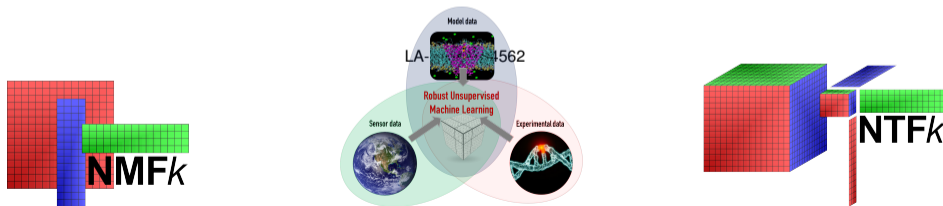


Physics-Informed Machine Learning Methods for Data Analytics and Model Diagnostics

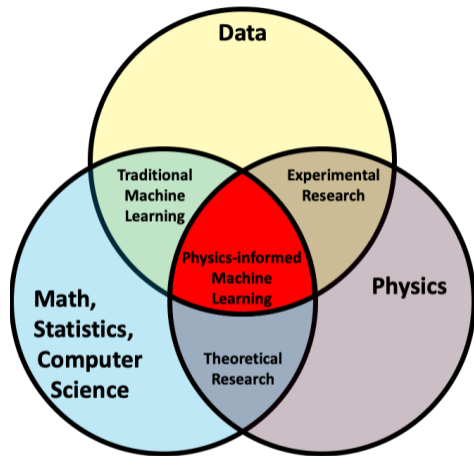
Velimir V. Vesselinov (monty) (vvv@lanl.gov)

Earth and Environmental Sciences Division, Los Alamos National Laboratory, NM, USA

<http://tensors.lanl.gov>



- ▶ We have developed a series of novel **Machine Learning** (ML) methods
- ▶ **Data Analytics**: Identify features (signals) in datasets
- ▶ **Model Analytics**: Identify processes (features) in model outputs
- ▶ **Physics-Informed Machine Learning**: Coupled Data/Model Analytics



- ▶ Extracting common **hidden features** present in data (observations/experiments) and model outputs
- ▶ Incorporating **physics constraints** in ML analyses
- ▶ Incorporating **physics processes** in ML analyses

- ▶ **PIML** is of critical importance for data- and physics-driven science and engineering applications
- ▶ **PIML** models can replace computationally intensive numerical simulators
- ▶ **PIML** models can be applied to predict complex processes when:
 - physics is not well understood
 - numerical models cannot be developed (due to physics complexity)
- ▶ Training **PIML** models is 1-2 orders of magnitude faster than traditional ML models

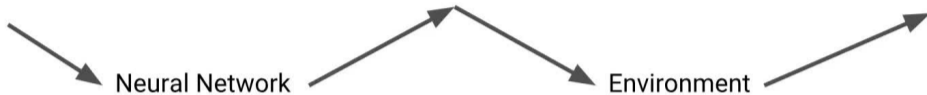


- ▶ **“Deep Learning is dead. Long Live Differentiable Programming!”**
Yann LeCun, Chief AI Scientist at Facebook, and one of the leading ML experts in the world
- ▶ **Differentiable Programming** for PIML can be done efficiently only in **julia**
C/C++, FORTRAN, Python, scikit-learn, TensorFlow, PyTorch cannot do it

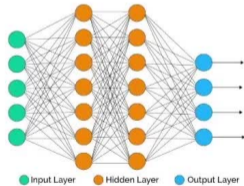
CartPole State

Control Parameters

Loss



$angle = -3^\circ$
 $velocity = 0.5^\circ/s$



$\{left, right\}$



$angle^2$

▶ **Field Data:**

- ▶ Groundwater contamination
- ▶ US Climate data
- ▶ Geothermal data
- ▶ Seismic data

▶ **Lab Data:**

- ▶ X-ray Spectroscopy
- ▶ UV Fluorescence Spectroscopy
- ▶ Microbial population analyses

▶ **Operational Data:**

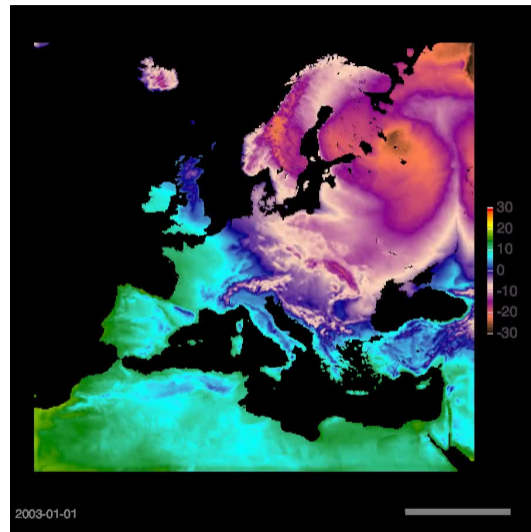
- ▶ LANSCE: Los Alamos Neutron Accelerator
- ▶ Oil/gas production

▶ **Model Outputs:**

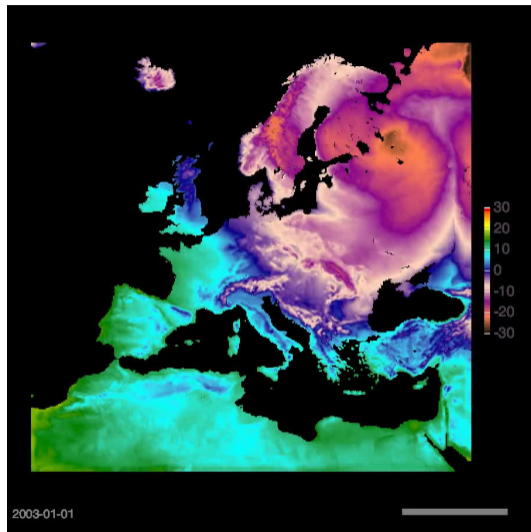
- ▶ Reactive mixing $A + B \rightarrow C$
- ▶ Phase separation of co-polymers
- ▶ Molecular Dynamics of proteins
- ▶ EU Climate modeling

- ▶ Vesselinov, Munuduru, Karra, O'Maley, Alexandrov, Unsupervised Machine Learning Based on Non-Negative Tensor Factorization for Analyzing Reactive-Mixing, **Journal of Computational Physics**, (in review), 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**, 10.1016/j.jconhyd.2017.11.002, 2017.
- ▶ Alexandrov, Vesselinov, Blind source separation for groundwater level analysis based on nonnegative matrix factorization, **WRR**, 10.1002/2013WR015037, 2014.

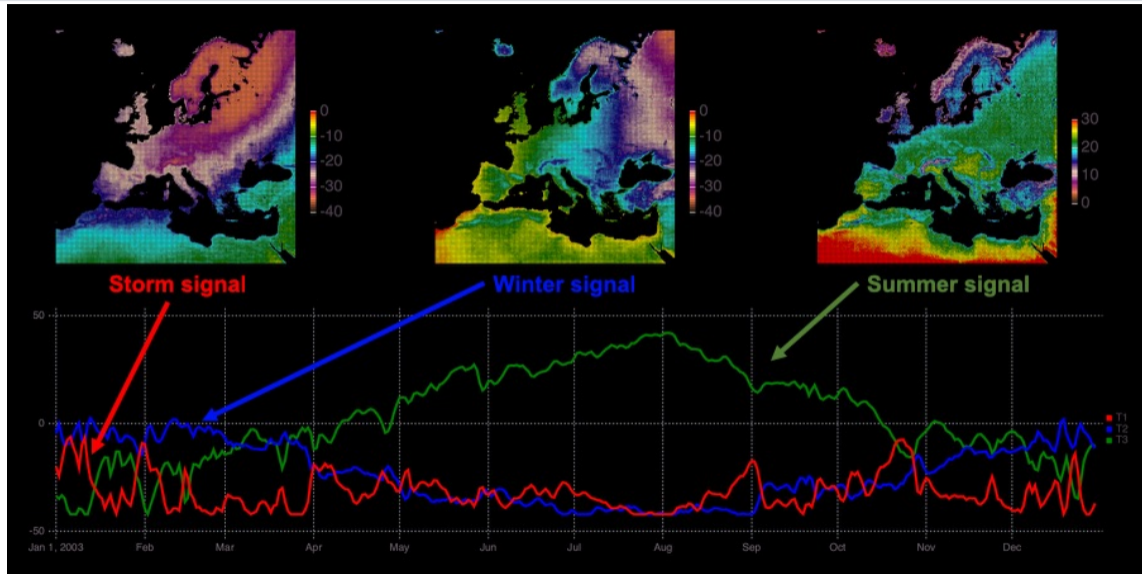
- ▶ **TerrSysMP**: coupled climate simulator (CLM/ParFlow) of atmospheric, surface and subsurface processes



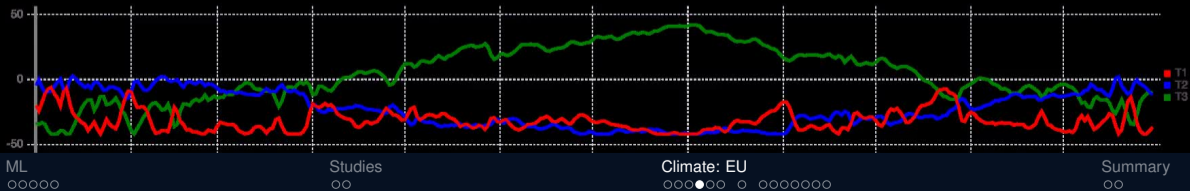
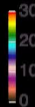
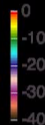
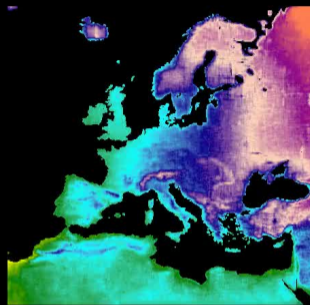
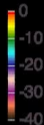
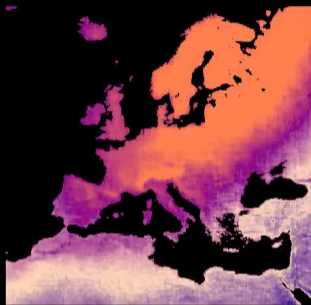
- ▶ Daily fluctuations in the air temperature [$^{\circ}C$]
- ▶ Tensor: $(424 \times 412 \times 365)$
(*columns* \times *rows* \times *days*)
- ▶ **NTF $_k$** applied to extract dominant hidden (latent) features



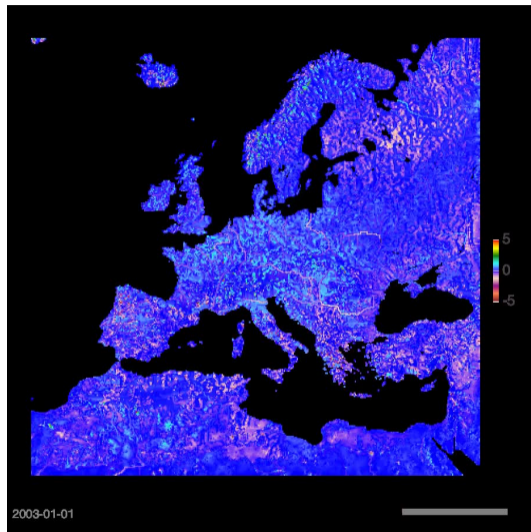
Climate model of Europe: 2003 temperature fluctuations represented by 3 features



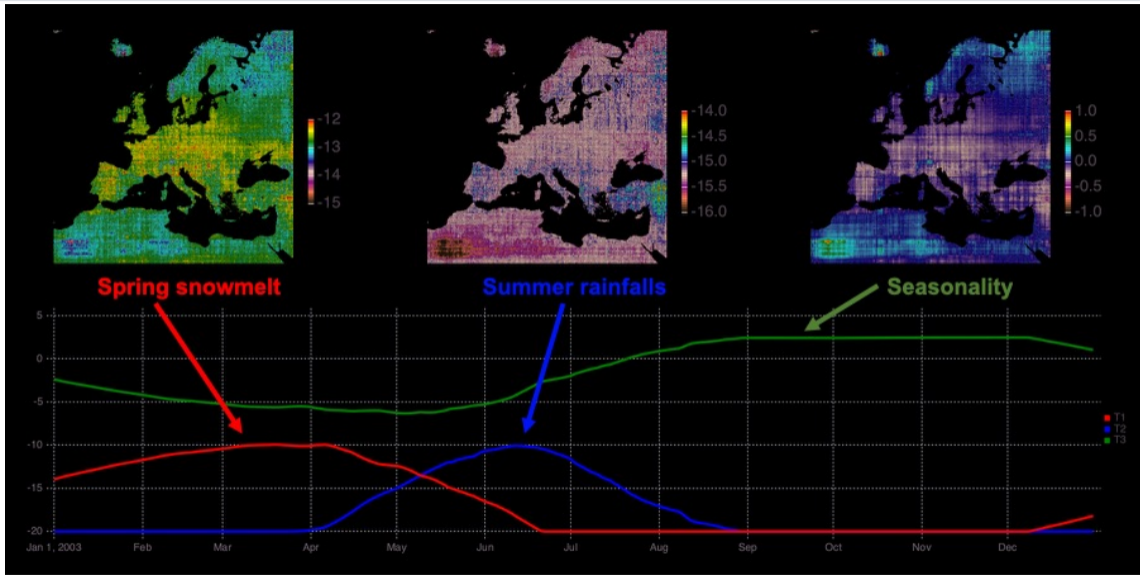
Climate model of Europe: 2003 temperature fluctuations represented by 3 features



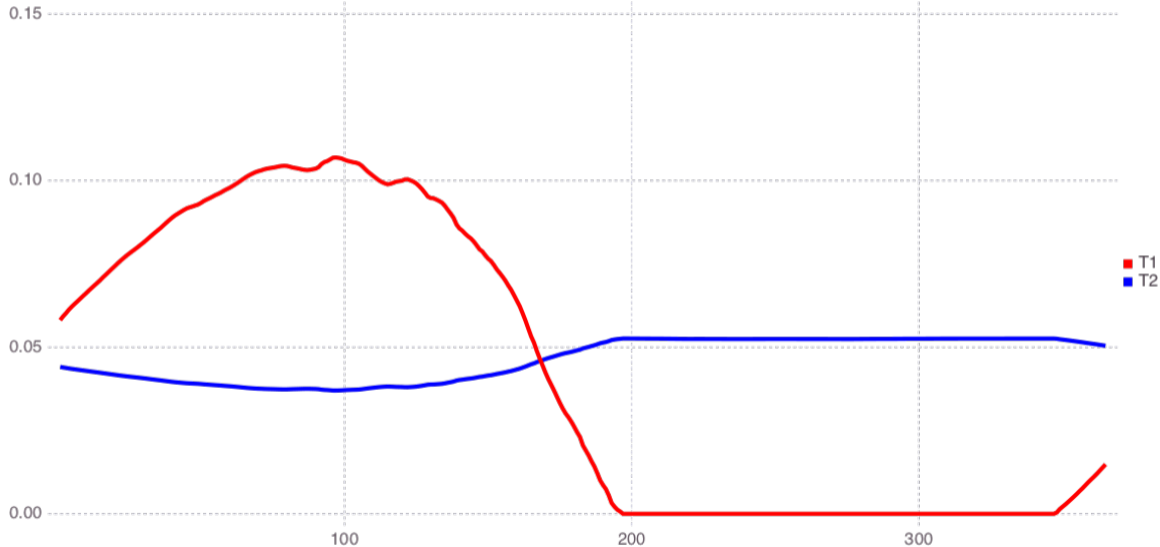
- ▶ Daily fluctuations in the water-table depth [m]
- ▶ Tensor: $(424 \times 412 \times 365)$
(*columns* \times *rows* \times *days*)
- ▶ **NTF $_k$** applied to extract dominant hidden (latent) features



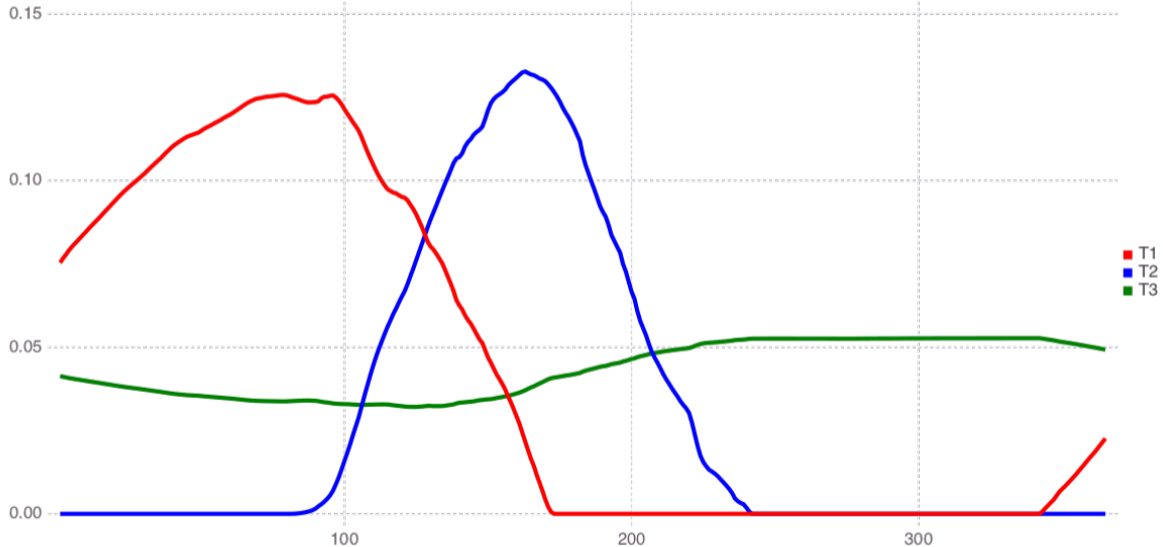
Climate model of Europe: 2003 water-table fluctuations represented by 3 features



Climate model of Europe: Water-table fluctuations represented by 2 features



Climate model of Europe: Water-table fluctuations represented by 3 features



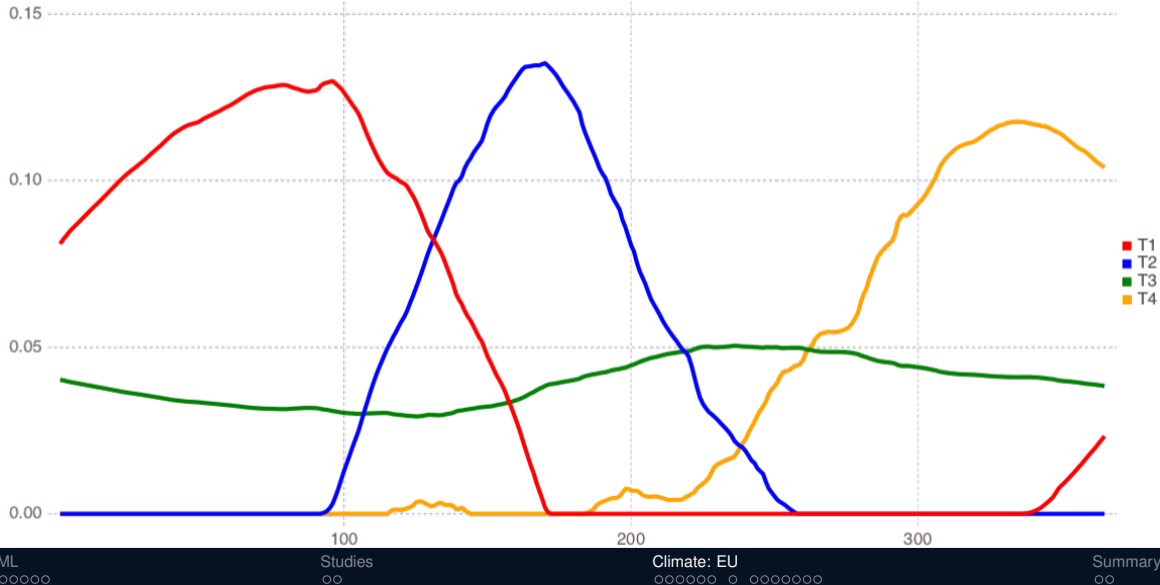
ML
○○○○○

Studies
○○

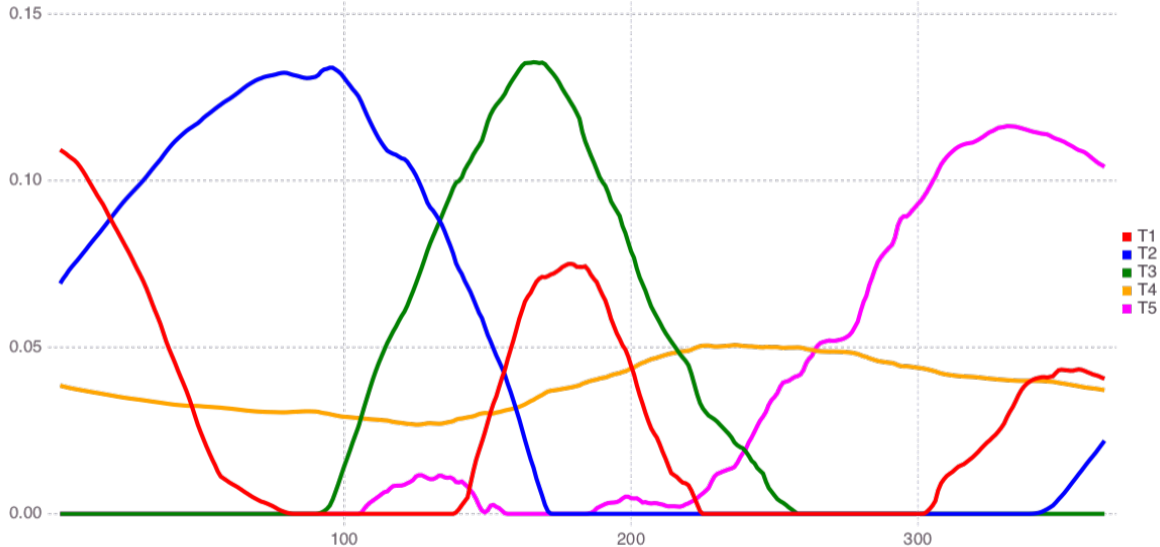
Climate: EU
○○○○○○○ ○ ○○○○○○○

Summary
○○

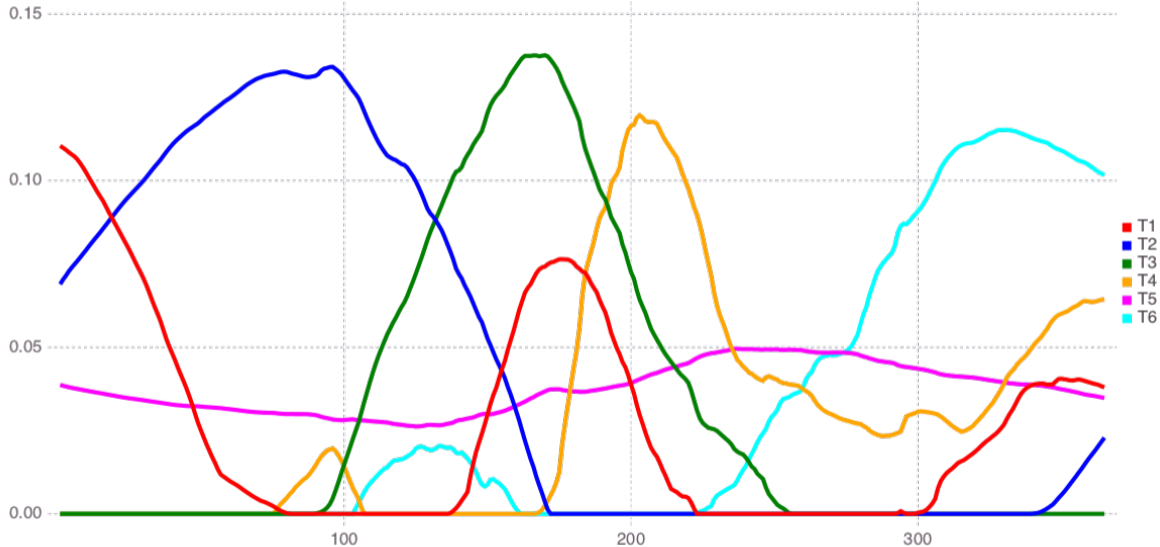
Climate model of Europe: Water-table fluctuations represented by 4 features



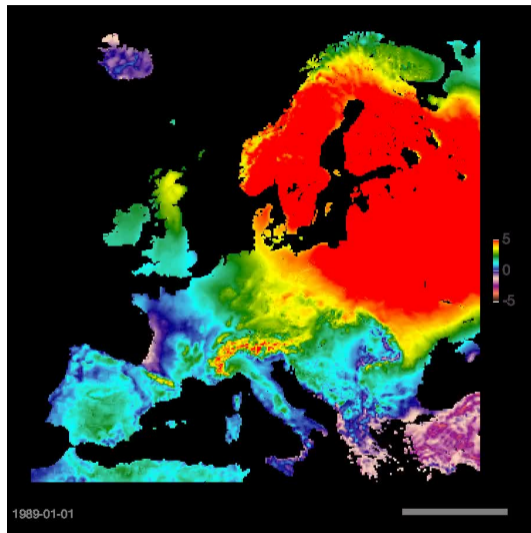
Climate model of Europe: Water-table fluctuations represented by 5 features



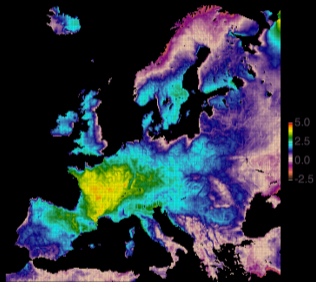
Climate model of Europe: Water-table fluctuations represented by 6 features



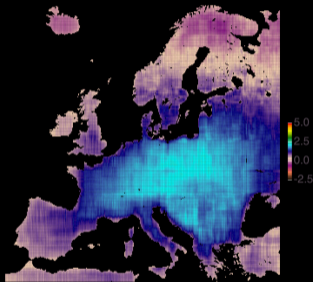
- ▶ Monthly fluctuations in the air temperature from 1989 to 2017 [$^{\circ}C$]
- ▶ Tensor: $(316 \times 316 \times 348)$
(*columns* \times *rows* \times *months*)
- ▶ **NTF $_k$** applied to extract dominant hidden (latent) features



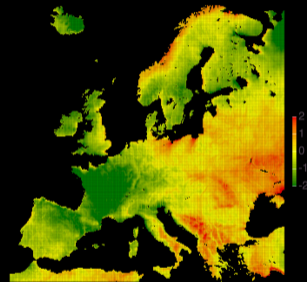
Original



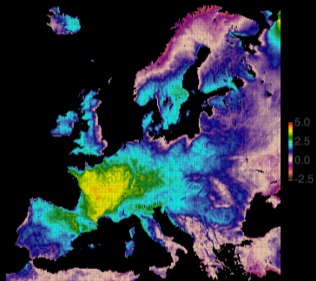
Reconstruction



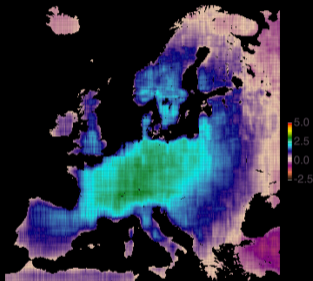
Error



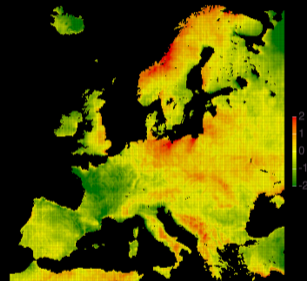
Original



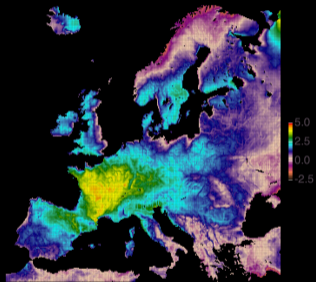
Reconstruction



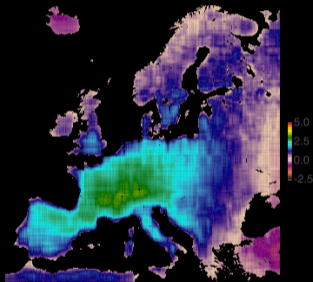
Error



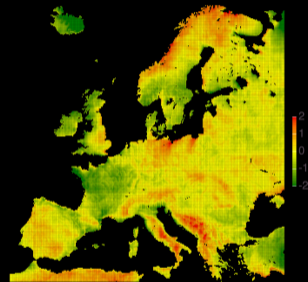
Original



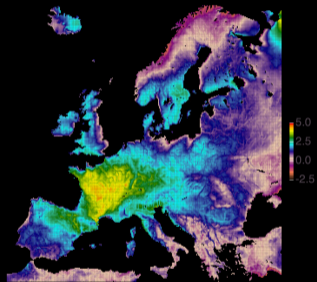
Reconstruction



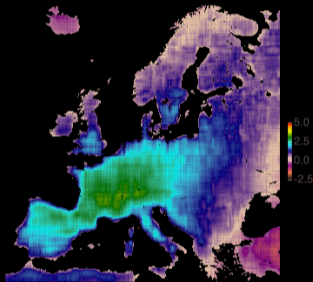
Error



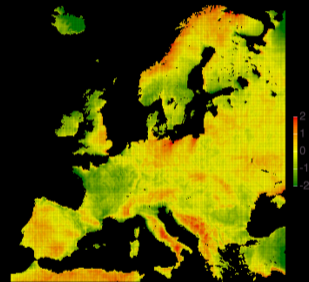
Original



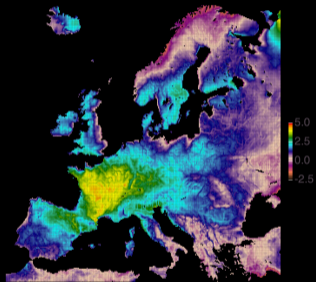
Reconstruction



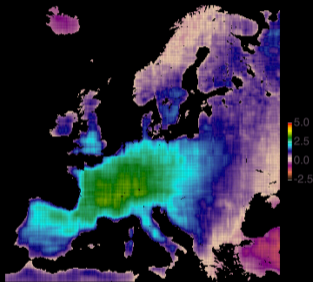
Error



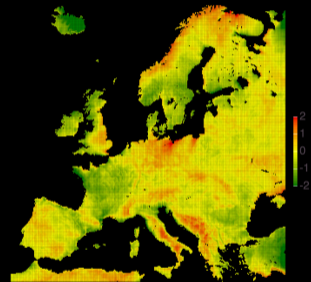
Original



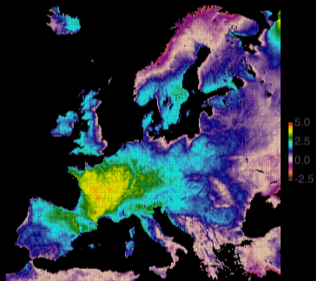
Reconstruction



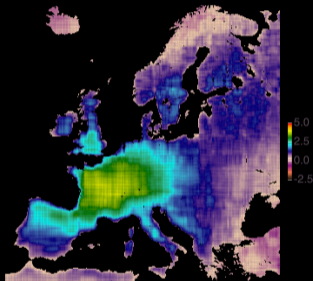
Error



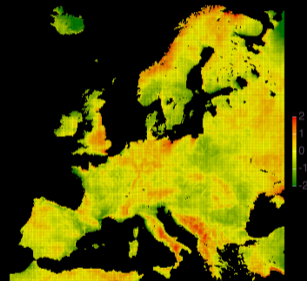
Original



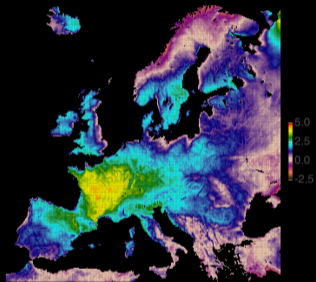
Reconstruction



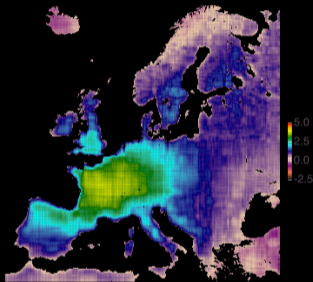
Error



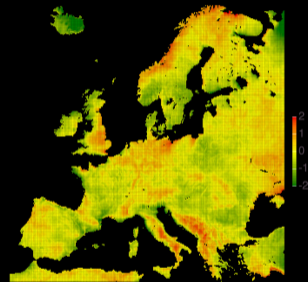
Original



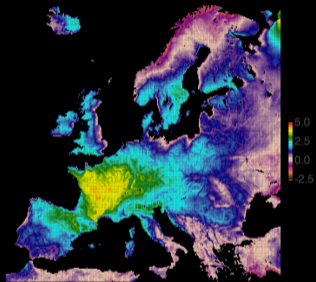
Reconstruction



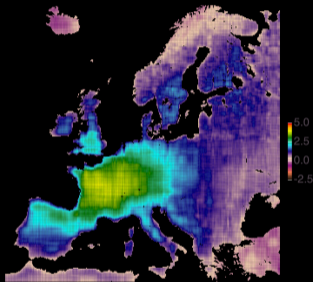
Error



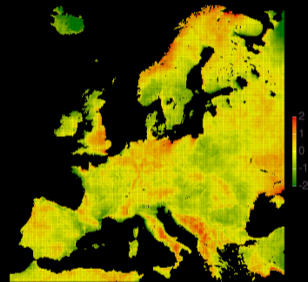
Original



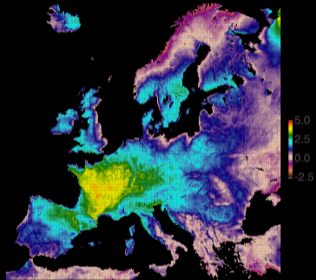
Reconstruction



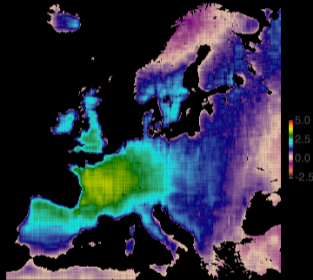
Error



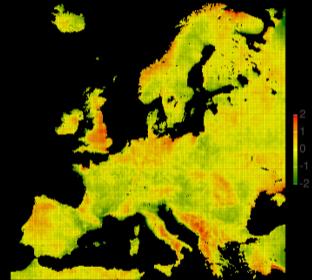
Original



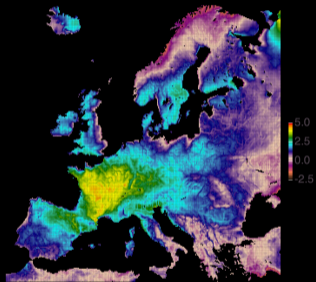
Reconstruction



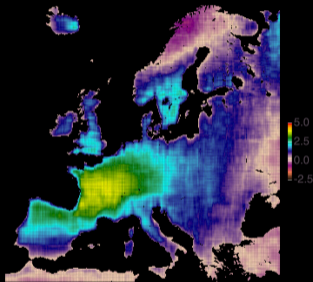
Error



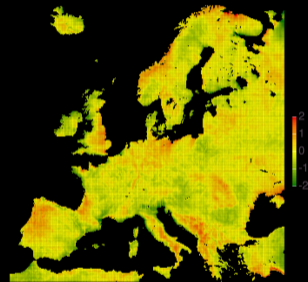
Original



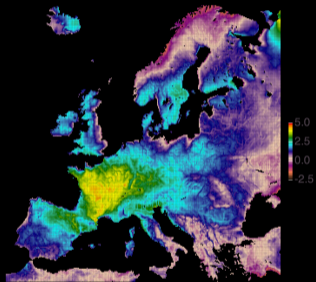
Reconstruction



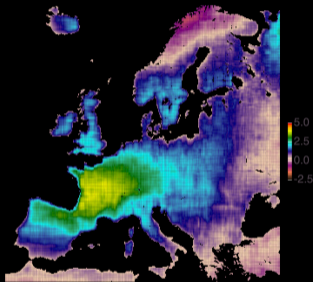
Error



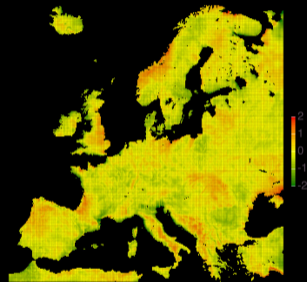
Original



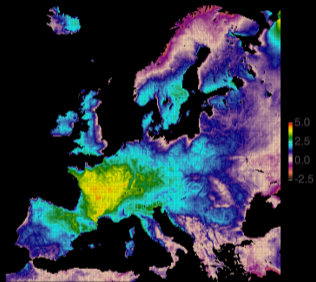
Reconstruction



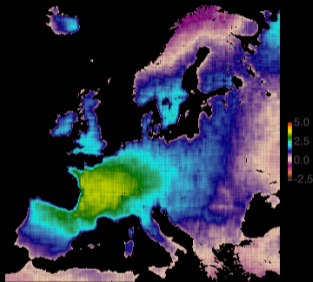
Error



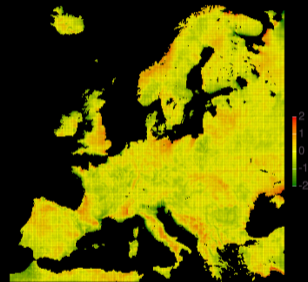
Original



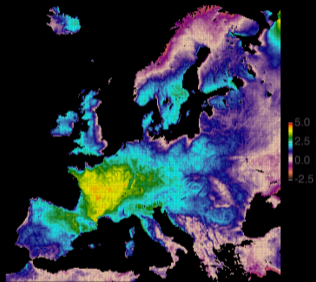
Reconstruction



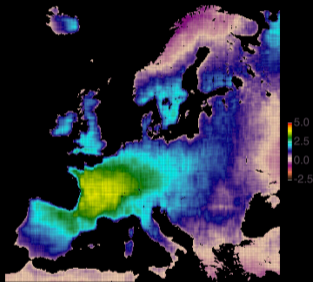
Error



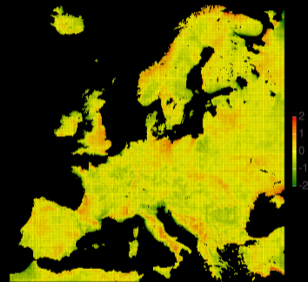
Original



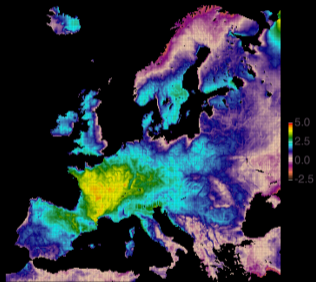
Reconstruction



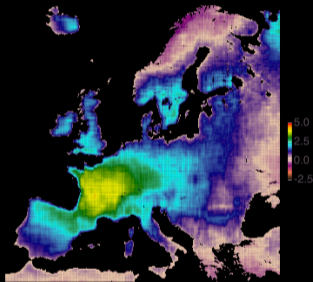
Error



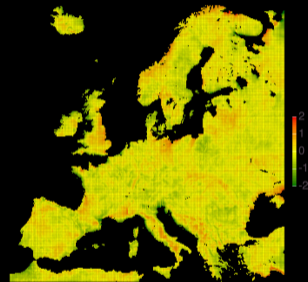
Original



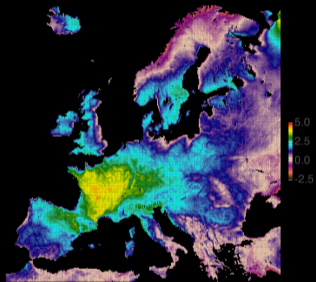
Reconstruction



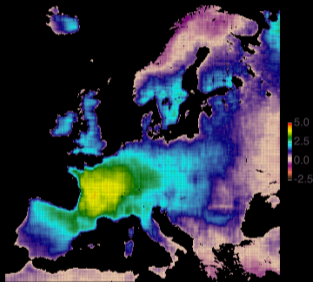
Error



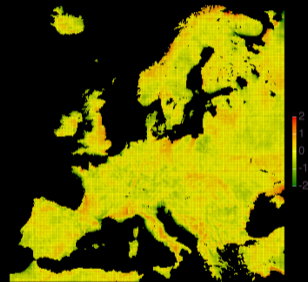
Original



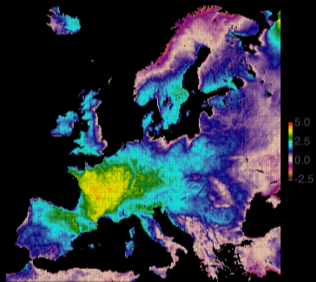
Reconstruction



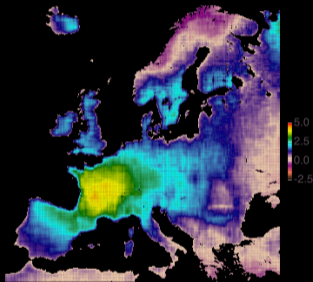
Error



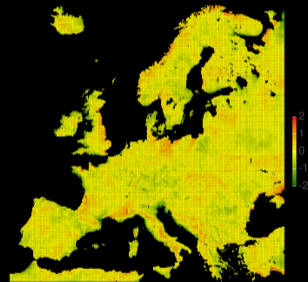
Original



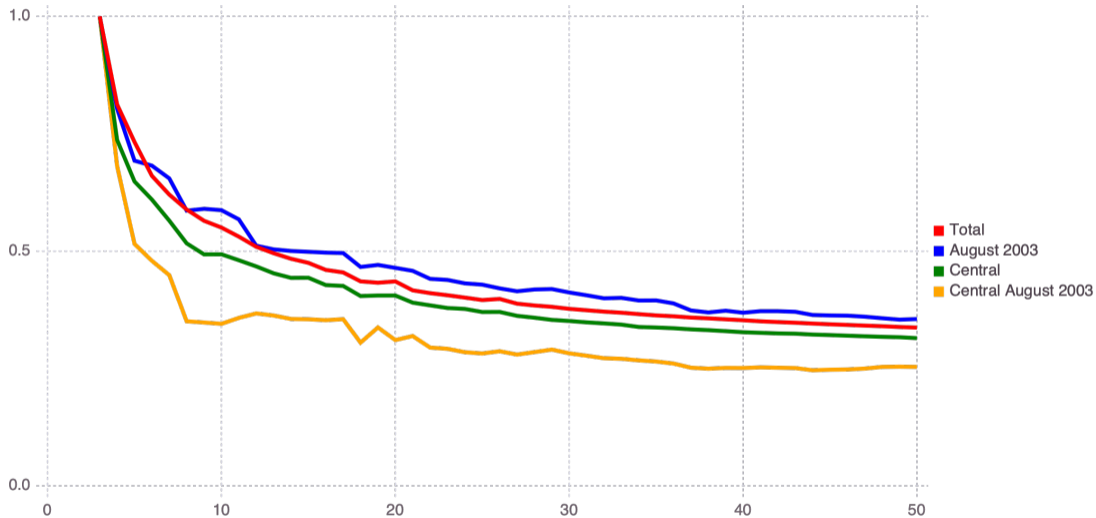
Reconstruction



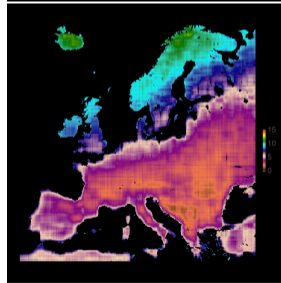
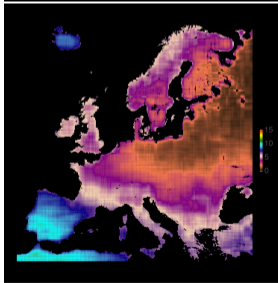
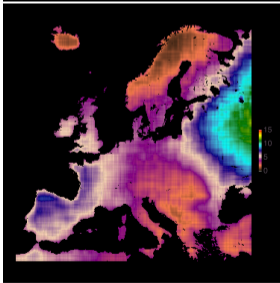
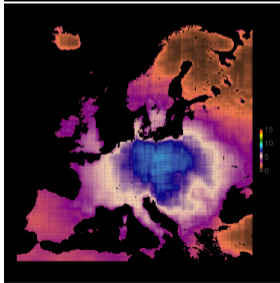
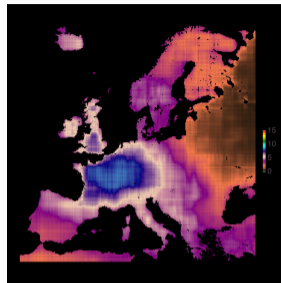
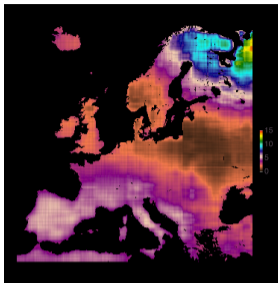
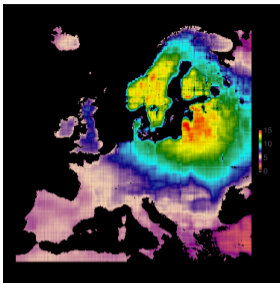
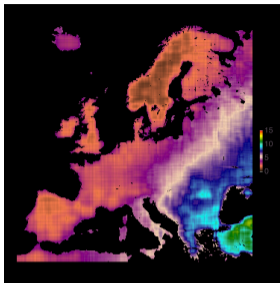
Error



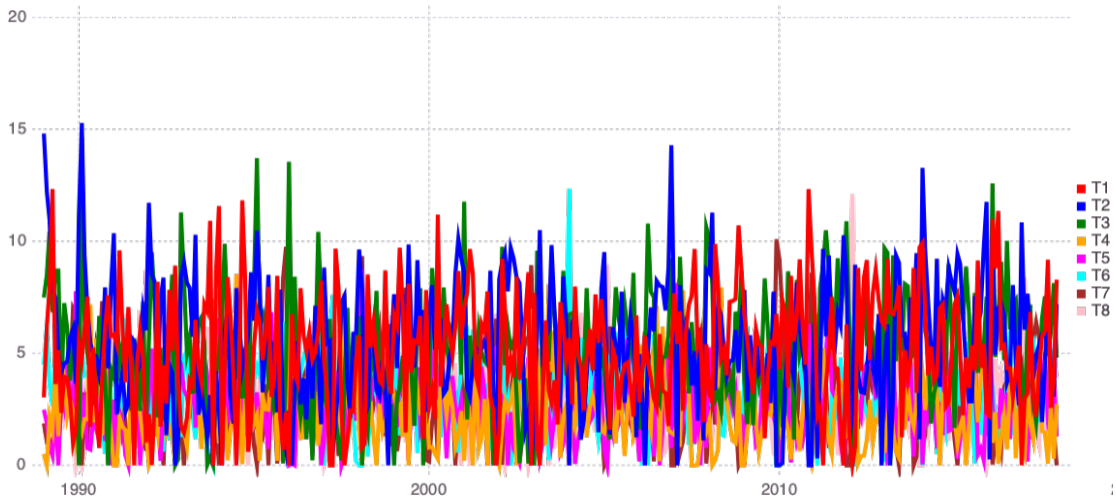
Climate model of Europe: air temperature reconstruction errors



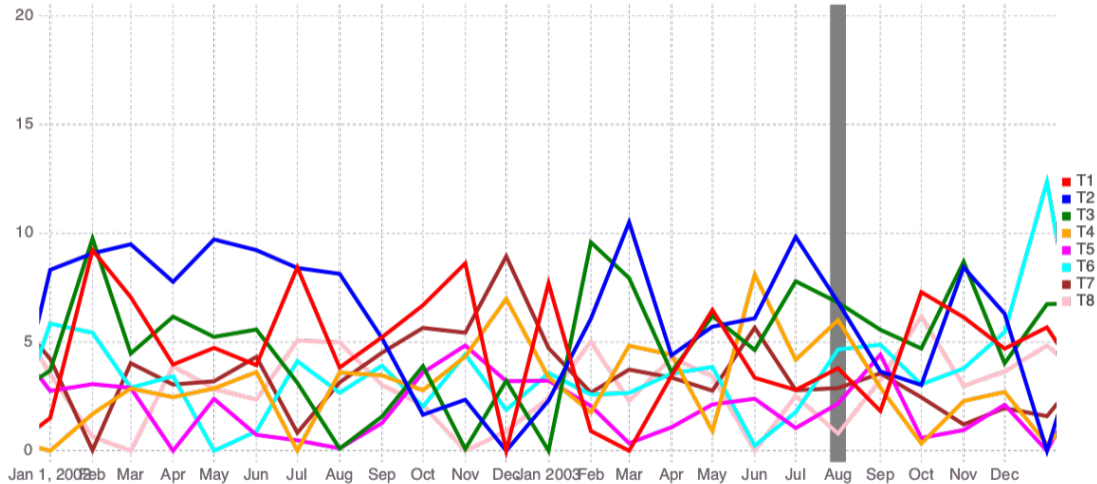
Climate model of Europe: air temperature features (8)



Climate model of Europe: air temperature features (8) 1989-2017



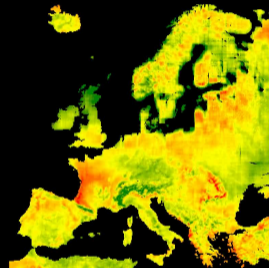
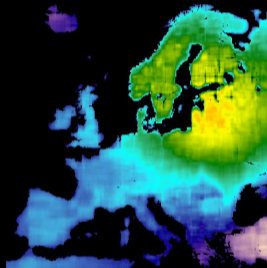
Climate model of Europe: air temperature features (8) 2002-2003



Original

Reconstruction

Error

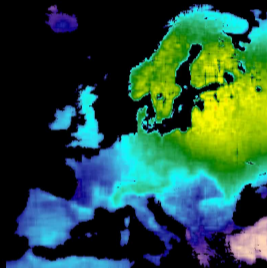


1889-01-01

Original

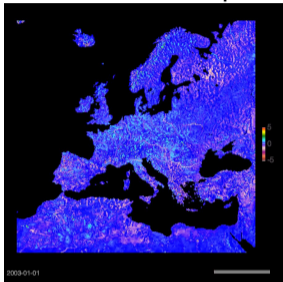
Reconstruction

Error

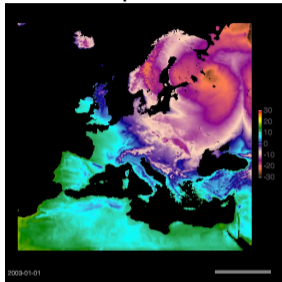


1889-01-01

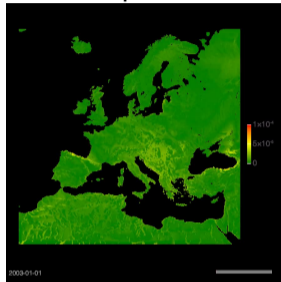
Water-table depth



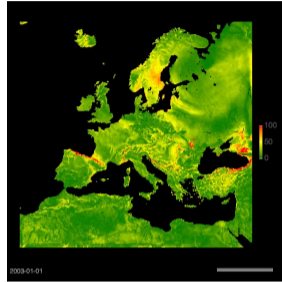
Temperature



Evaporation

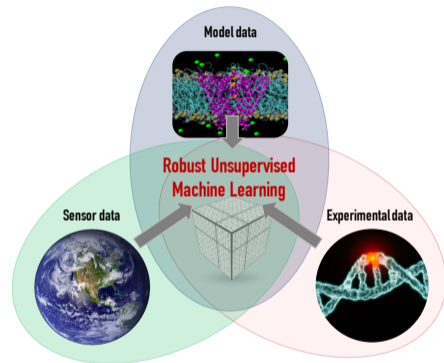


Sensible heat flux



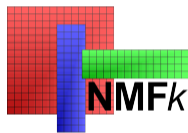
- ▶ Find interconnections and common dominant features in model outputs
- ▶ Evaluate impacts of different model setups (initial conditions, resolution, etc.)
- ▶ Find dominant processes impacting model predictions (e.g., occurrence of heat waves, climate impacts on groundwater resources, climate impacts of subsurface processes on atmospheric conditions)

- ▶ Developed a series of **novel** ML methods and computational tools
- ▶ Our ML methods have been applied to solve various real-world problems at various scales (from molecular to global) (brought breakthrough discoveries related to human cancer research)



► Codes:

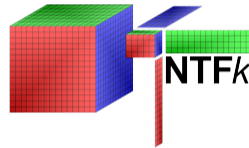
NMF k .jl



Mads.jl



NTF k .jl



► Examples:

http://madsjulia.github.io/Mads.jl/Examples/blind_source_separation

<http://tensors.lanl.gov>

<http://tensordecompositions.github.io>