

Статистическая устойчивость результатов ретроспективных исследований на основе геохронологического трекинга

Я. А. Ивакин

Санкт-Петербургский Федеральный исследовательский
центр Российской академии наук
yan_a_ivakin@mail.ru

С. Н. Потапычев

АО «Концерн «Океанприбор»
potapychev@mail.ru

Аннотация. Геохронологический трекинг получил широкое признание как соответствующий научно-методический инструментальный и эффективная информационная технология ретроспективных исследований в интересах обоснования и рационализации маршрутных сетей транспорта, логистики перевозок, анализа фактов миграции населения и перемещений отдельных исторических личностей и пр. На базе геохронотрекинга разработана процедура статистической проверки исследовательских гипотез об устойчивых тенденциях в развитии различных пространственно-временных процессов. Надежность и достоверность принятия той или иной гипотезы в рамках ретроспективного исследования определяется представительностью (репрезентативностью) объема исходных данных о географических перемещениях, рассматриваемых как выборка из генеральной совокупности. Статистическая значимость (устойчивость) результатов ретроспективного исследования на основе геохронологического трекинга зависит от достаточности учтенных исходных данных о перемещениях исследуемых объектов. Анализ указанной зависимости и выработке алгоритма оценки указанной устойчивости (значимости) посвящен данный доклад.

Ключевые слова: географические информационные системы; ГИС-технологии для ретроспективных исследований; геохронологический трек и трекинг; изоморфизм графов; рациональный алгоритм; междисциплинарные исследования на базе ГИС; статистическая устойчивость выводов

Ретроспективный статистически-значимый анализ перемещений в географическом пространстве есть база для принятия решений по организации различных пространственно-временных систем. Геохронологический трекинг получил широкое признание как соответствующий научно-методический инструментальный и эффективная информационная технология ретроспективных исследований в интересах обоснования и рационализации маршрутных сетей транспорта, логистики перевозок, анализа фактов миграции населения и перемещений отдельных исторических личностей и пр. Основные принципы, процедуры и алгоритмы геохронологического трекинга описаны в [1–3]. Его математическая сущность сводится к поиску и оценке статистической значимости изоморфизма соответствующих графов: итоговый граф

геохронотрекинга представляется как граф-базис в структуре которого выявляется подграф изоморфный заданному, т.е. устанавливается наличие взаимно однозначного отображения одного графа на подграф другого, при котором сохраняется отношение инцидентности [2]. Граф, на изоморфность к которому в составе базового графа геохронологического трекинга определяется подграф, топологически описывает ту или иную определенную гипотезу исследования об устойчивой особенности в перемещениях исторических личностей, объектов или других сущностей в географическом пространстве. Далее определяется степень устойчивости в признании гипотезы исследования о выявляемой особенности в перемещениях с использованием статистического аппарата доверительной вероятности и доверительных интервалов [3].

Вместе с тем, надежность и достоверность принятия той или иной гипотезы в рамках ретроспективного исследования определяется представительностью (репрезентативностью) объема исходных данных о географических перемещениях, рассматриваемых как выборка из генеральной совокупности. Статистическая значимость (устойчивость) результатов ретроспективного исследования на основе геохронологического трекинга зависит от достаточности учтенных исходных данных о перемещениях исследуемых объектов. Иными словами, для принятия исследовательских гипотез ретроспективного исследования на базе геохронотрекинга с заданной доверительной вероятностью должно быть обеспечено необходимое и достаточное (релевантное) число учтенных единичных геопространственных перемещений, рассматриваемых как единичные статистические испытания. Обоснованная выработка математико-статистического аппарата и методики увязывания доверительной вероятности принятия гипотез исследований на базе геохронотрекинга с исходным числом учитываемых перемещений составляет существо обеспечения статистической устойчивости (значимости) выводов указанных исследований.

Разработка и обоснование математико-статистического аппарата и методики определения необходимого и достаточного (релевантного) числа разовых испытаний в ходе ретроспективных исследований на базе геохронотрекинга для обеспечения требуемого уровня

Работа выполнена при поддержке РФФИ (проект №19-07-00006)

доверия к итоговым результатам осуществлено путем последовательной реализации следующих логических шагов:

- теоретическая разработка и адаптация к условиям исследования математико-статистических основ определения необходимых и достаточных объемов выборки из генеральной совокупности данных о единичных перемещениях в географическом пространстве для обеспечения заданного значения доверительной вероятности получаемых выводов проводимого ретроспективного исследования;
- интерпретация выделенного математико-статистического аппарата, как аппарата обеспечения требуемого уровня надежности получаемых выводов проводимого ретроспективного исследования, применительно к подходам и моделям геохронотрекинга;
- конкретизация и описательное представление алгоритма расчета релевантного числа учитываемых перемещений объектов – единичных испытаний в ходе ретроспективных исследований методом геохронотрекинга для обеспечения приемлемого уровня рисков при принятии итоговых решений.

Детализация существа указанных шагов позволяет раскрыть существо методики определения необходимого и достаточного числа разовых испытаний в ходе ретроспективных исследований на основе геохронотрекинга для обеспечения требуемого уровня доверия к результатам и выводам исследования, в целом.

В рамках постановки ретроспективных исследований на базе геохронотрекинга в работах [4,5] в качестве генеральной совокупности данных рассматривается теоретическое число выборочных значений учитываемых перемещений рассматриваемых объектов, обеспечивающее доверительную вероятность принятия решений о выводе частного исследования равным 1. Очевидно, что размер генеральной совокупности, при данной постановке, теоретически не ограничен, однако на практике объем данных положенных в основу геохронотрека всегда конечен и ограничен. Тогда, частная задача определения релевантного (т. е. необходимого и достаточного) объема выборки из генеральной совокупности данных о перемещениях в географическом пространстве за заданный промежуток времени для обеспечения заданного уровня доверия к выводам исследования сводится к математическому увязыванию значения доверительной вероятности правильного вывода в оценке исследуемого параметра с таким числом единичных испытаний, которое обеспечивает оценку требуемого доверительного интервала в разбросе искомого параметра. В своей информационно-логической сущности данная задача детально рассмотрена и теоретически решена в работах, посвященных информационным технологиям компьютерного моделирования, трудах по теории вероятности и прикладной статистике. Например, в таких как [6–10]. Существо теоретического решения данной задачи заключается в построении (оценке)

доверительного интервала для исследуемого параметра, обусловленного заранее заданной доверительной вероятностью, как некоторого двумерного функционала, определяемого разницей между теоретическим значением искомого параметра и выборочным значением накопленной статистики его оценок.

В работах [7, 8, 10] решение указанной частной математической задачи сведено к построению конкретизированных эллипсов рассеивания выборочных значений исследуемых параметров применительно к назначаемым величинам доверительной вероятности. Очевидно, что в условиях геохронологического трекинга, т. е. в условиях дискретного прироста объема значений данных по итогам разовых испытаний, а также объективной ограниченности генеральной совокупности в силу конечности учитываемых данных о географических перемещениях анализируемых объектов, сводная теоретическая картина решения указанной частной математической задачи может быть преобразована к дискретному варианту представления. В частности, соотношение числа r выбросов в экспериментально получаемых значениях параметра вне обоснованного эллипса среднеекватрического отклонения к общему числу единичных испытаний (наблюдений) эксперимента N в рамках эксперимента позволяет формализовать решение указанной частной задачи (для дискретного варианта) в виде, показанном на рисунке.

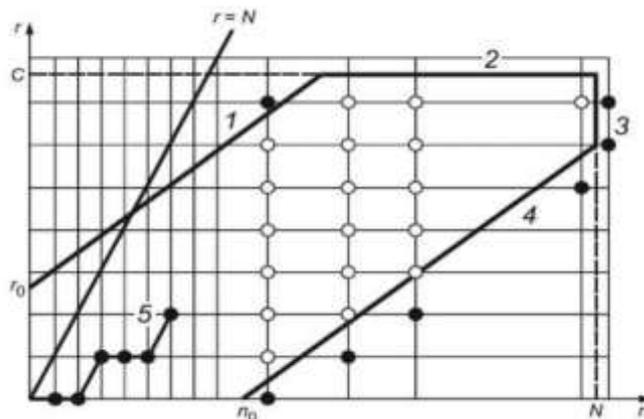


Рис. 1. Интерпретация наблюдаемых выборочных значений исследуемого параметра при накоплении его статистики в ходе геохронотрекинга

Так, на рисунке фигура, ограниченная линиями 1-2-3-4, является дискретной интерпретацией эллипса среднеекватрического отклонения, определяемого доверительным интервалом и интервалом допуска, соответствующих задаваемой доверительной вероятности правильного принятия решения в ретроспективном исследовании на базе геохронотрекинга. Соответственно, линии 1 и 2 обозначают границы зоны, за которыми регистрируемые экспериментальные значения параметра нельзя считать соответствующими теоретическому значению. Линии 3 и 4 обозначают границы зоны избыточного числа единичных испытаний при обоснованном принятии решения с заданной

доверительной вероятностью, что принимаемые значения параметра геохронотрека следует считать соответствующими теоретическому значению. Линия 5 обозначает дискретный процесс накопительного учета выбросов в экспериментально получаемых значениях исследуемого параметра вне обоснованного эллипса среднеквадратического отклонения при геохронотрекинге в зависимости от текущего суммарного числа одиночных испытаний (наблюдений) в ходе исследования. При этом, очевидно, что линия дискретного учета выбросов 5 в экспериментально получаемых значениях исследуемого параметра не может попадать в область, расположенную на рисунке выше границы $r = N$.

Приведенная на рисунке 1 интерпретация решения логико-математической задачи определения релевантного (т. е. необходимого и достаточного) объема выборки из генеральной совокупности данных многократных испытаний – перемещений для обеспечения заданного уровня доверия к выводам ретроспективного исследования позволяет рассматривать указанный объем выборки как ключевой показатель обеспечения требуемой надежности выводов проводимого исследования. То есть, релевантным является такое не избыточное число единичных испытаний (учитываемых перемещений в географическом пространстве), которое обеспечивает заданную, в виде доверительной вероятности, надежность результатов проведения ретроспективных исследований. Именно такое понимание задачи определения релевантного объема выборки из генеральной совокупности данных геохронотрекинга для обеспечения заданного уровня доверия к выводам исследования позволило интерпретировать её в рамках подходов к определению надежности статистических выводов в рамках эксперимента на базе ГИС.

В общенаучном смысле, надежность выводов исследования – это свойство такого объекта инфосферы, как вновь полученное знание, устойчиво и неизменно вырабатывать функционально пригодные и достоверные результаты при заданных начальных данных и входных условиях. Применительно к условиям проведения ретроспективных исследований на базе геохронотрекинга показателем с количественной мерой для оценки указанной надежности выступает доверительная вероятность истинности результатов частного испытания, исследования. При этом доверительная вероятность истинности результатов задается априорно и обеспечивается в ходе проведения ретроспективных исследований путем проведения релевантного (прежде всего, достаточного) числа элементарных испытаний.

Вышеописанный вариант рассмотрения статистической сущности проводимого исследования позволил интерпретировать её в рамках стандартизированного аппарата обеспечения и расчета показателей надежности в технике. Указанный аппарат разработан, апробирован и рекомендован к применению в рамках действующей национальной системы нормативно-технического регулирования. Он применим к предметной области ретроспективных исследований на базе геохронотрекинга в ГИС.

Решение задачи достижения необходимой статистической значимости (устойчивости) результатов ретроспективного исследования на основе геохронологического трекинга заключается в обеспечении достаточности учтенных исходных данных о перемещениях исследуемых объектов при построении соответствующего геохронотрека. При решении указанной задачи введена понятная и традиционная мера указанной значимости (устойчивости) результатов ретроспективного исследования в виде доверительной вероятности. Для различных градаций указанной вероятности и уровня риска в её принятии определен объем (релевантное количество) учитываемых перемещений объектов или артефактов в составе геохронотрека, рассматриваемый как суммарное число единичных испытаний, которое должно быть обеспечено при синтезе соответствующего указанного трека. При неперевышении выявленного соотношения указанного объема и числа фактов отклонений тех или иных единичных перемещений, гипотеза ретроспективного исследования принимается с искомой доверительной вероятностью.

Граничные условия для полученного решения задачи достижения статистической устойчивости результатов ретроспективного исследования на основе геохронологического трекинга определены как границы применимости приложений теории вероятности и математической статистики.

Дальнейшие направления совершенствования методики определения необходимого и достаточного (релевантного) числа разовых испытаний в ходе ретроспективных исследований на основе геохронотрекинга для обеспечения требуемого уровня доверия к результатам и выводам исследования связаны с её алгоритмизацией и автоматизацией, интеграцией в состав современных геоинформационных систем, ориентированных на прикладные исследования и решение пространственно-временных, аналитических задач в смежных областях.

Таким образом, приведенный вариант методики решения задачи достижения статистической устойчивости результатов ретроспективного исследования на основе геохронологического трекинга позволяет обеспечить и значительно расширить применимость научно-методического аппарата геохронологического трекинга на новые классы приложений. В свою очередь, данный факт позволяет расширить применимость математического аппарата проверки гипотез ретроспективных исследований на основе геохронологического трекинга для различных предметных областей и новых объектов изучения, добиться более эффективной его интеграции в соответствующие программные приложения для геоинформационных систем.

Разработка математико-статистического аппарата и методики определения необходимого и достаточного (релевантного) числа разовых испытаний в ходе ретроспективных исследований на базе геохронотрекинга позволит развить соответствующий научно-методический инструментарий и вытекающие из него информационные технологии ретроспективных и исторических

геопространственных исследований в интересах инженерии, логистики, а также гуманитарных наук. Также очевидна перспективность работ по развитию прикладной алгоритмики геохронотрекинга, как такового. К указанным работам следует отнести внедрение и интеграцию соответствующих информационных технологий искусственной интеллектуальности, интеграции и слияния информации, виртуализации и пр.

СПИСОК ЛИТЕРАТУРЫ

- [1] Рейхсфельд В.О., Еркова Л.Н. Оборудование производств основного органического синтеза и синтетического каучука. Киев: Наукова думка, 1984. 346 с. Ивакин Я. А. Рациональный алгоритм проверки гипотез ретроспективных исследований использования водного транспорта на базе геохронологического трекинга/ Я.А.Ивакин, С.Н. Потапычев, Р.Я. Ивакин // Вестник Государственного университета морского и речного флота имени адмирала С.О. Макарова. 2019. Т. 11. No 3. С. 448–460. DOI: 10.21821/2309-5180-2019-11-3-448-460.
- [2] Ивакин Я.А., Потапычев С.Н. Геохронологический трекинг – специализированный ГИС-инструментарий исторического исследования // Историческая информатика. Информационные технологии и математические методы в исторических исследованиях и образовании, 2016; № 1-2, с. 3–11.
- [3] Ивакин Я.А., Потапычев С.В. Информационная технология геохронологического трекинга для проверки гипотез ретроспективных исследований использования водного транспорта // Вестник Государственного университета морского и речного флота имени адмирала С. О. Макарова. 018. Т. 10. № 2. С. 452–461. DOI: 10.21821/2309-5180-2018-10-2-452-461.
- [4] Ивакин Р.Я., Ивакин Я.А., Потапычев С.Н. Оптимизированный алгоритм статистической проверки гипотез ретроспективных исследований на основе геохронологического трекинга // Труды учебных заведений связи. 2020. Т. 6. № 1. С. 86–93. DOI:10.31854/1813-324X-2020-6-1-86-93
- [5] Ивакин Я.А., Потапычев С.Н. Информационная технология исследований особенностей применения изделий гидроакустической техники на основе геохронологического трекинга // Информационные технологии и телекоммуникации. 2020. Том 8. № 2. С. 109–119. DOI 10.31854/2307-1303-2020-8-2-109-119.
- [6] Codescu M., Horsinka G., Kutz O., Mossakowski T., Rau R. DO-ROAM: Activity-Oriented Search and Transport Navigation with Open Street Map/ GeoSpatial Semantics // Proceedings of the 6th International Conference, GeoS 2015. 2015. Pp. 88-108.
- [7] Sigma Knowledge Engineering Environment [Электронный ресурс] - электронные данные, – режим доступа URL: <http://sigmakee.sourceforge.net>. Дата доступа: январь 2020 г.
- [8] Советов Б.Я., Яковлев С.А. Моделирование систем, 7-е изд. Москва: Издательство Юрайт, 2019. 343 с.
- [9] Юсупов Р.М., Заболотский В.П. Концептуальные и научно-методологические основы информатизации. СПб.: Наука, 2009. 541 с.
- [10] Советов Б.Я., Цехановский В.В. Информационные технологии. 6-е изд., перераб. Москва: Издательство Юрайт, 2016. 263 с.