OXFORD JOURNAL
OF STUDENT
SCHOLARSHIP

PyroCast-Morpheus: A Mnemonic Counterfactual Digital Twin for Pyrocumulonimbus Wildfire Prediction and Analysis

Avani Patil
Patilvavani@gmail.com

ABSTRACT

Wildfires continue to increase today as a result of climate change, human activity, and natural causes such as lightning. This has also been seen with a rise of Pyrocumulonimbus (PyroCb) clouds, which are a result of updrafts created from fires and wildfires. This paper describes and analyzes a digital twin model that predicts, simulates, and provides current state, memory retrieval, forecasting, and risk analysis for future scenarios. We trained and validated a digital twin model which processed 10 events across North America as 227 rows of 6-hour interval data. To provide more context for the analysis, the data is composed of ERA5 single level variables, pressure level variables, GOES satellite imagery, vegetation data and terrain data per event. This study is an expansion on current PyroCb prediction models and datasets in order to create a well-rounded pipeline of a digital twin model. Results showed the 6-hour EnKF synchronization reduced state RMSE by an average of 45%. Additionally, the fused memory twin was found to retrieve similar past events to the current wildfire state and had a higher prediction accuracy to the true observed path of wildfire. This digital twin can later be applied to a larger and worldwide dataset and can potentially enable teams to work on ground level operations more efficiently, plan containment, and rescue operations.

INTRODUCTION

Wildfires are unpredictable fires which can cause significant damage to lives and the surrounding environment. Scientists studied wildfire data for 21 years and saw that extreme wildfires are more frequent, intense, and larger. The biggest change in behavior was located in the temperate conifer forests of western U.S and the boreal forests of northern North America and Russia [15]. As climate change increases, warmer temperatures have contributed to higher fire activity. Therefore, wildfire modeling is extremely important because it helps firefighters, forest managers, and government leaders make key decisions surrounding wildfire management and prevention [4].

With wildfires becoming more common globally, Pyrocumulonimbus (PyroCb) clouds formation and observation have increased. PyroCb clouds begin as intense heat is released from fires and the surface of the Earth, creating hot air and smoke. The plume starts to grow as cool air mixes. When the plume

reaches high enough, a cloud forms with ice particles that condense on the smoke particles from the plume [14]. These thunderstorms can result in strong downbursts or dry storms that can both contribute to growing a fire and igniting new ones in different locations. Hence, these clouds can contribute to a higher wildfire risk, putting more lives and ecosystems in danger [9]. Due to PyroCb clouds contributing further to the uncontrollable nature of wildfires, a digital twin model can provide researchers with more useful predictions, conclusions, and knowledge. Past studies have found that PyroCb formation occurs because of a combination of atmospheric variables, wildfire and environmental conditions. Since wildfires are not caused by one variable and rapidly change, this strengthens the need for a twin model that gathers data from multiple sources [11].

Current wildfire research includes machine learning, simulation, and data assimilation in order to improve prediction and forecasting. Predicting pyroCb wildfires is a challenging task because pyroCb formation is due to a lot of variables that work together. A study by Krell et al created a dataset with atmospheric variables and fire conditions and tested whether multiple machine learning models could distinguish non pyroCb fires from pyroCb fires. More importantly, the study explored how predication changed based on different combinations of features and found that machine learning models have the potential to forecast pyroCb [7].

Wildfire spread models are important to understand and visualize physical conditions of how a wildfire will move, grow, or behave based on certain variables. A study in 2008 demonstrates the importance of data assimilation through the usage of ensemble Kalman filters (EnKF) in order to improve wildfire models and simulation. It developed a physics-based model and applied an EnKF to test whether it could combine temperature measurements with the running model. The study evaluated this through measurements and simulations and found that EnKF could correct fire state estimates [8]. Therefore, it showed data assimilation, specifically EnKF, can potentially be valuable in better model accuracy. Similarly, a study by Rocheux et al also wanted to test how effective EnKF was to improve wildfire spread prediction, except they used fire front observations to do so. They sequentially updated the fireline markers in order to have a more accurate and reliable fire front location to the actual conditions, and tested a controlled experiment, finding the simulations with EnKF could correct inaccurate fire location predictions [10].

A digital twin is useful for real world application as they are known to be highly accurate, precise, fast, and flexible. A digital model can use event observation to predict accurate outcomes while also providing an intricate simulation [3]. A study by Zhong et al also tried to see if a digital twin model could improve the efficiency of a current physics wildfire model while also testing if data assimilation could decrease prediction errors. After combining deep learning and assimilation, they found that the digital twin had the same efficiency, however, with increased speed and less prediction error [16]. This is essential in wildfire management because fires can change intensity, location, and unpredictability at any given time [3].
This paper expands on combining data assimilation methods with digital twins, focusing on pyroCb wildfire events in order to improve wildfire modeling and forecasting. As a result, this study presents PyroCast-Morpheus, a Mnemonic Counterfactual Twin, which combines a physics model, four memory

components, prediction engine with synchronization using EnKF, counterfactual simulation, and decision intelligence.

METHODOLOGY

PyroCb Events and Selection

The cohort of ten events are taken from the PyroCast database and are all located in North America. Table 1 provides detailed descriptions of each event [13].

ID	Dates	Location	Hrs	Regime
179	May 27–Jun 2, 2021	33.29°N, 108.59°W	138 h	High-elevation subtropical continental
180	Jun 12–17, 2021	33.20°N, 111.08°W	132 h	Arid continental
181	Jun 16–21, 2021	37.71°N, 113.79°W	132 h	Plateau
189	Jun 28–Jul 3, 2021	41.47°N, 122.33°W	132 h	Cascade
190	Jun 28–Jul 3, 2021	57.52°N, 123.00°W	132 h	Boreal cordillera
202	Jul 8–13, 2021	50.80°N, 95.03°W	138 h	Boreal lowland
216	Jul 17–21, 2021	64.12°N, 133.02°W	108 h	Subarctic
253	Mar 29–Apr 3, 2022	25.60°N, 80.43°W	138 h	Subtropical wetland
258	Jun 9–14, 2022	63.79°N, 153.40°W	132 h	Boreal
260	Jun 8–14, 2022	63.33°N, 155.60°W	138 h	Boreal

Table 1: *PyroCb event cohort (North America)*

Events were selected from the PyroCast database, which consists of a majority of North American and Australian events spanning from 2018-2023. Therefore, the cohort in this paper was chosen from the more recent available data. The events were also chosen such that their duration was less than two weeks in order to keep events duration variability to a minimum. Additionally, the GOES ABI did not cover Australia, which is why the cohort is focused on North American events.

Data Sources and Preprocessing

Downloading data sources and extraction was done using Python. Each extracted data source was merged into a final csv. The table below describes various sources used to collect data from. Data sources 1-3 from table two were sampled by 6 hour intervals, ranging from start time to end time of the event [6, 5, 2]. The terrain sources are spatial data sources based on event locations [1, 12]. In order to create one dataset, the geospatial data was duplicated and reused based on event id. The final dataset contains 6 hour intervals with 10 events and has 227 row entries. Data cleansing or missing value imputation was done to ensure that the data is ready for training. Cleansing was done by finding the median for a feature and then replacing it with the median where there was no value.

#	Source Name	Type of data	Reason for inclusion
1	ERA5 Single level	Surface meteorological variables, and available vegetation variables	Describes near surface weather conditions
2	ERA5 Pressure level	Atmospheric variables at different pressure levels	Provides details about atmospheric conditions
3	GOES 16 ABI	Satellite observations, bands	Uses visuals to extract information about fire and cloud conditions
4.	ALOS World 3D - 30m	Terrain	Gives land-surface details, used for events in Canada and Alaska specifically
5.	SRTM Digital Elevation Data Version 4	Terrain	Gives land-surface details, however data does not include parts of Canada and Alaska

Table 2: *Data sources information*

Digital Twin Architecture

Figure 1 illustrates the architecture of the digital twin model and describes the workflow. First, the twin uses the multisource dataset of 227 entries and observes the current state. These inputs are then checked for alignment and utilized to derive more features such as ventilation from BLH and near-surface windspeed. The model predicts its state and compares it to its next forecast with EnKF synchronizing the wildfire state. The innovation r is the difference between observation and the forecast provided by the digital twin. The divergence pressure, D, measures the error between model prediction and observation relative to its uncertainty. The purple box in the figure represents the episodic memory used to retrieve past historically similar events to help provide more context to the current forecasted wildfire. The counterfactual component of the twin compares a baseline scenario to 8 different what-if scenarios created by the model. The decision intelligence then ranks how risky each scenario is. The gray horizontal arrows represent the information flow and the green dashed arrows are the feedback/tasking.

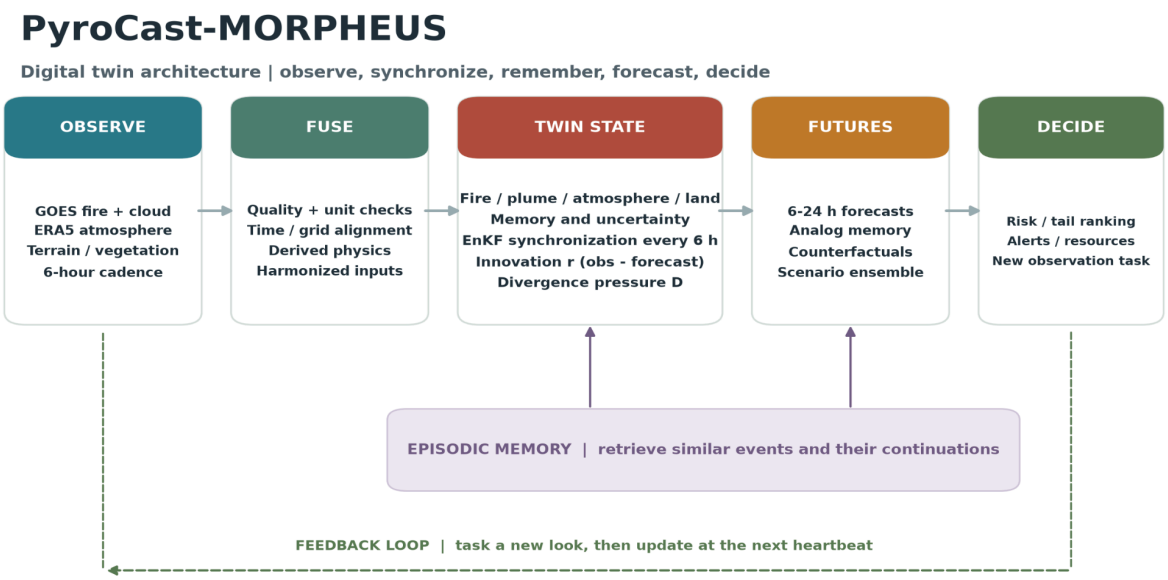


Figure 1: *PyroCast-MORPHEUS: Mnemonic, Observationally-coupled, Recursive, Physics-informed, Episodic, Uncertainty-calibrated, Self-learning Twin.*

Digital Twin Memory Components

Figure 2 below shows how the twin stores and accesses knowledge to gain more context of the wildfire. The key components include a consolidation engine, retrieval pathway, and write pathway. The sensory store is the current 6-hour observation data. The short term working memory includes the last 8 cycles of 6-hours of data. The episodic store can fetch specific event data from the past. The semantic store has memory on general regime archetypes, coupling priors, knowledge-graph edges. It helps the model understand what types of worlds exist and how they work. The consolidation engine works offline and organizes past events that seem more important. From there, the retrieval pathway gathers specific event data as needed for the current event context. The write pathway makes sure to update and store new changes and patterns in the model's memory.

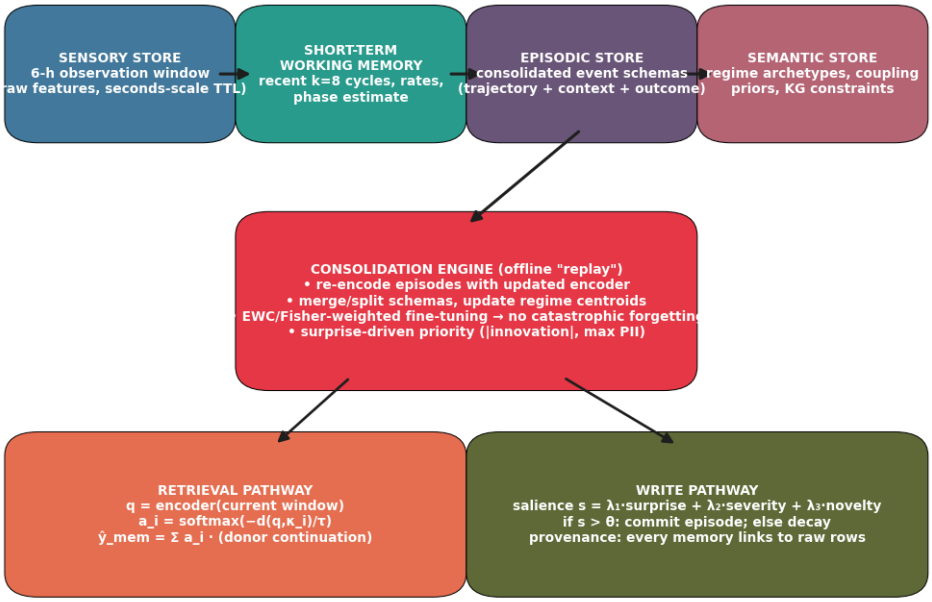


Figure 2: *Memory component of the twin*

Digital Twin Training

For all components of the twin, training is done using the leave-one-event-out (LOEO) method. Since this digital twin has 10 events, the model gets trained on 9 events at one time and tested on the last one. This gets repeated for each combination. Ridge and XGBoost are compared against the baseline persistence only model. Persistence only assumes that there is no change in variable values after a time step of 6 hours. The evaluation of these models uses root mean square error (RMSE) and is further discussed in the results section. Ridge is a linear estimator, so its inputs are standardized in the training. On the other hand, XGBoost is a nonlinear estimator, so it uses original feature values.

Key Terms

This digital twin has a few key terms listed below with definitions. Some key terms are derived variables, proxies that are utilized to explain the context of the wildfire state. The table below defines the term name, calculation, and interpretation. These values are calculated each 6-hour time step.

Variable/Component	Calculation	Interpretation
Fire Proxy	GOES ABI Band 7 (3.9 μm) radiance − Band 14 (11.2 μm) radiance	Represents fire activity at each timestep. A larger value means larger difference in fire channel and thermal channel.

Cloud-height Proxy	GOES ABI Band 14 (11.2 μm) radiance - Band 16 (13.3 μm) radiance	Represents a proxy for vertical cloud growth related to the fire.
Raw cloud BT	GOES ABI Band 14 (11.2 μm)	Raw thermal signal from cloud
Divergence Pressure	$D(t) = r^T S^{-1} r$	Represents how significantly the model prediction differs from the observation
Twin State	$x(t) = [x_t^F \mid x_t^P \mid x_t^A \mid x_t^L \mid x_t^M \mid x_t^U \mid x_t^T]$	Describes the wildfire context at a particular time step. The digital twin has six state components: Fire, plume, atmosphere, land, memory, and uncertainty.

Table 3: *General key terms*

EnKF Framework

A simple flow can be described by expanding upon the previous state, making a forecast, creating a probability distribution of the forecast, applying any corrections, and finally updating the current state. Additionally, table 4 provides more information on parameters used with their dimensions.
The EnKF system finds its first forecast:

$$x_t^{f,(i)} = Ax_{t-1}^{a,(i)} + b + \eta^{(i)}$$

where, A is the learned coupling matrix, which compares current and next time step observations, b is the intercept, and $\eta^{(i)} \sim N(0, Q)$ accounts for the model uncertainty.

The ensemble mean, $\bar{x}_t^f$, and the covariance, P_f, are found using the equations below.

$$\bar{x}_t^f = \frac{1}{N} \sum_{i=1}^{N} x_t^{f,(i)}$$

$$P_f = \frac{1}{N-1} \sum_{i=1}^{N} (x_t^{f,(i)} - \bar{x}_t^f)\left(x_t^{f,(i)} - \bar{x}_t^f\right)^T$$

The observation equation is listed below :

$$y_t = Hx_t + \epsilon,\ \epsilon \sim N(0, R)$$

where, H is the observation operator and $\epsilon^{(i)} \sim N(0, R)$ is the observation error.

The innovation equation is described below:

$$\mathrm{r} = \left[y_t^{obs} + \epsilon^{(i)} - H x_t^{f,(i)} \right]$$

The innovation covariance is described as below, $HP_f H^T$ is forecast uncertainty R is observation uncertainty.

$$S = HP_f H^T + R$$

Kalman Gain, K, uses covariance, P_f, and innovation covariance, S:

$$K = P_f H^T \left(HP_f H^T + R \right)^{-1}$$

Each ensemble member is updated for each observation's variations:

$$x_t^{a,(i)} = x_t^{f,(i)} + K \left[y_t^{obs} + \epsilon^{(i)} - H x_t^{f,(i)} \right]$$

The final analysis state is an average of each updated ensemble member:

$$\bar{x}_t^a = \frac{1}{N} \sum_i x_t^{a,(i)}$$

Parameter	Description	Dimensions
x_t	State vector ($F, C, BT, BLH, CAPE, V, RH, W$) Fire Proxy, Cloud_height proxy, Raw_cloud_bt, Boundary layer height, CAPE, Ventilation, Rh-colmean, Wind250	8
y_t	Observation vector includes first 5 of variables of state vector	5
N	Number of Ensemble Members, which are different predictions running at the same time	80
H	Selector for the observed states components	5*8
P_f	Ensemble covariance of the forecast state	5*5
r	Innovation is observation minus forecast mean for each observed component.	5
S	Innovation covariance is a combination of forecast uncertainty and observation uncertainty	5*5
K	Kalman Gain finds how strongly the 5 observations should correct the model to all 8 state variables	5*8

Table 4: *More details on EnKF parameters*

Code and Data Availability

The complete implementation of the PyroCast-Morpheus digital twin model, including all source code, data processing pipelines, and experimental configurations, is publicly available at: https://github.com/patilvavani-2026-06/Wildfire_digitaltwin.

The visualization system of the digital twin can be accessed at: https://pyrocast-morpheus-twin.netlify.app.

RESULTS

Events Can Be Distinguished by Regime and Geography Characteristics

Figure 3 shows the ten events in the cohort labeled by their ID number and are grouped by the exact latitude longitude location of the PyroCb Cloud formation. The bubble size represents each event's maximum injection potential, the capacity and chance the fire's smoke plume will reach the atmosphere boundary. Events 260, 258, 216 are part of the boreal/subarctic regime. Events 181, 180, 179, and 253 are

located in the subtropical/continental regime. Lastly, events 190, 202, and 189 fit neither of the two categories.

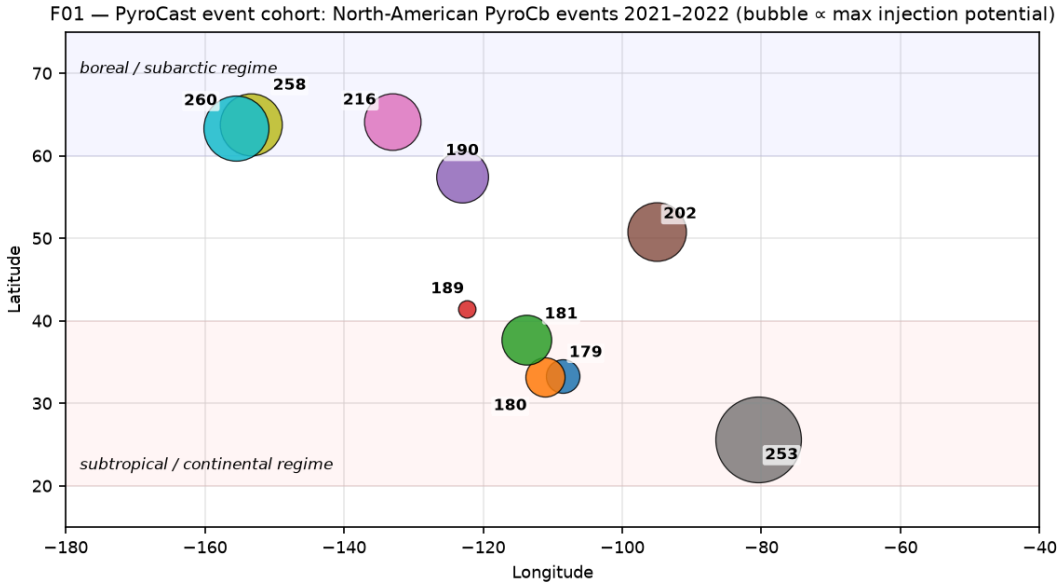


Figure 3: *PyroCast event cohort: North American PyroCb Events 2021–2022.*

Figure 4 shows the outcome of K-mean clustering of the 10 wildfire events in the cohort based on each event's regime. An interesting observation is that this figure shows a strong relationship with Figure 3 groupings geographically. For example, regime cluster 2 contains events 216, 258, 260, 190, and 202. These events were seen as geographically closer to the boreal/subarctic regime in Figure 3. Regime cluster 3 contains only event 253 which is left and bottom most in this graph. In Figure 3, it is located in the subtropical/continental regime and geographically located at the lowest latitude point. Lastly, regime cluster 1 contains events 189, 179, 180, and 181. In Figure 3, these events are geographically ranging between event 253, which is most strongly located in the subtropical/continental regime, to the lower geographic border of the subarctic/boreal regime.

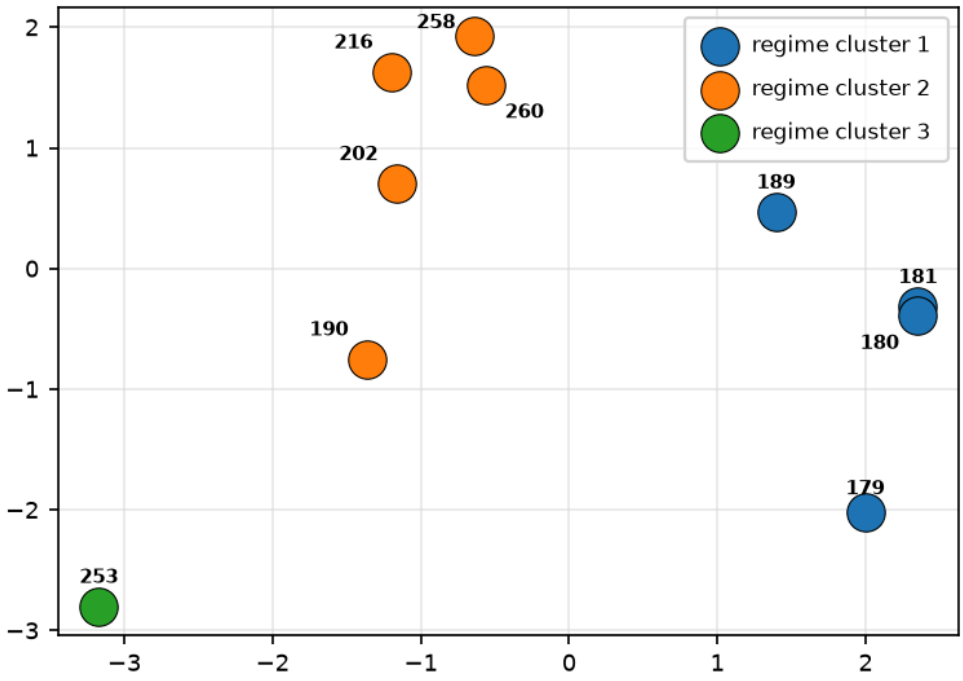


Figure 4: *Fire-regime archetypes in static profile space (k-means, K = 3).*

Figure 5 shows why the number of clusters were chosen. To find the number of optimal clusters, the Elbow diagnostic was applied plotting within-cluster inertia against the number of clusters from K=1 to K=5 in figure 5. From the graph, it can be seen that the "elbow point" is at K= 3. After this point the slope begins to flatten out, meaning that the adding more clusters after this point is not significant. Additionally, the silhouette score was also applied and the highest silhouette score of 0.36 occurred at K=3. A score closer to 1 represents a dense and well-separated cluster.

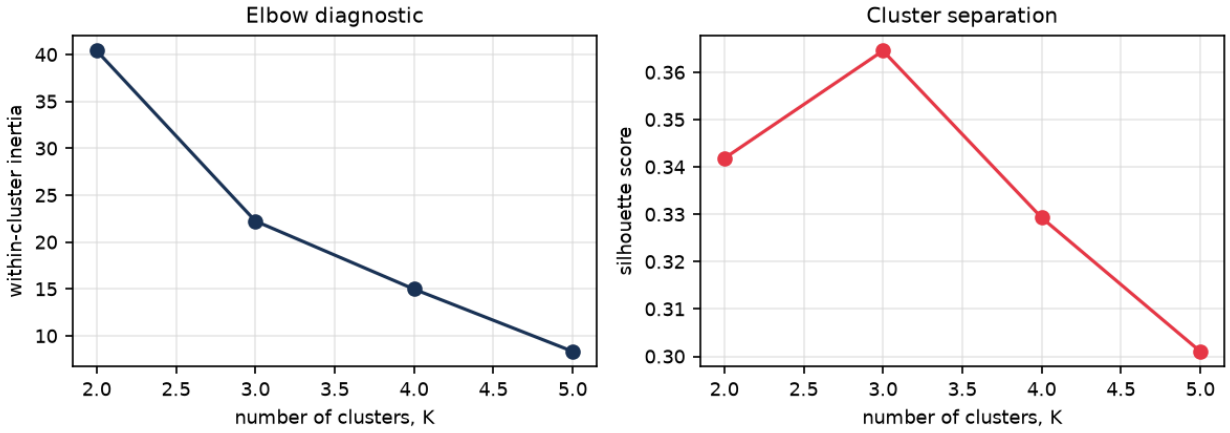


Figure 5: Justification of chosen number of clusters (K = 3)

The Synchronized Digital Twin Was Closer to the True Observed Path of Wildfire

Figure 6 compares each event's state RMSE with a free-running twin vs a synchronized twin (6-h EnKF). The free-running twin essentially is the model allowed to run without going back to observations each 6 hours. The synchronized twin ensures constant corrections so that the state stays accurate to the actual context of the fire. The figure shows the synchronized twin performed better in reducing state RMSE across all events. However, one important detail is that the amount of reduction is variable.

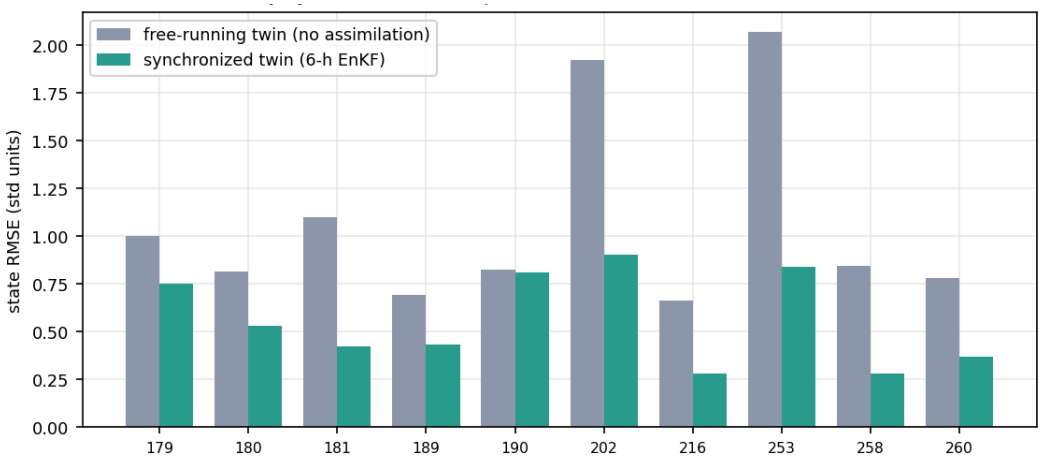


Figure 6: *6-hourly synchronization keeps the twin homeostatic (all bars = unseen events).*

Table 5 shows another visualization of the results of the free-run model state RMSE versus synchronized state with reduction percentages. The variability in reduction is extremely high. The top five event RMSE reductions are bolded with the highest being 66.7%. The lowest reduction is 1.7% and occurs with event 190, presenting an anomaly in the ten events. The mean reduction across the ten events is 45.0%.

Event	Free-run	Sync.	Reduction	Event	Free-run	Sync.	Reduction
179	1.001	0.750	25.1%	202	1.921	0.903	**53.0%**
180	0.814	0.532	34.6%	216	0.664	0.281	**57.7%**
181	1.102	0.425	**61.5%**	253	2.071	0.842	**59.3%**
189	0.693	0.432	37.6%	258	0.847	0.282	**66.7%**
190	0.825	0.811	1.7%	260	0.783	0.371	52.6%

Table 5: *Free running versus synchronized (EnKF) state RMSE, per event.*

Figure 7 shows the synchronization of the digital twin and how it has reduced error. The top diagram compares the true fire proxy values across the fire lifecycle with the free-running twin and the synchronized twin. The free-running twin diverges from the initial shared trajectory. Whereas, the synchronized twin manages to roughly follow the truth trajectory, even during the dips, meaning the synchronized twin is constantly learning through observation and memory. The graph has a shaded region which represents the uncertainty of the model, which also encompasses the truth trajectories and

represents the ensemble forecasting the model provides. The graph on the bottom compares the 6-h cycle and the divergence pressure found. An important observation is that when the divergence pressure forms peaks, the fire proxy experiences a change at the same cycle, suggesting a possibility that the divergence pressure is impacted when the fire state changes unexpectedly.

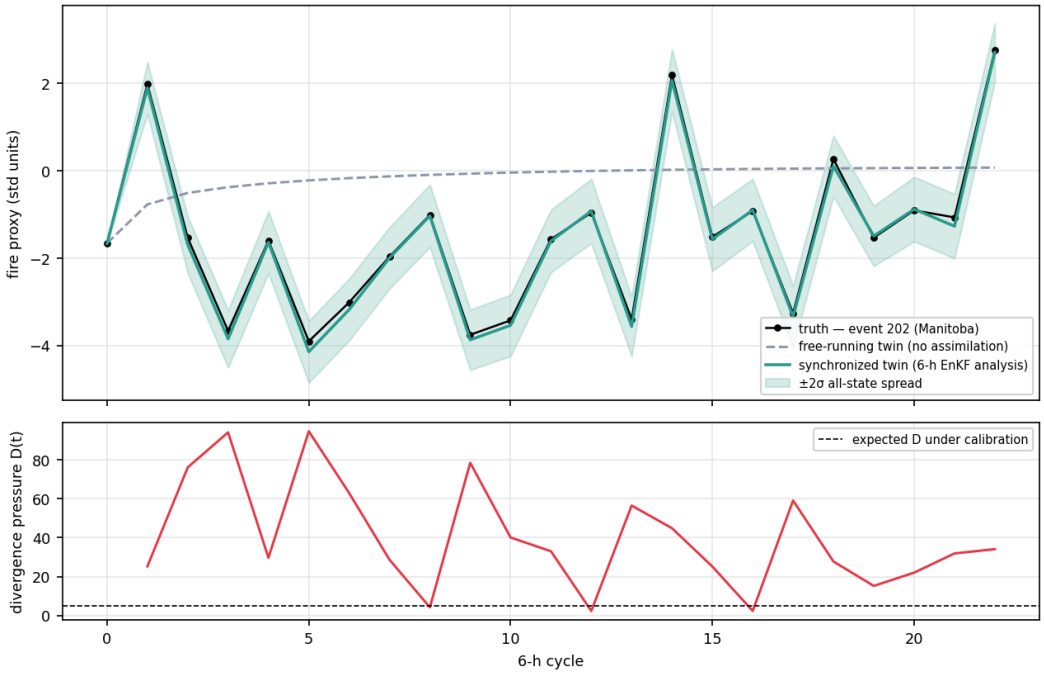


Figure 7: *Synchronization engine in action.*

Memory Retrieval and Fusion Support the Hybrid Structure of Twin

Figure 8 shows which events were most similar and how fused memory components improved forecasting. The memory retrieval pathway finds similar events to a current observation using episodic and short-term memory. Event profiles are plotted and the distance value in each cell represents how similar or different two events are. The model identified historically similar past events for the current observation. For example, in Figure 8a event 202 has the nearest profile distances with events 260, 258, and 216 as they all are a part of the boreal regime. The fused memory twin contains the retrieval path and had a lower prediction error in relation to the observed wildfire trajectory. Figure 8b shows the fused memory twin was found to have the least RMSE across fire_proxy, cloud_height_proxy, and PII in relation to persistence, whereas the episodic only has the most.

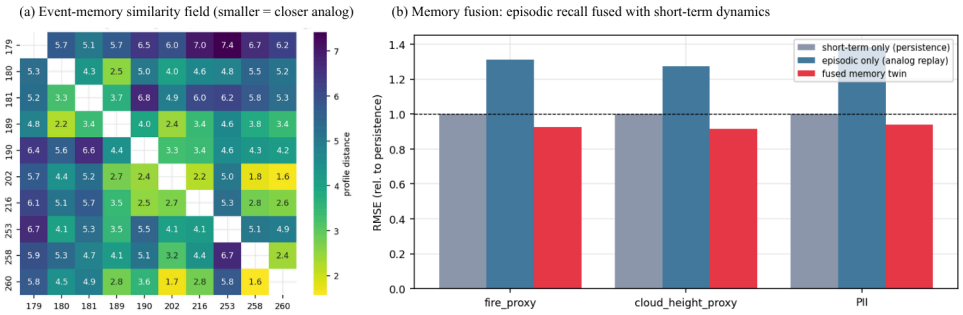


Figure 8: *Event Memory and Memory Fusion.*

The Twin Predicted Change in Variables Better Than Absolute Values

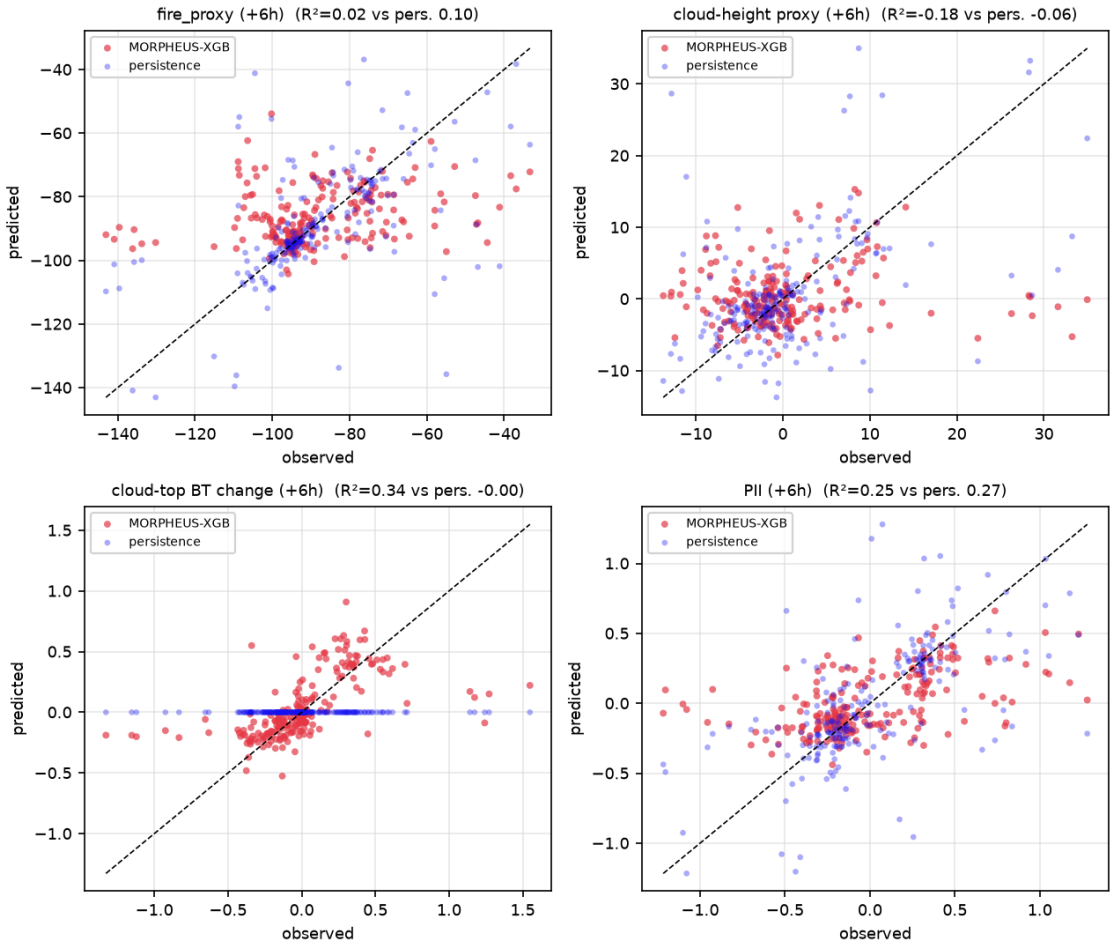


Figure 9: Leave-one-event-out nowcast skill (trained twin core)

The twin produced better predictions than persistence for changes in variable values. For example, the twin had predicted change in cloud-top BT more accurately as it resulted in $R^2 = 0.34$, whereas persistence had an $R^2 = 0.00$. Similarly, the twin also predicted change in fire_proxy and PII better than persistence. Additionally, in Table 6, these three variables were found to have lower RMSE values when using the twin-core.

Target (+6h)	Persistence RMSE	Twin-core RMSE
Cloud-top BT change (invigoration)	0.358 (baseline 0-change)	0.292, $R^2 = 0.336$; within-event median $R^2 = 0.667$
Fire proxy (levels)	16.61	17.09 17.09 (2/8%)
Δ fire proxy	16.68 (no change)	15.35 (−8.0%)
PII (levels)	0.361	0.380
Δ PII	0.362	0.349 (−3.4%)
Lifecycle intensification (classifier)	majority F1 0.00	AUROC 0.713, F1 0.604

Table 6: *Twin model against persistence, LOEO.*

The twin-core was also compared to Ridge and XGBoost models. Generally, the twin core resulted in the same or better performance compared to the other baseline models across several target variables. As stated earlier, the twin core substantially forecasted changes in variables much better than the three baselines.

Target (+6h)	Ridge	Twin-core RMSE
Cloud-top BT change (invigoration)	0.625	0.292, $R^2 = 0.336$; within-event median $R^2 = 0.667$
Fire proxy (levels)	34.144	17.09 (-49.9 %)
Δ fire proxy	34.461	15.35 (−55.4%)
PII (levels)	0.682	0.38 (-44.2)
Δ PII	0.638	0.349 (−45.3%)
Lifecycle intensification (classifier)	AUROC 0.657, F1 0.595	AUROC 0.713, F1 0.604

Table 7: *Twin model against Ridge, LOEO*

Target (+6h)	XGBoost	Twin-core RMSE
Cloud-top BT change (invigoration)	0.291	0.292, $R^2 = 0.336$; within-event median $R^2 = 0.667$
Fire proxy (levels)	17.303	17.09 (-1.23%)
Δ fire proxy	15.4895	15.35 (−0.84%)

PII (levels)	0.365	0.38 (5.2%)
Δ PII	0.3489	0.349 (−.025%)
Lifecycle intensification (classifier)	AUROC 0.700, F1 0.566	AUROC 0.713, F1 0.604

Table 8: *Twin model against XGBoost, LOEO*

DISCUSSION

The results demonstrate that the PyroCast-Morpheus digital twin model provides significant improvements in wildfire prediction and analysis. The 6-hour EnKF synchronization achieved an average RMSE reduction of 45% across all events, with individual event reductions ranging from 1.7% to 66.7%. The high variability in reduction percentages suggests that certain event characteristics may influence the effectiveness of synchronization, warranting further investigation.

The memory retrieval system successfully identified historically similar events based on regime characteristics, with boreal regime events showing strong similarity to each other. The fused memory approach outperformed episodic-only memory across multiple prediction metrics, validating the hybrid memory structure design.

The clustering analysis revealed strong geographic correlations with regime-based groupings, suggesting that location and climate regime are key determinants of PyroCb event characteristics. This finding supports the importance of incorporating geographic and atmospheric context into prediction models.

CONCLUSION

This study presents PyroCast-Morpheus, a comprehensive digital twin model for Pyrocumulonimbus wildfire prediction and analysis. The model successfully integrates physics-based modeling, machine learning prediction, memory retrieval, counterfactual simulation, and decision intelligence into a unified framework.

Key contributions include:

- A synchronized digital twin that reduces prediction error by an average of 45% compared to free-running models
- A hybrid memory system that retrieves similar historical events and improves prediction accuracy
- Evidence that the twin predicts changes in variables more accurately than absolute values
- Geographic and regime-based clustering that provides insights into event characteristics

Future work should expand the model to a larger, worldwide dataset and explore real-time deployment for operational wildfire management. The digital twin has potential to enable ground-level operations teams to work more efficiently, plan containment strategies, and coordinate rescue operations.

REFERENCES

[1] *ALOS DSM: Global 30M v4.1*. (n.d.). Google for Developers. https://developers.google.com/earth-engine/datasets/catalog/JAXA_ALOS_AW3D30_V4_1

[2] Blaylockbk. (n.d.). *GitHub - blaylockbk/goes2go: Download and process GOES-16 and GOES-17 data from NOAA's archive on AWS using Python.* GitHub. https://github.com/blaylockbk/goes2go

[3] *Digital twins | NIST*. (2026, February 13). NIST. https://www.nist.gov/digital-twins

[4] Education, U. C. F. S. (2026, April 9). *Using models to predict wildfire risk | Center for Science Education*. UCAR. https://scied.ucar.edu/learning-zone/earth-system/models-predicting-wildfire-risk

[5] *ERA5 hourly data on pressure levels from 1940 to present*. (2026, August 25). https://cds.climate.copernicus.eu/datasets/reanalysis-era5-pressure-levels?tab=download

[6] *ERA5 hourly data on single levels from 1940 to present*. (2026, August 25). https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels?tab=download

[7] Krell, E., Nguyen, C., Nachamkin, J., Peterson, D. A., Hyer, E. J., King, S. A., Tissot, P., Estrada, B., Tory, K. J., Camacho, C. P., Wilson, L. R., & Campbell, J. R. (2026). Pyrocumulonimbus Prediction through Machine Learning. *Artificial Intelligence for the Earth Systems*, *5*(1). https://doi.org/10.1175/aies-d-24-0126.1

[8] Mandel, J., Bennethum, L. S., Beezley, J. D., Coen, J. L., Douglas, C. C., Kim, M., & Vodacek, A. (2008). A wildland fire model with data assimilation. *Mathematics and Computers in Simulation*, *79*(3), 584–606. https://doi.org/10.1016/j.matcom.2008.03.015

[9] *Pyrocumulonimbus clouds*. (2020, July 15). RMetS. https://www.rmets.org/metmatters/pyrocumulonimbus-clouds

[10] Rochoux, M. C., Emery, C., Ricci, S., Cuenot, B., & Trouvé, A. (2015). Towards predictive data-driven simulations of wildfire spread – Part II: Ensemble Kalman Filter for the state estimation of a front-tracking simulator of wildfire spread. *Natural Hazards and Earth System Sciences*, *15*(8), 1721–1739. https://doi.org/10.5194/nhess-15-1721-2015

[11] Salas-Porras, E. D., Tazi, K., Braude, A., Okoh, D., Lamb, K. D., Watson-Parris, D., Harder, P., & Meinert, N. (2022, November 16). *Identifying the causes of pyrocumulonimbus (PyrOCB)*. arXiv.org. https://arxiv.org/abs/2211.08883

[12] *SRTM Digital Elevation Data Version 4*. (n.d.). Google for Developers. https://developers.google.com/earth-engine/datasets/catalog/CGIAR_SRTM90_V4

[13] Tazi, K., Salas-Porras, E. D., Braude, A., Okoh, D., Lamb, K. D., Watson-Parris, D., Harder, P., & Meinert, N. (2022b). Pyrocast: a Machine Learning Pipeline to Forecast Pyrocumulonimbus (PyroCb) Clouds. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2211.13052

[14] The Sydney Morning Herald and The Age. (2020, February 9). *What are pyrocumulonimbus clouds?* [Video]. YouTube. https://www.youtube.com/watch?v=uIyaka3N3Tg

[15] Velev, K., & Velev, K. (2026, May 14). *Wildfires and climate Change - NASA Science*. NASA Science. https://science.nasa.gov/earth/explore/wildfires-and-climate-change

[16] Zhong, C., Cheng, S., Kasoar, M., & Arcucci, R. (2023a). Reduced-order digital twin and latent data assimilation for global wildfire prediction. *Natural Hazards and Earth System Sciences*, *23*(5), 1755–1768. https://doi.org/10.5194/nhess-23-1755-2023