

МОДЕЛИРОВАНИЕ ПОВЕРХНОСТИ МОРСКОГО ДНА

Ландик А.С.¹

В данном исследовании предпринята попытка сравнения возможностей моделирования поверхностей с использованием языка статистики R относительно более традиционных и узкоспециализированных инструментов, что помогает понять как выбор инструмента влияет на процесс, скорость и результаты моделирования. Исследование основывается на упрощенном воспроизведении всех этапов моделирования и описания возможностей языка R на каждом из этапов (получение данных, обработка данных, исследование, моделирование, валидация результатов, публикация результатов).

В качестве показательных примеров использовались сгенерированные псевдослучайным алгоритмом матрицы высот (рис. 1) поверхности вулкана и дна озера. На этапе получения данных была продемонстрирована возможность использования большого количества популярных форматов (csv, excel, sas, rds, sql и др.), а так же использования языка запросов SQL для форматирования данных.

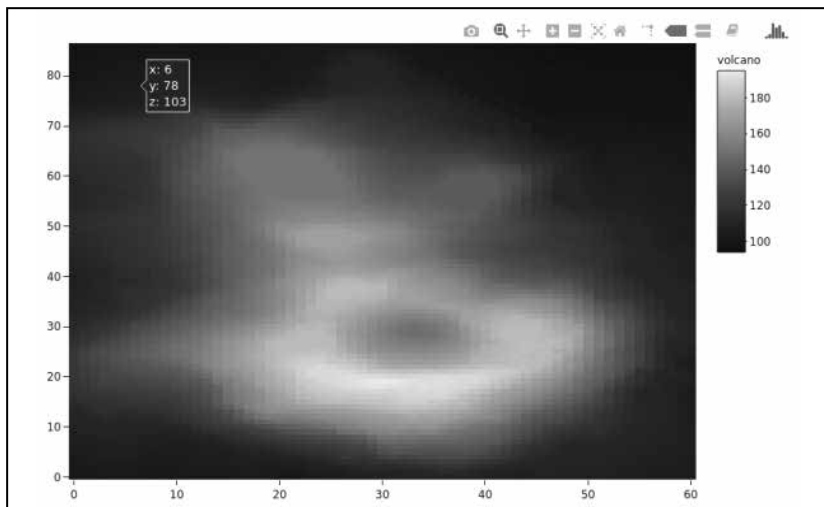


Рис. 1. Тепловая карта высот

¹ Ландик А.С. – МГУ им. Ломоносова, факультет биоинженерии и биоинформатики, Москва.

Этап обработки данных был освещен в виде работы с форматами данных `data.frame`, `data.table`, а также в бинарном формате. Была показана возможность быстрого преобразования и очистки данных для последующего исследования. Этап исследования представлен рассмотрением возможностей визуализации данных пакетами `ggplot2` и `plotly` для поиска, построения и проверки закономерностей и последующего построения модели. Этапы построения модели и визуализации были рассмотрены на примере генерации и моделирования поверхностей вулкана (рис. 2) и дна озера (рис. 3).

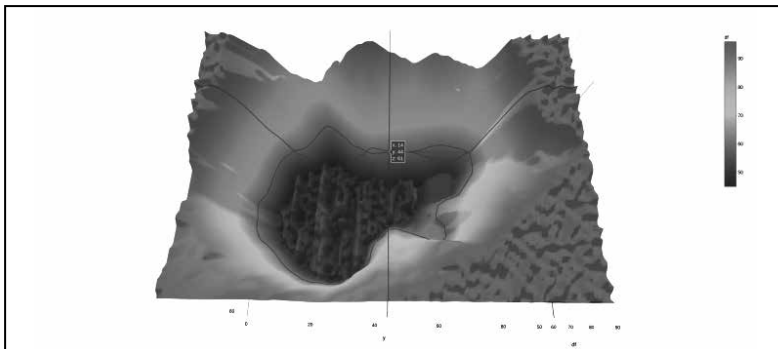


Рис. 2. Интерактивная 3D модель поверхности дна озера.
Цветовой градиент соответствует изменению высоты

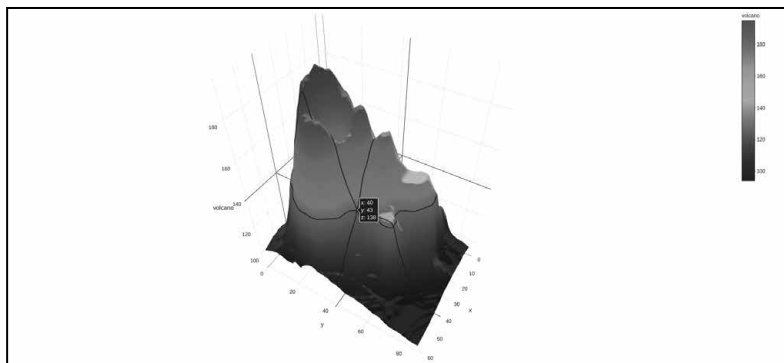


Рис. 3. Интерактивная 3D модель поверхности вулкана

Были показаны преимущества языка R в интерактивной работе с 3D изображением, отсутствие усложняющих работу не технического специалиста настроек сцены (проекция, перспектива, освещение,

кадрирование и др.), возможности быстрого внесения изменений в модель, возможности работы с несколькими отдельными моделями или несколькими субмоделями образующими слои единой модели (рис. 4). Возможности языка статистики R для проверки и оценки данных были показаны сравнением моделей построенных разными подходами, широкими методами статистики и отладки (поиска ошибок и выбросов). Получение и публикация результатов представлена использованием R Markdown (экспорт в веб, экспорт в популярные форматы: pdf, powerpoint, jpg, doc и др.) и R Notebook.

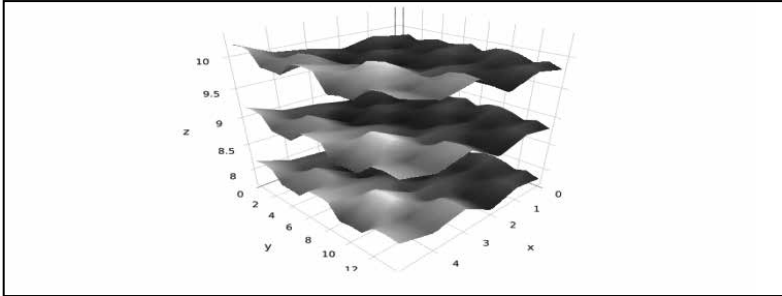


Рис. 4. Пример использования нескольких субмоделей на одной сцене для моделирования расслоения

Выводы

Язык R имеет низкий порог входа для не технического специалиста и широкий функционал для работы на всех этапах процесса моделирования. Наличие большого количества пакетов позволяют тонко настраивать инструментарий для каждого из этапов. Гибкость и скорость работы позволяют «на лету» проверять новые перспективные техники и гипотезы. С другой стороны, язык R имеет ряд ограничений относительно оптимизации работы с памятью, работы с масштабными проектами и надежности новых пакетов.

Литература

- [1] Maindonald J.H. and Braun W.J. 2003. Data Analysis and Graphics Using R – An Example-Based Approach. Cambridge University Press.
- [2] Шипунов А.Б., Балдин Е.М., Волкова П.А. и др.: Наглядная статистика. Используем R!, 2014, Электронное издание.
- [3] Роберт И. Кабаков: R в действии. Анализ и визуализация данных на языке R (перевод с английского), 2014, Электронное издание.