

Способ построения геопространственного индекса, использующего ZIPU-кривые.

Автор: Башков Алексей Александрович
НИИ ТП

ZIPU-кривые – класс кривых, заметающих пространство, использующий шаблоны Z, И, П, У.

Описаны в работах:

- 1) ОБОБЩЕННЫЙ АЛГОРИТМ ПОСТРОЕНИЯ ЗАМЕТАЮЩИХ КРИВЫХ. Башков А.А. Естественные и технические науки. 2024. № 10 (197). С. 151-162.
- 2) Башков А.А. Использование заметающих кривых для построения геопространственного индекса // Материалы 22-й Международной конференции «Современные проблемы дистанционного зондирования Земли из космоса». Москва: ИКИ РАН, 2024. С. 24. DOI 10.21046/22DZZconf-2024a

Основные шаги создания шаблонов

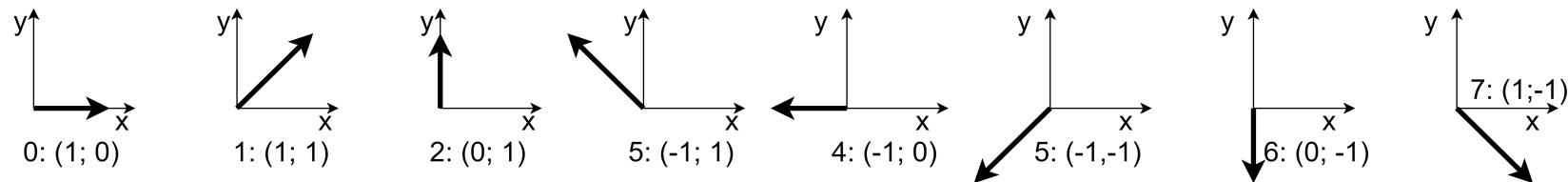
Основой эффективного использования ZIFU-кривых является подготовка шаблонов разного уровня декомпозиции с помощью построения базовых линий.

Алгоритм подготовки шаблона:

- I. Разделение области определения на блоки.
- II. Создание множества элементов базовой линии на разных уровнях детализации
- III. Построение базовой линии
- IV. Подбор элементов базовой линии, которые удовлетворяют критериям.

Базовые линии. Способ задания и кодирования.

Базовые линии задают обход соседних ячеек в произвольном порядке. Для двумерного случая должны кодироваться 8 направлений, для трехмерного – 26.



Оптимальный способ кодирования при сохранении и использовании - восьмеричное целое с ведущей единицей для обработки случая с ведущими нолями. Таким образом в 64разрядном двоичном числе удастся сохранить базовую линию состоящую из 21 участка.

Маски базовых линий (для областей до 8 на 8). Алгоритм создания

Сформировать два массива масок: по строкам L и по столбцам C

L[0]=0b00000000_00000000_00000000_11111111

L[1]=0b00000000_00000000_11111111_00000000

...

L[3]=0b11111111_00000000_00000000_00000000

C[0]=0b00000001_00000001_00000001_00000001

C[1]=0b00000010_00000010_00000010_00000010

...

C[7]=0b10000000_10000000_10000000_10000000

Для каждой базовой линии необходимо получить занимаемые координаты

XY =[(x1,y1),(x2,y2),...,(xN,yN)]

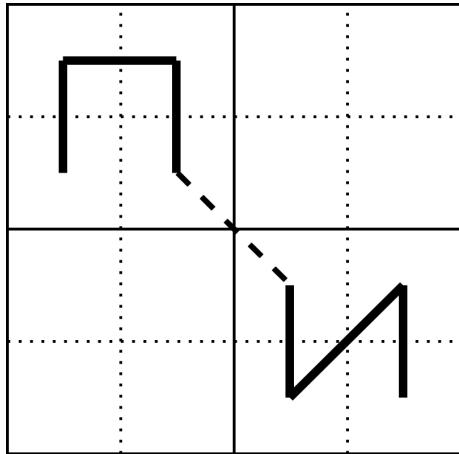
mask = 0

Для каждой пары x, y из XY:

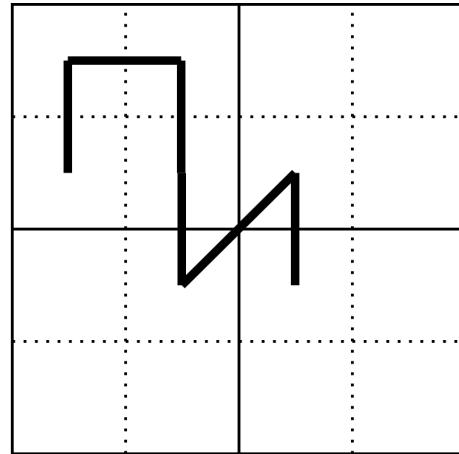
mask = mask ИЛИ (L[y] И C[x])

Способы соединения двух базовых линий

Базовые
линии для
символов
 Π (P_u) и И
(I_u)
 Π : 1206
И: 1616



Соединение с
промежуточным звеном:
 $1206+7+1616=12067616$



Соединение без
промежуточного звена:
 $1206+1616=1206616$

Соединение с промежуточным звеном используется при композиции нескольких базовых линий N длины L в базовую линию длиной L^*N . При создании базовых линий для блоков 6 на 6 использовалось 4 линии длиной 9.

Алгоритм проверки на совместимость двух базовых линий

x=0	x=1	x=2	x3
1	1	0	0
1	1	0	0
0	0	1	1
0	0	1	1

Пусть для базовой линии Π (P_u) – маска M_1 , а для базовой линии I (I_u) – M_2 ,
Тогда для возможности совмещения базовых линий должно выполниться 3 условия:
1) побитовая И для M_1 , M_2 равна 0
2) конечная точка для Π и начальная точка для I – смежные
3) отсутствует самопересечение у результирующей линии.

Определение критериев для отбора базовых линий

Для выбора оптимальных вариантов базовых линий для покрытия определенных участков планеты, были введены несколько критериев, учитывающие разные особенности, которые могут влиять на поиск данных.

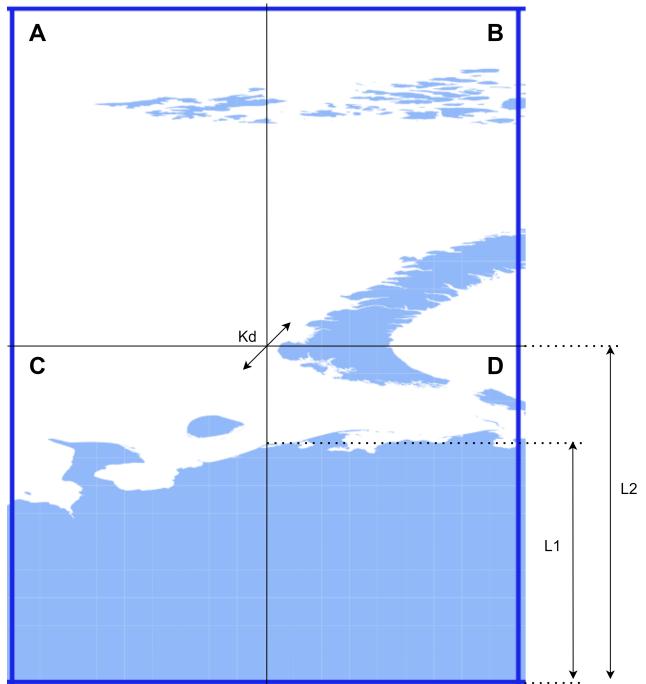
Некоторые из этих критериев обладают возможностью композиции/декомпозиции

$$K(La) + K(Lb) + K(Lab) = K(La+Lb),$$

La – базовая линия A, Lb – базовая линия B, Lab – переход от линии A к B, La+Lb – объединенная базовая линия

- 1) критерий максимального пересечения по суше (КПС);
- 2) критерий населения (КН);
- 3) критерий, использующий политическую карту мира (КПК);
- 4) критерий симметричности кривой (КС);
- 5) критерий минимизации суммарного номера соседних ячеек (КНС).

Критерий максимального пересечения по суше



Для перехода С-Д.

L_1 - длина линии пересечения с сушей в км.

L_2 - длина линии в км,

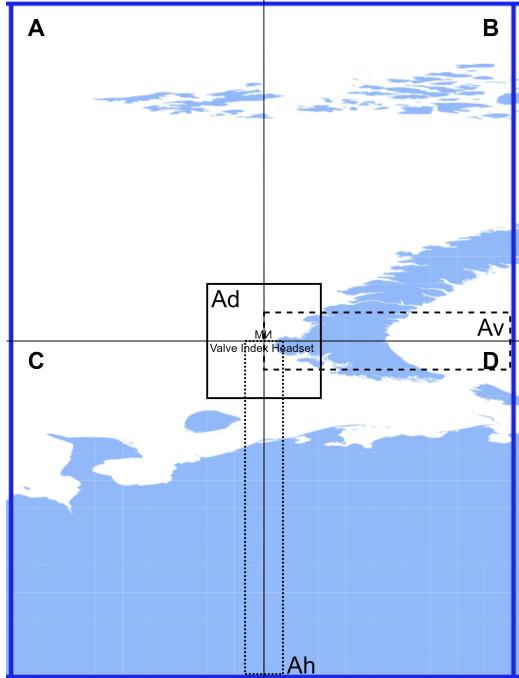
$NL = L_1/L_2$ (находится в диапазоне [0;1], большее-лучшее).

Но для этого критерия лучше перевести его в область начинающейся с 1, где 1-наилучшее значение:

$$K=1+(1-L_1/L_2)$$

Для диагонального случая - 2,5

Критерий населения



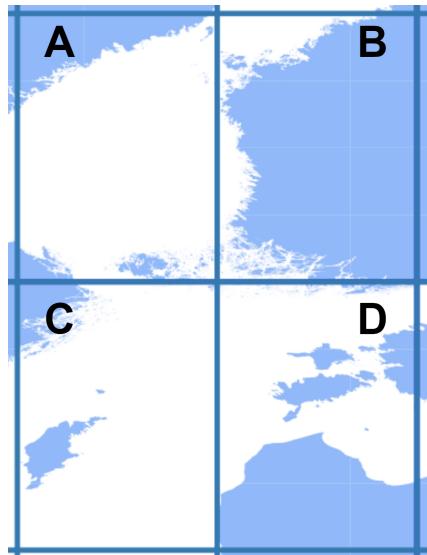
Между областями А, В, С, Д создаются области.
Возможны 2 варианта:

- 1) одинаковой площади $S(Ad) = S(Av) = S(Ah)$.
- 2) с одинаковым размером буфера, при этом
площади могут отличаться.

Зная плотность населения в этих областях,
можно вычислить примерную численность
населения в этих областях, которая и является
значением критерия.

Область значений критерия $[0; N]$, больше -
лучше.

Критерий, использующий политическую карту мира



Применение этого критерия основывается на том, что данные хранятся и обрабатываются с группировкой по странам. Область значений критерия $[0;N]$, где N -зависит от рассматриваемой области. Больше - лучше.

Для каждой области вычисляется пересечение с территориями стран.
Например, в квадратных градусах:

A: Aland-0.124, Sweden-1.969

B: Finland-8.972

C: Sweden-1.024

D: Latvia-2.969, Estonia-1.032, Finland-0.049, Lithuania-0.967

Для каждой страны вычисляется сколько раз встречается страна во всех ячейках области ABCD. C(Finland)=2, C(Sweden)=2, для остальных C=1

S(c) – множество стран, находящихся в ячейке c.

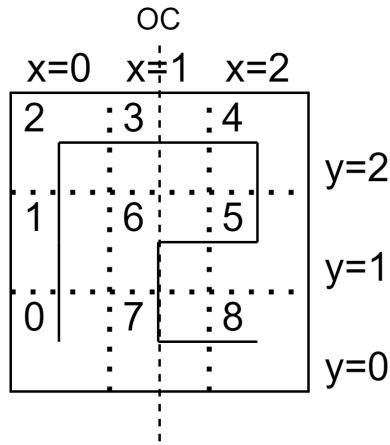
Псевдокод на примере перехода A-C:

Kac=0

```
for i in (S(a) and S(c)):  
    Kac+=C(i)
```

Переходы A-C, B-D имеют вес 2, остальные - 0

Критерий симметричности кривой



$$T[1] = (0; 1)$$

$$TS[1] = (2; 1)$$

$$T[7] = (1; 0)$$

$$D[1] = (1-2; 0-1)$$

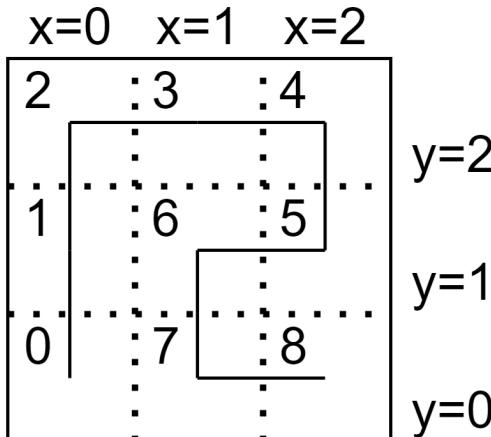
Шаблоны ZИPU в зависимости от поворота имеют определенную предполагаемую ось симметрии. Например, для шаблонов Z и И предпочтительно использовать симметрию относительно центральной точки, а для шаблонов П и U – осевую.

Критерий симметрии отражает то, насколько первая точка симметрична с последней, вторая – с предпоследней и тд.

Для каждой точки $T[i]$ на кривой вычисляется симметричная ей точка $TS[i]$, а также берется точка, которая симметрична ей относительно положения на кривой $T[j]$, где $j=L-1-i$, L – длина кривой. Затем вычисляется сумма длин векторов $D[i]$, соединяющих $TS[i]$ с $T[j]$.

Чем меньше полученное значение, тем более симметричную кривую мы получили.

Критерий минимизации суммарного номера соседних ячеек.



Для первой точки на кривой (с индексом 0):

$$C = (0;0)$$

$$N = [(0;1), (1;0), (1;1)]$$

$$K = |1-0| + |7-0| + |6-0| = 14$$

При выборе из большого числа примерно одинаковых вариантов можно применить критерий минимизации суммарного номера соседних ячеек.

Для каждой ячейки С на кривой L получить список соседних ячеек N, а затем рассчитать сумму модулей разностей положения ячеек N на кривой L и положения ячейки С на кривой L. Чем меньше полученное значение, тем лучше.

$$K=0$$

for C in L:

 for n in N:

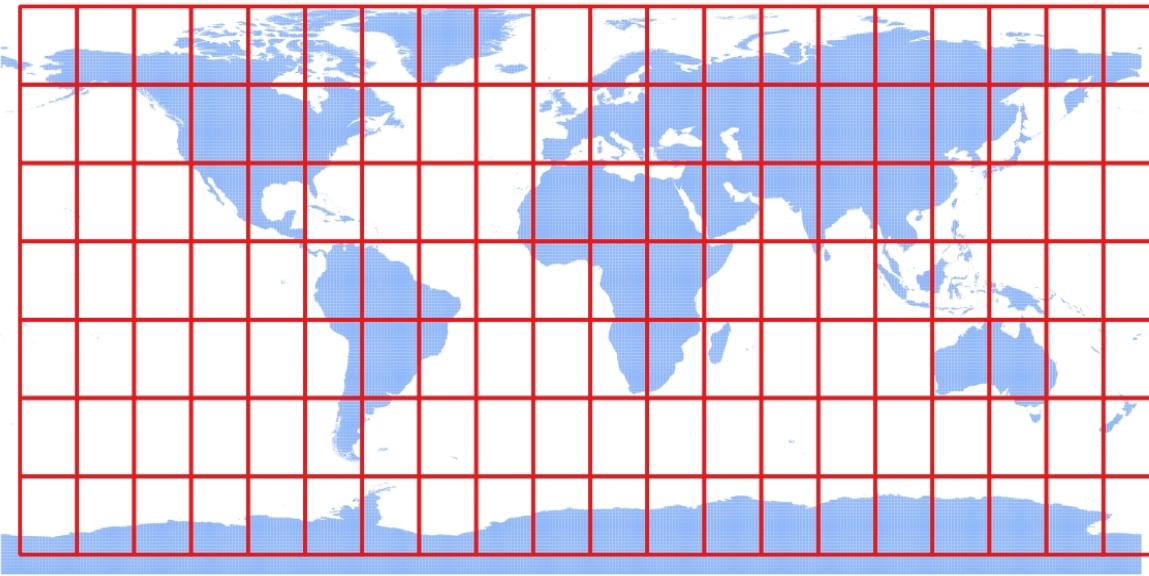
$$K += \text{abs}(\text{INDEX}(n) - \text{INDEX}(C))^{**P}$$

Степень P управляет чувствительностью критерия. В данном исследовании использовалась P=1

Алгоритм подготовки шаблонов

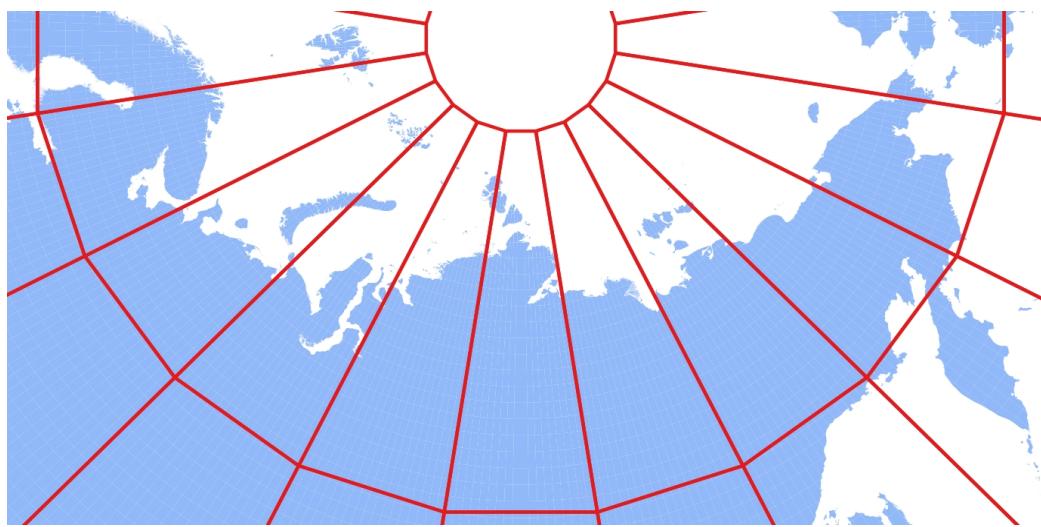
- Алгоритм подготовки шаблона:
- I. **Разделение области определения на блоки.**
- II. Создание множества элементов базовой линии на разных уровнях детализации
- III. Построение базовой линии
- IV. Подбор элементов базовой линии, которые удовлетворяют критериям.

Выбор размеров блоков

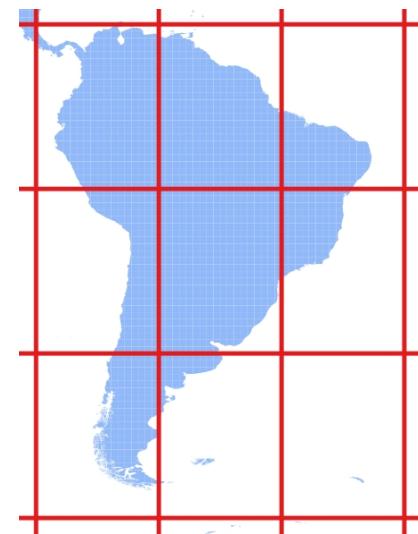


- Смещение блоков на 6 градусов (одну зону UTM) позволит оптимизировать обход некоторых участков суши. Каждый из 140 блоков имеет размеры 18 на 24 градуса и может быть разделен на регулярные ячейки 6 на 6 градусов.

Особенности блоков, смещенных на 6 градусов: Заполярье, Южная Америка

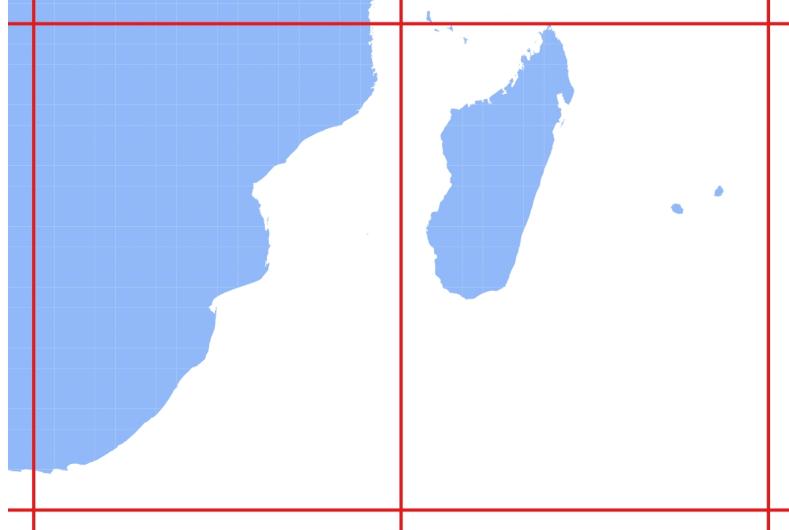


Восточная граница Таймырского полуострова присоединяется основному блоку. Полуостров Ямал полностью входит в один блок. Минимизируется область Чукотского АО, которая входит в блок содержащий Северную Америку.

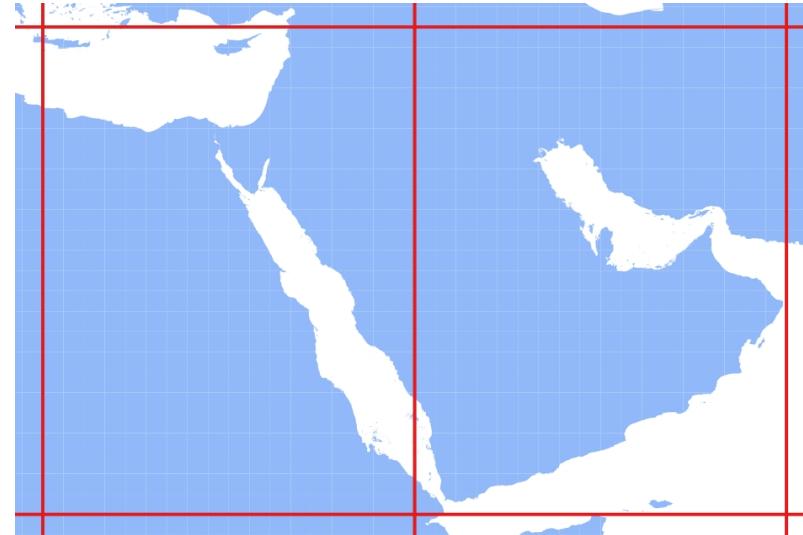


Южная Америка преимущественно покрывается восемью блоками

Особенности блоков, смещенных на 6 градусов: Южная Африка, Синайский полуостров

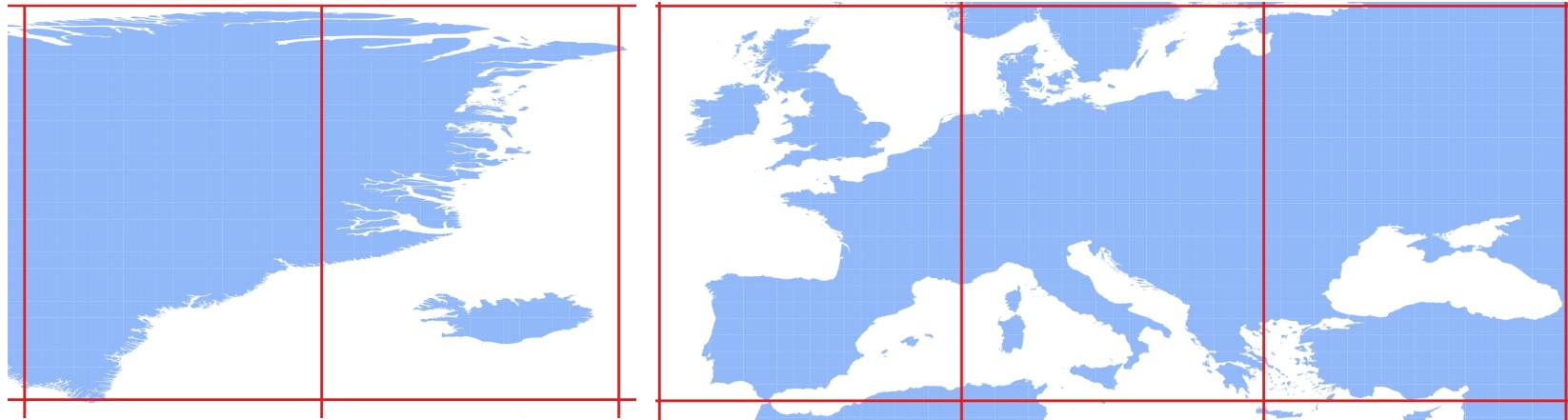


Южная Африка отделена от
Мадагаскара



Синайский полуостров
покрывается полностью двумя
блоками

Особенности блоков, смещенных на 6 градусов: Гренландия, Европа (Британия), Исландия



- 1) граница блока практически выравнена по восточной границе Гренландии;
- 2) уменьшаются водные области в блоке, покрывающем Западную Европу;
- 3) Исландия покрывается одним блоком;

Алгоритм подготовки шаблонов

- Алгоритм подготовки шаблона:
- I. Разделение области определения на блоки.
- II. Создание множества элементов базовой линии на разных уровнях детализации
- III. Построение базовой линии
- IV. Подбор элементов базовой линии, которые удовлетворяют критериям.

Способ декомпозиции задачи

2	3
1	4

Схема шаблона П (Ри)
2 на 2

2	2	2	3	3	3
2	2	2	3	3	3
2	2	2	3	3	3
1	1	1	4	4	4
1	1	1	4	4	4
1	1	1	4	4	4

Схема
шаблона П (Ри)
6 на 6 с делением
на одинаковые
блоки

Одна из схем
шаблона П (Ри)
6 на 6 с делением
по маске
 $m = [4,4,3,3,2,2]$

C=2	2	2	3	3	3	3
C=2	2	2	3	3	3	3
C=3	2	2	2	3	4	4
C=3	1	2	2	4	4	4
C=4	1	1	1	1	4	4
C=4	1	1	1	1	4	4

Количество вариантов масок для
 $S=6 - 141, S=8 - 38165$

Решать задачу многокритериального построения базовых линий длительностью 36 – сложно. Проводилась декомпозиция задачи на 4 подзадачи (A,B,C,D) меньшего размера. Ниже продемонстрирован способ соединения 4х базовых линий длиной 9 для заполнения блока 6 на 6 клеток.

Пусть размер блока $S=6$, тогда максимальная длина базовой линии $L=6^{**}2=36$,

{m} – множество перестановок с повторением элементов из множества $[2,...,S-2]$ где сумма элементов равна $L/2$

Условия подбора базовых линий A и B для маски m:

1-3) $M(A) \text{ OR } m = m, M(B) \text{ OR } m = m, M(A) \text{ AND } M(B) = 0;$

4) начало A входит во множество начальных точек, которое зависит от шаблона;

5) окончание B входит во множество конечных точек, которое зависит от маски M;

6) окончание A и начало B – соседние (и входят в множество точек, зависящее от маски M).

Создание набора символов для блоков.

Для каждой из 140 областей, были созданы наборы символов с размерами 6 на 6 ячеек для дальнейшей оптимизации:

Zu, Zr, Zd, Zl, Iu, Ir, Id, Il, Pu, Pr, Pd, Pl, Uu, Ur, Ud, Ul.

Созданы базы данных с группировкой по областям для глобального поиска.

Пример запроса к БД Sqlite :

```
p09 = 1-part
p11 = 1+part
cursor.execute('''select ?, ?,
                     s.`env`, s.smb, s.bl, s.start_xy,
                     s.ti, s.sym, s.su3, s.pop, s.neib,
                     ?, ?, ?
                from selector s
               where `env`=? and smb=?
                 and ti <= ?
                 and pop >= ?
                 and su3 >= ? ''',
               (part, name, min_ti, max_pop, max_su3,
                env, smb, min_ti*p11, max_pop*p09, max_su3*p09)
              )
```

где part – допуск к максимальному/минимальному значению критериев.

Алгоритм подготовки шаблонов

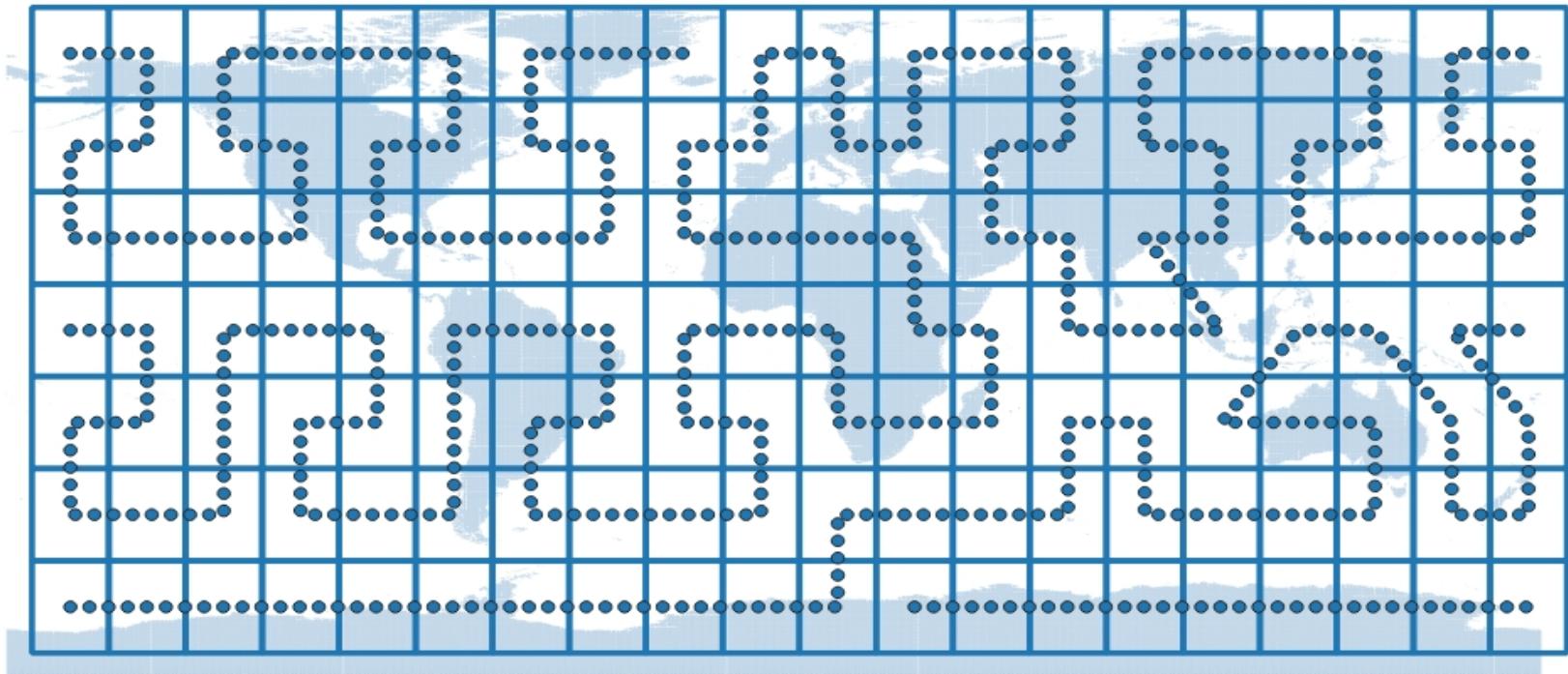
- Алгоритм подготовки шаблона:
- I. Разделение области определения на блоки.
- II. Создание множества элементов базовой линии на разных уровнях детализации
- **III. Построение глобальной базовой линии**
- IV. Подбор элементов базовой линии, которые удовлетворяют критериям.

Основные критерии построения глобальной базовой линии.

- 1) начало и конец базовой линии не должны совпадать
- 2) необходимо обеспечить высокую локальность данных по континентам, частям Света и странам
- 3) необходимо учитывать возможность перехода через антимеридиан
- 4) не использовать чрезмерное разбиение на блоки

Можно учитывать разные критерии при разбиении на блоки

Глобальная базовая линия для блоков ячеек размером 3 на 4 градуса

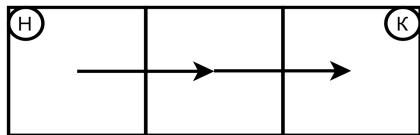


Алгоритм подготовки шаблонов

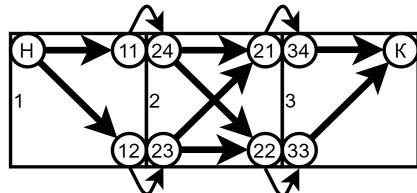
- Алгоритм подготовки шаблона:
- I. Разделение области определения на блоки.
- II. Создание множества элементов базовой линии на разных уровнях детализации
- III. Построение глобальной базовой линии
- IV. Подбор элементов базовой линии, которые удовлетворяют критериям.

Способ создания графов для оптимального заполнения области определения индекса

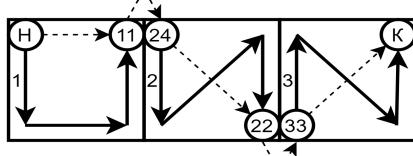
После создания наборов символов предлагается следующий способ размещения символов на базовых линиях: строится ориентированный граф с кратными дугами и далее выполняется поиск кратчайшего пути между начальной и конечной точками графа.



Базовая линия с отмеченными начальной и конечной точками



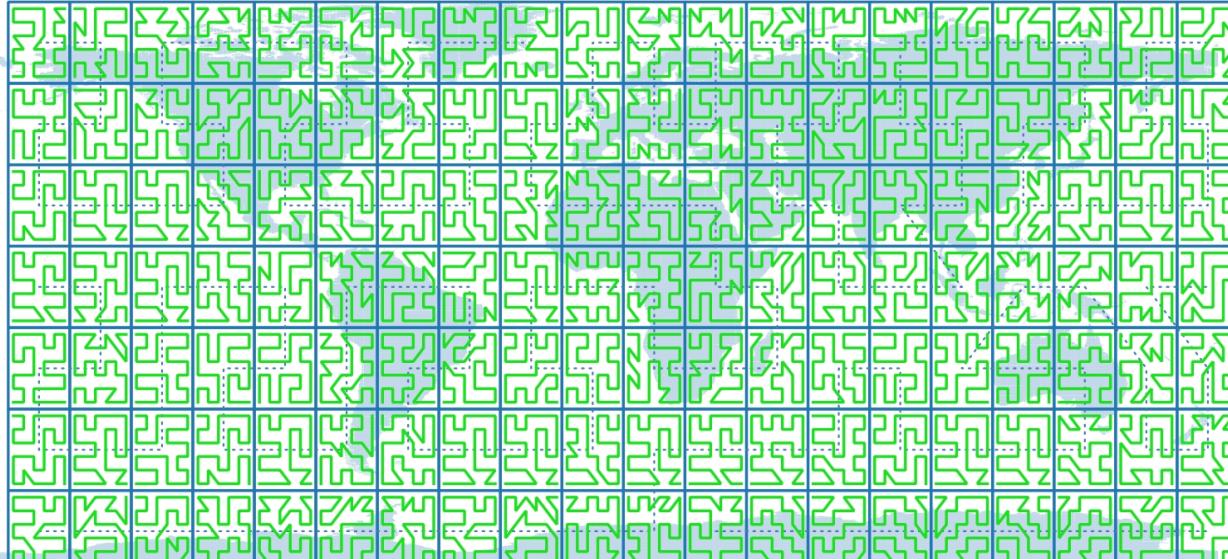
Полный граф для поиска оптимального пути.
Жирными стрелками указаны множества кратных вершин



Граф одного из вариантов пути: Uu, Iu, Zl.

Обход ячеек сетки 3 на 4 градуса.

На слайде показана текущая версия (34-2025-07-28) оптимального обхода ячеек сетки 3 на 4 градуса.

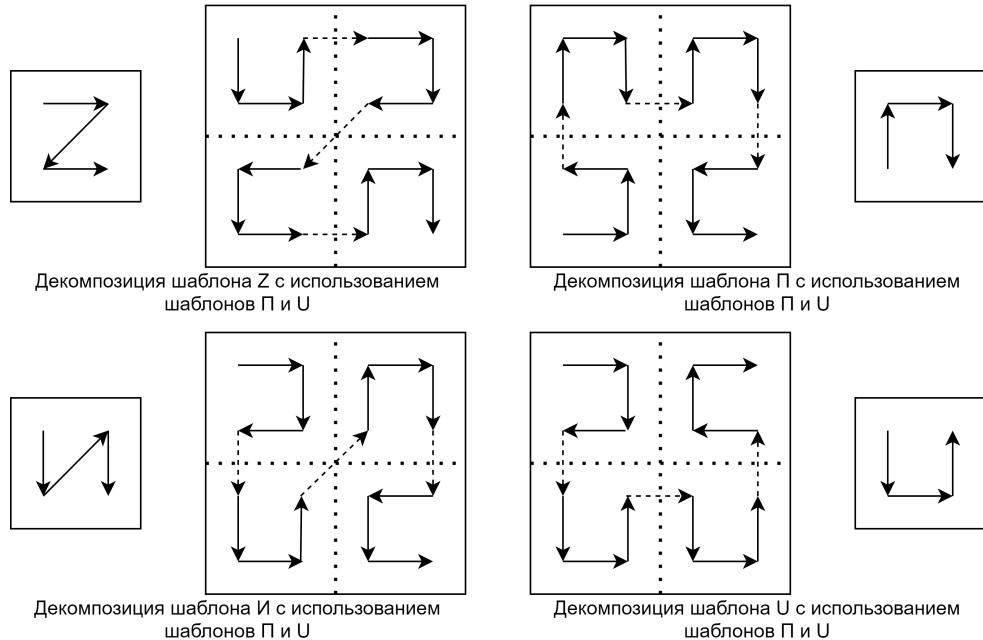


В качестве веса ребер использовался инвертированный критерий политической карты мира, а ребра при построении графа отбирались по тройке КПС, КН, КПК.

Дальнейшая оптимизация обхода выполняется аналогично, применяя алгоритм Дейкстры для ориентированных графов с кратными ребрами.

Переход от шаблонов Z и И к П и U. Использование преобразования Гильберта.

При формировании шаблонов для ZИPU (ZIPU)
- преобразования оптимизация обхода выполняется лишь на низких уровнях заметающей кривой. Для построения кривых более высокого порядка используются стандартные способы декомпозиции шаблонов.



Заключение

- 1.1. Предложен способ построения базовых кривых и ZIPU-кривых, удовлетворяющих некоторым критериям.
- 1.2. Продемонстрировано использование способа для градусной сетки 3 на 4 градуса.
2. Предложены способы вычисления критериев.

Спасибо за внимание

Использованные материалы:

- 1) ОБОБЩЕННЫЙ АЛГОРИТМ ПОСТРОЕНИЯ ЗАМЕТАЮЩИХ КРИВЫХ. Башков А.А. Естественные и технические науки. 2024. № 10 (197). С. 151-162.
- 2) Башков А.А. Использование заметающих кривых для построения геопространственного индекса // Материалы 22-й Международной конференции «Современные проблемы дистанционного зондирования Земли из космоса». Москва: ИКИ РАН, 2024. С. 24. DOI 10.21046/22DZZconf-2024a
- 3) Данные о плотности населения
https://neo.gsfc.nasa.gov/view.php?datasetId=SEDAC_POP
- 4) Данные о поверхности Земли
<https://osmdata.openstreetmap.de/data/land-polygons.html>
- 5) Данные о странах*:
https://naciscdn.org/naturalearth/packages/natural_earth_vector.sqlite.zip