

# Использование заметающих кривых для построения геопространственного индекса

Автор: Башков Алексей Александрович,  
Инженер, НИИ ТП

# Общий алгоритм поиска пересечения в геопространственном индексе.

- 1) все имеющиеся данные (полигоны и точки) кодируются совокупностью идентификаторов ячеек, которыми покрываются геометрии для поиска;
- 2) во время запроса на поиск запрашиваемая геометрия также покрывается ячейками и формируется множество идентификаторов по которым выполняется поиск;
- 3) идентификатор ячейки, обычно, является целым числом (индексом ячейки при обходе по кривой, заметающей пространство), то поиск соответствия выполняется намного быстрее. В таком случае используется либо выборка по значению, либо выборка по диапазону.

Задача нумерации ячеек может решаться с помощью кривых, заметающих пространство.

Вариант нумерации 1

1	4	7	10
2	5	8	11
3	6	9	12

Вариант нумерации 2

1	2	3	4
8	7	6	5
9	10	11	12

# Стандартные заметающие кривые

Основным свойством заметающих кривых является преобразование двумерного / трехмерного пространства в одномерное. Это позволяет использовать более эффективные алгоритмы поиска.

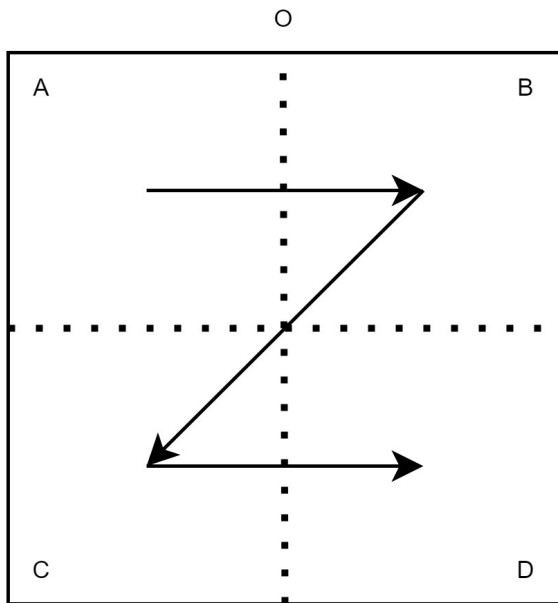
Пространственная локальность данных – близкие индексы на заметающей кривой означают пространственную близость двумерных областей.

При построении геопространственных индексов наиболее часто используются следующие заметающие кривые:

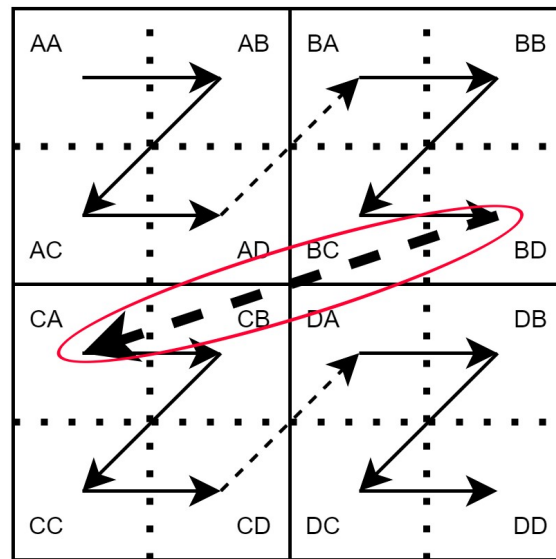
- кривая Лебега (Z-кривая, кривая Мортон)
- кривая Гильберта
- H-кривая

и некоторые другие.

# Кривая Лебега (Z-кривая)



а) деление пространства на первом шаге

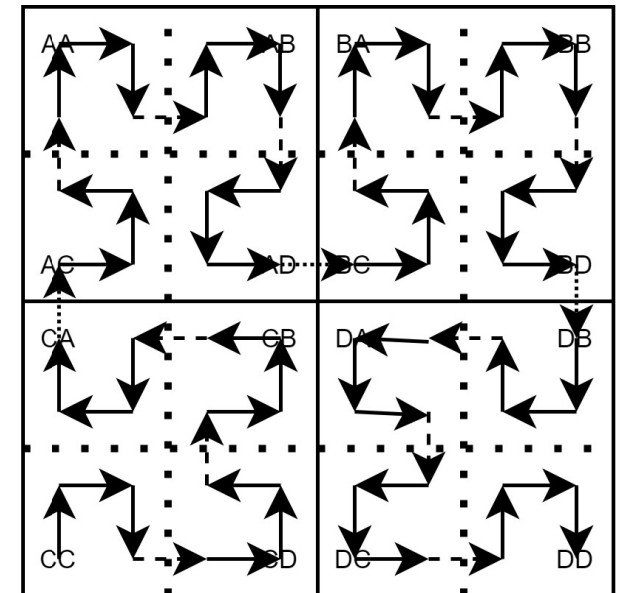
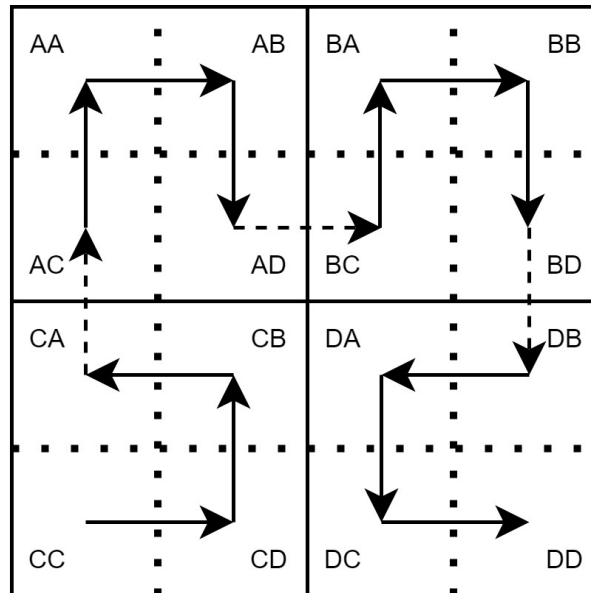
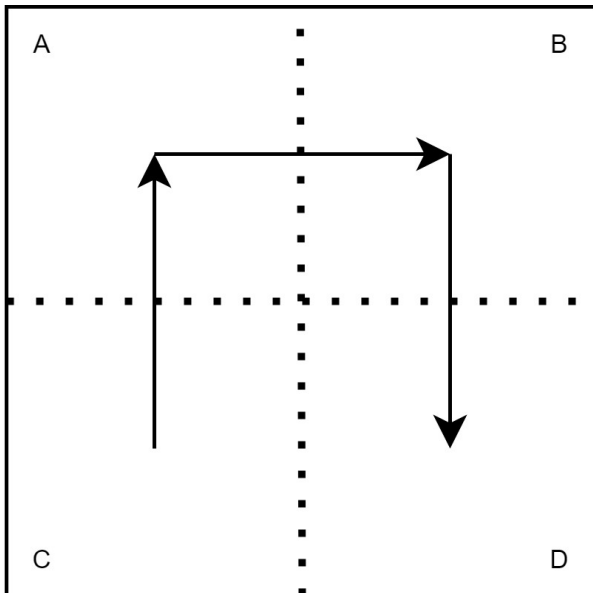


б) демонстрация деления на следующем шаге

В некоторых геопространственных индексах (Bing Maps) используется кривая Лебега, в которой базовой структурой является Z-образный обход четырех ячеек. Одним из недостатков которой является длинный переход между ячейками BD и CA. Расстояние между которыми в пределе доходит до ширины области определения.

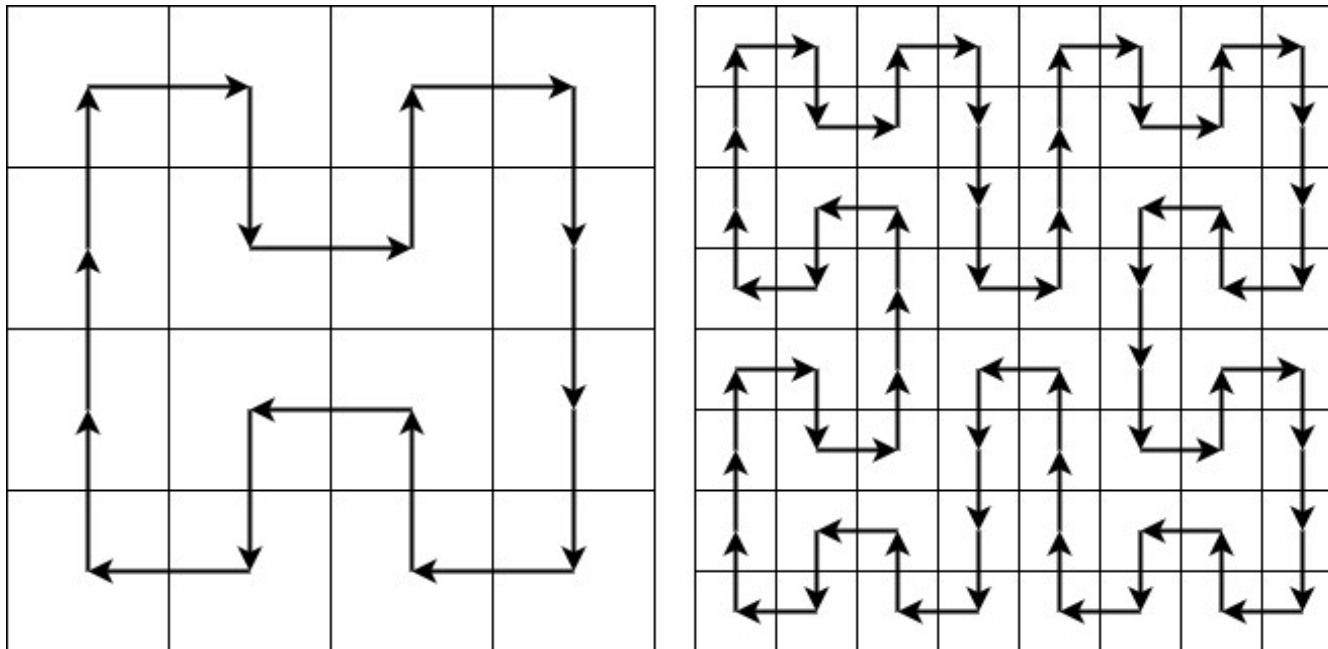
# Кривая Гильберта

Используется во многих современных индексах (например в S2Geometry), является функцией заметающей квадрат. Важная особенность - смежные ячейки на кривой, являются смежными и на заметаемой области.



# H-кривая

Разработана в 2020 году Игорем Нетаем. Смежные ячейки на кривой, являются смежными и на замкнутой области, кроме этого является циклической (начало и конец кривой являются смежными в пространстве.)



# Особенности стандартных заметающих кривых

- 1) заметаемая область является квадратом;
- 2) в случае с кривой Лебега (Z-кривая), имеет место нарушение локальности данных: когда области с соседними номерами на кривой не являются смежными на заметаемой области;

# Особенности данных ДЗЗ, которые не учитываются стандартными заметающими кривыми

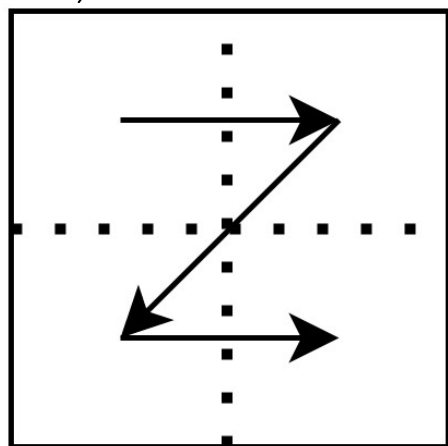
- 1) стандартные заметающие кривые не учитывают географические особенности заметаемой области (в области построения геопространственных индексов)
- 2) в зависимости от проекции заметаемая область является квадратом, прямоугольником или кругом
- 3) поверхность планеты неоднородна (есть океаны, материки, густо- и малозаселенные области)
- 4) имеются широкораспространенные проекции (например UTM) которые не обладают регулярным способом деления на зоны. Как правило, число зон по осям не равно степени двойки.



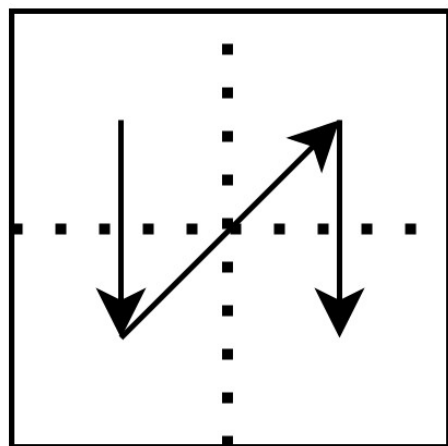
# Предлагаемый способ построения заметаящих кривых

Прямой и обратный алгоритм преобразования 2-мерного пространства в 1-мерное, а также эффективный алгоритм построения заметаящих кривых, опубликованы в журнале «Естественные и технические науки» в 2024 году.

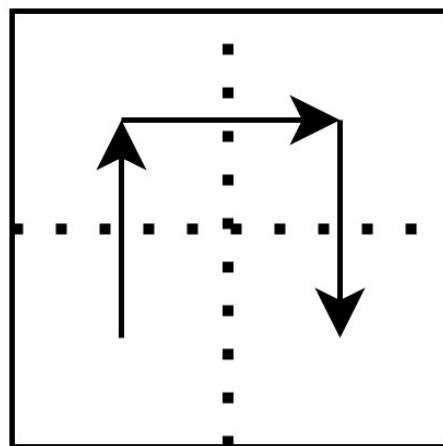
Базовыми элементами построения произвольной кривой являются 4 типа блоков Z, И, П, U:



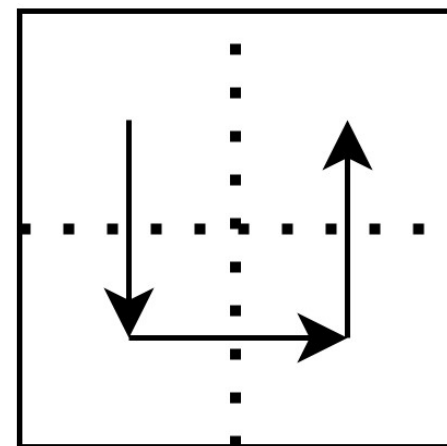
Блок Z



Блок И



Блок П



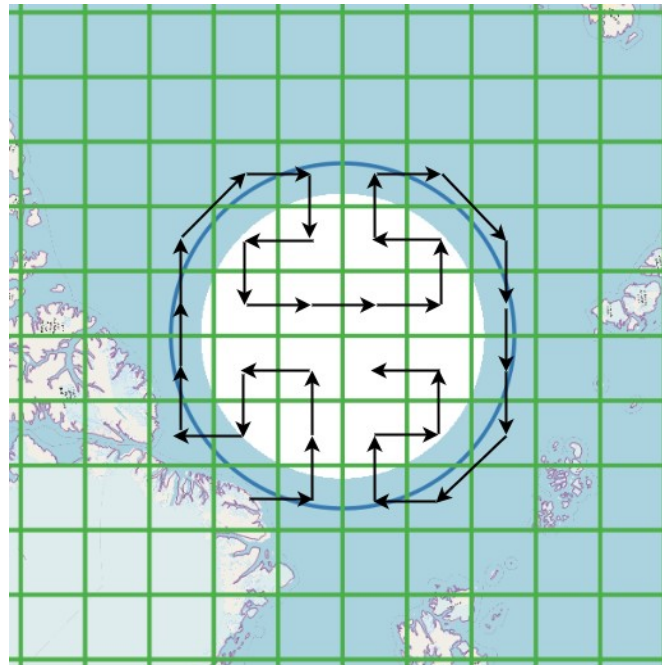
Блок U

Эти блоки могут быть повернуты на углы 0, 90, 180, 270 градусов

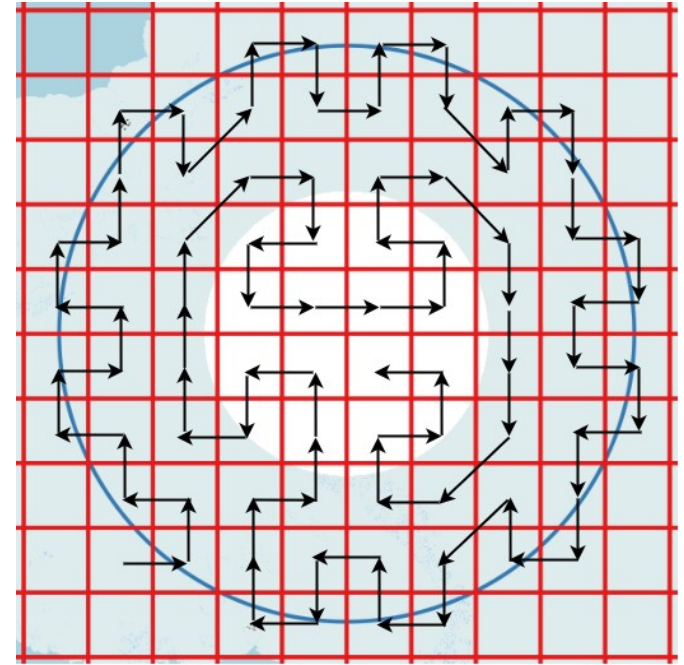
# Базовые кривые для обхода полярных областей

Ниже представлены варианты обхода полярных областей, которые обладают следующими свойствами:

- 1) конец (или начало) обхода находится на полюсах;
- 2) кривая для южной полярной области включает в себя кривую для северной;
- 3) шаг сетки един и составляет 250 000 м в проекциях EPSG:5041 и EPSG:5042

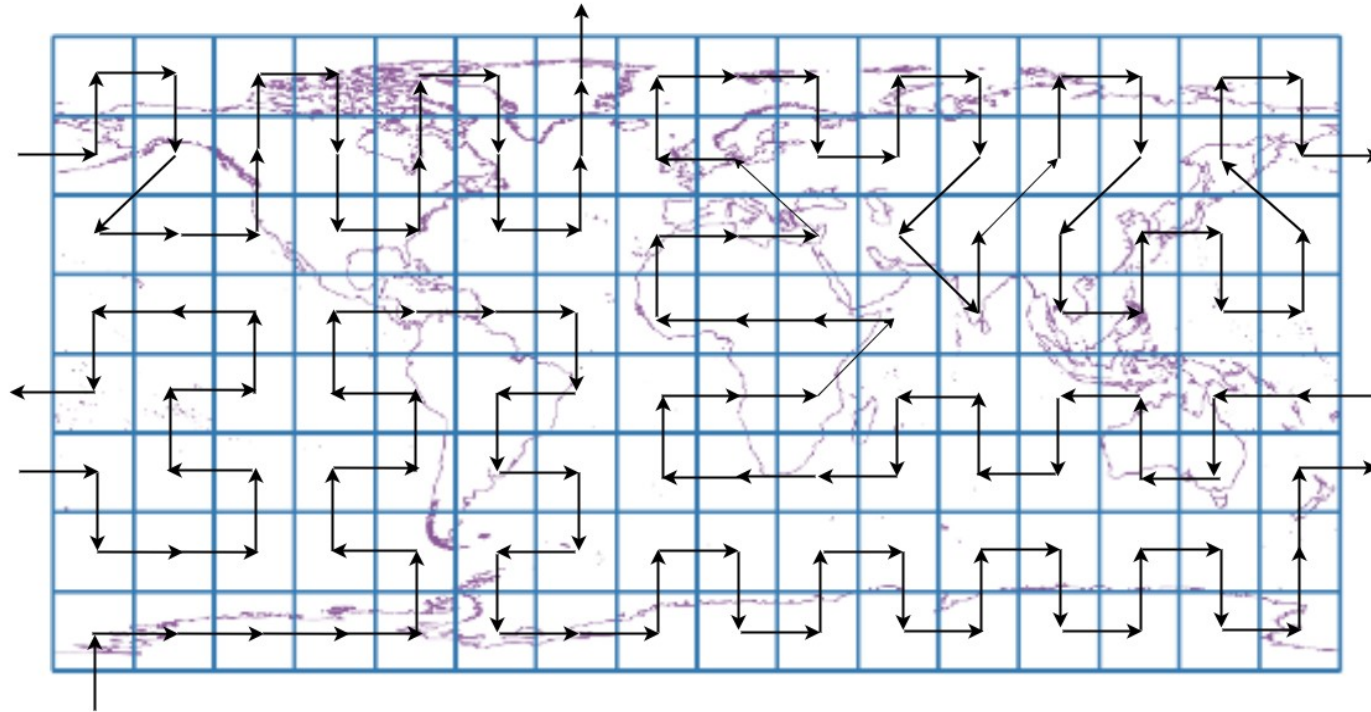


Северная полярная область



Южная полярная область

# Вариант построения базовой кривой для геодезической СК



Особенностью данного обхода является то, что участки заселенных континентов, являются близко расположенными на кривой. Вариант ручного заполнения.

# Особенности зон UTM. Способ построения регулярной сетки, совместимой с UTM.

Зоны над северной Европой имеют размеры (градусы по долготе на градусы по широте):

X31: 9 на 12,

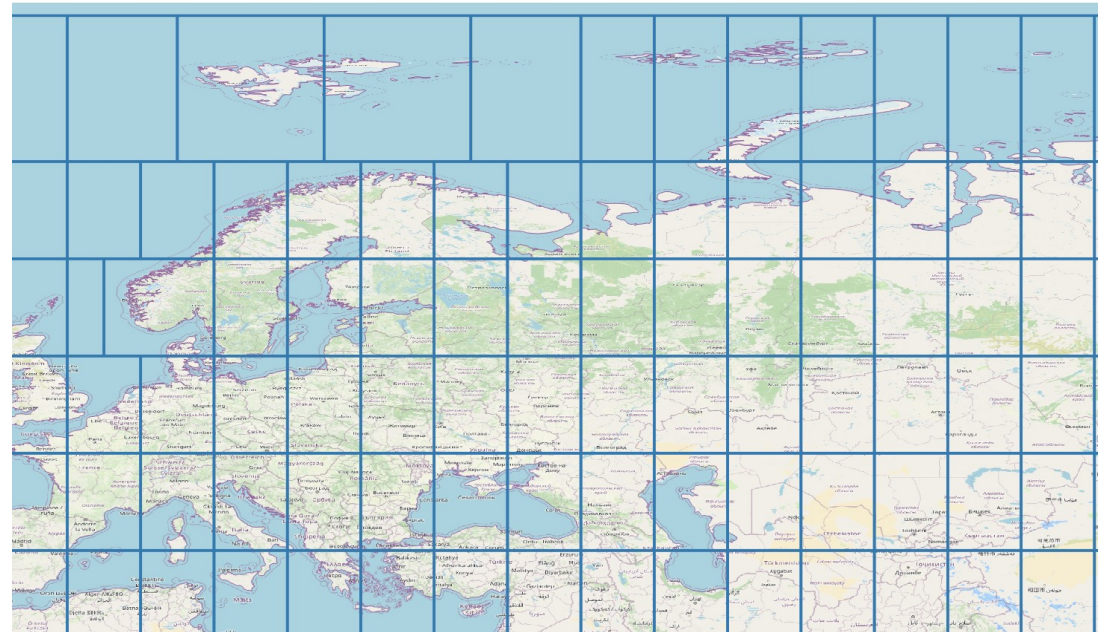
X33: 12 на 12,

X35: 12 на 12,

X37: 9 на 12,

V31: 3 на 8,

V32: 9 на 8.

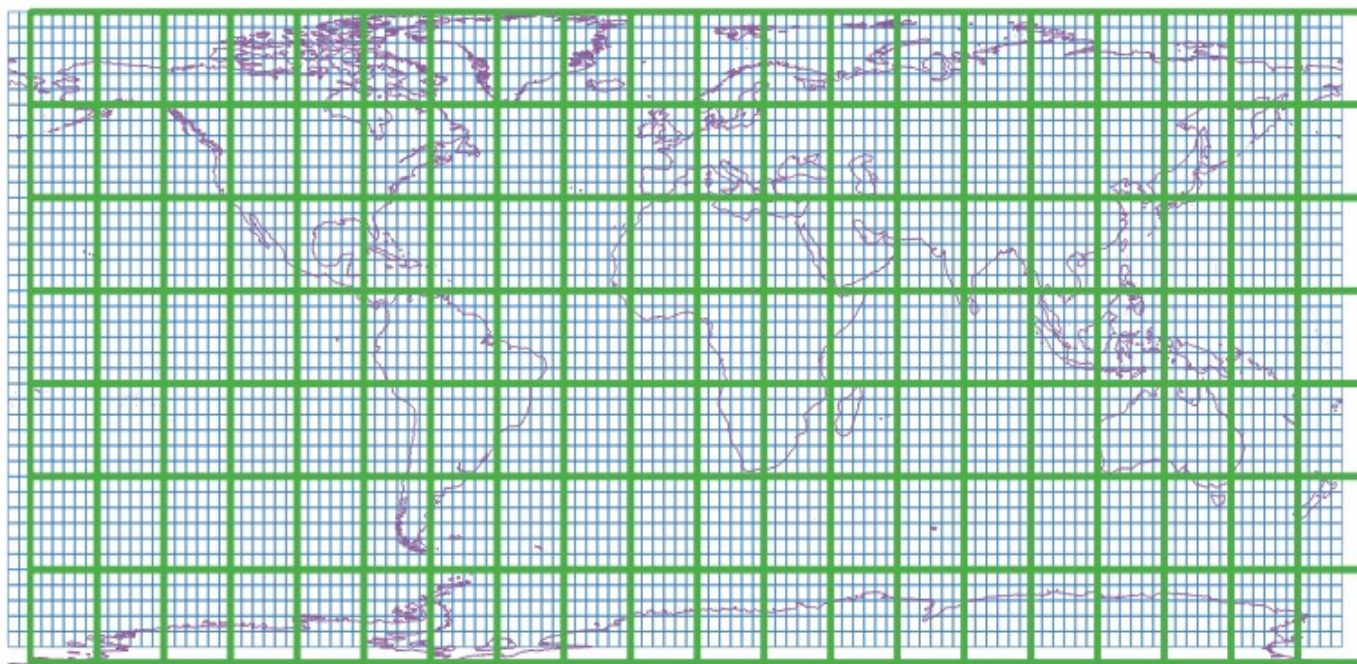


Для того, чтобы устранить эти исключения, можно перейти к сетке 3 на 4 градуса.

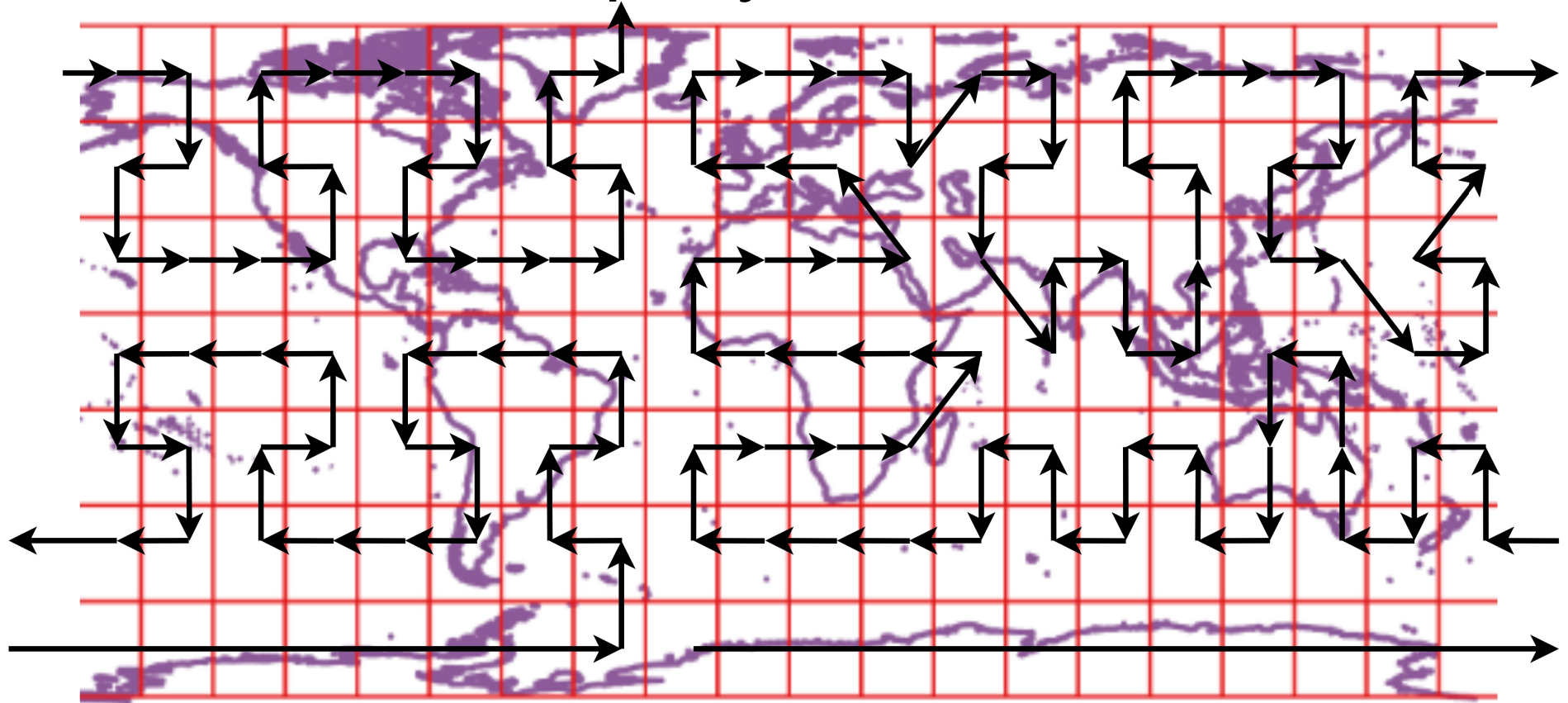
# 3x4-градусная глобальная сетка с блоками по 6x6 ячеек

После разбиения всей поверхности Земли на ячейки меньшего размера требуется выполнить группировку ячеек на блоки.

В данный момент выполняется исследование возможности группировки по 36 ячеек, что позволит с одной стороны упростить подбор варианта обхода блоков, а с другой – оставить некоторую свободу выбора обхода ячеек внутри блока.



# Вариант обхода блоков ячеек 3x4-градусной сетки



# Вариант обхода блока из ячеек 3x4-градусной сетки

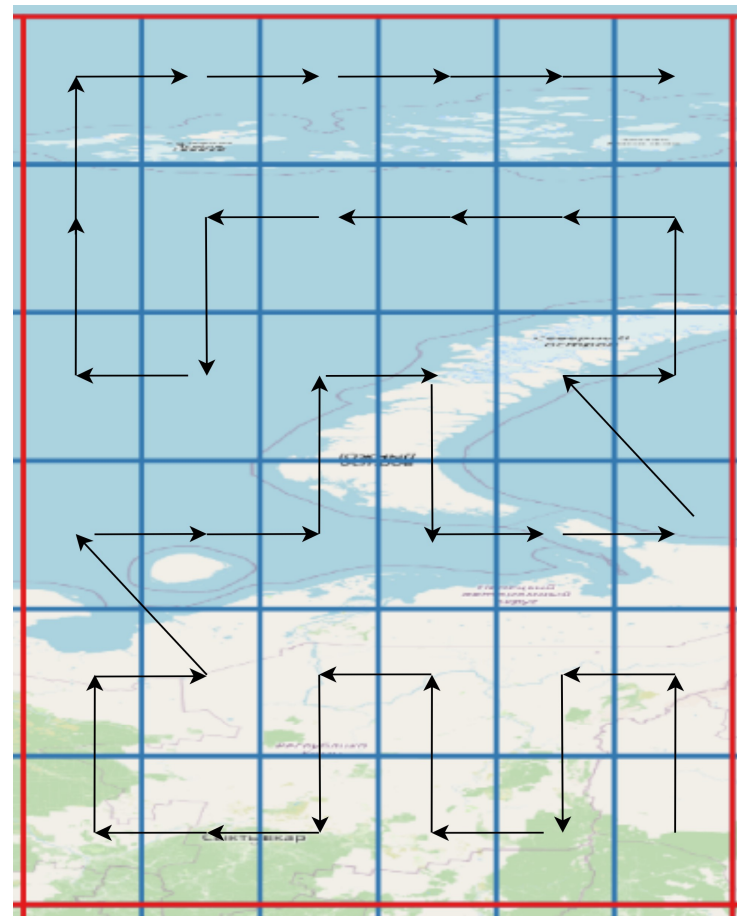
На рисунке продемонстрирован ручной вариант обхода ячеек 3 на 4 градуса объединенный в блок 6 на 6 ячеек.

Шаблон типа ПІ.

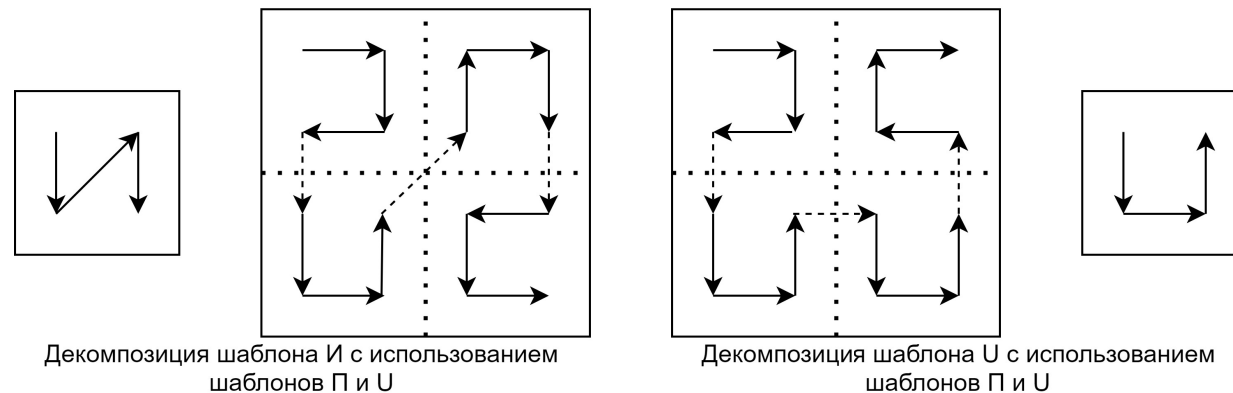
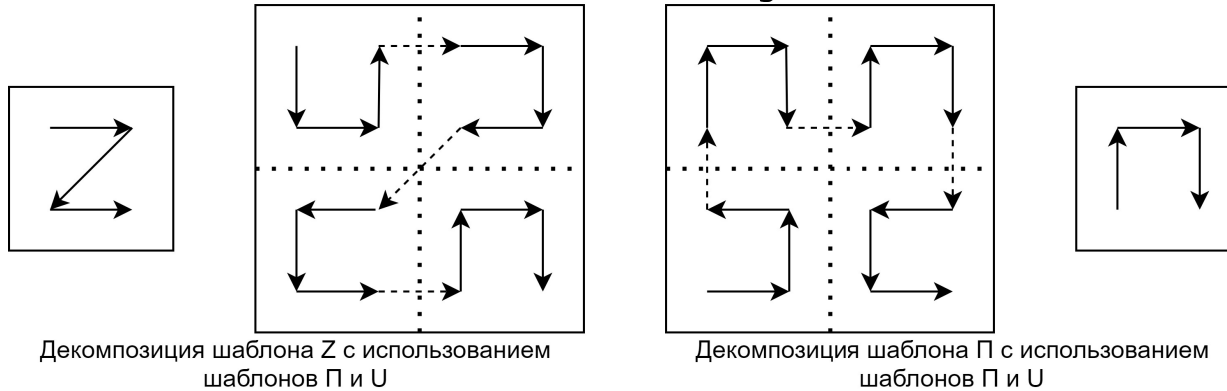
Имеется как минимум 3 способа отбора вариантов:

- 1) критерии симметрии кривой и ее участков;
- 2) критерий пересечения по суше для смежных ячеек на ломаной;
- 3) критерий минимума числа пересечений границ государств.

Данные критерии позволяют автоматизировать выбор шаблонов для заметания разных блоков.



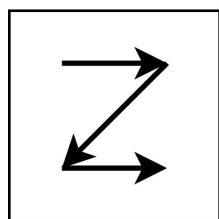
# Декомпозиция четырех базовых шаблонов на 4 дочерние области размером 2 на 2, используя шаблоны П и U



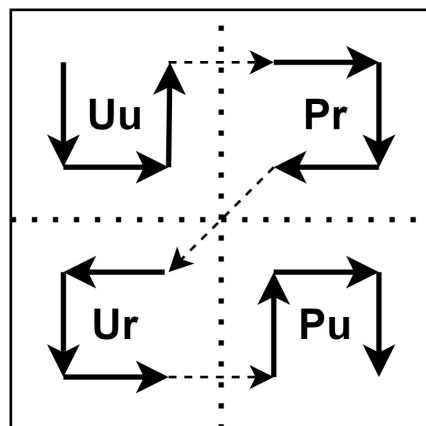
Представленный способ декомпозиции может использоваться на уровнях деления ячеек на 4 дочерние области с конфигурацией 2 на 2 и минимизацией числа используемых шаблонов. После иерархического перехода остаются только шаблоны П и U



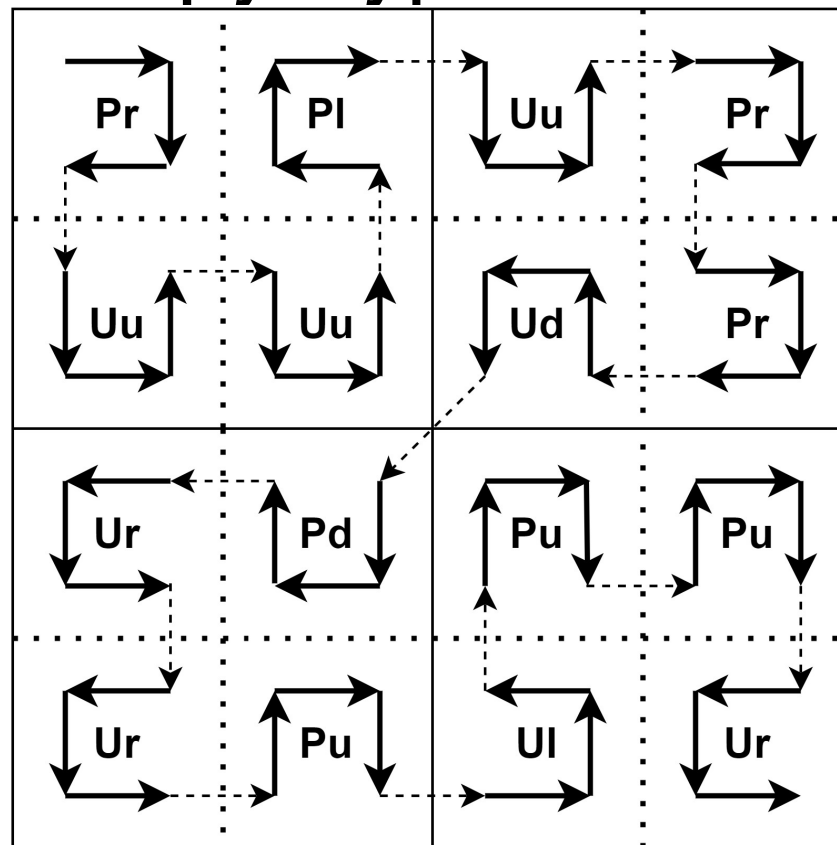
# Демонстрация двух последовательных иерархических переходов при делении на дочерние области со структурой 2 на 2



Базовый шаблон Z



Уровень декомпозиции 1



Уровень декомпозиции 2

# Заключение

Предложен способ индексации данных в различных проекциях, и область индексации отличной от квадрата. Способ обладает большой вариативностью в зависимости от интересующих параметров заметаемой области.

Геопространственные индексы, построенные с использованием ZIPU-кривых, позволяют на ранних этапах фильтрации получать информацию о континенте, позволяют упростить запрос на выборку информации.

Предложен способ оценки площади региона по покрытию, построены таблицы зависимости площади от уровня ячейки и её индекса по широте.

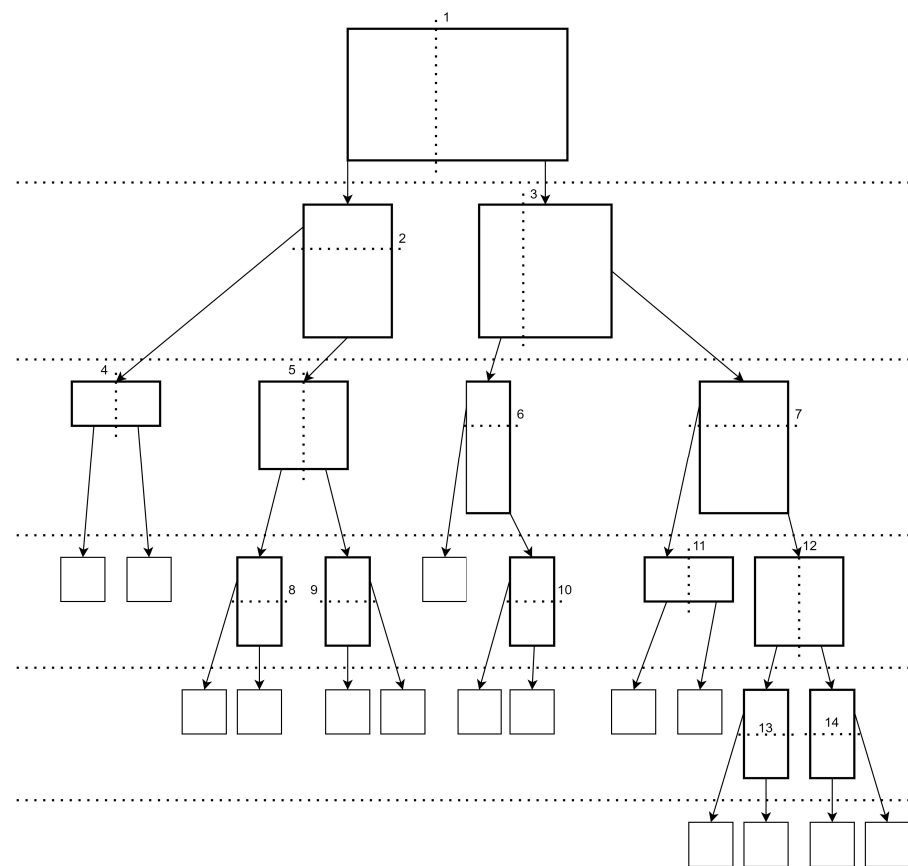
Спасибо за внимание!

# Иерархическое деление ячеек

Наиболее вычислительно эффективным является деление ячейки на 4 части, поэтому, начиная с определенного уровня детализации, при дальнейшем делении для достижения определенной точности покрытия лучше использовать степени двойки.

Однако, необходимо продемонстрировать иерархическое деление нестандартной области, например 3 на 5.

Для вычисления покрытия произвольной области сеткой с размерами 3 на 5 ячеек требуется выполнить до 28 пересечений с прямоугольником.



# Идея вычисления оценки площади по покрытию

Способ вычисления оценки площади по покрытию (без вычисления процента пересечения). Для регулярных сеток.

При получении покрытия региона ячейками мы знаем индекс/номер ячейки по оси широт. Если ячейка полностью входит в регион, то мы знаем точную площадь данной ячейки. Можем ограничить площадь региона снизу. Если учитывать ячейки не полностью входящие в регион, то мы можем ограничить площадь региона сверху.

$$S_{\min} = S_2 \cdot 4 + S_3 \cdot 4 + S_4 \cdot 4 + S_5 \cdot 4$$

$$S_{\max} = S_1 \cdot 6 + S_2 \cdot 6 + S_4 \cdot 6 + S_5 \cdot 6 + S_6 \cdot 6$$

