ACCORD

A Consortium for COnvection-scale modelling Research and Development

Towards a data-driven turbulence and cloud-precipitation microphysics

a review

ACCORD ASW, 4-8 April 2022, Ljubljana, Slovenia

Overview

- **Data driven modeling - general**
- **AI in NWP - current status**
- **Data-driven turbulence modeling**
- **Data-driven cloud-precipitation modeling**
- **Concluding remarks**

Goal

1. To **filter out and present** the most relevant literature about data-driven modeling in NWP, with special focus on turbulence and cloud-precipitation microphysics

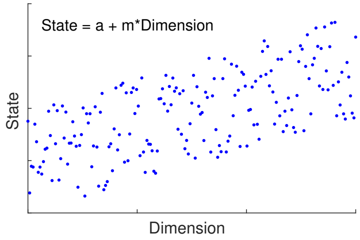
2. To **give an opinion** on future direction regarding the use of data-driven modeling in ACCORD

Motivation

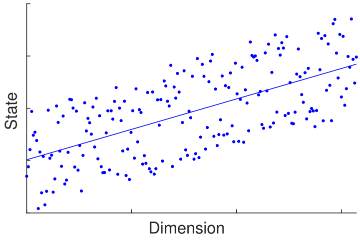
1. **Why data-driven modeling?**

- The availability of large amount of DATA and high-computing infrastructure makes solving inverse problems to be suitable in representing complex processes

Create a function for the state

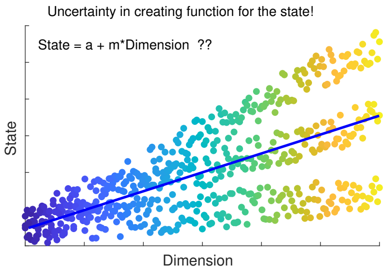


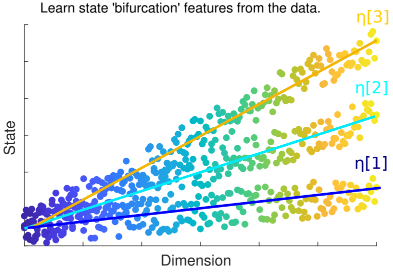
Fit the data to represnt the state



1. **Why data-driven modeling?**

- The availability of large amount of DATA and high-computing infrastructure makes solving inverse problems to be suitable in representing complex processes





Data-driven modeling - general

A general form of all data-driven models can be written as follows:

$$\tilde{\mathcal{M}} \equiv \mathcal{M}(\mathbf{w}; \mathcal{P}(\mathbf{w}); \mathbf{c}; \boldsymbol{\theta}; \boldsymbol{\delta}; \boldsymbol{\epsilon_\theta}),$$

- $\mathcal{M}$ represents a function (**a model**) of an array of independent variables **w** linked with a set of algebraic or differential operators $\mathcal{P}$ and parameters **c**
- $\boldsymbol{\theta}$ is the **data**
- $\boldsymbol{\epsilon_\theta}$ is the uncertainty in the data
- $\boldsymbol{\delta}$ represents the **discrepancy** of the model $\mathcal{M}$ to represent the *true* state; usually it is defined in terms of a **set of features** $\boldsymbol{\eta}$ given from a prior knowledge, constrains or directly from data

The idea of data-driven models is to use data to construct **calibrated models** $\tilde{\mathcal{M}} \neq \mathcal{M}$ and the goal is to predict quantities of interest $\mathbf{q}(\tilde{\mathcal{M}})$ based on proper model coefficients **c**.

Data-driven modeling - general

Finding a functional form of $\boldsymbol{\delta}(\boldsymbol{\eta})$:

$$\boldsymbol{\delta}(\boldsymbol{\eta}) = \boldsymbol{W}\boldsymbol{\eta} + \boldsymbol{\beta},$$

W and $\boldsymbol{\beta}$ are the weight matrix and the bias vector.

- solving optimization problem (e.g., *maximum a posteriori* (**MAP**) or *least squares* (**LS**) (e.g. Aster et al. 2018)
- the procedure is called training or **learning** of a model
- basic building block in **neural networks** - a part of the **machine learning (ML)** techniques, which belongs to a more general field of **artificial intelligence (AI)**.

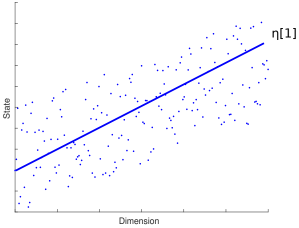


Figure: Solving linear problem.

Data-driven modeling - general

Finding a functional form of $\boldsymbol{\delta(\eta)}$:

$$\boldsymbol{\delta(\eta)} = \boldsymbol{W\eta} + \boldsymbol{\beta},$$

W and $\boldsymbol{\beta}$ are the weight matrix and the bias vector.

- solving optimization problem (e.g., *maximum a posteriori* (**MAP**) or *least squares* (**LS**) (e.g. Aster et al. 2018)
- the procedure is called training or **learning** of a model
- basic building block in **neural networks** - a part of the **machine learning (ML)** techniques, which belongs to a more general field of **artificial intelligence (AI)**.

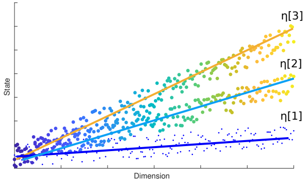


Figure: Solving complex (nonlinear) problem

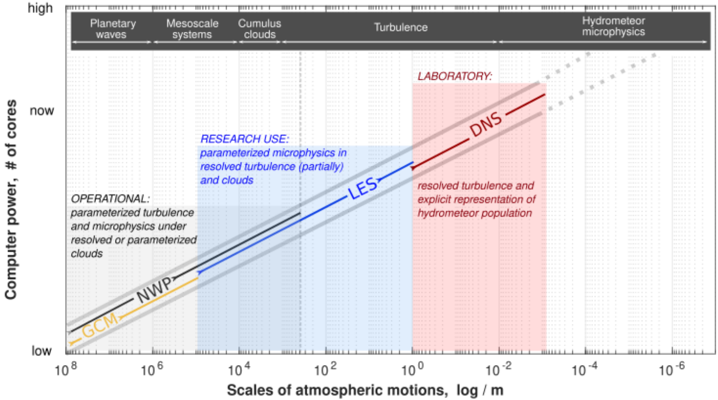
$$\boldsymbol{\delta(\eta)} = \boldsymbol{W^{(l)}}\sigma(\boldsymbol{W^{(l-1)}\eta} + \boldsymbol{\beta^{(l-1)}}) + \boldsymbol{\beta^{(l)}}$$

Deep Neural Networks

AI in NWP (Dewitte et al. 2021, a review)

- AI is already revolutionizing all components of the NWP systems, **from observations to post-processing**
- Aiming at **fully replacing the physics-based NWP** approaches (Kim et al. 2020; Zheng et al. 2020)
- My focus in on **turbulence and cloud-precipitation microphysics**

Turbulence and microphysics modeling



- Forecast uncertainty due to small-scale turbulence and cloud-precipitation microphysics parameterization

Turbulence modeling (Duraisamy et al. 2019, a review) in the Age of Data

Four levels of uncertainty (UL):

UL1: The **spatio-temporal averaging procedure** of the Navier-Stokes (NS) equations, resulting in undetermined system, $\langle\mathcal{NS}(\cdot)\rangle \neq \mathcal{NS}(\langle\cdot\rangle)$; **requires modeling assumptions** to be closed (the well-known **turbulence closure** problem)

UL2: Selection (design) of the functional form of the closure model:

$$\langle\mathcal{NS}(\cdot)\rangle = \mathcal{NS}(\langle\cdot\rangle) + \mathcal{M}(\cdot)$$

UL3: The functional forms within the model, typically representing some physical processes (e.g., convection, waves, boundary dynamics)

UL4: Parameters, coefficients or numbers used in the model

Turbulence modeling (Duraisamy et al. 2019, a review) in the Age of Data

The main idea is to derive a model $\tilde{\mathcal{M}} = \mathcal{M}(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is usually DNS data, which explicitly includes the predefined (or learned) discrepancy function $\boldsymbol{\delta}$.

UL4: **Calibrating model coefficients**

- Introducing simple discrepancies (Cheung et al. 2011; Edeling et al. 2014a; Edeling et al. 2014b; Lefantzi et al. 2015; Oliver and Moser 2011; Ray et al. 2016, 2018)

Turbulence modeling (Duraisamy et al. 2019, a review) in the Age of Data

The main idea is to derive a model $\tilde{\mathcal{M}} = \mathcal{M}(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is usually DNS or LES data, which explicitly includes the predefined (or learned) discrepancy function $\boldsymbol{\delta}$.

UL4: **Calibrating model coefficients**

- Introducing simple discrepancies (Cheung et al. 2011; Edeling et al. 2014a; Edeling et al. 2014b; Lefantzi et al. 2015; Oliver and Moser 2011; Ray et al. 2016, 2018)

UL3: **Accounting for the discrepancy in the Reynolds stress**

- ML to reconstruct discrepancies in the anisotropy tensor of the Reynolds stress (Tracey et al. 2013)
- Also magnitude and orientation of the Reynolds stress tensor (Wu et al. 2017)
- A comprehensive Physics-Informed Machine Learning (PIML) framework for predictive turbulence modeling (Wu et al. 2018)

Turbulence modeling (Duraisamy et al. 2019, a review) in the Age of Data

The main idea is to derive a model $\tilde{\mathcal{M}} = \mathcal{M}(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is usually DNS or LES data, which explicitly includes the predefined (or learned) discrepancy function $\boldsymbol{\delta}$.

UL4: **Calibrating model coefficients**

- Introducing simple discrepancies (Cheung et al. 2011; Edeling et al. 2014a; Edeling et al. 2014b; Lefantzi et al. 2015; Oliver and Moser 2011; Ray et al. 2016, 2018)

UL3: **Accounting for the discrepancy in the Reynolds stress**

- ML to reconstruct discrepancies in the anisotropy tensor of the Reynolds stress (Tracey et al. 2013)
- Also magnitude and orientation of the Reynolds stress tensor (Wu et al. 2017)
- A comprehensive Physics-Informed Machine Learning (PIML) framework for predictive turbulence modeling (Wu et al. 2018)

UL2 **Modeling the Reynolds stress tensor**

- Deep neural network algorithms opened a new path in the research of data-driven universal turbulence models (Beck et al. 2018)

Modeling cloud-precipitation micriphysics (Morrison et al. 2020, a review)

Uncertainty and challenges:

1. Uncertainty in **parameterizations of hydrometeor population** due the impossibility of modeling all hydrometeors individually in a cloud

2. Uncertainty in **process rates** due to limited cloud physics knowledge at the scale of individual hydrometeors

Modeling cloud-precipitation micriphysics (Morrison et al. 2020, a review)

Purely physical or theoretical models

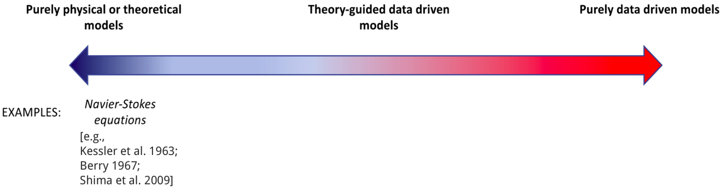
Theory-guided data driven models

Purely data driven models

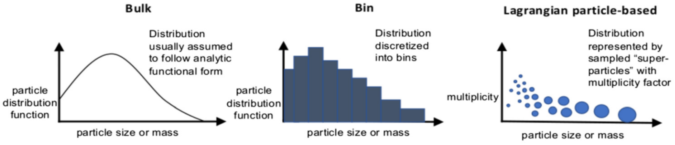
EXAMPLES:

Navier-Stokes equations
[e.g.,
Kessler et al. 1963;
Berry 1967;
Shima et al. 2009]

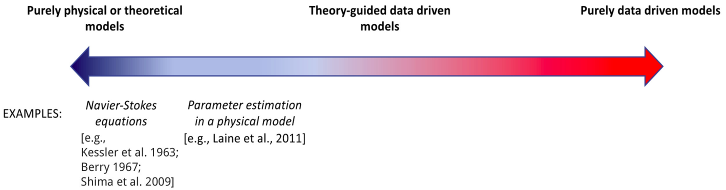
Modeling cloud-precipitation micriphysics (Morrison et al. 2020, a review)



- Purely physics: **bulk** (Kessler et al, 1963; ...), **bin** (Berry 1967; ...), **Lagrangian particle-based** (e.g., Shima et al. 2009; 2019)



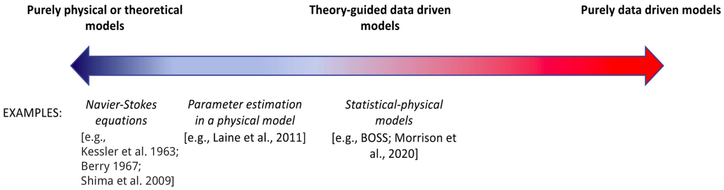
Modeling cloud-precipitation micriphysics (Morrison et al. 2020, a review)



- Purely physics: **bulk** (Kessler et al, 1963; ...), **bin** (Berry 1967; ...), **Lagrangian particle-based** (e.g., Shima et al. 2009; 2019)

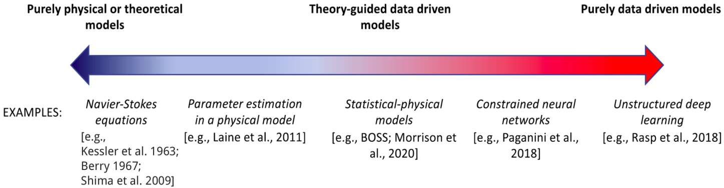
- Ensemble prediction and parameter estimation (Laine et al. 2011)

Modeling cloud-precipitation micriphysics (Morrison et al. 2020, a review)



- Purely physics: **bulk** (Kessler et al, 1963; ...), **bin** (Berry 1967; ...), **Lagrangian particle-based** (e.g., Shima et al. 2009; 2019)
- **Ensemble prediction** and parameter estimation (Laine et al. 2011)
- **Physics-guided data-driven** models (Morrison et al. 2020)

Modeling cloud-precipitation micriphysics (Morrison et al. 2020, a review)



- Purely physics: **bulk** (Kessler et al, 1963; ...), **bin** (Berry 1967; ...), **Lagrangian particle-based** (e.g., Shima et al. 2009; 2019)
- **Ensemble prediction** and parameter estimation (Laine et al. 2011)
- **Physics-guided data-driven** models (Morrison et al. 2020)
- **Purely data-driven** modeling (e.g. Rasp et al. 2018)

Concluding notes

- Data-driven methods are **important and extremely useful** tools for improving the NWP

- **Best results** are achieved when used **together with the physics**

- Data-driven models **highly-dependent on the DATA** (quality and quantity)

'Filtered' references

Aster, R. C., B. Borchers, and C. H. Thurber (2018). *Parameter estimation and inverse problems*. Elsevier.

Dewitte, S., J. P. Cornelis, R. Müller, and A. Munteanu (2021). "Artificial Intelligence Revolutionises Weather Forecast, Climate Monitoring and Decadal Prediction". In: *Remote Sensing* 13.16. DOI: 10.3390/rs13163209.

Duraisamy, K., G. Iaccarino, and H. Xiao (2019). "Turbulence Modeling in the Age of Data". In: *Annual Review of Fluid Mechanics* 51.1, pp. 357–377. DOI: 10.1146/annurev-fluid-010518-040547.

Morrison, H. et al. (2020). "Confronting the Challenge of Modeling Cloud and Precipitation Microphysics". In: *Journal of Advances in Modeling Earth Systems* 12.8, e2019MS001689. DOI: https://doi.org/10.1029/2019MS001689.