Навигация

[Пример. Озеро 1 2](#_Toc15911433)

[Описание данных 2](#_Toc15911434)

[График собственных значений (eigenvalue) 2](#_Toc15911435)

[График объяснённой дисперсии 3](#_Toc15911436)

[Факторный анализ 4](#_Toc15911437)

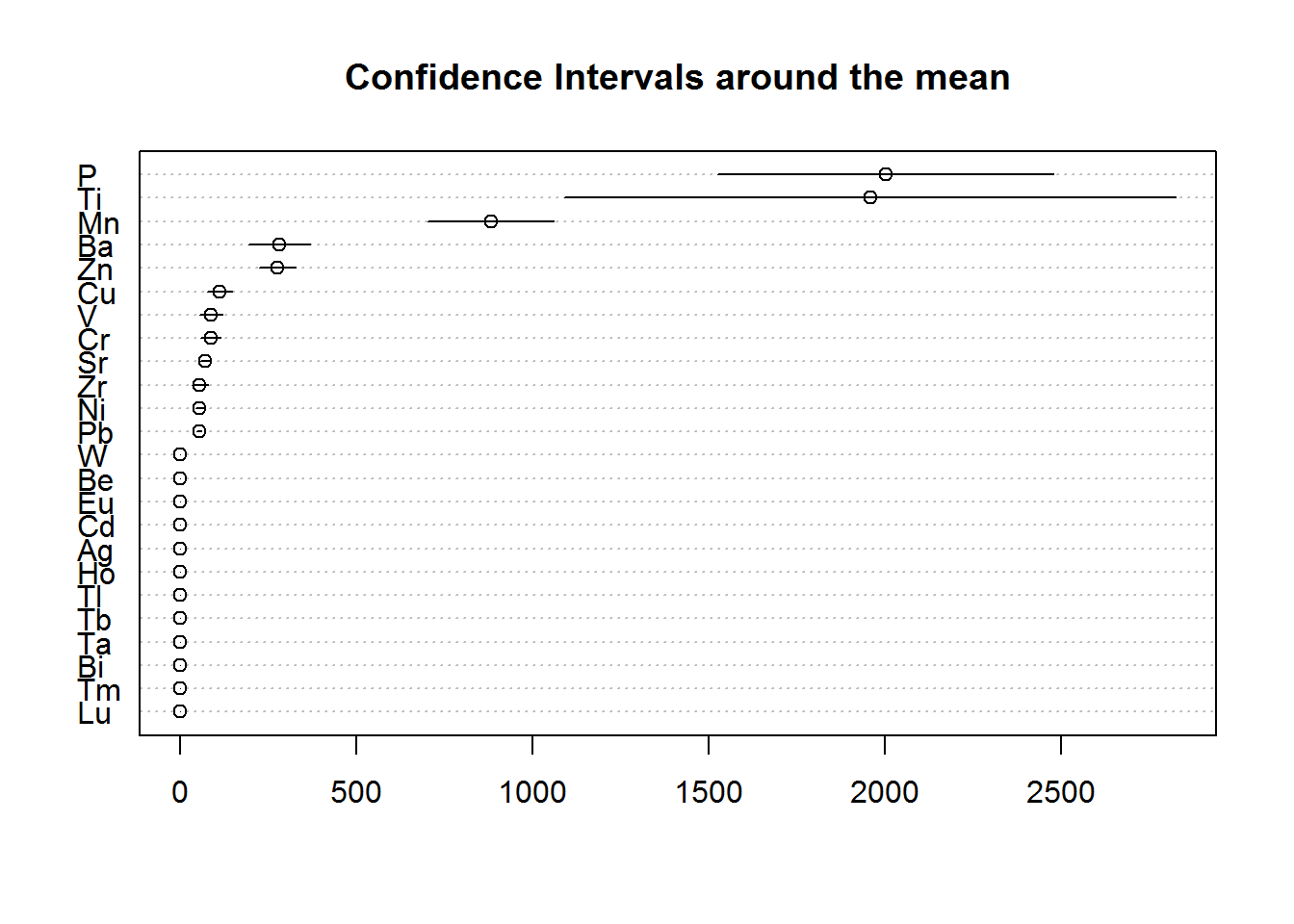
[Приложение 6](#_Toc15911438)

# Пример. Озеро 1

## Описание данных

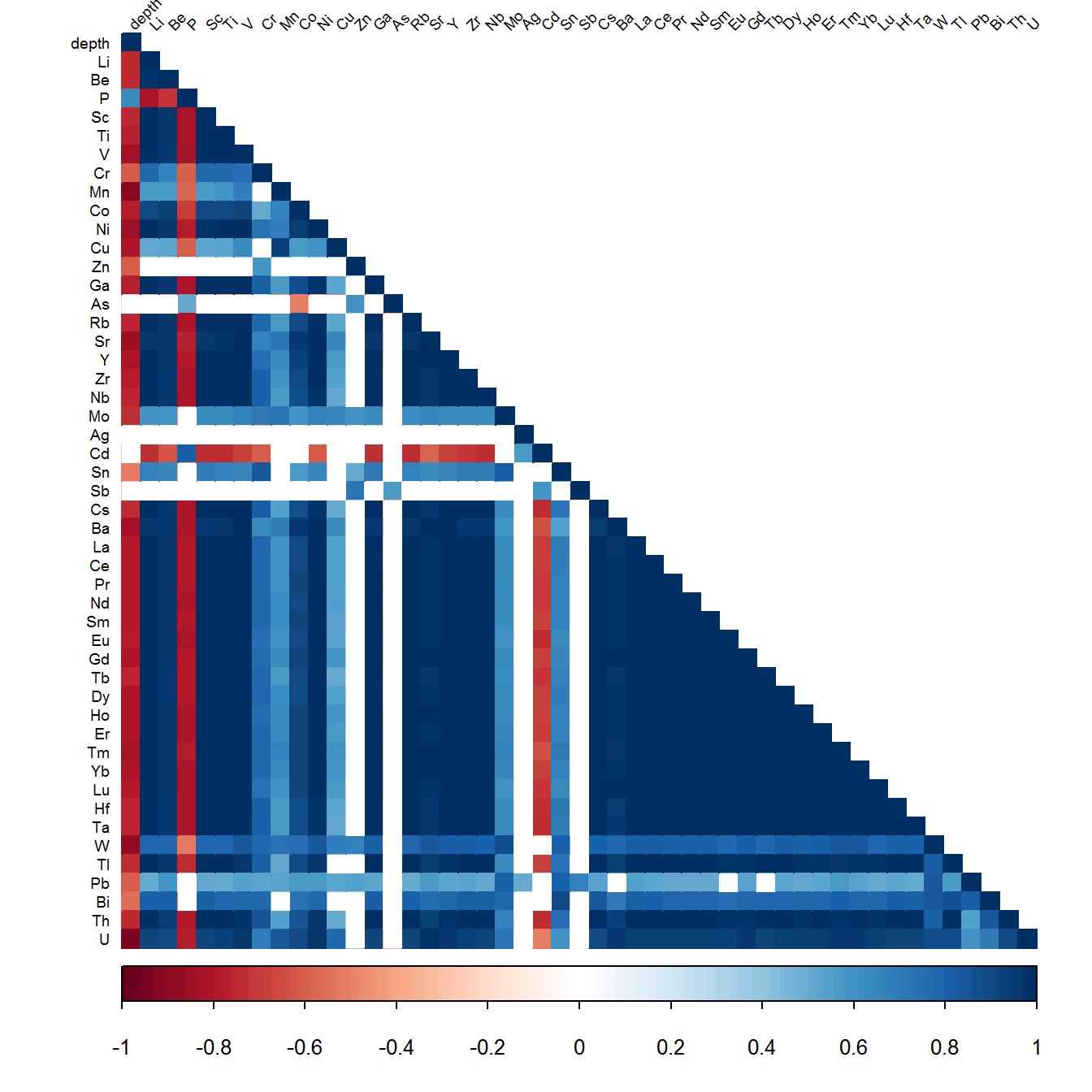
1. Разброс значений вокруг среднего для каждого хим. элемента.

Чем длиннее по горизонтальной оси – тем больше разброс значений.



1. Карта корреляций (закрашены парные корреляции сильнее 0,5).

Красное – отрицательная корреляция, синее – положительная.



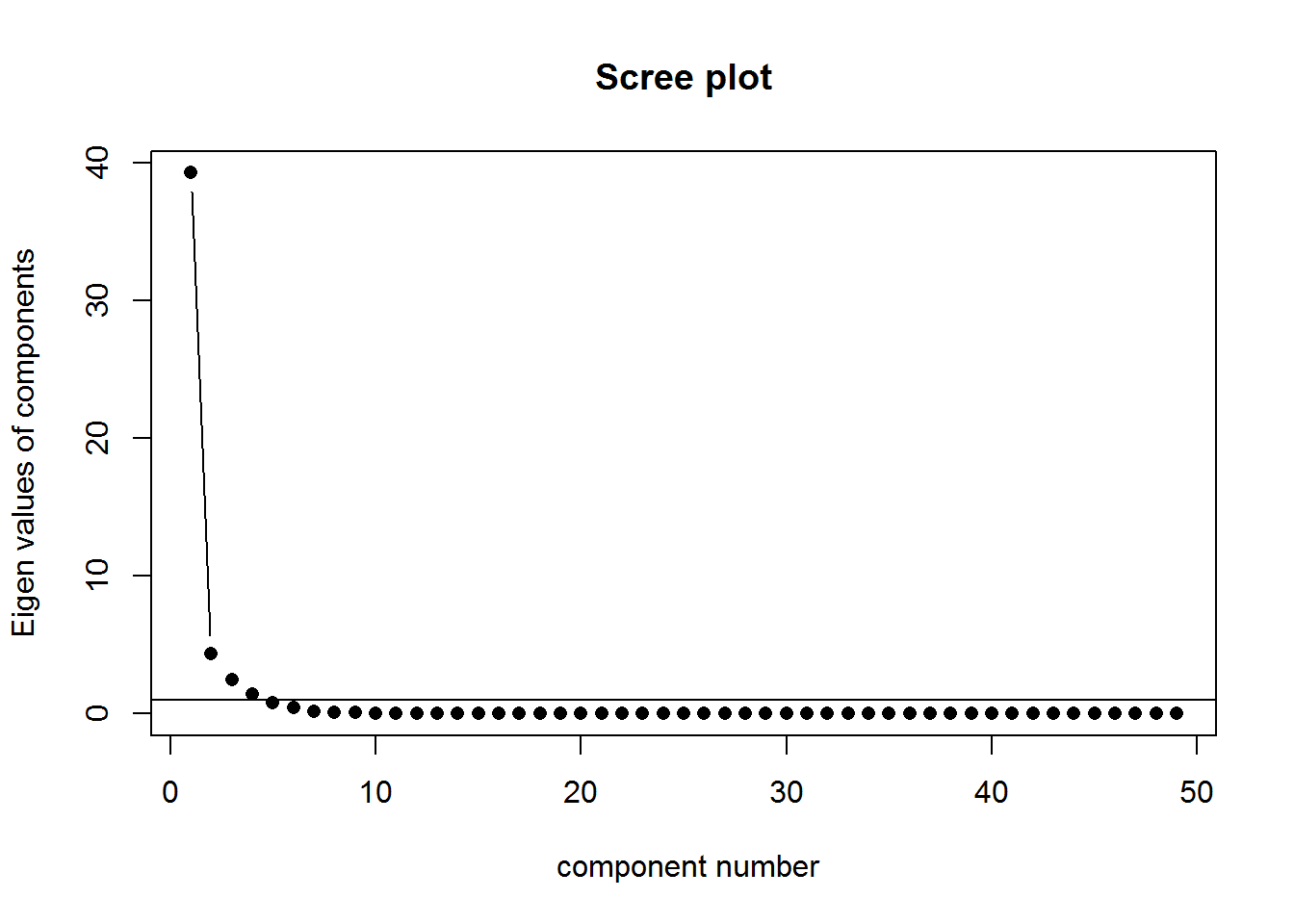
## График собственных значений (eigenvalue)

Автоматически итерационно (один за другим) строится факторный анализ (ФА) с количеством факторов от 1 до N (где N – это количество исходных хим элемеентов, столбцов). Для каждого полученного ФА берётся характеристика – собственное значение (eigenvalue рис1), либо дисперсия (рис2) – доля объяснённой вариации текущим фактором.

Проще говоря: шаг 1 – создаём фактор №1 (вместо имевшихся 49), с его помощью пытаемся описать имеющиеся у нас данные, оцениваем насколько хорошо получилось (получаем eigenvalue), отмечаем точкой на графике (х=1, у=40 означает, что созданный фактор №1 описывает около 40 исходных факторов). Шаг 2 – создаём дополнительный фактор №2 (теперь в ФА участвуют факторы №1 и №2), пытаемся описать исходные данные, получаем eigenvalue уже для фактора №2, отмечаем на графике (х=2, у=4 означает, что созданный фактор №2 описывает около 4 исходных факторов).

Горизонтальная линия на уровне у=1 показывает границу, выше которой каждый новый созданный фактор описывает больше 1 хим элемента (и его целесообразно включать в анализ), и ниже которой каждый новый созданный фактор описывает менее 1 хим элемента (и его не целесообразно включать). Носит скорее рекомендательный характер.

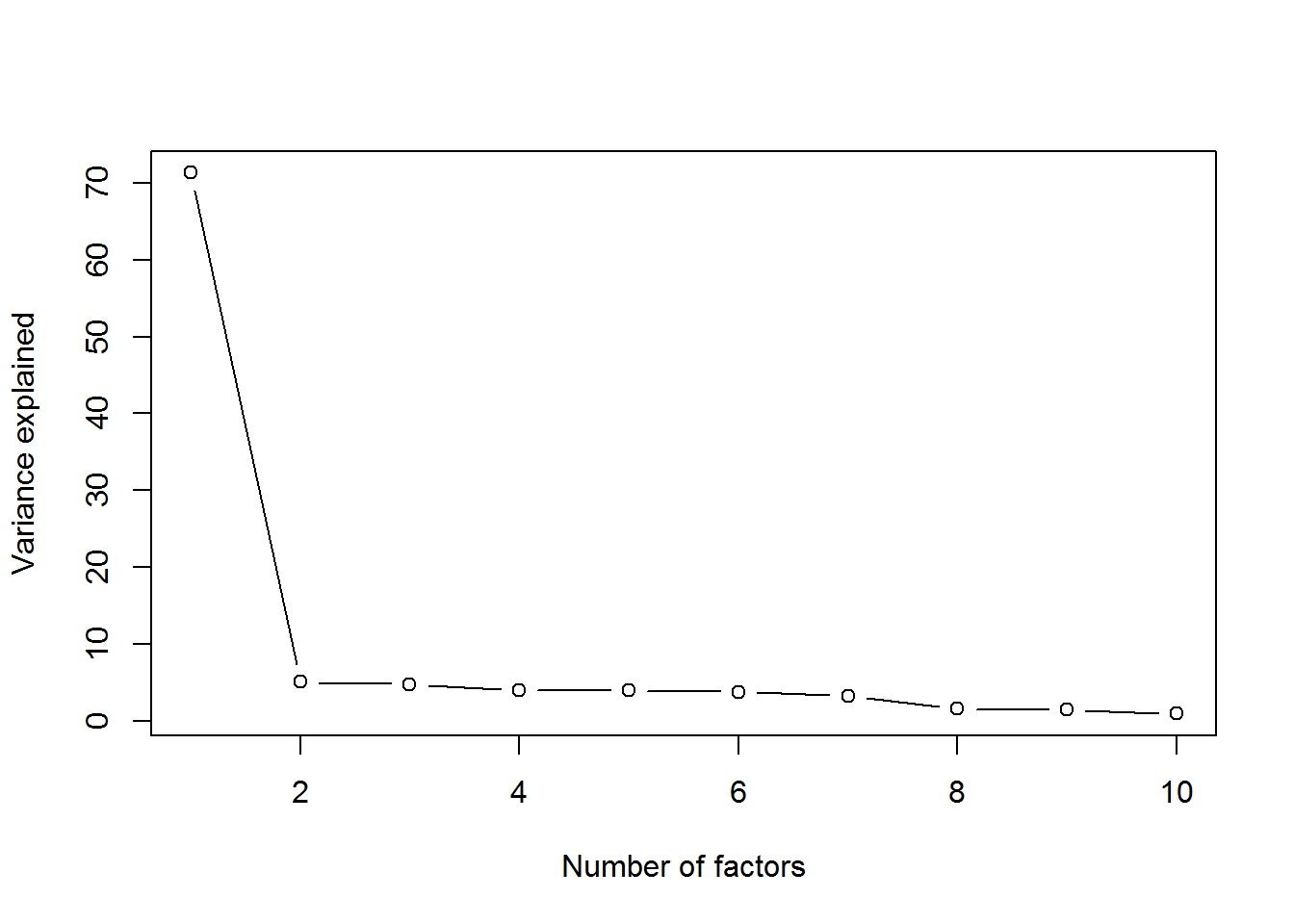
В нашем случае имеем, что eigenvalue для 4 факторов больше 1, то есть целесообразно включать в ФА до 4 факторов.



## График объяснённой дисперсии

График аналогичен предыдущему, но имеет иную детализацию – по горизонтальной оси рассматриваются ФА с количеством факторов до 10, по вертикальной указывается доля объяснённой дисперсии.

В нашем случае фактор №1 описывает 70% исходной информации, фактор №2 – 5% и тд. Так, видим, что факторы 2-6 объясняют примерно равные доли дисперсии, поэтому можем установить приемлемую для нас «планку» -- насколько точно мы хотим описать особенности (и, соответственно, насколько менее точно обобщить).



В итоге получаем несколько рекомендаций по выбору количества факторов:

* брать все факторы, eigenvalue которых больше 1 (или выше среднего);
* брать столько факторов, сколько могут объяснить заранее установленную «планку» дисперсии (например, >70%);
* по графику смотреть на резкое изменение наклона кривой (то есть когда наблюдается резкий спуск, например 1-2 факторы озера 1).

## Факторный анализ

Loadings – факторные нагрузки для каждого хим элемента (каким образом каждый фактор объединяет в себя исходные столбцы). Полная таблица для озёр (не совместно) [здесь](#нагрузки).

SS loadings – суммарная нагрузка фактора, Proportion Var – доля объяснённой вариации данным фактором, Cumulative Var – накопленная доля объяснённой вариации данным и предыдущими факторами.

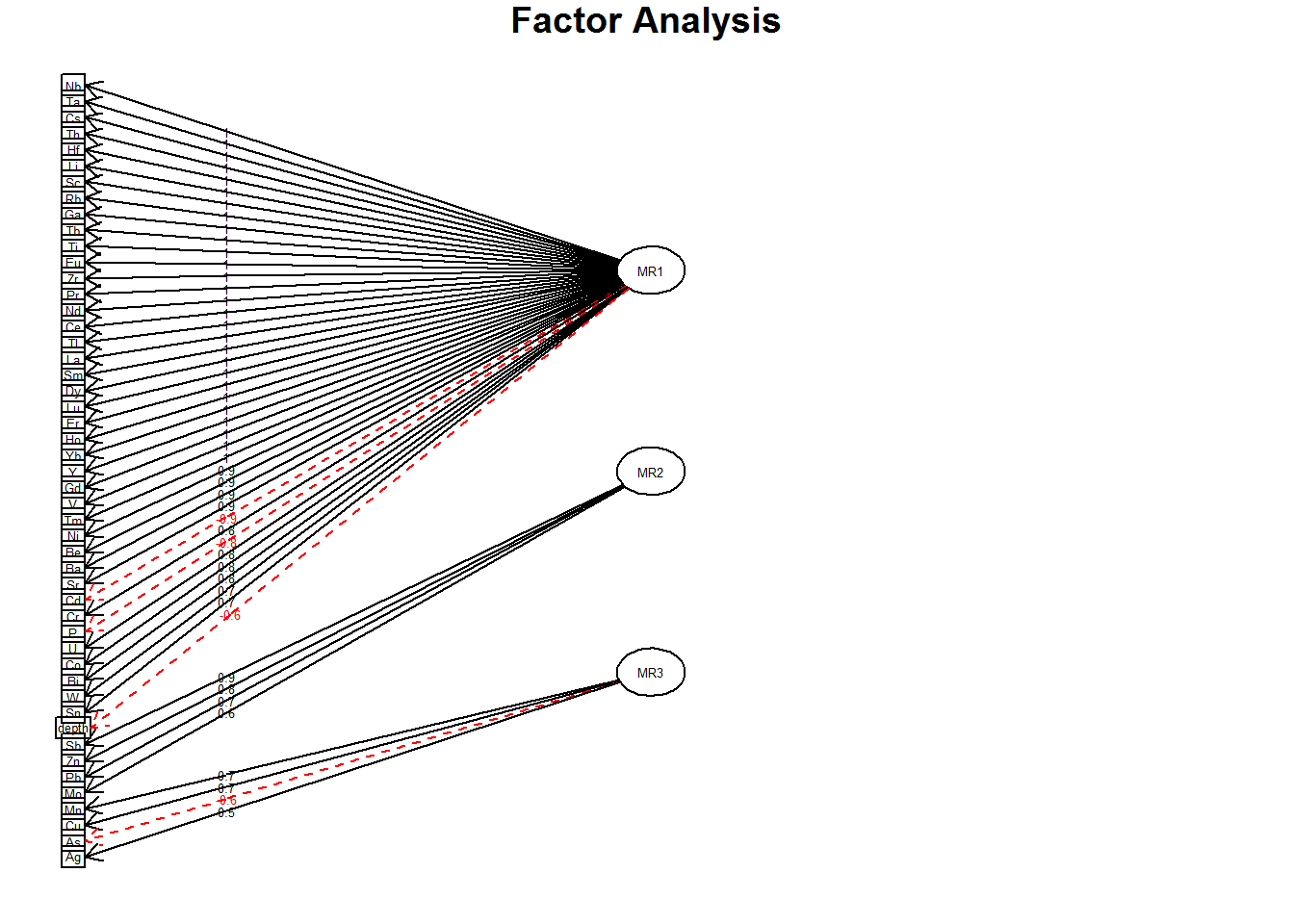
1 фактор описывает данные на 76.9%

1+2 факторы описывают данные на 85.6%

1+2+3 факторы описывают данные на 91%

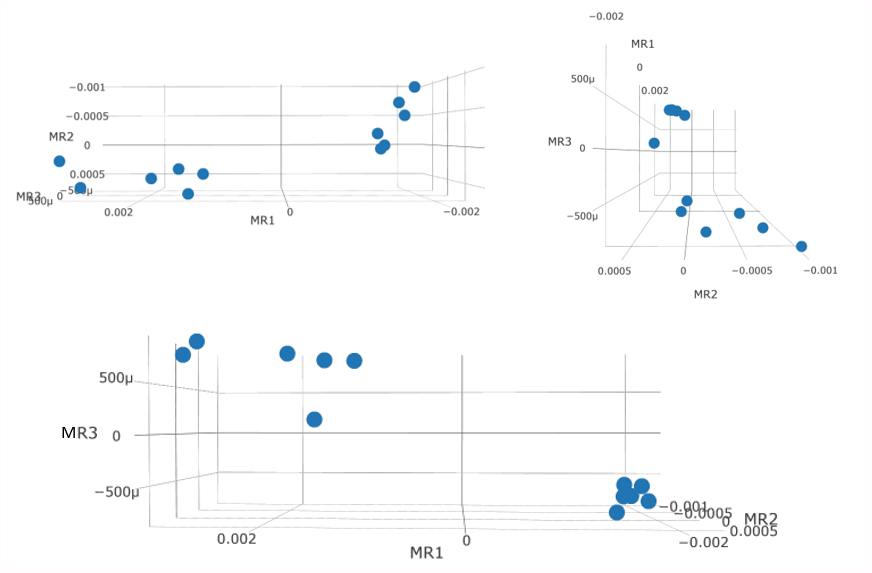
|  |  |  |  |
| --- | --- | --- | --- |
|  | Озеро 1 | | |
| MR1 | MR2 | MR3 |
| SS loadings | 37.66 | 4.307 | 2.613 |
| Proportion Var | 0.769 | 0.088 | 0.053 |
| Cumulative Var | 0.769 | 0.856 | 0.910 |

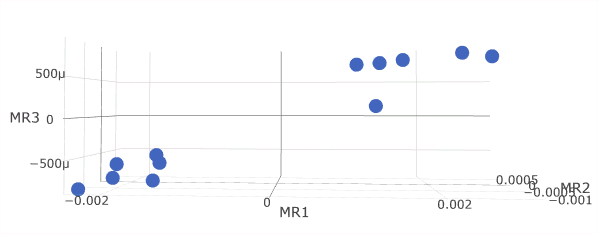
Какие исходные факторы (хим элементы) входят в каждый созданный фактор:



В исходном виде данные были представлены 49 переменными (хим элементами), а теперь представлены 3 факторами. Получается, что каждый выделенный фактор обобщает в себя несколько исходных (было 49 столбцов, стало 3). При этом количество строк не изменилось. Каждая строка – наблюдение, его теперь можно охарактеризовать 3 значениями (факторов), а эти 3 значения мы можем представить в виде координат в 3-мерном пространстве (до этого было 49 измерений, их не удастся нарисовать). Получается, что теперь каждый отобранный образец (наблюдение) можно превратить в точку на графике.

Если посмотреть на нижний рисунок, то видно что наблюдения разделились на 2 класса, причём разделение проходит по фактору №1 (MR1) в болшей степени, и по фактору №2 1 (MR2) в меньшей. На нижнем рисунке, справа сверху наблюдения имеют положительные значения по факторам №1 и №2, снизу слева – отрицательные. Если посмотреть на факторные нагрузки [здесь](#нагрузки), и с пониманием геохимии объяснить суть этих двух факторов, то можно так же объяснить разделение образцов на 2 класса. Пример (на уровне моего понимания я могу объяснить только глубину – depth, она имеет отрицательное вхождение в оба фактора): в наблюдениях, скопившихся сверху справа глубина отбора образца меньше, чем в наблюдениях, скопившихся снизу слева, -- иными словами, класс образцов слева взят с более глубоких мест озера.





# Приложение

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Озеро 1 | | | Озеро 2 | | Озеро 2 | | | Озеро 3 | | |
| MR1 | MR2 | MR3 | MR1 | MR2 | MR1 | MR2 | MR3 | MR1 | MR2 | MR3 |
| depth | -0.628 | -0.286 | -0.414 | -0.435 | -0.649 | -0.504 | -0.630 | -0.199 | -0.814 | 0.448 |  |
| Li | 1.010 |  |  | 0.871 | 0.182 | 0.861 | 0.182 | -0.203 | 1.003 | 0.156 |  |
| Be | 0.940 |  |  | 0.826 | 0.249 | 0.847 | 0.244 |  | 0.340 | 0.766 | 0.458 |
| P | -0.839 | 0.369 | -0.134 | -0.289 | -0.660 | -0.323 | -0.645 | 0.102 |  | -0.164 | 0.934 |
| Sc | 1.010 |  |  | 0.383 | 0.268 | 0.394 | 0.263 |  | -0.706 | 0.499 | -0.160 |
| Ti | 1.007 |  |  | 0.892 | 0.156 | 0.881 | 0.157 | -0.195 | 0.955 | 0.236 |  |
| V | 0.963 |  | 0.114 | 0.540 | 0.330 | 0.552 | 0.319 | -0.121 | 0.633 | -0.107 | 0.571 |
| Cr | 0.840 | 0.347 | -0.402 | 0.760 | 0.254 | 0.774 | 0.246 |  | 0.987 | 0.145 |  |
| Mn | 0.398 | 0.190 | 0.688 | -0.253 | -0.710 | -0.289 | -0.694 | 0.109 | 0.864 | -0.334 |  |
| Co | 0.828 |  | 0.296 | 0.732 | -0.179 | 0.798 | -0.200 | 0.527 | 0.751 |  | 0.518 |
| Ni | 0.947 |  | 0.130 | 0.506 | 0.439 | 0.528 | 0.423 | -0.109 | 0.849 |  | 0.316 |
| Cu | 0.356 | 0.124 | 0.688 | 0.141 | 0.726 | 0.205 | 0.695 |  | 0.339 | 0.425 | 0.729 |
| Zn | 0.148 | 0.771 |  | 0.148 | 0.496 | 0.210 | 0.473 | 0.208 | 0.646 | -0.288 | 0.478 |
| Rb | 1.010 |  |  | 0.924 | 0.101 | 0.908 | 0.104 | -0.199 | 1.010 |  |  |
| Sr | 0.913 |  | 0.213 | 1.020 | -0.287 | 1.014 | -0.278 | 0.114 | 0.884 | -0.102 | 0.261 |
| Y | 0.978 |  |  | 0.973 |  | 0.985 |  |  | -0.414 | 0.871 |  |
| Zr | 1.000 |  |  | 0.770 | 0.191 | 0.790 | 0.189 |  | 0.935 |  |  |
| Nb | 1.014 |  |  | 0.891 | 0.162 | 0.882 | 0.162 | -0.175 | 0.979 | 0.245 | -0.143 |
| Mo | 0.508 | 0.551 | 0.192 |  | 0.614 |  | 0.586 | 0.200 | 0.502 |  | 0.740 |
| Ag | -0.371 | 0.384 | 0.451 | -0.286 | -0.321 | -0.236 | -0.339 | 0.589 | -0.201 | 0.447 | 0.482 |
| Cd | -0.891 | 0.518 | 0.352 | -0.157 | 0.642 |  | 0.648 | 0.531 | 0.762 | -0.428 | 0.148 |
| Sn | 0.657 | 0.551 | -0.235 |  | 0.523 |  | 0.522 | 0.502 | 0.892 |  | 0.243 |
| Sb | -0.186 | 0.950 |  | -0.165 | 0.408 |  | 0.436 | 0.720 | 0.953 | -0.126 |  |
| Cs | 1.013 |  |  | 0.898 | 0.144 | 0.886 | 0.145 | -0.195 | 1.015 |  | -0.160 |
| Ba | 0.926 |  | 0.204 |  | -0.764 |  | -0.759 | 0.236 | 0.929 |  | 0.117 |
| La | 0.995 |  |  | 1.045 | -0.168 | 1.041 | -0.161 |  | 0.120 | 0.987 |  |
| Ce | 0.997 |  |  | 1.053 | -0.291 | 1.045 | -0.282 | 0.102 | 0.525 | 0.828 | -0.197 |
| Pr | 0.997 |  |  | 1.027 | -0.117 | 1.022 | -0.110 |  | 0.590 | 0.754 | -0.293 |
| Nd | 0.997 |  |  | 1.031 | -0.111 | 1.027 | -0.105 |  | 0.662 | 0.660 | -0.351 |
| Sm | 0.993 |  |  | 1.030 | -0.110 | 1.026 | -0.104 |  | -0.172 | 0.813 | -0.194 |
| Eu | 1.004 |  |  | 1.049 | -0.216 | 1.045 | -0.208 |  |  | 0.880 | -0.253 |
| Gd | 0.977 |  |  | 1.018 |  | 1.015 |  |  |  | 0.965 |  |
| Tb | 1.008 |  |  | 0.974 |  | 0.980 |  |  | -0.483 | 0.809 |  |
| Dy | 0.992 |  |  | 0.976 |  | 0.978 |  |  | -0.607 | 0.762 |  |
| Ho | 0.985 |  |  | 0.955 |  | 0.965 |  |  | -0.624 | 0.744 |  |
| Er | 0.986 |  |  | 0.960 |  | 0.960 |  |  | -0.536 | 0.796 |  |
| Tm | 0.962 |  |  | 0.923 | 0.101 | 0.925 | 0.101 |  | -0.534 | 0.759 |  |
| Yb | 0.981 |  |  | 0.947 |  | 0.948 |  |  | -0.134 | 0.837 | -0.223 |
| Lu | 0.990 |  |  | 0.899 | 0.160 | 0.903 | 0.158 |  | 0.283 | 0.896 | -0.170 |
| Hf | 1.011 |  |  | 0.843 | 0.205 | 0.852 | 0.203 |  | 0.941 |  |  |
| Ta | 1.014 |  |  | 0.905 |  | 0.895 |  | -0.139 | 0.986 | 0.106 |  |
| W | 0.679 | 0.548 | 0.183 | 0.159 | 0.639 | 0.213 | 0.618 |  | 0.838 |  | 0.388 |
| Tl | 0.996 |  | -0.102 | 0.767 | 0.293 | 0.786 | 0.286 |  | 0.964 |  | 0.107 |
| Pb | 0.342 | 0.719 | 0.233 | -0.196 |  | -0.114 |  | 0.783 | 0.965 | -0.154 |  |
| Bi | 0.764 | 0.385 | -0.101 |  |  |  |  | 0.698 | 0.971 | -0.139 |  |
| Th | 1.011 |  | -0.105 | 0.928 | 0.104 | 0.915 | 0.106 | -0.181 | 0.710 | 0.607 | -0.330 |
| U | 0.831 | 0.103 | 0.326 | 0.737 | 0.380 | 0.746 | 0.371 | -0.167 |  | 0.860 | 0.167 |