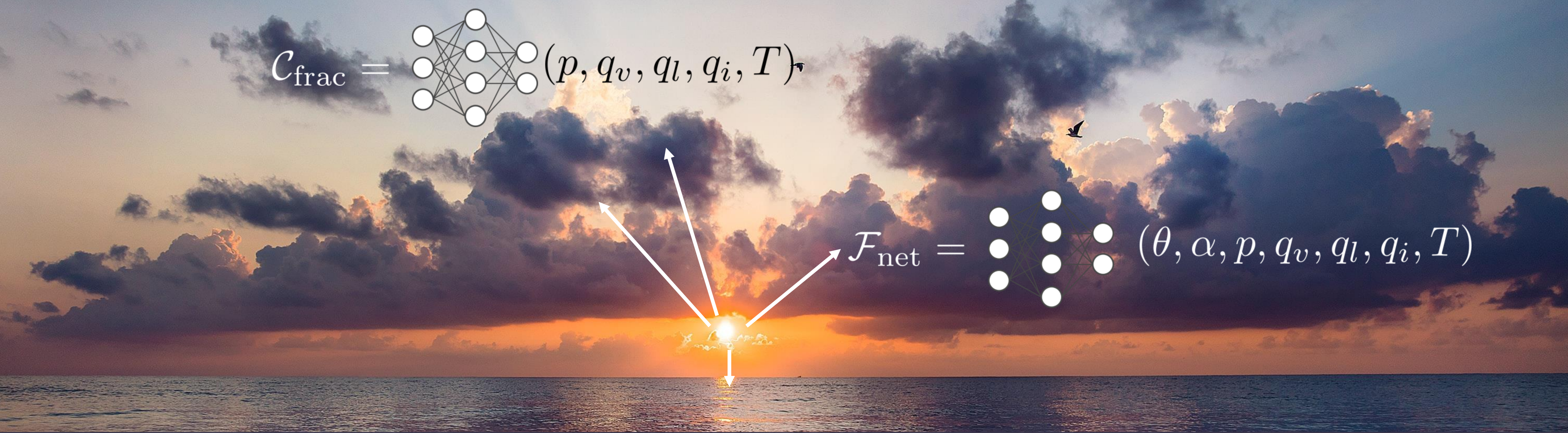

∂³AWN
∂ata-∂riven
Atmospheric & Water
∂yNamics
UNIL | Université de Lausanne
Climate Informatics 2023
April 19-21st
University of Cambridge, UK

∂³AWN
∂ata-∂riven
Atmospheric & Water
∂yNamics

Systematically Generating Hierarchies of Machine-Learning Models, from Equation Discovery to Deep Neural Networks

UNIL | Université de Lausanne



Presenter: Tom Beucler (UNIL) – AMS Annual 2023

Co-authors:

Arthur Grundner (DLR)

Sara Shamekh (Columbia)

Ryan Lagerquist (CIRA/NOAA)

Motivation: Added value of ML for weather/climate is measurable (↓RMSE), but challenging to understand

FourCastNet: A Global Data-driven High-resolution Weather Model using Adaptive Fourier Neural Operators

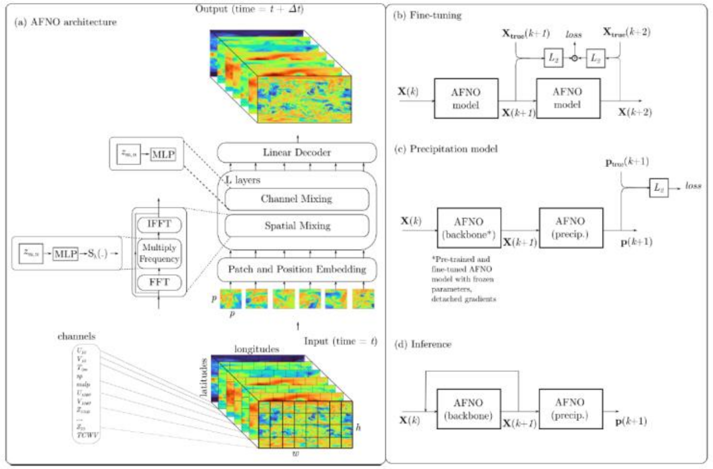
A Preprint

Jaideep Pathak NVIDIA Corporation Santa Clara, CA 95051
Shashank Subramanian Lawrence Berkeley National Laboratory Berkeley, CA 94720
Peter Harrington Lawrence Berkeley National Laboratory Berkeley, CA 94720
Sanjeev Raja University of Michigan Ann Arbor, MI 48109
Ashesh Chattopadhyay Rice University Houston, TX 77005
Morteza Mardani NVIDIA Corporation Santa Clara, CA 95051
Thorsten Kurth NVIDIA Corporation Santa Clara, CA 95051
David Hall NVIDIA Corporation Santa Clara, CA 95051
Zongyi Li California Institute of Technology Pasadena, CA 91125 NVIDIA Corporation Santa Clara, CA 95051
Kamyar Azizzadenesheli Purdue University West Lafayette, IN 47907
Pedram Hassanzadeh Rice University Houston, TX 77005
Karthik Kashinath NVIDIA Corporation Santa Clara, CA 95051
Animashree Anandkumar California Institute of Technology Pasadena, CA 91125 NVIDIA Corporation Santa Clara, CA 95051

February 24, 2022

Abstract

FourCastNet, short for *Four*ier Fore*Cast*ing Neural *Net*work, is a global data-driven weather forecasting model that provides accurate short to medium-range global predictions at 0.25° resolution. FourCastNet accurately forecasts high-resolution, fast-timescale variables such as the surface wind speed, precipitation, and atmospheric water vapor. It has important implications for planning wind energy resources, predicting extreme weather events such as tropical cyclones, extra-tropical cyclones, and atmospheric rivers. FourCastNet matches the forecasting accuracy of the ECMWF Integrated Forecasting System (IFS), a state-of-the-art Numerical Weather Prediction (NWP) model, at short lead times for large-scale variables, while outperforming IFS for small-scale variables, including precipitation. FourCastNet generates a week-long forecast in less than 2 seconds, orders of magnitude faster than IFS. The speed of FourCastNet enables the creation of rapid and inexpensive large-ensemble forecasts with thousands of ensemble-members for improving probabilistic forecasting. We discuss how data-driven deep learning models such as FourCastNet are a valuable addition to the meteorology toolkit to aid and augment NWP models.

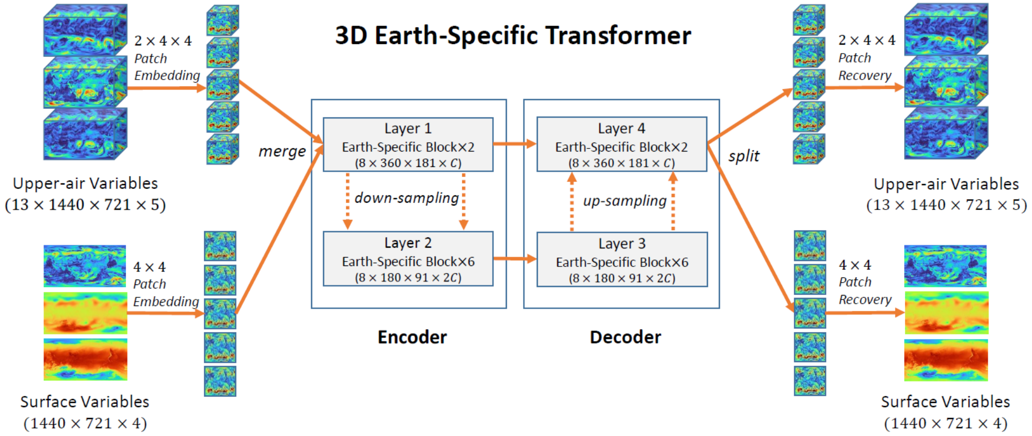


Pangu-Weather: A 3D High-Resolution System for Fast and Accurate Global Weather Forecast

Kaifeng Bi, Lingxi Xie, Hengheng Zhang, Xin Chen, Xiaotao Gu, and Qi Tian✉, *Fellow, IEEE*

Abstract—In this paper, we present Pangu-Weather, a deep learning based system for fast and accurate global weather forecast. For this purpose, we establish a data-driven environment by downloading 43 years of hourly global weather data from the 5th generation of ECMWF reanalysis (ERA5) data and train a few deep neural networks with about 256 million parameters in total. The spatial resolution of forecast is $0.25^\circ \times 0.25^\circ$, comparable to the ECMWF Integrated Forecast Systems (IFS). More importantly, for the first time, an AI-based method outperforms state-of-the-art numerical weather prediction (NWP) methods in terms of accuracy (latitude-weighted RMSE and ACC) of all factors (*e.g.*, geopotential, specific humidity, wind speed, temperature, *etc.*) and in all time ranges (from one hour to one week). There are two key strategies to improve the prediction accuracy: (i) designing a 3D Earth Specific Transformer (3DEST) architecture that formulates the height (pressure level) information into cubic data, and (ii) applying a hierarchical temporal aggregation algorithm to alleviate cumulative forecast errors. In deterministic forecast, Pangu-Weather shows great advantages for short to medium-range forecast (*i.e.*, forecast time ranges from one hour to one week). Pangu-Weather supports a wide range of downstream forecast scenarios, including extreme weather forecast (*e.g.*, tropical cyclone tracking) and large-member ensemble forecast in real-time. Pangu-Weather not only ends the debate on whether AI-based methods can surpass conventional NWP methods, but also reveals novel directions for improving deep learning weather forecast systems.

Index Terms—Numerical Weather Prediction, Deep Learning, Medium-range Weather Forecast.

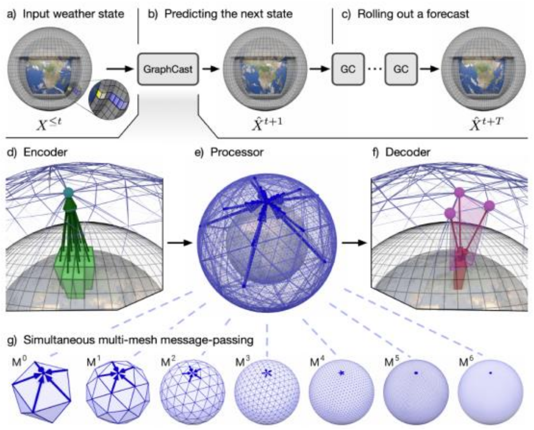


GraphCast: Learning skillful medium-range global weather forecasting

Remi Lam[*,1], Alvaro Sanchez-Gonzalez[*,1], Matthew Willson[*,1], Peter Wirnsberger[*,1], Meire Fortunato[*,1], Alexander Pritzel[*,1], Suman Ravuri[1], Timo Ewalds[1], Ferran Alet[1], Zach Eaton-Rosen[1], Weihua Hu[1], Alexander Merose[2], Stephan Hoyer[2], George Holland[1], Jacklynn Stott[1], Oriol Vinyals[1], Shakir Mohamed[1] and Peter Battaglia[1]
[*]equal contribution, [1]DeepMind, [2]Google

We introduce a machine-learning (ML)-based weather simulator—called "GraphCast"—which outperforms the most accurate deterministic operational medium-range weather forecasting system in the world, as well as all previous ML baselines. GraphCast is an autoregressive model, based on graph neural networks and a novel high-resolution multi-scale mesh representation, which we trained on historical weather data from the European Centre for Medium-Range Weather Forecasts (ECMWF)'s ERA5 reanalysis archive. It can make 10-day forecasts, at 6-hour time intervals, of five surface variables and six atmospheric variables, each at 37 vertical pressure levels, on a 0.25° latitude-longitude grid, which corresponds to roughly 25×25 kilometer resolution at the equator. Our results show GraphCast is more accurate than ECMWF's deterministic operational forecasting system, HRES, on 90.0% of the 2760 variable and lead time combinations we evaluated. GraphCast also outperforms the most accurate previous ML-based weather forecasting model on 99.2% of the 252 targets it reported. GraphCast can generate a 10-day forecast (35 gigabytes of data) in under 60 seconds on Cloud TPU v4 hardware. Unlike traditional forecasting methods, ML-based forecasting scales well with data: by training on bigger, higher quality, and more recent data, the skill of the forecasts can improve. Together these results represent a key step forward in complementing and improving weather modeling with ML, open new opportunities for fast, accurate forecasting, and help realize the promise of ML-based simulation in the physical sciences.

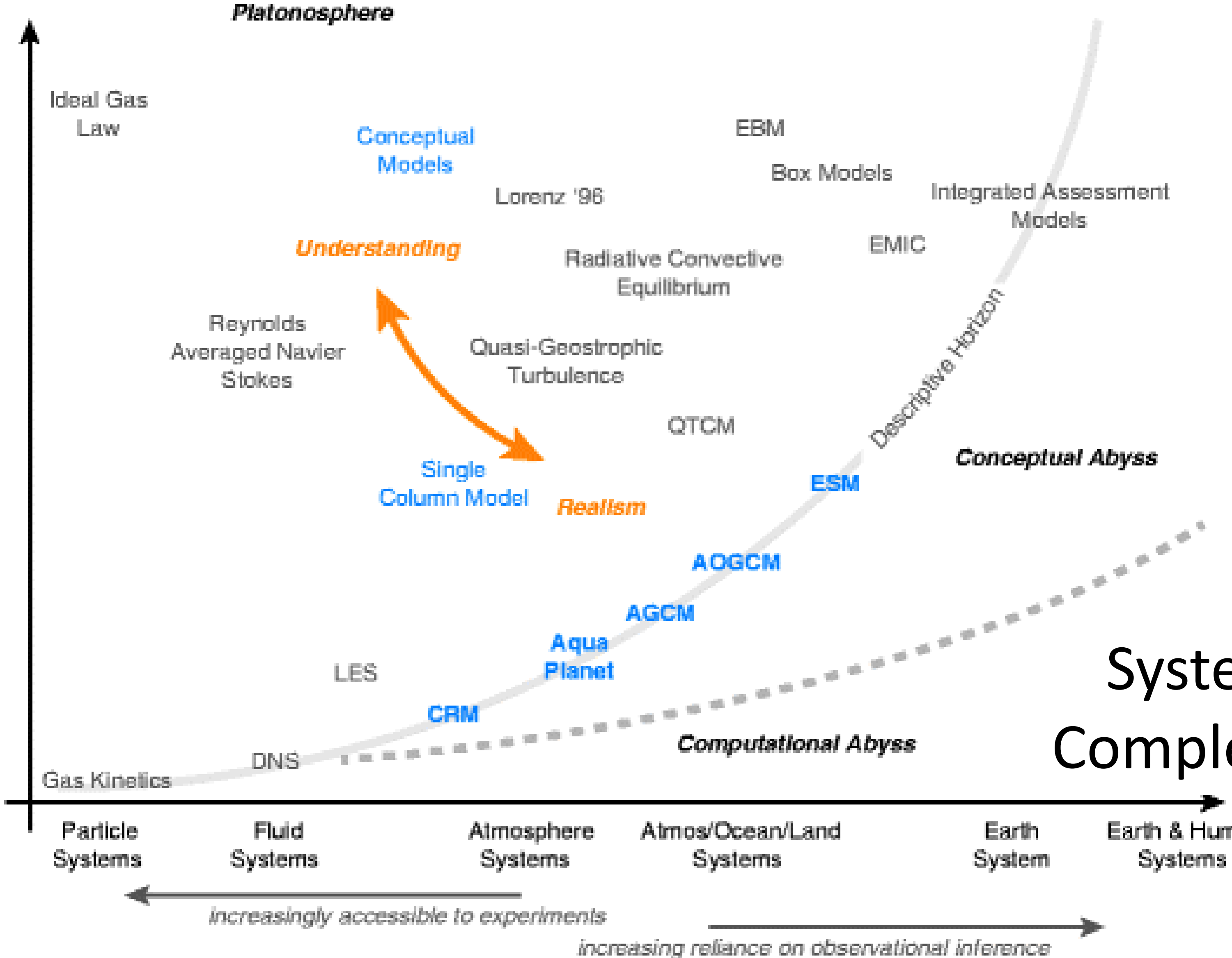
Keywords: Weather forecasting, ECMWF, ERA5, HRES, learning simulation, graph neural networks



See: Kurth et al. (2022), Keisler (2022), Pathak et al. (2022), Bi et al. (2022), Lam et al. (2022)

Analogy: Climate Model Hierarchies connect our fundamental understanding with model prediction

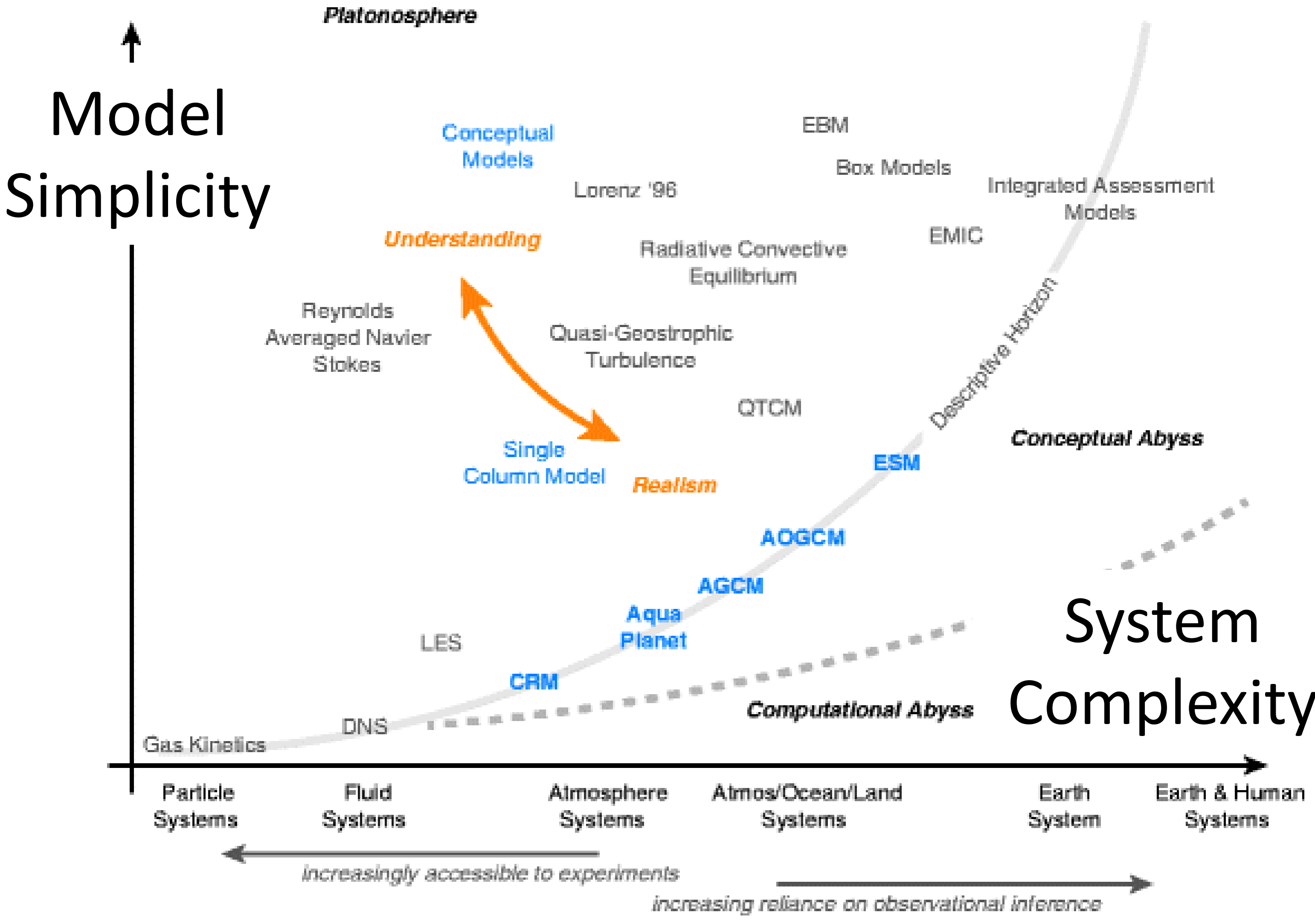
Model Simplicity



System Complexity

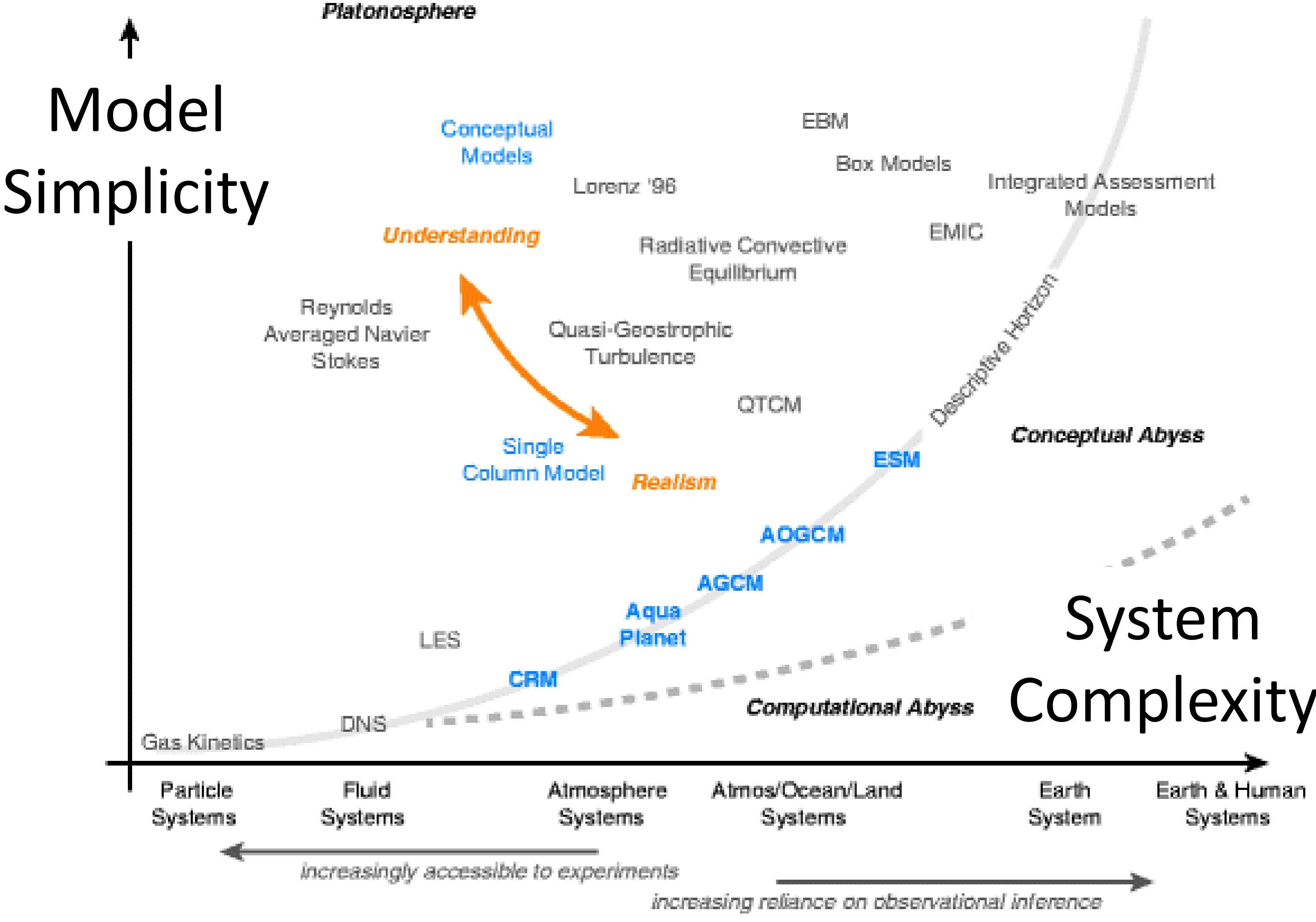
Source: Bony et al. (2013); See: Jeevanjee et al. (2017), Balaji (2022)

Idea: Promoting model hierarchies for ML models



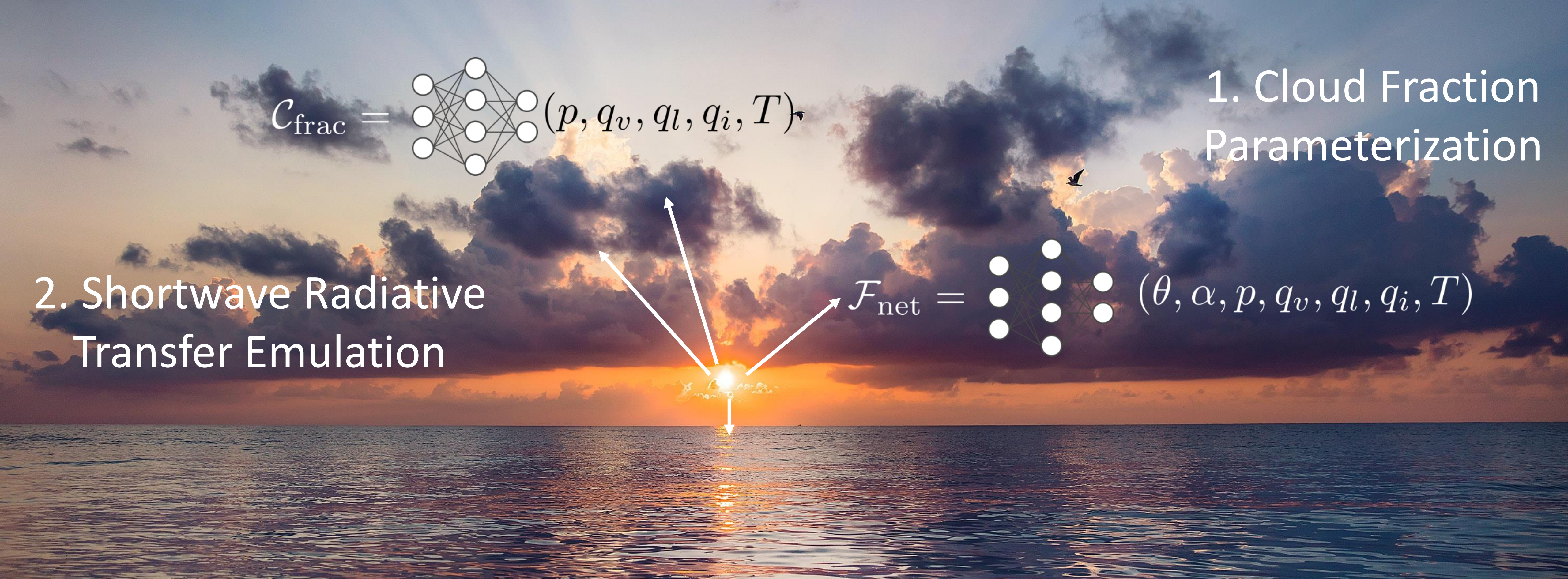
Source: Bony et al. (2013), xkcd; See: Jeevanjee et al. (2017), Balaji (2022)

Given a climate process for which we have reliable data, how can we systematically generate hierarchy of ML models?



Source: Bony et al. (2013), xkcd; See: Jeevanjee et al. (2017), Balaji (2022)

Outline: Generate Hierarchy for two atmospheric processes relevant for climate/weather predictions



Applications: 1) Data-driven model development, 2) Process understanding

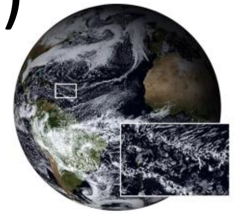
Methods: 1) Sequential Feature Selection, 2) Pareto Optimality

Movie from: Monsoon IV (Olbinski, 2017)

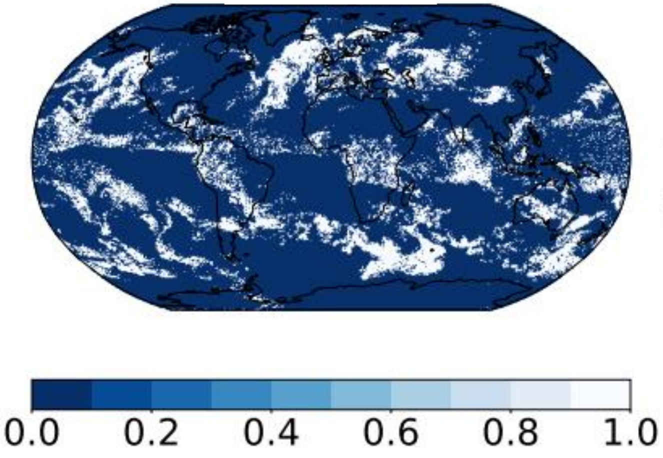
1. Improving Cloud Cover Parameterization in ICON (Unified German NWP/climate model)

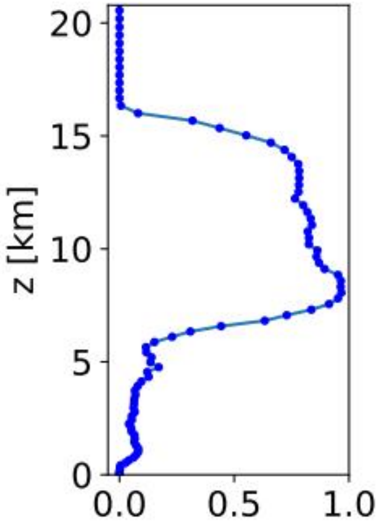
<u>Motivation</u>: Reduce cloud-related biases for climate projections

<u>Data</u>: 2.5km-res, 59-layer, global storm-resolving ICON runs (DYAMOND)



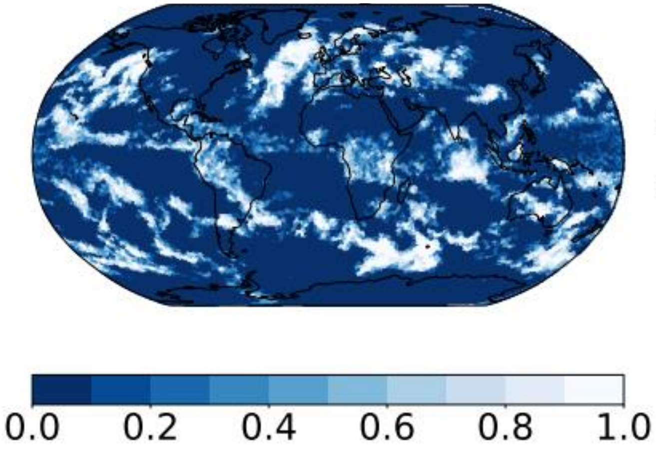
<u>Original Cloud Cover</u>

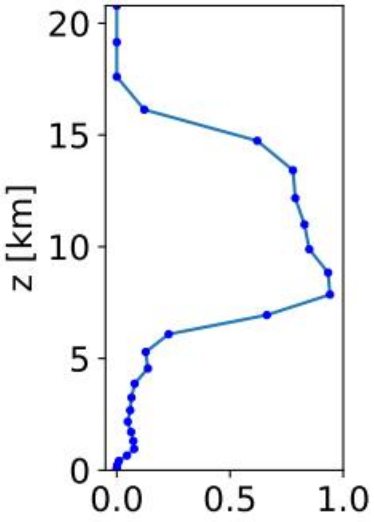




Coarse Graining

<u>80km-res "High-fidelity" Cl. Cov.</u>

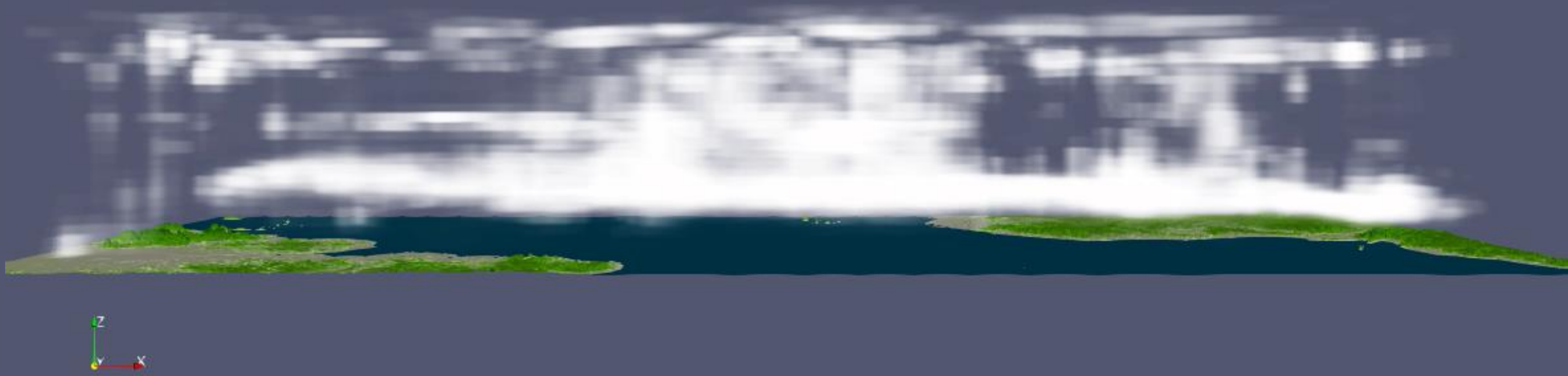




<u>Source</u>: Grundner, Beucler et al. (2022), Giorgetta et al. (2022), Stevens et al. (2019)

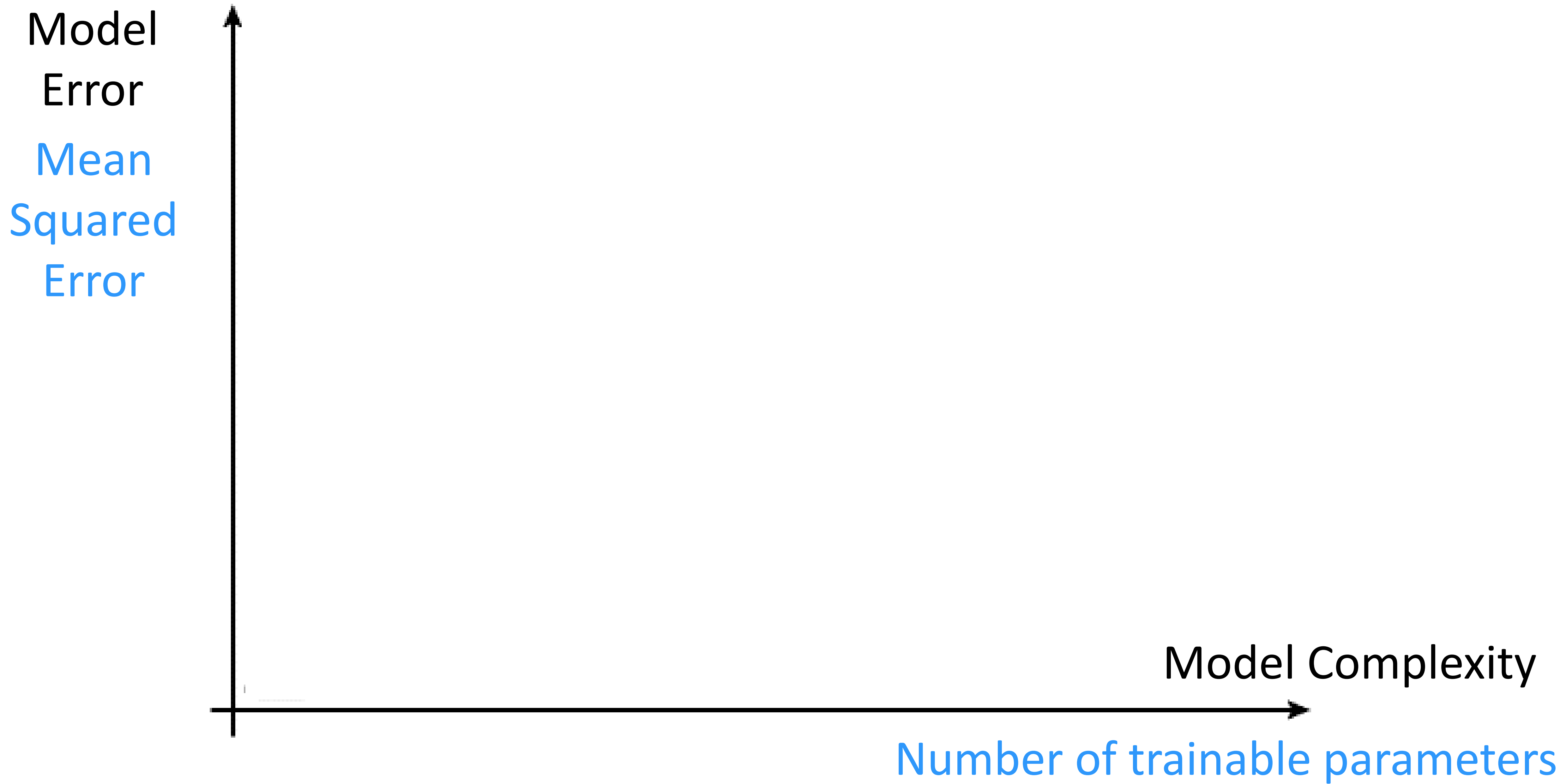
Neural Nets have root-mean squared errors < 7%

Neural Network Estimate

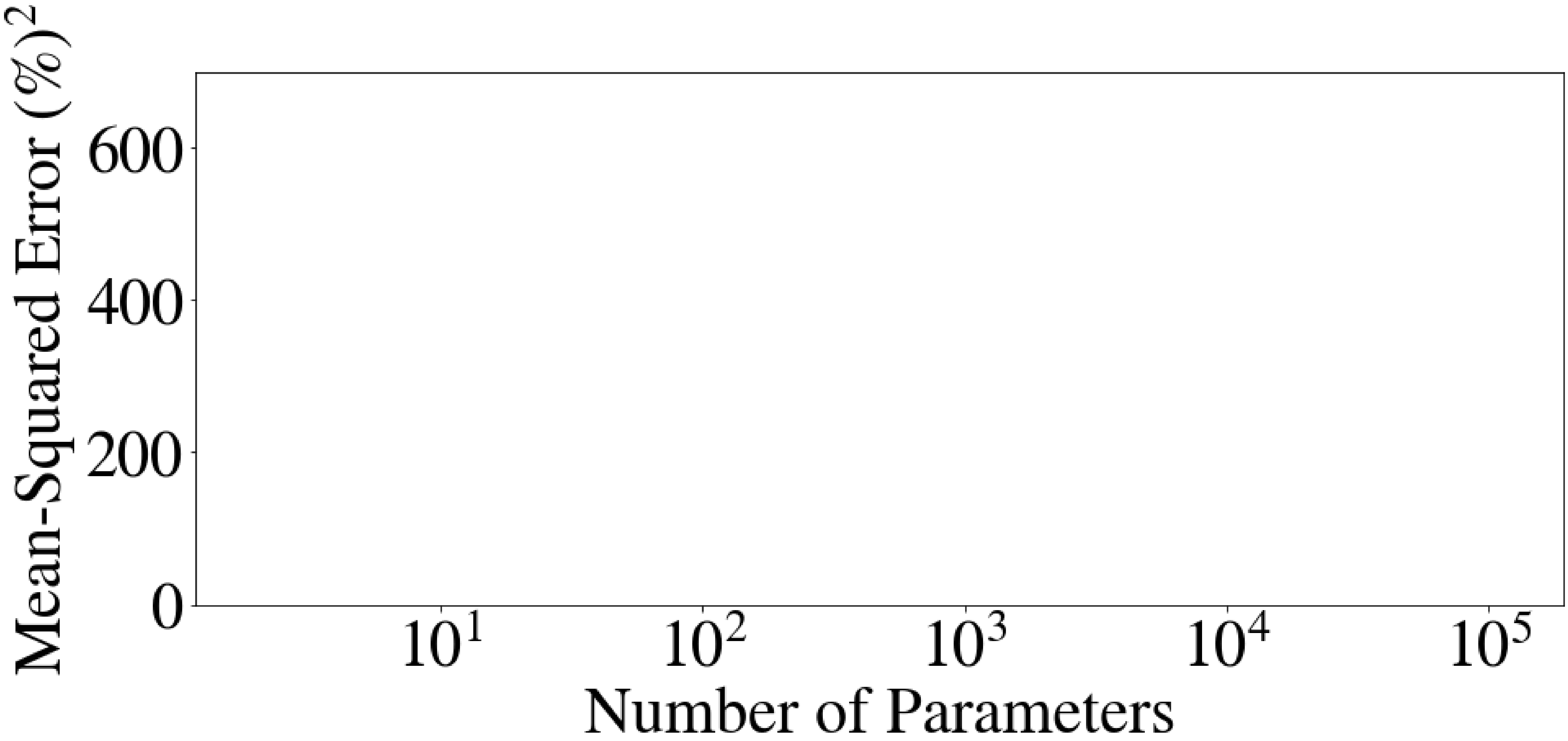


Reference (Coarse-Grained High-resolution simulation)

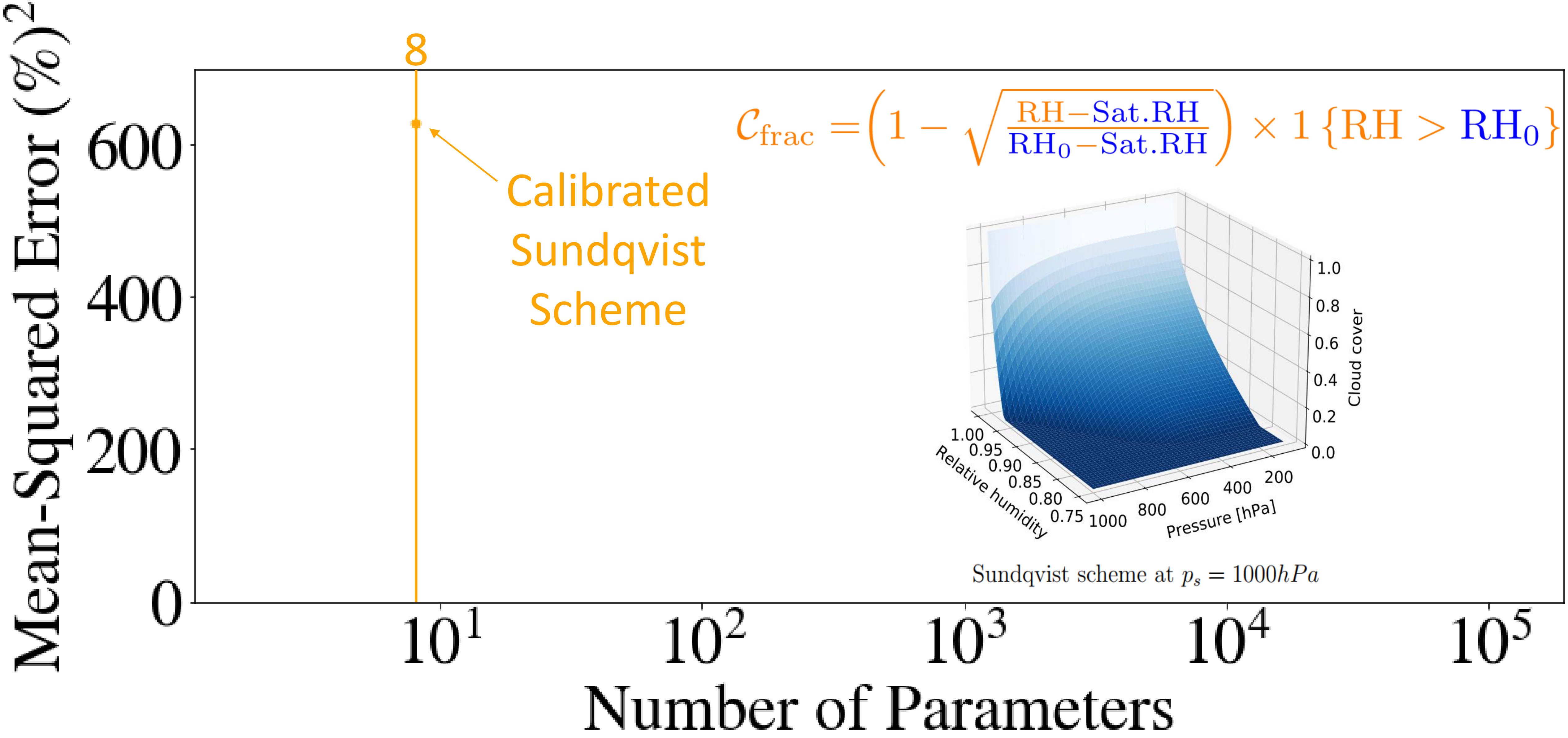
Analogy: Work in a well-defined (Complexity,Performance) plane



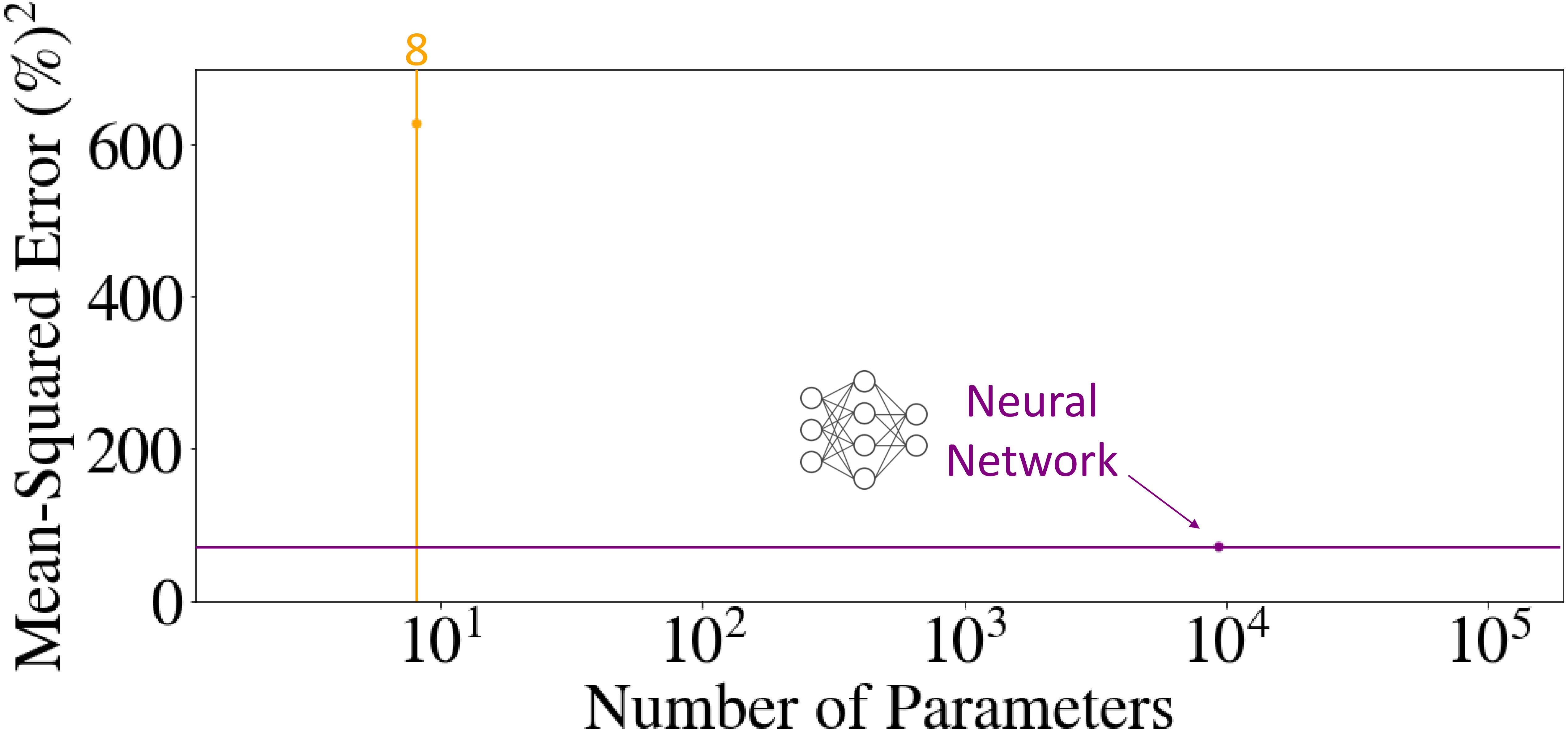
Improving Cloud Cover Parameterization using High-Res. ICON Data



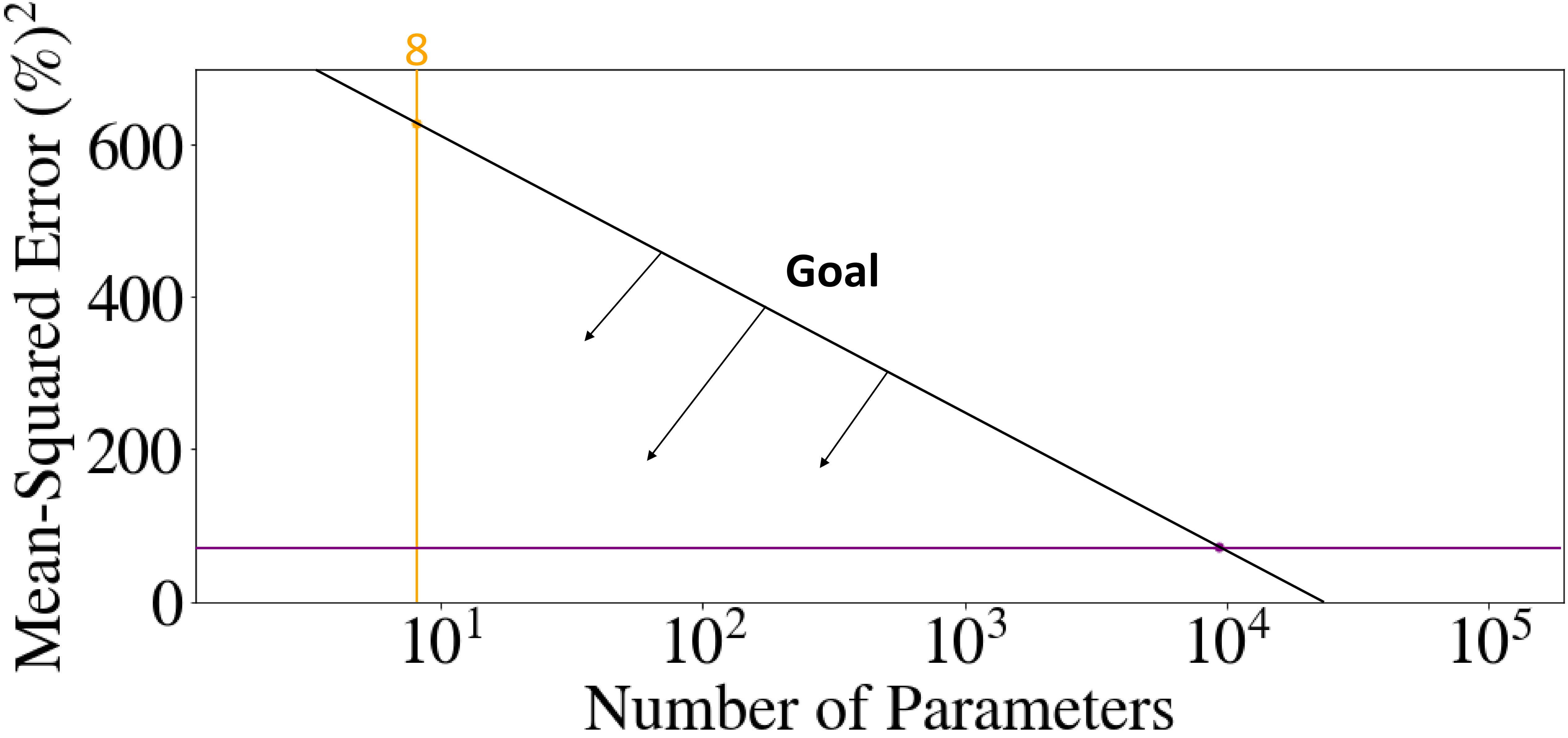
Improving Cloud Cover Parameterization using High-Res. ICON Data



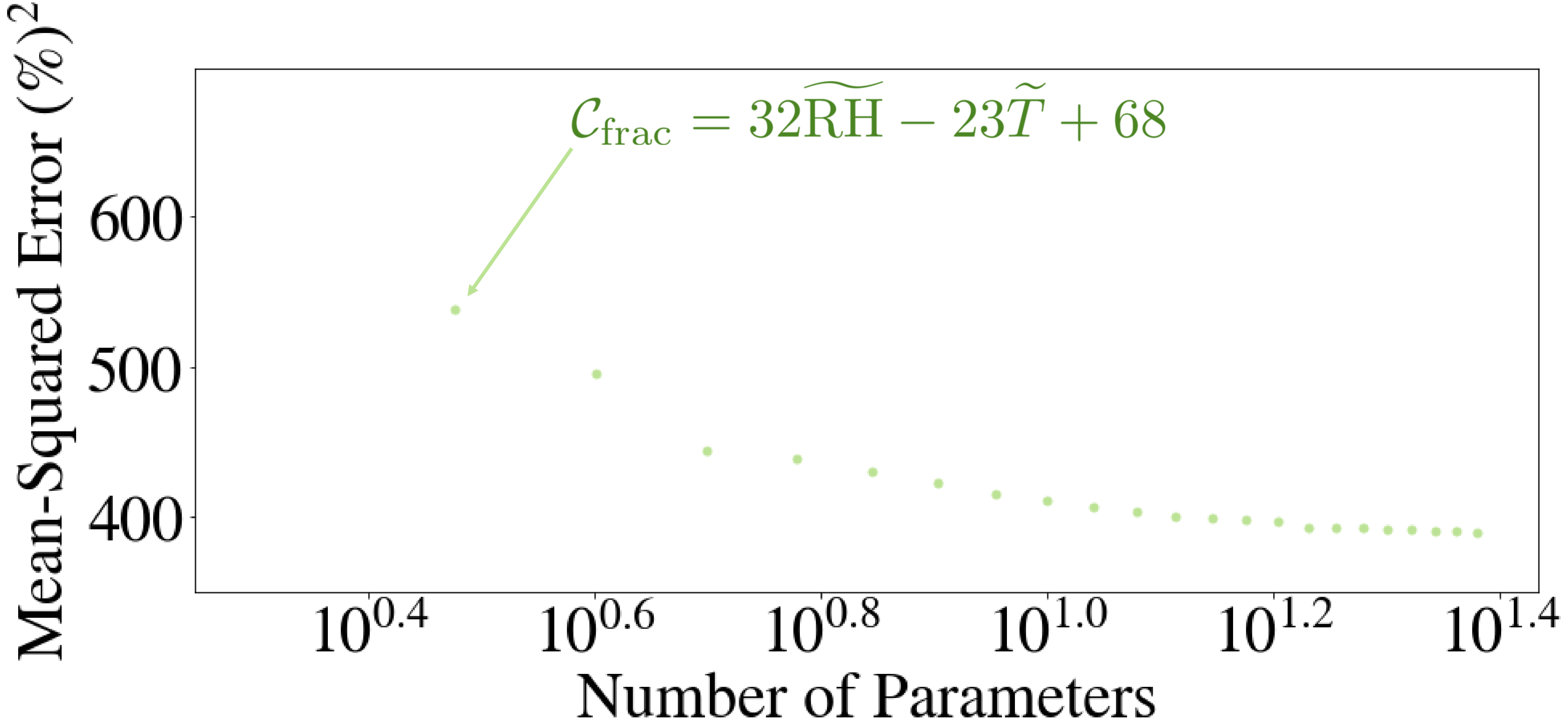
Improving Cloud Cover Parameterization using High-Res. ICON Data



How to jointly minimize error and complexity?

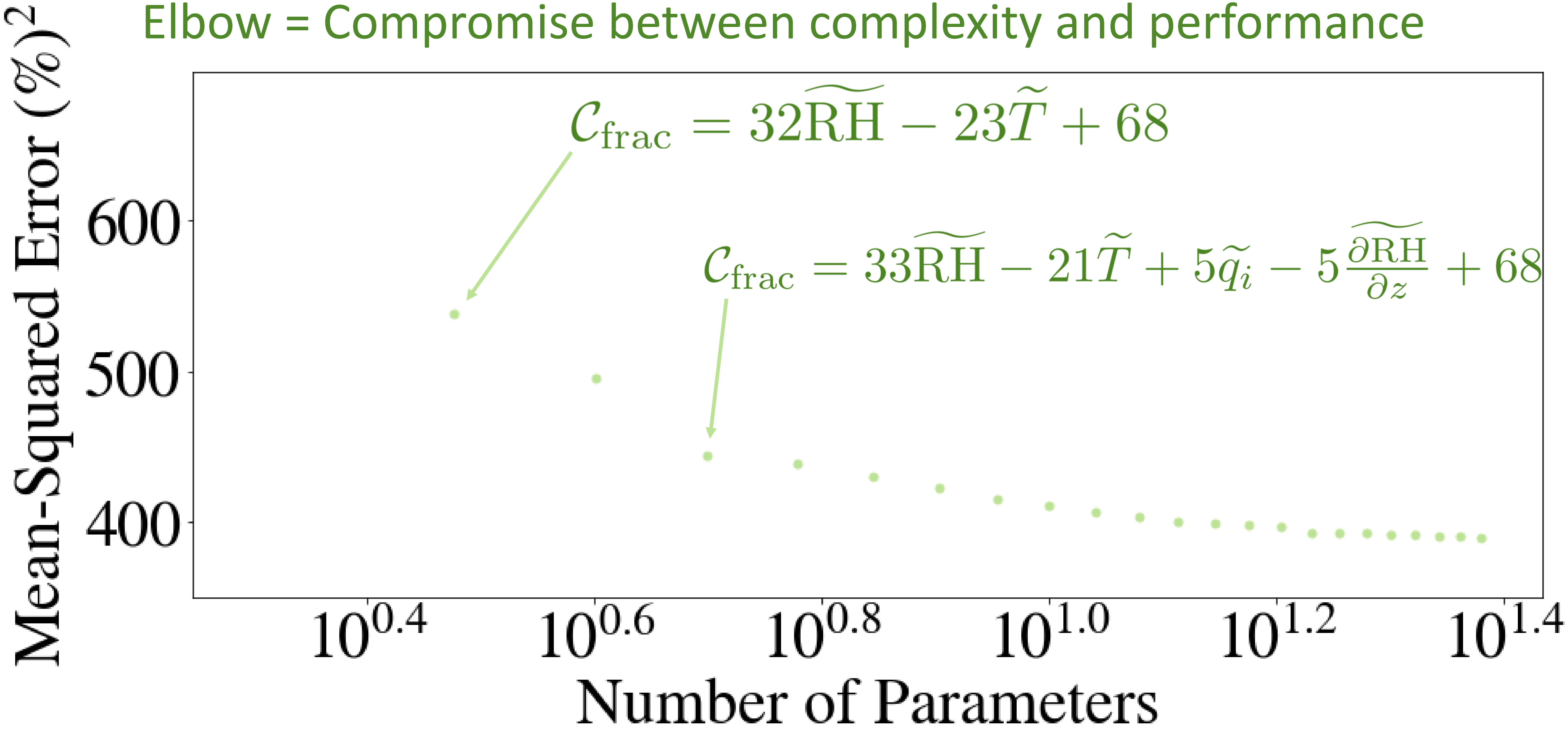


Sequential Feature Selection helps build a hierarchy of ML models by progressively increasing the **number of inputs**

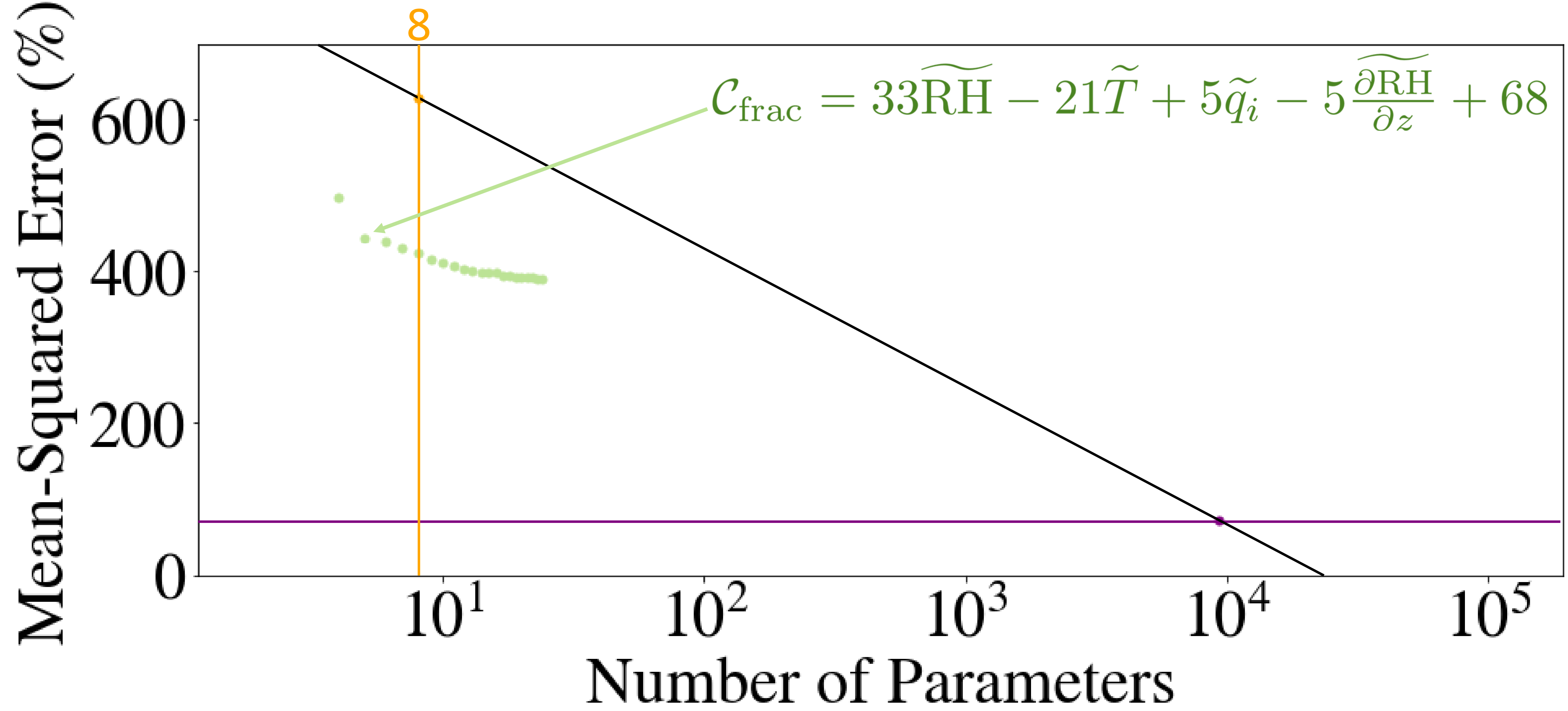


⚒️ Tool 1: Sequential Feature Selection helps build a hierarchy of ML models by progressively increasing the **number of inputs** ⚒️

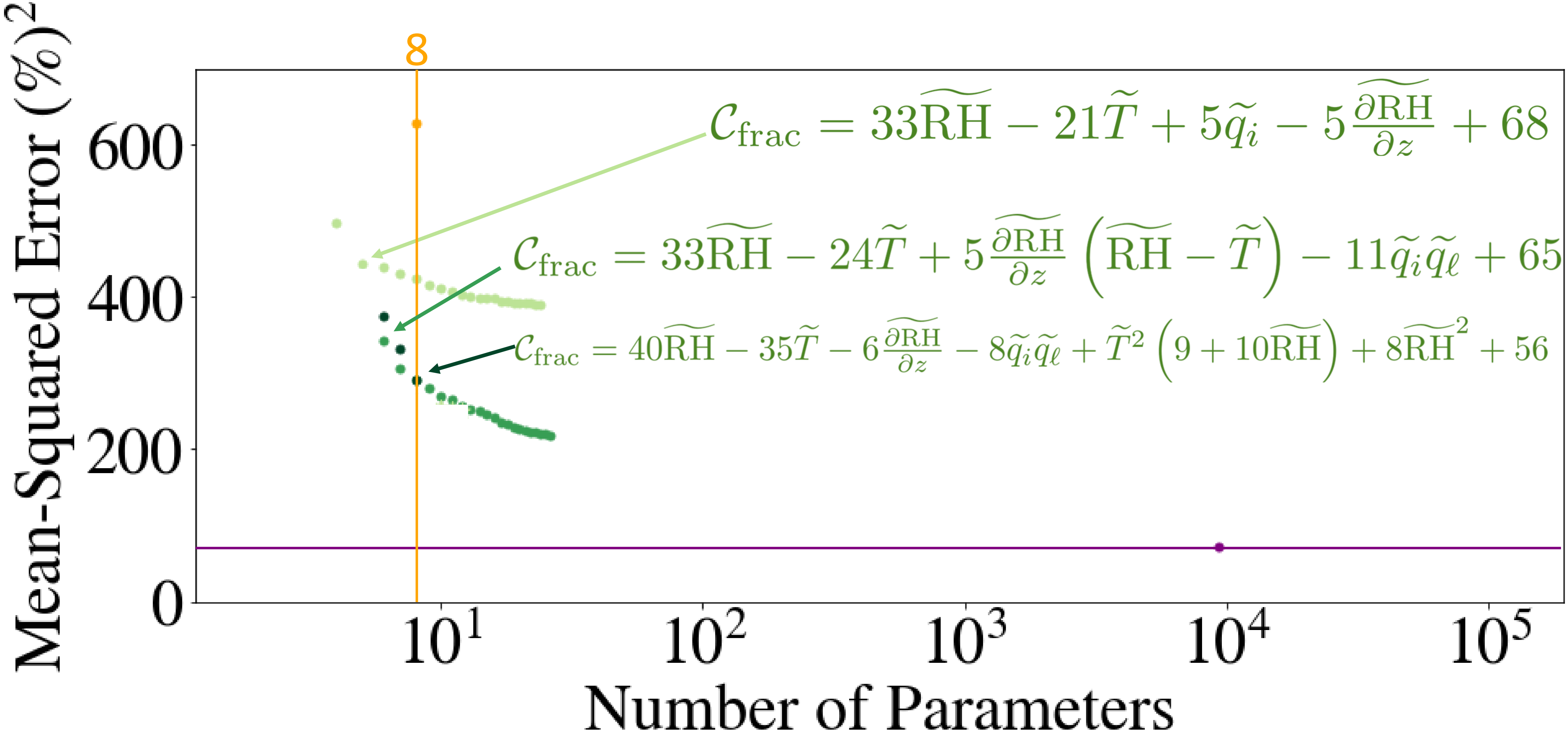
Elbow = Compromise between complexity and performance



Simple linear models are more appropriate than Sundqvist scheme to parameterize cloud fraction

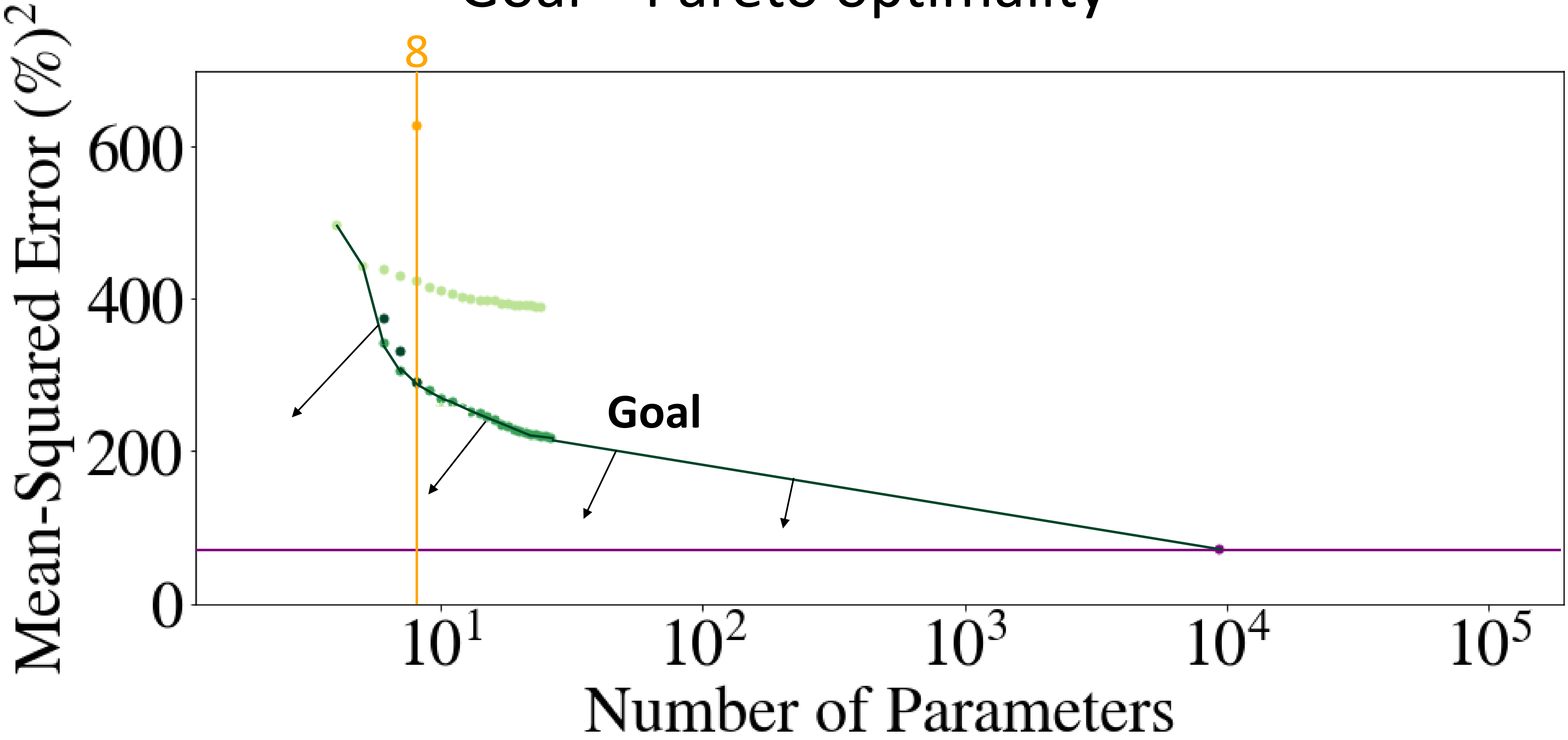


Tool 2: Increasing model complexity draws a Pareto frontier

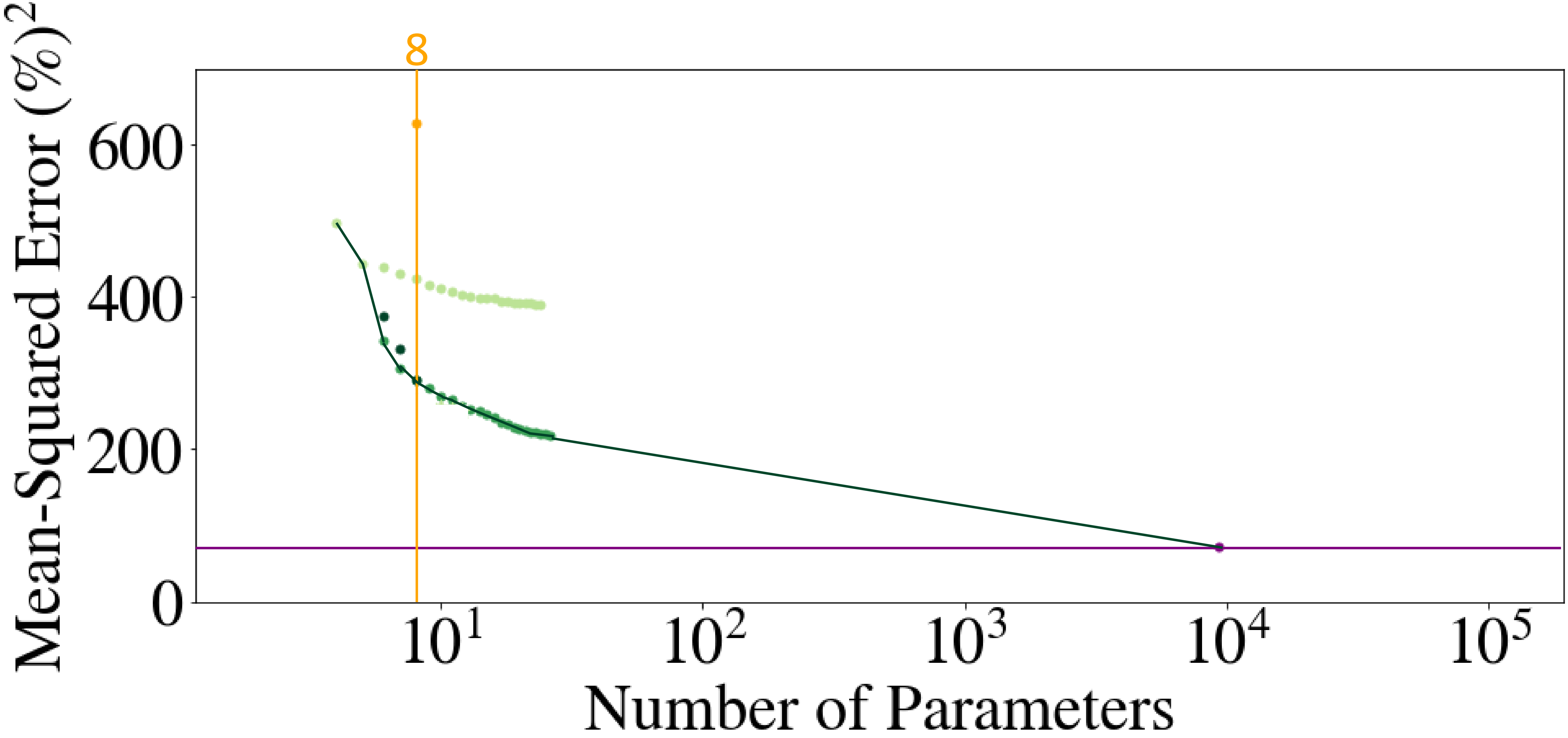


✅ <u>Application 1</u>: Data-Driven Model Development ✅

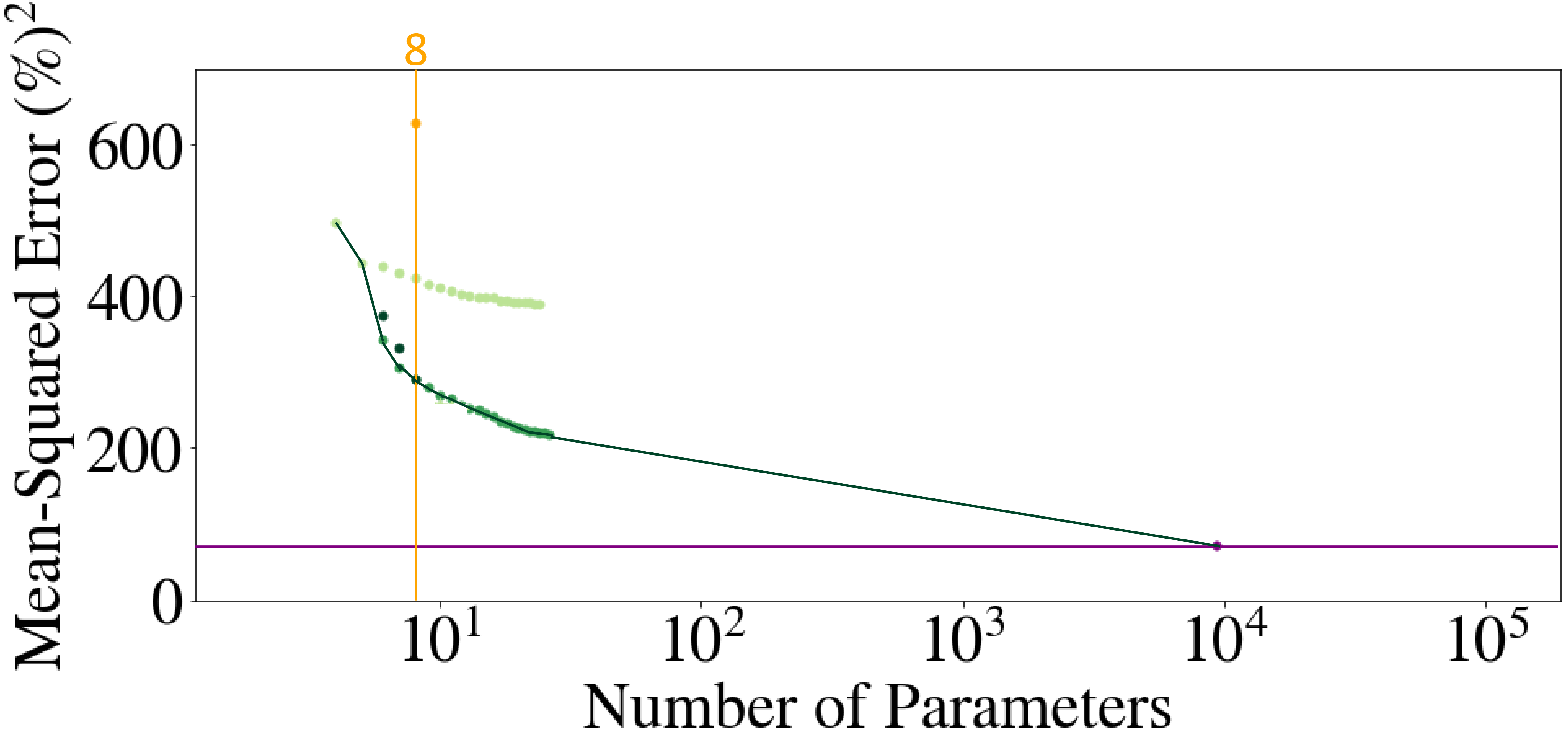
Goal = Pareto optimality



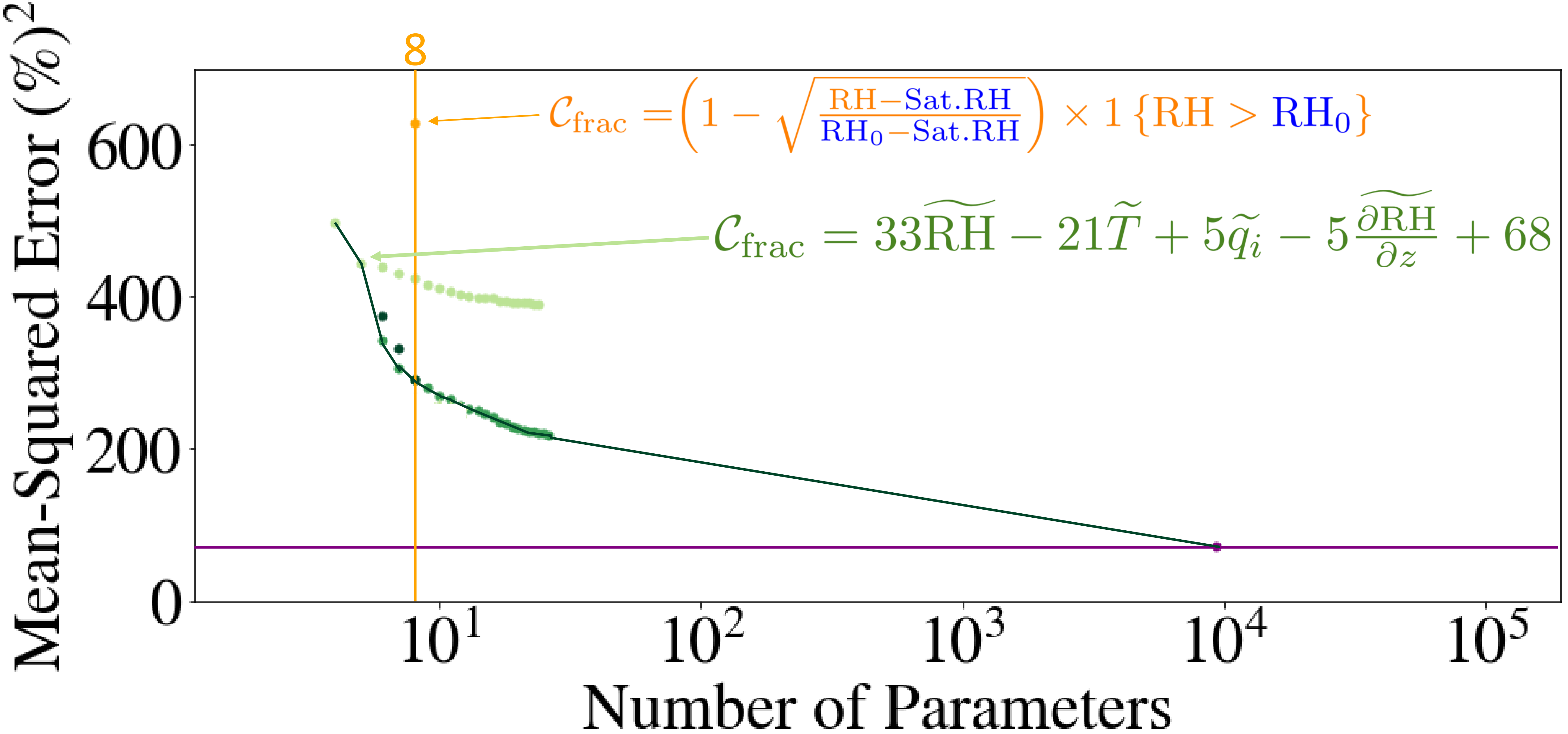
Application 2: Process Understanding



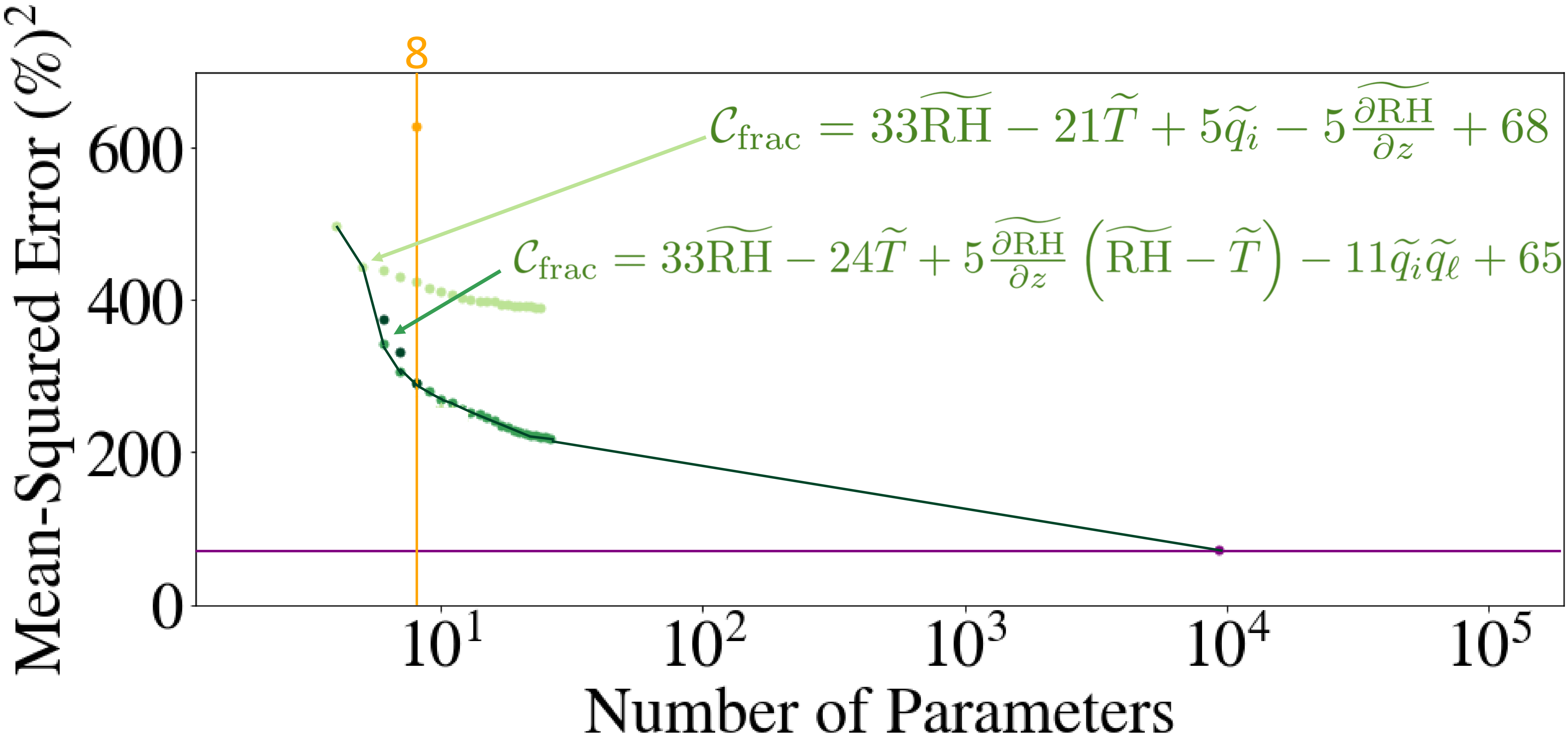
Why does increased complexity improve performance?



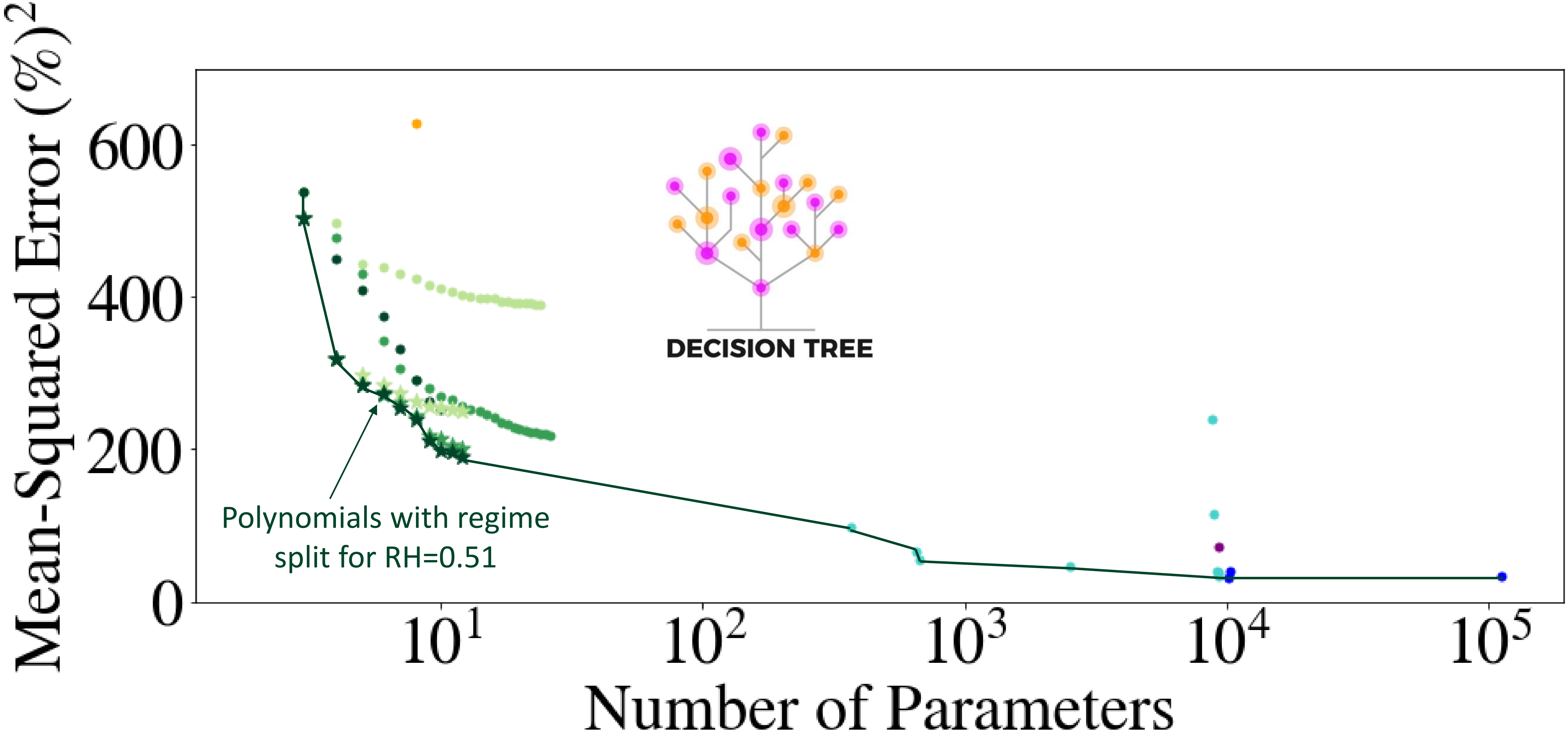
Concentration of *all* water species & temperature help accurately predict cloud fraction



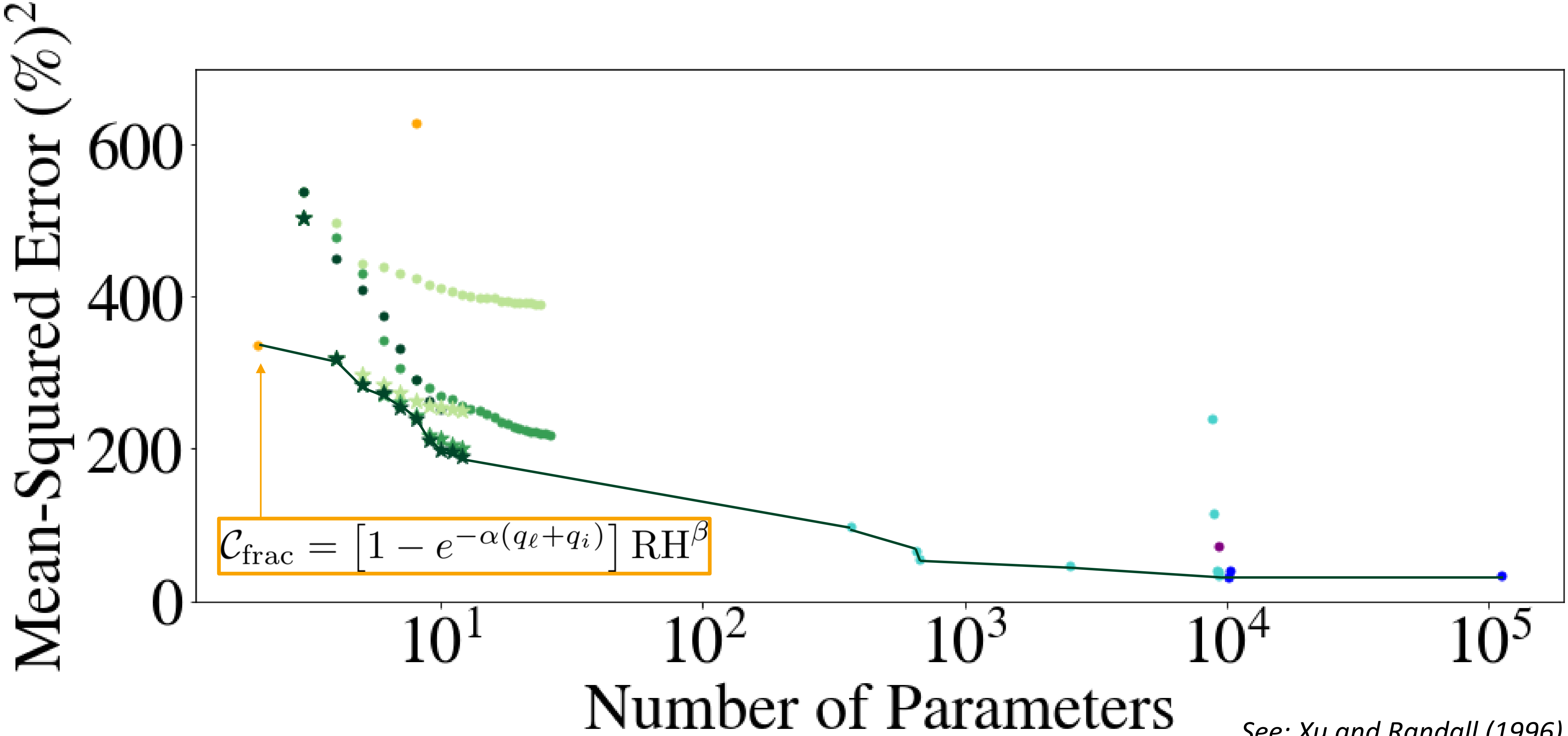
Condensates *nonlinearly* related to cloud fraction



Nonlinearities are crucial, even for relative humidity…

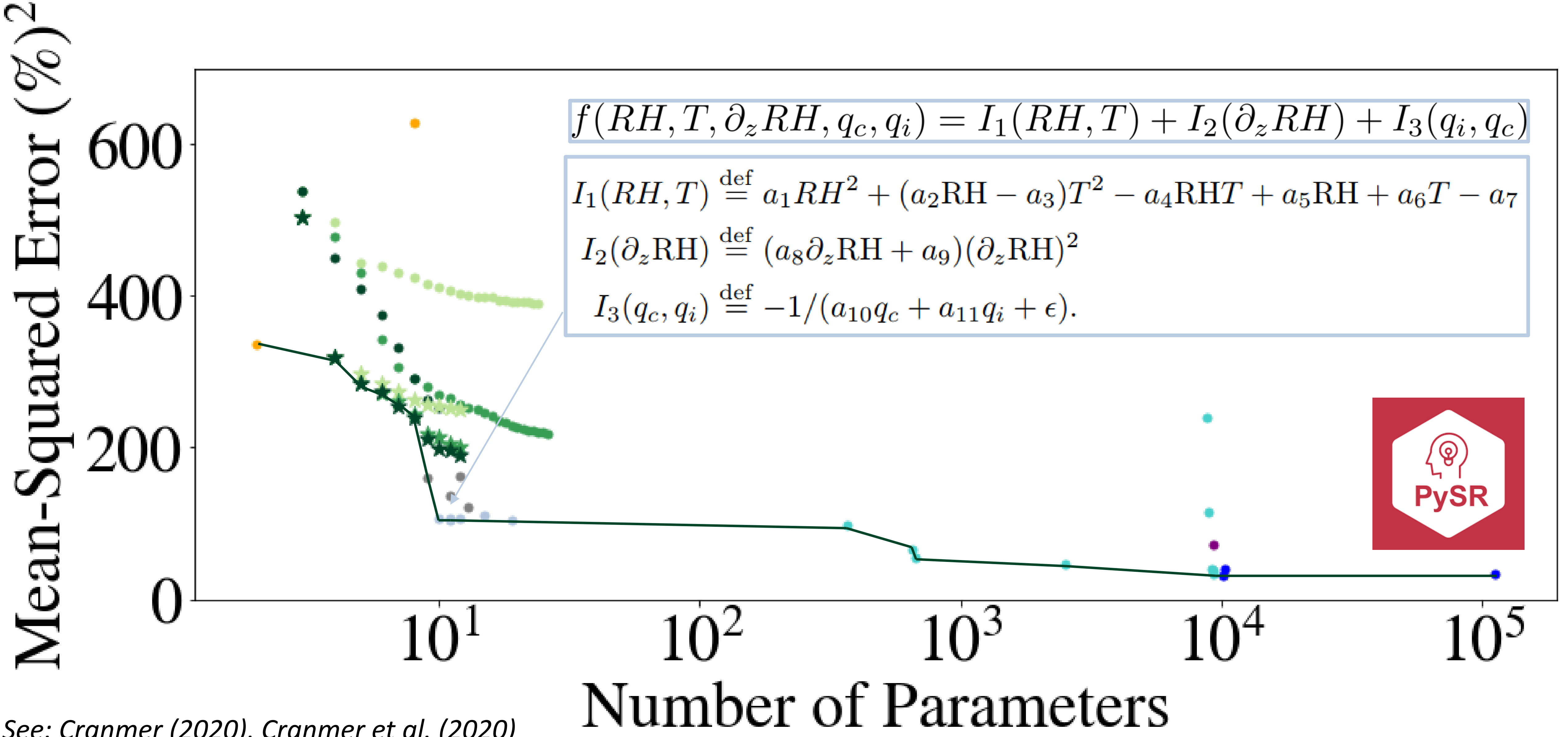


...explaining the success of a simple exponential scheme for cloud cover



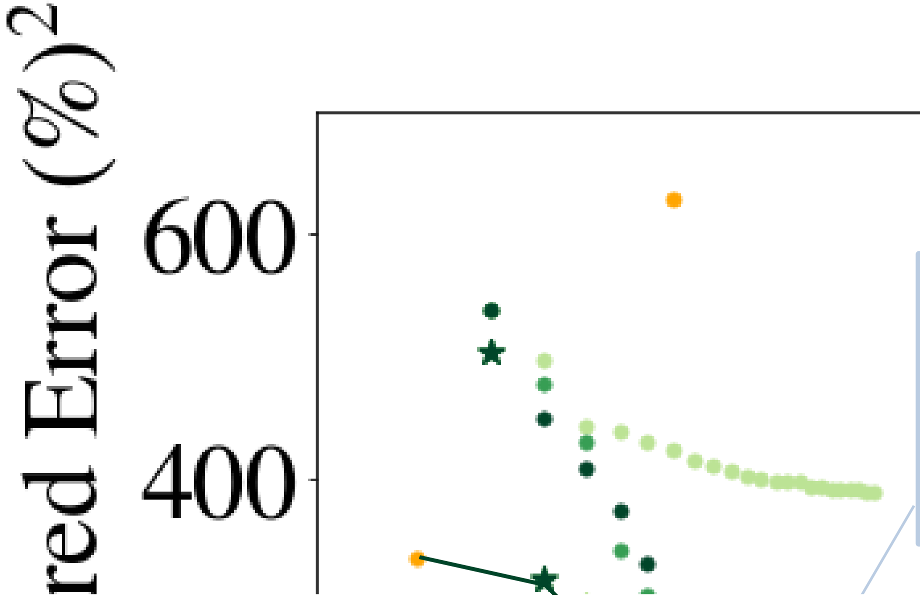
See: Xu and Randall (1996)

And guiding the discovery of new equations for cloud cover



<u>See</u>: Cranmer (2020), Cranmer et al. (2020)

And guiding the discovery of new equations for cloud cover
Example of **transparent machine learning**

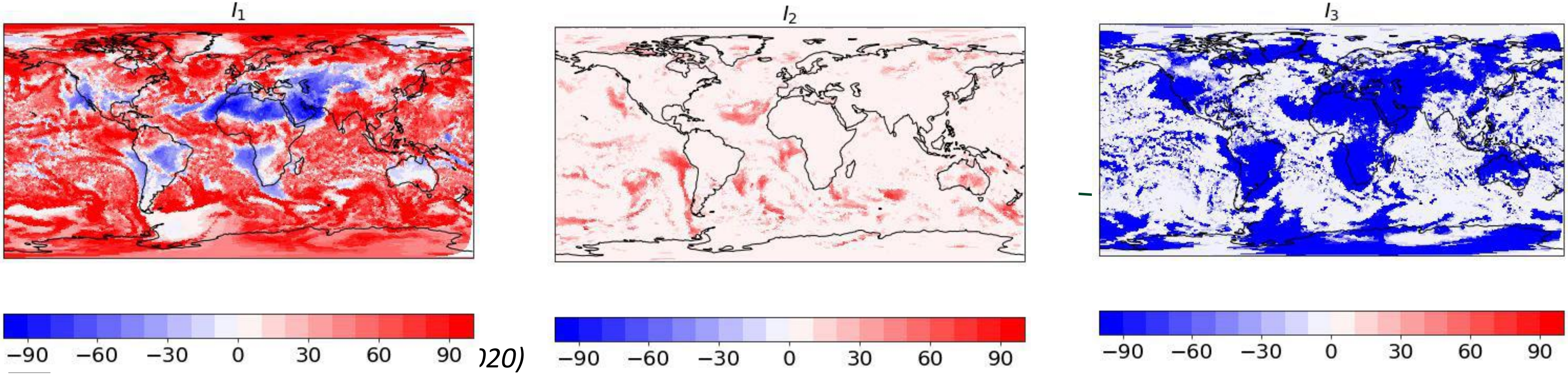


$$f(RH, T, \partial_z RH, q_c, q_i) = I_1(RH, T) + I_2(\partial_z RH) + I_3(q_i, q_c)$$

$$I_1(RH, T) \stackrel{\text{def}}{=} a_1 RH^2 + (a_2\text{RH} - a_3)T^2 - a_4\text{RH}T + a_5\text{RH} + a_6 T - a_7$$

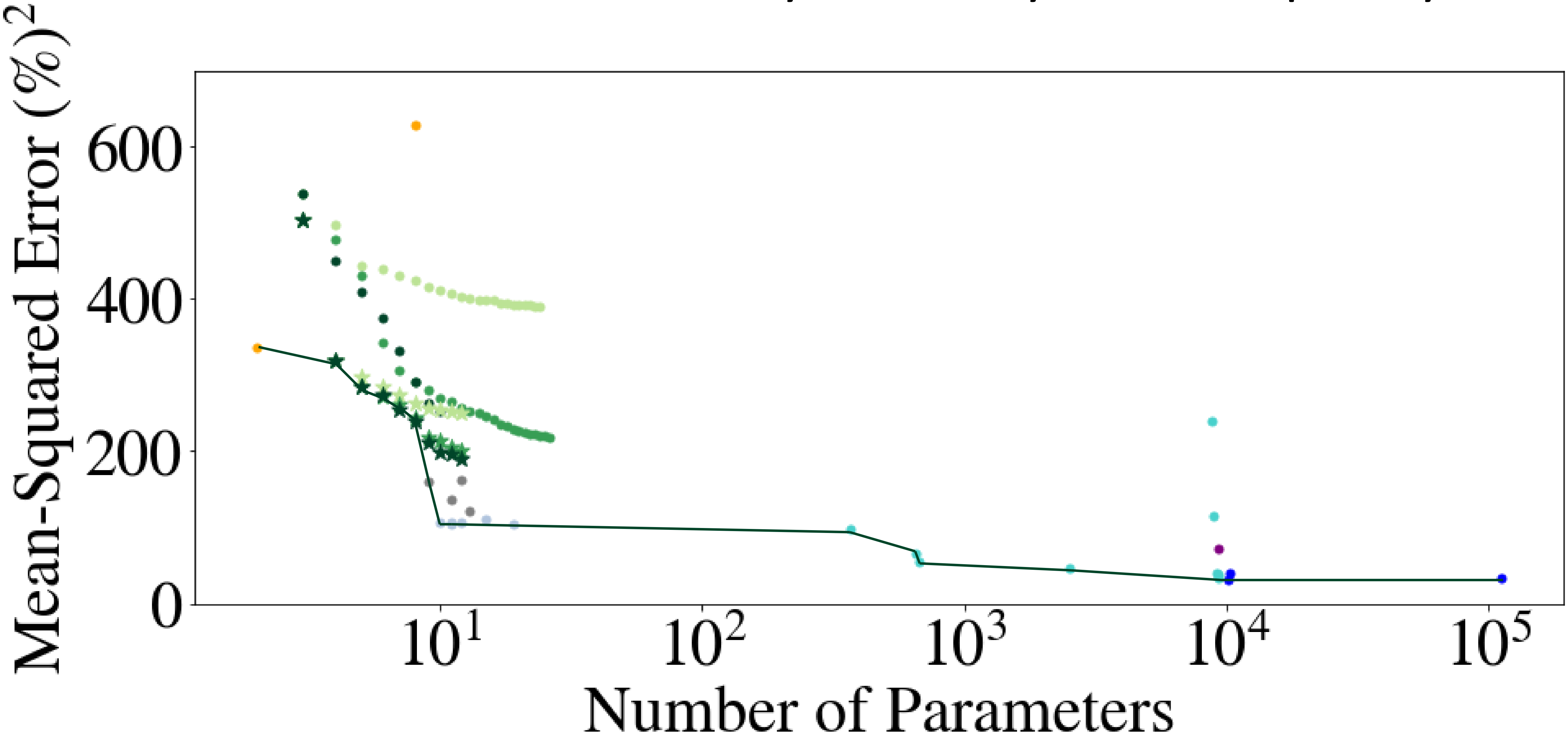
$$I_2(\partial_z \text{RH}) \stackrel{\text{def}}{=} (a_8 \partial_z \text{RH} + a_9)(\partial_z \text{RH})^2$$

$$I_3(q_c, q_i) \stackrel{\text{def}}{=} -1/(a_{10} q_c + a_{11} q_i + \epsilon).$$



✅ Application 2: Process Understanding ✅

Pareto frontier hierarchically unveils system complexity



2. Accelerating Shortwave Radiative Transfer for NWP by emulating high-fidelity rad. transfer model

Source: Lagerquist et al. (2021)

2. Accelerating Shortwave Radiative Transfer for NWP by emulating high-fidelity rad. transfer model

<u>Motivation</u>: Even correlated-k models (RRTM) are too slow for NWP

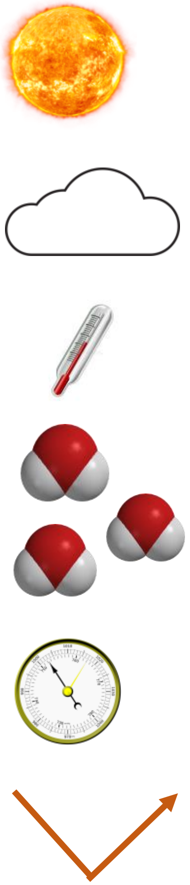
<u>Data</u>: Input derived from the Rapid Refresh model, then fed to RRTM

<u>Source</u>: Lagerquist et al. (2021), Krasnopolsky et al. (2010, 2020), Benjamin et al. (2016), Mlawer et al. (1997)

2. Accelerating Shortwave Radiative Transfer for NWP by emulating high-fidelity rad. transfer model

Motivation: Even correlated-k models (RRTM) are too slow for NWP

Data: Input derived from the Rapid Refresh model, then fed to RRTM



Zenith Angle

Vapor/Liq Paths Above/Below

Temperature

Vapor Concentration

Liquid Concentration

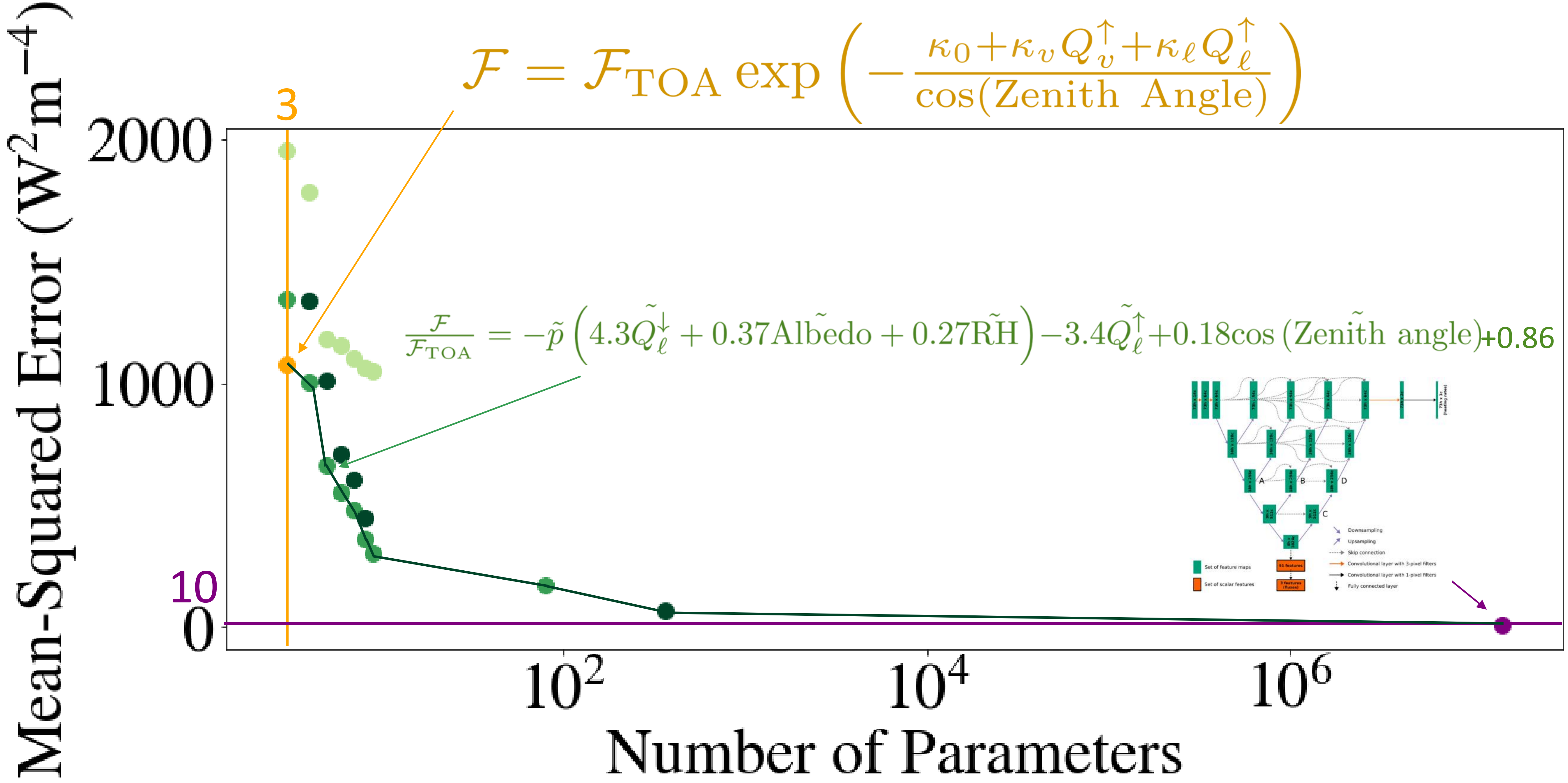
Pressure

Albedo

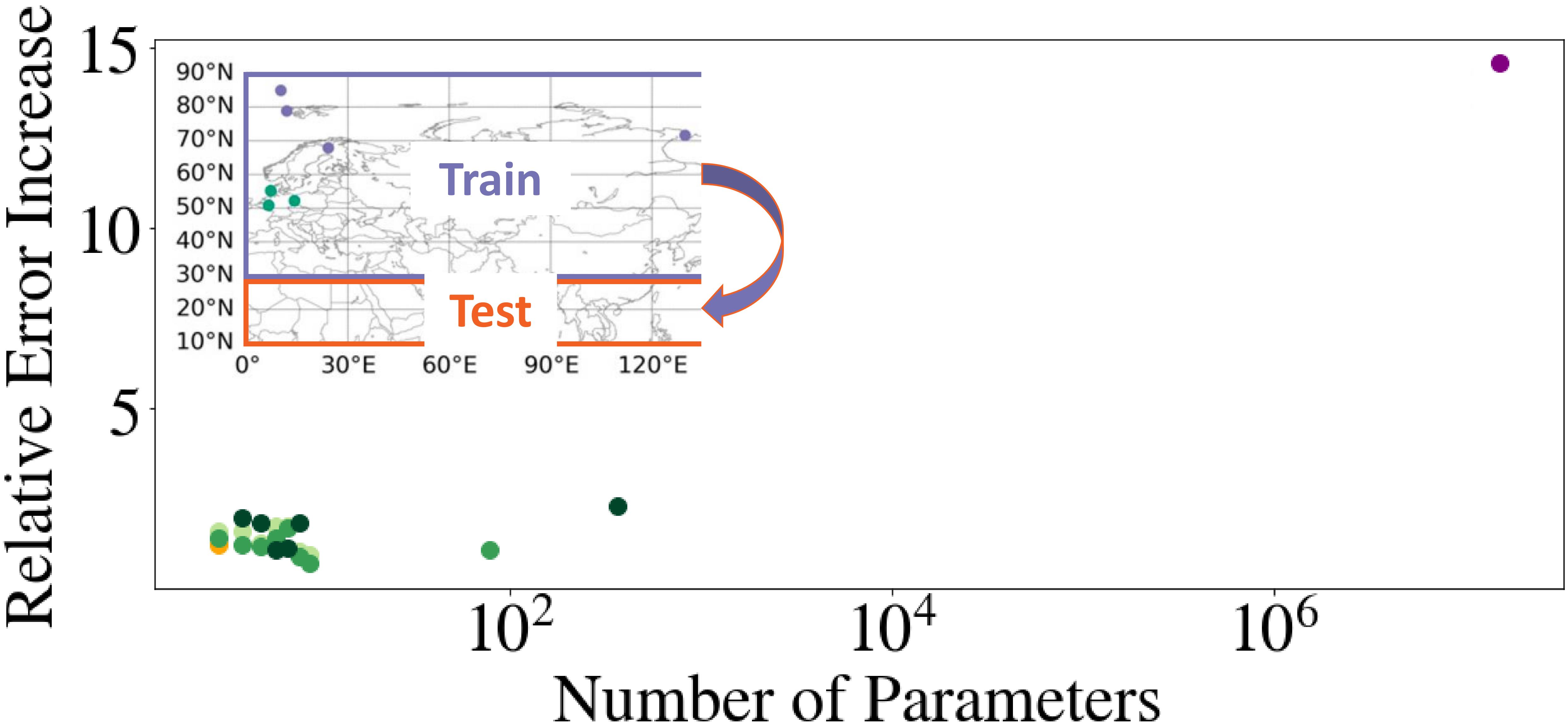
Goal: Data-driven model

Source: Lagerquist et al. (2021), Krasnopolsky et al. (2010, 2020), Benjamin et al. (2016), Mlawer et al. (1997), NASA Sci.

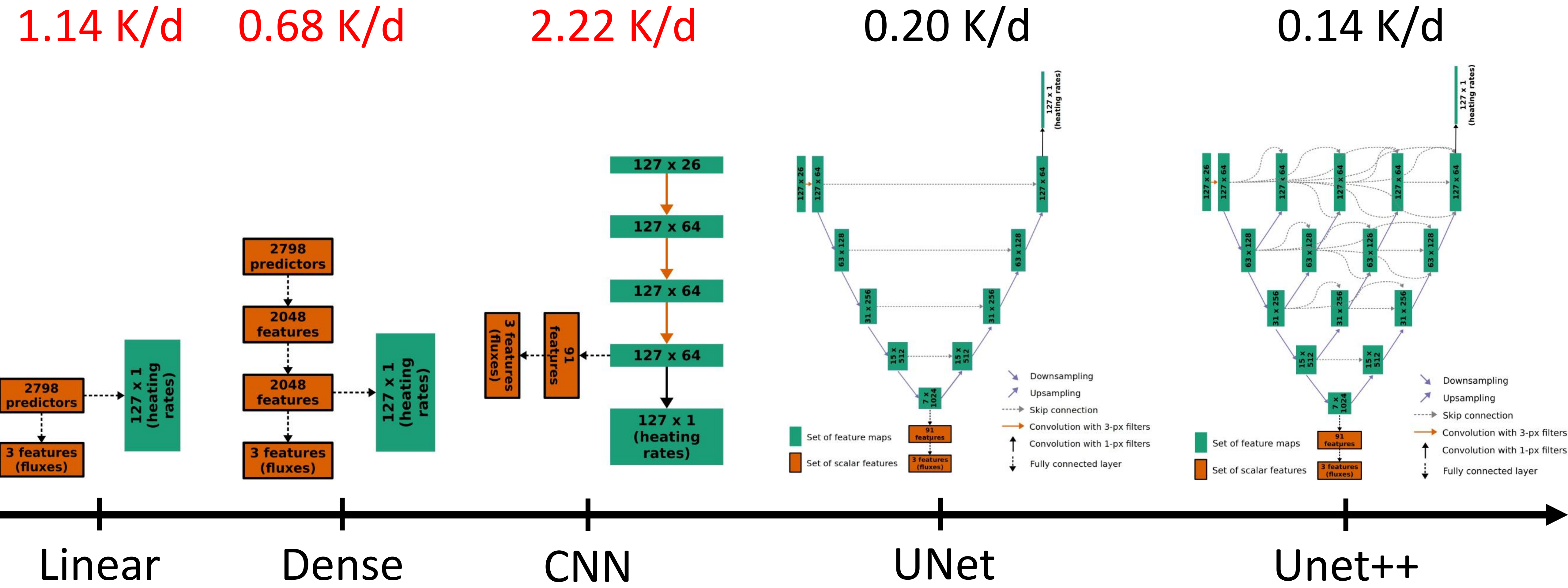
One-stream & quadratic models are Pareto-optimal



Parsimony Principle: "Models with less parameters tend to generalize better to out-of-distribution samples"



Tip: Explore various definitions of performance & complexity



Data Source: GFS (not RAP), See: AMS Talk 1A.5 by Ryan Lagerquist et al., Ukkonen (2022)

We can systematically build hierarchies of ML models to better understand their added value

<u>Applications</u>:

1. Guide data-driven model development by jointly minimizing error and complexity in a well-defined plane, indicating "Pareto-optimal" models
2. Further process understanding by hierarchically unveiling system complexity (key features/nonlinearity/space-time connectivity) by comparing models & investigating error statistics along Pareto frontier
3. Anticipate generalization to out-of-distribution (parsimony principle)

<u>Advantages</u>:

- Cleanly comparing existing schemes to powerful data-driven models
- Developing data-driven models directly applicable to NWP/climate
- Improving trustworthiness of ML models via hierarchical understanding*

Thank you!

tom.beucler@unil.ch

www.unil.ch/dawn

UNIL | Université de Lausanne

∂^3AWN

∂ata-∂riven Atmospheric & Water ∂yNamics

GitHub

arXiv

Data-Driven Equation Discovery of a Cloud Cover Parameterization

Arthur Grundner[1,2], Tom Beucler[3], Pierre Gentine[2], and Veronika Eyring[1,4]

arXiv 2304.08063

JAMES Journal of Advances in Modeling Earth Systems

Research Article | Open Access

Deep Learning Based Cloud Cover Parameterization for ICON

Arthur Grundner, Tom Beucler, Pierre Gentine, Fernando Iglesias-Suarez, Marco A. Giorgetta, Veronika Eyring

lms.ecmwf.int

Editorial Type: Article

Article Type: Research Article

Using Deep Learning to Emulate and Accelerate a Radiative Transfer Model

Ryan Lagerquist, David Turner, Imme Ebert-Uphoff, Jebb Stewart, and Venita Hagerty

Online Publication: 21 Sep 2021

Print Publication: 01 Oct 2021

DOI: https://doi.org/10.1175/JTECH-D-21-0007.1

Page(s): 1673–1696