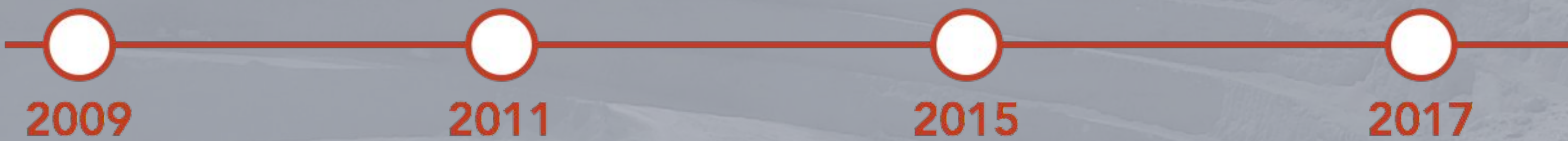


MACHINE LEARNING APLICADO A LAS EXPLORACIONES MINERAS



Alejandro Ehrenfeld



El AMTC fue creado en marzo de **2009** como parte del Programa de CONICYT para Centros Científicos y Tecnológicos de Excelencia



CONICYT
Comisión Nacional de Investigación Científica y Tecnológica



**CUALQUIER COSA QUE INTENTE
REEMPLAZAR AL EXPERTO
VA A FRACASAR**



¿QUÉ ES M.L.?



1. Datos



4. Modelos



2. Ruido



5. Entrenamiento



3. EDA



6. Validación



¿QUÉ ES M.L.?



Modelos

Supervised Learning

Classification

Regression

Grade prediction,
Image classification

Unsupervised Learning

Dimensionality Reduction

Clustering

Big data visualization,
Segmentation

Reinforcement Learning

Dimensionality Reduction

Control

Realtime decisions,
Learning tasks

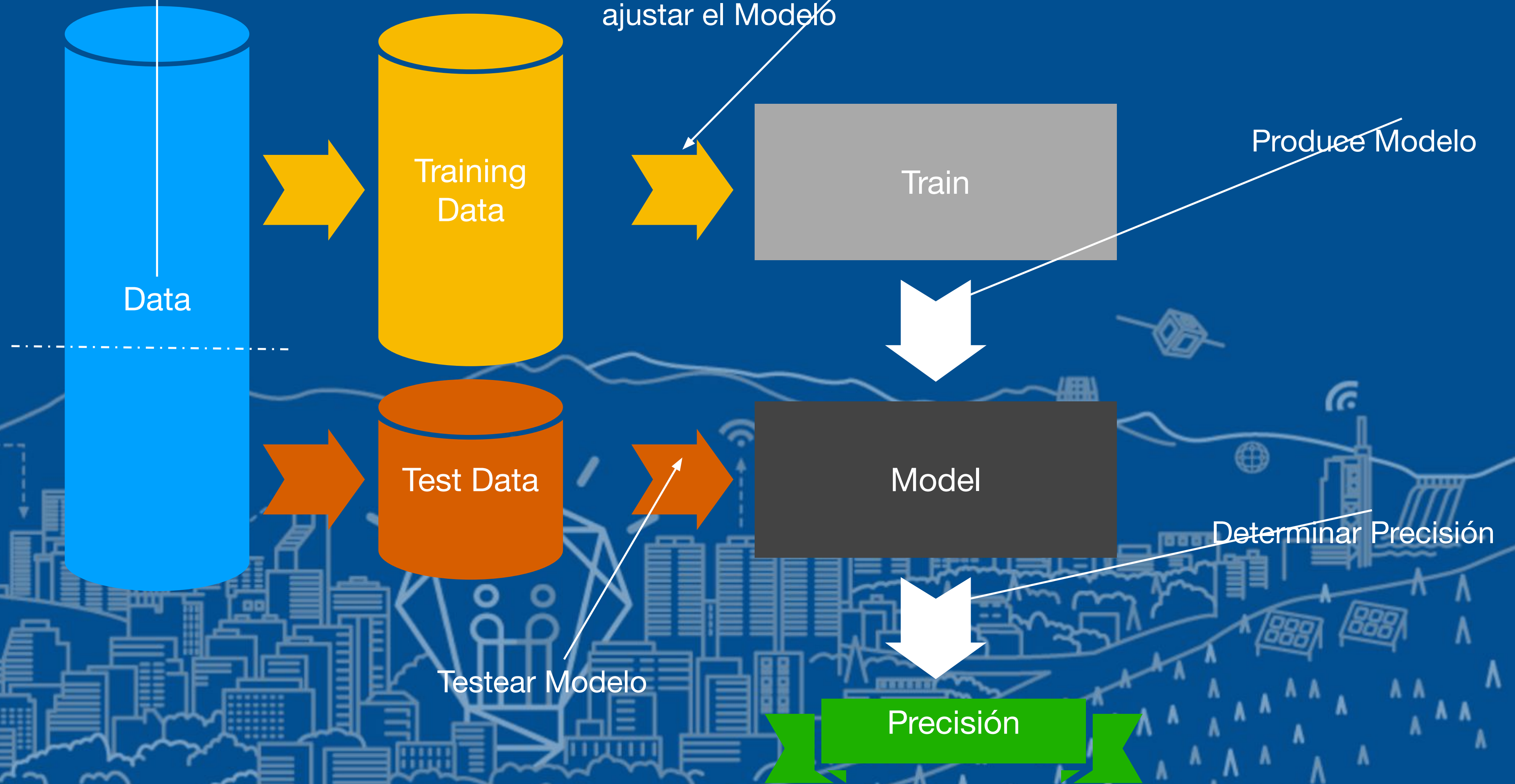
- El **modelo** depende absolutamente del problema a resolver, entender el problema, las variables disponibles, la cantidad de datos...
- Los resultados serán los algoritmos y el modelo a usar



Dividir los datos en Entrenamiento y Test

¿QUÉ ES M.L.?

Usar datos de entrenamiento para ajustar el Modelo



Testear Modelo

Produce Modelo

Determinar Precisión

Precisión

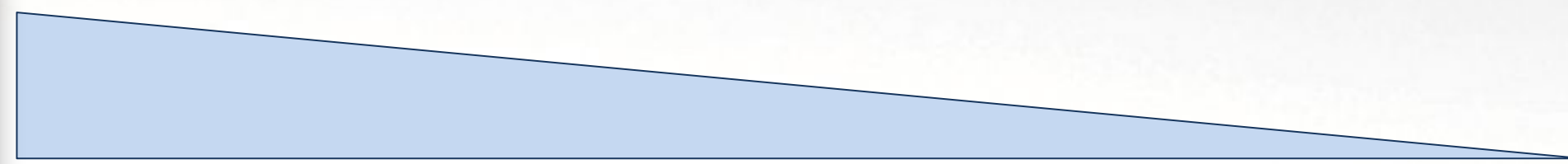
Ajuste del modelo, conjuntos de datos, análisis de resultados



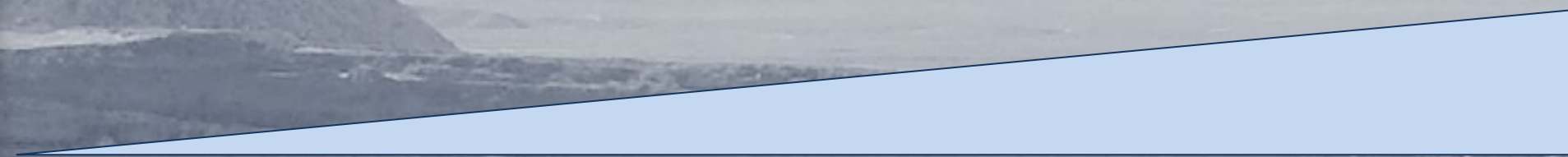
Entrenamiento



- Mucha incertidumbre
- Pocos datos



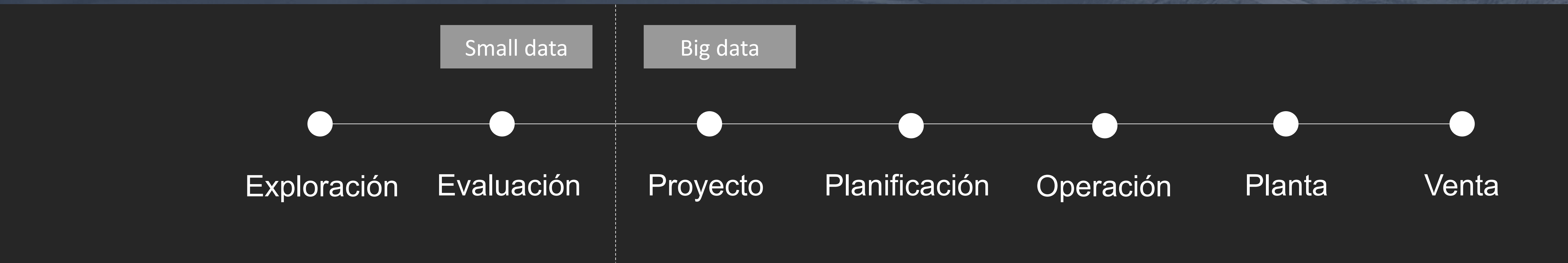
incertidumbre



datos



- Poca o nula incertidumbre
- Muchos datos





TARGET
DELINEATION



TD

GREEN FIELD



DRILL
TESTING



DT

BROWN FIELD



En el espacio (campo aleatorio)

$\vec{\theta}$

$$M(Z_{cu}) = \rho(Z_{cu} | \theta_1, \dots, \theta_n)$$

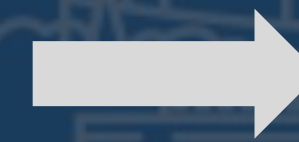
$\vec{\theta}$

1. Datos cuantitativos
2. Semi-cuantitativos (Sesgos en la medición)
3. Cualitativos

En nuestro caso:

- cuantitativos
- cualitativos

Datos
continuos
&
puntuales



Soporte común
Datos procesados &
continuos



RGB Images



Hyperspectral data



Topography



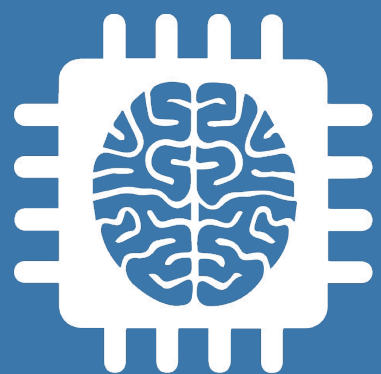
Topography



Drillhole Data



Advanced Geochemistry



MACHINE LEARNING



GEOESTADÍSTICA

Desde interpolación simple hasta sofisticadas técnicas de predicción

Oversegmentation

Data fusion

Scenario Simulation

Likelihood Evaluation

TD

Fertility Regression



Optimization

DT

Project milestones

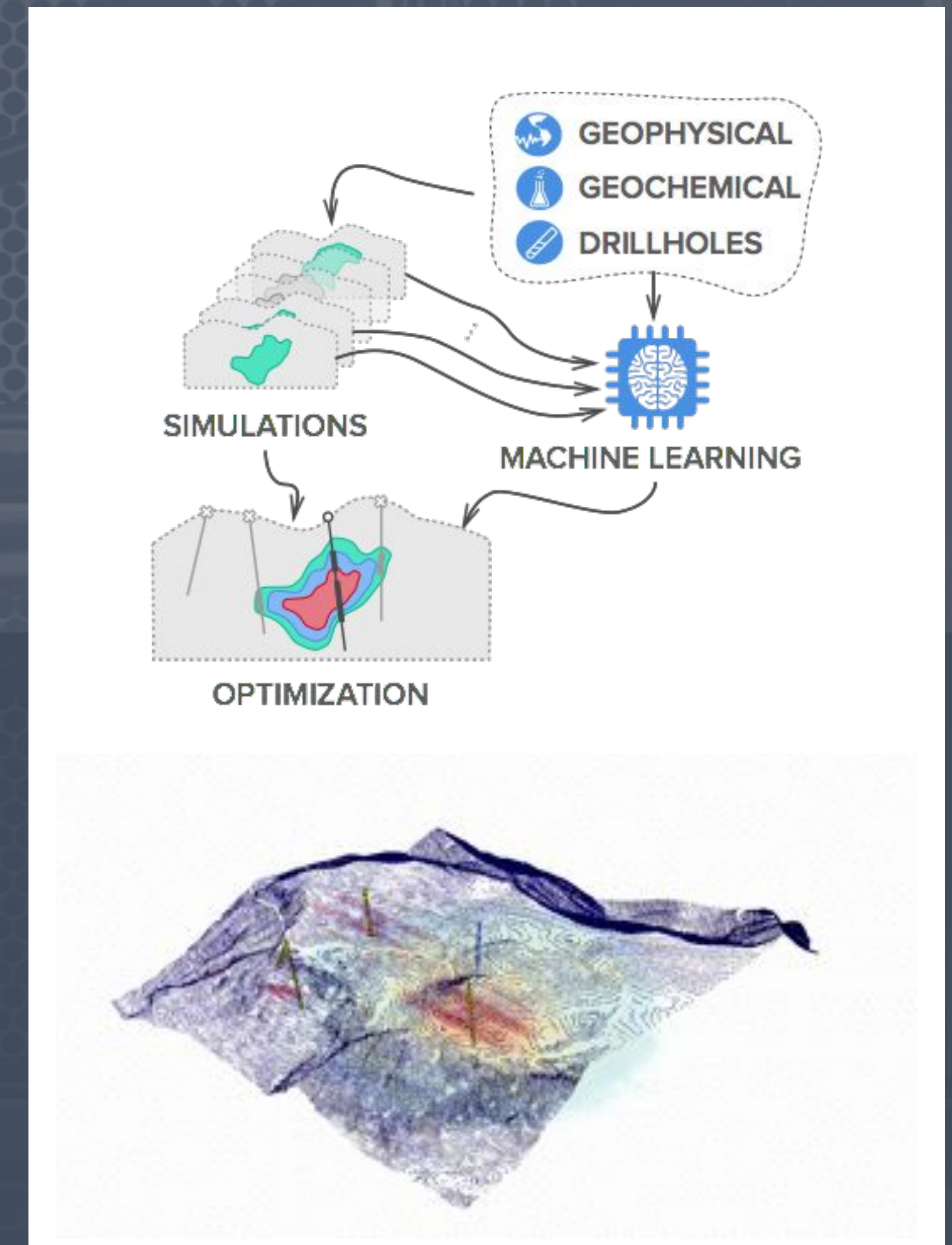
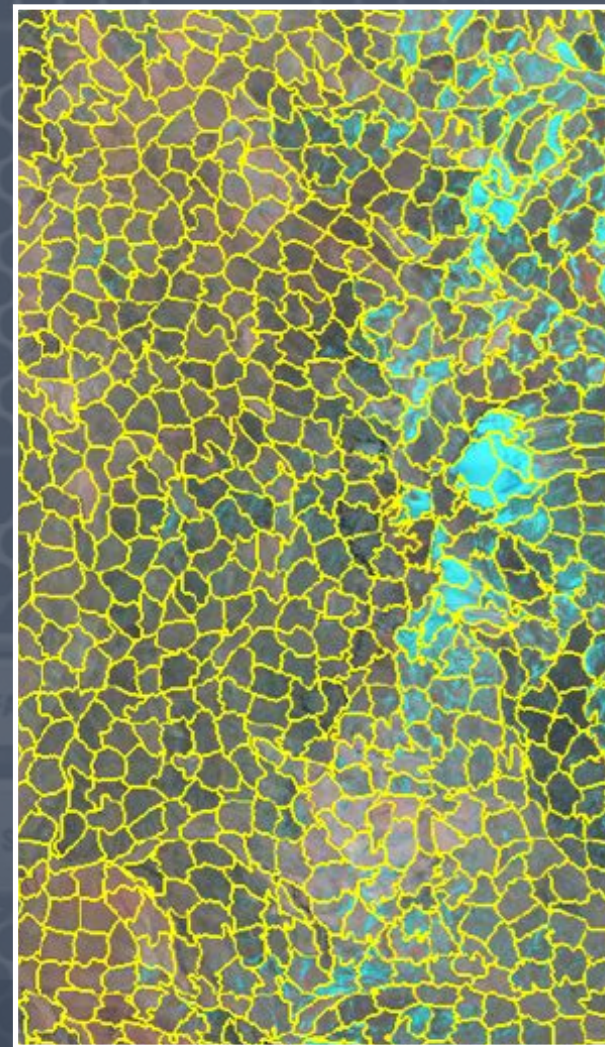
The system should be able to find:

- 1st drillhole positioning
- 5th drillhole positioning
- Target delineation

Scientific hypothesis

Can the system answer:

- Correlation between data (surface / underground)
- To find an optimal drillhole, using estimation with uncertainty?
- It is possible to handle uncertainty?
- Delineate target areas using only raw data?





Hipótesis científicas comprobadas:

- Es posible correlacionar datos de superficie con datos en profundidad o geología existente
- Es posible encontrar un sondaje óptimo
- Es posible manejar la incertidumbre
- Podemos proponer áreas target usando sólo datos crudos



Tipos de datos analizados



Drillhole
Data



Advanced
Geochemistry



Geophysical
Data



Advanced
Geology



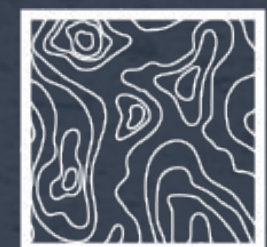
Sensor
Data



RGB
Images



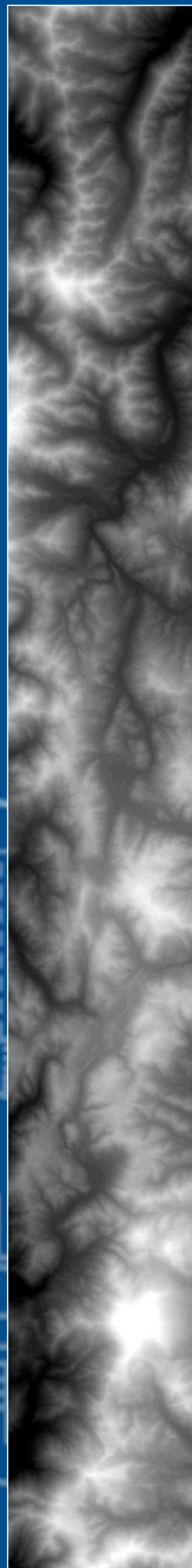
Hyperspectral
data



Topography



RGB



DEM



DEM
Gradient



+14
Bands

ASTER



Módulo

TD

Data fusion

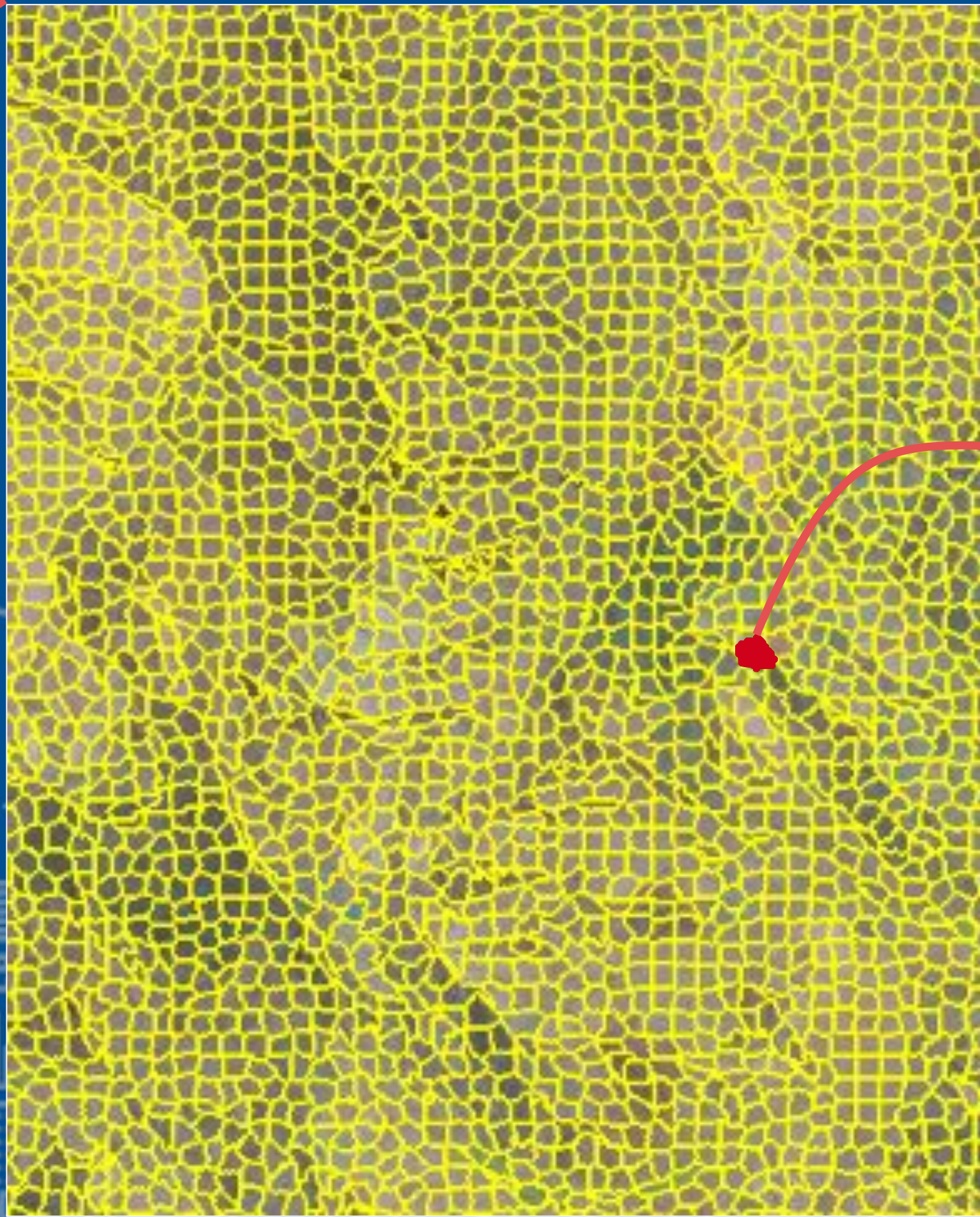
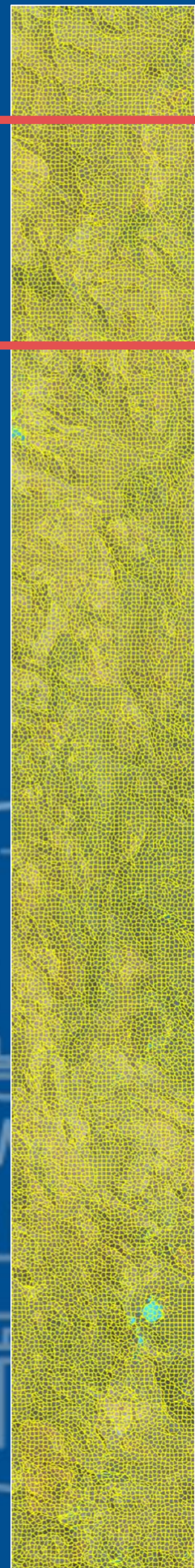
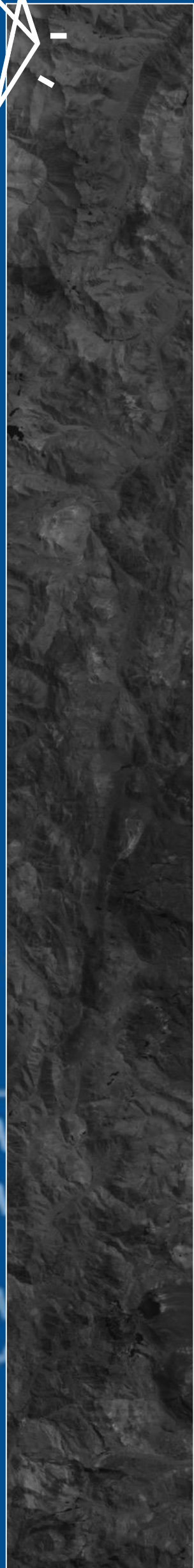
Pixel size:
37.5m

Cada pixel contiene
más de 30 capas:

- Topography
- Multispectral (14 bands)
- RGB (3 colors)
- Topografía
- Estructuras



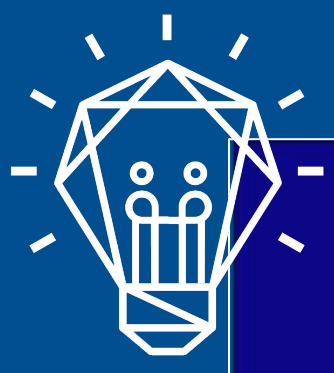
Oversegmentation



Super pixel size:
82.3 pixels (avg)
~115 000 m²

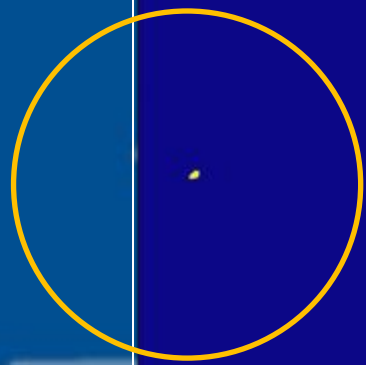
Multibanda

Segmentación

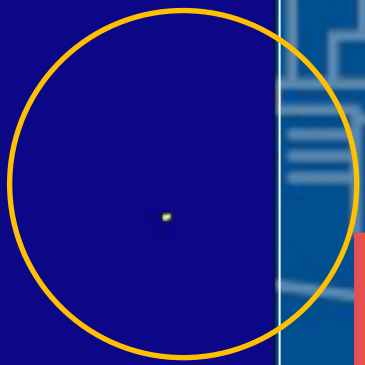


Fertility Regression

Depósito 2

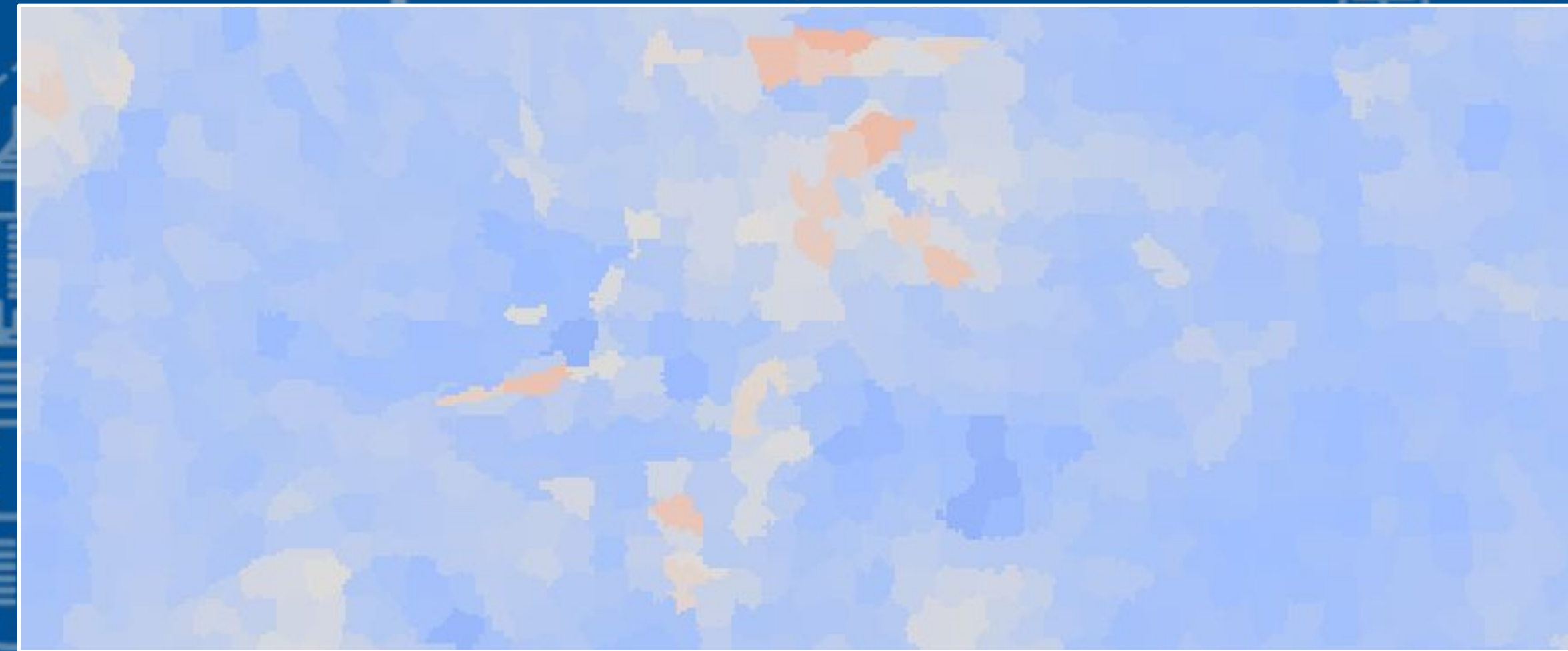


Depósito 1



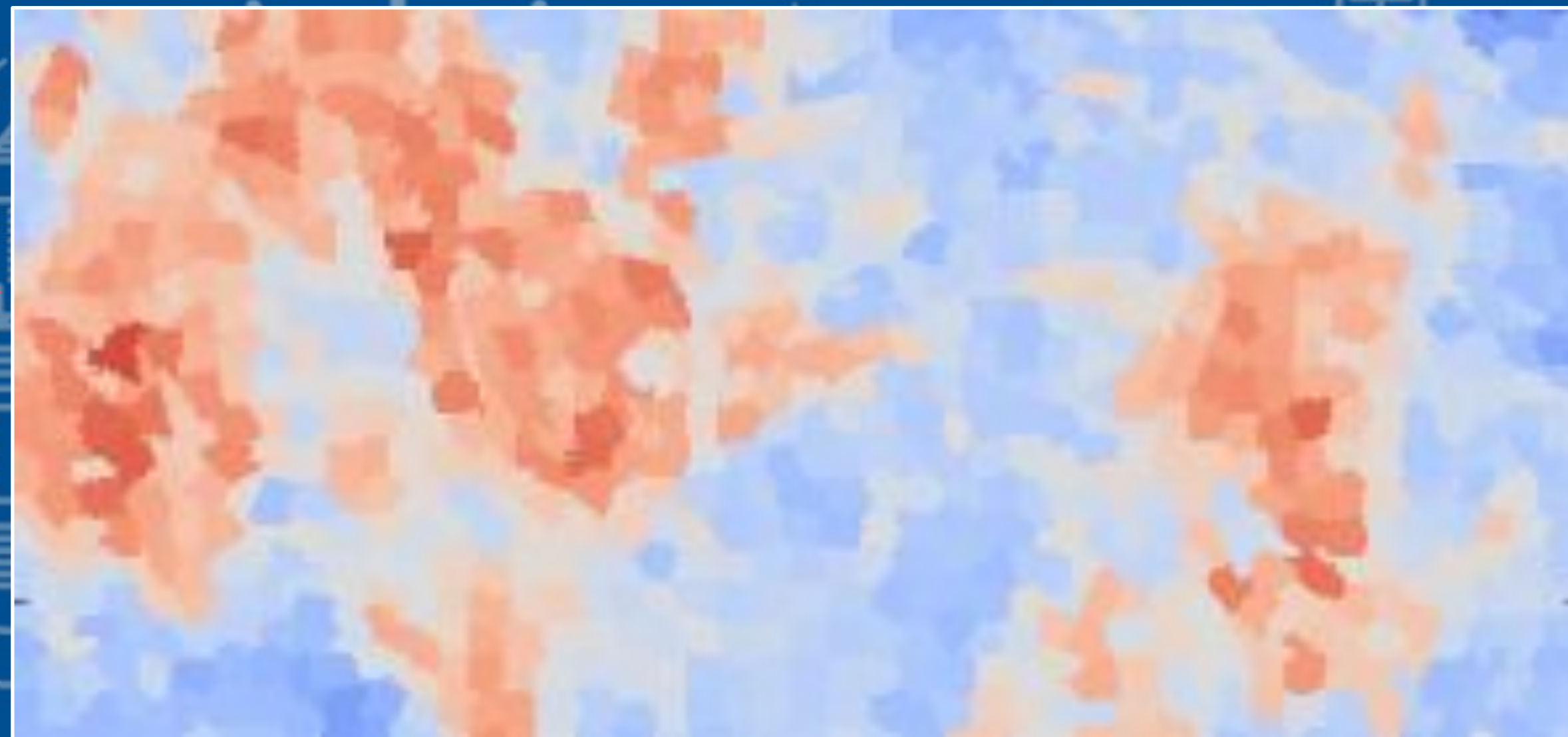
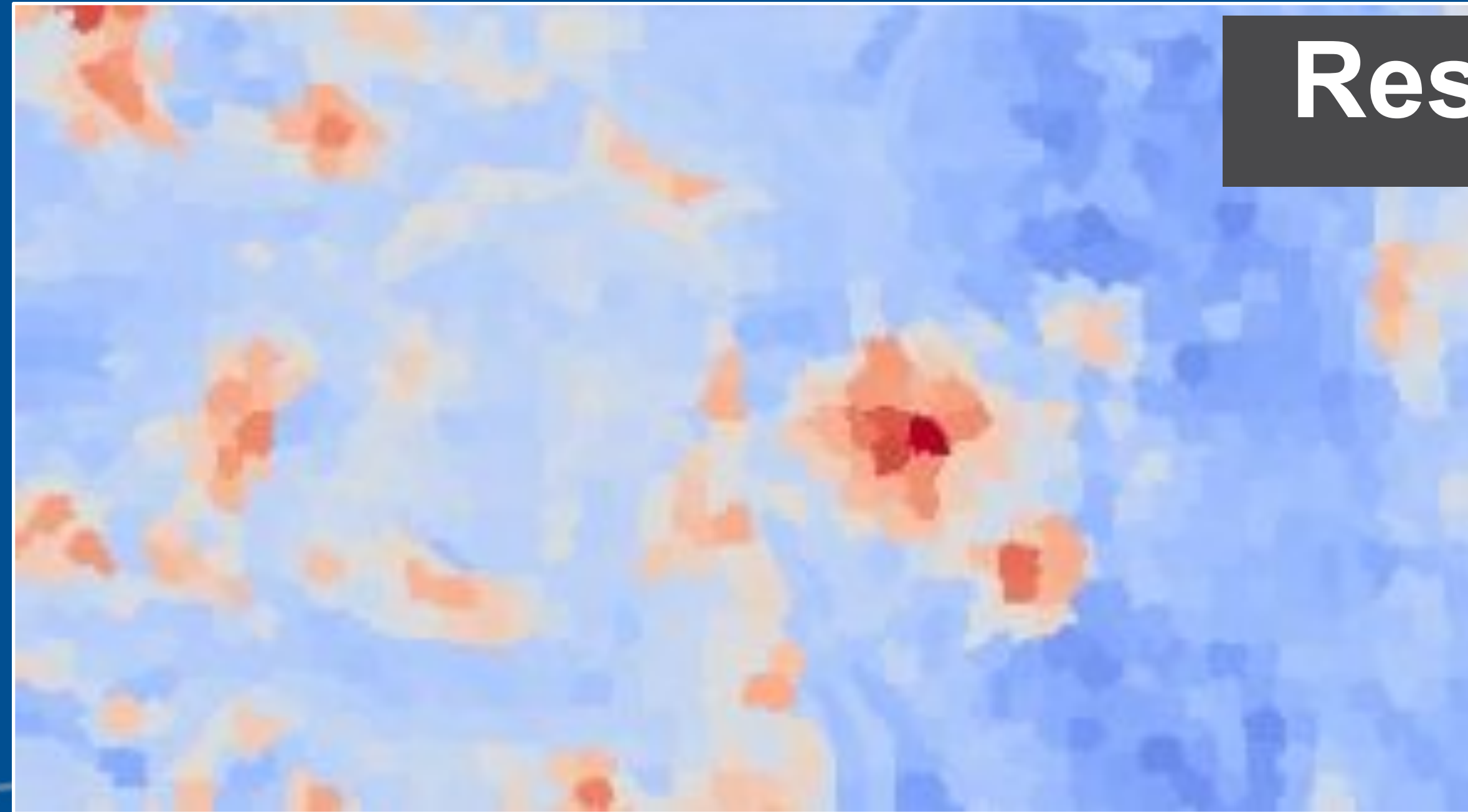
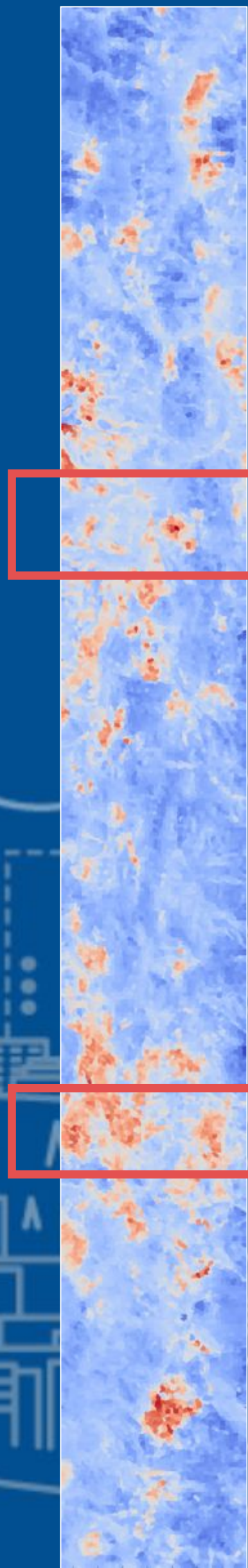
Referencia

Resultado





Resultados TD



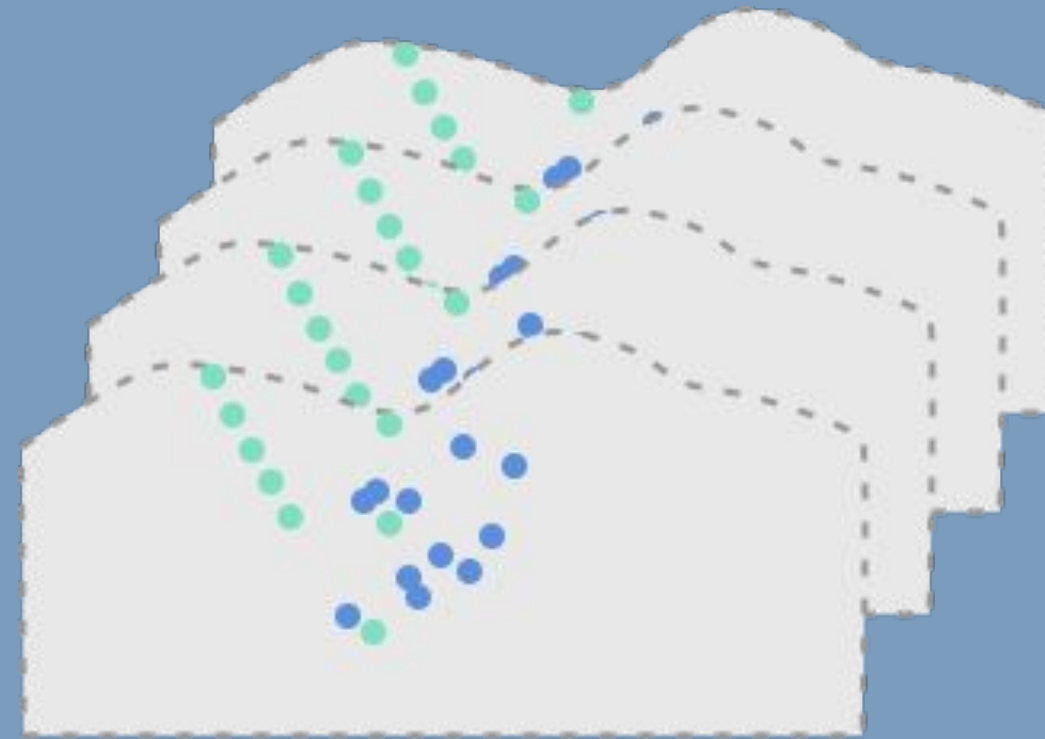
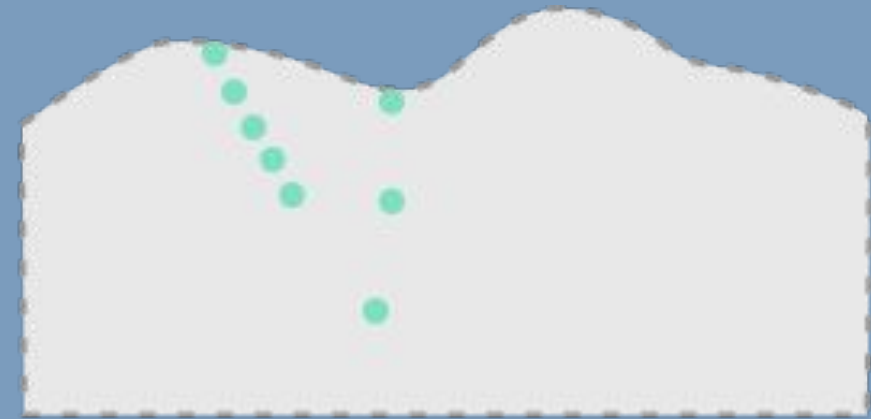
Referencia Resultado



Módulo

DT

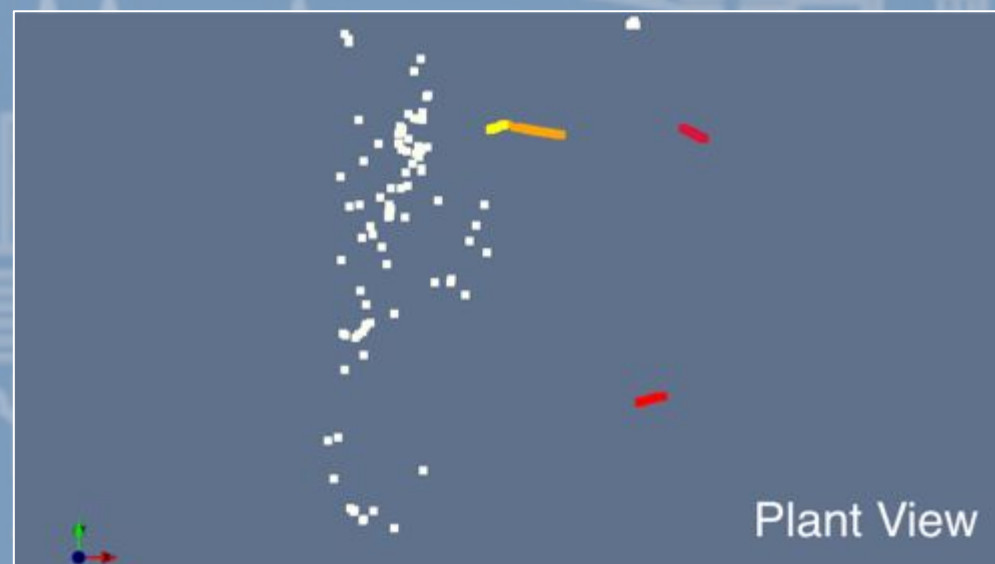
Scenario Simulation



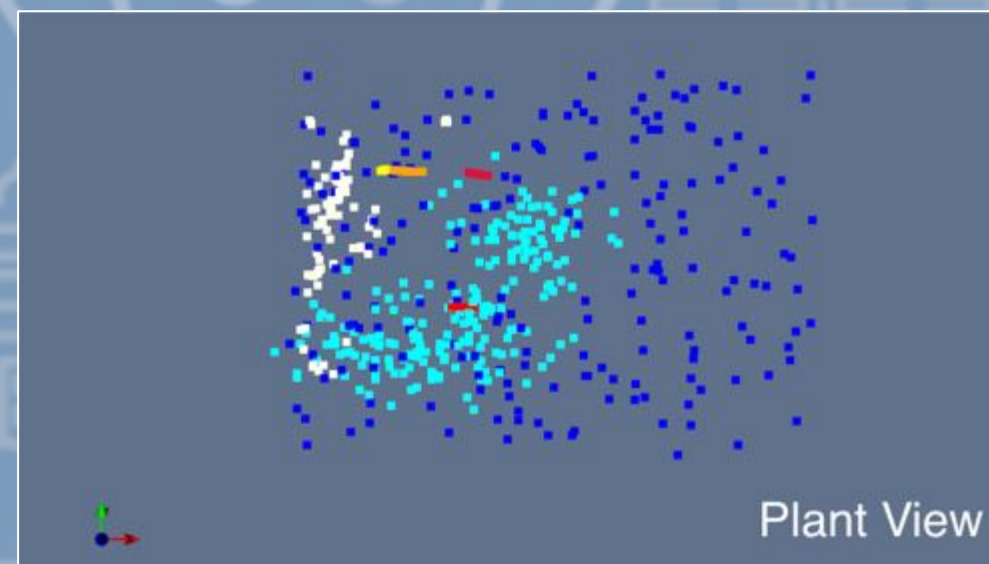
Conditioning data

Simulation of conditioning data

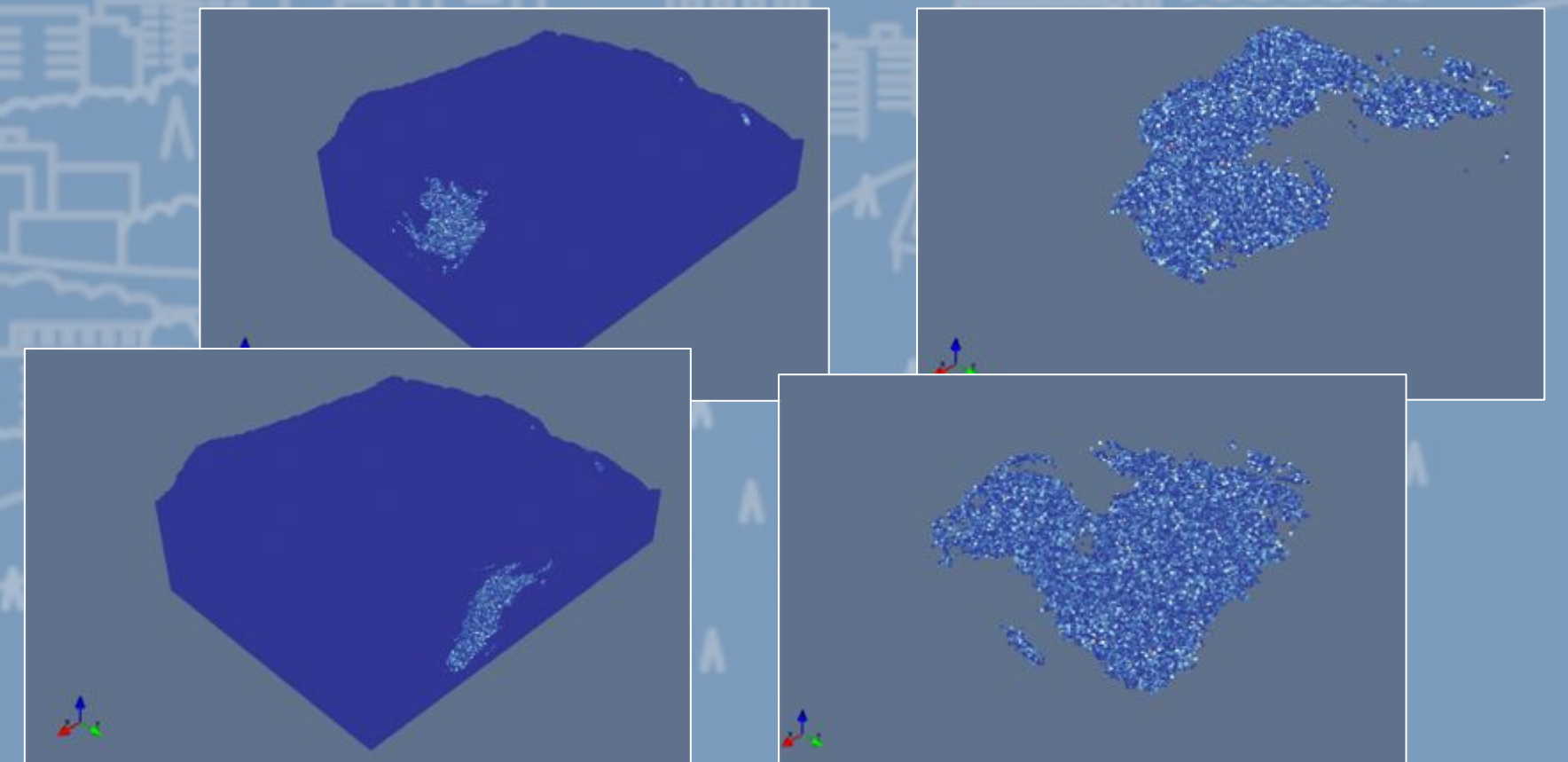
Simulation process



Plant View



Plant View



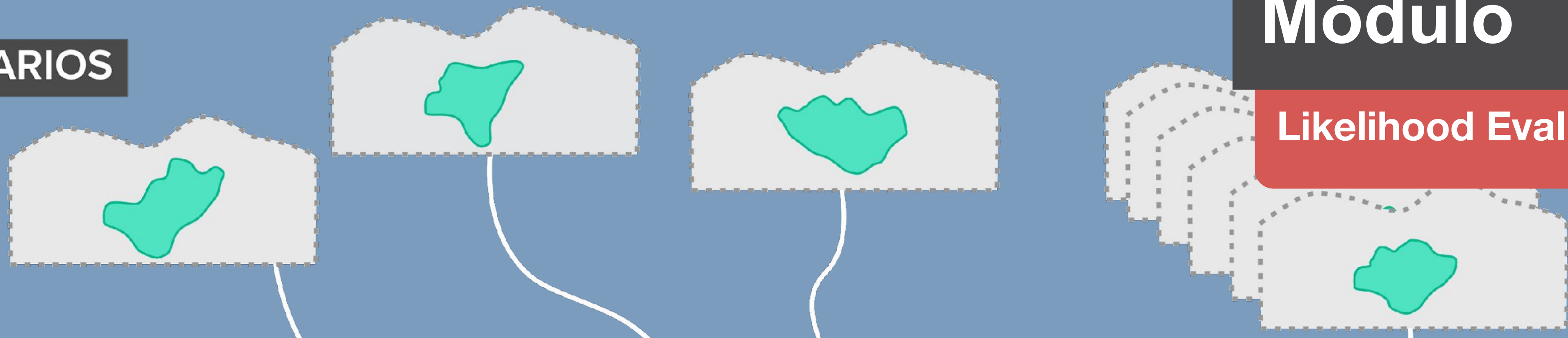


SCENARIOS

Módulo

DT

Likelihood Evaluation

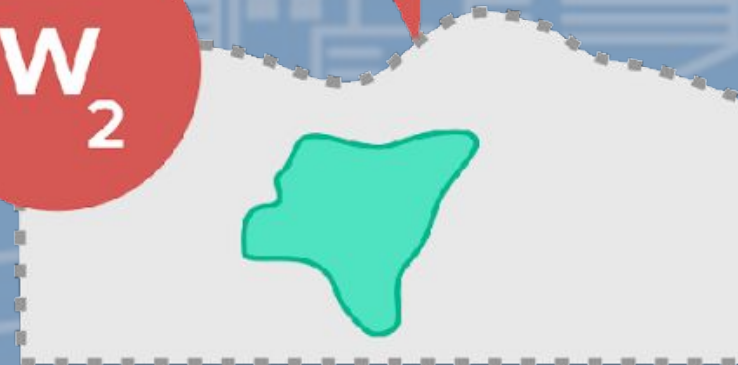


MACHINE LEARNING

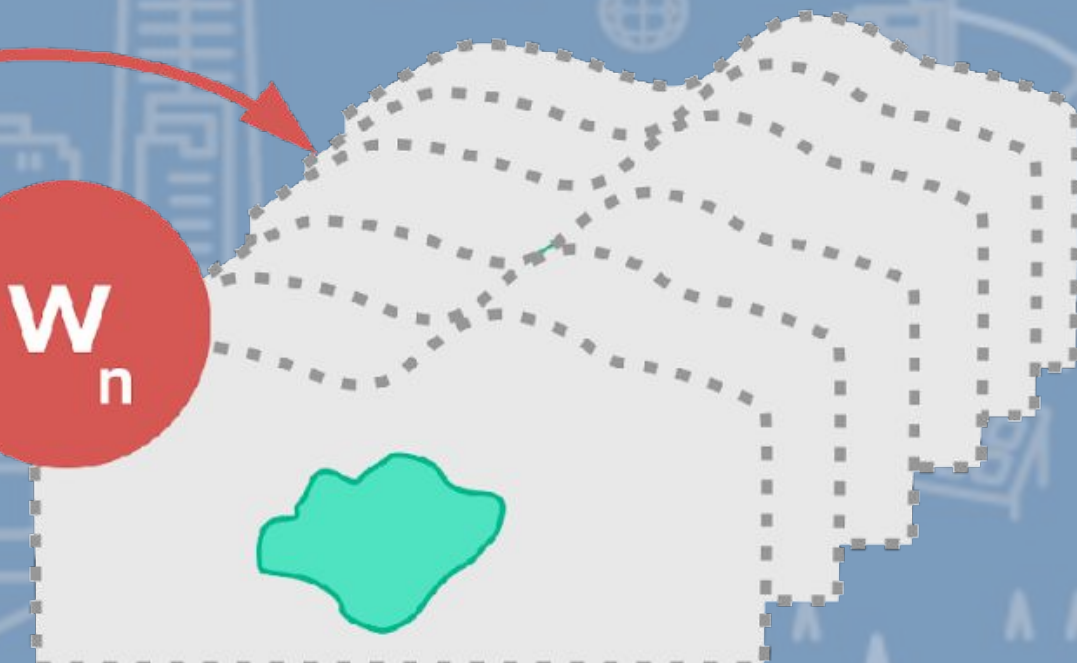
W_1

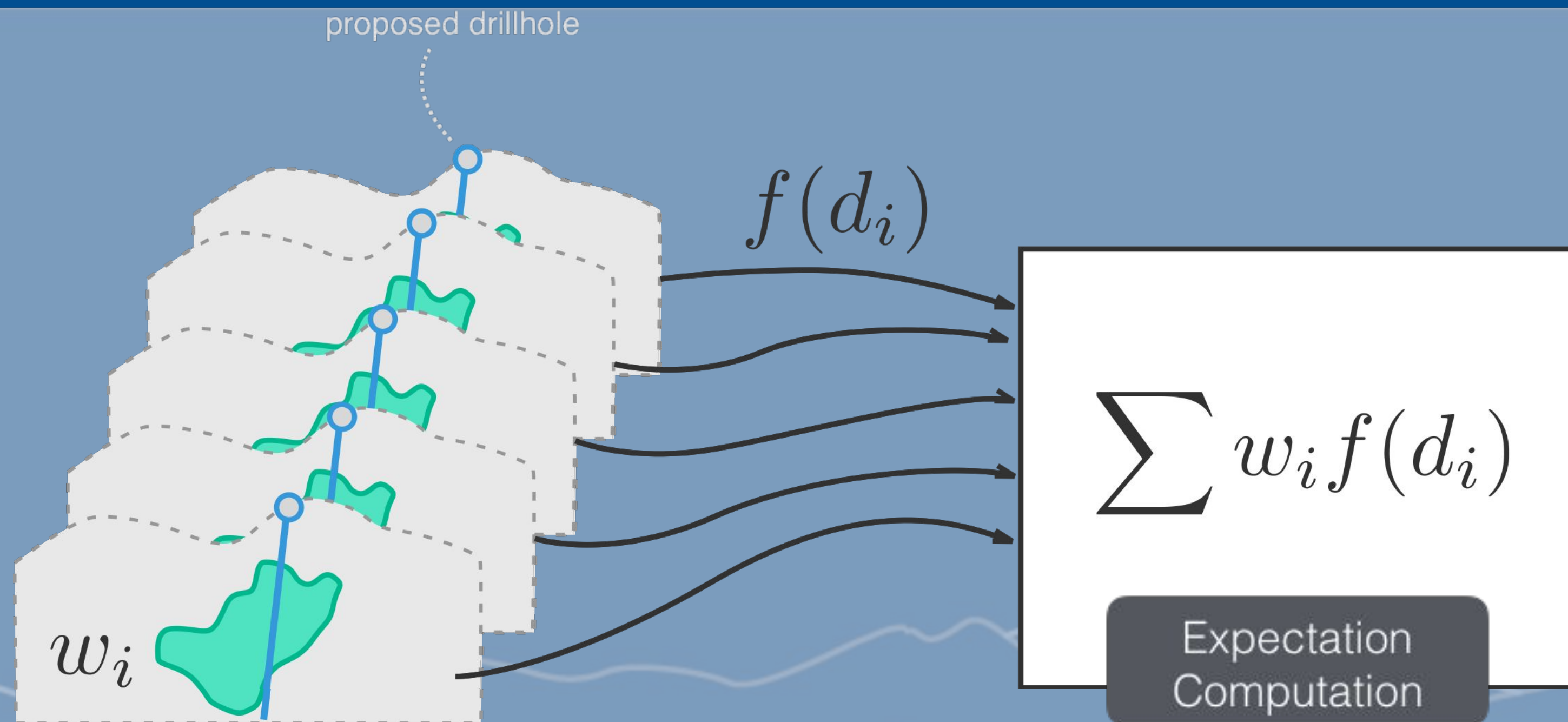


W_2



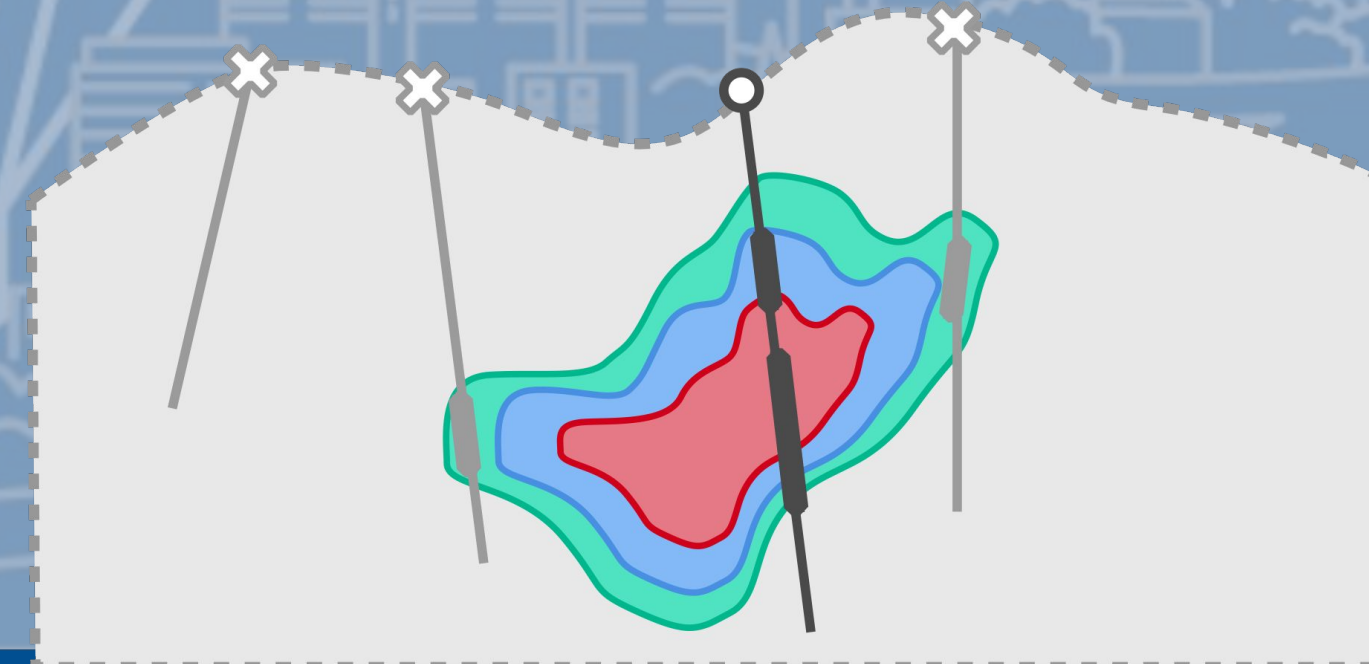
W_n





Target Function Computation

Después de un proceso iterativo se propone el mejor sondaje encontrado

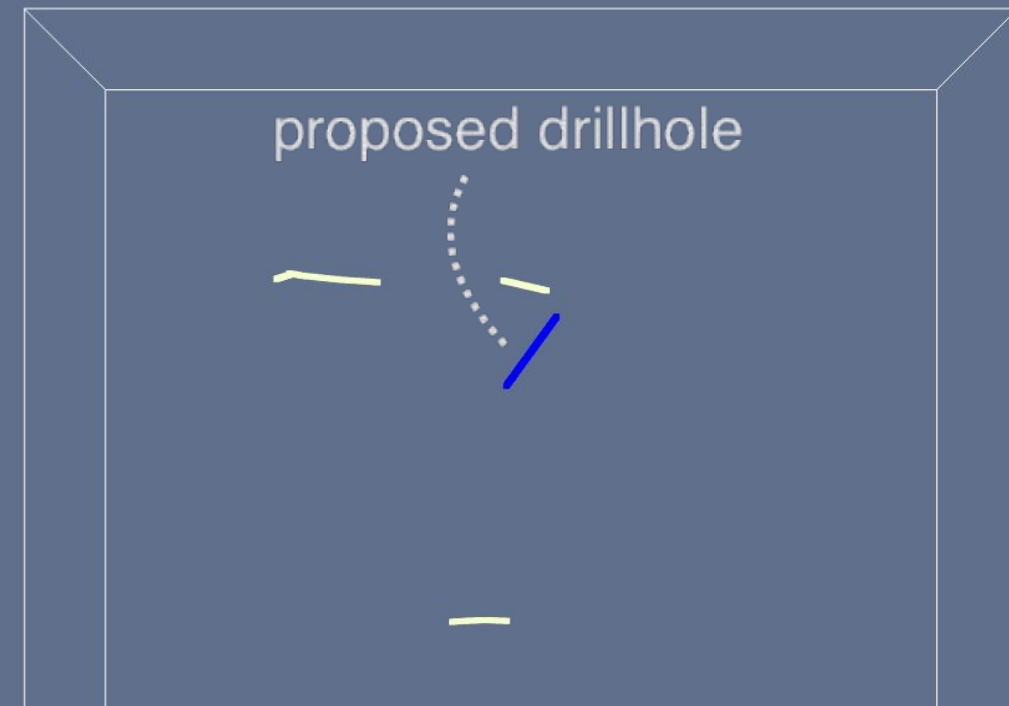




Usando modelo entrenado en deposito A

Resultados

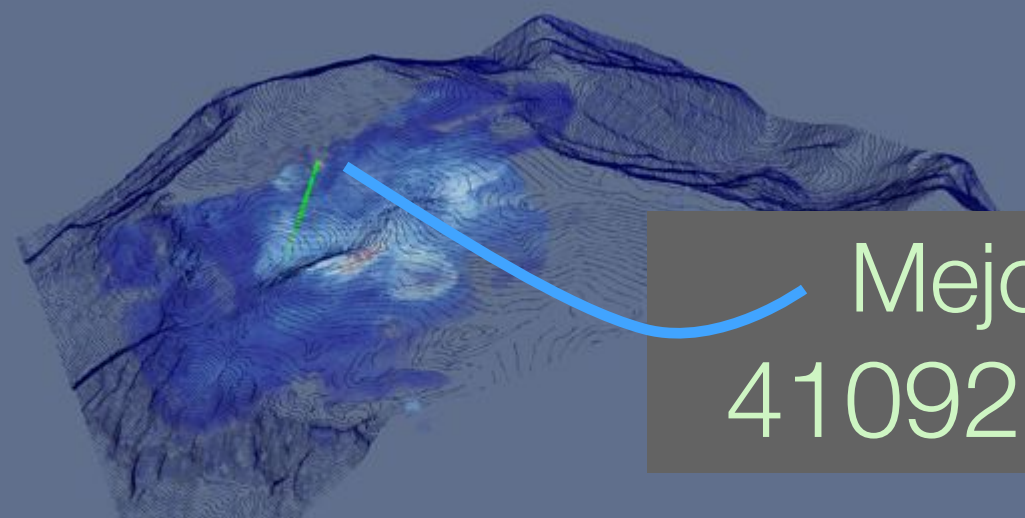
DT



Usando sólo el Deposito A, datos geoquímicos y los primeros 4 sondajes

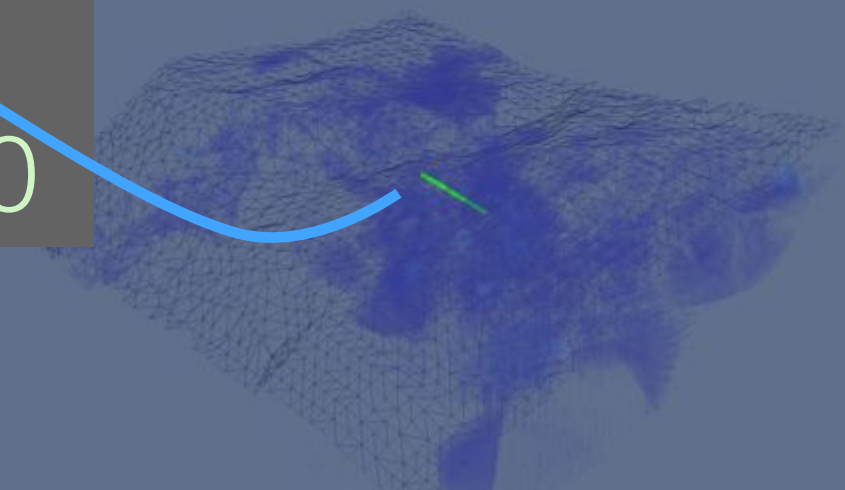


Simulación con sólo **20%** de los datos del Depósito B, datos geoquímicos y los primeros 5 sondajes



Mejor sondaje
410926, 6678989

Mejor sondaje en Deposito A
dado un modelo de bloques



Mejor sondaje
400330, 6755730

Mejor sondaje en Deposito B
en un modelo de bloques generado

Hipótesis científicas comprobadas:

- Es posible usar datos de superficie junto con datos en profundidad (datos geofísicos y algunos sondajes)
- Es posible sugerir potenciales targets de exploración
- Existe una forma de cuantificar incertidumbre y probabilidad
- Integramos el **conocimiento experto** como parte de nuestros modelos

Tipos de datos analizados



Drillhole
Data



Advanced
Geochemistry



Geophysical
Data



Advanced
Geology



Sensor
Data



RGB
Images



Hyperspectral
data



Topography



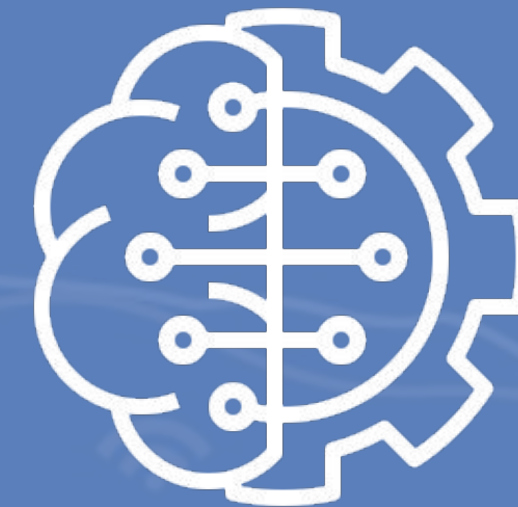
**ANALISIS
DE DATOS**



**PROCESAMIENTO
DE DATOS**



**ANALISIS DE
VARIABILIDAD**



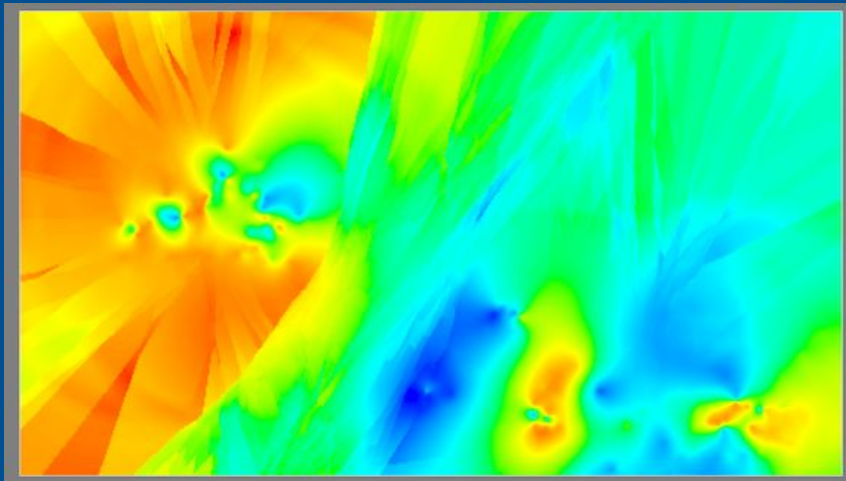
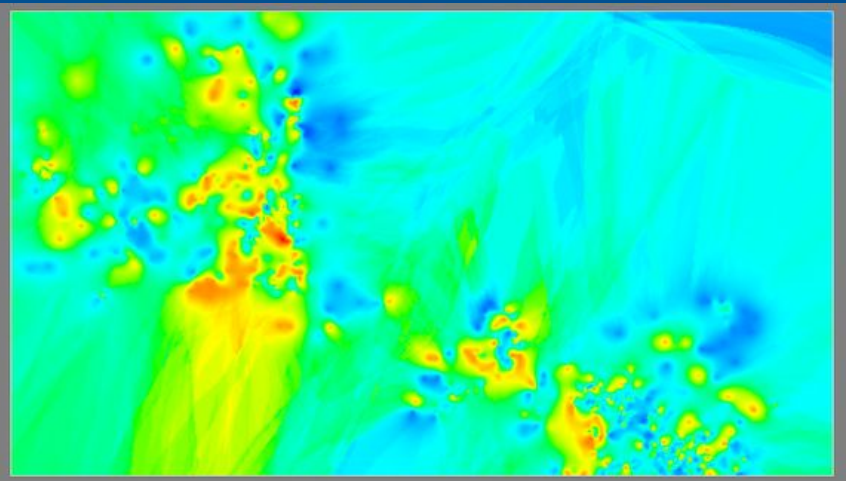
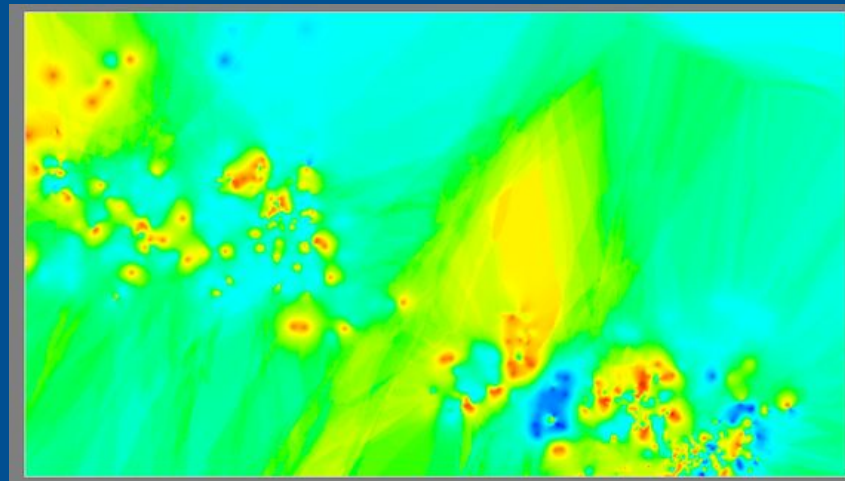
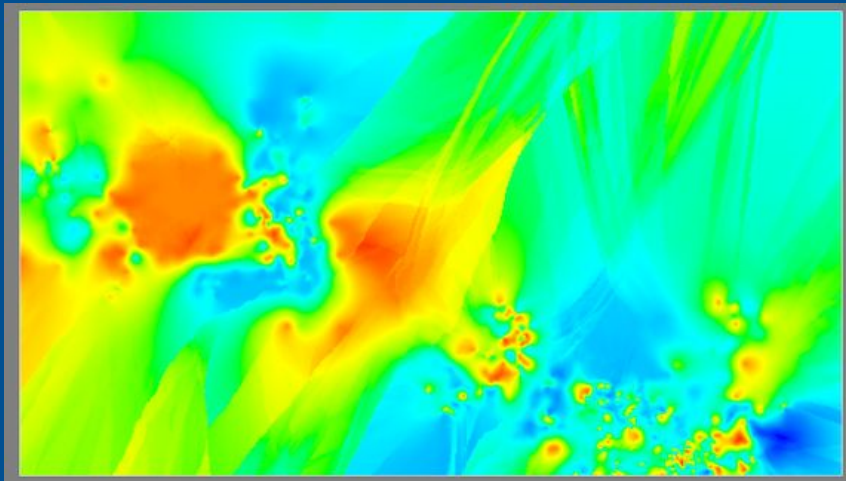
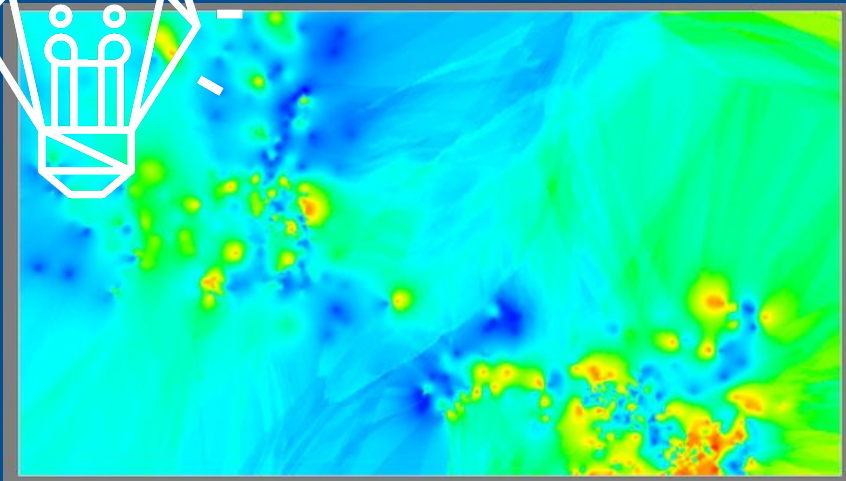
**VARIABLES
SINTETICAS**



**ANALISIS DE
INCERTIDUMBRE**



**MAPA DE
PROBABILIDAD**



Datos Procesados

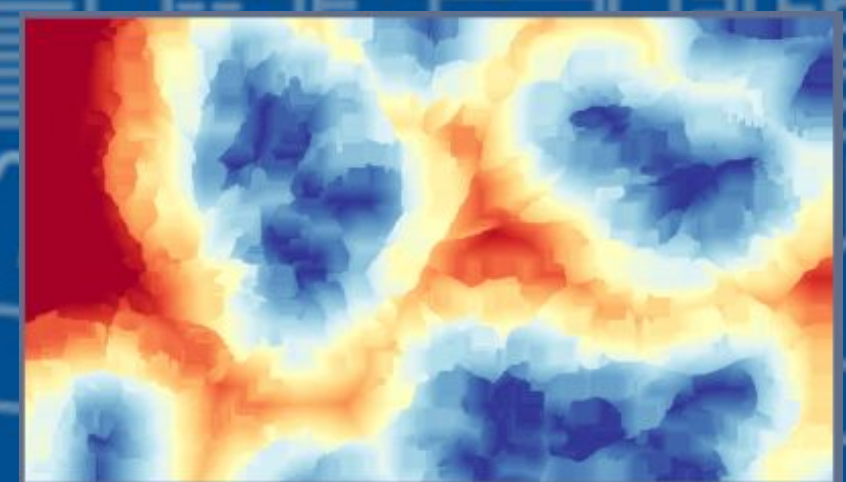
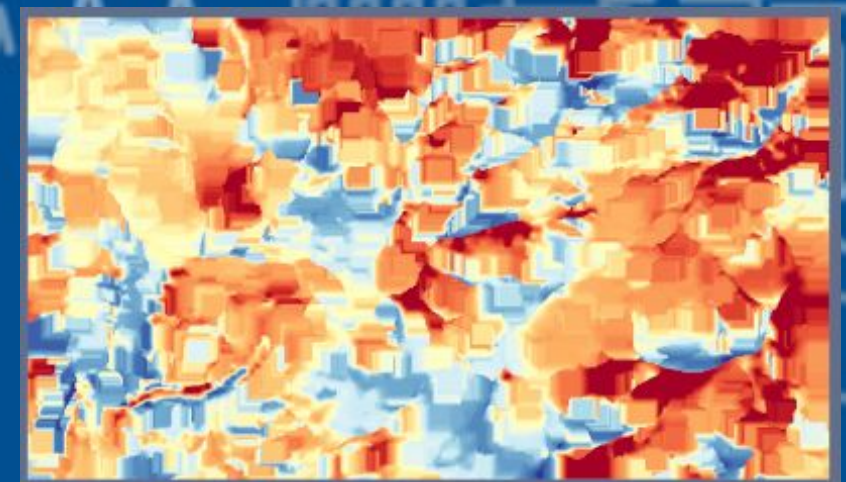
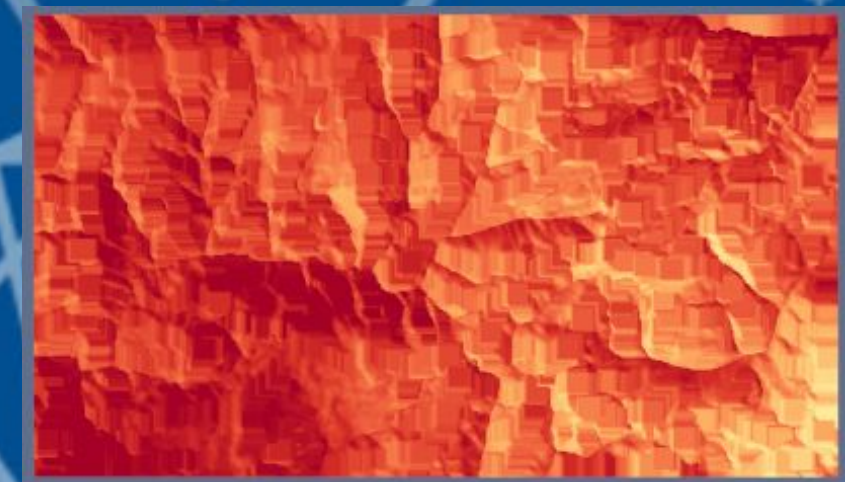
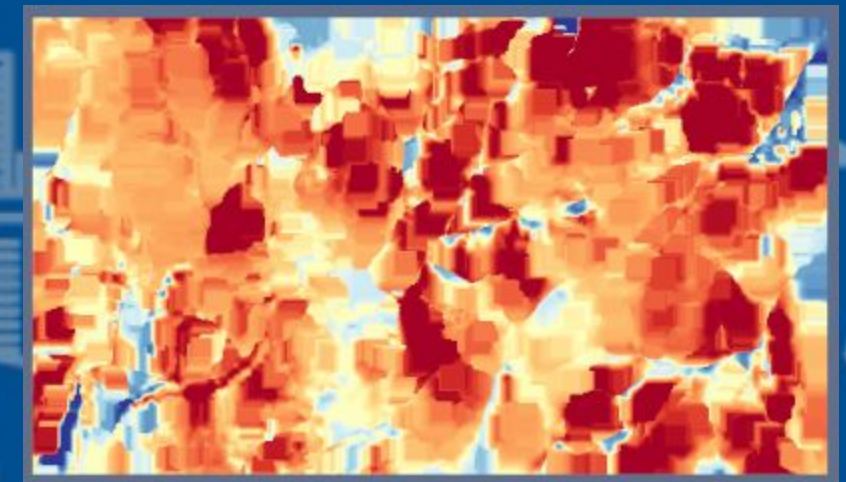
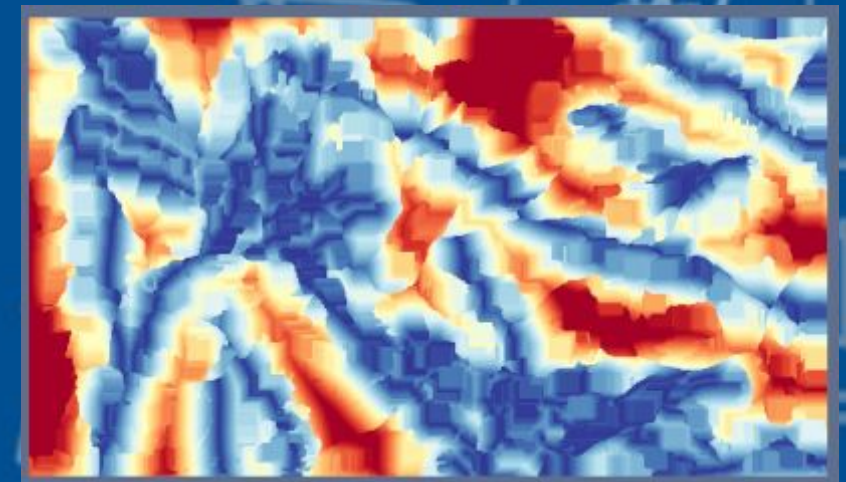


Advanced
Geochemistry



PROCESAMIENTO
DE DATOS

Datos Continuos



Advanced
Geology

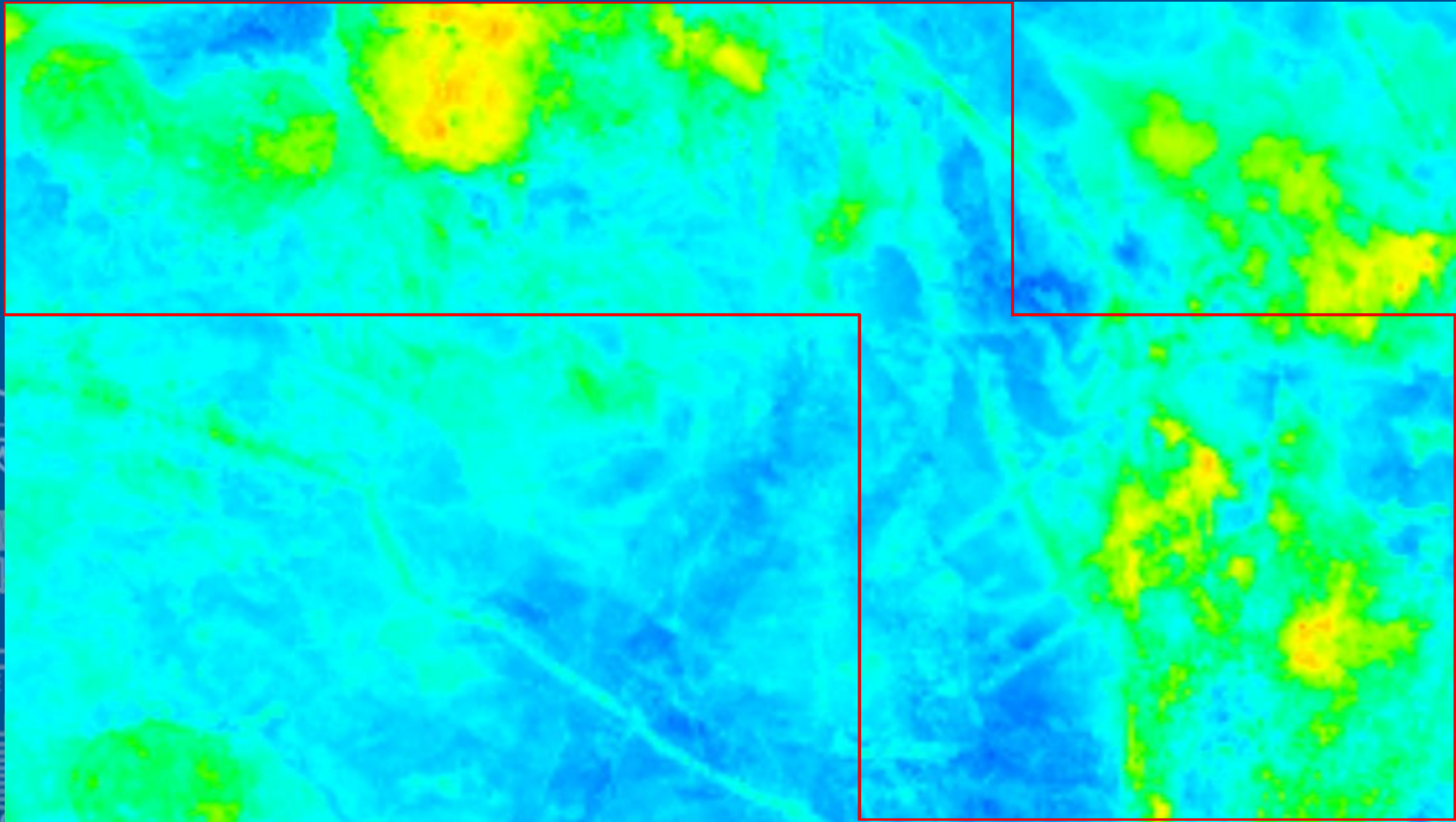
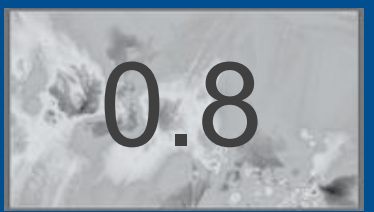
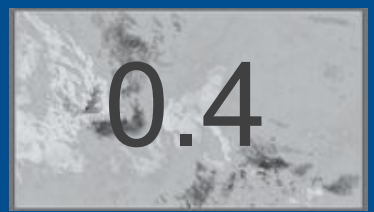
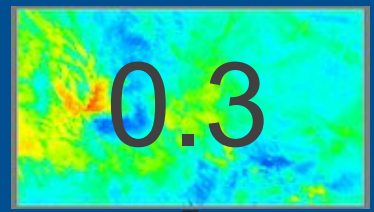
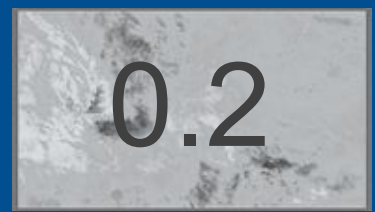
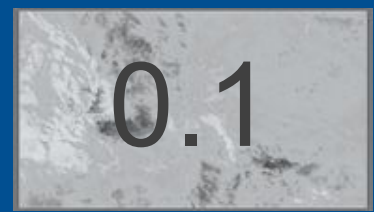
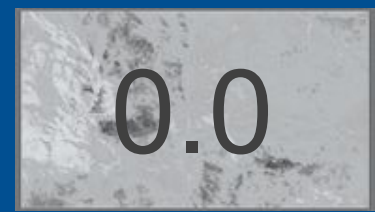


Geophysical
Data



Multispectral
Data

α



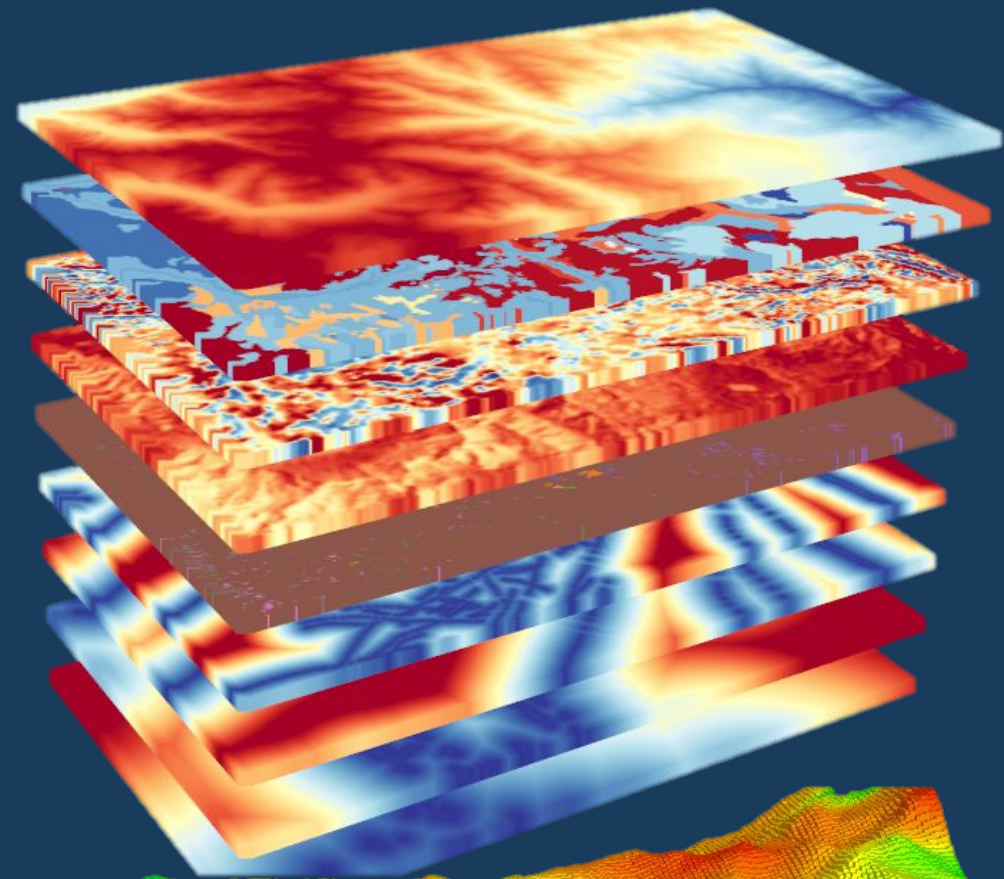
ANÁLISIS DE VARIABILIDAD

Advanced Geochemistry

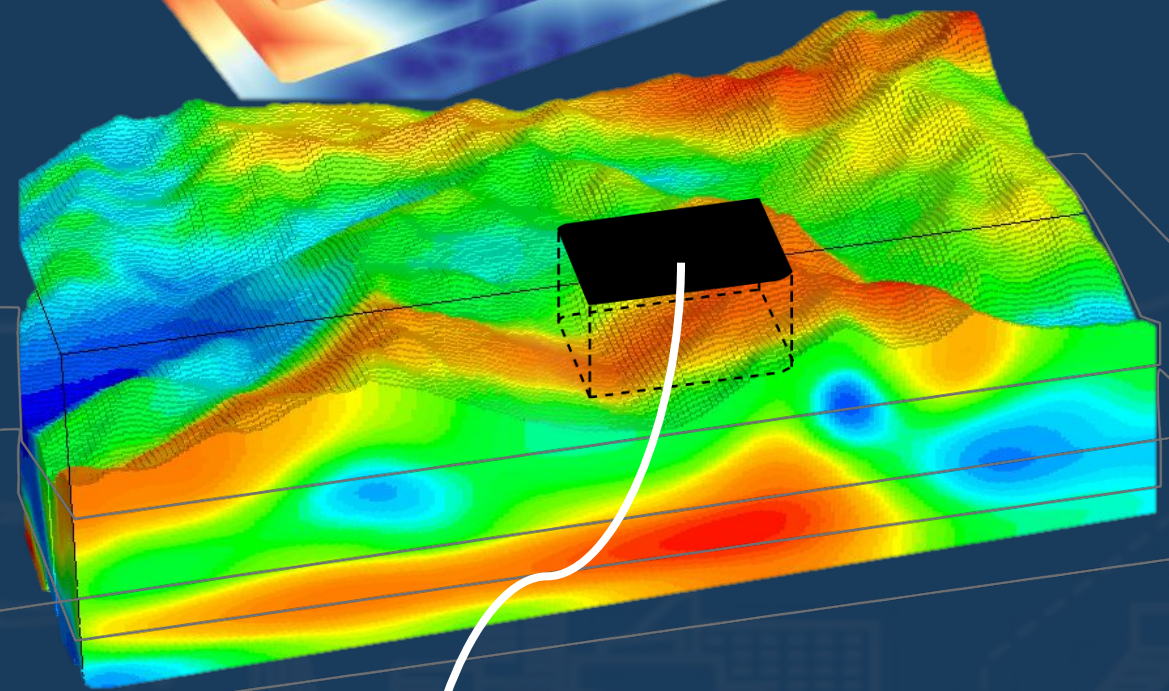
ANÁLISIS DE VARIABILIDAD | Mo



$\vec{\theta}$



Ψ



h

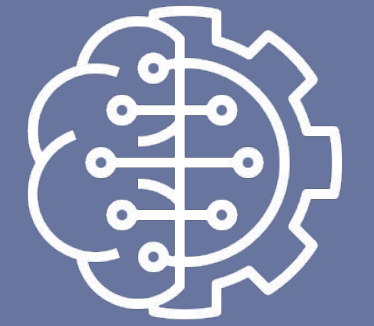
Y_{vs}

Modelo de bloques
geoestadístico usando
sondajes



$$\tilde{Y}_{vs}^h = f_h(\vec{\theta}; \Psi_h)$$

El modelo ML de la *variable sintética* por cota, usando información de la superficie, modelo geoestadístico y geofísica



VARIABLE
SINTETICA

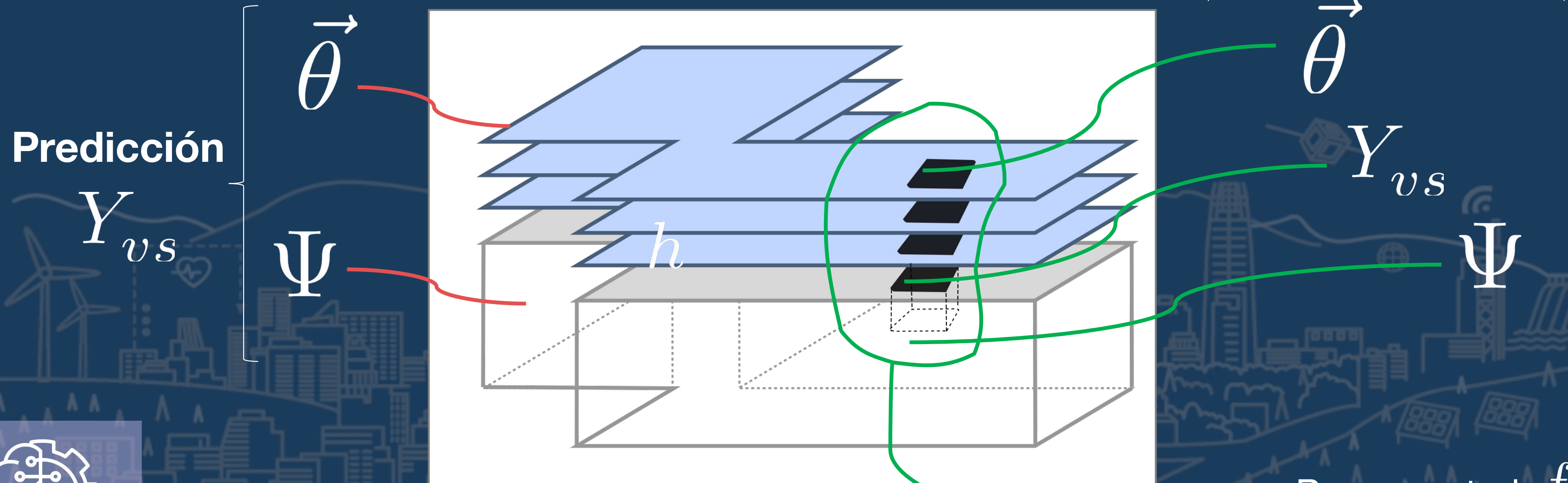
Our model:

$$M(Z_{cu}) = \rho(Z_{cu} | \theta_1, \dots, \theta_n)$$



$$\tilde{Y}_{vs}^h = f_h(\vec{\theta}; \Psi_h)$$

Training & Test



Predicción

Y_{vs}

$\vec{\theta}$

Ψ

h

$\vec{\theta}$

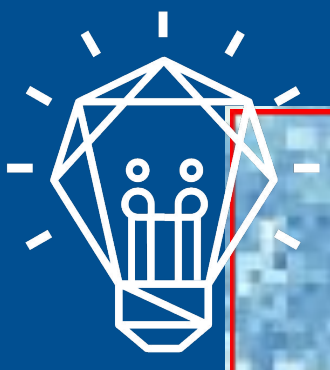
Y_{vs}

Ψ



VARIABLE
SINTETICA

Para construir f_h
dividimos este conjunto en
entrenamiento y test

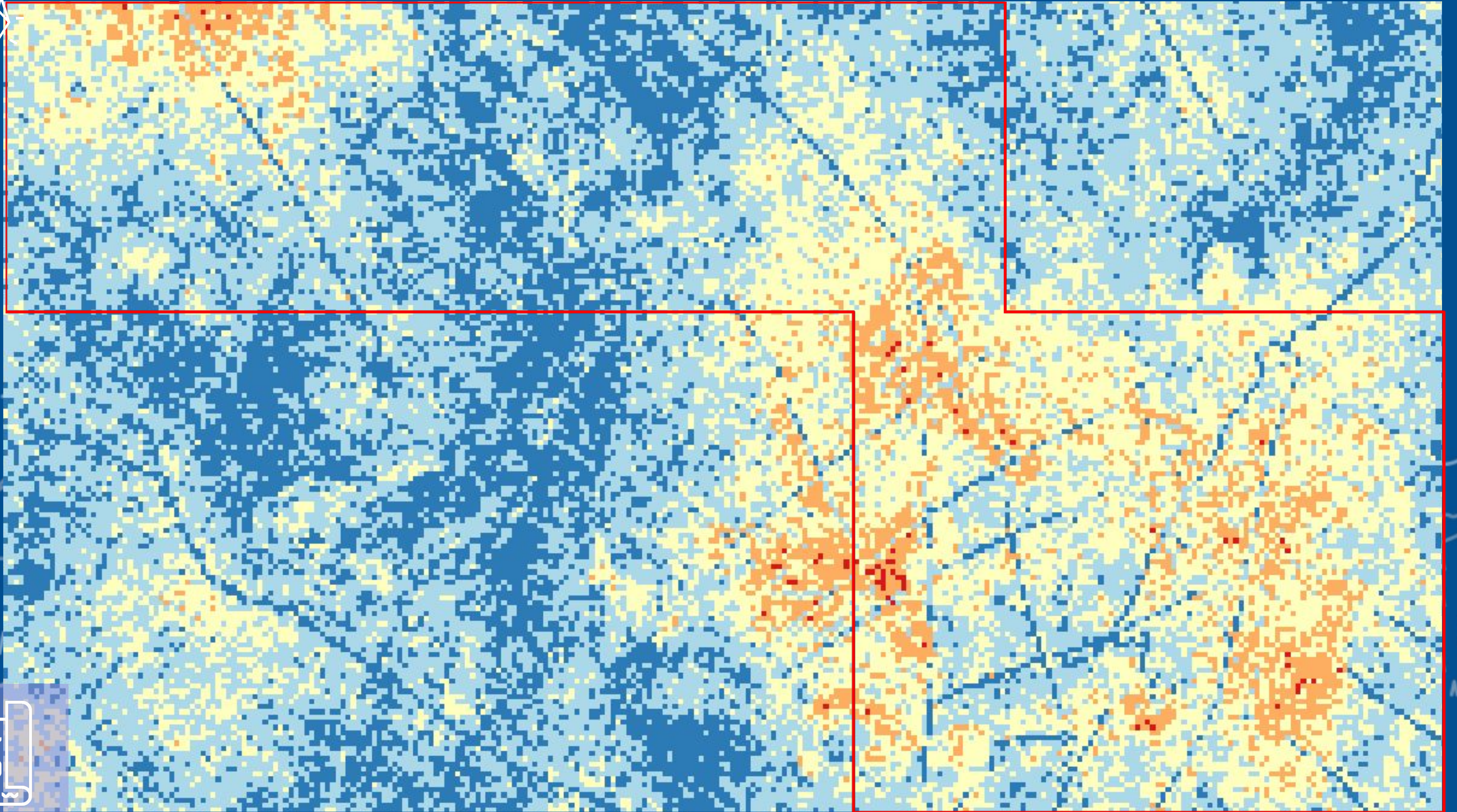
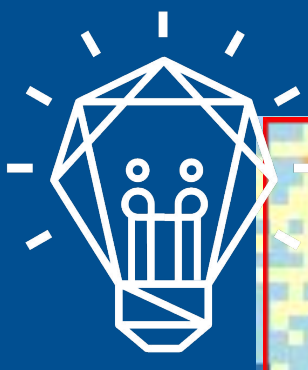


Resultado

- La **variable sintética** fue predicha usando modelos entrenados por cota (en profundidad) y aplicada a toda el área de estudio.
- Se muestra el promedio de esta variable en el área de estudio



SYNTHETIC
VARIABLE

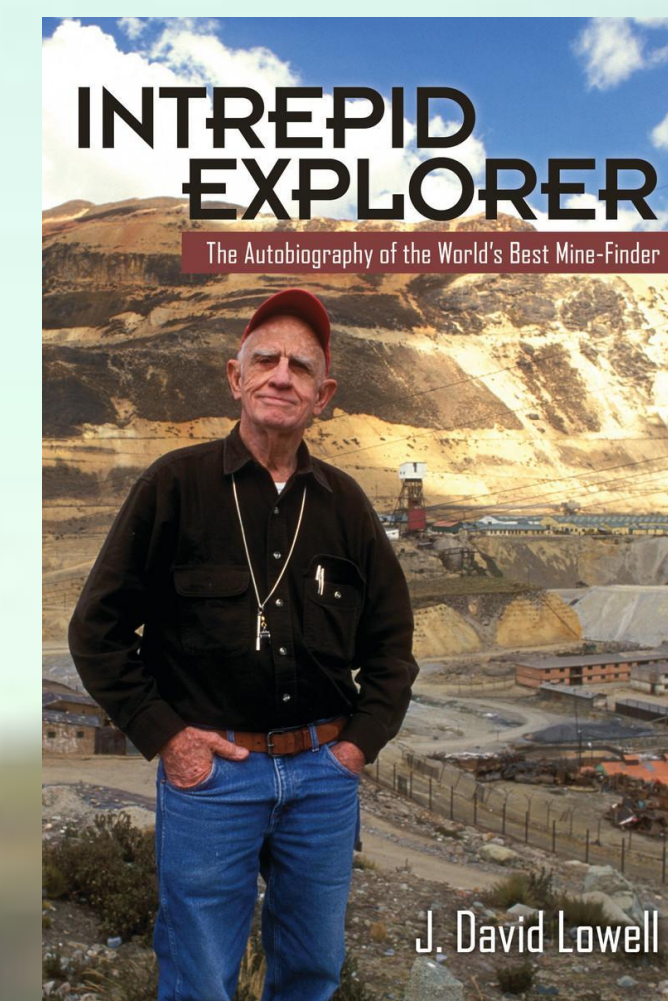


MAPA DE
PROBABILIDAD



*"Mines are found **in the field,**
not the office"*

— Dave Lowell —





fcfm

Ingeniería de Minas
FACULTAD DE CIENCIAS
FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE

ALGES
ADVANCED LABORATORY FOR
GEOSTATISTICAL SUPERCOMPUTING



amtc
ADVANCED MINING TECHNOLOGY CENTER

¿Preguntas?

MACHINE LEARNING APLICADO A LAS EXPLORACIONES MINERAS