبات متفاوت در مدلسازی بخش غیرخطی مقادیر خطای متفاوتی داشت.

مقایسه دو رویکرد

ارزیابی عملکرد دو رویکرد براساس آماره‌های معرفی شده بخش مواد و روش‌ها انجام شده است. مقایسه عملکرد دو مدل با آماره‌های بیان شده در شکل ۲ نشان داده شده است.

جدول ۳. نتایج تحلیل حساسیت رگرسیون بردار پشتیبان عملکرد گندم در استان کرمانشاه

نوع ترکیب	تابع کرنل	C	RMSE	تابع کرنل	C	RMSE	تابع کرنل	C	RMSE
ترکیب شماره ۱	خطی	C=1	۱/۰٤	شعاعی	C=1	۱/۲۹	سیگموئید	C=1	۰/۹۷
		C=0.03	۰/۹۸		C=0.03	۰/۹۳		C=0.03	۰/۹۲
		C=2	۱/۰٤		C=2	۱/۲۷		C=2	۱/۰٤
ترکیب شماره ۲	خطی	C=1	۰/۹۹	شعاعی	C=1	۱/۲	سیگموئید	C=1	۰/۹۳
		C=0.03	۱/۰۲		C=0.03	۰/۹۷		C=0.03	۱/۱
		C=2	۱/۰۳		C=2	۱/۱٦		C=2	۰/۹۳
ترکیب شماره ۳	خطی	C=1	۸/۰٤	شعاعی	C=1	۷/۸۱	سیگموئید	C=1	۸/٦
		C=0.03	۷/۹۸		C=0.03	٦/۸٤		C=0.03	۷/۰۷
		C=2	۸/۰۳		C=2	۷/۹		C=2	۸/۸
ترکیب شماره ٤	خطی	C=1	۹/۱۲	شعاعی	C=1	۸/۳	سیگموئید	C=1	۸/۲۳
		C=0.03	۸/٦٤		C=0.03	٦/۹		C=0.03	۷/۲٤
		C=2	۹/۰۸		C=2	۸/۳۷		C=2	٦/۵٤



شکل ۲. مقایسه عملکرد دو رویکرد مدلسازی با آماره‌های خطای متفاوت

براساس نتایج آماره‌های بیان شده در شکل۲، میزان کاهش RMSE از مدل ARIMA به مدل هیبرید در سری زمانی گندم (اصفهان)، ذرت دانه‌ای(اصفهان)، گندم (کرمانشاه) و ذرت دانه‌ای (کرمانشاه) به‌ترتیب ٤٥/٩٤، ٢٩/٤٨، ٥٩/٦٥ و ٤٠/٣١ درصد بود. میزان کاهش MAE از مدل ARIMA به مدل هیبرید در سری زمانی گندم (اصفهان)، ذرت دانه‌ای(اصفهان)، گندم (کرمانشاه) و ذرت دانه‌ای (کرمانشاه) به‌ترتیب ٥٢/٢٩، ٢٧/٦٤، ٦٦/٥١ و ٥٠ درصد بود. میزان کاهش UII از مدل ARIMA به مدل هیبرید در سری زمانی گندم (اصفهان)، ذرت دانه-ای(اصفهان)، گندم (کرمانشاه) و ذرت دانه‌ای (کرمانشاه) به‌ترتیب ٤٦، ٢٩/٠٩، ٦٢/٠٦ و ٣٩/٠٣٣ درصد بود. بنابراین بهبود مقادیر شبیه‌سازی‌شده توسط مدل ARIMA با تفکیک به دو بخش خطی و غیر خطی کاملا مشهود است و این نتیجه با نتیجه محققینی مانند Peter Zhang (۲۰۰۳)، Liang (۲۰۰۹)، Ruiz-Aguilar و همکاران (۲۰۱٤) و Chen and Wang (۲۰۰۷) کاملا مطابقت دارد. اگر متوسط مقادیر آماره‌ها با هر دو مدل ARIMA و هیبرید و با هر دو محصول در هر استان گرفته شود، نتایج به این‌صورت است: اصفهان RMSE= ۱/۳۷، MAE =۱/۱۷، UII= ۰/۲۸. کرمانشاه RMSE= ۱/۸۲، MAE =۱/٥٦، UII=۰/۳۲ بنابراین میزان آماره‌های خطا از استان اصفهان به کرمانشاه افزایش یافته است. اگر متوسط مقادیر آماره‌ها با هر دو مدل ARIMA و هیبرید و در هر دو استان برای هر محصول گرفته شود، نتایج به این‌صورت می‌باشد: گندم RMSE= ۱/٥۳، MAE =۱/۳۸، UII=۰/۳۹. ذرت دانه‌ای RMSE= ۱/٦۷، MAE =۱/۳٥، UII=۰/۲۱۸. میزان آماره RMSE در ذرت دانه‌ای افزایش یافته ولی دو آماره دیگر در ذرت دانه‌ای کاهش داشته‌اند. جهت تعیین بیش‌برآورد و کم‌برآورد مدل هیبرید از آماره نسبت خطای میانگین هندسی (GMER[1]) استفاده شد. مقادیر آماره در مدل هیبرید در شکل ۳ نشان داده شده است.

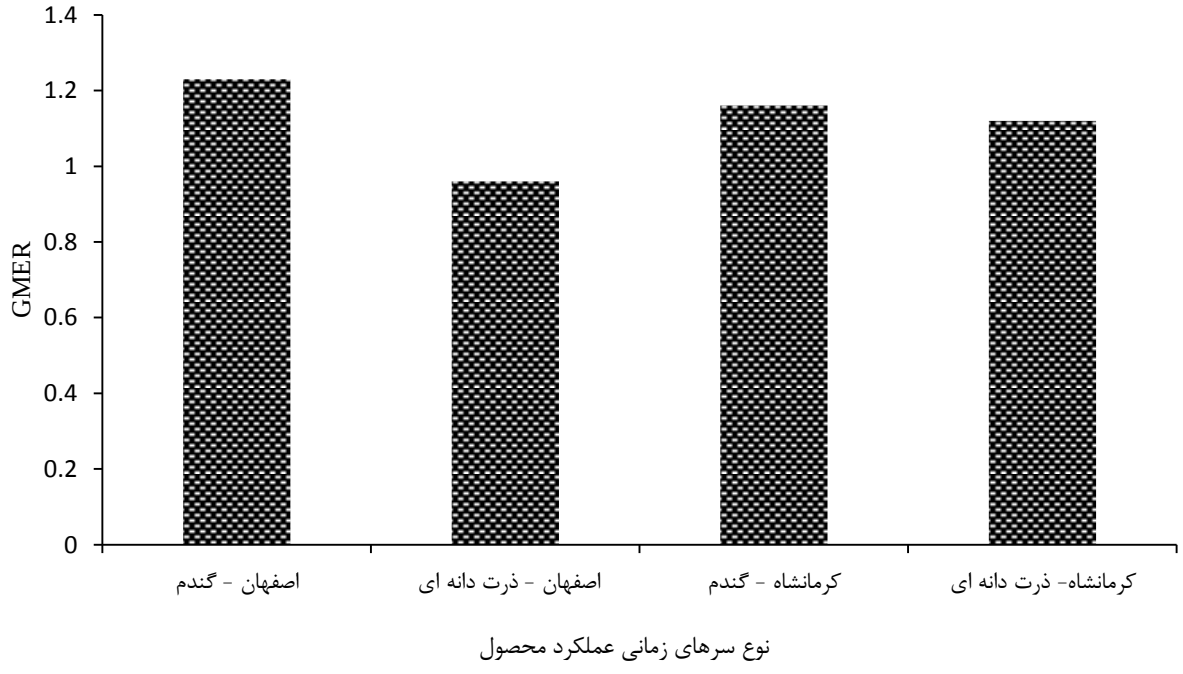
با توجه به این که مقدار GMER در هر چهار سری زمانی بزرگ‌تر از یک می باشد، نمایانگر بیش برآورد در داده‌های پیش‌بینی شده، است. در ادامه جهت مقایسه عملکرد دو مدل، از متوسط مقادیر عملکرد محصول در طی دوره صحت‌سنجی استفاده شد که نتایج درشکل ٤ آورده شده است.

[1] Geometric mean error ratio

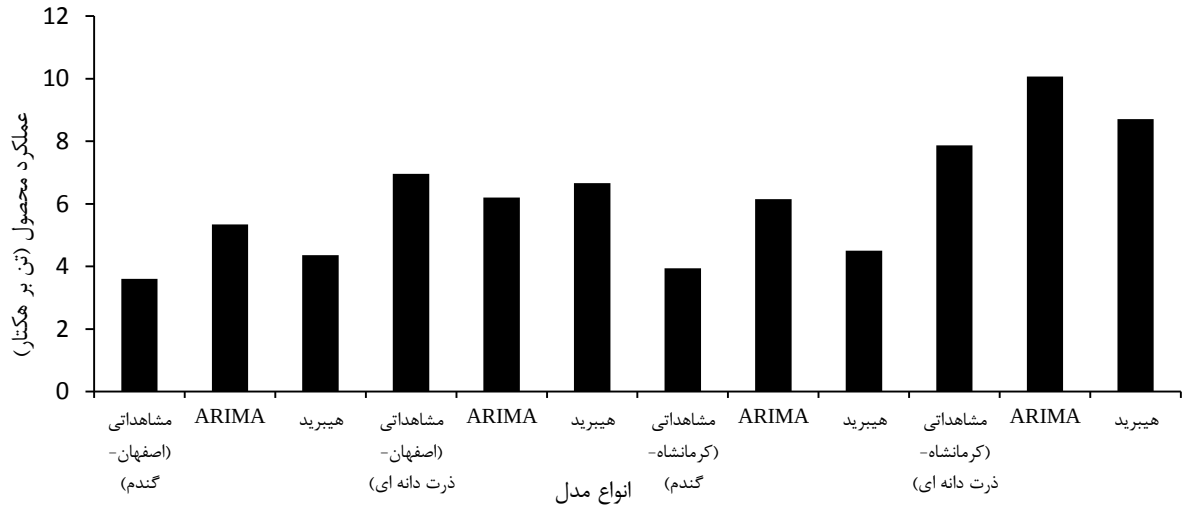
با توجه به شکل ٤، متوسط مقادیر عملکرد محصول حاصل از مدل هیبرید اختلاف کمی با مقادیر مشاهداتی دارد، در واقع با محاسبه متوسط مقادیر عملکرد محصول و مقایسه با داده‌های مشاهداتی، عملکرد قابل قبول مدل هیبرید در برابر مدل ARIMA اثبات می‌شود که کارایی بالای مدل هیبرید نسبت به مدل ARIMA در تحقیقات Ruiz-Aguilar و همکاران (۲۰۱٤)، Chen and Wang(۲۰۰۷) و Dı´az-Robles و همکاران (2008) هم مشاهده شد.

نتیجه‌گیری

پیش‌بینی عملکرد محصول با مدلی کارآمد در سیاست‌گذاری‌ها و تصمیم‌گیری‌های بخش کشاورزی از اهمیت چشمگیری برخوردار است. در این تحقیق از دو مدل ARIMA و هیبرید در تخمین عملکرد محصول استفاده شد. میزان آماره‌های خطا با استفاده از مدل هیبرید کاهش یافت که نمایانگر تاثیر تفکیک سری زمانی به دو بخش خطی و غیرخطی در روند مدلسازی است. مدل هیبرید با تلفیق نقاط قوت هر دو مدل کارایی بالایی را نسبت به حالت انفرادی هر مدل دارد، در واقع ادغام مدل خطی و غیرخطی توانایی در برگرفتن فرم‌های مختلف از ارتباطات درونی داده‌های سری زمانی را تسهیل می‌کند یعنی ساختار واقعی حاکم بر سری زمانی (بخش خطی و غیر خطی) را با دقت بیشتر مدلسازی می‌کند و الگوی بخش خطی و غیر خطی با دقت بیشتری آشکار می‌شود. نوع ترکیب مورد استفاده در مدلسازی بخش غیرخطی تاثیر چشمگیری در کاهش خطای مدل داشت، بنابراین یکی از مراحل مهم در مدلسازی هیبرید انتخاب ترکیب مناسب از باقی‌مانده‌ها و سری زمانی در بخش غیرخطی می‌باشد. متوسط مقادیر آماره‌ها با هر دو مدل و با تمام محصولات در هر استان بیانگر افزایش آماره‌های خطا در استان کرمانشاه است و یکی از دلایل این مساله را می‌توان در نوع اقلیم حاکم در منطقه و تاثیر آن در روند کاهشی یا افزایشی عملکرد محصول دانست. مقادیر GMER در هر چهار سری زمانی بزرگتر از یک بودند که بیانگر بیش-برآورد مدل هیبرید می‌باشد.



شکل ۳. مقادیر GMER در سری‌های زمانی مختلف و استان‌های متفاوت برای مدل هیبرید.



شکل ٤. مقایسه متوسط مقادیر عملکرد محصول مشاهداتی با شبیه‌سازی حاصل از عملکرد محصول

منابع مورد استفاده

پرویز، ل. و پیمایی، م. ۱۳۹۷. پیش بینی اقلیمی استوکستیکی عملکرد چهار گیاه گندم، جو، سیب زمینی و ذرت دانه‌ای در استانهای آذربایجان شرقی و غربی در راستای توسعه برنامه‌ریزی کشاورزی. نشریه تولید گیاهان زراعی، ۱۱(٤): ۱۱-۲٦.

زارع ابیانه، ح. ۱۳۹۲. بررسی نقش عوامل اقلیمی و خشکسالی بر تغییرپذیری عملکرد چهار محصول دیم در مشهد و بیرجند. دانش آب و خاک تبریز، ۲۳ (۱): ۳۹-٥٦.

Biswas, B., Dhaliwal L.K., Singh S.P. and Sandhu S.K. 2014. Forecasting wheat production using ARIMA model in Punjab**.** International Journal of Agricultural Sciences 10(1): 158-161.

Box, G.E.P. and Jenkins, G. 1970. Time Series Analysis, Forecasting and Control, Holden-Day, San Francisco, CA.

Byvatov, E., Fechner, U., Sadowski, J. and Schneider, G.2003. Comparison of Support Vector Machine and Artificial Neural Network Systems for Drug/Nondrug Classification. The Journal of Chemical Information and Computer Scientists 43: 1882-1889.

Chen, K.U. and Wang, C.H. 2007. A hybrid SARIMA and support vector machines in forecasting the production values of the machinery industry in Taiwan. Expert Systems with Applications 32: 254–264.

Dı ´ az-Robles, L.A., Ortega, J.C., Fu, J.S., Reed, G.D., Chow, J.C., Watson,J.G. and Moncada-Herrera, J.A.2008. A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile. Atmospheric Environment. Atmospheric Environment 42: 8331–8340.

Hamidi, O., Poorolajal., J., Sadeghifar, M., Abbasi,H., Maryanaji, Z., Faridi.H.R. and Tapak, L. 2014. A comparative study of support vector machines and artificial neural networks for predicting precipitation in Iran. Theoretical and Applied Climatology 119(3-4): 723–731.

Kumar Paul, R. and Sinha K.2016. Forecasting crop yield: a comparative assessment of ARIMAX and NARX model. RASHI 1(1):77-85.

Lecerf, R., Ceglar, A., Lopez-Lozano, R., Van Der Velde, M. and Baruth B. 2019. Assessing the information in crop model and meteorological indicators to forecast crop yield over Europe. Agricultural Systems. 168 : 191–202.

Liang, Y.H. 2009. Combining seasonal time series ARIMA method and neural networks with genetic algorithms for predicting the production value of the mechanical industry in Taiwan. Neural Computing & Applications 18: 833–841.

Markham, I.S. and Rakes, T.R. 1998.The e5ect of sample size and variability of data on the comparative performance of arti/cial neural networks and regression. Computers & Operations Research 25: 251–263.

Narasimha Murthy, K. V., Saravana,R. and Vijaya Kumar., K. 2018. Modeling and forecasting rainfall patterns of southwest monsoons in North–East India as a SARIMA process. Meteorology and Atmospheric Physics 130: 99–106.

Peter Zhang, G. 2003. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50: 159 – 175.

Ruiz-Aguilar, J. J., Turias, I. J., Jimenez-Come, M. J. and Mar Cerban, M. 2014. Hybrid Approaches of support vector regression and SARIMA models to forecast the inspections volume. International Conference of Hybrid Artificial Intelligence Systems 502-514.

Sharma, P.K., Dwivedi, S., Ali, L. and Arora. R.K. 2018. Forecasting Maize Production in India using ARIMA Model. Agro Economist - An International Journal 5(1): 01-06.

Suman, U. and Verma, S. 2016. Autoregressive Integrated Moving Average models for Sugarcane Yield Estimation in Haryana. International Journal of Computer & Mathematical Sciences 5(12): 33-38.

Verma, U., Koehler, W. and Goyal, M. 2012. A study on yield trends of different crops using ARIMA analysis. Environment and Ecology 30(4A): 1459-1463.

Zaynoddin, M., Bonakdari, H., Azari, A., Ebtehaj, I., Gharabaghi, B. and Riahi Madavar, H. 2018. Novel hybrid linear stochastic with non-linear extreme learning machine methods for forecasting monthly rainfall a tropical climate. Journal of Environmental Management 222: 190-206.

ISSN 2251-7480

Application of hybrid ARIMA and support vector regression model for improvement of time series forecasting

Laleh Parviz[1*] and Bahareh Saeedabadi[2]

1*) Associate Professor, Faculty of Agriculture, Azarbaijan Shahid Madani University, Tabriz, Iran.
Corresponding author email: laleh_parviz@yahoo.com
2) Graduated of BSc, Faculty of Agriculture, Azarbaijan Shahid Madani University, Tabriz, Iran.

Received: 02-11-2019 Accepted: 10-11-2020

Abstract

Accurate investigation related to the structure of time series plays an important role in increasing the accuracy of ARIMA forecasting. The aim of this research is to investigate the effect of modeling decomposition of linear and non linear parts of time series on ARIMA model results. The decomposition of wheat and maize yield time series (in Kermanshah and Esfahan provinces) in the linear part was related to ARIMA and in the non linear part was conducted with support vector regression (hybrid model). The kind of configuration of non linear part of hybrid model is more important for example in the maize time series of Kermanshah, the values of RMSE for configuration with residual was 1.52 and for time series configuration was 15.03. The decreasing of RMSE, MAE and UII for wheat time series of Esfahan with hybrid model was 45.94%, 52.29% and 46%, respectively which is indicative of hybrid model improvement. The value of GMER in all four time series was greater than one which indicates the overestimation of hybrid model. Comparison the average of each criteria with two models and crops in each province indicated the effect of climate on modeling process because the average of criteria in Esfahan province decreased rather to Kermanshah (RMSE decreasing= 24.72%, UII decreasing=12.24%). Therefore, decomposition of time series to linear and non linear parts of time series can increase the accuracy of ARIMA model results.

Keywords: ARIMA; Nonlinear; Support Vector Regression; Hybrid.