COMBINING MACHINE LEARNING WITH KNOWLEDGE-BASED MODELING FOR SCALABLE FORECASTING OF VERY LARGE, COMPLEX, SPATIOTEMPORALLY CHAOTIC PROCESSES

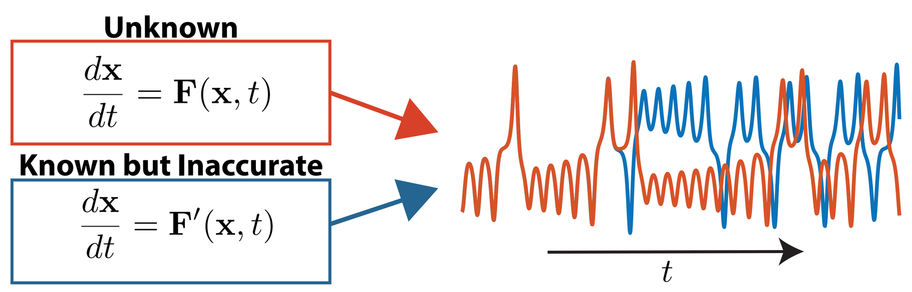
ALEXANDER WIKNER[1], JAIDEEP PATHAK[1], TROY ARCOMANO[2], ISTVAN SZUNYOGH[2], BRIAN HUNT[1], MICHELLE GIRVAN[1], EDWARD OTT[1]

[1]UNIVERSITY OF MARYLAND, COLLEGE PARK, [2]TEXAS A&M UNIVERSITY

OVERVIEW

GOAL: Improve forecasting of very large, complex spatiotemporally chaotic systems.

SETTING: A knowledge-based predictor is available, but its utility is limited by error in its forumation. In addition, we have access to a limited time series of measured system states.



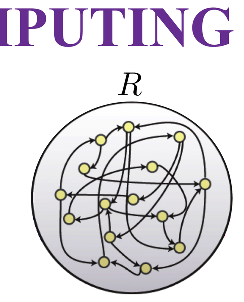
APPROACH: Use machine learning (reservoir computing[1,2,3]) and available training data in combination with the imperfect knowledge-based model to enhance prediction.

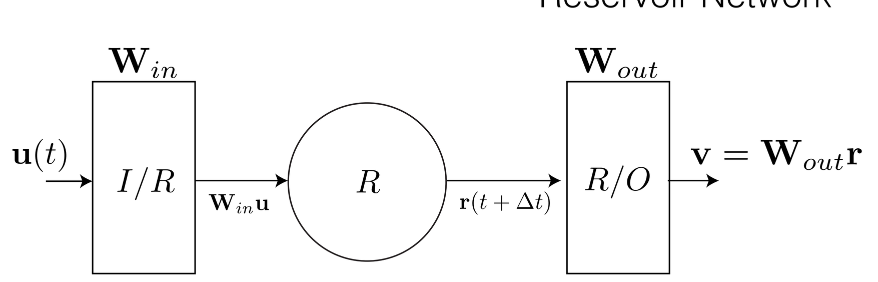
CHALLENGE: Feasibility and scalability of the machine learning to very large systems of interest.

APPLICATIONS: Forecasting the weather, ocean conditions, conditions in the solar wind, the magnetosphere & ionosphere, forest fire evolution, ecosystem response to climate variation, and neural activity.

OUR IMPLEMENTATION OF RESERVOIR COMPUTING

- Sparsely connected directed random network with D nodes.
- Adjacency matrix: $\mathbf{A}$.
- State of the i^{th} node: $r_i(t)$.
- State of the reservoir: $\mathbf{r}(t)$.





Train Data: Time series measurements of system state $\mathbf{u}(t)$ for $-T \le t \le 0$.

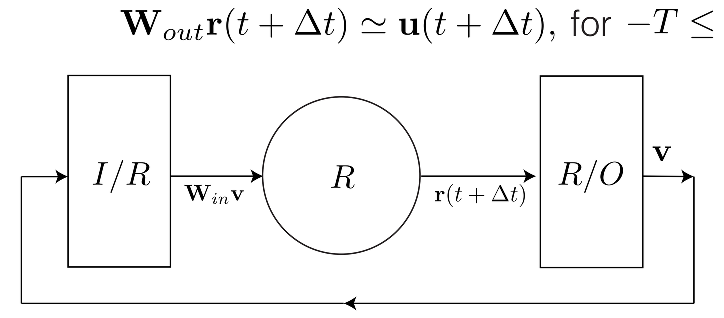
Training: (1) Input $\mathbf{u}(t)$ to reservoir I/O system,

$$\mathbf{r}(t+\Delta t) = \tanh[\mathbf{A}\mathbf{r}(t) + \mathbf{W}_{in}\mathbf{u}(t)].$$

(2) Record and store $[\mathbf{u}(t), \mathbf{r}(t)]$ for $-T \le t \le 0$

(3) Adjust the matrix $\mathbf{W}_{out}$ to achieve:

$$\mathbf{W}_{out}\mathbf{r}(t+\Delta t) \simeq \mathbf{u}(t+\Delta t), \text{ for } -T \le t \le 0$$

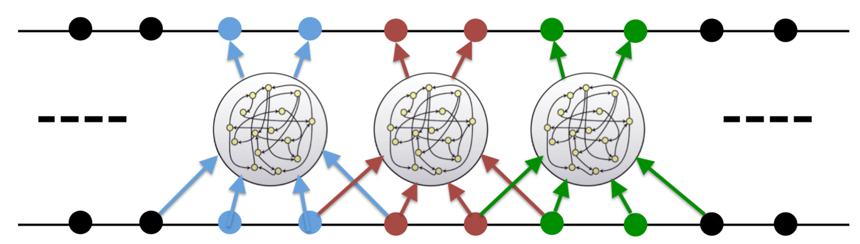


Prediction: A feedback connection is added so that the reservoir runs autonomously for $t > 0$.

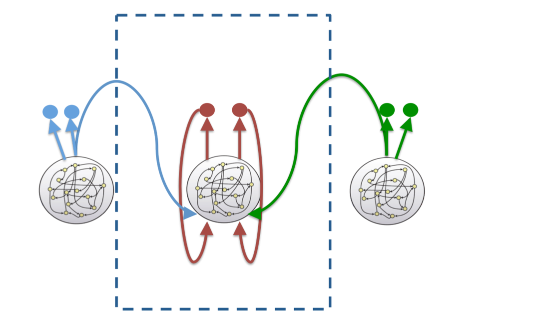
$$\mathbf{r}(t+\Delta t) = \tanh[\mathbf{A}\mathbf{r}(t) + \mathbf{W}_{in}\mathbf{W}_{out}\mathbf{r}(t)]$$

The prediction is $\mathbf{W}_{out}\mathbf{r}(t) = \mathbf{v}(t)$

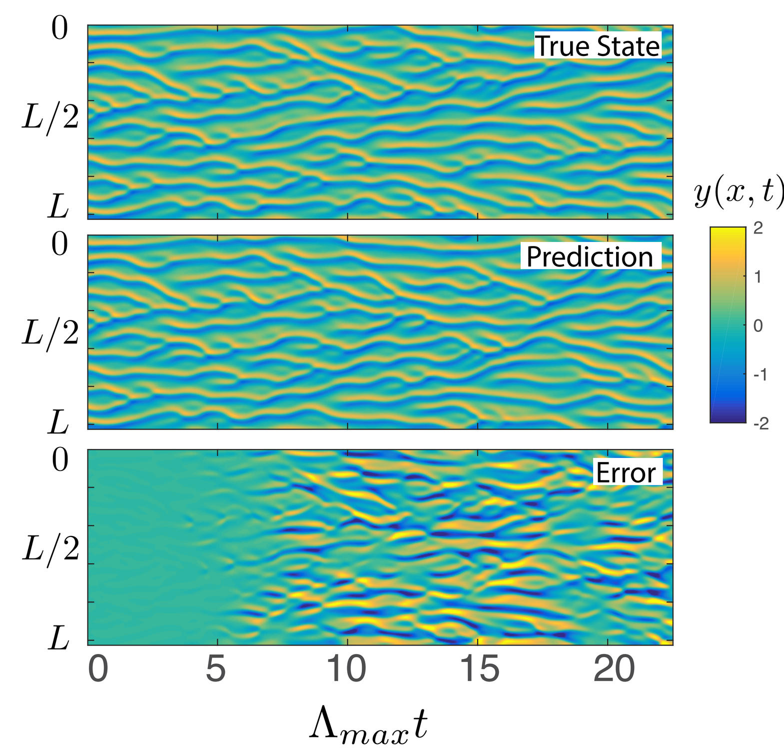
MACHINE LEARNING SCALABILITY (without a knowledge-based model): FORECASTING HIGH-DIMENSIONAL CHAOS



Parallelized Training



Parallelized Prediction with Feedback



Putative unknown test system: the Kuramoto-Sivashinky equation,

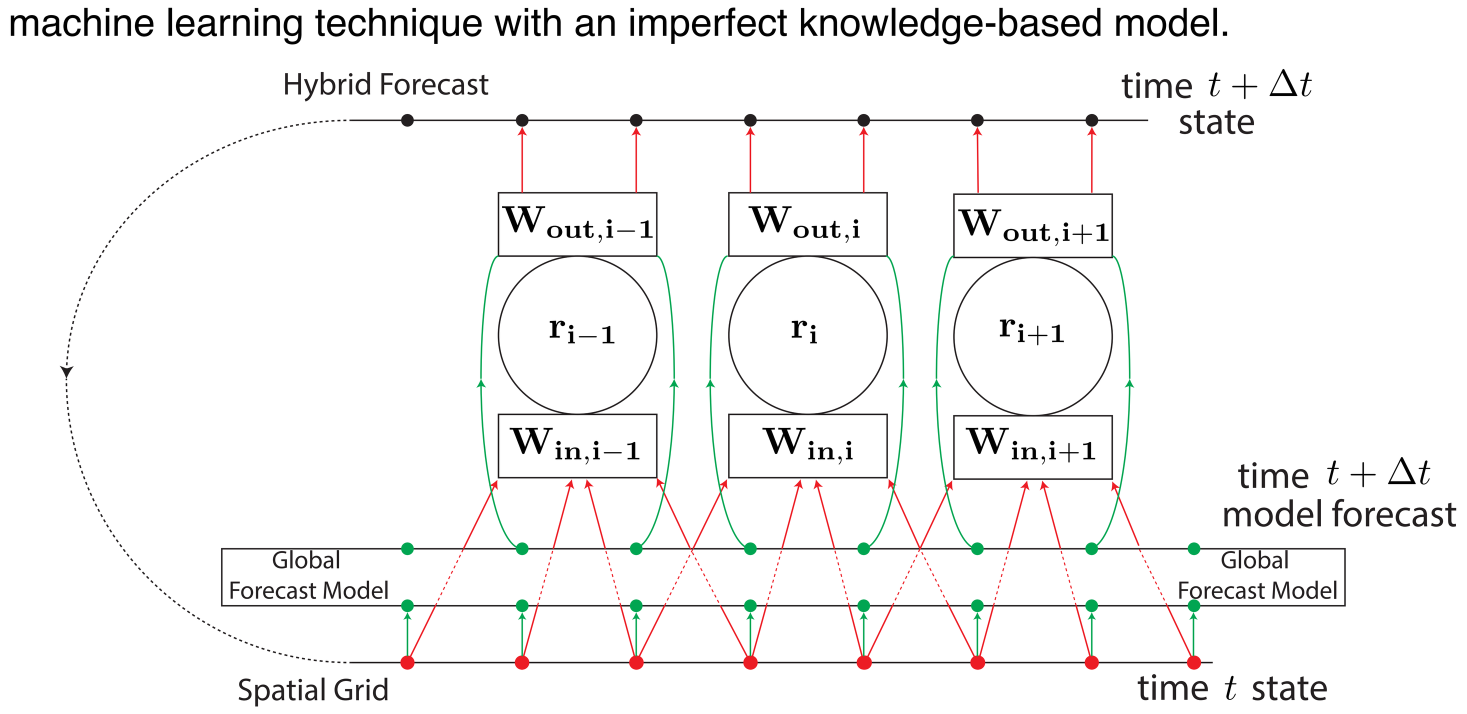
$$\frac{\partial y}{\partial t} = -y\frac{\partial y}{\partial x} - \frac{\partial^2 y}{\partial x^2} - \frac{\partial^4 y}{\partial x^4}$$
$$x \in [0, L)$$
$$y(x+L, t) = y(x, t)$$

with L = 100, D_{KY} = 23.
This example uses 32 reservoirs in parallel.

<u>The parallelized reservoir prediction scheme can scale to arbitrarily high-dimensional chaotic systems.</u>

SCALABLE HYBRID MACHINE LEARNING MODEL

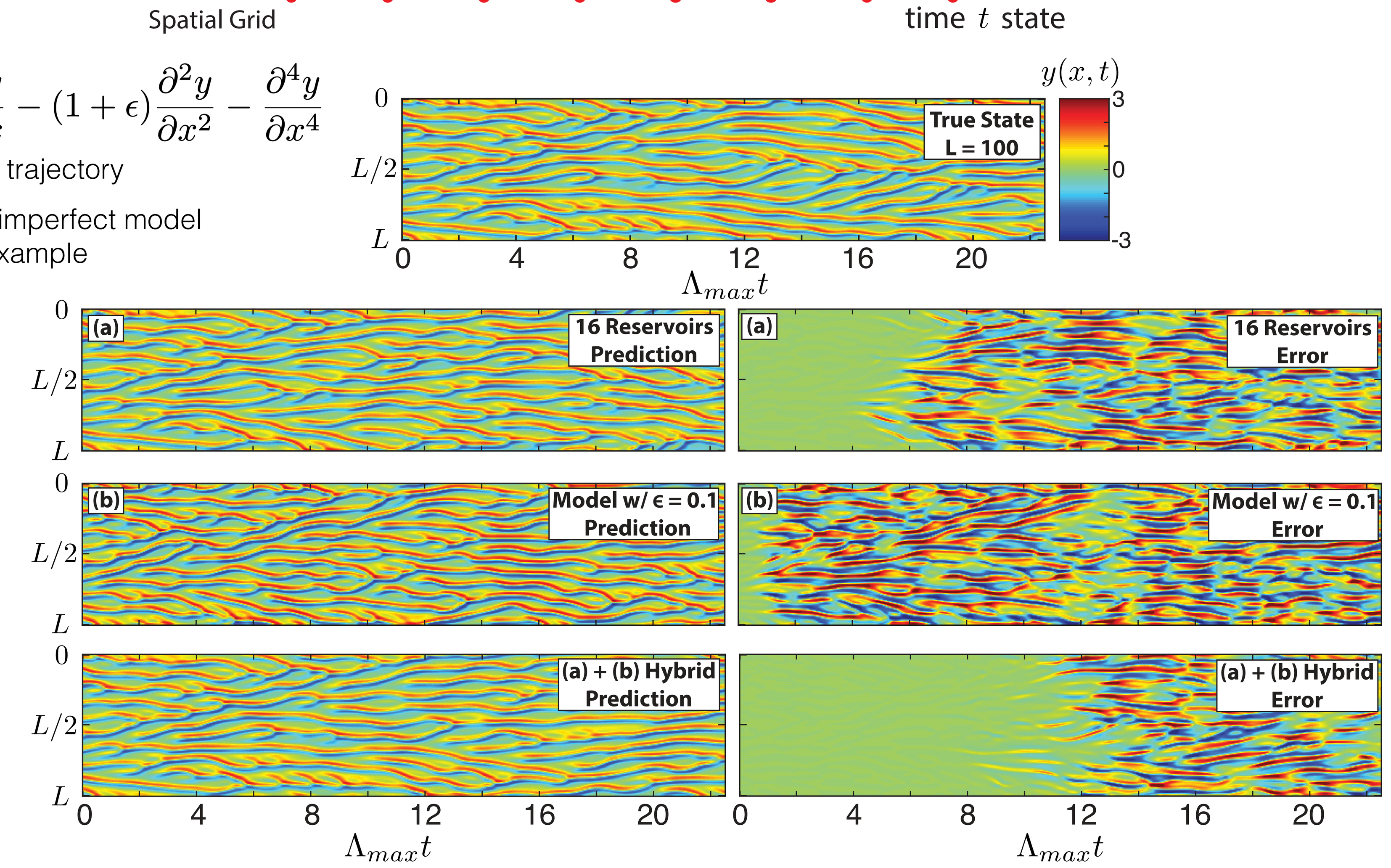
For scalable forecasting of very high dimensional chaotic dynamicals systems, we combine the parallelized machine learning technique with an imperfect knowledge-based model.

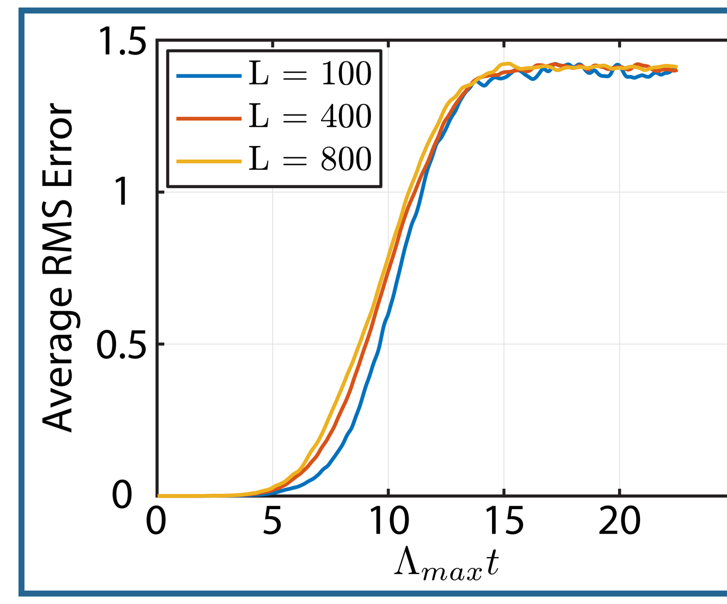


$$\frac{\partial y}{\partial t} = -y\frac{\partial y}{\partial x} - (1+\epsilon)\frac{\partial^2 y}{\partial x^2} - \frac{\partial^4 y}{\partial x^4}$$

$\epsilon = 0$ For true trajectory

$\epsilon = 0.1$ For the imperfect model in our example





DEMONSTRATION OF SCALABILITY

RMS Error average over 100 predictions of the KS equation vs. L (system size) with [(number of reservoirs)/L] held constant at 0.16 and the reservoir size held constant at 4000 nodes.

The RMS Error remains relatively invariant as L increases, indicating that our hybrid method can scale to large systems.

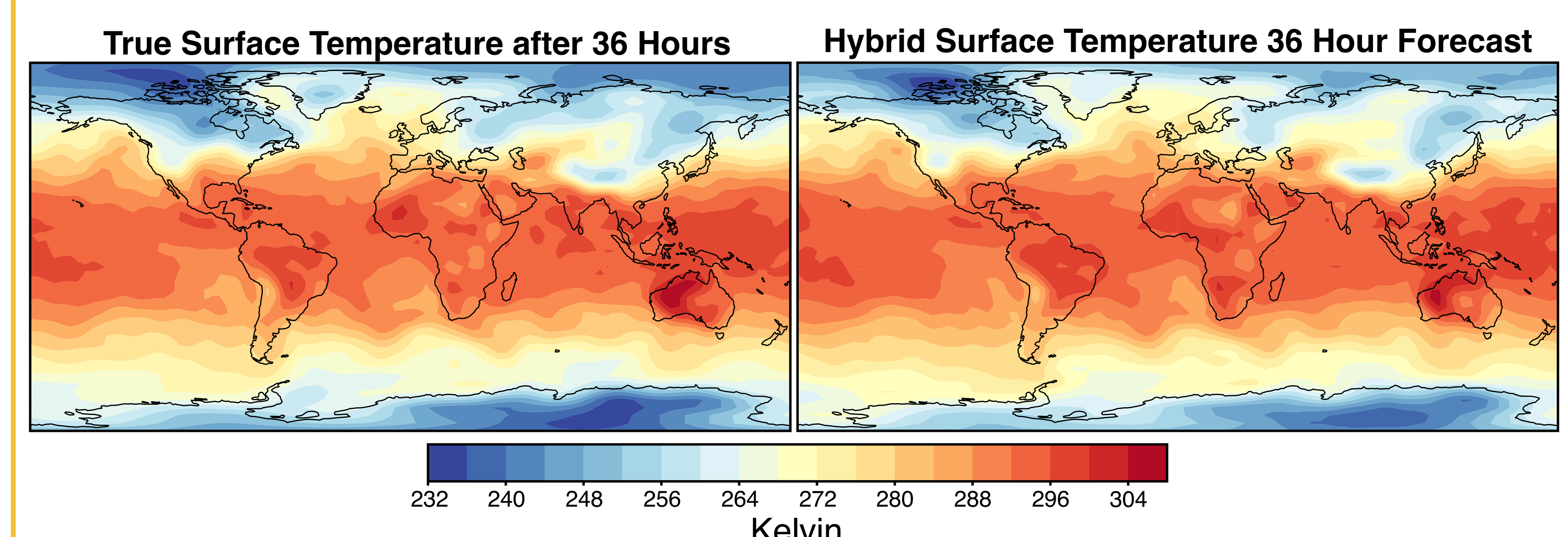
AN EXAMPLE: GLOBAL WEATHER FORECASTING

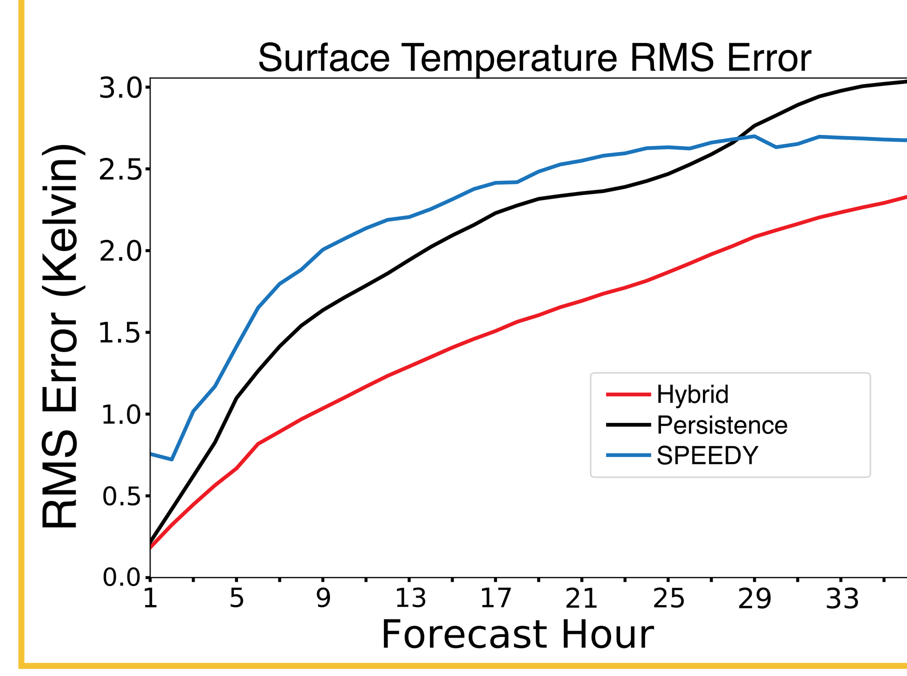
Training Data: 9 years of data[6] interpolated onto a 96 by 48 grid with 8 vertical layers

Model Variables: Temperature, horizontal wind vector, and specific humidity at each grid point, plus earth surface pressure, leading to a total of 152,064 system variables.

Machine Learning: 1,152 reservoirs, each with 5,000 nodes, are each trained to predict system variables in a 2 by 2 grid (512 variables per reservoir).

Knowledge-based Model: SPEEDY[7], a low-resolution atmospheric global model.





We compare the global RMS error between predictions of the surface temperature and the truth. We find that, compared to persistence from the initial condition, SPEEDY has no predictive skill, while the hybrid forecast is stable and has predictive skill for the investigated 36 hours.

REFERENCES

[1] Jaeger et al. **Science (2004)**
[2] Jaeger. **GMD Technical Report (2001)**.
[3] Maass et al. **Neural Computation (2002)**.
[4] Pathak et al. **Physical Review Letters** (2018).
[5] Pathak et al. **Chaos (2018)**.
[6] C3S, **https://cds.climate.copernicus.eu/cdsapp#!/home (2017)**.
[7] Molteni **Clim. Dyn. (2003)**, Kucharski et al. **Clim. Dyn. (2006)**.

ACKNOWLEGEMENTS

We gratefully acknowledge financial support from DARPA under contract number HR00111890044 and travel support from NSF under award DGE-1632976

COLLEGE OF COMPUTER, MATHEMATICAL, & NATURAL SCIENCES

TEXAS A&M UNIVERSITY