

УДК 504.064.2.001.18

## ПРОГНОЗИРОВАНИЕ СОДЕРЖАНИЯ АНОМАЛЬНО РАСПРЕДЕЛЕННОГО В ПОЧВЕ ХРОМА ГИБРИДНЫМИ МОДЕЛЯМИ НА ОСНОВЕ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ

© 2018 г. А.В. Шичкин, А.Г. Буевич, А.П. Сергеев,  
Е.М. Баглаева, И.Е. Субботина

*Институт промышленной экологии Уральского отделения Российской академии наук  
(ИПЭ УрО РАН),*

*ул. С. Ковалевской, д.20, г. Екатеринбург, 620219 Россия.*

*E-mail: and@ecko.uran.ru, bagalex3@gmail.com*

Поступила в редакцию 21.09.2017 г. После исправления 1.12. 2017 г.

Работа посвящена применению гибридной модели, сочетающей искусственные нейронные сети (ИНС) и кригинг, для прогнозирования аномально распределенного химического элемента хром (Cr). Известно, что сочетание в модели геостатистических интерполяционных подходов (кригинг) и нейронных сетей обеспечивает лучшую точность прогнозирования и производительность. Обобщенные регрессионные нейронные сети (Generalized regression neural networks – GRNN) и многослойные перцептроны (Multilayer perceptron – MLP) представляют собой классы нейронных сетей, широко используемых для моделирования в экологических исследованиях. В работе сравниваются две нейронные сети: GRNN и MLP, а также два комбинированных метода: GRNN с кригингом остатков (GRNNRK) и MLP с кригингом остатков (MLPRK). Исследование основано на реальных данных по поверхностному загрязнению почвы Cr в г. Новый Уренгой (Россия), полученных во время ранее проведенного скрининга. Структуры сетей выбраны во время компьютерного моделирования, основанного на минимизации RMSE. MLP и MLPRK показали наилучшую прогнозическую точность по сравнению с кригингом, GRNN и GRNNRK.

**Ключевые слова:** гибридные модели, искусственные нейронные сети, хром, кригинг остатков.

DOI: 10.7868/S0869780318030109

### ВВЕДЕНИЕ

Прогнозирование распределения химических элементов в верхнем слое почвы – важная задача исследований, связанных с экологическими проблемами во всем мире. Быстрая индустриализация за последние десятилетия способствовала значительному увеличению загрязнения почв в арктических и субарктических регионах России. Загрязнение почв, особенно на городских территориях, считается важным индикатором качества среды обитания в связи с возможными негативными последствиями для окружающей среды и здоровья людей.

Предварительный анализ эмпирических данных для различных географических областей показал гетерогенность пространственных распределений геохимических спектров [3]. Результаты мониторинга городских территорий сильно зависят от относительного положения и интен-

сивности источников выбросов, а также от особенностей строительства, метеорологических и гидрологических условий, изменчивости климата и других факторов. Эти процессы и факторы могут вызывать пространственную гетерогенность, а иногда и аномалии распределения загрязняющих веществ [8, 18, 19].

Один из наиболее широко используемых методов моделирования – интерполяция. Существует два основных типа используемых методов пространственной интерполяции: детерминированный и геостатистический. Детерминированный подход, при котором результаты точно определяются через известные отношения между состояниями и событиями, использует методы, которые вычисляют неизвестные значения, исходя из степени сходства. Методы геостатистической интерполяции (например, кригинг) используют статистические характеристики измеренных пятен вместе с пространственной автокорреляцией

между ними и учитывают пространственную конфигурацию пятен образца в месте прогнозирования. Кригинг – общий метод, используемый в пространственном прогнозировании, который оценивает значения для любой координаты без смещения и с минимальной дисперсией [7]. Среди интерполяторов, построенных на основе взвешенных средних, кригинг является лучшим интерполятором с несмещенной оценкой [10, 11]. Точность методов кригинга зависит от плотности и размера участков выборки, но часто невозможно получить необходимое количество образцов на обследуемом участке. Для устранения этого недостатка и повышения точности методов интерполяции при создании карт распределения с высоким разрешением требуется более эффективный метод.

В настоящее время одним из самых результативных методов являются искусственные нейронные сети (ИНС). ИНС предоставляют множество мощных методов решения задач прогнозирования, распознавания образов, анализа данных, контроля и многих других. Способность к обучению делает их незаменимыми при решении нестандартных и динамично меняющихся задач. Прогностическая точность, получаемая ИНС, часто выше, чем у других методов [8]. Модель ИНС может применяться к измеренным данным, полученным при мониторинге, может использоваться для прогнозирования содержания загрязняющих веществ в неконтролируемых местах [10, 15, 20, 21]. В традиционной модели ИНС пространственные координаты используются в качестве входов, а результат предсказания – как выход. Функциональная зависимость между входами и выходами устанавливается через сеть синаптических связей (весов). Эти веса определяются через процесс обучения с использованием итерационных процедур, которые иногда занимают много времени. Чтобы избежать проблемы локальных минимумов, приводящих к неоптимальному решению в процессе обучения, применимы некоторые алгоритмы оптимизации. Наиболее широко используется метод Левенберга-Маркварта.

В экологических исследованиях наиболее часто используется ИНС типа многослойного перцептрона (Multilayer perceptron – MLP). Благодаря широкому распространению этот тип сетей хорошо развит и показывает высокую производительность. Структура сети MLP описывается несколькими числами, относящимися к количеству нейронов в слоях: вход – скрытый слой – выход (например, 5–3–1 для трехслойного перцептрона означает 5 входных нейронов, 3 скрытых нейрона

в скрытом слое, 1 выходной нейрон). MLP широко используются в исследованиях, посвященных распределению химических элементов в почве [4, 6, 14], в частности тяжелых металлов [2]. Многие исследователи применяли MLP для оценки ресурсов [12, 17] и показали превосходство моделей MLP над геостатистикой.

Обобщенные регрессионные нейронные сети (Generalized regression neural networks – GRNN) также используются в качестве интерполяторов и известны как универсальные аппроксиматоры функций, которые могут научиться аппроксимировать любую непрерывную нелинейную функцию между наборами входов и выходов [16]. При обучении GRNN не требуют использования таких итерационных процедур, как алгоритмы обратного распространения ошибки. Они аппроксимируют функцию, составляя оценки непосредственно из набора данных обучения и минимизируя ошибку оценки путем расширения набора данных обучения.

Несмотря на очевидные преимущества нейронных сетей, они имеют и недостатки. Во-первых, большинство подходов к построению ИНС являются эвристическими и часто не приводят к однозначным решениям. Кроме того, проблемы, возникающие при подготовке учебной выборки, связаны со сложностью нахождения достаточного количества примеров обучения. Нейронные сети хорошо выявляют тренды, однако часто не могут идентифицировать небольшие колебания моделируемого признака. Комбинация различных методов способна нейтрализовать их слабые места и усилить их достоинства. В частности, интеграция кригинга в модель ИНС может привести к улучшению точности прогноза и уменьшению ошибок [5, 9, 11, 13].

Цель настоящей статьи – сравнение результатов прогноза уровня загрязнения хромом (Cr) верхнего слоя почвы на выбранном участке субарктической территории (г. Новый Уренгой, Россия), полученных с помощью двух моделей, основанных на ИНС (GRNN и MLP), а также двух гибридных моделей, объединяющих методы прогнозирования на основе ИНС и кригинга.

## МАТЕРИАЛЫ И МЕТОДЫ

Данные для исследования были получены по результатам почвенного обследования в г. Новый Уренгой, Ямало-Ненецкого автономного округа, Россия (N66.084°, E76.678°) [18, 19], где было отмечено аномальное распределение содержания

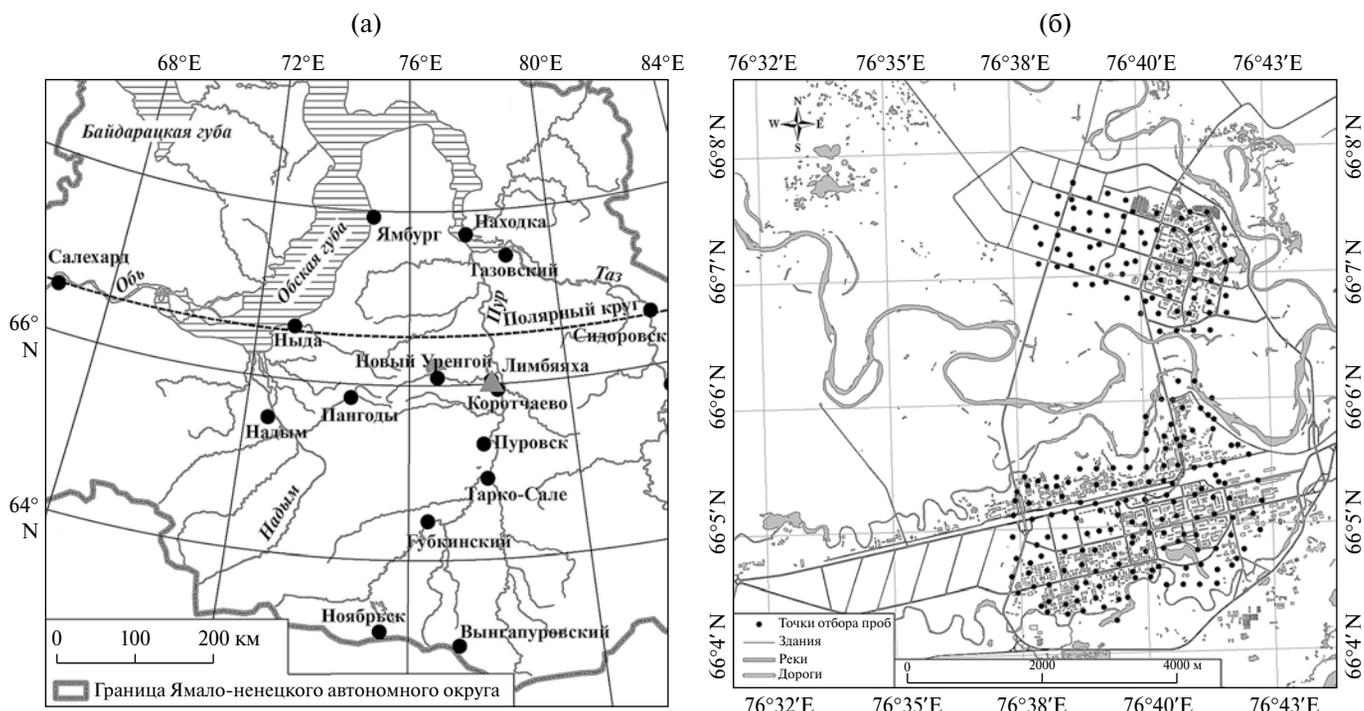


Рис. 1. Карта-схема места отбора проб: (а) Ямало-Ненецкий автономный округ, Россия; (б) г. Новый Уренгой, точками обозначены места отбора проб.

хрома (Cr) в поверхностном слое почвы. Площадь отбора проб составляла примерно  $8.5 \text{ км}^2$ . Ландшафт в месте отбора проб равнинный и покрыт торфяно-подзолисто-глеельвiallyно-гумусовым песчаным грунтом. Всего было собрано 150 образцов верхнего слоя почвы на глубину 0.05 м. Подробное пространственное расположение точек выборки показано на рис. 1.

Отбор проб почвы в черте города выполнен по точкам, расположенным в узлах квадратной сетки с шагом 250 м. Фактическое их расположение определялось при проведении опробования непосредственно на местности, исходя из необходимости отбора проб почвы на ненарушенных, естественных участках исследуемой территории. Географическая привязка осуществлялась с помощью GPS-приемника. Поверхность места предполагаемого отбора пробы почвы размечалась в виде квадрата со стороной 1 м. В вершинах, серединах сторон и в центре размеченного квадрата пробоотборником из нержавеющей стали диаметром 0.05 м отбирались от 5 до 9 кернов почвы на глубину 0.05 м. Отобранные керны объединялись в одну пробу и запаковывались в двойные полиэтиленовые пакеты для пищевых продуктов. На внутреннем пакете маркером наносился идентификатор пробы.

Подготовка проб почвы для количественного химического анализа проводилась в соответ-

ствии с действующими на данный момент нормативными документами и включала следующие стадии: сушку до воздушно-сухого состояния; просеивание на сите 1.0 мм для определения подвижных форм; измельчение до 0.074 мм для определения валовых форм. Количественный химический анализ подготовленных проб проводился в соответствии с действующими на момент проведения анализа нормативными документами в аккредитованной аналитической лаборатории ИПЭ УрО РАН.

Для выявления зависимости пространственного распределения концентрации Cr в поверхностном слое почвы от направления были построены вариограммные поверхности. Вариограммная поверхность — инструмент, который дает представление о поведении пространственной структуры в целом, представляет собой поверхность значений, вычисленных на регулярной сетке в пространстве лагов по формуле вариограммы, что позволяет сразу увидеть анизотропию и определить приоритетные направления для построения вариограмм. Следует отметить, что вариограммная поверхность обладает точечной симметрией относительно точки (0, 0).

Чтобы оценить содержание элемента и предсказать его распределение на исследуемом участке, авторы использовали две конкурирующие методики, основанные на гибридизации нейрон-

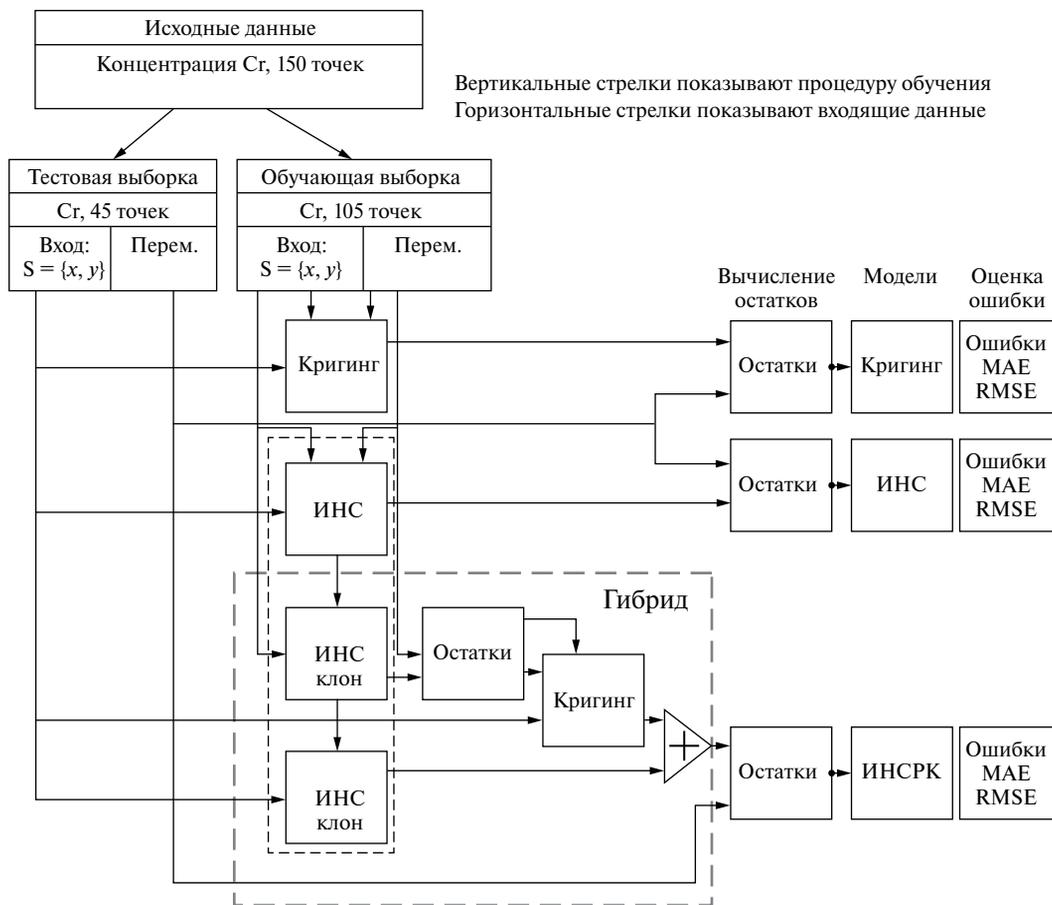


Рис. 2. Блок-схема гибридного подхода.

ной сети и геостатистических подходов: обобщенные регрессионные нейронные сети с кригингом остатков (GRNNRK) и многослойный персептрон с кригингом остатков (MLPRK). Схема гибридного подхода показана на рис. 2.

Все образцы (150 значений) случайным образом были разбиты на обучающий и тестовый наборы данных. Набор обучающих данных (105 значений) применялся для построения кригинга, обучения сетей и интерполяции распределения элементов на поверхности. Набор тестовых данных (45 значений) использовался только один раз, как набор данных для проверки каждой из моделей.

В качестве первого типа ИНС была выбрана MLP с алгоритмом обучения Левенберга-Маркварта. Структура сети определялась методом компьютерного моделирования. Выбор количества нейронов в скрытом слое осуществлялся посредством поиска минимальной среднеквадратичной ошибки (RMSE) для предсказания содержания элемента (Cr). Количество нейронов варьировалось от 2 до 25. Каждая сеть обучалась 500 раз, и выбиралась лучшая из них. Качество предсказания сети проверя-

лось коэффициентом корреляции и RMSE между результатами прогнозов сети и набором обучающих данных.

В качестве второй ИНС выбрана GRNN. Сеть GRNN имеет 105 входных нейронов в соответствии с 105 точками набора данных обучения. Основным параметром, характеризующим аппроксимационные свойства и точность интерполяции для сети GRNN, – параметр SPREAD (разброс значений параметра). SPREAD изменялся с 0 до 0.3 с шагом 0.01, всего моделирование проведено 300 раз. Построена зависимость между параметром SPREAD и стандартной ошибкой RMSE в преобразованном пространстве по направлению. После определения оптимального направления, для которого SPREAD и RMSE имеют минимальное значение, конечная сеть использовала параметр SPREAD с минимальной ошибкой. Преобразование пространства, в котором учитывается анизотропия в данных, определялось как:

$$\dot{h} = \begin{bmatrix} 1 & 0 \\ 0 & a_{min}/a_{max} \end{bmatrix} \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} h,$$

где  $h$  – вектор проекций лага на оси координат;  $a_{min}$ ,  $a_{max}$  – длины малой и большой оси эллипса;  $\theta$  – угол, определяющий направление большой оси эллипса. Угол измеряется в направлении запад–восток против часовой стрелки.

Для проверки методов, предложенных в исследовании, проводилось сравнение со стохастическим методом интерполяции Универсальным кригингом (Universal Kriging -UK).

Невязки (остатки) нейронной сети были определены следующим образом:

$$u(x_i) = Z(x_i) - Z_{ANN}(x_i),$$

где  $u(x_i)$  – остатки набора данных  $x_i$ ,  $Z(x_i)$  – измеренные значения,  $Z_{ANN}(x_i)$  – значения, оцененные нейронной сетью. Полученные остатки оценивались с использованием ординарного кригинга:

$$u_{OK}(x) = \sum \lambda_i r(x_i),$$

где  $u_{OK}(x)$  – оценочное значение в точке  $x$  с использованием ОК,  $\lambda_i(x)$  – оптимальные веса с условием  $\sum \lambda_i = 1$ , а  $u(x_i)$  – остатки нейронной сети для точки  $x_i$ . Окончательная оценка содержания элемента  $Y(x_i)$  получалась как сумма оценки нейронной сети и оценки остатков методом кригинга:

$$Y(x_i) = Z_{ANN}(x_i) + u_{OK}(x_i).$$

Прогностическая точность каждого выбранного подхода проверялась коэффициентом ранговой корреляции Спирмена ( $r$ ), средней абсолютной ошибкой MAE и RMSE между предсказанными и исходными данными из тестового набора:

$$MAE = \frac{\sum_{i=1}^n |x_{modi} - x_i|}{n},$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_{modi} - x_i)^2}{n}},$$

где  $x_{modi}$  – предсказанная концентрация (ANNs, кригинг),  $x_i$  – измеренная концентрация,  $n$  – количество точек.

**Таблица 1.** Описательная статистика по содержанию в пробах почвы Сг

Элемент	Минимальное значение, мг/кг	Максимальное значение, мг/кг	Среднее значение, мг/кг	Стандартное отклонение, мг/кг	CV	Асимметрия	Эксцесс	Медиана, мг/кг
Сг	25.8	1265	245	256	1.04	1.41	4.61	89.5

## РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Описательная статистика показана в табл. 1. Из табл. 1 видно, что распределение концентраций Сг имеет правый перекосяк и островершинный тип распределения. Средняя концентрация общего Сг в пятнах аномалии примерно в 10 раз выше остальных. Общая концентрация Сг за пределами аномалии не выше значений фоновых концентраций в Уральском регионе (Уральский кларк) и в мировых почвах (World Clarke), тогда как на участках аномалий превышает в 2.5 раза концентрации в Уральском кларке [1]. Образцы с аномальными концентрациями Сг образовывали произвольные пятна на исследуемом участке. Концентрации Сг в точках отбора проб составляют от 25.8 до 1265 мг/кг при среднем значении 245 мг/кг и стандартном отклонении в 256 мг/кг. Из-за асимметрии распределения значение медианы (89.5 мг/кг) лучше характеризует среднее значение содержания Сг в исследуемой области, чем среднее арифметическое (245 мг/кг).

Результат теста Колмогорова-Смирнова показывает, что концентрация Сг не имеет нормального распределения. Логарифмирование значений концентрации не привело к нормализации, распределение стало бимодальным. Чтобы свести к минимуму влияние экстремальных выбросов для оценок методом универсального кригинга, использовались преобразованные значения.

Для выявления анизотропии в данных, были построены вариограммы в шести направлениях (0°, 30°, 60°, 90°, 120° и 150°) (рис. 3а). На поверхности вариограммы видна анизотропия в направлении северо-запад – юго-восток (рис. 3б).

Окончательная конфигурация выбранной сети MLP составляет 2–5–1, таким образом, скрытый слой содержит 5 нейронов (рис. 4).

Очевидная зависимость коэффициента корреляции и среднеквадратической ошибки от параметра SPREAD для разных направлений также показывает присутствие анизотропии в данных (рис. 5).

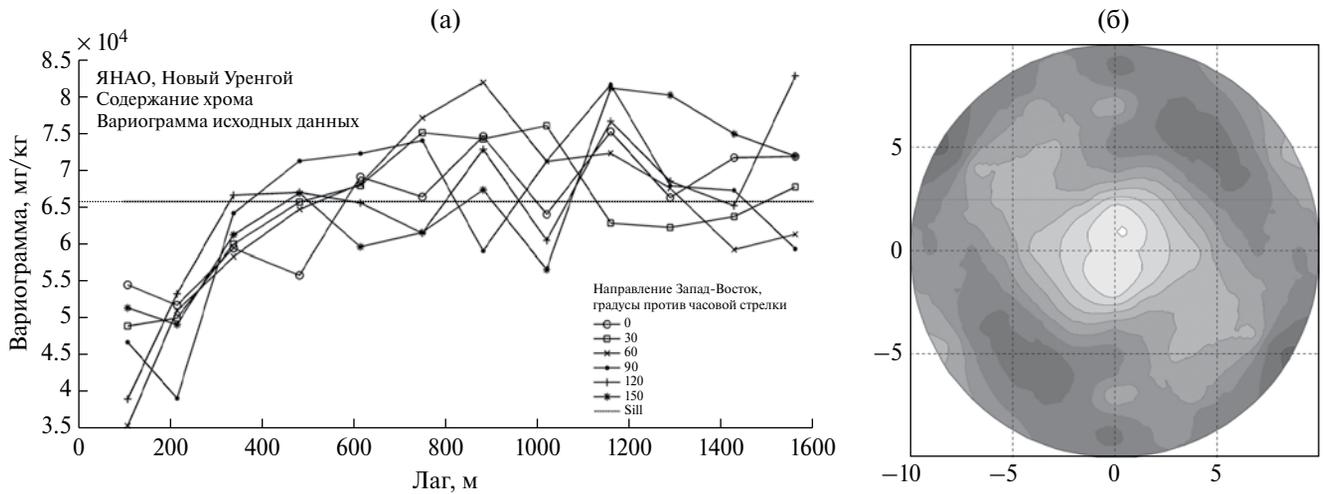


Рис. 3. Вариограммы в шести направлениях (а); вариограммная поверхность концентраций Cr (б) (угол от направления на восток, против часовой стрелки).

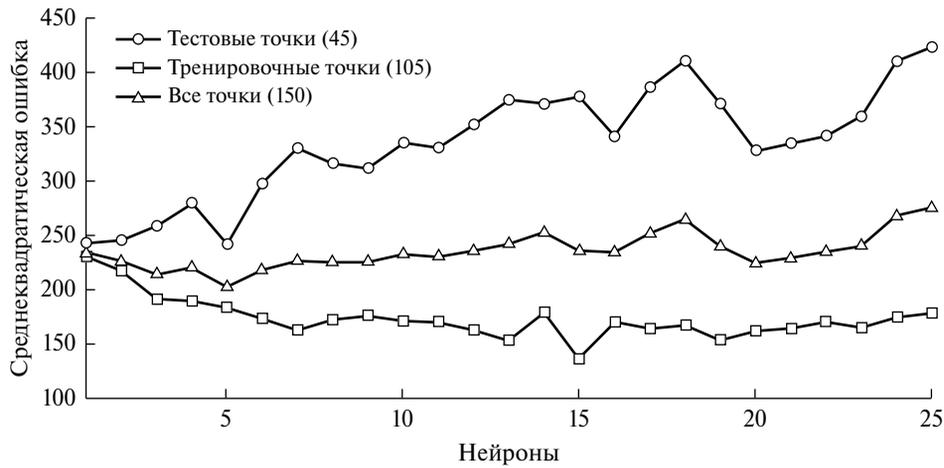


Рис. 4. Выбор структуры MLP на основе минимизации RMSE: среднеквадратическая ошибка (RMSE) нейронной сети для тестирования, обучения и общих данных для разного количества нейронов в скрытом слое для Cr.

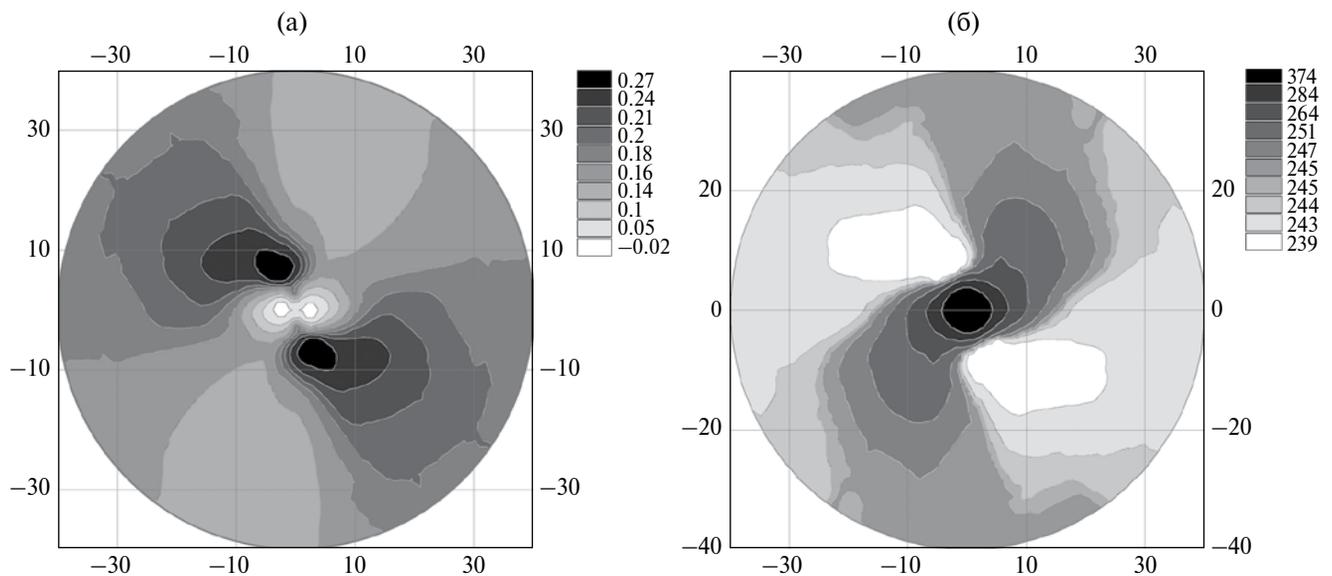


Рис. 5. Диаграмма зависимости коэффициента корреляции (а) и RMSE (б) от параметра SPREAD для разных направлений (от 0° до 180° с шагом 5°) (угол от направления на восток, против часовой стрелки).

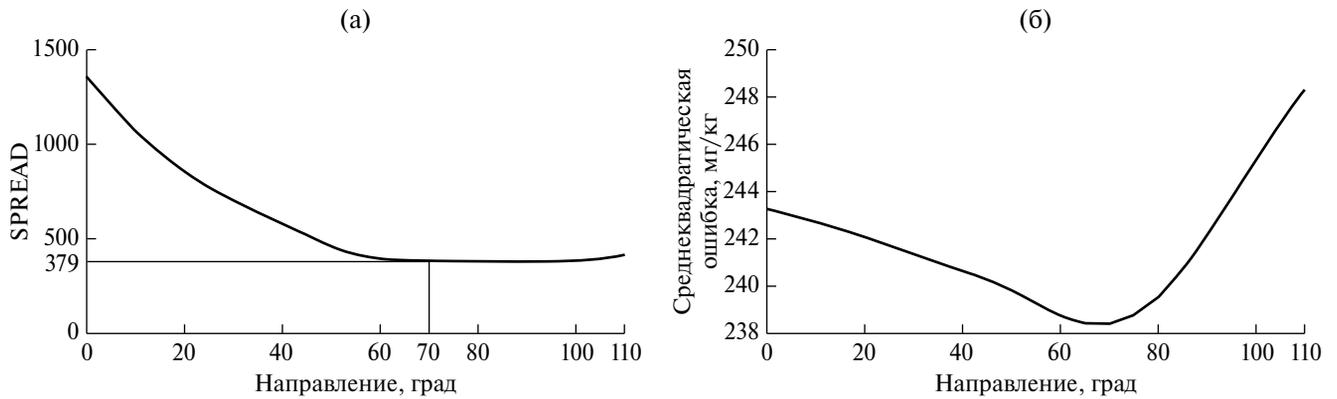


Рис. 6. Зависимость от направления анизотропного GRNN: а) SPREAD; б) RMSE (угол от направления на восток, против часовой стрелки).

Было определено оптимальное направление, для которого параметры SPREAD и RMSE имели минимальное значение (примерно 70°) (рис. 6). Полученные результаты использовались для построения оптимальной сети GRNN.

Сравнение методов показало превосходство моделей MLP и MLPRK в точности моделирования. Как видно из табл. 2, подход GRNNRK имел меньшую RMSE, чем модель GRNN (улучшение на 4.2%) и кригинг (улучшение на 8.6%). Нейрон-

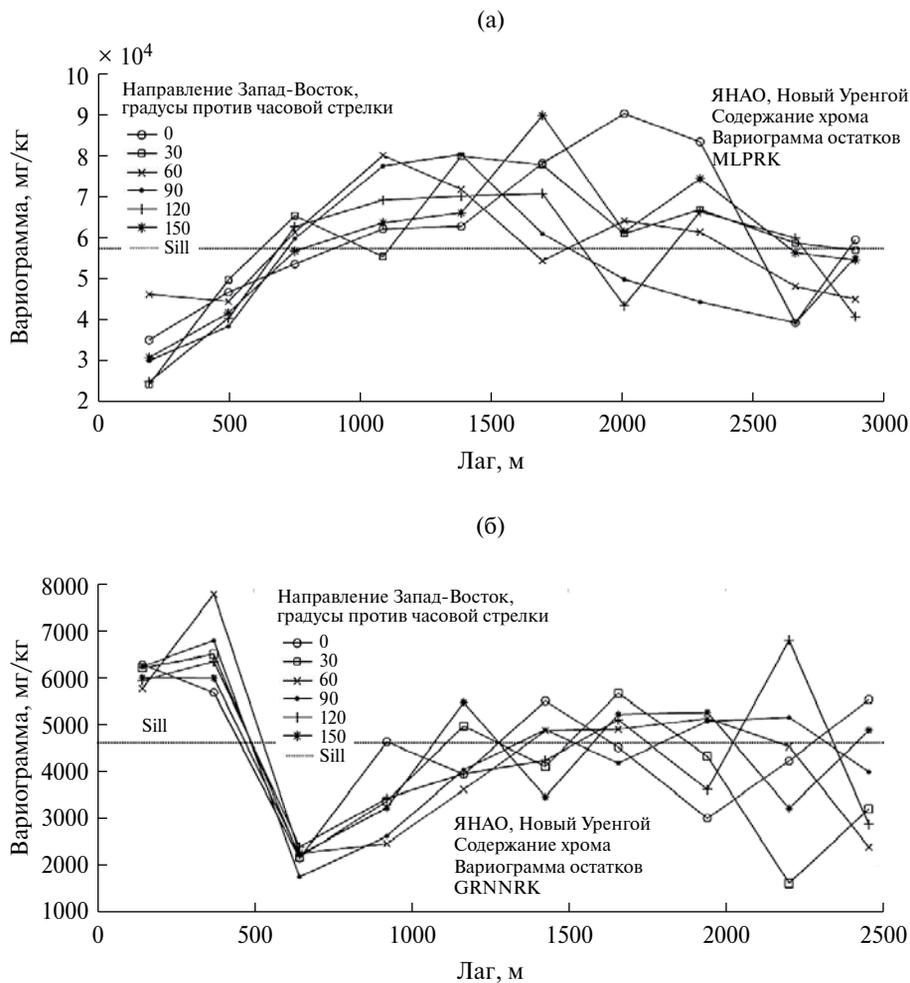


Рис. 7. Полувариограммы для остатков по различным направлениям для MLPRK (а) и GRNNRK (б) (угол от направления на восток, против часовой стрелки).

Таблица 2. Индексы оценки точности моделей

Модель	Коэффициент корреляции между наблюдаемыми и предсказанными значениями, $r$	p-level	MAE, мг/кг	RMSE, мг/кг
UK	0.12	$p < 0.05$	185.7	249.1
MLP	0.28	$p < 0.05$	186.2	242.0
GRNN(анизотропная)	0.26	$p < 0.05$	172.0	237.9
MLPRK	0.28	$p < 0.05$	<b>167.5</b>	<b>210.6</b>
GRNNRK(анизотропная)	0.40	$p < 0.05$	169.3	227.8

ные сети в целом оказались точнее, чем кригинг: GRNN и MLP имели меньшую RMSE (на 4.5% и 2.9%, соответственно).

В то же время установлено, что использование гибридного подхода MLPRK дает повышение точности прогнозирования (на основе RMSE) примерно на 13.0% относительно MLP и 15.5%

относительно кригинга, что соответствует результатам [4].

Изучение остатков подтверждает важность вариографии для анализа и моделирования пространственных данных с использованием алгоритмов обучения нейронной сети (MLP и GRNN). Хорошо обученные нейронные сети моделируют

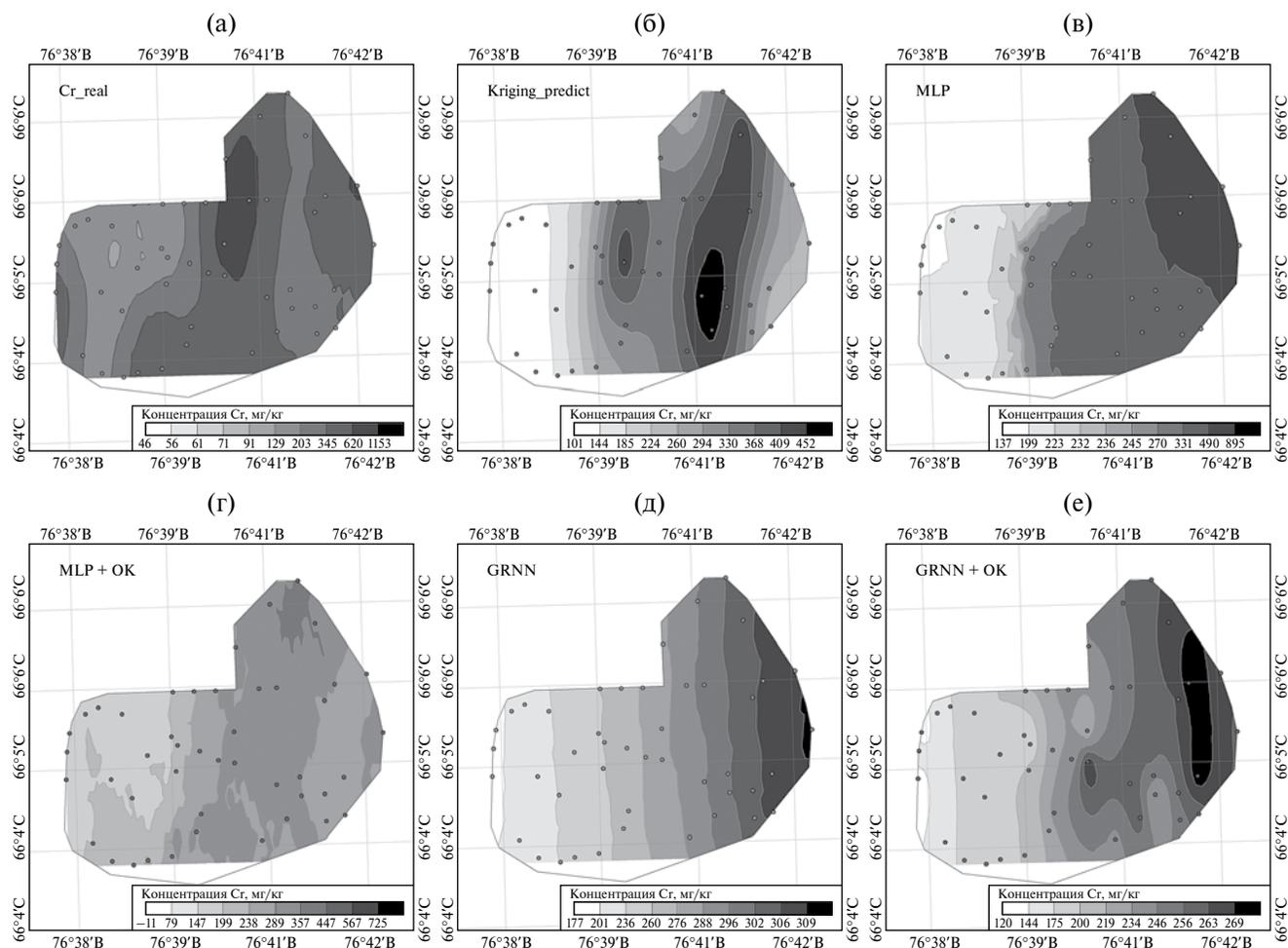


Рис. 8. Карты реального распределения концентраций хрома (а) и построенные с помощью моделей GRNN, MLP и комбинированных методов (б-е).

всю структурированную информацию. Таким образом, остатки не должны пространственно коррелироваться. Направленные вариограммы для остатков демонстрируют чистый эффект самородка — отсутствие пространственной корреляции (рис. 7).

Одним из возможных объяснений низкой точности универсального кригинга может быть сильная локальная вариация поверхностной концентрации Cr. Пространственные аномалии распределения концентрации Cr могут зависеть и от внутренних факторов (таких как тип и структура почвы) и от внешних факторов (антропогенное воздействие, климат и т.д.).

Карты реального распределения концентраций хрома и результаты предсказаний моделей показаны на рис. 8.

### ЗАКЛЮЧЕНИЕ

Целью этого исследования являлось сравнение моделей, основанных на ИНС и гибридного подхода, объединяющего преимущества ИНС и геостатистики (кригинга) в сильно гетерогенной среде на примере прогнозирования содержания аномально распределенного химического элемента (Cr) в почве в г. Новый Уренгой. Точность моделей проверялась путем их сравнительной оценки с другими моделями (кригинг, MLP, GRNN). Коэффициент корреляции ( $r$ ), средняя абсолютная погрешность (MAE) и среднеквадратичная ошибка (RMSE) применялись в качестве индикаторов предсказательной точности моделей.

Результаты показали, что модели на основе MLP более точны, чем модели, основанные на кригинге и GRNN. Применение гибридного подхода уменьшает ошибки предсказания ИНС, что повышает точность моделирования. По сравнению с другими методами наибольшее улучшение ошибки RMSE (15.5%) наблюдается в модели MLPRK.

Полученные результаты подтверждают возможности гибридных методов, которые могут быть использованы для повышения точности моделирования пространственного распределения концентраций химических элементов в верхнем слое почвы городских районов, которые характеризуются высокой гетерогенностью.

*Работа выполнена при поддержке проекта УрО РАН 15-11-2345-27.*

### СПИСОК ЛИТЕРАТУРЫ

1. Саев Ю.Е., Ревич Б.А., Янин Е.П. и др. Геохимия окружающей среды. М.: Недра, 1990. 335 с.
2. Anagu I., Ingwersen J., Utermann J., Streck T. Estimation of heavy metal sorption in German soils using artificial neural networks // *Geoderma*. 2009. 152. P. 104–112.
3. Chukanov V.N., Sergeev A.P., Ovchinnikov S.M., Medvedev A.N. Diagnostics of snow-cover contamination with soluble and insoluble metal impurities // *Russian Journal of Nondestructive Testing*. 2006. N42. P. 630–636.
4. Dai F., Zhou O., Lva Z., Wang X., Liu G. Spatial prediction of soil organic matter content integrating artificial neural network and ordinary kriging in Tibetan Plateau // *Ecological Indicators*. 2014. N45. P. 184–194.
5. Demyanov V., Kanevsky M., Chernov S., Savelieva E., Timonin V. Neural Network Residual Kriging Application for Climatic Data // *Journal of Geographic Information and Decision Analysis*. 1998. N2. P. 215–232.
6. Falamaki A. Artificial neural network application for predicting soil distribution coefficient of nickel // *Journal of Environmental Radioactivity*. 2013. N115. P. 6–12.
7. Goovaerts P. Geostatistics in soil science: State of the art and perspectives // *Geoderma*. 1999. N89. P. 1–45.
8. Guo G.H., Wu F., Xie F., & Zhang R. Spatial distribution and pollution assessment of heavy metals in urban soils from southwest China // *Journal of Environmental Sciences*. 2012. V. 24. Is. 3. P. 410–418.
9. Kanevski M., Arutyunyan R., Bolshov L., Demyanov V., Maignan M. Artificial neural networks and spatial estimations of Chernobyl fallout // *Geoinformatics*. 1995. 7(1–2). P. 5–11.
10. Kanevski M. Spatial Predictions of Soil Contamination Using General Regression Neural Networks // *International Journal of Systems Research and Information Systems*. 1999. V. 8. Is. 4. P. 241–256.
11. Kanevski M., Pozdnoukhov A., Timonin V. Machine Learning for Spatial Environmental Data. Theory. Applications and Software. EPFL Press. 2009. 380 p.
12. Koike K., Matsuda S., Suzuki T., Ohmi M. Neural Network-Based Estimation of Principal Metal Contents in the Ho-kuroku District, Northern Japan, for Exploring Ku-roko-Type Deposits // *Natural Resources Research*. 2002. N11. (2). P. 135–156.
13. Lakes T., Müller D., Krüger C. Cropland change in southern Romania: A comparison of logistic regressions and artificial neural networks // *Landscape Ecology*. 2009. 24(9). P. 1195–1206.
14. Li Y., Li C., Tao J.J., Wang L.D. Study on Spatial Distribution of Soil Heavy Metals in Huizhou City Based on BP-ANN Modeling and GIS // *Procedia Environmental Sciences*. 2011. 10. P. 1953–1960.

15. Liu F., He X., Zhou L. Application of generalized regression neural network residual kriging for terrain surface interpolation // Proc. SPIE 7492, International Symposium on Spatial Analysis, Spatial-Temporal Data Modeling, and Data Mining, 2009. 74925F.
16. Mohanty K., Majumdar T.J. Using artificial neural networks for synthetic surface fitting and the classification of remotely sensed data // International Journal of Applied Earth Observation and Geoinformation, 1999. N1(1). P. 78–84.
17. Samanta B., Ganguli R., Bandopadhyay S. Transactions of the Institution of Mining and Metallurgy. 2005. 114. P. 129–139.
18. Sergeev A.P., Baglaeva E.M., Shichkin A.V. Case of soil surface chromium anomaly of a northern urban territory – preliminary results // Atmospheric Pollution Research. 2010. V. 1. P. 44–49.
19. Sergeev A.P., Baglaeva E.M., Antonov K.L., Medvedev A.N., Rakhmatova A.Y. Anomalies of chromium surface distribution in urban soils from subarctic region of Russia // 15th International multidisciplinary scientific geoconference SGEM 2015. Water Resources. Forest, Marine and Ocean Ecosystems. Conference proceedings, V. II Soils, Forest Ecosystems, Marine and Ocean Ecosystems. 18–24 June, 2015. Bulgaria. P. 27–34.
20. Shaker R., Tofan L., Bucur M., Costache S., Sava D., Ehlinger T. Land cover and landscape as predictors of groundwater contamination: a neural-network modelling approach applied to Dobrogea, Romania // Journal of Environmental Protection and Ecology. 2010. N11(1), pp. 337–348.
21. Shaker R.R., Ehlinger T.J. Exploring non-linear relationships between landscape and aquatic ecological condition in southern Wisconsin: A GWR and ANN approach // International Journal of Applied Geospatial Research. 2014. N5(4). P. 1–20.
5. Demyanov, V., Kanevsky, M., Chernov, S., Savelieva, E., Timonin, V. Neural Network Residual Kriging Application for Climatic Data. *Journal of Geographic Information and Decision Analysis*, 1998, no. 2, pp. 215–232.
6. Falamaki, A. Artificial neural network application for predicting soil distribution coefficient of nickel. *Journal of Environmental Radioactivity*, 2013, no. 115, pp. 6–12.
7. Goovaerts, P. Geostatistics in soil science: state of the art and perspectives. *Geoderma*, 1999, no. 89, pp.1–45.
8. Guo, G.H., Wu, F., Xie, F., & Zhang, R. Spatial distribution and pollution assessment of heavy metals in urban soils from southwest China. *Journal of Environmental Sciences*, 2012, vol. 24, issue 3, pp. 410–418.
9. Kanevski, M., Arutyunyan, R., Bolshov, L., Demyanov, V., Maignan, M. Artificial neural networks and spatial estimations of Chernobyl fallout. *Geoinformatics*, 1995, no. 7(1–2), pp. 5–11.
10. Kanevski, M. Spatial predictions of soil contamination using general regression neural networks. *International Journal of Systems Research and Information Systems*, 1999, vol. 8, issue 4. pp. 241–256.
11. Kanevski, M., Pozdnoukhov, A., Timonin, V. Machine Learning for Spatial Environmental Data. Theory. Applications and Software, EPFL Press, 2009.
12. Koike, K., Matsuda, S., Suzuki, T., Ohmi, M. Neural network-based estimation of principal metal contents in the Ho-kuroku district, Northern Japan, for exploring Kuroko-type Deposits. *Natural Resources Research*, 2002, no. 11 (2), pp. 135–156.
13. Lakes, T., Müller, D., Krüger, C. Cropland change in southern Romania: A comparison of logistic regressions and artificial neural networks. *Landscape Ecology*, 2009, 24(9), pp. 1195–1206.
14. Li, Y., Li, C., Tao, J.J., Wang, L.D. Study on Spatial Distribution of Soil Heavy Metals in Huizhou City Based on BP-ANN Modeling and GIS. *Procedia Environmental Sciences*, 2011, no. 10, p. 1953–1960.
15. Liu, F., He, X., Zhou, L. Application of generalized regression neural network residual kriging for terrain surface interpolation. Proc. SPIE7492, International Symposium on Spatial Analysis, Spatial-Temporal Data Modeling, and Data Mining, 2009, 74925F.
16. Mohanty, K., Majumdar, T.J. Using artificial neural networks for synthetic surface fitting and the classification of remotely sensed data. *International Journal of Applied Earth Observation and Geoinformation*, 1999, no. 1(1), pp. 78–84.
17. Samanta, B., Ganguli, R., Bandopadhyay, S. Transactions of the Institution of Mining and Metallurgy, 2005, no. 114, pp. 129–139.
18. Sergeev, A.P., Baglaeva, E.M., Shichkin, A.V. Case of soil surface chromium anomaly of a northern urban territory – preliminary results. *Atmospheric Pollution Research*, 2010, vol. 1, pp. 44–49.

## REFERENCES

1. Saet, Yu.E., Revich, B.A., Yanin, E.P. *Geokhimiya okruzhayushchei sredy* [Environment geochemistry], Moscow, Nedra Publ., 1990, pp. 84–108 (in Russian).
2. Anagu, I., Ingwersen, J., Utermann, J., Streck, T. Estimation of heavy metal sorption in German soils using artificial neural networks. *Geoderma*, 2009, 152, pp.104–112.
3. Chukanov, V.N., Sergeev, A.P., Ovchinnikov, S.M., Medvedev, A.N. Diagnostics of snow-cover contamination with soluble and insoluble metal impurities. *Russian Journal of Nondestructive Testing*, 2006, no. 42, pp. 630–636.
4. Dai, F., Zhou, O., Lva, Z., Wang, X., Liu, G. Spatial prediction of soil organic matter content integrating artificial neural network and ordinary kriging in Tibetan Plateau. *Ecological indicators*, 2014, no. 45, pp. 184–194.

19. Sergeev, A.P., Baglaeva, E.M., Antonov, K.L., Medvedev, A.N., Rakhmatova A. Yu. Anomalies of chromium surface distribution in urban soils from subarctic region of Russia. 15th International multidisciplinary scientific geoconference SGEM 2015. Water Resources, Forest, Marine and Ocean Ecosystems. Conference proceedings, V. II Soils, Forest Ecosystems, Marine and Ocean Ecosystems. 18–24 June, 2015, Bulgaria, pp. 27–34.
20. Shaker, R., Tofan, L., Bucur, M., Costache, S., Sava, D., Ehlinger T. Land cover and landscape as predictors of groundwater contamination: a neural-network modeling approach applied to Dobrogea, Romania. *Journal of Environmental Protection and Ecology*, 2010, no. 11(1), pp. 337–348.
21. Shaker, R.R., Ehlinger, T.J. Exploring non-linear relationships between landscape and aquatic ecological condition in southern Wisconsin: A GWR and ANN approach. *International Journal of Applied Geospatial Research*, 2014, no. 5(4), pp. 1–20.

## FORECASTING THE CONTENT OF ABNORMALLY DISTRIBUTED CHROME IN SOIL BY HYBRID MODELS BASED ON ARTIFICIAL NEURAL NETWORKS

**A.V. Shichkin, A.G. Buevich, A.P. Sergeev, E.M. Baglaeva, I.E. Subbotina**

*Institute of Industrial Ecology Ural Branch of RAS, Ekaterinburg, Russia,  
ul. S. Kovalevskoi 20, Yekaterinburg, 620219 Russia*

The work is devoted to the use of a hybrid model combining artificial neural networks (ANNs) and kriging to predict an anomalously distributed chromium (Cr). It is known that the combination of geostatistical interpolation approaches (kriging) and neural networks in the model leads to a better accuracy in forecasting and performance. Generalized regression neural networks (GRNN) and multi-layer perceptron (MLP) are classes of neural networks widely used for modeling in environmental studies. In this paper, we have compared two neural networks, i.e., GRNN and MLP, as well as two combined methods: GRNN with residual scoring (GRNNRK) and MLP with residual killing (MLPRK). The study is based on the actual data sets on surface contamination of soil in Novy Urengoy, Russia, with chromium obtained from earlier performed screening. The network structures were chosen during the computer modeling based on the minimization of RMSE. MLP and MLPRK showed the best prognostic accuracy in comparison with kriging, GRNN and GRNNRK.

**Key words:** *hybrid models, artificial neural networks, chrome, residual kriging.*