

PMTools

beta v.0.9.0



Руководство пользователя

от 19 апреля 2021

Москва, 2021 г.

Условия использования и доп. информация

PMTools написана в 2020-2021 гг.

PMTools распространяется бесплатно и имеет открытый исходный код.

Автор: Ефремов И. В.

PMTools написана под руководством Веселовского Р. В.

По всем вопросам и пожеланиям, связанным с использованием PMTools, обращаться по адресу i1948374@yandex.ru (Иван Ефремов).

При работе с программой пользователь обязан ссылаться на соответствующую статью:
(ожидается...)

Лицензия MIT

Copyright © 2021

Оглавление

Введение	3
1. Взаимодействие с PMTools	6
1.1. Стартовое окно	6
1.2. Горячие клавиши	6
1.3. PCA	7
1.4. Directional statistics	9
1.5. Poles.....	11
1.6. Статистические тесты	12
1.6.1. Тест складки	12
1.6.2. Тест обращения	12
1.6.3. Тест на общее среднее	13
1.6.4. Тест галек	14
1.7. Окно настроек	14
1.8. Управление графикой	15
1.8.1. Контекстное меню.....	15
1.8.2. Масштабирование и экспорт	17
2. Методы статистической обработки палеомагнитных данных, реализованные в PMTools	20
2.1. Распределения векторов на сфере.....	20
2.2. Палеомагнитные статистические тесты	21
2.2.1. Тесты значимости	21
2.2.2. Полевые тесты	23
2.3. Метод Bootstrap.....	25
2.3.1. Модификации тестов	26

Введение

Сейчас PMTools это кроссплатформенная программа для анализа палеомагнитных данных, включающая полный цикл палеомагнитных операций — от компонентного анализа до нахождения виртуальных геомагнитных полюсов.

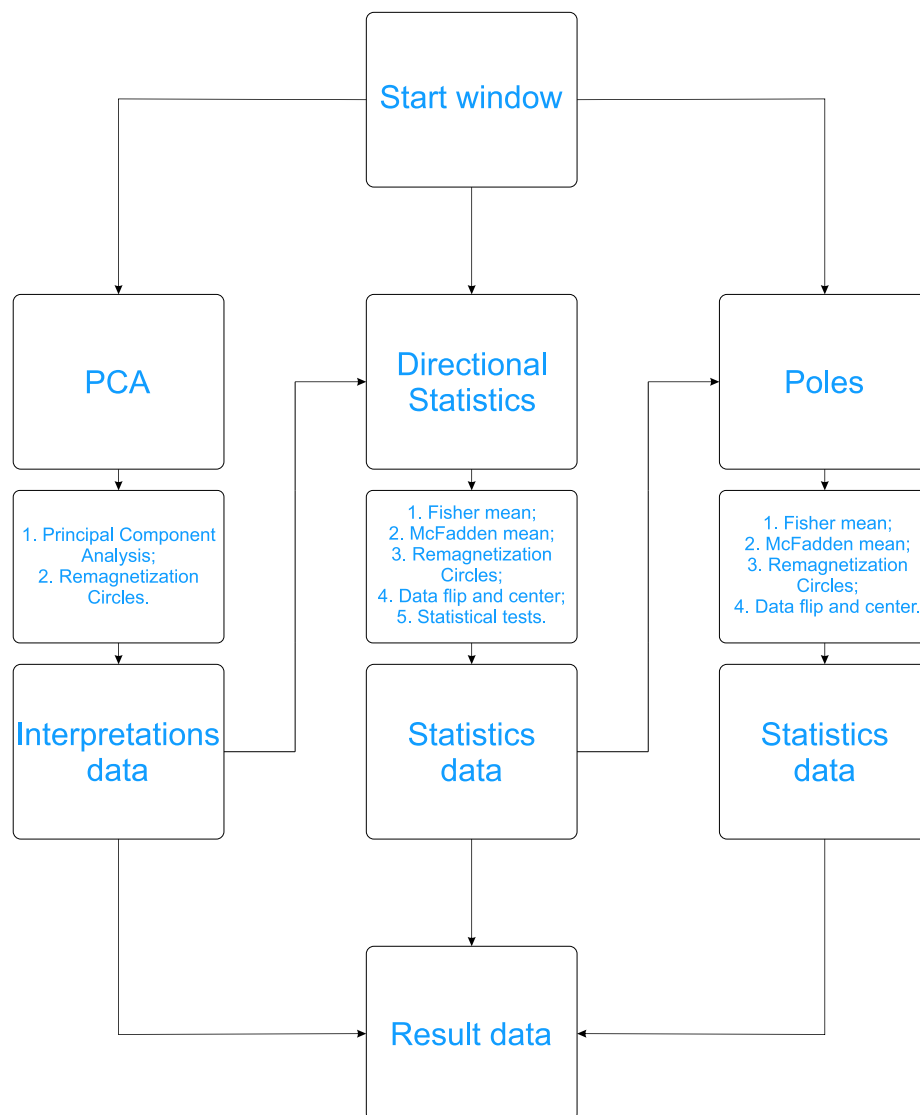


Рис. 0.1: Концепция программы PMTools

PMTools поддерживает множество распространённых форматов хранения палеомагнитных данных: `.cit`, `.dir`, `.jr6`, `.jra`, `.pmd`, `.pmm`, `.rs3`, `.rsf` и некоторые другие.

Реализована поддержка импорта и экспорта данных в MagIC-формате, что позволяет напрямую и с максимальным удобством взаимодействовать с базой палеомагнитных данных MagIC.

Также к ключевым особенностям программы можно отнести наличие горячих клавиш, при помощи которых можно использовать почти все имеющиеся функции. Важно отметить, что горячие клавиши максимально соответствуют таковым из программ Р. Энкина под DOS

[Enkin, 1996]. Использование горячих клавиш значительно упрощает и ускоряет процесс анализа палеомагнитных данных.

Другой важной особенностью программы является её графическая составляющая. Для создания графики в РМTools используется библиотека `highcharts.js`, позволяющая создавать интерактивные векторные графики и диаграммы, с которыми очень просто и удобно взаимодействовать. Вся графика может быть экспортирована как в векторном, так и в растровом формате, и без дополнительной обработки использоваться как графический материал для презентаций и публикаций.

Поскольку в РМTools реализован полный цикл палеомагнитных операций, было решено создать 3 окна, в каждом из которых реализован отдельный этап анализа данных: компонентный анализ (PCA), статистика направлений (*Directional statistics*) и расчёт полюсов (*Poles*). В каждом окне имеются свои индивидуальные особенности, потому о каждом из окон будет подробно рассказано далее. Помимо них будут рассмотрены и все остальные имеющиеся окна.

1. Взаимодействие с PMTools

1.1. Стартовое окно

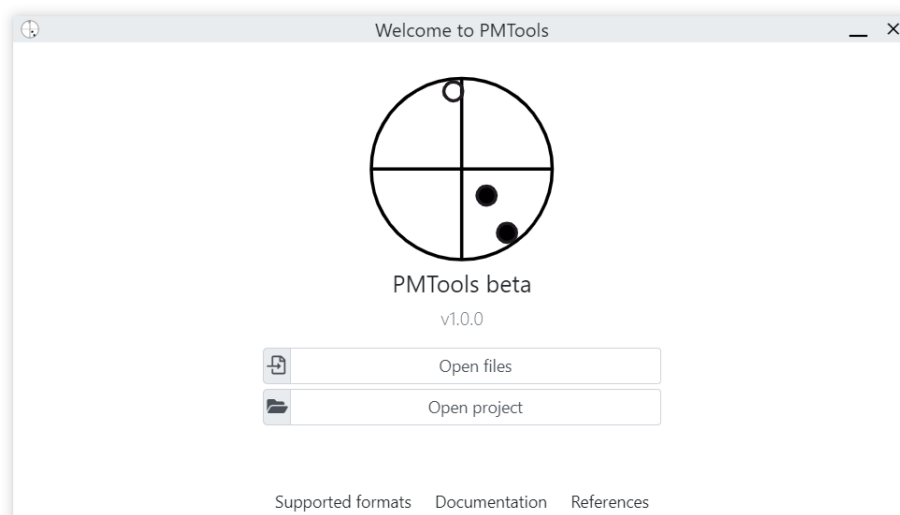


Рис. 1.1: Стартовое окно при первом запуске

При первом запуске PMTools откроется следующее окно (рис. 1.1). Здесь можно открыть файлы любого из поддерживаемых форматов нажав кнопку **Open files**, либо же открыть проект, нажав кнопку **Open project**. Проект — это `.json` файл, хранящий в себе всю информацию о рабочей сессии: все открытые файлы, все настройки графиков и диаграмм, все полученные статистические результаты. Таким образом, проект позволяет начать работу ровно с того места, где работа была приостановлена.

В нижней части окна находятся кнопки вызова окон с дополнительной информацией. Там вы можете посмотреть все доступные форматы данных — кнопка **Supported formats**. Также можете открыть документацию к программе — кнопка **Documentation**. А при нажатии на кнопку **References** можно увидеть всю информацию об авторах, лицензии и правах.

Если же программа запускается не в первый раз и при этом последняя сессия была автоматически сохранена (по умолчанию последняя сессия автоматически сохраняется в памяти программы, изменить это можно в настройках), то стартовое окно не откроется, но сразу же откроется программа с данными предыдущей сессии.

1.2. Горячие клавиши

Горячие клавиши в PMTools существуют как для каждого из окон, так и для всей программы в целом. Списки всех горячих клавиш можно увидеть на рис. 1.2.

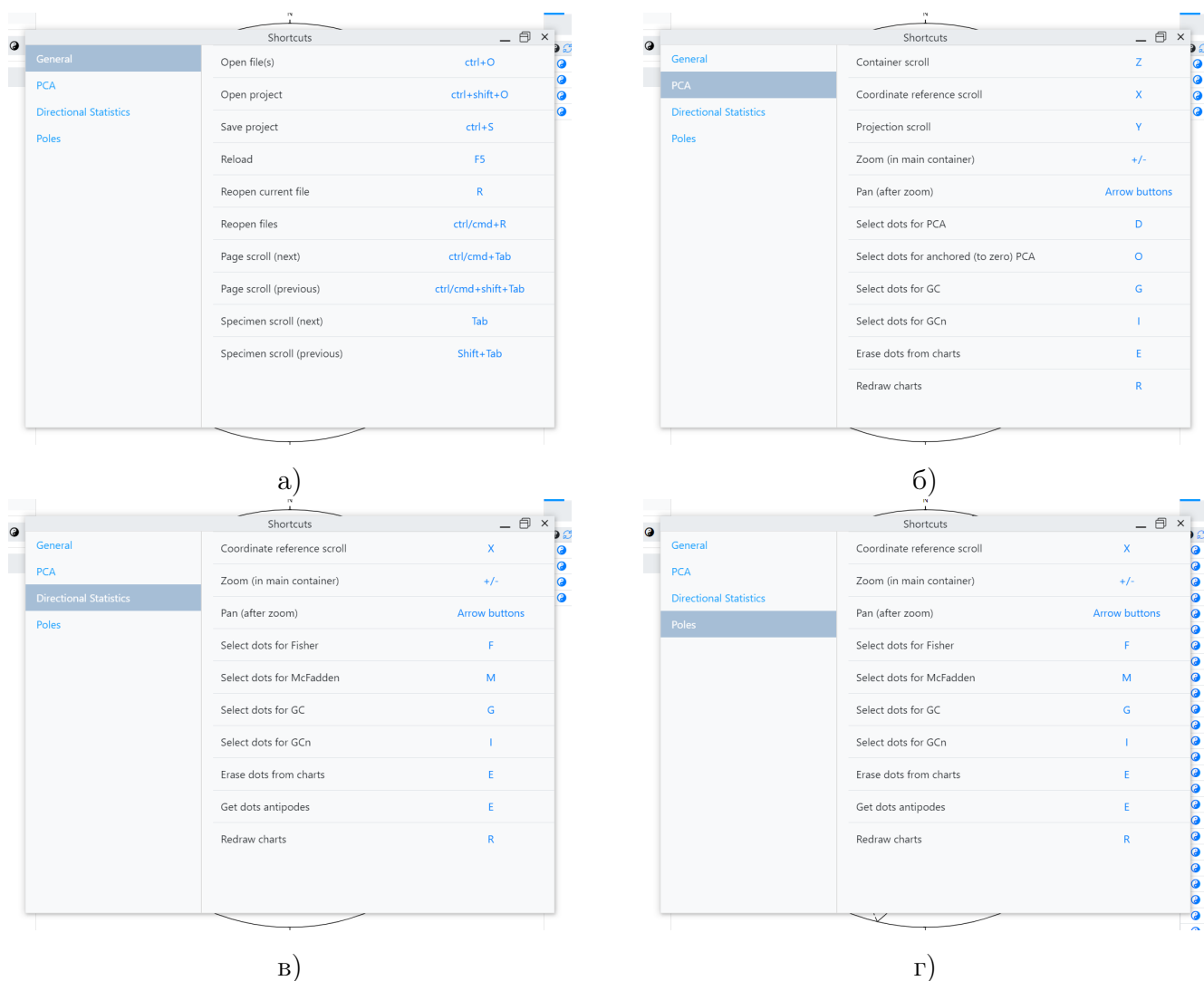




Рис. 1.2: Горячие клавиши для а) всей программы; б) окна PCA; в) окна Directional statistics; г) окна Poles

1.3. PCA

Импортировав первичные палеомагнитные данные (результаты магнитных чисток), откроется в окно для проведения компонентного анализа (рис. 1.3). Оно предназначено для выделения компонент естественной остаточной намагниченности.

В правой верхней части окна находится поле с информацией о пути к открытому файлу, ниже — строка выбора текущего файла (из всех открытых), и, далее, сокращённая таблица со входными данными. Чтобы увидеть полную таблицу со входными данными, необходимо нажать кнопку **Specimen data**. В таблице, помимо самих данных, первым столбцом находятся кнопки , позволяющие скрывать необходимые направления (точки). В заголовке столбца имеется кнопка , позволяющая скрыть все отображаемые точки и отобразить все скрытые.

Входные данные также отображаются на стереограмме, диаграмме Зийдервельда и гра-

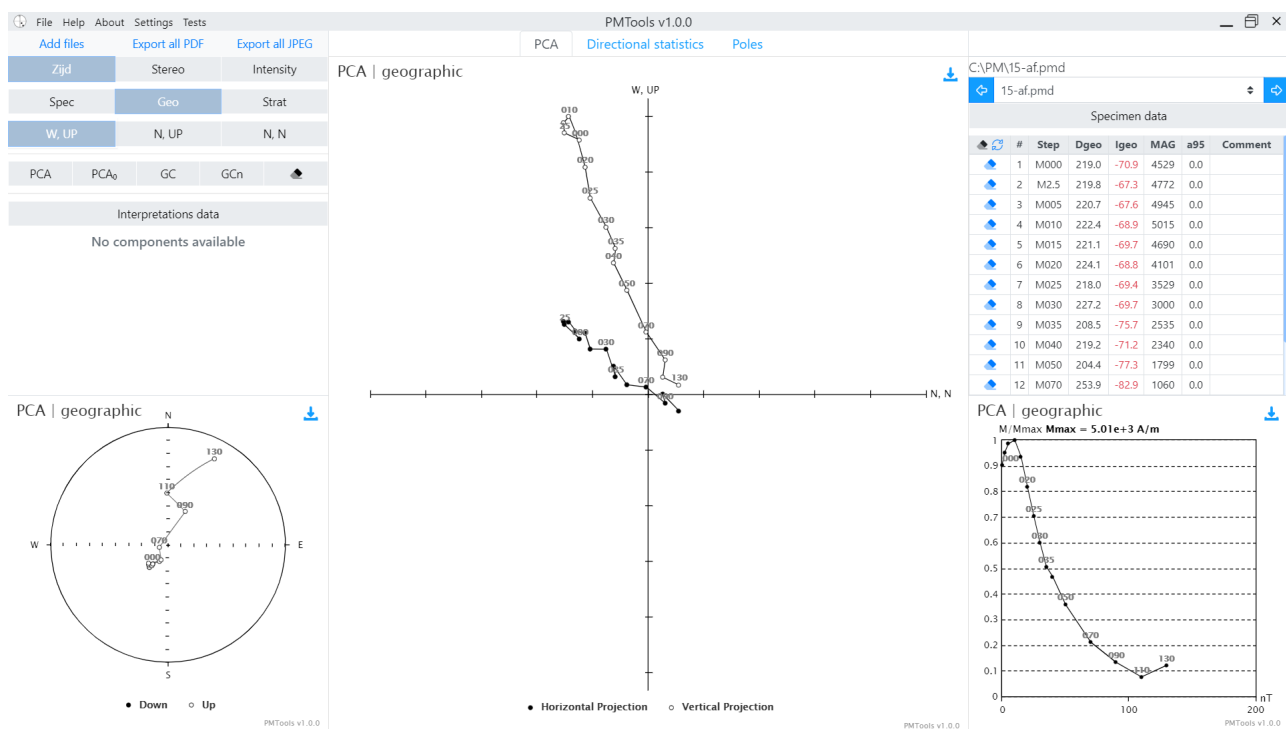



Рис. 1.3: Окно для компонентного анализа данных магнитных чисток

фике размагничивания. Подробнее о работе с графикой будет рассказано далее (см. раздел 2.9. Управление графикой).

В левой верхней части окна находится панель инструментов и управления данными. Через панель управления можно менять положение графиков и диаграмм — кнопки **Zijd**, **Stereo** и **Intensity** (горячая клавиша **Z**) отвечают за их положения. Там же можно изменить систему координат и проекцию на диаграмме Зийдервельда — кнопки **Spec**, **Geo**, **Strat** (горячая клавиша **X**) и кнопки с названиями направлений осей (горячая клавиша **Y**) соответственно. Ниже находятся кнопки, позволяющие выбрать направления (точки) для соответствующих статистических методов: **PCA** (Principle Component Analysis, **D**), **PCA₀** (Principle Component Analysis with zero anchoring, **O**), **GC** (Great Circles, **G**), **GCn** (Great Circle normalized, **I**). Также там имеется кнопка , позволяющая выбрать точки, которые необходимо скрыть (горячая клавиша **E**).

Выбор точек может производиться двумя путями — либо посредством строки ввода (по умолчанию), либо путём выбора начальной и конечной точек из выпадающих списков. Строка ввода позволяет выбирать не только начальную и конечную точки, она также позволяет выбирать точки по условию. Например, можно выбрать все точки с положительным наклоном: `inc > 0`, или точки в определённой области: `(dec > 180) && (dec < 270) && (inc > 40) && (inc < 80)`, или сделать выборку по комментарию: `comment == 'h'`. Для выбора могут быть использованы следующие наименования параметров: `id`, `code`, `step`, `n`, `id`, `index`, `dec`, `inc`, `MAG`, `mag`, `a95`, `MAD`, `mad`, `comment`. В качестве разделителя дробной части используется точка.

Ниже представлены все возможные команды для выбора точек и их краткие описания:

1. логические операторы: `&&` (и), `||` (или), `==` (равно), `!=` (не равно), `>` (больше), `<` (меньше);
2. фрагментация строки: `slice()` — позволяет получить фрагмент строчного параметра, например `code.slice(0,2) == "GC"`;
3. поиск символов:
 - (a) `includes()` — позволяет проверять наличие символа или определенной последовательности символов в строке, например `code.includes("GC") == true`;
 - (b) `startsWith()` — позволяет проверить, начинается ли строка с определённых символов, например `code.startsWith("GC") == true`;
 - (c) `endsWith()` — позволяет проверить, кончается ли строка определёнными символами, например `code.endsWith("Kir") == true`.

Результаты применения статистических методов для текущего файла записываются в таблицу ниже кнопки **Interpretations data**. При этом результаты со всех файлов автоматически заносятся в общую таблицу, которую можно увидеть, нажав на кнопку **Interpretations data**. Эту общую таблицу можно сохранить в `.csv` и `.xlsx` и открыть в следующем окне.

1.4. Directional statistics

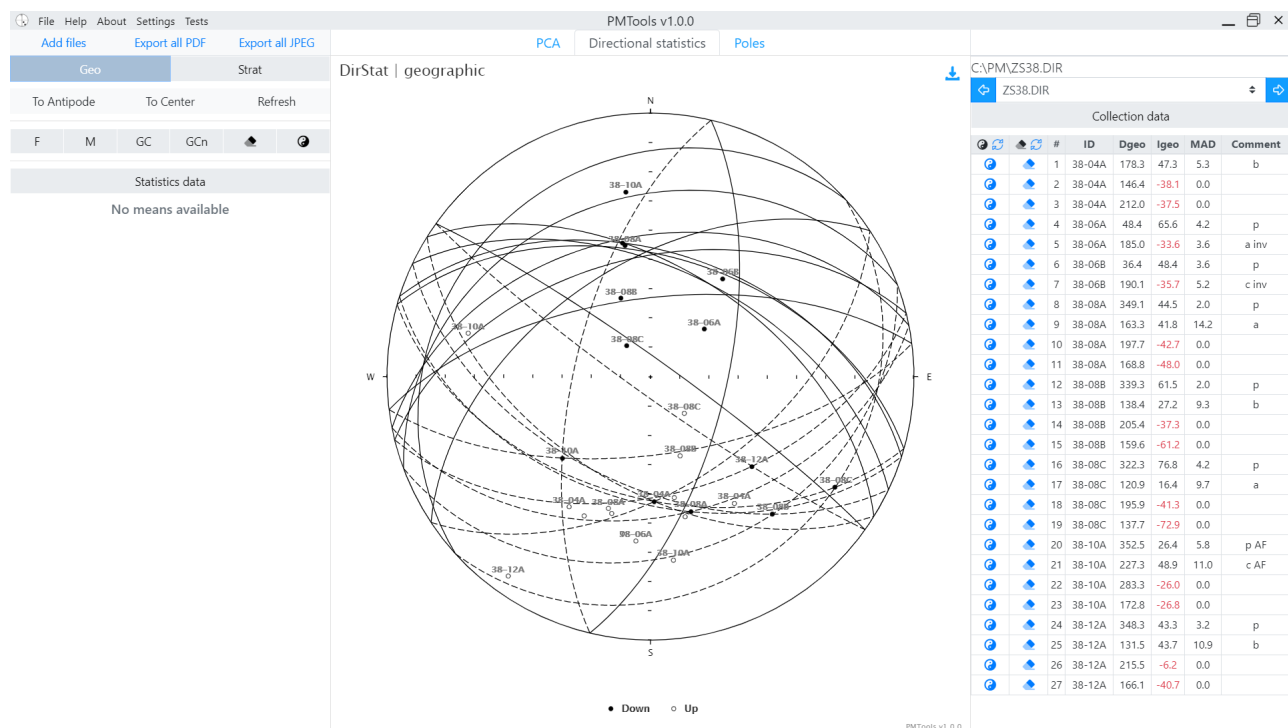







Рис. 1.4: Окно для статистики направлений выделенных компонент намагниченности

Полученные с прошлого этапа данные (выделенные компоненты намагниченности) отображаются в этом окне. В правой части находится путь к открытому файлу, строку выбора текущего файла (из всех открытых) и сокращённая таблица со входными данными. Чтобы увидеть полную таблицу со входными данными, необходимо нажать кнопку **Collection data**. В сокращённой таблице, помимо самих данных, первым столбцом находятся кнопки , позволяющие менять полярность направления. Вторым столбцом идут кнопки , позволяющие скрывать необходимые направления (точки). В заголовках столбцов имеется кнопка , позволяющая инвертировать применение соответствующего инструмента (скрытия или смены полярности направлений).

В центральной части окна располагается стереограмма, на которой отображаются входные данные. На ней могут отображаться два типа данных: выделенные направления и выделенные круги перемангничивания. Их отображение зависит от значения параметра в поле **Code**: если первые символы это **Dir**, тогда будет отображаться направление, а если **GC**, тогда это направление считается нормалью к кругу перемангничивания и на стереограмме отображается и круг перемангничивания, и нормаль к нему (регулируется через настройки, см. раздел 2.9.1 Контекстное меню).

В левой верхней части окна находится панель управления данными. Здесь можно менять систему координат — кнопки **Geo** и **Strat** (горячая клавиша **X**). Ниже находятся кнопки **To Antipode**, **To Center** и **Refresh**. Кнопка **To Antipode** автоматически делает распределение рассматриваемых данных унимодальным (рассчитывается главная компонента и относительно неё все направления обратной полярности переводятся в прямую полярность). Кнопка **To Center** позволяет центрировать рассматриваемые направления относительно последнего из рассчитанных средних направлений (работает исключительно при наличии рассчитанного среднего направления). Кнопка **Refresh** (горячая клавиша **R**) удаляет все изменения и полученные статистические данные для текущего файла, оставляя только исходные импортированные данные файла.

Статистические методы представлены следующими кнопками: **F** (Fisher statistics, **F**), **M** (McFadden and McElhinny statistics, — совмещение кругов перемангничивания и направлений для статистики Фишера, **M**), **GC** (Great Circles, **G**), **GCn** (Great Circle normalized, **I**). Также там имеются кнопки  и , позволяющие выбрать направления, которые необходимо скрыть (горячая клавиша **E**) и направления, у которых необходимо сменить полярность (горячая клавиша **A**), соответственно.

Результаты применения статистических методов для текущего файла записываются в таблицу ниже кнопки **Statistics data**. При этом результаты со всех файлов автоматически заносятся в общую таблицу, которую можно увидеть, нажав на кнопку **Statistics data**. Эту общую таблицу мы можем сохранить в **.csv** и **.xlsx** формате и открыть в следующем окне.

1.5. Poles

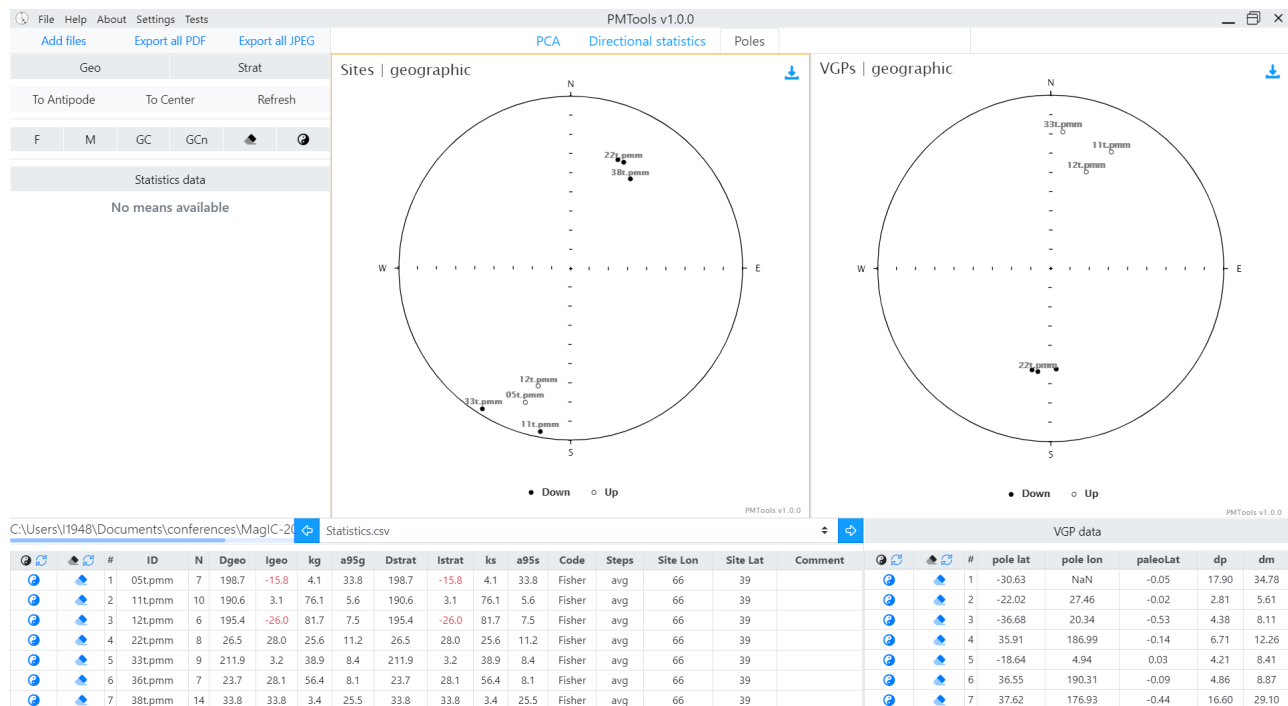


Рис. 1.5: Окно для статистической обработки данных по полюсам

Полученные с прошлого этапа данные (средние направления по сайтам) отображаются в этом окне. В нижней части окна находятся две таблицы: левая таблица — таблица со входными данными, правая — производная от левой таблицы таблица с рассчитанными виртуальными геомагнитными полюсами (VGP). Первые два столбца каждой из таблиц — кнопки и , позволяющие менять полярность направлений и скрывать их соответственно. В заголовках этих столбцов имеется кнопка , позволяющая инвертировать применение соответствующего инструмента (скрытия или смены полярности направлений). В левой таблице имеются столбцы **Site Lon** и **Site Lat**, значения в которых редактируемы. В них пользователь вносит координаты сайта, если они не были внесены заранее. Как только для направления заданы широта и долгота, автоматически в таблицу справа добавляется виртуальный геомагнитный полюс (Virtual Geomagnetic Pole, VGP), соответствующий этому направлению.

Данные с обеих таблиц отображаются на стереограммах выше: данные с левой таблицы на левой стереограмме, с правой - на правой стереограмме.

В левой верхней части экрана находится панель управления данными, полностью соответствующая той, что была в окне **Directional Statistics**.

1.6. Статистические тесты

В PMTools реализованы тест складки, тест обращения и тест галек. При этом для тестов складки и обращения используются их bootstrap-модификации. Для каждого из тестов имеется своё собственное окно, кнопки вызова которых находятся в верхней строке меню программы (в **Tests**). В качестве входных данных для всех тестов автоматически выбирается текущий файл в окне Directional Statistics.

1.6.1. Тест складки



Рис. 1.6: Окно для применения теста складки к текущему файлу

В верхней части окна находится дополнительная информация о самой процедуре теста. Далее идёт кнопка **Initialize**, при нажатии на которую начинается выполнение теста складки. Процесс выполнения теста показывается на прогресс-баре под кнопкой **Initialize**. В ходе теста выполняется 1000 симуляций, потому процесс не мгновенный и может занять несколько секунд (зависит от характеристик компьютера, на котором выполняется тест). Подробнее про суть теста можно узнать в разделе Модификации тестов.

1.6.2. Тест обращения

В верхней части окна находится дополнительная информация о самой процедуре теста. Далее идёт кнопка **Initialize**, при нажатии на которую начинается выполнение теста обращения. Процесс выполнения теста показывается на прогресс-баре под кнопкой **Initialize**. В ходе теста выполняется суммарно 2000 симуляций (по 1000 на данные прямой и обратной полярностей), потому процесс не мгновенный и может занять несколько секунд (зависит от характеристик компьютера, на котором выполняется тест). Подробнее про суть теста можно узнать в разделе Модификации тестов.

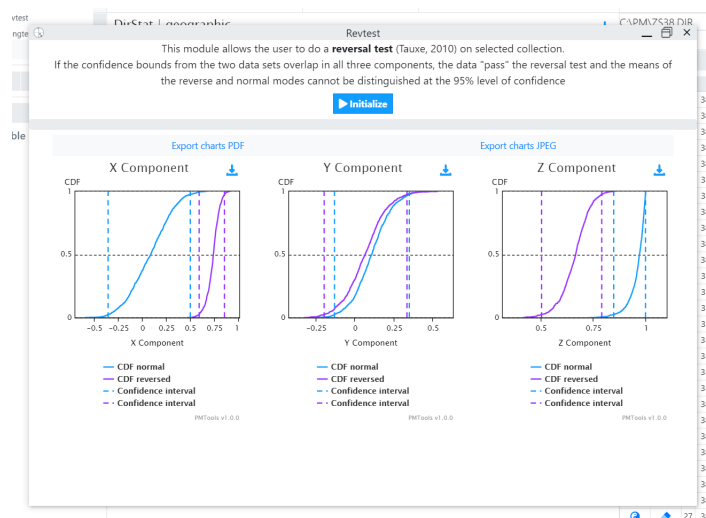


Рис. 1.7: Окно для применения теста обращения к текущему файлу

Экспортировать можно как все три графика сразу — кнопки **Export charts PDF** и **Export charts JPEG**, так и каждый из них по отдельности.

1.6.3. Тест на общее среднее

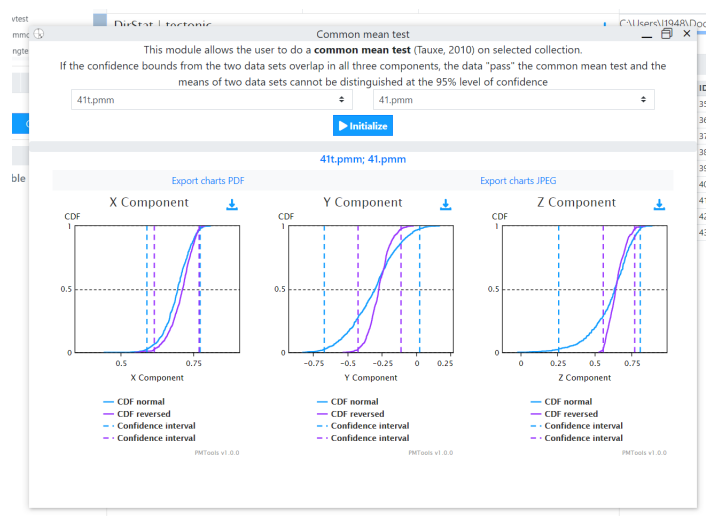


Рис. 1.8: Окно для применения теста на общее среднее

В верхней части окна находится дополнительная информация о самой процедуре теста. Ниже расположены поля выбора коллекций, которые будут сравниваться. Далее идёт кнопка **Initialize**, при нажатии на которую начинается выполнение теста складки. Процесс выполнения теста показывается на прогресс-баре под кнопкой **Initialize**. В ходе теста выполняется суммарно 2000 симуляций (по 1000 на каждый из наборов данных), потому процесс не мгновенный и может занять несколько секунд (зависит от характеристик компьютера, на котором выполняется тест). Подробнее про суть теста можно узнать в разделе Тесты значимости.

Экспортировать можно как все три графика сразу — кнопки **Export charts PDF** и **Export charts JPEG**, так и каждый из них по отдельности.

1.6.4. Тест галек

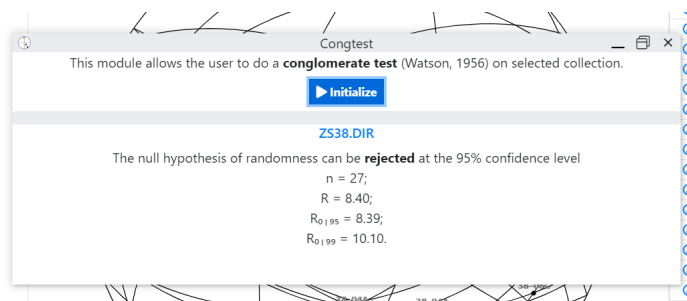


Рис. 1.9: Окно для применения теста галек (конгломератов) к текущему файлу

В верхней части окна находится дополнительная информация о самой процедуре теста. Далее идёт кнопка **Initialize**, при нажатии на которую начинается выполнение теста складки. Процесс выполнения теста показывается на прогресс-баре под кнопкой **Initialize**. Подробнее про суть теста можно узнать в разделе Полевые тесты.

1.7. Окно настроек

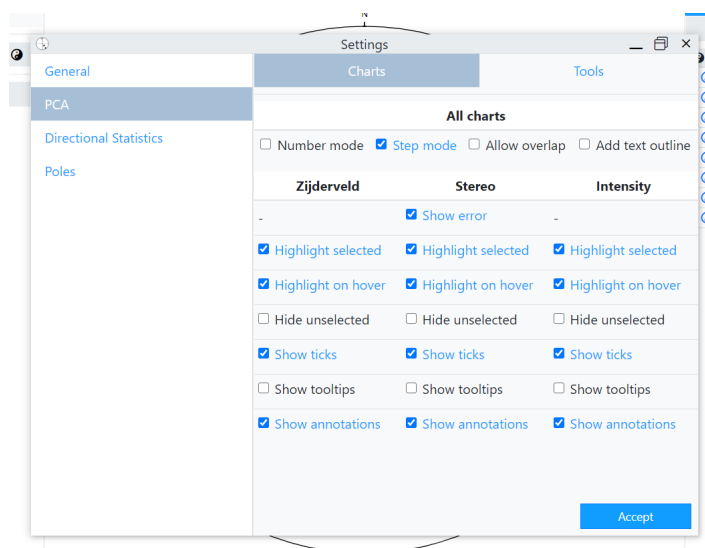


Рис. 1.10: Окно настроек PMTools

В PMTools имеется окно настроек, позволяющее настраивать отображение элементов графиков, изменять способ ввода нужных направлений при использовании панели управления данными, а также менять некоторые глобальные параметры программы.

Глобальными параметрами являются следующие:

1. **Auto-save current project** — при закрытии PMTools проект автоматически сохранится в её памяти (не в виде файла!) и будет автоматически открыт при следующем запуске PMTools;
2. **Append new files** — каждый новый файл будет добавляться к уже существующим, а не заменять их (не объединение файлов!);
3. **Dashed lines** — глобальная настройка для всех стереограмм, переводит все линии на стереограммах с $\text{inc} < 0$ из непрерывных в пунктирные.
4. **Theme** — визуальная тема PMTools, выбирается из следующих:
 - (a) **Light** — светлая, классическая тема;
 - (b) **Dark** — тёмная тема;

Для каждого из окон (PCA, Directional statistics, Poles) можно выбрать, использовать ли в качестве способа выбора направлений для обращения (☯), скрытия (🔒) и применения к ним статистических методов текстовое поле (Settings > PCA|Directional Statistics|Poles > Tools > Dots selector > Plain text), или же использовать поля выбора начального и конечного направлений (Settings > PCA|Directional Statistics|Poles > Tools > Dots selector > Select). Наглядно разница между этими способами выбора отображена на рис. 1.11

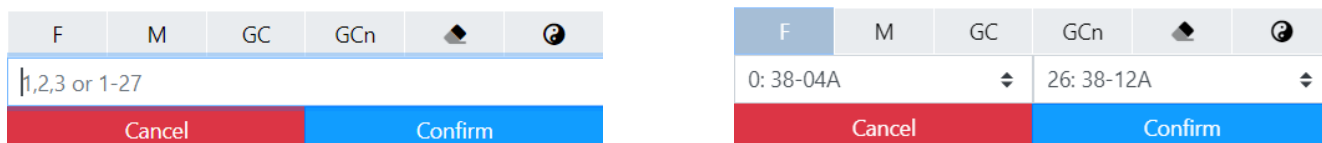


Рис. 1.11: Два способа выбора направлений: текстовый (слева) и графический (справа).

Также имеются настройки для графиков и диаграмм (рис. 1.10), индивидуальные для каждого из окон. Подробно об этих настройках можно узнать в разделе Контекстное меню

1.8. Управление графикой

1.8.1. Контекстное меню

Каждый график и каждая диаграмма в PMTools могут быть настроены под нужды пользователя. Посредством контекстного меню пользователь может быстро изменять отображение данных. Набор параметров в контекстном меню зависит от графика/диаграммы, откуда оно было вызвано. В частности, для стереограмм контекстное меню содержит дополнительные параметры (рис. 1.14), которых нет в контекстном меню для диаграммы Зийдервельда и графика размагничивания (рис. 1.12).

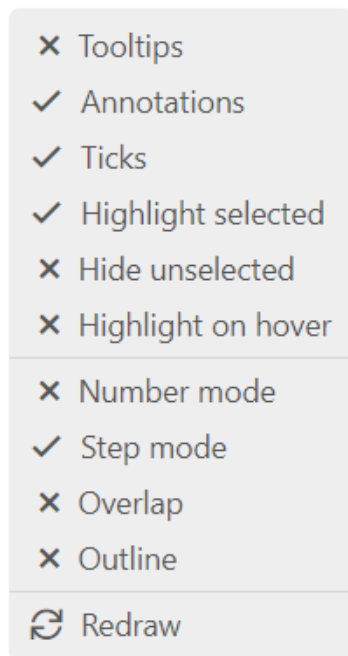


Рис. 1.12: Контекстное меню для диаграммы Зийдверельда и графика размагничивания

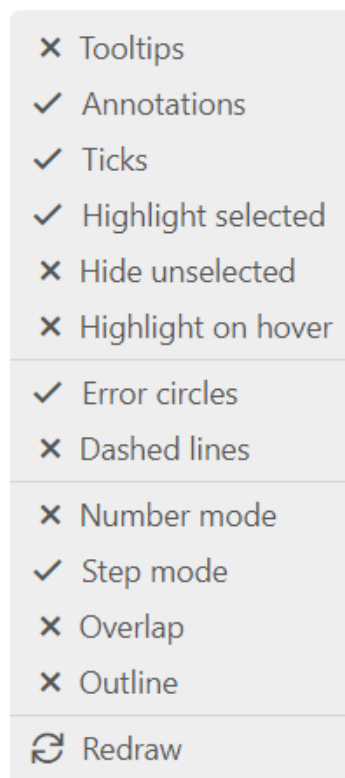


Рис. 1.13: Контекстное меню для стереограммы в окне PCA

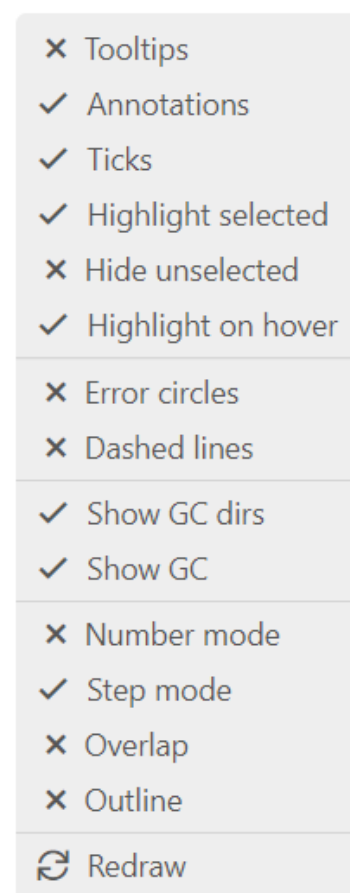


Рис. 1.14: Контекстное меню для стереограмм в окнах Directional Statistics и Poles


Далее рассматриваются все существующие пункты контекстных меню. Выделяются индивидуальные для каждого графика/диаграммы параметры:

1. **Tooltips** — эта настройка регулирует отображение всплывающих при наведении на направление данных о нём (рис. 1.15.1);
2. **Annotations** — эта настройка регулирует отображение подписей точек (рис. 1.15.2);
3. **Ticks** — эта настройка регулирует отображение линий или меток сетки (рис. 1.15.3);
4. **Highlight selected** — эта настройка регулирует отображение подсветки тех точек, которые были выбраны для статистического метода (рис. 1.15.4);
5. **Hide unselected** — эта настройка регулирует отображение тех точек, которые не были выбраны для статистического метода (рис. 1.15.5);
6. **Highlight on hover** — эта настройка регулирует отображение визуального выделения точек при наведении курсора на них (рис. 1.15.1);







7. **Error circles** — эта настройка регулирует отображение кругов доверия направлений (только для стереограмм, рис. 1.15.6);
8. **Dashed lines** — эта настройка регулирует отображение всех линий с отрицательным наклоном (только для стереограмм, рис. 1.15.7);
9. **Show GC dirs** — эта настройка регулирует отображение нормалей к кругам перемангничивания (только для стереограмм в окнах Directional statistics и Poles, рис. 1.15.8);
10. **Show GC** — эта настройка регулирует отображение кругов перемангничивания (только для стереограмм в окнах Directional statistics и Poles, рис. 1.15.9);

И общие для всех графиков и диаграмм параметры:


1. **Number mode** — эта настройка регулирует отображение порядкового номера направления в подписи точки (рис. 1.16.1);
2. **Step mode** — эта настройка регулирует отображение параметра Step или ID направления в подписи точки (рис. 1.16.2);
3. **Overlap** — эта настройка регулирует возможность подписей перекрывать друг друга при их близком расположении (рис. 1.16.3);
4. **Outline** — эта настройка регулирует наличие обводки вокруг подписей, повышающей, в некоторых случаях, их видимость (рис. 1.16.4);

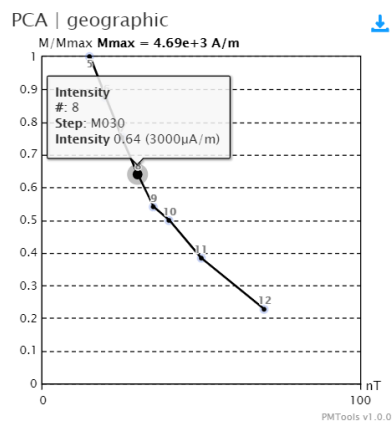
Кнопка  **Redraw** заново отрисовывает график или диаграмму, откуда было вызвано текущее контекстное меню.

1.8.2. Масштабирование и экспорт

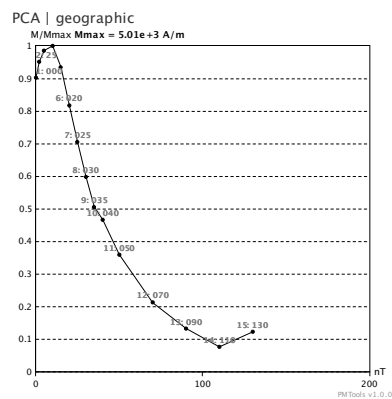
В PMTools имеется возможность масштабировать графику. Делать это можно как посредством горячих клавиш:  и , так и используя колесо мыши (или любой аналогичный по действию способ ввода). Перемещаться по графику или диаграмме можно также либо посредством горячих клавиш: , , , , либо используя мышь.

Экспорт графики может осуществляться двумя способами:

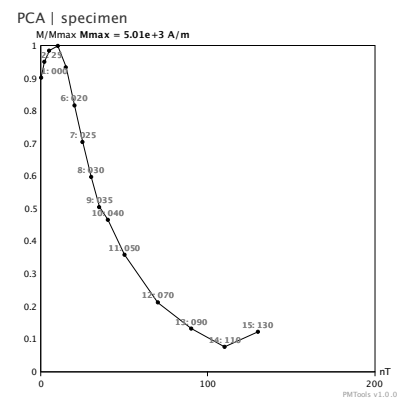
1. экспорт каждого графика или диаграммы по отдельности доступен через кнопку ;
2. экспорт всей графики со страницы, доступен через кнопки в левом верхнем углу каждого из окон (над панелью управления данными): **Export as PDF** и **Export as JPEG**.



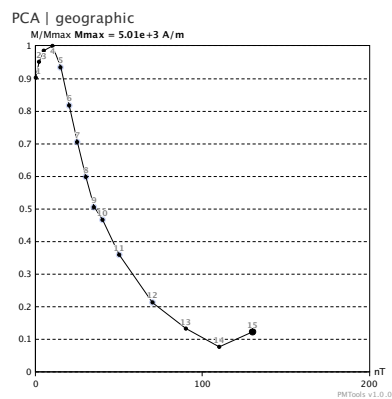
(1.15.1) График размагничивания. Пример включенной настройки Tooltips



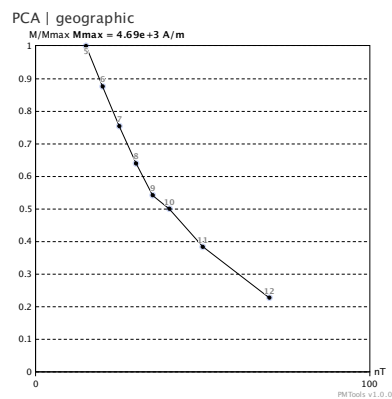
(1.15.2) График размагничивания. Пример включенной настройки Annotations



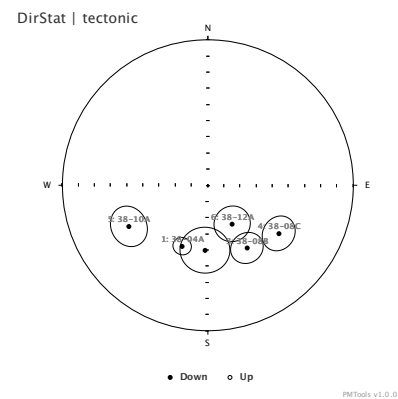
(1.15.3) График размагничивания. Пример выключенной настройки Ticks



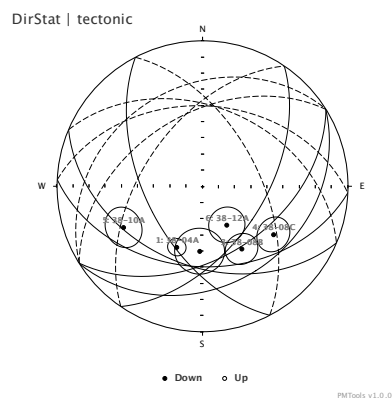
(1.15.4) График размагничивания. Пример включенной настройки Highlight



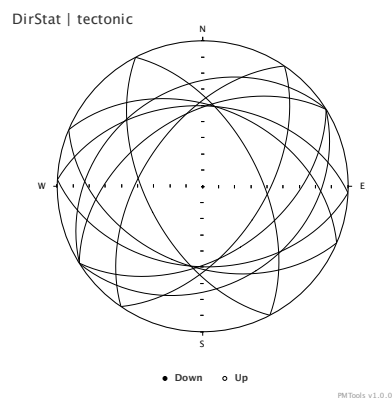
(1.15.5) График размагничивания. Пример включенной настройки Hide



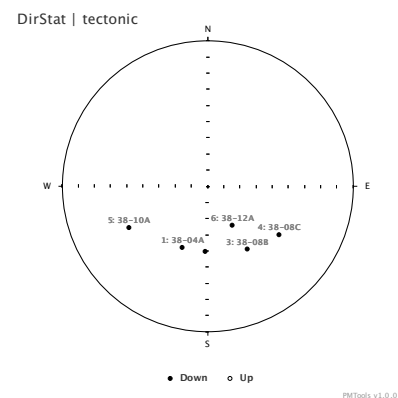
(1.15.6) Стереодиаграмма. Пример включенной настройки Error circles



(1.15.7) Стереодиаграмма. Пример включенной настройки Dashed lines

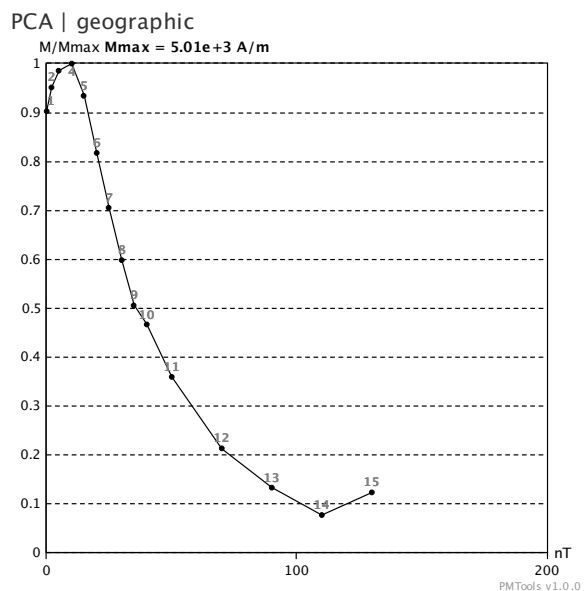


(1.15.8) Стереодиаграмма. Пример выключенной настройки Show GC dirs

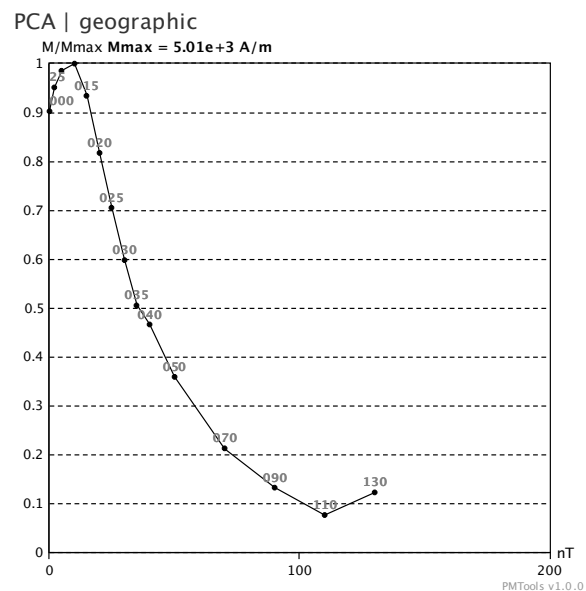


(1.15.9) Стереодиаграмма. Пример выключенной настройки Show GC

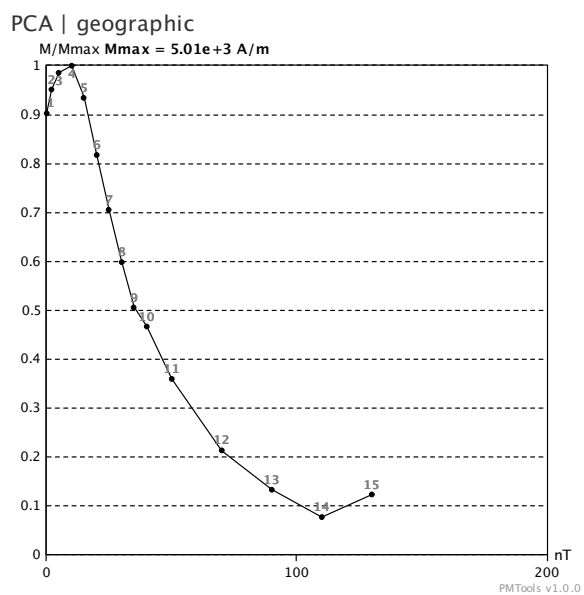
Рис. 1.15: Примеры индивидуальных настроек графиков



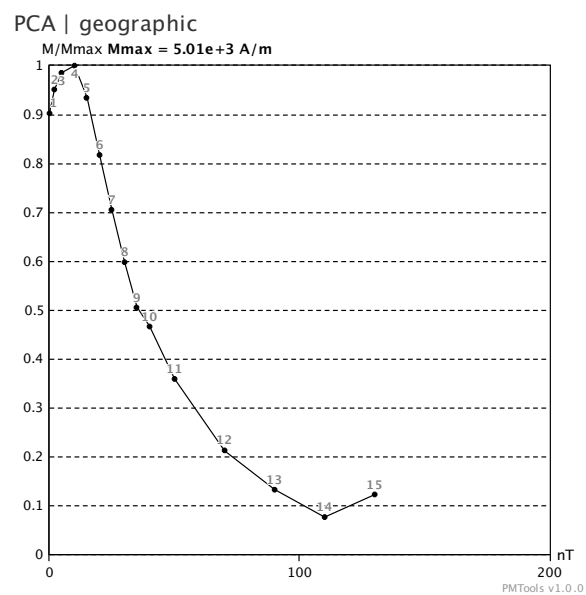
(1.16.1) График размагничивания. Пример включенной настройки Number mode



(1.16.2) График размагничивания. Пример включенной настройки Step mode



(1.16.3) График размагничивания. Пример включенной настройки Overlap



(1.16.4) График размагничивания. Пример включенной настройки Outline

Рис. 1.16: Примеры общих настроек графиков

2. Методы статистической обработки палеомагнитных данных, реализованные в PMTools

2.1. Распределения векторов на сфере

Полученные в ходе палеомагнитных исследований вектора естественной остаточной намагниченности (Natural Remanent Magnetization, NRM) \mathbf{J}_n всегда имеют некоторое распределение в пространстве. В большинстве случаев, лучше всего любое подобное распределение описывается распределением Бингхэма. Основной причиной этого является тот факт, что в сферической системе координат распределение Бингхэма не осесимметрично, то есть оно предполагает, что рассматриваемые данные находятся в пределах некоторого эллипса на поверхности сферы.

В ходе компонентного анализа, при выделении компонент естественной остаточной намагниченности, используется метод главных компонент [Kirschvink, 1980], в основу которого заложено распределение Бингхэма. Главным статистическим параметром для выделенных компонент является величина MAD (Maximum Angular Deviation). Для векторов она вычисляется следующим образом:

$$MAD = \arctan \left(\sqrt{\frac{\lambda_{\text{int}} + \lambda_{\text{min}}}{\lambda_{\text{max}}}} \right). \quad (2.1)$$

И следующим образом для плоскостей (кругов перемангничивания):

$$MAD = \arctan \left(\sqrt{\frac{\lambda_{\text{min}}}{\lambda_{\text{int}}} + \frac{\lambda_{\text{min}}}{\lambda_{\text{max}}}} \right), \quad (2.2)$$

где λ_{min} , λ_{int} , λ_{max} — собственные значения следующей симметрической матрицы [Kirschvink, 1980]:

$$\mathbf{H} = \begin{pmatrix} \sum (x_i - \bar{x})^2 & \sum (x_i - \bar{x})(y_i - \bar{y}) & \sum (x_i - \bar{x})(z_i - \bar{z}) \\ \sum (x_i - \bar{x})(y_i - \bar{y}) & \sum (y_i - \bar{y})^2 & \sum (y_i - \bar{y})(z_i - \bar{z}) \\ \sum (x_i - \bar{x})(z_i - \bar{z}) & \sum (y_i - \bar{y})(z_i - \bar{z}) & \sum (z_i - \bar{z})^2 \end{pmatrix}. \quad (2.3)$$

Затем, для полученной коллекции выделенных компонент NRM находится среднее направление (средний вектор). Обычно, на данном этапе используется распределение Фишера.

Распределение Фишера — аналог распределения Гаусса на сфере, то есть оно осесимметрично относительно своего генерального среднего направления. Главными параметрами этого распределения являются кучность k и радиус круга доверия α_{1-p} :

$$k = \frac{N - 1}{N - R}, \quad (2.4)$$

$$\alpha_{1-p} = \cos^{-1} \left(1 - \frac{N-R}{R} \left(\left(\frac{1}{p} \right)^{\frac{1}{N-1}} - 1 \right) \right) \quad (2.5)$$

где p — вероятность того, что среднее направление рассматриваемого набора данных лежит за пределами круга доверия (обычно принимают $p = 0.05$), N — число рассматриваемых направлений, а R — модуль вектора $\mathbf{R} = \sum_{i=1}^N \mathbf{j}_{ni}$, где \mathbf{j}_n — единичный вектор для \mathbf{J}_n .

Если же имеется смешанный набор данных — выделенных направлений и кругов перемангничивания, тогда вместо обычного распределения Фишера необходимо воспользоваться его модификацией [McFadden, 1988]. Суть модификации заключается в том, что итеративно производится поиск такого направления вдоль каждого из кругов перемангничивания, которое максимизирует длину искомого среднего направления.

На основе этих распределений и их параметров были созданы палеомагнитные статистические тесты, рассматриваемые далее.

2.2. Палеомагнитные статистические тесты

2.2.1. Тесты значимости

Используя распределение Фишера мы можем достаточно точно охарактеризовать наши данные, в частности мы можем узнать:

1. Носит ли распределение случайный характер?
2. Насколько один набор данных лучше сгруппирован, чем другой?

Разрешаются эти вопросы проведением т. н. тестов значимости, суть которых заключается в проверке нужной нам нулевой гипотезы. В ходе теста производится сравнение выборки с заранее заданной нулевой гипотезой, причём производится оно только на некотором уровне значимости — вероятностном уровне, на котором любые различия между выборкой и генеральной совокупностью носят исключительно случайный характер. Например, если уровень значимости составляет 5% и тест успешно пройден, то вероятность случайного соответствия выборки генеральной совокупности составляет 5%, а с вероятностью 95% выборка является частью рассматриваемой генеральной совокупности.

Тест на случайность

Имея средний вектор для выборки по Фишеру \mathbf{R} и зная размер выборки N можно понять, носит ли рассматриваемое распределение случайный характер. Тест Ватсона [Watson, 1956] заключается в расчёте параметра $R_0 = \sqrt{7.815 \cdot N/3}$ и сравнении его с модулем среднего направления выборки $R = \sqrt{(\sum_i x_i)^2 + (\sum_i y_i)^2 + (\sum_i z_i)^2}$. Если $R > R_0$, тогда нулевая гипотеза о случайности распределения может быть принята только на рассматриваемом уровне значимости, в противном случае распределение однозначно случайно.

Сравнение кучностей

Рассмотрим 2 выборки с параметрами N_1, k_1 и N_2, k_2 соответственно. Возьмём за нулевую гипотезу идею, что эти выборки принадлежат генеральным совокупностям с одинаковыми кучностями k , тогда соотношение k_1/k_2 будет отличаться из-за ошибок отбора образцов следующим образом [McElhinny, 1964]:

$$\frac{k_1}{k_2} = \frac{\text{var}(2(N_2 - 1))}{\text{var}(2(N_1 - 1))}, \quad (2.6)$$

где $\text{var}(2(N_2 - 1))$ и $\text{var}(2(N_1 - 1))$ — дисперсии с $2(N_2 - 1)$ и $2(N_1 - 1)$ степенями свободы. Это соотношение должно согласовываться с F-распределением, если предположение о генеральных совокупностях с одной кучностью k верное. Чтобы подтвердить это, соотношение должно быть близко к 1. А о том насколько сильно может отклониться оно от единицы, прежде чем это отклонение станет статистически значимым, позволяют судить таблицы для F-распределения.

Тест на общее среднее

Пусть имеется два средних направления, полученных как средние по Фишеру, и необходимо понять, насколько статистически значимо различие между ними. Наиболее простой и известный метод проверки значимости в этом случае — вариация F-теста, предложенная Ватсоном [Watson, 1956]. Рассмотрим две выборки единичных векторов размерами N_1, N_2 и со средними направлениями длинами R_1, R_2 соответственно. Тогда F-статистика имеет вид:

$$F = (N - 2) \frac{R_1 + R_2 - R}{N - R_1 - R_2}, \quad (2.7)$$

где $N = N_1 + N_2$, а R — среднее направление для объединённых выборок. Полученная F-статистика сопоставляется с табличными значениями для 2 и $2(N - 2)$ степеней свободы. Если она превышает табличное значение на выбранном уровне значимости, тогда рассматриваемые средние направления различны на этом уровне значимости.

Также существует альтернативный и более корректный в большинстве случаев способ определения, имеют ли два рассматриваемых набора данных общее среднее направление — V_w -тест Ватсона [Watson, 1983].

Пусть $R_1, R_2, (x_{11}, x_{12}, x_{13}), (x_{21}, x_{22}, x_{23})$ и k_1, k_2 — длины средних векторов, их декартовы координаты и кучности для двух выборок размерами N_1 и N_2 соответственно. Тогда средневзвешенная сумма и средневзвешенные значения вычисляются следующим образом:

$$S_w = k_1 R_1 + k_2 R_2, \quad (2.8)$$

$$\hat{X}_j = k_1 R_1 \bar{x}_{1j} + k_2 R_2 \bar{x}_{2j}, \quad (2.9)$$

где $j = 1, 3$, а $\bar{x} = \frac{1}{R} \sum_{i=1}^N x_{ij}$ — компонента среднего вектора. Зная средневзвешенные значения по всем координатным осям, можем найти средневзвешенное направление:

$$R_w = \sqrt{\hat{X}_1^2 + \hat{X}_2^2 + \hat{X}_3^2}, \quad (2.10)$$

И тогда величина V_w -статистики вычисляется как:

$$V_w = 2(S_w - R_w). \quad (2.11)$$

Таким образом, V_w -статистика растёт с ростом разницы между двумя рассматриваемыми направлениями. И если величина V_w -статистики превышает некоторое критическое значение $V_{w|crit}$, тогда нулевая гипотеза об общем среднем для рассматриваемых наборов данных отклоняется. Критическое значение при этом можно найти используя метод Монте-Карло. Алгоритм нахождения критического значения следующий:

1. создаются два набора данных размерами N_1, N_2 и кучностями k_1, k_2 , но с одинаковым средним направлением (данные соответствуют распределению Фишера);
2. рассчитывается V_w -статистика для созданных наборов данных;
3. шаги 2–3 повторяются необходимое число раз (например, $Q = 1000$);
4. полученное распределение V_w -статистик сортируется по возрастанию;
5. находится критическое значение $V_{w|crit}$, которое будет являться пороговым для установленного уровня значимости: пусть уровень значимости равен s , тогда критическое значение V_w будет $(1 - s)Q$ по счёту (с учётом сортировки);

Затем, рассчитав V_w -статистику для исследуемых данных, остаётся лишь сравнить её с $V_{w|crit}$. Например, пусть уровень значимости составляет 5%, и если найденное V_w превысит 95% сортированных по возрастанию симулированных V_w , тогда можно считать, что с вероятностью 95% рассматриваемые наборы данных не имеют общего среднего направления.

2.2.2. Полевые тесты

Всегда важно понимать, когда изучаемый образец приобрёл остаточную намагниченность. Для этого были созданы тест обжига (контакта), тест галек (конгломератов) и тест складки. В этой работе рассматриваются только последние два теста.

Тест галек

Тест галек позволяет датировать намагниченность изучаемого образца относительно момента переотложения породы в виде конгломератов. Для проведения теста необходимо удостовериться в следующем:

1. Гальки конгломератов представлены именно той породой, намагниченность которой изучается (внутриформационные конгломераты);
2. Намагниченность галек однородна по направлению (высокая кучность);

Сам по себе тест при этом является, по сути, проверкой того, носит ли распределение направлений намагниченности галек в пространстве случайный характер (равновероятно в пространстве). И, таким образом, математически полностью сводится к тесту Ватсона на случайность. Интерпретация у результатов теста однозначная: если распределение направлений намагниченности галек не случайное, тогда они приобрели намагниченность после переотложения в виде конгломератов.

Тест складки

Тест складки заключается в изучении истории складчатости геологического тела, из которого был отобран изучаемый образец. Он позволяет понять, когда была приобретена намагниченность - до или после смятия тела в складку. Для проведения этого теста достаточно отобрать образцы с обоих крыльев складки.

Математически этот тест сводится к сравнению кучностей рассматриваемого набора направлений до и после смятия в складку. Интерпретируются результаты теста следующим образом: если направления лучше сгруппированы до складчатости, чем после неё, значит намагниченность была приобретена после смятия в складку (тест считается пройденным); в противном случае результат противоположный.

В ходе теста сначала исследуется соотношение кучностей k до k_b и после k_a тектонических движений, а затем применяется тест значимости на сравнение кучностей, который позволяет понять, статистически значимо ли отличие соотношения k_a/k_b от 1. Если это соотношение превышало величину F-статистики для 5% уровня значимости, значит с вероятностью менее 5% увеличение k после тектонических движений не связано напрямую с тектоническими движениями, а является случайным. И, соответственно, это даёт 95% вероятность, что увеличение k после тектонических движений значимо и k_a/k_b значимо отличается от 1. Такой результат считается статистически значимым прохождением теста складки.

Тест обращения

Если набор палеомагнитных сайтов даёт приемлемое усреднение вековых вариаций в ходе интервалов как с нормальной полярностью, так и с обратной, и при этом не было приобретено вторичных компонент NRM, тогда среднее направление первичной NRM для сайтов с нормальной полярностью должно быть антипараллельным (антиподальным) к аналогичному среднему направлению для сайтов с обратной полярностью.

Тест обращения проверяет, насколько статистически значимо отклонение угла между направлениями с сайтов прямой и обратной полярности от 180° . Математически эта процедура

полностью может быть описана тестом Ватсона на сопоставление средних направлений, для этого достаточно в качестве двух наборов данных взять направления прямой и обратной полярности. При этом важно учитывать, что корректно применять этот тест только для Фишеровских распределений.

2.3. Метод Bootstrap

Многие имеющиеся палеомагнитные данные часто имеют распределения не осесимметричные и не бимодальные, их невозможно корректно описать распределением Фишера, а порой и распределением Бингхэма. Как следствие, необходимо найти непараметрический способ описания таких данных. Таким подходом будет применение метода бутстрэп (bootstrap) [Efron, 1979], который отлично подходит для палеомагнитных данных [Tauche, 1991].

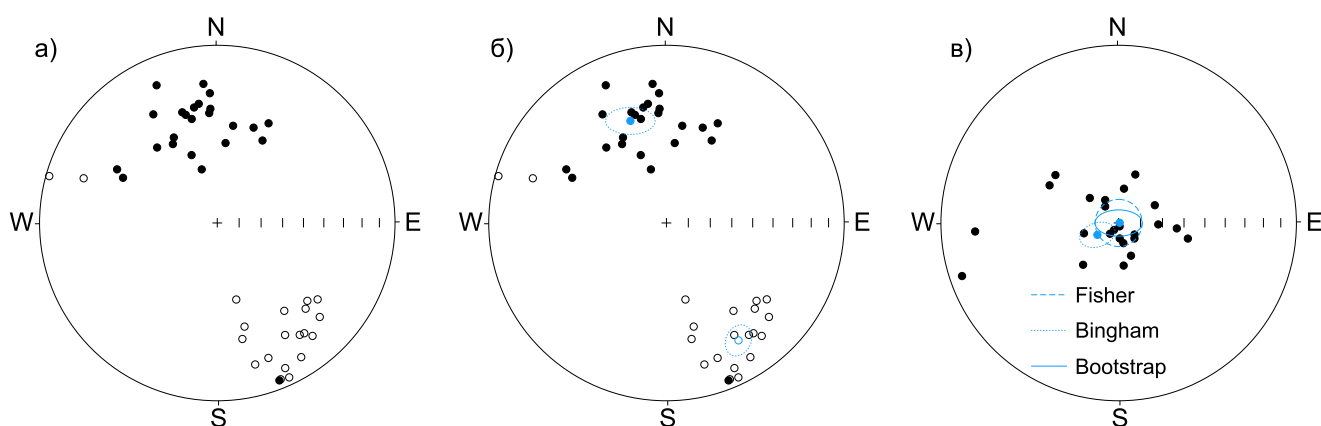


Рис. 2.1: а) некоторая выборка, не являющаяся бимодальной и осесимметричной одновременно б) те же данные, что и в а), но с рассчитанными средними по Бингхэму; в) Сравнение 95% доверительных областей, полученных по Фишеру, Бингхэму и с использованием бутстрэпа. Данные те же, что и в а) и б), но повернутые таким образом, что собственные векторы ориентационной матрицы сумм и квадратов служат осями: главное направление ориентированно по центру стереограммы, максимальное и минимальное — EW и NS соответственно [Tauche, 1991].

Допустим, что имеется набор сложным образом распределённых палеомагнитных векторов, размер выборки N и необходимо найти её среднее. Напрямую применить к этой выборке какое-либо известное распределение нельзя, поскольку любое из них будет плохо его описывать. В таком случае можно использовать бутстрэп:

1. случайным образом выбираются N векторов из выборки, причём каждый вектор может быть выбран неоднократно;
2. для полученного нового распределения рассчитывается средний вектор по Фишеру;
3. шаги 1–2 повторяются необходимое число раз (чем больше повторений, тем лучше будет итоговое распределение), сохраняя каждый раз результат шага 2;

4. полученное распределение средних принимается соответствующим распределению Бингхэма и, исходя из этого, рассчитываются все его параметры.

Пример применения описанного метода представлен на рис. 2.1.

2.3.1. Модификации тестов

Существует множество различных модификаций палеомагнитных тестов, но почти все из них так или иначе отходят от заложенных в основу этих тестов идей. Например, существуют как минимум 8 известных модификаций теста складки, причём все они уже сравнивались между собой в работе [Enkin, 2003]. И, по озвученной выше причине, все эти тесты и любые другие подобные им нельзя назвать классическими. Более того, в настоящее время ни один из таких тестов не используется широко, а большая их часть не используется вовсе. Тем не менее, существуют такие их модификации [Таухе, 2010], которые не изменяют базовые идеи тестов, а только лишь дают новые статистические способы их описания, лишённые недостатков тех способов, которые ранее использовались в качестве классических. К тому же, именно эти модификации тестов сейчас наиболее широко используются палеомагнитологами, являясь частью широко используемого за рубежом пакета анализа палеомагнитных данных PmagPY [Tauche et al., 2016].

Модификация теста складки

Классический тест складки, описанный ранее, имеет ряд недостатков:

1. исследуемые направления могут быть двух полярностей и при этом разбросаны по сфере;
2. исследуемые направления могут быть лучше всего сгущены не точно в географический или стратиграфической системе координат (не до и после смятия в складку), а где-то «между» (неполное «распрямление» складки);
3. все структурные изменения, приведшие складку к настоящему виду, доподлинно неизвестны и в большинстве случаев не могут быть известны в принципе.

Все эти недостатки не позволяют применять классический тест складки ко многим палеомагнитным данным, особенно к не соответствующим распределению Фишера. Чтобы избежать недостатков, связанных с распределением Фишера, вместо классической кучности можно использовать её аналог из симметрической матрицы сумм и квадратов:

$$\mathbf{T} = \begin{pmatrix} \sum x_i^2 & \sum x_i y_i & \sum x_i z_i \\ \sum x_i y_i & \sum y_i^2 & \sum y_i z_i \\ \sum x_i z_i & \sum y_i z_i & \sum z_i^2 \end{pmatrix}, \quad (2.12)$$

Здесь роль кучности выполняют относительные величины собственных значений λ : с ростом сгруппированности («кучности»), дисперсия вдоль главной оси возрастает, а вдоль других уменьшается. Таким образом, исследование величины λ_1 в ходе раскрытия складки позволит определить момент, когда достигается наибольшая сгруппированность направлений. И это сразу же решает второй из упомянутых выше недостатков: можно постепенно, с необходимой «скоростью» раскрывать складку, что позволяет выделить конкретный момент, когда была максимальная сгруппированность направлений.

Теперь, чтобы понять, является ли полученный результат статистически значимым, можно воспользоваться бутстрэпом. Рассчитав λ_1 необходимое число раз и построив его кумулятивное распределение, выделяем граничные значения процентов распрямления складки, внутри которых находится 95% распределения. И если эти границы включают в себя момент 100% распрямления складки, значит тест складки для рассматриваемых данных пройден.

Модификация теста на общее среднее

В первую очередь стоит обобщить V_w -тест на случай распределения, не соответствующего фишеровскому. Для этого вместо метода Монте-Карло необходимо воспользоваться бутстрэпом. При этом все средние направления переводятся в декартовы координаты и строятся графики их кумулятивного распределения. Для каждого распределения находим границы, содержащие 95% всех средних. Если хотя бы по одной из осей координат границы распределений не будут пересекаться, значит, с вероятностью 95%, у рассматриваемых наборов данных средние направления различаются.

Модификация теста обращения

Для модификации теста достаточно воспользоваться модифицированным тестом на общее среднее и взять в качестве двух наборов данных направления, соответствующие прямой полярности, и направления, соответствующие обратной.