CWE-8
June 7-11, 2026
London, ON, Canada

A Knowledge-Enhanced Two-Stage Transformer Framework for Wind Field Reconstruction from Sparse Measurements

Baoheng Li [a], Teng Wu [b]

[a]*The State University of New York at Buffalo, Buffalo, NY, USA, baohengl@buffalo.edu*
[b]*The State University of New York at Buffalo, Buffalo, NY, USA, tengwu@buffalo.edu*

SUMMARY

Pointwise measurement devices are widely used in wind tunnels and field studies, yet reconstructing complete spatiotemporal wind fields from sparse data remains challenging. Conventional methods (e.g., semi-empirical fitting, interpolation, data assimilation, and conditional simulation) struggle to recover turbulence structure in non-stationary flows. Projection-based methods (e.g., POD and DMD) can reconstruct dynamic wind fields but rely on low-rank assumptions. Leveraging data-driven advances, transformers are employed here to map low-spatial-resolution and high-temporal-resolution measurements to high-resolution spatiotemporal fields. To mitigate the curse of dimensionality, a two-stage transformer framework is proposed. The first stage estimates inflow from sparse measurements, and the second captures downstream evolution to recover the entire 3D wind field. The domain knowledge of divergence-free constraints is integrated into the loss to improve physical consistency. Trained using CFD data and validated with simulations and experiments from the UB multi-fan wind tunnel, the proposed framework achieves a mean absolute velocity error of 0.2m/s.

Keywords: *Wind field reconstruction, Sparse measurements, Transformer, Computational fluid dynamics, Domain knowledge*

1. INTRODUCTION

The aerodynamic loads of flexible structures depend strongly on the three-dimensional spatiotemporal characteristics of the wind field. However, wind data are typically obtained using pointwise measurement devices with limited locations. Under the sparsely sampled condition, it is difficult to directly capture the complete spatiotemporal wind field from measurements. While the high-fidelity computational fluid dynamics (CFD) tools may be able to provide sufficiently high spatial and temporal resolution of wind flow, their accuracy needs measurement data to be validated. In addition, their applications are limited by the extremely high computational demand. Therefore, reliable wind field reconstruction based on limited measurements is essential.

The semi-empirical profile fitting and interpolation-based methods are popularly used to reconstruct wind fields. However, these approaches are confined to planar reconstruction and cannot recover three-dimensional (3D) volumetric flows. Profile fitting assumes that the wind follows a specific empirical form (e.g., a power or logarithmic law) and estimates its parameters from sparse measurements. This method is only suitable for stationary wind fields with well-defined statistical profiles and cannot reproduce wind-velocity time series or non-stationary winds with time-varying statistics. Moreover, real field or wind-tunnel wind flows may not conform perfectly to a single empirical law, leading to notable discrepancies between reconstructed and actual fields. When no empirical formulas are available, interpolation is typically used based on the assumption that velocity depends solely on nearby measurements. This assumption conflicts with the complex multi-scale spatial correlations generated by turbulence.

The reconstruction of arbitrary spatiotemporal wind fields using pointwise measurement systems represents a typical sparse field reconstruction problem. However, common reconstruction approaches, including conditional estimation and simulation of random fields, data assimilation methods, and reduced-order modeling techniques, each have their own inherent limitations to varying degrees when applied to the wind case. The conditional estimation and simulation methods require a prescribed statistical model (e.g., covariance, spectral, or coherence functions) and calibrate their parameters from sparse data to generate a statistically consistent random field (Leoni et al., 2018). However, these models rely on low-order statistical closures and fail to capture the complexity of nonstationary flows encountered in field or wind-tunnel measurements. Nonstationary turbulence often exhibits non-Gaussianity, intermittency, cross-scale nonlinear coupling, and 3D coherent structures, which are predominantly reflected in higher-order statistics. As a result, conditional-estimation methods based on simplified spectral or covariance representations are insufficient for reconstructing detailed vortex structures in such flows. The data assimilation methods typically require a complete system model and calibration using sparse measurement observations (Xu et al., 2014); however, obtaining such a full-scale model is highly challenging in practical scenarios. In field measurements, it is very expensive and nearly impossible to construct a numerical framework that accurately represents the real conditions. In wind-tunnel applications, the complex geometry of flow-generation devices and associated aerodynamics also hinders the development of a comprehensive model, preventing the true experimental inflow from being reproduced. Consequently, data assimilation methods are limited in reconstructing wind fields under arbitrary spatiotemporal inflow conditions. Some studies infer the inflow as an additional state–parameter variable to compensate for modeling difficulties, but this requires iterative optimization that repeatedly adjusts the simulated flow based on sparse measurements. Such procedures are computationally expensive and prone to local optima, especially when the observations are sparse or the flow is non-stationary.

Several studies have explored projection-based methods to infer latent flow features from sparse measurements and reconstruct spatiotemporal wind fields (Geelen et al., 2024). For stationary-flow reconstruction, these methods rely on the low-rank and compressible nature of the flow data under a fixed domain and boundary condition, assuming that high-dimensional information can be represented by limited dominant modes. The commonly used dimensionality reduction technique is proper orthogonal decomposition (POD). For transient flows, system dynamics must also be modeled. However, the reduced features are not direct physical projections of the Navier–Stokes (NS) equations, making direct substitution into the NS system infeasible due to unclosed terms and instability. To overcome this issue, many studies adopt operator inference, where major NS operators (e.g., advection, diffusion) are reformulated and fitted with reduced-space data to build simplified dynamic systems. Alternatively, the dynamic mode decomposition (DMD) based on Koopman theory identifies dominant dynamic modes, yielding spectral features such as frequencies and growth rates, though it cannot explicitly recover the governing equations. Both approaches, however, remain valid only within the trained subspace. Despite their success, reduced-order models struggle to establish generalizable frameworks, as real wind conditions in both field and wind-tunnel environments rarely follow fixed modal structures. Therefore, a flexible approach capable of learning context-dependent representations from arbitrary wind conditions, without relying on low-rank assumptions, is essential.

Recent advances in sequence modeling, especially transformers, offer a promising solution to this challenge. Based on self-attention, the transformer encodes wind-velocity time series from sparse probes and captures global temporal–spatial dependencies to extract latent features. The decoder then maps these features to each spatial location to reconstruct the complete flow field. Historical inputs help the model learn dynamical patterns and provide contextual cues for accurate prediction. Unlike projection-based methods, the transformer jointly learns spatial structures and dynamics without predefined dimensionality-reduction paths. With multi-scenario training across different domain configurations and inflow conditions, it can develop a unified flow encoding capable of reconstructing time-resolved flow fields from sparse measurements across various flow regimes. However, the direct use of transformers for flow-field reconstruction still encounters the curse of dimensionality. Its global attention mechanism computes dependencies across all temporal and spatial positions, leading to high computational cost for large wind fields due to both dense spatial information and long-range temporal correlations associated with flow advection. To overcome the curse of dimensionality while fully leveraging the global modeling capability of transformers, this study proposes a two-stage spatiotemporal wind field reconstruction framework. The framework comprises two transformer-based modules. The first is an inflow reconstructor to recover the spatiotemporal wind field at the initial cross-section or segment from sparse probe measurements, and hence provides a complete inflow condition. The second is a streamwise propagator to reconstruct the wind field at the subsequent cross-section or segment based on the reconstructed field of the previous one, thereby modeling the downstream evolution of the flow. The schematic of the proposed two-stage transformer framework is shown in Figure 1.

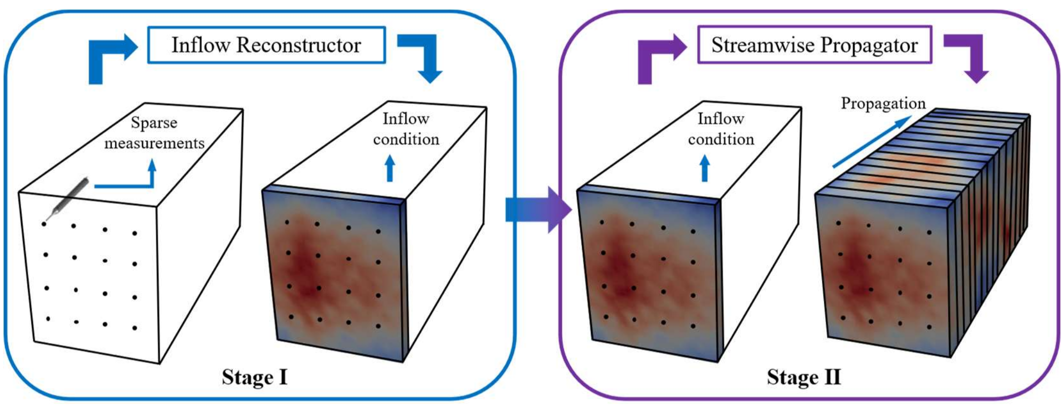

Figure 1: Two-stage transformer framework.

Essentially, this two-stage framework leverages structural similarity across streamwise locations for dimensionality reduction. Each segment is sufficiently narrow relative to the full field, greatly reducing input and output data points and shortening the time for information propagation. This two-stage design decreases both input sequence length and output dimensionality, leading to a smaller model and mitigating the curse of dimensionality. It also enables efficient data utilization during reconstruction. Specifically, any segment can provide sparse measurements for inflow reconstruction, and any adjacent pair can train the streamwise propagator, significantly improving training efficiency. Since the flow in different segments follows the NS equations under similar conditions, reconstructing the initial segment and propagating downstream yields a computationally efficient framework. Although iterative reconstruction introduces accumulated errors and downstream deviations, the reduced dimensionality enhances training stability and accuracy, keeping overall performance comparable to a global transformer. To further enhance the accuracy and physical consistency of the reconstructed flow fields and improve the generalization performance of models to unseen scenarios, domain knowledge of divergence-free constraints is incorporated into the loss functions of both models.

To evaluate the proposed knowledge-enhanced two-stage transformer framework for wind flow reconstruction, a 3D CFD model of the test section of the University at Buffalo (UB) multi-fan wind tunnel was developed. The framework was trained using sparse measurements and full-field data obtained under randomly generated inlet velocity profiles and validated against both CFD simulations and wind tunnel measurements. The reconstruction results under a randomly generated inlet condition presented in Figures 2 and 3 correspond to the validation against the CFD-simulated flow field, where the trained model achieves a mean absolute error (MAE) of approximately 0.2m/s in both the longitudinal (u) and vertical (v) velocity components. The preliminary results demonstrate the capability and potential of the proposed approach for spatiotemporal flow-field reconstruction under arbitrary inflow conditions.

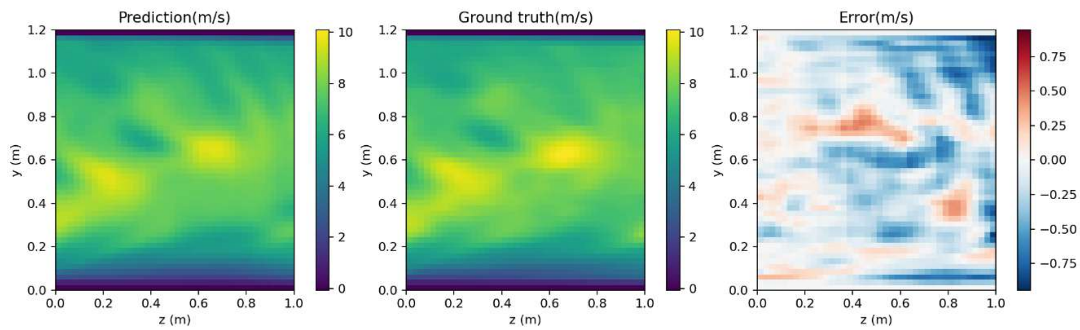

Figure 2: Comparison between predicted and CFD-simulated u-component velocity fields.

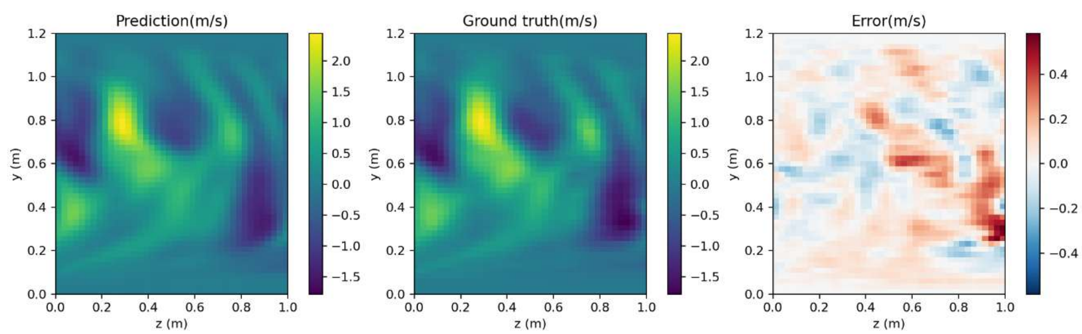

Figure 3: Comparison between predicted and CFD-simulated w-component velocity fields.

REFERENCES

Geelen, R., Balzano, L., Wright, S., Willcox, K., 2024. Learning physics-based reduced-order models from data using nonlinear manifolds. Chaos: An Interdisciplinary Journal of Nonlinear Science 34. https://doi.org/10.1063/5.0170105

Leoni, P.C.D., Mazzino, A., Biferale, L., 2018. Inferring flow parameters and turbulent configuration with physics-informed data-assimilation and spectral nudging. Phys. Rev. Fluids 3, 104604. https://doi.org/10.1103/PhysRevFluids.3.104604

Xu, Y.-L., Hu, L., Kareem, A., 2014. Conditional Simulation of Nonstationary Fluctuating Wind Speeds for Long-Span Bridges. J. Eng. Mech. 140, 61–73. https://doi.org/10.1061/(ASCE)EM.1943-7889.0000589