5th International Workshop on Deep Learning in Computational Physics

Identifying partial differential equations of land surface schemes in Earth system models with neural networks

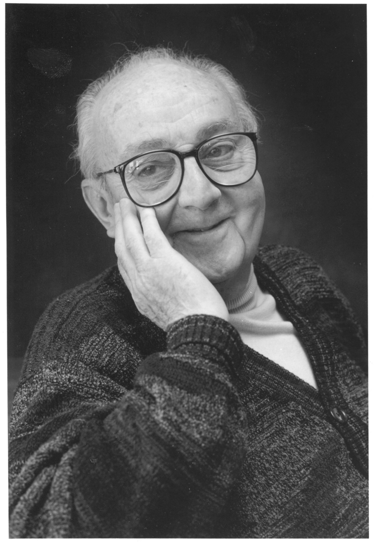
Mikhail Krinitskiy[1,2,*], Viktor Stepanenko[2,3,4], Ruslan Chernyshev[2,3]

[1] P.P. Shirshov Institute of Oceanology, Russian Academy of Sciences
[2] Research Computing Center, Lomonosov Moscow State University
[3] Faculty of Geography, Lomonosov Moscow State University
[4] Moscow Center for Fundamental and Applied Mathematics

[*] krinitsky@sail.msk.ru

CMIP6 climate models

"All models are wrong, but some are useful"



George Box

Climate models are one of the primary means for scientists to understand how the climate has changed in the past and may change in the future. These models simulate the physics, chemistry and biology of the atmosphere, land and oceans, and require some of the largest supercomputers in the world to generate their climate projections.

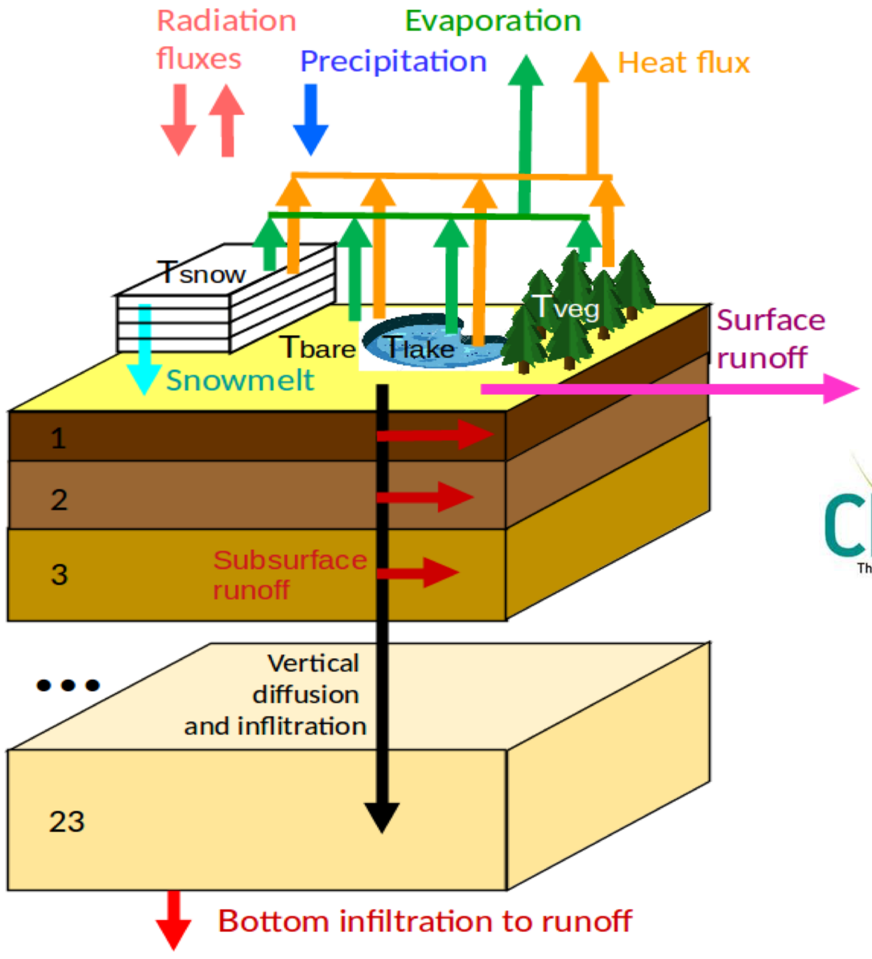
There is no an established climate model that is THE ONE and only.

There are **49 scientific groups** that run **over 100 distinct climate models** in CMIP6

Some of them are more accurate in reproducing sea surface temperature, others accurately reproduce snow cover, *etc.*

CMIP — coupled model intercomparison project

The INM RAS-MSU land surface scheme



Heat and moisture exchange in a cell of INM RAS-MSU land surface model

INM RAS-MSU land surface scheme

Climate model of Institute of Numerical Mathematics (INMCM)

Global numerical weather prediction SL-AV model

CMIP
The World Climate Research Programme's Coupled Model Intercomparison Project

- ❑ Heat, moisture, water vapor and ice dynamics in soil (23 layers)
- ❑ Snow cover with liquid water treatment (4 layers)
- ❑ Transpiration and photosynthesis by vegetation
- ❑ Wetland CH_4 model
- ❑ LAKE model
- ❑ River routing scheme
- ❑ Standalone version resolution is 0.5 x 0.5 deg.
- ❑ GLCC global land cover data

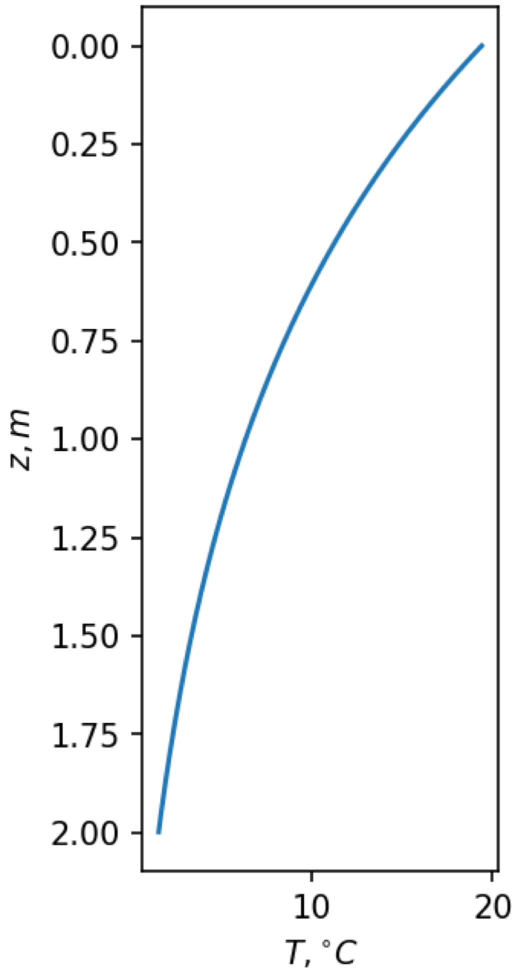
Land surface scheme

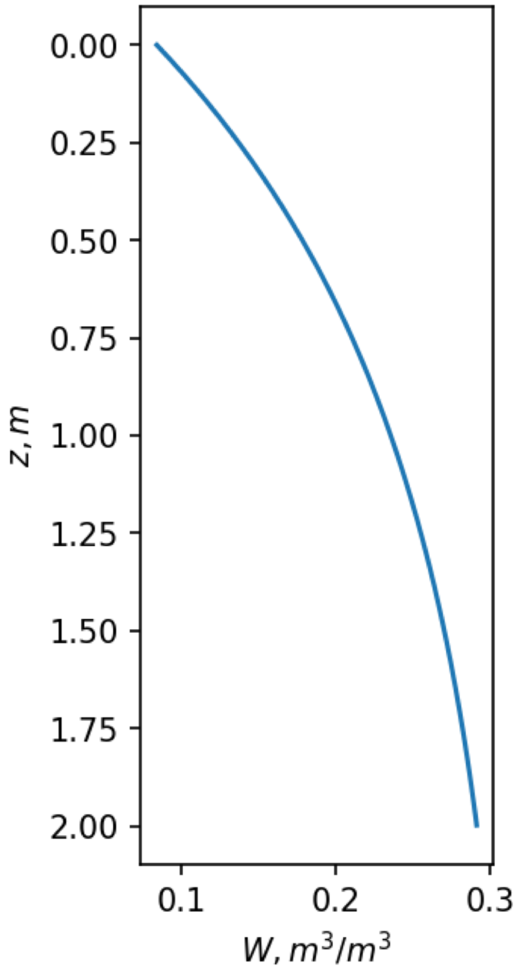
Thermal conductance equation

$$\rho c \frac{\partial T}{\partial t} = \frac{\partial}{\partial z} \lambda_T \frac{\partial T}{\partial z} + \cdots$$

Richards equation: water vapor dynamics equation

$$\frac{\partial W}{\partial t} = \frac{\partial}{\partial z}\left(\lambda_W \frac{\partial W}{\partial z}\right) + \frac{\partial \gamma_W}{\partial z} + \cdots$$





Land surface scheme

Thermal conductance (diffusion) equation

$$\rho c \frac{\partial T}{\partial t} = \frac{\partial}{\partial z} \lambda_T \frac{\partial T}{\partial z} + \cdots$$

$$\boxed{\frac{\partial T}{\partial t} = \frac{\partial}{\partial z} \lambda_T^* \frac{\partial T}{\partial z}}$$

heat conductivity coefficient

$$\lambda_T = \lambda_T(W, T)$$

Richards equation — for liquid water dynamics

$$\frac{\partial W}{\partial t} = \frac{\partial}{\partial z}\left(\lambda_W \frac{\partial W}{\partial z}\right) + \frac{\partial \gamma_W}{\partial z} + \cdots$$

$$\boxed{\frac{\partial W}{\partial t} = \frac{\partial}{\partial z}\left(\lambda_W \frac{\partial W}{\partial z}\right) + \frac{\partial \gamma_W}{\partial z}}$$

water diffusivity coefficient

$$\lambda_W = \lambda_W(W, T)$$

hydraulic conductivity coefficient

$$\gamma_W = \gamma_W(W, T)$$

Land surface scheme

Thermal conductance equation

$$\frac{\partial T}{\partial t} = \frac{\partial}{\partial z} \lambda_T \frac{\partial T}{\partial z}$$

Richards equation

$$\frac{\partial W}{\partial t} = \frac{\partial}{\partial z}\left(\lambda_W \frac{\partial W}{\partial z}\right) + \frac{\partial \gamma_W}{\partial z}$$

Important uncertainties of the land surface characteristics evolution are due to crude approximations of λ_T, λ_W and γ_W

$\lambda_T = k_{dry} + k_r\left(k_{sat} - k_{dry}\right), \quad k_r = \frac{kW}{1 + (k-1)W}$	(Cote and Conrad, 2005)
$\lambda_W(W) = \frac{\gamma_{max}(1-m)}{\alpha m(\theta_s - \theta_r)} W^{\frac{1}{2}-\frac{1}{m}} * \left[\left(1 - W^{\frac{1}{m}}\right)^{-m} + \left(1 - W^{\frac{1}{m}}\right)^{m} - 2\right]$ $\gamma_W(W) = \gamma_{max} W^{\frac{1}{2}} \left[1 - \left(1 - W^{\frac{1}{m}}\right)^{m}\right]^2$	(van Genuchten, 1980)

PDE identification

Thermal conductance equation

$$\boxed{\frac{\partial T}{\partial t} = \frac{\partial}{\partial z} \lambda_T \frac{\partial T}{\partial z}}$$

Richards equation

$$\boxed{\frac{\partial W}{\partial t} = \frac{\partial}{\partial z}\left(\lambda_W \frac{\partial W}{\partial z}\right) + \frac{\partial \gamma_W}{\partial z}}$$

Major **uncertainties** of the land surface characteristics evolution are due to **crude approximations** of λ_T, λ_W and γ_W

PDE identification: the task is to approximate the coefficients λ_T, λ_W and γ_W.

Our approach: to approximate the coefficients λ_T, λ_W and γ_W as parametric learnable functions of PDE solution.

PDE identification

$$\boxed{\frac{\partial X}{\partial t} = \frac{\partial}{\partial z}\lambda_X \frac{\partial X}{\partial z}}$$

as an example

$$\lambda_X = F(X, \theta)$$

$$\hat{\theta} = \underset{\Theta}{\operatorname{argmin}}\, \boldsymbol{\mathcal{L}}(X_{true}, F(X, \theta))$$

θ are the parameters
X is the PDE solution

$\boldsymbol{\mathcal{L}}$—?

PDE identification

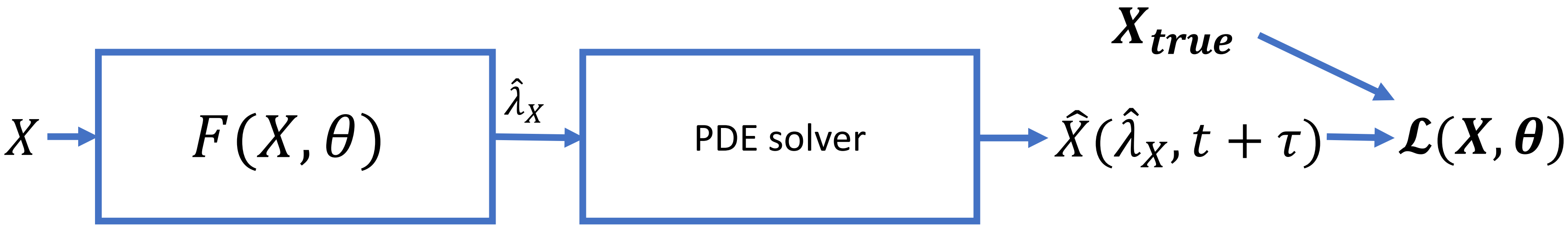
$$\boxed{\frac{\partial X}{\partial t} = \frac{\partial}{\partial z}\lambda_X \frac{\partial X}{\partial z}}$$

as an example

$$\lambda_X = F(X, \theta)$$

θ are the parameters
X is the PDE solution

$$\hat{\theta} = \underset{\Theta}{\operatorname{argmin}}\, \mathcal{L}(X_{true}, F(X, \theta))$$



$$\mathcal{L}(\boldsymbol{X}, \boldsymbol{\theta}) = \boldsymbol{MSE}\left(\widehat{\boldsymbol{X}}\left(\boldsymbol{F}(\boldsymbol{X}, \boldsymbol{\theta})\right), \boldsymbol{X}_{\boldsymbol{true}}\right)$$

PDE identification

$$\boxed{\frac{\partial X}{\partial t} = \frac{\partial}{\partial z}\lambda_X\frac{\partial X}{\partial z}}$$

as an example

$$\lambda_X = F(X,\theta)$$

θ are the parameters
X is the PDE solution

$$\hat{\theta} = \underset{\Theta}{\operatorname{argmin}}\, \mathcal{L}(X_{true}, F(X,\theta))$$

$$X \rightarrow \boxed{F(X,\theta)} \xrightarrow{\hat{\lambda}_X} \boxed{\text{PDE solver}} \rightarrow \hat{X}(\hat{\lambda}_X, t+\tau) \rightarrow \boldsymbol{\mathcal{L}(X,\theta)} \leftarrow \boldsymbol{X_{true}}$$

$$\boldsymbol{\mathcal{L}(X,\theta) = MSE\left(\hat{X}\left(F(X,\theta)\right), X_{true}\right)}$$

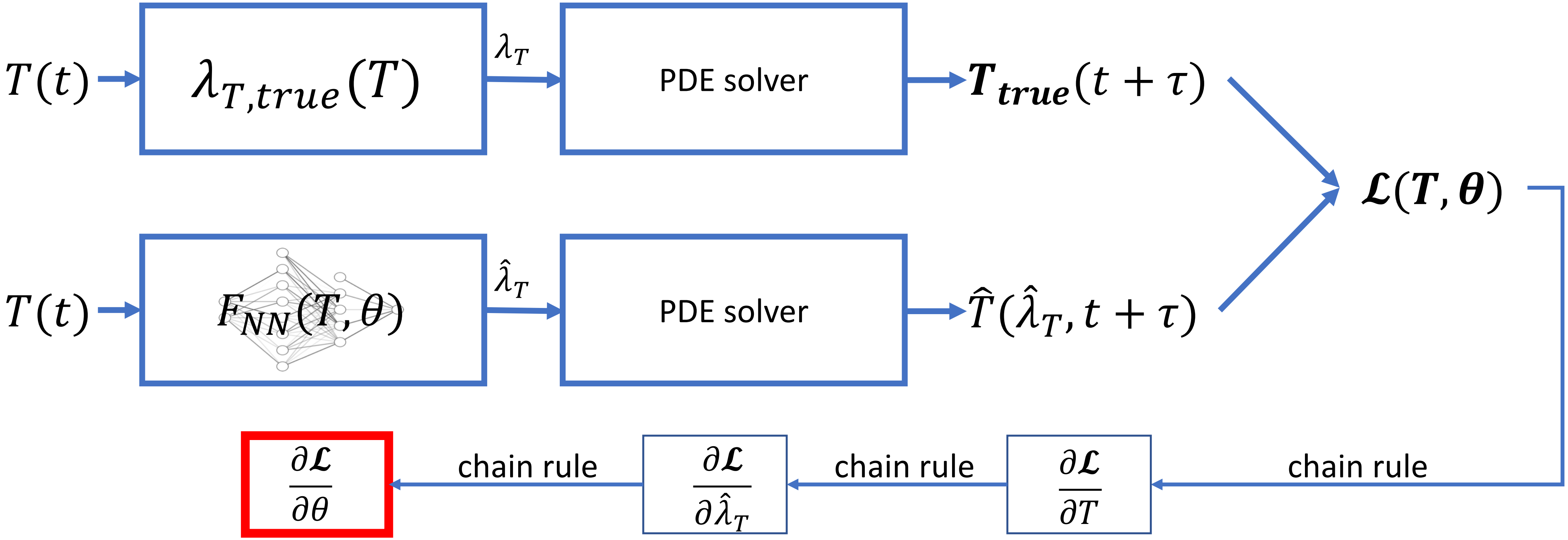
Chain rule: $$\frac{\partial \mathcal{L}(X,\theta)}{\partial \theta} = 2\left(\hat{X}\left(F(X,\theta)\right) - X_{true}\right) * \boldsymbol{\frac{\partial \hat{X}}{\partial F}} * \frac{\partial F}{\partial \theta}$$ Gradient optimization starts here

PDE identification: scenario 1

$$\frac{\partial T}{\partial t} = \frac{\partial}{\partial z}\lambda_T \frac{\partial T}{\partial z}$$

X – temperature T;
$\lambda_T = F_{NN}(T, \theta)$;

we model the evolution T_{true} of soil column using known formulae for $\lambda_{T,true}$. Assuming Dirichlet b.c.s.

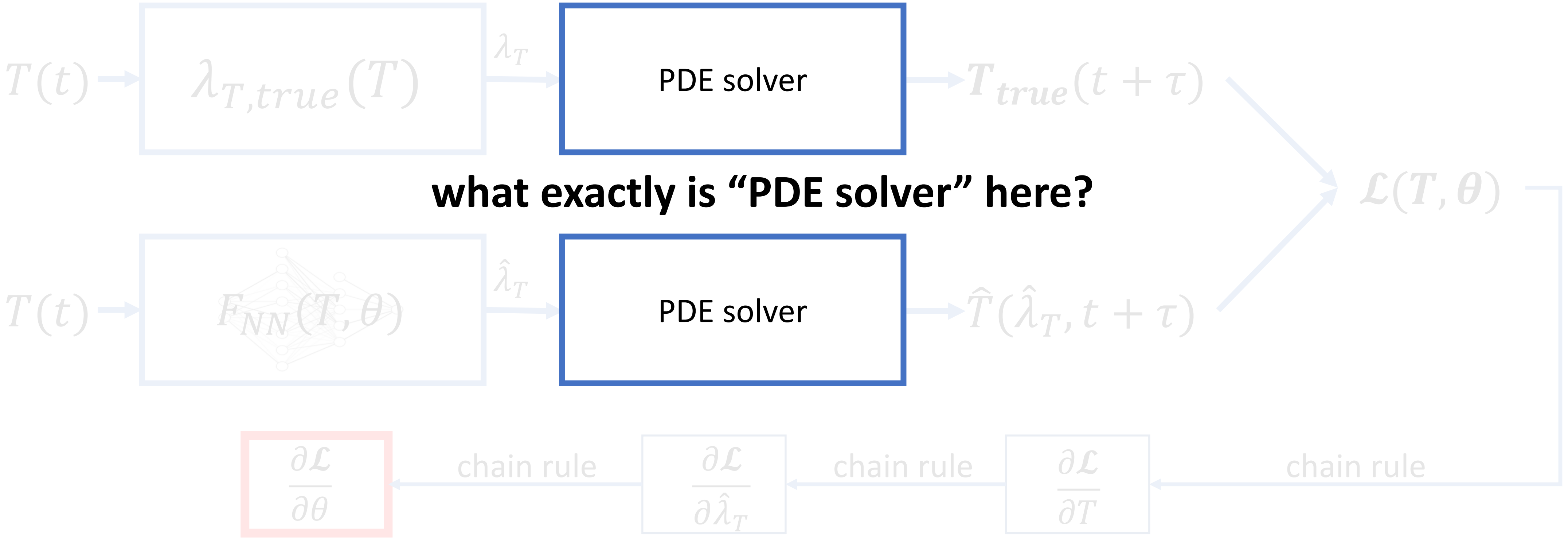


PDE identification: scenario 1

$$\frac{\partial T}{\partial t} = \frac{\partial}{\partial z}\lambda_T \frac{\partial T}{\partial z}$$

X – temperature T;
$\lambda_T = F_{NN}(T, \theta)$;

we model the evolution T_{true} of soil column using known formulae for $\lambda_{T,true}$



PDE identification: scenario 1

$$\frac{\partial T}{\partial t} = \frac{\partial}{\partial z}\lambda_T \frac{\partial T}{\partial z}$$

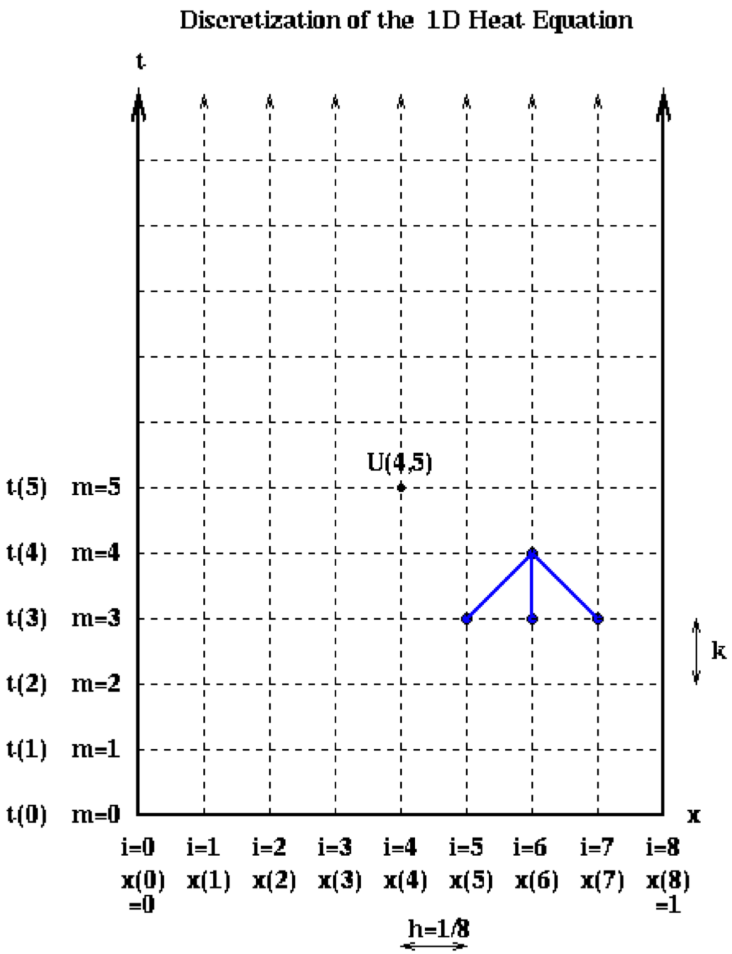
PDE solver:
explicit scheme

$$\left(\frac{\partial T}{\partial t}\right)_i = \frac{1}{\Delta z_i}\left(\frac{\boldsymbol{\lambda}\left(T^t_{i+\frac{1}{2}}\right)\left(T^t_{i+1} - T^t_i\right)}{\Delta z_i + \Delta z_{i+1}} - \frac{\boldsymbol{\lambda}\left(T^t_{i-\frac{1}{2}}\right)\left(T^t_i - T^t_{i-1}\right)}{\Delta z_{i-1} + \Delta z_i}\right)$$

$$T_i^{t+1} = T_i^t + \left(\frac{\partial T}{\partial t}\right)_i \Delta t$$

remember the goal: $$\frac{\partial \mathcal{L}(T,\theta)}{\partial \theta} = 2\left(\hat{T} - T_{true}\right) * \frac{\boldsymbol{\partial \hat{T}}}{\boldsymbol{\partial \lambda_T}} * \frac{\partial F_{NN}}{\partial \theta}$$

and gradient optimization starts here



PDE identification: scenario 1

we model the evolution T_{true} of soil column using known formulae for $\lambda_{T,true}$

$$\mathcal{L} = MSE\left(\hat{T}, T_{true}\right) = \sum\left(T_i^t + \widehat{\left(\frac{\partial T}{\partial t}\right)}_i \Delta t - T_i^t - \left(\frac{\partial T}{\partial t}\right)_i \Delta t\right)^2 =$$

$$= k\sum\left(\widehat{\left(\frac{\partial T}{\partial t}\right)}_i - \left(\frac{\partial T}{\partial t}\right)_i\right)^2$$

$$\boxed{\mathcal{L}^* = MSE\left(\widehat{\left(\frac{\partial T}{\partial t}\right)}, \left(\frac{\partial T}{\partial t}\right)\right)} = k\mathcal{L}$$

$$\underset{\theta}{\operatorname{argmin}}\, \mathcal{L}^* = \underset{\theta}{\operatorname{argmin}}\, \mathcal{L}$$

and gradient optimization starts here

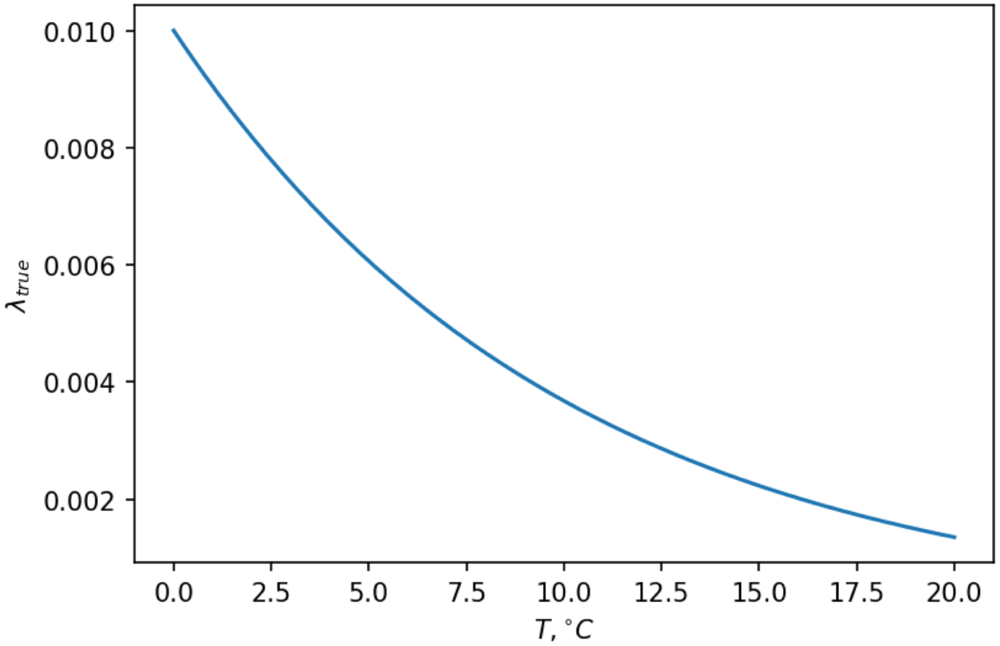
Quality measures

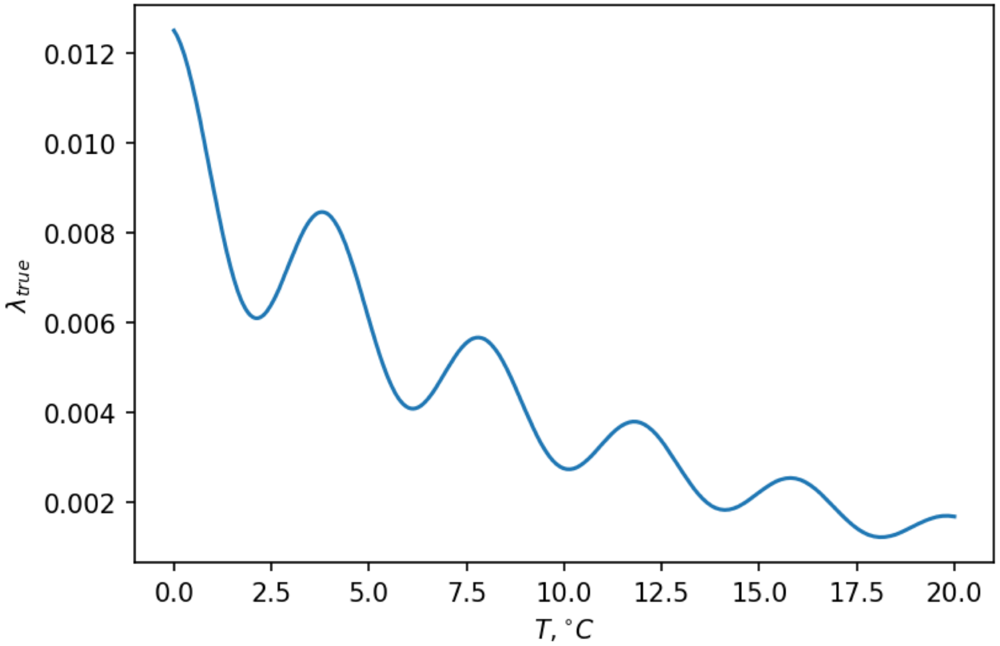
$$MAPE(\lambda_{NN}, \lambda_{true}) = \frac{1}{N}\sum\left|\frac{\lambda_{NN} - \lambda_{true}}{\lambda_{true} + \xi}\right|$$

$$RMSE\left(\frac{\widehat{\partial X}}{\partial t}_{NN}, \frac{\partial X}{\partial t}_{true}\right) = \sqrt{\frac{1}{N}\sum\left(\frac{\widehat{\partial X}}{\partial t}_{NN} - \frac{\partial X}{\partial t}_{true}\right)^2}$$

PDE identification: scenario 1

we model the evolution T_{true} of soil column using **known formulae** for $\lambda_{T,true}$



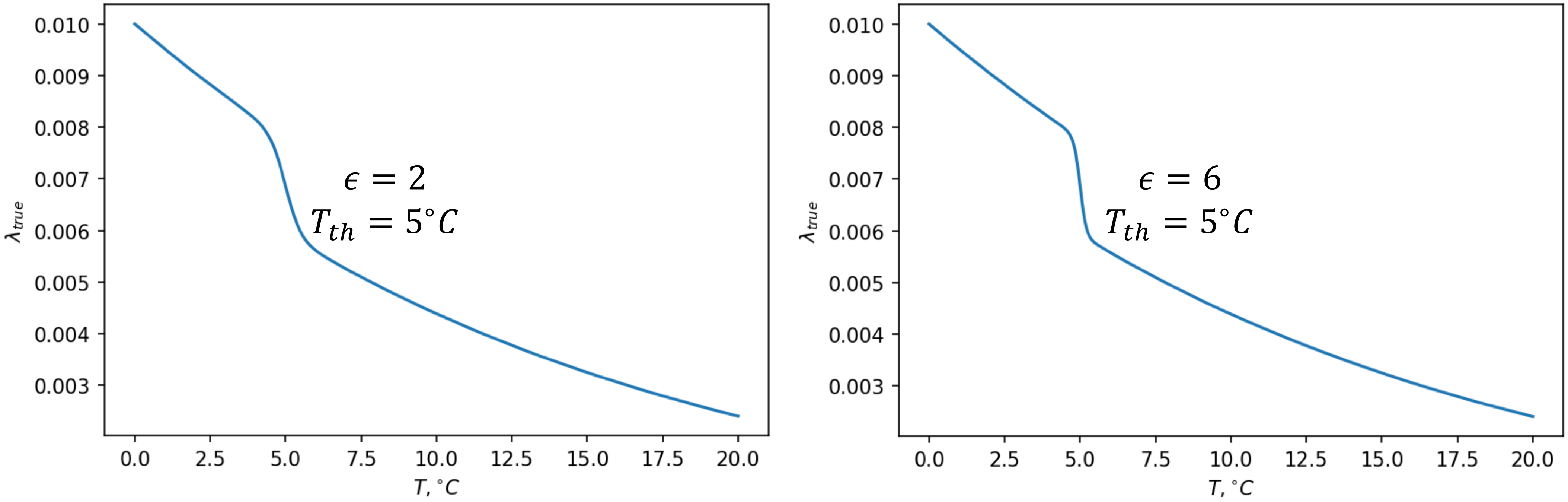


$$\lambda_T(T) = a\exp(bT)$$

$$\lambda_T(T) = a\exp(bT) * \frac{C + \cos(k\pi T)}{C}$$

PDE identification: scenario 1

we model the evolution T_{true} of soil column using **known formulae** for $\lambda_{T,true}$



$$\zeta = \frac{1}{2}\left(1 - \tanh\left(\epsilon(T - T_{th})\right)\right),$$

$$\lambda_T(T) = a_1 e^{b_1 T} * \zeta + a_2 e^{b_2 T} * (1 - \zeta)$$

Neural network optimization

$\lambda_T(T) = F_{NN}(T, \theta)$ — Fully-connected neural network

6 layers

Mish activation[*]

Approaches used:

Adam[**] optimizer;

SGDR[***] learning rate schedule;

Normal additive noise for improving generalization ability (zero-centered, uncorrelated)

Exponential decay of noise magnitude

Implementation:

(a) **Jax** + Haiku + Optax (FP32 only)

(b) **Tensorflow** (FP32 and mixed precision tested)

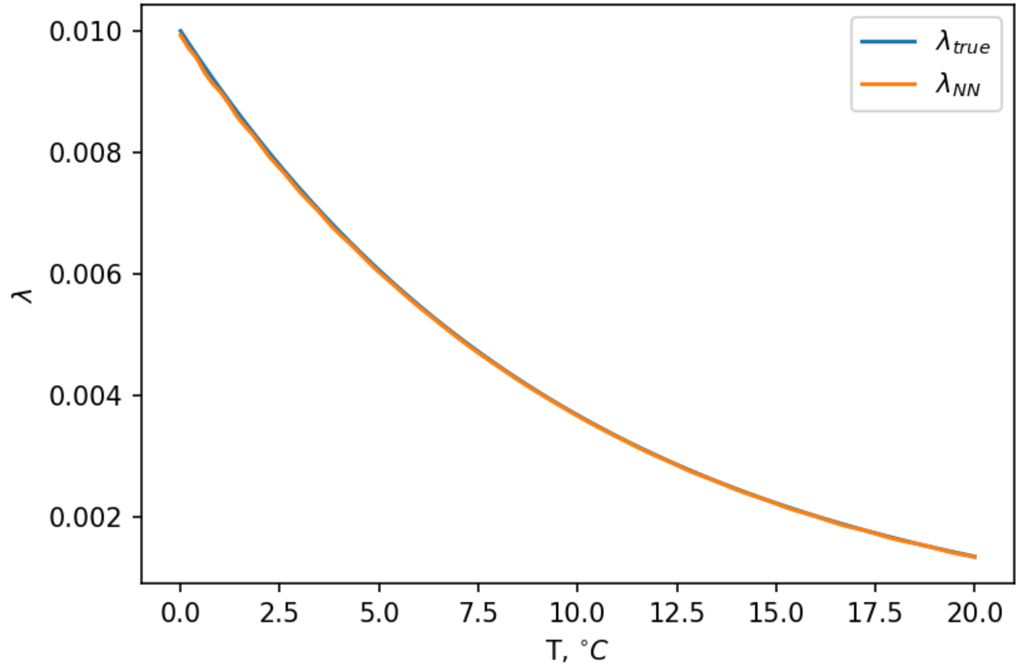
NVIDIA DGX Station at SAIL IORAS

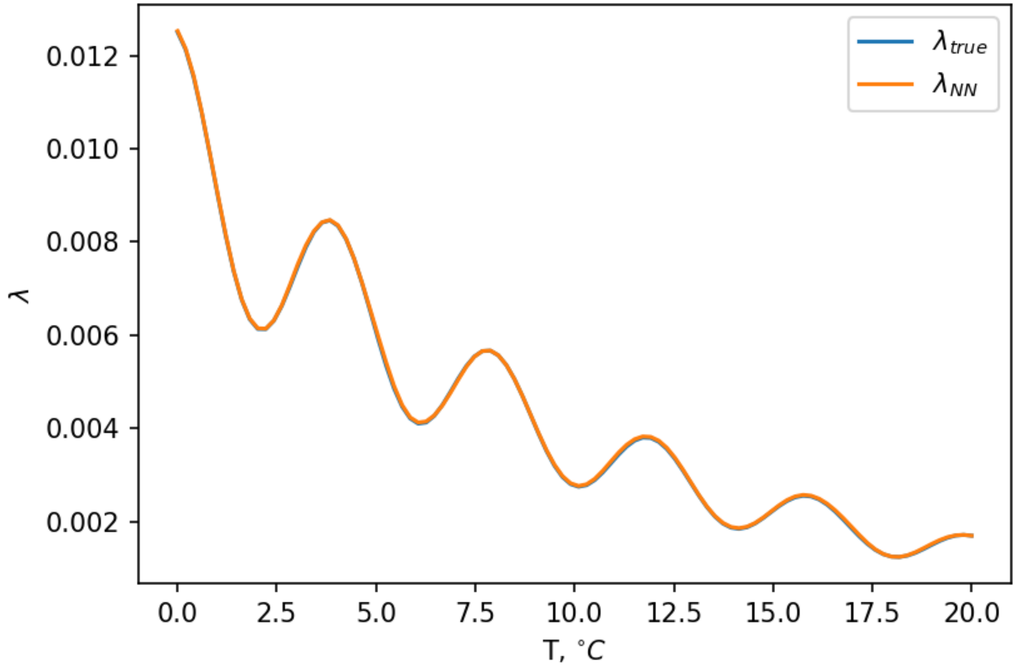
[*] Diganta Misra. "Mish: A Self Regularized Non-Monotonic Activation Function". arXiv:1908.08681 [cs, stat] (Aug. 13, 2020). http: //arxiv.org/abs/1908.08681

[**] Kingma D. P., Ba J. Adam: A Method for Stochastic Optimization // arXiv:1412.6980 [cs]. 2017. http: //arxiv.org/abs/1412.6980

[***] Loshchilov I., Hutter F. SGDR: Stochastic Gradient Descent with Warm Restarts // arXiv:1608.03983 [cs, math]. 2016. http: //arxiv.org/abs/1608.03983

Results: scenario 1





FP32

$MAPE(\lambda) = 8\times10^{-4}$

$RMSE\left(\frac{\partial T}{\partial t}\right) = 4.6\times10^{-6}$

Jax: 72 it/sec
TF: 30 it/sec

Mixed precision

FP32

$MAPE(\lambda) = 1.6\times10^{-3}$

$RMSE\left(\frac{\partial T}{\partial t}\right) = 7\times10^{-6}$

Jax: 71 it/sec
TF: 30 it/sec

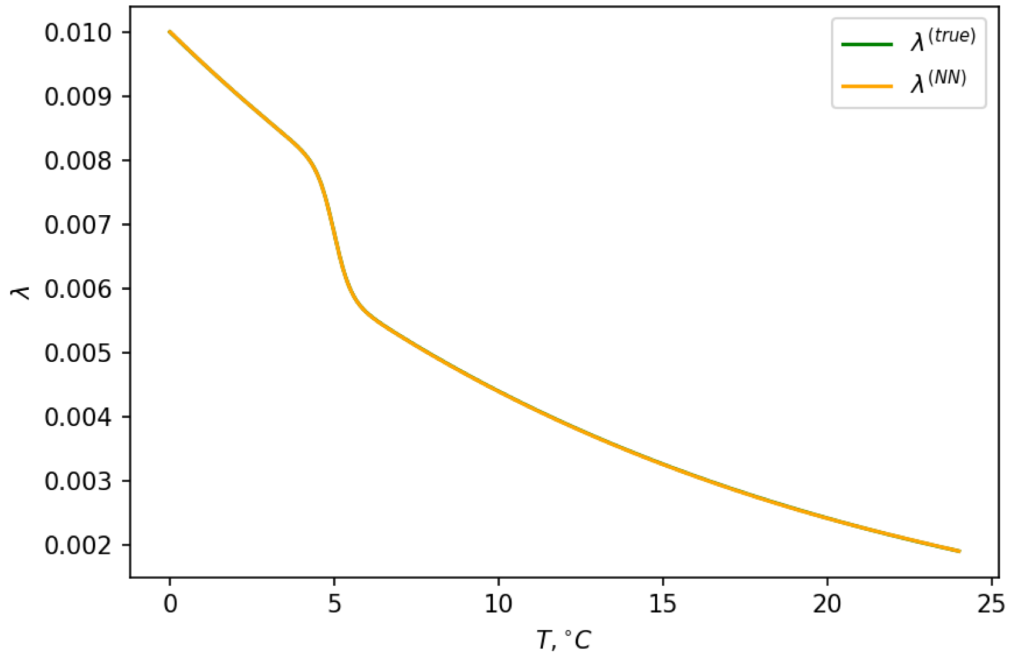
Mixed precision

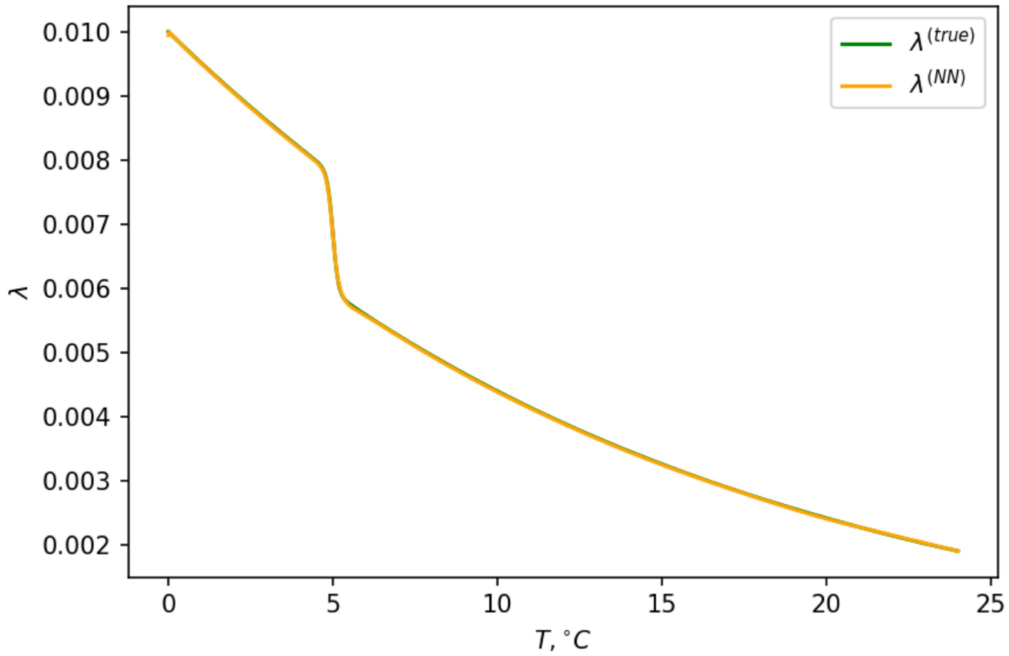
$MAPE(\lambda) = 1.0\times10^{-2}$

$RMSE\left(\frac{\partial T}{\partial t}\right) = 4.5\times10^{-2}$

TF: 74 it/sec

Results: scenario 1





FP32

$MAPE(\lambda) = 1.5\times10^{-4}$

$RMSE\left(\frac{\partial T}{\partial t}\right) = 3.6\times10^{-4}$

Jax: 70 it/sec
TF: 29 it/sec

Mixed precision

$MAPE(\lambda) = 1.2\times10^{-3}$

$RMSE\left(\frac{\partial T}{\partial t}\right) = 2\times10^{-1}$

TF: 75 it/sec

FP32

$MAPE(\lambda) = 1.75\times10^{-4}$

$RMSE\left(\frac{\partial T}{\partial t}\right) = 2\times10^{-4}$

Jax: 71 it/sec
TF: 29 it/sec

Mixed precision

$MAPE(\lambda) = 1.1\times10^{-3}$

$RMSE\left(\frac{\partial T}{\partial t}\right) = 1.9\times10^{-1}$

TF: 74 it/sec

PDE identification: scenario 2

$$\boxed{\frac{\partial W}{\partial t} = \frac{\partial}{\partial z}\lambda_W \frac{\partial W}{\partial z}}$$

X – liquid water W;
$\lambda_W = F_{NN}(W, \theta)$;

we model the evolution W_{true} of soil column using known formulae for $\lambda_{W,true}$. Assuming Dirichlet b.c.s.

Additional assumptions:
1. $W \geq 0; W \leq 1;$
2. $\lambda_W(0) = 0$
3. $\lambda_W(W) \geq 0$ for any W
4. $\frac{\partial \lambda_W}{\partial W} \geq 0$ for any W

regularization terms:

$$\mathcal{L}^{**} = \mathcal{L}^{*} + \alpha_1 F_{NN}^2(W = 0, \theta\ ) +$$

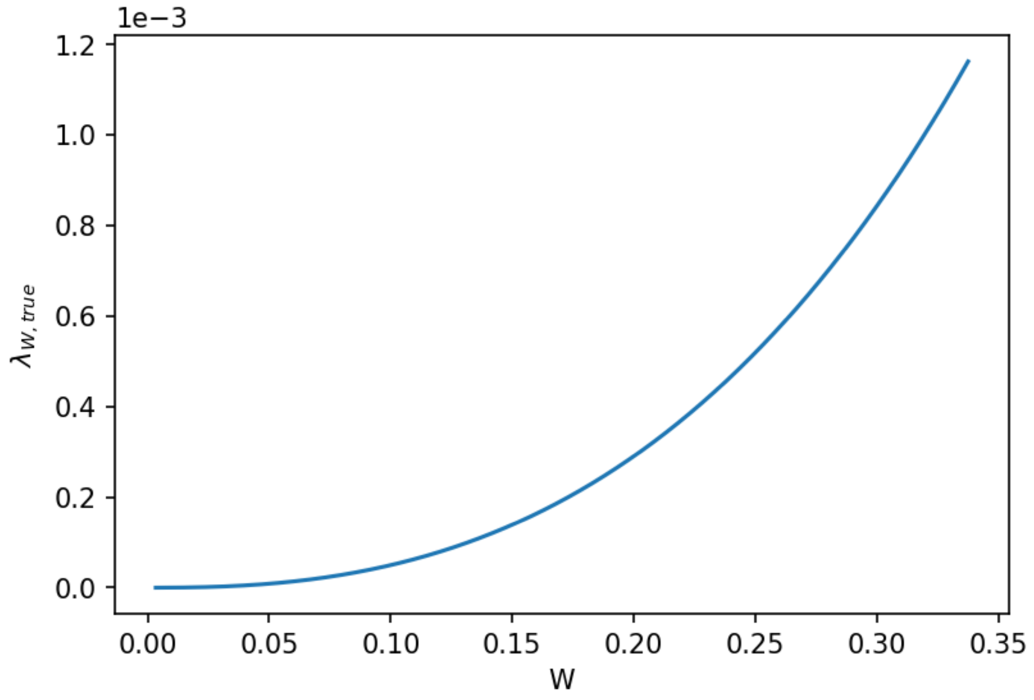
$$+\alpha_2 \sum_{\hat{\lambda}_W < 0} F_{NN}^2(W, \theta)$$

$$+\alpha_3 \sum_{\frac{\partial \hat{\lambda}_W}{\partial W} < 0} \left(\frac{\partial F_\lambda(W, \theta_\lambda)}{\partial W}\right)^2$$

PDE identification: scenario 2

we model the evolution W_{true} of soil column using **known formulae** for $\lambda_{W,true}$

$$\lambda_W(W) = \frac{\gamma_{max}(1-m)}{\alpha m(\theta_s - \theta_r)} W^{\frac{1}{2}-\frac{1}{m}} * \left[\left(1 - W^{\frac{1}{m}}\right)^{-m} + \left(1 - W^{\frac{1}{m}}\right)^{m} - 2\right]$$



Results: scenario 2

FP32

$MAPE(\lambda) = 2.3\times10^{-2}$

$RMSE\left(\frac{\partial W}{\partial t}\right) = 7.7\times10^{-4}$

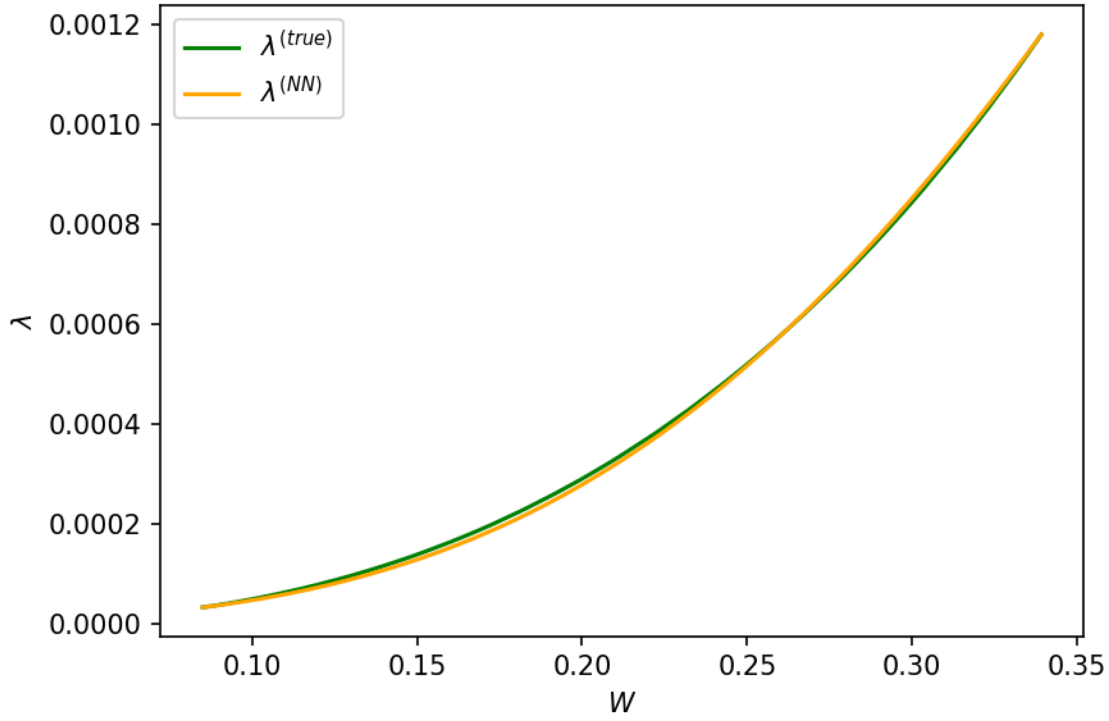
Jax: 72 it/sec
TF: 30 it/sec

Mixed precision

$MAPE(\lambda) = 3.2\times10^{-2}$

$RMSE\left(\frac{\partial W}{\partial t}\right) = 1.5\times10^{-4}$

TF: 76 it/sec



PDE identification: scenario 3

$$\boxed{\frac{\partial W}{\partial t} = \frac{\partial}{\partial z}\left(\lambda_W \frac{\partial W}{\partial z}\right) + \frac{\partial \gamma_W}{\partial z}}$$

Richards equation
X – liquid water W;
$\lambda_W = F_\lambda(W, \theta_\lambda)$;
$\gamma_W = F_\gamma(W, \theta_\gamma)$;

we model the evolution W_{true} of soil column using known formulae for $\lambda_{W,true}$ and $\gamma_{W,true}$. Assuming Dirichlet b.c.s.

Additional assumptions:

1. $W \geq 0; W \leq 1;$
2. $\lambda_W(0) = 0$
3. $\frac{\partial \lambda_W}{\partial W} \geq 0$ for any W
4. $\gamma_W \geq 0$
5. $\gamma_W(0) = 0$
6. $\frac{\partial \gamma_W}{\partial W} \geq 0$ for any W

regularization terms:

$$\mathcal{L}^{**} = \mathcal{L}^{*} + \alpha_1 F_\lambda^2(W = 0, \theta_\lambda) + \alpha_2 F_\gamma^2(W = 0, \theta_\gamma) +$$

$$+\alpha_3 \sum_{\hat{\lambda}_W < 0} F_\lambda^2(W, \theta_\lambda) + \alpha_4 \sum_{\hat{\gamma}_W < 0} F_\gamma^2(W, \theta_\gamma)$$

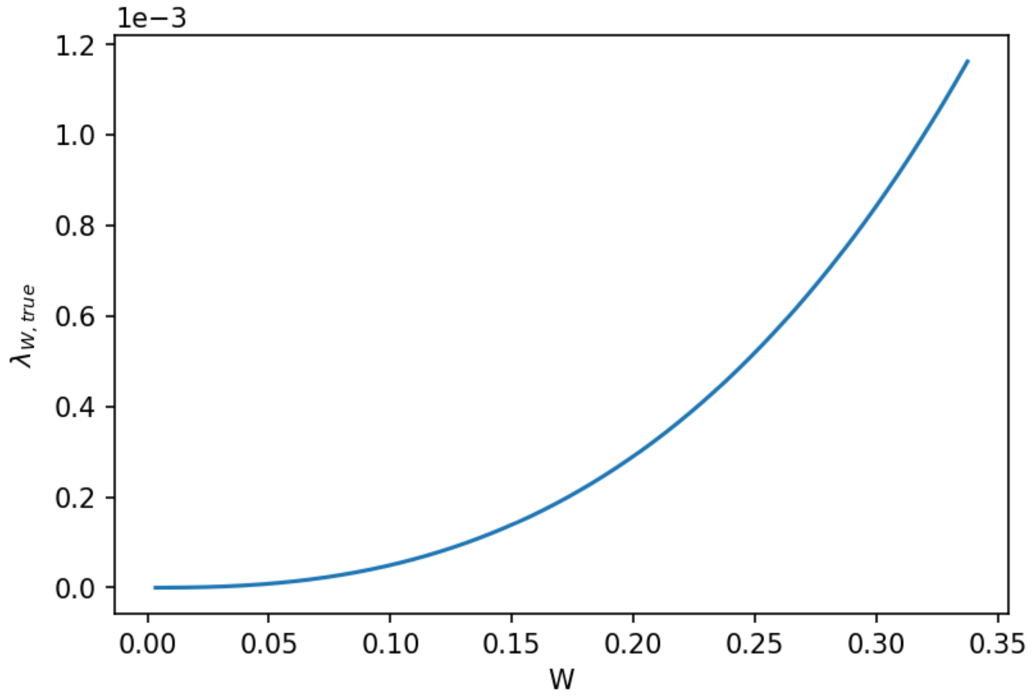
$$+\alpha_5 \sum_{\frac{\partial \hat{\lambda}_W}{\partial W} < 0} \left(\frac{\partial F_\lambda(W, \theta_\lambda)}{\partial W}\right)^2 + \alpha_6 \sum_{\frac{\partial \hat{\gamma}_W}{\partial W} < 0} \left(\frac{\partial F_\gamma(W, \theta_\gamma)}{\partial W}\right)^2$$

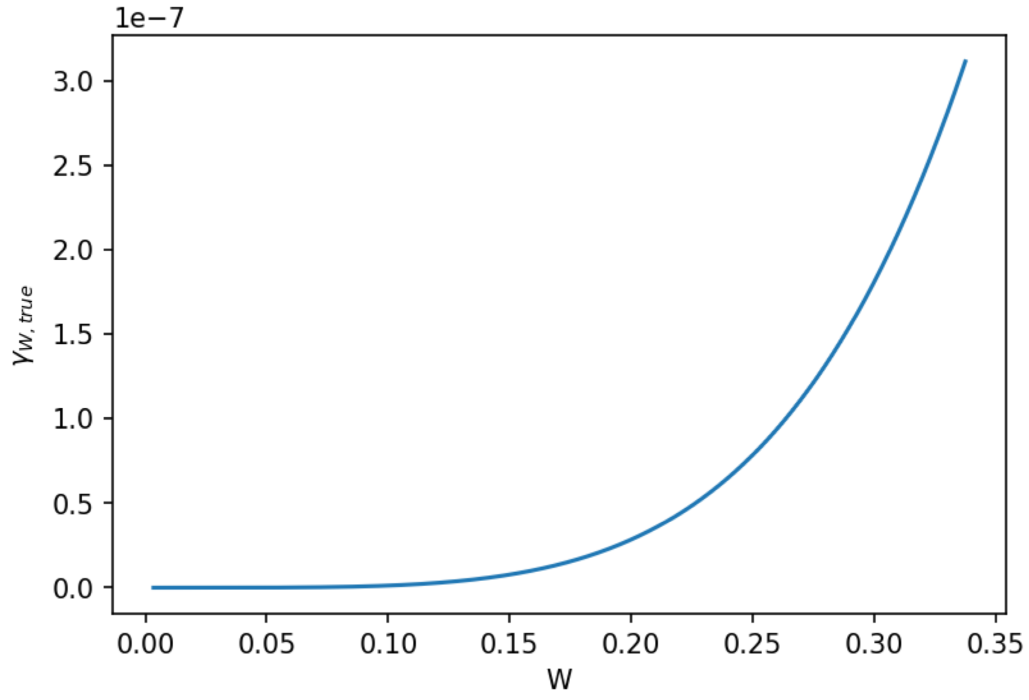
PDE identification: scenario 3

we model the evolution W_{true} of soil column using **known formulae** for $\lambda_{W,true}$ and $\gamma_{W,true}$

$$\lambda_W(W) = \frac{\gamma_{max}(1-m)}{\alpha m(\theta_s - \theta_r)} W^{\frac{1}{2}-\frac{1}{m}} * \left[\left(1 - W^{\frac{1}{m}}\right)^{-m} + \left(1 - W^{\frac{1}{m}}\right)^{m} - 2\right]$$

$$\gamma_W(W) = \gamma_{max} W^{\frac{1}{2}} \left[1 - \left(1 - W^{\frac{1}{m}}\right)^{m}\right]^2$$





Results: scenario 3

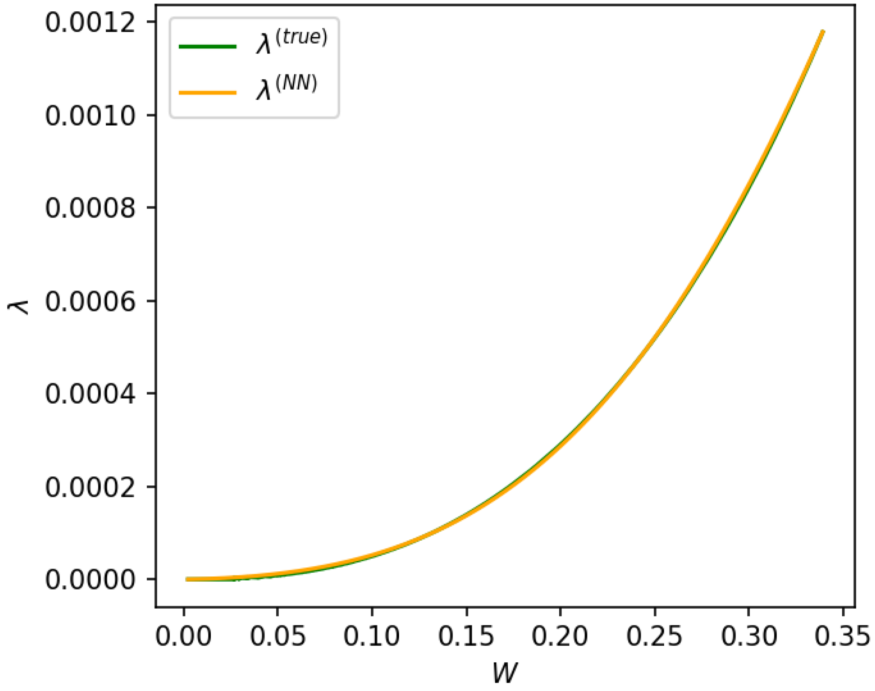
FP32

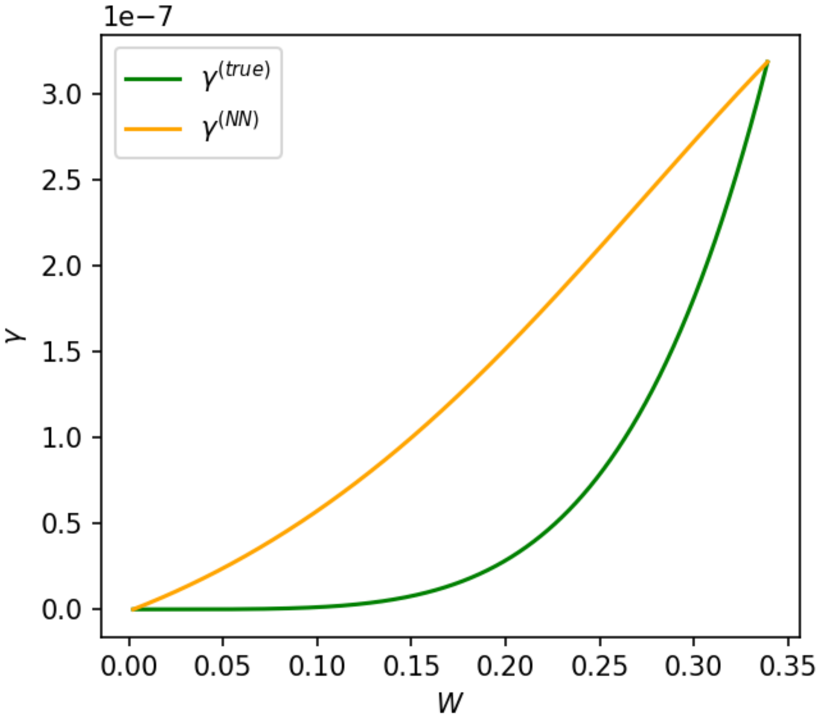
$MAPE(\lambda) = 4.4{\times}10^{-2}$

$MAPE(\gamma) = 0.94$

$RMSE\left(\frac{\partial W}{\partial t}\right) = 3.0{\times}10^{-4}$

TF: 1.3 it/sec *





* due to computationally expensive Hessian estimation: one needs to compute $\nabla_\theta \left(\frac{\partial F_{NN}(W,\theta)}{\partial W}\right)$ with backpropagation

Conclusions

- We propose a novel data-driven approach for identifying PDE systems using artificial neural networks;
- We state the problem of a PDE identification as an inverse problem since we need to infer some variables (PDE coefficients) based on the evolution of the system;
- Our approach is capable of approximating various complicated forms of PDE coefficients;
- Some flaws were detected related most likely to the well-known tendency of statistical models (neural networks included) to lose accuracy in heavy tails of training data distribution;
- Coefficients approximation quality does not seem to correlate strongly with the accuracy of PDE solution (yet to be assessed);
- Sensitivity to noise properties (considered simulating measurements errors) yet to be assessed.

Outlook

- Regularize the networks and modify training procedure to better fit the coefficients of varying **orders of magnitude;**
- Assess accurately the link between the coefficients approximation quality and PDE solution quality;
- Assess the sensitivity of the method to noise properties: spatial correlation, temporal correlation, systematic bias, *etc.*;
- Assess the sensitivity of our method to spatial resolution of the measurements;
- Apply our method to real data taken at MSU meteorological station and other locations with different types of soil.

Thank you