NONLINEAR PREDICTION OF MOBILE RADIO CHANNELS: MEASUREMENTS AND MARS MODEL DESIGNS

Torbjörn Ekman[1*] *Gernot Kubin*[2]

[1] Signals and Systems Group, Uppsala University, Pb 528, SE-751 20 Uppsala, Sweden (`torbjorn.ekman@signal.uu.se`)
[2] Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology
Gusshausstrasse 25 /389, A-1040 Vienna, Austria (`g.kubin@ieee.org`)

ABSTRACT

The rapid time variation of mobile radio channels is often modeled as a random process with second order moments reflecting vehicle speed, bandwidth and the scattering environment. These statistics typically show that there is little room for prediction of channel properties such as received power or complex taps of the impulse response coefficients, at least when linear predictor structures are considered. We use mutual information estimation to measure statistical dependencies in sequences of wideband mobile radio channel data and find significant nonlinear dependencies, far exceeding the linear component. Based on these upper limits for the predictability of channel evolution over time intervals up to 30 ms ahead, we develop practical nonlinear predictor systems using Multivariate Adaptive Regression Splines (MARS). We demonstrate computationally efficient schemes that increase the prediction horizon beyond 10 ms, compared to less than 4 ms with linear predictors at comparable prediction gains.

1. INTRODUCTION

Time variation of a mobile radio channel is induced by movement of a mobile transceiver through an environment with scatterers. The mobile travels through an interference pattern that causes fading. It is reasonable to assume that, within some small local area or time frame, the scattering geometry is time invariant. A low-dimensional deterministic mapping may thus exist from the channel transmission properties measured at one time instance to the same properties measured a moment later, when the mobile has moved a short distance.

A common approach is to assume that a large number of horizontal planar waves with vertical polarization arrives at the receiver from different directions in a plane [1]. The degree of complexity quickly grows when changing polarizations, different elevation angles and non-planar waves are taken into consideration. Instead of trying to parameterize the physics, we view the process as a dynamical system producing an observed time series at its output. For the mobile radio channel, this can be either a sequence of received power measurements or sequences of estimates of a single or multiple coefficients of the time-varying impulse response.

One of the powerful results from the theory of deterministic dynamical systems is Takens embedding theorem [2], which allows the reconstruction of a state space representation from a scalar signal alone. As a precondition, this signal must have been observed from the output of a dynamical system in steady state motion on a finite-dimensional manifold in its state space (i.e. on an attractor). The reconstruction is achieved by combining several lagged samples of the signal into a signal vector whose dimension is more than twice the attractor dimension. Following the trajectory of this vector over time, the state space evolution is reconstructed in this embedding space, allowing us to analyze and model many signal properties the same way as when using physical state space coordinates.

In this paper we study how far ahead in time the received power and the complex taps of the impulse response are predictable. This information is important for adaptive resource allocation and power control but could also be used for adaptive coding/modulation and equalization schemes. Mutual information measurements on measured channel impulse responses and on the received power provide upper bounds on the achievable prediction gain (PG). Linear and nonlinear predictors are tested on measurements to find practical measures on the achievable PG for different prediction intervals. Neural nets have been proposed for power prediction in wideband systems and have been demonstrated to perform better than linear predictors [3] for 1 ms horizons, whereas we go to 10-30 ms. Here we use MARS models [4] as the nonlinear systems used to identify the dynamical characteristics of the channel.

2. AN UPPER BOUND ON THE PREDICTION GAIN

In prediction of a discrete time series y_t, the values of y_{t+L} are forecasted using present and past values, represented by the vector $\mathbf{y}_t = [y_t, y_{t-\tau_1}, \dots, y_{t-\tau_{p-1}}]$. A predictor estimates y_{t+L} as

$$\hat{y}_{t+L|t} = \hat{f}(\mathbf{y}_t). \tag{1}$$

The prediction error is given as the difference between observed and predicted values

$$e_t = y_t - \hat{y}_{t|t-L}. \tag{2}$$

*This work was done while T. Ekman was with the Inst. of Commun. and Radio-Frequency Engin., Vienna Univ. of Technology. Support by the EU Marie Curie Research Training Grant is gratefully acknowledged. We thank Ericsson Radio Systems AB (Kista, Sweden) for supplying measurement data, J.M. Friedman for the MARS 3.6 software and H.-P. Bernhard for assistance with his mutual information estimation software.

The PG $G(L)$ is the ratio between the variance of y_t and the power of the resulting error e_t

$$G(L) = 10\log_{10}\frac{E\{(y_t - m_y)^2\}}{E\{e_t^2\}} \quad (3)$$

measured in dB. Here m_y denotes the average value of the time series, $m_y = E\{y_t\}$. Under the assumption of stationarity a tight upper bound on $G(L)$ for an unbiased predictor is given by [5]

$$G(L) \leq G_{max}(L) = 6.02(I(y_{t+L}; \mathbf{y}_t) + \Delta) \quad (4)$$

where

$$\Delta = \frac{1}{2}\log_2(2\pi e E\{(y_t - m_y)^2\}) - H(y_t) \quad (5)$$

is the difference between the differential entropy of a Gaussian variable with the same variance as y_t and the first-order entropy, $H(y_t)$. The mutual information $I(y_{t+L}; \mathbf{y}_t)$ is a measure of how much information the vector $\mathbf{y}_t$ contains about the value to be predicted, y_{t+L} [6]. $G_{max}(L)$ is a tight upper bound on the PG for any predictor forecasting y_{t+L} using $\mathbf{y}_t$. Both the differential entropy $H(y_t)$ and the mutual information $I(y_{t+L}; \mathbf{y_t})$ can be estimated from a single realization of a stationary ergodic process using the fast algorithm described in [7].

3. MARS MODELS

Multivariate Adaptive Regression Splines (MARS) have been proposed by Friedman [4] to build models of the relationships between a scalar response variable and multiple regression variables. Lewis and Stevens [8] used MARS for nonlinear threshold modeling of time series. For a predictor, the response variable is chosen as y_{t+L} and the regression variables are delayed values of the time series, $\mathbf{y}_t = [y_t, y_{t-\tau_1}, \ldots, y_{t-\tau_{p-1}}]$. To build the MARS model, the multi-dimensional regression variable space is partitioned into subregions by recursive one-dimensional splits where each split is associated with a one-dimensional, linear spline function. Unlike set-theoretic partitions, the subregions may overlap due to their specific (suboptimal) iterative construction: at each iteration step, all available subregions are split along all dimensions (one at a time) and only the one refinement that provides the greatest increase in regression accuracy is retained. The basis function for a subregion is formed by multiplication of all the associated splines. In mathematical terms the model is described as below. Assume that y_{t+L} can be described by the regression model

$$y_{t+L} = f(\mathbf{y}_t) + \epsilon. \quad (6)$$

The MARS estimate of the unknown function $f(\mathbf{y_t})$ is

$$\hat{f}(\mathbf{y}_t) = a_0 + \sum_{m=1}^{M} a_m B_m(\mathbf{y}_t) \quad (7)$$

where $\hat{f}(\mathbf{y}_t)$ is a sum of weighted basis functions, $\{B_m\}$, associated with subregions $\{m\}$. A basis function is formed by multiplication of truncated linear spline functions $\{T_{m,l}\}$:

$$B_m(\mathbf{y}_t) = \prod_{l=1}^{K_m} T_{m,l}(\mathbf{y_t}) \quad (8)$$

where K_m is the level of interaction for the basis function of region m, describing how many truncated splines are used to build B_m. Each truncated linear spline function $T_{m,l}$ works along one dimension only (that is one delay τ_v) and has a partitioning point at $y_{t-\tau_v} = t_{m,l}$, partitioning the regression variable space:

$$T_{m,l}(\mathbf{y}_t) = [s_{m,l}(y_{t-\tau_v} - t_{m,l})]_+ \quad (9)$$

where $s_{m,l} = \pm 1$ gives the left or right side of the knot $t_{m,l}$. In (9), $[\cdot]_+$ denotes the half-wave rectifier function, i.e. it takes the value of the argument if it is positive and is zero otherwise. For $s_{m,l} = +1$, $T_{m,l}(\mathbf{y}_t)$ is positive for $y_{t-\tau_v} - t_{m,l} > 0$ and zero otherwise. For $s_{m,l} = -1$ the inequality is turned the other way. The MARS algorithm iteratively builds up the model structure (by enlarging the set of subregions while proceeding to higher levels of interaction among regression variables) and adjusts the thresholds ($t_{m,l}$) and weighting coefficients (a_i) to fit the data, minimizing the mean-squared error.

The models are continuous input-output maps and can handle linear systems as well as nonlinear systems with more complex behavior such as limit cycles etc. MARS models are more efficient than neural networks (NN) such as standard multi-layer perceptrons with backpropagation for several reasons: they include a bottom-up strategy to build up the model structure until a certain level of accuracy is achieved (with 'optimal' interaction of inputs rather than pruning a highly redundant, fully interconnected NN), they are good in approximating (locally) linear mappings, their digital implementation is very simple and computationally efficient (only hard thresholds and multiply/adds, no sigmoids involving transcendental functions). A predictor based on a MARS model with the highest level of interaction set to K and no more than M basis functions has at most $M(K+1)+1$ additions, $M(K+1)$ multiplications, MK threshold operations and uses no more than $M(K+1)+1$ parameters.

4. EXPERIMENTS AND RESULTS

4.1. Measurements

Wideband radio channel measurements were performed at 1880 MHz, at distances of 200 to 2000 m from the base station antenna placed on a high roof-top. The mobile antenna was placed on a car driving in a suburban environment mostly at non-line of sight. Vehicle speed varied between 30 to 90 km/h. A total of 30 measurement runs were recorded at different positions. The measurements consisted of a repeatedly transmitted sequence of length 109 μs, resulting in 156 ms continuous received signal at each measurement location. The baseband sampling rate of the receiver was 6.4 MHz.

The impulse response is estimated using a 120 tap FIR-model, covering 18.75 μs, which is fitted by least squares to each repeated sequence. For the identification, the signal from a back-to-back measurement is used as reference. This resulted in 1430 consecutive complex impulse response estimates at each measurement location. The method of identification is unbiased even when the noise is colored and it is chosen to give estimates with low variance for the dominant taps. In 25 out of the 30 measurement runs, the largest peak of the power delay profile is at least 20 dB over the noise floor. Only those runs are retained for further analysis and modeling.

Three different bandwidths are examined, 5, 2.5 and 1.25 MHz, all wideband. The different bandwidths are achieved by filtering and subsampling. The resulting sampling rates are 6.4, 3.2 and 1.6 MHz respectively, for the three considered bandwidths.

4.2. Power Prediction

For prediction of the received power, a MARS model and two linear predictors are compared. In all the experiments the state variable vectors spans 15 ms which corresponds to a memory of 137 samples and the prediction interval L is varied from 1 ms up to 10 ms. All prediction experiments are done on detrended data with zero average. The MARS model for the power uses 6 delayed signal samples, uniformly spaced over the 15 ms memory, as regression variables. The maximum number of basis functions are 20 and the level of interaction 3. This structure is chosen to be rich enough to model a great variety of fading patterns. Linear predictors are two FIR filters, using either all 137 samples in the memory or only the 6 samples used by the MARS model.

To get an estimate of how predictable the time series are, the mutual information between the power at $t+L$ and two delayed measurements $\mathbf{y}_t = [y_t, y_{t-T}]$ is estimated. The reason to use only two delays is to keep the accuracy of the mutual information estimate high and still have the extra structural information the second delay gives. Using (4) an upper bound on the PG for any predictor using two lagged variables within the memory is estimated. The delay spacing T is chosen to give the highest PG within the memory ($T \leq 15$ ms). In Fig. 1, the average of the optimal delay spacing is presented for the different bandwidths and prediction intervals. Even though the prediction interval increases from 1 ms to 30 ms the average delay spacing giving the highest PG increases less than 5 ms. Thus when the prediction interval is increased the memory size does not have to increase accordingly.

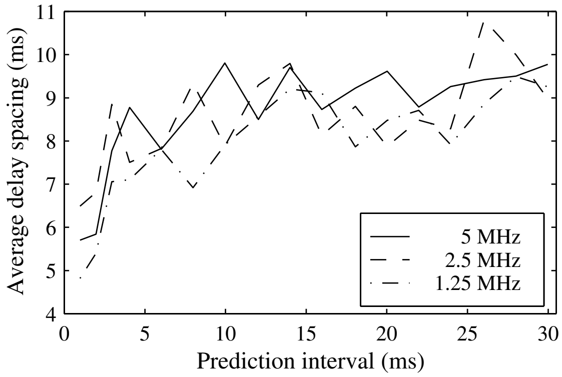

Figure 1. Average delay spacing over all measurements giving the highest PG for the power, estimated with the mutual information algorithm.

As seen in Fig. 2 the long linear filter has higher PG for short prediction intervals than the other predictors. This is because the long linear filter captures the local smoothness, which the subsampled short FIR filter and the MARS model are unable to do. Going to longer prediction intervals the dynamics of the system becomes important and the MARS predictor outperforms the linear predictors. The PG

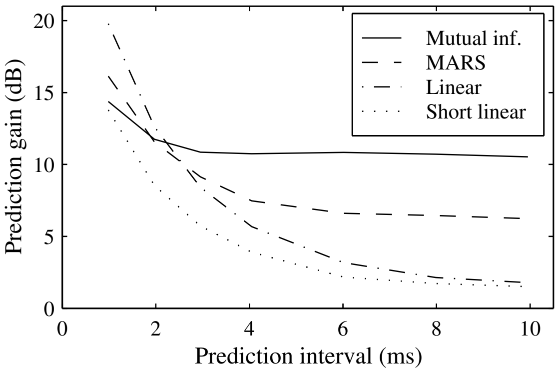

Figure 2. Average PGs for the power at 1.25 MHz bandwidth, using mutual information estimate of PG with two delays, MARS predictor (6 input variables), long linear predictor (137 inputs) and the short linear predictor (same 6 inputs as MARS).

bound for predictors using two lagged variables, estimated through the mutual information, first drops 3-4 dB when the prediction interval is increased from 1 to 4 ms and then remains fairly constant (also seen in Fig. 3). This indicates that when the dynamics are known it is possible to increase the prediction interval with only a small loss of accuracy. The drop in the PG bound when reducing the bandwidth

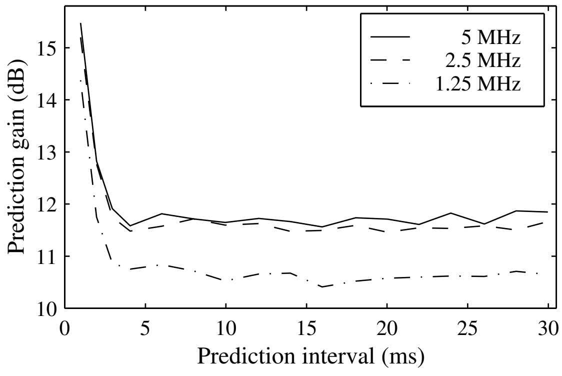

Figure 3. Average PG bound for the power over all power measurements, estimated from mutual information using two delays.

is due to the deeper fades and more rapid changes for lower bandwidths.

4.3. Prediction of Channel Parameters

Predictors for the largest taps in the complex impulse response, for each of the 25 measurements, are designed and tested. The number of taps used decreases with the bandwidth from 9 taps for 5 MHz to 5 taps for 1.25 MHz.

For MARS modeling of the complex taps, two models have to be built, one each for the real and imaginary parts. Both models use the same input of 6 delayed measurements of a single complex tap, separated in real and imaginary parts, resulting in 12 inputs. The 6 delays are uniformly spaced over the memory. The maximum number of basis functions are 20 and the level of interaction 2, allowing for multiplicative interactions. On average 57 real-valued parameters are used by the MARS predictor. The linear predictors have the same structure as for power prediction. Here the use of complex signals results in predictors with either 137 or 6 complex parameters. Not only the complex coefficients are predicted, also the power of the taps , that is the squared magnitude of the coefficients, are predicted as shown in Fig. 4. The MARS prediction captures the fading pattern even though the prediction interval is as long as 10 ms, which is more than the typical duration of a fade.

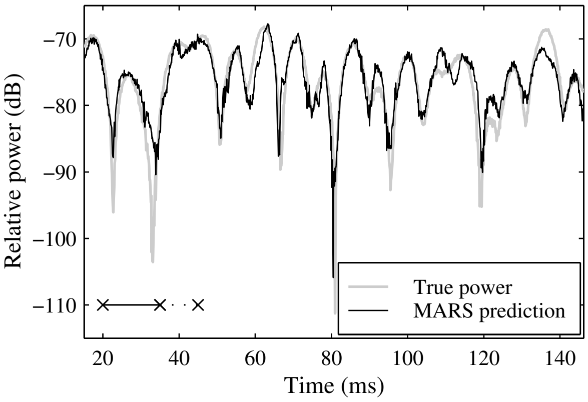

Figure 4. True and predicted squared magnitude (PG is 5.8 dB) for the largest tap in one measurement at 5 MHz bandwidth. The predicted power is calculated from the complex tap, predicted 10 ms ahead with a MARS predictor. The memory (solid line) and prediction interval (dotted line) are indicated in the lower left corner.

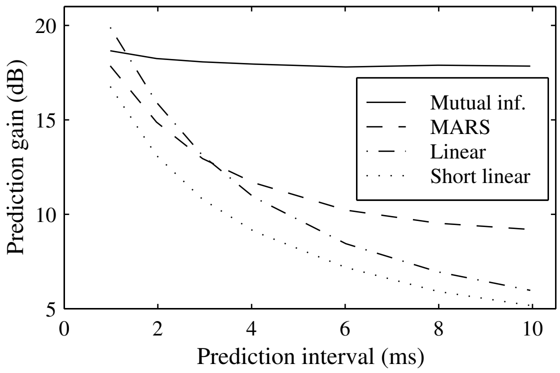

Figure 5. Average PG for the measured complex taps at 1.25 MHz bandwidth. Mutual information estimate of PG with two delays, MARS predictor, linear predictor and a short linear predictor using the same delayed variables as the MARS model.

The PG for the different coefficient predictors have a similar behavior as the power prediction (see Fig. 5). The MARS predictor on average gives 3-4 dB higher PG than the long linear predictor for prediction intervals of 10 ms, and the mutual information measurements indicate that it should be possible to gain significantly more with the right choice of nonlinear structure.

5. CONCLUSIONS

A nonlinear approach to prediction of received power and the complex coefficients in the impulse response of a mobile radio channel is proposed. MARS models, using only a few inputs from the memory of the signal, are able to capture the dynamics of the fading channel better than linear filters using all samples in the same memory. It is possible to predict received power and channel coefficients in the impulse response 10 ms ahead at prediction gains averaging around 6 dB for the power and 9-11 dB for the channel coefficients (the higher value is for 5 MHz bandwidth), at frequencies in the 1800 MHz band. Mutual information measurements on measured impulse responses and power indicate that even higher PG and longer prediction horizon should be attainable.

As the MARS predictor has a low complexity, comparable to a linear predictor, it is possible to use it on line. The updating of the model is more computationally demanding but can be done at a lower rate. How often the model has to be updated is a question for further research.

REFERENCES

[1] Jr W.C. Jakes, "Multipath interference," in *Microwave Mobile Communications*, Jr W.C. Jakes, Ed. IEEE, Piscataway, NY, 1974.

[2] S. Haykin and J. Principe, "Making sense of a complex world," *IEEE Signal Processing Magazine*, vol. 15, no. 3, pp. 66–81, May 1998.

[3] X.M. Gao, J.M.A. Tanskanen and S.J. Ovaska, "Comparison of linear and neural network-based power prediction schemes for mobile DS/CDMA systems," in *Proc. IEEE* 46^{th} *Vehic. Tech. Conf.*, Atlanta, GA, May 1996, pp. 61–65.

[4] J.M. Friedman, "Multivariate adaptive regression splines," *The Annals of Statistics*, vol. 19, no. 1, pp. 1–141, 1991.

[5] H.-P. Bernhard, "A tight upper bound on the gain of linear and nonlinear predictors for stationary stochastic processes," *IEEE Transactions on Signal Processing*, vol. 43, November 1998, to appear.

[6] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, Wiley, New York, 1991.

[7] H.P. Bernhard and G. Kubin, "A fast mutual information calculation algorithm," in *Signal Processing VII: Theories and Applications*, M.J.J. Holt et al., Ed., vol. 1, pp. 50–53. Elsevier, Amsterdam, Sept. 1994.

[8] P.A.W. Lewis and J.G. Stevens, "Nonlinear modeling of time series using multivariate adaptive regression splines (MARS)," *Journal of the American Statistical Association*, vol. 86, no. 416, pp. 864–877, Dec. 1991.