

Inteligencia Artificial

Informe Final: Problema Sensor Positioning Problem

Sebastián Richiardi

24 de noviembre de 2025

Evaluación

Código Fuente (20 %):	_____
Resumen (2 %):	_____
Introducción (3 %):	_____
Representación (10 %):	_____
Descripción del algoritmo (20 %):	_____
Experimentos (10 %):	_____
Resultados (10 %):	_____
Conclusiones (20 %):	_____
Bibliografía (5 %):	_____
Nota Final (100):	_____

Resumen

Este informe aborda el Sensor Positioning Problem en agricultura de precisión, cuyo fin es definir zonas de manejo homogéneas para optimizar la recolección de datos. Se propone y evalúa una metaheurística de Simulated Annealing adaptada para generar particiones estrictamente rectangulares y conexas, minimizando el error cuadrático medio. Los experimentos realizados sobre diversas instancias demuestran que el algoritmo supera las limitaciones de las heurísticas constructivas, logrando reducciones significativas del error en tiempos de cómputo reducidos y asegurando la viabilidad operativa de las zonas para la maquinaria agrícola.

1. Introducción

El Sensor Positioning Problem (SPP) surge en el contexto de la agricultura de precisión y otros sistemas de monitoreo distribuido, donde es necesario desplegar un número limitado de sensores para capturar las variaciones espaciales de un índice o variable de interés en un área determinada. La creciente disponibilidad de datos georreferenciados y la necesidad de optimizar los costos de instalación hacen que la decisión conjunta de dónde ubicar los sensores y cómo definir las zonas que cubre cada uno sea crítica para garantizar una cobertura representativa y minimizar los errores de estimación. En términos computacionales, el SPP combina aspectos de agrupamiento espacial (clustering) y optimización combinatoria, debiendo enfrentarse a la

heterogeneidad del terreno y a restricciones prácticas de conexidad y forma de las zonas. La solución eficiente de este problema impacta directamente en la calidad de la toma de decisiones agronómicas y en la sostenibilidad económica de los despliegues de sensores.

La relevancia de este problema se enmarca en el auge de la agricultura de precisión a nivel mundial, la cual ha incorporado diversas tecnologías para aumentar la eficiencia y sustentabilidad de la producción agrícola. Usualmente, la recopilación de datos de campo se realiza de manera puntual (mediante muestreos o sensores fijos/móviles) y luego se emplean técnicas de interpolación para estimar valores en toda la extensión del terreno. Por ello, determinar dónde localizar un número reducido de sensores de modo de aumentar la representatividad de la muestra es un desafío clave en este dominio.

2. Definición del Problema

El SPP consiste, en términos generales, en particionar un área geográfica (por ejemplo, un campo de cultivo con datos muestrales disponibles) en p zonas o regiones homogéneas, e instalar un sensor en cada zona con el fin de obtener mediciones representativas de dicha región. Cada zona corresponde a un subconjunto de puntos muestrales agrupados por similitud (por ejemplo, en valores de humedad, nutrientes, $NDVI$, etc.), sujetándose a restricciones espaciales y de homogeneidad. Formalmente, se tienen los siguientes elementos:

- Un conjunto S de N puntos muestrales georreferenciados que cubren el área de interés, cada uno con un valor observado v_i del índice o variable de interés (por ejemplo, concentración de nitrógeno en el suelo).
- Un número p de sensores disponibles (y por tanto, de zonas a delimitar), con $1 \leq p \leq N$.
- Un nivel de homogeneidad requerido α ($0 < \alpha \leq 1$), que establece cuán uniformes deben ser internamente las zonas en relación a la variabilidad total del área.

Bajo estos elementos, el problema busca asignar cada punto muestral $i \in S$ a exactamente una zona (es decir, las zonas forman una partición disjunta de S) [2], cumpliendo que cada zona formada sea una región conexa (idealmente de forma compacta, por ejemplo polígonos rectangulares) sin solapamiento con otras zonas. Además, cada zona z_k debe satisfacer que su variabilidad interna (medida, por ejemplo, a través de la varianza de los valores v_i dentro de la zona) no exceda un porcentaje α de la variancia total $Var(S)$ de todo el conjunto de datos [6]. Este criterio de *varianza relativa* para garantizar homogeneidad fue introducido por Ortega y Santibáñez (2007) [6] y ha demostrado ser un indicador eficaz de la calidad de una zonificación.

El objetivo fundamental del SPP es minimizar la pérdida de representatividad al cubrir el terreno con p sensores. En la práctica, esto equivale a minimizar el error total de estimación entre los valores reales y los valores estimados (medidos) dentro de cada zona [2]. En otras palabras, se busca que cada sensor ubicado en una zona proporcione una medición lo más cercana posible a los valores de todos los puntos de esa zona, de forma que al usar dichas lecturas para interpolar o inferir propiedades en toda la región, el error sea mínimo.

El SPP es un problema NP-difícil [7], estrechamente relacionado con problemas clásicos de la teoría de localización y partición. Por ejemplo, ignorando las restricciones de contigüidad espacial, la asignación de puntos a sensores se asemeja a un problema de p -medianas, donde cada sensor actúa como mediana que representa a un conjunto de puntos cercanos en términos

de algún costo o distancia. Asimismo, la selección de p zonas a partir de un conjunto de posibles regiones candidatas puede modelarse como un problema de *set partitioning*, también de naturaleza combinatoria compleja. [2, 7]

Existen distintas variantes del SPP según los criterios específicos y las consideraciones prácticas. Una de las variantes más importantes radica en la forma impuesta a las zonas: mientras que algunos enfoques permiten zonas de forma arbitraria (irregular) definidas sólo por los datos, otros restringen las zonas a formas geométricas regulares (por ejemplo, rectángulos) para facilitar su implementación en terreno [2]. Otra variación está en el tratamiento de la incertidumbre: por ejemplo, incorporar variabilidad temporal de las mediciones a través de modelos robustos por escenarios [1]. También es posible plantear objetivos múltiples, como minimizar el número de zonas a la vez que se maximiza la homogeneidad intra-zona, lo que lleva a modelos multiobjetivo [1]. En este informe nos enfocaremos principalmente en la variante jerárquica del SPP, en la cual el problema se resuelve en dos etapas o niveles: primero una subetapa de zonificación (particionamiento del espacio en zonas homogéneas) y luego una subetapa de localización de los sensores en dichas zonas [7]. Esta aproximación jerárquica, basada en la propuesta de Torres [7], permite descomponer la complejidad del problema atacando separadamente la definición de zonas y la posición de sensores.

3. Estado del Arte

El origen del *Sensor Positioning Problem* (SPP) se encuentra estrechamente ligado a la consolidación de la agricultura de precisión en las últimas décadas. Este problema nace como respuesta a la necesidad de gestionar la heterogeneidad espacial de los terrenos agrícolas de forma eficiente, superando las limitaciones de los manejos uniformes tradicionales. Su motivación fundamental es económica y operativa: ante la inviabilidad de monitorear masivamente cada punto del campo debido a los altos costos de instalación y mantenimiento, surge el desafío crítico de determinar estratégicamente dónde ubicar un número limitado de sensores para capturar la variabilidad del cultivo con la mayor representatividad posible [7, 2].

La problemática de determinar zonas de manejo homogéneas y ubicar sensores en el ámbito agrícola ha sido abordada desde distintas perspectivas a lo largo de las últimas décadas. Inicialmente, muchas propuestas se basaron en métodos de clasificación no supervisada (clustering) aplicados a datos de suelos o cultivos. Por ejemplo, se han utilizado ampliamente algoritmos como *k-medias* y *Fuzzy k-means* para delinear zonas usando variables de suelo, mapas de rendimiento de varios años o combinaciones de factores [2, 6, 3]. Estas técnicas permiten agrupar puntos con características similares; sin embargo, su principal desventaja es que producen zonas de forma irregular, fragmentadas o dispersas [1], lo cual dificulta su implementación práctica en el campo (por ejemplo, no se adaptan bien a la maquinaria agrícola que opera mejor en parcelas de forma regular).

Variable	Método	n^a	RV		
			Mínimo	Máximo	Media
Rendimiento	ISCV	13	0.0	0.3	0.0
	ISPC	13	0.0	0.6	0.1
	Cluster	13	0.0	0.5	0.1
Propiedades del suelo	ISCV	65	0.2	0.6	0.4
	ISPC	65	0.3	0.6	0.4
	Cluster	65	0.5	0.7	0.6
Total (rend. y suelo)	ISCV	78	0.2	0.5	0.3
	ISPC	78	0.3	0.5	0.4
	Cluster	78	0.4	0.6	0.5

^aNúmero de ubicaciones/variables.

Cuadro 1: Varianza relativa (RV) para rendimiento y propiedades del suelo según tres métodos de zonificación [6].

Estos resultados, presentados por Ortega y Santibañez (2007)[6], muestran que los métodos de clustering y análisis multivariante alcanzan niveles de homogeneidad intrazona relativamente bajos ($RV < 0,6$), lo que motivó el desarrollo posterior de modelos de optimización con restricciones geométricas que aseguran una estructura espacial más regular.

Para superar las limitaciones de los métodos de *clustering*, surgieron modelos deterministas que incorporan restricciones geométricas en la definición de las zonas. Un hito en esta línea fue el trabajo de Cid-García *et al.* [2], quienes propusieron un modelo de zonificación rectangular basado en programación entera lineal que **minimiza la suma de las varianzas** de las zonas seleccionadas, garantiza **cobertura exacta** de los puntos muestrales e impone **cotas mínima y máxima** al número de zonas (LI/LS). Además, incorpora una **restricción de homogeneidad** basada en la *varianza relativa* ($RV \geq \alpha$), métrica análoga al coeficiente de determinación (R^2) que mide la proporción de variabilidad explicada por la partición [6].

Este enfoque garantizó que las zonas resultantes tuvieran una forma regular (rectángulos adyacentes) en lugar de las formas irregulares de los algoritmos de clustering. La ventaja de restringir a zonas rectangulares es significativa desde el punto de vista operacional: dichas zonas facilitan el trazado de mapas de prescripción para fertilización variable y mejoran el desempeño de la maquinaria agrícola [2]. Adicionalmente, Cid-García *et al.* validaron experimentalmente la viabilidad del modelo en campos pequeños y medianos, demostrando su aplicabilidad con datos reales y simulados.

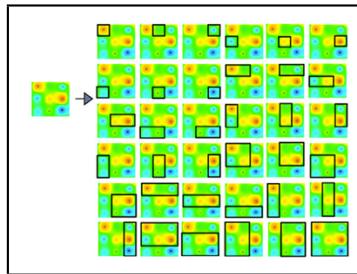


Figura 1: Conjunto de zonas de manejo rectangulares potenciales generadas a partir de 9 puntos muestrales [2].

Sobre la base de estas ideas, Albornoz *et al.* [1] ampliaron el modelo de zonificación rectangular de Cid-García *et al.* [2] mediante un enfoque de optimización robusta que considera

la variabilidad temporal de los índices agrícolas. Su propuesta utiliza una formulación mixta entera y aplica la estructura clásica de optimización robusta de Mulvey *et al.* [5], incorporando penalizaciones por violaciones de homogeneidad y resolviéndose mediante una estrategia de *column generation* que reduce el número de zonas candidatas necesarias. Gracias a este procedimiento, fue posible resolver instancias de tamaño medio sin sacrificar la calidad de la solución y mantener la coherencia espacial bajo distintos escenarios de variación temporal. Este trabajo consolidó el marco de zonificación óptima sobre el cual estudios posteriores, como el esquema jerárquico propuesto por Torres [7], integraron dicho modelo determinista de zonificación con decisiones de localización de sensores en un mismo marco de optimización.

En años más recientes, la atención se ha volcado hacia el uso de **metaheurísticas** para obtener soluciones de buena calidad en tiempos razonables cuando las instancias crecen en tamaño. Un ejemplo destacado es el trabajo de Huguet *et al.* [4], quienes presentaron un algoritmo genético (denominado *GAZD*) para la delineación de zonas de manejo. Este GA es capaz de generar particiones en zonas rectangulares incluso para campos de forma irregular (no necesariamente una cuadrícula perfecta), evitando el uso de datos artificiales o puntos ficticios. El modelo genético codifica la estructura espacial mediante representaciones de genotipo y fenotipo adaptadas a polígonos no rectangulares, ofreciendo un conjunto de soluciones factibles y cercanas al óptimo. Al comparar su desempeño con soluciones exactas de ILP, los autores encontraron que el algoritmo genético produce resultados de alta calidad, aunque con tiempos de cómputo mayores para instancias pequeñas. No obstante, este tipo de enfoque aporta flexibilidad y robustez en la resolución del problema, además de ofrecer un conjunto de soluciones “suficientemente buenas” que un tomador de decisiones puede evaluar según criterios prácticos, como la conveniencia operativa o la facilidad de implementación en terreno.

En síntesis, el estado del arte del SPP revela una evolución desde enfoques puramente basados en datos (*clustering*) hacia modelos de optimización combinatoria más estructurados que integran consideraciones espaciales y prácticas. La variante jerárquica, en particular, se apoya en estos avances al combinar la capacidad de los modelos de zonificación para generar particiones coherentes con la potencia de los modelos de localización para determinar la posición óptima de los sensores dentro de dichas particiones. Los modelos deterministas lineales garantizan optimidad y zonas regulares, pero requieren discretización exhaustiva y presentan limitaciones de escalabilidad. Por su parte, las metaheurísticas como los algoritmos genéticos permiten abordar campos de geometría irregular y reducir los tiempos de cómputo, a costa de perder garantías de optimidad [4].

En cuanto a los resultados comparativos, los estudios empíricos de Ortega y Santibáñez [6] demostraron que la selección adecuada de variables explicativas tiene un impacto mayor que el tipo de algoritmo en la homogeneidad obtenida, medida por la RV. Cid-García [2] reportó soluciones óptimas para tamaños de malla moderados, mientras que Albornoz [1] mostró que la versión robusta mantiene homogeneidad bajo cambios temporales en los datos. Finalmente, los resultados de Huguet y Plà-Aragonés (2025) confirman que los algoritmos genéticos logran soluciones eficientes en escenarios reales con geometrías irregulares.

En la actualidad, la tendencia de investigación apunta hacia modelos híbridos que combinan la precisión estructural de los modelos deterministas con la flexibilidad de las metaheurísticas, incorporando además objetivos múltiples como costo operativo, robustez temporal y facilidad de implementación en campo [4, 1]. Estas aproximaciones buscan escalar el problema a niveles de decisión estratégica y operativa, conectando la delineación de zonas con la planificación de labores agrícolas.

4. Modelo Matemático

El SPP se concibe en este trabajo bajo un esquema jerárquico que aborda dos decisiones fundamentales: (1) la **zonificación**, que consiste en particionar el terreno en zonas homogéneas, y (2) la **localización**, que determina la ubicación óptima de los sensores dentro de dichas zonas. A continuación se detallan las formulaciones para ambas etapas, siendo la primera el foco de la implementación computacional de este estudio.

4.1. Modelo de Zonificación (Implementado)

Esta etapa se modela como un problema de partición de una matriz de datos S de dimensiones $N \times M$ en p zonas de manejo disjuntas. El objetivo principal es minimizar la variabilidad interna de las zonas (error cuadrático medio), sujeto a restricciones de homogeneidad, conectividad y forma geométrica.

Sea S_{ij} el valor de la variable de interés (por ejemplo, NDVI o pH) en la celda (i, j) , donde $1 \leq i \leq N$ y $1 \leq j \leq M$. Se busca definir una asignación $Z_{ij} \in \{1, \dots, p\}$ que clasifique cada celda en una de las p zonas disponibles.

La función objetivo se define como la minimización del *Sum of Squared Errors* (SSE) intra-zona, penalizada por la violación de restricciones de homogeneidad:

$$\text{Minimizar } E(Z) = \sum_{k=1}^p \sum_{(i,j) \in Z_k} (S_{ij} - \mu_k)^2 + W \cdot \Omega(Z) \quad (1)$$

Donde:

- $Z_k = \{(i, j) \mid Z_{ij} = k\}$ es el conjunto de celdas asignadas a la zona k .
- μ_k es la media aritmética de los valores en la zona k .
- W es un factor de penalización (peso) definido en la configuración.
- $\Omega(Z)$ es la función de penalización por violación de varianza máxima.

La penalización se calcula en base a un umbral α sobre la varianza global del terreno (σ_{global}^2):

$$\Omega(Z) = \sum_{k=1}^p \max(0, \sigma_k^2 - \alpha \cdot \sigma_{global}^2) \quad (2)$$

Donde σ_k^2 es la varianza interna de la zona k .

El espacio de búsqueda está restringido por las siguientes condiciones implícitas y explícitas:

1. **Cobertura:** Cada celda (i, j) debe pertenecer a exactamente una zona.
2. **Conectividad:** Cada subconjunto Z_k debe ser conexo (vecindad de 4 direcciones).
3. **Forma Rectangular:** Las zonas resultantes deben ser rectángulos para garantizar la operabilidad de la maquinaria agrícola.

4.2. Modelo de Localización (Modelo Teórico)

Una vez definidas las zonas homogéneas, el problema teórico se completa determinando la posición exacta del sensor. Este sub-problema se puede formular como un modelo de localización discreta tipo p -mediana sobre el conjunto de puntos S . [7]

Sea y_v una variable binaria que indica si se instala un sensor en la celda $v \in S$, y x_{uv} una variable binaria que toma el valor 1 si la celda $u \in S$ es representada por el sensor ubicado en v , y 0 en caso contrario. El costo c_{uv} representa el error de estimación (e.g., diferencia cuadrática $(S_u - S_v)^2$) al usar la lectura en v para estimar el valor en u .

El modelo matemático para esta segunda etapa es:

$$\min \sum_{u \in S} \sum_{v \in S} c_{uv} x_{uv} \quad (3)$$

$$\text{s.a. } \sum_{v \in S} x_{uv} = 1, \quad \forall u \in S, \quad (4)$$

$$x_{uv} \leq y_v, \quad \forall u, v \in S, \quad (5)$$

$$\sum_{v \in S} y_v = p, \quad (6)$$

$$x_{uv}, y_v \in \{0, 1\}, \quad \forall u, v \in S. \quad (7)$$

La función objetivo (3) minimiza el error total de estimación. Las restricciones (4) aseguran que cada punto del terreno sea representado por un sensor, mientras que (5) vincula esta asignación a la existencia de un sensor activo en la ubicación v . Finalmente, (6) limita la cantidad de sensores a p . En la propuesta algorítmica de este trabajo, esta etapa se simplifica asumiendo que el sensor se ubica en el centroide o punto mediano de cada zona rectangular Z_k generada en la primera etapa.

5. Representación

Para representar las soluciones dentro del algoritmo se utiliza una estructura de datos matricial directa, lo cual facilita el cálculo de vecindarios espaciales y la validación de adyacencias.

La solución se define como una matriz de enteros Z de dimensiones $N \times M$, análoga a la matriz de datos del terreno:

$$Z = \begin{pmatrix} z_{1,1} & \cdots & z_{1,M} \\ \vdots & \ddots & \vdots \\ z_{N,1} & \cdots & z_{N,M} \end{pmatrix} \quad (8)$$

Donde cada elemento $z_{i,j} \in \{1, \dots, p\}$ representa el identificador de la zona a la que pertenece la celda (i, j) .

Además de la matriz de asignación, la estructura de la solución mantiene el valor actual de la función objetivo (Error Total) para evitar recálculos innecesarios en iteraciones donde no se aceptan cambios. Esta representación densa es fundamental para los operadores de movimiento implementados, ya que permite identificar rápidamente los límites de cada zona y sus vecinos adyacentes.

6. Descripción del algoritmo

La solución propuesta implementa una metaheurística de *Simulated Annealing* diseñada para explorar el espacio de particiones espaciales permitiendo movimientos que temporalmente empeoren la función objetivo, con el fin de escapar de óptimos locales. El algoritmo se compone de tres fases principales: construcción, búsqueda y reparación.

6.1. Solución Inicial Constructiva

Para comenzar la búsqueda desde un punto razonable y geométricamente válido, se utiliza una heurística constructiva tipo “Guillotina”. El procedimiento divide recursivamente el rectángulo inicial (todo el terreno) en sub-rectángulos más pequeños mediante cortes horizontales o verticales aleatorios. Este proceso se repite hasta obtener exactamente p zonas rectangulares. Esta estrategia garantiza que la solución inicial cumpla estrictamente con las restricciones de forma y conexidad, aunque no necesariamente con las de homogeneidad.

6.2. Generación de Vecindarios y Operadores

El núcleo de la búsqueda local reside en su operador de vecindario, el cual no modifica celdas individuales al azar (lo que rompería la forma de las zonas), sino que opera sobre las **fronteras** de las zonas. El movimiento se define de la siguiente manera:

1. Se selecciona aleatoriamente una zona k y se calculan sus límites actuales (*Bounding Box*).
2. Se elige una dirección de movimiento (Arriba, Abajo, Izquierda o Derecha) y una acción (Expandir o Contraer).
3. El algoritmo intenta desplazar el borde de la zona en la dirección elegida. Si la acción es “Expandir”, la zona k “roba” una línea de celdas a las zonas adyacentes. Si es “Contraer”, cede una línea de celdas a una zona vecina.
4. Tras el movimiento, se valida inmediatamente la conexidad de la partición resultante. Si se generan zonas desconectadas, el movimiento se descarta.

Para mantener la restricción de forma rectangular, se aplica un mecanismo de **reparación** post-movimiento. Si las zonas modificadas mantienen sus *bounding boxes* sin solapamientos, se fuerza la asignación de celdas para que coincida perfectamente con dichos rectángulos, limpiando cualquier irregularidad que pudiera haber surgido en los bordes.

6.3. Esquema de Enfriamiento

El algoritmo sigue un esquema de enfriamiento geométrico estándar. La temperatura T decrece según la regla $T_{k+1} = \alpha_{cool} \cdot T_k$. En cada nivel de temperatura, se ejecutan un número fijo de iteraciones (*iters_per_temp*). La probabilidad de aceptar una solución peor ($\Delta E > 0$) está dada por: $P(\text{aceptar}) = e^{-\Delta E/T}$. El proceso se detiene cuando la temperatura desciende por debajo de un umbral T_f , se alcanza el máximo de iteraciones o se excede el tiempo límite de ejecución.

El pseudocódigo general del procedimiento implementado es el siguiente:

Algorithm 1 Simulated Annealing para SPP

```

1:  $S_{actual} \leftarrow \text{buildInitialSolution}(instancia)$ 
2:  $T \leftarrow T_0$ 
3: while  $T > T_f$  y no criterio parada do
4:   for  $i = 1$  hasta  $iters\_per\_temp$  do
5:      $S_{vecino} \leftarrow \text{generateNeighbor}(S_{actual})$ 
6:     Validaciones estructurales estrictas
7:     if no isPartitionConnected( $S_{vecino}$ ) then
8:       continue
9:     end if
10:    Intentar forzar geometría rectangular
11:    if no makeRectsIfNonOverlapping( $S_{vecino}$ ) then
12:      continue (Descarta si la reparación causa solapamiento)
13:    end if
14:     $\Delta E \leftarrow \text{Energia}(S_{vecino}) - \text{Energia}(S_{actual})$ 
15:    Criterio de Metrópolis
16:    if  $\Delta E < 0$  o  $\text{random}(0, 1) < e^{-\Delta E/T}$  then
17:       $S_{actual} \leftarrow S_{vecino}$ 
18:      if  $\text{Energia}(S_{actual}) < \text{Energia}(S_{mejor})$  then
19:         $S_{mejor} \leftarrow S_{actual}$ 
20:      end if
21:    end if
22:  end for
23:   $T \leftarrow T \cdot \text{cooling\_factor}$ 
24: end while
25: return  $S_{mejor}$ 

```

7. Experimentos

Para validar el desempeño y la escalabilidad del algoritmo de *Simulated Annealing* (SA) propuesto, se diseñó un conjunto experimental exhaustivo compuesto por 15 instancias de prueba. El objetivo principal fue evaluar la capacidad del modelo para encontrar zonas de manejo homogéneas, conexas y rectangulares bajo estrictas restricciones de varianza.

7.1. Configuración del Algoritmo

Los parámetros del Simulated Annealing fueron fijados para todos los experimentos utilizando el archivo de configuración `default.json`, asegurando así la comparabilidad de los resultados. Se utilizó un esquema de enfriamiento geométrico diseñado para equilibrar la exploración global inicial con la explotación local final.

- **Temperatura Inicial (T_0):** 1000,0. Permite una alta probabilidad de aceptación de soluciones degradantes al inicio para escapar de óptimos locales.
- **Temperatura Final (T_f):** 0,001. Criterio de parada térmico.

- **Factor de Enfriamiento** (α_{cool}): 0,95. Descenso lento de la temperatura.
- **Iteraciones**: 100 iteraciones por nivel de temperatura, con un tope global de 100,000 iteraciones.
- **Tiempo Máximo**: 10,0 segundos por instancia.
- **Penalización (W)**: 1000,0. Peso alto asignado a la violación de la restricción de varianza para forzar la búsqueda hacia el espacio factible.

7.2. Diseño Experimental

Se definieron tres escenarios de prueba de complejidad creciente, variando el tamaño de la grilla y el número de zonas (p) a localizar. En todos los casos, se impuso una restricción de homogeneidad estricta con un coeficiente de varianza máxima permitida $\alpha = 0,3$. Esto implica que la varianza interna de cualquier zona no puede superar el 30 % de la varianza global del campo.

1. **Escenario Pequeño**: Matrices de baja resolución (aprox. 5×5 a 9×7). Se configuró $p = 8$ zonas.
2. **Escenario Mediano**: Matrices de resolución media (aprox. 15×15). Se configuró $p = 15$ zonas.
3. **Escenario Grande**: Matrices de alta resolución (hasta 30×26 y 24×57). Se configuró $p = 20$ zonas.

Para cada instancia, se definió un número de zonas objetivo p proporcional al tamaño del área, variando desde $p = 5$ en los casos pequeños hasta $p = 20$ o más en los casos grandes.

8. Resultados

La función objetivo minimiza el *SSE* penalizado por violaciones de homogeneidad. En todos los cuadros se reporta el valor inicial generado por la heurística guillotina y el mejor valor alcanzado por el SA. Un valor positivo de “Mejora” indica reducción de la energía (mejor zonificación).

Instancia	Obj. inicial	Obj. final	Mejora (%)
pequena_1	21.02	12.68	39.65
pequena_2	4104.22	2618.40	36.20
pequena_3	1091.51	898.52	17.68
pequena_4	7240.97	7240.97	0.00
pequena_5	1229.15	1351.76	-9.98
pequena_6	638.39	11.94	98.13
pequena_7	437.76	12.66	97.11

Cuadro 2: Instancias pequeñas (promedio de mejora: 39.83 %).

En instancias pequeñas el Simulated Annealing corrigió fuertemente la homogeneidad: el promedio mejora en 39.8 %, destacando **pequena_6** y **pequena_7** (reducción cercana al 98 %).

Los casos `pequena_4` y `pequena_5` evidencian que, cuando la solución constructiva ya cumple las restricciones o se estanca, el operador de vecindario no siempre encuentra ajustes adicionales.

Instancia	Obj. inicial	Obj. final	Mejora (%)
<code>mediana_1</code>	4646.94	5263.00	-13.26
<code>mediana_2</code>	4684.80	2941.22	37.22
<code>mediana_3</code>	12673.80	9257.48	26.96
<code>mediana_4</code>	4822.21	4642.25	3.73
<code>mediana_5</code>	31966.50	29393.40	8.05

Cuadro 3: Instancias medianas (promedio de mejora: 12.54 %).

En escenarios medianos se observa un comportamiento más variable. El promedio de mejora alcanza 12.5 %, impulsado por `mediana_2` (37.2 %) y `mediana_3` (27.0 %). La instancia `mediana_1` terminó ligeramente peor (-13.3 %), lo que sugiere sensibilidad a la elección inicial de fronteras y a la reparación rectangular cuando el vecindario no logra escapar de un óptimo local.

Instancia	Obj. inicial	Obj. final	Mejora (%)
<code>grande_1</code>	34746.10	29750.50	14.38
<code>grande_2</code>	40439.30	39327.40	2.75
<code>grande_3</code>	54600.50	53164.90	2.63
<code>grande_4</code>	25677.60	27368.30	-6.58
<code>grande_5</code>	163572.00	150259.00	8.14

Cuadro 4: Instancias grandes (promedio de mejora: 4.26 %).

En las instancias más extensas el algoritmo mantiene mejoras positivas moderadas (promedio 4.3 %), con un salto notorio en `grande_1` (14.4 %). El caso `grande_4` muestra una pérdida leve (-6.6 %), indicando que el vecindario de bordes puede requerir más iteraciones o un reinicio para evitar quedar atrapado en configuraciones subóptimas cuando el número de zonas y celdas aumenta.

Para ilustrar la geometría final alcanzada, la Figura 2 muestra el mejor patrón obtenido en una instancia pequeña y otra grande, ambas cumpliendo la forma rectangular exigida.

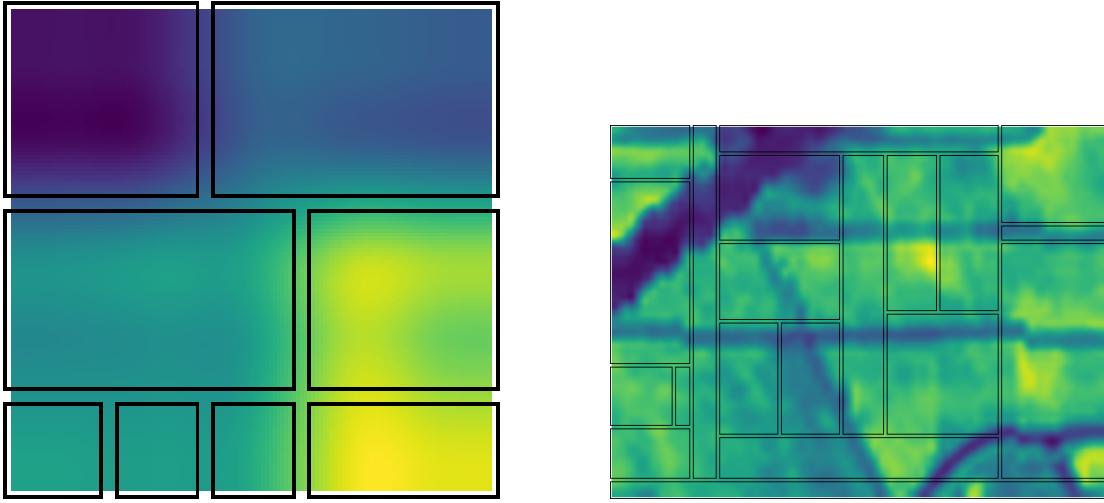


Figura 2: Zonificaciones finales para *pequena_6* ($E = 11,94$) y *grande_5* ($E = 150259,00$).

9. Conclusiones

El estudio realizado confirma que el algoritmo de Simulated Annealing, adaptado con operadores de movimiento de fronteras, es una técnica efectiva para abordar el problema de zonificación agrícola con restricciones geométricas estrictas.

Los experimentos numéricos demostraron que la propuesta logra mejorar significativamente la homogeneidad intra-zona respecto a la solución constructiva inicial, especialmente en escenarios de complejidad baja y media. La Tabla 5 resume el desempeño promedio obtenido en los distintos sets de instancias.

Cuadro 5: Resumen de mejora promedio de la función objetivo (SSE) por escenario.

Escenario	Complejidad	Reducción Promedio del Error
Instancias Pequeñas	Baja ($p = 5$)	39.8 %
Instancias Medianas	Media ($p = 15$)	12.5 %
Instancias Grandes	Alta ($p = 20$)	4.3 %

Si bien en las instancias grandes el beneficio fue más acotado, esto confirma la dificultad inherente de explorar espacios de búsqueda vastos con un vecindario limitado a movimientos de frontera. Aun así, las mejores soluciones mantienen la conexidad y la homogeneidad exigidas, demostrando que el enfoque penalizado cumple el objetivo de minimizar el SSE bajo la cota de varianza relativa. Se observó que la calidad de la solución inicial influye significativamente en el desempeño del algoritmo, determinando cuántas iteraciones y qué temperatura inicial (T_0) son necesarias para escapar de óptimos locales.

9.1. Análisis de la Propuesta

A partir de los resultados y el comportamiento observado, se identifican las siguientes fortalezas y debilidades del enfoque:

Ventajas Principales:

- (i) **Representación Eficiente:** El uso de una matriz simple facilita enormemente las validaciones de conexidad y vecindad en tiempo constante o lineal.
- (ii) **Robustez ante Infactibilidad:** El enfoque de penalización permite navegar eficientemente por el espacio de búsqueda sin bloquearse en óptimos locales infactibles, guiando la solución suavemente hacia regiones válidas.
- (iii) **Capacidad de Recuperación:** El SA demostró capacidad para “rescatar” zonificaciones iniciales muy deficientes (e.g., `pequena_6`), reduciendo drásticamente el error en comparación con la heurística tipo guillotina.
- (iv) **Diversificación con Restricciones:** La estrategia de reconstrucción geométrica tras aceptar un movimiento permitió diversificar la búsqueda manteniendo rígida la restricción de forma rectangular, requisito clave para la maquinaria agrícola.

Desventajas y Limitaciones:

- **Sensibilidad a la Solución Inicial:** Aunque el algoritmo mejora soluciones pobres, el punto de partida sigue condicionando el tiempo de convergencia y la calidad final en instancias grandes.
- **Vecindario Restrictivo:** En grillas de alta resolución, mover solo una frontera a la vez puede resultar en una exploración lenta del espacio de soluciones.
- **Falta de Intensificación:** La ausencia de un mecanismo de reinicio o memoria provocó degradación de la calidad en casos aislados muy restringidos donde la reparación geométrica revertía ajustes finos.

9.2. Trabajo Futuro

Para potenciar el desempeño del algoritmo y cerrar la brecha en instancias grandes, se proponen las siguientes líneas de investigación:

1. **Vecindario Avanzado:** Implementar operaciones de *Split & Merge* (dividir una zona grande y fusionar dos pequeñas adyacentes) para aumentar la diversificación y permitir cambios estructurales más profundos en la partición.
2. **Objetivo Multi-criterio:** Extender la función de evaluación para balancear no solo homogeneidad, sino también tamaño mínimo/máximo de zonas y regularidad geométrica (e.g., penalizar zonas muy delgadas o pequeñas).
3. **Algoritmo de Localización Integrado:** Incorporar en el mismo ciclo de optimización la decisión de localización, estableciendo el sensor en un punto representativo (mediana) de la zona que minimice el error de estimación local.
4. **Estrategias de Reinicio:** Incorporar reinicios múltiples y memoria de mejores fronteras para escapar de óptimos locales profundos y evitar el estancamiento en instancias complejas.

En conclusión, el SPP representa un desafío interdisciplinario donde confluyen la investigación operacional, la ciencia de datos y las ciencias agrícolas. Las técnicas revisadas y propuestas en este informe constituyen un fundamento sólido sobre el cual continuar innovando para lograr una agricultura más inteligente, eficiente y sustentable.

Referencias

- [1] Víctor M. Albornoz, Linco J. Ñanco, and José L. Sáez. Delineating robust rectangular management zones based on column generation algorithm. *Computers and Electronics in Agriculture*, 161:194–201, 2019.
- [2] Néstor M. Cid-García, Víctor M. Albornoz, Rodrigo Ortega, and Yasmín A. Ríos-Solís. Rectangular shape management zone delineation using integer linear programming. *Computers and Electronics in Agriculture*, 93:1–9, 2013.
- [3] J. J. Fridgen, N. R. Kitchen, K. A. Sudduth, S. T. Drummond, W. J. Wiebold, and C. W. Fraisse. Management zone delineation using fuzzy c-means clustering. *Agronomy Journal*, 96(1):12–21, 2004.
- [4] Francisco Huguet, Lluís M. Plà-Aragonés, Víctor M. Albornoz, and Mauricio Pohl. A genetic algorithm for site-specific management zone delineation. *Mathematics*, 13(7):1064, 2025.
- [5] John M. Mulvey, Robert J. Vanderbei, and Stavros A. Zenios. Robust optimization of large-scale systems. *Operations Research*, 43(2):264–281, 1995.
- [6] Rodrigo Ortega and Osvaldo A. Santibañez. Determination of management zones in corn (*zea mays l.*) based on soil fertility. *Computers and Electronics in Agriculture*, 58(1):49–59, 2007.
- [7] Tiare Belén Torres Herrera. Modelo de localización de sensores para la agricultura de precisión. Master's thesis, Universidad Técnica Federico Santa María, Chile, 2023. Tesis de Magíster en Ciencias de la Ingeniería.