

СТАТИСТИЧЕСКИЕ ФУНКЦИИ РАСПРЕДЕЛЕНИЯ ГИДРОХИМИЧЕСКИХ ПОКАЗАТЕЛЕЙ КАЧЕСТВА ВОДЫ ПОВЕРХНОСТНЫХ ВОДНЫХ ОБЪЕКТОВ

© 2012 г. А.П. Лепихин^{1,2}, А.А. Возняк²

¹ Горный институт Уральского отделения Российской академии наук, г. Пермь

² Камский филиал ФГУП «Российский научно-исследовательский институт комплексного использования и охраны водных ресурсов», г. Пермь

Ключевые слова: гидрохимические показатели, качество воды, статистики гидрохимических показателей, функции распределения гидрохимических показателей.

Рассмотрен механизм формирования функций распределения химических показателей качества воды. Показано, что изменчивость распределения показателей качества воды характеризуется очень существенной асимметрией и не может быть аппроксимирована нормальным распределением, за исключением растворенного кислорода, гидрокарбонатов и кремния. Предлагается схема построения расчета



А.П. Лепихин



А.А. Возняк

точечных показателей качества воды на основе материалов регулярного гидрохимического мониторинга качества воды водных объектов.

Общие положения

В простейшем случае изменение содержания рассматриваемых ингредиентов в какой-либо заданной точке водного объекта представляет собой одномерный случайный процесс, который на определенном временном интервале может рассматриваться как стационарный относительно первого и второго моментов. Случайный процесс достаточно полно, как известно, описывается совокупностью функций распределения и автокорреляции. Положенная в основу действующего

мониторинга поверхностных водных объектов дискретная схема наблюдений с шагом $T_d \gg T_i$ (где T_i – характерное время изменения i -го ингредиента под воздействием естественных или антропогенных факторов) делает практически невозможным анализ автокорреляционной функции. Соответственно, продолжительности непрерывного «стояния» определенных концентраций нормируемых поллютантов в действующих российских нормативно-методических документах, связанных с регламентацией антропогенных воздействий на водные объекты [1, 2], в отличие от ряда зарубежных стандартов, в явном виде не учитываются.

Поэтому в данном приближении характеристики изменения содержания химических веществ в воде водных объектов могут быть описаны достаточно полно на уровне одномерных функций распределения.

Использование статистических функций распределения гидрохимических показателей качества воды требуется при решении следующих задач:

- оценка корректности и обоснованности использования методов параметрического статистического анализа при анализе гидрохимической информации;

- построение эффективных, устойчивых показателей «характерных» значений содержания лимитирующих показателей качества воды как естественных, так и искусственных водных объектов;

- установление эффективных «статистических» фильтров отбраковки «испорченных», некорректных результатов анализов проб воды;

- оценка (расчет) экстремальных значений заданной обеспеченности лимитирующих показателей качества воды.

В настоящее время традиционные схемы статистического анализа гидрохимических показателей качества воды основываются, как правило, на параметрических методах статистики.

Данная ситуация приводит к тому, что несмотря на то, что статистические колебания химических показателей качества воды характеризуются существенной асимметрией, традиционные статистические оценки строятся на основе среднеарифметических значений. При этом они часто оказываются неэффективными из-за неустойчивости определяемых на их основе характеристик.

В настоящее время достаточно широко обсуждается вопрос о том, что функции распределения химических показателей качества воды не подчиняются нормальному закону распределения [3–5].

Поэтому при анализе эмпирических функций распределения, в первую очередь, проверялась их согласуемость с нормальным распределением. При этом в качестве показателей «нормальности» распределения использовались следующие критерии:

- Колмогорова–Смирнова;
- асимметрии;
- Шапиро–Уилка.

Основные статистические расчеты были проведены на основе использования лицензионного программного продукта Statistica 6, а имитационное статистическое моделирование – с использованием программного продукта Mathcad 14, которые позволяют достаточно просто и эффективно выполнить требуемые расчеты.

Анализ эмпирических статистик химических показателей качества воды на реках Камского бассейна

В качестве исходной информации использовались материалы сетевых ежемесячных наблюдений 1974–2000 гг., полученные территориальным подразделением Росгидромета в бассейне Верхней Камы.

Всего рассматривалось свыше 68 створов с общим количеством контролируемых показателей качества воды не менее 23 для каждого створа.

В связи с тем, что в работе использовались различные массивы информации за достаточно длительный промежуток времени, на этапе статистической обработки оценивалась однородность гидрохимических рядов с использованием критериев Стьюдента и Манн–Уитли. Дальнейший анализ проводился по однородным выборкам.

«Нормальность» распределения очень существенно влияет на характер статистической обработки рассматриваемых рядов наблюдений. Поэтому в настоящее время предложено достаточно большое количество различных критериев проверки нормальности распределения, использующих различные свойства данного распределения [6].

В данной работе использованы три критерия оценки «нормальности» распределения рассматриваемых выборок: отношение $\frac{C_s}{\sigma_{C_s}}$, Колмогорова–Смирнова, а также критерий Шапиро–Уилка [6], (где C_s – коэффициент асимметрии рассматриваемой выборки, σ_{C_s} – среднеквадратическое отклонение его оценки). Так как для нормального распределения $C_s=0$, а ошибку вычисления выборочных C_s в

первом приближении можно принять как имеющую нормальное распределение, то если $\frac{C_s}{\sigma_{C_s}} \geq 2$, то с уровнем значимости 0,05 гипотеза о нормальном распределении рассматриваемой выборки может быть отвергнута.

Следует особо отметить критерий Шапиро–Уилка, основанный на отношении оптимальной линейной несмещенной дисперсии к ее обычной оценке методом максимального правдоподобия. По оценке [6] данный критерий является одним из наиболее эффективных критериев проверки нормальности распределения случайных величин. Нулевая гипотеза нормальности распределения случайных величин отклоняется, если величины критериев $W < W_{крN}(\alpha)$.

Как следует из [6], при $N > 50$, $\alpha = 0,05$, $W_{кр} = 0,95$ можно принять, что при $W \geq 0,95$ ряд с уровнем значимости 0,05 может считаться имеющим нормальное распределение.

Выполненный анализ эмпирических выборок химических показателей качества воды для рассматриваемых водных объектов позволил оценить частоту (вероятность) приемлемости использования нормального распределения для описания их статистических колебаний (табл. 1).

Таблица 1. Частота (вероятность) описания отдельных химических показателей качества воды на водных объектах Камского бассейна нормальным распределением

№ п/п	Наименование веществ	Процентное соотношение, %	№ п/п	Наименование веществ	Процентное соотношение, %
1	Взвешенные вещества	50	13	Кремний	85
2	Кислород	100	14	Фосфор	30
3	Магний	47	15	Железо	45
4	Хлориды	36	16	Медь	64
5	Сульфаты	37	17	Цинк	40
6	Минерализация	68	18	Марганец	50
7	Гидрокарбонаты	72	19	Фенол	47
8	ХПК	50	20	Кальций	72
9	БПК ₅	62	21	Нефтепродукты	41
10	Азот аммонийный	43	22	СПАВ	43
11	Азот нитритов	54	23	Фосфаты	62
12	Азот нитратов	56			

Характерной особенностью распределения химических показателей качества воды на всех рассматриваемых створах является существенная асимметрия оцениваемых показателей. Низкой асимметрией и значительно высокими отклонениями от критических значений показателя Шапиро–Уилка характеризуются распределения содержания кислорода (100 %), гидрокарбонатов (72 %) и кремния (85 %). Распределение других химических показателей качества воды очень существенно отличается от нормального (см. табл. 1), их $W < 0,95$.

Схемы формирования статистических колебаний химических показателей качества воды в русловых потоках

В общем случае содержание химических веществ в водных объектах определяется совокупностью внешних, по отношению к водному объекту, факторов и внутриводоемных процессов. При этом принципиальное значение на формирование структурных колебаний химических показателей качества воды водного объекта оказывает их связь с расходом водотока. Внутриводоемные по механизмам изменения содержания лимитирующих ингредиентов разделяются на процессы:

- разбавления;
- трансформации, самоочищения;
- миграции.

При этом сами поллютанты в зависимости от их поведения в водных объектах разделяются на консервативные, неконсервативные, гетерофазнонеконсервативные.

Изменение содержания в водных объектах консервативных веществ происходит только вследствие процессов разбавления, для неконсервативных определяющими факторами, наряду с процессами разбавления, становятся процессы трансформации и самоочищения [4, 7].

Для гетерофазнонеконсервативных ингредиентов характерной особенностью является их способность активно мигрировать между отдельными звеньями системы <вода> <взвешенные наносы> <донные отложения>.

Таким образом, если изменения содержания консервативных загрязняющих веществ в водные объекты определяются только интенсивным их поступлением в водный объект и их разбавлением, при этом характер внутриводоемных процессов не зависит от характера поллютанта, то для неконсервативных и, тем более,

гетерофазнонеконсервативных веществ интенсивность внутриводоемных процессов существенно зависит от характера веществ.

В отдельном частном случае, когда роль внутриводоемных процессов в трансформации и миграции поллютантов минимальна, имеем

$$C(t) \sim \frac{q(Q(t), t)}{Q(t)}, \quad (1)$$

где $q(Q) \sim \int_0^L j(Q, x, C) dx$ – интенсивность поступления поллютанта в водный объект,

г/с;

j – удельная интенсивность поступления поллютанта, г/м·с.

Для дальнейшего анализа без существенной потери общности целесообразна степенная аппроксимация $q(Q) \sim P_{qQ} \cdot Q^{\alpha_{qQ}}$. При $\alpha_{qQ} \cong 1$ величина $C(Q)$ перестает зависеть от расхода водотока, а определяется параметром P_{qQ} , представляющим в данном случае эффективную концентрацию поступающих в водоток поллютантов. Данная ситуация характерна для консервативных поллютантов, в первую очередь, хлоридов, магния, калия, натрия.

В более общем случае, имеем $C(Q(t)) \approx \frac{P_{qQ}}{Q^{1-\alpha_{qQ}}} = \frac{P_{qQ}}{Q^\beta}$. При величине $1 - \alpha_{qQ} = \beta$

параметр α_{qQ} может быть как положительной величиной ($\alpha_{qQ} > 0$), когда интенсивность привноса поллютантов возрастает при увеличении Q , так и отрицательной, если интенсивность поступления i -го поллютанта падает при увеличении Q .

Если изучению характера функций распределения расходов воды в реках посвящены многочисленные исследования, то процессы формирования и статистические структуры $C(t)$ изучены хуже.

Если известна интегральная функция распределения $F_j(q)$ процесса $q(Q(t), t)$ распределения $P_Q(Q)$, то нетрудно определить функции распределения и $P_C(C)$, используя интеграл свертки:

$$P_C(C) = \int_0^\infty P_Q(\zeta) \cdot F_j(C \cdot \zeta) d\zeta \quad (2)$$

Если $q(t)$ и $Q(t)$ описываются Гамма-распределением, то $P(C) = \frac{C^{u-1} \cdot (1+C)^{u+v}}{B(u+v)}$ –

Бета-распределение второго рода, при $u=1$, $v=1$ вырождается в распределение Коши–Лоренца.

Если $q \sim Q$, в этом случае $C(Q) = \frac{G_0}{Q^\beta}$, где G_0 – интенсивность поступления от источника загрязнения. При этом, как известно, $P(C) = \left| \frac{d\varphi^{-1}(C)}{dx} \right| \cdot P_Q(\varphi^{-1}(C))$.

Соответственно, имеем

$$P(C) = q_0^{-\left(\frac{1}{\beta}\right)} \cdot C^{\left(\frac{1}{\beta}-1\right)} \cdot P_Q\left(\frac{C^{\frac{1}{\beta}}}{q_0^{\frac{1}{\beta}}}\right) \cdot 1/\beta \quad (3)$$

Как правило, для описания распределения расходов воды водотоков-приемников используется гамма-распределение. Характерной особенностью данного распределения является то, что оно характеризуется «тяжелым хвостом», т. е. вероятность экстремальных значений значительно повышена.

В этом случае динамика содержания рассматриваемых поллютантов в воде в водных объектах определяется совокупностью инерционных и флуктуационных внутриводоемных процессов

$$\frac{dC}{dt} = f(C, t) + \sigma(C, t) \cdot \zeta(t) \quad (4)$$

где $f(C, t)$ – функция, определяющая динамику поведения загрязняющих веществ в водном объекте, в простейшем случае описывается как $f(C, t) = K \cdot C$;

$\zeta(t)$ – составляющая флуктуационного воздействия;

$\sigma(C, t)$ – ее среднее квадратическое отклонение.

При этом обычно соблюдается условие

$$\frac{T_H}{T_\zeta} \gg 1,$$

где T_H , T_ζ – характерные времена, соответственно,

$$\text{действия инерционных факторов } T_H = \left(\frac{1}{f(C, t)} \left(\frac{\partial f(C, t)}{\partial t} \right) \right)^{-1},$$

$$\text{действия флуктуационных факторов } T_\zeta = \int_0^\infty R_{\zeta\zeta}(\tau) d\tau;$$

$R_{\zeta\zeta}(\tau)$ – автокорреляционная функция $\zeta(t)$.

В случае нормального распределения возмущающей функции для $C(t)$ допустима аппроксимация в виде диффузионного процесса, а его стационарная функция распределения будет иметь следующий вид [8]:

$$P(C) = \frac{1}{K} \cdot \exp\left(-\int_0^C \frac{f(e)}{\sigma^2(e)} \cdot de\right). \quad (5)$$

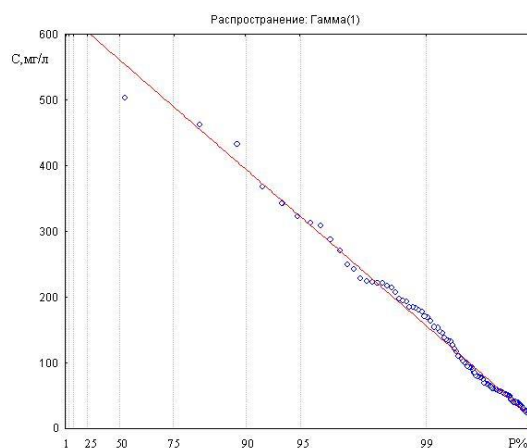
Константа K определяется из условия нормирования функции $P(C)$ как $\int_0^{\infty} P(C) = 1$.

Отсюда следует, что в простейшем случае для химических показателей качества воды при $f(e)=f_0=const$ и $\sigma(e)=\sigma_0=const$ имеет место экспоненциальное распределение, а при $f(e)=\alpha \cdot e$ и $\sigma(e)=\sigma_0=const$ – распределение Гаусса. В более общем случае при $f(e)=f_0+f_1e$ и $\sigma(e)=\sigma_0+\sigma_1e+\sigma_2e^2$ стационарное распределение вероятности $P(C)$ удовлетворяет уравнению

$$\begin{aligned} \frac{dP(C)}{dC} &= \frac{C - C_0}{b_0 + b_1C + b_2C^2} P(C); \\ C_0 &= \frac{\sigma_1 - f_0}{f_1 - 2\sigma_2}; & b_0 &= \frac{\sigma_0}{\sigma_1 - 2\sigma_2}; \\ b_1 &= \frac{\sigma_1}{\sigma_1 - 2\sigma_2}; & b_2 &= \frac{\sigma_2}{\sigma_1 - 2\sigma_2}. \end{aligned} \quad (6)$$

Решением для уравнения (6) при различных значениях C_0, b_0, b_1, b_2 является так называемое распределение Пирсона – в частном случае, когда $b_2=0$ – это двухпараметрическое гамма-распределение (рисунок а).

a



б

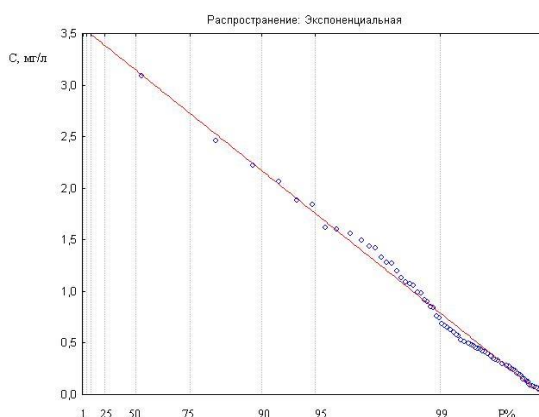


Рисунок. Интегральная функция распределения по створу р. Быстрый Танып – г. Чернушка: *a* – хлоридов; *б* – азота аммонийного.

Рассмотрим, как будет изменяться характер функции распределения $P(C)$ в зависимости от структуры функции $f(C)$ и $\sigma(C)$. В простейшем случае при $f(C) \sim k$, т. е. когда данная функция является независимой от концентрации рассматриваемых загрязнителей в воде водного объекта, как нетрудно видеть из соотношения (5), имеем место экспоненциальное распределение

$$P(C) = \frac{\sigma^2}{k} \exp\left(-\frac{k}{\sigma^2} \cdot C\right), \quad (7)$$

которое имеет фиксированное значение коэффициента вариации $C_v=1$ и асимметрии $C_s=2$ (рисунок *б*).

Если содержание ингредиентов в водных объектах описывается кинетическим уравнением первого порядка $f(C) = -\alpha(C - \bar{C})$, исходя из особенностей

условий формирования характерно, в первую очередь, только для кислорода [8], то оно описывается нормальным распределением. Как уже отмечалось, «нормальность» распределения содержания кислорода и кремния (см. табл. 1) является характерной особенностью динамики гидрохимического состава на абсолютном большинстве рассматриваемых створов наблюдений.

В целом характер функции гидрохимических показателей качества воды в зависимости от структуры функции $f(C)$ при $\sigma = const$ представлен в табл. 2.

Таблица 2. Типы распределения гидрохимических показателей качества воды в зависимости от структуры функции $f(C)$ при $\sigma(C) = const$

<i>Структура функции $f(C)$</i>	<i>Плотность статистического распределения</i>
$f(C) = const$	Показательное распределение (распределение Лапласа)
$f(C) = \alpha \cdot C$	Нормальное распределение (распределение Гаусса–Лапласа)
$f(C) \sim \alpha/C$	Степенное распределение (распределение Парето)
$f(C) \sim -\alpha \cdot \ln(C)$	Гамма-распределение
$f(C) = \exp(-\alpha \cdot C)$	Распределение экстремального значения

Построение оптимальных точечных расчетных оценок фоновых концентраций

При решении широкого круга практических задач требуются точечные оценки содержания химического показателя в воде водного объекта.

В общем виде точечную оценку рассматриваемых значений можно провести, используя схему средних значений Колмогорова:

$$x_{\kappa} = \frac{1}{N} \left[\varphi^{-1} \left(\sum_{i=1}^N \varphi(x) \right) \right].$$

В достаточно общем виде задача состоит в нахождении такой оценочной функции φ_j [5], чтобы параметр $C\varphi_j$, оцениваясь на основе соотношения

$$\sum_{i=1}^N \varphi_j(c_i - C\varphi_j) = 0 \quad (8)$$

по заданной выборке объема N , характеризовался бы минимальной дисперсией.

В простейшем случае, если рассматриваемая выборка:

– стационарна;

- статистически независима;
- описывается нормальным законом распределения,

то наиболее эффективны оценки, построенные с использованием $\varphi(x)=x$.

Для других функций распределений характер функции $\varphi(x)$, при котором будут наблюдаться оптимальные оценки средних значений, может существенно отличаться.

При этом возникает задача или нахождения оптимальных функций φ_j для характерных распределений рассматриваемых ингредиентов, или использования оценок на основе квантилей, являющихся непараметрическими оценками.

Простейшую непараметрическую оценку характерного значения рассматриваемой характеристики имеем при $\varphi_j(x)=sgn(x)$. В этом случае в качестве $C\varphi_j$ получаем медиану рассматриваемой выборки.

Численным критерием эффективности применяемого точечного показателя является дисперсия его оценки по заданной выборке объема N . Как известно, дисперсия среднеарифметического значения \bar{x} задается соотношением

$$D(\bar{x}) = \frac{\sigma^2}{N} \quad (9)$$

(при дисперсии σ^2 рассматриваемой выборки объема N).

Согласно [9] дисперсия координаты любого квантиля соответствующей вероятности p определяется следующим соотношением:

$$D[x_p] = \frac{p(1-p)}{\{N[p(x_p)]^2\}} \quad (10)$$

где $p(x_p)$ – плотность распределения в точке x_p ;

x_p – координата квантиля порядка p .

Для медианы нормального распределения, соответственно, имеем:

$$D[x_p] = \frac{\sigma^2(2 \cdot \pi)}{4 \cdot N} = \frac{\sigma^2}{N} \left(\frac{\pi}{2} \right) \quad (11)$$

Из соотношений (9) и (11) следует, что оценки, построенные на основе средних арифметических характеристик, в $\pi/2$ раз более эффективны, чем на основе медианы. Однако данное утверждение справедливо, как уже отмечалось, только для нормального закона распределения.

Относительная погрешность оценки центра распределения с использованием медианы по сравнению со среднеарифметическим, определяемая как $R = \frac{D(\bar{x})}{D(\hat{x})}$, зависит от характера распределения (табл. 3).

Таблица 3. Значения среднеквадратичных отклонений оценки среднеарифметических значений и медианы для различных функций распределения

<i>№ n/n</i>	<i>Тип распределения</i>	<i>R</i>
1	Нормальное распределение	0,64
2	Коши распределение	∞
3	Распределение Лапласа	2

Естественно, что оценка центра распределения в виде средних выборочных \bar{x} эффективна лишь для одномодальных распределений, близких к гауссовскому, тогда как оценки на основе медианы более эффективны для островершинных распределений.

По оценке [10] для плосковершинных и двухмодальных распределений более эффективна оценка в виде центра сгибов $x_c = \frac{(x_{0,25} + x_{0,75})}{2}$, а при весьма ограниченной выборке – в виде центра размахов $x_p = \frac{(x_1 + x_N)}{2}$, где x_1, x_N – первые и последние члены упорядоченной выборки.

При сопоставлении различных оценок центра распределения необходимо учитывать также их чувствительность к грубым ошибкам измерений или передачи информации. Параметрические статистики являются более чувствительными к этим погрешностям, чем непараметрические, построенные с использованием квантилей, в т. ч. медиана.

Поэтому, учитывая сложный характер статистических распределений химических показателей качества воды водных объектов, оценка их характерных значений должна проводиться не на основе параметрического показателя – среднеарифметических значений, а на основе квантилей.

Выводы

В целом, проведенный анализ эмпирических функций распределения показал:

– распределение Гаусса, как правило, описывает распределение содержания в воде растворенного кислорода и химических показателей, формирующихся, в основном, за счет равномерно распределенных факторов;

– изменчивость распределения всех других показателей качества воды характеризуется очень существенной асимметрией и не может быть аппроксимирована нормальным распределением;

– техногенез, как правило, существенно усиливает асимметрию распределения химических показателей [5];

– точечные показатели качества воды из-за существенной асимметрии распределения целесообразно строить на основе квантилей значений порядка, в том числе медианы, а не среднеарифметического значения, как это имеет место в действующих нормативно-методических документах.

СПИСОК ЛИТЕРАТУРЫ

1. Методические указания по разработке нормативов допустимых воздействий. Утв. Приказом МПР России № 328 от 12.12.2007.
2. РД 52.24.622-2001 Методические указания. Проведение расчетов фоновых концентраций химических веществ в воде водотоков.
3. Банникова О.А., Бычкова Е.Н. К вопросу об установлении региональных нормативов качества воды // Водное хозяйство России. 2011. № 6. С. 54–68.
4. Лепихин А.П., Мирошниченко С. А. Анализ функциональных зависимостей между содержанием поллютантов, обусловленных техногенными факторами, и расходами водотоков-приемников // Водохозяйственные проблемы. Вып. 2. Екатеринбург. 1997. С. 41–60.
5. Лепихин А.П., Мирошниченко С. А. Особенности задания «фоновой» концентрации в естественных водотоках // Водное хозяйство России. 2002. Т. 3, № 3. С. 247–262.
6. Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников. М.: ФИЗМАТЛИТ, 2006. 816 с.

7. *Никаноров А.М.* Гидрохимия: Учебник. 2-е изд., перераб. и доп. СПб: Гидрометеоздат, 2001. 444 с.
8. *Тихонов В.И., Миронов М.А.* Марковские процессы. М.: «Сов. радио», 1977. 488 с.
9. *Крамер Г.* Математические методы статистики. М.: Мир, 1975. 646 с.
10. *Новицкий П.В., Зограф И.А.* Оценка погрешностей результатов измерений. Л.: Энергоатомиздат, 1991. 301 с.

Сведения об авторах:

Лепихин Анатолий Павлович, д. г. н., профессор, заведующий лабораторией, Горный институт Уральского отделения Российской академии наук (УрО РАН), 614007, г. Пермь, ул. Сибирская, 78; директор, Камский филиал ФГУП РосНИИВХ, 614007, г. Пермь, ул. Народовольческая 33; e-mail: lepin49@mail.ru

Возняк Анна Анатольевна, к. г. н. старший научный сотрудник, Камский филиал ФГУП РосНИИВХ, 614007, г. Пермь, ул. Народовольческая 33, e-mail: AAVoznyak@gmail.com