Towards transparent data-driven parametrisations for weather and climate modelling

Honours research project proposal

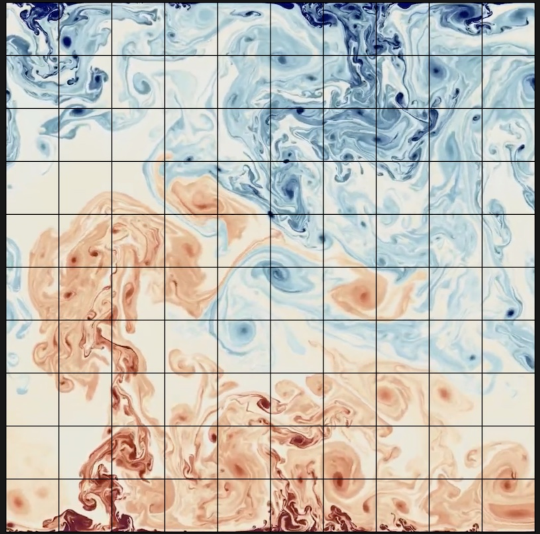
Thomas Schanzer

School of Physics
Climate Change Research Centre and ARC Centre of Excellence for Climate Extremes

University of New South Wales, Sydney, Australia

Thursday 30 March 2023

Models cannot explicitly resolve:



Stephan Lenz (2018), YouTube, CC BY License,
https://youtu.be/BJKiuwpdprQ

Huw A. Ogilvie (2005), Flickr, CC BY 2.0 License,
https://flic.kr/p/6eYnw

The general problem: never perfect, always looking for better ways

The parametrisation problem has this generic form:

$$\begin{cases} \partial_t(\text{resolved variable}) = S_\mathrm{r}(\text{resolved variables}) + C_\mathrm{r}(\text{unresolved variables}) \\ \partial_t(\text{unresolved variable}) = S_\mathrm{u}(\text{unresolved variables}) + C_\mathrm{u}(\text{resolved variables}) \end{cases}$$

$$\downarrow$$

$$\partial_t(\text{resolved variable}) \approx S_\mathrm{r}(\text{resolved variables}) + \underbrace{P(\text{resolved variables})}_{\text{parametrised tendency}}$$

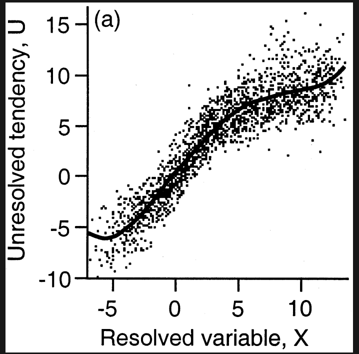
How do we construct P?

- Traditionally: simple conceptual models
- Data-driven methods:
  - Machine learning
  - Other statistical models and regressions

1. The million-dollar question
2. Traditional example: convection schemes like Zhang-McFarlane
3. Conceptual methods subject to biases (e.g. drizzle problem)
4. Machine learning: black box
5. Simpler methods tested using toy models (e.g. Lorenz 96)

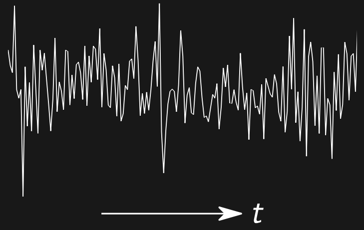
Three key ingredients for parametrisation

Deterministic base

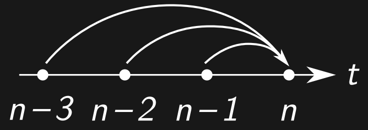


Wilks, D. S. (2005). "Effects of stochastic parametrizations in the Lorenz '96 system". Q. J. R. Meteorol. Soc. 131(606), 389-407. doi: 10.1256/qj.04.03.

Stochasticity



Memory



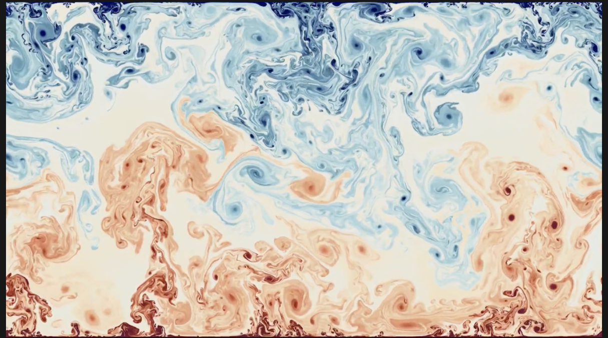
1. Explain construction of Wilks' plot
2. Stochasticity: capture variability
3. Memory: noise values not independent, unresolved tendencies persist over time, reminiscent of Takens' embedding theorem

I will ask:

- Do toy model parametrisation methods generalise to real systems?
- What level of predictability is achievable if the system has *inherent* stochasticity and/or memory?

1. Toy models are good but...
2. Even if the system is deterministic in principle, there may be completely unknown degrees of freedom, or we might not have a complete understanding of the physics

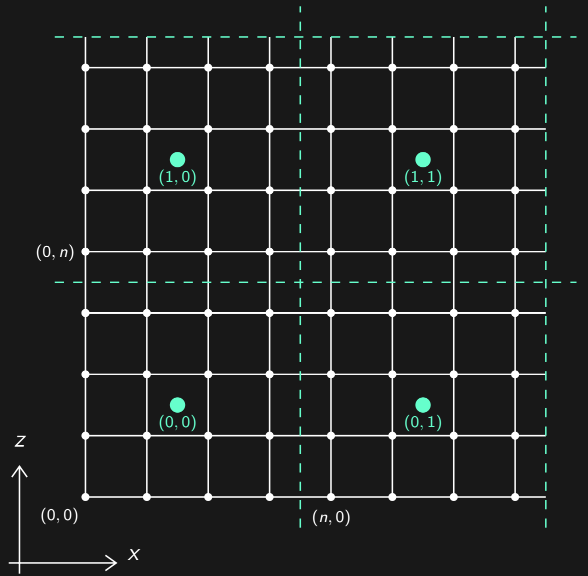
I will consider 2D Rayleigh-Bénard convection



$$\begin{cases} \partial_t \boldsymbol{u} + (\boldsymbol{u}\cdot\boldsymbol{\nabla})\boldsymbol{u} = -\dfrac{1}{\rho_0}\boldsymbol{\nabla} p + \nu\nabla^2\boldsymbol{u} + g\alpha(T-T_0)\hat{\boldsymbol{z}} \\ \partial_t T + (\boldsymbol{u}\cdot\boldsymbol{\nabla})T = D_T\nabla^2 T \\ \boldsymbol{\nabla}\cdot\boldsymbol{u} = 0 \end{cases}$$

Explain boundary conditions and typical steady behaviour

The equations are decomposed into coarse and fine parts



$$u \to \bar{u} + u'$$
$$w \to \bar{w} + w'$$
$$T \to \bar{T} + T'$$

Puts RBC into same form as toy models!

I will perform the following experiments:

1. Integrate full coupled system as truth, divide data into training and evaluation sets
2. Parametrise the model using different schemes:
   2.1 Trivial (i.e., nothing)
   2.2 Deterministic regression model
   2.3 Regression + stochasticity
   2.4 Regression + stochasticity + memory (e.g., autoregressive)
   2.5 More advanced methods in literature
3. Repeat (1-2) when the “truth” system has inherent:
   3.1 Stochastic forcing
   3.2 Memory
   3.3 Both of the above

I will perform the following experiments:

1. Integrate full coupled system as truth, divide data into training and evaluation sets
2. Parametrise the model using different schemes:
   - 2.1 Trivial (i.e., nothing)
   - 2.2 Deterministic regression model
   - 2.3 Regression + stochasticity
   - 2.4 Regression + stochasticity + memory (e.g., autoregressive)
   - 2.5 More advanced methods in literature
3. Repeat (1-2) when the "truth" system has inherent:
   - 3.1 Stochastic forcing
   - 3.2 Memory
   - 3.3 Both of the above

I will perform the following experiments:

1. Integrate full coupled system as truth, divide data into training and evaluation sets
2. Parametrise the model using different schemes:
   2.1 Trivial (i.e., nothing)
   2.2 Deterministic regression model
   2.3 Regression + stochasticity
   2.4 Regression + stochasticity + memory (e.g., autoregressive)
   2.5 More advanced methods in literature
3. Repeat (1-2) when the "truth" system has inherent:
   3.1 Stochastic forcing
   3.2 Memory
   3.3 Both of the above

Parametrisation performance will be assessed by:

- Short-term forecast accuracy:
  - RMSE
- Long-term average “climate” prediction:
  - Statistical moments of the coarse variables (mean, variance, skewness, etc.)
  - Closeness of their PDFs (e.g., Kolmogorov-Smirnov statistic)

I envision the following outcomes:

- For each parametrisation scheme:
  - Quantitative assessment of performance, strengths, weaknesses
  - Determination of ideal settings
  - Scalability from toy models to more complex systems
  - How do these change if the system has inherent stochasticity/memory?
- Concluding recommendations on scheme choice and settings for different modelling objectives (i.e., forecasting vs. climate statistics)
- Implications for parametrisation development in weather/climate models

I envision the following outcomes:

- For each parametrisation scheme:
  - Quantitative assessment of performance, strengths, weaknesses
  - Determination of ideal settings
  - Scalability from toy models to more complex systems
  - How do these change if the system has inherent stochasticity/memory?
- Concluding recommendations on scheme choice and settings for different modelling objectives (i.e., forecasting vs. climate statistics)
- Implications for parametrisation development in weather/climate models

I envision the following outcomes:

- For each parametrisation scheme:
  - Quantitative assessment of performance, strengths, weaknesses
  - Determination of ideal settings
  - Scalability from toy models to more complex systems
  - How do these change if the system has inherent stochasticity/memory?
- Concluding recommendations on scheme choice and settings for different modelling objectives (i.e., forecasting vs. climate statistics)
- Implications for parametrisation development in weather/climate models