ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

HPC y Supercómputo Aplicado a la Evaluación de Recursos

Julián Ortiz
Álvaro Parra
Exequiel Sepúlveda

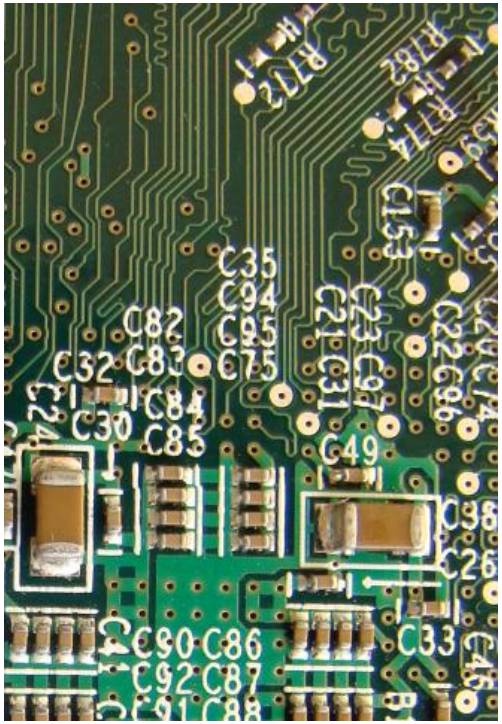
5° Seminario de Acercamiento Tecnológico
Codelco Digital: Minería del Futuro

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

amtc
ADVANCED MINING TECHNOLOGY CENTER

Temas

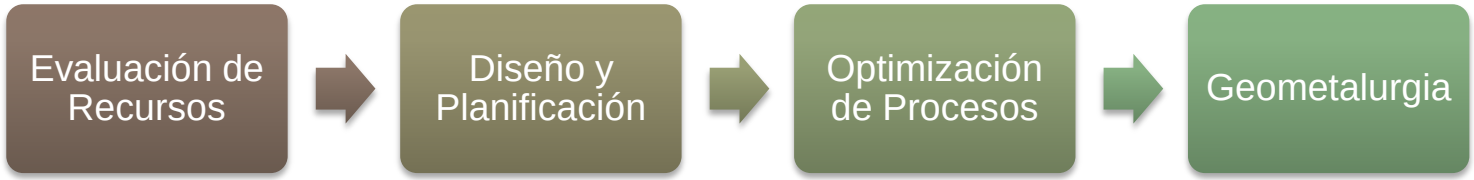
- Contexto
- ALGES – Laboratorio Avanzado de Aplicaciones de Supercómputo para Geoestadística
  - Objetivo
  - Proyectos
  - Equipo
- Proyectos
  - Captura de datos
  - Modelamiento numérico
- SC y HPC
- Aplicaciones
- Conclusiones



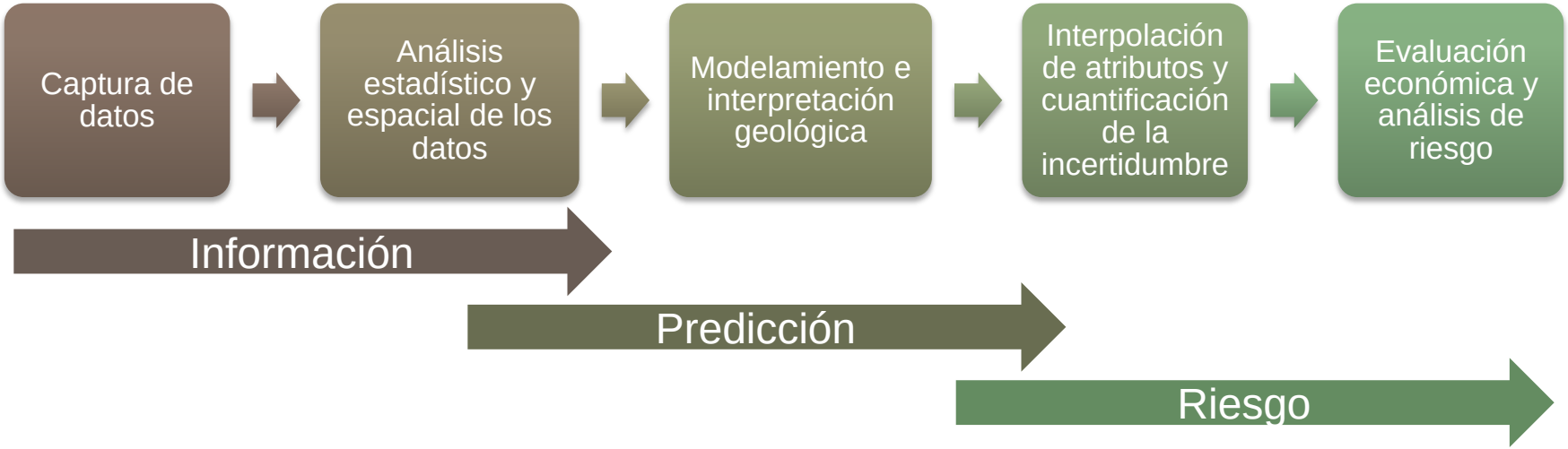
ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

Contexto

- Evaluación de recursos es clave en etapas posteriores de proyectos mineros



- La cuantificación de recursos requiere de las siguientes etapas:



ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

ALGES – Objetivo

Laboratorio Avanzado para Aplicaciones de Supercómputo en Geoestadística

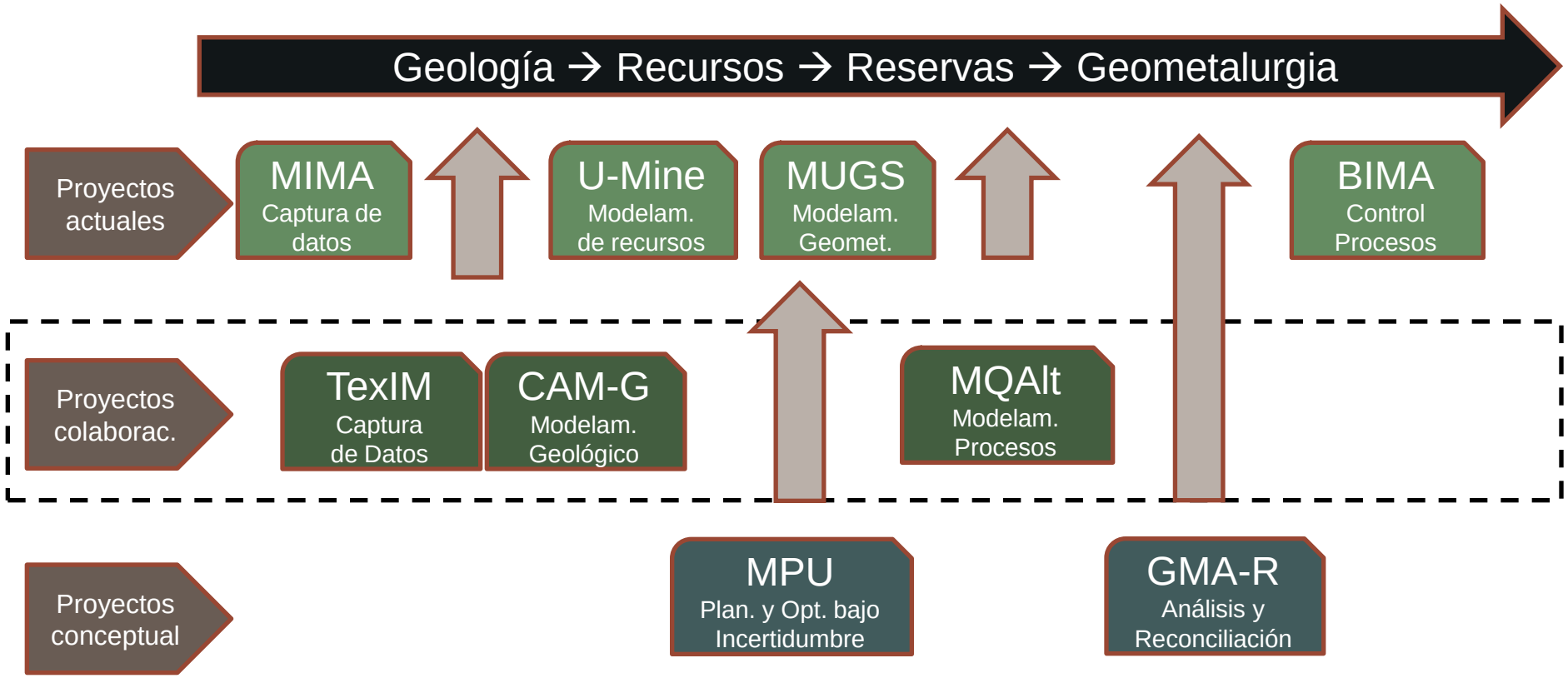
SC / HPC	Data Mining	Estadística	Geoestadística
Análisis de Imágenes	Modelos Estocásticos	Desarrollo de Software	Transf. Tecnológica

- El objetivo del laboratorio ALGES es desarrollar las herramientas y modelos para la caracterización de fenómenos geológicos
- Tareas
  - Predicción de variables categóricas y continuas
  - Cuantificación de incertidumbre para toma de decisiones
  - Modelamiento multivariable (estimación y simulación)
  - Desarrollo de software y modelamiento matemático:
    - Análisis de imágenes de datos geológicos, mineros y metalúrgicos
    - Supercómputo para modelamiento geoestadístico
    - Optimización y toma de decisiones
- AMTC – Centro Avanzado de Tecnología para la Minería

ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

ALGES – Proyectos

- La visión es entender, modelar e integrar todas las variables relevantes al negocio → modelos geológicos, geotécnicos y geometalúrgicos.



ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

ALGES – Equipo

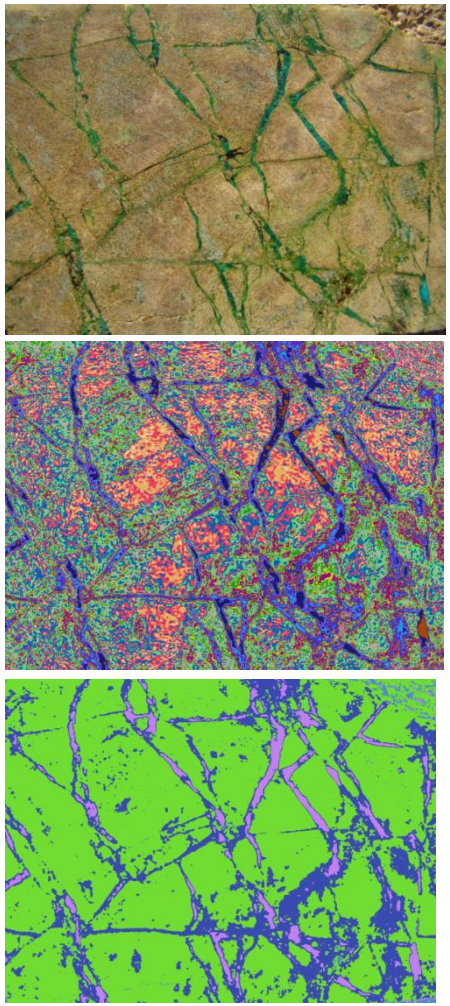
- 3 académicos jornada completa
- 1académico jornada parcial
- 3 investigadores asociados
- 3 tesistas de Magíster
- 8 estudiantes de pregrado

- Fuerte vínculo industrial
- Equipo de investigadores multidisciplinario con mucha experiencia:
  - **Ciencias de la Computación** → Supercómputo / HPC / Computación Científica
  - **Tecnologías de la Información** → Administración y Análisis de Datos Corporativos
  - **Astronomía** → Data Mining / Análisis de Imágenes
  - **Matemáticas** → Modelos Estocásticos / Geoestadística
  - **Minería** → Muestreo / Geoestadística / Optimización / Modelamiento Numérico
  - **Metalurgia** → Flotación / Control de Procesos / Modelamiento de Proceso

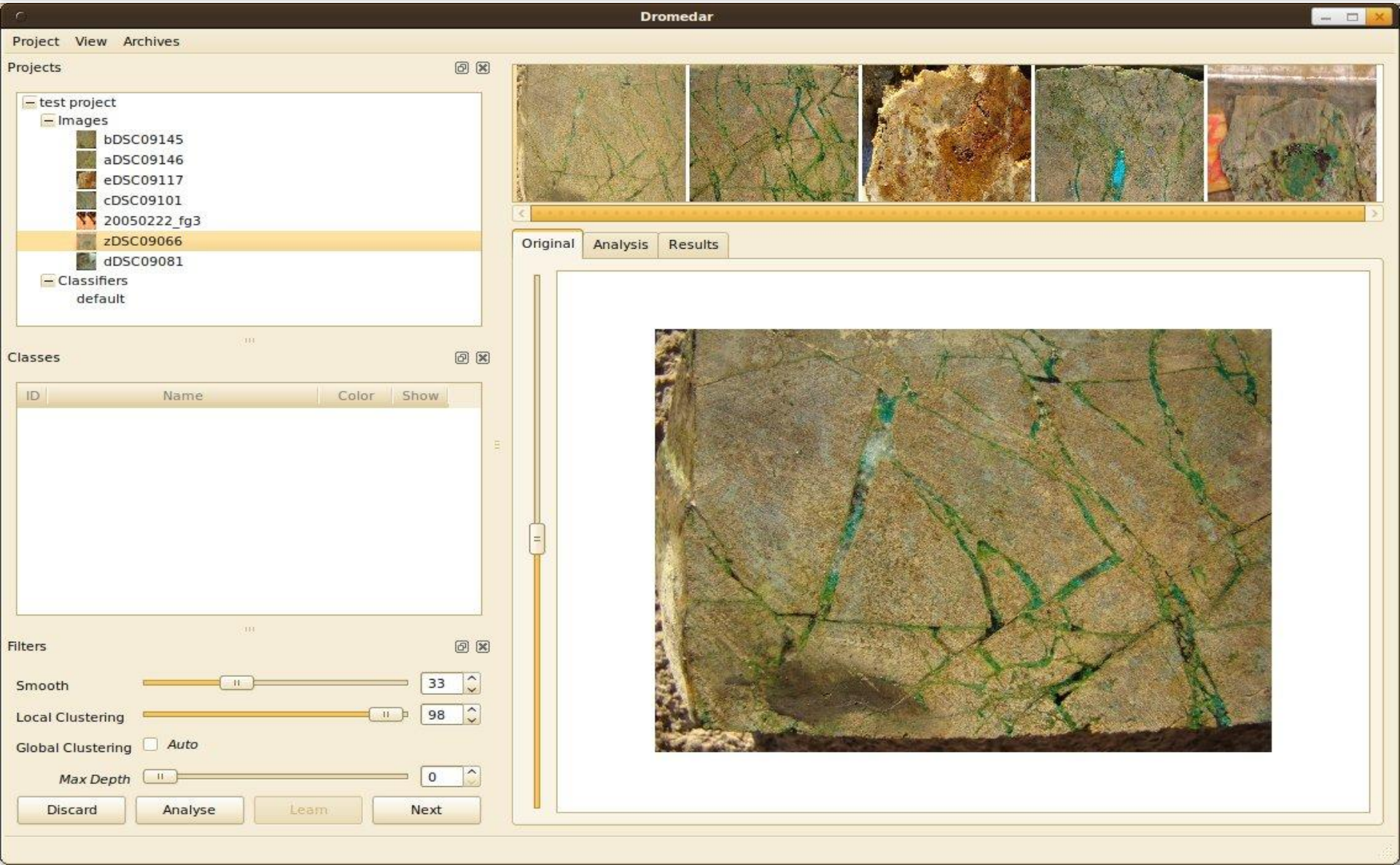
ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

Captura de datos

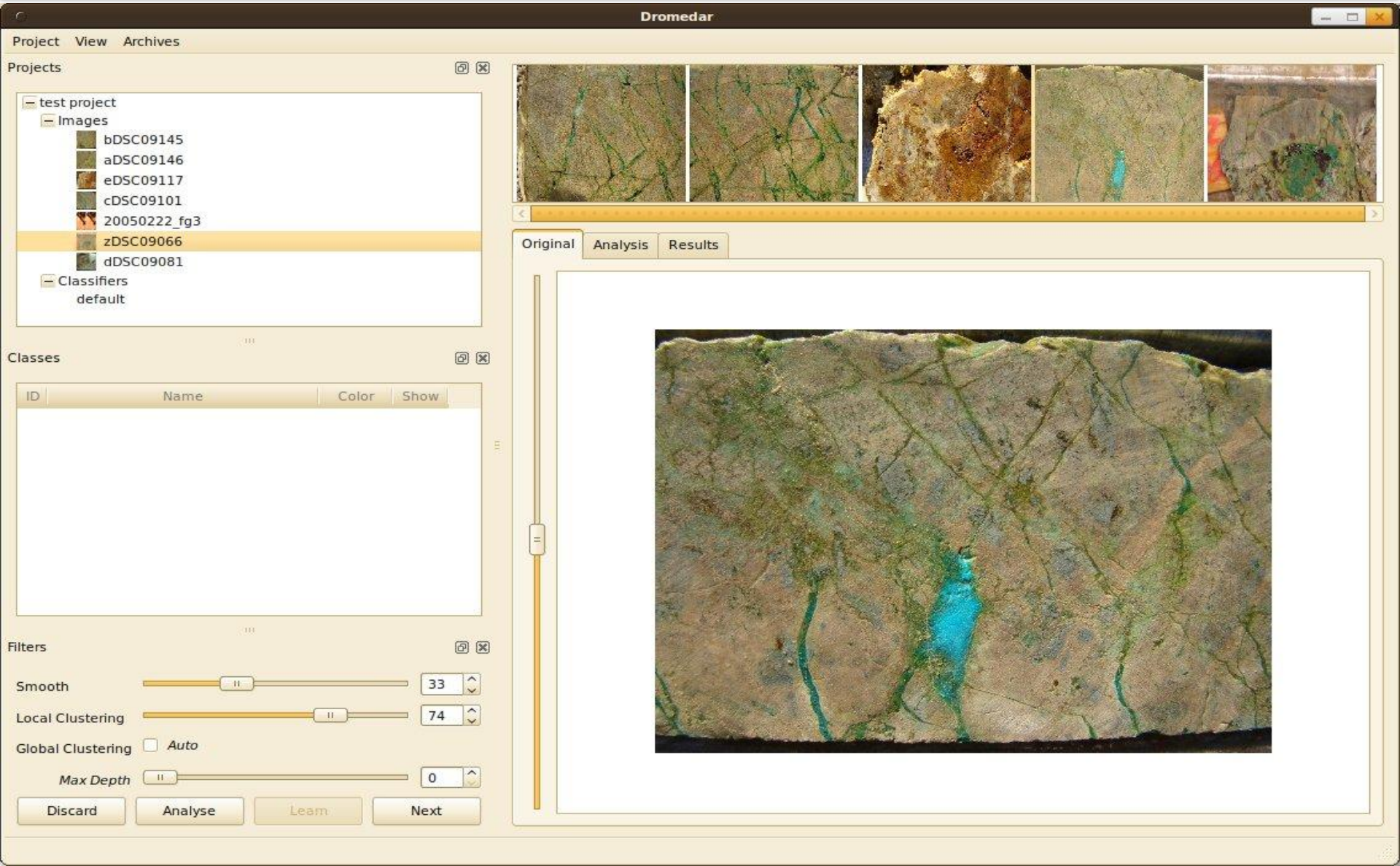
- Captura de datos
  - Información de sondajes es la base de la construcción de modelos de recursos
  - Mapeo geológico permite identificar / cuantificar variables → información es subutilizada
  - Idea: incorporar mayor información a los modelos
    - Geo-referenciada
    - Cuantitativa
    - Rápida
  - Sistema semi-supervisado de captura de información de mineralogías mediante análisis de imágenes.
    - Extensión a alteraciones, litologías, texturas



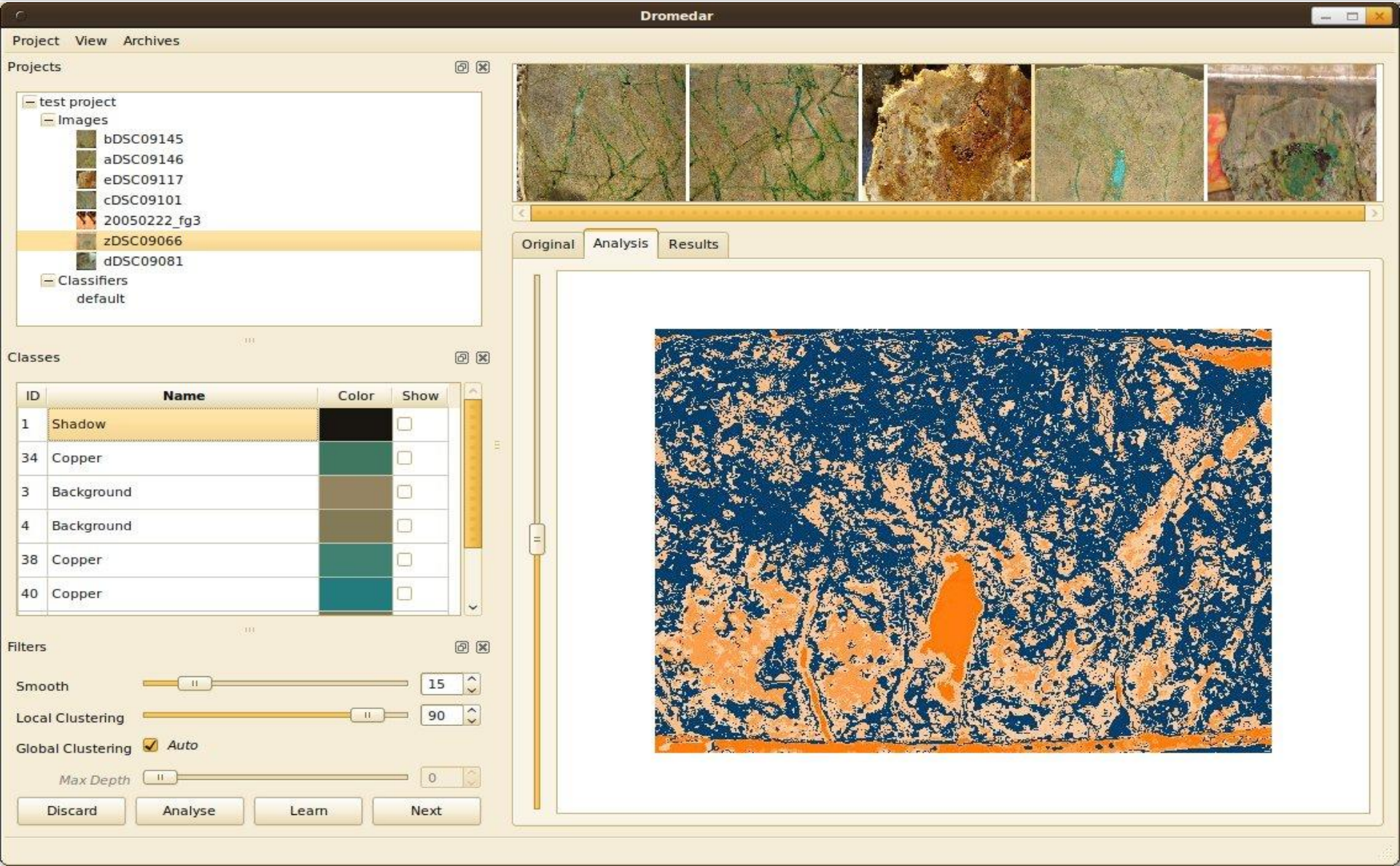
Captura de datos



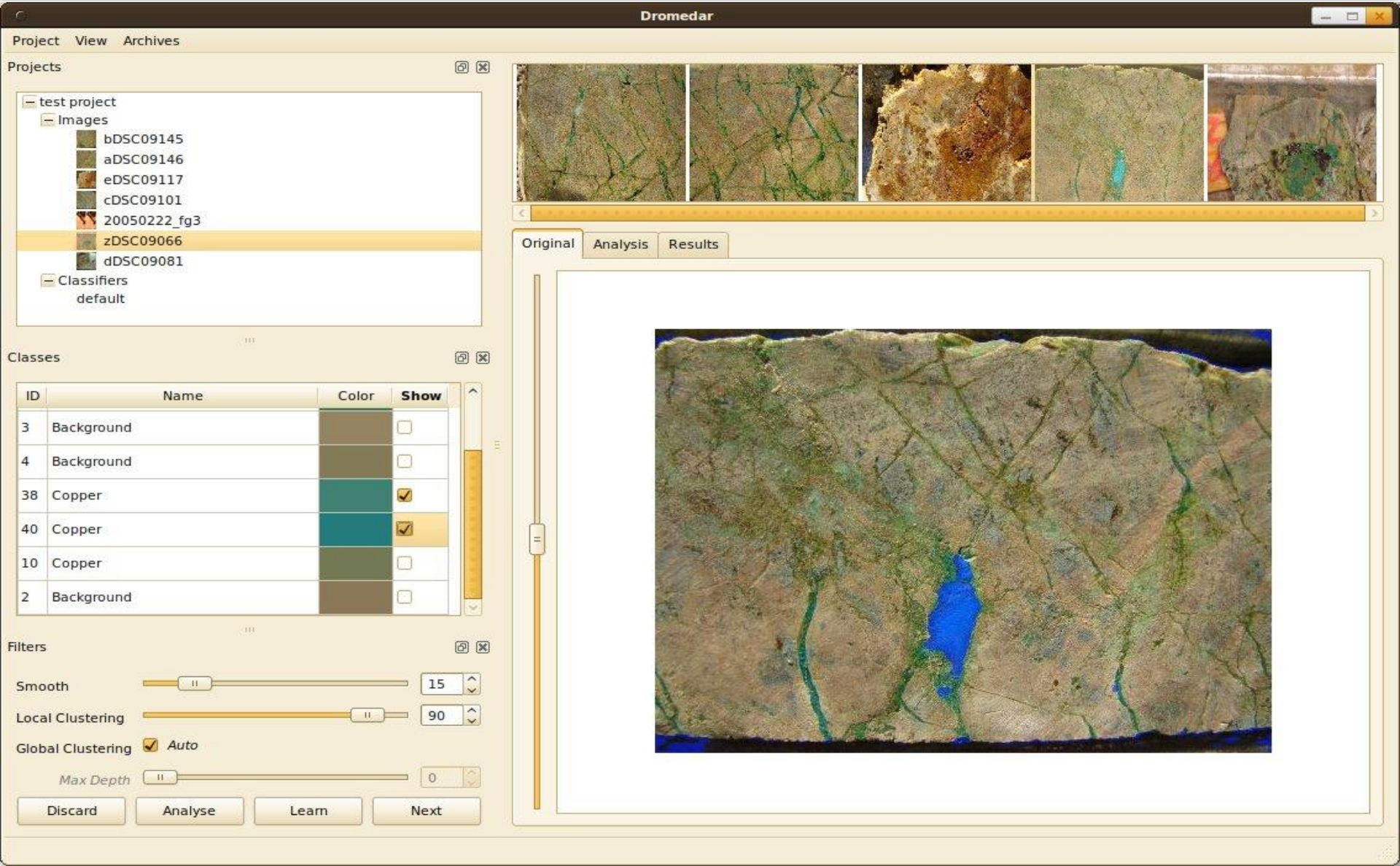
Captura de datos



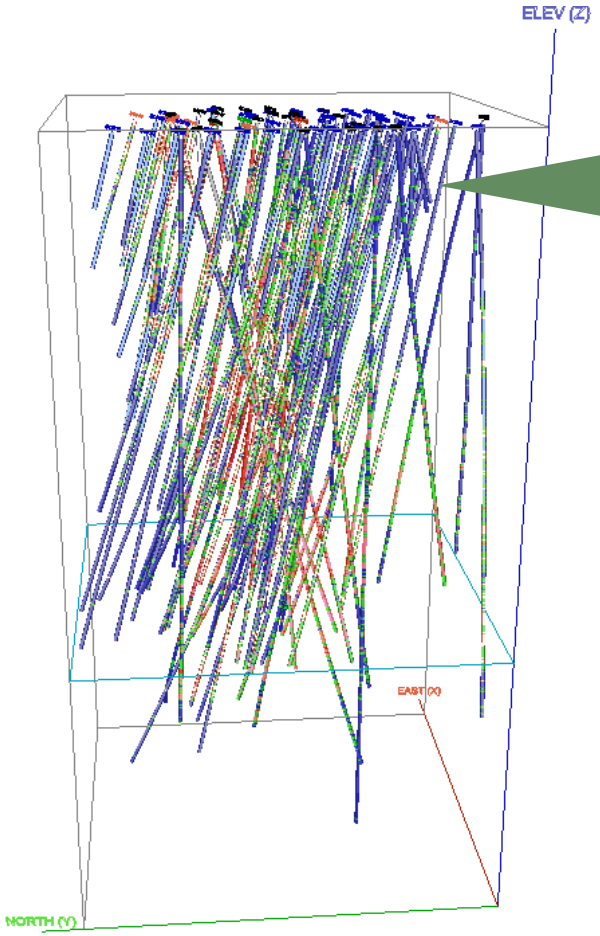
Captura de datos



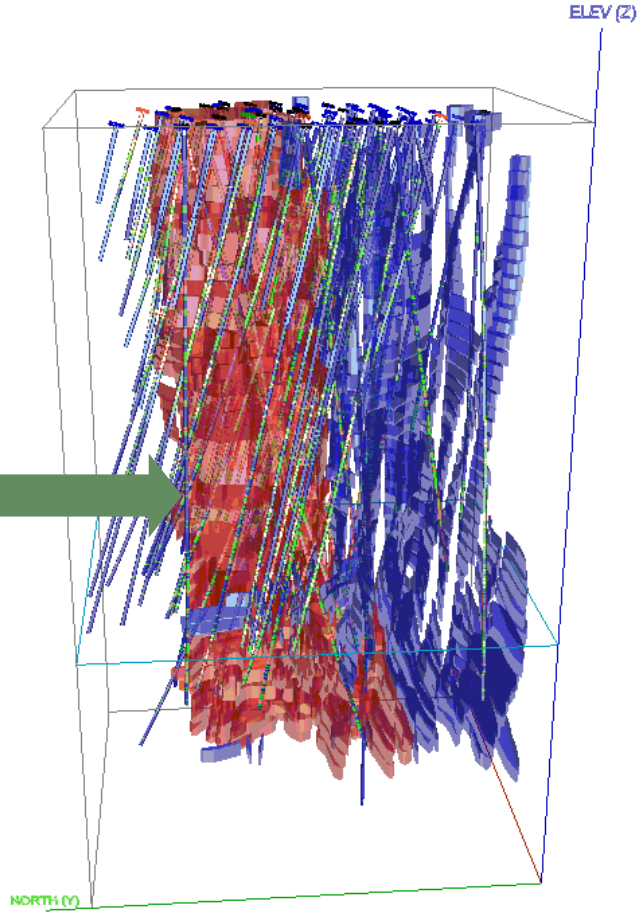
Captura de datos



Captura de datos



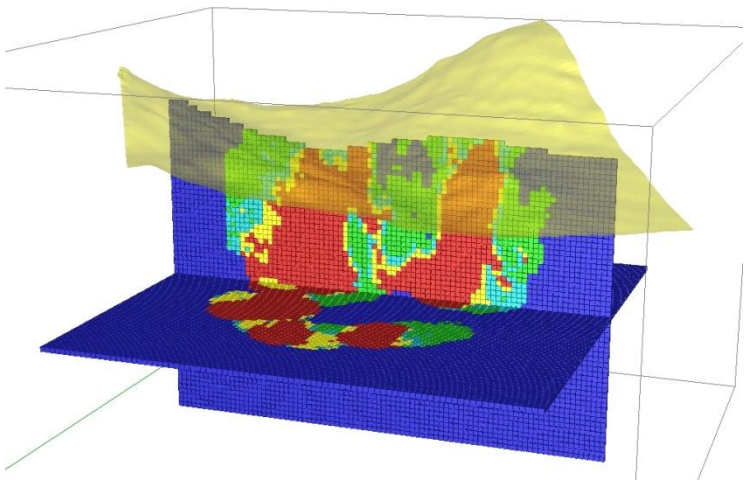
- Leyes
- Mineralogía cuantitativa
- Minerales de alteración
- Textura
- Tipos litológicos
- Atributos geometalúrgicos

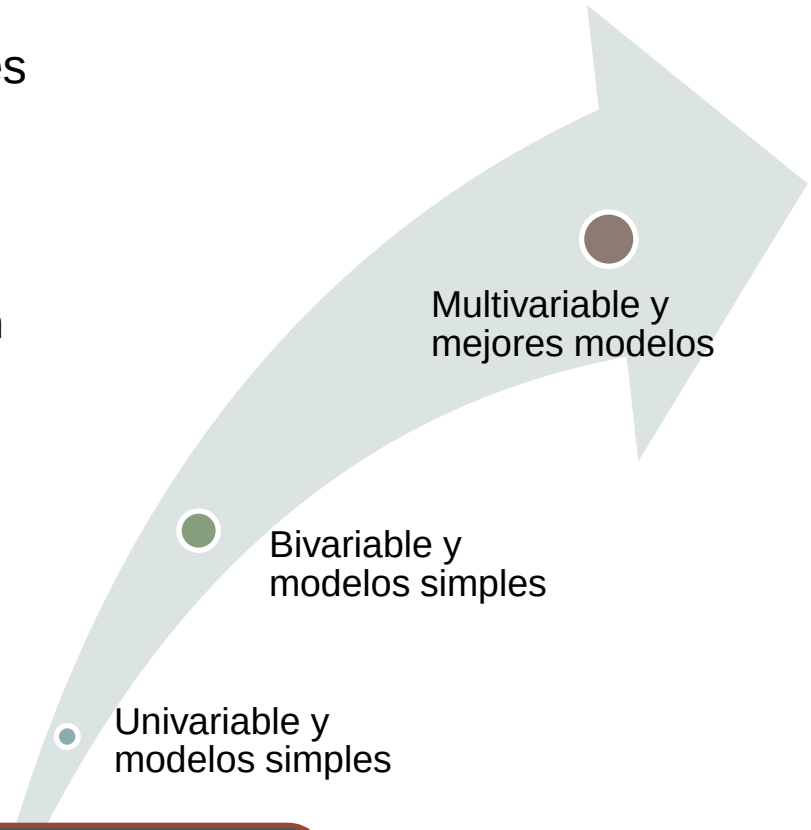


ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

Modelamiento numérico de geología y recursos

- Procedimiento estándar:
  - Captura de datos: mapeo de sondajes
  - Análisis e interpretación de datos:
    - Controles de mineralización
    - Definición de unidades geológicas
    - Definición de unidades de estimación
  - Modelo geológico
  - Estimación y categorización



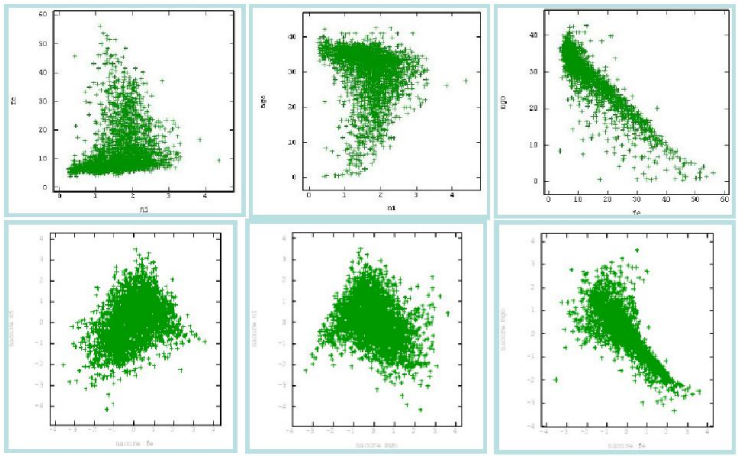


- Modelo geológico
- Modelo de recursos

ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

Modelamiento numérico de geología y recursos

- Más variables conduce a mejores modelos
  - Estimación y Simulación multivariable
- Aplicaciones
  - Flexibilidad operacional (Número de frentes)
  - Requerimientos de mezclas (Stocks, Arsénico,…)
  - Modelos consistentes (CuT/CuS)



Multivariable y mejores modelos

Bivariable y modelos simples

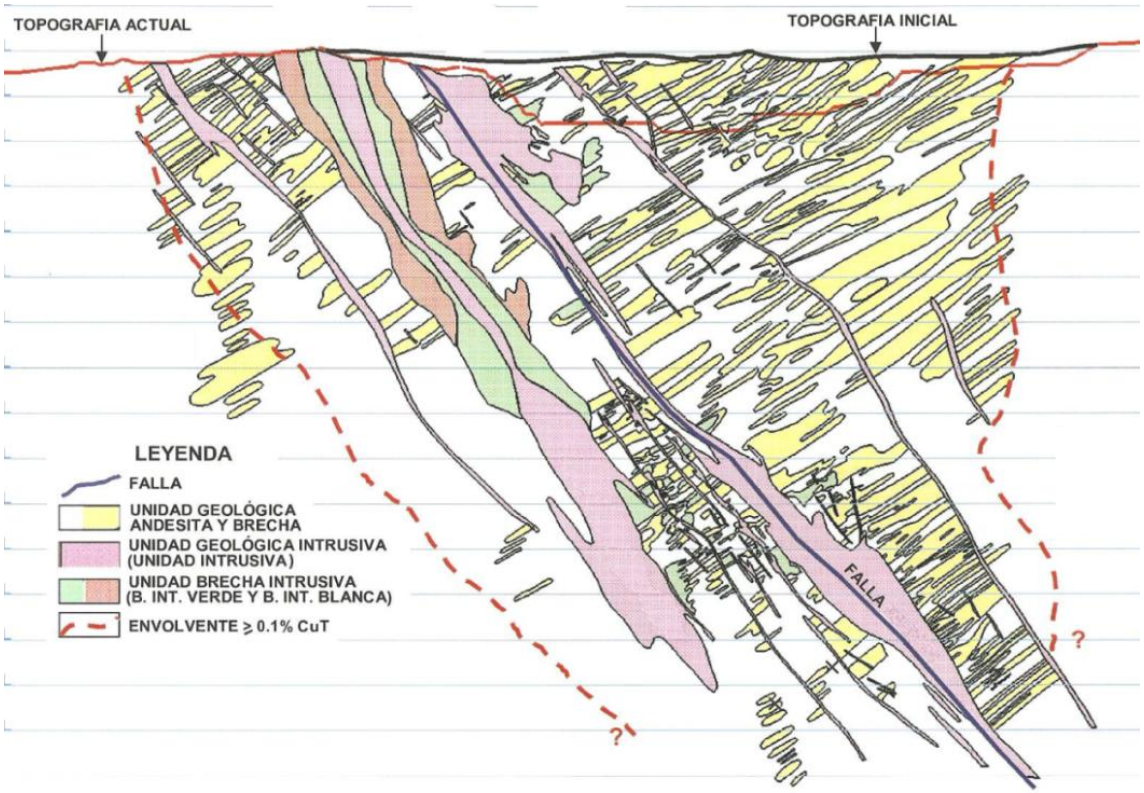
Univariable y modelos simples

- Modelo geológico
- Modelo de recursos

ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

Modelamiento numérico de geología y recursos

- Modelamiento e interpretación geológica
  - Proceso lento → Meses
  - Difícil de automatizar
  - Sujeto a error (por interpretaciones)



Información cuantitativa
- Min / Lito / Alt
- Texturas

Relaciones de dependencia
- Multivariables
- Restricciones mineralógicas

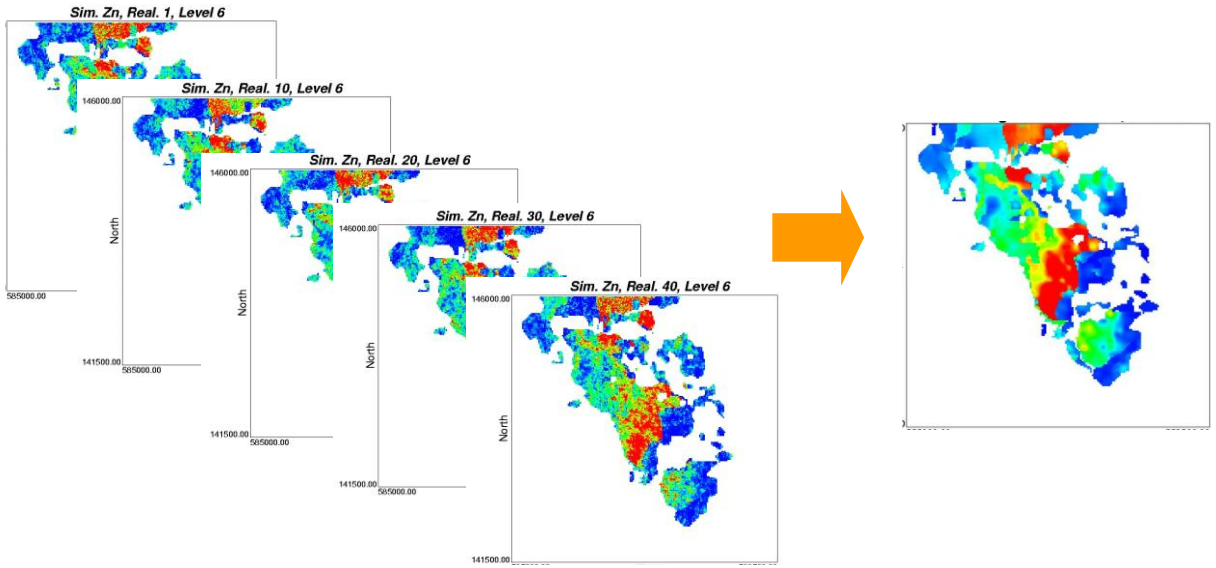
Inferencia y modelo
- Predicción
- Simulación

ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

Modelamiento numérico de geología y recursos

- Modelamiento de recursos y categorización
  - Basado en modelo geológico
    - Determinístico o estocástico
  - Debe considerar relaciones multivariables
  - Estimación y cuantificación de incertidumbre

- Incertidumbre se traduce en riesgo



Inferencia
- Unidades de estimación
- Estadísticas de poblaciones

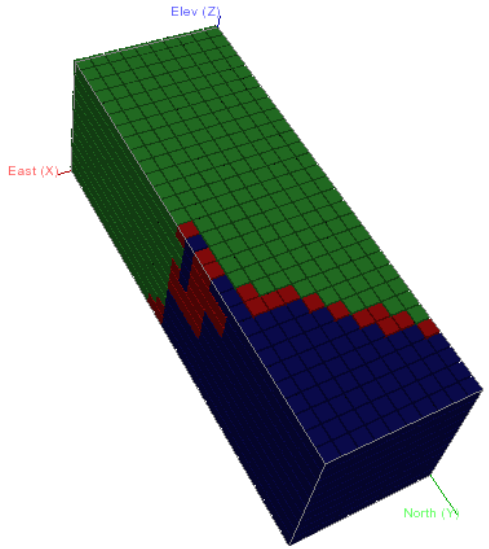
Estimación e incertidumbre
- Modelo consistente
- Mide error esperado

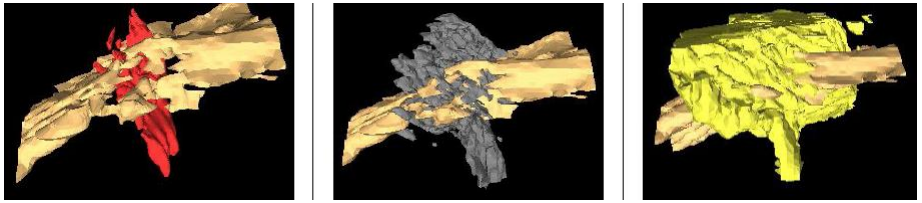
Análisis de riesgo
- Depende de la pregunta
- Costo de incertidumbre

ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

Modelamiento numérico de geología y recursos

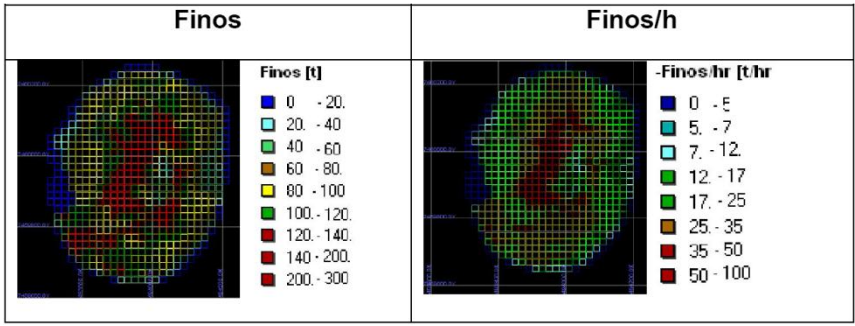
- En resumen:
  - Modelamiento geológico asistido por computador

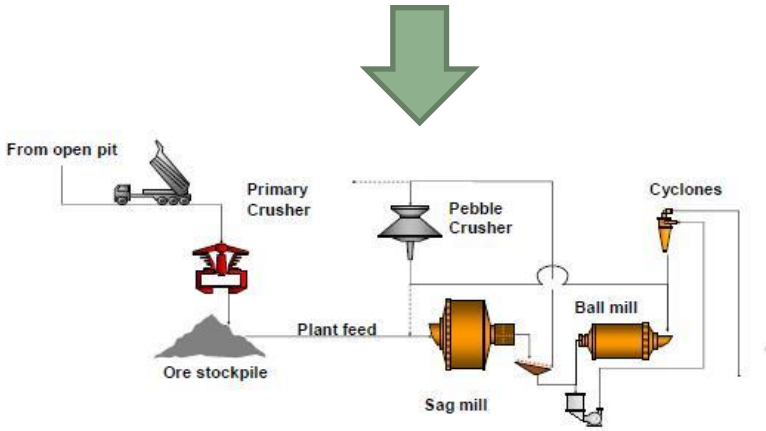




Modelamiento numérico de geología y recursos

- En resumen:
  - Modelamiento geológico asistido por computador
  - Interpolación y cuantificación de incertidumbre
    - Kriging y simulación
    - Impacto en geometalurgia

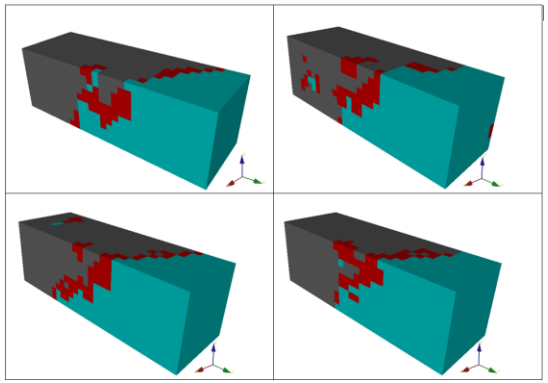


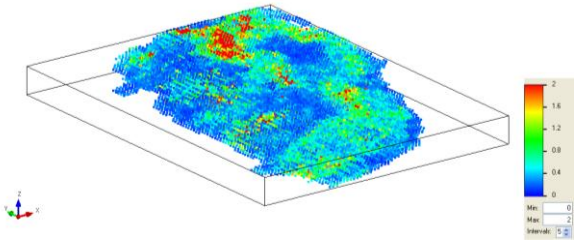


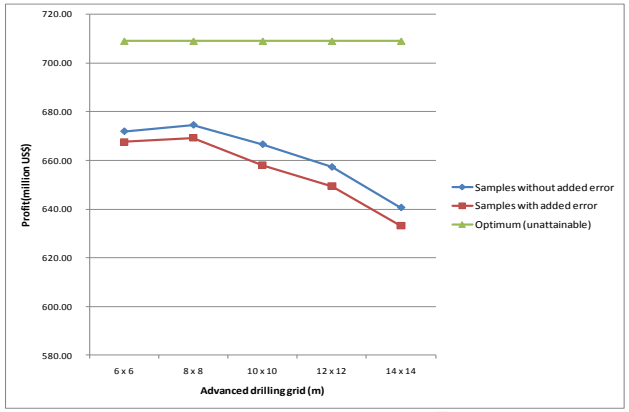
ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

Modelamiento numérico de geología y recursos

- En resumen:
  - Modelamiento geológico asistido por computador
  - Interpolación y cuantificación de incertidumbre
    - Kriging y simulación
    - Impacto en geometalurgia
  - Evaluación económica y análisis de riesgo
    - Incertidumbre en geología (tonelaje, posición y extensión de unidades), leyes y atributos geometalúrgicos
    - Depende del uso que se le dará al modelo: diseño, planificación, evaluación de procesos, optimización de operaciones unitarias
    - Análisis de riesgo **requiere de herramientas ad hoc** para el problema

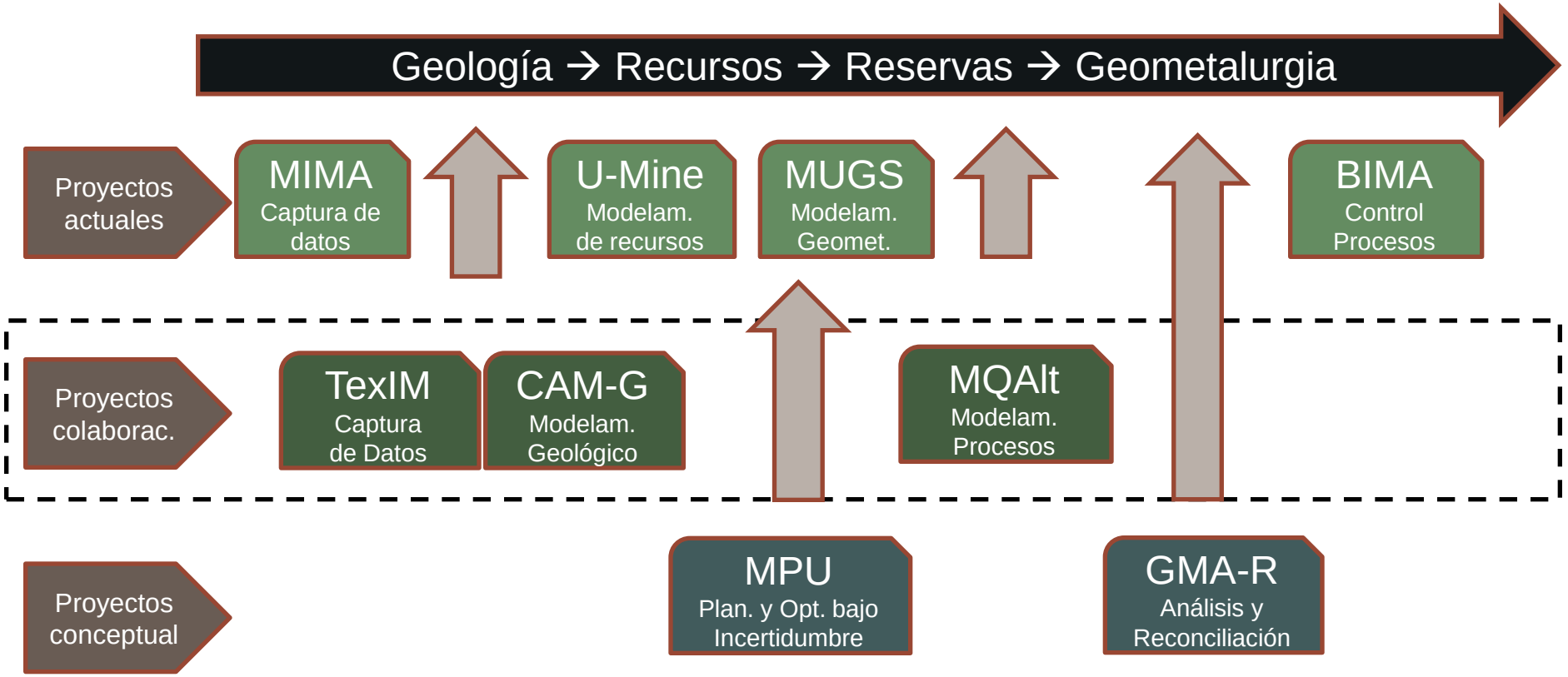






ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

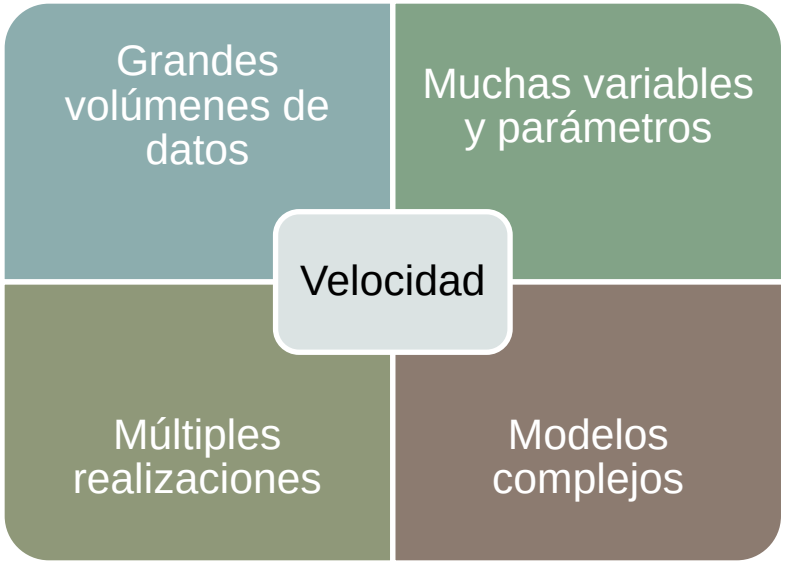
Proyectos



ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

SC y HPC – Problemática

- Problemática
  - Grandes modelos
    - Intensidad de cálculo
    - Millones de bloques
    - Muchas variables
  - Modelos sensibles a muchos parámetros →actualización y validación
  - Múltiples simulaciones → análisis de incertidumbre
  - Tiempos razonables de procesamiento
- Soluciones:
  - Supercómputo (SC)
  - Computación de alto desempeño (HPC)



SC y HPC – Situación actual

- Algoritmos secuenciales
  - Optimizados para correr en sólo una CPU
  - Corren sólo en un computador
  - Difícilmente escalables → “hardware más potente”
- El hardware ya no es un problema ($)
  - PCs de escritorio con 4 o más CPUs y con GPU
  - Clusters
  - Cloud Computing
- Algoritmos secuenciales + hardware ← cota

- Software insuficiente:
  - Construcción de modelos distritales o modelos de detalle con más de 10^7 bloques
    - Codelco Norte
    - Río Blanco
    - Sierra Gorda
  - Capacidad de diseño y optimización insuficiente

ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

SC y HPC – Definiciones

- Supercómputo
  - Resolver problemas de gran escala principalmente por volumen de datos y complejidad de los algoritmos
  - Se resume en “correr grandes problemas con capacidad masiva”
- HPC (Computación de alto desempeño)
  - Uso de supercomputadores para resolver problemas de alta demanda de cómputo
  - Se resume en “correr problemas lo más rápido posible”

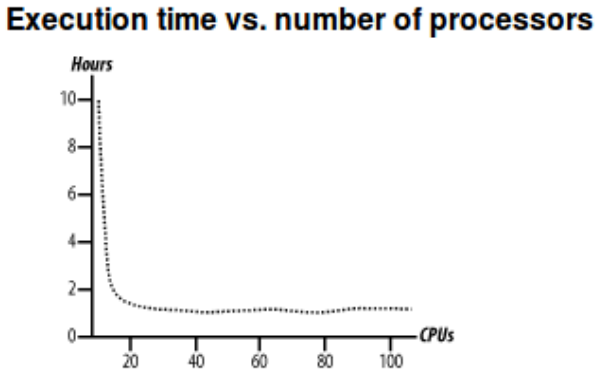
SC → Masivo

HPC → Afinado

ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

SC y HPC – Definiciones

- Conceptos
  - Speedup: $$S_N = \frac{T_s}{T_p}$$
  - Nos gustaría tener un speedup lineal, pero…
    - No todos los algoritmos son paralelizables
    - Paralelización conlleva sobrecosto
    - Ley de Amdahl

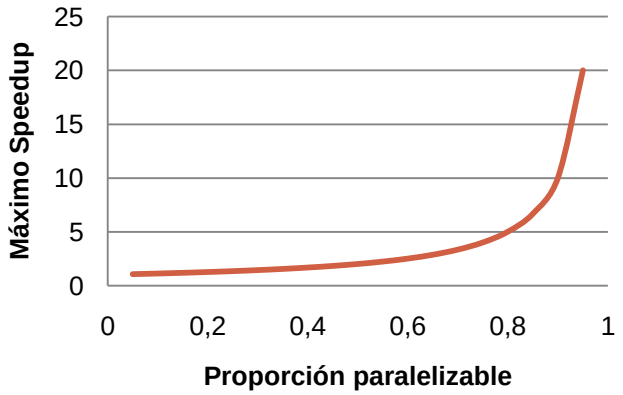


- Ley de Amdahl:
  - El máximo Speedup teórico: $$S_N = \frac{1}{(1-P) + \frac{P}{N}}$$
  - P: Número de procesadores
  - T_s Tiempo utilizando algoritmo secuencial
  - T_p Tiempo utilizando algoritmo paralelo

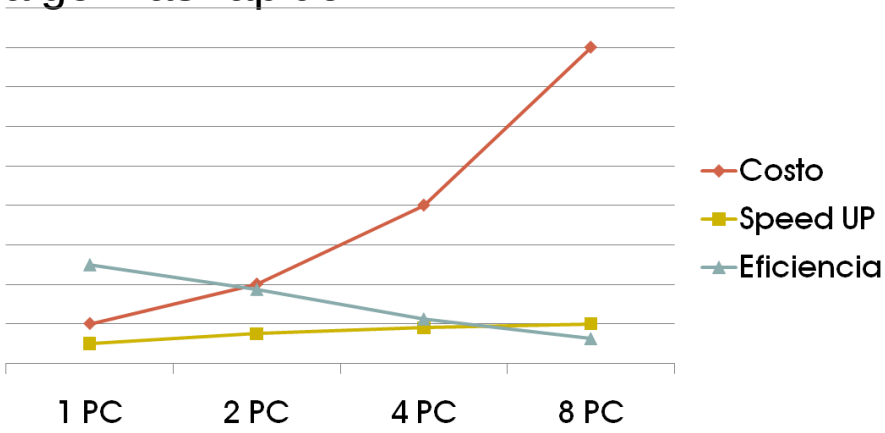
SC y HPC – Definiciones

- Para un algoritmo con un 80% (P=0,8) paralelizable, el máximo Speedup que puede alcanzar es sólo 5

P	S
0,1	1,11
0,2	1,25
0,3	1,43
0,4	1,67
0,5	2,00
0,6	2,50
0,7	3,33
0,8	5,00
0,9	10,00



- Hay un “tope” económico de usar más computadores con el fin de hacer algo más rápido



No basta con tener más procesadores…

ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

Enfoques

1. Algoritmos eficientes
   - Indexación espacial inteligente
   - Bases de datos numéricas
2. Algoritmos paralelizados
   - Capacidad de usar múltiples CPU (era multicore)
3. Aceleración por hardware
   - Uso de una o más GPU
   - Discos de estado sólido
   - Mucha RAM
4. Problemas distribuibles
   - Particionar el problema
   - Ejecutar cada partición en un conjunto de computadores
5. Cluster
   - Muchos computadores dedicados al cómputo
   - Red de alto rendimiento
   - Hardware especializado en cómputo

Enfoques

- Escalabilidad
  - En volumen de datos
  - En eficiencia

Casi nunca lineal

- Algoritmos paralelos
  - Los algoritmos secuenciales no son necesariamente paralelizables
  - Pensar en el problema y en la escalabilidad

No 100% paralelizables

- Algoritmos distribuidos
  - Lidiar con restricciones de Hardware
  - Utilizar todos los recursos disponibles
  - Pensar en el problema y en la escalabilidad

Ejemplo: Análisis de Carga

- Cálculo de memoria RAM en un cokriging
  - Grilla de dimensión 1000x1000x250 nodos → 250 millones de nodos
  - Memoria por nodo:
    - N variables de 32 bits cada una → 4N Bytes
    - Posición nodo: coordenada de 3 dimensiones → 12 Bytes
  - Memoria necesaria:
    - $10^3 \cdot 10^3 \cdot 250 \cdot 4(N + 3)$ Bytes $\approx (N+3) \cdot 2^{10} \cdot 2^{10} \cdot 2^{10}$
    - (N+3) Giga Bytes
  - Memoria requerida relacionada con número de nodos
  - Para modelos grandes es necesario distribuir la carga

ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

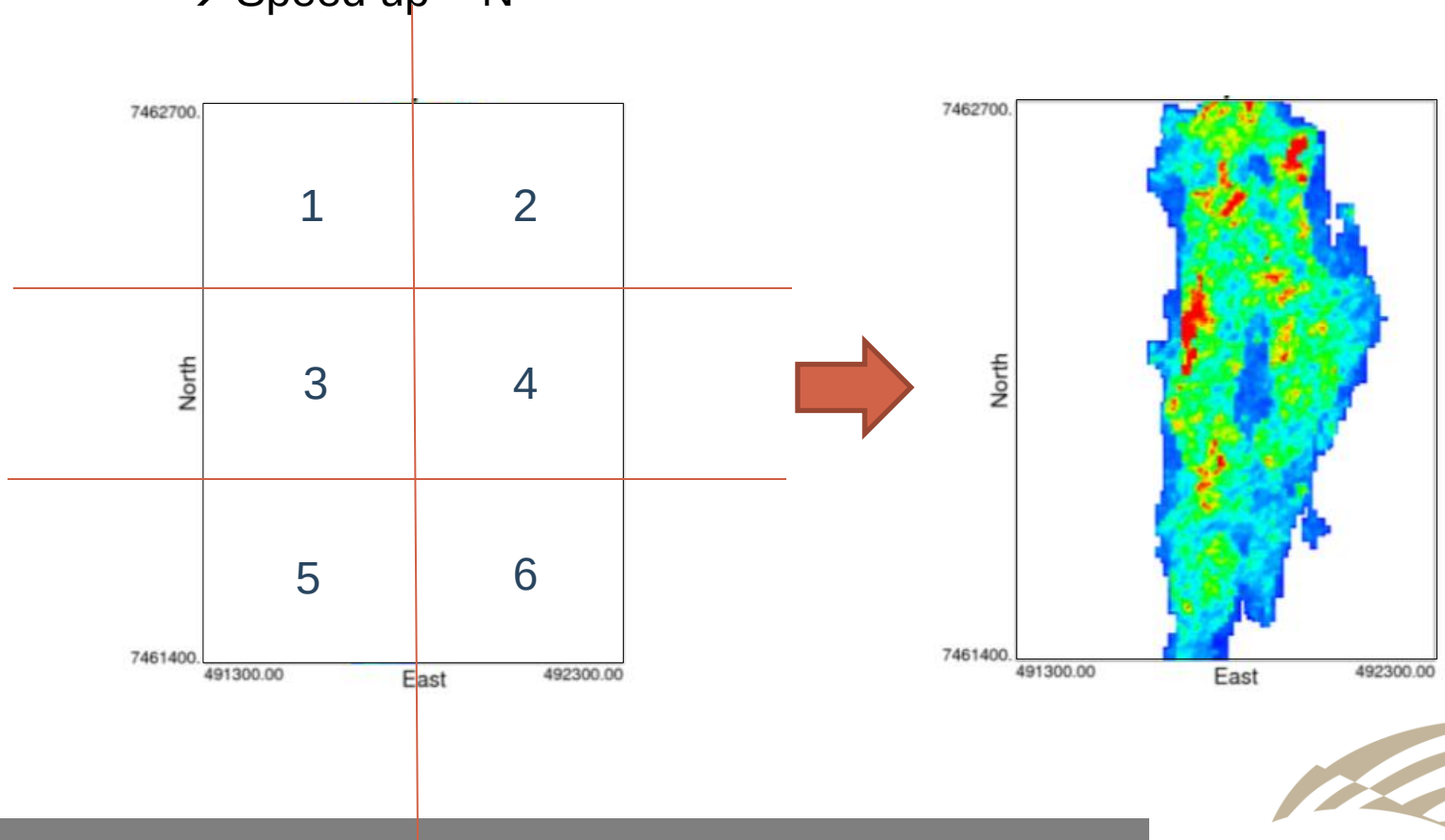
Algunas aplicaciones en desarrollo

- En evaluación de recursos, los problemas típicos son:
  - Búsquedas espaciales de datos vecinos a usar en la interpolación o simulación
  - Solución de sistemas lineales de ecuaciones
  - Cálculos de variogramas modelados

- ¿Cómo lo estamos enfrentando?
  - Estimación por kriging y cokriging distribuyendo el problema
  - Simulación secuencial paralelizada
  - Búsqueda CPU-GPU
  - Bandas rotantes: altamente paralelizable y multivariable

Algunas aplicaciones en desarrollo

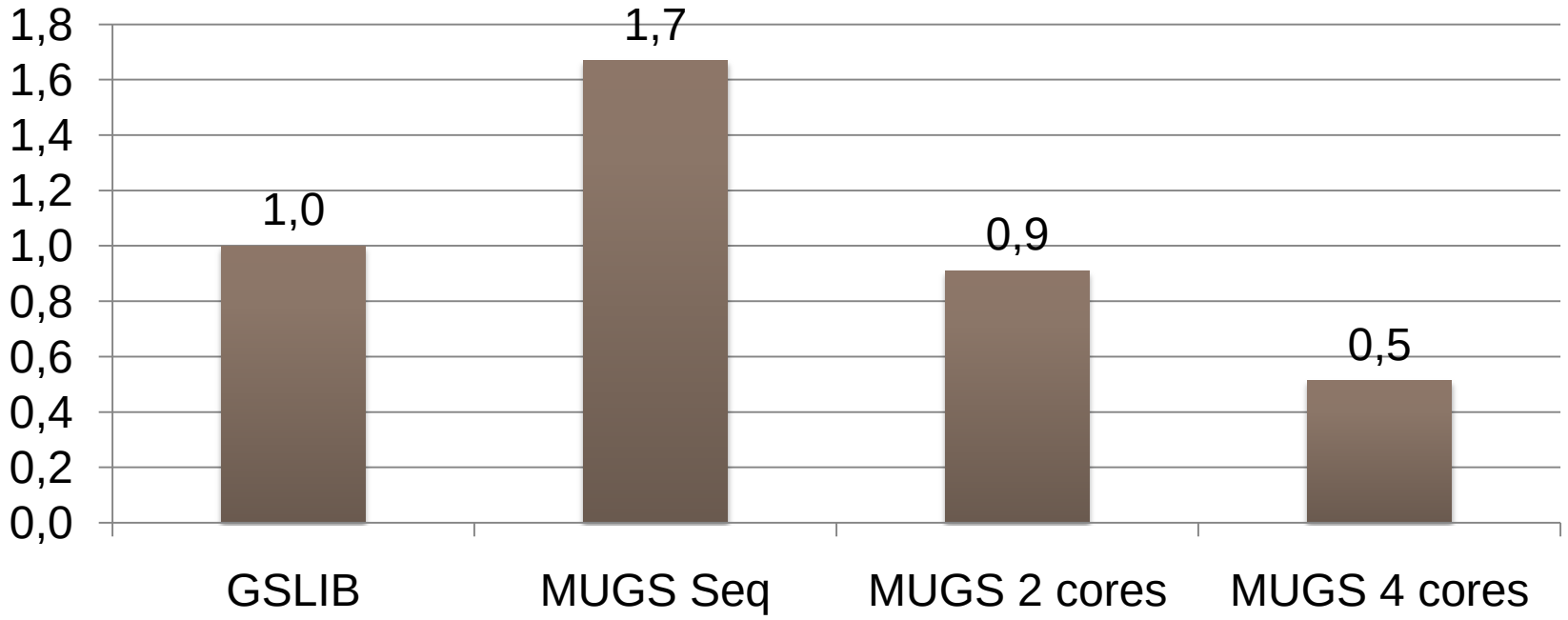
- Estimación por kriging y cokriging
  - Dominio se divide en sectores
  - Cada sector se estima con todos los datos y en una CPU independiente → Speed up ~ N



ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

Algunas aplicaciones en desarrollo

- Cokriging comparado con GSLIB
  - Primer nivel de optimización de código
  - Mejor desempeño con 2 o más cores

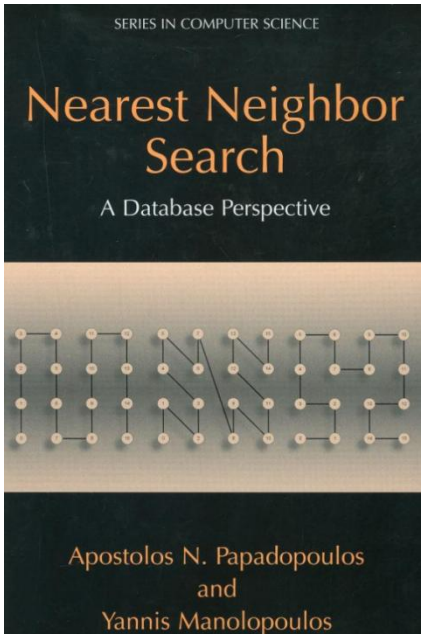


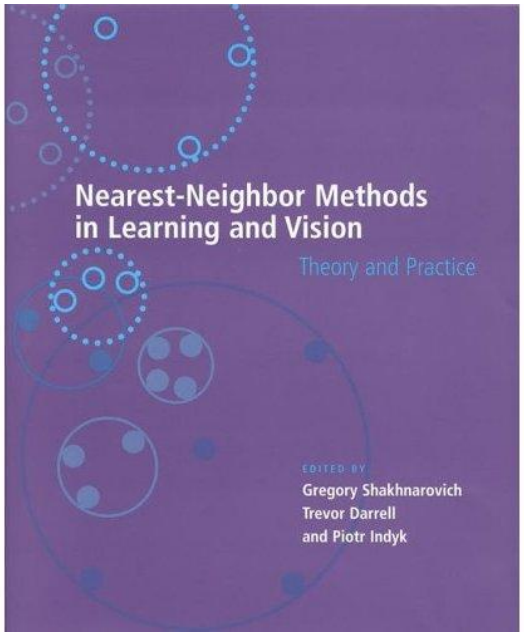
Es necesario romper el paradigma de la programación secuencial: **el software es más lento en modo secuencial, pero es escalable**

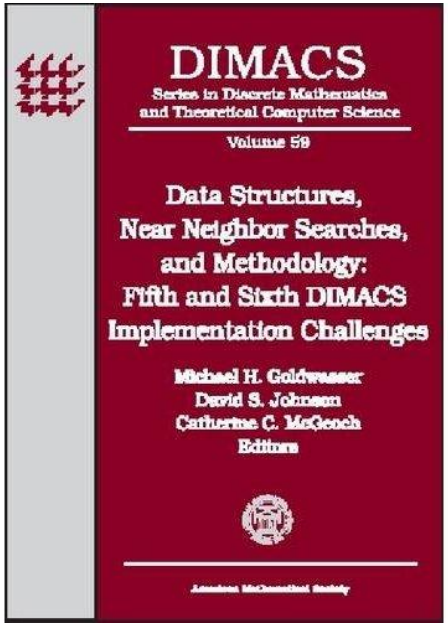
ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

Algunas aplicaciones en desarrollo

- Búsqueda espacial:
  - Problema: "buscar los K vecinos más cercanos, dado N datos disponibles"



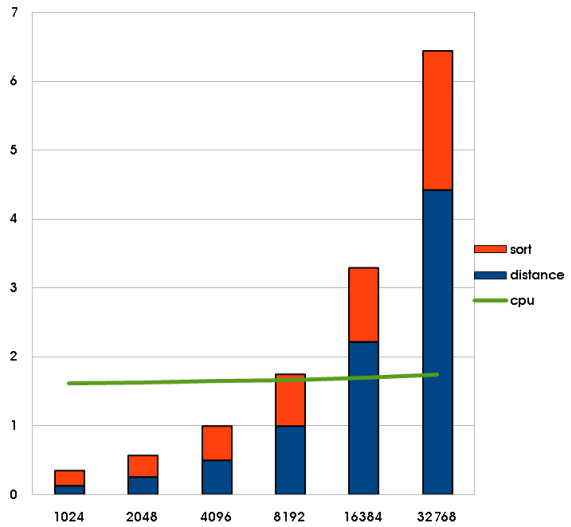


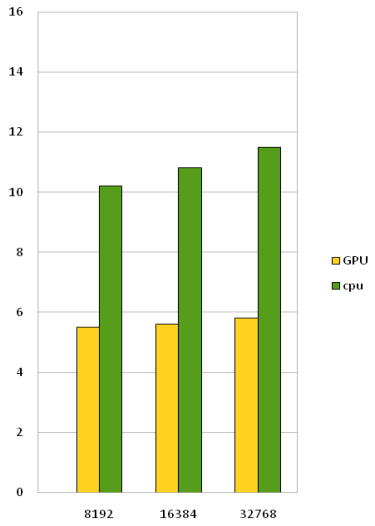


ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

Algunas aplicaciones en desarrollo

- Búsqueda espacial:
  - Problema: “buscar los K vecinos más cercanos, dado N datos disponibles”
  - Solución: dividir cálculos en filtros (CPU) + cálculo masivo de distancias (GPU)





- Resultados preliminares:
  - Aplicables a grandes bases de datos
  - Crecimiento en tiempo de búsqueda no es exponencial
  - Falta paralelizar componente CPU

ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

Algunas aplicaciones en desarrollo

Algoritmo de simulación condicional multivariable

1. Preparar simulación (paralelizable)

   Simular direcciones, frecuencias y fases para método expectral, y parámetros de funciones aleatorias

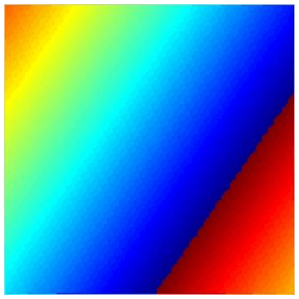
2. Simulación no condicional en cada sitio con dato (paralelizable)

   Proyectar sitio sobre direcciones, calcular valor simulado y combinar

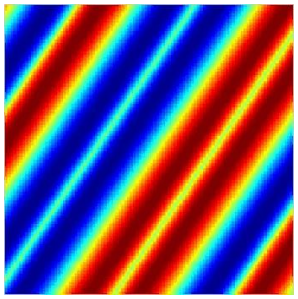
3. Simulación condicional en bloques (paralelizable)

   Buscar K datos condicionantes, proyectar y obtener valor simulado en punto; condicionar, transformar y combinar para obtener simulación de bloques

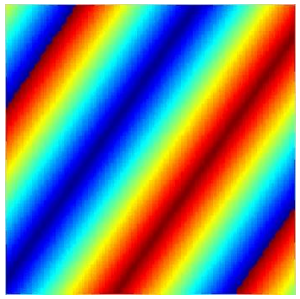
Estable



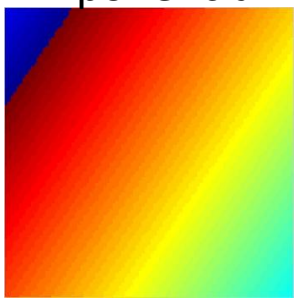
Cúbico



Esférico



Exponencial



ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

Conclusiones

- SC y HPC ofrecen potencial para resolver problemas que hoy no pueden ser resueltos con software y hardware disponibles
- Los proyectos presentados tienen el potencial de:
  - Poder evaluar varios enfoques de estimación / parámetros
  - Rápida validación
  - Capacidad para realizar muchas simulaciones en poco tiempo y analizar escenarios
  - Capacidad para actualizar modelos de manera rutinaria
- Aún hay mucho trabajo por hacer:
  - Modelamiento geológico
  - Interfaz humano-computador
  - Utilización de los resultados para toma de decisiones:
    - Simuladores de procesos mineros: planif CP, MP, LP, extracción, sistemas de despacho, etc.
    - Diseño, optimización y planificación minera

Conclusiones

- Cambio de paradigma:
  - Nuevo software escalable
    - Secuencialmente puede ser más lento
    - Rapidez por cómputo masivo y paralelizable
  - “big iron vs. cluster computing”
- SC y HPC debe facilitar el trabajo:
  - 80-90% automatizado
  - 20-10% supervisado por humano
- Visión global apunta a desarrollo de modelos orientados a la geometalurgia

ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

HPC y Supercómputo Aplicado a la Evaluación de Recursos

Julián Ortiz
Álvaro Parra
Exequiel Sepúlveda

5° Seminario de Acercamiento Tecnológico
Codelco Digital: Minería del Futuro

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

amtc
ADVANCED MINING TECHNOLOGY CENTER