

Методика выбора методов прогнозирования временных рядов киберфизических систем предприятий ТЭК

Д. П. Плахотников

Санкт-Петербургский государственный электротехнический университет
«ЛЭТИ» им. В. И. Ульянова (Ленина), Санкт-Петербург, Россия

dimapl21@yandex.ru

Аннотация. На текущий момент создано множество моделей прогнозирования временных рядов. Исходные данные, генерируемые в процессе эксплуатации киберфизических систем, могут быть использованы для прогнозирования будущего состояния системы. Представлены способы получения и обработки данных параметров киберфизических систем, их очистки и построения прогнозной модели для разных методов. Методы сравнивались по качеству прогноза и длительности построения модели. В результате выявлены наиболее рациональные методы прогнозирования временных рядов киберфизических систем в зависимости от прогнозируемого параметра.

Ключевые слова: киберфизические системы, большие данные, анализ данных, прогнозирование

Для цитирования: Плахотников Д. П. Методика выбора методов прогнозирования временных рядов киберфизических систем предприятий ТЭК // Изв. СПбГЭТУ «ЛЭТИ». 2022. Т. 15, № 7. С. 20–27. doi: 10.32603/2071-8985-2022-15-7-20-27.

Original article

Methodology for Choosing Methods for Forecasting Time Series of Cyber-Physical Systems of Fuel and Energy Complex Enterprises

D. P. Plakhotnikov

Saint Petersburg Electrotechnical University, Saint Petersburg, Russia

dimapl21@yandex.ru

Abstract. At the moment, many time series forecasting models have been created. The initial data generated during the operation of cyber-physical systems can be used to predict the future state of the system. The article presents methods for obtaining and processing data on the parameters of cyber-physical systems, their cleaning and building a predictive model for different methods. The methods are compared in terms of the quality of the forecast and the duration of the model building. As a result, the most rational methods for forecasting the time series of cyber-physical systems depending on the predicted parameter were identified.

Keywords: cyber-physical systems, big data, data analysis, forecasting

For citation: Plakhotnikov D. P. Methodology for Choosing Methods for Forecasting Time Series of Cyber-Physical Systems of Fuel and Energy Complex Enterprises // LETI Transactions on Electrical Engineering & Computer Science. 2022. Vol. 15, no. 7. P. 20–27. doi: 10.32603/2071-8985-2022-15-7-20-27.

Введение. С появлением концепции киберфизических систем связывают переход к индустрии 4.0, появление которой началось с повсеместным внедрением данных систем в производство и промышленность [1].

Особенность четвертой промышленной революции заключается в сближении производствен-

ных технологий, новых информационных технологий (сбор и аналитика «больших» данных) и появление технологии искусственного интеллекта (включающие машинное обучение и глубокое обучение). Происходит интеллектуализация промышленности и производства. Оборудование становится «умным» благодаря сенсорам и датчи-

кам, которые собирают информацию о физическом мире, и благодаря сбору этих данных можно выполнять необходимые расчеты [2].

Принципиальная особенность киберфизических систем заключается в том, что в процессе своей работы они генерируют большие и разнообразные объемы информации – данные с датчиков, сенсоров, систем контроля и других источников. Эта информация в дальнейшем может быть использована для прогнозирования различных параметров, чтобы предугадать будущую нагрузку на киберфизическую систему и предотвратить возможные проблемы, тем самым снизив участие человека в процессе управления.

Киберфизические системы. Термин «киберфизические системы» впервые был применен в 2006 г. Он обозначал комплексы, которые содержали объекты реального мира, искусственные подсистемы и контроллеры [3]. Данные системы должны были обеспечить оптимизацию процессов производства за счет автоматизации управления ими.

Киберфизические системы – это следующий этап эволюции распределенных систем, основное отличие которых состоит в том, что киберфизические системы контролируются с помощью специальных алгоритмов и интегрированы через сети связи (например, через глобальную сеть Интернет или корпоративную локальную сеть).

Встроенные системы реального времени, автоматизированные системы управления и распределенные вычислительные системы – это предшественники современных киберфизических систем.

На рис. 1 представлен их эволюционный путь [4].

В большинстве случаев под киберфизическими системами сейчас понимают традиционные робототехнические системы и роботизированные производственные линии. Такие системы решают весьма сложные технологические задачи и, как правило, оснащены набором интеллектуальных функций, но круг задач, решаемых ими, фиксирован.

Основная идея создания киберфизических систем заключается в глубокой интеграции объектов реального мира и проходящих в них физических процессов (за счет применения различных датчиков, сенсоров) с вычислительными блоками, которые представляют собой различные процессы, сервисы, приложения, управляющие реальными физическими объектами. Произошло «оцифровывание» процессов производства и реализации продукции.

Структурная схема киберфизической системы представлена на рис. 2 [5].



Рис. 1. Эволюционный путь киберфизических систем

Fig. 1. Evolutionary path of cyber-physical systems

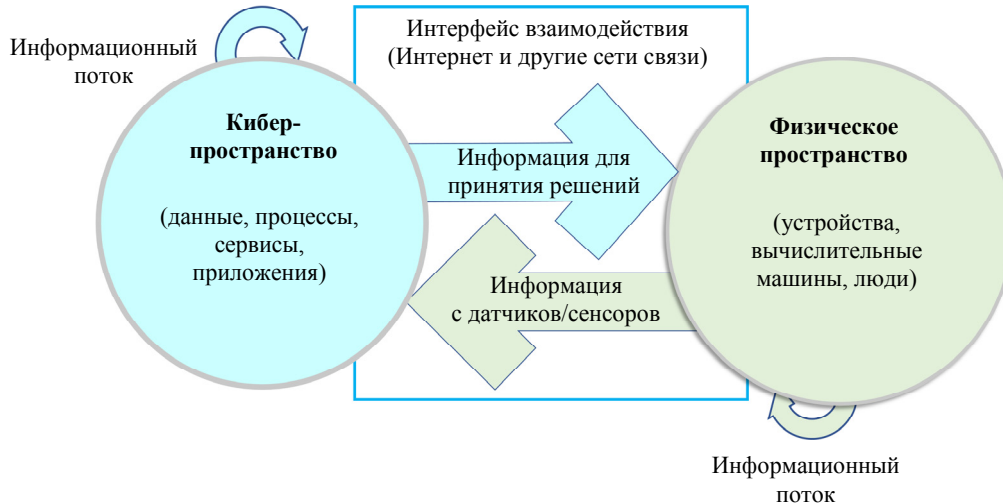


Рис. 2. Структурная схема киберфизической системы
Fig. 2. Block diagram of a cyber-physical system

Постановка задачи. Основная задача состоит в поиске рационального метода прогнозирования временных рядов киберфизических систем для различных параметров. Задача состоит из трех этапов:

1. Получение и очистка данных.
2. Применение алгоритмов прогнозирования и подсчет метрик.
3. Сравнение точности прогноза и времени выполнения расчета для каждого параметра.

Предлагаемая методика выбора заключается в минимизации средней абсолютной ошибки и времени построения модели.

Средняя абсолютная процентная ошибка (MAPE) – это одна из мер точности прогноза, определяемая формулой [6]

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|,$$

где t – текущая точка; n – количество подогнанных точек; A_t – фактическое значение в текущей точке; F_t – прогнозируемое значение n .

Данный показатель чаще всего используется в качестве функции потерь для регрессии.

Длительность выполнения прогноза представляет собой фактическое время, прошедшее от запуска функции до ее завершения, т. е. разность между временем завершения задачи и временем ее начала [7].

Для оценки среднего времени проводится несколько построений модельных прогнозов, после чего высчитывается среднее время прогноза. В данном случае было сделано 10 построений.

Далее обе меры нормализуются с помощью формулы

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)},$$

где x' – это нормализованное значение, которое располагается на отрезке $[0; 1]$; $\min(x)$, $\max(x)$ – минимальное и максимальное значения меры.

Далее необходимо найти максимальную оценку с помощью определенных весовых коэффициентов, учитывающих как точность прогноза, так и длительность его построения. В зависимости от приоритета точности над длительностью весовые коэффициенты могут меняться. В данной методике выбора оценка представлена следующей предлагаемой эмпирической формулой:

$$\text{Оценка} = (0.8MAPE(\text{норм.}) + 0.2T(\text{норм.}))^{-1}.$$

Получение и обработка данных. В качестве исходных были получены данные из транзакционной базы данных киберфизической системы предприятия топливно-энергетического комплекса. Данные были получены с помощью разработки скрипта загрузки данных для ETL-процесса.

Данные представляют собой транзакционную таблицу процессинговой базы данных предприятия топливно-энергетического комплекса и содержат следующие параметры:

- метка времени транзакции (дата и время);
- объем реализации продукции;
- метка времени старта реализации;
- метка времени завершения реализации.

Эти параметры собираются непосредственно с объекта реализации топливно-энергетического комплекса за счет специальных контроллеров.

Табл. 1. Пример исходных данных
Tab. 1. Initial data example

Дата данных («date»)	Объем реализации продукции («sum_release»), тыс. м ³	Количество реализаций продукции («count_release»), тыс. м ³	Медиана длительности реализации продукции («mid_duration_release»), тыс. м ³	Средняя скорость реализации продукции («avg_speed_release»), тыс. м ³
2019-01-01	89 735.309	2229	5.5073501	7.241275
2019-01-02	104 142.73	2667	5.4002748	7.149088
...
2022-05-13	277820.786	9207	5.6292366	7.0928477
2022-05-14	244648.228	8173	5.2968092	7.1509133

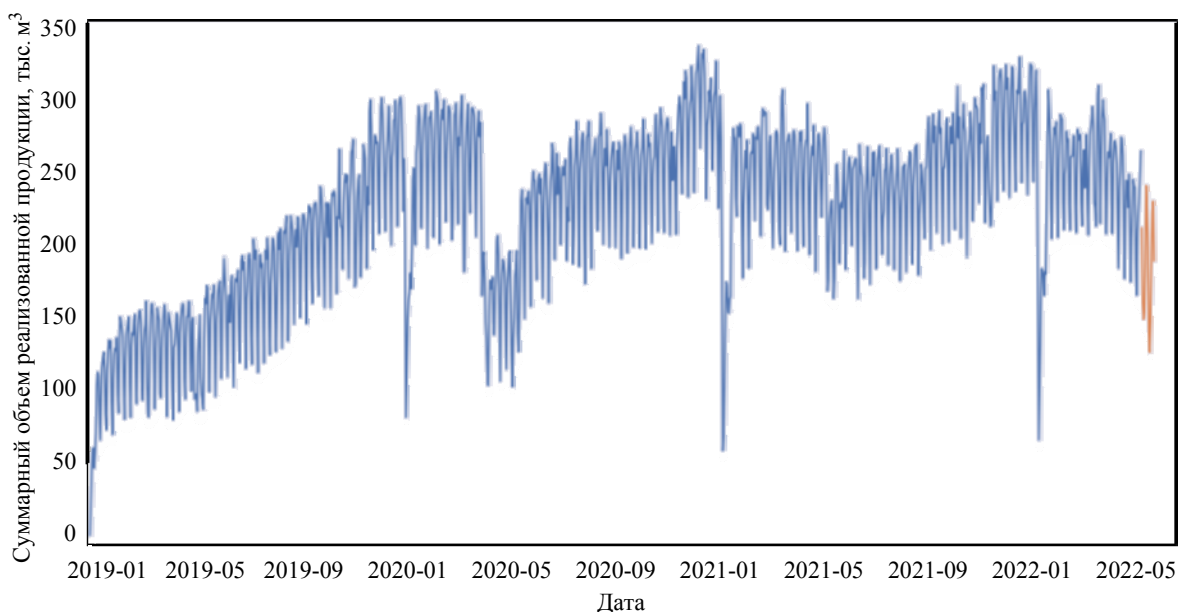


Рис. 3. Набор данных для прогноза
Fig. 3. Forecast dataset

К ним подключены датчики температуры, давления, отпущенного объема. Также эти датчики содержат вычислительный блок и с их помощью происходит управление объектом реализации продукции.

Первый этап обработки заключался в очистке данных. Были удалены данные с пустыми объемами реализации, а также с объемами ниже 5-го и выше 95-го перцентилей. Была вычислена оценка длительности реализации как разница меток времени старта и завершения реализации продукции.

Полученные данные были сгруппированы посуточно, для каждого суток были рассчитаны медиана длительности реализации, суммарная реализация и средняя скорость реализации продукции. Формат данных был стандартизирован. Пример исходных данных для прогнозирования представлен в табл. 1.

Набор содержит 1231 строку.

Методы прогнозирования данных. Помимо классических методов прогнозирования временных рядов, ориентированных на линейные отно-

шения, распространены следующие модели прогнозирования:

1. Autoregression (авторегрессия).
2. Moving Average (скользящее среднее).
3. Autoregressive Moving Average (авторегрессионное скользящее среднее).
4. Autoregressive Integrated Moving Average (авторегрессионное интегрированное скользящее среднее).
5. Seasonal Autoregressive Integrated Moving Average (сезонное авторегрессионное интегрированное скользящее среднее).
6. Simple Exponential Smoothing (простое экспоненциальное сглаживание).
7. Holt Winter’s Exponential Smoothing (экспоненциальное сглаживание Холта–Винтера).

Проведение эксперимента. После загрузки исходных данных для прогнозирования они были разделены на два набора данных – тестовый и валидационный наборы данных временных рядов параметров киберфизических систем. В качестве тестового набора были использованы данные с

Табл. 2. Результаты параметров прогноза объема реализации продукции
 Tab. 2. The results of the parameters of the forecast of the volume of sales

Метод прогнозирования	Средняя абсолютная ошибка, %	Средняя длительность прогнозирования, с	Оценка прогноза
Autoregression (авторегрессия)	0.12034561544837583	0.010322394599999996	5.872263061
Moving Average (скользящее среднее)	0.10101472956349483	0.49403181180000005	7.881713595
Autoregressive Moving Average (авторегрессионное скользящее среднее)	0.12121096417006806	0.77281421249999999	2.646095997
Autoregressive Integrated Moving Average (авторегрессионное интегрированное скользящее среднее)	0.1216275323503491	0.41017202419999999	3.490832227
Seasonal Autoregressive Integrated Moving-Average (сезонное авторегрессионное интегрированное скользящее среднее)	0.1216275323503491	0.3981382101000001	3.529724914
Simple Exponential Smoothing (простое экспоненциальное сглаживание)	0.19182756725232344	0.016607510900000122	1.247429402
Holt Winter's Exponential Smoothing (экспоненциальное сглаживание Холта–Винтера)	0.19182756725232344	0.01678438729999989	1.247357213

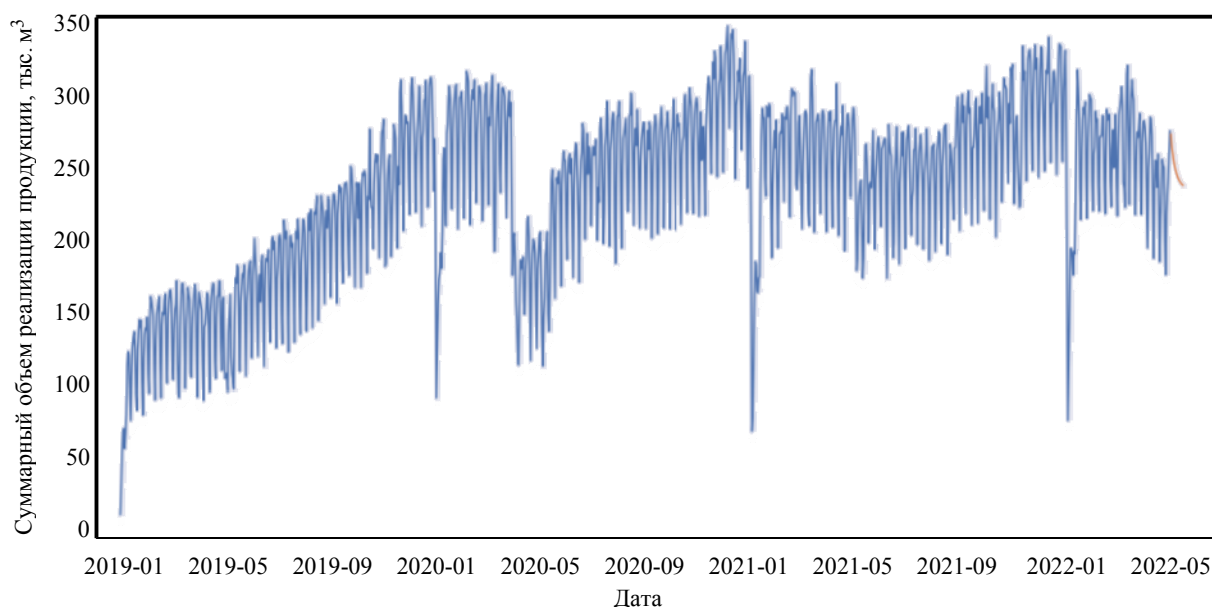


Рис. 4. Прогнозирование объема реализации методом авторегрессии
 Fig. 4. Forecasting the volume of sales by the autoregressive method

01.01.2019 по 30.04.2022, а в качестве валидационного – данные с 01.05.2022 по 14.05.2022. Визуально они отображены на рис. 3.

Далее были проведены эксперименты с описанными методами прогнозирования. Каждый эксперимент был проведен для следующих параметров:

- объем реализации продукции (табл. 2);
- количество реализованной продукции (табл. 3);
- медиана длительности реализации продукции (табл. 4);
- средняя скорость реализации продукции (табл. 5).

Как видно из таблицы, наиболее рационален метод скользящего среднего, поскольку он одновременно быстр и точен (рис. 4).

При прогнозе параметра количества реализации продукции наиболее рационально точен также метод скользящего среднего (рис. 5).

Наиболее рациональным методом прогнозирования длительности является экспоненциальное сглаживание Холта–Винтера (HWES) (рис. 6).

По эффективности наиболее рациональный метод – также экспоненциальное сглаживание Холта–Винтера (рис. 7).

Обсуждение результатов. В результате исследования было показано, что в зависимости от параметров киберфизической системы наиболее рациональный метод прогнозирования меняется в пределах одной и той же системы. Выбор метода прогноза при прочих равных условиях связан с

Табл. 3. Результаты параметров прогноза количества реализации продукции
 Tab. 3. The results of the parameters of the forecast of the count of sales

Метод прогнозирования	Средняя абсолютная ошибка, %	Средняя длительность прогнозирования, с	Оценка прогноза
Autoregression (авторегрессия)	0.15555809345115965	0.01199998269999867	4.166278142
Moving Average (скользящее среднее)	0.12953814184656734	0.21548906639999926	12.52543009
Autoregressive Moving Average (авторегрессионное скользящее среднее)	0.1515676735721979	0.5217576412	2.480081548
Autoregressive Integrated Moving Average (авторегрессионное интегрированное скользящее среднее)	0.21626322633743214	0.3149406354999996	1.088308942
Seasonal Autoregressive Integrated Moving-Average (сезонное авторегрессионное интегрированное скользящее среднее)	0.21626322633743214	0.32137454699999923	1.085327308
Simple Exponential Smoothing (простое экспоненциальное сглаживание)	0.21193330442501027	0.01779046750000077	1.311767404
Holt Winter's Exponential Smoothing (экспоненциальное сглаживание Холта–Винтера)	0.21193330442501027	0.01478994999999941	1.313796241

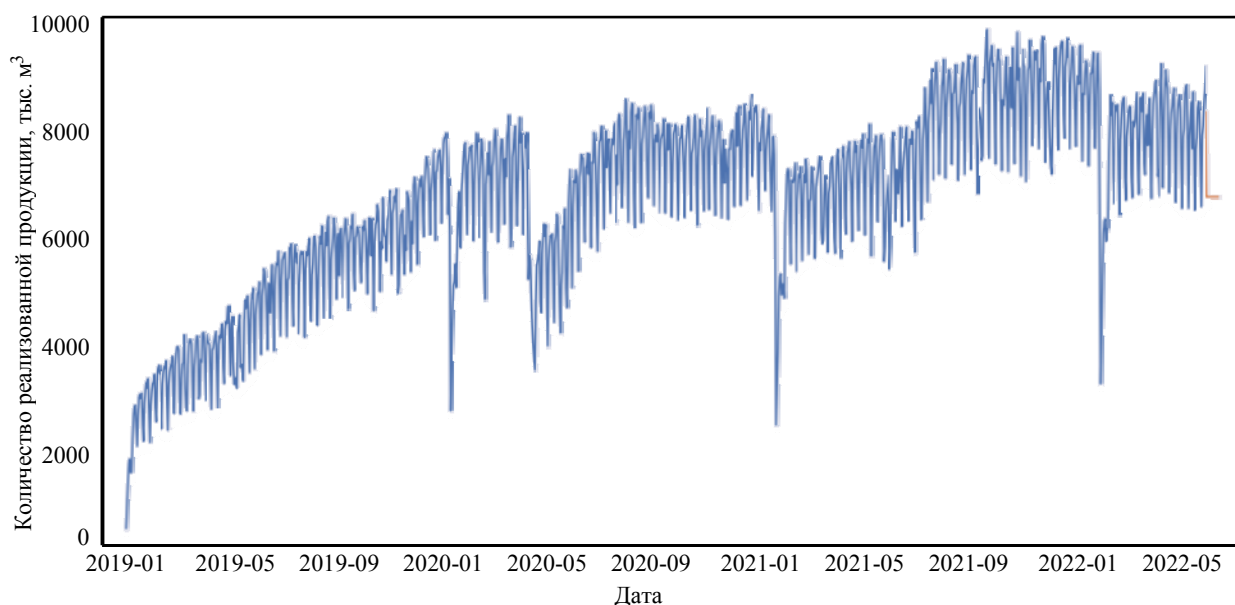


Рис. 5. Прогнозирование количества реализации методом авторегрессии
 Fig. 5. Predicting the number of sales using the autoregressive method

Табл. 4. Результаты параметров прогноза медианы длительности реализации продукции
 Tab. 4. The results of the forecast parameters of the median duration of product sales

Метод прогнозирования	Средняя абсолютная ошибка, %	Средняя длительность прогнозирования, с	Оценка прогноза
Autoregression (авторегрессия)	0.06177970084194706	0.015417504599999975	1.3915236
Moving Average (скользящее среднее)	0.06409124602561414	0.4594285999999996	1.0751947
Autoregressive Moving Average (авторегрессионное скользящее среднее)	0.06213721097752398	0.6981746409000001	1.0738594
Autoregressive Integrated Moving Average (авторегрессионное интегрированное скользящее среднее)	0.05269180069418859	0.4814713859000001	1.8681948
Seasonal Autoregressive Integrated Moving-Average (сезонное авторегрессионное интегрированное скользящее среднее)	0.05269180069418859	0.5192791668000002	1.8303248
Simple Exponential Smoothing (простое экспоненциальное сглаживание)	0.04136309889269209	0.026095306100000215	319.70867
Holt Winter's Exponential Smoothing (экспоненциальное сглаживание Холта–Винтера)	0.04136309889269209	0.017659080399999992	1522.9401

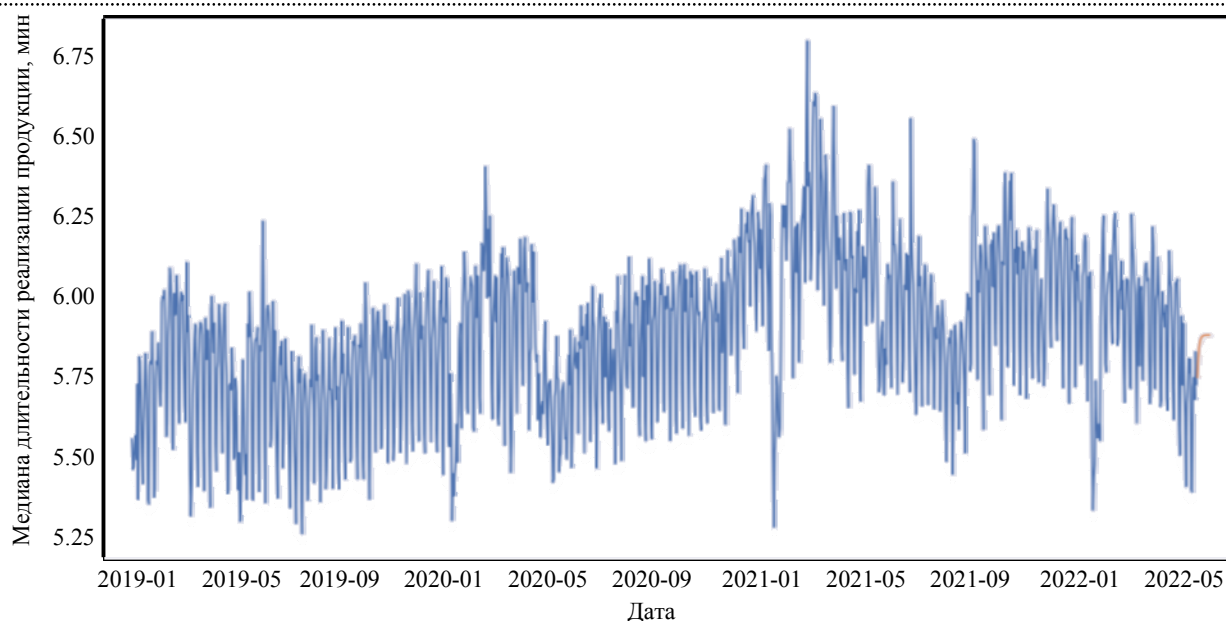


Рис. 6. Прогнозирование медианы длительности методом экспоненциального сглаживания Холта–Винтера
 Fig. 6. Predicting the median duration by Holt Winter's exponential smoothing

Табл. 5. Результаты параметров прогноза средней скорости реализации продукции
 Tab. 5. The results of the forecast parameters of the average speed of product sales

Метод прогнозирования	Средняя абсолютная ошибка, %	Средняя длительность прогнозирования, с	Оценка прогноза
Autoregression (авторегрессия)	0.025957466495268748	0.008429018299989367	1.8859133
Moving Average (скользящее среднее)	0.03367904798388213	0.03367904798388213	1.2438559
Autoregressive Moving Average (авторегрессионное скользящее среднее)	0.012180709234128083	1.2863856022000164	4.0167579
Autoregressive Integrated Moving Average (авторегрессионное интегрированное скользящее среднее)	0.010924614962210805	0.2719689002999985	21.589234
Seasonal Autoregressive Integrated Moving-Average (сезонное авторегрессионное интегрированное скользящее среднее)	0.010924614962210805	0.2630662425000082	22.258767
Simple Exponential Smoothing (простое экспоненциальное сглаживание)	0.010779332482743544	0.0183659556000066	643.03343
Holt Winter's Exponential Smoothing (экспоненциальное сглаживание Холта–Винтера)	0.010779332482743544	0.01763730069999383	693.91692

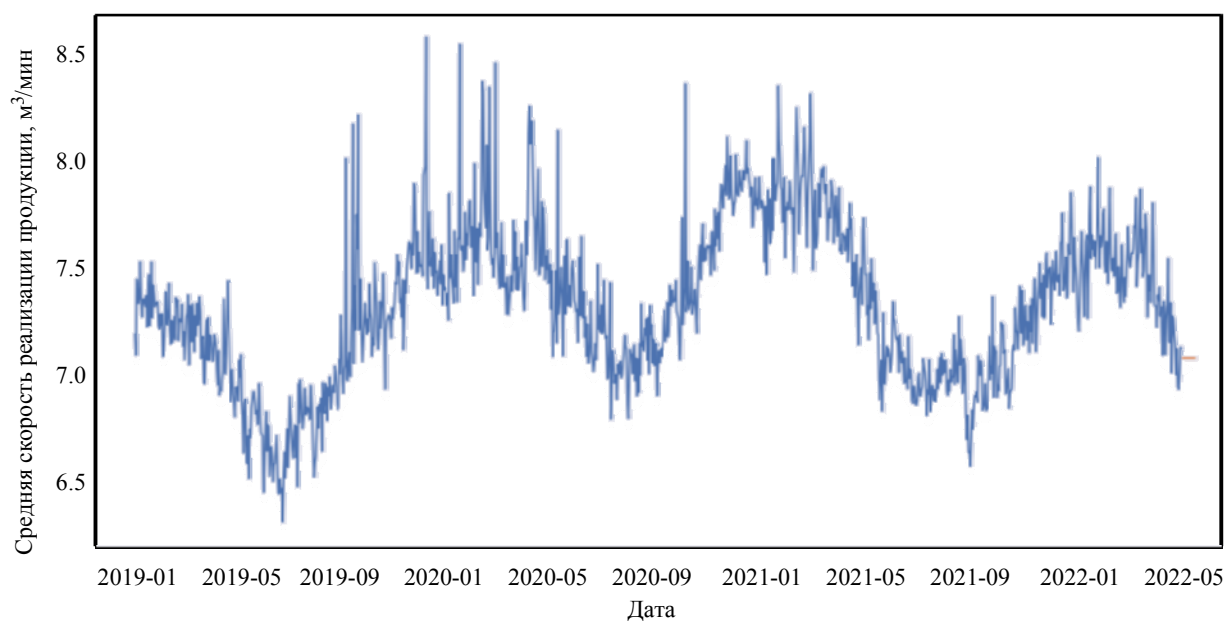


Рис. 7. Прогнозирование средней скорости реализации методом экспоненциального сглаживания Холта–Винтера
 Fig. 7. Predicting the average implementation velocity by Holt Winter's exponential stroking

временем прогнозирования и средней абсолютной ошибкой, поскольку фактор времени для киберфизических систем – один из наиболее критичных.

Выводы и заключение. Прогнозирование параметров киберфизических систем позволяет предугадать будущее состояние киберфизической системы, а полученные данные могут стать основой для принятия решений. При выборе метода

прогнозирования необходимо опираться не только на точность метода, но и на время построения прогнозной модели. В зависимости от особенностей прогнозируемого параметра используемый метод может быть изменен для более точного или более рационального, учитывающего фактор времени прогноза.

Список литературы

1. Analysis and design of cyber-physical systems. A hybrid control systems approach / R. G. Sanfelice, D. Rawat, J. Rodrigues, I. Stojmenovic // *Cyber-Physical Systems: From Theory to Practice*. CRC Press, 2016. P. 3–31.
2. Мокир Дж. Просвещенная экономика. Великобритания и промышленная революция 1700–1850 гг. М.: Изд-во Ин-та Гайдара, 2017. 792 с.
3. Ястреб Н. А. Индустрия 4.0: киберфизические системы и интернет вещей. Сб. науч. ст. / под ред. Н. А. Ястреб. Вологда: ВоГУ, 2015. Вып. 2. С. 136–143.
4. Кудж С. А., Цветков В. Я. Сетевое управление и киберфизические процессы // *Образовательные ресурсы и технологии*. 2017, № 2 (19). С. 86–92.

5. Plakhotnikov D. P., Kotova E. E. The use of artificial intelligence in cyber-physical systems // 2020 XXIII Intern. Conf. on Soft Computing and Measurements (SCM), St. Petersburg, Russia, 2020. P. 238–241. doi: 10.1109/SCM50615.2020.9198749.
6. Hyndman R. J., Koehler A. B. Another look at measures of forecast accuracy // *Intern. J. of Forecasting*. 2006, № 22(4). P. 679–688.
7. Nicol J. Fundamentals of real-time distributed simulation. URL: <https://www.goodreads.com/book/show/13830263-fundamentals-of-real-time-distributed-simulation> (дата обращения 22.04.22).

Информация об авторе

Плахотников Дмитрий Петрович – аспирант СПбГЭТУ «ЛЭТИ».
E-mail: dimapl21@yandex.ru

References

1. Sanfelice R. G., Rawat D., Rodrigues J., Stojmenovic I. Analysis and Design of Cyber-Physical Systems. A Hybrid Control Systems Approach // *Cyber-Physical Systems: From Theory to Practice*. CRC Press, 2016. P. 3–31.
2. Mokir Dzh. Prosveshchyonnaya ekonomika. Velikobritaniya i promyshlennaya revolyuciya 1700–1850 gg. M.: Izd-vo In-ta Gajdara, 2017. 792 s. (In Russ.).
3. YAstreb N. A. Industriya 4.0: kiberfizicheskie sistemy i internet veshchej. Sb. nauch. st. / pod red. N. A. YAstreb. Vologda: VoGU, 2015. Vyp. 2. S. 136–143. (In Russ.).
4. Kudzh S. A., Cvetkov V. YA. Setecentricheskoe upravlenie i kiberfizicheskie processy // *Obrazovatel'nye resursy i tekhnologii*. 2017. № 2 (19). S. 86–92. (In Russ.).

5. Plakhotnikov D. P., Kotova E. E. The Use of Artificial Intelligence in Cyber-Physical Systems // 2020 XXIII Intern. Conf. on Soft Computing and Measurements (SCM), St. Petersburg, Russia, 2020. P. 238–241. doi: 10.1109/SCM50615.2020.9198749.
6. Hyndman R. J., Koehler A. B. Another Look at Measures of Forecast Accuracy // *Intern. J. of Forecasting*. 2006, № 22(4). P. 679–688.
7. Nicol J. Fundamentals of Real-Time Distributed Simulation. URL: <https://www.goodreads.com/book/show/13830263-fundamentals-of-real-time-distributed-simulation> (data obrashcheniya 22.04.22).

Information about the author

Dmitriy P. Plakhotnikov – postgraduate student of Saint Petersburg Electrotechnical University.
E-mail: dimapl21@yandex.ru

Статья поступила в редакцию 23.05.2022; принята к публикации после рецензирования 31.05.2022; опубликована онлайн 13.09.2022.

Submitted 23.05.2022; accepted 31.05.2022; published online 13.09.2022.