Channel Prediction Using Ordinary Differential Equations for MIMO systems

Lei Wang, Guanzhang Liu, Jiang Xue, *Senior Member, IEEE*, and Kat-Kit Wong, *Fellow, IEEE*

***Abstract*—Channel state information (CSI) estimation is part of the most fundamental problems in 5G wireless communication systems. In mobile scenarios, outdated CSI will have a serious negative impact on various adaptive transmission systems, resulting in system performance degradation. To obtain accurate CSI, it is crucial to predict CSI at future moments. In this paper, we propose an efficient channel prediction method in multiple-input multiple-output (MIMO) systems, which combines genetic programming (GP) with higher-order differential equation (HODE) modeling for prediction, named GPODE. In the first place, the variation of one-dimensional data is depicted by using higher-order differential, and the higher-order differential data is modeled by GP to obtain an explicit model. Then, a definite order condition is given for the modeling of HODE, and an effective prediction interval is given. In order to accommodate to the rapidly changing channel, the proposed method is improved by taking the rough prediction results of Autoregression (AR) model as a priori, i.e., Im-GPODE channel prediction method. Given the effective interval, an online framework is proposed for the prediction. To verify the validity of the proposed methods, We use the data generated by the Cluster Delay Line (CDL) channel model for validation. The results show that the proposed methods has higher accuracy than other traditional prediction methods.**

***Index Terms*—MIMO, 5G, Channel prediction, Genetic programming, Ordinary differential equation.**

I. Introduction

In fifth-generation (5G) mobile communication systems, multiple-input multiple-output (MIMO) efficiently achieves high data rates for multi-user communications by increasing the number of transmitter and receiver antennas [1], [2]. However, these results rely on the accurate understanding of the channel state information (CSI) between the transmitter and receiver. CSI describes the channel state of a communication link. In frequency division duplex (FDD) system, the CSI estimated at the receiving end is fed back to the sending end, and the CSI obtained at the sending end may be out of date before actual use, which is mainly due to feedback delay. Although time division duplex (TDD) system can take advantage of channel reciprocity to avoid feedback, processing delays can still cause CSI inaccuracy, especially in high-mobility scenarios. It has been widely proven that the outdated CSI seriously deteriorates system performance [3]–[7]. In 5G systems, this problem will become more serious [8].

To cope with the problem of CSI degradation, researchers have proposed a large number of solutions [9]–[11]. Channel prediction, if done reliably, can significantly reduce the performance loss. Existing channel prediction methods can be mainly divided into three categories: autoregressive (AR) model [11], [12], a sum of sinusoids (SOS) model [13]–[15] and neural network model [16]–[19].

The AR model uses a weighted linear combination of past and current CSI to predict future CSI. Duel-Hallen [11] used the AR model to track channel changes and calculate future CSI based on outdated CSI. Since the AR model is a linear prediction model [12], it has a good prediction effect under slow-changing channels, but its performance is limited by fast-changing channels. The wireless channel can also be modeled as the sum of complex sinusoidal signals. Unlike the AR model, the SOS model does not directly estimate CSI but utilizes outdated CSI for parameter estimation. J. Hwang [14] and M. Chen [15] used multiple signal classification algorithm (MUSIC) to estimate physical parameters and studied estimating signal parameter via rotational invariance techniques (ESPRIT) to predict future CSI, respectively. When the wireless channel parameters do not change with time, the SOS model has superior prediction performance. However, in actual communication systems, channel parameters will change over time, which leads to performance degradation over time. In recent years, the neural network has become the principal method of prediction. The prediction methods based on deep convolutional neural network (CNN) [17], real-valued recurrent neural network (RNN) [18], and long short-term memory network (LSTM) [19] were proposed, and their prediction performance was evaluated in some simulated communication cases. It has self-adaptive[1] ability, and the ability to capture complex nonlinear data rules is strong. However, it has the issues of high data volume requirements, complex parameter settings and poor model interpretation[2], so it is not appropriate for online channel prediction.

In scientific experiments and engineering designs, ordinary differential equations (ODEs) are used to describe time-related nonlinear and complex systems, and their theoretical analysis can provide guidance for actual systems [20]–[22]. H. Helge et

This work was supported in part by the National Key R&D Program of China (2020YFA0713900), the Foundamental Mathematical Theory of Network Comunication, the major key project of PCL, (PCL2021A12), and the Joint Project of Industries and Universities Shaanxi (S2021-YF-GXZD 0076).

L. Wang, G. Liu and J. Xue are with the Xi'an Jiaotong University, Shaanxi, China.

K.-K. Wong is with the Department of Electronic and Electrical Engineering, University College London, Edinburgh, London WCIE 7JE, U.K.

Correspondence author is Jiang Xue (x.jiang@xjtu.edu.cn, 86-29-82665732)

[1]Artificial neural network can obtain the weights and structure of the network through training and learning, showing strong self-learning ability and adaptability to the environment, that is, it can automatically adjust the network structure, node weights, step length, etc.

[2]Artificial neural networks are often referred to as "black boxes", which lack a clear connection between network parameters and approximate mathematical functions.

al. proposed a prediction band based on speedy integration to evaluate the uncertainty of the ODE model on cellular signals [23]. Yang et al. proposed a complex-valued ODE modeling for time series identification [24]. A differential polynomial neural network (DPNN) with ODEs substitutions to predict short-term power load was proposed by Zjavka et al. [25]. Cao et al. used genetic programming (GP) to evolve ODE from the observed time series [26]. The main idea is to embed genetic algorithm (GA) into GP for the discovery and optimization of the model structure. It also demonstrates that the method based on GP has many advantages over other modeling methods. Based on the CSI sequence analysis, CSI in the moving scene is complex and nonlinear. Traditional methods cannot accurately capture these characteristics. Therefore, we express the change characteristics of CSI with high-order differential, and model the high-order differential data through genetic programming, and obtain an explicit model for CSI prediction. An online prediction framework is proposed for long-term prediction, and the simulation results indicate the effectiveness of the proposed algorithm.

In this paper, we focus on the channel prediction problem in time-varying MIMO environments. The main contributions of this paper are as follows:

- GP and higher-order differential equation (HODE) are proposed to model the channel of MIMO systems.
- Use the AR model prediction results as a priori to improve the proposed method performance.
- Use the mean square error (MSE) as a tool to select the order of higher-order differential equation.
- According to the effective prediction interval given by the channel model, an online prediction framework is proposed. The online channel prediction has higher accuracy than offline channel prediction in real communication environments.

The rest of this paper is organized as follows. Section II briefly introduces the system model and nonlinearity. The channel prediction method based on the GPODE modeling is presented in Section III. In Section IV, the performance analysis and simulation results are provided. Finally, section V concludes this paper.

Notations: Throughout this paper, lowercase boldface letters represent vectors, while uppercase boldface letters represent matrices. $\mathbf{A}^T$, $\mathbf{A}^{-1}$ indicate the matrix transpose and inverse of matrix $\mathbf{A}$, respectively. $\mathbb{C}^{m\times n}$ denotes the vector space of all $m\times n$ complex matrices. $||\cdot||$, $j=\sqrt{-1}$ and $x^{(i)}$ represent the L_2-form operator, the imaginary unit and the derivative of x (i is the order of the derivative), respectively.

II. SYSTEM MODEL AND NONLINEARITY

This section mainly introduces the system model and nonlinearity of time-varying channel.

A. System model

Considering the time-varying MIMO systems with N_t and N_r antennas at the transmitter and receiver, respectively. The signal $\mathbf{y}(t)$ received at time t through time-varying channel $\mathbf{H}(t)$ is modeled as

$$\mathbf{y}(t)=\mathbf{H}(t)\mathbf{s}(t)+\mathbf{n}(t), \tag{1}$$

where $\mathbf{y}(t)=[y_1(t),y_2(t),...,y_{N_r}(t)]^T$ is the $N_r\times 1$ vector of received signals at time t, $\mathbf{s}(t)=[s_1(t),s_2(t),...,s_{N_t}(t)]^T$ is an $N_t\times 1$ vector of transmitted signals, $\mathbf{n}(t)$ represents the vector of independent and identically distributed (i.i.d) noise, whose elements follow the complex Gaussian distribution with zero mean and variance σ_n^2. Also, $\mathbf{H}(t)=[h_{n_rn_t}(t)]\in\mathbb{C}^{N_r\times N_t}$ is the matrix of time-varying channel impulse responses at time t, and $h_{n_rn_t}(t)\in\mathbb{C}^{1\times 1}$ denotes the gain of the fading channel between the (n_t,n_r) -antenna pair ($1\le n_t\le N_t$, $1\le n_r\le N_r$). Due to feedback and delay, the CSI obtained at the transmitter may be out of date, i.e., $\mathbf{H}(t)\neq\mathbf{H}(t+\tau)$ for $\tau>0$, resulting in severe degradation of the performance of the communication system [1]–[7], [27].

The simulation channel model considered in this paper is the Cluster Delay Line (CDL) model. The CDL channel model is used for generating realistic radio channel impulse responses for system-level simulations of mobile radio networks, compliance with 3GPP standard (TR 38.901) [28].

As in [28], the CDL channel from the u-th receiver antenna to the s-th transmitter antenna through the n-th path at the t-th time slot can be expressed as

$$\begin{aligned}h_{u,s,n}(t)=\sqrt{\frac{P_n}{M}}\sum_{m=1}^{M}&\begin{bmatrix}F_{rx,u,\theta}\left(\theta_{n,m,ZOA},\phi_{n,m,AOA}\right)\\F_{rx,u,\phi}\left(\theta_{n,m,ZOA},\phi_{n,m,AOA}\right)\end{bmatrix}^T\\&\begin{bmatrix}\exp\left(j\Phi_{n,m}^{\theta\theta}\right)&\sqrt{\kappa_{n,m}^{-1}}\exp\left(j\Phi_{n,m}^{\theta\phi}\right)\\\sqrt{\kappa_{n,m}^{-1}}\exp\left(j\Phi_{n,m}^{\Phi\theta}\right)&\exp\left(j\Phi_{n,m}^{\phi\phi}\right)\end{bmatrix}\\&\begin{bmatrix}F_{tx,s,\theta}\left(\theta_{n,m,ZOD},\phi_{n,m,AOD}\right)\\F_{tx,s,\phi}\left(\theta_{n,m,ZOD},\phi_{n,m,AOD}\right)\end{bmatrix}\\&\exp\left(\frac{j2\pi(\widehat{r}_{rx,n,m}^T\cdot\overline{d}_{rx,u})}{c}\right)\\&\exp\left(\frac{j2\pi(\widehat{r}_{tx,n,m}^T\cdot\overline{d}_{tx,s})}{c}\right)\\&\exp\left(j2\pi\frac{(\widehat{r}_{rx,n,m}^T\cdot\overline{v})}{c}t\right),\end{aligned} \tag{2}$$

where P_n is the power of the n-th path, m is a sub-diameter index in the diameter cluster, M is the number of sub-diameters contained in the diameter cluster n, $\overline{v}$ is the velocity of UE, $F_{rx,u,\theta}$ and $F_{rx,u,\phi}$ are the field patterns of the u in the direction of the spherical basic vector θ and ϕ, respectively, $F_{tx,s,\theta}$ and $F_{tx,s,\phi}$ are the corresponding field modes of the s in the direction of the spherical basic vector θ and ϕ , respectively, $\{\Phi_{n,m}^{\theta\theta},\Phi_{n,m}^{\theta\phi},\Phi_{n,m}^{\Phi\theta},\Phi_{n,m}^{\phi\phi}\}$ are random phases, $\kappa_{n,m}^{-1}$ is the cross-polarization power ratio and c is the wavelength of the carrier frequency, $\widehat{r}_{rx,n,m}$ is a spherical unit vector with an elevation angle of $\theta_{n,m,ZOA}$ and an arrival angle of $\phi_{n,m,AOA}$, and $\widehat{r}_{tx,n,m}$ is a spherical unit vector with elevation angle $\theta_{n,m,ZOD}$ and departure angle $\phi_{n,m,AOD}$. Each path comes from a different scatterer and is composed of multiple sub-paths.

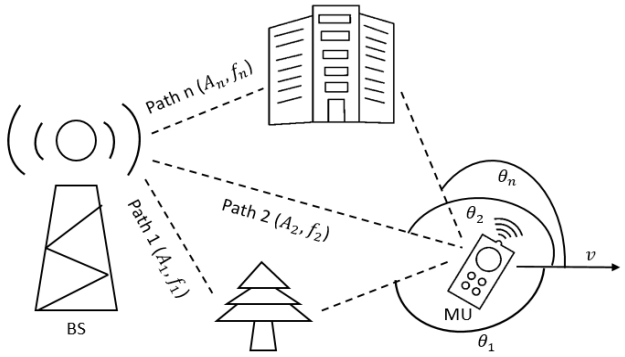

Fig. 1: Channel model.

In this channel, the moving speed only affects the doppler frequency offset. For mathematical convenience, with fixed antenna and diameter cluster, we can write it as

$$h(t) = \sum_{m=1}^{M} A_m \exp(j2\pi f_m t), \tag{3}$$

where $f_m = \frac{\widehat{r}_{rx,n,m}^T \cdot \overline{v}}{c}$ is the doppler shift.

Fig. 1 shows the relationship between base stations (BS), mobile users (MU) and scatterers in a multi-path fading environment. The communication channel between BS and MU is distorted by interference of multi-path and movement of communication ends and/or scatterers. Now, we re-express $h_{n_t n_r}(t)$ in (3) as the superposition of complex sinusoids

$$h_{n_t n_r}(t) = h_{n_t n_r R}(t) + j h_{n_t n_r I}(t), \tag{4}$$

$$h_{n_t n_r R}(t) = \sum_{m=1}^{M} A_m \cos(2\pi f_m t), \tag{5}$$

$$h_{n_t n_r I}(t) = \sum_{m=1}^{M} A_m \sin(2\pi f_m t). \tag{6}$$

We further write $\mathbf{h}(t) = [h_{n_t n_r I}(t) \quad h_{n_t n_r R}(t)]$.

B. Nonlinearity of time-varying channel

The characteristics of data are of great importance to the selection of the prediction model, so it is necessary to analyze the characteristics of time-varying channel.

In this section, we briefly analyze the stability (Augment Dickey-Fuller (ADF) test in [29]), nonlinear (Brock-Dechert Scheinkman (BDS) test in [30]), and complexity (sample entropy in [31]) for time-varying channel data. We use the CDL-C (Urban Macro) scenario that complies with the 3GPP standard (TR 38.901) [28] to generate the channel, the initial phase is a random variable with uniformly distributed over $[-\pi, \pi]$, the subcarrier frequency is 3.5 GHz. The UE velocity is $v = 10$ km/h. Two groups of data are selected from the generated channel data to test the real part and the imaginary part, respectively, marked 1-4. The statistical analysis results are shown in Table 1.

According to the statistical analysis, we have the following conclusions. For the stationary test, the smaller the p-value, the greater the probability of rejecting the null hypothesis (non-stationary sequences), which are stationary sequences in the Table 1. The nonlinear test shows that the data generation is related to the nonlinear system. The complexity can be seen from sample entropy calculation, and the higher the value, the higher the complexity. Therefore, according to the results of timing analysis, we find that the CSI generated by the channel model is nonlinear and complex. The non-linearity of the channel is obtained based on the analysis of experimental data and channel model, which is mainly caused by the scatterers and the moving speeds. Therefore, it is very important to design a prediction model that can model the nonlinearity of the channel.

TABLE I: Channel characteristics analysis results.

	Stability (p value)	Nonlinearity	Complexity
1	Stability (5.79e-13)	Nonlinearity	0.314
2	Stability (0)	Nonlinearity	0.411
3	Stability (2.42e-23)	Nonlinearity	0.532
4	Stability (3.58e-24)	Nonlinearity	0.536

III. THE PROPOSED CHANNEL PREDICTION METHOD

In this section, we first propose a modeling based on high-order ordinary differential equations (HODE), which will be utilized to directly model nonlinearities. Then, we propose a HODE method based on GP to solve the dynamic modeling problem of HODE, called GPODE. That is, we carry out differential processing on the continuous observation data with the scale of n, and use the self-organization and self-learning characteristics of GP to model the processed data, and get the expression of the equation. Its structure can be described with a flowchart as shown in Fig. 2.

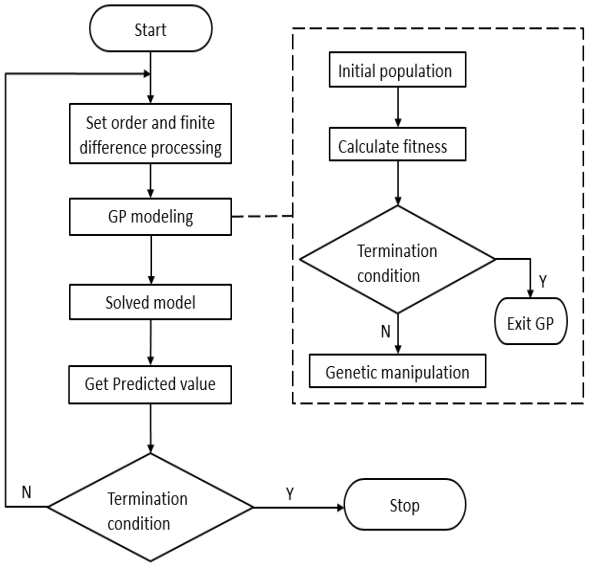

Fig. 2: GPODE flowchat.

A. Proposed model

Let $\mathbf{x}$ be an n-dimensional column vector, whose Euclidean norm is defined as

$$||\mathbf{x}|| = \sqrt{\sum_{i=1}^{n} x_i^2},$$

where x_i is the i-th component of $\mathbf{x}$.

We suppose that the sample observations for a set of dynamic systems $\mathbf{H}$ at the past successive n time steps can be written as

$$\mathbf{H} = (h(t_0), h(t_1), ..., h(t_{n-1})), \tag{7}$$

where $t_i = t_0 + i * \triangle t (i = 0, 1, 2, ..., n-1)$, t_0 denotes the starting time, $\triangle t$ denotes the interval between two observations. The modeling problem of HODE for $\mathbf{H}$ is to find an m-order ODE model,

$$\frac{d^m \mathbf{H}^*}{dt^m} = f(t, h(t), h^{(1)}(t), ..., h^{(m-1)}(t)), \tag{8}$$

where $\mathbf{H}^*$ is the numerical solution, to describe the system such that $||\mathbf{H}^* - \mathbf{H}||$ is minimized and the values of $\mathbf{H}$ at the next τ time steps $\{h(t+1), ..., h(t+\tau)\}$ can be predicted to based on the model, where m is the derivative order and f is composed of some elementary functions, including triangle functions, exponential functions and power functions.

Here, $||\mathbf{H}^* - \mathbf{H}||$ is defined as

$$||\mathbf{H}^* - \mathbf{H}|| = \sqrt{\sum_{i=0}^{n} [h^*(t_i) - h(t_i)]^2}. \tag{9}$$

The structure of the HODE model defined is flexible, and can be in the form of a complex nonlinear ODE with variable coefficients. We define n as the number of modeling samples, τ as the stepsize of prediction. It is normally called short-term prediction for $\tau = 1$, long-term prediction for $\tau \geq 2$.

B. Finite difference approximations

For the model shown in (8), first perform the following transformation

$$(h(t), h^{(1)}(t), ..., h^{(m-1)}(t)) = (z_0(t), z_1(t), ..., z_{m-1}(t)). \tag{10}$$

According to variable substitution, the HODE shown in (8) is transformed into the ODEs system shown in (10)

$$\begin{cases} z_0^{(1)} = z_1 \\ z_1^{(1)} = z_2 \\ \quad \vdots \\ z_{m-2}^{(1)} = z_{m-1} \\ z_{m-1}^{(1)} = f(t, z_1(t), z_2(t), ..., z_{m-1}(t)), \end{cases} \tag{11}$$

where t represents all time points of $t_1 \sim t_n$.

In order to obtain the fluctuation rule of the data, it is necessary to carry out finite difference processing on the data, calculate the approximate value of the derivative [32]. Taking the calculation of the first derivative as an example, the approximate value of the derivative is obtained by the following numerical difference formula.

The data at time t_0 is processed as

$$h^{(1)}(t_i) = \frac{-h(t_{i+2}) + 4h(t_{i+1}) - 3h(t_i)}{2h} + O(h^2); \tag{12}$$

for time $t_1 \sim t_{i-2}$:

$$h^{(1)}(t_i) = \frac{h(t_{i+1}) - h(t_{i-1})}{2h} + O(h^2); \tag{13}$$

for time t_{i-1}:

$$h^{(1)}(t_i) = \frac{h(t_{i-2}) - 4h(t_{i-1}) + 3h(t_i)}{2h} + O(h^2), \tag{14}$$

where h represents the unit time interval, and $h(t_i)$ represents the data at time i. The calculation of high-order are analogous to ensure that data error at different times is always $O(h^2)$[3].

Carry on the m-order difference approximation to $\mathbf{H}$ to obtain matrix $\mathbf{Z}^{n\times(m+1)}$:

$$\mathbf{Z} = \begin{pmatrix} h(t_0) & h^{(1)}(t_0) & \cdots & h^{(m)}(t_0) \\ h(t_1) & h^{(1)}(t_1) & \cdots & h^{(m)}(t_1) \\ \vdots & \vdots & & \vdots \\ h(t_{n-1}) & h^{(1)}(t_{n-1}) & \cdots & h^{(m)}(t_{n-1}) \end{pmatrix}. \tag{15}$$

C. GP

1) Encoding of the model population:

After the HODE is transformed into ordinary differential equations, only the m-th equation, which is essentially the right-hand function form of the original HODE, can affect the structure of the model. When the population is initialized, some such individuals are generated along with the program, and each body is expressed in the additive tree form unique to the GP.

The structure of the additive tree model of ODE system is evolved by using the tree structure based on an evolutionary algorithm. Therefore, we encode the right-hand side of an HODE as an additive tree individual. For example, the following second-order ODE is described

$$h^{(2)} = R + \exp(t) + h * h^{(1)} + \sin\left(\frac{h}{h^{(1)}}\right), \tag{16}$$

where R is random constant number. The corresponding equation

$$z_2 = R + \exp(t) + z_0 * z_1 + \sin\left(\frac{z_0}{z_1}\right), \tag{17}$$

can be represented as an additive tree, as illustrated in Fig. 3.

The instruction set Q_0 and the operator set Q_1 are given to generate the additive tree.

$$\begin{cases} Q_0 = +2, +3, ..., +N, \\ Q_1 = F \cup T = \{-, *, /, \sin, \cos, \exp, \log, t, R, z_i\}, \end{cases} \tag{18}$$

[3]The numerical differentiation method we choose in our paper is the three-point derivation, which mainly uses the forward and backward difference and the central difference. The maximum error of this method is $O(h_2)$, so maintaining the $O(h_2)$ can obtain the optimal approximate derivative value.

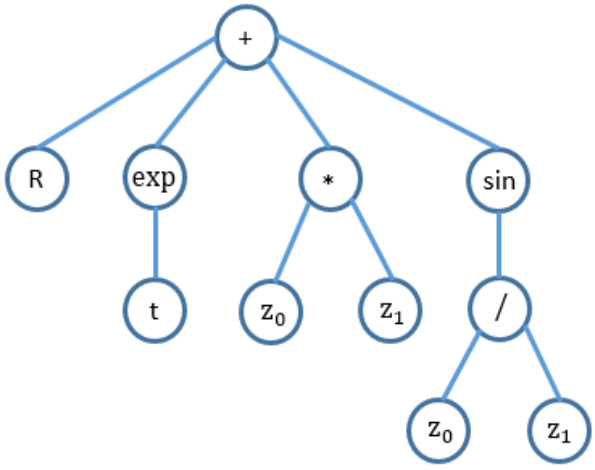

Fig. 3: Example of a HODEs in the form of the additive tree model.

where $F = \{+, -, *, /, \sin, \cos, \exp, \log\}$ and $T = \{t, z_i, R\}$ represent functions and terminal set. $+N, -, *, /$, $\sin, \cos$, $\exp, \log$, t, z_i and R denote addition, subtraction, multiplication, protected division ($\forall a, b \in R$: when $b = 0, a/0 = 1$), sine, cosine, exponent, protected logarithm ($\forall a \in R, a \neq 0$: $\log(a) = \log(\text{abs}(a))$), sample points and constant, which taking $N, 2, 2, 1, 1, 1, 1, 0, 0$ and 0 parameters, respectively [33]. N is an integer (the maximum value of items at the right end of HODE), Q_0 is the instruction and the root node, and the instructions for the other nodes are selected from the operation set Q_1. We select the function set according to the established channel model.

2) Fitness evaluation: A fitness function maps the ODE to a scalar, and a real fitness value that reflects the ODE's performances on a given task. Suppose that the corresponding system of ODEs of an arbitrary individual p_i in the model population has the general form of (11), then the fitness calculation of q_i can be written by:

- Let $\mathbf{H}^*$ and $\triangle\mathbf{H}$ be both m-dimensional column vectors, $\mathbf{Z}^*$ is a $(n+1) \times m$ empty matrix, assign the first row of $\mathbf{Z}$ to that of $\mathbf{Z}^*$;
- Take the element of the 1-th row of $\mathbf{Z}$ as the initial value, and use a numerical algorithm to integrate each individual one step to generate the element of the i-th row of $\mathbf{Z}^*$;
- Let $\mathbf{H}^* = \mathbf{Z}_1^*$, ($\mathbf{Z}_1^*$ denotes the vector composed of the first column of $\mathbf{Z}^*$);
- Fitness $(q_i) = ||\triangle\mathbf{H}||$, $\triangle\mathbf{H} = \mathbf{H} - \mathbf{H}^*$.

It's clear that the less the fitness, the better the individual. Furthermore, due to the diversity of the systems of HODE generated randomly, some of them may not be stable and will give rise to overflow during the fitness calculation. In this case, we return a large number of the fitness value as a penalty so that these unreasonable models can be eliminated from the population soon.

3) Selection strategy and design of genetic operators: The genetic operator of GP is briefly introduced in the following. The genetic operator includes selection operator, crossover operator and mutation operator. The three operators and their operation methods are as follows:

- Selection: Adopt a tournament selection strategy, and select a certain number of individuals from the population at a time (put back into the samples), then the best one is selected to enter the progeny population, and this operation is repeated until the new population reaches the original population size.
- Crossover: The first two parent nodes are selected according to the preset crossover probability, a non-terminal node in the hidden layer is randomly selected for each additive tree, and then the selected subtrees are swapped to get two new subtrees.
- Mutation: Randomly select a mutation point from the parent body, and replace the mutation point with a randomly generated subtree as the root node to generate mutation offspring.

D. Model solving

The HODE model obtained by GPODE algorithm is a kind of nonlinear model, so it is usually solved by a numerical method. The idea of this method is to transform the HODE into a set of equations composed of several first-order equations, and then solve the first-order equation one by one by using the numerical method, which is the odeint solver toolbox in Python [34]. Then, the numerical solution of the HODE is obtained by using the results recursively.

The initial value of the ODEs:

$$\begin{pmatrix} z_1(t_0) \\ \vdots \\ z_m(t_0) \end{pmatrix} = \begin{pmatrix} h'(t_{n-1}) \\ \vdots \\ h^{m-1}(t_{n-1}) \end{pmatrix}. \tag{19}$$

Through numerical methods, the equations shown in (11) are solved recursively from bottom to top, and finally the predicted value $z_0(t+\tau)$ is obtained, where τ is the predicted step size.

E. The prediction framework

In this section, we mainly give the effective prediction interval and two training methods.

1) Determine HODE order: The order of an ODE is also a particularly important choice. When the order of the equation is over high, its numerical solution is more complicated and the stability is worse. We choose the simple and direct method to determine the order, namely, the mean square error (MSE) as the order criterion. From second-order to fifth-order, the most appropriate order was selected to build the model. MSE is defined as

$$\text{MSE} = \frac{1}{N}\sum_{t=1}^{N}(h^*(t) - h(t))^2, \tag{20}$$

where N is prediction length.

2) Analysis on the interval of effective prediction: Due to the variation of doppler, the effective prediction in coherent time should be considered. According to the derivation of the effective prediction interval in [35], [36], it is assumed that $h(t)$ is differentiable of m-order, we can obtain similar interval of effective prediction (IEP) as

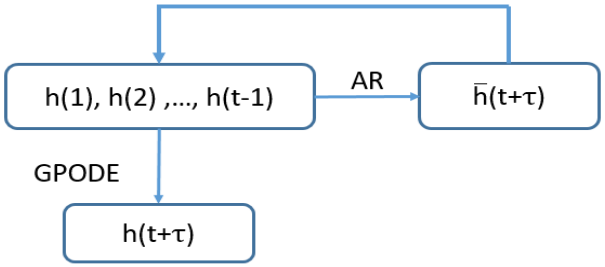

Fig. 4: The framework for improved GPODE.

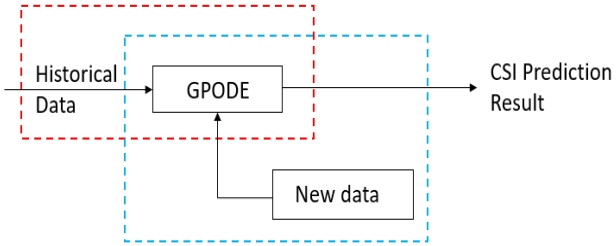

Fig. 5: The framework for online CSI prediction .

$$T_{IEP} = \frac{1}{\beta}\mathrm{In}\left(1+\frac{\delta}{2\pi A_m w_m}\right) \propto \frac{\delta}{w_m^2}, \tag{21}$$

where $\beta = 2\pi w_m$, and $\delta = |h^{'}(t+T_{IEP}) - h^{'}(t)|$ is the threshold.

It can be seen that T_{IEP} is determined by the maximum doppler frequency and the user's movement direction. Therefore, the faster the user moves in the direction of the scatterer, the smaller the T_{IEP}.

3) Improved GPODE: Due to user movement speed increase, the channel fading volatility increases gradually, it is difficult to capture the future state of the moment. Therefore, we improve the finite difference data processing of the verification set. Firstly, the AR channel prediction method is used to preliminary predict the information of the past time, and the information after AR prediction is taken as the information before the initial value of the approximate derivative, which is called the improved GPODE (Im-GPODE).

The specific framework is shown in Fig. 4, where $\overline{h}(t+\tau)$ is the AR model prediction result.

4) Training framework: We consider two prediction frameworks, one is to train a model based on certain data and use the model to predict the subsequent data, which we call it offline prediction for one-step prediction. The other is to continuously update the prediction model according to the effective prediction interval, called online prediction, which is suitable for multi-step prediction. The online framework updates the model by adding the latest data, making predictions in real time. The offline framework can only input all the historical data at one time and use the historical data to predict, and cannot be updated in real time, so it is better to use the online prediction framework for multi-step prediction. Fig. 5 is an online prediction framework diagram.

F. computational complexity

Last but not least, we evaluate the computational complexity of the proposed channel prediction algorithm. The number of complex multiplications is taken as the measurement, the number of generation is g, and the population size is p. For GPODE, the computational complexity mainly includes the calculation of approximate derivative value, the generation of equation and the number of generation. Therefore, the computational complexity of GPODE is $O((p+gp-1)N_tN_rn)$.

Algorithm 1 The Channel Prediction Framework of MIMO.

Input: $\mathbf{H}(t),\ t=1,2,\cdots,N$
Functions set $\mathrm{F}=\{+,-,*,/,\sin,\cos,\exp,\log\}$
Output: $\mathbf{H}(t+\tau)$

for $n_t=1;\ n_t\le N_t;\ n_t++$; **do**
  for $n_r=1;\ n_r\le N_r;\ n_r++$; **do**
    Obtain the sub-channels $h_{n_tn_r}(t)$ following equation (4);
    for $m=2;\ m\le 5; m++$; **do**
      The transformation matrix (15) is obtained by equation (12)∼(14);
      Randomly initialize the population;
      while $i=0;\ i<Generation$ **do**
        Evaluate fitness $q(i)$;
        The next generation parent is selected according to fitness and selection strategy;
        Recombine the next generation parent by using genetic operators (crossover and mutation);
        Evaluate the next generation parent's fitness;
        i=i+1;
      end while
      The optimal individual using equation (20) to calculate the MSE;
      Compare MSE to select the optimal HODE;
    end for
    Predict the next time value $h_{n_tn_r}(t+\tau)$ using equation (19);
  end for
end for
Combine the prediction $\mathbf{H}(t+\tau)$.
return $\mathbf{H}(t+\tau)$.

IV. SIMULATION RESULTS

In this section, we first describe the simulation scenario and parameter settings. Then, we evaluate and analyze the performance of the proposed channel prediction methods through simulations. Some illustrative simulation results in terms of the prediction trends and the predictive accuracy are presented.

A. Simulation configuration

In all simulation, we use simulated fading channels under the CDL-C scenario in the 3GPP standard (TR 38.901) [28] to characterize the performance of the proposed channel prediction methods, where the carrier frequency is 3.5GHz,

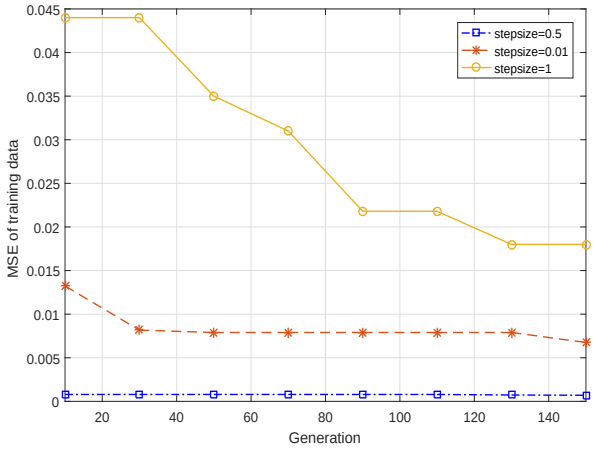

Fig. 6: MSE of training data with different step sizes.

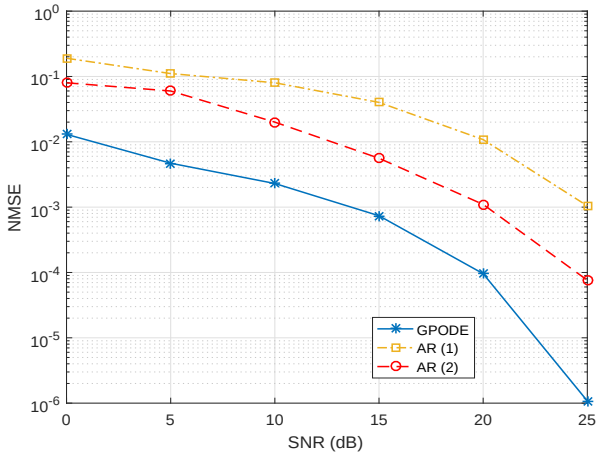

Fig. 7: The NMSE of the proposed method and AR.

the random phase is uniformly distributed within $(-\pi, \pi)$, the signal to noise power ratio (SNR) is set as 20 dB, and the time slot duration is 50ms. We consider that the communication between the BS and MU moves away from the BS at a speed of v.

The system parameters are summarized in Table II while we explicitly state parameter values if they are different from the table. Population size is the number of individuals in the initial population, Generation is the maximum generation, Crossover rate is the pre-defined crossover probability, Mutation rate is the pre-defined mutation probability, and Stepsize is the step size of the numerical method. The selection of these parameters is experienced, and different parameter selection affects the convergence rate of the methods. In order to test the effect of stepsize, we tried experiments with different stepsizes. Fig. 6 illustrates the MSE of training data for different stepsizes of 0.01, 0.5, and 1 for every 20 generations. As evident, the error of HODE increases significantly when the stepsize is about 1, while too small stepsize such as 0.01, can provide smaller errors in the beginning, but offer little improvement in the evolutionary process of HODE. Therefore, we choose the optimal iteration step size of 0.5.

TABLE II: Parameters for experiment.

Parameter	Value
Environment	UMi
Carrier frequency	3.5GHz
MIMO	2×2 ULA
Sampling rate	1MHz
Population size	50
Generation	200
Crossover rate	0.5
Mutation rate	0.1
Stepsize	0.5

In our simulation, we compare the following methods:

- **True Value**: Actual CSI.
- **Outdated**: No prediction is required, and the CSI at the previous moment is directly used as the CSI at the current moment.
- **AR**: Autoregressive prediction introduced in Section III.
- **GPODE**: GPODE trained with the measurements $\{h(1), ..., h(t)\}$ without any processing.
- **Im-GPODE**: Im-GPODE developed in Section V.

B. Prediction analysis

We mainly analyze our methods by prediction accuracy and prediction trend. For short-time prediction, we choose offline training method, while for long-time prediction, we choose online training method.

1) Prediction Accuracy: To evaluate the accuracy of the proposed methods, We define the normalized mean square error (NMSE) as the performance metric

$$\mathbf{NMSE} = \frac{\sum_{\mathbf{i=1}}^{\mathbf{N}} (\widetilde{\mathbf{h}}_{\mathbf{i+1}} - \mathbf{h}_{\mathbf{i+1}})}{\sum_{\mathbf{i=1}}^{\mathbf{N}} \mathbf{h}_{\mathbf{i+1}}}, \tag{22}$$

where N is prediction length, $\widetilde{h}_{i+1}$, h_{i+1} are the predicted and true channel data, respectively.

Firstly, we compare this channel prediction method with AR(1) and AR(2) channel prediction methods for short-time prediction of NMSE under different SNR. Our method is similar to AR(1) method in the structure, but as can be seen from Fig. 7, its accuracy is higher than that of AR(2) method. We observe that the NMSE performance of the proposed GPODE channel prediction method is significantly better than that the traditional channel prediction methods. The proposed channel prediction algorithm can effectively reduce CSI errors. By increasing the SNR, the NMSE of 0 can be closed to by the proposed GPODE when SNR = 25 dB. Therefore, the proposed method is superior in terms of both the accuracy and timeliness. It is effective to use the proposed channel prediction method to predict time-varying CSI in MIMO environment.

Then, the effect of the different BS antenna on channel prediction is evaluated, and the results are shown in Fig. 8. We configure the number of BS antennas as SISO to 32 antennas. It can be seen that with the increase of the number of antennas, the NMSE of CSI will increase due to the large error

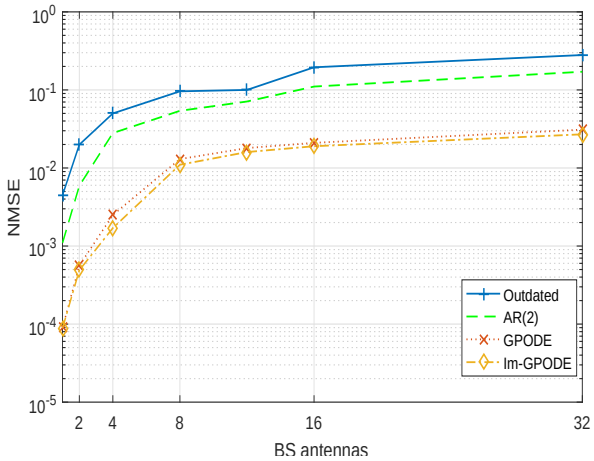

Fig. 8: The NMSE with different number of BS antennas.

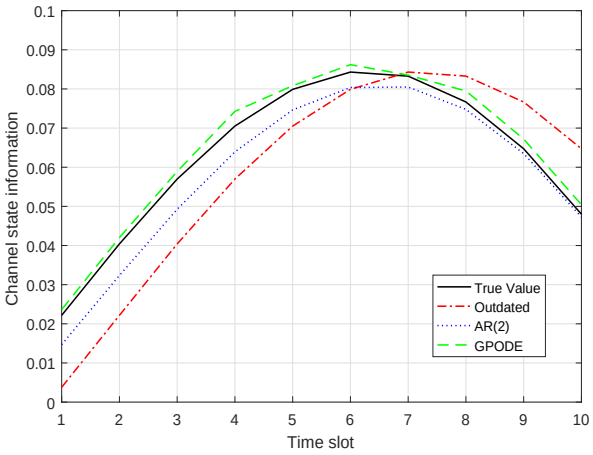

Fig. 10: Prediction trends of the different method in a short period.

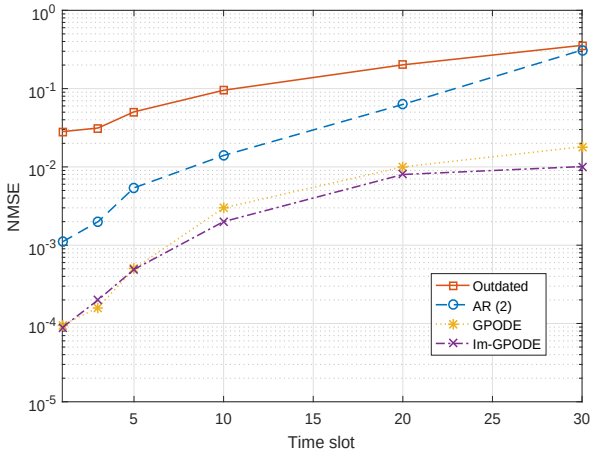

Fig. 9: The NMSE of the long time prediction.

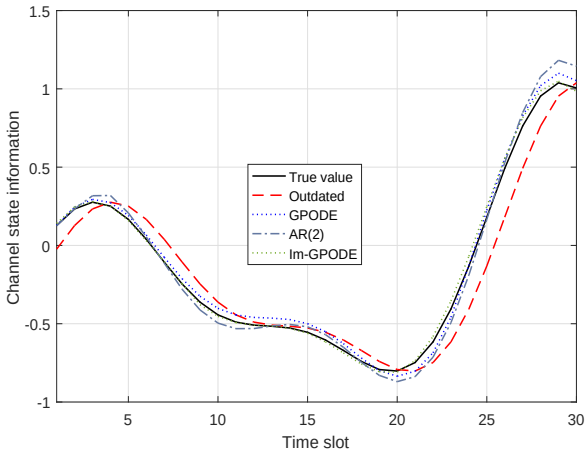

Fig. 11: Prediction trends of the different method in a long period.

accumulation. The accuracy of our method is also improved when the number of antennas is increased.

In order to verify that the framework of online training has significant gain for long-term prediction, we compared the predicted step size. As can be seen from Fig. 9, for long-term prediction, our method have gained compared with the traditional AR channel prediction method. The gain of the improved method is more obvious, and when the predicted step size is longer, the gain is also obvious. Our methods are online sliding window prediction, so the error of stacking will be appropriately reduced. However, for the CSI prediction, the improvement of short-time prediction accuracy can improve the transmission performance. This fully demonstrates that our methods are effective.

2) Prediction Trends: In order to illustrate the prediction trend of our proposed channel prediction methods, channel data are selected for MU to leave the BS at different speed.

Fig. 10 shows the short-term results, where $v = 10km/h$ is used. The long-term results are shown in Fig. 11, where $v = 20km/h$. For trend prediction, an online learning framework is adopted, and the prediction model is constantly updated by sliding window to predict CSI. For the long-term prediction, we use the improved model for prediction because the speed changes too fast.

The results show that the proposed algorithm can track the channel changes smoothly and provide reliable channel prediction when the channel coefficients are monotonously increasing/decreasing in a short time. However, when the channel coefficient fluctuates significantly over a long period of time, the predicted channel coefficient will vibrate due to flutter interference. The improved GOPDE adds the prior of AR prediction, so it combines AR's prediction of the linear part of GPODE's modeling of the nonlinear part, making it able to track the change in the channel. Notably, (21) also shows that T_{IEP} decreases significantly with increasing speed, so more sophisticated signal processing techniques should be used to improve prediction performance.

V. CONCLUSION

In this paper, we studied the channel prediction problem for MIMO systems with mobile user, by leveraging genetic programming techniques. Firstly, we analyzed the time-

varying channel characterization for MIMO systems. Secondly, we proposed two channel prediction methods: the GPODE channel prediction method, and the improved by adding AR predictors, i.e., the Im-GPODE channel prediction method. Then, we gave effective prediction interval and order determination method of HODE. For the long-term prediction, we presented two training modes: online and offline. The numerical results demonstrated that both the proposed GPODE and the Im-GPODE channel prediction methods can enhance the accuracy of channel prediction methods compared with existing methods, and for long-term prediction, the online training mode was more accurate than the offline training mode, the effectiveness of our algorithm was verified in the CDL channel.

REFERENCES

[1] R. Zi, X. Ge, J. Thompson, C. Wang, H. Wang, and T. Han, "Energy efficiency optimization of 5G radio frequency chain systems," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 758–771, 2016.

[2] K. J. Kim, M. Pun, and R. A. Iltis, "Channel prediction for limited feedback precoded mimo-ofdm," in *2008 42nd Annual Conference on Information Sciences and Systems*.

[3] Q. Wang, L. J. Greenstein, L. J. Cimini, D. S. Chan, and A. Hedayat, "Multi-user and single-user throughputs for downlink MIMO channels with outdated channel state information," *IEEE Wireless Communications Letters*, vol. 3, no. 3, pp. 321–324, 2014.

[4] J. Kim, J. Choi, and J. M. Cioffi, "Cooperative distributed beamforming with outdated CSI and channel estimation errors," *IEEE Transactions on Communications*, vol. 62, no. 12, pp. 4269–4280, 2014.

[5] P. Aquilina and T. Ratnarajah, "Performance analysis of IA techniques in the MIMO IBC with imperfect CSI," *IEEE Transactions on Communications*, vol. 63, no. 4, pp. 1259–1270, 2015.

[6] Y. Teng, M. Liu, and M. Song, "Effect of outdated CSI on handover decisions in dense networks," *IEEE Communications Letters*, vol. 21, no. 10, pp. 2238–2241, 2017.

[7] X. Yu, W. Xu, S. H. Leung, and J. Wang, "Unified performance analysis of transmit antenna selection with OSTBC and imperfect CSI over nakagami-m fading channels," *IEEE Transactions on Vehicular Technology*, no. 99, 2017.

[8] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge, U.K.: Cambridge Univ. Press,2005.

[9] W. Jiang, T. Kaiser, and A. J. H. Vinck, "A robust opportunistic relaying strategy for co-operative wireless communications," *IEEE Transactions on Wireless Communications*, vol. 15, no. 4, pp. 2642–2655, 2016.

[10] D. J. Love, R. W. Heath, V. K. N. Lau, D. Gesbert, B. D. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 8, pp. 1341–1365, 2008.

[11] A. Duel-Hallen, "Fading channel prediction for mobile radio adaptive transmission systems," *Proceedings of the IEEE*, vol. 95, no. 12, pp. 2299–2313, 2007.

[12] K. E. Baddour and N. C. Beaulieu, "Autoregressive modeling for fading channel simulation," *IEEE Transactions on Wireless Communications*, vol. 4, no. 4, pp. 1650–1662, 2005.

[13] R. O. Adeogun, P. D. Teal, and P. A. Dmochowski, "Extrapolation of MIMO mobile-to-mobile wireless channels using parametric-model-based prediction," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 10, pp. 4487–4498, 2015.

[14] Jeng-Kuang Hwang and J. H. Winters, "Sinusoidal modeling and prediction of fast fading processes," in *IEEE GLOBECOM 1998 (Cat. NO. 98CH36250)*, vol. 2, 1998, pp. 892–897.

[15] Ming Chen, M. Viberg, and T. Ekman, "Two new approaches to channel prediction based on sinusoidal modelling," in *IEEE/SP 13th Workshop on Statistical Signal Processing*, 2005, pp. 697–700.

[16] W. Jiang and H. D. Schotten, "Neural network-based fading channel prediction: A comprehensive overview," *IEEE Access*, vol. 7, pp. 118 112–118 124, 2019.

[17] P. Dong, H. Zhang, G. Y. Li, N. NaderiAlizadeh, and I. S. Gaspar, "Deep CNN for wideband mmwave massive mimo channel estimation using frequency correlation," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 4529–4533.

[18] W. Liu, L-L Yang, and L. Hanzo, "Recurrent neural network based narrowband channel prediction," in *2006 IEEE 63rd Vehicular Technology Conference*, vol. 5, 2006, pp. 2173–2177.

[19] C. Luo, J. Ji, Q. Wang, X. Chen, and P. Li, "Channel state information prediction for 5G wireless communications: A deep learning approach," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 1, pp. 227–236, 2020.

[20] B. Tra, B. Xue, and M. Zhang, "Genetic programming for feature construction and selection in classification on high-dimensional data," *Memetic Computing*, 2015.

[21] K. Krawiec, *Behavioral Program Synthesis with Genetic Programming*. Springer International Publishing, 2016.

[22] M. M. Mostafa and A. A. El-Masry, "Oil price forecasting using gene expression programming and artificial neural networks," *Economic Modelling*, vol. 54, pp. 40–53, 2016.

[23] H. Helge, K. Clemens, T. Jens, and K. Daniel, "Fast integration-based prediction bands for ordinary differential equation modelsfast integration-based prediction bands for ordinary differential equation models," *Bioinformatics (Oxford, England)*, 2016.

[24] B. Yang and W. Bao, "Complex-valued ordinary differential equation modeling for time series identification," *IEEE Access*, vol. 7, pp. 41 033–41 042, 2019.

[25] L. Zjavka and V. Snasel, "Short-term power load forecasting with ordinary differential equation substitutions of polynomial networks," *Electric Power Systems Research*, vol. 137, no. aug., pp. 113–123, 2016.

[26] H. Cao, L. Kang, Y. Chen, and J. Yu, "Evolutionary modeling of systems of ordinary differential equations with genetic programming," *Genetic Programming and Evolvable Machines*, vol. 1, no. 4, pp. 309–337, 2000.

[27] A. M. Tulino, A. Lozano, and S. Verdu, "Impact of antenna correlation on the capacity of multiantenna channels," *IEEE Transactions on Information Theory*, vol. 51, no. 7, pp. 2491–2509, 2005.

[28] "5g; study on channel model for frequencies from 0.5 to 100 GHz : Tr 38.901 version 15.0.0," 2018.

[29] D. Dickey and W. Fuller, "Distribution of the estimators for autoregressive time series with a unit root," *Journal of the American Statal Association*, vol. 79, no. 366, pp. 355–367, 1979.

[30] W. D. D. W.A.Broock and J. Scheinkan, "A test for independence based on the correlation dimension," *Working papers*, vol. 15, no. 3, pp. 197–235, 1995.

[31] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American journal of physiology. Heart and circulatory physiology*, vol. 278, no. 6, pp. 2039–2049, 2000.

[32] I. R. Khan and R. Ohba, "Taylor series based finite difference approximations of higher-degree derivatives," *Journal of Computational and Applied Mathematics*, vol. 154, no. 1, pp. 115–124, 2003.

[33] Y. Chen, J. Yang, Y. Zhang, and J. Dong, "Evolving additive tree models for system identification," vol. 3, 01 2005.

[34] G. V. Rossum and F. L. Drake, "Python tutorial release 2. 1.1," 2001.

[35] W. Peng, M. Zou, and T. Jiang, "Channel prediction in time-varying massive mimo environments," *IEEE Access*, vol. 5, pp. 23 938–23 946, 2017.

[36] W. Peng, W. Li, W. Wang, X. Wei, and T. Jiang, "Downlink channel prediction for time-varying FDD massive MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 5, pp. 1090–1102, 2019.