

УДК 621.391:537.86:519

ДЕКОМПОЗИЦІЯ МОДЕЛІ НЕСТАЦІОНАРНОГО ВИПАДКОВОГО ПРОЦЕСУ

DOI 10.36994 / 2788-5518-2023-01-05-11

Тихонов В.А., д.ф.-м.н., проф. Харківський національний університет радіоелектроніки, Україна.
vyacheslav.tykhonov@nure.ua

Безрук В.М., д.т.н., проф. Харківський національний університет радіоелектроніки, Україна.
valeriy_bezruk@ukr.net

Анотація: Розглядаються особливості виконання декомпозиції моделі нестационарного випадкового процесу, яка полягає в роздільному визначенні довільного тренду, сезонної складової та стаціонарної складової. Це дозволяє вирішувати завдання прогнозування, що означає на кілька кроків уперед передбачення значень випадкового процесу. Показано деякі проблеми, що виникають при використанні моделі авторегресії-проінтегрованого ковзного середнього для декомпозиції моделі випадкового нестационарного процесу. Наведено результати досліджень на реальних даних, що демонструють процес декомпозиції моделі нестационарного випадкового процесу.

Ключові слова: випадковий процес, модель, декомпозиція, тренд, сезонна складова, стаціонарна складова, прогнозування.

DECOMPOSITION OF THE NON-STATIONARY RANDOM PROCESS MODEL

Vyacheslav Tikhonov, Dr.habil., Prof. Kharkiv National University of Radio Electronics, Ukraine.
vyacheslav.tykhonov@nure.ua

Valeriy Bezruk, Dr.habil., Prof. Kharkiv National University of Radio Electronics, Ukraine.
valeriy_bezruk@ukr.net

Abstract: In the work, the peculiarity of the task of constructing the decomposition of a non-stationary model of a random process is studied using the examples of the analysis of two types of real data. The peculiarities of the decomposition of the model of a non-stationary random process are considered. It consists in the separate definition of an arbitrary trend, a seasonal component and a stationary component. This allows solving, for example, forecasting tasks. This task means several steps ahead in predicting the values of a random process. Some problems that arise when using the autoregression-integrated moving average (ARIMA) model for the decomposition of the model of a random non-stationary process are shown. The results of research on real data are given. They demonstrate the process of decomposition of a model of a non-stationary random process. The analyzed classes of random non-stationary processes with a trend allow us to solve two main problems with acceptable accuracy. The first task involves the decomposition of a model of a random non-stationary process. It includes the calculation of an arbitrary trend, a quasi-periodic seasonal component and a stationary component of the process. The second task involves the actual forecasting of a non-stationary time series using the ARIMA model. In the ARIMA model, it is assumed that the trends are deterministic and are linear, quadratic or higher order polynomials. The seasonal component is periodic and has equal counts in each period, which can be removed through the subtraction period. If the seasonal component does not satisfies these requirements, then when it is removed by difference operators, problems arise and the ARIMA model may lose accuracy.

Key words: random process, non-stationary model, decomposition, trend, seasonal component, stationary component, forecasting

Вступ

При вирішенні практичних завдань обробки реальних даних, які мають випадковий характер, виникає необхідність у роздільному визначенні тренду, сезонної складової та стаціонарної складової аналізованого випадкового процесу. Питання побудови математичних моделей випадкових процесів розглядаються у ряді робіт [1-3]. Однак цим питанням не приділено належної уваги. У даній статті детально розглядаються практичні особливості виконання декомпозиції моделі нестационарного випадкового процесу, яка полягає визначенні тренду, сезонної складової та стаціонарної складової процесу. При цьому може вирішуватися завдання прогнозування, тобто на кілька кроків уперед передбачати значення випадкового процесу. Наведено результати досліджень на реальних даних, які демонструють процес декомпозиції

моделі нестационарного випадкового процесу. Досліджено питання побудови моделі авторегресії-проінтегрованого ковзного середнього (АРПКС) та її зв'язок з декомпозицією нестационарного процесу. Розглянуто вплив оператора взяття різниці складових нестационарного процесу на точність декомпозиції та прогнозування. Показано, що точна компенсація тренду та сезонної складової можлива у разі їх детермінованості та строгої періодичності. У разі невиконання цих вимог модель АРПКС та декомпозиція неточні, а похибка компенсуються моделлю стаціонарної складової, що представляється у загальному випадку моделлю авторегресії-ковзного середнього.

Виконання досліджень

В роботі проведено дослідження особливостей завдання декомпозиції нестационарної моделі випадкового процесу на прикладах аналізу двох типів реальних даних. Проаналізовані класи випадкових нестационарних процесів з трендом дозволяють вирішувати з прийнятною точністю дві основні задачі. До першої задачі відноситься декомпозиція моделі випадкового нестационарного процесу, тобто обчислення довільного тренду, квазіперіодичної сезонної складової та стаціонарної складової процесу. До другої задачі належить власне прогнозування нестационарного часового ряду з використанням моделі АРПКС. У моделі АРПКС вважається, що тренди детерміновані і є лінійними, квадратичними або є поліномами вищого порядку. Сезонна складова періодична і має рівні відліки в кожному періоді, які можна видалити через період віднімання. Якщо сезонна складова не задовольняє цим вимогам, то при її видаленні операторами взяття різниці виникають проблеми і модель АРПКС може втратити точність. Приклади реальних нестационарних даних з трендом, що досить сильно відрізняються, представлені на рис. 1 та рис. 2. Перший випадковий процес (рис. 1) є трафік передачі даних у системі мобільного зв'язку [4]. Довжина часового ряду складає 365 відліків, що відповідає 365 дням. Він має значний зростаючий тренд. Дані містять сезону складову з періодом, рівним приблизно 7 днів, і стаціонарний процес, який описується в загальному випадку моделлю АРКС. Другий випадковий процес (рис. 2) є даними про середньомісячні температури, зафіксовані у метеорологічній обсерваторії міста Пулково. Повна довжина ряду становить 3108 значень температури, отриманих обсерваторією за 259 років з 1752 по 2010 роки. Дані містять слабкий тренд, яскраво виражені сезонні коливання температур з періодом 12 місяців та стаціонарну складову процесу. У даній статті шляхом усереднення знайдено тренди процесів, що представлені на рис. 1 та рис. 2. Для цього щоденний трафік передачі даних (рис. 1) представлено у вигляді ряду середньотижневих даних:

$$x^n[t] = \overline{x_0[t]}, \overline{x_1[t]} \dots \overline{x_{m-1}[t]}, \quad (1)$$

де

$$\overline{x_k[t]} = \frac{1}{n} \sum_{i=1}^n x_i[t], \quad n = 7. \quad (2)$$

У процесах (1) та (2) час визначено як номер даних середнього значення трафіку за тиждень. Для оцінки довготривалого тренду середньорічні температури, що представлені на рис. 2, також перетворені за допомогою виразу (1), де середньорічні значення температур обчислюються як:

$$\overline{x_k[t]} = \frac{1}{n} \sum_{i=1}^n x_i[t], \quad n = 12. \quad (3)$$

Для даних у виразі (3) час визначено як послідовність років.

Для отримання трендів середньотижневі значення даних (1) згладжені низькочастотним фільтром, що ковзає. Щоб накласти тренд на вихідні дані, тобто щоб довжина вихідних даних та даних, отриманих усередненням збігалася, продовжені значення трендів на 7 або 12 відліків відповідно. Тоді отримаємо часові ряди трендів, що дорівнюють початковій довжині даних.

Отримані тренди показані на рис. 1 та рис. 2. Зауважимо, що процедура ковзного усереднення не дає можливості отримати на початку часового ряду і в його кінці таке ж точне усереднення як і у середині часового ряду. Це зв'язано з очевидними особливостями процедури усереднення,

коли на початку часового ряду відсутні необхідні для усереднення попередні значення часового ряду, а кінці ряду – наступні значення. Як видно із графіків на рис. 1 та рис. 2, знайдені тренди мають довільну форму, а не детерміновану (лінійну, параболічну), як це визначено в моделі АРПКС. У цілому нині тренди неоднорідні, тк як у них є ділянки з різною швидкістю зміни.

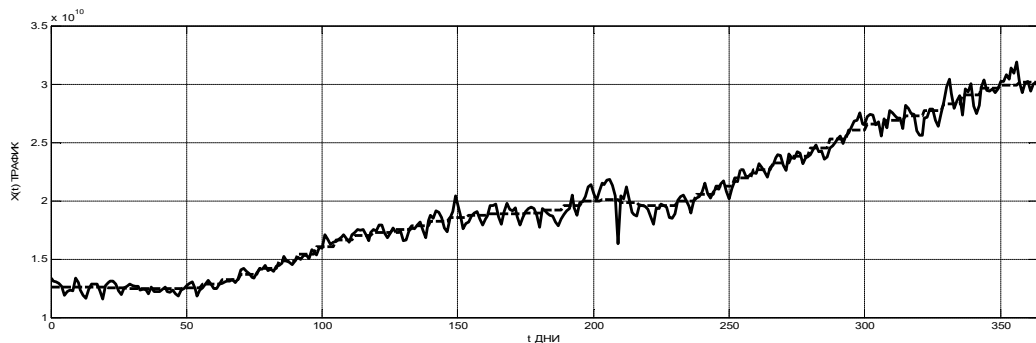
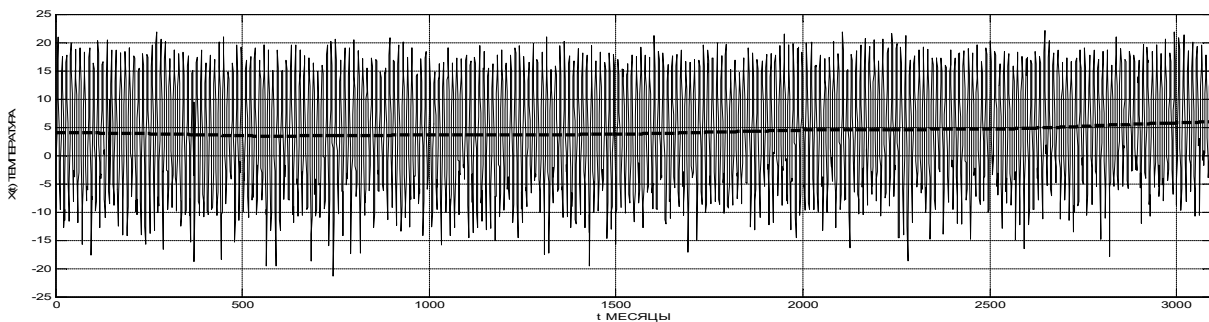


Рис. 1. Трафік передачі даних у системі стільникового мобільного зв'язку та оцінка його тренду

Аналіз тренду на рис. 2 показує, що з 1752 до 1803 температура падала з $4,05^{\circ}\text{C}$ до $3,45^{\circ}\text{C}$. Спад температури склав $0,6^{\circ}\text{C}$. З 1804 року спостерігалось зростання температури з $3,45^{\circ}\text{C}$ до $5,96^{\circ}\text{C}$



у 2010 році. Приріст температури за цей термін становив $2,51^{\circ}\text{C}$

Рис.2. Дані середньомісячних температур, зафіксованих у метеорологічній обсерваторії міста Пулково, та оцінка тренду

Як показує аналіз рис. 1 та рис. 2, отримані тренди мають довільний характер.

Модель АРПКС нестационарних даних та їх декомпозиція

Розглянемо особливості побудови моделей АРПКС для нестационарних процесів. Відповідно до [1] моделюємий нестационарний процес може мати детермінований тренд. У рамках моделі нестационарних процесів сезонна складова має бути періодичною функцією з певним періодом. Завданням побудови моделі АРПКС є видалення з вихідних даних тренду та сезонної складової за допомогою відповідних операторів взяття різниці та побудова моделі лінійного передбачення АРКС стаціонарного залишку. Після цього можна вирішувати завдання прогнозування [1] шляхом відновлення тренду та сезонної складової нестационарного процесу, а також побудови моделі АРКС стаціонарної складової.

Модель АРПКС нестационарного процесу представляється у вигляді мультиплікативного процесу $\omega[t]$ [2]:

$$\omega[t] = \nabla^d \nabla_s x[t] = (1-z)^d (1-z^s) x[t] = \nabla^d \omega_1[t]. \quad (4)$$

Оператори взяття різниці, що входять до (4), враховують сезону складову та тренд шляхом їх віднімання. Для видалення сезонної складової застосовується операція $\nabla_s = 1 - z^s$, где z^s - оператор зсуву, дія якого визначається виразом $z^s x[t] = x[t-s]$. Процес видалення сезонної складової за допомогою оператора зсуву визначається виразом:

$$\omega_1[t] = \nabla_s x[t] = (1 - z^s)x[t]. \quad (5)$$

Детермінований тренд у $\omega_1[t]$ визначається оператором $\nabla^d = (1 - z)^d$. Тоді процес без тренда можна представити у вигляді:

$$\omega[t] = (1 - z)^d \omega_1[t]. \quad (6)$$

Як правило, обмежуються найпростішими випадками лінійного та квадратичного трендів. Для цих випадків вважають $d = 1$, $d = 2$. Тоді:

$$\begin{aligned} \omega[t] &= \nabla \omega_1[t] = \omega_1[t] - \omega_1[t-1], \\ \omega[t] &= \nabla^2 \omega_1[t] = \nabla(\omega_1[t] - \omega_1[t-1]) = (\omega_1[t] - 2\omega_1[t-1] + \omega_1[t-2]). \end{aligned}$$

Таким чином, тренд (5) та сезонна складова (6), що визначаються операторами зсуву, видаляються. Під час прогнозування ці детерміновані складові відновлюються.

При ширшому аналізі нестационарних процесів інтерес представляє також процедура знаходження довільного тренду та сезонної складової, що виходить за межі побудови АРПКС моделі. При цьому форма тренду та властивості сезонної складової можуть змінюватися у ширших межах, ніж у моделі АРПКС. Це завдання називають декомпозицією моделі нестационарного часового ряду, що полягає у його розкладанні на складові [3].

Декомпозиція у моделі нестационарного процесу розуміється як представлення у вигляді:

$$x[t] = m[t] + c[t] + y[t], \quad (7)$$

где $x[t]$ - нестационарний процес, $m[t]$, $c[t]$, $y[t]$ - відповідно тренд, сезонна складова та корельована стаціонарна складова.

Похибки декомпозиції моделі нестационарного процесу

Зауважимо, що оператор $\nabla_s = 1 - z^s$, що усуває сезонні коливання в (4), діє не тільки на цю складову процесу, а й на трендову складову. При цьому стаціонарна складова нестационарного процесу також змінюється. Після дії оператором $\nabla^d = (1 - z)^d$ на тренд інші складові моделі декомпозиції нестационарного процесу можуть змінитися. Найбільший вплив оператора видалення сезонної складової. Після видалення сезонної складової оператором $\nabla_s = 1 - z^s$ декомпозиція (7) моделі нестационарного процесу набуває вигляду:

$$\nabla_s x[t] = m[t] - m[t-s] + c[t] - c[t-s] + y[t] - y[t-s], \quad (8)$$

де $\nabla_s x[t]$ - нестационарний процес без сезонної складової.

Якщо сезонна складова строго детермінована, то $c[t] = c[t-s]$ та декомпозиція (8) спрощується [2] :

$$\nabla_s x[t] = m[t] - m[t-s] + y[t] - y[t-s]. \quad (9)$$

Однак, як видно із рис. 1 та рис. 2, сезонна складова не детермінована і тому вираз (9) є несправедливим для цих даних.

На рис. 3 та рис. 4 показані дані трафіку та середньомісячних температур після видалення сезонних складових. На даних, що представлені на рис. 3, не видно тренд, який добре проглядається на рис. 1. За графіком на рис. 4 також важко визначити довготривалий тренд температур. Таким чином, видалення сезонної складової оператором $\nabla_s = 1 - z^s$ з вихідних даних також суттєво впливає і на тренд.

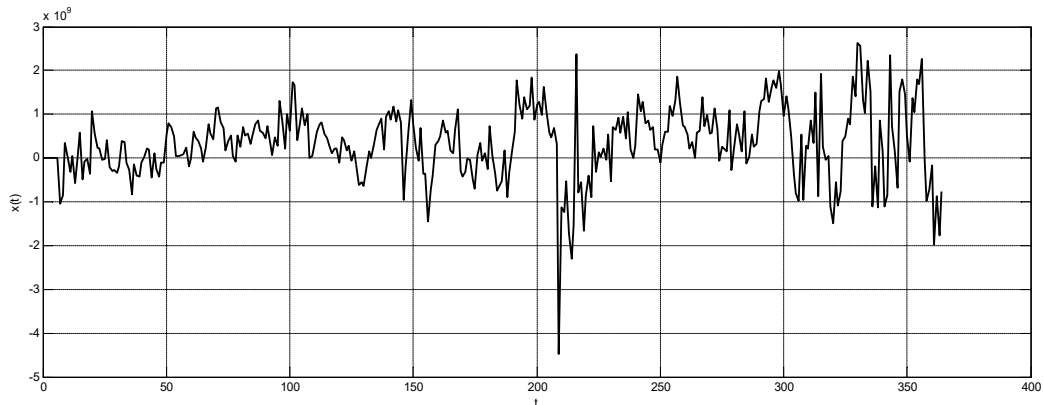


Рис. 3 Трафік передачі даних у системі стільникового мобільного зв'язку після видалення сезонної складової

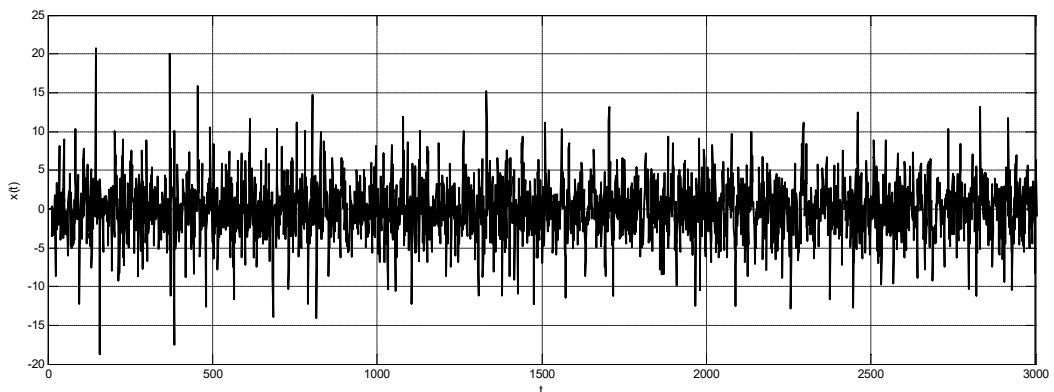


Рис. 4. Вибірка даних середньомісячних температур після видалення сезонної складової

Операція видалення тренду оператором $\nabla^d = (1 - z)^d$ при $d = 1$ менше спотворює дані. Однак слід мати на увазі, що дія оператора $\nabla = (1 - z)$ по суті зводиться до диференціювання випадкового процесу. Таким чином, доводиться досліджувати властивості не безтрендового процесу, а його дискретної похідної. На рис. 5 та рис. 6 показані дані трафіку та середньомісячних температур після видалення тренду. Аналіз даних на рис. 5 та рис. 6 показує, що тренд дійсно видаляється.

Наявність сезонної складової даних, можна визначити за графіками спектральної щільності потужності (СЩП) аналізованих даних [3]. Вони показані на рис. 7 та рис. 8. Аналіз графіка спектра, поданого на рис. 7 показує, що сезонна складова трафіку даних менше виражена. Форма спектру показує, що в отриманих після диференціювання даних трафіку, є постійна складова, сезонна складова з періодом 6,8 днів, а також слабкі коливання з періодами 3,4 і 2,3 дня. Як видно із рис. 8 більш явно сезонна складова виражена в даних температур. З аналізу спектра, показаного на рис. 8, слідує, що період коливань сезонної складової дорівнює 11,9 днів. Зазначимо, що для деяких форм сезонної складової диференціювання не змінює період коливань.

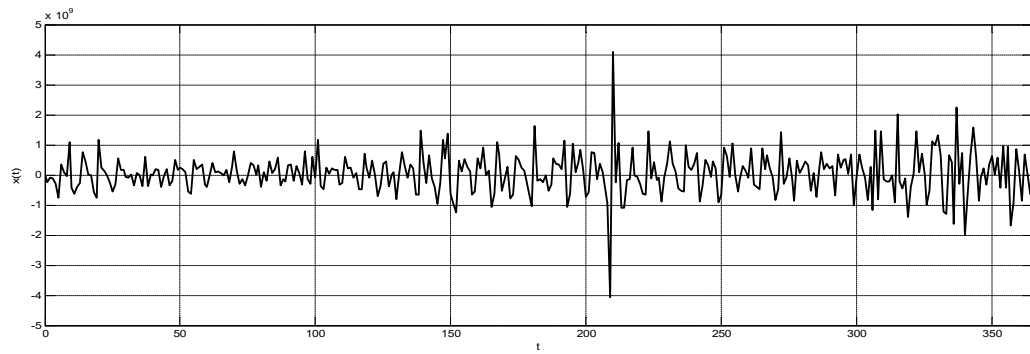


Рис. 5. Трафік передачі даних у системі мобільного зв'язку після видалення тренда

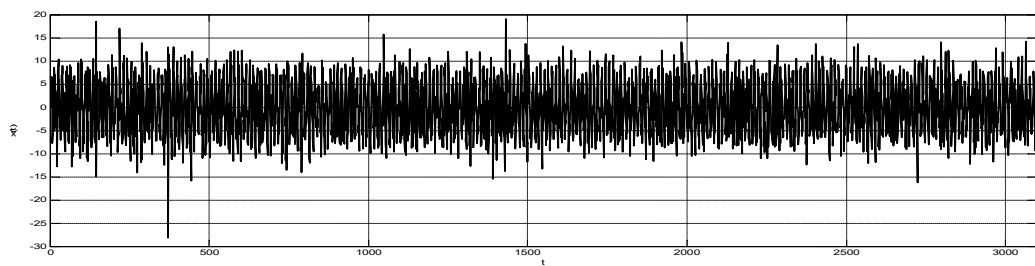


Рис. 6. Дані про середньомісячні значення температури після видалення тренду

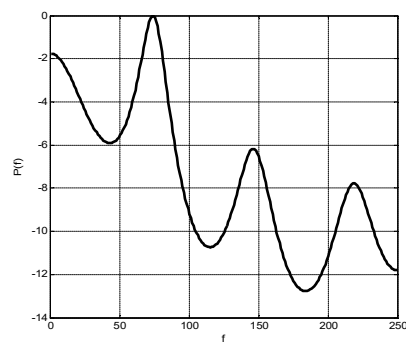


Рис. 7. СЩП після видалення тренду в даних трафіку у системі мобільного зв'язку

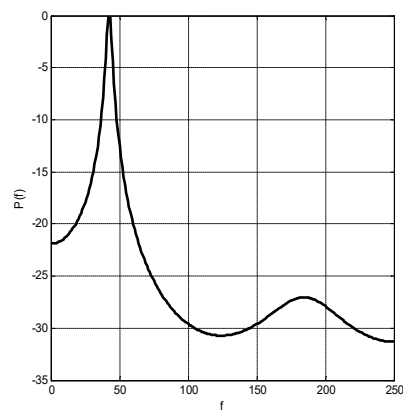


Рис. 8. СЩП після видалення тренду з даних температур

Висновки

У роботі розглянуто теоретичні і практичні особливості декомпозиції моделі нестационарного процесу, яка полягає у роздільному визначенні тренду, сезонної складової та стаціонарної складової. Це дозволяє вирішити задачу прогнозування випадкового процесу. Проведено дослідження особливостей задачі декомпозиції нестационарної моделі випадкового процесу на прикладах аналізу двох типів реальних даних. Розглянуто випадковий процес, що являє собою трафік передачі даних у системі мобільного зв'язку. Розглянуто також випадковий процес, що являє собою дані про середньомісячні температури, зафіксовані у метеорологічній обсерваторії міста Пулково. Показано деякі проблеми, що виникають при використанні моделі у вигляді АРПКС та декомпозицію моделі для аналізу нестационарних даних, що містять трендову, сезонну та стаціонарну складові. Використання в моделі АРПКС операторів, що визначають ці складові, обмежує допустимі властивості сезонної складової та тренду. Вплив на модель аналізованого процесу операторів зсуву продемонстровано на задачі декомпозиції нестационарного процесу. Застосування цих операторів зводиться до диференціювання вихідних даних, що спотворює характеристики вихідного процесу та призводить до неточностей. Для виділення тренду використані усереднені за деякий період дані з подальшим усередненням низькочастотним фільтром, що ковзає. Наведені результати декомпозиції моделей для вибраних реальних даних, що мають нестационарний характер.

Література

1. Бокс Дж., Дженкинс Г. Анализ временных рядов: Пер. с. англ. М. : Мир, 1974. 406 с.
2. Brockwell P. J., Davis R. A. Introduction to Time Series and Forecasting. Springer, 2002. 434 p.
3. Марпл С. Л. Цифровой спектральный анализ и его приложения. М. : Мир, 1990. 584 с.
4. Тихонов В. А., Чеботарёва Д. В. Прогнозирование потока данных в сетях мобильной связи // *Наукоємні технології в інфокомунікаціях*. Матеріали III Міжнародн. наук.-практ. конф. Харків : Друкарня «Мадрид», 2019, С. 126-127.

References

1. Boks Dzh., Dzhenskyns H. Analyz vremennykh riadov: Per. s. anhl. M. : Myr, 1974. 406 s.
2. Brockwell P. J., Davis R. A. Introduction to Time Series and Forecasting. Springer, 2002. 434 p.
3. Marpl S. L. Tsyfrovoi spektralnyi analiz y eho prylozheniya. M. : Myr, 1990. 584 s.
4. Tykhonov V. A., Chebotarëva D. V. Prohnozyrovanye potoka dannykh v setiakh mobylnoi svyazy // *Naukoiemni tekhnolohii v infokomunikatsiiakh*. Materialy III Mizhnarodn. nauk.-prakt. konf. Kharkiv : Drukarnia «Madrid», 2019, S. 126-127.