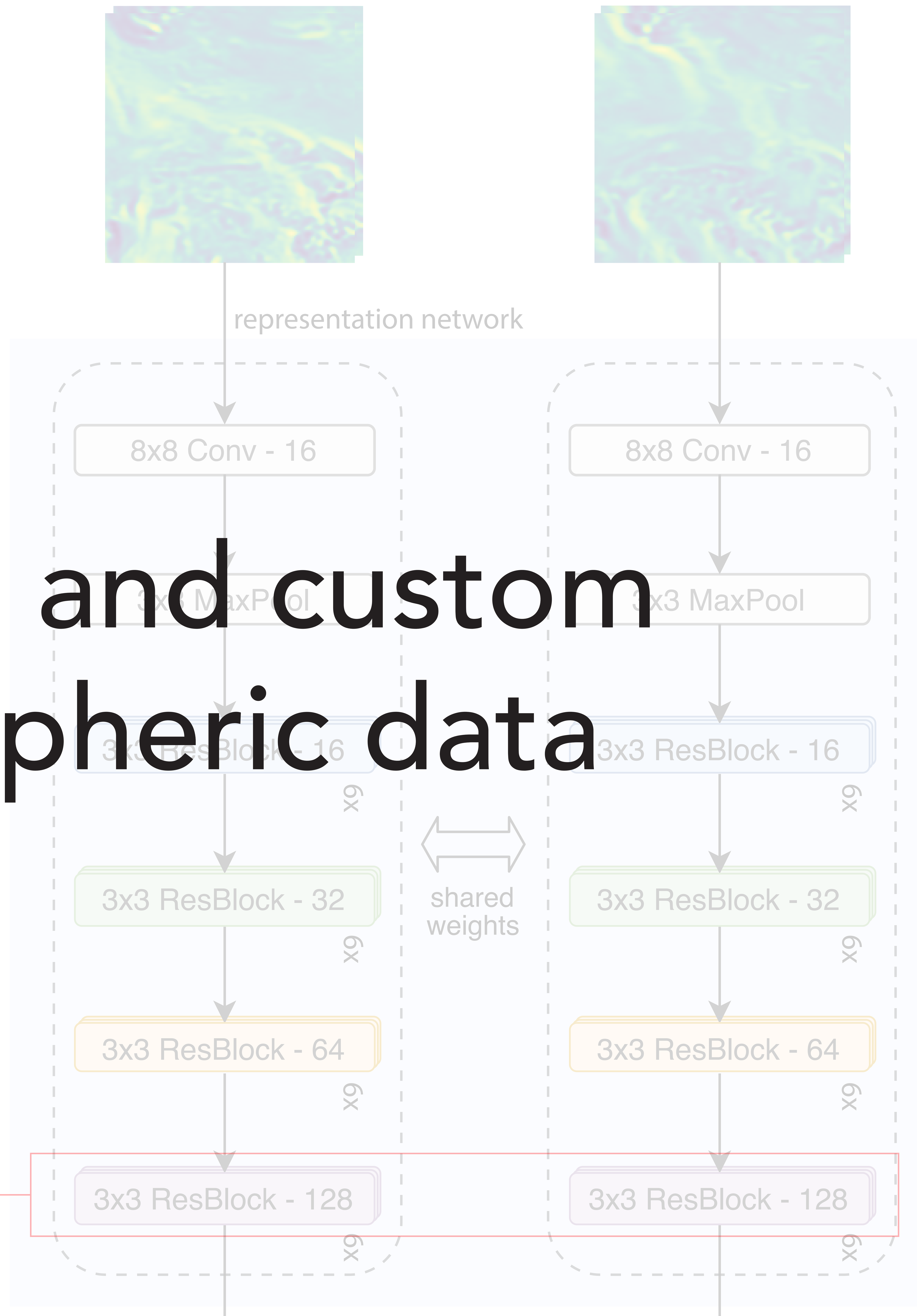
Representation learning and custom loss functions for atmospheric data

Christian Lessig (joint work with Sebastian Hoffmann)

Otto-von-Guericke-Universität Magdeburg

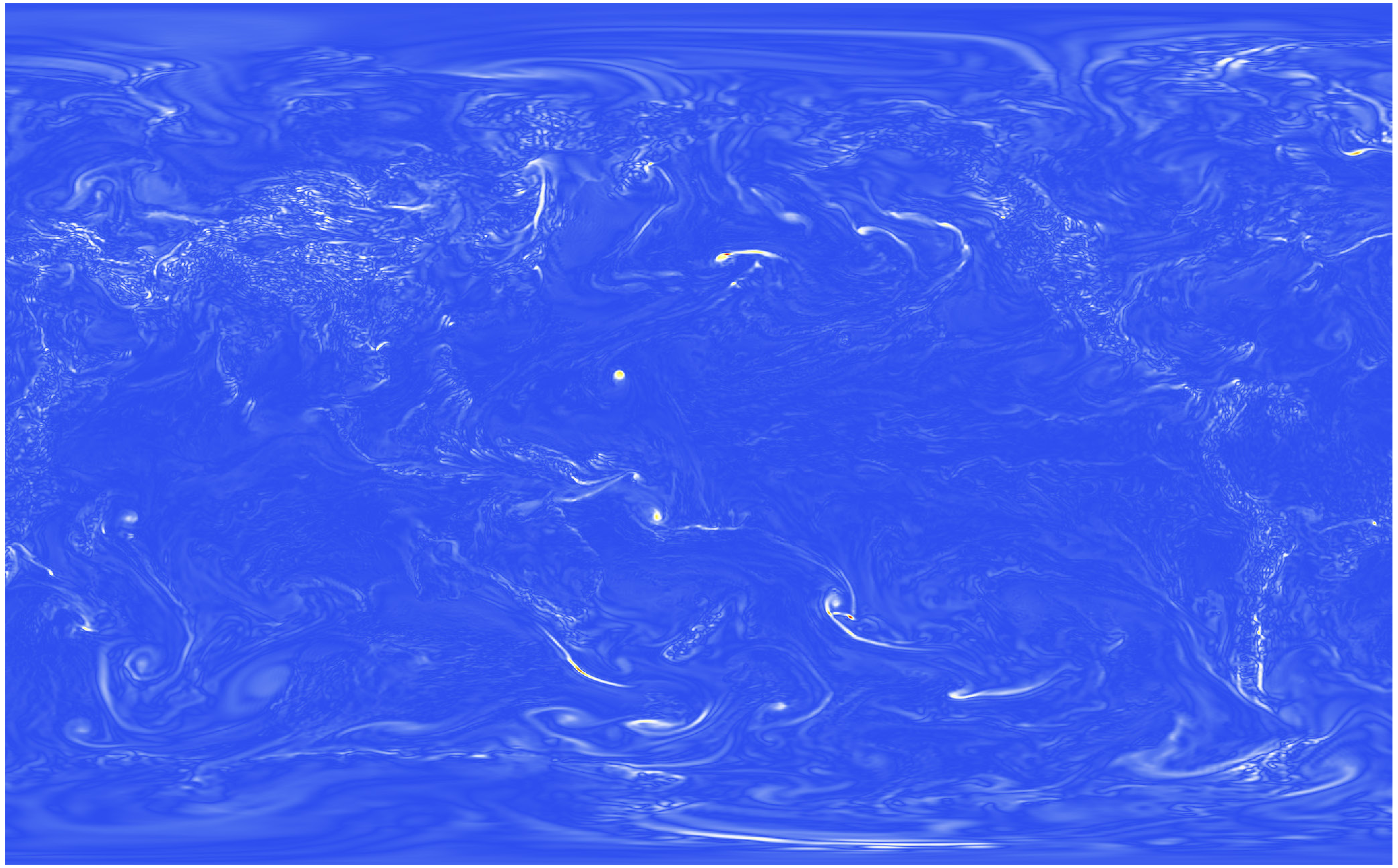


$$\mathcal{L} = \sum_{j=1}^{N} \| h_j - \tilde{h}_j \|$$

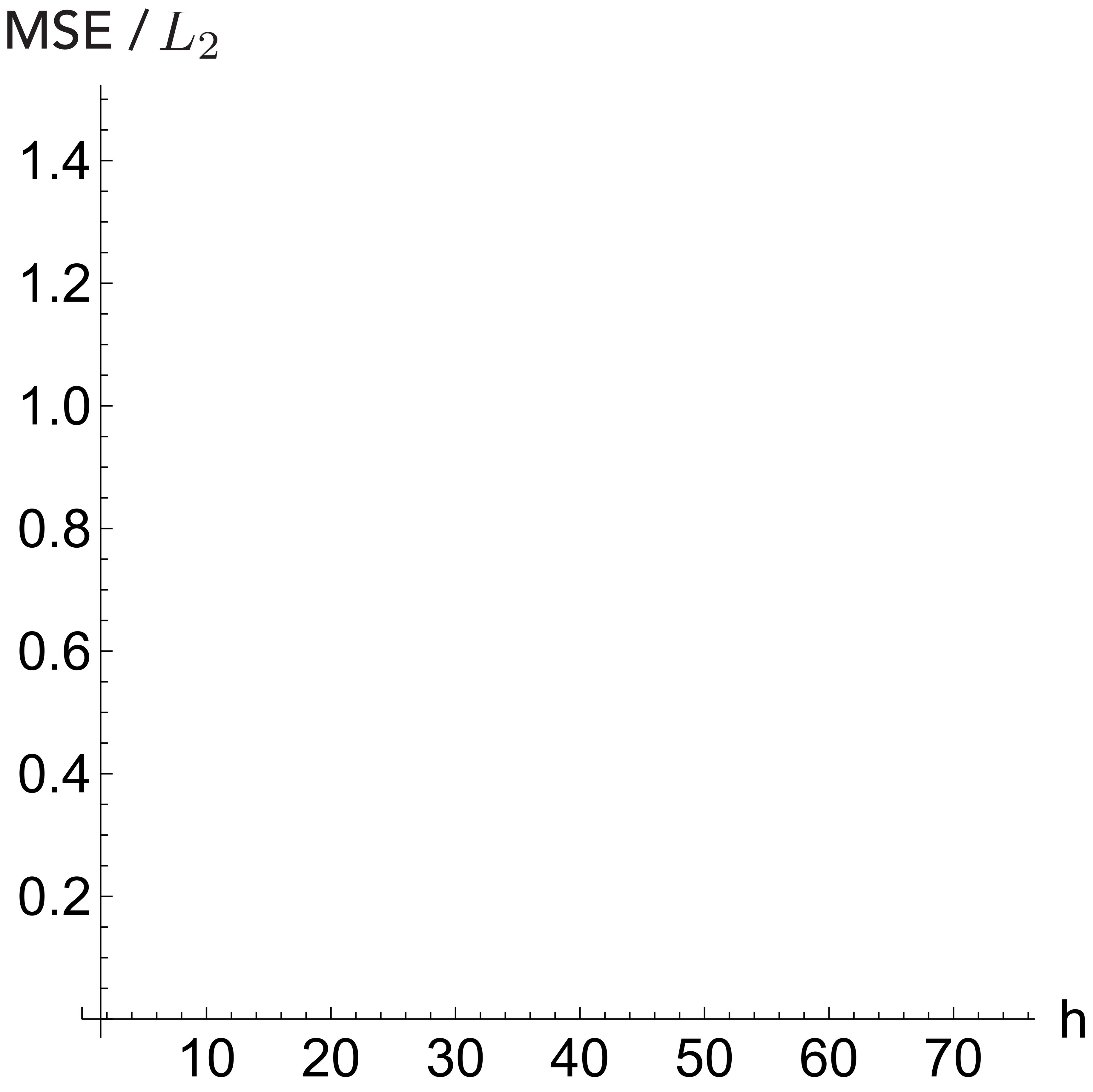
Motivation

- Learning relies on informative and discriminative loss functions

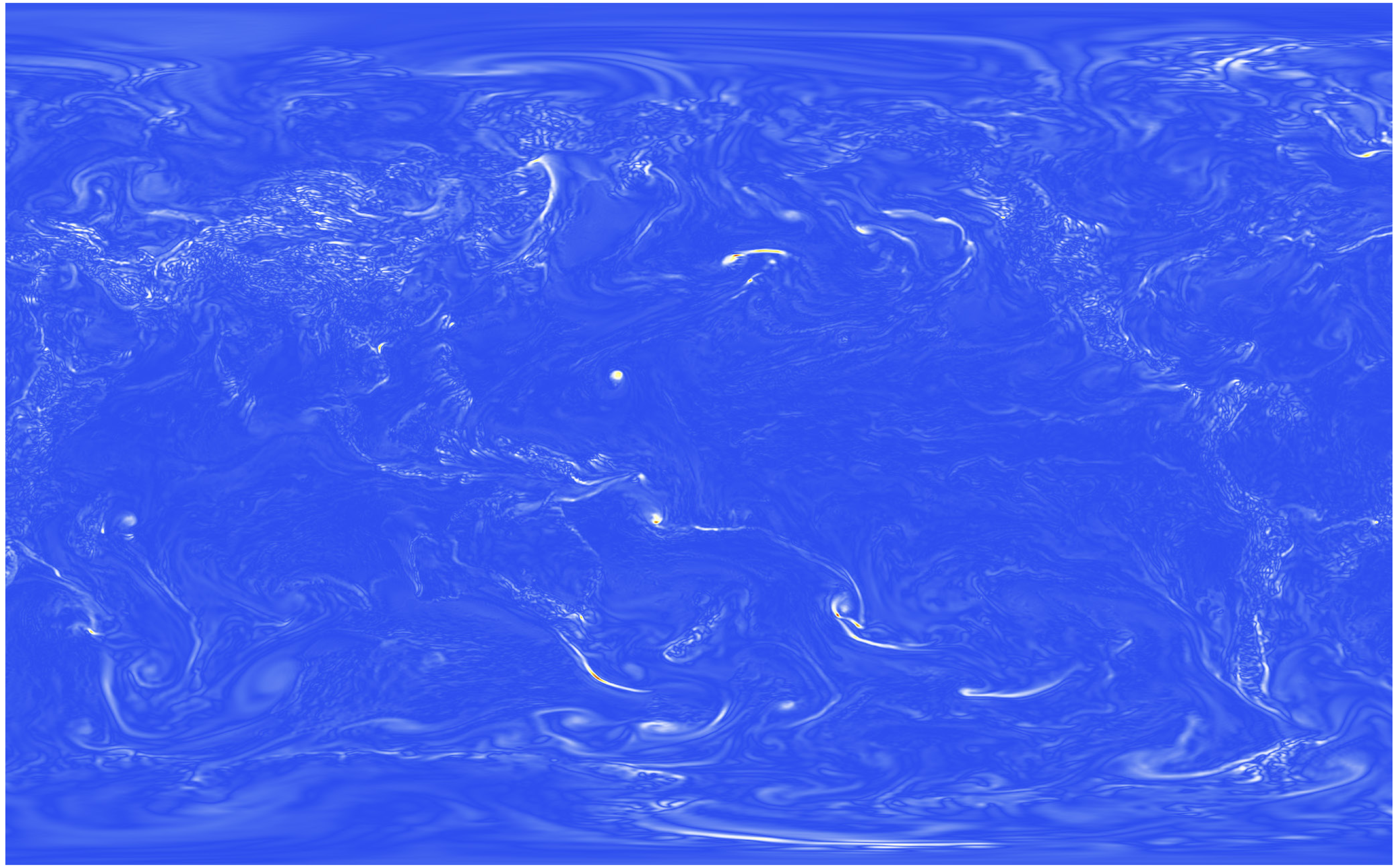
Motivation



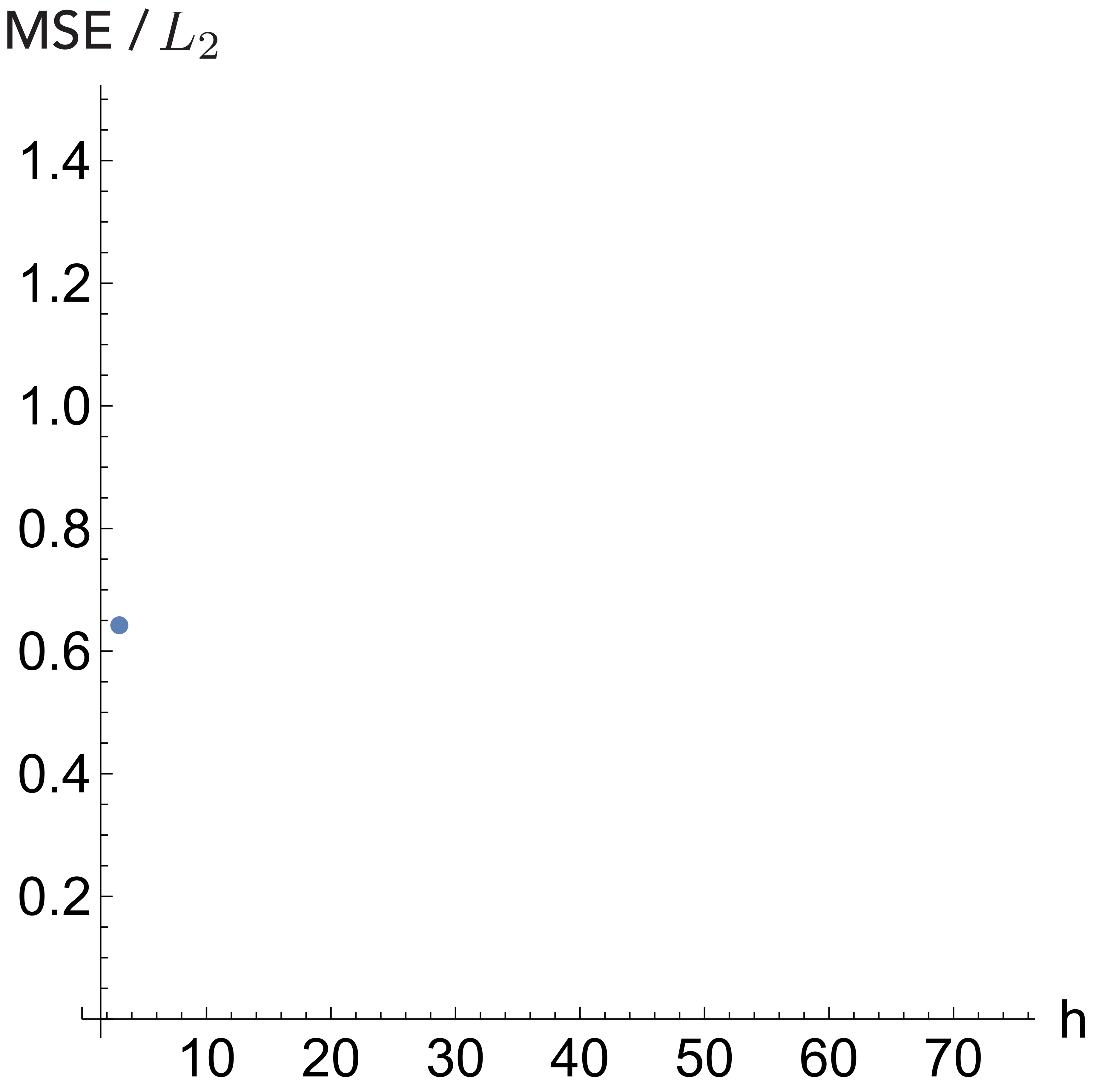
vorticity



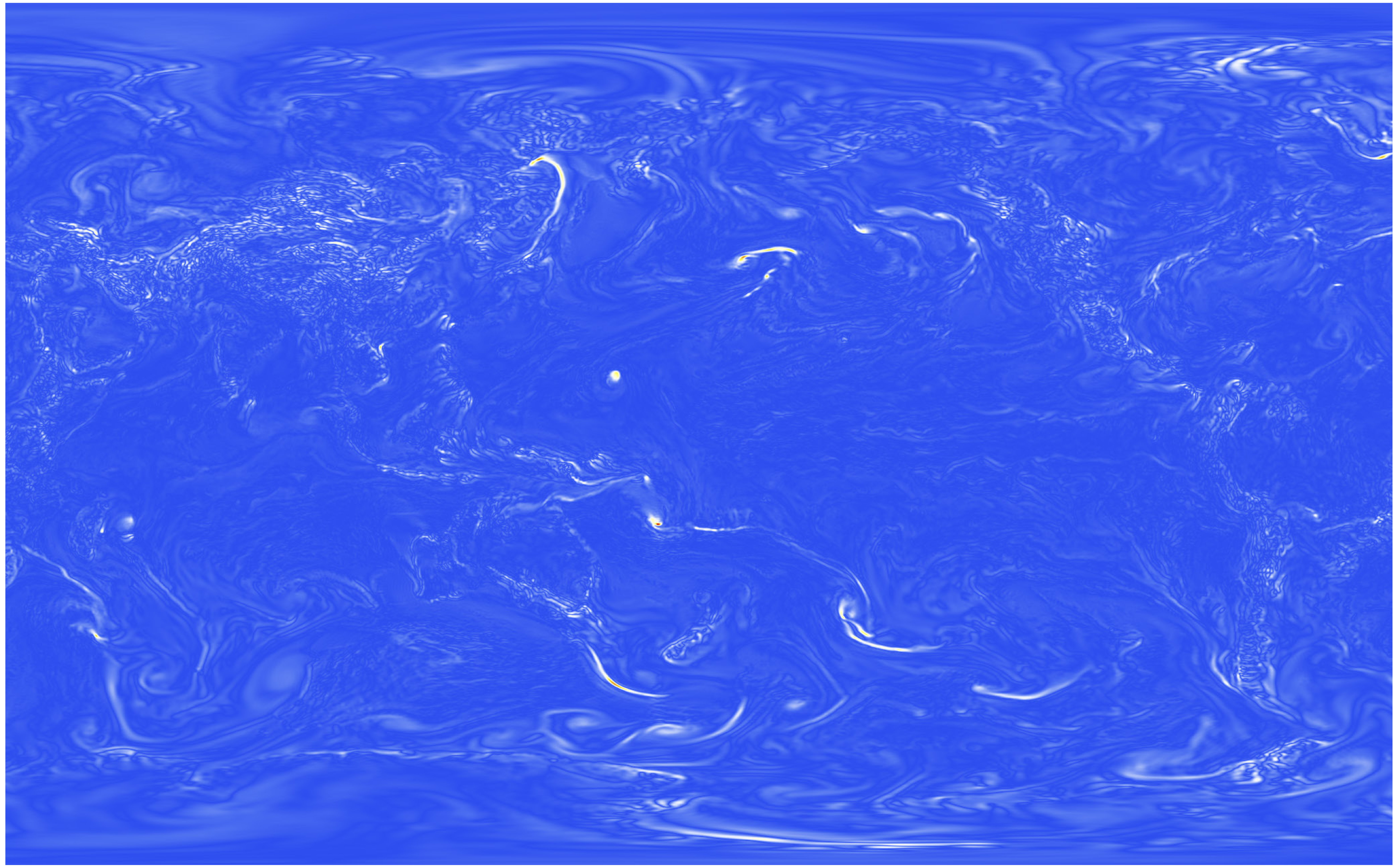
Motivation



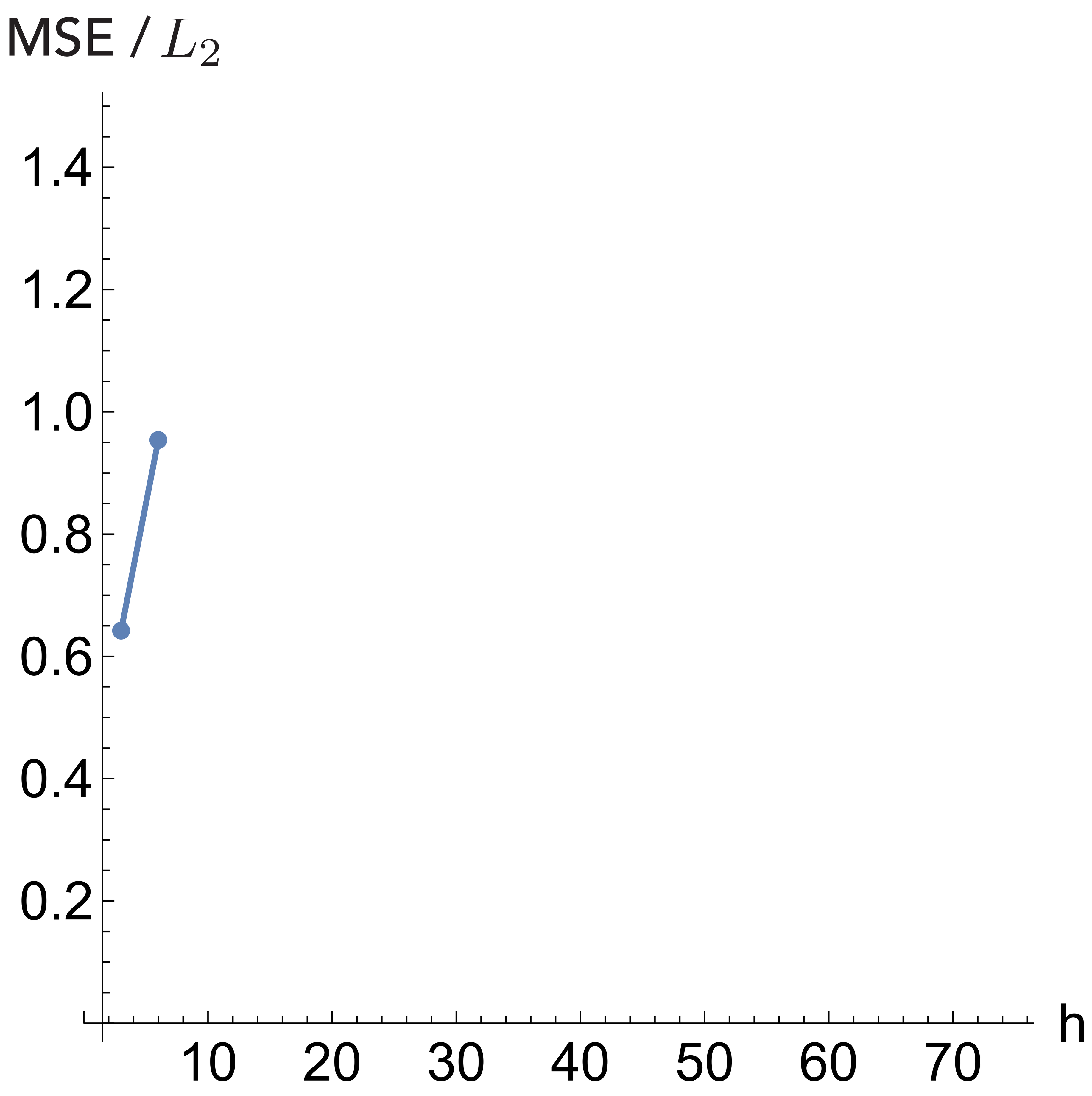
vorticity



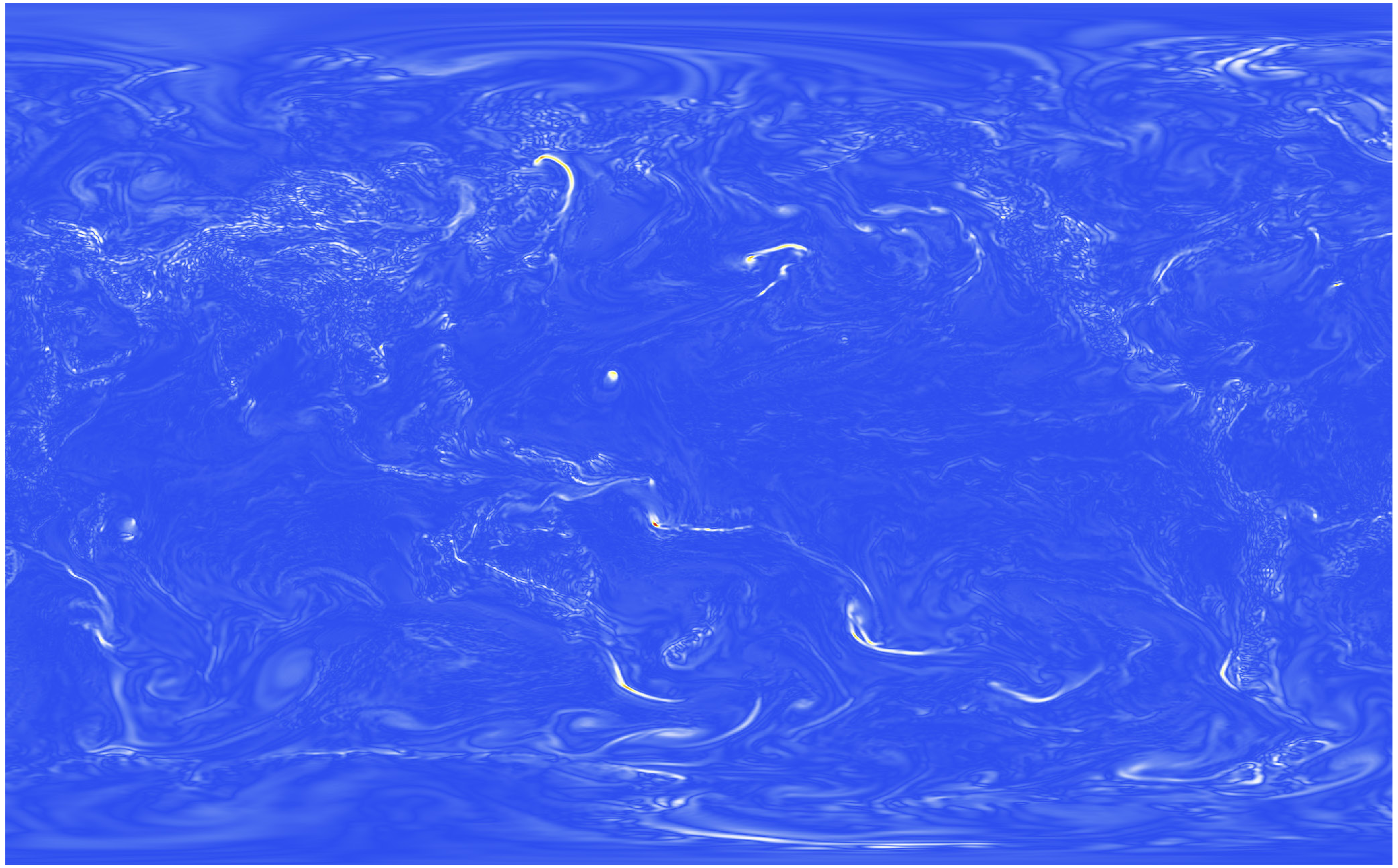
Motivation



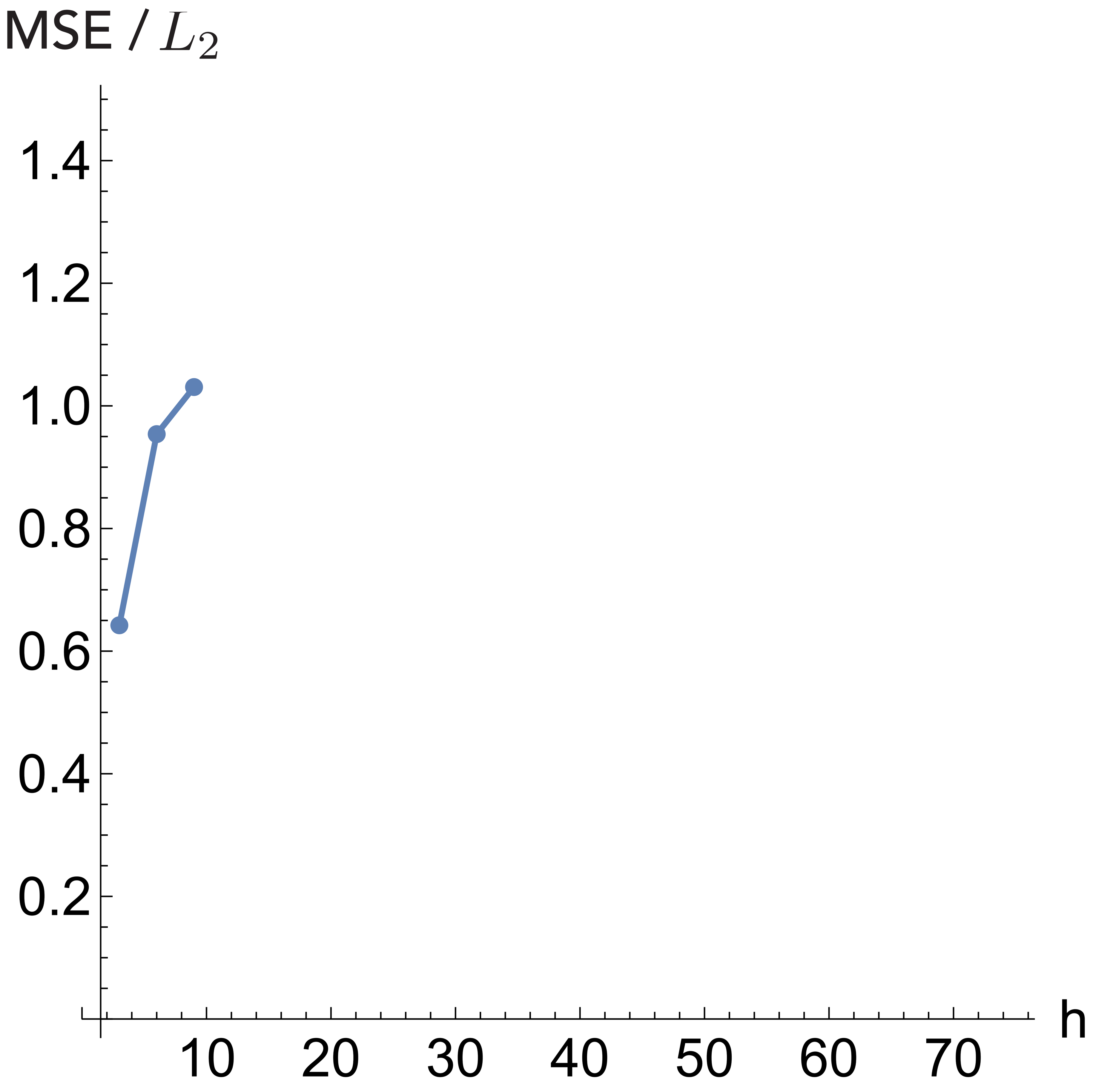
vorticity



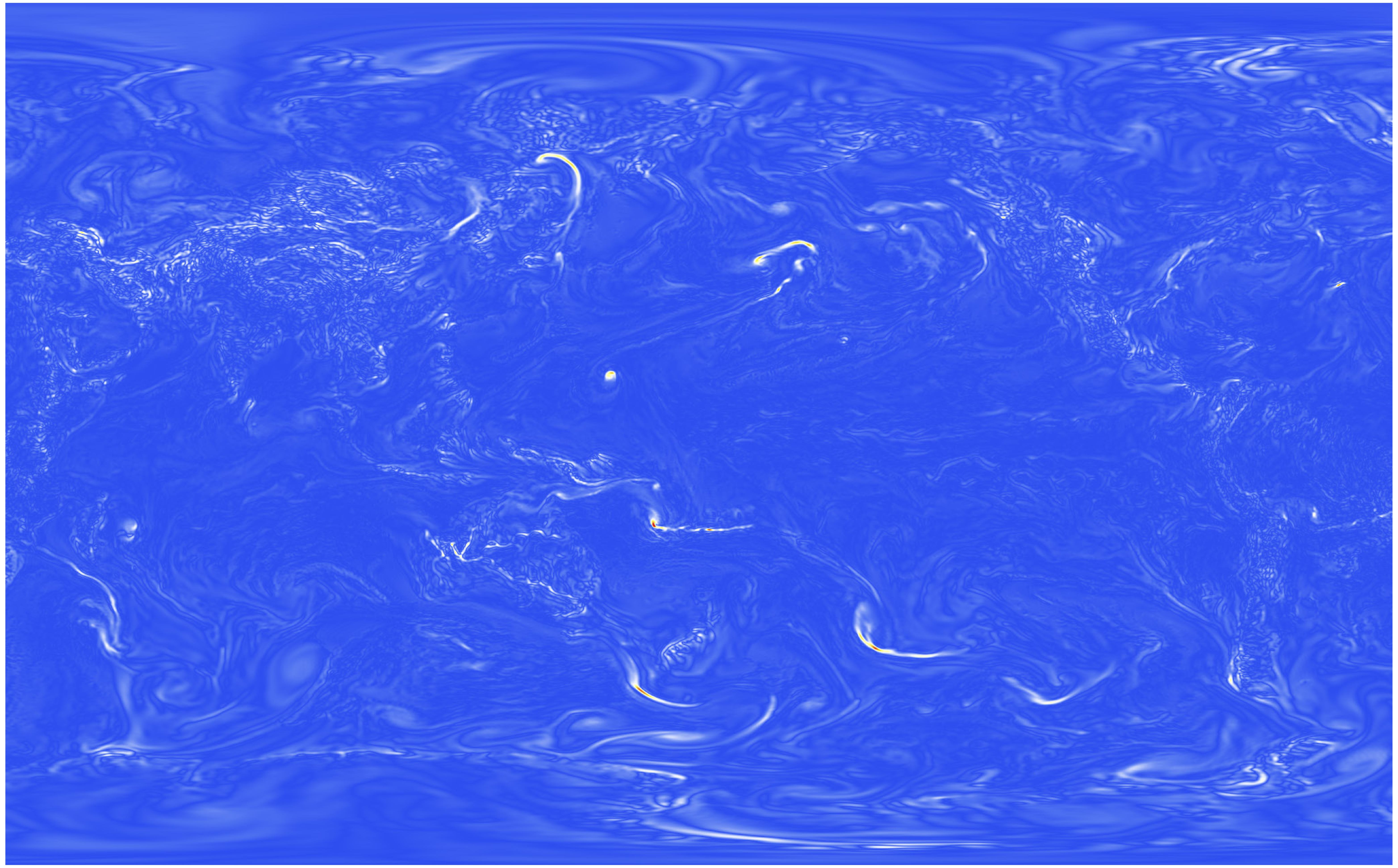
Motivation



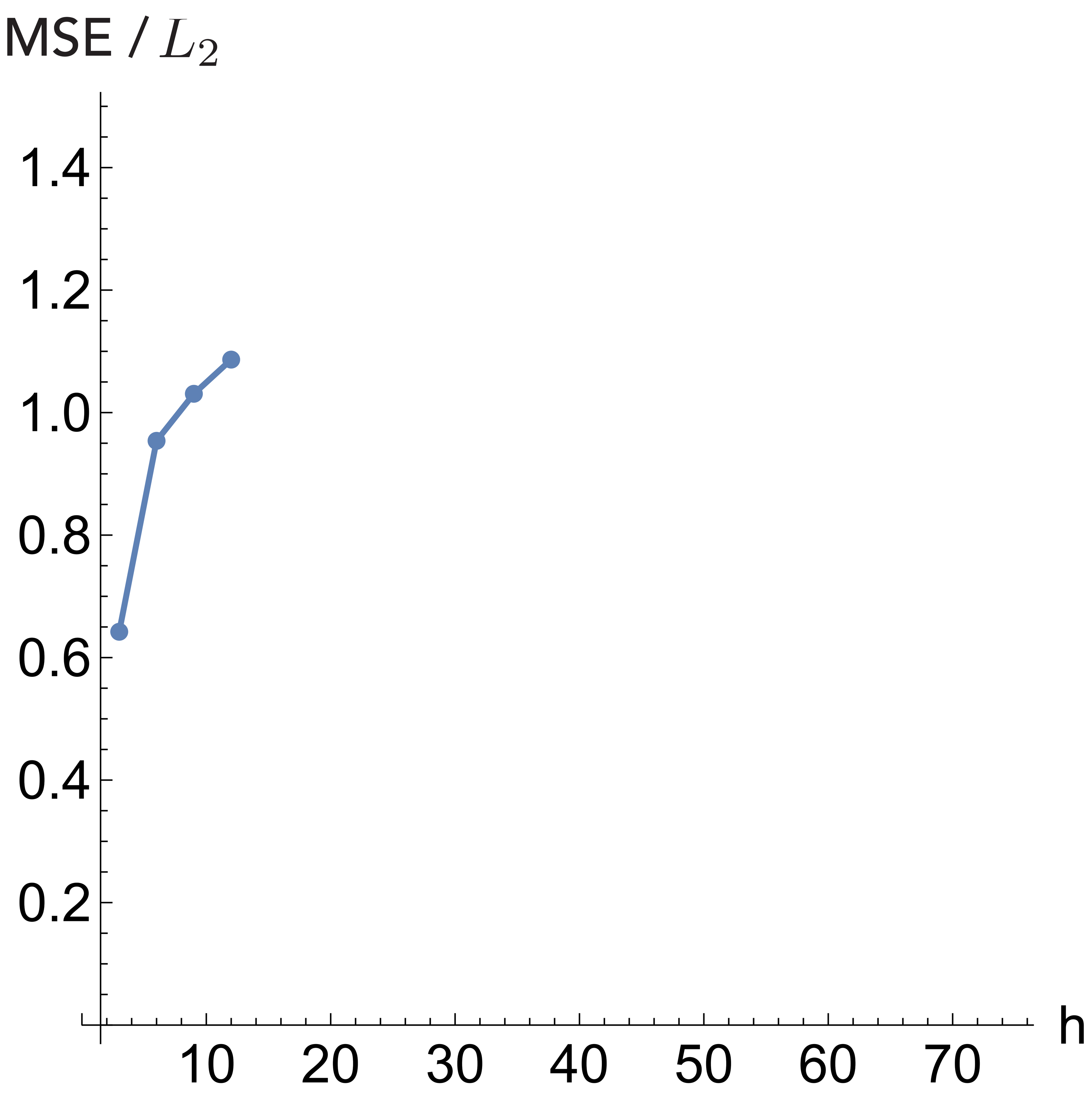
vorticity



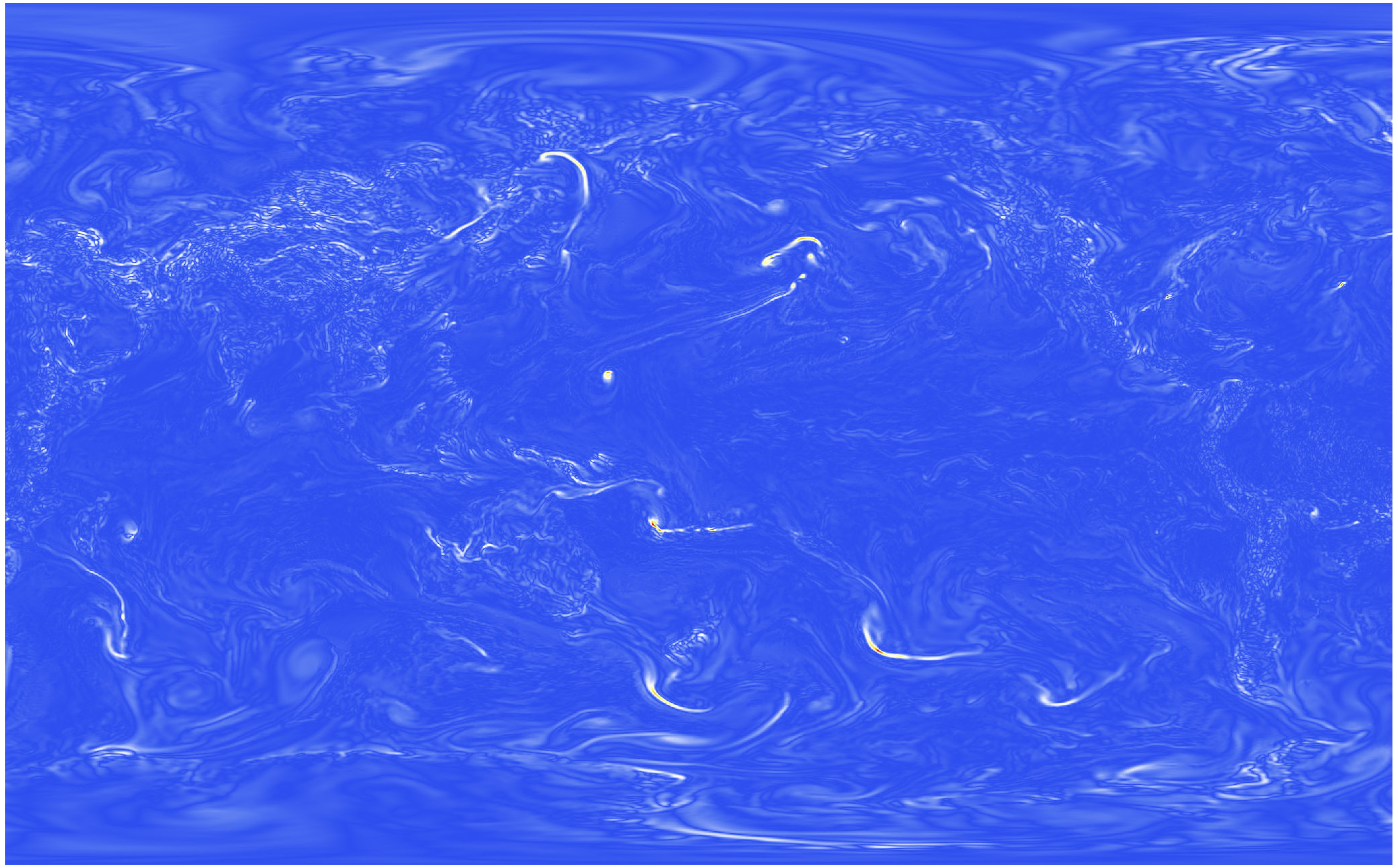
Motivation



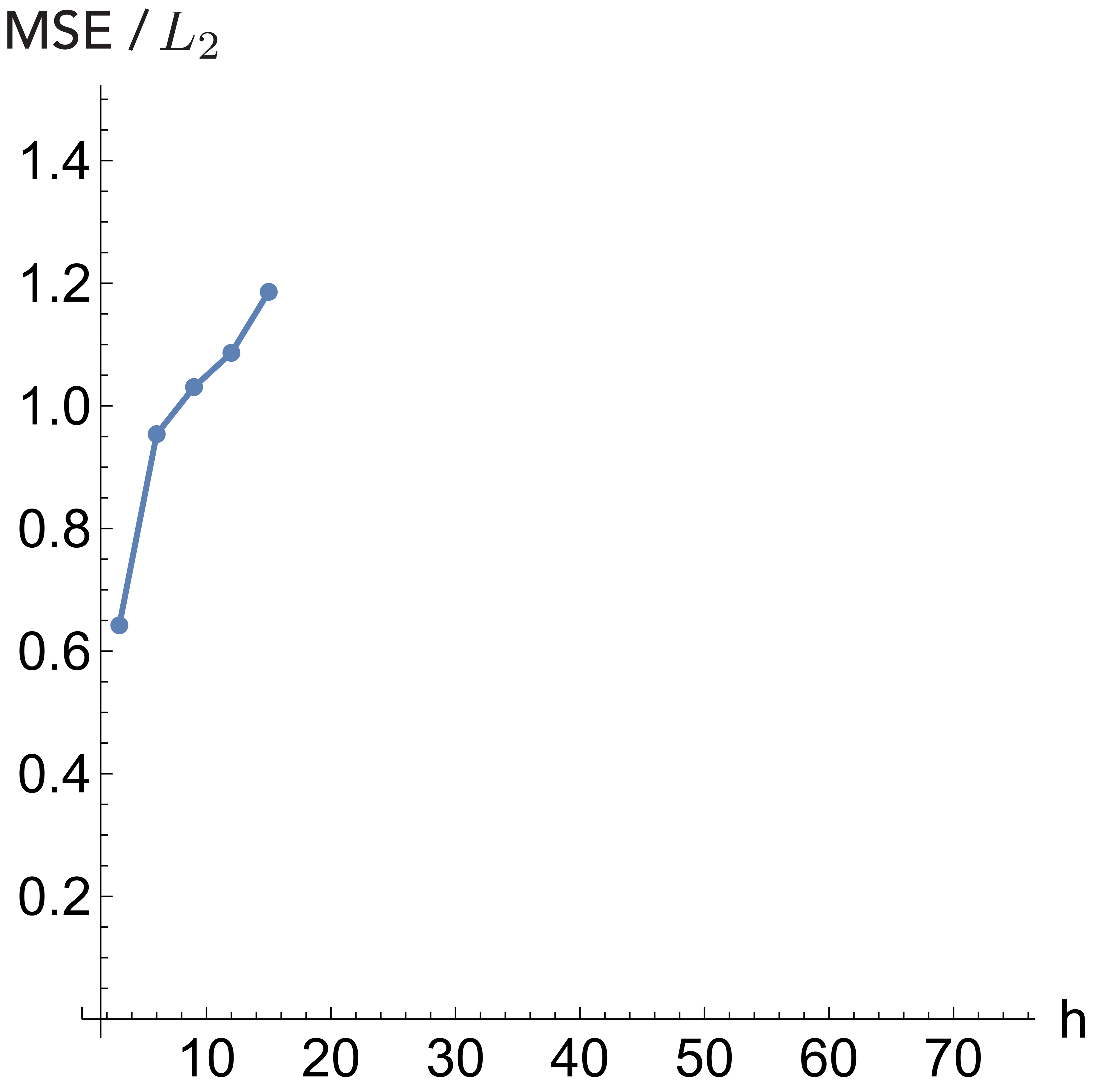
vorticity



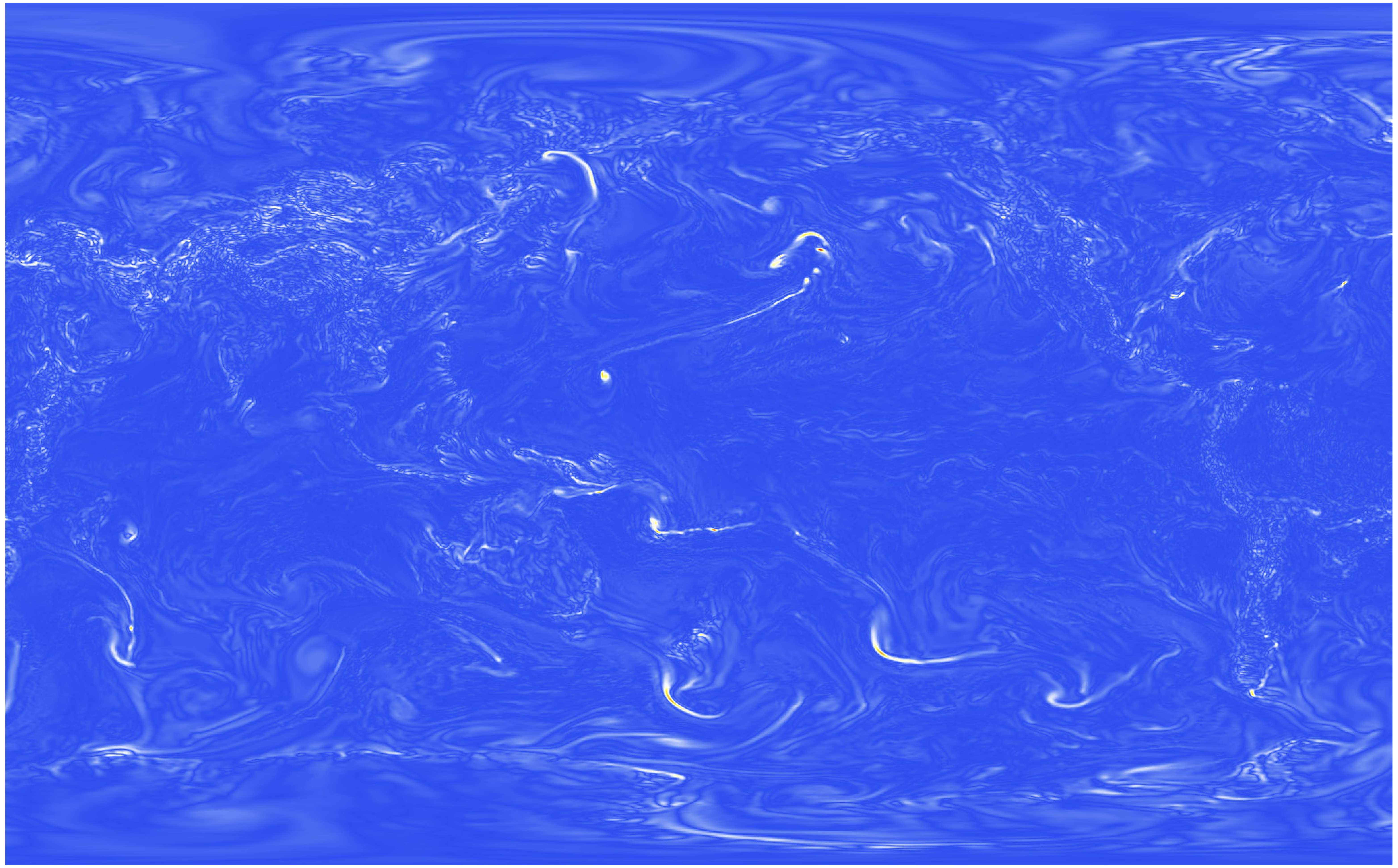
Motivation



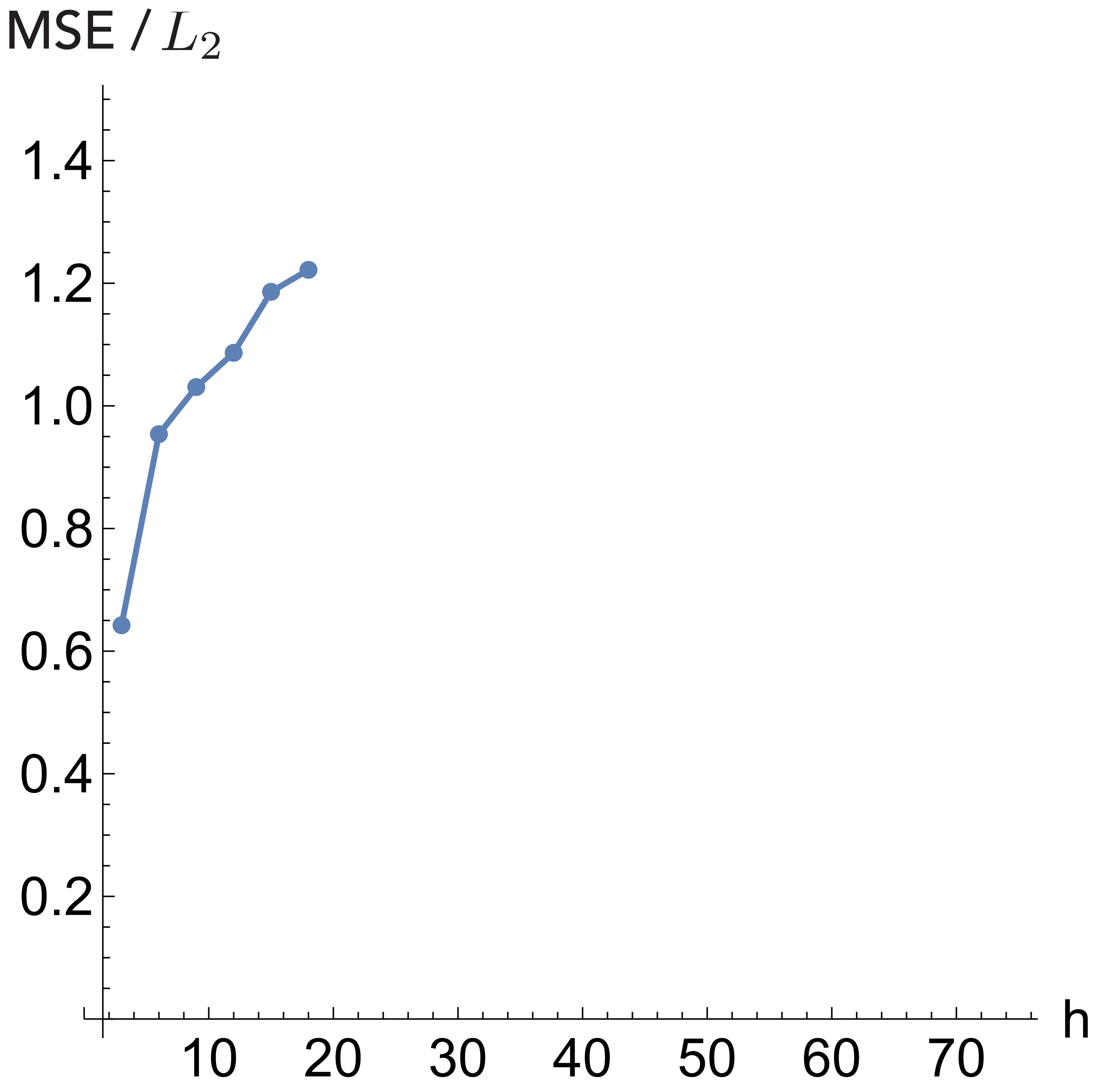
vorticity



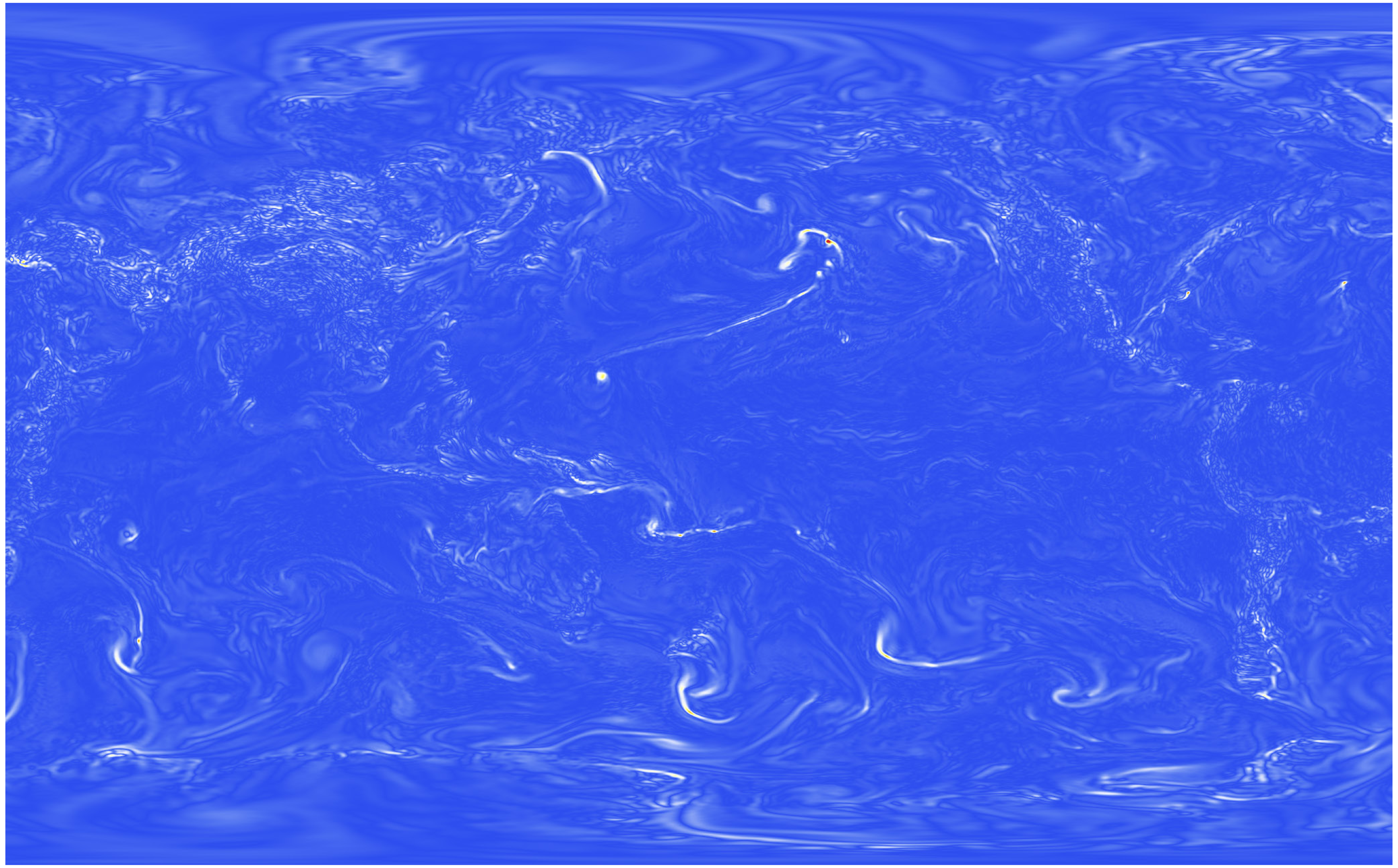
Motivation



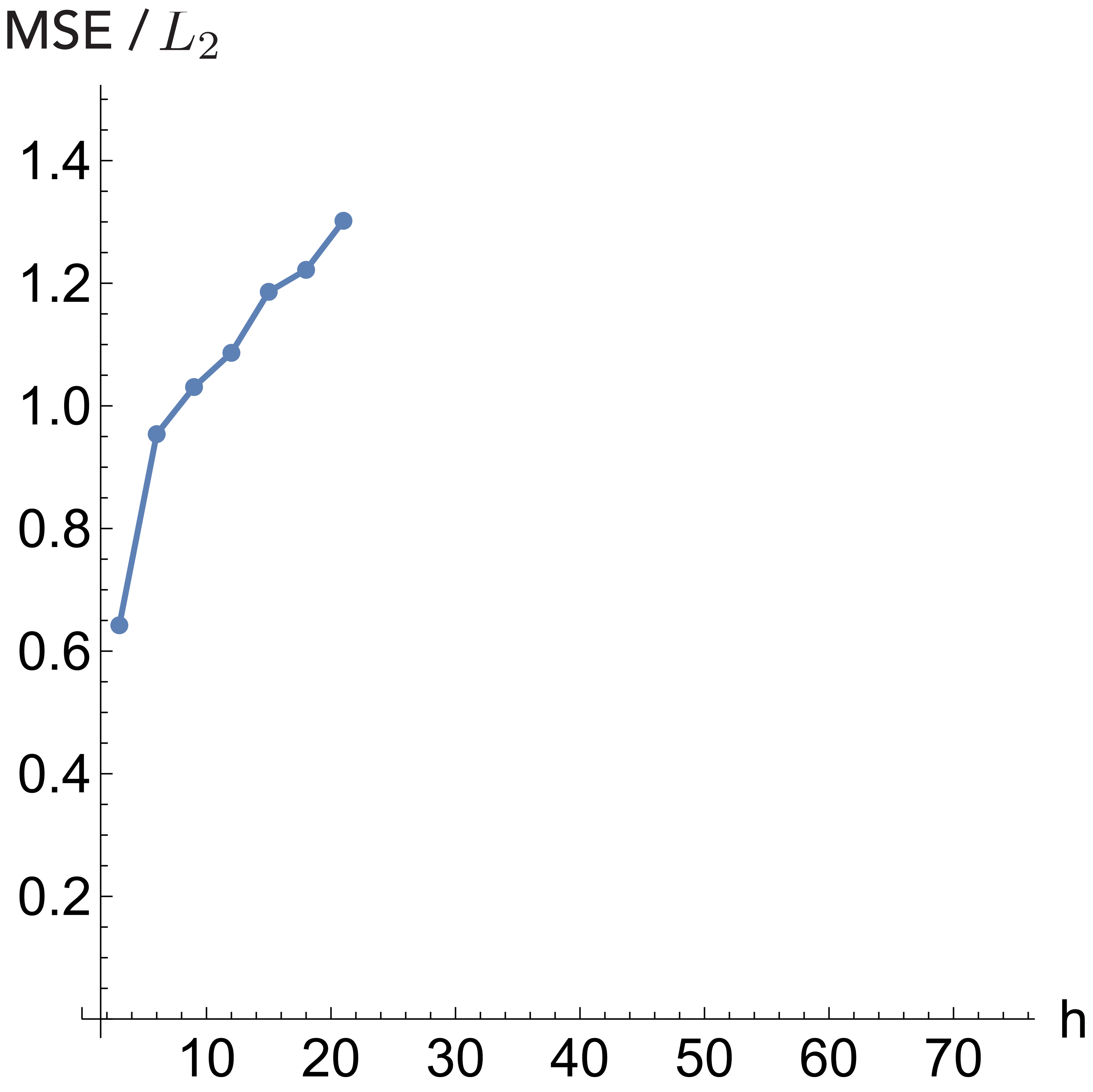
vorticity



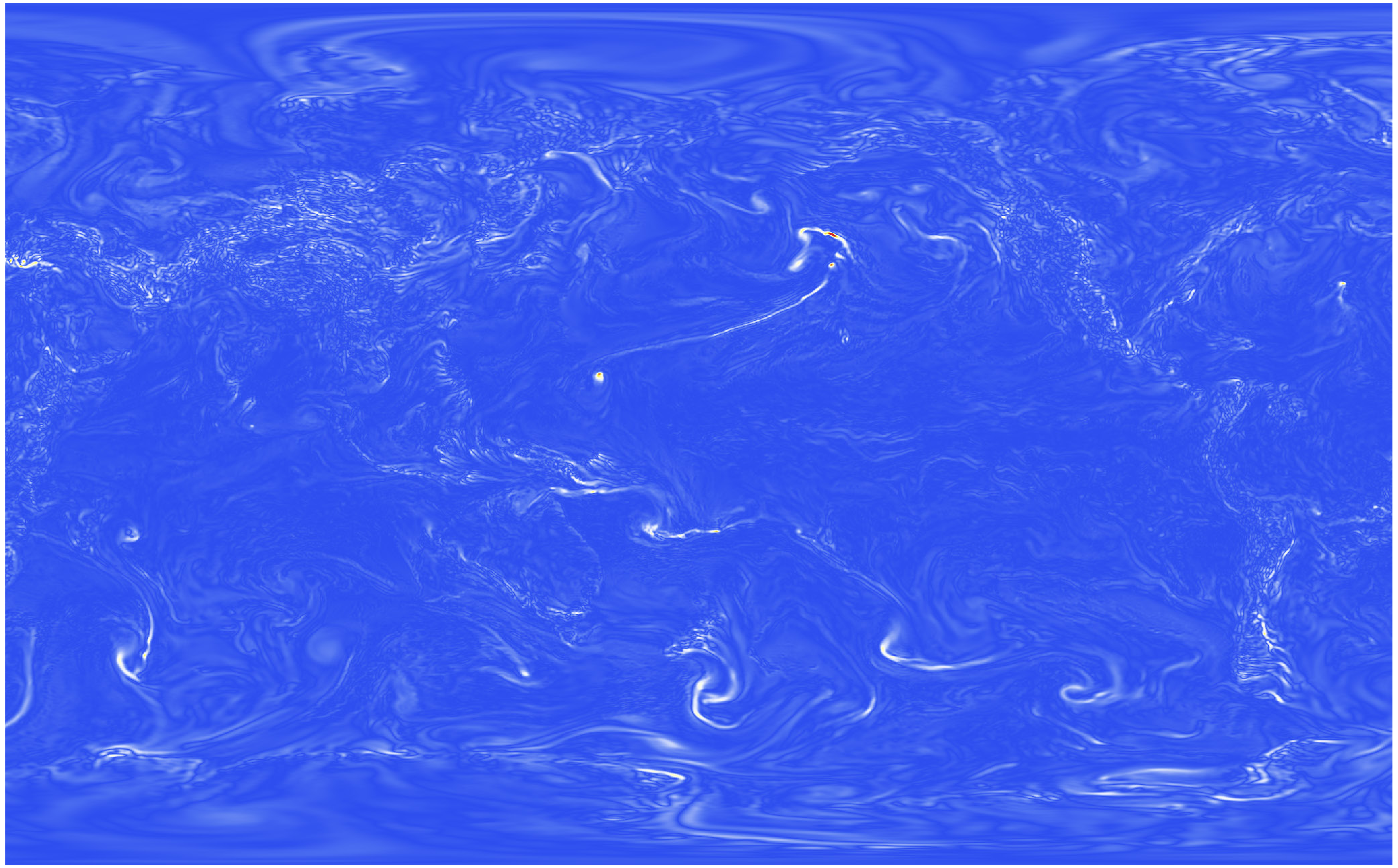
Motivation



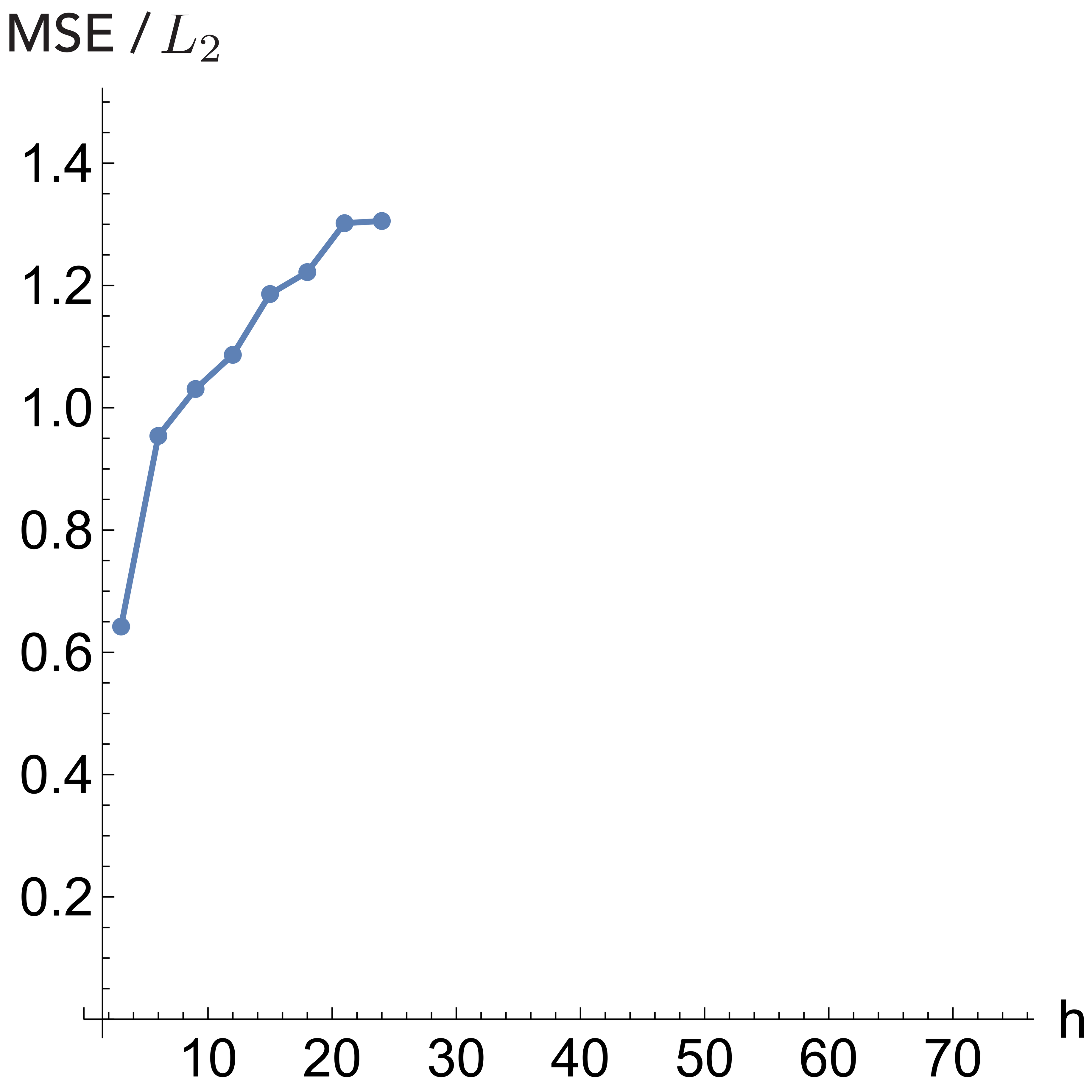
vorticity



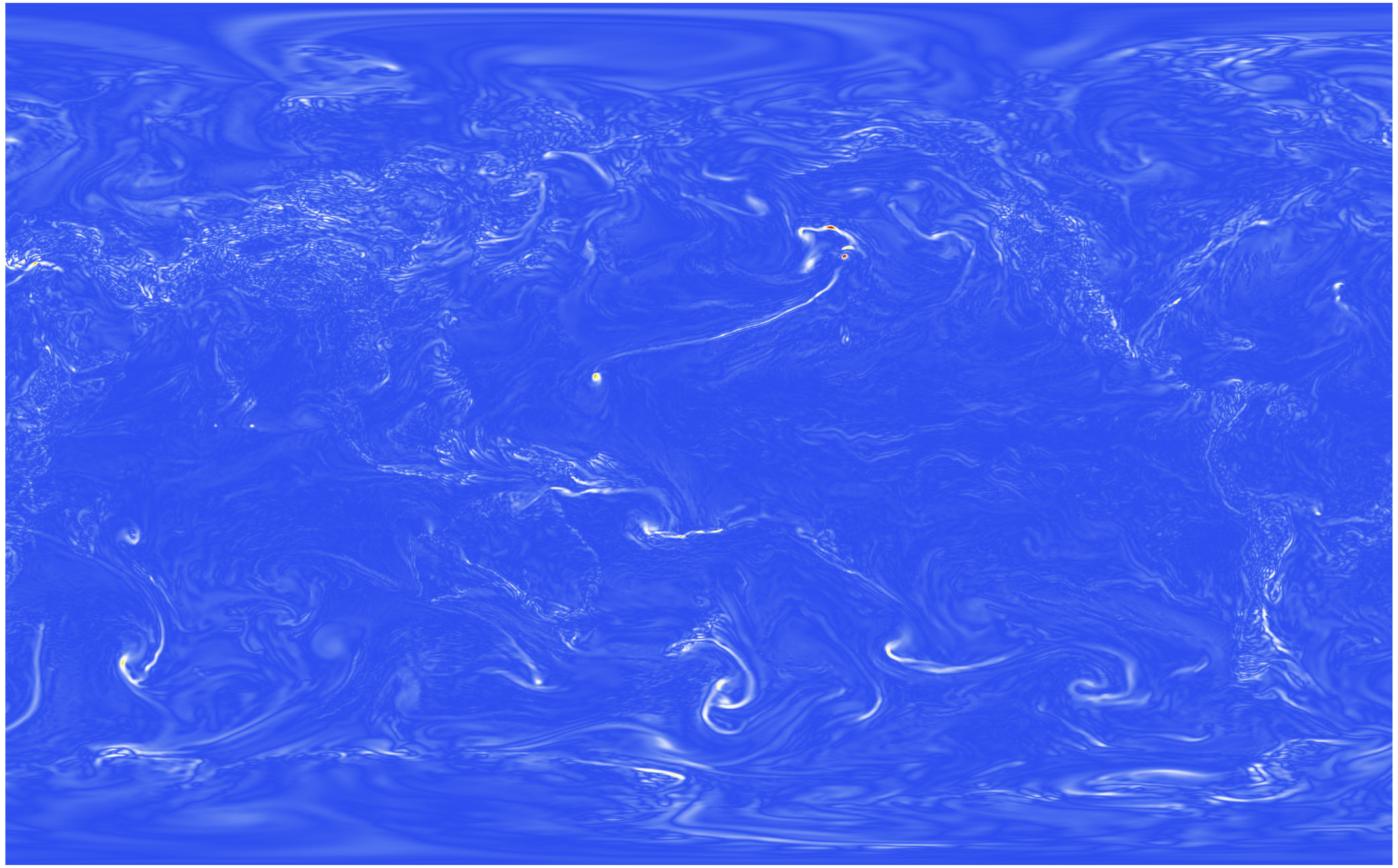
Motivation



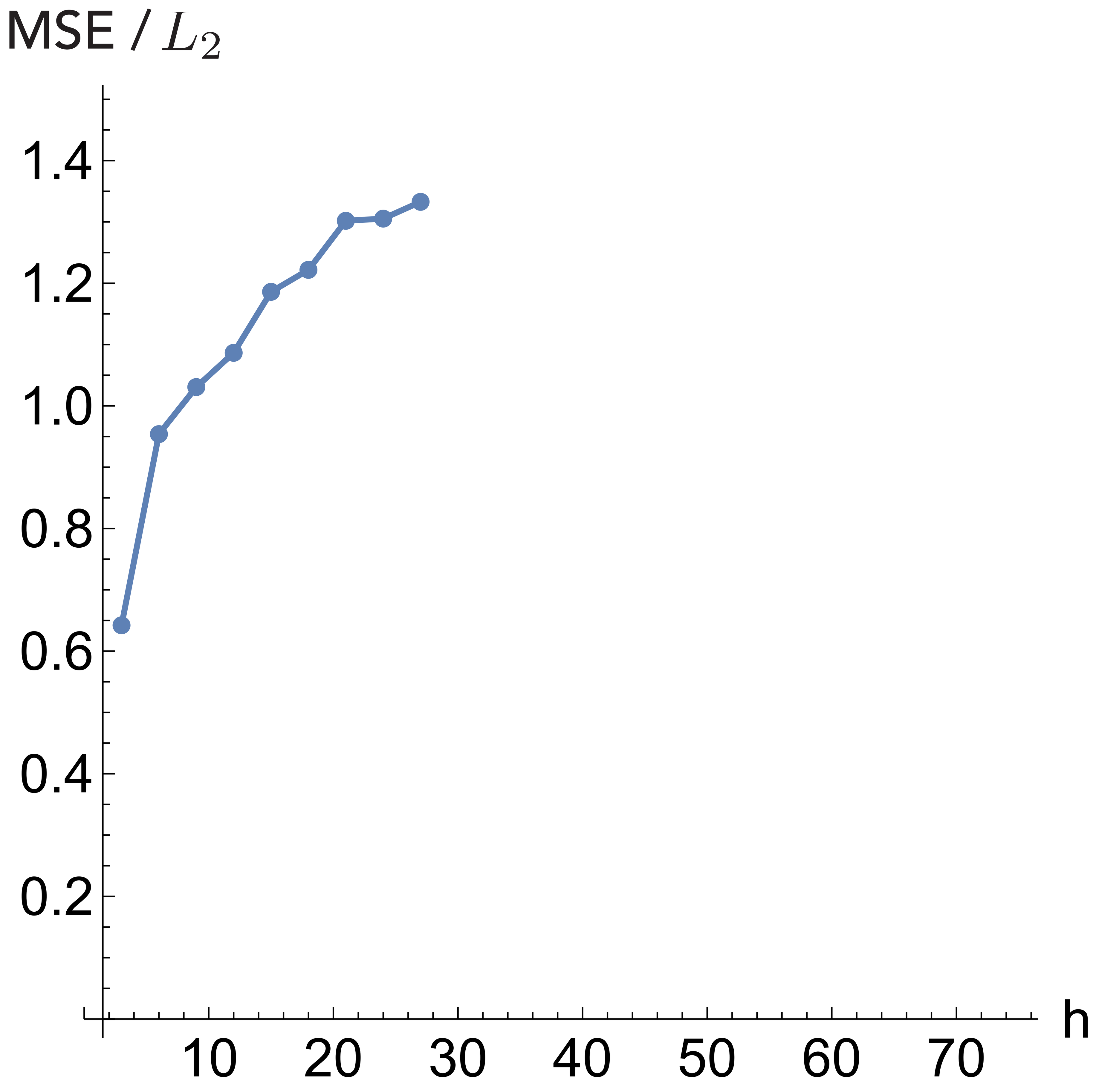
vorticity



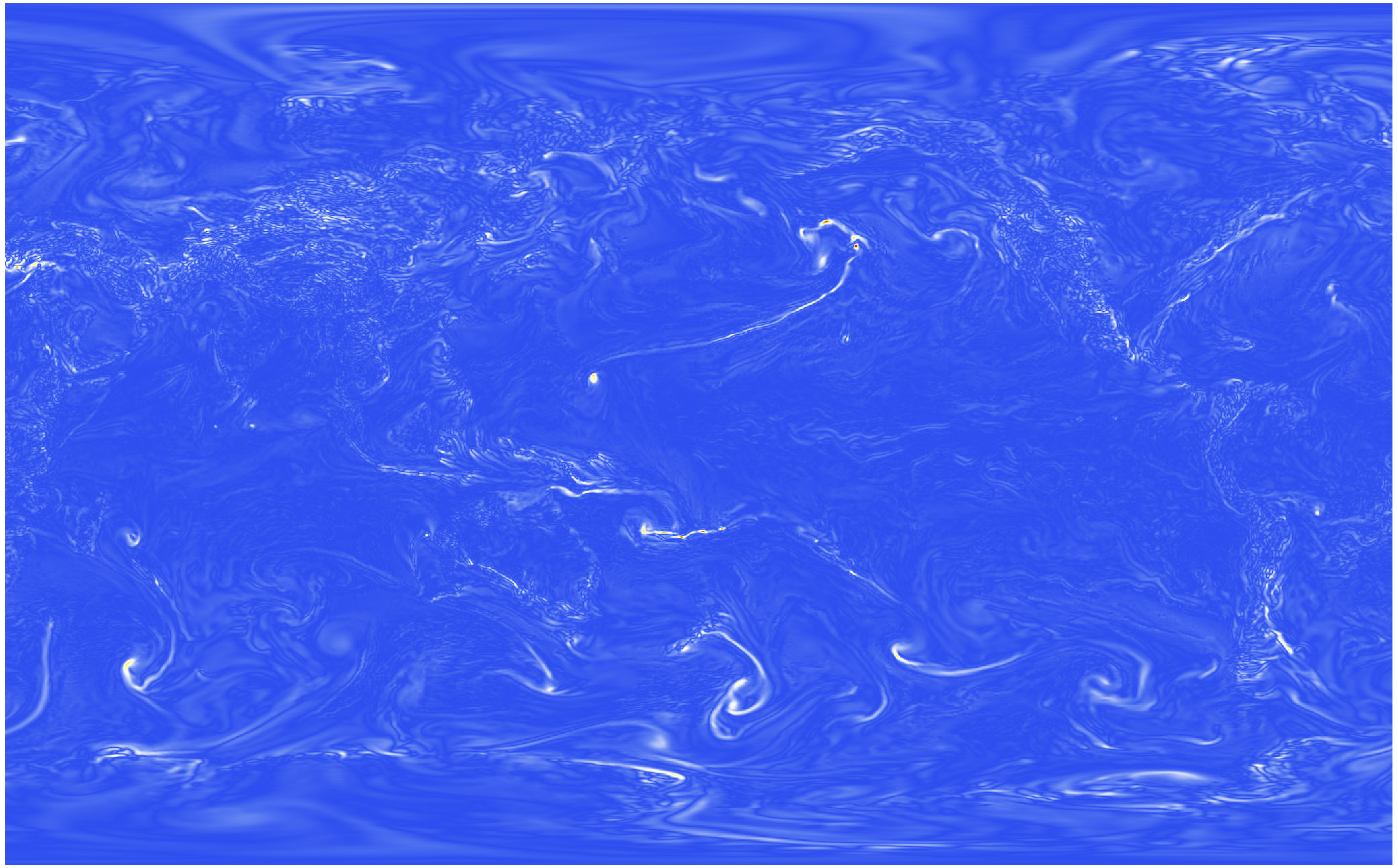
Motivation



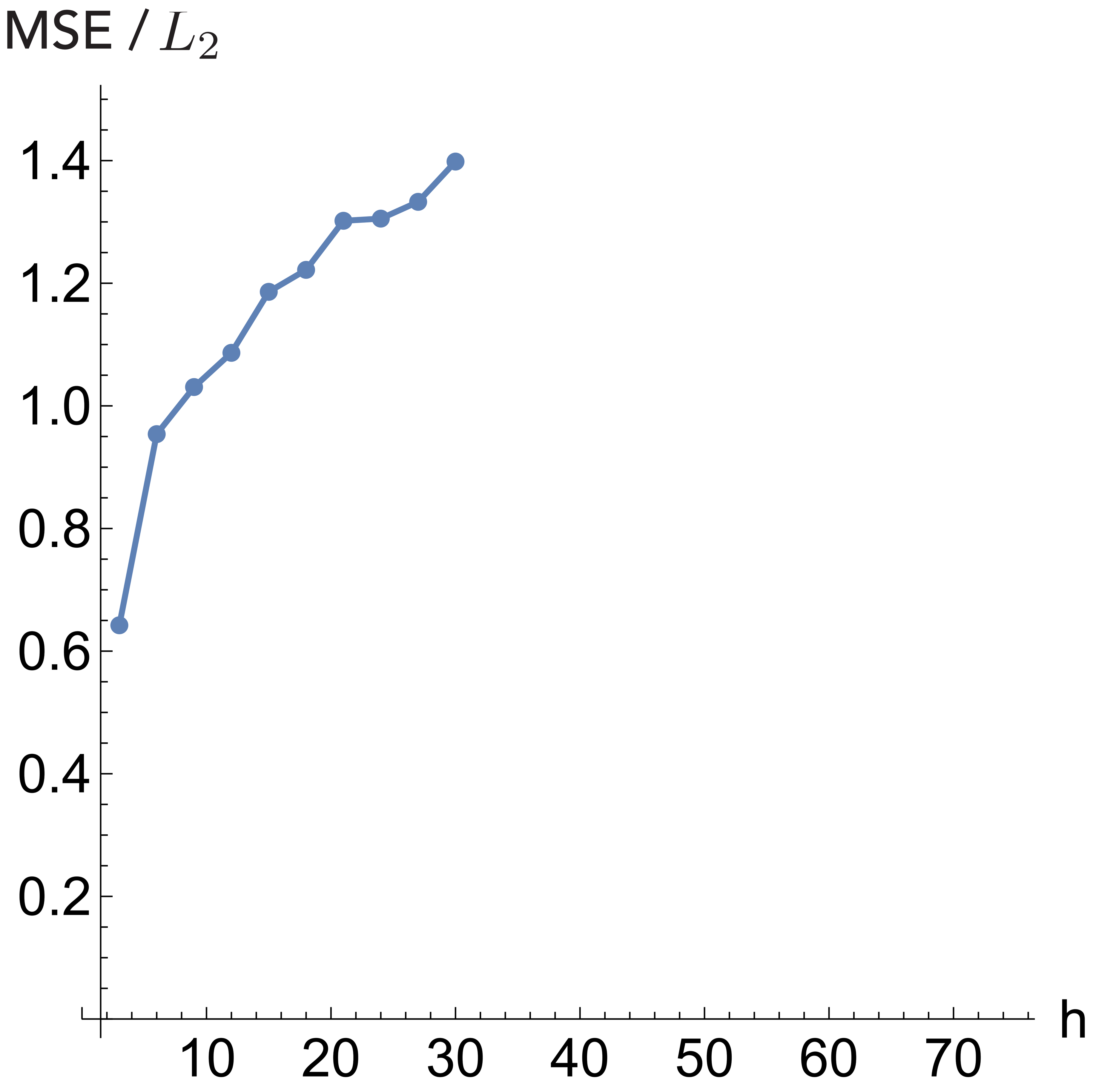
vorticity



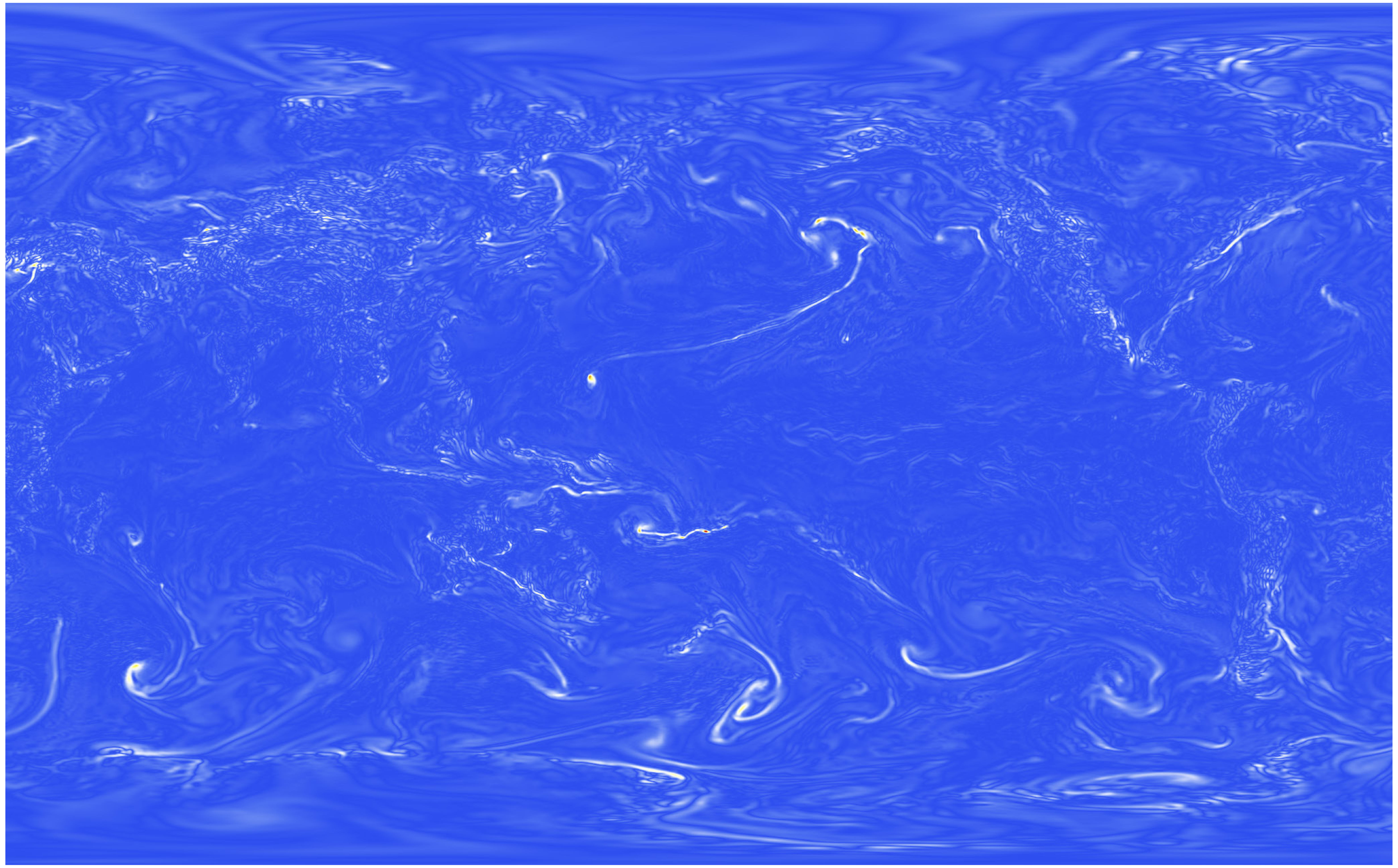
Motivation



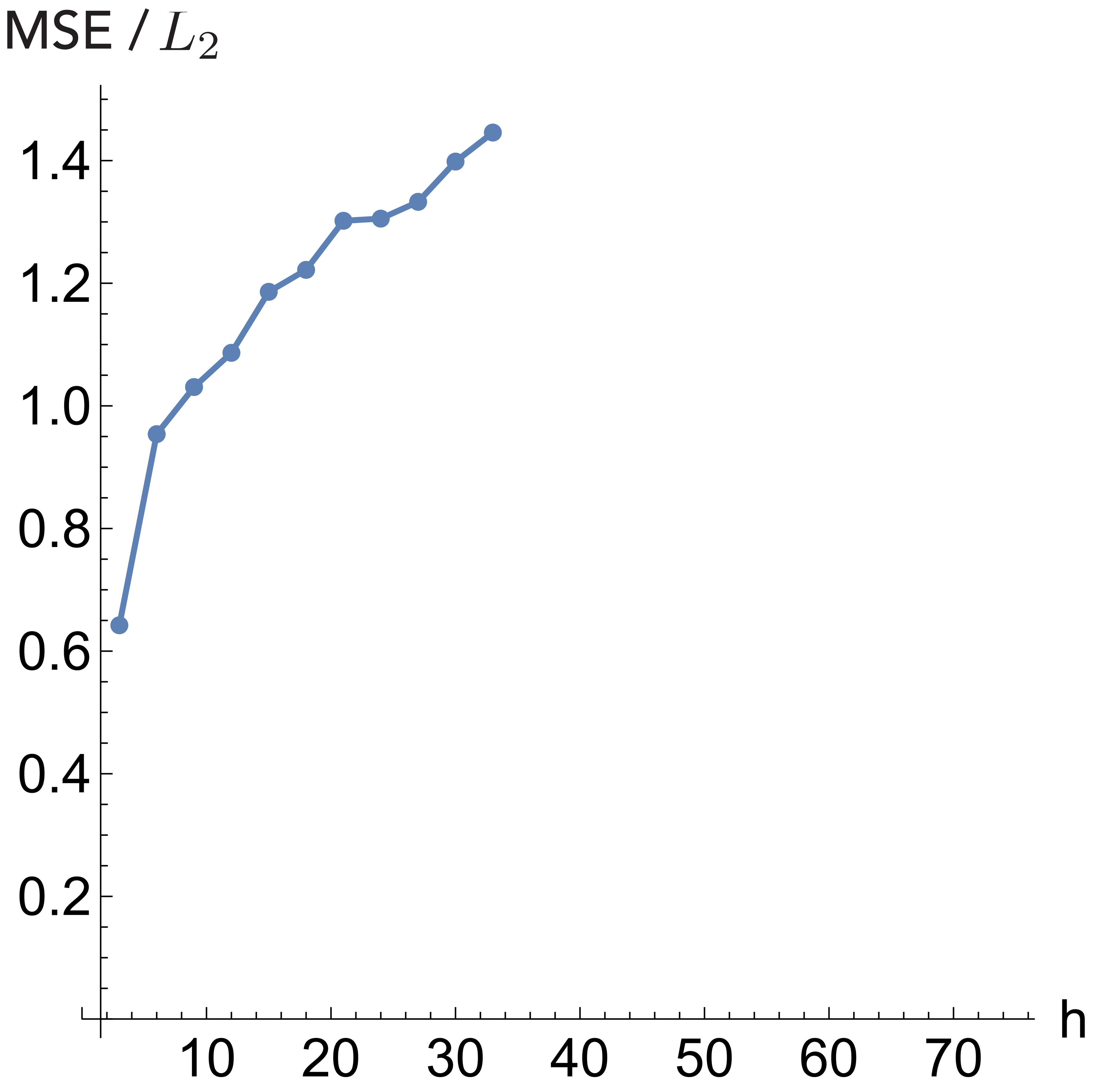
vorticity



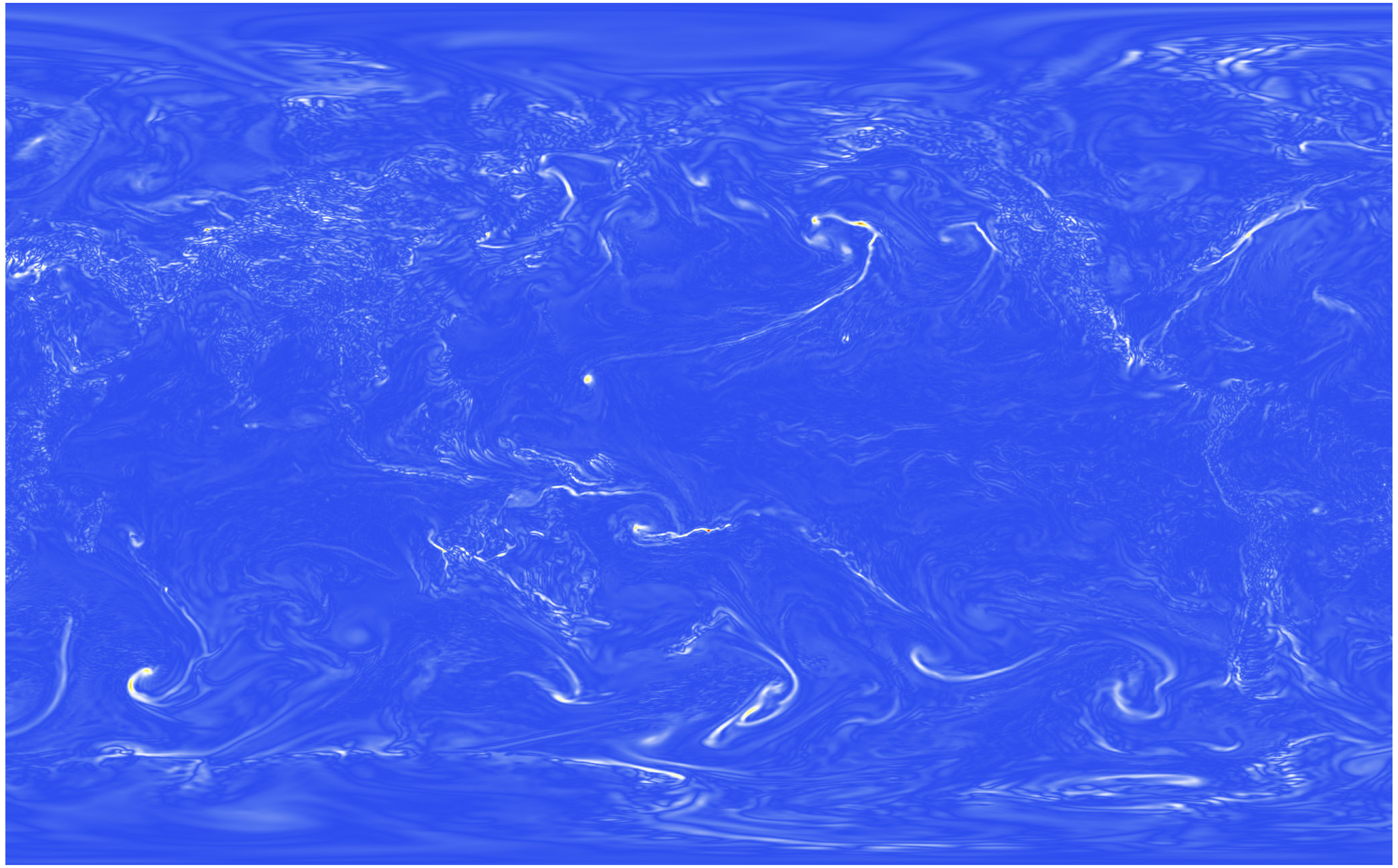
Motivation



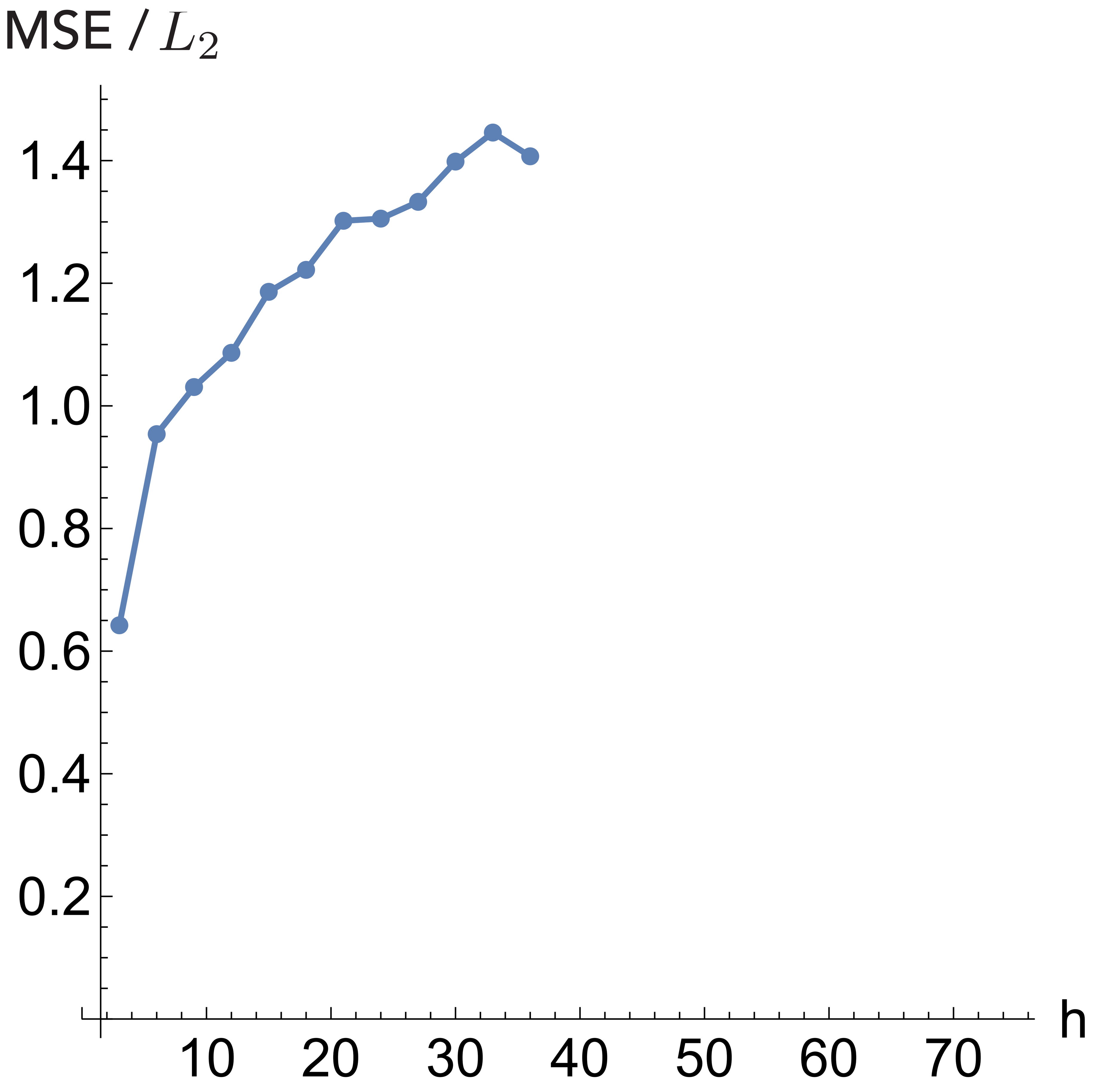
vorticity



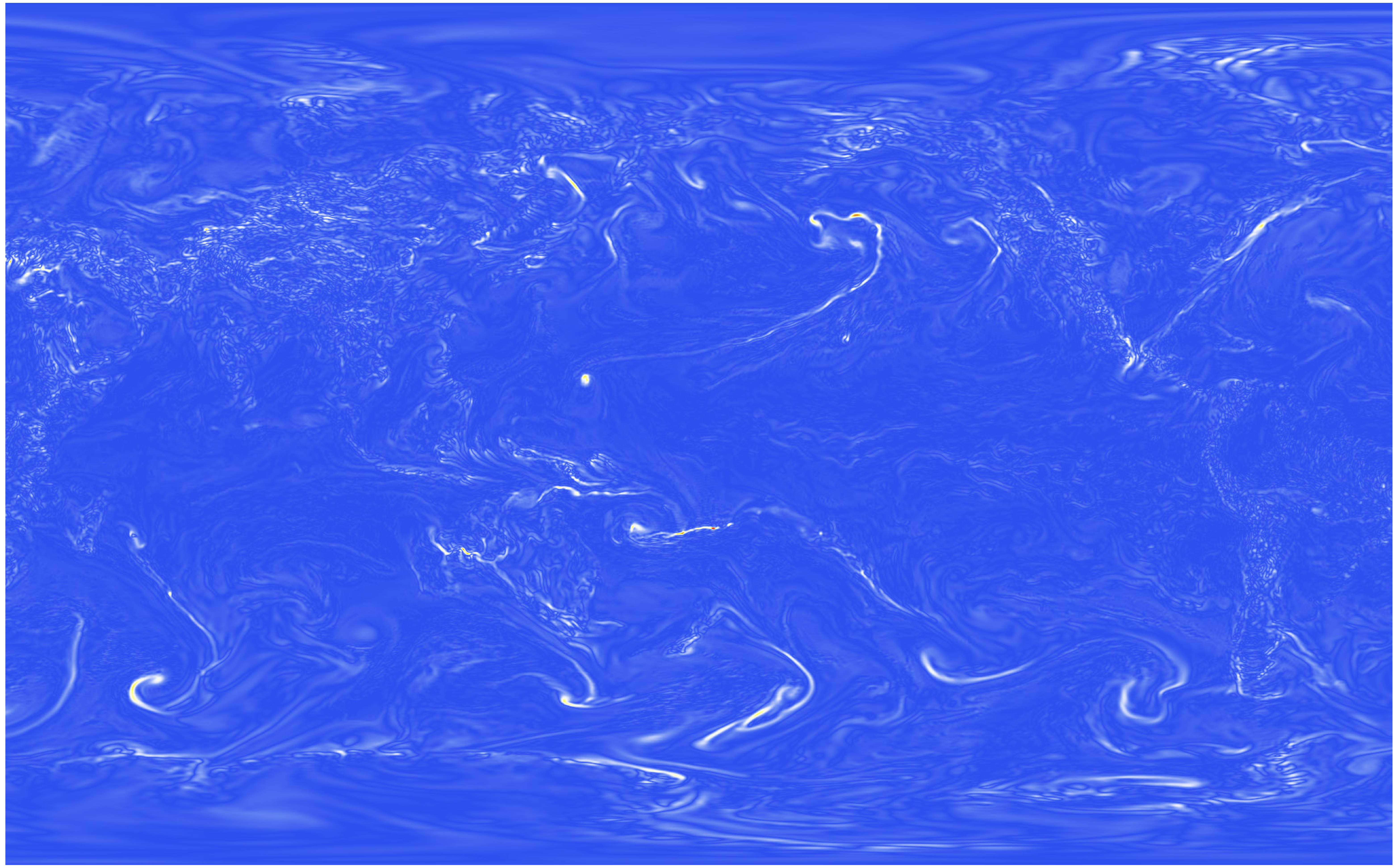
Motivation



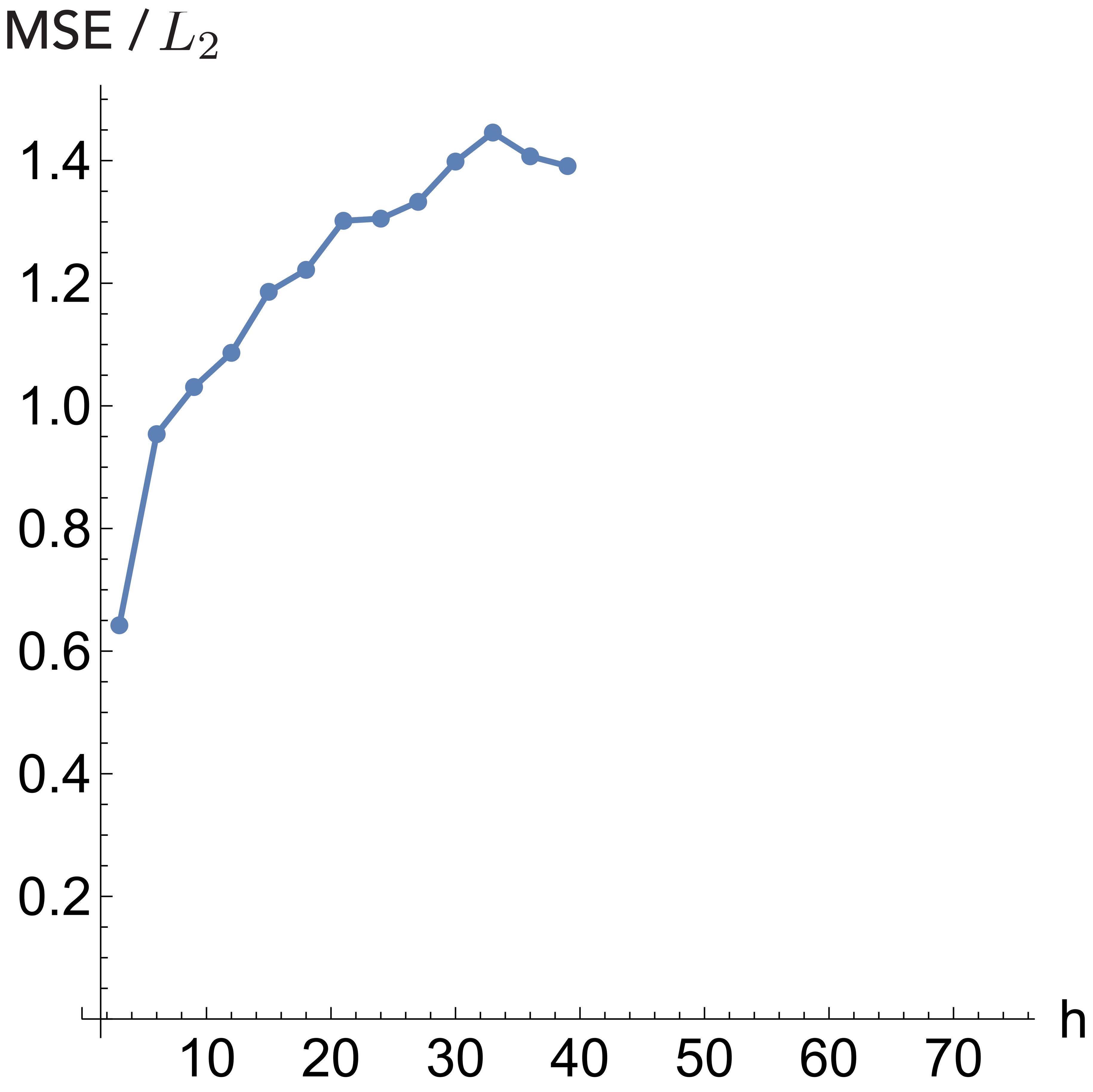
vorticity



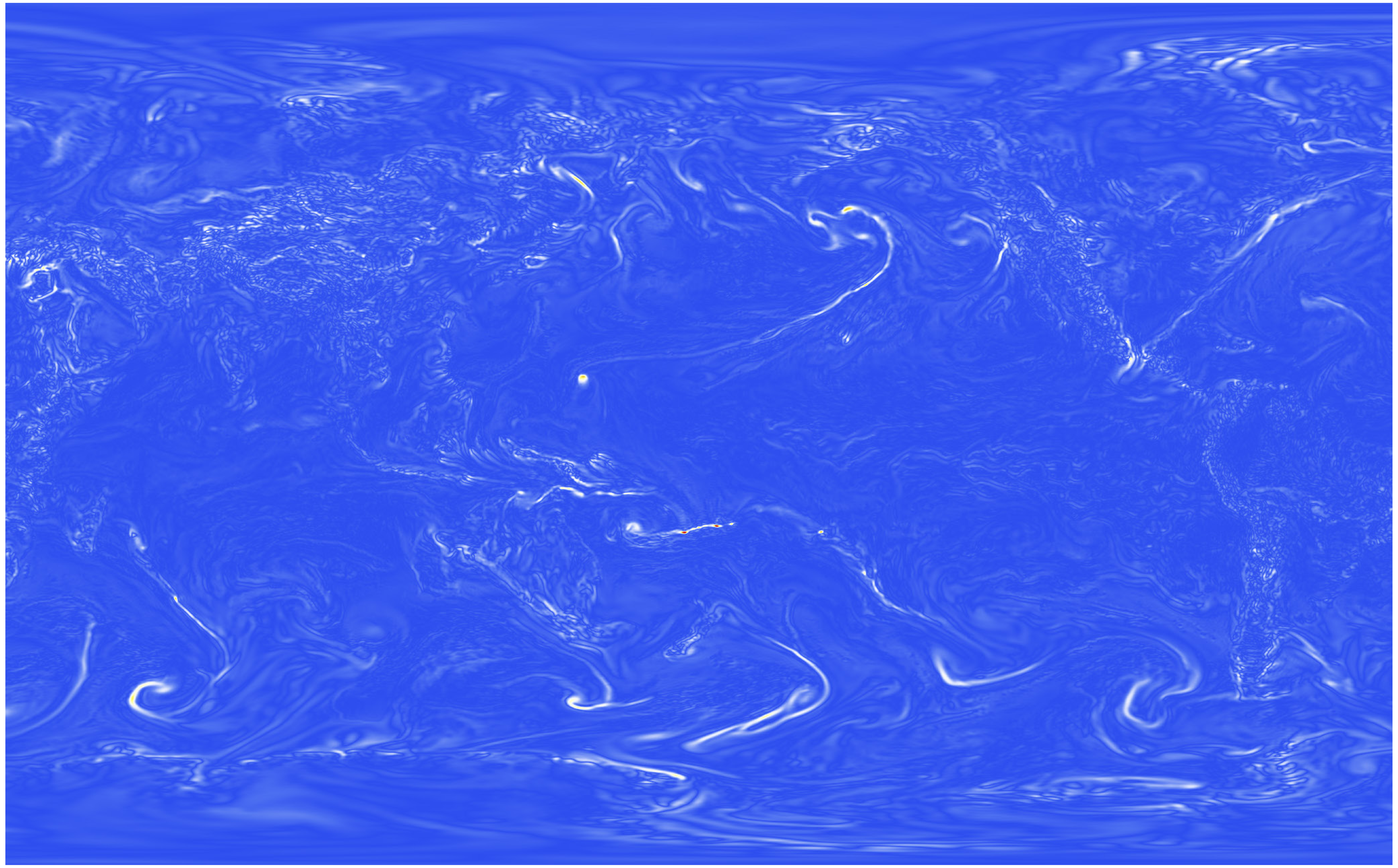
Motivation



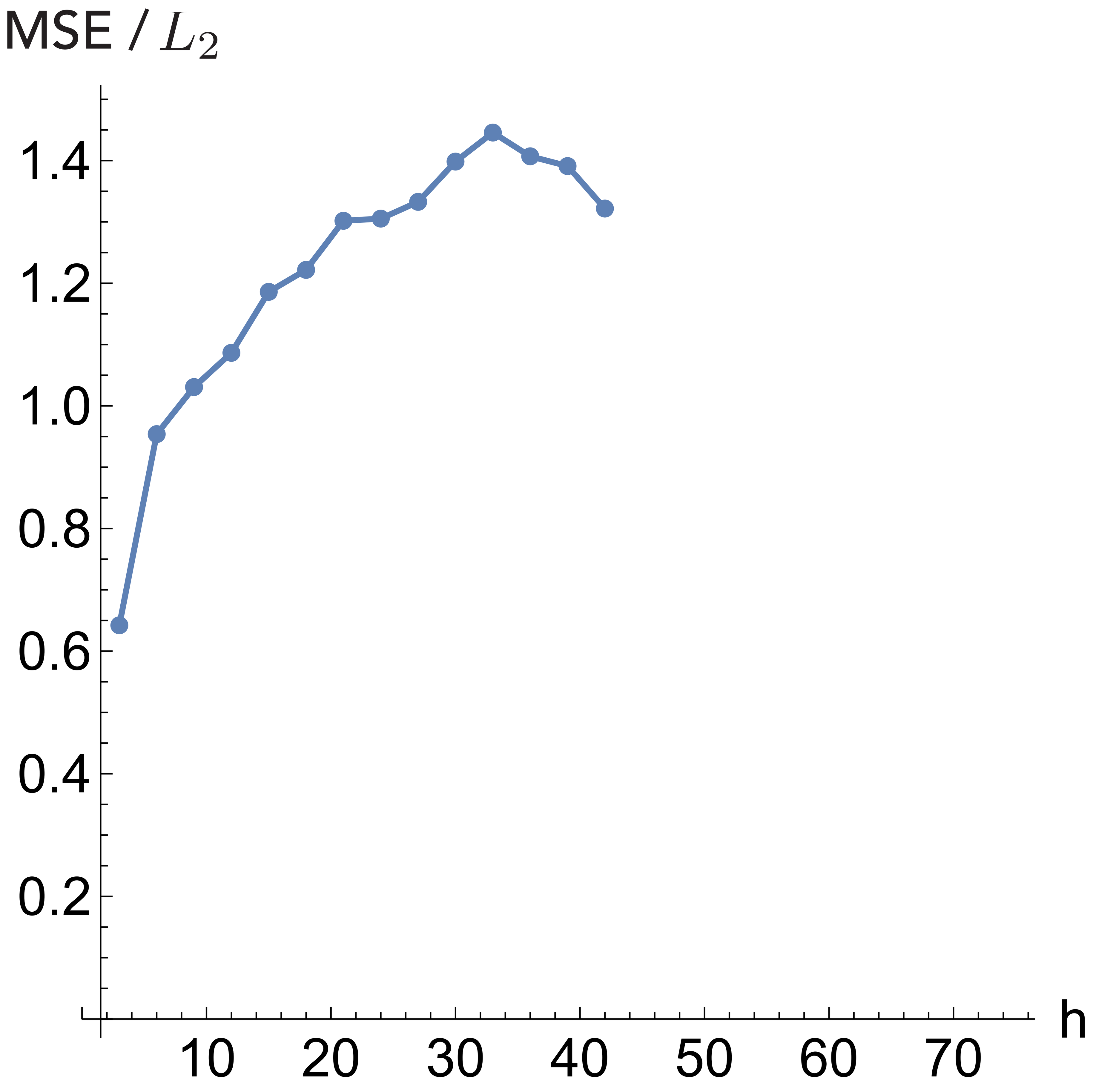
vorticity



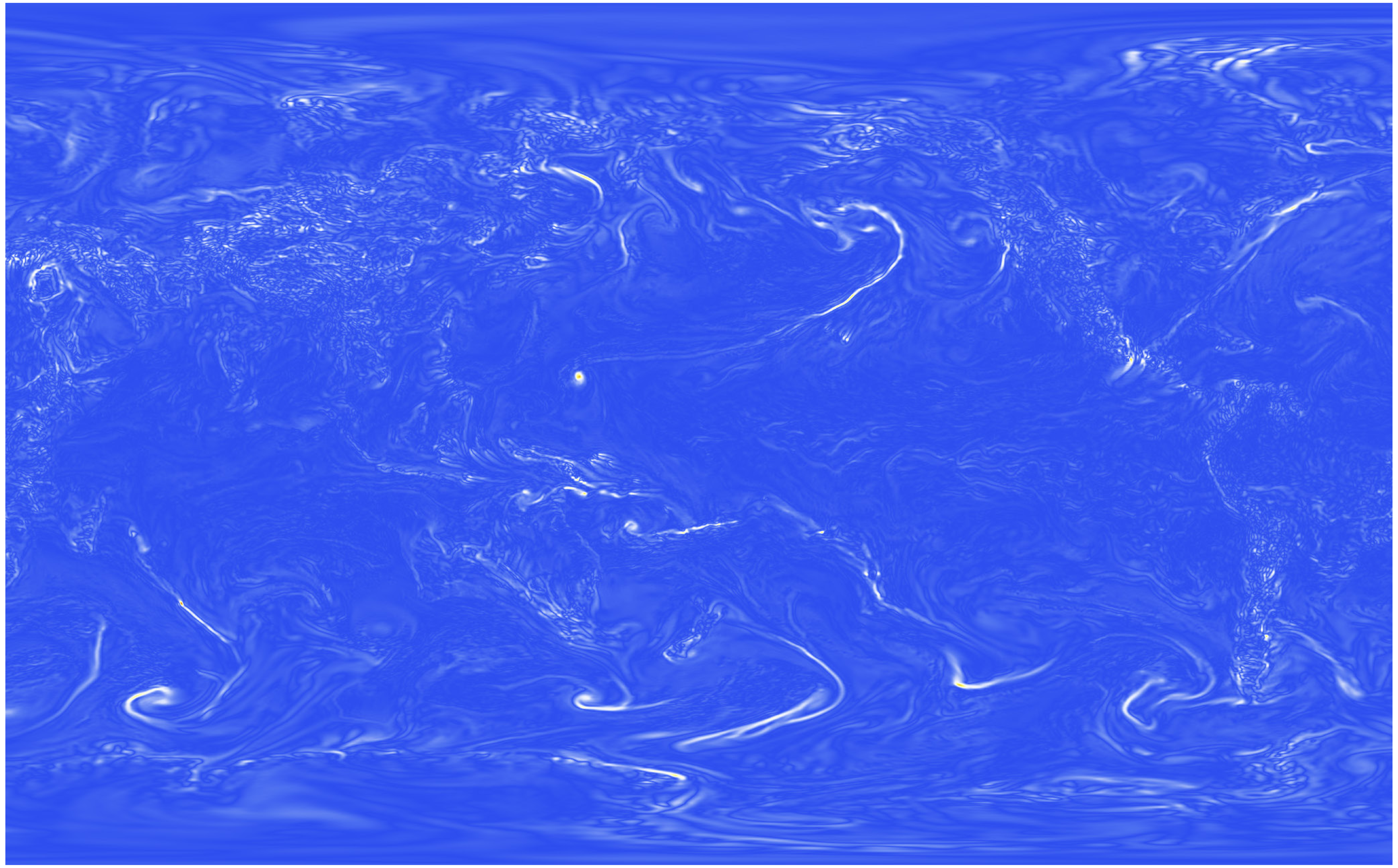
Motivation



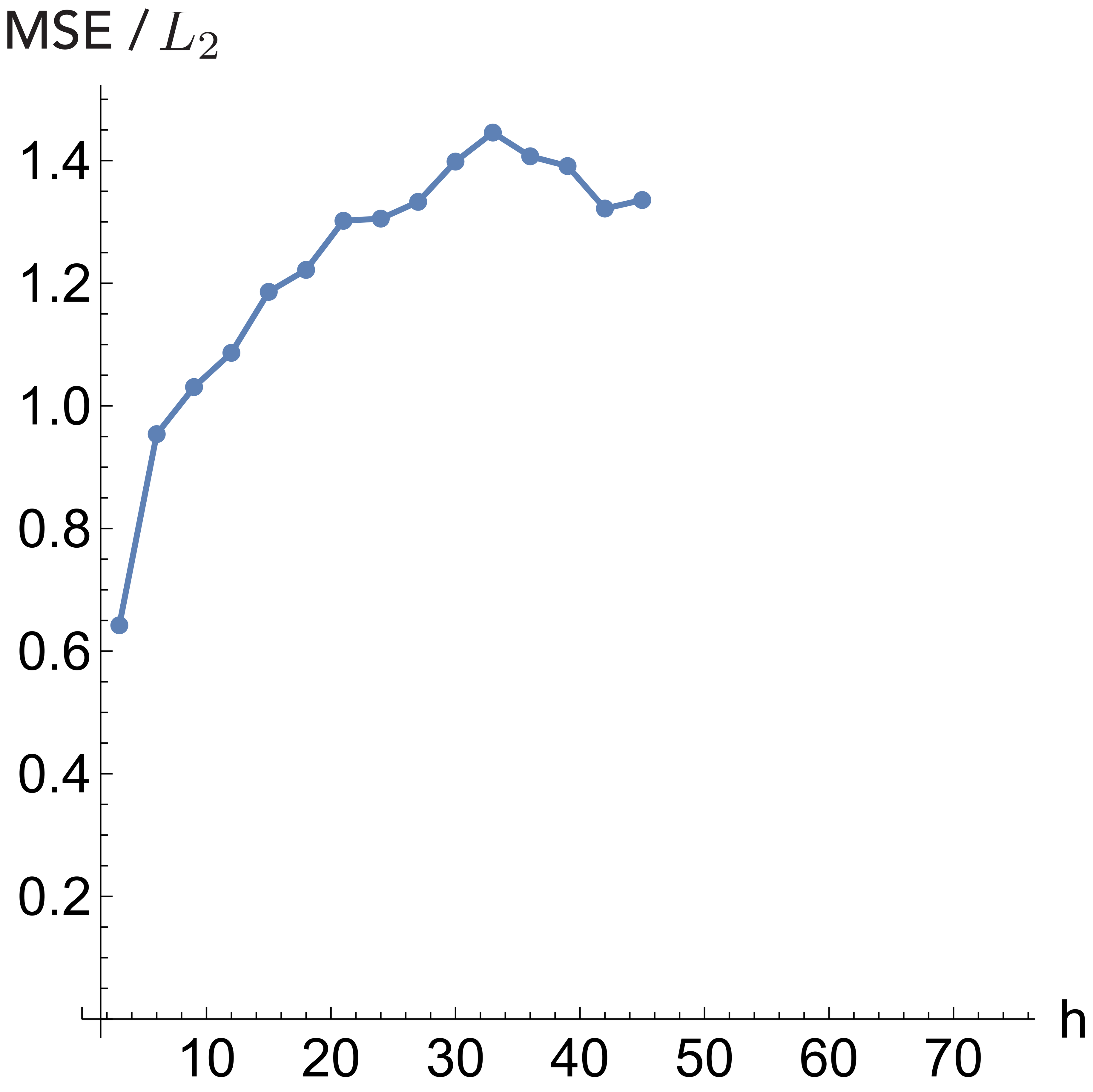
vorticity



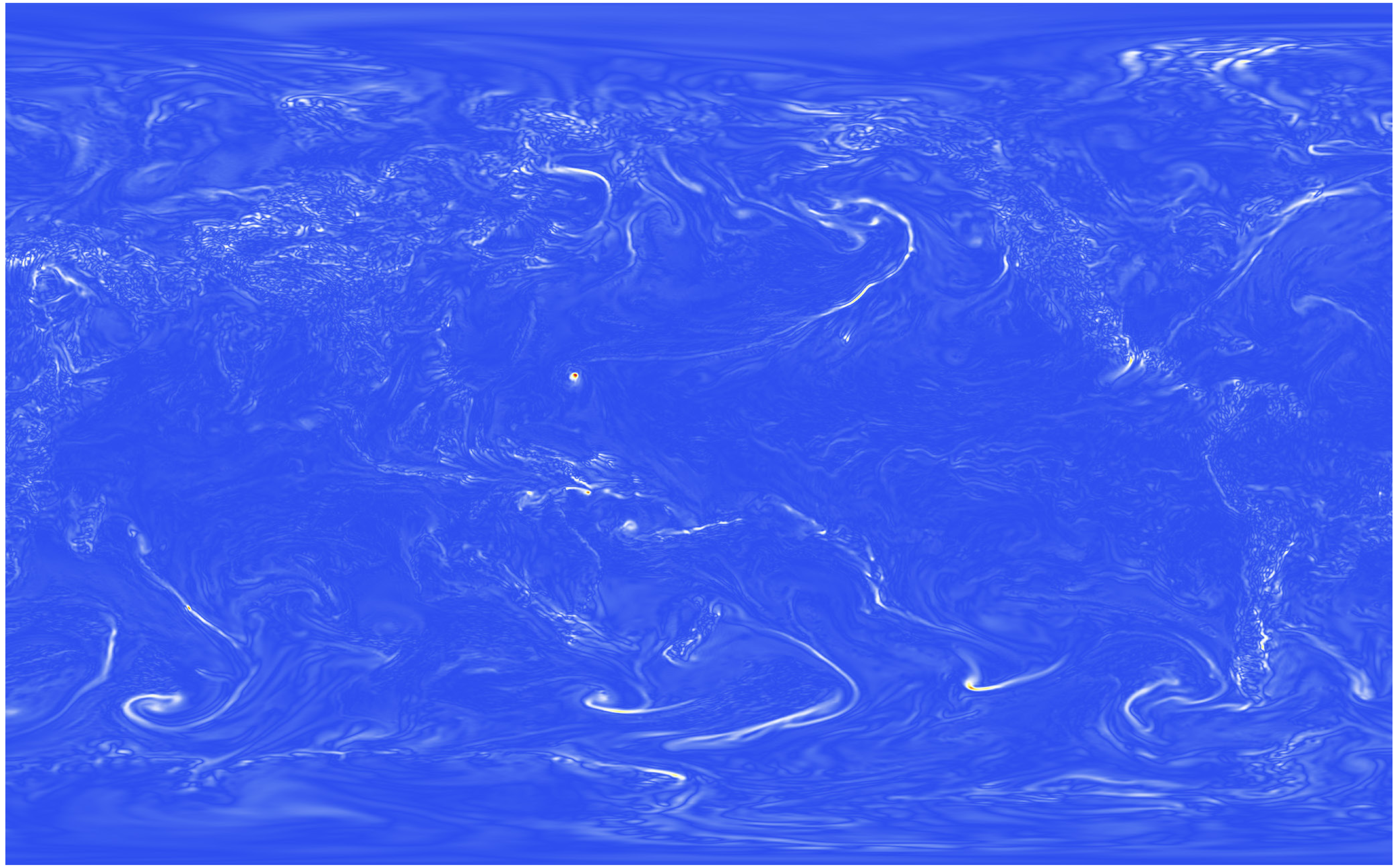
Motivation



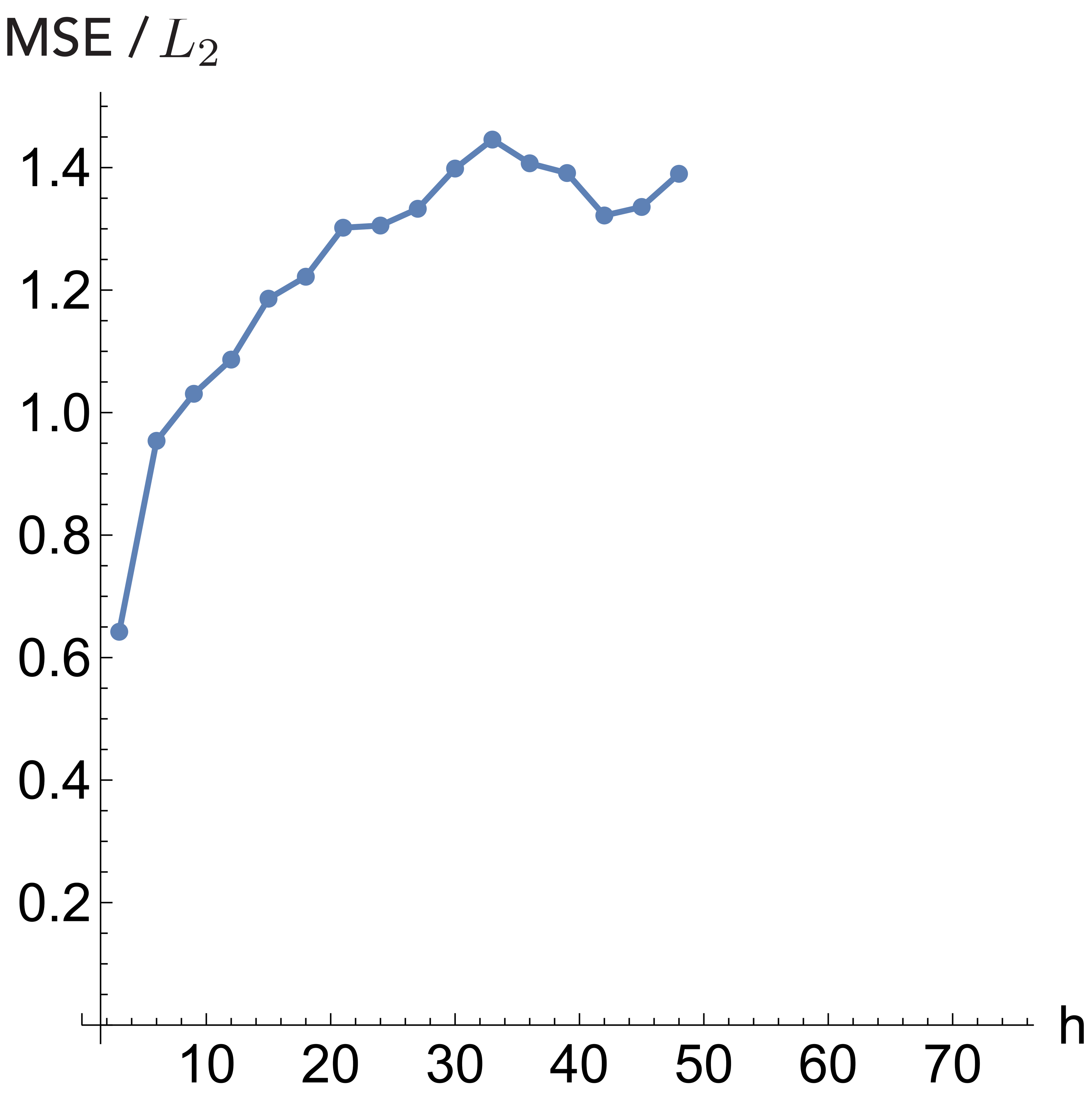
vorticity



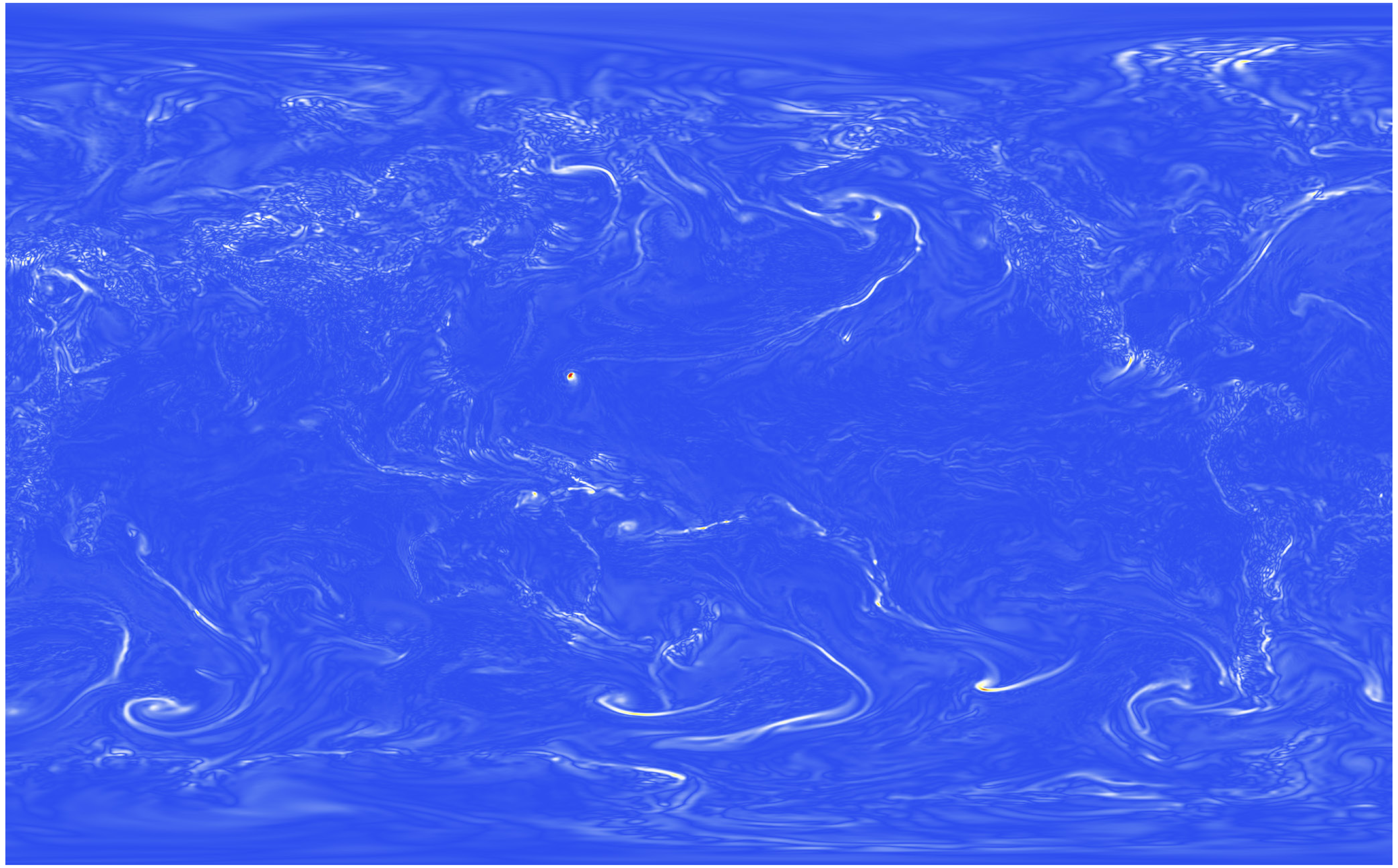
Motivation



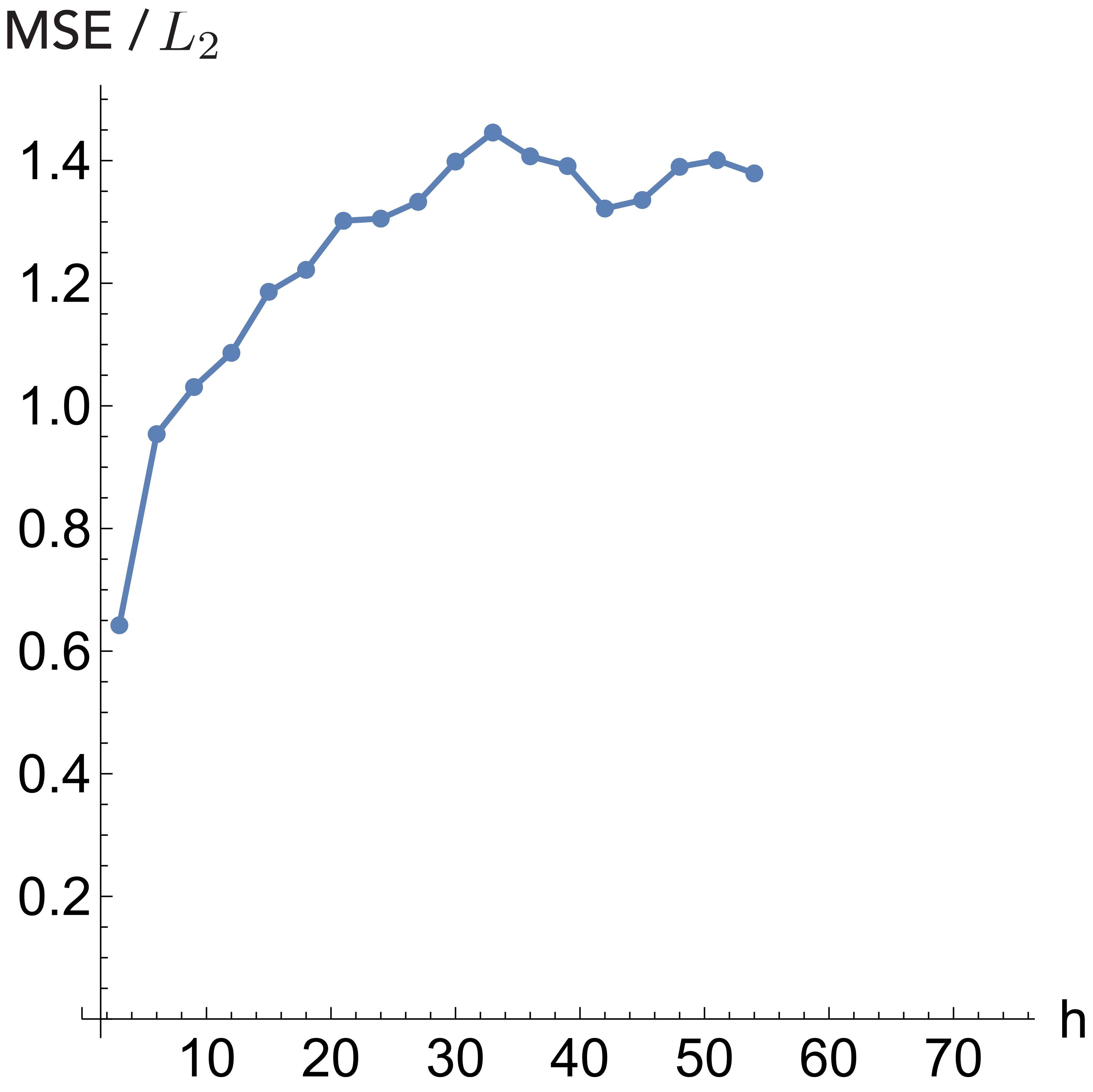
vorticity



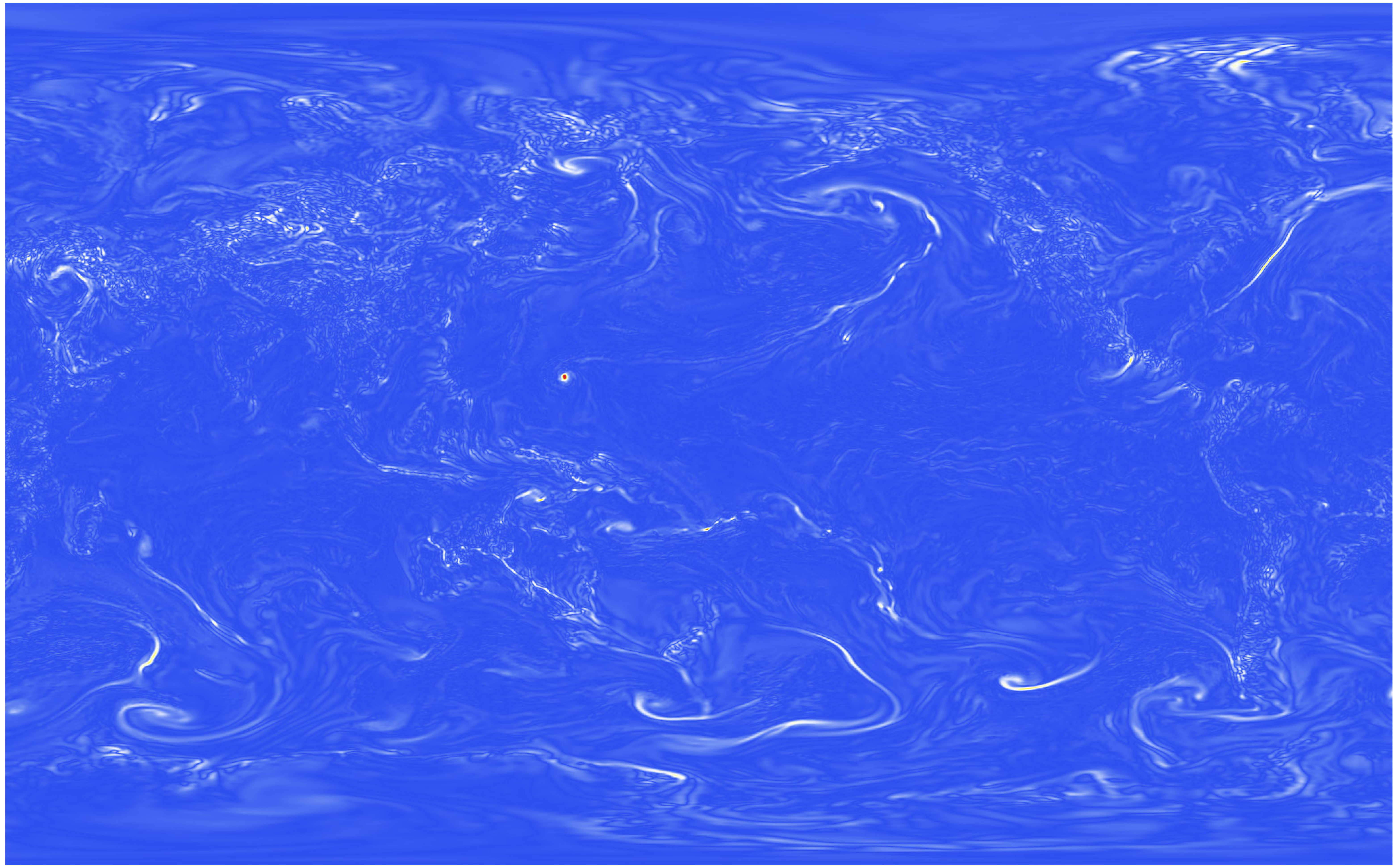
Motivation



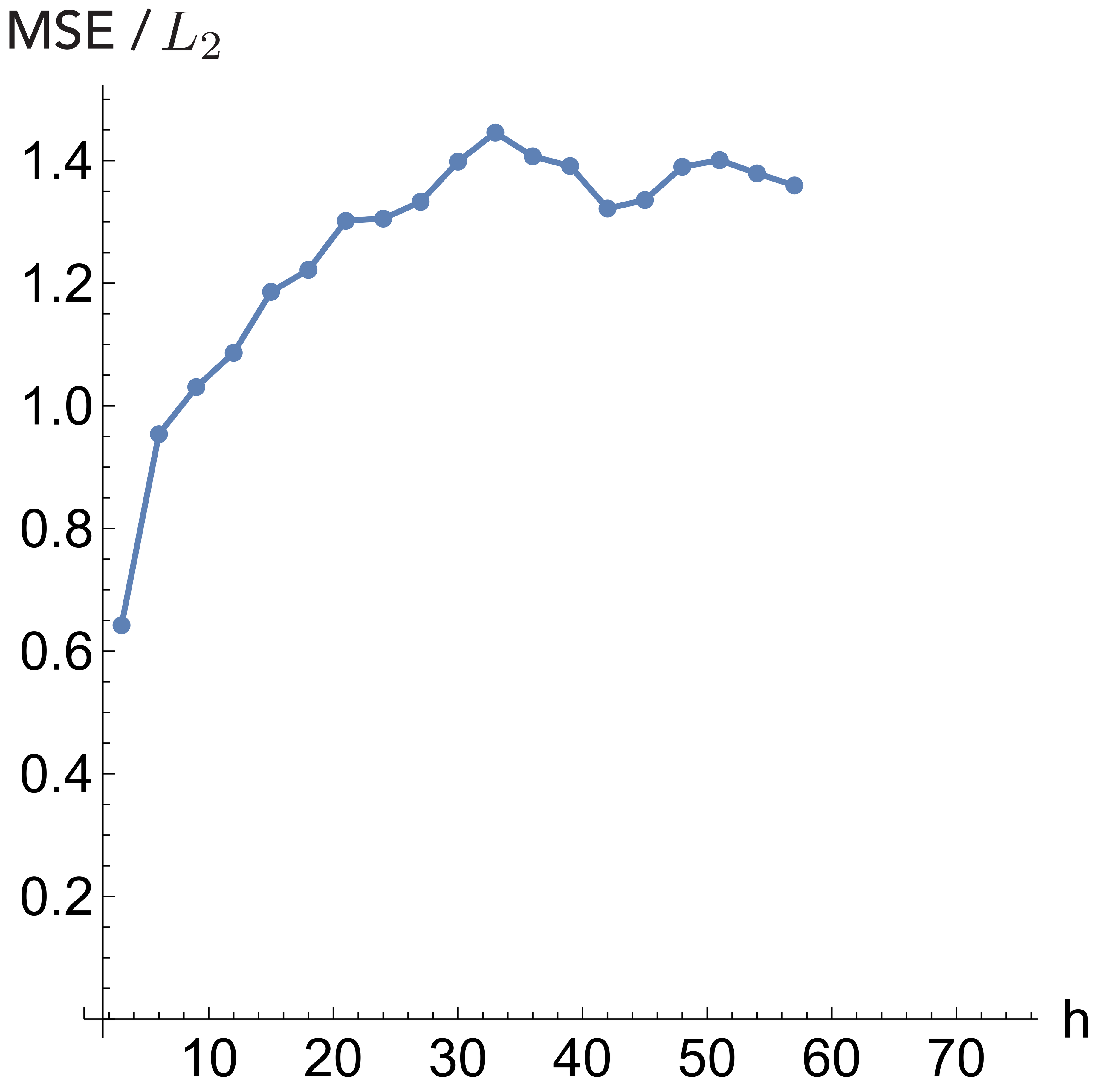
vorticity



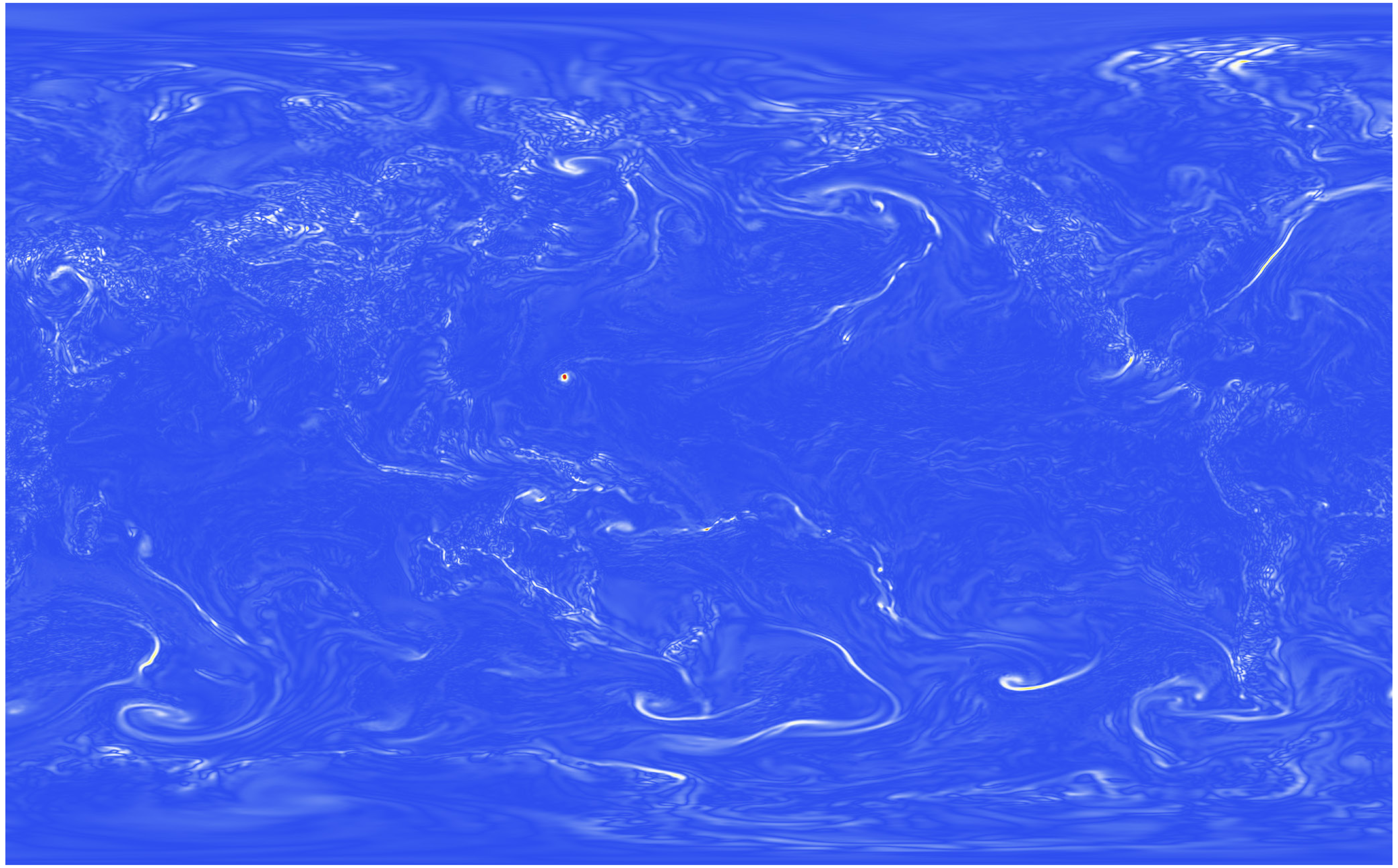
Motivation



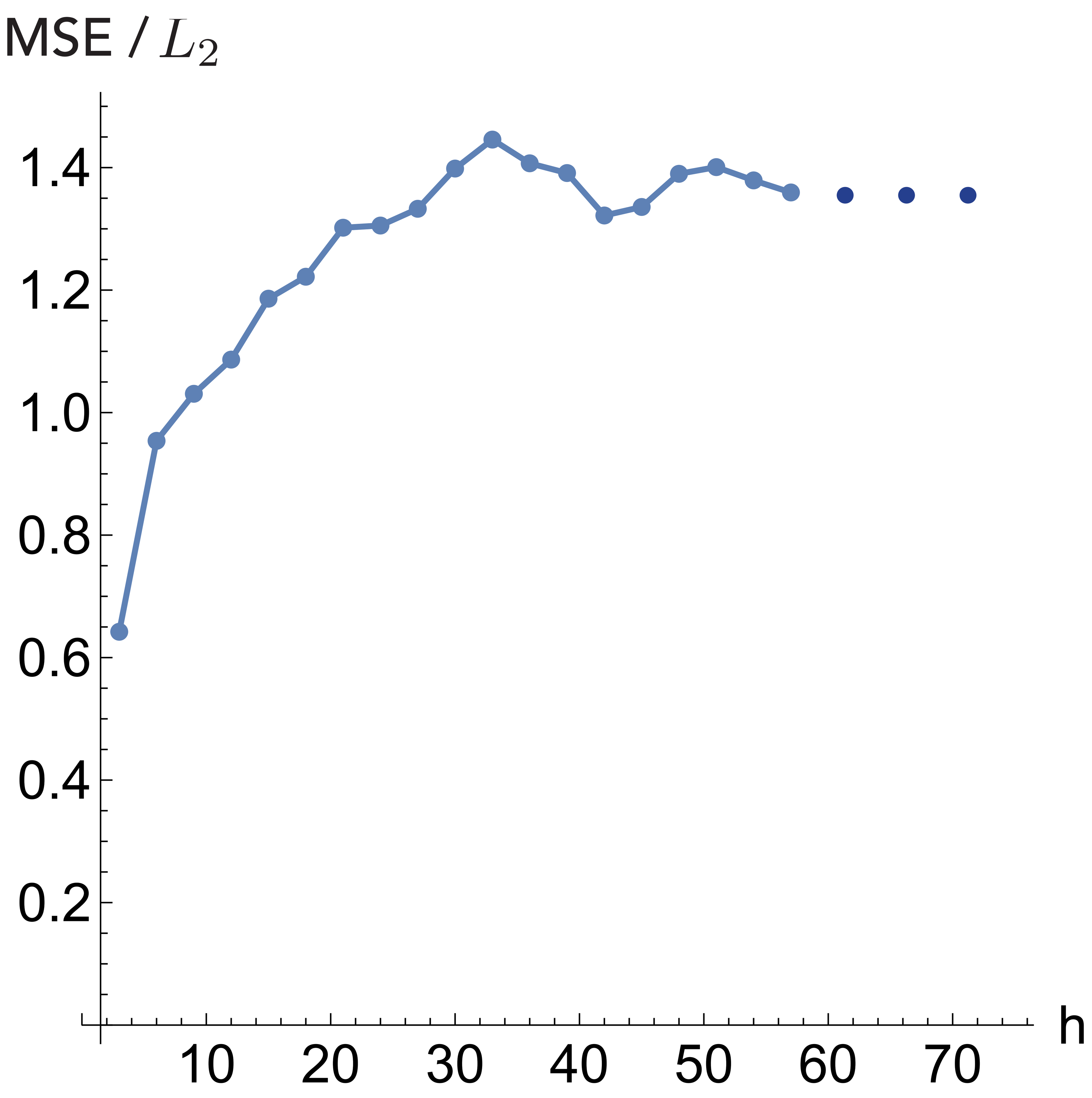
vorticity



Motivation



vorticity



Motivation

- Learning relies on informative and discriminative loss functions
  - Mathematically grounded norms are often not particularly effective

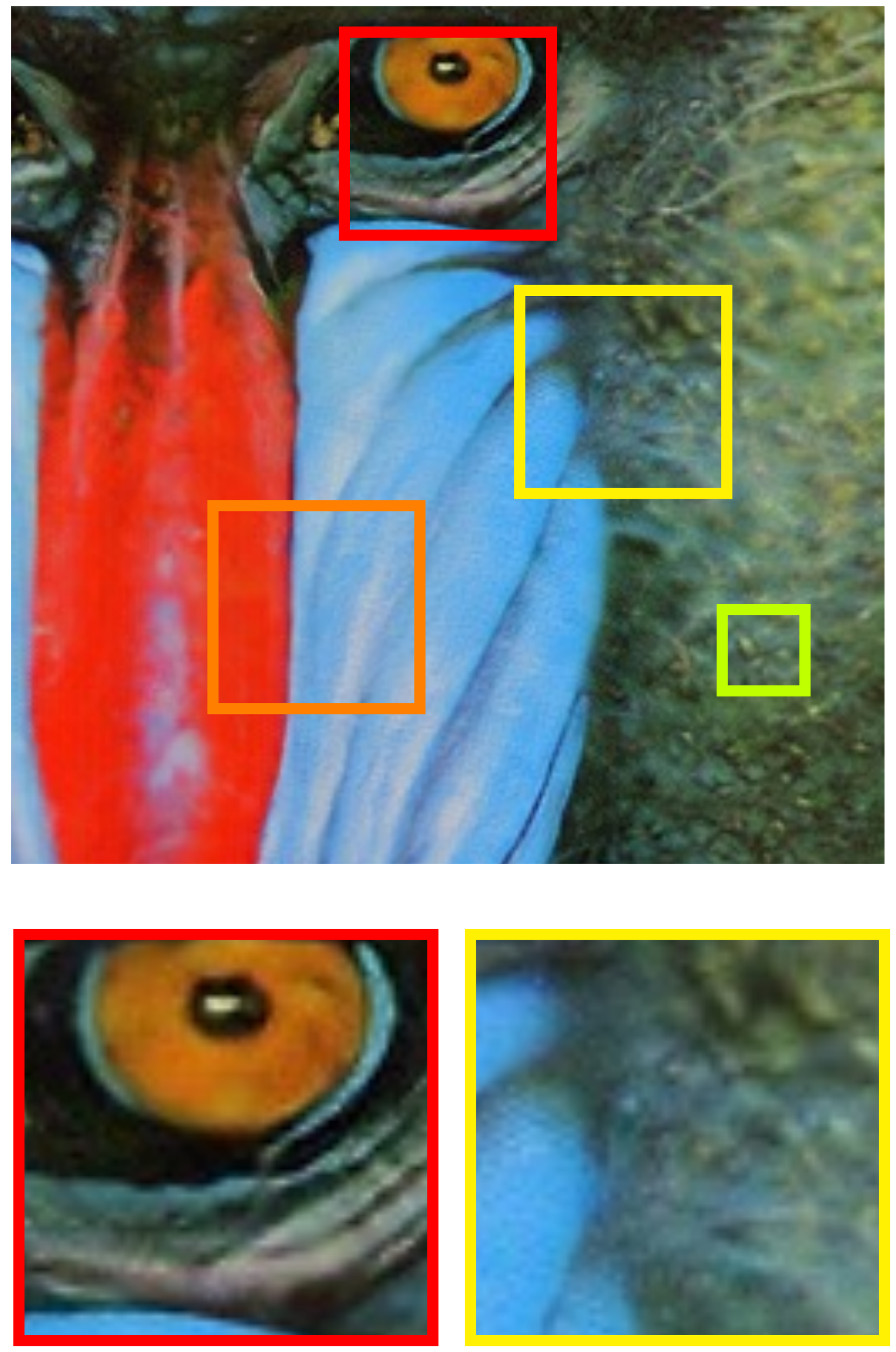
Motivation

- Learning relies on informative and discriminative loss functions
  - Mathematically grounded norms are often not particularly effective
  - Similar issues for natural images

Motivation

- GAN-based super-resolution:[1]

SRGAN-MSE

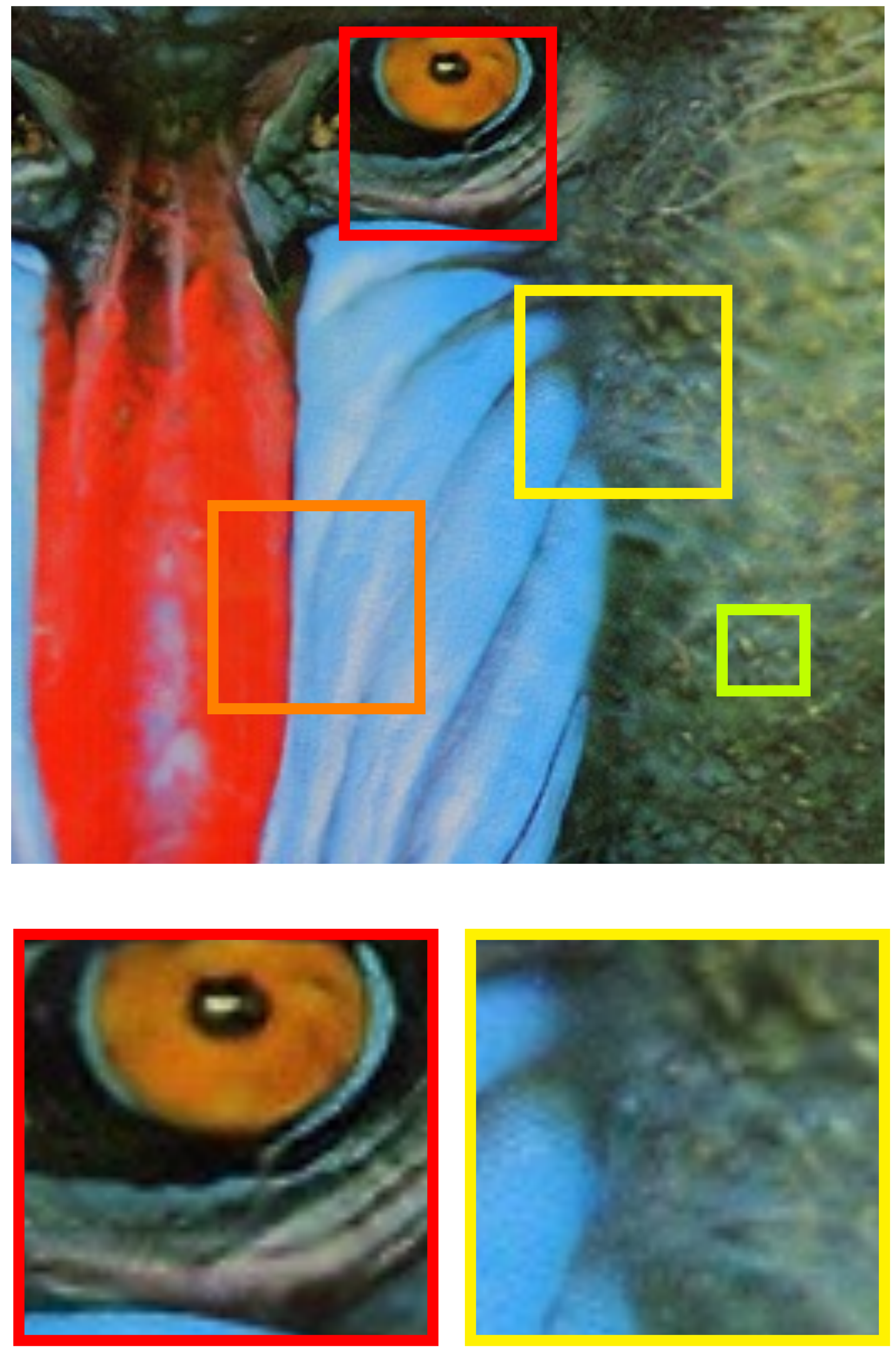


[1] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017.

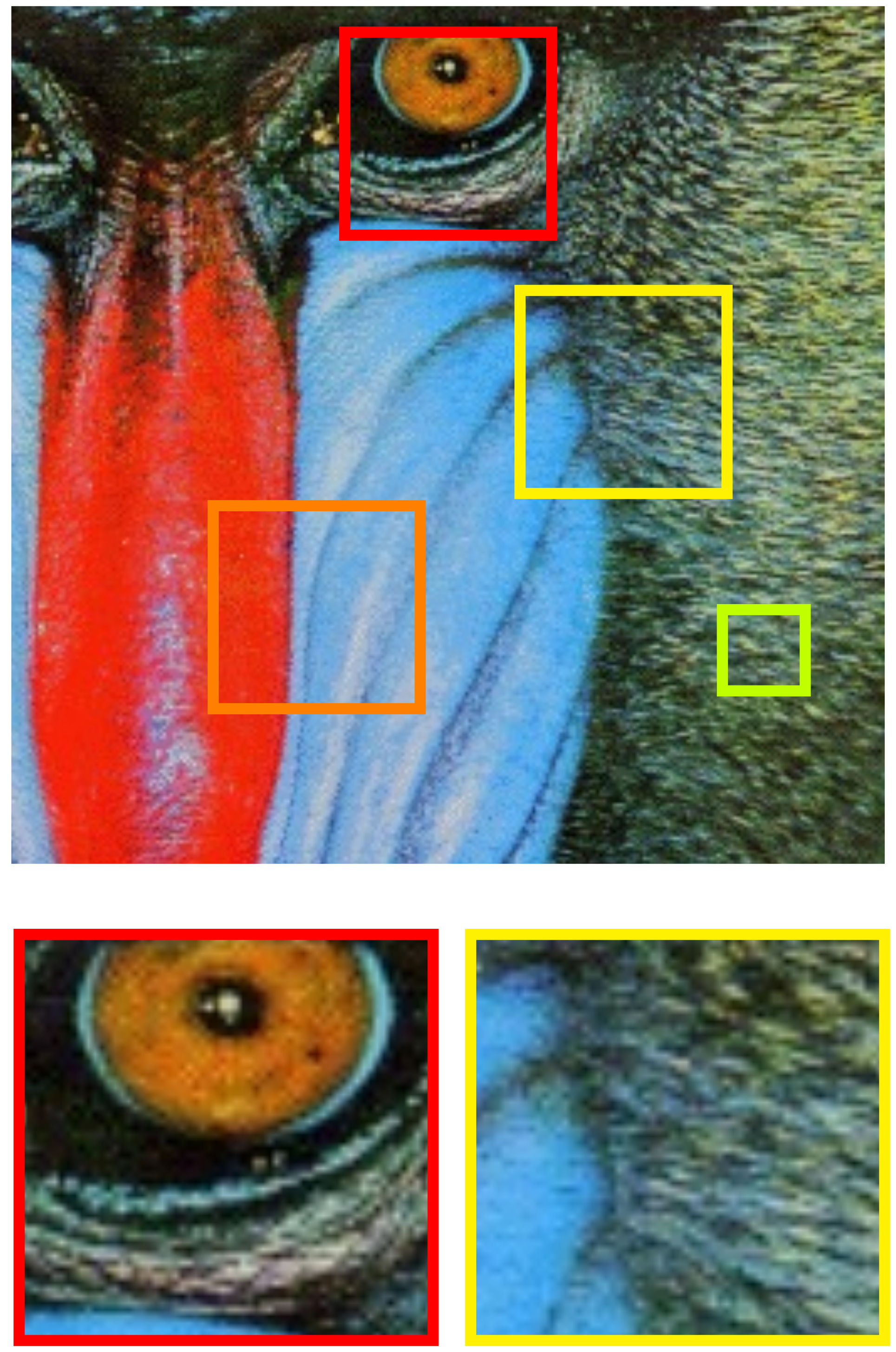
Motivation

- GAN-based super-resolution:[1]

SRGAN-MSE



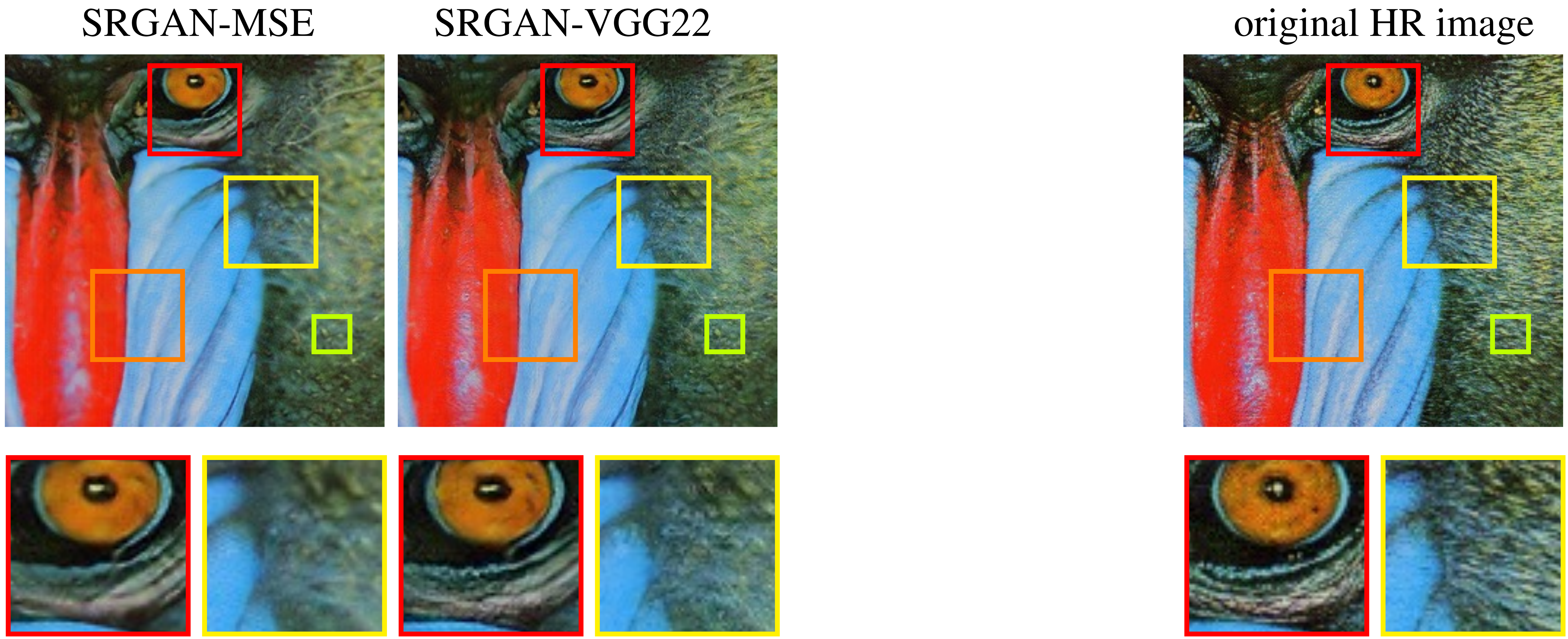
original HR image



[1] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017.

Motivation

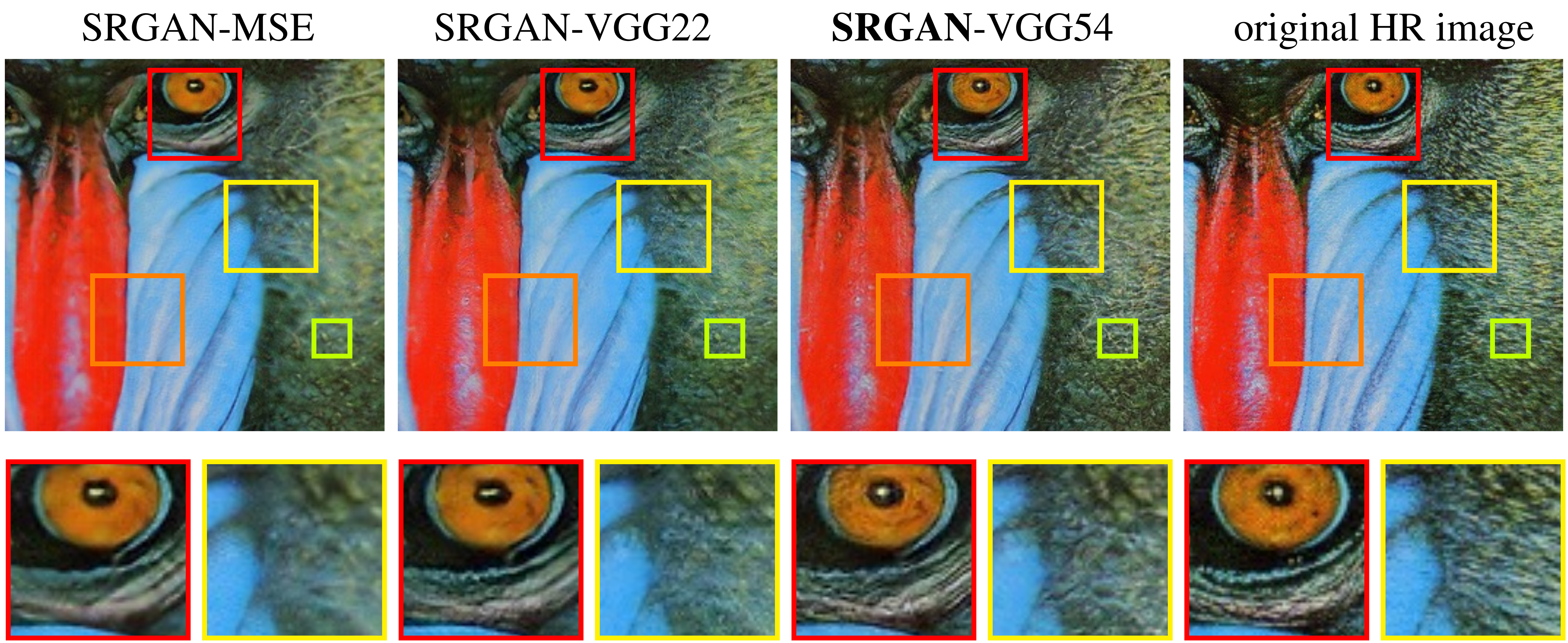
- GAN-based super-resolution:[1]



[1] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017.

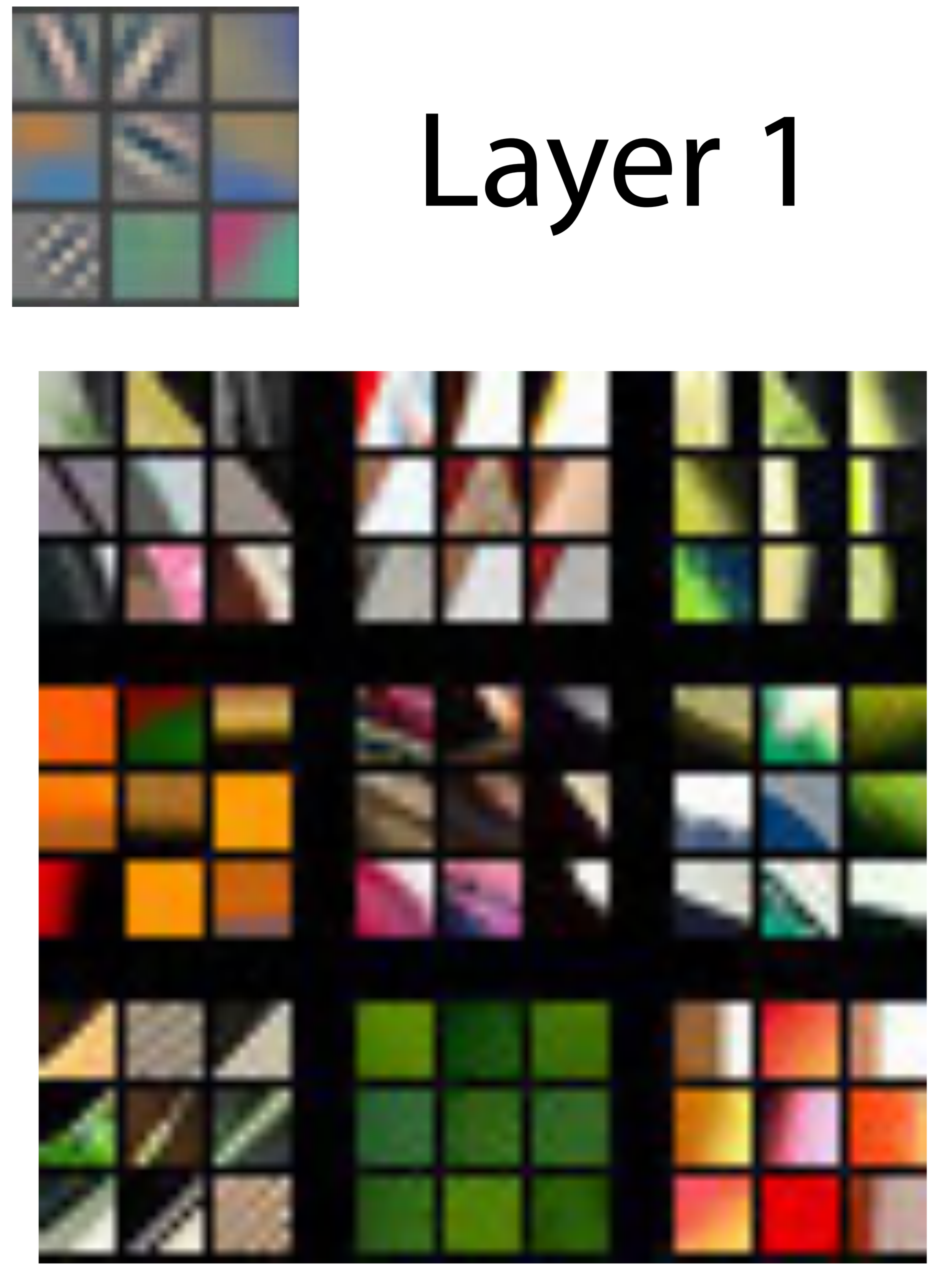
Motivation

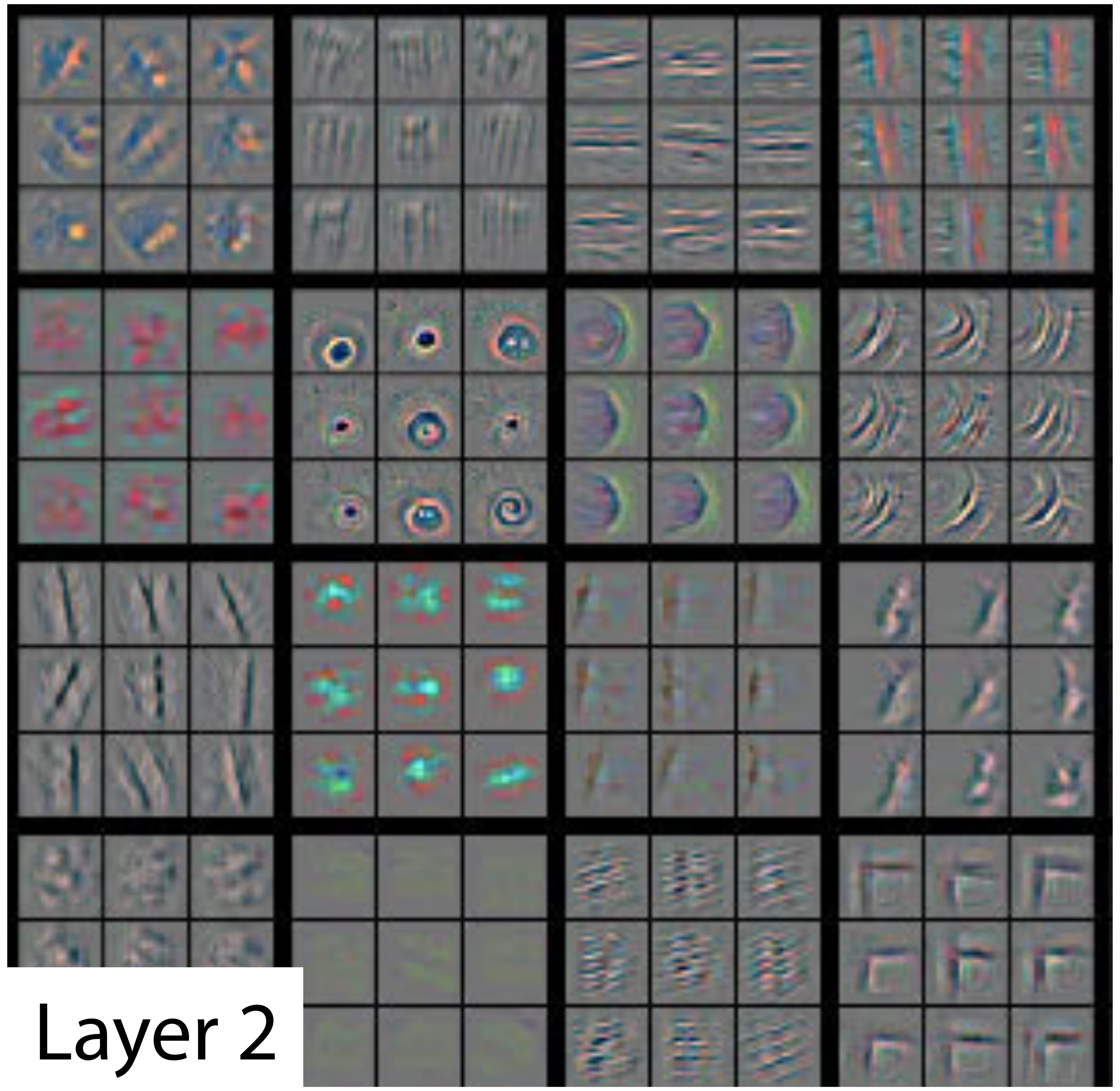
- GAN-based super-resolution:[1]



[1] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017.

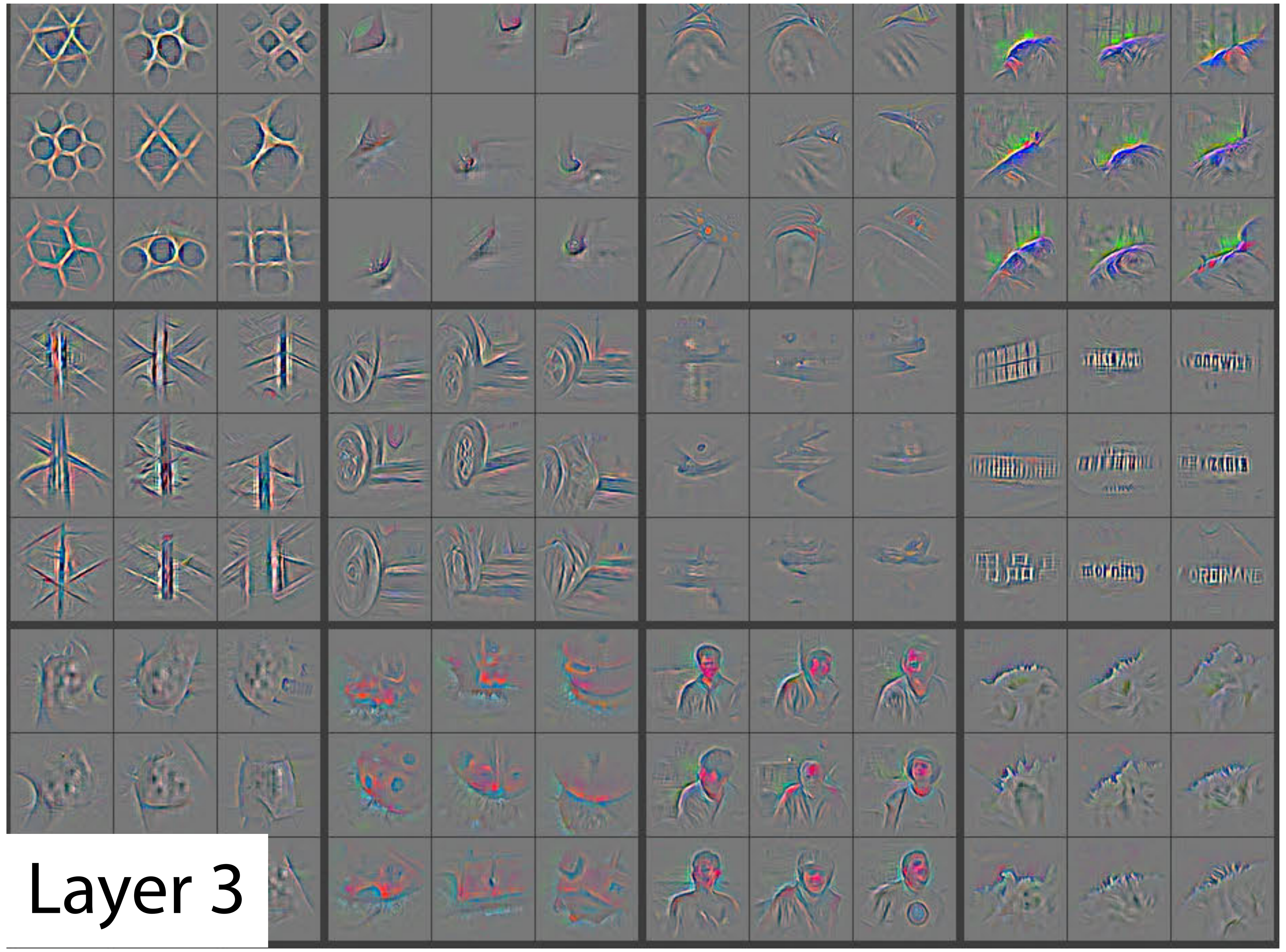
Motivation





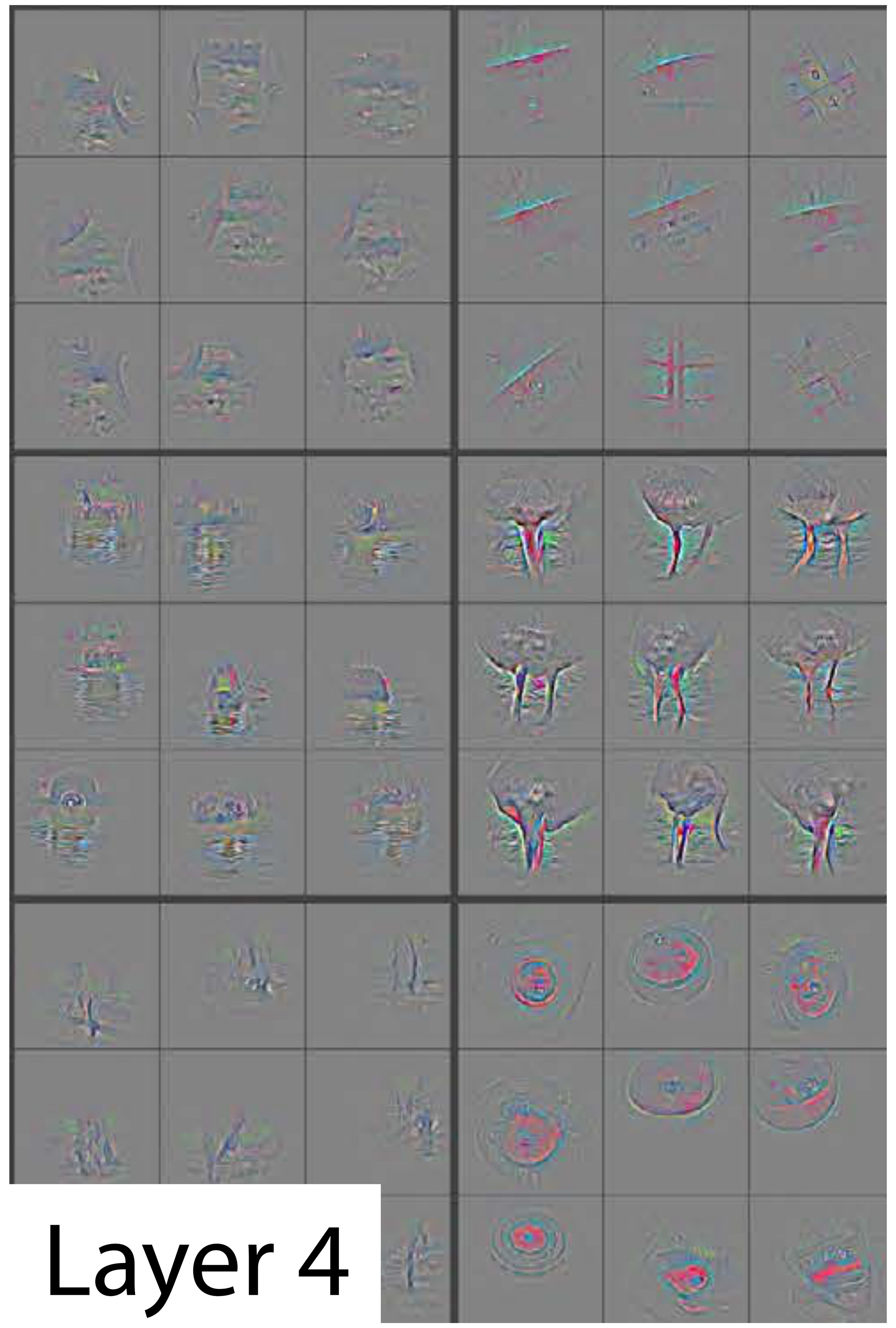
From M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional networks. In D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, editors, Computer Vision – ECCV 2014, pages 818–833, Cham, 2014. Springer International Publishing.

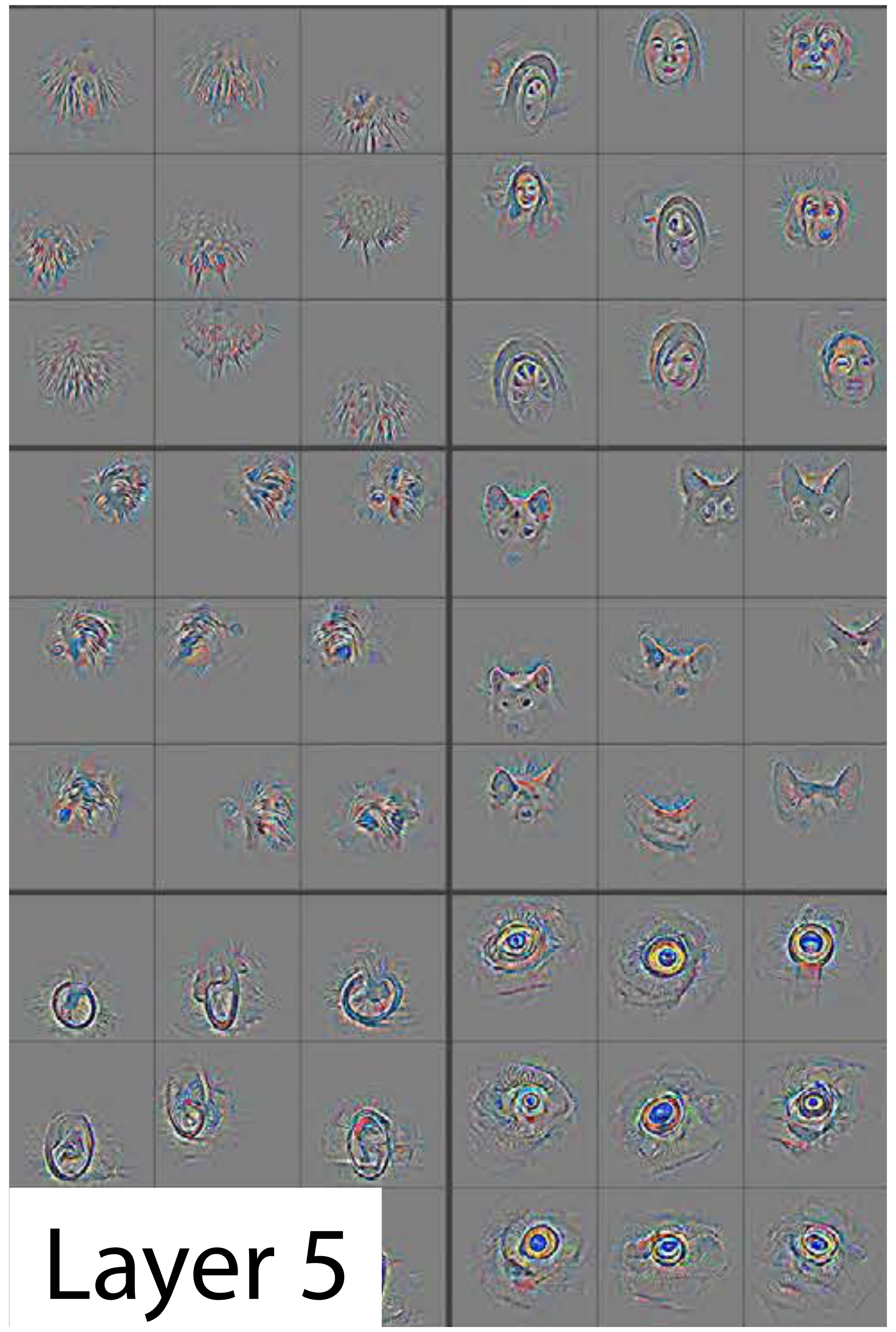
Motivation



From M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional networks. In D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, editors, Computer Vision – ECCV 2014, pages 818–833, Cham, 2014. Springer International Publishing.

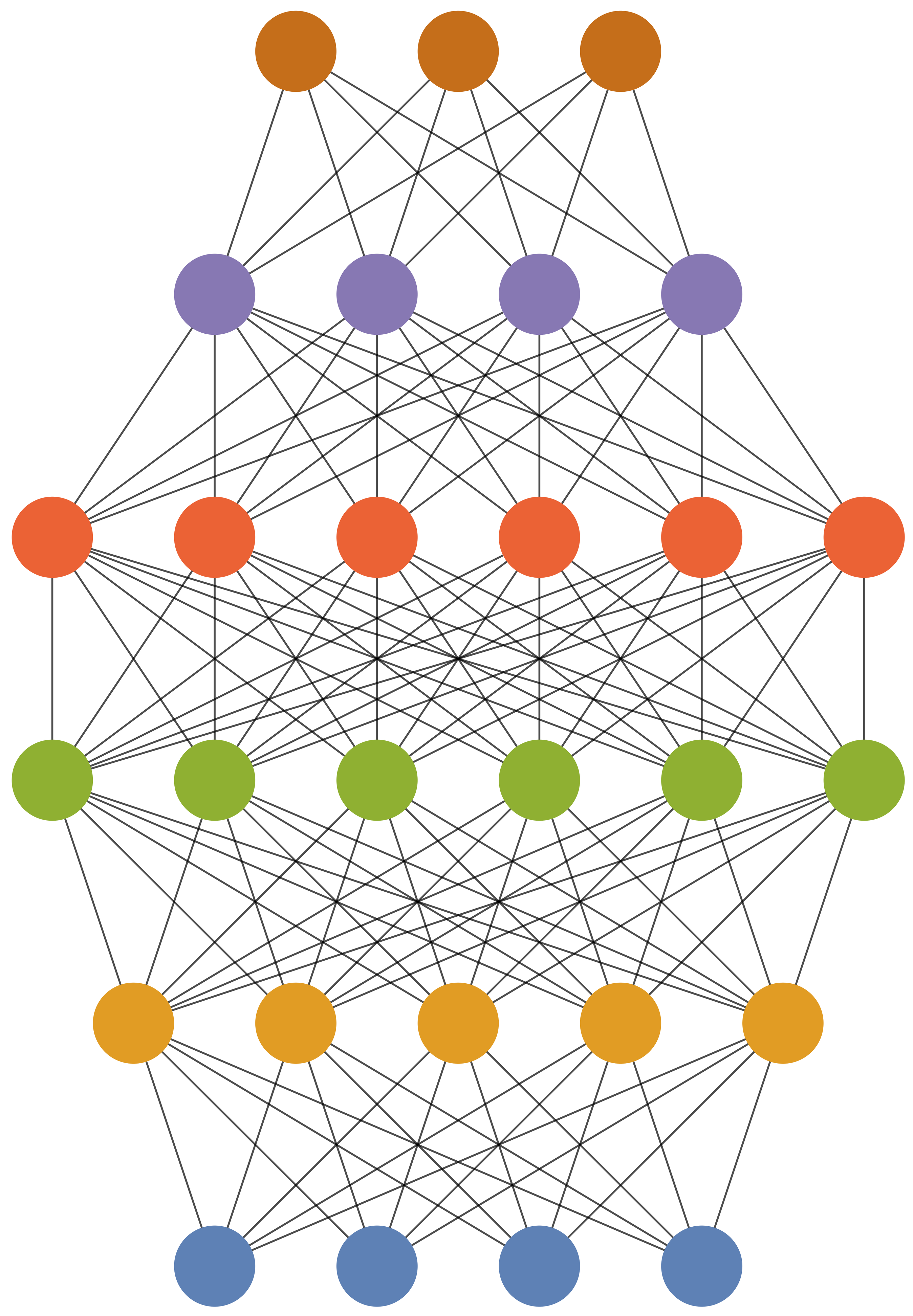
Motivation



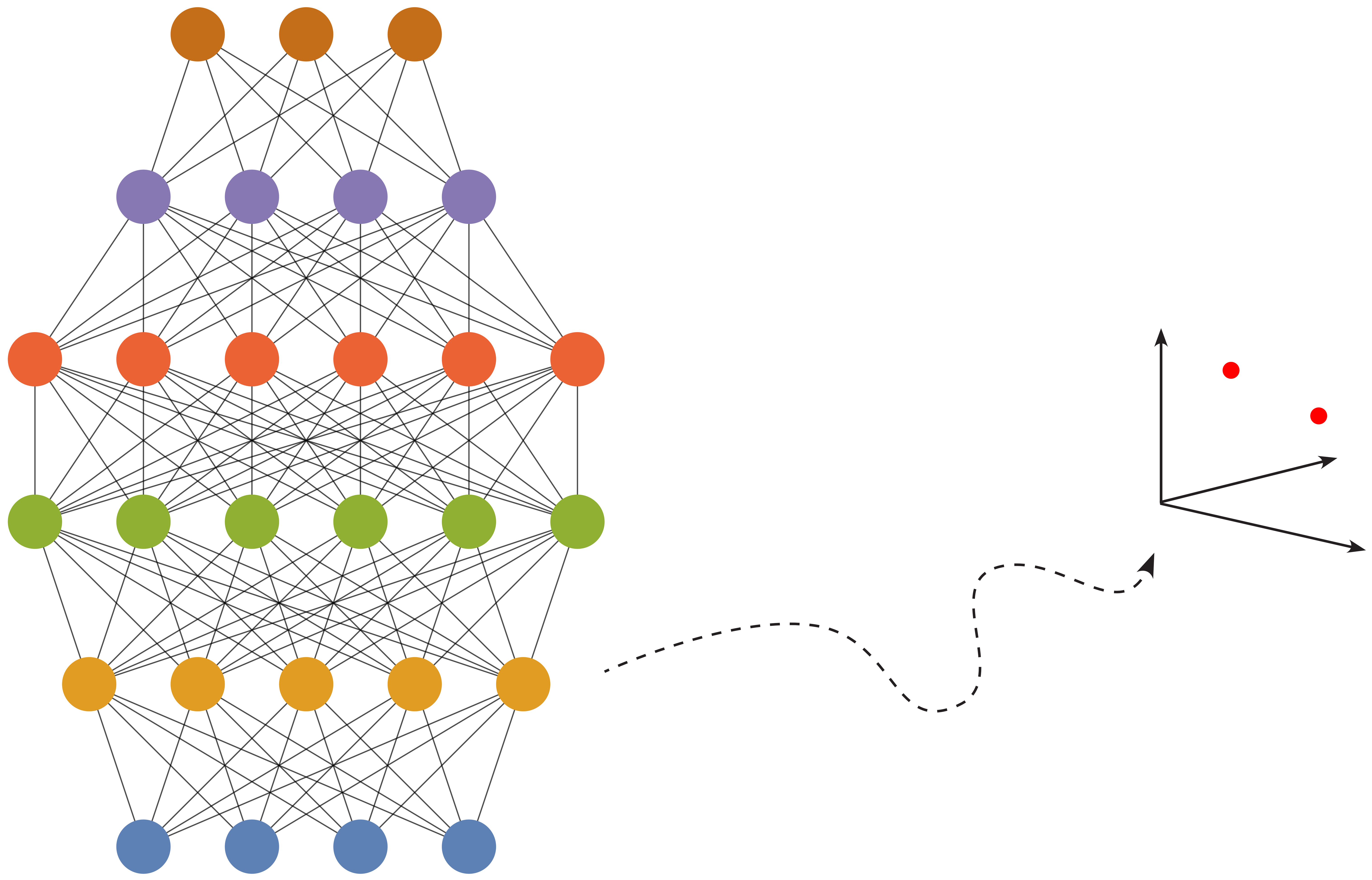


From M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional networks. In D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, editors, Computer Vision – ECCV 2014, pages 818–833, Cham, 2014. Springer International Publishing.

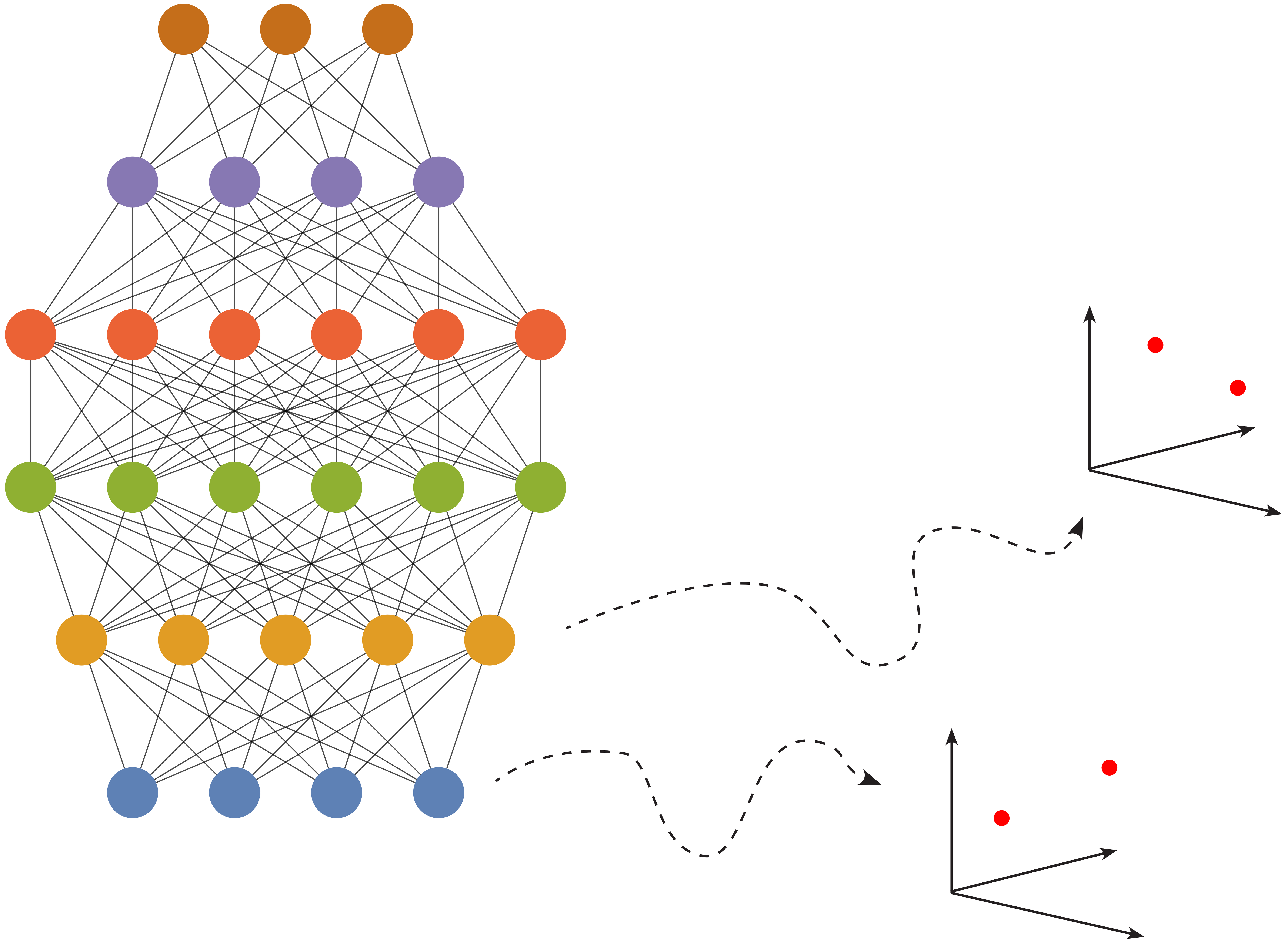
Motivation



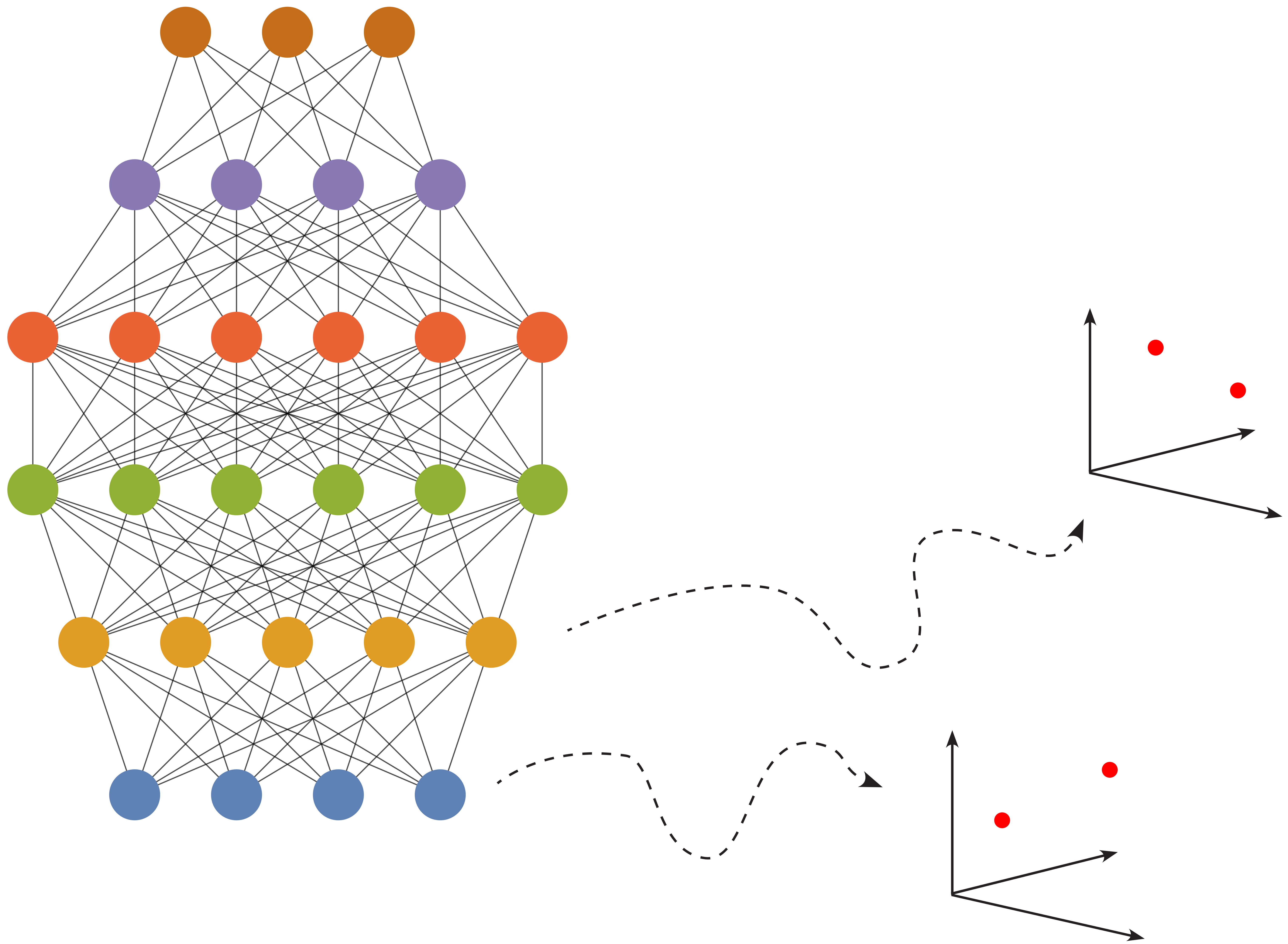
Motivation



Motivation

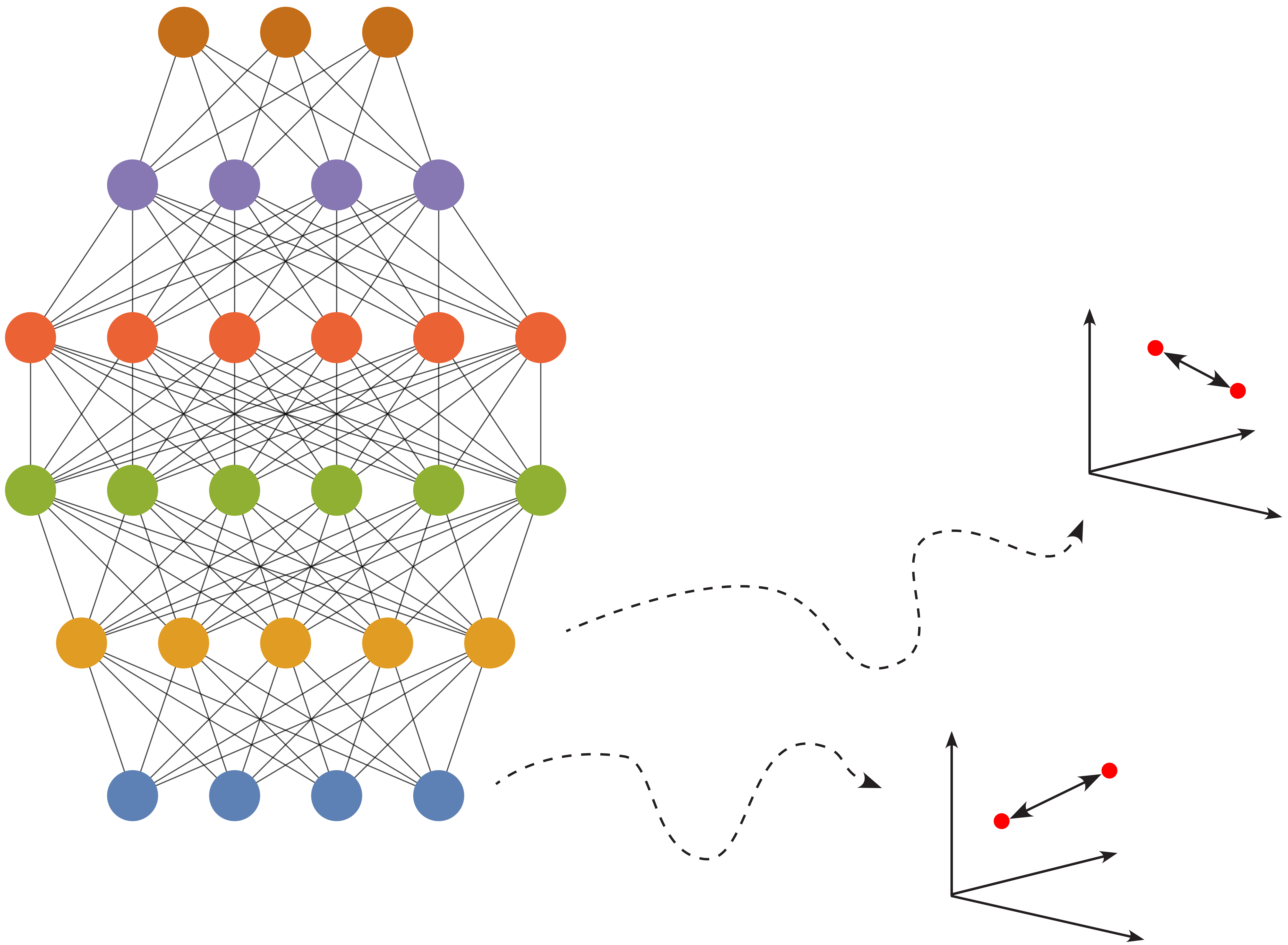


Motivation



Feature space
is task / domain
specific

Motivation

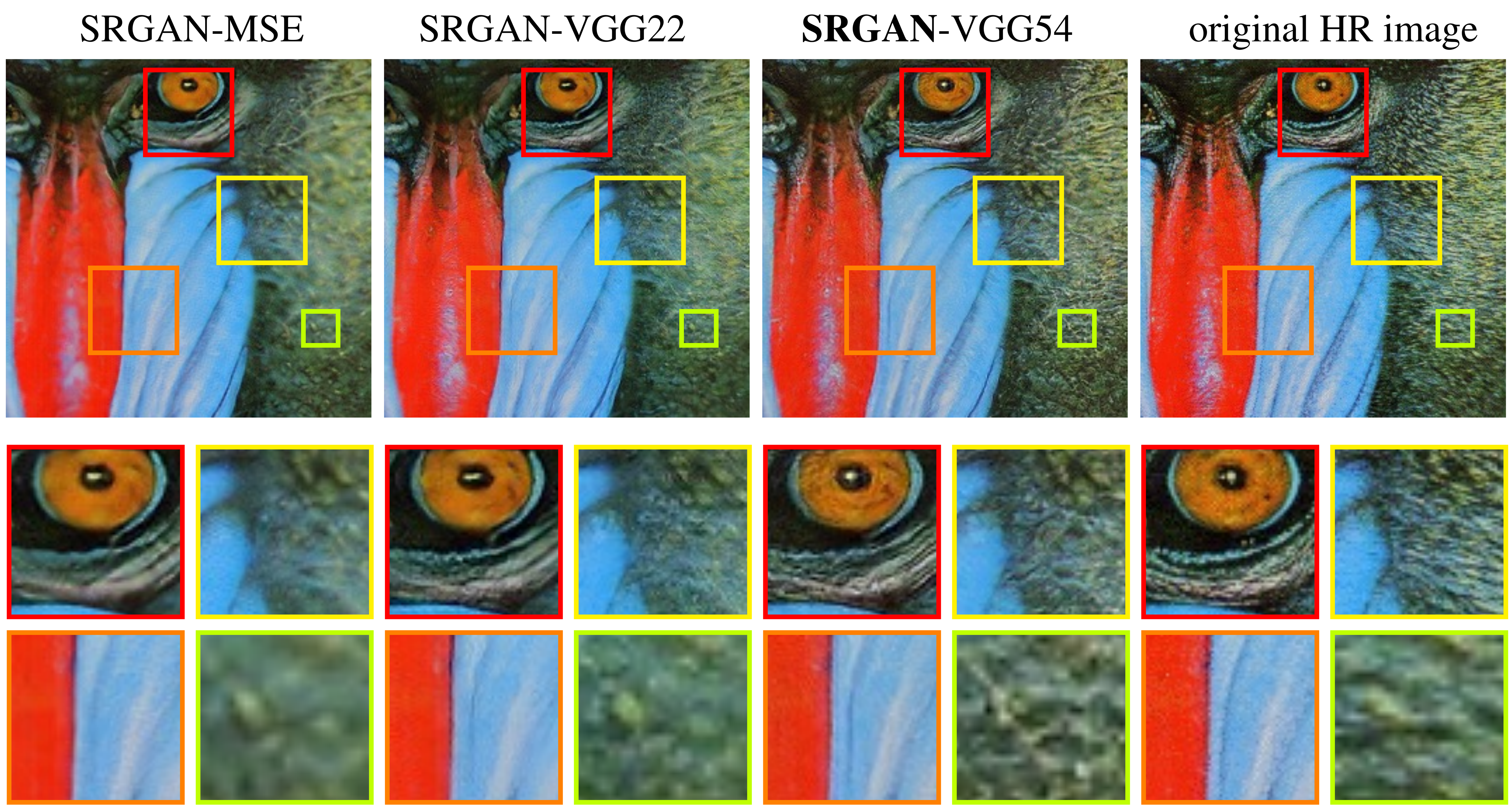


Feature space is task / domain specific

Compute norm there!

Motivation

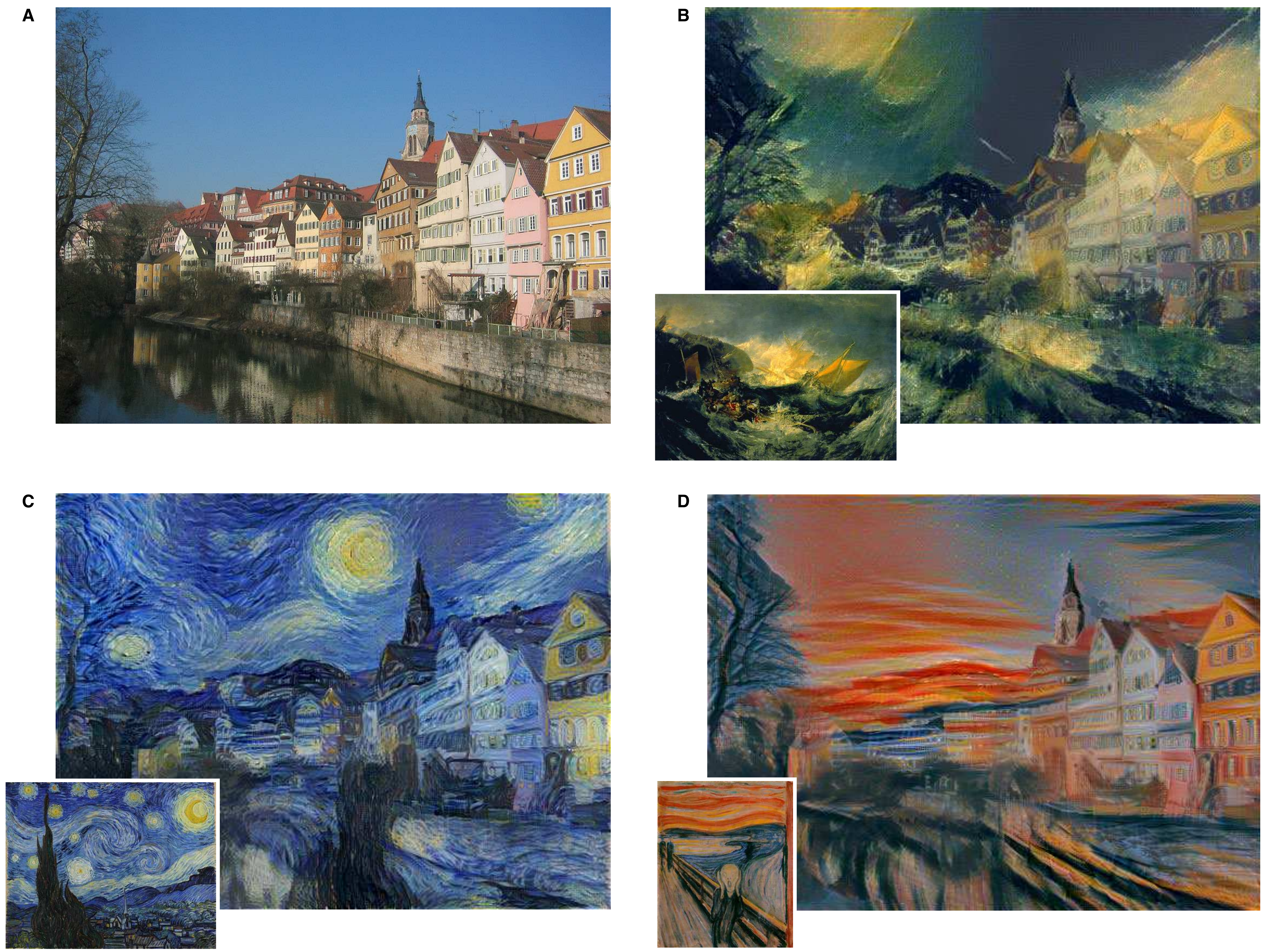
- GAN-based super-resolution:[1]



[1] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017.

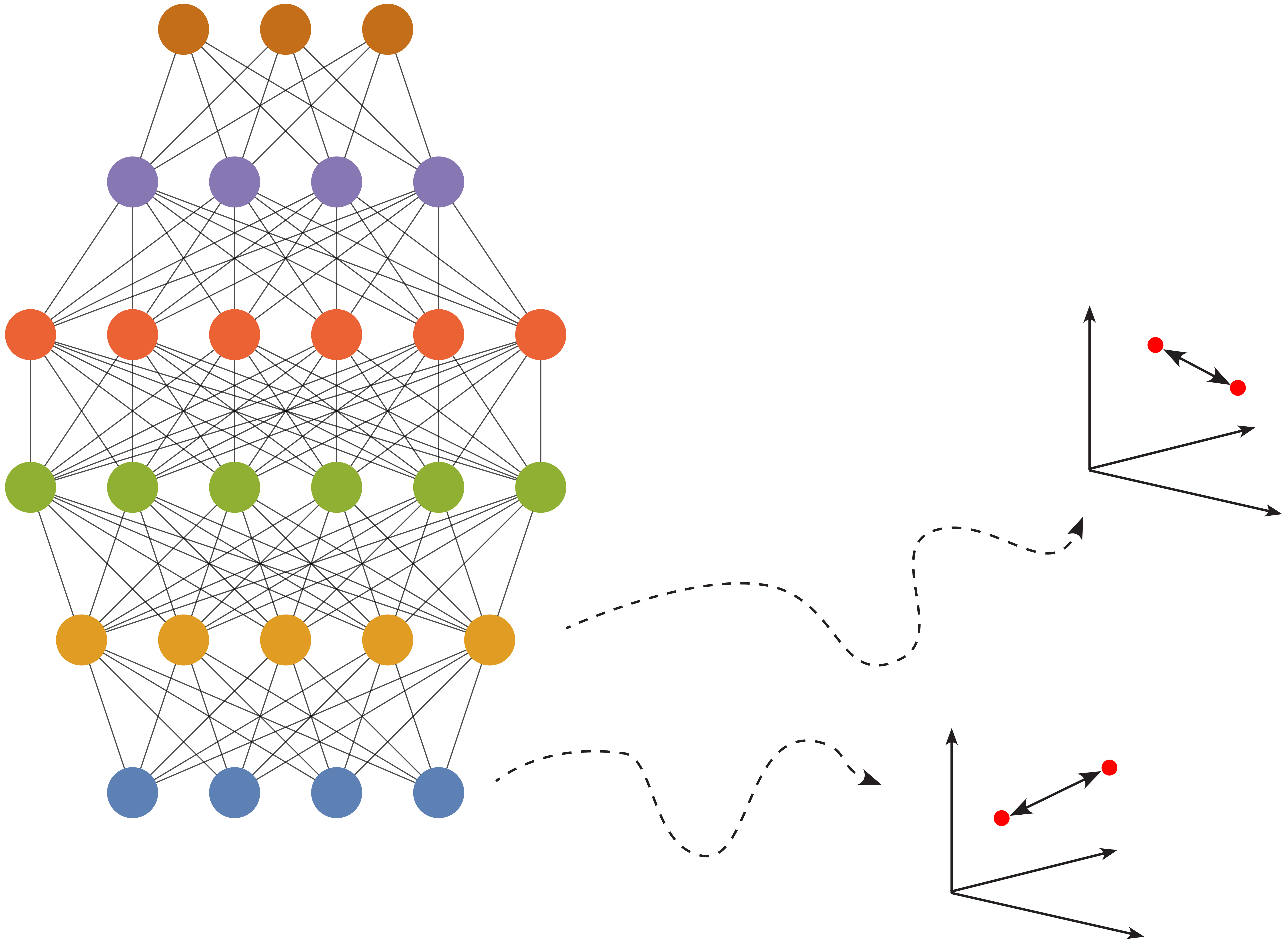
Motivation

- Neural style transfer:[1]



[1] L. A. Gatys, A. S. Ecker, and M. Bethge. Image style transfer using convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016.

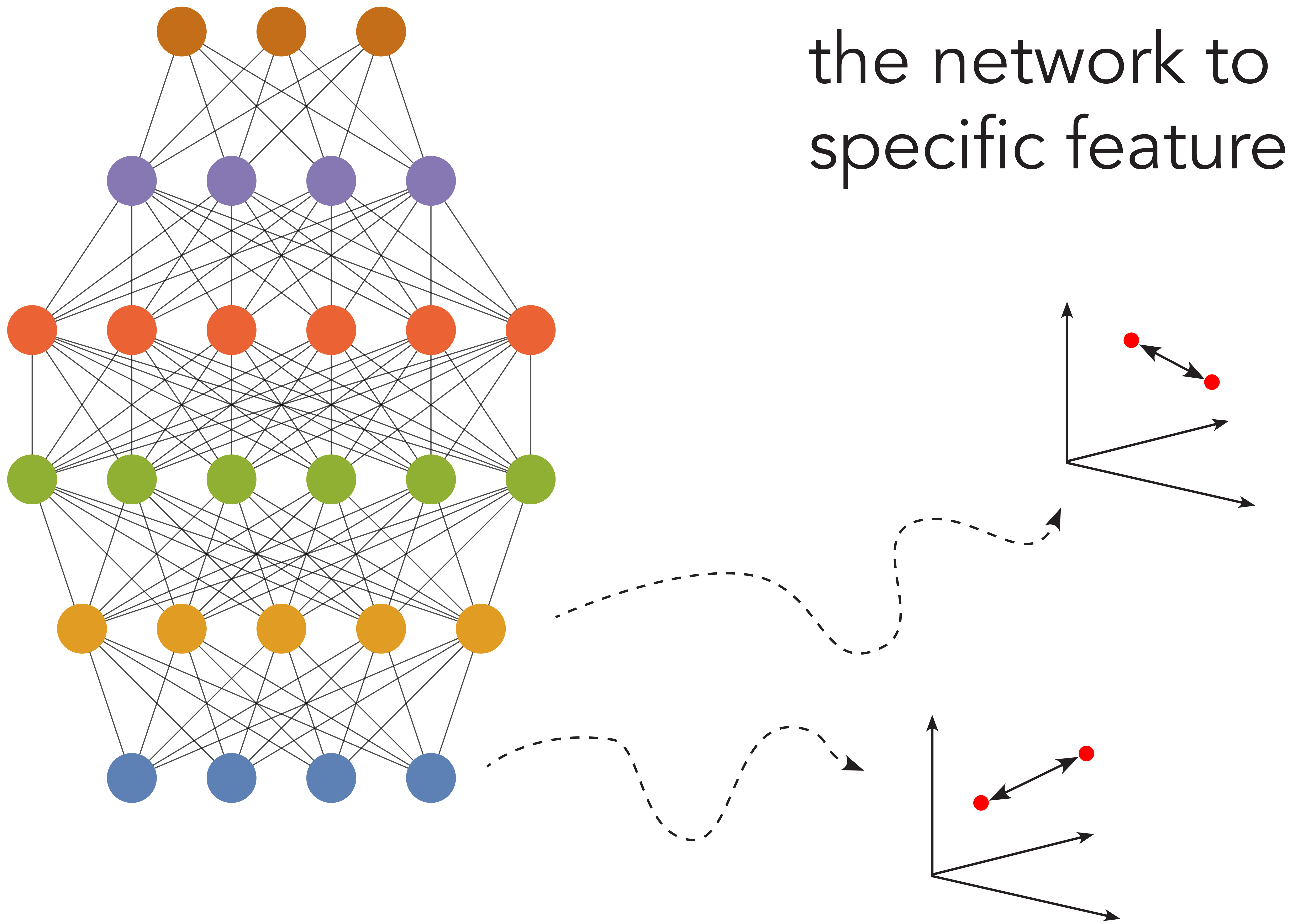
Motivation



Feature space is task / domain specific

Compute norm there!

Motivation



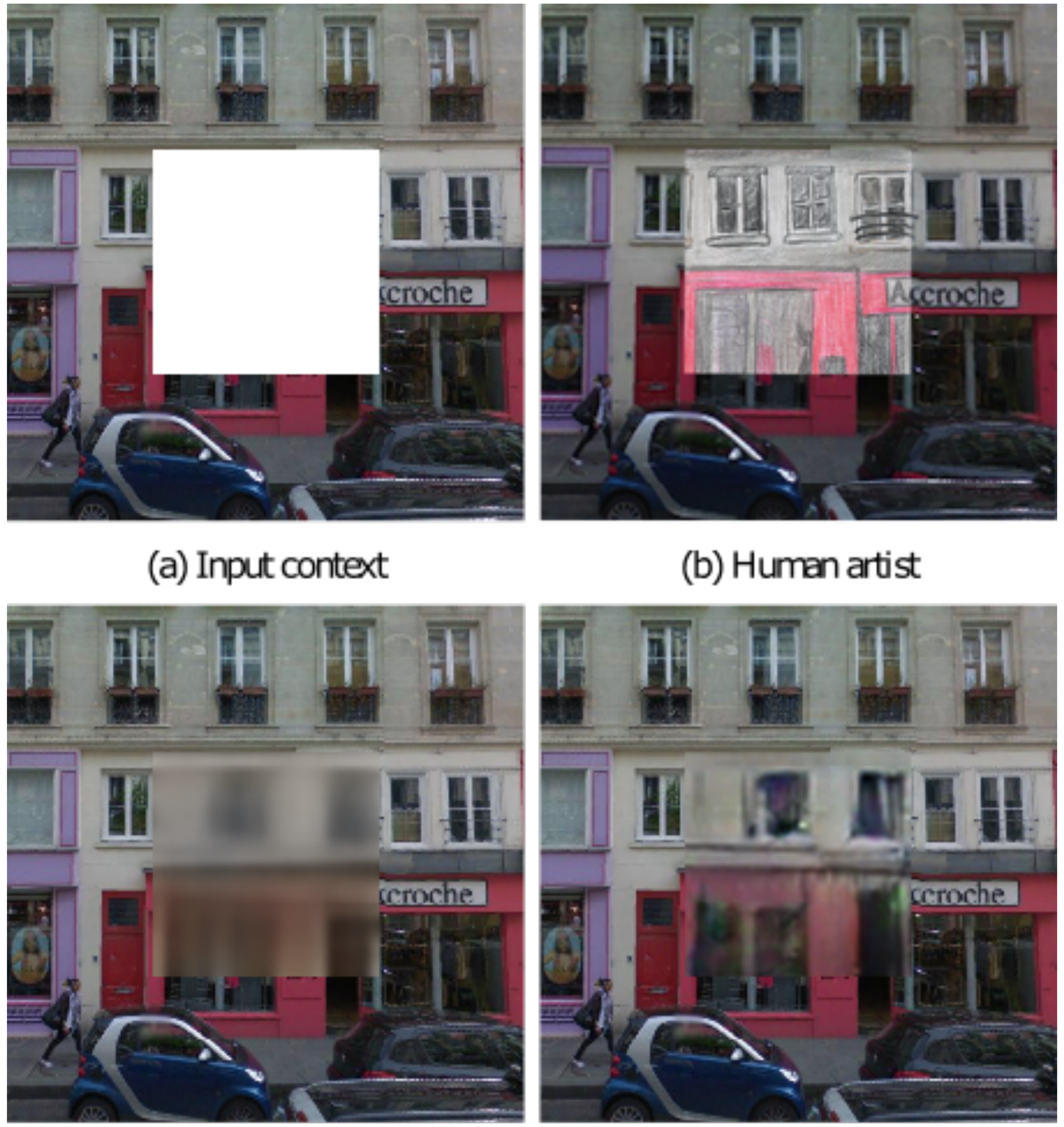
Train on *some* task that requires the network to learn domain specific features

Feature space is task / domain specific

Compute norm there!

Motivation

- Pretext task: inpainting of randomly deleted image parts[1]



(a) Input context

(b) Human artist

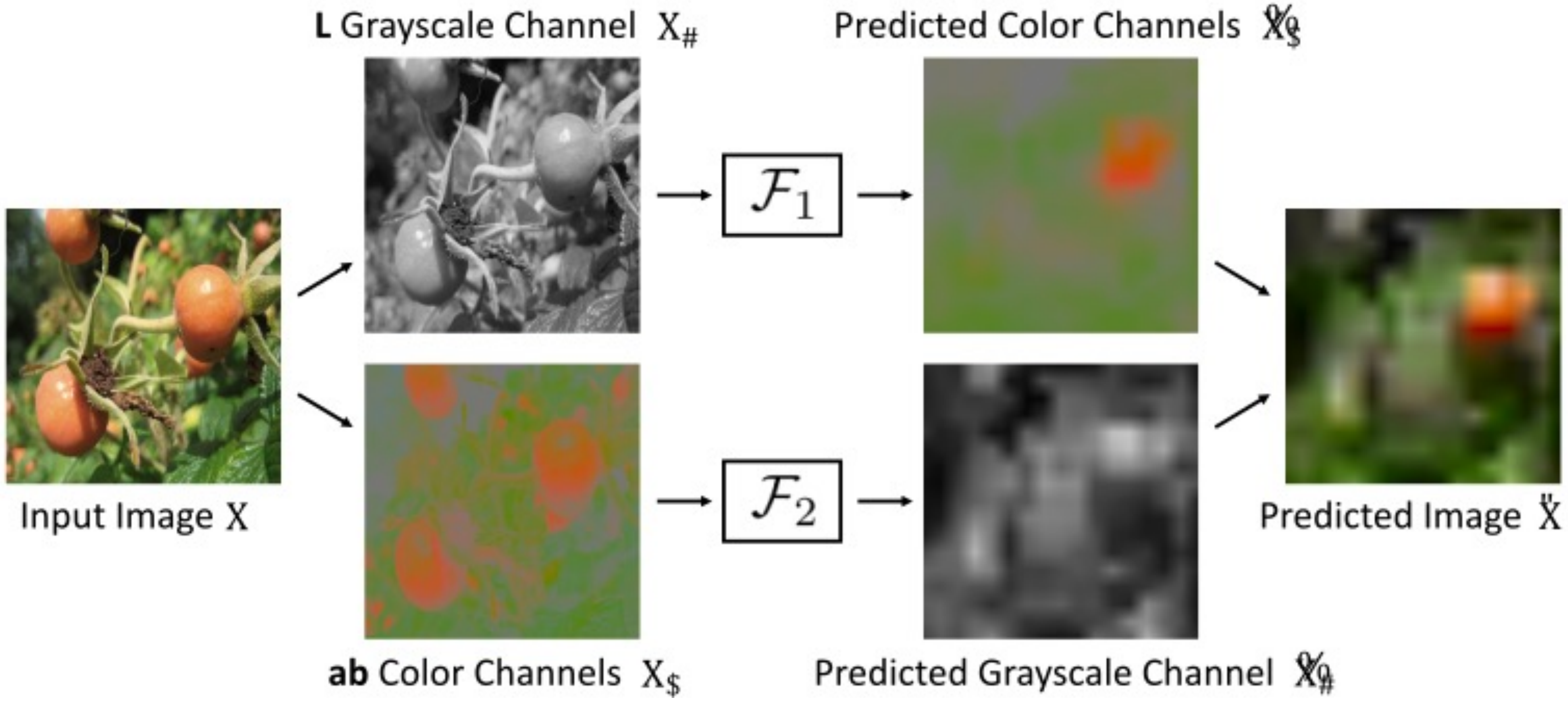
(c) Context Encoder (L 2 loss)

(d) Context Encoder (L 2 + Adversarial loss)

[1] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros. Context encoders: Feature learning by inpainting. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016.

Motivation

- Pretext task: predicting deleted color and gray scale channels[1]



[1] R. Zhang, P. Isola, and A. A. Efros. Split-brain autoencoders: Unsupervised learning by cross-channel prediction. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017.

Motivation

How can we adapt these ideas to atmospheric data?

Atmospheric data

- Wind field, vorticity, divergence, temperature, geopotential height, precipitation, ...

Atmospheric data

- Wind field, vorticity, divergence, temperature, geopotential height, precipitation, ...
- Image-like in grid representation
  - With usual issues but good starting point

Atmospheric data

- Wind field, vorticity, divergence, temperature, geopotential height, precipitation, ...
- Image-like in grid representation
  - With usual issues but good starting point
- ERA5 provides well curated data set for training
  - Contains effects we cannot model

Atmospheric data

- Wind field, vorticity, divergence, temperature, geopotential height, precipitation, …
- Image-like in grid representation
  - With usual issues but good starting point
- ERA5 provides well curated data set for training
  - Contains effects we cannot model
  - But unlabelled

AtmoDist[1]

- Custom distance metric for vorticity + divergence (wind velocity vector field)

[1] S. Hoffmann and C. Lessig. Towards representation learning for atmospheric data. In NEURIPS 2021 Workshop on Climate Change (poster), 2021.

AtmoDist[1]

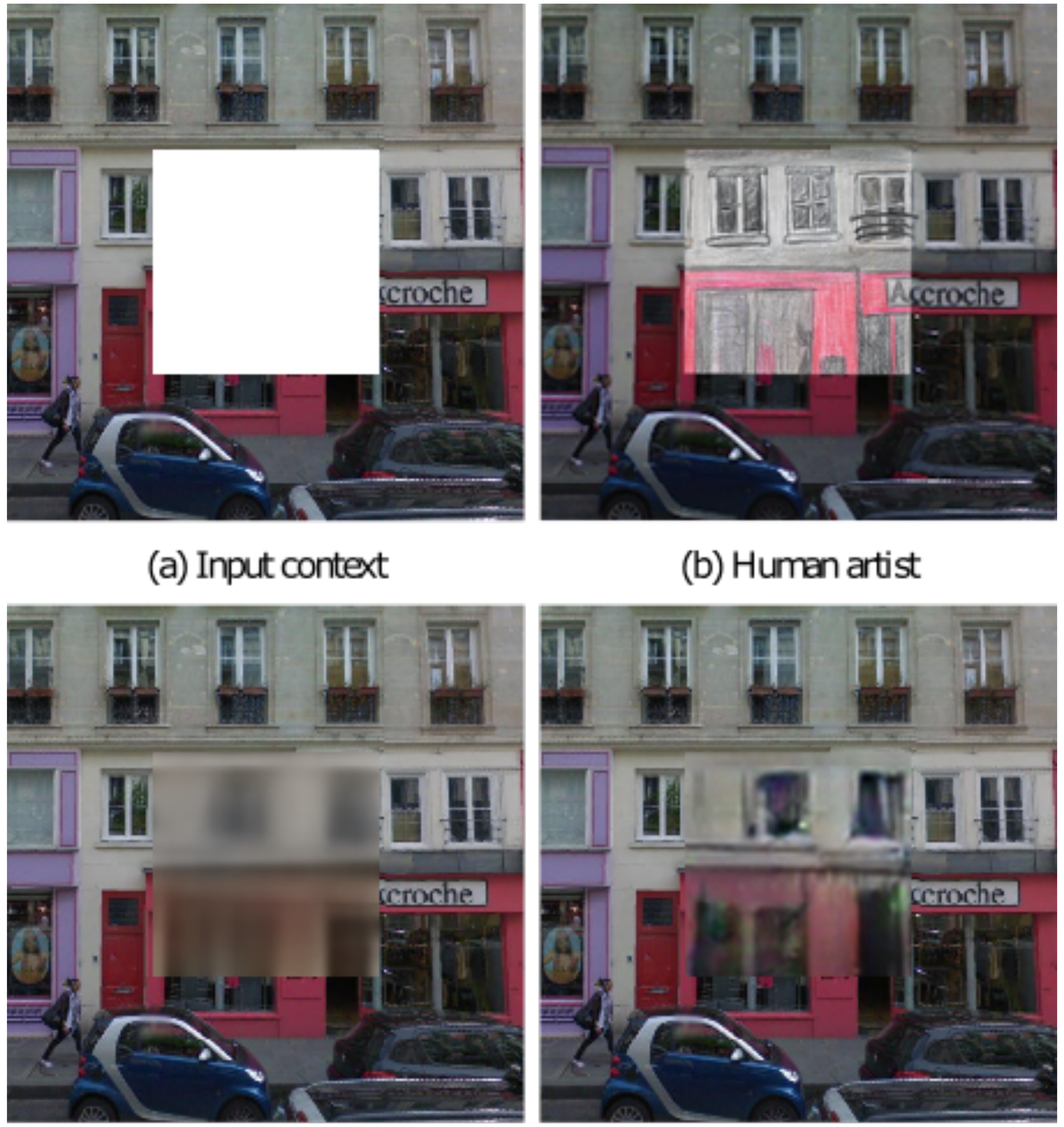
- Custom distance metric for vorticity + divergence (wind velocity vector field)
- GAN-based super-resolution / downscaling as validation application
  - Recent work by Stengel et al.[2] as baseline

[1] S. Hoffmann and C. Lessig. Towards representation learning for atmospheric data. In NEURIPS 2021 Workshop on Climate Change (poster), 2021.

[2] K. Stengel, A. Glaws, D. Hettinger, and R. N. King. Adversarial super-resolution of climatological wind and solar data. Proceedings of the National Academy of Sciences, 117(29):16805–16815, 2020.

AtmoDist

What pretext task can we use?

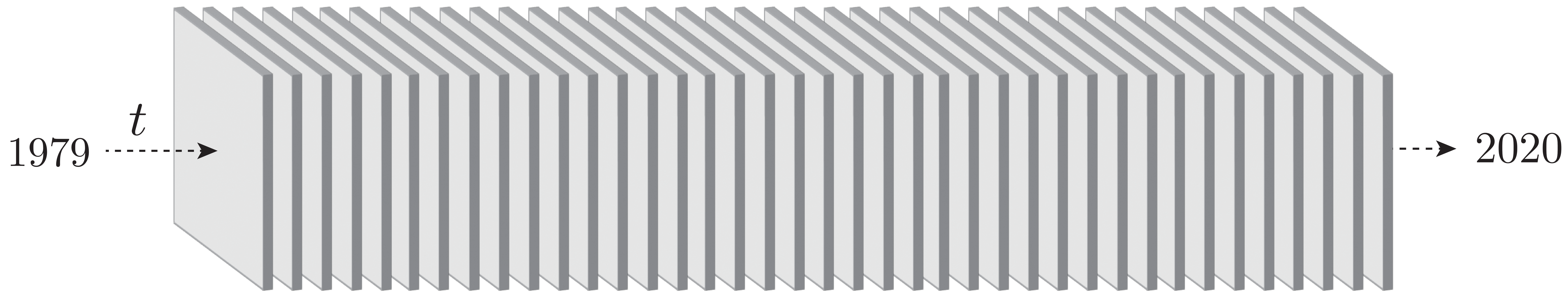


(a) Input context (b) Human artist

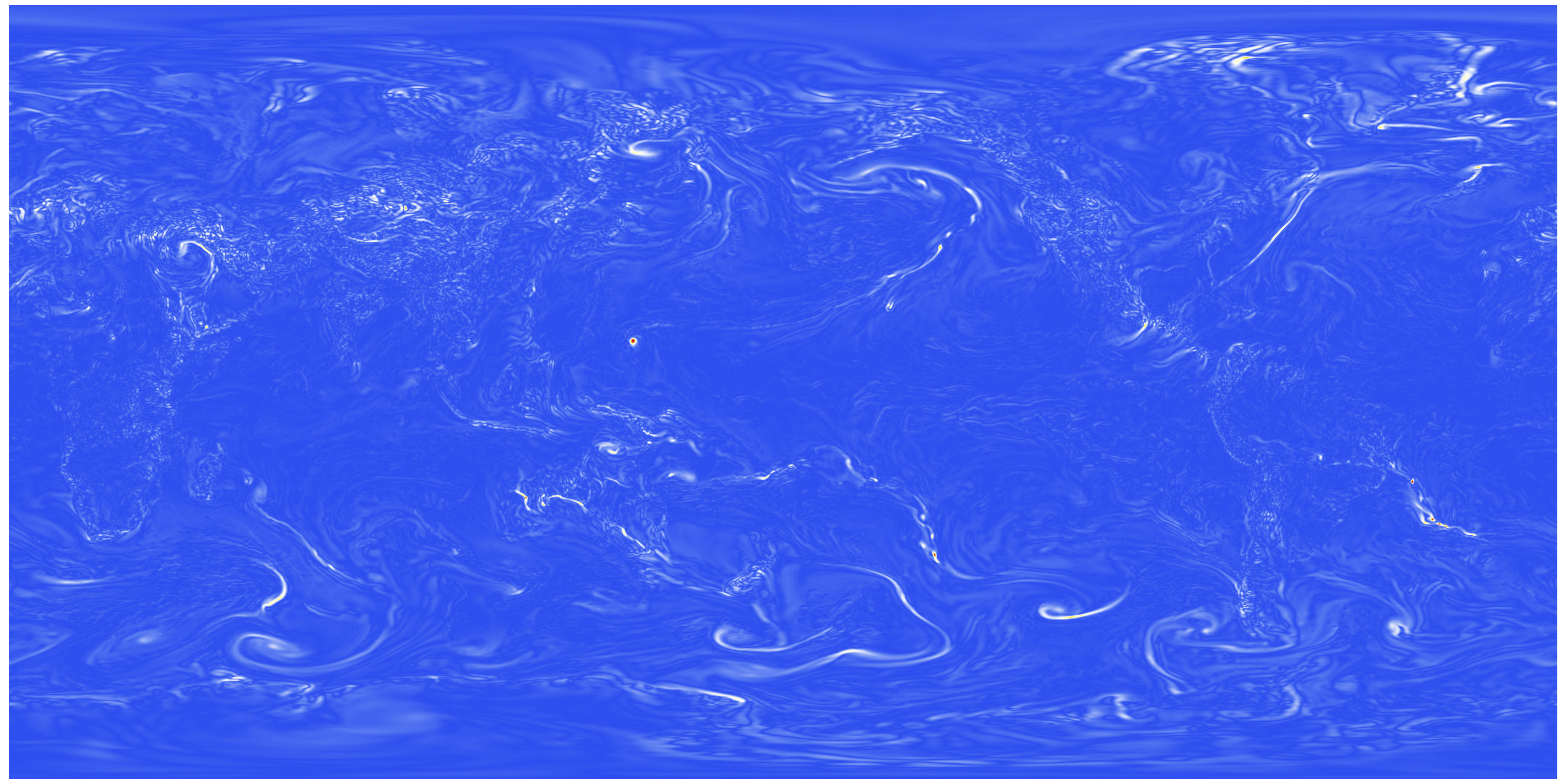
(c) Context Encoder (L2 loss) (d) Context Encoder (L2 + Adversarial loss)

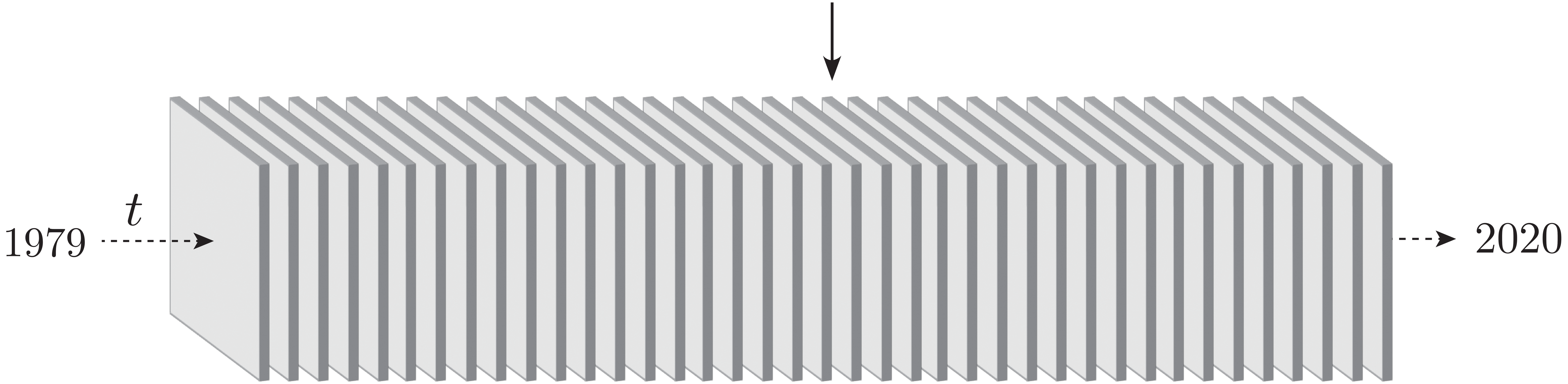
[1] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros. Context encoders: Feature learning by inpainting. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016.

AtmoDist

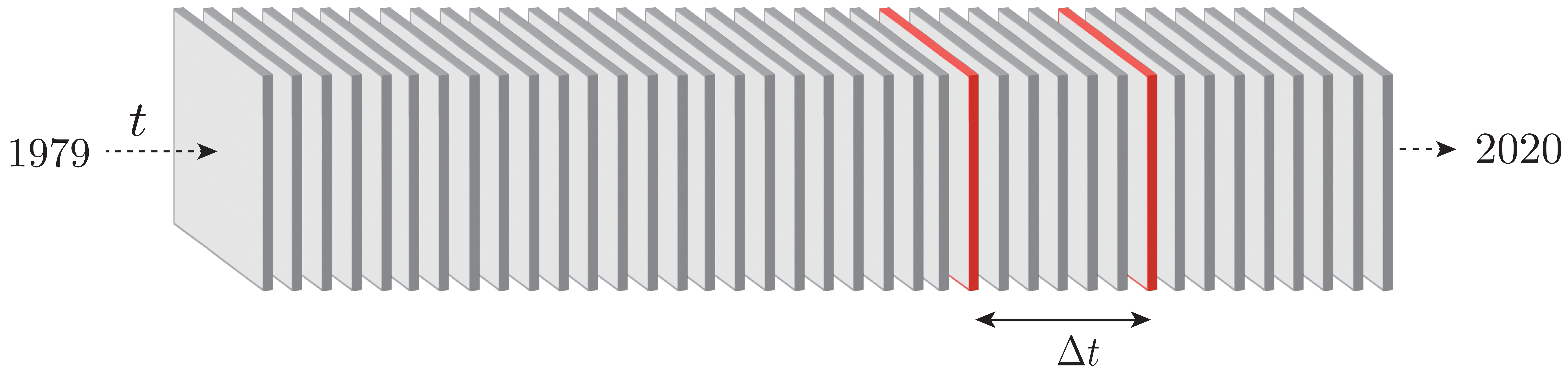


AtmoDist

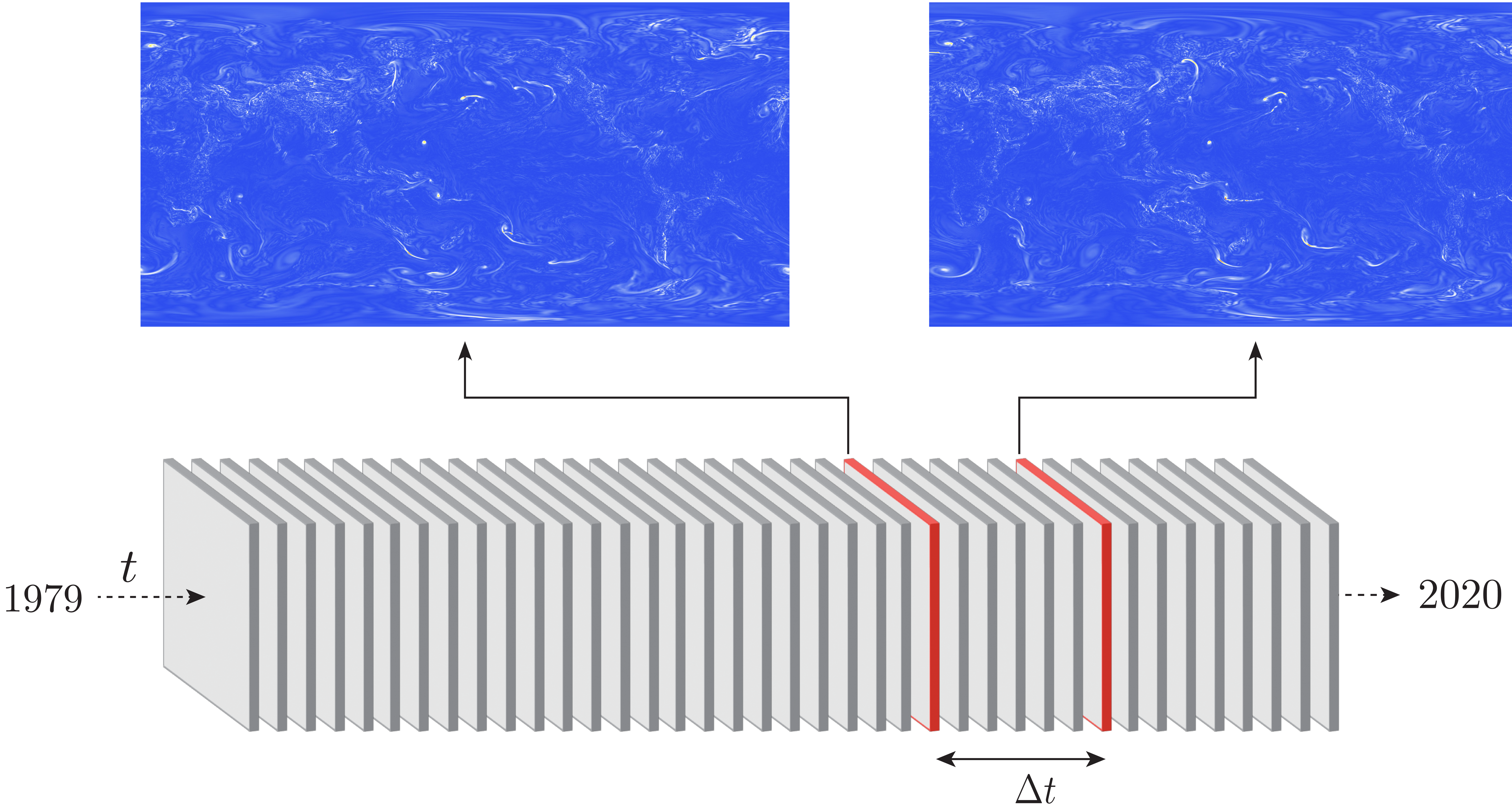




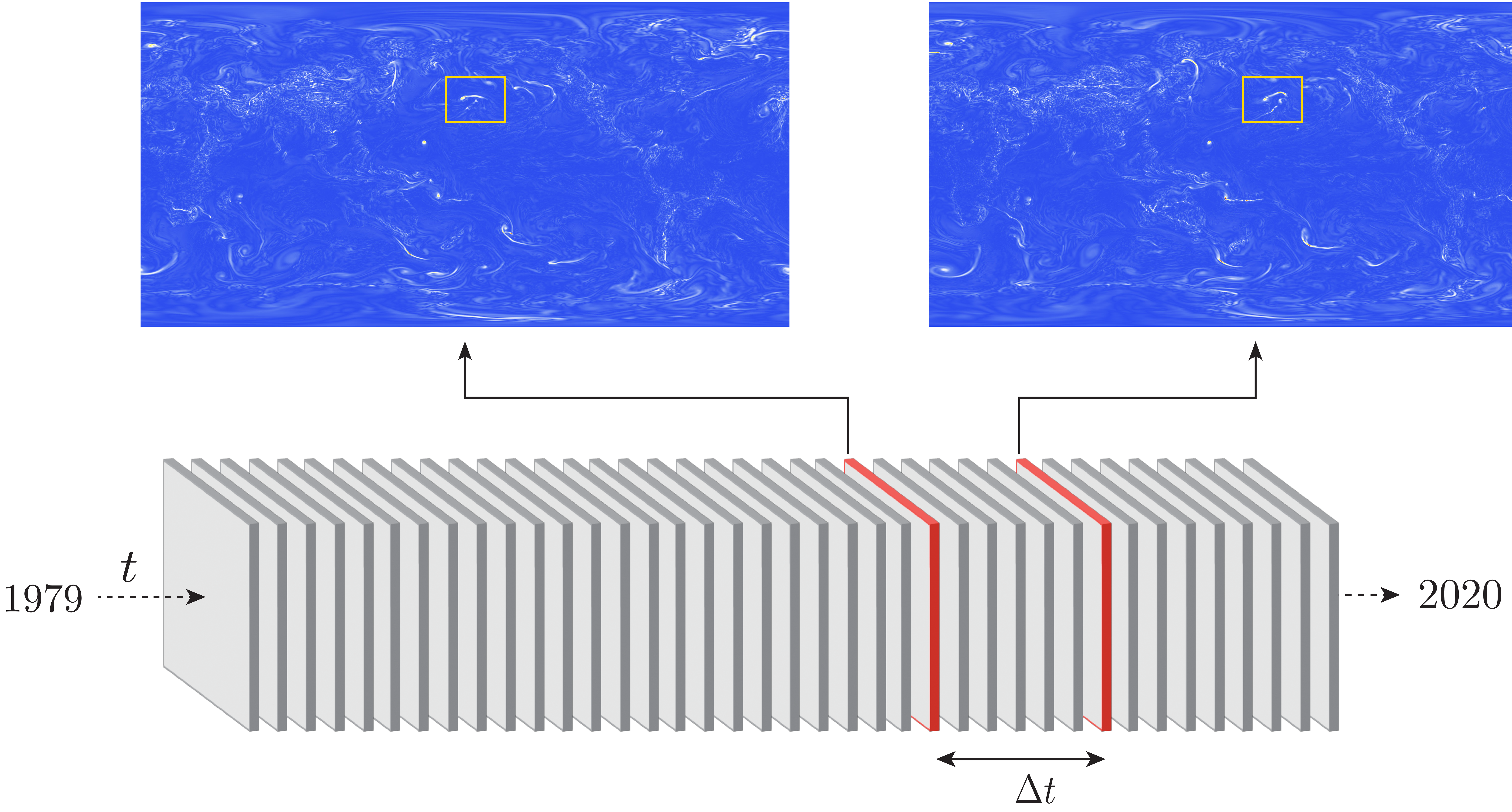
AtmoDist



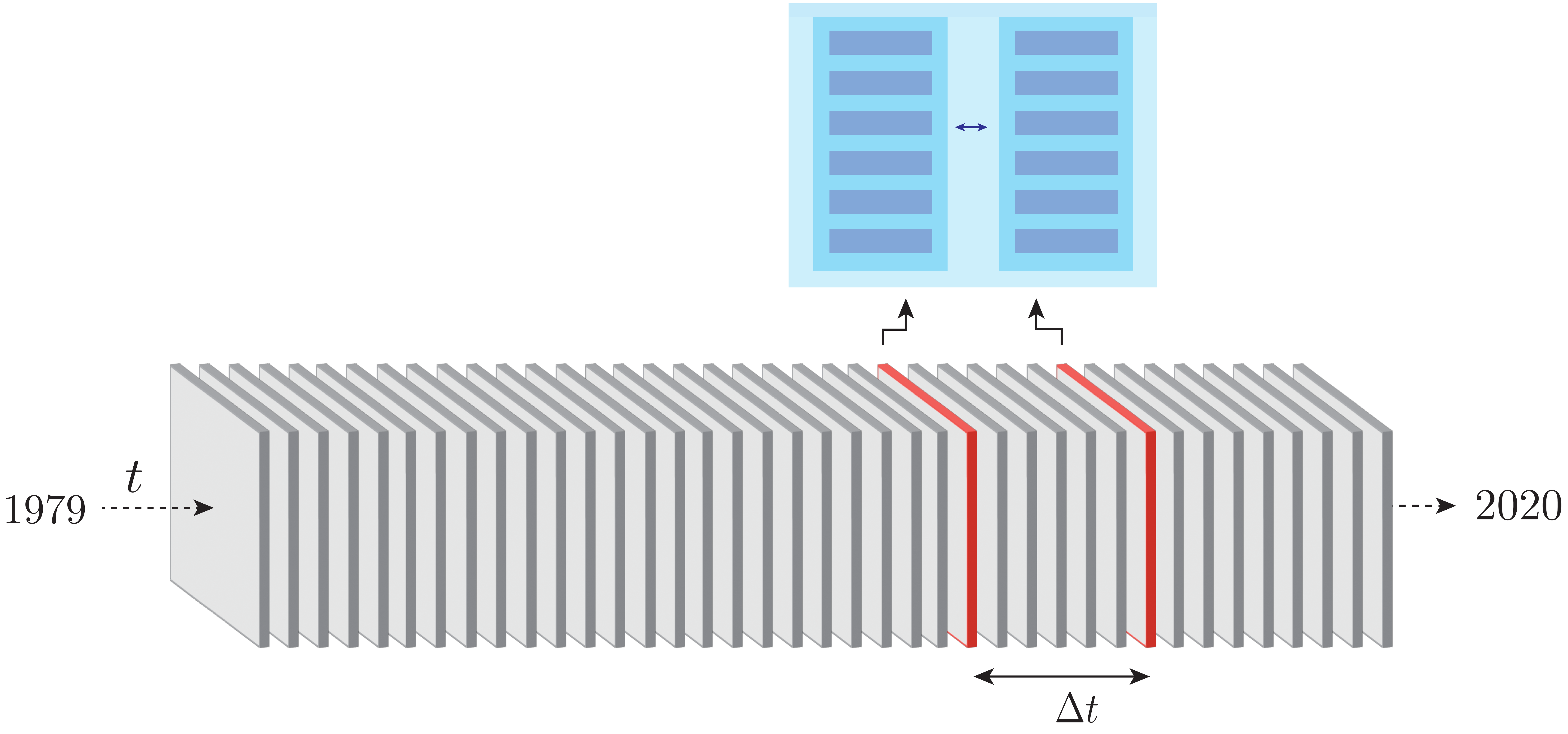
AtmoDist



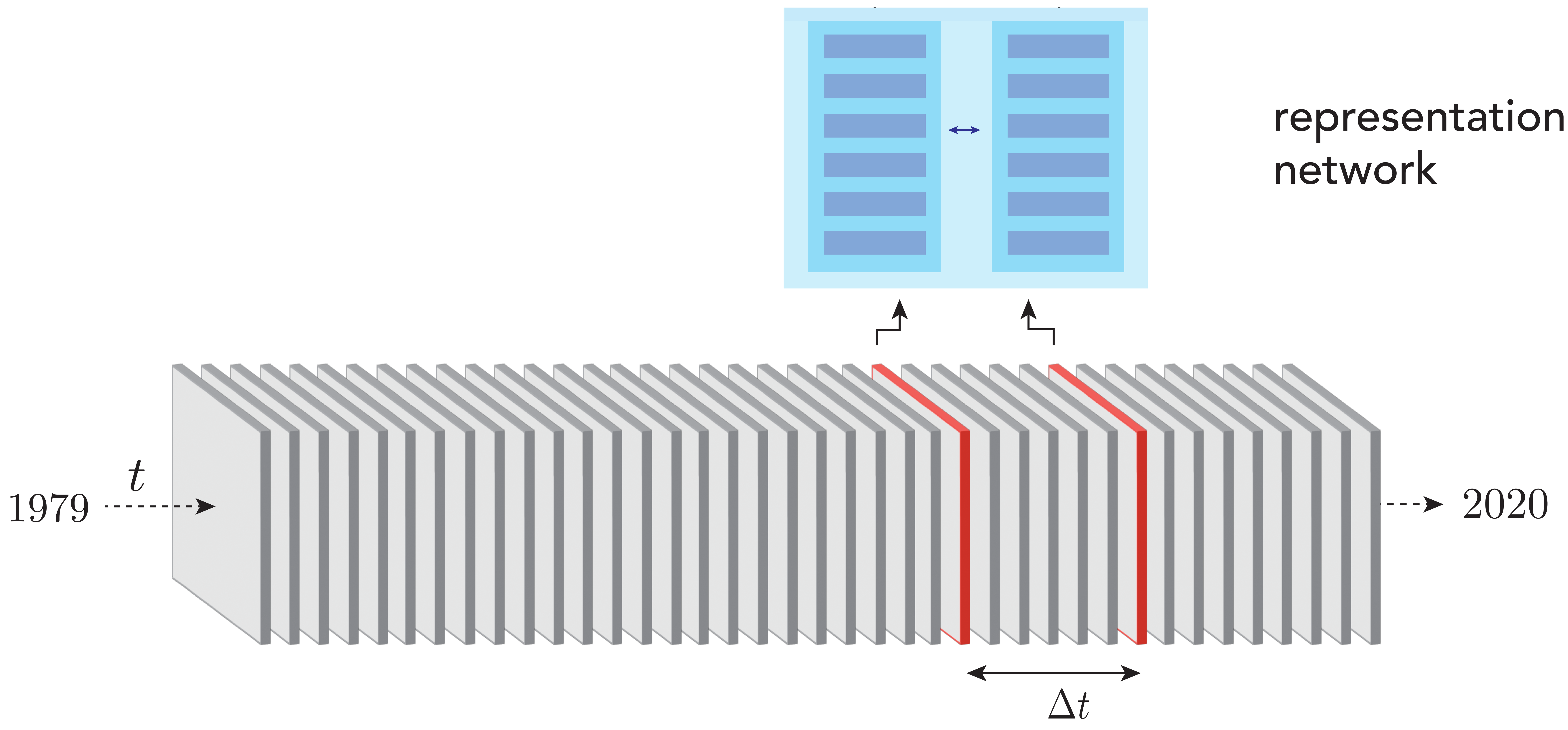
AtmoDist



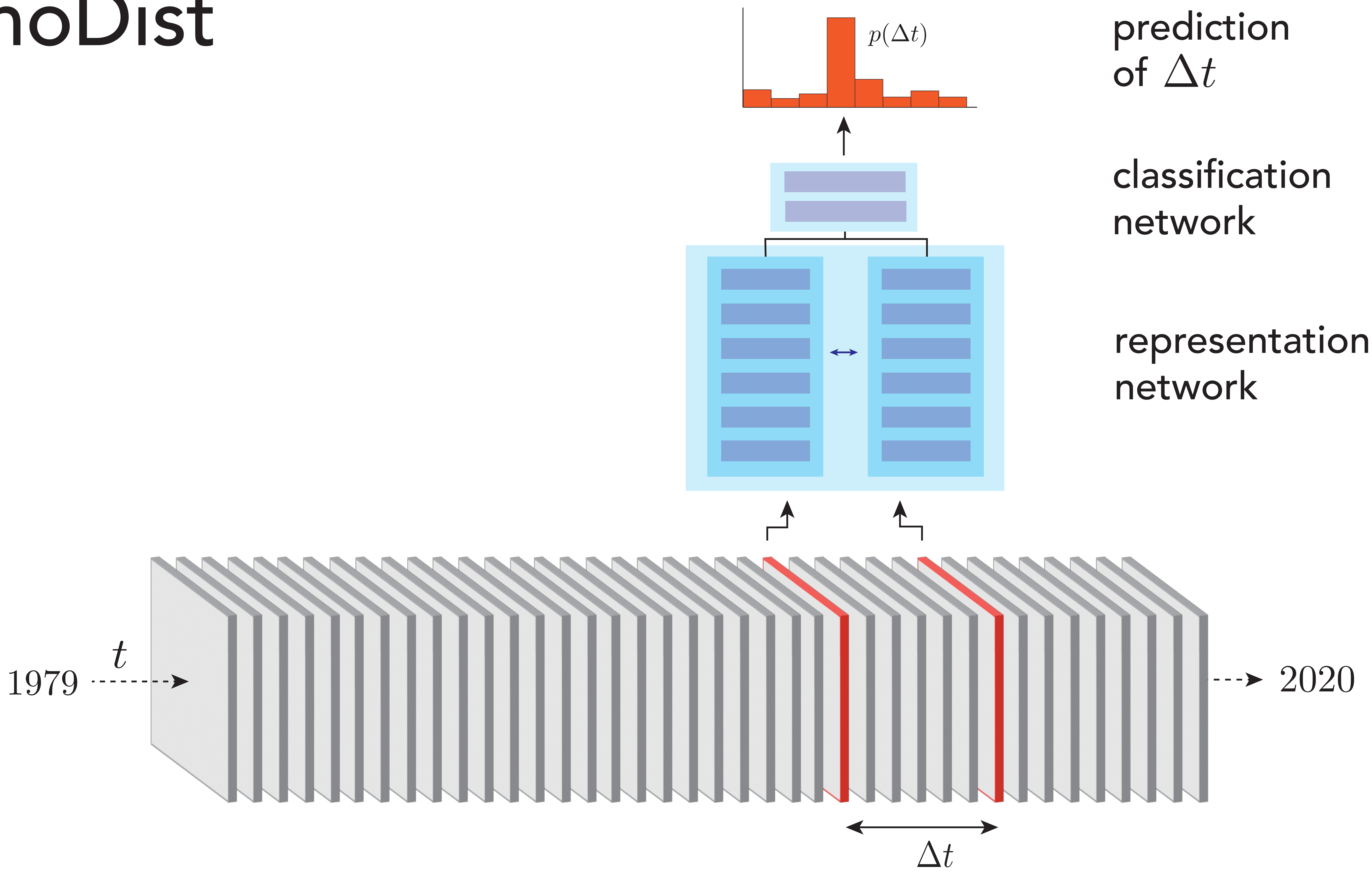
AtmoDist



AtmoDist



AtmoDist

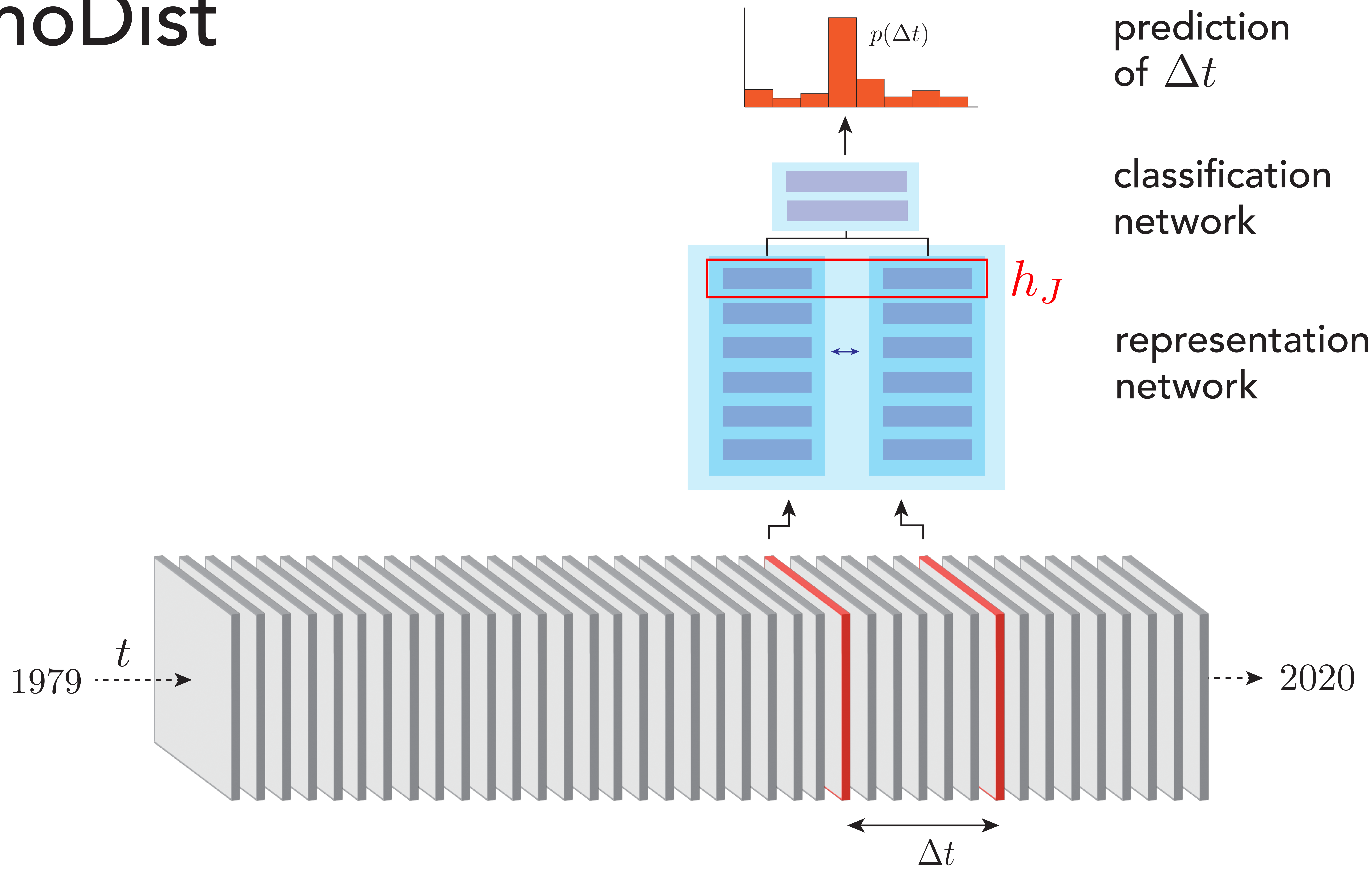


prediction of Δt

classification network

representation network

AtmoDist

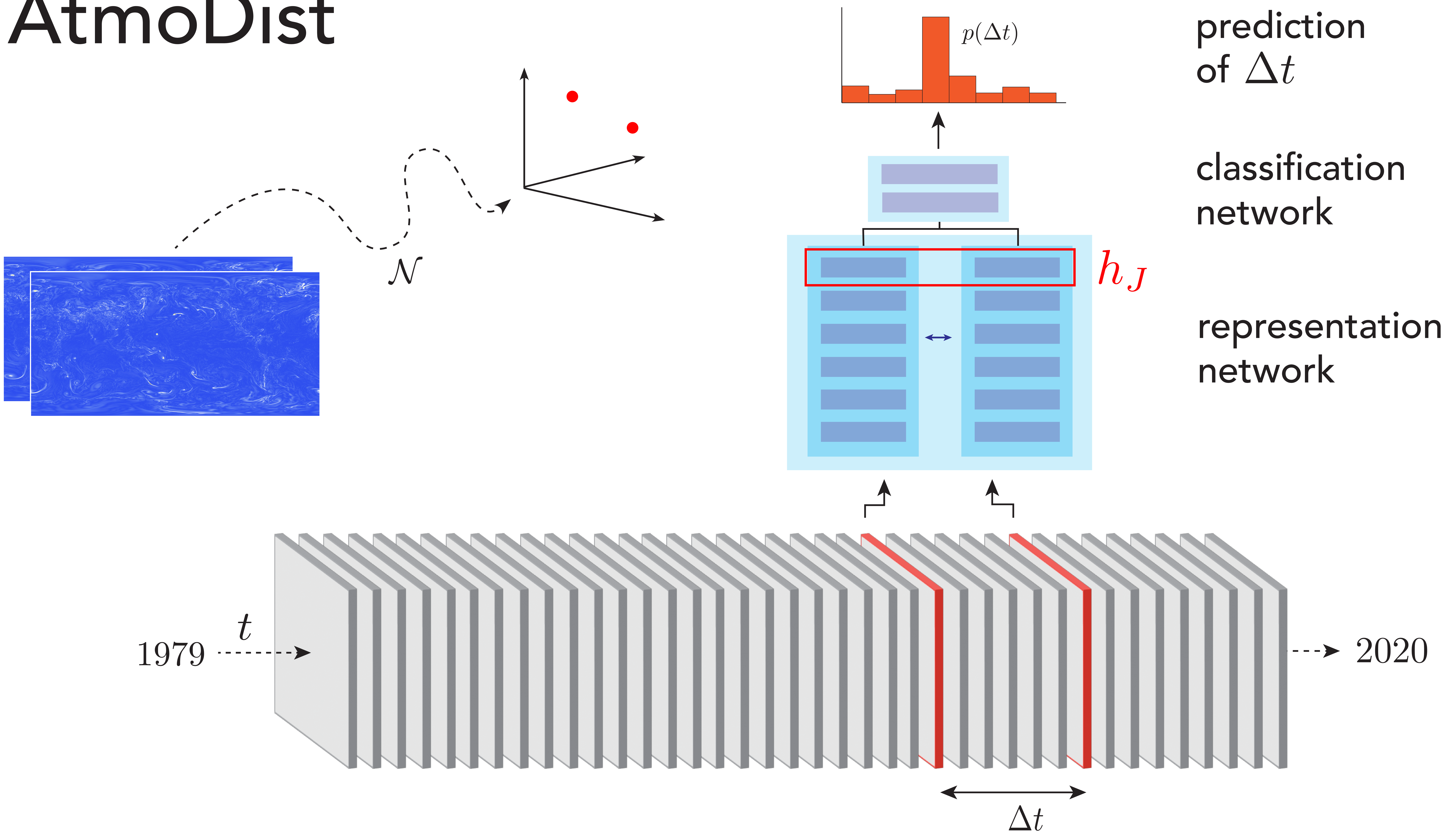


prediction of Δt

classification network

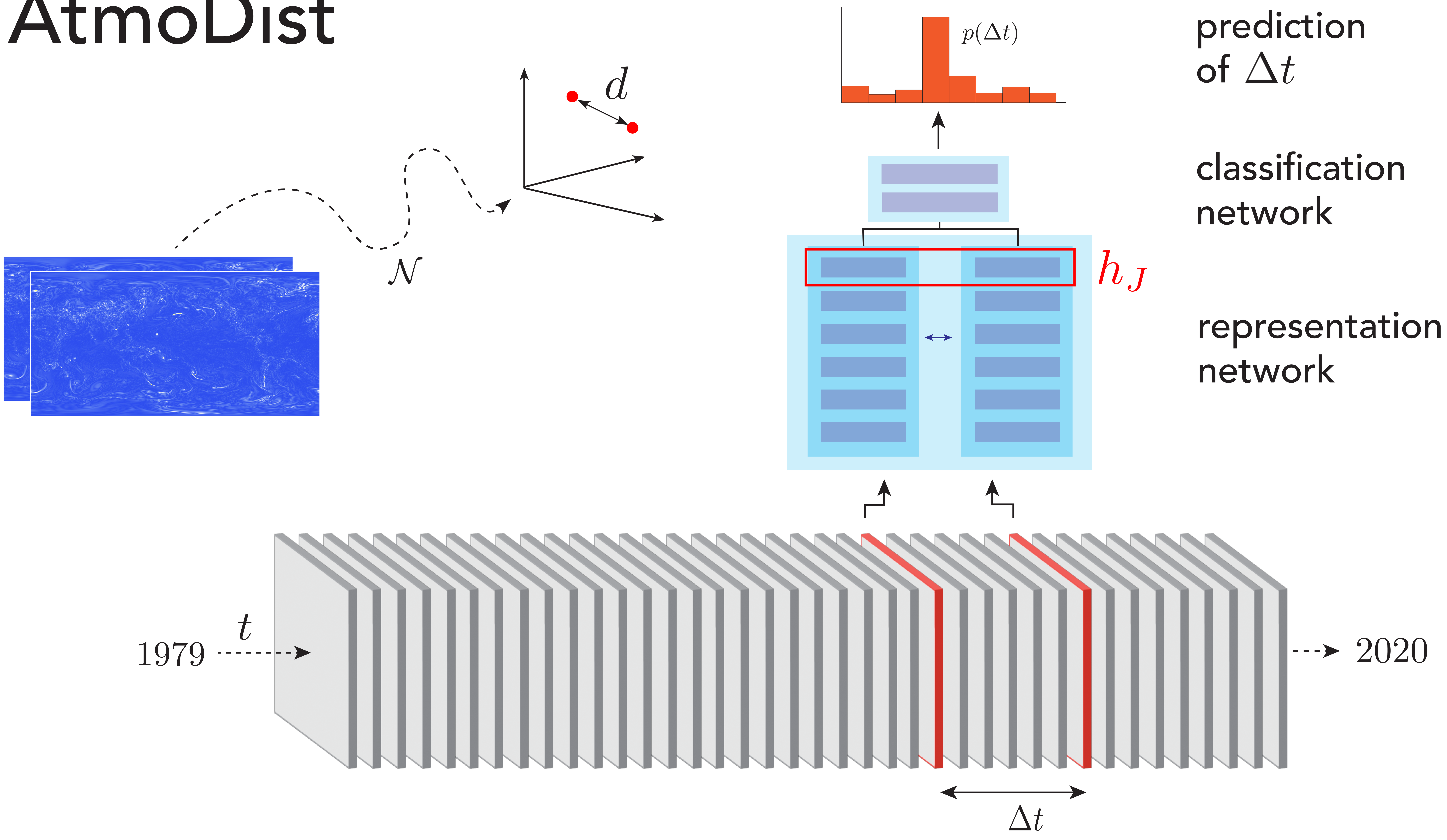
representation network

AtmoDist



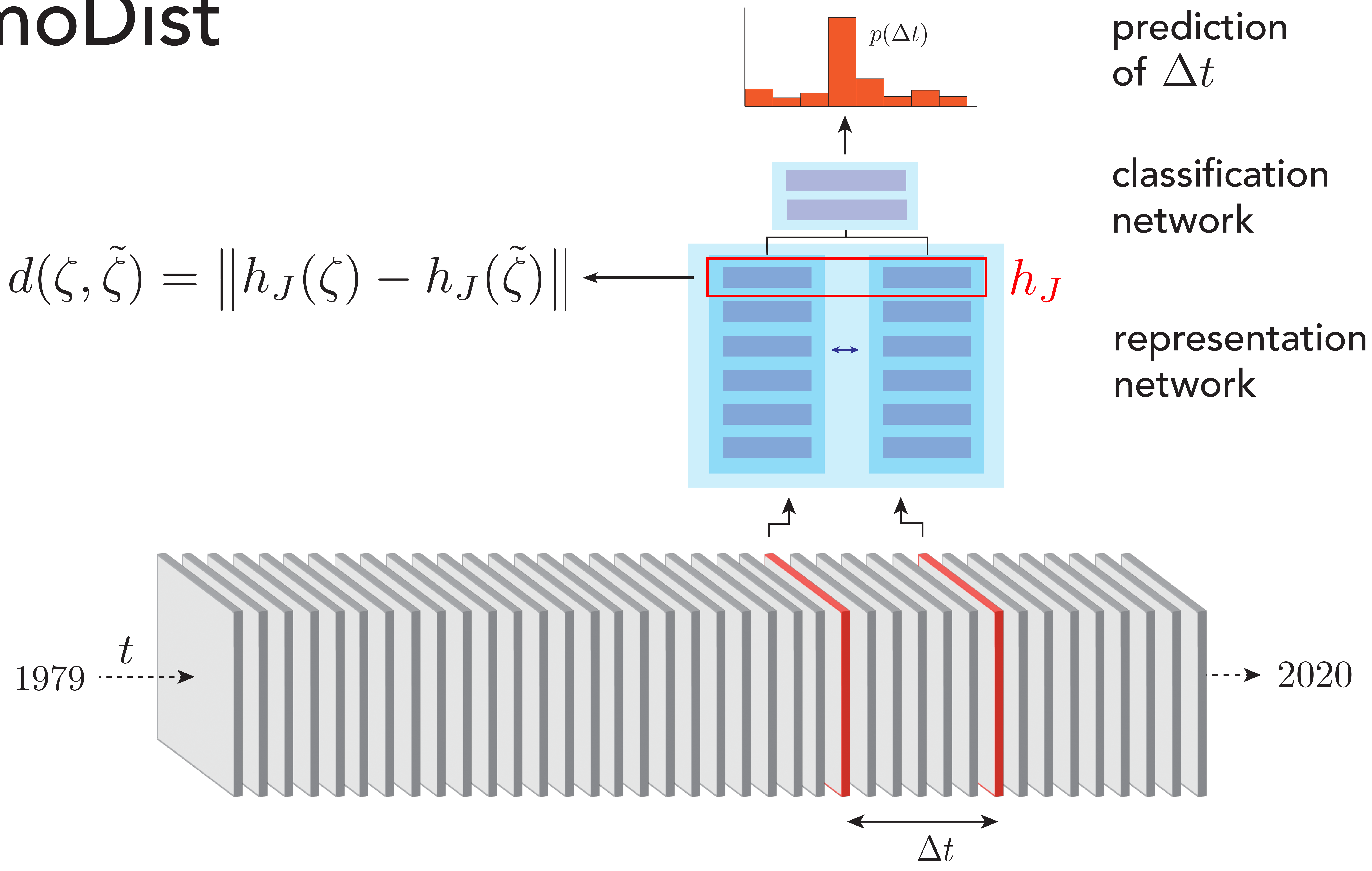
$p(\Delta t)$
prediction of Δt
classification network
h_J
representation network
$\mathcal{N}$
t
1979
2020
Δt

AtmoDist



d
$\mathcal{N}$
$p(\Delta t)$
prediction of Δt
classification network
h_J
representation network
1979
t
2020
Δt

AtmoDist



$p(\Delta t)$
prediction of Δt
classification network
$d(\zeta, \tilde{\zeta}) = \left\| h_J(\zeta) - h_J(\tilde{\zeta}) \right\|$
h_J
representation network
t
1979
2020
Δt

AtmoDist: data

- ERA5[1] reanalysis 1979-2006
  - Training: 1979-1998; Evaluation: 1999-2006 (58,440 training slices and 17,536 evaluation ones)

[1] Hersbach et al., The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146 (730), 2020.

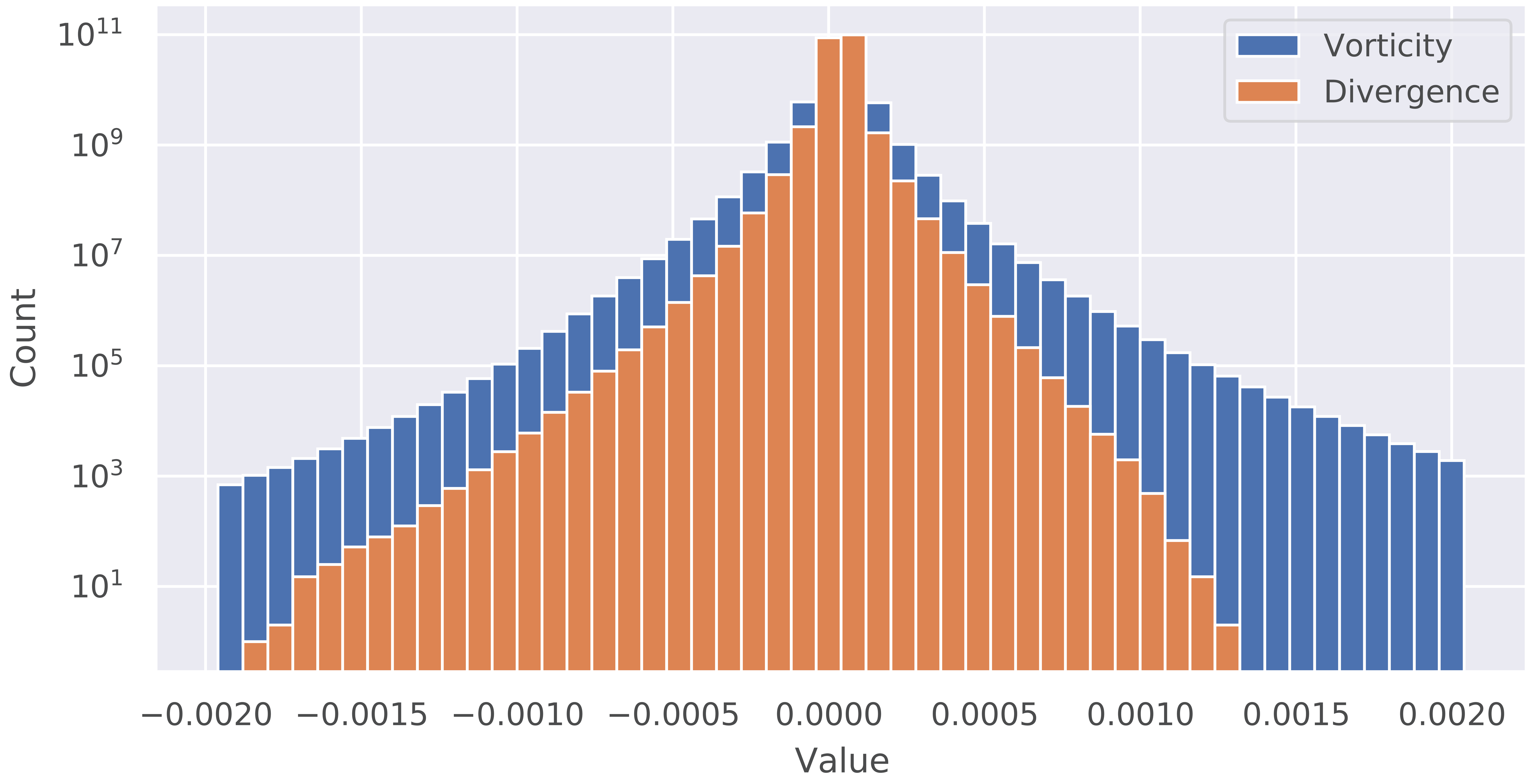
AtmoDist: data

- ERA5[1] reanalysis 1979-2006
  - Training: 1979-1998; Evaluation: 1999-2006 (58,440 training slices and 17,536 evaluation ones)
  - Vorticity and divergence

[1] Hersbach et al., The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146 (730), 2020.

AtmoDist: data

- Distribution of data:



AtmoDist: data

- ERA5[1] reanalysis 1979-2006
  - Training: 1979-1998; Evaluation: 1999-2006 (58,440 training slices and 17,536 evaluation ones)
  - Vorticity and divergence
  - 1280 × 2560 grids sampled into 160 x 160 patches

[1] Hersbach et al., The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146 (730), 2020.

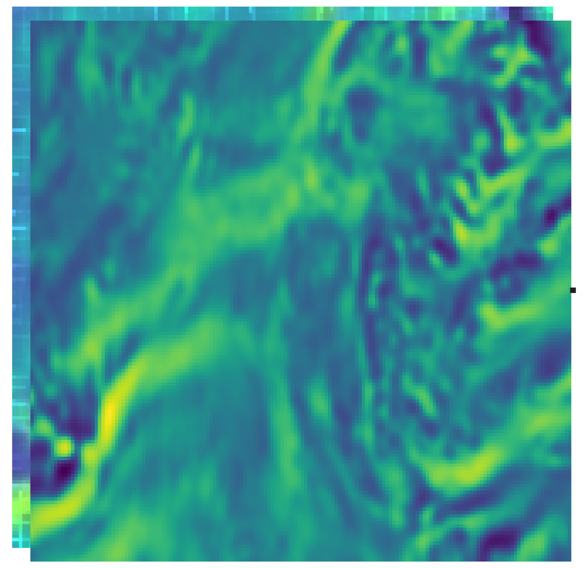
AtmoDist: data

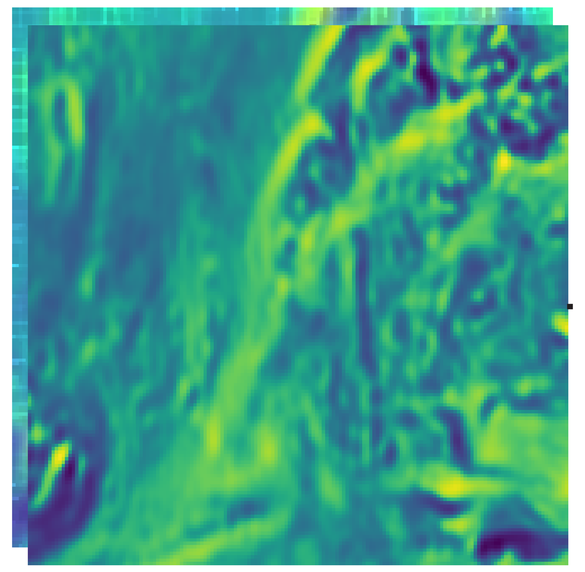
- ERA5[1] reanalysis 1979-2006
  - Training: 1979-1998; Evaluation: 1999-2006 (58,440 training slices and 17,536 evaluation ones)
  - Vorticity and divergence
  - 1280×2560 grids sampled into 160 x 160 patches
  - One vertical layer ($\approx$ 883 hPa)

[1] Hersbach et al., The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146 (730), 2020.

AtmoDist: network

vorticity,
divergence

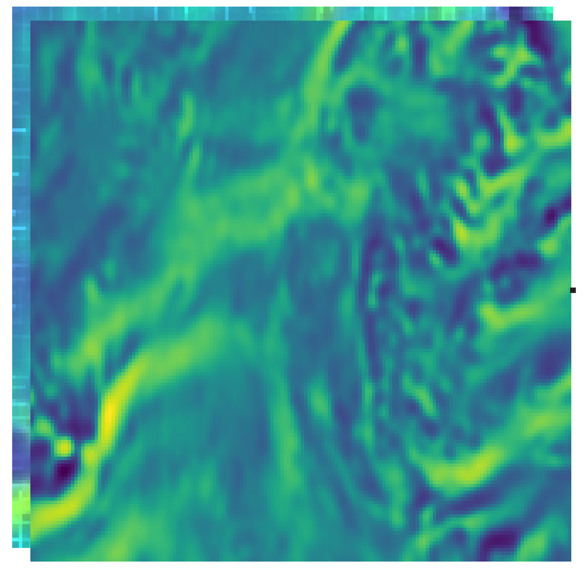




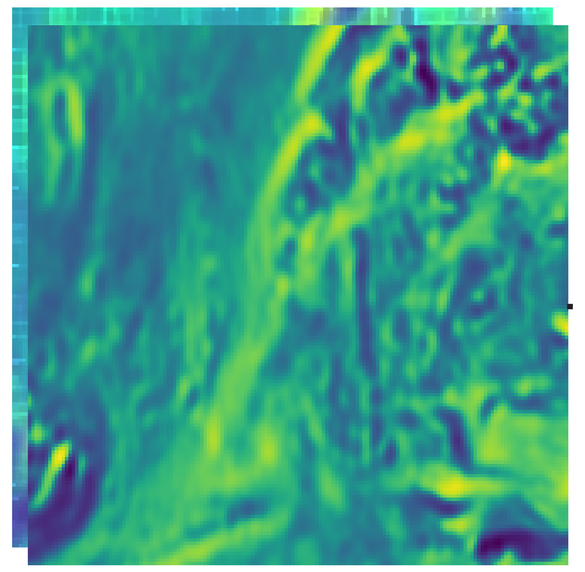
AtmoDist: network

vorticity,
divergence

t



$t + \Delta t$

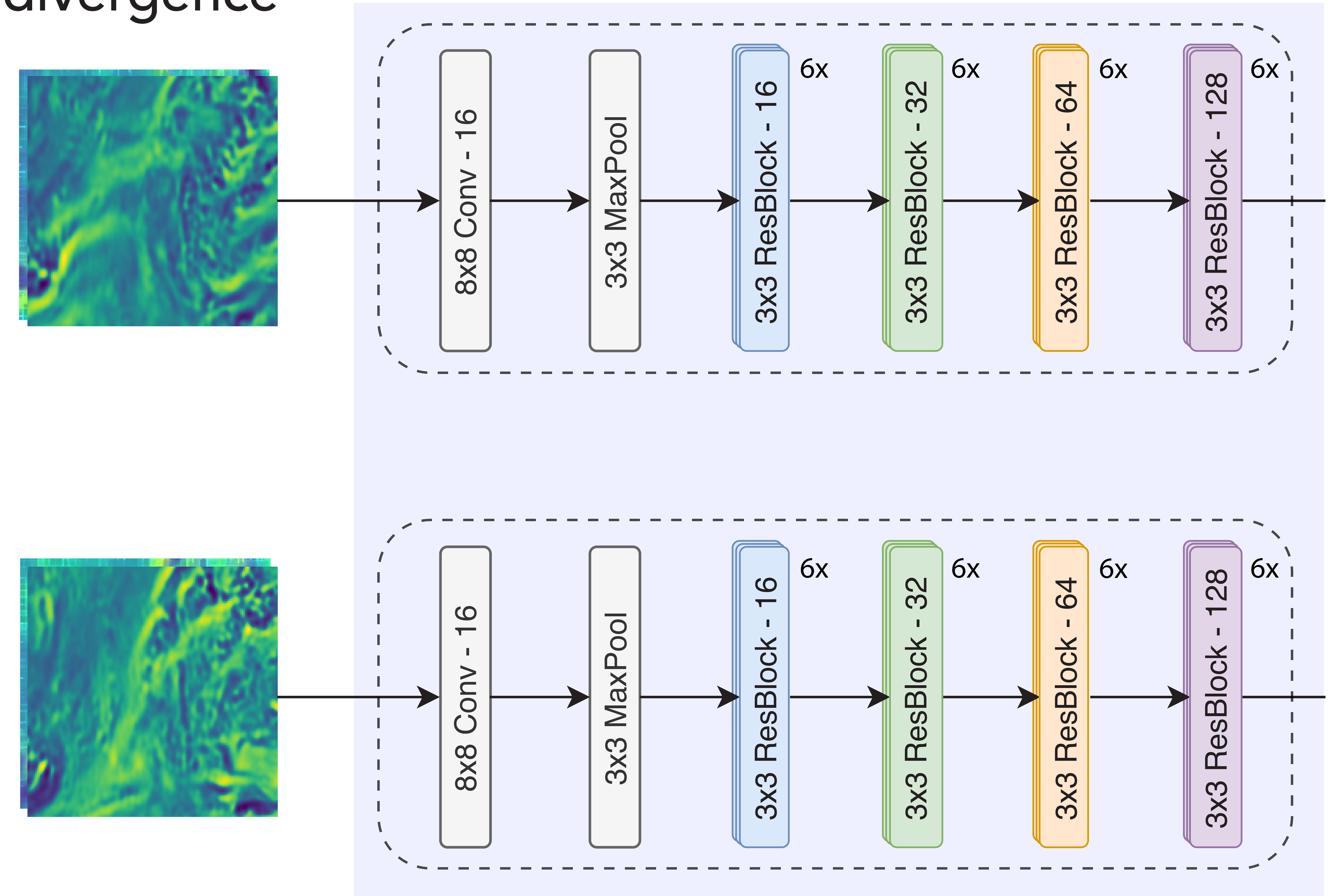


AtmoDist: network

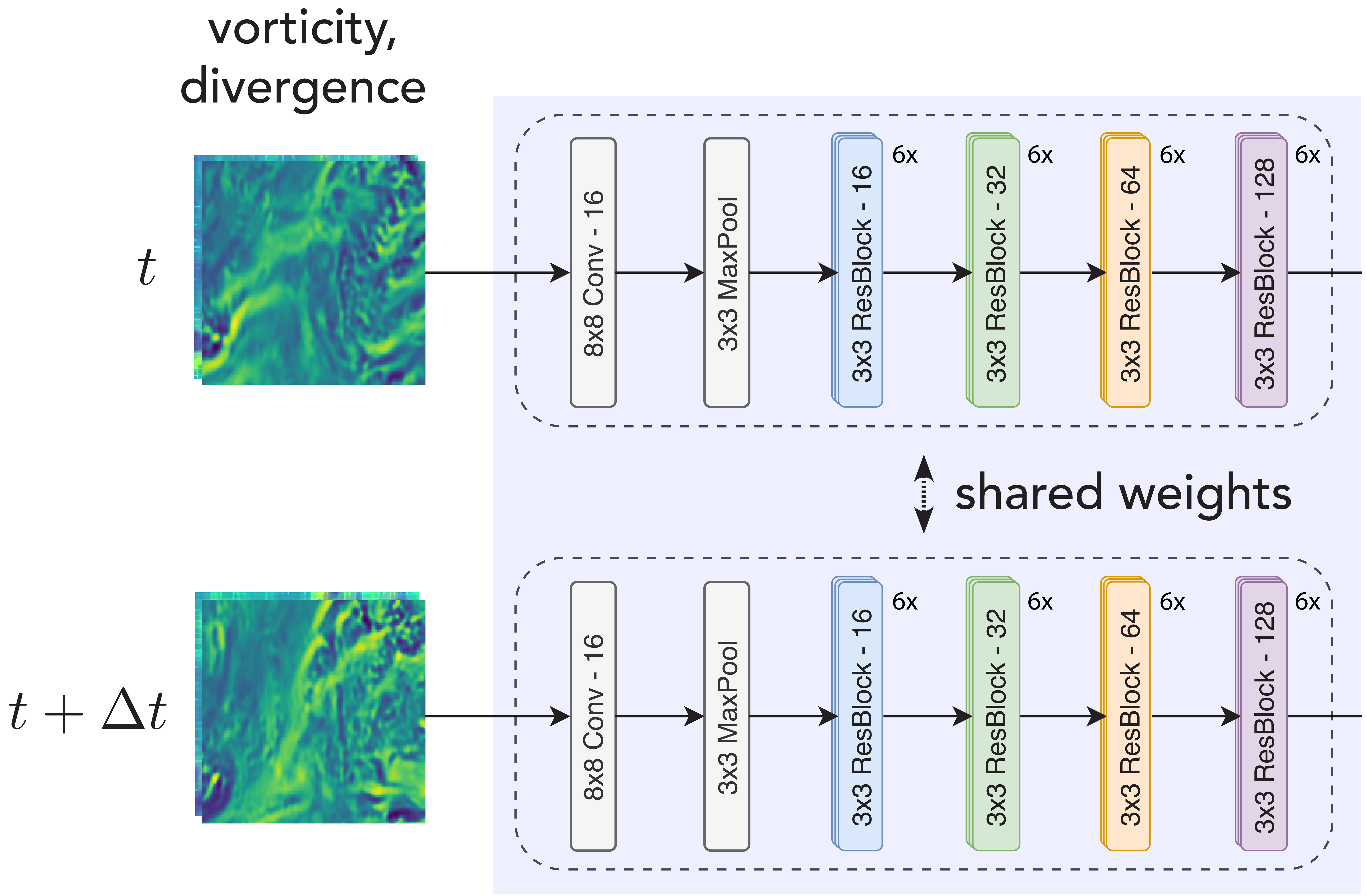
vorticity,
divergence

t

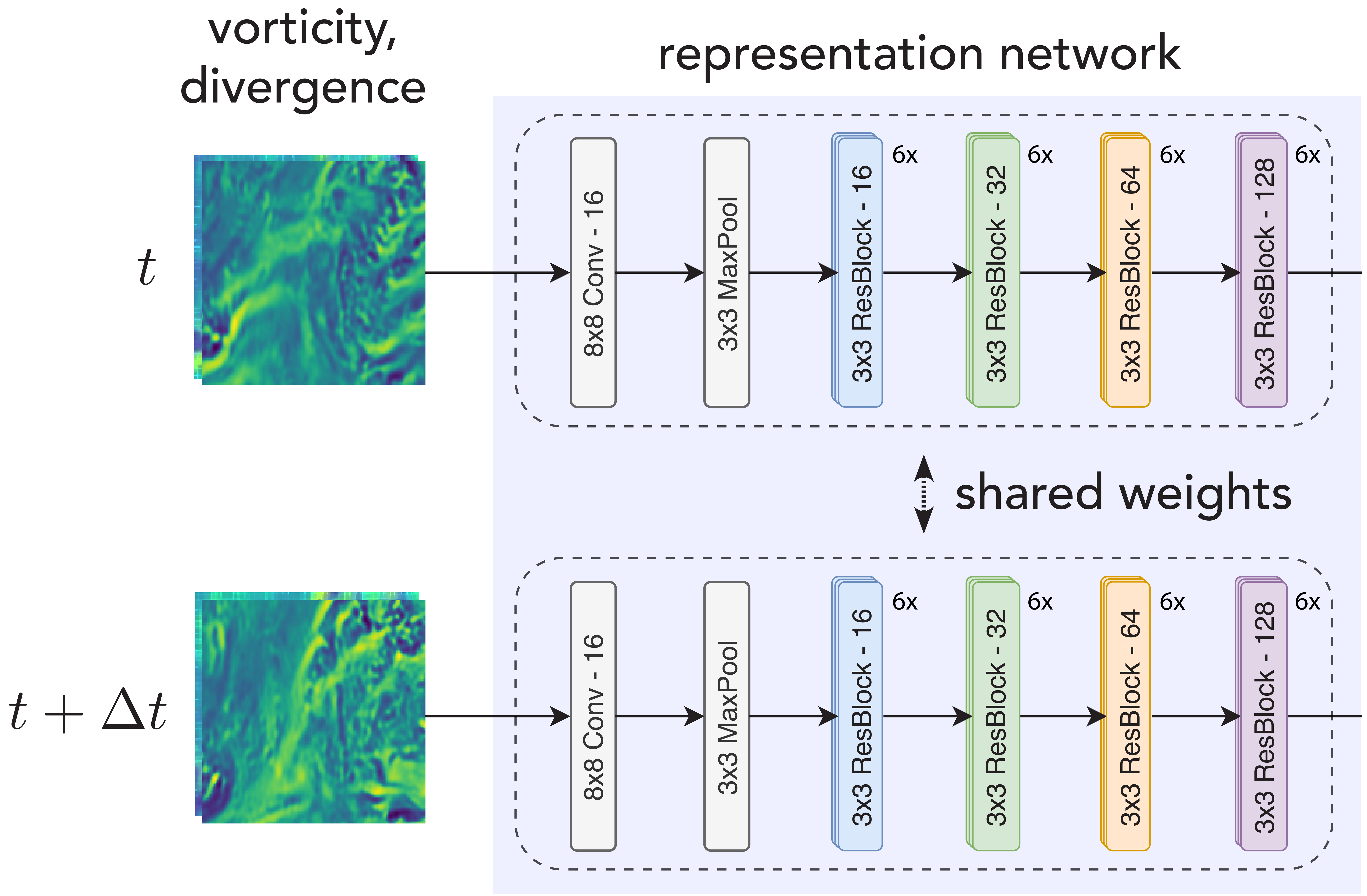
$t + \Delta t$



AtmoDist: network

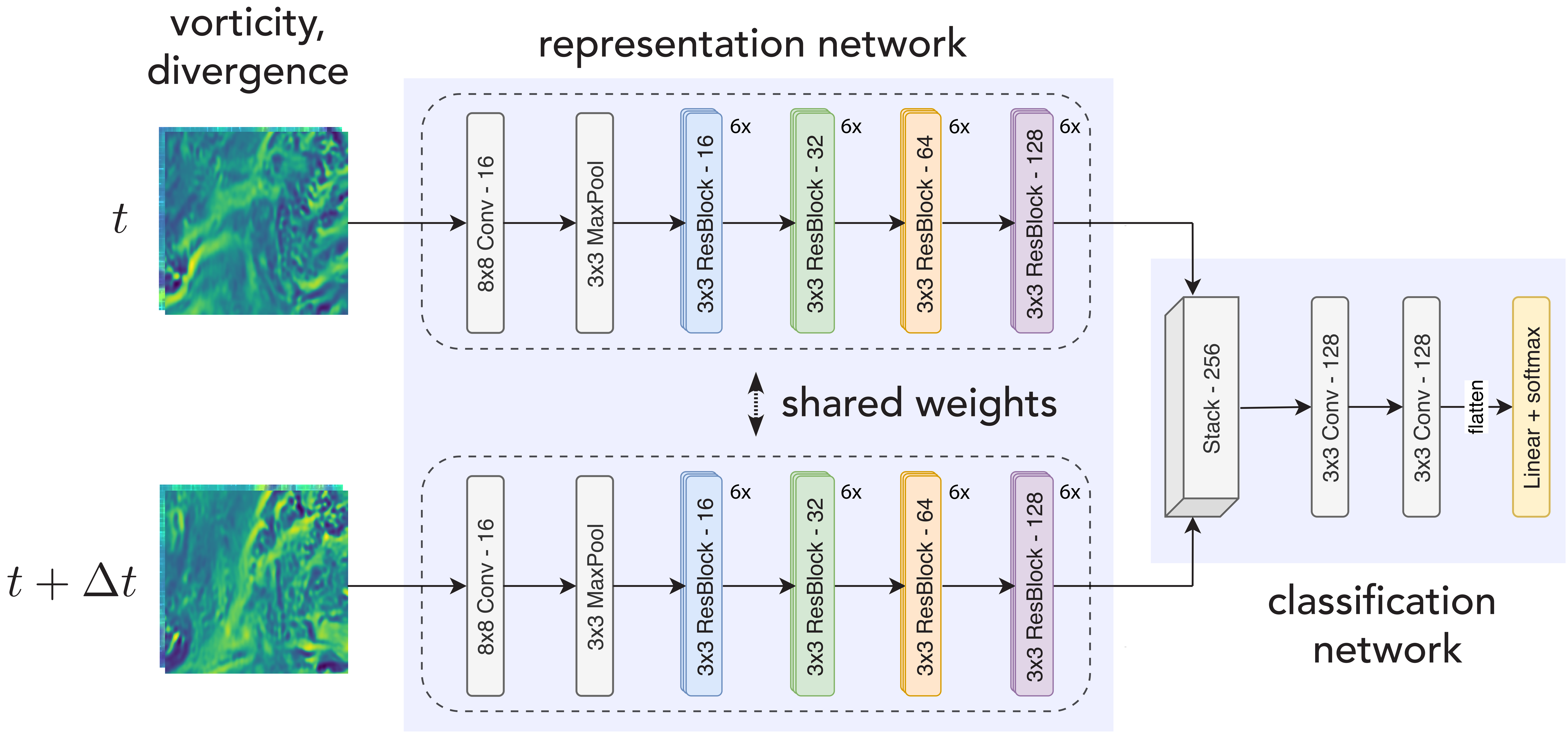


AtmoDist: network



2.27 million parameters

AtmoDist: network

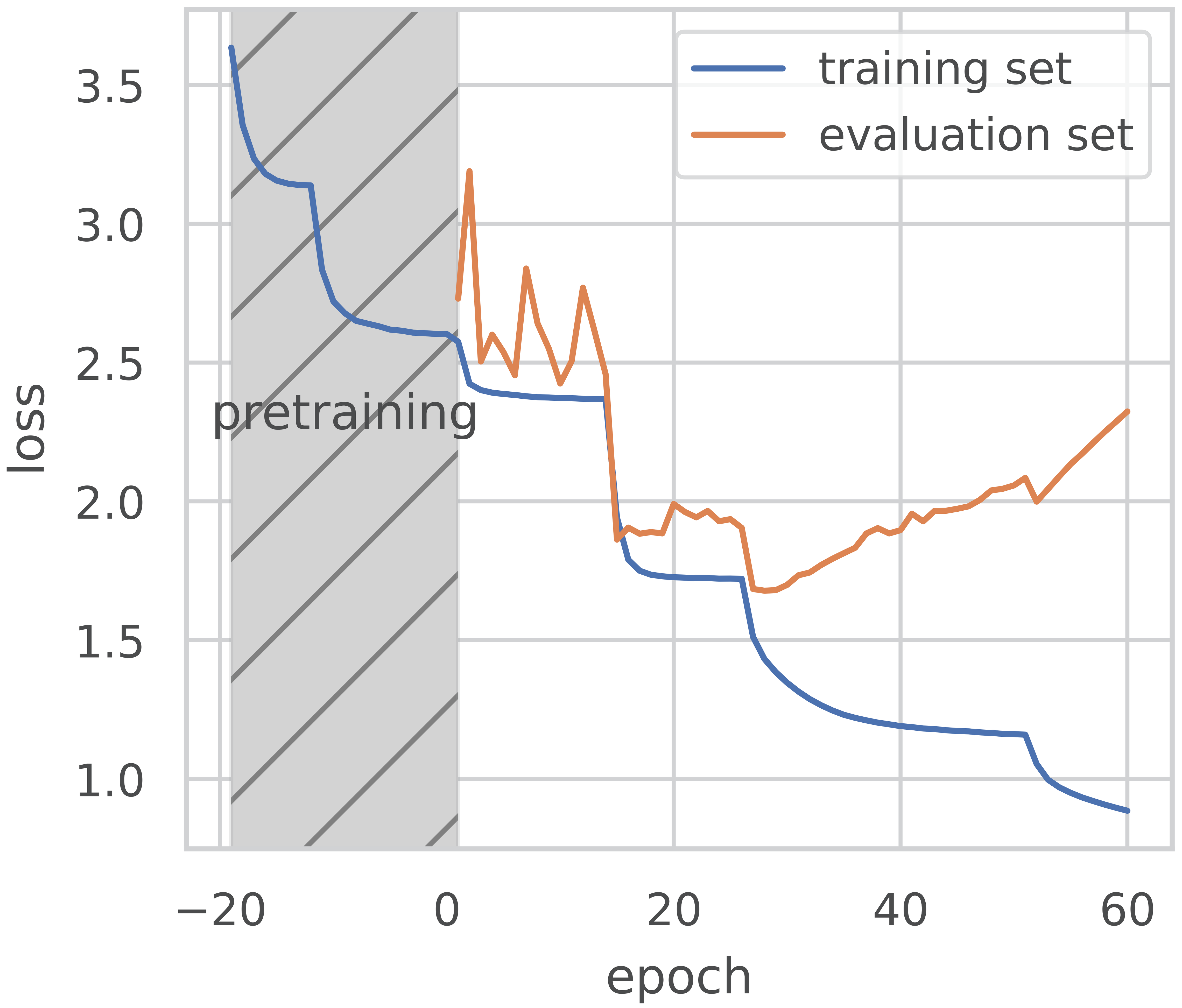


AtmoDist: training

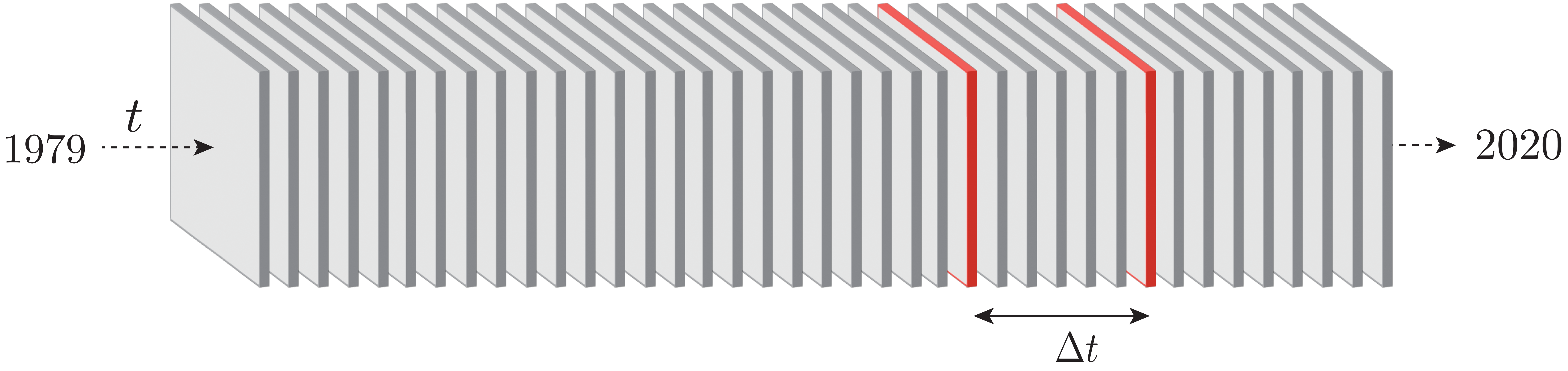
- 23 classes with max. time difference of 69 hours
- Pre-training with subset of data to improve convergence

AtmoDist: training

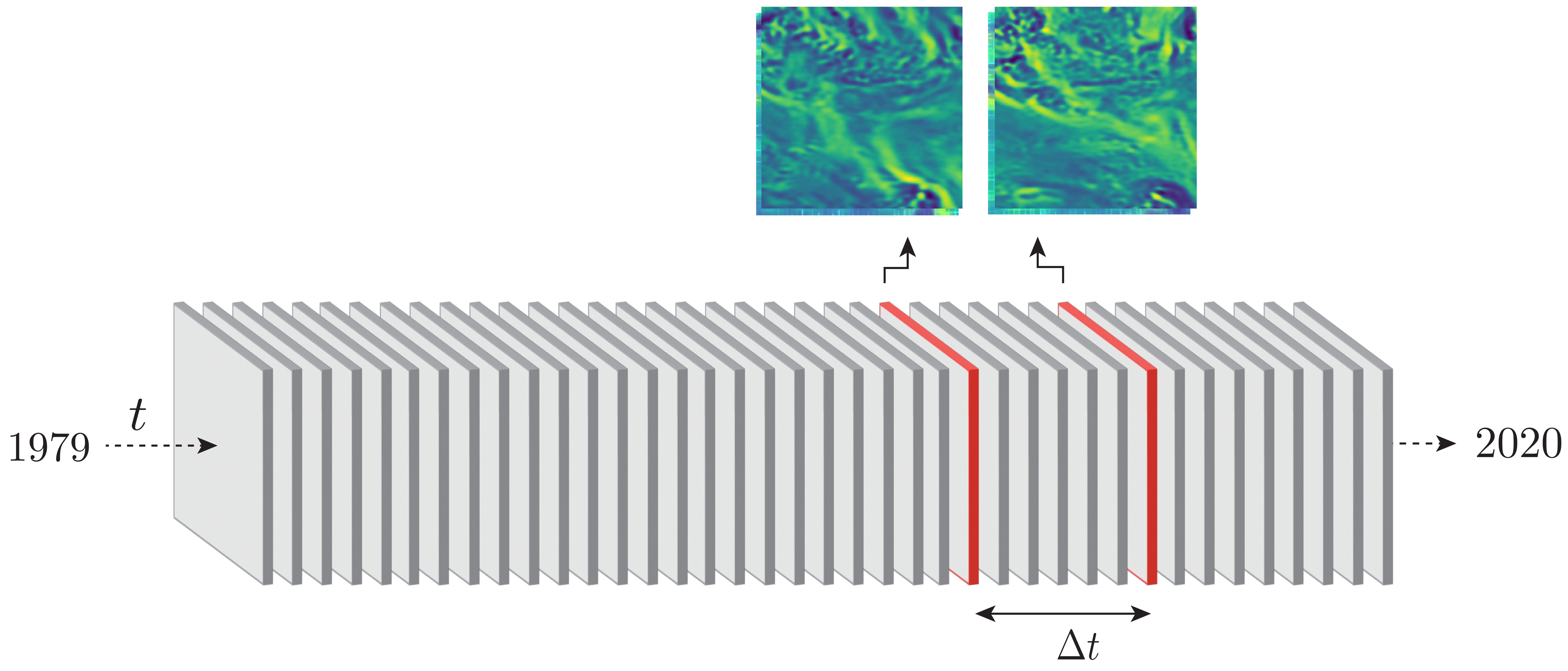
- 23 classes with max. time difference of 69 hours
- Pre-training with subset of data to improve convergence



AtmoDist: evaluation

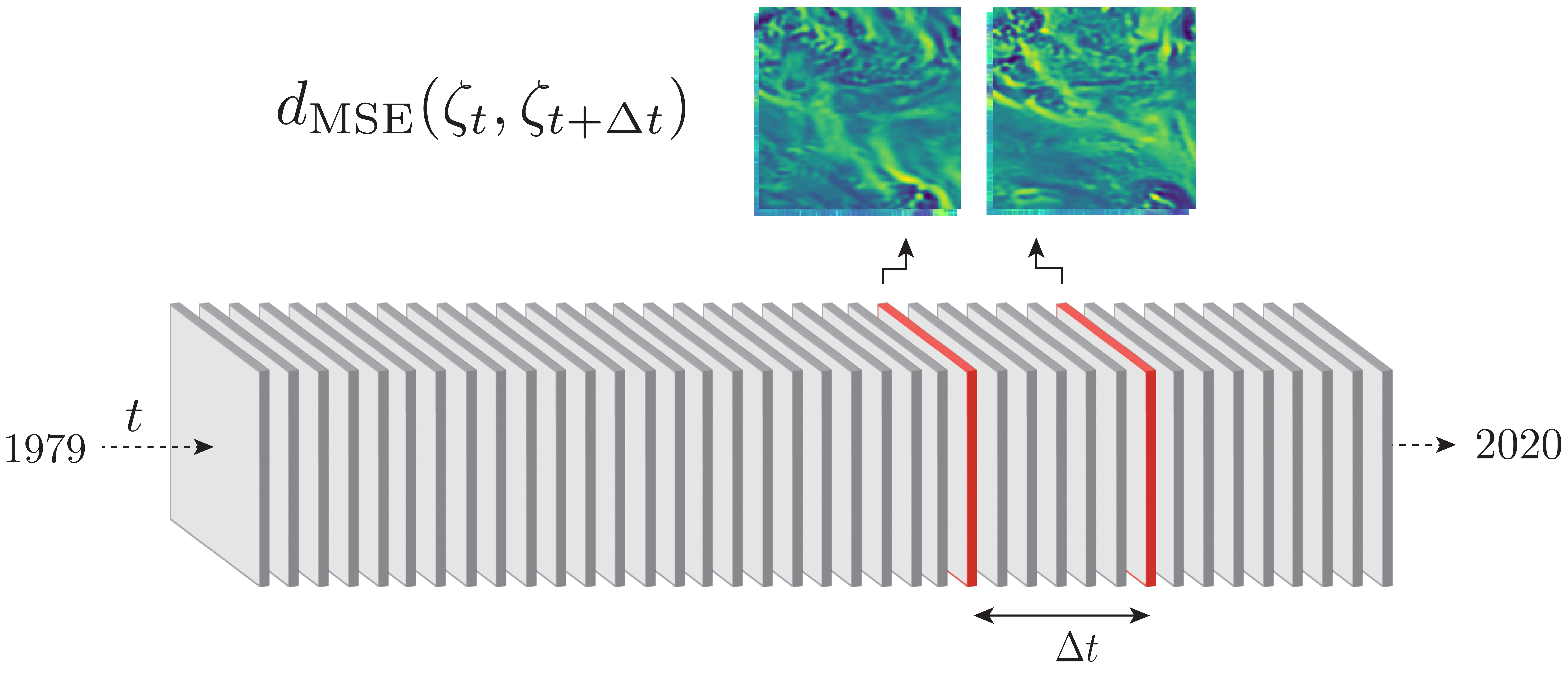


AtmoDist: evaluation

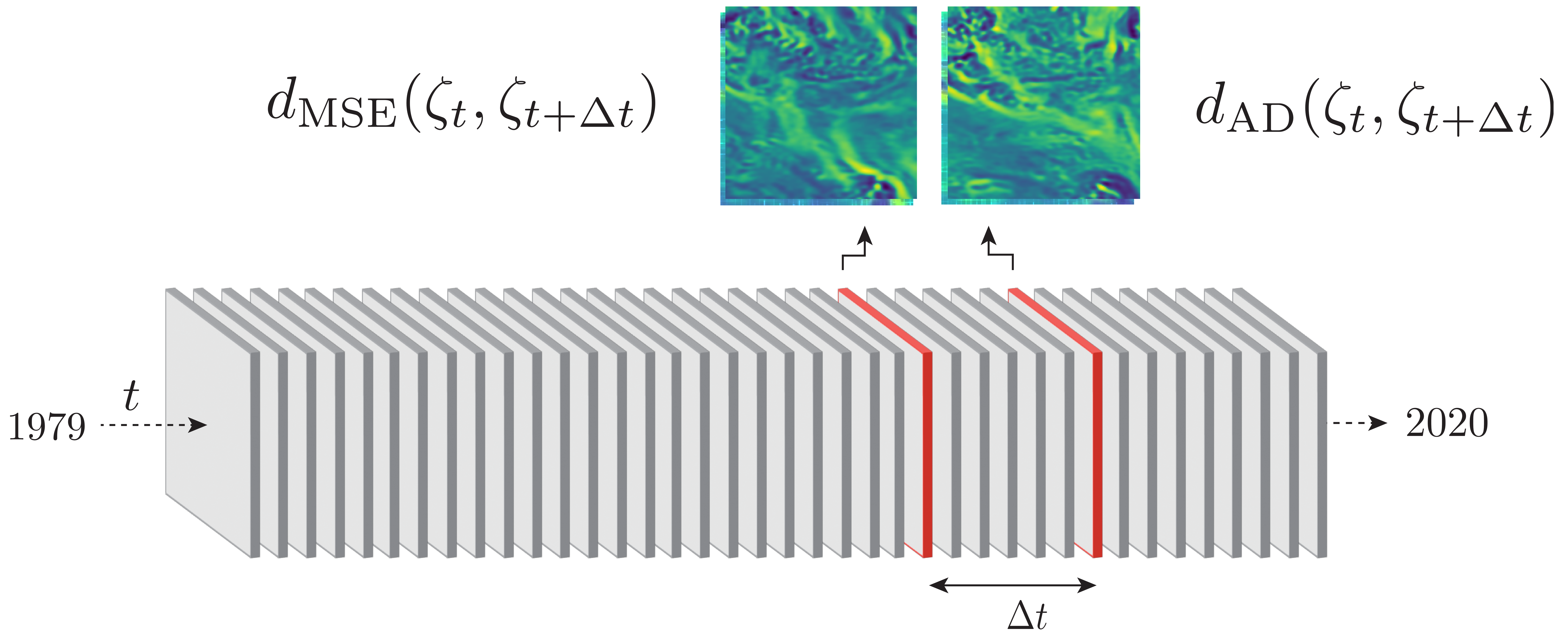


AtmoDist: evaluation

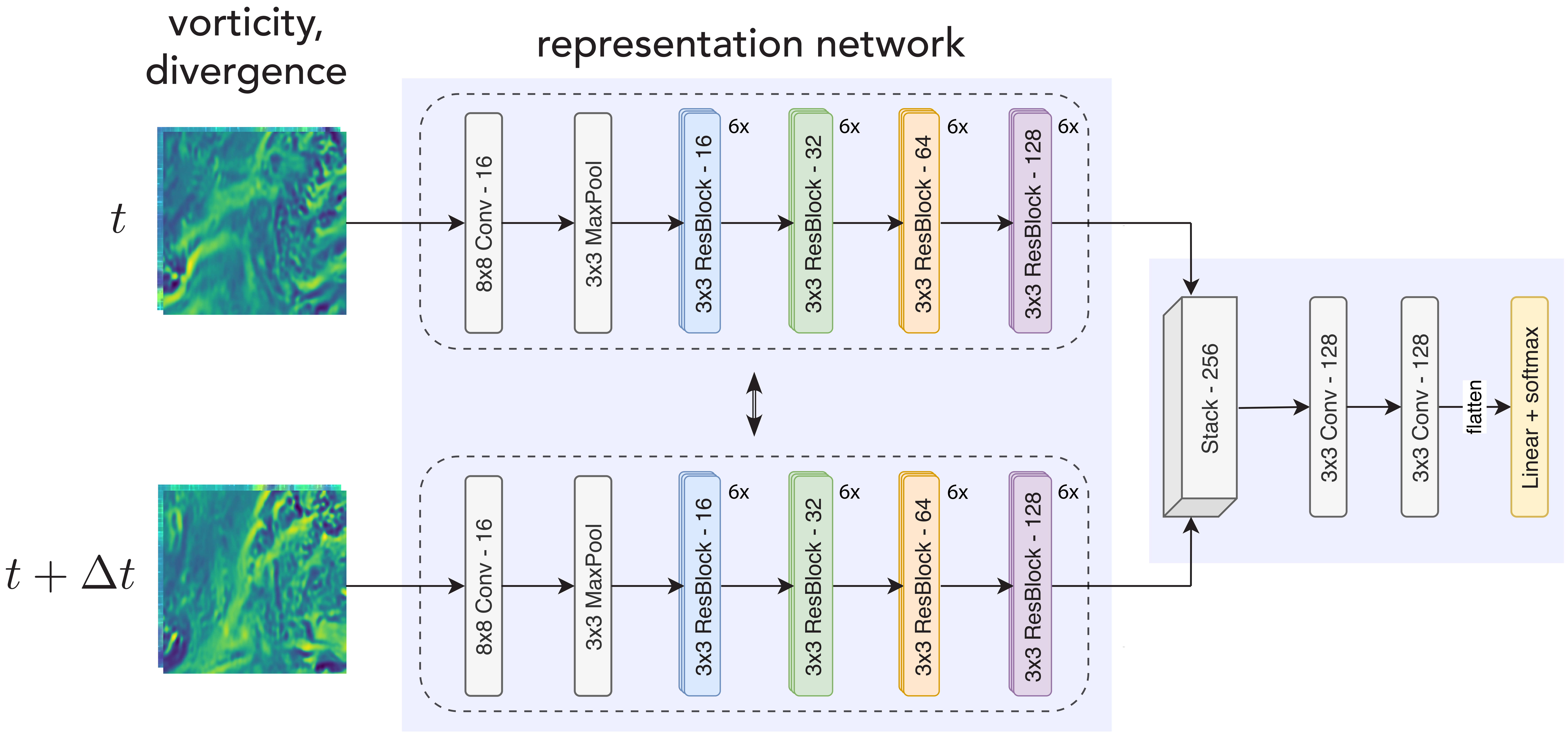
$$d_{\mathrm{MSE}}(\zeta_t, \zeta_{t+\Delta t})$$



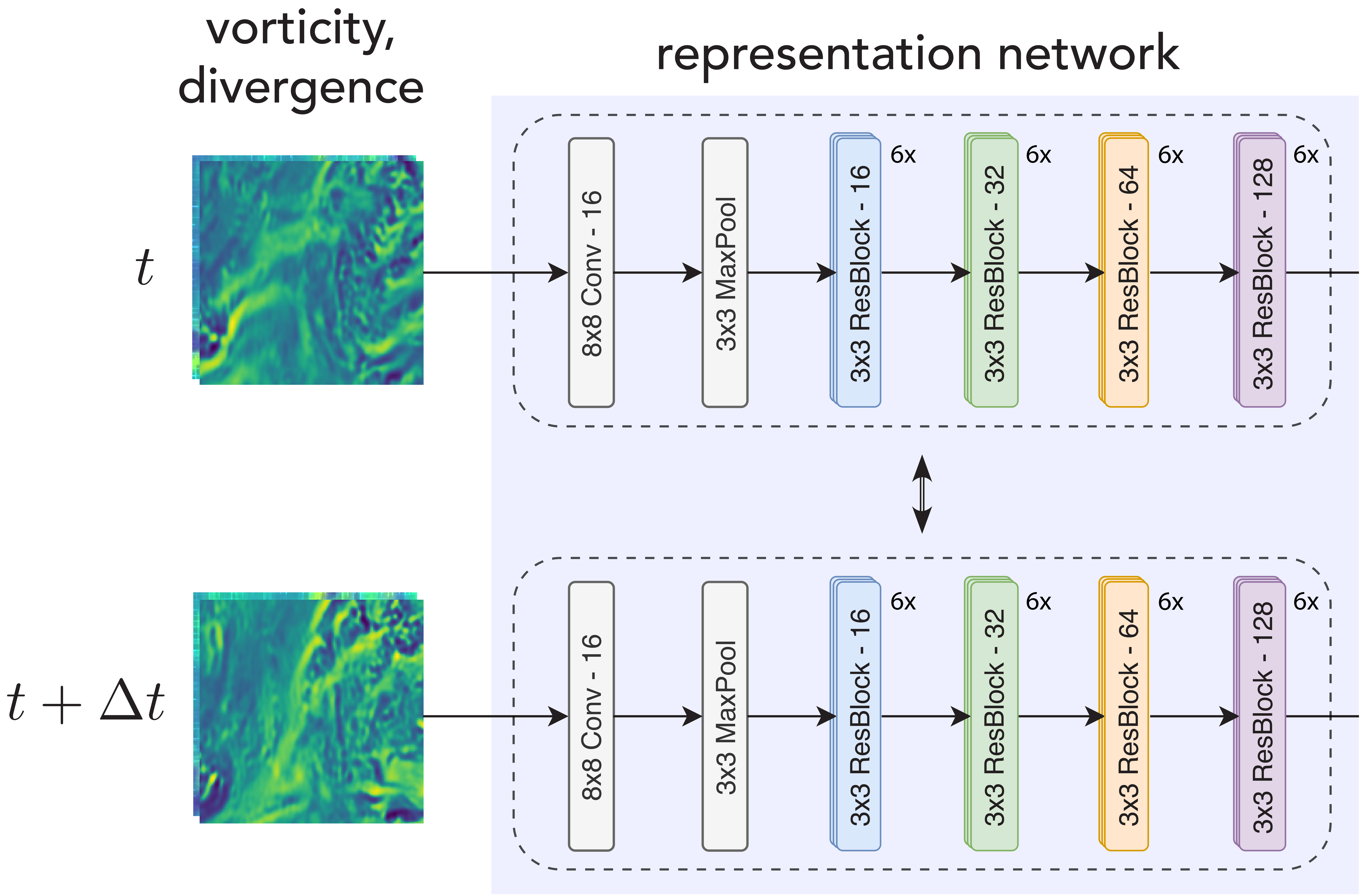
AtmoDist: evaluation



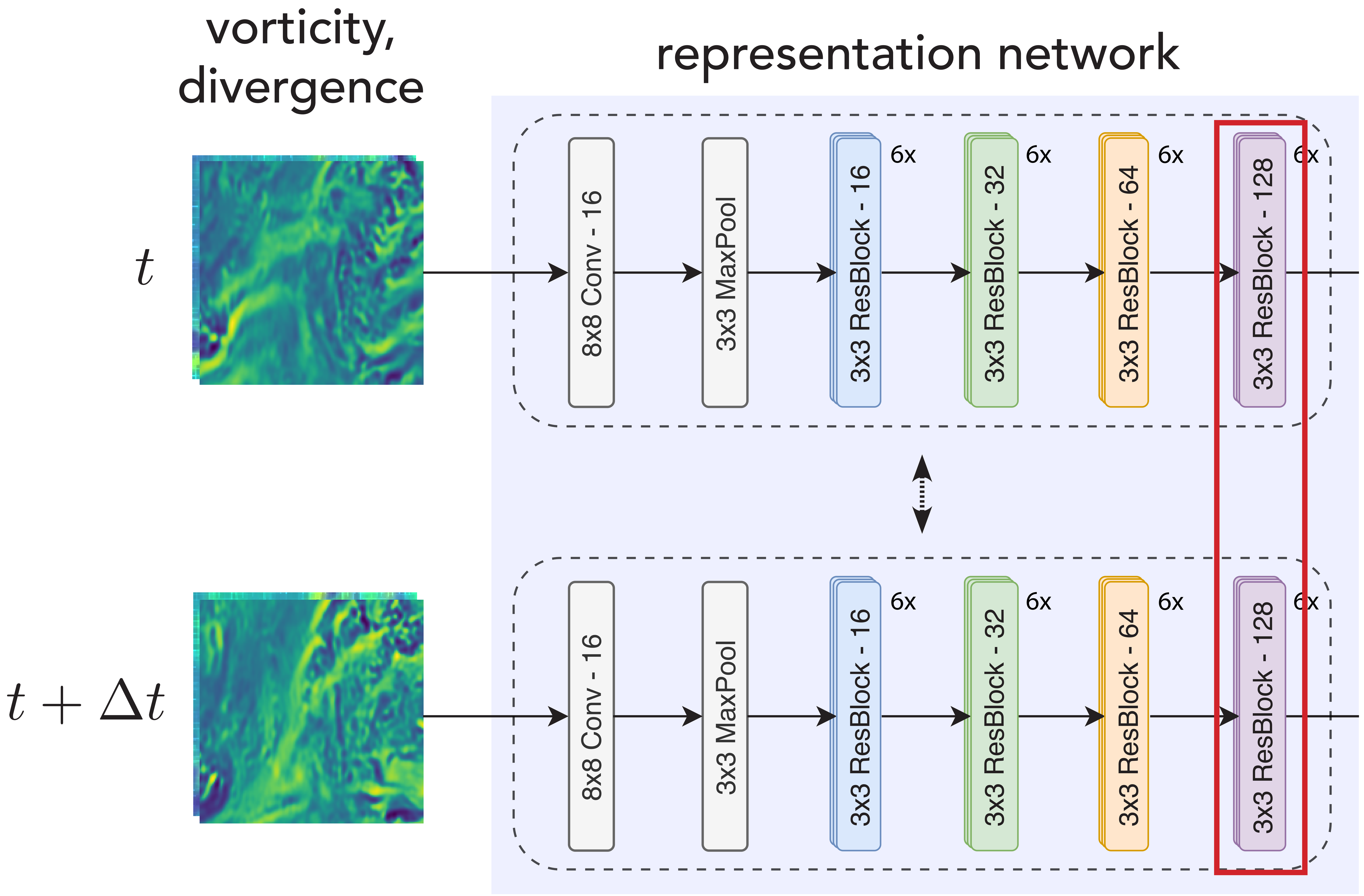
AtmoDist: evaluation



AtmoDist: evaluation

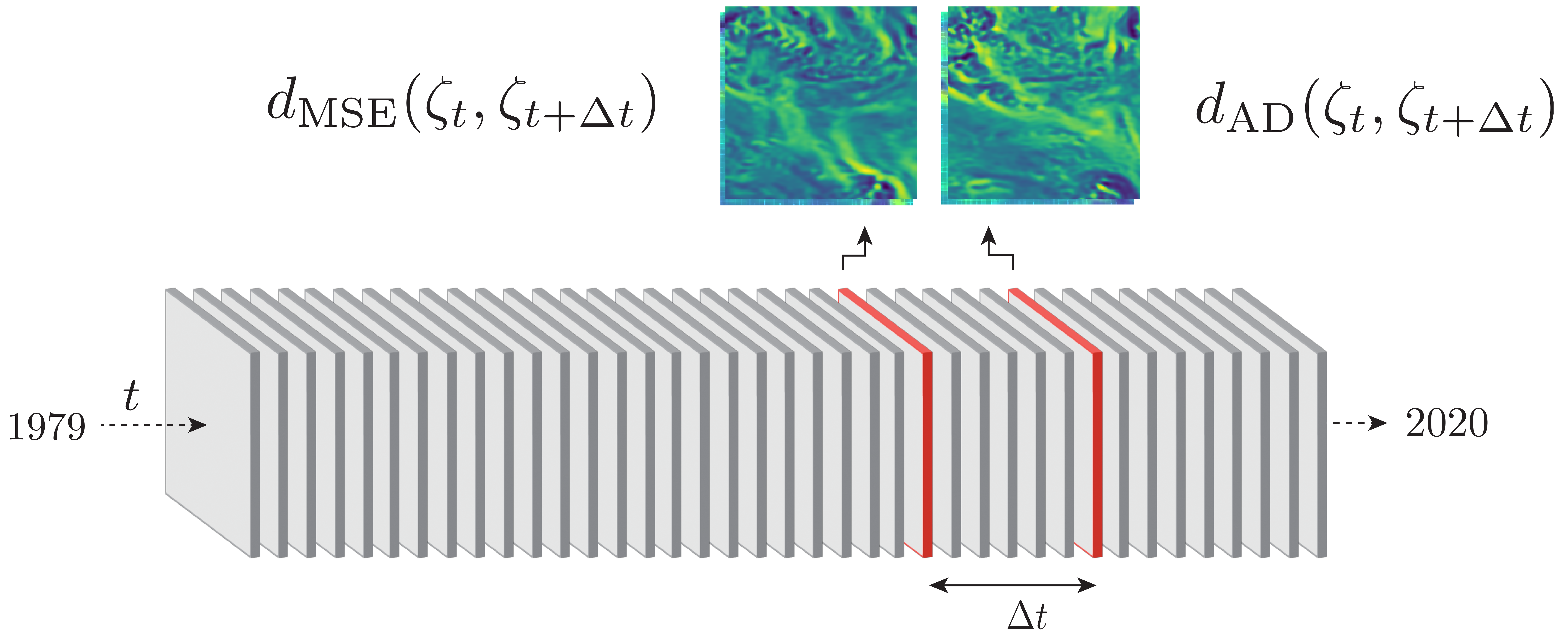


AtmoDist: evaluation

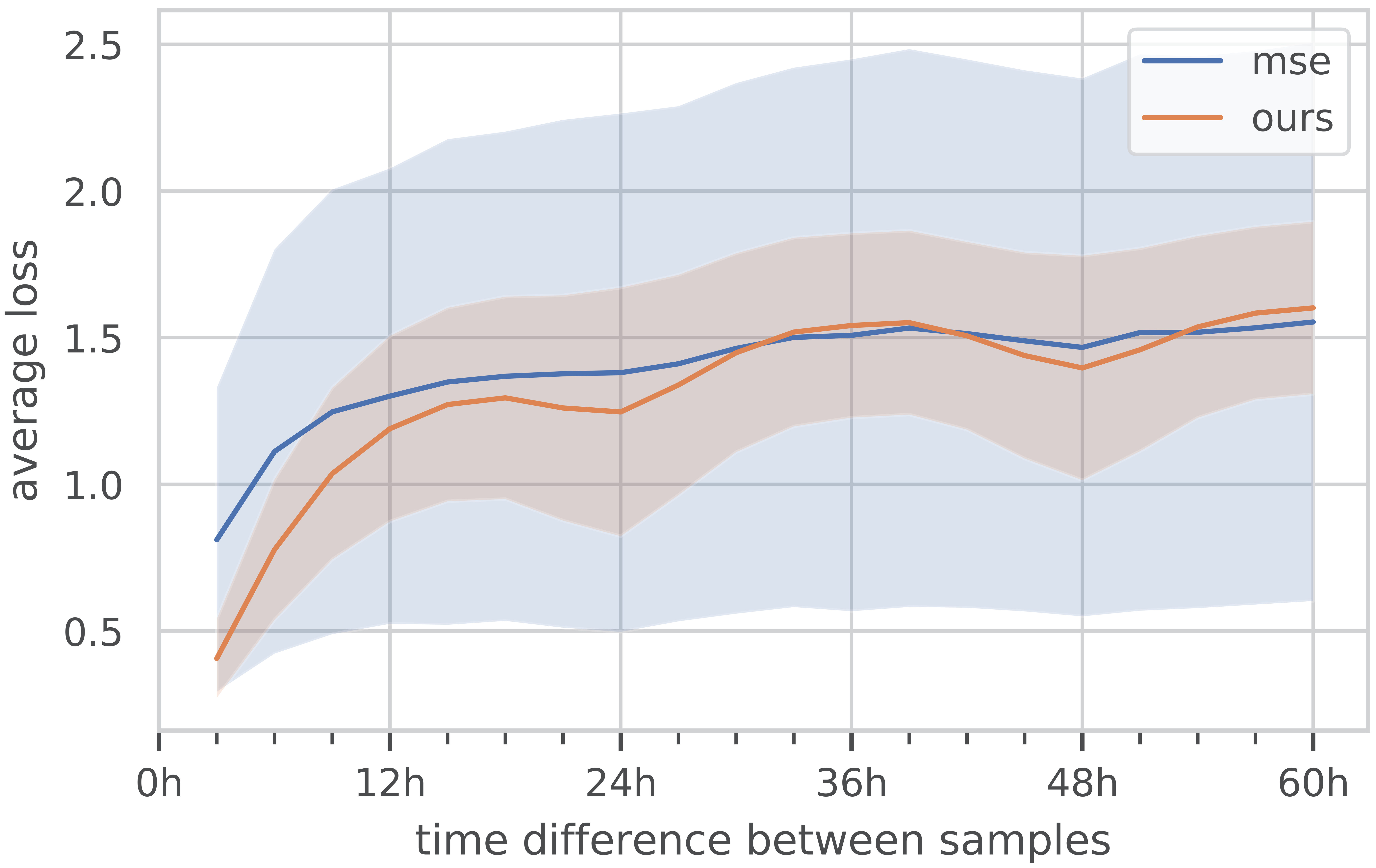


AtmoDist: evaluation

AtmoDist: evaluation



AtmoDist: evaluation



Super-resolution using AtmoDist

- Objective: down-scaling / upsampling of coarse fields

Super-resolution using AtmoDist

- Objective: down-scaling / upsampling of coarse fields
- Comparison using and with the GAN of Stengel et al.[1]

[1] K. Stengel, A. Glaws, D. Hettinger, and R. N. King. Adversarial super-resolution of climatological wind and solar data. Proceedings of the National Academy of Sciences, 117(29):16805–16815, 2020.

[2] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of CVPR, July 2017.

Super-resolution using AtmoDist

- Objective: down-scaling / upsampling of coarse fields
- Comparison using and with the GAN of Stengel et al.[1]
  - GAN is based on SRGAN[2] for natural images

[1] K. Stengel, A. Glaws, D. Hettinger, and R. N. King. Adversarial super-resolution of climatological wind and solar data. Proceedings of the National Academy of Sciences, 117(29):16805–16815, 2020.

[2] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of CVPR, July 2017.

Super-resolution using AtmoDist

- Objective: down-scaling / upsampling of coarse fields
- Comparison using and with the GAN of Stengel et al.[1]
  - GAN is based on SRGAN[2] for natural images
  - Our content loss replaces mean squared error

[1] K. Stengel, A. Glaws, D. Hettinger, and R. N. King. Adversarial super-resolution of climatological wind and solar data. Proceedings of the National Academy of Sciences, 117(29):16805–16815, 2020.

[2] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of CVPR, July 2017.

Super-resolution using AtmoDist

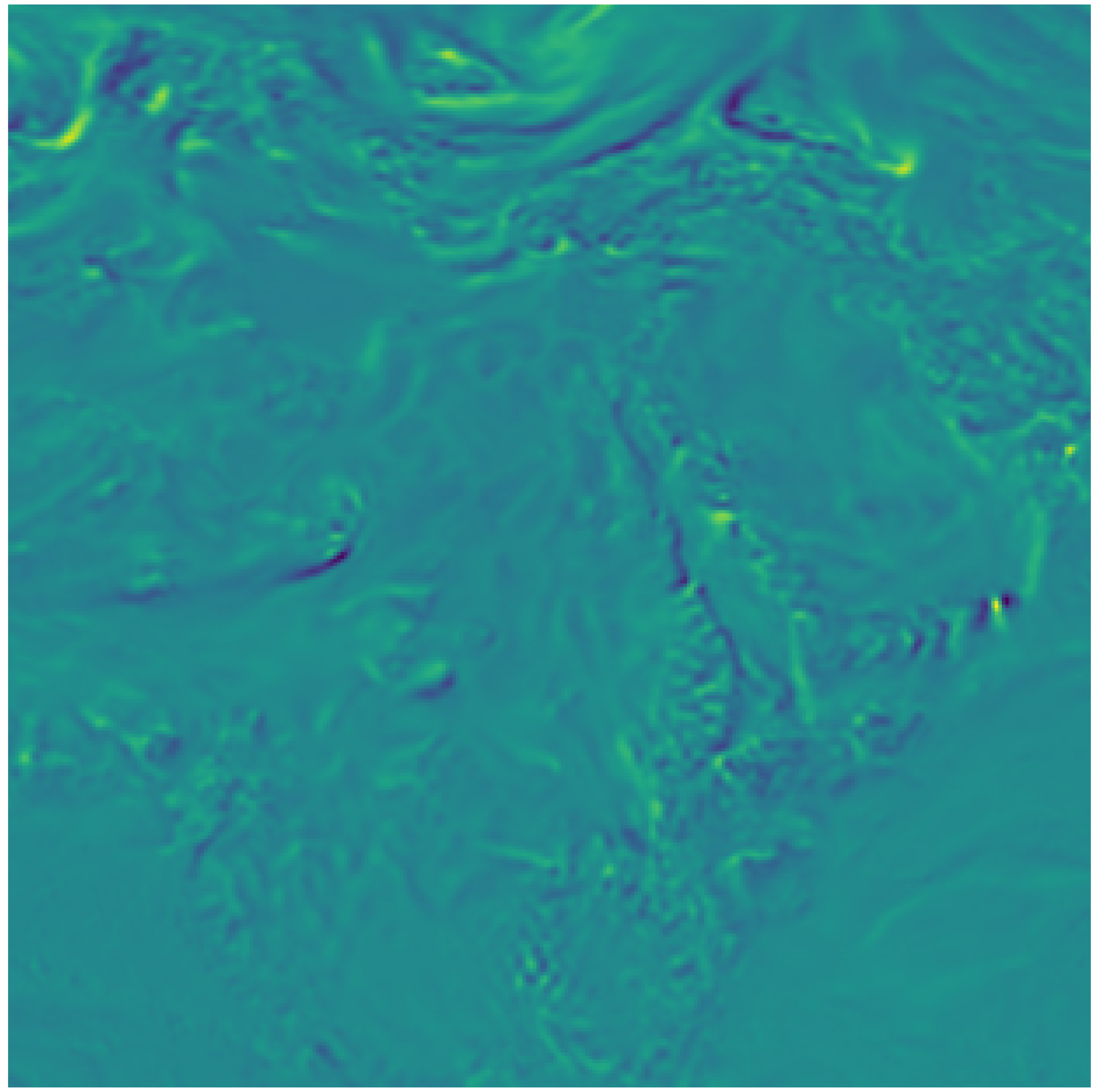
- Objective: down-scaling / upsampling of coarse fields
- Comparison using and with the GAN of Stengel et al.[1]
  - GAN is based on SRGAN[2] for natural images
  - Our content loss replaces mean squared error
  - Only 4X super-resolution in our work

[1] K. Stengel, A. Glaws, D. Hettinger, and R. N. King. Adversarial super-resolution of climatological wind and solar data. Proceedings of the National Academy of Sciences, 117(29):16805–16815, 2020.

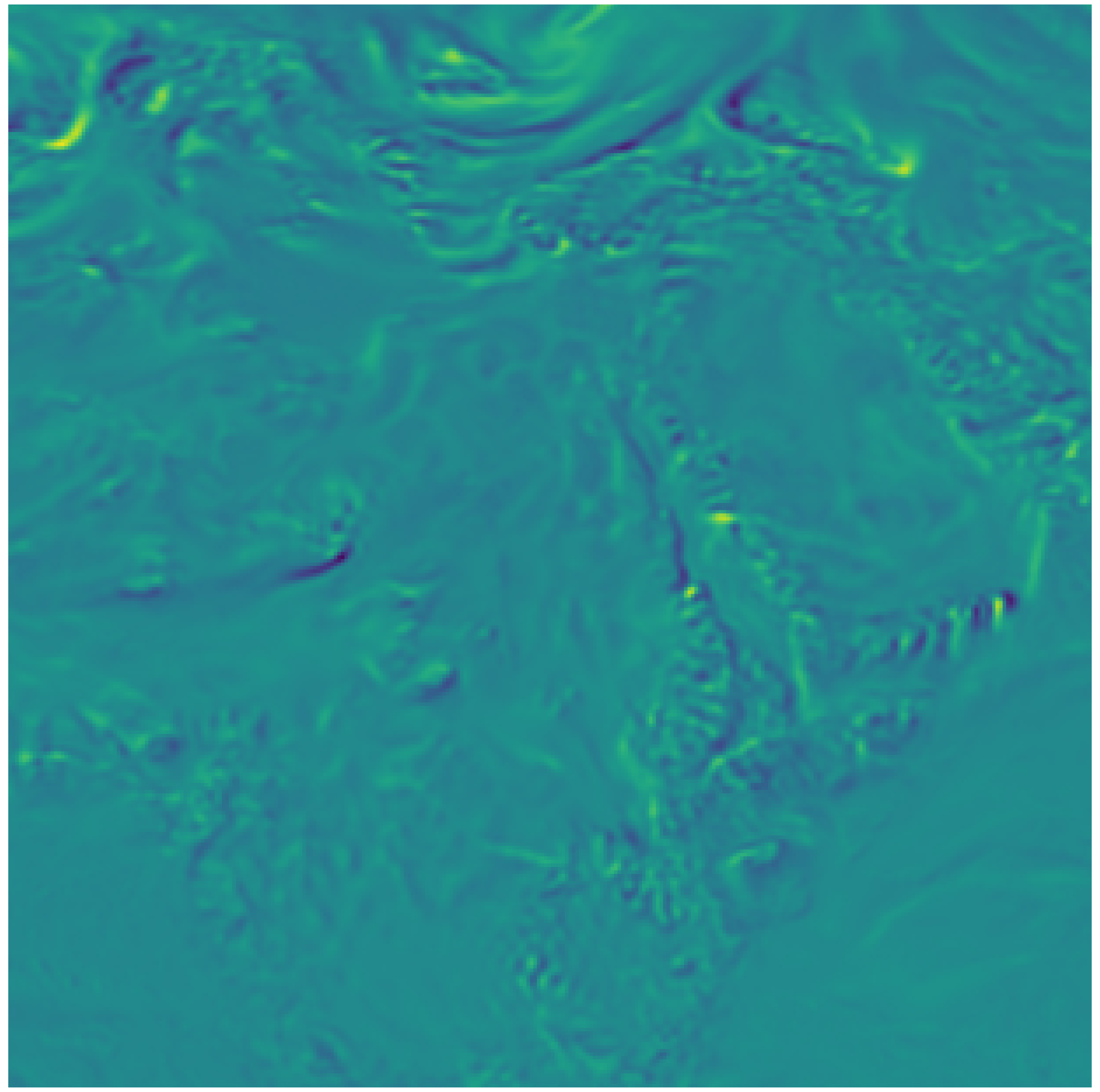
[2] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of CVPR, July 2017.

Super-resolution using AtmoDist

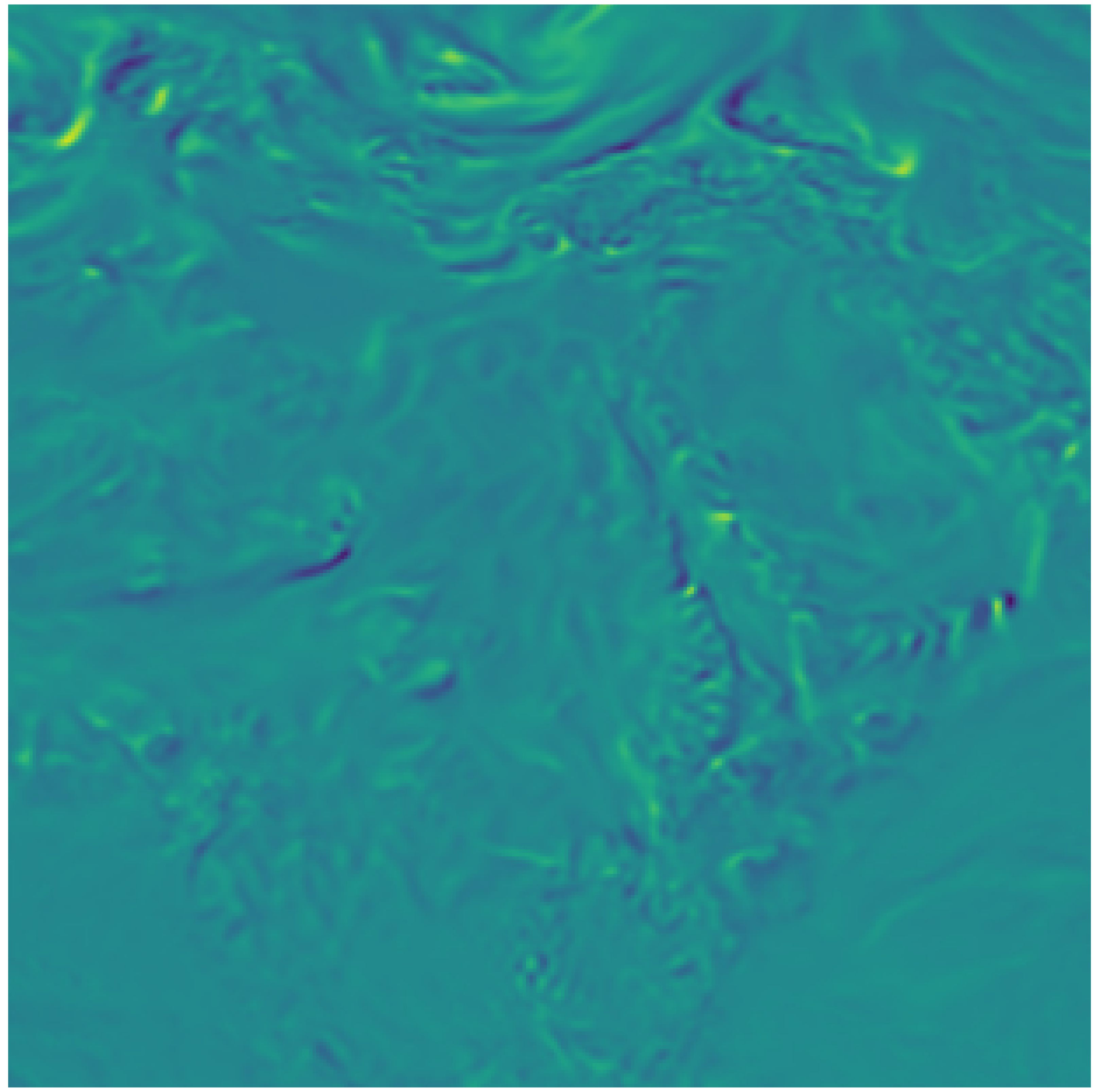
ours



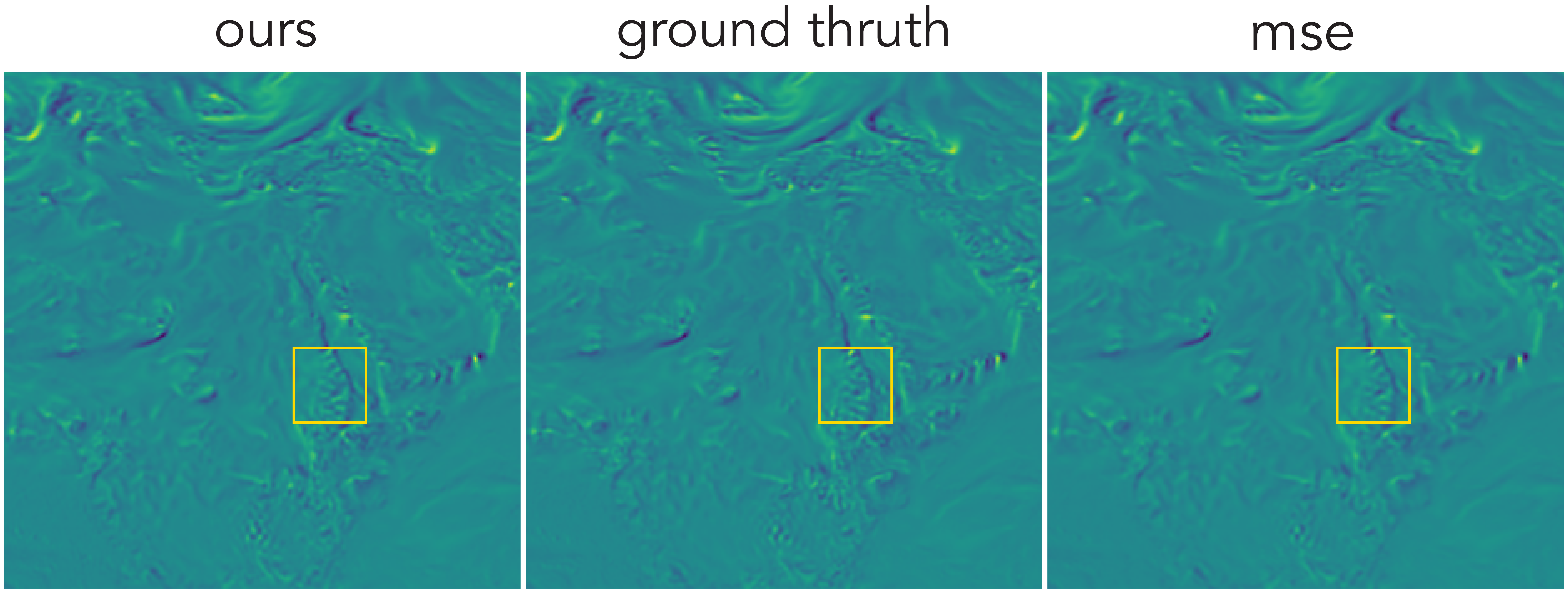
ground thruth



mse

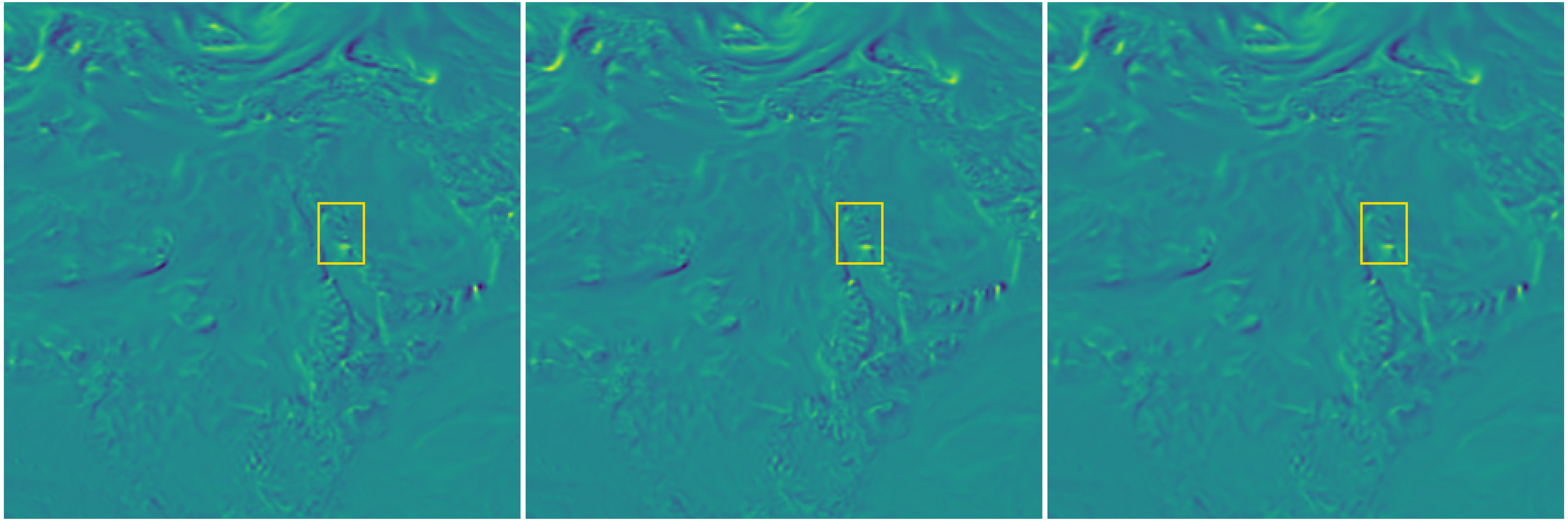


Super-resolution using AtmoDist



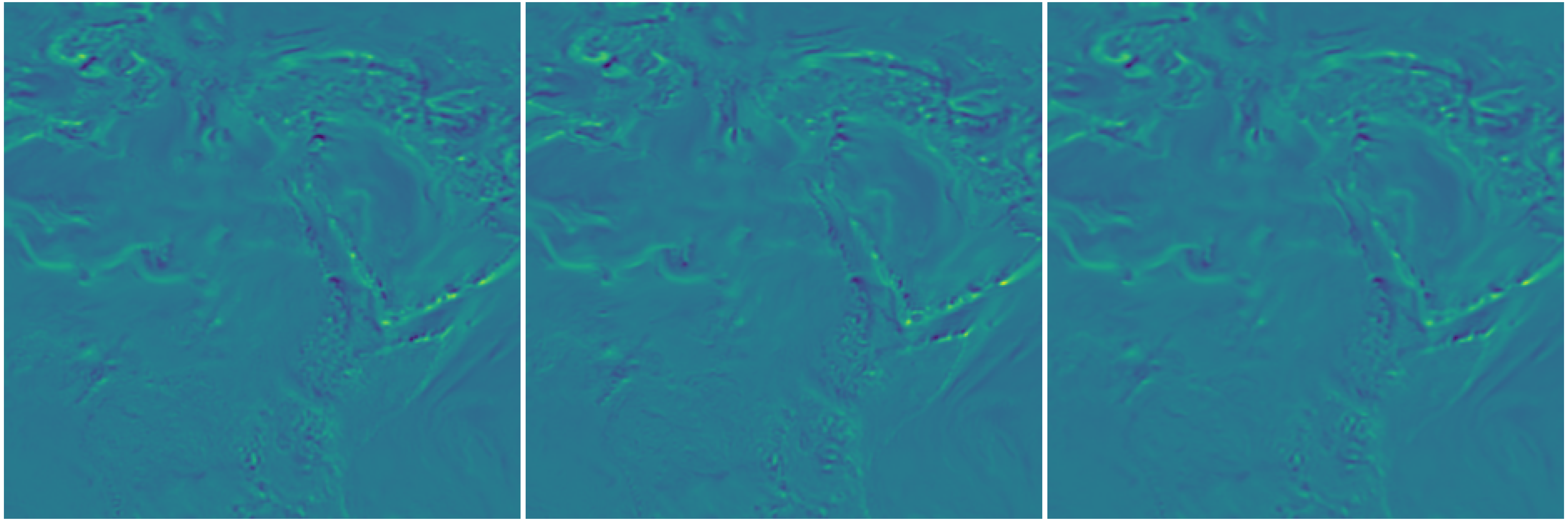
Super-resolution using AtmoDist

ours ground thruth mse

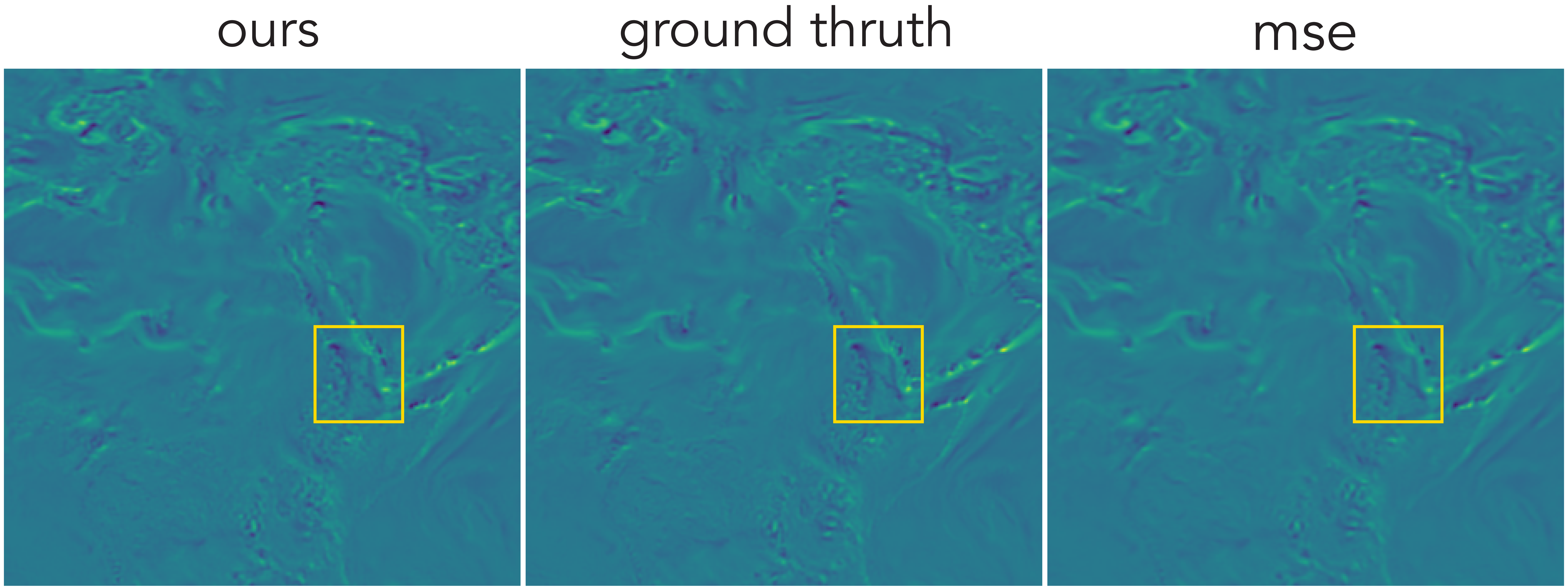


Super-resolution using AtmoDist

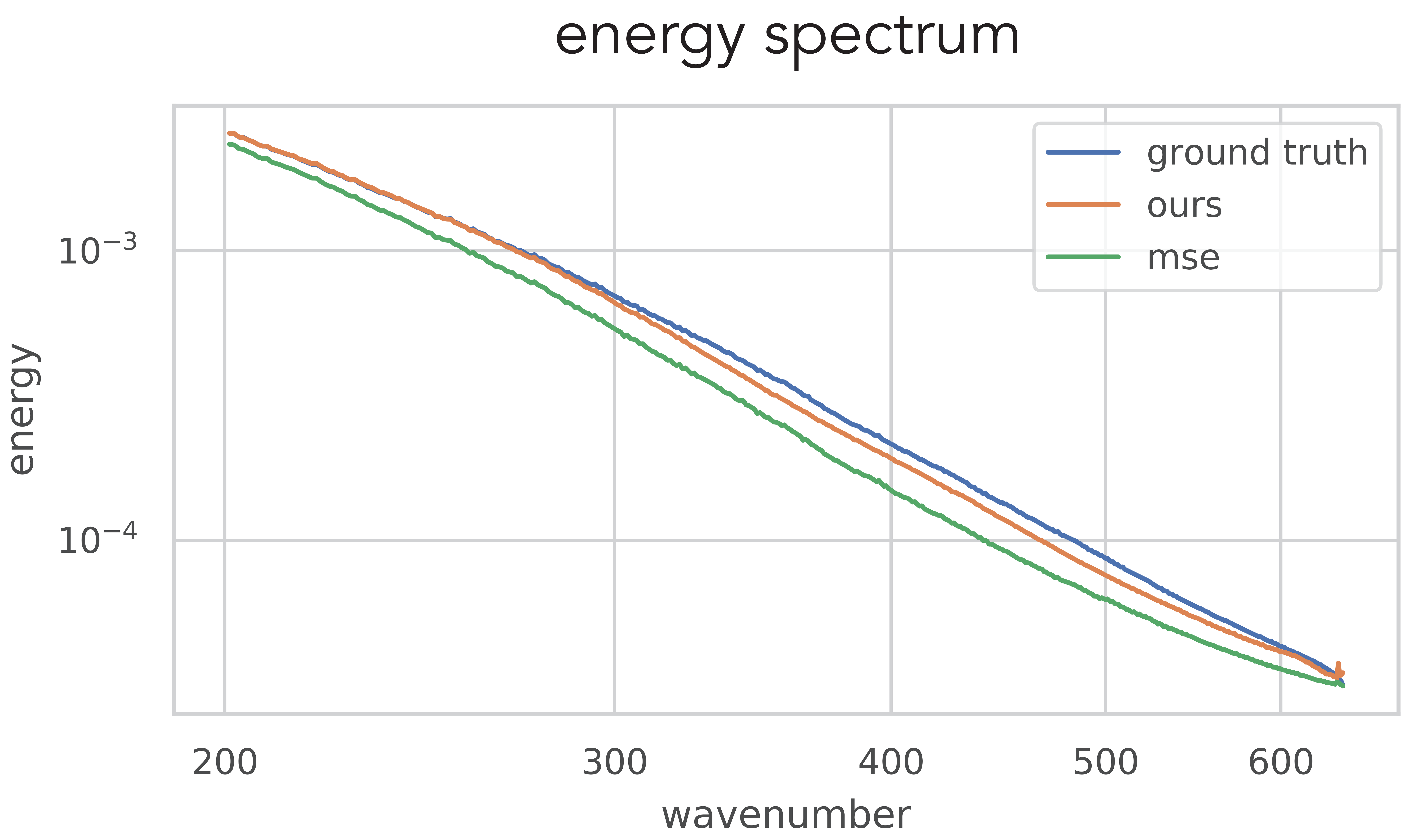
ours ground thruth mse



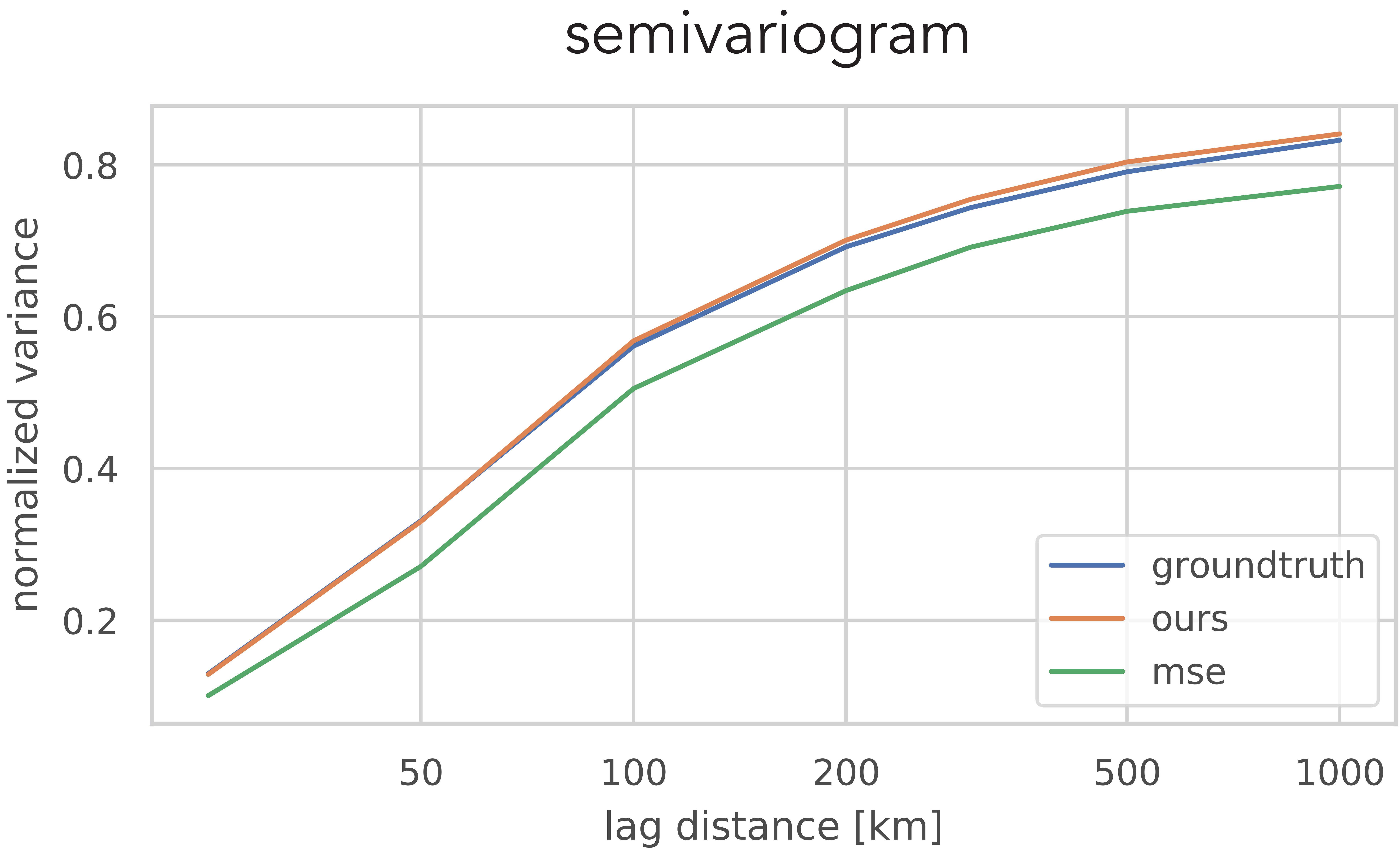
Super-resolution using AtmoDist



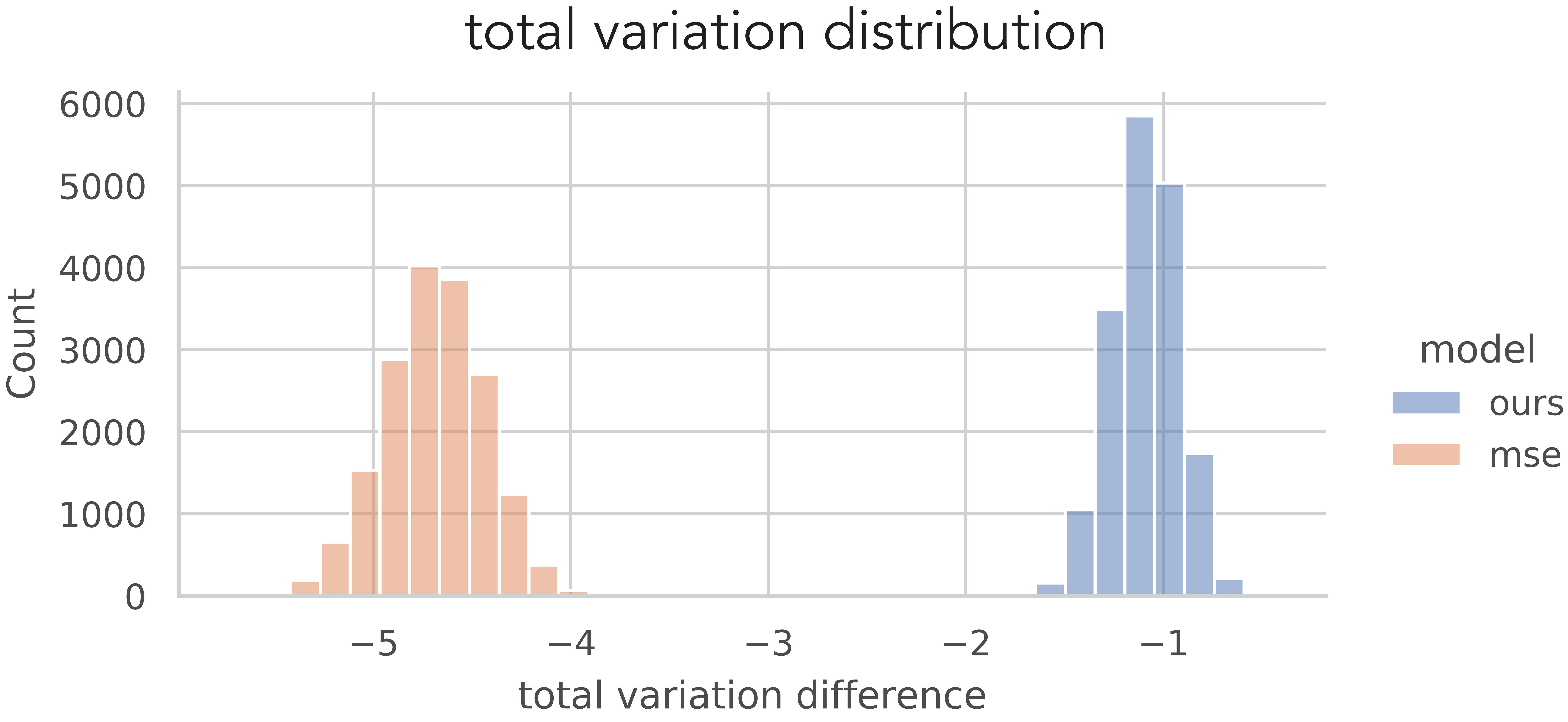
Super-resolution using AtmoDist



Super-resolution using AtmoDist



Super-resolution using AtmoDist

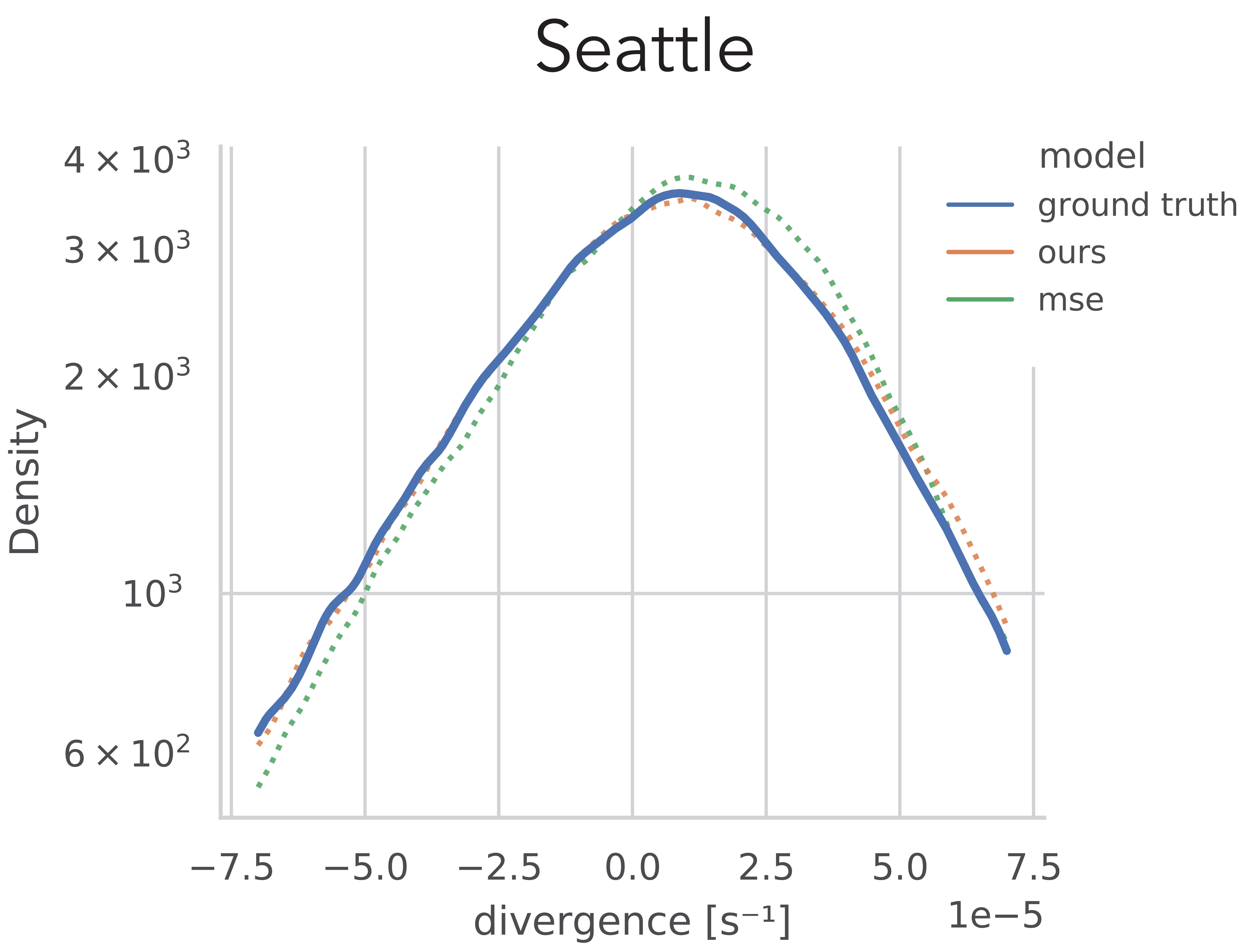


Super-resolution using AtmoDist

- Local statistics by averaging over super-resolution predictions for entire reanalysis data set
- 150 big cities as locations

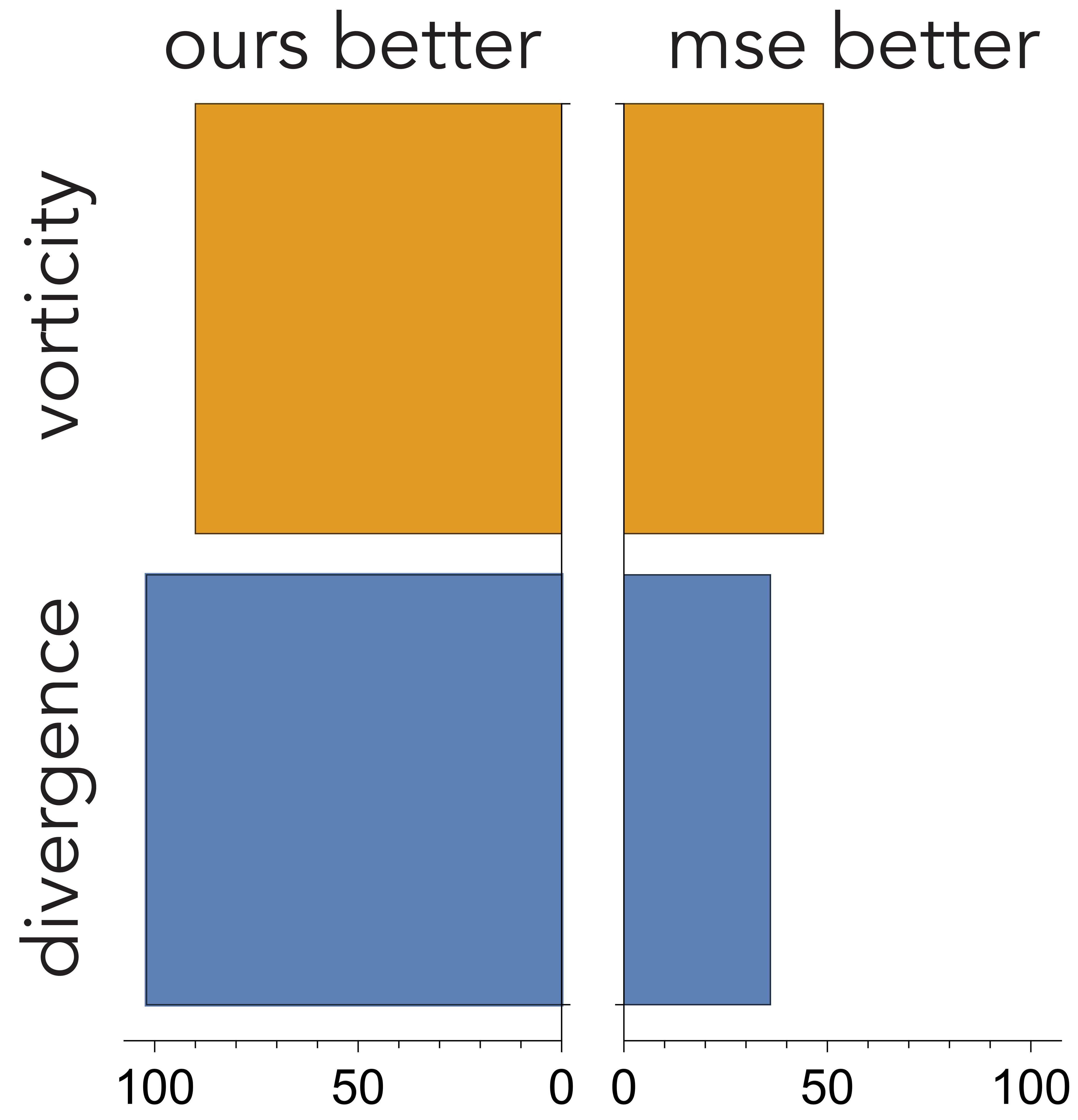
Super-resolution using AtmoDist

- Local statistics by averaging over super-resolution pre-dictions for entire reanalysis data set
- 150 big cities as locations



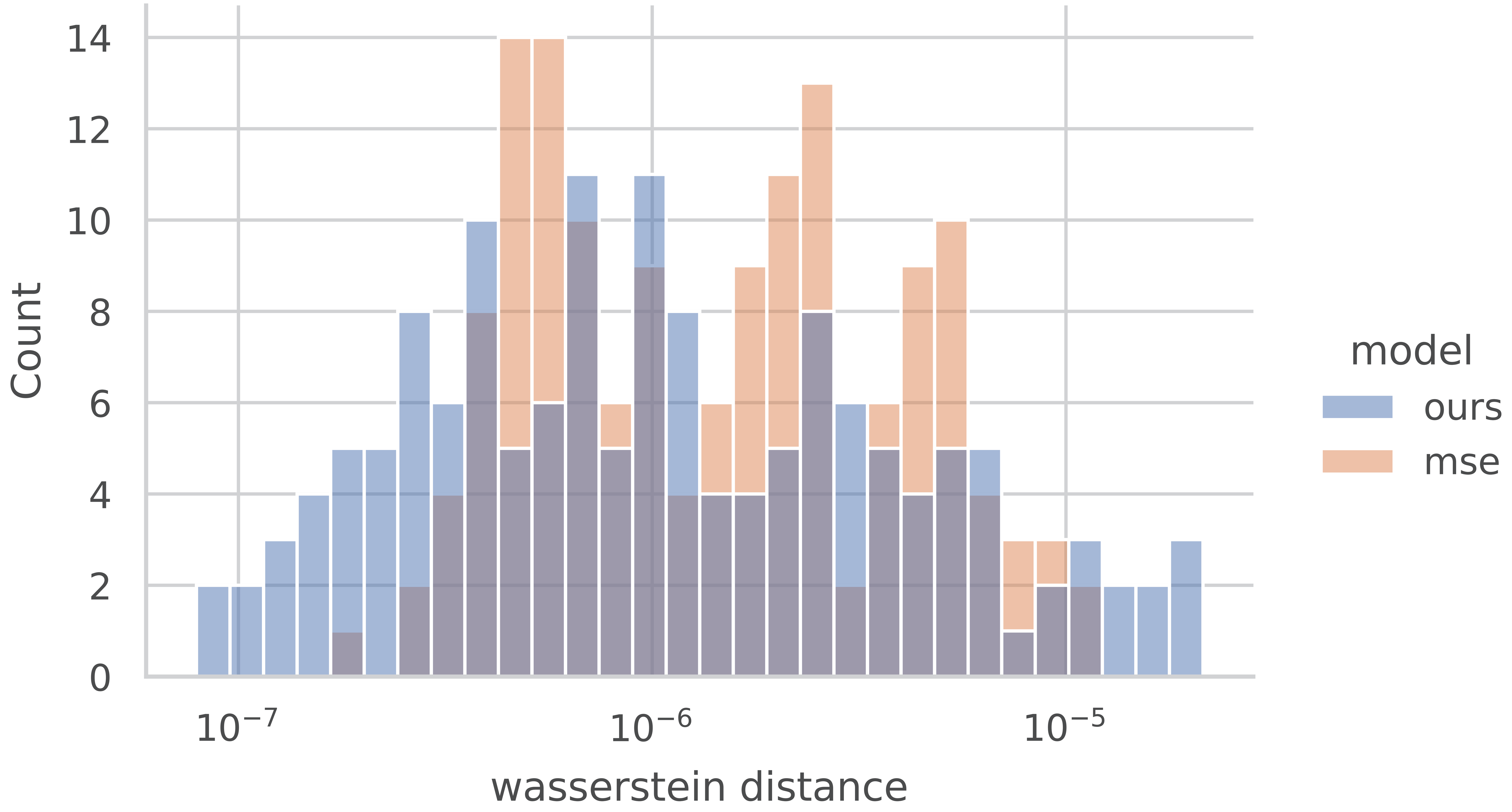
Super-resolution using AtmoDist

- Local statistics by averaging over super-resolution pre-dictions for entire reanalysis data set
- 150 big cities as locations



Super-resolution using AtmoDist

Wasserstein distance to ground truth for local statistics



Summary

- Custom loss functions for atmospheric data are possible and useful
  - Prediction of time difference is an effective pretext task
  - Applicable to wide range of data sets

Summary

- Custom loss functions for atmospheric data are possible and useful
  - Prediction of time difference is an effective pretext task
  - Applicable to wide range of data sets
- Super-resolution using AtmoDist improves quantitative and qualitative results
  - Local statistics still need more work

Outlook: applications

- Super-resolution / downscaling
  - Larger amplification factor
  - Adapted to local regions / specific phenomena

Outlook: applications

- Super-resolution / downscaling
  - Larger amplification factor
  - Adapted to local regions / specific phenomena
- Detection of atmospheric patterns
  - Extreme events, blocking, ...

Outlook: applications

- Super-resolution / downscaling
  - Larger amplification factor
  - Adapted to local regions / specific phenomena
- Detection of atmospheric patterns
  - Extreme events, blocking, ...
- Hybrid dynamical core for GCMs

Outlook: applications

- Super-resolution / downscaling
  - Larger amplification factor
  - Adapted to local regions / specific phenomena
- Detection of atmospheric patterns
  - Extreme events, blocking, ...
- Hybrid dynamical core for GCMs
- Applications directly addressing climate change

Outlook: custom loss functons

- Other / more atmospheric fields
  - Separate dynamic variables and tracers?
  - More vertical layers

Outlook: custom loss functons

- Other / more atmospheric fields
  - Separate dynamic variables and tracers?
  - More vertical layers
- Transfer learning / refinement to local regions
  - Train with only local data in a final training phase

Outlook: custom loss functons

- Other / more atmospheric fields
  - Separate dynamic variables and tracers?
  - More vertical layers
- Transfer learning / refinement to local regions
  - Train with only local data in a final training phase
- Shift in distribution (e.g. global warming)
  - Parametrized model?

Outlook: representation learning

- Un-/semi-supervised learning has significant potential for atmospheric data
  - Side-steps need for labelling of data
  - Generically used for pre-training with natural images

Outlook: representation learning

- Un-/semi-supervised learning has significant potential for atmospheric data
  - Side-steps need for labelling of data
  - Generically used for pre-training with natural images
- Integrate heterogeneous data sources
  - Reanalysis and simulations
  - High frequency, high resolution satellite data

Outlook: theoretical foundations

- Theoretical integration of analytic models and machine learning
  - E.g. predict effective network architecture
- Consistency, stability, and convergence for (hybrid) simulations
- Predict effectiveness of learning / data for climate dynamics

AtmoDist: https://arxiv.org/abs/2109.09076

Slides via KITP Online Talks repository
and on my hompage