

Laying the groundwork for a greener future: ML for characterizing and managing geologic reservoirs

Velimir (“monty”) Vesselinov
Hope Jasperson
Trais Kliphuis



Motivation

- Geologic reservoirs are critical for our renewable energy future
- Reservoirs are applied for different geologic engineering projects:
 - Water Supply
 - Oil & Gas Extraction
 - Carbon Sequestration
 - Geothermal Production
 - Energy Storage
 - Mineral Mining
 - Rare-Earth-Element Extraction
 - Hydrogen Production
 - Waste Disposal
- Multi-use of reservoirs is frequently needed
(e.g., simultaneous geothermal production and energy storage)

Motivation

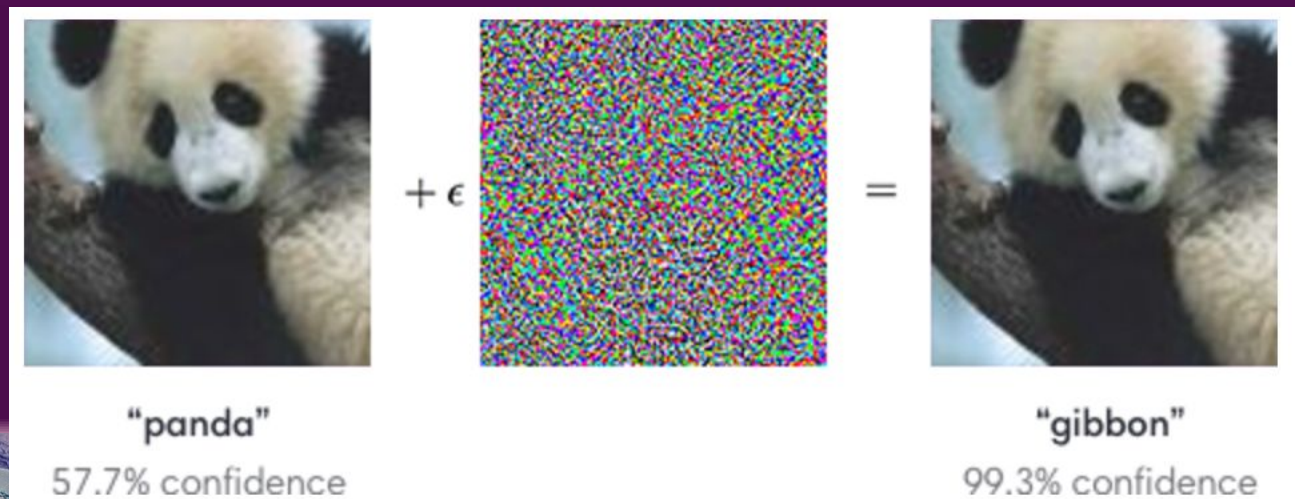
- Characterization and utilization of geologic reservoirs are challenging
- Data are limited
- Some of the governing processes are poorly understood
- Negative environmental impacts can be triggered by reservoir engineering activities:
 - public health and safety
 - induced seismicity
 - ground-surface deformation
 - wildlife and ecosystem impacts
 - contamination
 - fluid leaks

Motivation

- Optimization of the reservoir use and minimization of environmental impact are critical
- Data assimilation and model prediction are challenging
- ML methods and tools can help
- To address these challenges, we develop ML methods and tools that will be provided as a SaaS (Software-as-a-Service) on the cloud
- Our work focuses on:
 - Unsupervised (self-supervised) ML
 - Supervised ML
 - Science-Informed ML (SIML)

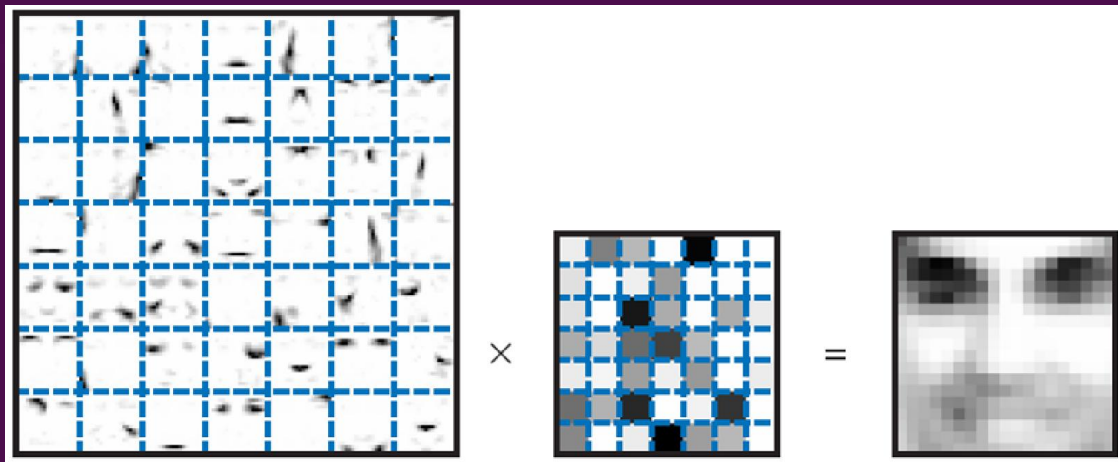
Supervised ML

- learns everything from data
- cannot discover/learn something that is not known already
- requires large training datasets
- highly impacted by noise in the data
- ML models are difficult to interpret \Rightarrow black boxes
- can recognize cats and dogs but cannot recognize horses if not pre-trained



Unsupervised ML

- extracts information (features) from data automatically
- produces unbiased analyses
- identifies features characteristic for different animals (cats, dogs, horses, etc.)
- finds features discriminating between different geologic reservoirs/events
- can help data labelling for training of supervised ML

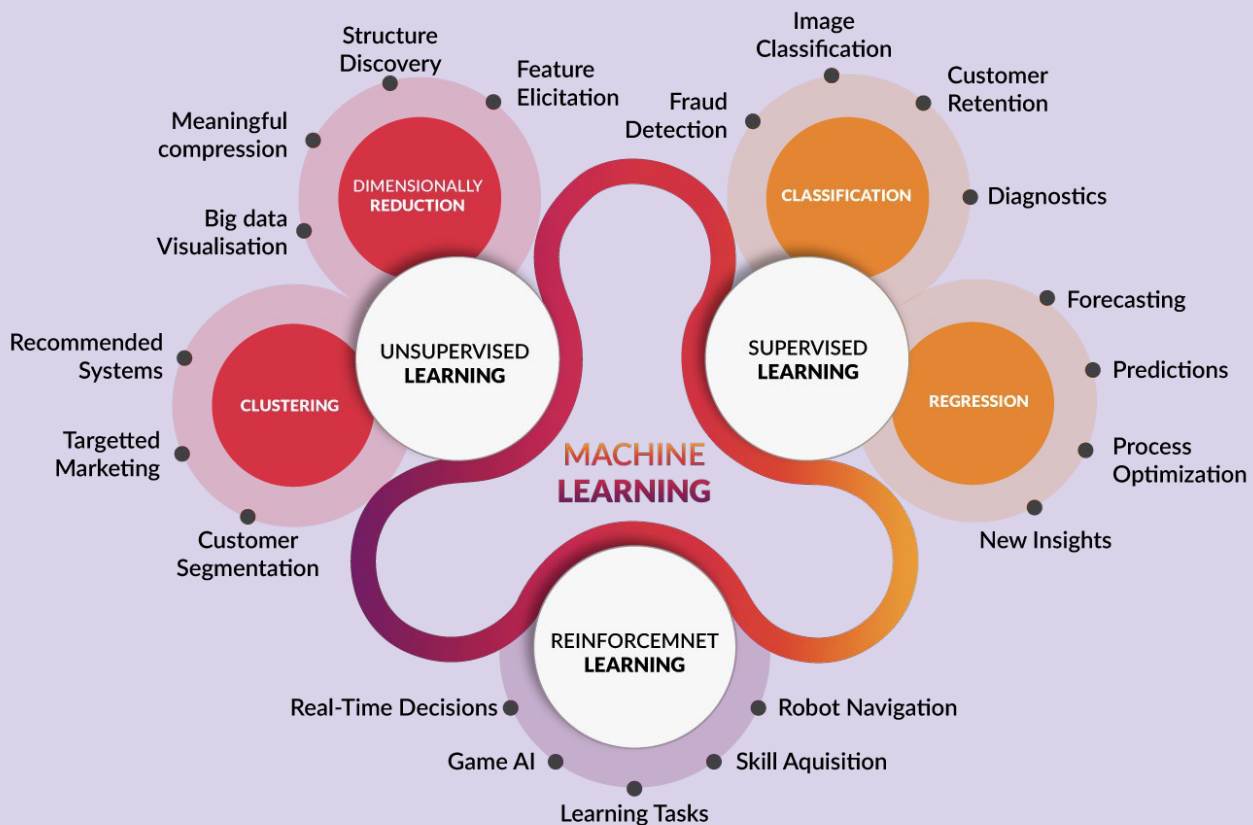


Science (Physics) Informed ML (SIML)

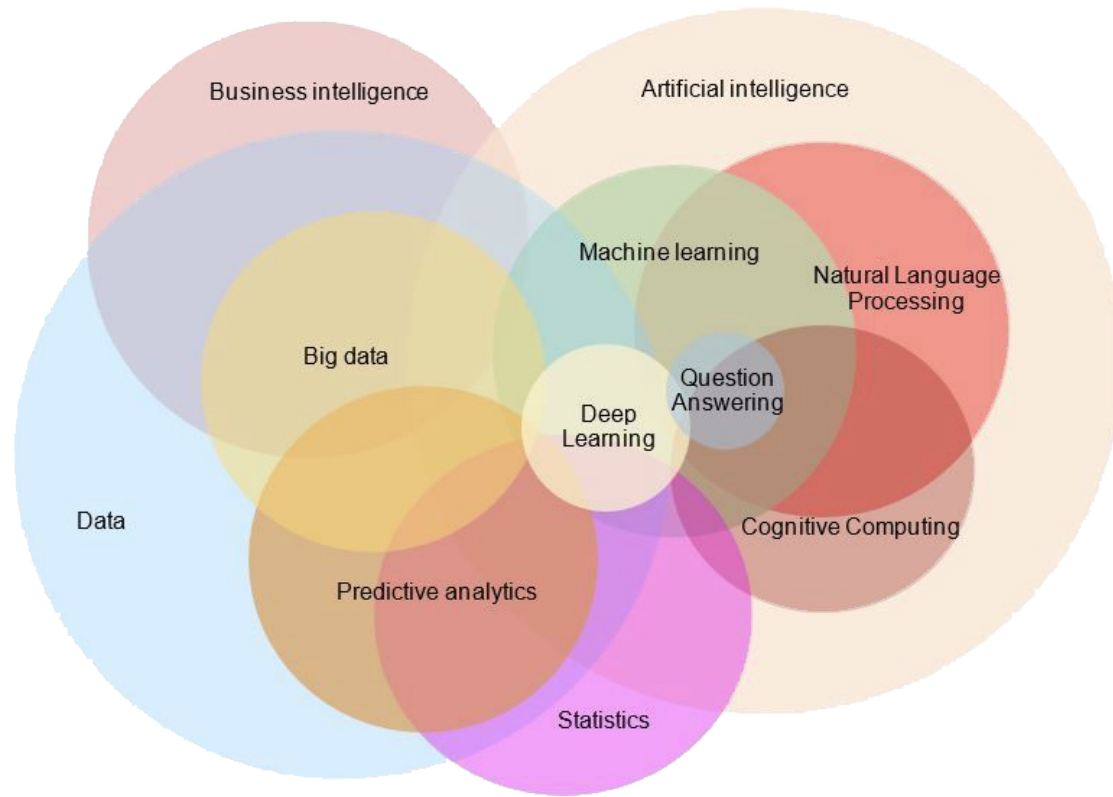
- learns from data but includes preconceived science knowledge
- SIML frameworks are problem specific
- subject-matter expertise is needed before and after ML is performed
- increases efficiency, accuracy, and robustness (scientific defensibility)
- requires differentiable programming

The Julia logo consists of the word "julia" in a lowercase, black, sans-serif font. Above the letters are five colored circles: a blue circle above the 'j', a green circle above the 'u', an orange circle above the 'l', a purple circle above the 'i', and a light purple circle above the 'a'.

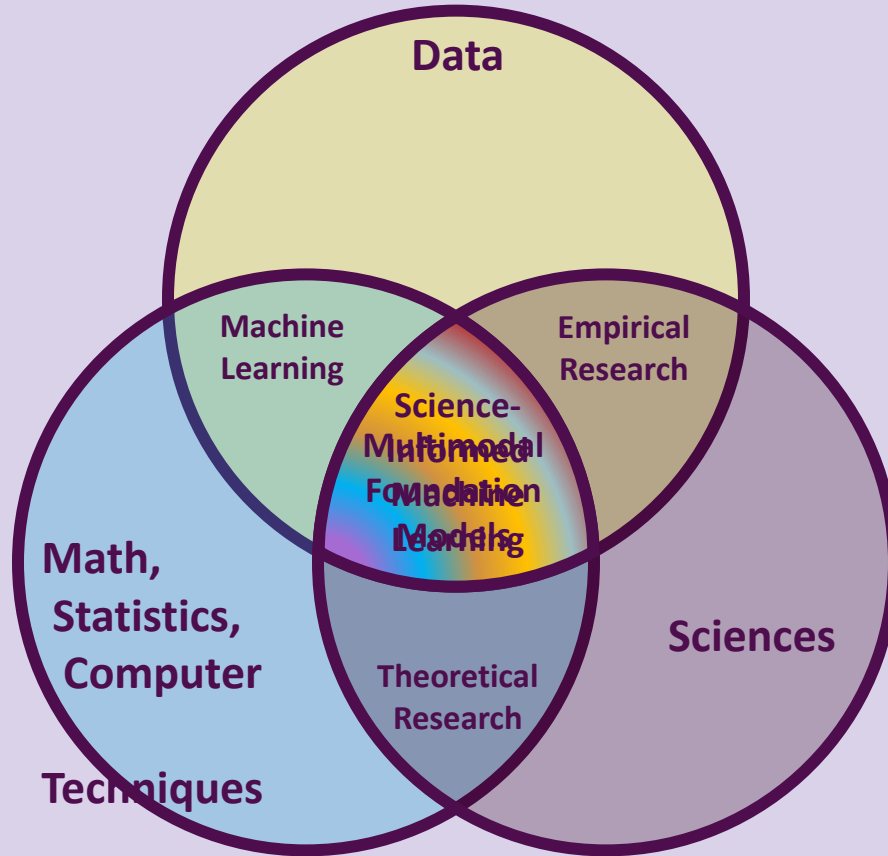
ML methods

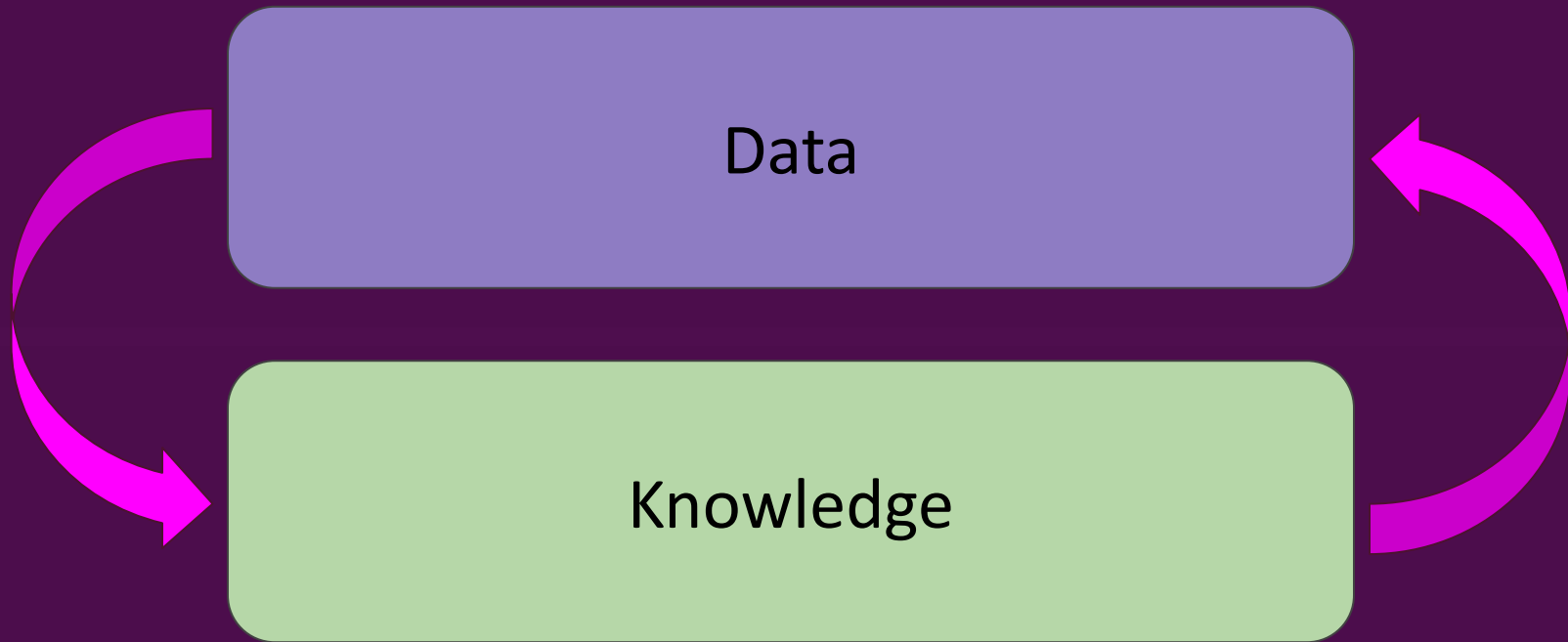


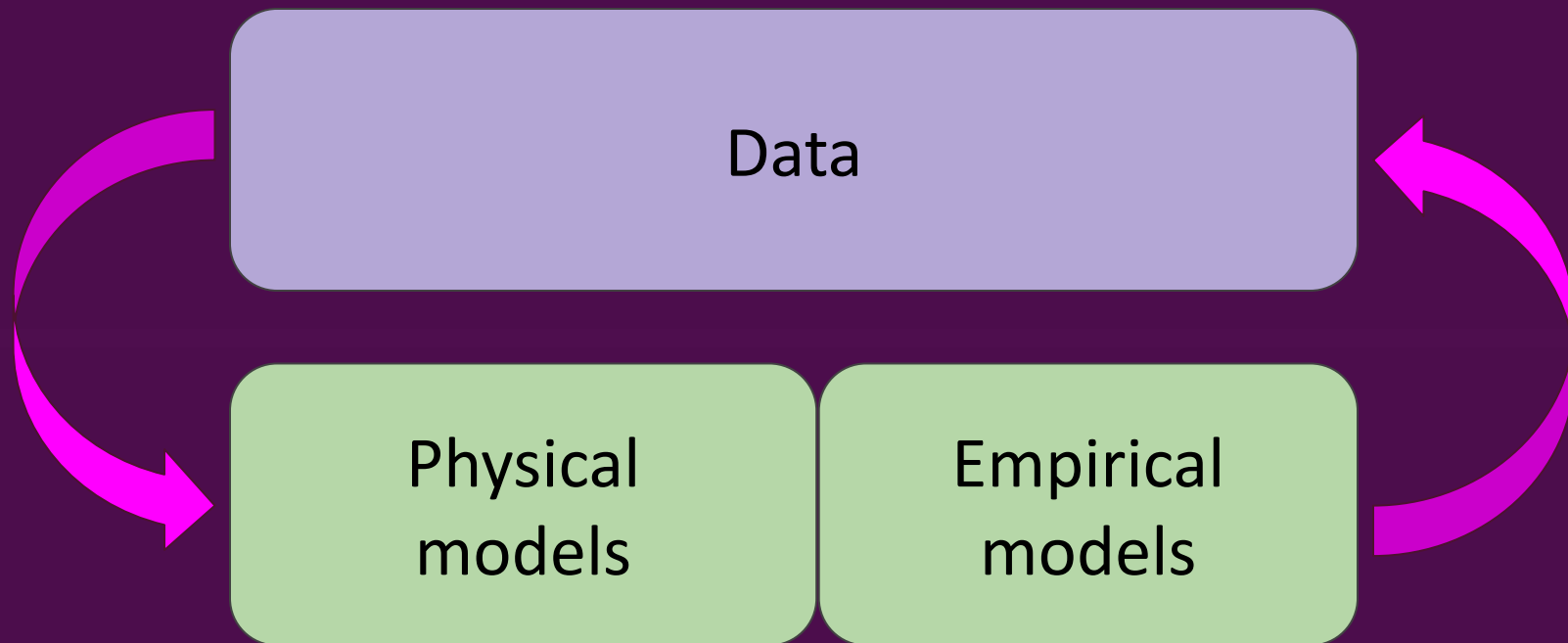
ML Venn diagrams (there are so many) ...

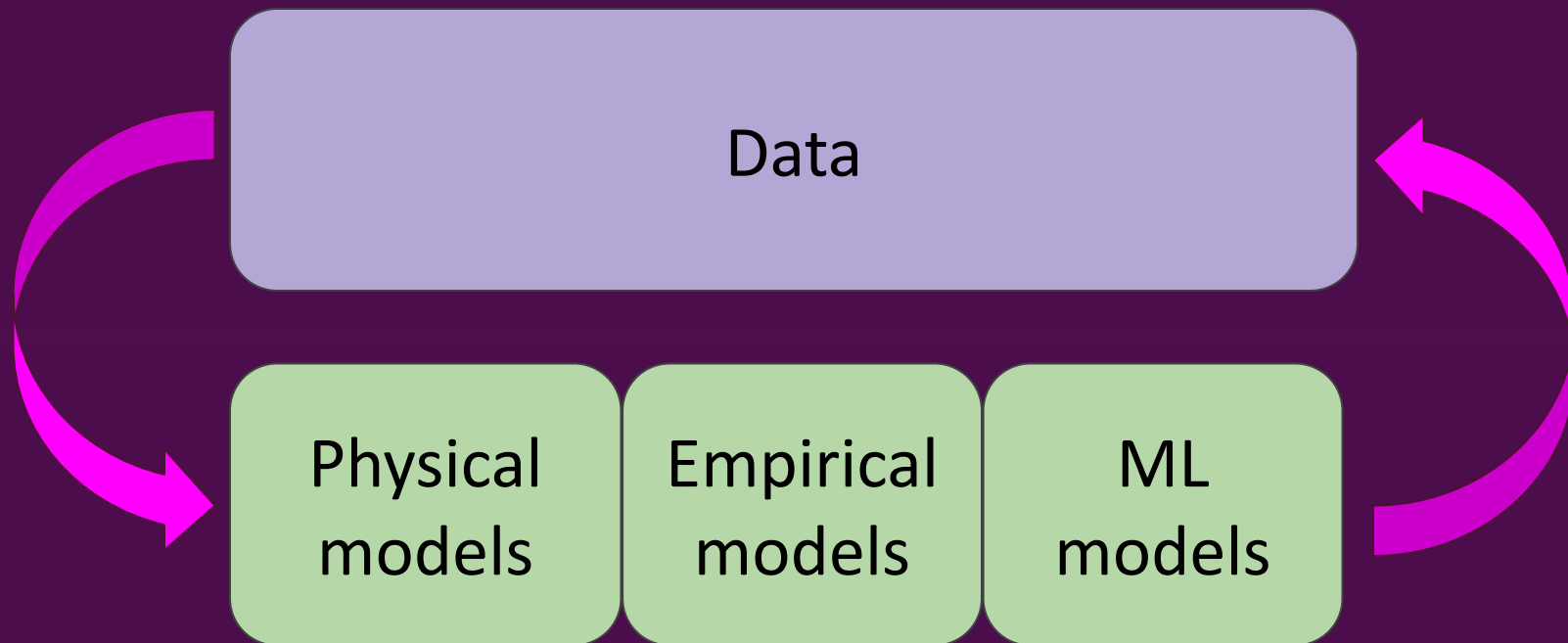


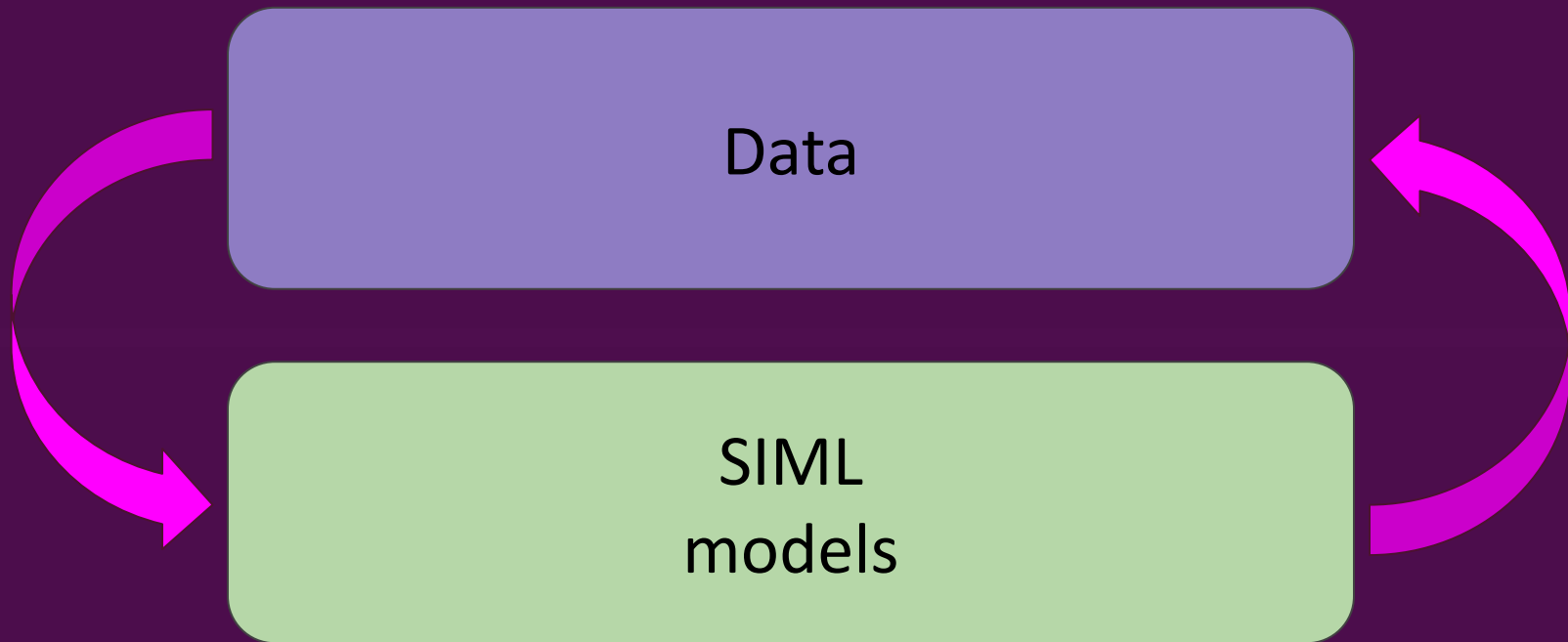
Science-Informed ML



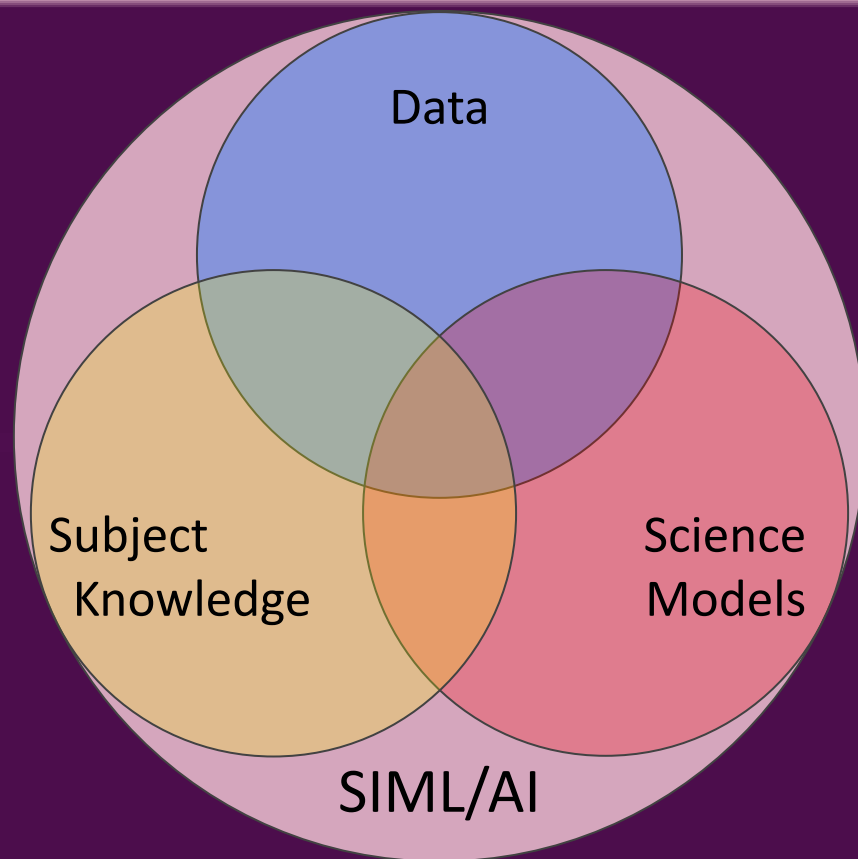








Science-Informed AI/ML



Subject knowledge:

- Domain knowledge & expertise
- Best practices
- Operating procedures & instructions
- Text/voice interactions
- Execution processes & logistics

Science models:

- Theories, laws, axioms
- Numerical/analytical simulators

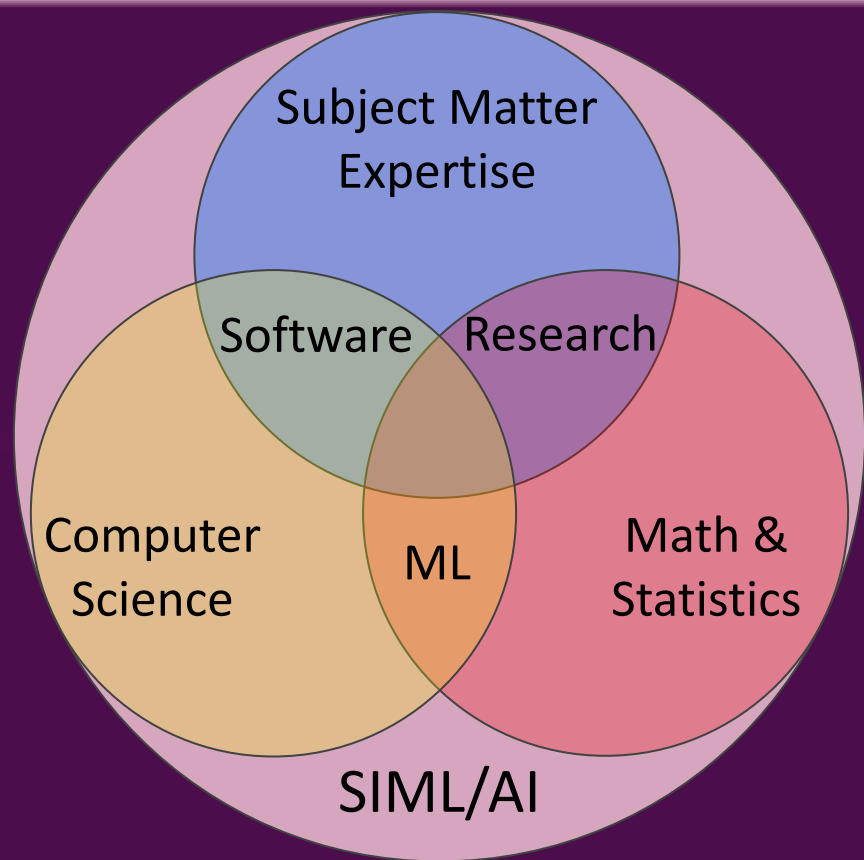
Data + Subject knowledge:

- Scientific/engineering discoveries
- Events detection
- Inference

Data + Science models:

- Data-driven validation of science-based models
- Discover new governing processes from data analysis

Science-Informed AI/ML

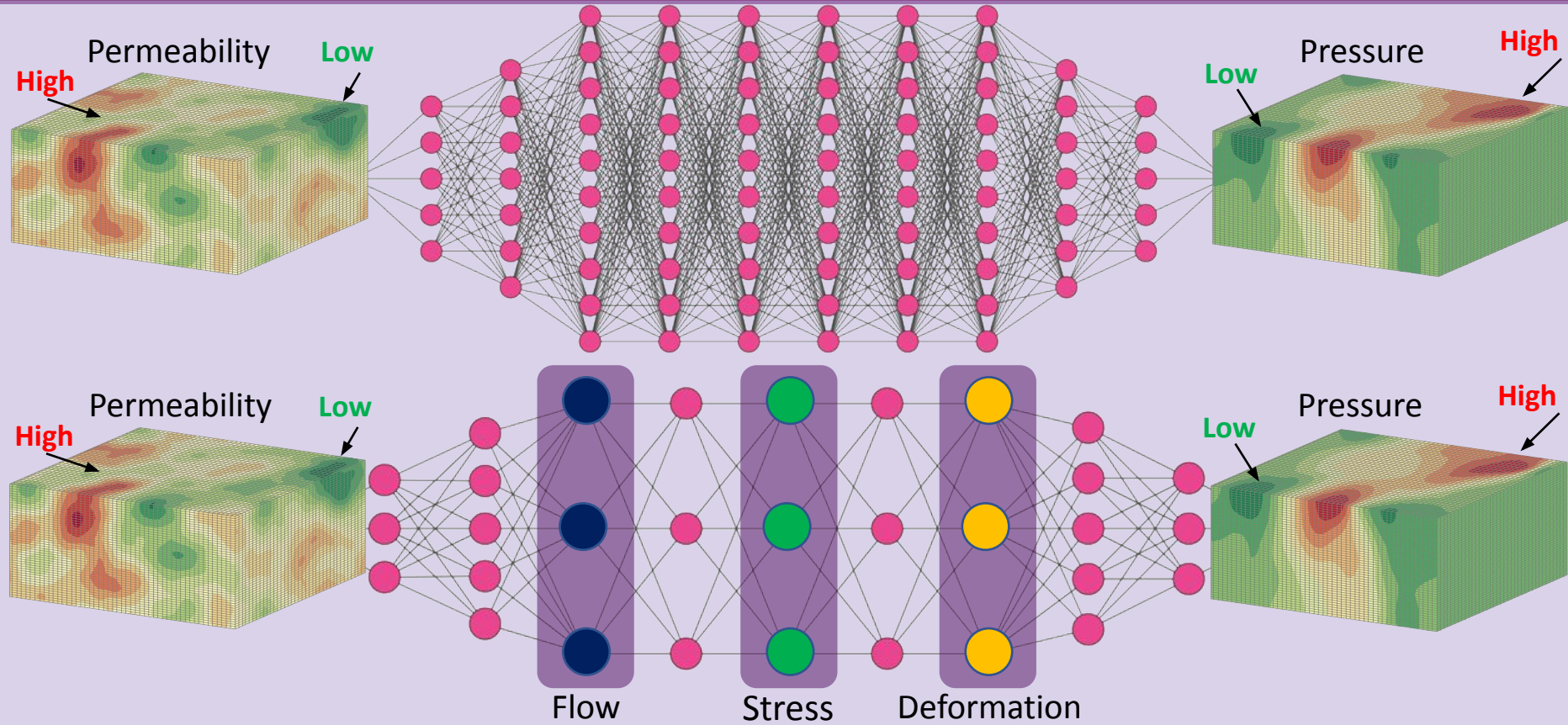


How to incorporate science knowledge/information in ML?

- Implicitly in the training process
(e.g., merging real-world and synthetic data, neural operators)
- Implicitly in the loss function
(e.g., adding regularization penalties and constraints)
- Explicitly in the ML machinery
(e.g., embedding numerical models in the neural networks)



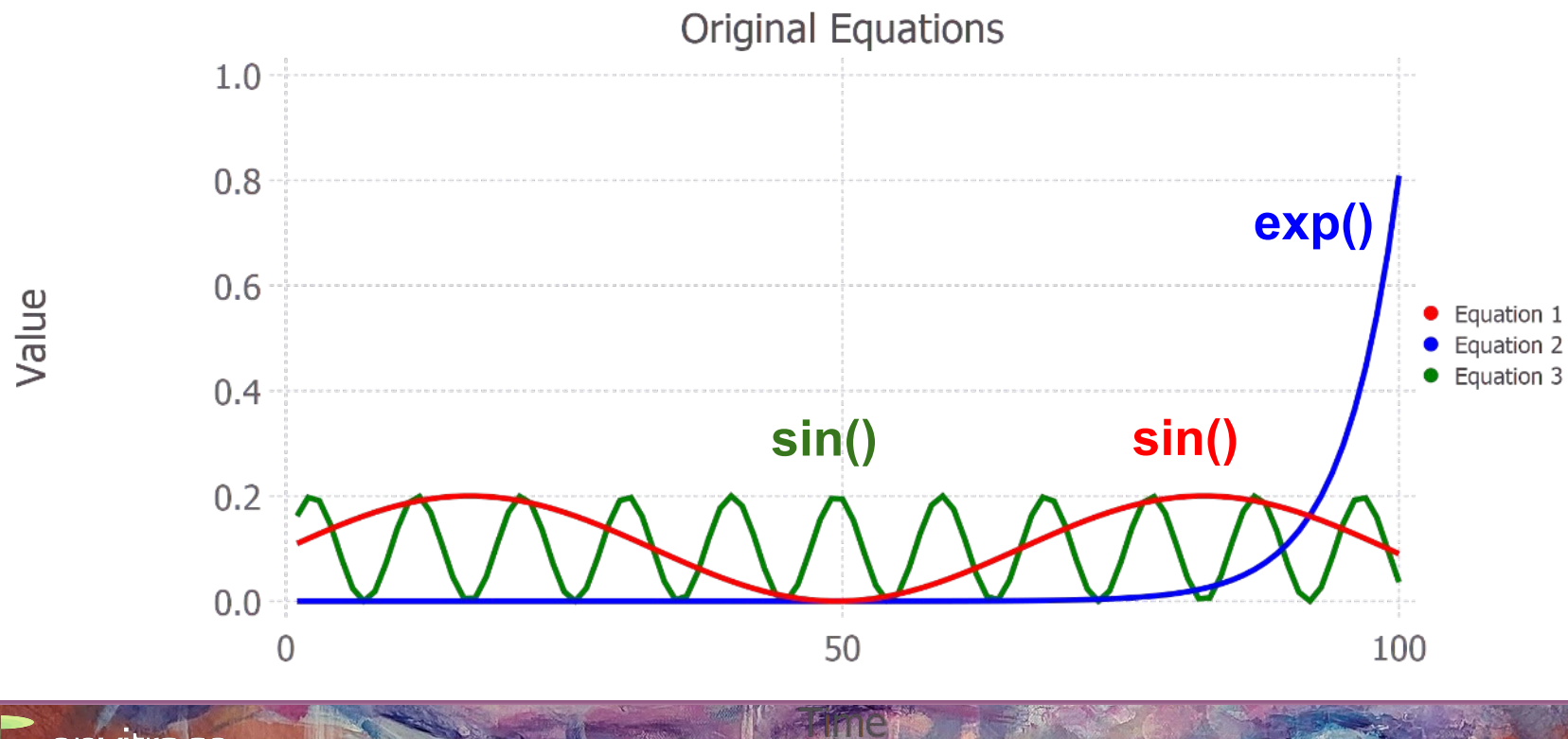
Explicit SIML



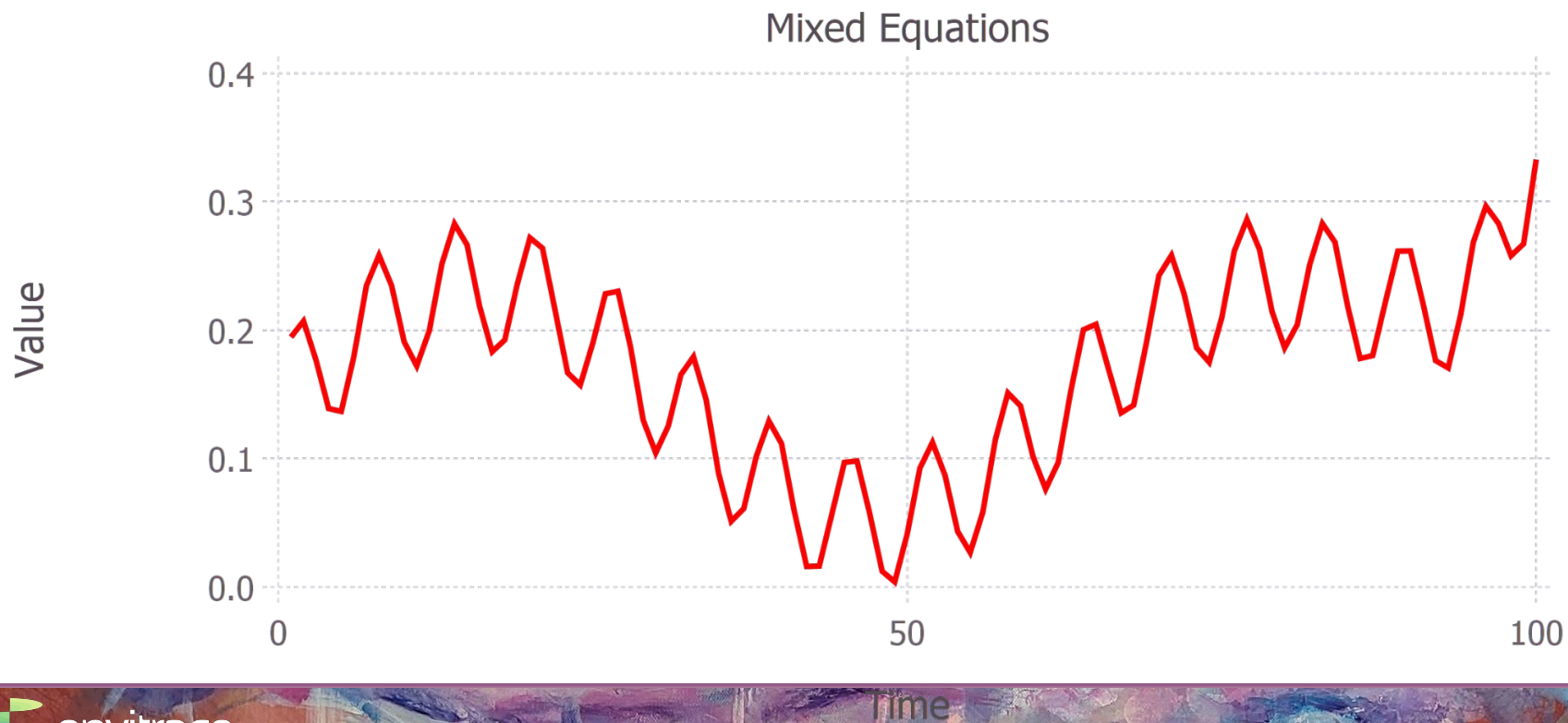
SIML examples

- Groundwater contaminant transport
 - Vesselinov et al., Nonnegative Tensor Factorization for Contaminant Source Identification, Journal of Contaminant Hydrology, [10.1016/j.jconhyd.2018.11.010](https://doi.org/10.1016/j.jconhyd.2018.11.010)
 - Vesselinov et al., Contaminant source identification using semi-supervised machine learning, Journal of Contaminant Hydrology, [10.1016/j.jconhyd.2017.11.002](https://doi.org/10.1016/j.jconhyd.2017.11.002)
- Oil & Gas Production

Traditional vs Science-informed ML (SIML)

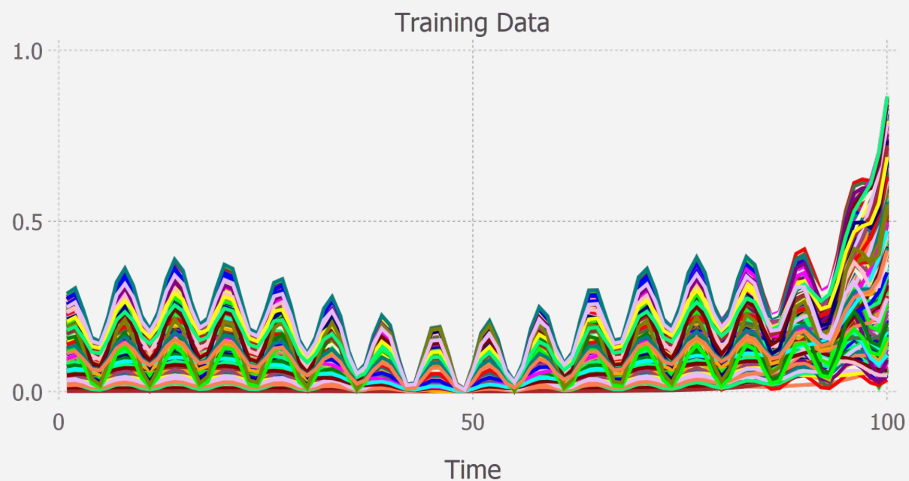


Traditional vs Science-informed ML (SIML)

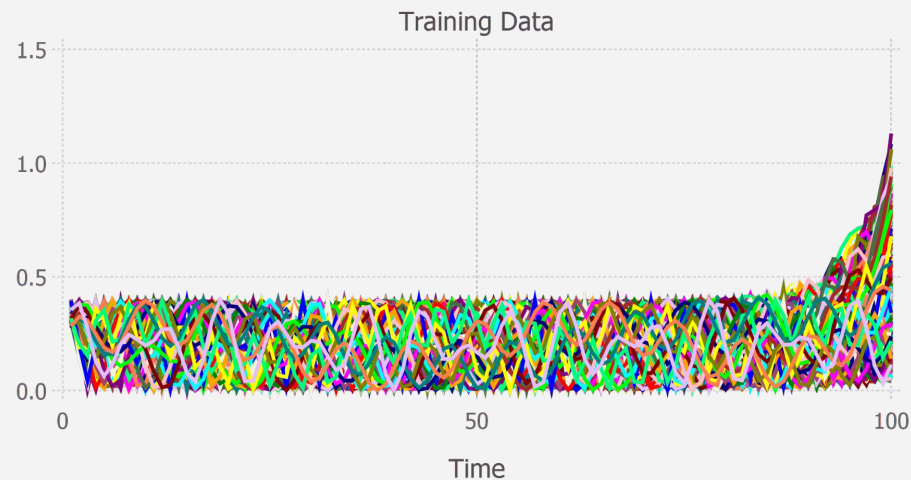


Traditional vs Science-informed ML (SIML)

Training Data (500 random samples) for 2 different cases:



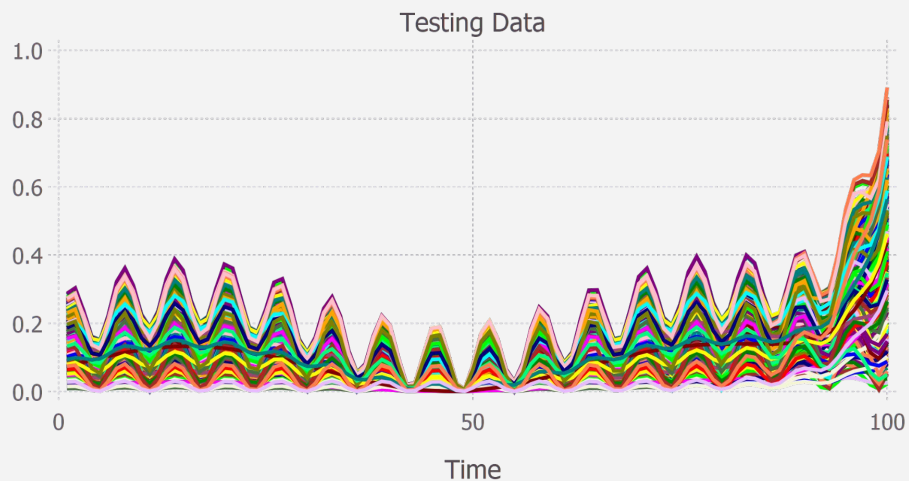
Case: "Amplitudes"



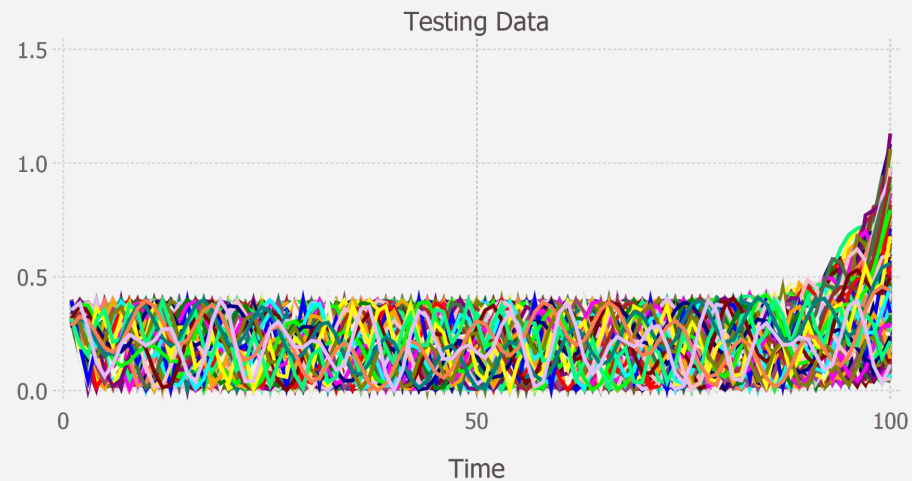
Case: "Frequencies"

Traditional vs Science-informed ML (SIML)

Testing Data (500 random samples)



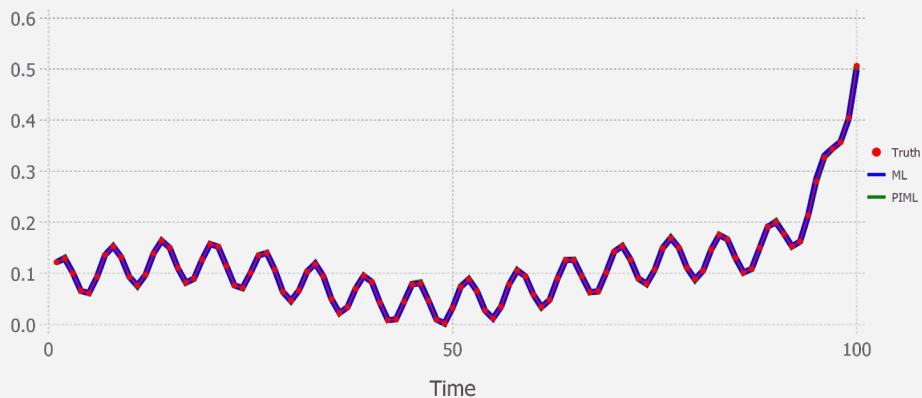
Case: "Amplitudes"



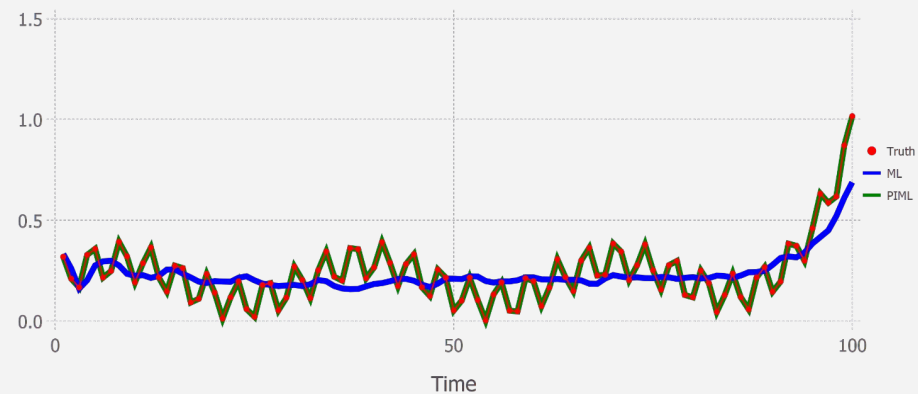
Case: "Frequencies"

Traditional vs Science-informed ML (SIML)

Comparisons between the true end estimated models using ML and SIML



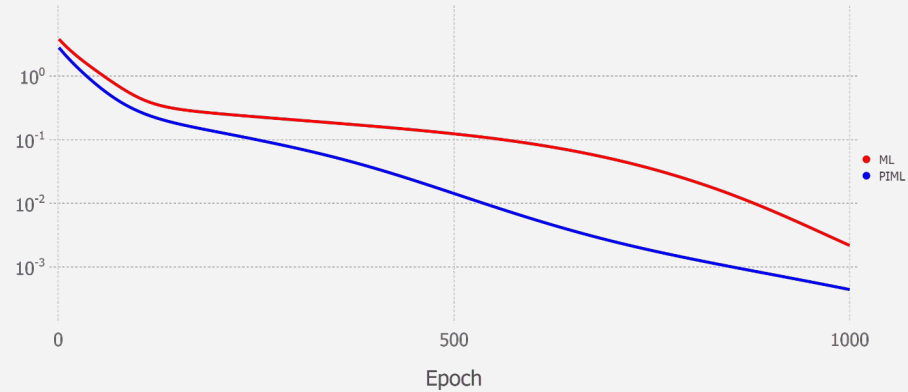
Case: "Amplitudes"



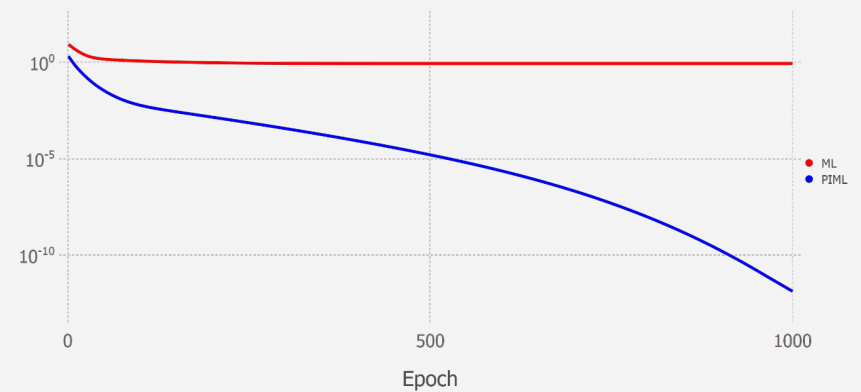
Case: "Frequencies"

Traditional vs Science-informed ML

Loss function decline over training epochs



Case: "Amplitudes"



Case: "Frequencies"

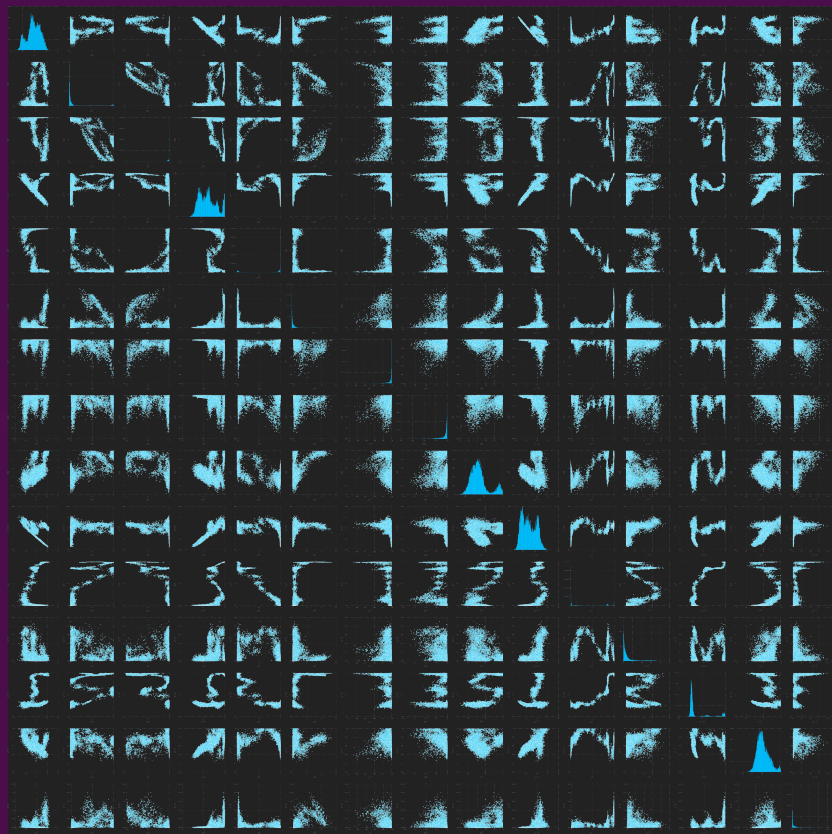
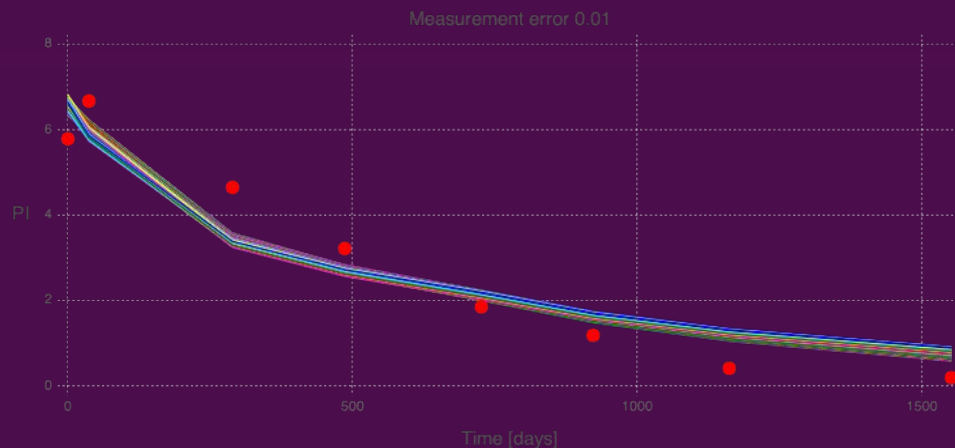
SIML for Oil & Gas Production

- GIMI.jl (Geomechanics-Informed Machine-Learning Intelligence; proprietary)
- Science-information provided by:
 - Synthetic data
 - MCMC outputs for 4 numerical models: 469 MCMC runs * 4 models
 - 51 inputs (static and transient) => 8 transient outputs
 - Geomechanics equations (regarding key flow and stress relations; 16)
 - Proprietary analytical expressions (4)
- Data for a series of existing oil production well analyzed (proprietary)



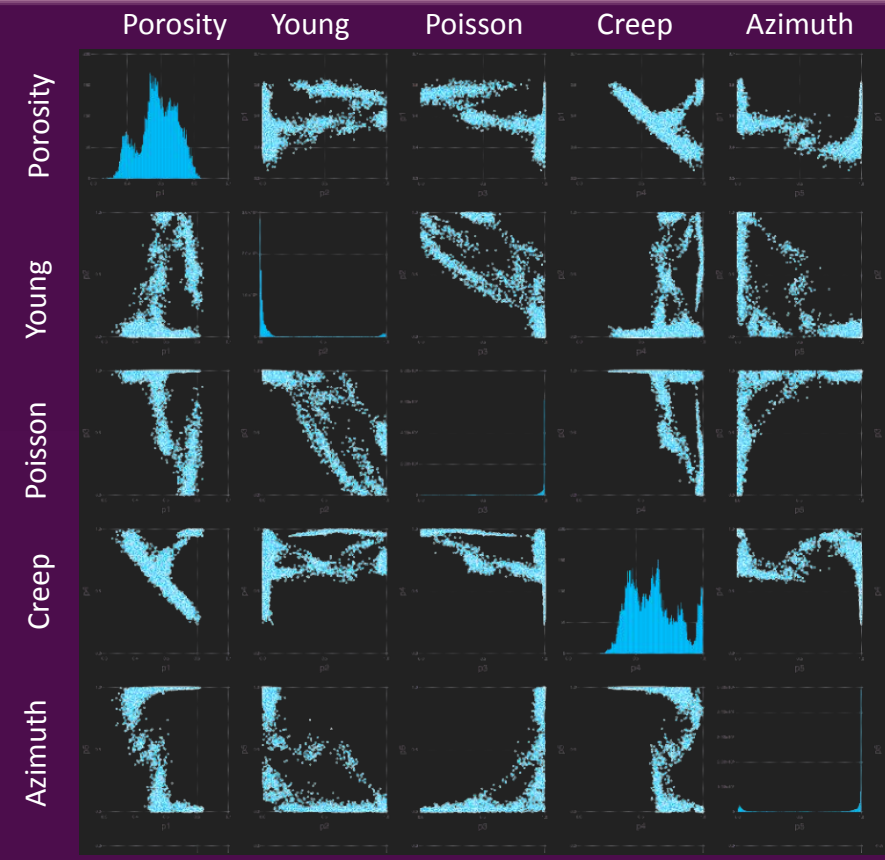
SIML for Oil & Gas Production

- History matching performed
- 15 ML parameters estimated
- Posterior parameter uncertainties evaluated
- Affine Invariant MCMC
- <https://github.com/madsjulia/AffineInvariantMCMC.jl>

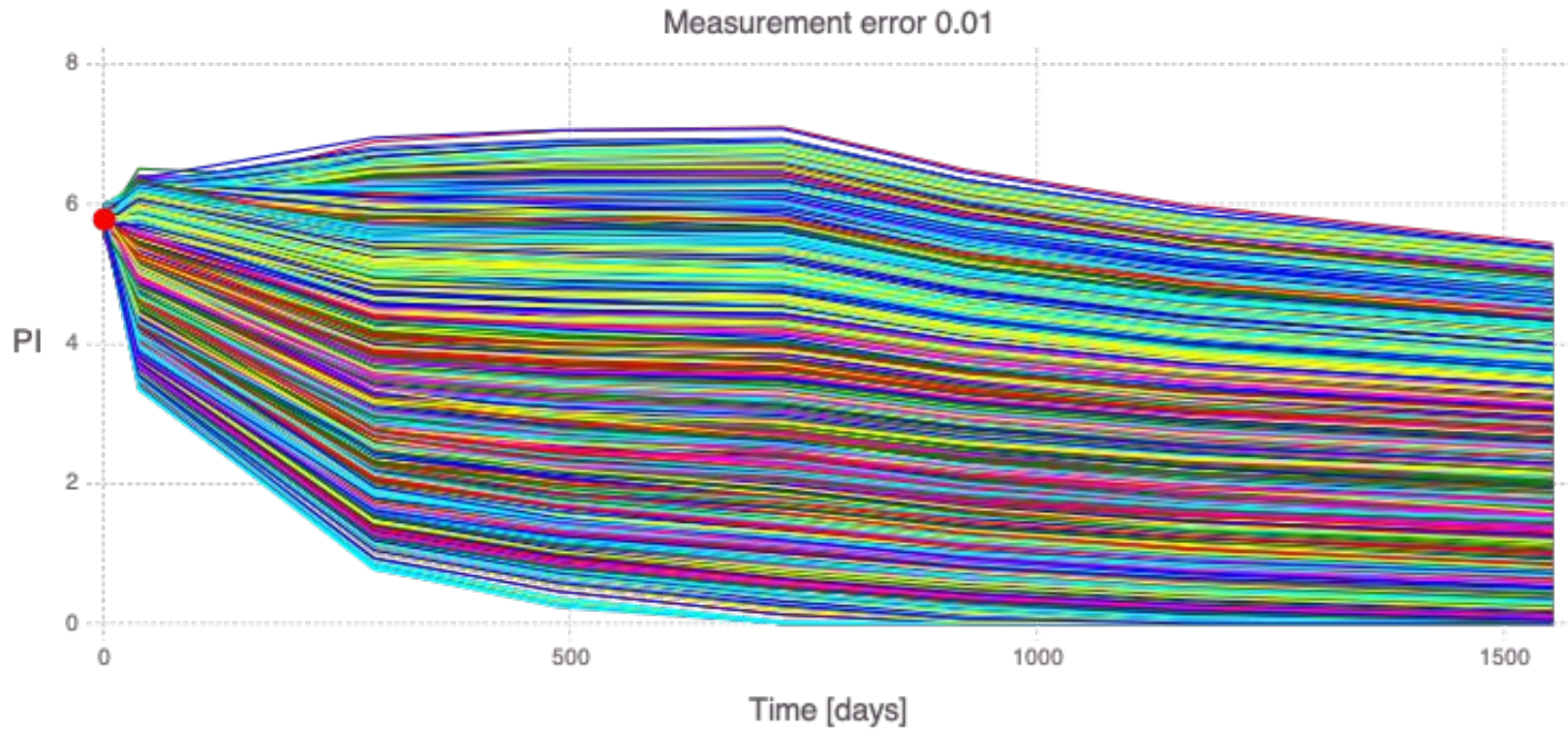


SIML for Oil & Gas Production: Parameter prior vs. posterior

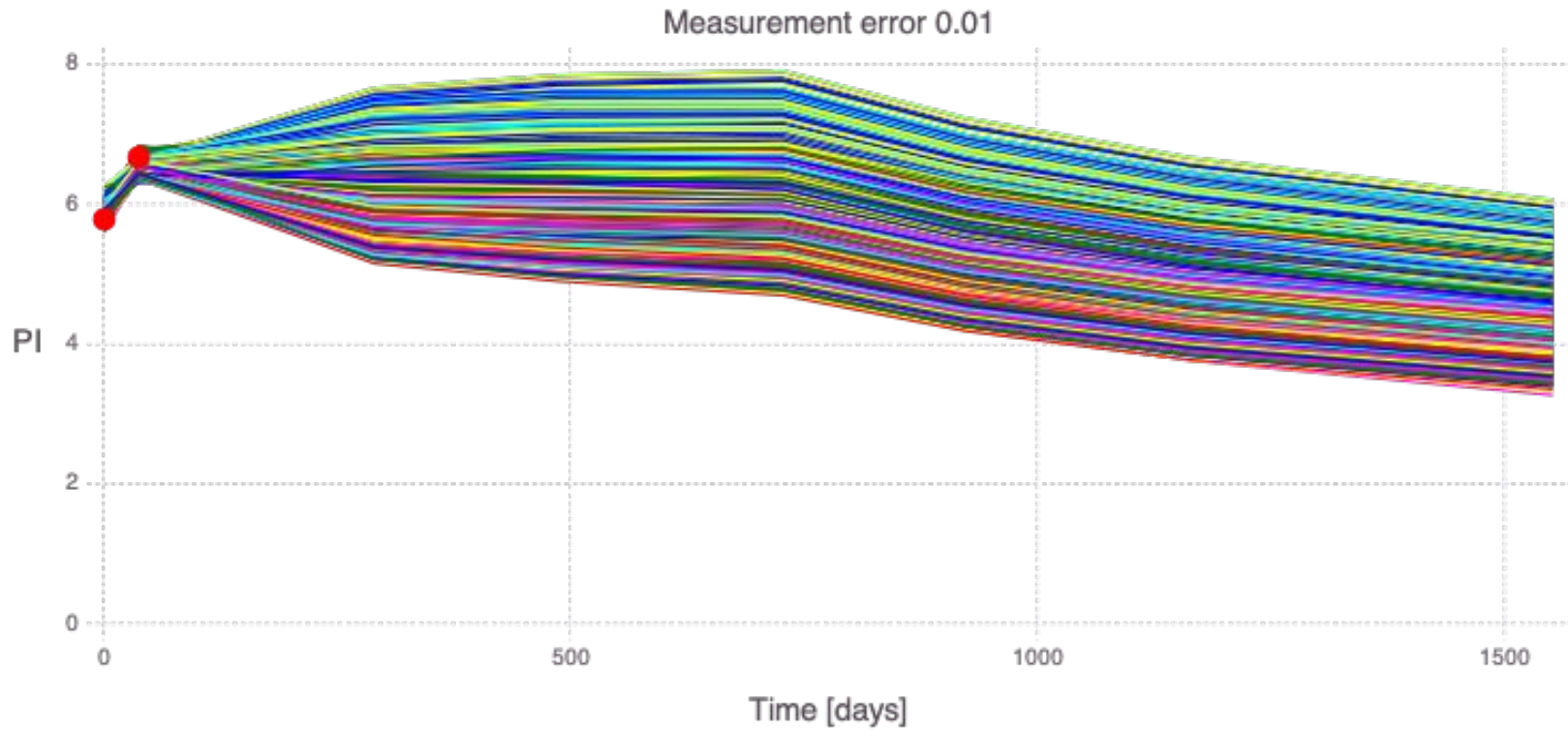
Model Inputs	Minimum	Maximum
Porosity	0.22	0.24
Young's Modulus	650,000	1,000,000
Poisson's Ratio	0.22	0.29
Creep Percentage	0	0.5
Well Azimuth	0	90



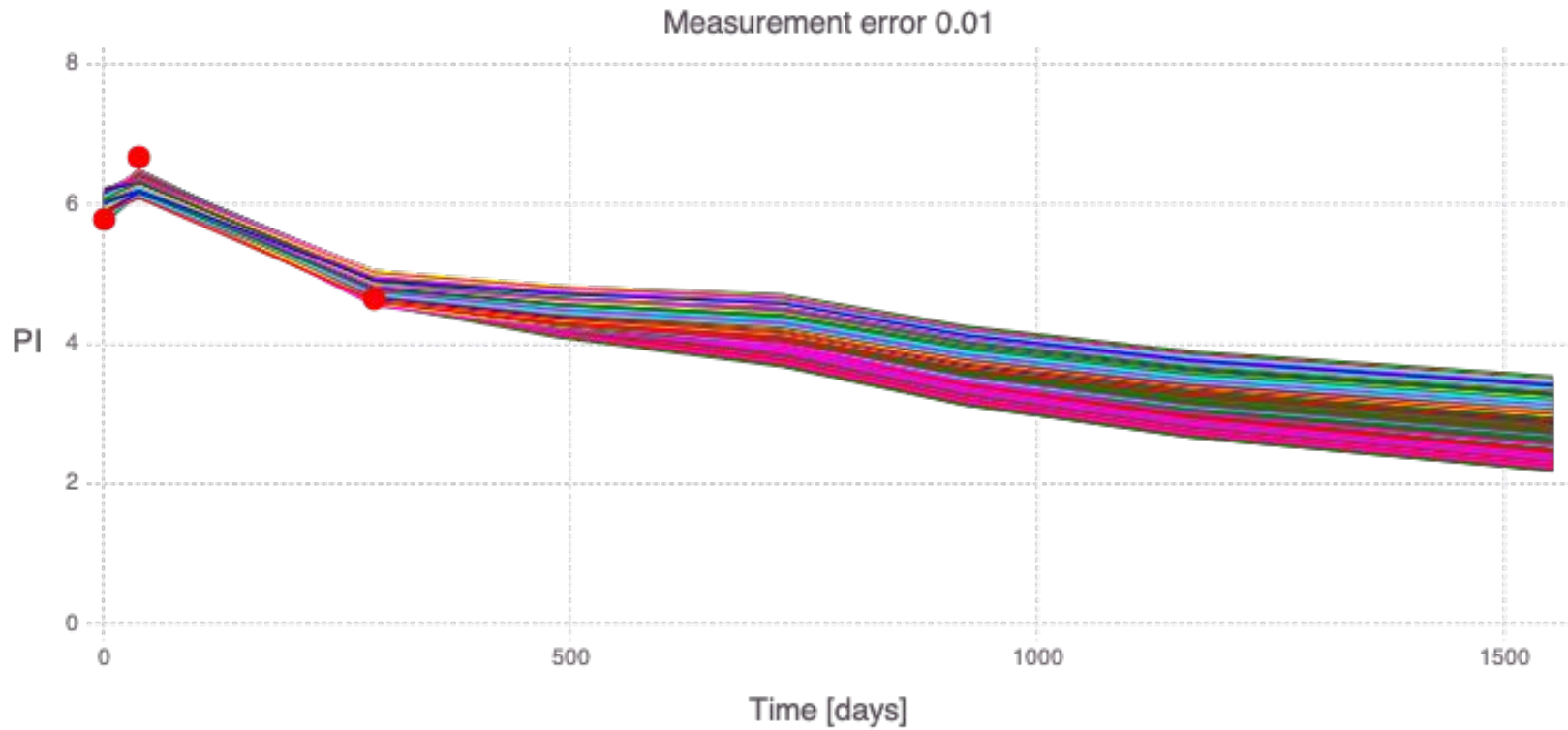
SIML for Oil & Gas Production: Predictive uncertainties



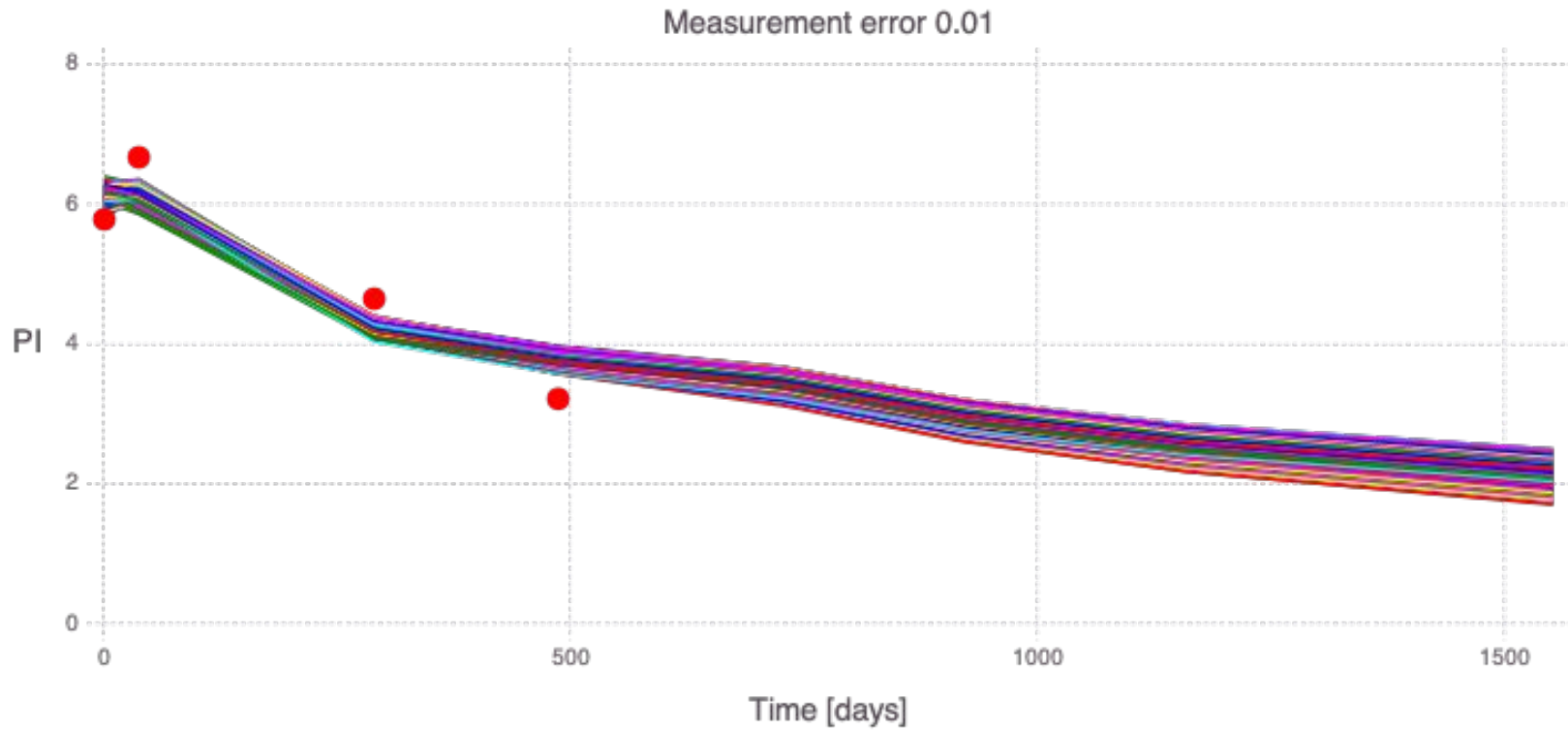
SIML for Oil & Gas Production: Predictive uncertainties



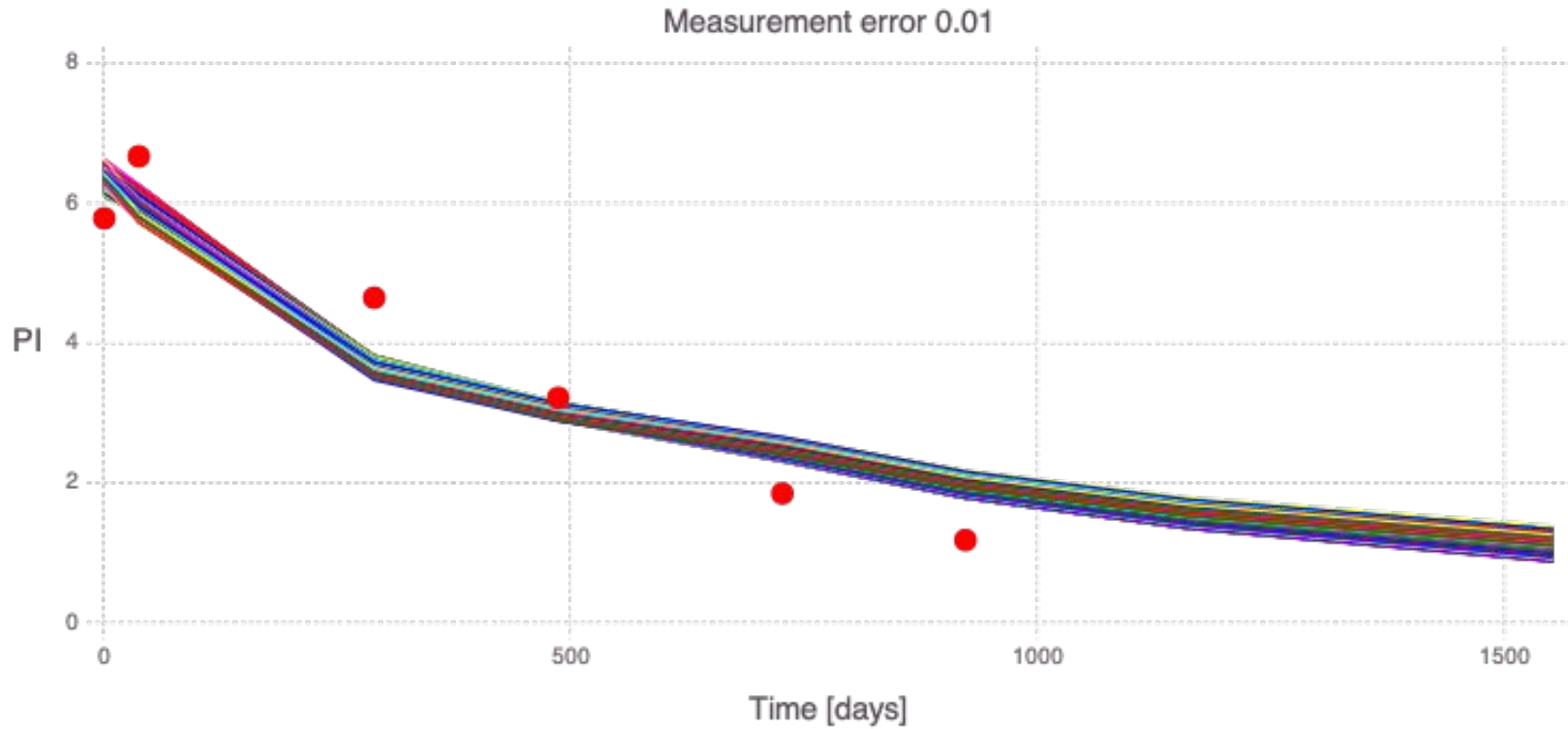
SIML for Oil & Gas Production: Predictive uncertainties



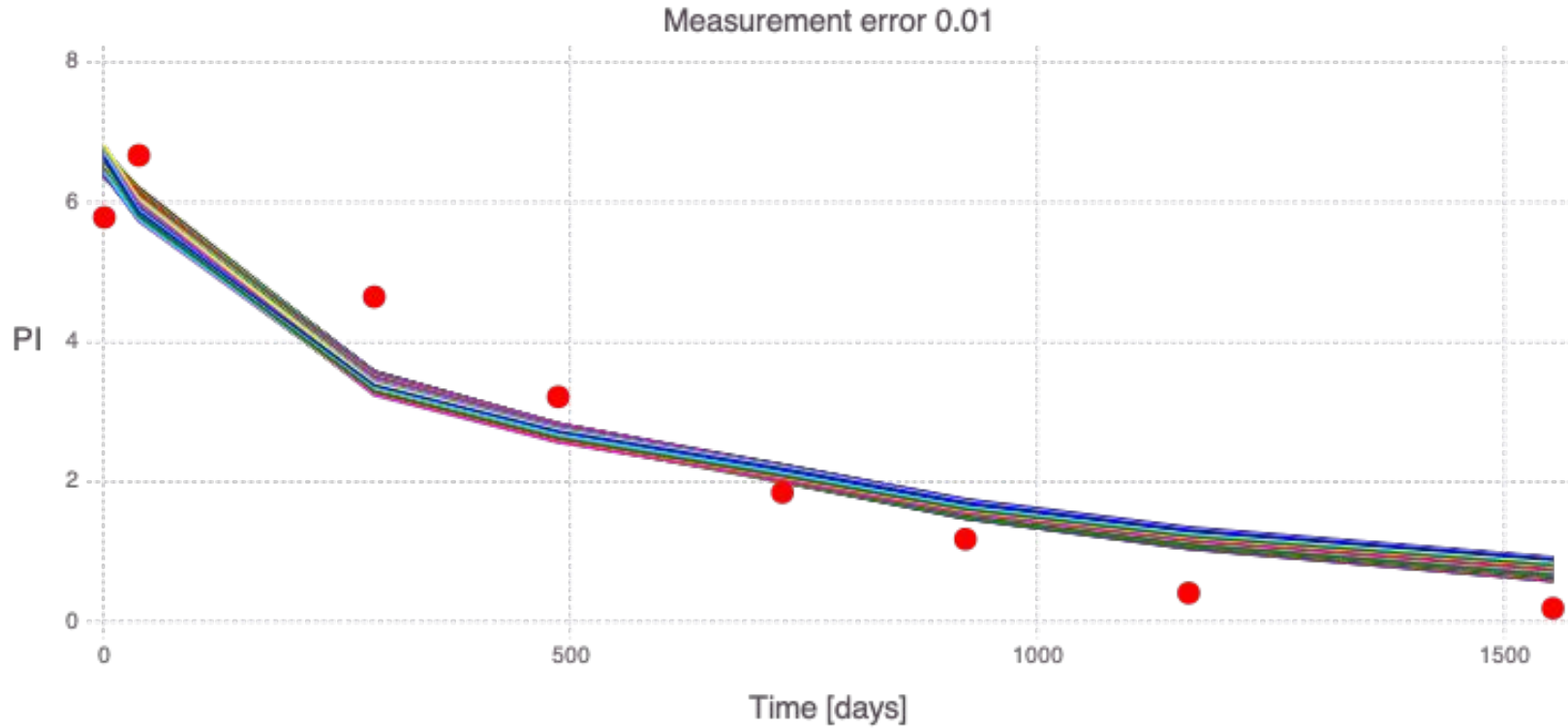
SIML for Oil & Gas Production: Predictive uncertainties



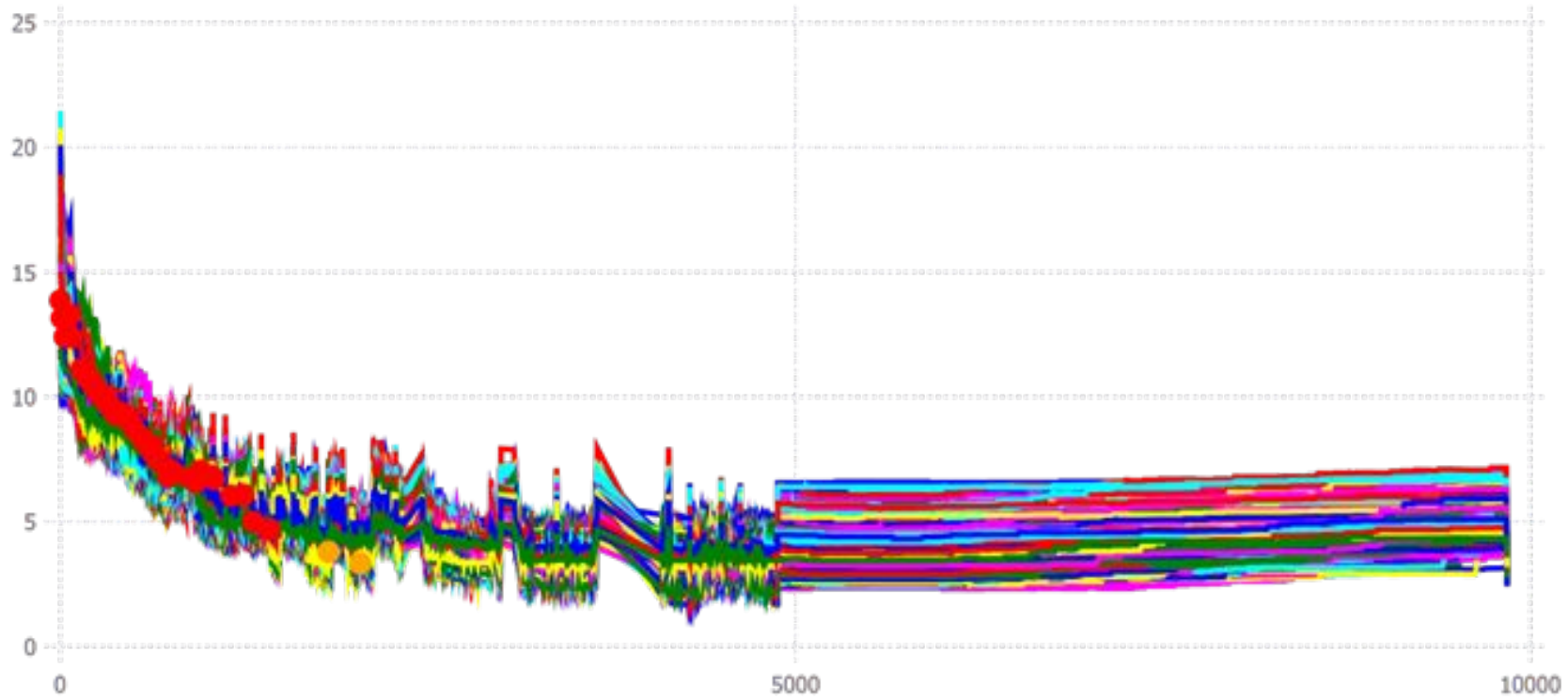
SIML for Oil & Gas Production: Predictive uncertainties



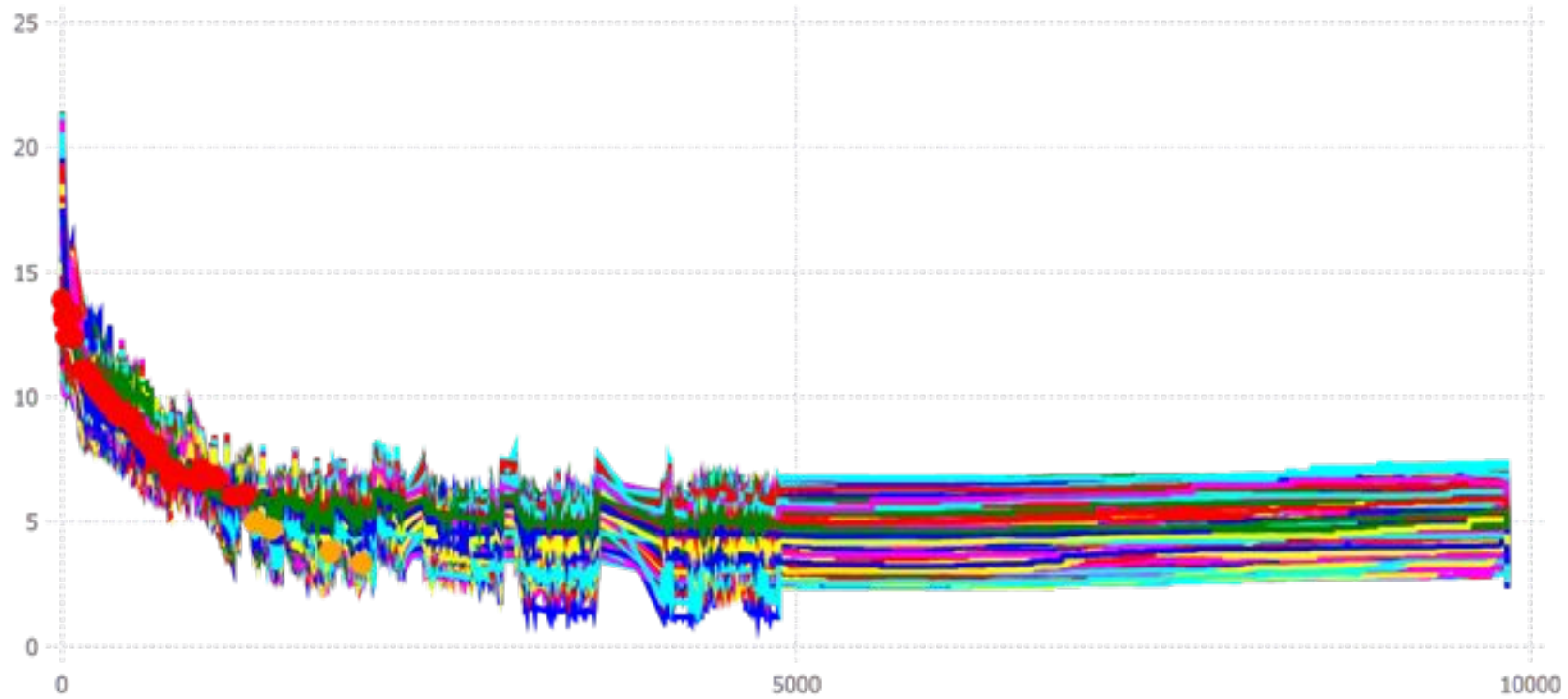
SIML for Oil & Gas Production: Predictive uncertainties



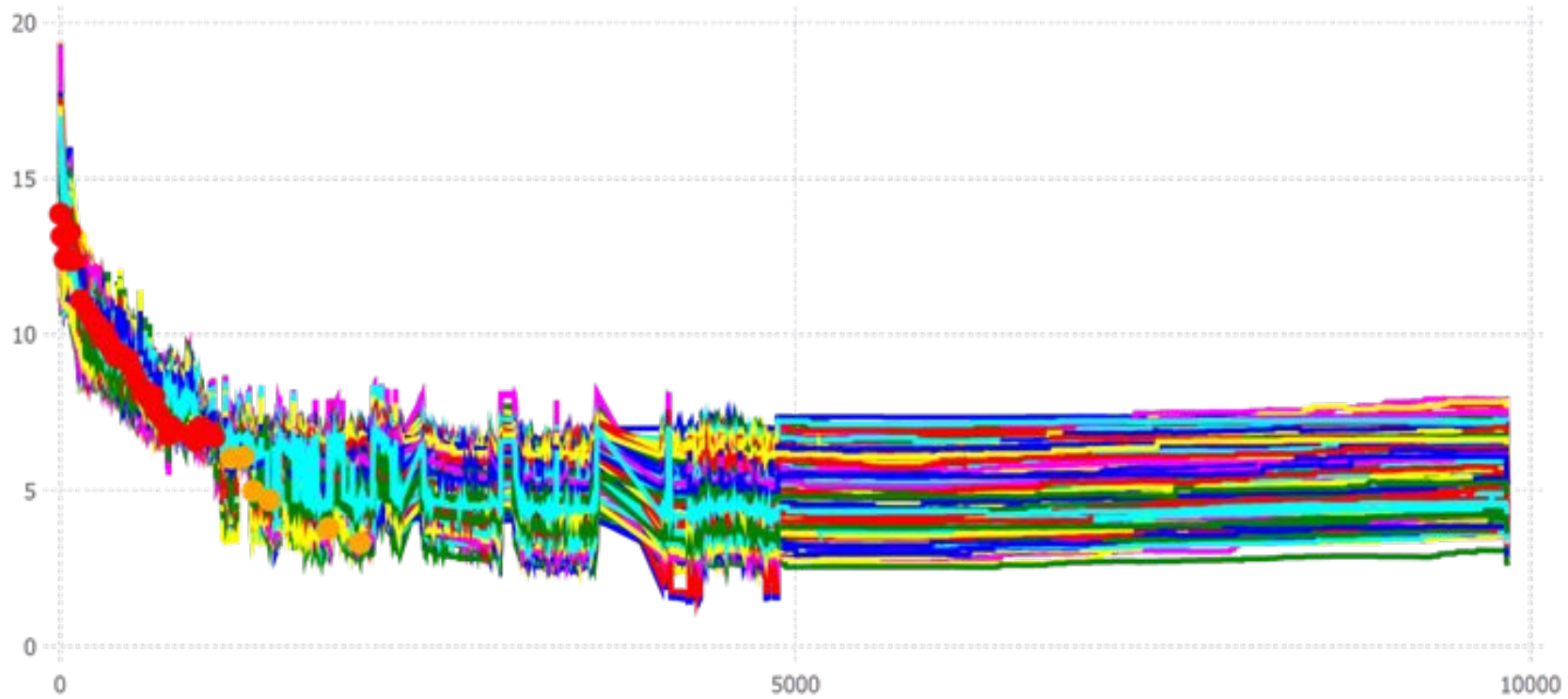
SIML for Oil & Gas: Prediction of Productivity Index (PI)



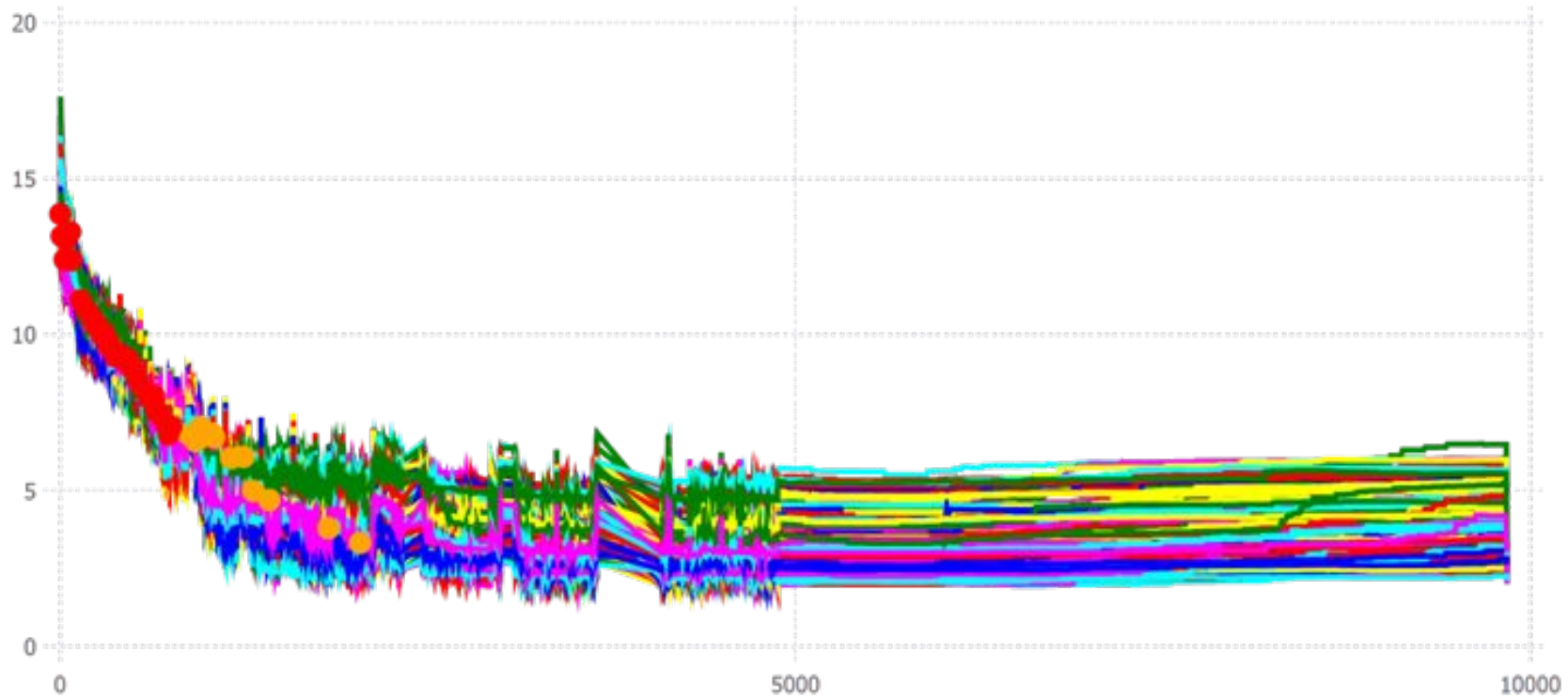
SIML for Oil & Gas: Prediction of Productivity Index (PI)



SIML for Oil & Gas: Prediction of Productivity Index (PI)



SIML for Oil & Gas: Prediction of Productivity Index (PI)

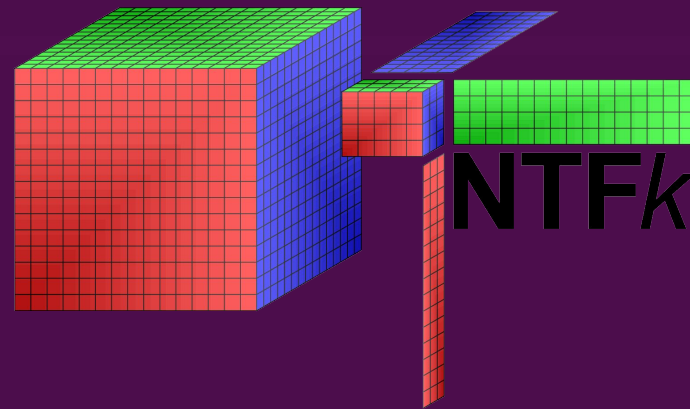
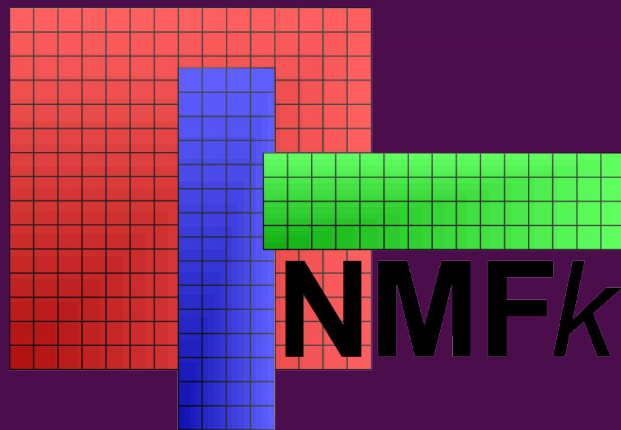


Unsupervised (self-supervised) examples

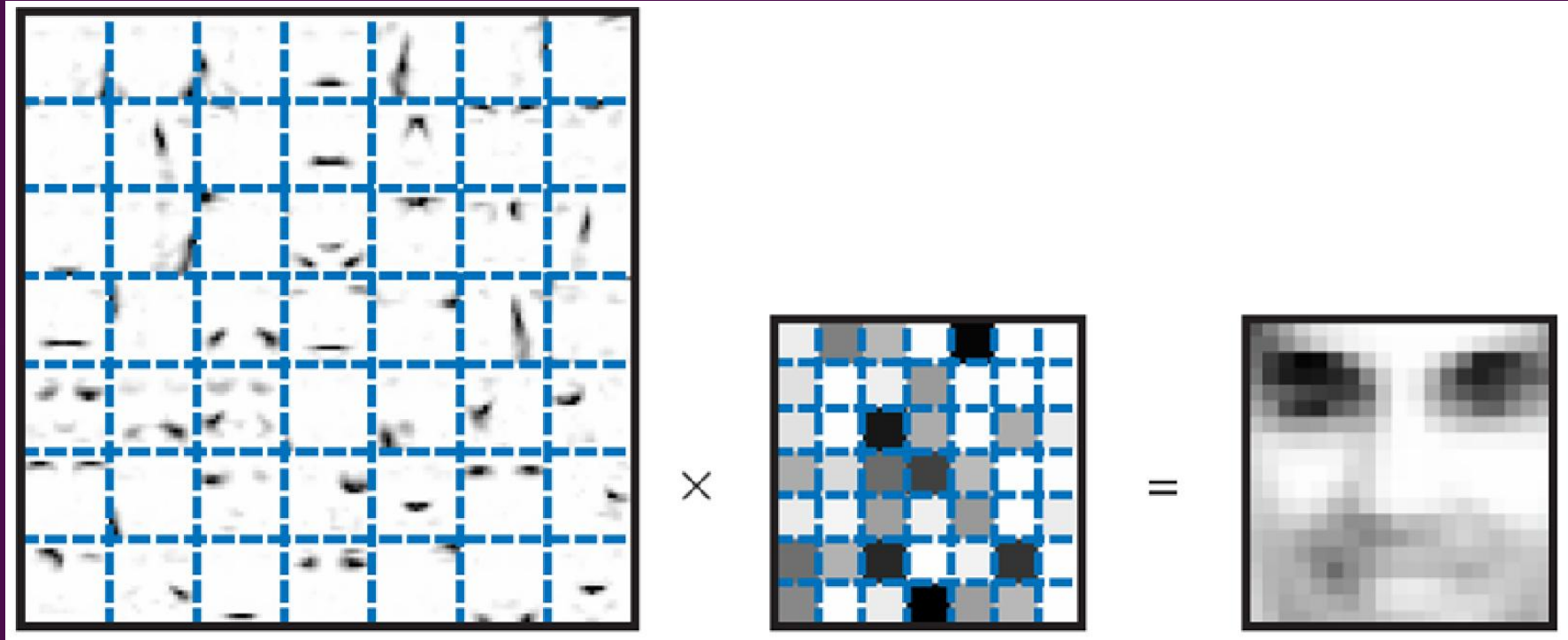
- Geysers
- Oklahoma seismicity
- Geologic reservoir (geothermal) exploration and utilization

Factorization Methods

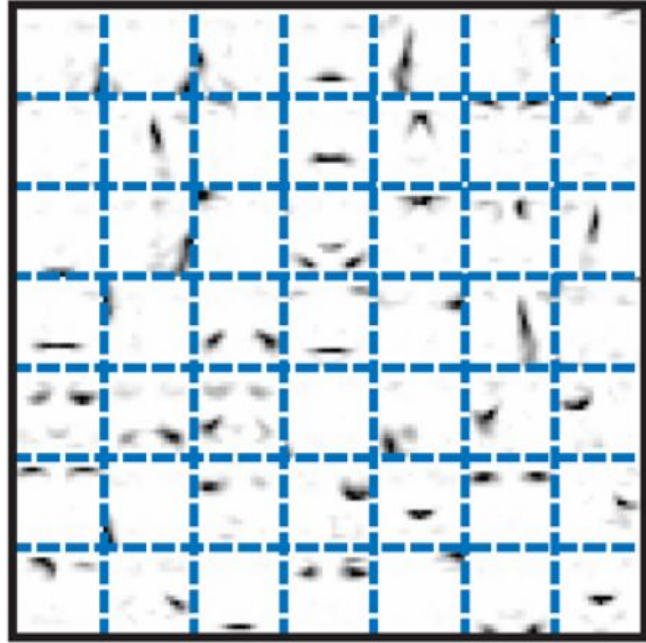
- Nonnegative Matrix Factorization
- Nonnegative Tensor Factorization
- <https://github.com/SmartTensors>



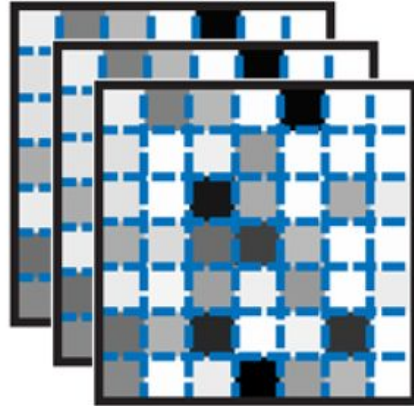
Nonnegative Matrix Factorization



Nonnegative Tensor Factorization



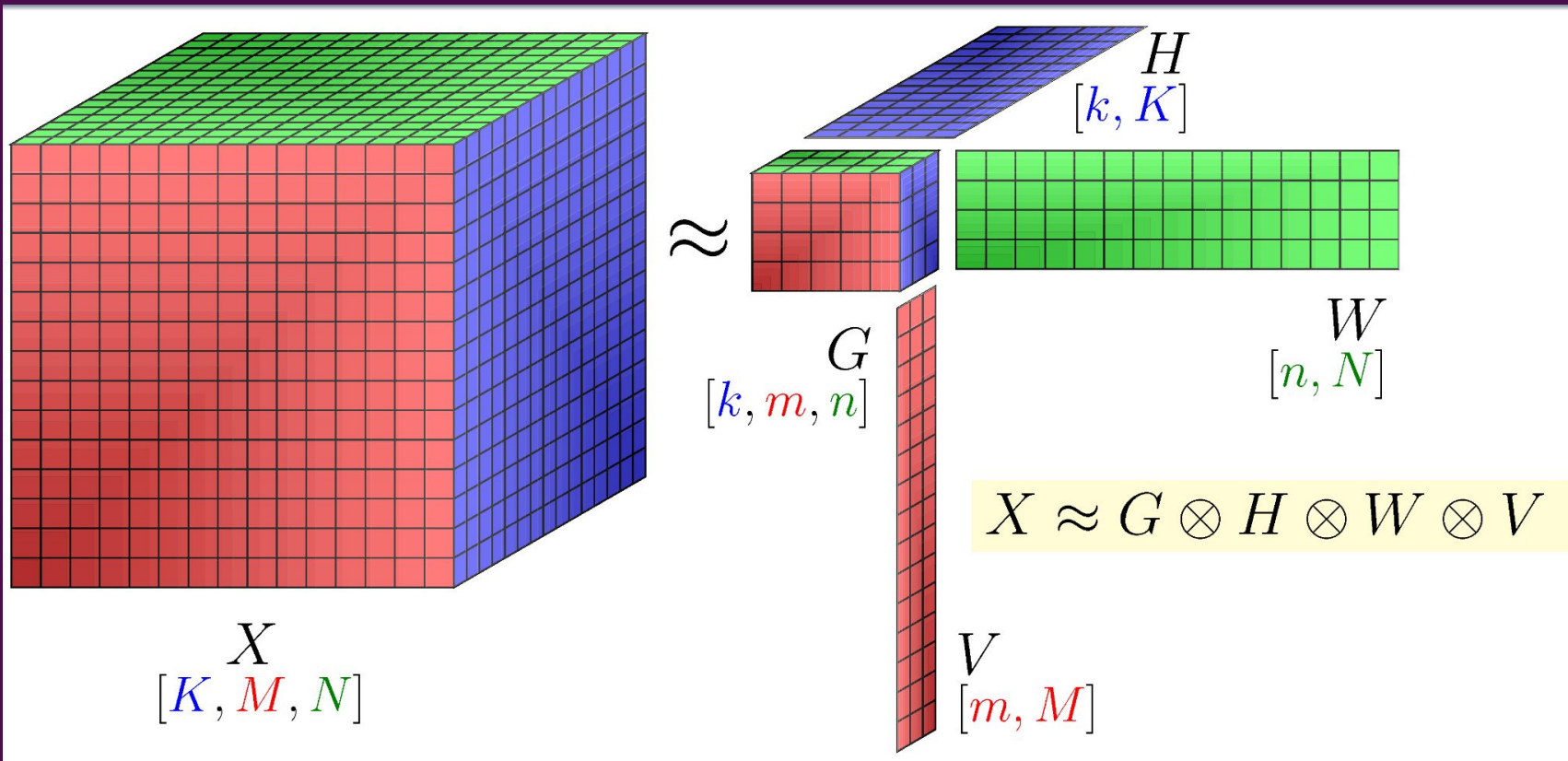
\otimes



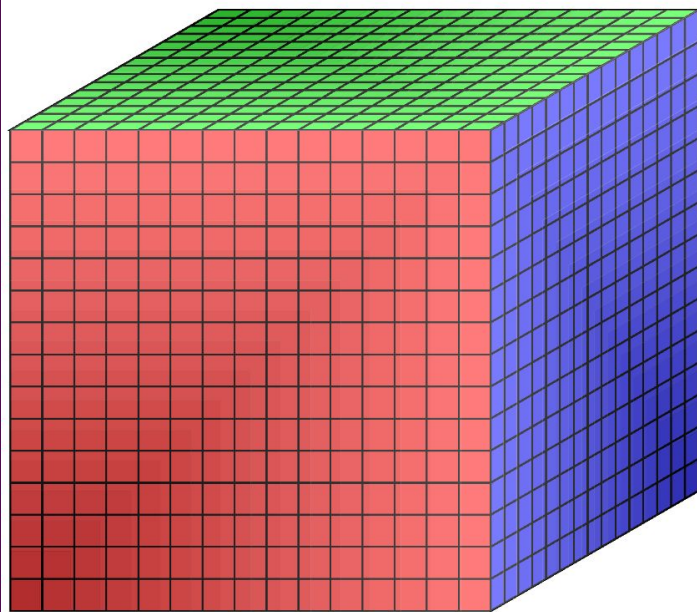
=



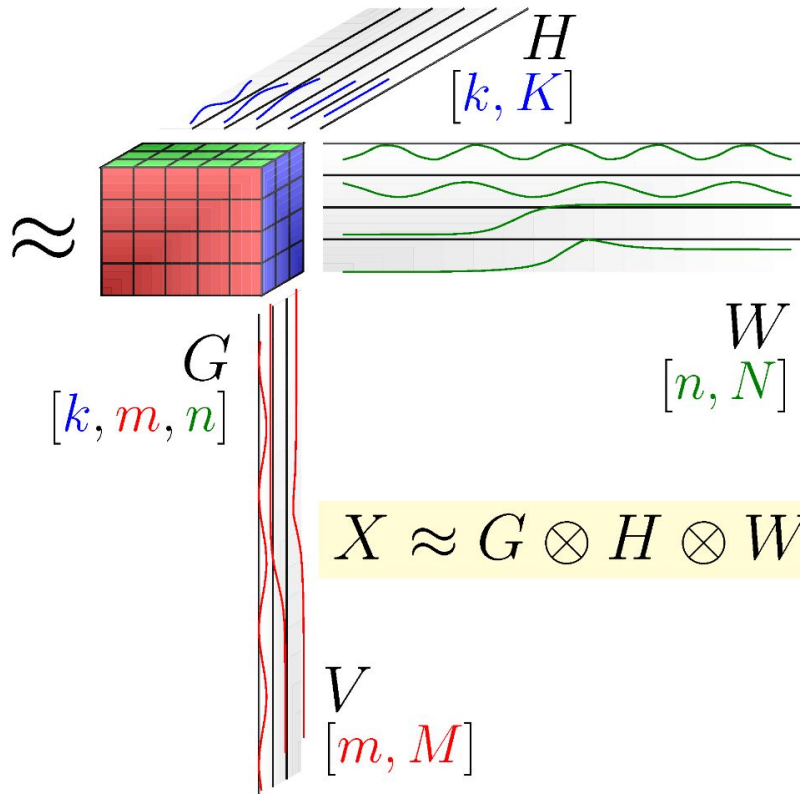
Nonnegative Tensor Factorization



Nonnegative Tensor Factorization

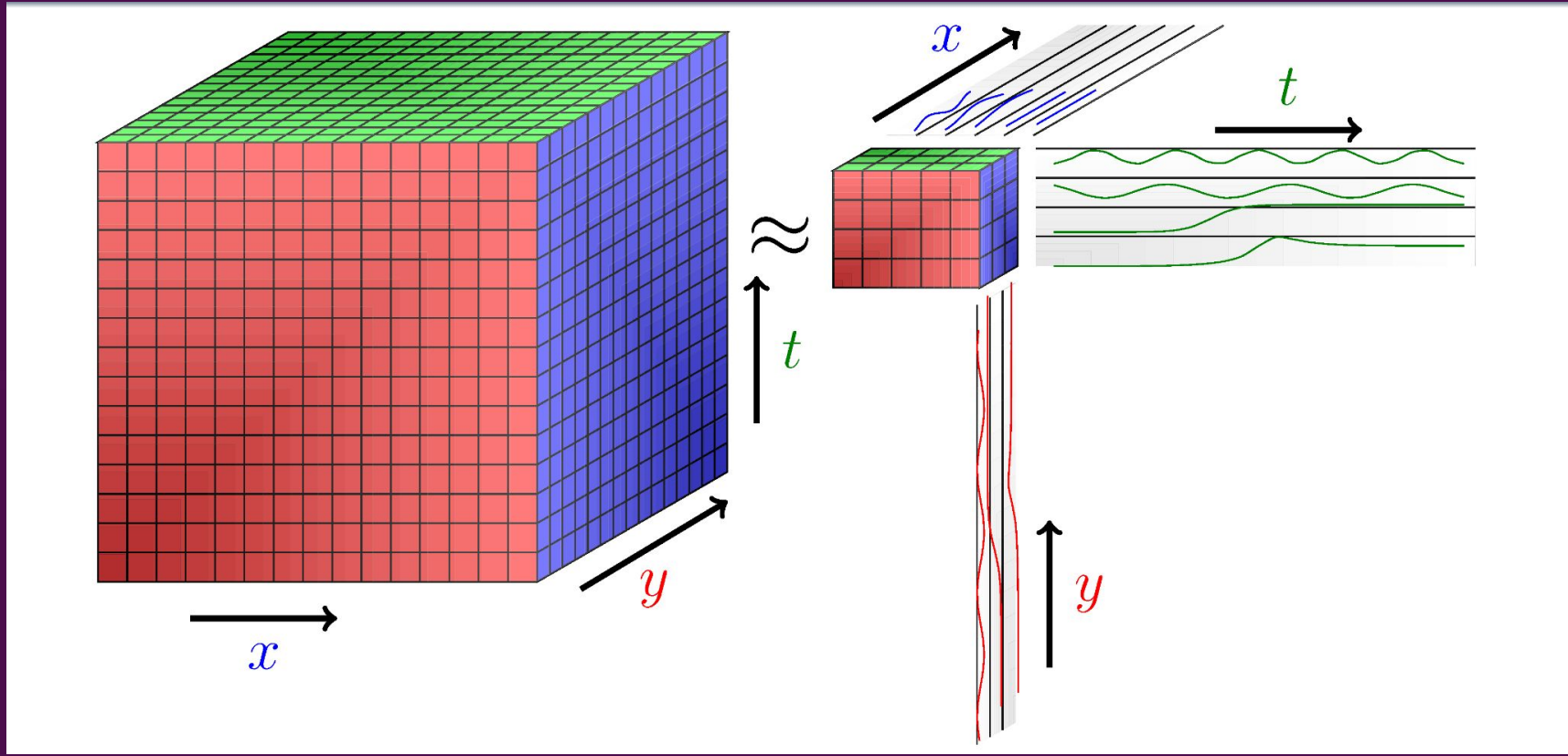


X
 $[K, M, N]$

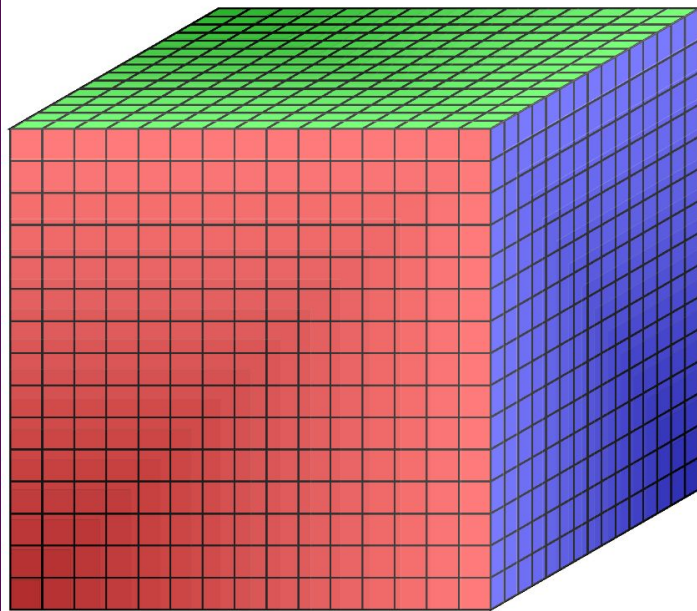


$$X \approx G \otimes H \otimes W \otimes V$$

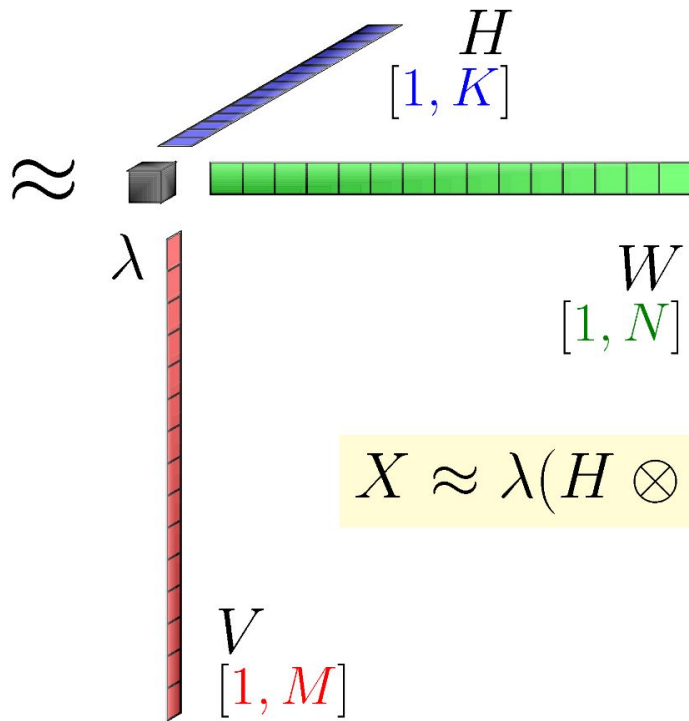
Nonnegative Tensor Factorization



Nonnegative Tensor Factorization

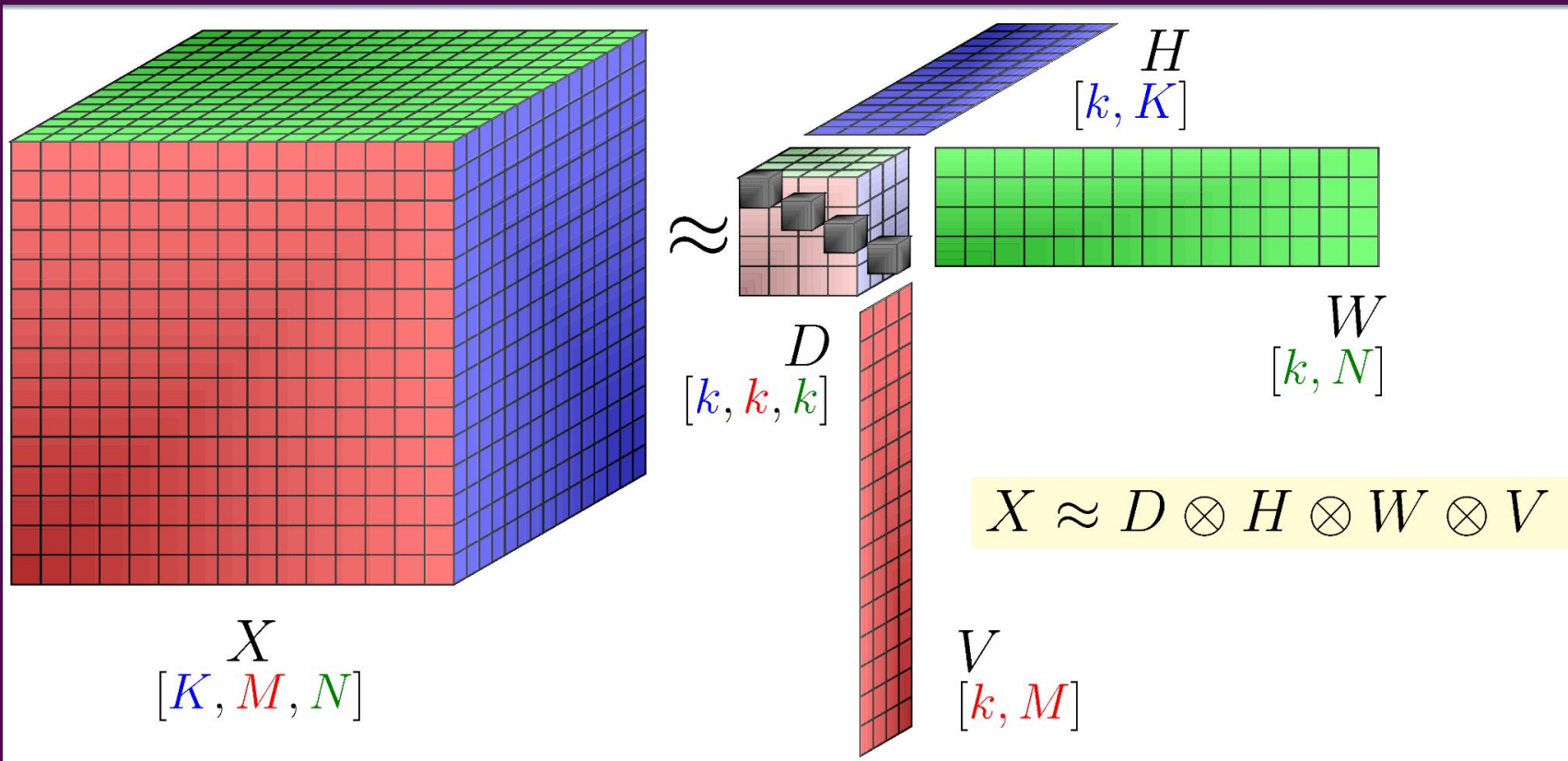


X
 $[K, M, N]$

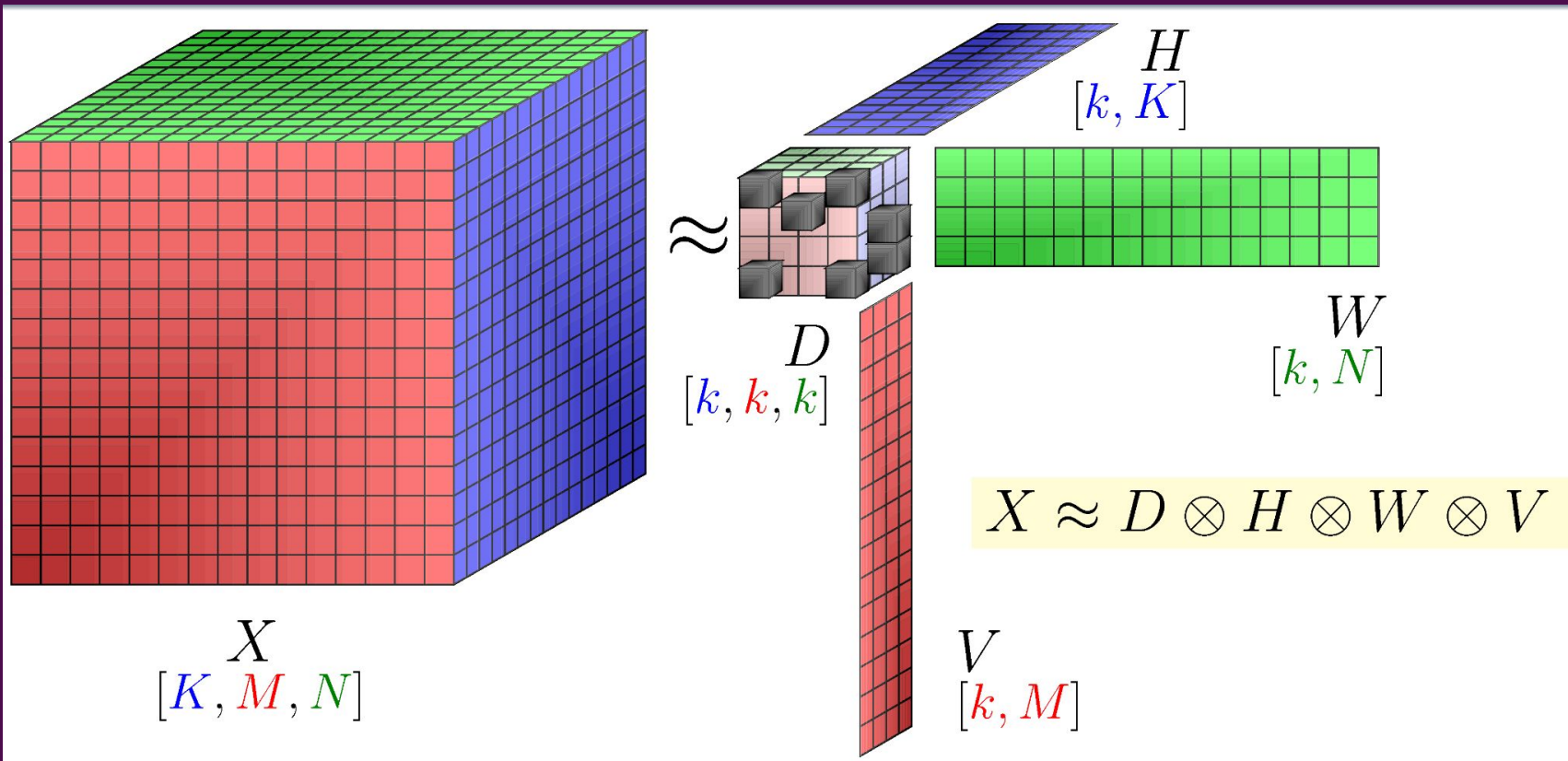


$$X \approx \lambda(H \otimes W \otimes V)$$

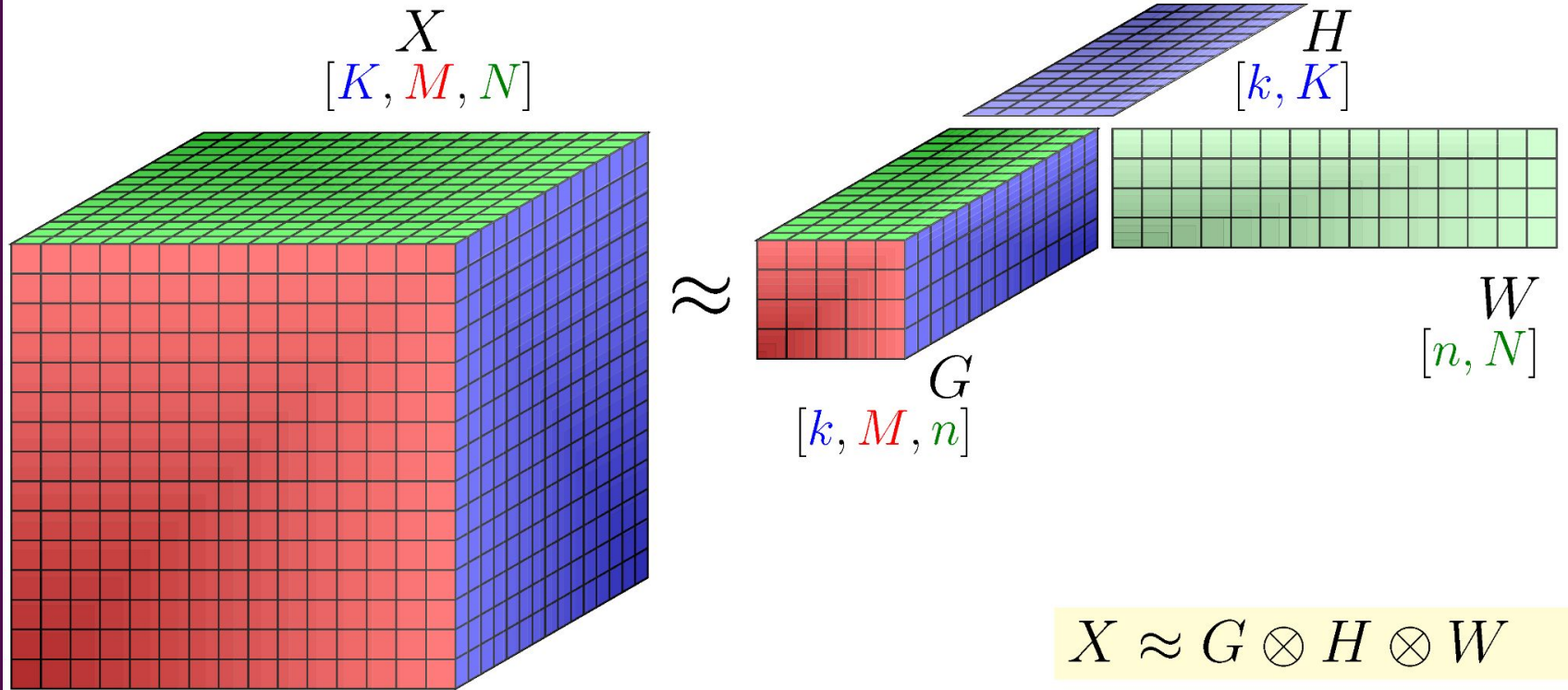
Nonnegative Tensor Factorization



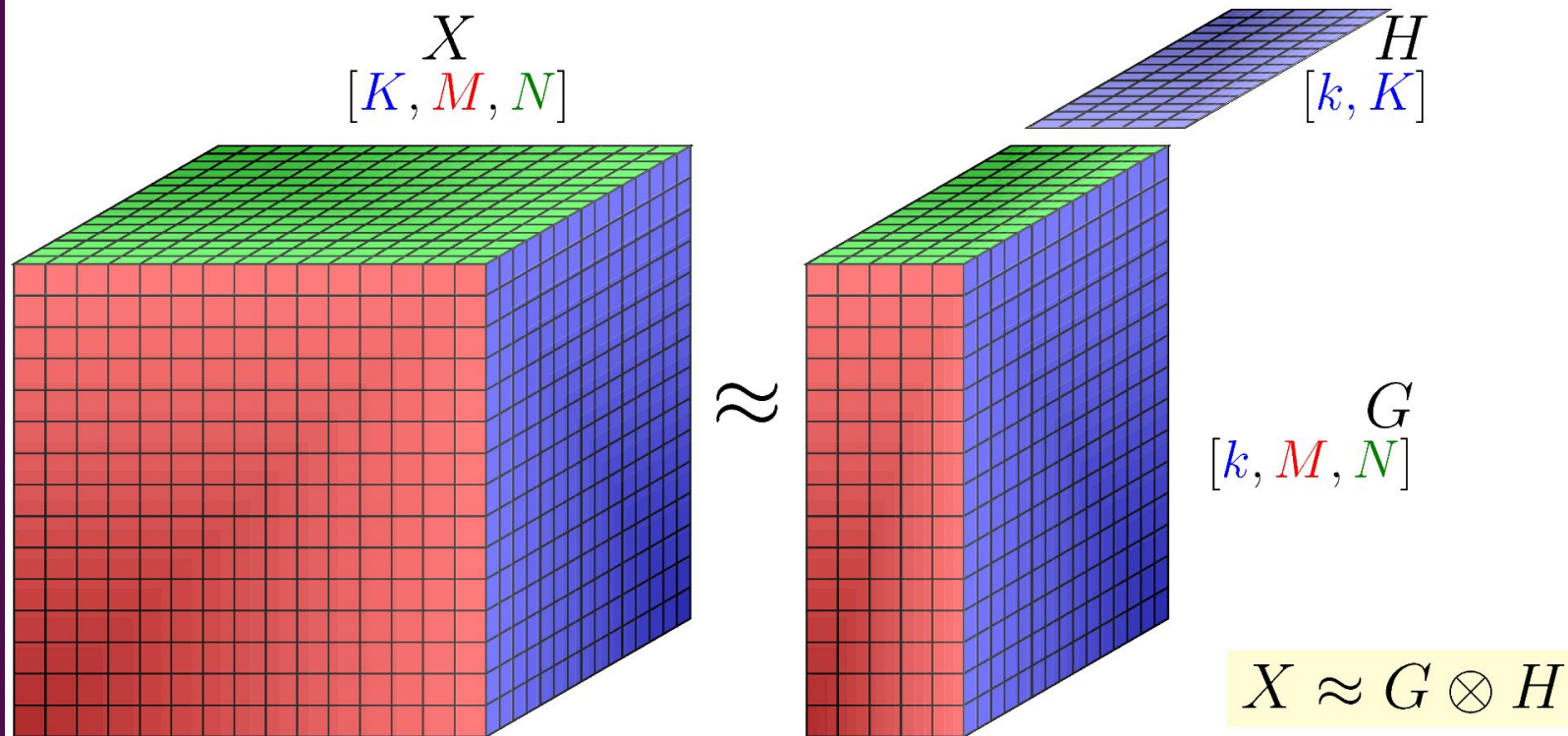
Nonnegative Tensor Factorization



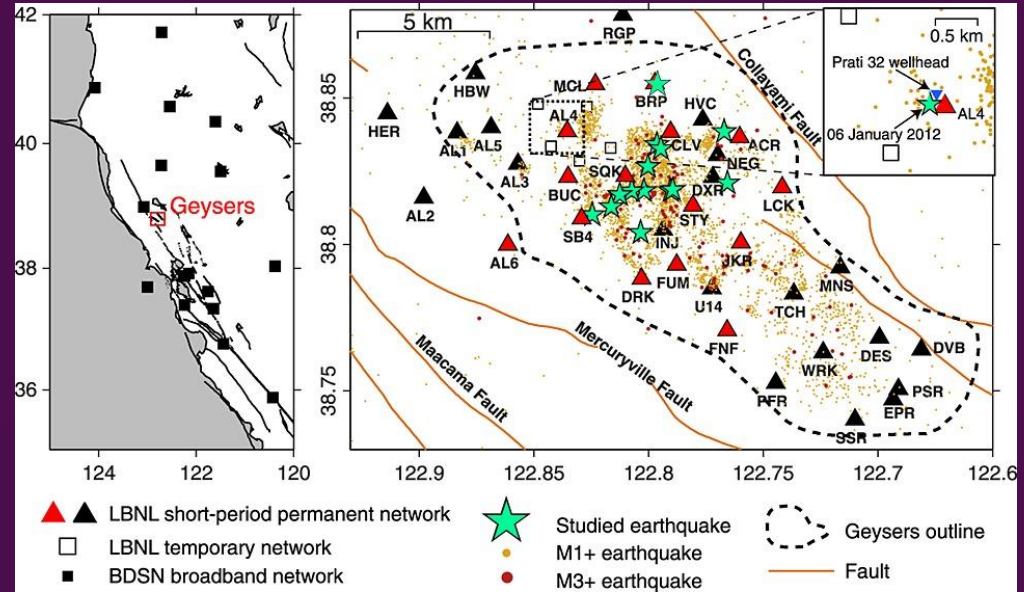
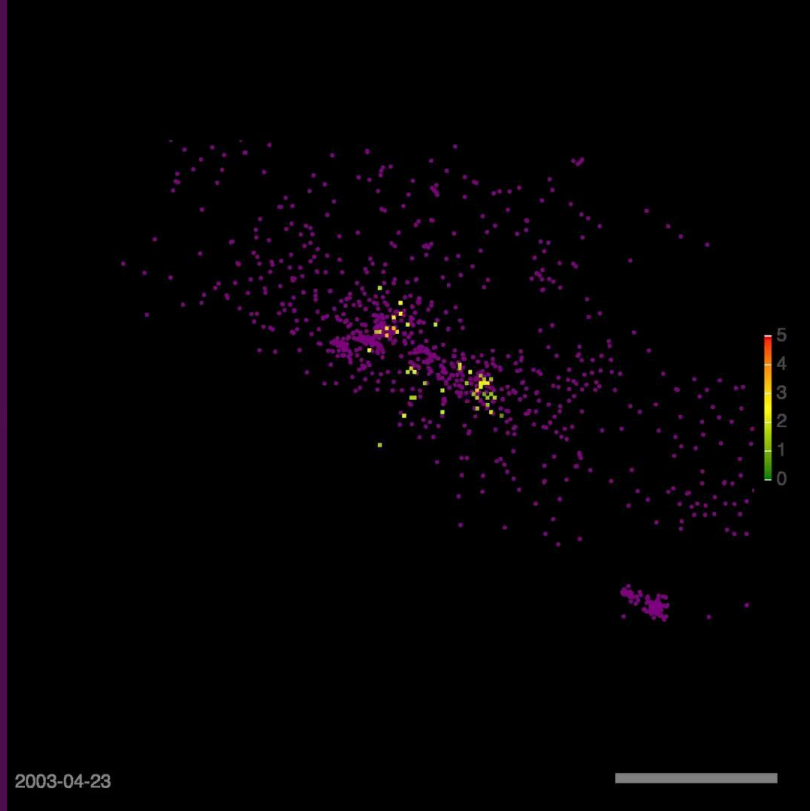
Nonnegative Tensor Factorization



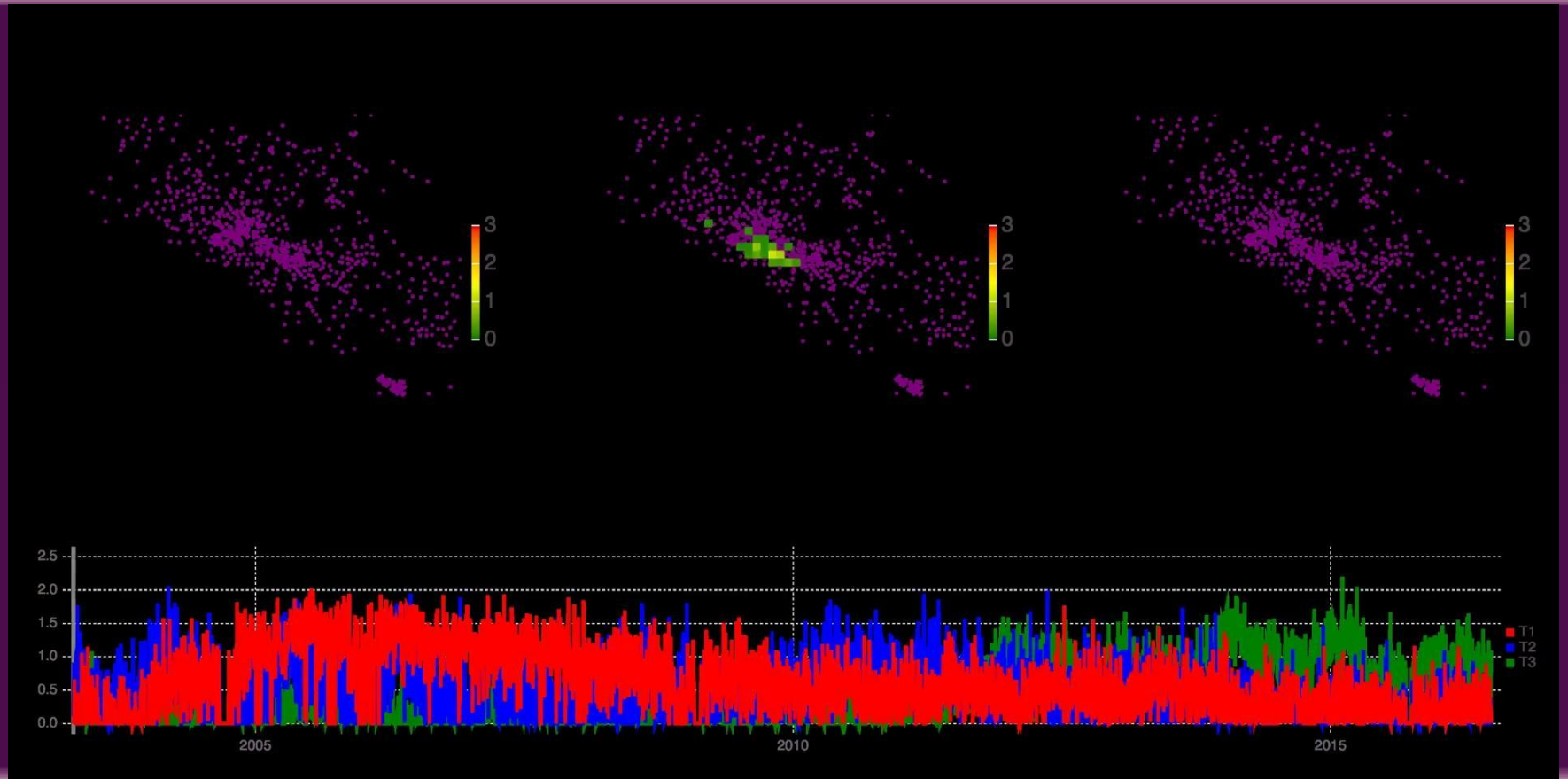
Nonnegative Tensor Factorization



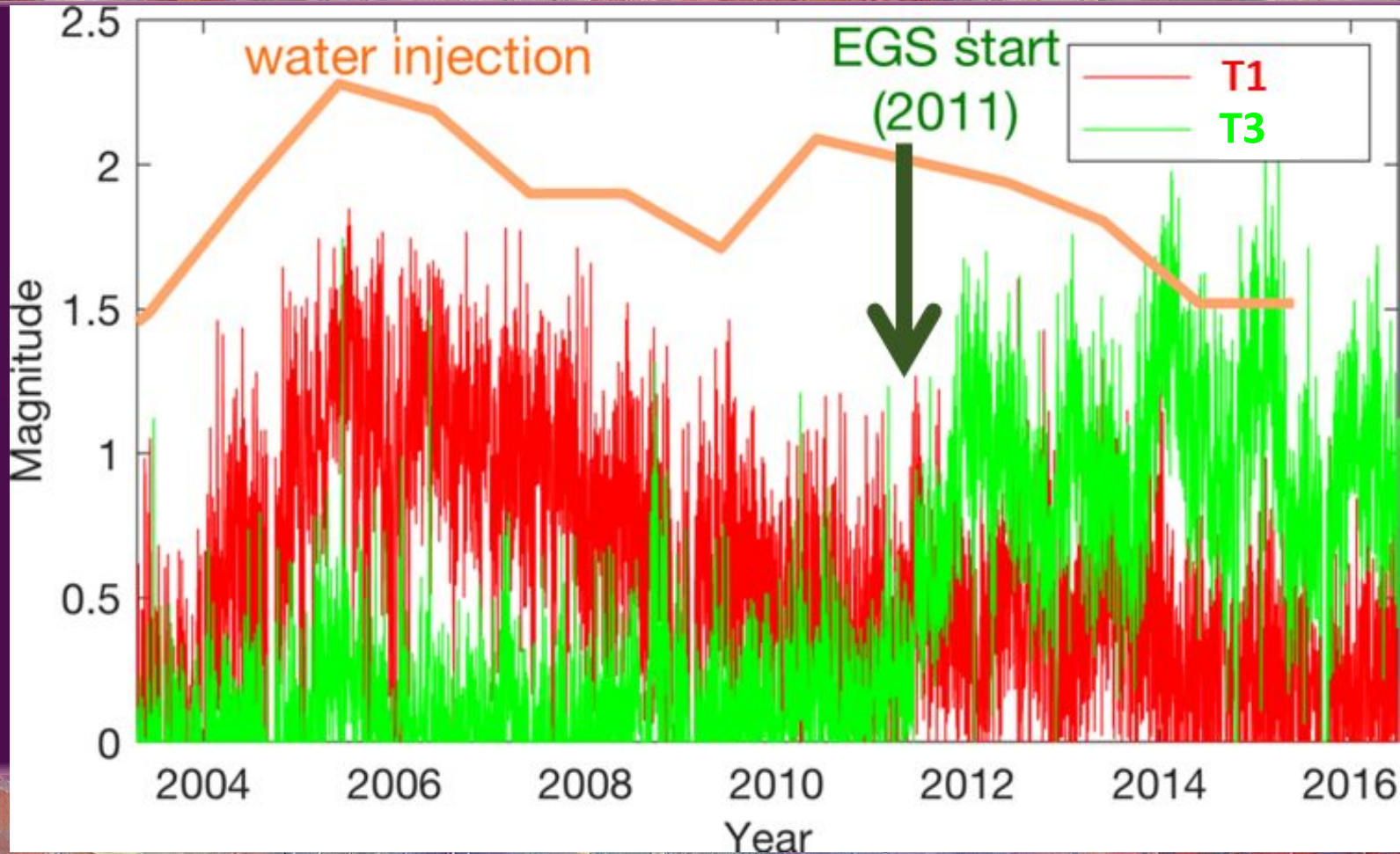
Geysers Induced Seismicity



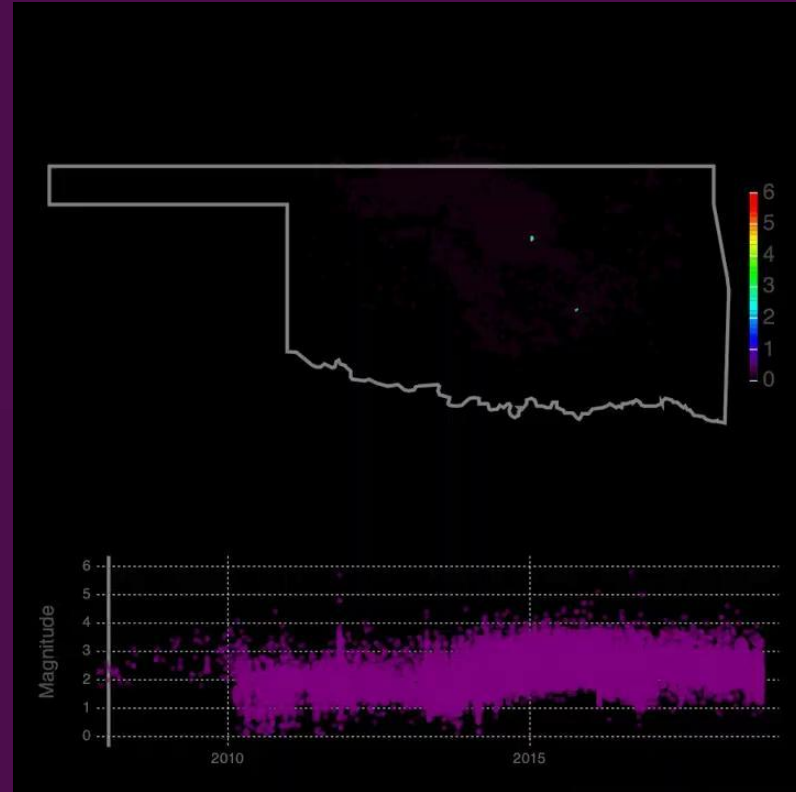
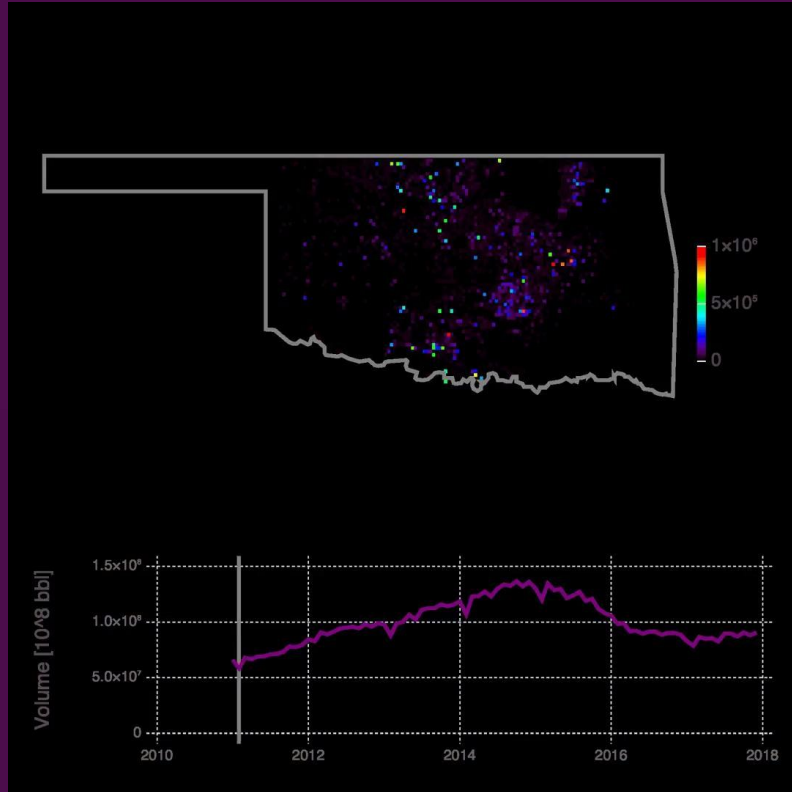
Geysers Induced Seismicity



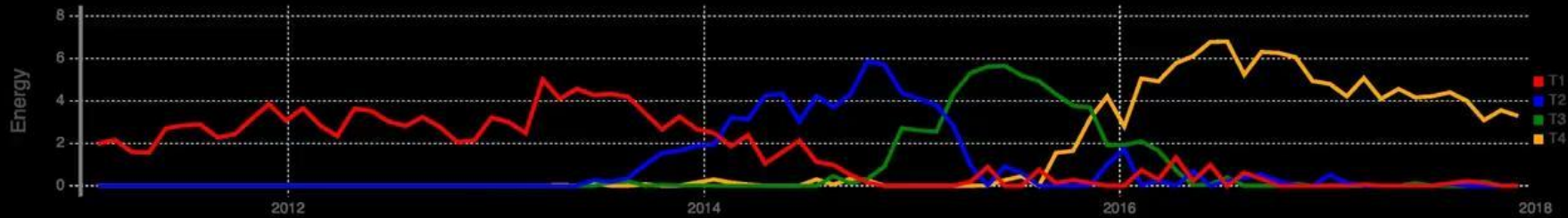
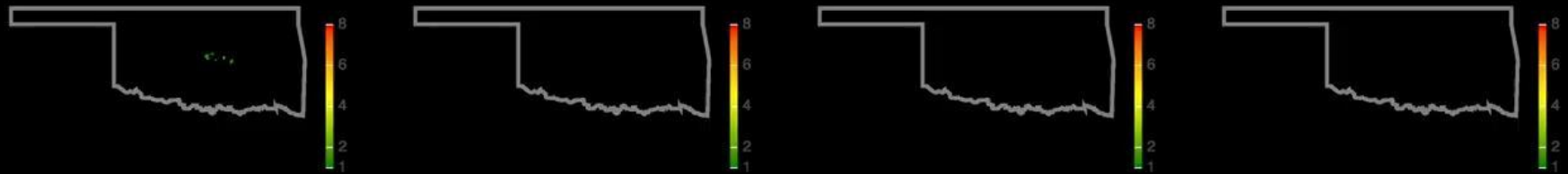
Geysers Induced Seismicity



Oklahoma Induced Seismicity



Oklahoma Induced Seismicity (Seismic events)

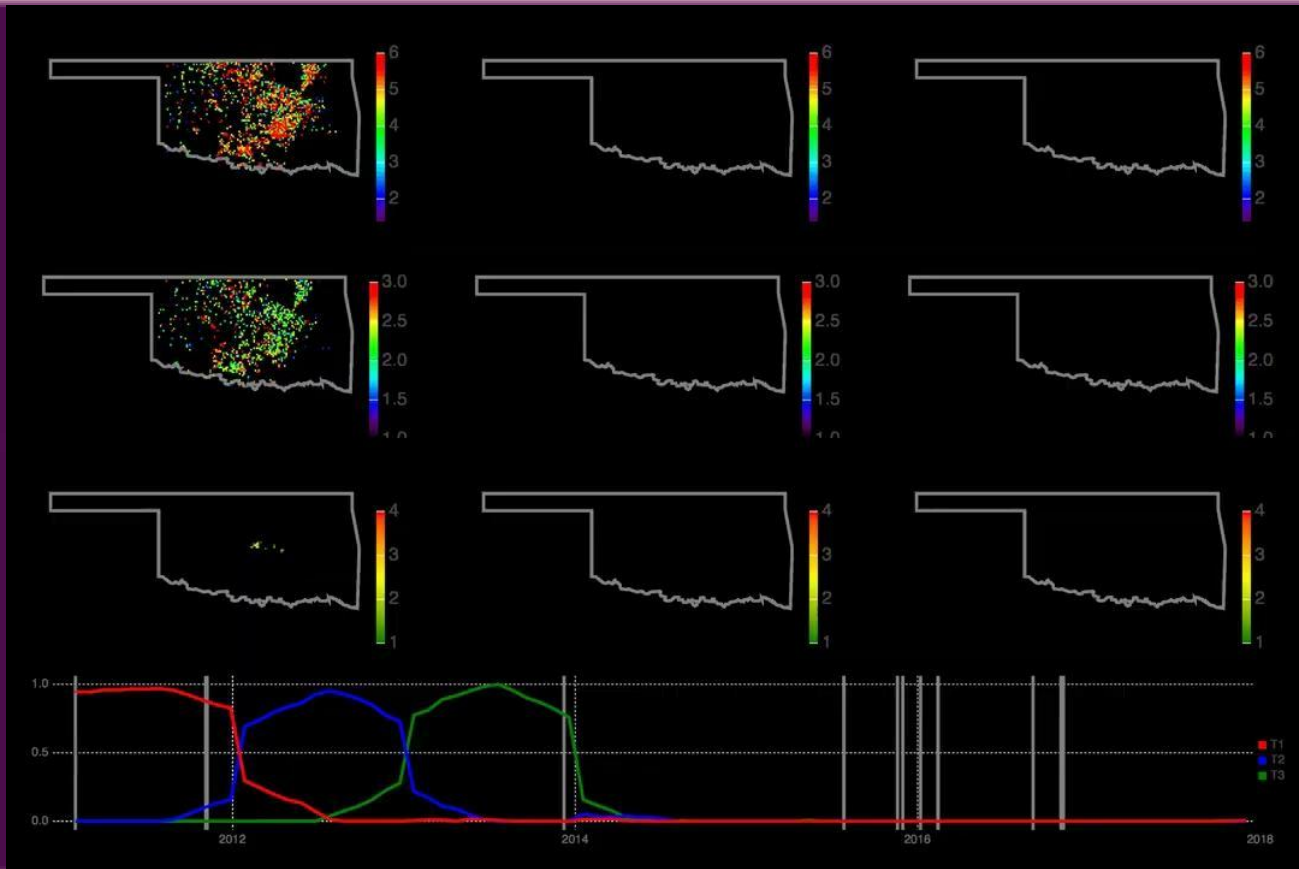


Oklahoma Induced Seismicity (Water disposal)

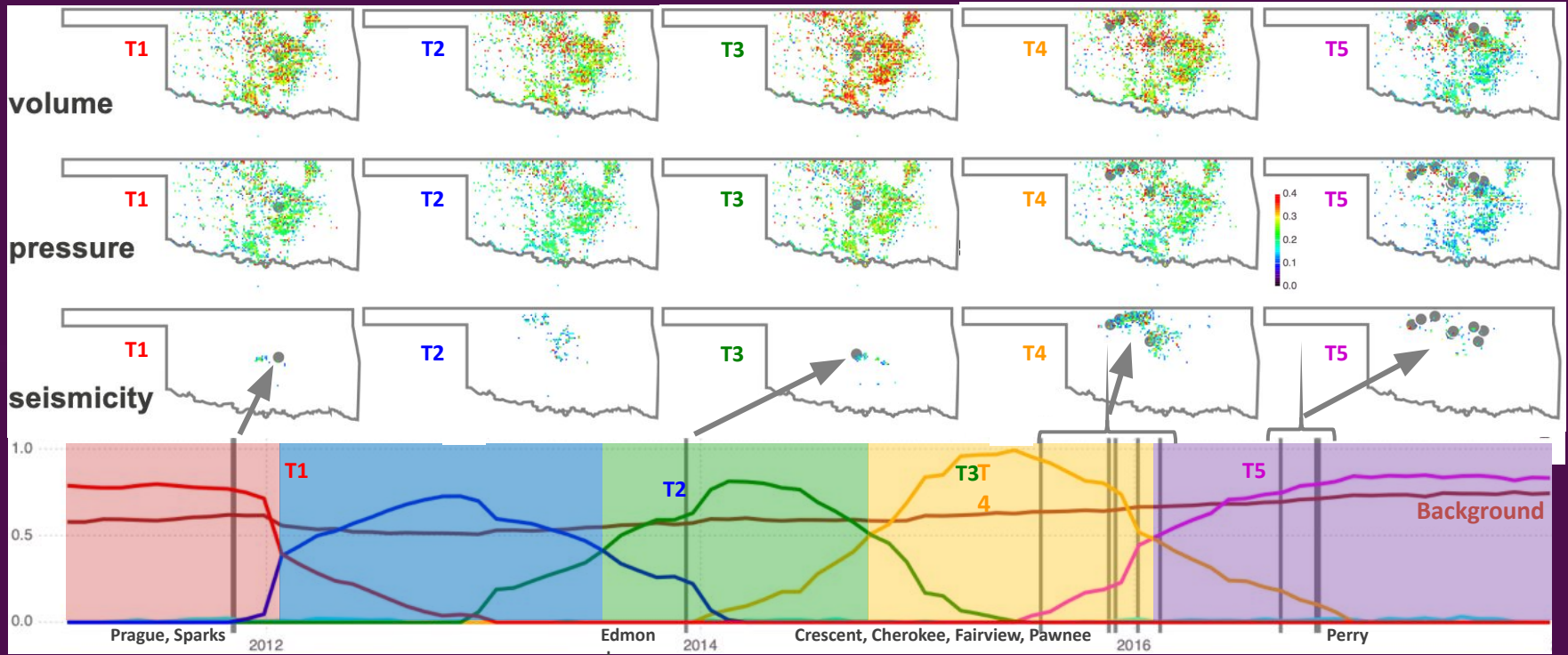
Water Disposal Volume

Reservoir Pressure

Seismicity



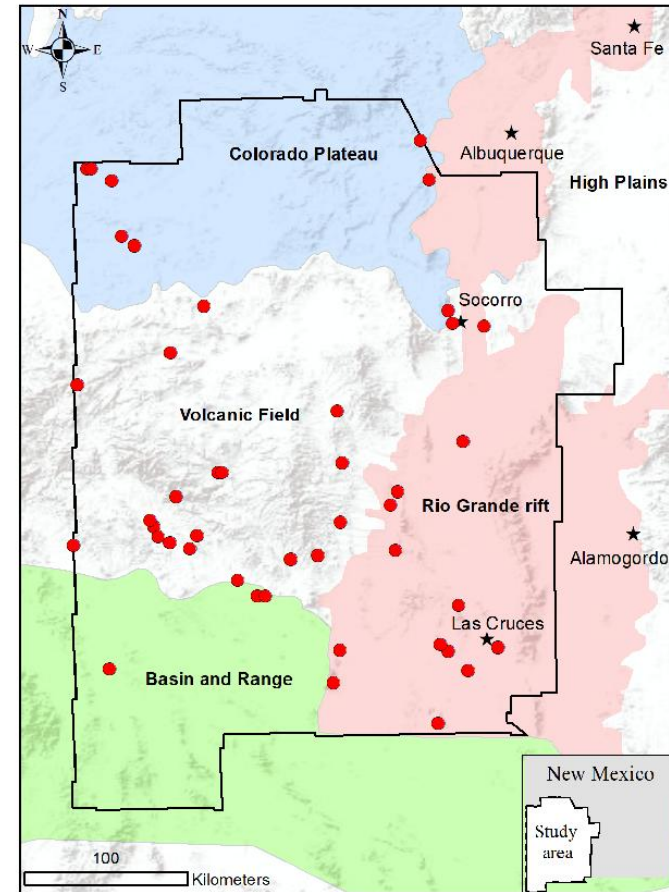
Oklahoma Induced Seismicity



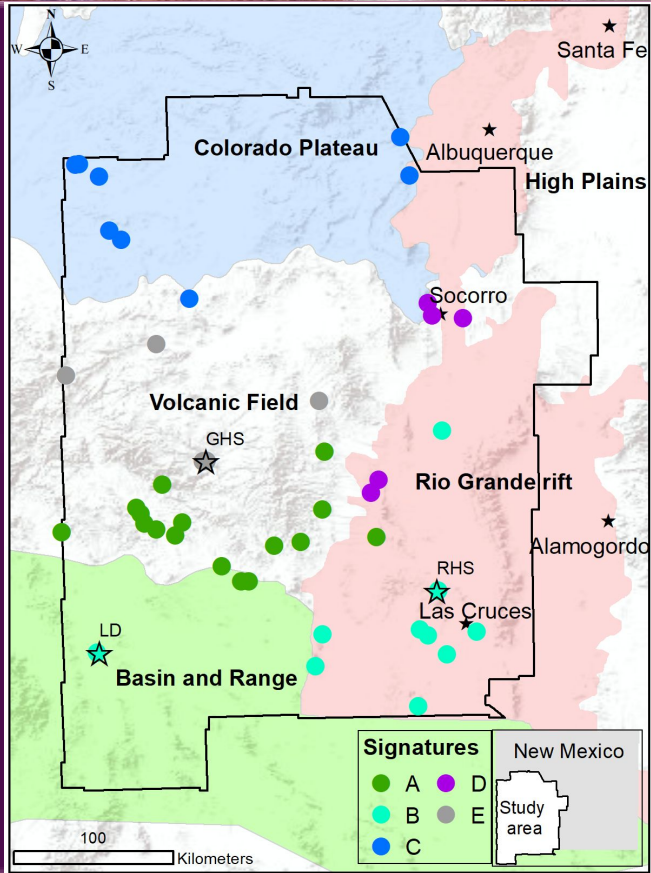
- Interactive cloud-based software
- Applies unsupervised and supervised science-informed ML
- Supports characterization and utilization of geothermal
- Provides:
 - Data management and visualization
 - ML analyses
 - Prospectivity assessment and mapping
 - Real-time ML demonstrations
- <http://geotgo.com>
- <http://98.60.118.144:8989>
- Expanding the capabilities of GeoThermalCloud
<https://github.com/SmartTensors/GeoThermalCloud.il>



GeoTGo: Southwest New Mexico Geothermal Analysis



GeoTGo: Southwest New Mexico Geothermal Analysis



Signature A Southern volcanic field
Shallow heat flow

Signature B Rio Grande Rift
Deep heat flow

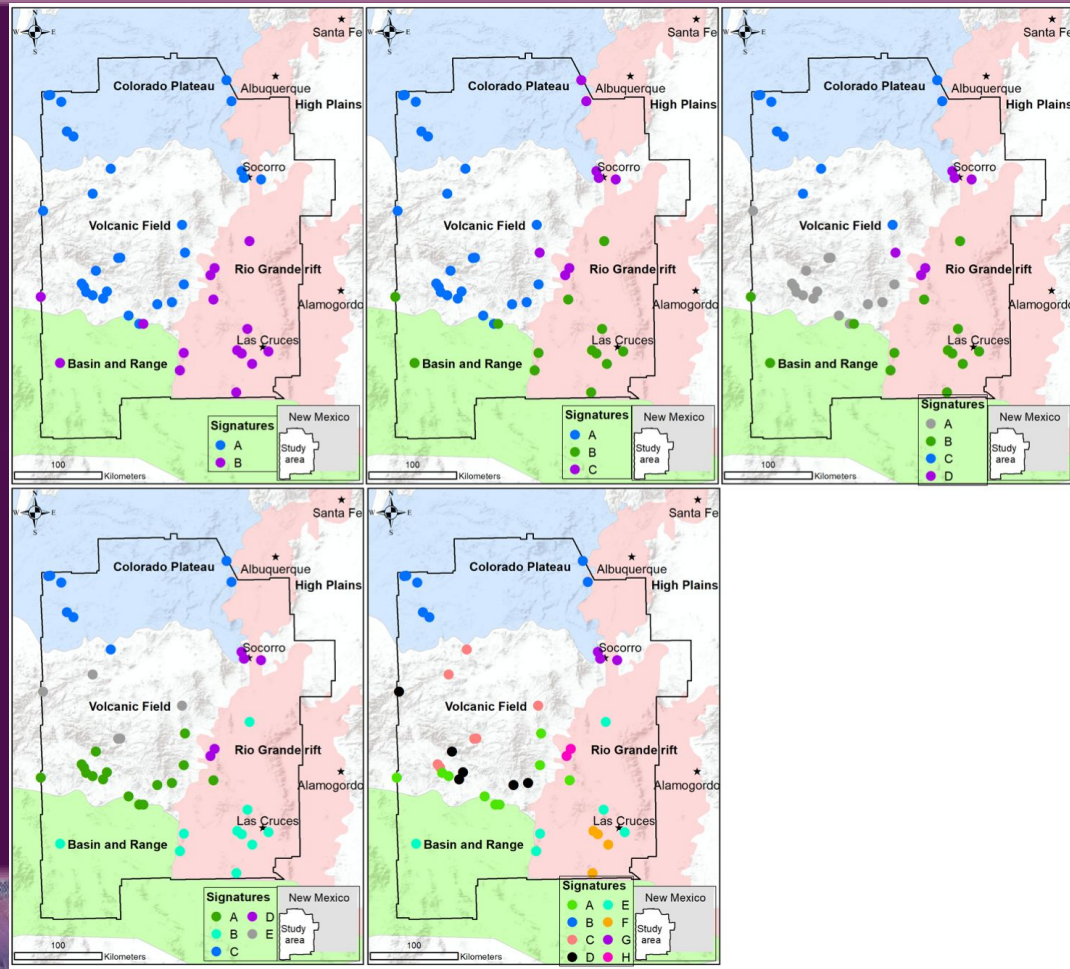
Signature C Colorado Plateau
Thick crust

Signature D Central Rio Grande Rift
Tectonics

Signature E Northern volcanic field
Vertical hydraulics

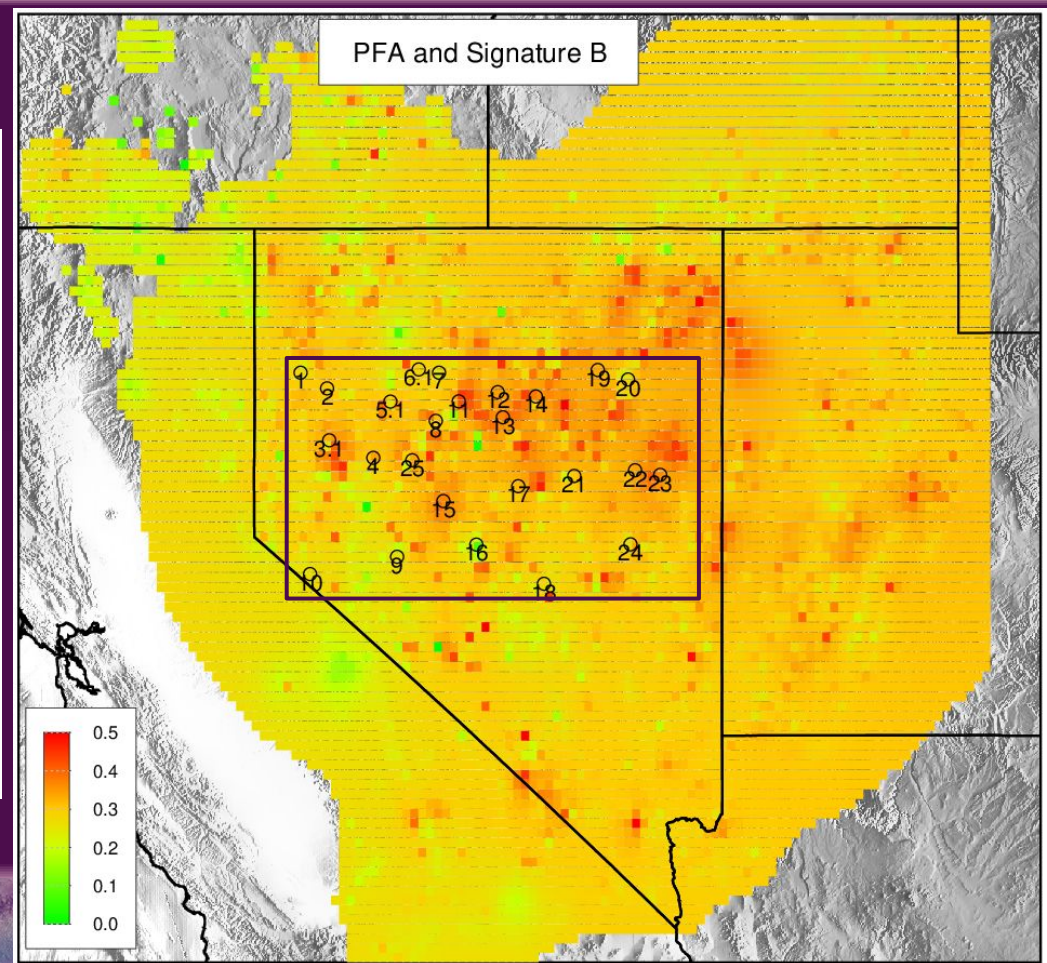
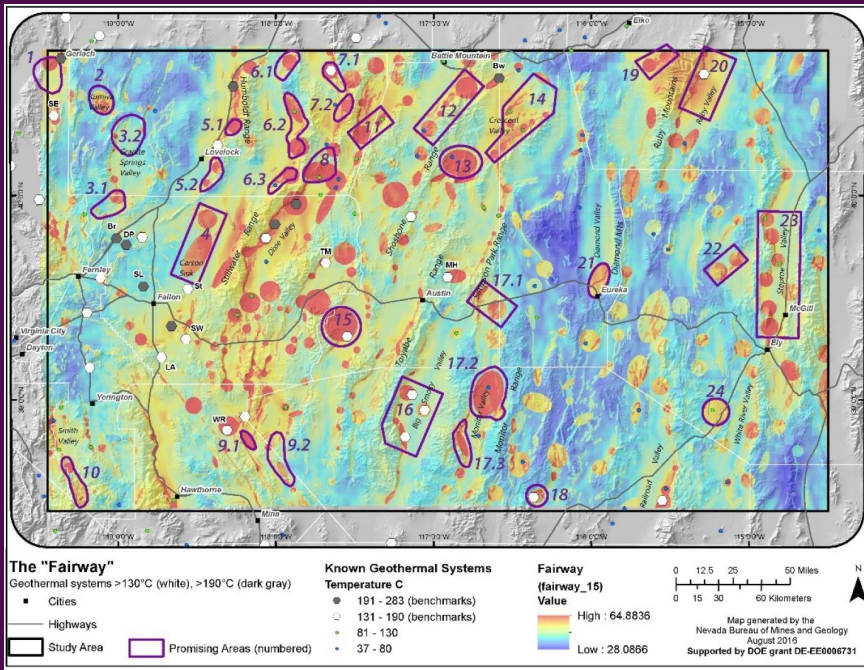
GeoTGo: Southwest New Mexico Geothermal Analysis

- 2, 3, 4, 5, and 8 signatures also can explain the dataset
- 5 signatures are optimal
- 2, 3, and 4 signatures are undefitting
- 8 signatures are overfitting
- Nevertheless, results for 2, 3, 4, 5, and 8 signatures provide data categorization consistent with regional physiographic provinces



GeoTGo: Great Basin Geothermal Predictions vs PFA

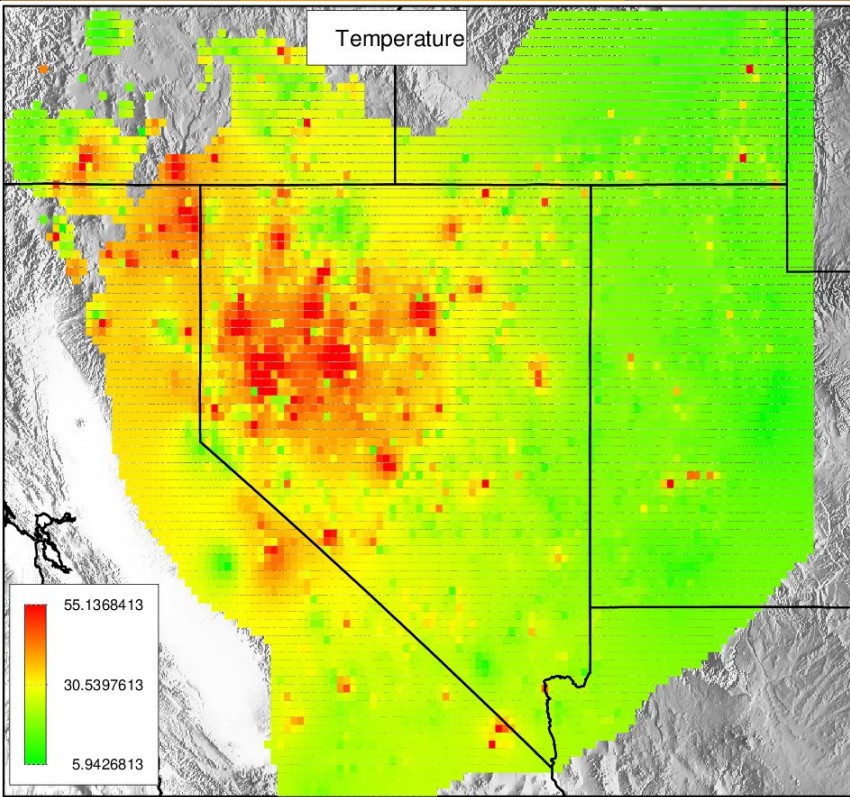
High-prospective areas estimated using PFA
(Faulds et al. 2017)



Source: <https://gbcge.org/recent-projects/nevada-play-fairway-phase-2>

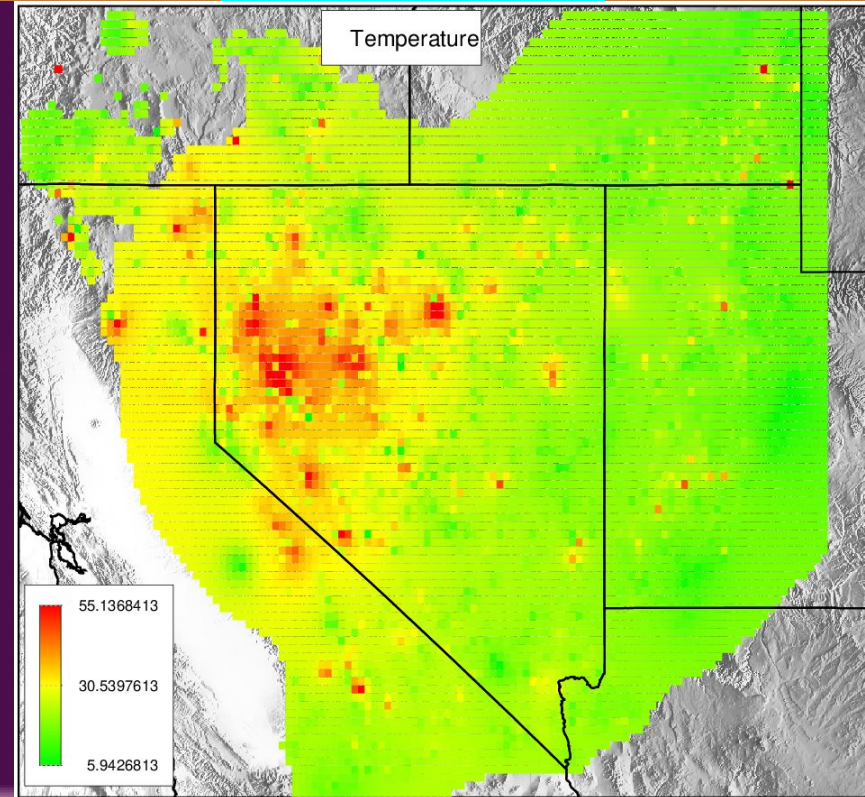
GeoTGo: Great Basin Geothermal Predictions

Data



Temperature

ML prediction

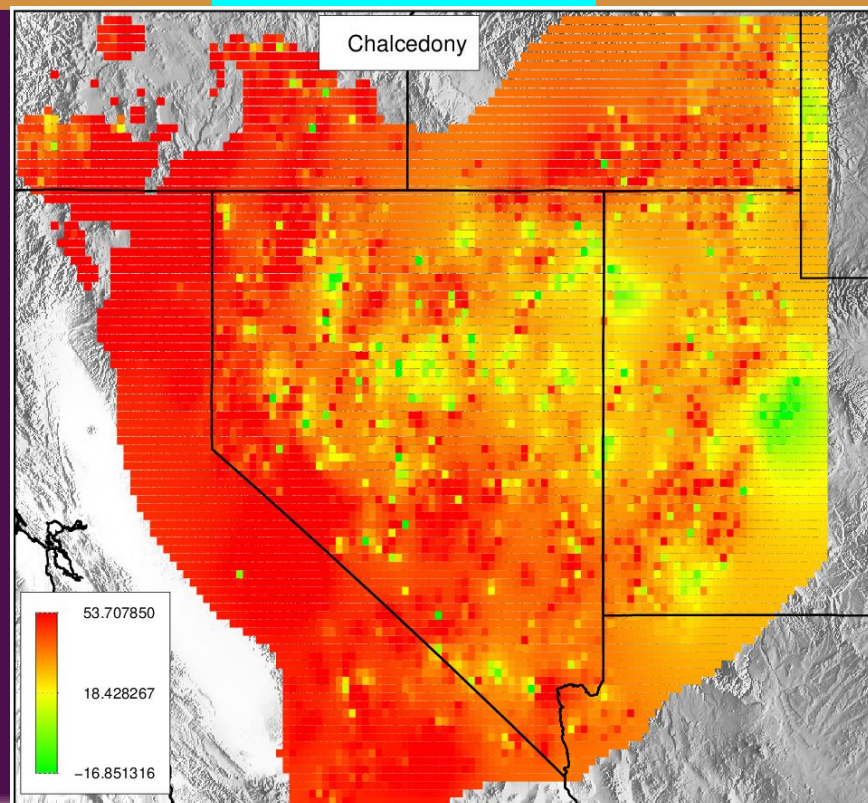
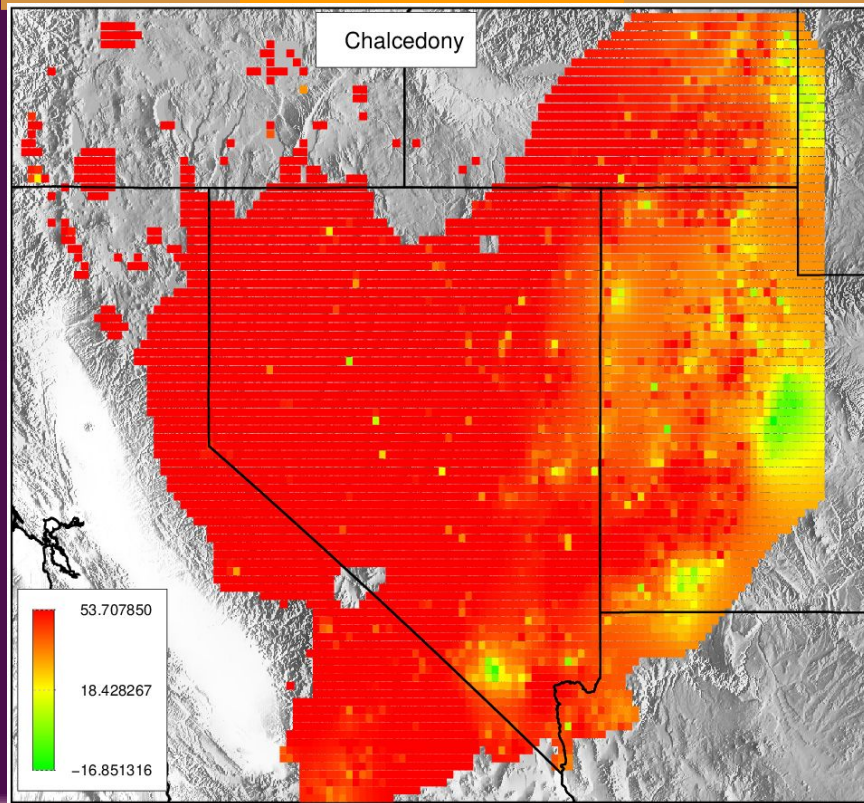


GeoTGo: Great Basin Geothermal Predictions

Data

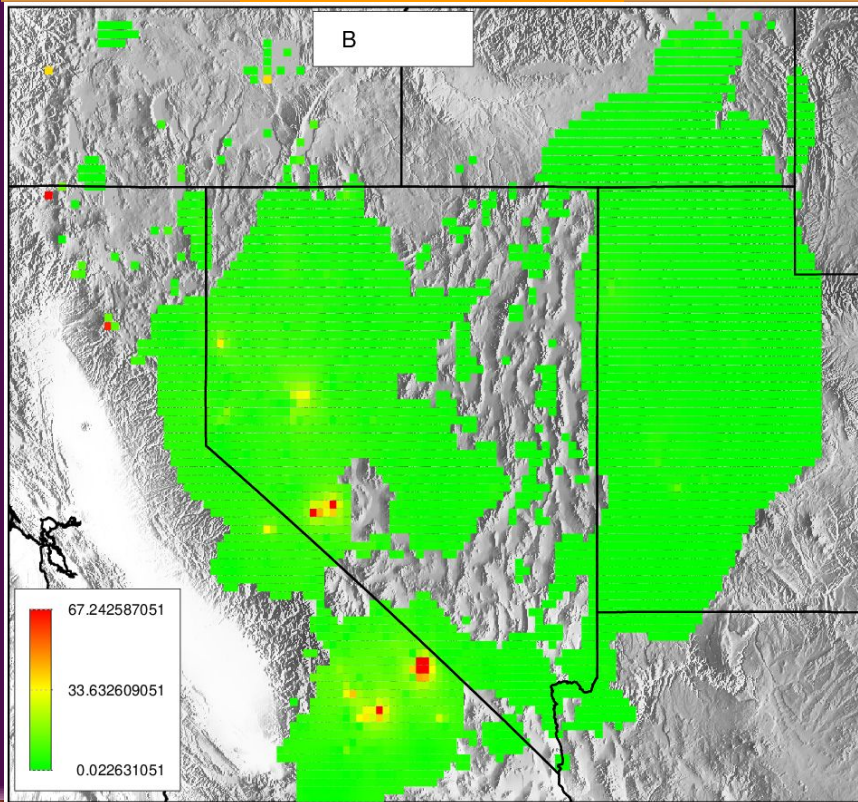
Chalcedony

ML prediction



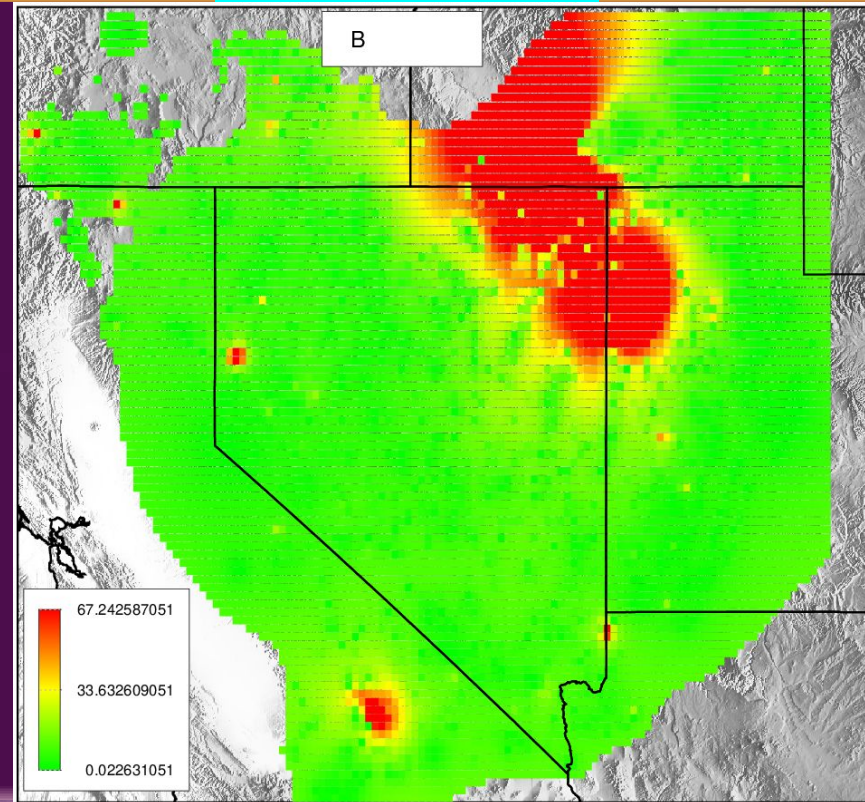
GeoTGo: Great Basin Geothermal Predictions

Data



B

ML prediction

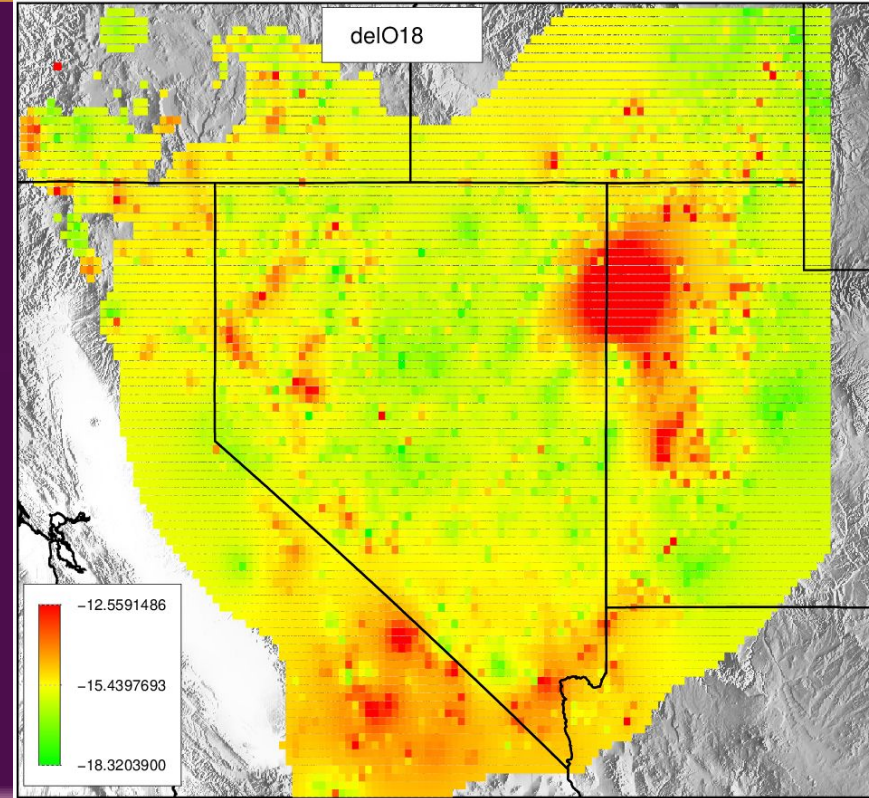
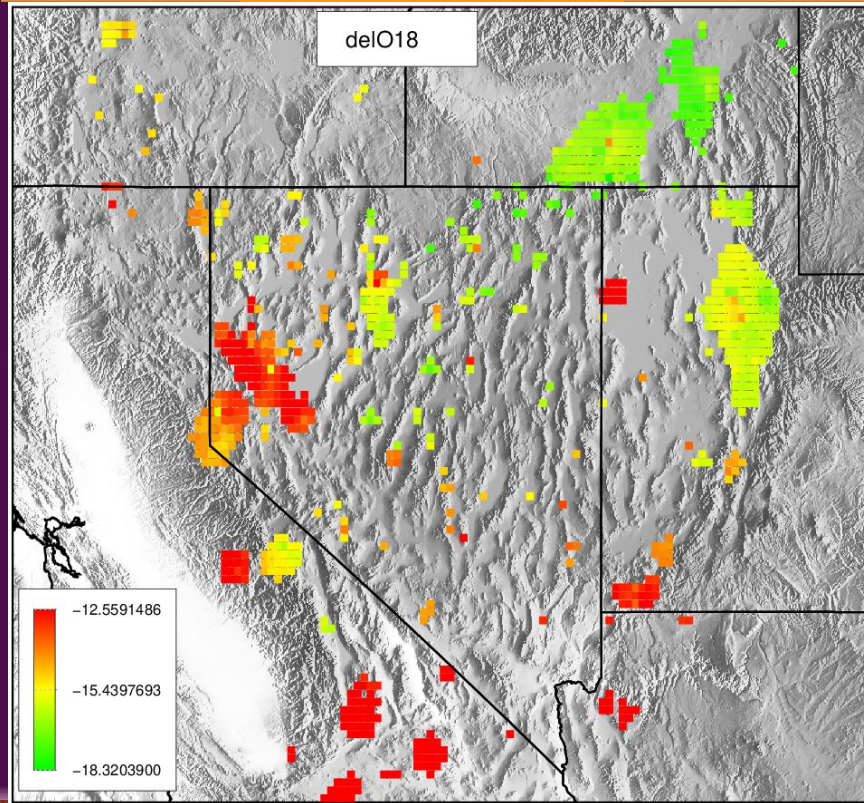


GeoTGo: Great Basin Geothermal Predictions

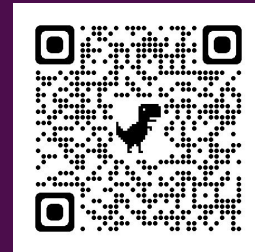
Data

delO18

ML prediction



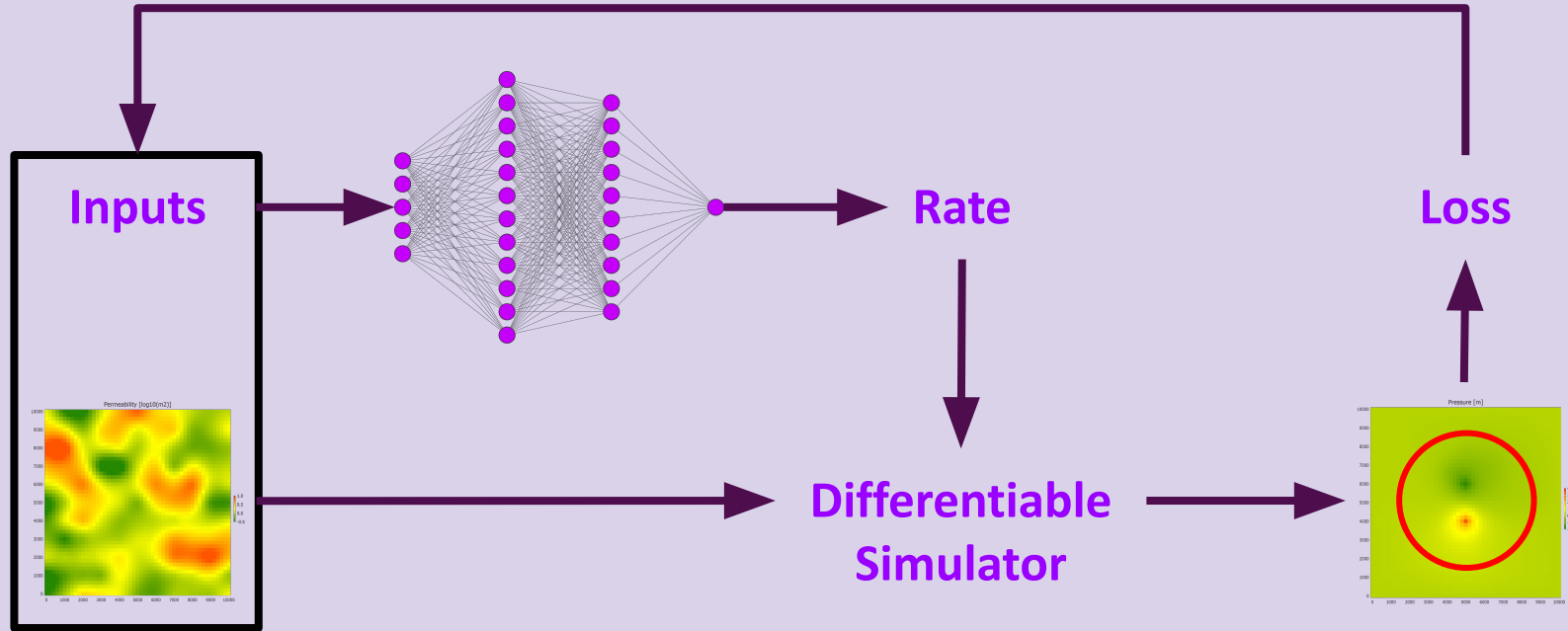
- Interactive cloud-based software
- Applies unsupervised and supervised science-informed ML
- Supports characterization and utilization of reservoirs in multi-use geologic engineering projects
- Provides:
 - Data management and visualization
 - ML analyses
 - Prospectivity assessment and mapping
 - Real-time ML demonstrations
- <http://geo-ml.com>
- <http://98.60.118.144:8988>



SIML workflow in GeoML

- Alternative uses (fluid injection only vs coupled injection/extraction)
 - Different number of wells
 - Radius around the well system within which pressure changes need to be controlled (minimized)
 - Maximum allowed pressure change
 - Reservoir heterogeneity (permeability, porosity, etc.)
-
- Based on the provided information, GeoML computes optimal injection/extraction rates
 - For a given reservoir heterogeneity structure, creates a prospectivity map

SIML training in GeoML



Our AI/ML Projects & Software



ChemML: Contaminant characterization and remediation
(DOE SBIR)



GeoTGo: Development and utilization of geothermal resources
(DOE SBIR)



GeoML: Analysis and interpretation of geochemical and geophysical data
(DOE SBIR)



DAPE: Data Assimilation and Model Parameterization for Earth Science
(NASA SBIR)



AI-PREDICT: Predicting Dynamics of Fracture Systems and Induced Seismicity
(DOE SBIR)



UrbanAI: Urban Energy Solutions Using AI
(DOE SBIR)

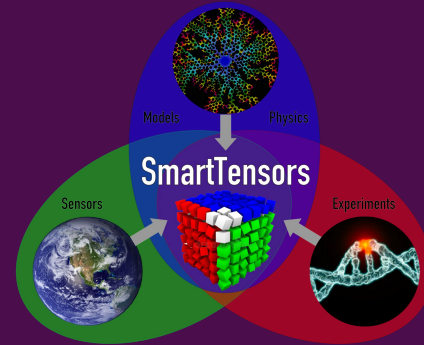
Software

- SmartTensors: AI/ML framework unsupervised, supervised, science-informed ML
 - SmartTensors.com
 - <https://github.com/SmartTensors>
 - [NMFk](#), [NTFk](#), ...
- MADS: Model Analysis & Decision Support <https://github.com/madsjulia/Mads.jl>
- ChemML: Characterization and remediation of contamination sites
 - <https://github.com/EnviTrace/ChemML>
- GeoTGo: Characterization and utilization of geothermal energy resources
 - GeoTGo.com
 - <https://github.com/EnviTrace/GeoTGo>
- GeoML: Characterization and multi-use utilization of geologic reservoirs
 - Geo-ML.com
 - <https://github.com/EnviTrace/GeoML>

Software support

- Examples
- Tests
- Notebooks
- Documentation
- User interfaces (dashboards)

**R&D
100**



Summary

- Harnessing science information and knowledge in AI/ML is critical
- Development of AI/ML models is an art
- Development of Science-Informed AI/ML (SIML) methods and tools is even more challenging
- Our software is designed to apply SIML on the cloud for characterization and improved understanding/prediction of environmental/earth-science problems

