

ПРИМЕНЕНИЕ СГЛАЖИВАНИЯ ДАННЫХ ПО МЕТОДУ СКОЛЬЗЯЩЕЙ СРЕДНЕЙ ПРИ МОДЕЛИРОВАНИИ ДИНАМИКИ ЗАБОЛЕВАЕМОСТИ ТУБЕРКУЛЕЗОМ

¹С.А.Лаптёнок, ²Ю.Д.Синкевич, ¹А.А.Кологривко,
²О.И.Родькин, ¹С.А.Хорева, ¹Ю.В.Кляусова

¹Белорусский национальный технический университет,
пр. Независимости, 65, 220013, г. Минск, Республика Беларусь
²«Международный государственный экологический институт им. А.Д.Сахарова»
Белорусского государственного университета,
ул. Долгобродская, 23/1, г. Минск, 220070, г. Минск, Республика Беларусь

Представлены результаты оценки эффективности сглаживания данных по методу скользящей средней при моделировании динамики заболеваемости туберкулезом населения Ошмянского района Гродненской области Республики Беларусь за период с 1960 по 2021 гг.

Ключевые слова: моделирование динамики заболеваемости; сглаживание данных; метод скользящей средней; эффективность; туберкулез.

Экспоненциальное сглаживание по методу скользящей средней. В связи с интенсивным ростом темпов производства и увеличением объемов потребления в глобальном масштабе в последние десятилетия все активнее проявляется тенденция расширения спектра биолого-экологических процессов, возникающих в результате негативного антропогенного воздействия на природные системы и комплексы. Течение такого рода процессов подконтрольно человеку в значительно меньшей степени, чем процессов, протекающих в производственной, экономической, социальной сфере и т.п. [1]. Этим обусловлена особая актуальность изучения динамики их развития в Республике Беларусь, где ситуация осложняется влиянием дополнительных факторов, самым значительным из которых является воздействие ионизирующих излучений от радиоактивных изотопов, выброшенных в результате взрыва из активной зоны реактора четвертого блока Чернобыльской АЭС.

Принятие эффективных управленческих решений по предотвращению возникновения нежелательных процессов и минимизации негативных последствий невозможно без обоснованного прогноза динамики их развития, который формируется на основе стохастической (вероятностной) модели.

Одним из методов моделирования динамических процессов является метод аппроксимации различными функциями, заключающийся в подборе функциональных зависимостей, наиболее

точно описывающих изменения изучаемых показателей с течением времени. Полученные аппроксимирующие функции экстраполируют на 1, 2 или более перспективных временных интервала (в зависимости от степени точности аппроксимации) и получают вероятностный прогноз развития изучаемого процесса. Таким образом, чем точнее подобранная функция аппроксимирует исходный временной ряд, тем более точным будет прогноз.

В ряде случаев точность аппроксимации динамики реальных процессов может быть ограничена вследствие ее нерегулярности: результаты последовательных измерений не укладываются в графики определенных функций, имеют место «нелогичные» выбросы как в сторону роста, так и в сторону уменьшения анализируемых показателей. Следовательно, подбор адекватной аппроксимационной модели усложняется. В таких случаях прибегают к предварительной обработке динамических данных, заключающейся в их сглаживании. Наиболее часто используются метод экспоненциального сглаживания, сглаживание по методу наименьших квадратов и др.

Одним из методов, которые могут применяться, в частности, для обработки динамических рядов, является метод сглаживания по скользящей средней. Не уступая в точности конкурирующим методам, данный метод позволяет значительно сократить время, необходимое для получения результата, за счет простоты вычислений. Суть метода заключается в следующем.

Пусть имеется исходный ряд значений $x_1, x_2, x_3, \dots, x_n$. Для получения результирующего ряда применяется выражение

$$x_i = \frac{x_{i+1} + x_{i-1}}{2} \quad (1)$$

Таким образом, рассчитываются значения x_2, x_3, \dots, x_{n-1} , причем при вычислении нового значения каждого следующего элемента используется новое значение предыдущего. Например, новое значение x_3 вычисляется как средняя арифметическая нового значения x_2 и исходного значения x_4 . Начальный и конечный элементы последовательности определяются уже для полностью сглаженного ряда соответственно выражением

$$x_{i-1} = \frac{2x_i + x_{i+1} - x_{i+3}}{2} \quad (2)$$

при $i = 2$ (для первого элемента) и выражением

$$x_{i+1} = \frac{2x_i + x_{i-1} - x_{i-3}}{2} \quad (3)$$

при $i = n-1$ (для последнего элемента).

Если после такой обработки тенденция развития анализируемого процесса проявляется все же недостаточно, процедуру сглаживания повторяют уже с использованием полученного, сглаженного ряда. Каждое повторение цикла сглаживания называют итерацией.

Для оценки эффективности экспоненциального сглаживания по методу скользящей средней при предварительной обработке динамических показателей в качестве объекта исследования были выбраны показатели динамики заболеваемости туберкулезом населения Ошмянского района Гродненской области Республики Беларусь с 1960 по 2021 год (табл. 1).

Очевидно, что в течение исследуемого периода (1960–2021 гг.) среднегодовые значения интенсивных показателей изменялись нерегулярно, и выявление общей тенденции значительно затруднено. Тем более затруднен и подбор аппроксимирующей функции.

В результате процедуры сглаживания исходного ряда по методу скользящей средней

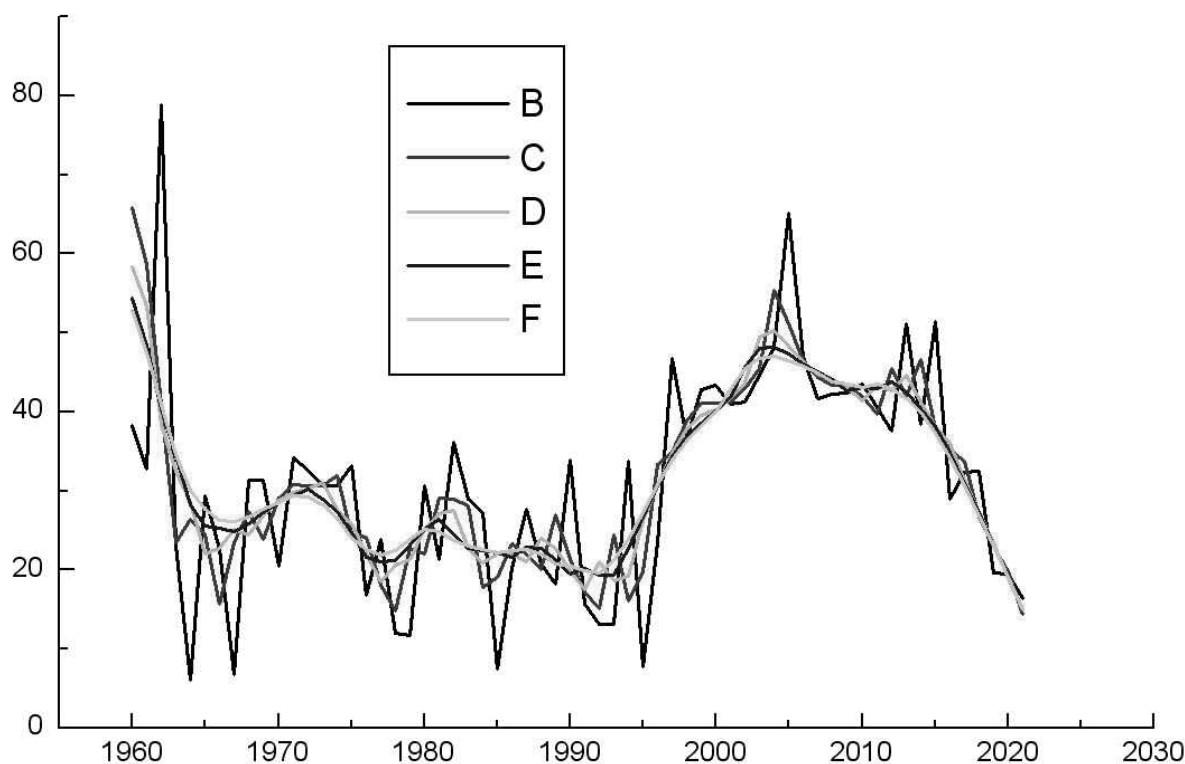


Рис. Графическое отображение интенсивных показателей заболеваемости туберкулезом населения Ошмянского района Гродненской области Республики Беларусь и результатов сглаживания методом скользящей средней
(на графике: В – динамика интенсивного показателя заболеваемости, С, D, E, F – результаты сглаживания данной динамики по первой, второй, третьей и четвертой итерациям соответственно)

**Среднегодовая динамика уровня заболеваемости населения
Ошмянского района Гродненской области Республики Беларусь**

Год	Интенсивный показатель	Первая итерация	Вторая итерация	Третья итерация	Четвертая итерация
1960	38,3	65,84	58,34	54,45	52,86
1961	32,8	58,60	53,37	48,38	47,43
1962	78,9	40,90	38,41	40,40	40,93
1963	23,2	23,45	32,42	34,43	34,59
1964	6,0	26,43	28,47	28,25	30,05
1965	29,4	24,51	22,06	25,51	27,67
1966	22,6	15,66	22,77	25,29	26,28
1967	6,8	23,48	25,08	24,90	26,07
1968	31,3	27,39	24,51	25,85	26,70
1969	31,3	23,94	26,79	27,33	27,62
1970	20,5	29,07	28,82	28,55	28,50
1971	34,2	30,84	29,77	29,38	29,37
1972	32,6	30,72	30,21	30,23	29,17
1973	30,6	30,66	31,07	28,97	28,31
1974	30,6	31,93	27,72	27,45	26,51
1975	33,2	24,36	25,92	24,72	24,05
1976	16,8	24,13	22,00	21,58	22,57
1977	23,9	18,07	18,44	21,09	21,88
1978	12,0	14,88	20,59	21,19	22,53
1979	11,7	22,74	21,30	23,19	23,83
1980	30,6	22,02	25,18	25,13	25,10
1981	21,3	29,06	27,08	26,36	24,81
1982	36,1	28,98	27,59	24,53	23,75
1983	28,9	28,09	22,69	22,69	23,07
1984	27,2	17,80	20,85	22,39	22,65
1985	7,5	19,00	22,10	22,23	22,15
1986	20,2	23,35	22,06	21,65	22,50
1987	27,7	22,02	21,08	22,85	22,64
1988	20,7	20,11	24,05	22,78	22,01
1989	18,2	27,01	22,70	21,37	20,73
1990	33,9	21,35	19,96	19,46	20,49
1991	15,7	17,23	17,56	20,24	19,95
1992	13,1	15,16	21,02	19,41	19,62
1993	13,1	24,48	18,58	19,30	21,19
1994	33,8	16,14	19,18	22,76	23,94
1995	7,8	19,77	26,23	26,68	27,32
1996	23,4	33,29	30,59	30,70	30,82
1997	46,8	34,94	34,73	34,32	33,86
1998	36,6	38,87	37,93	36,91	36,24
1999	42,8	41,14	39,50	38,62	38,19
2000	43,4	41,07	40,34	40,13	40,03
2001	41,0	41,18	41,64	41,87	42,86
2002	41,3	42,94	43,61	45,69	45,43
2003	44,7	45,57	49,50	48,00	46,81
2004	48,2	55,39	50,32	48,19	47,03
2005	65,2	51,14	48,37	47,26	46,55
2006	46,9	46,42	46,34	46,06	45,82
2007	41,7	44,31	44,85	45,10	44,95
2008	42,2	43,36	44,14	44,07	43,84
2009	42,4	43,43	43,05	42,74	43,47
2010	43,5	41,96	41,41	43,09	43,17
2011	40,5	39,78	43,43	42,88	43,47
2012	37,6	45,44	42,67	43,77	42,99
2013	51,1	41,92	44,67	42,51	41,67
2014	38,4	46,66	41,25	40,35	39,92
2015	51,4	37,83	38,18	38,17	37,34
2016	29,0	35,12	36,00	34,76	34,17
2017	32,4	33,81	31,35	30,99	30,70
2018	32,5	26,70	27,22	27,24	26,90
2019	19,6	23,10	23,49	23,09	22,97
2020	19,5	19,75	18,94	19,05	19,04
2021	16,4	14,40	15,01	15,10	15,17

Результаты статистического сравнения значений интенсивного показателя заболеваемости и первой итерации сглаживания по методу скользящей средней (+ 35 - 27 n=62)

Год	Интенсивный показатель	Первая итерация	Знак
1960	38,3	65,84	+
1961	32,8	58,60	+
1962	78,9	40,90	-
1963	23,2	23,45	+
1964	6,0	26,43	+
1965	29,4	24,51	-
1966	22,6	15,66	-
1967	6,8	23,48	+
1968	31,3	27,39	-
1969	31,3	23,94	-
1970	20,5	29,07	+
1971	34,2	30,84	-
1972	32,6	30,72	-
1973	30,6	30,66	+
1974	30,6	31,93	+
1975	33,2	24,36	-
1976	16,8	24,13	+
1977	23,9	18,07	-
1978	12,0	14,88	+
1979	11,7	22,74	+
1980	30,6	22,02	-
1981	21,3	29,06	+
1982	36,1	28,98	-
1983	28,9	28,09	-
1984	27,2	17,80	-
1985	7,5	19,00	+
1986	20,2	23,35	+
1987	27,7	22,02	+
1988	20,7	20,11	-
1989	18,2	27,01	+
1990	33,9	21,35	-
1991	15,7	17,23	+
1992	13,1	15,16	+
1993	13,1	24,48	+
1994	33,8	16,14	-
1995	7,8	19,77	+
1996	23,4	33,29	+
1997	46,8	34,94	-
1998	36,6	38,87	+
1999	42,8	41,14	-
2000	43,4	41,07	-
2001	41,0	41,18	+
2002	41,3	42,94	+
2003	44,7	45,57	+
2004	48,2	55,39	+
2005	65,2	51,14	-
2006	46,9	46,42	-
2007	41,7	44,31	+
2008	42,2	43,36	+
2009	42,4	43,43	+
2010	43,5	41,96	-
2011	40,5	39,78	-
2012	37,6	45,44	+
2013	51,1	41,92	-
2014	38,4	46,66	+
2015	51,4	37,83	-
2016	29,0	35,12	+
2017	32,4	33,81	+
2018	32,5	26,70	-
2019	19,6	23,10	+
2020	19,5	19,75	+
2021	16,4	14,40	-

Результаты статистического сравнения значений интенсивного показателя заболеваемости и второй итерации сглаживания по методу скользящей средней (+ 32 - 30 n=62)

Год	Интенсивный показатель	Третья итерация	Знак
1960	38,3	54,45	+
1961	32,8	48,38	+
1962	78,9	40,40	-
1963	23,2	34,43	+
1964	6,0	28,25	+
1965	29,4	25,51	-
1966	22,6	25,29	+
1967	6,8	24,90	+
1968	31,3	25,85	-
1969	31,3	27,33	-
1970	20,5	28,55	+
1971	34,2	29,38	-
1972	32,6	30,23	-
1973	30,6	28,97	-
1974	30,6	27,45	-
1975	33,2	24,72	-
1976	16,8	21,58	+
1977	23,9	21,09	-
1978	12,0	21,19	+
1979	11,7	23,19	+
1980	30,6	25,13	-
1981	21,3	26,36	+
1982	36,1	24,53	-
1983	28,9	22,69	-
1984	27,2	22,39	-
1985	7,5	22,23	+
1986	20,2	21,65	+
1987	27,7	22,85	-
1988	20,7	22,78	+
1989	18,2	21,37	+
1990	33,9	19,46	-
1991	15,7	20,24	+
1992	13,1	19,41	+
1993	13,1	19,30	+
1994	33,8	22,76	-
1995	7,8	26,68	+
1996	23,4	30,70	+
1997	46,8	34,32	-
1998	36,6	36,91	+
1999	42,8	38,62	-
2000	43,4	40,13	-
2001	41,0	41,87	+
2002	41,3	45,69	+
2003	44,7	48,00	+
2004	48,2	48,19	-
2005	65,2	47,26	-
2006	46,9	46,06	-
2007	41,7	45,10	+
2008	42,2	44,07	+
2009	42,4	42,74	+
2010	43,5	43,09	-
2011	40,5	42,88	+
2012	37,6	43,77	+
2013	51,1	42,51	-
2014	38,4	40,35	+
2015	51,4	38,17	-
2016	29,0	34,76	+
2017	32,4	30,99	-
2018	32,5	27,24	-
2019	19,6	23,09	+
2020	19,5	19,05	-
2021	16,4	15,10	-

**Результаты статистического сравнения значений третьей и четвертой итерации
сглаживания интенсивного показателя заболеваемости по методу
скользящей средней (+ 26 - 36 n=62)**

Год	Третья итерация	Четвертая итерация	Знак
1960	54,45	52,86	-
1961	48,38	47,43	-
1962	40,40	40,93	+
1963	34,43	34,59	+
1964	28,25	30,05	+
1965	25,51	27,67	+
1966	25,29	26,28	+
1967	24,90	26,07	+
1968	25,85	26,70	+
1969	27,33	27,62	+
1970	28,55	28,50	-
1971	29,38	29,37	-
1972	30,23	29,17	-
1973	28,97	28,31	-
1974	27,45	26,51	-
1975	24,72	24,05	-
1976	21,58	22,57	+
1977	21,09	21,88	+
1978	21,19	22,53	+
1979	23,19	23,83	+
1980	25,13	25,10	-
1981	26,36	24,81	-
1982	24,53	23,75	-
1983	22,69	23,07	+
1984	22,39	22,65	+
1985	22,23	22,15	-
1986	21,65	22,50	+
1987	22,85	22,64	-
1988	22,78	22,01	-
1989	21,37	20,73	-
1990	19,46	20,49	+
1991	20,24	19,95	-
1992	19,41	19,62	+
1993	19,30	21,19	+
1994	22,76	23,94	+
1995	26,68	27,32	+
1996	30,70	30,82	+
1997	34,32	33,86	-
1998	36,91	36,24	-
1999	38,62	38,19	-
2000	40,13	40,03	-
2001	41,87	42,86	+
2002	45,69	45,43	-
2003	48,00	46,81	-
2004	48,19	47,03	-
2005	47,26	46,55	-
2006	46,06	45,82	-
2007	45,10	44,95	-
2008	44,07	43,84	-
2009	42,74	43,47	+
2010	43,09	43,17	+
2011	42,88	43,47	+
2012	43,77	42,99	-
2013	42,51	41,67	-
2014	40,35	39,92	-
2015	38,17	37,34	-
2016	34,76	34,17	-
2017	30,99	30,70	-
2018	27,24	26,90	-
2019	23,09	22,97	-
2020	19,05	19,04	-
2021	15,10	15,17	+

(итерация I) во всех случаях значительно сократилось рассеяние значений относительно главной тенденции (рис.), Тем не менее, сохранившаяся нерегулярность динамики показателей не позволяет адекватно аппроксимировать их какой-либо функцией.

В ходе двух последующих итераций (II и III) были получены следующие результаты.

После второй итерации степень рассеяния значений уменьшилась по сравнению с результатами первой, и появилась возможность адекватного определения аппроксимирующей функции. Третья итерация внесла незначительные уточнения в ранее полученные результаты.

Статистическое сравнение значений интенсивного показателя заболеваемости и результатов сглаживания проводилось с использованием критерия знаков [2–4].

Для объема совокупности $n=62$ количество альтернативных разностей при принятии решения о достоверности различия ($P \leq 0.05$) не должно превышать 21 [2–4].

По результатам анализа представленных в табл. 1–4 данных с использованием критерия зна-

ков [2–4] достоверных различий не выявлено. Следовательно, исходный вариационный ряд адекватно моделируется методом сглаживания по скользящей средней.

Матрицы корреляции интенсивного показателя заболеваемости и результатов сглаживания по методу скользящей средней по Пирсону, по Кендаллу и по Спирмену представлены в табл. 5–7 соответственно.

Значимость коэффициентов корреляции для таблиц 5–7 составляет $P < 0.01$.

По результатам анализа матриц корреляции (табл. 5–7), следует вывод о том, что корреляция между значениями с каждой итерацией повышается. При дальнейшем увеличении количества итераций коэффициенты корреляции (по Пирсону, по Кендаллу, по Спирмену) будут асимптотически стремиться к единице – то есть, сглаживание по четвертой и последующим итерациям не имеет практического смысла.

Исходя из вышеизложенного, можно заключить, что применение экспоненциального сглаживания методом скользящей средней может эффективно использоваться для предварительной обра-

Таблица 5

Матрица корреляции интенсивного показателя заболеваемости и результатов сглаживания по методу скользящей средней (по Пирсону)

	Исх	И1	И2	И3	И4
Исх	1	0,702	0,703	0,728	0,721
И1	0,702	1	0,963	0,945	0,934
И2	0,703	0,963	1	0,990	0,981
И3	0,728	0,945	0,990	1	0,981
И4	0,721	0,934	0,981	0,981	1

Таблица 6

Матрица корреляции интенсивного показателя заболеваемости и результатов сглаживания по методу скользящей средней (по Кендаллу)

	Исх	И1	И2	И3	И4
Исх	1	0,561	0,553	0,555	0,550
И1	0,561	1	0,803	0,779	0,754
И2	0,553	0,803	1	0,899	0,852
И3	0,555	0,779	0,899	1	0,938
И4	0,550	0,754	0,852	0,938	1

Таблица 7

Матрица корреляции интенсивного показателя заболеваемости и результатов сглаживания по методу скользящей средней (по Спирмену)

	Исх	И1	И2	И3	И4
Исх	1	0,780	0,785	0,771	0,785
И1	0,780	1	0,933	0,915	0,902
И2	0,785	0,933	1	0,980	0,963
И3	0,771	0,915	0,980	1	0,992
И4	0,785	0,902	0,963	0,992	1

ботки динамических рядов, описывающих развитие динамических процессов, с целью подготовки их к подбору адекватных аппроксимирующих функций. Необходимый эффект достигается в результате двух, реже трех циклов итераций. При возрастании количества итераций эффект сглаживания становится незначительным, снижаясь с каждым следующим циклом.

Литература

1. Лаптёнок, С.А. Системный анализ геоэкологических данных в целях митигации чрезвычайных ситуаций / С.А.Лаптёнок. – Минск: БНТУ, 2013. – 287 с.
2. Бубнов, В.П. Решение задач экологического менеджмента с использованием методологии системного анализа / В.П.Бубнов, С.В.Дорожко, С.А.Лаптёнок. – Минск: БНТУ, 2009. – 266 с.
3. Моделирование экологических процессов. Учебное пособие / С.А.Лаптёнок, И.В.Лазар, Н.М.Новикова [и др.]. – Минск: ИВЦ Минфина, 2022. – 144 с.
4. Johnson, R. Elementary Statistics. 4th edition / R.Johnson. – Boston: Duxbury Press, 1984. – 557 p.

USE OF DATA SMOOTHING BY MOVING AVERAGE METHOD IN TUBERCULOSIS MORBIDITY DYNAMICS MODELING

¹S.A.Laptyonok, ²Yu.D.Sinkevich,
¹A.A.Kologrivko, ²O.I.Rodzkin, ¹S.A.Horeva,
¹Yu.V.Klyausova

¹Belarusian National Technical University, 65, Nezavisimosti Ave., 220013, Minsk, Republic of Belarus

²International Sakharov Environmental Institute of Belarusian State University, 23/1, Dolgobrodskaya Str., 220070, Minsk, Republic of Belarus

Results are presented of effectiveness evaluating of data smoothing using the moving average method in dynamics modeling of tuberculosis incidence in population of Oshmyany District (Grodno Region, Republic of Belarus) for the period 1960–2021.

Keywords: modeling of morbidity dynamics; data smoothing; moving average method; efficiency; tuberculosis.

Сведения об авторах:

Лаптёнок Сергей Антонович, канд. техн. наук, доцент; Белорусский национальный технический университет, факультет горного дела и инженерной экологии, доцент кафедры «Инженерная экология»; тел.: (+37529) 6257828; e-mail: 267413@mail.ru.

Синкевич Юлия Дмитриевна; «Международный государственный экологический институт имени А.Д.Сахарова» Белорусского государственного университета, студент; тел.: (+37529) 6940079; e-mail: cherskayau@yandex.by.

Кологривко Андрей Андреевич, канд. техн. наук, доцент; Белорусский национальный технический университет, декан факультета горного дела и инженерной экологии; тел.: (+37529) 6840510; e-mail: kologrivko@tut.by.

Родькин Олег Иванович, д-р биол. наук, доцент; «Международный государственный экологический институт имени А.Д.Сахарова» Белорусского государственного университета, директор; тел.: (+37529) 3871501; e-mail: aleh.rodzkin@rambler.ru.

Хорева Светлана Алексеевна, д-р биол. наук, профессор; Белорусский национальный технический университет, факультет горного дела и инженерной экологии, профессор кафедры «Инженерная экология»; тел.: (+37529) 6857939; e-mail: horeva.svetlana@mail.ru.

Кляусова Юлия Владимировна, канд. с.-х. наук, доцент; Белорусский национальный технический университет, факультет горного дела и инженерной экологии, доцент кафедры «Инженерная экология»; тел.: (+37529) 1679237; e-mail: yuliya-klaus@mail.ru.

Поступила 21.09.2023 г.