Optimizing a Hydrological Model with a text generating Neural Network

Simon Schmitz
SC-HPC

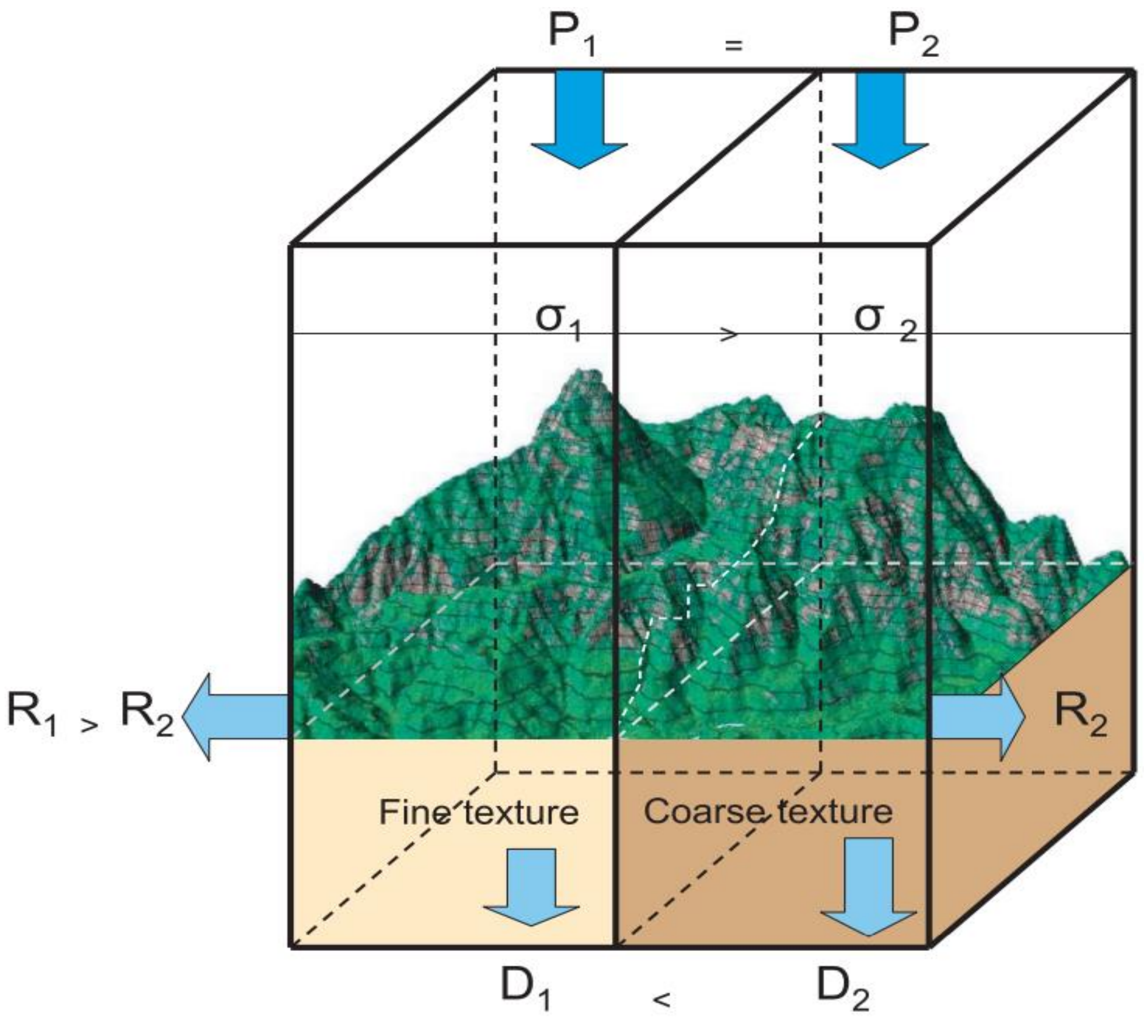
Introduction

- A hydrological model aims to describe the physical processes related to the land surface hydrology

- Based on
  - precipitation P
  - orography σ
  - fine/coarse soil texture

  the hydrological model describes:
  - run-off R
  - drainage D

- The hydrological model in this work (HTESSEL, 10.21957/yzyeh0vlw) is part of the physical parameterization package of the European Centre for Medium-Range Weather Forecasts (ECMWF)

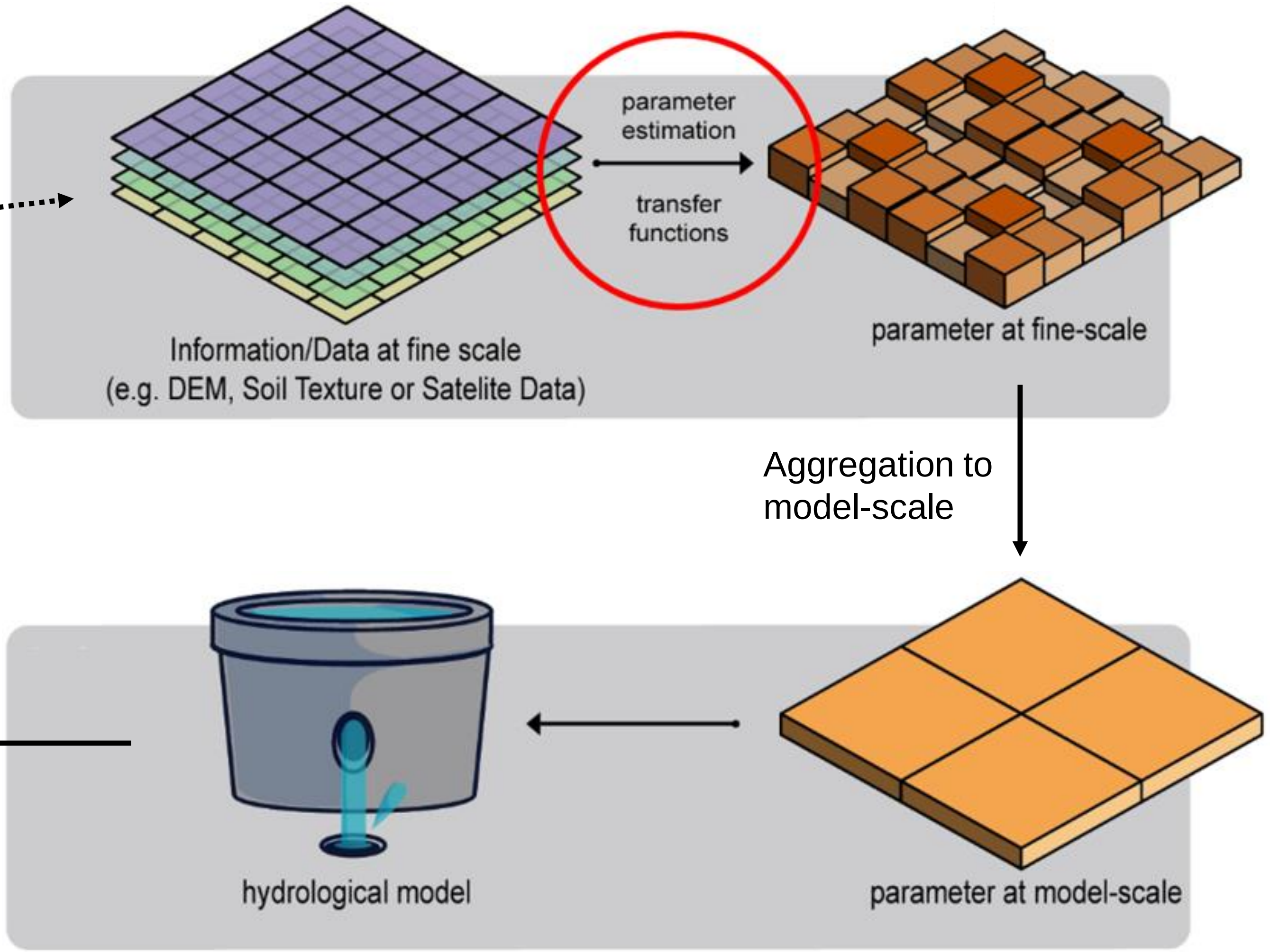


10.21957/yzyeh0vlw

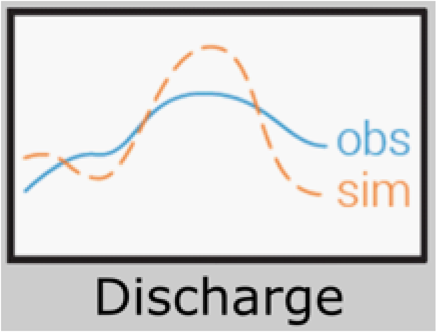
Application of the Hydrological Model

Input Data:

- Clay content in %
- Sand content in %
- Soil organic carbon in g per kg
- …

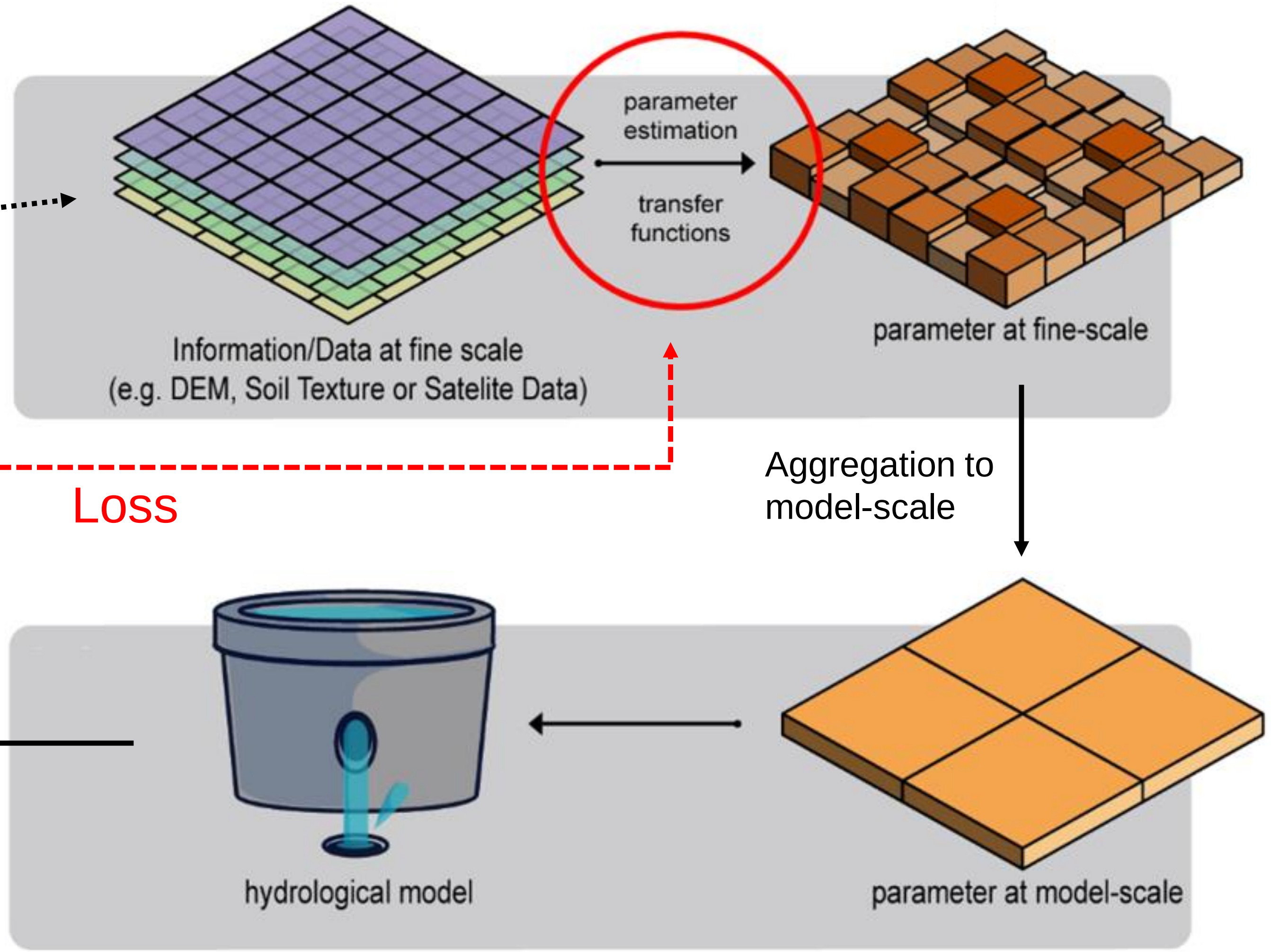


Output



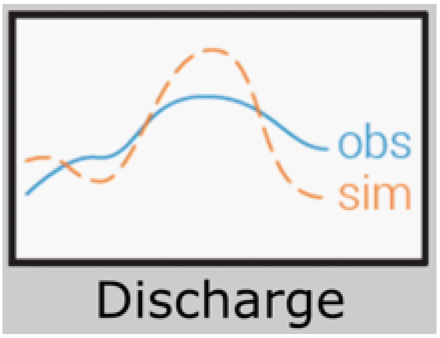
Application of the Hydrological Model

Input Data:
- Clay content in %
- Sand content in %
- Soil organic carbon in g per kg
- …



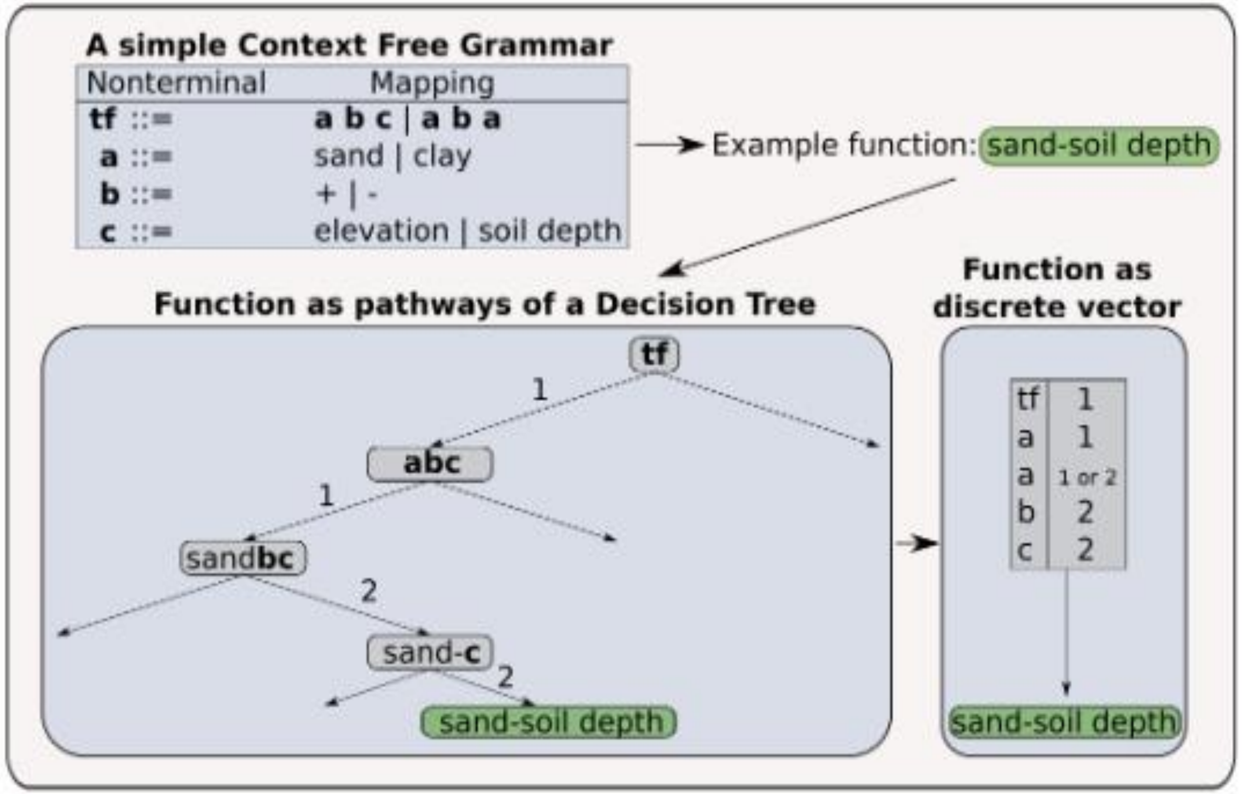
Loss

Output



How to select the Transfer Functions?

- Need to define a realm of possible transfer functions (TFs)
- Use a Context Free Grammar (CFG) to define search space of vector representations of possible TFs
- Problem: no distance metric reflects closeness of TFs



https://doi.org/10.1029/2020WR027385

How to select the Transfer Functions?

- The following TFs produce the same parameter distributions but their semantics are quite different:
  - 0.3 * sand + 1.3
  - -0.35 * clay + 1.6

- Create a numerical space which defines distance between TFs by their
  1. Semantic closeness
  2. Closeness in the resulting parameter distribution (described via cumulative probabilities)

- Any continuous global optimization method can be applied

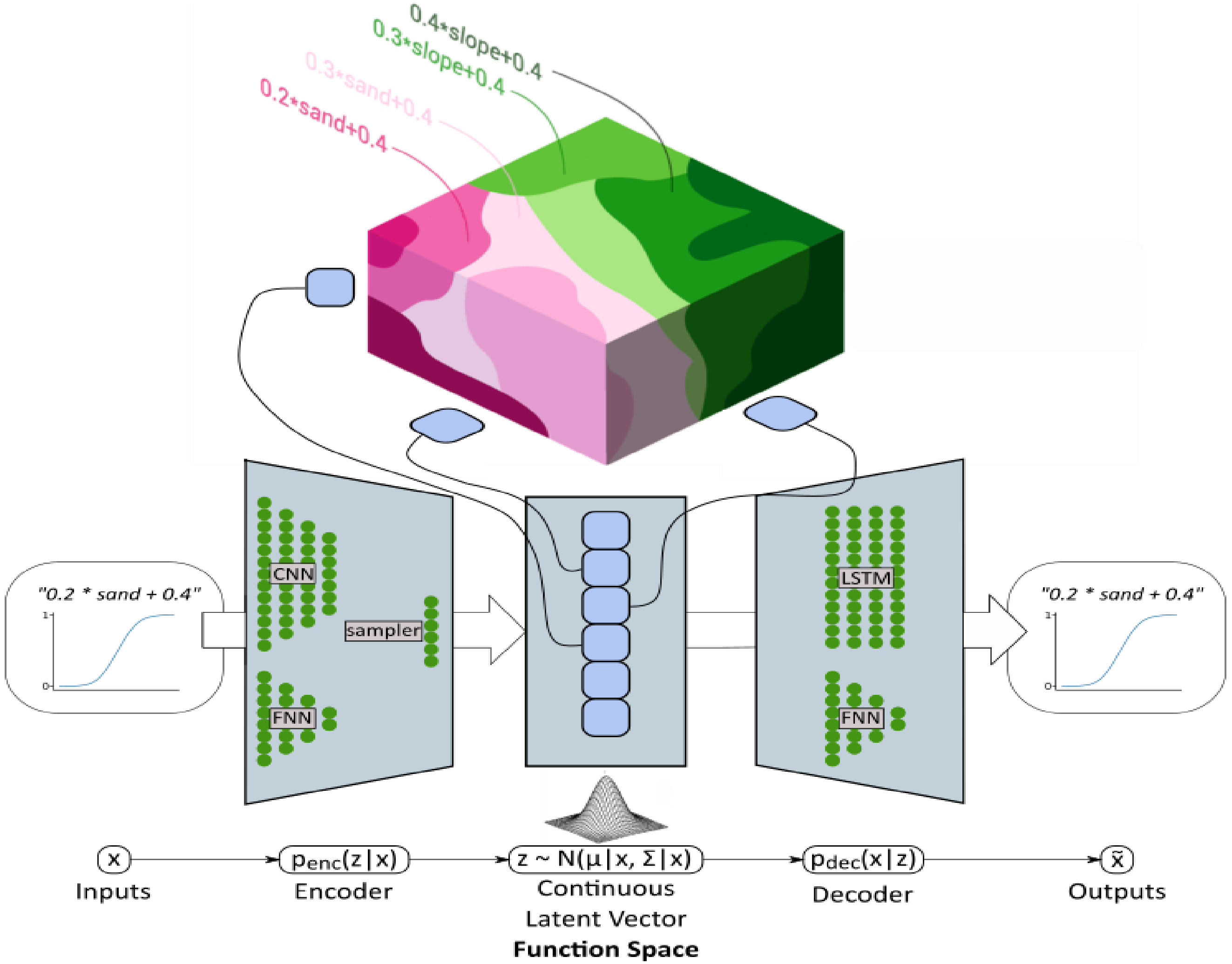
- How to create this numerical space?
  - Function Space Optimization (https://doi.org/10.1029/2020WR027385)

Function Space Optimization (FSO)

- Train a Variational Autoencoder (VAE) to encode the TFs into a latent space

- Semantic information and parameter distribution information is compressed into a 6 dimensional latent space

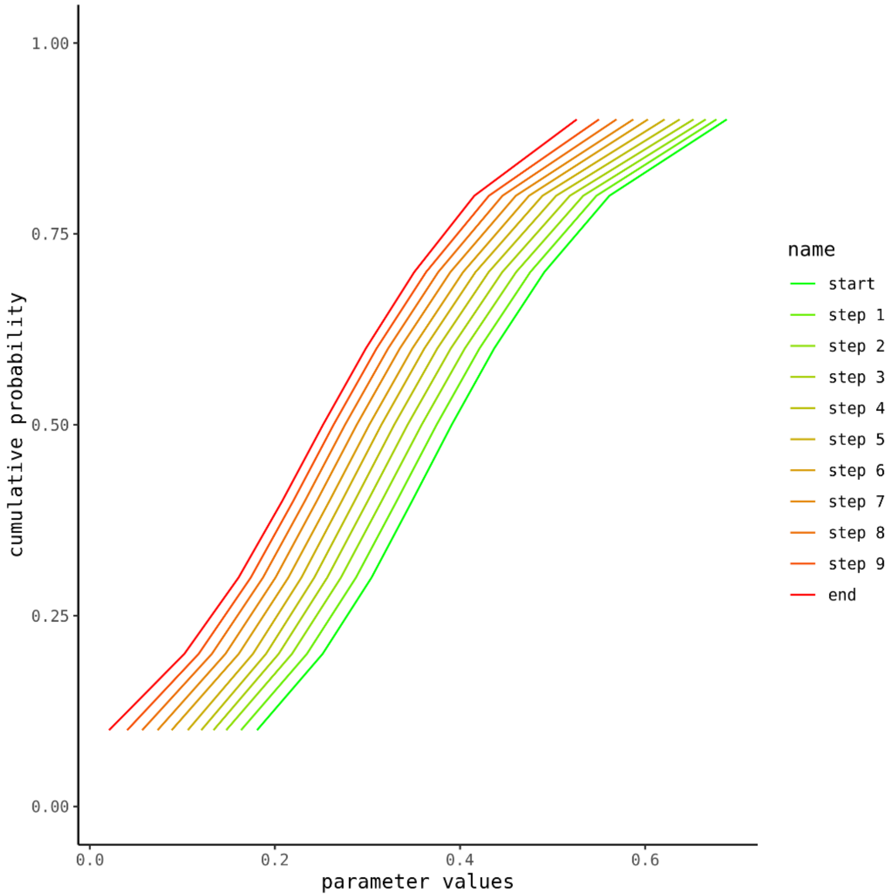
- The decoder part of the VAE can be used to sample new TFs from the latent space



https://doi.org/10.1029/2020WR027385

Results

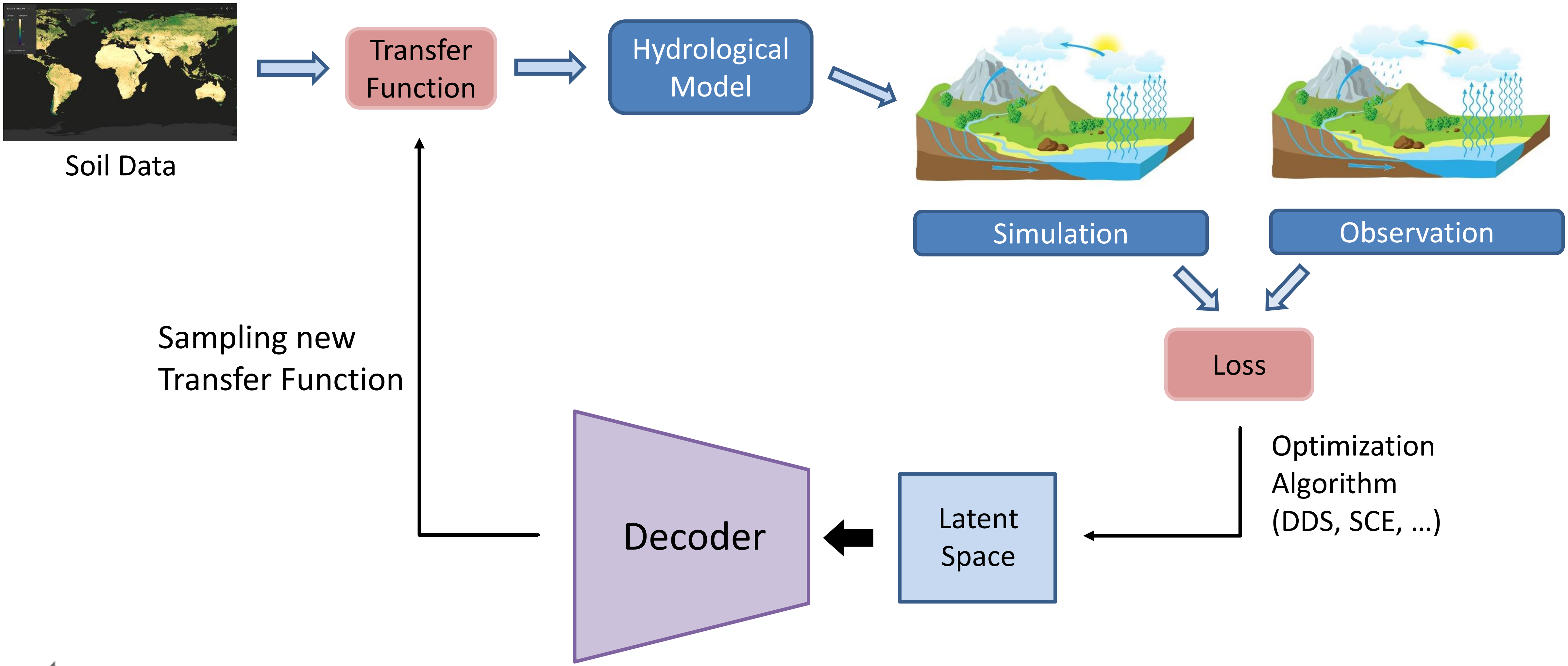
- Randomly picked a start and end point in the latent space and walked in equal steps from start to end



Sampled Functions

0.0 "BLDFIE_M * SNDPPT_M + 3.00 * TEXMHT_M / 3.00„
0.1 "CLYPPT_M * CRFVOL_M - 3.00 * TEXMHT_M ^ 3.00„
0.2 "(asin(CRFVOL_M) - TEXMHT_M) ^ 3.00„
0.3 "BLDFIE_M * CLYPPT_M - SLTPPT_M ^ 3.00„
0.4 "asin(BLDFIE_M - SLTPPT_M) ^ 3.00„
0.5 "CLYPPT_M * (PHIHOX_M - SLTPPT_M) ^ 3.00„
0.6 "CLYPPT_M * (BLDFIE_M - CRFVOL_M) ^ 3.00„
0.7 "BLDFIE_M * ORCDRC_M - sin(ORCDRC_M) ^ SNDPPT_M„
0.8 "BLDFIE_M * SLTPPT_M„
0.9 "SLTPPT_M * (BLDFIE_M * SLTPPT_M - 3.00)„
1.0 "BLDFIE_M * (BLDFIE_M * SLTPPT_M ^ 1.00)"

Overview of Optimization Procedure



Conclusion

- Optimization of the hydrological model is important for weather forecast

- Optimization is dependent on the selection of a transfer function

- Needed to transform the application of the hydrological model into an optimizable problem

- Trained a variational autoencoder to encode TFs based on:
  - Semantic closeness
  - Closeness in the resulting parameter distribution

- Used latent space and decoder part of the VAE to search for new TFs