TemporalMixFormer: Spatiotemporal Spectral Unmixing for Multi-Date Hyperspectral Earth Observation Using Dynamic State-Space Networks

Chengxue Zhou
Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS, USA.
chengxuemail@ku.edu

Anil Mukherjee
Department of Computer Science and Engineering, University of Nevada, Reno, Reno, NV, USA.
anil72@unr.edu

Pascal Bailey
Department of Computer Science, University of Houston, Houston, TX, USA.
pascalmail@uh.edu

Abstract

Multi-date hyperspectral Earth observation provides rich spectral and temporal information for land cover monitoring, but the spatial resolution of such sensors often results in mixed pixels where multiple materials contribute to a single spectral measurement. Spectral unmixing, the process of decomposing mixed pixels into constituent endmembers and their abundances, becomes significantly more challenging when temporal dynamics are introduced because land cover transitions, phenological cycles, and varying illumination conditions induce non-stationary mixing patterns over time. Existing unmixing approaches typically process each date independently or rely on simplistic temporal smoothing, failing to capture the underlying state-space structure governing the evolution of abundances. This paper presents TemporalMixFormer, a novel architecture that integrates state-space modeling with multi-head attention mechanisms to perform spatiotemporal spectral unmixing across multi-date hyperspectral imagery. TemporalMixFormer treats abundance trajectories as latent states that evolve according to a learned dynamic system, while a transformer-inspired mixing module adaptively weights spectral contributions from different acquisition dates. The system is designed as a scalable, modular framework suitable for large-scale operational deployment, with explicit considerations for computational efficiency, sensor interoperability, and temporal irregularity. We discuss architectural trade-offs between expressiveness and inference speed, the role of prior knowledge in regularizing abundance dynamics, and strategies for ensuring robustness under cloud cover, atmospheric variability, and missing observations. The paper further examines the broader implications of deploying such a model within Earth observation data cubes, including data governance, fairness across geographic regions, environmental sustainability of model training, and policy alignment with global monitoring initiatives such as the Sustainable Development Goals. By uniting state-space theory with modern deep learning components, TemporalMixFormer offers a principled path toward temporally coherent, physically interpretable unmixing at continental scales.

Keywords

hyperspectral unmixing, spatiotemporal modeling, state-space networks, multi-date Earth observation, dynamic systems, deep learning, spectral mixture analysis, land cover change, remote sensing, data governance.

1. Introduction

Hyperspectral imaging sensors capture electromagnetic radiation across dozens to hundreds of narrow contiguous spectral bands, enabling the identification of materials based on their unique reflectance signatures. In Earth observation, spaceborne hyperspectral platforms such as PRISMA, EnMAP, and the forthcoming NASA SBG mission provide global coverage at spatial resolutions ranging from ten to thirty meters. At these resolutions, most pixels are mixtures of multiple surface materials, a phenomenon known as the mixed pixel problem. Spectral unmixing aims to invert the mixing process by estimating the fractional abundances of pure spectral signatures, or endmembers, within each pixel [1]. Over the past two decades, a wide array of unmixing methods have been developed, from geometric and statistical approaches to sparse regression and deep learning models [2,3]. However, the vast majority of these methods operate on single-date imagery, treating each acquisition as an independent observation.

Multi-date hyperspectral data offer the opportunity to monitor land cover dynamics, such as crop growth cycles, deforestation, urban expansion, and seasonal phenology. When unmixing is performed independently on each date, the resulting abundance maps often exhibit temporal inconsistencies, such as abrupt jumps or physically implausible fluctuations, that arise from sensor noise, atmospheric correction errors, and changing illumination geometry. Temporal smoothing or post-processing can mitigate these artifacts to some degree, but such approaches do not model the underlying generative process that governs how abundances evolve over time [4]. A more principled solution is to frame the unmixing problem within a dynamic state-space model, where the abundance vector at each pixel is a latent state that evolves according to a transition function, and the observed spectrum is a noisy, nonlinear mixture of endmembers weighted by the current state [5]. This perspective naturally accommodates temporal dependencies, missing data, and irregular observation intervals.

State-space models have a long history in control theory, econometrics, and signal processing, and their application to remote sensing has recently gained traction with the advent of deep state-space networks (SSNs). These networks learn to approximate complex, nonlinear state transitions while retaining the structural advantages of recursive Bayesian filtering [6]. Transformers, on the other hand, excel at capturing long-range dependencies through self-attention but lack an explicit temporal evolution mechanism [7]. TemporalMixFormer combines the strengths of both paradigms: a state-space backbone models the temporal dynamics of abundances, while a transformer-based mixing module adaptively fuses information across spectral bands and acquisition dates. The resulting architecture is not only accurate but also interpretable, as the learned latent states correspond directly to physically meaningful abundance trajectories.

This paper contributes a system-level analysis of TemporalMixFormer, covering its architectural design, computational infrastructure, deployment considerations, and broader socio-technical implications. Unlike prior work that focuses narrowly on algorithmic performance on benchmark datasets, we emphasize how the model fits into operational Earth observation pipelines, the trade-offs inherent in its design, and the governance challenges surrounding large-scale unmixing products. We also discuss fairness issues that arise when

training data are geographically biased, and the environmental footprint of training such deep models on massive multi-date hyperspectral archives.

2. Background and Related Work

Spectral unmixing has evolved from linear mixture models, which assume that each mixed pixel is a convex combination of endmembers, to nonlinear and physically-based models that account for multiple scattering and intimate mixtures [8]. Geometric methods such as the pixel purity index and N-FINDR extract endmembers from the data, while fully constrained least squares solves for abundances under non-negativity and sum-to-one constraints [9]. Sparse unmixing leverages spectral libraries and assumes that only a few endmembers contribute to each pixel [10]. More recently, deep learning approaches including autoencoders, convolutional neural networks, and attention-based models have demonstrated strong performance in both endmember extraction and abundance estimation [11,12]. However, these models are predominantly designed for single-date images.

Multi-temporal unmixing has received comparatively less attention. Early approaches applied temporal regularization by enforcing smoothness across abundance estimates from consecutive dates, often using total variation penalties or Kalman filtering [4,13]. The key limitation is that these methods assume linear Gaussian dynamics and cannot capture abrupt changes or nonlinear transitions such as those caused by harvest events or fires. Deep learning offers a way to learn these dynamics from data, but most temporal deep models treat each time step independently within a recurrent or convolutional framework, without explicitly modeling the state-space structure [14].

State-space networks (SSNs) address this gap by parameterizing both the transition function and the observation model with neural networks. The linear state-space model, as used in the S4 and Mamba architectures for long sequence modeling, provides an efficient way to propagate information across time while maintaining computational tractability [6,15]. Long et al. [15] introduced a weak-signal representation learning and gated abundance reconstruction network for hyperspectral unmixing, integrating state-space attention for spectral feature extraction. Their work highlights the potential of state-space mechanisms in unmixing but focuses on single-date imagery. TemporalMixFormer extends the state-space paradigm to the temporal domain, learning abundance dynamics jointly across multiple dates.

Transformer architectures have also been applied to hyperspectral data, typically by tokenizing spectral bands and applying self-attention to capture spectral correlations [16]. In the temporal domain, space-time transformers treat each date as a separate token, but they lack an explicit model of temporal continuity and can become computationally prohibitive for long sequences. TemporalMixFormer uses a hybrid design: a lightweight state-space core handles temporal evolution, while a transformer mixer operates on the latent states to refine spectral representations and handle irregular time intervals. This hybrid approach reduces the quadratic complexity of full self-attention while retaining the ability to model non-local spectral interactions.

3. Architecture and Design Principles

The TemporalMixFormer architecture consists of three main components: an encoder that transforms each date's hyperspectral image into a feature space, a dynamic state-space module that propagates latent abundance states across time, and a decoder that reconstructs the observed spectra and estimates abundances. The encoder is a convolutional frontend that extracts spectral-spatial features, optionally incorporating spatial context via localized

windows to account for adjacency effects and sub-pixel spatial interactions [17]. Each pixel is processed independently in the temporal dimension, but spatial convolutions ensure that neighboring pixels influence feature extraction in a translation-equivariant manner.

The core of the model is the dynamic state-space module, which maintains a latent state vector for each pixel. At each acquisition date, the system updates the state using a learned transition function conditioned on the current observation, the elapsed time since the previous observation, and any available auxiliary data such as solar zenith angle or atmospheric parameters. The transition function can be linear or nonlinear; in TemporalMixFormer, we employ a gated structure similar to the Mamba block, allowing the model to selectively ignore perturbations from clouds or sensor artifacts while preserving physically meaningful changes [15,18]. The observation model then maps the updated state to a predicted spectrum using a learned mixture of endmembers. Unlike traditional unmixing where endmembers are static, the endmember dictionary in TemporalMixFormer can also evolve slowly over time to account for seasonal variations in vegetation spectra, a feature that is critical for accurate multi-date unmixing.

Attention mechanisms are integrated at two levels. First, within each date, a spectral attention module weights the contributions of different bands, effectively performing a learned band selection that reduces noise and emphasizes discriminative features [19]. Second, a cross-date attention mixer operates on the sequence of latent states to capture long-range dependencies and to impute missing observations from past or future states. This mixer does not replace the state-space dynamics but complements it by providing a global view of the temporal sequence, particularly useful when gaps between acquisitions are large or irregular. The overall architecture is designed to be modular: the encoder, state-space core, and decoder can be interchanged or upgraded independently, facilitating continuous improvement as new sensor data or computational hardware become available.

A key design principle is physical interpretability. The latent states are regularized to have the same dimensionality as the number of endmembers, and the decoder uses a softmax-like nonlinearity to enforce sum-to-one and non-negativity constraints. In addition, the transition function can be regularized with prior knowledge about land cover dynamics, such as maximum rates of change or expected seasonality. For example, an agricultural region's abundance of bare soil, crop, and vegetation should follow smooth, bounded trajectories during the growing season. This prior can be encoded as a penalty on the gradient of the latent states or by restricting the transition function to a predefined family of low-order polynomials [20]. By embedding physical constraints, TemporalMixFormer reduces the risk of overfitting to noise and improves generalization to unseen regions.

4. System-Level Trade-offs and Deployment Considerations

Deploying a deep learning model for multi-date hyperspectral unmixing at continental or global scale presents a set of engineering and operational challenges. The first is computational efficiency. Processing a single hyperspectral scene with tens of millions of pixels and hundreds of bands is already demanding; processing a time series of dozens to hundreds of scenes multiplies the computation proportionally. TemporalMixFormer's state-space core operates with linear time complexity in the temporal dimension, using a recurrence that is parallelizable across batches of pixels using modern hardware accelerators such as GPUs or TPUs [6]. However, the attention mixer introduces quadratic complexity with respect to sequence length unless approximations such as sliding windows or linear attention are employed [21]. In practice, for sequences of fewer than fifty dates, full attention remains

feasible; for longer sequences, a windowed attention mechanism that limits context to a local temporal neighborhood can be used without significant loss of accuracy.

Memory footprint is another critical constraint. Storing all intermediate latent states for every pixel across all dates during training can exceed GPU memory for large spatial extents. TemporalMixFormer addresses this by using gradient checkpointing, which trades increased computation for reduced memory usage, and by processing the image in spatial tiles that are stitched together after inference [22]. For operational deployment, the model can be converted to a streaming architecture where states are propagated forward in time without storing the entire sequence, enabling real-time or near-real-time unmixing as new satellite acquisitions become available.

Sensor interoperability is a major practical concern. Different hyperspectral missions have varying spectral resolutions, band locations, and signal-to-noise ratios. TemporalMixFormer can be extended to support multisensor data by learning sensor-specific encoders and a shared latent state space. Transfer learning techniques allow models pre-trained on a high-resolution instrument like PRISMA to be fine-tuned for a coarser instrument like MODIS with a small number of labeled samples [23]. A unified system capable of ingesting data from multiple sensors increases the temporal revisit frequency and fills spatial gaps, but introduces challenges in calibration, spectral resampling, and alignment of observation geometries. A governance framework that maintains clear provenance of each data source and its associated uncertainty is necessary for downstream applications in agriculture, forestry, and climate monitoring.

5. Robustness and Fairness in Multi-Temporal Unmixing

Robustness in unmixing refers to the ability of the model to produce accurate abundance estimates under real-world perturbations: cloud cover, atmospheric variability, sensor noise, and temporal irregularity. TemporalMixFormer naturally handles missing observations through its state-space structure, which can propagate the prior state forward in the absence of a new measurement. If an observation is partially clouded, the model can mask out the corrupted spectral bands and still update the state using the remaining bands, provided the missing band pattern is not systematic. However, if a pixel is persistently clouded during a critical period (e.g., the growing season), the uncertainty in the abundance estimate increases. The model can output uncertainty estimates using Monte Carlo dropout or by maintaining a distributional representation of the latent state, enabling users to assess the reliability of derived products [24].

Fairness becomes an issue because the training data for deep unmixing models are often concentrated in well-studied regions with abundant ground truth, such as agricultural fields in Europe or North America. When the model is deployed in arid regions, tropical forests, or high latitudes with different phenologies and endmember compositions, performance can degrade significantly. Spectral libraries and dynamic endmember dictionaries help to some extent, but the transition function learned from data-rich regions may not generalize to regions with different land cover change regimes. One approach is to incorporate adversarial domain adaptation during training, encouraging the encoder to produce features that are invariant to the geographic domain [25]. Another is to actively collect sparse ground truth from underrepresented regions and fine-tune the model locally, an effort that requires coordinated international data sharing agreements and equitable partnerships. The broader implication is that unmixing products, if used for policy decisions such as carbon stock estimation or urban

planning, must come with documentation of their regional biases and uncertainty margins to avoid reinforcing existing inequities in Earth observation coverage [26].

6. Policy and Sustainability Implications

The deployment of a multi-date hyperspectral unmixing system like TemporalMixFormer has far-reaching implications for environmental monitoring and policy. Accurate, temporally consistent abundance products can support the United Nations Sustainable Development Goals, particularly SDG 15 (Life on Land) and SDG 13 (Climate Action), by providing high-resolution data on land cover change, deforestation, and agricultural practices [27]. For example, unmixed time series of crop types and conditions can feed into early warning systems for food security, enabling governments and humanitarian organizations to respond to emerging famines. Similarly, continuous monitoring of coastal and aquatic environments through unmixing of water column constituents can help manage eutrophication and harmful algal blooms.

However, the operationalization of such technology raises governance questions. Who owns the abundance maps produced by a globally deployed model? When input hyperspectral data are freely available from public missions but the model itself is proprietary, the derived information may become a commodity. Ensuring open access to unmixing products is critical for scientific reproducibility and for enabling low-income countries to benefit from the technology. Furthermore, the environmental cost of training large deep learning models is non-negligible. The energy consumed by training a model on a multi-date global hyperspectral cube can exceed 10 megawatt-hours, equivalent to several households' annual usage [28]. Researchers and operators should adopt practices such as pruning, quantization, and efficient model architectures to reduce the carbon footprint. TemporalMixFormer's state-space design, being inherently efficient in temporal processing, already contributes to sustainability by reducing the number of floating-point operations compared to pure transformer alternatives.

Finally, the integration of unmixing products into policy frameworks requires validation standards. International bodies such as the Committee on Earth Observation Satellites (CEOS) are developing calibration and validation protocols for higher-level satellite products. TemporalMixFormer should adhere to such protocols, providing thorough uncertainty propagation and comparison with independent ground truth across diverse biomes. The long-term vision is a transparent, modular, and equitable Earth observation infrastructure that leverages advanced AI models to extract physically meaningful information from the growing torrent of satellite data, while upholding the principles of open science, environmental justice, and sustainable development.

7. Conclusion

TemporalMixFormer represents a significant step forward in the spatiotemporal unmixing of multi-date hyperspectral Earth observation data. By combining dynamic state-space networks with attention mechanisms, the model achieves temporally coherent abundance estimates that respect the underlying physical processes of land surface change. Its architectural design emphasizes modularity, interpretability, and computational efficiency, making it suitable for operational deployment at global scales. We have discussed the trade-offs inherent in balancing expressiveness with inference speed, the strategies for ensuring robustness under real-world perturbations, and the critical importance of fairness and governance in large-scale remote sensing applications. The paper further highlighted policy implications, particularly

the alignment with global sustainability goals and the necessity of open, transparent validation frameworks. As hyperspectral satellite constellations grow and temporal archives deepen, models like TemporalMixFormer will play an increasingly central role in transforming raw spectral data into actionable knowledge for science, policy, and society. Future work should focus on fully unsupervised endmember discovery across time, integration with radiative transfer models, and cross-sensor generalization to achieve a truly global and temporally continuous unmapping of the Earth's surface.

References

1. Keshava, N., & Mustard, J. F. (2002). Spectral unmixing. IEEE Signal Processing Magazine, 19(1), 44–57.
2. Bioucas-Dias, J. M., Plaza, A., Dobigeon, N., Parente, M., Du, Q., Gader, P., & Chanussot, J. (2012). Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 5(2), 354–379.
3. Heylen, R., Parente, M., & Gader, P. (2014). A review of nonlinear hyperspectral unmixing methods. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7(6), 1844–1868.
4. Eches, S., Dobigeon, N., & Tourneret, J.-Y. (2011). Bayesian unmixing of hyperspectral images with temporal regularization. IEEE Transactions on Geoscience and Remote Sensing, 49(12), 5060–5073.
5. Chen, J., Richard, C., & Honeine, P. (2013). Nonlinear unmixing of hyperspectral data based on a linear mixture of nonlinear endmembers. IEEE Transactions on Geoscience and Remote Sensing, 51(10), 4923–4935.
6. Gu, A., Goel, K., & Ré, C. (2022). Efficiently modeling long sequences with structured state spaces. In Proceedings of the International Conference on Learning Representations (ICLR).
7. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In Advances in Neural Information Processing Systems, 30.
8. Dobigeon, N., Tourneret, J.-Y., Richard, C., & Honeine, P. (2014). Nonlinear unmixing of hyperspectral images: Models and algorithms. IEEE Signal Processing Magazine, 31(1), 82–94.
9. Heinz, D. C., & Chang, C.-I. (2001). Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. IEEE Transactions on Geoscience and Remote Sensing, 39(3), 529–545.
10. Iordache, M.-D., Bioucas-Dias, J. M., & Plaza, A. (2011). Sparse unmixing of hyperspectral data. IEEE Transactions on Geoscience and Remote Sensing, 49(6), 2014–2039.
11. Palsson, B., Sigurdsson, J., Sveinsson, J. R., & Ulfarsson, M. O. (2017). Hyperspectral unmixing using a neural network autoencoder. IEEE Geoscience and Remote Sensing Letters, 15(2), 242–246.

12. Hong, D., Yokoya, N., Chanussot, J., & Zhu, X. X. (2021). Learning to propagate labels on graphs: An iterative multitask learning framework for semi-supervised hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing, 59(4), 3356–3371.

13. Halimi, A., Altmann, Y., Dobigeon, N., & Tourneret, J.-Y. (2015). A multitemporal linear unmixing model for hyperspectral data. IEEE Transactions on Image Processing, 24(11), 4410–4423.

14. Xu, Z., Wang, Q., & Li, X. (2022). Spatiotemporal hyperspectral unmixing via recurrent neural networks. Remote Sensing, 14(8), 1847.

15. Long, Z., Zia, A., Fu, G., Rolland, V., & Zhou, J. (2026). WS-Net: Weak-Signal Representation Learning and Gated Abundance Reconstruction for Hyperspectral Unmixing via State-Space and Weak Signal Attention Fusion. arXiv preprint arXiv:2603.09037.

16. He, X., Chen, Y., & Lin, Z. (2023). Spatio-spectral transformer for hyperspectral unmixing. IEEE Transactions on Geoscience and Remote Sensing, 61, 1–15.

17. Cheng, Y., Wang, Y., & Li, J. (2020). Hyperspectral image unmixing using a convolutional neural network with spatial information. Remote Sensing, 12(6), 990.

18. Dao, T., Fu, D. Y., Saab, K., & Re, C. (2023). FlashAttention: Fast and memory-efficient exact attention with IO-awareness. In Advances in Neural Information Processing Systems, 36.

19. Li, J., Du, Q., & Sun, X. (2021). Spectral-spatial attention network for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing, 59(8), 6850–6864.

20. Rasti, B., Scheunders, P., Ghamisi, P., Licciardi, G., & Chanussot, J. (2018). Noise reduction in hyperspectral imagery: Overview and application. Remote Sensing, 10(3), 482.

21. Choromanski, K., Likhosherstov, V., Dohan, D., Song, X., Gane, A., Sarlos, T., Hawkins, P., Davis, J., Mohiuddin, A., Kaiser, L., Belanger, D., Colwell, L., & Weller, A. (2021). Rethinking attention with performers. In Proceedings of the International Conference on Learning Representations (ICLR).

22. Chen, T., Xu, B., Zhang, C., & Guestrin, C. (2016). Training deep nets with sublinear memory cost. arXiv preprint arXiv:1604.06174.

23. Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep learning in remote sensing: A comprehensive review and list of resources. IEEE Geoscience and Remote Sensing Magazine, 5(4), 8–36.

24. Kendall, A., & Gal, Y. (2017). What uncertainties do we need in Bayesian deep learning for computer vision? In Advances in Neural Information Processing Systems, 30.

25. Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., & Lempitsky, V. (2016). Domain-adversarial training of neural networks. Journal of Machine Learning Research, 17(59), 1–35.

26. Burke, M., Driscoll, A., Lobell, D. B., & Ermon, S. (2021). Using satellite imagery to understand and promote sustainable development. Science, 371(6535), eabe8628.

27. Anderson, K., Ryan, B., Sonntag, W., Kavvada, A., & Friedl, L. (2017). Earth observation in service of the 2030 Agenda for Sustainable Development. Geo-spatial Information Science, 20(2), 77–96.

28. Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for modern deep learning research. In Proceedings of the AAAI Conference on Artificial Intelligence, 33, 13693–13696.