A Comparison of Wireless Channel Predictors: Artificial Intelligence versus Kalman Filter

Wei Jiang*† and Hans Dieter Schotten†*
*German Research Center for Artificial Intelligence (DFKI)
Trippstadter Street 122, Kaiserslautern, 67663 Germany
Emails: {wei.jiang, hans.schotten}@dfki.de
†Institute for Wireless Communication and Navigation, University of Kaiserslautern
Building 11, Paul-Ehrlich Street, Kaiserslautern, 67663 Germany
Emails: {wei.jiang@dfki.uni-kl.de, schotten@eit.uni-kl.de}

***Abstract*—Accurate channel state information (CSI) is a prerequisite to reap the benefits of fading-adaptive wireless communications. In practice, however, the available CSI is generally outdated due to processing and feedback delays, which deteriorate system's performance severely. Channel prediction that is able to forecast future CSI provides a promising solution. In addition to statistical methods, namely modelling a time-varying channel as an autoregressive process and using a Kalman filter to predict, artificial intelligence techniques with the capability of time-series prediction are also being discussed recently. This paper compares performance and complexity of these two kinds of predictors. The numerical results on prediction accuracy measured by mean squared error in both noiseless and noisy Rayleigh fading channels, together with their achieved performance in a transmit antenna selection system, are comparatively illustrated.**

I. Introduction

Providing accurate channel state information (CSI) at the transmitter can remarkably improve the performance of wireless techniques, e.g., antenna selection [1], multiple-input multiple-output (MIMO) [2], massive MIMO [3], relaying [4], physical layer security [5] and ultra-reliable transmissions [6]. However, these techniques suffer from severe performance degradations if the available CSI is inaccurate, which generally happens due to feedback and processing delays. In the era of fifth generation (5G) mobile systems, such a dilemma becomes more challenging. On the one hand, the revolutionary 5G applications, such as Tactile Internet [7] and automated driving, impose a demand on more reliable, secure and higher available wireless connectivity. On the other hand, getting accurate CSI becomes harder in some 5G deployment scenarios, e.g., millimeter wave and high-speed trains.

To combat with CSI inaccuracy, a large number of mitigation algorithms and protocols have been proposed in the literature. These methods sacrifice scarce wireless resources (power, time, frequency, etc.) to *passively* compensate for the performance loss [4]. In contrast, as an *active* approach that improves the quality of CSI directly without the cost of wireless resource, channel prediction has attracted interest from researchers. In addition to statistical methods [8]–[10] that model a fading channel as an autoregressive (AR) process and employ a Kalman filter (KF) to predict, Artificial Intelligence (AI) techniques with the capability of time-series prediction is also being discussed recently. In [11], the authors proposed to use a recurrent neural network (RNN) to build a predictor for narrow-band channels. This predictor was further extended to MIMO channels by [12] and [13]. In [14], the authors of this paper proposed to apply a real-valued RNN to implement a multi-step MIMO channel predictor and further illustrated its achievable performance in a multi-antenna system in [15].

Although the feasibility of AI-based channel prediction has been justified in [11]–[15], to the best of our knowledge, a comparison on performance and complexity between AI- and KF-based predictors is still missing. Also, their achievable performance in wireless systems, especially in noisy channels, were not reported in the literature. To fill this gap, this paper compares prediction accuracy measured by mean squared error (MSE) between KF- and RNN-based predictors in both noiseless and noisy Rayleigh channels. The performance in terms of outage probability achieved by a transmit antenna selection (TAS) system with the aid of channel prediction is also provided. Besides, the computational complexity of these two predictors in terms of the amount of complex-valued multiplication is compared.

The rest of this paper is organized as follows: Section II discusses the principles of these two predictors. Section III provides an example of applying a predictor in a TAS system. Section IV compares their performance and complexity via numerical results. Finally, Section V remarks this paper.

II. MIMO Channel Predictors

Without loss of generality, a frequency-flat fading MIMO system with N_r receive and N_t transmit antennas is given by

$$\mathbf{y}(t) = \mathbf{H}(t)\mathbf{x}(t) + \mathbf{z}(t), \tag{1}$$

where $\mathbf{y}(t)$ represents N_r received symbols at time t, $\mathbf{x}(t)$ corresponds to N_t transmit symbols, $\mathbf{z}$ is additive white Gaussian noise, $\mathbf{H}(t)=[h_{n_r n_t}(t)]_{N_r \times N_t}$ stands for the channel matrix at time t, and $h_{n_r n_t} \in \mathbb{C}^{1\times 1}$ denotes the complex-valued channel coefficient between transmit antenna n_t and receive antenna

*This work was supported by German Federal Ministry of Education and Research (BMBF) under the TACNET4.0 project with grant no. *KIS15GTI007*.

n_r, where $1 \leqslant n_r \leqslant N_r$ and $1 \leqslant n_t \leqslant N_t$. Owing to feedback and processing delays, the CSI used to select adaptive parameters at the transmitter may substantially differs from the CSI at the instant of using these selected parameters to transmit signals, namely $\mathbf{H}(t) \neq \mathbf{H}(t+\tau)$, where τ denotes the delay. It was widely recognized [1]–[4] that the outdated CSI severely deteriorates systems' performance. The aim of channel prediction is to forecast a channel matrix $\hat{\mathbf{H}}(t+\tau)$ that is as close as possible to its actual value $\mathbf{H}(t+\tau)$ for the upcoming instant $t+\tau$. Relying on traditional statistical methods [8]–[10], a fading channel is modelled as an autoregressive process and a Kalman filter is employed to build a linear predictor. Recently, recurrent neural network [16], an AI technique has a strong capability on time-series prediction, is also being discussed. The mechanisms of these two kinds of predictors are detailed as follows:

A. KF-based Predictor

According to [17], a complex autoregressive process of order p denoted by $\mathrm{AR(p)}$ can be generated by

$$x[n] = \sum_{k=1}^{p} a_k x[n-k] + w[n], \tag{2}$$

where $w[n]$ is zero mean complex Gaussian noise with the variance of σ_p^2, and $\{a_1, a_2, \ldots, a_p\}$ denote the AR model coefficients. The corresponding power spectral density (PSD) of the $\mathrm{AR(p)}$ process is written as

$$P_a(f) = \frac{\sigma_p^2}{\left|1 + \sum_{k=1}^{p} a_k e^{-2\pi j f k}\right|^2}. \tag{3}$$

For the Rayleigh channel, the theoretical PSD of a fading signal has the following form:

$$P_r(f) = \begin{cases} \dfrac{1}{\pi f_d \sqrt{1 - \left(\frac{f}{f_d}\right)^2}}, & |f| \leqslant f_d \\ 0, & f > f_d \end{cases}, \tag{4}$$

where f_d is the maximum Doppler shift. The corresponding discrete-time autocorrelation function is

$$R[n] = J_0(2\pi f_m |n|), \tag{5}$$

where $f_m = f_d T_s$ indicates the maximal Doppler shift normalized by the sampling rate $f_s = 1/T_s$, and $J_0(\cdot)$ is the zeroth-order Bessel function of the first kind. According to [17], an arbitrary spectrum can be closely approximated by an AR model with sufficiently large order. The basic relationship between a desired autocorrelation function $R[n]$ and an $\mathrm{AR(p)}$ model parameters can be given in matrix form by

$$\mathbf{v} = \mathbf{R}\mathbf{a}, \tag{6}$$

where

$$\mathbf{R} = \begin{bmatrix} R[0] & R[-1] & \cdots & R[-p+1] \\ R[1] & R[0] & \cdots & R[-p+2] \\ \vdots & \vdots & \ddots & \vdots \\ R[p-1] & R[p-2] & \cdots & R[0] \end{bmatrix}, \tag{7}$$

$$\mathbf{a} = \begin{bmatrix} a_1 & a_2 & \cdots & a_p \end{bmatrix}^T, \tag{8}$$

$$\mathbf{v} = \begin{bmatrix} R[1] & R[2] & \cdots & R[p] \end{bmatrix}^T. \tag{9}$$

Substituting (7)-(9) into (6), the coefficients $\{a_1, a_2, \ldots, a_p\}$ is determined. Thus, the KF-based predictor is depicted as

$$\hat{\mathbf{H}}[t+1] = \sum_{k=1}^{p} a_k \mathbf{H}[t-k+1] \tag{10}$$

Note that (10) can only provide a one-step-ahead prediction, in comparison with a multi-step-ahead prediction enabled by the RNN-based predictor.

B. RNN-based Predictor

The RNN is an AI technique that has shown a strong capability in time-series prediction [16]. Fig.1 illustrates the internal structure of an RNN-based multi-step multi-antenna channel predictor. It is composed by three layers: an output layer with N_o neurons, a hidden layer with N_H neurons and an N-dimensional input layer that consists of external input and feedback from the output. At time t, the corresponding channel matrix $\mathbf{H}(t)$, as well as its d-step delays $\mathbf{H}(t-1), ..., \mathbf{H}(t-d)$, are fed into the RNN as the external input. In order to adapt to the input layer, channel matrices need to be vectorized as:

$$\mathbf{h}_v = \vec{\mathbf{H}} = [h_{11}, h_{12}, ..., h_{N_r N_t}]. \tag{11}$$

Such an input structure can be implemented through a tapped delay line with d taps and a Matrix-to-Vector (M2V) module as shown in the figure. Feeding a recurrent component denoted by $\hat{\mathbf{h}}_v(t) = [\hat{h}_{11}(t), ..., \hat{h}_{N_r N_t}(t)]$ back to the input layer, together with the external input, the input vector at time t can be expressed by $\mathbf{i}(t) = [\mathbf{h}_v(t), \mathbf{h}_v(t-1), ..., \mathbf{h}_v(t-d), \hat{\mathbf{h}}_v(t)]$. The output of the RNN is thus a prediction for D steps ahead, i.e., $\hat{\mathbf{h}}_v(t+D)$, which can be transformed into a predicted channel matrix $\hat{\mathbf{H}}(t+D)$ by a Vector-to-Matrix (V2M) module.

The behaviour of an RNN is decided by its weights and transfer functions. Each connection between the output of a neuron in the predecessor layer and the input of a neuron in the successor layer is assigned with a weight. As shown in Fig.1, w_{ln} denotes the weight connecting the n^{th} input and the l^{th} hidden neuron, while c_{ml} is the weight for hidden neuron l and output neuron m, where $1 \leqslant n \leqslant N$, $1 \leqslant l \leqslant N_H$ and $1 \leqslant m \leqslant N_o$. The transfer function typically falls into one of three models: linear, threshold and Sigmoid. In general, the Sigmoid function is chosen in each hidden neuron, which is defined as

$$S(x) = \frac{1}{1 + e^{-x}}. \tag{12}$$

Then, the output of the l^{th} hidden neuron at time t is

$$\Omega_l(t) = S\left(\mathbf{w}_l \cdot \mathbf{i}(t)\right) \tag{13}$$

where $\mathbf{w}_l \cdot \mathbf{i}(t)$ is the dot product of the input vector $\mathbf{i}(t)$ and the weight vector $\mathbf{w}_l = [w_{l1}, ..., w_{lN}]$ related to the l^{th} hidden

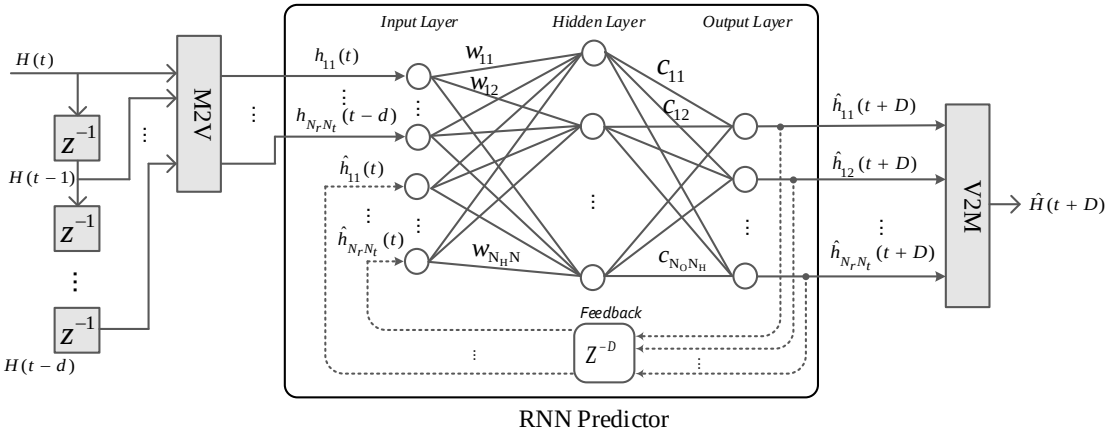

Fig. 1. Schematics of an RNN-based multi-step multi-antenna channel predictor.

neuron. Given the output neuron with a linear transfer function, the m^{th} output denoting a prediction value for time $t+D$ is

$$y_m(t+D) = \sum_{l=1}^{N_H} c_{ml}\Omega_l(t), \tag{14}$$

where the index m stands for $m=(n_r-1)N_T+n_t$. Substituting (13) into (14), the prediction value at the m^{th} output neuron is

$$\hat{h}_{n_r n_t}(t+D) = \sum_{l=1}^{N_H} c_{((n_r-1)N_T+n_t)l} S\left(\mathbf{w}_l \cdot \mathbf{i}(t)\right). \tag{15}$$

The operation of the RNN-based predictor is divided into two phases: training and prediction. Once a network's parameters, such as the number of layers and neurons, have been determined, it is ready to be trained. Providing a training dataset, i.e., a series of channel samples, the RNN processes each sample and compares its resulting prediction against the desired value. The prediction error is propagated back through the network so as to update the weights iteratively until a convergence condition reaches. Afterwards, the trained RNN can be applied to predict unknown channels. The details of training methods for an RNN can refer to the literature such as [18].

III. Prediction-assisted TAS System

To further shed light on channel prediction, a multi-antenna system with transmit antenna selection is utilized as an exemplary application. As illustrated in Fig.2, a frequency-flat fading MIMO system with N_t transmit antenna candidates and N_r receive antennas is considered. The signal transmission is organized as blocks with a length of S symbols per block. The antenna-specific pilot symbols are inserted at the header of every block in a time-division multiplexing manner so as to independently estimate the experienced fading. The pilot symbols constitute the first N_t entries, while the remainder is data symbols. Relying on these pilots, a channel matrix of size $N_r \times N_t$ for the t^{th} block, i.e., $\mathbf{H}(t)=[h_{n_t}(t)]_{N_r \times N_t}$, can be estimated at the receiver.

Assuming L out of N_T transmit antennas are selected, corresponding to a total of $\binom{N_t}{L}$ possible choices. Denoting the channel matrix with the dimension of $N_r \times L$ for the j^{th} choice as $\mathbf{H}_j(t)$, where $1 \leqslant j \leqslant \binom{N_t}{L}$, which is a subset of $\mathbf{H}(t)$. The traditional TAS system directly applies $\mathbf{H}(t)$ to select antennas with the largest channel gain:

$$J = \arg\max_{1 \leqslant j \leqslant \binom{N_t}{L}} \|\mathbf{H}_j(t)\|^2, \tag{16}$$

where $\|\cdot\|$ denotes the Frobenius norm of a matrix. The receiver feeds the selected index J back to the transmitter through a feedback channel. The transmitter activates antennas belonging to choice J to transmit the data symbols at block $t+D$, where D corresponds to the minimal number of blocks needed to absorb the feedback delay. Due to the channel fading, the channel matrix $\mathbf{H}(t)$ used to select antennas may substantially differ from $\mathbf{H}(t+D)$ at the block of data transmission. In other words, $\mathbf{H}_j(t)$ that has the largest channel gain at block t may vary after D blocks later, even in a deep fade, leading to a severe performance loss. With the assistance of channel prediction, the best antenna choice can be selected in terms of the predicted CSI rather than the outdated CSI. At time t, as depicted in Fig.2, the channel matrix $\mathbf{H}(t)$ are fed into the RNN to predict the upcoming CSI $\hat{\mathbf{H}}(t+D)$, which is employed to select transmit antennas following

$$\hat{J} = \arg\max_{1 \leqslant j \leqslant \binom{N_t}{L}} \|\hat{\mathbf{H}}_j(t+D)\|^2 \tag{17}$$

IV. Performance and Complexity Comparisons

In this section, the performance and complexity of the KF- and RNN-based predictors are compared by Monte-Carlo simulation. The numerical results on prediction accuracy measured by MSE, as well as outage probability achieved by a prediction-assisted TAS system, over independent and identically distributed (*i.i.d*) multi-antenna channels are illustrated. Each subchannel is assumed to follow Rayleigh frequency-flat fading, where the channel coefficient h is zero-mean circularly-symmetric complex Gaussian random variable with the variance of 1, i.e., $h \sim \mathcal{CN}(0,1)$. The sampling rate is set to 100KHz that generally satisfies a frequency-flat fading assumption and the maximal Doppler shifts up to 300Hz are tested. The signal transmission is organized in block-wise, with a block size

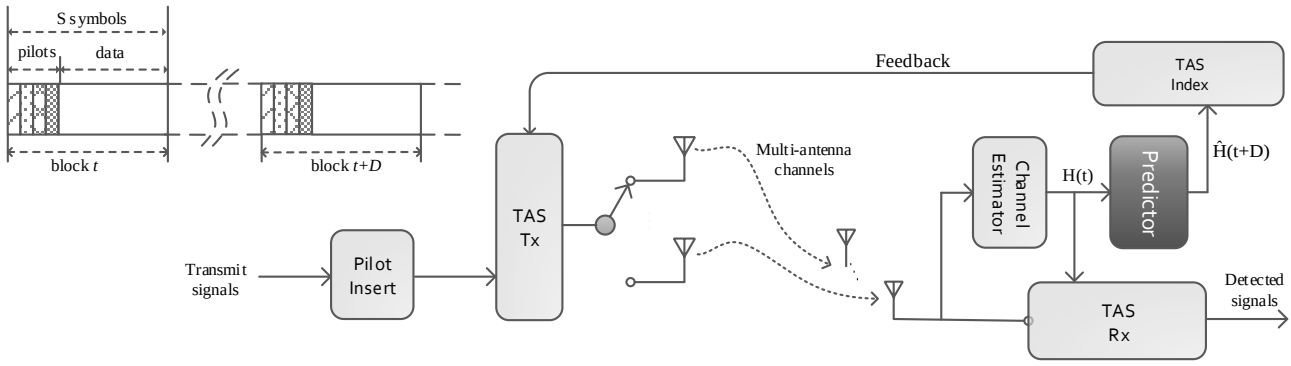

Fig. 2. Illustration on a prediction-assisted TAS system, where the current CSI $\mathbf{H}(t)$ is input into a channel predictor to forecast the upcoming CSI $\hat{\mathbf{H}}(t+D)$. The antennas are selected according to $\hat{\mathbf{H}}(t+D)$, rather than $\mathbf{H}(t)$ in a traditional TAS system.

TABLE I
SIMULATION PARAMETERS

Parameters	Values
Sampling rate	f_s=100KHz
Maximum Doppler shifts	$f_d < 300$Hz
MIMO configuration	4×1
Block size	S=50 symbols
Channels	*i.i.d.* Rayleigh frequency-flat fading
Neural Network	3-layer RNN
Training algorithm	Levenberg-Marquardt [18]
Size of training data	1000 samples per subchannel
Size of testing data	10^7 samples in total
Number of hidden neuron	$N_H = 10$

amounts to S=50 that includes N_t pilot symbols inserted at the head of each block. Through the observation in the simulation, the optimal number of hidden neurons is N_L=10 and the length of tapped delay line is d=3. The simulation parameters are summarized in Table I.

A. Computational Complexity

Let us use the number of operations in complex multiplication as the metric for computational complexity. To conduct one time prediction, as calculated from (13) and (14), the hidden and output layers need to carry out NN_H and N_oN_H times multiplication operations, respectively, amounting to a complexity of Ω_r=$N_H(N+N_o)$. As shown in Fig.1, the number of required input neurons is proportional to both the number of subchannels and the delayed taps d, namely N=$(d+2)N_rN_t$. The number of output neurons equals to the number of subchannels N_o=N_rN_t. Then, the required multiplication operations per prediction can be calculated by Ω_r=$(d+3)N_HN_rN_t$. In contrast, the complexity of the KF-based predictor is Ω_k=pN_rN_t as derived from (10). It is recommended by simulation results that p=4 is the optimal filter order, analogous to d=3 and N_H=10 for the RNN. In a 4×1 MIMO system, for example, we have Ω_k=16 and Ω_r=240. That is too say, the KF-based predictor is simpler than the RNN-based predictor.

Further, it is meaningful to make clear how many computing resources are required. The prediction rate, i.e., the number of predictions conducted per second, can be calculated by f_p=f_s/S with the assumption of one prediction per block, resulting in f_p=2000 in terms of Table I. Thus, the required number of multiplication operations per second for the KF- and RNN-based predictor are $\Omega_k f_p$=3.2×10^4 and $\Omega_r f_p$=4.8×10^5, respectively. Compared with the capability of current digital signal processor (DSP), e.g., TI 66AK2x that provides more than 10^4 Million Instructions executed Per Second (MIPS), their required computing resources are less than 0.01%. Even if in a massive MIMO system with a dimension of 32×4, their required computing resources occupy only about 0.1%. Hence, the complexity of channel prediction is quite reasonable in comparison with the hardware capability, which is very promising from the practical perspective.

B. Performance

To train the neural network, we get a training dataset that consists of a series of CSI extracted from consecutive 10^3 data blocks, i.e., $\{\mathbf{H}(t)|1\leqslant t\leqslant 10^3\}$. The training process starts from an initial state where all weights are randomly set. At iteration t, feeding the channel matrix $\mathbf{H}(t)$ into the RNN, the resultant output is compared with the desired value and the prediction error $\hat{\mathbf{H}}(t+D)-\mathbf{H}(t+D)$ is backpropagated to update the weights by training algorithms such as Levenberg-Marquardt [18]. This process is iteratively carried out until the RNN reaches a certain convergence condition. For evaluating the trained RNN, a testing dataset consisting of CSI extracted from consecutive K=10^7 blocks, i.e., $\{\mathbf{H}(k)|1\leqslant k\leqslant K\}$, is built. At the k^{th} data block, the predictor generates a predicted channel matrix $\hat{\mathbf{H}}(k+D)$ to approximate its actual value $\mathbf{H}(k+D)$. To measure the accuracy of prediction, mean squared error that is defined as MSE=$\frac{1}{K}\sum_{k=1}^{K}\left\|\hat{\mathbf{H}}(k)-\mathbf{H}(k)\right\|^2$ is employed as a performance metric. In contrast, the KF-based predictor does not need a training process. Its filter coefficients can be figured out according to (6) given f_d and f_s. Note that the filter is

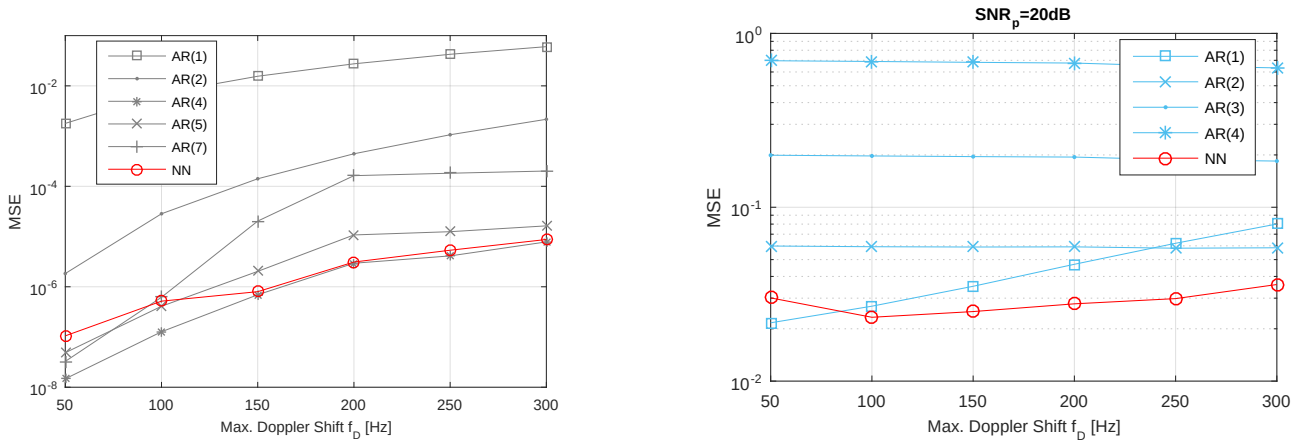

Fig. 3. Prediction accuracy of two predictors with different maximal Doppler shifts in (a) noiseless and (b) noisy Rayleigh channels.

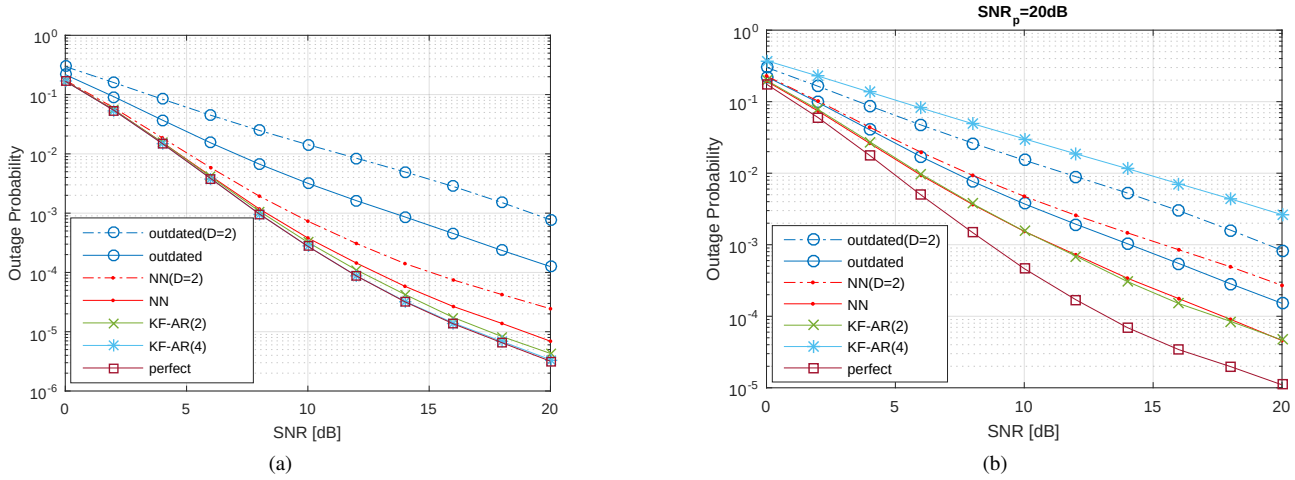

Fig. 4. Outage probabilities achieved by the prediction-assisted TAS system in (a) noiseless and (b) noisy *i.i.d.* multi-antenna Rayleigh channels.

applied to predict the CSI of pilots, rather than that of data symbols. Hence, the sampling rate used in (5) should equal to the rate of prediction f_p. Once the coefficients are determined, we can apply the KF-based predictor given in (10) on the testing dataset to derive its MSE.

Fig.3 compares the prediction accuracy of two kinds of predictors in both noiseless and noisy Rayleigh channels. For simplicity, only the MSE results for the single-antenna case N_t=N_r=1 are illustrated since those of the multi-antenna case are equivalent due to the assumption of *i.i.d.* channels. First, we assume that the applied CSI is perfect without taking into account estimation errors. As shown in Fig.3a, the channel predictors have very high prediction accuracy in a noiseless channel. Although the MSE results get worse with the growth of f_d, the RNN-based predictor still achieves a result of nearly 10^{-5} in a high Doppler shift of f_d=300Hz. It is observed that the performance of the KF-based predictor is affected by the filter order p. Starting from p=1, which has a weak capability with an MSE greater than 10^{-1} in f_d=300Hz, the prediction accuracy can be improved with the increase of the filter order until p=4. As shown in the figure, $\mathrm{AR}(4)$ is the optimal KF-based predictor in a noiseless channel. It has comparable accuracy as the RNN-base predictor except in the cases of f_d=50 and 100Hz where $\mathrm{AR}(4)$ slightly outperforms. The accuracy becomes worse after the filter order increases to greater than p>4. That is to say, a larger order does not necessarily correspond to a performance improvement.

In practice, the available CSI is imperfect because additive noise cannot be avoided in the process of channel estimation. Under the assumption that the signal-to-noise ratio (SNR) of pilots is $\mathrm{SNR_p}$=20dB, a training and testing dataset contain noisy CSI are built. Applying these datasets to retrain the RNN and evaluate the predictors, the results reveal that noise has an obvious impact on prediction accuracy. As shown in Fig.3b, the MSE results achieved by the RNN-based predictor deteriorate from better than 10^{-5} in the noiseless channel to larger than 10^{-2}. The best accuracy of the KF-based predictor is this time achieved by $\mathrm{AR}(1)$, which is still inferior to the RNN-based

predictor. As shown in the figure, the accuracy gets worse with the increase of the filter order. That is because the problem of error propagation becomes severe with a larger filter order when the applied CSI contains estimation error.

In addition to prediction accuracy, evaluating the performance of two predictors applied in wireless systems is more practically meaningful. Hence, the numerical results on outage probability achieved by a prediction-assisted TAS system having N_t=4 transmit antennas and N_r=1 receive antenna in *i.i.d.* MIMO channels with f_d=100Hz are obtained. Three different antenna selection modes are compared: i) *The perfect mode* that chooses transmit antennas for block $t+D$ according to the perfect CSI at that block, i.e., $\mathbf{H}(t+D)$, although it is practically unavailable. ii) *The outdated mode* in traditional TAS systems where the outdated CSI $\mathbf{H}(t)$ is applied. iii) With the aid of channel prediction, *the prediction mode* makes a decision based on the predicted CSI $\hat{\mathbf{H}}(t+D)$. The default prediction step is set to D=1 since the KF-based predictor can only make one-step-ahead prediction, while D=2 steps prediction for the RNN-based predictor is also tested. In noiseless channels, as shown in Fig.4a, applying channel prediction receives a remarkable performance gain. Observing the level of outage probability at 10^{-4}, the prediction mode has an SNR gain of nearly 9dB over the outdated mode. Using the performance of the perfect mode as a benchmark, the KF-based predictor with p= 4 can achieve the optimal performance and the RNN-based predictor has a suboptimal performance with a small SNR loss of around 1dB. When the predictor step is increased to D=2, the RNN-based predictor gets a larger performance gain over the outdated mode, amounting to an SNR gain of approximately 10dB at the outage probability of 10^{-3}.

Meanwhile, the results of outage probability in noisy channels are derived by means of reusing the datasets employed in Fig.3b. This case, the optimal KF-based predictor is $\mathrm{AR}(2)$, which receives almost the same performance as the RNN-based predictor. Observing the outage probability of 10^{-3}, the prediction mode has a performance gap of around 2.5dB away from the optimal result, whereas it outperforms the outdated mode with an SNR gain of greater than 3dB. In the case of D=2, the RNN-based predictor obtains a gain of around 4dB. Although the noise has a strong negative effect on prediction accuracy, its impact on the performance of the prediction-assisted TAS system is mild. Compared Fig.4b with Fig.4a, the outage probability of the RNN-based predictor degrades from 7×10^{-6} to 4.5×10^{-5} at the SNR of 20dB. However, the KF-based predictor performs unstably in noisy channels and, as exemplified by the curve of $\mathrm{AR}(4)$ in Fig.4b, is sometimes vulnerable to the noise. In a nutshell, the RNN-based predictor not only reaps a comparable performance gain as the KF-based predictor but also exhibits better flexibility (in multi-step predictions) and robustness against additive noise.

V. Conclusions

This paper compared the complexity and performance of two kinds of wireless channel predictors built from the Kalman filter and artificial intelligence. The complexity of the RNN-based predictor is higher than the KF-based predictor, but its required computing resource is still marginal compared with DSP's capability. The numerical results on prediction accuracy and outage probability achieved by a prediction-assisted TAS system revealed that applying channel prediction has a remarkable performance benefit. Although the KF-based predictor is simpler and has a comparable prediction capability as the RNN-based predictor, the latter exhibits better flexibility in multi-step predictions and stronger robustness against additive noise. Moreover, other advanced AI techniques such as long-short term memory and deep learning, may provide a greater potential in channel prediction, which are worth taking efforts to explore in the next step.

References

[1] Y. Huang *et al.*, "Secure transmission in MIMO wiretap channels using general-order transmit antenna selection with outdated CSI," *IEEE Trans. Commun.*, vol. 63, no. 8, pp. 2959–2971, Aug. 2015.

[2] G. Caire *et al.*, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2845–2866, Jun. 2010.

[3] W. Peng *et al.*, "Channel prediction in time-varying Massive MIMO environments," *IEEE Access*, vol. 5, pp. 23 938–23 946, Nov. 2017.

[4] W. Jiang *et al.*, "A robust opportunistic relaying strategy for co-operative wireless communications," *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 2642–2655, Apr. 2016.

[5] A. Weinand *et al.*, "Providing physical layer security for mission critical machine type communication," in *Proc. of IEEE ETFA*, Berlin, Germany, Sep. 2016.

[6] H. D. Schotten *et al.*, "Availability indication as key enabler for ultra-reliable communication in 5G," in *Proc. of European Conf. on Net. and Commun. (EUCNC)*, Bologna, Italy, Jun. 2014.

[7] TACNET 4.0 project. [Online]. Available: http://www.tacnet40.de

[8] T. Eyceoz *et al.*, "Deterministic channel modeling and long range prediction of fast fading mobile radio channels," *IEEE Commun. Lett.*, vol. 2, no. 9, pp. 254–256, Sep. 1998.

[9] H. P. Bui *et al.*, "Performance evaluation of a Multi-User MIMO system with prediction of time-varying indoor channels," *IEEE Trans. Antennas Propag.*, vol. 61, no. 1, pp. 371–379, Aug. 2012.

[10] J.-Y. Wu and W.-M. Lee, "Optimal linear channel prediction for LTE-A uplink under channel estimation errors," *IEEE Trans. Veh. Technol.*, vol. 62, no. 8, pp. 4135–4142, Oct. 2013.

[11] W. Liu *et al.*, "Recurrent neural network based narrowband channel prediction," in *Proc. IEEE Vehicular Tech. Conf. (VTC)*, Melbourne, Australia, May 2006.

[12] C. Potter *et al.*, "MIMO beam-forming with neural network channel prediction trained by a novel PSO-EA-DEPSO algorithm," in *Proc. IEEE Intl. Joint Conf. on Neural Networks (IJCNN)*, Hong Kong, China, Jun. 2008.

[13] K. T. Truong and R. W. Heath, "Fading channel prediction based on combination of complex-valued neural networks and chirp Z-transform," *IEEE Trans. Neural Netw.*, vol. 25, no. 9, pp. 1686–1695, Sep. 2014.

[14] W. Jiang and H. D. Schotten, "Multi-antenna fading channel prediction empowered by artificial intelligence," in *Proc. IEEE Vehicular Tech. Conf. (VTC)*, Chicago, USA, Aug. 2018.

[15] ——, "Neural network-based channel prediction and its performance in multi-antenna systems," in *Proc. IEEE Vehicular Tech. Conf. (VTC)*, Chicago, USA, Aug. 2018.

[16] J. Connor *et al.*, "Recurrent neural networks and robust time series prediction," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 240–254, Mar. 1994.

[17] K. Baddour and N. Beaulieu, "Autoregressive modeling for fading channel simulation," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1650–1662, Jul. 2005.

[18] X. Fu *et al.*, "Training recurrent neural networks with the Levenberg Marquardt algorithm for optimal control of a grid-connected converter," *IEEE Trans. Neural Netw.*, vol. 26, no. 9, pp. 1900–1912, Sep. 2015.