Cross-Sensor Temporal Fusion for Crop Monitoring Using Weak-Signal Attention and State-Space Modeling

Matteo Willis
Department of Computer Science and Engineering, University of Nevada, Reno, Reno, NV, USA.
matteowillis05@unr.edu

Robert M. Bergman
Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO, USA.
robert.work@missouri.edu

Sagar Gumar
Department of Computer Science, Colorado State University, Fort Collins, CO, USA.
contactsagar@colostate.edu

Abstract

Accurate and timely crop monitoring is essential for global food security, sustainable agriculture, and climate adaptation. Recent advances in Earth observation have produced a rich and heterogeneous archive of satellite and aerial sensor data, including multispectral, hyperspectral, synthetic aperture radar, thermal infrared, and LiDAR modalities. However, the effective exploitation of these diverse data streams remains constrained by challenges of temporal irregularity, spatial resolution mismatches, and the presence of weak signals that precede critical phenological transitions. This paper introduces a comprehensive system-level framework for cross-sensor temporal fusion that integrates weak-signal attention mechanisms with state-space modeling to address these limitations. The proposed architecture operates as a layered pipeline: first, multi-sensor data are aligned and normalized through a unified preprocessing stage; second, a weak-signal attention module selectively amplities subtle spectral and temporal anomalies that are indicative of early stress, nutrient deficiency, or water deficit; third, a state-space model captures the continuous-time dynamics of crop development, accommodating irregular revisit intervals and sensor gaps. The system is designed to be deployed on hybrid cloud-edge infrastructures, balancing computational load with latency requirements for near-real-time applications. The paper further examines structural trade-offs inherent in sensor selection, model complexity, and data governance, and discusses implications for robustness under adversarial conditions, sustainability of large-scale AI inference, and equity in access to precision agriculture technologies. Policy recommendations are offered to promote open data standards, ensure algorithmic fairness, and align monitoring capabilities with global agricultural mandates. This work contributes both a conceptual architecture and a critical evaluation of the socio-technical conditions under which cross-sensor fusion can fulfill its potential.

Keywords

Cross-sensor fusion, temporal modeling, weak-signal attention, state-space models, crop monitoring, remote sensing, agricultural AI, sustainability.

1. Introduction

The transformation of global agriculture into a data-driven enterprise has been catalysed by the proliferation of Earth observation satellites, unmanned aerial vehicles, and in-situ sensor networks. These platforms generate vast and heterogeneous datasets that, if properly integrated, can provide unprecedented insights into crop health, yield prediction, and resource management. Yet the very diversity that enriches these observations also introduces profound methodological challenges. Optical sensors are impeded by cloud cover, synthetic aperture radar (SAR) suffers from speckle noise, thermal imagery captures only surface temperature with coarse spatial resolution, and hyperspectral sensors yield high spectral dimensionality but limited temporal revisit. The need to fuse these disparate modalities across time is not merely a technical convenience but a necessity for robust, year-round monitoring [1,2].

Crop growth is a continuous, non-linear process influenced by weather, soil conditions, and management practices. Phenological transitions—such as emergence, anthesis, and senescence—are often preceded by subtle changes in spectral reflectance or canopy structure that standard vegetation indices fail to capture. These weak signals, buried in noise and overshadowed by dominant seasonal trends, hold the key to early detection of stress and disease [3,4]. Convolutional and recurrent neural networks have been applied to time-series remote sensing data, yet they struggle with irregularly spaced observations and require large volumes of labelled training data. Transformer-based architectures have improved temporal attention, but their quadratic complexity and fixed context windows limit scalability for long-term, high-frequency monitoring [5].

State-space models (SSMs) offer an alternative paradigm that explicitly represents the latent dynamics of crop systems. By modelling observations as a function of an evolving hidden state, SSMs naturally handle missing data and variable sampling intervals. When coupled with attention mechanisms that can amplify weak signals, the combined framework becomes capable of extracting actionable information from sparse and noisy multi-sensor inputs [6,7]. This paper presents a system-level architecture that integrates weak-signal attention and state-space modeling for cross-sensor temporal fusion in crop monitoring. The contribution is not a single algorithmic novelty but a coherent design that addresses the full stack—from sensor calibration to policy implications. The remainder of the paper is structured as follows. Section 2 reviews related work in remote sensing fusion, temporal modeling, and attention mechanisms. Section 3 details the architectural framework. Sections 4 and 5 respectively explore weak-signal attention and state-space modeling in depth. Section 6 discusses system-level trade-offs involving robustness, sustainability, and fairness. Section 7 considers deployment and governance, and Section 8 concludes.

2. Background and Related Work

Crop monitoring using remote sensing has long relied on vegetation indices derived from multispectral sensors such as Landsat and Sentinel-2. These indices, including NDVI, EVI, and LAI, provide proxies for green biomass and chlorophyll content, but they are insensitive to early stress signals that manifest in narrow spectral bands or in thermal emissions [1,8]. Hyperspectral sensors, such as those on the PRISMA and EnMAP missions, offer hundreds of contiguous bands that can detect subtle biochemical changes, but their spatial coverage and temporal frequency are often limited. SAR sensors, particularly C-band and L-band, are sensitive to canopy structure and moisture content, and can penetrate cloud cover, but they are prone to speckle and require sophisticated filtering [9]. LiDAR provides three-dimensional structural information, but is typically collected from airborne platforms with limited spatial extents. The fusion of these modalities has been explored through early, intermediate, and late fusion strategies, with intermediate (feature-level) fusion showing the greatest promise for preserving complementary information [5,10].

Temporal modeling of crop phenology often employs curve-fitting methods (e.g., double logistic, Savitzky-Golay) to reconstruct smooth time series from irregular observations. More recently, deep learning approaches such as long short-term memory networks (LSTMs) and temporal convolutional networks have been applied to predict yield or classify crop types from time-series data [11]. However, these methods assume uniform sampling intervals and require considerable computational resources for long sequences. State-space models, particularly those based on linear dynamical systems and Kalman filters, have been used for data assimilation in agricultural models, but they often assume Gaussian noise and linear transitions. The recent development of structured state-space models (e.g., S4, Mamba) has revived interest in SSMs for deep learning, offering linear-time inference and the ability to model long-range dependencies [6,7]. These models are especially well-suited to the irregular and multi-rate nature of remote sensing time series.

Attention mechanisms have revolutionised sequence modeling, with the Transformer architecture enabling the capture of long-range dependencies through self-attention. In remote sensing, attention has been applied to spatial-spectral feature extraction and temporal fusion [3,4]. Weak-signal attention, a variant that emphasizes subtle deviations from a baseline, has been proposed for anomaly detection in time series. This mechanism is particularly relevant for crop monitoring, where the goal is often to detect small changes that precede visible symptoms. By learning to focus on those time steps or spectral bands where the signal-to-noise ratio is low but informative, weak-signal attention can improve early detection without requiring large amounts of labelled data [12]. The system described in this paper builds on these foundations, integrating weak-signal attention within a state-space framework to achieve robust cross-sensor fusion.

3. Architectural Framework for Cross-Sensor Temporal Fusion

The proposed architecture is designed as a modular pipeline that ingests multi-sensor data streams, aligns them in space and time, extracts features via weak-signal attention, and models temporal dynamics via a state-space model. The first stage, data ingestion and preprocessing, handles a variety of input sources including multispectral (Sentinel-2, Landsat 8/9, MODIS), hyperspectral (PRISMA, EnMAP, DESIS), SAR (Sentinel-1, RADARSAT-2), thermal (ECOSTRESS, Landsat thermal bands), and LiDAR (GEDI, airborne surveys). Each sensor has distinct spatial resolution, swath width, radiometric calibration, and temporal revisit characteristics. The preprocessing stage performs geometric registration to a common grid (e.g., 10 m for Sentinel-2, resampled to 30 m for others), cloud and shadow masking using quality bands, and atmospheric correction using physically based models such as MODTRAN or 6SV [1,8]. For SAR, speckle filtering using Goldstein or Lee filters is applied [9]. The output is a set of coregistered, sensor-specific data cubes with known acquisition timestamps.

The second stage, temporal alignment, addresses the irregularity of observations. Rather than resampling to a fixed time grid (which introduces interpolation artefacts), the system maintains a list of observation events with their timestamps. A learned or handcrafted mapping function transforms each sensor's data into a common feature space. This mapping can be a shallow neural network that predicts a set of universal vegetation attributes (e.g., leaf area index, canopy water content, land surface temperature) from each sensor's radiance or reflectance values. The mapping is trained on limited in-situ validation data or through physics-guided simulations [13]. The output is a time-indexed series of feature vectors, each with an associated quality flag and uncertainty estimate. This design allows the downstream state-space model to handle missing observations naturally.

The fusion core combines weak-signal attention with a state-space model. First, a weak-signal attention module processes the feature time series to produce an attended representation that amplifies subtle deviations from a learned phenological baseline [12]. This baseline is derived from historical data or from a long-term moving average of the same sensor features. The attention weights are computed by a neural network that takes as input the local context (e.g., a window of several observations) and outputs a soft mask that emphasizes time points where the feature vector diverges significantly from the baseline, but only for those features that have low absolute magnitude. A thresholding mechanism prevents large outliers or noise from being misattributed as weak signals. The attended features are then fed into a state-space model that maintains a latent state capturing the true phenological trajectory [6,7]. The state evolves continuously in time according to a linear or nonlinear transition function, and observations are generated from the latent state via an emission function that accounts for sensor-specific noise and quality. Inference is performed using a Kalman filter or a variational inference procedure for nonlinear SSMs.

The final stage produces output products: smoothed time series for each predicted variable (e.g., NDVI, canopy water content, temperature), detection flags for early stress or disease, and uncertainty estimates. The entire pipeline can be deployed on a hybrid cloud-edge infrastructure, with preprocessing and spatial alignment performed on edge nodes (e.g., within satellite ground stations or local servers), and the attention-SSM inference executed on cloud clusters for larger regions. Latency requirements for near-real-time applications (e.g., irrigation scheduling) demand that the SSM inference be optimised, possibly using approximate inference or streaming updates [14].

4. Weak-Signal Attention Mechanisms

Standard attention mechanisms, such as those used in Transformers, assign high weight to positions that are most relevant for predicting a target. In crop monitoring, the most relevant signals may not be the strongest—they are often weak precursors to rapid changes. For example, a slight decrease in the red-edge reflectance (indicative of chlorophyll degradation) may precede a visible yellowing by several days. Similarly, a small increase in canopy temperature (due to stomatal closure) can signal water stress long before leaf wilting occurs. Weak-signal attention is designed to focus on these subtle deviations while ignoring strong seasonal patterns that are already well modeled by the state-space dynamics [3,12].

The architecture implements weak-signal attention as a separate module after feature extraction but before the state-space model. Given a sequence of feature vectors (each time step corresponds to an observation from any sensor), the module first computes a baseline trajectory for each feature dimension. This baseline can be an ensemble mean from previous years for the same geographic location, or a running median from recent observations. The deviation signal is computed as the difference between the observed feature and the baseline. Then, a score is assigned to each time step based on two criteria: the magnitude of the deviation relative to the local noise level, and the persistence of the deviation across consecutive observations. Time steps with large, persistent deviations that are small in absolute value (i.e., weak signals) receive high attention scores. Strong signals (e.g., large drough-induced changes) are already captured by the state-space model and do not need amplification; they may even be downweighted to prevent overfitting [4].

The attention weights are produced by a small neural network that takes as input a window of deviation signals and noise estimates. The network is trained on labelled examples where early stress events have been identified from ground-truth surveys. The training objective is to maximize the correlation between attention weights and the time steps closest to the onset of the event. Because weak signals are rare, the module uses a contrastive learning objective that distinguishes time steps just before a known event from those far away [15]. The attended features are then a weighted sum of the original feature vectors, with the weights renormalized to maintain statistical consistency. This attention-enhanced representation is fed into the state-space model, which further smooths the time series and produces predictions of future states and observations.

One key trade-off lies in the choice of the window length for computing the baseline. If the window is too long, seasonal changes may be misinterpreted as baseline and weak signals may be discounted; if too short, the baseline may follow stress signals too closely, reducing sensitivity. Adaptive baselines that adjust based on the recent history of sensor availability are under exploration [12]. Furthermore, the weak-signal attention module introduces additional computational overhead, especially when applied to high-dimensional hyperspectral features. In practice, dimensionality reduction (e.g., via principal component analysis or autoencoders) is applied before attention to maintain efficiency [5].

5. State-Space Modeling for Temporal Dynamics

State-space models provide a principled framework for modeling time series with irregular observations and sensor-specific noise. In the context of crop monitoring, the latent state represents the true physiological condition of the crop (e.g., canopy biomass, water content, leaf area index, pigment concentrations) as a continuous function of time. The state evolves according to a process model that captures phenological progression, influenced by known driving variables such as temperature, precipitation, and solar radiation. Observations from different sensors are generated from the state conditional on sensor-specific bias, noise, and missing data mechanisms [6,7].

The state-space model used in the proposed architecture is based on a linear stochastic differential equation with a drift term that incorporates a seasonal component and a local trend. This formulation is computationally tractable and can be estimated efficiently using Kalman filtering and smoothing. The process model includes a velocity term (first derivative of state) to capture growth rates, and parameters that encode the typical growth curve of the crop (e.g., derived from a crop model such as DSSAT or WOFOST [16]). The emission model for each sensor is a linear function of the state plus Gaussian noise; for hyperspectral sensors, the emission function may be nonlinear (e.g., a neural network) to map from latent variables to reflectance in multiple bands. Inference is performed using an extended Kalman filter for nonlinear cases, or a particle filter for highly non-Gaussian scenarios.

A major advantage of state-space models is their ability to handle missing data without imputation. When a sensor observation is missing (e.g., due to cloud cover or sensor outage), the model simply updates the state using the process model alone. This is crucial for crop monitoring in many parts of the world where optical imagery is frequently obscured by clouds.

Moreover, the model can integrate observations from multiple sensors with different temporal frequencies: a high-frequency but noisy SAR observation and a low-frequency but accurate hyperspectral observation can be combined to yield a better state estimate than either alone [17].

The state-space modeling component also provides uncertainty quantification for all predictions, which is essential for decision support in agriculture. Farmers and policymakers need to know not only the expected yield but also the confidence intervals. The Kalman filter yields a posterior covariance that can be propagated forward to obtain prediction intervals. This uncertainty information can be used to guide targeted ground-truthing or to prioritize areas for more frequent satellite acquisitions. The computational scalability of the linear SSM makes it suitable for regional-scale deployments with thousands of fields. For large areas, the model can be run in parallel across a distributed computing infrastructure, with each field or grid cell treated as an independent time series [14].

6. System-Level Implications: Robustness, Sustainability, and Fairness

The deployment of a cross-sensor temporal fusion system at scale raises significant concerns regarding robustness, sustainability, and fairness. Robustness must be considered across multiple dimensions: sensor failure, adversarial attacks, and environmental extremes. If one satellite sensor fails or is decommissioned, the state-space model can continue to operate using remaining data, but with increased uncertainty. The system should be designed with graceful degradation, such that output products retain validity even when one data source is unavailable. Adversarial attacks, such as the deliberate injection of misleading observations into the data stream (e.g., spoofing a drought signal), are a growing concern as agricultural AI becomes economically valuable. Weak-signal attention may be particularly vulnerable because it amplifies small differences; an adversary could craft small perturbations that mimic early stress. Defensive strategies include anomaly detection on the incoming sensor metadata, robust training of the attention network with adversarial examples, and ensembling multiple partial models [18].

Sustainability of the computational infrastructure is another critical dimension. Large-scale attention mechanisms and state-space inference can be energy-intensive, particularly when processing high-resolution hyperspectral cubes over millions of hectares. The trade-off between model accuracy and energy consumption must be openly evaluated. The proposed architecture reduces energy by performing feature extraction (which is relatively lightweight) on edge devices, and by using a linear state-space model that requires only matrix operations rather than transformer self-attention [7]. However, the weak-signal attention module adds overhead. Future work should explore pruning, quantization, and knowledge distillation to reduce the carbon footprint of the fusion pipeline. Furthermore, the use of cloud computing centres powered by renewable energy can mitigate environmental impact, but this is not always available in many agricultural regions [19].

Fairness in access to crop monitoring technologies is a pressing socio-technical issue. High-resolution satellite data (e.g., from commercial constellations) can be expensive, and the required computational infrastructure may be out of reach for smallholder farmers in developing countries. The fusion system should be designed to work with freely available data (Sentinel, Landsat, MODIS) as a baseline, and to incorporate higher-resolution commercial data only when available. Even with free data, there is a digital divide in internet connectivity and computing skills. The deployment strategy must include capacity building, open-source software, and integration with local agricultural extension services. Moreover, the algorithmic bias of attention models must be examined: if the training data oversamples regions with large industrial farms, the weak-signal detection may not generalize to diverse cropping systems in other climates [20]. Ensuring representative training datasets and participatory model validation can mitigate this risk.

7. Deployment and Policy Considerations

Deploying a cross-sensor temporal fusion system at national or global scales requires careful attention to infrastructure, data governance, and regulatory frameworks. The proposed architecture can be implemented as a cloud-based service (e.g., Google Earth Engine or Amazon SageMaker) with APIs for users to upload their own in-situ data or to request specific output products. However, latency constraints for near-real-time applications (e.g., irrigation automation, pest outbreak alerts) may necessitate edge computing nodes that run the weak-signal attention and SSM inference locally. Hybrid architectures in which edge devices perform preprocessing and anomaly detection, and cloud servers execute full seasonal modeling, offer a balanced solution [14].

Data governance is a major consideration. Multi-sensor data come from different agencies and private companies with varying access restrictions. The system must comply with data licensing agreements (e.g., Copernicus open data policy, Landsat data policy) and respect privacy concerns when integrating field-level data from individual farms. A metadata layer that tracks provenance, processing steps, and uncertainty is essential for reproducibility and accountability. Open standards for data formats (e.g., Cloud Optimized GeoTIFF, Zarr) and APIs (e.g., STAC) facilitate interoperability [21].

Policy interventions can accelerate the adoption of such monitoring systems. Governments may subsidise the cost of high-resolution satellite imagery for smallholders, or invest in low-cost ground sensor networks that can cross-validate satellite products. Regulatory frameworks for agricultural AI should require transparency in algorithmic decisions, especially when used for loan eligibility or insurance payouts. The system's predictions of crop failure could have profound economic consequences; thus, a human-in-the-loop validation process should be mandated [22]. Finally, international collaboration on sharing satellite data and Earth observation standards, as exemplified by the Group on Earth Observations (GEO), is essential for a global food security early warning system. The proposed system aligns with the goals of the UN Sustainable Development Goals (SDG 2 Zero Hunger and SDG 13 Climate Action) by enabling timely, equitable, and sustainable monitoring [23].

8. Conclusion

Cross-sensor temporal fusion for crop monitoring presents both formidable technical challenges and transformative opportunities. This paper has proposed a system-level architecture that integrates weak-signal attention with state-space modeling to overcome the limitations of existing approaches. By amplifying subtle spectral and thermal anomalies that precede visible symptoms, and by modeling the continuous-time dynamics of crop growth, the system offers earlier and more reliable detection of stress, disease, and nutrient deficiencies. The architectural design explicitly accounts for sensor heterogeneity, irregular sampling, and uncertainty, making it suitable for operational deployment. Beyond the technical details, the paper has examined critical system-level implications related to robustness, sustainability, and fairness, and has offered policy recommendations to guide responsible implementation. As Earth observation data continue to multiply and artificial intelligence methods advance, the fusion of multiple sensors with appropriate temporal modeling will become an indispensable tool for sustainable agriculture. The proposed framework provides a foundation for future research and practice in this vital domain.

References

1. Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing, 114, 24–31.
2. Zhu, X., Cai, F., Tian, J., & Woodcock, C. E. (2018). Spatiotemporal fusion of multisource remote sensing data: Literature survey, taxonomy, principles, applications, and future directions. Remote Sensing, 10(4), 527.
3. Li, Y., Huang, J., & Zhang, L. (2021). Weakly supervised attention learning for crop disease detection in remote sensing. IEEE Transactions on Geoscience and Remote Sensing, 60, 1–14.
4. Chen, J., Wang, C., & Yan, J. (2022). Temporal attention for early crop stress detection using satellite image time series. Remote Sensing of Environment, 279, 113123.
5. Yang, J. X., Wang, J., Li, Z., Sui, C., Long, Z., & Zhou, J. (2025). HSLiNets: Evaluating Band Ordering Strategies in Hyperspectral and LiDAR Fusion. IEEE Geoscience and Remote Sensing Letters.
6. Gu, A., Goel, K., & Ré, C. (2022). Efficiently modeling long sequences with structured state spaces. In International Conference on Learning Representations (ICLR).
7. Smith, J., Warrington, A., & Turner, R. (2023). State-space models for deep learning: A survey. Journal of Machine Learning Research, 24(1), 1–67.
8. Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. Remote Sensing of Environment, 8(2), 127–150.
9. Moreira, A., Prats-Iraola, P., Younis, M., Krieger, G., Hajnsek, I., & Papathanassiou, K. P. (2013). A tutorial on synthetic aperture radar. IEEE Geoscience and Remote Sensing Magazine, 1(1), 6–43.
10. Hao, P., Wang, L., & Niu, Z. (2020). Intermediate fusion of optical and SAR data for crop classification. International Journal of Applied Earth Observation and Geoinformation, 92, 102157.
11. You, J., Li, X., Low, M., Lobell, D., & Ermon, S. (2017). Deep Gaussian processes for forecasting crop yields. In Advances in Neural Information Processing Systems 30.
12. Xiong, Z., Ye, X., Yaman, B., Cheng, S., Lu, Y., Luo, J., ... & Ren, L. (2026). UniDrive-WM: Unified Understanding, Planning and Generation World Model For Autonomous Driving. arXiv preprint arXiv:2601.04453.
13. Jiang, Z., Huete, A. R., Didan, K., & Miura, T. (2008). Development of a two-band enhanced vegetation index without a blue band. Remote Sensing of Environment, 112(10), 3833–3845.
14. Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. Computers and Electronics in Agriculture, 147, 70–90.

15. Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. In International Conference on Machine Learning (ICML).
16. Jones, J. W., Hoogenboom, G., Porter, C. H., Boote, K. J., Batchelor, W. D., Hunt, L. A., ... & Ritchie, J. T. (2003). The DSSAT cropping system model. European Journal of Agronomy, 18(3-4), 235–265.
17. Crow, W. T., & Ryu, D. (2009). Recent advances in the use of remote sensing for soil moisture estimation. In Advances in Water Resources, 32(12), 1595–1601.
18. Finlayson, S. G., Bowers, J. D., Ito, J., Zittrain, J. L., Beam, A. L., & Kohane, I. S. (2019). Adversarial attacks on medical machine learning. Science, 363(6433), 1287–1289.
19. Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 3645–3650.
20. Tsaftaris, S. A., & Ntouskos, V. (2019). Fair and interpretable machine learning for precision agriculture. AI Magazine, 40(3), 40–54.
21. Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment, 202, 18–27.
22. Zook, M., Barocas, S., Boyd, D., Crawford, K., Keller, E., Gangadharan, S. P., ... & Pasquale, F. (2017). Ten simple rules for responsible big data research. PLOS Computational Biology, 13(3), e1005399.
23. United Nations. (2015). Transforming our world: The 2030 agenda for sustainable development. A/RES/70/1.