EMD empowered neural network for predicting spatio-temporal non-stationary channel in UAV communications

Qiuyun Zhang, Qiumei Guo, Hong Jiang, Xinfan Yin, Muhammad Umer Mushtaq, Ying Luo & Chun Wu

School of Information Engineering, Southwest University of Science and Technology, 59 Qinglong Road, Mianyang, 621010, Sichuan, China
Qiuyun Zhang, Qiumei Guo, Hong Jiang, Ying Luo & Chun Wu

College of Intelligence Science and Technology, National University of Defense Technology, No. 137 Yanwachi Street, Changsha, 410073, Hunan, China
Xinfan Yin

Department of Computer Science, University of Pretoria, cnr Lynnwood Road and Roper Street Hatfield, Pretoria, 0082, Gauteng, South Africa
Muhammad Umer Mushtaq

Correspondence to Qiuyun Zhang. Email: zhangqiuyun@swust.edu.cn

Abstract

This paper introduces a novel prediction method for spatio-temporal non-stationary channels between unmanned aerial vehicles (UAVs) and ground control vehicles, essential for the fast and accurate acquisition of channel state information (CSI) to support UAV applications in ultra-reliable and low-latency communication (URLLC). Specifically, an empirical mode decomposition (EMD)-empowered spatio-temporal attention neural network is proposed, referred to as EMD-STANN. The STANN sub-module within EMD-STANN is designed to capture the spatial correlation and temporal dependence of CSI. Furthermore, the EMD component is employed to handle the non-stationary and nonlinear dynamic characteristics of the UAV-to-ground control vehicle (U2V) channel, thereby enhancing the feature extraction and refinement capabilities of the STANN and improving the accuracy of CSI prediction. Additionally, we conducted a validation of the proposed EMD-STANN model across multiple datasets. The results indicated that EMD-STANN is capable of effectively adapting to diverse channel conditions and accurately predicting channel states. Compared to existing methods, EMD-STANN exhibited superior predictive performance, as indicated by its reduced root mean square error (RMSE) and mean absolute error (MAE) metrics. Specifically, EMD-STANN achieved a reduction of 24.66% in RMSE and 25.46% in MAE compared to the reference method under our simulation conditions. This improvement in prediction accuracy provides a solid foundation for the implementation of URLLC applications.

Keywords

Neural network · EMD · Channel prediction · UAV · Non-stationary

1 Introduction

1.1 Motivation

Ultra-reliable and low-latency communications (URLLC) will be the backbone of the upcoming sixth-generation (6G) systems and will facilitate mission-critical scenarios [1, 2]. Such as unmanned aerial vehicles (UAVs) control information delivery, vehicle-to-vehicle (V2V) communications, industrial automation, etc [3,4,5]. However, the stringent reliability and latency requirements for URLLC systems present significant hurdles for both industry and academia [3]. In recent years, UAV technology has garnered significant attention due to its extensive applications across military, public safety, and civilian sectors, becoming a key area for URLLC applications. Future UAV networks must ensure full coverage, strong reliability, low latency, and high security to meet the needs of various applications [6].

Nevertheless, implementing URLLC for UAVs in high-mobility application scenarios is extremely challenging. This is because acquiring channel state information (CSI) accurately and quickly is particularly problematic in UAV wireless communications, especially in high mobility application scenarios with rapidly varying channels. High mobility and multipath effects are evident in UAV wireless communication systems, making the channel highly non-stationary [7], which causes temporal variation in the channel distribution. As shown in [8], the actual channels for the second half of the slots in the same frame are likely to undergo significant changes. Therefore, the timely and accurate acquisition of CSI is crucial for achieving URLLC in high-mobility scenarios.

To guarantee the quality of CSI obtained in high-mobility application scenarios, the main technical means include: adding more reference signals [9], and adopting more advanced channel estimation or interpolation methods[8]. However, while increasing the number of reference signals can enhance the quality of CSI, it also consumes more spectral resources, which reduces spectral efficiency. Furthermore, even with advanced channel interpolation schemes, the number of reference signals still needs to increase rapidly with movement speed, which may lead to a decrease in effective data rate and additional energy consumption. Meanwhile, they typically require high computational complexity, making them unsuitable for wireless communication systems that adopt new technologies such as massive multiple-input-multiple-output (MIMO), orthogonal frequency-division multiplexing (OFDM), and millimeter wave (mmWave) [10]. To overcome the above problems, channel prediction is considered a promising method for future wireless communication systems to obtain CSI [11,12,13,14,15]. It achieves prediction of future

channel states by studying channel characteristics in both time and space domains, providing a possible solution to improve CSI quality in high mobility scenarios.

1.2 Related work

In the field of wireless communication, channel prediction technology is crucial for improving communication quality, reducing energy consumption, and enhancing network performance. Based on the fact that there is a certain temporal correlation between channel samples [16] and spatial consistency, i.e., closely located links experience similar channel statistical properties [17, 18]. Currently, there is an extensive body of literature on channel prediction research. Existing methods for wireless channel prediction can be generally divided into two categories: statistical model-based methods and artificial intelligence (AI) based data-driven methods [11, 12].

Although statistical model-based methods have theoretical advantages, their limitations are gradually becoming apparent in practical applications. Those methods rely on theoretical channel models and strict assumptions, such as quasi-static and wide-sense stationary (WSS) [15]. However, due to multipath and Doppler effects, the actual channel evolution is often very complex. This complexity makes it difficult for channel prediction methods based on statistical models to adapt to rapidly changing channels in high-mobility application scenarios, resulting in a mismatch between the model and the actual channel [8, 19].

In recent years, with the continuous development of artificial intelligence technology and its integration with wireless communication technology [20, 21], AI-based data-driven channel prediction methods have attracted considerable attention recently with their more flexible non-linear expression ability and more accurate prediction capabilities [13, 14]. For example, utilizing the robust predictive capabilities of Recurrent Neural Network (RNN), a variety of forecasting algorithms based on RNN [22] or its improved architectures such as Long Short-Term Memory (LSTM) [23,24,25,26] and Gated Recurrent Units (GRU) [27, 28] have been proposed, which leverage the possible temporal correlation between current or future channels and past channel states for prediction. Nevertheless, in dynamic communication networks, channel state is closely related to both time and the spatial environment. Although aforementioned time-series prediction methods perform well in stationary scattering environments, they may not fully account for the spatial correlation of channels, which may lead to significant deviations in prediction models due to the inability to extract spatial features.

Given the circumstances mentioned aforementioned, to obtain more channel features and improve prediction accuracy, convolutional neural networks (CNN) have been introduced into channel prediction. By relying on the feature extraction and high-dimensional information processing capabilities of CNN, further extraction of the short-term temporal relationships within the channel is achieved [14, 29], and direct processing of higher-dimensional channel data is conducted to obtain additional features [30]. Furthermore, recent studies have integrated channel spatial correlation into channel prediction models. For instance, the authors in [13] and [15, 31] have proposed a Conv-CLSTM model, leveraging the combination of CNNs and convolutional long short-term memory networks to capture both the spatial correlation and temporal dependence of CSI. Reference [16] introduces a method utilizing a spatio-temporal autoregressive model to extract the correlations between elements of the Angle Delay Channel Response Matrix (ADCRM) for channel prediction purposes.

Although the above methods can effectively predict channel states, they are mainly aimed at ground communication networks. For UAV communication channels, they may not necessarily be applicable due to the high mobility and three-dimensional deployment capabilities of UAVs. At present, research on UAV communication channels mainly focuses on channel modeling and resource optimization allocation, but there are still some positive explorations and efforts in UAV channel prediction. Such as the authors in [21] and [22] proposed UAV-to-ground control vehicle (U2V) channel prediction methods based on LSTM and RNN, respectively, to reduce the cost of channel estimation in UAV application scenarios. In [32], a learning method based on the Gaussian Process (GP) is proposed to model and predict channel fading in specific environments. In [33], the authors classify the multipath component (MPC) parameters along UAV trajectories using Euclidean distance and implement prediction based on an autoregressive model. In [34], a combined autoregressive integrated moving average model (ARIMA) and GP channel prediction method was proposed to achieve accurate and fast-tracking of channel changes. Due to the non-stationary and nonlinear characteristics of U2V channels [35], the aforementioned prediction methods still have some drawbacks. 1) Only the temporal samples of channel features are considered in the prediction, however, when the speed of the mobile terminal changes, the position of the mobile terminal also changes, which leads to the changes of channel features. Thus, the prediction of ignoring the position or speed information of the terminal is not reliable. 2) The differential method [13, 34] currently used may not effectively remove nonlinear trends, and there is also a risk of losing useful information, especially high-order information.

In summary, while AI-based prediction methods are effective for obtaining channel states, further research is needed to adapt to the non-stationary and nonlinear characteristics of U2V channels. Additionally, fully leveraging the spatio-temporal correlation of channels for more accurate predictions is essential, particularly in UAV communication systems designed for URLLC applications.

1.3 Contribution and organization

Accurate prediction of channel states is crucial for both the successful deployment and operation of URLLC in UAV communication systems. This paper aims to address the challenge of quickly and accurately obtaining the states of non-stationary U2V channels in urban scenarios, which is essential for supporting URLLC applications. Given the complex and dynamic characteristics of the U2V channel, there is a need for an effective prediction method. We introduce a novel prediction method that leverages a spatio-temporal attention neural network to forecast the trends of spatio-temporal non-stationary U2V channels. The proposed model integrates an empirical mode decomposition (EMD) module, aiming to handle the non-stationary and nonlinear dynamic characteristics of the U2V channel. This method reduces the complexity of the original data while preserving its essential features, achieving adaptive decomposition of U2V channel state characteristics and information expansion. Concurrently, we constructed a spatio-temporal neural network with an enhanced attention mechanism to better capture the diverse temporal scale features and inherent spatio-temporal dependencies of U2V channels, thereby achieving accurate prediction. Our contributions can be summarized as follows:

1) We analyzed the spatio-temporal characteristics of the U2V channel and designed an adaptive channel feature decomposition module utilizing EMD to address its non-stationary and nonlinear characteristics of the U2V channel.

2) We propose a spatio-temporal neural network learning framework that incorporates a CNN-LSTM enhanced attention mechanism to capture the spatio-temporal correlation of non-stationary U2V channels for channel prediction.

3) A non-stationary channel dataset for dynamic scenarios was generated using the ray tracing method, and the applicability of the proposed method in UAV applications was analyzed and compared with LSTM-based, CNN-based, and CNN-LSTM-based models based on this dataset.

The rest of this paper is structured as follows: Section 2 provides a detailed depiction of the communication framework utilized by UAVs in the context of URLLC applications, simultaneously analyzing the characteristics of U2V channels and formulating the channel

prediction problem. Section 3 offers a detailed description of the proposed EMD-STANN non-stationary channel prediction model, including preprocessing and learning frameworks. Subsequently, Section 4 presents the simulation results and their corresponding analysis. Finally, Section 5 provides a summary of this paper and discusses prospects for future work.

2 System model and problem formulation

2.1 Application scenarios

In this subsection, the channel model for U2V communication is presented, followed by the problem formulation for non-stationary channel prediction.

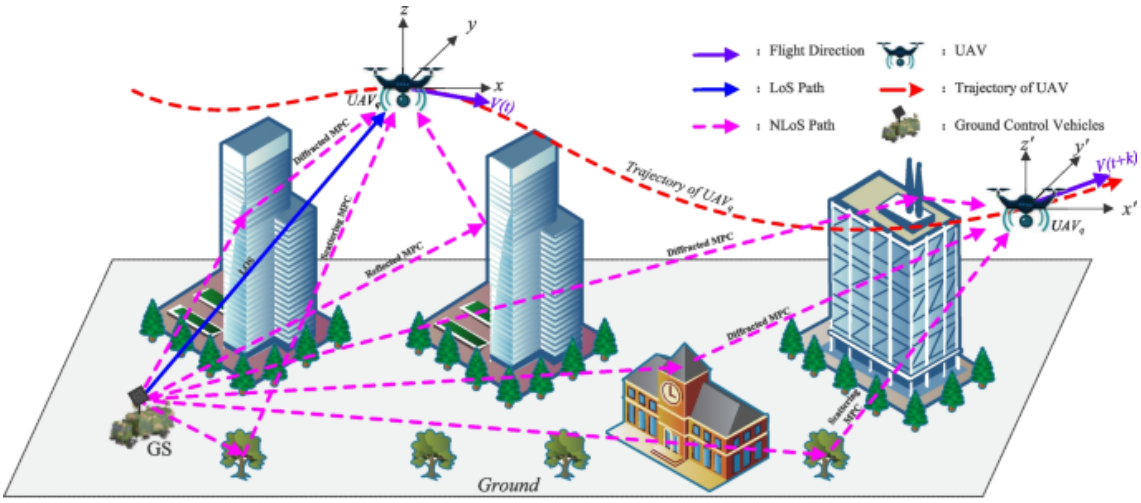


Fig. 1

UAV-enabled communication for search applications

In this paper, we consider an urban application environment where UAVs are utilized for performing latency-sensitive and reliable transmission tasks. As illustrated in Fig. 1, fixed-position ground control vehicles direct N UAVs to follow their respective 3D trajectories for performing search tasks. We assume that the UAV is equipped with an omnidirectional single antenna, while the control vehicle is equipped with an M-element uniform rectangular array (URA) antenna. Frequency division multiplexing (FDM) technology is used to exchange data and information between UAVs and the control vehicle, preventing mutual interference among different UAVs. Due to the presence of a large number of scatterers in urban environments, such as buildings and trees, the wireless propagation between UAVs and control vehicles contains a significant amount of MPCs. Moreover, the MPCs worsen the situation, as they change with the relative position

MPCs. Moreover, the MPCs worsen the situation, as they change with the relative position between the UAV and the control vehicle given the UAV's constantly shifting position in this application.

2.2 U2V channel model and characterization

As shown in Fig. 1, UAVs fly in 3D space, exhibiting varying heights and velocities. This results in time-varying multipath propagation that depends on both height and velocity [36]. The rapid movement of UAVs subjects the U2V channel to the effects of both large-scale (such as path loss and shadowing) and small-scale fading. According to research [17, 36, 37], considering large-scale factors and small-scale fading, the U2V channel matrix is given by the formula (1).

$$\mathbf{H}(t,\tau) = \sqrt{PL \times SH} \times \mathbf{H}_s(t,\tau) \tag{1}$$

Where *PL* denotes path loss, its value is determined by the carrier frequency f_c and the distance $d(t)$ between the transmitter and the receiver. *SH* denotes the shadowing, which may be caused by buildings, trees, undulating ground, or other objects completely or partially blocking the LoS path in the U2V propagation environment. Both of these factors are closely related to the current position of the UAVs. The small-scale fading is represented as a complex matrix $\mathbf{H}_s(t,\tau) = [h_{pq}(t,\tau)]_{M\times N}, p, q \in \mathbb{R}$, where $h_{pq}(t,\tau)$ is the CIR between the *p*-th Tx antenna element of the control vehicle and the *q*-th UAV at any time *t*, and expressed as the summation of the line-of-sight (LOS) and Non-line-of-sight (NLOS) components [38, 39], and can be written as the formula (2).

$$\begin{aligned} h_{pq}(t,\tau) = &\sqrt{\frac{K(t)}{K(t)+1}} h_{pq}^{LoS}(t,\tau)\,\delta(\tau-\tau_{LoS}(t)) + \\ &\sqrt{\frac{1}{K(t)+1}} \sum_{n=1}^{N(t)} \sum_{m_n=1}^{M_n(t)} h_{pq,n,m_n}^{NLoS}(t,\tau)\,\delta(\tau-\tau_n(t)-\tau_{m_n}(t)) \end{aligned} \tag{2}$$

Where the term $K(t)$ is the Rician factor, which is a measure of the strength of the direct signal in a communication system. $h_{pq}^{LoS}(t,\tau)$ and $h_{pq,n,m_n}^{NLoS}(t,\tau)$ represent the LoS and NLoS components, respectively. They are functions of the departure azimuth angle (AAoD), departure elevation angle (EAoD), arrival azimuth angle (AAoA), and arrival elevation angle (EAoA) for each propagation path. The detailed relationships can be found in [19] and will not be reiterated here. $\delta(\cdot)$ is the Dirichlet function, and $\tau_{LoS}(t)$ is the LoS delay between Tx and Rx. The variables $N(t)$ and $M_n(t)$ denote the total number of clusters and the number of rays within the *n*-th cluster, respectively. The symbol $\tau_n(t)$ represents the delay associated with the *n*-th cluster and $\tau_{m_n}(t)$ refers to the relative

delay of the m-th ray within that cluster. These parameters are all time variables as they change with the position of the UAV$_q$.

Due to the mobility of UAVs, the wireless propagation environment will present time-varying and non-stationary properties in space and time domains [19, 36]. The spatio-temporal geometric positional relationship between the UAVs and the control vehicle is shown in Fig. 2. Considering a specific UAV$_q$ and a particular cluster n to demonstrate the variations in the U2V channel. The position of the q-th UAV is a function of time, at any time t can be expressed as $P_q(t) = [x_q(t), y_q(t), z_q(t)], q \in \mathbb{R}$, where $x_q(t)$ and $y_q(t)$ represent the latitude and longitude coordinates, and $z_q(t)$ represents flight altitude, which can be obtained through onboard global positioning system (GPS) and other sensors.

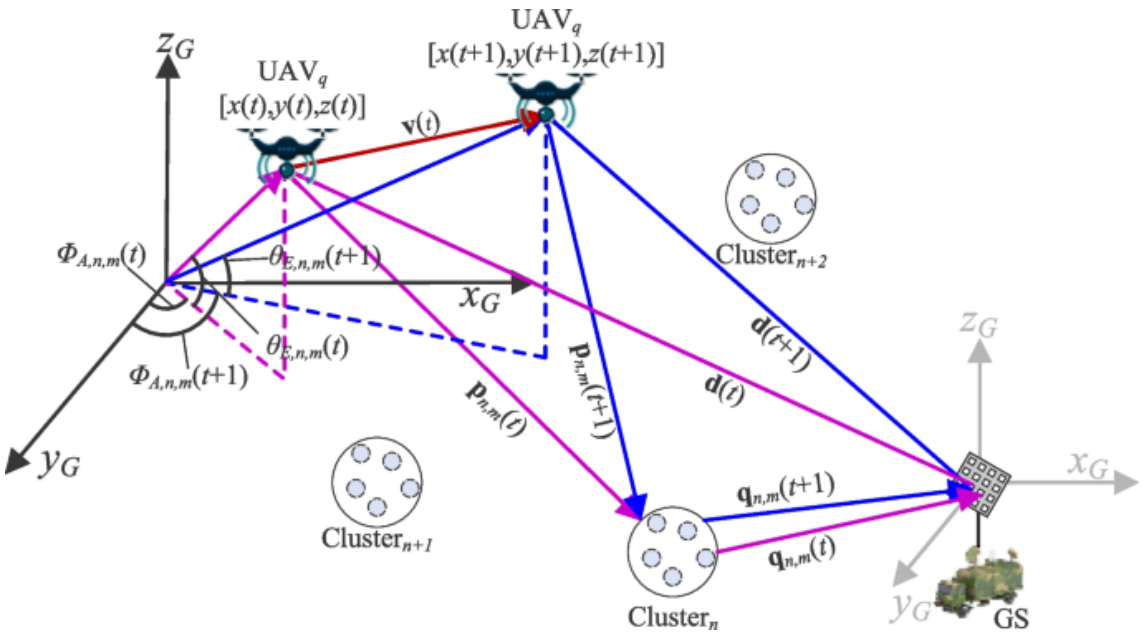


Fig 2

Illustration of the U2V channel with spatio-temporal variation process

To simplify, let's consider the m-th propagation path in cluster n as an example. At the time t, the AAoD and EAoD of the m-th ray emitted by the UAV$_q$ in cluster n are $\Phi_{A,n,m}(t)$ and $\theta_{E,n,m}(t)$, respectively. Using $d_{n,m}$ to represent the propagation length of the m-th propagation path in cluster n, can be obtained by the formula (3).

$$d_{n,m} = \|\mathbf{d}(t)\|_2 + c \times [\tau_n(t) + \tau_{m_n}(t)] \tag{3}$$

Where c represents the speed of light, and $\|\mathbf{d}(t)\|_2$ is the LOS propagation length. The arrival vector $\mathbf{p}_{n,m}(t)$ and propagation length $\|\mathbf{p}_{n,m}(t)\|_2$ from UAV$_q$ to cluster n can be calculated through the formula (4).

$$\begin{aligned} \|\mathbf{p}_{n,m}(t)\|_2 &= \frac{d_{n,m}^2 - \|\mathbf{d}(t)\|_2^2}{2\left(d_{n,m} + \mathbf{d}^T(t)\,\mathbf{p}_{n,m}(t)\right)}, \\ \mathbf{p}_{n,m}(t) &= \begin{bmatrix} \cos\Phi_{A,n,m}(t)\cos\theta_{E,n,m}(t) \\ \sin\Phi_{A,n,m}(t)\cos\theta_{E,n,m}(t) \\ \sin\theta_{E,n,m}(t) \end{bmatrix} \end{aligned} \tag{4}$$

Combining it with the position of UAV$_q$ at the time t, the corresponding position of cluster n can be determined. Then, at the time t+1, the AAOD $\Phi_{A,n,m}(t+1)$, EAoD $\theta_{E,n,m}(t+1)$, and arrival vector $\mathbf{p}_{n,m}(t+1)$ of the m-th ray can be updated according to the formulas (5),(6),and(7).

$$\begin{aligned} \mathbf{p}_{n,m}(t+1) &= -\mathbf{v}(t+1) + \mathbf{p}_{n,m}(t) \\ &= \begin{bmatrix} \cos\Phi_{A,n,m}(t)\cos\theta_{E,n,m}(t) \\ \sin\Phi_{A,n,m}(t)\cos\theta_{E,n,m}(t) \\ \sin\theta_{E,n,m}(t) \end{bmatrix} - \begin{bmatrix} x(t+1) - x(t) \\ y(t+1) - y(t) \\ z(t+1) - z(t) \end{bmatrix} \end{aligned} \tag{5}$$

$$\Phi_{A,n,m}(t+1) = \arctan\frac{x(t+1)}{y(t+1)} \tag{6}$$

$$\theta_{E,n,m}(t+1) = arc\frac{z(t+1)}{\|\mathbf{p}_{n,m}(t+1)\|_2} \tag{7}$$

Similarly, we can establish a comparable relationship between the control vehicle and cluster n. The analysis above indicates that signal propagation in the U2V channel exhibits spatio-temporal dependencies. Furthermore, literature research suggests that scatterers tend to cluster around their centers within 3D channels, allowing for spatial consistency to be achieved based on clustering positions [17]. These features enable us to use previous time snapshots to predict future channel states.

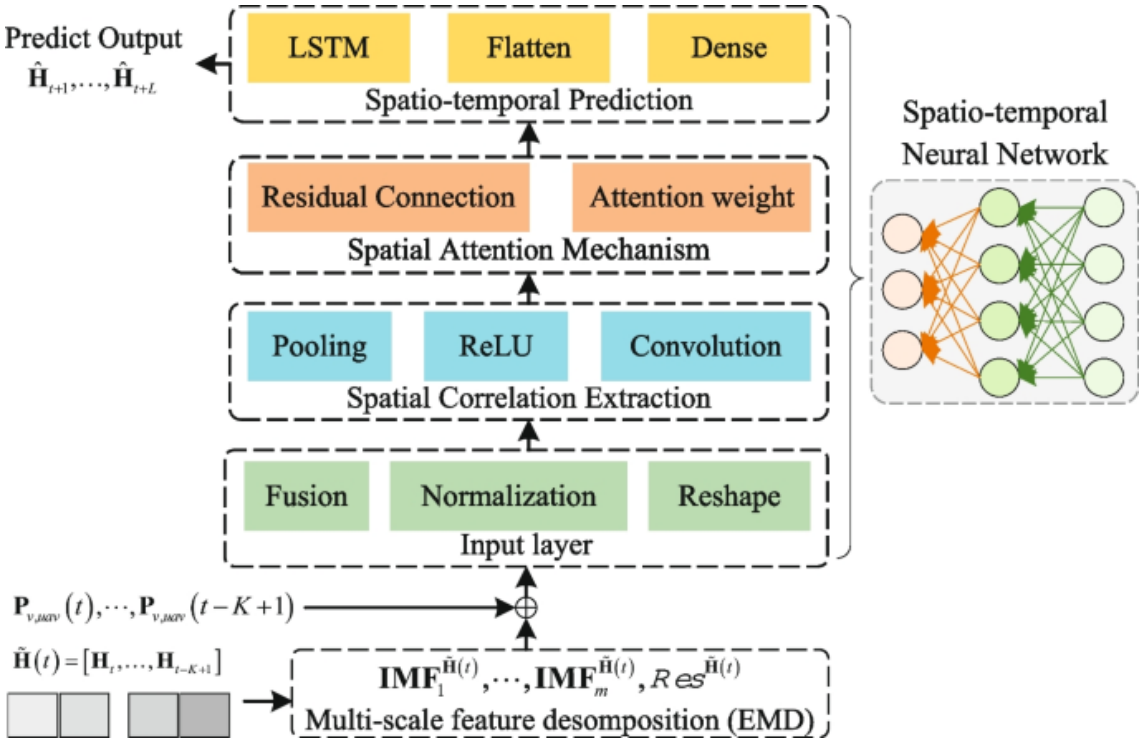


Fig. 3

The proposed predicted framework

2.3 U2V channel prediction problem formulation

To predict time-varying and non-stationary U2V channels and ensure URLLC applications in urban environments, this paper proposes a channel prediction method that utilizes historical channel data. Channel prediction is typically regarded as a time series analysis problem, where the characteristics of a channel are related to its historical state and can be described using the (8).

$$\hat{\mathbf{H}}_{t+1},\cdots,\hat{\mathbf{H}}_{t+L}=\zeta\left(\mathbf{H}_t,\cdots,\mathbf{H}_{t-K+1}\right) \tag{8}$$

Where $\hat{\mathbf{H}}_{t+i}, i\in\{1,..,L\}$ denotes the (t+i)-th CSI prediction value, $\mathbf{H}_{t-j}, j\in\{0,..,K-1\}$ refers to the historical data, with a length of K, and $\zeta(\cdot)$ represents the mapping relationship between them. According to Subsection 2.2, not only is there a temporal correlation between the future state of the U2V channel and historical data, but also a complex spatial dependency relationship, which is directly related to the relative positions of the UAVs and the ground control vehicle. It becomes imperative to simultaneously account for both spatial and temporal correlations. We use the (9) to represent the relationship between ground control vehicles and the UAV_q.

$$\mathbf{P}_{v,uav}(t) = \begin{bmatrix} x_v & y_v & z_v \\ x_q(t) & y_q(t) & z_q(t) \end{bmatrix} \tag{9}$$

Where x_v, y_v, z_v are the 3-D coordinates of the ground control vehicle, by adding these coordinates to the prediction model, the problem of time series prediction transforms into a spatio-temporal prediction problem. Consequently, the formula (8) can be rewritten as the formula (10).

$$\hat{\mathbf{H}}_{t+1}, \cdots, \hat{\mathbf{H}}_{t+L} = \Theta\left\{\mathbf{H}_t, \cdots, \mathbf{H}_{t-K+1}; \mathbf{P}_{v,uav}(t), \cdots \mathbf{P}_{v,uav}(t-K+1)\right\} \tag{10}$$

Where $\Theta(\cdot)$ represents the spatio-temporal prediction model. To achieve the target of channel prediction, the key is to determine the function $\Theta(\cdot)$. However, in high-mobility application scenarios, the non-stationary and nonlinear characteristics of U2V channels must be carefully considered.

3 Non-stationary U2V channel prediction method

In order to achieve the prediction of non-stationarity U2V channels, in this section, we propose an EMD empowered spatio-temporal neural network model. We begin by outlining the overall network structure of the proposed framework, followed by a detailed introduction to the spatio-temporal correlation extraction, the attention mechanism module, and the feature fusion prediction mechanism.

3.1 Channel spatio-temporal prediction framework

As described in Subsection 2.3, we construct the non-stationary U2V channel prediction problem as a spatio-temporal sequence prediction model. The prediction framework proposed in this article is illustrated in Fig. 3, consisting of two parts: feature decomposition and spatio-temporal neural network. The Feature decomposition is used to handle the non-stationary and nonlinear characteristics of U2V channel CSI sequences and to extract local features at multi-scales. The prediction network is composed of an input layer, a spatial feature extraction module, an attention module, a temporal dependency extraction module, and an output layer. It learns intrinsic spatio-temporal dependencies from the current input's multidimensional feature sequence and constructs a mapping model between historical and future states.

More specifically, the workflow of the proposed framework is as follows:

(i) Firstly, let $\tilde{\mathbf{H}}(t)$ represent the historical sequence $\mathbf{H}_t, \cdots, \mathbf{H}_{t-K+1}$. It is then decomposed into multi-dimensional features consisting of multiple Intrinsic Mode Functions (IMFs) with different frequencies and a residual using EMD, expressed as $\mathbf{IMF}_1^{\tilde{\mathbf{H}}}(t), \cdots, \mathbf{IMF}_m^{\tilde{\mathbf{H}}}(t), Res^{\tilde{\mathbf{H}}}(t)$.

(ii) Then, the obtained feature sequence $\mathbf{IMF}_1^{\tilde{\mathbf{H}}}(t), \cdots, \mathbf{IMF}_m^{\tilde{\mathbf{H}}}(t), Res^{\tilde{\mathbf{H}}}(t)$ is combined with the original input $\mathbf{H}_t, \cdots, \mathbf{H}_{t-K+1}$ and positional information $\mathbf{P}_{v,uav}(t), \cdots \mathbf{P}_{v,uav}(t-K+1)$ to form new input features for the spatio-temporal neural network.

(iii) Subsequently, a spatial feature extraction module, which consists of 2 CNNs, 2 activation layers (ReLU), and a pooling layer, extracts the U2V channel state space relationship from the new feature data generated in step (ii). Meanwhile, the convolutional layer encodes the data, thereby mitigating the impact of temporal variations on the data structure.

(iv) Next, a dot product attention module is used to suppress irrelevant or noisy information and capture the spatial long-term dependencies.

(v) Following this, an LSTM layer to identify and extract inherent temporal long-term dependencies in the U2V channel state feature.

(vi) Finally, the predicted output $\hat{\mathbf{H}}_{t+1}, \cdots, \hat{\mathbf{H}}_{t+L}$ is provided by a dense layer.

3.2 EMD module

EMD is an adaptive time-frequency analysis technique that excels at handling data characterized by nonlinearity and non-stationarity [40]. Unlike Fourier transforms constrained by specific basis functions, EMD does not require pre-defined basis functions and operates solely on the data's intrinsic time-scale features to perform signal decomposition. Moreover, EMD does not deviate from the time domain, ensuring that the length of the decomposed signal components remains consistent with the original signal. The EMD process involves decomposing the signal from higher to lower frequencies, yielding a finite set of IMFs and a residual component within the time-scale spectrum as the (11). Each IMF must satisfy the following criteria:

a) The difference between the number of extreme points and the number of zero crossings must be zero or one.

b) At any given time, the average of the envelopes defined by the local maximum and the local minimum must be zero.

$$\tilde{\mathbf{H}}(t)=\sum_{i=1}^{m}\mathbf{IMF}_{i}^{\tilde{\mathbf{H}}}(t)+Res^{\tilde{\mathbf{H}}}(t) \tag{11}$$

In this paper, EMD is used to preprocess the original channel state data. For a given U2V channel spatio-temporal series $\tilde{\mathbf{H}}(t)$, the specific decomposition process of EMD can be described as Algorithm 1.

1: **Input:** The original series $\tilde{\mathbf{H}}(t)$ and initialize $\Delta\tilde{\mathbf{H}}(t)=\tilde{\mathbf{H}}(t)$.
2: **while** Stopping criterion is not met **do**
Step 1: Determine the local maximum $\tilde{\mathbf{H}}_{\max}(t)$ and minimum $\tilde{\mathbf{H}}_{\min}(t)$ points of the original series $\tilde{\mathbf{H}}(t)$.
Step 2: Obtain the lower $\tilde{\mathbf{H}}_{-}(t)$ and upper envelopes $\tilde{\mathbf{H}}_{+}(t)$ of the original series $\tilde{\mathbf{H}}(t)$.
Step 3: Calculate the mean envelope $\bar{\mathbf{H}}(t)=\left[\tilde{\mathbf{H}}_{-}(t)+\tilde{\mathbf{H}}_{+}(t)\right]/2$.
Step 4: Calculate the difference $\Delta\tilde{\mathbf{H}}(t)=\tilde{\mathbf{H}}(t)+\bar{\mathbf{H}}(t)$, and residual component $\mathbf{R}_i(t)=\tilde{\mathbf{H}}(t)-\Delta\tilde{\mathbf{H}}(t)$.
3: **if** $\Delta\tilde{\mathbf{H}}(t)$ meets a) and b) **then**
4: Let $\tilde{\mathbf{H}}(t)=\mathbf{R}_i(t)$, $\mathbf{IMF}_i=\Delta\tilde{\mathbf{H}}(t)$ and go back to Step 1.
5: **else if** $R_l(t)$ is monotonic or constant **then**
6: break;
7: **else**
8: Let $\tilde{\mathbf{H}}(t)=\Delta\tilde{\mathbf{H}}(t)$, and go back to Step 1.
9: **end if**
10: **end while**
11: **Output:** m IMFs and a Res.

Algorithm 1

Empirical Mode Decomposition (EMD) Algorithm.

It can be seen from the decomposition process that EMD has the characteristics of direct, adaptive, and without prior knowledge decomposition when processing signals, and ensures that useful information in the original data is not lost. Through EMD operation, we can obtain the multi-scales feature sequence $\mathbf{IMF}_1^{\tilde{\mathbf{H}}}(t),\cdots,\mathbf{IMF}_m^{\tilde{\mathbf{H}}}(t),Res^{\tilde{\mathbf{H}}}(t)$. Then, it combines the original data and location information to form new input features, which are subsequently passed to the spatio-temporal prediction network.

3.3 Spatial correlation extraction

Spatial correlation is influenced by the distance between antennas and the characteristics of the propagation medium, as described in [13]. This paper develops a CNN-based spatial feature extractor that simplifies the network architecture, reduces the number of parameters, and uses pooling layers to reduce the dimensionality of the feature maps, effectively capturing spatial features while preserving temporal characteristics to improve prediction performance.

Let $\left\{\widehat{\mathbf{H}}_t\right\}_{t=1}^{K} \in \mathbb{R}^{n\times K}$ denotes the input features series, which is obtained by fusing the original data, EMD decomposition results, and positional information through the input layer. It is a CSI matrix that includes *K* time steps, and let *n* represent the feature dimension. Assuming that there are *r* convolution kernels, each kernel with a size of $n\times s$ and a sliding step size of 1, and the weight matrix is represented by $\mathbf{W}_{r_k} \in \mathbb{R}^{s\times n}, r_k \in \{1, 2, \cdots, r\}$. For each of the *s* time steps in the sequence vector, feature extraction is performed, yielding a feature vector that can be calculated using the formula (12).

$$\widehat{\mathbf{H}}_i^{'} = f\left(\mathbf{W}_{r_k}^{T} \otimes \widehat{\mathbf{H}}_{i:i+s-1} + b_0\right), i \in \{1, 2, \ldots, K-s+1\} \tag{12}$$

Where $f(\cdot)$ is a nonlinear activation function and a bias term is represented by $b_0 \in R$. Upon the extraction of sequence data from a sample by a convolution kernel, a feature map $\widehat{\mathbf{H}}_t^{'}$ with the shape of $(K-s+1)\times 1$ will be obtained according to the formula (13).

$$\widehat{\mathbf{H}}^{'}{}_t = \left[\widehat{H}^{'}{}_1, \widehat{H}^{'}{}_2, \ldots, \widehat{H}^{'}{}_{K-s+1}\right]^T \tag{13}$$

The input sequence data is convolved with r convolution kernels to obtain *r* feature maps. Subsequently, a pooling operation is carried out according to the formula (14), employing a pooling size of 2 and a sliding step size of 2, to obtain a three-dimensional feature tensor. This tensor is then expanded to produce a feature map with dimension $r \times [(K-s+1)/2]$.

$$\widehat{\mathbf{H}}^{''}{}_t = MaxPool\left(\widehat{H}^{'}{}_1, \widehat{H}^{'}{}_{i+1}\right), \{i = 1, 3, 5, \ldots, K-s\} \tag{14}$$

3.4 Spatial attention module

Although CNNs can extract spatial correlations, capturing long-term ones requires additional convolution operations, significantly increasing computational complexity. To address these structural limitations, this paper introduces an attention mechanism that uses concurrent weighting to maintain global context within a single layer, thereby improving efficiency [41]. The attention module captures long-term spatial dependencies by analyzing interactions across the feature map and identifying key features. Moreover, it can adaptively adjust weights based on historical data to enhance the accuracy of future channel predictions. Specifically, for early historical data that are most relevant to future status, the attention mechanism automatically assigns greater weight to these early parts, allowing the model to focus its learning on them.

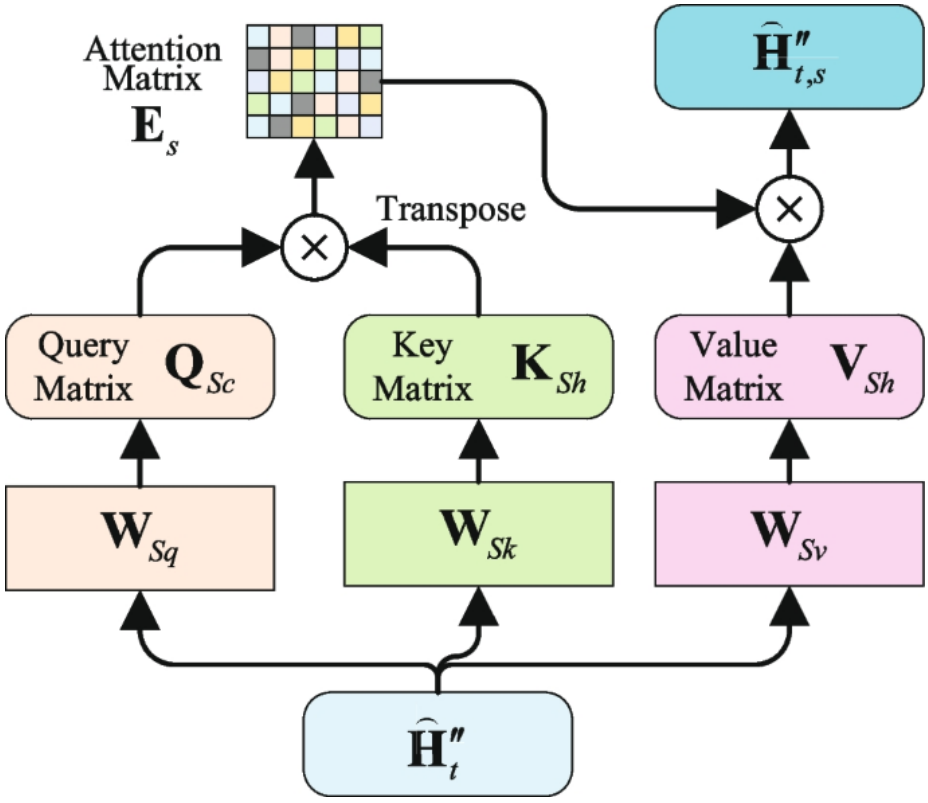


Fig. 4

The illustration of the attention mechanism

The architecture of the attention module is shown in Fig. 4. To implement the attention mechanism, a normalized attention matrix is introduced to distribute attention unevenly across inputs. The output of the module is calculated as a weighted combination of features from each sequence, with the weights indicating the attention levels assigned to those sequences. More prominent input elements are assigned greater weights, and the ultimate output is computed by allocating attention to the inputs based on the weights within the attention matrix. Specifically, the spatial features obtained by the spatial feature extractor are mapped into different feature spaces by applying three different linear transformations as formula (15).

$$\begin{aligned}
\mathbf{K}_{Sh} &= \mathbf{W}_{Sk}\widehat{\mathbf{H}}''^{T}_{t}, \in \mathbb{R}^{d_k \times r} \\
\mathbf{Q}_{Sh} &= \mathbf{W}_{Sq}\widehat{\mathbf{H}}''^{T}_{t}, \in \mathbb{R}^{d_k \times r} \\
\mathbf{V}_{Sh} &= \mathbf{W}_{Sv}\widehat{\mathbf{H}}''^{T}_{t}, \in \mathbb{R}^{r \times K'}
\end{aligned} \tag{15}$$

The matrices $\mathbf{W}_{Sq} \in \mathbb{R}^{d_k \times K'}$, $\mathbf{W}_{Sv} \in \mathbb{R}^{r \times K'}$, and $\mathbf{W}_{Sk} \in \mathbb{R}^{d_k \times K'}$ represent the respective linear transformations that are adjustable through training. Additionally, the symbols, d_k are used to signify the feature dimension of the key vector and $K' = (K - s + 1)/2$ denote the length of the input spatial feature channels. The attention module uses the scaled dot product attention, and the normalized attention weight between the *i*-th point and the *j*-th point can be calculated by the following formula (16).

$$e_{i,j} = \frac{\exp\left(\left(\mathbf{W}_{Sq}\widehat{H}''_{i}\right)^{T}\left(\mathbf{W}_{Sk}\widehat{H}''_{j}\right)\right)}{\sum_{j=1}^{K'}\exp\left(\left(\mathbf{W}_{Sq}\widehat{H}''_{i}\right)^{T}\left(\mathbf{W}_{Sk}\widehat{H}''_{j}\right)\right)}, i, j \in \{1, 2, \ldots, K'\} \tag{16}$$

Then, the attention matrix between the input is calculated by the matrix multiplication between the query matrix $\mathbf{Q}_{Sh}^{T}$ and the key matrix $\mathbf{K}_{Sh}$. As the dimensionality of the key vector's feature space grows, there is an increase in the variance of its inner product with the query matrix. Consequently, the resultant product requires a corresponding adjustment by the scaling factor $\sqrt{d}$. The attention matrix can be calculated by formula (17).

$$\mathbf{E}_S = Soft\max\left(\frac{\mathbf{Q}_{Sh}^{T}\mathbf{K}_{Sh}}{\sqrt{d}}\right) \in {}^{K' \times K'} \tag{17}$$

Every column within the attention matrix $\mathbf{E}_S$ represents a probability vector, where each element is non-negative and the sum of all elements equals one. When there is a closer match between the key vector $\mathbf{K}_{Sh}$ and the query vector $\mathbf{Q}_{Sh}$, the associated attention weight is more significant. Therefore, by summing the features across all positions, the aggregated feature $\widehat{H}''_j$, corresponding to the i-th location, can be calculated by the formula (18), each weighted by their respective attention scores.

$$\widehat{H}''_{t,i} = \sum_{j=1}^{K'} e_{i,j} \left(\mathbf{W}_{Sv} \widehat{H}''_j \right) \in \mathbb{R}^{r \times 1} \tag{18}$$

An increased attention weight $e_{i,j}$ results in a more significant contribution from the associated value vector $\mathbf{W}_{Sv}\widehat{H}''_j$ to the final output. Therefore, the feature output of the attention module can be expressed by (19).

$$\widehat{\mathbf{H}}''_{t,s} = \mathbf{V}_{Sh} Soft\max \left(\frac{\mathbf{Q}_{Sh}^T \mathbf{K}_{Sh}}{\sqrt{d}} \right) \in \mathbb{R}^{r \times K'} \tag{19}$$

3.5 Spatio-temporal prediction

Having obtained the spatial correlation characteristics of U2V channels, we further extract their temporal dependencies to achieve more accurate predictions. Unlike feedforward neural networks such as CNNs, the LSTM network is a type of RNN that incorporates feedback connections to utilize information from previous inputs. The output of an LSTM is determined by both the current input and the hidden states from prior inputs, enabling it to capture temporal correlations within a dataset. To address the challenges of gradient explosion and vanishing during training, LSTMs incorporate forgetting gates, input gates, and output gates within each unit. These gates help maintain long-term memory by selectively retaining or discarding information. Consequently, in this paper, we employ an LSTM layer to capture long-term dependencies from the output of the attention module. The mathematical expression of this calculation process can be succinctly described as follows:

$$\mathbf{i}_t = \sigma \left(\mathbf{W}_{ix} \mathbf{X}_t + \mathbf{W}_{ih} \mathbf{H'}_{t-1} + \mathbf{b}_i \right) \tag{20}$$

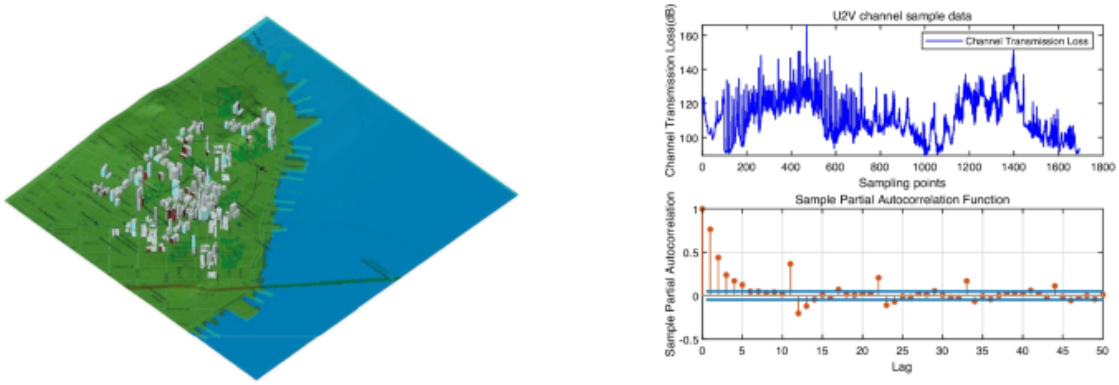
$$\mathbf{f}_t = \sigma \left(\mathbf{W}_{fx} \mathbf{X_t} + \mathbf{W}_{fh} \mathbf{H'}_{t-1} + \mathbf{b}_f \right) \tag{21}$$

$$\mathbf{o}_t = \sigma \left(\mathbf{W}_{ox} \mathbf{X_t} + \mathbf{W}_{oh} \mathbf{H'}_{t-1} + \mathbf{b}_o \right) \tag{22}$$

$$\tilde{\mathbf{c}}_t = \tanh(\mathbf{W_{cx}}\mathbf{X_t} + \mathbf{W}_{ch}\mathbf{H'}_{t-1} + \mathbf{b}_c) \tag{23}$$

$$\mathbf{c}_t = \mathbf{f}_t \otimes \mathbf{c}_{t-1} + \mathbf{i}_t \otimes \tilde{\mathbf{c}}_t \tag{24}$$

$$\mathbf{H'}_t = \mathbf{o}_t \otimes \tanh(\mathbf{c}_t) \tag{25}$$



(a) The simulation scenario (b) Sample data

Fig. 5

Simulation scenarios and sample data

Where σ represents the sigmoid function, which processes the input data $\mathbf{X}_t$ at time t, where $\mathbf{X}_t$ is the output of the attention module, i.e. $\widehat{\mathbf{H}}''_{t,s}$. The cell state at the previous time step is denoted by $\mathbf{c}_{t-1}$, and the hidden state by $\mathbf{H'}_{t-1}$. The candidate value for updating the cell state is $\tilde{\mathbf{c}}_t$, and the operator $\otimes$ denotes the Hadamard product. The forgetting gate $\mathbf{f}_t$ selects the information to be retained from $\mathbf{c}_{t-1}$, while the input gate $\mathbf{i}_t$ modulates the new input to update the cell state. The output gate $\mathbf{o}_t$ determines the portion of $\mathbf{c}_t$ that contributes to the current hidden state $\mathbf{H'}_t$.

By integrating the acquired spatio-temporal features with the original data, we can obtain richer information, which may enhance the accuracy of our predictions. The symbol $\mathbf{Z}_t$ is used to represent the aggregation of raw data and spatio-temporal features. In this paper, this aggregation is achieved through simple merging, as described by formula (26). Finally, a fully connected layer is used to derive the final predicted output, which represents the future channel states as described by the formula (27).

$$\mathbf{Z}_t = \left[\mathbf{H}_t, \widehat{\mathbf{H}}''_{t,s}\right] \tag{26}$$

$$\hat{\mathbf{H}}_{t+1} = Dense\left(\mathbf{Z}_t\right) \tag{27}$$

Table 1 Scenario Simulation Parameters

Parameter Category	Parameter name	Value	Parameter Category	Parameter name	Value
UAV	Velocity	50m/s	Scene	Geographic location	SanFrancisco
	Height	500m		Number of facets	39741
	Trajectory	Custom		Number of spheres	1
Atmosphere	Pressure(Pa)	1013.25		Rang(m)	9784*9784
	Temperature(°C)	22.2	Model	Sampling number	1695
	Humidity(%RH)	50		Sampling point spacing	2m
	Attenuation (dB/km)	0.00645		Number of paths/point	25
Receiver	Antenna model	Omni-directional		Calculated number of reflections	6
	Polarization	Vertical		Ray spacing (degrees)	0.25
	Antenna Gain	3.5dBi		Phi (degrees)	0~360
Transmitter	Transmission power	23dBm		Theta (degrees)	0~180
	Frequency	2GHz		Ray tracing algorithm	SBR
	Phase(degrees)	0		Propagation model	Full 3D

4 Simulation and numerical results

This section first describes data collection and explains the experimental parameters and evaluation metrics. Based on the collected data samples, we evaluated the performance of the proposed prediction model and compared it with existing methods: LSTM-based [23], CNN-based [29], and CNN-LSTM-based [10, 15]. Additionally, to verify the robustness of the algorithm, we conducted testing and validation on various datasets, including both authoritative public datasets and datasets from papers.

Table 2 Parameters used for EMD-STANN configuration and training

Parameters name	Value
Optimization method	Adam
Historical buffer length	5 19
Convolution kernel size	3*1
Number of convolution kernels	8
MaxEpochs	500
LearnRate	0.001
Batch size	50
Loss function	Mean Square Error

4.1 Data collection

Firstly, we generate the training and verification datasets using a 3D ray-tracing simulation software Wireless InSite (Remcom) [42]. We constructed an outdoor search application in urban shown in Fig. 5(a) and generated the non-stationary wireless channel data, scenario simulation parameters are set as shown in Table 1. Without loss of generality, we select the channel receiving transmission path loss as CSI and generate the dataset with a length of 1695 for the particular scenarios. Figure 5(b) illustrates the datasets of channel parameters, i.e. channel transmission loss, for the subchannel from the ground control vehicle antenna to the UAV antenna. We calculated the partial autocorrelation function (PACF) of the sampled data with different lags, and the results are shown in Fig. 5(b). It is not difficult to see from the figure that the data has PACF with obvious tails, further demonstrating the non-stationary characteristics of the U2V channel.

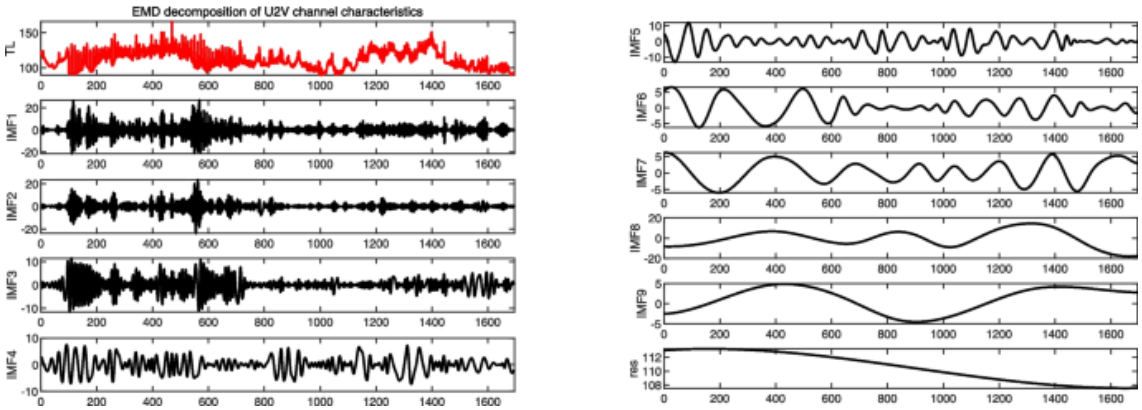
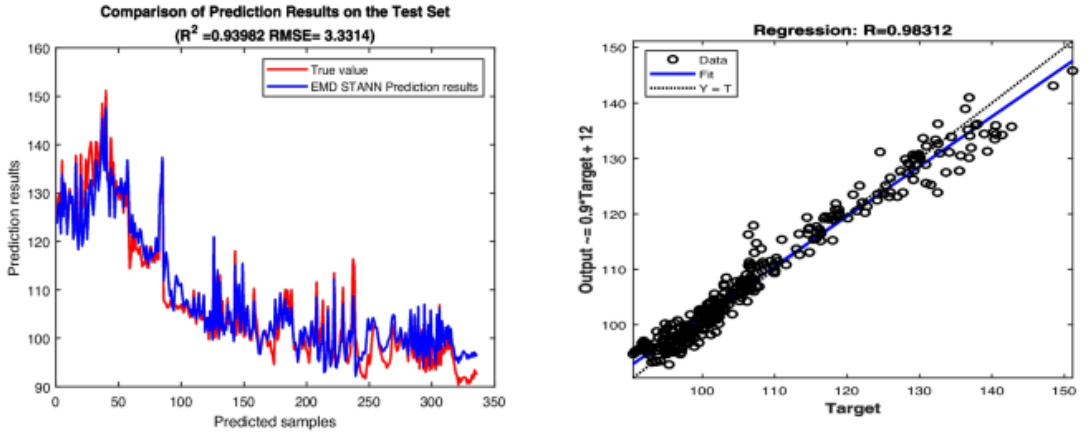


Fig. 6

EMD decomposition of U2V channel characteristics



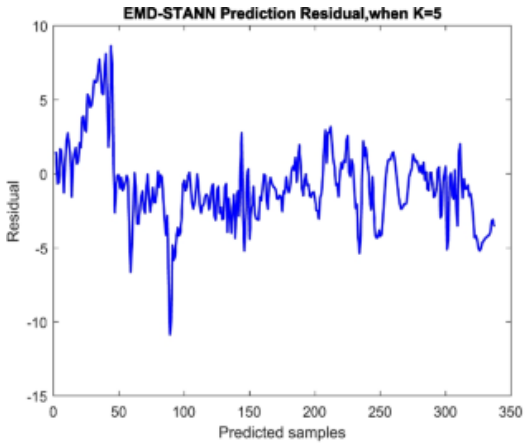
(a) Prediction results on the test dataset (b) Regression results on the test dataset

Fig. 7

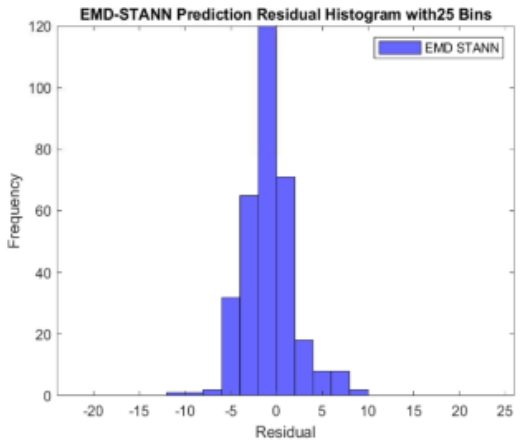
EMD-STANN prediction results, when K=5

4.2 Experimental setting

In our experiment, we introduced data into the model via data streams, dynamically creating training and validation datasets at a fixed ratio of 8:2. To train the EMD-STANN, the parameters defined in Table 2 are applied. Meanwhile, we conducted comparative experiments with the prediction methods in references [21, 29, 43].



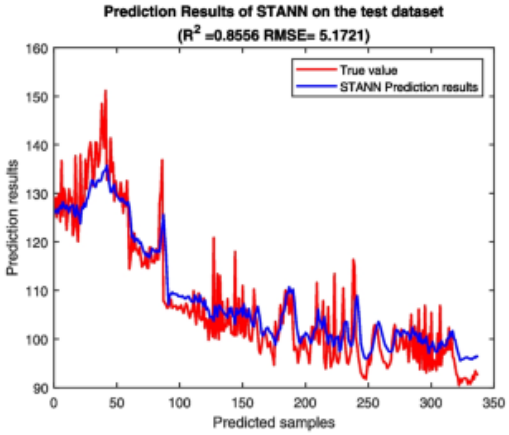
(a) Prediction residuals on the test dataset



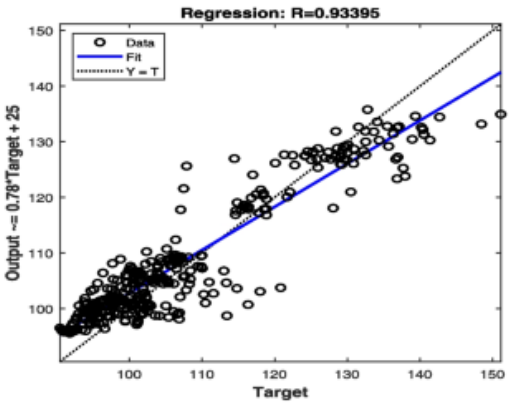
(b) Residual Histogram on the test dataset

Fig. 8

EMD-STANN prediction residuals, when K=5



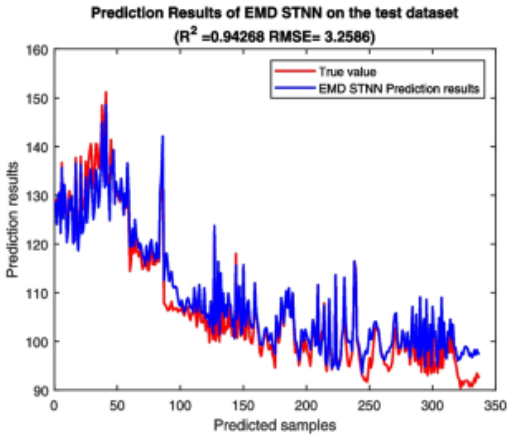
(a) Prediction results on the test dataset



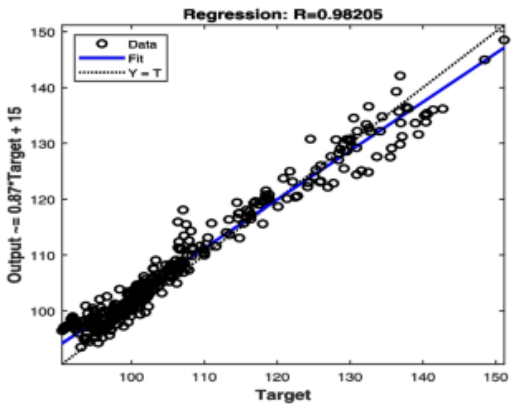
(b) Regression results on the test dataset

Fig. 9

STANN prediction results, when K=5



(a) Prediction results on the test dataset



(b) Regression results on the test dataset

Fig. 10

EMD-STNN prediction results, when K=5

To assess the effectiveness of the proposed prediction method in this work, as reference [15, 28, 29, 44], the root mean square error (RMSE) and mean absolute error (MAE) were selected as metrics to quantify the prediction performance. The corresponding calculation formulas are presented in the formula (28). Where N represents the sample length, $\mathbf{H}_{t+i}$ is a perfect estimate of the U2V non-stationary data transmission channel, feedback from the receiver, and $\hat{\mathbf{H}}_{t+i}$ is the predicted value. The lower the RMSE or MAE within the predictive model's metrics, the less the divergence among the forecasted values, indicating a higher degree of predictive accuracy.

$$\begin{aligned} RMSE &= \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left\|\mathbf{H}_{t+i}-\hat{\mathbf{H}}_{t+i}\right\|^2} \\ MAE &= \frac{1}{N}\sum_{i=1}^{N}\left|\mathbf{H}_{t+i}-\hat{\mathbf{H}}_{t+i}\right| \end{aligned} \tag{28}$$

4.3 Simulation results and analysis

In this subsection, we demonstrate the training and prediction performance of the proposed EMD-STANN prediction model based on the U2V channel data collected in Subsection 4.1. Meanwhile, the contribution of each component in the model was demonstrated through ablation experiments, further validating the rationality of the proposed model. In addition, to verify the model's superiority and robustness, it was compared with other methods and tested across various datasets.

4.3.1 Effect of the prediction

As described in Section 3, this paper employs EMD to address the nonlinear and non-stationary characteristics of the data. The EMD process decomposes the data into a series of IMFs with distinct frequency attributes, as illustrated in Fig. 6. This decomposition facilitates a more detailed analysis and understanding of the signals' internal structure. Compared to the original signal, the decomposed data exhibits stationarity, and the changing trends of certain local features are clearly highlighted.

The prediction results of the EMD-STANN model on the non-stationary U2V channel dataset are presented in Fig. 7. Figure 7(a) demonstrates that EMD-STANN effectively tracks the changes in U2V channels on the test set. Additionally, the proximity of the linear fitting line (solid blue line) to the dashed line in Fig. 7(b) confirms the model's ability to closely follow the output targets. Furthermore, from the perspective of residual prediction, the prediction error is relatively concentrated, with 95% of the errors falling between -5 and +5, as shown in Fig. 8. These results indicate that the EMD-STANN model exhibits good prediction performance.

4.3.2 Ablation study

In the proposed framework, the EMD and attention modules are pivotal components. To ascertain the indispensability of these two modules, we undertook model ablation studies. These studies were conducted with consistent parameter configurations, loss functions, and training schemes. Figure 9 presents the prediction results of the model on the U2V channel dataset following the removal of EMD. As shown in Fig. 9(a), while the model can track the general trend of the U2V channel, it loses much detailed information. The RMSE increases from 2.78 to 5.1715, and the fitting effect deteriorates, as illustrated in Fig. 9(b).

The prediction results of the model on the U2V dataset, after the attention mechanism was removed, are depicted in Fig. 10. From the prediction results in Fig. 10(a), it can be seen that while the model's predictive performance has decreased, more detailed features have been retained. This further underscores the crucial role of EMD within the model. Compared to the removal of EMD, the performance indicators decreased less, and the RMSE increased from 2.78 to 3.2586. The fitting effect also decreases synchronously, as shown in Fig. 10(b).

The performance metrics from the ablation experiments are detailed in Table 3. After removing the EMD and attention modules from the EMD-STNN model, the RMSE index increased by 46.24% and 14.67%, respectively. This indicates that the inclusion of EMD and attention modules significantly enhances the model's capacity for precise prediction of non-stationary channels.

Table 3 Ablation study for the EMD and attention sub-module, when K=5

Model	RMSE	MAE	Performance changes(RMSE)
EMD-STANN	2.7800	2.0948	–
STANN(without EMD module)	5.1715	4.0795	–46.24%
EMD-STNN(without attention module)	3.2580	2.5766	–14.67%

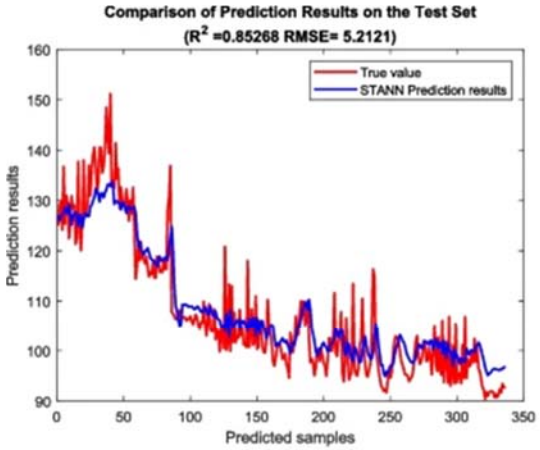
Table 4 Comparison of prediction results for different historical lengths(RMSE)

K	LSTM	CNN	CNN-LSTM	LSTM-Attn	CNN-Attn	CNN-LSTM Attn	STANN	EMD-STANN	Reduction %
5	6.7674	6.7425	7.3880	6.6540	6.8608	6.4173	5.1715	**2.7800**	**46.24%**
6	6.7887	6.7472	7.5710	6.7817	6.8264	6.3094	5.0384	**3.1364**	**37.75%**
7	6.7690	6.7768	7.4726	6.8117	6.7710	6.3932	5.1380	**3.1871**	**37.97%**
8	6.8194	6.7307	7.5645	6.6300	6.8447	6.4040	5.2127	**3.3314**	**36.09%**
9	6.8917	6.7432	7.6325	6.6716	6.7947	6.3679	5.0958	**3.4704**	**31.90%**
10	7.0856	7.2103	7.4594	7.1186	7.1402	7.0063	5.1477	**3.4502**	**32.98%**
11	7.0382	7.0236	7.5779	6.8933	7.1001	6.9894	5.5305	**4.0411**	**26.93%**
12	7.0370	7.2012	7.5739	7.1465	7.0199	7.0773	5.4013	**4.5935**	**14.96%**
13	7.2429	7.2506	7.4752	6.8863	7.2160	7.0330	5.5238	**3.3154**	**39.98%**
14	7.0796	7.2200	7.3935	7.0210	7.1957	7.0683	5.4174	**3.7179**	**31.37%**
15	7.2803	7.2336	7.5834	7.0576	7.0434	7.0821	5.4698	**4.8759**	**10.86%**
16	7.3077	7.2185	7.7600	6.9786	7.2232	7.0001	5.5882	**4.2929**	**23.18%**
17	7.4092	7.1945	7.7040	6.9419	7.0660	7.0517	5.6559	**4.9888**	**11.79%**
18	7.4759	7.2177	7.5351	7.0211	7.1268	7.1288	5.7630	**5.1001**	**11.50%**
19	7.4459	7.2584	7.6201	7.0519	7.1485	6.9474	**5.6770**	5.9610	-5.00%
20	7.1781	7.2602	7.5302	7.0786	7.2119	7.2122	5.7727	**5.4201**	**6.11%**

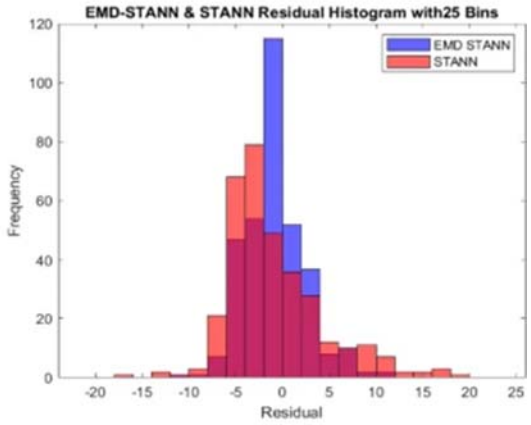
4.3.3 Performance comparison

To demonstrate the effectiveness of EMD-STANN, we compared its prediction performance with several state-of-the-art methods on U2V channels, including LSTM-based [23], CNN-based [29], and CNN-LSTM-based [10, 15]. All comparative analyses were conducted using a fixed historical data length K=5. To demonstrate the prediction performance of different methods, the results of each method are presented in separate subgraphs in Figs. 11(a) to 11(l). The comparative analysis shows that while these existing methods can capture the general trends in the U2V channel, they have limitations in effectively addressing their non-stationary characteristics. This is evident from their coefficient of determination R^2. Furthermore, residual analysis indicates that these methods have a larger prediction error range and higher RMSE values compared with the EMD-STANN method, further indicating that EMD-STANN has better performance for non-stationary U2V channels.

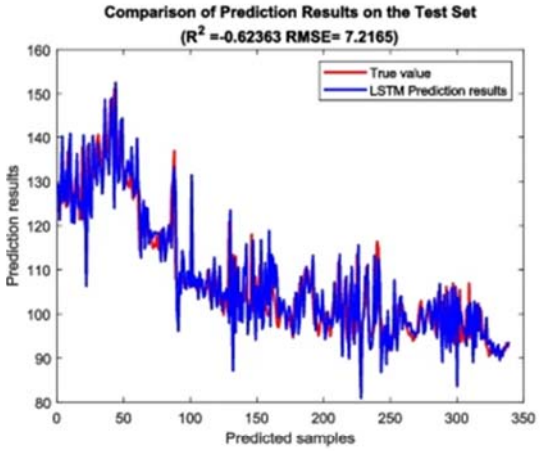
Figure 12 compares the prediction results of different methods for the next 1 to 35 time slots. It is evident from the figure that the EMD-STANN proposed in this paper outperforms the reference method, demonstrating an ability to closely track the trend of channel changes without the prediction lag observed in other methods. In addition, it is not difficult to see from the results that the prediction error of EMD-STANN increases with the increase of prediction step size. However, given the continuous variations in channel conditions, the practicality of such long-step size predictions is limited. Therefore, in practical applications, we prioritize short-term prediction results. To address this issue, with the improvement of computing power, the model can be adjusted to adapt to the dynamic changes of the channel based on re-estimation of the channel.



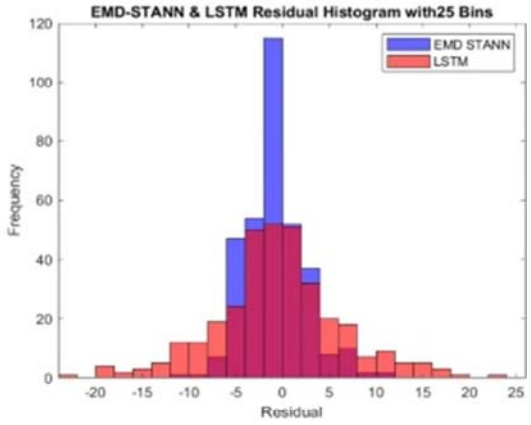
(a) STANN Predition results



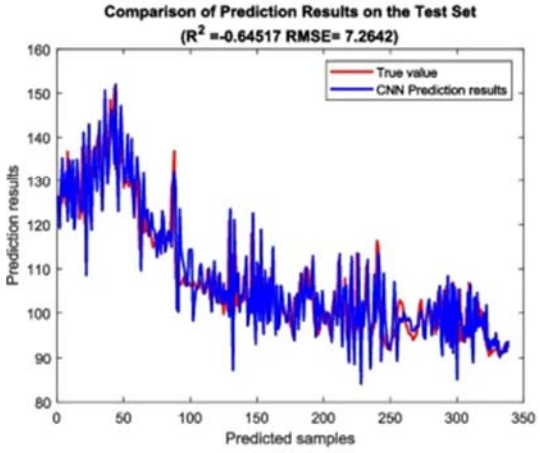
(b) STANN prediction residual



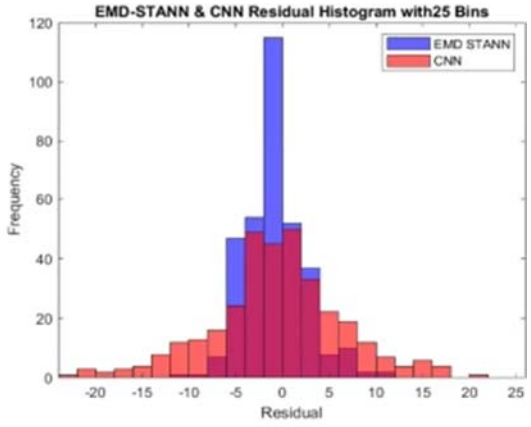
(c) LSTM Predition results



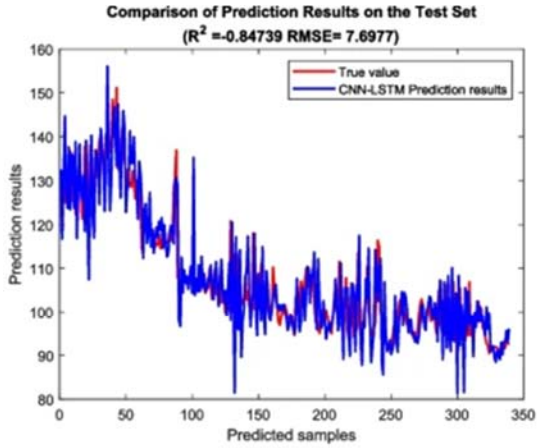
(d) LSTM prediction residual



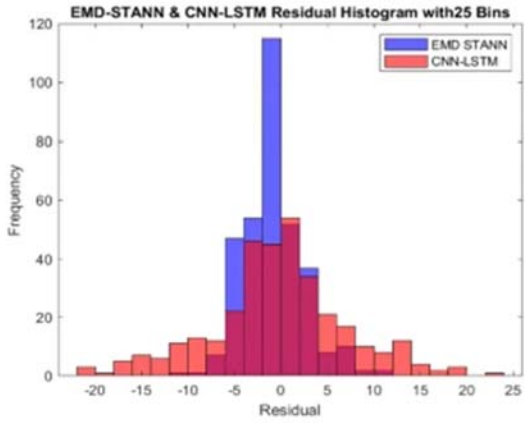
(e) CNN Predition results



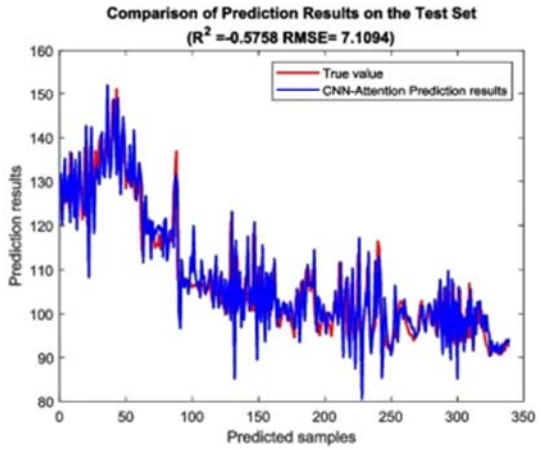
(f) CNN prediction residual



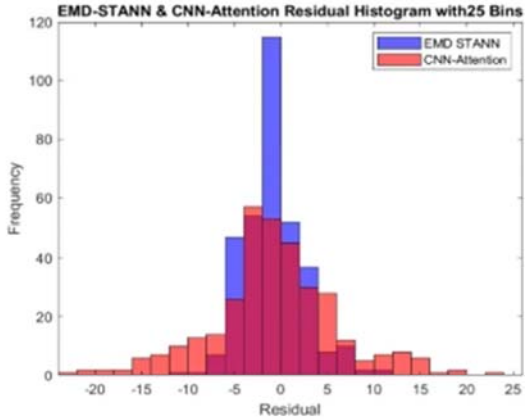
(g) CNN-LSTM Predition results



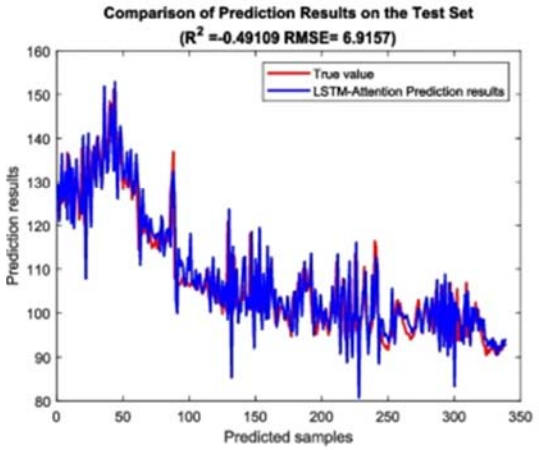
(h) CNN-LSTM prediction residual



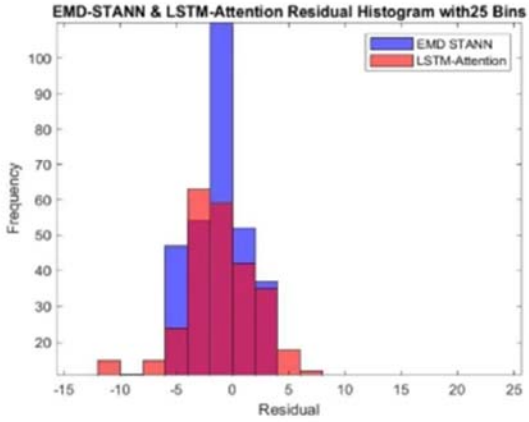
(i) CNN-Attention Predition results



(j) CNN-Attention prediction residual



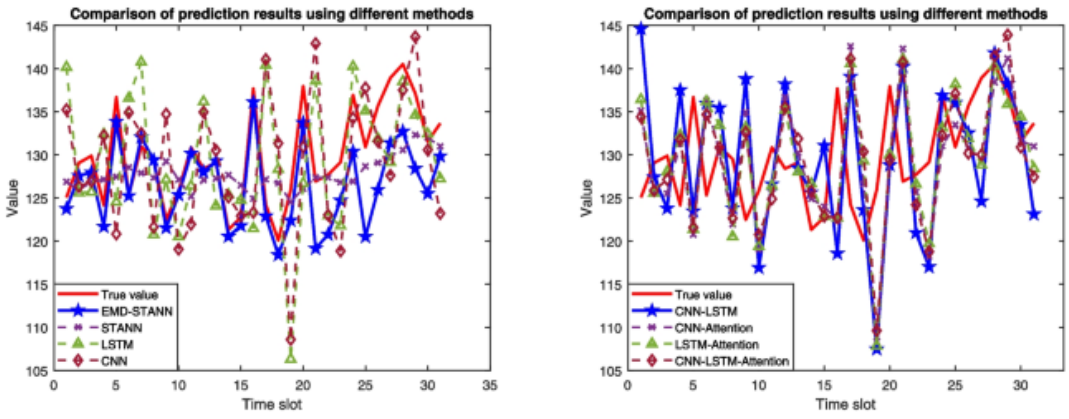
(k) LSTM-Attention Predition results



(l) LSTM-Attention prediction residual

Fig. 11

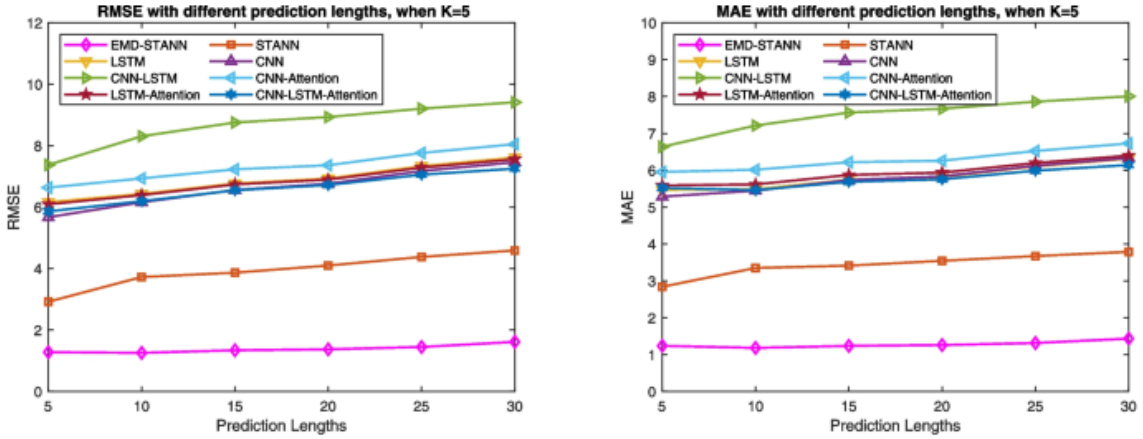
Predicted results from different methods



(a) (b)

Fig. 12

Comparison of prediction results of different methods in the next 1-35 time slots



(a) (b)

Fig. 13

Statistical analysis of RMSE and MAE for different prediction lengths

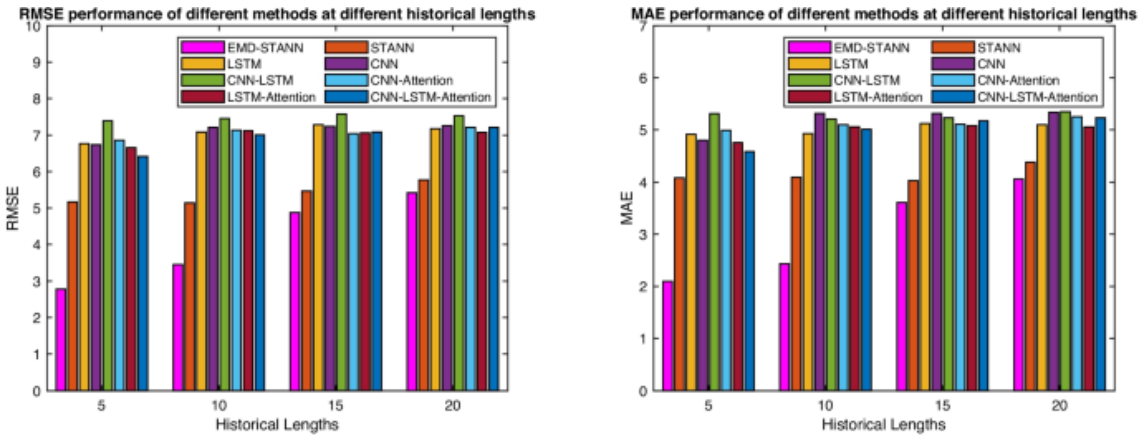


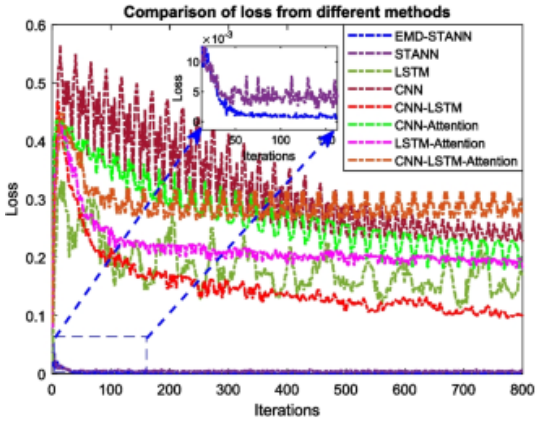
Fig. 14

Statistical analysis of RMSE and MAE for different historical lengths

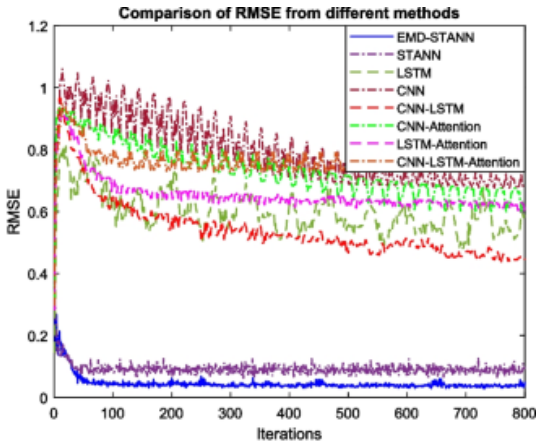
Subsequently, we conducted a statistical analysis of the predictive performance metrics RMSE and MAE for the proposed EMD-STANN and other reference methods. The statistical results are presented in Figs. 13 and 14, which respectively illustrate the performance indicators under varying prediction step sizes and historical data lengths. As shown in Fig. 13(a) and (b), the EMD-STANN method demonstrates significantly lower RMSE and MAE values compared to other reference methods across different prediction step sizes. This indicates a substantially reduced cumulative error effect for EMD-STANN. As shown in Fig. 14(a) and (b), when using historical data of different lengths for prediction, the EMD-STANN prediction method exhibits better performance than other methods. Especially when predicting with less historical data, EMD-STANN has more pronounced advantages, which is beneficial for practical applications. Furthermore, we present detailed prediction indicators in the form of tables for different historical data lengths, as shown in Tables 4 and 5, respectively, and highlight the optimal results in bold.

Table 5 Comparison of prediction results for different historical lengths(MAE)

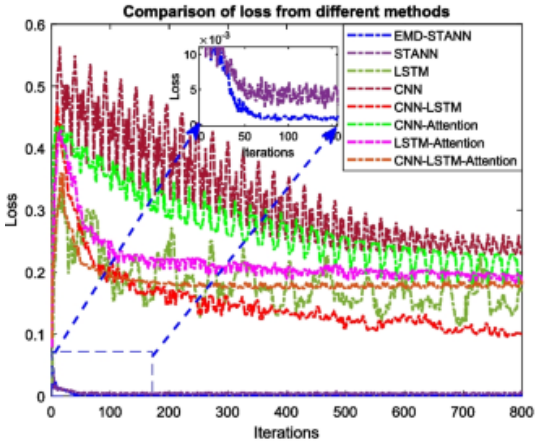
K	LSTM	CNN	CNN-LSTM	LSTM-Attn	CNN-Attn	CNN-LSTM Attn	STANN	EMD-STANN	Reduction %
5	6.7674	6.7425	7.3880	6.6540	6.8608	6.4173	5.1715	**2.7800**	**46.24%**
6	6.7887	6.7472	7.5710	6.7817	6.8264	6.3094	5.0384	**3.1364**	**37.75%**
7	6.7690	6.7768	7.4726	6.8117	6.7710	6.3932	5.1380	**3.1871**	**37.97%**
8	6.8194	6.7307	7.5645	6.6300	6.8447	6.4040	5.2127	**3.3314**	**36.09%**
9	6.8917	6.7432	7.6325	6.6716	6.7947	6.3679	5.0958	**3.4704**	**31.90%**
10	7.0856	7.2103	7.4594	7.1186	7.1402	7.0063	5.1477	**3.4502**	**32.98%**
11	7.0382	7.0236	7.5779	6.8933	7.1001	6.9894	5.5305	**4.0411**	**26.93%**
12	7.0370	7.2012	7.5739	7.1465	7.0199	7.0773	5.4013	**4.5935**	**14.96%**
13	7.2429	7.2506	7.4752	6.8863	7.2160	7.0330	5.5238	**3.3154**	**39.98%**
14	7.0796	7.2200	7.3935	7.0210	7.1957	7.0683	5.4174	**3.7179**	**31.37%**
15	7.2803	7.2336	7.5834	7.0576	7.0434	7.0821	5.4698	**4.8759**	**10.86%**
16	7.3077	7.2185	7.7600	6.9786	7.2232	7.0001	5.5882	**4.2929**	**23.18%**
17	7.4092	7.1945	7.7040	6.9419	7.0660	7.0517	5.6559	**4.9888**	**11.79%**
18	7.4759	7.2177	7.5351	7.0211	7.1268	7.1288	5.7630	**5.1001**	**11.50%**
19	7.4459	7.2584	7.6201	7.0519	7.1485	6.9474	**5.6770**	5.9610	-5.00%
20	7.1781	7.2602	7.5302	7.0786	7.2119	7.2122	5.7727	**5.4201**	**6.11%**



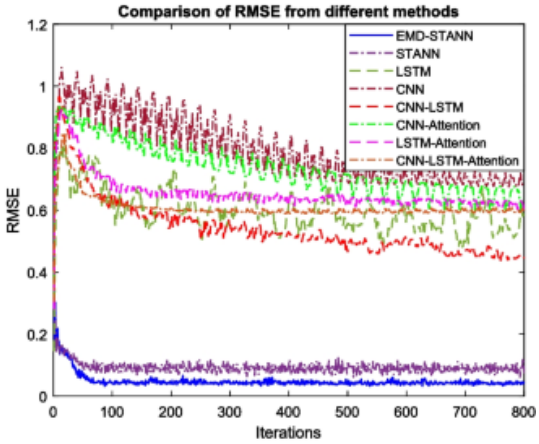
(a) Loss function with K=5



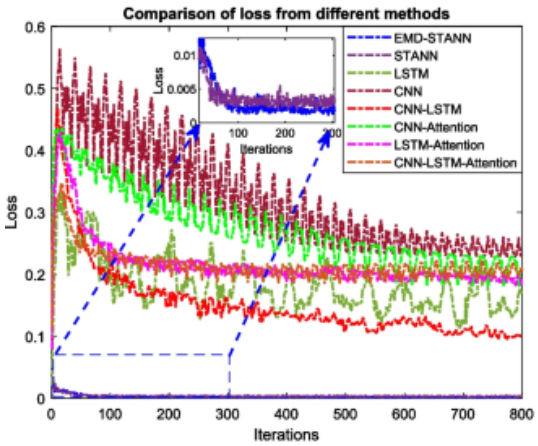
(b) RMSE with K=5



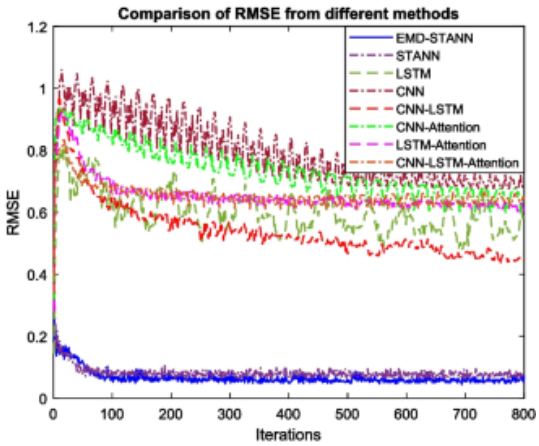
(c) Loss function with K=10



(d) RMSE with K=10



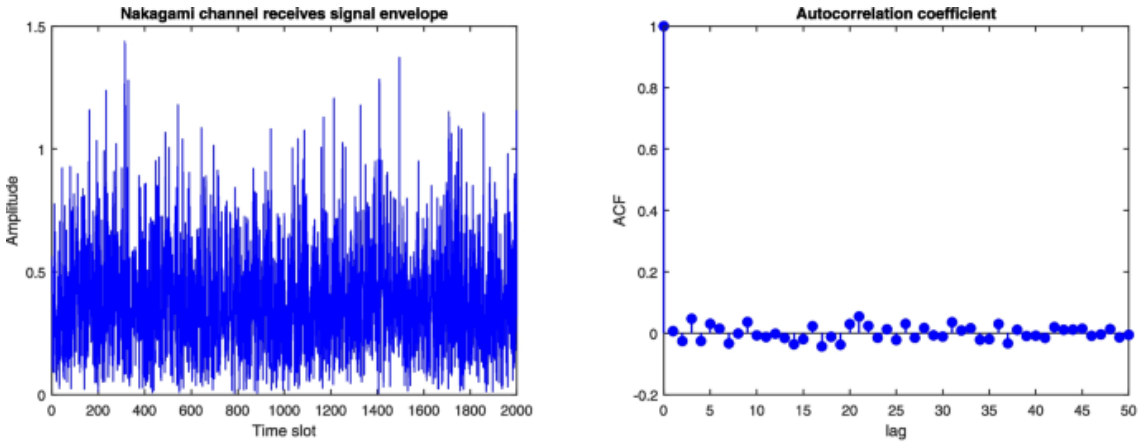
(e) Loss function with K=20



(f) RMSE with K=20

Fig. 15

The loss function and root mean square error of channel prediction methods with a historical length of K=5,10,20



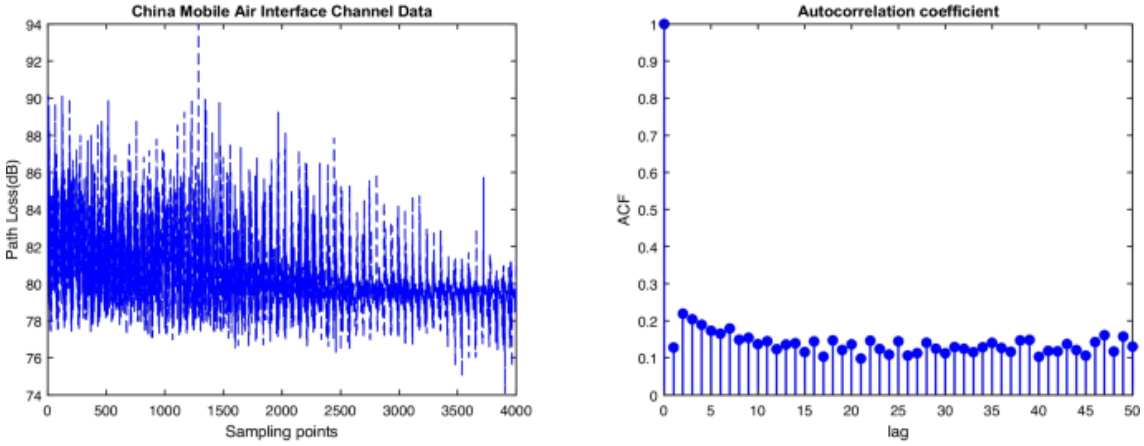
(a) (b)

Fig 16

Nakagami channel dataset

In addition, we compared the convergence performance of different methods during the training process, including loss values and prediction metrics. Figure 15 illustrates the convergence for historical data lengths of 5, 10, and 20. The training parameters for the comparative methods were set based on the literature, whereas EMD-STANN and STANN were trained using the parameters listed in Table 2. It is evident from Fig. 15(a), (c), and (e) that, compared to other methods, the EMD-STANN and STANN models achieve faster loss convergence within 50 to 100 iterations. Furthermore, the locally enlarged areas reveal that EMD-STANN demonstrates better performance and minimal post-convergence fluctuations. Meanwhile, the convergence of RMSE exhibits similar performance, as shown in Fig. 15(b), (d), and (f) .

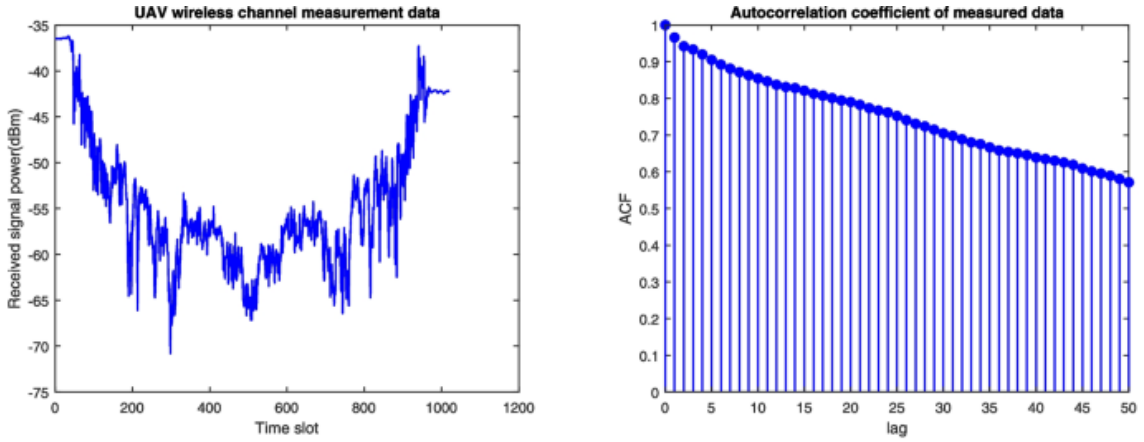
The analysis shows that the EMD-STANN model proposed in this paper outperforms others, as effective channel feature decomposition boosts prediction accuracy for non-stationary channels.



(a) (b)

Fig. 17

China mobile air interface channel dataset



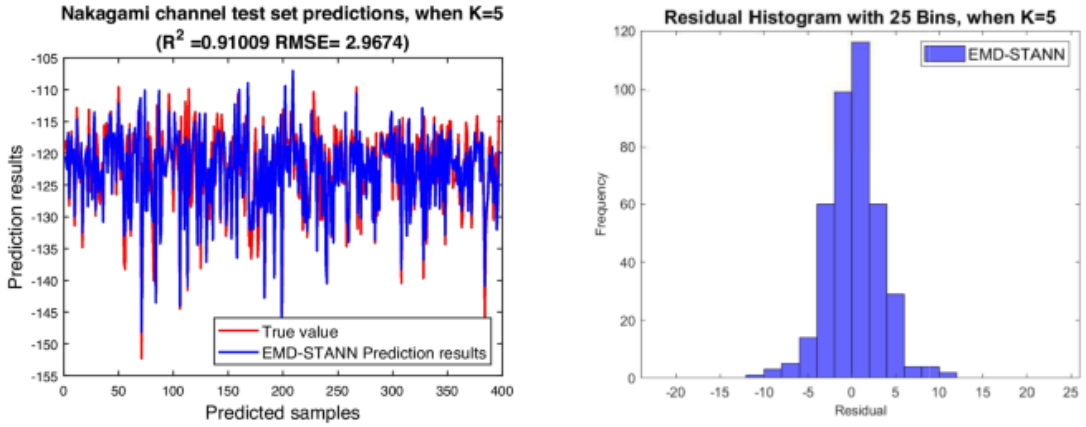
(a) (b)

Fig. 18

Measurement dataset

4.3.4 Robustness verification

To further validate the effectiveness of the proposed model, we conducted a comprehensive evaluation of its performance across three distinct types of datasets: a simulated Nakagami channel dataset, a certified air interface dataset publicly provided by China Mobile [45], and an actual test dataset from a campus environment. Specifically, for the campus environment dataset, the AD9361-based Universal Software Radio Peripheral (USRP) is mounted on the DJI drone as the transmitter. Meanwhile, the 4051E spectrum analyzer is used as the ground receiver to collect the power of the USRP transmission signals. The time-domain waveforms and ACF analysis results for the three datasets are displayed in Figs. 16, 17 and 18. The ACF analysis results reveal that all datasets exhibit non-stationary characteristics and strong temporal correlations.

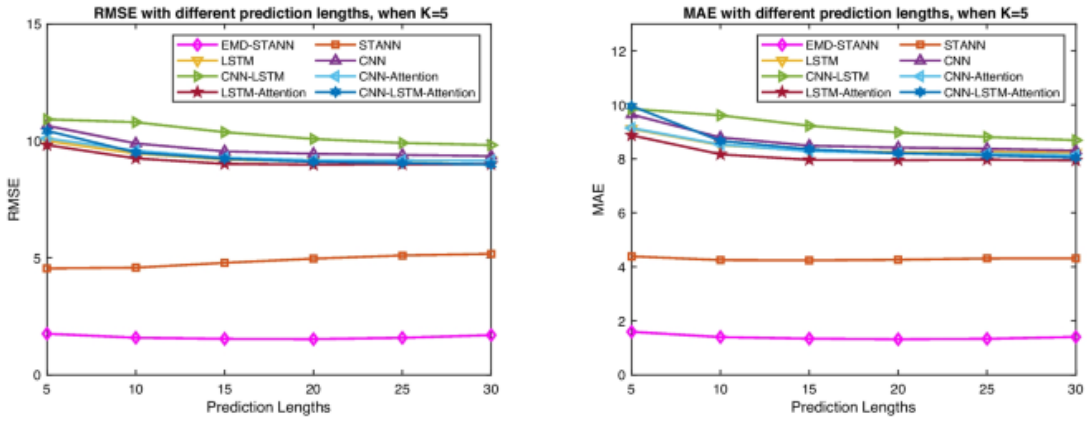


(a) (b)

Fig. 19

Prediction results of Nakagami channel dataset

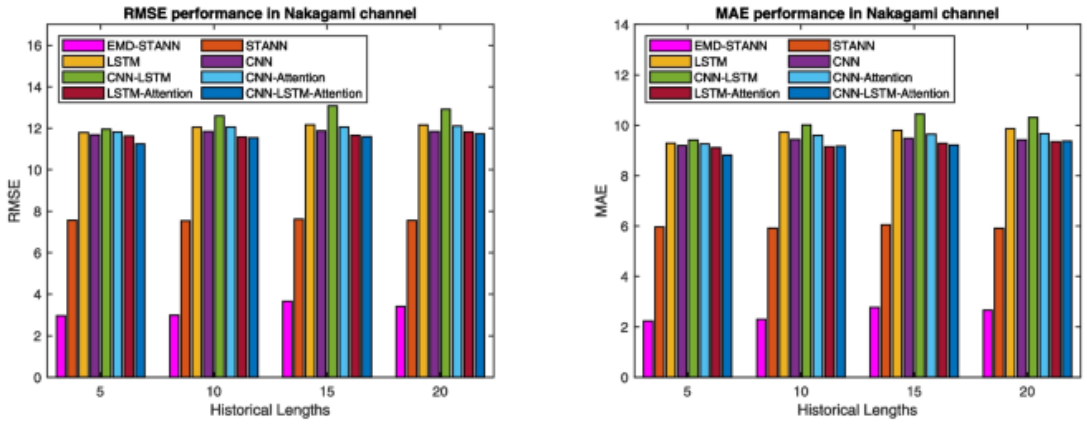
Figures 19, 20 and 21 illustrate the predictive performance of the EMD-STANN method on the Nakagami channel dataset. This dataset exhibits stationary characteristics, and according to the prediction results, the EMD-STANN method proposed in this paper is still applicable and has better prediction performance than other methods.



(a) (b)

Fig. 20

RMSE and MAE statistical analysis of different prediction lengths for the Nakagami channel dataset



(a) (b)

Fig. 21

RMSE and MAE statistical analysis of different historical lengths for the Nakagami channel dataset

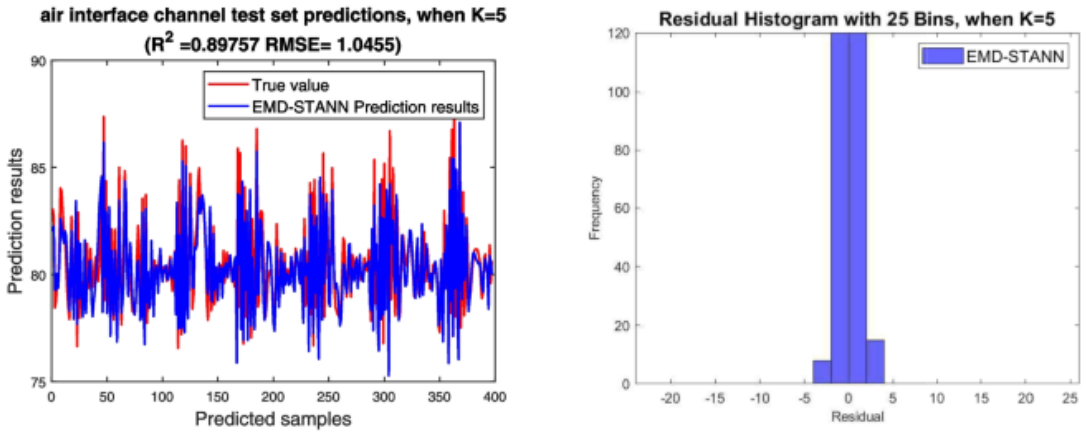


Fig. 22

Prediction results of the China Mobile Air Interface channel dataset

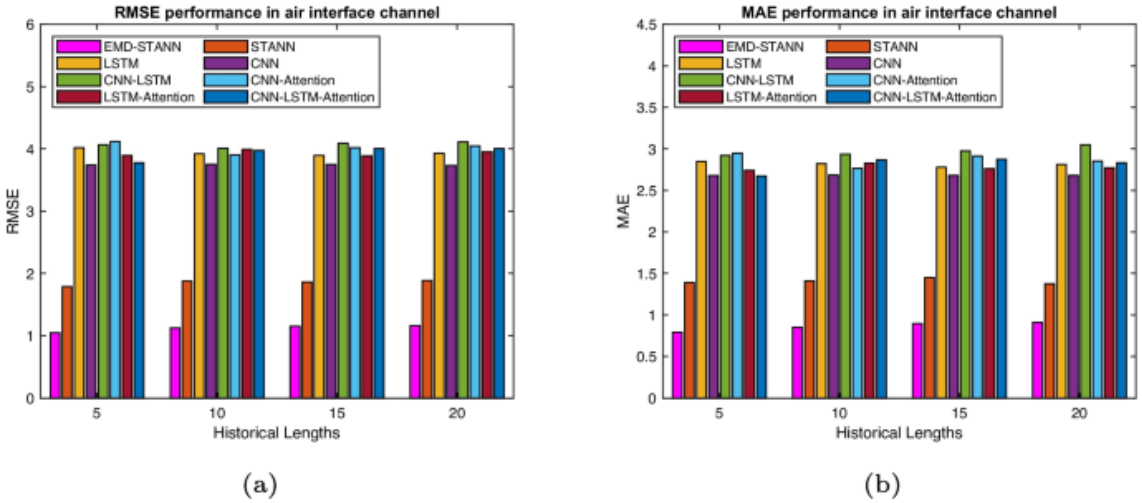
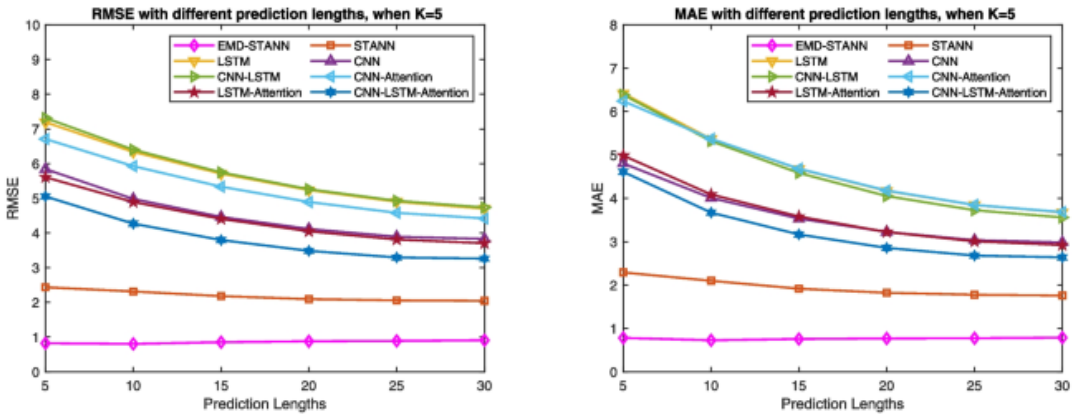


Fig. 23

RMSE and MAE statistical analysis of different prediction lengths for the air interface channel dataset

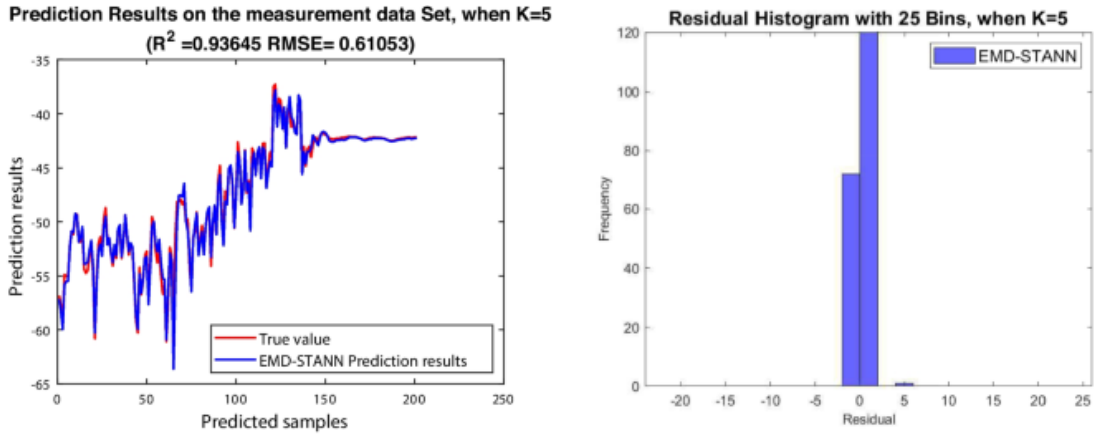
Figures 22, 23 and 24 illustrate the predictive performance of the EMD-STANN method on the China mobile air interface channel dataset. According to the statistical comparison of various prediction indicators, it is evident that the EMD-STANN method proposed in this paper outperforms other methods.



(a) (b)

Fig. 24

RMSE and MAE statistical analysis of different historical lengths for the air interface channel dataset



(a) (b)

Fig. 25

Prediction results of measurement dataset

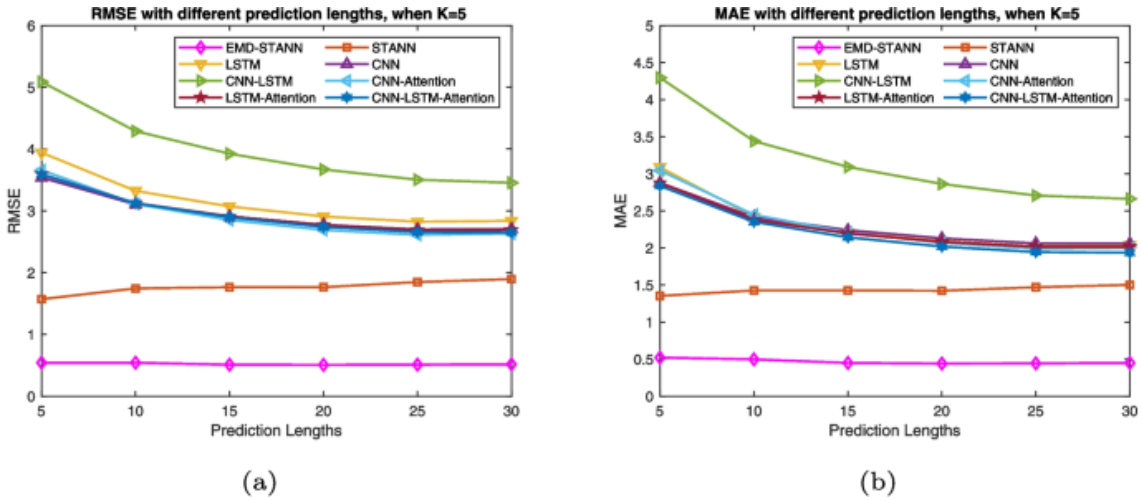


Fig. 26

RMSE and MAE statistical analysis of different prediction lengths for measurement dataset

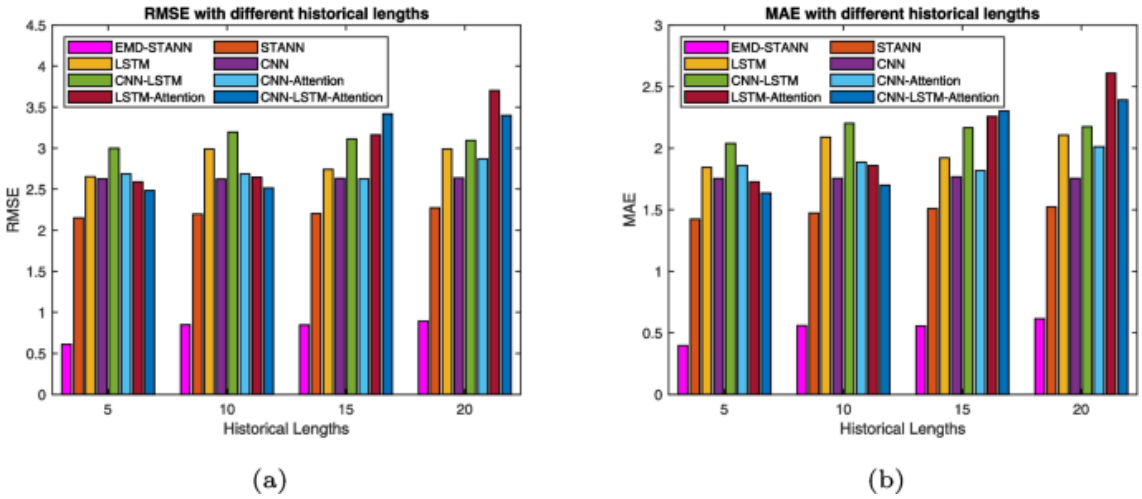


Fig. 27.

RMSE and MAE statistical analysis of different historical lengths for measurement dataset

Figures 25, 26 and 27 illustrate the prediction performance of the EMD-STANN method on the measured dataset. The measured dataset exhibits a high degree of long-term persistence, with autocorrelations that decay slowly over time, as shown in Fig. 18(b). In this situation, according to the statistical results of various prediction indicators, the EMD-STANN method proposed in this paper still demonstrates better performance in practical environments.

Based on the prediction performance across various datasets and the comparison results with other methods, we believe that the EMD-STANN method demonstrates greater robustness and better adaptability to different channels.

4.4 Complexity analysis

In this subsection, we will analyze the computational complexity of the proposed prediction scheme. As previously described, the prediction method proposed in this paper mainly consists of 1 EMD module, 2 CNN layers, 1 LSTM, 1 Attention module, and 1 output layer, which are cascaded with each other. Therefore, the overall complexity can be expressed as the sum of the complexities of each component by the formula(29).

$$\begin{aligned}&\mathcal{O}\left(N_{EMD-STANN}\right)=\mathcal{O}\left(N_{EMD}\right)+\mathcal{O}\left(2*N_{CNN}\right)+\\&\mathcal{O}\left(N_{LSTM}\right)+\mathcal{O}\left(N_{Attention}\right)+\mathcal{O}\left(N_{Dense}\right)\end{aligned}\tag{29}$$

In the worst-case scenario, if each IMF requires multiple iterations to satisfy the condition, then for a signal containing N data points, each iteration requires $\mathcal{O}\left(N\right)$time to find the extremum point and perform interpolation. Therefore, the total time complexity can be approximated as $\mathcal{O}\left(kN\right)$, where k is the average number of iterations required for each IMF in the decomposition process. The time complexity of a single convolutional layer is related to the size of the input features, the number of channels, and the size of the convolution kernel. Specifically, time complexity can be expressed as $\mathcal{O}\left(\sum_{l=1}^{D}M_l^2\times K_l^2\times C_{l-1}\times C_l\right)$, where M_l represents the edge length of the l-th layer output feature map, K_l is the edge length of the l-th layer convolution kernel, C_{l-1} and C_l indicates the number of input and output channels of the l-th layer, respectively. The core of the attention mechanism is to obtain the weighted representation of each element in the sequence by calculating the relationship between query, key, and value. In the model, it is necessary to calculate the correlation between each element in the sequence and all other elements in the sequence, so its complexity is $\mathcal{O}\left(n^{2d}\right)$, where n is the length of the sequence and d is the feature dimension of each element. The time complexity of the LSTM layer is $\mathcal{O}\left(m^2\right)$, where m is the hidden layer dimension of LSTM units. The complexity of a fully connected output layer is related to its input-output length, which can be expressed as $\mathcal{O}\left(k_{in}\times k_{out}\right)$. Then, the time complexity can be expressed as $\mathcal{O}\left(kN\right)+\mathcal{O}\left(\sum_{l=1}^{D}M_l^2\times K_l^2\times C_{l-1}\times C_l\right)+\mathcal{O}\left(n^{2d}\right)+\mathcal{O}\left(m^2\right)+\mathcal{O}\left(k_{in}\times k_{out}\right)$. Based on the previous parameter settings, the time complexity can be calculated using the formula (30).

$$\begin{aligned}\mathcal{O} &= \mathcal{O}(9N) + \mathcal{O}(K \times 3 \times 1 \times 32 + (K-2) \times 3 \times 1 \times 64) \\ &\quad + \mathcal{O}\left((K-4)^2\right) + \mathcal{O}\left(64^2\right) + \mathcal{O}\left(32^2\right) \\ &\quad + \mathcal{O}(32 + 64 \times (K-2)/2) \\ &= \mathcal{O}(9N) + \mathcal{O}(288K - 384) + \mathcal{O}\left(K^2 - 8K + 16\right) \\ &\quad + \mathcal{O}\left(64^2\right) + \mathcal{O}\left(32^2\right) + \mathcal{O}(32K - 32)\end{aligned} \tag{30}$$

We compared the complexity with other methods, including training time, and testing time, as shown in Table 6. These results were obtained using the Intel (R) Core (TM) i7-9700K and NVIDIA GeForce RTX 2070 computing platform. In terms of prediction time, EMD-STANN exhibits relatively high complexity, with the time required for single-step prediction being approximately 0.135 milliseconds. However, EMD-STANN significantly outperforms other methods in enhancing prediction accuracy, which can provide reliable support for the communication strategies necessary to implement URLLC, such as beamforming, transmission power adjustment, and bandwidth allocation. With improvements in computing power, mobile terminals utilizing higher-performance platforms or parallel computing technologies, such as NVIDIA Jetson Xavier NX and Huawei Atlas 200, are expected to reduce prediction times further to meet the 3GPP R16 specification's URLLC latency requirements of between 0.5 and 1 millisecond.

Table 6 Comparison of time cost for different methods

Method	Training stage	Testing stage
EMD-STANN	119.4074s	1.3463e-4s
STANN	98.1710s	9.5071e-5s
LSTM	133.9242s	5.7735e-5s
CNN	40.5974s	1.2808e-4s
CNN-LSTM	201.2440s	9.3501e-5s
CNN-Attention	132.6971s	7.2763e-5s

5 Conclusion

In this paper, we proposed an EMD-STANN model to enhance the prediction accuracy of spatio-temporal non-stationary channels in high-mobility UAV scenarios, which is crucial for the implementation of URLLC applications. Research establishes a mathematical model by analyzing channel characteristics and uses EMD modules to handle non-stationary and nonlinear problems in the U2V channel. Furthermore, the deep neural network model combining LSTM, CNN, and attention mechanism effectively captures the spatio-temporal

correlation of channel features, improving prediction accuracy. The effectiveness of EMD-STANN has been validated through extensive testing on various datasets and comparisons with existing methods.

Despite these advancements, EMD-STANN faces challenges in unknown application environments due to its reliance on supervised learning and the need for historical data. To address this, future work will focus on researching model adaptation methods, such as adjusting model structures or parameters to fit various application scenarios. Concurrently, we will explore the use of model compression, pruning techniques, and parallel computing to optimize algorithms, thereby reducing computational complexity and improving prediction efficiency to meet the needs of practical applications.

Data Availability

The datasets in this study have been stored on GitHub and can be accessed through the following link: https://github.com/zyq5258/UAV-non-stationary-channel. Any interested researcher can access the data while complying with the corresponding terms. Further information can be obtained by contacting the first author via email for assistance (zhangqiuyun@swust.edu.cn). The data of this study will be stored in the above link for a long time and updated regularly as needed.

Contributions

All authors contributed to the study's conception and design. Qiuyun Zhang played a key role by conceptualizing the study, designing experiments, analyzing data, and writing the initial manuscript. Qiumei Guo conducted the investigations and provided valuable resources for the study. She also participated in experimental design and data analysis. Hong Jiang provided overall guidance and support throughout the project, including overseeing the research process, as well as revising the manuscript. Xinfan Yin guided unmanned aerial vehicle simulation experiments and flight tests. Mushtaq Muhammad Umer assists with paper writing, proofreading, and formatting. Ying Luo provided partial funding and guidance. Chun Wu contributed to the visualization of the data. All authors reviewed and contributed to the final manuscript, providing critical feedback and revisions.

Competing Interests

The authors have no competing interests to declare that are relevant to the content of this article.

Acknowledgements

This work was supported in part by the Natural Science Foundation of Sichuan Province (Grant No. 2023NSFSC1373) and in part by the Sichuan Science and Technology Program (Grant No. 2024NSFSC0476).

References

1. Ranjha A, Javed MA, Srivastava G et al (2023) Intercell interference coordination for uav enabled urllc with perfect/imperfect csi using cognitive radio. IEEE Open J Commun Soc 4:197–208. https://doi.org/10.1109/ojcoms.2022.3232888

2. Ranjha A, Javed MA, Piran MJ, et al (2024) Toward facilitating power efficient urllc systems in uav networks under jittering. pp 3031–3041, https://doi.org/10.1109/TCE.2023.3305550

3. Ranjha A, Kaddoum G, Dev K (2022) Facilitating urllc in uav-assisted relay systems with multiple-mobile robots for 6g networks: A prospective of agriculture 4.0. IEEE Trans Indust Inf 18(7):4954–4965. https://doi.org/10.1109/TII.2021.3131608

4. Chen K, Wang Y, Fei Z, et al (2020) Power limited ultra-reliable and low-latency communication in uav-enabled iot networks. In: 2020 IEEE Wireless Communications and Networking Conference (WCNC), pp 1–6,https://doi.org/10.1109/WCNC45663.2020.9120565

5. Ranjha A, Kaddoum G (2021) Urllc facilitated by mobile uav relay and ris: A joint design of passive beamforming, blocklength, and uav positioning. IEEE Int Things J 8(6):4618–462. https://doi.org/10.1109/JIOT.2020.3027149

6. Cai Y, Jiang X, Liu M et al (2022) Resource allocation for urllc-oriented two-way uav relaying. IEEE Trans Veh Technol 71(3):3344–334. https://doi.org/10.1109/TVT.2022.3143174

7. Zou Z, Careem M, Dutta A, et al (2023) Joint spatio-temporal precoding for practical non-stationary wireless channels. IEEE Trans Commun pp 1–1. https://doi.org/10.1109/tcomm.2023.3241326

8. Jiang H, Cui M, Ng DWK et al (2022) Accurate channel prediction based on transformer: Making mobility negligible. IEEE J Select Areas in Commun 40(9):2717–2732. https://doi.org/10.1109/JSAC.2022.3191334

9. Chen S, Hu J, Shi Y et al (2020) A vision of c-v2x: Technologies, field testing, and challenges with chinese development. IEEE Int Things J 7(5):3872–3881

10. Luo C, Ji J, Wang Q et al (2018) Channel state information prediction for 5g wireless communications: A deep learning approach. IEEE Trans Netw Sci Eng 7(1):227–236

11. Peng F, Zhang S, Jiang Z et al (2023) A novel mobility induced channel prediction mechanism for vehicular communications. IEEE Trans Wire Commun 22(5):3488–3502. https://doi.org/10.1109/TWC.2022.3219052

12. Jiang W, Schotten HD (2019) Neural network-based fading channel prediction: A comprehensive overview. IEEE Access 7:118112–118124

13. Zhou T, Zhang H, Ai B et al (2022) Deep-learning-based spatial–temporal channel prediction for smart high-speed railway communication networks. IEEE Trans Wire Commun 21(7):5333–5345. https://doi.org/10.1109/twc.2021.3139384

14. Shao Y, Zhao MM, Li L, et al (2022) Deep learning based channel prediction for ofdm systems under double-selective fading channels. In: 2022 International Symposium on Wireless Communication Systems (ISWCS). IEEE, pp 1–6

17. Bian J, Wang CX, Gao X et al (2021) A general 3d non-stationary wireless channel model for 5g and beyond. IEEE Trans Wire Commun 20(5):3211–3224. https://doi.org/10.1109/twc.2020.3047973

15. Liu G, Hu Z, Wang L et al (2022) Spatio-temporal neural network for channel prediction in massive mimo-ofdm systems. IEEE Trans Commun 70(12):8003–8016. https://doi.org/10.1109/tcomm.2022.3215198

16. Wu C, Yi X, Zhu Y et al (2021) Channel prediction in high-mobility massive mimo: From spatio-temporal autoregression to deep learning. IEEE J Select Areas in Commun 39(7):1915–1930. https://doi.org/10.1109/JSAC.2021.3078503

17. Bian J, Wang CX, Gao X et al (2021) A general 3d non-stationary wireless channel model for 5g and beyond. IEEE Trans Wire Commun 20(5):3211–3224. https://doi.org/10.1109/twc.2020.3047973

18. Lyu Y, Liang C, Chen J, et al (2024) Channel measurements and analysis for fixed-wing uav-to-vehicle communications at 2.7 ghz in rural area. IEEE Antenna Wire Prop Lett pp 1–5. https://doi.org/10.1109/LAWP.2024.3453497

19. Chang H, Wang CX, Liu Y et al (2021) A novel nonstationary 6g uav-to-ground wireless channel model with 3-d arbitrary trajectory changes. IEEE Int Things J 8(12):9865–9877. https://doi.org/10.1109/JIOT.2020.3018479

20. Sun R, Cheng N, Li C et al (2024) Knowledge-driven deep learning paradigms for wireless network optimization in 6g. IEEE Netw 38(2):70–78. https://doi.org/10.1109/MNET.2024.3352257

21. Yu J, Liu X, Gao Y et al (2022) Deep learning for channel tracking in irs-assisted uav communication systems. IEEE Trans Wire Commun 21(9):7711–7722

22. Shehzad MK, Rose L, Assaad M (2019) Rnn-based twin channel predictors for csi acquisition in uav-assisted 5g+ networks. In: 2021 IEEE Global Communications Conference (GLOBECOM). IEEE, pp 1–6

23. Zhu Y, Dong X, Lu T (2019) An adaptive and parameter-free recurrent neural structure for wireless channel prediction. IEEE Trans Commun 67(11):8086–8096

24. Mattu SR, Theagarajan LN, Chockalingam A (2022) Deep channel prediction: A dnn framework for receiver design in time-varying fading channels. IEEE Trans Veh Technol 71(6):6439–6453

25. Varshney R, Gangal C, Sharique M et al (2023) Deep learning based wireless channel prediction: 5g scenario. Procedia Comput Sci 218:2626–2635

26. Wei Y, Zhao MM, Liu A et al (2022) Channel tracking and prediction for irs-aided wireless communications. IEEE Trans Wire Commun 22(1):563–579

27. Kulkarni A, Seetharam A, Ramesh A et al (2019) Deepchannel: Wireless channel quality prediction using deep learning. IEEE Trans Veh Technol 69(1):443–456

28. Jiang W, Schotten HD (2020) Deep learning for fading channel prediction. IEEE Open J Commun Soc 1:320–332

29. Xiong L, Zhang Z, Yao D (2022) A novel real-time channel prediction algorithm in high-speed scenario using convolutional neural network. Wire Netw pp 1–14

30. Zhang Y, Wang J, Sun J et al (2020) Cv-3dcnn: Complex-valued deep learning for csi prediction in fdd massive mimo systems. IEEE Wireless Commun Lett 10(2):266–270

31. Chu L, Burghal D, Neuman M, et al (2024) Context-conditioned spatio-temporal predictive learning for reliable v2v channel prediction arxiv:2409.09978

32. Ladosz P, Oh H, Zheng G et al (2020) Gaussian process based channel prediction for communication-relay uav in urban environments. IEEE Trans Aerospace and Electron Syst 56(1):313–325. https://doi.org/10.1109/TAES.2019.2917989

33. Khawaja W, Ozdemir O, Guvenc I (2021) Channel prediction for mmwave ground-to-air propagation under blockage. IEEE Antenna Wireless Prop Lett 20(8):1364–1368. https://doi.org/10.1109/LAWP.2021.3078268

34. Zhang Q, Yang T, Wu C et al (2023) A non-stationary channel prediction method for uav communication network with error compensation. Eng Appl Art Intell 123:106206

35. Pang M, Zhu Q, Wang CX et al (2023) Geometry-based stochastic probability models for the los and nlos paths of a2g channels under urban scenarios. IEEE Int Things J 10(3):2360–2372. https://doi.org/10.1109/JIOT.2022.3211524

36. Colpaert A, Cui Z, Vinogradov E et al (2024) 3d non-stationary channel measurement and analysis for mamimo-uav communications. IEEE Trans Veh Technol 73(5):6061–6072. https://doi.org/10.1109/TVT.2023.3340447

37. Lyu Y, Wang W, Chen P (2024) Fixed-wing uav based air-to-ground channel measurement and modeling at 2.7ghz in rural environment. IEEE Trans Antenna Prop pp 1–1. https://doi.org/10.1109/TAP.2024.3428337

38. Cheng X, Li Y, Wang CX et al (2020) A 3-d geometry-based stochastic model for unmanned aerial vehicle mimo ricean fading channels. IEEE Int Things J 7(9):8674–8687. https://doi.org/10.1109/JIOT.2020.2995707

39. Huang Z, Cheng X (2021) A general 3d space-time-frequency non-stationary model for 6g channels. IEEE Trans Wireless Commun 20(1):535–548. https://doi.org/10.1109/twc.2020.3026356

40. Chen T, Gao S, Zheng S et al (2023) Emd and vmd empowered deep learning for radio modulation recognition. IEEE Trans Cogn Commun Netw 9(1):43–57. https://doi.org/10.1109/TCCN.2022.3218694

41. Lin Z, Li M, Zheng Z et al (2020) Self-attention convlstm for spatiotemporal prediction. Proceed AAAI Conf Art Intell 34(7):11531–11538

42. Remcom (2023) Wireless insite. https://www.remcom.com/wireless-insite-em-propagation-software

43. Joo J, Park MC, Han DS et al (2019) Deep learning-based channel prediction in realistic vehicular communications. IEEE Access 7:27846–27858. https://doi.org/10.1109/ACCESS.2019.2901710

44. Zhang Y, Wu Y, Liu A et al (2021) Deep learning-based channel prediction for leo satellite massive mimo communication system. IEEE Wireless Commun Lett 10(8):1835–1839. https://doi.org/10.1109/LWC.2021.3083267

45. Beijing University of Posts and Telecommunications (2024) AI Air Interface Channel Data. https://doi.org/10.12448/3ch6-w717, Published by China Mobile