# Can machine learning replace numerical models for partial differential equations?

Christian Lessig
Otto-von-Guericke-Universitaet Magdeburg

# Atmospheric dynamics

Coarse scale gas dynamics described by primitive equations (Navier-Stokes eqs.)

$$\frac{dv}{dt} = -(\frac{1}{\rho})\nabla p - g(\frac{r}{r}) + (\frac{1}{\rho})\left[\nabla \cdot (\mu \nabla v) + \nabla(\lambda \nabla \cdot v)\right]$$

$$c_v \frac{dT}{dt} + p\frac{d\alpha}{dt} = q + f$$

$$\frac{d\rho}{dt} + \rho \nabla \cdot v = 0$$

$$p = nT$$

# Atmospheric dynamics

Coarse scale gas dynamics described by primitive equations (Navier-Stokes eqs.)

$$\frac{dv}{dt} = -(\frac{1}{\rho})\nabla p - g(\frac{r}{r}) + (\frac{1}{\rho})\left[\nabla \cdot (\mu \nabla v) + \nabla(\lambda \nabla \cdot v)\right]$$

$$c_v \frac{dT}{dt} + p\frac{d\alpha}{dt} = q + f$$

$$\frac{d\rho}{dt} + \rho \nabla \cdot v = 0$$

$$p = nT$$

Needs to be complemented with equations describing radiative transfer, cloud formation, interaction with soil, ocean, biosphere, ...

# Atmospheric dynamics

Numerical solution:

- Primitive equation discretized using standard approaches (spectral, finite volume, finite elements)
- “Parametrizations” to model large range of other phenomena
- Two closure problems:
  - Turbulence of gas phase
  - Many phenomena do not even have equations or models are far too expensive in a global simulation
- Global simulation has billions of DOF
  - Exascale computing

# Atmospheric dynamics

High societal relevance:

- Weather forecasting
- Climate projections
- Significant government and private investments in modeling and simulation

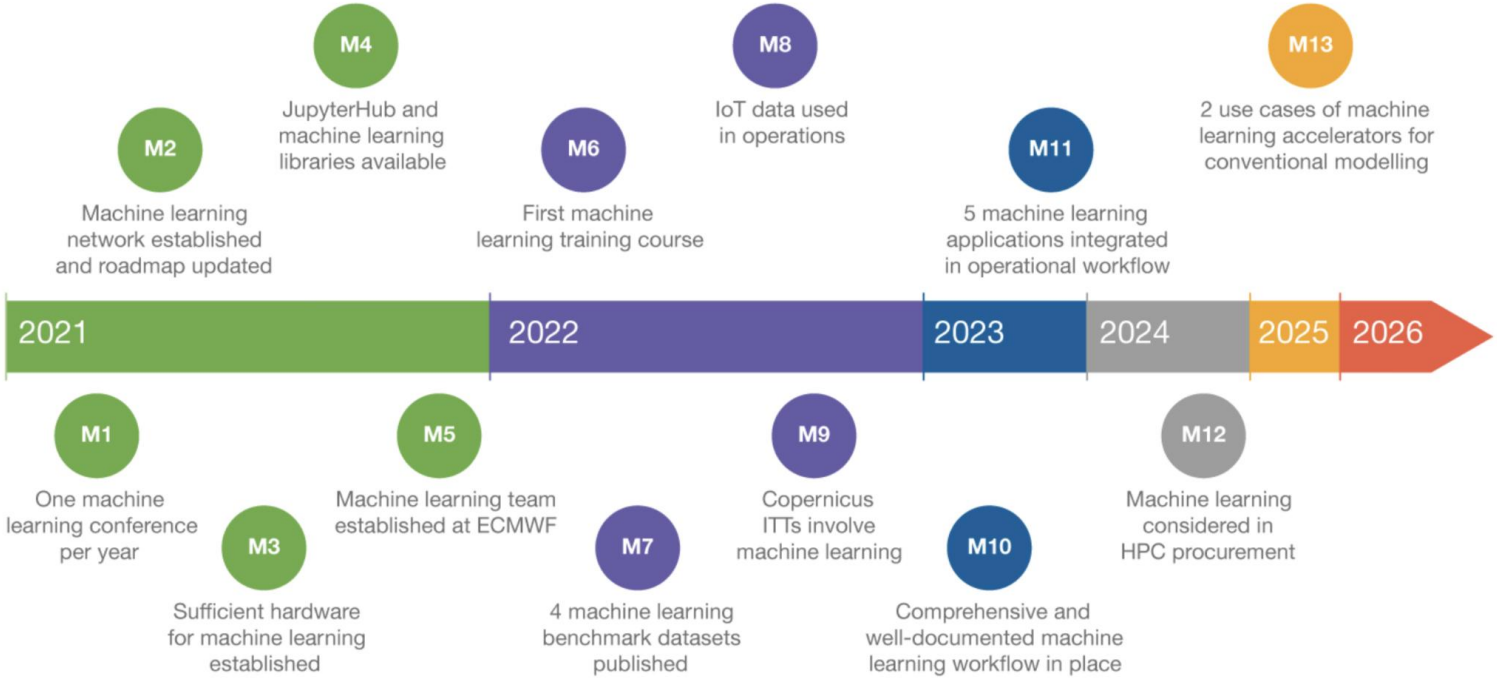
# Atmospheric dynamics

High societal relevance:

- Weather forecasting
- Climate projections
- Significant government and private investments in modeling and simulation

Which role can machine learning play?

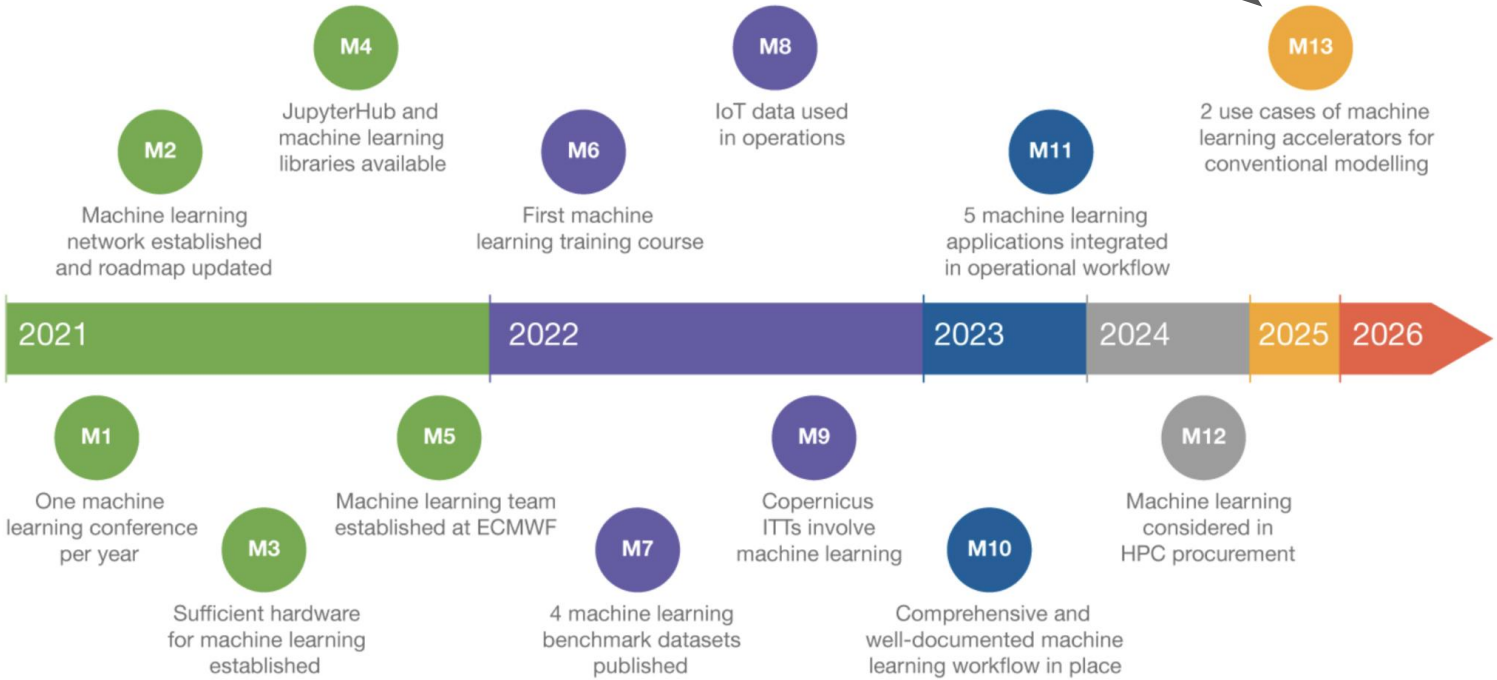
# The plan




ECMWF machine learning roadmap

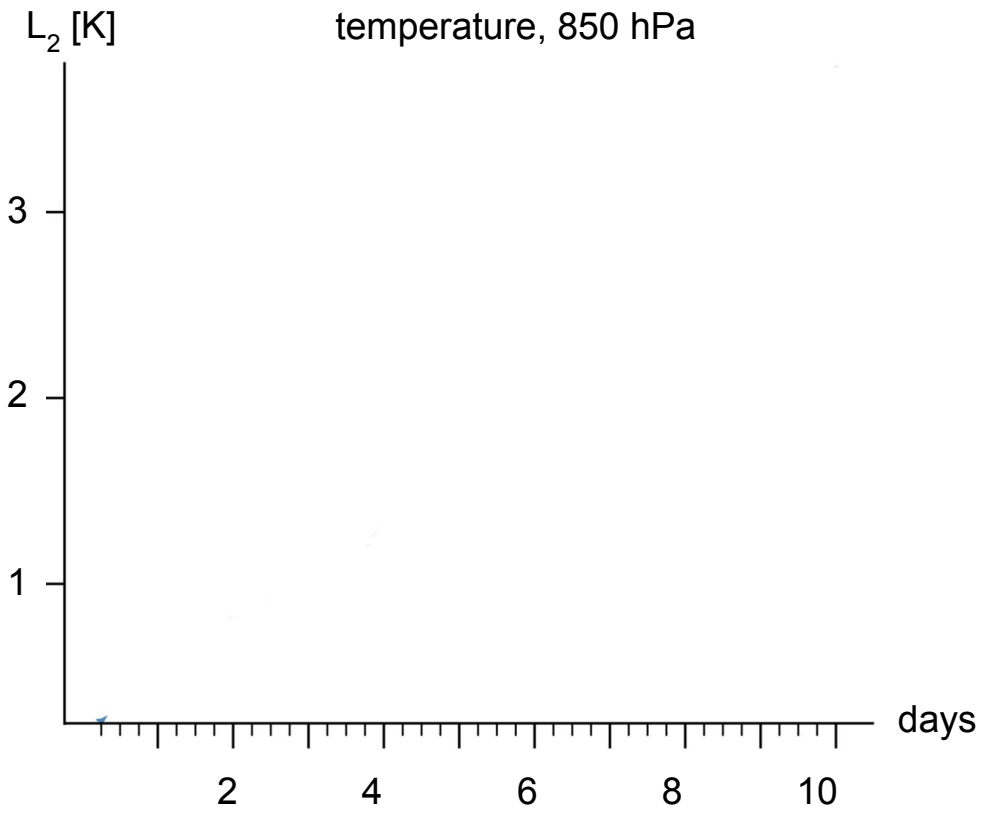
# The plan

Two applications of machine learning speed up conventional modelling

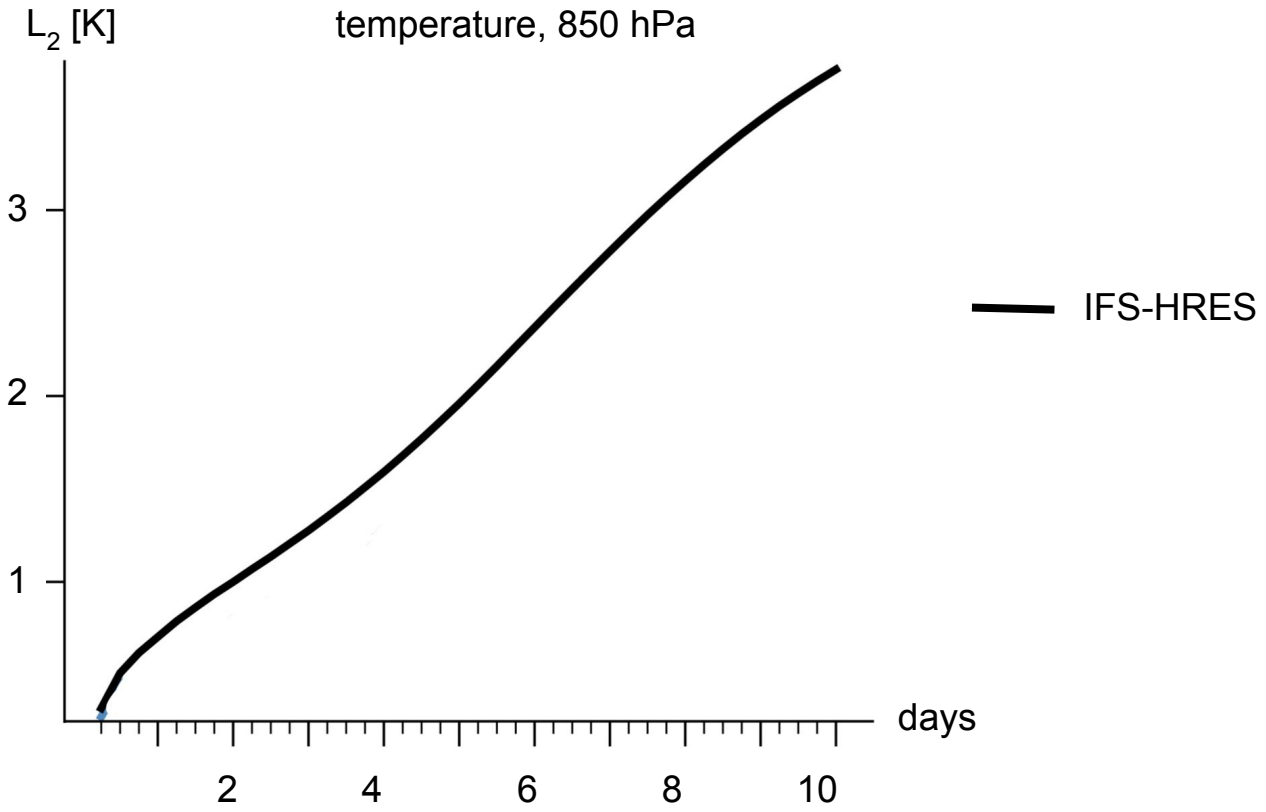



ECMWF machine learning roadmap

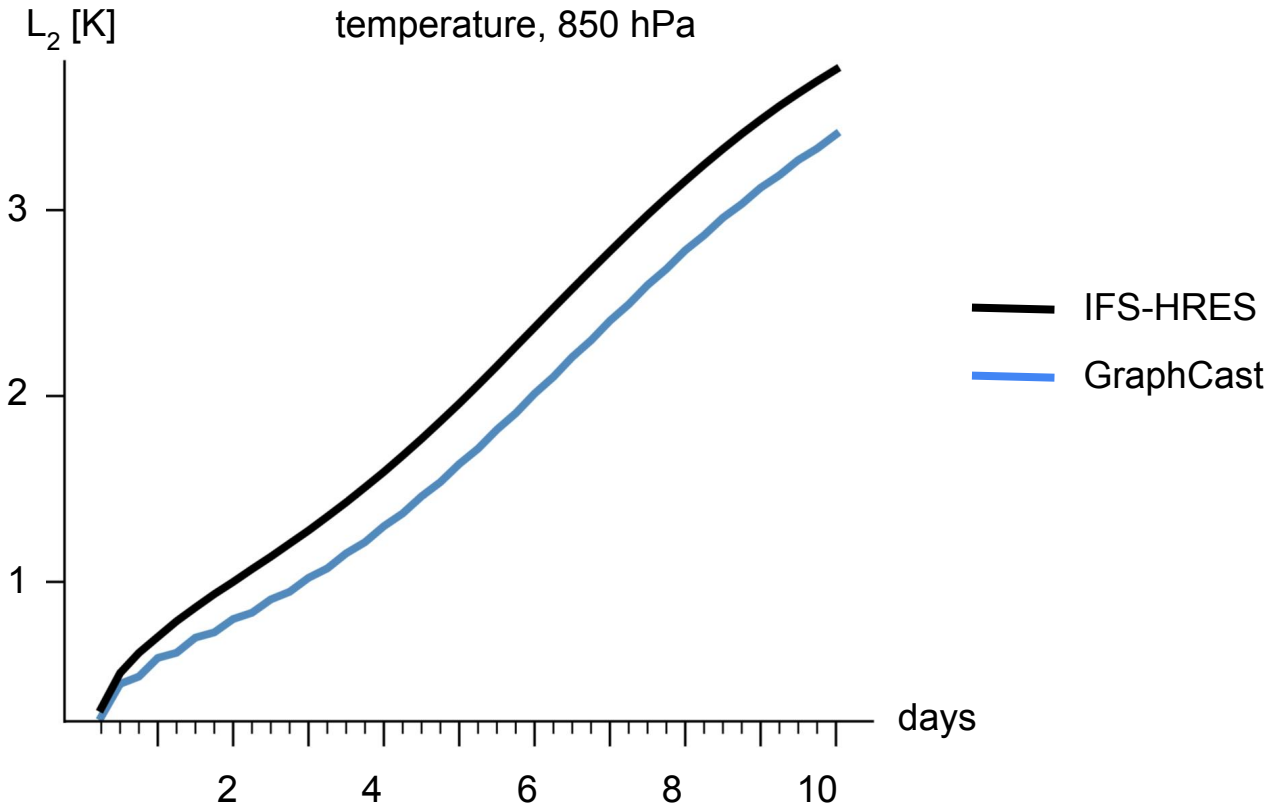
# How it is going



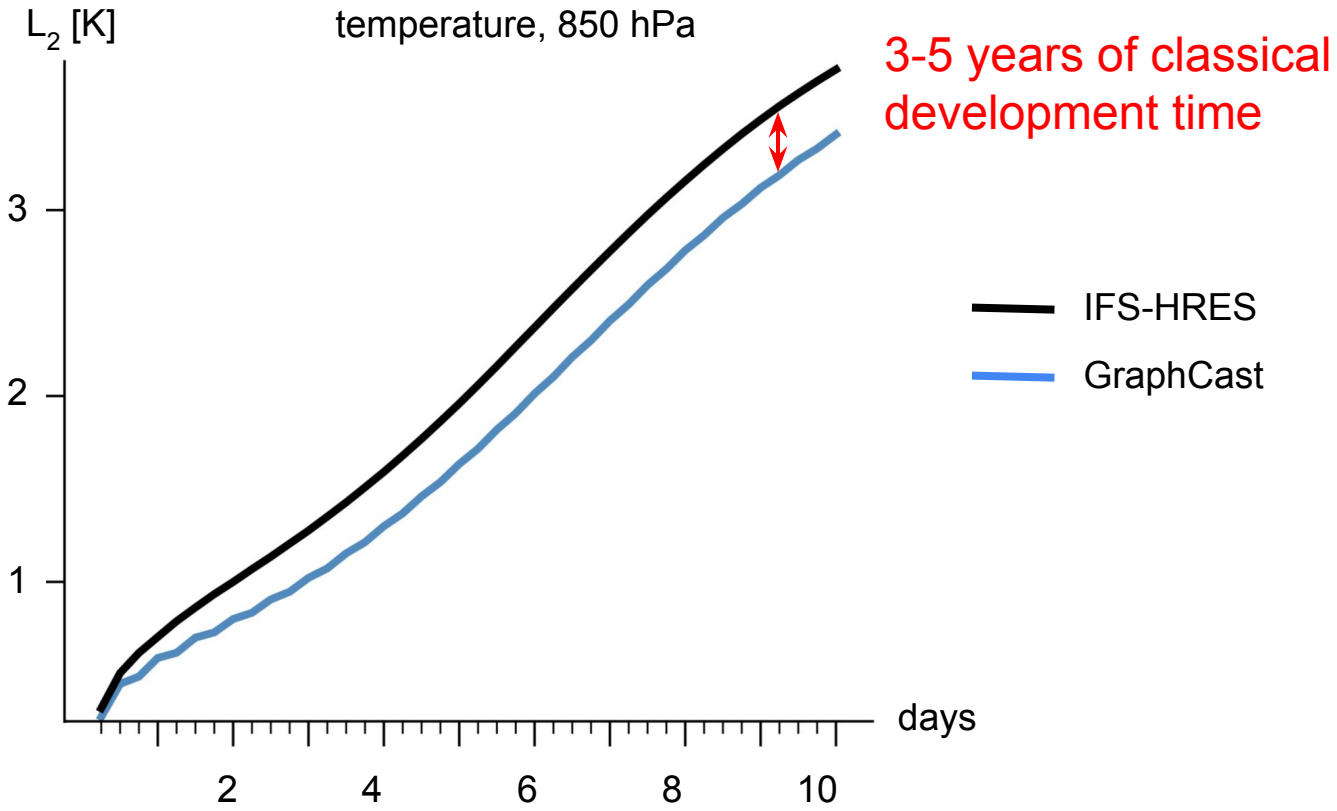
# How it is going



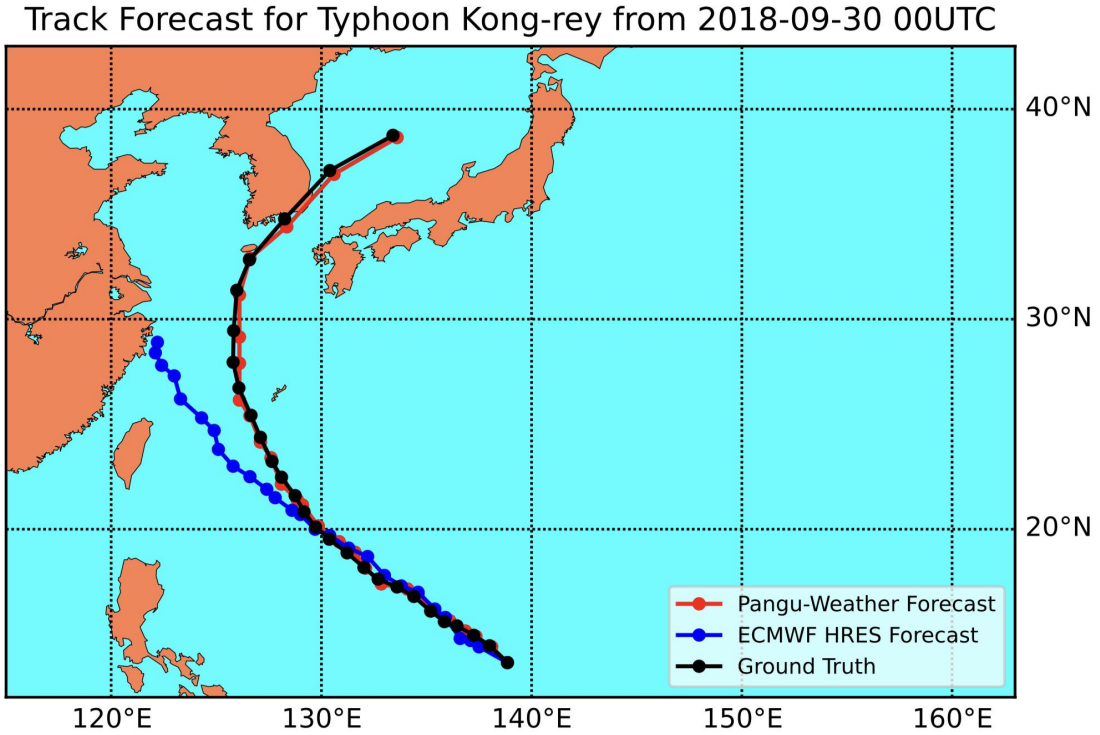
# How it is going



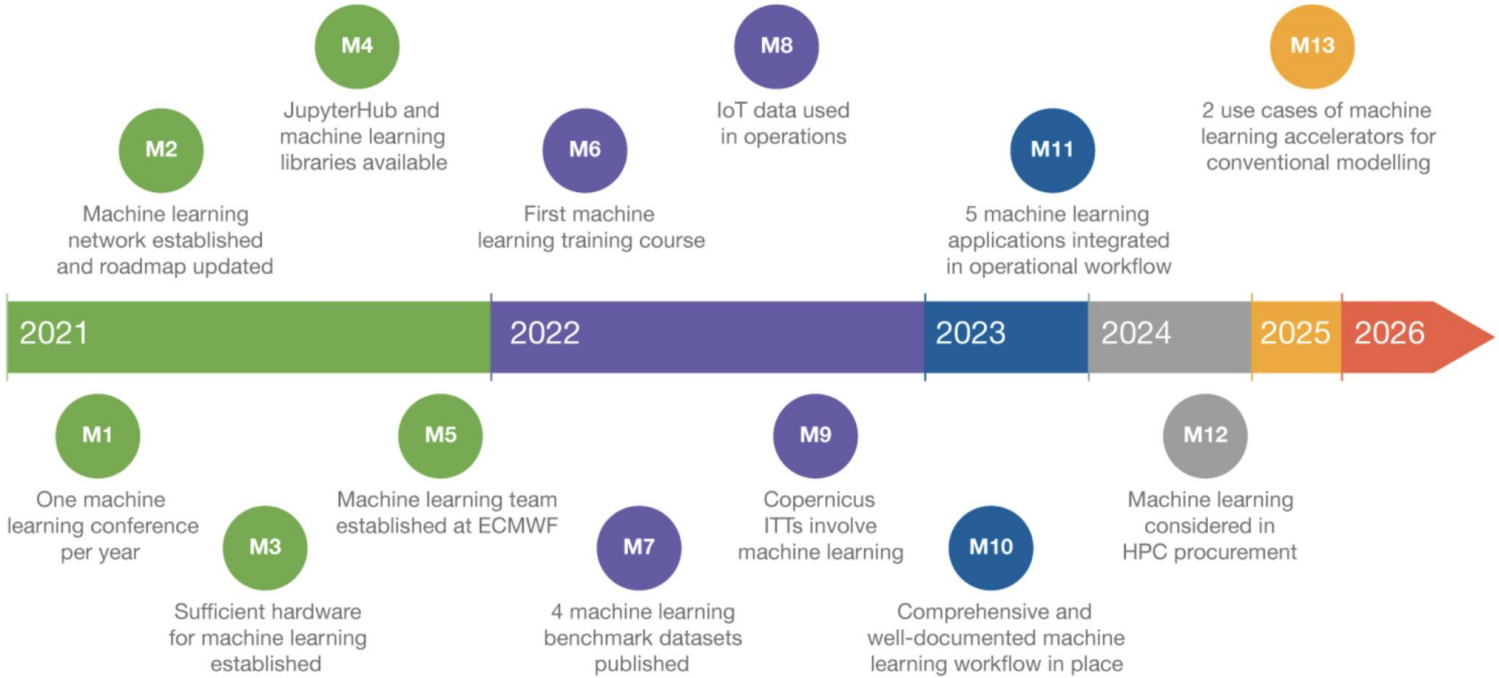
# How it is going



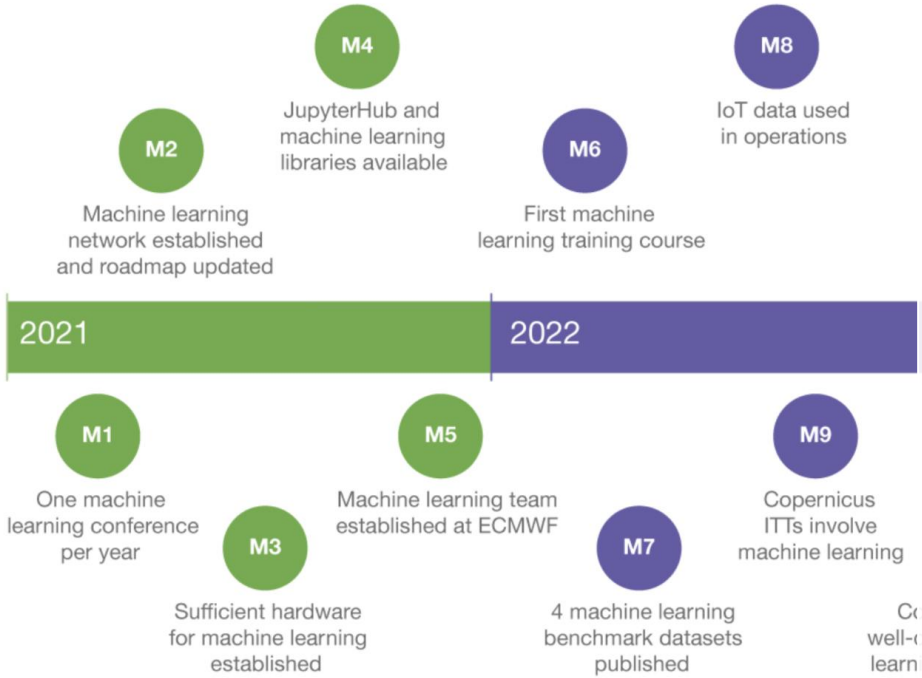
# How it is going



# How it is going

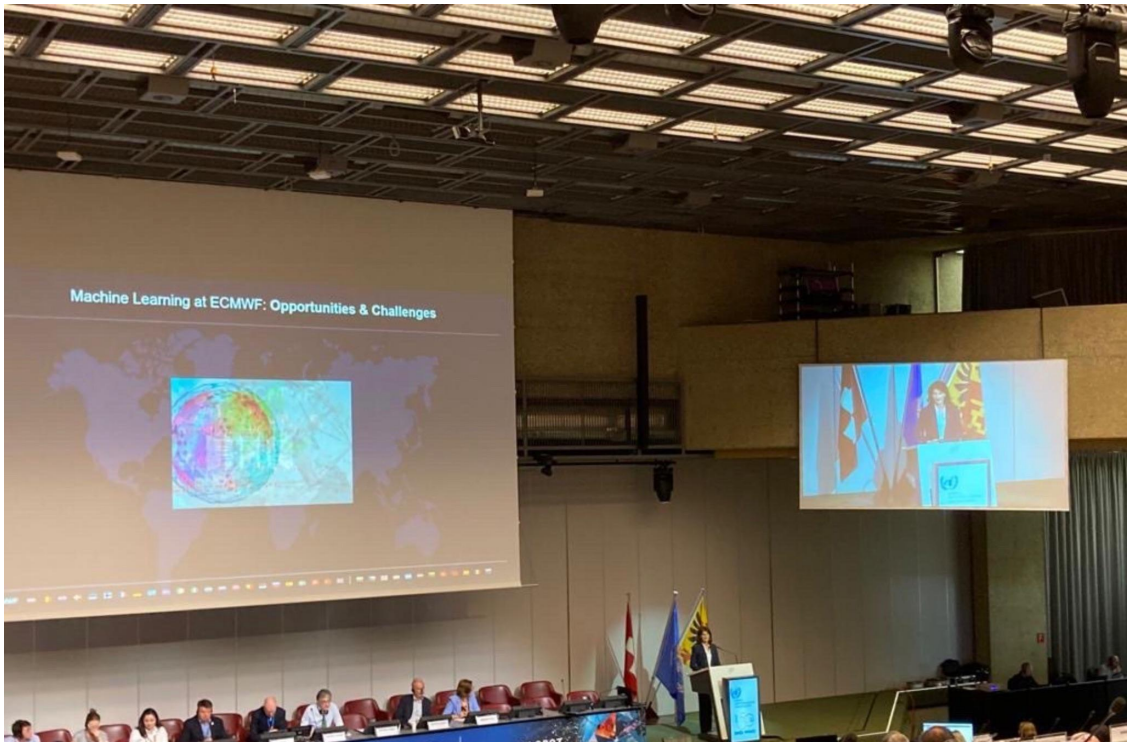


# How it is going




**Machine learning methods substantially outperform classical numerical schemes**

# How it is going




19th WMO congress (last weekend)

# What enables GraphCast, PanguWeather, …?

# What enables GraphCast, PanguWeather, …?

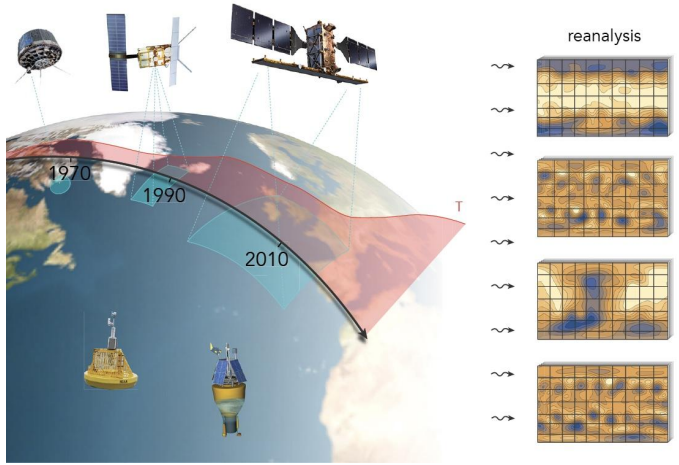
All purely data-driven

- Trained on ERA5 reanalysis: post-processed observations
- Very large dataset (6 PB) readily amenable to machine learning

# What enables GraphCast, PanguWeather, …?

All purely data-driven

- Trained on ERA5 reanalysis: post-processed observations
- Very large dataset (6 PB) readily amenable to machine learning



# What enables GraphCast, PanguWeather, …?

All purely data-driven

- Trained on ERA5 reanalysis: post-processed observations
- Very large dataset (6 PB) readily amenable to machine learning

All use state-of-the-art machine learning models

- Transformer, graph neural networks, neural differential operators
- But all different ones

# What enables GraphCast, PanguWeather, …?

All purely data-driven

- Trained on ERA5 reanalysis: post-processed observations
- Very large dataset (6 PB) readily amenable to machine learning

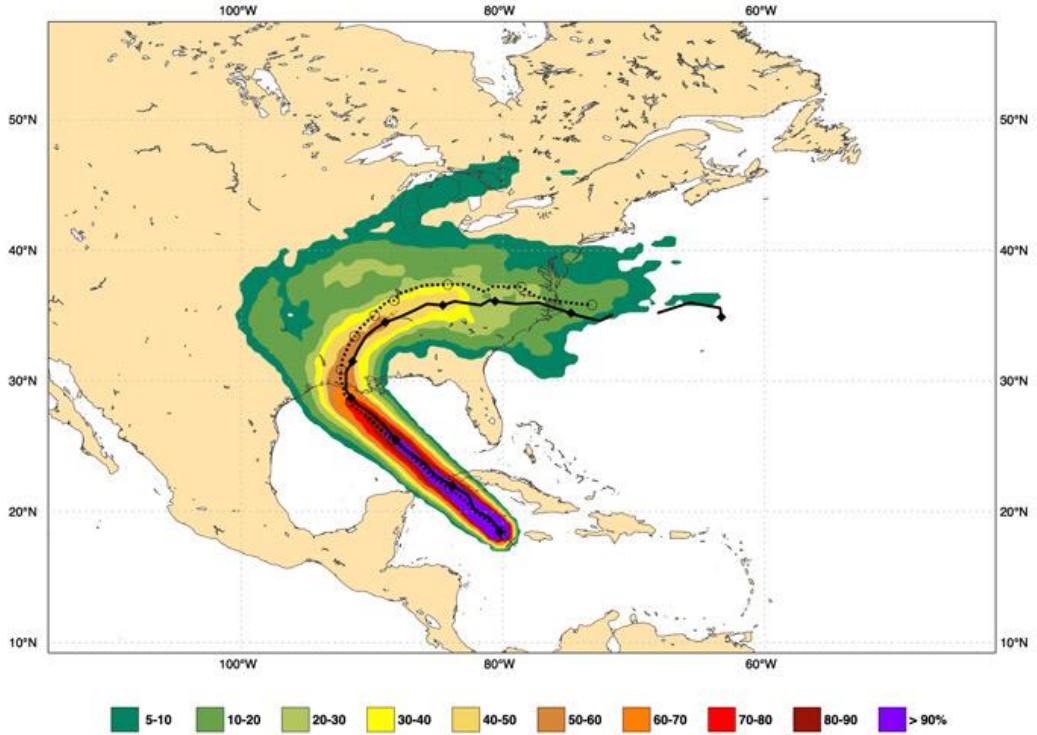
All use state-of-the-art machine learning models

- Transformer, graph neural networks, neural differential operators
- But all different ones

All by big industry labs

- Weather forecasting has clear metrics/benchmarks for comparison
- High societal impact

# Dynamics are not really described by primitive equations



# Theoretical formulation

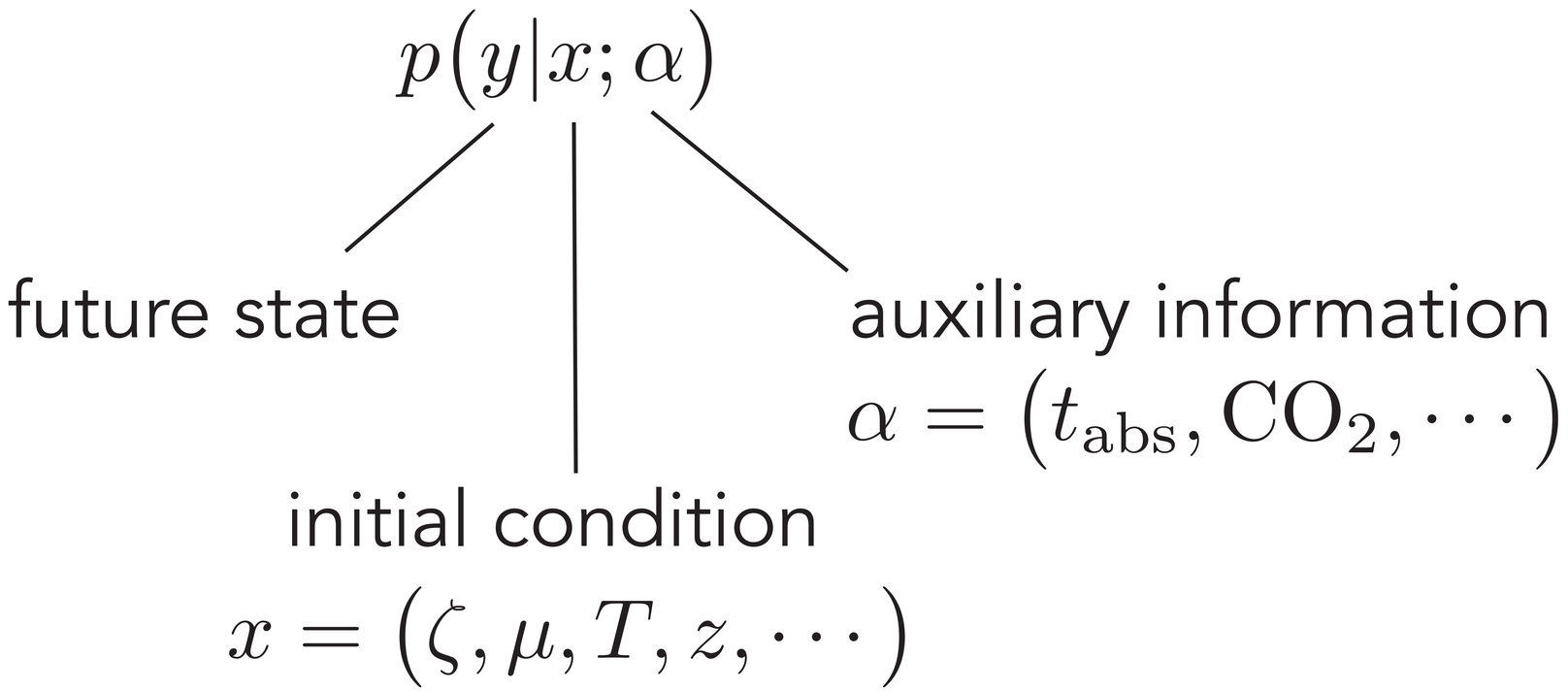
- Atmosphere as stochastic dynamical system:

$$p\big(y|x;\alpha\big)$$

© Christian Lessig, 2023

# Theoretical formulation

- Atmosphere as stochastic dynamical system:




© Christian Lessig, 2023

# Theoretical formulation

- Atmosphere as stochastic dynamical system:

$$p\big(y|x;\alpha\big)$$

  - Forecasting, downscaling, interpolation: $p\big(y|x;\alpha\big)$

© Christian Lessig, 2023

# Theoretical formulation

- Atmosphere as stochastic dynamical system:

$$p\big(y|x;\alpha\big)$$

  - Forecasting, downscaling, interpolation: $p\big(y|x;\alpha\big)$
  - Counterfactuals: $p\big(y|x;\hat{\alpha}\big)$

© Christian Lessig, 2023

# Theoretical formulation

- Atmosphere as stochastic dynamical system:

$$p\big(y|x;\alpha\big)$$

  - Forecasting, downscaling, interpolation: $p\big(y|x;\alpha\big)$
  - Counterfactuals: $p\big(y|x;\hat{\alpha}\big)$
  - Climate: $p_\alpha(y) = \int p\big(y|x;\alpha\big)\, p(x)\, \mathrm{d}x$

© Christian Lessig, 2023

# Theoretical formulation

- Atmosphere as stochastic dynamical system:

$$p(y|x;\alpha)$$ highly complex, in-stationary distribution

  - Forecasting, downscaling, interpolation: $p(y|x;\alpha)$
  - Counterfactuals: $p(y|x;\hat{\alpha})$
  - Climate: $p_\alpha(y) = \int p(y|x;\alpha)\, p(x)\, \mathrm{d}x$

© Christian Lessig, 2023

# Theoretical formulation

- Atmosphere as stochastic dynamical system:

$$p\big(y|x;\alpha\big)$$

- Numerical statistical atmospheric model:

$$\tilde{p}\big(\tilde{y}|\tilde{x};\tilde{\alpha}\big)$$

© Christian Lessig, 2023

# Theoretical formulation

- Atmosphere as stochastic dynamical system:

$$p\big(y|x;\alpha\big)$$

- Numerical statistical atmospheric model:

$$\tilde{p}\big(\tilde{y}|\tilde{x};\tilde{\alpha}\big) \approx p\big(y|x;\alpha\big)$$

© Christian Lessig, 2023

# Theoretical formulation

- Atmosphere as stochastic dynamical system:

$$p\big(y|x;\alpha\big)$$

- Numerical statistical atmospheric model:

$$\tilde{p}\big(\tilde{y}|\tilde{x};\tilde{\alpha}\big) \approx p\big(y|x;\alpha\big)$$

approx. future state  approx. initial condition

© Christian Lessig, 2023

# Theoretical formulation

- Atmosphere as stochastic dynamical system:

$$p\big(y|x;\alpha\big)$$

- Numerical statistical atmospheric model:

$$\tilde{p}\big(\tilde{y}|\tilde{x};\tilde{\alpha}\big) \approx p\big(y|x;\alpha\big)$$

highly complex, instationary distribution

© Christian Lessig, 2023

# Theoretical formulation

- Atmosphere as stochastic dynamical system:

$$p\big(y|x;\alpha\big)$$

- Numerical statistical atmospheric model:

very large neural network

$$\tilde{p}\big(\tilde{y}|\tilde{x};\tilde{\alpha}\big) \approx p\big(y|x;\alpha\big)$$

highly complex, instationary distribution

© Christian Lessig, 2023

# Theoretical formulation

- Atmosphere as stochastic dynamical system:

$$p\big(y|x;\alpha\big)$$

- Numerical statistical atmospheric model:

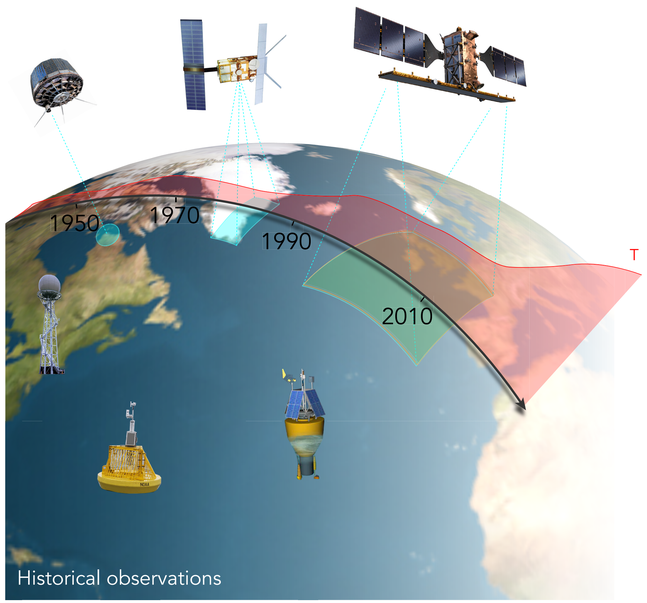
very large neural network

$$\tilde{p}_\theta\big(\tilde{y}|\tilde{x};\tilde{\alpha}\big) \approx p\big(y|x;\alpha\big)$$

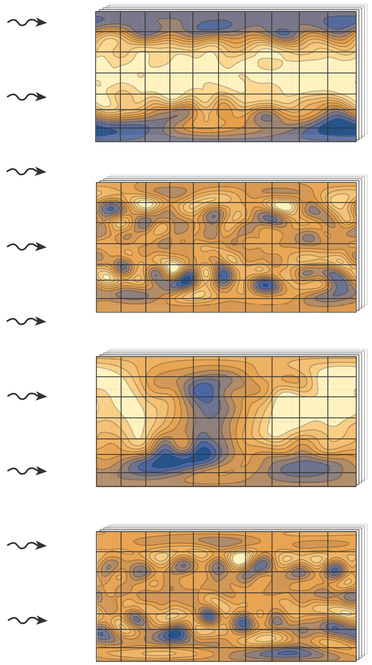
highly complex, instationary distribution

© Christian Lessig, 2023

# AtmoRep

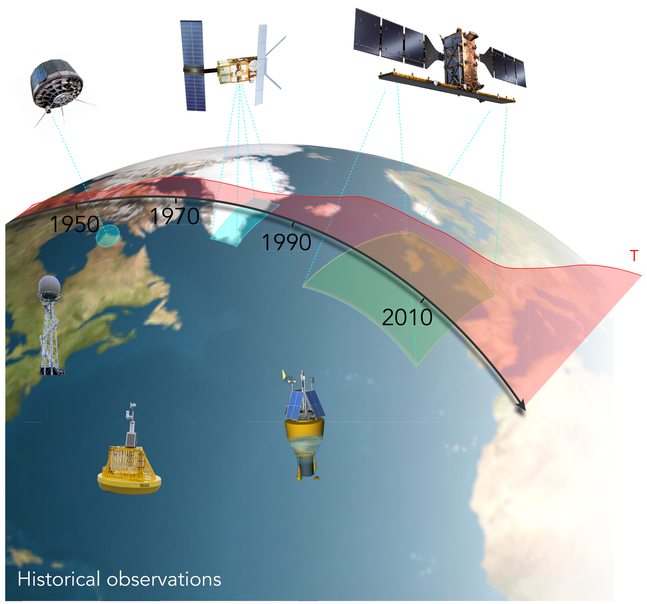



ERA5 reanalysis

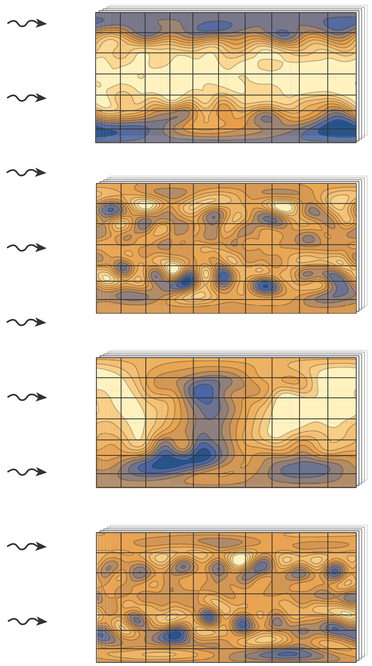


© Christian Lessig, 2023

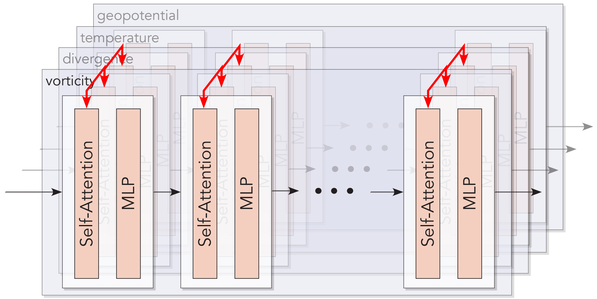
# AtmoRep




ERA5 reanalysis

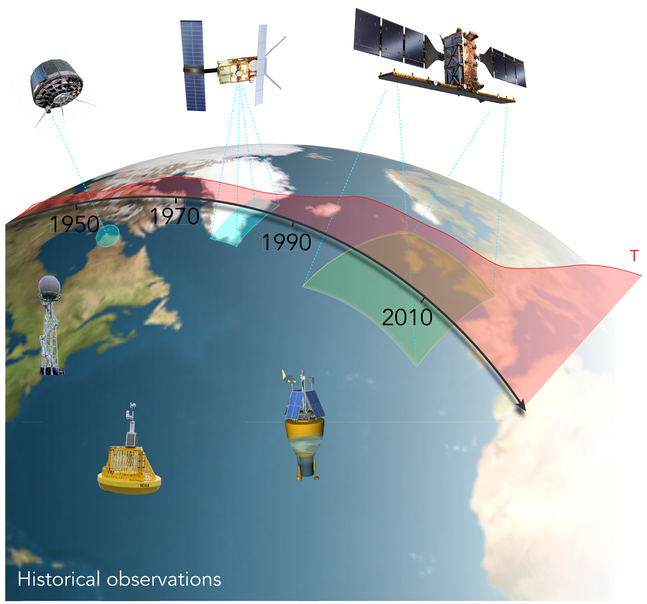


$$p_\theta(y|x)$$

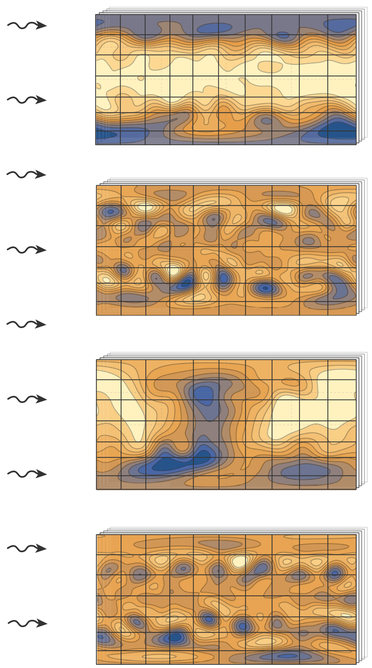



© Christian Lessig, 2023

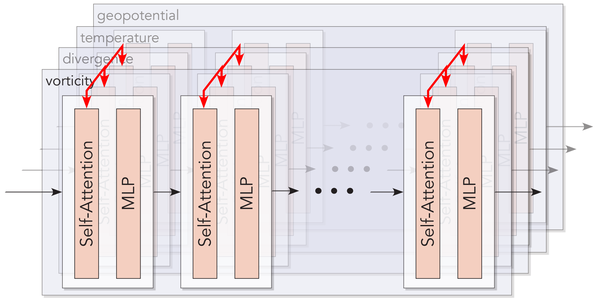
# AtmoRep




ERA5 reanalysis



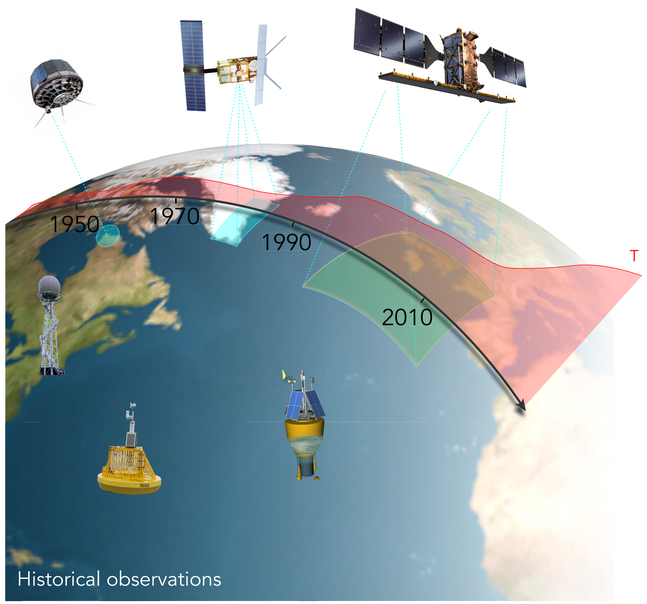
$$p_\theta(y|x)$$



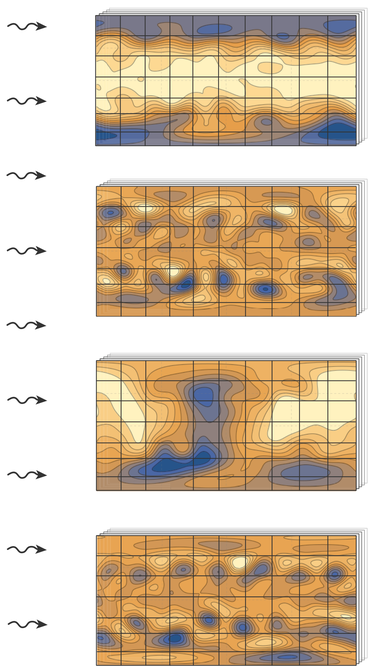

large transformer
with 3.5 x $10^9$ parameters

© Christian Lessig, 2023

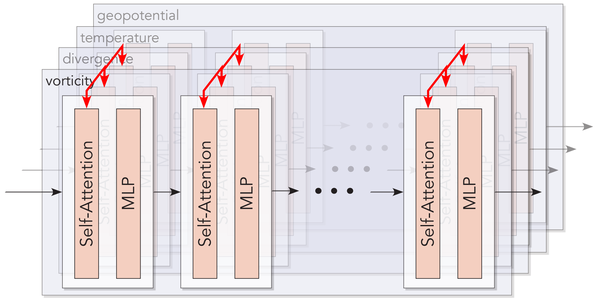
# AtmoRep




ERA5 reanalysis

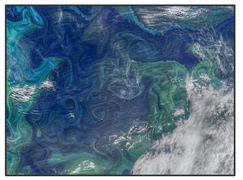


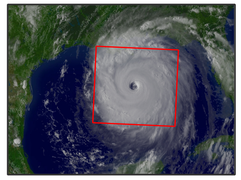
$p_\theta(y|x)$

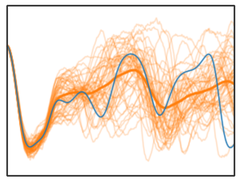


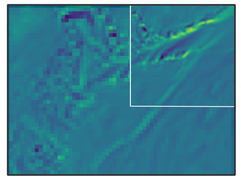

large transformer
with 3.5 x $10^9$ parameters

applications


Forecasting

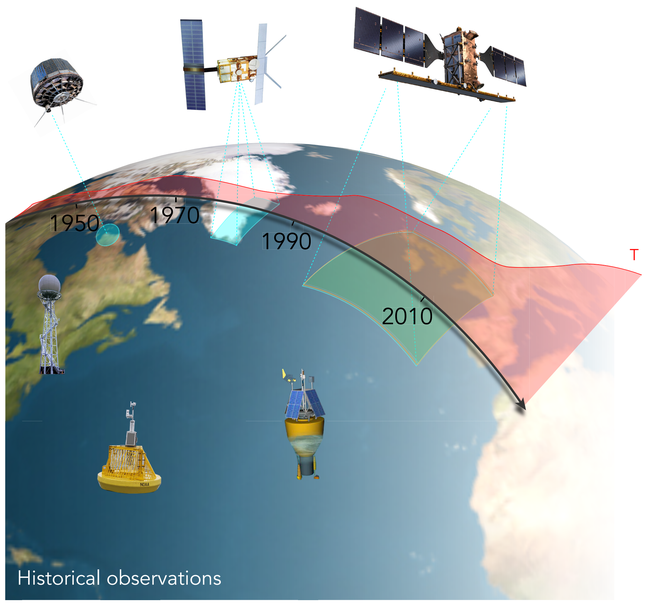

Impact analysis


Statistical analysis

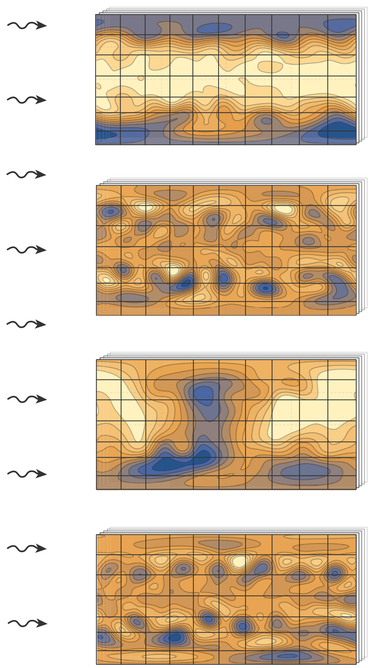

Downscaling

© Christian Lessig, 2023

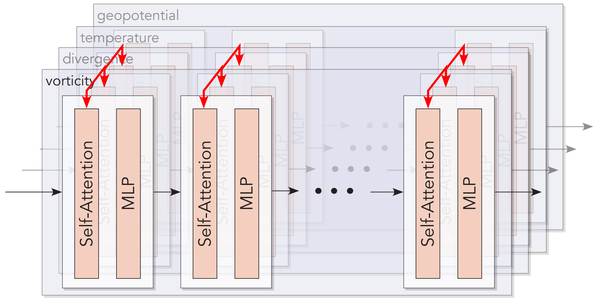
# AtmoRep




ERA5 reanalysis

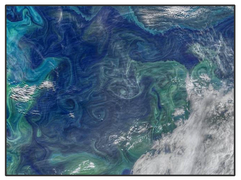


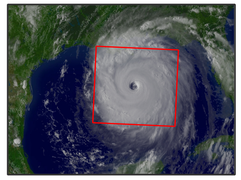
$p_\theta(y|x)$

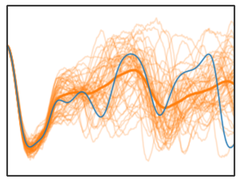


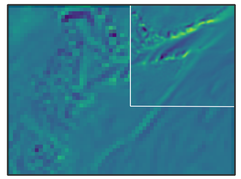

large transformer
with 3.5 x $10^9$ parameters

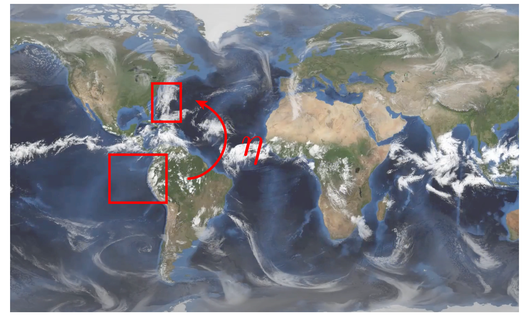
applications


Forecasting


Impact analysis


Statistical analysis


Downscaling



scientific insight

© Christian Lessig, 2023

# AtmoRep network architecture

- Transformer-based network architecture
  - Scales well to very large data-sets
  - Generative model (with decoder)
  - Effective for wide range of data sets
  - Attention maps provide (physical) interpretability

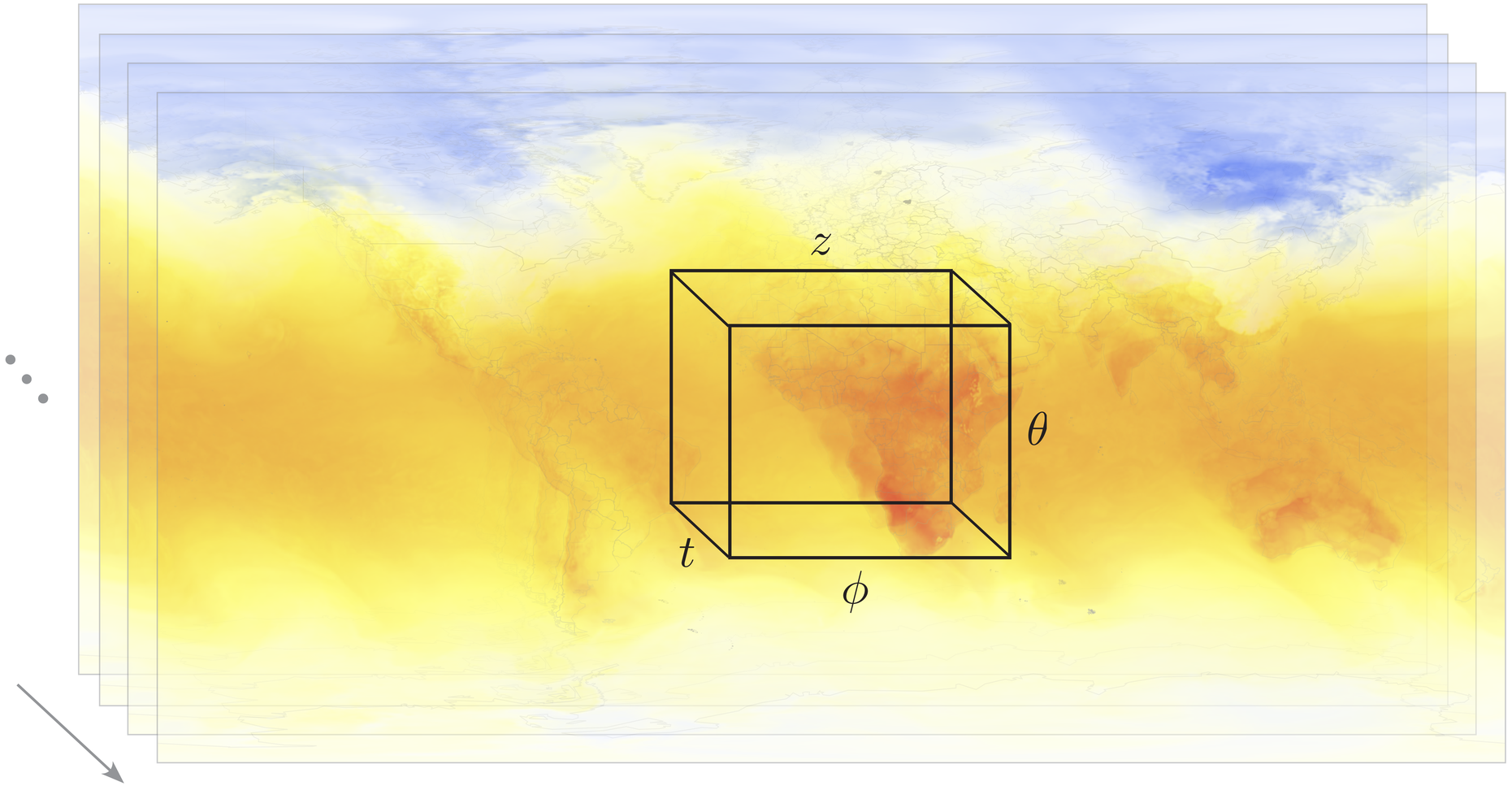
© Christian Lessig, 2023

# What is a token?

- Key property of transformer: set of tokens (=“small “inputs) is processed simultaneously
  - Language: token is word or sub-word
  - Images: token is small image patch (e.g. 16x16 pixels)
- Attention relates tokens

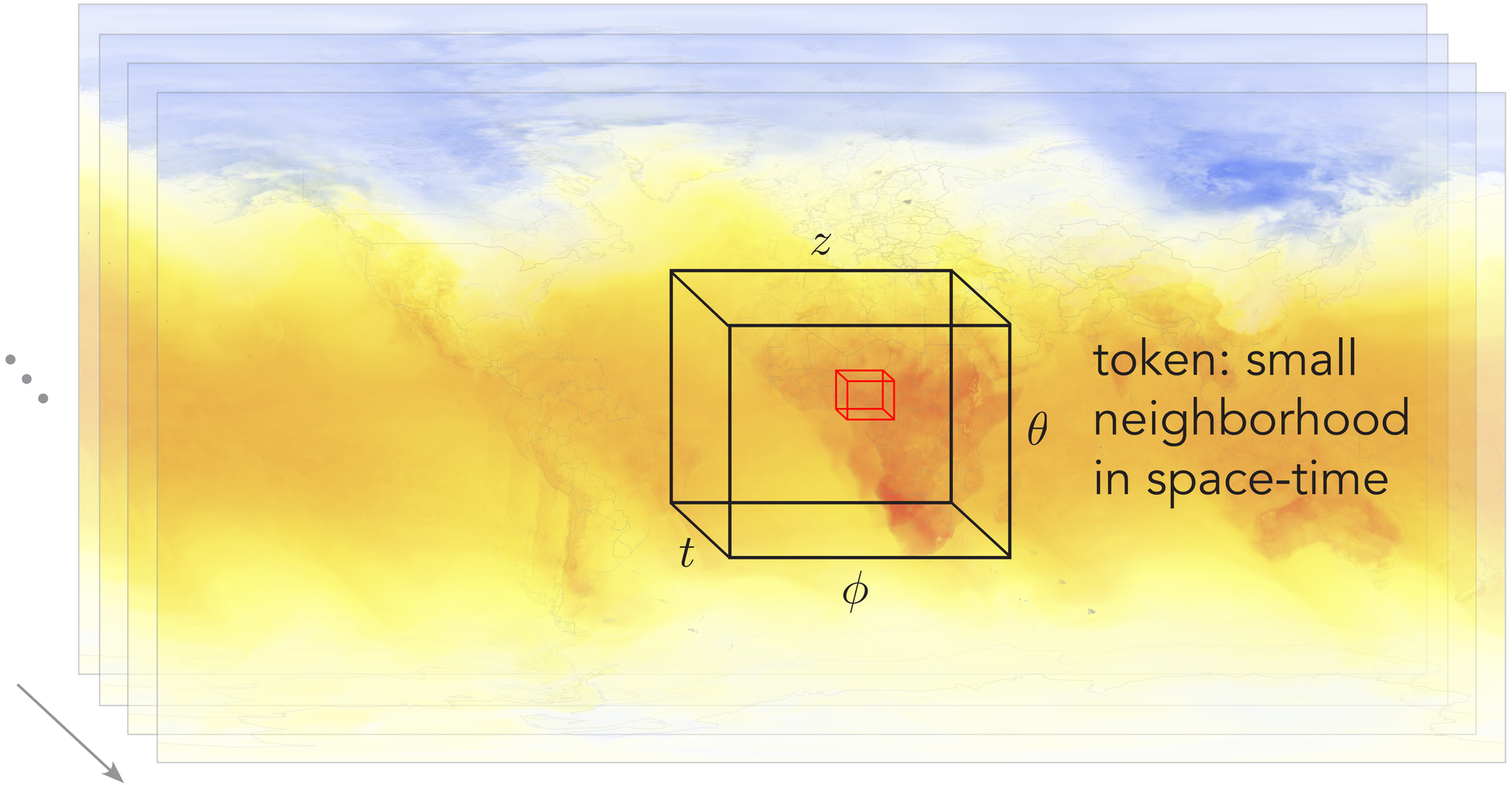
© Christian Lessig, 2023

# What is a token?



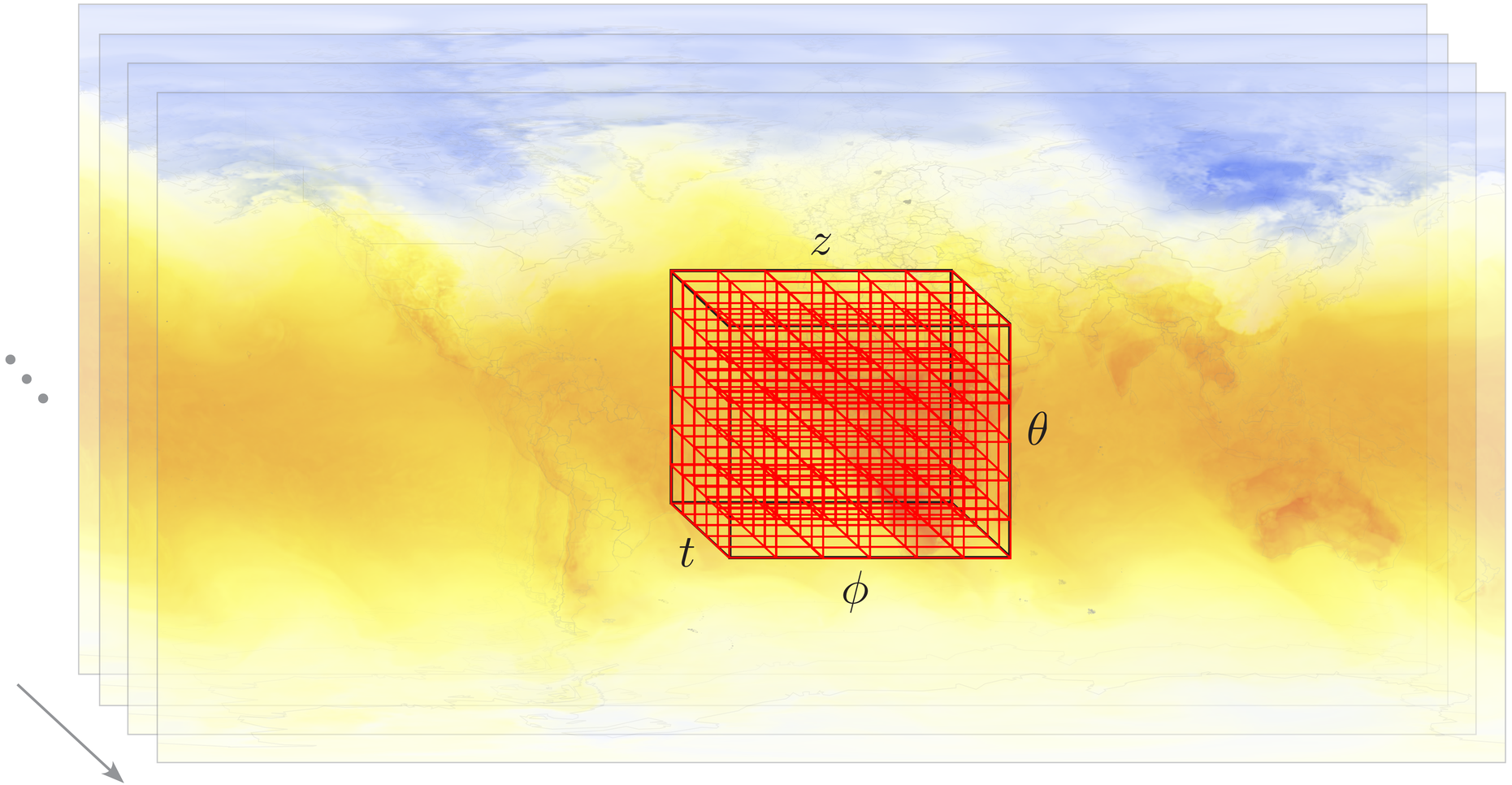

© Christian Lessig, 2023

# What is a token?




© Christian Lessig, 2023

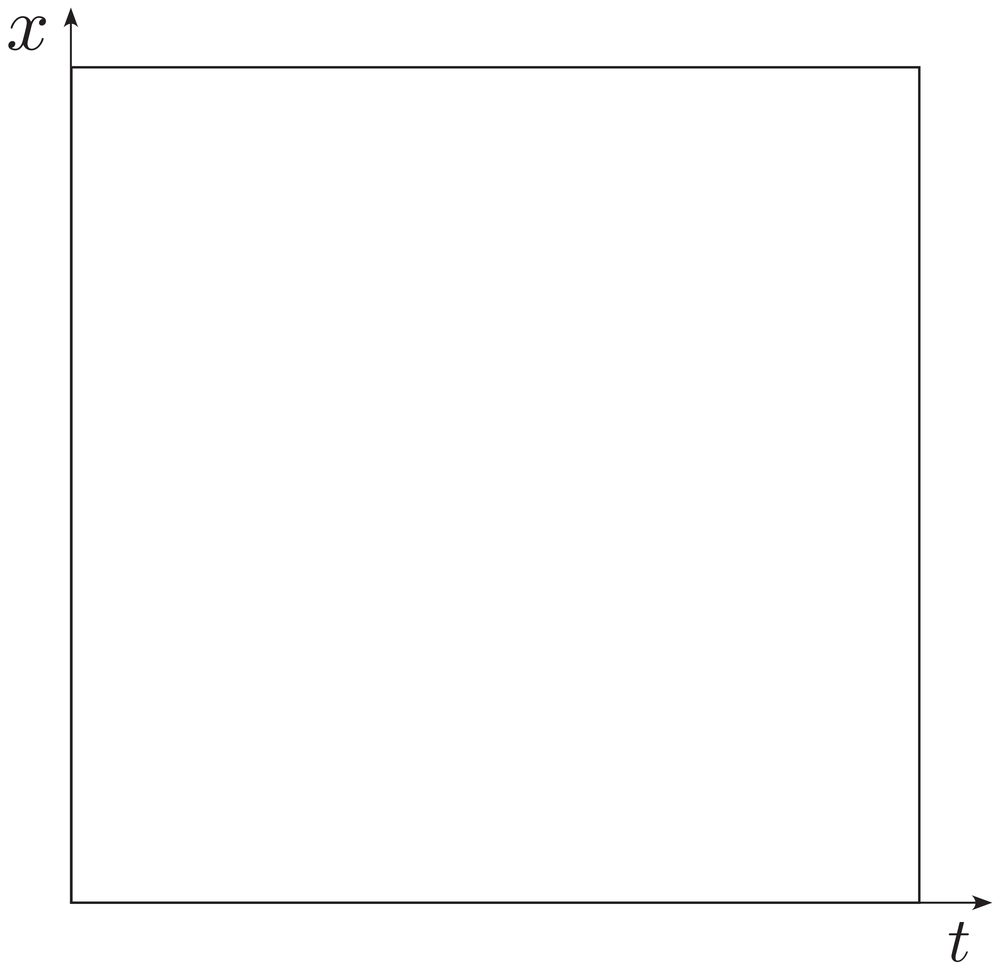
# What is a token?




© Christian Lessig, 2023

# What is a token?

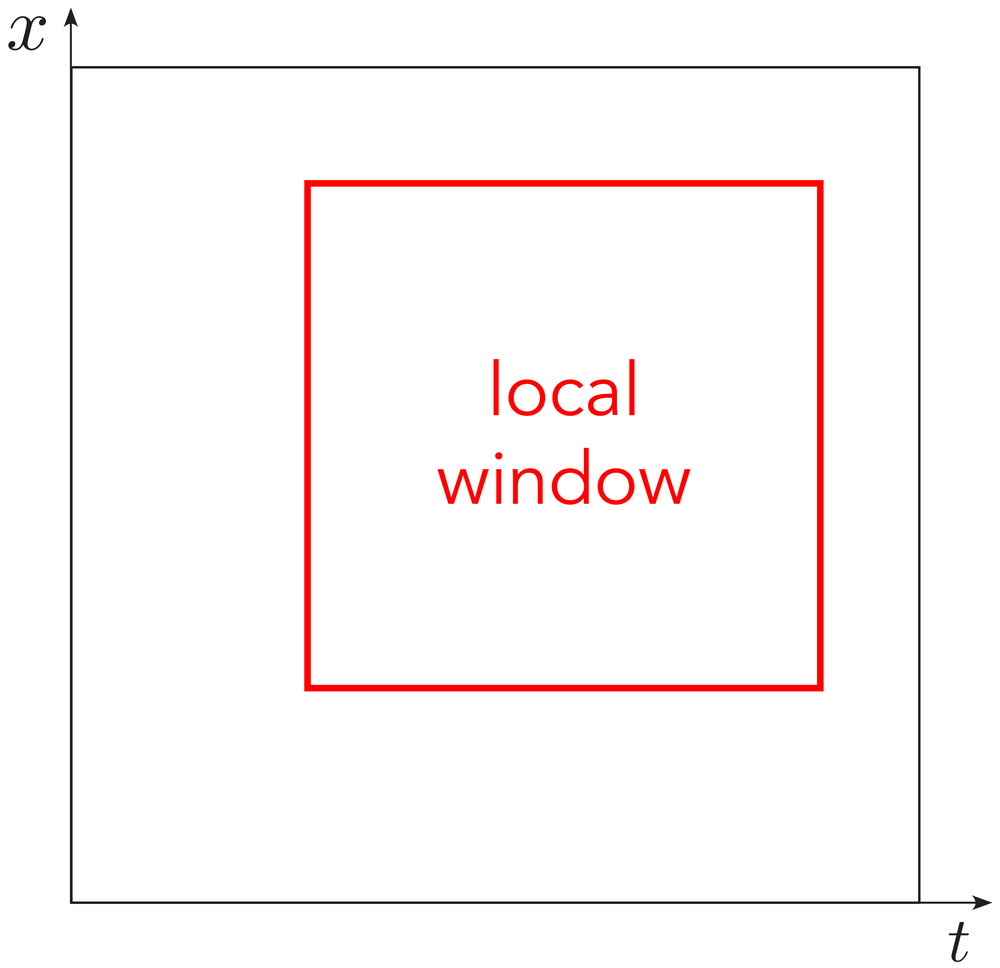
- Flatland view:




© Christian Lessig, 2023

# What is a token?

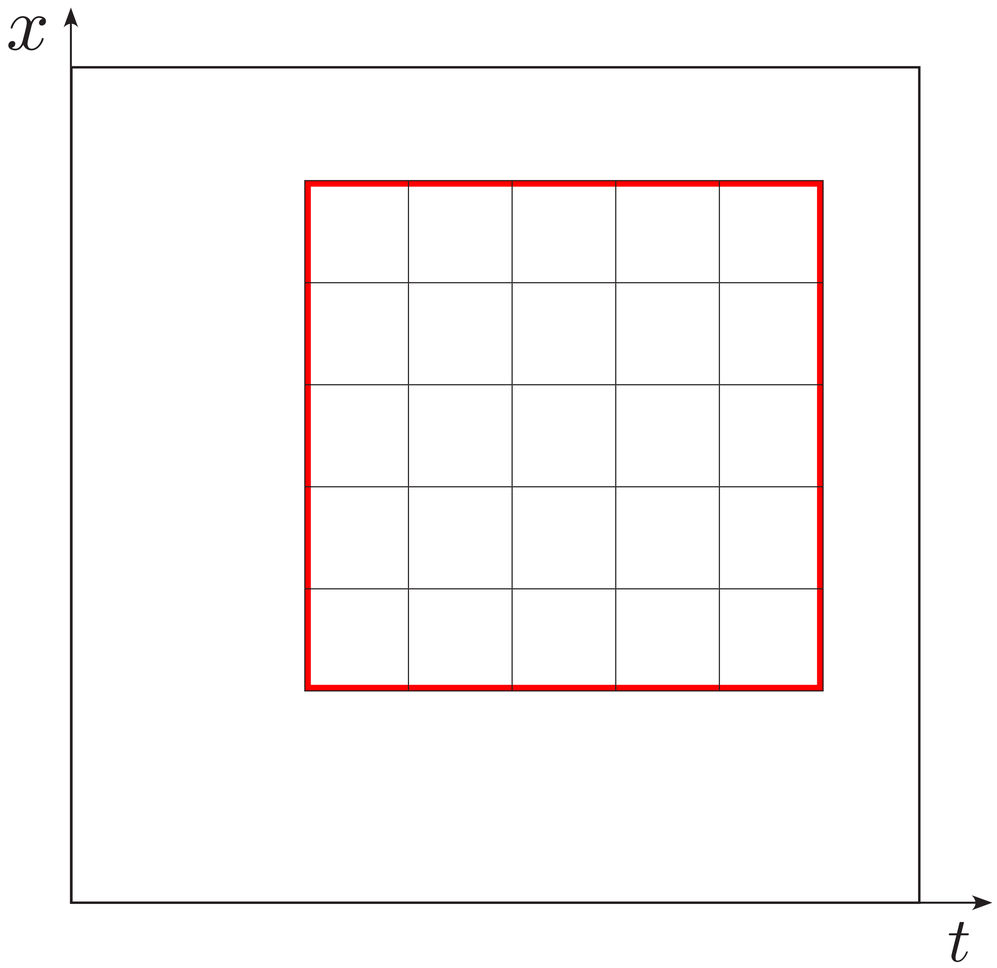
- Flatland view:




© Christian Lessig, 2023

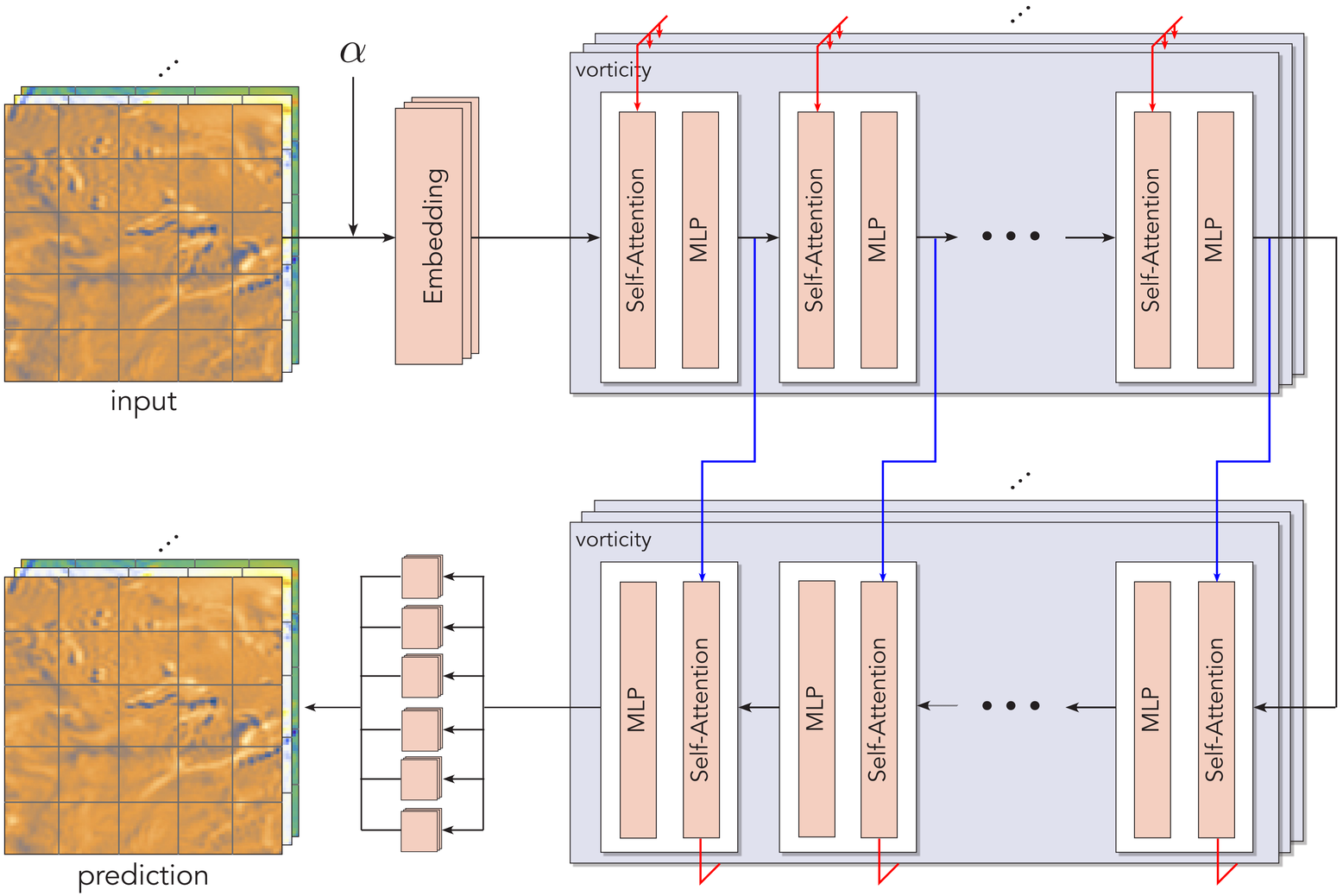
# What is a token?

- Flatland view:




© Christian Lessig, 2023

# Multiformer




© Christian Lessig, 2023

# Training objective

- Numerical statistical atmospheric model

$$p_\theta\left(y|x,\alpha\right)$$

© Christian Lessig, 2023

# Training objective

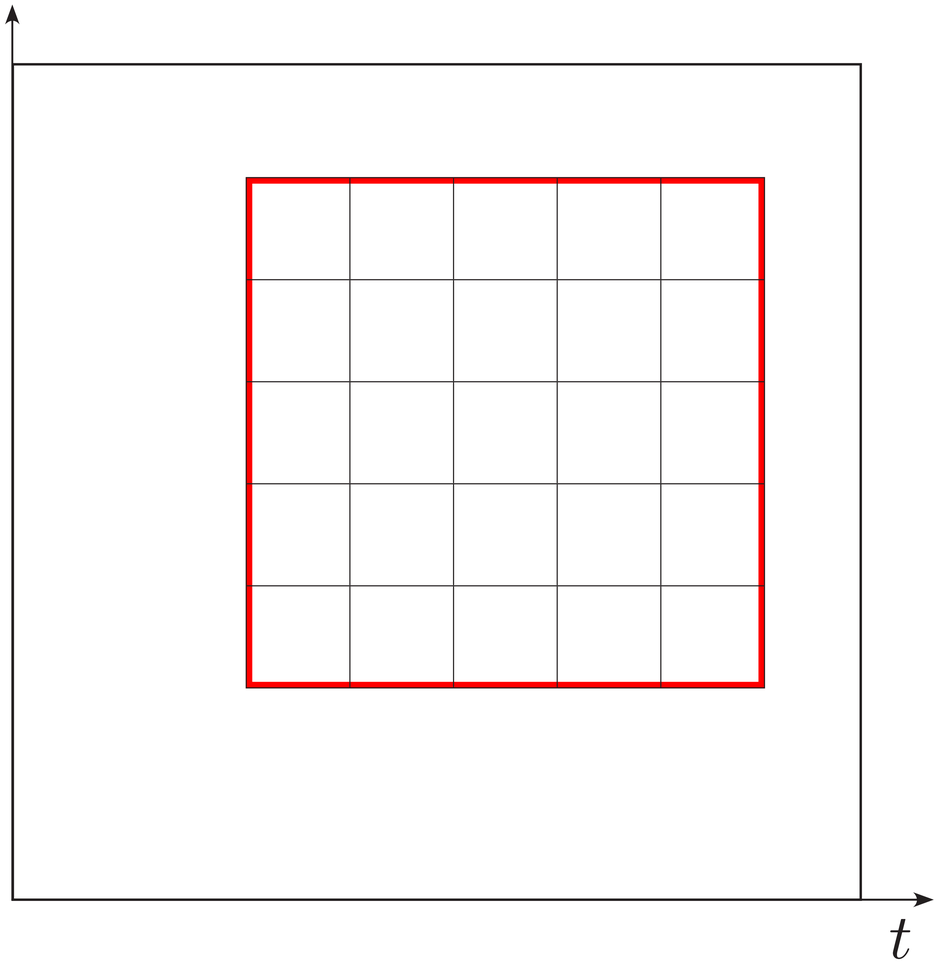
- Numerical statistical atmospheric model

$$p_\theta\left(y|x,\alpha\right)$$

  - Training should model spatio-temporal relationship between arbitrary state $x$ and $y$

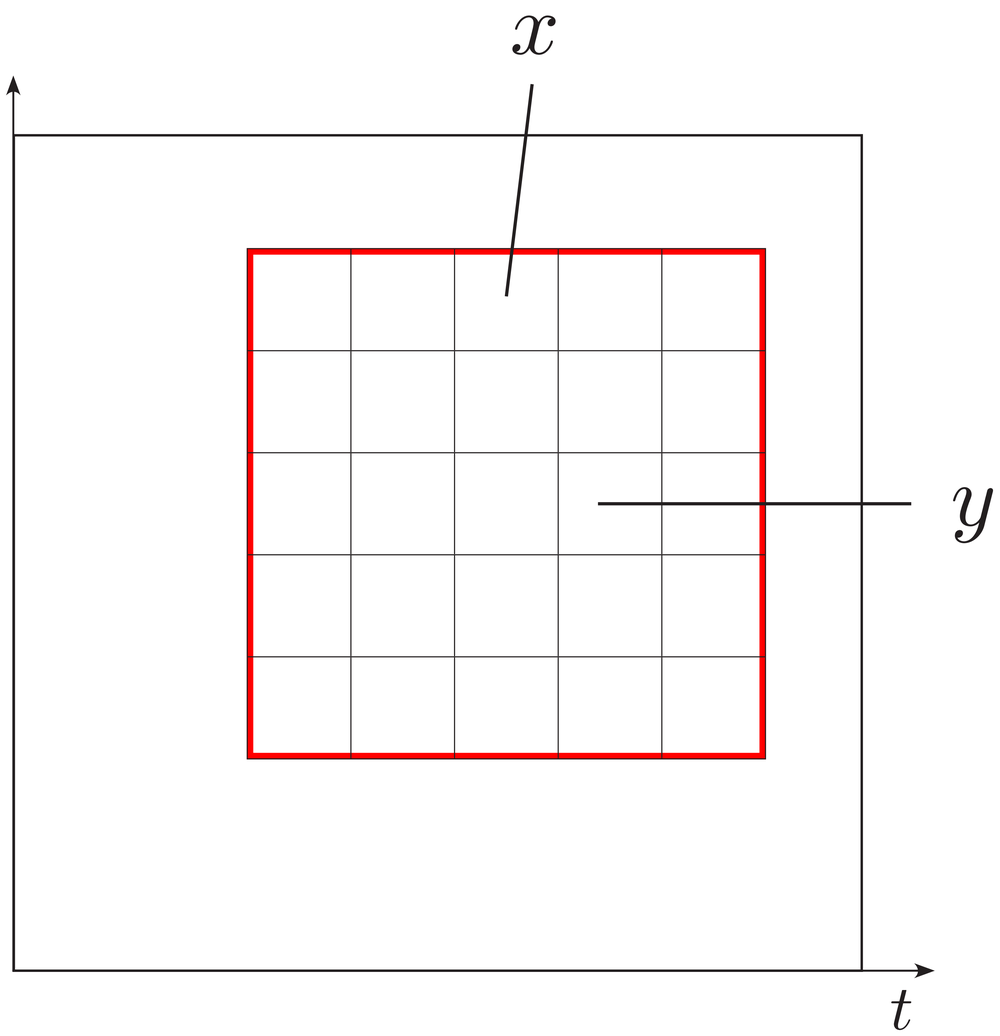
© Christian Lessig, 2023

# Training objective



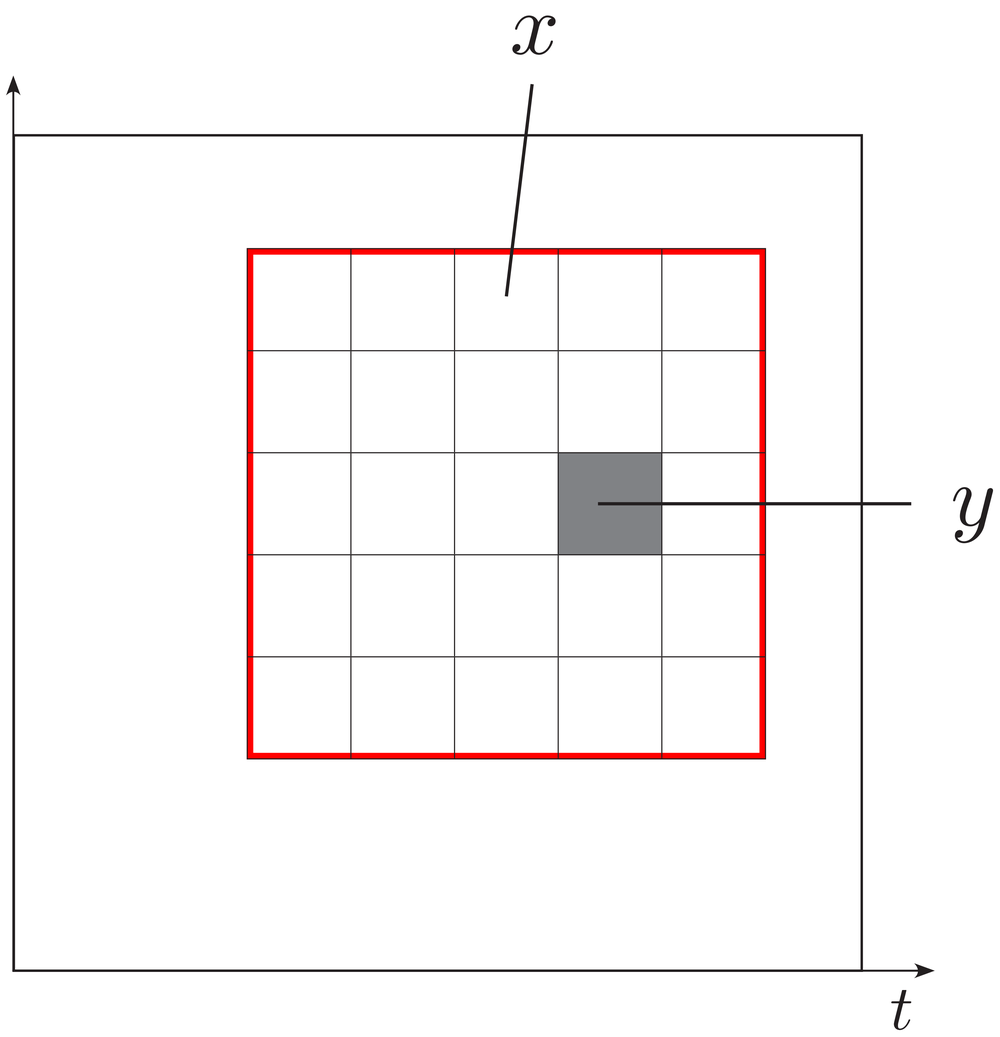

© Christian Lessig, 2023

# Training objective



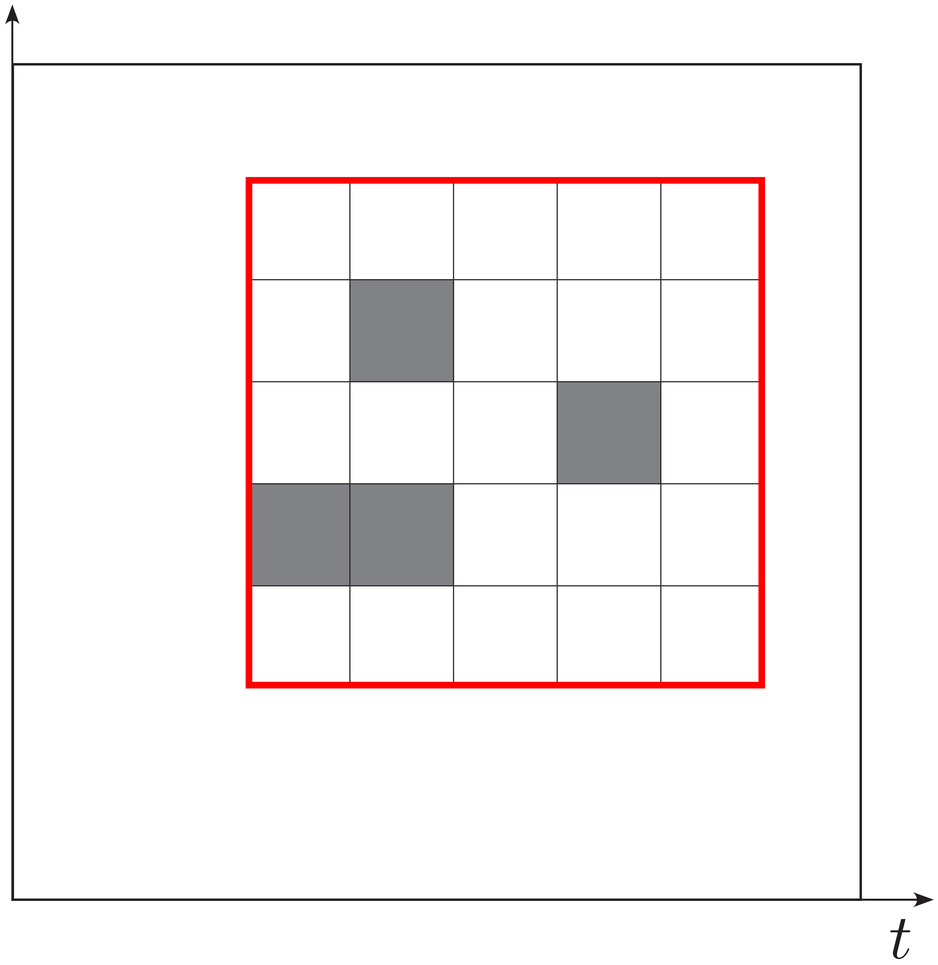

© Christian Lessig, 2023

# Training objective




© Christian Lessig, 2023

# Training objective




masked
token model:

training to predict
randomly masked
information

© Christian Lessig, 2023

# Training objective

- Numerical statistical atmospheric model:

$$p_\theta\left(y|x,\alpha\right)$$

approximate initial state

© Christian Lessig, 2023

# Training objective

- Numerical statistical atmospheric model:

$$p_\theta\left(y|x,\alpha\right)$$

approximate initial state

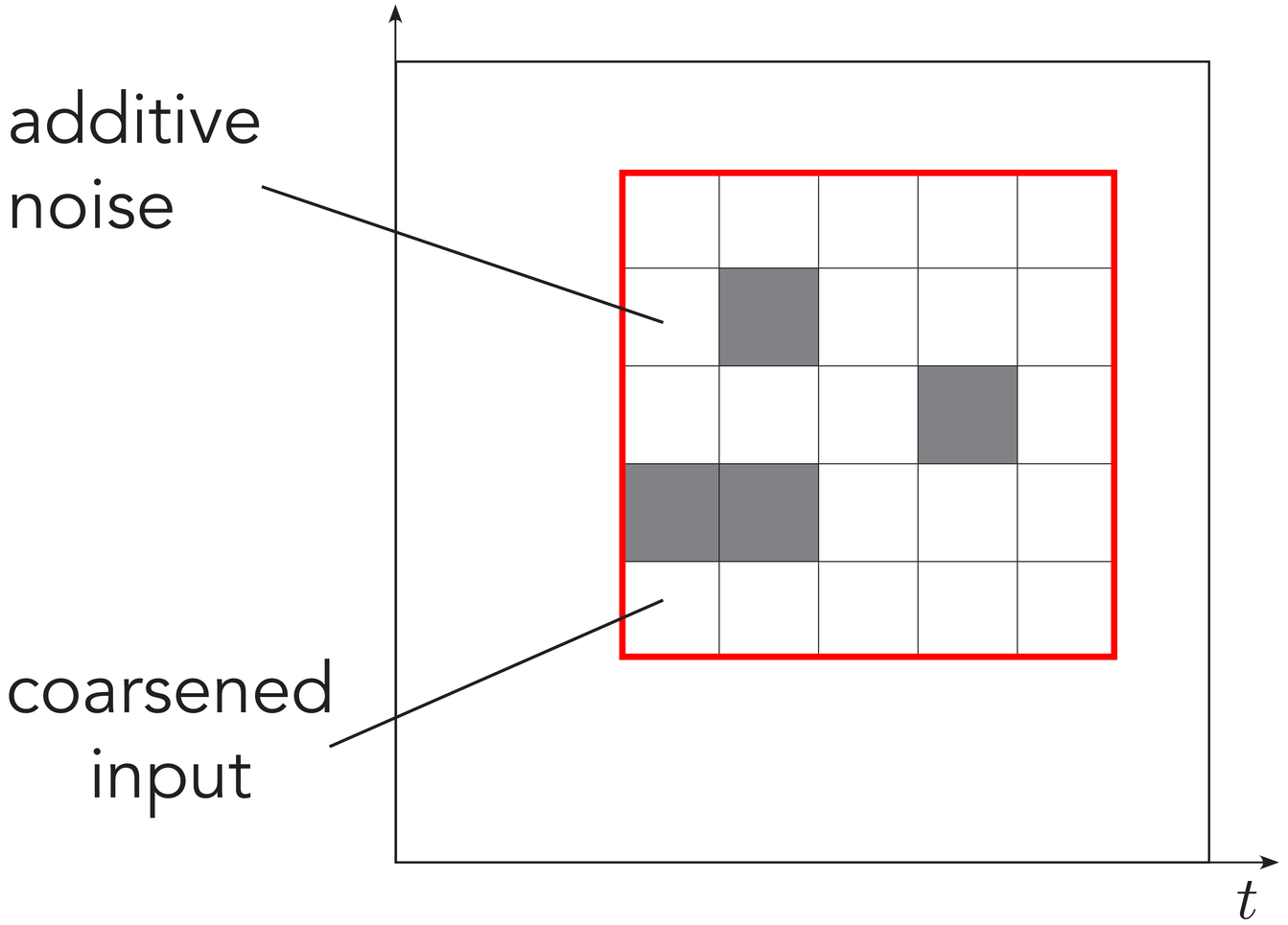
=> BERT: encourage robust, probabilistic model by randomly applying distortions to input

---

J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. BERT: pre-training of deep bidirectional transformers for language understanding. In NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers), pages 4171–4186. Association for Computational Linguistics, 2019.

© Christian Lessig, 2023

# Training objective



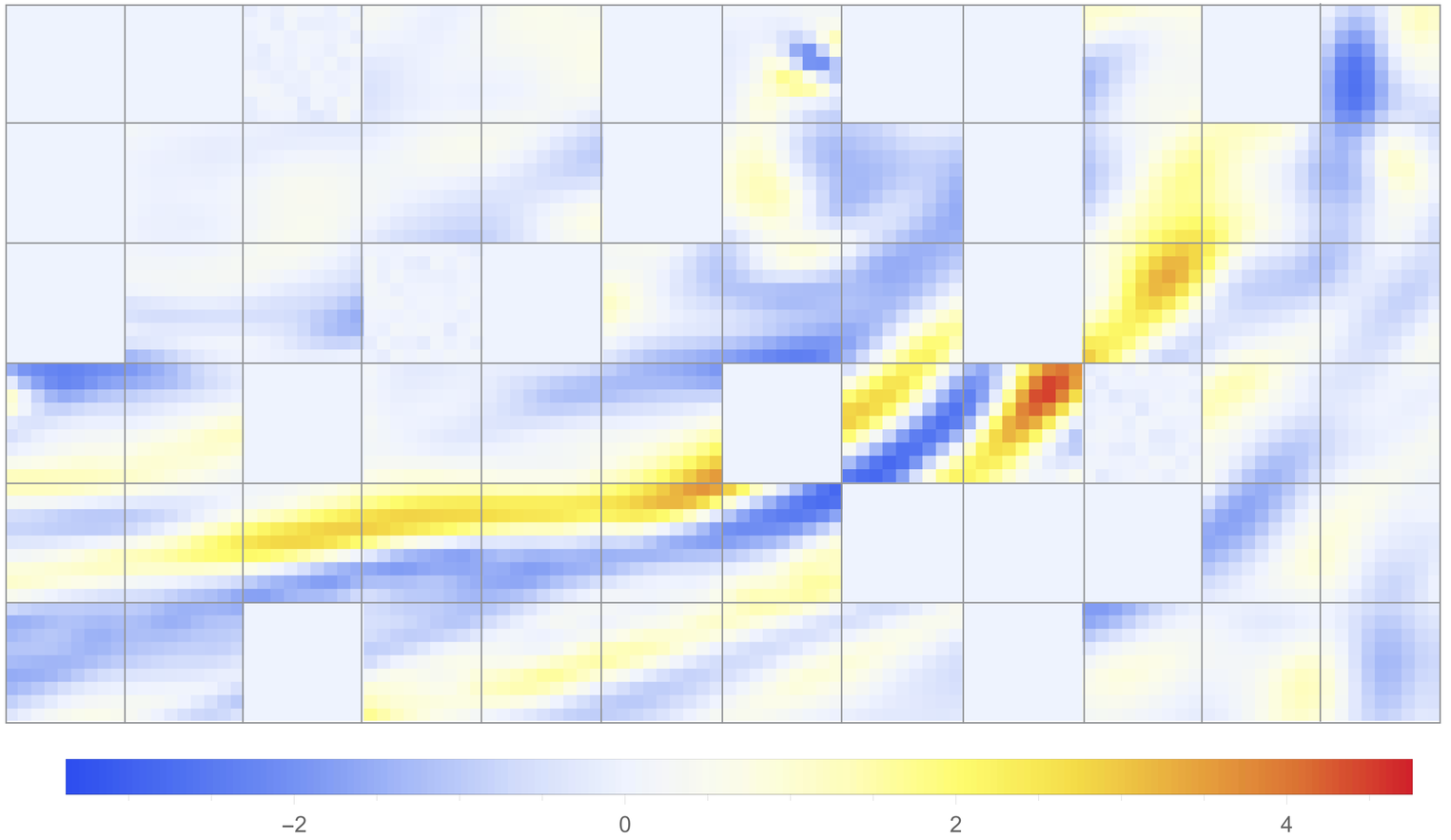

masked
token model:

training to predict
randomly masked
information

© Christian Lessig, 2023

# Training objective

divergence, ml=96



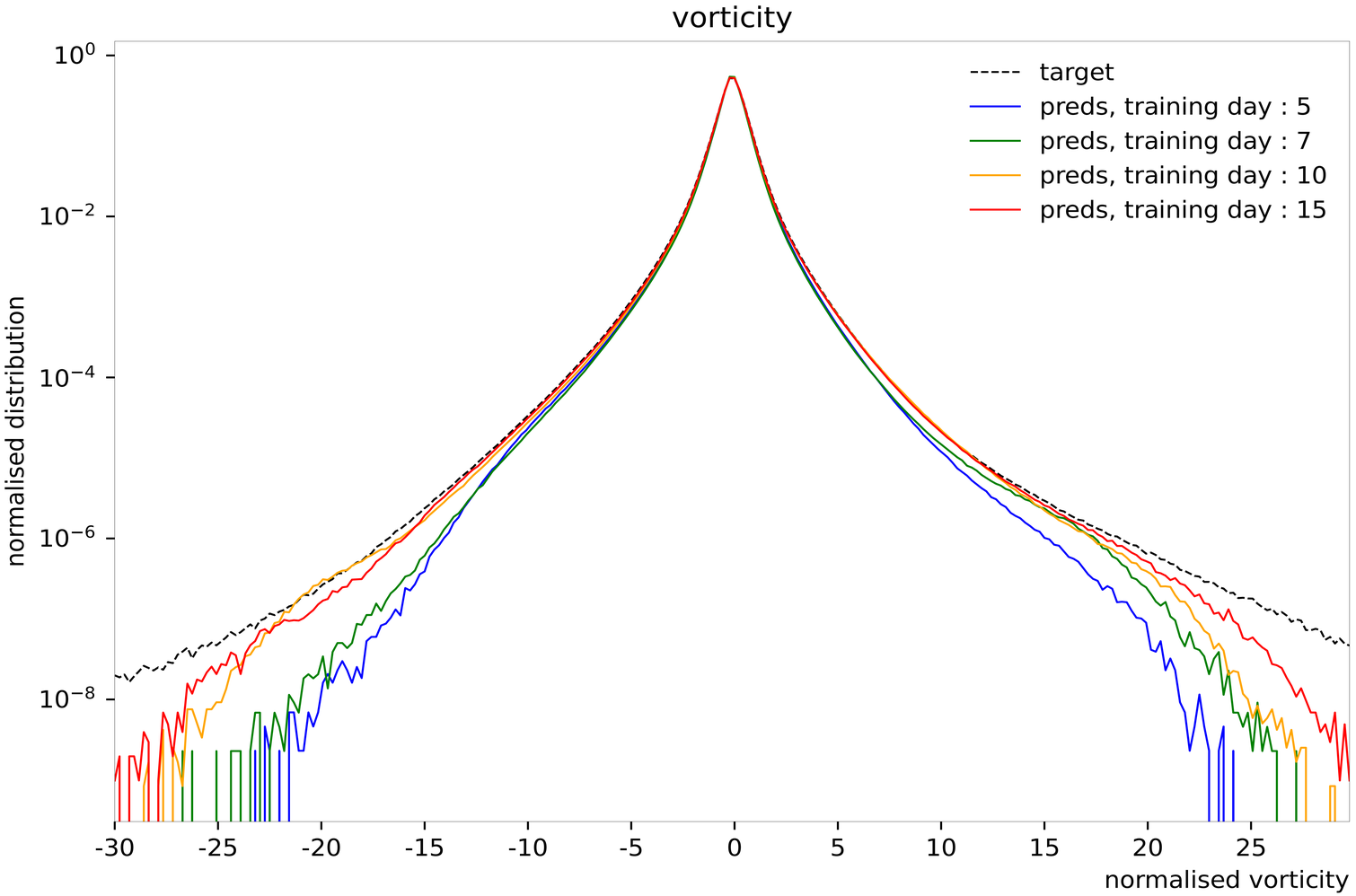

© Christian Lessig, 2023

# Training objective

- BERT-style masked token model
  - Mixture of masking, noise-pertubation and coarsening to learn robust, probabilistic representation
  - Masking ratio is sampled up to, e.g., 0.5, 0.75

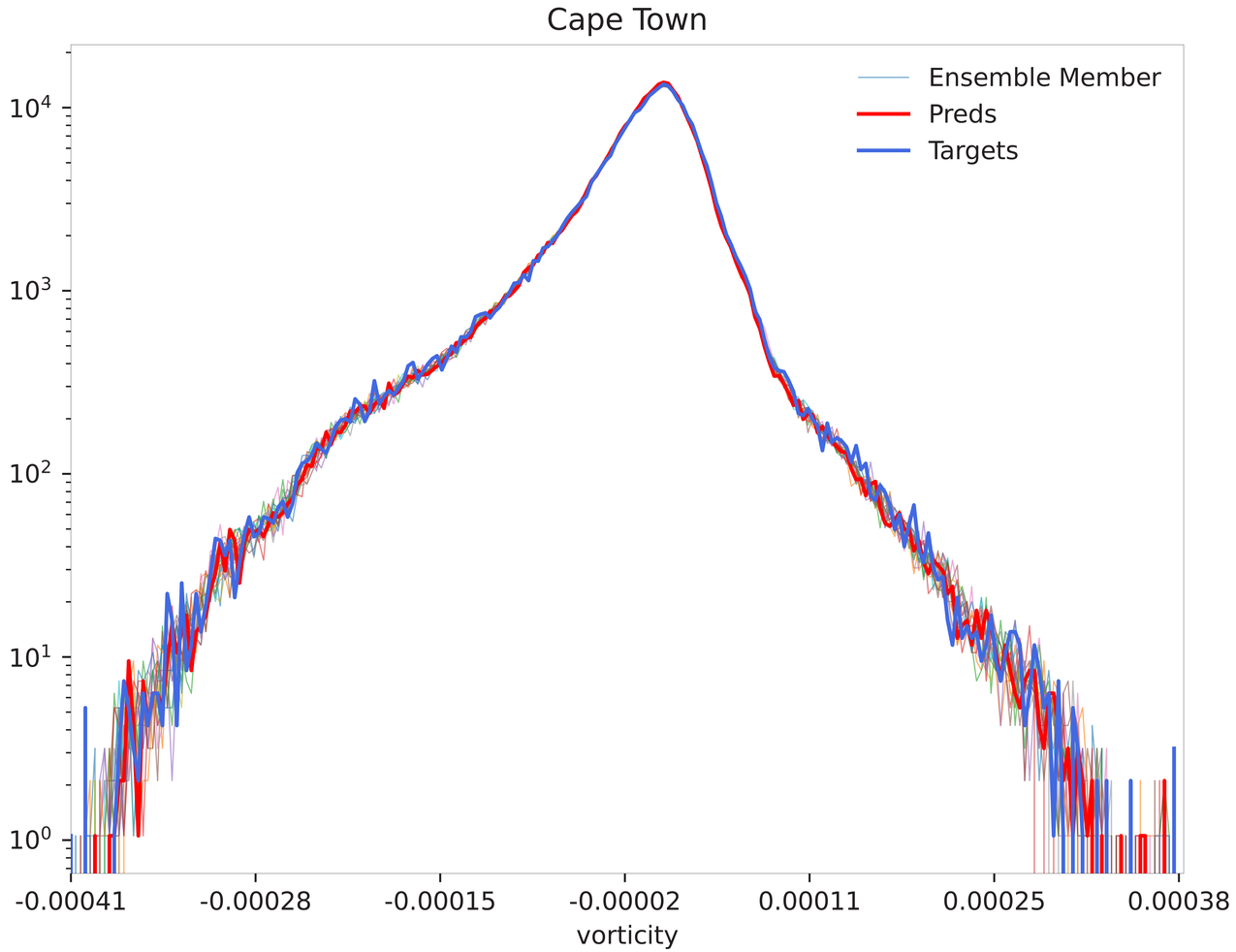
© Christian Lessig, 2023

# Pre-training results



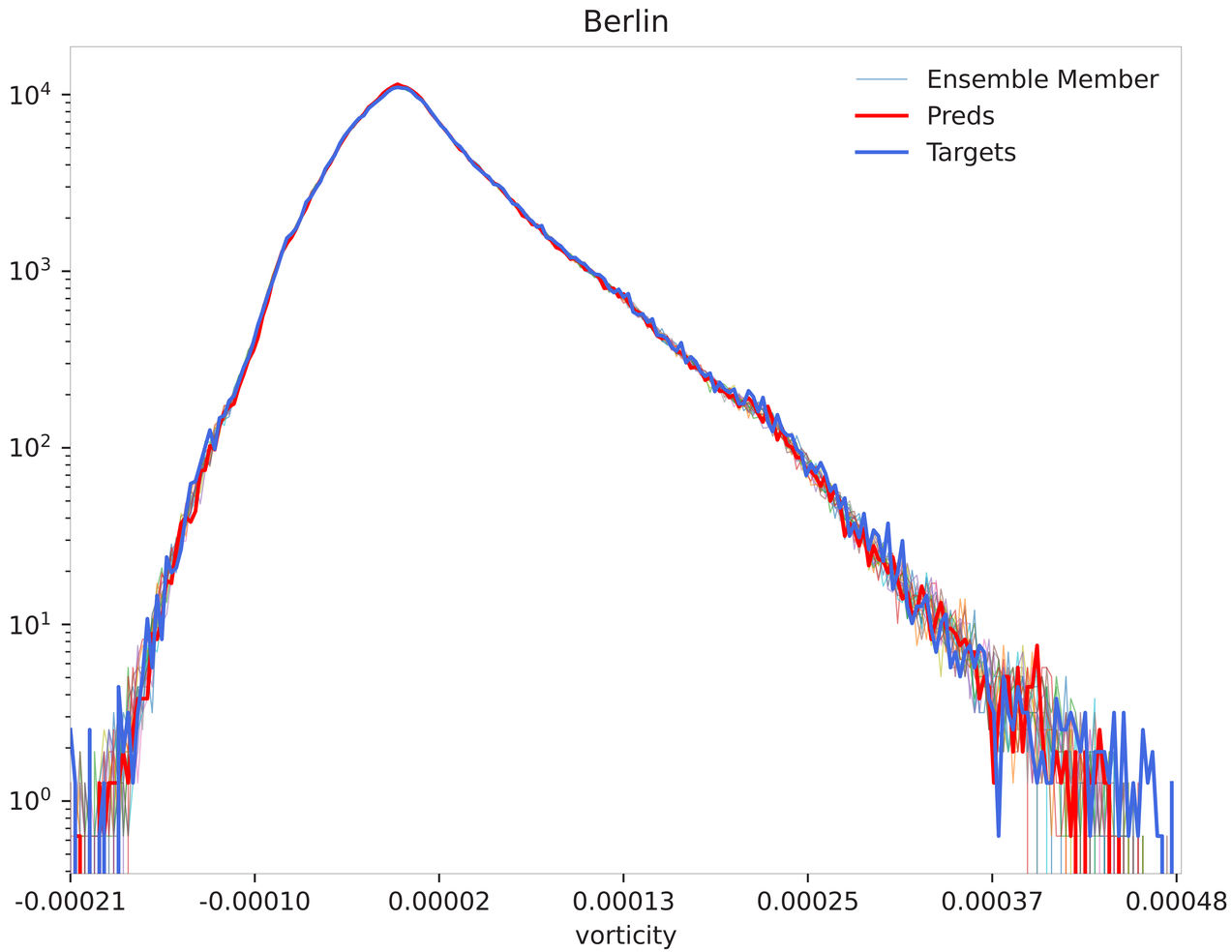

© Christian Lessig, 2023

# Pre-training results




© Christian Lessig, 2023

# Pre-training results




© Christian Lessig, 2023

# Intrinsic Capabilities

© Christian Lessig, 2023

# Zero shot capabilities

- Numerical statistical atmospheric model:

$$p_\theta\left(y|x,\alpha\right)$$

- Model directly includes important applications: forecasting, downscaling, temporal interpolation, ...

© Christian Lessig, 2023

# Zero shot capabilities

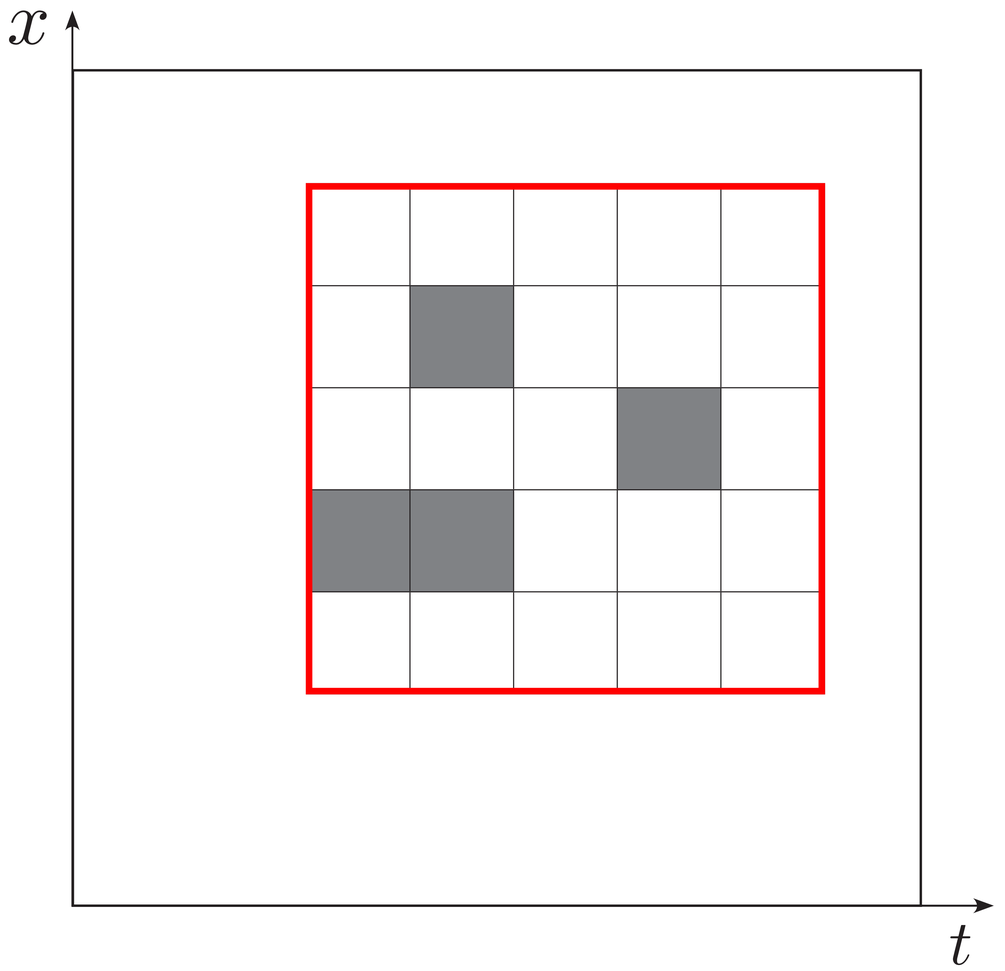
- Numerical statistical atmospheric model:

$$p_\theta\left(y|x,\alpha\right)$$

- Model directly includes important applications: forecasting, downscaling, temporal interpolation, ...
- Training with random masking contains many of these applications as special cases

© Christian Lessig, 2023

# Zero shot capabilities

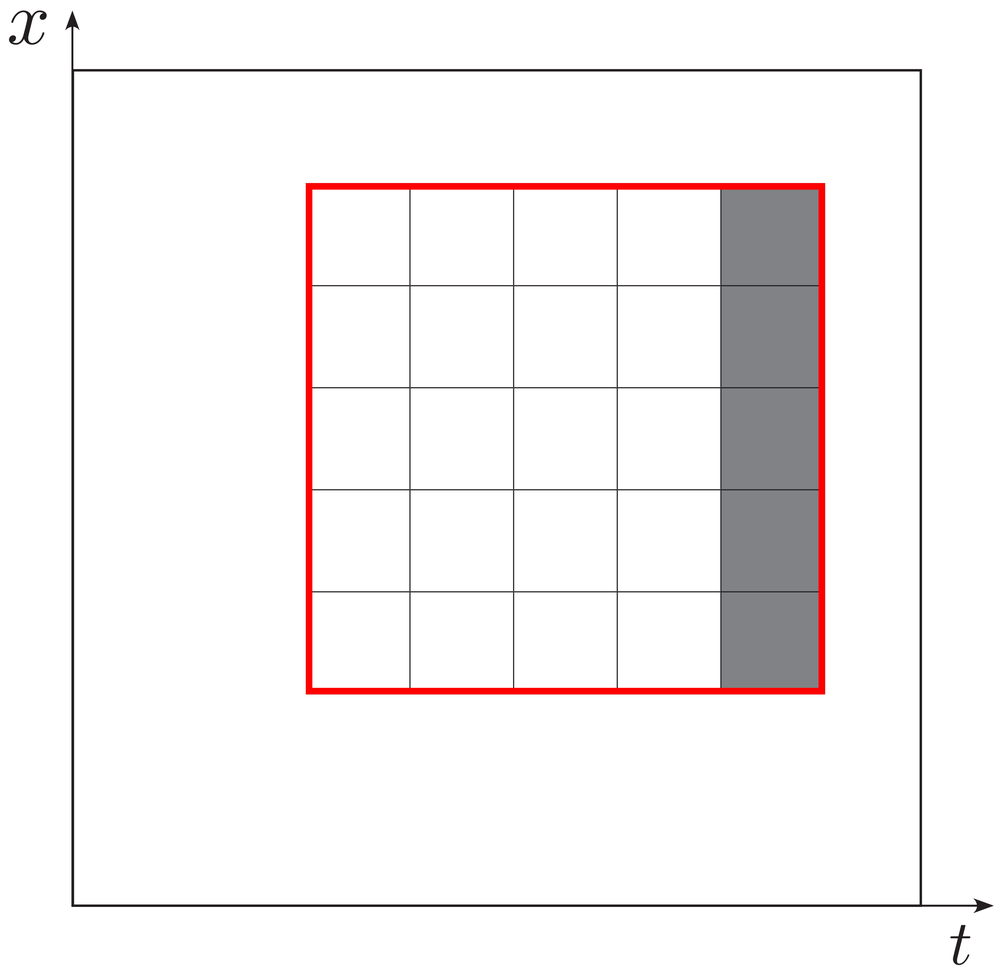



Training task:

predict randomly masked neighbor-hoods in space-time

© Christian Lessig, 2023

# Zero shot capabilities



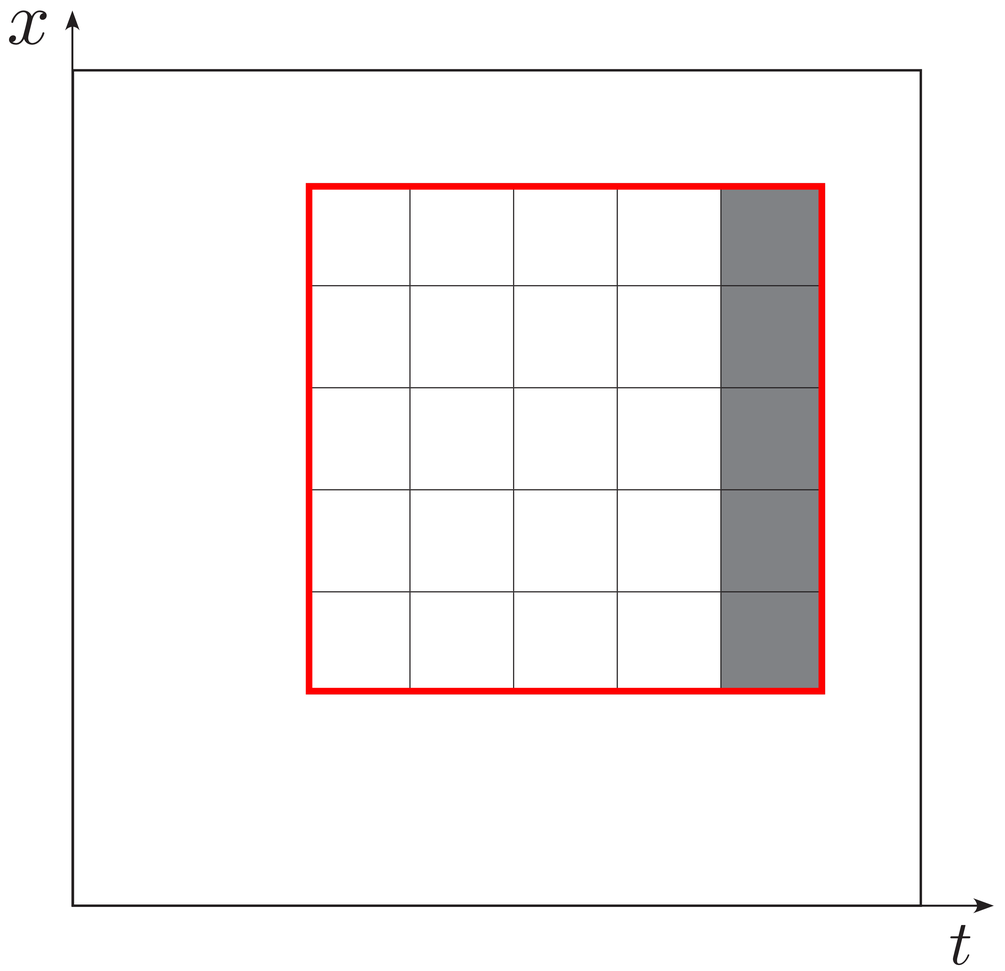

Training task:

predict randomly masked neighbor-hoods in space-time

© Christian Lessig, 2023

# Zero shot capabilities

Forecasting




Training task:

predict randomly masked neighbor-hoods in space-time

© Christian Lessig, 2023

# Zero shot capabilities

## The three settings we explore for in-context learning

**Zero-shot**

The model predicts the answer given only a natural language description of the task. No gradient updates are performed.

```
1  Translate English to French:        ← task description
2  cheese =>                           ← prompt
```

## Traditional fine-tuning (not used for GPT-3)

**Fine-tuning**

The model is trained via repeated gradient updates using a large corpus of example tasks.

```
1  sea otter => loutre de mer          ← example #1
```




T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei. *Language models are few-shot learners*. In Advances in Neural Information Processing Systems, volume 33, pages 1877–1901. Curran Associates, Inc., 2020.

© Christian Lessig, 2023

# Zero shot capabilities

## Large language model: $p_\theta\left(y|x\right)$

### The three settings we explore for in-context learning

**Zero-shot**

The model predicts the answer given only a natural language description of the task. No gradient updates are performed.

```
1  Translate English to French:      ← task description
2  cheese =>                         ← prompt
```

### Traditional fine-tuning (not used for GPT-3)

**Fine-tuning**

The model is trained via repeated gradient updates using a large corpus of example tasks.

```
1  sea otter => loutre de mer        ← example #1
```

↓

**gradient update**

↓

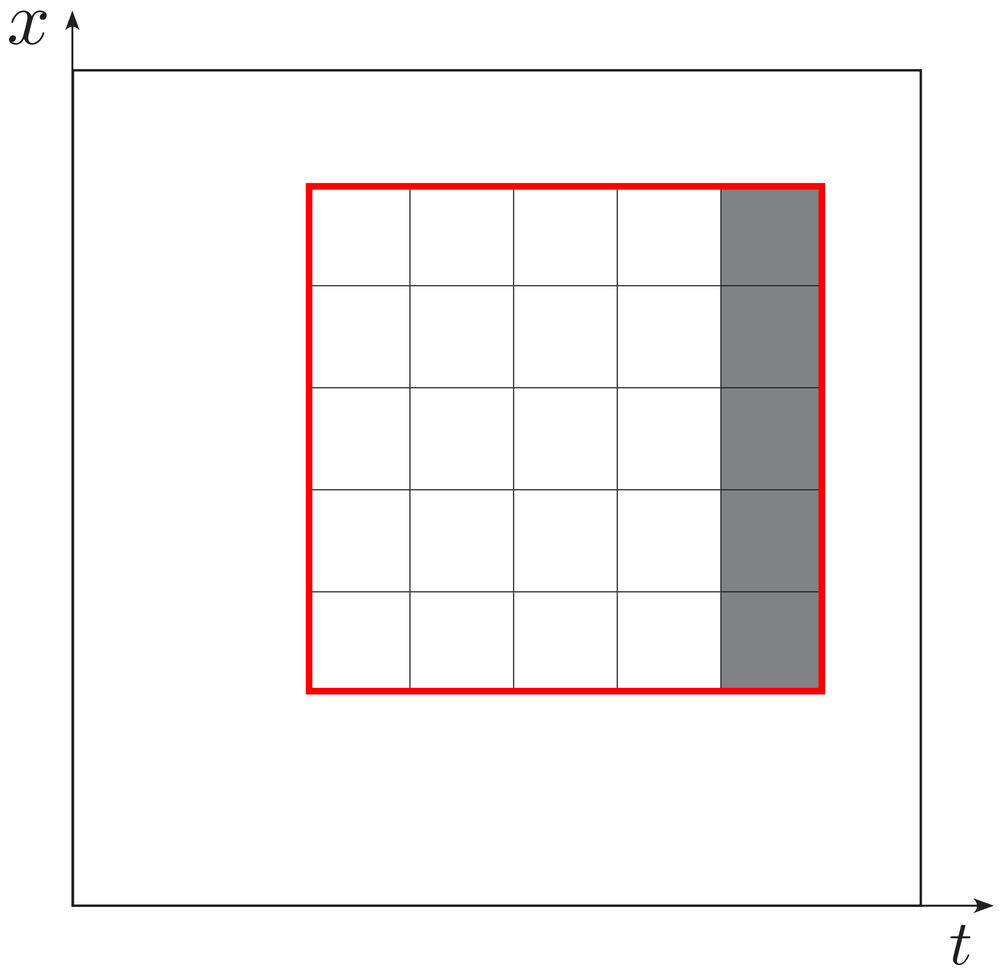
---

T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei. *Language models are few-shot learners*. In Advances in Neural Information Processing Systems, volume 33, pages 1877–1901. Curran Associates, Inc., 2020.

© Christian Lessig, 2023

# Zero shot capabilities

Forecasting

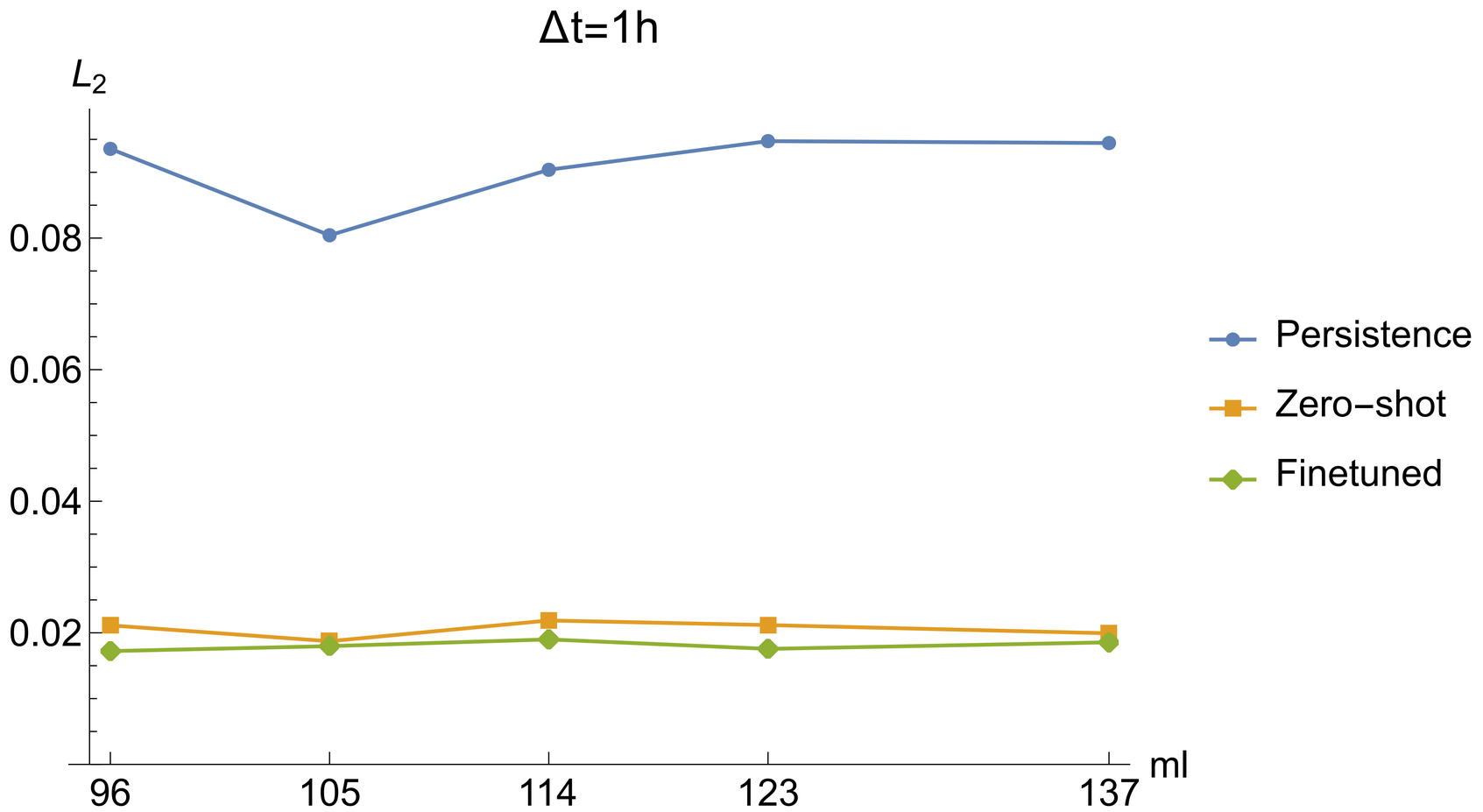



Training task:

predict randomly masked neighbor-hoods in space-time

© Christian Lessig, 2023

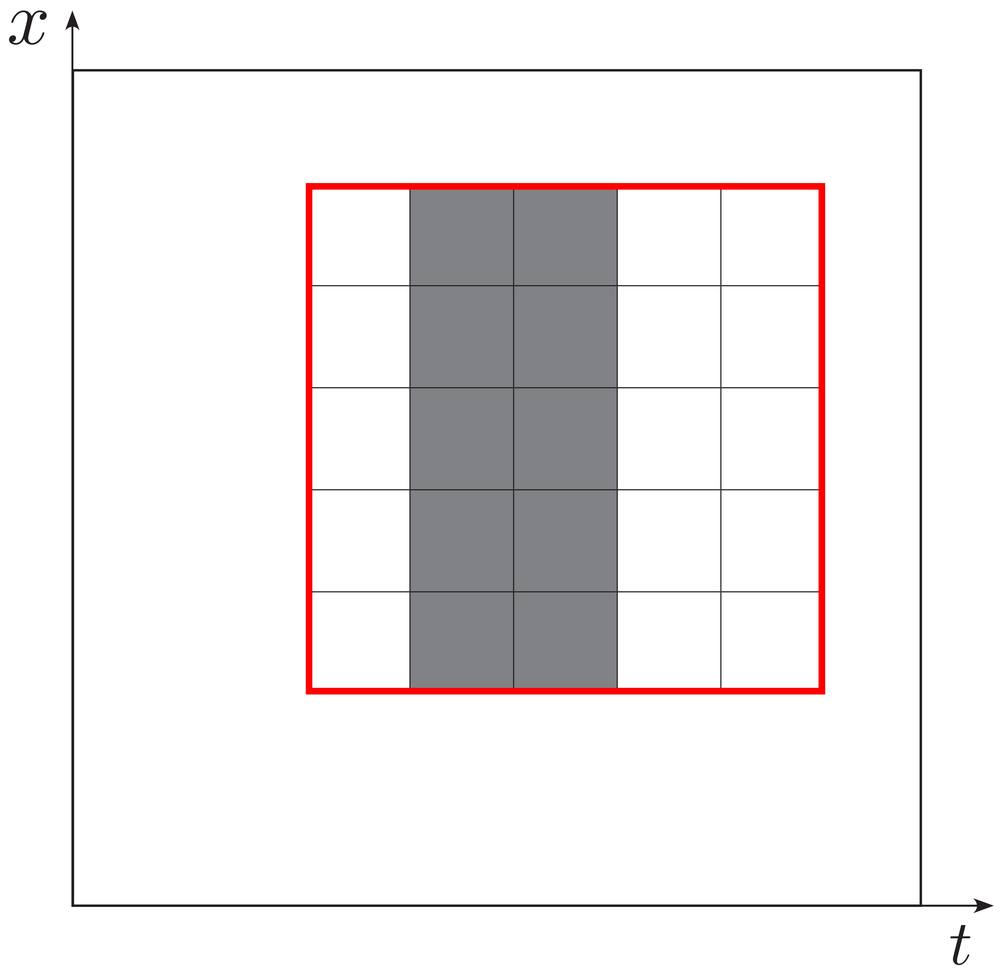
# Zero shot capabilities




© Christian Lessig, 2023

# Zero shot capabilities

Temporal interpolation

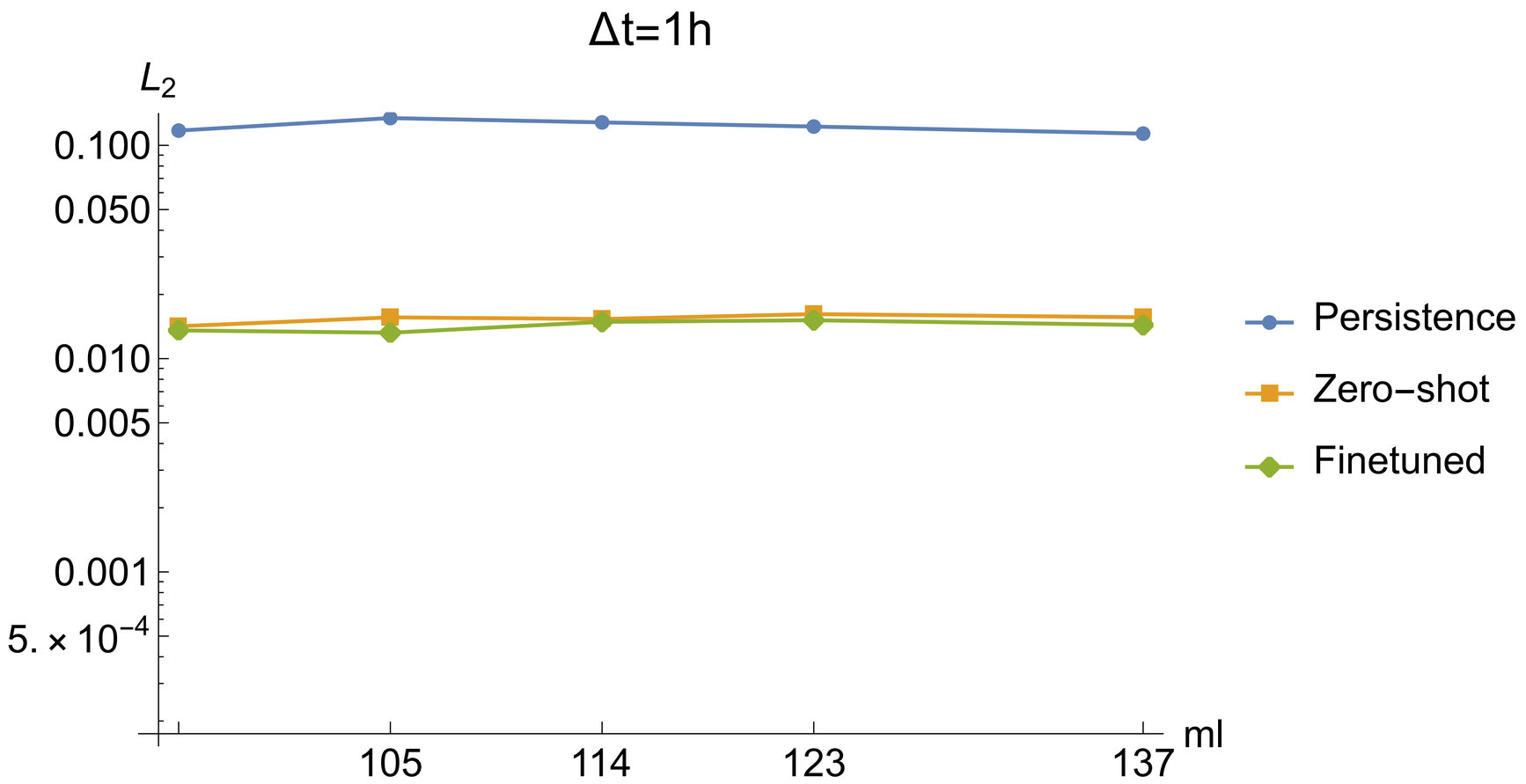



Training task:

predict randomly masked neighbor-hoods in space-time

© Christian Lessig, 2023

# Zero shot capabilities

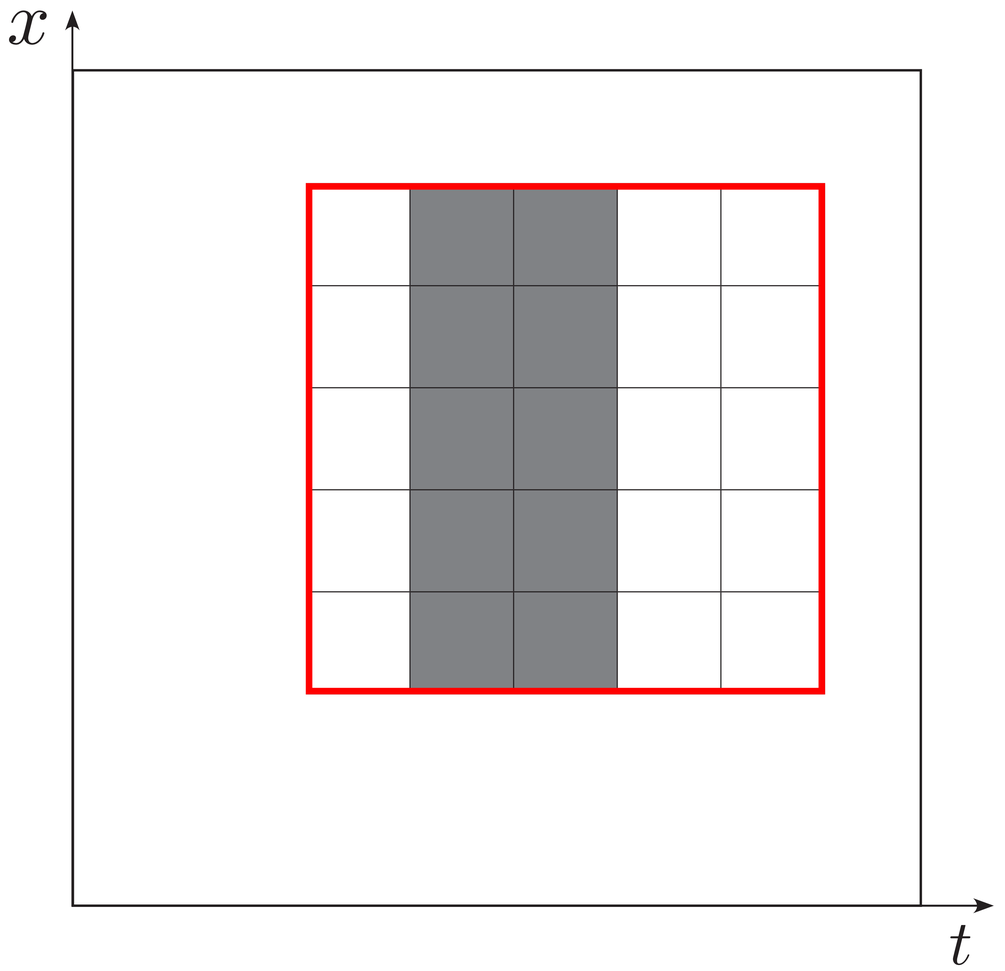



© Christian Lessig, 2023

# Zero shot capabilities

Also:

- spatial interpolation (missing data)
- downscaling
- ...




Training task:

predict randomly masked neighborhoods in space-time

© Christian Lessig, 2023

# Model correction

- Numerical statistical atmospheric model:

$$p_\theta\left(y|x,\alpha\right)$$

© Christian Lessig, 2023

# Model correction

- Numerical statistical atmospheric model:

$$p_\theta\left(y|x,\alpha\right)$$

approximate initial state

© Christian Lessig, 2023

# Model correction

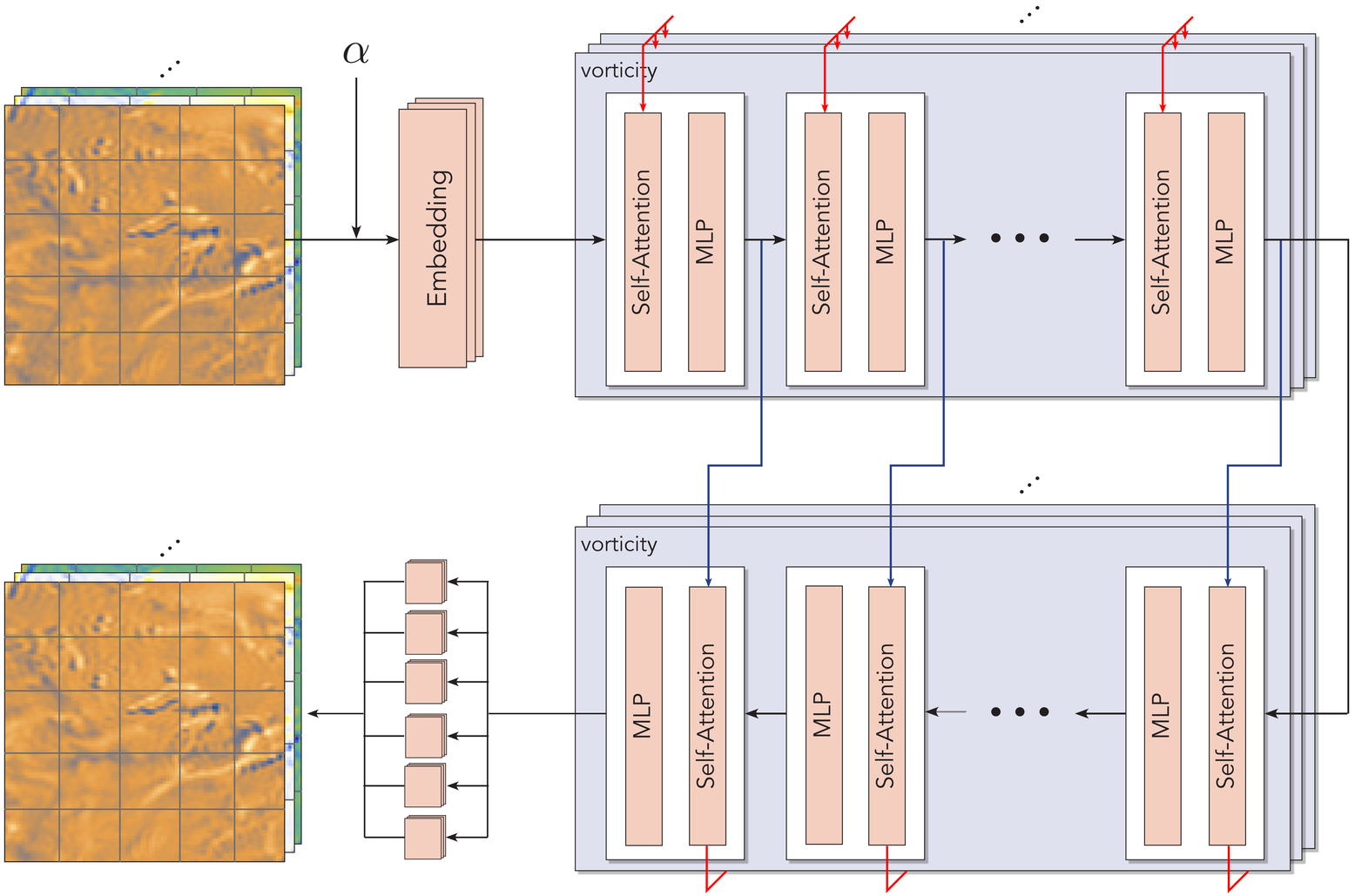
- Numerical statistical atmospheric model:

$$p_\theta\left(y|x,\alpha\right)$$

approximate initial state (training with masking and noise, lower resolution)

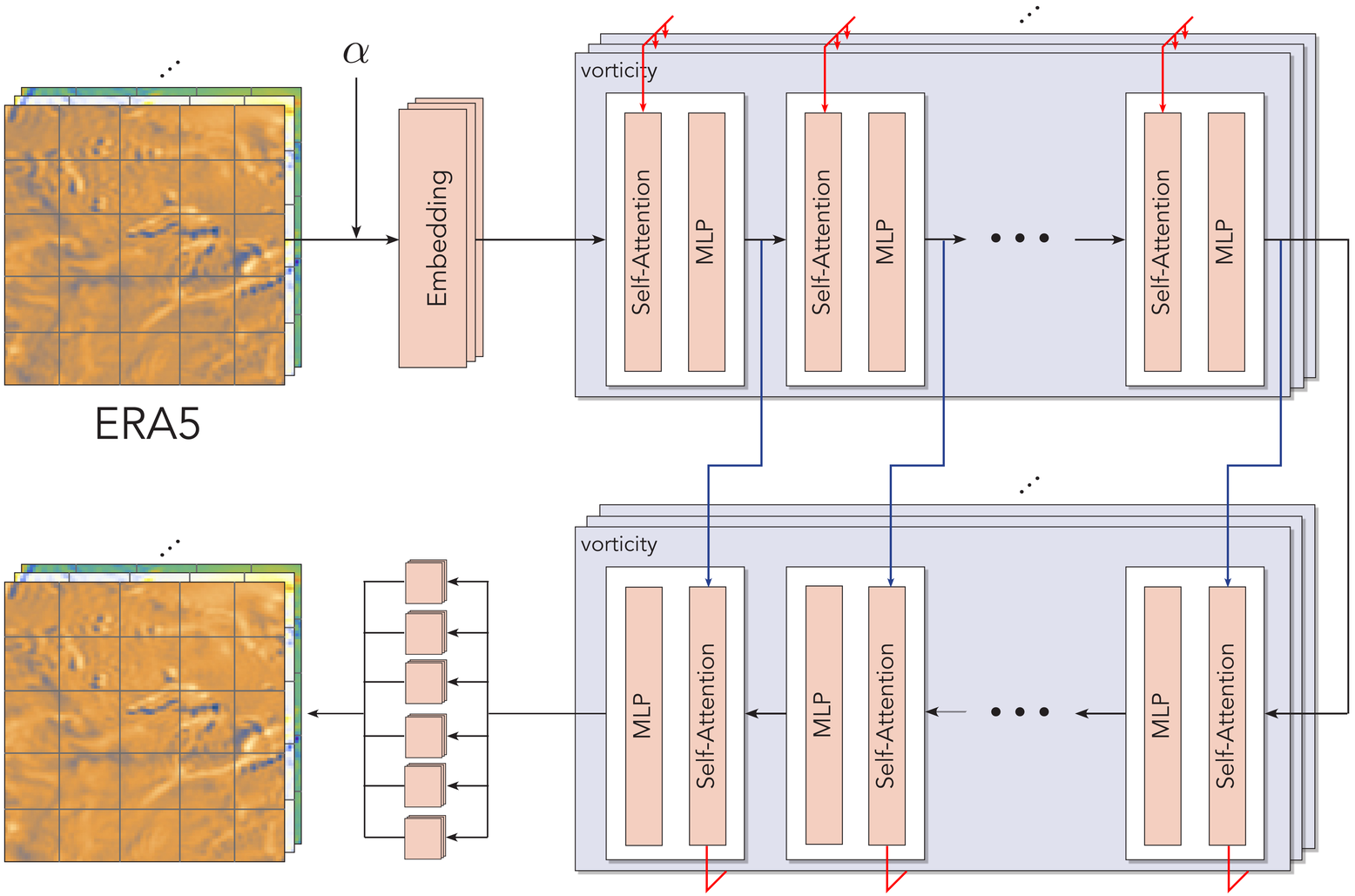
© Christian Lessig, 2023

# Model correction



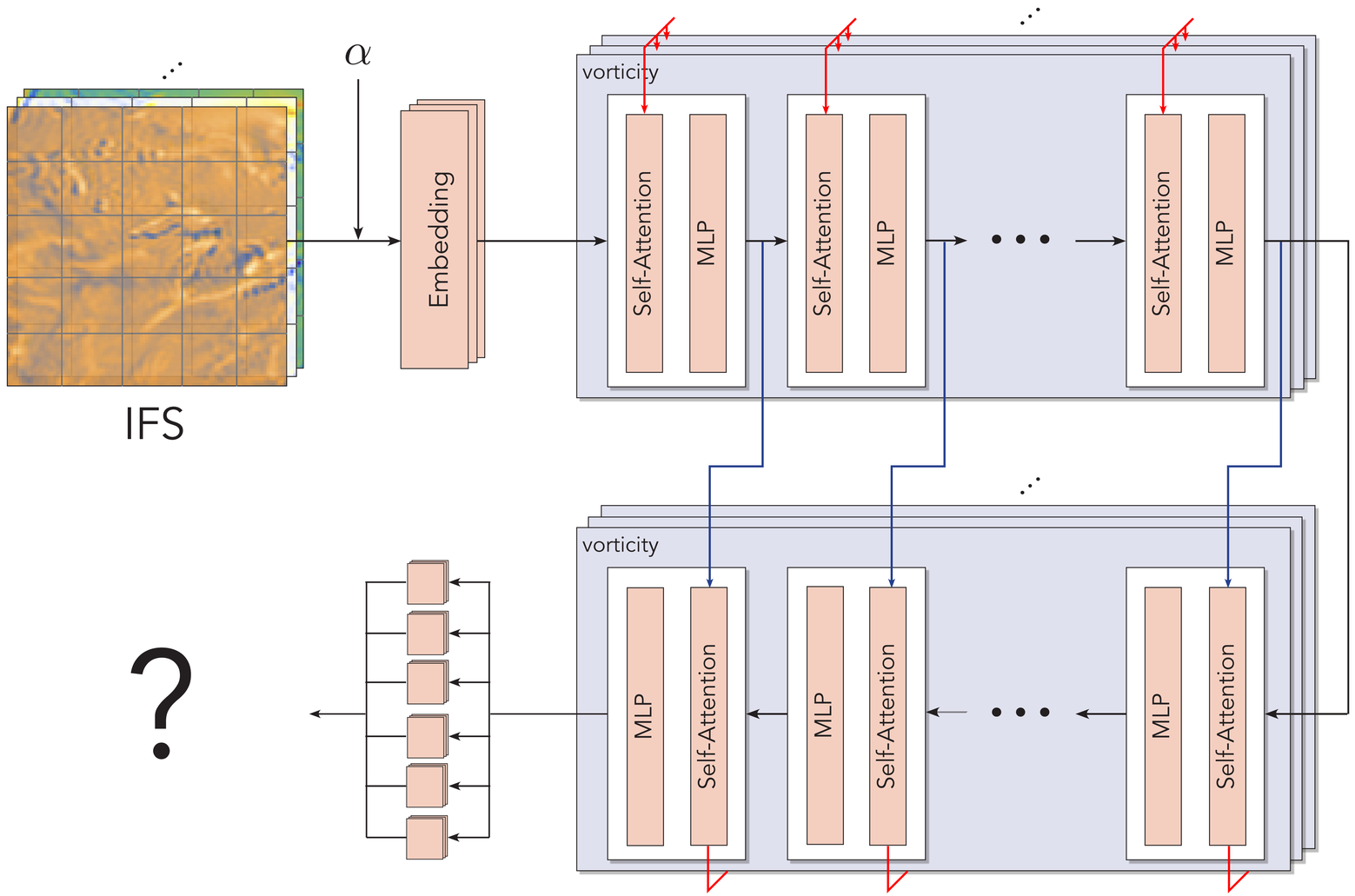

© Christian Lessig, 2023

# Model correction



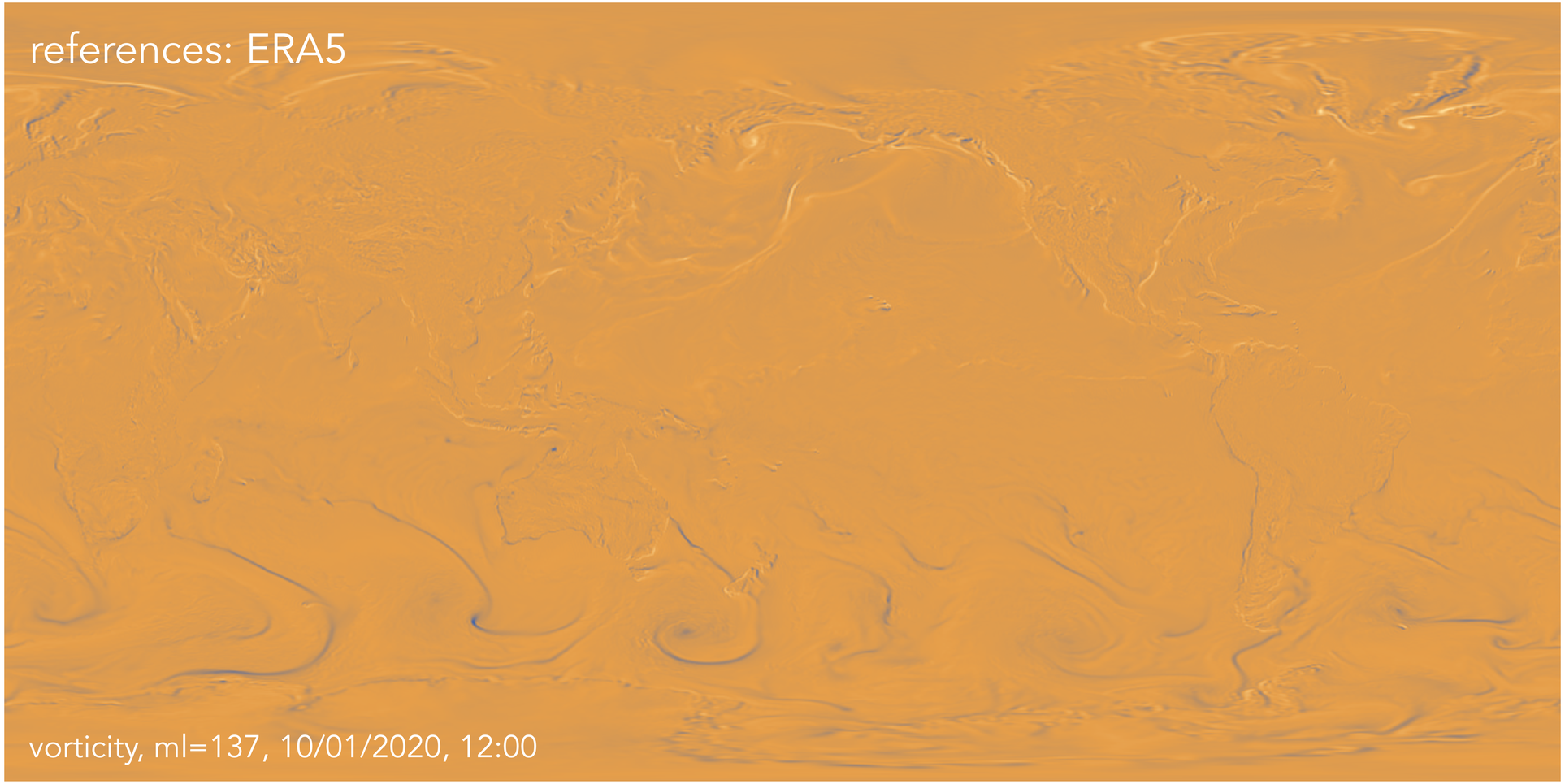

© Christian Lessig, 2023

# Model correction



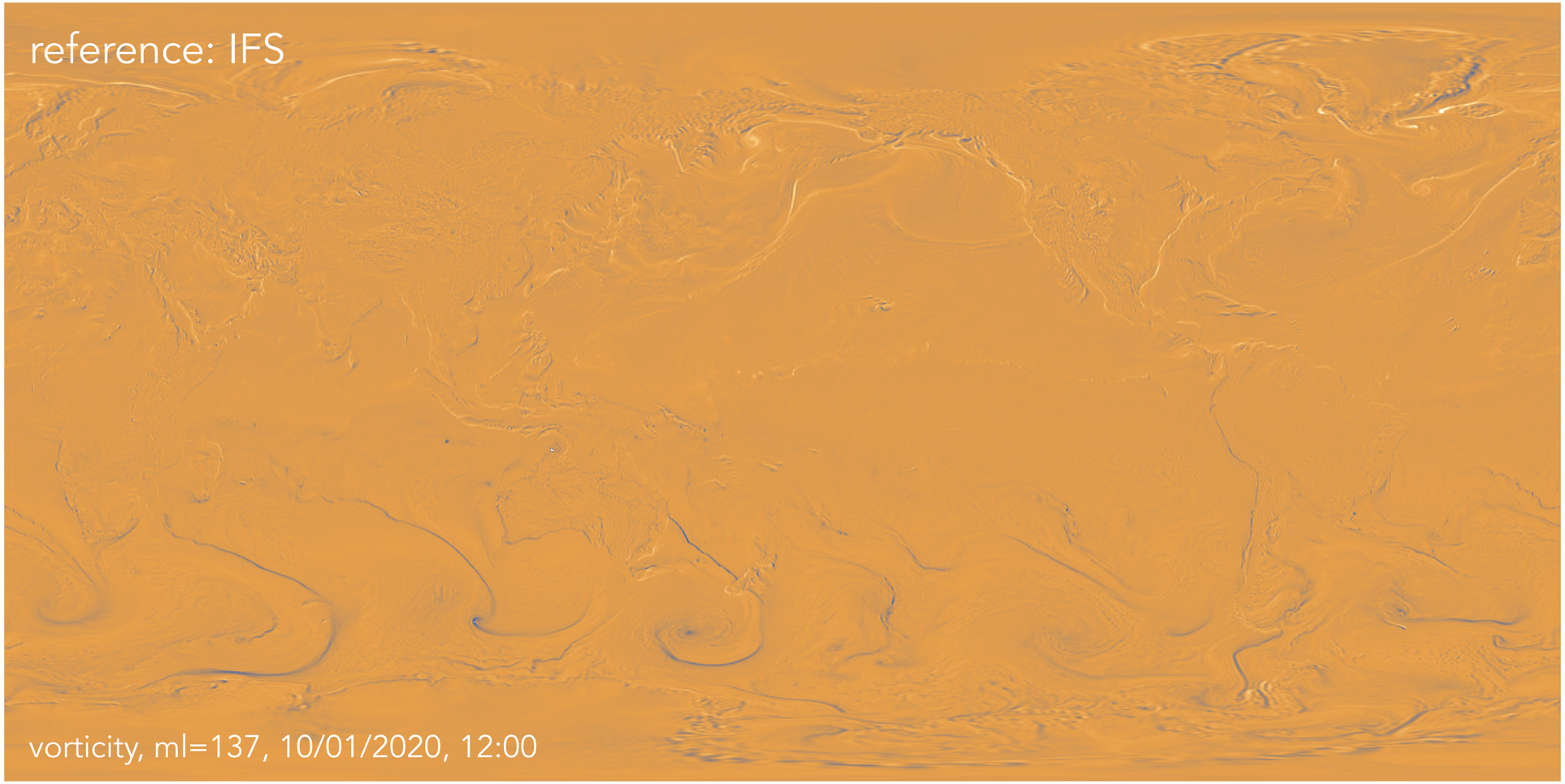

© Christian Lessig, 2023

# Model correction



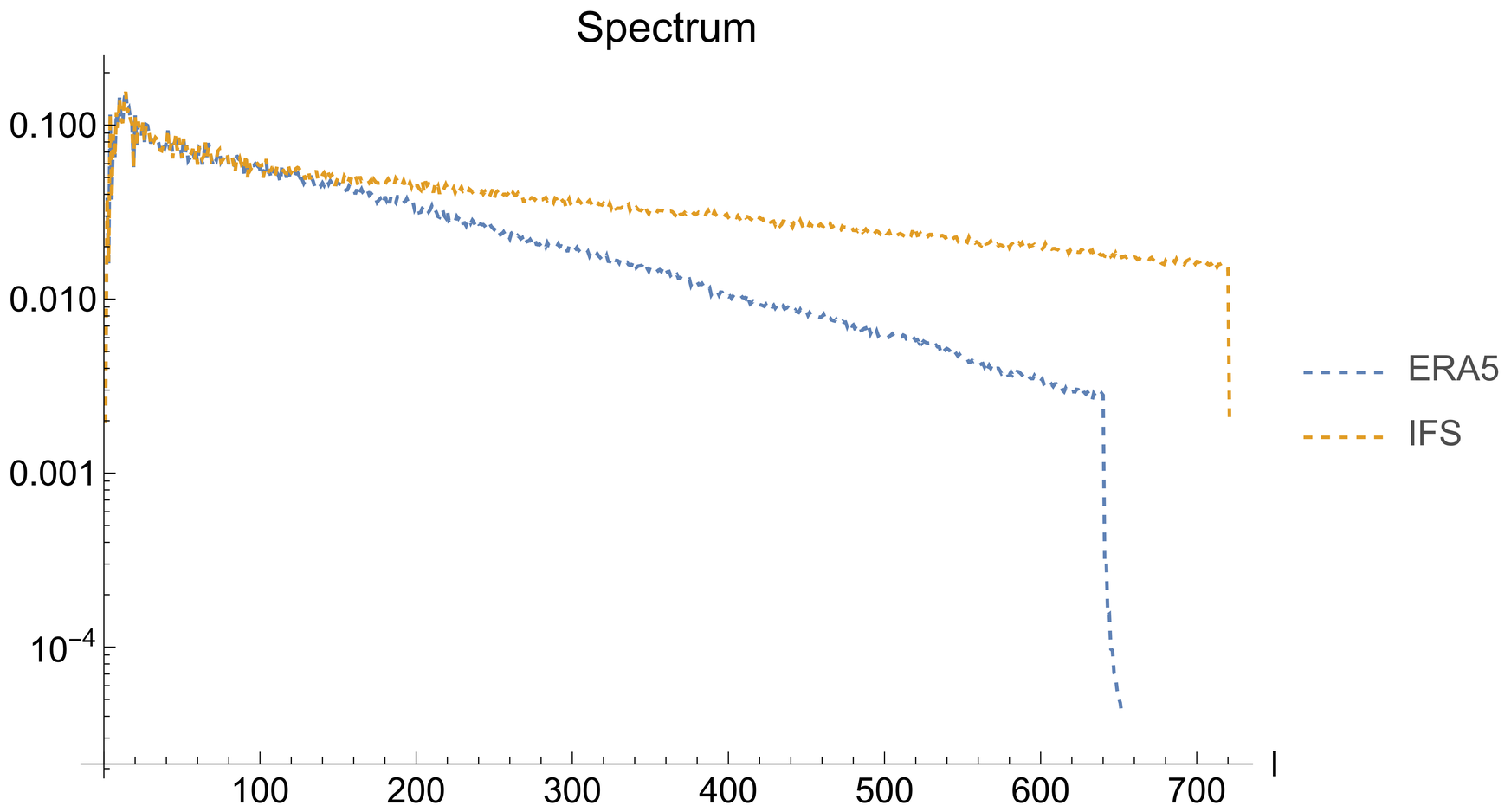

© Christian Lessig, 2023

# Model correction



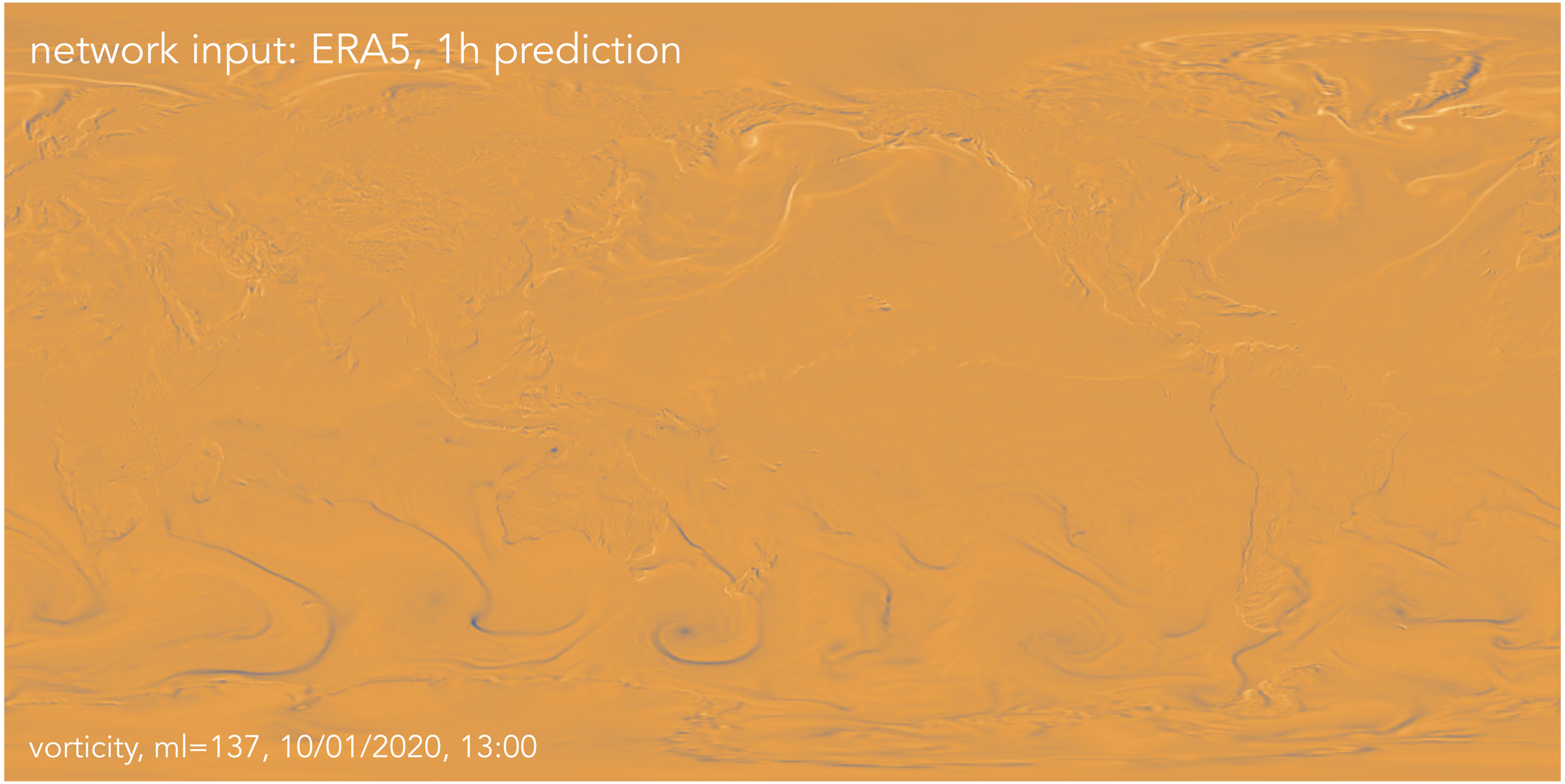

© Christian Lessig, 2023

# Model correction



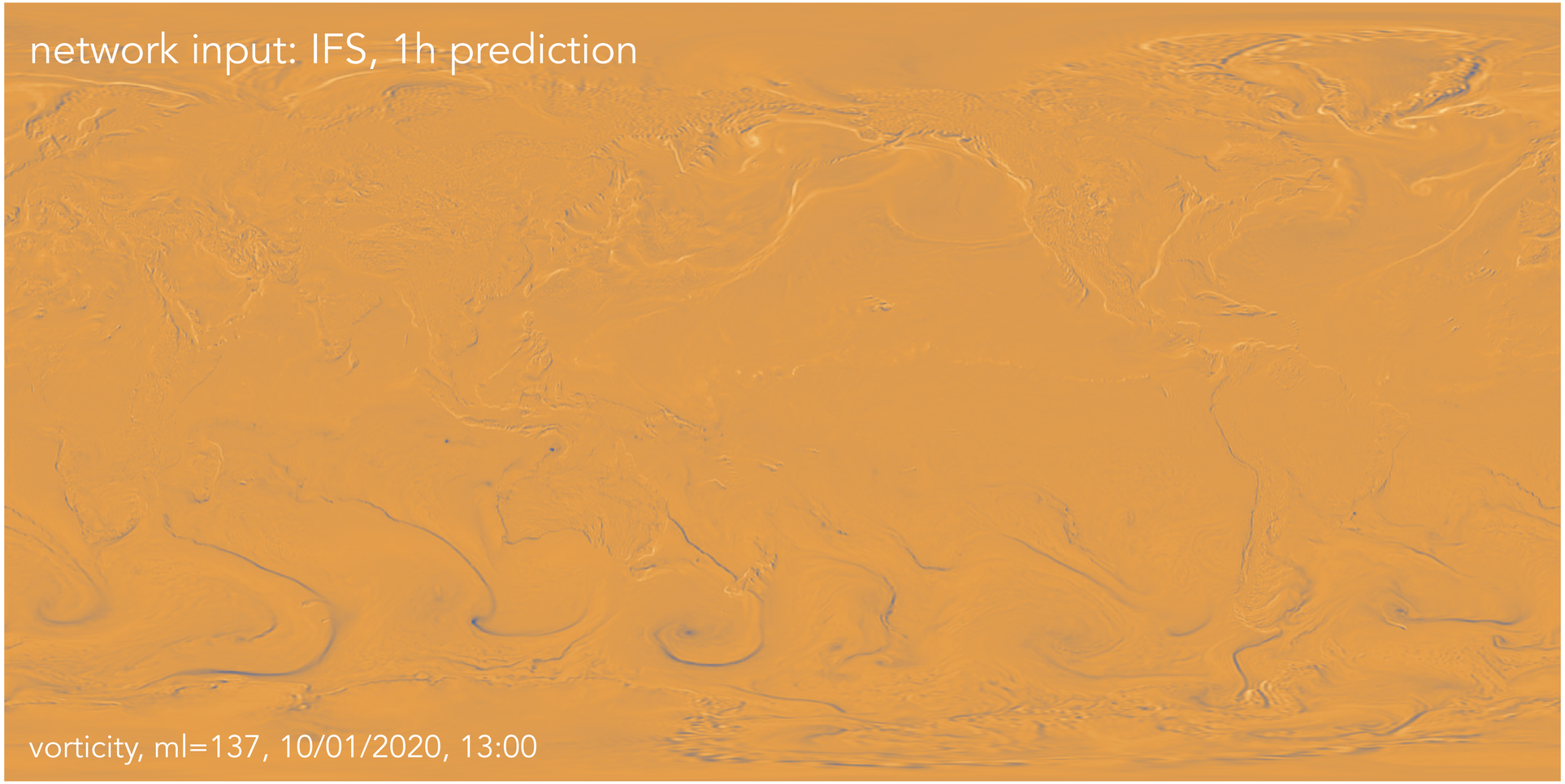

© Christian Lessig, 2023

# Model correction




© Christian Lessig, 2023

# Model correction



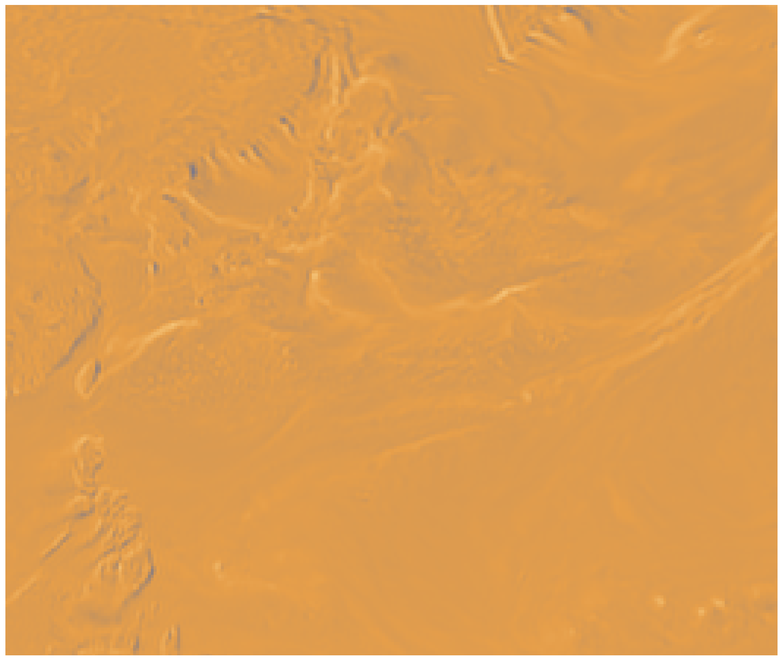

© Christian Lessig, 2023

# Model correction

network input: ERA5

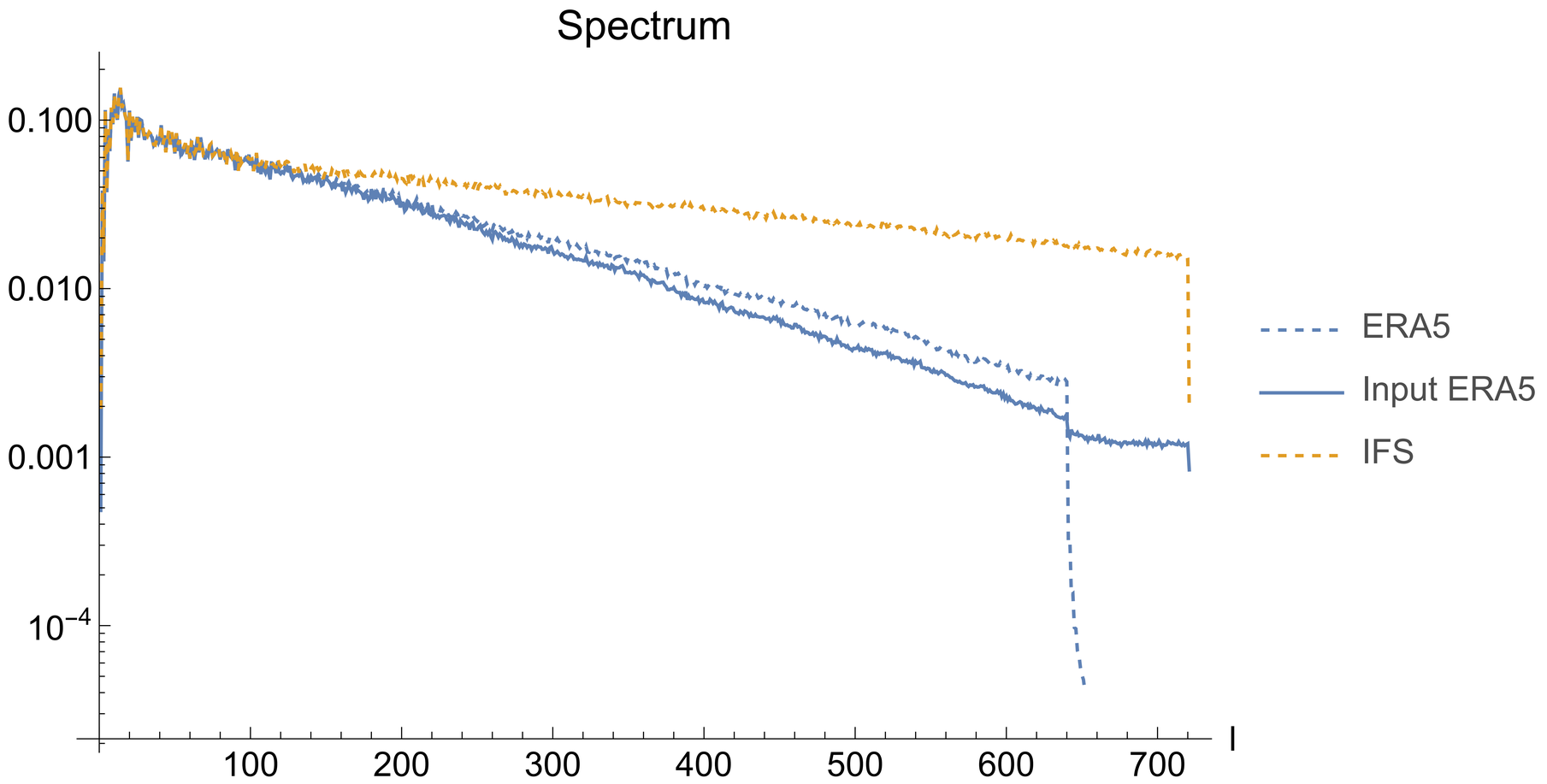


network input: IFS



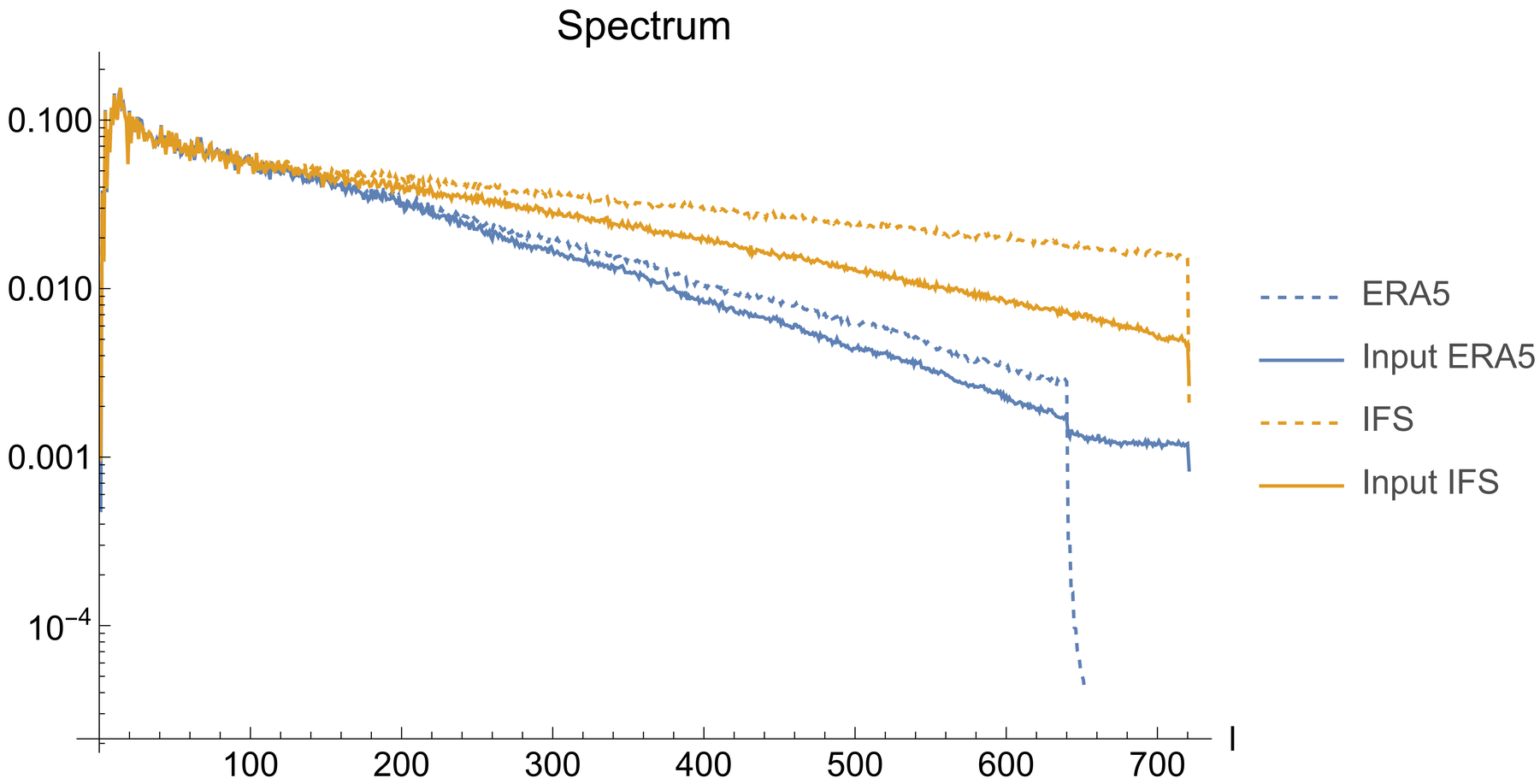
© Christian Lessig, 2023

# Model correction




© Christian Lessig, 2023

# Model correction




© Christian Lessig, 2023

# Counterfactuals

- Numerical statistical atmospheric model:

$$p_\theta\left(y|x,\alpha\right)$$

© Christian Lessig, 2023

# Counterfactuals

- Numerical statistical atmospheric model:

$$p_\theta\left(y|x,\alpha\right)$$

$$\alpha = \left(\mathrm{year}, \mathrm{day}, \mathrm{hour}, \mathrm{ml}, \theta, \phi, \mathrm{res}\right)$$

© Christian Lessig, 2023

# Counterfactuals

- Numerical statistical atmospheric model:

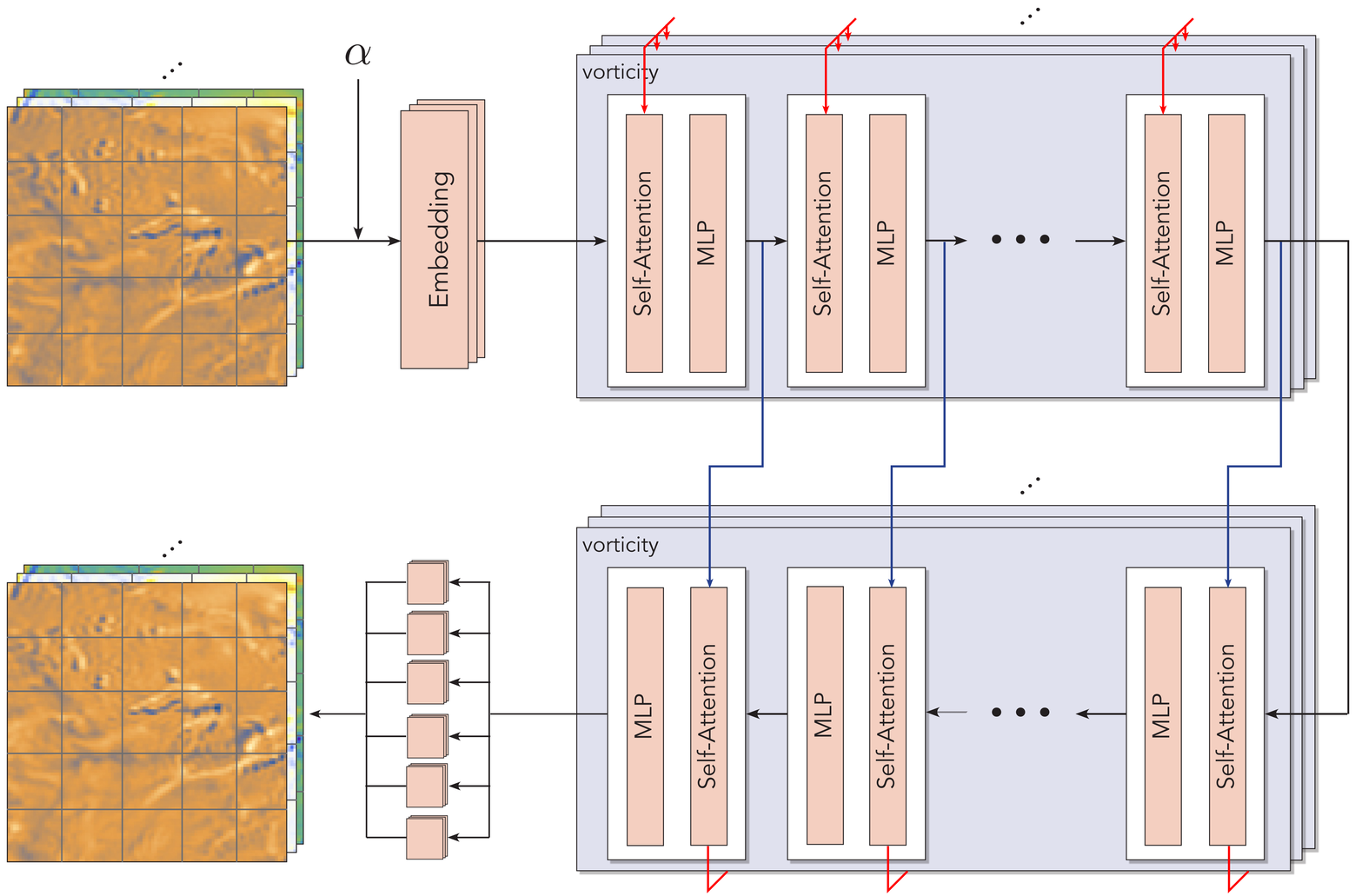
$$p_\theta\left(y|x,\alpha\right)$$

$$\alpha = \left(\mathrm{year}, \mathrm{day}, \mathrm{hour}, \mathrm{ml}, \theta, \phi, \mathrm{res}\right)$$

- Counterfactual: keep initial conditions $x$ but manipulate the auxiliary information $\alpha$

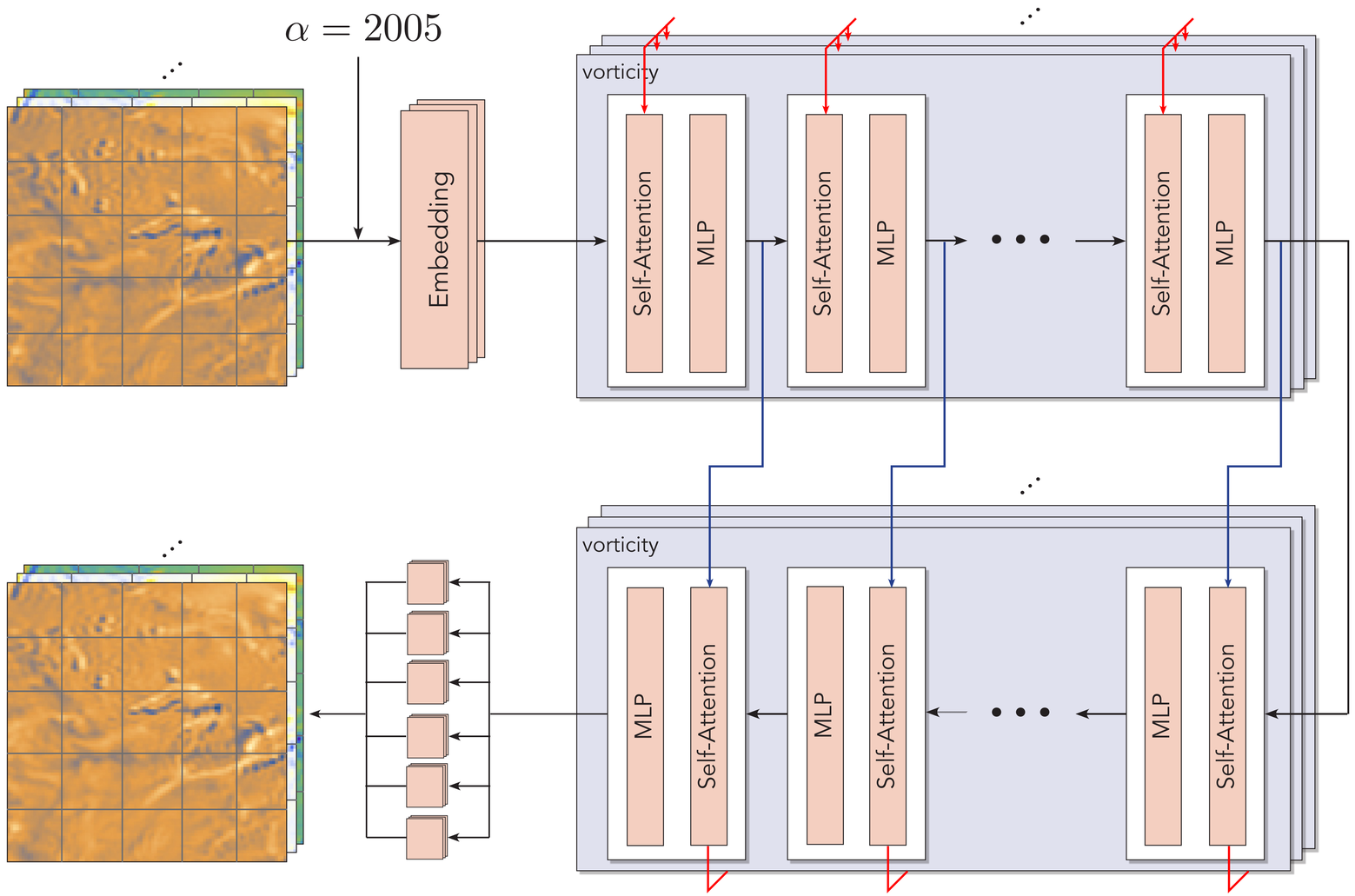
© Christian Lessig, 2023

# Counterfactuals



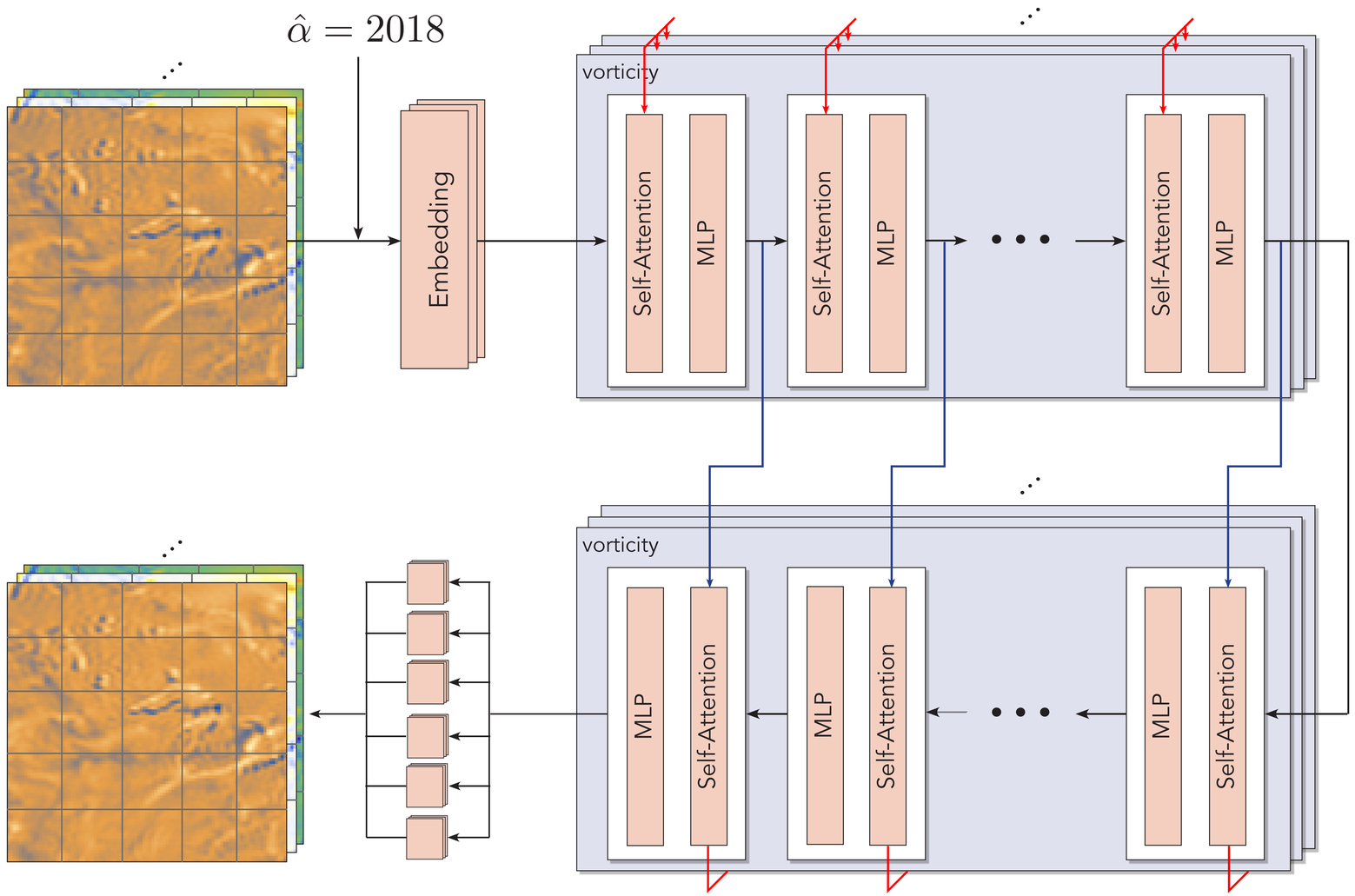

© Christian Lessig, 2023

# Counterfactuals



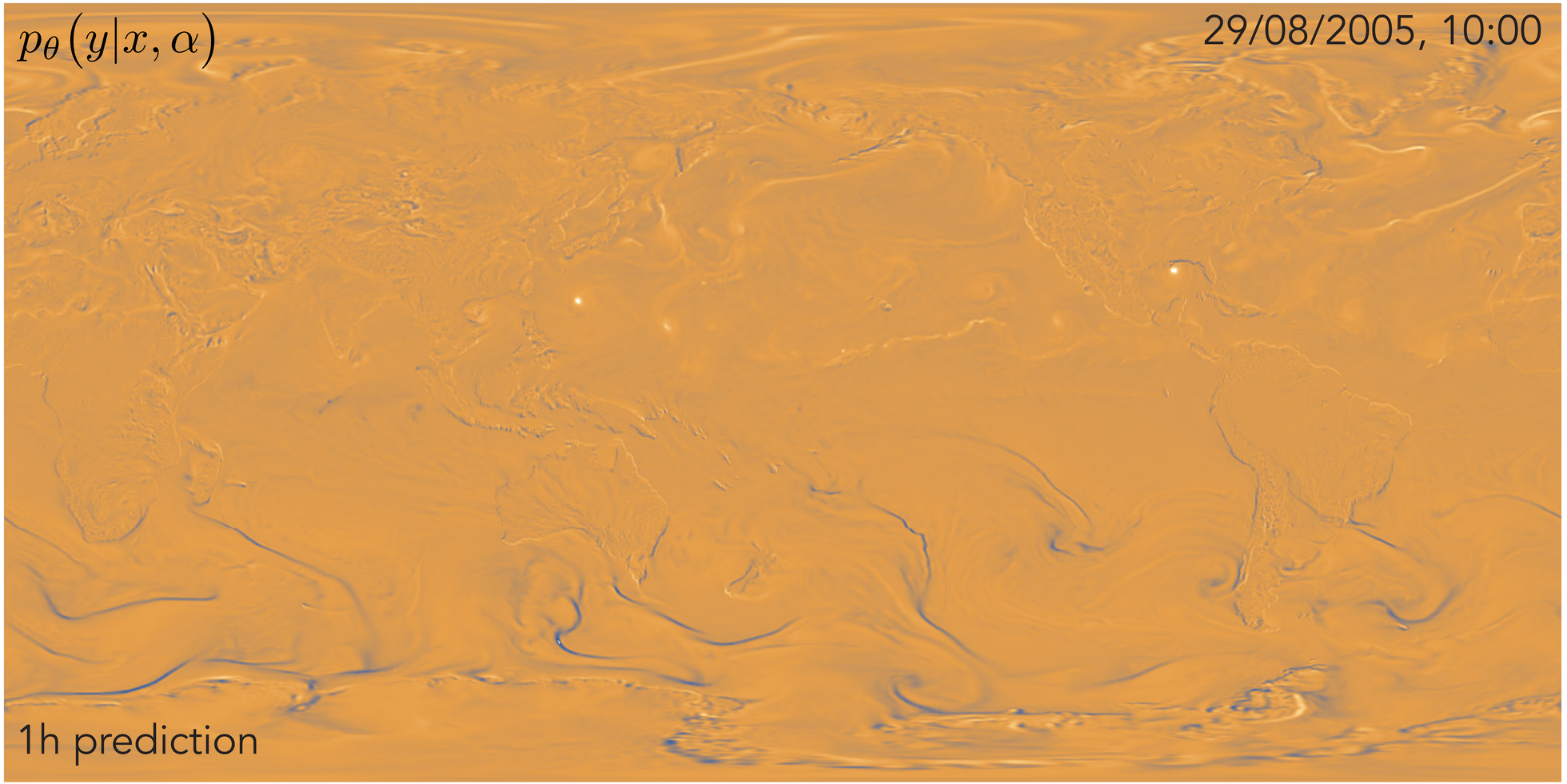

© Christian Lessig, 2023

# Counterfactuals



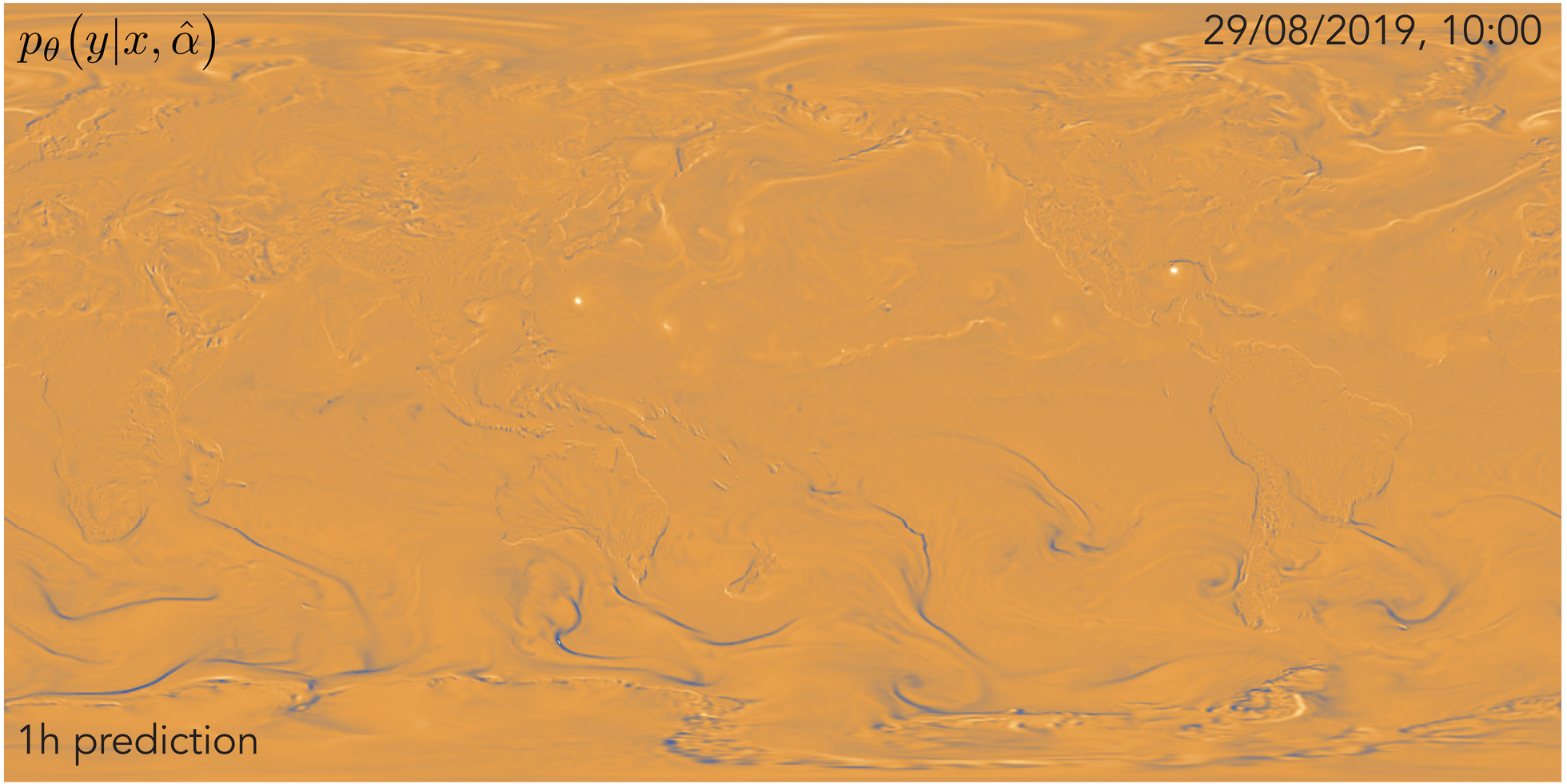

© Christian Lessig, 2023

# Counterfactuals




© Christian Lessig, 2023

# Counterfactuals




© Christian Lessig, 2023

# Counterfactuals




© Christian Lessig, 2023

# Can machine learning replace numerical models for partial differential equations?

# Can machine learning replace numerical PDEs?

Perhaps that is not the question: how can we solve problems with machine learning?

# Can machine learning replace numerical PDEs?

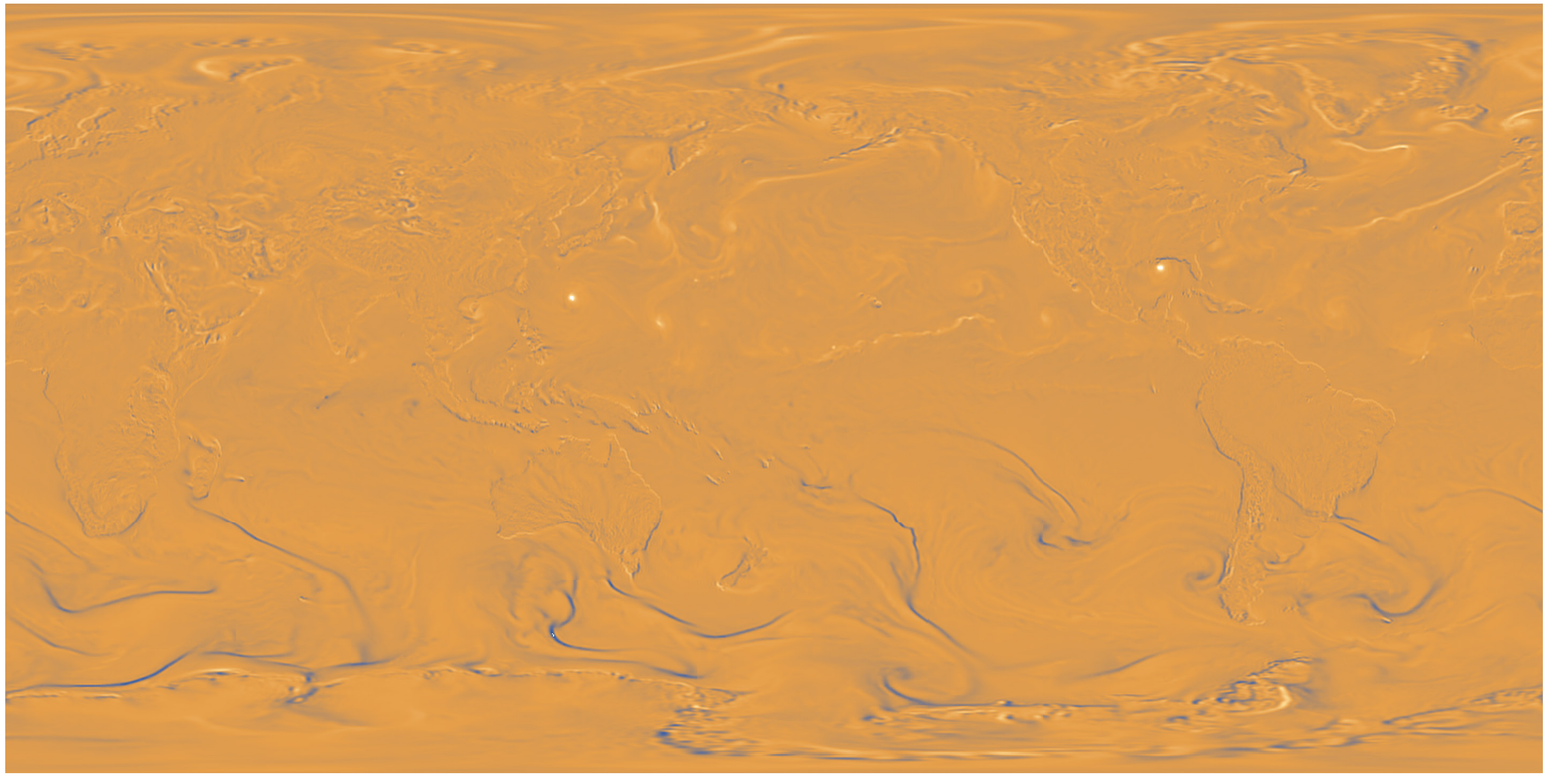
Perhaps that is not the question: how can we solve problems with machine learning?

- Think of machine learning as complementary to existing approaches
  - Consider problems where existing approaches are sub-optimal, e.g. closure problems
- Integration of observational/measurement data
  - Growing exponentially in many domains
- Start machine learning with appropriate mathematical modeling
  - Find problems where machine learning fits rather than vice versa
- Develop domain specific but task independent neural networks

# Questions

# Medium range forecasting

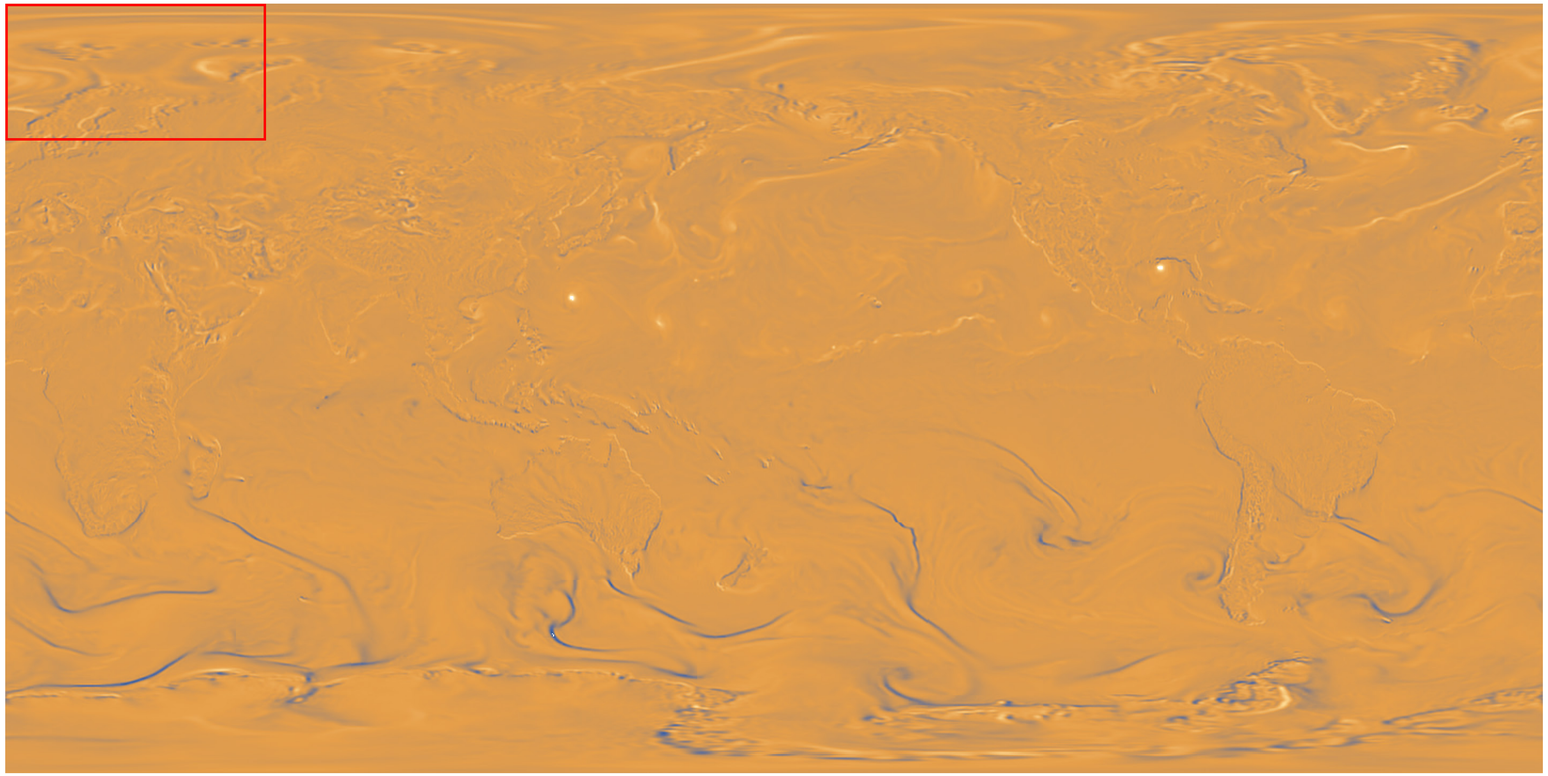
- How to do global forecasts with a local model?



© Christian Lessig, 2023

# Medium range forecasting

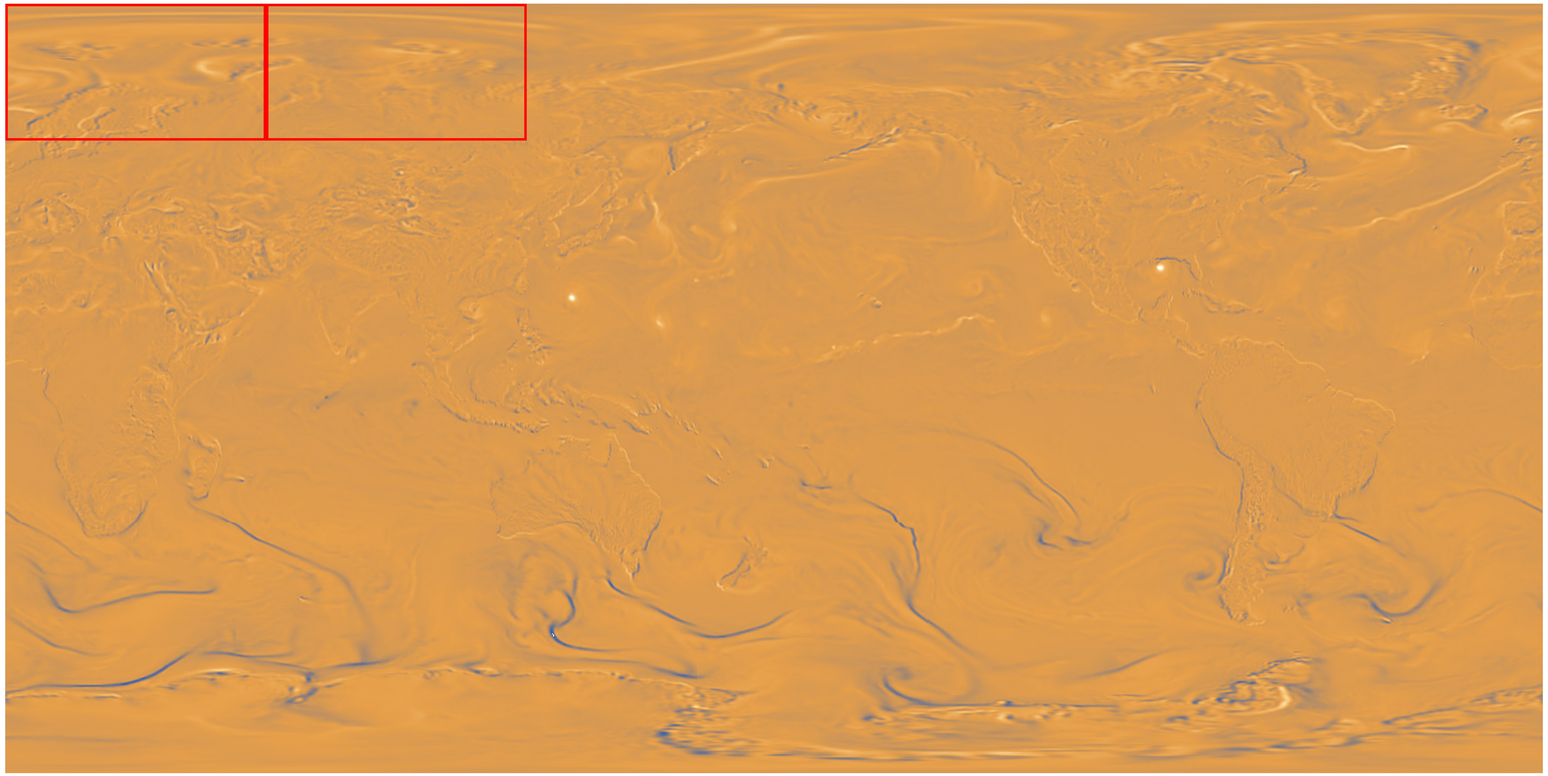
- How to do global forecasts with a local model?



© Christian Lessig, 2023

# Medium range forecasting

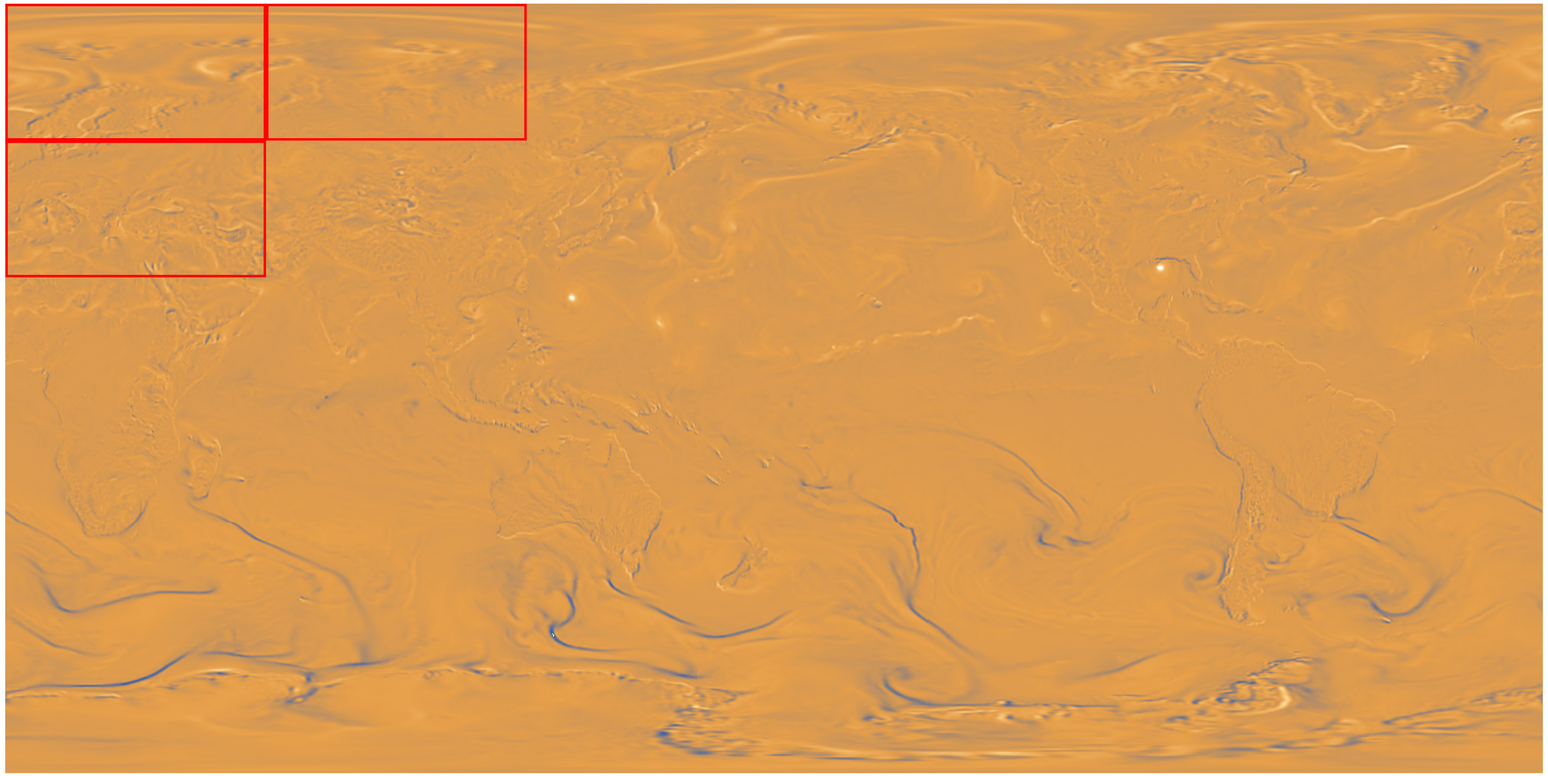
- How to do global forecasts with a local model?



© Christian Lessig, 2023

# Medium range forecasting

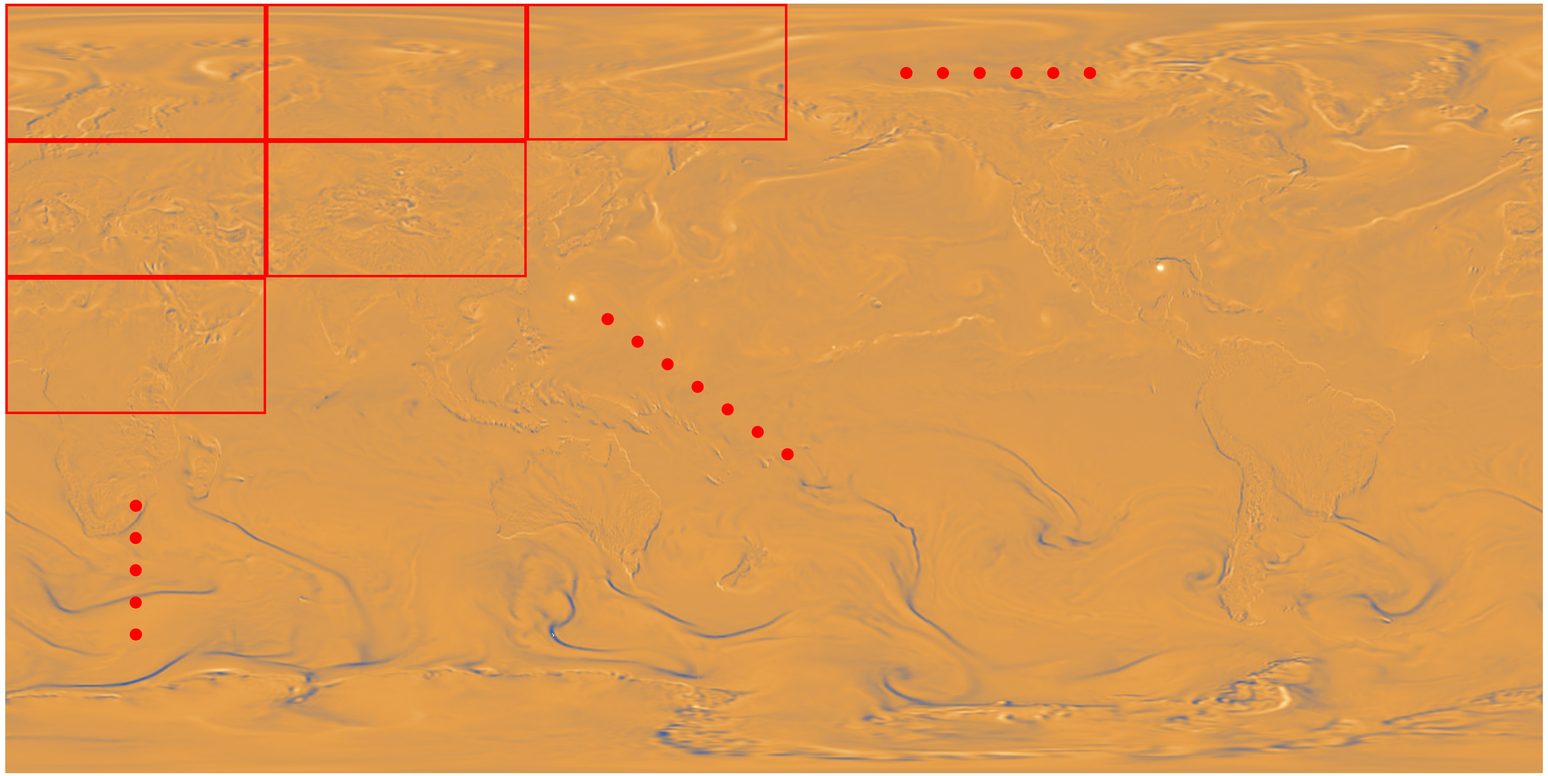
- How to do global forecasts with a local model?



© Christian Lessig, 2023

# Medium range forecasting

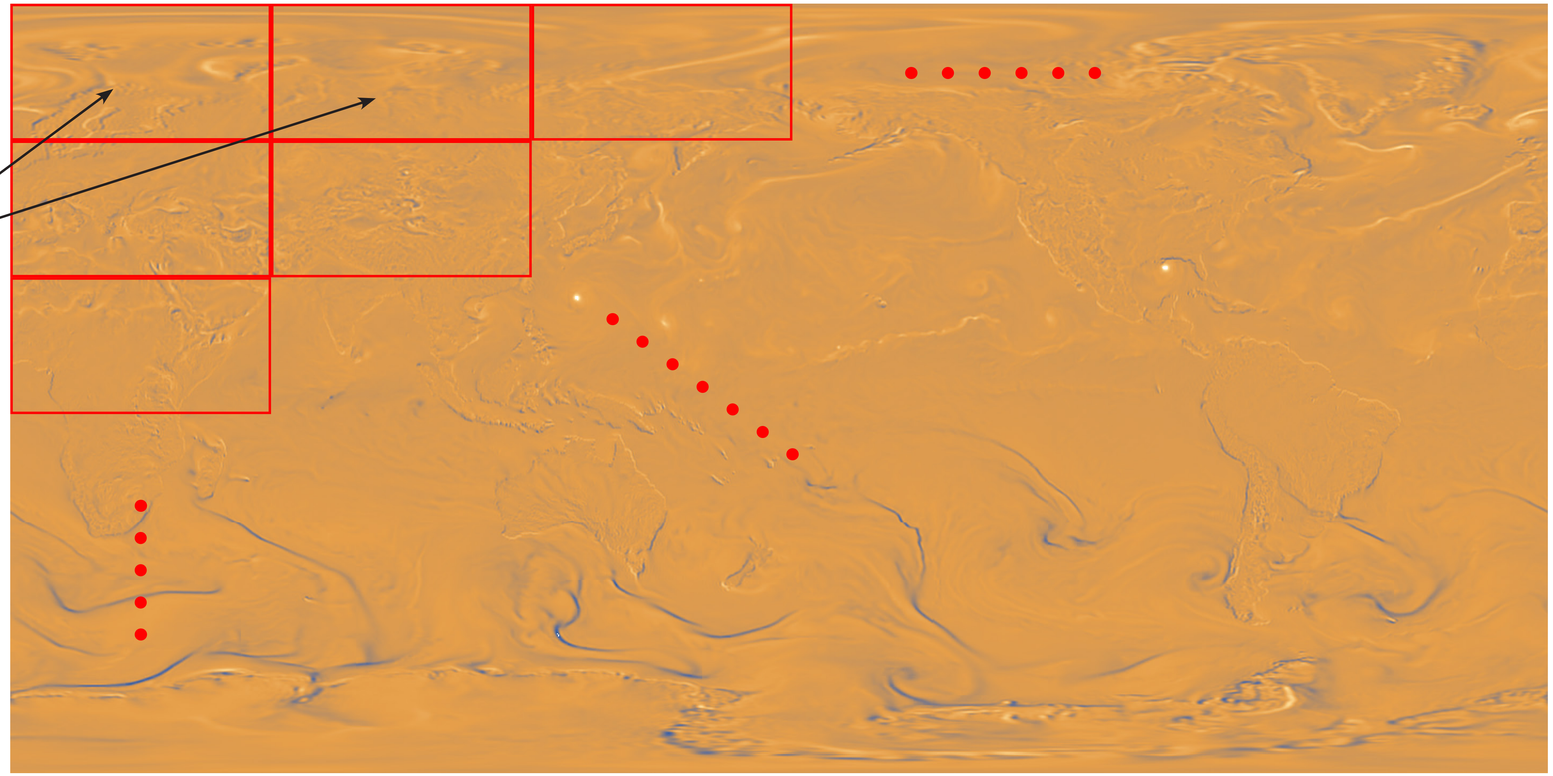
- How to do global forecasts with a local model?



© Christian Lessig, 2023

# Medium range forecasting

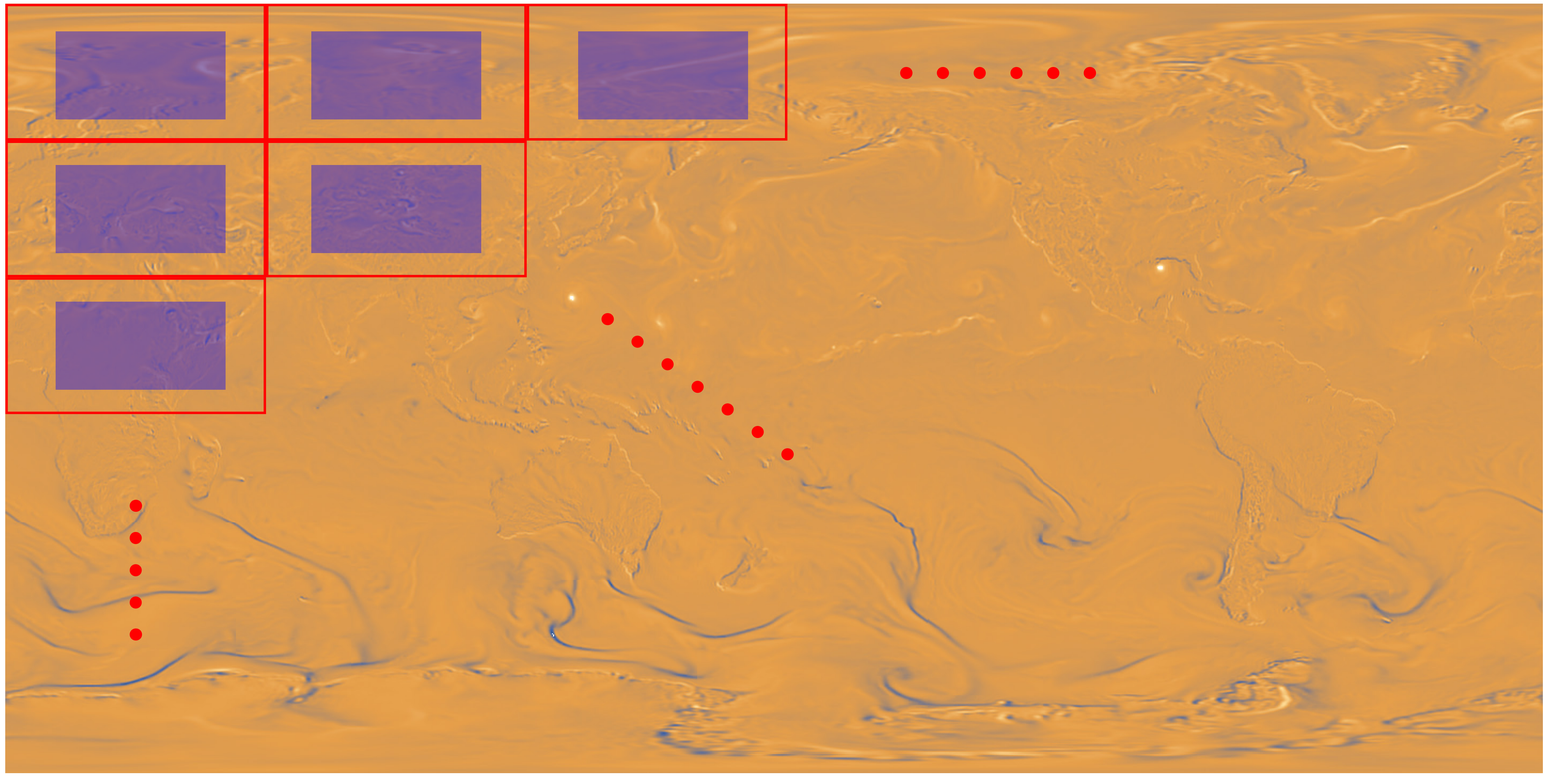
- How to do global forecasts with a local model?




© Christian Lessig, 2023

# Medium range forecasting

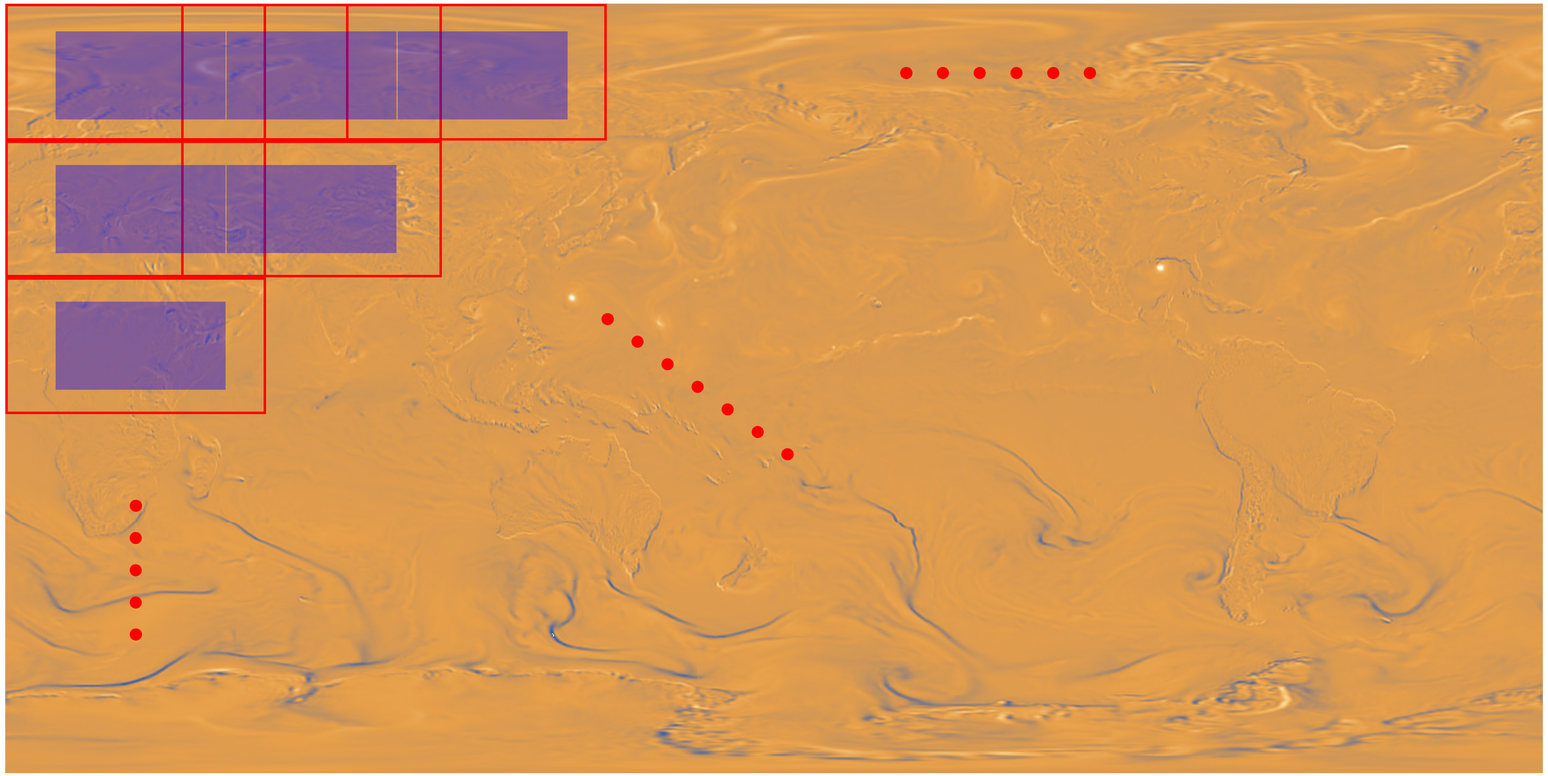
- How to do global forecasts with a local model?



© Christian Lessig, 2023

# Medium range forecasting

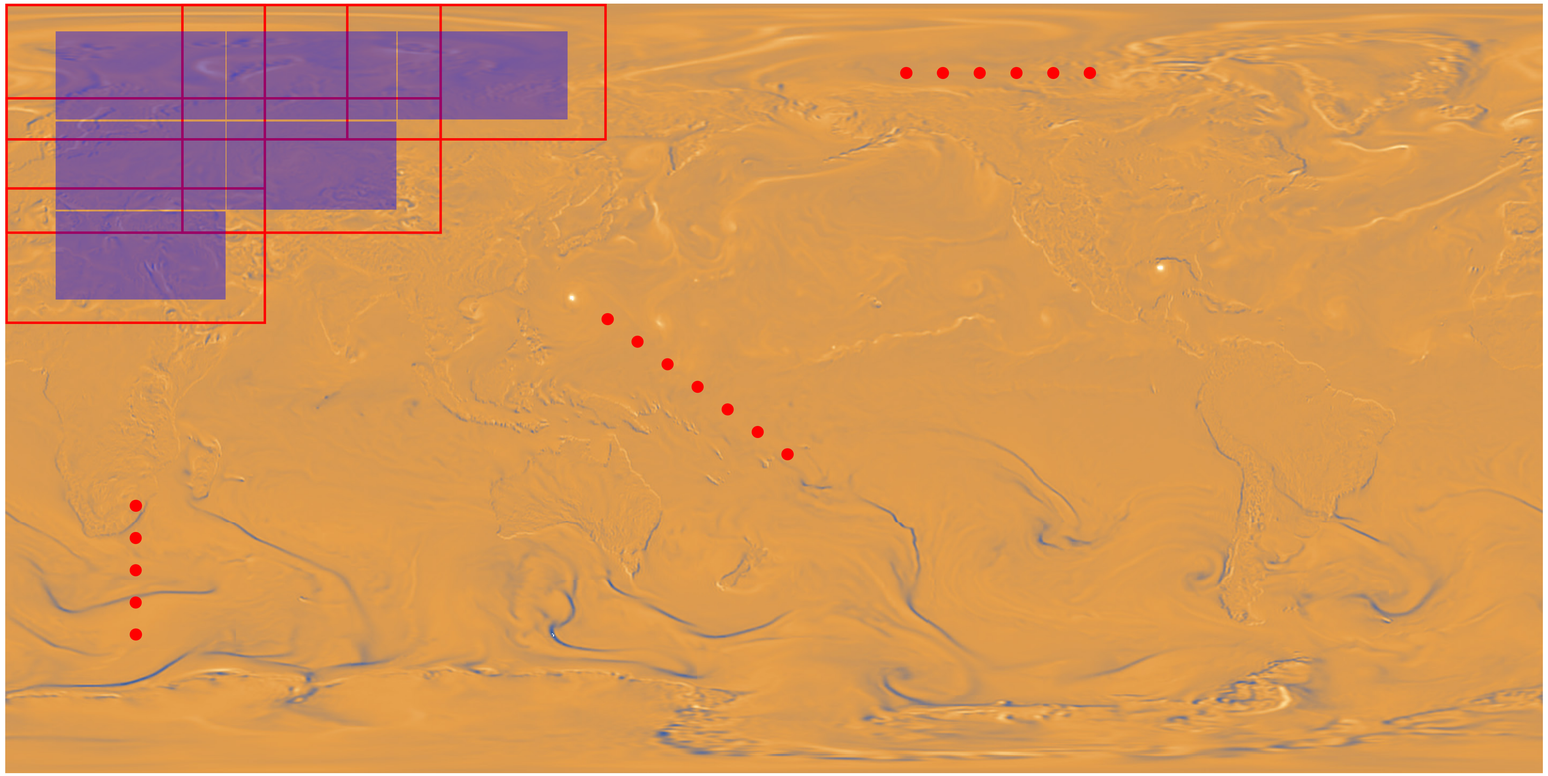
- How to do global forecasts with a local model?



© Christian Lessig, 2023

# Medium range forecasting

- How to do global forecasts with a local model?



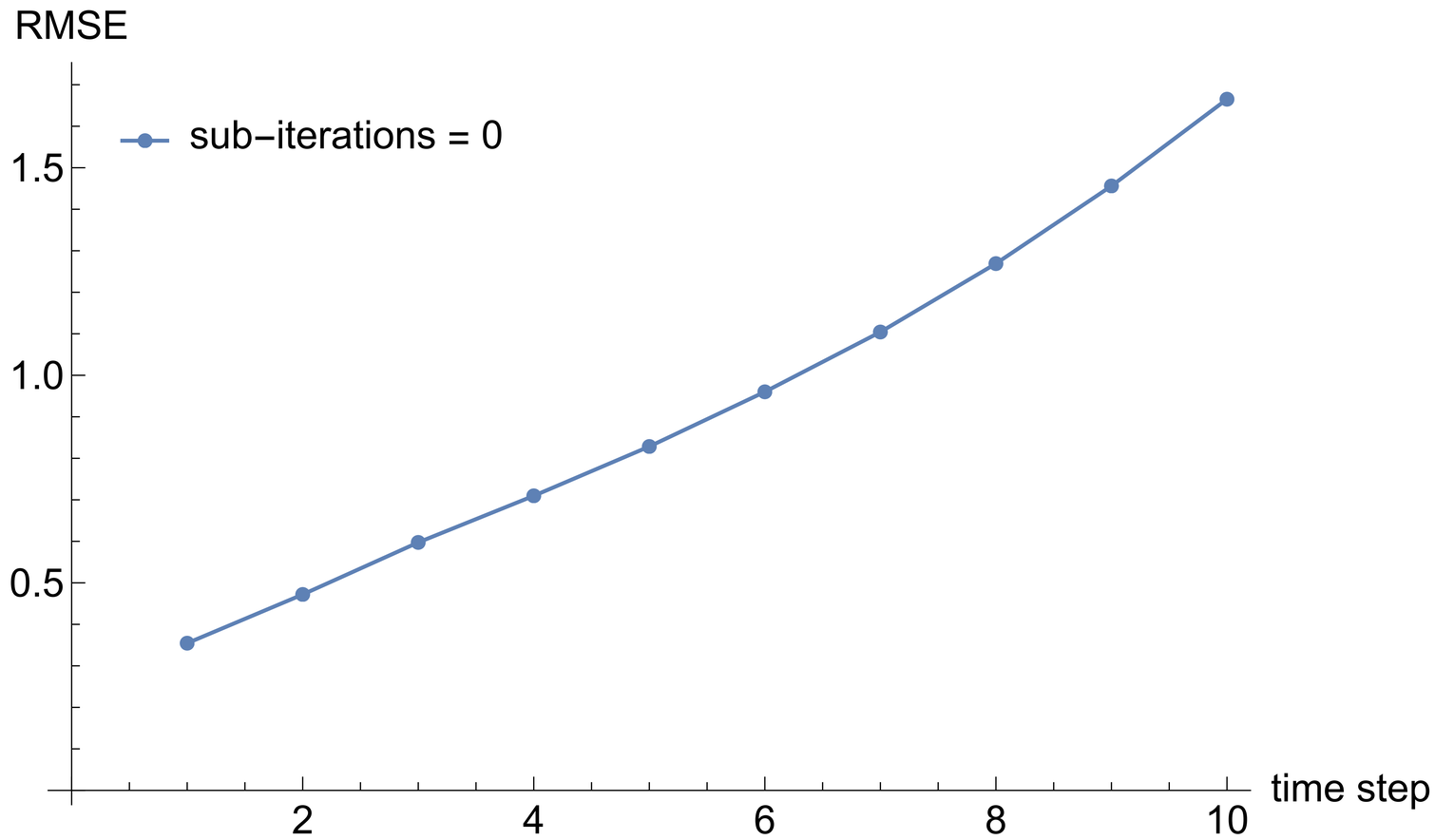
© Christian Lessig, 2023

# Medium range forecasting

Video

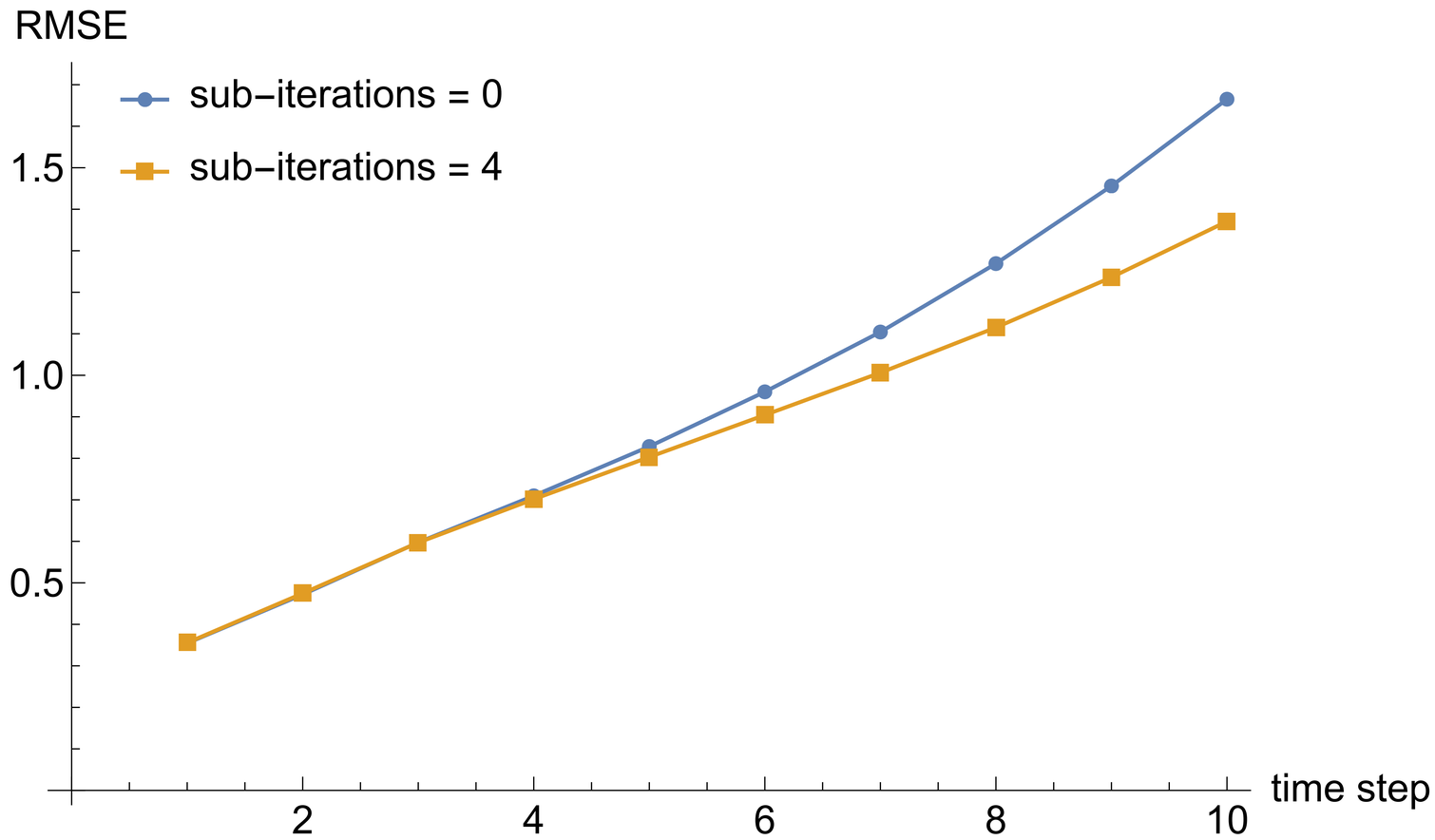
© Christian Lessig, 2023

# Medium range forecasting




© Christian Lessig, 2023

# Medium range forecasting



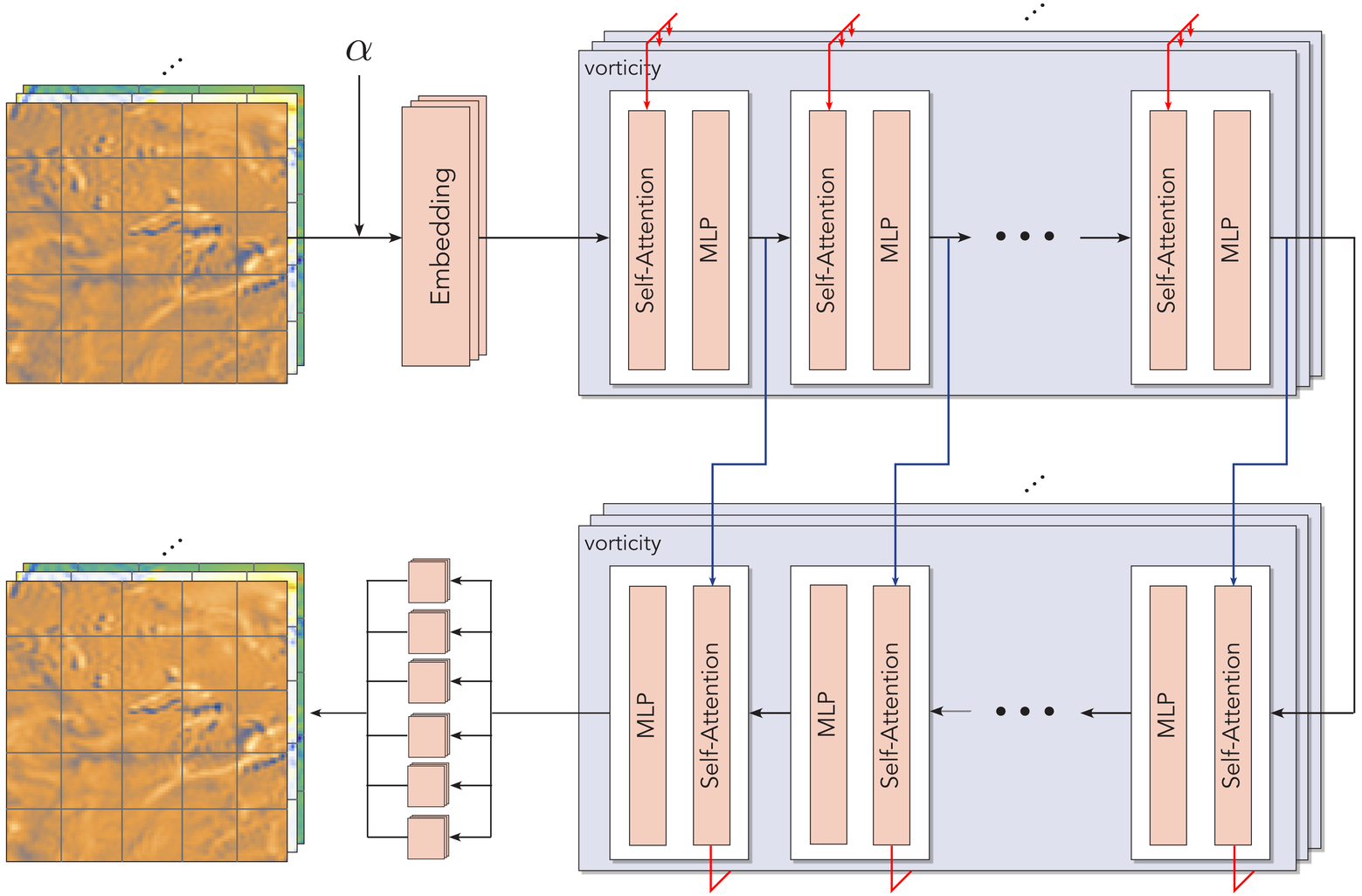

© Christian Lessig, 2023

# Downscaling

- Generate higher-resolution field from coarser input

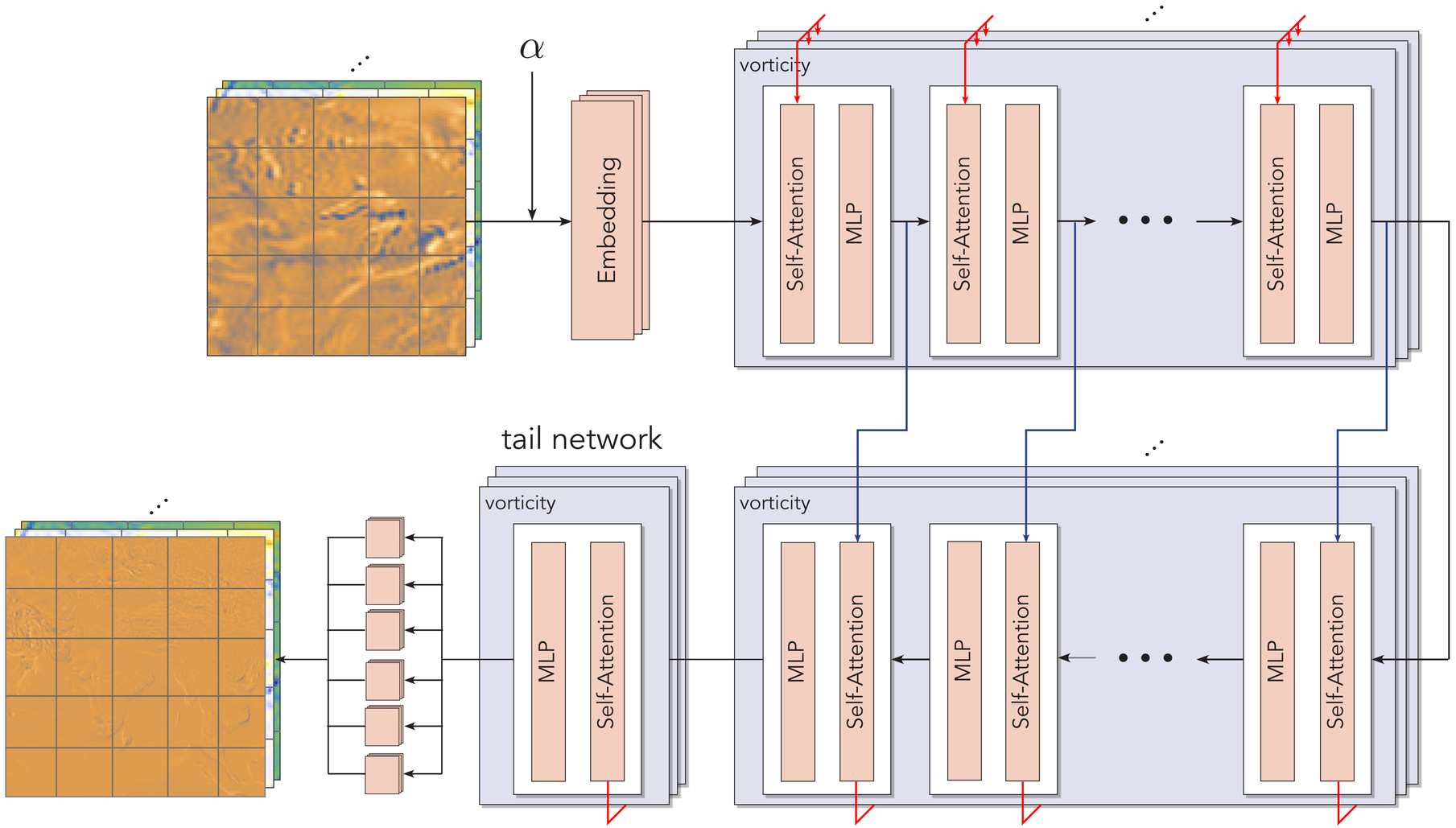
© Christian Lessig, 2023

# Downscaling



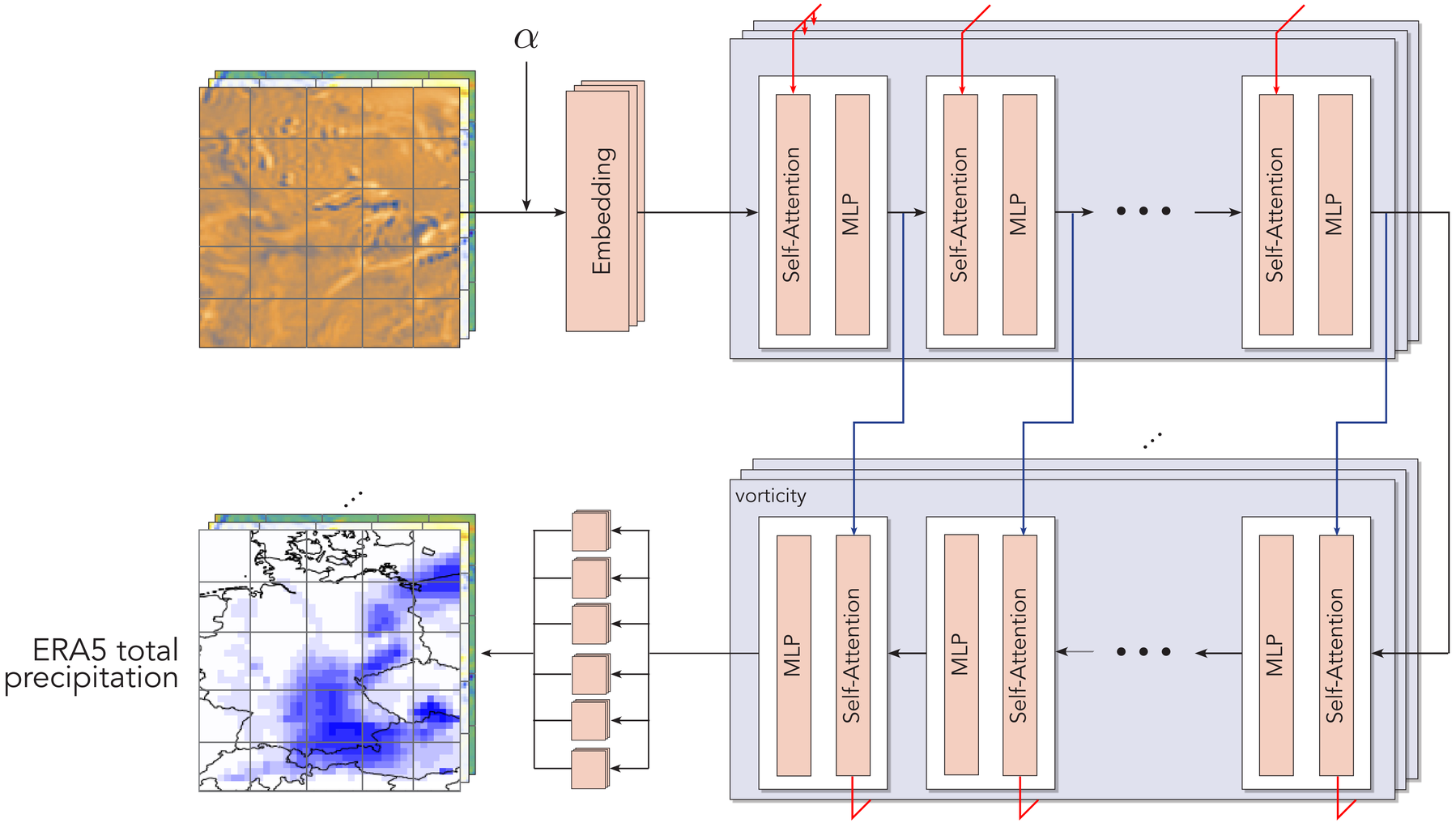

© Christian Lessig, 2023

# Downscaling



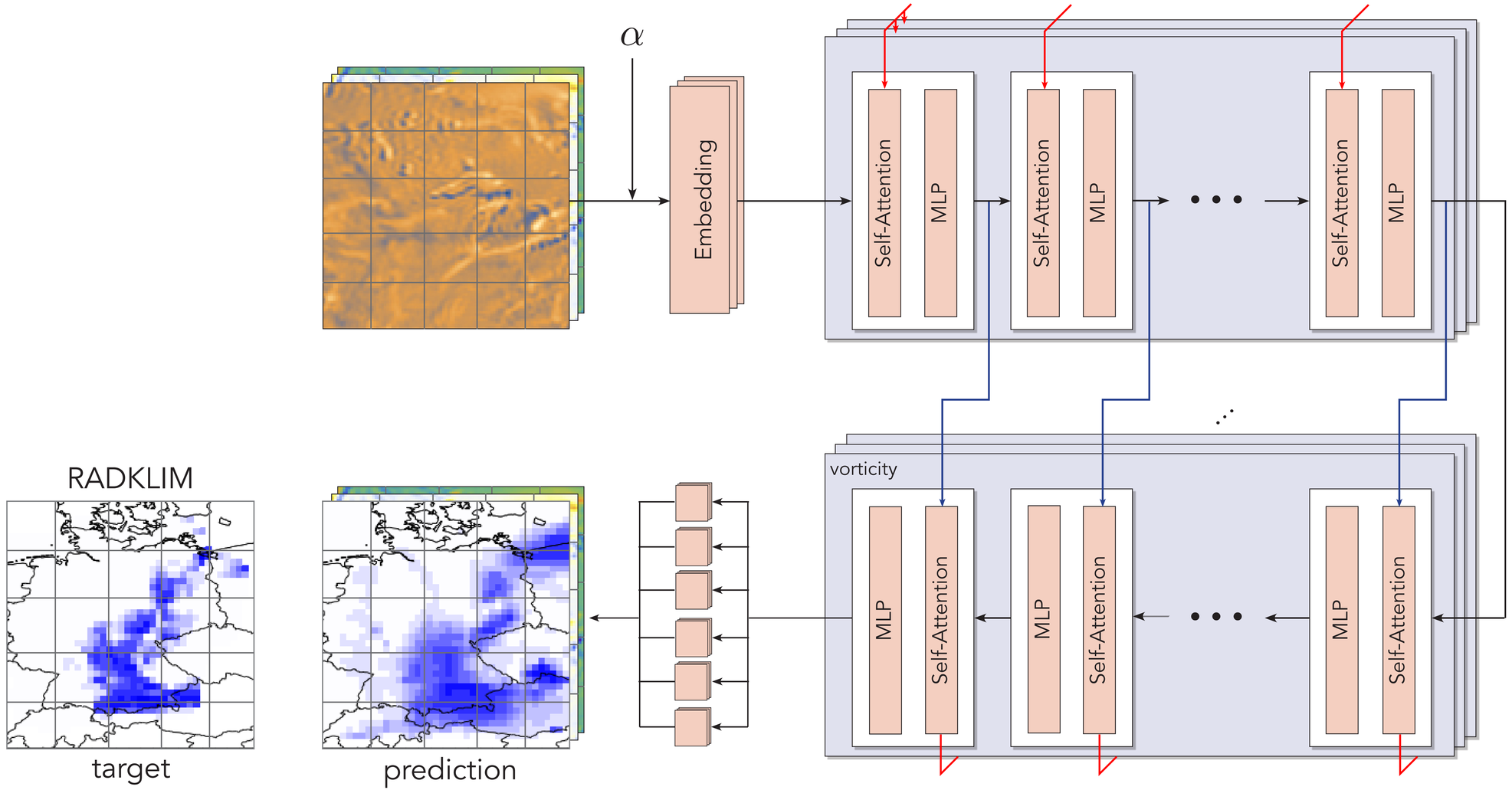

© Christian Lessig, 2023

# Bias correction



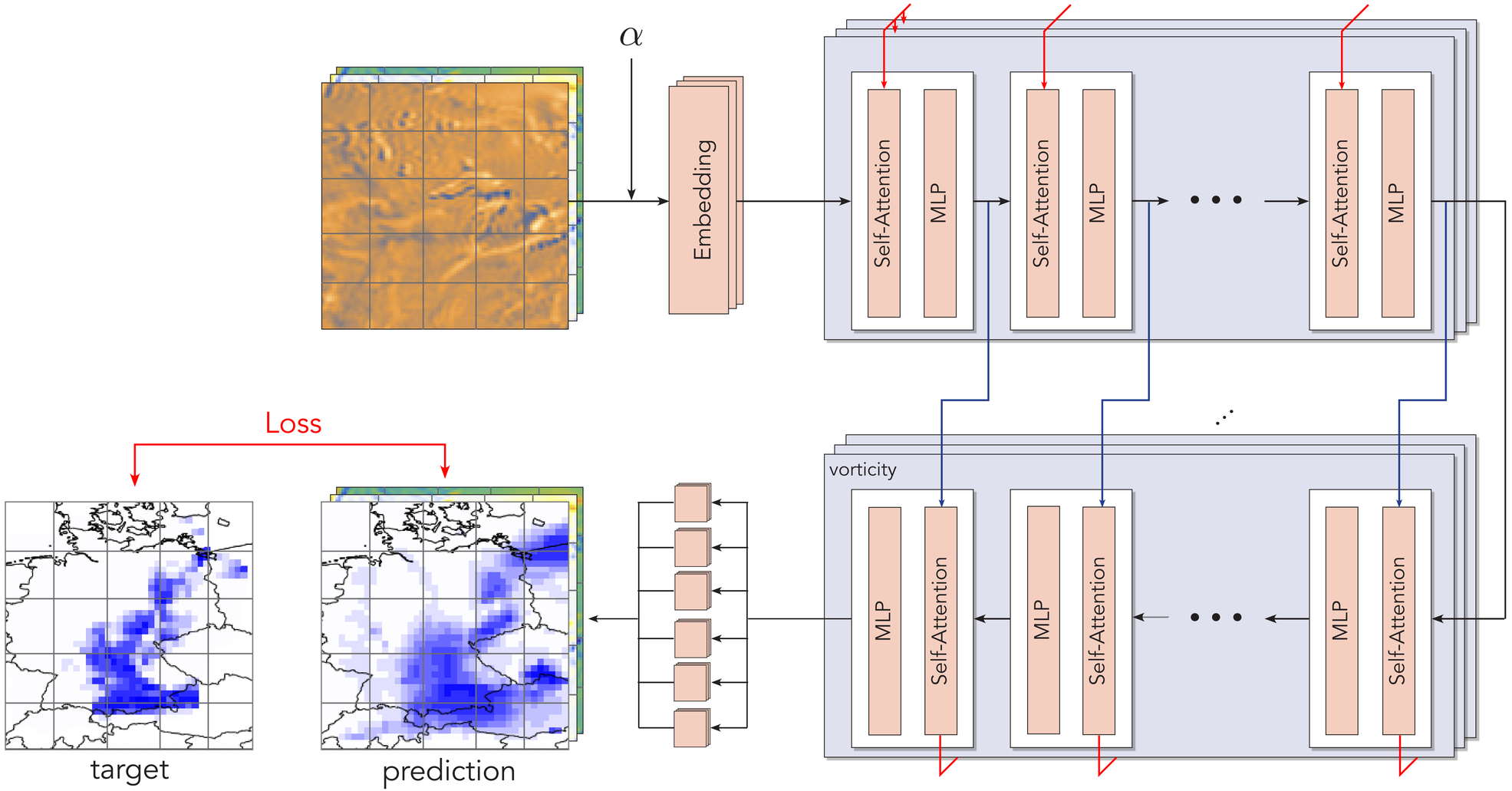

© Christian Lessig, 2023

# Bias correction




© Christian Lessig, 2023

# Bias correction




© Christian Lessig, 2023

# Outlook

© Christian Lessig, 2023

# Scaling

- Scale network to more levels, fields, 1950-...,
- Include longer-range effects by training on SST, soil moisture
- Train across multiple steps to better capture long-range interactions

© Christian Lessig, 2023

# Scaling

- Scale network to more levels, fields, 1950-...,
- Include longer-range effects by training on SST, soil moisture
- Train across multiple steps to better capture long-range interactions

How large does a numerical stochastic atmospheric model have to be to absorb ERA5?

© Christian Lessig, 2023

# Training on observations

- Neural networks are models that work well on heterogenous and noisy data
- Fine-tune a pre-trained model with observations instead of training from scratch

© Christian Lessig, 2023

# Training on observations

- Neural networks are models that work well on heterogenous and noisy data
- Fine-tune a pre-trained model with observations instead of training from scratch

Can one continuous update a model with observations?

How to handle and propagate uncertainties?

© Christian Lessig, 2023

# GPT$^{x}$

- LLMs can generate runable code (python, C++, ...)

© Christian Lessig, 2023

# GPT$^x$

- LLMs can generate runable code (python, C++, ...)

=> Model can autonomously gather new data and optimize its input

=> Model can fine-tune itself

© Christian Lessig, 2023

# GPT$^x$

- LLMs can generate runable code (python, C++, ...)

=> Model can autonomously gather new data and optimize its input

=> Model can fine-tune itself

What is the equivalent for a stochastic atmospheric model?

© Christian Lessig, 2023

# Summary

© Christian Lessig, 2023

# AtmoRep

- Numerical statistical atmospheric model
  - Complementary to classical GCMs and ESMs
  - Represented by very large neural network
  - Very long training leads to continuous improvement

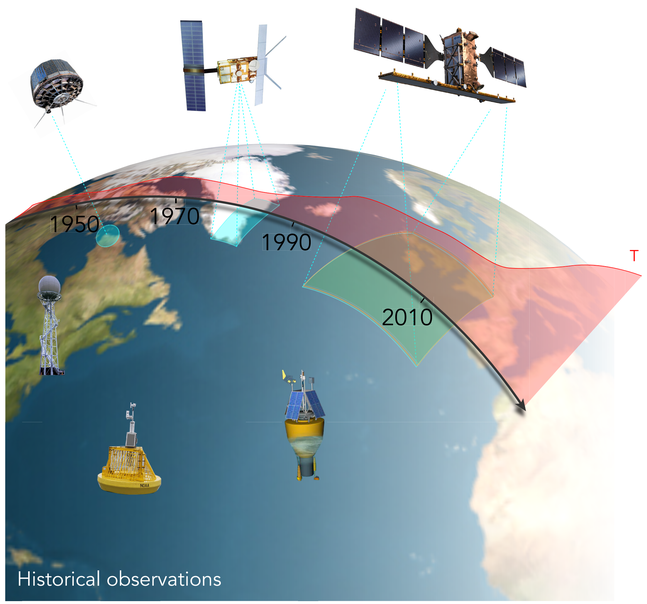
© Christian Lessig, 2023

# AtmoRep

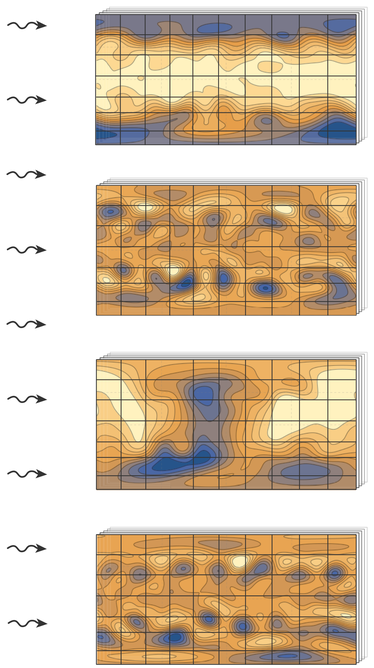
- Numerical statistical atmospheric model
  - Complementary to classical GCMs and ESMs
  - Represented by very large neural network
  - Very long training leads to continuous improvement
- BERT-type training leads to versatile intrinsic capabilites
  - Forecasting, temporal interpolation, model correction, ...
  - Straight forward extension to various applications

© Christian Lessig, 2023

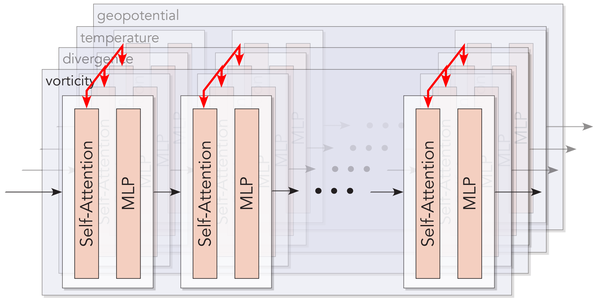
# AtmoRep



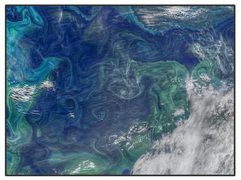

ERA5 reanalysis

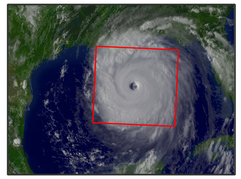


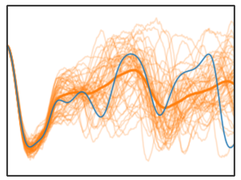
$p_\theta(y|x)$

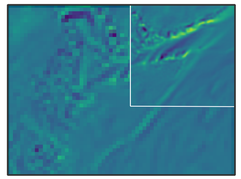


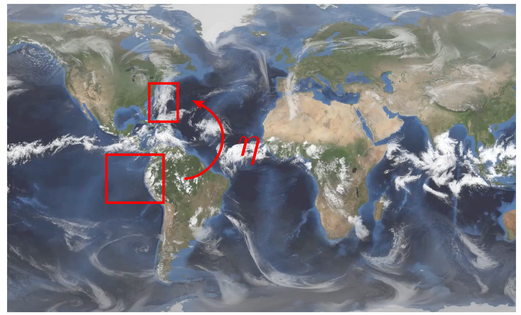

applications


Forecasting


Impact analysis


Statistical analysis


Downscaling

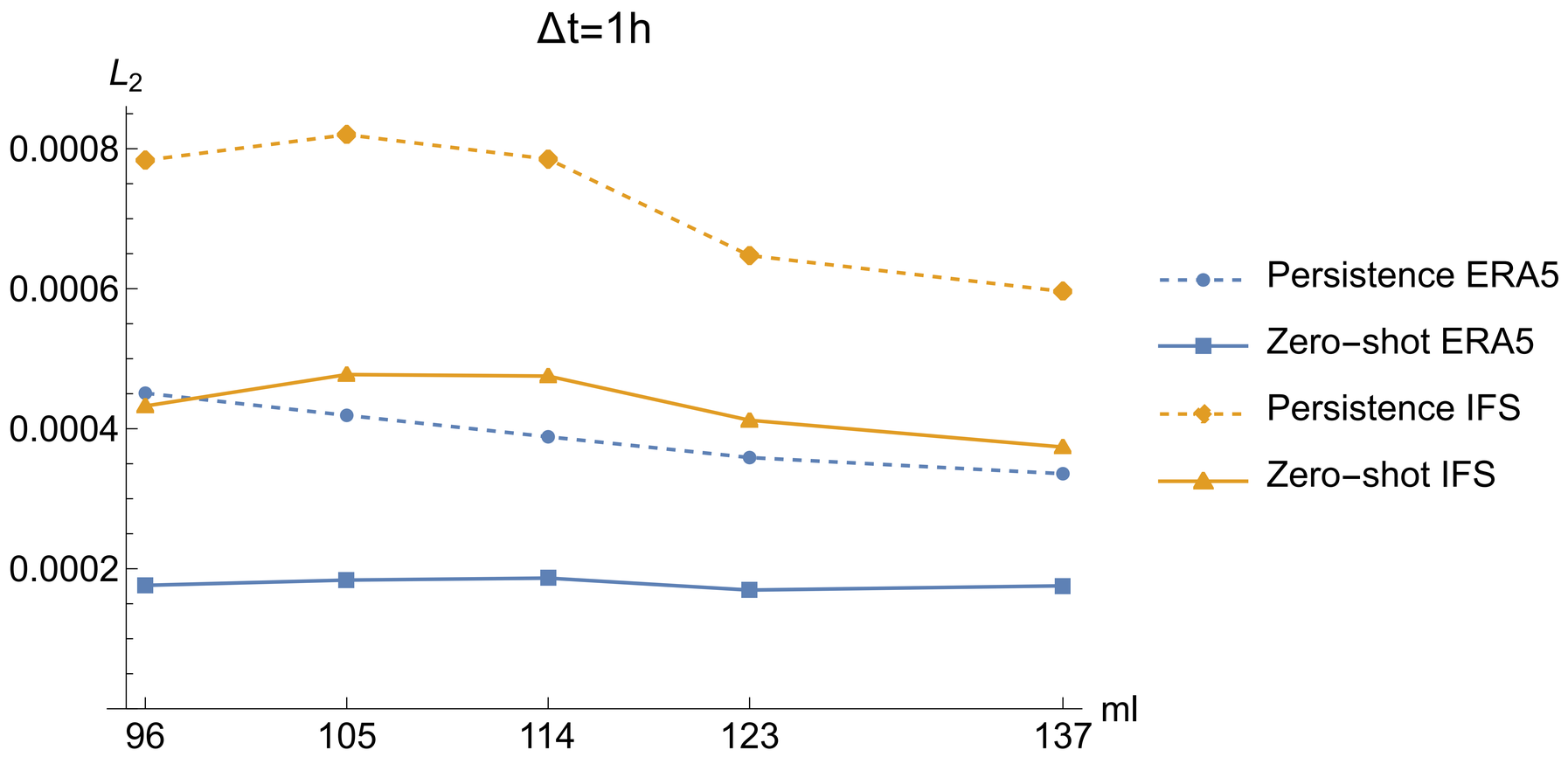


scientific insight

© Christian Lessig, 2023

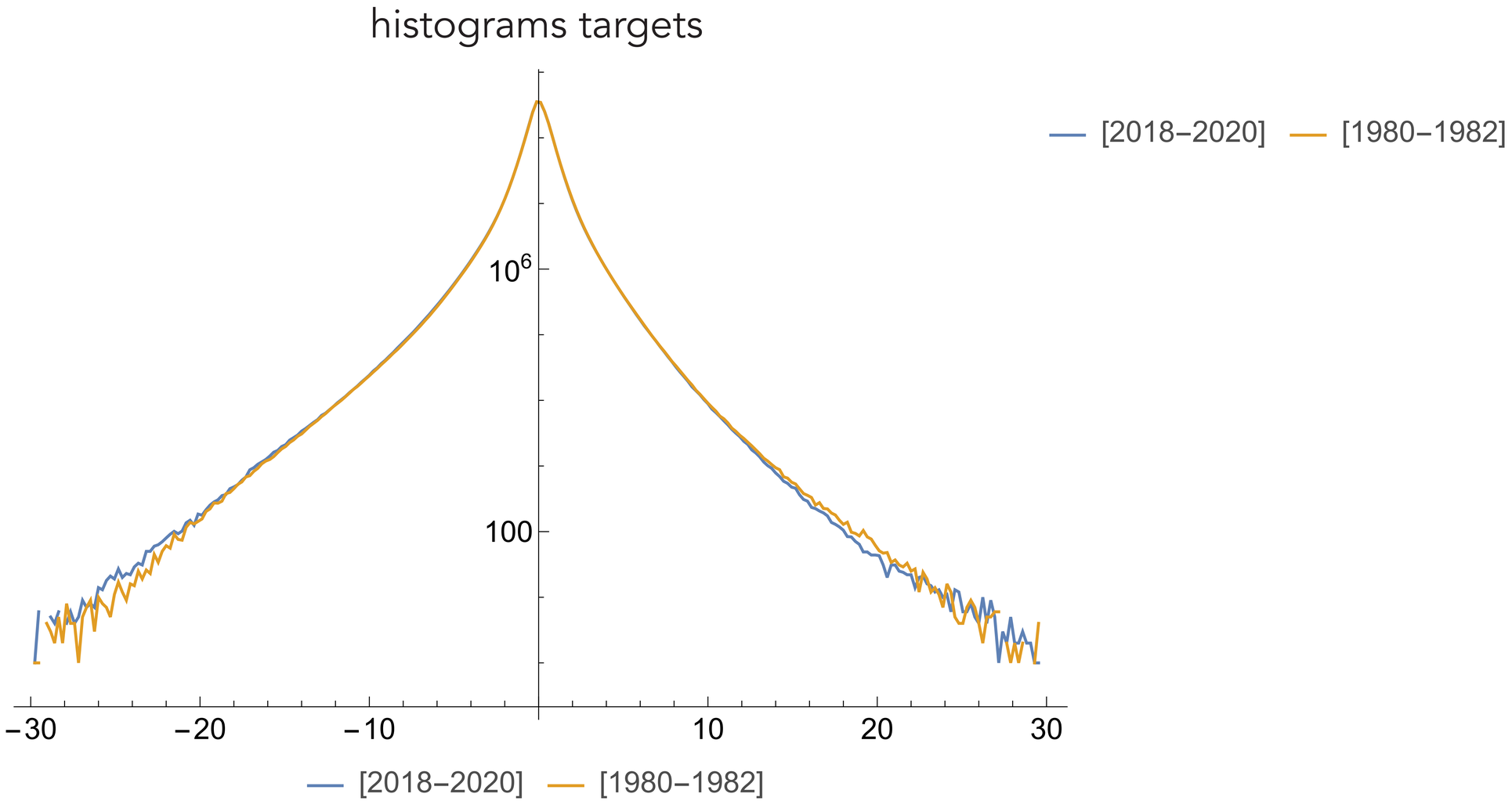
© Christian Lessig, 2023

# Model correction



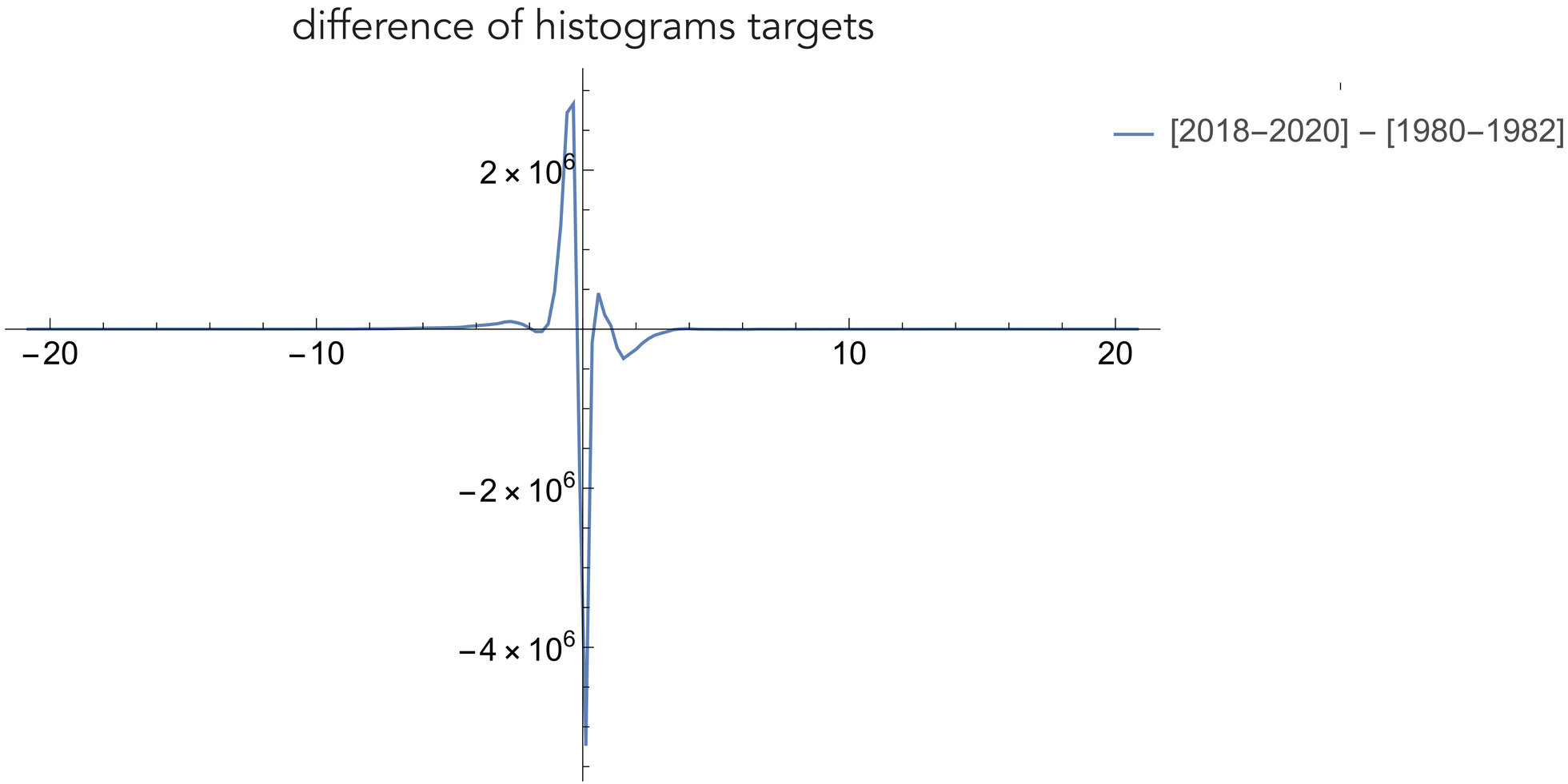

© Christian Lessig, 2023

# Counterfactuals



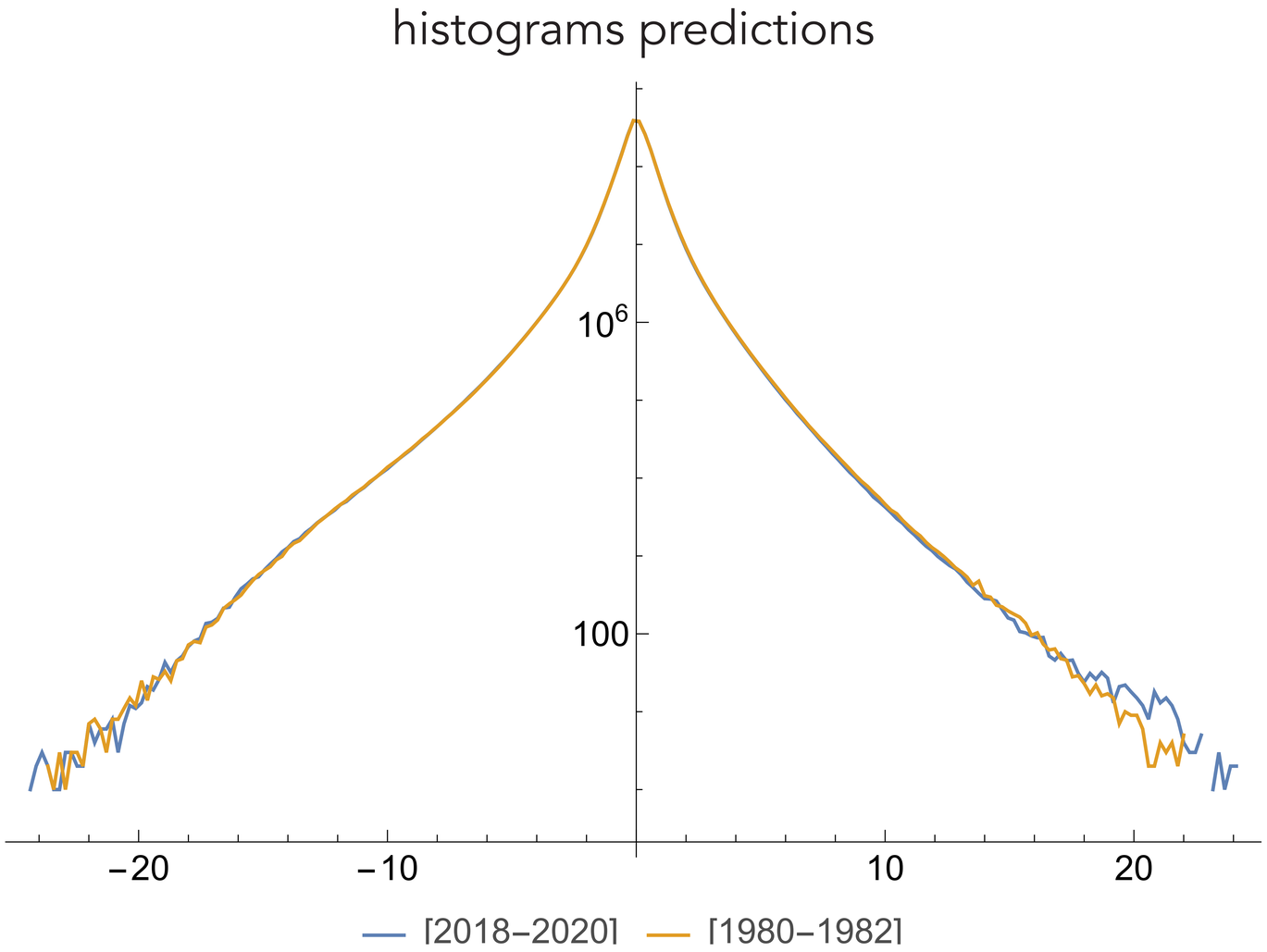

© Christian Lessig, 2023

# Counterfactuals



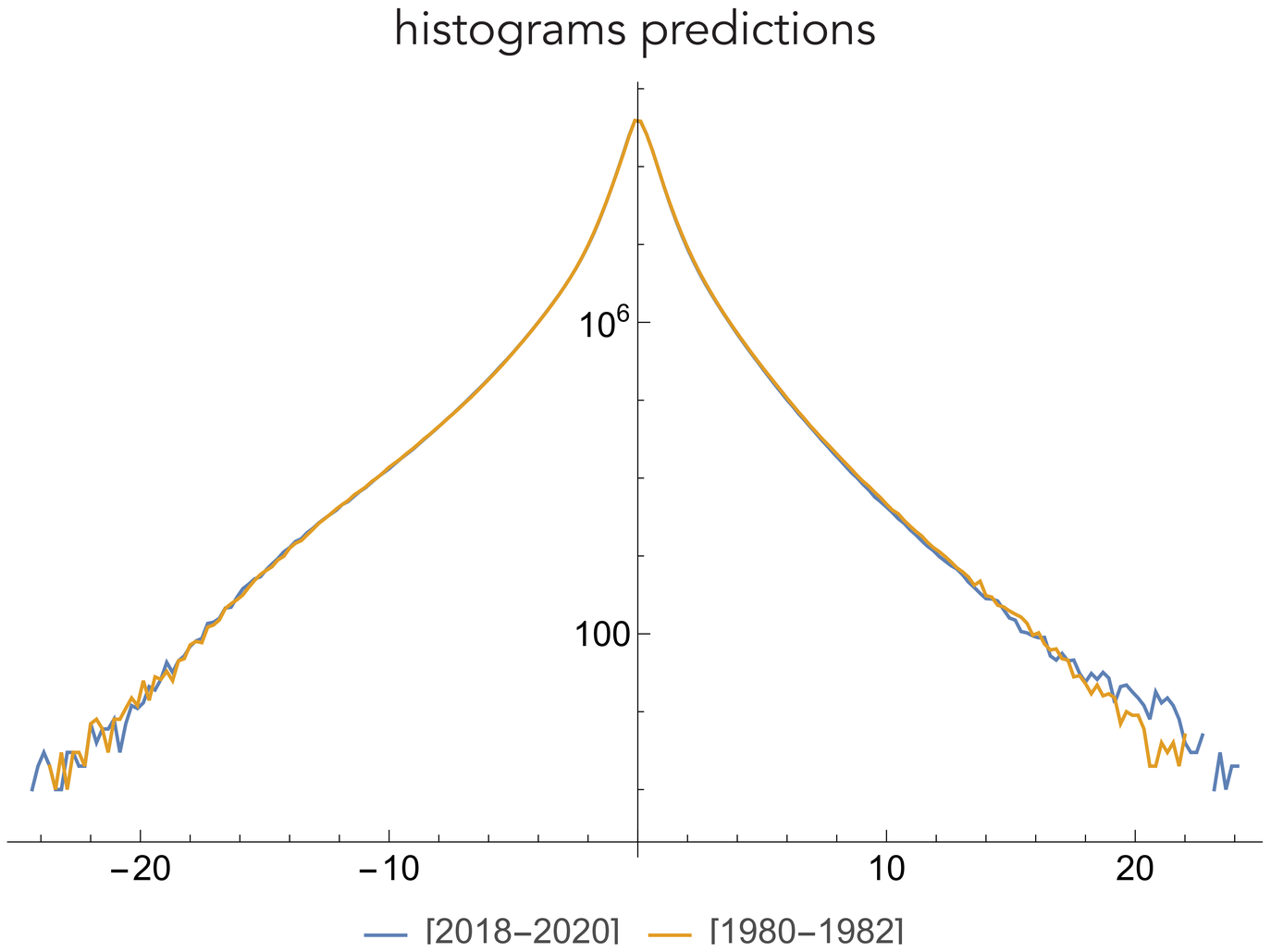

© Christian Lessig, 2023

# Counterfactuals




© Christian Lessig, 2023

# Counterfactuals



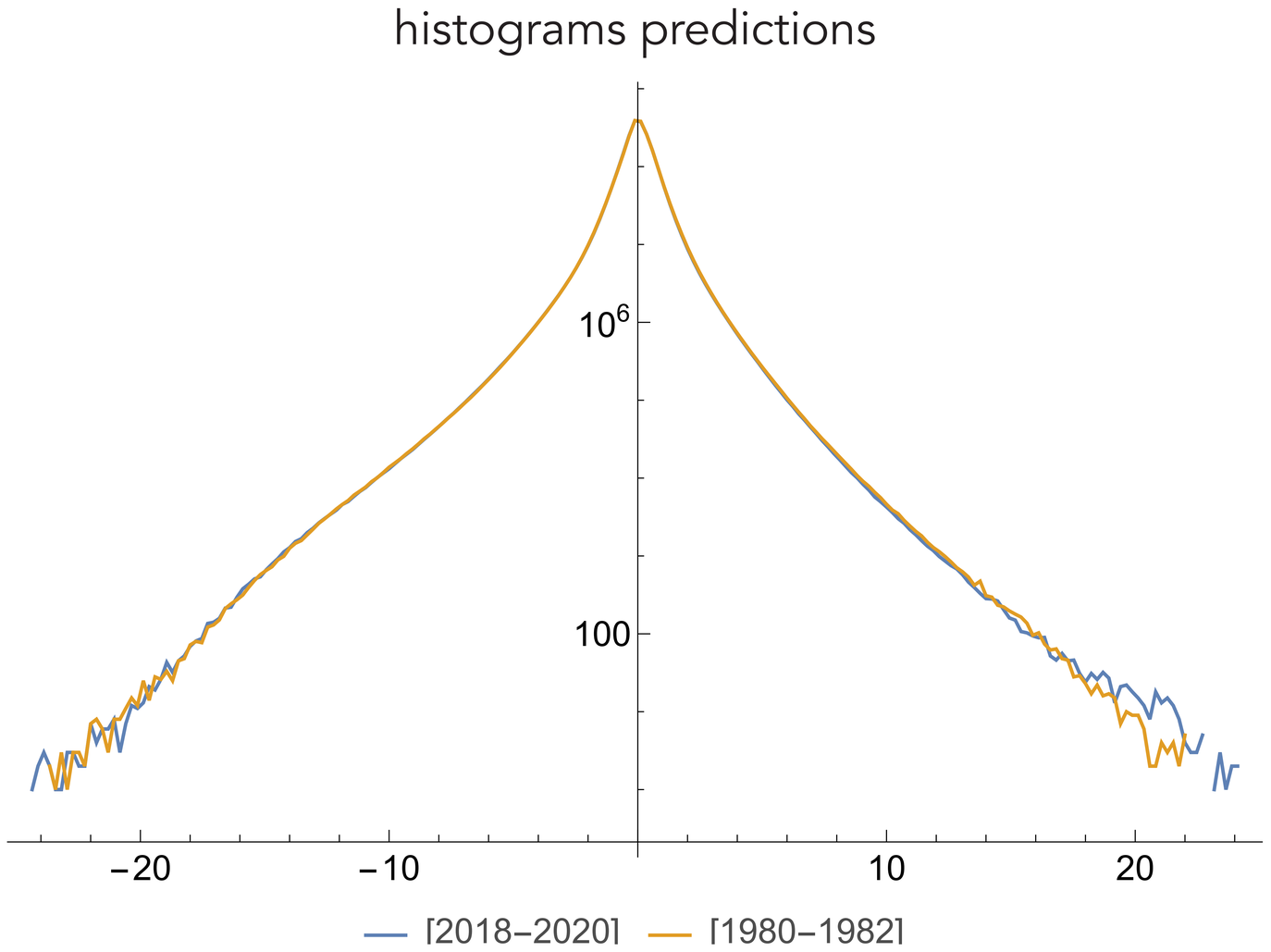

[2018–2020]

initial conditions *but*

$\alpha_{\text{year}} = [1980, 1982]$

© Christian Lessig, 2023

# Counterfactuals




[2018–2020]

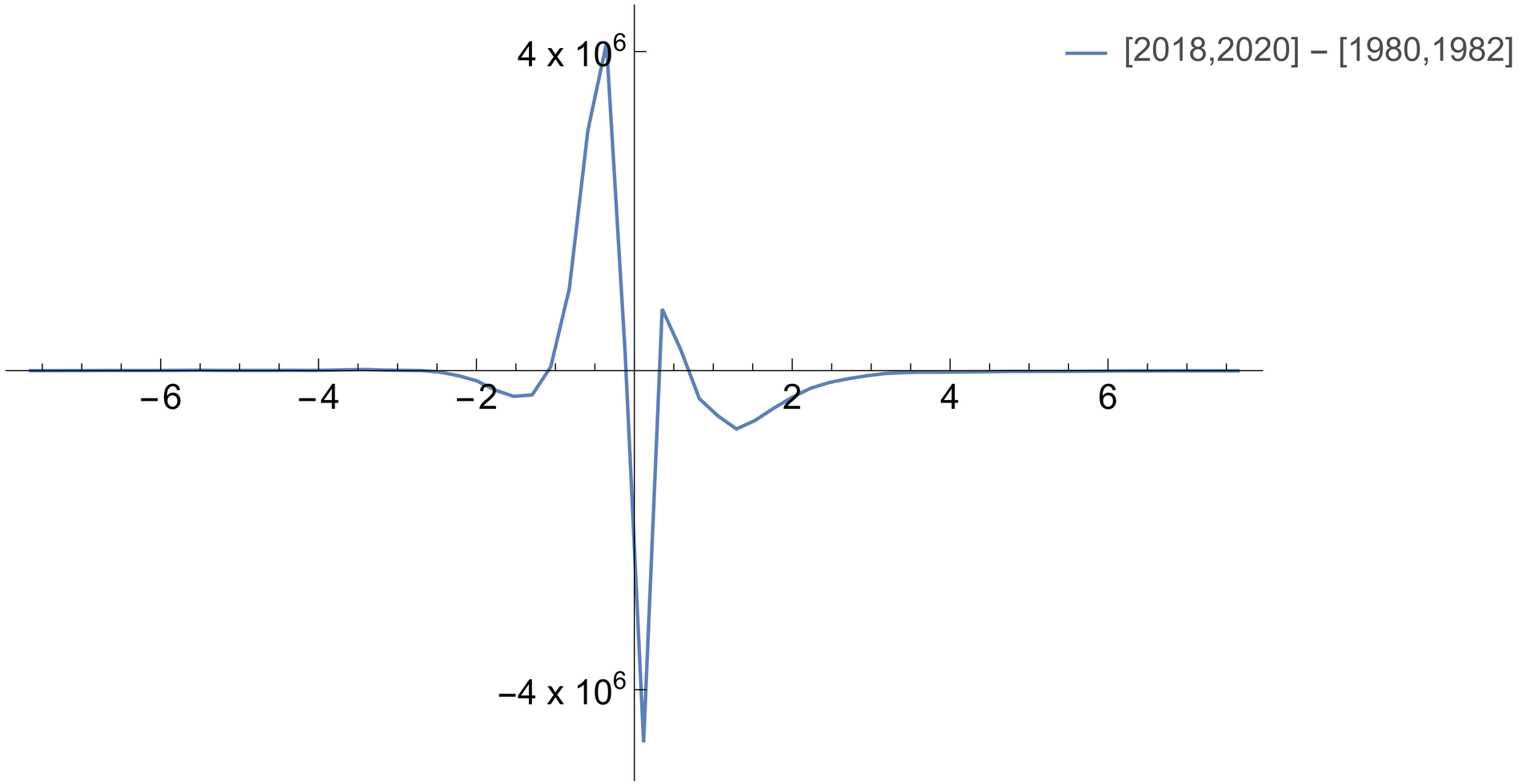
initial conditions *but* $\alpha_{\text{year}} = [1980, 1982]$

[1980–1982]

initial conditions *but* $\alpha_{\text{year}} = [2018, 2020]$

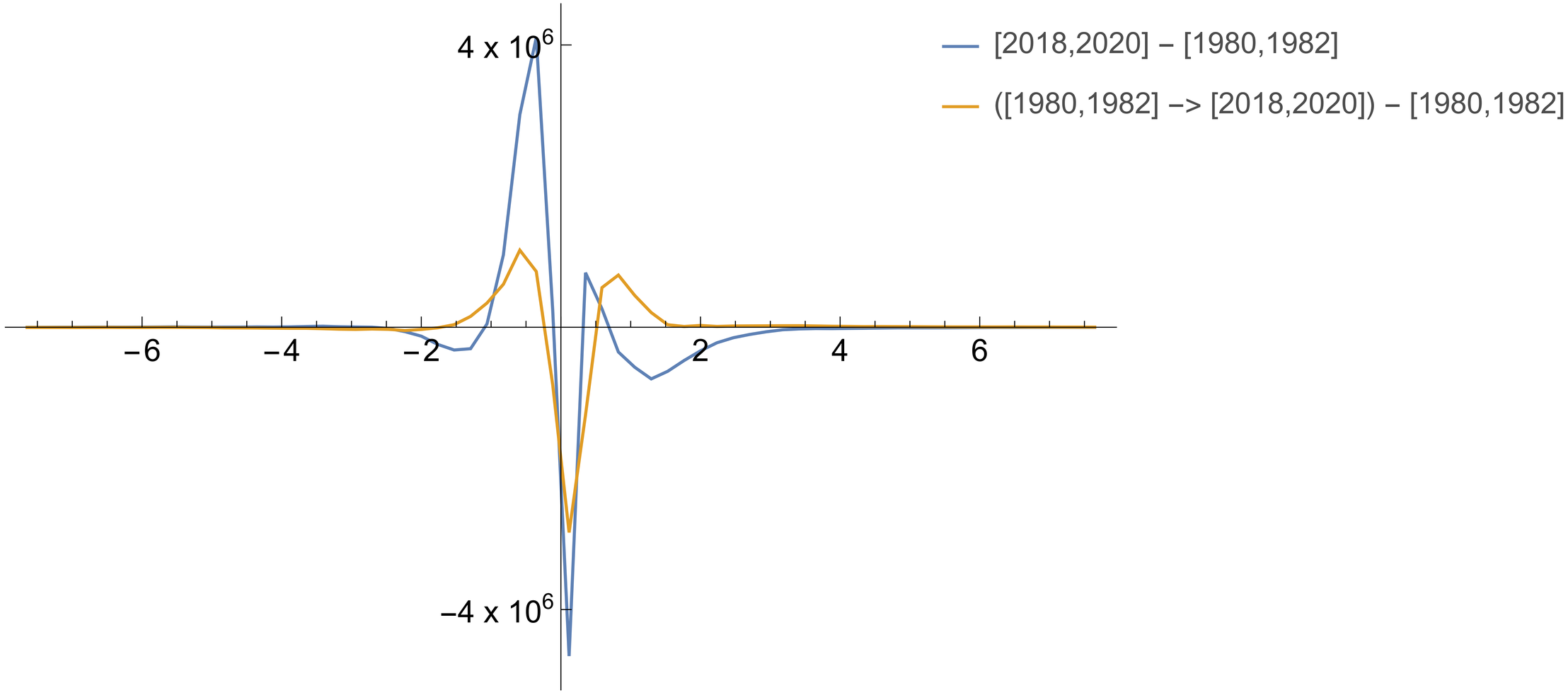
© Christian Lessig, 2023

# Counterfactuals



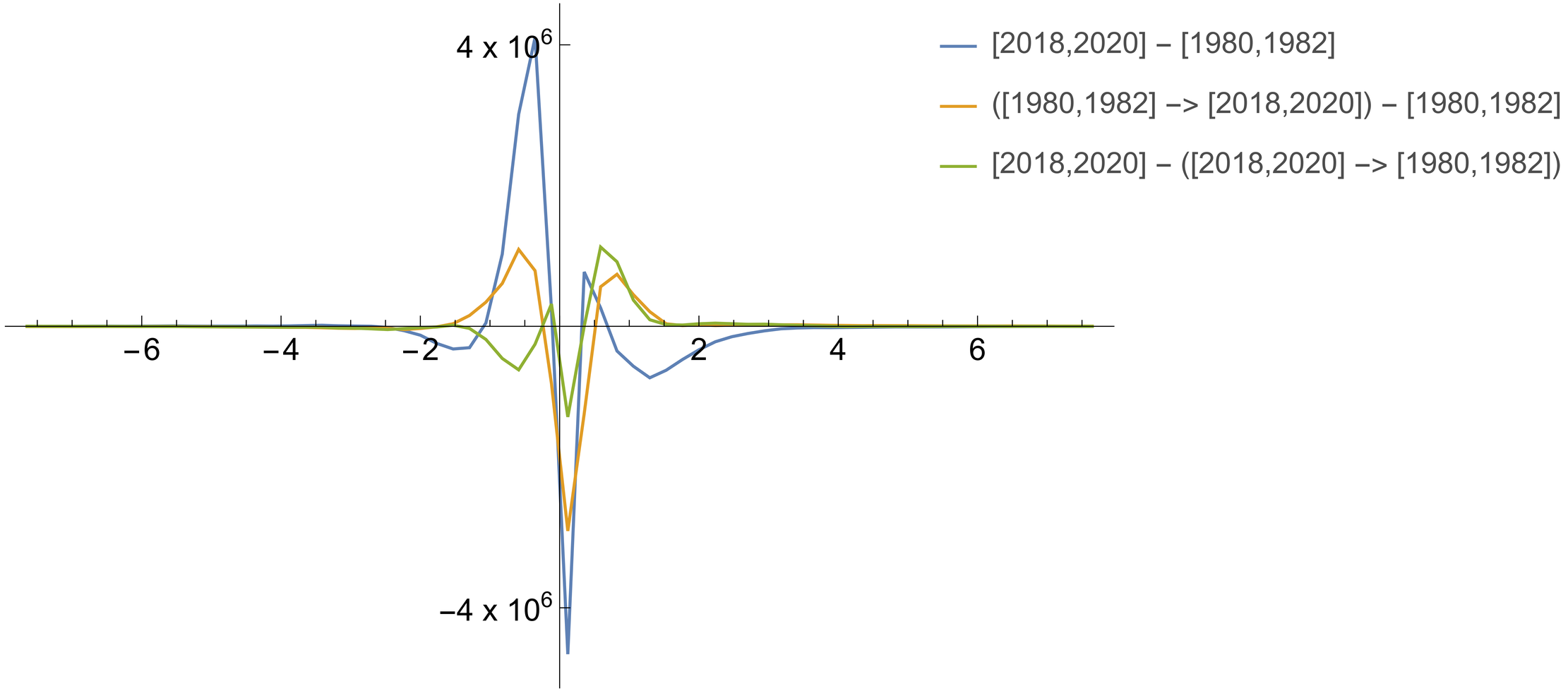

© Christian Lessig, 2023

# Counterfactuals



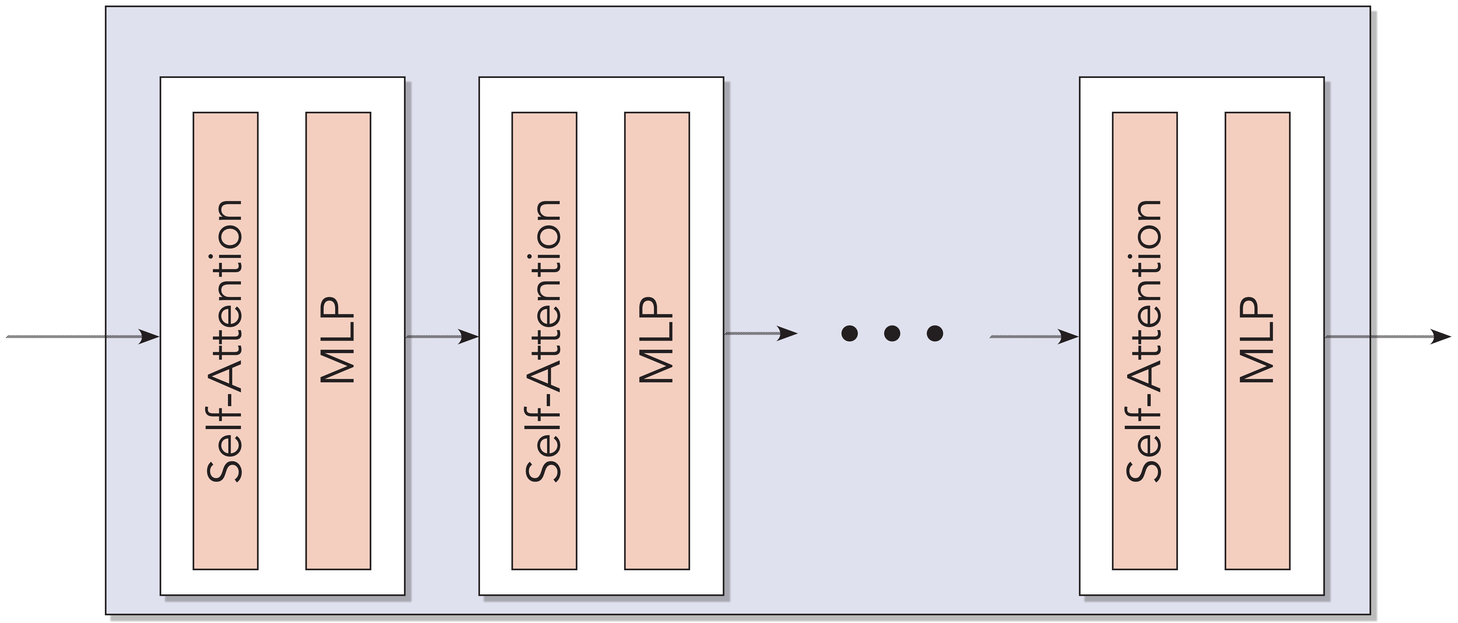

© Christian Lessig, 2023

# Counterfactuals




© Christian Lessig, 2023

# Multiformer

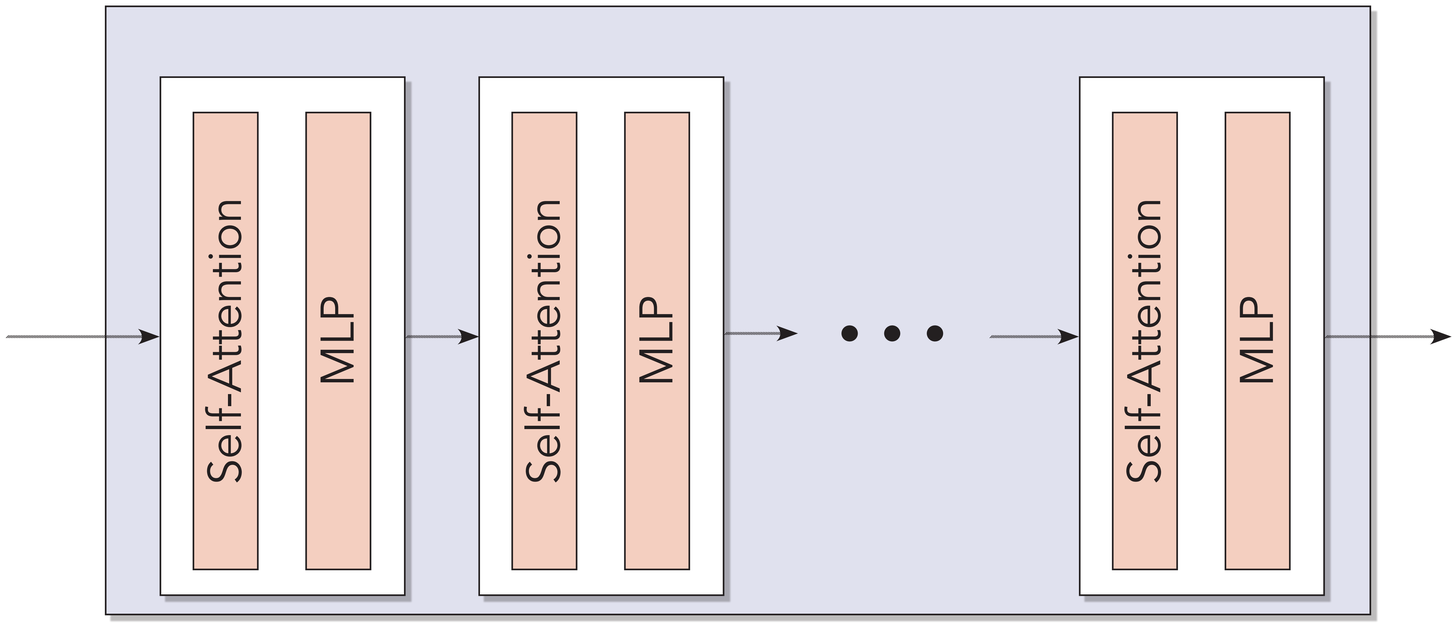



© Christian Lessig, 2023

# Multiformer

Self attention

$$\sigma\left(Q\,K^T\right)V$$

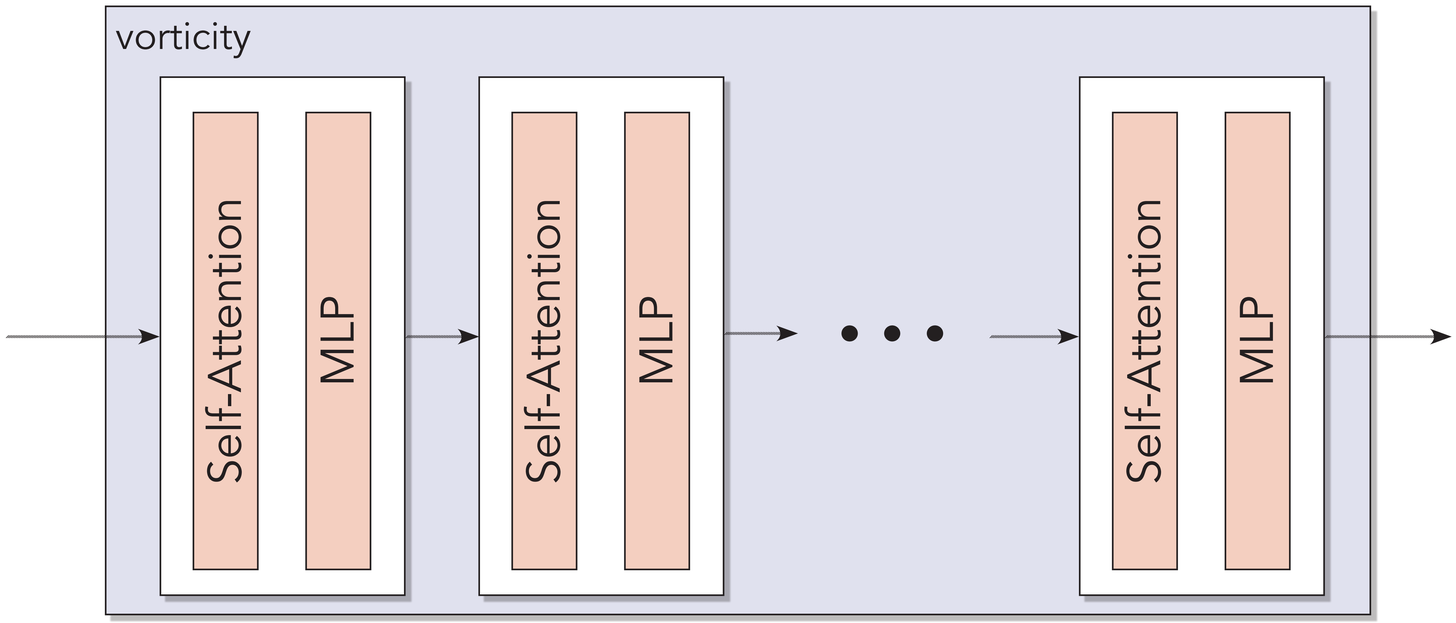



© Christian Lessig, 2023

# Multiformer

Self
attention

$$\sigma\left(Q\,K^T\right)V$$

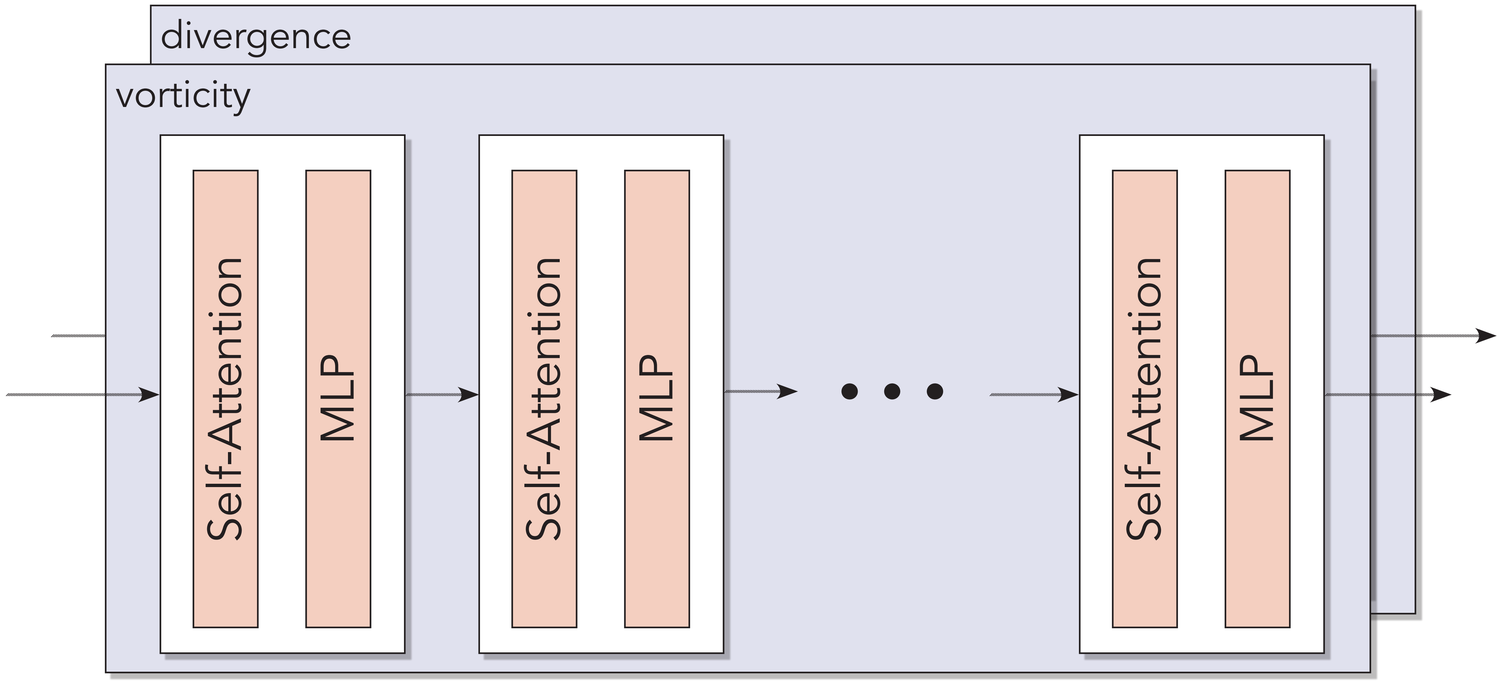



© Christian Lessig, 2023

# Multiformer

Self
attention

$$\sigma\left(Q\,K^T\right)V$$

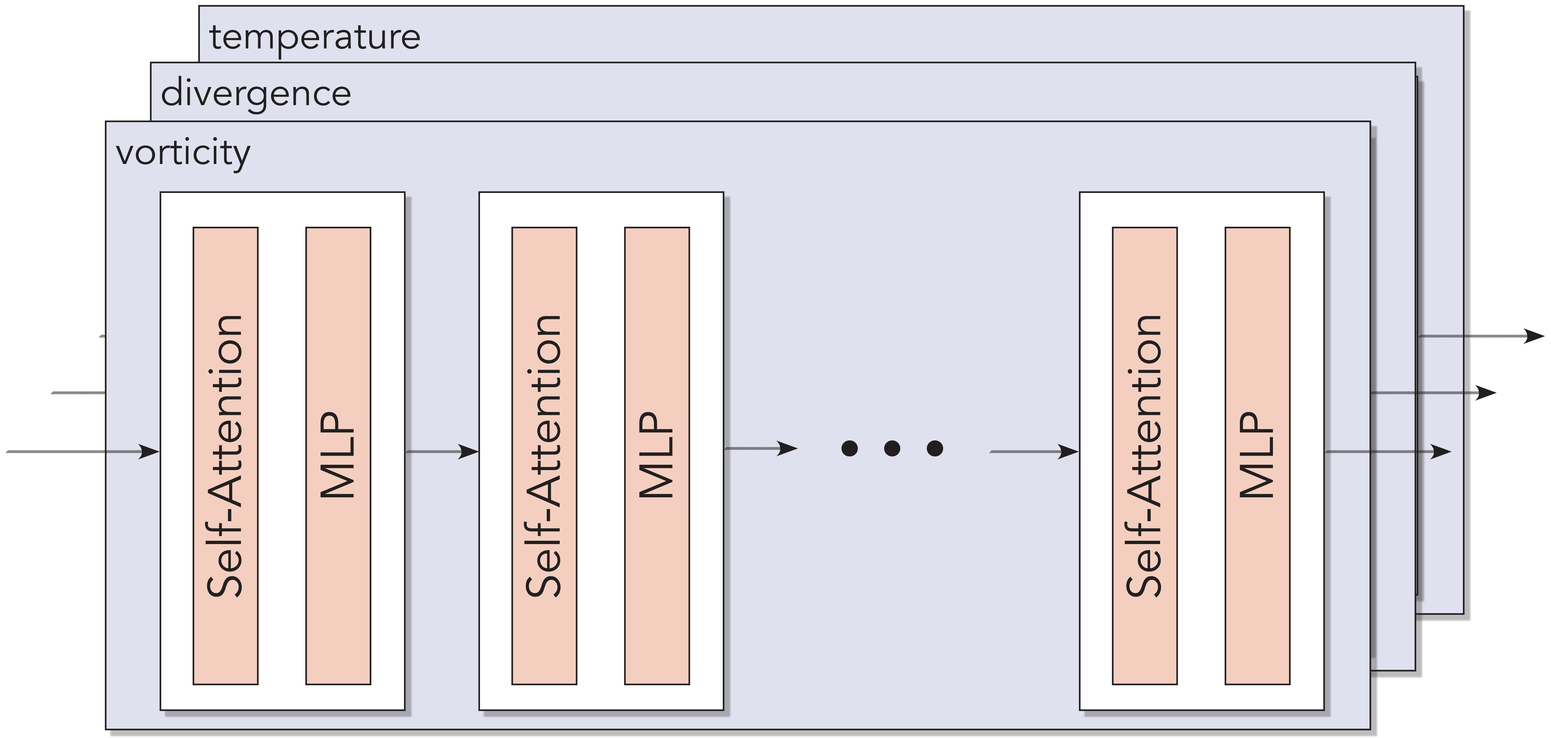



© Christian Lessig, 2023

# Multiformer

Self
attention

$$\sigma\left(Q\,K^T\right)V$$

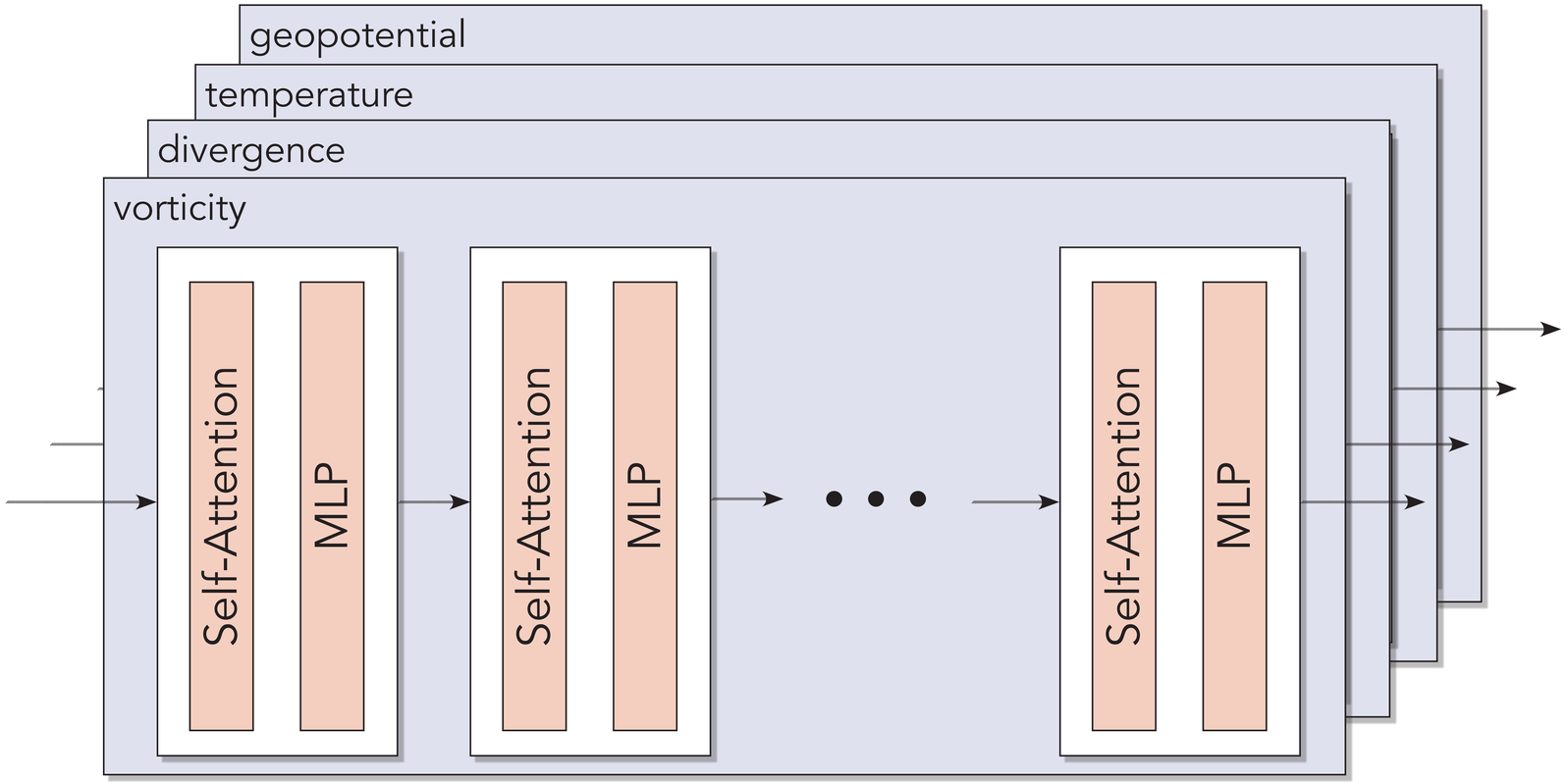



© Christian Lessig, 2023

# Multiformer

Self
attention

$$\sigma\left(Q\,K^T\right)V$$

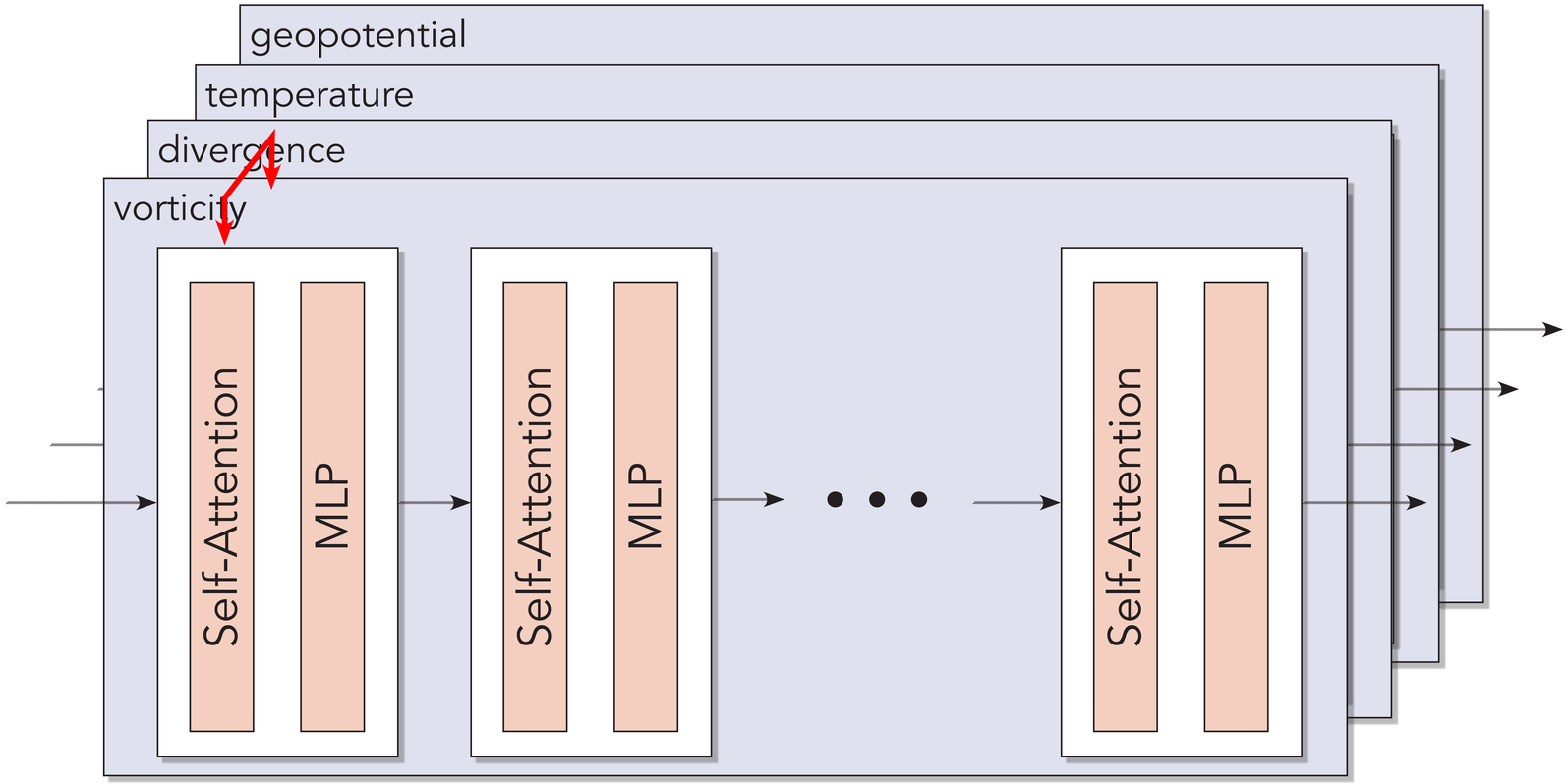



© Christian Lessig, 2023

# Multiformer

Self
attention

$$\sigma\left(Q\,K^T\right)V$$

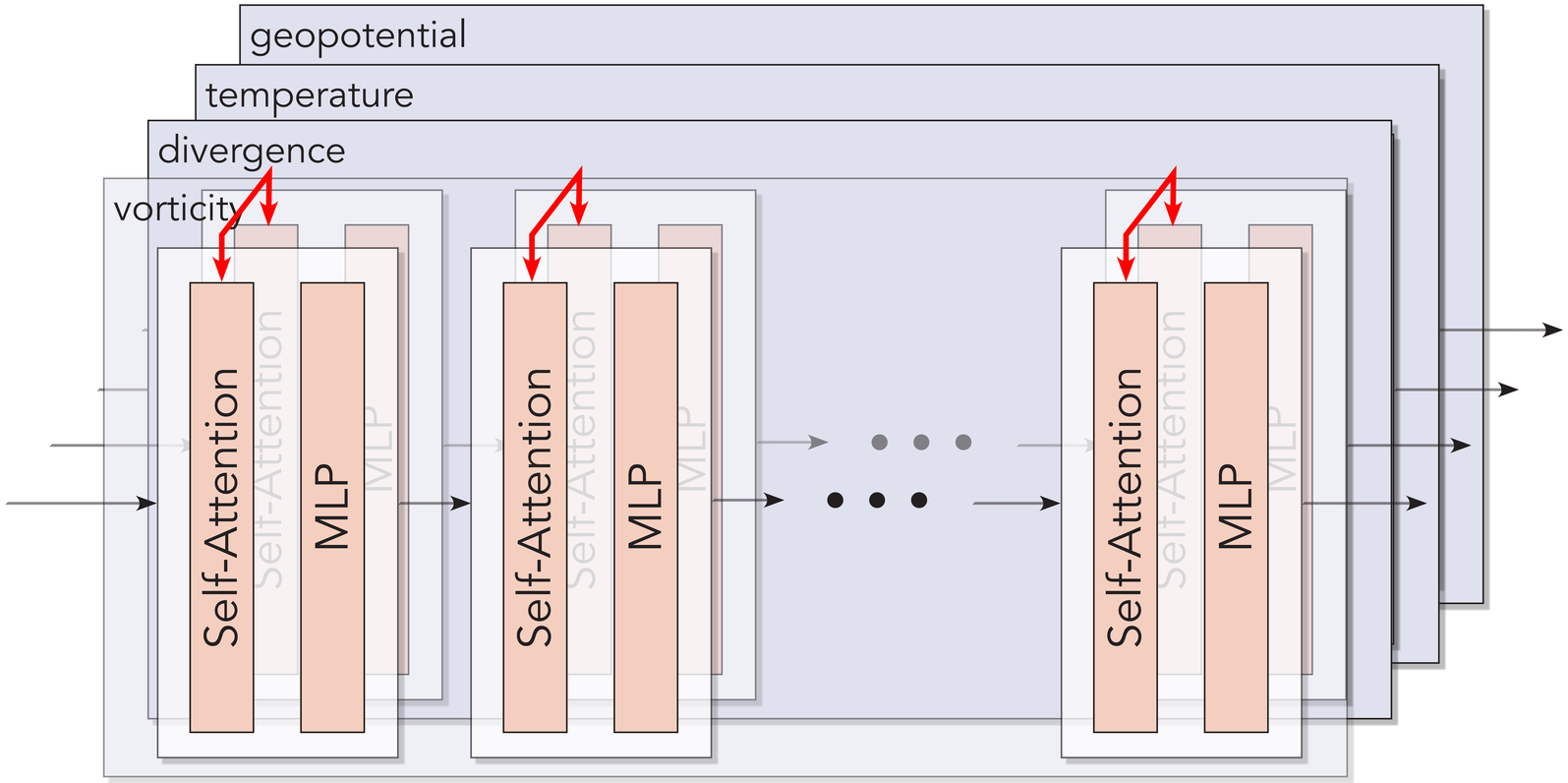



© Christian Lessig, 2023

# Multiformer

Self
attention

$$\sigma\left(Q\,K^T\right)V$$




© Christian Lessig, 2023

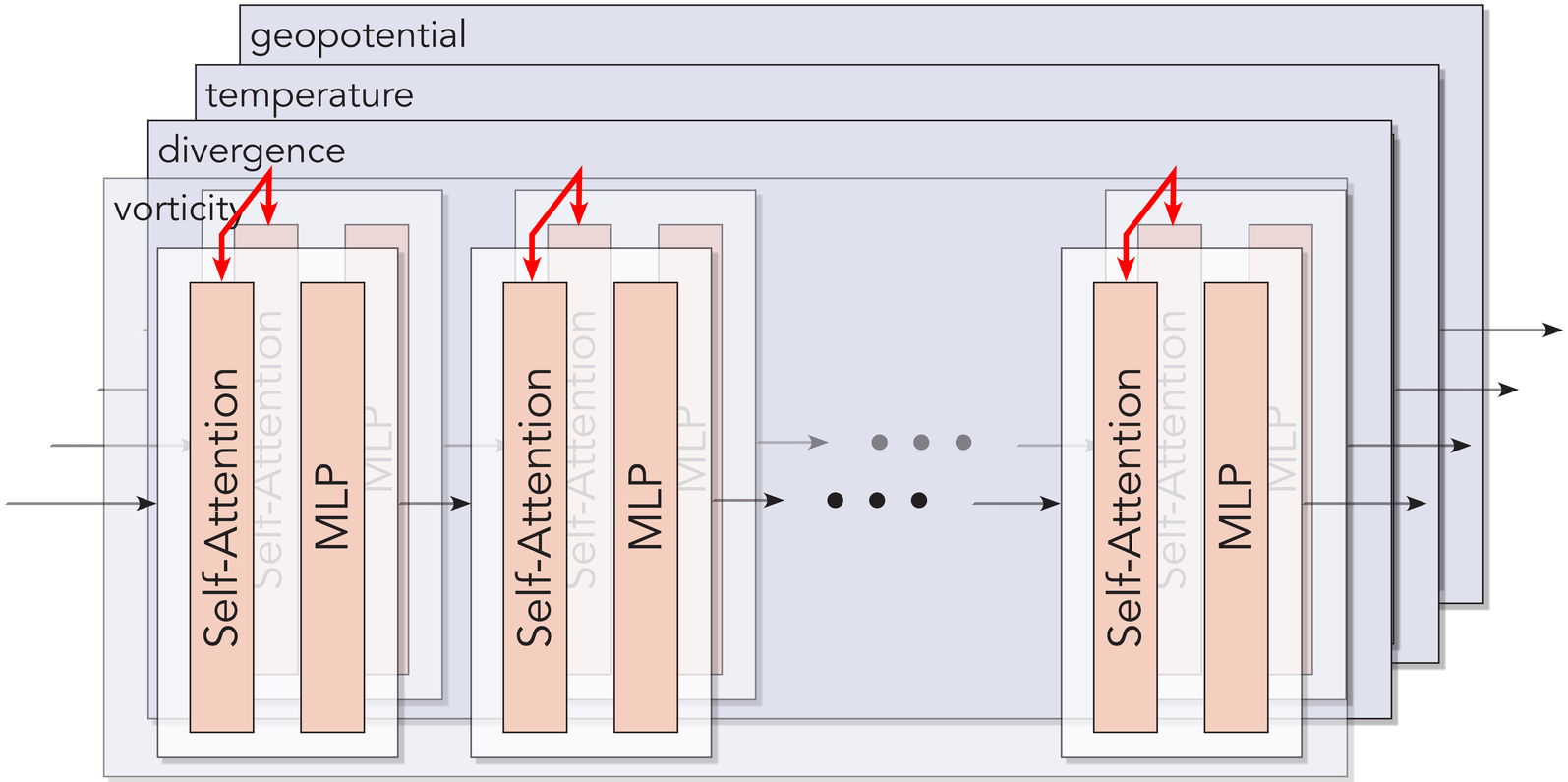
# Multiformer

Self attention

$$\sigma\left(Q\,K^T\right)V$$

Cross attention

$$\sigma\left(Q_\zeta\,K_\mu^T\right)V_\mu$$




© Christian Lessig, 2023

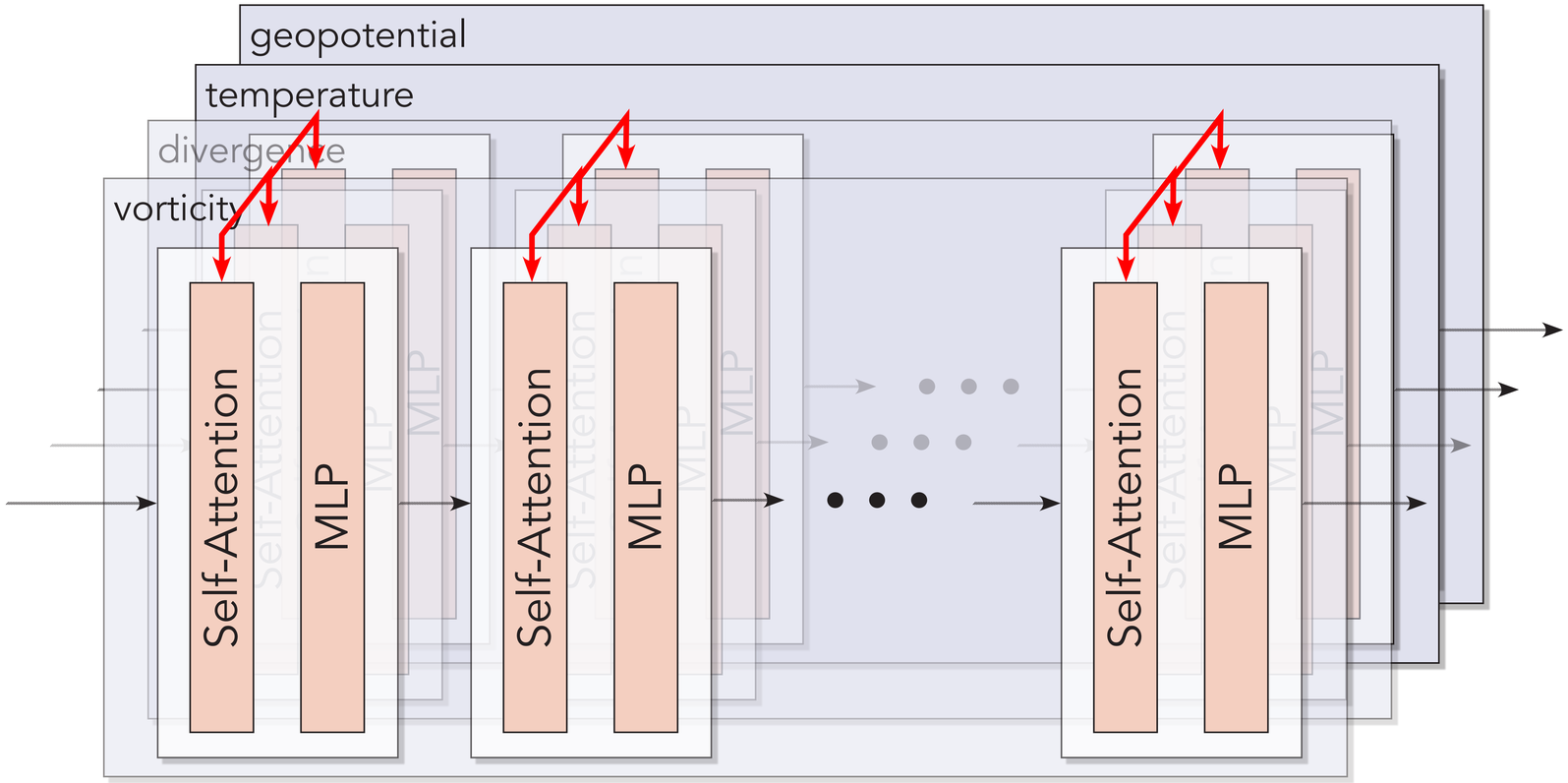
# Multiformer

Self attention

$$\sigma\left(Q\,K^T\right)V$$

Cross attention

$$\sigma\left(Q_\zeta\,K_\mu^T\right)V_\mu$$




© Christian Lessig, 2023

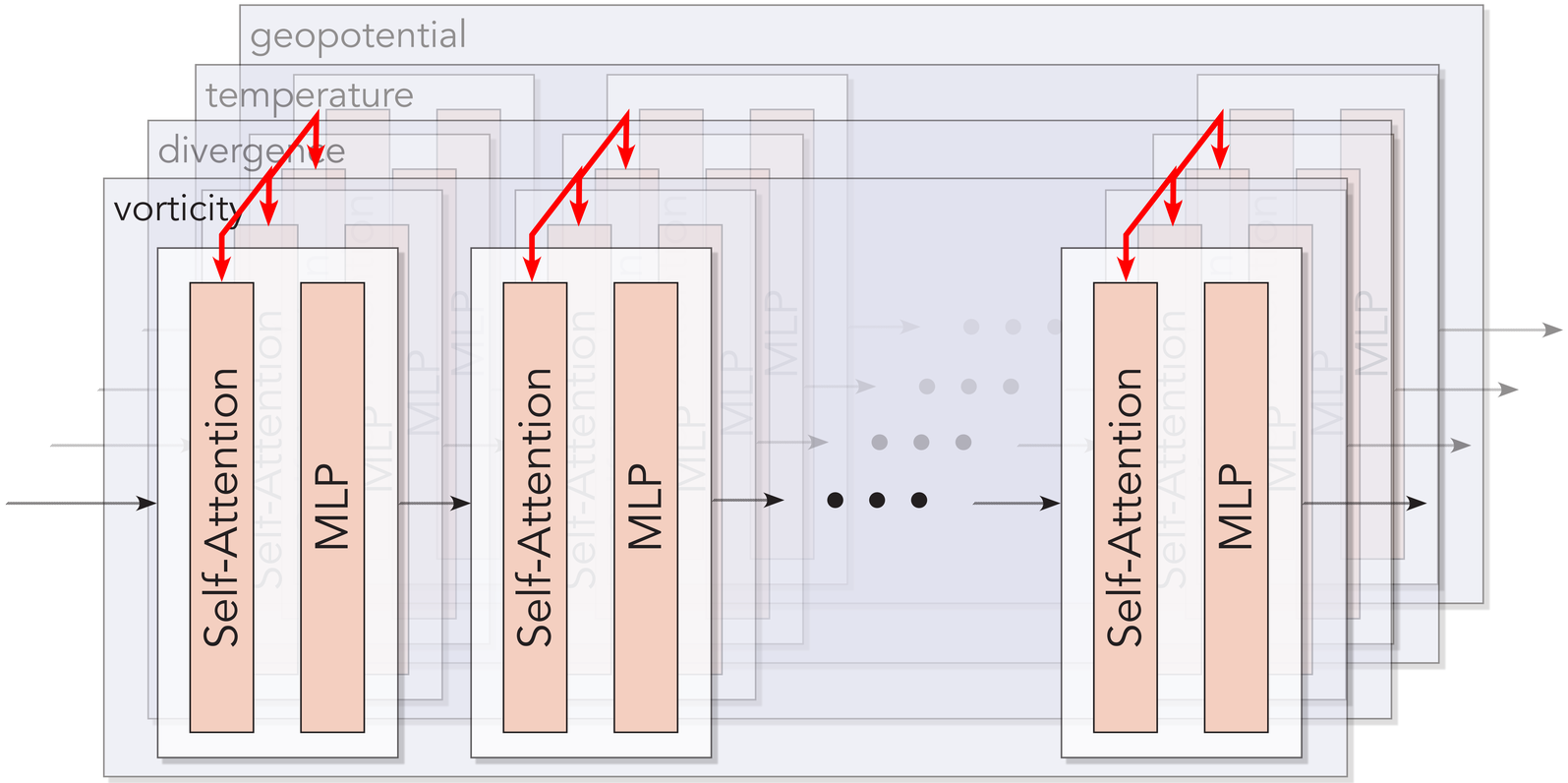
# Multiformer

Self attention

$$\sigma\left(Q\,K^T\right)V$$

Cross attention

$$\sigma\left(Q_\zeta\,K_\mu^T\right)V_\mu$$




© Christian Lessig, 2023

# Multiformer

- Plug-and-play of fields
  - Fields can be added/removed with limited (or no) computational effort

© Christian Lessig, 2023

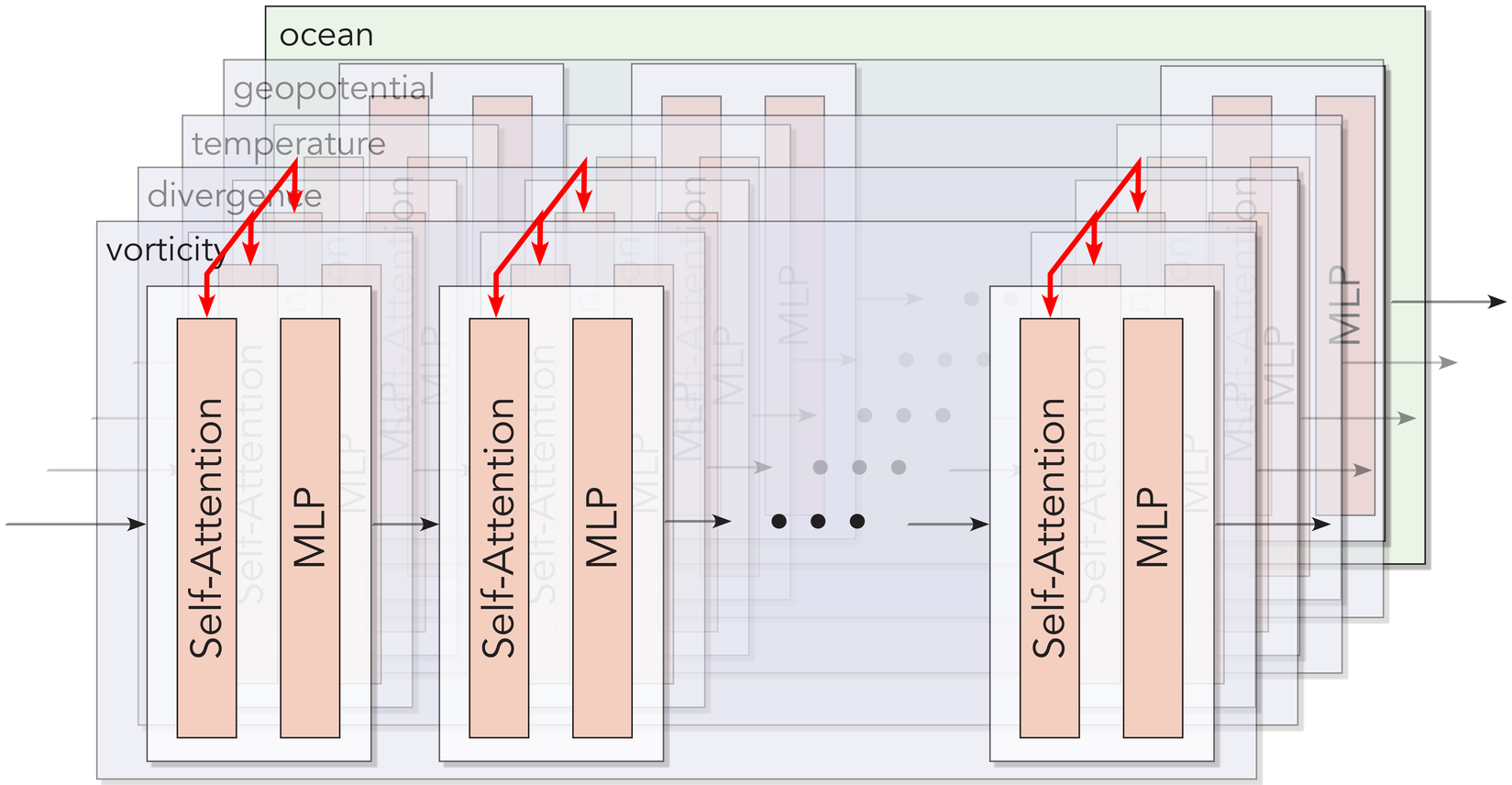
# Multiformer

Self attention

$$\sigma\left(Q\,K^T\right)V$$

Cross attention

$$\sigma\left(Q_\zeta\,K_\mu^T\right)V_\mu$$




© Christian Lessig, 2023

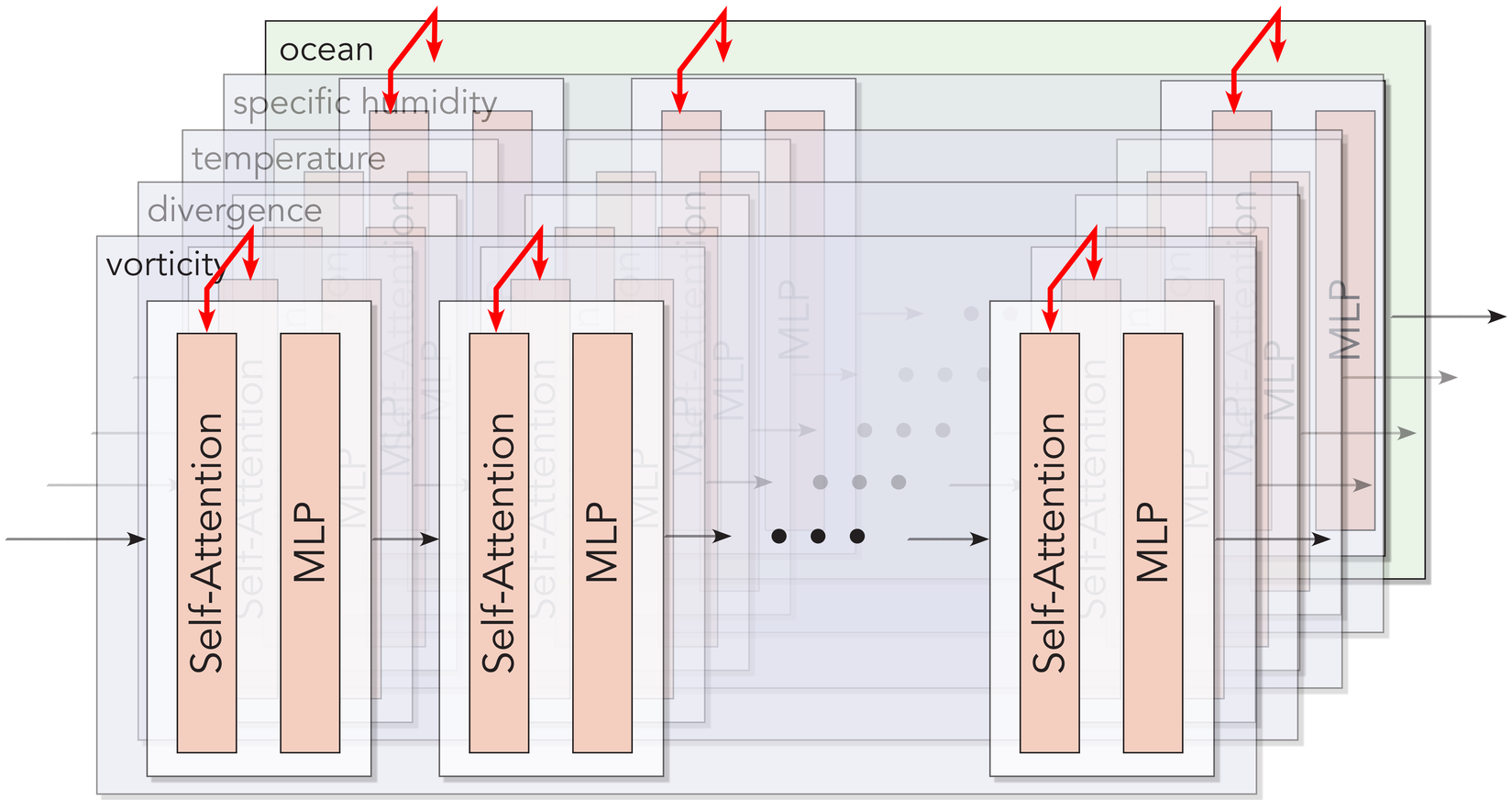
# Multiformer

Self attention

$$\sigma\left(Q\,K^T\right)V$$

Cross attention

$$\sigma\left(Q_\zeta\,K_\mu^T\right)V_\mu$$



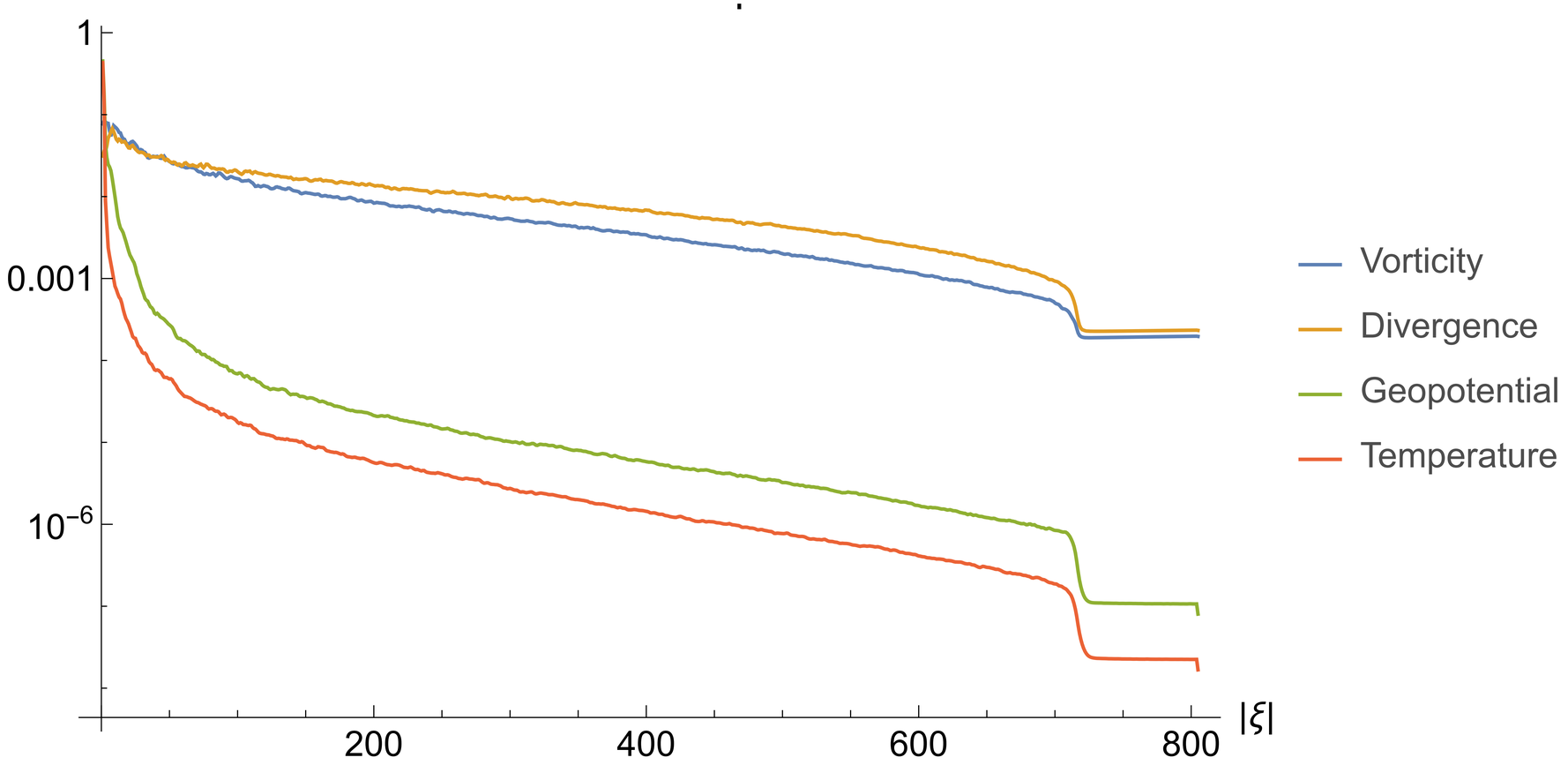

© Christian Lessig, 2023

# Multiformer

- Plug-and-play of fields
  - Fields can be added/removed with very limited computational effort
- Cross-attention allows for explicit introspection of interaction between fields
- Different physical fields with different properties have separate latent spaces (and transformations for these)

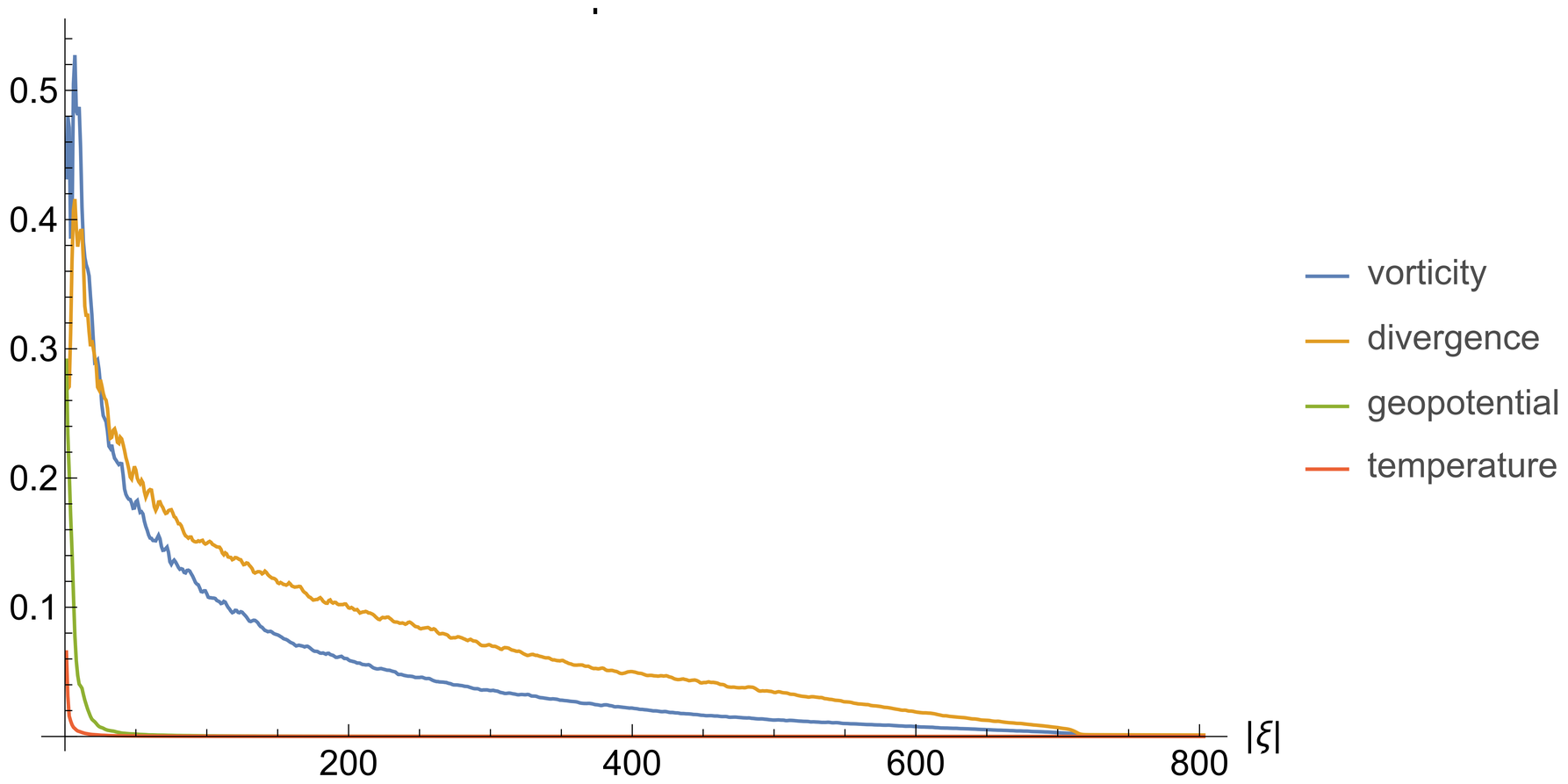
© Christian Lessig, 2023

# AtmoRep data



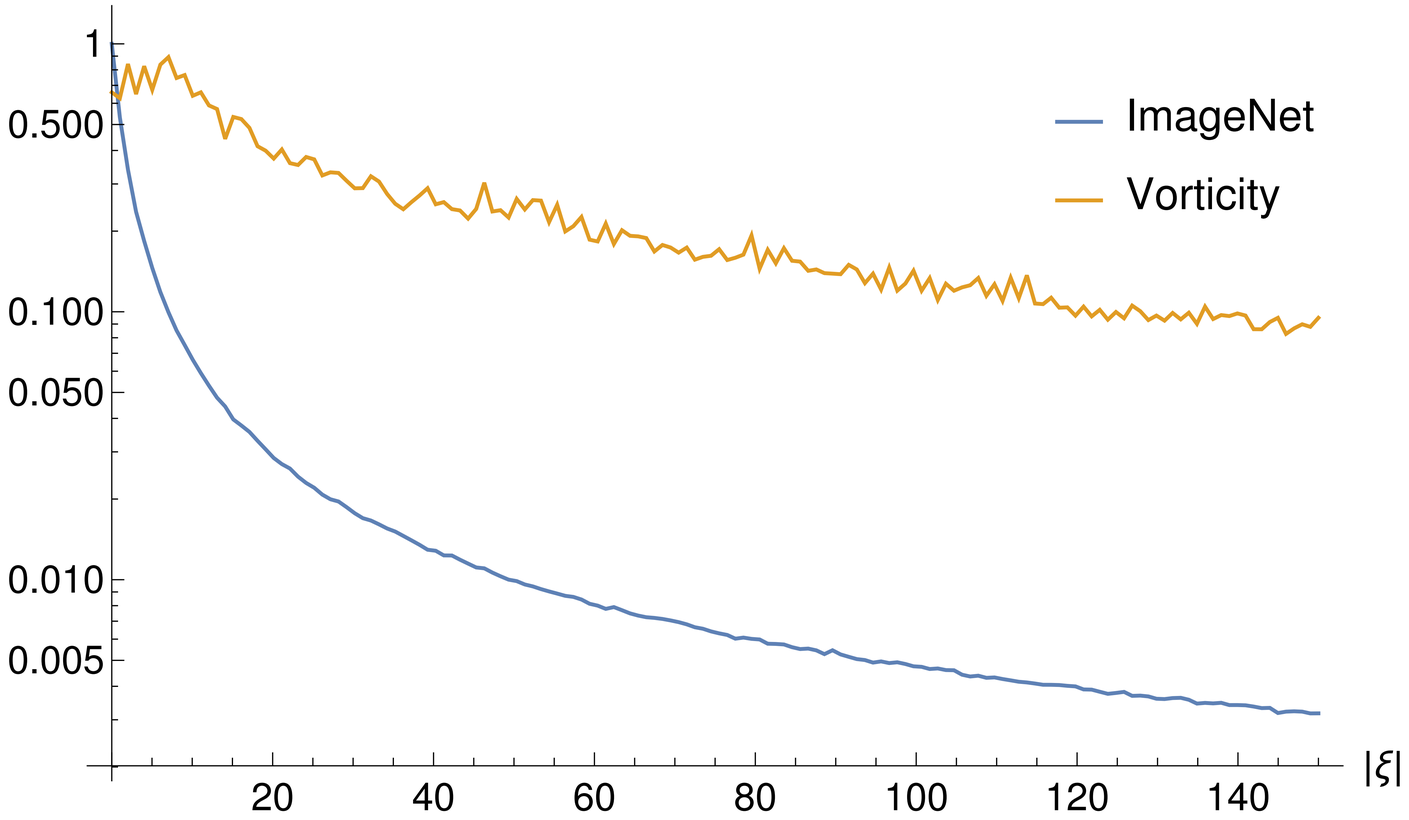

© Christian Lessig, 2023

# AtmoRep data




© Christian Lessig, 2023

# ERA5 versus ImageNet




© Christian Lessig, 2023

# ERA5 versus ImageNet

stream function
velocity potential

$\Leftrightarrow$

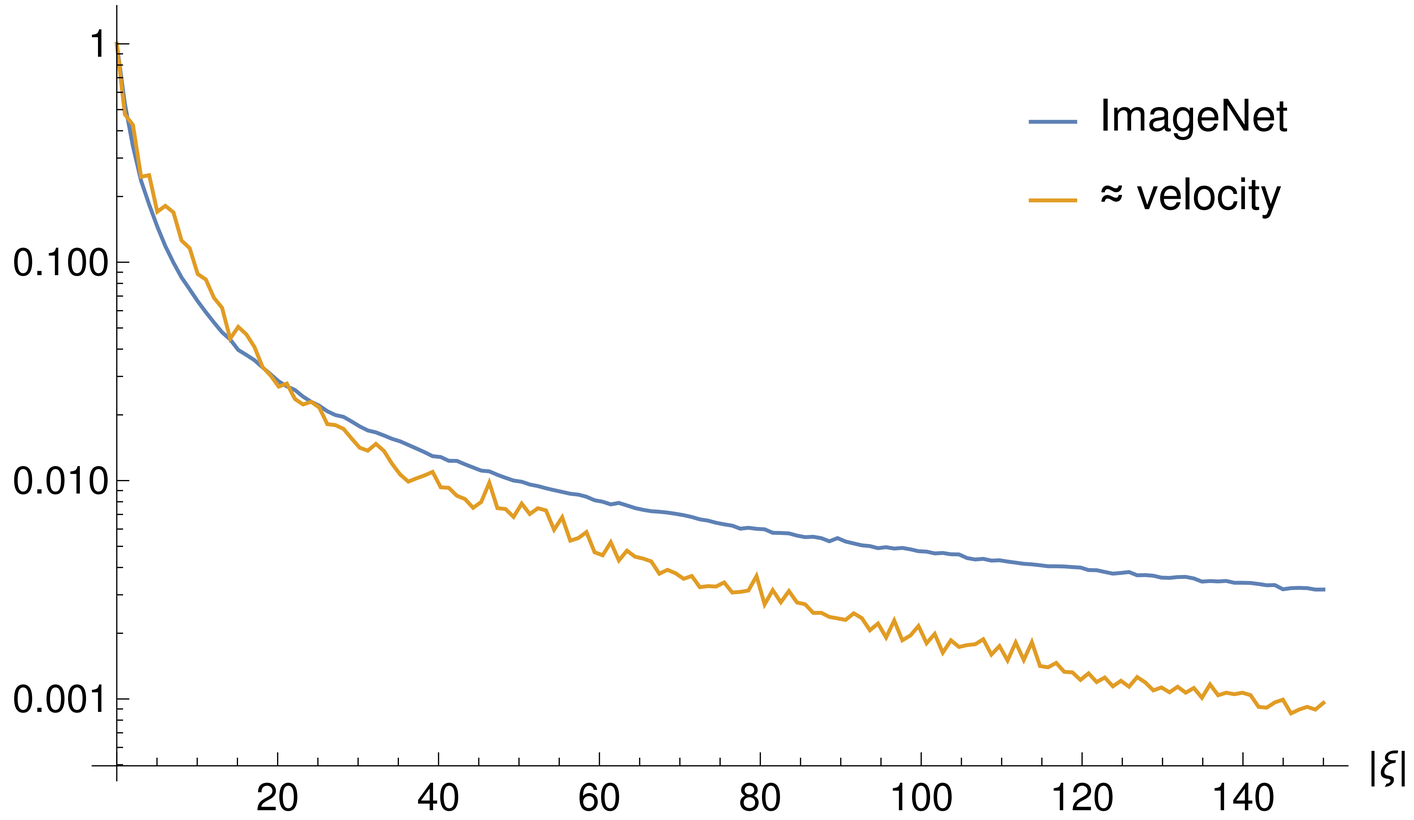
velocity
vector field

$\Leftrightarrow$

vorticity
divergence

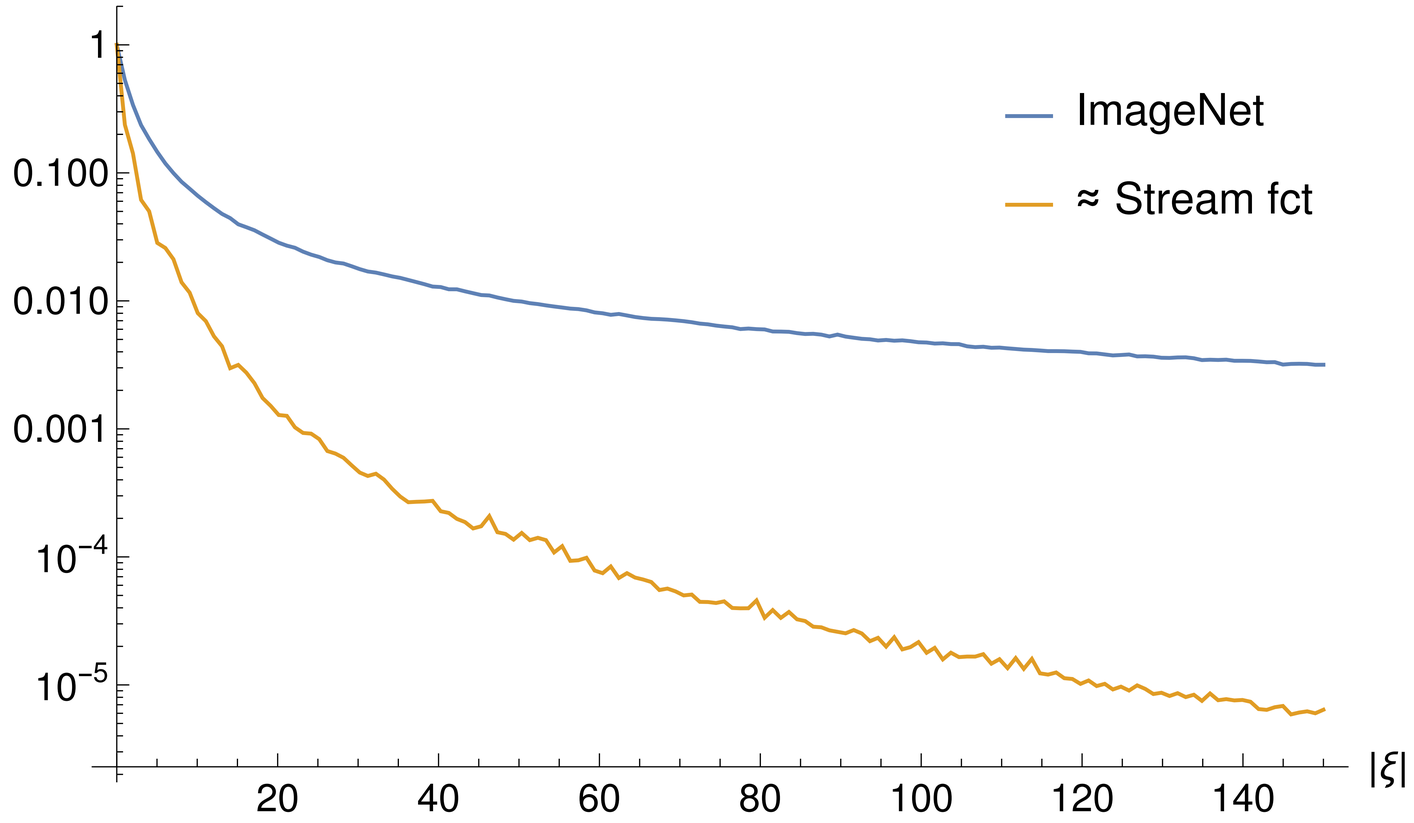
© Christian Lessig, 2023

# ERA5 versus ImageNet



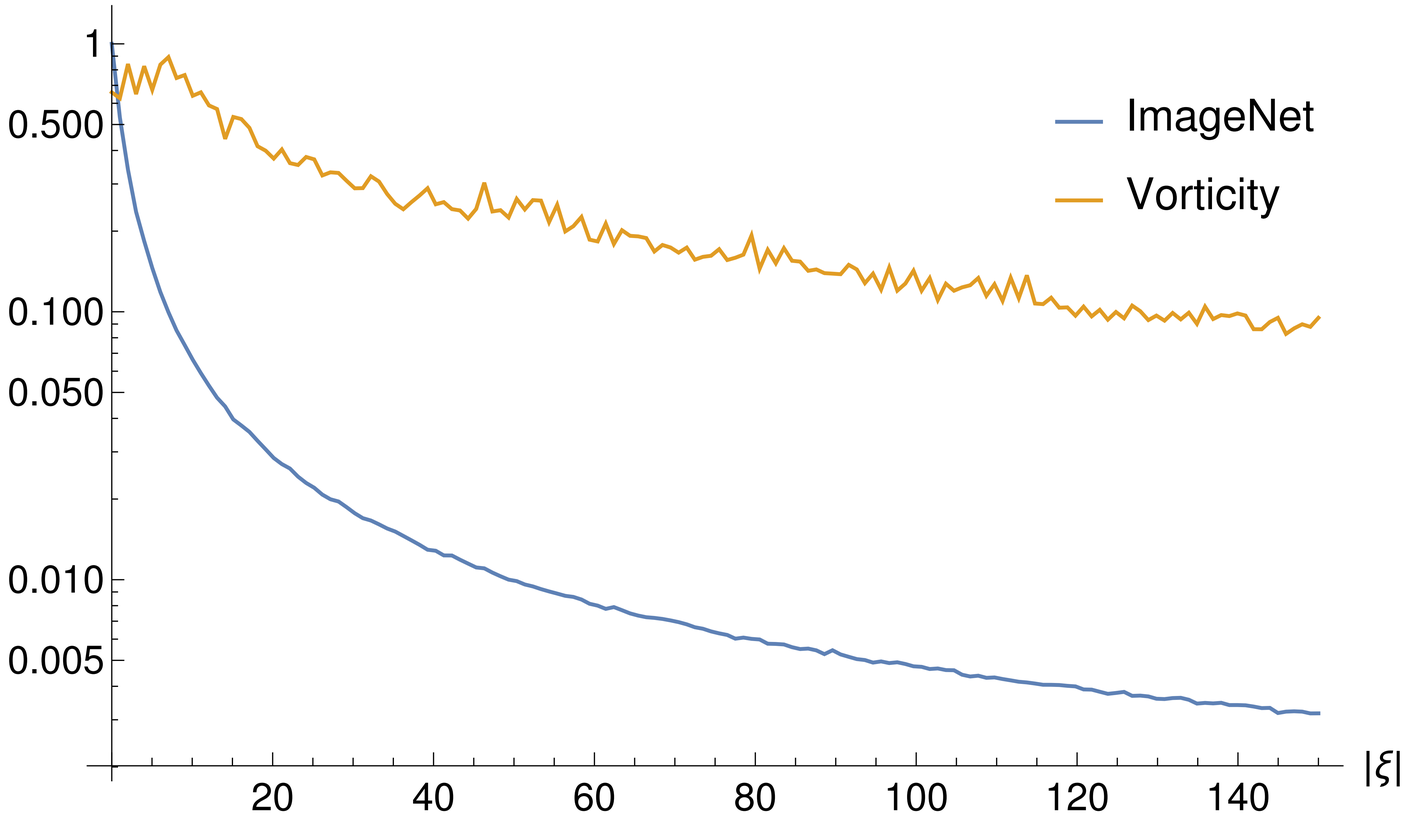

© Christian Lessig, 2023

# ERA5 versus ImageNet




© Christian Lessig, 2023

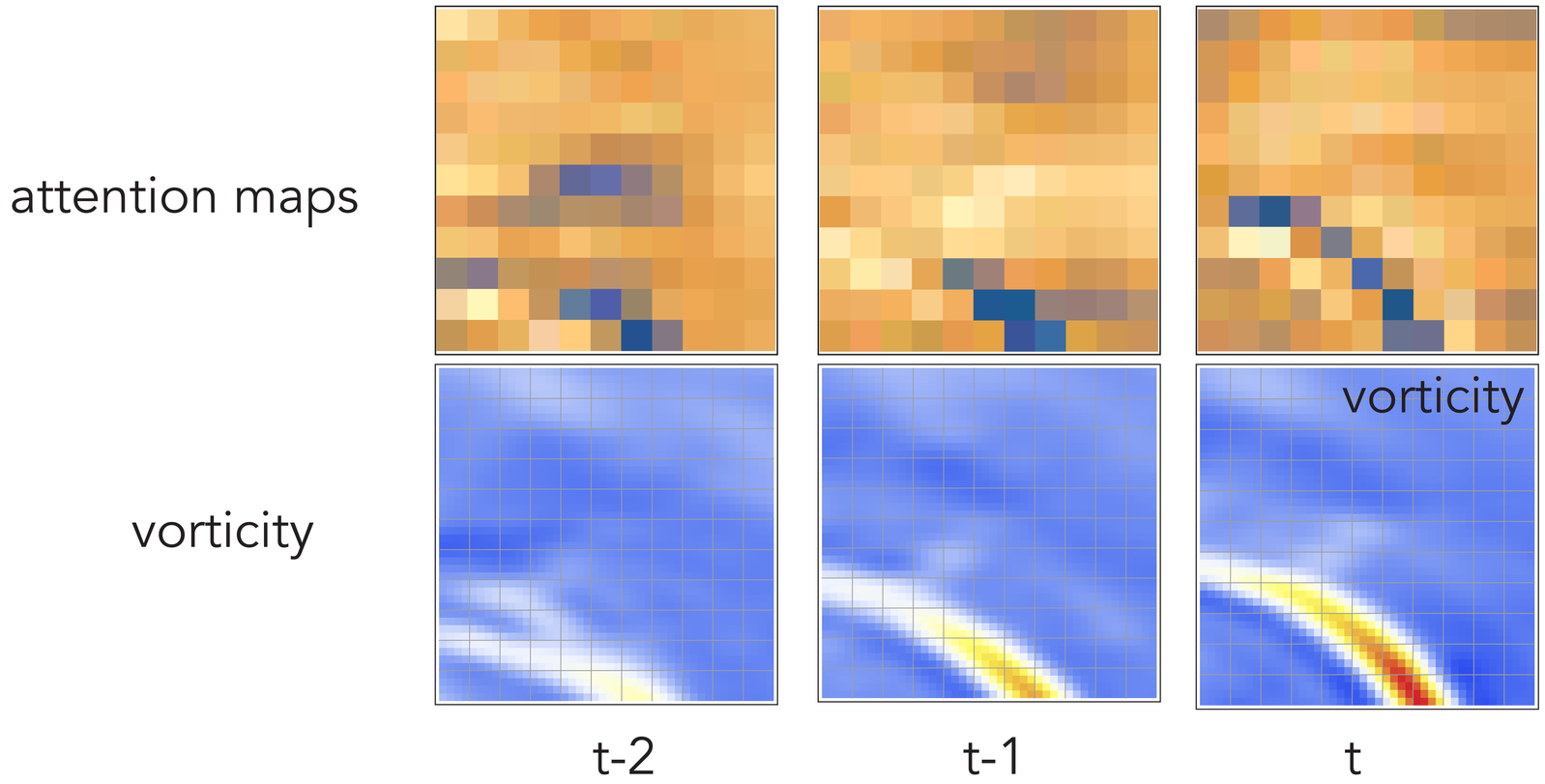
# AtmoRep data




© Christian Lessig, 2023

# Statistical loss

- Attention maps:




© Christian Lessig, 2023