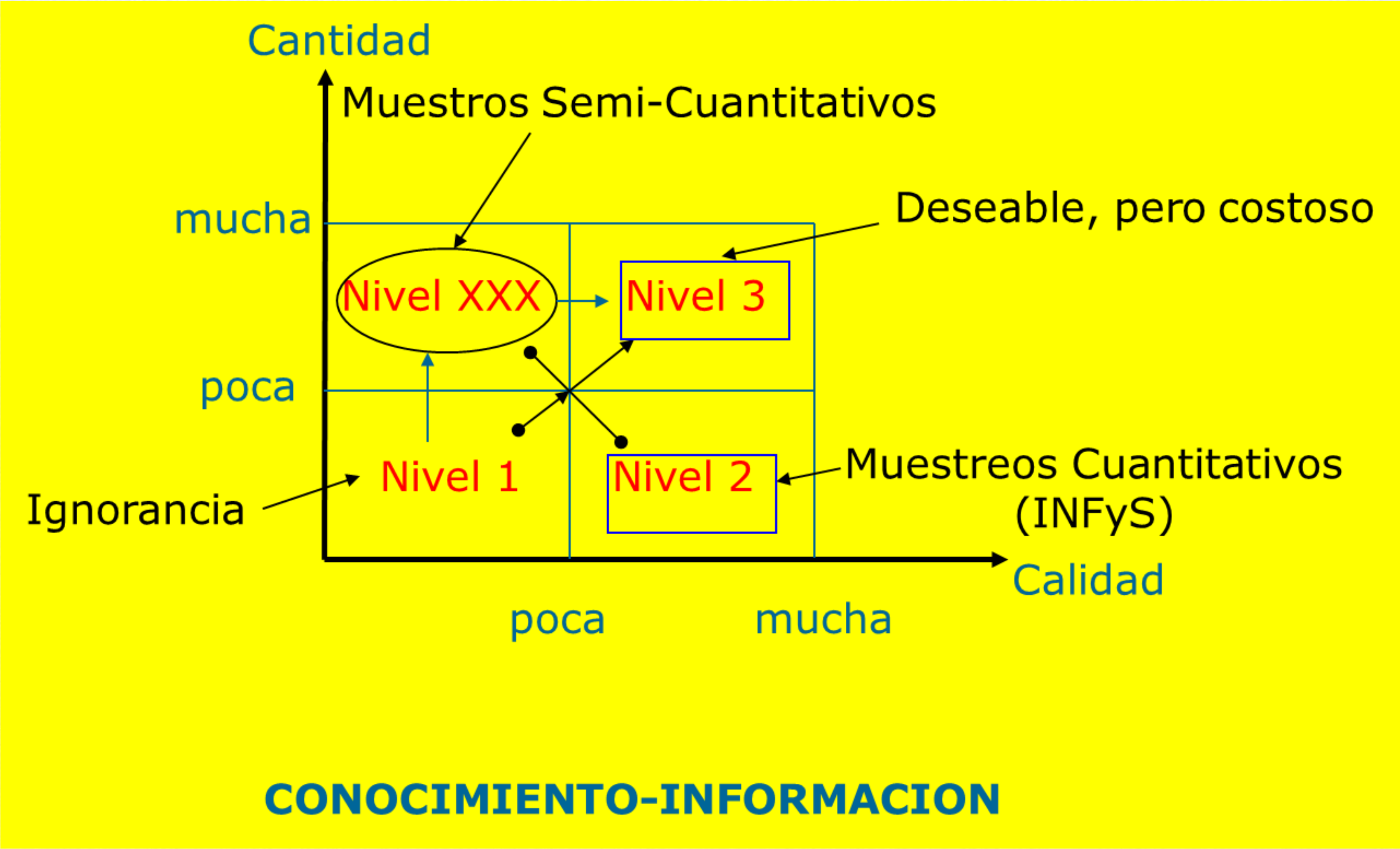
PMC

Programa Mexicano del Carbono

FUSIÓN DE INFORMACIÓN MULTI-FUENTE DE INVENTARIOS DE CARBONO

Bases Teóricas - Conceptuales

ESTRATEGIA GENERAL - INFORMACIÓN



GUIAS IPCC-CMNUCC

Los principios de la CMNUCC para la estimación y reporte de los inventarios de emisiones de GEI deben seguir los siguientes criterios:

- **Transparencia.** **Todas las hipótesis y metodologías usadas en los inventarios deben ser explicadas claramente y documentadas en forma apropiada, de tal forma que cualquiera pueda verificar si son correctas.**

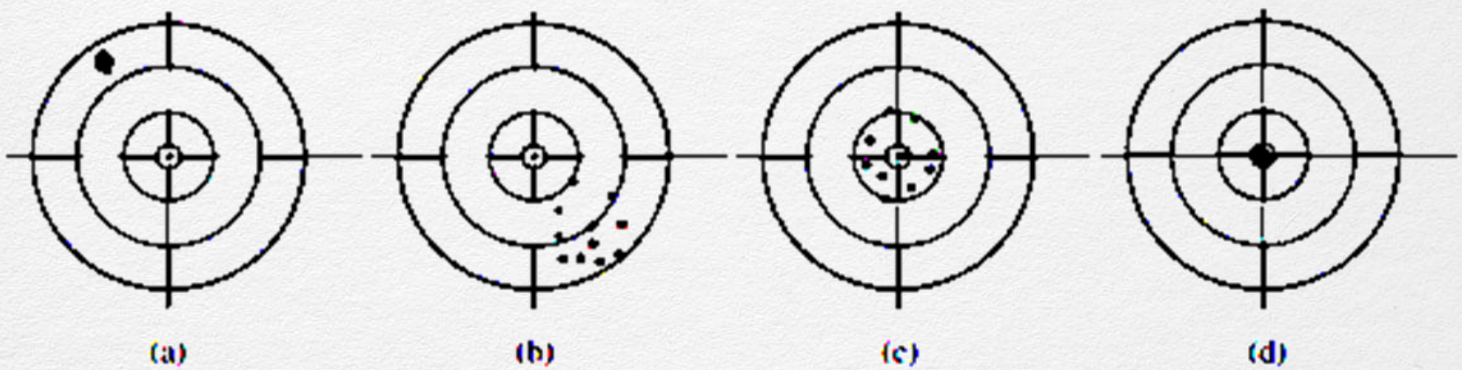
- **Consistencia.** **Un inventario debe ser internamente consistente en todos sus elementos en relación a los inventarios de otros años. Un inventario es consistente si los mismos conjuntos de datos y metodologías son usados a lo largo del tiempo. Bajo ciertas circunstancias, las estimaciones usando diferente metodologías para diferentes años pueden ser consideradas consistentes si pueden ser calculadas en forma transparente**

- **Comparabilidad.** **Las estimaciones de las emisiones deben ser comparables entre las partes (de la CMNUCC). Para este propósito, las partes deben seguir las metodologías y formatos estándares del PICC, acordados en la CMNUCC para la compilación y reporte inv.**

- **Completes.** **Los estimados deben incluir todas las categorías de uso del suelo acordadas, todos los gases y todos los almacenes.**

- **No sesgo (“accuracy”).** **En el sentido de que las estimaciones no deben estar arriba o abajo del valor verdadero, de tal forma que puedan ser evaluadas y que las incertidumbres puedan ser reducidas cuando esto es práctico. Deben usarse las metodologías apropiadas, de acuerdo con las guías de buenas prácticas del PICC, para promover que las estimaciones en los inventarios no sean sesgadas y para cuantificar las incertidumbres para mejorar los inventarios futuros.**

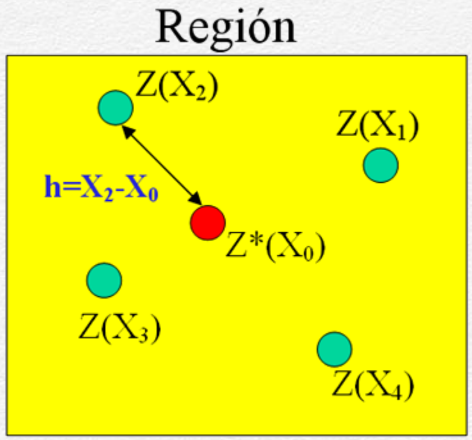
INCERTIDUMBRE



Esquematización del sesgo y la precisión de las estimaciones. (a) sesgada, pero precisa; (b) sesgada e imprecisa; (c) no sesgada, pero imprecisa; y (d) no sesgada y precisa.

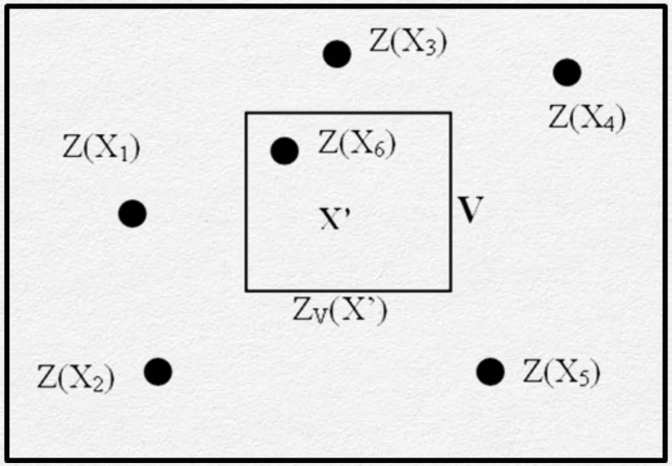
PROBLEMA DE ESTIMACIÓN

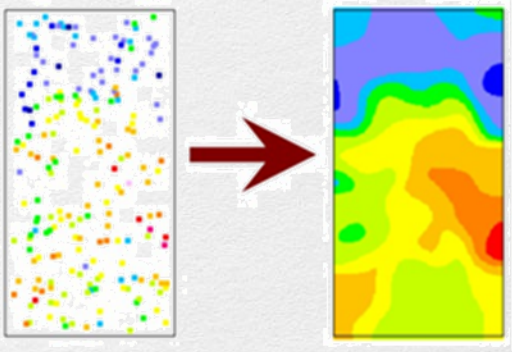
Escala fija



$$Z^*(X_0) = \sum_{i=1}^{n} \lambda_i Z(X_i)$$

Escala variable





CUMPLIENDO CON IPCC-CMNUCC

No sesgo => $E[Z - Z^*] = 0$

Precisión => $VAR[Z - Z^*]$ mínima

Z = Valor real

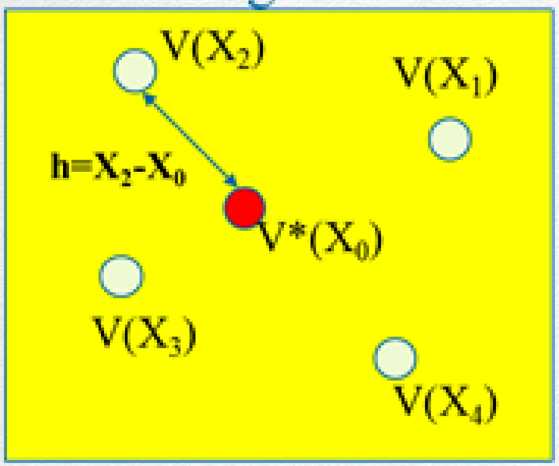
Z^* = Valor estimado

PROBLEMAS:

+ ¿Mejor estimación con respecto a?
+ Dependiente de modelo (ex ante)
+ Métrica de error
+ No sujeto a validación (ex post)

GEOSTADISTICA

Región

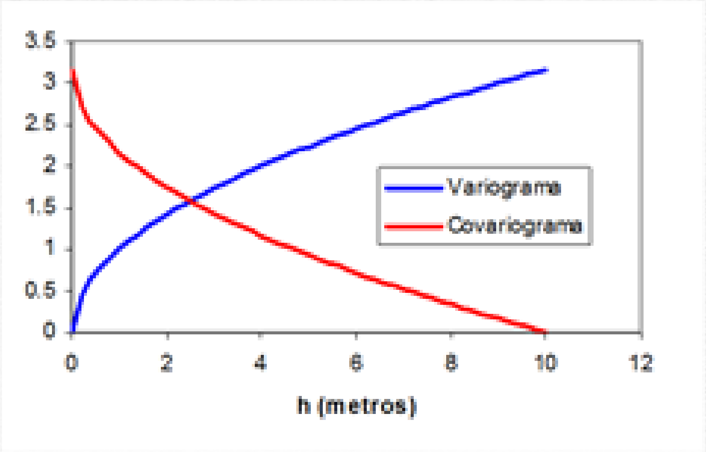


$$V^* = \sum \lambda_i V_i$$

Funciones de estructura:

$$\gamma(h) = E\left[\{V(X) - V(X+h)\}^2\right]$$

$$C(h) = E[V(x)V(x+h)] - E[V(X)]E[V(X+h)]$$

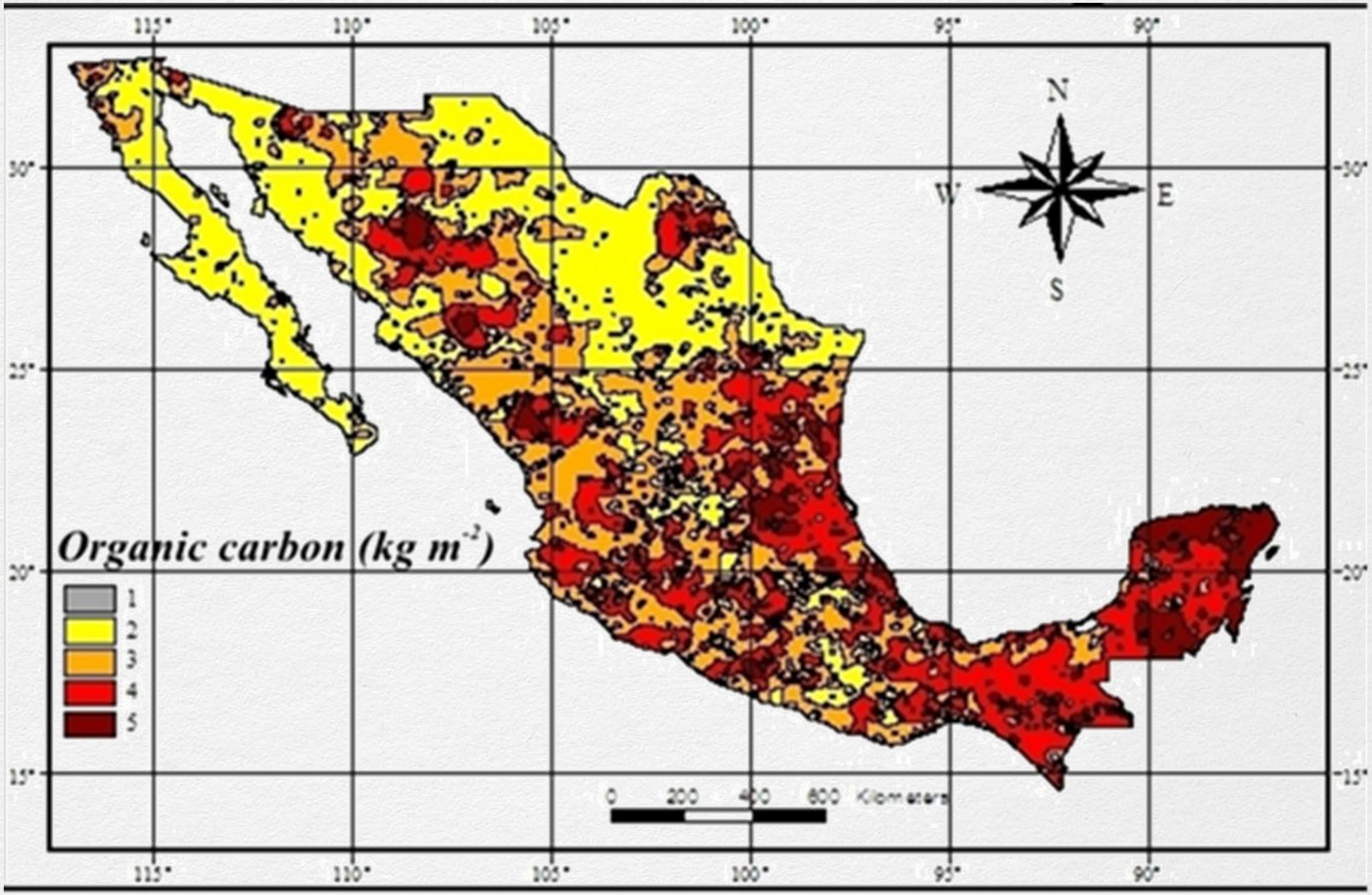


Estimación:

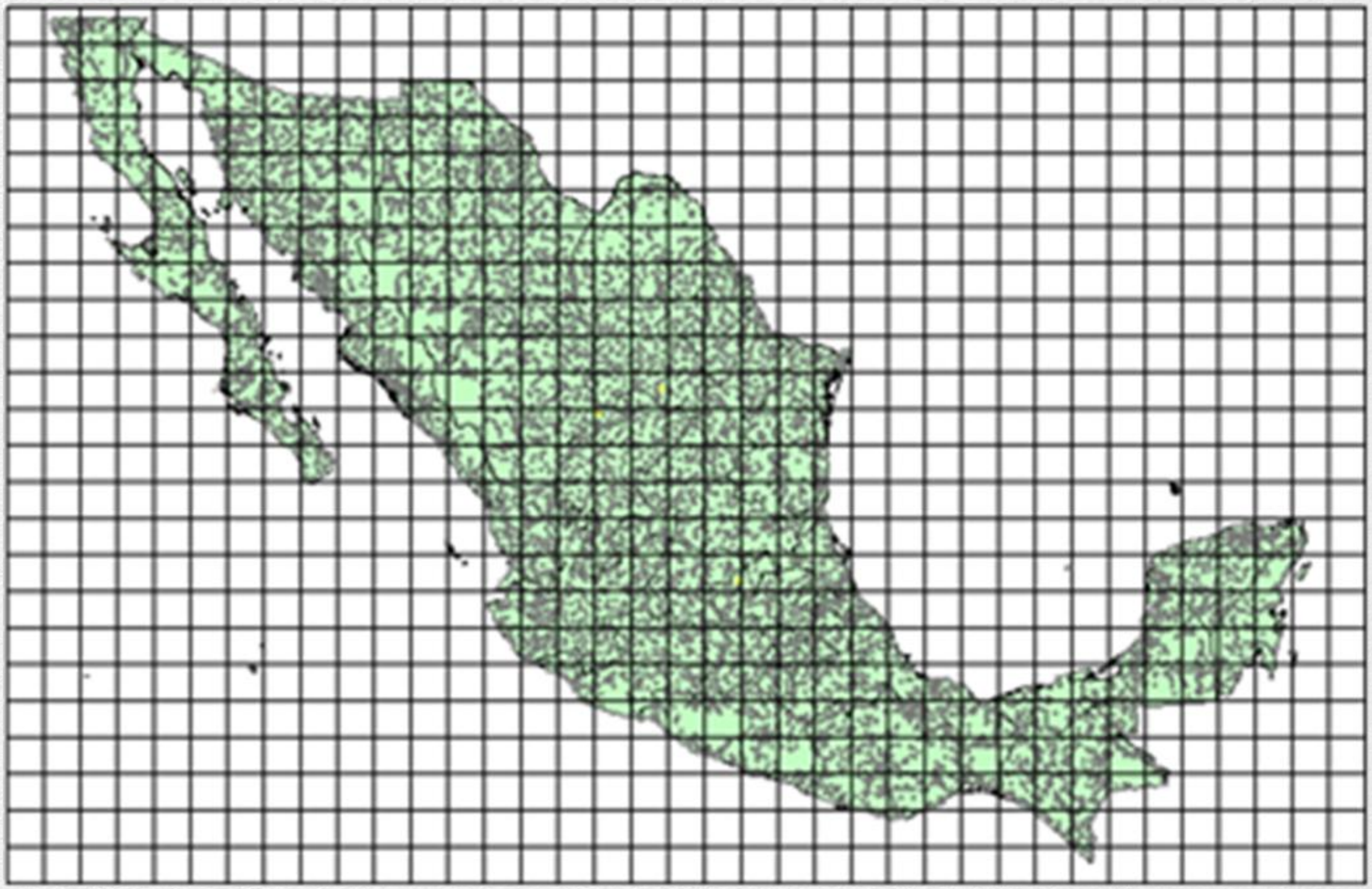
Minimizar $E\left[\{V^*(x) - V(x)\}^2\right]$

Suujeto a : $\sum \lambda_i = 1$

GEOSTADISTICA Y «BULL EYES»

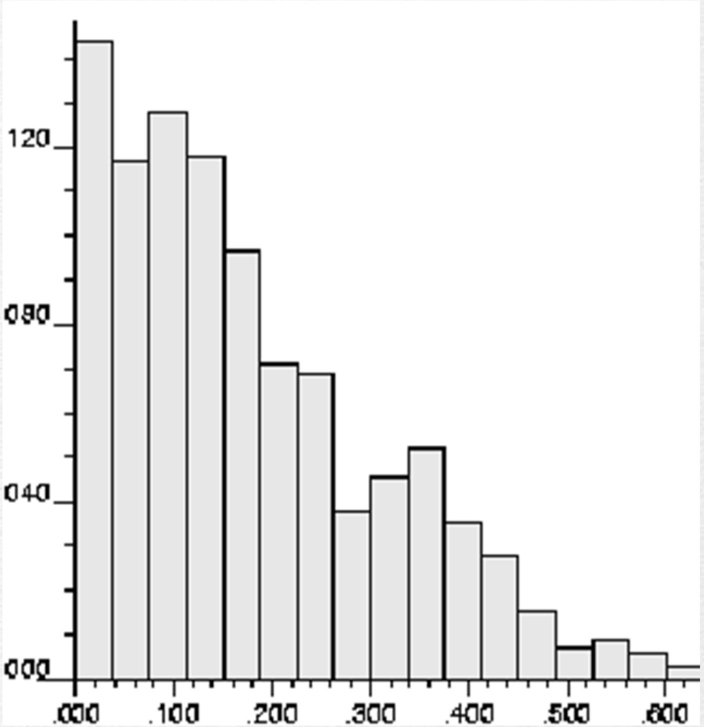


RASTERS



ESTIMACIONES DE CARBONO

Estimaciones de C en cada celda de malla nacional



ESTIMACIONES LOCALES vs GLOBALES

- Las estimaciones globales usan los datos como están para hacer estimaciones (suponen que los rangos de influencia de los datos son iguales para valores grandes o pequeños)

- Estimaciones globales => Valor anómalo alto => Sobre estimaciones (e.g. pepitas de oro) – no cumple guías IPCC

- Estimaciones locales consideran que los rangos de influencia son dependientes del valor de la variable:
  - Valores altos con rango de influencia espacial corto
  - Valores bajos con rango de influencia espacial largo

- Problema: se requieren muchos datos para calibrar

GEOESTADISTICA INDICADORA => NO LINEALIDAD Y NO GAUSIANIDAD

$$I_V = \begin{Bmatrix} 1,\ \text{if } V(x) \le v_i \\ 0, \text{sif } V(x) > v_i \end{Bmatrix}$$

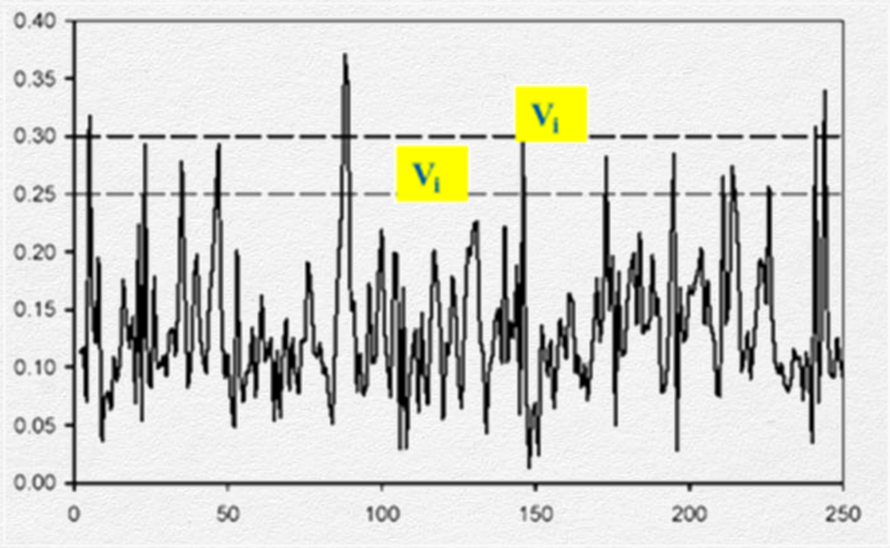
Función de estructura :

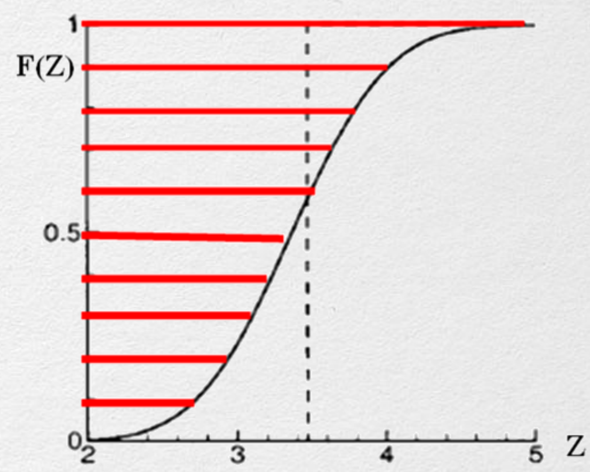
γ_I y C_I

Estimación: $I^* = \sum \lambda_i I_i$

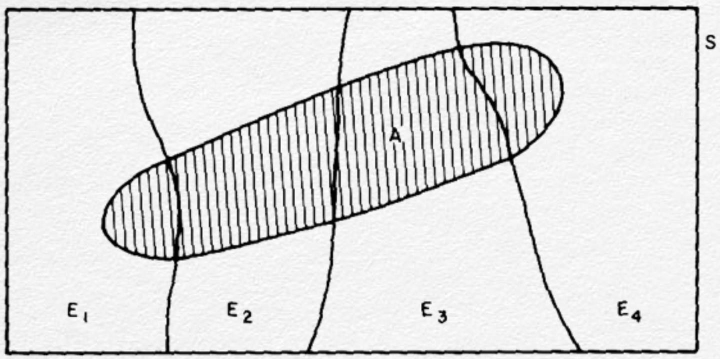
Minimizar $E\left[\left\{I^*(x) - I(x)\right\}^2\right]$

Sujeto a : $\sum \lambda_i = 1$





ENFOQUE BAYESIANO



$$P(A|E_i) = \frac{P(E_i \mid A)P(A)}{P(E_i)}$$

$$f(x \mid y) = \frac{f(y \mid x) f_x(x)}{f_y(y)}$$

$$f(z \mid e) = \frac{f(e \mid z) f_z(z)}{f_E(e)}$$

$f(z|e) =$ distribución a posterio

$f(e|z) =$ función de verosimilitud

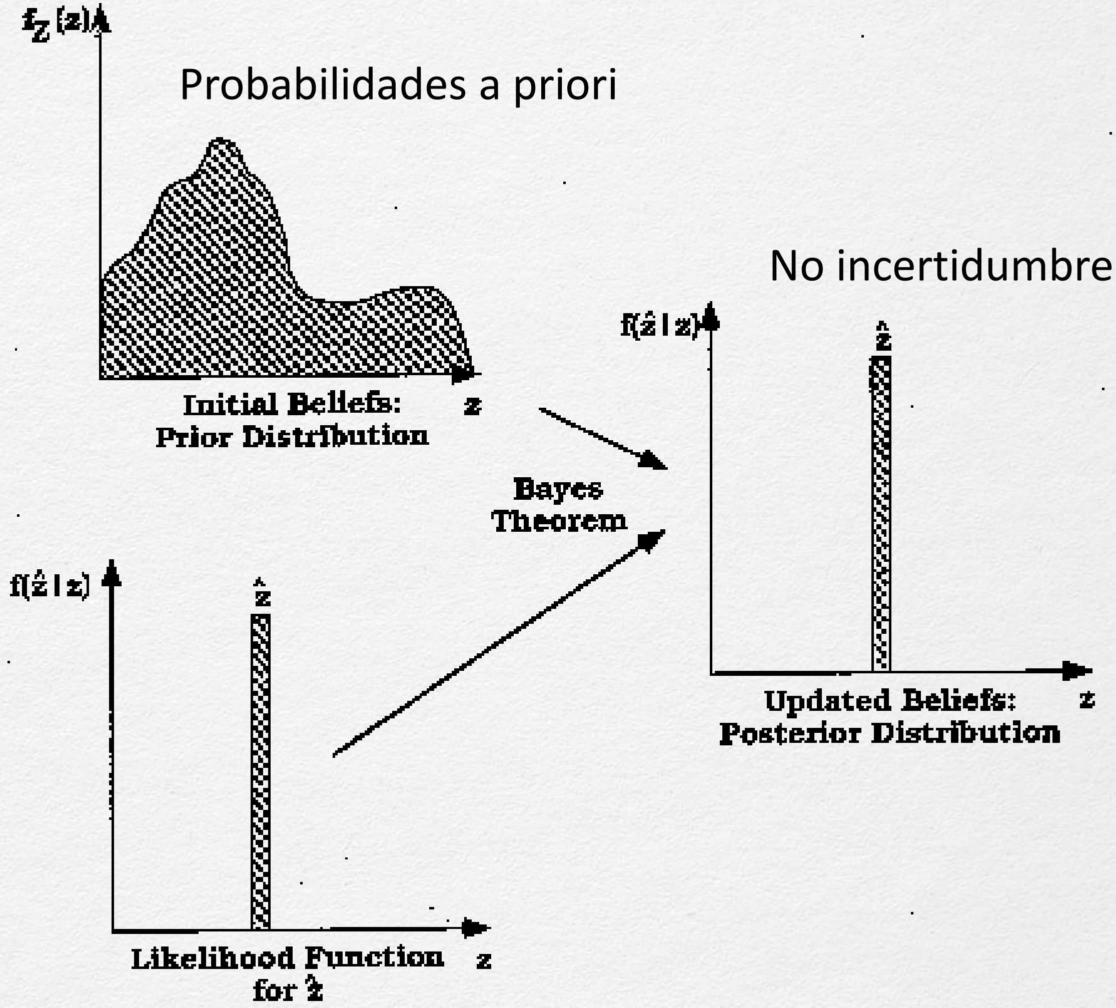
$=$ calificación de la calidad de la informa

del experimento E

CASO DE MEDICIONES (DATOS DUROS)

Probabilidades a priori

No incertidumbre)



DATOS BLANDOS

- Datos blandos (estimaciones indirectas o semi-cuantitativas):
  - Entre X y Y
  - Mayor que X
  - Menor que Y
  - Clase o rango (e.g textura del suelo)
  - Distribución de probabilidad
- Esquema operacional: Codificación de indicadores =>

$$I_V = \begin{cases} 1, \text{ if } \mathrm{V(x)} \le \mathrm{v_i} \\ 0, \text{ sif } \mathrm{V(x)} > \mathrm{v_i} \end{cases}$$

INDICADORES EN FUNCION DE «CUTOFFS» O UMBRALES

$$I_V = \begin{cases} 1, \text{ if } V(x) \le v_i \\ 0, \text{sif } V(x) > v_i \end{cases}$$

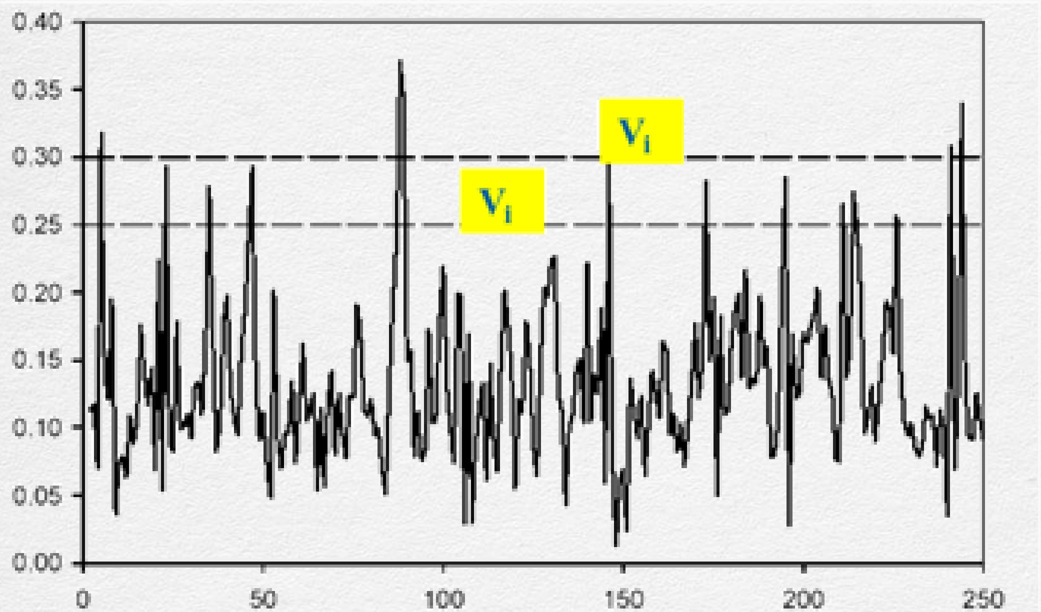
Función de estructura :

γ_I y C_I

Estimación: $I^* = \sum \lambda_i I_i$

Minimizar $E\left[\{I^*(x) - I(x)\}^2\right]$

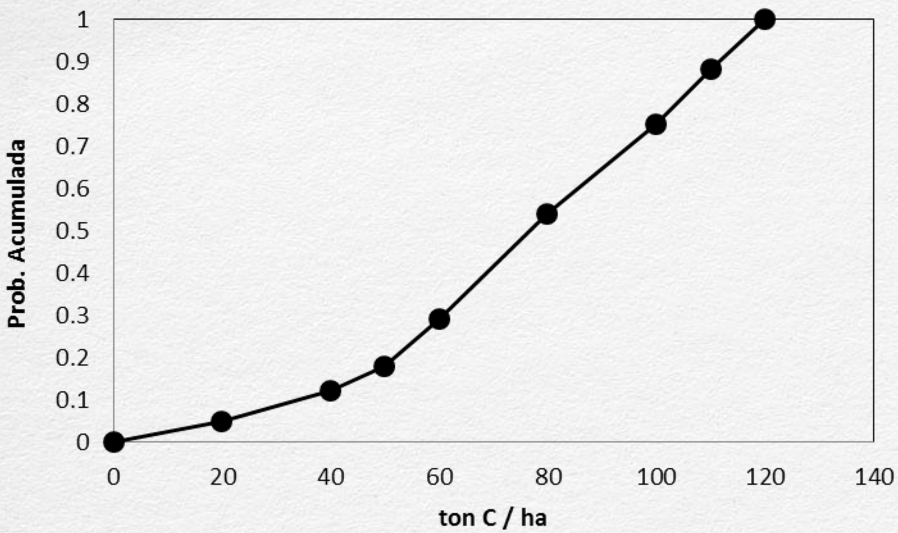
Sujeto a : $\sum \lambda_i = 1$



Problemas =>

+ Número de muestras
+ Escala

EJEMPLO



Límite	Prob. Acum.
0	0
20	0.05
40	0.12
50	0.18
60	0.29
80	0.54
100	0.75
110	0.88
120	1

PARA UNA REGIÓN DADA O TIPO DE USO DEL SUELO/VEGETACIÓN:

+ Definir el numero de limites Nc, con umbrales z_c (Nc+1 clases)
+ Los limites están en función del rango de la variable y conocimiento
+ En el ejemplo se supone que 120 es el limite superior (no puede haber mayores)

TIPOS DE DATOS

TYPE OF DATA		FORMAT	UNCERTAINTY MEASURE
HARD DATA		single value $z(x)$	no uncertainty
SOFT DATA	Type-A	imprecise single value $z^{\wedge}(x)$	quality index, probability level
	Type-B	$\lvert z_{min}(x), z_{max}(x)\rvert$	interval width
	Type-C	probability distribution	probability distribution

CODIFICACIÓN DATOS DUROS (EJEMPLO $z(x_i) = 64.5$)

0 para toda $z < 64.5$
1 para toda $z \geq 64.5$

Limite z_c	Prob. Acum.
0	0
20	0
40	0
50	0
60	0
80	1
100	1
110	1
120	1

CODIFICACIÓN DATOS BLANDOS TIPO A (EJEMPLO $z(x_i) = 46.7$)

0 para toda $z < 46.7$
1 para toda $z \geq 46.7$

Limite z_c	Indicador
0	0
20	0
40	0
50	1
60	1
80	1
100	1
110	1
120	1

CODIFICACIÓN DATOS BLANDOS TIPO B (EJEMPLO $67.3 \leq z(x_i) \leq 108.4$)

0 para toda $z < 67.3$

1 para toda $z \geq 108.4$

Indefinido en el intervalo

Limite z_c	Indicador
0	0
20	0
40	0
50	0
60	0
80	?
100	?
110	1
120	1

CODIFICACIÓN DATOS BLANDOS TIPO C (DISTRIBUCIÓN DE PROBABILIDAD)

Limite z_c	Indicador
0	0
20	0.05
40	0.12
50	0.18
60	0.29
80	0.54
100	0.75
110	0.88
120	1

PRUEBAS DE HIPOTESIS

	H_0 is TRUE	H_0 is FALSE
Accept H_0	Correct Decision (p_1)	Type II Error (p_2)
Reject H_0	Type I Error ($1-p_1$)	Correct Decision ($1-p_2$)

$$p_1(x,z_c) = P\left[\hat{Z}(x) \le z_c \mid Z(x) \le z_c\right] = P\left[\hat{I}(x,z_c) = 1 \mid I(x,z_c) = 1\right]$$

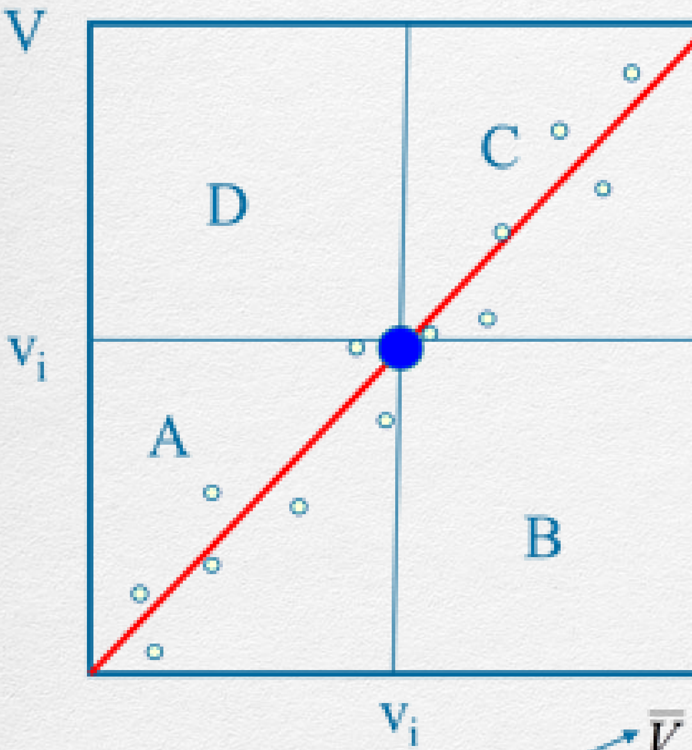
$$p_2(x,z_c) = P\left[\hat{Z}(x) \le z_c \mid Z(x) > z_c\right] = P\left[\hat{I}(x,z_c) = 1 \mid I(x,z_c) = 0\right]$$

$$L_{cc}\left(\hat{i},i\right) = P\left[\hat{i}(x,z_c) \mid i(x,z_c)\right]$$

INCERTIDUMBRES

carbono

Geoestadística Indicadora Bayesiana



Variable indirecta (satelital u otra)

$$I_V = \begin{cases} 1, \text{ if } \mathrm{V(x)} \leq \mathrm{v_i} \\ 0, \text{ if } \mathrm{V(x)} > \mathrm{v_i} \end{cases}$$

$$P\{\overline{V}(x) \leq \mathrm{v}_i \mid V(x) \leq \mathrm{v}_i\} = p_1$$

$$P\{\overline{V}(x) \leq \mathrm{v}_i \mid V(x) > \mathrm{v}_i\} = p_2$$

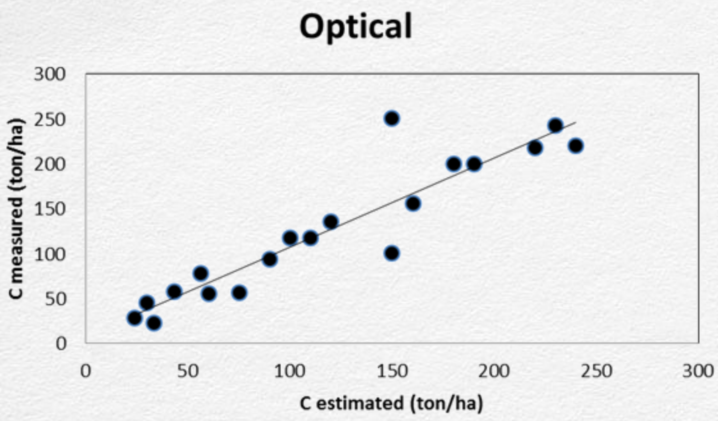
Calidad de la información=>

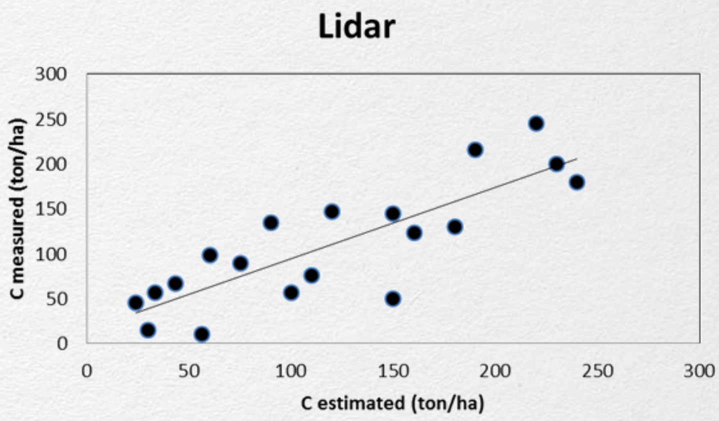
$$p_1 = \frac{A}{A+D}$$

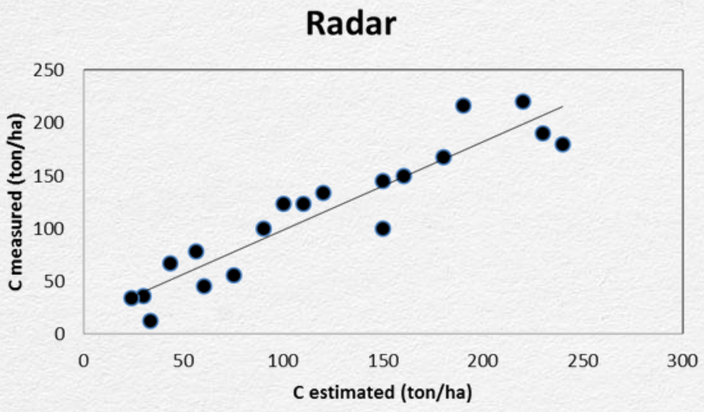
$$p_2 = \frac{B}{B+C}$$

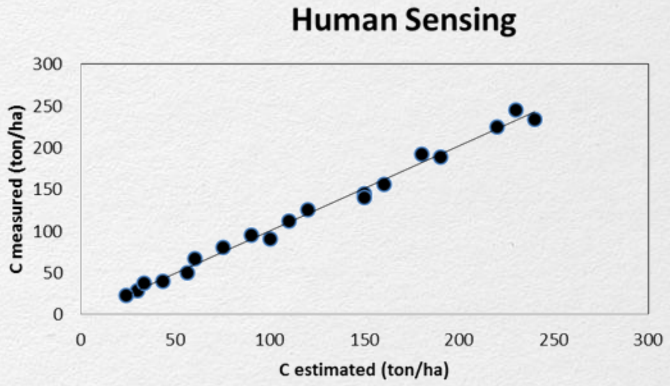
Con $\mathrm{p_1}$ y $\mathrm{p_2}$ => Función de covarianza indicadora

CALIBRACIÓN p_1 Y p_2









FUSIÓN DE INFORMACIÓN

$$C_I(h,v_i) = \frac{C_{I\bar{I}}(h,v_i)}{p_1(v_i) - p_2(v_2)}$$

$$C_I^*(h,v_i) = w\left[C_I^*(h,v_i)\right]_{Nh} + (1-w)\left[\frac{C_{I\bar{I}}^*(h,v_i)}{p_1^*(v_i) - p_2^*(v_i)}\right]_{Nh,Ns}$$

$$I^* = \sum \lambda_i I_i + \sum \beta_j \bar{I}_j$$

APROXIMACION DE MARKOV => LOS DATOS BLANDOS CO-COLOCADOS CON DATOS DUROS NO SE USAN (PREVALECEN LOS DUROS)

SOFTWARE

- UNCERTAIN
- GS-LIB
- Muchos otros
- Libros de Geoestadística y tutoriales => muchos
- «La practica hace al maestro», claro si tiene conocimiento

OTROS METODOS DE INTERPOLACION – FUSION DE INFORMACION

- Se requieren medidas de error (e.g. varianzas de estimación)
  - Métodos del inverso de la distancia a una potencia (sin medidas de varianza). Posible extensión: las potencias en función de los valores de la variable (estimaciones locales)
- Pesos para fusión de información => **($p_1 – p_2$)**, sujeto a $p_1 > p_2$ (de otro modo no se aporta información). Datos duros: $p_1 = 1$, $p_2 = 0$

¿PREGUNTAS?