A faint, light gray world map is visible in the background of the slide, centered behind the text.

**Шести семинар „Математика и примени“, 17 Март 2023,
Институт за математика, Природно-математички факултет,
Универзитет „Св. Кирил и Методиј“, Скопје**

**МЕТОД НА ИНТЕРПОЛАЦИЈА СО ИНВЕРЗНО
ТЕЖИНСКО РАСТОЈАНИЕ И НЕГОВА ПРИМЕНА
ВО ГЕОГРАФСКИ ИНФОРМАЦИСКИ СИСТЕМИ
(ГИС)**

Ива Лазова

**Институт за математика, Природно-математички факултет,
Универзитет „Св. Кирил и Методиј“, Скопје**

План на предавањето

2

- Вовед
- Просторна анализа
- Математичката позадина и постојните пристапи
- Приод кој користи тежински просеци
 - Избор на експонент
 - Недостатоци на чисто инверзно тежинско растојание
- Подобрување на функцијата за тежинско растојание
 - Избирање блиски точки
 - Вклучување насока
 - Определување наклон
 - Намалување на пресметковната грешка
 - Својства на $f(T)$
- Примена за вредноста на m_{10} честичките
- Заклучок

1. Вовед

ГИС Е “ИНСТИТУЦИОНАЛЕН ОБЈЕКТ КОЈ ОДРАЗУВА ОРГАНИЗАЦИСКА СТРУКТУРА ШТО ИНТЕГРИРА ТЕХНОЛОГИЈА СО **БАЗА НА ПОДАТОЦИ**, СТРУЧНОСТ И ПОСТОЈАНА ФИНАНСИСКА ПОДДРШКА”, ОСВЕН ТОА Е И “МОЌЕН ИНСТРУМЕНТАРИУМ ЗА СОБИРАЊЕ, СКЛАДИРАЊЕ, ПРЕЗЕМАЊЕ, ПРЕОБРАЗУВАЊЕ И ПРИКАЖУВАЊЕ ПРОСТОРНИ ПОДАТОЦИ ОД РЕАЛНИОТ СВЕТ ЗА ОДРЕДЕНИ НАМЕНИ”.

- ГИС Е “ОРГАНИЗИРАНА АКТИВНОСТ СО КОЈА
- СЕ МЕРАТ АСПЕКТИТЕ НА ГЕОГРАФСКИТЕ ПОЈАВИ И ПРОЦЕСИ;
 - ОВИЕ МЕРИ СЕ ПРЕТСТАВУВААТ ОБИЧНО ВО ФОРМА НА КОМПЈУТЕРСКА БАЗА НА ПОДАТОЦИ ЗА ДА СЕ НАГЛАСАТ ПРОСТОРНИТЕ СЛОЕВИ, ЕНТИТЕТИ И ОДНОСИ;
 - СЕ ВРШАТ ОПЕРАЦИИ ВРЗ ОВИЕ ПРЕТСТАВУВАЊА СО ЦЕЛ ДА СЕ ИЗВЕДАТ ДРУГИ МЕРЕЊА И ДА СЕ ОТКРИЈАТ НОВИ ОДНОСИ ПРЕКУ ИНТЕГРИРАЊЕ РАЗЛИЧНИ ИЗВОРИ; И
 - ОВИЕ ПРЕТСТАВУВАЊА СЕ ПРЕОБРАЗУВААТ ЗА ДА СЕ УСОГЛАСАТ СО ДРУГИ РАМКИ НА ЕНТИТЕТИ И НА ОДНОСИ”.

2. Просторна анализа

- Просторната анализа е процес на манипулација со просторни информации за да се извлечат нови информации и значење од оригиналните податоци. Обично просторната анализа се врши со Географски Информациски Системи (ГИС).
- Често во географијата, каде се користат емпириски теренски податоци, се појавува потреба за интерполација на неправилно распоредени податоци со цел конструирање на непрекинатата површина.
- Со цел да се прикажат овие податоци во некој вид на контурна карта или перспективен приказ, за да се споредат со податоци од истиот регион врз основа на други податочни точки, или да се анализираат за екстремни, градиенти или други цели, се дефинира непрекинатата интерполациска функција, која точно се совпаѓа во дадените вредности.
- Интерполацијата во софтверскиот пакет QGIS за ГИС поддржува методи за:
 - интерполација со **триаголна неправилна мрежа** (TIN, Triangular Irregular Network), и
 - интерполација со **инверзно тежинско растојание** (IDW, Inverse Distance Weighting)
- Методот TIN, најчесто се користи за податоци за надморска височина. Додека методот IDW, се користи за интерполирање на други видови податоци, како што се: концентрации на минерали, популации, температурни карти, карти за просечна количина врнежи, итн.

3. Математичка позадина и постојните пристапи

- Во суштина, се посакува оперативно решение на проблемот на дводимензионална интерполација од неправилно распоредени податочни точки. Се претпоставува дека е даден конечен број на тројки (x_i, y_i, z_i) , N , каде што x_i, y_i , се локациските координати на податочната точка D_i , а z_i е соодветната вредност на податоците.
- Се бара интерполационата функција $z = f(x, y)$ која доделува вредност на која било локација $T(x, y)$ во рамнината. Оваа дводимензионална функција на интерполација треба да биде „глатка“ (непрекината и барем еднаш диференцијабилна), да поминува низ наведените точки (т.е., $f(x_i, y_i) = z_i$), и да се исполнети интуитивните очекувања на корисникот за феноменот што се истражува. Понатаму, функцијата треба да биде соодветна за компјутерска апликација по разумна цена.
- Постојните методи на дво-димензионална точна интерполација се од два типа: една единствена, глобална функција, која често е премногу сложена; или едно соодветно дефинирано множество од едноставни, локални, функции кои се соодветно усогласени на нивните граници. Функцијата што се презентира во овој труд е од вториот тип, така конструирана да потдоменот на секоја локална функција е автоматски дефиниран, и функцијата е непрекинато диференцијабилна дури и на поврзувањата на локалните функции.

4. Приод кој користи тежински просеци

- Била тестирана една почетна функција од инверзно растојание; за неа вредноста во било која точка T во рамнината била тежински просек од вредностите во податочните точки $D_i, i = 1, 2, \dots, N$.
- Нека z_i е вредноста во податочната точка D_i , и $d[T, D_i]$ е Евклидското растојание меѓу T и D_i . Онаму каде што референтната точка T се подразбира, ознаката $d[T, D_i]$ ќе ја скратиме со d_i . Интерполираната вредност во T која ја користи првата интерполирачка функција е:

$$f_1(T) = \begin{cases} [\sum_{i=1}^N (d_i)^{-p} z_i] / \sum_{i=1}^N (d_i)^{-p} & \text{ако } d_i \neq 0 \text{ за сите } D_i \text{ (} p > 0 \text{)} \\ z_i & \text{ако } d_i = 0 \text{ за некои } D_i. \end{cases} \quad (1)$$

- Забележете дека како што T се приближува кон податочната точка $D_i, d_i \rightarrow 0$, и i –тите изрази во броителот и именителот ги надминуваат сите граници додека останатите изрази остануваат ограничени. Според тоа $\lim_{T \rightarrow D_i} f_1(T) = z_i$ како што е потребно, и според тоа функцијата $f_1(T)$ е непрекината.

4.1. Избор на експонент

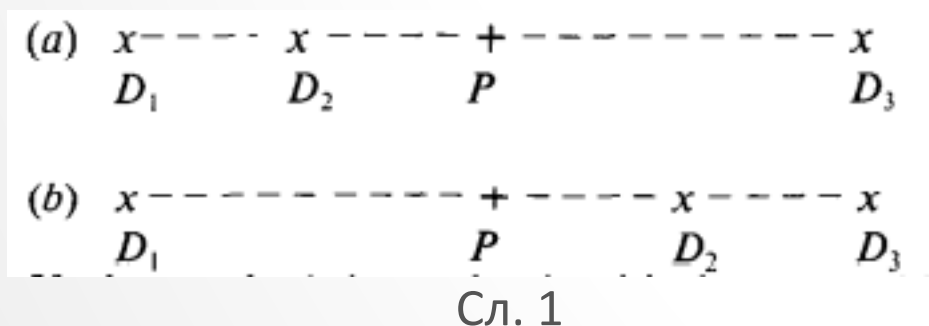
- Користејќи ги координатите $T(x, y)$ и $D_i(x_i, y_i)$, парцијалното диференцирање води до:

$$\frac{\partial}{\partial x} f_1(x, y) = \frac{\sum_{i=1}^N \sum_{\substack{j=1 \\ i \neq j}}^N (d_i)^{-p-2} (d_j)^{-p} (x-x_i) z_i (z_i - z_j)}{[\sum_{i=1}^N (d_i)^{-p}]^2} \quad (2)$$

- Заменувањето на $(x - x_i)$ со $(y - y_i)$ во горното равенство го дава соодветниот израз за $\frac{\partial}{\partial y} f_1(x, y)$. За T во соседството на D_i , $\frac{\partial}{\partial x} f_1(x, y)$ се однесува како $(x - x_i)(d_i)^{p-2}$, или како $(d_i)^{p-1}$.
- За $p = 1$, обата леви и десни парцијални изводи постојат; тие се генерално ненулни и спротивни по знак. За $p < 1$, не постојат изводи.
- Затоа имаме услов $p > 1$. Емпириските тестови покажале дека повисоките експоненти ($p > 2$) имаат тенденција да ја прават површината релативно рамна (константна) близу до сите податочни точки, со многу стрмни градиенти над малите интервали меѓу податочните точки. Експонентот од $p = 2$ не само што дава навидум задоволителни емпириски резултати за целите на општо мапирање на површина и опис, но исто ја претставува најлесната пресметка.
- Во Евклидови координати, $(d_i)^{-2} = 1/[(x - x_i)^2 + (y - y_i)^2]$ (3)

4.2. Недостатоци на чисто инверзно тежинско растојание 8

- Кога бројот на податочни точки е голем, пресметувањето на $z = f_1(T)$ станува пропорционално подолго. На крај методот ќе стане неефикасен или непрактичен.
- Само растојанијата до T од податочните точки D_i , а не и насоката се земаат предвид. Заради тоа следниве две конфигурации на колинеарни точки на пример би давале идентични интерполирани вредности кај T :



- Нултите парцијални изводи добиени во секоја податочна точка D_i укажуваат дека добивме интерполирана функција од две променливи со локални екстреми во секоја од точките D_i , што е неповолен тип на функција.
- Пресметувачката грешка станува значајна во соседството на точките D_i , бидејќи предоминантните изрази резултираат од разликата од два речиси еднакви броеви.

5. Подобрување на функцијата за тежинско растојание ⁹

5.1. Избирање блиски точки

- N - вкупен број податочни точки
- A - област на најголемиот многуаголник ограден со податочните точки

$$\pi r^2 = 7 \left(\frac{A}{N} \right). \quad (4)$$

- Сега нека $C_T = \{D_i | d_i \leq r\}$ и $n(C_T) =$ бројот на елементи во C_T и нека $0 \leq d_{i_1} \leq d_{i_2} \leq \dots \leq d_{i_N}$.
- Сега дефинираме

$$C_T^n = \{D_{i_1}, D_{i_2}, \dots, D_{i_n}\} \quad (n \leq N) \quad \text{и} \quad r'(C_T^n) = \min \{d_{i_j} | D_{i_j} \notin C_T^n\} = d_{i_{n+1}} \quad (5)$$

$$\text{• Конечно, нека } C'_T = \begin{cases} C_T^4 & \text{ако } 0 \leq n(C_T) \leq 4 \\ C_T & \text{ако } 4 < n(C_T) \leq 10 \\ C_T^{10} & \text{ако } 10 < n(C_T) \end{cases} \quad (6) \quad \text{и} \quad r'_T = \begin{cases} r'(C_T^4) & \text{ако } n(C_T) \leq 4 \\ r & \text{ако } 4 < n(C_T) \leq 10 \\ r'(C_T^{10}) & \text{ако } 10 < n(C_T) \end{cases} \quad (7)$$

5. Подобрување на функцијата за тежинско растојание ¹⁰

5.1. Избирање блиски точки

- Уште, подобрениот тежински израз $s_i = s(d_i)$ го дефинираме:

$$s(d) = \begin{cases} \frac{1}{d} & \text{ако } 0 < d \leq \frac{r'}{3} \\ \frac{27}{4r'} \left(\frac{d}{r'} - 1\right)^2 & \text{ако } \frac{r'}{3} < d \leq r' \\ 0 & \text{ако } r' < d \end{cases} \quad (8)$$

- Конечно првото подобрување е:

$$f_2(T) = \begin{cases} \left[\frac{\sum_{D_i \in C'} (s_i)^2 z_i}{\sum_{D_i \in C'} (s_i)^2} \right] & \text{ако } d_i \neq 0 \text{ за сите } D_i \\ z_i & \text{ако } d_i = 0 \text{ за некои } D_i \end{cases} \quad (9)$$

5.2. Вклучување насока

- Насочениот тежински израз за секоја податочна точка D_i во близина на T е дефиниран со

$$t_i = \left[\sum_{D_j \in C'} s_j [1 - \cos(\angle D_i T D_j)] \right] / \left[\sum_{D_j \in C'} s_j \right] \quad (10)$$

- Косинусот на аголот $\angle D_i T D_j$ може да биде пресметан преку скаларен производ како

$$\left[(x - x_i)(x - x_j) + (y - y_i)(y - y_j) \right] / d_i d_j. \quad (11)$$

- Може да се дефинира насока на броење, нова тежинска функција $w_i = (s_i)^2 \cdot (1 + t_i)$. Второто подобрување на функцијата за интерполација сега е:

$$f_3(T) = \begin{cases} \left[\sum_{D_i \in C'} w_i z_i \right] / \left[\sum_{D_i \in C'} w_i \right] & \text{ако } d_i \neq 0 \text{ за сите } D_i \\ z_i & \text{ако } d_i = 0 \text{ за некои } D_i \end{cases} \quad (12)$$

5.3. Определување наклон

- A_i и B_i се константи кои ги определуваме за секоја точка D_i и кои се посакуваните наклони во x и y насоките во D_i . Нека $C_i'' = C_{D_i}' \setminus \{D_i\}$. Тогаш

$$A_i = \frac{\sum_{D_j \in C_i''} w_j \frac{(z_j - z_i)(x_j - x_i)}{(d[D_j, D_i])^2}}{\sum_{D_j \in C_i''} w_j} \quad \text{и} \quad B_i = \frac{\sum_{D_j \in C_i''} w_j \frac{(z_j - z_i)(y_j - y_i)}{(d[D_j, D_i])^2}}{\sum_{D_j \in C_i''} w_j} \quad (13)$$

- Следно, v е параметар кој го ограничува максималниот ефект на наклонот.

$$v = \frac{0,1[\max\{z_i\} - \min\{z_i\}]}{[\max\{(A_i^2 + B_i^2)\}]^{\frac{1}{2}}} \quad (14)$$

- Дефинираме зголемување Δz_i за секој $D_i \in C_T'$ како функција од T .

$$\Delta z_i = [A_i(x - x_i) + B_i(y - y_i)] \left[\frac{v}{v + d_i} \right] \quad (15)$$

5.3. Определување наклон

- Пресметувањето на парцијалните изводи во D_i дава

$$\frac{\partial}{\partial x} (\Delta z_i) \Big|_{\substack{x=x_i \\ y=y_i}} = A_i \quad \text{и} \quad \frac{\partial}{\partial y} (\Delta z_i) \Big|_{\substack{x=x_i \\ y=y_i}} = B_i \quad (17)$$

- Така, интерполационата функција дефинирана подолу ќе ги даде посакуваните парцијални изводи A_i и B_i во D_i и се однесува инаку како f_3 (12).
- Конечно, третото подобрување на тежинската функција е:

$$f_4(T) = \begin{cases} [\sum_{D_i \in C'} w_i (z_i + \Delta z_i)] / [\sum_{D_i \in C'} w_i] & \text{ако } d_i \neq 0 \text{ за сите } D_i \in C' \\ z_i & \text{ако } d_i = 0 \text{ за некои } D_i \in C' \end{cases} \quad (18)$$

5.4. Намалување на пресметковната грешка

- Пресметковниот проблем е избегнат со утврдување на некое ϵ , во зависност од компјутерската прецизност, и дефинирање $f(T) = z_i$ во рамки на ϵ - околината на D_i .
- Ако неколку податочни точки се појават во рамки на ϵ - околината $N_\epsilon(T)$ на T , тогаш нивните вредности се земаат во просек.
- Така, конечната интерполациона функција е според тоа дадена со:

$$f(T) = \begin{cases} [\sum_{D_i \in C'} w_i (z_i + \Delta z_i)] / [\sum_{D_i \in C'} w_i] & \text{ако } d_i > \epsilon \text{ за сите } D_i \\ [\sum_{D_i \in N_\epsilon(T)} z_i] / [\sum_{D_i \in N_\epsilon(T)} 1] & \text{ако } d_i < \epsilon \text{ за некои } D_i \end{cases} \quad (19)$$

5.5. Својства на $f(T)$

- Интерполационата функција $f(T)$ (19) дефинирана погоре ги исполнува експлицитните барања кои се почетно поставени. Површината дефинирана со $f(T)$ е непрекинато диференцијабилна секаде освен на бариерите каде е потенцирана непрекинатоста. Површината ги претпоставува бараните вредности во сите податочни точки.
- Практично, пресметувањето на интерполираните вредности директно е разумно. Парцијалните изводи кај податочните точки се базирани врз забелешката какви тие “би требало” да бидат, т.е. просек од првите (Њутнови) поделени разлики од блиските познати вредности (13).
- Конечната функција на интерполација може да има екстремни насекаде не само во податочните точки. Обсегот на интерполационата функција е до 10% поголем на секој крај од опсегот на почетните вредности z_i .

6. Примена за вредноста на pm10 честичките

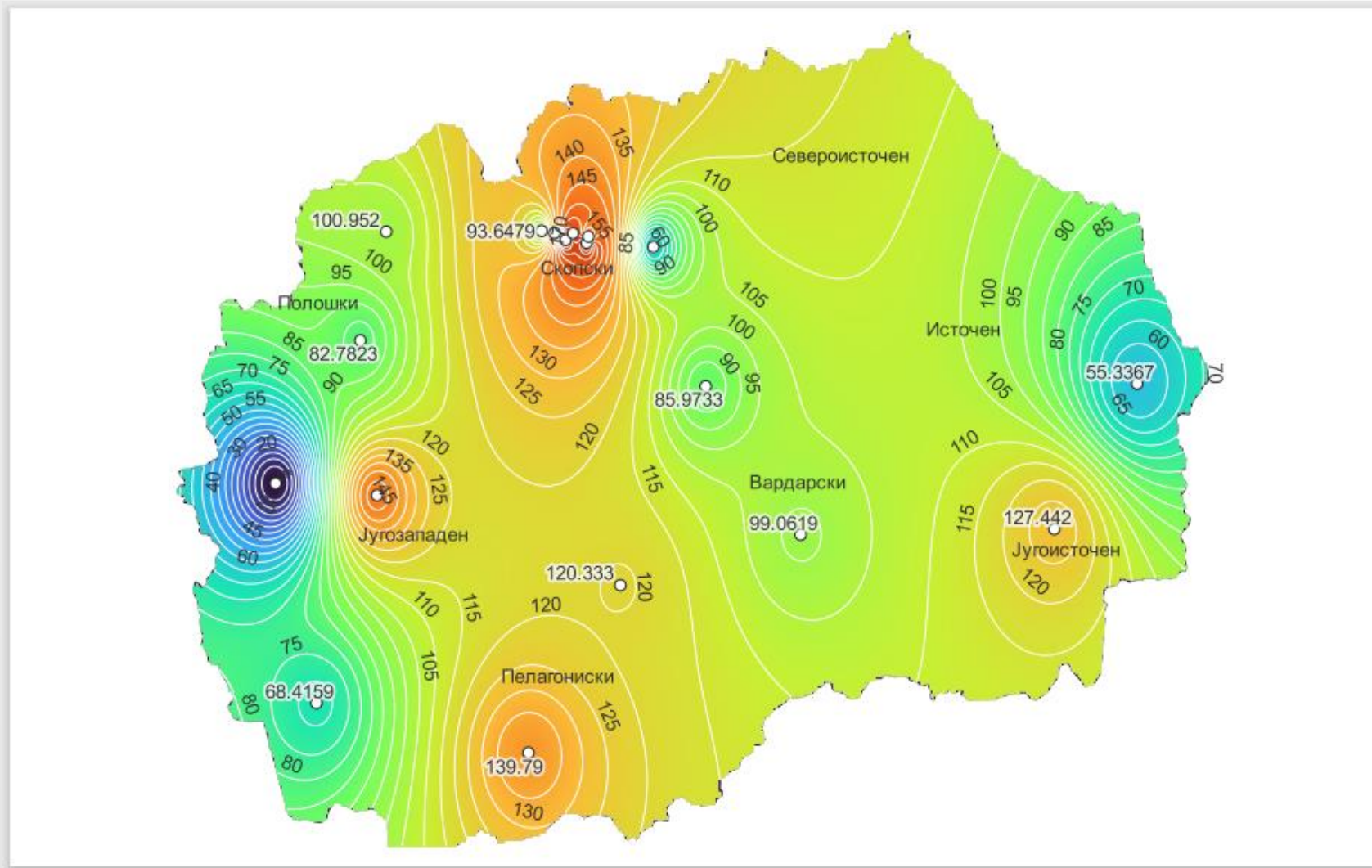
- Во овој труд разгледуваме проблем каде се дадени податоци за загадувањето во Република С. Македонија и целта ни е да изработиме карта на загадувањето со која ќе можеме да го предвидиме загадувањето во произволна точка на територијата. Според горекажаното, ќе користиме метод на интерполација со инверзно тежинско растојание (IDW).
- Применуваме интерполација со инверзно тежинско растојание (IDW), за изработка на карта на загадувањето во С. Македонија, врз основа на мерни податоци од 18 мерни станици низ С. Македонија. Податоците за вредноста на pm10 честичките се однесуваат на ден 06.01.2023 (дадени во Табела 1). Како резултат на извршената интерполација добиени се контури кои се претставени на картата.
- Во софтверскиот пакет **QGIS**, интерполацијата се изработува со помош на алатката за интерполација од менито *Processing Toolbox*. Потребните операции се:
 - ✓ Креирање на контури од податоци за точки.
 - ✓ Маскирање на вредности без податоци од растерски слој (raster layer).
 - ✓ Додавање натписи (labels) на векторски слој (vector layer).

6. Примена за вредноста на pm10 честичките

Мерни станици	Вредност на pm10	Latitude	Longitude
Центар-СК	189.596	41.9897330365848	21.424852542002352
Карпош-СК	125.996	42.002757509429365	21.3968197926856
Лисиче-СК	194.427	41.98385932547682	21.477586470230847
Гази Баба-СК	126.106	41.99593248357466	21.48213549533233
Ректорат-СК	158.938	42.002071850721336	21.44368334775833
Миладиновци	58.0625	41.97563541593915	21.645418313986767
Ѓ.Петров-СК	93.6479	42.00676678023977	21.364596284367924
Битола	139.79	41.02799927868185	21.326353203144006
Кичево	147.804	41.511251586087475	20.950758122201616
Лазарополе	4.8087	41.533357452307044	20.697385926095645
Тетово	100.952	42.006040470838016	20.97341742203057
Гостивар	82.7823	41.80195673106927	20.908662468857948
Прилеп	120.333	41.3412273343385	21.55690935103475
Велес	85.9733	41.71339795092725	21.773889323967477
Кавадарци	99.0619	41.43281437381098	22.007219883748327
Струмица	127.442	41.436289180106996	22.639963720565653
Берово	55.3367	41.705660435316304	22.85417925841553
Охрид	68.4159	41.12130686981299	20.8007923251227

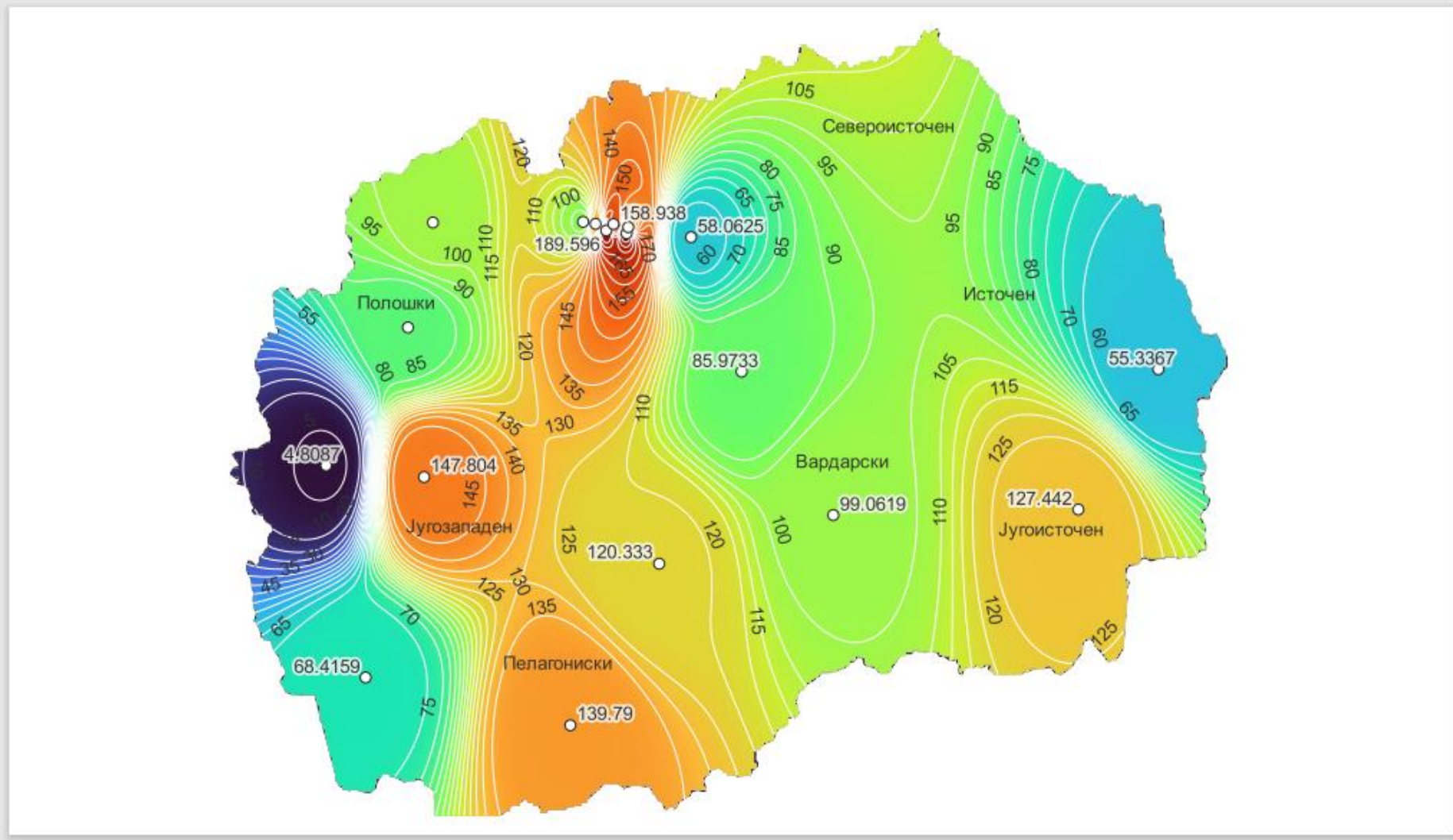
Табела 1. Суспендирани честички pm10 (microgram/m³) на ден 06.01.2023

6. Примена за вредноста на pm10 честичките



Сл 2. Карта на загадувањето во С. Македонија, врз основа на мерни податоци од одреден број мерни станици (мерните податоците за вредноста на pm10 честичките се однесуваат на ден 06.01.2023 од 18 мерни станици), користејќи интерполација со IDW со Коефициентот за растојание $P = 2$.

6. Примена за вредноста на pm10 честичките



Сл 3. Карта на загадувањето во С. Македонија, врз основа на мерни податоци од одреден број мерни станици (мерните податоците за вредноста на pm10 честичките се однесуваат на ден 06.01.2023 од 18 мерни станици), користејќи интерполација со IDW со Коефициентот за растојание $P = 5$.

6. Примена за вредноста на pm_{10} честичките

- Кога се користи IDW методот, треба да се избере вредност за Коефициентот за растојание P . Како што може да се види, поголем коефициент (Сл. 3) значи дека е потребно поголемо растојание за *вредностите на површината* да станат различни од блиските точки. Помал коефициент (Сл. 2) значи дека *вредностите на површината* брзо ќе се менуваат како што се зголемува растојанието. Ова може да предизвика една нагла промена во вредностите и е склоно кон „ефектот на биковите очи (bull's-eye effect)“ што создава кружни области на површината. Најдобро е да се креираат неколку различни површини и да се прилагоди оваа бројка, за да одговара на бараната анализа.
- Значи: „*Колку е поголем коефициентот P , толку помал ефект ќе имаат мерните точки ако тие се далеку од непознатата точка за време на процесот на интерполација*“. Ова може да се протолкува како: „*Колку е повисок бројот на коефициентот P , толку е побрзо опаѓањето на ефектот на мерните точки со растојанието*“. Кога е $P = 2$ значи побавно опаѓање со растојанието, а кога е $P = 5$ значи побрзо опаѓање со растојанието. Во било кој од случаите, целокупната теорија е: влијанието на блиските мерни (т.е. податочни) точки станува послабо колку тие се подалеку.

7. Заклучок

- Интерполационата функција е применета во ГИС, кои се користат за просторна анализа. Како предности на ГИС можеме да се наведат: 1) Заштеда на трошоци поради поголема ефикасност, 2) Подобро донесување одлуки, 3) Подобрена комуникација, 4) Географско менаџирање.
- Во ГИС, достапноста на податоците е главен проблем; ако податоците не се достапни, тогаш ГИС не се од корист. Недостаток при користењето на ГИС е тоа што може да ги прикажат резултатите како посигурни отколку што тие се во суштина, а грешките и претпоставките може да бидат сокриени, што доведува до недостиг од преиспитување на резултатите.
- Можеби некое подобрување би се добило ако ГИС во иднина се доразвие до способност за автоматско пресметување на грешката, па не би добиле фиксна карта на очекувани вредности во секоја точка, туку опсег во кои граници очекуваните вредности можат да се движат во произволна точка од областа на истражување.

7. Заклучок

- За крај за заинтересираните лица, линк до софтверскиот пакет QGIS:

<https://www.qgis.org/en/site/forusers/download.html>

- И уште еден линк за QGIS Tutorials со детални објаснувања за разните можности на QGIS:

<https://www.qgistutorials.com/en/>

ВИ БЛАГОДАРАМ НА ВНИМАНИЕТО

ИВА ЛАЗОВА | IVALAZOVA123@GMAIL.COM