

Л.И.Великанова, Н.М.Лыченко

Институт машиноведения и автоматизации НАН КР, Бишкек, Кыргызстан

E-mail: ludvel@mail.ru, nlychenko@mail.ru

МУЛЬТИРЕГРЕССИОННЫЕ И ОБОБЩЕННО-РЕГРЕССИОННЫЕ НЕЙРОСЕТЕВЫЕ МОДЕЛИ КРАТКОСРОЧНОГО ПРОГНОЗА ЗАГРЯЗНЕНИЯ PM2.5 В Г. БИШКЕК С УЧЕТОМ МЕТЕОРОЛОГИЧЕСКИХ ПАРАМЕТРОВ

Разработаны модели краткосрочного прогноза загрязнения атмосферного воздуха г. Бишкек частицами PM2.5 для летнего периода времени на основе линейной мультирегрессионной модели и обобщенно-регрессионной нейросетевой модели (GRNN) и с учетом метеорологических параметров. При разработке нейросетевой модели предложено декомпозировать исходные данные по срокам измерений. Это позволило уменьшить ошибку прогноза в сравнении с мультирегрессионной моделью.

Ключевые слова: прогноз концентраций PM2.5, мультирегрессионные модели, обобщенно-регрессионная нейронная сеть GRNN, ошибка прогноза.

Введение. Прогноз загрязнений воздуха в приземном слое атмосферы – актуальная и одна из основных природоохранных задач, которой в мире в последнее время уделяется все большее внимание. Зная ожидаемые показатели загрязнения, можно планировать мероприятия по защите населения от неблагоприятного воздействия загрязнителей, находящихся в атмосферном воздухе, решать задачи управления процессами загрязнения, ограничивая, например, воздействия источников загрязнений. Поскольку для г. Бишкек проблема загрязнения атмосферного воздуха стоит крайне остро, то и построение прогностических моделей на основе актуальных для города данных – важная задача, требующая исследования. В [1] опубликованы результаты построения прогностической модели из класса ARIMA-моделей (авторегрессионных интегрированных моделей со скользящим средним), которая может быть использована для прогноза концентрации PM2.5 на основе данных наблюдений за концентрациями за прошедшие четверо суток – для зимнего периода и двое суток – для летнего периода. Однако интересной представляется задача прогноза загрязнений с учетом метеорологических факторов. В литературе достаточно широко представлены соответствующие публикации, краткий обзор некоторых из них приведен в [2]. В [2] проведена оценка влияния метеорологических факторов (таких, как скорость ветра, температура, относительная влажность воздуха, температура точки росы, интенсивность осадков и атмосферное давление) на процесс загрязнения воздуха г. Бишкек частицами PM2.5 в период с 9 февраля по 25 ноября 2019 г. Показано, что существует слабая и умеренная корреляция (как положительная, так и отрицательная) между концентрациями PM2.5 и метеорологическими параметрами, измеренными в текущий и прошлые сроки. Представляется интересным использование этой информации в прогностических моделях на основе множественной регрессии и нейросетевых технологий [3-6], позволяющих одновременно учитывать большое количество влияющих параметров, и наиболее полно отражать характерные для исследуемой системы причинно-следственные связи между параметрами.

В настоящей работе представлены результаты построения модели краткосрочного прогноза загрязнения воздуха г. Бишкек частицами PM2.5 с учетом метеорологических

факторов для летнего периода времени (24/03/2019 – 31/11/2019) на основе обобщенно-регрессионной нейронной сети (GRNN) и сравнения эффективности полученной модели с линейной мультирегрессионной моделью. Информационная база исследования формировалась на основе данных загрязнения частицами Pm2.5 атмосферного воздуха в городе Бишкек [7] и архивных данных метеослужбы [8].

Модели и методы. *Мультирегрессионные модели.* Общее уравнение линейной множественной регрессии может быть выражено как [9]:

$$y_t = b + k_1 x_{1t} + k_2 x_{2t} + \dots + k_m x_{mt} + e_t \quad (1)$$

где b – константа регрессии, k_1, k_2, \dots, k_m – коэффициенты регрессии, m – количество независимых переменных (факторов, предикатов). Значения константы и коэффициентов определяются с использованием метода наименьших квадратов, который минимизирует ошибку $e_t = (y_t - y_{ft})$ или остатки модели. Для определения показателя качества соответствия линейной модели обычно используется коэффициент детерминации R^2 или множественный коэффициент корреляции R :

$$R^2 = 1 - \frac{\sum_{t=1}^n (y_t - y_{ft})^2}{\sum_{t=1}^n (y_t - \bar{y})^2}.$$

Для построения регрессионной модели в настоящем исследовании был использован метод ступенчатой (пошаговой) регрессии, суть которого заключается в отборе из большого количества предикатов x небольшой подгруппы переменных, которые вносят наибольший вклад в вариацию зависимой переменной y . При этом в модель регрессии (1) последовательно включаются факторы, и после этого проверяется их значимость, т.е. определяется, насколько уменьшается сумма квадратов остатков и увеличивается величина R^2 (прямой метод). Либо в уравнение регрессии включаются все потенциальные независимые переменные и затем устраняются те, которые не являются статистически значимыми на основе t-критерия Стьюдента. Либо используются оба метода, при этом выбор переменной прекращается, когда переменные больше не соответствуют критериям ввода и удаления. Фактор является незначимым, если его включение в уравнение регрессии только изменяет значение коэффициентов регрессии, не уменьшая значительно среднюю сумму квадратов остатков. Если при включении в модель соответствующего факторного признака величина множественного коэффициента корреляции R увеличивается, а коэффициент регрессии не изменяется (или меняется незначительно), то данный признак существен и его включение в уравнение регрессии необходимо. При построении мультирегрессионной модели концентрации PM2.5 рассматривались как зависимые переменные. Метеопараметры: температура воздуха, температура точки росы, влажность воздуха, скорость ветра, давление, интенсивность осадков рассматривались как независимые переменные.

Нейросетевые модели GRNN. Модели на основе искусственных нейронных сетей (ИНС) активно используются в прогнозировании, поскольку они обладают особенностью, состоящей в том, что сеть умеет находить зависимости между данными путем анализа большого количества подобных примеров или паттернов. Существуют различные варианты архитектуры нейронных сетей, однако в настоящем исследовании использовалась обобщенно-регрессионная нейронная сеть GRNN (Generalized Regression Neural Network) [10], что обусловлено следующими ее преимуществами:

- возможность моделирования нелинейных связей между входными и выходными параметрами;
- архитектура сети фиксирована и не нуждается в определении;
- высокая скорость обучения.

Архитектура сети GRNN показана на рисунке 1. Первый слой – это радиальный базисный слой с числом нейронов, равным числу элементов Q обучающего множества; в качестве начального приближения для матрицы весов выбирается массив \mathbf{P}^1 ; смещение \mathbf{b}^1 устанавливается равным вектор-столбцу с элементами $0.8326/spread$. Функция $dist$ вычисляет расстояние между вектором входа и вектором веса нейрона; вход функции активации \mathbf{n}^1 равен поэлементному произведению взвешенного входа сети на вектор смещения; выход каждого нейрона первого слоя \mathbf{a}^1 является результатом преобразования вектора \mathbf{n}^1 радиальной базисной функцией $radbas$. Если вектор веса нейрона равен транспонированному вектору входа, то взвешенный вход равен 0, а выход функции активации – 1. Если расстояние между вектором входа и вектором веса нейрона равно $spread$, то выход функции активации будет равен 0.5. Выход первого слоя можно выразить формулой

$$A^1 = radbas(\|W - X\| \cdot B),$$

где A^1 – выход первого слоя; функция $radbas$ – радиально-базисная функция; W – матрица весов первого слоя сети; X – входной вектор; B – вектор порогов первого слоя. Второй слой – это линейный слой с числом нейронов, также равным числу элементов Q обучающего множества, причем в качестве начального приближения для матрицы весов LW^{21} выбирается массив \mathbf{T} . Во втором слое используется блок `normprod` для вычисления нормированного скалярного произведения строки массива весов LW^{21} и вектора входа \mathbf{a}^1 в соответствии со следующим соотношением:

$$n^2 = \frac{LW^{21}a^1}{sum(a^1)}$$

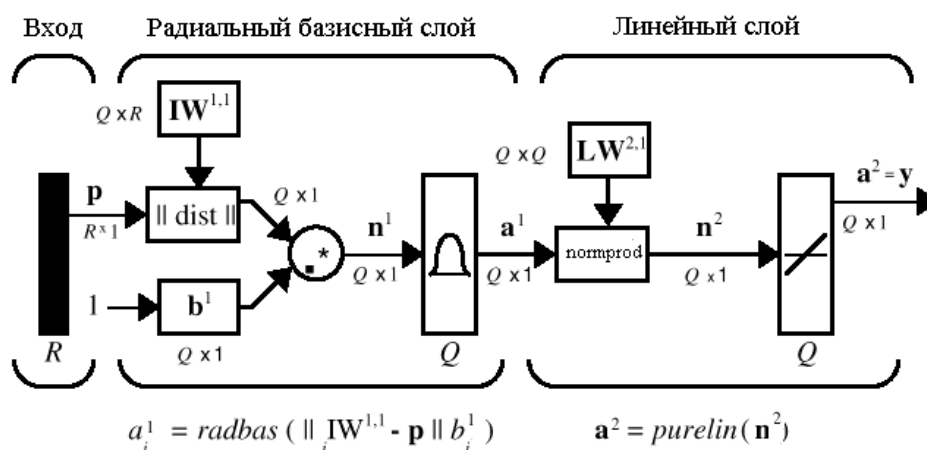


Рисунок 1 – Архитектура сети GRNN

Предположим [10], что имеем вектор входа \mathbf{p}_i , близкий к одному из векторов входа \mathbf{p} из обучающего множества. Этот вход \mathbf{p} генерирует значение выхода слоя \mathbf{a}_i^1 , близкое к 1. Это приводит к тому, что выход слоя 2 будет близок к \mathbf{t}_i . Если параметр влияния $spread$ мал, то радиальная базисная функция характеризуется резким спадом и диапазон входных значений, на который реагируют нейроны скрытого слоя, оказывается весьма малым. С увеличением параметра $spread$ наклон радиальной базисной функции становится более гладким, и в этом случае уже несколько нейронов реагируют на значения вектора входа. Тогда на выходе сети формируется вектор, соответствующий среднему нескольких целевых векторов, соответствующих входным векторам обучающего множества, близких к данному вектору входа. Чем больше значение параметра $spread$, тем большее число нейронов участвует в формировании среднего значения, и в итоге функция, генерируемая сетью, становится более гладкой.

Для оценки эффективности моделей и сравнения их прогностических свойств были использованы следующие оценки [9].

1. Средняя квадратичная ошибка (Mean Squared Error, *MSE*):

$$MSE = \frac{1}{N} \sum_{t=1}^N (e_t)^2,$$

где $e_t = (y_t - y_{ft})$ - ошибка прогноза или остатки модели.

Эта ошибка сильно подвержена влиянию выбросов (из-за квадрата в формуле). Поэтому в результате использования этого критерия предпочтение отдаётся менее точным моделям (посредственно спрогнозировавшим данные, но при этом не допустившим ни одной большой ошибки).

2. Квадратный корень из средней квадратичной ошибки (Root Mean Squared Error, *RMSE*):

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (e_t)^2}.$$

3. Средняя абсолютная ошибка (Mean Absolute Error, *MAE*):

$$MAE = \frac{1}{N} \sum_{t=1}^N |e_t|.$$

Эту ошибку трудно интерпретировать, с ее помощью можно проводить сравнение только между моделями по одному ряду данных.

4. Средняя абсолютная процентная ошибка (Mean Absolute Percentage Error, *MAPE*):

$$MAPE = \frac{1}{N} \sum_{t=1}^N \frac{|e_t|}{y_t} 100\%$$

Это коэффициент, не имеющий размерности, с очень простой интерпретацией: ошибка составила $n\%$ от фактических значений. Из-за деления на фактические значения этот коэффициент чувствителен к масштабу переменных (при y_t , близких к нулю, *MAPE* стремительно возрастает).

Построение линейной мультирегрессионной модели. Описание исходных данных для исследования представлено в [2]. Восстановление пропущенных значений временных рядов осуществлялось согласно алгоритму, предложенному в [11]. Архивные данные были разделены на 2 части: обучающее множество – данные измерения $PM_{2.5}$ и метеорологических параметров по г. Бишкек за период с 24.03.2019 по 31.07.2019 г.; тестовое множество – данные измерения за период с 01.08.2019 по 18.08.2019 г.

Из таблиц 1–6 [2] видно, что для летнего периода времени (24/03/2019 – 31/10/201) наилучшие взаимосвязи обнаружены между концентрациями $PM_{2.5}$ и измерениями температуры воздуха T в предшествующий срок измерений $i-1$ ($R=0.1768$), температуры точки росы Td в срок измерений $i-2$ ($R=-0.1857$), интенсивностью осадков PR в срок измерений $i-2$ ($R=-0.1899$), влажностью RH в срок измерений $i-1$ ($R=-0.2493$), скоростью ветра в срок измерений $i-5$ ($R=-0.1280$). Корреляция между концентрациями $PM_{2.5}$ и давлением незначительная: $R=-0.0795$. Используя эту информацию в качестве исходной, на обучающей выборке 24/03/2019 – 29/08/2019 была построена линейная мультирегрессионная модель (см. таблицу 1), учитывающая все перечисленные относительно существенные факторы (температура точки росы, температура воздуха, относительная влажность, скорость ветра, интенсивность осадков):

$$PM_{2.5}(i) = b + k_1 * Td(i-2) + k_2 * T(i-1) + k_3 * PR(i-2) + k_4 * Ws(i-5) + k_5 * RH(i-1).$$

Таблица 1 – Линейные мультирегрессионные модели

Teaching set: 24/03/2019 – 31/07/2019									
$PM2.5(i) = b + k_1 * Td(i-2) + k_2 * T(i-1) + k_3 * PR(i-2) + k_4 * Ws(i-5) + k_5 * RH(i-1)$									
b	k_1	k_2	k_3	k_4	k_5	R	$MAPE$	MAE	MSE
13.1054	-0.4092	0.4569	-1.9980	-0.0671	0.0550	0.3183	0.284	7.440	134.4
$PM2.5(i) = b + k_1 * Td(i-2) + k_2 * T(i-1) + k_3 * PR(i-2) + k_4 * PM2.5(i-1) + k_5 * PM2.5(i-2)$									
b	k_1	k_2	k_3	k_4	k_5	R	$MAPE$	MAE	MSE
10.4621	-0.4235	0.4930	-2.9192	0.0537	0.0432	0.3514	0.2772	5.867	70.11
Testing set: 01/08/2019-18/08/2019									
$PM2.5(i) = b + k_1 * Td(i-2) + k_2 * T(i-1) + k_3 * PR(i-2) + k_4 * PM2.5(i-1) + k_5 * PM2.5(i-2)$									
b	k_1	k_2	k_3	k_4	k_5	R	$MAPE$	MAE	MSE
10.4621	-0.4235	0.4930	-2.9192	0.0537	0.0432	0.3514	0.2087	4.395	43.57

В таблице 1 также представлены коэффициенты множественной корреляции R между концентрациями $PM2.5$ и метеорологическими параметрами, средняя квадратическая ошибка модели MSE , средняя абсолютная ошибка MAE , средняя абсолютная процентная ошибка $MAPE$.

Затем, используя метод пошаговой множественной линейной регрессии, получена итоговая регрессионная модель, также представленная в таблице 1:

$$PM2.5(i) = b + k_1 * Td(i-2) + k_2 * T(i-1) + k_3 * PR(i-2) + k_4 * PM2.5(i-1) + k_5 * PM2.5(i-2).$$

В этой модели исключены факторы: скорость ветра Ws и относительная влажность RH как имеющие наименьшие регрессионные коэффициенты. Для учета инерционности процесса загрязнения в модель включены измерения $PM2.5$ в $(i-1)$ -ый и $(i-2)$ -ой сроки измерения.

Полученная модель применена для восстановления концентрации $PM2.5$ на тестовой выборке: с 05/09/2019 по 15/09/2019. На рис. 5 представлены вычисленные на основе этой модели и измеренные (наблюденные) значения концентраций.

Построение нейросетевой модели. На первом этапе были исследованы различные варианты обучения и моделирования прогнозных значений $PM2.5$ с использованием вышеприведенных параметров (исключая давление, как несущественный фактор) с различной историей и прогнозом метеослужбы на всем диапазоне сроков измерения i (полная выборка, $i = 3, 6, 9, 12, 15, 18, 21, 24$ час). Данные были разделены на 2 части:

- обучающее множество – данные измерения $Pm2.5$ и метеорологических параметров по г. Бишкек за период с 24.03.2019г. по 31.07.2019г. (1040 сроков измерения);
- тестовое множество – данные измерения за период с 01.08.2019 по 18.08.2019 г. (144 срока измерения).

Было проведено обучение ИНС GRNN для 10 вариантов входных векторов (см. таблицу 2). Размеры векторов варьировались от 5 до 14 признаков. Здесь, так же, как и в [2], $i-1, i-2$ – сроки предыдущих измерений (на 3 часа и на 6 часов ранее).

Таблица 2. Варианты наборов признаков во входных векторах при обучении и моделировании ИНС

№ варианта	Контролируемые параметры						Размер входного вектора
	Pm2.5	Ws	PR	Td	T	RH	
1	<i>i-1</i>	<i>i</i>	<i>i</i>	<i>i</i>		<i>i</i>	5
2	<i>i-1</i>	<i>i</i>	<i>i</i>		<i>i</i>	<i>i</i>	5
3	<i>i-1</i>	<i>i</i> <i>i-1</i>	<i>i</i> <i>i-1</i>	<i>i-1</i>		<i>i-1</i>	7
4	<i>i-1</i>	<i>i</i> <i>i-1</i>	<i>i</i> <i>i-1</i>		<i>i-1</i>	<i>i-1</i>	7
5	<i>i-1</i>	<i>i</i> <i>i-1</i>	<i>i</i> <i>i-1</i>	<i>i</i> <i>i-1</i>		<i>i</i> <i>i-1</i>	9
6	<i>i-1</i>	<i>i</i> <i>i-1</i>	<i>i</i> <i>i-1</i>	<i>i</i> <i>i-1</i>	<i>i</i> <i>i-1</i>	<i>i</i> <i>i-1</i>	11
7	<i>i-1</i> <i>i-2</i>	<i>i</i> <i>i-1</i>	<i>i</i> <i>i-1</i>	<i>i</i> <i>i-1</i>		<i>i</i> <i>i-1</i>	10
8	<i>i-1</i> <i>i-2</i> <i>i-3</i> <i>i-4</i>	<i>i</i> <i>i-1</i>	<i>i</i> <i>i-1</i>	<i>i</i> <i>i-1</i>			10
9	<i>i-1</i> <i>i-2</i> <i>i-3</i> <i>i-4</i>	<i>i</i> <i>i-1</i>	<i>i</i> <i>i-1</i>	<i>i</i> <i>i-1</i>		<i>i</i> <i>i-1</i>	12
10	<i>i-1</i> <i>i-2</i> <i>i-3</i> <i>i-4</i>	<i>i</i> <i>i-1</i>	<i>i</i> <i>i-1</i>	<i>i</i> <i>i-1</i>	<i>i</i> <i>i-1</i>	<i>i</i> <i>i-1</i>	14

Результаты моделирования PM2.5 при использовании полной выборки показали неудовлетворительные результаты для всех вариантов модели. Среднеквадратическая ошибка *MSE* принимала значения от 59,93 до 81,38, средняя абсолютная процентная ошибка *MAPE* – от 25,38% до 31,10%. Одна из причин этого в том, что данные не вполне отвечают требованиям, предъявляемым к обучающим множествам нейронных сетей. Исходные данные обязательно должны быть непротиворечивы. Между тем, для одинаковых входных векторов нередко имеются резко отличающиеся значения функции цели, что объясняется влиянием дополнительных природных и техногенных факторов, например, направление ветра, возгорание отдельных объектов и т.п. Для решения этой проблемы необходимо увеличить размерность пространства признаков (количество компонент входного вектора, соответствующего образцу). Но при увеличении размерности пространства признаков может возникнуть ситуация, когда число примеров может стать недостаточным для обучения сети, и она вместо обобщения просто запомнит примеры из обучающей выборки и не сможет корректно функционировать. Решением этой проблемы может стать увеличение количества входных векторов обучающего множества за счет новых наблюдений.

Кроме того, как видно из рисунка 1 [2], диапазон значений PM2.5 достаточно широкий. Способность нейронной сети к обобщению на таком диапазоне также может привести к потере точности обучения. Для уменьшения разброса значений PM2.5 была

проведена декомпозиция значений параметров по срокам измерения. Размеры обучающих и тестовых множеств сократились в 8 раз (130 примеров обучающих множеств и 18 – тестовых). На рисунке 2 приведены графики значений PM2.5 тестового множества в сроки i ($i=1,8$). Из этого рисунка наглядно видно, что в большинстве случаев все значения PM2.5 в летний период находятся в пределах принятой в Кыргызстане нормы (35), за исключением срока 21 час, в который достаточно много значений эту норму превышают. Основное влияние на концентрацию в воздухе PM2.5 в г. Бишкек в летнее время оказывает интенсивность транспортной нагрузки.

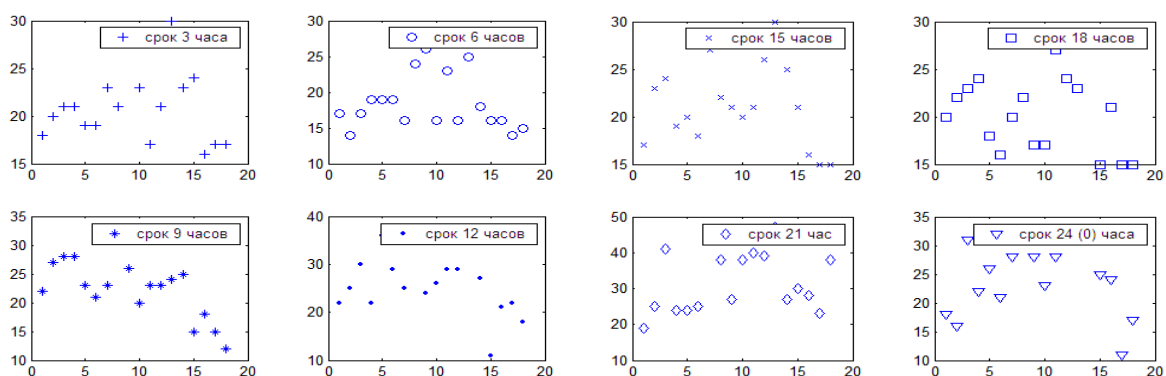


Рисунок 2 – Значения Pm2.5 за тестовый период в сроки измерения метеоданных i ($i=1,8$)

Для каждого срока прогнозирования i значения Pm2.5 были сформированы обучающие и тестовые множества для 10 вариантов набора параметров, и проведено обучение и моделирование всех ИНС $_{ij}$ ($i=1,8, j=1,10$).

В таблице 3 приведены средние абсолютные процентные ошибки прогнозирования PM2.5. Здесь С.С. – среднесуточная ошибка, С.М. – средняя для модели. Как видно, ошибка прогноза существенно уменьшилась.

Таблица 3 – Средние абсолютные процентные ошибки (в %) прогнозирования Pm2.5 (декомпозиция по срокам измерения)

Сроки измерения	3	6	9	12	15	18	21	24(0)	С.С.
№ варианта	MAPE								
1	17,43	18,68	35,17	26,04	17,70	22,63	28,65	41,22	25,94
2	16,20	18,54	36,17	24,49	16,33	21,87	41,76	28,32	25,46
3	17,01	19,82	24,01	24,77	12,88	19,57	32,03	46,93	24,63
4	16,85	20,84	23,22	27,78	15,99	17,31	33,26	44,53	24,97
5	13,57	14,29	31,13	26,67	18,41	19,25	32,67	49,45	25,68
6	17,00	15,82	27,01	22,55	19,03	22,25	24,23	47,31	24,40
7	19,61	21,19	33,66	26,06	19,42	18,86	20,77	32,42	24,00
8	21,33	15,83	33,05	24,85	24,13	20,91	29,63	29,36	24,89
9	18,20	16,30	34,68	22,05	22,21	20,09	31,80	32,04	24,67
10	18,88	16,61	36,32	25,81	19,65	20,73	30,55	34,10	25,33
С.М.	17,77	17,69	30,65	24,97	19,33	21,91	29,45	39,46	

Из таблицы 3 видно, что самая маленькая ошибка прогноза соответствует наименьшей транспортной загруженности (3 и 6 часов местного времени). Затем следует период от 12 до 18 часов, и самые большие ошибки прогноза в периоды максимального движения транспорта.

На рис. 3 представлены графики наблюдаемых значений концентраций PM2.5 и прогнозные значения, полученные на моделях с полными данными и данными, декомпозированными по срокам измерений: а) – для 3-х часов (модель 1); б) – для 6 часов (модель 7).

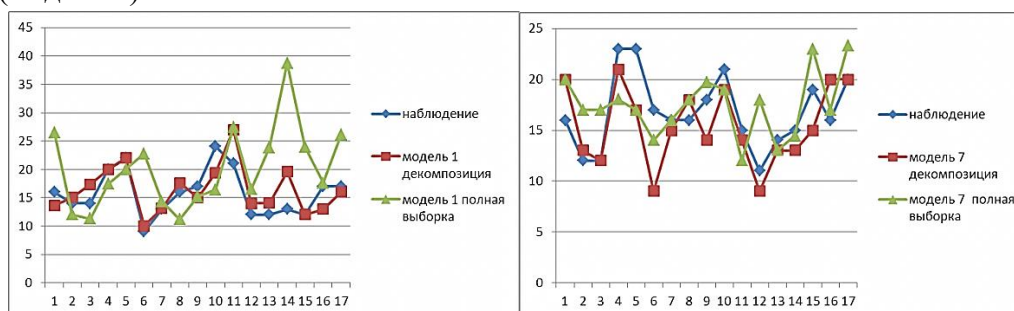


Рисунок 3 – Графики наблюдаемых значений концентраций PM2.5 и прогнозные значения в срок 3 часа (а) и в срок 6 часов (б)

Рассмотрим влияние размерности входных векторов на ошибку прогноза на примере столбцов со сроками 3 и 6 часов таблицы 3. Наименьшую ошибку для этих сроков имеет вариант 5. Обучающее множество для этого варианта состоит из 9-ти размерных векторов, учитывающих прогноз метеопараметров и их предшествующие значения на один срок. Увеличение размерности этих векторов с целью учета истории значений параметров вызывает увеличение ошибки (варианты 6–10). Это связано с недостаточным количеством примеров в обучающих выборках. Существенное увеличение количества наблюдений позволит в дальнейшем рассматривать варианты с большей размерностью векторов.

Варианты 1 и 2, 3 и 4 попарно отличаются между собой использованием температуры воздуха или точки росы. Это не вызывает существенного отличия ошибок, однако использование одновременно обоих этих параметров ошибку увеличивает (вариант 6), что связано с увеличением размерности векторов. Сравнение вариантов 8 и 9 показывает, что добавление во входные векторы параметра «влажность воздуха», хотя и увеличивает размерность векторов, но в некоторые сроки уменьшает ошибки, т.е. является информативно значащим параметром.

В итоге, проверка 10 моделей на выборке с декомпозицией по срокам измерения выявила, что для сроков измерения 3, 6, 15 и 18 часов наиболее приемлемым является 5-ый вариант модели прогноза концентрации Pm2.5. Для увеличения точности прогноза по всем срокам необходимо проводить дальнейшие наблюдения за контролируемыми параметрами и наращивать количество примеров обучающего множества.

Обсуждение результатов. Сравнение полученных мультирегрессионной и нейросетевой GRNN моделей прогноза производилось на выборке с 5.09.2019 по 15.09.2019. Ошибки представлены в таблице 4. Как видно, средняя абсолютная процентная ошибка *MAPE* нейросетевой модели незначительно больше мультирегрессионной. Однако средняя абсолютная ошибка, среднеквадратическая ошибка –

меньше. Собственно средняя ошибка прогноза $e = (y_t - y_{ft})$ у нейросетевой модели значительно меньше. Это иллюстрируют и графики (рисунок 5), на котором представлены наблюдаемые (измеренные) значения концентраций PM2.5 и их прогнозные значения на 3 часа вперед (что соответствует одному сроку наблюдений).

Таблица 4. Ошибки прогнозирования мультирегрессионной и нейросетевой моделей.

Период прогноза: 5.09.2019 – 15.09.2019	MAPE	MAE	MSE	$mean(y_t - y_{ft})$
Мультирегрессионная модель (MR-Model)	0.1935	4.6084	52.7004	5.7681
Нейронная сеть GRNN (ANN-Model)	0.1980	4.3332	43.6020	0.7970

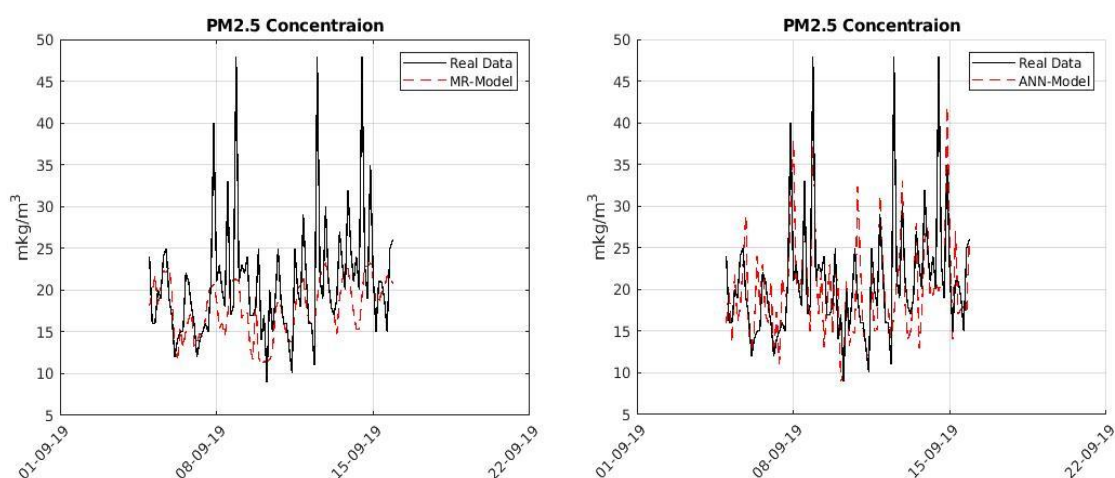


Рисунок 5 – Сравнение мультирегрессионной и нейросетевой GRNN моделей.

Таким образом, прогнозная модель на основе ИНС GRNN более предпочтительная с точки зрения точности прогноза, чем линейная мультирегрессионная, построенная на основе анализа корреляций концентрации загрязнений PM2.5 и метеопараметров. При этом идея декомпозировать данные концентраций PM2.5 и метеопараметров по срокам измерения оправдала себя, поскольку позволила снизить ошибку прогноза. Однако прогноз загрязнения атмосферного воздуха частицами PM2.5 по г. Бишкек проводился по данным лишь одного сенсора, показания которого могут быть зашумлены локальными явлениями (сжигание листьев или бытового мусора, скопление автомобилей в зоне измерения, подметание улицы и т.п.) и не типичны для данного срока измерения и для города в целом. Для полноценного прогноза необходимо снабжение такими сенсорами различных районов города. Это позволит сгладить последствия локальных явлений и уменьшить ошибку прогноза.

Литература

1. Верзунов С.Н., Лыченко Н.М. Анализ и ARIMA-модели динамики изменения концентрации PM_{2.5} в атмосферном воздухе г. Бишкек // Проблемы автоматизации и управления. – 2019, №1 (36). – С. 21–30.
2. Лыченко Н.М. Регрессионный анализ метеорологических факторов и концентраций частиц PM_{2.5} в атмосферном воздухе г. Бишкек // Проблемы автоматизации и управления. – 2019. – №2(37). – С. 5–15.
3. Рощупкин Э.В. Нейросетевая модель прогноза загрязнения атмосферного воздуха // Известия ТулГУ. Технические науки. Вып.4.Ч.2, 2010. – С. 266–271.
4. Добровольская Л.А., Ключев Д.С. Прогнозирование степени загрязнения атмосферного воздуха в промышленном регионе // Вісник Приазовського державного університету, серія: Технічні науки, Вип. 36, 2018. – С. 216–222
5. Беляев Н.Н., Славинская Е.С., Кириченко Р.В. Прогноз загрязнения атмосферного воздуха выбросами автотранспорта с учетом химической трансформации вредных веществ // Наука та прогрес транспорту. Вісник Дніпропетровського національного університету залізничного транспорту, №3, 2017, с. 15–20.
6. Якшина Н.В. Нейросетевое моделирование процессов загрязнения окружающей среды // Диссертация на соискание ученой степени кандидата физико-математических наук, Екатеринбург, 2007.
7. AirNow Department of State // [https://airnow.gov/index.cfm?action=airnow.global_summary#U.S. Department of State \\$Bishkek](https://airnow.gov/index.cfm?action=airnow.global_summary#U.S.Department.of.State.Bishkek), (дата обращения: 30.11.2019).
8. Сайт «Погода и Климат» <http://www.pogodaiklimat.ru>, (дата обращения: 30.11.2019).
9. Современное прогнозирование. [Электронный ресурс]. URL: <https://forecasting.svetunkov.ru/etextbook/> (дата обращения: 30.09.2019).
10. Медведев В.С., Потемкин В.Г. Нейронные сети. МАТЛАБ 6 // Пакеты прикладных программ, Кн.4, 2002, 496 с.
11. Великанова Л.И., Лыченко Н.М. Восстановление пропусков в температурных временных рядах на базе нейросетевых моделей // Проблемы автоматизации и управления. – 2018. №2(35). – С. 51–58.