

ANALYTICAL REVIEW OF MODERN INFORMATION TECHNOLOGIES FOR COLLECTION, PROCESSING AND ANALYSIS OF DATA APPLIED FOR MONITORING ATMOSPHERIC POLLUTION AIR

O. D. Sokolova, A. V. Materukhin*

Institute of Computational Mathematics and Mathematical Geophysics SB RAS,
630090, Novosibirsk, Russia

*Moscow State University of Geodesy and Cartography
105064, Moscow, Russia

DOI: 10.24411/2073-0667-2020-10002

The article is an analytical review of scientific publications on the topic of modern information technologies in the field of data collection, processing and analysis for monitoring air pollution in large cities. Modern large cities are the source of most emissions of harmful substances into the atmosphere. To monitor atmospheric pollution parameters, as a rule, either data obtained from few and fairly far from each other regular environmental monitoring stations are used, or data obtained as a result of expensive short-term field studies. This is not enough to obtain adequate information about the spatial distribution of pollutants or to identify sources of pollution. To monitor atmospheric pollution parameters with high spatial and temporal resolution are required.

Over the past two decades, progress has been made in the development of small-sized sensor devices with the added ability to determine their location, the so-called geosensors. The use of wireless geosensor networks for monitoring air quality is considered in many scientific publications. The aim of the review was to analyze modern scientific publications in this field of research. The problems of research are related to the tasks of managing of spatio-temporal data streams. The review presents articles that describe mathematical models and optimization algorithms for data collection processes, as well as their processing. Today, the scientific community has already reached a consensus about which information technologies are most preferable at this level — data mining methods are the most promising tool for analyzing and predicting air pollution. The level of spatial and temporal granularity of data necessary for the application of these methods can be achieved using geosensor networks.

The main problems that arise during the collection and transmission of data in a wireless geosensor network are problems of ensuring energy efficiency, speed and accuracy of transmission. For the effective functioning of monitoring networks, it is necessary to develop adequate mathematical models and methods that will optimize the data collection process. The review analyzes various approaches to solving this problem (for example, the use of mobile sinks or clustering of many geosensors). The review includes publications in which it is proposed to improve monitoring through the use of mobile nodes. The possibility of using Vehicle ad-hoc Sensor Networks (VANET) for monitoring air quality has been offered quite often in recent years — sensors can be installed on cars that move in the transport network of a city. Thanks to mobility, the sensor node can take measurements in different places, thereby reducing the requirements for a large number of sensor nodes for monitoring the territory.

The bottleneck of most research projects in this area is the lack of the possibility of large-scale deployment of the proposed solutions. Therefore, the review focuses on the use of simulation to verify

the effectiveness of various technological solutions. Some systems for modeling data streams from geosensors (for example, the MobGeoSensSimulation system), which measure air quality parameters over a large urban area and send data to a processing center, are analyzed.

The analysis of scientific publications carried out in the article suggests that the main trend in research in the field of monitoring atmospheric air pollution is the development of technological solutions associated with the use of wireless geosensor networks, and the current trend is the transition from stationary to mobile geosensors. Such a transition will allow the use of data mining methods with greater efficiency. Various solutions described in the modern scientific literature and intended for application at various levels of the atmospheric monitoring system are analyzed: at the level of data collection, at the level of data processing and at the level of data analysis. Another identified research trend is the development and study of computer modeling systems for monitoring systems.

Key words: monitoring of air pollution, wireless geo-sensor networks.

References

1. Nittel S. A survey of geosensor networks: Advances in dynamic environmental monitoring // *Sensors*. 2009. N 9(7). P. 5664–5678.
2. Duckham Matt, *Decentralized Spatial Computing. Foundations of Geosensor Networks* // Springer. 2014.
3. Yakubajlik O. E., Gosteva A. A., Erunova M. G., Kadochnikov A. A., Matveev A. G., Pyataev A. S., Tokarev A. V. *Razrabotka sredstv informacionnoj podderzhki nablyudenij za sostoyaniem okruzhayushchej prirodnoj sredy* // *Vestnik KemGU*. 2012. T. 2. N 4(52). S. 136–141.
4. Materuhin A. V. *Teoreticheskie osnovy i metodologiya obrabotki potokov prostranstvenno-vremennyh dannyh: dissertaciya ...doktora tekhnicheskikh nauk: 25.00.35 / Materuhin Andrej Viktorovich; [Mesto zashchity: Mosk. gos. un-t geodezii i kartografii]. Moskva, 2018.*
5. Materuhin A. V. *Matematicheskaya model' processov obrabotki potokov prostranstvenno-vremennyh dannyh v geoinformacionnoj sisteme* // *Izvestiya vysshih uchebnyh zavedenij. Geodeziya i aerofotos"emka*. 2018. T. 62. N 2. S. 237–243.
6. Gheyas I. A., Smith L. S. *A Neural Network Approach to Time Series Forecasting* // *Proceedings of the World Congress on Engineering, London, 2009. Vol. 2. P. 1292–1296.*
7. Rahman, N. H. A.; Lee, M. H.; Suhartono; Latif, M. T. *Artificial neural networks and fuzzy time series forecasting: An application to air quality* // *Qual. Quant.* 2015. N 49. P. 2633–2647.
8. Andreeva E. S., Lazareva E. O., Lipovickaya I. N. *Prognoz urovnya zagryazneniya atmosfernogo vozduha v Sankt-Peterburge s primeneniem algoritma prinyatiya reshenij* // *Vestnik Moskovskogo universiteta. Seriya 5: Geografiya*. 2019. N 2. S. 55–60.
9. Dincer, N. G., Akkus, O. *A new fuzzy time series model based on robust clustering for forecasting of air pollution* // *Ecol. Inform.* 2018. N 43. P. 157–164.
10. Lu Bai, Jianzhou Wang, Xuejiao Ma, Haiyan Lu. *Air Pollution Forecasts: An Overview* // *International Journal of Environmental Research and Public Health*, 2018.
11. Mishra, D.; Goyal, P. *NO2 forecasting models Agra* // *Atmos. Pollut. Res.* 2015. N 6. P. 99–106.
12. Tsujitaa W., Yoshinoa A., Ishidab H., and Moriizumi T. *Gas sensor network for air-pollution monitoring* // *Sensors Actuators B, Chem.* 2005. Vol. 110. N 2. P. 304–311.
13. Savinyh V. P., Majorov A. A., Materuhin A. V. *Postroenie prostranstvennoj modeli zagryazneniya vozduha na osnove ispol'zovaniya potokov dannyh ot setej geosensorov* // *Geodeziya i kartografiya*. 2017. N 12. T. 78. S. 39–43.
14. Rayane El Sibai, Yousra Chabchoub, Raja Chiky, Jacques Demerjian, Kablan Barbar. *Assessing and Improving Sensors Data Quality in Streaming Context* // *9th International Conference, Computational Collective Intelligence ICCCI 2017, Nicosia, Cyprus, September 27–29, 2017, Proceedings, Part II, 2017, Vol. 10449. P. 590–599.*

15. Rai A., Kumar P., Pilla F., Skouloudis A., Sabatino S., Ratti C., Yasar A., and Rickerby D. End-user perspective of low-cost sensors for outdoor air pollution monitoring // *Sci. Total Environ.* 2017. Vol. 607–608. P. 691–705.
16. Duckham Matt, Nussbaum Doron, Sack Jorg Rudiger, Santoro Nicola. Efficient, Decentralized Computation of the Topology of Spatial Regions // *IEEE Transactions on Computers.* 2011. Vol. 60. N 8. P. 1100–1113.
17. Shahov V. V., Yurgenson A. N., Sokolova O. D. Effektivnyj metod generacii sluchajnyh geometricheskikh grafov dlya modelirovaniya besprovodnyh setej // *PDM.* 2016. N 4 (34). S. 99–108.
18. Bieringer P. E.; Young G. S.; Rodriguez L. M.; Annunzio, A. J.; Vandenberghe F., Haupt S. E. Paradigms and commonalities in atmospheric source term estimation methods // *Atmospheric Environment.* Elsevier BV. 2017. N 156. P. 102–112.
19. Issartel J.-P. Emergence of a tracer source from air concentration measurements, a new strategy for linear assimilation // *Atmospheric Chemistry and Physics, Copernicus GmbH,* 2005. N 5. P. 249–273.
20. Mamonov A. V. Tsai Y.-H. R. Point source identification in nonlinear advection-diffusion-reaction systems // *Inverse Problems, IOP Publishing,* 2013. N 29. 035009.
21. Penenko V. V.; Penenko, A. V.; Tsvetova E. A. Gochakov A. V. Methods for Studying the Sensitivity of Air Quality Models and Inverse Problems of Geophysical Hydrothermodynamics // *Journal of Applied Mechanics and Technical Physics, Pleiades Publishing Ltd,* 2019. N 60. P. 392–399.
22. Penenko A., Mukatova Z. S., Penenko V., Gochakov A., Antokhin P. Numerical study of the direct variational algorithm of data assimilation in urban conditions // *Atmospheric and ocean optics.* 2018. N 31. P. 456–462.
23. Sun C., Li V., Lam J., and Leslie I. Optimal Citizen-Centric Sensor Placement for Air Quality Monitoring: A Case Study of City of Cambridge, the United Kingdom // *IEEE Access,* 2019. Vol. 7.
24. Sokolova O. D., YUrgenson A. N. Zadacha optimal'nogo razmeshcheniya ustrojstv analiza informacionnyh potokov v setyah // *Problemy informatiki.* 2010. N 2. S. 33–42.
25. Säaskilahti Kirsti, Kangaskorte Riitta, Luimula Mika & Yli-Hemminki, Juha. Collecting and visualizing wireless geosensor data using mobile devices // *ACM International Conference Proceeding Series.* 2010. 10.1145/1823854.1823882.
26. Ilka A. R., Gilberto C., Renato A., Antnio M. V. M. Data-Aware Clustering for Geosensor Networks Data Collection // *Anais XIII Simpsio Brasileiro de Sensoriamento Remoto, INPE,* 2007. P. 6059–6066.
27. Shahov V. V., Migov D. A., Sokolova O. D. Besprovodnye sensornye seti, osnashchennye sredstvami polucheniya energii iz okruzhayushchej sredy // *Problemy informatiki.* 2014. N 4. S. 69–79.
28. Materuhin A. V., SHahov V. V., Sokolova O. D. Modeli processov sbora prostranstvenno-vremennyh dannyh s ispol'zovaniem mobil'nyh stokov // *Geodeziya i kartografiya.* 2018. T. 79. N 12. S. 22–28.
29. Boubrima A., Bechkit W., Rivano H. Optimal WSN deployment models for air pollution monitoring // *IEEE Trans. Wireless Commun.* 2017. Vol. 16. N 5. P. 2723–2735.
30. Kaivonen Sami, Ngai Edith. Real-time air pollution monitoring with sensors on city bus // *Digital Communications and Networks.* 2019.
31. You-Chiun Wang, Guan-Wei Chen. Efficient Data Gathering and Estimation for Metropolitan Air Quality Monitoring by Using Vehicular Sensor Networks // *IEEE Transactions On Vehicular Technology.* 2017. Vol. 66. N 8. P. 7234–7248.
32. Penza M., Suriano D., Villani M. G., Spinelle L., and Gerboles M. Towards air quality indices in smart cities by calibrated low-cost sensors applied to networks // in *Proc. IEEE Conf. Sensors,* 2014. P. 2012–2017.

33. Shakhov V., Sokolova O. Towards Air Pollution Detection with Internet of Vehicles // Proceedings of 15th International Asian School-Seminar Optimization Problems of Complex Systems. OPCS-2019. P. 183–187.
34. Murty, R. N.; Mainland, G.; Rose, I.; Chowdhury, A. R.; Gosain, A.; Bers, J.; Welsh, M. CitySense: An Urban-Scale Wireless Sensor Network and Testbed // In Proceedings of the 2008 IEEE Conference on Technologies for Homeland Security, Waltham, MA, USA, 12–13 May 2008. P. 583–588.
35. Shu-Chiung Hua, You-Chiun Wang, Chiuan-Yu Huang, Yu-Chee Tsenga. Measuring air quality in city areas by vehicular wireless sensor networks // The Journal of Systems and Software. 2011. N 84. P. 205–212.
36. Materukhin A. V., Gvozdev O. G., Maiorov A. A. and Sokolova O. D. Simulation of Spatio-Temporal Data Streams from Geosensors Located on Mobile Objects // 15th International Asian School-Seminar Optimization Problems of Complex Systems (OPCS), Novosibirsk, Russia, 2019. P. 179–182. DOI: 10.1109/OPCS.2019.8880188.

АНАЛИТИЧЕСКИЙ ОБЗОР СОВРЕМЕННЫХ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ В ОБЛАСТИ СБОРА, ОБРАБОТКИ И АНАЛИЗА ДАННЫХ, ПРИМЕНЯЕМЫХ ДЛЯ МОНИТОРИНГА ЗАГРЯЗНЕНИЯ АТМОСФЕРНОГО ВОЗДУХА

О. Д. Соколова, А. В. Матерухин*

Институт вычислительной математики и математической геофизики СО РАН,
630090, Новосибирск, Россия

*Московский государственный университет геодезии и картографии
105064, Москва, Россия

УДК 528

DOI: 10.24411/2073-0667-2020-10002

Статья представляет собой аналитический обзор научных публикаций на тему современных информационных технологий в области сбора, обработки и анализа данных для мониторинга загрязнения атмосферного воздуха в условиях крупных городов. Показано, что главным трендом исследований в этой области является разработка технологических решений, связанных с использованием геосенсорных сетей для создания систем мониторинга. Рассмотрены различные решения, описанные в современной научной литературе и предназначенные для применения на различных уровнях системы мониторинга атмосферы: на уровне сбора данных, на уровне обработки данных и на уровне анализа данных.

Ключевые слова: мониторинг загрязнения воздуха, беспроводные геосенсорные сети.

Введение. В настоящее время в крупных городах проживает более 50 % населения планеты, при этом именно мегаполисы являются источником большинства выбросов в атмосферу вредных веществ, таких, как оксид углерода или диоксид азота. Распространение таких загрязнений в воздушной среде приводит к негативным последствиям для окружающей среды в целом, а также для жизни и здоровья людей. В настоящее время для мониторинга состояния воздуха в городах используются, как правило, либо данные, получаемые от немногих и довольно далеко расположенных друг от друга станций регулярного экологического мониторинга, либо данные, которые были получены в результате дорогостоящих краткосрочных полевых исследований. Этого недостаточно для получения адекватной информации о пространственном распределении загрязняющих веществ или выявления источников загрязнения, даже при использовании сложных моделей рассеивания. Как отмечено в работах [1–3], для мониторинга загрязнения воздуха в гетерогенных средах, таких как городская среда, которые характеризуются наличием большого количества различных источников загрязняющих выбросов, требуется производить

Работа выполнена в рамках программы фундаментальных исследований СО РАН (проект № 0315-2016-0006) и при финансовой поддержке гранта РФФИ и Правительства Новосибирской области (грант № 19-47-540007 p_a).

измерения с большим пространственным и временным разрешением. За последние два десятилетия был достигнут беспрецедентный прогресс в разработке малоразмерных сенсорных устройств с добавленными возможностями определения своего местоположения — геосенсоров. Использование беспроводных сетей геосенсоров для мониторинга качества воздуха в крупных городах рассматривается во многих научных публикациях. Однако эти публикации часто ограничиваются описанием какого-либо одного аспекта такого рода применения сетей геосенсоров, не давая читателю общей картины исследований в этой области. Целью написания данного обзора являлось проведение анализа результатов исследований в области применения современных технологий (в частности, беспроводных сетей) для мониторинга состояния атмосферы.

1. Технологические проблемы и решения на уровне обработки и анализа данных. В последнее время интерес исследователей сместился со статического изучения геопространственных феноменов, при котором во внимание принимаются только пространственные характеристики изучаемого явления, к более продвинутым средствам выявления динамических отношений в исследуемых геосистемах. Как показано в [4], технология получения пространственно-временных данных с помощью систем геосенсоров предоставляет исследователям возможность получать данные с невозможной ранее пространственно-временной детализацией, а технический прогресс в области разработки геосенсоров существенно удешевляет использование этой технологии. Проблематика исследований в этой области связана с управлением потоками пространственно-временных данных от сети геосенсоров. Одной из основных задач являлась проблема обработки неограниченно масштабируемых потоков. В работе [5] представлена математическая модель процессов обработки потоков пространственно-временных данных в геоинформационной системе, а работа [4] описывает методологию обработки, которая обеспечивает независимость времени выполнения пространственных запросов к входящим потокам данных от интенсивности этих потоков.

Верхний уровень систем мониторинга состояния атмосферного воздуха — это технологии анализа данных. На сегодняшний день в научном сообществе уже сложился консенсус по поводу того, какие именно информационные технологии являются наиболее предпочтительными на этом уровне — в настоящее время для анализа и прогнозирования загрязнения воздуха наиболее популярными и перспективными инструментами являются методы интеллектуального анализа данных. Существует множество факторов, влияющих на такую сложную систему, как атмосфера, а взаимосвязь этих факторов плохо изучена и также является сложной. Поэтому при прогнозировании загрязнения воздуха в больших городах следует выбирать разные методы прогнозирования с учетом площади города и типов загрязняющих веществ. Среди применяемых методов можно выделить как наиболее часто упоминаемые следующие: методы, основанные на деревьях принятия решений; методы регрессионного анализа; применение искусственных нейронных сетей.

Одна из ранних попыток использовать искусственные нейронные сети для прогнозирования загрязнения воздуха описана в [6].

В [7] показано, что точность нейросетевых моделей прогнозирования выше, чем у других статистических моделей. Однако, несмотря на преимущества применения моделей, основанных на методах машинного обучения, по-прежнему не существует эффективных решений для повышения точности прогнозирования и увеличения длительности сроков.

В статье [8] предлагается решение на основе использования дерева принятия решения. В качестве экспериментальных данных авторами была использована база данных

погодных условий по данным метеорологических наблюдений, выполненных на станции, расположенной в г. Санкт-Петербург. В этой работе также были использованы данные наблюдений за состоянием атмосферного воздуха (период 2006–2014 гг.), полученные от 10 стационарных постов, расположенных в 8 административных районах города, 4 раза в сутки. Для оценки связи синоптического процесса и уровня загрязнения воздуха был использован коэффициент взаимной сопряженности Пирсона-Чупрова. По мнению авторов статьи [8], подход на основе дерева принятия решения позволяет определить ожидаемый уровень загрязнения атмосферного воздуха для теплого и холодного периодов года дневных и ночных часов, с заблаговременностью в 12 часов, с оправдываемостью 84–91%.

Для прогнозирования загрязнения воздуха в Турции была предложена модель Fuzzy Time Series (FTS) (нечеткого временного ряда), основанная на алгоритме кластеризации Fuzzy K-Medoid (FKM) [9]. По мнению авторов статьи [9], модели типа FTS имеют преимущество по сравнению с другими методами, используемыми при прогнозировании загрязнения воздуха, поскольку не требуют каких-либо предположений о вероятностном распределении данных временных рядов, а также обеспечивают успешные результаты даже в случае малого набора наблюдений. Эффективность рассматриваемого в этой статье метода прогнозирования загрязнения воздуха была продемонстрирована на временном ряде, состоящем из показаний концентраций SO_2 , измеренных на 65 станциях мониторинга в Турции в период с 2013 по 2016 г. Однако в этой статье отмечено, что существующие модели FTS, основанные на нечеткой кластеризации, не обладают достаточной точностью прогнозирования на наборах данных, содержащих аномальные выбросы.

Статья [10] предлагает использовать гибридную систему, включающую в себя два или более метода прогнозирования загрязнения воздуха. Такая гибридная система описана в [11]. Система была предназначена для прогнозирования почасовых концентраций диоксида азота на территории комплекса Тадж-Махал в городе Агра, Индия. Использовался метод главных компонент (РСА), чтобы найти корреляции между параметрами метеорологического прогнозирования и загрязнителями воздуха. Выделенные на этом этапе значимые параметры метеорологического прогнозирования были затем взяты в качестве входных параметров для обучения многослойной искусственной нейронной сети. Проведенное исследование показало, что обученная искусственная нейронная сеть не могла хорошо прогнозировать в периоды, в которые наблюдались высокие концентрации загрязнителя. Как правило, такие всплески вызваны антропогенными факторами.

Несмотря на различие применяемых методов интеллектуального анализа, все они, как правило, в качестве входных данных используют одни и те же типы данных. В ходе проведенного анализа было выявлено, что для прогнозирования уровня загрязнения воздуха используются в основном данные о текущем качестве воздуха, текущих погодных условиях, прогнозе погоды, а также времени суток и дня недели. Стоит также отметить, что методы интеллектуального анализа, как правило, требуют обширных данных, наиболее полно отображающих исследуемый объект. Использование только данных, полученных от стационарных постов наблюдения за концентрацией примесей в воздухе, с последующим применением математических моделей, накладывает ограничения на эффективное использование интеллектуальных методов анализа.

Таким образом, можно сделать вывод, что на настоящий момент главной проблемой в использовании методов интеллектуального анализа данных о загрязнении атмосферного воздуха является необходимость получения большого количества данных с высокой пространственной и временной детальностью.

2. Технологические проблемы и решения на уровне сбора данных. Процесс мониторинга окружающей среды состоит из регулярного сбора и анализа потоков данных, получаемых от датчиков. Необходимый для эффективного применения методов интеллектуального анализа данных уровень пространственной и временной детальности данных может быть достигнут с помощью использования сетей геосенсоров. За последние два десятилетия был достигнут большой прогресс в разработке сенсорных устройств с возможностями определения своего местоположения — геосенсоров, интеллектуальных датчиков, которые измеряют значения величин, распределенных в пространстве [12]. Так называемые геосенсорные сети (GSN) представляют собой специализированные приложения технологии беспроводных сенсорных сетей в географическом пространстве, которые обнаруживают и отслеживают явления и процессы в окружающей среде [13]. По сравнению со стационарными пунктами мониторинга, сеть с множеством датчиков обеспечивает более детальный мониторинг качества воздуха. Собираемые сенсорами данные о состоянии атмосферы передаются в аналитические центры для дальнейшей обработки. Доставка данных в режиме реального времени увеличивает точность мониторинга и прогноза состояния атмосферы. К собранным данным применяются различные методы извлечения данных, чтобы вывести агрегированную статистику, полезную для обнаружения и прогнозирования аномалий. Полученные результаты тесно связаны с качеством собранных данных. На самом деле, данные часто содержат шумные, ошибочные и пропущенные значения. Низкое качество данных приводит к ошибочным результатам. Одно из решений для преодоления этой проблемы представлено в статье [14]. Авторы представляют концепцию качества данных датчиков и предлагают полную систему управления. Описана архитектура системы Sensors Data Quality Management System, предложены методы для обнаружения данных с шумом и для улучшения их качества, что приводит к повышению точности измерений.

Качество геосенсоров, которые собирают данные, остается одной из главных проблем, препятствующих широкому распространению недорогих сенсорных технологий. Неверные данные дают искаженную картину об уровнях загрязнения воздуха. В статье [15] авторы анализируют рабочие характеристики нескольких недорогих датчиков для мониторинга частиц и газов, суммируют их возможности и ограничения. Для обеспечения точности данных рекомендуется двухэтапный процесс калибровки датчика, который включает лабораторную калибровку, дополненную проверками, выполняемыми пользователем в условиях окончательного развертывания. Для больших сенсорных сетей, где калибровочные проверки нецелесообразны, предлагается использовать статистические методы обеспечения качества данных. Дальнейшие усовершенствования и принятие сложных математических и статистических методов для калибровки датчиков, обнаружения неисправностей и обеспечения точности данных действительно могут реализовать преимущества недорогой сети мониторинга загрязнения воздуха в мегаполисах.

3. Математические методы и алгоритмы в решении задач оптимизации сбора и передачи информации. Большое количество научных публикаций (в основном зарубежных) на тему оптимизации функционирования сетей сбора информации о состоянии атмосферы посвящено разработке математических моделей и методов для решения возникающих задач. В качестве моделей для передачи данных в беспроводных сетях обычно используют случайные геометрические графы, в частности, UDG-графы [2, 16]. Расположение узлов (датчиков) моделируется точками на Евклидовой плоскости, а область распространения сигнала находится в круге с радиусом, равном радиусу действия сиг-

нала. Алгоритм генерации псевдослучайных UDG-графов (например, с ограничением на степень вершин, диаметр графа) предложен в [17].

Для оценки и прогнозирования уровня загрязнения атмосферы на основе данных мониторинга и математических моделей переноса и трансформации примесей используются алгоритмы решения обратных задач и задач усвоения данных. Также часто в исследованиях рассматриваются обратные задачи идентификации источников. Обзор алгоритмов идентификации источников представлен в [18]. В частности, используются алгоритмы на основе ансамблей решений сопряженных уравнений [19–21], которые допускают естественное распараллеливание. Алгоритмы усвоения данных решают последовательности обратных задач по мере поступления новых данных измерений. В контексте оперативных оценок особый интерес представляют алгоритмы усвоения, работающие "в реальном времени". Примером является алгоритм квази-независимого усвоения данных на отдельных шагах схемы расщепления [22].

В исследовании, описанном в статье [23], авторы решают задачу оптимального размещения датчиков, ориентированного на быстрое оповещение информацией граждан, то есть, с учетом мест наибольшего скопления людей в мегаполисе. Показано, что задача NP-трудна, но целевая функция обладает хорошим неубывающим и субмодульным свойством. Описаны эффективный жадный алгоритм и его варианты, которые могут быть приняты с гарантированным коэффициентом аппроксимации $(1 - 1/e)$. Подобная оценка для задачи размещения мониторинговых устройств была получена ранее, в статье [24].

Основные задачи, которые возникают при сборе и передаче данных в беспроводной сети геосенсоров — энергоэффективность, скорость и точность передачи. Так как больше всего энергии тратят узлы, расположенные рядом со стоками, то возникает дисбаланс энергии. Различные подходы к решению этой проблемы (например, использование мобильных стоков, кластеризация множества геосенсоров) описаны в [25, 26]. Одна из таких задач связана с понятием энергоэффективности беспроводных узлов. Разработка и внедрение современных сенсорных сетей, узлы в которых оснащены средствами для получения энергии из окружающей среды, привели к появлению новых задач, в которых необходимо учитывать стохастический характер и ненадежность источников энергии. Следовательно, требуется модификация существующих протоколов, используемых в традиционных сенсорных сетях. В статье [27] приводится обзор исследовательских проектов по данной тематике, а также предложены новые алгоритмы нахождения оптимальных мест для осуществления беспроводной подзарядки группы сенсоров при помощи мобильного подзаряжающего устройства. Алгоритмы основаны на методах решения известных задач из теории графов: поиск p -медиан и размещение p -центров.

Для эффективного функционирования сетей мониторинга необходима разработка адекватных математических моделей и методов, которые позволят оптимизировать процесс сбора данных (например, минимизировать энергозатраты). В статье [28] описана задача повышения эффективности потребления энергии в процессе сбора пространственно-временных данных с помощью сети геосенсоров. При передаче данных от всех узлов сети к стокам, через узлы, находящиеся рядом со стоками, проходит значительно больший объем трафика, чем через другие узлы. Таким образом, между узлами возникает дисбаланс энергопотребления, что приводит к уменьшению времени работы всей сети, то есть из-за отдельных узлов возникает необходимость в подзарядке. В статье предложено использовать в качестве ресурса для оптимизации возможную мобильность стоков. Авторы предложили математическую модель для анализа времени жизни беспроводной сети гео-

сенсоров, использующей мобильные стоки. Предложенный подход позволяет увеличить время жизни беспроводной сети геосенсоров с помощью нахождения оптимальных маршрутов движения мобильных стоков, собирающих информацию с узлов.

Использование беспроводных сетей для мониторинга загрязнения воздуха призвано решить две основные задачи — регулярное картирование и обнаружение высоких концентраций загрязнения. Для адекватности решения необходимо оптимальное размещение датчиков, чтобы получить точную информацию о загрязнении воздуха. В статье [29] авторы используют подходы к моделированию, основанные на методах целочисленного линейного программирования (ILP). С помощью модели ILP авторы описывают регулярное картирование загрязнения воздуха на основе прогнозируемых карт загрязнения. Еще две модели ILP разработаны для обнаружения пересечения порогов загрязнения. Модели основаны на прогнозируемых данных и на множестве собранных данных о выбросах с описанием источников загрязнения.

Улучшить мониторинг с помощью сенсорной сети можно, например, за счет использования мобильных узлов. Данный подход использовался в том числе и для организации мониторинга состояния атмосферы. В Упсале (Швеция) для мониторинга загрязнения воздуха применялись беспроводные сети, где в качестве мобильного носителя сенсора использовали общественный транспорт [30]. Основной целью этого эксперимента было сравнить данные сенсоров, размещенных на автобусах, полученных за длительный период времени, с данными других мониторинговых устройств. В указанной статье не исследуется вероятность ошибки и скорость обнаружения присутствия опасного загрязнения. Существует дефицит моделей для соответствующих вычислений.

В последнее время активно исследуются возможности использования автомобильной беспроводной сети (Vehicle Sensor Networks) для мониторинга качества воздуха в крупных городах — датчики устанавливаются на автомобилях, которые периодически замеряют данные о состоянии воздуха [31]. Благодаря мобильности, сенсорный узел может проводить измерения в разных местах, тем самым снижая требования к большому количеству сенсорных узлов для мониторинга территории мегаполиса. Здесь возникает задача составления баланса между точностью мониторинга и стоимостью связи: как регулировать расписание передачи информации мобильными узлами для удовлетворения качества мониторинга при одновременном снижении расходов на связь. Например, в районе, где много автомобилей с датчиками мониторинга, можно уменьшить частоту сообщений некоторых узлов, чтобы исключить возможное дублирование. С другой стороны, на территориях, где концентрация CO_2 резко меняется, узлу может потребоваться увеличить частоту передачи сообщений для повышения точности [32].

В статье [33] рассматривается вопрос использования беспроводной сети с мобильными геосенсорами, т. е. датчиками, закрепленными на движущихся объектах (транспортных средствах). Рассмотрено два варианта движения транспортных средств: время между появлением транспортных средств в исследуемой зоне является константой, либо распределено по экспоненциальному закону. Наличие повышенного уровня загрязнения в исследуемой области определяется во время пребывания транспортного средства в этой области с некоторой вероятностью. Рассматриваются методы оценки указанного значения вероятности. Также получена функция распределения для скорости обнаружения превышения порога загрязнения. Результаты могут применяться для управления системами мониторинга, использующими мобильные сенсоры.

4. Имитационное моделирование как средство проверки эффективности различных технологических решений. Узким местом большинства исследовательских проектов в рассматриваемой области является отсутствие крупномасштабного развертывания предложенных решений. Еще в [34] было отмечено, что большинство исследовательских групп оценивает свои идеи с помощью моделирования или небольших тестовых развертываний, результатом чего могут быть потенциальные проблемы в реальных крупномасштабных приложениях.

В [31, 35] авторы с помощью имитационного моделирования анализируют эффективность некоторых технологических решений, разработанных в рамках реализации концепции беспроводных сенсорных сетей с датчиками на автомобилях (vehicular sensor networks), и методы агрегирования таких данных.

В работе [36] описана система компьютерного моделирования обобщенной сети геосенсоров MobGeoSensSimulation (MGSS) для моделирования потоков данных от множества мобильных геосенсоров, которые измеряют показатели качества воздуха на большой территории мегаполиса и отправляют данные в центр обработки с некоторой частотой. Система была реализована на языке программирования Python 3, с использованием библиотек NumPy и SciPy. Система была сконфигурирована для имитационного моделирования потоков данных от геосенсоров, размещенных на общественном транспорте г. Москвы. Моделируется распространение загрязнения воздуха от источников, а также движение геосенсоров по реальным маршрутам городского транспорта. По получаемым данным в режиме реального времени восстанавливалось распределение концентрации загрязнения. Было установлено, что имитационное моделирование в разработанной системе компьютерного моделирования MGSS потоков пространственно-временных данных от мобильных геосенсоров может использоваться как средство для структурной оптимизации сети геосенсоров. Авторами предложено решение для задачи вычисления при заданных ограничениях оптимального количества транспортных средств с установленными датчиками для построения пространственной модели загрязнения исследуемой территории.

Заключение. Проведенный в настоящей статье анализ научных публикаций позволяет утверждать, что главным трендом исследований в области мониторинга загрязнения атмосферного воздуха в условиях крупного города является разработка технологических решений, связанных с использованием беспроводных геосенсорных сетей, причем текущая тенденция — это переход от стационарных к мобильным геосенсорам. Такой переход позволит использовать интеллектуальные методы анализа данных с большей эффективностью. Однако масштабирование геосенсорных сетей привело к появлению новых проблем, не все из которых решены. В частности, до сих пор не утратили своей актуальности проблемы оптимизации на уровне сети сбора данных. Другим выделенным исследовательским трендом является разработка и исследование систем компьютерного моделирования систем мониторинга загрязнения атмосферного воздуха. Было отмечено, что научные группы, работающие в этой области, активно используют имитационное моделирование беспроводных геосенсорных сетей не только как средство тестирования полученных технологических решений, но и как средство их оптимизации.

Список литературы

1. Nittel S. A survey of geosensor networks: Advances in dynamic environmental monitoring // *Sensors*. 2009. N 9(7). P. 5664–5678.
2. Duckham Matt, Decentralized Spatial Computing. Foundations of Geosensor Networks // Springer. 2014.
3. Якубайлик О. Э., Гостева А. А., Ерунова М. Г., Кадочников А. А., Матвеев А. Г., Пятаев А. С., Токарев А. В. Разработка средств информационной поддержки наблюдений за состоянием окружающей природной среды // *Вестник КемГУ*. 2012. Т. 2. № 4(52). С. 136–141.
4. Матерухин А. В. Теоретические основы и методология обработки потоков пространственно-временных данных: диссертация . . . доктора технических наук: 25.00.35 / Матерухин Андрей Викторович; [Место защиты: Моск. гос. ун-т геодезии и картографии]. Москва, 2018.
5. Матерухин А. В. Математическая модель процессов обработки потоков пространственно-временных данных в геоинформационной системе // *Известия высших учебных заведений. Геодезия и аэрофотосъемка*. 2018. Т. 62. № 2. С. 237–243.
6. Gheyas I. A., Smith L. S. A Neural Network Approach to Time Series Forecasting // *Proceedings of the World Congress on Engineering, London, 2009. Vol. 2. P. 1292–1296*.
7. Rahman, N. H. A.; Lee, M. H.; Suhartono; Latif, M. T. Artificial neural networks and fuzzy time series forecasting: An application to air quality // *Qual. Quant.* 2015. N 49. P. 2633–2647.
8. Андреева Е. С., Лазарева Е. О., Липовицкая И. Н. Прогноз уровня загрязнения атмосферного воздуха в Санкт-Петербурге с применением алгоритма принятия решений // *Вестник Московского университета. Серия 5: География*. 2019. № 2. С. 55–60.
9. Dincer, N. G., Akkus, O. A new fuzzy time series model based on robust clustering for forecasting of air pollution // *Ecol. Inform.* 2018. N 43. P. 157–164.
10. Lu Bai, Jianzhou Wang, Xuejiao Ma, Haiyan Lu. Air Pollution Forecasts: An Overview // *International Journal of Environmental Research and Public Health*, 2018.
11. Mishra, D.; Goyal, P. NO2 forecasting models Agra // *Atmos. Pollut. Res.* 2015. N 6. P. 99–106.
12. Tsujitaa W., Yoshinoa A., Ishidab H., and Moriizumi T. Gas sensor network for air-pollution monitoring // *Sensors Actuators B, Chem.* 2005. Vol. 110. N 2. P. 304–311.
13. Савиных В. П., Майоров А. А., Матерухин А. В. Построение пространственной модели загрязнения воздуха на основе использования потоков данных от сетей геосенсоров // *Геодезия и картография*. 2017. № 12, том 78. С. 39–43.
14. Rayane El Sibai, Yousra Chabchoub, Raja Chiky, Jacques Demerjian, Kablan Barbar. Assessing and Improving Sensors Data Quality in Streaming Context // *9th International Conference, Computational Collective Intelligence ICCCI 2017, Nicosia, Cyprus, September 27–29, 2017, Proceedings, Part II, 2017, Vol. 10449. P. 590–599*.
15. Rai A., Kumar P., Pilla F., Skouloudis A., Sabatino S., Ratti C., Yasar A., and Rickerby D. End-user perspective of low-cost sensors for outdoor air pollution monitoring // *Sci. Total Environ.* 2017. Vol. 607–608. P. 691–705.
16. Duckham Matt, Nussbaum Doron, Sack Jorg Rudiger, Santoro Nicola. Efficient, Decentralized Computation of the Topology of Spatial Regions // *IEEE Transactions on Computers*. 2011. Vol. 60. N 8. P. 1100–1113.
17. Шахов В. В., Юргенсон А. Н., Соколова О. Д. Эффективный метод генерации случайных геометрических графов для моделирования беспроводных сетей // *ПДМ*. 2016. № 4 (34). С. 99–108.
18. Bieringer P. E.; Young G. S.; Rodriguez L. M.; Annunzio, A. J.; Vandenberghe F., Haupt S. E. Paradigms and commonalities in atmospheric source term estimation methods // *Atmospheric Environment*. Elsevier BV. 2017. N 156. P. 102–112.

19. Issartel J.-P. Emergence of a tracer source from air concentration measurements, a new strategy for linear assimilation // *Atmospheric Chemistry and Physics*, Copernicus GmbH, 2005. N 5. P. 249–273.
20. Mamonov A. V. Tsai Y.-H. R. Point source identification in nonlinear advection-diffusion-reaction systems // *Inverse Problems*, IOP Publishing, 2013. N 29. 035009.
21. Penenko V. V.; Penenko, A. V.; Tsvetova E. A. Gochakov A. V. Methods for Studying the Sensitivity of Air Quality Models and Inverse Problems of Geophysical Hydrothermodynamics // *Journal of Applied Mechanics and Technical Physics*, Pleiades Publishing Ltd, 2019. N 60. P. 392–399.
22. Penenko A., Mukatova Z. S., Penenko V., Gochakov A., Antokhin P. Numerical study of the direct variational algorithm of data assimilation in urban conditions // *Atmospheric and ocean optics*. 2018. N 31. P. 456–462.
23. Sun C., Li V., Lam J., and Leslie I. Optimal Citizen-Centric Sensor Placement for Air Quality Monitoring: A Case Study of City of Cambridge, the United Kingdom // *IEEE Access*, 2019. Vol. 7.
24. Соколова О. Д., Юргенсон А. Н. Задача оптимального размещения устройств анализа информационных потоков в сетях // *Проблемы информатики*. 2010. № 2. С. 33–42.
25. Sääskilahti Kirsti, Kangaskorte Riitta, Luimula Mika & Yli-Hemminki, Juha. Collecting and visualizing wireless geosensor data using mobile devices // *ACM International Conference Proceeding Series*. 2010. 10.1145/1823854.1823882.
26. Ilka A. R., Gilberto C., Renato A., Antnio M. V. M. Data-Aware Clustering for Geosensor Networks Data Collection // *Anais XIII Simpsio Brasileiro de Sensoriamento Remoto*, INPE, 2007. P. 6059–6066.
27. Шахов В. В., Мигов Д. А., Соколова О. Д. Беспроводные сенсорные сети, оснащенные средствами получения энергии из окружающей среды // *Проблемы информатики*. 2014. № 4. С. 69–79.
28. Матерухин А. В., Шахов В. В., Соколова О. Д. Модели процессов сбора пространственно-временных данных с использованием мобильных стоков // *Геодезия и картография*. 2018. Т. 79. № 12. С. 22–28.
29. Boubrima A., Bechkit W., Rivano H. Optimal WSN deployment models for air pollution monitoring // *IEEE Trans. Wireless Commun.* 2017. Vol. 16. N 5. P. 2723–2735.
30. Kaivonen Sami, Ngai Edith. Real-time air pollution monitoring with sensors on city bus // *Digital Communications and Networks*. 2019.
31. You-Chiun Wang, Guan-Wei Chen. Efficient Data Gathering and Estimation for Metropolitan Air Quality Monitoring by Using Vehicular Sensor Networks // *IEEE Transactions On Vehicular Technology*. 2017. Vol. 66. N 8. P. 7234–7248.
32. Penza M., Suriano D., Villani M. G., Spinelle L., and Gerboles M. Towards air quality indices in smart cities by calibrated low-cost sensors applied to networks // in *Proc. IEEE Conf. Sensors*, 2014. P. 2012–2017.
33. Shakhov V., Sokolova O. Towards Air Pollution Detection with Internet of Vehicles // *Proceedings of 15th International Asian School-Seminar Optimization Problems of Complex Systems. OPCS-2019*. P. 183–187.
34. Murty, R. N.; Mainland, G.; Rose, I.; Chowdhury, A. R.; Gosain, A.; Bers, J.; Welsh, M. CitySense: An Urban-Scale Wireless Sensor Network and Testbed // In *Proceedings of the 2008 IEEE Conference on Technologies for Homeland Security*, Waltham, MA, USA, 12–13 May 2008. P. 583–588.
35. Shu-Chiung Hua, You-Chiun Wang, Chiuan-Yu Huang, Yu-Chee Tsenga. Measuring air quality in city areas by vehicular wireless sensor networks // *The Journal of Systems and Software*. 2011. N 84. P. 205–212.
36. Materukhin A. V., Gvozdev O. G., Maiorov A. A. and Sokolova O. D. Simulation of Spatio-Temporal Data Streams from Geosensors Located on Mobile Objects // *15th International Asian*

School-Seminar Optimization Problems of Complex Systems (OPCS), Novosibirsk, Russia, 2019. P. 179–182. DOI: 10.1109/OPCS.2019.8880188.



Соколова Ольга Дмитриевна — канд. техн. наук, старший научный сотрудник Института вычислительной математики и математической геофизики СО РАН. Научные интересы: математическое моделирование, современ-

ные сети передачи данных, беспроводные сенсорные сети, задачи оптимизации. E-mail: olga@rav.sscs.ru;

Olga Sokolova is a Senior Researcher of Institute of Computational Mathematics and Mathematical Geophysics (ICMMG SB RAS). PhD in computer science. Research interests: modern data networks, wireless sensor networks, simulation, optimization problems.

Матерухин Андрей

Викторович — доктор техн. наук, доцент Московского государственного университета геодезии и картографии; e-mail: crypto.miigaik@gmail.com;

Матерухин Андрей Викторович окончил факультет вычислительной математи-



ки и кибернетики МГУ им. М. В. Ломоносова по специальности „Прикладная математика“. Защитил диссертацию на соискание ученой степени кандидата технических наук по специальности „Геоинформатика“, диссертацию на соискание ученой степени доктора технических наук по этой же специальности в Московском государственном университете геодезии и картографии (МИИГАиК).

Основные научные интересы: обработка и анализ больших пространственных данных и потоков пространственно-временных данных.

Maturukhin Andrei Victorovich — graduated from the faculty of Computational Mathematics and Cybernetics of Lomonosov Moscow State University (specialty “applied mathematics”); earned the degree of Ph.D. (specialty Geoinformatics) in Moscow State University of Geodesy and Cartography (MIIGAiK); the degree of Dr. habil. is conferred in MIIGAiK (specialty Geoinformatics). Current position — Associate Professor at the Department of Data Acquisition Systems, MIIGAiK.

Major research interests: processing and analysis of big spatial data and spatio-temporal data streams.

Дата поступления — 22.02.2020