

Anàlisi de dades geoespacial· PAC 3

Exercicis

Exercici 1: Anàlisi i Visualització de Patrons de Punts. Aquest exercici es tracta de generar i visualitzar patrons de punts, en particular, *agregats*, *aleatoris* i *regulars* usant funcions de R.

Agregat de punts

És un patró on tenim zones amb una densitat de punts més elevada que en altres, és a dir, podem dir que tenim *clústers* de punts.

Generem un patró d'agregats amb la funció `rnorm()`. Considerem que tenim uns 51 punts agrupats en 3 clústers (de 17 punts), centrats (argument *mean*) per exemple en els punts (0, 0), (0, 5) i (5, 5) amb una agrupació o *variància* de 0.6.

```
num_mostres_grup <- 17
sigma <- 0.6 # Variància de la distribució
mean1 <- 0 # centre
mean2 <- 5 # centre

set.seed(12345) # Per la reproductibilitat

# DataFrame del primer clúster(17 punts), centrat a (0, 0)
grup1 <- data.frame(
  x = rnorm(num_mostres_grup, mean = mean1, sd=sigma),
  y = rnorm(num_mostres_grup, mean = mean1, sd=sigma))

# Segon clúster, centrat a (5, 5)
grup2 <- data.frame(
  x = rnorm(num_mostres_grup, mean = mean2, sd=sigma),
  y = rnorm(num_mostres_grup, mean = mean2, sd=sigma))

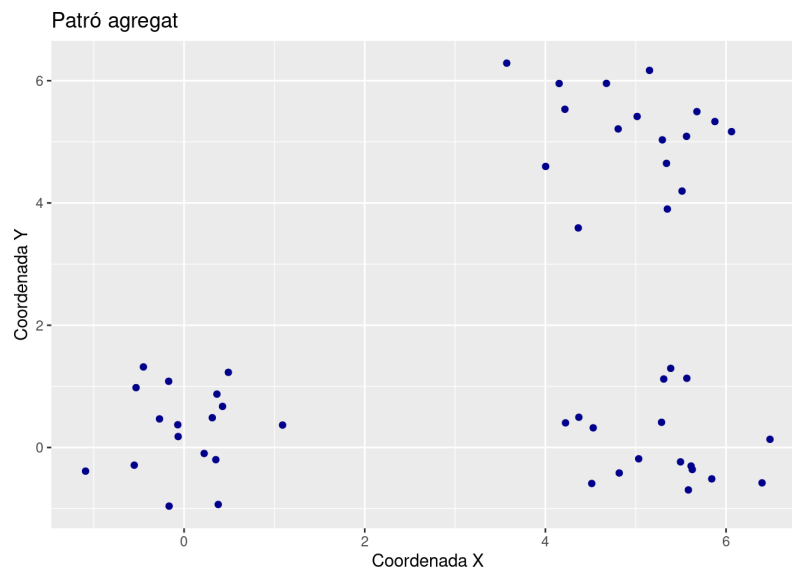
# Tercer clúster, centrat a (5, 0)
grup3 <- data.frame(
  x = rnorm(num_mostres_grup, mean = mean2, sd=sigma),
  y = rnorm(num_mostres_grup, mean = mean1, sd=sigma))

# Ajuntem tots els punts en un sol dataframe
agregats <- rbind(grup1, grup2, grup3)
```

Podem visualitzar-ho amb `ggplot()` i `geom_point()` de la llibreria `ggplot2`:

```
library(ggplot2)

ggplot(agregats, aes(x = x, y = y)) +
  geom_point(color = "darkblue") +
  labs(
    title = "Patró agregat",
    y = "Coordenada Y",
    x = "Coordenada X"
  )
)
```



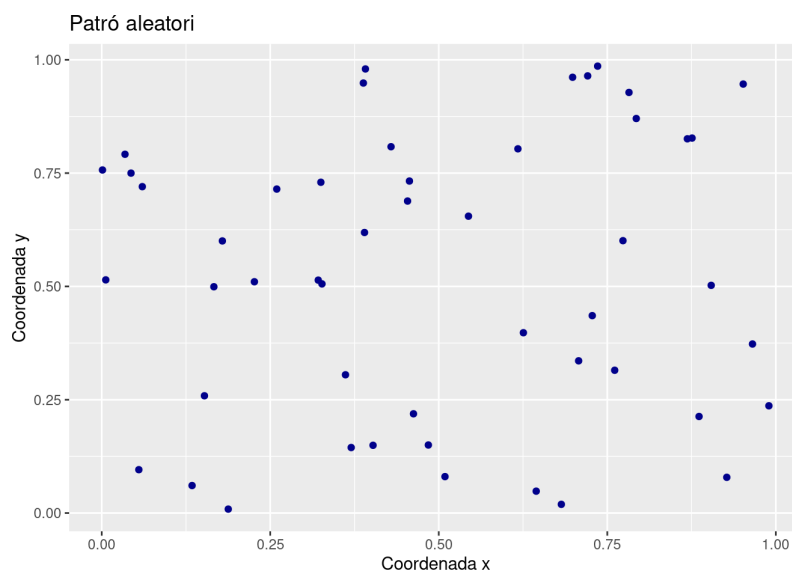
Patró aleatori

En aquest cas tenim una distribució aleatòria i uniforme en tot l'espai. Fem servir la funció `runif()` per generar uns 50 punts distribuïts en una regió entre (0, 0) i (1, 1).

```
num_mostres <- 50 # número de mostres
set.seed(12345) #reproductibilitat
aleatoris <- data.frame(
  x = runif(num_mostres),
  y = runif(num_mostres))
```

Tal com hem fet abans, ho podem visualitzar amb `ggplot()` i `geom_point()`.

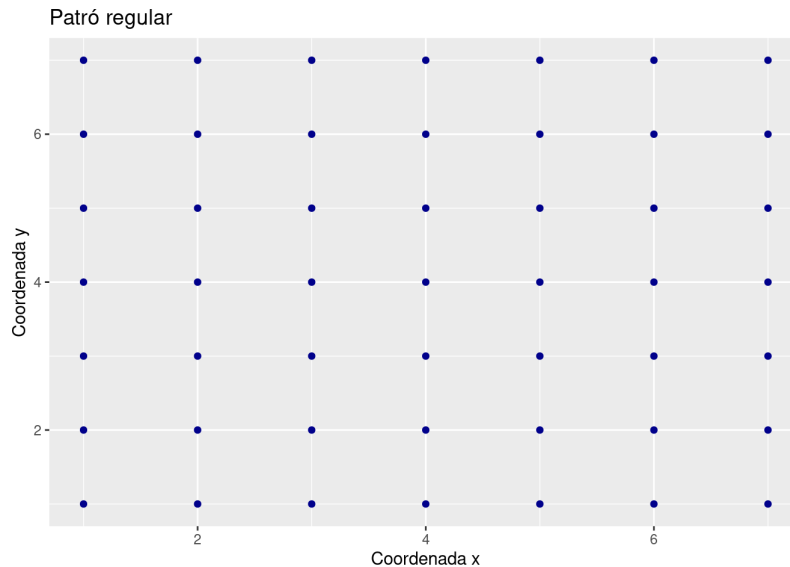
```
ggplot(aleatoris, aes(x = x, y = y)) +  
  geom_point(color = "darkblue") +  
  labs(  
    title = "Patró aleatori",  
    y = "Coordenada y",  
    x = "Coordenada x"  
  )
```



Patró regular

En aquest cas tenim un patró repetitiu en l'espai. Generem un conjunt de 49 punts (7×7), amb la funció `rep()`.

```
regulars <- data.frame(  
  x = rep(1:7, 7),  
  y = rep(1:7, each = 7)  
)  
ggplot(regulars, aes(x = x, y = y)) +  
  geom_point(color = "darkblue") +  
  labs(  
    title = "Patró regular",  
    y = "Coordenada y",  
    x = "Coordenada x"  
  )
```



Refs.:

RDocumentation. (n.d.). <https://www.rdocumentation.org/>

Exercici 2: Càlcul de Mesures Centrogràfiques. Generem un conjunt de punts amb `rnorm()` tal com hem fet en l'exercici anterior i calculem el *centroide* (o *centre geomètric*) i el *centre mitjà ponderat*, visualitzant el resultat.

Podem generar un nou conjunt de punts agregats o bé podem aprofitar el conjunt que hem generat a l'exercici 1. Triem aquesta darrera opció, i.e., el nostre conjunt de punts és agregats.

- Per calcular el *centre geomètric* fem servir la llibreria `aspace` i la funció `calc_mnc()`.

```
library(aspace)

geom_center <- calc_mnc(id = 1, weighted = FALSE,
                        weights = NULL, points = agregats)
```

Observem que el centre geomètric per N mostres es pot calcular:

$$\bar{X} = \frac{\sum_{i=1}^N x_i}{N} \quad , \quad \bar{Y} = \frac{\sum_{i=1}^N y_i}{N}$$

Podem comprovar que és efectivament el que hem trobat, (tenim $N = 17 \cdot 3 = 51$ punts):

```
geom_center$ATTRIBUTES # Resultat de calc_mnc()
sum(agregats$x)/51     # Resultat 'manual'
sum(agregats$y)/51
```

De les dues maneres trobem que el punt del centre és $\bar{X} = 3.430841$, $\bar{Y} = 1.846658$.

- La següent mesura centrogràfica que volem calcular és el *centre mitjà ponderat*, en aquest cas assignem a cada punt un pes. Podem fer el càlcul amb la mateixa funció `calc_mnc()`, indicant els pesos w_i i també amb les expressions:

$$\bar{X}_w = \frac{\sum_{i=1}^N w_i x_i}{\sum_{i=1}^N w_i} \quad , \quad \bar{Y}_w = \frac{\sum_{i=1}^N w_i y_i}{\sum_{i=1}^N w_i}$$

Podríem generar un pesos amb un cert patró per tal de simular alguna situació, (per exemple diferents pesos pels diferents clústers), però el que farem és generar 51 pesos aleatoris.

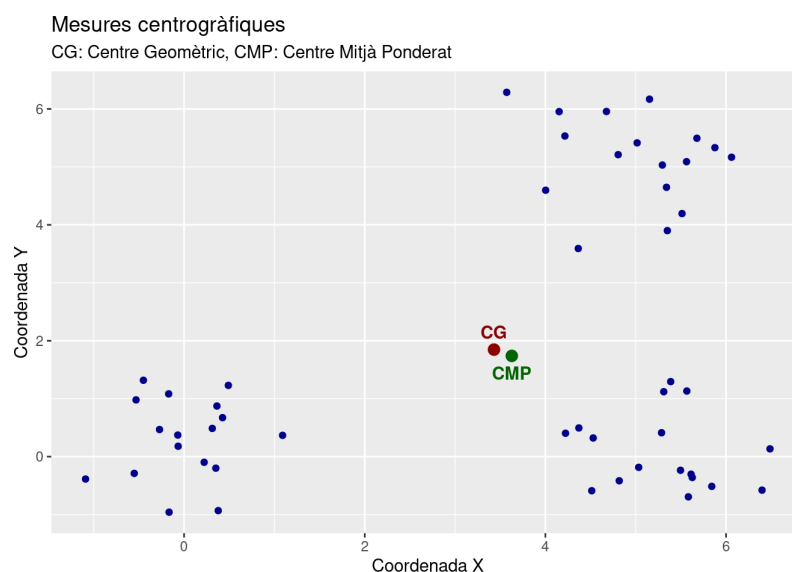
```
# Generem un vector de 51 nombres aleatoris
pesos <- runif(n = 51, min = 0, max = 1)

# Càlcul del centre mitjà ponderat
pond_center <- calc_mnc(id = 1,
                        weighted = TRUE,
                        weights = pesos,
                        points = agregats)

# Comprovació
pond_center$ATTRIBUTES
sum(agregats$x * pesos)/sum(pesos)
sum(agregats$y * pesos)/sum(pesos)
```

Obtenim el mateix resultat, $\bar{X}_w = 3.628007$, $\bar{Y}_w = 1.738414$.

Podem fer la visualització:



Observem que els dos punts obtinguts estan molt a prop, això és degut a que hem generat pesos aleatoris i uniformes per cada punt.

El codi del gràfic anterior és:

```
ggplot(agregats, aes(x = x, y = y)) +
  geom_point(color = "darkblue") +
  geom_point(geom_center$ATTRIBUTES,
    mapping = aes(x = CENTRE.x, y = CENTRE.y),
    color = 'darkred',
    size = 3) +
  geom_text(geom_center$ATTRIBUTES,
    mapping = aes(x = CENTRE.x, y = CENTRE.y,
      label='CG'),
    nudge_y=0.3,
    color='darkred', fontface='bold') +
  geom_point(pond_center$ATTRIBUTES,
    mapping = aes(x = CENTRE.x, y = CENTRE.y),
    color = 'darkgreen',
    size = 3) +
  geom_text(pond_center$ATTRIBUTES,
    mapping = aes(x = CENTRE.x, y = CENTRE.y,
      label='CMP'),
    nudge_y=-0.3,
    color='darkgreen', fontface='bold') +
  labs(
    title = "Mesures centrogràfiques",
    subtitle = "CG: Centre Geomètric, CMP: Centre Mitjà Ponderat",
    y = "Coordenada Y",
    x = "Coordenada X"
  )
```

Exercici 3: Exploració de l'Autocorrelació Espacial. Es tracta de definir què és l'autocorrelació espacial i per què és important en l'anàlisi de dades espacials i discutir breument un exemple visual d'aplicació.

Podem dir que l'*autocorrelació espacial* és una mesura que ens descriu el grau en que observacions, fenòmens o objectes en localitzacions espacials (punts, àrees, celes, ...) veïns tinguin semblança o valors semblants per una variable donada.

Pot ser positiva (valors semblants), negativa (valors diferents) o zero (sense relació) i els indicadors d'autocorrelació espacial que més es fan servir són l'*índex I de Moran* i el *paràmetre c de Geary*.

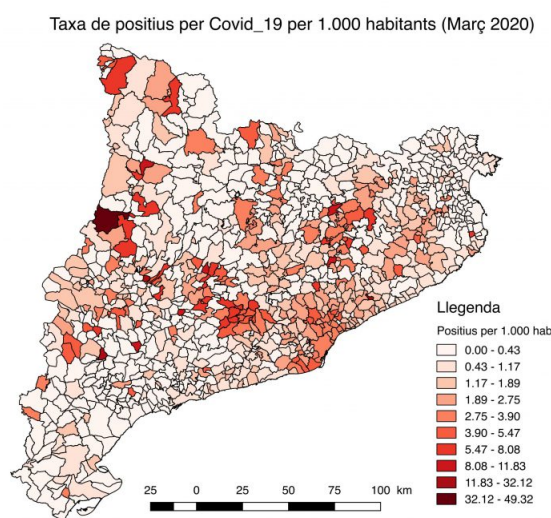
Hem vist diferents patrons espacials, però d'una manera general, fenòmens o observacions propers a l'espai presenten una dependència amb valors semblants per una o més variables, comparat amb observacions més llunyanes, (primera llei de Tobler o principi d'autocorrelació espacial).

És aquesta dependència espacial la que mesurem amb l'autocorrelació espacial i ens dona el grau en que una variable està relacionada amb ella mateixa.

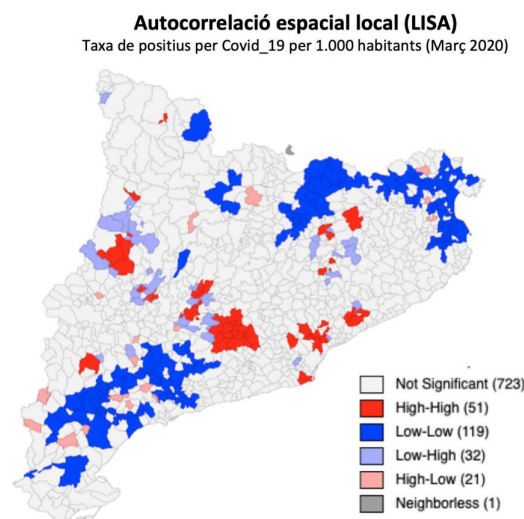
Alguns autors defineixen l'autocorrelació espacial com una extensió de la correlació temporal, amb la diferència que aquesta té una sola dimensió i la primera en té dues com a mínim, una forma que pot ser complexa i la dificultat de definir el que entenem per proximitat.

L'autocorrelació espacial és important perquè ens indica que les observacions no són independents, hi ha influències entre objectes i propietats espacials que mereixen una anàlisi més profunda i a la vegada ens pot dir que potser hi ha informació redundant en el nostre anàlisi de dades.

Com a exemple podem considerar *la geografia de la COVID-19*. En el següent mapa podem veure la situació de la COVID-19 el març de 2020.



Es poden posar de manifest diferents clústers (llocs positius per sobre de l'esperat) usant mesures d'autocorrelació. Podem comparar si l'indicador de la variable *casos positius per cada 1000 habitants* en un determinat municipi està relacionat amb els valors d'aquesta variable en municipis veïns. Això ens ho mostra el següent mapa:



Hi podem veure zones vermelles amb nivells de COVID alts, envoltades de zones vermelles, zones

blaves amb nivells baixos envoltades de municipis amb nivells baixos i zones grises sense cap relació significativa.

Referències:

Geografia de la COVID-19. (2020, April 3). Diari Digital De La URV. <https://diaridigital.urv.cat/es/opinions/geografia-de-la-covid-19/>

Spatial autocorrelation — R Spatial. (n.d.). <https://rspatial.org/analysis/3-spauto.html>

Exercici 4: Descripció i Aplicació de Mètodes d'Interpolació. Considerem tres mètodes d'interpolació (Polígons de Thiessen, Triangulació de Delaunay i Distància Inversa Ponderada) i exemples de les seves aplicacions pràctiques.

Donat un conjunt de valors d'una variable, entenem per *interpolació espacial*, el procés mitjançant el qual fem estimacions del valor d'aquesta variable en posicions de les quals la desconexim.

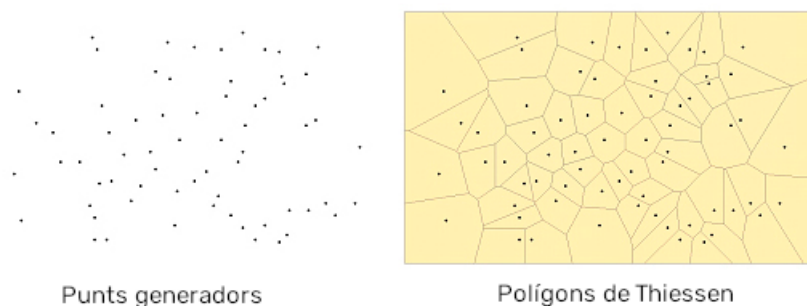
Normalment se sol partir d'un conjunt de punts distribuïts de forma irregular i s'obté una superfície en forma de malla regular.

Polígons de Thiessen

També coneguts com a *diagrames de Voronoi*, (una de les aplicacions més conegudes és la que va fer John Snow per determinar l'origen del brot de cólera a Londres el 1854).

Amb aquest mètode, assignem a cada posició el valor desconegut d'una variable igual al valor d'aquesta variable en el punt més proper que sí coneixem.

Partim d'un conjunt de punts amb valors de la variable coneguts (*llavors*) i anem fent rectes mediatrises, obtenint una zona tessellada que cobreix tota la superfície que ens interessa, de manera que cada punt està envoltat d'una àrea en que la variable té el mateix valor.

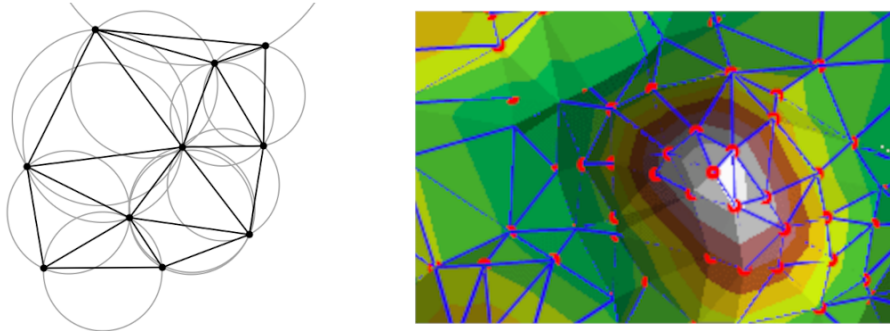


Val a dir que aquest mètode només és aconsellable per variables categòriques discretes (en el cas de John Snow va considerar les morts per cólera).

Triangulació de Delaunay

És una tècnica per crear una malla irregular de triangles (*TIN*), contigus, no superposats a partir d'un conjunt de punts.

Fem cercles que circumscriuen els triangles, de manera que no contenen cap altre punt del conjunt en el seu interior. Obtenim una representació *vectorial* d'una superfície on cada triangle té el mateix pendent i orientació.



A diferència del model ràster, que normalment és una malla regular formada per quadrats, el model *TIN* ens permet modelar amb més precisió superfícies heterogènies, pel fet de ser irregulars i amb un emmagatzematge menys exigent.

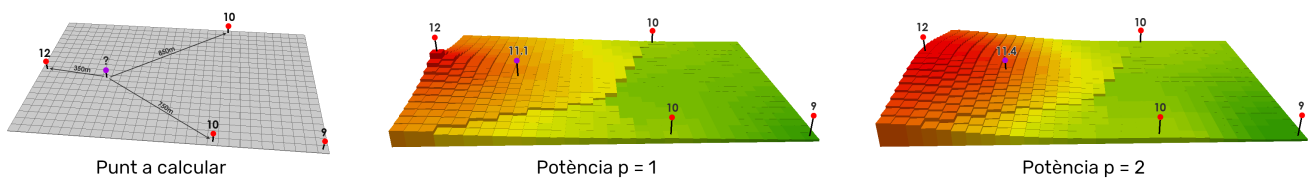
Distància inversa ponderada (IDW)

En aquest cas construïm una malla regular on cada posició té un valor mig que s'obté dels valors ponderats dels punts coneguts més propers. Aquest mètode fa ús de l'autocorrelació en el seu càlcul.

El pes de cada punt on fem el càlcul és proporcional a la inversa de la distància a cada punt de la malla, elevat a una potència:

$$\bar{Z}_j = \frac{\sum_{i=1}^n z_i / d_{ij}^p}{\sum_{i=1}^n 1 / d_{ij}^p}$$

Normalment s'indica el nombre n de punts més propers o els punts situats per sota d'una distància determinada. Podem veure un exemple de com calculem el valor del punt amb un interrogant en la següent figura, considerant 3 punts més propers.



L'expressió per calcular el valor, considerant les distàncies donades (en metres) i considerant $p = 1$ i $p = 2$, és:

$$\bar{Z}_? = \frac{(12/350^p) + (10/750^p) + (10/850^p)}{(1/350^p) + (1/750^p) + (1/850^p)}$$

Observem que per potències més grans els pesos dels punts propers augmenten. Els valors d'interpolació que obtenim estaran entre el màxim i mínim dels valors dels punts que fem servir.

Exemples pràctics d'aplicació

Tot seguit veurem algunes aplicacions pràctiques de cada un dels mètodes que hem vist.

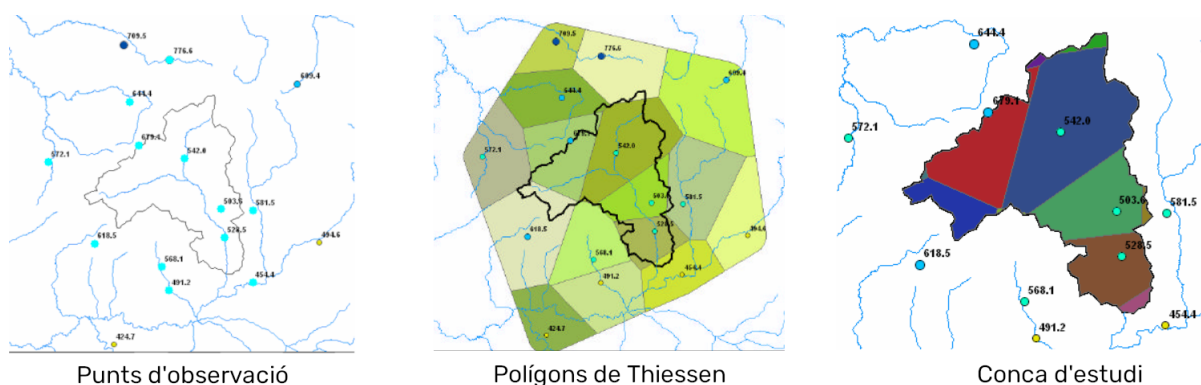
Polígons de Thiessen

Val a dir que els polígons de Thiessen són els diagrames de Voronoi aplicats en geofísica, meteorologia i en geografia. Reben aquest nom pel meteoròleg nord-americà Alfred H. Thiessen (1911) que ho va aplicar a l'anàlisi de la distribució espacial d'observacions meteorològiques com les precipitacions.

Les aplicacions d'aquest mètode parteixen de diferents punts d'observació dels quals tenim enregistraments i a partir d'ells es fan estimacions. Així podem citar entre d'altres, els següents exemples:

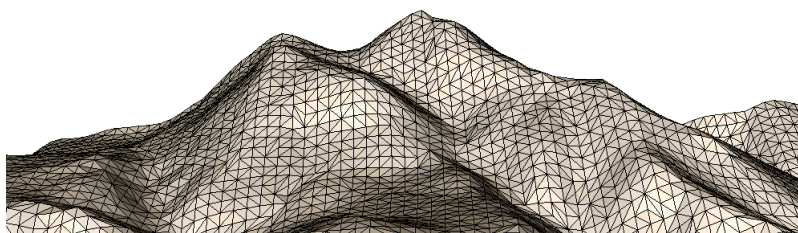
- En climatologia, s'utilitzen per a calcular totals o mitjanes de paràmetres com les precipitacions per a àrees a partir dels punts de les estacions meteorològiques.
- En ecologia, els polígons de Thiessen permeten interpolat valors de variables com la coberta forestal o el patró de creixement dels boscos per a àrees a partir de punts d'observació, i també generar mapes de models de combustible per a models de simulació d'incendis forestals.

Gràficament, pel cas del càlcul de precipitació d'una conca, apartir d'uns punts d'observació:



Triangulació de Delaunay

Potser l'aplicació més usual d'aquest mètode és el modelatge topogràfic on tenim diferències d'altures. Hi ha diferents llibreries de software que ens permeten generar superfícies amb aquest mètode d'una manera simplificada i optimitzada que representen un terreny.

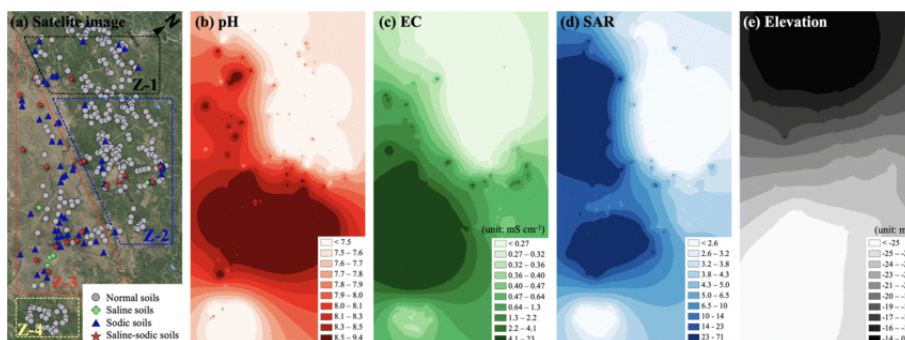


Distància inversa ponderada (IDW)

En quant al seu ús, podem dir que cal tenir en compte que treballem amb mitjanes ponderades i per tant si estem modelant un terreny, no podem crear carenes o valls si aquestes no han estat registrades. També pot passar que obtinguem un efecte d'ull de bou al voltant dels punts. Finalment aquest mètode funciona bé per conjunts de dades grans.

Com a exemple d'aplicacions podem citar:

- Geologia: per fer una estimació del gruix de capes sedimentàries, la profunditat de la roca base i la ubicació de dipòsits minerals.
- Seguiment ambiental, com ara l'estimació de concentracions de contaminants. Es poden detectar zones on cal una actuació precisament per aquesta contaminació. Per exemple, en la següent figura podem veure, a partir d'una imatge de satèl·lit(a) i amb una anàlisi amb interpolació IDW, el pH(b), la conductivitat elèctrica (c), la relació d'adsorció de Na (d) i l'elevació del terreny (e).



Referències:

Applications of inverse distance weighting in geostatistics - FasterCapital. (n.d.). FasterCapital. <https://fastercapital.com/topics/applications-of-inverse-distance-weighting-in-geostatistics.html>

GISGeography. (2024, March 9). *Inverse Distance weighting (IDW) interpolation.* GIS Geography. <https://gisgeography.com/inverse-distance-weighting-idw-interpolation/>

Martinez, F. J. S. (2007, January 1). *Cálculo de la precipitación areal mediante sistemas de información geográfica.* EOI. <https://www.eoi.es/es/savia/publicaciones/77678/calculo-de-la-precipitacion-areal-mediante-sistemas-de-informacion-geografica>

Polígons de Thiessen. Institut Cartogràfic i Geològic de Catalunya. (n.d.). <https://www.icgc.cat/Ciutada/Informa-t/Diccionaris/Poligons-de-Thiessen> Imatge: https://www.researchgate.net/figure/Spatial-analysis-results-of-salinity-factors-using-IDW-interpolation-for-satellite-images_fig4_348847563