

И.В. Чечушков¹, А.В. Епимахов²

I.V. Chechushkov, A.V. Epimakhov

¹Институт истории и археологии УрО РАН, г. Екатеринбург

²Южно-Уральский государственный университет, г. Челябинск,

epimakhovav@susu.ru

Варианты исследования мобильности по данным изотопии стронция (анализ фоновых значений)

Mobility study options based on strontium isotope data (analysis of background values)

Работа рассматривает некоторые методы анализа лабораторных данных соотношения $^{87}\text{Sr}/^{86}\text{Sr}$, включая метод базисных интервалов и основные методы картографирования вариативности, а также ограниченность их применения. Метод базисных интервалов заключается в вычислении среднего и интервала в две сигмы на основании образцов из геологической зоны. Недостатком метода является необходимость получения большой выборки и отказ от изучения вариативности. Метод картографирования доменов заключается в нанесении средних значений для зоны на карту с цветовой кодификацией. Диаграмма Вороного позволяет ограничить зоны различных значений и демонстрирует гетерогенность выборки. Геостатистические методы интерполяции позволяют произвести вычисления неизвестных числовых значений по известным данным в силу пространственной автокорреляции. Интерполированные модели имеют тенденцию к сглаживанию данных, маскируя неоднородность. Наконец, машинное обучение построено на алгоритме регрессии случайного ветвления, заключающегося в том, что входные данные проходят через каждое дерево решений, а выходные данные всех деревьев усредняются.

This paper discusses some methods for analyzing laboratory data on the ratio of $^{87}\text{Sr}/^{86}\text{Sr}$, including the baseline method, the primary variability mapping methods, and their limitations. The baseline is the mean and two-sigma interval based on samples from the geological zone. The disadvantage of the method is the need to obtain a large sample and reject the variability study. The method of domain mapping is to plot the mean values for a zone on a color-coded map. The Voronoi map allows to limit the zones of different values and demonstrate the sample heterogeneity. Geostatistical interpolation methods calculate unknown values from known data due to spatial autocorrelation. Interpolated models tend to smooth out the data, masking inhomogeneity. Finally, machine learning is built on a random forest regression algorithm. The input is passed through each decision tree, and the output of all the trees is averaged.

Исследование мобильности (в отличие от миграций) сравнительно недавно вошло в практику изучения археологических материалов России, в основном благодаря внедрению методов геохимии. В данной работе мы обратимся к вариантам анализа изотопии стронция, хотя спектр методов и элементов значительно шире [Somerville, Beasley, 2023; Price, 2023]. Причиной сужения тематики обзора является сравнительно разнообразное использование именно этой части геохимических методов. Кроме того, наметилась некоторая разница в подходах к формированию стратегии такого рода исследований, в том числе и в принципах сопоставления фоновых значений с археобиологическими материалами. Принципы отбора и картирования прямо сопрягаются с решаемыми задачами, а также характером итоговых заключений и возможностью их проверки [Slovak, Paytan, 2012; Holt et al., 2021]. По нашему мнению, накопленный опыт требует обобщения, что и является основной целью данной работы. Мы надеемся, что это будет полезно исследователям, планирующим подобные проекты, и поможет четче осознать возможности и ограничения каждой из стратегий.

Типы образцов для измерения фоновых значений. Любые варианты исследования мобильности на основе соотношения $^{87}\text{Sr}/^{86}\text{Sr}$ построены на сравнении результатов измерения археобиологических материалов с известными локальными (фоновыми) значениями. Практика реализации совместных проектов археологов и геохимиков в разных частях света показала, что даже при наличии национальных справочных карт о составе изотопов их прямое использование имеет ограниченный потенциал в решении археологических проблем [West et al., 2009].

Так или иначе приходится обращаться к отбору и анализу фоновых образцов. В качестве последних используются разные источники: вода, почва, растительность, слабо мигрирующие организмы и пр. Каждый из них потенциально имеет проблемы искажения (атмосферная влага, золотой перенос, современное антропогенное воздействие и т.д.) Ученые склоняются к заключению, что вследствие большой массы элемента изотопы стронция не фракционируются, т.е. локальный органический материал способен обеспечить надежные заключения об окружающей среде. Главной проблемой в этом варианте остается обеспечение именно локального характера фонового образца (животных или растений). Для аридной зоны под большим вопросом использование крупных мигрирующих копытных (поэтому чаще речь идет о мелких); не меньшие, хотя и разные, проблемы могут быть связаны с анализом костей домашних животных прошлого и современности. Так, наш опыт реализации проекта РФ «Миграции человеческих коллективов и индивидуальная мобильность в рамках мультидисциплинарного анализа археологической информации (бронзовый век Южного Урала)» показал, что невозможно гарантировать «чистоту» отбора не столько в связи с перемещением современных домашних животных, сколько в свете широкого использования разнотипных кормов и добавок, которые потенциально могут являться источником привнесенного стронция. При этом, даже сигнал по образцам животных в условиях свободного выпаса будет отображать некое усреднение значений, впрочем, как и древние образцы. Ясно, что мы имеем дело с сигналом, формирующимся в течение месяцев или лет, который суммирует разные типы потребляемых продуктов.

Еще одна версия формирования и анализа выборки фоновых образцов для интерполяции карт вариативности – вычисление усредненных значений разнотипных образцов. Учитывая перечисленные проблемы для большинства типов, такой вариант, по нашему мнению, никак не добавляет определенности в заключениях. В этой связи мы считаем полезным ориентироваться на разные источники, которые анализируются и обобщаются по отдельности с сопоставлением уже итоговых результатов.

Пространство отбора и моделирование карты фоновых значений. Несколько упрощая ситуацию, можно выделить два варианта стратегии: «снизу-вверх» и «сверху вниз» [Price, 2023, p. 254]. Иными словами, мы можем исходить из материалов конкретного объекта и их сравнения с единичными локальными образцами либо из необходимости предварительного построения карты фоновых значений для некой территории (с возможностью наложения данных по разным объектам в дальнейшем). Выбор стратегии прямо зависит от поставленных задач и возможностей исследователя, определяя весь дизайн проекта.

Нетрудно догадаться, что первый сценарий реализуется гораздо чаще, в том числе и для территории Урала [Анкушева и др., 2021; Киселева и др., 2021; Корочкова и др., 2023; Shishlina et al., 2020 и др.]. В результате выборка археобиологических значений оценивается с точки зрения однородности/неоднородности, а полученные группы (или группа, при отсутствии значимых различий в выборке) сопоставляются с серией фоновых образцов, связанных с локусом. Этот вариант исследований при относительной

доступности имеет заметные ограничения в информативности. По сути, это попытка ответить на вопрос: «Местный или неместный индивид (особь)?» Однако, даже этот вариант требует серий измерений для снижения риска влияния случайностей выборки. Ограничения в этом вопросе могут быть связаны не только с дороговизной анализов, но и наличием образцов в коллекции. Процедура отбора фоновых образцов подчиняется самым общим рекомендациям, но оценить статистически достоверность выводов невозможно.

В период становления методики полученные значения изотопных измерений археобиологических объектов нередко сопоставлялись с литологией подстилающих пород, для которых величины соотношения $^{87}\text{Sr}/^{86}\text{Sr}$ получены аналитическими методами либо для них построены прогностические модели. Сегодня такая практика признана неэффективной [Holt et al., 2021]. В этой связи неизбежно встает вопрос о вычислении местных или региональных базисных интервалов и карт на основе эмпирических данных. Методы построения существенно разнятся, поэтому на этапе планирования необходимо определить состав фоновых образцов и методов выбора локусов пробоотбора; осуществить выбор метода расчета базисных интервалов и/или интерполяции карт вариативности, а также определить способы верификации результатов; и разработать методы сопоставления фоновых и археобиологических измерений. Ответы на эти вопросы должны быть представлены и в ходе презентации результатов.

Одним из наиболее часто применяемых является метод расчета базисного интервала (линии; англ. baseline), т.е. локального значения соотношений изотопов стронция на основании отбора значительного количества образцов из нескольких локусов в пределах геологической зоны. В идеале, образцы для геоархеологического исследования должны состоять из элементов диеты древнего населения и животных, включая воду. Почва также широко используется в качестве источника, однако в литературе обозначены различия в результатах в зависимости от метода получения вытяжки [Frank et al., 2022]. Среднее значение рассчитывается для каждого типа образца или для всей выборки. В качестве доверительного интервала рекомендуется использовать значение в две сигмы [Ladegaard-Pedersen et al., 2021]. Основанный таким образом интервал в две сигмы ($\pm 2\sigma$; около 95 % измерений) вокруг среднего используется для определения принадлежности археологического образца к геологической зоне. Очевидным недостатком этого метода является необходимость получения достаточно большой выборки для каждой геологической зоны для уменьшения влияния случайных искажений ($n \geq 40$ в классической статистике). Кроме того, метод не предполагает картирования вариативности, а полагается на соотношение археологических памятников с геологическими зонами.

Метод картографирования доменов (номинальный подход), сходен с методом базисных интервалов, с тем лишь отличием, что усредненные значения для литологической или географической зоны наносятся на карту с цветовой кодификацией [Evans et al., 2010].

Сходным подходом является построение диаграммы Вороного или полигонов Тиссена с отличием в том, что они не требуют усреднения значений. Оба подхода предполагают разбиение плоскости карты на зоны с образованием множества точек, более близких к одному из элементов выборки, чем к любому другому элементу выборки. Такие карты позволяют ограничить зоны различных значений исходя из избранного аналитического шага и, таким образом, демонстрируют гетерогенность выборки. Данный подход не позволяет продемонстрировать градиентного перехода между значениями [Adams et al., 2018].

Геостатистические методы интерполяции широко применяются в картографии, и в этом смысле интерполяция изоскейпов ничем не отличается от интерполяции карт на основании любых других данных. В самом очевидном виде, геостатистика представляет собой процесс интерполяции измеренных значений для местоположений, в которых измерения не проводились. Другими словами, это способ вычисления неизвестных числовых значений на карте чего-либо по известным данным. Предполагается, что неизвестные значения близки измеренным в силу пространственной автокорреляции – явления, обуславливающего высокую вероятность того, что расположенные рядом друг с другом феномены имеют схожие пространственные значения. Методы геостатистики исходят из аналогичной идеи – получения пространственных данных о неизвестных точках путем интерполяции данных об известных точках [Чечушков, 2015]. К их числу относятся следующие методы: ближайшего соседа, обратных взвешенных расстояний, обычного кригинга, Байесовского кригинга, и т.п. Важно отметить, что в случае изоскейпов предположение об автокорреляции может быть ошибочным, т.к. литологические зоны могут быть резко очерчены и соседствующие зоны могут иметь сильно различающиеся значения. Интерполированные методами геостатистики модели имеют тенденцию к сглаживанию данных, маскируя неоднородность, которая может существовать в биосфере [Holt et al., 2021]. По этой причине, важным является проверка степени гетерогенности выборки другими методами.

Все описанные выше методы картографирования доступны для использования как в коммерческом программном обеспечении (ArcGIS, Surfer), так и в свободно-распространяемых ГИС (QGIS).

Сравнительно новым методом построения модели вариативности стронция (и любых других элементов) является применение машинного обучения для комбинирования различных типов данных [Bataille et al., 2015]. В основе этого подхода лежит алгоритм регрессии случайного ветвления (англ. *random forest regression*). Он сочетает в себе концепции случайного ветвления с анализом регрессии для предсказания неизвестных значений. В модели регрессии случайного ветвления создается набор деревьев возможных решений. Каждое дерево решений строится путем выбора случайного подмножества обучающих данных и случайного подмножества признаков. Обучение включает независимое построение нескольких деревьев, а затем объединение их прогнозов. Прогностические модели на основе регрессионной модели случайного ветвления построены на том, что входные данные проходят через каждое дерево решений, а выходные данные всех деревьев усредняются. Это усреднение помогает уменьшить индивидуальные смещения деревьев и повысить общую точность прогнозирования. Преимущества машинного обучения заключаются в надежности получаемого прогноза; нелинейности (фиксация сложных нелинейных отношений; выявление функций, оказывающих наибольшее влияние на прогноз; масштабируемость [Breiman, 2001; Cutler et al., 2007]. Для построения таких моделей используются программные пакеты языка R.

Итог данной работы можно свести к нескольким тезисам. Поиски оптимальной методики исследования продолжают, часть специалистов скептически оценивает саму возможность сформировать универсальный алгоритм, т.к. велики влияние локальных факторов (состав и разнообразие коренных пород, характер растительности, степень антропогенного вмешательства в структуру ландшафта и многое другое) и вариативность решаемых проблем. Тем не менее, основные параметры проекта должны быть четко сформулированы уже на этапе планирования. Выбор стратегии картирования зависит от конкретных задач и возможностей получения выборок. Независимо от варианта стратегии максимизация количества и разнообразия фоновых образцов биодоступного строн-

ция существенно повышает не только достоверность сделанных заключений, но и возможность их проверки, включая статистические методы. В представлении результатов во всех случаях должны быть полноценно описаны процедуры и возможные ограничения информативности в интерпретации.

Работа выполнена при финансовой поддержке РФФ, проект № 20-18-00402-П «Миграции человеческих коллективов и индивидуальная мобильность в рамках мультидисциплинарного анализа археологической информации (бронзовый век Южного Урала)» (рук. Епимахов А.В.)

Литература

Анкушева П.С., Алаева И.П., Садыков С.А., Ын Ч.Я., Анкушев М.Н., Зазовская Э.П., Рассадников А.Ю. «Степные коридоры» алакульских скотоводов: результаты изотопных и палеоботанических исследований на поселении Чебаркуль III // Уральский исторический вестник. 2021. № 3(72). С. 26–38.

Киселева Д.В., Анкушева П.С., Анкушев М.Н., Окунева Т.Г., Шагалов Е.С., Касьянова А.В. Определение фоновых изотопных отношений биодоступного стронция для рудника бронзового века Новотемирский // Краткие сообщения Института археологии. 2021. № 263. С. 176–187.

Корочкова О.Н., Спиридонов И.А., Стефанов В.И. Погребение эпохи энеолита в горнолесном Зауралье // Археология, этнография и антропология Евразии. 2023. Т. 51. № 2. С. 57–65.

Чечушков И.В. Методы геостатистики в изучении поселенческих памятников бронзового века // Этнические взаимодействия на Южном Урале: мат. VI Всерос. науч. конф. Челябинск: ЧОКМ, 2015. С. 90–96.

Adams S., Grün R., McGahan D., Zhao Jian Xin, Feng Yuexing, Nguyen Ai, Willmes M., Quaresimin M., Lobsey B., Collard M., Westaway M. C. A strontium isoscape of north east Australia for human provenance and repatriation // Geoarchaeology. 2019. Vol. 34(3). P. 231–251.

Breiman L. Random forests // Machine learning. 2001. Vol. 45(1). P. 5–32.

Bataille C.P., von Holstein I.C.C., Laffoon J.E., Willmes M., Liu Xiao-Ming, Davies G.R. A bioavailable strontium isoscape for Western Europe: A machine learning approach // PloS One. 2018. Vol. 13(5). e0197386.

Cutler D.R., Edwards T.C., Beard K.H., Cutler A., Hess K.T., Gibson J., Lawler J.J. Random forests for classification in ecology // Ecology. 2007. Vol. 88(11). P. 2783–2792.

Evans J.A., Montgomery J., Wildman G., Boulton N. Spatial variations in biosphere $^{87}\text{Sr}/^{86}\text{Sr}$ in Britain // Journal of the geological society. 2010. Vol. 167(1). P. 1–4.

Holt E., Evans J.A., Madgwick R. Strontium ($^{87}\text{Sr}/^{86}\text{Sr}$) mapping: A critical review of methods and approaches // Earth-Science Reviews. 2021. Vol. 216. 103593.

Price T.D. Conclusions and Recommendations // T.D. Price (Ed.) Isotopic Proveniencing and Mobility. Interdisciplinary Contributions to Archaeology. Cham: Springer Nature Switzerland AG, 2023. P. 241–267.

Shishlina N., Roslyakova N., Kolev Yu., Bachura O.P., Kuznetsova O.V., Kiseleva D., Retivov V.M., Tereschenko E. Animals, metal and isotopes: Mikhailo-Ovsyanka I, the Late Bronze Age mining site of the steppe Volga region. Archaeological Research in Asia. 2020. Vol. 24. 100229.

Slovak N.M., Paytan A. Applications of Sr Isotopes in Archaeology // M. Baskaran, (Ed.) Handbook of Environmental Isotope Geochemistry. Advances in Isotope Geochemistry. Berlin, Heidelberg: Springer, 2012. P. 743–768.

Somerville A.D., Beasley M.M. Exploring Human Behavior Through Isotopic Analyses: Tools, Scales, and Questions // M.M. Beasley, A.D. Somerville (Eds.) Exploring Human Behavior Through Isotope Analysis. Applications in Archaeological Research. Cham: Springer Nature Switzerland AG, 2023. P. 9–32.

West J.B., Bowen G.J., Dawson T.E. & Tu K.P. (Eds.) Isoscapes: Understanding movement, pattern, and process on Earth through isotope mapping. Dordrecht, Heidelberg, London, NY: Springer, 2009. 495 p.