GeoThermalCloud

Cloud Fusion of Big Data and Multi-Physics Models using Machine Learning for Discovery, Exploration and Development of Hidden Geothermal Resources

Project PI: Velimir ("monty") Vesselinov

GeoThermalCloud

LA-UR-21-24124

Project Motivation

- **Geothermal exploration and production are challenging, expensive and risky**
- **Diverse datasets available** (public and proprietary; satellite, airborne surveys, vegetation/water sampling, geological, geophysical, etc.)
- **How to utilize these datasets for geothermal exploration unknown due to imperfect understanding of how:**
  - **physical geothermal processes impact subsurface conditions**
  - **geothermal subsurface conditions are represented in observations**
- **ML is here to help …** (discover how geothermal conditions are represented in these datasets)

Project Goals

- Develop a general (flexible) open-source cloud-based ML framework for geothermal exploration
- Apply ML to discover and extract new (unknown/hidden) geothermal signatures in existing large datasets
- Categorize geothermal datasets and generate labels
- Identify high-value data acquisition strategies
- Fuse big data and multi-physics models
- Test and validate developed ML framework

Project Partners

- LANL
- Stanford University
- Google
- Descartes Labs
- University of Texas-Austin (Bureau of Economic Geology)

Phase 2 partners will include:

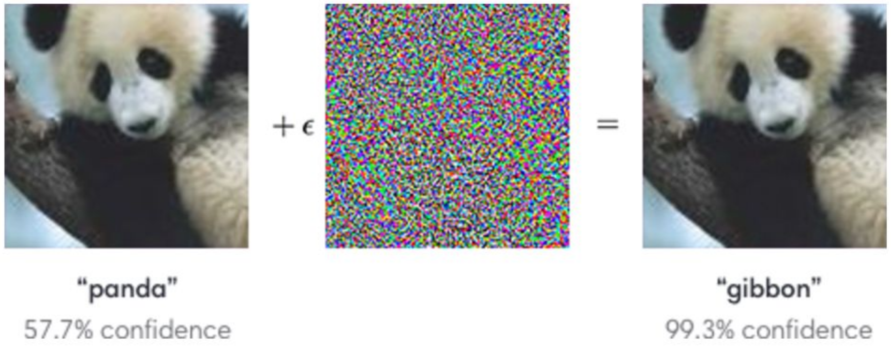
- Julia Computing
- PNNL
- Chevron

Machine Learning (ML) methods

- **Supervised**
- **Unsupervised**
- **Physics-informed**

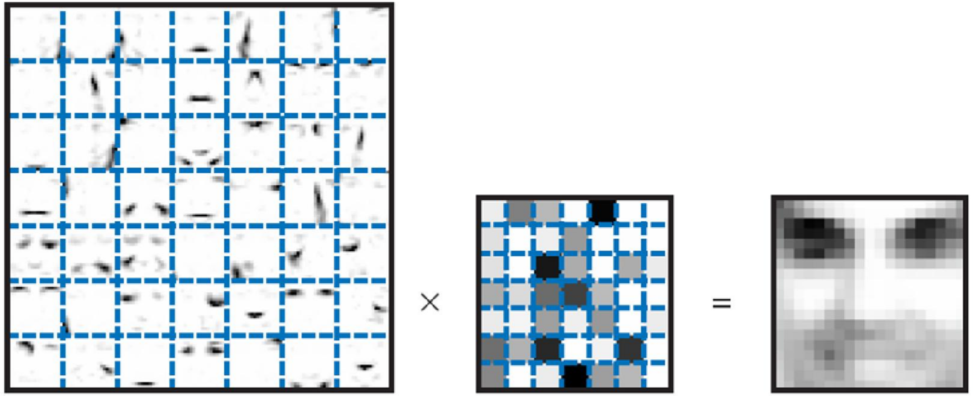
Supervised ML

- learns everything from data
- requires prior “labeling” (i.e., knowledge about the processed data)
- SME needed before ML is performed
- cannot discover/learn something that is not known already
- requires large training datasets
- highly impacted by noise in the data
- neural networks are difficult to interpret ⇒ black box analyses
- can recognize cats and dogs but cannot recognize horses if not pre-trained



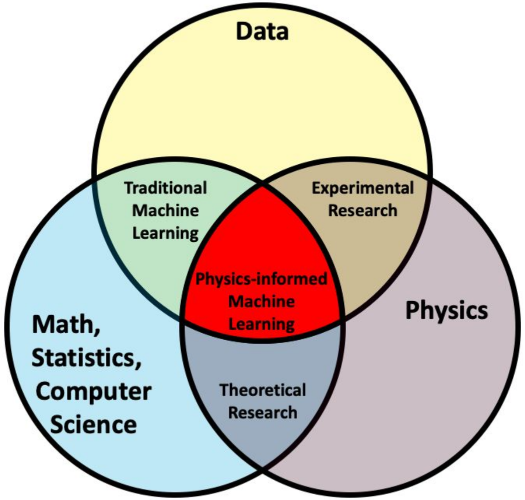
Unsupervised ML

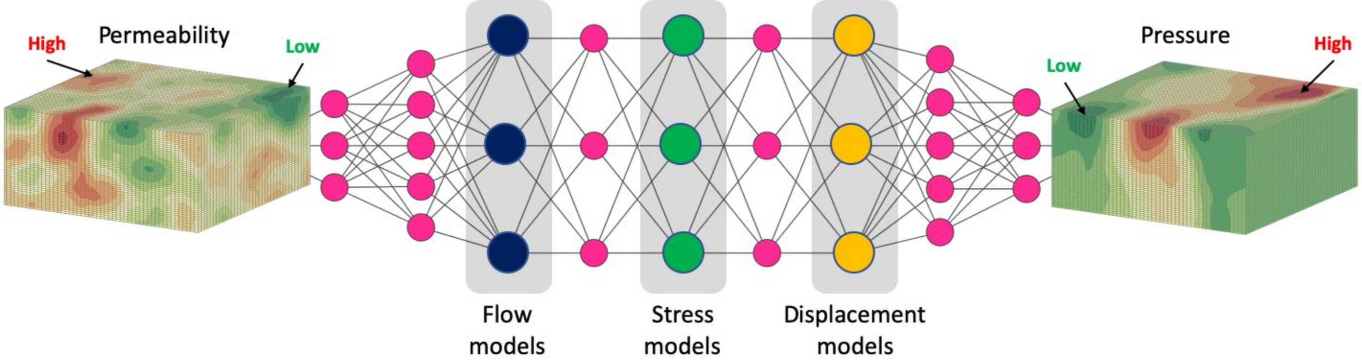
- extracts information (features/signatures) from data automatically
- applicable for both categorization and prediction
- produces unbiased analyses not impacted by data labeling, subject-matter-expert (SME) opinions, and physics assumptions
- identifies features that distinguish images of animals (e.g., cats, dogs, horses, etc.) or geothermal features
- categorizes data and SMEs can identify (“label”) animals (or geothermal features)
- SME needed after ML is performed



Physics-informed ML (PIML)

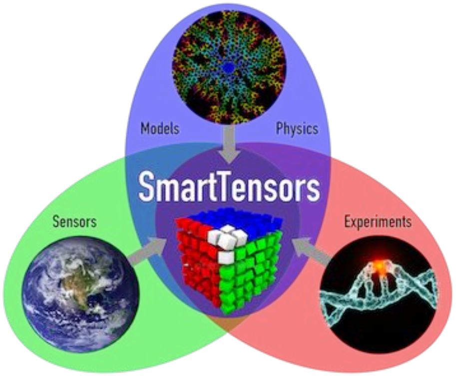
- learns from data but includes preconceived science knowledge
- physics information embedded in the ML framework or added as penalties
- PIML neural networks are problem specific
- needs SME inputs related to the analyzed problem
- SME needed before and after ML is performed
- increases efficiency, accuracy, and robustness
- requires differentiable programming (julia)





SmartTensors

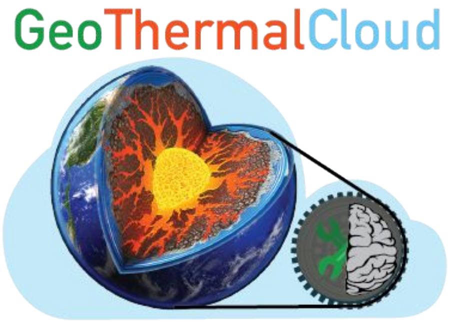
- SmartTensors open-source framework incorporates novel LANL-developed patented ML methods and tools based on matrix/tensor factorization
- SmartTensors can perform unsupervised and physics-informed ML (PIML)
- Non-negativity and physics constraints can be added ⇒ provide explainability
- SmartTensors extensively tested and validated
- … and applied for diverse problems (from COVID-19 to wildfires and text mining)
- Can efficiently process large datasets (TB's) utilizing GPU's & TPU's (Chennupati et al., 2020)
- Coded in julia; orders of magnitude faster than Python, R and MATLAB; allows differentiable programming
- SmartTensors is actively developed
- SmartTensors nominated for an R&D 100 award

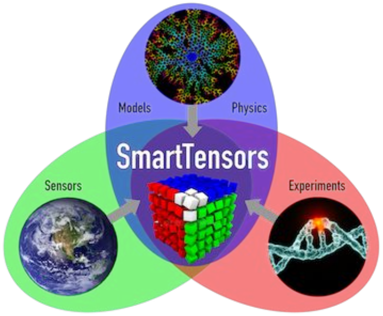


GeoThermalCloud + SmartTensors

- **GeoThermalCloud incorporates SmartTensors ML tools**

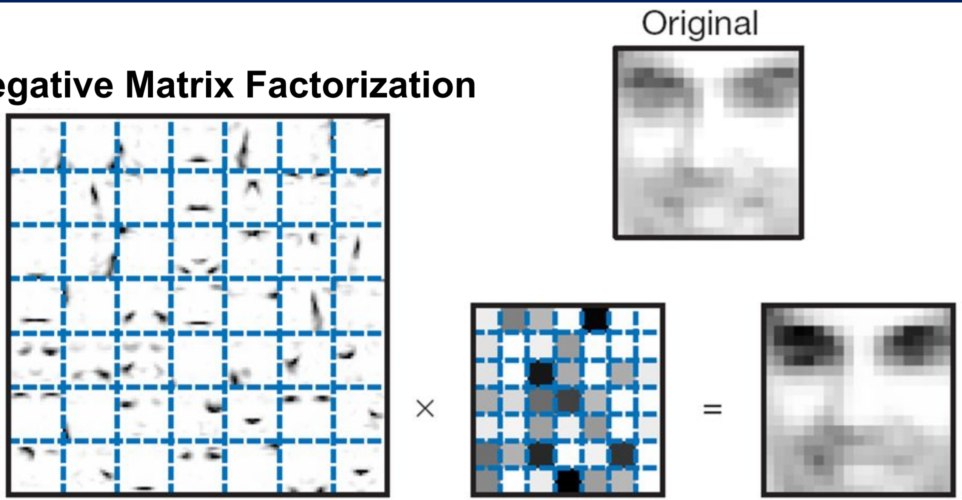
- **GDR: https://gdr.openei.org/submissions/1297**
- **GitHub:**
  - **https://github.com/SmartTensors**
  - **https://github.com/SmartTensors/GeoThermalCloud.jl**
- **JuliaHub cloud computing and data management (Phase 2)**
  - **Julia Computing: leading research company in the area of PIML and data analytics**



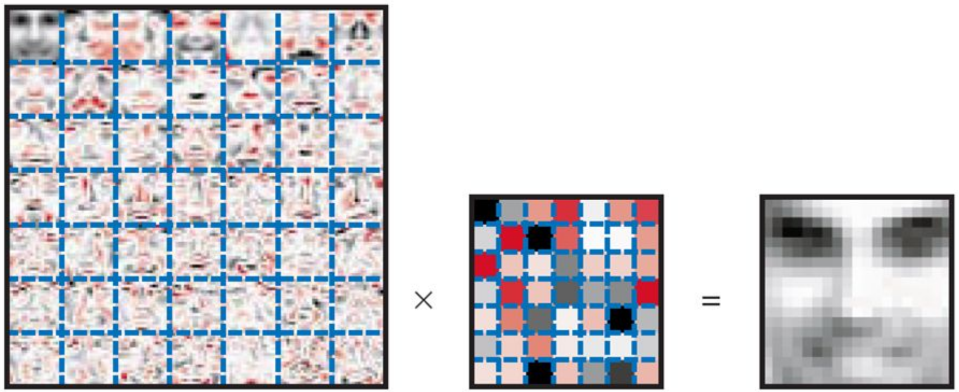


NMF vs PCA

NMF: Nonnegative Matrix Factorization



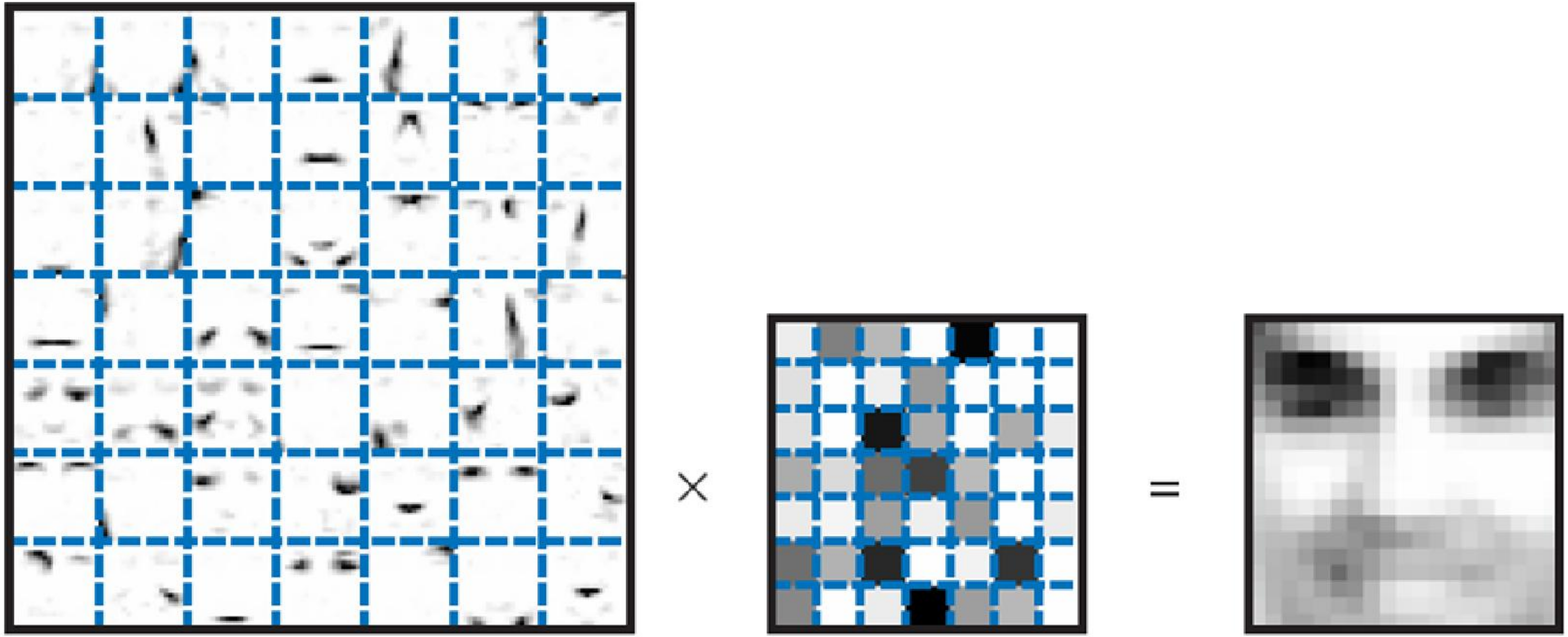
PCA: Principal Component Analysis



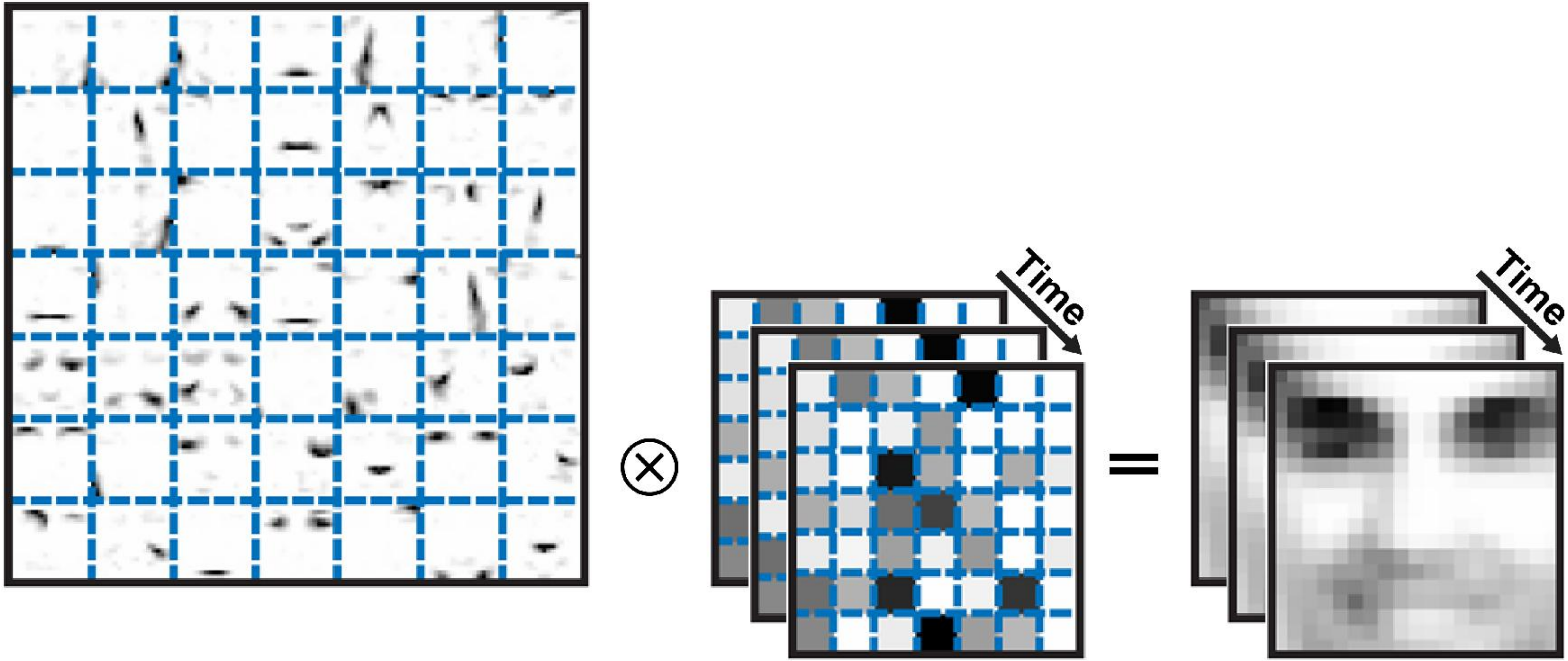
Nonnegativity constraints provide meaningful and interpretable results (and sparsity)

Lee & Seung, 1999

Nonnegative matrix factorization



Nonnegative tensor factorization



Machine Learning for unmixing waters

Let us assume there are 4 buckets representing 4 different groundwater types

Machine Learning for unmixing waters

- Water from the 4 buckets is mixed in unknown fashion in the subsurface
- Mixing is caused by various ill-defined processes

Machine Learning for unmixing waters

- Water compositions of the original water types (buckets) are unknown
- Groundwater mixtures observed in the monitoring wells are only known

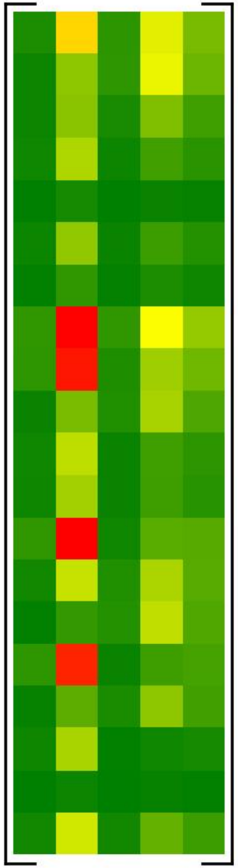
Machine Learning for unmixing waters

- Using observed mixtures (even if data gaps), the bucket composition can be estimated
- Water unmixing can be done using Machine Learning (ML)

Vesselinov et al, 2016, Contaminant source identification using semi-supervised machine learning, J. Contam. Hydrol.

Vesselinov et al, 2018, Nonnegative tensor factorization for contaminant source identification. J. Contam. Hydrol.

Nonnegative matrix factorization



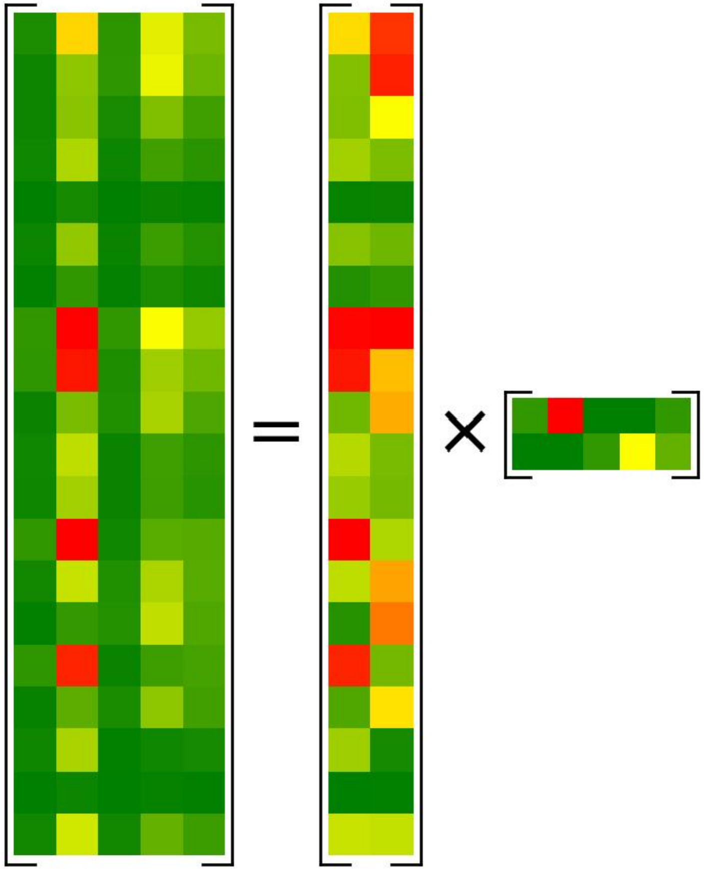
$\boldsymbol{X}$

$[20 \times 5]$

$\boldsymbol{X}$ – **data** matrix

[**attributes** $\times$ **locations**]

Nonnegative matrix factorization



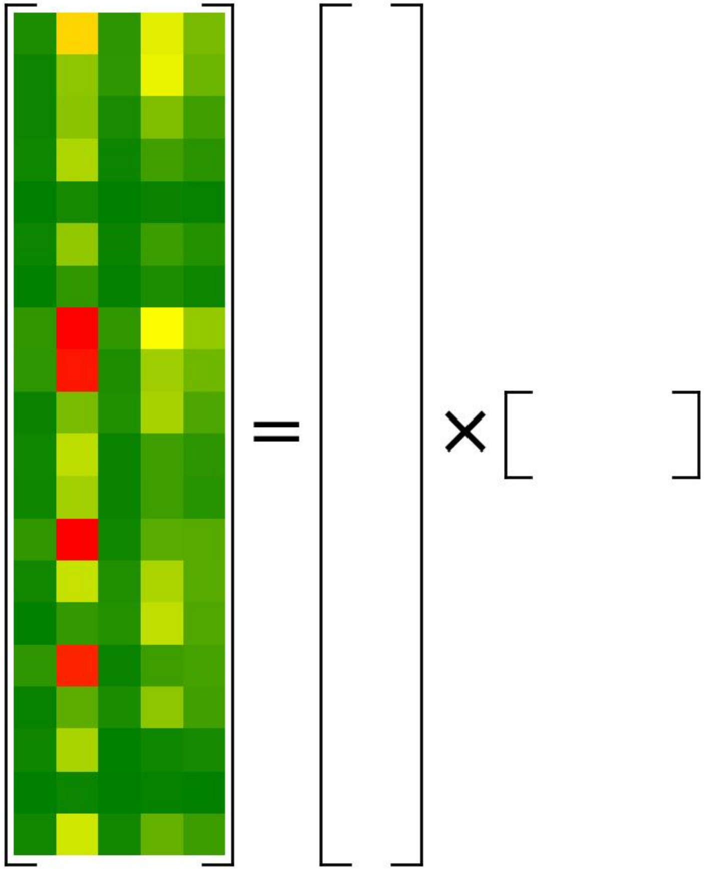
$$\boldsymbol{X} = \boldsymbol{W} \times \boldsymbol{H}$$
$$[20 \times 5] = [20 \times 2] \times [2 \times 5]$$

$\boldsymbol{X}$ – **data** matrix
[**attributes** × **locations**]

$\boldsymbol{W}$ – **feature** (**signal**) matrix
[**attributes** × **signatures**]

$\boldsymbol{H}$ – **mixing** matrix
[**signatures** × **locations**]

Nonnegative matrix factorization



$$\boldsymbol{X} = \boldsymbol{W} \times \boldsymbol{H}$$

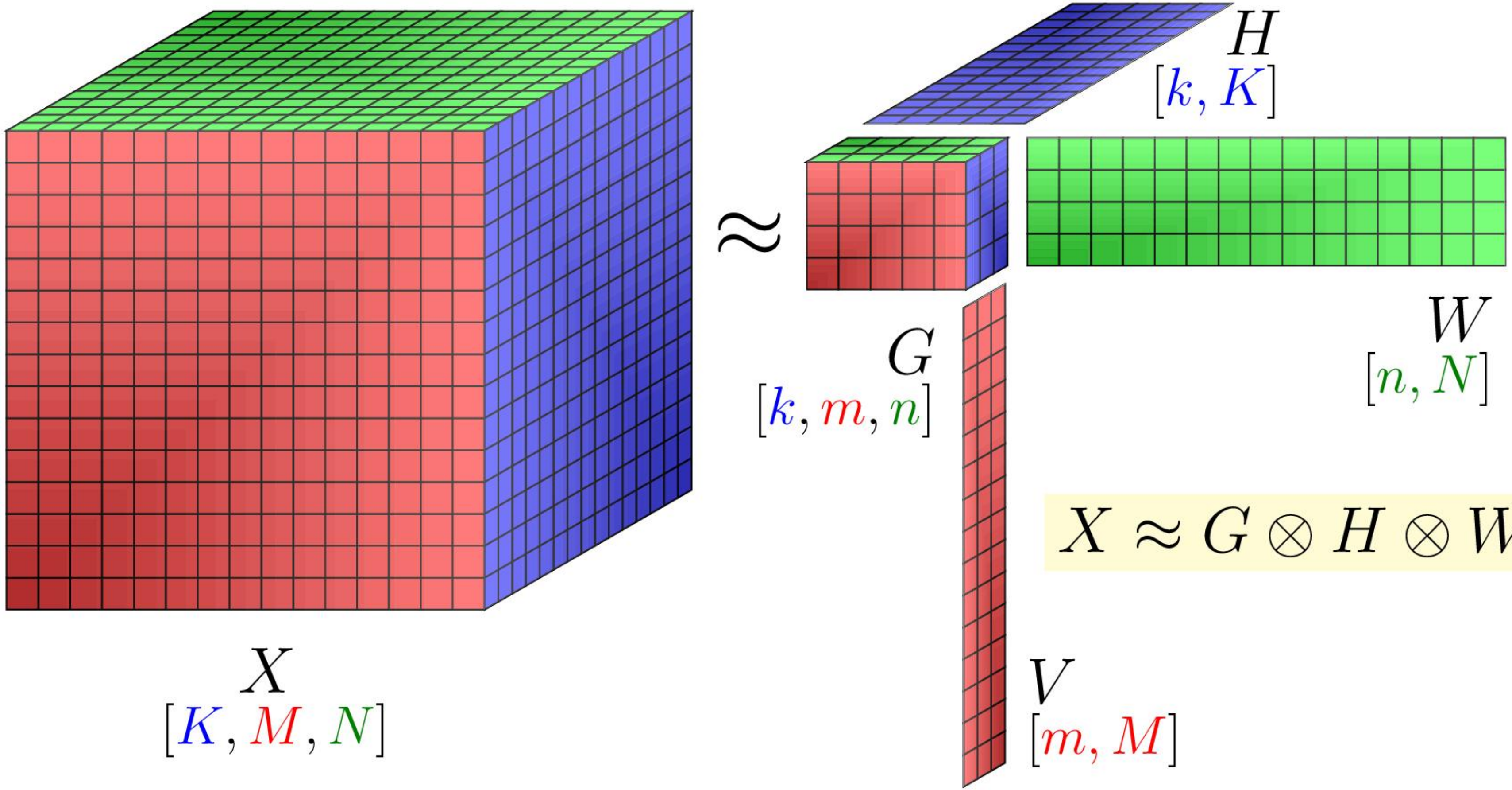
$$[20 \times 5] = [20 \times \textbf{?}] \times [\textbf{?} \times 5]$$

$\Rightarrow$ 100 **knowns**

$\Rightarrow$ **unknown** number of signatures (2 or more)

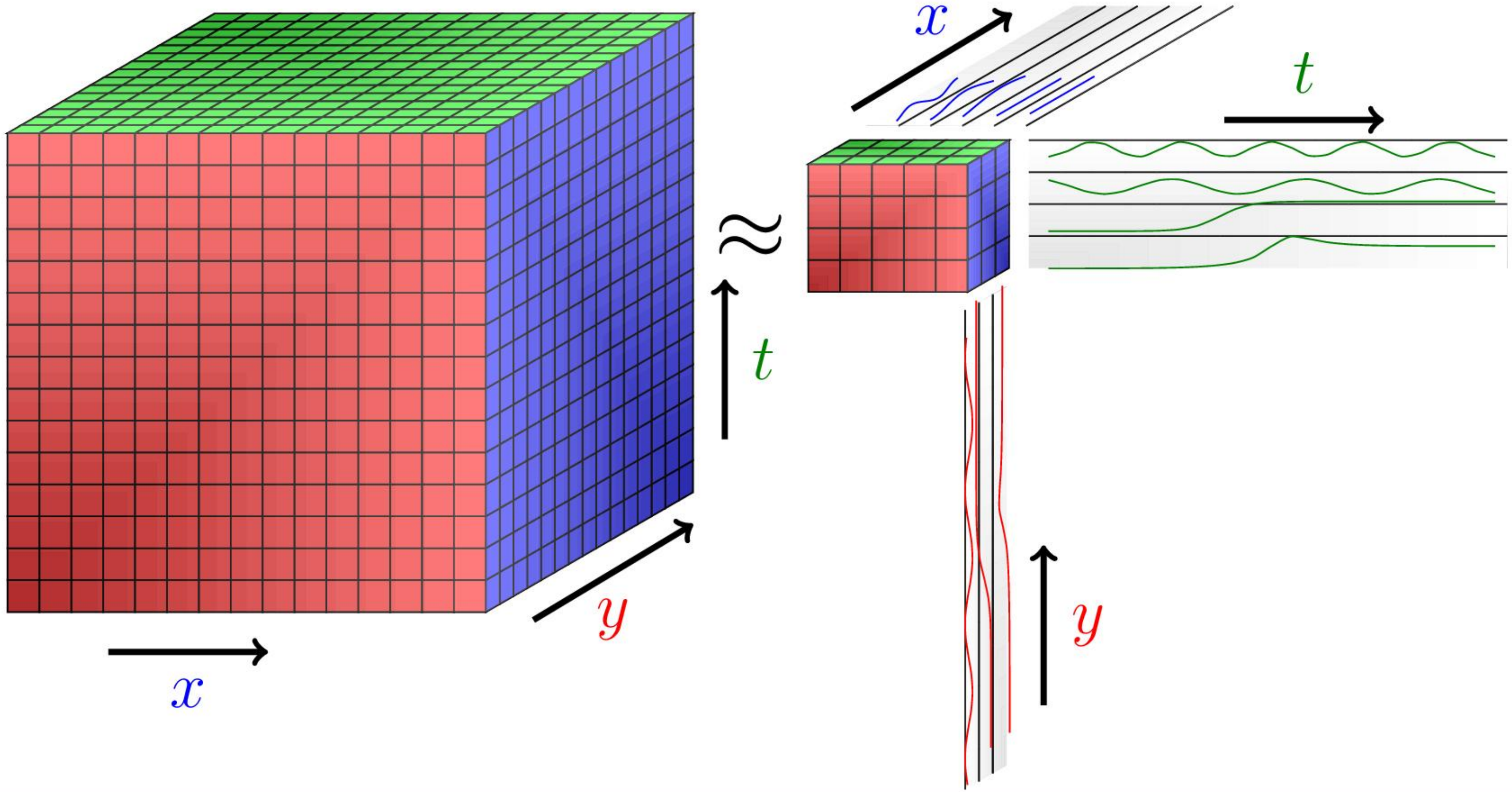
$\Rightarrow$ **unknown** matrix elements of $\boldsymbol{W}$ and $\boldsymbol{H}$ (50 or more)

Nonnegative tensor factorization

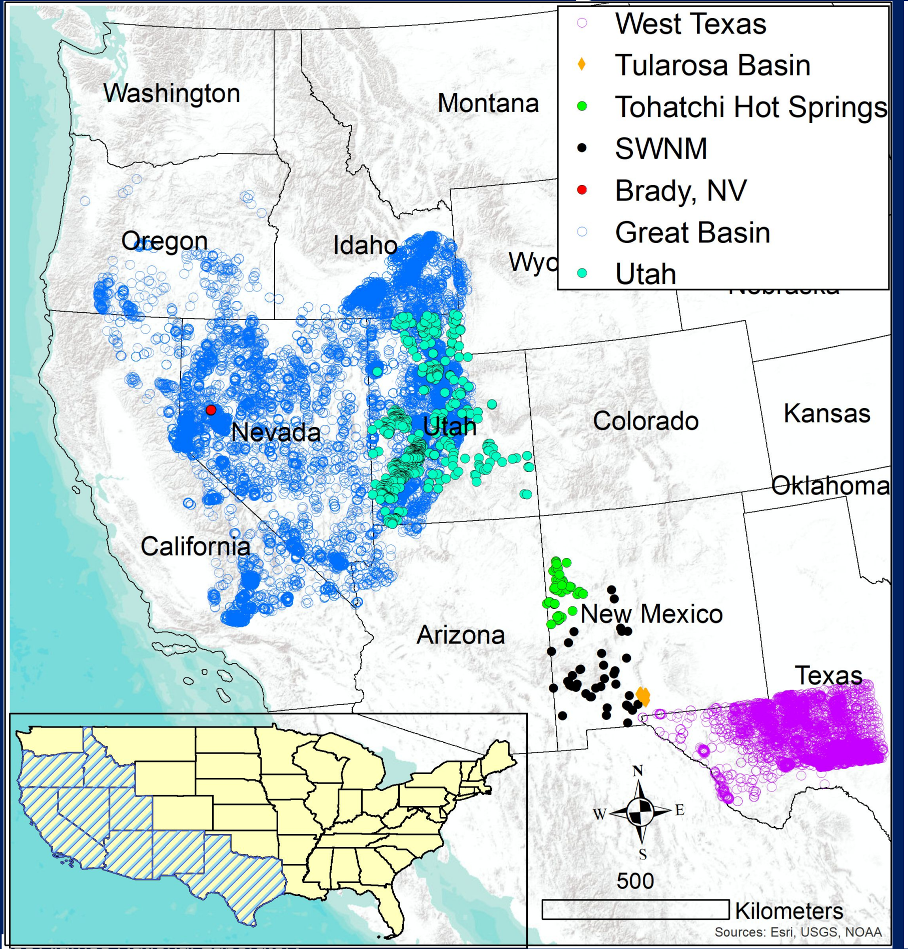


$$X \approx G \otimes H \otimes W \otimes V$$

Nonnegative tensor factorization

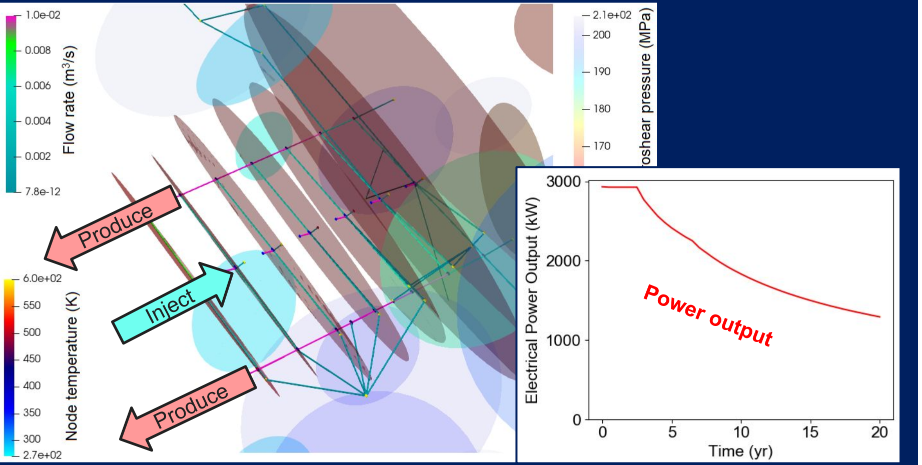


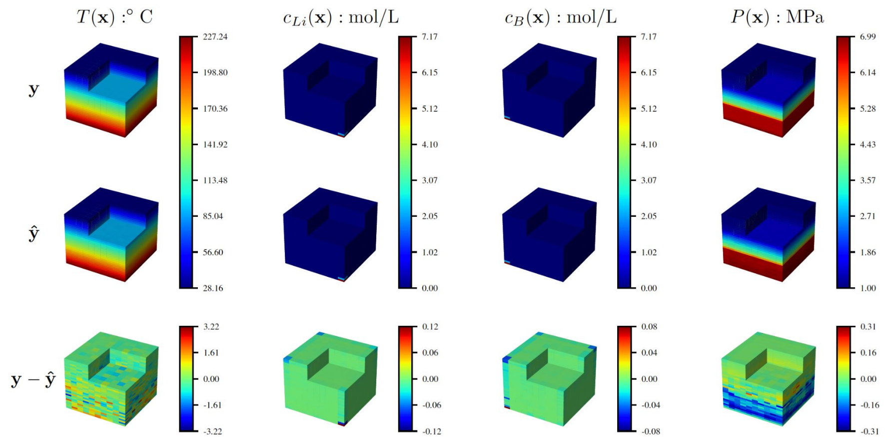
Datasets analyzed: Field



- Analyzed datasets include geothermal, geophysical, geomechanical, geochemical, geological attributes
- Covering various regions/conditions: NV, UT, CA, OR, ID, NM, TX, HI

Datasets analyzed: Synthetic





- **EGS energy production at UtahForge**

  ML using LANL's code GeoDT
  to optimize geothermal energy production

- **SWNM geothermal systems**

  ML using LANL's code PFLOTRAN
  to characterize geothermal heat sources

GeoDT Synthetic Dataset

GeoDT is a novel LANL-developed multi-physics modeling tool (Frash, 2021)

GeoDT designed to rapidly predict the performance of geothermal systems including induced seismicity

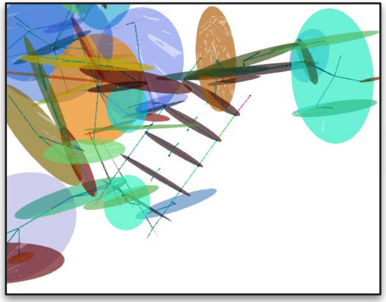
GeoDT accounts for:

1. *Well system design (spacing, dip, well diameter, etc.)*
2. *Natural fractures*
3. *Hydraulic fractures*
4. *State of stress*
5. *Stress-dependent fracture properties*

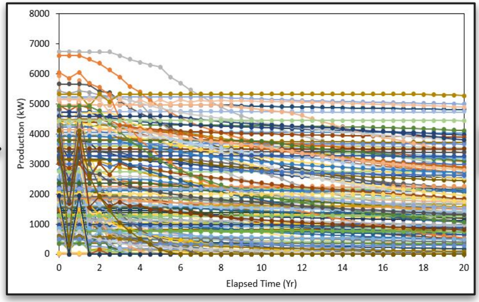
GeoDT integrated with GeoThermalCloud to find optimal behavioral trends

GeoThermalCloud + GeoDT analysis

- **Goals:**
  - **Find relations between site conditions and production transients**
  - **Identify site parameters that increase energy production**
  - **Characterize the impact of state of stress on the geothermal production**
  - **Develop ML model to efficiently predict the system behavior**
- **GeoDT predicts geothermal performance based on attainable site data**
- **GeoThermalCloud "separates" impacts of physics processes in model outputs to identify multivariate factors that control geothermal production**



GeoDT →



NMF*k* →

Output	A	B	C	D	Signal
Cumulatiove injection rate	0.51	0.06	0.00	0.88	D
Cumulative production rate	0.08	0.94	0.25	0.00	B
Boundary outflow rate	0.50	0.98	0.16	0.75	B
Boundary inflow rate	0.04	0.00	0.00	0.00	A
Production rate / Injection rate	0.48	1.00	0.18	0.75	B
Maximum induced earthquake magnitude	0.79	0.32	0.31	0.19	C
Pressure of injected fluid	0.00	0.00	0.41	0.11	C
Enthalpy of injected fluid	0.00	0.05	0.85	0.33	C
Number of fractures intercepting injectors	0.35	0.11	0.03	0.67	D
Number of fractures intercepting producers	0.15	0.18	0.00	0.60	D
Number of stimulated hydraulic fractures	0.37	0.03	0.01	0.96	D
Number of stimulated natural fracutres	0.20	0.00	0.05	0.00	A
Production mass flow rate	0.10	0.94	0.25	0.00	B
	Well spacing	Well design	Stress	Well dip	

GeoDT Model Parameters (25 analyzed)

Los Alamos NATIONAL LABORATORY EST. 1943

Note: a partial list of ~100 GeoDT input parameters that will be analyzed (Phase 2)

Category	Parameter	Unit	Nominal	Min	Max	Category
Thermal effects	Reservoir depth	m	6000	2000	10000	
Thermal effects	Geothermal gradient	K/km	50	20	60	
Thermal effects	Rock thermal conductivity	W/mK	2.5	2.1	2.8	
Thermal effects	Rock volumetric specific heat capacity	kJ/m3K	2063	1900	2200	
	Rock elastic modulus	GPa	50	30	90	Stress effects
	Rock Poisson's ratio	m/m	0.3	0.15	0.4	Stress effects
	Rock shear modulus	GPa	19.2	13.0	32.1	Stress effects
	Minimum lateral pressure coefficient	Pa/Pa	0.5	0.3	0.9	Stress effects
	Intermediate pressure coefficient	Pa/Pa	0.75	0.3	1.5	Stress effects
Fracture geometry	Fracture count set#1	Count	10	0	60	
Fracture geometry	Fracture count set#2	Count	10	0	60	
Fracture geometry	Fracture count #set3	Count	10	0	60	
Fracture geometry	Fracture roughness	-	0.8	0.7	1	
	Well spacing	m	300	100	800	Well design
	Well length	m	600	400	1600	Well design
	Well azimuth	deg	0	-90	90	Well design
	Well dip	deg	0	0	90	Well design
	Borehole/Casing radius	m	0.076	0.051	0.178	Well design
Combined effects	Reservoir pore pressure	MPa	57.8	19.3	96.2	
Combined effects	Reservoir temperature	C	325	50	635	
Combined effects	Overburden stress	Pa	158.9	45.1	274.7	
Combined effects	Intermediate stress	Pa	133.6	27.1	363.9	
Combined effects	Minimum stress	Pa	108.4	27.1	256.8	
	Cohesion	MPa	10	5	15	Strength effects
	Friction Angle	Degrees	35	20	50	Strength effects

GeoDT: ML results (Parameters ⇒ Signals)

Parameters are automatically identified to be correlated

Parameter	A	B	C	D	Signal
Reservoir depth	0.18	0.00	1.00	0.03	C
Geothermal gradient	0.44	0.75	0.16	0.26	B
Rock thermal conductivity	0.25	0.89	0.05	0.41	B
Rock volumetric specific heat capacity	0.17	0.84	0.08	0.41	B
Rock elastic modulus	0.25	0.92	0.00	0.39	B
Rock Poisson's ratio	0.14	0.90	0.08	0.42	B
Rock shear modulus	0.25	0.84	0.00	0.36	B
Minimum lateral pressure coefficient	0.00	0.85	0.12	0.49	B
Intermediate pressure coefficient	0.03	0.86	0.12	0.47	B
Fracture count set#1	0.21	0.95	0.04	0.39	B
Fracture count set#2	0.16	0.83	0.12	0.38	B
Fracture count set#3	0.21	0.91	0.05	0.37	B
Fracture roughness	0.20	0.85	0.10	0.37	B
Well spacing	1.00	0.57	0.19	0.00	A
Well length	0.23	0.97	0.05	0.35	B
Well azimuth	0.23	0.87	0.06	0.40	B
Well dip	0.22	0.00	0.00	1.00	D
Borehole/Casing radius	0.03	1.00	0.00	0.50	B
Reservoir pore pressure	0.26	0.08	0.79	0.05	C
Reservoir temperature	0.18	0.00	1.00	0.03	C
Overburden stress	0.18	0.00	1.00	0.03	C
Intermediate stress	0.04	0.30	0.54	0.21	C
Minimum stress	0.00	0.23	0.64	0.19	C
Cohesion	0.18	0.98	0.01	0.42	B
Friction Angle	0.19	0.89	0.07	0.37	B
	Well spacing	System+ reservoir	Stress	Well dip	

EGS design interplays with reservoir properties (elastic, heat capacity, gradient, etc.)

Well spacing and well orientation are both crucial attributes

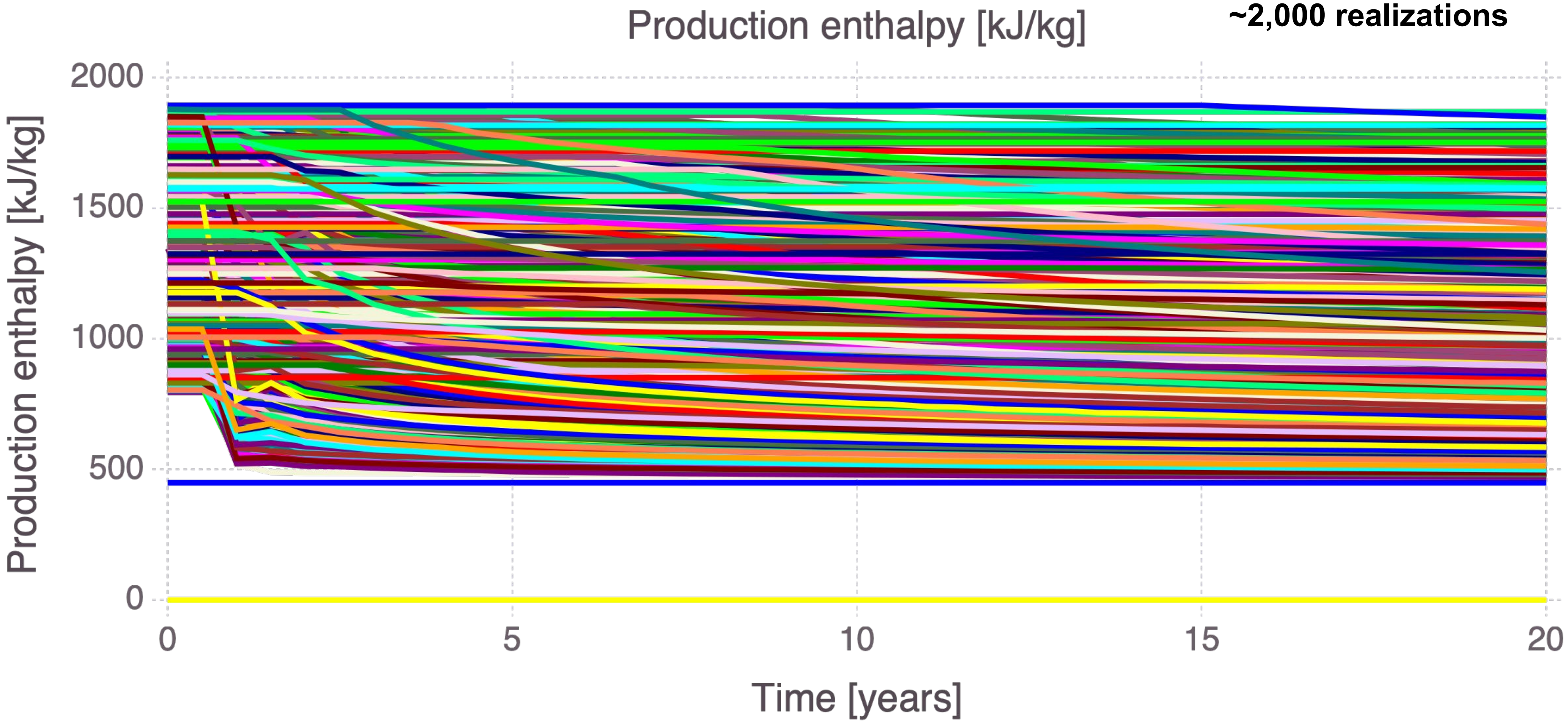
Well diameter is more important than originally anticipated

As to be expected, rock temperature has a crucial role

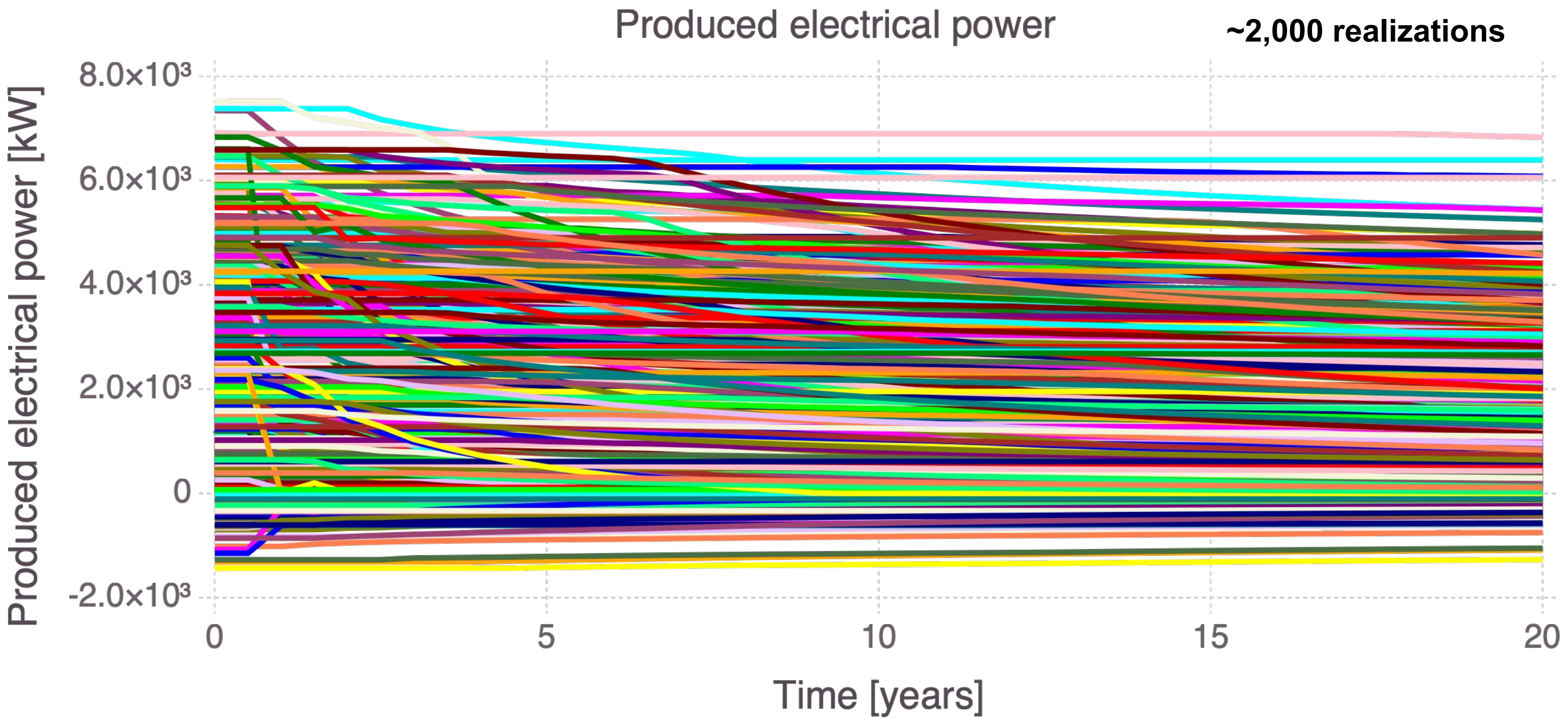
All three components of the stress tensor are important

GeoDT Model Outputs

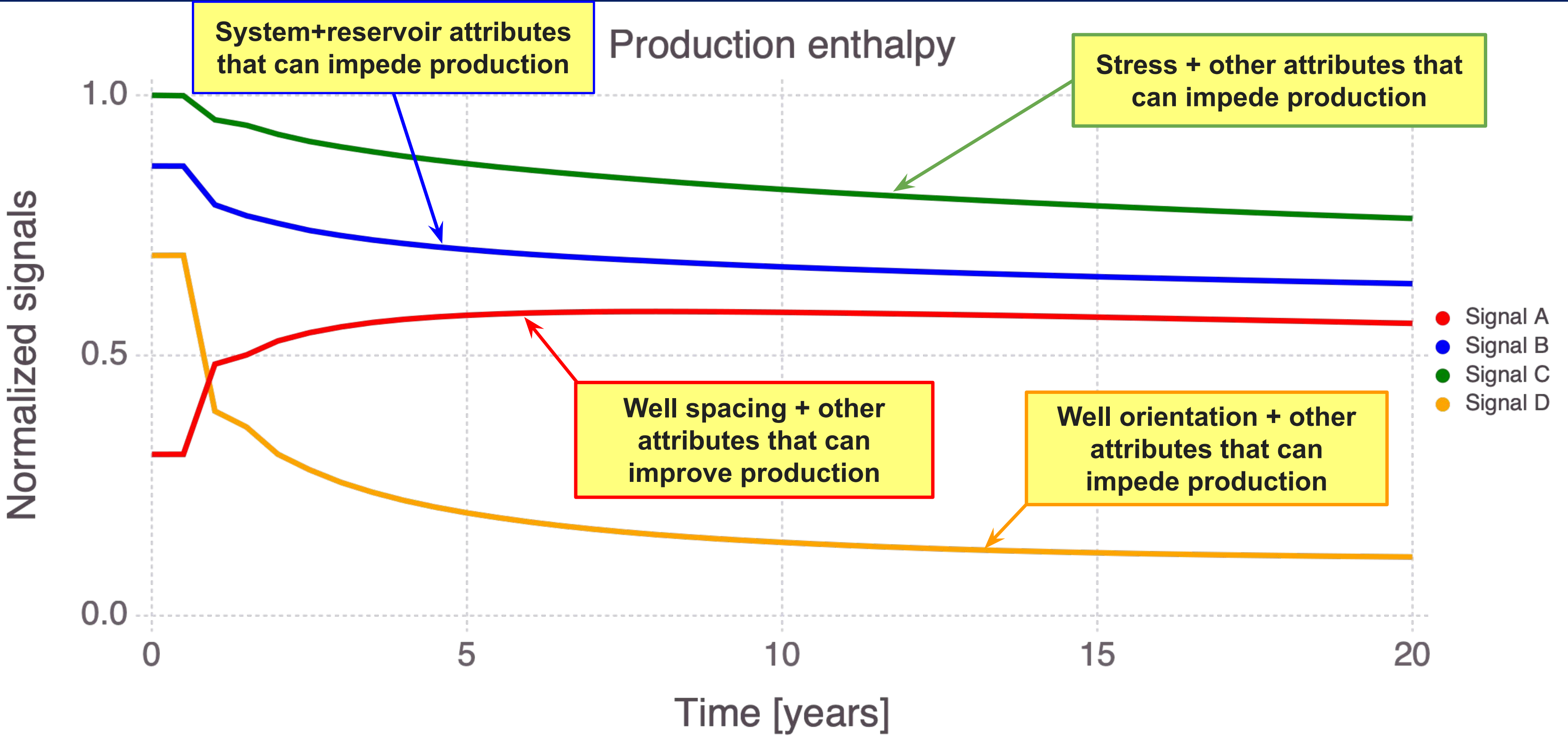
~2,000 realizations



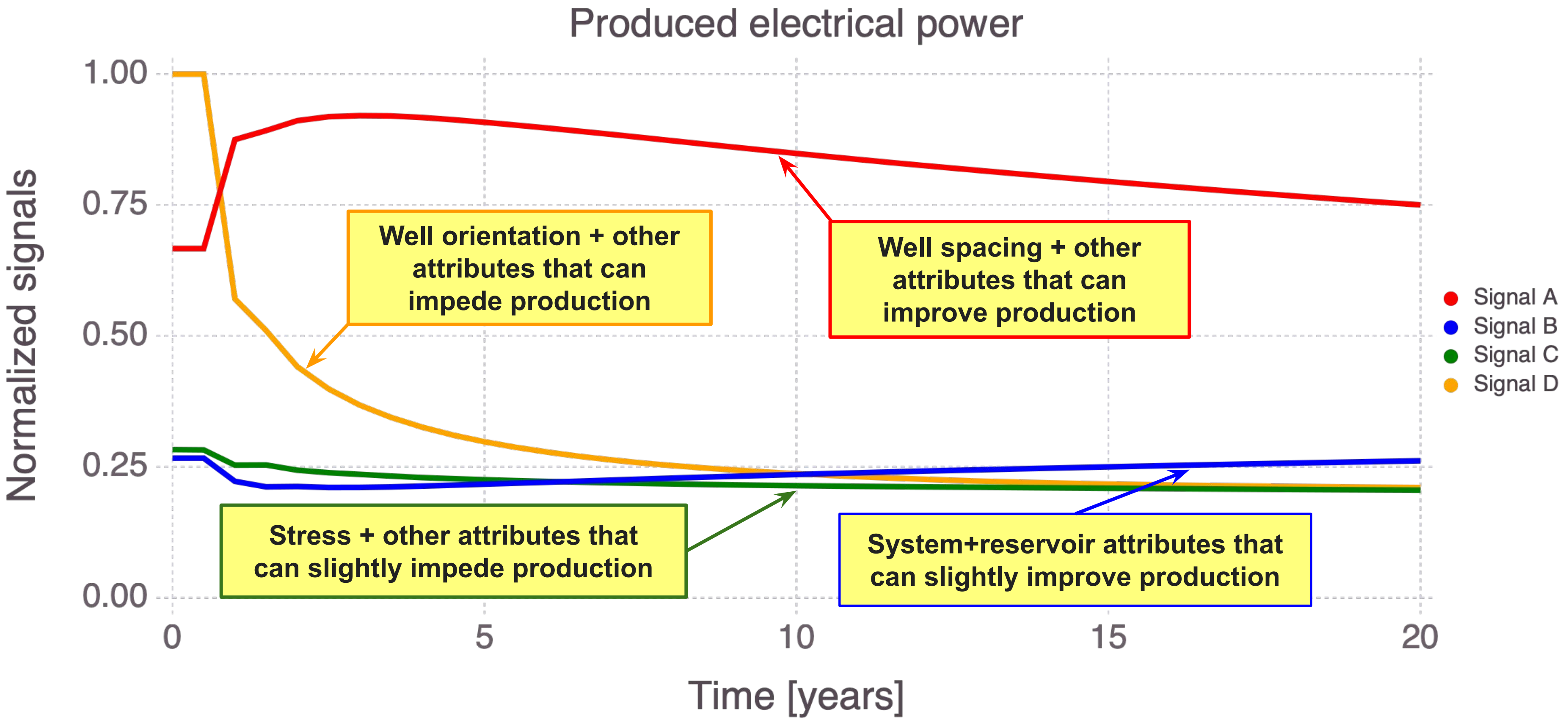
GeoDT Model Outputs



GeoDT: ML results (4 signals represent heat production)



GeoDT: ML results (2 signals represent power production)



GeoDT: ML results (Outputs ⇒ Signals)

Output	A	B	C	D	Signal
Cumulative injection rate	0.51	0.06	0.00	0.88	D
Cumulative production rate	0.08	0.94	0.25	0.00	B
Boundary outflow rate	0.50	0.98	0.16	0.75	B
Boundary inflow rate	0.04	0.00	0.00	0.00	A
Production rate / Injection rate	0.48	1.00	0.18	0.75	B
Maximum induced earthquake magnitude	0.79	0.32	0.31	0.19	C
Pressure of injected fluid	0.00	0.00	0.41	0.11	C
Enthalpy of injected fluid	0.00	0.05	0.85	0.33	C
Number of fractures intercepting injectors	0.35	0.11	0.03	0.67	D
Number of fractures intercepting producers	0.15	0.18	0.00	0.60	D
Number of stimulated hydraulic fractures	0.37	0.03	0.01	0.96	D
Number of stimulated natural fracutres	0.20	0.00	0.05	0.00	A
Production mass flow rate	0.10	0.94	0.25	0.00	B
	Well spacing	System+ reservoir	Stress	Well dip	

GeoDT: ML results (Outputs ⇒ Signals)

Output	A	B	C	D
Cumulative injection rate	0.51	0.06	0.00	0.88
Cumulative production rate	0.08	0.94	0.25	0.00
Boundary outflow rate	0.50	0.98	0.16	0.75
Boundary inflow rate	0.04	0.00	0.00	0.00
Production rate / Injection rate	0.48	1.00	0.18	0.75
Maximum induced earthquake magnitude	0.79	0.32	0.31	0.19
Pressure of injected fluid	0.00	0.00	0.41	0.11
Enthalpy of injected fluid	0.00	0.05	0.85	0.33
Number of fractures intercepting injectors	0.35	0.11	0.03	0.67
Number of fractures intercepting producers	0.15	0.18	0.00	0.60
Number of stimulated hydraulic fractures	0.37	0.03	0.01	0.96
Number of stimulated natural fracutres	0.20	0.00	0.05	0.00
Production mass flow rate	0.10	0.94	0.25	0.00
	Well spacing	System+ reservoir	Stress	Well dip

System+reservoir properties are strongly linked to leakoff risk

Well spacing is a strong factor for induced seismicity

Well orientation strongly controls interaction of natural and stimulated fractures

Stress is interlinked in a complex way to system performance

GeoDT: ML results (Outputs ⇒ Signals)

Output	A	B	C	D	Signal
Cumulative injection rate	0.51	0.06	0.00	0.88	D
Cumulative production rate	0.08	0.94	0.25	0.00	B
Boundary outflow rate	0.50	0.98	0.16	0.75	B
Boundary inflow rate	0.04	0.00	0.00	0.00	A
Production rate / Injection rate	0.48	1.00	0.18	0.75	B
Maximum induced earthquake magnitude	0.79	0.32	0.31	0.19	C
Pressure of injected fluid	0.00	0.00	0.41	0.11	C
Enthalpy of injected fluid	0.00	0.05	0.85	0.33	C
Number of fractures intercepting injectors	0.35	0.11	0.03	0.67	D
Number of fractures intercepting producers	0.15	0.18	0.00	0.60	D
Number of stimulated hydraulic fractures	0.37	0.03	0.01	0.96	D
Number of stimulated natural fracutres	0.20	0.00	0.05	0.00	A
Production mass flow rate	0.10	0.94	0.25	0.00	B
	Well spacing	System+ reservoir	Stress	Well dip	

System+reservoir properties are strongly linked to leakoff risk

Well spacing is a strong factor for induced seismicity

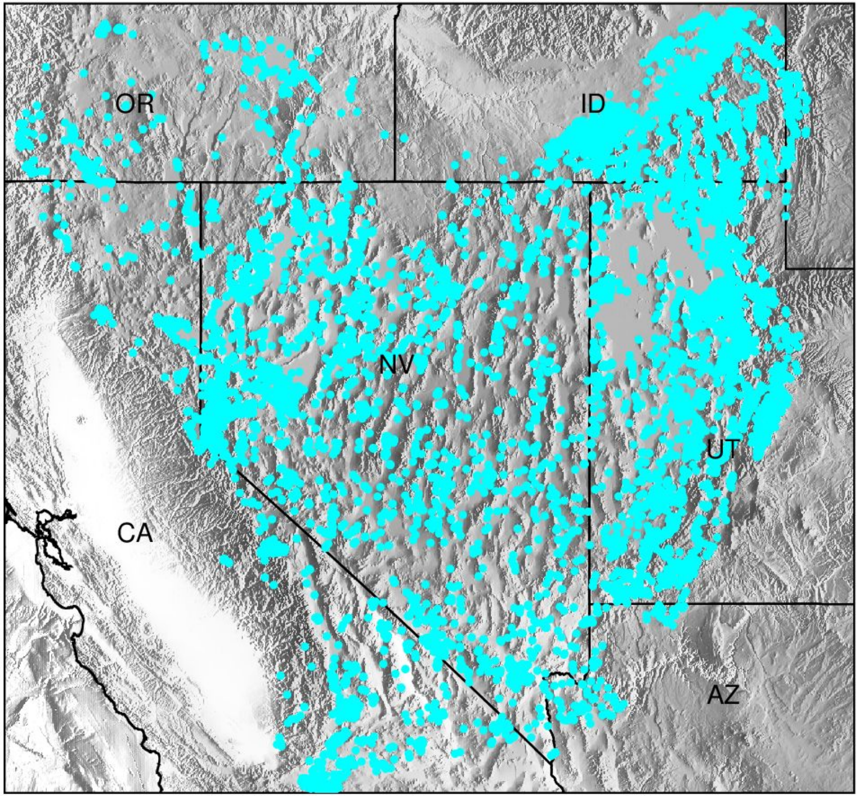
Well orientation strongly controls interaction of natural and stimulated fractures

Stress is interlinked in a complex way to system performance

ML analyses of GeoDT simulations:

- ML "separates" the impacts of different physical processes in the GeoDT model outputs (predicted heat, power production, etc.)
- Identifies interrelationship between GeoDT model outputs and GeoDT model inputs
- Well spacing and well orientation are identified to be critical parameters impacting energy production and induced seismicity

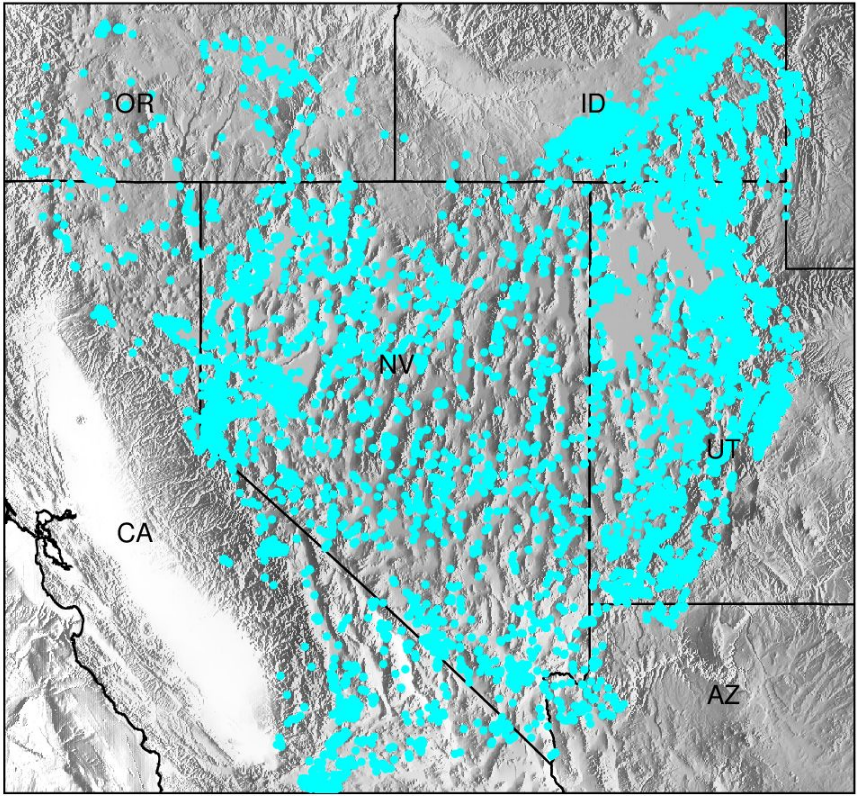
Great Basin



Study area with 14,341 data points

- **Great Basin includes multiple geothermal reservoirs**
- **Great Basin has huge geothermal potential**
- **Better understanding of local/regional spatial patterns in various geothermal-related attributes observed throughout the Great Basin region is needed**
- **> 14,000 locations at which geothermal-related data are available**

Great Basin: Why geochemistry?

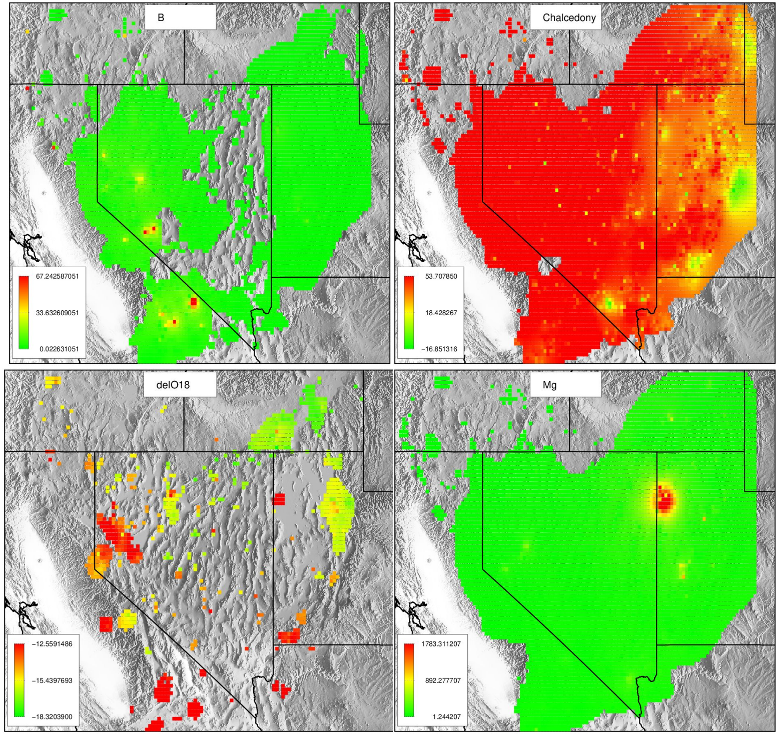


Study area with 14,341 data points

- **Geochemical data are easier to collect compared to other geothermal-related attributes**
- **Geochemistry can be applied to infer geothermal conditions** (e.g., reservoir temperatures, conditions, reservoir boundaries, and heat source type)
- **Geochemistry also captures water / rock interactions and water mixing**

Great Basin: Data Attributes

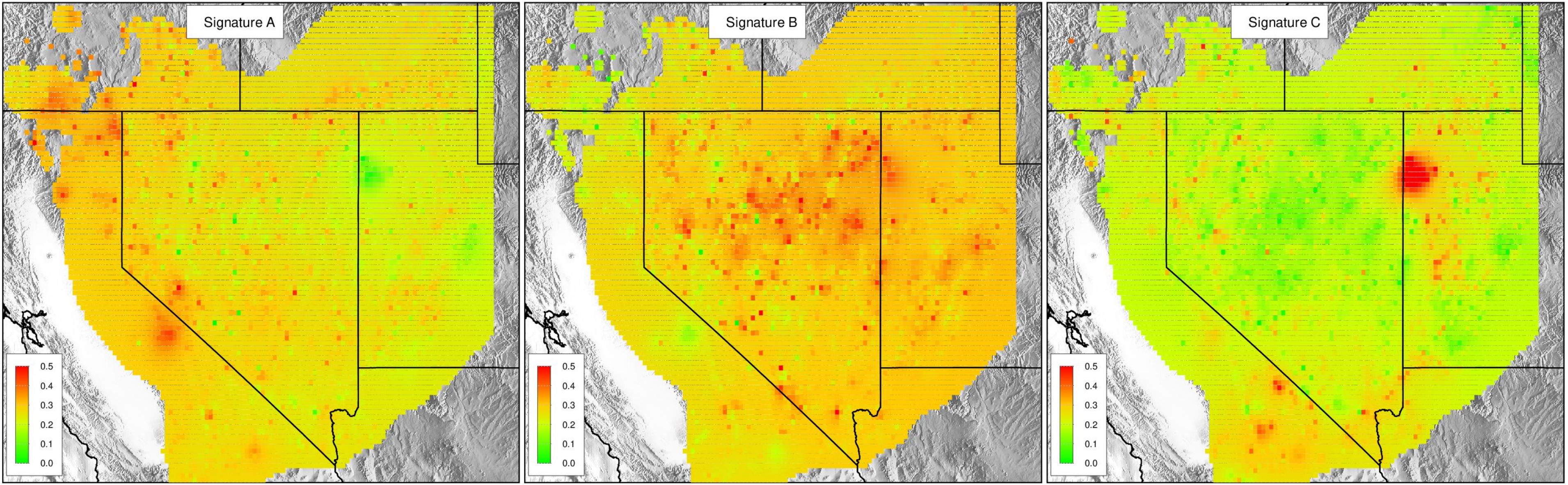
Attribute	Missing (%)
Groundwater temperature (℃)	2.6
Quartz geothermometer (℃)	39.1
Chalcedony geothermometer (℃)	39.1
pH	35.0
TDS (total dissolved solids) (PPM)	87.8
Al^{3+} (PPM)	90.5
B^{+} (PPM)	61.7
Ba^{2+} (PPM)	82.4
Be^{2+} (PPM)	88.5
Br^{-} (PPM)	86.4
Ca^{2+} (PPM)	33.6
Cl^{-} (PPM)	29.2
HCO_3^{-} (PPM)	76.1
K^{+} (PPM)	40.8
Li^{+} (PPM)	80.3
Mg^{2+} (PPM)	34.8
Na^{+} (PPM)	38.2
$\delta^{18}O$ (‰)	89.7



- 18 data attributes
- 14,341 locations
- Data gaps
- ML analyses are fast
- Will be redone to process updated/extra datasets (Phase 2)

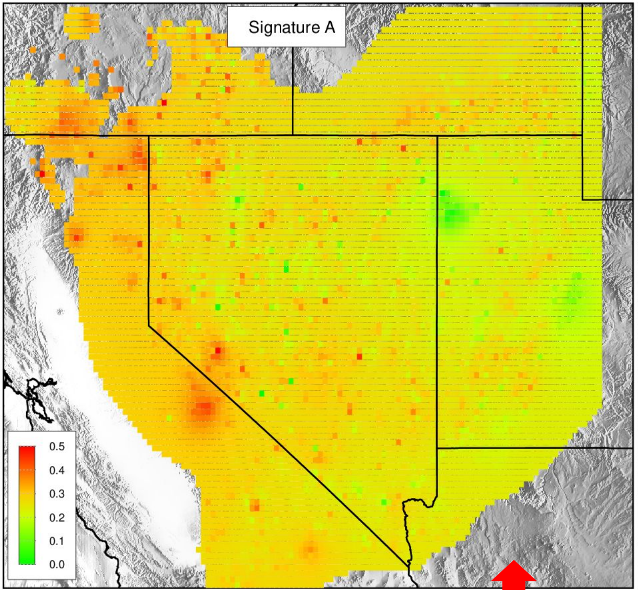
Great Basin: Geothermal Signatures

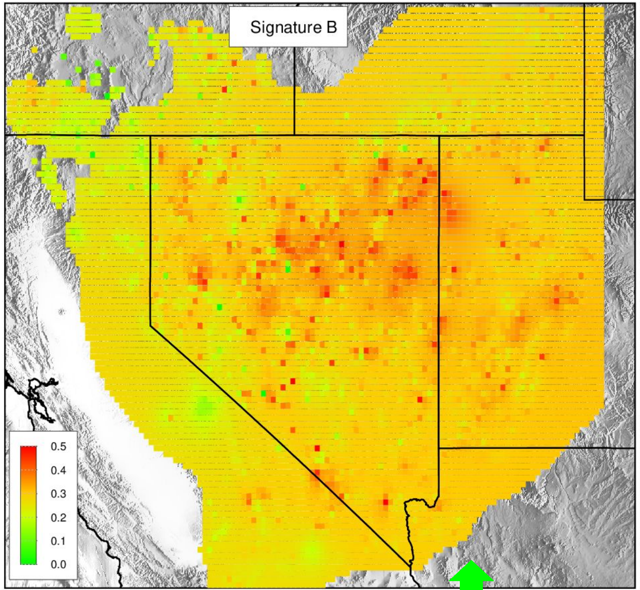
Signature A: Low T | Signature B: High T | Signature C: Mid T

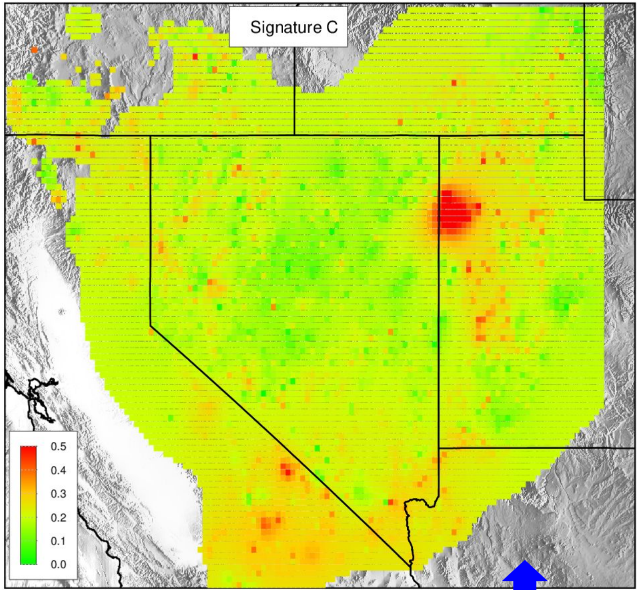


Great Basin: Geothermal Signatures

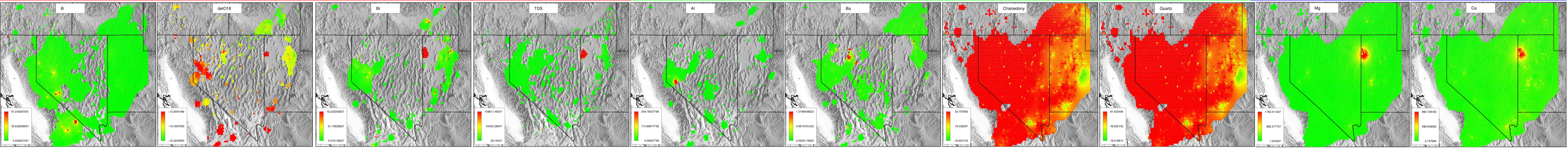
Signature A: Low T | **Signature B: High T** | **Signature C: Mid T**





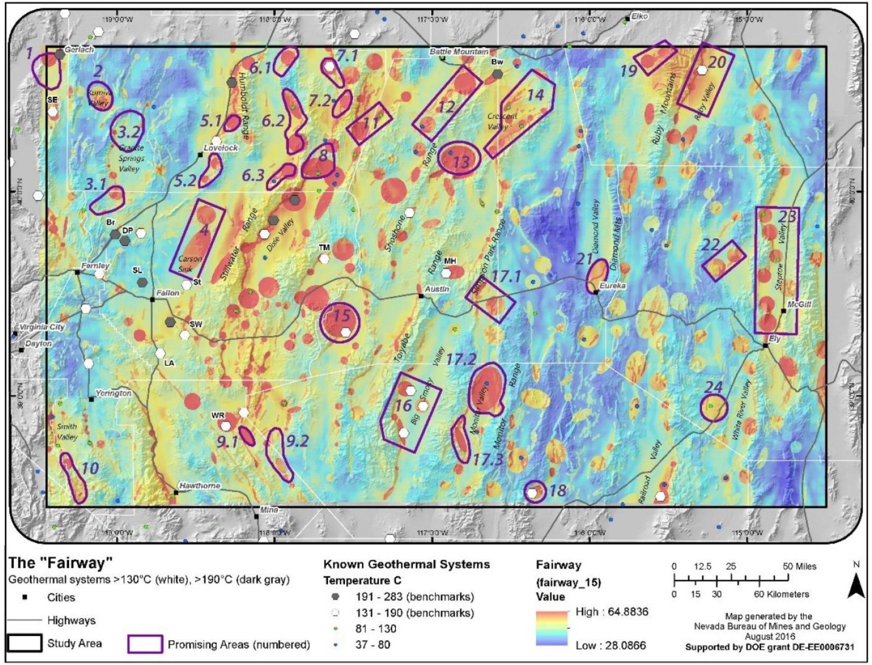


B | **δO18** | **Br** | **TDS** | **Al** | **Be** | **Chalcedony** | **Quartz** | **Mg** | **Ca**

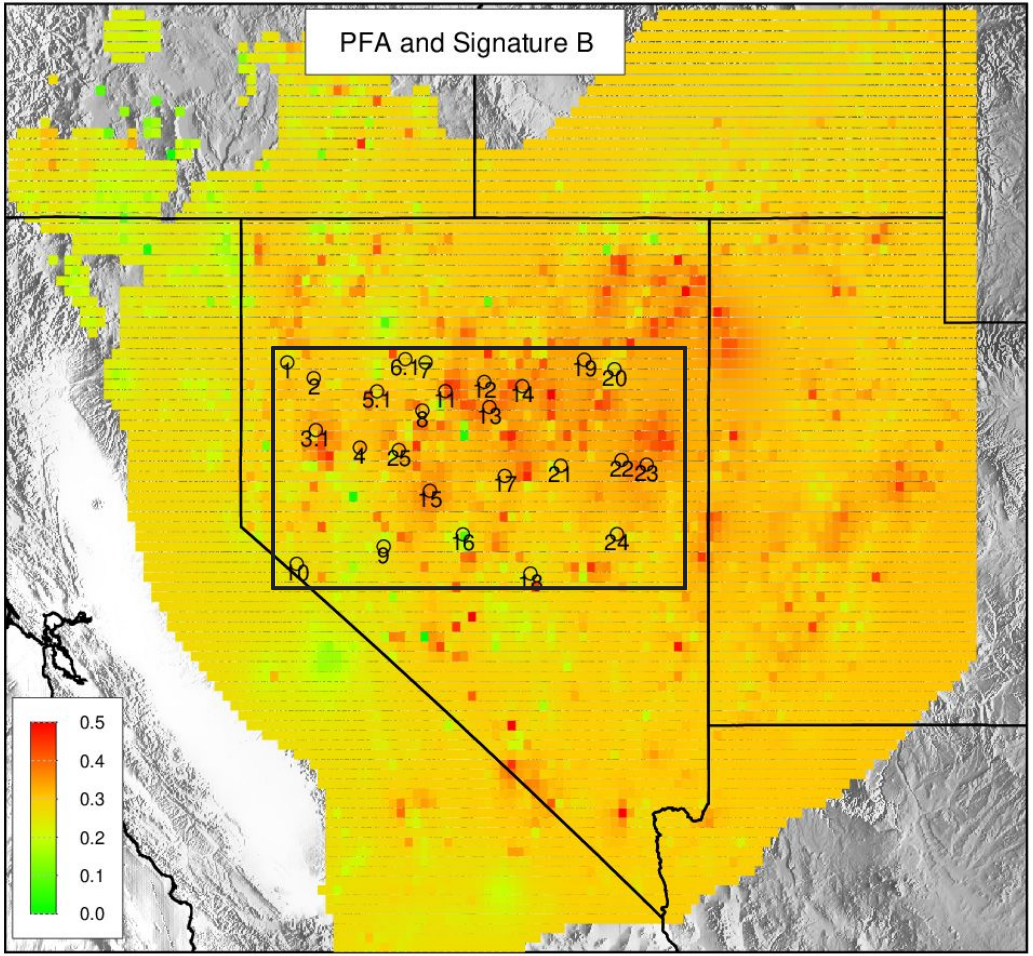


Comparison between PFA and GeoThermalCloud

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

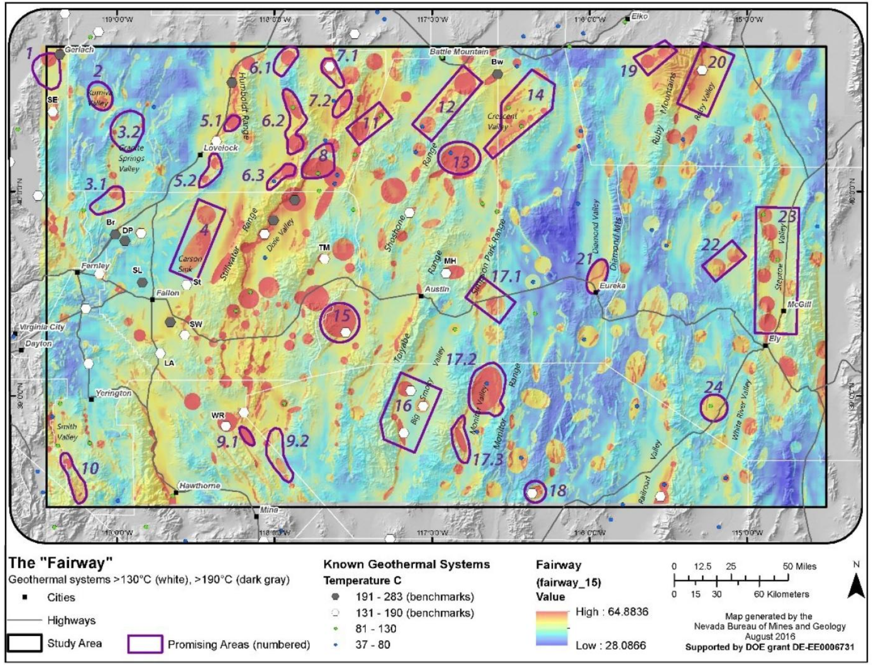


Source: https://gbcge.org/recent-projects/nevada-play-fairway-phase-2/

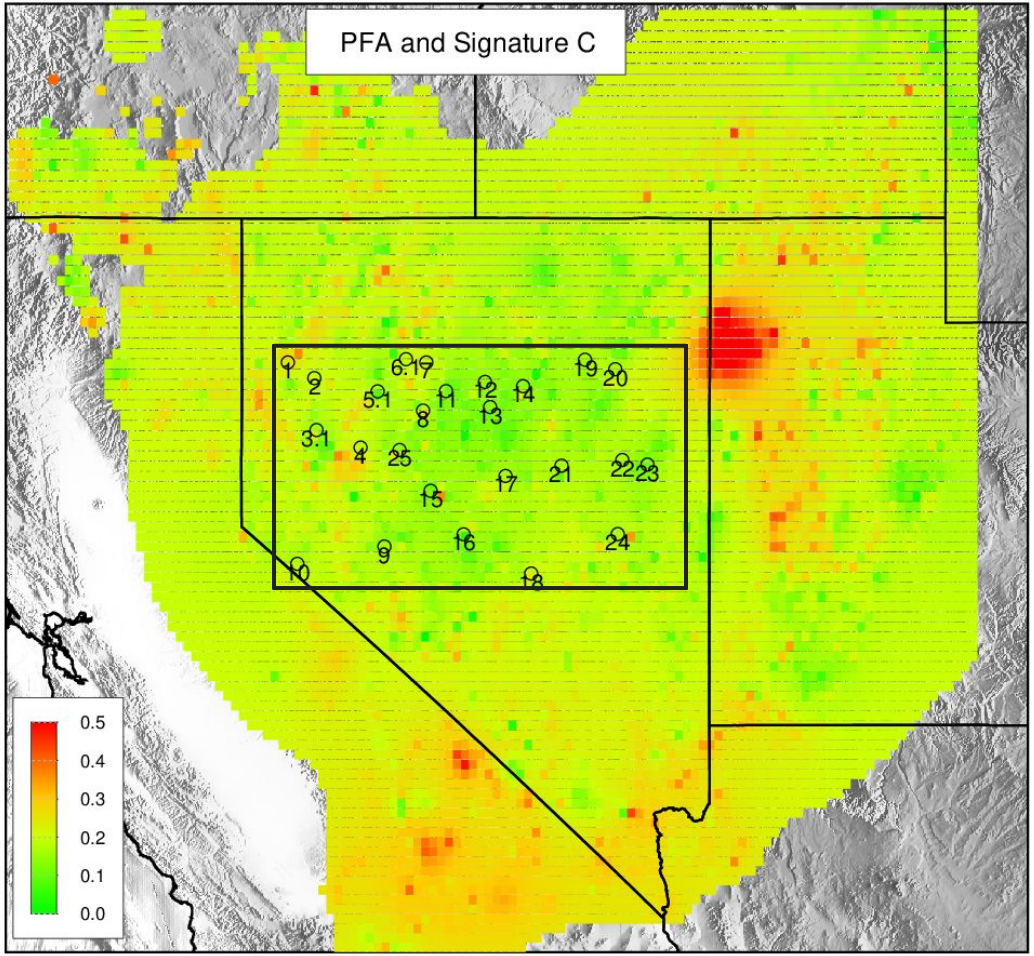


Comparison between PFA and GeoThermalCloud

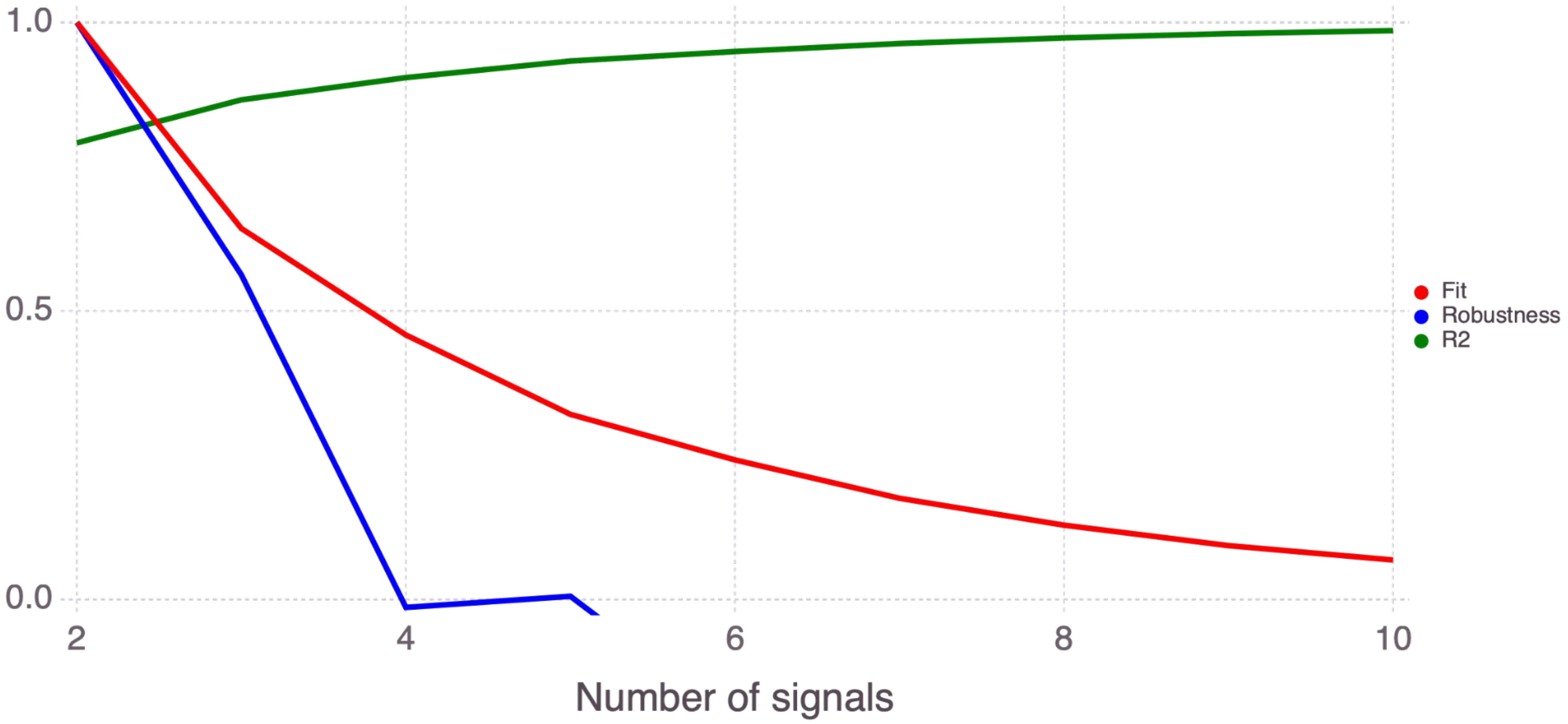
High-prospective areas estimated using PFA (Faulds et al. 2017)



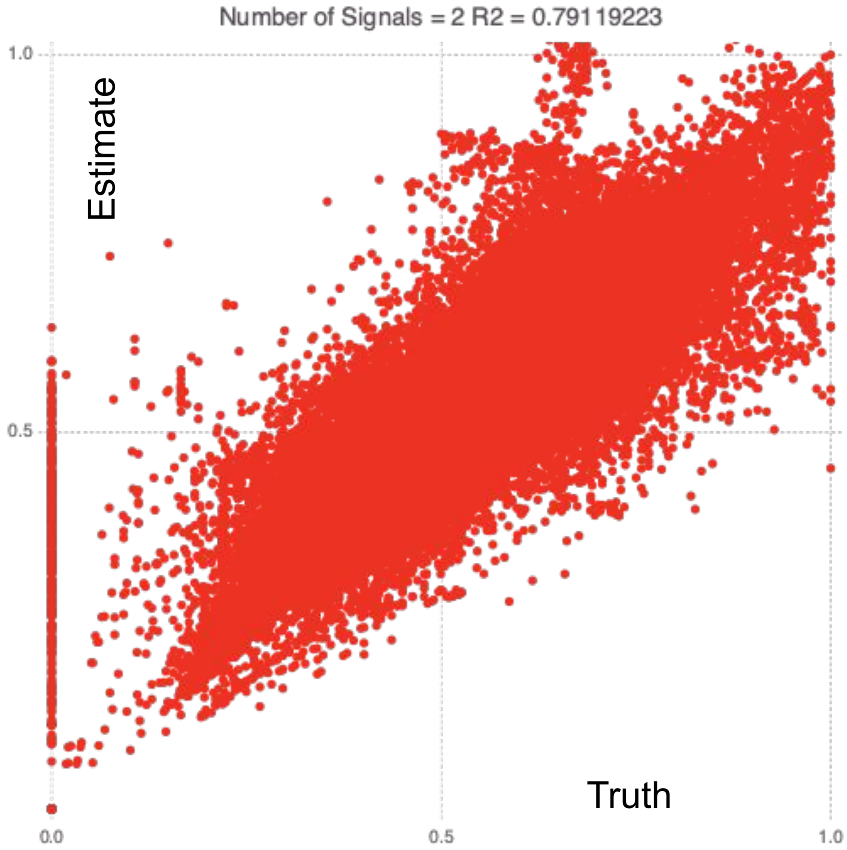
Source: https://gbcge.org/recent-projects/nevada-play-fairway-phase-2/

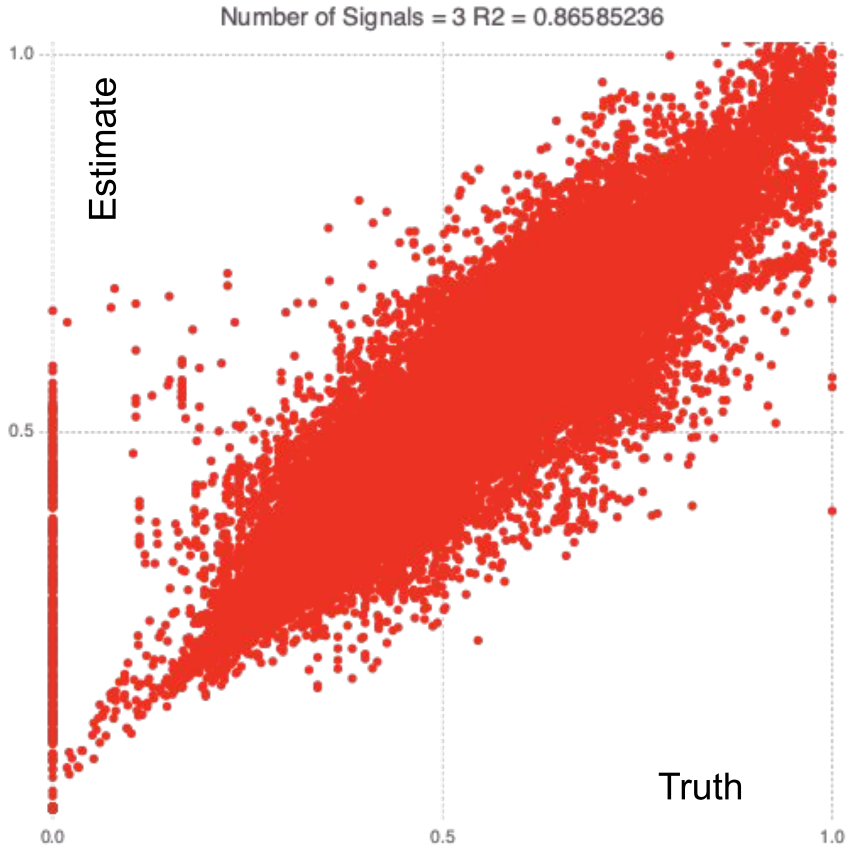


Great Basin: Signature Selection

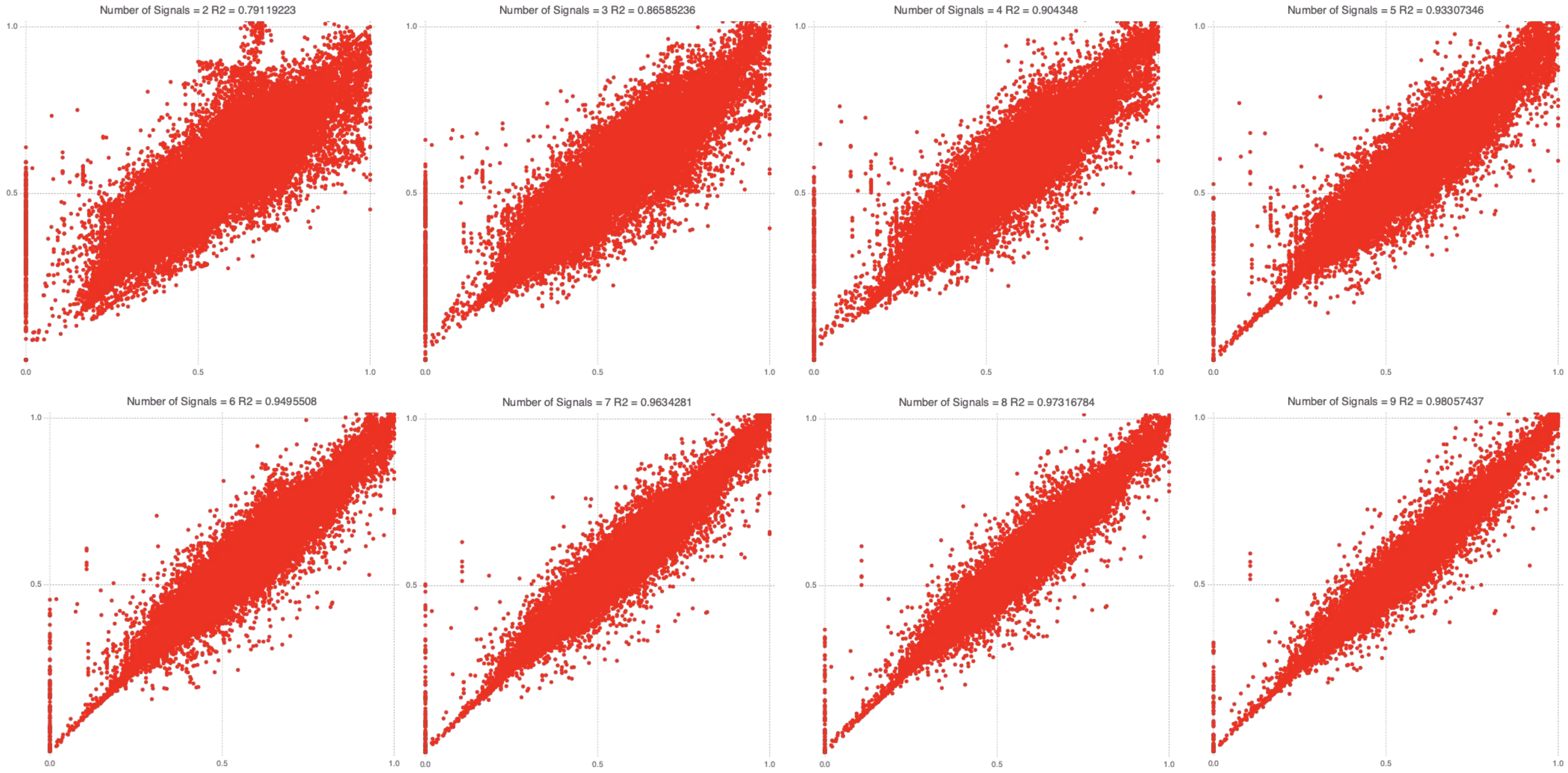


Great Basin: Signature Selection



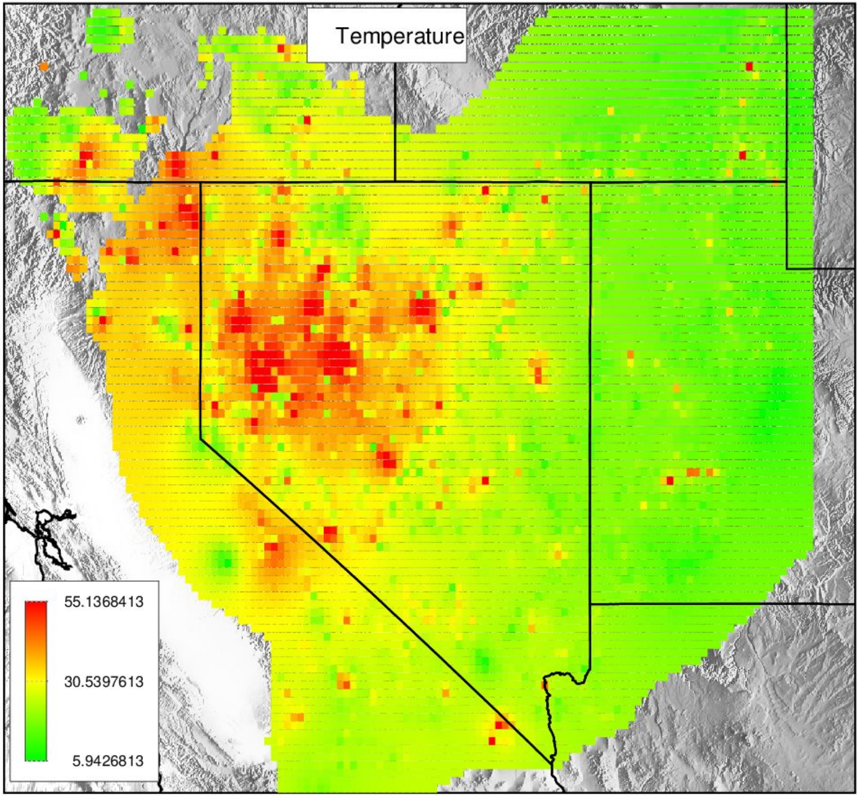


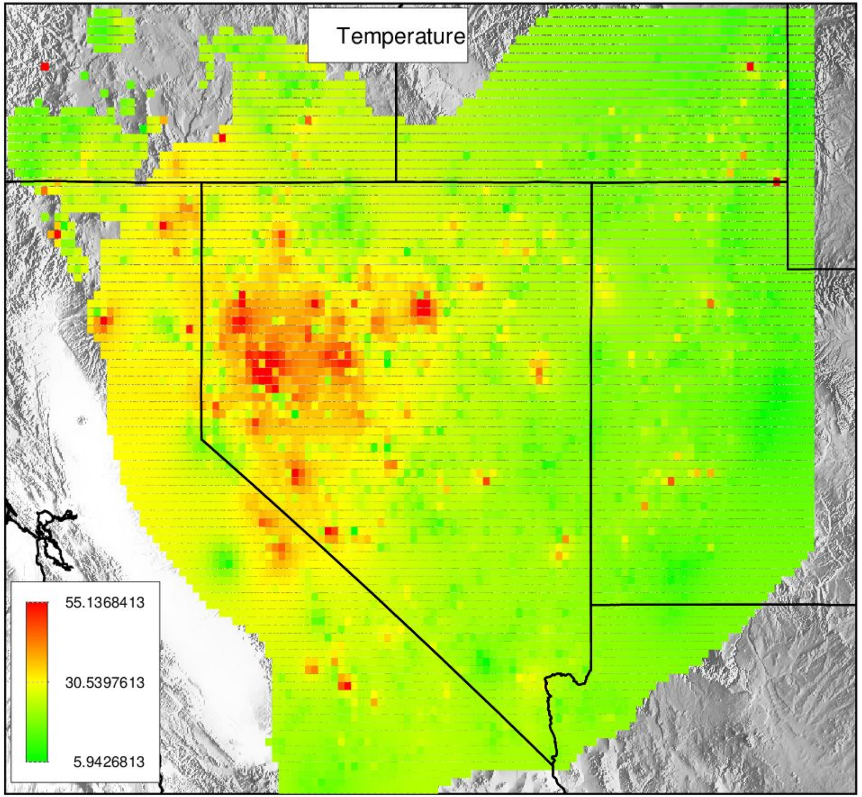
Great Basin: Signature Selection



Great Basin Geothermal Predictions

Data input | **Temperature** | **ML output**



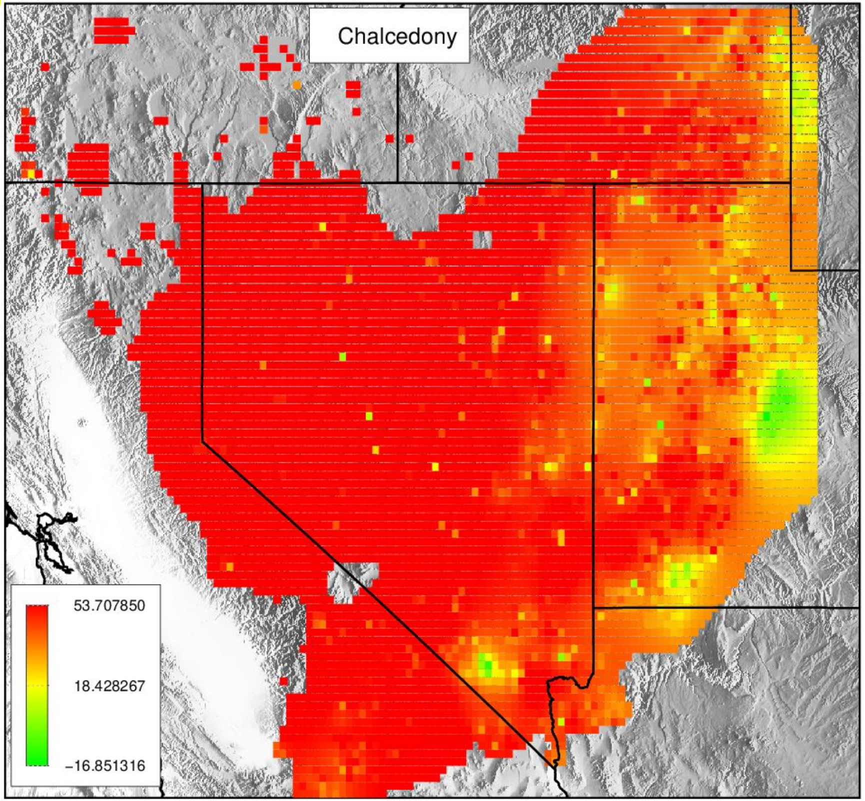


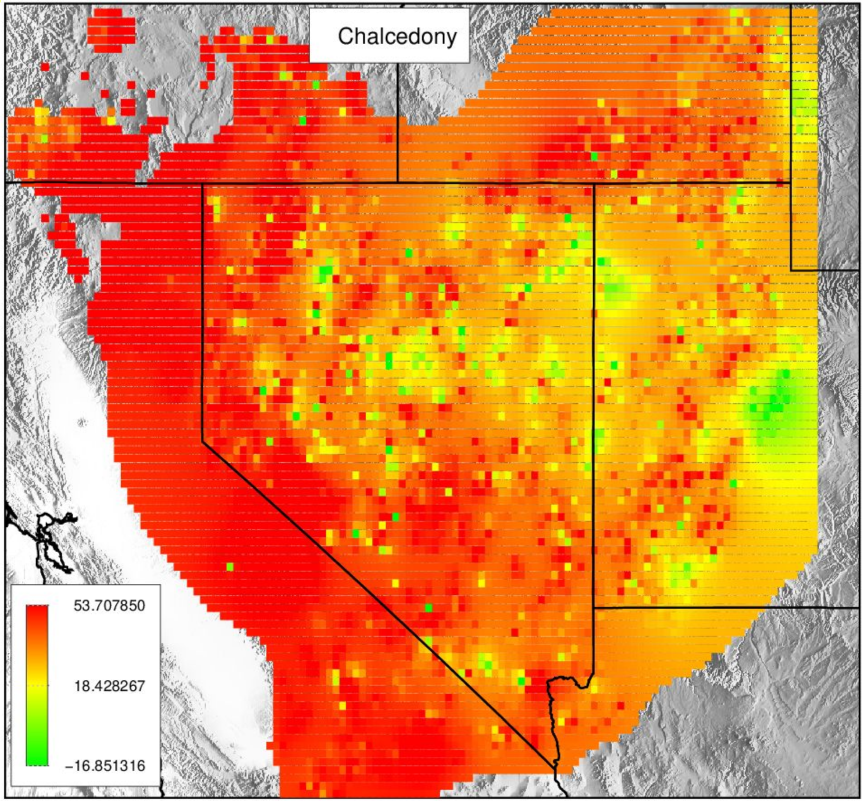
Great Basin Geothermal Predictions

Data input

Chalcedony

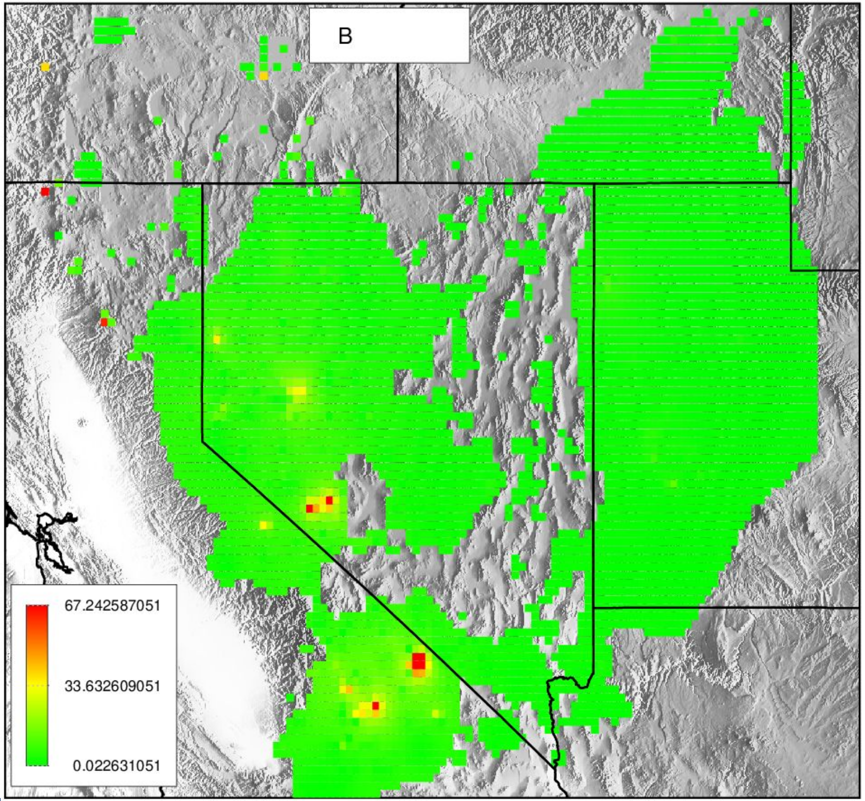
ML output

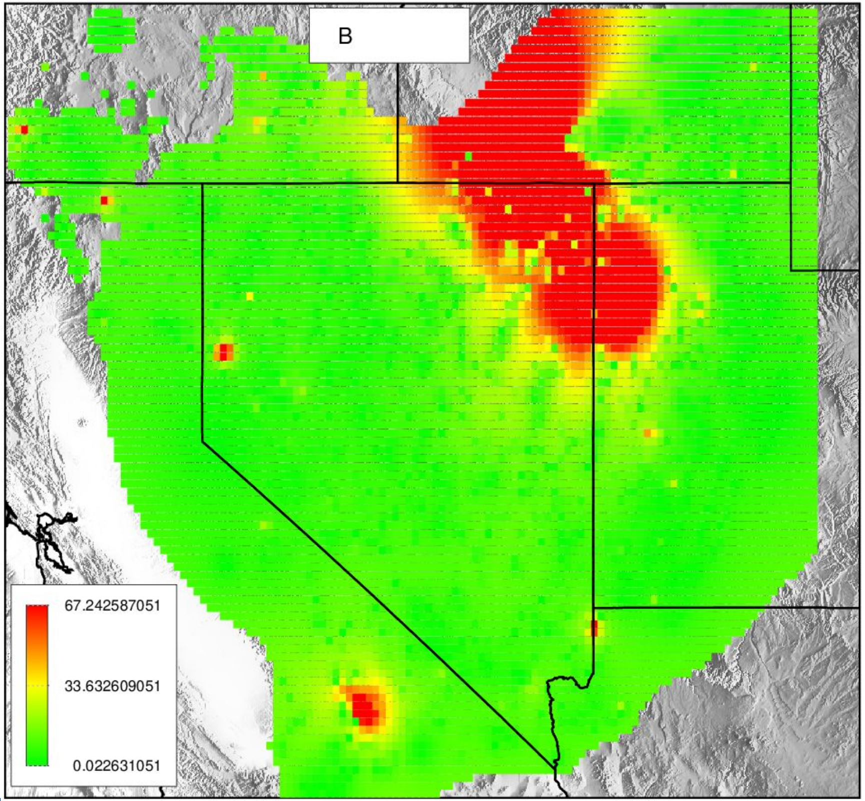




Great Basin Geothermal Predictions

Data input **B** **ML output**



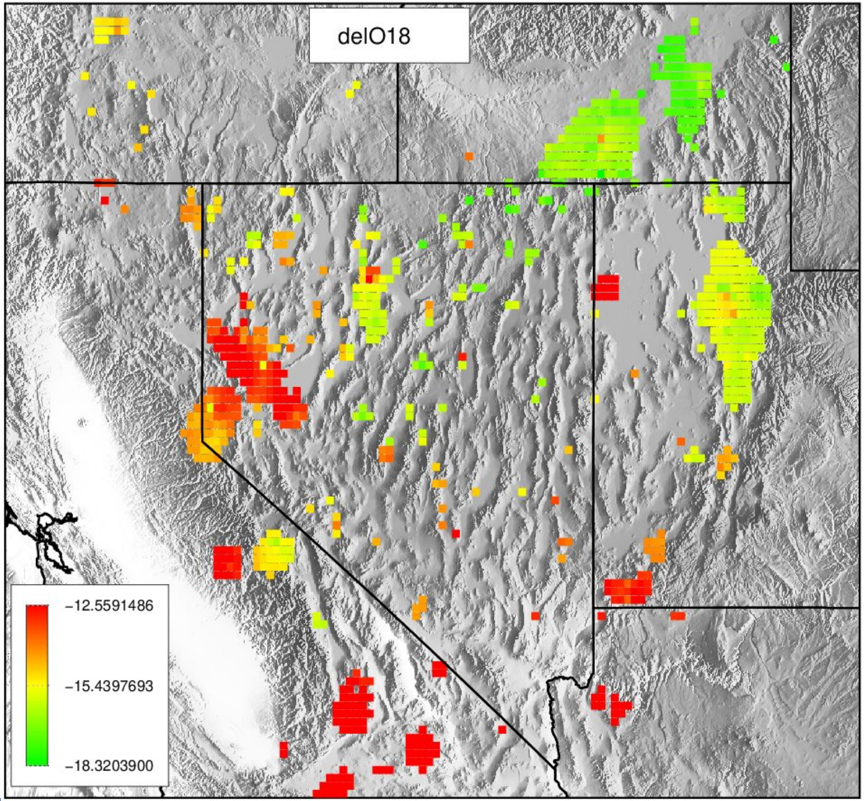


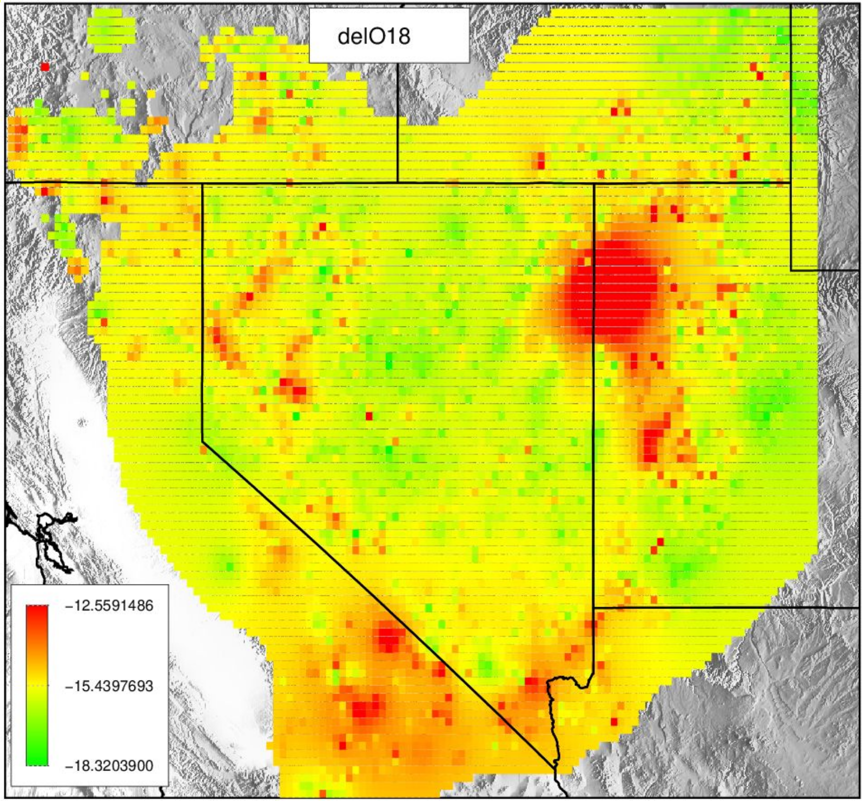
Great Basin Geothermal Predictions

Data input

delO18

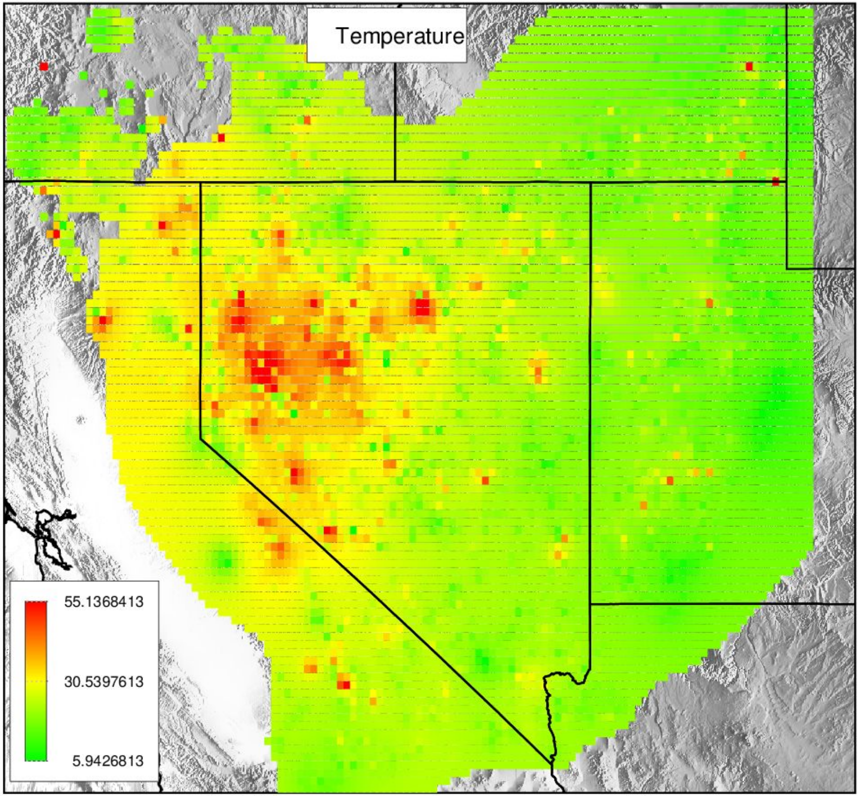
ML output

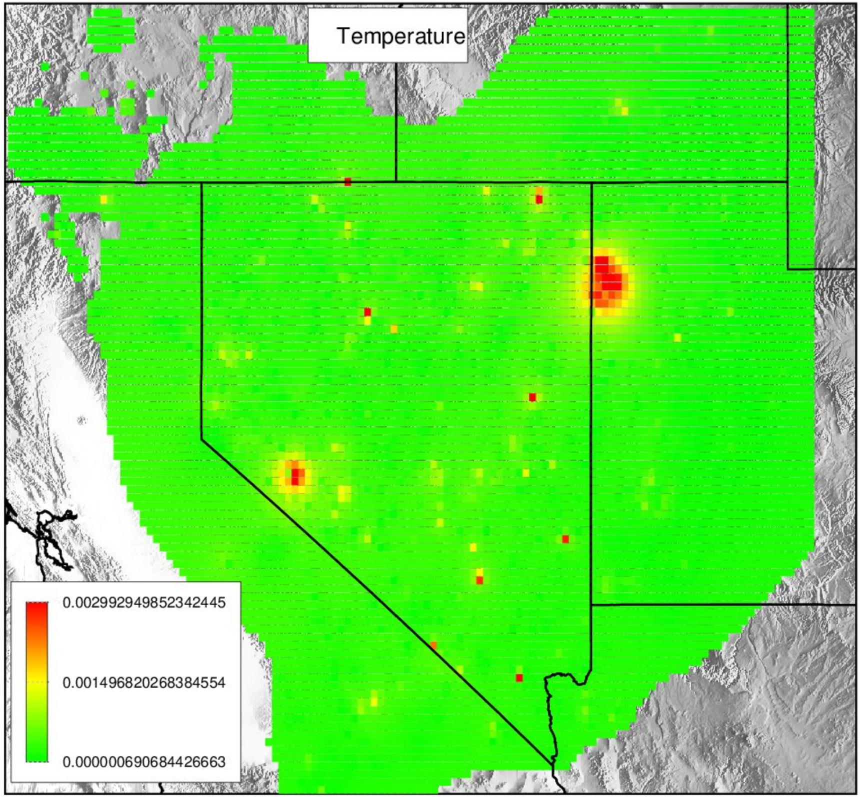




Great Basin Geothermal Predictions

ML output | **Temperature** | **ML uncertainty**

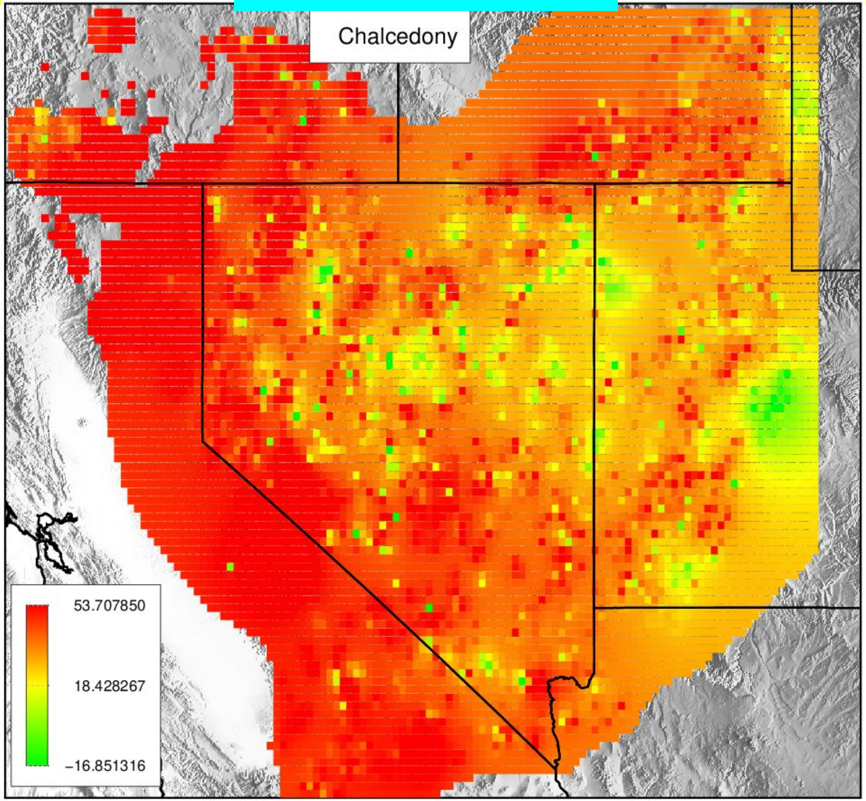




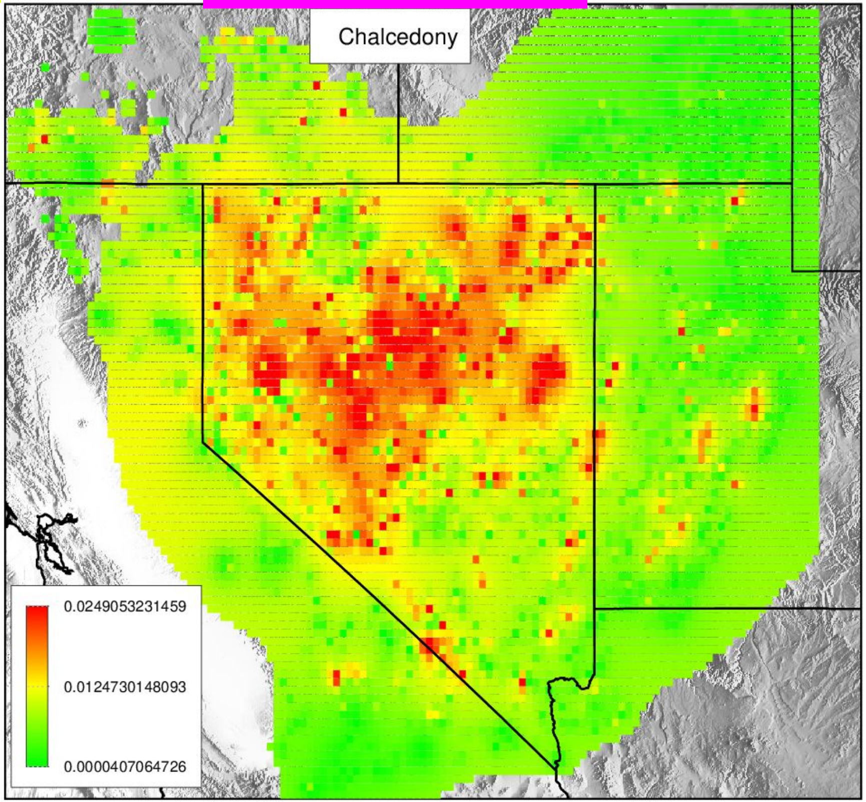
Great Basin Geothermal Predictions

Chalcedony

ML output



ML uncertainty

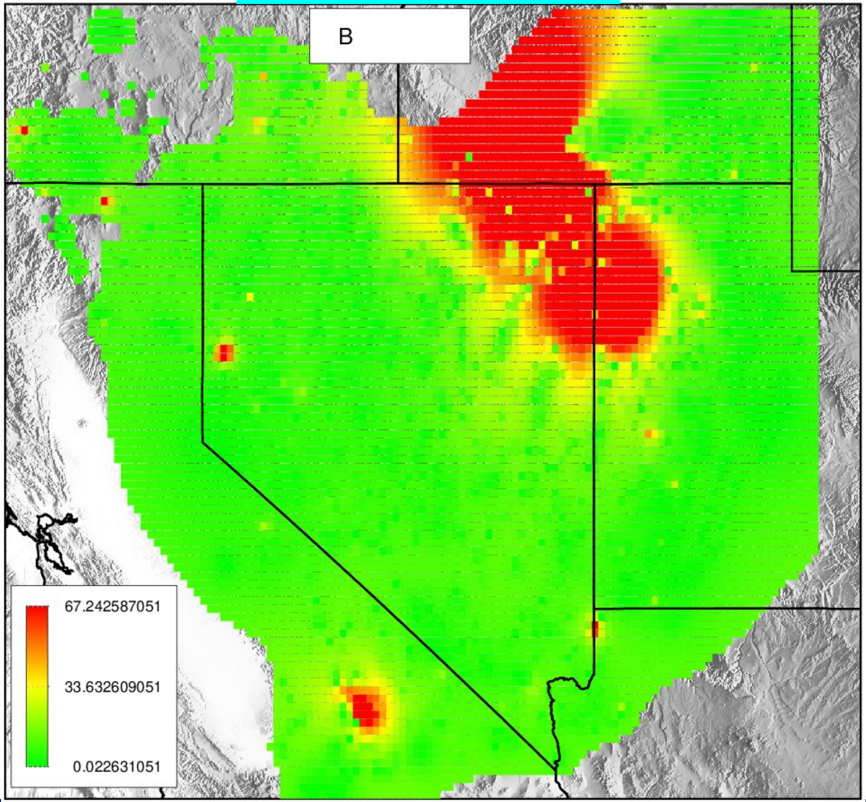


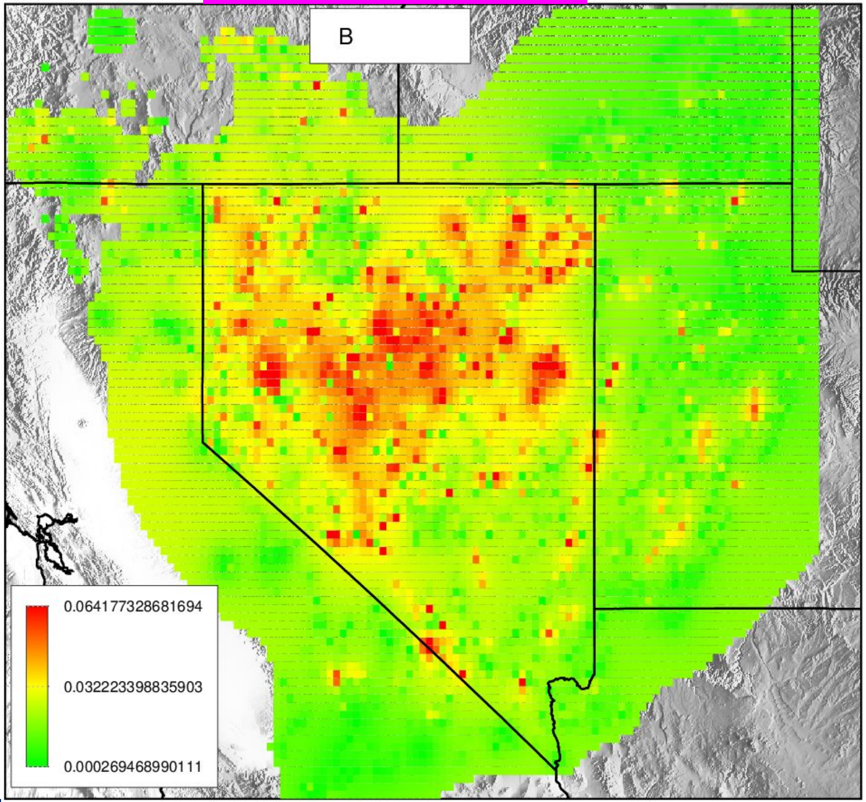
Great Basin Geothermal Predictions

ML output

B

ML uncertainty



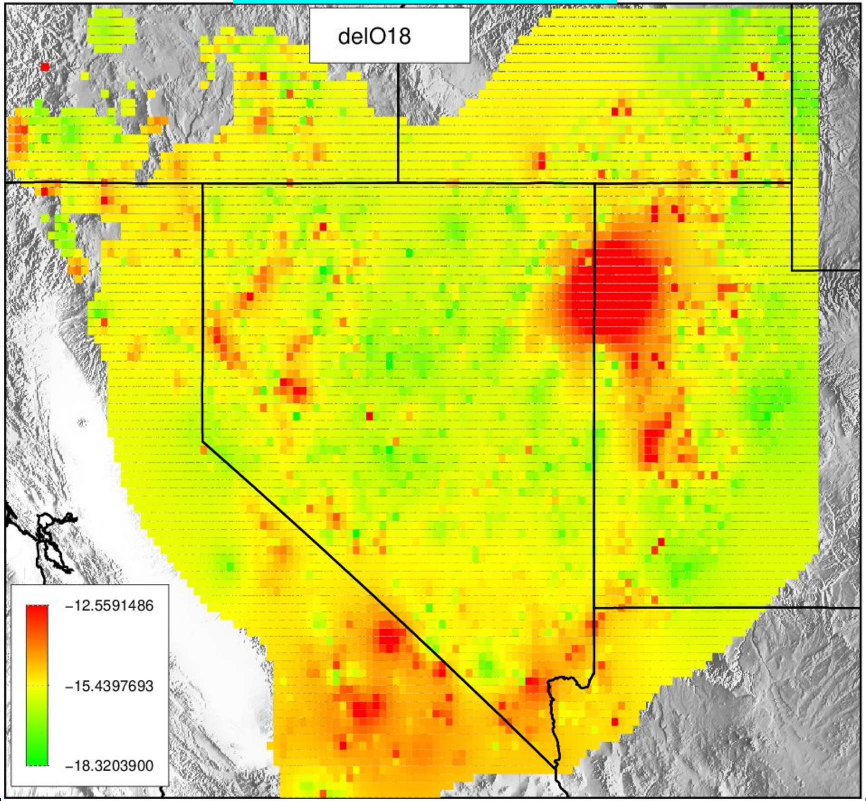


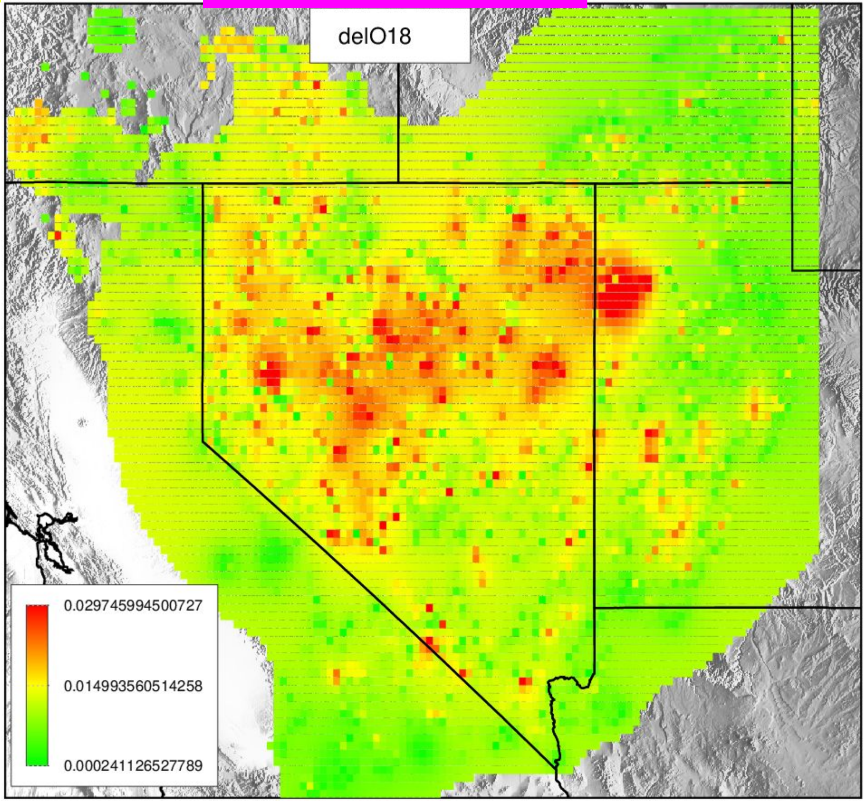
Great Basin Geothermal Predictions

ML output

delO18

ML uncertainty



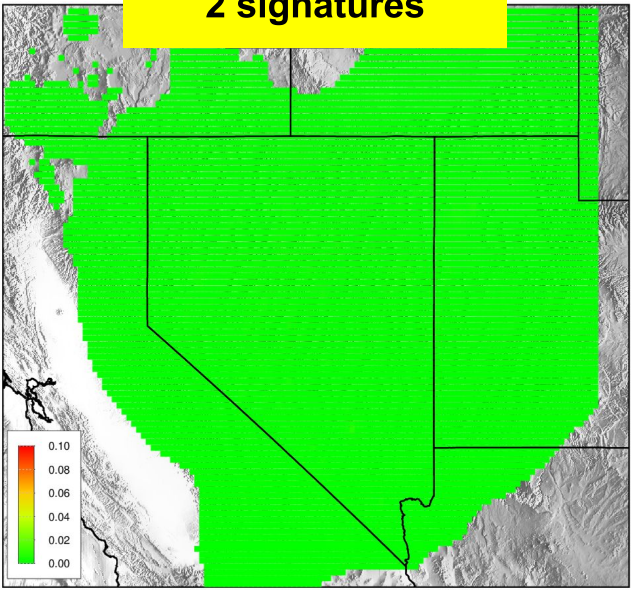


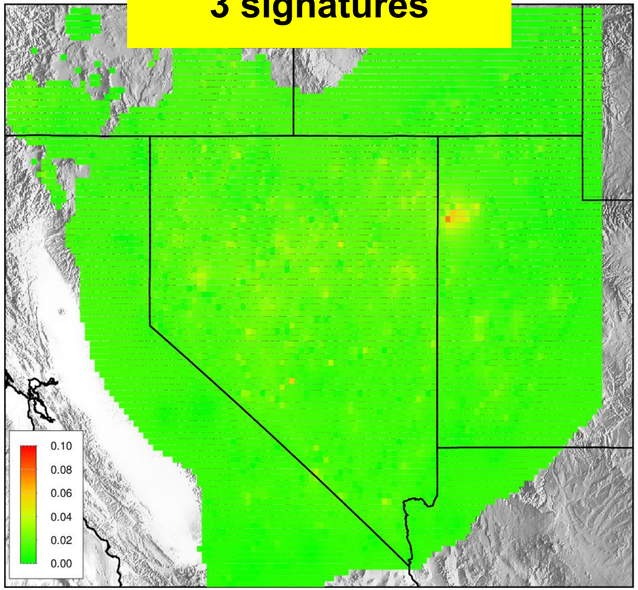
Great Basin Geothermal Predictions

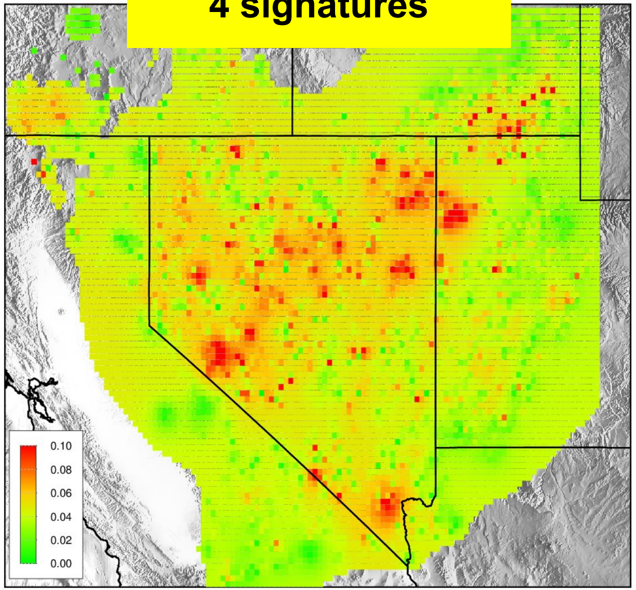
delO18

ML uncertainty

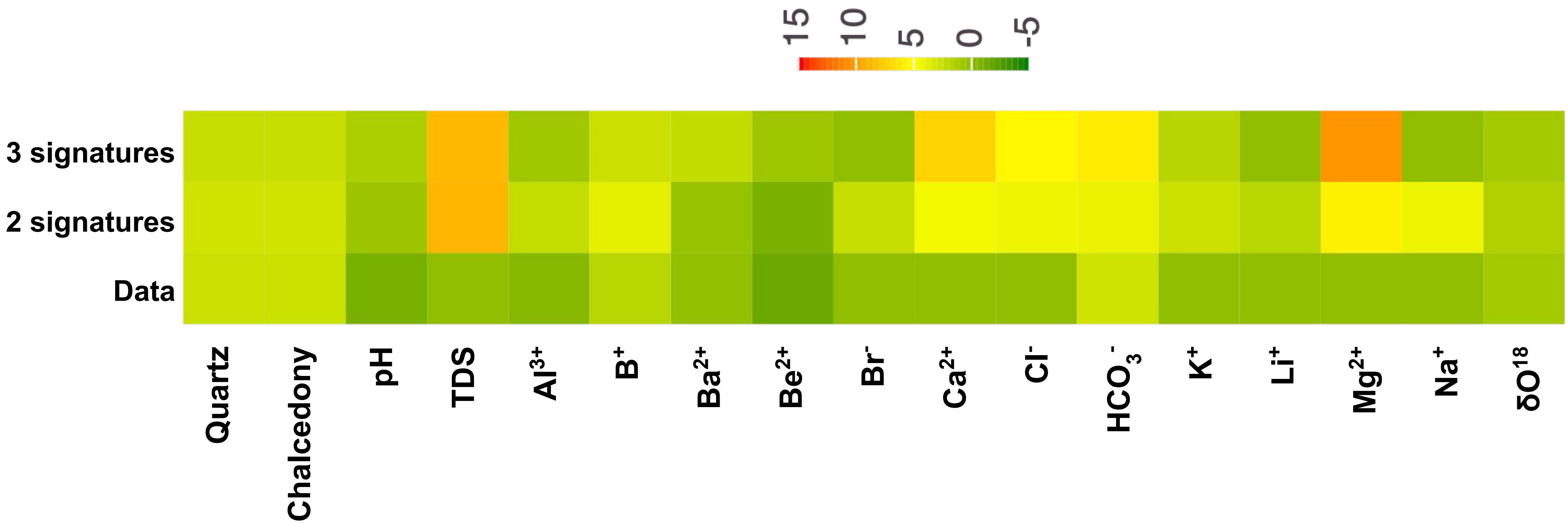
Underfitting		Overfitting
2 signatures	**3 signatures**	**4 signatures**



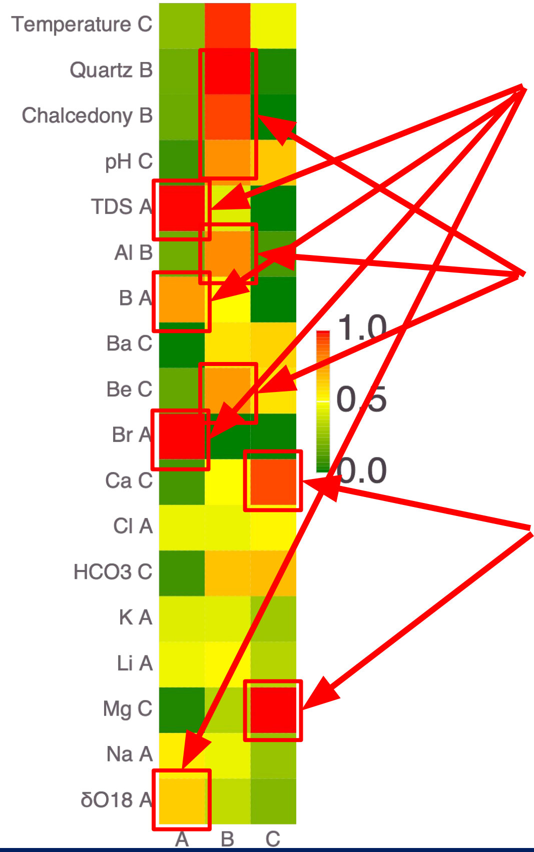




Great Basin: Covariances to temperature

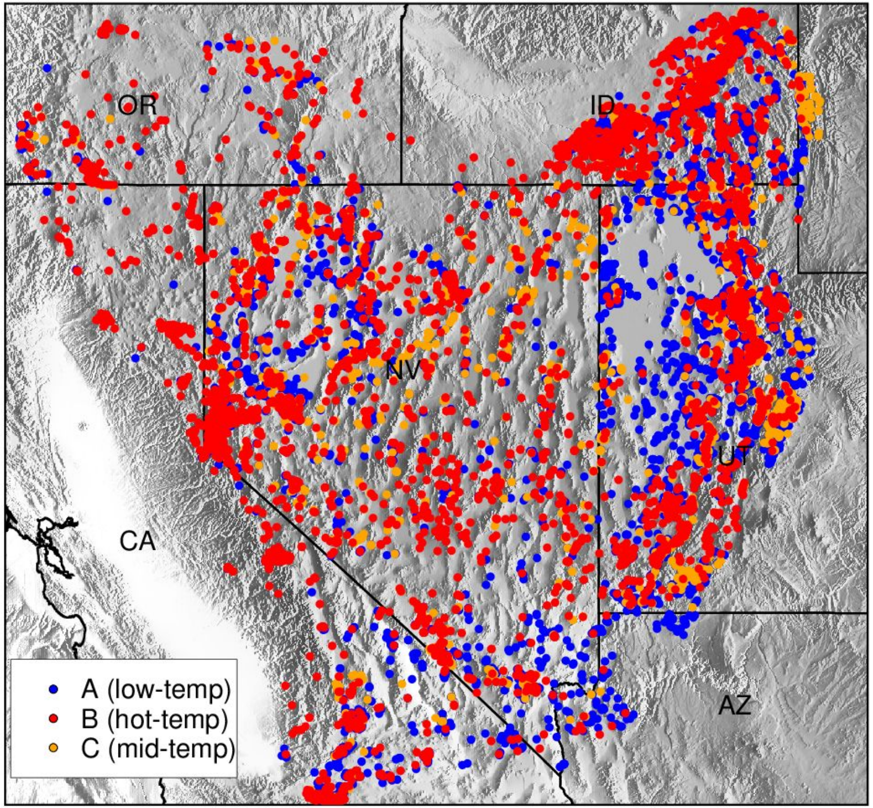


Great Basin: ML extracted Geothermal Signatures



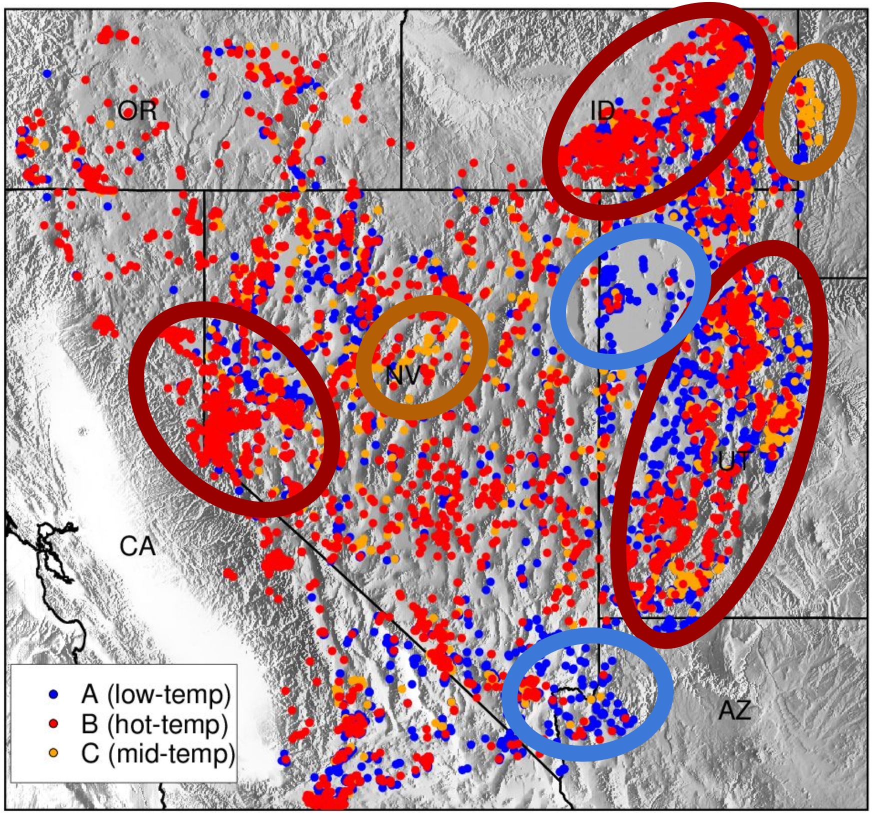
- **A:** Low-temperature resource
  - TDS, Br, B, δO18
- **B:** High-temperature resource
  - Quartz and Chalcedony geothermometers, pH, Al, Be
- **C:** Medium-temperature resource
  - Mg, Ca

Great Basin: Geothermal Signatures



- Our ML analyses also estimate the spatial distribution of hidden geothermal signatures
- **A**: Low-temperature resource
  - TDS, B, Br, δO18
- **B**: High-temperature resource
  - Al, Be, Quartz and Chalcedony geothermometers
- **C**: Medium-temperature resource
  - Mg, Ca

Great Basin: Geothermal Signatures



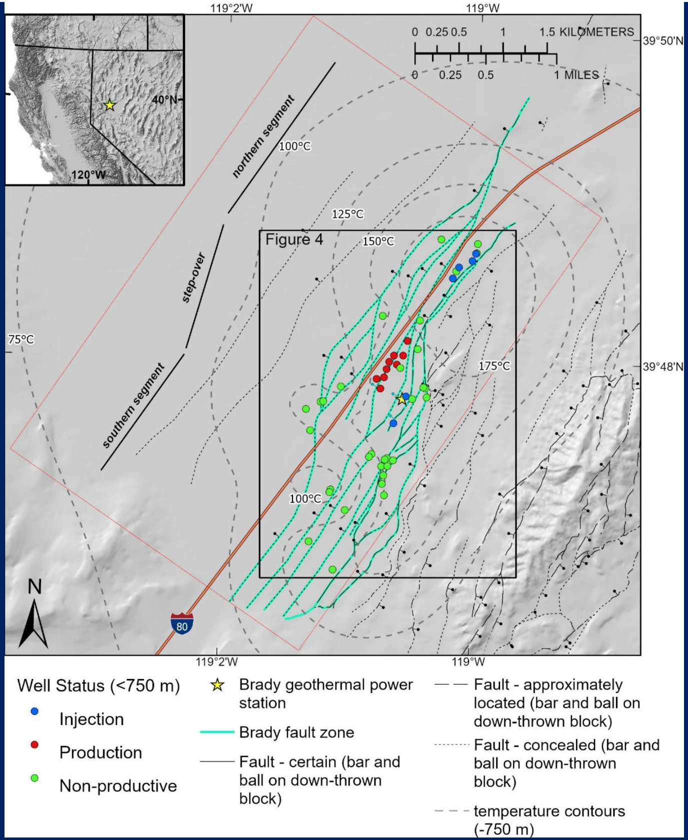
- Our ML analyses also estimate the spatial distribution of hidden geothermal signatures
- **A**: Low-temperature resource
  - TDS, B, Br, δO18
- **B**: High-temperature resource
  - Al, Be, Quartz and Chalcedony geothermometers
- **C**: Medium-temperature resource
  - Mg, Ca

Great Basin ML predictive uncertainties

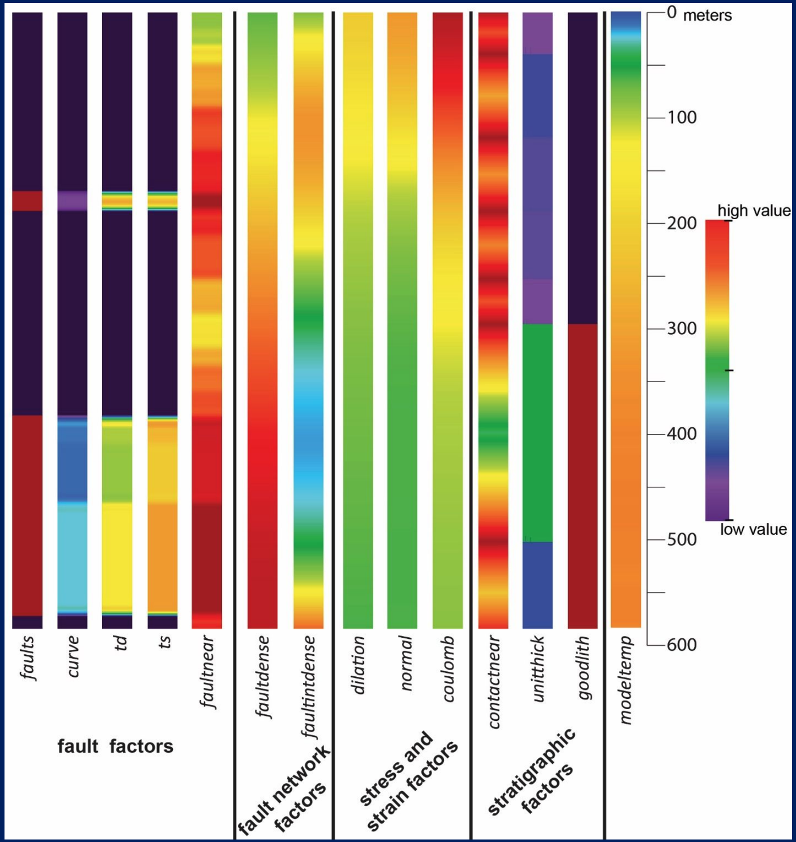
- Developed ML model is also applied to predict temperature based on all other attributes
- Artificial noise (mimicking measurement errors) at different levels is added
- Accuracy of the blind temperature predictions are evaluated (r^2)

Training percent	Noise level [%]			
	100%	50%	20%	10%
90%	0.675	0.823	0.939	0.976
80%	0.616	0.769	0.919	0.951
50%	0.574	0.749	0.870	0.917
20%	0.565	0.714	0.838	0.887
10%	0.441	0.623	0.755	0.876

Brady site, Nevada



Study area: 47 wells / 14 attributes

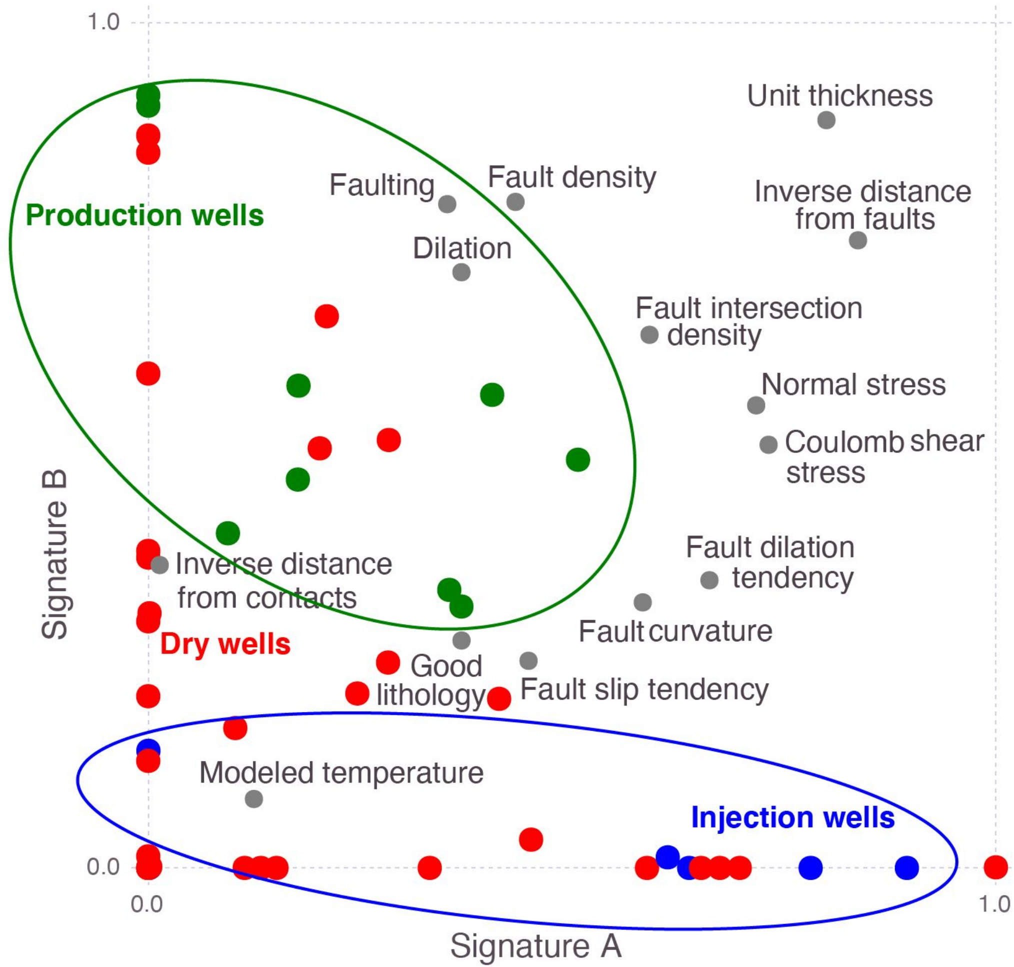


Data attributes along one of the wells

Brady site

Nevada

Extracted hidden geothermal signatures B & A separate production and injection wells



(Paper submitted in collaboration with USGS)

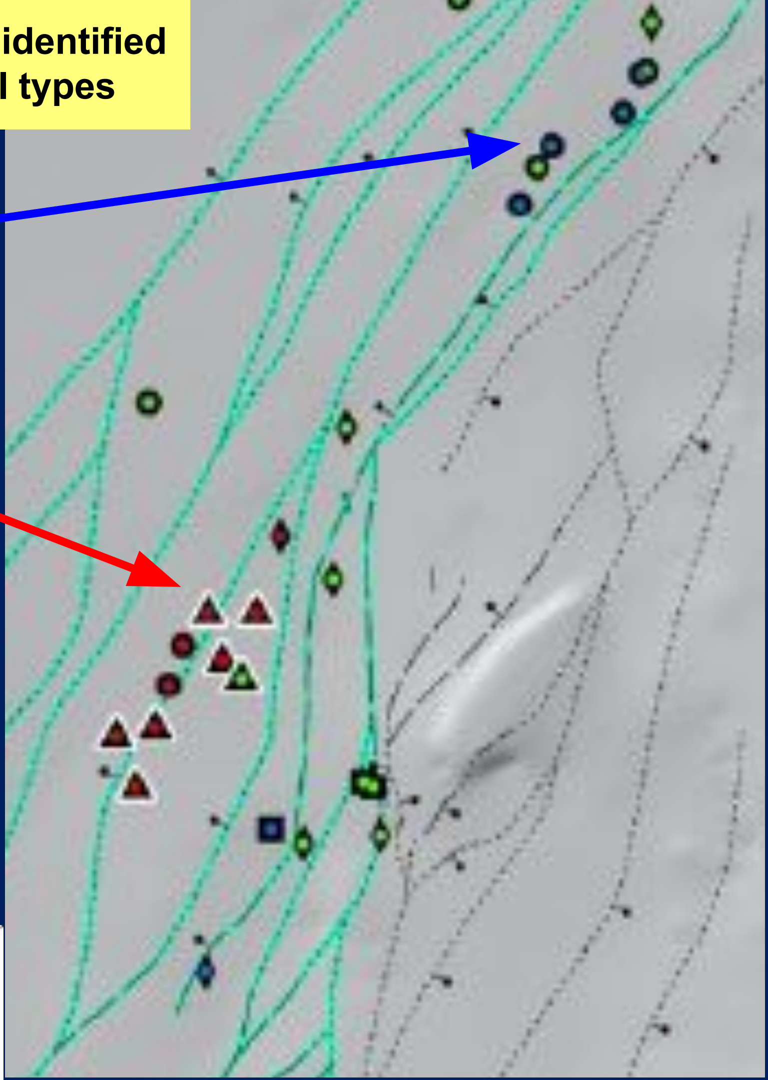
Brady site, Nevada

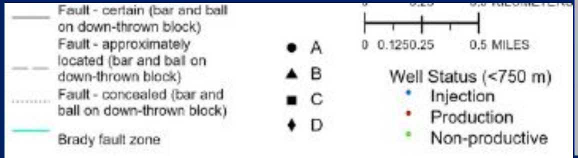
ML identified well types

Signature A: injection wells

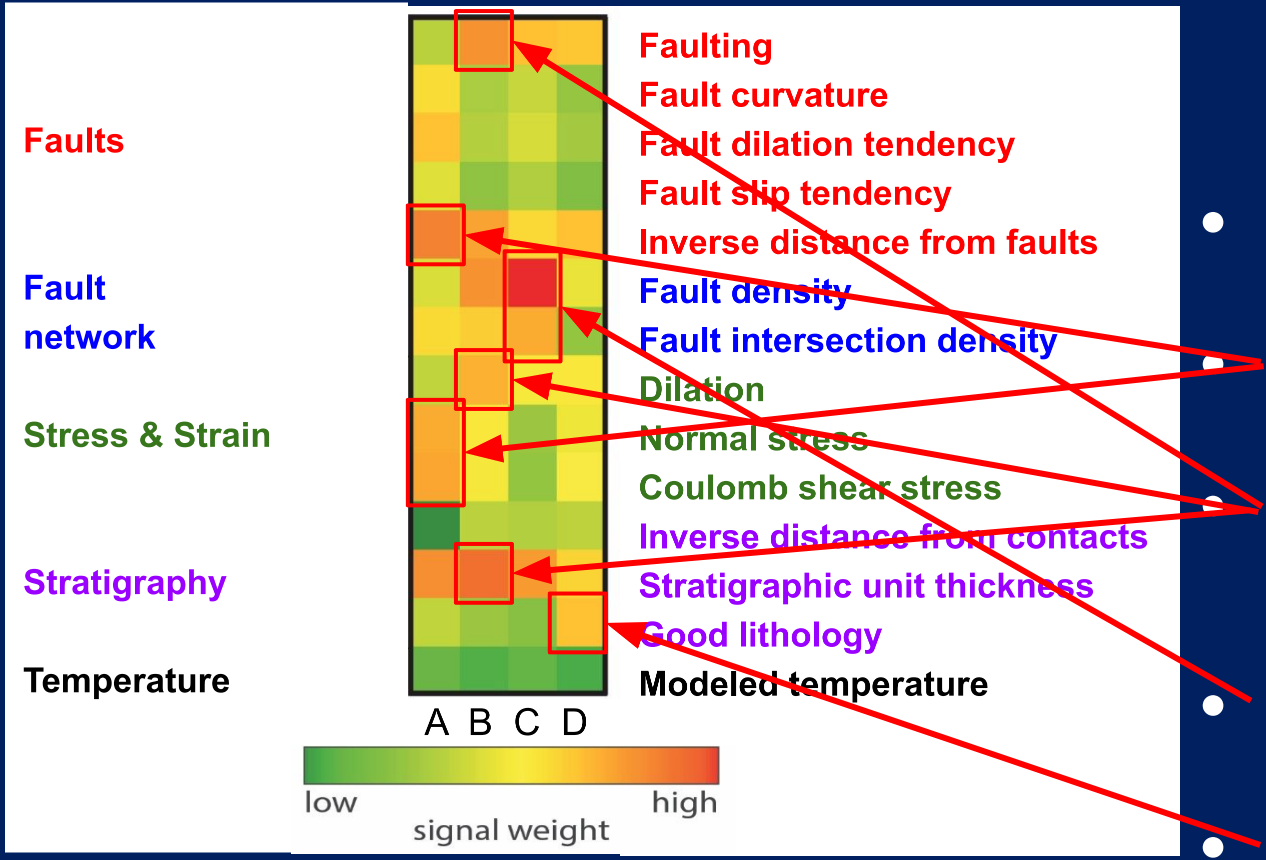
Signature B: production wells

Signatures C and D: dry wells



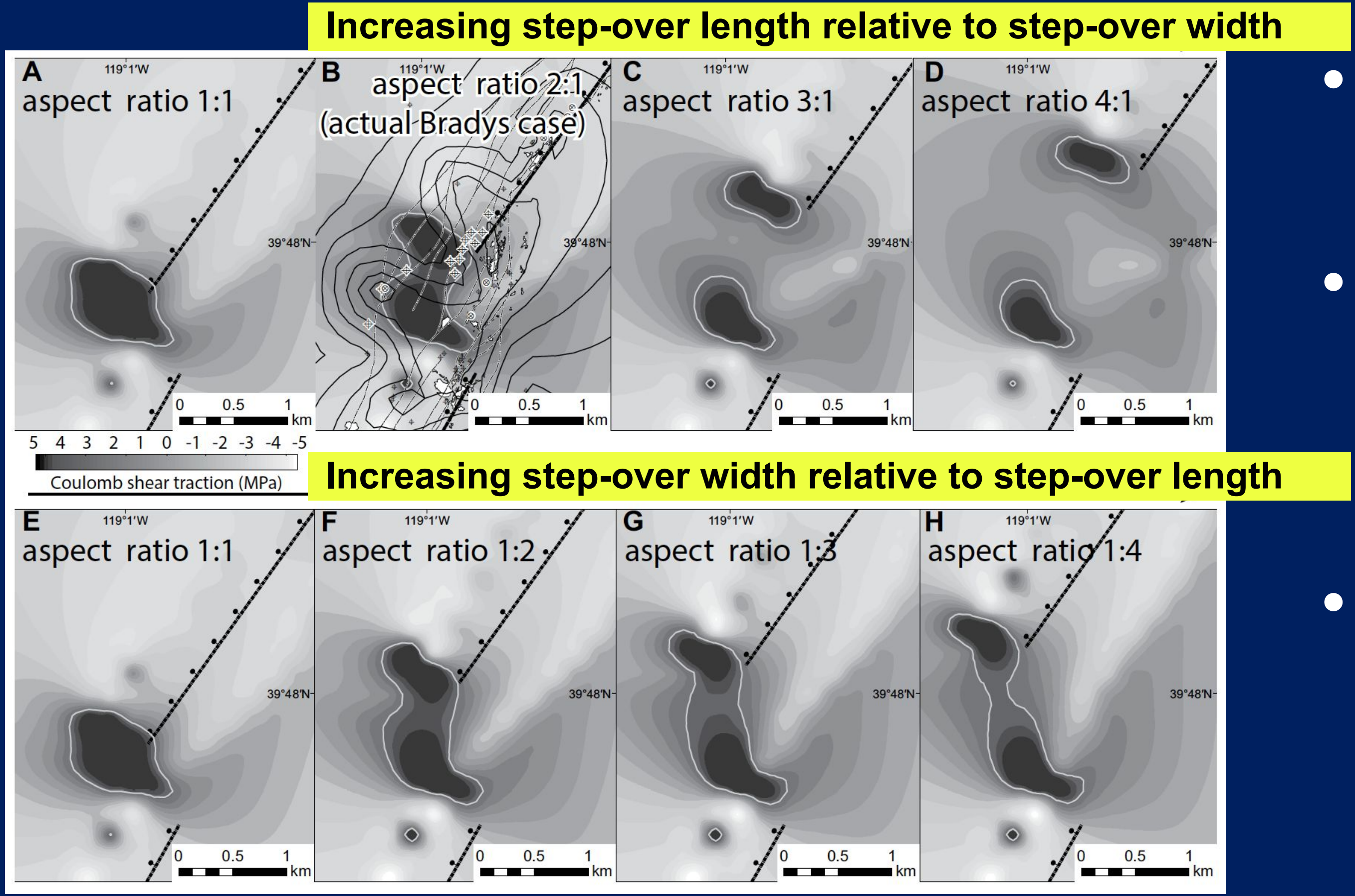


Brady site, Nevada



- **Analyzed dataset is a 3D tensor:**
  47 wells
  14 attributes
  750 vertical depths (1 m)
- **4 geothermal signatures extracted**
- **A: Stresses, Inverse distance from faults**
- **B: Stratigraphy unit thickness, Faulting, Dilation**
- **C: Fault density, fault intersection density**
- **D: Good lithology**

Brady site: State of Stress Impacts



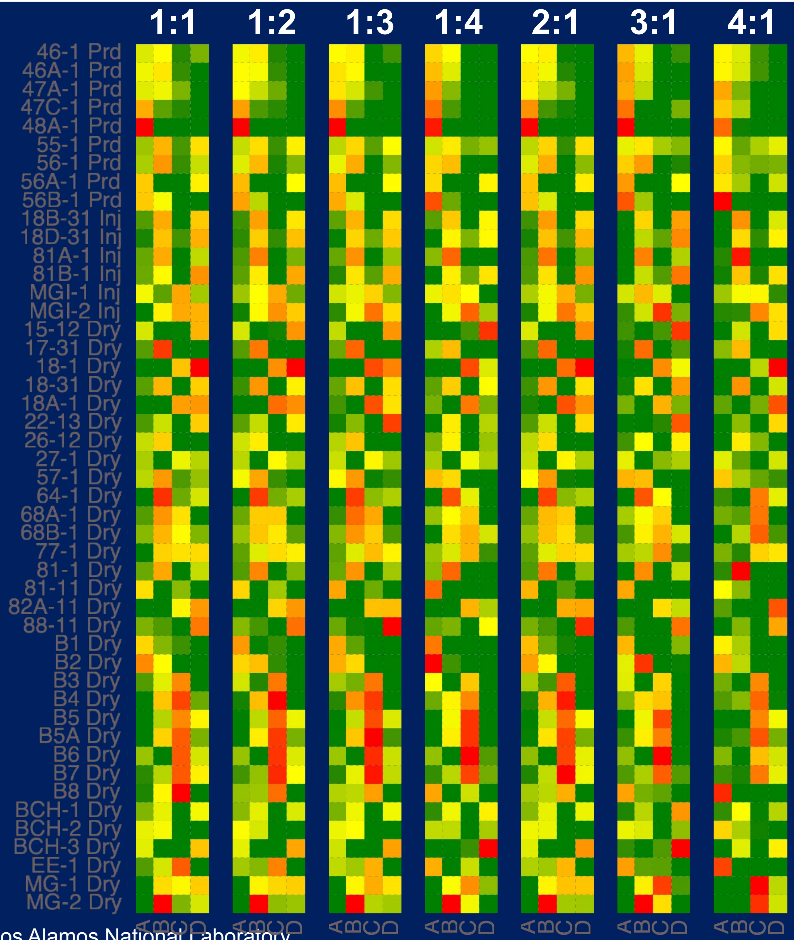
- Coulomb shear traction estimated at 1000 m depth
- Dark colors represent high Coulomb shear traction on optimally oriented normal faults as a result of slip
- Aspect ratio of 2:1 most probably characterizes the state of stress at the Brady case

Brady site: State of Stress Impacts

- Stress ratios at the site are unknown
- A series of stress ratios are modeled and after that analyzed using our ML methods
- Based on reconstruction errors and attribute categorizations, the ML blindly identified the 2:1 stress ratio as the most probable to represent site conditions
- In fact, this is the most probable stress ratio at the site (2:1) based on previous studies

Relative Improvement in Reconstruction Error						
1:1	1:2	1:3	1:4	2:1	3:1	4:1
156	155	237	243	262	0	140

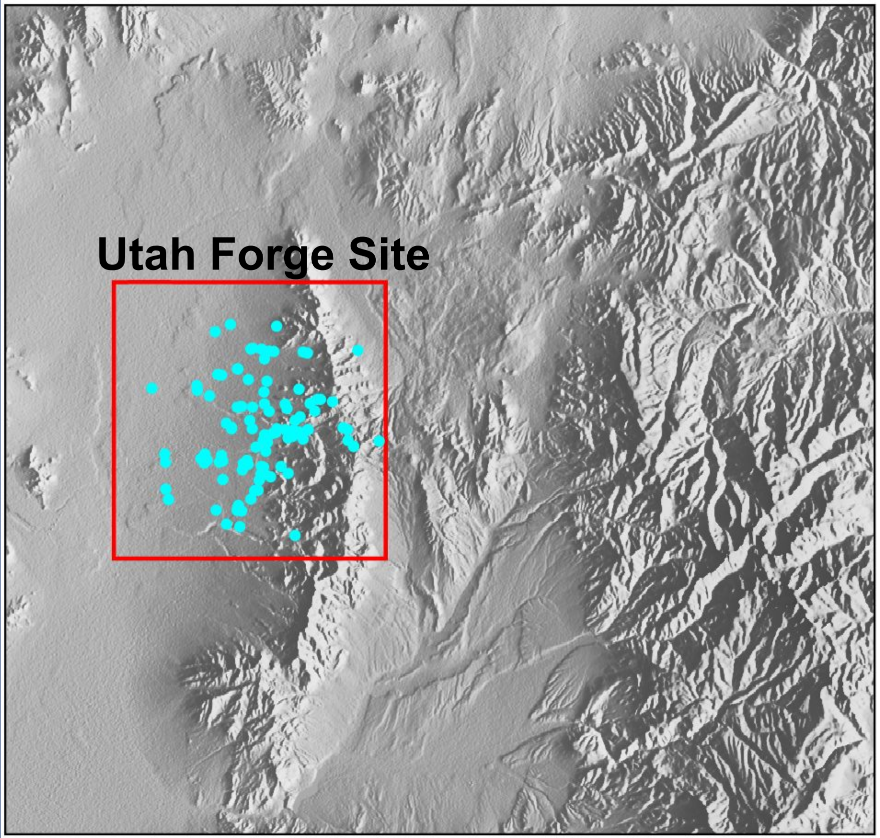
Brady site: State of Stress Impacts



- Stress ratios at the site are unknown
- A series of stress ratios are modeled and after that analyzed using our ML methods
- Based on reconstruction errors and attribute categorizations, the ML blindly identified the 2:1 stress ratio as the most probable to represent site conditions
- In fact, this is the most probable stress ratio at the site (2:1) based on previous studies

Relative Improvement in Reconstruction Error						
1:1	1:2	1:3	1:4	2:1	3:1	4:1
156	155	237	243	262	0	140

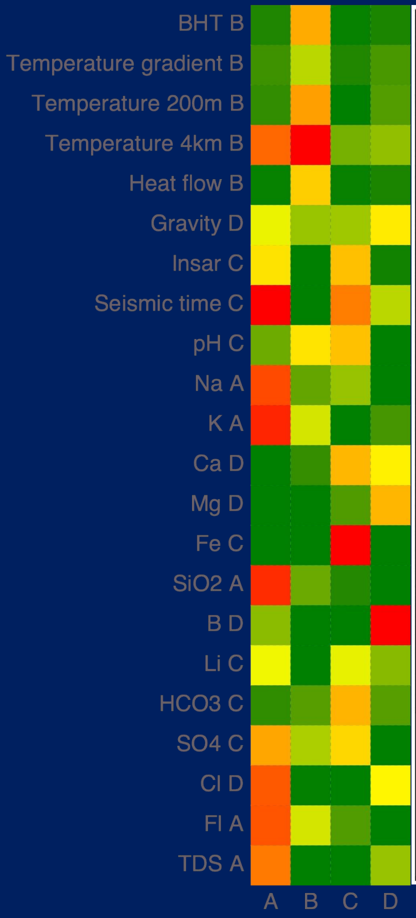
Utah Forge

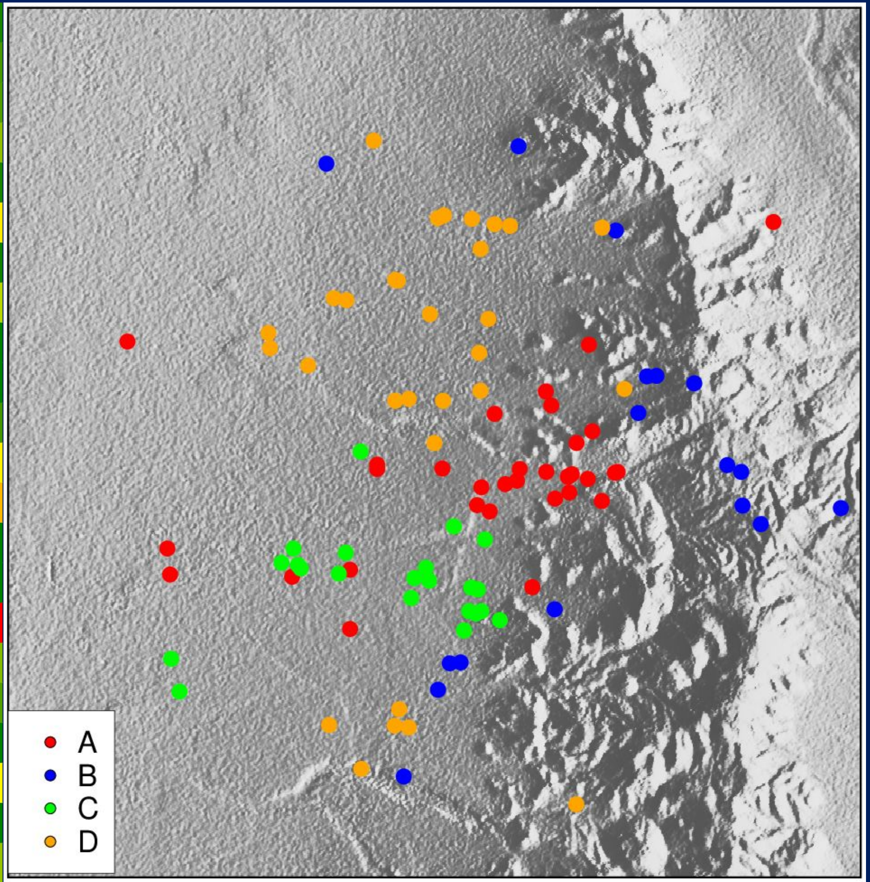


- Data from 102 locations
- 22 attributes including satellite (InSAR), geophysical (gravity, seismic), geochemical, and geothermal attributes
- 4 hidden geothermal signatures extracted
- Two of the signatures (A and B) are related to favorable geothermal conditions

(Will present @ GRC, 2021)

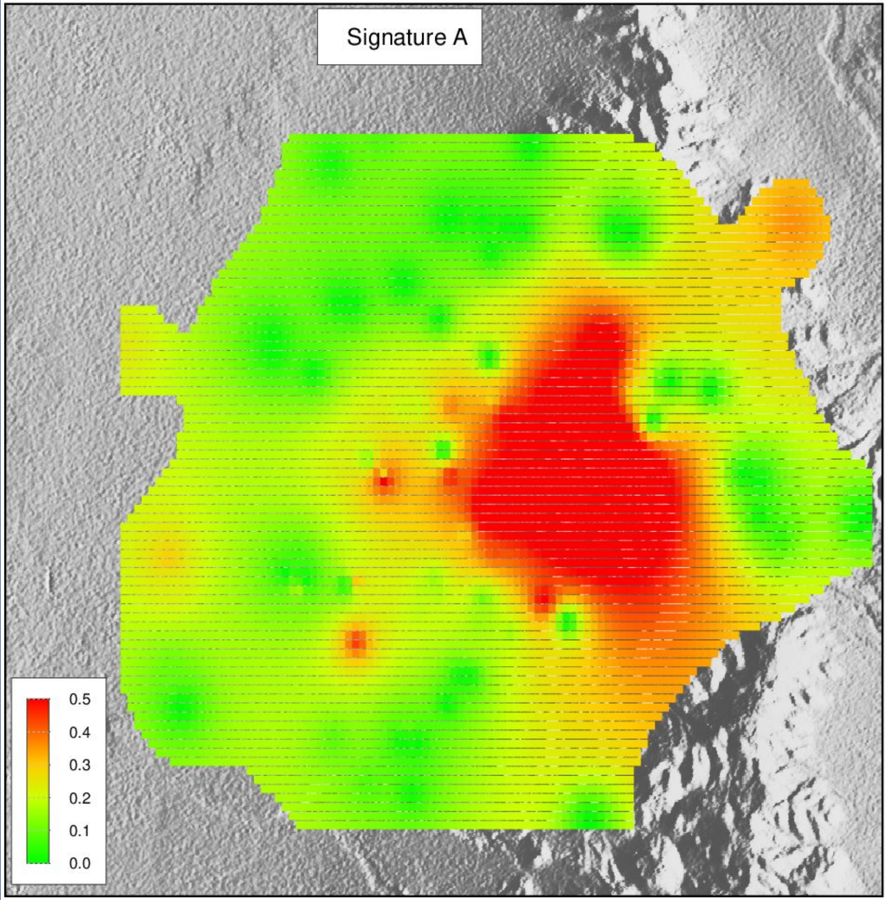
Utah Forge

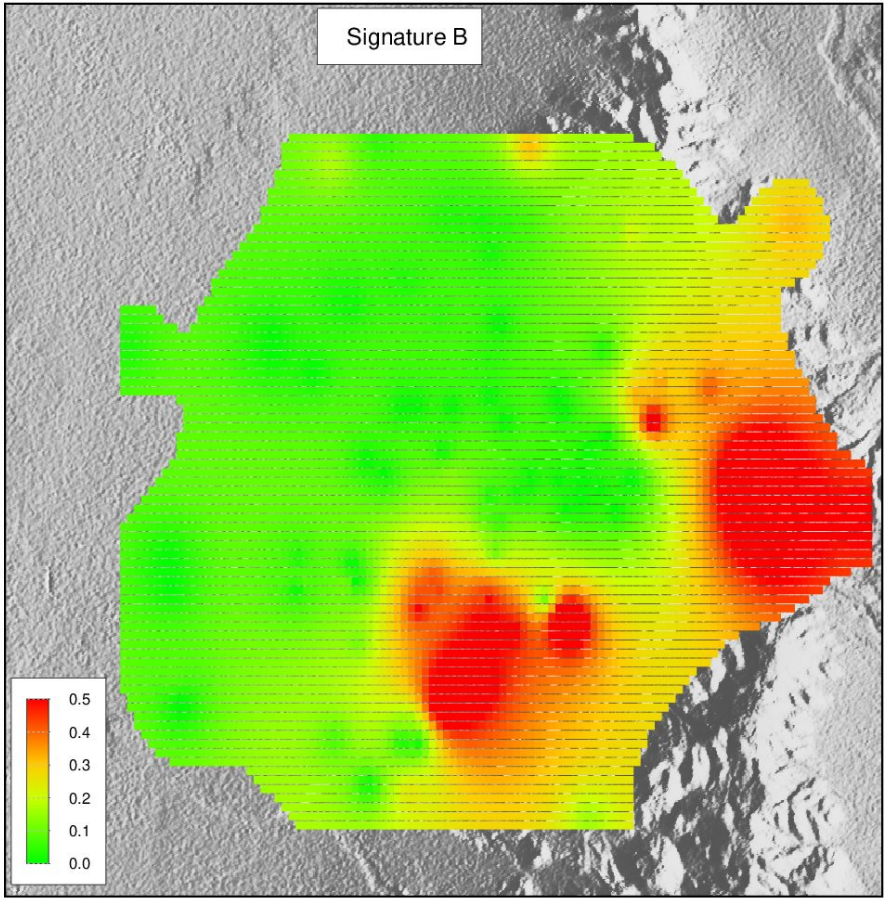




- Four hidden geothermal signatures are extracted
- Signatures A and B are related to favorable geothermal conditions
- However, Signatures A and B are very different
- Signature A key attributes are gravity, seismic, and specific geochemical species
- Signature A is NOT detected by BHT, gradient, head flow, and shallow temperature data

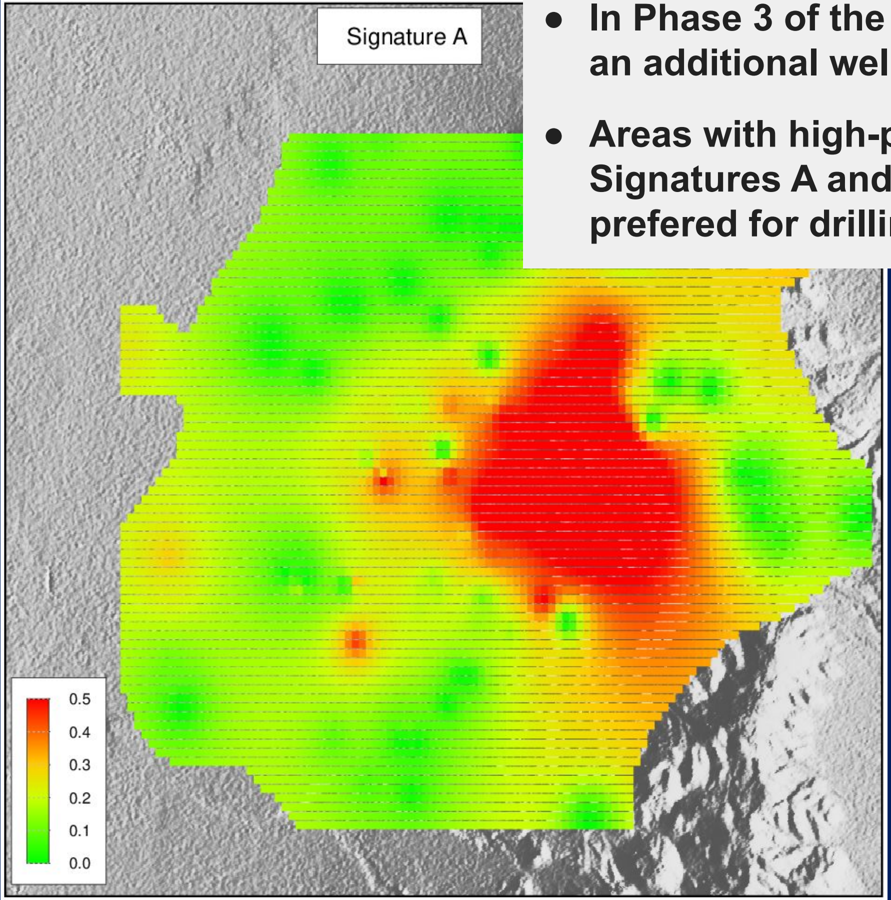
Utah Forge: Prospectivity maps

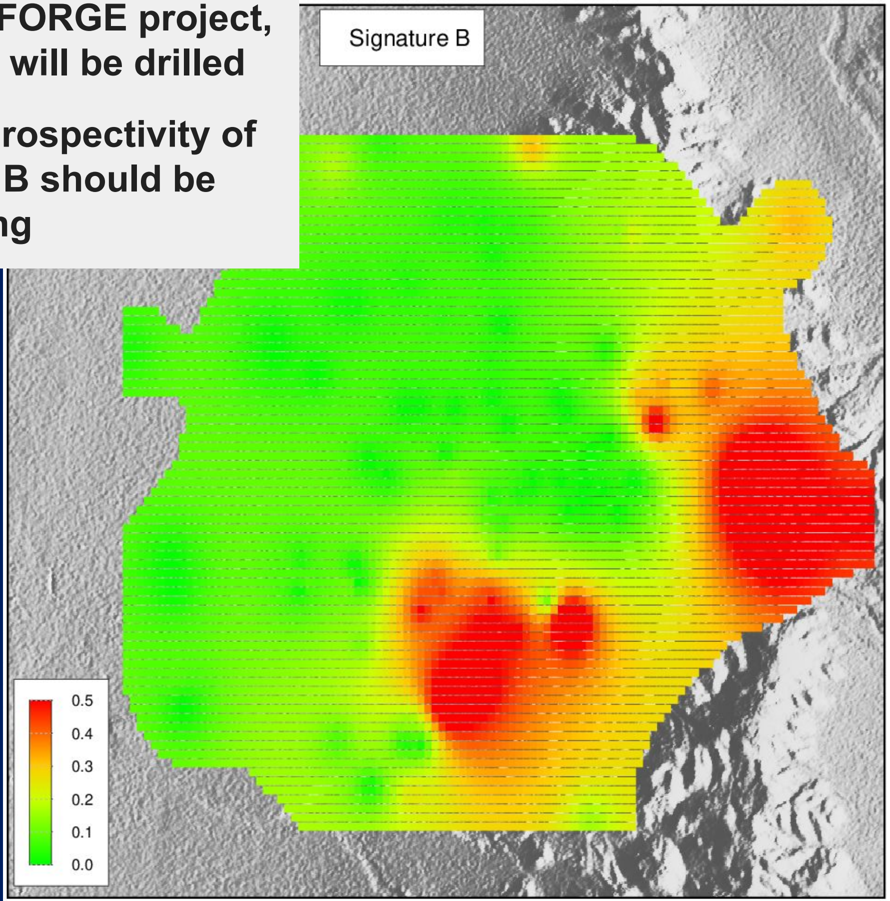




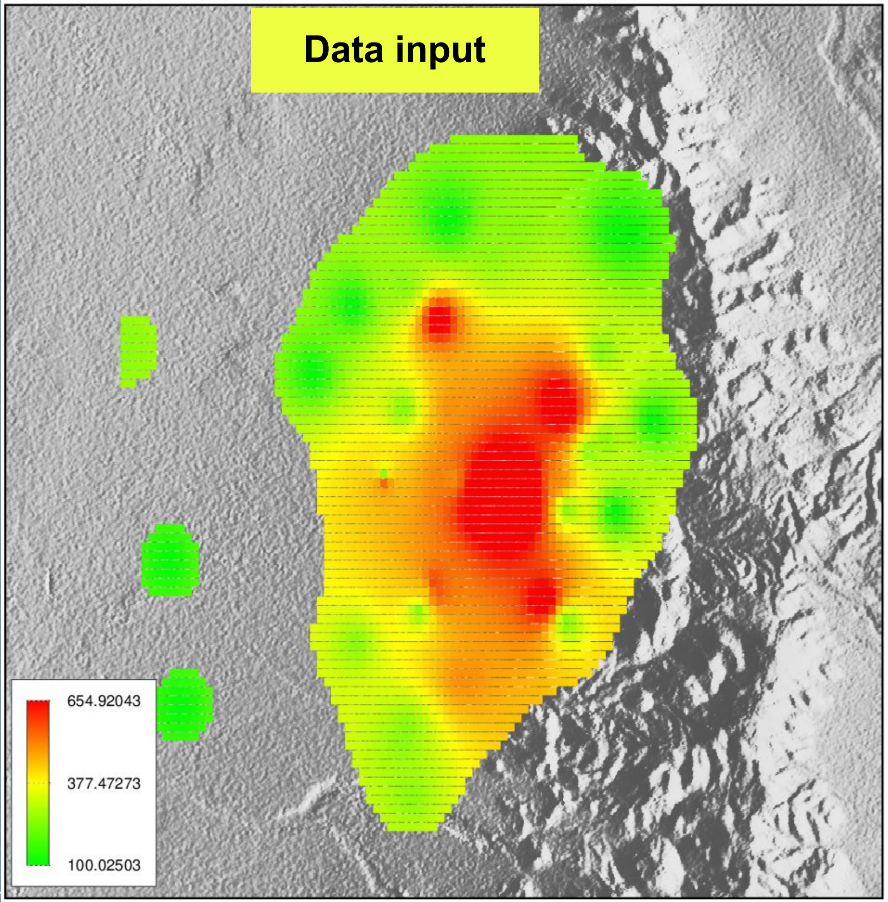
Utah Forge: Prospectivity maps

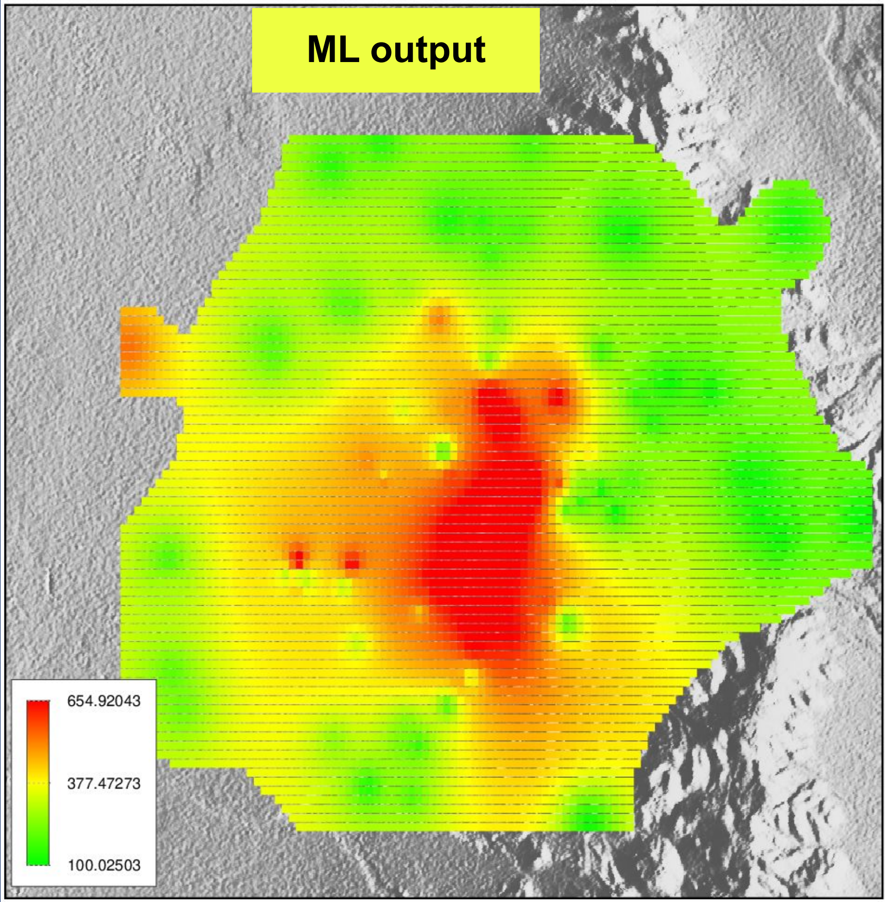
- In Phase 3 of the FORGE project, an additional well will be drilled
- Areas with high-prospectivity of Signatures A and B should be prefered for drilling





Utah Forge: Heat flux (mW/m^2) maps

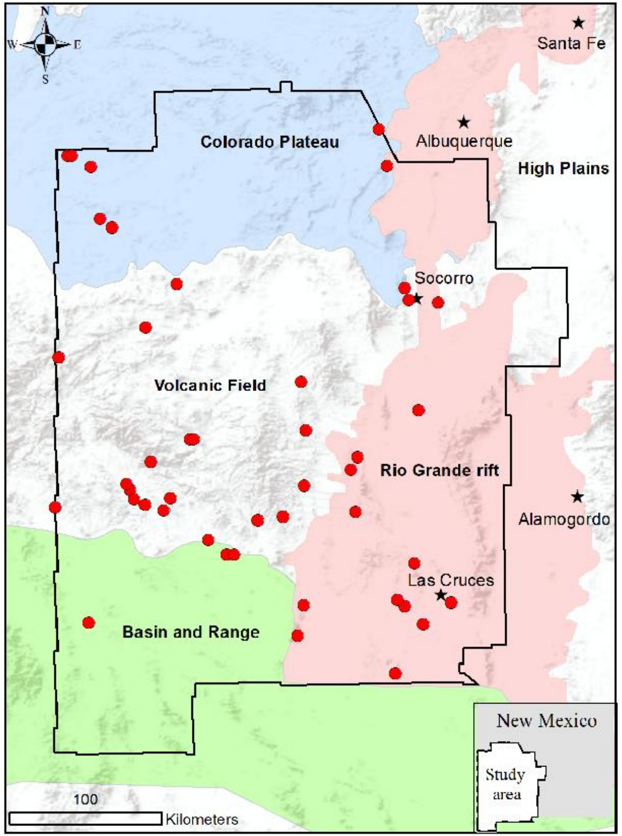




SWNM geothermal exploration

Southwest NM

(Stanford & GRC, 2020)



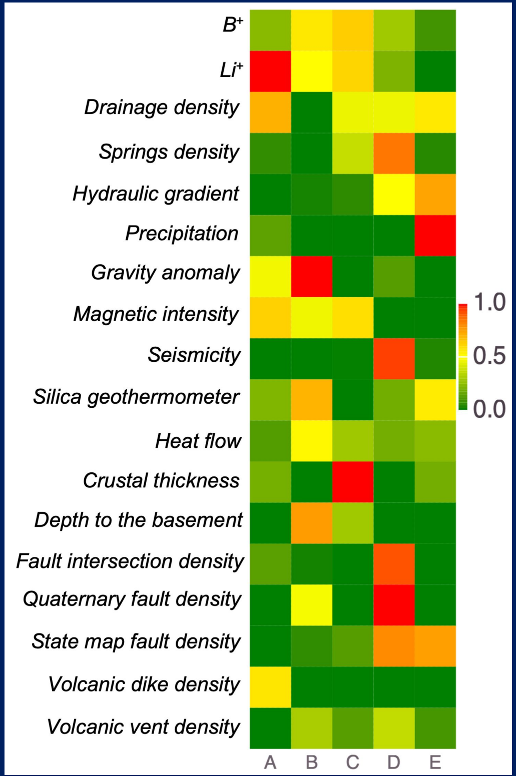
SWNM dataset

X = 44 x *18*
B⁺ concentration
Li⁺ concentration
Drainage density
Springs density
Hydraulic gradient
Precipitation
Gravity anomaly
Magnetic intensity
Seismicity
Silica geothermometer
Heat flow
Crustal thickness
Depth to the basement
Fault intersection density
Quaternary fault density
State map fault density
Volcanic dike density
Volcanic vent density

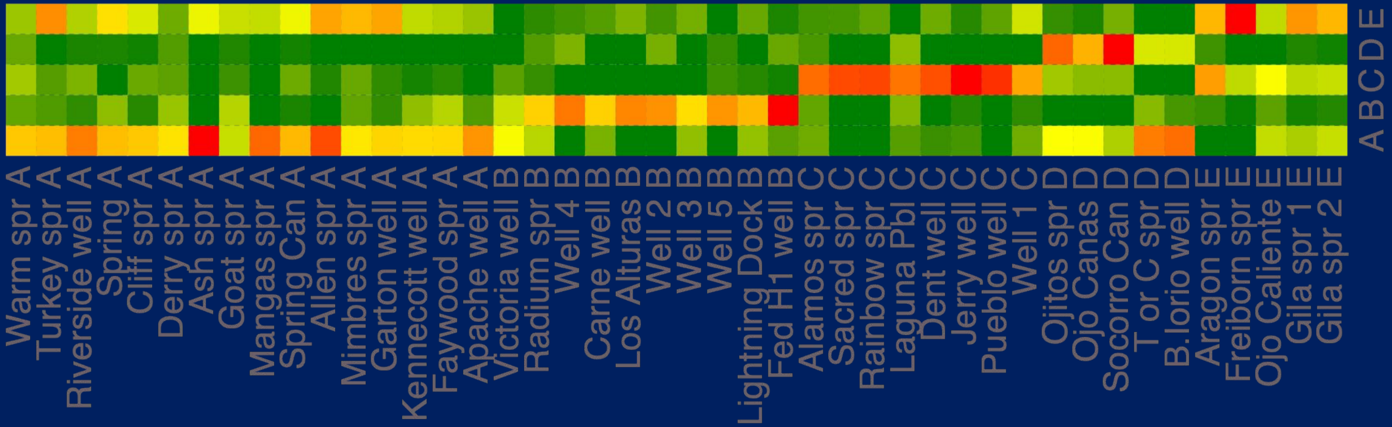
Location	Boron	Gravity	Magnet	Dikes	Drain	Fault	Qfault	Seism	NMFlt	Springs	Vents	Lithium	Precip	Silica	Δh	Qheat	Crust	Bsmt
Alamos Spring	-0.2	-203.3	136.2	0.431	7.4	0.000	0.00	0.004	16.2	0.010	0.003	-3.1	264.8	16.5	5.6	4.6	38.7	1439
Allen Springs	-3.2	-189.3	184.6	3.625	17.3	0.000	0.01	0.002	15.6	0.003	0.001	-4.0	514.5	24.0	13.9	4.4	32.5	51
Apache Tejo Warm Springs well	-1.8	-181.2	15.0	3.807	17.3	0.001	0.03	0.001	0.7	0.003	0.000	-8.6	326.3	52.0	4.7	4.6	30.7	24
Aragon Springs	1.5	-229.1	-317.7	0.010	19.0	0.000	0.00	0.000	41.1	0.005	0.003	-7.5	387.0	56.5	4.0	4.5	38.8	1486
Ash Spring	-2.7	-193.2	66.6	4.914	17.0	0.000	0.00	0.002	9.3	0.003	0.000	-5.0	492.0	29.3	4.1	4.4	32.2	-92
B. Iorio 1 well	-2.1	-196.5	-48.2	1.936	18.8	0.057	21.02	0.000	9.1	0.003	0.003	-2.6	260.4	59.4	0.9	4.0	30.9	-188
Cliff Warm Spring	-2.5	-199.1	-47.1	1.290	22.8	0.001	2.58	0.002	11.0	0.002	0.001	-6.9	364.2	64.2	1.8	4.2	33.1	-191
Dent windmill well	-2.1	-230.8	89.3	0.000	13.4	0.000	0.00	0.000	0.0	0.005	0.000	-7.3	341.7	19.7	2.4	4.7	43.5	865
Derry Warm Springs	-1.5	-161.6	197.0	0.659	18.3	0.007	9.16	0.000	15.9	0.002	0.000	-7.5	276.1	37.4	3.0	4.6	30.0	-120
Faywood Hot Springs	-2.6	-172.1	-49.8	0.939	16.6	0.002	2.81	0.000	1.9	0.003	0.000	-4.8	346.4	67.2	4.2	5.5	30.0	619
Federal H 1 well	-0.4	-132.0	35.0	0.000	5.8	0.004	20.31	0.001	7.2	0.000	0.015	-5.0	253.8	78.7	2.7	4.9	27.3	2906
Freiborn Canyon Spring	-2.5	-225.0	-242.0	0.401	13.1	0.000	0.00	0.001	19.8	0.001	0.004	-12.6	538.6	49.8	13.0	4.6	38.4	1138
Garton well	-3.2	-196.8	35.6	0.150	18.0	0.000	0.00	0.000	28.9	0.002	0.001	-5.0	489.9	70.0	4.3	3.9	30.9	-266
Gila Hot Springs 1	-1.9	-221.6	-149.3	0.127	24.2	0.000	0.00	0.001	25.5	0.003	0.003	-7.8	422.6	69.9	6.6	4.4	34.0	413
Gila Hot Springs 2	-1.8	-222.9	-138.8	0.112	24.7	0.000	0.00	0.001	23.7	0.003	0.003	-6.7	425.9	70.8	3.2	4.6	33.9	519
Goat Camp Spring	-2.1	-159.2	-29.7	0.751	10.0	0.001	2.22	0.007	10.6	0.002	0.001	-8.0	344.0	68.9	5.8	4.4	32.4	19
Jerry well	-0.8	-219.6	172.4	0.111	15.5	0.000	0.00	0.000	6.3	0.004	0.005	-7.9	243.9	13.4	1.0	4.4	42.3	1190
Kennecott Warm Springs well	-2.4	-178.3	-69.9	1.422	17.8	0.002	1.76	0.000	1.1	0.003	0.000	-6.9	355.0	66.1	4.3	5.0	30.0	409
Laguna Pueblo	0.4	-204.2	62.5	0.406	8.6	0.004	4.58	0.006	14.6	0.018	0.005	-3.3	259.7	42.9	2.6	4.4	37.2	1506
Lightning Dock	-1.0	-168.0	-168.1	0.086	4.6	0.008	8.40	0.002	4.3	0.000	0.000	-3.9	291.5	107.3	0.8	5.0	29.8	1800
Los Alturas Estates	-1.5	-141.4	-127.5	0.004	7.6	0.003	0.05	0.002	6.6	0.001	0.000	-12.7	265.3	71.9	2.2	6.3	27.4	4321
Mangas Springs	-2.6	-201.0	-227.1	3.503	20.2	0.000	0.91	0.002	11.5	0.002	0.000	-4.5	393.5	53.6	0.3	4.2	32.4	-178
Mimbres Hot Springs	-2.3	-200.6	43.4	0.670	15.4	0.002	1.13	0.000	19.0	0.004	0.000	-3.8	445.9	68.3	9.1	4.9	31.0	50
Ojitos Springs	-1.6	-202.1	-7.5	1.342	19.6	0.044	19.74	0.037	31.0	0.020	0.005	-4.5	257.5	57.6	7.2	4.5	33.0	-255
Ojo Caliente	-2.6	-226.5	-168.4	0.000	20.5	0.000	0.00	0.000	8.3	0.004	0.000	-2.9	333.6	48.4	3.5	5.5	33.8	2415
Ojo De las Canas	-1.7	-188.5	-85.8	0.839	22.3	0.036	12.55	0.036	28.0	0.013	0.003	-6.0	270.5	14.2	4.0	4.5	31.8	101
Pueblo windmill well	-1.2	-228.8	315.9	0.029	15.2	0.000	0.00	0.000	6.1	0.004	0.003	-12.0	265.8	18.3	2.9	4.3	42.5	1027
Radium Hot Springs	-0.8	-151.4	-7.8	0.010	8.8	0.013	11.40	0.003	10.6	0.001	0.000	-5.3	264.2	63.6	0.3	5.4	28.2	1191
Rainbow Spring	-1.7	-227.1	-48.5	0.000	11.0	0.000	0.00	0.001	0.0	0.006	0.000	-7.0	307.8	21.7	3.3	4.7	43.9	755
Riverside Store well	-1.3	-196.1	-102.9	1.562	22.6	0.000	2.50	0.002	11.7	0.002	0.001	-2.4	356.1	60.8	0.9	4.3	32.9	-165
Sacred Spring	-1.8	-228.4	-80.4	0.000	10.9	0.000	0.00	0.001	0.0	0.006	0.000	-7.0	298.4	21.2	1.3	4.6	43.9	742
Socorro Canyon	-1.8	-204.7	-136.5	1.203	21.1	0.051	28.88	0.034	33.8	0.020	0.005	-6.7	284.1	44.6	11.1	5.0	32.6	-229
Spring	-4.1	-183.5	334.5	0.218	20.1	0.011	1.81	0.000	20.1	0.001	0.006	-6.8	361.9	117.2	5.1	3.8	31.5	-104
Spring Canyon Warm Spring	-2.1	-194.2	117.3	2.293	21.9	0.000	1.50	0.002	12.7	0.002	0.000	-8.3	361.7	51.6	5.8	4.2	32.6	-57
Truth or Consequences spring	-1.1	-168.2	-54.3	2.175	18.4	0.064	20.51	0.000	10.3	0.003	0.002	-3.3	265.9	55.3	0.6	4.3	31.0	304
Turkey Creek Spring	-3.2	-196.4	54.8	0.984	19.2	0.001	3.69	0.002	28.1	0.002	0.002	-3.7	493.4	81.3	5.8	4.4	33.6	56
Victoria Land and Cattle Co. well	-1.8	-165.9	-65.4	0.478	6.4	0.003	0.06	0.001	0.9	0.001	0.000	-2.9	253.0	43.0	1.9	4.1	30.7	2014
Warm Springs	-2.1	-193.3	113.5	0.220	19.0	0.029	2.63	0.000	16.5	0.004	0.003	-2.5	314.6	56.0	5.4	4.3	32.7	1252
Well 1	-1.4	-230.7	-31.3	1.190	15.7	0.000	0.75	0.001	22.1	0.004	0.002	-6.6	345.4	49.0	1.7	4.4	40.0	1961
Well 2	-1.2	-162.5	0.8	0.000	4.5	0.008	24.24	0.003	11.8	0.000	0.006	-10.1	279.5	70.5	1.7	4.8	27.8	2993
Well 3	-2.5	-140.0	31.7	0.839	2.1	0.001	2.11	0.001	5.0	0.001	0.000	-7.3	369.0	51.0	4.1	4.3	28.0	3073
Well 4	-1.3	-161.7	-56.1	0.000	3.4	0.008	28.49	0.003	10.6	0.000	0.006	-10.0	274.3	94.0	1.9	4.7	27.7	3373
Well 5	-1.9	-167.2	-29.9	0.000	2.5	0.008	15.48	0.002	3.1	0.000	0.005	-6.8	243.8	47.0	0.3	4.0	27.4	5460
Well south of Carne	-2.4	-156.7	-129.6	0.457	4.3	0.000	2.11	0.002	6.0	0.001	0.000	-6.8	269.7	87.1	1.4	4.5	28.4	2761

SWNM geothermal signatures

W



H

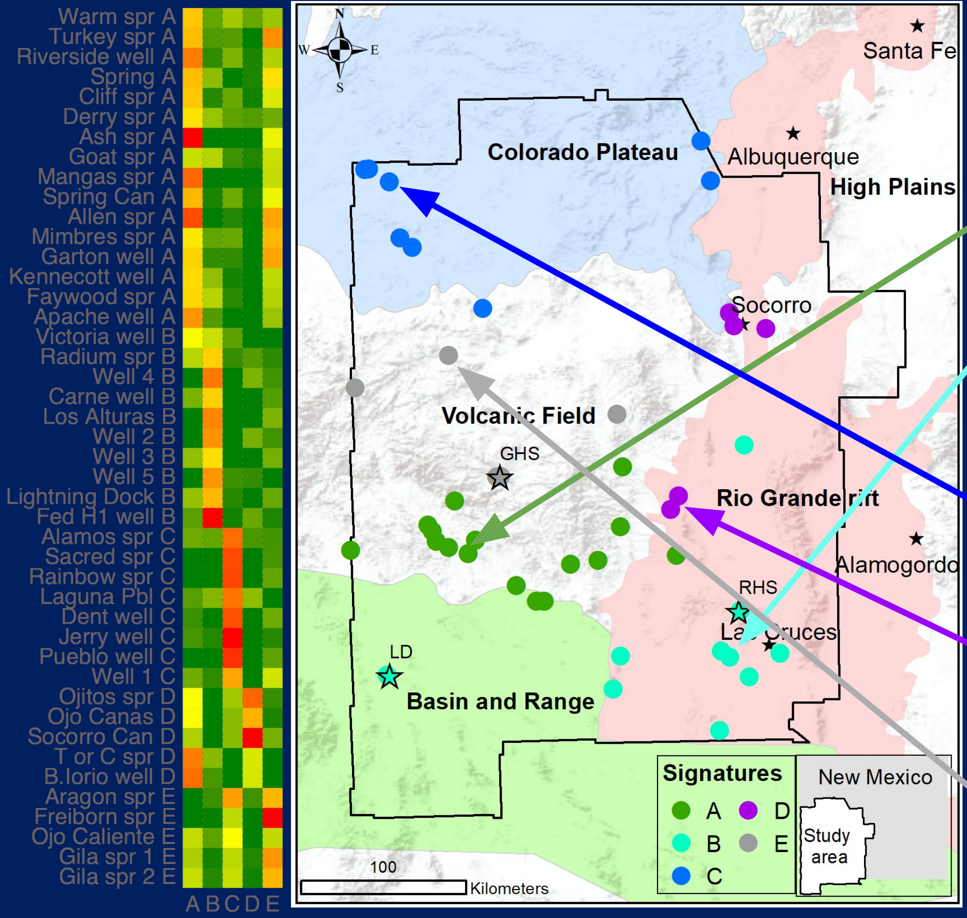


X* = *W* x *H

W: attribute matrix

H: location matrix

SWNM physiographic provinces



Physiographic associations:

Signature A: Southern volcanic field

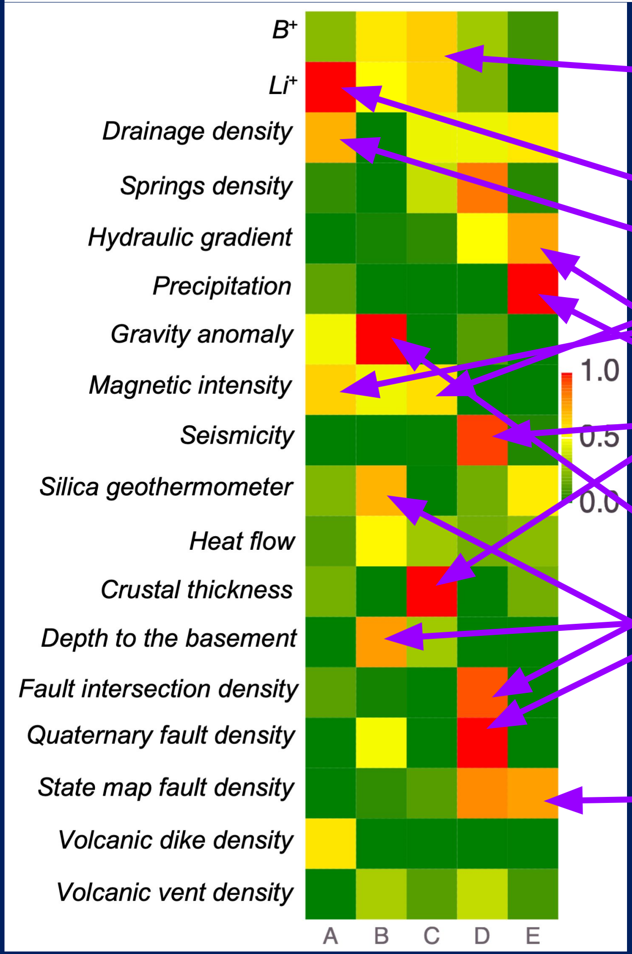
Signature B: Rio Grande Rift

Signature C: Colorado Plateau

Signature D: Central Rio Grande Rift

Signature E: Northern volcanic field

SWNM signature interpretation



Signature A:

Li^+ concentration
Drainage density
Magnetic intensity
Volcanic dike density
Gravity anomaly

Signature B:

Gravity anomaly
Depth to the basement
Silica geothermometer
B^+ and Li^+ concentrations
Magnetic intensity
Quaternary fault density
Heat flow

Signature C:

Crustal thickness
Magnetic intensity
B^+ and Li^+ concentrations
Drainage density

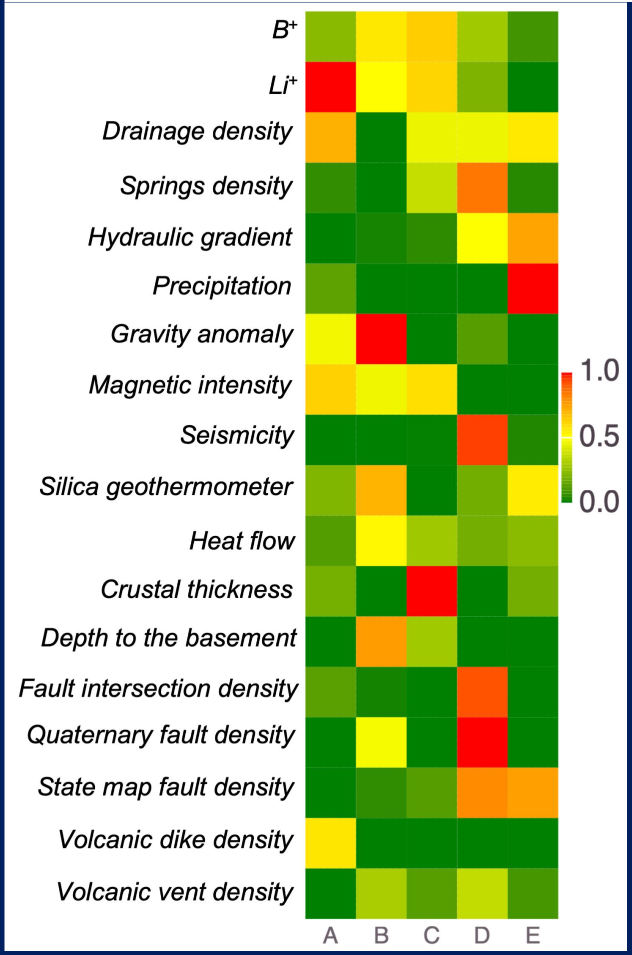
Signature D:

Quaternary fault density
Fault intersection density
Seismicity
State map fault density
Spring density
Hydraulic gradient
Drainage density

Signature E:

Precipitation
Hydraulic gradient
State map fault density
Drainage density
Silica geothermometer

SWNM signature interpretation



Signature A:
Shallow heat flow

Li^+ concentration
Drainage density
Magnetic intensity
Volcanic dike density
Gravity anomaly

Signature B:
Deep heat flow

Gravity anomaly
Depth to the basement
Silica geothermometer
B^+ and Li^+ concentrations
Magnetic intensity
Quaternary fault density
Heat flow

Signature C:
Thick crust

Crustal thickness
Magnetic intensity
B^+ and Li^+ concentrations

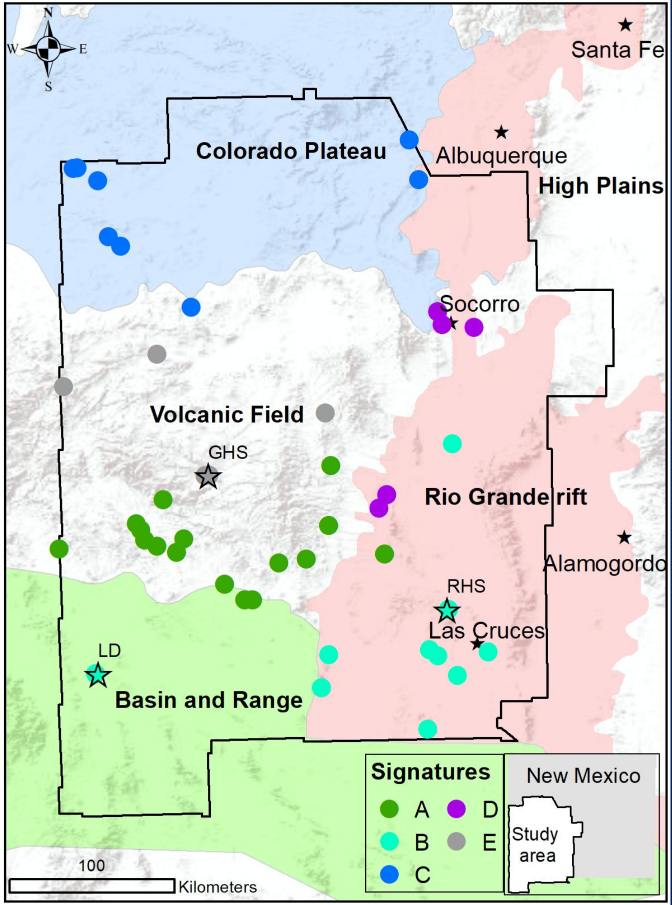
Signature D:
Tectonics

Quaternary fault density
Fault intersection density
Seismicity
State map fault density
Spring density

Signature E:
Vertical hydraulics

Precipitation
Hydraulic gradient
State map fault density

SWNM geothermal signatures



Signature A: Southern volcanic field
Shallow heat flow

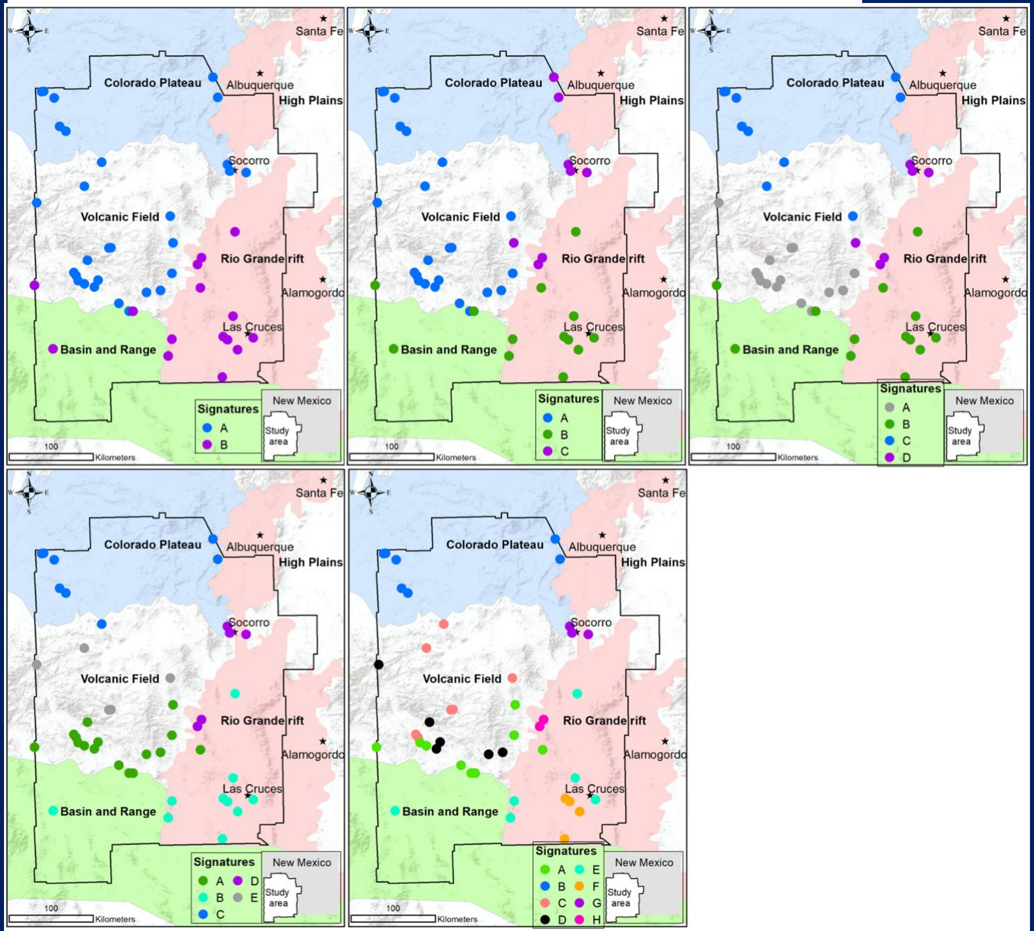
Signature B: Rio Grande Rift
Deep heat flow

Signature C: Colorado Plateau
Thick crust

Signature D: Central Rio Grande Rift
Tectonics

Signature E: Northern volcanic field
Vertical hydraulics

SWNM geothermal signatures



- 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

Conclusions:

- GeoThermalCloud is developed for ML analyses of geothermal datasets
- Our novel open-source ML methods have successfully extracted hidden geothermal signatures
- We were able to provide physical explanation of these signatures
- ML was applied to label datasets related to the geothermal signatures
- GeoThermalCloud capabilities were demonstrated on 9 field and 2 synthetic datasets

Conclusions:

- **Great Basin**: Low-, medium-, high-temperature hydrothermal systems, their dominant characterization attributes, and their spatial distribution identified using geochemistry data
- **Brady site**: Successfully defined relations between well types (production, injection, non-production) and attributes characterizing site conditions (faulting, geology, state of stress)
- **Utah FORGE**: Analyzed site prospectivity and proposed drilling location for future geothermal field exploration
- **SWNM**: Identified low- and medium-temperature hydrothermal systems, found dominant attributes and spatial distribution for each hydrothermal system; demonstrated blind predictions of provinces

Conclusions:

- **Tularosa Basin**: Identified low-, medium-, and high-temperature hydrothermal systems, found dominant attributes and spatial distribution for each hydrothermal system
- **Hawaii**: Analyzed four islands data separately and identified low-, medium-, and high-temperature hydrothermal systems
- **Tohatchi Springs**: Identified low- and medium-temperature hydrothermal systems, found dominant attributes and spatial distribution for each hydrothermal system
- **West Texas**: Subdivided the region into three areas; the western portion has higher geothermal potential at a lower depth than the middle and eastern portions

Conclusions:

- **EGSCollab: Field experiment data processed to extract dominant temporal patterns observed in 49 data streams; erroneous measurement attributes and periods automatically identified; interrelated data streams automatically identified**

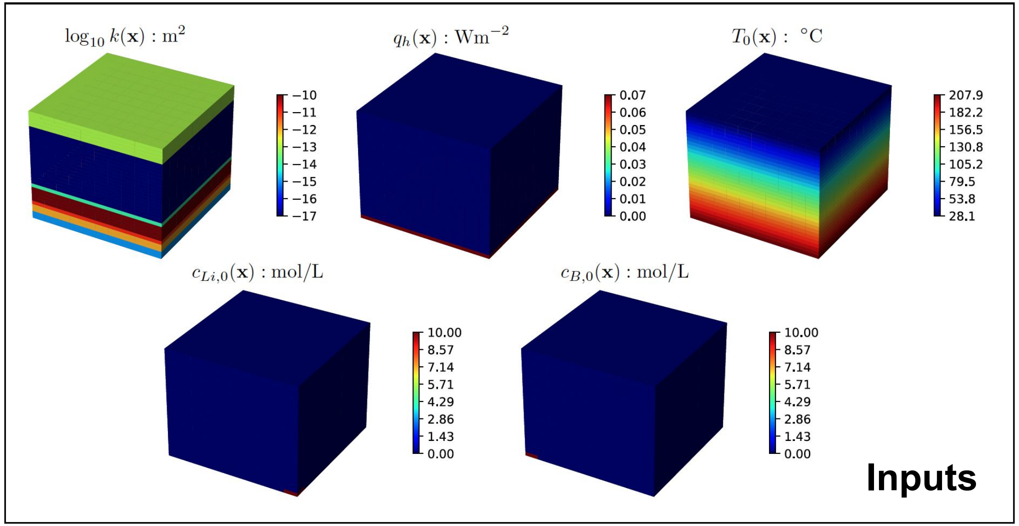
Conclusions:

- GeoDT multiphysics code is developed to rapidly predict the performance of geothermal systems
- GeoDT predicts the impact of attainable site data on geothermal performance
- GeoThermalCloud “separates” the impacts of different physical processes in the GeoDT model outputs
- GeoDT+GeoThermalCloud capabilities demonstrated on a synthetic dataset
- Well spacing and well orientation are identified to be critical parameters impacting energy production and induced seismicity

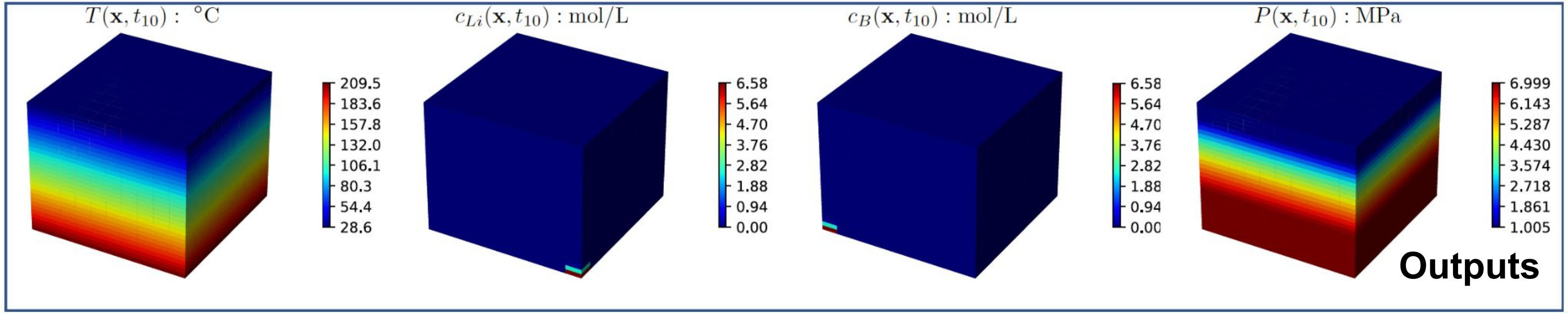
GeoThermalCloud:
Extra slides summarizing more geothermal studies

- **Synthetic SWNM ML analyses**
- **Tularosa Basin**
- **Tohatchi hot springs**
- **West Texas**
- **Hawaii Islands**
- **EGS Collab**

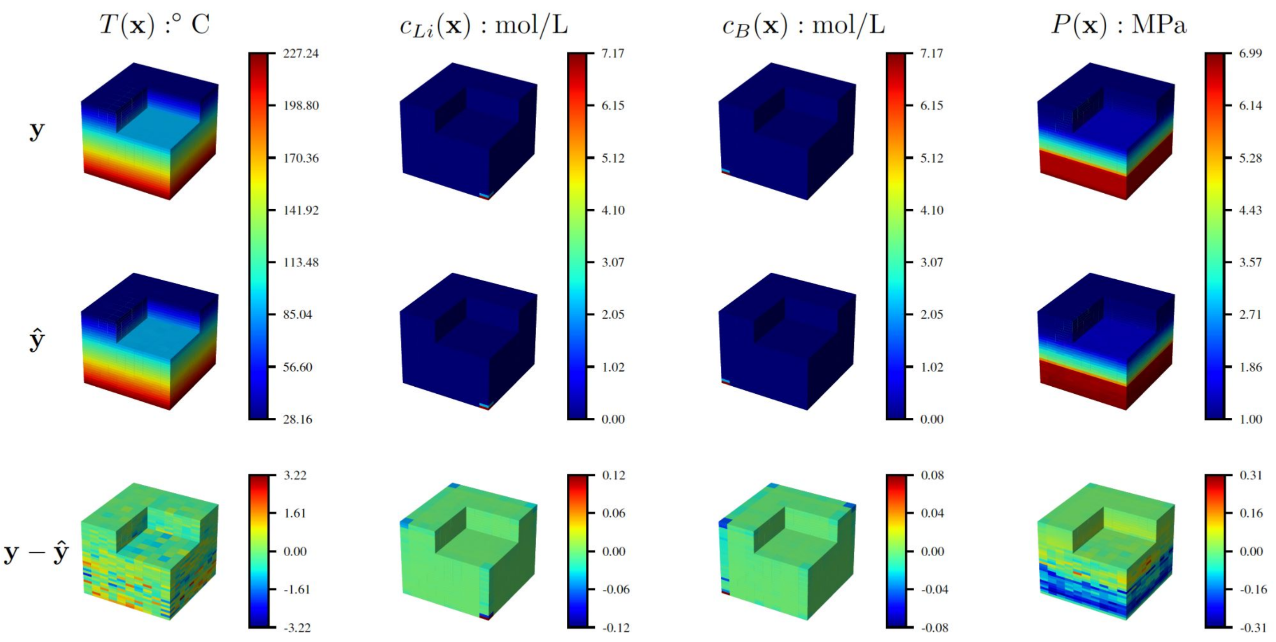
Synthetic ML analyses: SWNM site



A 3D thermo-hydrologic-chemical model was built
Variable model inputs: permeability, thermal gradient, and heat flux
Outputs: Temperature, Li and B concentrations



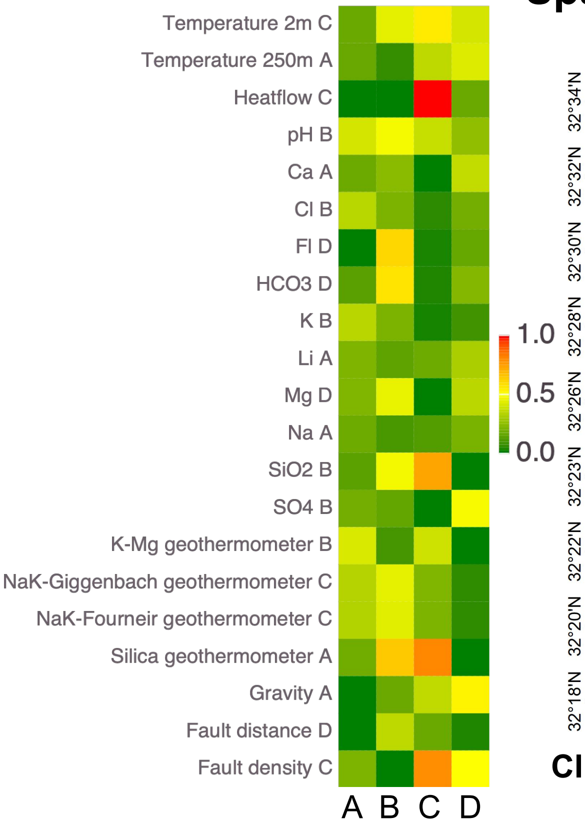
Synthetic ML analyses: Results



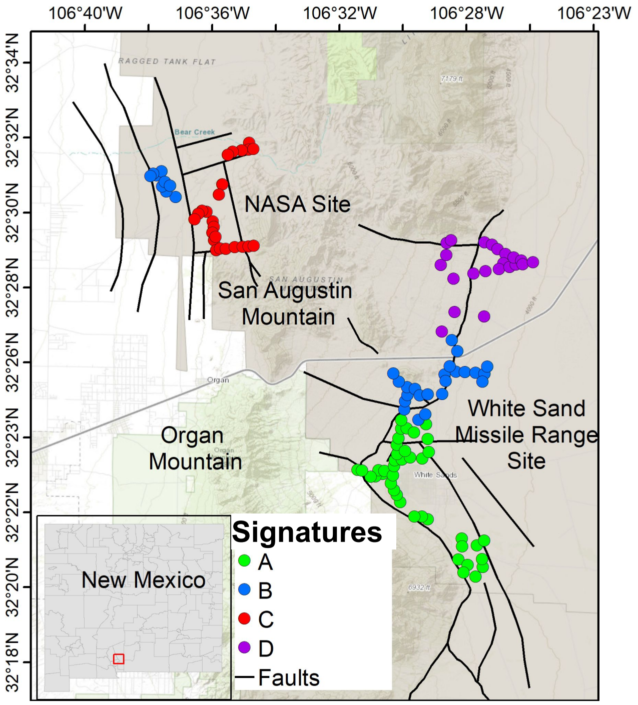
- **2,000 simulations**
- **1,600 pairs for training**
- **400 pairs for testing**
- **3D Convolutional encoder-decoder NN was built using (3,6,3) dense blocks**
- **NN trained for 200 epoches**
- **L_1 norm loss function is used for gradient descent optimization**

Tularosa Basin: Results

Hidden signals



Spatial distribution of signatures

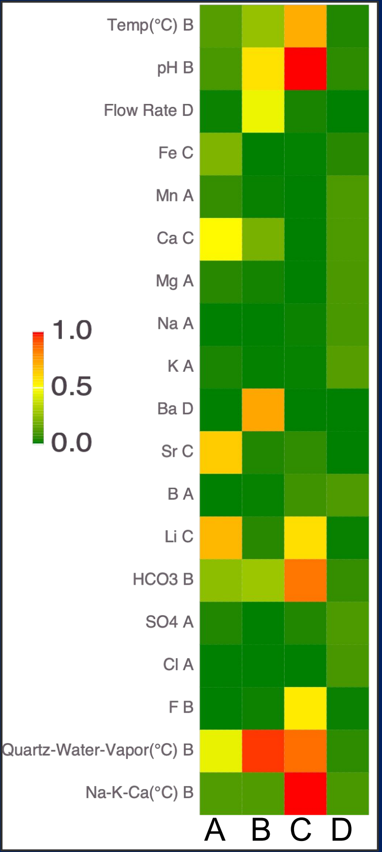


Clustered points represent data locations

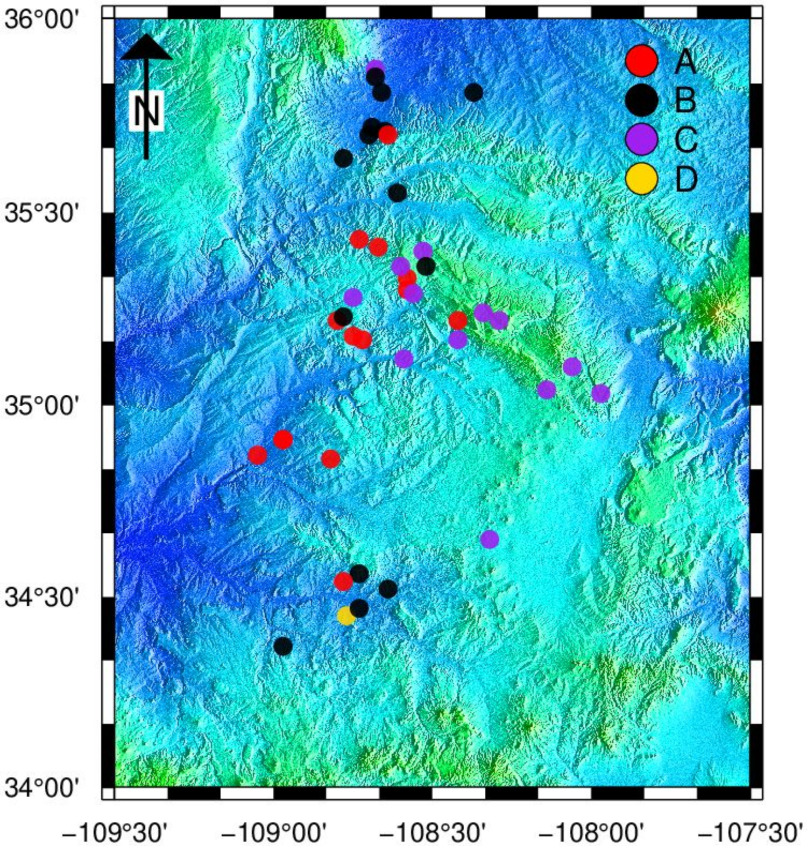
- **Tularosa basin (South New Mexico) has favorable geological structures for geothermal exploration**
- **We investigate a total of 21 attributes collected for PFA** [https://gdr.openei.org/submissions/928]
- **Signature C defines the hidden potential geothermal resources**
- **Signature C key attributes are heat flow, SiO_2, silica geothermometer and fault density**

Tohatchi hot spring area, NM

Hidden signals



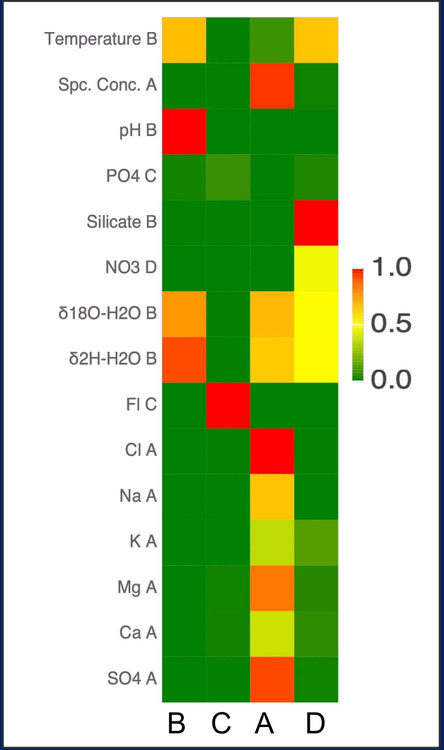
Spatial distribution of signatures



- Tohatchi hot springs in NM are favorable for hot dry rock geothermal exploration
- We investigated 19 attributes observed at 41 wells
- Signature C defines the hidden potential geothermal resources
- Signature C key attributes are pH, Li^{+} HCO$_3^{-}$, F^{+}, Quartz-water-vapor geothermometer and Na-K-Ca geothermometer

(supported by NM SBA, LANL)

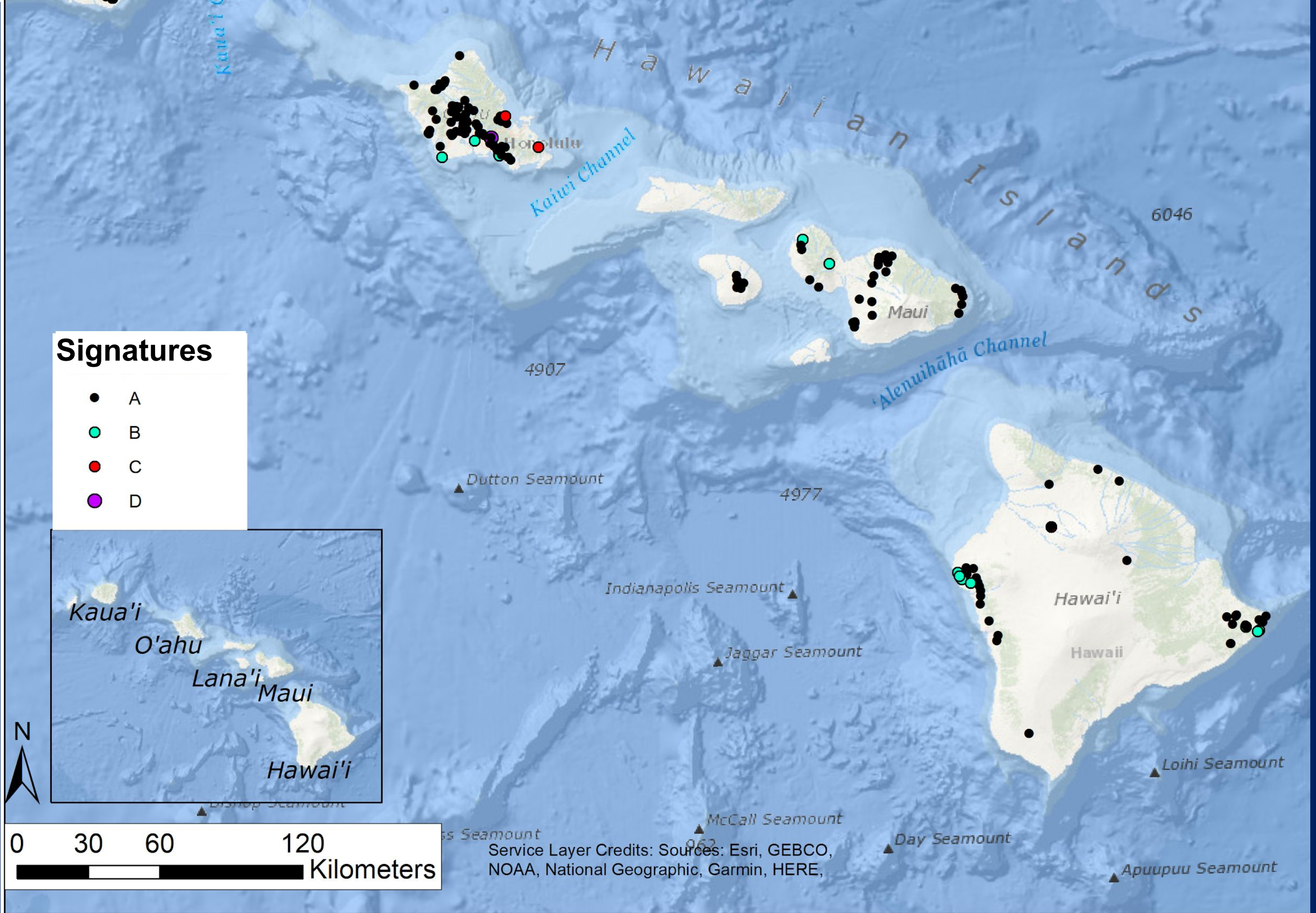
Hawaii Islands



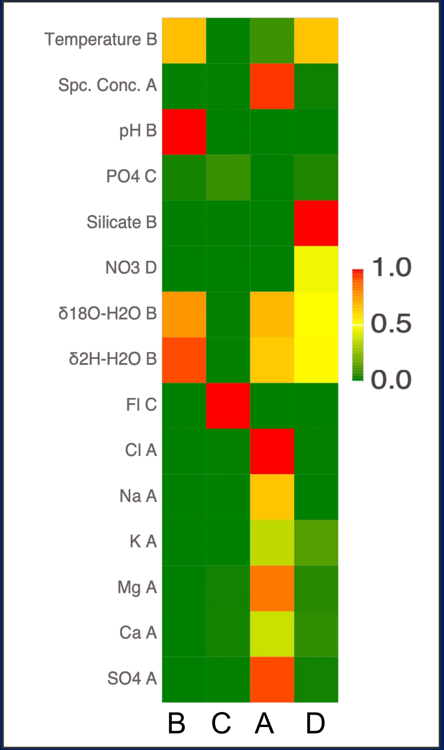
Hidden signatures

Los Alamos National Laboratory

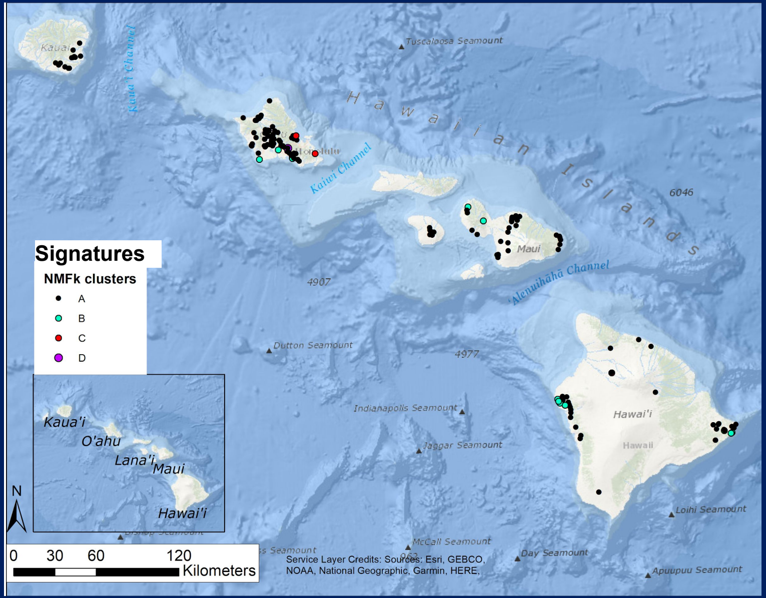
Spatial distribution of signatures



Hawaii Islands



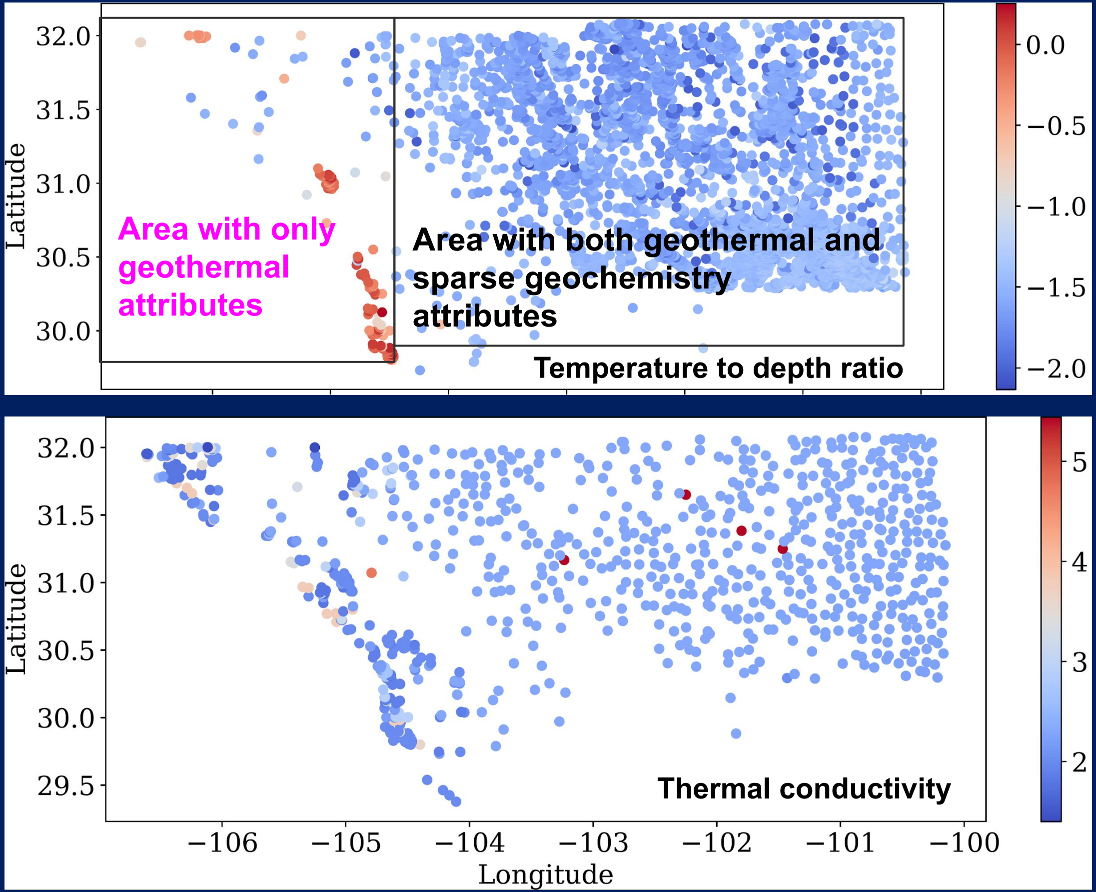
Hidden signatures



Spatial distribution of signatures

- **Four geothermal signatures characterize Hawaii islands**
- **Signatures B and D relate with groundwater temperature**
- **Their dominant attributes are:**
  - **pH**
  - **$\delta^{18}O$**
  - **$\delta^{2}H$**
  - **Silicate**

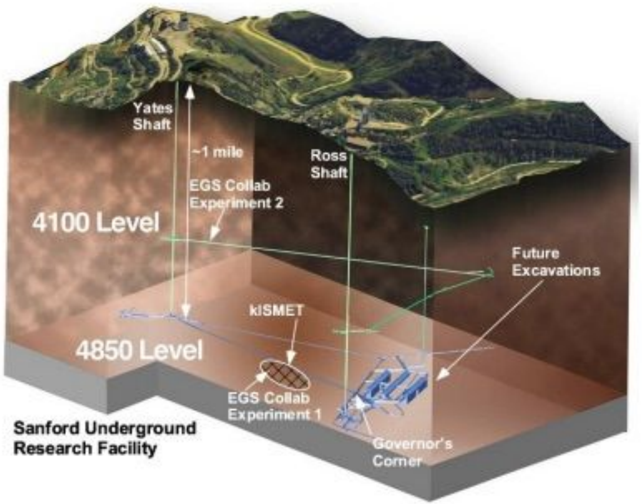
West Texas

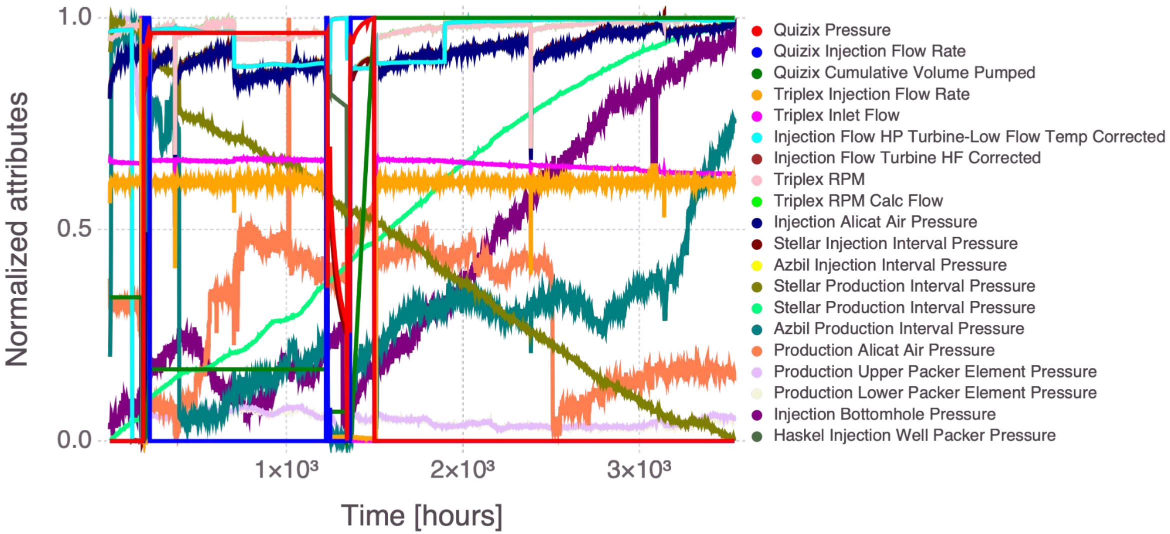


- Bottomhole temperature to depth ratio is higher in the western vs eastern areas
- Thermal conductivity is marginally higher in the west portion
- Temperature to depth ratio and Thermal conductivity demonstrate that western area has potential geothermal systems at a lower depth than the middle and eastern areas
- In Phase II, we will divide the dataset and perform transfer learning

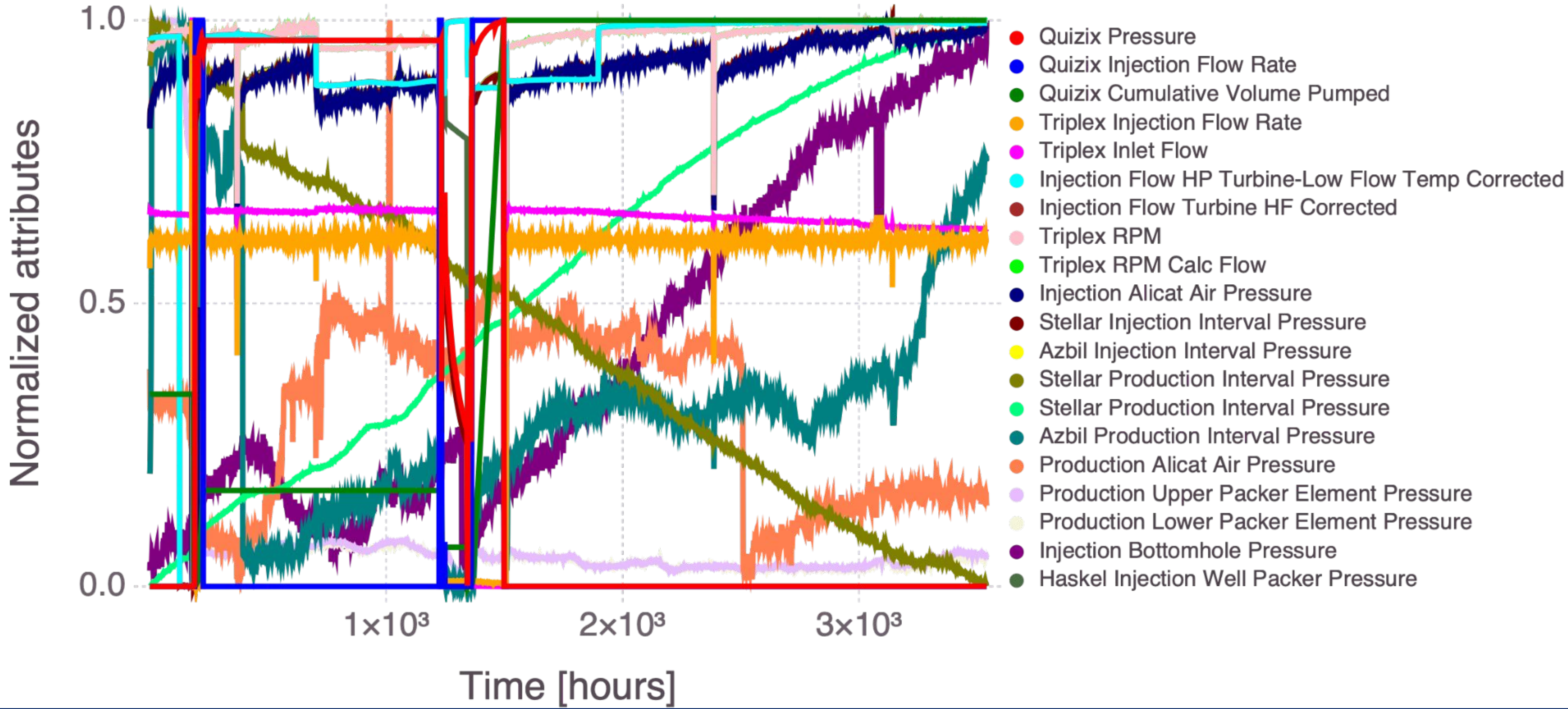
EGS Collab: Field experimental data

- Hourly field data collected during field experiments are analyzed
- Measurement interrelation are hard to understand (49 attributes processed)
- Some measurements are erroneous due to equipment failures

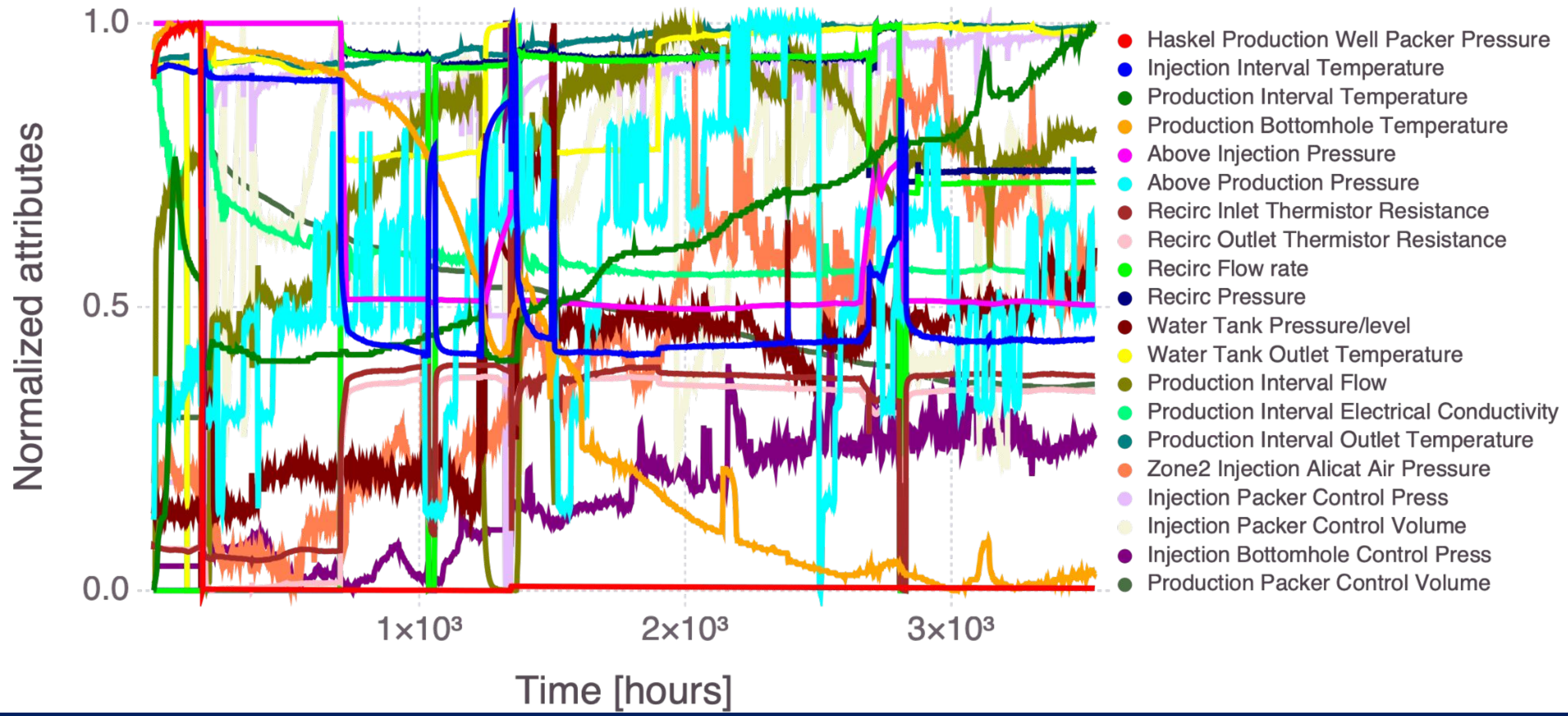




EGS Collab: Field experimental data

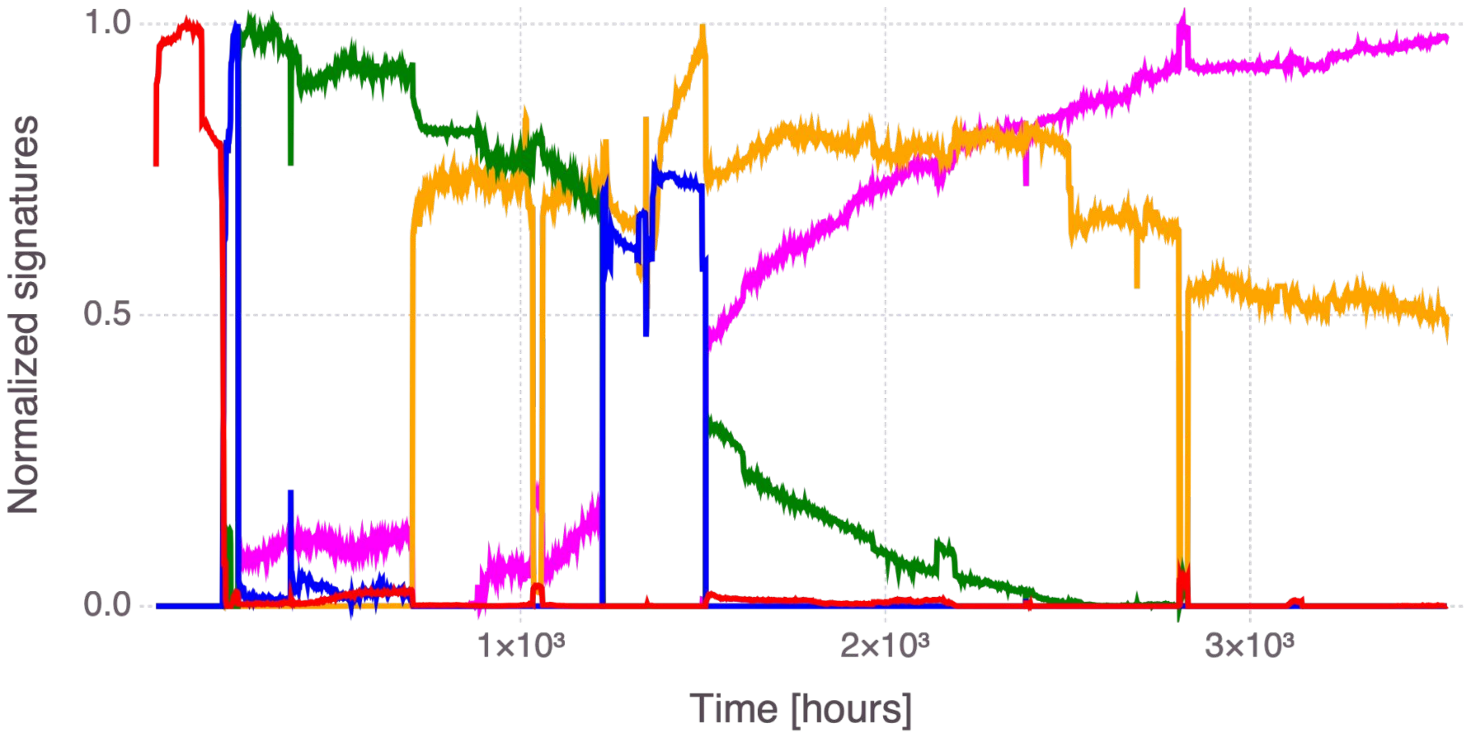


EGS Collab: Field experimental data



EGS Collab: ML results

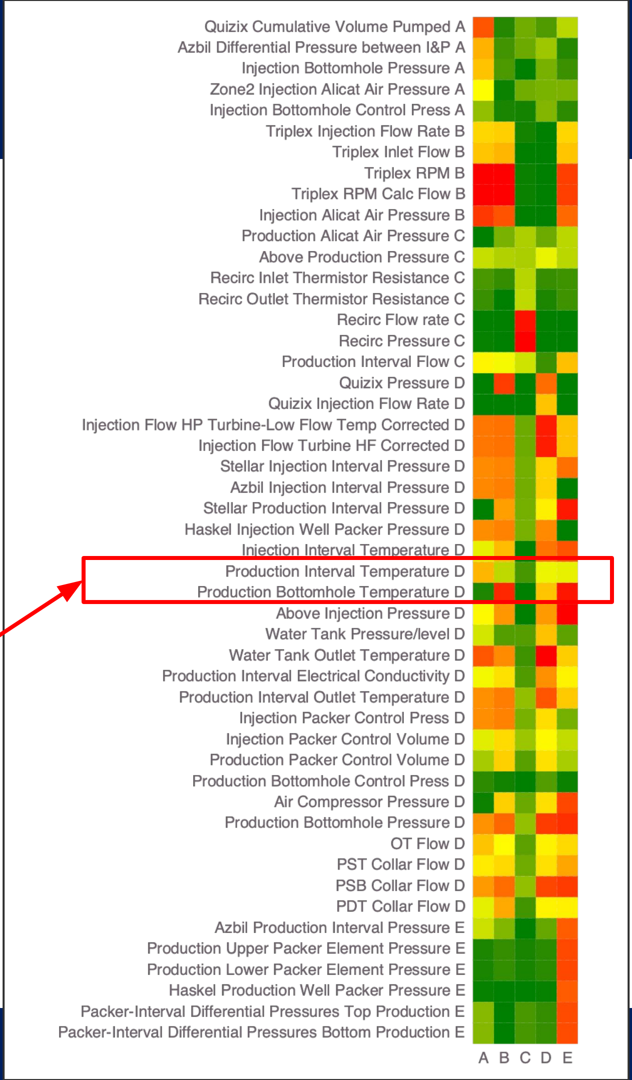
- **5 signatures "represent" the field experiment**



A: injection flow
B: injection pump rate
C: production temperature/pressure
D: mixed processes
E: production flow rate

EGS Collab: ML results

- Each signature is related to a series of measurement attributes
- Importance (weights) of attributes are evaluated
- Interrelated measurement attributes are identified (i.e., representing similar processes)
- Erroneous measurement attributes are identified (e.g., “Production Interval Temperature”)



EGS Collab: Interrelated measurements

Signatures / Measurement attributes	Weights
Signature A	
Quizix Cumulative Volume Pumped	0.919
Azbil Differential Pressure between I&P	0.707
Injection Bottomhole Pressure	0.664
Signature B	
Triplex RPM Calc Flow	1.0
Triplex RPM	1.0
Injection Alicat Air Pressure	0.921
Signature C	
Recirc Pressure	1.0
Recirc Flow rate	0.994

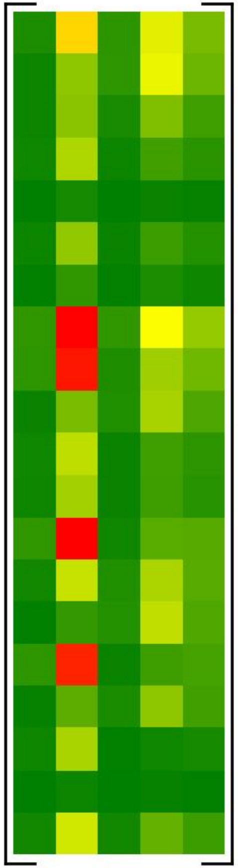
EGS Collab: Interrelated measurements

Signatures / Measurement attributes	Weights
Signature D	
Water Tank Outlet Temperature	1.0
Injection Flow Turbine HF Corrected	0.988
Injection Flow HP Turbine-Low Flow Temp Corrected	0.988
Production Bottomhole Pressure	0.957
PSB Collar Flow	0.943
Production Interval Outlet Temperature	0.918
Quizix Pressure	0.874
Injection Interval Temperature	0.86
Signature E	
Production Lower Packer Element Pressure	0.941
Production Upper Packer Element Pressure	0.94

GeoThermalCloud: Slides summarizing more about methodology

- **NMFk**
- **NTFk**

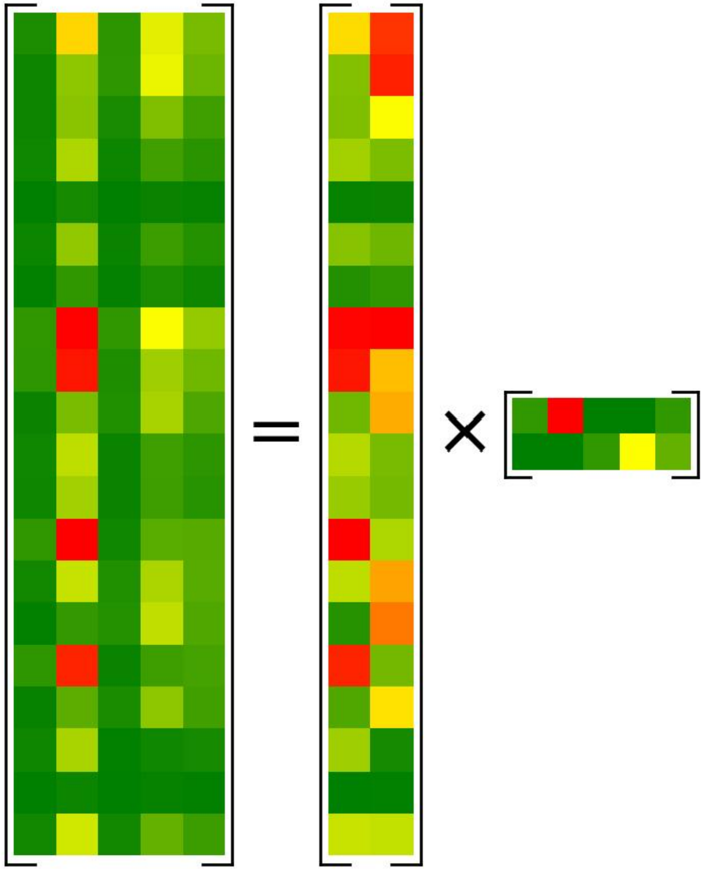
Nonnegative matrix factorization



$\boldsymbol{X}$
$[20 \times 5]$

$\boldsymbol{X}$ – **data** matrix
[**attributes** $\times$ **locations**]

Nonnegative matrix factorization



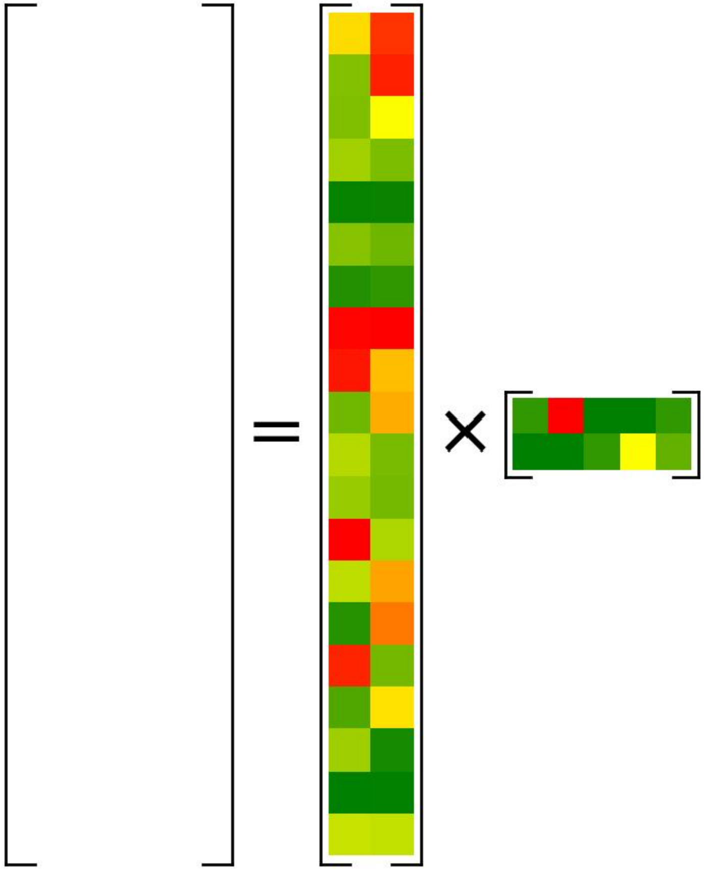
$$\boldsymbol{X} = \boldsymbol{W} \times \boldsymbol{H}$$
$$[20 \times 5] = [20 \times 2] \times [2 \times 5]$$

$\boldsymbol{X}$ – data matrix
[attributes × locations]

$\boldsymbol{W}$ – feature (signal) matrix
[attributes × signatures]

$\boldsymbol{H}$ – mixing matrix
[signatures × locations]

Nonnegative matrix factorization



$$\boldsymbol{X} = \boldsymbol{W} \times \boldsymbol{H}$$

$$[20 \times 5] = [20 \times 2] \times [2 \times 5]$$

$\boldsymbol{X}$ – data matrix

[attributes $\times$ locations]

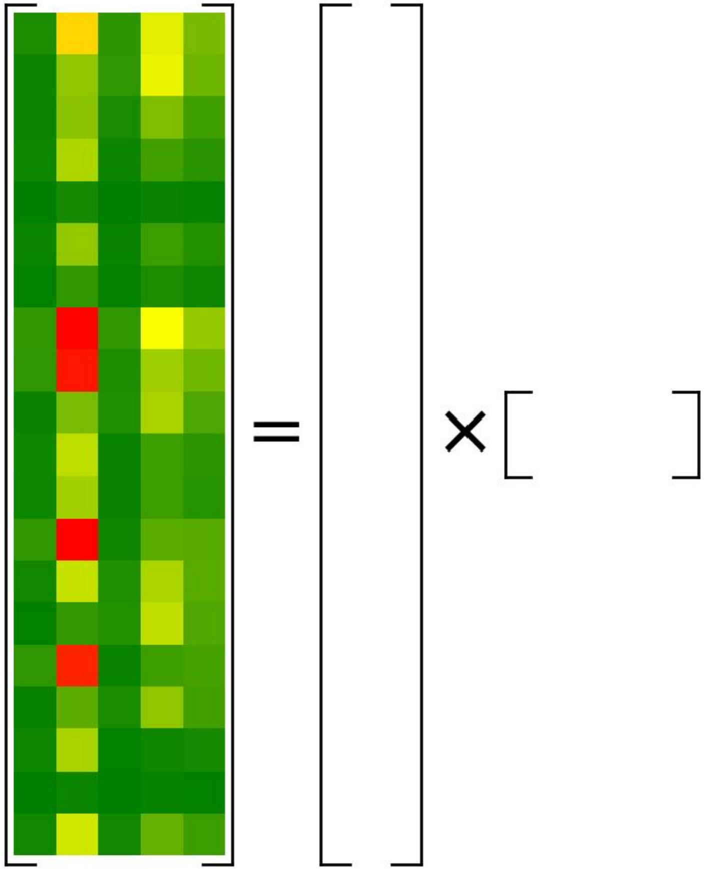
$\boldsymbol{W}$ – feature (signal) matrix

[attributes $\times$ signatures]

$\boldsymbol{H}$ – mixing matrix

[signatures $\times$ locations]

Nonnegative matrix factorization



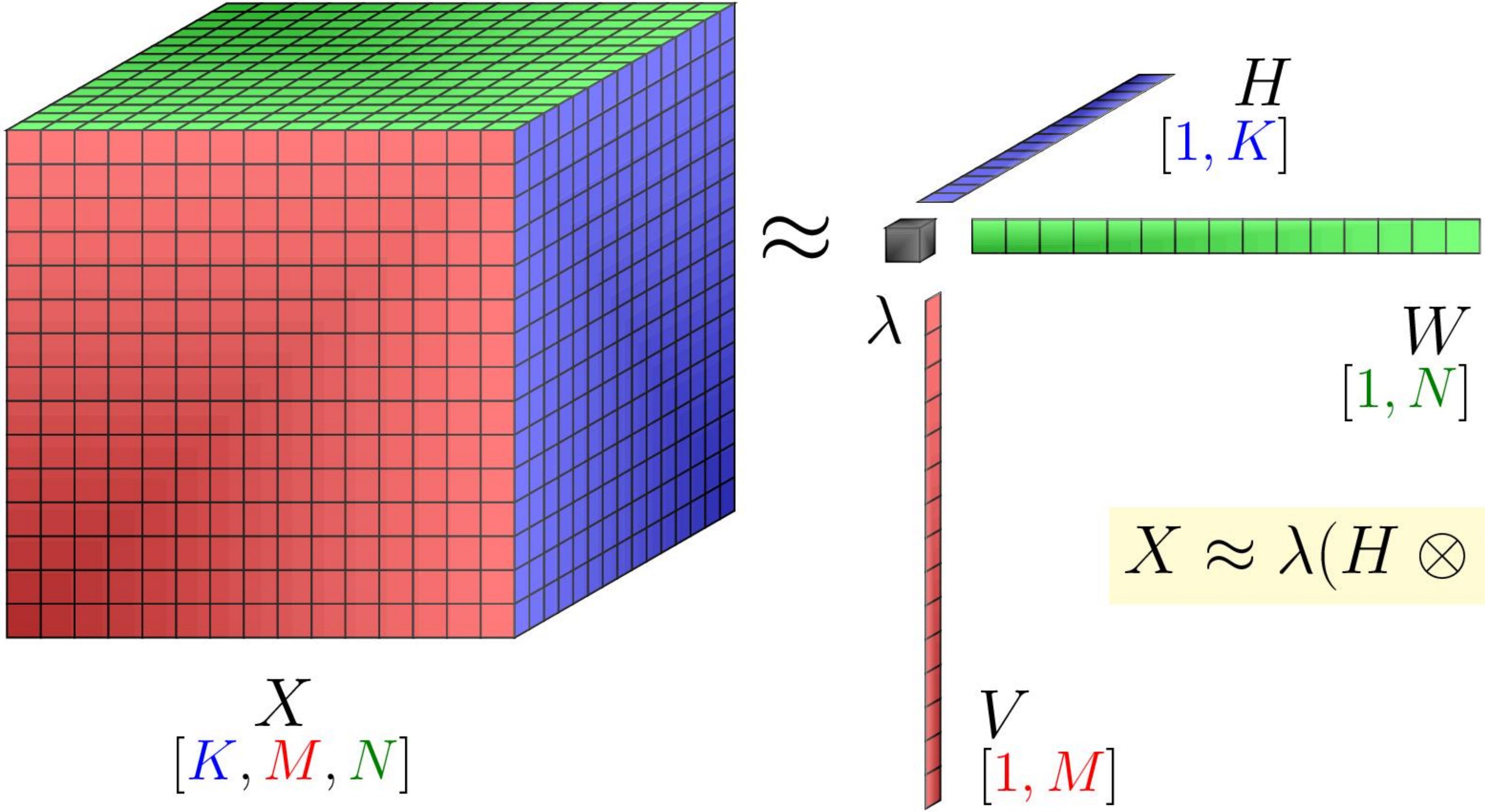
$$\boldsymbol{X} = \boldsymbol{W} \times \boldsymbol{H}$$
$$[20 \times 5] = [20 \times \textbf{?}] \times [\textbf{?} \times 5]$$

$\Rightarrow$ 100 **knowns**

$\Rightarrow$ **unknown** number of signatures (2 or more)

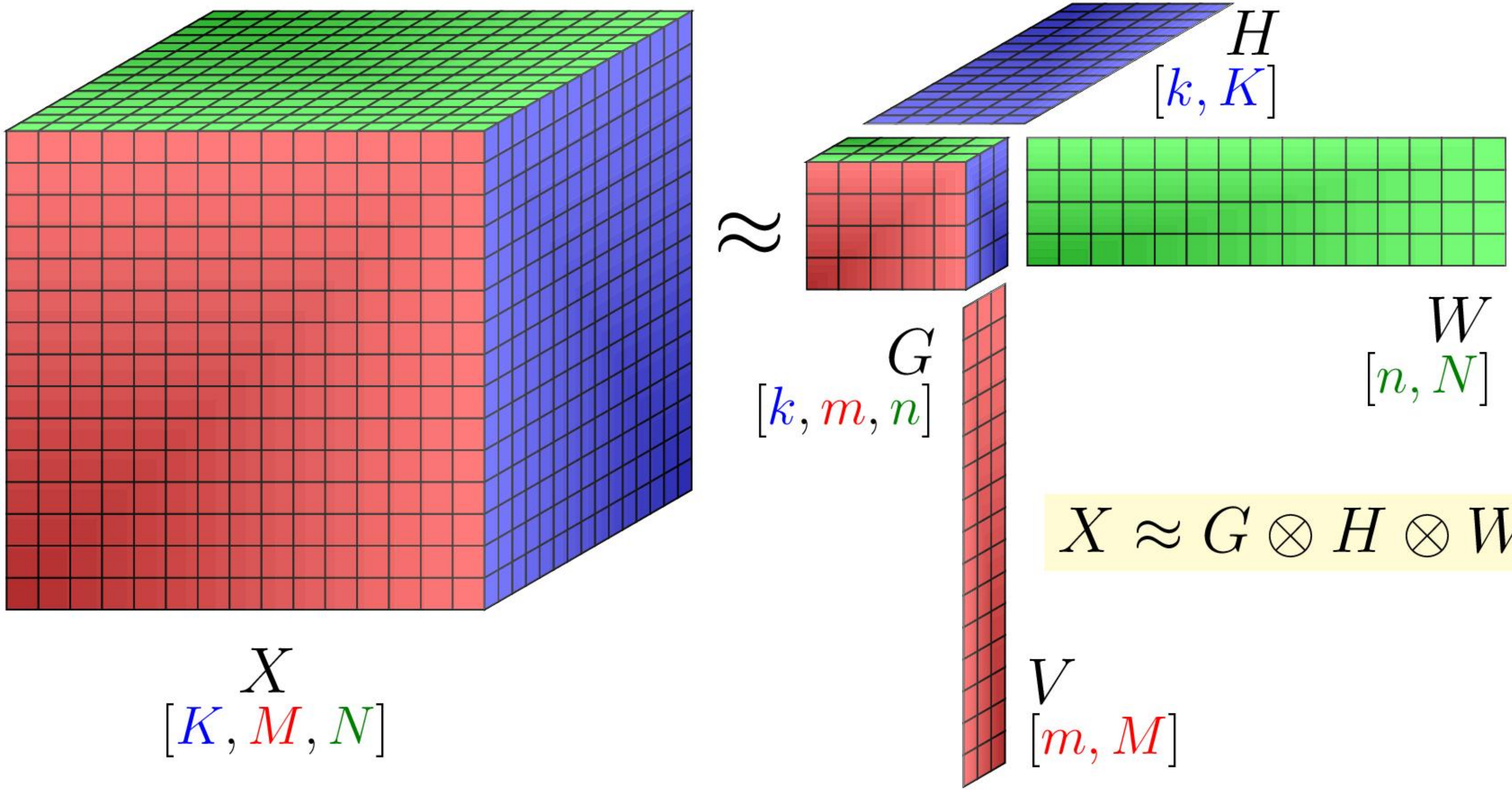
$\Rightarrow$ **unknown** matrix elements of $\boldsymbol{W}$ and $\boldsymbol{H}$ (50 or more)

Nonnegative tensor factorization



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

Nonnegative tensor factorization



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

Nonnegative tensor factorization