



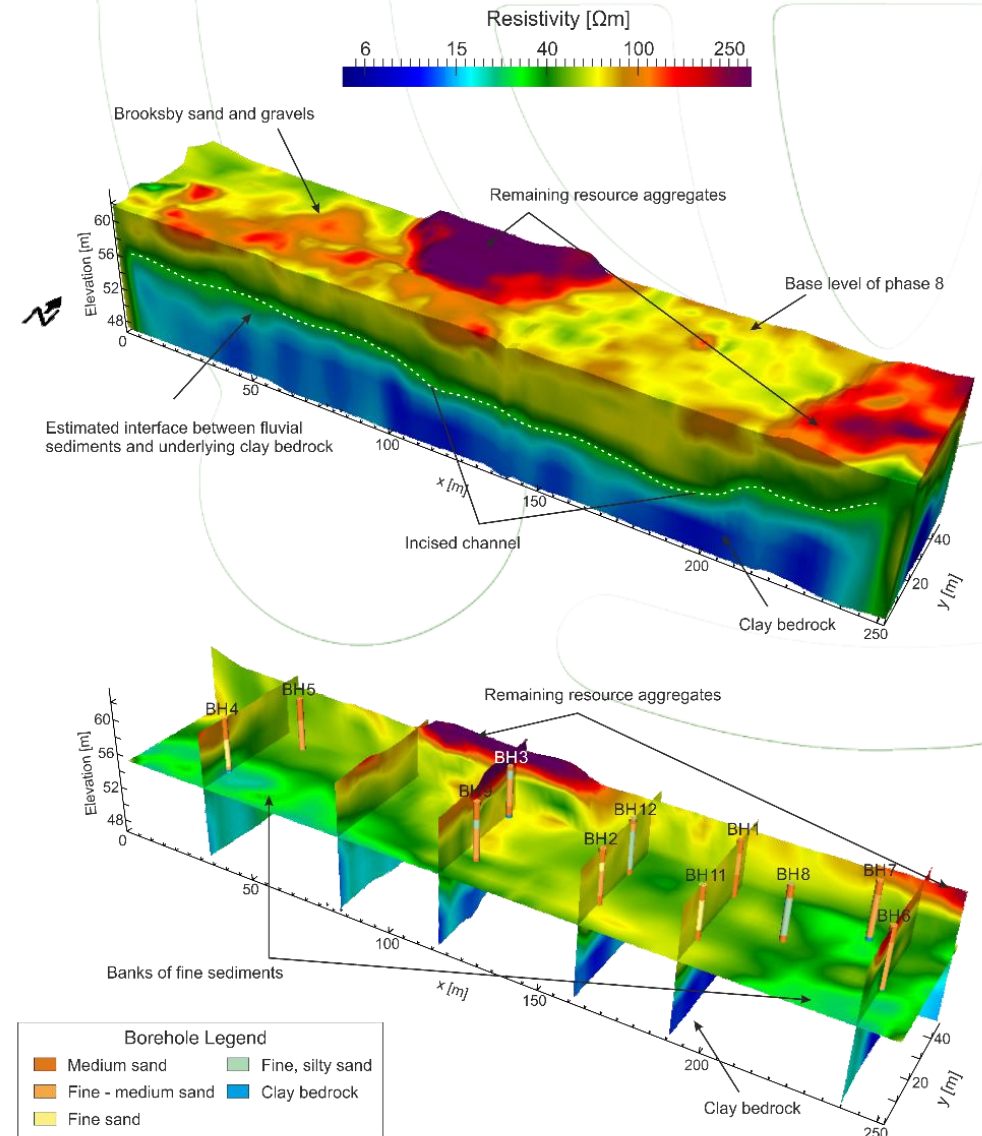
Machine Learning for exploiting and assimilating environmental data

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System
(D.D. n. 130/2022 - CUP B53C22002150006) Funded by EU - Next Generation EU PNRR-
Mission 4 “Education and Research” - Component 2: “From research to business” - Investment
3.1: “Fund for the realisation of an integrated system of research and innovation infrastructures”



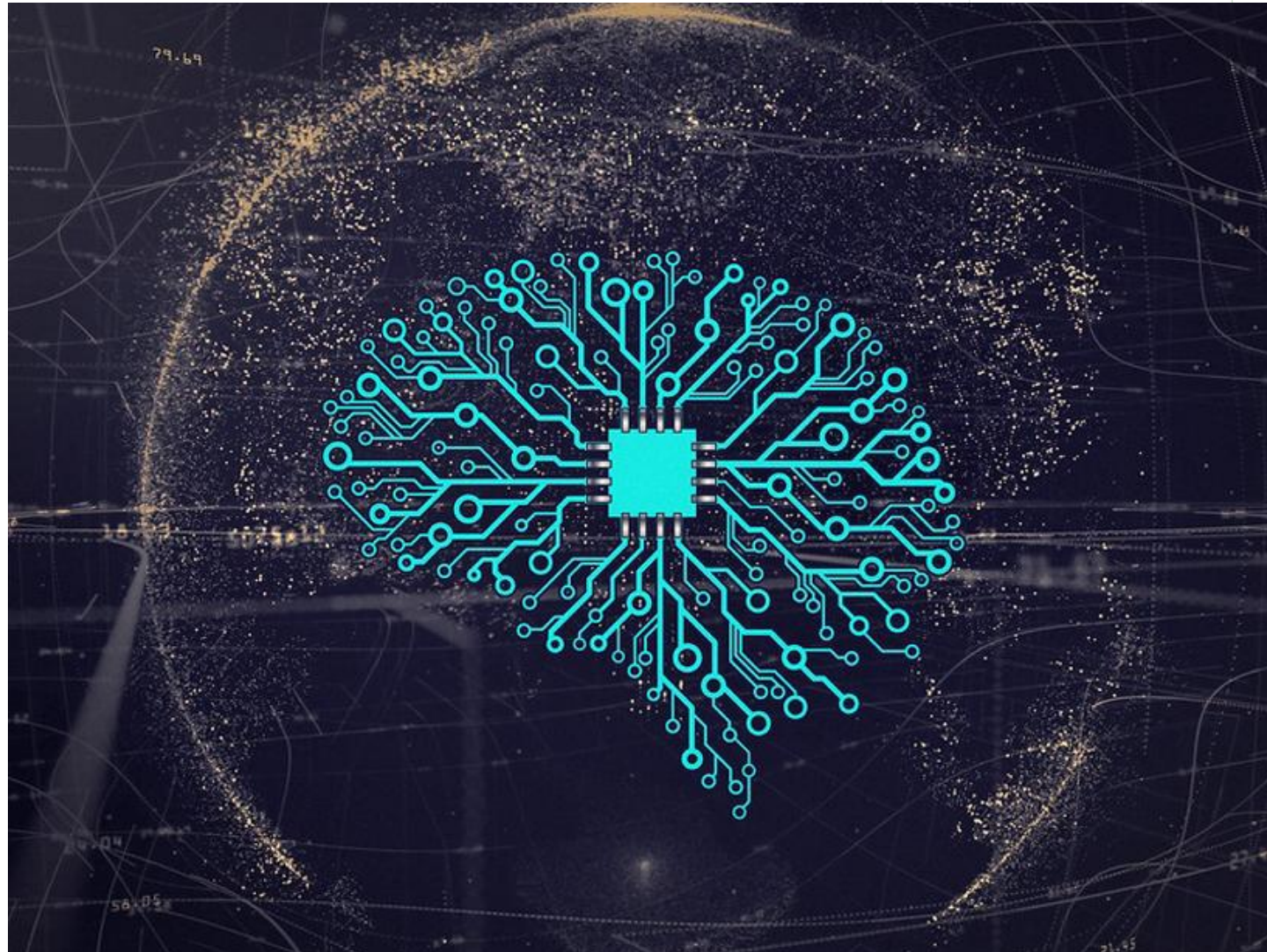
Outline of the day

- **Fundamentals of electrical resistivity measurements**
 - Electrical properties of soils and rocks
 - Translating geoelectrical data into hydrogeomechanical properties
- **Principles of Electrical Resistivity & Electromagnetic Measurements**
 - Applications on ground and airborne geophysical characterization
- **Geoelectrical monitoring: Principles and applications**
 - Inversion approaches, practical considerations
 - Monitoring applications
- **Machine Learning for exploiting and assimilating environmental data**



Machine Learning in the Geosciences

Unlocking Earth's secrets with AI



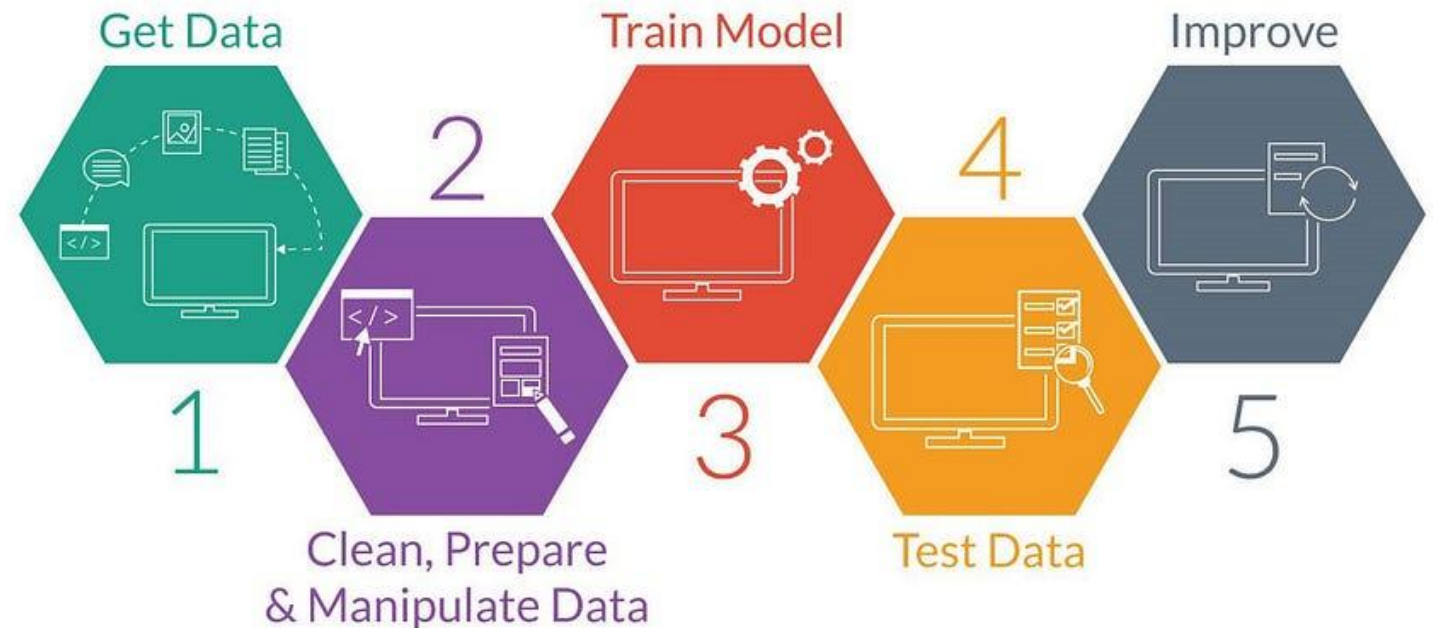
What is Machine Learning

Basic concepts

Definition: A subfield of **artificial intelligence** that allows computers to **learn from data** without explicit programming.

Key Concepts

- **Algorithms:** Mathematical models that enable computers to learn patterns and make predictions (e.g., random forest)
- **Data:** The fuel for machine learning models. Large and diverse datasets are crucial.
- **Training:** The process of feeding data to the algorithm, allowing it to adjust its parameters and improve its performance.
- **Prediction:** Once trained, the model can make predictions or classifications on new, unseen data.



Why Machine Learning in Geosciences

Basic concepts



Challenges in the Geosciences

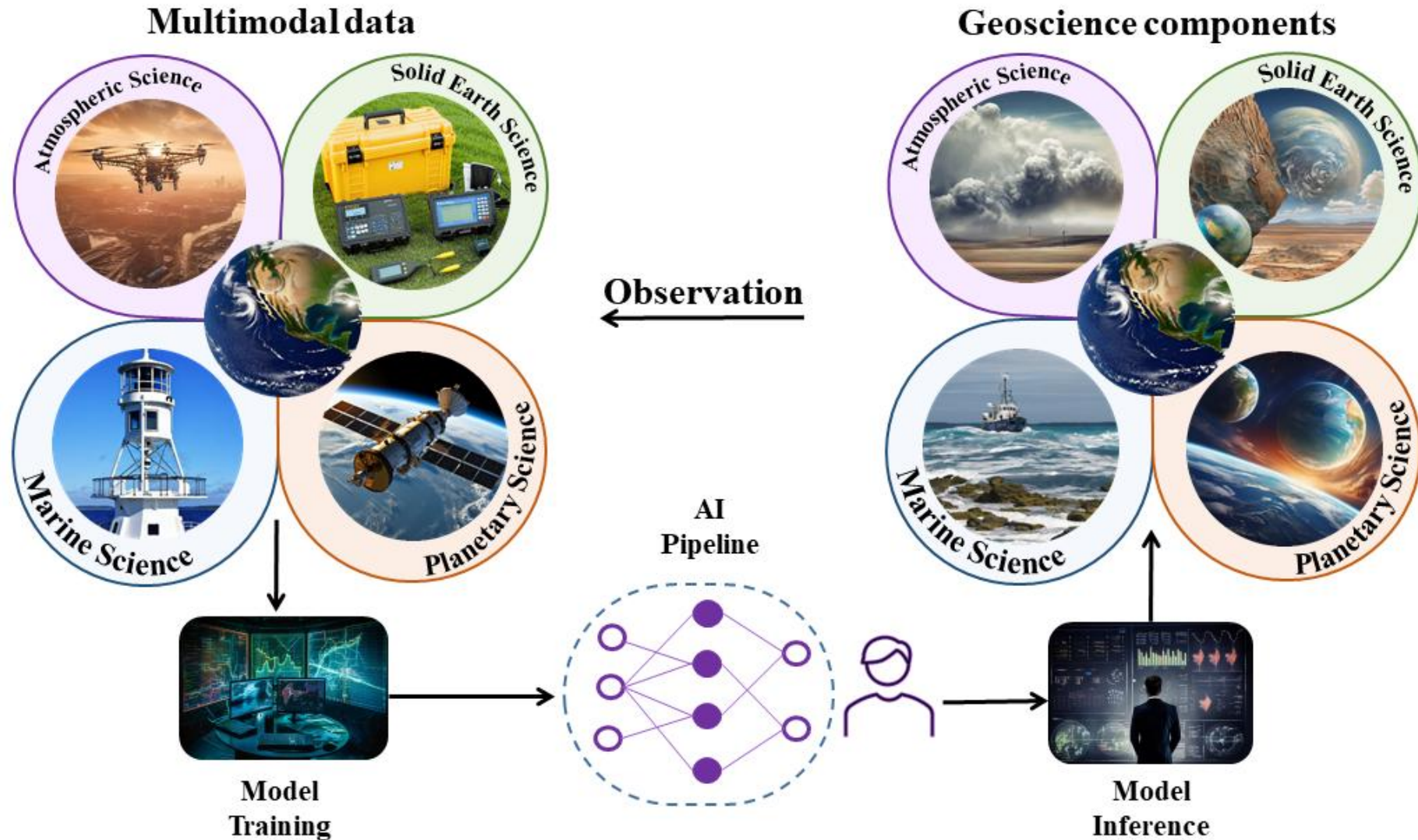
- **Massive datasets:** From satellite imagery to seismic data, geoscientists deal with enormous amounts of information.
- **Complex systems:** Earth systems are intricate and interconnected, making them difficult to understand and model.
- **Data heterogeneity:** Data comes from various sources, often with different formats and quality.

Benefits of Machine Learning

- **Data analysis:** Efficiently process and extract insights from large datasets.
- **Pattern recognition:** Identify hidden patterns and relationships in complex data.
- **Prediction:** Forecast future events, such as extreme weather events, landslides, and climate change.
- **Decision support:** Provide valuable insights for decision-making in areas like resource exploration and environmental management.

Machine Learning in the Geosciences

Unlocking Earth's secrets with AI



Types of Machine Learning

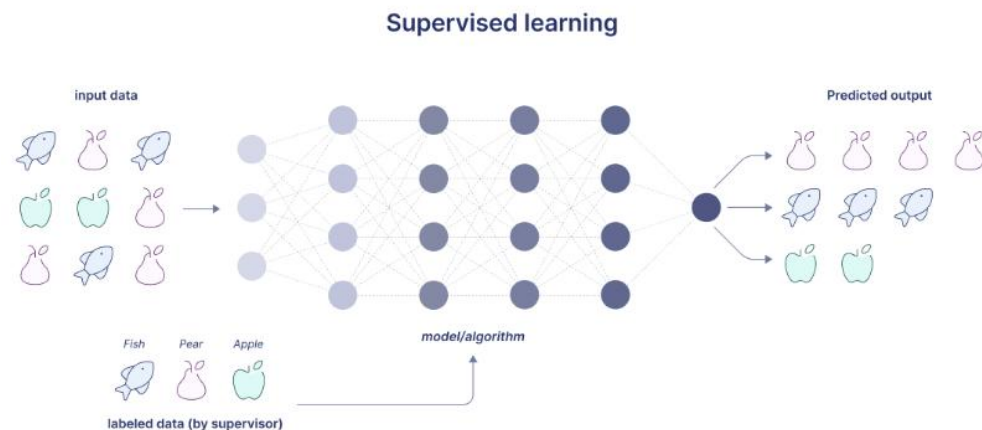
Supervised and unsupervised learning

Supervised learning

Trains on labeled data, where the algorithm learns to map inputs to known outputs.

Examples:

- Classification: Predicting the type of rock or mineral based on its properties.
- Regression: Forecasting future sea-level rise based on historical data.

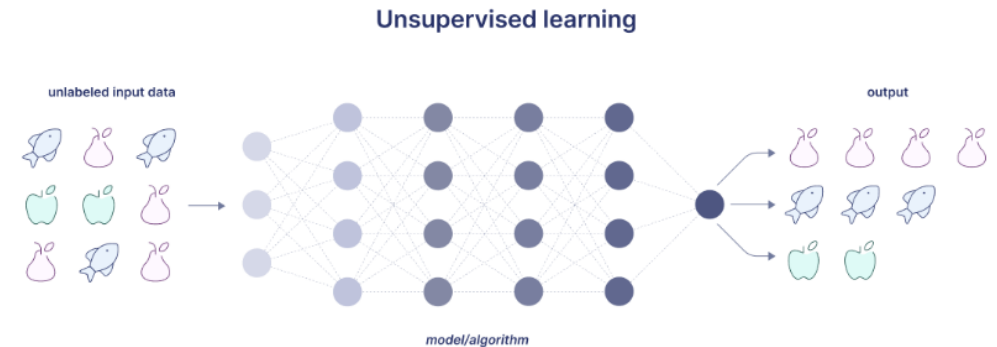


Unsupervised learning

Trains on unlabeled data, allows to discover hidden patterns and structures.

Examples:

- Clustering: Grouping similar earthquakes together based on their characteristics.
- Dimensionality reduction: Reducing the complexity of high-dimensional data.



Applications of Machine Learning in Geosciences

Large range of possible applications



Earth Observation

Image analysis: Classifying land cover, detecting changes in vegetation, and monitoring natural disasters.

Geophysics

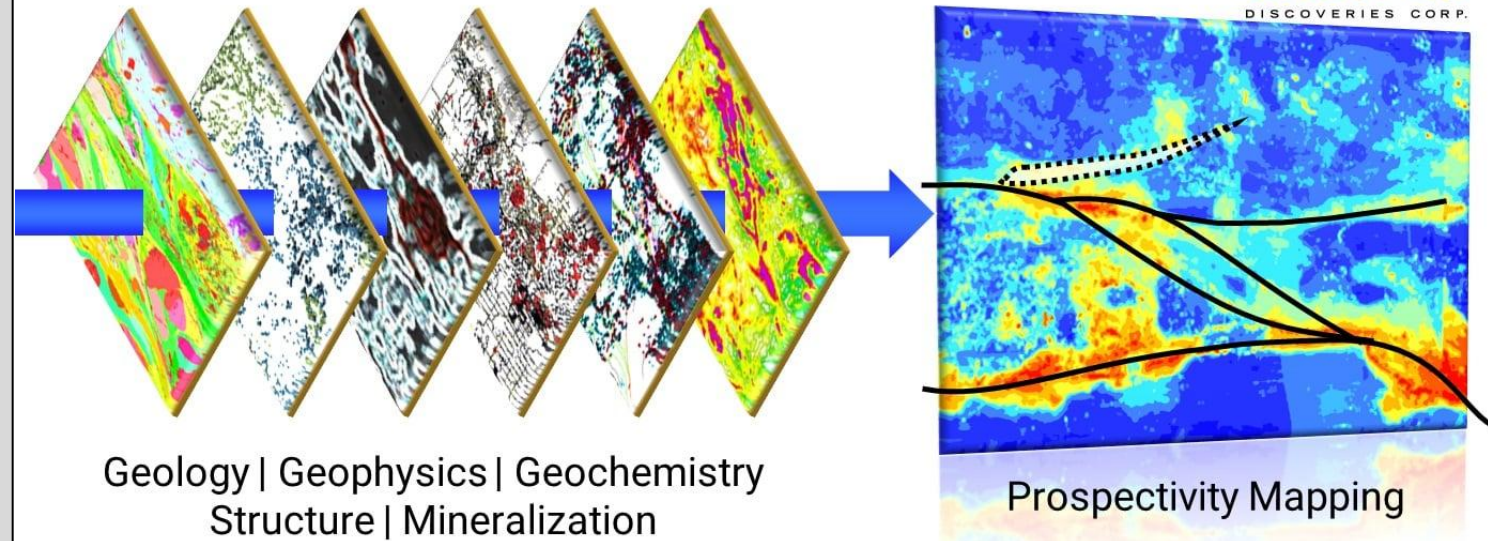
Seismic interpretation: Identifying oil and gas reservoirs, predicting earthquakes.

Hydrology

Water resource management: Forecasting floods, predicting water quality.

Climate Science

Climate modeling: Improving climate models, predicting future climate scenarios.



Machine Learning in the Geosciences

Limitations



**High Computation
Power**



Generation Latency



Data Scarcity



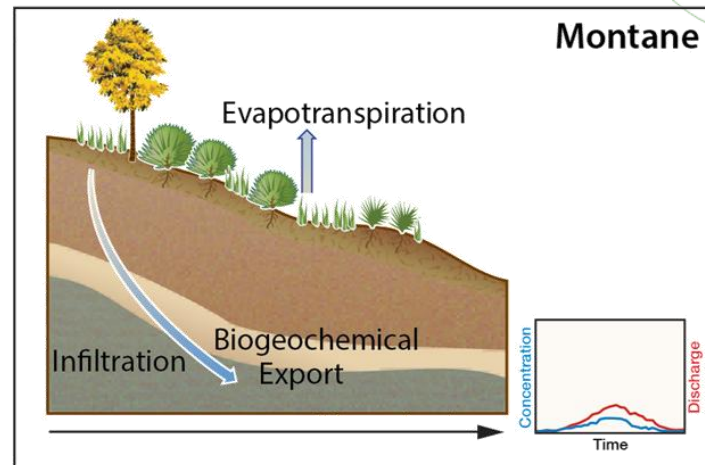
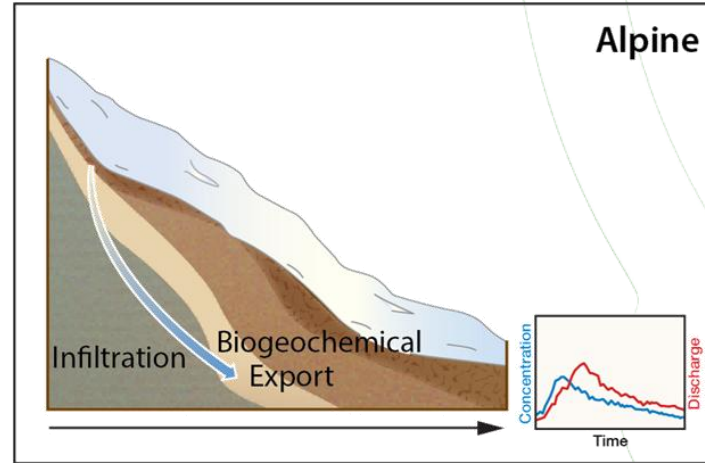
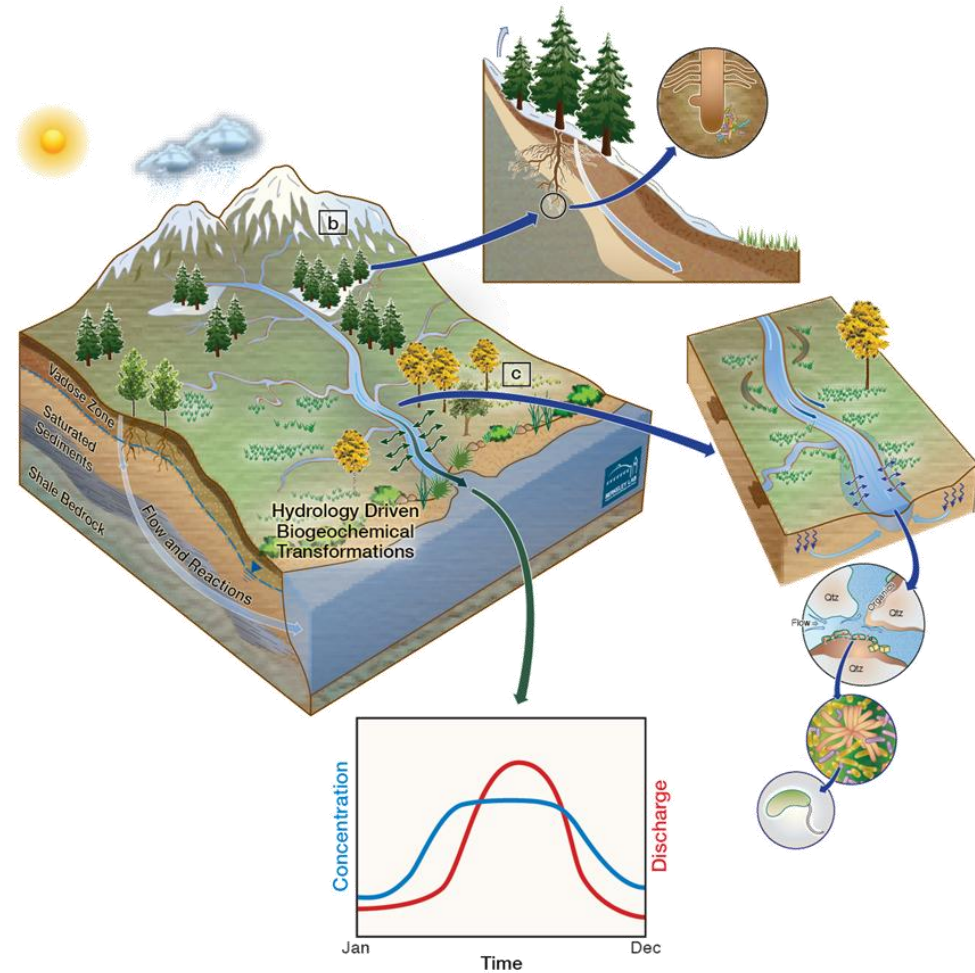
**Explainability &
Trustworthiness**



**Fake Data
Generation & Misuse**

Hillslope Hydrology

An integral part to understanding the functioning of mountainous watersheds



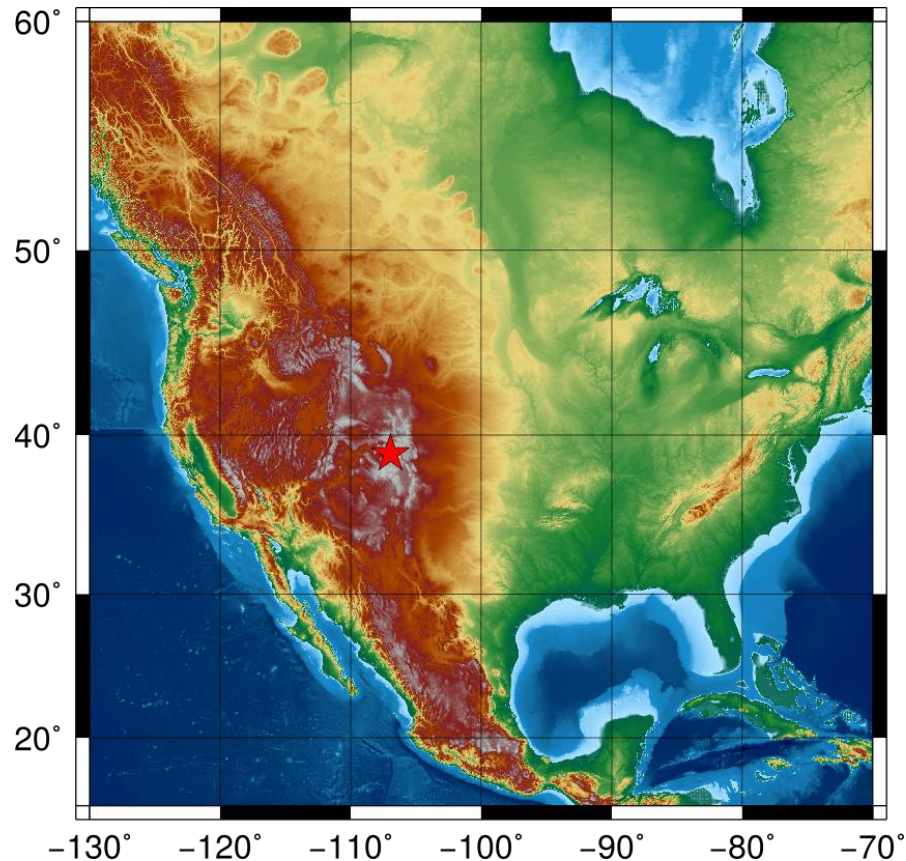
How do mountainous watersheds retain and release water, nutrients, carbon and metals over episodic to decadal timeframes?

Understanding **bedrock characteristics and flow** is critical to answering this question.

EESA16-013

The East River Watershed

A collaborative platform for the scientific community



Facts:

- **Area:** ~ 300 km²
- **Average Elevation:** 3266 m
- **Topographic relief:** 1420 m

- **Pronounced gradients in:**
 - Hydrology
 - Geomorphology
 - Life zone (alpine, subalpine, montane)

The East River is a **major tributary** to the Gunnison River and in turn the **Colorado River**

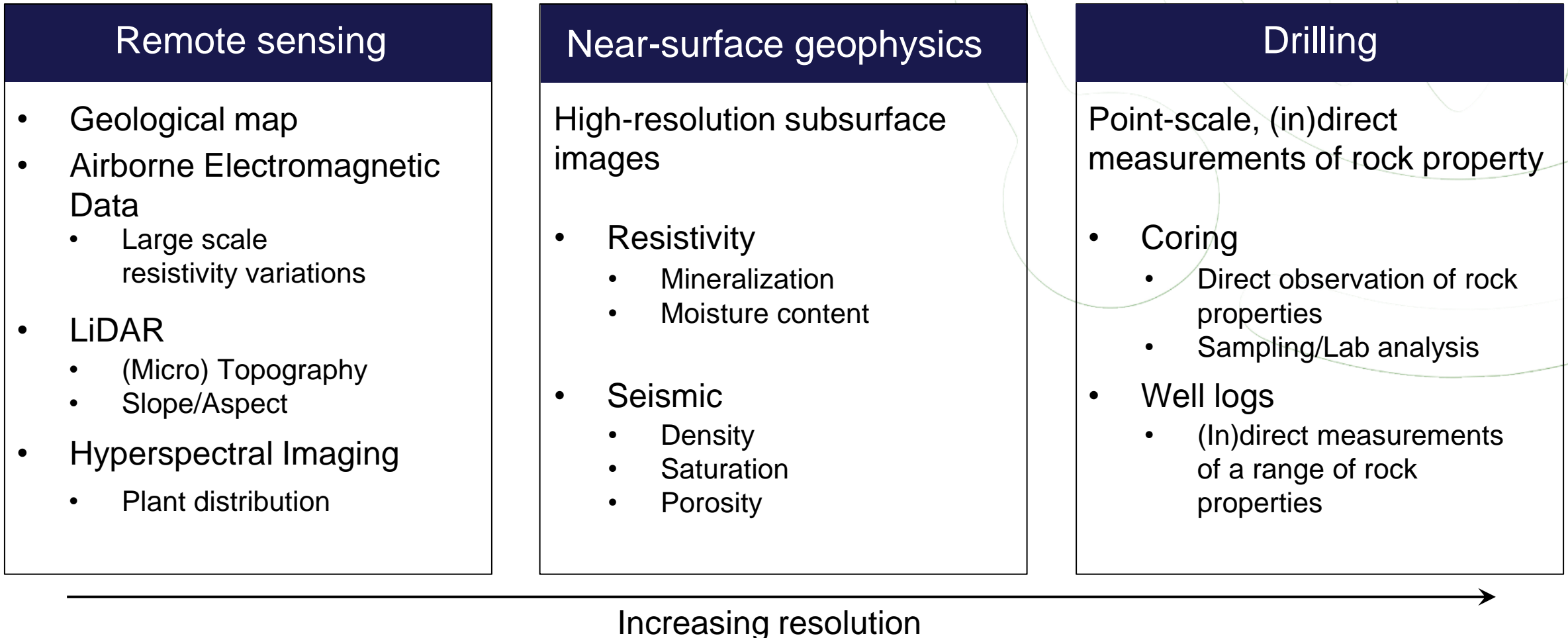
- Supplies water to **40 Million** people and **20,000 km²** of agricultural land



Upscaling bedrock properties

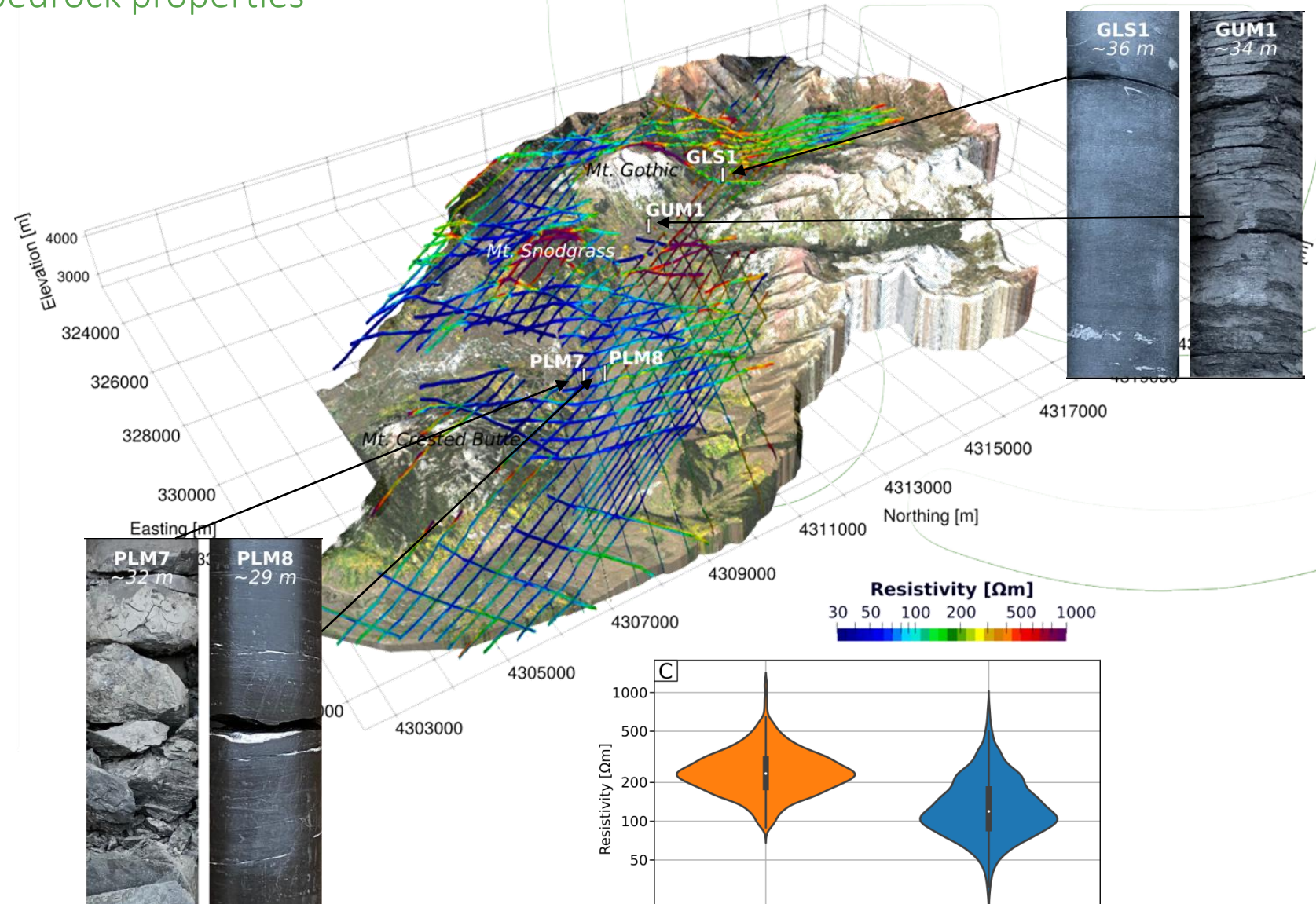
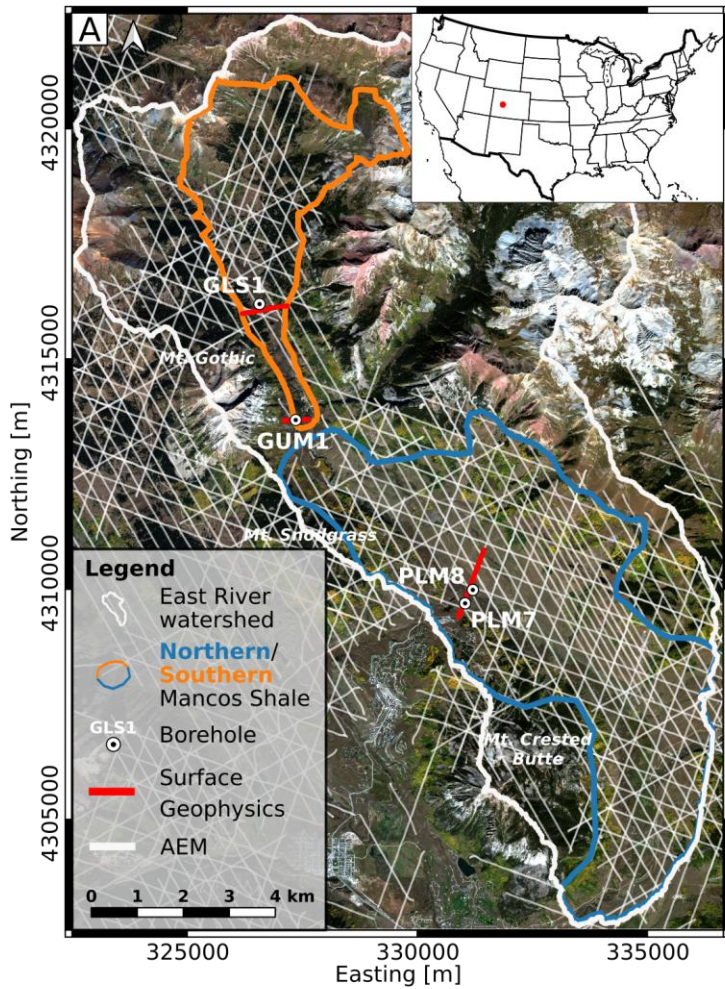
Multi-disciplinary approach

Large-scale characterization of bedrock properties by combining measurements of coarse to fine resolution



Shale properties vary throughout the watershed

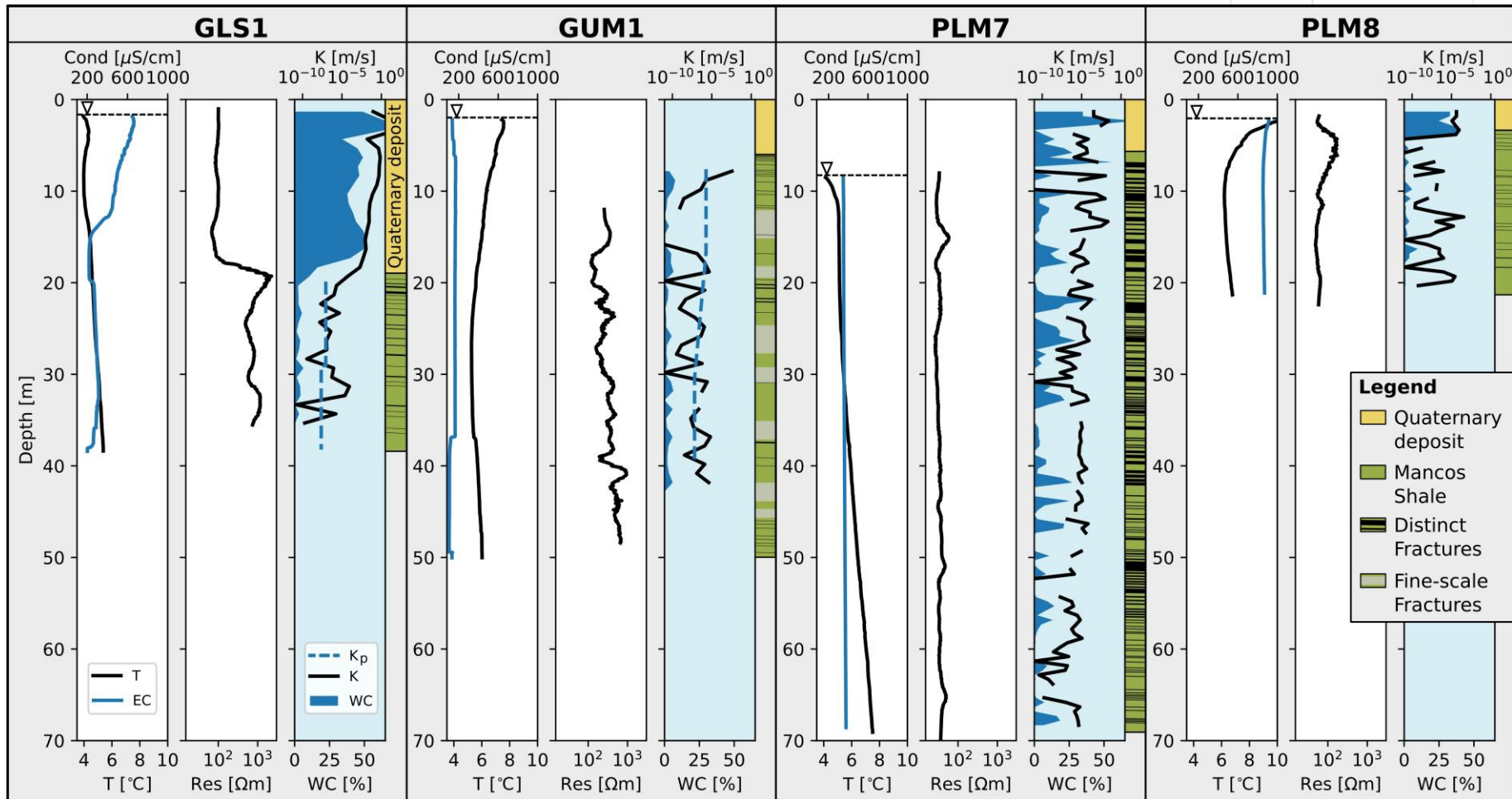
Electrical Resistivity as proxy for bedrock properties



Uhlemann et al. (2022) *Surface parameters and bedrock properties covary across a mountainous watershed: Insights from machine learning and geophysics*. Science Advances 8(12)

Shale properties vary throughout the watershed

Borehole logging highlight small scale variability

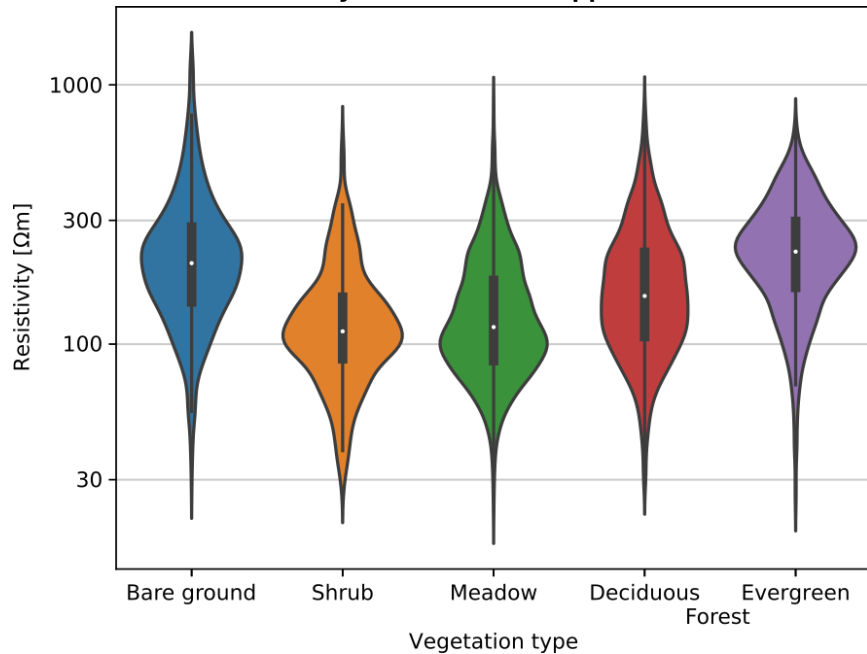


- Variability in groundwater EC linked to gradient in metamorphism of the bedrock
- Mancos Shale hydrological properties are fracture dominated

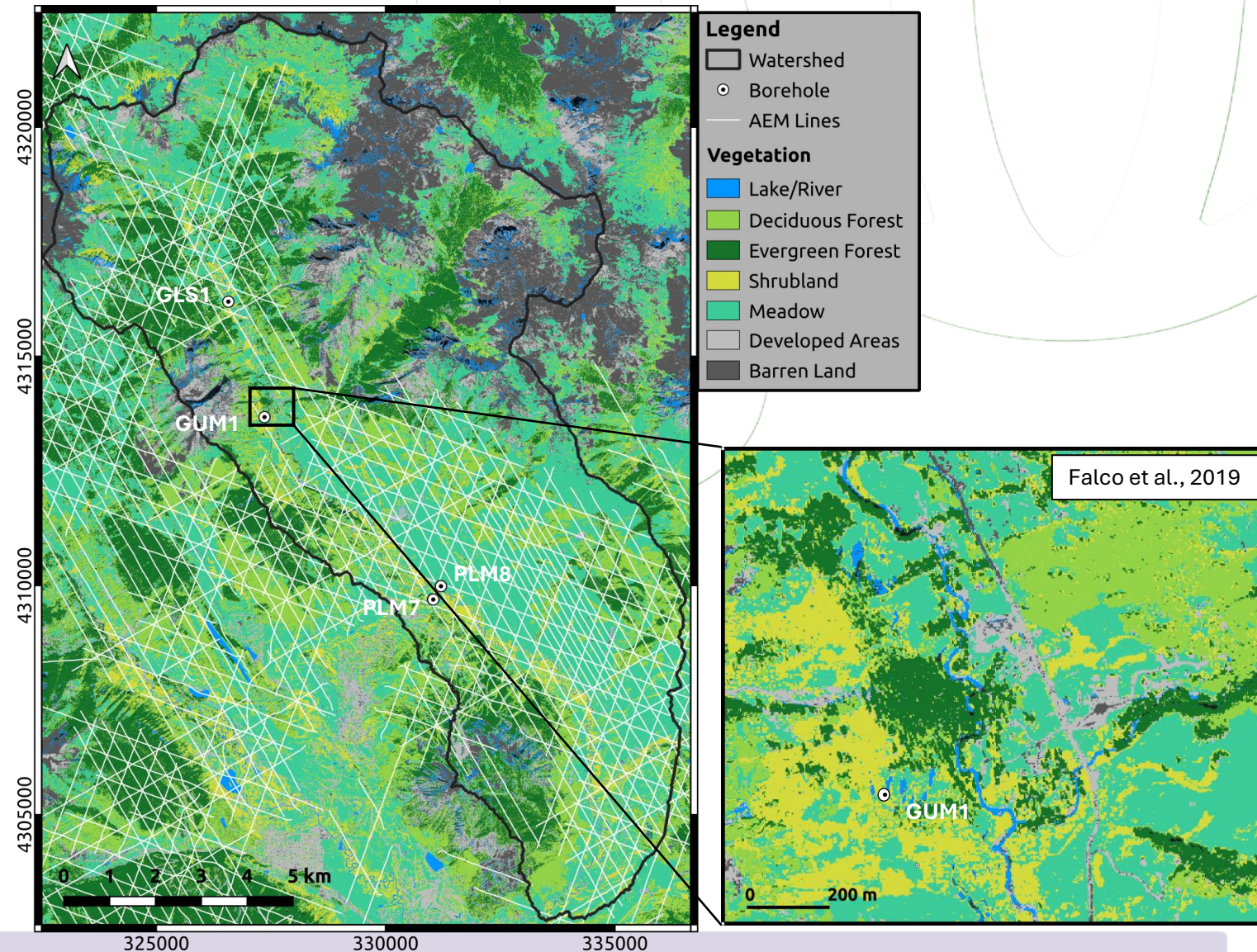
Linking surface and subsurface properties

Investigating links between vegetation and bedrock electrical resistivity

Resistivity variation in the upper 20 m



- Detailed vegetation map derived from hyperspectral data
- Low subsurface resistivity linked to vegetation with shallow roots
- Forest grows on more resistive = competent rock



Upscaling bedrock properties

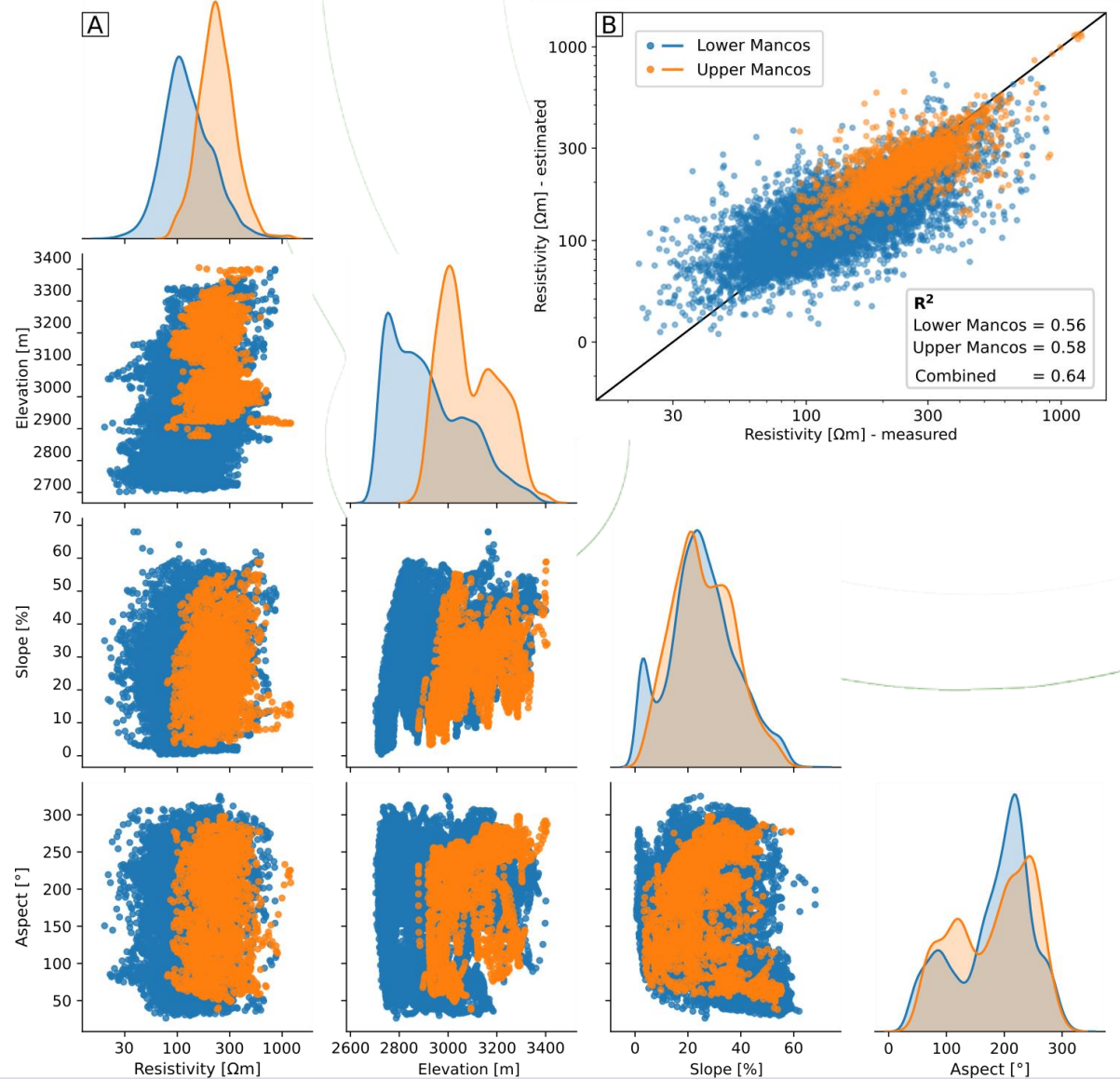
Feature extraction and prediction of bedrock properties

LiDAR data provides high-resolution topographic information that can be linked to bedrock properties

- Features = input data for ML
- Performances of various ML algorithms tested

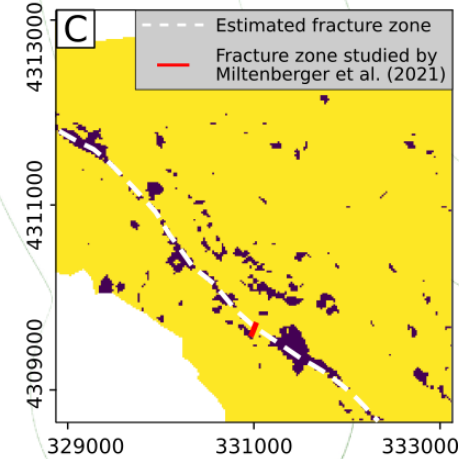
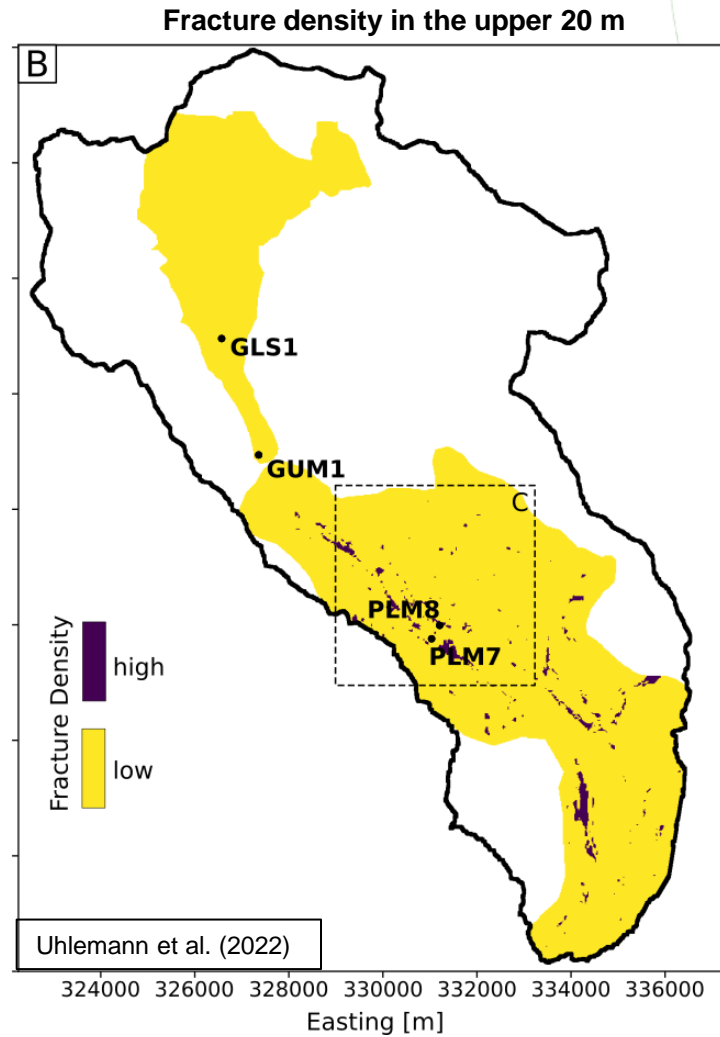
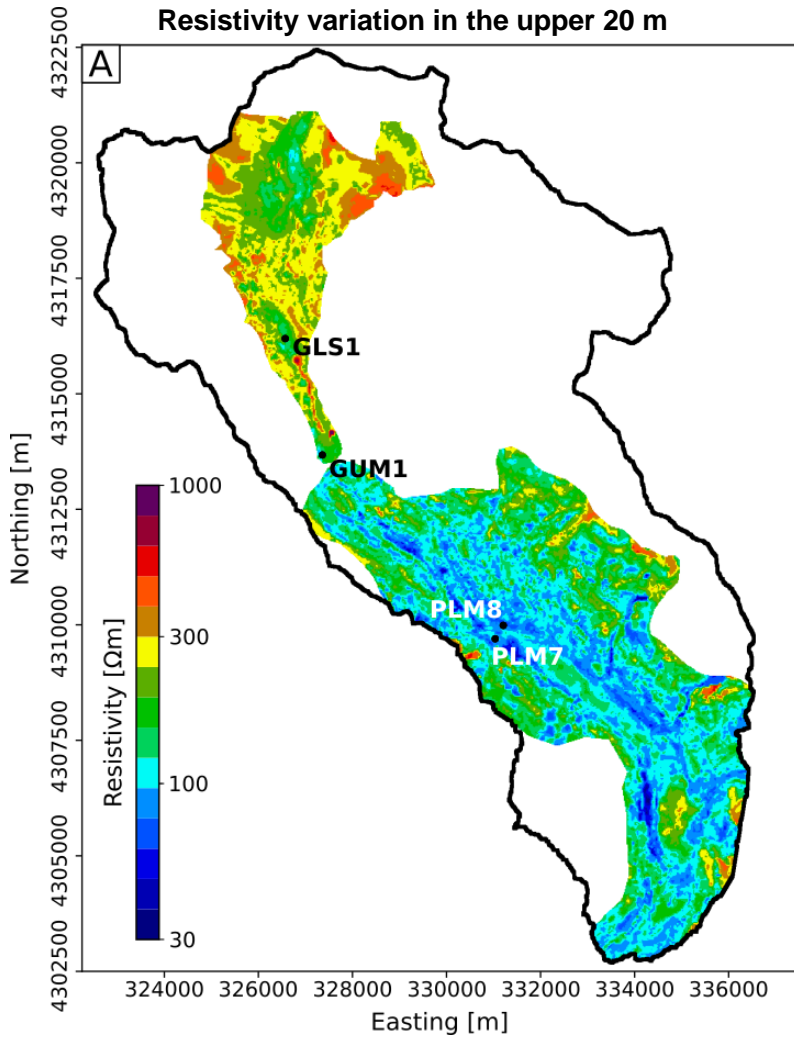
Extremely Randomized Trees regression to predict properties across the watershed

- Explains 64% of data variability – supports validity of many geomorphological models
- Differences related to inherent bedrock properties, like grade of metamorphism, mineralization, etc.



Upscaling bedrock properties

Estimating subsurface variability from remotely sensed data



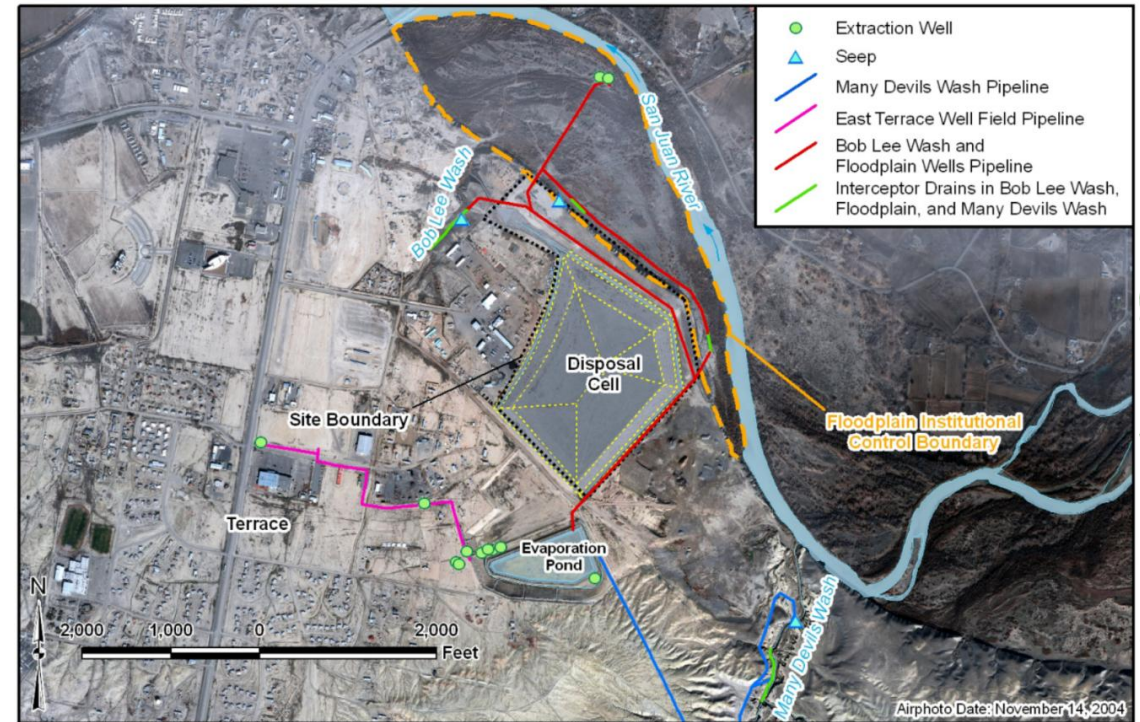
- Delineates previously unmapped fault line, associated to fracture zone
 - **Major control on local hydrology**
- Pronounced resistivity gradient can be associated to different geochemical environments
 - **Impacts seasonal variations in concentration – discharge relationships**

The Shiprock, NM, Uranium Milling Site

Site history and compliance strategy

Study site: The Shiprock, NM, Uranium Milling Site

- Former uranium- and vanadium-ore processing facility within the Navajo Nation in the northwest corner of New Mexico
- Operational between 1954 and 1968, creating process-related wastes and radioactive tailings
- In 1983 DOE and the Navajo Nation entered into an agreement for site cleanup
- **Compliance Strategy:**
 - Milling-related water from the eastern groundwater system of the terrace and the floodplain is pumped to an evaporation pond
 - Water samples are collected semi-annually
 - **Future: Construction of a water treatment unit for extracted groundwater and modular storage tanks for brine evaporation**



Geophysical measurements to improve the site model

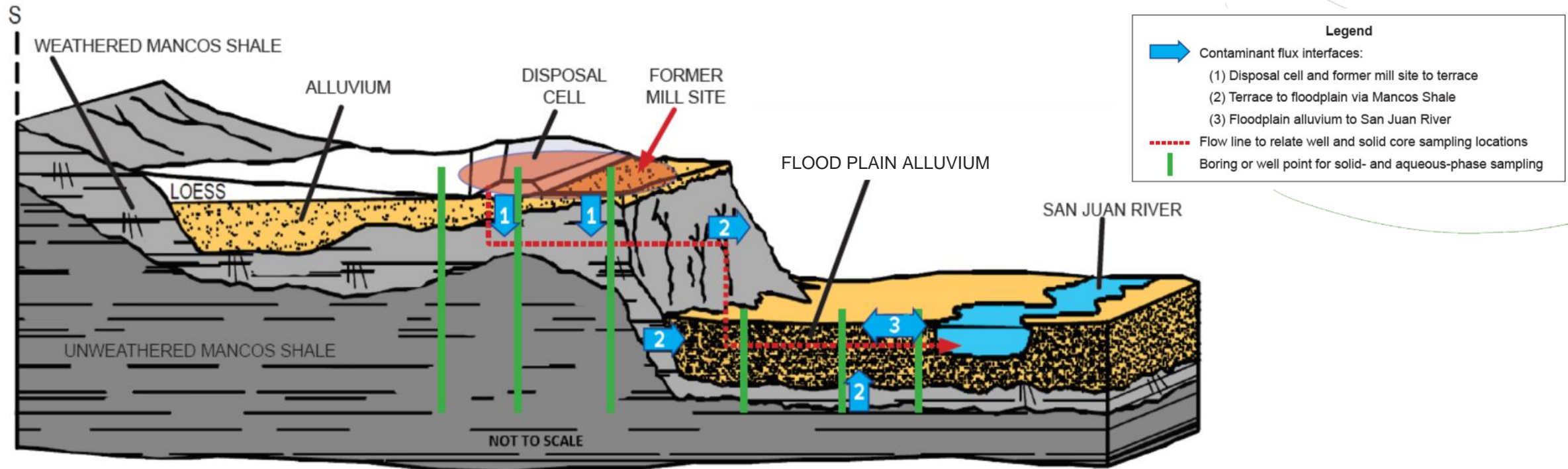
Conceptual site model and survey goals



Project goals

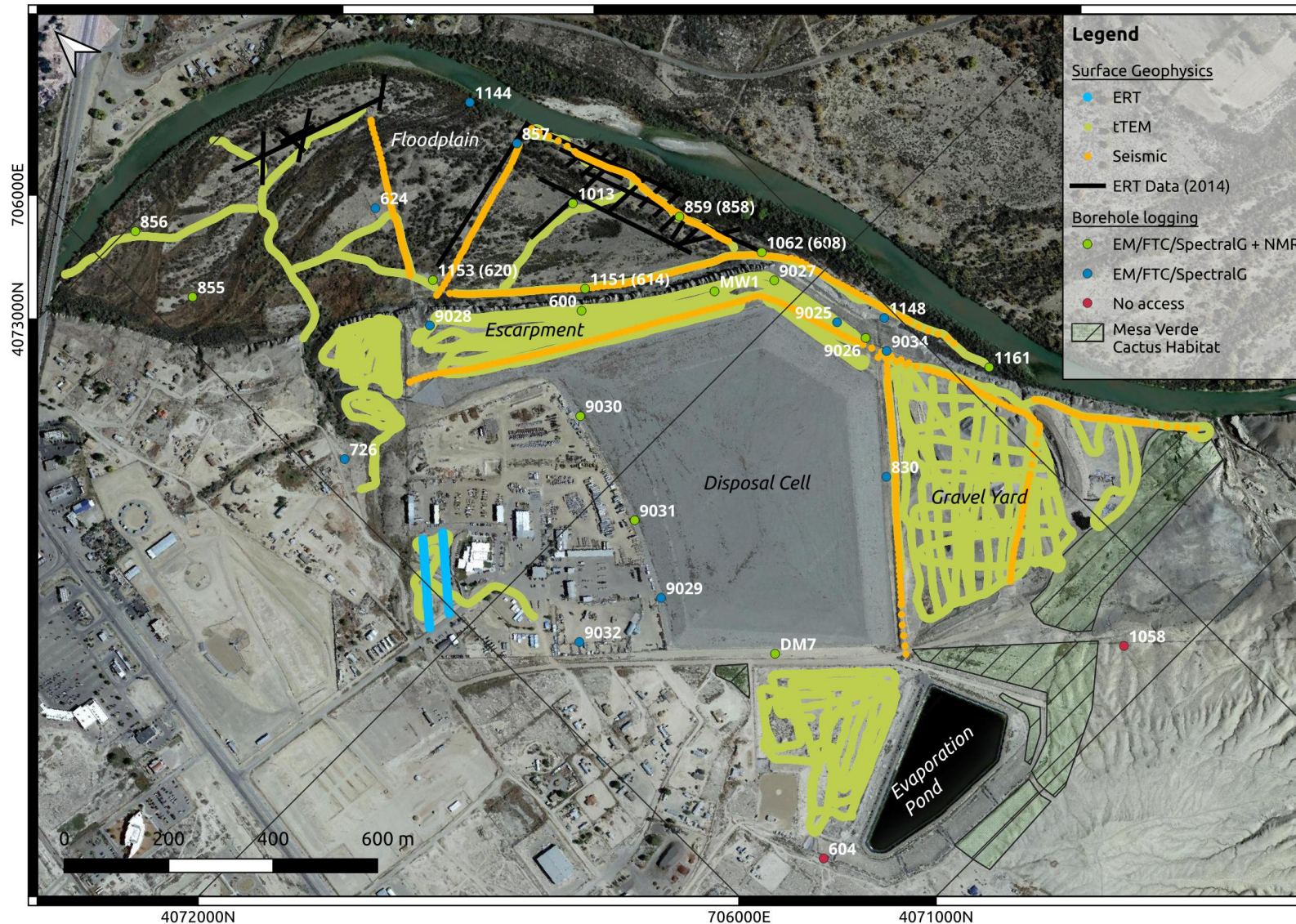
Identify location of saturated fractures and preferential flow paths transmitting dissolved contaminants from terrace to flood plain (incl. areas of elevated, anomalous electrical resistivity)

Establish a 3D hydrostratigraphic model, delineating various lithologies and providing estimates of hydraulic parameters



Dense sampling of subsurface conditions

Map showing location of surface and borehole geophysical measurements



Available Data:

Geophysics

- 20.2 km of tTEM data
- 24 ERT profiles (2 m electrode spacing)
- 9 seismic profiles (3.7 km, 5 - 10 m shot spacing)
- 26 boreholes with pore fluid EC&T, Spectral gamma, EM log
- 15 boreholes with NMR log

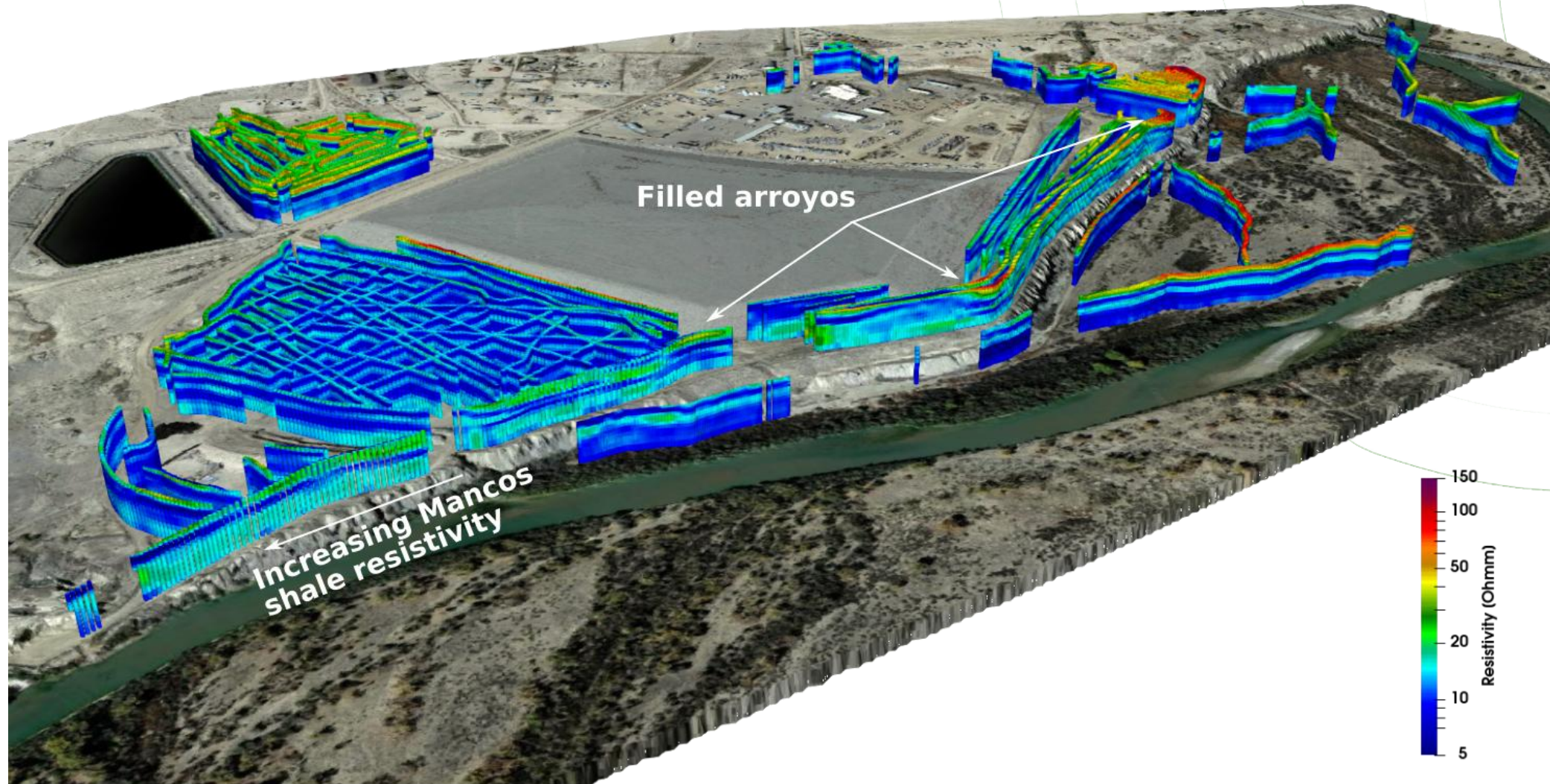
Lithology and Groundwater

- 65 boreholes with lithology logs within 25 m of tTEM & ERT data
- Groundwater depth during survey from 26 boreholes

Lots of data → great data set for ML!

Using electrical resistivity to map the lithology

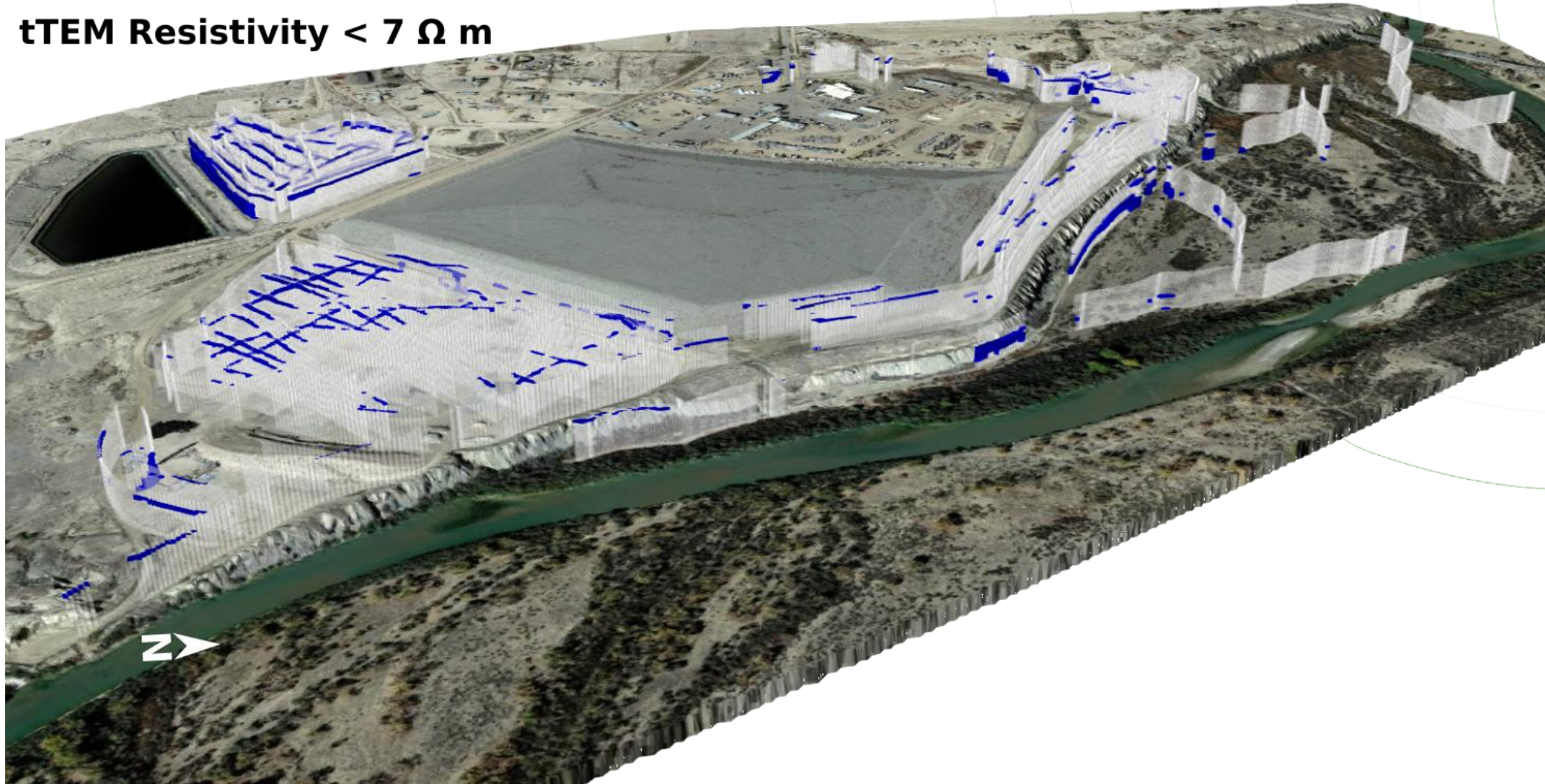
tTEM derived subsurface resistivity distribution using spatially-constrained inversion



Low resistivity feature may indicate contaminant location

Very low values mapped close to evaporation pond and on the floodplain

tTEM Resistivity < 7 Ω m

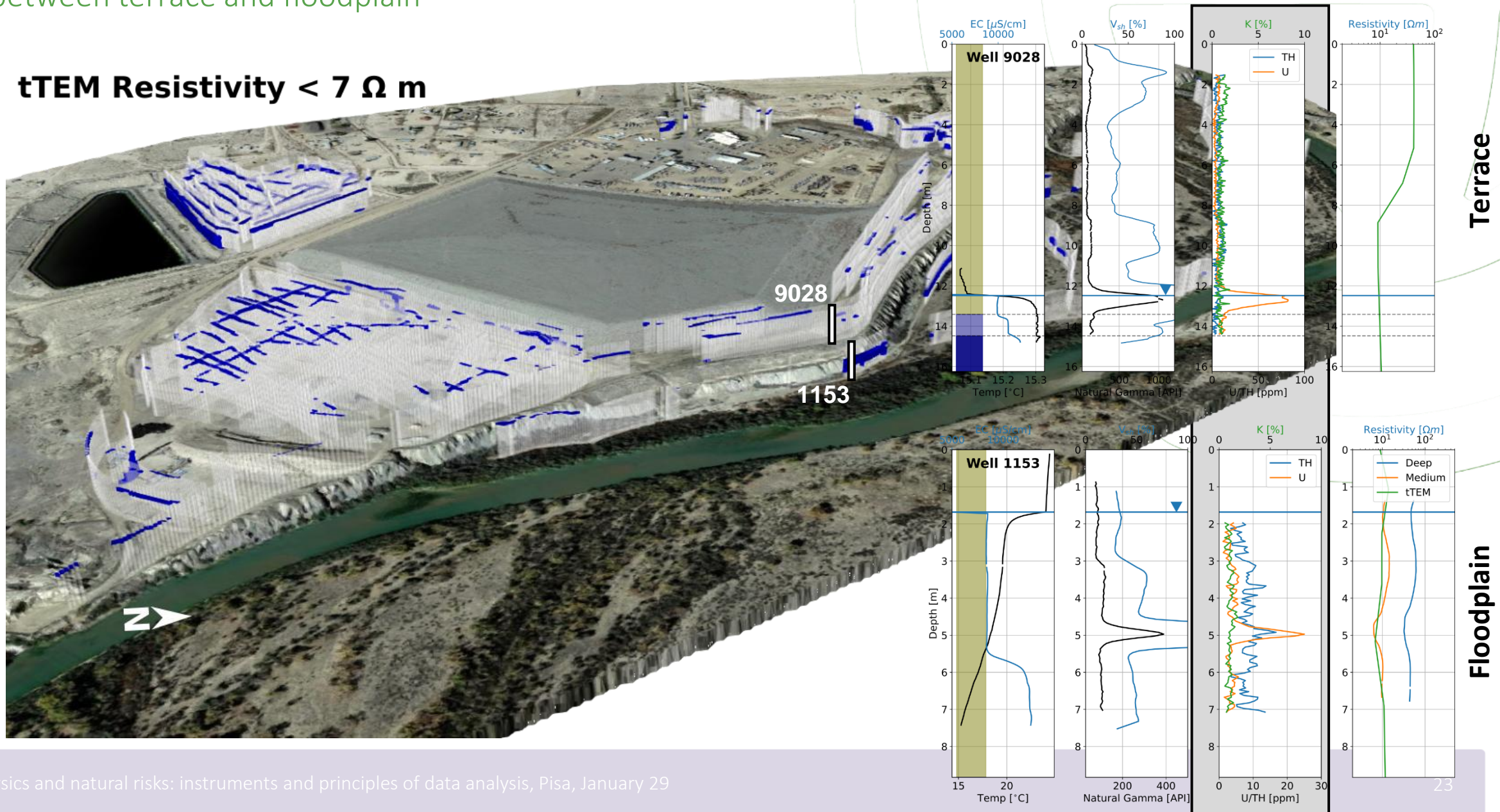


Borehole logging shows elevated contaminant concentration

Link between terrace and floodplain



tTEM Resistivity < 7 Ω m



Linking tTEM measurements with lithological logs

Exploiting feature relationships to predict subsurface lithology

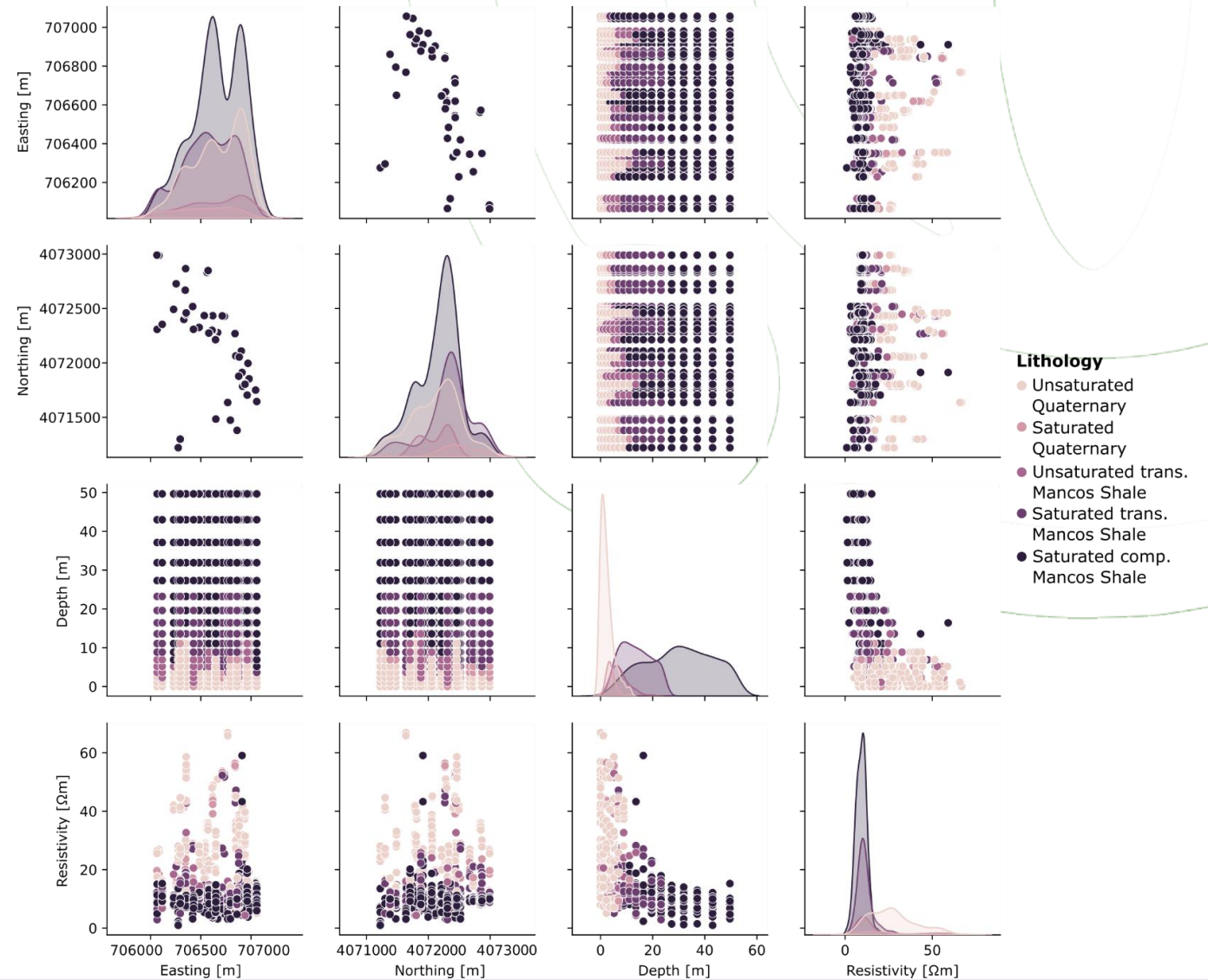
Extracted resistivity and lithology profile from tTEM data at 65 borehole locations

Classified lithology into 5 types:

- Unsaturated/saturated Quaternary deposits
- Unsaturated/saturated transmissive Mancos Shale
- Saturated non-transmissive Mancos Shale

Trained ML algorithms on standardized, balanced (SMOTE & Random oversampling) data set

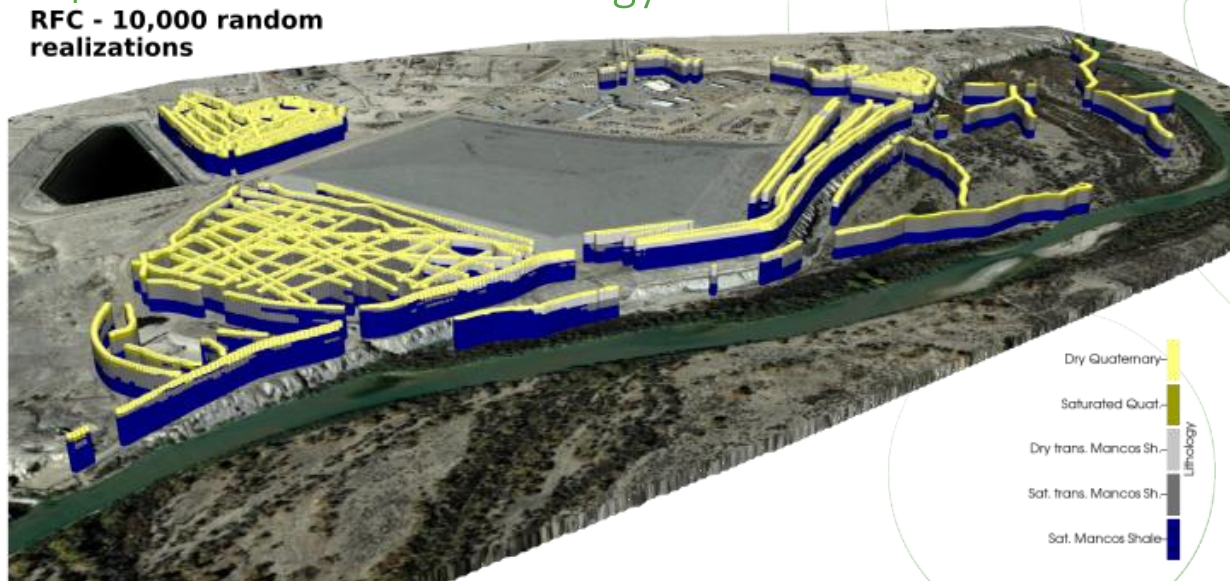
- Random Forest and Neural Net (Multi-layer perceptron) showed best performance
- Prediction uncertainty quantified by subsampling, training, and predicting 10,000 times



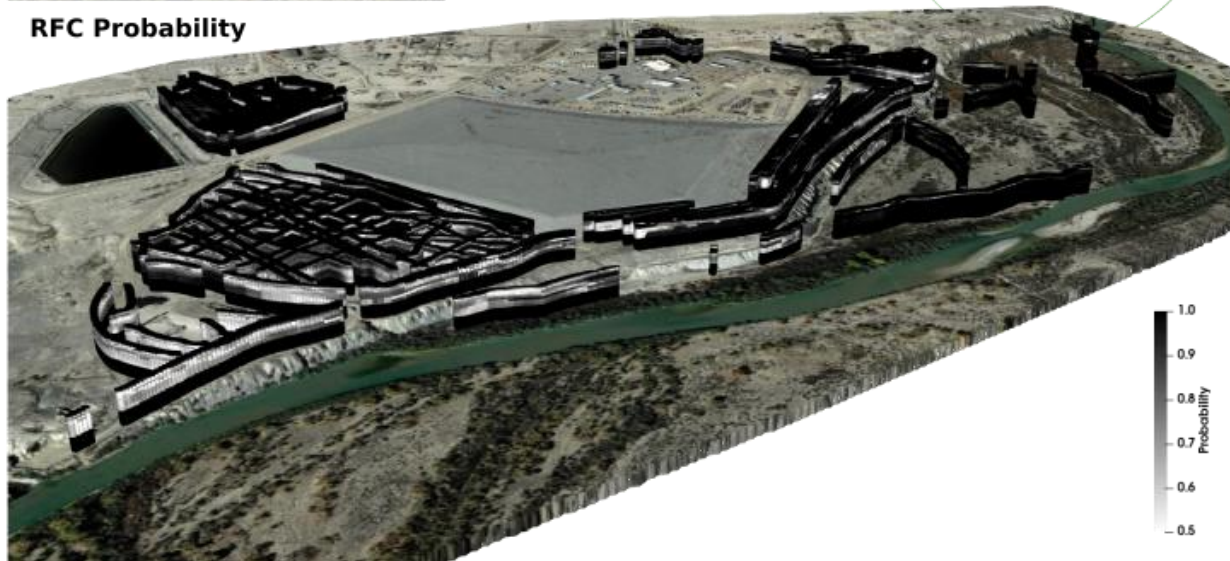
Linking tTEM measurements with lithological logs

Exploiting feature relationships to predict subsurface lithology

Predicted lithology

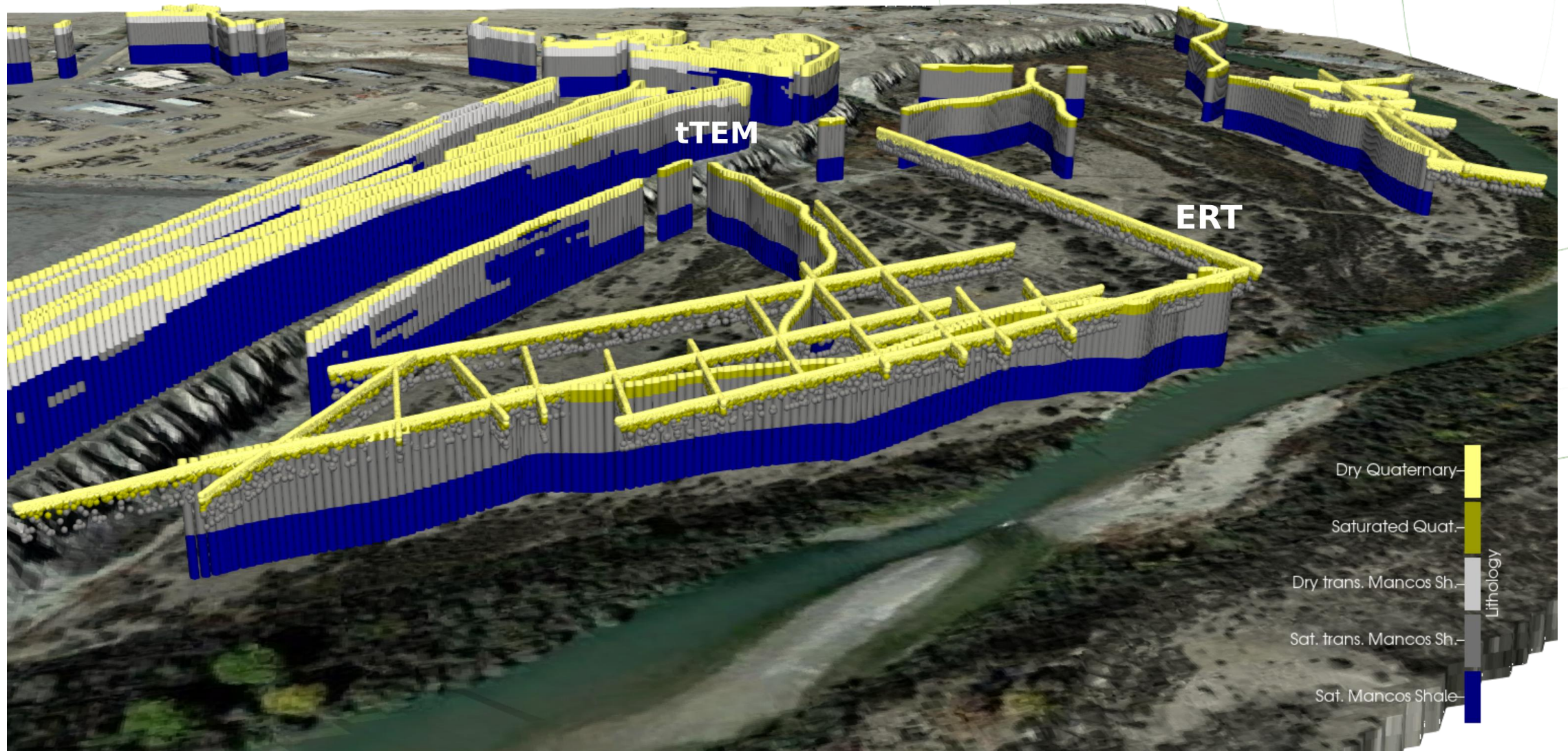


Prediction probability



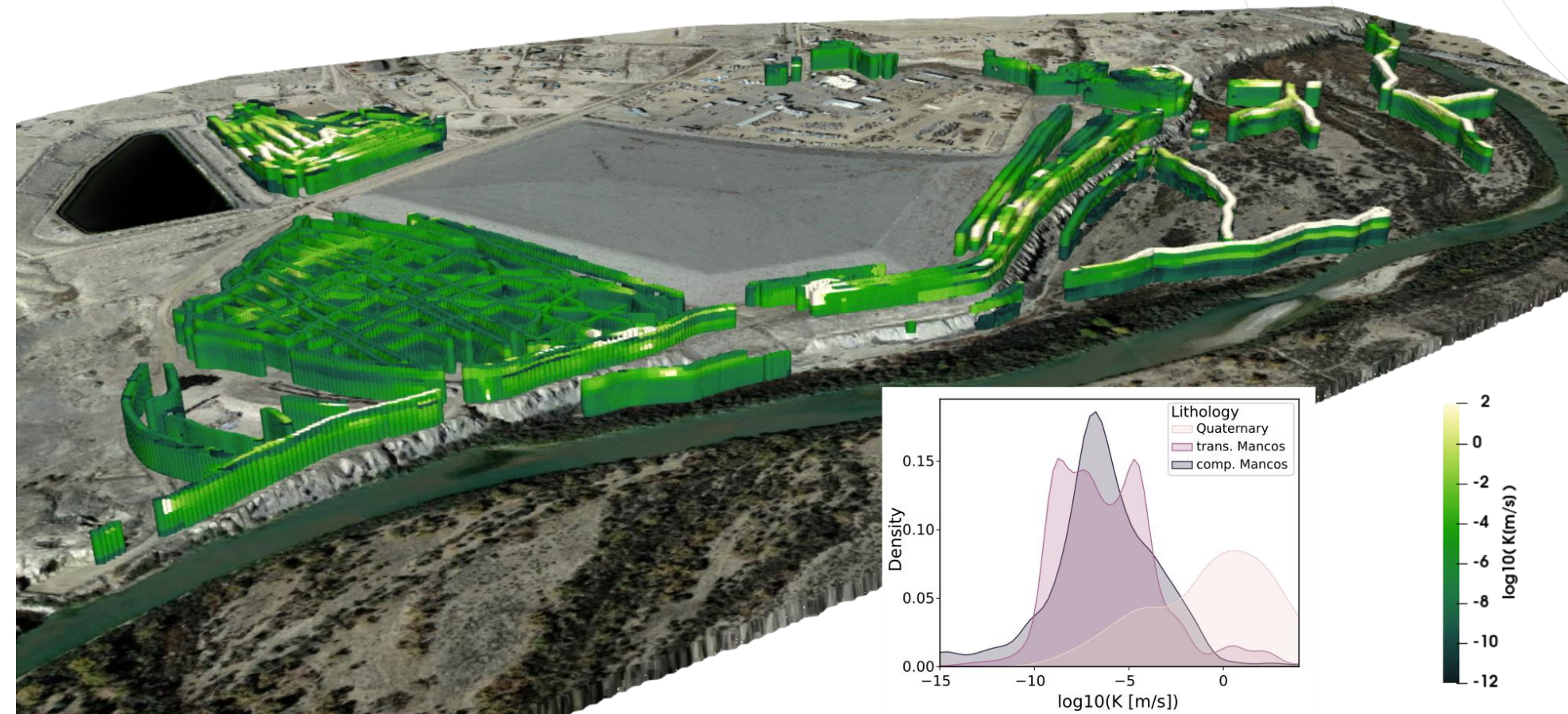
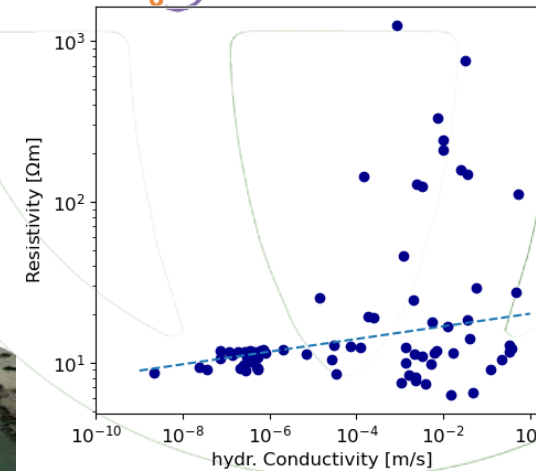
Linking tTEM measurements with lithological logs

Comparison of lithological prediction from tTEM and ERT data



Estimating hydraulic conductivity

Linking borehole logs and tTEM resistivity to obtain estimated of K_{sat}



Geophysics for supporting hydrological model development



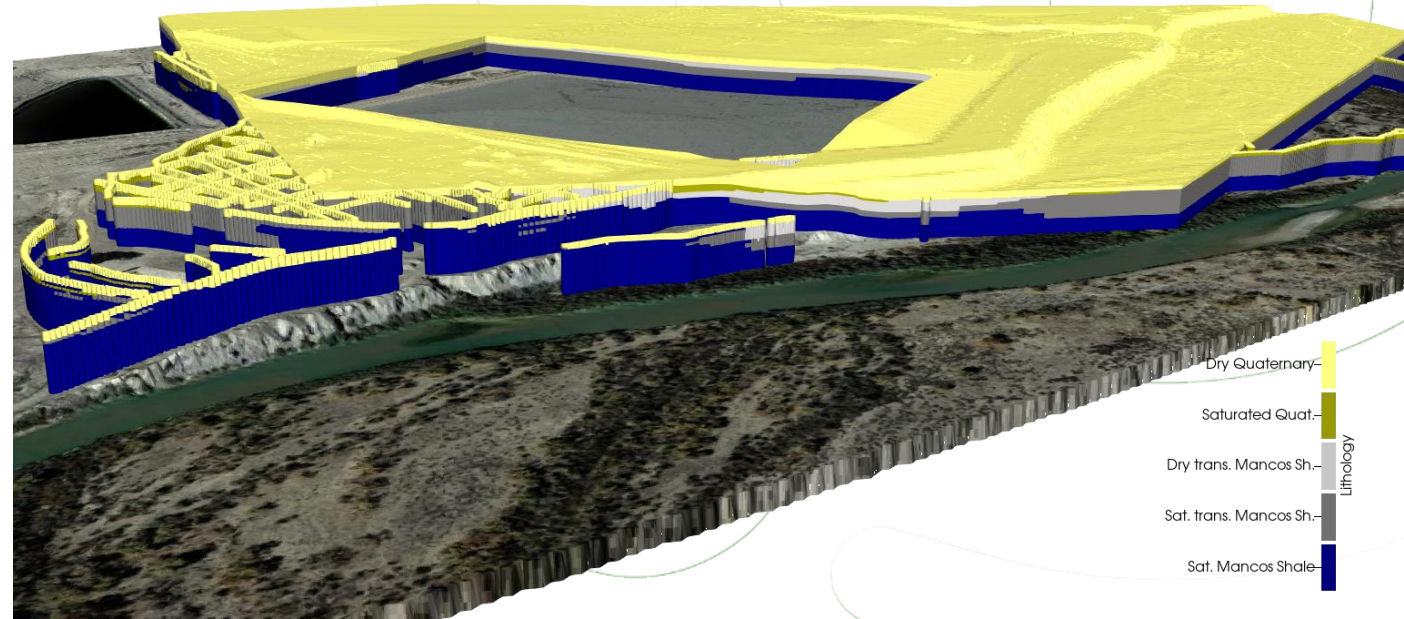
High-resolution surface and borehole geophysics to obtain 3D property distribution

Integrated and ML-supported multi-geophysics approach successful in achieving project goals

- Created 3D hydrostratigraphic model
- Provided distribution of hydraulic parameters

Low resistivity anomalies indicate areas of potential remaining contamination

Randomized ML approach allows to provide prediction probability, highlighting areas of uncertainty



Machine learning for data assimilation and interpretation

Summary



Machine learning is an efficient tool when analysing:

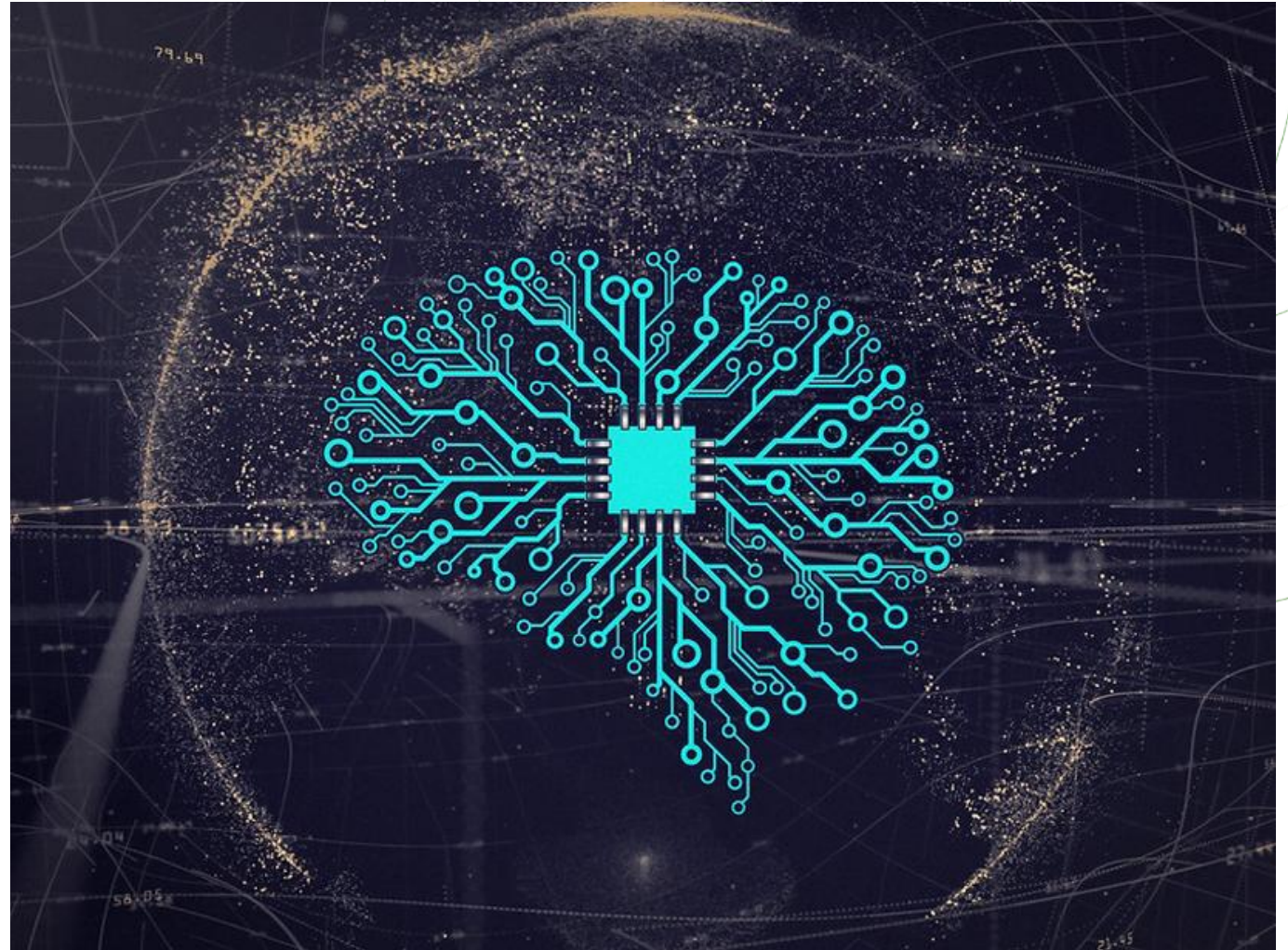
- Large data sets
- Multivariate datasets

Be careful about input features:

- Information content (linear dependence between various features)
- Potential biases

**Skewed input data can seriously affect predictions
- make sure you work with balanced datasets**

**Carefully test and optimize various ML approaches
– are they providing just a good fit, or are they transferable?**





THANKS!

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System
(D.D. n. 130/2022 - CUP B53C22002150006) Funded by EU - Next Generation EU PNRR-
Mission 4 “Education and Research” - Component 2: “From research to business” - Investment
3.1: “Fund for the realisation of an integrated system of research and innovation infrastructures”

