

УДК 519.87

DOI 10.33764/2411-1759-2025-30-4-15-25

## Формирование групп статистически однородных исходных данных для улучшения качества построения прогнозных математических моделей кинематического типа

В. С. Хорошилов<sup>1</sup>✉, Н. Н. Кобелева<sup>1</sup>

<sup>1</sup>Сибирский государственный университет геосистем и технологий, г. Новосибирск, Российская Федерация

e-mail: Khoroshilovvs@mail.ru

**Аннотация.** Одной из ключевых проблем при построении прогнозных математических моделей, включая и модели кинематического типа, является неоднородность исходных данных. Неоднородность может проявляться в различных формах, включая вариативность, наличие выбросов, асимметричные распределения и другие сложности, которые значительно ухудшают качество прогнозов и усложняют процесс моделирования. Для повышения точности прогнозных моделей необходимо разделять исходные данные на группы, обладающие схожими статистическими характеристиками, что позволяет снизить влияние различных источников ошибок и улучшить стабильность результатов. Это особенно актуально при работе с неоднородными массивами данных, где традиционные методы анализа часто оказываются недостаточными для выявления скрытых закономерностей деформационного процесса. Статистическая однородность отдельных реализаций наблюдаемых деформационных марок предполагает детерминировано-вероятностную природу развития процесса деформации инженерных сооружений и ее корректное применение при построении прогнозной математической модели, что должно обеспечиваться статистическим критерием репрезентативности объема выборки. В статье рассматриваются этапы выполнения исследования на основе дисперсионного анализа для корректного выделения групп статистически однородных исходных данных, что позволяет в дальнейшем улучшить качество построения прогнозных математических моделей. С этой целью исследуются возможности применения таких ключевых статистических показателей, как коэффициенты корреляции и вариации, величина дисперсии, средние значения и другие показатели.

**Ключевые слова:** деформации сооружений, дисперсионный анализ, статистически однородные данные, арифметические средние, коэффициент вариации, математическое моделирование, прогнозная модель

### Для цитирования:

Хорошилов В. С., Кобелева Н. Н. Формирование групп статистически однородных исходных данных для улучшения качества построения прогнозных математических моделей кинематического типа // Вестник СГУГиТ. – 2025. – Т. 30, № 4. – С. 15–25. – DOI 10.33764/2411-1759-2025-30-4-15-25

### Постановка задачи

В рамках построения прогнозных математических моделей кинематического типа при выделении статистически однородной выборки из ряда измеренных реализаций процесса преследуется цель – определить после-

довательность проведения анализа для выделения подобной выборки и уточнить применяемые ключевые статистические показатели при решении данной задачи. Важность подобного исследования заключается в возможности более корректного выделения групп статистически однородных исходных данных,

что в дальнейшем позволяет улучшить качество построения прогнозных математических моделей и обеспечить более качественное прогнозирование перемещений деформационных марок для исследуемого процесса поведения инженерного сооружения.

Для решения данной задачи предполагается определить основные этапы, которые позволяют с использованием ряда статистических показателей выявить структуру данных в выборке. Это может включать как выделение групп данных, обладающих схожими статистическими характеристиками, так и выявление закономерностей в распределении измеренных величин. С этой целью исследовались возможности применения таких ключевых статистических показателей, как коэффициенты корреляции и вариации, величина дисперсии, средние значения и другие показатели структуры данных:

- коэффициент корреляции позволяет оценить степень взаимосвязи между двумя переменными. При выявлении групп со схожими статистическими характеристиками на основе коэффициентов корреляции можно определить, какие переменные взаимосвязаны и могут быть сгруппированы вместе;

- величина дисперсии позволяет оценить разброс значений в выборке относительно их среднего значения. Группы данных с близкими значениями дисперсии могут быть объединены вместе как имеющие схожую изменчивость;

- анализ средних значений в различных группах данных позволяет определить, есть ли статистически значимые различия между ними. Группы с близкими средними значениями могут быть рассмотрены как имеющие схожие характеристики;

- коэффициент вариации позволяет выявить группы с близкими значениями, что отражает схожий уровень изменчивости.

### *Состояние вопроса*

Рассматривая инженерное сооружение как объект исследования, можно отметить, что оно самым тесным образом взаимодействует как с грунтовым основанием, так и с внешней средой. При этом подсистема

«сооружение» формируется в результате определенной упорядоченности, складывающейся из его объемно-планировочных и конструктивных решений. Это проявляется в конкретных типах строительных конструкций и их размерах, видах материалов несущих конструкций и их свойствах в процессе искусственного изготовления. Немаловажную роль здесь играют взаимное расположение элементов конструкций и возникающие при этом связи в результате планомерного, т. е. упорядоченного размещения конструкций в процессе строительства, что приводит к их инженерно-заданной работе. Поэтому подобная преднамеренная упорядоченность является основанием для реализации расчетов деформации сооружения с применением детерминистических методов. Подсистема «основание» определяется, главным образом, инженерно-геологическими условиями грунтового основания, т. е. условиями залегания и сложения грунтов конкретного основания под инженерное сооружение [1]. Вследствие постоянного давления сооружения на грунтовое основание происходит изменение объема грунта в процессе движения воды по его порам и возникающего внутреннего сопротивления грунта собственному разрушению. А появляющиеся при длительной нагрузке реологические процессы ведут к проявлениям таких свойств грунта, как ползучесть, снижение прочности и т. п. Однако в силу сложности и многообразия свойств грунтов изучить детально каждое отдельное основание под сооружение, его физико-механические свойства и глубину сжимаемой толщи грунтов представляется просто невозможным. В силу этих обстоятельств в теории механики грунтов принято считать случайной природу основания сооружения, а для вычисления необходимых характеристик грунтов применяются вероятностно-статистические методы. В подсистему «внешняя среда» включаются всевозможные нагрузки и воздействия на сооружение. Природа нагрузок достаточно разнообразна, но общепризнанным является то обстоятельство, что внешние условия эксплуатации сооружения представляют собой случайные про-

цессы. Из всего вышесказанного можно сделать вывод о смешанной (детерминированной и случайной) природе системы: сооружение – грунтовое основание – внешняя среда. Но в то же время, так как в конкретной структуре преобладают жестко установленные связи между конструктивными элементами сооружения, грунтовым основанием и условиями проектирования, все это приводит к значительной детерминированности процесса деформации инженерного сооружения. Подобные условия функционирования сооружения дают возможность сформулировать следующее положение: «из строго детерминированных единичных явлений складывается статистическая закономерность, с помощью которой можно получить достоверные выводы чисто математическими средствами» [2].

Геодезические данные наблюдений за деформациями сооружений обладают уникальной информацией, а именно, они позволяют выявить преобладающую детерминированную составляющую процесса деформации сооружения, что в определенной степени устраняет возможную неопределенность процесса. Это достигается правильностью выбора местоположения контролируемых деформационных марок, т. е. их расположение должно в полной мере характеризовать перемещение как отдельных наиболее характерных точек инженерного сооружения, так и основных его осей. А зная величины перемещения характерных точек, всегда есть возможность рассчитать нужные количественные характеристики любого вида деформации. При этом погрешности геодезических измерений имеют, как правило, случайный характер, а величины их значений, в большинстве случаев, ничтожно малы в сравнении с величинами измеряемых деформаций.

Применение корреляционной теории случайных функций позволяет достаточно корректно осуществить необходимое математическое моделирование процессов деформации инженерных сооружений. При этом, обеспечение необходимых для этого таких условий, как нормальность и линейность необходимой для анализа выборки из ряда геодезических наблюдений за положением де-

формационных марок, дает возможность правомерно использовать известные методы и процедуры оценки параметров математических моделей, гарантируя их математическую строгость [2, 3]. В качестве дополнительного условия, а в нашем случае наблюдается детерминировано-вероятностная природа развития деформации сооружений, для выбора периода основания прогноза из ряда реализаций процесса при построении прогнозной модели необходимо добиться формирования статистической однородности результатов измерений осадок сооружений из различных циклов наблюдений. По существу, данное условие реализует собой статистический критерий репрезентативности объема исследуемой выборки, что приводит к повышению уровня качества и надежности конечных прогнозных значений. В нашем случае математические модели и методы оценивания параметров ориентированы на нормальное распределение описываемых процессов деформаций. Статистическая проверка гипотезы о нормальности распределения значений величины деформации в каждом сечении процесса нами осуществляются приближенным методом с использованием коэффициентов асимметрии и эксцесса [4]. В случаях, когда выявляется аномальность распределения в большей части сечений процесса, производится его центрирование и нормализация путем экспоненциального или логарифмического преобразования.

Линейность процессов деформаций сооружений необходимо оценивать в соответствии с ее неоднозначными понятиями, чтобы правильно определиться с выбором вида математической модели. Предварительная оценка линейности производится по графикам изменения средних значений и дисперсии наблюдаемых перемещений деформационных марок, которые сравниваются с характером развития воздействующих факторов и другого прогнозного фона. Более глубокий анализ линейности основывается на изучении физической сущности моделируемых процессов. При построении кинематических моделей нелинейный тренд, как правило, исключается и учитывается в дальнейшем совместно с линеаризован-

ным центрированным процессом. Если таким путем не удастся учесть нелинейность, то следует линеаризовать процесс известными математическими методами, как это показано нами в работе [3], или же применять нелинейную динамическую модель.

В работах [2, 3, 5] представлены примеры формирования однородных групп на основе последовательного перебора различных реализаций деформационных марок. Первоначально, за однородную группу принимаются все наблюдаемые реализации исследуемого процесса. Далее для выяснения принадлежности  $i$ -й реализации  $x_e(t_j)$

к той же генеральной совокупности рассчитывается обобщенное расстояние Махалонбиса, а затем определяется коэффициент Хотеллинга с использованием выражения:

$$T^2 = \frac{n}{n+1} D^2, \quad (1)$$

по которой находится статистика:

$$R = \frac{1}{1 + \frac{1}{n-1} T^2}. \quad (2)$$

Статистика  $R$  подчиняется  $B$ -распределению:

$$B\left(y, \frac{n-N}{2}, \frac{N}{2}\right), N < n. \quad (3)$$

В дальнейшем выделяются однородная и отличающаяся сформированные группы в смысле описанного выше критерия Хотеллинга. Затем отличающаяся группа вновь подвергается анализу на однородность и т. д. Уровень значимости при проверке данной статистической гипотезы выбирается в зависимости от требуемой точности прогнозирования. При этом корректность прогнозирования в последующем возможно выполнить на основе контрольных прогнозов по построенным моделям. Условие, необходимое для выполнения контрольного прогнозирования, заключается в том, чтобы формирование статистически однородных групп исходных данных проводилось только тогда, когда число выбранных циклов

наблюдений для выбора периода основания прогноза было меньше числа реализаций процесса, т. е. числа деформационных марок.

В работе [3] нами представлен подобный процесс формирования статистически однородных групп деформационных марок для основного фрагмента из пяти реализаций. Подобная задача выполнялась и для дополнительного фрагмента с целью продемонстрировать возможные неблагоприятные последствия некачественной группировки деформационных марок также для пяти реализаций процесса. В то же время отдельные этапы осуществления данного процесса формирования статистически однородных групп необходимо уточнить и дополнить.

На начальном этапе отбора необходимого статистически однородного материала для каждого фрагмента из пяти реализаций процесса анализировались графики построенных изображения изолиний осадки сооружения. При этом, в случае близости изолиний деформационных марок и подобных им изолиний в одном и том же цикле наблюдений и к таким же изолиниям в других выбранных циклах наблюдений, принималось решение о допустимой статистической однородности значений осадочных марок. Следующий этап предполагал визуальный анализ построенных графиков осадок деформационных марок с целью выделения участков, для которых обеспечивается линейность развития процесса осадки сооружения. Данный анализ осуществлялся поэтапно для каждого из двух сформированных фрагментов – основного и дополнительного. На последующих этапах уже рассчитывали такие параметры, как средние значения осадки в каждом отдельном цикле наблюдений  $\tilde{m}_x(t_j)$ , стандарт осадки  $\tilde{\sigma}_x(t_j)$  и коэффициент вариации осадки  $\tilde{v}_x(t_j)$ . Все результаты вычисленных значений параметров осадок деформационных марок из двух сформированных фрагментов – основного и дополнительного – показаны в табл. 1 и 2.

Таблица 1

Вычисленные значения параметров осадок деформационных марок основного фрагмента

Циклы	2	3	4	5	6	7	8	9	10	11	12
$t_j$	0,5	1,5	2,5	3,4	4,5	5,5	6,5	13,1	16,8	19,8	22,8
$\tilde{\sigma}_x(t_j)$	2,42	2,77	1,95	1,82	1,79	2,17	2,17	1,79	1,79	1,92	1,95
$\tilde{m}_x(t_j)$	9,8	20,2	45,6	54,4	<b>58,2</b>	<b>59,8</b>	<b>60,8</b>	67,2	72,2	74,2	78,4
$\tilde{\nu}_x(t_j)$	0,249	0,137	0,043	0,039	<b>0,031</b>	<b>0,036</b>	<b>0,034</b>	0,027	0,025	0,026	0,025

Таблица 2

Вычисленные значения параметров осадок деформационных марок дополнительного фрагмента

Циклы	2	3	4	5	6	7	8	9	10	11	12
$t_j$	0,5	1,5	2,5	3,4	4,5	5,5	6,5	13,1	16,8	19,8	22,8
$\tilde{\sigma}_x(t_j)$	2,17	4,18	5,89	7,08	7,65	7,86	7,90	9,57	10,43	10,43	10,47
$\tilde{m}_x(t_j)$	7,8	19,0	39,2	<b>45,2</b>	<b>47,8</b>	<b>49,4</b>	<b>49,6</b>	54,8	59,6	62,6	66,8
$\tilde{\nu}_x(t_j)$	0,278	0,220	0,150	0,157	0,160	0,159	0,159	0,175	0,175	0,167	0,157

Сходство в изменениях рассчитанных значений величин оценок параметров изучаемой осадки деформационных марок для сформированных групп основного и дополнительного фрагментов проявилось в следующем [3, 6]. И в основном и дополнительном фрагментах выявились изменения в значениях коэффициента вариации осадки  $\tilde{\nu}_x(t_j)$  как реакция на понижение строительной нагрузки сооружения уже с 4-го цикла наблюдений. И если величины средних значений осадки деформационных марок в 6, 7 и 8-м циклах наблюдений основного фрагмента и в 5, 6, 7 и 8-м циклах дополнительного фрагмента оказались близко расположенными по отношению друг к другу, то значения коэффициентов вариации существенно изменились по величине, а именно, от величин (0,031–0,036) для основного фрагмента до значений (0,16–0,18) – в дополнительном. По нашему мнению, причиной такого изменения величин коэффициентов вариации осадки явилось возрастание неравномерности осадки самого сооружения, т. е. изменилась степень деформированности фунда-

ментной плиты. В связи с этим можно утверждать, что близость значений величины средних значений осадки деформационных марок в основном и дополнительном фрагментах и степень изменения коэффициентов вариации осадки в выбранных циклах наблюдений можно рассматривать как признак статистической однородности осадки на протяжении этих циклов. Основное различие проявилось лишь в том, что величины коэффициента вариации осадки (0,16–0,18) свидетельствуют о дополнительной опасности и возможности трещинообразования в фундаменте для участка дополнительного фрагмента при достаточно небольшой величине его средней осадки.

### Результаты исследования

Предпосылки, необходимые для обоснованного конструктивного применения вероятностно-статистических методов при построении математической модели кинематического типа для изучения исследуемого деформационного процесса, определяются тем, что строящаяся прогнозная модель разраба-

тывается в рамках корреляционной теории случайных функций. В этом случае следует, прежде всего, проверять линейность и нормальность распределения моделируемого процесса, а при необходимости обеспечить выполнение данных условий путем линеаризации и нормализующих преобразований или надлежащим объединением групп реализаций из деформационных марок и корректно выбранных сечений процесса. При этом терминология линейности деформационного процесса для кинематической и динамической моделей интерпретируется нами по-разному. Для кинематической модели под линейностью процесса деформации мы понимаем строго линейную функциональную зависимость вариации изучаемой величины осадки деформационных марок от изменения временного фактора [3]. В динамических моделях понятие линейности предполагает выполнение условия сохранения принципов пропорциональности между входным воздействием и выходным сигналом; при этом выполнение подобного условия осуществляется с помощью линейного оператора. В качестве примера простейшего линейного оператора может выступать экспонента. Особое внимание должно уделяться размещению наблюдаемых деформационных марок. Предполагается, что подобное размещение должно включать наиболее важные, т. е. исключительно характерные точки, отражающие основные проявления деформации или наоборот, наиболее «слабые» узлы системы «сооружение – грунтовое основание – внешняя среда». Местоположение этих характерных точек и «слабых» узлов формируется на первоначальном этапе перед началом строительства сооружения на основе имеющейся априорной информации, например, результатов предварительных исследований неоднородности состава грунтов в основании под сооружение, расчетных значений возможных нагрузок на основание, расчетной пространственной жесткости сооружения и т. п. При этом следует учитывать и то обстоятельство, какие нужные количественные характеристики деформации подлежат изучению.

Среднее значение и коэффициент вариации являются двумя различными статистическими показателями, но они могут быть связаны друг с другом, особенно при анализе изменчивости данных в выборке по отдельным циклам наблюдений [7, 8]. Среднее значение определяется как сумма всех значений результатов измерений (количество реализаций процесса) по отношению к количеству этих реализаций как в отдельном цикле, так и по всем циклам наблюдений. Обобщенное среднее по всем циклам наблюдений в этом случае характеризует преобладающую тенденцию в оценке распределения исходных данных, т. е. показывает, где предположительно располагается обобщенная величина из всех реализаций процесса. Коэффициент вариации, определяемый как величина степени изменчивости в виде отношения средней квадратической погрешности к арифметическому среднему в отдельных циклах наблюдений, характеризует уровень изменчивости величин осадок деформационных марок в зависимости от степени неравномерности этих осадок, т. е. чем больше величина коэффициента вариации, тем выше степень этой неравномерности. Наличие высокого коэффициента вариации может также указывать на необходимость дополнительного исследования для более полного понимания причин такой неравномерности осадок и корректной интерпретации результатов анализа. В то же время существует определенная связь между средним значением и коэффициентом вариации, которая заключается в том, что высокий уровень изменчивости данных (высокий коэффициент вариации) может указывать на то, что среднее значение выборки менее репрезентативно или вариативно. Например, если у выборки очень высокий коэффициент вариации, это может означать, что среднее значение не является хорошим показателем центральной тенденции из-за большого разброса данных. Следовательно, при анализе данных стоит обращать внимание как на среднее значение, так и на коэффициент вариации, чтобы полнее оценить особенности выборки и изменчивость данных.

Применение дисперсионного анализа для формирования статистически однородных

групп деформационных марок из различных циклов наблюдений для корректного конструирования математической модели кинематического типа при условии выполнения нормальности и линейности является достаточно эффективным подходом, так как обеспечивает хорошую точность оценок параметров, высокую мощность теста (возможность обнаружения статистически значимых различий) и позволяет сделать более информативные выводы [9]. Для проведения анализа дисперсий при исследовании статистической однородности измерений в качестве воздействующего фактора выбирался временной фактор между отдельными циклами измерений. Анализ включал следующие шаги:

– формулирование гипотез (нулевая: средние значения измерений в различных циклах равны, т. е. временной фактор не оказывает существенного влияния на изменение средних – *выборка является статистически однородной*; альтернативная: средние значения измерений в различных циклах различны);

– проведение дисперсионного анализа: вычисление средних значений результатов измерений осадок в каждом цикле и обобщенного среднего; расчет межгрупповой дисперсии (межгрупповой суммы квадратов) и внутригрупповой дисперсии (внутригрупповой суммы квадратов); вычисление статистики  $F_{\text{выч}}$ ;

– принятие решения: осуществляется сравнение рассчитанного значения статистики  $F_{\text{выч}}$  с допустимой критической величиной, определенной по таблице распределения Фишера для выбранного уровня значимости (как правило, значение выбирается:  $\alpha = 0,05$ ) и степеней свободы. В случае, если расчетное значение статистики  $F_{\text{выч}}$  оказывается меньше критического  $F_{\text{крит}}$ . – нулевая гипотеза принимается, т. е. временной фактор не оказывает существенного влияния на изменение средних – *выборка из всех реализаций процесса является статистически однородной*. Если же расчетное значение статистики превышает критическое значение, то нулевая гипотеза отвергается как свидетельство о наличии статистически существенных различий между

средними значениями измерений в различных циклах наблюдений. В этом случае желательнее подобрать новую группу деформационных марок для последующего исследования.

Сформулированная выше последовательность действий для выявления статистически однородной выборки была реализована в процессе проведения дисперсионного анализа в среде Excel, результаты которого представлены на рис. 1 и 2.

С одной стороны, как было отмечено выше, средние значения осадок деформационных марок основного фрагмента в 6, 7 и 8-м циклах наблюдений оказались близки по отношению друг к другу. В то же время достаточно небольшие значения коэффициентов вариации осадки (0,031–0,036), наблюдавшиеся начиная с 4-го цикла и в последующем преимущественно снижающиеся (см. табл. 1), свидетельствуют о низком уровне деформируемости выбранного для исследования основного фрагмента фундаментной плиты. Это позволяет сделать вывод о достаточно корректном выборе циклов наблюдений для периода основания прогноза при построении прогнозной математической модели. Однако, при этом следует понимать, что рассчитанные средние значения осадок деформационных марок еще не представляют собой объективную информацию и недостаточно полно отражают сформированные группы марок по отдельным циклам измерений, так как не учитываются их разброс и разнообразие. Эту недостающую информацию позволяют дополнить рассчитанные значения коэффициентов вариации осадки. С другой стороны, невысокие значения коэффициентов вариации осадки, как характеристика степени неравномерности осадки [10], свидетельствуют о более высокой точности прогнозирования по построенной прогнозной модели (при использовании для аппроксимации тренда в виде экспоненциальной зависимости, результаты долгосрочного прогнозирования для основного фрагмента очень хорошо укладываются в предвычисленный коридор погрешности прогноза [3, 6].

ОДНОФАКТОРНЫЙ ДИСПЕРСИОННЫЙ АНАЛИЗ - ОСНОВНОЙ ФРАГМЕНТ					
Параметры	Номера осадочных марок				
	22	23	24	25	26
Цикл/время	6/4,5	60	60	57	58
	7/5,5	62	62	59	59
	8/6,5	63	63	60	58
ГИПОТЕЗА=групповые средние равны общей средней (мат. ожид. близки в циклах)					
X6/4,5: сред. в цикле	58,20				
X7/5,5: сред. в цикле	59,80				
X8/6,5: сред. в цикле	60,80				
Хср. - общее:	59,600				
Квадрат откл-я от общ.ср.	0,1600	0,1600	6,7600	12,9600	2,5600
	5,7600	5,7600	0,3600	6,7600	0,3600
	11,5600	0,1600	2,5600	0,1600	
Сумма кв.откл.-ий - общее	11,5600				
сумма кв.откл. в группах	3,2400	3,2400	1,4400	4,8400	0,0400
	4,8400	4,8400	0,6400	7,8400	0,6400
	4,8400	4,8400	0,6400	7,8400	0,6400
Внутри груп.сум.кв.откл.-ий	11,5600				
Межгруп.сум.кв.откл.-ий	11,5600				
Проверка	11,5600				
Числ.степ.своб. - полн	14	число всех вариан.-1			
Числ.степ.своб. - межгруп.	2	число всех групп - 1			
Числ.степ.своб.внутригруп.	12	число всех вар.-числ.груп			
Ср.кв.драт между групп.	8,6000				
Ср.кв.драт внутри групп	4,2000				
Fвыч.	2,0476				
Fкрит.	3,8853				
Вывод: Fвыч. < Fкрит.	Матем. ожидания в циклах близки друг к другу - однородность выборки подтверждается				

Рис. 1. Результаты дисперсионного анализа в Excel для основного фрагмента

Для дополнительно фрагмента сложилась несколько иная картина. Местоположение наблюдаемых деформационных марок было выбрано на мало осевшем участке фундаментной плиты сооружения. Однако, при этом для данного участка оказалось характерным в разных местах наличие зон сгущенных изолиний. В свою очередь, это обстоятельство привело к тому, что, с одной стороны, подобное местоположение деформационных марок позволило сформировать статистически однородную группу из марок, так как в результате выполненных расчетов величины средних значений осадок в 5, 6, 7 и 8-м циклах оказались близко расположенными по отношению друг к другу (см. табл. 2). С другой стороны, выявилось существенно значимое влияние неравномерности осадки сооружения на сформированную группу марок дополнительного фрагмента, т. е. непосредственно возросла степень деформированности фрагмента – возрастание коэффициента вариации осадки до величины (0,16–0,18). И как результат, с одной стороны, – близость средних значений обеспечила правильность и корректность выбора периода основания прогноза для построения прогнозной математической модели; с другой, – возрастание коэффициента вариации осадки (как возросшая неравномерная деформированность фрагмен-

та) не позволило построить корректную прогнозную модель, т. е. тем самым ухудшилось качество прогнозирования по прогнозной модели (результаты долгосрочного прогноза для дополнительного фрагмента лежат в пределах удвоенного стандарта предвычисленного коридора погрешности прогноза [6]).

ОДНОФАКТОРНЫЙ ДИСПЕРСИОННЫЙ АНАЛИЗ - ДОПОЛНИТЕЛЬНЫЙ ФРАГМЕНТ						
Параметры	Номера осадочных марок					
	33	34	35	36	37	
Цикл/время	5/3,4	53	51	37	39	
	6/4,5	56	54	39	41	
	7/5,5	58	56	40	43	
	8/6,5	58	56	40	43	
ГИПОТЕЗА=групповые средние равны общей средней (мат. ожид. близки в циклах)						
R=групповые средние отличаются от общей следней (мат. ожид. различаются в циклах)						
X5/3,4: сред. в цикле	45,20					
X6/4,5: сред. в цикле	47,80					
X7/5,5: сред. в цикле	49,40					
X8/6,5: сред. в цикле	49,60					
Хср.общ:	48,000					
Квадр.откл. от общ.ср.	25,000	9,000	121,000	81,000	4,000	
	64,000	36,000	81,000	49,000	1,000	
	100,000	64,000	64,000	25,000	4,000	
	100,000	64,000	64,000	25,000	9,000	
Сумма кв.откл. общее	11,5600					
сумма кв.откл. в группах	60,8400	33,6400	67,2400	38,4400	0,6400	
	67,2400	38,4400	77,4400	46,2400	1,4400	
	73,9600	43,5600	88,3600	40,9600	0,3600	
	70,5600	40,9600	92,1600	43,5600	1,9600	
Внутр.груп.сум.кв.откл.	11,5600					
Межгруп.сум.кв.откл.	62,0000					
Проверка	11,5600					
Числ.степ.своб.полн	19	число всех вариан.-1				
Числ.степ.своб.межгруп.	3	число всех групп - 1				
Числ.степ.своб.внутригруп.	16	число всех вар.-числ.груп				
Ср.кв.драт между групп.	20,6667					
Ср.кв.драт внутри групп	58,0000					
Fвыч.	0,3563		Fкрит.	3,2389		
Вывод: Fвыч. < Fкрит.	Матем. ожидания в циклах близки друг к другу - однородность выборки подтверждается					

Рис. 2. Результаты дисперсионного анализа в Excel для дополнительного фрагмента

### Заключение

*Интерпретация коэффициента вариации.* Для построения математической модели при прогнозировании деформаций сооружения при выборе периода основания прогноза рост или понижение величины коэффициента вариации в сформированной выборке дает следующий результат: рост коэффициента вариации указывает на увеличение изменчивости деформаций сооружения во времени. Это может быть связано с различными факторами, например, такими как произошедшие изменения в условиях эксплуатации сооружения, воздействие внешних факторов или другие причины. Поэтому в ряде случаев оказывается целесообразным учесть возможное увеличение изменчивости в математической модели для более точного прогнозирования, т. е. желательно сформировать другую группу

деформационных марок при проведении дальнейшего исследования. Понижение же значения коэффициента вариации в выборке означает уменьшение изменчивости деформаций сооружения во времени. Это может быть связано, например, с улучшением условий эксплуатации сооружения, выполнением принятых мер по снижению степени изменчивости или корректным выбором группы деформационных марок для более точного прогнозирования.

*Интерпретация средних значений:* если средние значения оказались близки в различных циклах наблюдений, это может свидетельствовать о стабильности или однородности средних показателей в этих циклах, что предполагает использование сформированной выборки из деформационных марок для корректного выбора периода основания прогноза при построении кинематической мо-

дели. И в то же время, необходимо помнить, что средние значения еще недостаточно полно описывают исходные данные и не учитывают разброс и разнообразие значений. При анализе данных стоит обращать внимание как на средние значения, так и на величину коэффициента вариации, чтобы полнее оценить особенности выборки и изменчивость данных.

### **Финансирование**

Исследования проведены в процессе выполнения государственного задания Минобрнауки России (тема «Автоматический геодезический мониторинг природной среды и инженерных сооружений средствами малобюджетных высокоточных датчиков вертикальных перемещений в условиях Крайнего Севера», № FEFS-2023-0003).

### **БИБЛИОГРАФИЧЕСКИЙ СПИСОК**

1. Пособие к СНиП 2-02-01–83 (Пособие по проектированию оснований зданий и сооружений) НИИОСП им. Н. М. Герсеванова. – М. : Стройиздат, 1984. – 376 с.
2. Гуляев Ю. П. Прогнозирование деформации сооружений на основе результатов геодезических наблюдений : монография. – Новосибирск : СГГА, 2008. – 256 с. – ISBN 978-5-87693-290-7. – EDN SAQQCR.
3. Гуляев Ю. П., Хорошилов В. С. Математическое моделирование. Анализ и прогнозирование деформаций сооружений по геодезическим данным на основе кинематической модели : учеб. пособие. – Новосибирск : СГГА, 2012. – 91 с. – ISBN 978-5-87693-505-2. – EDN QNQDXD.
4. Румшицкий Л. З. Элементы теории вероятностей. – М. : Наука. – 1971. – 256 с.
5. Khoroshilov V. S., Kobeleva N. N., Sycheva N. V. Mathematical modeling of the high-rise buildings deformation development process in Moscow (Vosstania square) // IOP Conference Series: Materials Science and Engineering : International Scientific Conference "Construction and Architecture: Theory and Practice of Innovative Development" - Hydrometeorological and Geodetic Research in the Building Area, Kislovodsk, 01–05 октября 2019 года. Vol. 698, 4. – Kislovodsk: Institute of Physics Publishing, 2019. – P. 044004. – DOI 10.1088/1757-899X/698/4/044004. – EDN QWILTL.
6. Хорошилов В. С., Квашенко И. Ю., Носков М. Ф. Особенности выбора деформационных марок для построения кинематической модели при изучении деформаций сооружений // Изв. вузов. «Геодезия и аэрофотосъемка». – 2013. – № 4/С. – С. 58–61. – EDN UIYCCD.
7. Гмурман В. Е. Теория вероятностей и математическая статистика. – М. : Высшая школа, 2003. – 523 с. – ISBN 5-06-004214-6. – EDN QJLKXP.
8. Николаев С. А. Статистические исследования осадок инженерных сооружений. – М. : Недра, 1983. – 112 с.
9. Шеффе Г. Дисперсионный анализ / пер. с англ. Б. А. Севастьянова и В. П. Чистякова. – М. : Госуд. изд-во «Физико-математическая литература». 1963. – 625 с.
10. Прогноз скорости осадок оснований сооружений / Под. ред. Н. А. Цитовича. – М. : Стройиздат, 1967. – 239 с.

## Об авторах

Валерий Степанович Хорошилов – доктор технических наук, доцент, профессор кафедры космической и физической геодезии.

Наталья Николаевна Кобелева – кандидат технических наук, доцент кафедры космической и физической геодезии.

Получено 18.12.2024

© В. С. Хорошилов, Н. Н. Кобелева, 2025

## Formation of statistically homogeneous groups of initial data to improve the quality of predictive mathematical models of kinematic type

V. S. Khoroshilov<sup>1</sup>✉, N. N. Kobeleva<sup>1</sup>

<sup>1</sup> Siberian State University of Geosystems and Technologies, Novosibirsk, Russian Federation

e-mail Khoroshilovvs@mail.ru

**Annotation.** One of the key challenges when constructing predictive mathematical models, including those of a kinematic type, is the heterogeneity of raw data. Heterogeneity can be expressed in such ways as variability, presence of outliers, asymmetric distributions, and other complexities that significantly reduce forecast accuracy and complicate the modeling process. To enhance model performance, it is essential to divide raw data into groups sharing similar statistical characteristics, which helps minimize errors introduced by diverse sources and improves result stability. It is particularly important when dealing with heterogeneous datasets, when traditional analysis methods often prove insufficient for uncovering hidden patterns within deformation processes. Statistical homogeneity of individual implementations of observed deformation marks implies a deterministic and probabilistic nature of the engineering structure deformation process, and its correct application in developing a predictive mathematical model must be ensured by a statistical criterion of sample representativeness. The stages of conducting research based on variance analysis aimed at correctly identifying groups of statistically homogeneous initial data, which allows further improvement of the quality of predictive mathematical models, are discussed in the paper. For this purpose, the possibilities of using key statistical indicators such as correlation coefficients, variation ratios, dispersion magnitudes, mean values, and others are investigated.

**Keywords:** deformations of structures, analysis of variance, statistically homogeneous data, arithmetic averages, coefficient of variation, mathematical modeling, forecast model

## REFERENCES

1. Building regulations 2-02-01-83 (1984). *Posobie k SNIp 2-02-01-83 (Posobie po proektirovaniyu osnovaniy zdaniy i sooruzheniy) [Manual for designing foundations of buildings and structures]*: Moscow: NIIOSP im. N.M. Gersevanov.
2. Gulyaev, Yu. P. (2008). *Prognozirovanie deformatsii sooruzheniy na osnove rezul'tatov geodezicheskikh nablyudeniy [Forecasting the deformation of structures based on the results of geodetic observations]*. Novosibirsk: SSGA Publ., 256 p. ISBN 978-5-87693-290-7. EDN SAQQCR. [in Russian].
3. Gulyaev, Yu. P. Khoroshilov, V. S. (2012). *Matematicheskoe modelirovanie. Analiz i prognozirovaniye deformatsiy sooruzheniy po geodezicheskim dannym na osnove kinematicheskoy modeli [Mathematical modeling. Analysis and forecasting of deformations of structures using geodetic data based on a kinematic model]*: Novosibirsk: SSGA, Publ., 91 p ISBN 978-5-87693-505-2. EDN QNQDXD. [in Russian].

4. Rumshiskii, L. Z. (1971). *Elementy teorii veroyatnostey [Elements of probability theory]*: Moscow: Science: 256 p. [in Russian].
5. Khoroshilov V. S., Kobeleva N. N., Sycheva N. V. (2019) Mathematical modeling of the high-rise buildings deformation development process in Moscow (Vosstania square) *IOP Conf. Series: Materials Science and Engineering*. 698 , 044004. DOI 10.1088/1757-899X/698/4/044004. EDN QWILTL.
6. Khoroshilov, V. S., Kvashenko, I. Yu. & Noskov, M.F. (2013). *Osobennosti vybora deformatsionnykh marok dlya postroeniya kinematicheskoy modeli pri izuchenii deformatsiy sooruzheniy [Features of the choice of deformation marks for building a kinematic model when studying the deformations of structures]*. *Izvestiya vuzov. Geodeziya i aerofotos"emka [Izvestiya Vuzov. Geodesy and Aerophotography]*, 4/S, 58–61 EDN UIYCCD [in Russian].
7. Gmurman, V. E. (2003). *Teoriya veroyatnostey i matematicheskaya statistika [Probability Theory and Mathematical Statistics]*: Moscow: Graduate School: 523 p. ISBN 5-06-004214-6. EDN QJLKXP [in Russian].
8. Nikolaev, S. A. (1983). *Statisticheskie issledovaniya osadok inzhenernykh sooruzheniy [Statistical studies of settlement of engineering structures]*: Moscow: Bosom: 112 p. [in Russian].
9. Scheffe, G. (1963). *Dispersionnyy analiz [Dispersion analysis]*. (B. A. Sevast'yanov and V. P. Chistyakov, Trans). Moscow: Publishing house "Physical and Mathematical Literature". 625 p. [in Russian].
10. Tsitovich, N. A. Ed. (1967). *Prognoz skorosti osadok osnovaniy sooruzheniy [Forecast of the rate of settlement of foundations of structures]*. Moscow: Stroyizdat: 239 p. [in Russian].

#### Author details

Valery S. Khoroshilov – D. Sc., Associate Professor, Professor of the Department of Space and Physical Geodesy.

Natalia N. Kobeleva – Ph. D., Associate Professor, Department of Space and Physical Geodesy.

Received 18.12.2024

© V. S. Khoroshilov, N. N. Kobeleva, 2025