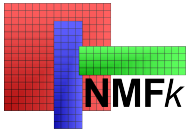
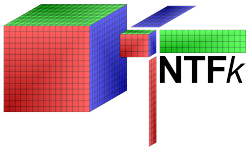
LA-UR-19-32569
Approved for public release; distribution is unlimited.

Title:	Machine learning analyses for characterization of oil, gas and water production from unconventional tight-rock reservoirs
Author(s):	Vesselinov, Velimir Valentinov
Intended for:	AGU Fall meeting, 2019-12-09 (San Francisco, New Mexico, United States)
Issued:	2019-12-17

Machine learning analyses for characterization of oil, gas and water production from unconventional tight-rock reservoirs

Velimir V. Vesselinov (monty) (vvv@lanl.gov)
Earth and Environmental Sciences Division
Los Alamos National Laboratory, NM, USA
http://tensors.lanl.gov





Machine Learning (ML) methods

- **Supervised** ML: learns everything from data
  ⇒ requires big training datasets
  ⇒ highly impacted by noise
  ⇒ cannot discover something that we do not know already

- **Physics-informed** ML: learns from data but includes preconceived knowledge about the governing processes
  ⇒ requires smaller training datasets
  ⇒ produces better predictability with lower uncertainty
  ⇒ robust to data noise

- **Unsupervised** ML: extracts features from data that can be applied for categorization and prediction
  ⇒ unbiased analyses not impacted by data labeling and physics assumptions

Machine Learning (ML) methods

- **Supervised** ML: learns everything from data
  ⇒ requires big training datasets
  ⇒ highly impacted by noise
  ⇒ cannot discover something that we do not know already

- **Physics-informed** ML: learns from data but includes preconceived knowledge about the governing processes
  ⇒ requires smaller training datasets
  ⇒ produces better predictability with lower uncertainty
  ⇒ robust to data noise

- **Unsupervised** ML: extracts features from data that can be applied for categorization and prediction
  ⇒ unbiased analyses not impacted by data labeling and physics assumptions

Machine Learning (ML) methods

- **Supervised** ML: learns everything from data
  ⇒ requires big training datasets
  ⇒ highly impacted by noise
  ⇒ cannot discover something that we do not know already

- **Physics-informed** ML: learns from data but includes preconceived knowledge about the governing processes
  ⇒ requires smaller training datasets
  ⇒ produces better predictability with lower uncertainty
  ⇒ robust to data noise

- **Unsupervised** ML: extracts features from data that can be applied for categorization and prediction
  ⇒ unbiased analyses not impacted by data labeling and physics assumptions

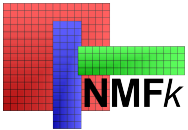
Unsupervised Machine Learning Applications: Data Analytics / Model Diagnostics

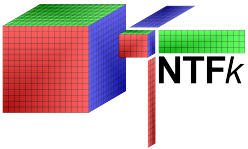
- Feature extraction (**FE**)
- Blind source separation (**BSS**)
- Detection of disruptions / anomalies
- Image recognition
- Separate physics processes
- Discover unknown dependencies and phenomena
- Develop reduced-order/surrogate models
- Identify dependencies between model inputs and outputs
- Guide development of physics models representing the data
- Make predictions
- Optimize data acquisition
- “Label” datasets for supervised ML analyses

Nonnegative Matrix/Tensor Factorization

- Novel LANL-patented, open-source, unsupervised Machine Learning (ML) methods and computational techniques
- Based in matrix/tensor factorization coupled with custom k-means clustering and nonnegativity/sparsity constraints:
  - NMFk: Nonnegative **Matrix** Factorization
  - NTFk: Nonnegative **Tensor** Factorization
  - `https://github.com/TensorDecompositions`
- Capable to efficiently process large datasets (TB's) utilizing GPU's, TPU's & FPGA's

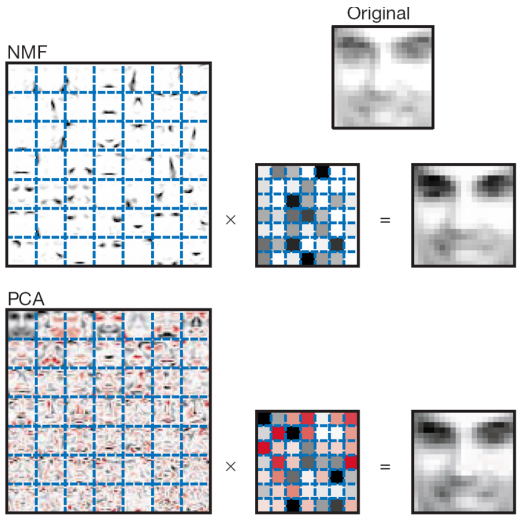
  $\Rightarrow$ julia, Flux.jl, Knet.jl, AutoOffLoad.jl, Zygote.jl, TensorFlow, PyTorch, MXNet





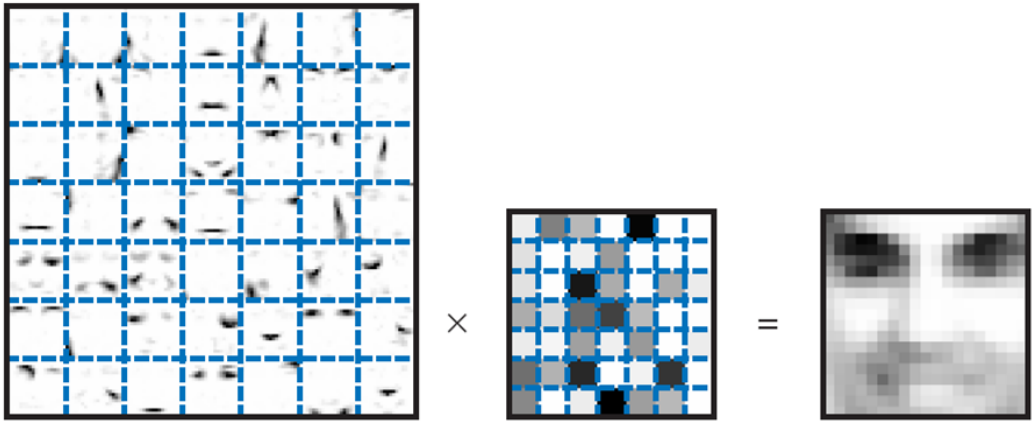
Why nonnegativity?

- NMF vs PCA (Lee & Seung, 1999)
- NMF: Nonnegative Matrix Factorization
- PCA: Principal Component Analysis

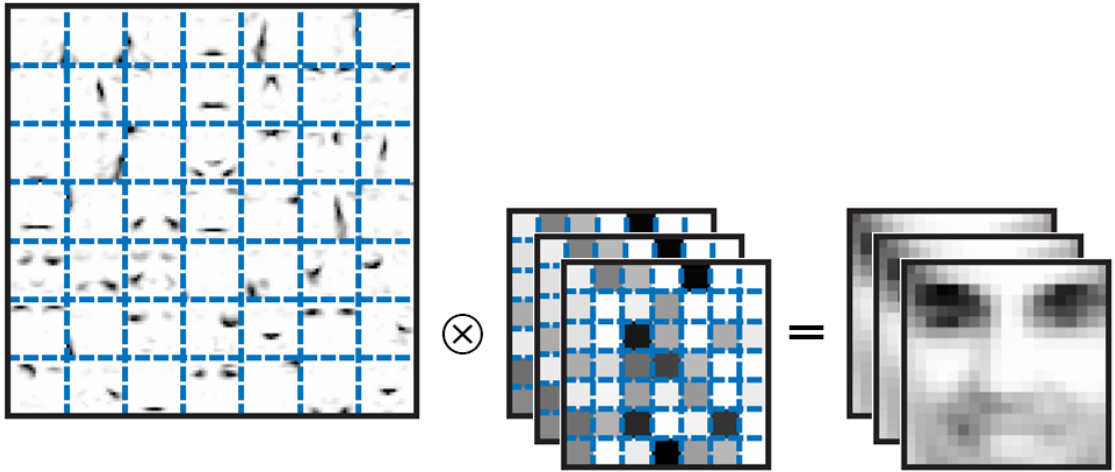


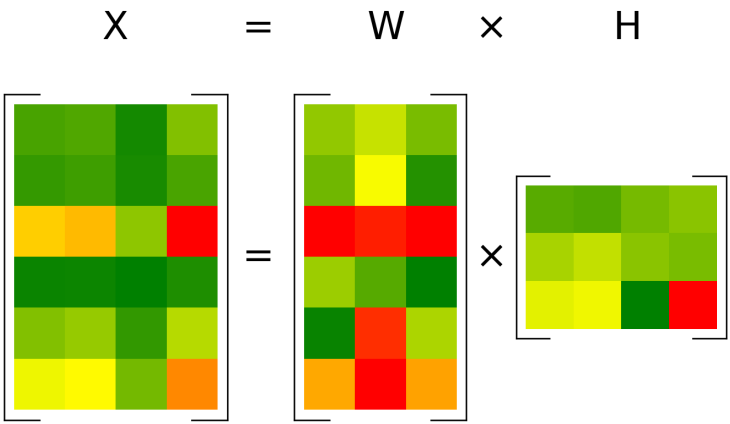
Nonnegativity constraints provide meaningful and interpretable results (+sparsity)

NMF: Nonnegative Matrix Factorization



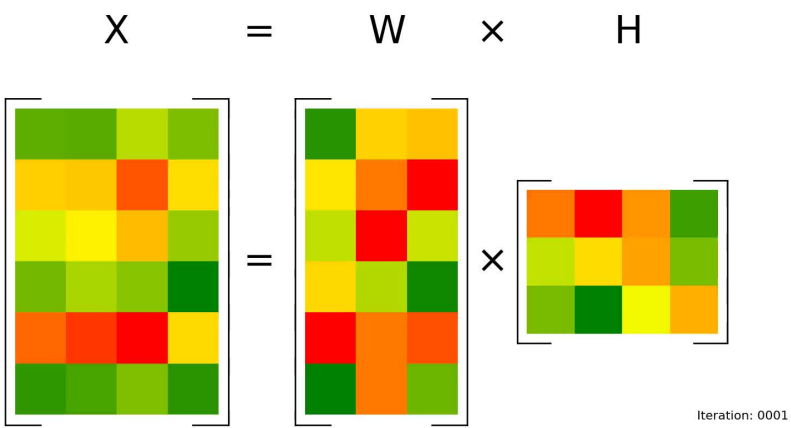
NTF: Nonnegative Tensor Factorization



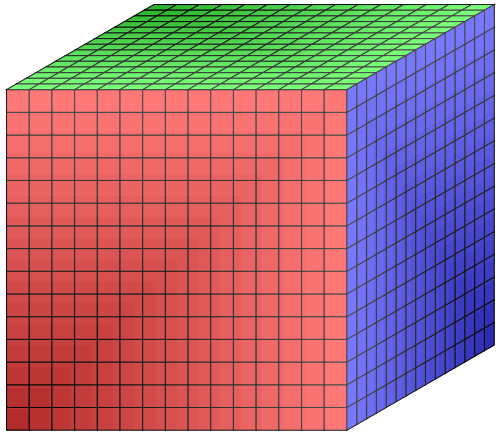


24 knowns $(6 \times 4) \rightarrow 30$ unknowns $(6 \times 3) + (3 \times 4)$
number of features k is also unknown (here, $k = 3$))

NMFk: Factorization process



Tensor Decomposition (3D case): Rank-1 tensor



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

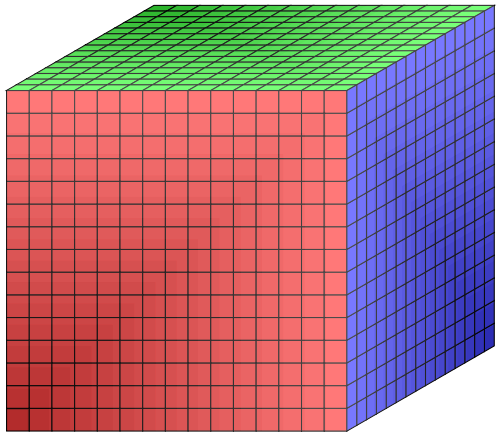
X
$[K, M, N]$

H
$[1, K]$

W
$[1, N]$

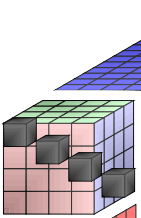
V
$[1, M]$

Tensor Decomposition (3D case): Rank-4 tensor

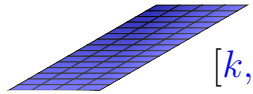


X
$[K, M, N]$

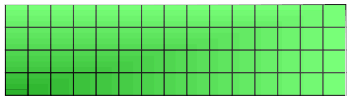
$\approx$



D
$[k, k, k]$



H
$[k, K]$

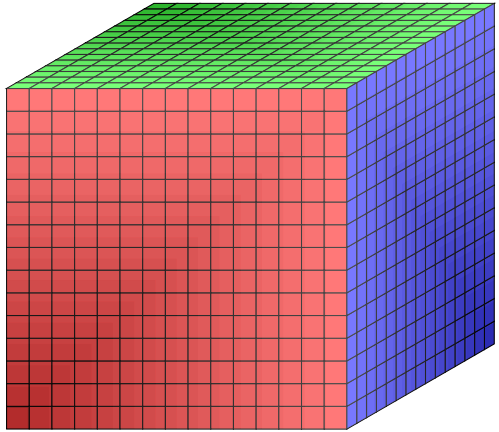


W
$[k, N]$

V
$[k, M]$

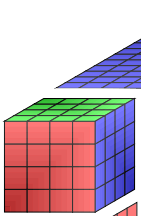
$$X \approx D \otimes H \otimes W \otimes V$$

Tensor Decomposition (3D case): Multirank-(4,4,4) tensor

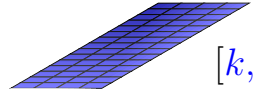


X
$[K, M, N]$

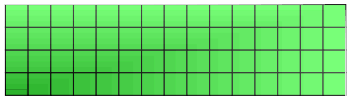
$\approx$



G
$[k, k, k]$



H
$[k, K]$

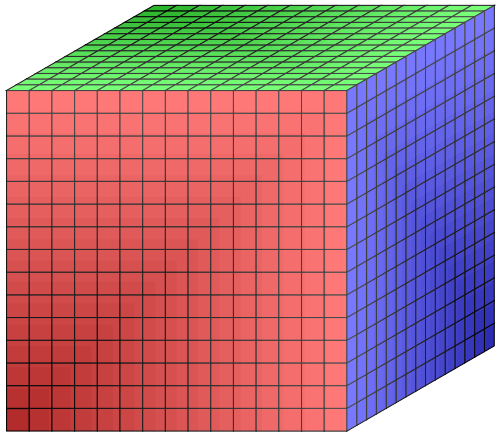


W
$[k, N]$

V
$[k, M]$

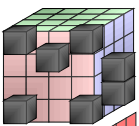
$$X \approx G \otimes H \otimes W \otimes V$$

Tucker Tensor Decomposition (3D case): Tucker-3 Multirank-(3,2,4)

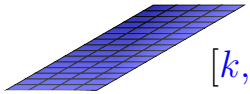


$$X \\ [K, M, N]$$

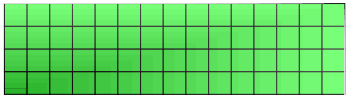
$$\approx$$



$$D \\ [k, k, k]$$



$$H \\ [k, K]$$



$$W \\ [k, N]$$

$$V \\ [k, M]$$

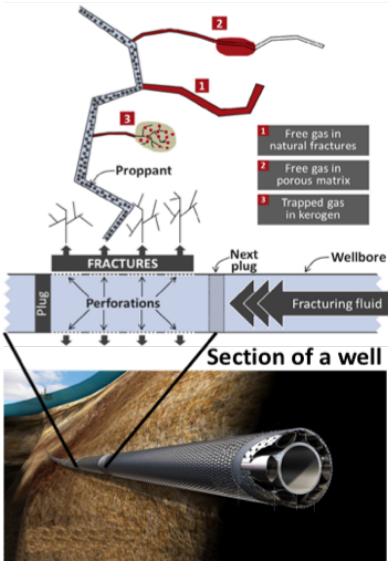
$$X \approx D \otimes H \otimes W \otimes V$$

- **Field Data**:
  - Contamination
  - Climate
  - Geothermal
  - Seismic
  - Oil/gas production
- **Lab Data**:
  - X-ray Spectroscopy
  - UV Fluorescence Spectroscopy
  - Microbial population analyses
  - Isotope fractionation
- **Operational Data**:
  - LANSCE: Los Alamos Neutron Accelerator
  - Oil/gas production
- **Model Outputs**:
  - Reactive mixing $A + B \rightarrow C$
  - Phase separation of co-polymers
  - Molecular Dynamics of proteins
  - Climate modeling

- Vesselinov, Munuduru, Karra, O'Maley, Alexandrov, Unsupervised Machine Learning Based on Non-Negative Tensor Factorization for Analyzing Reactive-Mixing, **Journal of Computational Physics**, Special issue: Machine Learning, 2019.
- Stanev, Vesselinov, Kusne, Antoszewski, Takeuchi, Alexandrov, Unsupervised Phase Mapping of X-ray Diffraction Data by Nonnegative Matrix Factorization Integrated with Custom Clustering, **Nature Computational Materials**, 2018.
- Vesselinov, O'Malley, Alexandrov, Nonnegative Tensor Factorization for Contaminant Source Identification, **Journal of Contaminant Hydrology**, 2018.
- O'Malley, Vesselinov, Alexandrov, Alexandrov, Nonnegative/binary matrix factorization with a D-Wave quantum annealer, **PLOS ONE**, 2018.
- Vesselinov, O'Malley, Alexandrov, Contaminant source identification using semi-supervised machine learning, **Journal of Contaminant Hydrology**, 2017.
- Alexandrov, Vesselinov, Blind source separation for groundwater level analysis based on nonnegative matrix factorization, **WRR**, 2014.

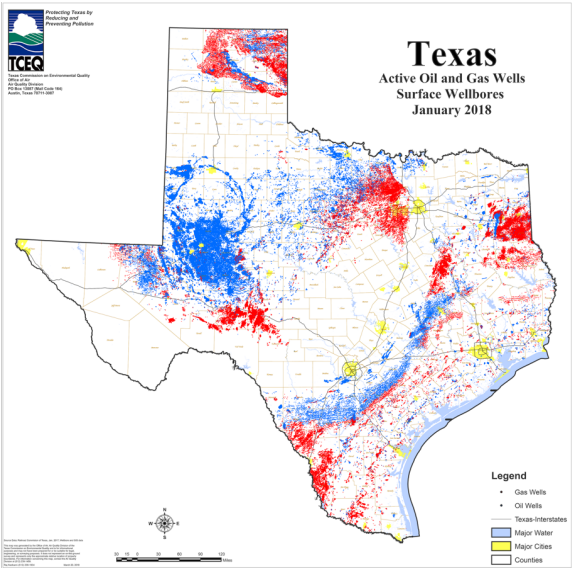
Oil/Gas Production

- Oil/Gas production from unconventional reservoirs extracts a small portion of the available resources (<10%)
- Oil/Gas production is challenging to predict and optimize
- Physics processes during well development (including hydrofracking) and extraction are poorly understood and challenging to simulate
- Alternative is to learn to predict system behavior based on the observed oil/gas production at existing wells

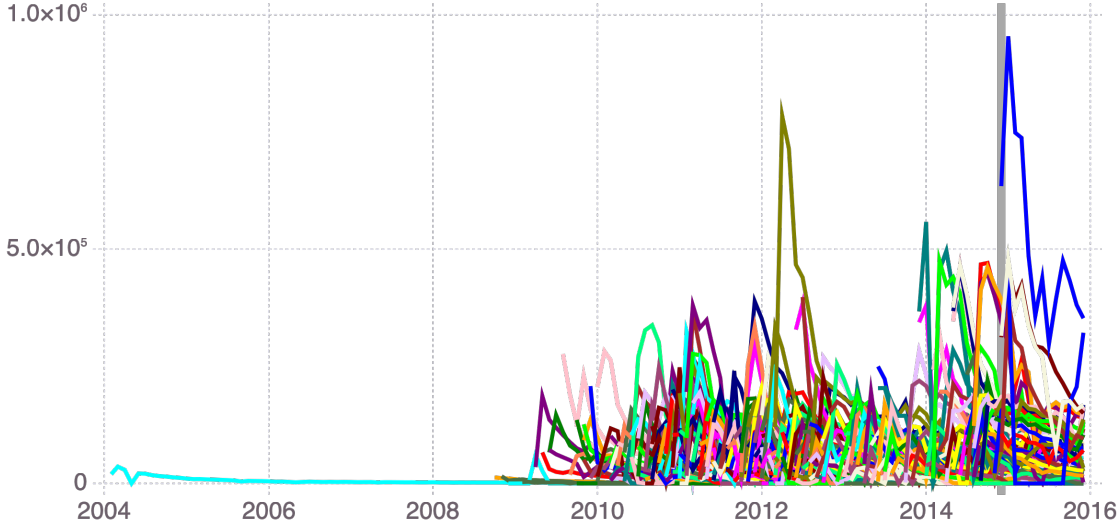


Oil/Gas Production Data

- Large public datasets are available representing unconventional oil and gas production (U.S. and world wide)
- Data represent monthly production rates (oil, gas, water) + many other well attributes
- $\sim$ 2,000,000 wells in U.S.
- > 300,000 wells in Texas
- > 20,000 wells in Eagle Ford Shale Play
- 327 gas wells in Eagle Ford Shale Play selected for preliminary analyses



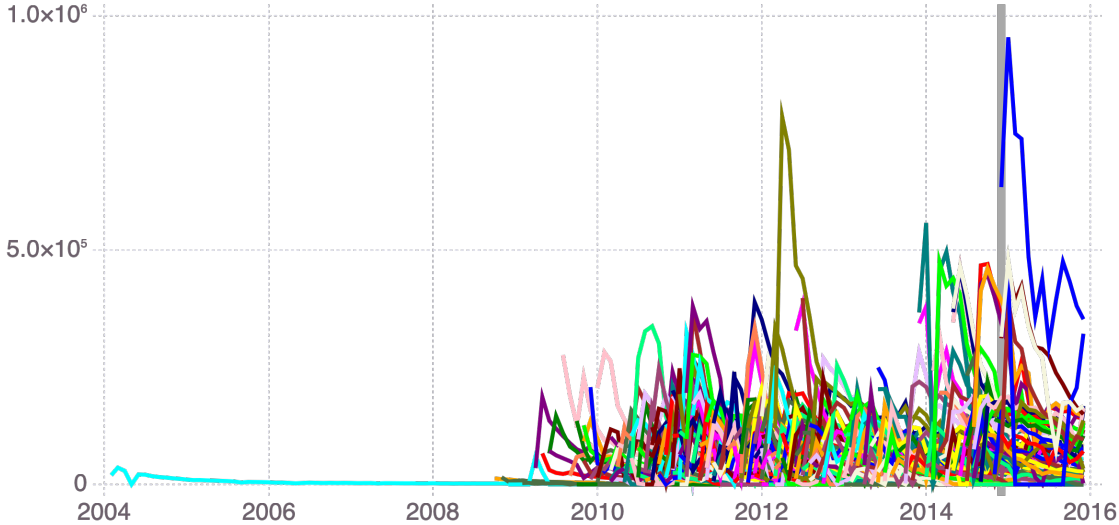
Eagle Ford Shale Play: Monthly production volumes [MCF] of 327 gas wells



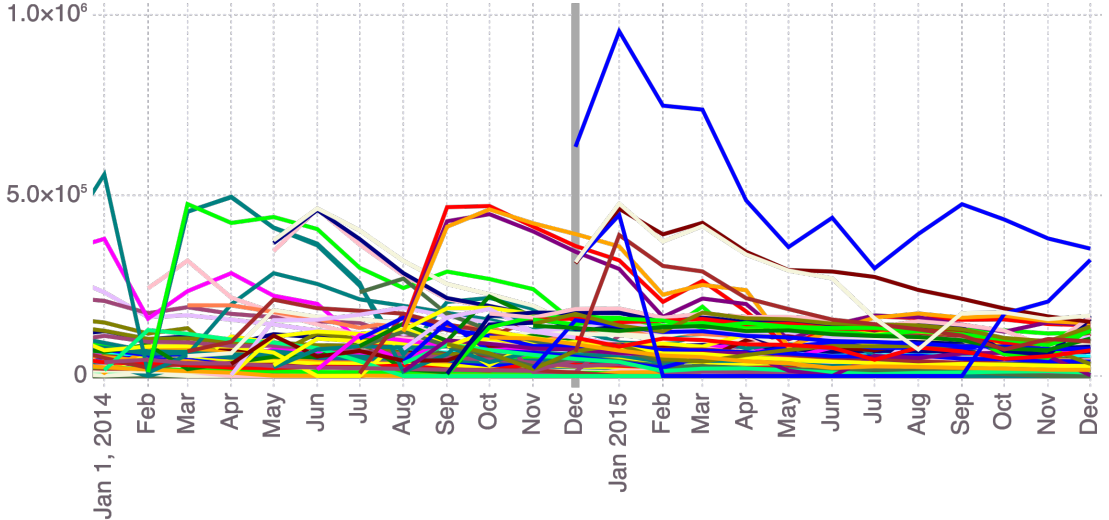
ML analysis of Oil/Gas Production Data

- Use all the data up to a given cutoff date (e.g. 2015)
- Apply ML to learn behavior of the “known“ well transients
  - Identify and group wells which behave similarly (having similar production transients)
  - Discover the optimal number of **master decline curves** required to represent the observed transients
  - **master decline curves** = production **features** or **signatures**
- Apply ML to predict **blindly** the unknown production transients beyond the cutoff
- Prediction is obtained by discovering to which type (group) the wells producing beyond the cutoff belong
- i.e., discovering what combinations of the **master decline curves** can represent the wells producing beyond the cutoff
- ML analyses performed using **NMFk**/**NTFk**

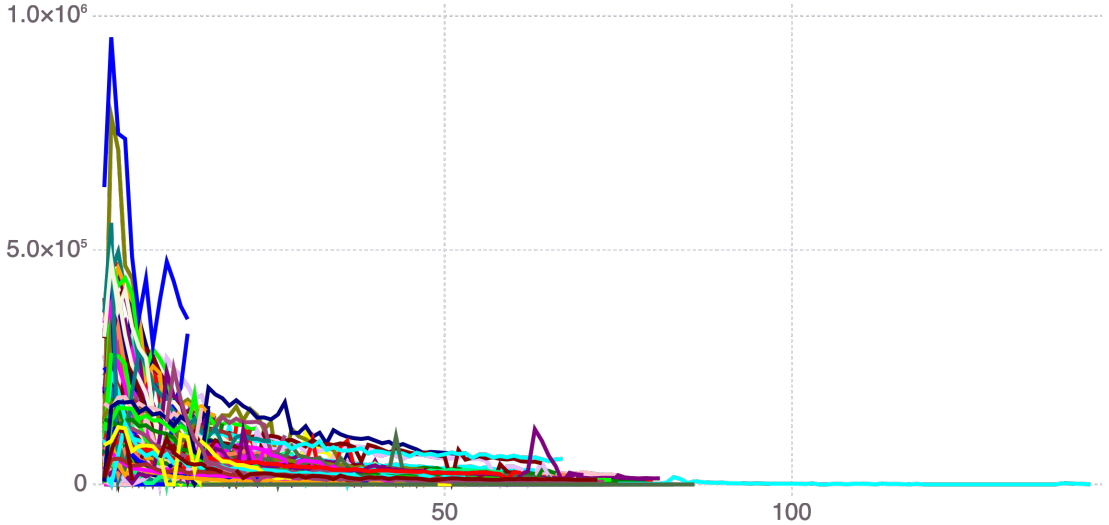
Eagle Ford Shale Play: Monthly production volumes [MCF] of 327 gas wells



Eagle Ford Shale Play: Monthly production volumes [MCF] of 327 gas wells

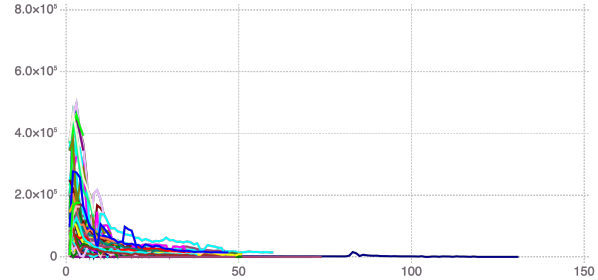


Eagle Ford Shale Play: Monthly production volumes [MCF] of 327 gas wells

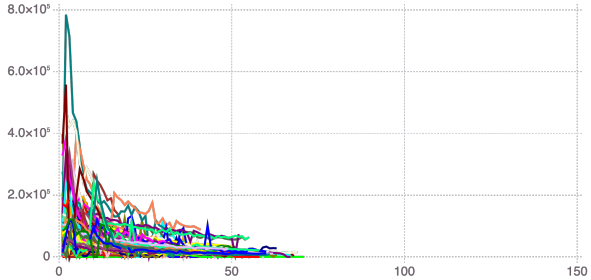


Eagle Ford Shale Play: Wells split into 2 groups

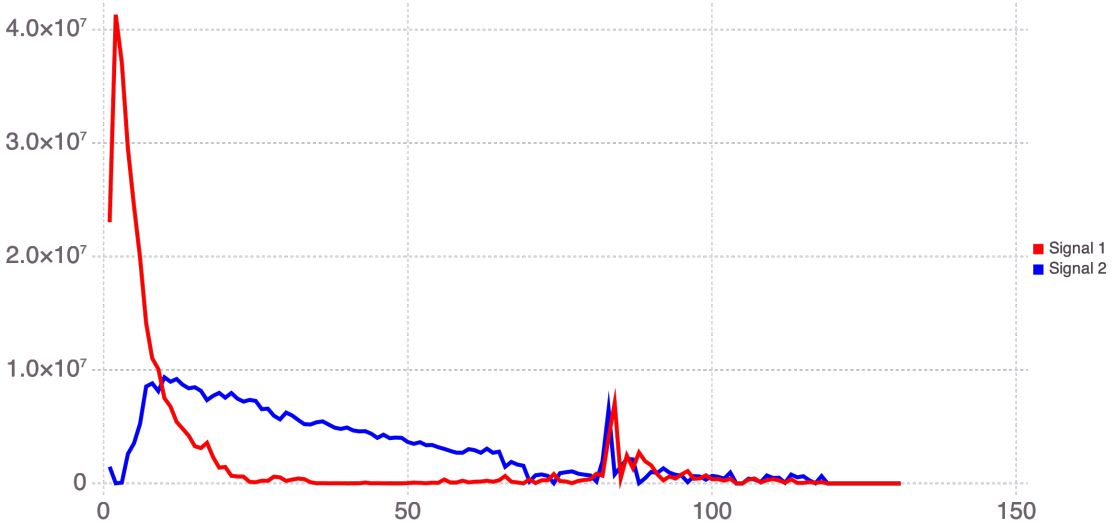
‘Fast” declining (135)



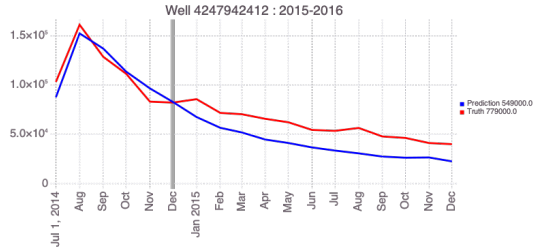
‘Slow” declining (192)

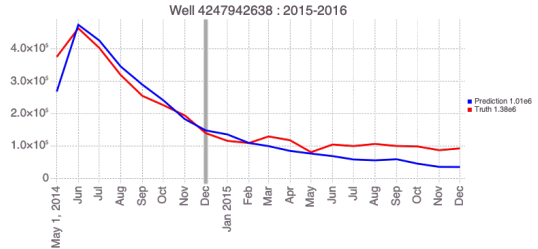


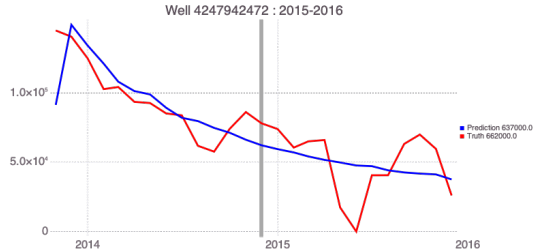
Eagle Ford Shale Play: Master Decline Curves [MCF over months]

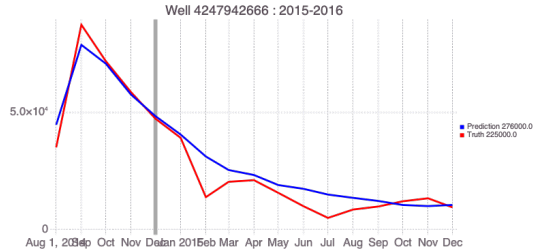


Eagle Ford Shale Play: Blind predictions beyond 2015

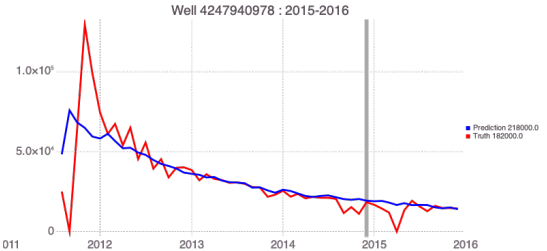


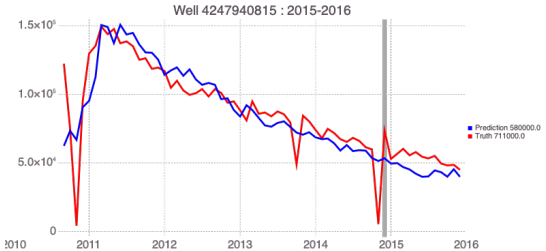


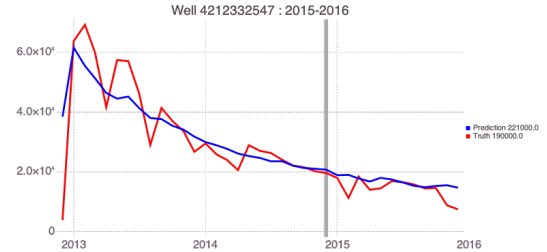


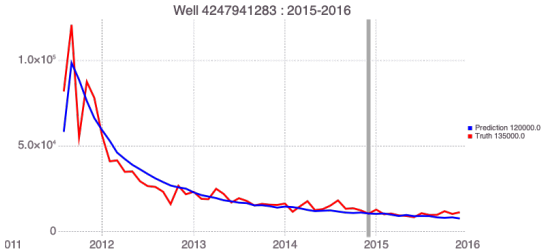


Eagle Ford Shale Play: Blind predictions beyond 2015



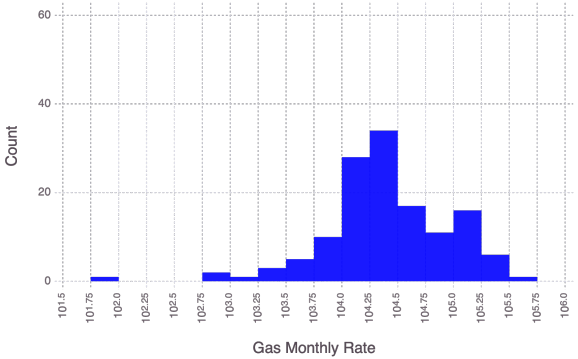




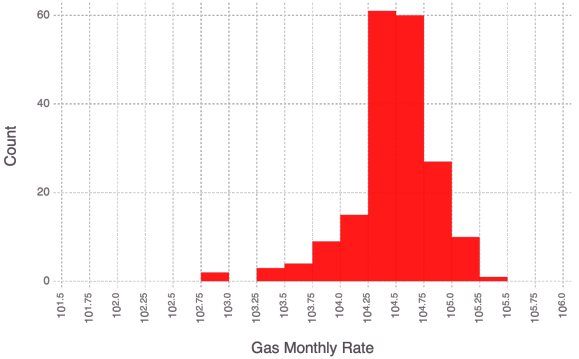


Monthly rate histograms

‘Fast” declining (135)

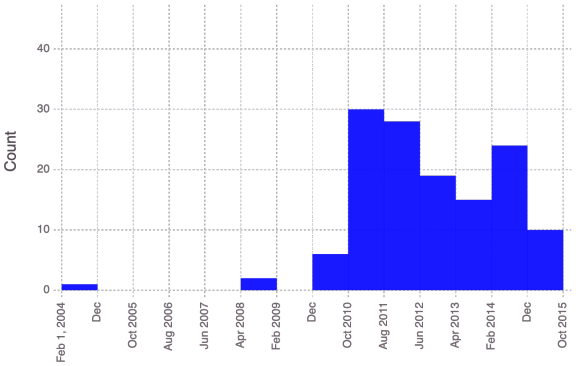


‘Slow” declining (192)

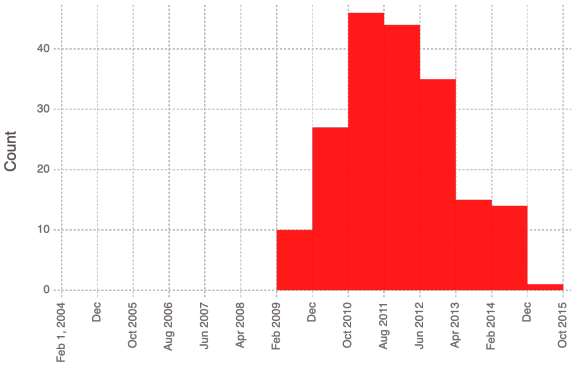


Drilling date histograms

‘Fast” declining (135)



‘Slow” declining (192)

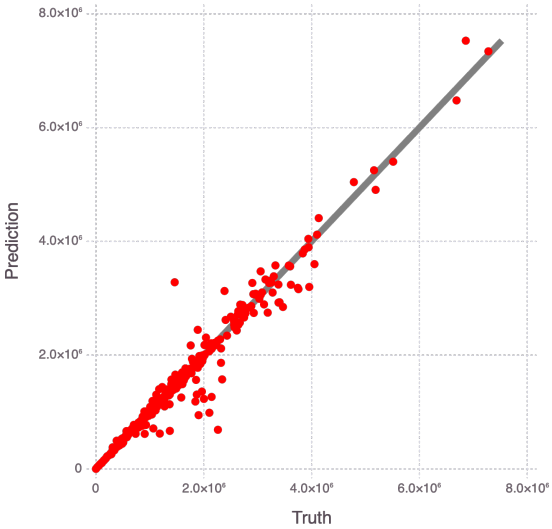


Eagle Ford Shale Play: Wells split into 2 groups

- Other well attributes also differ between the 2 groups
- For example:
  - Operators
  - Proppant mass
  - Injected fluid volumes
  - ... work in progress

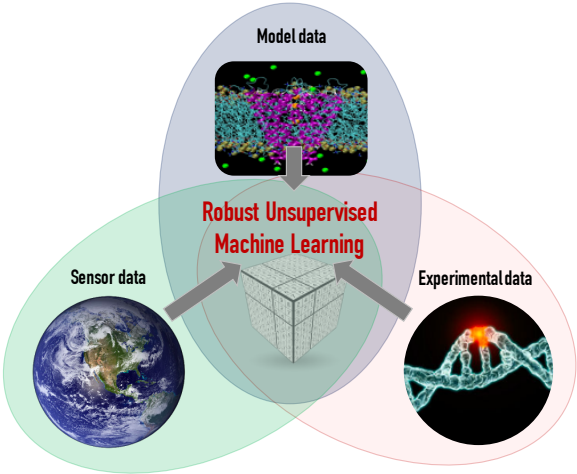
Eagle Ford Shale Play: Blind predictions beyond 2015

- 300 wells continue producing beyond 2015
- $r^2 = 0.96$



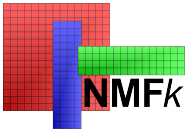
Summary

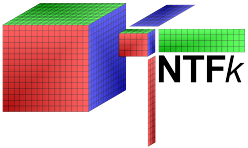
- Developed **novel** unsupervised and physics-informed ML methods and computational tools
- Our ML methods have been used to solve various real-world problems (brought breakthrough discoveries related to human cancer research)
- julia



- Codes:

NMFk MADS NTFk





- Examples:

http://madsjulia.github.io/Mads.jl/Examples/blind_source_separation
http://tensors.lanl.gov
http://tensordecompositions.github.io
https://github.com/TensorDecompositions
https://hub.docker.com/u/montyvesselinov

julia