Normalizing Flows: Uncertainty Quantification in Earth Machine Learning

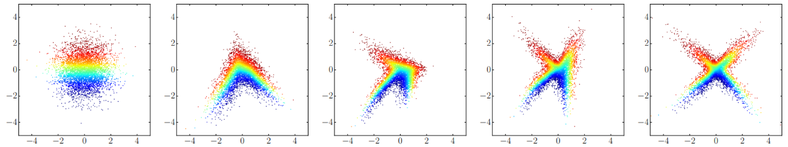
Peter Van Katwyk[1] , Karianne Bergen[1, 2]

[1] Department of Earth, Environmental, and Planetary Sciences, Brown University, [2] Data Science Initiative, Brown University

BROWN

MOTIVATION

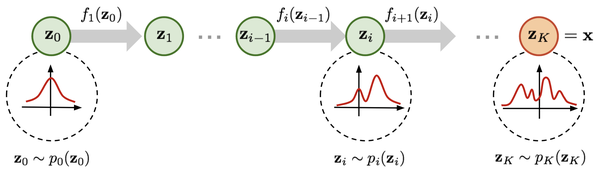
- **Machine learning (ML) is a powerful tool** that is commonly applied to **geophysical problems**
- **Uncertainty quantification** in machine learning models, which is achieved through probabilistic machine learning, is **essential in computer-aided scientific decision making** and gaining insights into **underlying data distributions** and complex earth systems
- **A major limitation** of traditional machine learning models is that they **only provide a point estimate** for the predicted value and **fail to provide uncertainty** associated with that prediction
- **Normalizing flows (NF)** are a class of **probabilistic generative neural network models** that accurately predict as well as quantify uncertainty by approximating the underlying data distribution



Example of a 4-step flow transforming samples from a standard-normal base density (left) to a cross-shaped target density (right), modified from Papamakarios, 2021

METHOD: Normalizing Flows

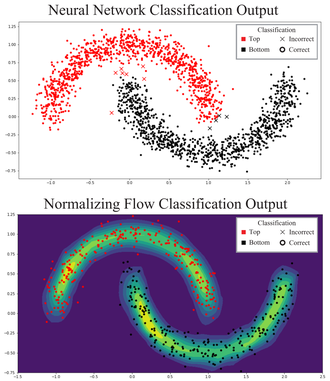
- A normalizing flow consists of **transformations from a simple probability distribution** (e.g. a standard normal distribution) into a more **complex distribution** (e.g. data distribution) through a series of tractable, differentiable, and **invertible functions**
- The unknown density can then be approximated and sampled by **sampling from the initial** simple probability distribution and **applying the trained transformations**



Normalizing flow transformation from a simple distribution (left) to a complex distribution (right), Weng, 2018

EXAMPLE: Uncertainty & Density Estimation

- **Objective**
  - Develop a model to classify each point in the scikit-learn synthetic "Moons" dataset
  - Generate uncertainty quantification using density estimation for each classification
- **Neural Network Classifier vs. Normalizing Flows**
  - Use standard dense neural network architecture to classify moons (0=top, 1=bottom)
  - Train a conditional normalizing flow, which learns distributions of each classified label, to classify as well as provide log probability uncertainty quantification and compare



Color map: Viridis - blues indicate high uncertainty, yellows indicate low.

- **Neural Network**
  - Two-layer dense network
  - Accuracy: **98.72%**
  - Training time: **26.37 seconds**
  - **Single, point-estimate classification**
- **Normalizing Flow**
  - Masked Autoregressive Flow
  - Accuracy: **100.00%**
  - Training time: **68.19 seconds**
  - **Multiple outputs**
    - **Point-estimate classification**
    - **Log-probability uncertainty**

Presenter Information:

Peter is a first-year Ph.D. student at Brown University studying Scientific Machine Learning. He specializes in using probabilistic ML to quantify uncertainty in natural phenomena. His goal is to bridge the gap between data science and domain expertise within the earth sciences to create more trustworthy, accurate, and predictive ML models.

Contact:

Email: peter_van_katwyk@brown.edu
Linkedin: https://www.linkedin.com/in/pvankatwyk/
GitHub: https://github.com/pvankatwyk

Resources:

Code: https://github.com/pvankatwyk/NF_SAGEGAGE

References:

Bergen, Karianne J., et al. "Machine learning for data-driven discovery in solid Earth geoscience." Science 363.6433 (2019): eaau0323.

Edwards, T.L., Nowicki, S., Marzeion, B. et al. Projected land ice contributions to twenty-first-century sea level rise. Nature 593, 74–82 (2021).

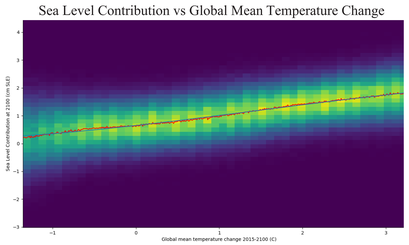
Papamakarios, George, Theo Pavlakou, and Iain Murray. "Masked autoregressive flow for density estimation." Advances in neural information processing systems 30 (2017).

Papamakarios, George, et al. "Normalizing Flows for Probabilistic Modeling and Inference." J. Mach. Learn. Res. 22.57 (2021): 1-64.

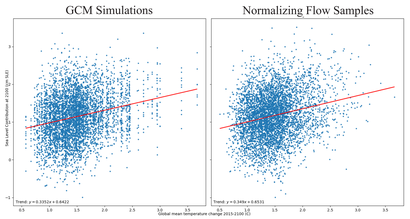
Weng, Lilian. "Flow-Based Deep Generative Models." lilianweng.github.io, 13 Oct. 2018, https://lilianweng.github.io/posts/2018-10-13-flow-models/

RESULTS: Uncertainty in Climate Emulators

- Understanding and quantifying uncertainty in climate modeling is crucial for understanding current climate trajectories and projecting scenarios into the future
- Emulators are used to approximate the large-scale, computationally expensive Global Climate Model (GCM) simulations
- Edwards et al., 2019 emulates ice sheet and glacial models to predict the land ice contribution to global mean sea level change using a Gaussian Process (GP)



Emulating the effect of global air temperature change on sea level contribution. Uncertainty bounds are estimated using normalizing flows. Color map: Viridis - blues indicate high uncertainty, yellows indicate low.



Sampling from unknown GCM simulation density using normalizing flows. Trends plotted for comparison.

- **Gaussian Processes** can be effective but…
  - they are **slow to train** - $O(n^3)$
  - they have **limited expressiveness** outside standard kernels
- Ongoing results show that **normalizing flow emulators** are an efficient alternative to GP models because they **remove the computational bottleneck** while maintaining predictive power
- Normalizing flow emulators can be used to both **sample** from the GCM simulation distribution and **provide uncertainty** on sea level contribution predictions over given ranges of potential global temperature changes

FUTURE WORK: Geophysical Applications

- Normalizing flows could be a useful tool in geophysical applications because…
  - Geophysical datasets contain quantitative and continuous measurements
  - Data is often high-dimensional
  - Aleatoric uncertainty (due to natural stochasticity) is a driving factor in geophysical outcomes
- Potential geophysical applications include **uncertainty quantification in computer-aided interpretation** (P&S wave arrival times, seismic horizons), **data processing** (cleaning, interpolating), and more