

Systèmes d'Information Géographique

<https://go.epfl.ch/sig>

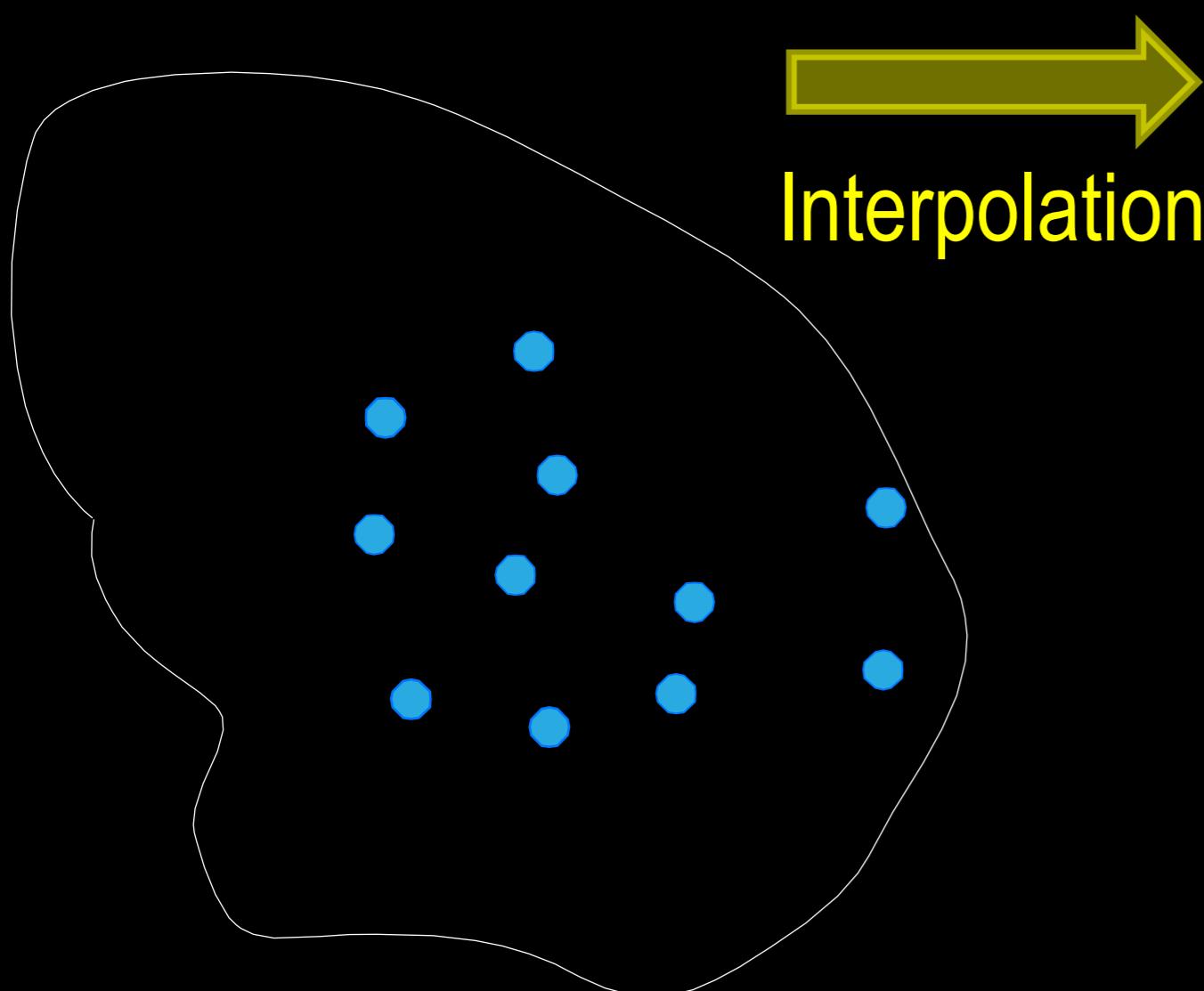
Phénomènes spatiaux continus - Interpolation

Stéphane Joost, Gabriel Kathari (GEOME-LGB)

Phénomènes spatiaux continus – Interpolation

Echantillon

Variable mesurée en certains points



Estimation de la valeur en un point non mesuré



- Les points échantillonnés doivent fournir des informations utiles sur le domaine auquel ils appartiennent
- On peut obtenir des informations utiles sur la zone étudiée en utilisant des indicateurs statistiques comme la médiane ou la moyenne, mais globales!
- L'intérêt principal de l'échantillonnage est d'estimer la valeur de la variable en tout point (local) du domaine par interpolation
- Une interpolation est une inférence, donc une prédiction d'une valeur de la variable en un point non mesuré

Méthodes d'interpolation déterministes

- Aucune étude statistique nécessaire sur le comportement de la variable (contrairement au krigage et à la variographie)
- Le choix de la méthode d'interpolation est basé sur une connaissance experte préalable du phénomène (et des méthodes existantes)

Méthodes déterministes globales versus locales

Méthode globale

- Prend en compte tous les points mesurés

Régression polynomiale

Méthodes locales

- Prend en compte uniquement un nombre limité de points mesurés, au voisinage du point à estimer
- Permet aussi d'attribuer des poids particuliers, notamment en fonction de la distance

Plus proche voisin

Réseau de triangles

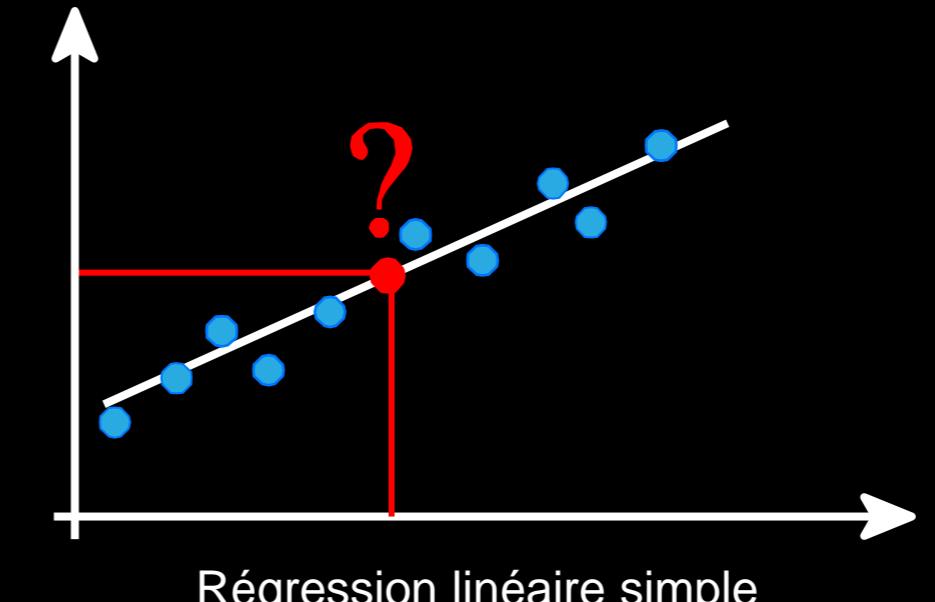
Moyenne pondérée

Méthode globale: régression polynômiale

- Polynôme qui minimise les écarts aux points de mesure (moindres carrés)
- Création d'une surface de tendance
- Degré inférieur à 4: modèle souvent trop simple pour fournir une approximation réaliste du phénomène
- Degré supérieur à 5: difficile à calculer

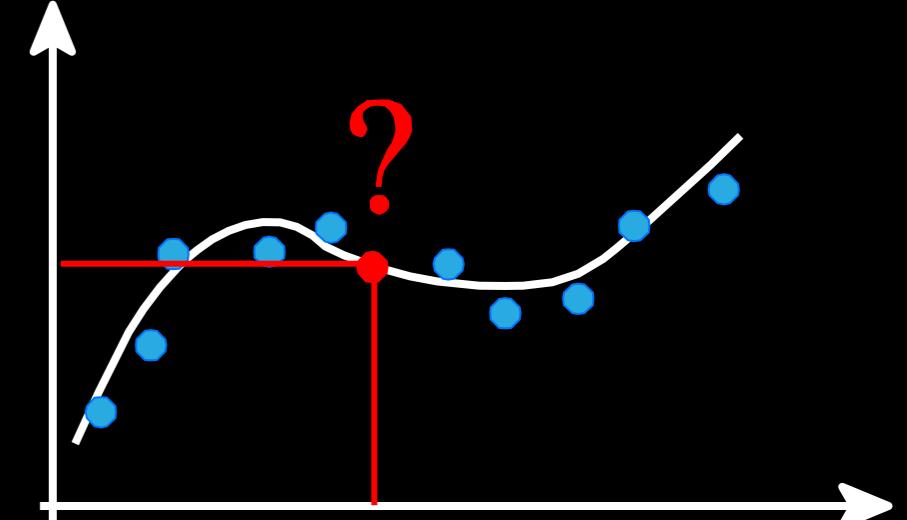
→ Méthodes locales souvent préférées

Degré 1

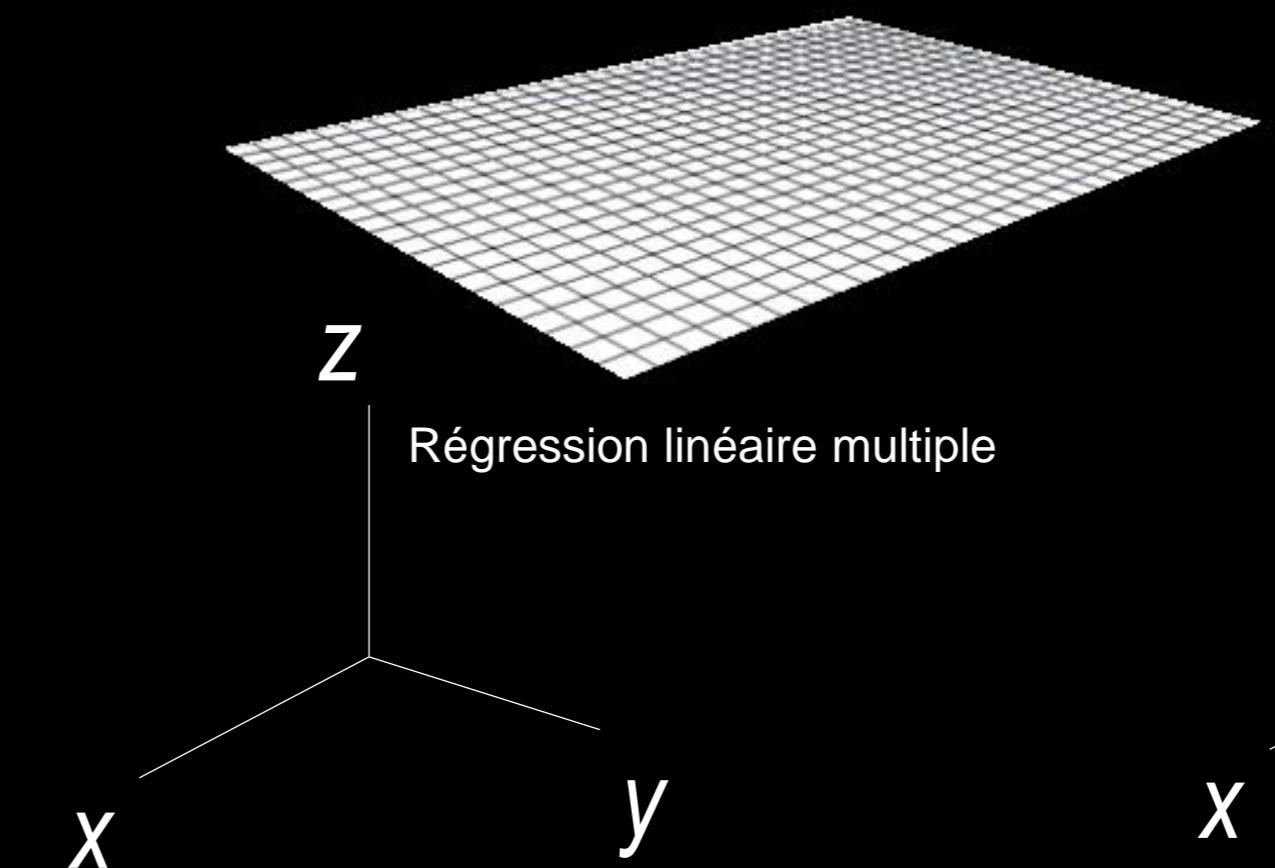


Régression linéaire simple

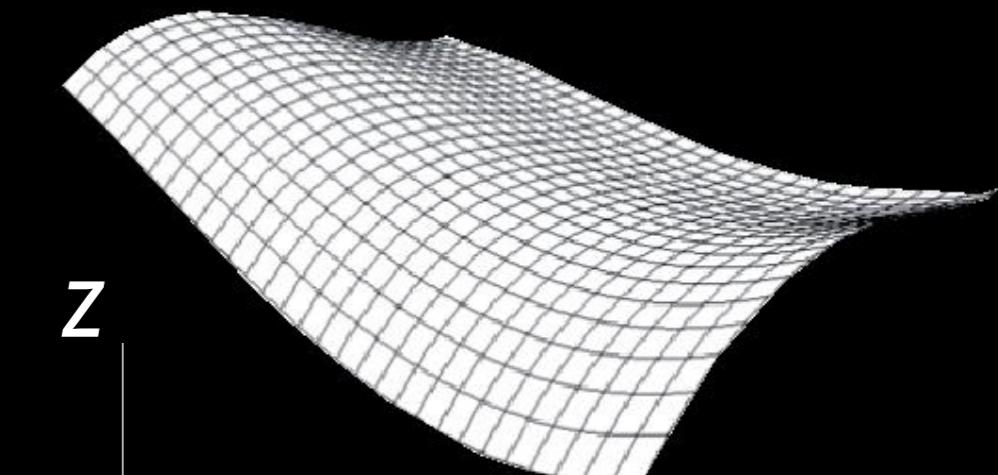
Degré > 1



Régression polynomiale simple



Régression linéaire multiple



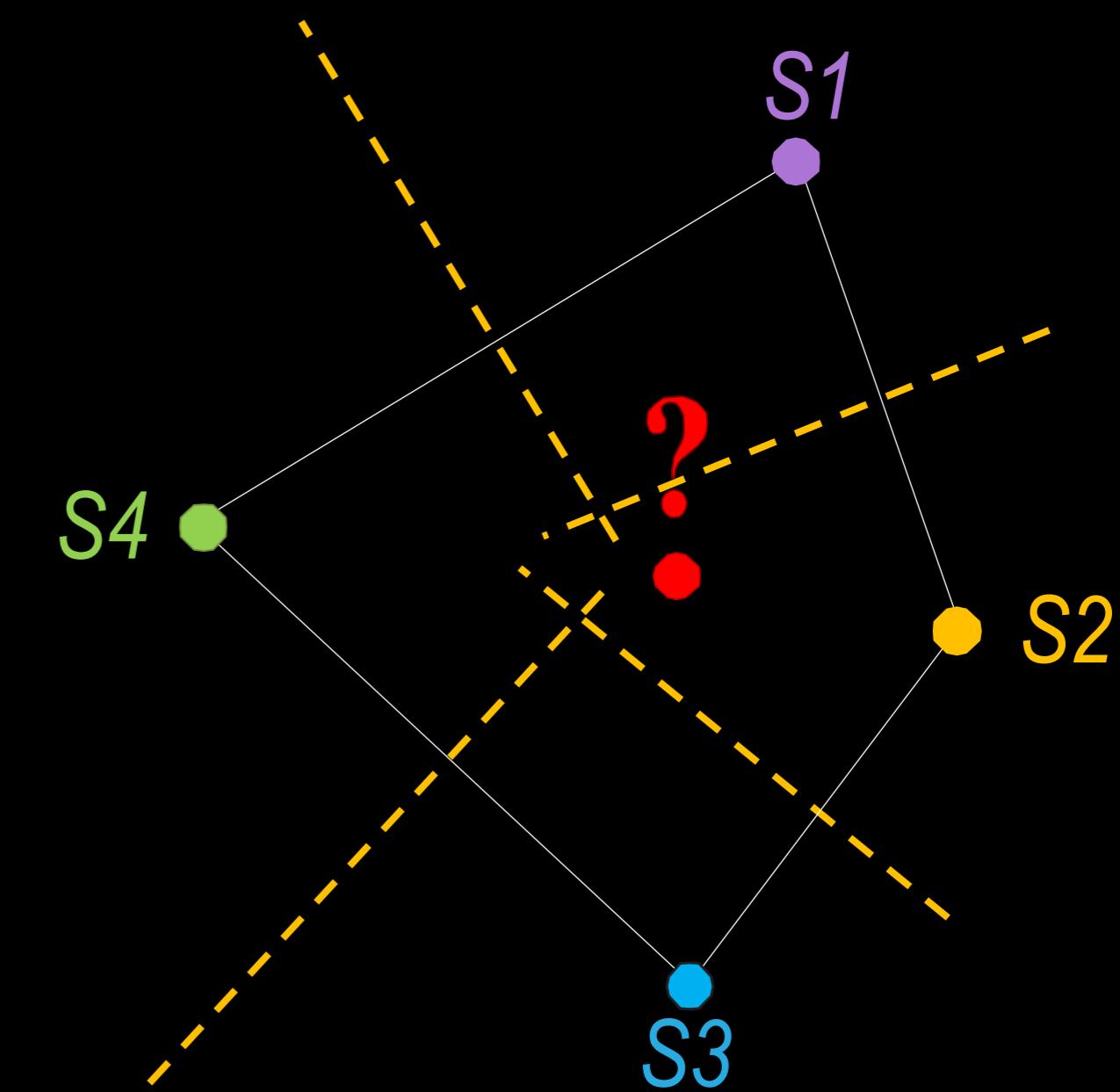
Régression polynomiale multiple

Méthodes locales

- Nombre limité de points
- Pondération des points en fonction de leur distance au point à prédire
- Plusieurs méthodes dont le choix dépend de:
 - La densité d'information
 - La précision souhaitée
 - L'objectif de l'interpolation
 - La connaissance experte préalable du phénomène

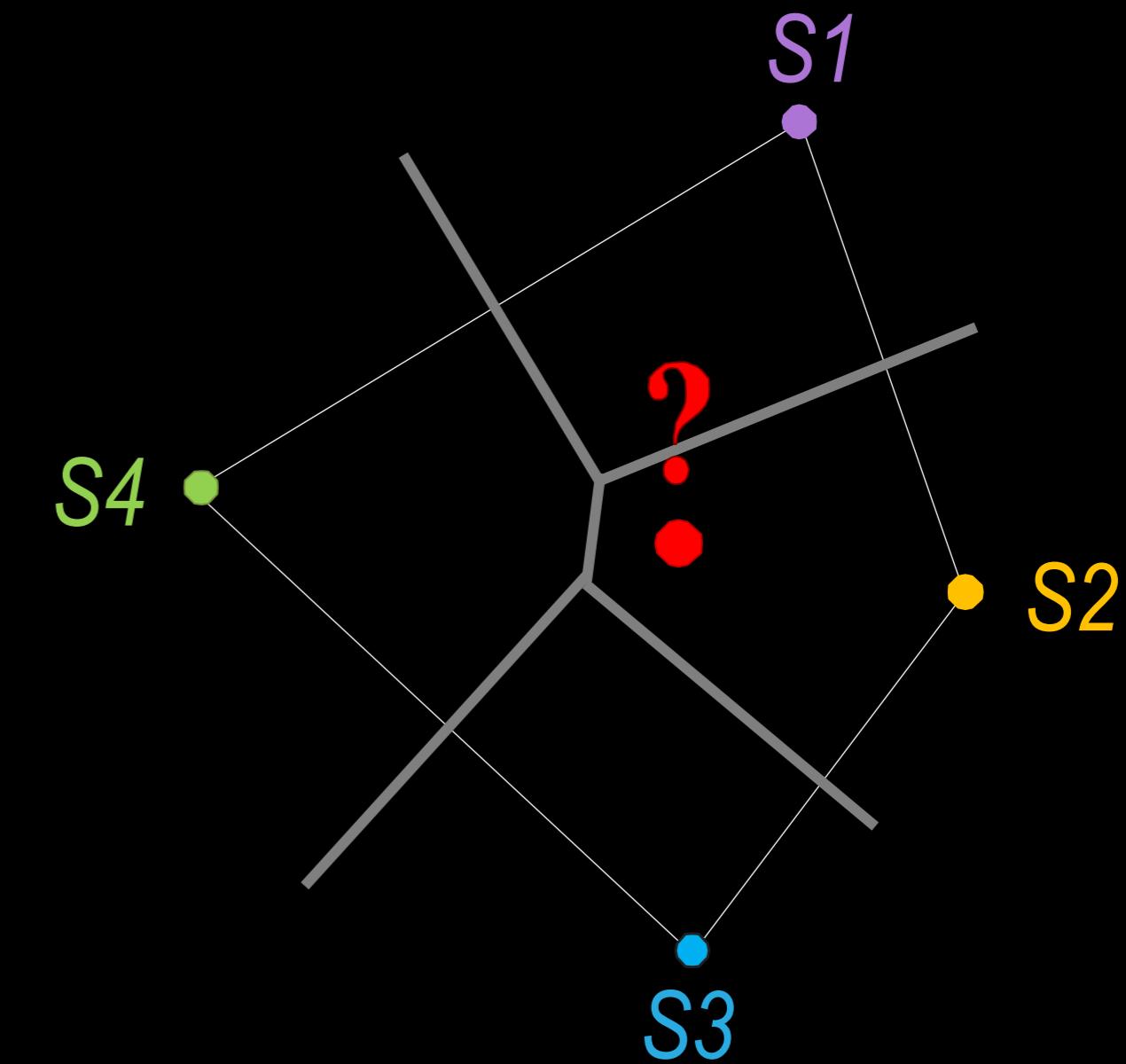
Plus proche voisin

- Valeur estimée = valeur du point mesuré le plus proche
- Polygones de Thiessen (médiatrices des segments reliant les points mesurés)



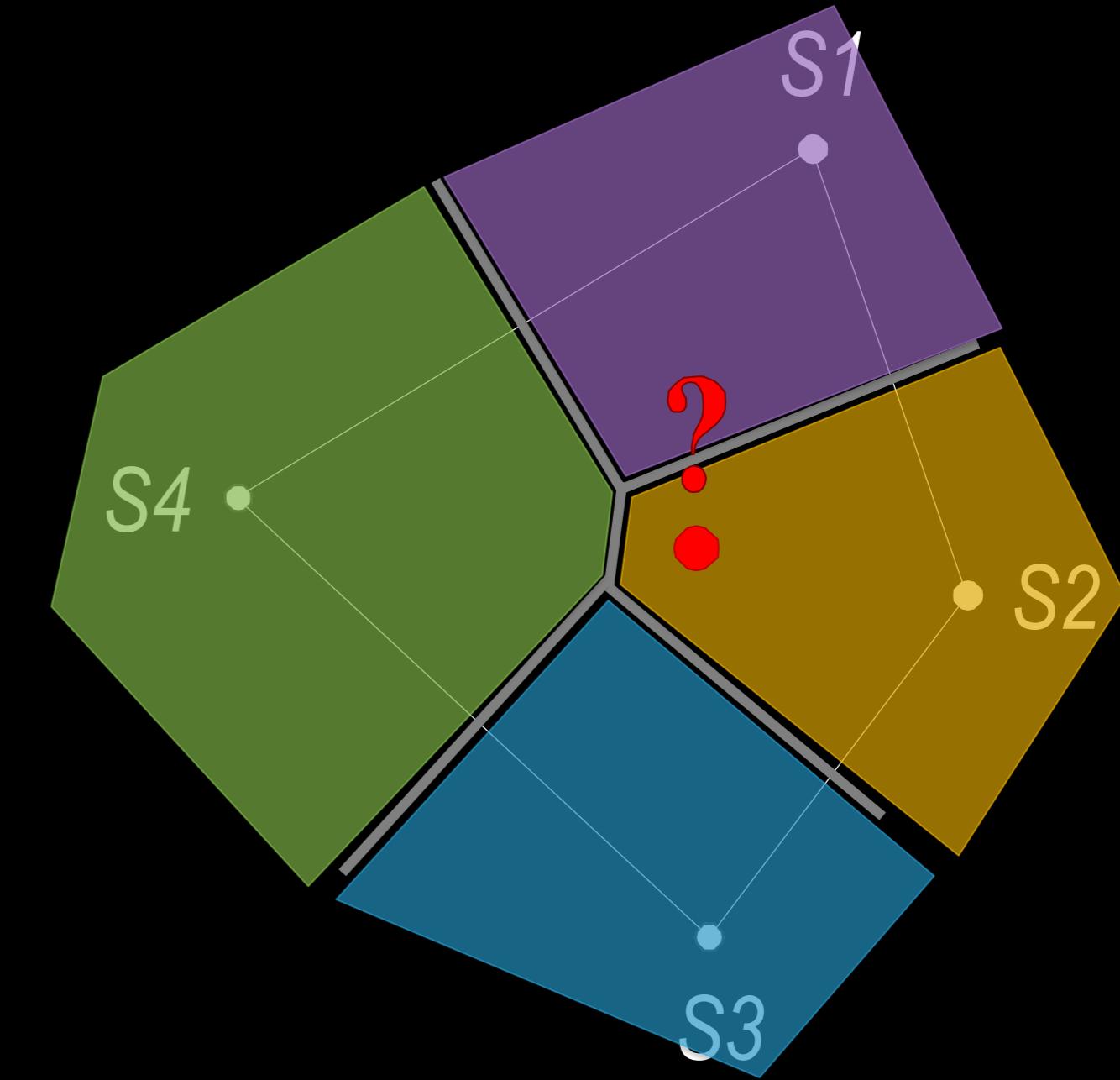
Plus proche voisin

- Valeur estimée = valeur du point mesuré le plus proche
- Polygones de Thiessen (médiatrices des segments reliant les points mesurés)



Plus proche voisin

- Valeur estimée = valeur du point mesuré le plus proche
- Polygones de Thiessen (médiatrices des segments reliant les points mesurés)
- Les points à estimer qui tombent à l'intérieur d'un même polygone obtiennent la même valeur

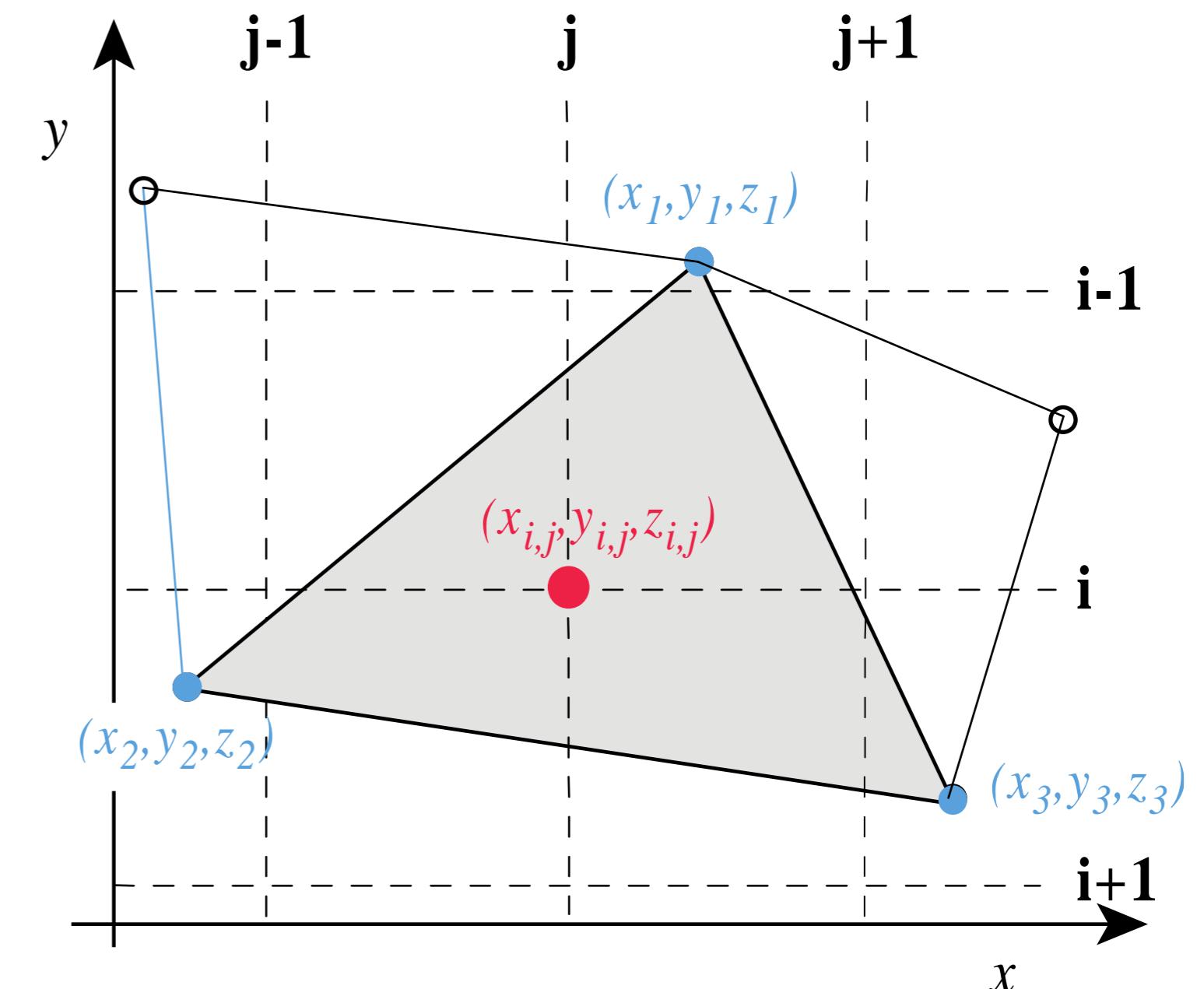


→ valeur ? = valeur S2

Réseau de triangles (triangulated irregular network TIN)

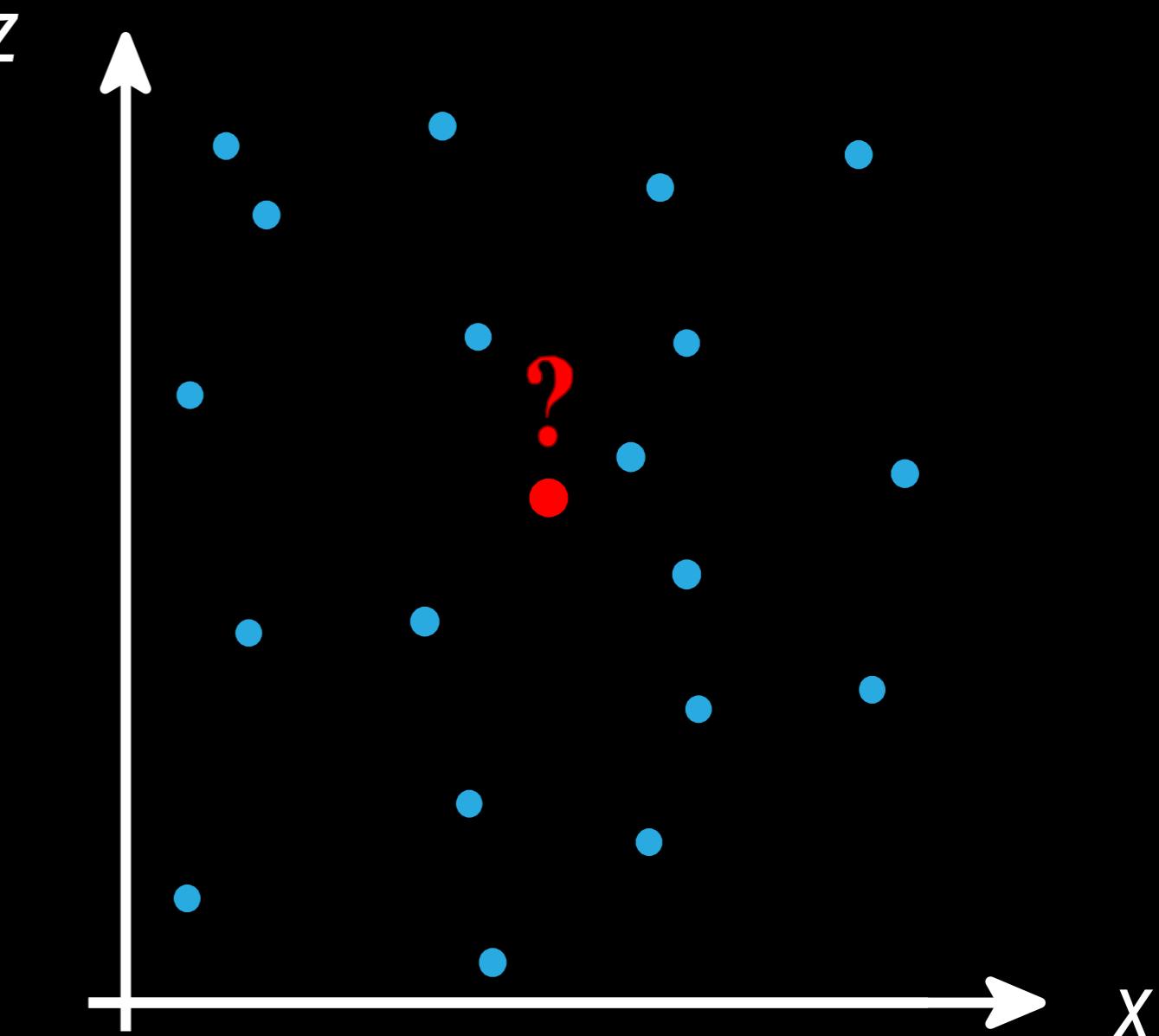
- Un TIN est formé en reliant les points d'échantillonnage entre eux pour former un réseau de mailles triangulaires
- La création des mailles triangulaires est réalisée par la méthode de Delaunay - création de triangles dont les angles ne sont pas supérieurs à 90°
- La variation du paramètre à l'intérieur de chaque triangle est supposée linéaire
- L'interpolation est réalisée en établissant l'équation du plan formé par chacune des faces triangulaires
- Il y a 3 inconnues, les coefficients a, b, c. Ils sont déterminés par les valeurs connues des 3 sommets du triangle

$$z = ax + by + c$$



Moyenne mobile pondérée (inverse distance weighting IDW)

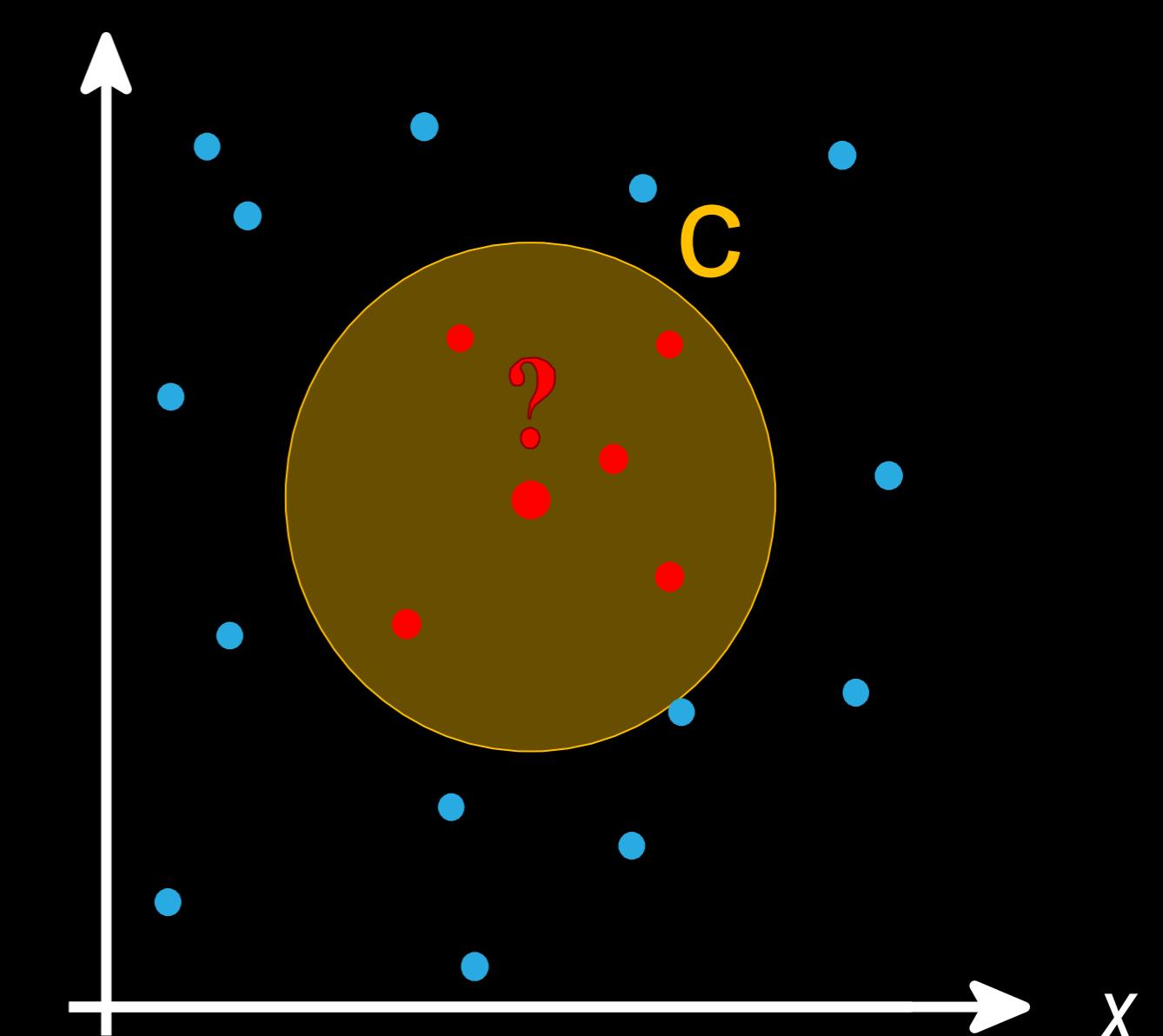
- Estimation en s'appuyant sur les valeurs mesurées de quelques points du voisinage
- En leur attribuant un poids en fonction de la distance (un point éloigné a moins d'influence qu'un point rapproché)



Moyenne mobile pondérée (IDW)

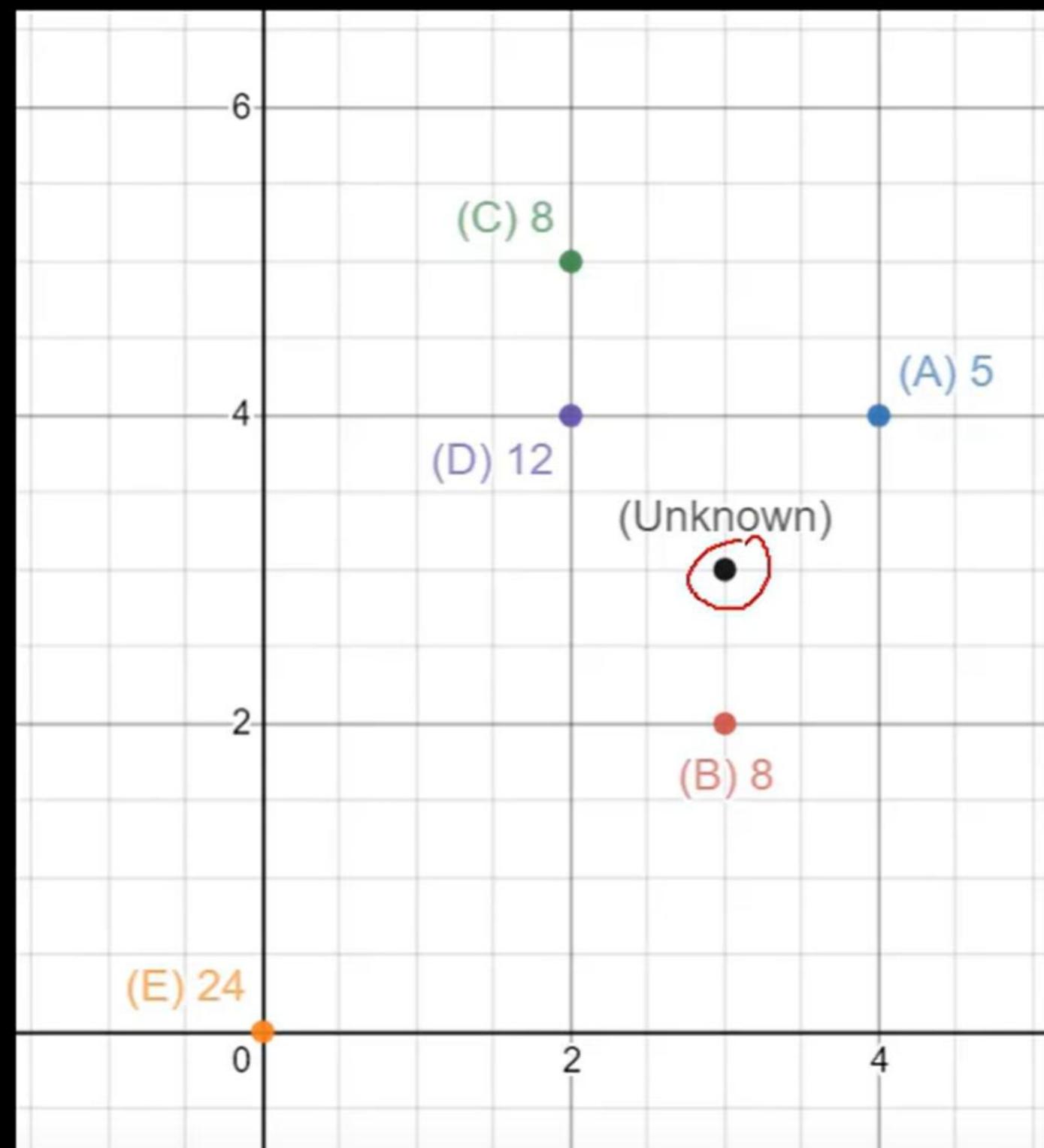
- Estimation en s'appuyant sur les valeurs mesurées de quelques points du voisinage
- Paramètres à définir:
 - Taille du voisinage (**C**) ou nombre de points pris en compte
 - Poids: les points plus éloignés ont moins d'influence que les points proches
 - Peut tenir compte de l'orientation d'un phénomène (anisotropie)

→ valeur en **?** = moyenne pondérée des voisins



$$\hat{z}(s_n) = \frac{\sum_{\alpha=1}^n \omega_{\alpha} z_{\alpha}}{\sum \omega_{\alpha}}$$

IDW exemple



$$\hat{z}(u) = \frac{\sum \omega_i Z_i}{\sum \omega_i}$$

$$\hat{z}(u) = \frac{29.25}{3.09} = 9.45$$

Point	Valeur	Distance	ω_i	$\omega_i \cdot Z_i$
A	5	1.414	1/1.414	3.53
B	8	1	1/1	8
C	8	2.23	1/2.23	3.57
D	12	1.41	1/1.41	8.49
E	24	4.24	1/4.24	5.66

$$\omega_i = \frac{1}{D}$$

Evaluation de l'incertitude - Validation croisée

- Pas de moyen d'estimer l'incertitude directement
- **Validation croisée**
 - Partager aléatoirement l'échantillon en deux
 - Un groupe pour l'interpolation
 - Un groupe pour les tests de validation
 - Comparer la valeur interpolée à la valeur mesurée en chaque site test
 - En déduire l'incertitude (différence des moyennes quadratiques des distributions)
 - Réitérations successives (10 tirages par exemple)

Merci pour votre attention !

