

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS FÍSICAS

DEPARTAMENTO DE ARQUITECTURA DE COMPUTADORES Y ATOMÁTICA



TRABAJO DE FIN DE GRADO

Código de TFG: ACA07

Control de sistemas multiagente y sus aplicaciones

Control of multi-agent systems and their application

Supervisor/es: Lía García Pérez, Juan Jiménez

Juan Bosco Román Camacho

Grado en Física

Curso académico 2023-2024

Convocatoria extraordinaria

Calificación: 9,9

Búsqueda de máximos globales mediante Kriging con un sistema multiagente.

Global extremum seeking by Kriging with a multi-agent system

Resumen:

Encontrar el máximo de cualquier distribución en el mundo físico es un problema de gran interés hoy en día. Desde encontrar fugas de petróleo en el océano a fuentes de radiación en zonas contaminadas. Es muy valioso entender estos comportamientos desde el punto de vista de los sistemas multiagente. Para ello es necesario medir un número suficiente de puntos, eligiéndolos estratégicamente de la mejor manera posible para lograr la mayor eficiencia. Medir con suficiente precisión magnitudes en un terreno impredecible se puede resolver mediante el uso de un enjambre de agentes simples y con una capacidad de movimiento y comunicación muy limitada, contando con un alto número de ellos como ventaja. Además, pueden ser muy pequeños y capaces de comunicarse localmente en entornos donde es difícil la comunicación a larga distancia, consiguiendo a pesar de ello un comportamiento global organizado y eficiente.

Palabras Clave: Sistema multiagente. Kriging, Problema de búsqueda de foco, sistema autónomo, inferencia gaussiana

Abstract:

Finding the maximum of any distribution in the physical world is a problem of great interest nowadays. From detecting oil leaks in the ocean to identifying sources of radiation in contaminated areas, understanding these behaviors from a multi-agent systems perspective is highly valuable. Achieving this requires measuring a sufficient number of points, strategically selected to maximize efficiency. Precisely measuring magnitudes in unpredictable terrain can be addressed by employing a swarm of agents, simple in design with limited movement and communication capabilities, yet advantageous in their high numbers. Furthermore, these agents can be small and capable of local communication in environments where long-distance communication is challenging, still achieving organized and efficient global behavior.

Keywords: multiagent system, kriging, autonomous system, source search problem, gaussian inference

Índice

1. Introducción	4
1.1. Objetivos	4
1.2. Metodología	4
2. Preliminares	4
2.1. Requisitos de consistencia: Marginalización y condicionamiento.	6
3. Formulación del problema de búsqueda	7
4. El cerebro de los agentes: Estimador kriging	8
4.1. Campo aleatorio gaussiano: Intuición	8
4.2. Kriging	9
4.3. Procesos gaussianos para kriging: deducción de las ecuaciones	10
4.4. Distribución posterior	12
4.5. Kriging simple y universal	14
5. Optimización de los hiperparámetros de la función de covarianza	14
5.1. Hipótesis de estacionariedad de primer y segundo orden	15
6. Método de selección de puntos de muestreo	16
6.1. Método 1	16
6.2. Método 2	17
7. Ley de control	17
8. Implementación del modelo y algoritmo	18
9. Resultados de la simulación	18
9.1. Resultados en una dimensión con $N = 2$	18
9.2. Tasa de éxito y escalabilidad del algoritmo en el caso 1D	20
9.3. Resultados en dos dimensiones con $N = 3$	20
10. Conclusiones y futuras extensiones del trabajo	21

1. Introducción

La búsqueda de máximos —o focos de emisión— de campos escalares es un problema de gran interés con aplicaciones en gran número de campos. Encontrar el origen de fugas midiendo la concentración de químicos en la superficie del agua, focos de contaminación atmosférica, o incluso la búsqueda de recursos en el subsuelo son misiones de gran relevancia hoy en día. Entre las soluciones más conocidas se encuentran la escalada de gradiente, que utiliza el gradiente del campo para ascender hacia el máximo, y la búsqueda directa, que explora el espacio de búsqueda sistemáticamente para encontrar la mejor dirección de ascenso. El primero puede ser eficiente pero tiende a converger en máximos locales, mientras que el segundo es preciso pero ineficiente al explorarse todo el campo. Ambos métodos se describen en [9]. Se propone un método diferente, que puede evitar estos problemas, mediante un sistema multiagente que realiza una interpolación espacial del campo, basado en [5]. Se impone un radio máximo de comunicación entre agentes, de manera que el grafo de comunicación no sea totalmente conexo, evitando la sobrecarga de la red cuando escala el número de agentes.

Tras el establecimiento de los objetivos del trabajo en 1.1 y la metodología usada en 1.2, se exponen definiciones básicas sobre estadística y variable aleatoria en 2. En 3 se formula el problema de búsqueda que se resolverá mediante un sistema multiagente, capaz de realizar un modelo kriging del campo. El funcionamiento de este modelo, que cada agente es capaz de realizar de manera individual, se explica con detalle en 4 y 5. Los agentes toman medidas en localizaciones específicas que se deciden mediante dos métodos diferentes expuestos en 6. Además, cada agente debe seguir su propia dinámica, determinada por una ley de control, que se expone en 7. Por último, en 8 se trata el algoritmo que se usará en las simulaciones, cuyos resultados se discuten en 9.

1.1. Objetivos

El objetivo de este trabajo es encontrar el máximo global de un campo desconocido utilizando un Sistema Multiagente. Cada agente puede moverse, según una dinámica individual especificada, para realizar mediciones con las que calcula un modelo Kriging. El kriging es un método de interpolación que, partiendo de una suposición previa sobre la covarianza del campo estimado, proporciona la mejor predicción lineal no sesgada¹ junto con su incertidumbre de estimación. Con esta información, cada agente explora el campo sólo en áreas con incertidumbre alta, mejorando la estimación actual del máximo global y ganando criterio para proceder con la búsqueda del máximo global a partir del modelo interpolado.

1.2. Metodología

Se utilizará *python* para las simulaciones. El modelo de kriging se resuelve mediante la librería orientada a geoestadística *Pykrige* [1]. Los campos aleatorios gaussianos con la librería *gstools* [8]. Además, hemos elaborado íntegramente todas las figuras presentadas en este trabajo con *python*.

2. Preliminares

Definición 1 (Variable aleatoria). *Formalmente, una variable aleatoria X es una función que va del espacio muestral Ω (el conjunto de todos los posibles resultados de un experimento aleatorio) al*

¹La mejor predicción lineal no sesgada (BLUP por sus siglas en inglés) es un término que tiene relevancia en la derivación usual de kriging, sin embargo, en este trabajo se usará una derivación diferente que no necesita esta definición explícitamente. Para ver más sobre BLUPs y conceptos más avanzados sobre kriging se recomienda [13] y [11]

conjunto de los números reales \mathbb{R} , permitiendo el análisis probabilístico y estadístico. Ver figura 1

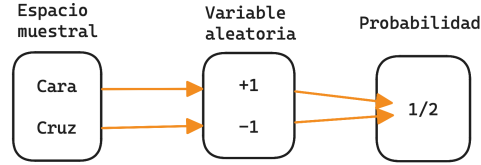


Fig. 1: Ejemplo esquemático de variable aleatoria para el proceso aleatorio de tirar una moneda

Definición 2 (proceso estocástico). *Un proceso estocástico es una colección o familia de variables aleatorias X_i , ordenadas según un subíndice i . Normalmente i se identifica con el tiempo aunque en este trabajo denotará una posición espacial.*

Definición 3 (valor esperado). *Sea X una variable aleatoria discreta con rango $R_X = \{x_1, x_2, x_3, \dots, x_N\}$ (finito o numerablemente infinito). El valor esperado de X , denotado por $E[X]$, se define como*

$$E[X] = \sum_{x_i \in R_X} x_i P(X = x_i),$$

donde $P(X = x_i)$ es la probabilidad de que X tome el valor x_i .

Definición 4 (Media). *De la definición 3, si $P(X = x_i) = \frac{1}{N}$ (constante) entonces $\mu \equiv E[X]$ es la media de la variable aleatoria X de manera que se calcula como:*

$$\mu = \frac{1}{N} \sum_{x_i \in R_X} x_i \quad (1)$$

Definición 5 (covarianza entre variables aleatorias). *La covarianza entre dos variables aleatorias X e Y se define como*

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] \quad (2)$$

Si X e Y toman los valores discretos x_i y y_i para $i = 1, 2, \dots, n$ con probabilidad $P[X = x_i] = 1/n$ y $P[Y = y_i] = 1/n$ respectivamente entonces

$$\text{Cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n (x_i - E[X])(y_i - E[Y]) \quad (3)$$

Definición 6 (Matriz de covarianza). *Si \mathbf{X} es un vector aleatorio de dimensión n , es decir, $\mathbf{X} = (X_1, \dots, X_n)^T$ donde X_i para $i = 1, 2, \dots, n$ son variables aleatorias, se puede construir la matriz de covarianza denotada por K cuyos componentes son:*

$$K_{ij} = \text{Cov}(X_i, X_j)$$

y en particular, cuando $i = j$

$$K_{ii} = \text{Cov}(X_i, X_i) = \text{Var}(X_i) \equiv \sigma_{X_i}^2 \quad (4)$$

Donde Var es la varianza, que es un caso particular de la covarianza cuando dos variables aleatorias son idénticas y σ_{X_i} es la desviación típica que, a diferencia de la varianza, tiene las mismas unidades que X_i . Entonces se tiene:

$$K = \begin{pmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) & \cdots & \text{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_n, X_1) & \text{Cov}(X_n, X_2) & \cdots & \text{Var}(X_n) \end{pmatrix} \quad (5)$$

Definición 7 (Distribución gaussiana unidimensional). *La distribución de probabilidad de una variable aleatoria X se denomina gaussiana o normal si tiene una densidad de probabilidad*

$$p(x; a, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-a)^2/2\sigma^2}$$

Una distribución normal es una buena aproximación en la naturaleza siempre que la variable aleatoria relevante sea la suma de un gran número de variables aleatorias independientes, siendo la más grande de estas pequeña en comparación con la suma total (véase el teorema del límite central).

Definición 8 (Distribución gaussiana multidimensional²). *Se dice que una variable aleatoria vectorial $X = [X_1 \cdots X_n]^T$ tiene una distribución normal (o Gaussiana) multivariada con media $\mu \in \mathbb{R}^d = (\mu_1, \dots, \mu_d)^T$ tal que $E[X_i] = \mu_i$ y matriz de covarianza $K \in \mathbb{R}^{d \times d}$ tal que $K_{ij} = \text{Cov}[X_i, X_j]$, si su función de densidad de probabilidad está dada por*

$$p(x; \mu, K) = \frac{1}{(2\pi)^{n/2} |K|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu)^T K^{-1}(x - \mu)\right),$$

y se representa como $X \sim \mathcal{N}(\mu, K)$. Estos dos parámetros determinan de manera única la distribución Gaussiana. $|K|$ es el determinante de K . K es simétrica y definida positiva.

2.1. Requisitos de consistencia: Marginalización y condicionamiento.

Supóngase un conjunto de variables aleatorias gaussianas. Se entiende por requisito de consistencia la condición de que cualquier subconjunto finito de variables aleatorias seleccionadas de dicho conjunto siga teniendo una distribución gaussiana. En particular considérese una distribución de probabilidad normal $P(X, Y)$ con X, Y vectores de variables aleatorias representando subconjuntos de la variable original.

$$P_{X,Y} = \begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}(\mu, K) = \mathcal{N}\left(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} K_{XX} & K_{XY} \\ K_{YX} & K_{YY} \end{bmatrix}\right) \quad (6)$$

Donde μ es la media y K la matriz de covarianza. Mediante la marginalización se obtiene información parcial de la distribución multivariable:

$$X \sim \mathcal{N}(\mu_X, K_{XX}) \quad Y \sim \mathcal{N}(\mu_Y, K_{YY}) \quad (7)$$

En otras palabras, la distribución del subconjunto no se ve afectada por el estudio del conjunto. Marginalizar es proyectar la distribución conjunta sobre una de las dimensiones. En decir, cada distribución X e Y solo depende de su correspondiente entrada en μ y Σ . La operación se realiza como:

$$p_X(x) = \int_y p_{X,Y}(x, y) dy = \int_y p_{X|Y}(x | y) p_Y(y) dy \quad (8)$$

Otra operación importante, que será la de mayor utilidad en este trabajo, es el condicionamiento. Se utiliza para determinar la probabilidad de una variable en función de otra variable. Al igual que la marginalización, esta operación también es cerrada y produce una distribución gaussiana distinta a la original. El condicionamiento se define por:

$$\begin{aligned} X | Y &\sim \mathcal{N}(\mu_X + K_{XY} K_{YY}^{-1} (Y - \mu_Y), K_{XX} - K_{XY} K_{YY}^{-1} K_{YX}) \\ Y | X &\sim \mathcal{N}(\mu_Y + K_{YX} K_{XX}^{-1} (X - \mu_X), K_{YY} - K_{YX} K_{XX}^{-1} K_{XY}) \end{aligned} \quad (9)$$

²también denominada multivariada o conjunta en la bibliografía

Nótese que tanto la media como la matriz de covarianza varían en la distribución condicionada, efecto que se observa visualmente en la figura 2. Las distribuciones gaussianas tienen la propiedad algebraica de ser cerradas bajo el condicionamiento y la marginalización, es decir, si aplicamos estas operaciones a una gaussiana, las distribuciones resultantes también son gaussianas (ver figura 2), de manera que se pueden obtener expresiones analíticas.

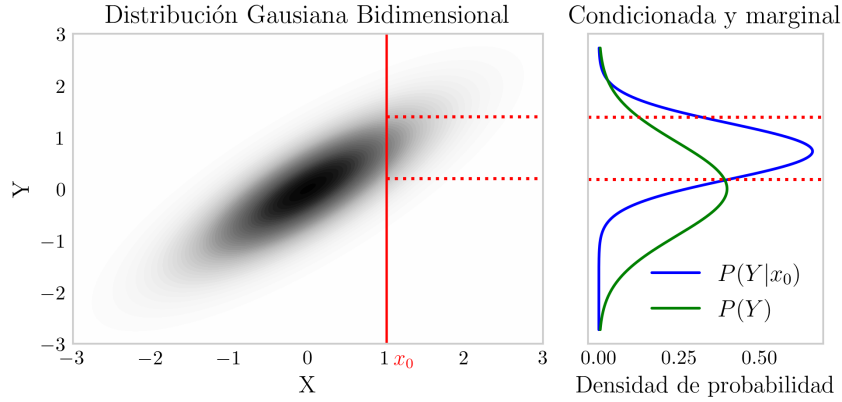


Fig. 2: Visualización de condicionamiento y marginalización en una distribución gaussiana bidimensional. El condicionamiento se puede entender como tomar una *rodaja* de la distribución conjunta. Al fijar una de las variables se impone una restricción sobre las otras.

3. Formulación del problema de búsqueda

Considérese un campo escalar desconocido, continuo y estacionario f definido sobre el espacio compacto $D \subset \mathbb{R}^m$ cuyo máximo global está asociado a un único argumento. El objetivo es encontrar dicho máximo usando una flota de N agentes idénticos con coordenadas espaciales \mathbf{x} contenidas en D . El agente queda indexado por las variables $i \in 1 \dots N$ y el instante de tiempo por k . Los agentes tienen la capacidad de:

1. Medir el campo f para cada tiempo t_k siendo la medida³ $f(\mathbf{x}(t_k))$.
2. No medir, donde $\mathcal{M}(t_k)$ es el conjunto de agentes que miden en t_k
3. Moverse siguiendo la dinámica: $M\ddot{\mathbf{x}}_i + C(\mathbf{x}_i, \dot{\mathbf{x}}_i)\dot{\mathbf{x}}_i = \mathbf{u}_i$ siendo $C(\mathbf{x}_i, \dot{\mathbf{x}}_i)$, M constantes y \mathbf{u}_i la función de control, que incluye evasión de colisiones entre agentes.
4. Comunicarse con su subconjunto de vecinos $\mathcal{N}_i(t_k) = \{j \mid \|\mathbf{x}_i - \mathbf{x}_j\| \leq r\}$ (agentes j a un radio r , que es la distancia máxima de comunicación).
5. Guardar datos $(\mathbf{x}, f(\mathbf{x}))$ medidos por el agente o recibidos de vecinos que han medido en todos los tiempos hasta el actual en $S_i(t_k) = \bigcup_{l=0}^k \{f(\mathbf{x}(t_l)), \mathbf{x}_j(t_l) \mid j \in \mathcal{N}_i(t_l) \cap \mathcal{M}(t_l)\}$
6. Construir un modelo de kriging —modelo de interpolación espacial \hat{f} — en cada instante de tiempo necesario a partir de los datos en memoria $S_i(t_k)$ y calcular su máximo.

³Por simplicidad en la notación, se omitirá la dependencia temporal del estado de los agentes, siempre que no lleve a confusión.

4. El cerebro de los agentes: Estimador kriging

En esta sección se lleva a cabo la explicación y derivación del modelo de interpolación/regresión⁴ que utilizan los agentes para modelar el campo. Existen dos perspectivas diferentes que resultan en las mismas ecuaciones predictivas, y que, en el fondo, se basan en los mismos principios⁵, sin embargo las derivaciones resultan muy diferentes. Estas dos aproximaciones son las de kriging, usada en el ámbito de la geoestadística, y la de procesos gaussianos (GP), usada en aprendizaje automático. La bibliografía es muy confusa, normalmente confundiendo los nombres. La referencia [14] contiene una aclaradora comparativa entre GP y Kriging. En este trabajo se considera que se puede obtener lo mejor de los dos mundos y por ello se ha hecho una derivación de las ecuaciones del kriging siguiendo la perspectiva de procesos gaussianos.⁶

4.1. Campo aleatorio gaussiano: Intuición

Un campo aleatorio gaussiano —sinónimo de proceso gaussiano— definido en una cuadrícula de n puntos se construye a partir de un vector— n -dimensional— extraído de una distribución gaussiana n -dimensional (Def. 8). Esta gaussiana tiene una media y matriz de covarianza que la caracterizan completamente. La existencia de la matriz de covarianza expresa que, de alguna manera, los elementos del vector extraído de la gaussiana están relacionados entre sí. Cada componente del vector tiene su propio proceso aleatorio pero, como proviene de la misma gaussiana multidimensional, todos tienen la misma forma.

En la naturaleza muchos procesos espaciales y espaciotemporales se pueden modelar como campos aleatorios gaussianos. Un ejemplo ilustrador es la superficie del agua. Si se mide la altura de cada punto a lo largo del tiempo y se realiza un histograma, se observa una distribución gaussiana con una cierta media. Si se mide la altura de un punto muy cercano en los mismos tiempos se observará que también sigue una distribución gaussiana, y que, además, existe cierta covarianza (ecuación 3) entre ellos, es decir, cuando el primer punto se encuentra alto, el segundo también —ambos están en la misma ola—. Existe una covarianza espaciotemporal entre puntos cercanos. Se puede entender el concepto de longitud de correlación como el tamaño de la ola (Véase la figura 3), que en la sección 4.2 tendrá un papel importante como el parámetro ℓ .

Aunque la altura de cada punto de la superficie siga su propia distribución gaussiana, existe una covarianza entre todas las distribuciones y, una vez construida la gaussiana conjunta, esto se expresa mediante la matriz de covarianza.

El ejemplo anterior es un caso de campo con variación espaciotemporal, sin embargo, imagine-mos que la superficie del agua se ‘congela’ en un determinado instante del tiempo, entonces se puede aplicar la misma idea para tratar el campo únicamente con variación espacial, con una diferencia importante: solo se puede acceder a una medida en cada localización. Esta restricción presenta el problema de que no se puede hacer estadística en cada punto, de manera que no se puede estimar ni la media ni la varianza en cada punto y por tanto, no se puede construir su distribución.

La clave está en que, aunque no se pueda conocer exactamente esta matriz de covarianza, se puede construir artificialmente, optimizándola con las medidas experimentales. Esto lo estudia la

⁴Si se asumen medidas sin ruido el Kriging es un interpolador exacto, es decir, la predicción pasa por las medidas.

⁵Los GP se basan en razonamientos bayesianos, haciendo una suposición previa sobre la forma del campo y actualizando la incertidumbre con las medidas. Los enfoques no bayesianos o frecuentistas determinan el modelo predictivo según un criterio de optimización a posteriori, como minimizar la suma de los errores al cuadrado

⁶kriging es un tipo específico de regresión de GP, pero no todos los métodos de regresión de GP son kriging

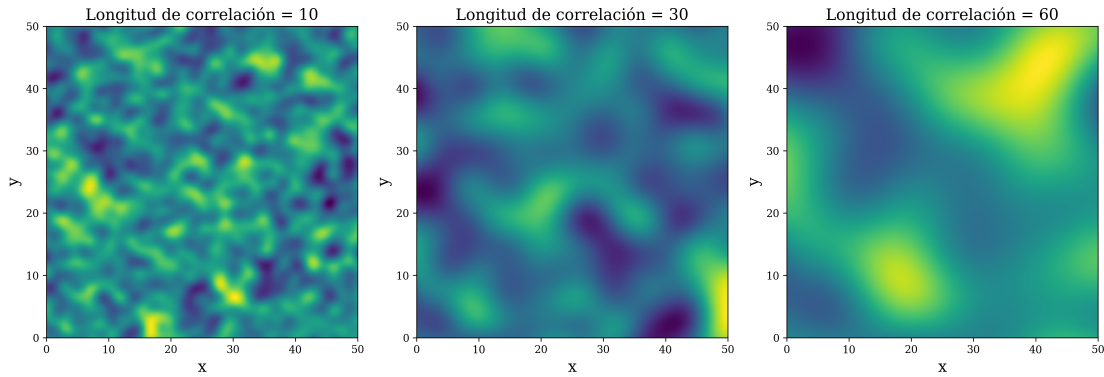


Fig. 3: Ejemplos de tres realizaciones de tres campos aleatorios gaussianos diferentes, con distintas longitudes de correlación. Cada campo es una realización de una gaussiana de dimensión 100×100 , graficado por tanto en una cuadrícula 100×100 .

geoestadística, que resuelve este último problema mediante algunas suposiciones que se tratan en detalle en la sección 5.

4.2. Kriging

El kriging (Matheron 1963 [7], aunque recibe su nombre del ingeniero de minas sudafricano Danie G. Krige.) utiliza las relaciones entre las ubicaciones de los datos y sus valores para proporcionar un modelo no paramétrico⁷ con la mejor predicción lineal no sesgada, asumiendo principalmente que los datos se pueden modelar como un campo aleatorio gaussiano (ver figura 3) y que por tanto tienen las siguientes propiedades:

1. Autocorrelación espacial. El kriging funciona asumiendo que la variación espacial se puede descomponer en tres componentes principales, como se puede ver en la figura 4:
 - a) Variación determinista o tendencia general predominante, que puede deberse a gradientes topográficos, gradientes hidráulicos, variaciones en la intensidad de la fuente, entre otros factores. Nótese que la tendencia marca la media de b).
 - b) Variación aleatoria espacialmente correlacionada: fluctuaciones aleatorias con media 0 correlacionadas espacialmente. Los valores de los datos cercanos entre ellos están más correlacionados que los valores entre datos más distantes. Esto es el campo aleatorio gaussiano.
 - c) Variación espacialmente no correlacionada (ruido): Esta variación representa el error o la variabilidad no explicada por la tendencia o la variación espacialmente correlacionada.
2. Estacionaridad: la distribución conjunta de probabilidad no varía en todo el espacio de estudio, es decir, la media y la varianza de los valores son constantes en todo el campo espacial. Se explicará con más detalle en la sección 5.1 .

Existen varias formas de kriging que permiten relajar esta última suposición:

- a) Kriging Simple: Supone media constante y conocida.
- b) Kriging Ordinario: Supone media constante desconocida (a estimar) y datos sin tendencia.

⁷Un modelo no paramétrico se puede entender como un modelo con un número infinito de parámetros. A diferencia de los modelos de regresión no lineales que son paramétricos el kriging es no paramétrico.

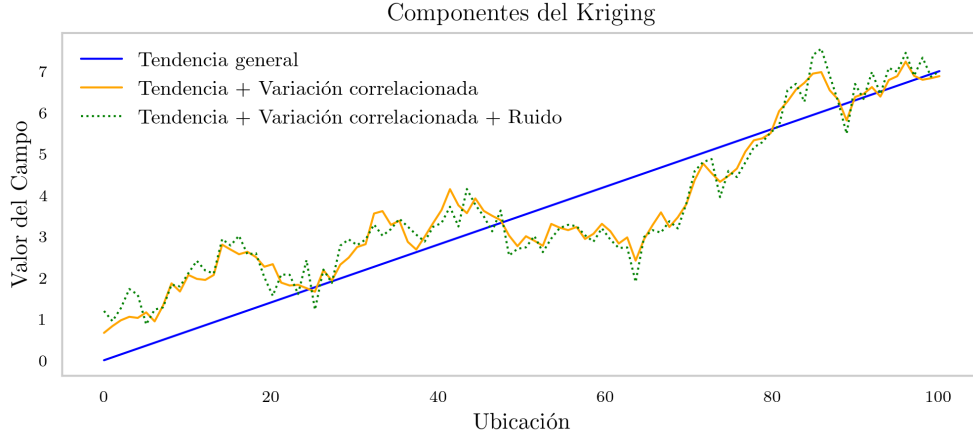


Fig. 4: Visualización de las componentes del kriging. La variación correlacionada ha sido generada como un campo aleatorio gaussiano.

- c) Kriging Universal: relaja la suposición de estacionaridad al permitir que la media (y no la varianza) de los valores difiera de manera determinista siguiendo una tendencia predominante.

4.3. Procesos gaussianos para kriging: deducción de las ecuaciones

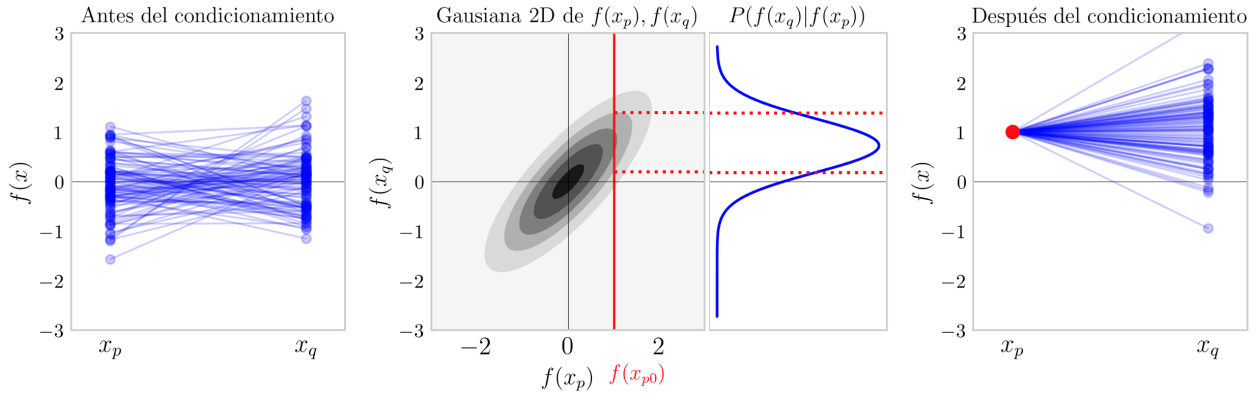


Fig. 5: Funcionamiento esquemático de un proceso gaussiano para dos valores $f(x_p)$ y $f(x_q)$ asociados a las localizaciones x_p y x_q . En medio se muestra la gaussiana conjunta de $f(x_p)$ y $f(x_q)$. En la izquierda se muestran varias realizaciones superpuestas de esta gaussiana conjunta. la línea roja vertical marca el valor medido de $f(x_p)$ al que se condiciona $f(x_q)$. A la derecha el valor de $f(x_p)$ queda fijado y el valor de $f(x_q)$ sigue la distribución marginalizada.

En esta sección, se derivará el kriging simple desde la perspectiva de procesos gaussianos, que se considera más intuitiva. Se seguirá el desarrollo mostrado en [10]. Algunas nociones complementarias han sido tomadas de [6].

Consideremos primeramente el caso con solo dos localizaciones ($n = 2$) \mathbf{x}_p y \mathbf{x}_q para los cuales no se conoce el valor del campo aleatorio gaussiano $f(\mathbf{x}_p)$ y $f(\mathbf{x}_q)$ —la variable aleatoria es el valor de $f(\mathbf{x})$ en la localización \mathbf{x} —. Primeramente se elige la media del campo como 0, aunque más adelante se generalizará a una función arbitraria. Se construye para $f(\mathbf{x}_p)$ y $f(\mathbf{x}_q)$ una gaussiana bidimensional ($n = 2$) con media nula y, si se conociese, la matriz de covarianza K_{pq} . Si ahora

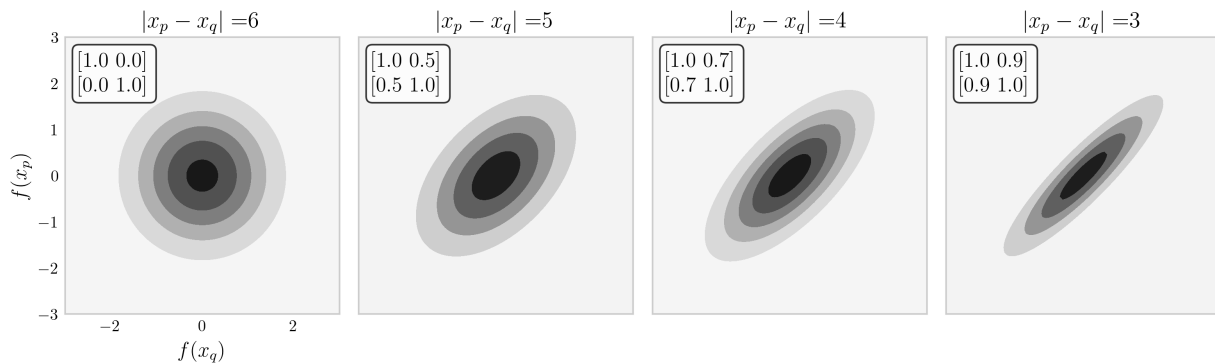


Fig. 6: Gaussiana bidimensional para los posibles valores de $f(x_p)$ y $f(x_q)$ con matriz de covarianza 2×2 . Solo se muestran los ejes en una de las figuras por simplicidad. Cuanto menor es la distancia $|x_p - x_q|$ entre los argumentos x_p y x_q mayor es la covarianza entre $f(x_p)$ y $f(x_q)$ que visualmente se entiende como un escalado de la gaussiana, de manera que los valores extraídos de esta serán más similares entre sí

el valor de \mathbf{x}_p se especificara —a través de una medida⁸—, se podría condicionar la distribución a esta medida, y se podría obtener la distribución predictiva para \mathbf{x}_p , como se muestra en la figura 5.

El principal problema es que como, como se ha visto en la sección 4.1 como solo hay una realización de las variables aleatorias $f(\mathbf{x}_p)$ y $f(\mathbf{x}_q)$ en sus respectivas localizaciones \mathbf{x}_p y \mathbf{x}_q no se puede calcular su covarianza directamente con (3), y por tanto no se puede construir la matriz de covarianza K_{pq} . Sin embargo, la clave está en que se puede construir un modelo de covarianza que refleje y se base en las creencias a priori sobre los datos. En la figura 6 se muestra un posible modelo que refleja la hipótesis de que a medida que aumenta la distancia entre pares de localizaciones x_p y x_q , la covarianza entre sus realizaciones asociadas $f(x_p)$ y $f(x_q)$ disminuya. En el caso general $f(\mathbf{x}_1), \dots, f(\mathbf{x}_n)$ componen una gaussiana n -dimensional.

Aquí está la suposición más importante: La covarianza entre los datos depende únicamente de la distancia entre sus argumentos (localizaciones), no de sus valores. En este trabajo, la función de covarianza modelo tendrá la forma de una gaussiana:

$$\text{cov}(f(\mathbf{x}_p), f(\mathbf{x}_q)) \approx k(\mathbf{x}_p, \mathbf{x}_q) = \tau \exp\left(-\frac{1}{2} (|\mathbf{x}_p - \mathbf{x}_q|/\ell)^2\right) \quad \ell, \tau > 0 \quad (10)$$

ℓ ⁹ y τ son hiperparámetros¹⁰ que marcan la longitud característica y la amplitud del proceso respectivamente. Estos hiperparámetros son constantes en todo el campo espacial por las hipótesis de estacionariedad (ver sección 5.1). Para esta función de covarianza en particular, observamos que la covarianza es casi unitaria entre variables cuyas localizaciones correspondientes están muy cerca, y disminuye a medida que aumenta la distancia. A partir de la función de covarianza modelo se construye la matriz de covarianza modelo (fig 7 Izquierda) tomando pares de puntos \mathbf{x}_p y \mathbf{x}_q . En la figura 6 se observa como a medida que aumenta la distancia entre los puntos \mathbf{x}_p y \mathbf{x}_q su covarianza disminuye, reflejando la suposición anterior.

⁸Una medida es lo que se denomina una realización del campo, es decir, de todos los posibles valores que se podrían haber dado para el campo en una localización, se realiza solo uno, que es el que se mide. Una medida en un punto es una realización de una variable aleatoria asociada a ese punto

⁹se recuerda que ℓ se corresponde con el tamaño de la ola en la analogía de la sección 4.1

¹⁰Un hiperparámetro, en este caso, es un parámetro cuyo valor se establece antes del proceso de regresión y no durante, como sería el caso de los parámetros de un modelo paramétrico

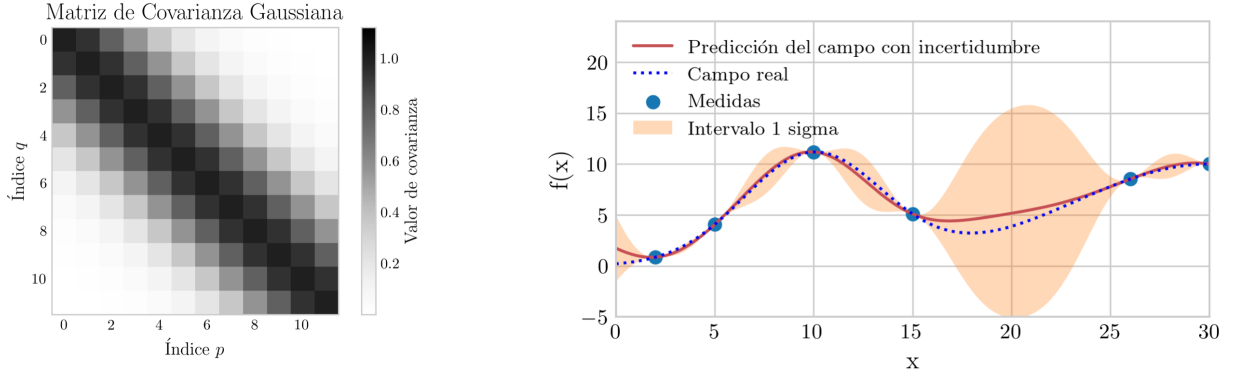


Fig. 7: **izquierda:** Visualización de la matriz de covarianza resultante de la función de covarianza 10 aplicada a 10 puntos, de manera que hay 10×10 pares de puntos, y por tanto la matriz es 10×10 . La intensidad de color es proporcional a la covarianza entre pares de puntos, nótese que en la diagonal el valor de la covarianza es máximo ya que $x_p = x_q$ **Derecha:** Ejemplo de regresión kriging con incertidumbre en cada punto representada como una banda naranja.

Una vez especificada la función de covarianza modelo y optimizados sus hiperparámetros (ver sección 5) se construye la distribución $f(\mathbf{x}) \sim \mathcal{N}(0, k(\mathbf{x}_p, \mathbf{x}_q))$, que es una distribución de funciones porque se pueden extraer muestras de la distribución evaluadas en cualquier número de puntos. Si elegimos un conjunto arbitrario de localizaciones a interpolar X_* y se escribe la matriz de covarianza K —a partir de la función de covarianza— elemento a elemento, podemos generar un vector gaussiano aleatorio $f(X_*) = \{f(\mathbf{x}_1), \dots, f(\mathbf{x}_n)\}$ y graficar los valores generados $f(X_*)$ en función de los argumentos (localizaciones) X_* . El resultado, como cabría esperar, serán realizaciones del campo aleatorio gaussiano modelo $f(x)$.

$$\mathbf{f}_* \sim \mathcal{N}(\mathbf{0}, K(X_*, X_*)) \quad (11)$$

Donde,

$$K_{p,q} = k(\mathbf{x}_p, \mathbf{x}_q) \equiv k(|\mathbf{x}_p - \mathbf{x}_q|) \quad (12)$$

se construye a partir de la ecuación (10) manteniendo la ordenación espacial de los índices. Entonces de manera genérica, un proceso gaussiano queda definido por su media $\mu(\mathbf{x})$ y función de covarianza $k(\mathbf{x}, \mathbf{x}')$.

$$f(\mathbf{x}) \sim \mathcal{PG}(\mu(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \quad (13)$$

4.4. Distribución posterior

Hasta ahora no se han incluido las medidas (evidencia). Para incorporar la información que aportan las medidas consideramos primeramente el caso de medidas sin ruido $\{(\mathbf{x}_p, f(\mathbf{x}_p)) \mid p = 1, \dots, n\}$, donde ahora se escribe, por simplicidad $\mathbf{f} \equiv f(\mathbf{x}_p)$. La distribución prior conjunta de las medidas, \mathbf{f} , y los puntos de test \mathbf{f}_* es:

$$\begin{bmatrix} \mathbf{f} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(X, X) & K(X, X_*) \\ K(X_*, X) & K(X_*, X_*) \end{bmatrix}\right) \quad (14)$$

Se tienen n medidas y n_* puntos de test. $K(X, X_*)$ es la matriz de covarianza $n \times n_*$ entre medidas y puntos test. De igual forma se tienen $K(X, X)$, $K(X_*, X_*)$ y $K(X_*, X)$. Para obtener la distribución posterior de funciones se debe restringir la prior conjunta para que contenga solo funciones que pasan por los puntos de medidas. A priori esto podría ser una complicación, pero como se ha visto en los requisitos de consistencia, el condicionando de la distribución gaussiana

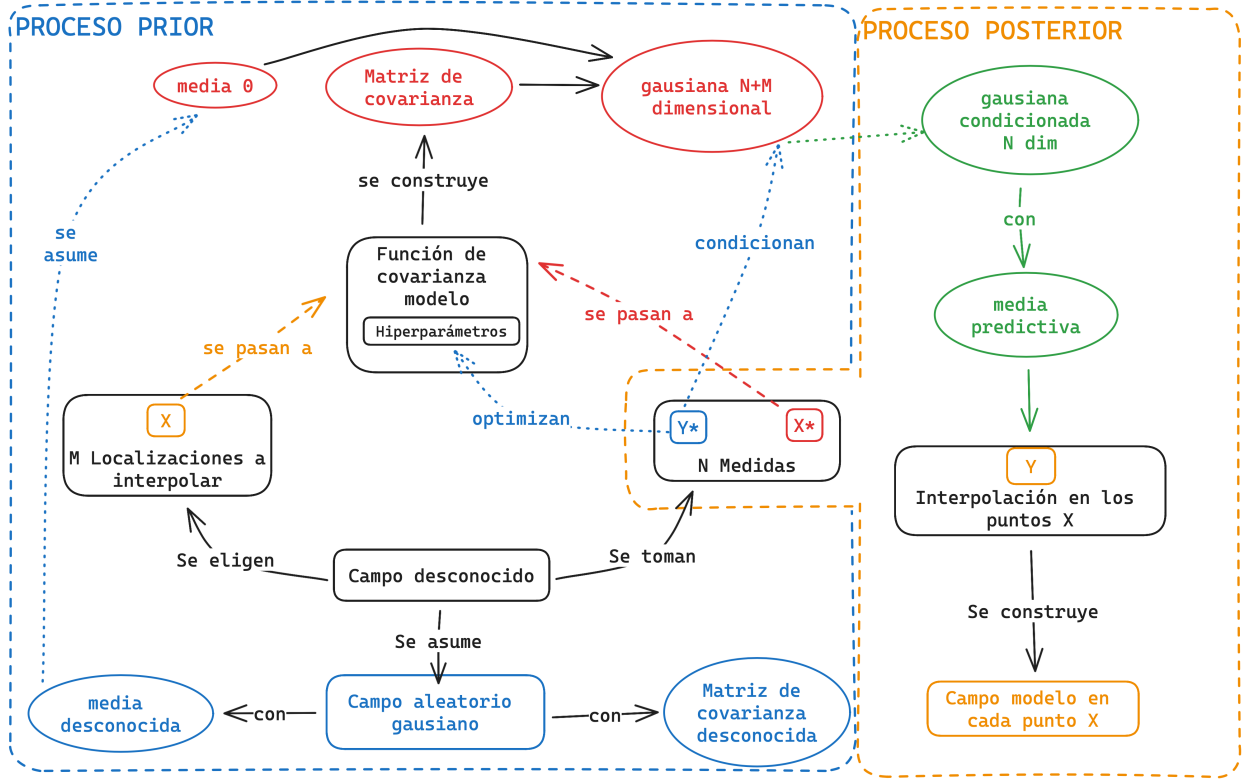


Fig. 8: Esquema de proceso gaussiano equivalente a kriging simple

conjunta prior a las medidas se puede expresar analíticamente, de manera que el único problema computacional a resolver es invertir la matriz K :

$$\mathbf{f}_* | X_*, X, \mathbf{f} \sim \mathcal{N}(\bar{\mathbf{f}}_*, \text{cov}(\mathbf{f}_*)), \text{ Donde} \quad (15)$$

$$\bar{\mathbf{f}}_* \triangleq \mathbb{E}[\mathbf{f}_* | X, \mathbf{f}, X_*] = K(X_*, X) K(X, X)^{-1} \mathbf{f} \quad (16)$$

$$\text{cov}(\mathbf{f}_*) = K(X_*, X_*) - K(X_*, X) K(X, X)^{-1} K(X, X_*) \quad (17)$$

Nótese que $\bar{\mathbf{f}}_*$ es la media predictiva, es decir, se predice la media en la localización a estimar. La varianza en la localización \mathbf{x}_q —que se interpreta como la incertidumbre de predicción para $f(\mathbf{x}_q)$ — es el elemento diagonal $[\text{cov}(\mathbf{f}_*)]_{qq}$. Se pueden generar tantos valores de \mathbf{f}_* (correspondientes a localizaciones donde se quiere interpolar X_*) como se deseen evaluando la media y función de covarianza y generando vectores a partir de la distribución posterior. Para incluir ruido $y = f(\mathbf{x}) + \varepsilon$ tal que $\varepsilon \sim \mathcal{N}(\mathbf{0}, \sigma_n^2)$ solo hay que hacer el cambio $K(X, X) \rightarrow K(X, X) + \sigma_n^2 I$ en las ecuaciones anteriores.

Una vez se especifica el modelo de función de covarianza, todos los términos de las ecuaciones anteriores son conocidos. Se pueden completar las matrices evaluando la función de covarianza en todos los pares de puntos disponibles: argumentos de las medidas y puntos a predecir. Se recuerda que $K(X, X)$ solo depende de la localización de las medidas — no de su valor — $K(X_*, X_*)$ solo depende de la distribución espacial de puntos a interpolar y $K(X_*, X)$ depende de ambos. La idea queda expuesta esquemáticamente en la figura 8. En la figura 9 derecha se presenta un ejemplo de regresión con aplicando las ecuaciones (16) y (17).

4.5. Kriging simple y universal

El kriging universal es un proceso gaussiano semiparamétrico en el que se considera la media como determinista, y no como una variable estocástica. Para incluir una función media determinada $m(\mathbf{x})$ simplemente se aplica el proceso gaussiano con media nula (kriging simple) a la diferencia entre las observaciones y la función de media fija:

$$\mathbf{f} \rightarrow \mathbf{y} - \mathbf{m}(X) \quad (18)$$

$$f(\mathbf{x}) \sim \mathcal{PG}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \quad (19)$$

Entonces $\text{cov}(\mathbf{f}_*)$ no varía pero la media predictiva $\bar{\mathbf{f}}_*$ queda:

$$\bar{\mathbf{f}}_* = \mathbf{m}(X_*) + K(X_*, X) K^{-1}(\mathbf{y} - \mathbf{m}(X)) \quad (20)$$

Sin embargo, en la práctica, la media en una región pequeña solo se puede conocer si hay repeticiones del fenómeno, como ocurre con los procesos espacio-temporales, o cuando el número de datos es tan grande que permite estimar la media casi a la perfección. Por esto en muchos casos puede ser más conveniente especificar algunas funciones de base fijas cuyos coeficientes, $\boldsymbol{\beta}$ se infieren mediante una simple regresión de mínimos cuadrados a las medidas. Esto es el kriging universal, que asume el modelo:

$$g(\mathbf{x}) = f(\mathbf{x}) + \mathbf{h}(\mathbf{x})^\top \boldsymbol{\beta} \quad (21)$$

Donde $f(\mathbf{x}) \sim \mathcal{PG}(0, k(\mathbf{x}, \mathbf{x}'))$ es un proceso gaussiano con media nula encargado de trazar la correlación espacial de los datos. $\mathbf{h}(\mathbf{x})$ son funciones base y $\boldsymbol{\beta}$ parámetros, encargados de representar la variabilidad global o tendencia de los datos; es, por tanto, la función de media determinista¹¹. Los datos medidos siguen una tendencia cercana a un modelo lineal global con sus errores modelados con el GP.

En el ejemplo mostrado en la figura (4) la tendencia es lineal $\mathbf{h}(\mathbf{x})^\top \boldsymbol{\beta} = \mathbf{x}\boldsymbol{\beta}$. Sustrayendo la tendencia a los datos originales, se obtienen los errores, $\varepsilon(\mathbf{s})$, que se asumen autocorrelacionados y aleatorios con media 0 y por tanto pueden ser modelados por el proceso gaussiano $f(\mathbf{x})$. De esta manera se combina un modelo frecuentista o determinista (tendencia), que modela cualquier tendencia espacial en los datos en el kriging universal, con un modelo estocástico (proceso gaussiano) que modela la variabilidad espacial de los datos. Se podría haber usado cualquier otra función de base, como un polinomio cúbico o una función periódica mientras se adapte a la tendencia observada.

En resumen, el kriging universal realiza una regresión a las medidas con las coordenadas espaciales como variables explicativas. Sin embargo, en lugar de asumir que los errores $\varepsilon(\mathbf{s})$ son independientes, se modelan como autocorrelacionados. Las ecuaciones resultantes son similares al kriging simple y ordinario, y se encuentran, con la misma notación, en [10].

5. Optimización de los hiperparámetros de la función de covarianza

En la sección 4.3 se ha visto que los elementos de todas las matrices de covarianza se obtienen usando la función de covarianza modelo. Para optimizar los hiperparámetros de esta función se realiza una regresión lineal de mínimos cuadrados¹² con la función de covarianza modelo a los valores de las covarianzas experimentales.

¹¹En la derivación usual de procesos gaussianos, se toma $\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{b}, B)$, donde B es la matriz de covarianza proveniente de la incertidumbre de los parámetros de la media, de manera que se pueden marginalizar los parámetros, obteniendo otro proceso gaussiano, sin embargo, en el kriging la media se considera determinista y $B=0$. Ver [10]

¹²Otras técnicas usadas son máxima verosimilitud y métodos bayesianos

5.1. Hipótesis de estacionariedad de primer y segundo orden

Para calcular la covarianza experimental entre pares de puntos, si se tuvieran $n > 1$ realizaciones en cada localización $\mathbf{f}(\mathbf{x}_p) = \{f(x_{p_1}), \dots, f(x_{p_n})\}$, $\mathbf{f}(\mathbf{x}_q) = \{f(x_{q_1}), f(x_{q_2}) \dots f(x_{q_n})\}$ se podría usar (1) para encontrar $E[\mathbf{f}(\mathbf{x}_p)] = \mu_p$ y $E[\mathbf{f}(\mathbf{x}_q)] = \mu_q$. Así, con (3) se obtiene directamente la covarianza entre $\mathbf{f}(\mathbf{x}_p)$ y $\mathbf{f}(\mathbf{x}_q)$ como:

$$\text{Cov}(\mathbf{f}(\mathbf{x}_p), \mathbf{f}(\mathbf{x}_q)) = \frac{1}{n} \sum_{i=1}^n (f(x_{p_i}) - \mu_p) \cdot (f(x_{q_i}) - \mu_q)$$

El problema reside en que, en fenómenos espaciales solo hay una realización del campo en cada localización $\mathbf{f}(\mathbf{x}_p) = f(x_p)$ y $\mathbf{f}(\mathbf{x}_q) = f(x_q)$ y la expresión anterior se reduce a:

$$\text{Cov}(f(x_p), f(x_q)) = E[\{f(x_p) - \mu(x_p)\} \cdot \{f(x_q) - \mu(x_q)\}]$$

La primera mitad de la solución es asumir que los valores esperados son los mismos en todas las localizaciones del campo, de manera que se puede calcular el valor esperado del campo como el valor esperado del conjunto de medidas disponibles. $\mu = E[\{f(x_p), \dots, f(x_q), \dots\}]$ Esto se conoce como estacionariedad de primer orden. Entonces la expresión anterior se reduce a:

$$\text{Cov}(f(x_p), f(x_q)) = E[\{f(x_p) - \mu\} \cdot \{f(x_q) - \mu\}]$$

La segunda mitad de la solución es asumir que la covarianza entre medidas depende solo de la separación de su localización asociada, entonces se puede calcular la covarianza de un conjunto medidas cuyas localizaciones están separadas por la misma distancia, asumiendo que son realizaciones de una misma variable aleatoria. De esta manera se está calculando la autocovarianza¹³.

$$\begin{aligned} \text{Cov}[f(\mathbf{x}), f(\mathbf{x} + \mathbf{h})] &= E[\{f(\mathbf{x}) - \mu\} \cdot \{f(\mathbf{x} + \mathbf{h}) - \mu\}] \\ &= E[\{f(\mathbf{x})\} \cdot \{f(\mathbf{x} + \mathbf{h})\} - \mu^2] \\ &\equiv C(\mathbf{h}) \end{aligned}$$

Esto implica una varianza (4) constante puesto que para $\mathbf{h} = 0$

$$C(0) = \text{Var}(f) = \sigma_f^2$$

La varianza constante se denomina estacionariedad de segundo orden.

De esta forma se puede calcular el valor esperado del vector aleatorio con las múltiples realizaciones de este. Solo hay una realización del campo en cada punto, pero cada punto es una realización diferente, entonces se pueden considerar todas las medidas como realizaciones de una misma variable aleatoria vectorial.

Por esto no es posible construir una matriz de correlación experimental, como se hizo para la función de covarianza modelo, porque no está definida la correlación entre pares de puntos, sino entre pares de conjuntos de puntos distanciados por \mathbf{h} . En la práctica esto se hace tomando todos los pares de medidas cuya separación de argumentos es aproximadamente \mathbf{h} . Un ejemplo de se muestra en la figura 9.

¹³No confundir con autocorrelación $\rho(\mathbf{h}) = C(\mathbf{h})/C(0)$, que es la autocovarianza normalizada por la varianza total del campo $C(0) = \sigma_f^2$, que es la covarianza en un punto y, como se ha explicado, se considera constante en todo el campo.

$$S(\mathbf{h}) = \{(p, q) \mid \|\mathbf{x}_p - \mathbf{x}_q\| \approx \mathbf{h}\} \quad (22)$$

$$C(h) = \frac{1}{N} \sum_{p, q \in S(\mathbf{h})} [\{f(\mathbf{x}_p) - \mu\} \cdot \{f(\mathbf{x}_q) - \mu\}]$$

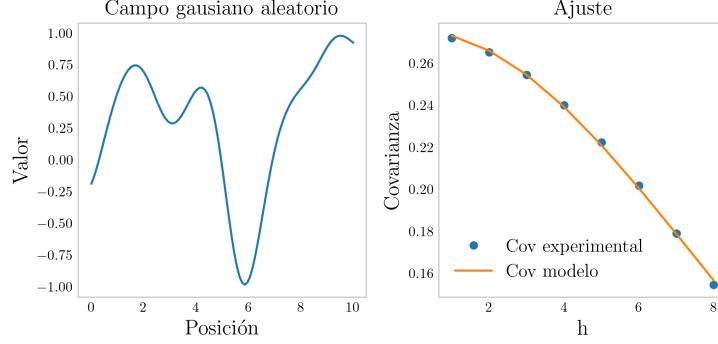


Fig. 9: Ejemplo de ajuste de la función de covarianza modelo (10) a los valores de la covarianza experimental obtenidos de un campo gaussiano aleatorio unidimensional. En el caso multidimensional se suele asumir que el campo es isótropo para que el covariograma—que es el conjunto de puntos de las covarianzas experimentales graficado— sea igual en todas direcciones, aunque se pueden definir funciones de covarianza dependientes del ángulo.

6. Método de selección de puntos de muestreo

Como se expone en [5] los agentes solo realizan una medida cuando han llegado a su respectivo punto de muestreo. Una vez allí retransmiten la medida a sus vecinos. Cada agente actualiza su modelo del campo solo cuando ha recibido información nueva (de un vecino que ha llegado a su respectivo punto de muestreo). Los agentes añaden la información recibida a su memoria $S_i(t_k)$ aumentando su conocimiento del campo con el tiempo. Así puede ser que un agente recalculé su siguiente punto de muestreo \mathbf{x}_{k+1}^d sin haber llegado al anterior \mathbf{x}_k^d . De esta forma es menor el número de mediciones necesarias.

Para lograr la máxima eficiencia, la selección de puntos de muestreo no debe ser arbitraria. Cada agente debe ser capaz de especificar su siguiente punto de muestreo usando los datos de su memoria, teniendo en cuenta las restricciones de su dinámica. Un correcto método de elección de siguientes puntos de muestreo distribuye la carga de exploración entre los agentes. Los dos métodos expuestos a continuación se extraen de [5].

6.1. Método 1

El primer método es el más simple. El siguiente punto de búsqueda es el valor máximo del campo modelado $\hat{\phi}_{i,k}$ por el agente i en el tiempo t_k

$$\mathbf{x}_i^d(t_k) = \arg \max_{\mathbf{x} \in D} \left\{ \hat{f}_{i,k}(\mathbf{x}) \right\} \quad (23)$$

Este método tiene como principal desventaja que el siguiente punto de muestreo será igual para grupos de agentes vecinos entre ellos, de manera que convergen rápidamente formando grupos que

son efectivamente como un solo agente al aportar todos la misma información. Cuando se dan estas agrupaciones de puntos se rompe la homogeneidad en la distribución de las medidas y el modelo puede ser poco fiable.

6.2. Método 2

El segundo método es más sofisticado e intenta distribuir mejor la carga de búsqueda. Supongamos que la estimación del máximo de f realizada por el agente i en el tiempo t_k es:

$$M_{\text{máx}}^i(t_k) = \max_{\mathbf{x} \in S_i(t_k)} \left\{ \hat{f}_{i,k}(\mathbf{x}) \right\} \quad (24)$$

Y el siguiente punto de muestreo para el agente i se calcula como:

$$\mathbf{x}_i^d(t_k) = \arg \min_{\mathbf{x} \in D} \left\{ J_i^{(k)}(\mathbf{x}) \right\} \quad (25)$$

$$J_i^{(k)}(\mathbf{x}) = \|\mathbf{x}_i(t_k) - \mathbf{x}\|^2 - \sum_{j \in \mathcal{N}_i(t_k)} \alpha \|\mathbf{x}_j(t_k) - \mathbf{x}\|^2 \quad (26)$$

Donde α y b son dos parámetros positivos.

Con la siguiente condición que restringe el subconjunto de D que potencialmente contiene el máximo:

$$\hat{f}_{i,k}(\mathbf{x}) + b\sigma_{f,i,k}(\mathbf{x}) > M_{\text{máx}}^i(t_k) \quad (27)$$

$\sigma_{f,i,k}$ es la varianza —que se entiende como la incertidumbre— asociada al campo modelo estimado $\hat{f}_{i,k}$ cuyo valor se obtiene, como se ha explicado en la sección 4.4, de los elementos diagonales de (17). b es un parámetro llamado ‘curiosidad’ pues de su valor depende la relevancia de las regiones con mayor incertidumbre. Así se busca un siguiente punto de muestreo lo suficientemente cerca a la posición actual del agente $\mathbf{x}_i(t_k)$ y lo suficientemente lejos de las posiciones de otros agentes vecinos $\mathbf{x}_j(t_k)$, $j \in \mathcal{N}_i(t_k)$. La convergencia al máximo no está garantizada con este método. En la sección 9.2 se muestran algunos resultados de convergencia para el caso unidimensional.

7. Ley de control

Cuando un agente calcula su siguiente punto de muestreo $\mathbf{x}_i^d(t_k)$, debe modificar su trayectoria para llegar a dicho punto y tomar una medida allí. Cada agente se mueve hacia \mathbf{x}_i^d con velocidad terminal $\dot{\mathbf{x}}_i^d$, evitando colisionar con otros agentes. Esto se consigue usando la ley de control:

$$\mathbf{u}_i = C(\mathbf{x}_i, \dot{\mathbf{x}}_i \dot{\mathbf{x}}_i) - k_3 (\mathbf{x}_i - \mathbf{x}_i^d) + 2k_2 \sum_{j=1}^N (\mathbf{x}_i - \mathbf{x}_j) \frac{1}{q} \exp\left(-\frac{(\mathbf{x}_i - \mathbf{x}_j)^T (\mathbf{x}_i - \mathbf{x}_j)}{q}\right) - k_1 (\dot{\mathbf{x}}_i - \dot{\mathbf{x}}_i^d) \quad (28)$$

Donde k_1, k_2, k_3, q son parámetros positivos.

Esta ley de control se puede entender como un potencial experimentado por los agentes centrado en \mathbf{x}_i^d . El término con k_3 es el de mayor importancia, pues controla la atracción de los agentes hacia el pozo del potencial. Nótese que la aceleración del agente siempre es paralela a la fuerza atractiva, esto podría no ser así si se trata de vehículos que tienen que maniobrar para redireccionarse. El término con k_1 define un potencial de velocidades, es decir, una velocidad terminal. El término con k_2 define un potencial de repulsión gaussiano, de manera que podemos interpretar q como la

desviación típica, que en este caso controla la anchura del potencial de repulsión.

La ley de control lleva a los agentes asintóticamente a la posición deseada $\mathbf{x}_i^d(t_k)$ con velocidad nula $\dot{\mathbf{x}}_i^d(t_k) = 0$, El agente toma la medida cuando $\|\mathbf{x}_i(t_k) - \mathbf{x}_i^d(t_k)\| < \epsilon$. Es importante que el agente llegue al punto de muestreo con velocidad nula pues modela la posible tardanza de medición, como, por ejemplo, que el agente deba realizar una perforación en la roca.

8. Implementación del modelo y algoritmo

Por simplicidad, se usará kriging ordinario. Kriging universal se puede implementar fácilmente con [8]. El modelo de kriging se resuelve mediante la librería de python orientada a geoestadística *Pykrige* [1]. La condición $S_i(t_k) \neq S_i(t_{k-1})$ se realiza comparando la longitud de los arrays de memoria en t_k y t_{k-1} . Además, los agentes empiezan a calcular el modelo de kriging cuando tienen más de dos medidas en su memoria, es decir: $|S_i(t_k)| > 2$. El esquema del algoritmo se muestra en algoritmo 1.

Algorithm 1 Algoritmo de búsqueda de máximo

```

for every time instant  $t_k$  do
  for each agent  $i$  do
    if  $\mathbf{x}_i(t_k) = \mathbf{x}_i^d(t_k)$  then
      Toma una medida  $y_i$  en  $\mathbf{x}_i(t_k)$ 
      Retransmite  $\{y_i, \mathbf{x}_i(t_k)\}$  a los vecinos  $j \in \mathcal{N}_i(t_k)$ 
      Actualiza  $S_i(t_k)$ 
    end if
    if  $S_i(t_k) \neq S_i(t_{k-1})$  and Longitud( $S_i(t_k)$ )  $> 2$  then
      Actualiza el modelo de kriging
      utiliza un método de selección de siguiente punto de muestreo para hallar  $\mathbf{x}_i^d(t_k)$ 
    else
       $\mathbf{x}_i^d(t_k) = \mathbf{x}_i^d(t_{k-1})$ 
    end if
    calcula la función de control  $\mathbf{u}_i(t_k)$  (28)
  end for
end for

```

9. Resultados de la simulación

En esta sección se analizan los resultados obtenidos de varias simulaciones realizadas en *python* que implementan el algoritmo 1. Las posiciones iniciales de los agentes son aleatorias y su velocidad inicial nula. La simulación termina una vez los agentes se encuentran a una determinada distancia del máximo $d < \sigma$. La búsqueda termina cuando no se satisface la condición (27) para ningún punto del dominio D . La simulación termina cuando los agentes convergen al máximo estimado tras finalizar la búsqueda. Para ambas simulaciones se usará el segundo método de selección de puntos de muestreo, expuesto en la sección 6.2.

9.1. Resultados en una dimensión con $N = 2$

Para una primera aproximación se ha considerado el problema en una dimensión, del que se puede extraer una valiosa información y entendimiento intuitivo que se aplicará a dimensiones superiores. La limitación más notable es que para poder converger al máximo en una dimensión no

puede haber repulsión entre agentes. El objetivo principal es la visualización del campo modelo para cada agente con su incertidumbre y el campo real simultáneamente, en la misma gráfica.

Por simplicidad visual —para que los agentes generen el mismo campo modelo— los 2 agentes tienen un radio de comunicación infinito. En el caso 2D se usarán tres agentes. El menor tiempo de ejecución en el caso 1D facilita la elección óptima de los parámetros de la ley de control, pudiéndose realizar múltiples simulaciones para comparar. En el caso 2D se verá como el tiempo de cálculo aumenta notablemente dificultando esta tarea, sin embargo, se usarán los valores de los parámetros optimizados en el caso 1D.

El campo sintético que los agentes exploran está definido en $D = [0, 50]$ se ha construido a partir de dos gaussianas con máximos en $x = 30$ y $x = 10$ siendo este último el argumento del máximo global. Aunque esta función no tenga, a primera vista, aspecto de campo aleatorio gaussiano, cualquier función puede ser entendida como una realización de un campo aleatorio gaussiano con una apropiada matriz de covarianza y cumplirá las condiciones expuestas en la sección 4.2. Para esta función¹⁴ no se conoce su media constante —nula o no— por tanto no se debería usar kriging simple, que como se ha visto asume media constante conocida. Por otro lado, se asume que tampoco tiene tendencia general de manera que no es necesario usar kriging universal. Entonces, la solución óptima en términos de eficiencia y coste computacional será el kriging ordinario, que asume media constante desconocida.

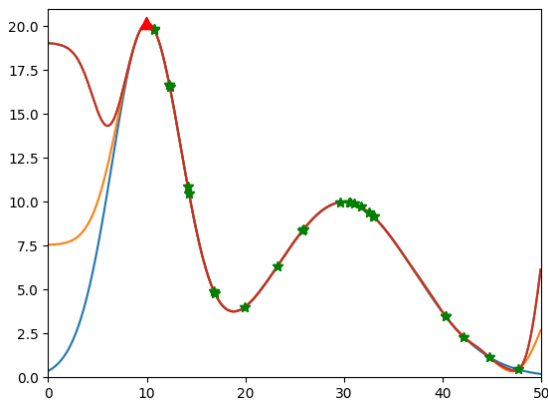


Fig. 10: Campo modelo en naranja, campo con incertidumbre en rojo y campo real en azul

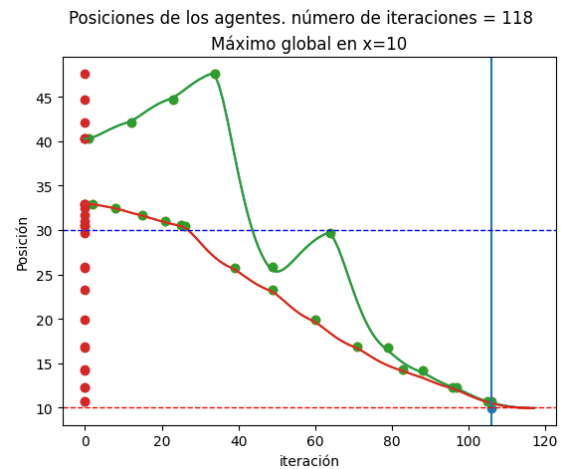


Fig. 11: Distancia al máximo con puntos de medición

En la figura 10 se puede observar como la incertidumbre en la zona (0-10), a pesar de no estar explorada, el modelo representa correctamente el máximo. Los agentes han explorado solo las zonas necesarias para descartar el mínimo local y converger al global. La línea vertical azul en la figura 11 marca la iteración en la que no se satisface la condición (27) para ningún punto en el dominio y por tanto se da la búsqueda por terminada, y los agentes convergen al máximo, finalizando la simulación. Cuando el agente realiza una medida se marca como un punto verde en el espacio iteración-posición. Los puntos rojos son las proyecciones de las medidas en el eje, de manera que se puede observar claramente la distribución de medidas en el espacio.

¹⁴Se utilizan los términos función y campo como sinónimos en este trabajo

N	#éxitos	#fallos	% de éxito	tiempo de ejecución (s)	#iteraciones
2	55	45	55	0.2232	83.65
3	66	34	66	0.1921	50.19
4	64	36	64	0.2140	42.90
5	72	28	72	0.2363	37.95
6	63	37	63	0.2475	30.38
7	70	30	70	0.2093	24.38
8	69	31	69	0.3311	25.97
9	65	35	65	0.2542	18.78

Cuadro 1: Resultados tras realizar 100 simulaciones para cada N de agentes usando el método 2. Para los tiempos de ejecución y número de iteraciones se ha tomado la media de las 100 simulaciones.

9.2. Tasa de éxito y escalabilidad del algoritmo en el caso 1D

Para comprobar la efectividad del algoritmo se han realizado 100 simulaciones. Se considera una búsqueda del máximo exitosa cuando el argumento del máximo calculado al finalizar la búsqueda y argumento del máximo real difieren en un valor ϵ , en este caso $\epsilon = 1$. El tiempo de ejecución¹⁵ es el tiempo de búsqueda (ver sección 9 para ver la diferencia entre final de la simulación y final de búsqueda). La tasa de éxito se define como:

$$\% \text{ Éxito} = \frac{\text{n}^{\circ} \text{ éxitos}}{\text{n}^{\circ} \text{ éxitos} + \text{n}^{\circ} \text{ fallos}} * 100 \quad (29)$$

En la tabla 1 se observa que el incremento de tiempo de ejecución por haber más agentes y por tanto tener que calcular mas veces el modelo de kriging se compensa con el hecho de que se muestrean más puntos y el máximo se encuentra antes, de manera que se mantiene constante la tasa de éxito y el tiempo de ejecución, disminuyendo el número de iteraciones. En esta simulación el radio de comunicación de los agentes es infinito, a medida que vaya creciendo el número de agentes la ralentización por el aumento de la complejidad del grafo de comunicaciones será más notable. La elección del número de agentes dependerá entonces de la movilidad que tengan en el medio y de si es importante el número de iteraciones necesarias.

9.3. Resultados en dos dimensiones con $N = 3$

El resultado en dos dimensiones es más aplicable a fenómenos espaciales reales, pero aumenta notablemente la complejidad y el tiempo de cálculo, mientras que disminuye la probabilidad de convergencia. En dos dimensiones sí se puede incorporar la repulsión entre agentes. En vez de un campo sintético con expresión simple, como la combinación de dos gaussianas en el caso 1D, se ha optado por generar un campo aleatorio gaussiano sintético, más realista, como el de la figura 3. Además, se usarán tres agentes.

La visualización del campo real y campo modelo ya no se puede hacer simultáneamente como en el caso 1D, por ello hay tres gráficas. En la figura 12a se muestra el campo gaussiano sintético generado y sobre el que los agentes realizan la exploración y cuyo máximo se encuentra en la parte superior derecha, marcado con una cruz. Al final de la simulación los agentes han realizado el modelo presentado en la figura 12b, que recoge las suficientes características del campo, como para converger al máximo. En la figura 12c se observa como la incertidumbre es más baja en la zona cercana al máximo permitiendo finalizar la simulación con el suficiente criterio.

¹⁵Nótese que, en un sistema real, cada agente realizaría propio cálculo, pero es útil estudiar el tiempo de cálculo del sistema en su conjunto.

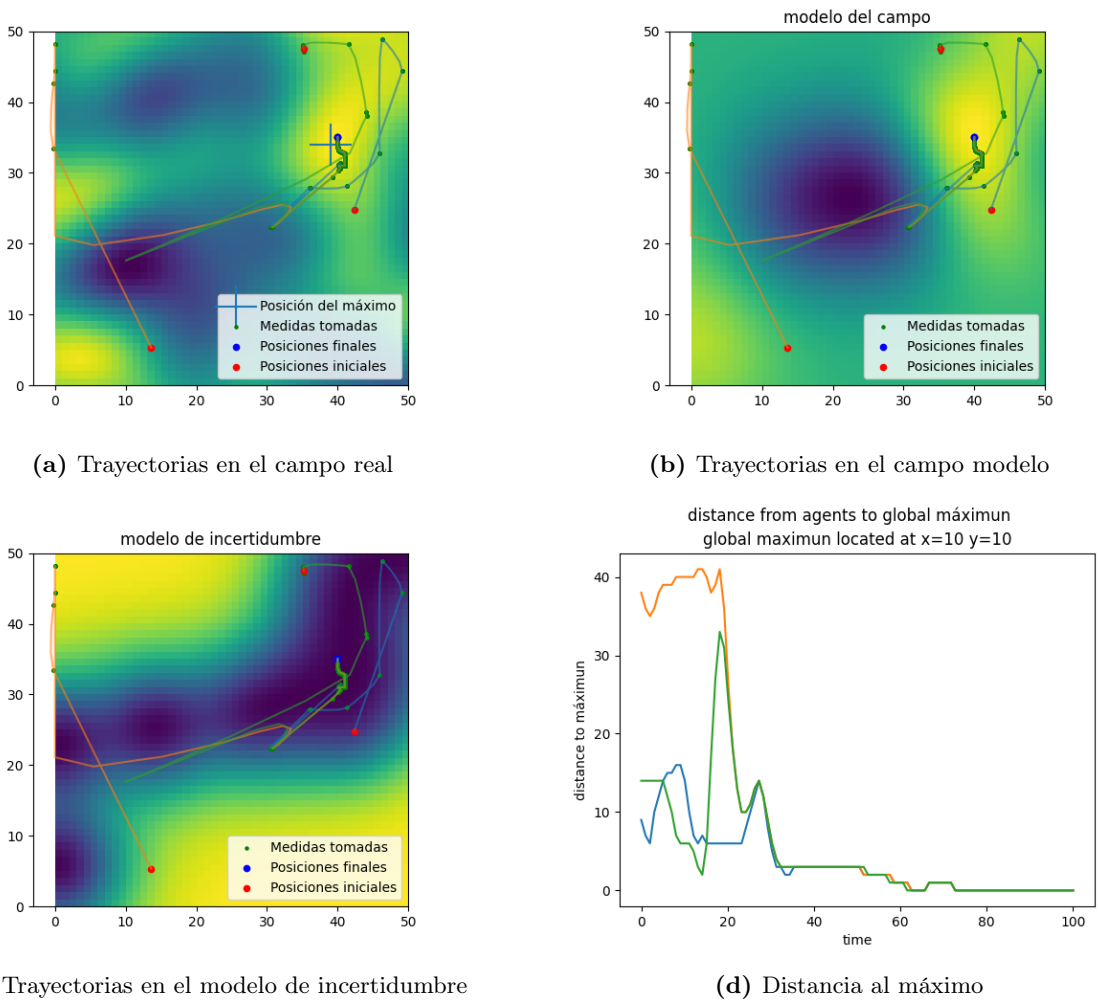


Fig. 12: Resultados de una simulación en dos dimensiones

Otro fenómeno interesante se observa cuando se siguen las trayectorias de los agentes. Tras realizar una medida el cambio de dirección es inmediato, hecho que es poco probable en la realidad, donde los agentes físicos, como aviones de ala fija, necesitarían maniobrar para alinearse con la nueva dirección. Esto es resultado de la ley de control (28) y se ha predicho este fenómeno en la sección 7.

Respecto a la eficiencia del método, se ha encontrado que los tiempos de ejecución son hasta dos órdenes de magnitud superiores al caso 1D. La convergencia es notablemente más baja que en el caso 1D, aunque no se ha hecho un estudio exhaustivo debido al largo tiempo de ejecución.

10. Conclusiones y futuras extensiones del trabajo

En este trabajo se ha visto que muchas distribuciones de fenómenos espaciales siguen la forma de un campo aleatorio gaussiano, y con el método propuesto se explota esta propiedad para encontrar el máximo. Además, se ha comprobado que un sistema multiagente puede encontrar el foco de estas distribuciones y se ha observado que para el caso 1D el tiempo de ejecución se mantiene aproximadamente constante aunque aumente el número de agentes. La simulación 2D también ha demostrado ser efectiva aunque la tasa de convergencia es notablemente menor mientras que el

tiempo de ejecución es hasta dos órdenes de magnitud mayor que en el caso 1D.

Como futuras mejoras, se contempla mejorar el modelo 2D, para estudiar una dinámica más realista para los agentes, que tenga en cuenta maniobras de viraje propias de aviones de ala fija o uniciclos. Se abre también la posibilidad de estudiar kriging 3D, más útil para estudios atmosféricos con baja variabilidad temporal. Además, otros métodos más sofisticados de selección de siguientes puntos de muestreo pueden considerarse. Se contempla también el tratamiento de fenómenos con variación espaciotemporal, como fenómenos atmosféricos o la superficie del agua en reservas con alta movilidad del medio. A nivel didáctico, muchas de las gráficas presentadas han sido construidas de manera que en un futuro podrían convertirse en animaciones o hacerse interactivas.

Referencias

- [1] Sebastian Müller Benjamin S. Murphy. Pykrige, May 28, 2024.
- [2] András Bárdossy. *Introduction to Geostatistics*. Institute of Hydraulic Engineering University of Stuttgart. 2012.
- [3] J.P. Chilès and P. Delfiner. *Geostatistics: Modeling Spatial Uncertainty*. Wiley Series in Probability and Statistics. Wiley, 2009.
- [4] Jochen Görtler, Rebecca Kehlbeck, and Oliver Deussen. A visual exploration of gaussian processes. *Distill*, 2019. <https://distill.pub/2019/visual-exploration-gaussian-processes>.
- [5] Arthur Kahn, Julien Marzat, Hélène Piet-Lahanier, and Michel Kieffer. Global extremum seeking by kriging with a multi-agent system. *IFAC-PapersOnLine*, 48, 07 2015.
- [6] Marcin Ligas and Marek Kulczycki. Simple spatial prediction - least squares prediction, simple kriging, and conditional expectation of normal vector. *Geodesy and Cartography*, 59:69–81, 01 2010.
- [7] G. Matheron. Principles of geostatistics. *Economic Geology*, 58(8):1246–1266, 1963.
- [8] S. Müller, L. Schüler, A. Zech, and F. Heße. *GSTools v1.3: a toolbox for geostatistical modelling in python*. *Geoscientific Model Development*, 15(7):3161–3182, 2022.
- [9] M.E.J. Newman. *Computational Physics*. CreateSpace Independent Publishing Platform, 2013.
- [10] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian processes for machine learning*. Adaptive computation and machine learning. MIT Press, 2006.
- [11] Matthias Schonlau. *Computer experiments and global optimization*. PhD thesis, CAN, 1997. AAINQ22234.
- [12] Schonlau, Matthias. *Computer experiments and global optimization*. PhD thesis, 1997.
- [13] M.L. Stein. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer Series in Statistics. Springer New York, 2012.
- [14] S. K. Suryasentana and B. B. Sheil. Demystifying the connections between gaussian process regression and kriging. pages 1–8, September 2023. 9th International SUT OSIG Conference : Innovative Geotechnologies for Energy Transition ; Conference date: 12-09-2023 Through 14-09-2023.