Analog Forecasting of Extreme Weather Events: Applications of Structurally Equivariant Deep Learning and Data Assimilation

Ashesh Chattopadhyay, Ebrahim Nabizadeh, Pedram Hassanzadeh
Rice University, Houston

Mustafa Mustafa, Karthik Kashinath
Lawrence Berkeley National Laboratory, NERSC

Motivation

Data-driven Forecasting

- Numerical models (NWP)
- Multi-ensemble runs
- Data-assimilation

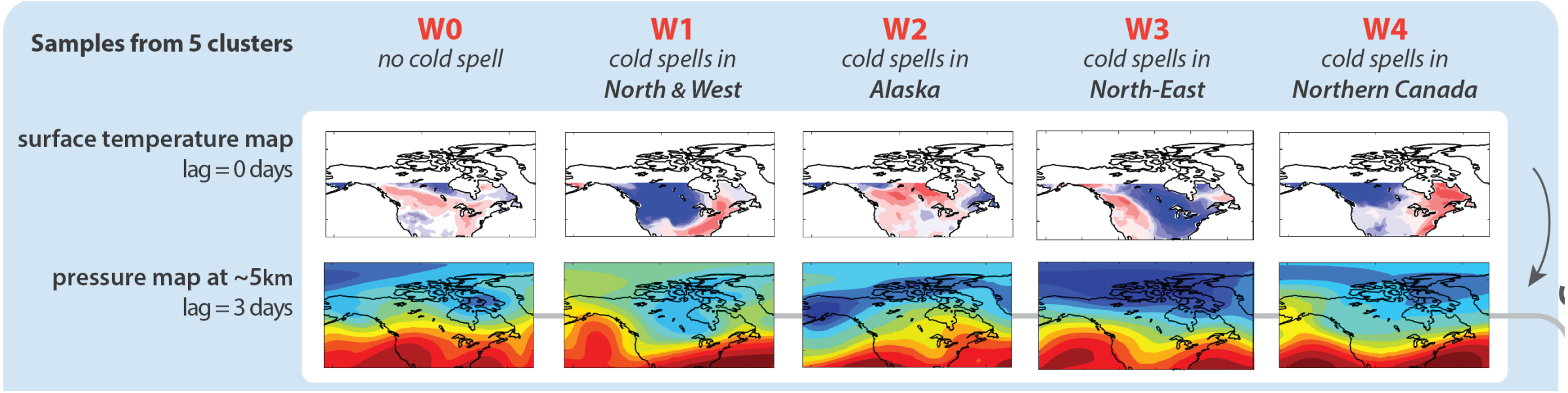
Weyn et al., JAMES 2019;
Scherr, GRL 2018;
Scherr et al., GMD 2019;
Chattopadhyay et al., NPG 2020

Extreme Weather Prediction

- Might be difficult for long lead times
- Precursors

Chattopadhyay et al., JAMES 2020

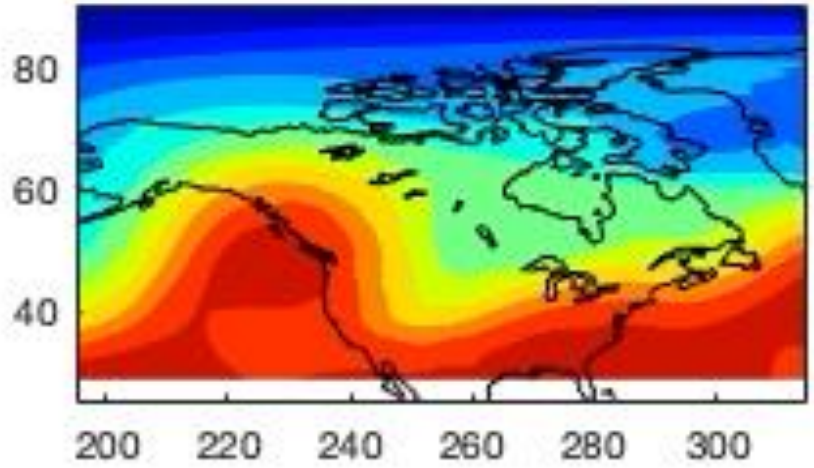
Extreme Weather Prediction



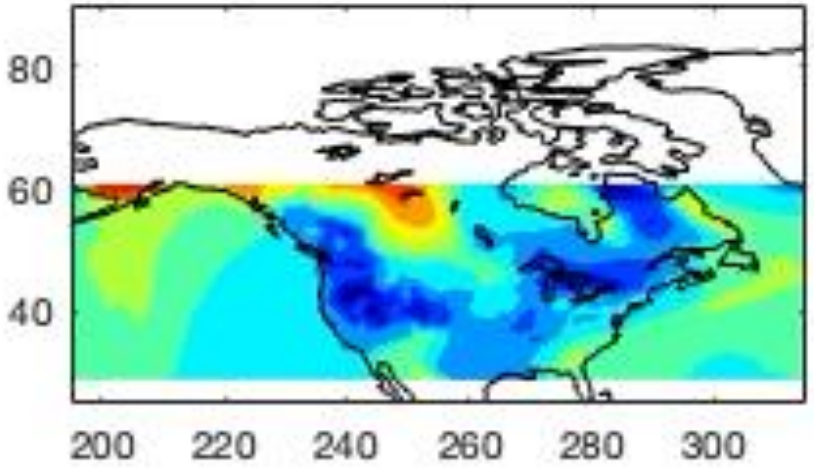
- CESM, 40 members, data from 1920-2005
- Heat Waves/Cold Spells between 30^o – 60^o North and 195^o – 315^o East

Chattopadhyay et al., JAMES 2020

Data-driven extreme weather prediction

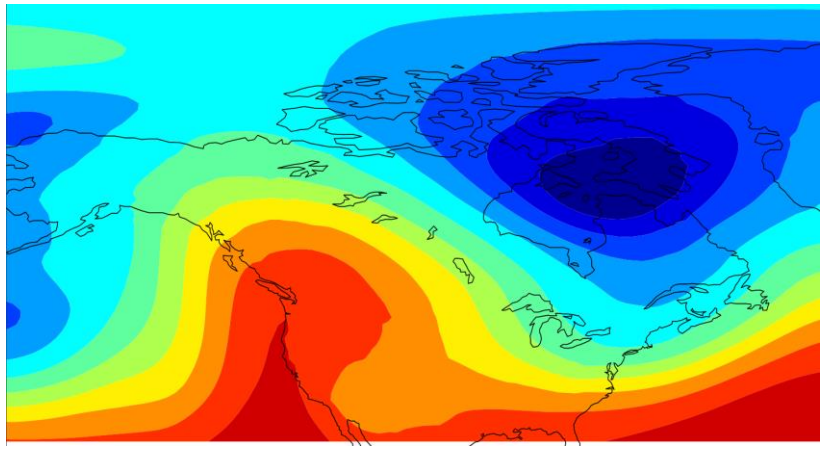


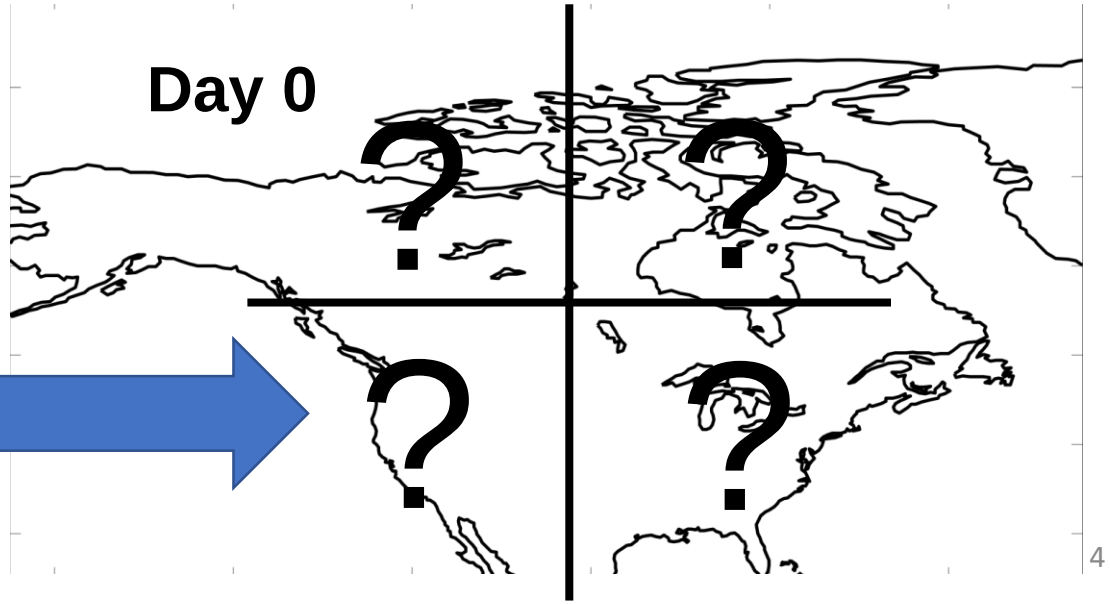
Z500 (day -3)



T2m (day 0)

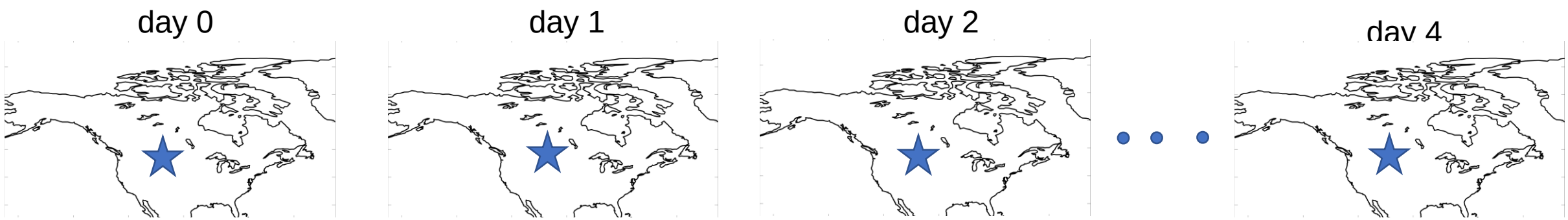
Day -3





Defining Onset of Extreme Events

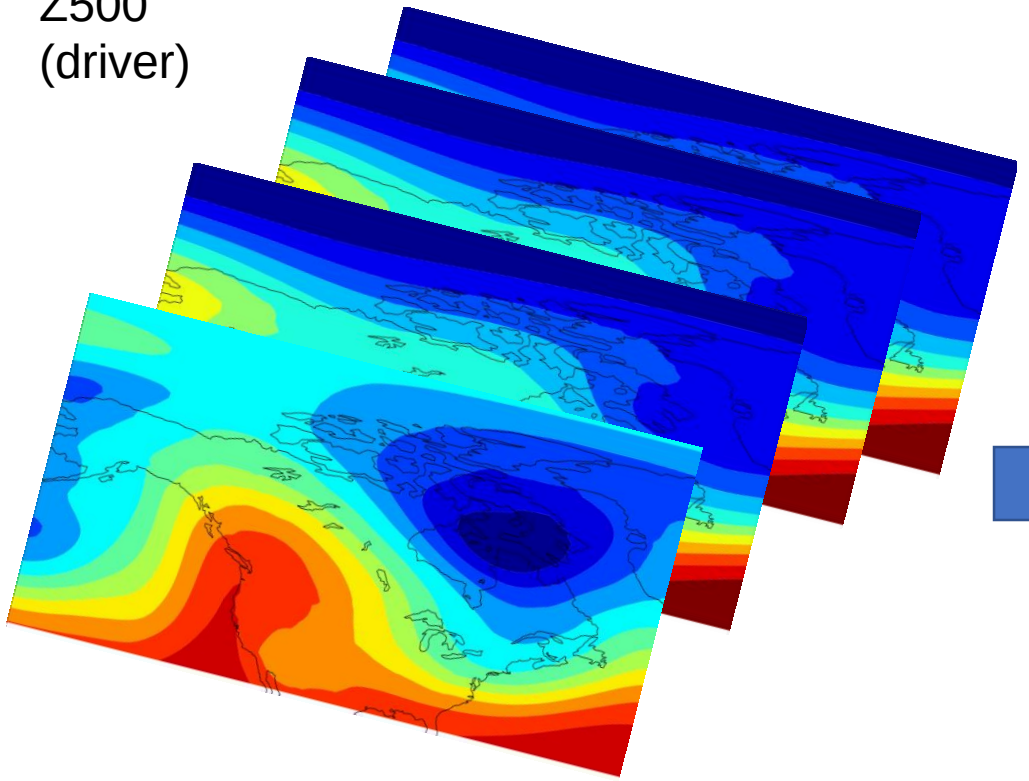
Extreme temperature



Onset

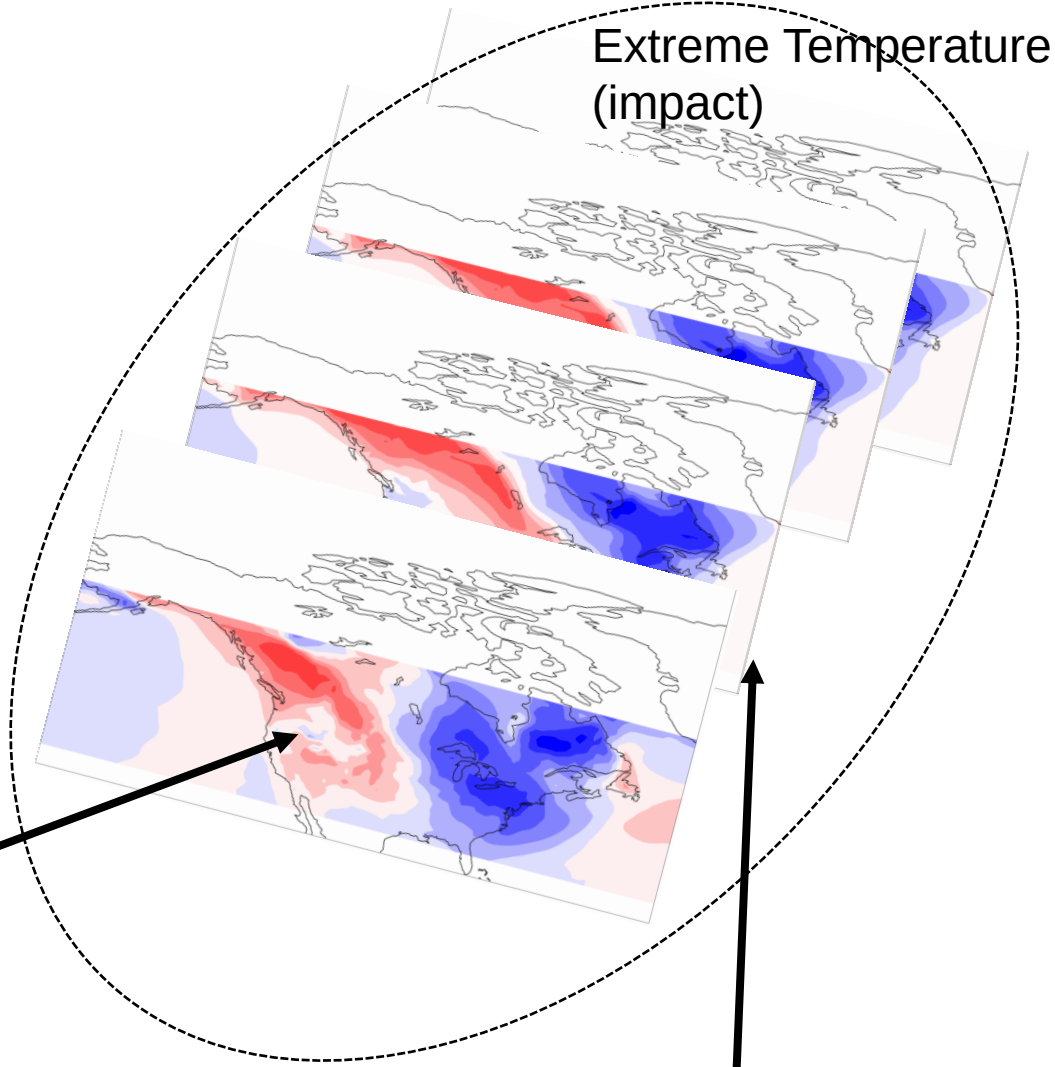
Labeling

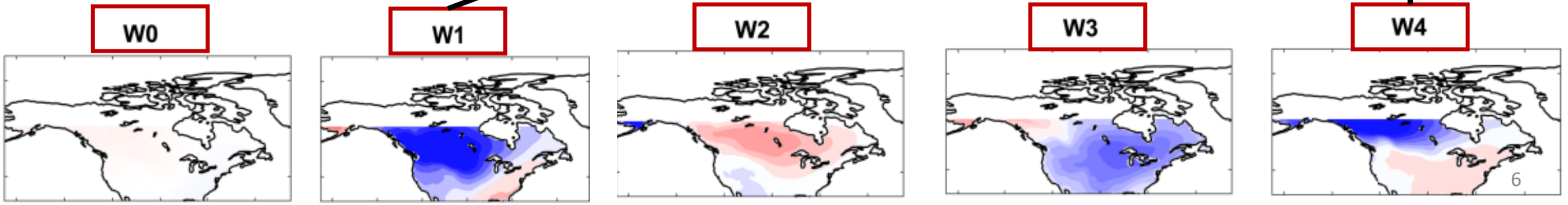
Z500
(driver)



Day -3

Extreme Temperature
(impact)

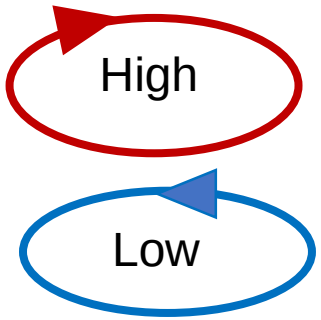


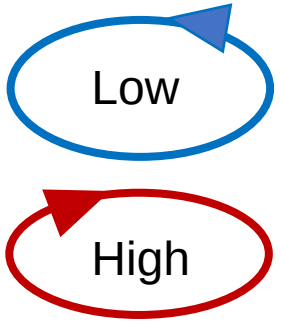


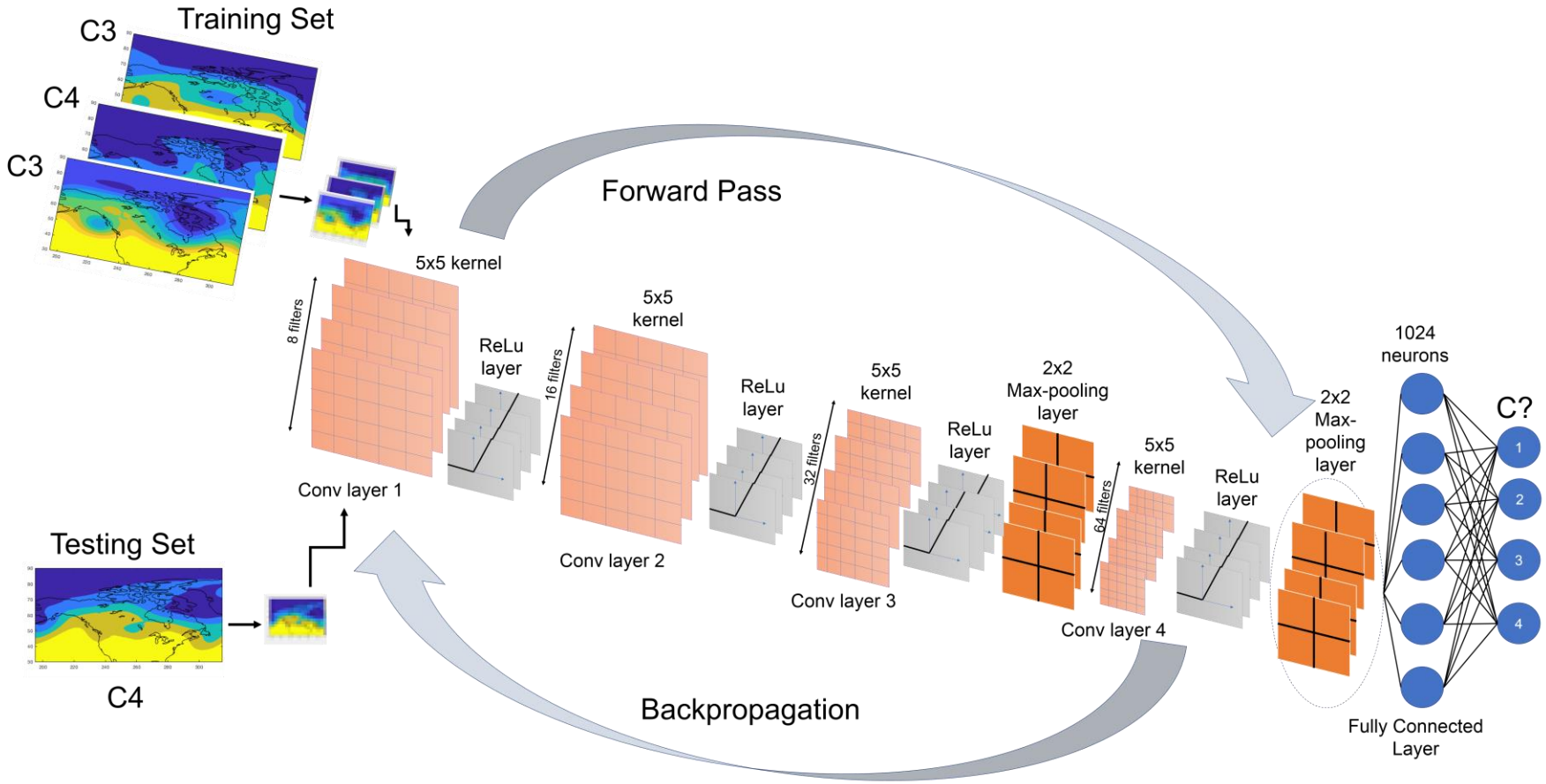
Convolutions: Is it really so trivial?

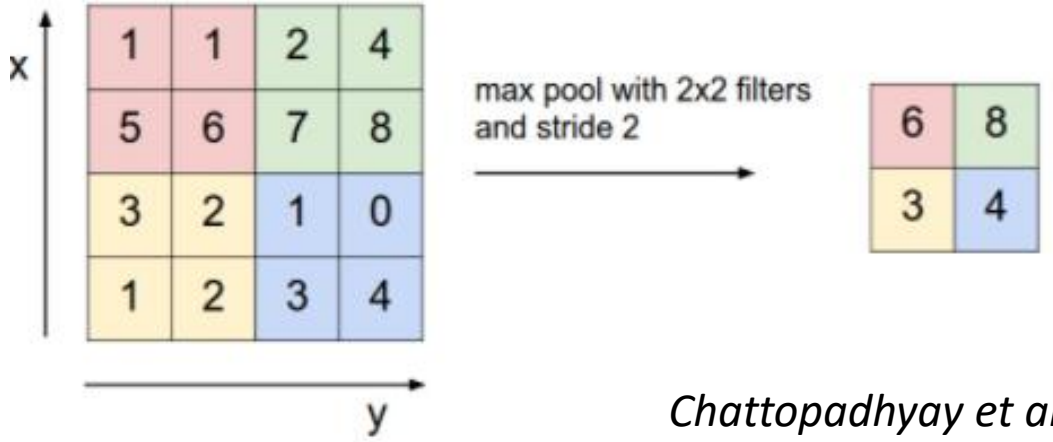
This must be a face!

This must be a face as well!



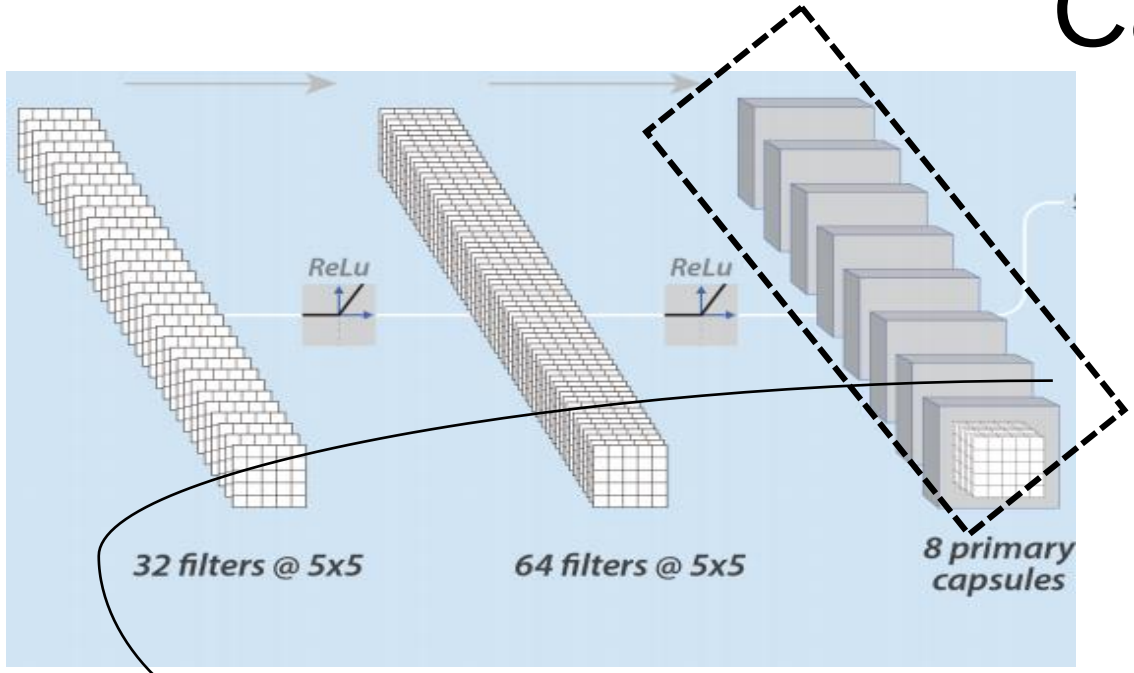


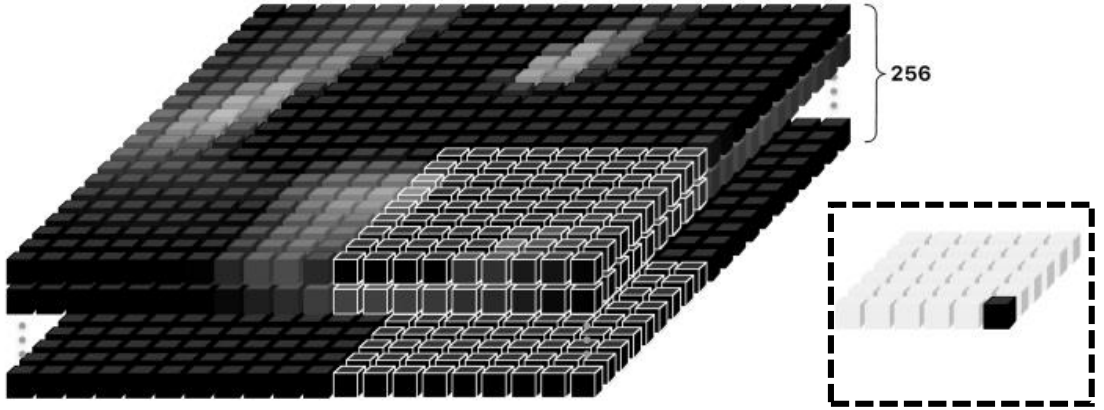


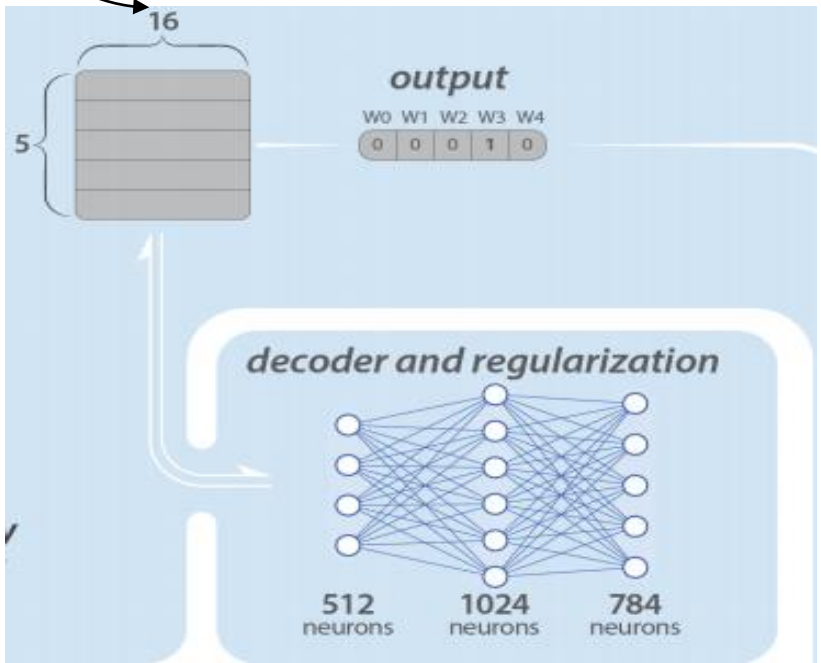


Chattopadhyay et al., Scientific Reports, 2020

Capsules



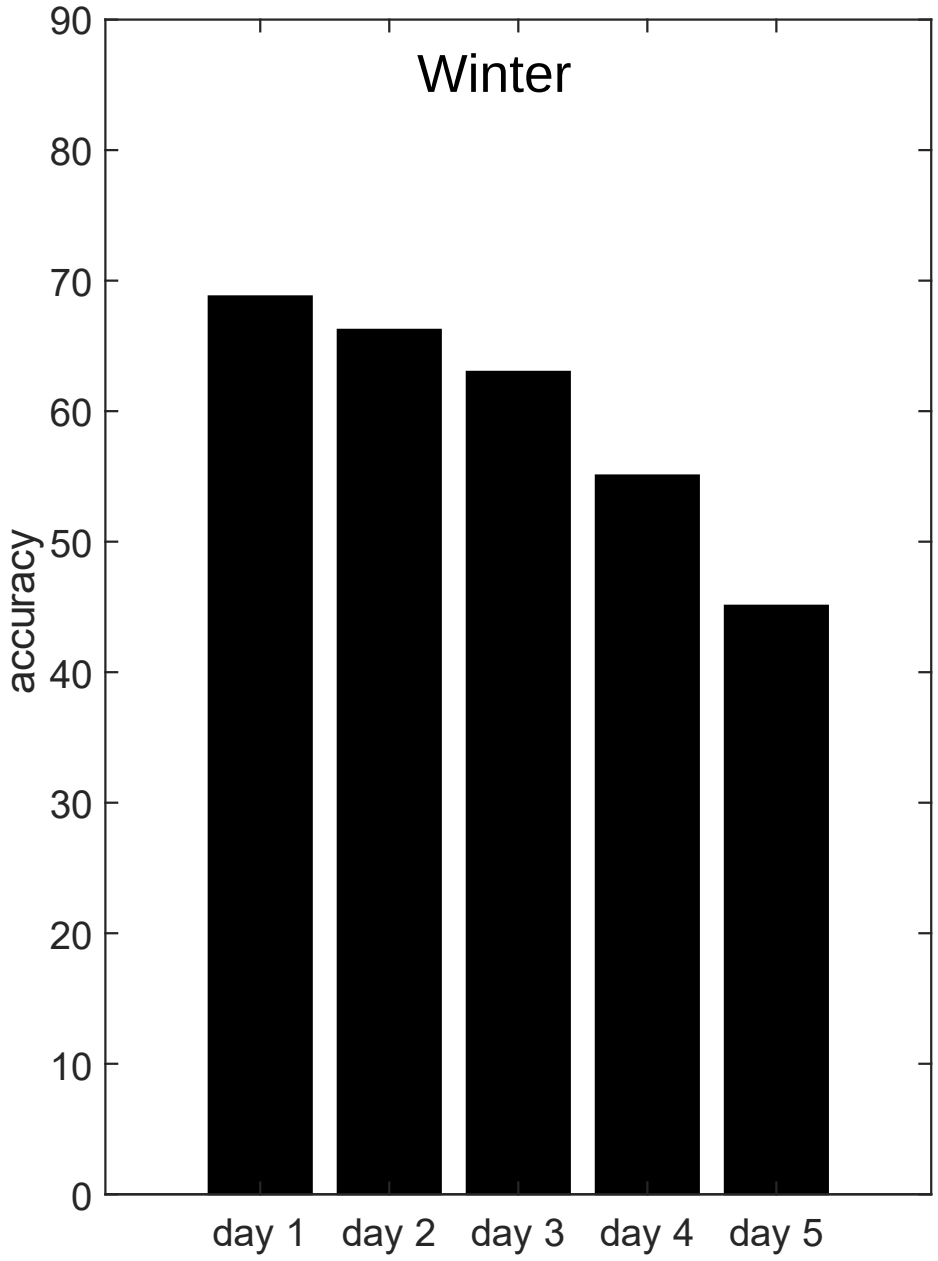


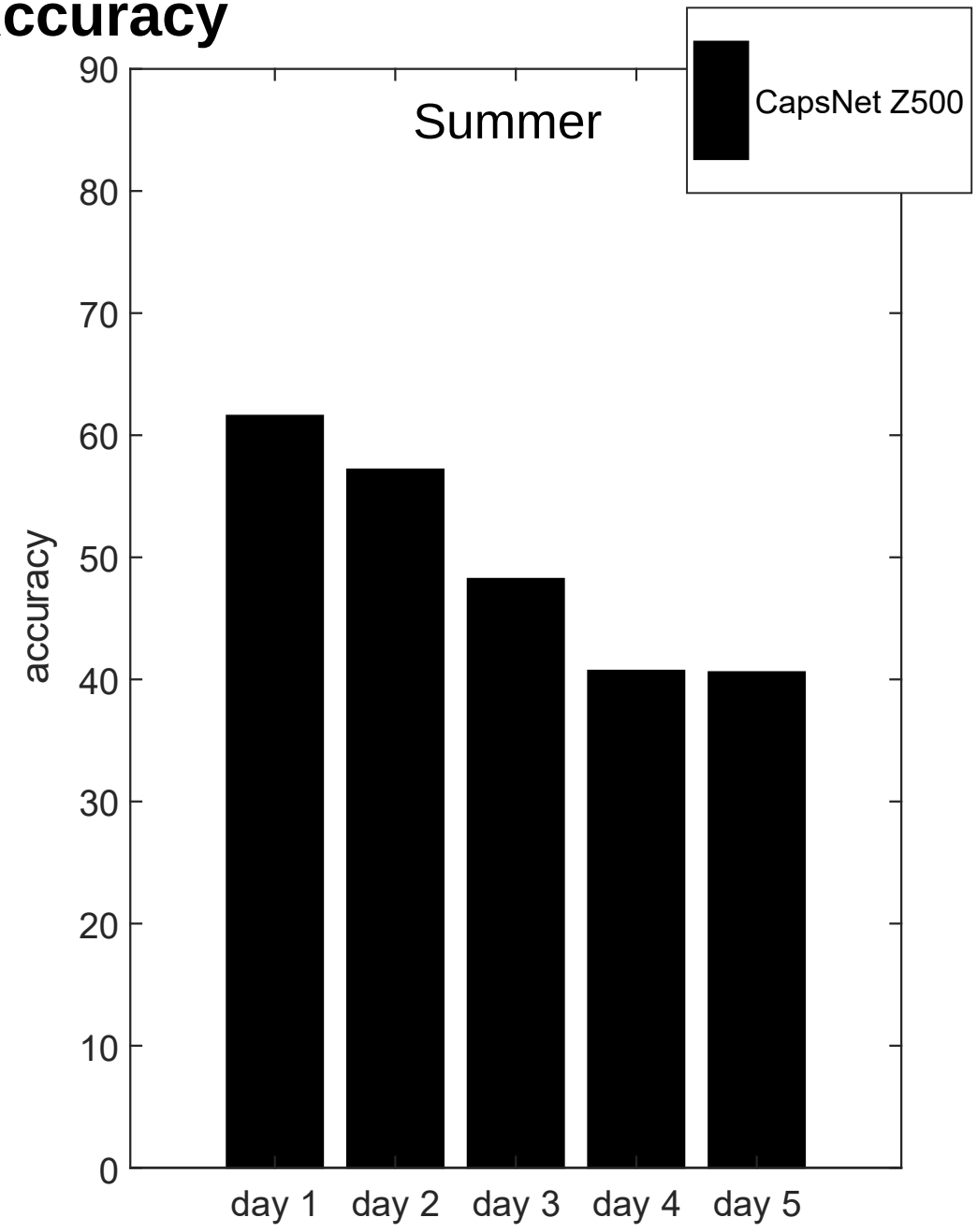


Affine Transform	$\hat{\mathbf{u}}_{j\|i} = \mathbf{W}_{ij}\mathbf{u}_i$
Weighting	$\mathbf{s}_j = \sum_i c_{ij}\hat{\mathbf{u}}_{j\|i}$
Sum	
Nonlinear Activation	$\mathbf{v}_j = \frac{\|\mathbf{s}_j\|^2}{1+\|\mathbf{s}_j\|^2}\frac{\mathbf{s}_j}{\|\mathbf{s}_j\|}$

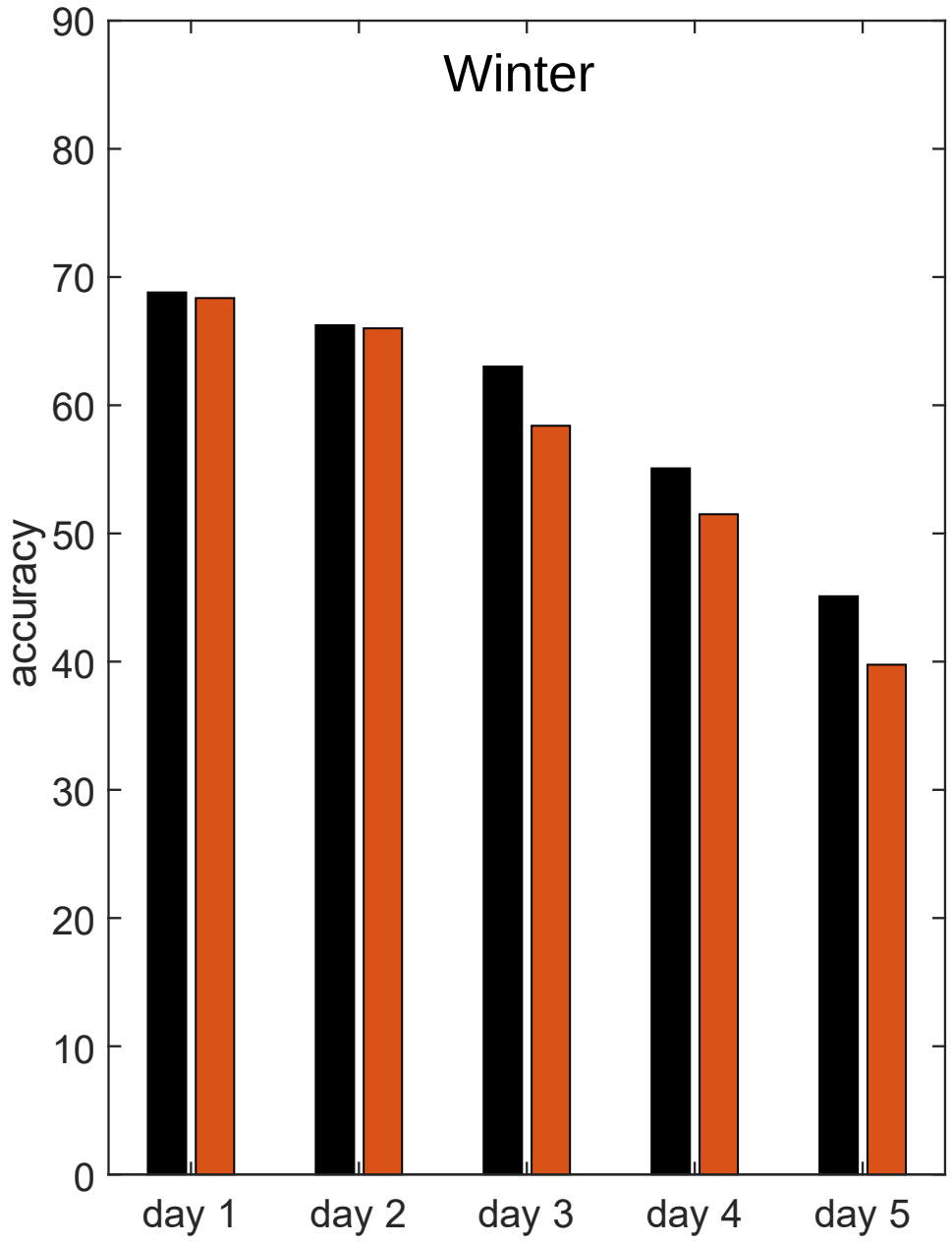
Sabour et al., Neurips 2018

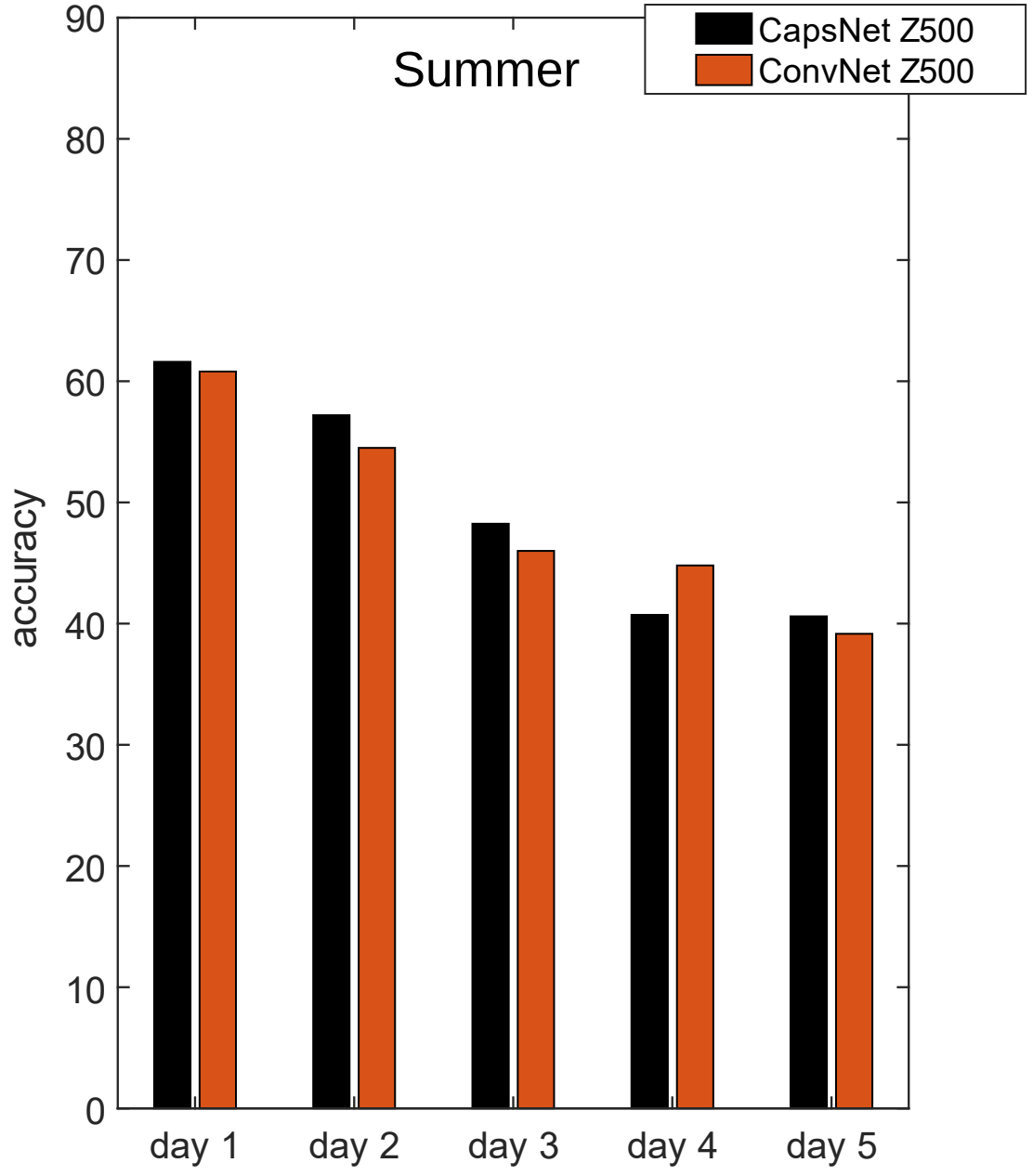
Results: Accuracy



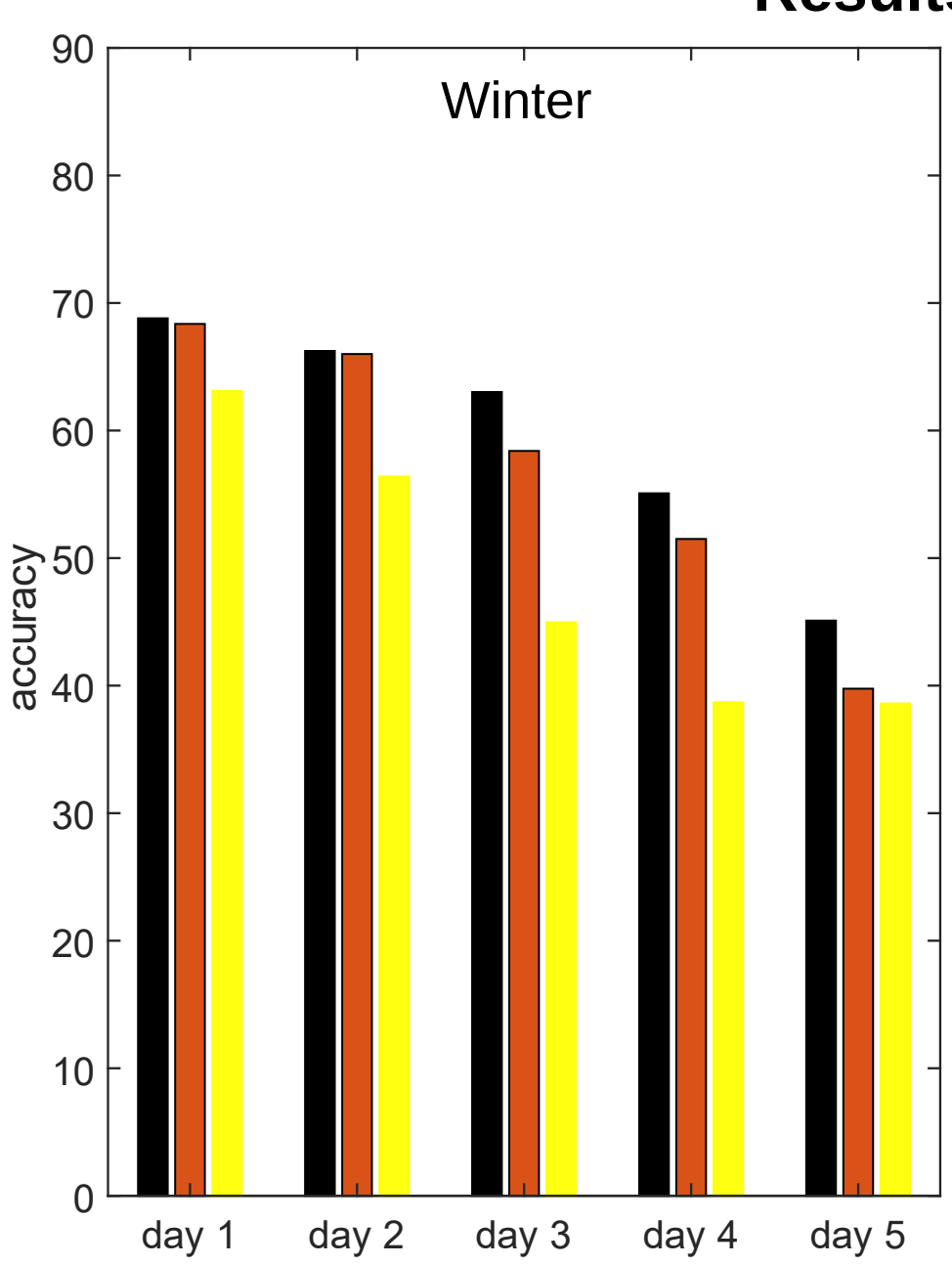


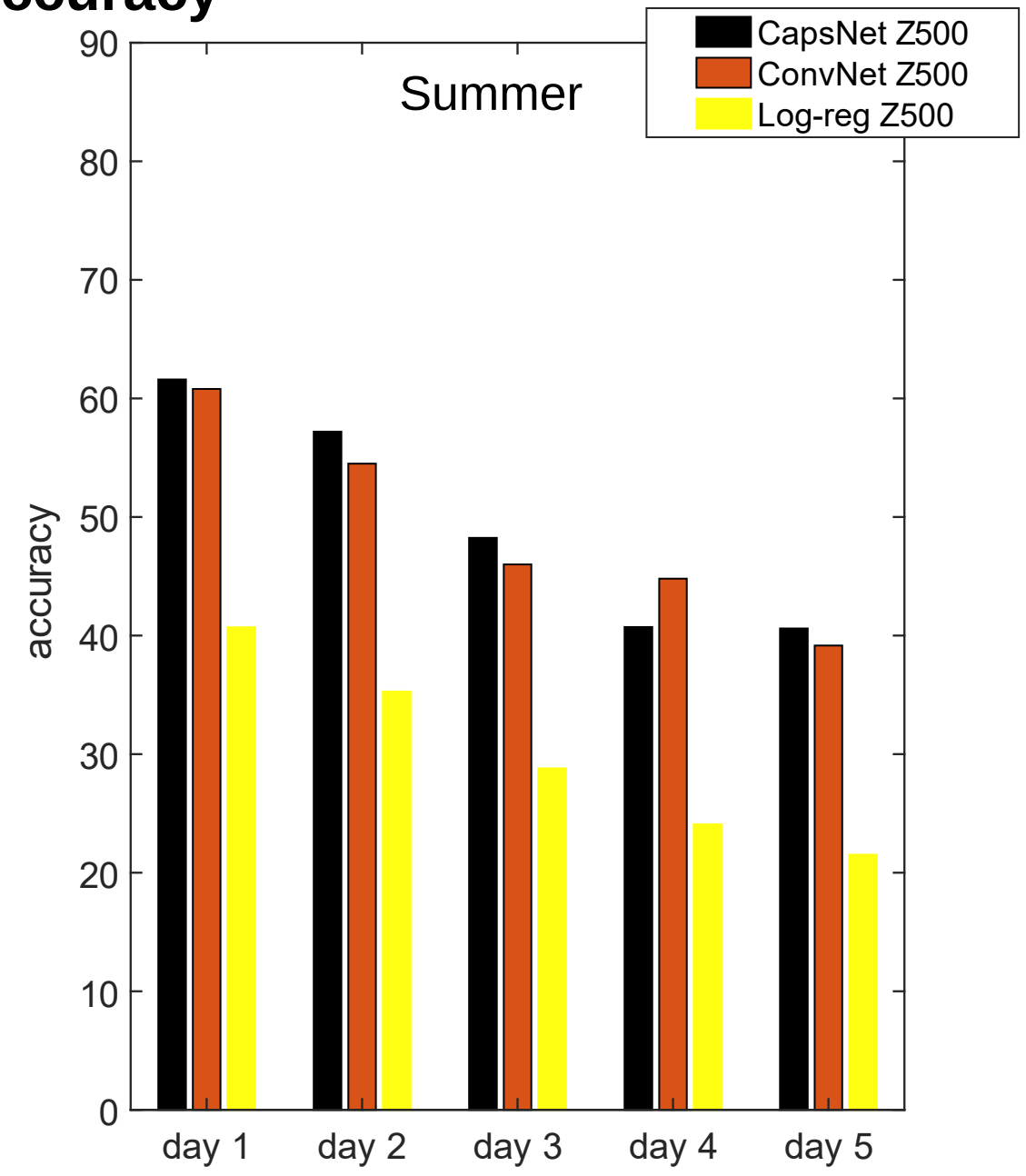
Results: Accuracy



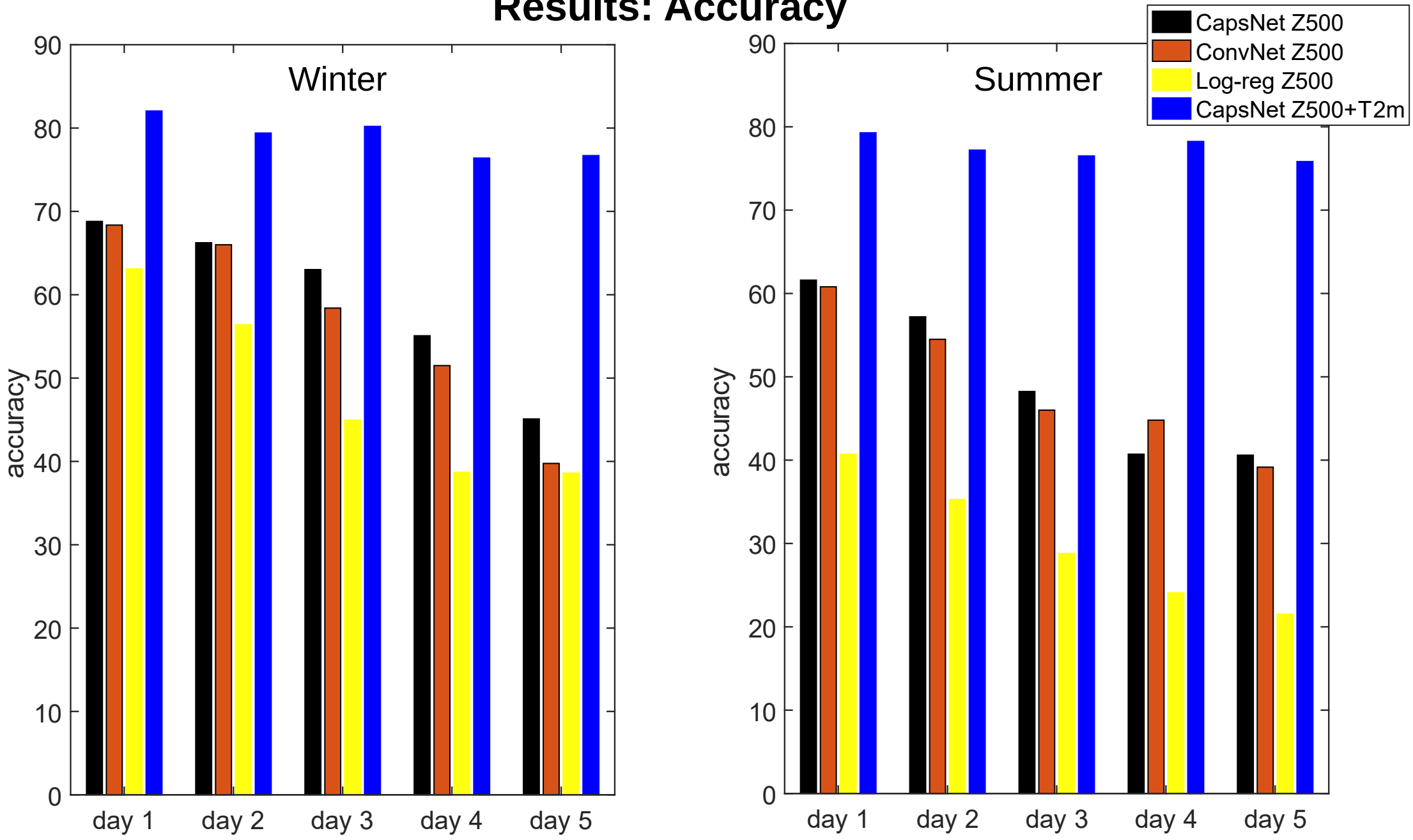


Results: Accuracy

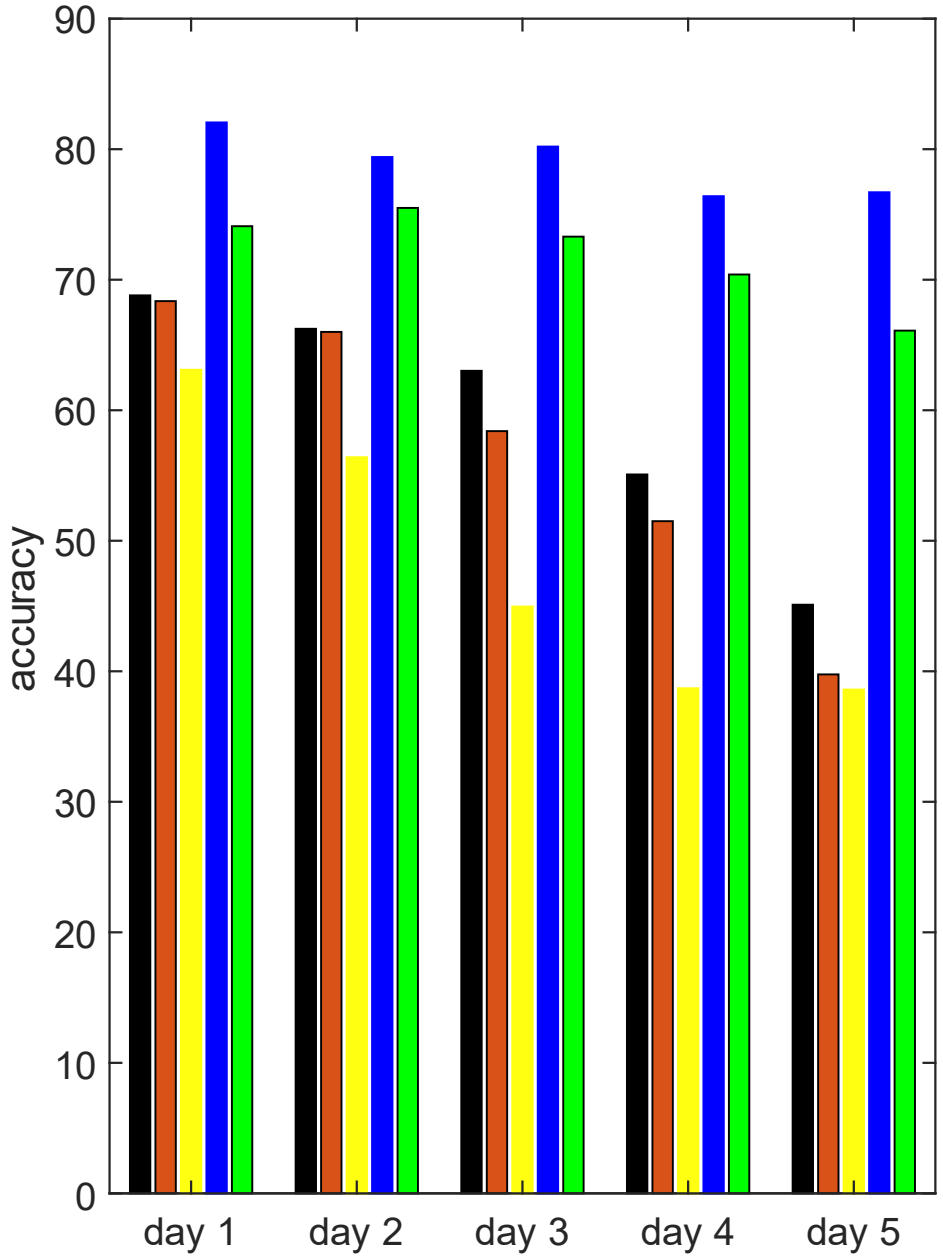


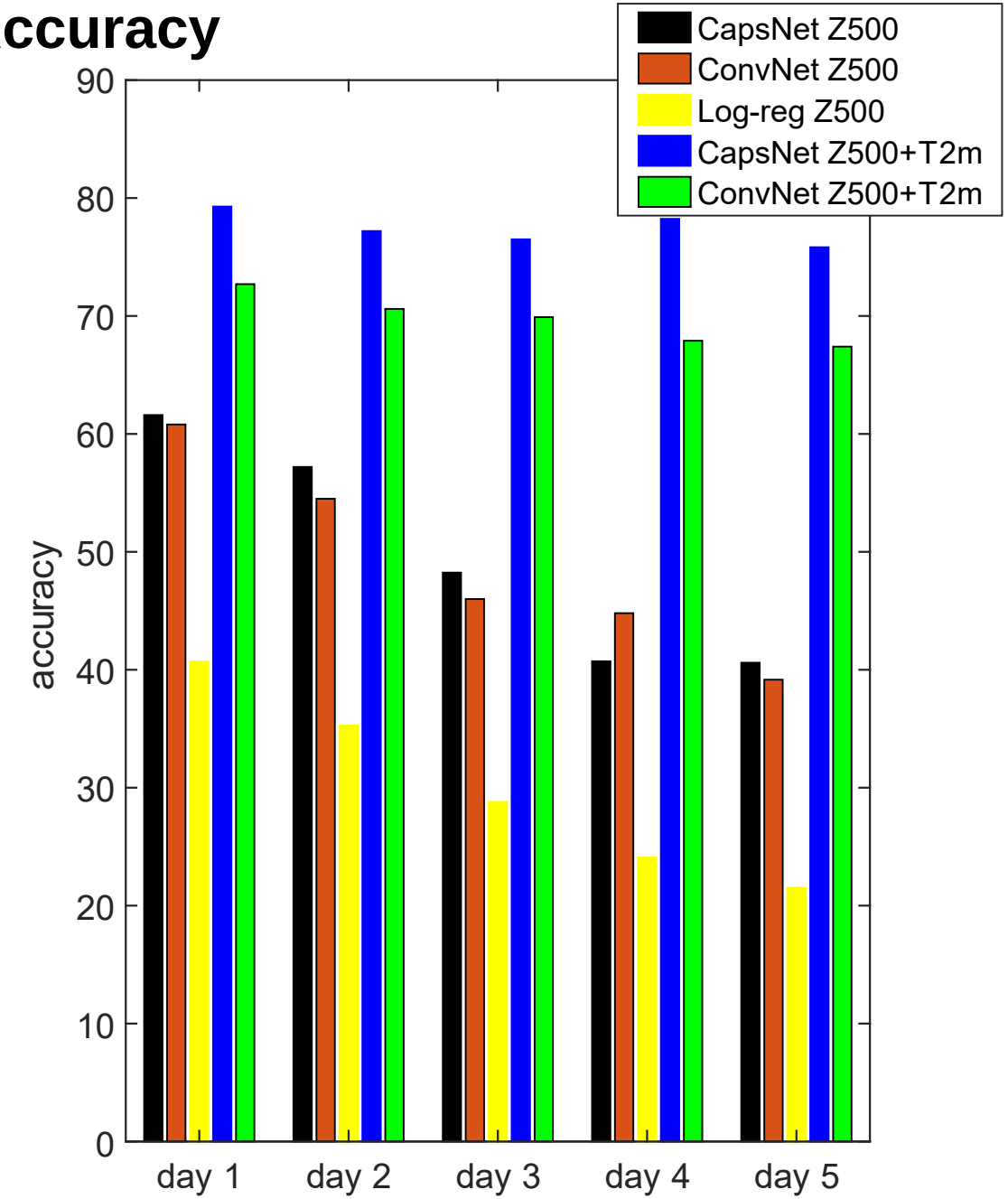


Results: Accuracy

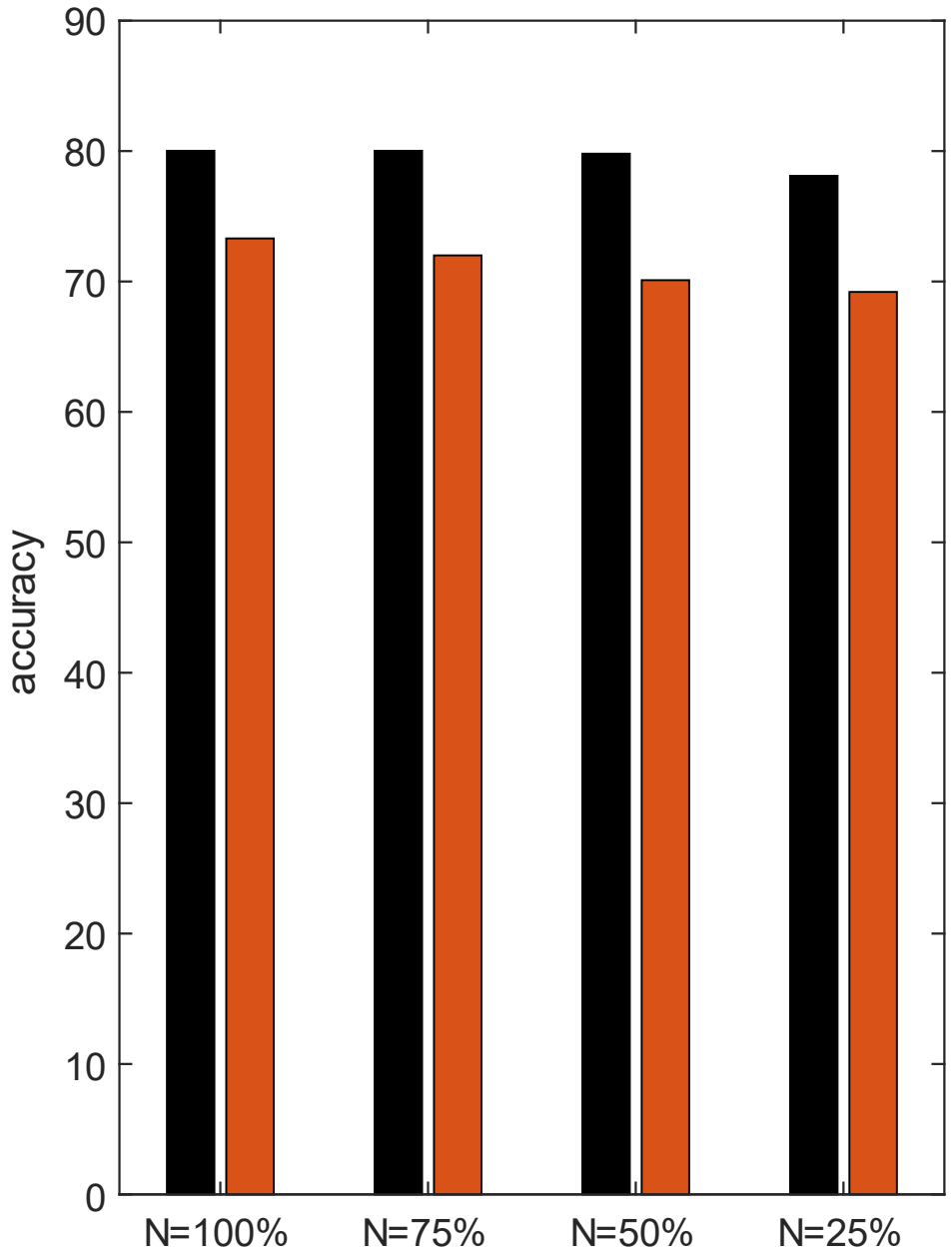


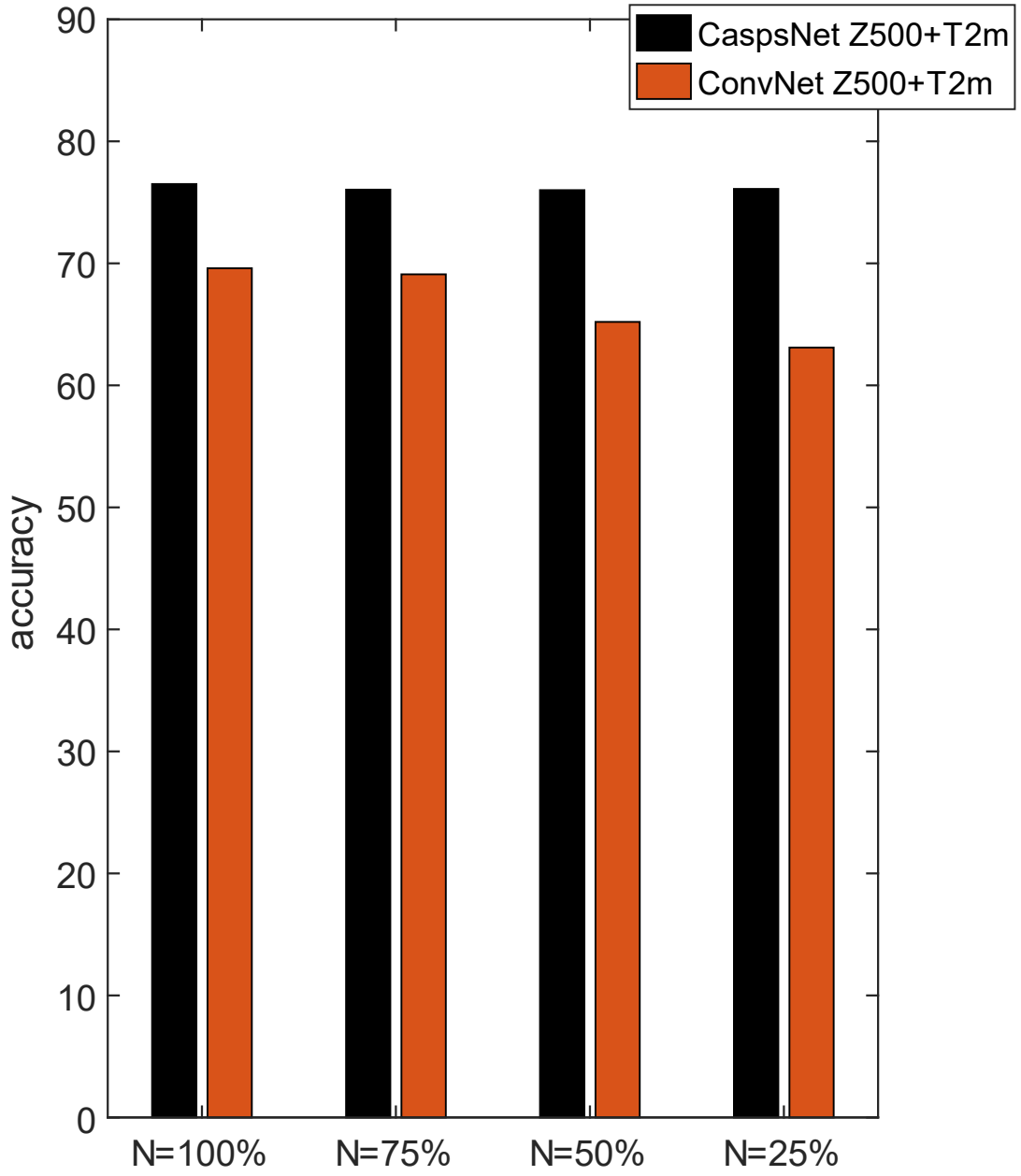
Results: Accuracy





Poor Data Regime: day 3 accuracy





Data-driven extreme weather prediction

-- Precursor?

-- Can we look at past patterns and search of relevant analogs?

-- Impact-based auto-labeling: expert labeling is expensive, difficult, and prone to biases

-- Not predict evolution of Z500, predict the impact directly

-- Re-analysis is in a data poor regime: how robust can deep learning be on low training samples?

Motivation

Day-to-day Data-drive Forecasting

- Numerical models (NWP)
- Multi-ensemble runs
- Data-assimilation

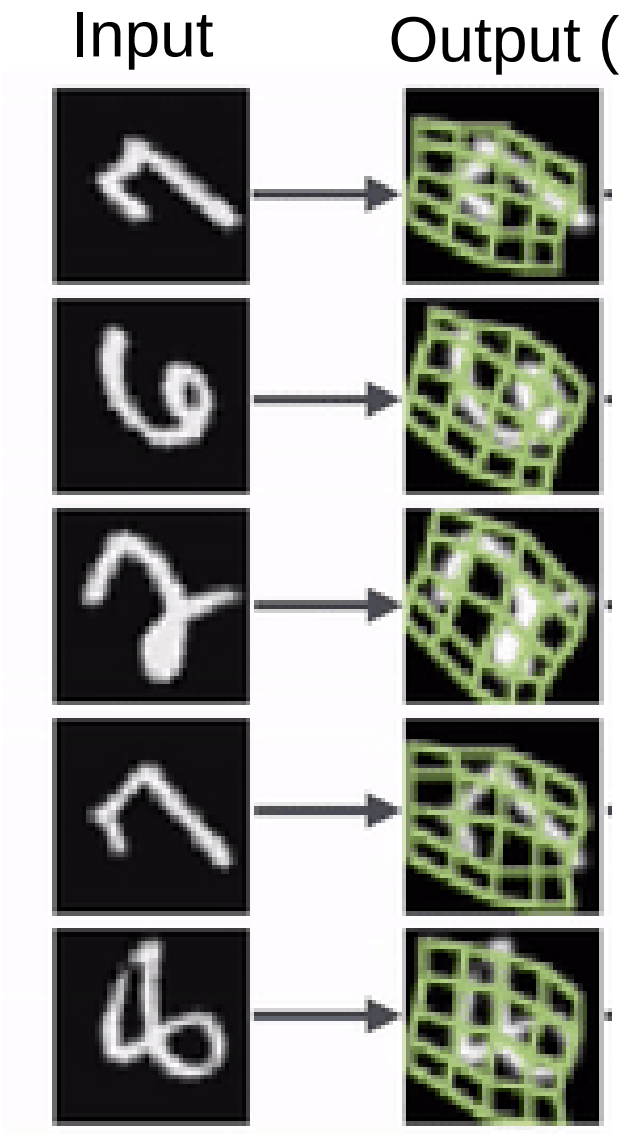
Weyn et al., JAMES 2019;
Scherr, GRL 2018;
Scherr et al., GMD 2019;
Chattopadhyay et al., NPG 2020

Extreme Weather Prediction

- Might be difficult for long lead times
- Precursors

Chattopadhyay et al., JAMES 2020

Capturing Rotational Features

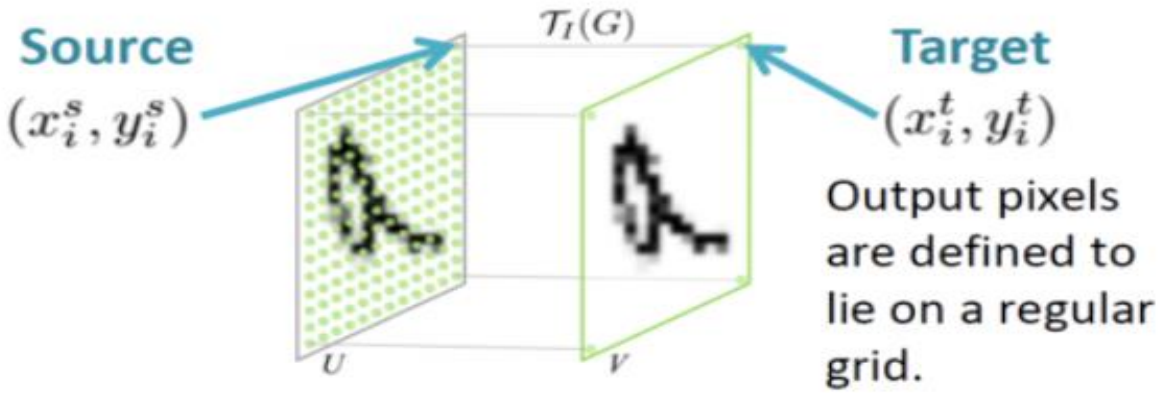


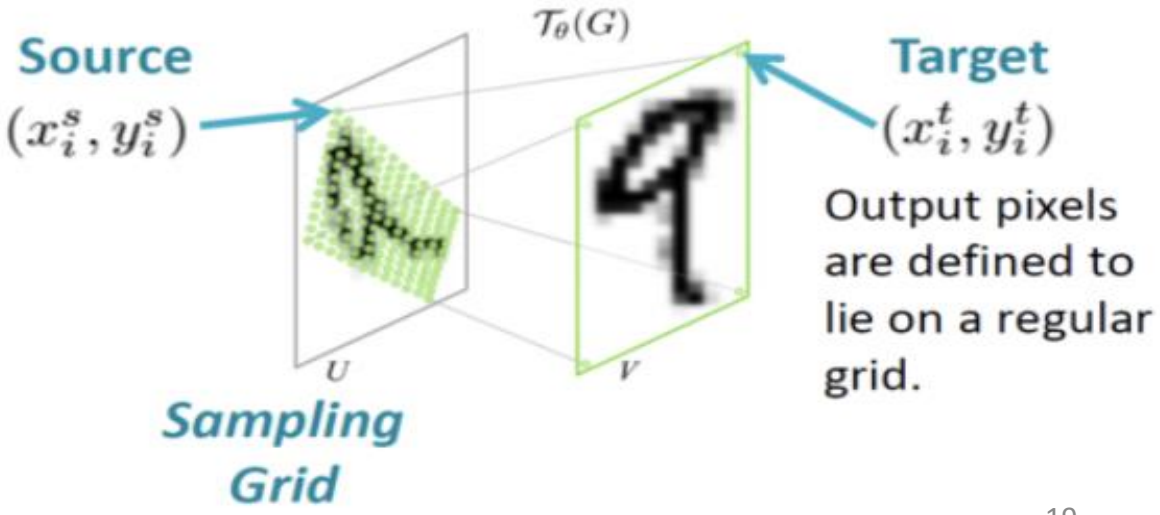
- Pooling layers are effective for dimension reduction but cannot capture large distortions in feature space (e.g., rotational variations)
- Need to capture rotational variances for effective spatio-temporal modeling involving turbulent vortices
- Can we incorporate differentiable modules that can capture rotational transformations inside a Conv encoder?

Affine Transformation and Image Sampling

$$\begin{bmatrix} x^s \\ y^s \end{bmatrix} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{bmatrix} \begin{bmatrix} x^t \\ y^t \\ 1 \end{bmatrix}$$

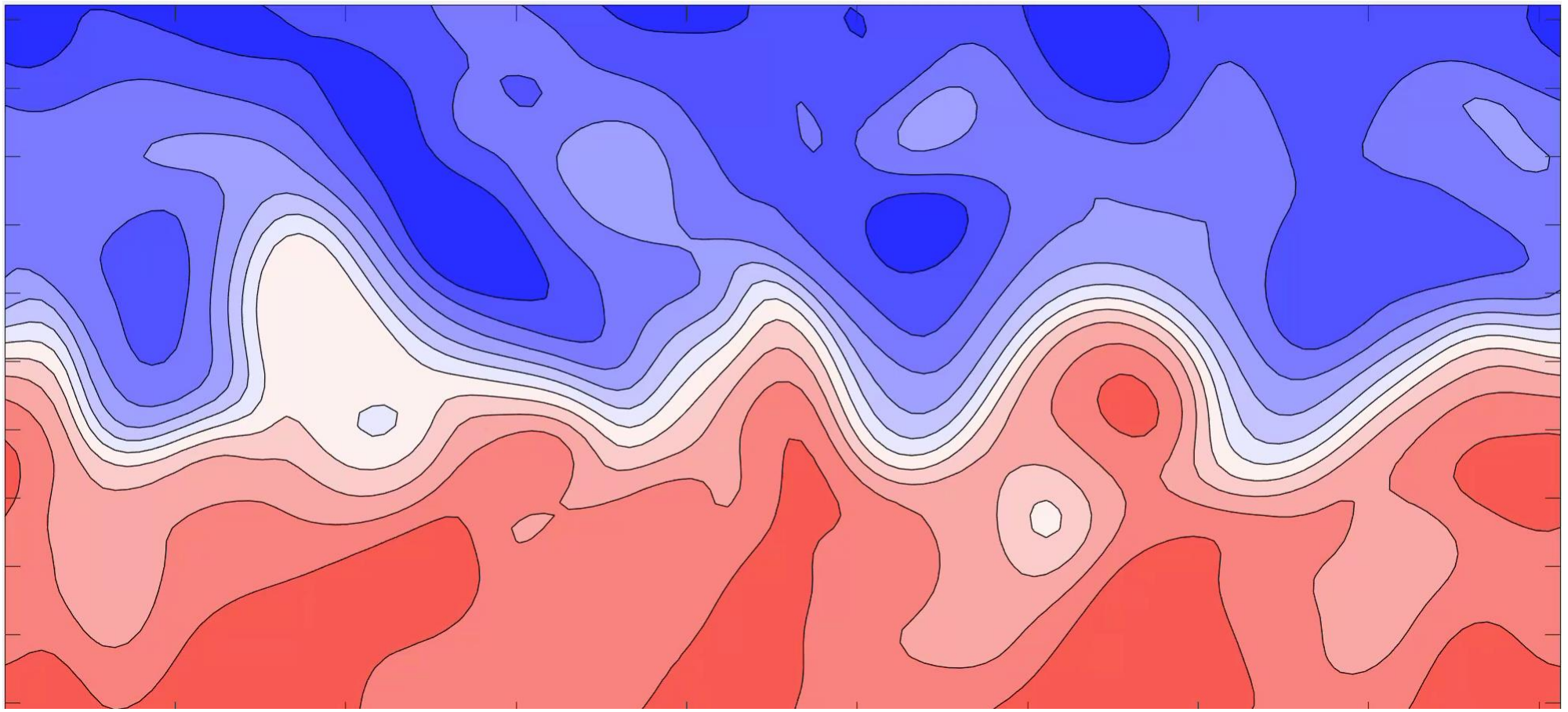
$$V_i = \sum_m^H \sum_n^W U_{n,m} k(x_i^s - m; \Phi_x) k(y_i^s - n; \Phi_y)$$





Test Bed

2-Layer Quasi Geostrophic Equations (QG)



Lutsko et al., JAS 2015; Nabizadeh et al., GRL 2019,

Problem Statement

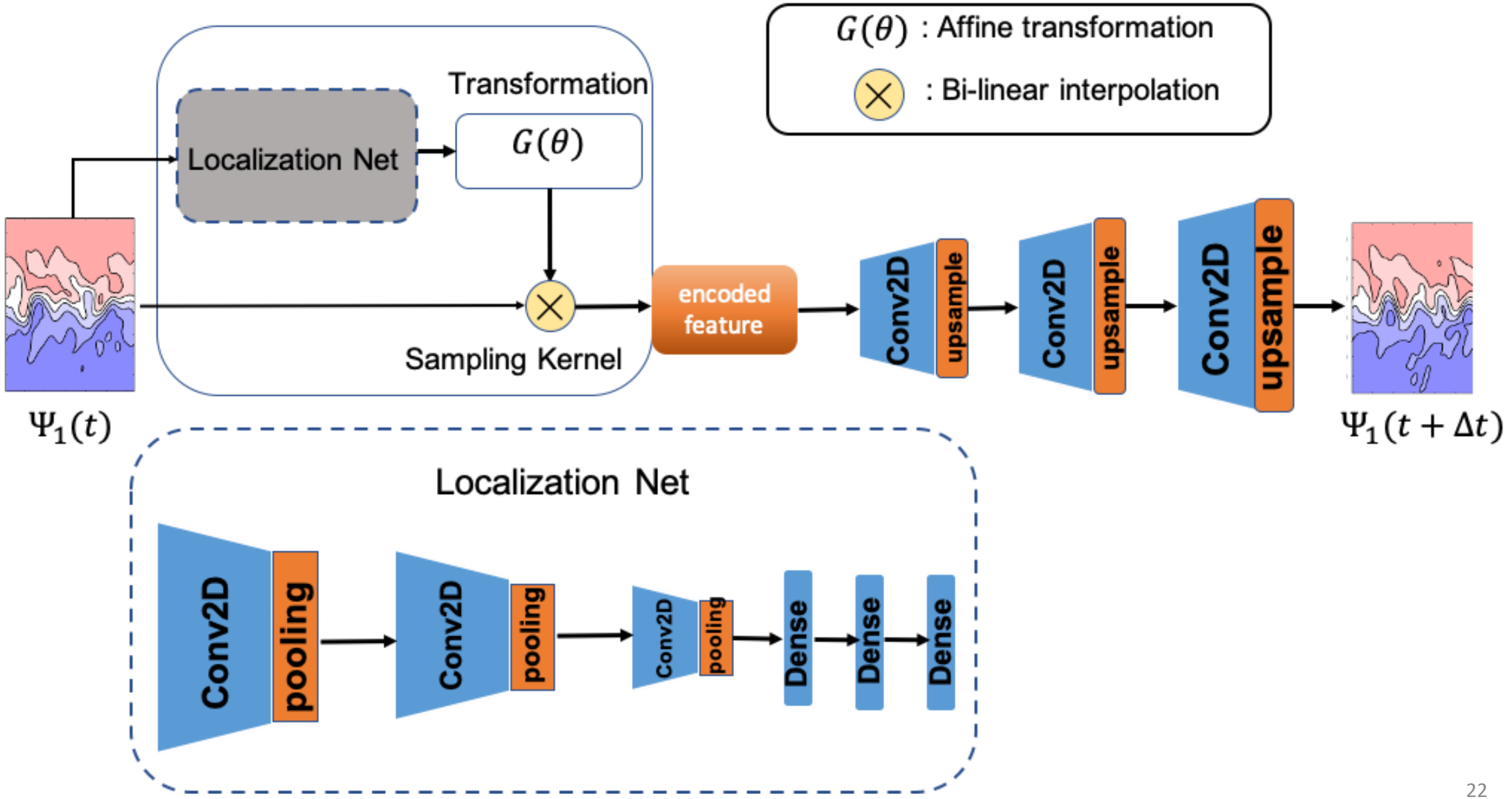
$$\Psi = \begin{pmatrix} \Psi_1 \\ \Psi_2 \end{pmatrix}$$

*Given $\Psi_1(t)$, **alone**, for training, can we predict the short-term evolution and the long-term statistics of Ψ_1*

- *Capture rotational features of coherent structures across all scales*

- *Ensure a non-drifting jet-stream at the center*

Deep Learning Architecture



Loss functions

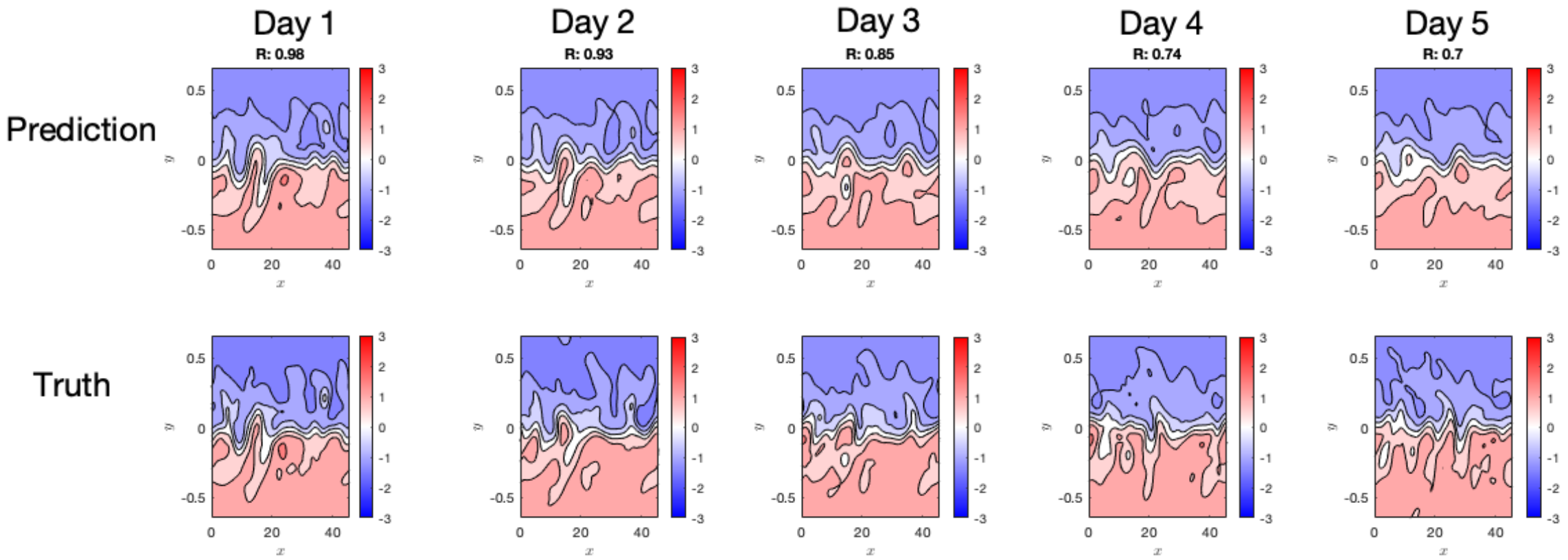
$$\mathcal{L}_1 = \frac{(1-\gamma)}{N}\left(\sum_N \Psi_1^{pred} - \Psi_1^{true}\right)^2 + \frac{\gamma}{N}\sum_N \Psi_1^{pred}$$

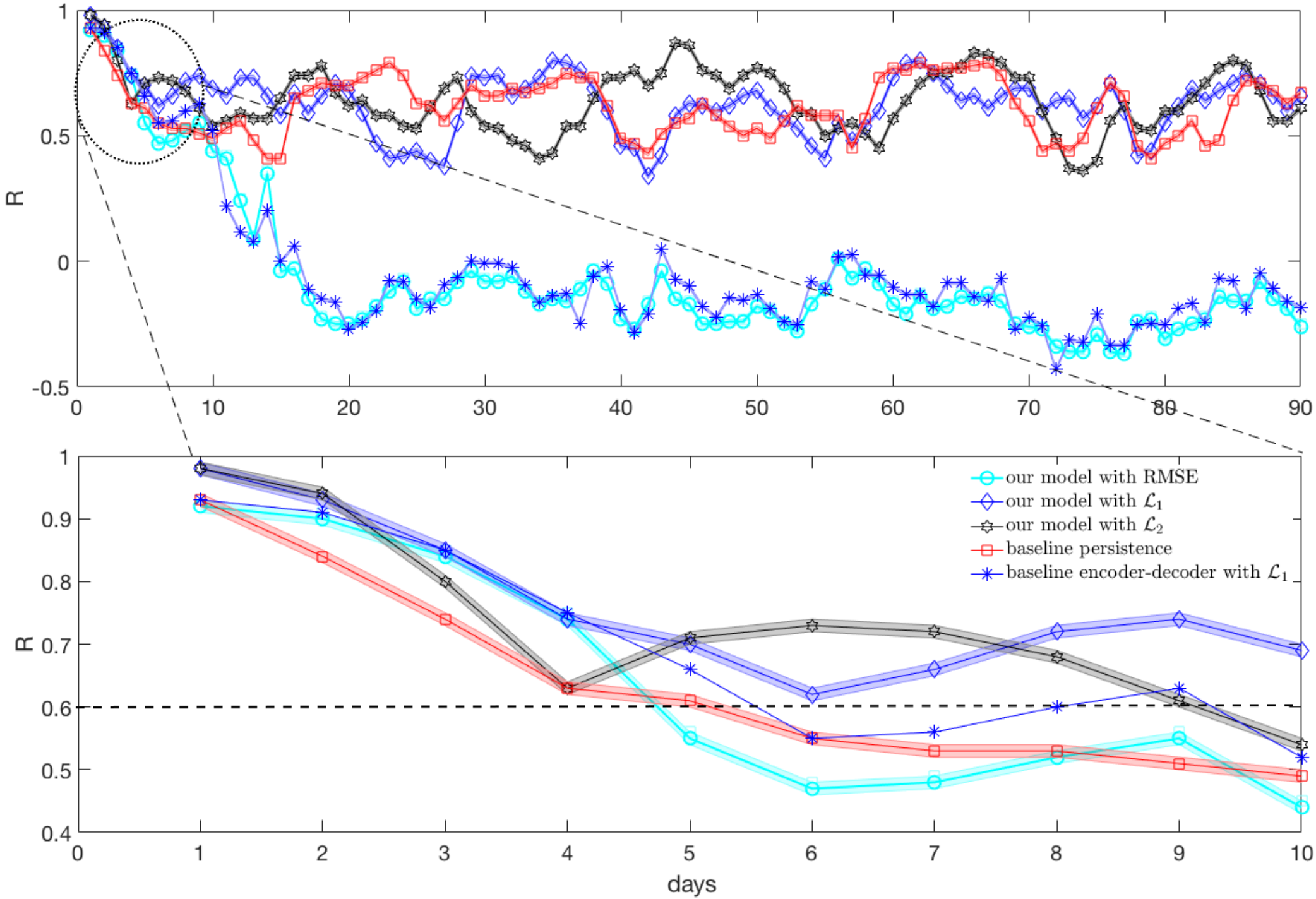
Forces the long-term mean of jet-stream to remain in the center

$$\mathcal{L}_2 = \frac{(1-\gamma)}{N}\sum_N\left(\Psi_1^{pred} - \Psi_1^{true}\right)^2 + \frac{\gamma}{N}\sum_N\left(\nabla_{xy}\Psi_1^{pred} - \nabla_{xy}\Psi_1^{true}\right)^2$$

Captures the boundaries of PV more accurately

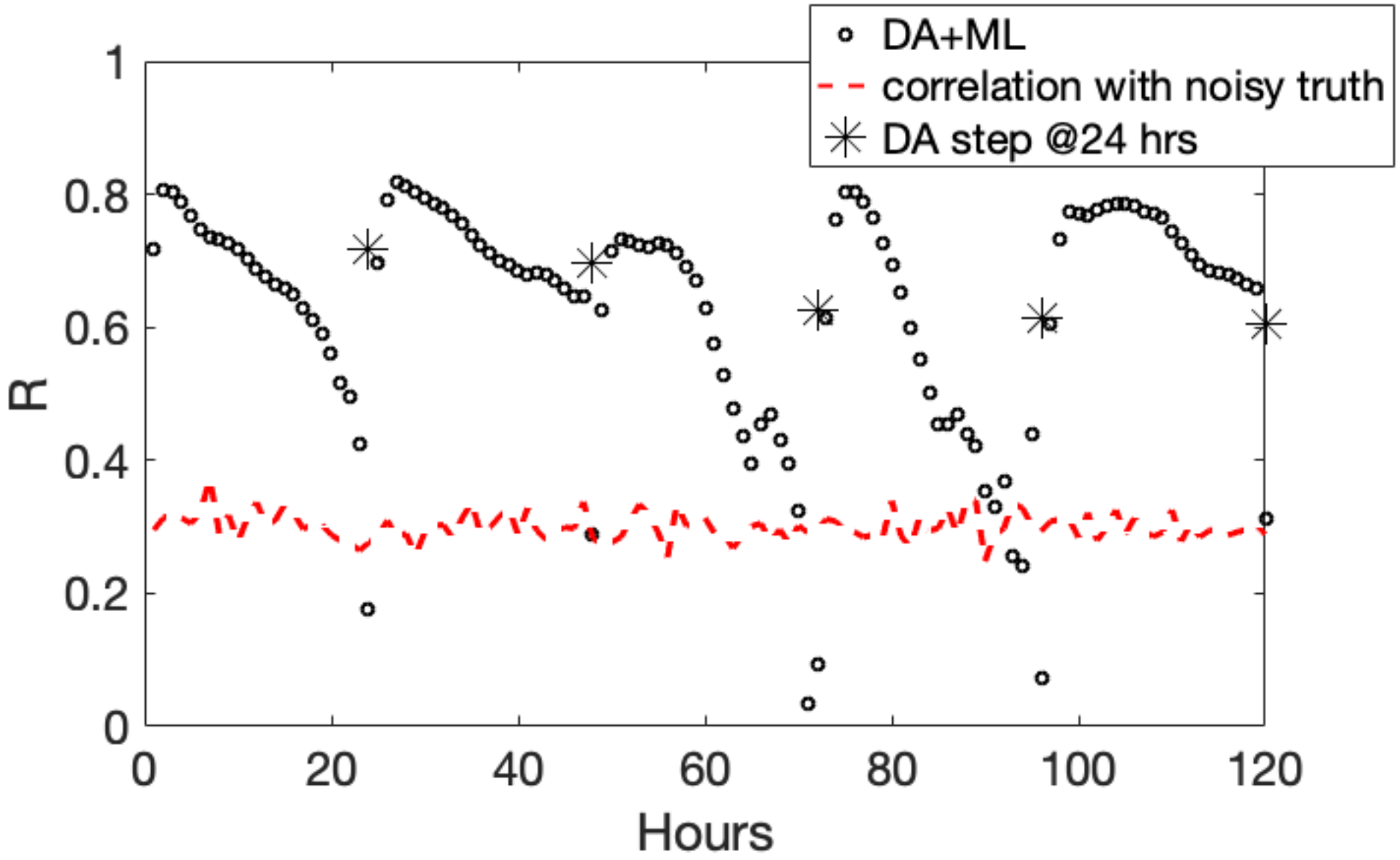
Results


Day 1
R: 0.98
Day 2
R: 0.93
Day 3
R: 0.85
Day 4
R: 0.74
Day 5
R: 0.7
Prediction
Truth
x
y


R
days
our model with RMSE
our model with $\mathcal{L}_1$
our model with $\mathcal{L}_2$
baseline persistence
baseline encoder-decoder with $\mathcal{L}_1$

Looking Ahead…

Weather Forecasting ? DA + ML



Conclusion

- Data-driven forecasting can enable multi-ensemble runs, integrated with DA can maintain trajectory
- Preserving Equivariance, symmetry, etc., can marginally improve prediction
- Extreme weather is challenging; data-driven models can augment NWP efforts, provide early warnings
- Long lead time predictions, precursors, etc.