UPPSALA UNIVERSITY

Prediciton of Mobile Radio Channels

Torbjörn Ekman

November 2000

Signals and Systems
Uppsala University
Uppsala, Sweden

Submitted to the Faculty of Science and Technology, Uppsala University in partial fulfillment of the requirements for the degree of Technical Licentiate in Signal Processing.

Printed in Sweden by Lindbergs Grafiska HB, Uppsala 2000
Distributed by Signals and Systems, Uppsala University, Uppsala, Sweden

To whom it may concern

Abstract

Prediction of the coefficients of mobile radio channel is of interest for a range of applications such as power control, adaptive resource allocation, adaptive coding and modulation. Power control in e.g. WCDMA requires short-term prediction over only a small fraction of the distance between two dips in the short-term fading pattern. Radio resource allocation and planning would require accurate and more long-term prediction, the longer the better.

The performance of different predictors for the mobile radio channel are evaluated partly on simulated data, using a spherical wave propagation model, but mainly on measured broadband channel impulse responses from a suburban environment. The focus is mainly on adaptive and non-adaptive linear FIR predictors but quadratic-Volterra and MARS predictors are also studied.

The received power of a mobile radio channel is predicted as the sum of the squared magnitudes of the predicted individual complex taps in the channel impulse response. The linear adaptive iterated subsampled FIR predictor generally produces excellent predictions of both complex taps and total power for short ranges, that is up to 0.1 wavelengths. The performance of power predictors is reduced markedly at ranges over half a wavelengths. The advantage over using just the average power for prediction then becomes small, so we can not claim that the investigated predictors are efficient for these prediction ranges.

Acknowledgments

It is all about communication, connecting people. I would especially like to express my gratitude for the communication with and connection to the following people.

My patient supervisors, Prof. Anders Ahlén and docent Mikael Sternad for thoroughness and enthusiasm, respectively. Prof. Gernot Kubin for extending my views with good humour. All three have been highly involved in the work leading to this thesis.

Dr. Andreas Jakobsson for invaluable support. M.Sc. Nilo Casimiro Ericsson for applications. Dr. Claes Tidestav for relevance. Dr. Hans-Peter Bernhard for analysis method. Dr. Erik Lindskog for intriguing discussions. Prof. J. Bach Andersen for clarification. Prof. Petre Stoica for his fast mind and thourough knowledge. Sören Anderson, Henrik Asplund and Jan-Erik Berg at Ericsson Radio Systems AB, for providing measurement data.

All the people at the groups Signals and Systems, System and Control at Uppsala University, where I always have felt at home. The colleagues at the Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology, where I spent some of my most exciting years, hitherto. The Digital Signal Processing group at the University of Oslo, for their kind hospitality.

Kathrine Vangen for making it all worth while.

Torbjörn Ekman
Oslo, November 2000.

Contents

Chapter 1

Introduction

Mobile radio communication makes it possible to communicate with people on the move. The communication is facilitated by the transmission of radio signals between a fixed base station and a mobile station, which often is a mobile phone. The problem with radio communication is that the receiving condition depends on where you are and changes as you move along. This is due to that the channel changes depending on the position of the mobile station. The channel and the level of noise (mostly interference from other transmitters) sets the limit for the amount of information that can be transmitted between the base station and the mobile station.

Today mobile communication is much more than just talking in a phone. The extensive use of text messages such as SMS, shows that people actually are interested in using their mobile phones as information terminals. It is assumed that, in the near future, there will be a demand for mobile internet and high bit-rate applications. There is thus a need to send as much data over a given mobile radio channel as possible. If the channel is known in advance, then the transmission scheme could be adjusted to fully exploit the current receiving condition and thus increase the average amount of transmitted data. This idea could be even further exploited if many users simultaneously share the same base-station and radio frequency.

1.1 Mobile Radio Channels

There are two different types of variation in the receiving conditions: *large-scale fading* and *small-scale fading*.[1] This thesis treats the problem of pre-

[1] Also called slow and fast fading respectively

dicting the small-scale fading. In both cases it is the environment of scatterers that determines the receiving conditions.

Large-scale fading determines the average received power at a geographical point. It is caused by e.g. shadowing of the transmission from the base station from buildings or other large objects. Large-scale fading thus remains fairly constant when the mobile station moves a short distance.

Small-scale fading causes dramatic variations in the received power when the mobile station moves a fraction of a wavelength of the transmitted radiowave. The small-scale fading is caused by the multiple paths for the transmitted radio signal. The transmitted signal can travel in indirect paths via e.g. reflections from, or diffraction around, buildings or other man-made or natural objects. This causes the transmitted radio signal to arrive at the receiver with different delays from different directions, an effect called *multipath propagation*. The radio waves from the different paths interact and form constructive and destructive interference, corresponding to high or low received power at different locations in space. The time variability of a mobile radio channel is induced by the movement of a mobile transceiver through a multipath environment. The small-scale fading is thus due to that the mobile station travels through an interference pattern.

A common approach to describe the channel is to assume that the transmitted radio waves, causing the interference pattern, act as plane wave fronts from different directions, due to the multipath propagation. A statistical description of the mobile radio channel has then been given by Jakes [1] and Clarke [2]. The properties of the channel is deduced from a scattering propagation model which assumes that the field incident to the receiver antenna is composed of an infinite number of randomly phased azimuthal plane waves of arbitrary azimuthal angles.

1.2 Prediction

If the number of contributing wave fronts are limited and the plane wave assumption is valid, then the contribution to the channel from each wave front can be modeled as a complex sinusoid. This is an approximation with limited validity, as we will see in Chapter 2. Given that the channel acts as a sum of sinusoids the phase shift and attenuation of each path can be estimated. Then the prediction can be obtained by propagating the estimated sinusoids. This is the most common approach to prediction of mobile radio channels. Proposed approaches to the estimation of the sinusoids are root-MUSIC [3] and a modified ESPRIT algorithm [4]. A different approach is to model the sum of sinusoids as an AR-process. The estimated AR-parameters are then used to derive a linear predictor [5]. Overviews of these methods are found in [6] and [7].

The approach presented in this thesis is to use delay embeddings of measured data for model design, focusing *directly* on the prediction. The performance will not be limited by the validity of a theoretical model, but by the ability of the used predictor to model the dynamical behavior in the data. Here, as in [5], linear FIR-predictors are used, but nonlinear predictors are also studied.

The problem of power prediction is closely related to channel prediction as it is the transmission capability of the channel that limits the received power. Neural nets have been proposed for power prediction in wideband systems [8] and have been demonstrated to perform better than linear predictors [9] for 1 ms prediction horizons. Power prediction is desired to achieve better power control, which is crucial in e.g. a CDMA system.

To be able to exploit the communication link as much as possible it would be useful to know the receiving conditions in advance. The ability to predict the channel is thus important for adaptive resource allocation and power control but could also be of use for adaptive coding and modulation [12], [13]. For scheduling and adaptive resource allocation longer prediction horizons, on the order of 5-10 ms are of interest [14].

1.3 Outline of the Thesis

The thesis discusses the problem of predicting the time varying channel impulse responses of mobile radio channels. The only information used for the prediction are noisy estimates of past impulse responses.

As mentioned above the channel can be modeled as the sum of complex sinusoids. In Chapter 2 the approximations leading to this model are discussed. Relaxing the constraints of constant speed and plane wave fronts by including circular motion and spherical wave propagation, leads to models with quadratic terms in the phase functions of the multipath components. The most critical deviation is due to close scatterers causing spherical waves. It is in those cases more appropriate to model the channel impulse response as a weighted sum of complex sinusoids and linear chirps.

The performance of different predictors are evaluated partly on simulated data, using a spherical wave propagation model described in Chapter 2. However, most of the evaluations are performed on measured channel impulse responses.[2] In Chapter 3 the recording and identification of channel impulse responses are treated. The data base consist of a number of Broadband channel-sounder measurements are performed at a number of locations in a suburban environment. The impulse responses are modeled as FIR filters which are estimated from the measured data using least squares. It is demonstrated that the variance of the estimation error can be estimated as the noise floor in the power delay profile or the Doppler spectrum. To reduce the noise on the estimated impulse responses, we propose to use simple Wiener-smoothers, that give a very small delay.

In Chapter 4 we study the performance of time invariant predictors. Three predictor structures are examined: a linear FIR-predictor, a quadratic Volterra-predictor and a MARS predictor. The non-linear predictors show good modeling properties but poor generalization capabilities.

In Chapter 5 we investigate the extension of the time invariant linear predictor to the adaptive predictor. An adaptive Kalman filter is used to estimate the coefficients of the predictor. The Kalman filter uses an AR1I-hypermodel for the variation of the predictor coefficients. The performance of the adaptive predictor on measured impulse responses is discussed in Chapter 6. The linear adaptive iterated subsampled FIR predictor generally produces excellent predictions of both complex taps and total power for short ranges, that is up to 0.1 wavelengths. The prediction gain is still good for 0.2-0.3 wavelengths but falls off rapidly with increasing prediction range.

For short ranges the proposed adaptive predictor gives satisfactory prediction accuracy. This thesis does however not provide a viable method for very long range prediction of mobile radio channels.

[2]The author is grateful to Ericsson Radio Systems in Kista for providing these measurements.

Chapter 2

Model of the Channel

This chapter begins with an overview of the commonly used continuous-time and discrete-time models of terrestrial radio communication channels fore mobile communication. Such models are based on assumptions of planar wave-fronts, constant vehicle velocity and propagation via reflectors and scattering objects. The channel is then described as a weighted sum of complex sinusoids with fixed frequencies. The approximations and assumptions in the derivation of this model are discussed.

From linear systems theory it is known that such a channel can be perfectly extrapolated in space and time, i.e. predicted, if the number of sinusoids are known and the predictor is properly tuned.

In Section 2.3, the limitations of this simple model is discussed and a class of models with more general validity is proposed. By taking spherical wavefronts and non-constant vehicle velocities into account, a model is obtained in which the sinusoids may have time-varying frequencies, such as chirps. While it would be unrealistic to expect unbounded predictability in such scenarios, it is still reasonable to believe that prediction over a large fraction of a wavelength of the carrier frequency may be attainable. The rest of this thesis constitutes an initial exploration of methods with which such long-range predictions might be realized.

We will in the following always refer to the moving transmitter/receiver as the mobile station. Whether it acts as a receiver or transmitter does not matter, as the channel is the same in both directions due to reciprocity. Thus both the up- and down-links are treated.

2.1 Multipath Propagation via Reflectors and Scatterers

In the classical papers by Clarke [2] and Jakes [1], a statistical passband description of the mobile radio channel is developed. The model treats a scenario where N plane waves are arriving at the receiver from random directions. The different propagation paths cause the waves to have different attenuation and phase shifts. In Figure 2.1, a scenario with a mobile receiving waves from two different directions is depicted. The base-station antenna acts as a point source emitting a spherical electromagnetic wave. On large distances from the antenna the spherical waves are locally perceived as plane waves. In Figure 2.1 we have no line of sight due to the shadowing of a building. The buildings further away will also contribute, although the longer propagation paths result in a higher path loss. In Figure 2.1

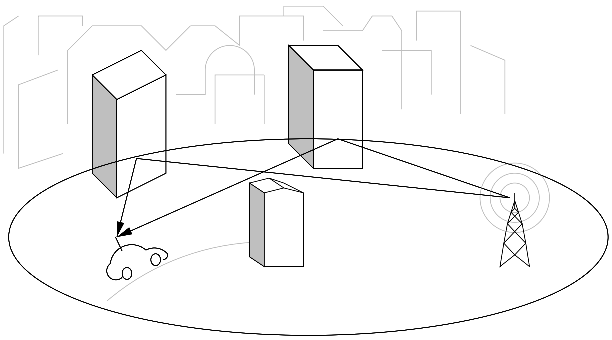

Figure 2.1: Mobile receiving two reflected rays that have the same path distance.

two buildings act as *remote reflectors* and the corresponding wave-fronts are considered to be almost plane at the mobile station. A lamp-post nearby the mobile would act as a *close scatterer*. Close scatterers cause spherical wave-fronts at the mobile station, whereas rays that are scattered around corners of houses close by cause cylindrical wave-fronts. Further away from the scatterers and corners the corresponding spherical and cylindrical waves are locally perceived as plane waves. The characteristic size and distance of objects that make them behave as reflectors rather than scatterers is discussed in Appendix 2.A.

2.2 The Continuous Time Channel for Plane Wave Fronts

The voltage, $V(\mathbf{r}, t)$, which can be measured at an antenna is a function of the spatial electric field and the antenna properties. The incident field at the receiver, $\mathbf{E}(\omega, \mathbf{r}, t)$, is a function of the angular frequency ω, the position $\mathbf{r}$ and time t.[1] We represent a transmitted single-frequency radio signal as a complex exponential,[2] $e^{j\omega t}$. The open circuit voltage at the receiving antenna can then be approximated as the product between $e^{j\omega t}$ and a space-dependent transfer function, $H(\omega, \mathbf{r})$, such that

$$V(\mathbf{r}, t) = H(\omega, \mathbf{r}) e^{j\omega t}. \tag{2.1}$$

The wavefronts contributing to the electric field can be thought of as rays with different delays coming from different directions. The transfer function $H(\omega, \mathbf{r})$ is therefore approximated as the summation of the contributions from reflectors and scatterers according to

$$H(\omega, \mathbf{r}) = \sum_{n=1}^{N} a_n(\mathbf{r}) e^{j(\psi_n(\mathbf{r}) - k r_n(\mathbf{r}))}. \tag{2.2}$$

Here $k r_n(\mathbf{r})$ is the (scalar) electrical distance (with $k = 2\pi/\lambda = \omega/c$ being the wave number, where c is the velocity of light) and $r_n(\mathbf{r})$ is the physical distance of path n. The number of contributors N may be arbitrary large. The effective complex amplitude of contributor n is $a_n(\mathbf{r}) e^{j\psi_n(\mathbf{r})}$. It includes, among other things, antenna pattern weighting effects and path attenuation. The attenuation $a_n(\mathbf{r})$ and the phase shift $\psi_n(\mathbf{r})$ depend on physical parameters such as path distance and the texture of the reflectors. These parameters are fairly constant over short distances. Therefore we shall regard $a_n(\mathbf{r})$ and $\psi_n(\mathbf{r})$ as being space independent, at least over distances of a few wavelengths. Thus we set

$$a_n(\mathbf{r}) = a_n, \tag{2.3}$$
$$\psi_n(\mathbf{r}) = \psi_n. \tag{2.4}$$

In addition we also assume a_n to be independent of ω, at least over the bandwidth of the system around the carrier frequency.

[1] The origin of the coordinate system for the position is arbitrary.

[2] The electric and magnetic fields are the real and imaginary parts respectively

An alternative description of the exponential term of (2.2) uses the scalar product between the wave-vector of the nth path $\mathbf{k}_n(\mathbf{r})$, and the position $\mathbf{r}$,

$$H(\omega, \mathbf{r}) = \sum_{n=1}^{N} a_n(\mathbf{r}) e^{j(\varphi_n(\mathbf{r}) - \mathbf{k}_n(\mathbf{r}) \cdot \mathbf{r})}. \tag{2.5}$$

The phase $\varphi_n(\mathbf{r})$ in (2.2) includes the phase shift due to reflection and/or scattering just as $\psi_n(\mathbf{r})$ but $\varphi_n(\mathbf{r})$ also includes a spatially invariant contribution that depends on the choice of the origin of the coordinate system. The two expressions (2.2) and (2.5) are fully equivalent but can be used to illustrate different properties of the channel.

A more detailed description of the exponential term of (2.2) will be derived next.

The electrical distance and the phase of a received sinusoid

The model (2.1), (2.2) is valid in a local region if the phase, $\psi_n(\mathbf{r})$, and amplitude $a_n(\mathbf{r})$, for each scatterer remain constant in that region. Define a right hand coordinate system with the x-axis along the direction of movement and the y-axis pointing to the left in the plane on which the mobile station is moving (the z-axis will then point up from the plane). Thus the position along the direction of movement is x ($\mathbf{r} = [x, 0, 0]$).

The electrical distance between the base station and the mobile along path n will be a function of x. It can be expressed as

$$kr_n(x) = \frac{2\pi}{\lambda} r_n(x) = \omega \frac{r_n(x)}{c} = \omega \tau_n(x), \tag{2.6}$$

where $r_n(x)$ as before is the location-dependent physical distance between the base station and the mobile station along path n and $\tau_n(x)$ is the corresponding path delay.

The difference in electrical distance between two points corresponds to the phase difference for a sinusoid carrier signal. We will below derive an expression for the difference in electrical distance when the mobile moves a distance x at an angle θ_n to the direction of the scatterer/reflector which is causing the incoming wave front. The path distance from the closest point of reflection will then change from $r_n(0)$ to $r_n(x)$ according to

$$r_n^2(x) = r_n^2(0) + x^2 - 2xr_n(0)\cos(\theta_n). \tag{2.7}$$

A second-order Taylor expansion of the square root of this expression around $x = 0$ gives

$$r_n(x) \approx r_n(0) - x\cos(\theta_n) + \frac{x^2 \sin^2(\theta_n)}{2r_n(0)}. \tag{2.8}$$

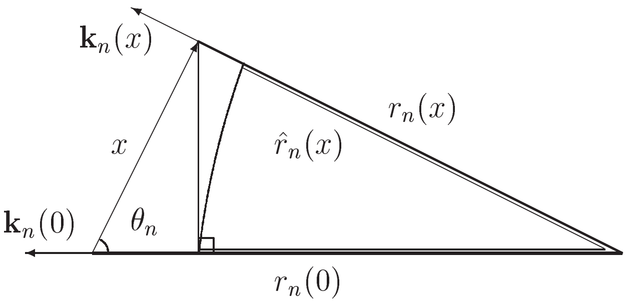

Figure 2.2: The change of path distance when moving the mobile. Here $\hat{r}_n(x)$ is the approximate and $r_n(x)$ is the true distance after moving the distance x.

The corresponding change of electrical distance (the change of phase) from position 0 to x is thus approximated by

$$kr_n(x) - kr_n(0) \approx \frac{\omega}{c}\left(-x\cos(\theta_n) + \frac{x^2\sin^2(\theta_n)}{2r_n(0)}\right) \tag{2.9}$$

If the last term in (2.8) is disregarded, then the common linear approximation, cf. Figure 2.2, is obtained as

$$\hat{r}_n(x) = r_n(0) - x\cos(\theta_n). \tag{2.10}$$

This is the plane wave approximation which is valid if the mobile is moving over distances which are small in comparison to the distance to the reflector. For a plane wave the sides in the triangle in Figure 2.2 would be parallel (and thus not forming a triangle) as in Figure 2.3. The corresponding change of

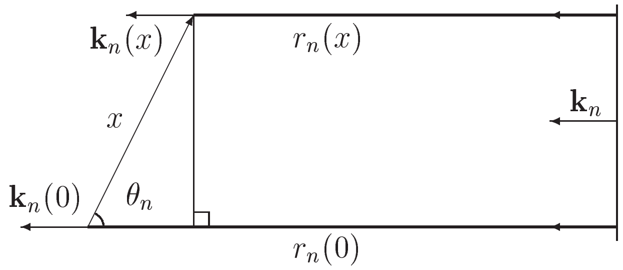

Figure 2.3: The change of path distance when moving the mobile a distance x in a plane wave scenario.

electrical distance is then

$$kr_n(x) - kr_n(0) = \omega\tau_n(x) - \omega\tau_n(0) \approx -\frac{\omega x\cos(\theta_n)}{c}, \tag{2.11}$$

where $\tau_n(\cdot)$ denotes the path delay associated with $r_n(\cdot)$. The change of electrical distance (2.11) can also be easily derived using (2.5) assuming plane waves, that is constant wave-vectors $\mathbf{k}_n$, as

$$\begin{aligned}\mathbf{k}_n \cdot \mathbf{r}(x) - \mathbf{k}_n \cdot \mathbf{r}(0) &= \frac{\omega}{c}[-\cos\theta_n, \sin\theta_n, 0] \cdot [x, 0, 0] - 0 \\ &= -\frac{\omega x \cos(\theta_n)}{c}, \end{aligned} \tag{2.12}$$

where the position vectors are $\mathbf{r}(x) = [x, 0, 0]$, $\mathbf{r}(0) = [0, 0, 0]$ and the wave-vector is $\mathbf{k}_n = \frac{\omega}{c}[-\cos\theta_n, \sin\theta_n, 0]$. The scalar product representation of the phase is especially handy in the plane wave case.

To summarize, movement of the transceiver will change the phase according to (2.9). For plane waves the first order approximation (2.11) will be valid.

A continuous-time baseband description

A frequency component ω of a radio signal with nonzero bandwidth can be expressed as the baseband frequency, ω_b, shifted by the carrier frequency, ω_c

$$\omega = \omega_b + \omega_c. \tag{2.13}$$

The antenna voltage (2.1) can then be expressed as

$$V(\mathbf{r}, t) = H_p(\omega_b + \omega_c, \mathbf{r})e^{j(\omega_b + \omega_c)t}, \tag{2.14}$$

where the subindex p on the transfer function denotes that it is the passband channel. The radio transmission is performed at the high frequencies of the passband but it is more convenient to do the signal processing on the symbols in the baseband. The received signal $V(\mathbf{r}, t)$ is transfered from the passband to the baseband by multiplication with $e^{-j\omega_c t}$. The corresponding baseband signal $y(\mathbf{r}, t)$ can be expressed as

$$\begin{aligned} y(\mathbf{r}, t) = V(\mathbf{r}, t)e^{-j\omega_c t} &= H_p(\omega_b + \omega_c, \mathbf{r})e^{j\omega_b t} \\ &= H_b(\omega_b, \mathbf{r})e^{j\omega_b t}. \end{aligned} \tag{2.15}$$

The passband channel $H_p(\omega_b + \omega_c, \mathbf{r})$ is thus equivalent to the baseband channel $H_b(\omega_b, \mathbf{r})$ within the band-limits of the system. The model (2.2) for the passband channel is also a model for the baseband channel,

$$H_b(\omega_b, \mathbf{r}) = \sum_{n=1}^{N} a_n(\mathbf{r})e^{j(\psi_n(\mathbf{r}) - kr_n(\mathbf{r}))}. \tag{2.16}$$

For the wave number we obtain

$$k = k_b + k_c = \frac{\omega_b}{c} + \frac{\omega_c}{c}. \tag{2.17}$$

By using the linearization (2.11) of equation (2.9), the electrical distance can be rewritten as

$$kr_n(x) \approx \omega_c \tau_n(0) + \omega_b \tau_n(0) - k_c \cos(\theta_n)x - k_b \cos(\theta_n)x, \tag{2.18}$$

where the dependence of the electrical distance on the local spatial movement is collected in the last two terms. The first term is constant and may be included in a combined complex attenuation and phase factor α_n that collects all (nearly) space-independent factors. The second term is similar to the first term, but since it depends on the baseband frequency we need it for the transfer function description. The fourth term will be small relative to the third term if the relative bandwidth is small, i.e., if ω_b/ω_c is small. In an application with a carrier frequency at 1800 MHz and a bandwidth of 5 MHz the relative bandwidth will be less than 0.3% and the fourth term can thus be neglected. The nearly time-invariant attenuation and phase factors are collected in a factor α_n as

$$\alpha_n = a_n e^{j(\psi_n - \omega_c \tau_n(0))}. \tag{2.19}$$

Thus, the baseband transfer function (2.16) can for planar waves and straight line motion with constant velocity of the mobile station, be approximated using (2.18) and (2.19) as

$$H(\omega, x) = \sum_{n=1}^{N} \alpha_n e^{j(-\omega \tau_n(0) + \frac{\omega_c}{c} x \cos(\theta_n))}, \tag{2.20}$$

where the subscript b is dropped so that ω from now on will denote *baseband frequency*. The phase shifts described by the exponential terms $\frac{\omega_c}{c} x \cos(\theta_n)$ will cause the rapid variation of the channel when moving short distances. Since $\omega_c/c = 2\pi\lambda_c$, these terms are significant already when the traveled distances is a small fraction of the carrier wavelength.

The parameters a_n, ψ_n and θ_n are also space dependent but on a different scale, governed by the overall geometry, the distance to and the structure of the reflectors. It is reasonable to assume that these parameters will remain fairly constant over at least a small number of wavelengths, whereas the channel varies significantly.

As there exists a parameterized description of the process, where x is the only variable, it is conceivable that there exists a mapping from observations of the channel coefficients to the corresponding coefficients at a location within in a distance of a few wavelengths. Thus, it is likely that the channel parameters can be predicted reasonably well on this small geometrical scale of a few wavelengths. The utilized predictor may, but does not have to, rely on estimates of the parameters a_n, ψ_n and θ_n.

Time dependent transfer function

We have described how the channel depends on the position of the mobile station. By introducing the velocity of the mobile station we can also describe the channel as a function of time. If the vehicle drives straight ahead along the x-axis, at constant velocity, then the traveled distance depends on time as $x = vt$, where v is the speed of the vehicle. This introduces a rotation of each term in the channel (2.20) with an angular frequency

$$\omega_{D_n} = \omega_c \frac{v \cos(\theta_n)}{c}, \tag{2.21}$$

that corresponds to the Doppler-shift of the carrier frequency due to the motion.[3] The maximum Doppler shift ω_D, encountered when driving straight towards the transceiver, is

$$\omega_D = \omega_c \frac{v}{c} = \frac{2\pi v}{\lambda}. \tag{2.22}$$

If we introduce any form of acceleration or change of direction, i.e., driving in a curve, the Doppler shift will become time dependent.

When the Doppler shift is introduced in equation (2.20) the following time-frequency domain representation is obtained

$$H(\omega, t) = \sum_{n=1}^{N} \alpha_n e^{j(-\omega\tau_n(0) + \omega_{D_n} t)}. \tag{2.23}$$

The time domain equivalent of this baseband description of the continuous-time channel is [15]

$$h(\tau, t) = \sum_{n=1}^{N} \alpha_n e^{j\omega_{D_n} t} \delta(\tau - \tau_n(0)), \tag{2.24}$$

where $\delta(\tau)$ is Dirac's delta function and α_n is given by (2.19). The time-varying channel impulse response is thus described as the sum of N complex

[3]The Doppler shift, ω_{D_n}, is caused by the movement through the wave pattern. The wavelength is perceived as shorter (or longer) when traveling towards (or away from) a wave front.

sinusoids, with fixed frequencies between plus/minus the maximum Doppler frequency ω_D.

It is useful to study the Doppler domain representation of equation (2.24), defined as

$$\begin{aligned} H(\tau, \Omega) &= \int_{-\infty}^{+\infty} \sum_n \alpha_n e^{j\omega_{D_n} t} \delta(\tau - \tau_n(0)) e^{-j\Omega t} dt \\ &= 2\pi \sum_n \alpha_n \delta(\tau - \tau_n(0)) \delta(\Omega - \omega_{D_n}), \end{aligned} \tag{2.25}$$

where Ω is the Doppler frequency (in rad/s). In the Doppler domain not only the different delays $\tau_n(0)$ for the incoming waves, as in the impulse response, but also their Doppler frequency ω_{D_n} can be studied. Using the Doppler domain representation (2.25), two rays that arrive with the same delay τ_n from different directions, as in Figure 2.1, can be separated by their different Doppler frequencies.

Let us recapitulate under what conditions the expressions (2.23)-(2.25) are valid.

- The channel models obtained in (2.23)-(2.25) are valid only locally, that is for a few wavelengths and corresponding short times (t) when moving the transceiver.
- Narrow relative bandwidth is presupposed. This holds for most mobile radio systems.
- Under the assumption that the scatterers are fixed and not too close to the receiver, the linearization (2.8) without the quadratic term, resulting in equation (2.23) is valid. This is the plane wave approximation.
- The velocity of the mobile is assumed constant, which excludes curves and accelerations.
- Furthermore, the models do not take polarization effects into consideration.

The use of a spherical wave propagation model is discussed next.

2.3 Sampled Channel with Time Varying Frequencies

The plane-wave approximation can be justified only at a large distance from the wave source. The assumption of spherical wave propagation may be more realistic in the following cases:

- A relatively close primary source visible via the direct path,
- A relatively close primary source visible via one or multiple reflections, modeled by the mirror-image of the primary source.
- A secondary source induced by a distant-source wave front impinging on a close point scatterer (such as a lamp-post).

Using a more elaborate model based on ray-optics including the last term in (2.8), we will see that in the presence of close scatterers the description of the channel as a sum of time invariant weighted complex sinusoids is an oversimplification. This motivates the use of adaptive and nonlinear predictors.

The sampled channel: Several reflectors may contribute to each tap

We shall repeat and expand the discussion performed in the last section but now in the time domain and for sampled channels. Recall equations (2.2) and (2.17) with $\omega_b = \omega$ in a base-band formulation. Let the position $\mathbf{r}$ depend on time. Thus we obtain

$$H(\omega, \mathbf{r}(t)) = \sum_{n=1}^{N} a_n(\mathbf{r}(t)) e^{j(\psi_n(\mathbf{r}(t)) - (\omega_c + \omega) r_n(\mathbf{r}(t))/c)}. \tag{2.26}$$

The nth path delay is $\tau_n(t) = r_n(t)/c$. We can view the expression (2.26) as a function of time directly by dropping the position $\mathbf{r}(t)$ and just using t as variable. The time-varying impulse response $h(\tau, t)$ for a baseband channel at time t in a multi-path environment can thus be described by

$$h(\tau, t) = \sum_{n=1}^{N} a_n(t) e^{j(\psi_n(t) - \omega_c \tau_n(t))} \delta(\tau - \tau_n(t)), \tag{2.27}$$

where ω_c is the carrier frequency (in rad/s) and $a_n(t)$ is a time-varying attenuation factor covering antenna effects, path loss and attenuation due to reflection and scattering for the nth path [15]. The phase shift caused by reflectors and scatterers is described by $\psi_n(t)$ whereas $\tau_n(t)$ denotes the propagation delay for the nth path. The expression (2.27) corresponds to (2.24) without the approximations due to linearization of the change of path distance and small relative bandwidth that render the Doppler frequency description of the phase.

Let $g(\cdot)$ be a time invariant impulse response due to pulse shaping and receiver filtering and let the symbol interval be T. The discrete-time channel impulse response can then be described by an FIR-filter with the kth tap given by [16]

$$\begin{aligned} h_k(t) &= \int_0^{MT} g(Tk-\tau)h(\tau,t)d\tau \\ &= \sum_{n=1}^{N} g(Tk-\tau_n(t))a_n(t)e^{j(\psi_n(t)-\omega_c\tau_n(t))}, \end{aligned} \tag{2.28}$$

where MT covers the length of the continuous-time impulse response. Note that N may be arbitrary large. In an ideal noiseless and lossless environment the radio waves could be reflected between objects forever, resulting in paths of unbounded length, requiring an IIR description of the channel. In practice we can assume that the impulse response $h(\tau,t)$ will be of finite length, as the long paths are sufficiently attenuated, through propagation loss and loss at the reflecting/scattering surfaces, to fall below the background noise level. With an effective support of $g(\cdot)$ on the closed interval $[-KT, KT]$, the number of reflectors and scatterers contributing to the kth tap will be limited to paths with delays in the interval $[\max(0, T(k-K)), T(k+K)]$. The limited number of contributors is an advantage when the tap is to be predicted.

Effective source

If we base our modeling on ray optics and omit the effects of diffraction and Fresnel optics, a *scatterer* can be modeled as a *secondary effective source* induced by a wave front whereas a *reflector* generates a *secondary effective source* as the *mirror image* of the emitting source. Thus both scatterers and the mirror images can be viewed as secondary effective sources emitting *spherical wavefronts*. Even with this simplification we will encounter a model where the phases of the rotating channel coefficients are nonlinear functions of time. Such coefficients can not be predicted accurately by a linear predictor.

Path delays and phase

To simplify the expression of the phase in (2.28) we separate time-invariant and time depending factors. The path delay $\tau_n(t)$ can be decomposed into the sum of a time-varying delay from the effective source to the mobile ($\tau_n^{MS}(t)$) and a time-invariant path delay from the base station to produce the secondary effective source, τ_n^{BS} as

$$\tau_n(t) = \tau_n^{MS}(t) + \tau_n^{BS}. \tag{2.29}$$

For a scatterer τ_n^{BS} is the path delay from the base station to the scatterer and $\tau_n^{MS}(t)$ is the path delay from the scatterer to the mobile station. For a reflector the secondary source is the mirrored image of the primary source. The path delay from the secondary source to the mobile station is thus equal to the path delay from the primary source to the mobile station, that is $\tau_n^{MS}(t) = \tau_n(t)$. Accordingly is $\tau_n^{BS} = 0$ for reflections.

The geometry for the scatterers and reflectors are assumed to be time-invariant. The phase shift due to the time invariant delay, τ_n^{BS}, can be included in a complex attenuation factor which is now defined as

$$\alpha_{n,k}(t) = g(Tk - \tau_n(t))a_n(t)e^{j(\psi_n(t) - \omega_c \tau_n^{BS})}, \tag{2.30}$$

while the time-varying term

$$\phi_n(t) \triangleq -\omega_c \tau_n^{MS}(t) = -\frac{2\pi}{\lambda} r_n^{MS} = -\frac{2\pi}{\lambda} ||\mathbf{r}_n^{MS}(t)||, \tag{2.31}$$

remains in the exponential factor of (2.28). Note that $a_n(t)$ and $\psi_n(t)$ and thus $\alpha_{n,k}(t)$ are assumed to be time-varying on a much slower time scale than $\phi_n(t)$. The discrete-time channel (2.28) can thus be expressed as

$$h_k(t) = \sum_n \alpha_{n,k}(t) e^{j\phi_n(t)}. \tag{2.32}$$

In (2.31), $\mathbf{r}_n^{MS}(t)$ is a vector in space pointing from the nth effective source to the mobile station (see Figure 2.4) and λ is the wavelength. Consequently the norm of $\mathbf{r}_n^{MS}$ is nothing but the physical distance r_n^{MS}. For spherical waves the position vector $\mathbf{r}_n^{MS}$ is parallel to the wave vector, $\mathbf{k}_n(t)$. The phase $\phi_n(t)$ is thus solely a function of the electrical distance to the effective source. When the distance changes by as little as one wavelength the phase $\phi_n(t)$ changes by 2π, causing the effect of small-scale (fast) fading. In the model based on the plane wave-approximation (2.24) the phase $\phi_n(t)$ has only a linear dependence on time. In the following we will derive how much the linearized model can deviate from a phase modeled using spherical waves.

Straight-line motion

Consider, as before, the simplest mobile dynamics, a straight-line motion at constant velocity $\mathbf{v}$. For notational convenience we denote the initial position by $\mathbf{r}_n^{MS}(0) = \mathbf{r}_n^{MS}$ without explicit time index. We then have the phase function

$$\phi_n(t) = -\frac{2\pi}{\lambda} ||\mathbf{r}_n^{MS}(t)|| = -\frac{2\pi}{\lambda} ||\mathbf{r}_n^{MS} + \mathbf{v}t||. \tag{2.33}$$

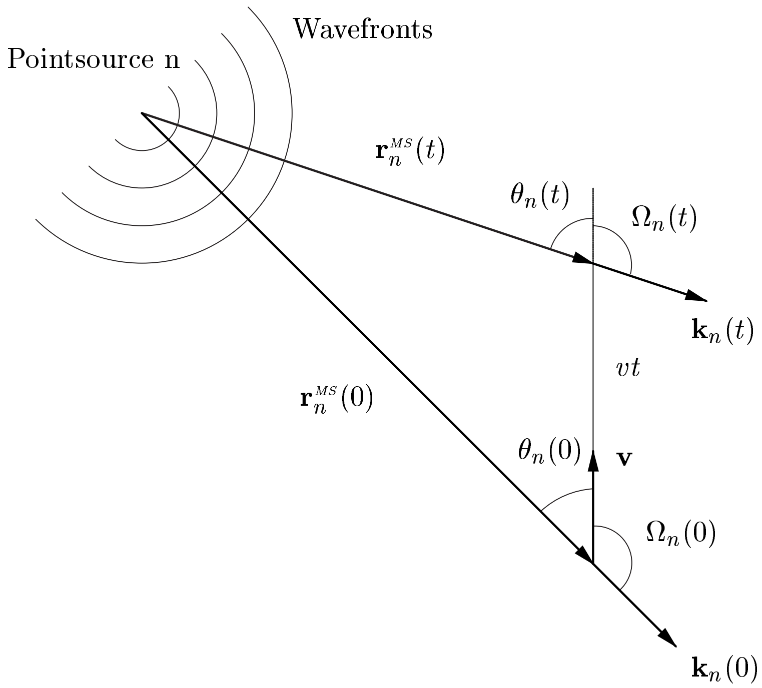

Figure 2.4: The change of path distance, $\mathbf{r}_n^{MS}(t)$, when moving with constant velocity, $\mathbf{v}$. Here $\Omega_n(t)$ is the angle between the velocity vector $\mathbf{v}$ and the wave vector $\mathbf{k}_n(t)$, whereas the angle $\theta_n(t) = \pi - \Omega_n(t)$ is the angle between $\mathbf{v}$ and the direction to the effective source at time t (that is $-\mathbf{r}_n^{MS}(t)$).

Let $\Omega_n(t)$ denote the angle between the wave vector $\mathbf{k}_n(t)$ (or equivalently the position vector $\mathbf{r}_n^{MS}(t)$) and the velocity vector $\mathbf{v}$, see Figure 2.4. The angle towards the effective source, also called the angle of incidence, is

$$\theta_n(t) = \pi - \Omega_n(t).$$

Furthermore let

$$\theta_n = \theta_n(0),\ r_n^{MS}(0) = ||\mathbf{r}_n^{MS}(0)||,\ v = ||\mathbf{v}||,\ r_n^{MS}(t) = ||\mathbf{r}_n^{MS}(t)||.$$

Then, by using the cosine theorem, the phase function at the position $\mathbf{r}_n^{MS}(t)$ at time t can be rewritten as

$$\phi_n(t) = -\frac{2\pi r_n^{MS}}{\lambda}\sqrt{1 - 2\frac{vt}{r_n^{MS}}\cos\theta_n + \left(\frac{vt}{r_n^{MS}}\right)^2}. \tag{2.34}$$

We use a second-order approximation

$$\sqrt{1+y} \approx 1 + y/2 - y^2/8,$$

with y representing the sum of the two last terms under the square root sign of (2.34). Furthermore, we neglect terms in vt/r_n^{MS} of higher order than two since $vt/r_n^{MS} \ll 1$ and thus obtain

$$\begin{aligned}\phi_n(t) &\approx -\frac{2\pi r_n^{MS}}{\lambda}\left[1 - \frac{vt}{r_n^{MS}}\cos\theta_n + \frac{1}{2}\left(\frac{vt}{r_n^{MS}}\right)^2 \sin^2\theta_n\right] \\ &= -\frac{2\pi r_n^{MS}}{\lambda} + \omega_{D_n} t - \frac{\pi r_n^{MS}}{\lambda}\left(\frac{vt}{r_n^{MS}}\right)^2 \sin^2\theta_n, \end{aligned} \qquad (2.35)$$

with the Doppler frequency $\omega_{D_n} = 2\pi v \cos(\theta_n)/\lambda$ as in (2.21). The first term, the initial phase, is a time invariant phase shift and can be included in the complex attenuation factor $\alpha_{n,k}(t)$. The second term represents the phase rotation at the Doppler frequency, encountered in Section 2.2. Beyond this linear increase of the phase, we obtain a third, quadratic term originating from the sphericity of the wave fronts. The quadratic time dependence of the phase can be interpreted as a *chirp*, whereas a strict linear time dependence would represent a complex sinusoid with time-invariant frequency.

2.3.1 The instantaneous frequency

The instantaneous frequency, that is the time derivative of the phase function, can offer further insights in to the relation between phase and Doppler frequency. The instantaneous frequency can be derived directly from (2.34) and will then be expressed in the initial angle θ_n, speed and time. It can also be derived by operating on (2.33) to obtain an instantaneous frequency expressed in the instantaneous angle $\theta_n(t)$. In the latter approach the time-derivative of $r_n^{MS}(t)$ is needed. An infinitesimal step in time, Δt, result in a change of path distance as

$$r_n^{MS}(t + \Delta t) = ||\mathbf{r}_n^{MS}(t) + \mathbf{v}\Delta t||. \qquad (2.36)$$

As Δt is infinitesimally small we can use the approximation (2.10) in (2.36),

$$r_n^{MS}(t + \Delta t) = r_n^{MS}(t) - v\cos\theta_n(t)\Delta t. \qquad (2.37)$$

The time-derivative of the path-length is thus

$$\frac{dr_n^{MS}(t)}{dt} = \lim_{\Delta t \to 0} \frac{r_n^{MS}(t + \Delta t) - r_n^{MS}(t - \Delta t)}{2\Delta t} = -v\cos\theta_n(t). \qquad (2.38)$$

The instantaneous frequency is thus given by

$$\begin{aligned}\dot{\phi}_n(t) &= -\frac{2\pi}{\lambda}\frac{d}{dt}||\mathbf{r}_n^{MS}(t)|| = \frac{2\pi}{\lambda} v\cos\theta_n(t) \\ &= \omega_D \cos\theta_n(t) = \omega_{D_n}(t), \end{aligned} \qquad (2.39)$$

where ω_D is the maximum Doppler frequency (2.22) and $\omega_{D_n}(t)$ is the instantaneous Doppler frequency for path n. Using the approximate expression (2.35), we obtain

$$\dot{\phi}_n(t) \approx \frac{2\pi}{\lambda}\left[v\cos\theta_n - \frac{t}{r_n^{MS}}\left(v\sin\theta_n\right)^2\right] = \omega_D\cos\theta_n - \frac{2\pi t}{\lambda r_n^{MS}}\left(v\sin\theta_n\right)^2. \tag{2.40}$$

The instantaneous frequency can thus either be described as an instantaneous Doppler shift depending on the momentary angle of incidence $\theta_n(t)$ as in (2.39) or approximately by the difference between the Doppler shift at time $t = 0$ and a time dependent correction, as in (2.40).

Phase as a function of instantaneous angle of incidence

When comparing the expression (2.39) to (2.40), one can be led to assume that if the instantaneous Doppler frequency was used instead of the Doppler frequency corresponding to the angle at $t = 0$, in (2.35), then there would be no need of a correction term. This is however not the case as we will see below. To express the distance $\mathbf{r}_n^{MS}(t)$ as a function of the instantaneous angle of incidence $\theta_n(t)$, we can by using the cosine theorem to obtain

$$r_n^{MS2} = (vt)^2 + r_n^{MS}(t)^2 + 2vtr_n^{MS}(t)\cos\theta_n(t). \tag{2.41}$$

Solving for $r_n^{MS}(t)$ we obtain

$$r_n^{MS}(t) = -tv\cos\theta_n(t) + r_n^{MS}\sqrt{1-\left(\frac{vt}{r_n^{MS}}\right)^2\sin^2\theta_n(t)}. \tag{2.42}$$

The phase function at the position $\mathbf{r}_n^{MS}(t)$ at time t can thus be expressed as

$$\phi_n(t) \quad = \quad \omega_{D_n}(t)t - \frac{2\pi r_n^{MS}}{\lambda}\sqrt{1-\left(\frac{vt}{r_n^{MS}}\right)^2\sin^2\theta_n(t)}, \tag{2.43}$$

or using the same approximation as in (2.35) as

$$\phi_n(t) \quad \approx \quad -\frac{2\pi r_n^{MS}}{\lambda} + \omega_{D_n}(t)t + \frac{\pi r_n^{MS}}{\lambda}\left(\frac{vt}{r_n^{MS}}\right)^2\sin^2\theta_n(t). \tag{2.44}$$

The initial phase is obviously the same as in (2.35). In the second term the difference lies in that the instantaneous Doppler frequency $\omega_{D_n}(t)$ is used instead of the initial Doppler frequency ω_{D_n}. Observe that even though we use the instantaneous Doppler frequency there is a quadratic correction term in the phase, in this case involving $\sin^2\theta_n(t)$ instead of $\sin^2\theta_n$.

The sine of the instantaneous angle of incidence can however by application of the sine and cosine theorem, be expressed in terms of the time and the initial values of distance and angle, as

$$\sin^2\theta_n(t) = \frac{r_n^{MS2}\sin^2\theta_n}{r_n^{MS2}(t)} = \frac{r_n^{MS2}\sin^2\theta_n}{r_n^{MS2}+(vt)^2-2r_n^{MS}vt\cos\theta_n}. \tag{2.45}$$

Using the relation (2.45) we can show that the term $(\frac{vt}{r_n^{MS}})^2\sin^2\theta_n(t)$ occurring in the correction terms of (2.43) and (2.44) is bounded:

$$0 \leq \left(\frac{vt}{r_n^{MS}}\right)^2 \sin^2\theta_n(t) \leq 1. \tag{2.46}$$

Thus the correction term in (2.43) is limited whereas the correction term in (2.35) grows continuously in magnitude.

2.3.2 Linearized model

Assuming plane incoming waves, that is waves with a constant angle of incidence $\theta_n(t) = \theta_n$, and a time invariant complex attenuation factor $\alpha_{n,k}$ in (2.32), we obtain the following commonly used approximation [1]

$$h_k(t) = \sum_{n=1}^{N} \alpha_{n,k} e^{j\omega_{D_n} t}. \tag{2.47}$$

This is the sampled version of (2.24). This model has been used as a basis expansion model for blind equalization [17] and also for long-range prediction of mobile radio channels by Duel-Hallen and co-workers [5].

As we have seen in (2.44) the Doppler frequencies actually vary, and therefore a more suitable model is

$$h_k(t) = \sum_{n=1}^{N} \alpha_{n,k}(t) e^{j\omega_{D_n}(t)t}, \tag{2.48}$$

where $\omega_{D_n}(t)$ is the instantaneous Doppler frequency for the nth path and $\alpha_{n,k}(t)$ is the instantaneous complex attenuation. Both these parameters can be assumed to vary slowly.

Phase error in sinusoid models

The linear phase model, that is the deterministic sinusoid model (2.47) will accumulate a phase error when used in a situation with close scatterers where the plane wave approximation is non-valid. The size of this phase error defines the maximum time interval over which the linear deterministic model can be used as an approximation in the spherical-wave case. The difference in phase can be approximated by the second order term in (2.35)

$$\Delta\phi_n(t) = -\frac{\pi v^2}{\lambda r_n^{MS}} t^2 \sin^2\theta_n = -\frac{\omega_D v}{2r_n^{MS}} t^2 \sin^2\theta_n. \tag{2.49}$$

The largest deviation of the phase in the model (2.47) relative to (2.34) occurs for a *transversal vehicle movement* at $\theta_n = \pi/2$, when the velocity vector is orthogonal to the direction of the incident wave at time $t = 0$ and the nominal Doppler shift vanishes.

Example 2.1

Consider a situation where there is only one path contributing to the channel and the velocity vector is orthogonal to the direction of the incident waves at time $t = 0$. The contribution by the path to the channel tap is a chirp $h(t) = \alpha e^{i\zeta t^2}$ (with $\zeta = -\pi v^2/(\lambda r_{MS})$). The true parameters, α and ζ, at time $t = 0$ are now used either in the linear sinusoid model (2.47) or the model (2.32), with the quadratic phase expression (2.34), and the time is propagated. Two different predictions of $h(t)$ into the future are then obtained. The phase of $h(t)$ at time $t = T$ is ζT^2, and this will also be the result in the model containing the quadratic term. The linear phase model, with the accumulated phase error of $\Delta\phi(T)$, will predict the contribution to be

$$\hat{h}(t) = \alpha e^0 = \alpha e^{i(\zeta T^2 + \Delta\phi(T))}. \tag{2.50}$$

The channel tap prediction error, for the predictor using the linear model, at time $t = T$ is then

$$\begin{aligned} \varepsilon(T) &= \alpha(e^{i\zeta T^2} - e^{i(\zeta T^2 + \Delta\phi(T))}) = \alpha e^{i(\zeta T^2 + \Delta\phi(T)/2)}(-2i)\sin\left(\frac{\Delta\phi(T)}{2}\right) \\ &= \alpha e^{i(\zeta T_2 - \Delta\phi(T)/2)}(-2i)\sin\left(\frac{\Delta\phi(T)}{2}\right) \end{aligned} \tag{2.51}$$

and the relative power of the error is

$$\frac{|\varepsilon(T)|^2}{|\alpha|^2} = 4\sin^2\frac{\Delta\phi(T)}{2}. \tag{2.52}$$

Compare this to the result if we would predict the tap by its mean, $\hat{h}(T) = 0$, which would give a relative error of 1 and a prediction gain of 0 dB.

The time limit for how long the linear model can be used as a predictor, with better performance than the mean, is obtained by

$$4sin^2 \frac{\Delta\phi(T)}{2} \leq 1, \tag{2.53}$$

that is

$$|\Delta\phi(T)| \leq \pi/3. \tag{2.54}$$

Thus, predictions based on the extrapolation of a sinusoidal model will perform better than the zero predictor only up to prediction horizon T where the accumulated phase error has grown to $\pi/3$.

Based on the above example for a situation with just one path with a tap acting as a chirp, the time interval $T_{\pi/3}$ denotes the time over which a linear predictor based on the model (2.47) actually renders predictions better than using the mean as the prediction. Defining this time interval, $T_{\pi/3}$, for the worst case $\theta_n = \pi/2$, as the the interval after which the phase approximation has grown to 60°, $|\ \Delta\phi_n(t)|_{T_{\pi/3}}| = \pi/3$, we obtain

$$T_{\pi/3} = \sqrt{\frac{\lambda r_n^{MS}}{3v^2}}. \tag{2.55}$$

For this time span, the above second-order approximation (2.35) can be well justified as we have

$$\left.\frac{vt}{r_n^{MS}}\right|_{t=T_{\pi/3}} = \sqrt{\frac{\lambda}{3r_n^{MS}}} \ll 1 \tag{2.56}$$

where the last inequality holds because even a close point scatterer will be many wavelengths away from the mobile transceiver in outdoor mobile radio scenarios.

Example 2.2

A simple scenario including a vehicle driving at 90 km/h past a close point scatterer, 10 m beside the road, demonstrates the limitations of the linear model. With a carrier frequency of 1800 MHz the resulting $T_{\pi/3}$ is as short as 29 ms. Thus, the linear deterministic model is certainly inadequate for an estimation window of 150 ms. This is the time span we have in our real world channel measurements, to be presented in the following chapters. This is also true for prediction horizons of 10 to 30 ms, when there are close scatterers. However, this problem does not exist for more distant (primary and reflected) sources, say, at $r_n^{MS} = 250$ m and beyond where we obtain $T_{\pi/3} = 145$ ms.

2.3.3 Statistical description of the linearized model

As the parameters in the model (2.28) are hard to estimate when N is large, the common approach [1] is instead to use the linearized model (2.47) and to assume that N is infinite, the damping factors $\{a_n\}$ are mutually uncorrelated and identically distributed, and that the phases $\{\psi_n\}$ are mutually uncorrelated with uniform distribution in the interval $[0, 2\pi[$. The complex attenuation factor $\alpha_{n,k}$ in (2.47) then has a circular complex Gaussian distribution.

A snapshot of a complex valued channel coefficient, in a sampled version of the time domain channel, is thus distributed as

$$h_k(t) \sim CN(0, \sigma_k^2), \tag{2.57}$$

where $CN(\cdot)$ denotes a circular complex Gaussian distribution. The modulus of $h_k(t)$ is then Rayleigh distributed.

Furthermore, all angles of incidence are assumed to be time-invariant and equally probable. That is, the angles θ_n are mutually uncorrelated with uniform distribution in the interval $[0, 2\pi[$. The Doppler frequency is given as $\omega_D \cos\theta_n$. As the probability distribution for $\cos\theta_n$ has high peaks at ± 1, the Doppler frequencies ω_{D_n} are likely to be close to the limits $\pm\omega_D$. This cause the well known bathtub shape of the Doppler spectrum for a Rayleigh fading tap shown in Figure 2.5. This theoretical Doppler spectrum for a Rayleigh fading tap is a good approximation for taps in a narrow-band channel, when the assumptions of a high number of reflectors are met. For broad-band channels the spatial resolution is higher and fewer reflectors contribute to each tap. This cause the Doppler spectra for broad-band channels to have much more fine structure, as illustrated in Appendix A.

2.3.4 Path loss

The path loss for a scattered or reflected path is proportional to the path distance as [18]

$$a_n(t) \propto (r_n^{MS}(t) \otimes r_n^{BS}(t))^{-\gamma/2} \tag{2.58}$$

where $r_n^{MS}(t)$, $r_n^{BS}(t)$ denotes the distance from the nth scatterer/reflector to the mobile station and the base station respectively and γ is the power

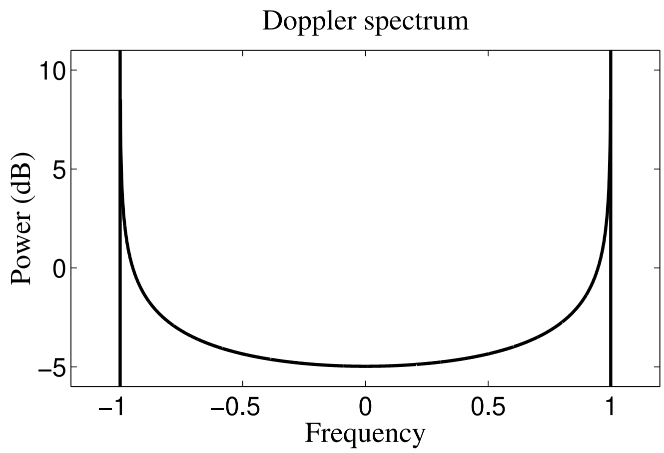

Figure 2.5: Doppler spectrum for a Rayleigh fading tap. Frequency is normalized by the Doppler frequency ω_D and the power is normalized by the power of the tap. There is no power outside the the bounds set by the Doppler frequency (± 1).

attenuation exponent. In (2.58) $r_n^{MS}(t) \otimes r_n^{BS}(t)$ denotes $r_n^{MS}(t) \cdot r_n^{BS}(t)$ for scattering and $r_n^{MS}(t) + r_n^{BS}(t)$ for specular reflection. The power attenuation exponent corresponding to the path loss in free space is $\gamma = 2$. Even fairly close scatterers will give contributions with amplitudes which are an order of magnitude weaker than the specular reflections. Thus, it is appropriate to use the linear model (2.47) in situations where there are direct line of sight (LOS) or large contributions from specular reflections, that is strong reflections from large buildings. In all other cases the effect of spherical waves from nearby scatterers introduces a significant deviation from the linear model (2.47).

2.3.5 Curves and plane waves

Phase functions containing quadratic terms occur when the vehicle accelerates or makes a turn, even if the plane-wave approximation holds. Assume that the mobile moves at constant tangential speed v around a circle of radius R with its center at the origin of the coordinate system such that

$$||\mathbf{r}(t)|| = R, \tag{2.59}$$

where $\mathbf{r}(t)$ denotes the position of mobile station. As the mobile station now change direction we use a fixed coordinate system as shown in Figure 2.6.

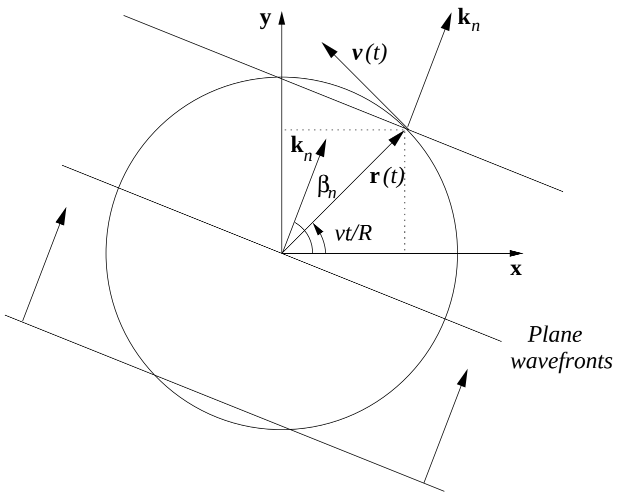

Figure 2.6: Plane waves and a vehicle driving with velocity v in a curve with radius R.

Let β_n be the angle between the constant wave vector $\mathbf{k}_n$ (plane wave) and the initial position vector $\mathbf{r}(0)$ as in Figure 2.6. Recall that the phase could be calculated as the scalar product between wave vector and position vector added to a time invariant phase factor (2.5). We take $\phi_n(t)$ to be the time varying part of the phase in (2.31). If we initiate $\phi(t)$ to be zero when the position vector and the wave vector are orthogonal, then the phase function at time t is given by

$$\begin{aligned}\phi_n(t) &= -\mathbf{k}_n \cdot \mathbf{r}(t) \\ &= -\frac{2\pi}{\lambda}[\cos\beta_n, \sin\beta_n, 0] \cdot R[\cos(vt/R), sin(vt/R), 0] \\ &= -\frac{2\pi}{\lambda}R\cos(vt/R - \beta_n). \end{aligned} \tag{2.60}$$

The phase does not increase as a linear function of time here. Instead the Doppler shift oscillates (slowly and) symmetrically around 0. The maximum deviation from a linear phase function occurs at the time instances when the cosine function has its maximum curvature, which coincide with the cosine maxima. These are reached when the vehicle moves *transversal to the incident wave*, that is when $vt/R = \beta_n$, where the instantaneous Doppler shift vanishes but the corresponding chirp rate has its maximum. Consider the time interval $T_{\pi/3}$ over which a phase error of $\pi/3$ is accumulated. To this end, a second-order approximation of the cosine function is used around its maximum, i.e.

$$\phi_n(t) \approx -\frac{2\pi}{\lambda} R \left[1 - \frac{1}{2} (v\tau/R)^2 \right], \tag{2.61}$$

where τ is relative to the time where the cosine maximum is reached during the circular motion. Again we have obtained a second-order polynomial phase behavior. The constant term corresponds to the linear deterministic model which, for transversal motion, would not show any Doppler shift at all. The quadratic term represents the deviation from the linear behavior. Using $T_{\pi/3}$ as the measure for how fast these deviations occur we obtain

$$T_{\pi/3} = \sqrt{\frac{\lambda R}{3v^2}}. \tag{2.62}$$

For this time span, the above second-order approximation can be well justified as we have

$$\left. \frac{v\tau}{R} \right|_{\tau = T_{\pi/3}} = \sqrt{\frac{\lambda}{3R}} \ll 1 \tag{2.63}$$

where the last inequality holds because the minimal turning circle diameter of any car is orders of magnitudes larger than the radio wavelengths used in digital mobile communications systems.

Example 2.3

We evaluate equation (2.62) for some typical speeds and sizes of curves. For v =90 km/h and R =100 m we have $T_{\pi/3}$ =92 ms (2.3 m of traveled distance) and, for v =20 km/h and R =10 m, $T_{\pi/3}$=131 ms (0.73 m of traveled distance). We conclude that normal deviations from the exact straight-line motion may result in significant phase errors of the linear deterministic fading model over estimation time windows of the order of 0.1 s.

2.4 Channel Simulation

A simple simulation model including the effects of scatterers in the vicinity of the mobile station can be derived from (2.32). The path losses are modeled as in (2.58) and the phase shifts ψ_n in (2.30) are taken to be random variables uniformly distributed in $[0, 2\pi]$. The time-varying phases $\phi_n(t)$ are calculated as in (2.33).

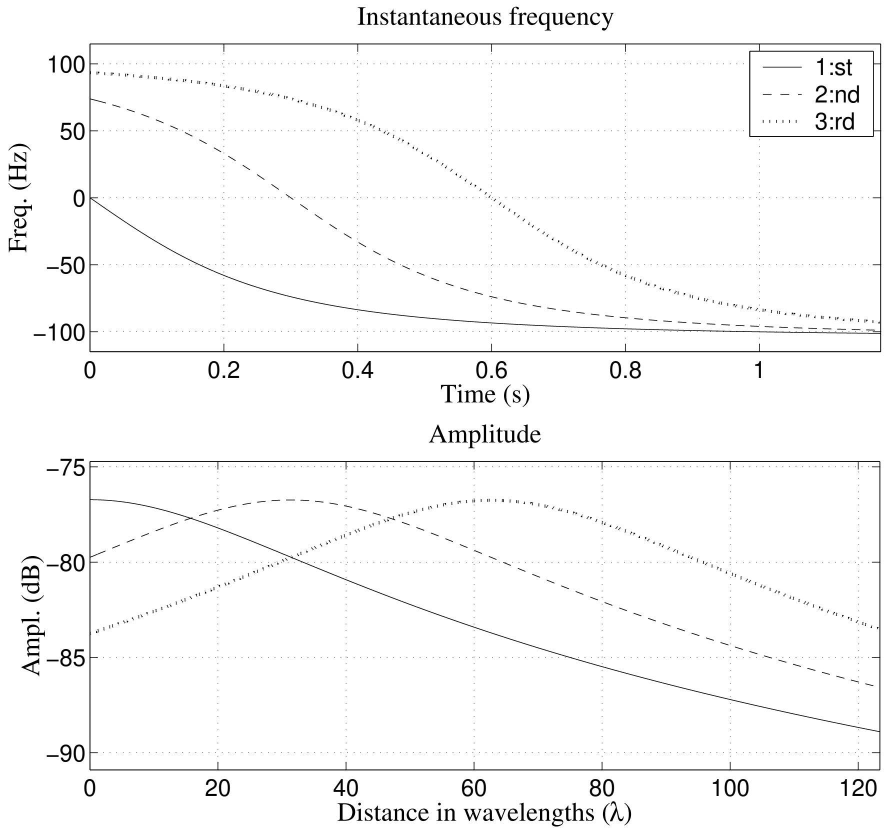

Figure 2.7: The simulated instantaneous frequencies for 3 scatterers and their amplitudes. The maximum Doppler frequency is 104.4 Hz. The upper and lower x-axises show time and traveled distance respectively.

The following simulation scenario is studied: A mobile station is passing three point scatterers which are located 5 m from the road and 5 m apart. The vehicle speed is 60 km/h and the carrier frequency 1880 MHz, as in the measurements described in Section 3.1. The wavelength is thus 16 cm and the maximum Doppler frequency is 104 Hz. No line-of-sight or dominant reflector is assumed. Figure 2.7 shows the instantaneous frequency for the different scattered wavefronts and the corresponding amplitudes. When the mobile passes a scatterer the Doppler shift vanishes. This happens at times

0 s, 0.3 s and 0.6 s for the different contributing scatterers. At the same time the corresponding chirp rate has its maximum. For the contribution of one scatterer at the time for the passage $T_{\pi/3}$ is 31 ms.

In Figure 2.8 the accumulated power from the scatterers are shown when the second scatterer is passed. The interference between the wave fronts cause fades. The fading dips are spaced roughly one wavelength apart. The depth and width of the fades vary significantly.

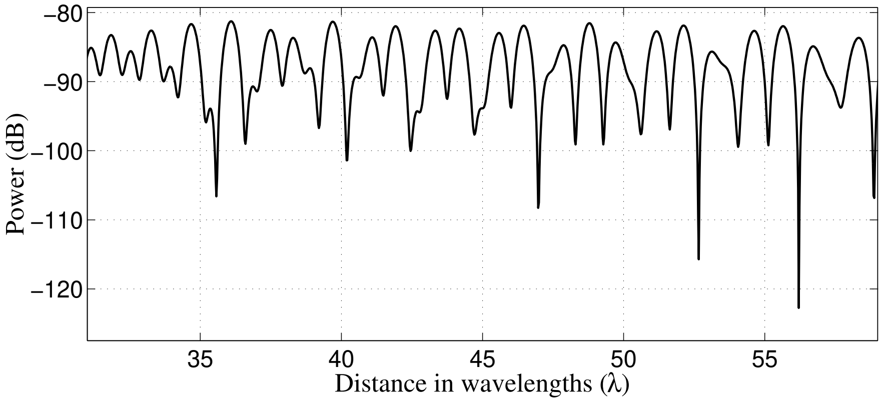

Figure 2.8: The received power (the summed contribution from the scatterers), from 0.3 s to 0.5 s. That is when driving past the second scatterer. The x-axis is scaled in traveled distance measured in wavelengths. At 60 km/h it takes 9.1 ms to travel one λ of 16 cm.

2.5 Conclusion

We have analyzed two deviations from the conventional plane-wave and constant-velocity model for mobile radio multi-path channels. Each of the two assumptions of spherical wave propagation and circular motion leads to a quadratic term in the phase functions of the multi-path components (2.31).

The most critical deviation from the conventional model was found in spherical wave propagation due to close point scatterers. Over the considered time intervals, the phase functions can be represented quite accurately by quadratic polynomials. Thus, the channel impulse response is well modeled as a weighted sum of complex sinusoids and linear chirps.

The quadratic behavior is obtained as approximations, (2.35) and (2.61), which are justified only for relatively small time intervals. Within the range of validity of these approximations, we could verify that the linear deterministic model for the fading pattern (i.e., the sum-of-sinusoids model for the channel parameters (2.47)) incurs phase errors of the order of $\pi/3$ within a time window that would in practice be required for estimating the model parameters, that is time windows on the order of 100 ms, or a few wavelengths in space. This motivates why adaptive linear and/or nonlinear predictors should be used.

2.A The Required Size of a Reflector

To classify an object as a reflector or scatterer we need the concept of Fresnel zones. The first Fresnel zone is defined as the ellipse where the path length

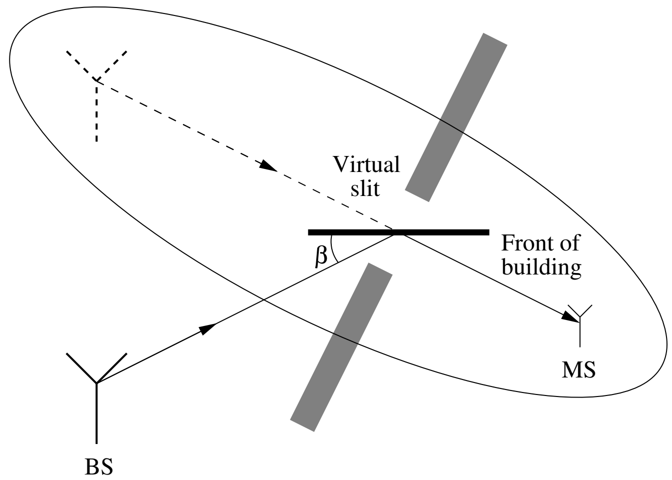

Figure 2.9: A building of width L reflects the waves from the antenna. The angle of incidence is β. The ellipse show the first Fresnel zone. The building acts as a slit of width $L \sin \beta$ for the virtual reflected source.

from the mirror image via a point on the ellipse exceeds the direct path from the mirrored image to the mobile station by $\lambda/2$. A reflector can be viewed as a slit for the mirror image. If the slit opening allows at least the whole first Fresnel zone through, the object act as an reflector (see Figure 2.9). The direct path is here the path from the virtual mirror source to the receiver. The path length is the sum of the distance from the base station r_{BS} (or the secondary source) to the reflecting object and the distance from the object to the mobile station r_{MS}. The width of the object L, has thus to be large enough to let the first Fresnel zone through. The slit is tilted at an angle β, the angle of incidence, and the effective width of the slit is thus $L \sin \beta$. The smallest size for an object to be considered as a reflector is

$$L > \frac{2}{\sin \beta} \sqrt{\frac{r_{BS} + r_{MS}}{r_{BS} r_{MS}}} \lambda. \tag{2.64}$$

Smaller objects will cause noticeable diffraction, spreading energy not only in the angle of reflection. The diffraction thus cause a similar effect as that of a scatterer.

Example 2.4

Consider a scenario with an object acting as a reflector in the nth path. The carrier frequency is 1880 MHz (the wavelength λ is thus 16 cm) and the object is on a distance of 490 m from the base station and only 10 m from the mobile station. This object can be no smaller then 2.5 m to act as a reflector (that is for the most beneficial angle $\beta = \pi/2$ or 90°). For an angle of reflection of 45° the object has to be at least 3.5 m wide to act as a reflector. An object further away from the mobile, 400 m from the base station and 100 m away from the mobile station, at an angle of 45° would have to be at least 10 m wide. Most objects along a road, as cars, are thus too small to act as reflectors. A building is often large enough to be considered as a reflector but most façades are not homogeneous reflecting surfaces. This causes a blend of reflection, diffraction and scattering from buildings.

Chapter 3

Measurements

In this chapter the recording and identification of channel impulse responses are treated. The data base consist of a number of channel-sounder measurements at different locations. The channel-sounder is a correlation channel-sounder, transmitting a sequence with good correlation properties to give the best possible estimated channel impulse responses. The system is shown in Figure 3.1. The sequence $s(t)$ is sent through a transmitter over the mobile radio channel $\mathbf{h}(t)$ to the receiver, where the signal $y(t)$ is received. While no co-channel interference is present in the measurements, some noise $n(t)$ from the environment and the equipment corrupts the received signal.

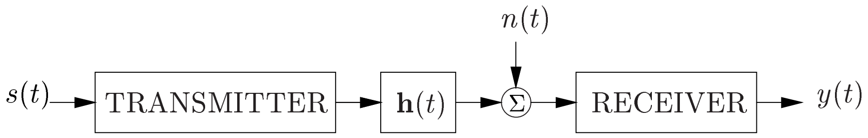

Figure 3.1: Channel sounder. The air interface, $\mathbf{h}(t)$, is modeled as a time varying FIR-filter.

In Section 3.2 the procedure to identify FIR-channels from the measurements are described. The channels are estimated by the least squares method. The power delay profile and the effect of measurement noise is covered in Section 3.2.3. The aim of this chapter is to understand the underlying nature of the measured channel impulse responses as well as developing methods for distinguishing between the channel dynamics and the noise.

3.1 Radio Channel Database

Measurements

The mobile radio channel data studied in this thesis were recorded in Kista, a suburb outside Stockholm. Wideband radio channel measurements were collected at 1880 MHz, at distances of 200 to 2000 m from the base station antenna placed on the roof of a high building. The mobile antenna was placed on a car driving in the suburban area mostly without line of sight. The vehicle velocity varied between 30 and 90 km/h. In total, 25 usable measurement runs were recorded at different positions. The measurements consist of 156.4 ms long recordings of the received signal at each measurement location. The transmitted signal consisted of 1430 repetitions of a sequence of length 109.4 μs. As the baseband sampling rate of the receiver was 6.4 MHz, each transmitted sequence of 109.4 μs resulted in 700 recorded samples.

Channel sampling

The impulse response of the channel is modeled by an FIR-filter, with parameters estimated by a block based least squares (LS) method. For each repetition of the transmitted sequence, that is 700 received samples, a new channel is identified, resulting in 1430 consecutive impulse responses at each measurement location. The *channel sampling frequency* is thus 6.4/700 MHz $\approx$ 9.1 kHz. Since the highest Doppler frequency in the measurements, that occurs for velocities around 90 km/h, is about 160 Hz ($f_D = f_c v/c$), the channel sampling frequency is sufficient to avoid aliasing.

The block identification of the channel coefficients introduces an error, due to the time variation of the parameters of the channel during the identification interval, and an error due to the measurement noise. The size of the error due to time variation of the estimated coefficients, depends on the time-frequency (TF) product between the length of the identification interval (the block) and the Doppler frequency. With the chosen block-length the bias will be negligible. The problem of reducing the noise on the identified parameters by filtering is treated in Section 3.3.

The time-span covered by the FIR-filter has to encompass all contributing paths. It was found that 120 taps, thus a time span of 18.75 μs, was sufficient. That is, only paths shorter than 5.6 km were found to contribute to the measurements. Note that 700 recorded samples are used to identify 120 complex parameters in the FIR-model.

An identification procedure using $s(t)$ and $y(t)$ in Figure 3.1 would result in an estimate of the channel convolved with the transmitter and receiver filters. Under the condition that the transmitter and receiver are linear and that the channel is slowly time varying, the order of the components can be interchanged as in Figure 3.2. To obtain a reference signal for identification

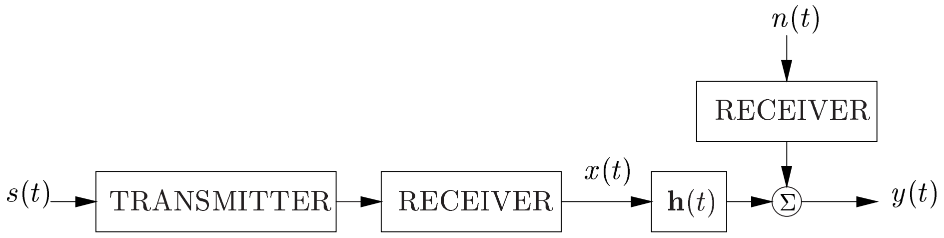

Figure 3.2: Equivalent channel sounding system. Used as a model in identification of the channel, $\mathbf{h}(t)$*.*

of the impulse responses, a back-to-back measurement is performed. The transmitted sequence, $s(t)$, is then sent through the transmitter directly connected to the receiver with a cable, as in Figure 3.3. On the receiver side the 700 samples of reference signal, $x(t)$, is obtained for the system without the air-interface. By using the back-to-back measurement for identification we avoid identifying the transmitter and receiver filters and obtain an estimate of the impulse response of the air interface.

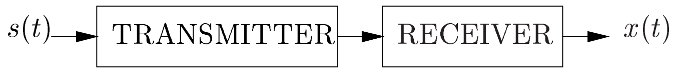

Figure 3.3: Back-to-back measurement on the channel sounder to obtain the reference signal $x(t)$*.*

Span of scales involved

The most important properties of the measurement and the required properties of the predictors to be constructed are summarized in the Tables 3.1 and 3.2 respectively.

Note the large difference in scale. The entries of Table 3.1 are explained in the following.

- The carrier frequency is 1880 MHz and the wavelength is 15.9 cm. The waves will thus interact with objects larger than ≈10 cm.
- The baseband sampling rate is 6.4 MHz. That is, we obtain a sample every 0.156 μs, which corresponds to a spatial resolution of 46.6 m.

Parameter	Temporal measure	Spatial structure
Carrier frequency	1880 MHz	15.95 cm
Baseband sampling rate	6.4 MHz	46.6 m
Impulse response length	18.75 μs	5.6 km
Max delay spread	3 μs	900 m
Channel update rate	109μs	0.9-2.7 mm

Table 3.1: Table of the parameters governing the measurements of the impulse responses. The vehicle velocity is 30-90 km/h (roughly 8-25 m/s). An interval in the spatial properties refers to the different vehicle velocities.

Parameter	Temporal measure	Spatial structure
Adaption interval	156 ms	1.3-3.9 m
Prediction range	1-10 ms	0.8-25 cm
Prediction memory	15 ms	12.5-37.5 cm

Table 3.2: Table of the parameters governing the prediction of the impulse responses (see Chapters 4-6).

Thus, the main contribution to each tap comes from reflections with path distances that differ less than 50 m. Thus, the geometry on the 10 m scale is involved.

- The impulse response length is 18.75 μs. This allows path-lengths from the transmitter to the receiver of 5.6 km. Thus, the estimated impulse response accounts for reflections over the 10^3 m scale.

- The measured delay spread is a measure of the average path delay, which is different at each location. It is less than 3 μs in all of the 25 measurements. Thus, the difference in path distances is less than 900 m.

- The channel update rate is 109 μs which corresponds to a new estimate of the impulse response every 0.9-2.7 mm at vehicle velocities in the range 30-90 km/h. That is we obtain at least 60 channel samples per wavelength. Thus we are working on the 10^{-3} m scale. With such small intervals, stationarity over the channel identification interval is well justified.

The goal is to predict the channel into the near future. In Table 3.2 three properties of the predictor are listed and we will discuss how these properties connect to the scales of the channels.

- The predictor is trained on a data set, which we shall call the *adaption interval*. The adaption interval for the predictors based block methods, can be selected as the whole measurement, that is 156 ms (or 1430 impulse responses) which corresponds to 1.3-3.9 m of traveled distance at vehicle velocities in the range 30-90 km/h. In other words it comprises 8-25 wavelengths. The geometry of the radio environment, and the parameters known from Chapter 2, governing the channel evolution can be assumed to be fairly constant over such a distance, but as is evident from Section 2.3 there are scenarios where a linear channel model is unsuitable even for such short intervals. Thus, nonlinear and adaptive predictors will be investigated.

- The prediction range defines how far ahead we aim to predict the channel. A prediction range of 1-10 ms, the longer the better, is desired to perform adaptive resource allocation. In other words, we thus want to predict 0.8-25 cm ahead, that is, 0.05-1.6 λ, or even longer.

- The prediction memory describes how much of the past data that is used in the prediction. With a memory of 15 ms, channel samples from the last 12.5-37.5 cm, that is 0.8-2.3 λ, are used for prediction. Predictors generally perform at their best if the memory contains at least one full cycle of the dominant oscillation [19] . If we take the Doppler frequency to be the dominant frequency, the memory should cover at least one wavelength.

The large separation of scales is a key property that makes it possible to find valid prediction models for the channel.

3.2 Identification Procedure

In this work, the radio channel is described by a discrete time transfer function, i.e., a discrete time impulse response. This transfer function, denoted $\{h_k(t)\}$, is time varying and the goal of the identification procedure is to estimate the time-dependent parameters in $\{h_k(t)\}$ as accurately as possible. The received signal is described as a convolution between the transmitted (puls-shaped) signal, $x(t)$, and the impulse response $\{h_k(t)\}$ corrupted by an additive noise, $n(t)$:

$$y(t) = \sum_{k=0}^{\infty} h_k(t)x(t-k) + n(t). \tag{3.1}$$

Over short time intervals, the time-varying channel can be approximately described by short segments of time invariant impulse responses. A further simplification is to assume that the transfer function, $\{h_k(t)\}$, can be described by a time invariant finite impulse response (FIR) model of length M (following the result in (2.28)), in each time interval. Equation (3.1) is then reduced to

$$y(t) = \sum_{k=0}^{M-1} h_k x(t-k) + n(t), \tag{3.2}$$

where M has to be chosen large enough to encompass all significant contributing paths. By expressing (3.2) using in the unit delay operator q^{-1} ($q^{-1}x(t) = x(t-1)$), we obtain

$$y(t) = H(q^{-1})x(t) + n(t), \tag{3.3}$$

where $H(q^{-1}) = \sum_k h_k q^{-k}$. This model is valid for short time segments, where short means in relation to the channel variation[1]. Here batches of N data samples are used to obtain a description of the impulse response $\{h_k\}$ for each batch. The choice of N is a trade off between bias and variance. A short batch of data results in high time resolution, but as fewer data points are used, the variance of the estimates will be higher than for a longer batch. If, on the other hand, a long batch is used, then the channel parameters will change during the estimated interval, which causes bias.

In this work we advocate the use of short batches to estimate the impulse response FIR parameters. Thus we obtain practically no bias. Furthermore we use of filtering of the estimated parameters of the channel, to reduce the variance (see Section 3.3). In the following sections a few estimation methods are reviewed.

3.2.1 Empirical Transfer Function Estimate

The convolution in (3.2) can be expressed as a multiplication in the frequency domain, i.e.,

$$Y(\omega) = H(\omega)X(\omega) + N(\omega). \tag{3.4}$$

The empirical transfer function estimate (ETFE) of $H(\omega)$ is then simply given as [20]

$$\hat{H}(\omega) = Y(\omega)/X(\omega) \tag{3.5}$$

[1] If the channel is slowly or not at all varying, then the time segments may be long.

and the impulse response estimate $\hat{h}_k$ is given by the inverse Fourier transform of $\hat{H}(\omega)$. This method of identification in the frequency domain usually gives results with unnecessarily high variance but due to its simplicity it is still worth attention.

The variance in the frequency domain of an estimated transfer function will depend on signal to noise ratio in the frequency domain [21] . The power spectrum of the reference signal (from the back-to-back measurement), as seen in Figure 3.4, shows that the transmitted signal in the measurements does not excite frequencies beyond 2.5 MHz from the center frequency. This will result in low accuracy of the estimated channel impulse response for the highest frequencies, regardless of which identification method is chosen.

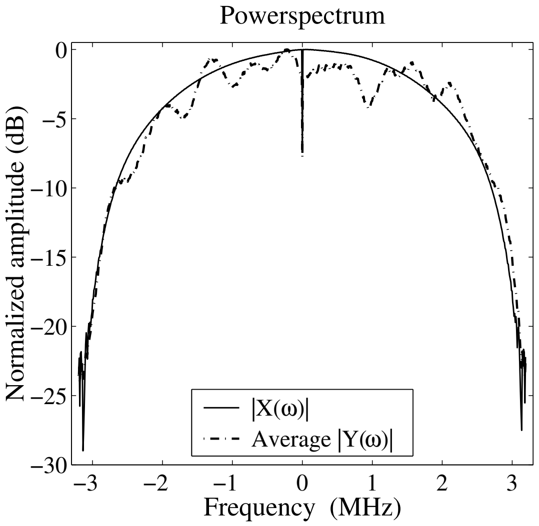

Figure 3.4: Normalized power spectrum of the transmitted signal $|X(\omega)|$ and the average normalized power spectrum of the received signals at one measurement location (real data).

3.2.2 The Least Squares Method

The use of the empirical transfer function estimate would cause an unnecessarily high variance on the estimated parameters. The problem is particularly apparent for taps with low amplitude. To obtain a better overall estimate of the channels, a least squares estimator [20] will be used.

Write equation (3.2) in vector form as

$$y(t) = \mathbf{x}(t)^T \mathbf{h} + n(t), \tag{3.6}$$

where

$$\mathbf{x}(t) = [x(t)\ x(t-1) \ldots\ x(t-M+1)]^T \tag{3.7}$$

is the vector of transmitted data[2] and

$$\mathbf{h} = [h_0\ h_1 \ldots h_{M-1}]^T \tag{3.8}$$

is the true parameter vector and $n(t)$ is a noise term assumed to have zero mean and variance σ_n^2. The total number of observations in one batch is N and the number of unknown parameters in $\mathbf{h}$ is M. Form the Toeplitz matrix for the transmitted data

$$\mathbf{X} = [\mathbf{x}(1) \ldots \mathbf{x}(N)]^T \tag{3.9}$$

and a received data vector

$$\mathbf{y} = [y(1) \ldots y(N)]^T. \tag{3.10}$$

Similarly let $\mathbf{n}$ denote the vector of the N noise samples,

$$\mathbf{n} = [n(1) \ldots n(N)]^T. \tag{3.11}$$

The Equation (3.6) can then for $t = 1 \ldots N$ be formulated as

$$\mathbf{y} = \mathbf{X}\mathbf{h} + \mathbf{n} \tag{3.12}$$

and the off-line LS estimate of the parameters of the channel is

$$\hat{\mathbf{h}} = \mathbf{X}^\dagger \mathbf{y} = \mathbf{h} + \mathbf{X}^\dagger \mathbf{n} = \mathbf{h} + \mathbf{v}, \tag{3.13}$$

where $\mathbf{X}^\dagger = (\mathbf{X}^H\mathbf{X})^{-1}\mathbf{X}^H$ denotes the Moore-Penrose inverse of the matrix $\mathbf{X}$ and $(\cdot)^H$ denotes complex conjugate transpose.[3] The estimation error will be $\mathbf{v} = \mathbf{X}^\dagger\mathbf{n}$.

Under the assumption that the noise $n(t)$ in (3.6) is independent of the term $\mathbf{x}(t)^T\mathbf{h}$ and that $\mathbf{h}$ is time invariant, the covariance matrix of the estimated parameters $\hat{\mathbf{h}}$ is given by

$$\begin{aligned} \mathrm{cov}(\hat{\mathbf{h}}) &= E\{(\hat{\mathbf{h}} - \mathbf{h})(\hat{\mathbf{h}} - \mathbf{h})^H\} = E\{\mathbf{v}\mathbf{v}^H\} \\ &= E\{\mathbf{X}^\dagger\mathbf{n}\mathbf{n}^H\mathbf{X}^{\dagger H}\} = \mathbf{X}^\dagger\mathbf{R}_n\mathbf{X}^{\dagger H}, \end{aligned} \tag{3.14}$$

[2]Since the transmitted sequence is repeated, the samples with negative index in $\mathbf{x}(t)$ are simply collected from the end of the preceding transmitted batch.

[3]The vector $\mathbf{X}^H\mathbf{y}/M$ is an estimate of the cross-covariance vector, $\mathbf{r}_{yx}$, and $\mathbf{X}^H\mathbf{X}/M = \hat{\mathbf{R}}_x$ is the sample auto-covariance matrix of $\mathbf{x}(t)$. The LS solution in (3.13) is thus nothing but the Wiener-Hopf equations with this particular choice of estimators for the covariances.

where the noise covariance matrix, $\mathbf{R}_n$, is defined as

$$\mathbf{R}_n = E\{\mathbf{n}\mathbf{n}^H\}. \tag{3.15}$$

The average power of the parameter vector, here approximated as the expectation,

$$E\{\hat{\mathbf{h}}\hat{\mathbf{h}}^H\} = E\{\mathbf{h}\mathbf{h}^H\} + \mathbf{X}^\dagger \mathbf{R}_n \mathbf{X}^{\dagger H}, \tag{3.16}$$

is of interest in the following discussion about the power delay profile.

When the noise is white, the noise covariance has the simple form $\mathbf{R}_n = \sigma_n^2 \mathbf{I}$ and the estimated parameter covariance in equation (3.14) reduces to

$$\mathrm{cov}(\hat{\mathbf{h}}) = \sigma_n^2 (\mathbf{X}^H \mathbf{X})^{-1}. \tag{3.17}$$

An unbiased estimate of σ_n^2 for white noise can be found as

$$\hat{\sigma}_n^2 = \boldsymbol{\varepsilon}^H \boldsymbol{\varepsilon}/(N - M). \tag{3.18}$$

where ϵ is the equation error:

$$\boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{X}\hat{\mathbf{h}} = \mathbf{y} - \mathbf{X}(\mathbf{h} + \mathbf{v}) = \mathbf{X}\mathbf{X}^\dagger \mathbf{n}. \tag{3.19}$$

3.2.3 Power Delay Profile Estimates

The power delay profile (PDP) is an average measure at one location, for how the received power is distributed over delays, that is, the time average of the squared amplitude of the taps. Through the PDP the effective length of the channel can be observed and it can also be used to measure the delay spread. Here we will use the PDP to obtain an estimate of the variance of the estimated parameters of the channel, where estimation errors will act as noise on the sequence of channel estimates.

The PDP is estimated as the average over one measurement location of the estimated power in the taps $[|\hat{h}_0|^2 \dots |\hat{h}_{M-1}|^2]$. Assume the channel tap to be time invariant over the estimation window of length N and assume the noise to be independent of the regressors. The expected value for the power in one tap, say tap k, estimated by the LS procedure from data corrupted by *white noise* is then obtained by using (3.17) in (3.16):

$$E\{|\hat{h}_k(t)|^2\} = \sigma_n^2 [(\mathbf{X}^H \mathbf{X})^{-1}]_{kk} + E\{|h_k(t)|^2\} \tag{3.20}$$

where $(\cdot)_{kk}$ denotes the element on row and column k and the variance of the white noise is σ_n^2. The *noise floor* is the power level at those parameters that in reality are zero. Thus, if $h_k(t) = 0, \forall t$, then $E\{|\hat{h}_k(t)|^2\} =$

$\sigma_n^2[(\mathbf{X}^H\mathbf{X})^{-1}]_{kk}$. As the PDP is formed through averaging over the available realizations, in our case 1430 realizations, a rather accurate description of the noise floor $\mathbf{n}_f$, is given by

$$\mathbf{n}_f = \sigma_n^2 \operatorname{diag}[(\mathbf{X}^H\mathbf{X})^{-1}] \tag{3.21}$$

where diag$(\cdot)$ forms a column vector from the diagonal elements of a matrix. The noise floor for tap k will be given by the kth element of this noise floor vector $\mathbf{n}_f$. We cannot expect to obtain any reliable estimates of the taps close to the noise floor. It could even be beneficial to exclude those delays from the estimation procedure [22].

Verification of the noise floor for white measurement noise

To verify that the level of the noise floor, and thus the variance of the estimation error, is as expected we compare a measured PDP with a Monte Carlo simulated profile. Here we use a dB scale with an arbitrary reference to display the measured data. This way of presenting measured data will be used throughout the thesis.

The simulated channel has a PDP formed after a pattern given by the measured PDP. Some taps in the simulated channel should carry no signal power, so values in the pattern PDP under an ad hoc chosen threshold is set to zero. The square root of the measured PDP, with values under a threshold of -101.4 dB set to zero, is thus used as a pattern for the simulated channel parameters. The threshold level is chosen a little bit above the observed noise-floor in Figure 3.5 to include all the visible peaks from the measured PDP into the pattern.

For each Monte Carlo trial the taps in the channel are taken as the pattern multiplied with normal distributed random complex numbers with zero mean and variance one. The obtained input channel has independent taps and the desired PDP. The input signal used in the measurements are transmitted through the simulated channel and the received signal is corrupted by an additive white Gaussian noise, the measurement noise. The noise has the same variance as the one estimated from the measurements using equation (3.18). The simulated channel is then identified using equation (3.13), just as for the measurements.

A total of 100 Monte Carlo trials are performed with different noises and channel parameters. In Figure 3.5 and 3.6 the PDPs for one measurement location and the average result from the 100 Monte Carlo trials are plotted respectively. In the measured data, Figure 3.5, there is significant deviation between the noise floor calculated as in (3.21) (dash-dotted curve) and the

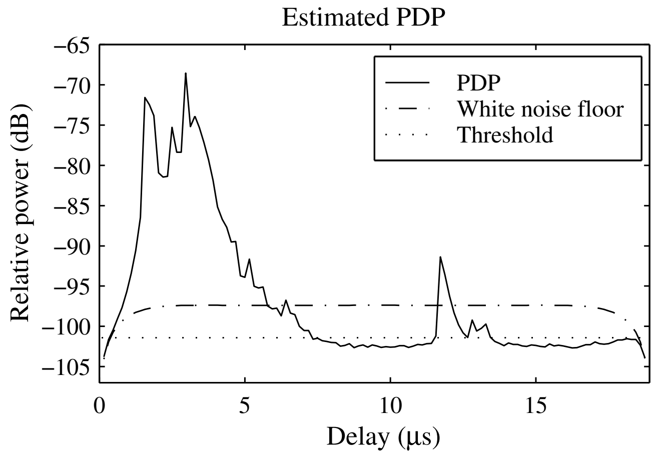

Figure 3.5: The estimated PDP for measurement 23 and the calculated noise floor under the white noise assumption (3.21). The part of the PDP under the threshold is set to zero in the simulation.

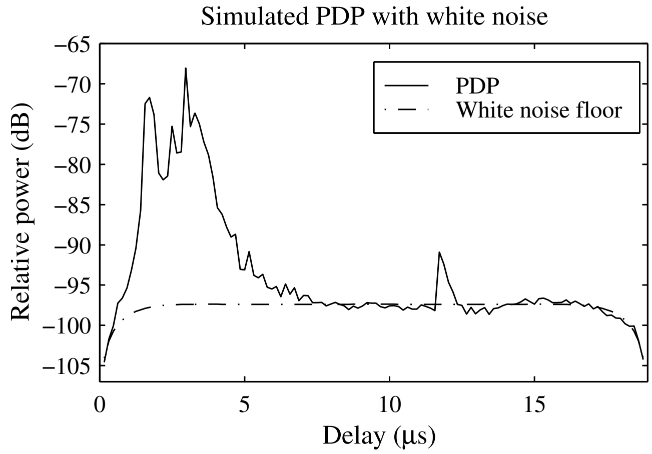

Figure 3.6: The simulated PDP with white noise. Measurement 23 is used as a pattern for generating a set of channels used for Monte Carlo simulations.

average power of the taps with the least power (dotted curve). This is not the case in the simulated PDP with white noise. As can be seen in Figure 3.6 the estimated noise floor for the simulated PDP coincides with the power level for the smallest estimated taps, those that in the pattern are zero and thus carry no signal power and only noise power, as expected from the theory. The difference between the result expected from theory and the measured PDP (in Figure 3.5) indicates that some assumptions about the data or the model are wrong. A possible error source could be the presence of colored measurement noise.

Noise floor for colored noise

To investigate the presence of colored noise we identify the channel using a prediction error method (PEM) with a more general model structure, including a noise-filter $C(q^{-1})$

$$y(t) = H(q^{-1})x(t) + C(q^{-1})v(t) \tag{3.22}$$

where $v(t)$ is a white noise [21]. In this example we identify a moving average noise-filter $C(q^{-1})$, of degree five. Even though the identified channel parameters $h_k(t)$ showed large variations over the time t, the identified noise-filter parameters were rather stable as can be seen in Figure 3.7.[4] The average parameters for the noise coloring filter $C(q^{-1})$ are $[1.0, 0.20 +$

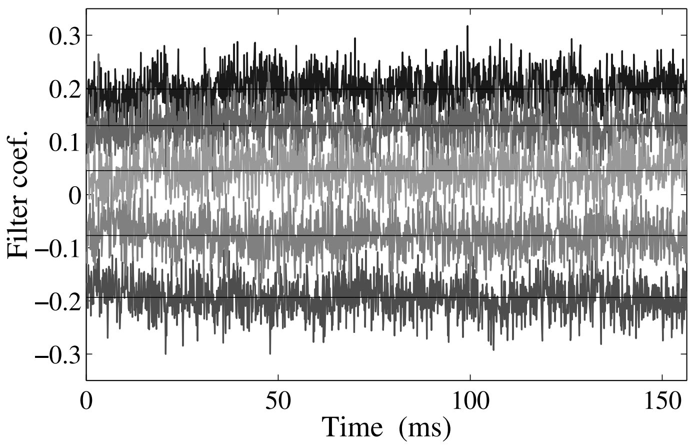

Figure 3.7: Real part of the identified noise-filter parameters from measurement 23. The darker the parameter the smaller the delay. The solid lines are the means.

$0.03i, -0.2, 0.1, -0.08, 0.05]$, a rapidly decaying impulse response resulting in a weakly colored noise. The assumption of white noise is thus not valid.

The LS estimate $\hat{\mathbf{h}}$ remains unbiased even with colored noise, but the expected value for the power will change from what is given in equation

[4]This was somewhat unexpected, as the properties of environmental noise would change just as fast as the mobile radio channel when the mobile moves. If, on the other hand, the noise is produced internally, in the receiver, the noise filter could be close to time-invariant.

(3.20) to

$$E\{|\hat{h}_k(t)|^2\} = [\mathbf{X}^\dagger \mathbf{R}_n \mathbf{X}^{\dagger *}]_{kk} + E\{|h_k|^2\}. \tag{3.23}$$

The noise floor is then altered to

$$\mathbf{n}_f = \mathrm{diag}(\mathbf{X}^\dagger \mathbf{R}_n \mathbf{X}^{\dagger *}), \tag{3.24}$$

where $\mathbf{R}_n$ is the covariance matrix (3.11),(3.15) of $n(t) = C(q^{-1})v(t)$.[5] The noise floor will be lower for the colored noise than for white noise with the same variance σ_n^2, since the variance of $v(t)$ will then be lower than than the variance of $n(t)$. To calculate the noise floor as in (3.24) the noise covariance $\mathbf{R}_n$ must be estimated.

When an additive colored noise is introduced in the simulations, a similar effect as for the measured data can be obtained, see Figure 3.8. The simulation is performed in the same manner as previously described, using LS for estimating the impulse responses, but the noise is now generated as a white noise filtered by a 40 tap long exponentially decaying noise filter. This is an ad hoc choice, built on the assumption that the echoes represented by the noise term decay rapidly. (If there is internal colored noise in the receiver, there are no echoes. Still the rapidly decaying impulse response is similar to that observed for the noise filter in measurements.) The taps in the noise-filter are selected as $c_k = (b_k + 1)e^{-k/2}/b$, where b_k is the absolute value of a normed normal distributed random variable (zero mean and variance one). The normalization factor b is selected so that the filter does not change the variance of the noise, i.e. $\sum |c_k|^2 = 1$. The variance of the colored noise is the same as for the white noise in the previous simulation. The colored measurement noise results in a lower noise floor than white noise with the same variance, when using LS-estimation for the impulse response.

The average parameters for the estimated noise-filter and the variance estimate from the PEM have been used to estimate the covariance matrix $\hat{\mathbf{R}}_n$ for the colored measurement noise on the measured data. The resulting noise floor, obtained as in (3.24), is in agreement with the observed noise floor in the measured PDP (estimated by LS).

From experience with measurements and simulations we can conclude that the observed lowest levels in the PDP can be used as an estimate of the channel tap estimation error variance. Thus, an estimated tap can be modeled as the true value and an estimation error, as

$$\hat{h}_k(t) = h_k(t) + v_k(t), \tag{3.25}$$

[5] The equation (3.20) is just a special case of (3.23) with $\mathbf{R}_n = \sigma_n^2 \mathbf{I}$.

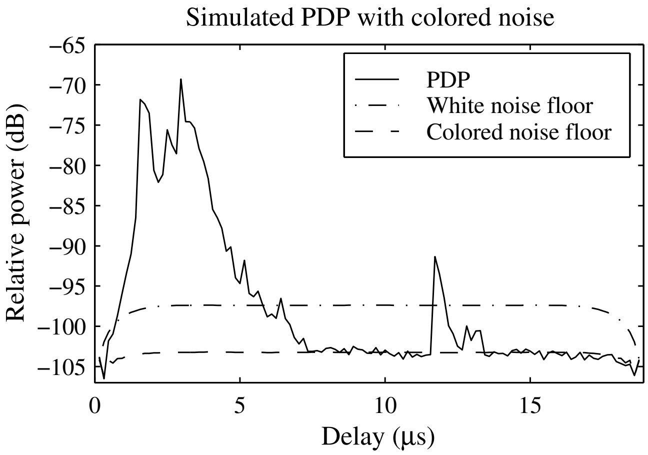

Figure 3.8: Simulated PDP, estimated by LS, with colored noise. The colored noise has the same variance as estimated from the measurement. The coloring noise channel impulse response is a rapidly decaying exponential function.

where the variance of the estimation error $v_k(t)$ is given by the noise floor. There is thus no need to estimate the variance using (3.24). We have also seen that calculating the noise floor using (3.21) can result in an overestimation of the noise floor. The estimated error variance will be used in the sequel in a noise-reduction smoother, for filtering of the estimated channel parameters.

3.2.4 Choice of Identification Procedure

All the tested methods, ETFE, LS and PEM give roughly the same results for the taps that contain the most energy. To find the best method we look at the lowest level (the noise floor) in the estimated PDP. As discussed in Section 3.2.3, the noise floor in the PDP can be linked to the variance of the estimation error.

The ETFE produced estimated channel impulse responses where the lowest level in the PDPs are a few dB over the lowest levels produced by the LS estimation (this level varies from measurement to measurement though). The PEM, taking the colored noise into account, also results in higher lowest level in the PDP than the LS-method. This is because PEM estimates more parameters, which causes a higher variance in the estimates. In addition it is generally hard to estimate the noise-filter, $C(q^{-1})$ in (3.22), with good accuracy from a limited amount of data. Thus, we do not gain anything in accuracy for the estimated impulse responses by using PEM instead of LS. As discussed in Section 3.2.3, the lowest level in the PDP can be linked to the variance of the estimation error. Since both the ETFE and PEM methods produce estimates with higher variance than LS for the taps with low amplitude, the LS-method is selected for identification of the channel.

3.3 Reduction of Noise on Estimated Channel Impulse Responses

The LS-estimated channel tap vector forms a time series with sampling rate 109 μs. As outlined in Section 3.2.3, this time series will be corrupted by noise, with a power indicated by the noise floor of the power delay profile (PDP). By filtering the sequence of tap estimates the noise level can be reduced, and the noise floor lowered, as seen in Figure 3.9. The dashed line is the estimated PDP after that noise reduction is applied to each tap. The method for noise reduction will be explained in the following.

The same noise floor as in the PDP can be observed in the Doppler spectrum depicted in Figure 3.10.[6] It is the average over the Doppler spectra for the taps, for the same measurement as for the PDP in Figure 3.9. Most of the power is in the band of ± 60 Hz, where 60 Hz is the Doppler frequency $v/\lambda = f_D$. There are also lower peaks around -145 Hz, 30 dB below the largest peaks. Outside the band [-240,160] Hz the average Doppler spectrum is close to the level of the noise floor observed in the PDP. This effect could be caused by white noise on the estimated channel taps.[7] Still, we can not say anything about the color of the noise inside the band [-240,160] Hz.

Due to the approximate band limitation to $\pm f_D$ of the frequency content of a tap and the high oversampling in relation to the Doppler frequency, it is possible to reduce the noise level. An ideal low-pass filter cutting off all frequencies outside $[-f_D \ f_D]$ would give a large noise reduction [24] but it would not include the part of the tap energy outside the Doppler band. It would also be hard to implement without long delays. The best noise reduction should optimize amplification with regard to the signal power and

[6] The Doppler spectrum is estimated using a Capon method, (MASC) [23], to avoid the high side-lobes and low resolution in ordinary Fourier transform estimation of the spectrum.

[7] Even though the measurement noise in the 6.4 MHz sampled measurements used for channel identification was colored, the estimation noise on the consecutive LS-estimated taps can be close to white.

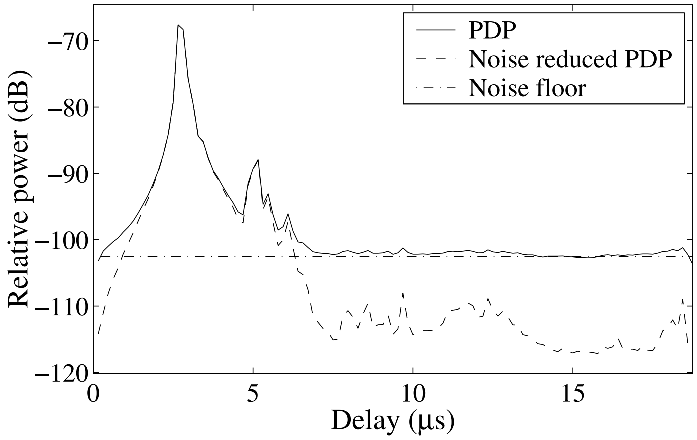

Figure 3.9: Power delay profile (PDP) and noise floor for one measurement location. The noise reduced PDP is estimated from the taps, after the proposed noise reduction.

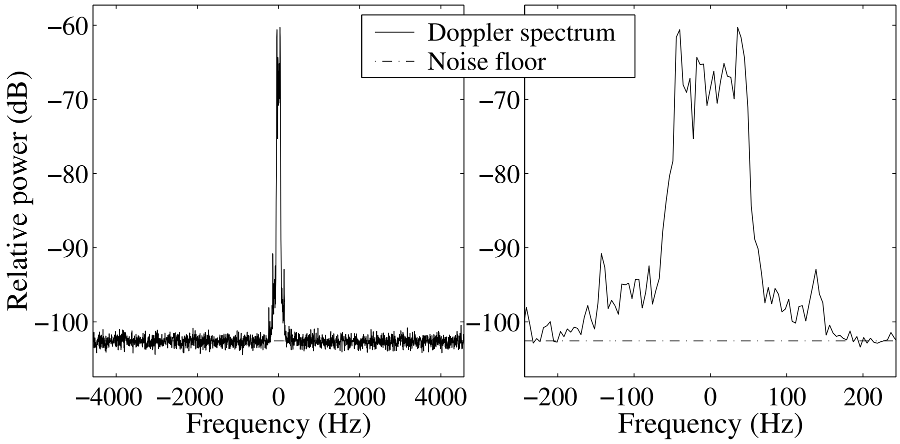

Figure 3.10: Average Doppler spectrum (zoomed to the right) and noise floor for the same measurement as used in Figure 3.9.

the noise power at each frequency and still be able to retain a low delay. In the following we will present such a method.

Design of noise reducing Wiener-smoother

To be able to design a smoothing filter that suppresses the noise and lets the signal through with as little distortion as possible, we need to know how the power of the signal is distributed in the frequency domain. By using crude ARMA models for the channel dynamics, a robust smoother can be derived.

A channel tap, $h_k(t)$, can be modeled as a white noise, $e(t)$, filtered through a low-pass filter,

$$h_k(t) = \frac{B(q^{-1})}{A(q^{-1})} e(t). \tag{3.26}$$

Here the time index t denotes samples taken at the channel sampling rate 109 μs and not as previously at the symbol rate 6.4 MHz. Even though this filter cannot produce the fine structure of the power-spectrum of a tap, it is sufficient for indicating what levels of signal power that could be expected for different frequencies. Thus, the frequency response of $B(q^{-1})A^{-1}(q^{-1})$ scaled by the variance of the driving noise $e(t)$ serve as a model for the Doppler spectrum of the tap $h_k(t)$. Given the filter $B(q^{-1})A^{-1}(q^{-1})$ and the variance of the tap $h_k(t)$, the variance of the driving noise $e(t)$ can be calculated.

An additive measurement noise (the estimation error), $v(t)$, corrupts the tap. The Wiener fixed-lag smoother,

$$\hat{h}_k(t-m|t) = \frac{Q_0(q^{-1})}{R_0(q^{-1})}(h_k(t) + v(t)), \tag{3.27}$$

is designed to optimally reduce the effect of the measurement noise, at a given SNR [25]. As the Doppler spectrum is flat outside the Doppler frequency, the noise, $v(t)$ is modeled as white. The estimated level of the noise floor is used as the variance of the measurement noise $v(t)$ in (3.27). The procedure results in an estimate of the tap, delayed with the smoothing lag m, see Figure 3.11.[8] The smoother is designed to reduce the noise

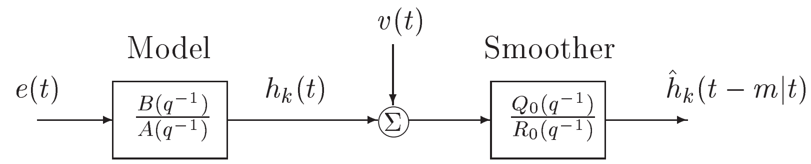

Figure 3.11: The model for smoothing design for noise reduction of a tap.

(estimation error) without introducing unnecessary bias.

[8] In an online application the last available samples of the channel will need smoothing lags shorter than m.

If the model $B(q^{-1})A^{-1}(q^{-1})$ is a low-pass filter with cut-off frequency at f_D and the noise is white, then the smoother $Q_0(q^{-1})R_0^{-1}(q^{-1})$ will be a low-pass filter that has an approximately flat frequency response with amplification one, and a linear phase in the passband $[-f_D\ f_D]$. Thus all frequencies within the passband will be left unchanged whereas power outside the passband will be attenuated.

Performance of noise reduction on simulated data

The performance of the noise reduction is studied on two different types of simulated data, all with the maximum Doppler frequency $f_D = 104.4$ Hz (this corresponds to a vehicle velocity of 60 km/h at a carrier frequency of 1880 MHz) and channel sampling at $f_s = 9.14$ kHz.

The first set of simulated data consist of an approximately Rayleigh distributed time series formed through a weighted summation of 100 complex sinusoids with frequencies taken as $f_D \cos\theta_n$ and where the angle θ_n is a stochastic variable drawn from the uniform distribution $[0\ 2\pi[$. The weights in the summation are complex and Gaussian distributed.

The second set is the simulation set in Section 2.4, modeling a car traveling by three close scatterers. Noise is added to give different SNR conditions and the SNR improvement after noise reduction is measured for a range of smoothing-lags. In both cases 6000 data points are used and 100 Monte Carlo trials are performed with different noise realizations.

The filter $B(q^{-1})A^{-1}(q^{-1})$ in (3.26) is chosen as a Chebyshev type one low-pass filter of degree 4 with 4 dB passband ripple and the cut-off frequency at the Doppler frequency f_D. The filter serves as the rough approximate model for the dynamics of the taps. For each SNR and smoothing-lag, m, a Wiener-smoother is designed according to [25]. In Table 3.3 average gain in SNR, over both Monte Carlo trials and data sets, is presented.

By increasing the smoothing-lag the performance is improved, but we increase the delay in the prediction so a longer prediction range will be needed. Thus, the smoothing-lag should be kept as low as possible. As can be seen from Table 3.3, there is little to gain by increasing the smoothing-lag above 5 samples. Actually, for higher SNRs, that is above 10 dB, most of the gain is already achieved with a smoothing-lag of 3 samples. To increase the smoothing-lag from 3 to 5 samples increase the gain in SNR about 1 dB. A further increase from 5 to 30 samples, result in less than 3 dB improvement, which is relatively small when the SNR is around 20 dB. The SNR gain saturates when the smoothing lag is further increased, since then all the noise outside the band-limits is suppressed and the remaining noise

	Smoothing-lag						
SNR	0	3	5	7	10	20	30
0	10.4	11.6	12.4	13.1	14.0	15.0	15.1
5	9.2	10.8	11.8	12.7	13.7	14.5	14.7
10	8.1	10.2	11.5	12.5	13.5	13.9	14.3
20	6.4	9.6	11.2	12.1	12.5	13.0	13.6
30	5.0	9.3	10.8	11.2	11.3	12.4	12.5

Table 3.3: Table over the average gain in SNR (dB) for the two simulation sets after Wiener smoothing, for different SNR and smoothing lags. The difference between the two sets was less than ±0.1 dB, so the total average is shown here.

within the band-limits can not be filtered away. To reduce the noise within the band-limits more involved signal models, taking the fine structure of the power-spectrum into account, have to be used instead of the simple low-pass model.

The average gain in SNR differed less than ±0.1 dB between the two data sets. As the results are so similar for two different types of signals, we can assume that the gain in SNR will be close to those indicated in the table for most mobile radio channels with f_D =104.4 Hz and for channel sampling at 9.14 kHz. The gain in SNR decreases when the ratio $2f_D/f_s$ decreases as then there is less noise outside the frequency band occupied by the signal, that can be suppressed. A perfect low-pass filter cutting away everything outside a perfectly band-limited signal would increase the SNR by $10\log_{10}(f_s/2f_D)$ dB, in this case 16.4 dB.

The properties of a Wiener smoother approaches those of an ideal low-pass filter when $m \to \infty$ and when the design is performed for high SNR's.

Noise reduction on estimated channels

To reduce the noise level in the wide-band measurements we use a Chebyshev type one low-pass filter of degree 4 with 4 dB passband ripple. The cut-off frequency is set to the Doppler frequency for a vehicle velocity of 105 km/h, to encompass all the possible Doppler frequencies in the measurements. (A Chebyshev filter with high passband ripple has a faster roll-off than a Butterworth filter of the same order. As we will see, the ripple shows no significant effect on the final smoother.)

Using the Chebyshev filter together with the knowledge about the noise level (from the PDP), a Wiener-smoother, with smoothing lag of 5 samples (introducing a delay of 0.5 ms), is designed. In the left part of Figure 3.12

the theoretical power spectrum of the signal $B(q^{-1})A^{-1}(q^{-1})e(t)$, that is the frequency response of the Chebyshev filter scaled by the variance of the driving noise, is plotted together with the Doppler spectrum for the tap corresponding to the second peak of the PDP in Figure 3.9. The tap is

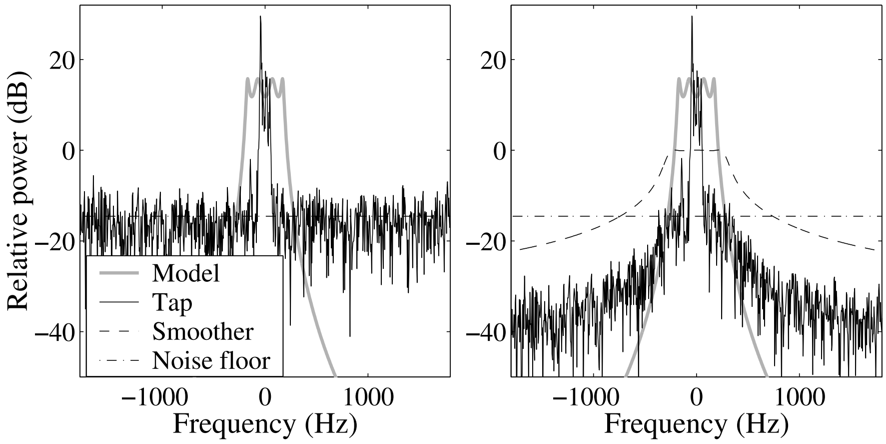

Figure 3.12: Power spectrum of the model and the Doppler spectrum for the tap corresponding to the second peak of the PDP in Figure 3.9, to the left. To the right the corresponding Doppler spectrum of the tap after noise reduction using the smoother (with dashed frequency response). The tap is normalized to have unit variance and the SNR is estimated to be 14 dB. The full bandwidth is ±4.6 kHz out of which the Figure shows ±1.8 kHz.

scaled to have unit variance. The right-hand part of Figure 3.12 shows the Doppler spectrum after noise reduction using the smoother. The smoother has 0 dB amplification (and close to linear phase) in the passband, even though the Chebyshev filter has a significant amount of passband ripple, and it attenuates the out of band power. Thus, the relevant part of the tap is left unaltered.

The smoother has a slightly broader passband than the Chebyshev filter. This is due to a trade-off between bandwidth and linearity of the phase in the passband. For a lower noise level the bandwidth would increase, to obtain a more linear phase in the passband whereas for a higher noise level the bandwidth of the smoother would decrease, to diminish the effect of the noise power. The reduction of bandwidth is done at the cost of a worse performance of the phase response. A longer smoothing-lag increases the performance especially for high noise levels.[9]

[9]To completely avoid a non-linear phase in the passband, FIR-filters can be used as

Noise reduction could alternatively be performed by the predictor and the optimization of the predictor would then include the attainment of sufficient noise reduction. However, there is a clear advantage in separating the noise reduction from the prediction. In the noise reduction is it possible to exploit the high oversampling of the channel and to use a simple model for the dynamics of the taps, to obtain a pre-designed low complexity smoother. For predictors with a prediction ranges of 5 ms, a long memory, on the order of at least 10-20 ms, is needed. The predictor is therefore preferably sub-sampled to keep the complexity low. Such a sub-sampled structure can not exploit the oversampling for noise reduction. Thus, by separating the noise reduction from the prediction we keep the complexity low while we can still use predictors with structures directly designed to suite the dynamics of the taps.

in [3]. However the FIR-filter requires many more parameters and introduces a longer delay than does the Wiener-smoother, at a comparable performance.

3.A Bandwidth Reduction on Measured Channels

Radio frequencies are a scarce resource. To exploit the radio frequencies as efficiently as possible the signal that is transmitted has to be band-limited to avoid interference among the frequency bands. In mobile communication systems the bandwidth is generally not as wide as 6.4 MHz. Lower bandwidth also means lower spatial resolution in the impulse response which means that more scatterers and reflectors contribute to each tap. To study the behavior of the channel in a system with a lower bandwidth, the bandwidth of the measured channel has to be reduced. To create artificial channel-data with smaller desired bandwidths than in the measured data, the identified impulse responses are band-limited by filtering and then sub-sampled.

Due to the band-limiting there is redundancy in the information provided by the taps. The impulse response of the channel can thus be sub-sampled, that is, if we reduce the bandwidth a factor K, we only need to use every Kth tap in the impulse response, without loss of information. The low-pass filtering is performed over the taps of an estimated channel impulse response. The corresponding frequency domain transfer function, that originally occupies the band from -3.2 to 3.2 MHz, is thus band-limited by the filter to occupy the desired bandwidth. This produced no visible edge effects on the estimated impulse response when the identified impulse response covers the time where power is received. That is, outside the range of the impulse response the power is close to zero.

A more direct approach would be to low-pass filter the transmitted and received signals prior to identification. As the low-pass filter has a high stop-band attenuation the signals are close to band-limited. However this operation can cause numerical problems in the LS estimation of the band-limited channel. In the LS estimation the inverse of the sample covariance matrix estimate of the transmitted signal[10] is used to calculate the LS estimate of the channel. The sample covariance matrix has approximately a Toeplitz structure. The eigenvalues of a sequence of Toeplitz matrices constructed from the autocorrelation of a discrete random process are asymptotically distributed like the samples of the process' power spectrum [26]. As the transmitted signal is band-limited, the power in the stop-band will be very low in relation to the pass-band. The eigenvalue spread will thus be high and the covariance matrix will be ill conditioned. This causes numerical problems in the LS estimation.[11]

[10] $\hat{\mathbf{R}}_x = \mathbf{X}^H\mathbf{X}/M$ is a maximum likelihood estimate of the covariance matrix for $x(t)$

[11] The problem can be avoided by sub-sampling of the signals *after* filtering to widen the

To avoid numerical problems and to simplify the calculations we have chosen to perform the band-limiting filtering on the estimated channel impulse responses instead of prior to the estimation.

The band-limiting filters are 10th order Bessel-filters with the 3dB bandwidth specified to 5, 2.5 and 1.25 MHz. In communication applications filters with much more rapid roll off than those used here (see Figure 3.13) are used. A common requirement is that the filter gain has to be below -70dB outside the specified bandwidth. With this definition the bandwidths are roughly 6, 5 and 4 MHz respectively.

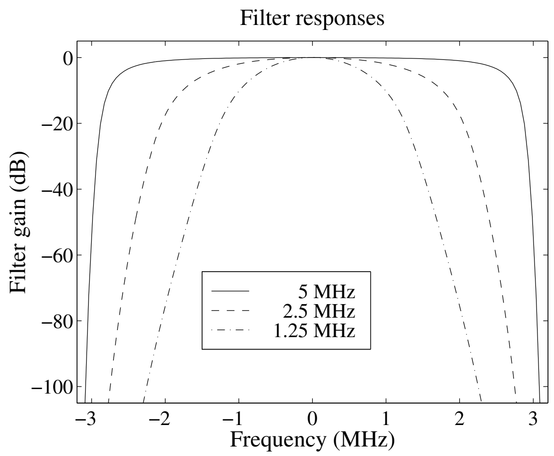

Figure 3.13: Filter gain for the Bessel-filters used for band limiting.

pass-band to cover the whole band. This would limit the possible choices of bandwidths to $6.4/K$ MHz where $K = 1, 2, \ldots$.

Chapter 4

Block Prediction Methods

The estimated impulse responses of the mobile radio channel are noisy observations of an underlying physical reality. In this work we are interested in the prediction of the impulse response, given previous observations of it. The taps of the impulse response can be treated as continuous time signals that are sampled through the estimation procedure, where the samples of the channel are corrupted by noise. It is the underlying noiseless channel that determines the conditions of transmission and reception. Using the techniques presented in Chapter 3 we reduce the effect of the noise in the estimated channel impulse responses. This results in a discrete time series of estimated taps in the impulse response, with only small deviations from the parameters of the underlying channel, at the sampling instances. This discrete time series of taps is then used for prediction of the channel.

In prediction of a discrete time series $y(t)$, the values of $y(t+L)$ are forecasted using present and past values. When the predictor is limited to work with a finite time window T and a finite number of measurements p of the time series, the input to the predictor can be represented by the vector

$$\mathbf{y}(t) = [y(t-\tau_1)\ldots y(t-\tau_p)]^T. \tag{4.1}$$

The delays $[\tau_1 \ldots \tau_p]$ may be chosen in any manner over the finite interval T. With the delays in ascending order, the effective length of the memory is $\tau_p \leq T$. The representation in (4.1) can be regarded as the outcome of a subsampling of the acquired data. It enables the predictor to use a long memory, yet using a limited number of coefficients. For notational convenience (and without loss of generality) we define $\tau_1 = 0$.

Assume that $y(t+L)$ can be described by the regression model

$$y(t+L) = f(\mathbf{y}(t)) + \epsilon(t), \tag{4.2}$$

where $\epsilon(t)$ represents the part of the signal $y(t+L)$ which cannot be described by the regression of data up to time t. If the function $f(\cdot)$ is time-invariant or slowly varying, an estimate of the function based on a block of observations, could serve as a predictor. A predictor estimate of $y(t+L)$ would then be

$$\hat{y}(t+L) = \hat{f}(\mathbf{y}(t)), \tag{4.3}$$

where $\hat{f}(\cdot)$ is the predictor function.

In this chapter a number of different predictors, with structures that might make them suitable for prediction of the mobile radio channel, are presented, together with some design issues.

The *linear FIR-predictor* introduced in Section 4.2 is closely related to the AR-spectrum of the channel taps and can perfectly predict weighted sums of complex exponentials.

If close scatterers or other causes of deviation from the exponential model (2.47) are to be taken into account, then we have to include nonlinearities in the predictor. The simplest extension from linear to nonlinear predictors is to include quadratic terms in the regressor. The *quadratic predictor*, or truncated Volterra series, is introduced in Section 4.3. It is found to be unsuitable as it is hard to obtain estimates of the parameters that have good generalizing properties. The reason is that quadratic nonlinearity is in many applications too strong.

Two other alternatives are neural nets and MARS (*multivariate adaptive regression splines*) [10]. MARS is a structure for fitting of nonlinear functions. It can be interpreted as a gain-scheduling algorithm, where different regions of the regression space has its own simple model. We have here, in Section 4.4, chosen to use and investigate MARS instead of neural nets, as MARS has been demonstrated to perform comparably well on small training sets and can produce good models with fewer parameters than neural nets [27]. However, it will turn out that the MARS predictor, as well as the quadratic predictor, shows poor generalization properties in the prediction of mobile radio channels.

4.1 An Upper Bound on the Prediction Gain

A question that arises when trying to predict a time series is whether the signal is predictable at all. If the time series is stationary and generated by a linear process then the auto-covariance gives a good indication of the ability to predict the signal. If the signal is generated by a nonlinear process, then the auto-covariance is insufficient as an indicator. However information theory offers the *mutual information* measure, a measure on the time series that can be viewed as a generalized covariance. This will provide limits for the predictability of the time series.

The prediction error of the predictor in (4.3) is given as

$$\varepsilon(t+L) = y(t+L) - \hat{y}(t+L|t), \tag{4.4}$$

where $\hat{y}(t+L|t)$ denotes the predicted value at time t, given data up to time t. To quantify the prediction quality we introduce the term *prediction gain*. The prediction gain $G(L)$, which is measured in dB, is defined as the ratio of the variance of $y(t)$ to the power of the resulting prediction error $\varepsilon(t)$

$$G(L) = 10\log_{10}\frac{E\{|y(t)-m_y|^2\}}{E\{|\varepsilon(t)|^2\}}, \tag{4.5}$$

where

$$m_y = E\{y(t)\}. \tag{4.6}$$

Under the assumption that the time series $\{y(t)\}$ is stationary, an upper bound on $G(L)$ for any unbiased predictor is given by Bernhard in [28]

$$G(L) \leq G_{max}(L) = 6.02[I(y(t+L);\mathbf{y}(t)) + \Delta] \tag{4.7}$$

where

$$\Delta = \frac{1}{2}\log_2(2\pi \mathrm{e} E\{|y(t)-m_y|^2\}) - H(y(t)) \tag{4.8}$$

is the difference between the differential entropy of a Gaussian variable with the same variance as $y(t)$ and the first-order entropy, $H(y(t))$. Here $I(y(t+L);\mathbf{y}(t))$ denotes the mutual information, which is a measure of how much information the vector $\mathbf{y}(t)$ contains about the value to be predicted, that is $y(t+L)$ [29]. The constant 6.02 (or exact $20\log_{10}(2)$) is just a scaling factor from bits to dB and e is the natural base. $G_{max}(L)$ is an upper bound on the prediction gain for any predictor forecasting $y(t+L)$ based on $\mathbf{y}(t)$. It does not imply that there actually exist a predictor that can achieve the bound.

The mutual information between two processes X and Y is defined as

$$I(X;Y) = \int\int p_{X,Y}(x,y)\log_2\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}dxdy, \tag{4.9}$$

where the functions $p_{X,Y}(x,y)$,$p_X(x)$,$p_Y(y)$ are, respectively, the joint probability density of (X,Y) and the marginal densities of X and Y.

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 [30]. To obtain accurate estimates often requires large amounts of data. The upper bound on mutual information (4.7) does not help us to find the optimal predictor. Instead its main use is as a generalized correlation analysis, to indicate the predictability of the signal.

A simple example of the advantage of using the mutual information instead of correlation analysis is given in [31].

Example 4.1

Let an input signal $x(t)$ and a noise $n(t)$ be independent and generated by white, zero mean, Gaussian processes with variances σ_x^2 and σ_n^2 respectively. Consider a nonlinear channel where the received signal is given by

$$y(t) = a|x(t)| + n(t). \tag{4.10}$$

The cross correlation function

$$r_{yx}(\tau) = E\{y(t+\tau)x(t)\} = 0, \tag{4.11}$$

will be zero for all τ. On the other hand the mutual information function will measure the full dependence between the signals. We can use equation (4.9) and argue as follows: The information not contained in the input signal $x(t)$ about the output signal $y(t)$ is contained in the noise, $n(t)$. If we calculate the auto-mutual information $I(y(t);y(t))$, that is the total information content of $y(t)$, and subtract the information of the noise, what remains is the mutual information between $x(t)$ and $y(t)$. This can be formalized as

$$I(x(t);y(t+\tau)) = I(y(t);y(t+\tau)) - I(n(t);n(t+\tau))$$
$$= \begin{cases} -\int\limits_{-\infty}^{\infty} p_Y(u)\log_2 p_Y(u)du - \frac{1}{2}\log_2\left(2\pi e\sigma_n^2\right) & ,\tau = 0 \\ 0 & ,\text{otherwise.} \end{cases}$$

Here $p_Y(y)$ is the marginal probability density function of the received signal. Only relying on linear correlation analysis, which is blind to nonlinearities, would give the false impression that no knowledge about $y(t)$ can be obtained from knowledge about $x(t)$. Mutual information analysis on the other hand, detects the connection and tells us how much information the input signal $x(t)$ contains about the received signal $y(t)$.

The predictor[1] that achieves the bound for the prediction gain for estimating $y(t)$ given $x(t)$ is $a|x(t)|$. However, it does not exist a predictor for $x(t)$ given $y(t)$ that achieves the bound. It is only possible to predict the envelope of $x(t)$.

4.1.1 Predictability of Received Power in the Mobile Radio Channel

To get an estimate of how predictable the received power is from the obtained measurements, the mutual information between the power at time $t+L$ and two delayed measurements of the power $\mathbf{y}(t) = [y(t), y(t-\tau)]$ is estimated for the measured powers from all locations. The reason to use only two delays is to keep the accuracy of the mutual information estimate high in our limited data sets and yet have the extra structural information offered by the second delay. Using equation (4.7), an upper bound on the prediction gain for any predictor using two lagged variables is estimated. The measurements in the database presented in Section 3.1 are used.

To see how the length of the memory affects the prediction, we vary the delay spacing. (As we use only two delayed variables the memory length is the same as the delay spacing.) All different delay spacings in the range $\tau = [1, 15]$ ms (with steps of 1 ms) are evaluated and we select the delay spacing τ which gives the highest prediction gain bound. In Figure 4.1, the average over all measurements 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 prediction gain, increases less than 5 ms. Thus, *when the prediction interval is increased the memory size does not have to increase proportionaly.*

The average of the upper prediction gain bounds is presented in Figure 4.2. For predictors using two lagged variables, estimated through the mutual information, the upper bound on the prediction gain first drops 3-4 dB when the prediction interval is increased from 1 to 4 ms and then remains fairly constant. This indicates that when the dynamics are known it would be possible to increase the prediction interval with only a small loss of accuracy. The high bound for the prediction gain for long ranges can be an effect of a hidden determinism in the time series. Thus, it can be worth the

[1] As the time index for $y(t)$ and $x(t)$ are the same we would in the strict sense perform estimation, not prediction, of one of the signals given the other.

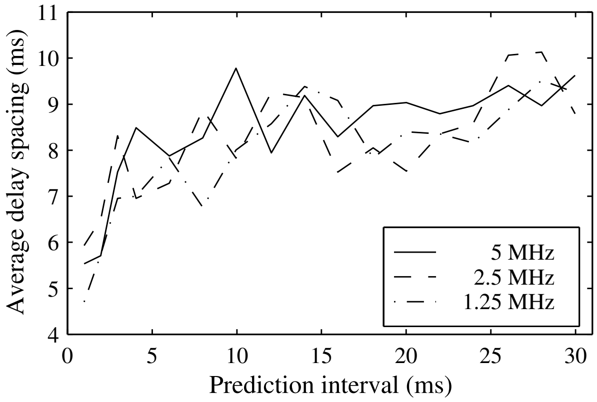

Figure 4.1: Average delay spacing τ in $\mathbf{y}(t) = [y(t), y(t-\tau)]$, over all measurements giving the upper prediction gain bound for the power, estimated with the mutual information algorithm.

effort to look for good predictors. The drop in the prediction gain bound

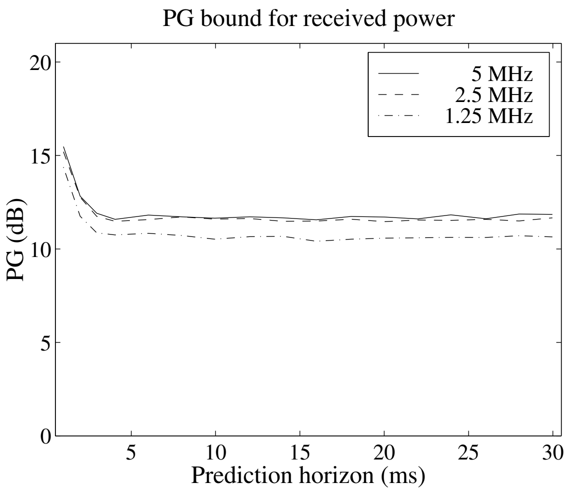

Figure 4.2: Upper bound on the prediction gain for received power at different bandwidths.

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

The received power is the sum of the contributions from many taps. Experiments indicate that it is easier to predict the power by prediction of the complex taps and then summing the squared envelopes of the predictions, than to predict the power directly. Unfortunately is it not possible to perform a study of the mutual information between the complex taps and the total power on the data at hand. This is due to that, even using only two delayed samples, the number of variables in the analysis would be to high for the method to obtain a reasonable estimate of the bound for the prediction gain. If as few as ten taps contribute to the received power, then there are ten complex values and thus 20 real values that influence the power. With two delays this corresponds to an X, in (4.9), of 40 variables. Here Y is just the received power. It is not feasible to estimate the mutual information for that many variables, with data sets of a length on the order of 10^3.

To summarize, the mutual information determines an upper bound on the prediction gain. However it is important to remember that there might not exist any realizable predictor that achieves this bound.

4.2 Linear FIR Filter Predictors

A linear FIR prediction of a zero mean complex-valued time series $y(t)$, based on measured data confined by (4.1) can be expressed as

$$\hat{y}(t+L) = \hat{f}(\mathbf{y}(t)) = \sum_{n=1}^{p} b_n y(t-\tau_n) \tag{4.12}$$

$$= \mathbf{b}^T \mathbf{y}(t), \tag{4.13}$$

where $\mathbf{b} = [b_1 \dots b_p]^T$ are the complex-valued weights for the different delayed measurements. In the following we will discuss how to estimate $\mathbf{b}$ from $y(t)$. For time series with nonzero mean, such as the received power in our case, we estimate the mean separately and work on the detrended data.

4.2.1 FIR filters minimizing the MSE

Under the assumption that the taps of the radio channel can be described by a weighted sum of complex exponentials (2.47), a linear FIR-filter can give perfect predictions [32]. The linear FIR filter of a pre-specified length, that minimizes the mean square of the prediction error is given by the LS solution, similar to the one described in Section 3.2.2. Assume that the time series actually can be described by the regression model

$$y(t+L) = \mathbf{b}_0^T \mathbf{y}(t) + \epsilon(t), \tag{4.14}$$

where $\epsilon(t)$ is uncorrelated with $y(t)$ and represents the part of $y(t+L)$ that cannot be described by a linear regression. The relation between the FIR-predictor and the ordinary AR-model $A(q^{-1})y(t) = \epsilon(t)$, is straightforward. The predictor can be expressed as the polynomial filter $B(q^{-1})$ given as

$$B(q^{-1}) = \sum_{n=1}^{p} b_n q^{-\tau_n}, \tag{4.15}$$

with τ_n expressed in samples. As before q^m is the shift operator shifting a sequence m steps forward. Equation (4.14) can then be written as

$$q^L y(t) = B(q^{-1})y(t) + \epsilon(t). \tag{4.16}$$

Multiply both sides with q^{-L} and collect the $y(t)$ components on the left hand side to obtain

$$(1 - q^{-L}B(q^{-1}))y(t) \triangleq A(q^{-1})y(t) = q^{-L}\epsilon(t) = \epsilon_A(t), \tag{4.17}$$

where we introduced $\epsilon_A(t) = q^{-L}\epsilon(t)$, to remove the irrelevant time shift in the innovation noise.

The one step ahead predictor, based on all the samples in the available memory, can thus be interpreted as a Doppler spectrum estimator where the spectrum is modeled by

$$\phi(\omega) = \frac{\sigma^2}{|A(\omega)|^2}, \tag{4.18}$$

with σ^2 being a positive scalar and where $A(\omega)$ is the polynomial

$$A(\omega) = 1 - b_1 \mathrm{e}^{-j\omega(\tau_1+L)} - \ldots - b_p \mathrm{e}^{-j\omega(\tau_p-L)}. \tag{4.19}$$

The AR equation, $A(q^{-1})y(t) = \epsilon_A(t)$ in (4.17), is able to model signals having spectra with narrow peaks by placing some zeros of the polynomial $A(q^{-1})$ close to the unit circle [32]. Since the Doppler spectrum of a tap in a mobile radio channel often has a rather peaky structure, as each wave front is distinguished by its Doppler frequency, the linear AR-structure might be pertinent for designing a predictor.

Thus, there exist a close relationship between the Doppler spectrum of a tap and the corresponding predictor coefficients. This relation has been used for prediction on simulated data, for sums of complex exponentials, in [5].

In our context it is crucial that the number of predictor coefficients, p, is large enough to model the main peaks occurring in the Doppler spectrum. Using an LS estimate of the FIR-predictor taps, no prior knowledge of Doppler frequencies and amplitudes has to be assumed. If, on the other hand, such information is available, then it can be advantageous to choose a predictor structure taking advantage of it. The approaches of [4], [3] and [6] are based on prior estimation of the peaks.

In [33] and [34] the introduction of feedback loops in a low order FIR-predictor is proposed to obtain a trade off between a low noise gain and a fast step response. This gives the predictor a faster response to changes. How this would affect the long term predictions is a topic for future research.

4.2.2 Complex exponential modeling using FIR filters

One approach to the prediction of future tap values is to first estimate the coefficients of the parameterization of $h_k(t)$ in equation (2.47) at time t and then use these coefficients to calculate a prediction of $h_k(t+L)$. In the Doppler spectrum associated with $h_k(t)$ there will be peaks at the Doppler frequencies corresponding to the incoming wave fronts. This approach, but for prediction in the frequency domain, was investigated by Bach-Andersen in [4] using ESPRIT and by Hwang and Winters [3] using MUSIC.

In the following we present a method with least squares estimates of the weight coefficients, based on the assumption that all the Doppler frequencies are known. This condition will later be relaxed with the introduction of orthogonal basis functions. The procedure has an interpretation as a pre-designed linear filter and is thus, as we shall see below, deemed to have worse performance than a predictor using the direct least squares estimate of the predictor coefficients. The only benefit is that of lower complexity. Prediction using the pre-designed linear filters can also be interpreted as phase advance of parts of the DFT spectrum.

Assume that a tap, here denoted $y(t)$, can be modeled as in (2.47) with an additive noise as

$$y(t) = \sum_{n=1}^{N} \alpha_n e^{j\omega_n t} + n(t). \tag{4.20}$$

Assuming that the frequencies $\{\omega_n\}$ are known, then (4.20) is a weighted sum of known basis functions. When a signal is generated by known functions of time, corrupted by additive noise, is it possible to estimate the weights for these functions and then use the estimate for prediction.

The basis functions can be collected in the vector

$$\mathbf{f}(t) = [e^{j\omega_1 t} \dots e^{j\omega_N t}] \tag{4.21}$$

and with the weights also ordered in a vector

$$\boldsymbol{\alpha} = [\alpha_1 \dots \alpha_N]^T, \tag{4.22}$$

we can write the sum in (4.20) as the vector multiplication

$$y(t) = \mathbf{f}(t)\boldsymbol{\alpha} + n(t). \tag{4.23}$$

Form the matrix of basis functions at $p+1$ time instances

$$\mathbf{F}(t) = \begin{bmatrix} \mathbf{f}(t) \\ \mathbf{f}(t-1) \\ \vdots \\ \mathbf{f}(t-p+1) \end{bmatrix}, \tag{4.24}$$

where $p \geq N$. If we use all the samples in the memory, that is no subsampling, then the data vector (4.1) is given by

$$\mathbf{y}(t) = [y(t), y(t-1) \dots y(t-p+1)]^T \tag{4.25}$$

and

$$\mathbf{y}(t) = \mathbf{F}(t)\boldsymbol{\alpha} + \mathbf{n}(t). \tag{4.26}$$

The least squares estimate of $\boldsymbol{\alpha}$ using the p latest samples, stacked in $\mathbf{y}(t)$, is readily obtained as

$$\hat{\boldsymbol{\alpha}} = \mathbf{F}^{\dagger}(t)\mathbf{y}(t). \tag{4.27}$$

The corresponding L step ahead predictor is

$$\hat{y}(t+L) = \mathbf{f}(t+L)\hat{\boldsymbol{\alpha}} \tag{4.28}$$

To be able to evaluate the performance of this prediction procedure we reformulate it as a linear filtering problem.

Theorem 4.1 *Consider the tap model (4.20)-(4.23) with complex exponentials as basis functions. The least squares estimation (4.27) of the coefficients $\boldsymbol{\alpha}$, in a data window of length p, followed by prediction using (4.28) is then equivalent to the linear filtering*

$$\hat{y}(t+L) = \mathbf{b}^T\mathbf{y}(t),$$

with the time invariant filter

$$\mathbf{b}^T = \mathbf{f}(L)\mathbf{F}^{\dagger}(0). \tag{4.29}$$

Proof 4.1 *See Appendix 4.A.1.*

The sensitivity of the filter to broadband noise is given by the *noise gain*. For an FIR-filter the noise gain is given by the sum of the squared amplitudes of the filter coefficients, i.e.,

$$NG = \mathbf{b}^H \mathbf{b}. \tag{4.30}$$

It is preferable that the noise gain is low as the predictor will then be insensitive to additive noise. As seen in the following Example 4.2, the use of basis functions with frequencies that cause the basis functions to be non-orthogonal to each other cause a high noise gain. If instead we constrain the frequencies, in the filter design, to only take certain discrete values

$$\omega_n = \frac{2\pi k_n}{N}, \tag{4.31}$$

then the basis functions will be orthogonal over the estimation interval, that is

$$\mathbf{F}^H \mathbf{F} = \mathbf{I}, \tag{4.32}$$

which also can be expressed as

$$\sum_{k=0}^{p-1} e^{-i\omega_n k} e^{-i\omega_m k} = \delta_{n-m}, \ \forall n, m. \tag{4.33}$$

We can, instead of the true frequencies, use the frequencies corresponding to orthogonal basis functions. The chosen frequencies belong to $[-\omega_D \ \omega_D]$, where ω_D is the maximal Doppler frequency, as most of the energy appear in that band. The choice of frequencies, forming an orthogonal basis, corresponds to those of the discrete Fourier transform (DFT). The whole procedure can be interpreted as taking the DFT of a window of data and then phase advancing the complex exponentials in the spectrum for a subset of frequencies (see Appendix 4.A.1).

This method uses least square estimates of the coefficients for the basis functions. The estimated coefficients are optimal in the MSE sense in that they give the best *model* for the data in the memory. However, what we really want are *predictors* that minimize the prediction error and not the model error so this is a sub-optimal approach.

Example 4.2

Mobile radio data
Figure 4.3 shows the Doppler spectrum for the largest tap $h_k(t)$, in one measurement. Each peak in the spectrum corresponds to a complex exponential at that frequency. We can construct a simplified model of the time series $h_k(t)$ as the sum of the contribution of only the largest peaks in the spectrum. The time series $h_k(t)$ of 1430 samples, shown in Figure 4.4, can be reconstructed from the DFT.

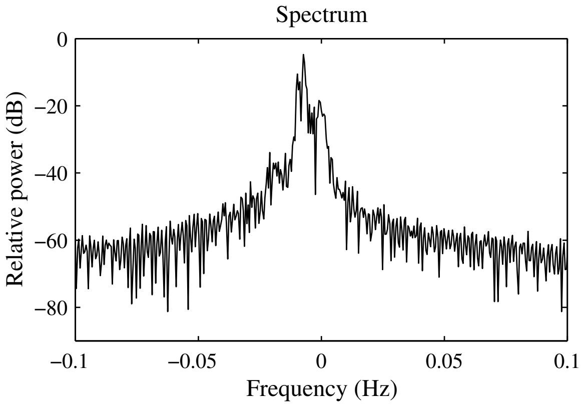

Figure 4.3: The power spectral density of the tap depicted in Figure 4.4. The sampling frequency is normalized to 1 Hz. Here 0.1 Hz (normalized frequency) corresponds to the Doppler frequency 457 Hz.

If using only the frequencies corresponding to the 20 largest values in the spectrum, then the normalized reconstruction error is as small as -19 dB. (Using 25 frequencies only gives a minor improvement of 2 dB.) Thus we can expect a rather good performance of linear predictors, as they rely on the property that a few complex exponentials can describe the tap.

The 20 frequencies are used in a predictive filter design (see Section 4.A.1 in Appendix 4.A) with a single sided Hamming window for the data. Two filter lengths are considered, $N = 1024$ and $N = 512$, both with $L = 9$ samples (i.e., 1 ms). The performance of these predictors are compared to those designed using orthogonal basis functions based on 20 frequencies as close as possible to the estimated frequencies and a short ($N = 137$) filter designed using least squares directly on the predictor co-

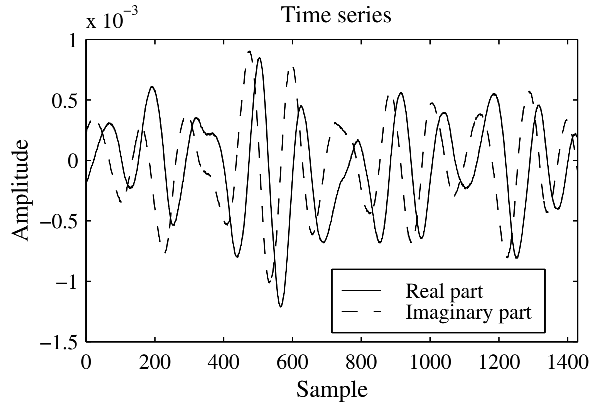

Figure 4.4: The real(solid line) and imaginary (dashed line) part of the tap in Example 4.2. The channel sampling period is 109 μs.

efficients to minimize the prediction error over the whole data set. The results are shown in Table 4.1. Of the pre-designed filters, only the long

Design method	**Filter length**	**NG [dB]**	**PG [dB]**
Nonorthogonal basis	512	12.8	-10
Nonorthogonal basis	1024	4.2	3.1
Orthogonal basis	512	-15	3.8
Orthogonal basis	1024	-15	10.7
LS FIR predictor	137	3.9	28.8

Table 4.1: Performance of different linear predictors on one tap. Noise gain (NG) and prediction gain (PG) for different designs and filter lengths. Prediction 9 samples (1 ms) ahead.

($N = 1024$) using orthogonal basis functions gives a reasonable prediction performance, still far below that of the short filter, designed using least squares directly on the predictor coefficients.

The poor performance of even the best linear predictor based on basis functions is due to minimization of the wrong criterion. In the basis function method the modeling error is minimized as opposed to the LS FIR predictor which minimizes the prediction error.

4.3 Truncated Volterra Series

If we assume that there are significant contributions to the channel from close point scatterers, the performance of the linear prediction approach will be inadequate, cf. Section 2.3. This is a motivation for investigating an extension of the linear predictor. Assuming that the nonlinear function governing the time evolution of the channel is smooth, then a truncated Taylor series expansion of the function might be sufficient to achieve the desired prediction gain. This Taylor expansion can be obtained by using Volterra series [35]. The quadratic Volterra system is described by

$$y(t+L) = a + \sum_{n=1}^{p} b_n y(t-\tau_n) + \sum_{k=1}^{p}\sum_{l=k}^{p} c_{kl} y(t-\tau_k) y(t-\tau_l). \tag{4.34}$$

This structure is from here on referred to as a quadratic filter.

The most general quadratic filter for complex signals includes not only the quadratic terms of (4.34) but also all possible combinations of conjugate products. However in the examples studied we have found no need for the conjugate products.

To include quadratic terms increase the complexity of the predictor. A linear predictor with the same subsampling as the quadratic, has p coefficients, whereas the quadratic predictor has $1 + \frac{3}{2}p + \frac{1}{2}p^2$ coefficients that have to be estimated.

For both the linear and the quadratic predictor we use the LS method to estimate the parameters of the predictors. To avoid numerical problems, the data is normalized so that linear and quadratic terms are of the same order of magnitude. This is done by dividing the signal by its standard deviation. (It is always preferable to work on normalized signals in applications where calculations are made with a limited number of bits.) Still, the covariance matrix for the regressors may be ill-conditioned. This is a common problem with quadratic filters, resulting in low numerical accuracy and poor generalization capability for the predictor. To avoid this we regularize the solution [21] by adding small positive constants to the diagonal of the estimated covariance matrix (diagonal loading) prior to the LS estimation.

The LS solution for quadratic filters relies on accurate estimates of fourth order moments. This can limit the use of quadratic filters in a noisy time-varying environment such as a mobile radio channel.

Example 4.3

Mobile radio data
Here we use a quadratic filter for prediction of one complex tap taken from a measured channel with a bandwidth of 6.4 MHz. Let the memory (T) be 25 ms (229 samples) and the prediction range be 10 ms (91 samples). Six delayed samples evenly spaced in the memory are used to form the regression variables, that is $p = 6$. Only the quadratic terms (not the conjugated) and the linear terms are used. A constant 1 is also appended to the regression variables to estimate the coefficient a in equation (4.34). A total of 28 complex coefficients are used in the Volterra filter. All the

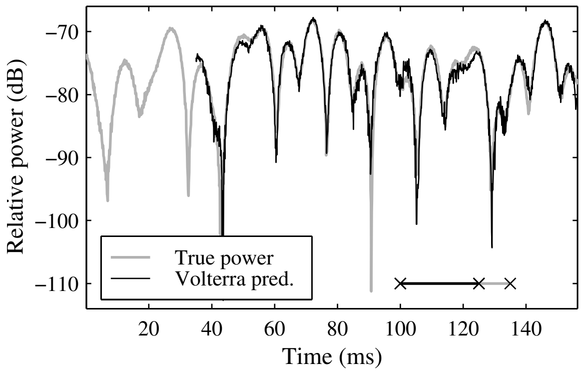

Figure 4.5: True and predicted squared magnitude (prediction gain is 12.8 dB) for the largest tap in one measurement at 5 MHz bandwidth. The predicted power is calculated from the complex tap, predicted 10 ms (91 samples) ahead with a Volterra predictor. The memory (x-x solid line) and prediction interval (x-x gray line) are indicated in the lower right corner. The prediction is only evaluated on the training data.

available 1430 data samples are used for estimating the coefficients for the predictor. The modeling performance is then evaluated on the training data.

The method performs worse when the memory is reduced from 25 to 15 ms (229 to 137 samples). An LS optimized linear FIR-filter, designed using all the samples in the memory, which requires 229 complex parameters, achieves a prediction gain of only 1.3 dB for the squared magnitude of the tap. On the other hand the quadratic predictor achieves a prediction gain of 12.8 dB. In both cases the prediction of the squared magnitude of the tap was calculated by first predicting the complex tap and then taking the square of the absolute value of the prediction value.

Based on the modeling performance, exemplified above, the quadratic predictor is a candidate for further investigation and comparative evaluation in Section 4.5. Even though the quadratic predictor shows good modeling properties, the generalization properties are poor as will be seen in Section 4.5.2.

4.4 MARS Modeling

Multivariate Adaptive Regression Splines (MARS) have been proposed by Friedman [10] to build models of the relationships between a scalar response variable and multiple regression variables. Lewis and Stevens [36] 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, $\mathbf{y}(t) = [y(t-\tau_1) \ldots y(t-\tau_p)]^T$. 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 [37], 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) but only the 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

$$y(t+L) = f(\mathbf{y}(t)) + \epsilon(t). \tag{4.35}$$

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

$$\hat{f}(\mathbf{y}(t)) = a_0 + \sum_{m=1}^{M} a_m B_m(\mathbf{y}(t)) \tag{4.36}$$

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 the basis function of the parent region $B_{P(m)}(\mathbf{y}(t))$ and a *truncated linear spline* *function* $T_m(\mathbf{y}(t))$,

$$B_m(\mathbf{y}(t)) = B_{P(m)}(\mathbf{y}(t)) \times T_m(\mathbf{y}(t)). \tag{4.37}$$

The *level of interaction* K_m of region m describes how many truncated splines are used to build B_m. A child has one level of interaction higher than its parent, that is $K_m = K_{P(m)} + 1$. Each truncated linear spline function T_m works along one dimension only (that is one delay τ_v) and has a partitioning point in the parent region at $y(t - \tau_v) = p_{P(m),l}$, partitioning the regression variable space as

$$T_m(\mathbf{y}(t)) = [s_m(y(t - \tau_v) - p_{P(m),l})]_+, \tag{4.38}$$

where $[\cdot]_+$ denotes the half-wave rectifier function, i.e. it takes the value of the argument if it is positive and is zero otherwise. In (4.38), $s_m = (-1)^m$ gives the left or right side of the threshold $p_{P(m),l}$. The index l on the threshold denotes which partition of parent region $P(m)$ that result in the child m. For $s_m = +1$, $T_m(\mathbf{y}(t))$ is positive for $y(t - \tau_v) - p_{P(m),l} > 0$ and zero otherwise. For $s_m = -1$ the inequality is turned the other way.

The MARS algorithm builds up the model structure iteratively (by enlarging the set of subregions while proceeding to higher levels of interaction among regression variables) and adjusts the thresholds ($p_{P(m),l}$) and weighting coefficients (a_m) to fit the data, while minimizing the mean square error. How a MARS model structure might look is illustrated in Figure 4.6.

Example 4.4

Basis function in a MARS model.
To illustrate how basis functions in MARS can look like we evaluate the basis function $B_{11}(\mathbf{x})$ in Figure 4.6. Region 11 is a child of region 8 and it corresponds to the left hand side of the first (and only) partition of region 8. Using (4.37) and (4.38) we can write the basis function as

$$B_{11}(\mathbf{x}) = B_8(\mathbf{x}) \times [p_{8,1} - x_3]_+. \tag{4.39}$$

If we expand $B_8(\mathbf{x})$ and evaluate the half-wave rectifier we then can write the basis functions as

$$\begin{aligned} B_{11}(\mathbf{x}) &= [x_1 - p_{0,3}]_+ \times [p_{8,1} - x_3]_+ \\ &= \begin{cases} p_{8,1}x_1 + p_{0,3}x_3 - x_1x_3 - p_{0,3}p_{8,1} \ , & x_1 > p_{0,3}, x_3 < p_{8,1} \\ 0 \ , & \text{otherwise.} \end{cases} \end{aligned}$$

The basis function consist of linear terms and a product between two variables, thus the level of interaction is equal to two. Only for $B_9(\mathbf{x})$ and $B_{10}(\mathbf{x})$ the level of interaction is three. We can from the partitioning tree see that x_3 is of less importance to the model than x_1 and x_2, as only the last partition uses x_3.

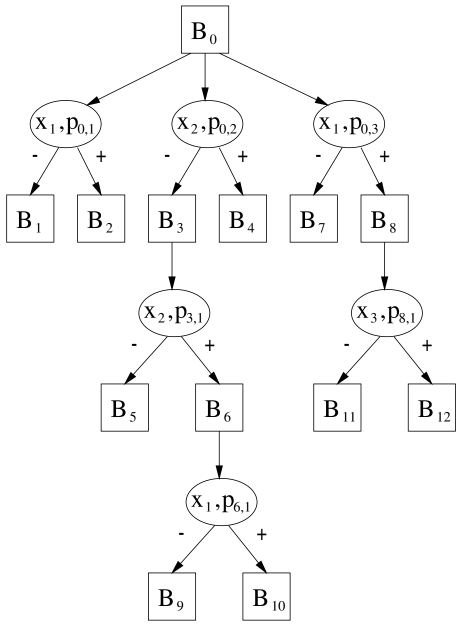

Figure 4.6: MARS model structure for three dimensional input $\mathbf{x} = (x_1, x_2, x_3)$*, where* $x_n(t) = y(t - \tau_n)$*, defined over the domain* B_0*. The level of interaction is* $K = 3$ *and the number of basis functions is* $M = 12$*. The model has six thresholds,* $p_{P(m),l}$*.*

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, such as the standard multi-layer perceptron trained with back-propagation, 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 neural net) and they are good at approximating (locally) linear mappings.

Furthermore their digital implementation are very simple and computationally efficient (only hard thresholds, multiplications and additions, 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 requires at most $M(K+1)+1$ additions, $M(K+1)$ multiplications, MK threshold operations, and uses no more than $M(K+1)+1$ parameters.

Example 4.5

Prediction of a mobile radio channel tap using a MARS model. For MARS modeling of the complex taps, two models have to be built; one for the real and one for the imaginary part. 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. The

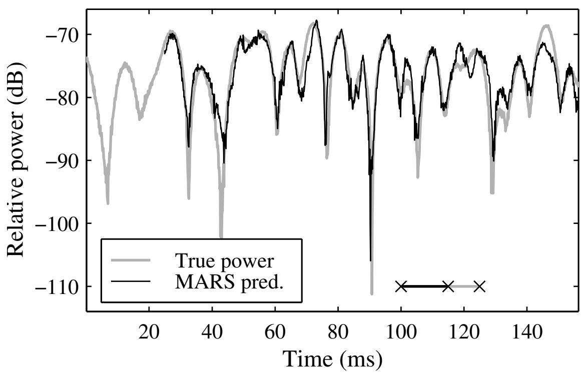

Figure 4.7: True and predicted squared magnitude (prediction gain is 5.8 dB) for the largest tap in one measurement at 5 MHz bandwidth. The predicted power is calculated from the complex tap, predicted 10 ms (91 samples) ahead with a MARS predictor. The memory (solid line x-x) and prediction interval (gray line x-x) are indicated in the lower right corner. The predictor is evaluated on the data it is trained on.

MARS prediction captures the fading pattern in Figure 4.7 even though the prediction interval is as long as 10 ms, which is more than the typical duration of a fade. The same data as in Figure 4.5 is used and as there, the modeling performance is evaluated on the training data.

The MARS model offers a very flexible structure for prediction. However, on mobile radio data, as in the example above, over-fitting often occurs. This leads to good modeling with few parameters but the validation fails.

4.5 Prediction of a Channel Tap

Each tap of the sampled channel impulse response can approximately be described by the model (2.32) as a sum of complex sinusoids with time varying frequency and amplitude. The prediction of the value of individual complex taps can be of interest in some equalizers, but often only the power, i.e., the squared absolute value, is of interest. As the complex tap contains more information than the power signal, it will be more suitable for prediction. In the subsequent subsections the presented predictors are tested on individual complex taps. First a small simulation study is conducted to illustrate some inherent problems of prediction. Then a combined modeling and validation study on measured channel impulse responses demonstrates the poor generalization properties of the quadratic predictors. The generalization properties for the MARS predictor, not shown here, are equally poor.

4.5.1 Simulation

For the simple simulation scenario in Section 2.4, with a mobile station driving past three close scatterers at 60 km/h, the performance of a quadratic and two linear predictors are evaluated. The quadratic predictors use different numbers of delayed samples with delays evenly spaced over the interval $[-200\ 0]$ as seen in Figure 4.8. This corresponds to a memory length of 22 ms, as the channel sampling period is 109 μs. With a sub-sampling factor of 25 the used delays $[\tau_1, \ldots, \tau_p]$ are $[0, 25, 50, 75, 100, 125, 150, 175, 200]$. The prediction range is $L = 50$ samples (5.5 ms). The channel is normalized to have a variance equal to one and a very small noise term is added, resulting in an SNR of 120 dB. The first 2000 samples (219 ms) of the data record are used to build the models and the following 500 samples (55 ms) are used for validation.

To improve the numerical accuracy and generalization capability of the predictors, we use diagonal loading of the covariance matrix for the regressors. In the quadratic predictor the coefficients corresponding to the quadratic terms need more regularization than those corresponding to the linear terms. This can be achieved by using different loading in different parts of the estimated covariance matrix. Here we add 10^{-5} to the diagonal

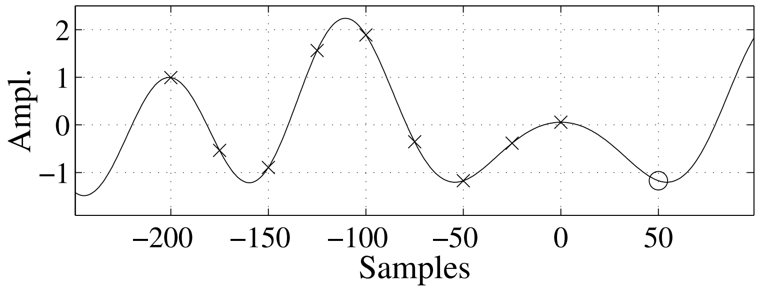

Figure 4.8: Short segment of a simulated tap time series. The 'x's denotes the samples used by the predictor and the 'o' the predicted sample. Here the samples used for the predictor are spaced 25 samples apart, thus a sub-sampling factor of 25 is used.

elements corresponding to quadratic terms and 10^{-9} for the linear. The same diagonal loading as for the linear terms in the quadratic predictor is used when estimating the coefficients for the linear predictors.

The different predictor structures and their performance are summarized in Table 4.2.

Predictor	Subsampling	# param.	PG mod. [dB]	PG val. [dB]
Quad.	25	55	36	6
Linear	25	9	22	1
Linear	5	41	38	12

Table 4.2: Table of the results for the different predictors on the simulated data with an SNR of 120 dB. The prediction range is 50 samples.

To use the quadratic predictor with a sub-sampling factor of 5 is not feasible as it would result in a predictor with over 900 coefficients. Our experience is that the quadratic predictor shows a high sensitivity to noise and to the choice of the diagonal loading terms. When the noise level increases the quadratic predictor loses its generalization capability and fails in the validation.

The large difference between the prediction gain for the model and the validation set for the linear predictors is due to the nonlinearity caused by the quadratic phase terms, cf. (2.32, 2.35). As can be seen in Figure 2.7, both the amplitudes and the instantaneous Doppler frequencies of the three contributing paths change significantly from the training set (the first 0.22 s) to the validation set (from 0.22 s to 0.27 s). The linear predictors perform best when the Doppler frequencies are constant. Neither the linear nor the quadratic predictor seems to be able to model this nonlinearity. Still, the linear predictor with the lower sub-sampling seems more robust than the quadratic predictor.

4.5.2 Prediction of Estimated Channels

In the following the result of prediction of the channel coefficient-estimates from the data base of measured broadband data at 1880 MHz, with 6.4 MHz bandwidth, is presented for different prediction ranges. Noise reduction, as described in Section 3.3, is applied to the estimated impulse responses prior to prediction. In the noise reduction the model for the dynamics is a 4:th order Butterworth filter, with a cut-off frequency corresponding to a speed of 110km/h, to assure that pass-band is wide enough to encompass the whole interesting range of Doppler frequencies.

For the measured impulse responses, only those taps with SNR above 10 dB are used to evaluate the performance of the predictors. As the noise level is reduced using a smoothing filter we can assume, relying on the simulation results in Section 3.3, that the effective SNR is at least 20 dB. The same predictors as in Section 4.5.1 were used on the measurements, except that no diagonal loading was used when estimating the linear terms of the quadratic and linear predictors. The memory thus consist of 201 samples.

In prediction the memory should contain at least one full cycle of the dominant oscillation to obtain the best prediction performance [19]. With the chosen memory length the traveled distance, during the length of a window, should be at least one wavelength. Thus the length corresponds to at least one oscillation period at the maximum Doppler frequency. (For a Rayleigh fading channel the maximum Doppler frequency is the frequency of the dominant oscillation.)

We vary the prediction range L from 25 to 100 samples, that is 2.7-11 ms. Of the 1430 samples available for each tap in a measured impulse response, 10 samples are lost due to the smoothing (5 due to the smoothing lag and another 5 due to the initialization of the smoothing-filter). Of the remaining 1420 samples the first $1200 + L$ are used for modeling and the remaining $220 - L$ samples constitute the validation set. The need for a long training set result in a very short validation set, but the amount of validation data is sufficient to show the poor generalization performance of the quadratic predictor.

A total of 532 taps, from 25 estimated channel impulse responses, are used in the study. The achieved prediction gain is averaged (in dB) for all the used taps. The prediction gain for the model and validation data is shown in Figure 4.9. Even though the quadratic predictor gives a much higher

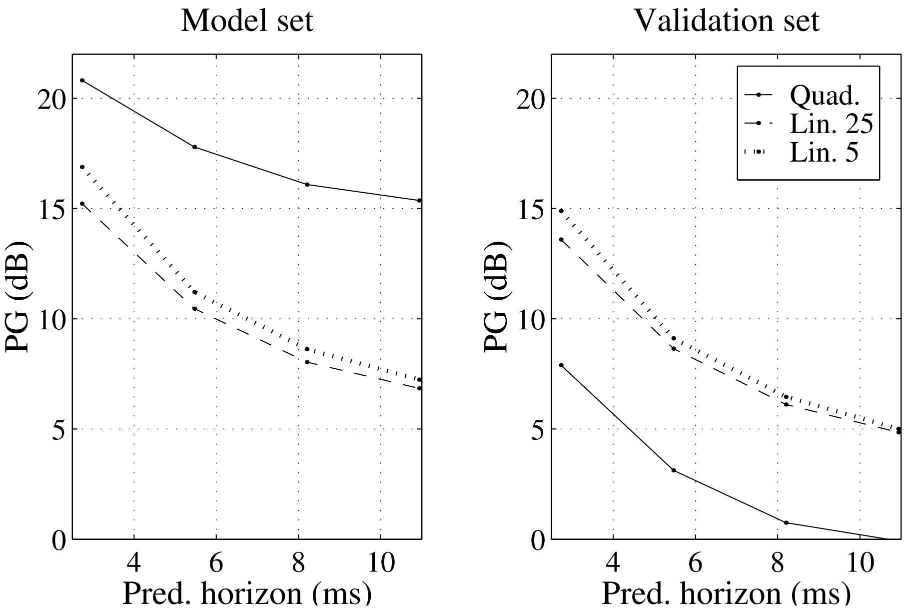

Figure 4.9: Average prediction gain for different prediction ranges for both model and validation set with sub-sampling factor 25 and 5 respectively. The quadratic predictor has sub-sampling factor 25.

prediction gain on the modeling set it fails to generalize to the validation set, whereas the linear predictors loose only 2 dB.

There is no major difference in performance between using 41 or 9 parameters (sub-sampling factor 5 instead of 25) for the linear prediction. Neither does an increase of the memory to 300 samples (33 ms), using subsample factors 5 and 25, result in a significant improvement of the performance of the linear predictors on these data sets.

Noise reduction

The noise reduction reduces the bandwidth of a single tap and it might clean the signal from features that actually are interesting. To evaluate the effect of noise reduction on prediction performance, we compare a predictor trained on data after noise reduction with a predictor trained on the noisy data set. Both predictors are evaluated on the original noisy data set. If the predictor performs better when optimized based on noise reduced data in the regressors as compared to using noisy data, then there is no structure in the removed "noise" that the predictor can use.

Example 4.6

Two taps from one of the measurements are here used to demonstrate the properties of the noise reduction combined with prediction. The two taps have an estimated SNR of 20 dB and 10 dB and the prediction gain for two prediction ranges, $L=50$ and $L=10$ samples (that is 5.5 and 1.1 ms), are evaluated.

The interesting property is the increase in prediction gain, ΔPG, that the use of noise reduced regressors give in prediction of the noisy taps. The noise is reduced in the same manner as for the validation study above. For the linear filters the noise gain (NG)[2], defined as the squared sum of the filter coefficients, shows how sensitive the predictor is to noise. The difference in noise gain between a linear predictor designed based on noise reduced data and the same predictor designed based on noisy data is ΔNG.

The results are presented in Table 4.3. The noise gain for a quadratic predictor is signal dependent and can thus not be calculated from the coefficients of the predictor. For the tap with an SNR of 20 dB the predic-

		Quad.	Lin. sub.25 [dB]		Lin. sub.5 [dB]	
SNR [dB]	L	ΔPG [dB]	ΔPG	ΔNG	ΔPG	ΔNG
20	50	3.8	2.0	4.6	1.7	9.9
	10	2.5	2.5	2.3	1.9	10.7
10	50	1.7	1.1	8.3	0.7	9.8
	10	1.4	1.5	5.2	0.9	8.8

Table 4.3: Table of the results for the prediction in Example 4.6. The gain, using noise reduced data to predict noisy data, as compared to using nosy data for the prediction, for all the predictors is tabulated. The noise gain is also included for the linear predictor. (As NG is not defined for the quadratic predictor, only ΔPG is displayed.)

tion gain increases the most, when using noise reduced regressors in the predictors as compared to using the noisy data. For the noisier tap this improvement is about half as large (in dB). The improvement is higher

[2]The noise gain is the increase of variance for a white noise signal, run through the predictor.

for the quadratic predictor than for the linear, which is natural as the quadratic terms result in high noise sensitivity.

The linear filter can use its coefficients to achieve noise reduction, which result in a low noise gain. If the noise is already reduced, then the linear filter can use it coefficients to achieve better prediction gain and the noise gain increases, as seen in Table 4.3. For the linear predictor using 41 coefficients (sub-sampling by a factor of five), the increase in noise gain is high but the improvement of prediction gain is smaller than for the linear predictor using 9 coefficients (sub-sampling by a factor of 25). With 41 coefficients the linear filter can both predict and reduce noise, whereas 9 coefficients are sufficient only for prediction of the linear dependencies in the tap when there are small amounts of noise.

As can bee seen from the example, the noise reduction actually increases the predictability of the noisy data. It is thus beneficial to first reduce the amount of noise and then apply the predictors.

4.5.3 Modeling

In the previous section it was demonstrated that the generalization properties for the studied quadratic predictors are poor. The same holds for the MARS predictor. In this section we address only the ability of the different predictor structures to mimic the dynamics of a given data set and not their feasibility as predictors. We obtain a measure for how well the channel coefficients could have been predicted, using the best possible setting of the parameters in the predictors.

A predictor using a small number of parameters and still achieving a low modeling error (i.e., good prediction gain), has the potential to efficiently model the dynamics of the system. A structure that has poor modeling capabilities is also a poor predictor.

The bandwidth of the channels is syntheticly reduced, as described in Chapter 3.A, to 5 MHz. No noise reduction is applied. Instead only the nine largest taps are used to obtain a high SNR. The predictors are then trained and evaluated on each tap. A total of 225 taps, from 25 different estimated channel impulse responses, are evaluated. The prediction range varies from L =1 ms (9 samples) up to L =10 ms (91 samples), to cover the range from short to long prediction ranges. The prediction is made on the complex taps and the prediction gain (in dB) is averaged over the 225 taps. The average performance of the predictors are displayed in Figure 4.10. All the predictors use a memory of 15 ms (137 samples).

However, the results are deceptive as they do not describe how good the predictors would work on successive validation data. We know from Section 4.5.2 that the validation properties are poor for the nonlinear predictors.

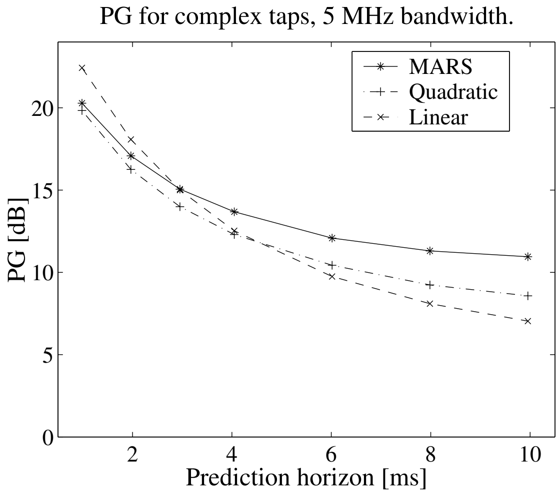

Figure 4.10: Average PG for the different predictors proposed in the previous section. All the predictors use the same memory length (137 samples or 15 ms).

The MARS and the quadratic predictor use only six delayed samples in the memory whereas the linear predictor uses all the 137 samples. The MARS model has the level of interaction equal to 2 and uses no more than 20 basis functions. As the MARS model used here only allows for quadratic interaction it can be interpreted as a gain-scheduled sparse quadratic predictor[3] (see Section 4.4).

Dependence on the prediction range

For short prediction ranges the linear predictor performs better than all the other predictors. As the linear predictor is not sub-sampled, as the other predictors, it is more suitable to model local smoothness as the parameters vary slowly. On a longer time scale the slow dynamics of the system becomes the more important factor. The nonlinear predictors *model* this dynamics better than the linear predictor and thus perform better on the training set.

[3]With sparse we mean that not all quadratic interactions are used, just an efficient subset.

The effect of noise
The estimated channel impulse responses are treated as if they were noise free and the only cause to change of a tap is due to the interference pattern. Still, there is noise in the measurements and a tap with a low power level is more affected by this noise than one with a high level. The noise on the estimated channel coefficients is propagated from the received data to the coefficients by the identification procedure as described in Section 3.2.3. A high noise level degrades the performance of the predictors, especially the ones that work only on a few delayed samples, as the averaging effect is then reduced.

The quadratic predictor shows only a minor improvement in modeling performance as compared to the linear on these data. This was not the case for the models in Section 4.5.2, where the difference is large. This is due to the high sensitivity to noise of the quadratic predictor, as no noise reduction was performed on the data here.

4.6 Concluding Remarks

There are two main causes for failure of a predictor.

1. Noise.

2. The predictor is unable to model the dynamics of the process.

The first cause can be addressed by applying noise reduction to the data and to choose a predictor that is robust against noise. To avoid the second problem we need to use a structure that either is tailored for the problem or is general enough to model the dynamics.

A linear filter can be given a design that is robust against noise and it has a close connection to the linearized model of the channel dynamics (2.47). The studied nonlinear predictors, on the other hand, are sensitive to noise and have more general structures. Even though the nonlinear predictors, with few parameters, give good models for a given data set the validation demonstrates that the studied nonlinear structures are of limited use for prediction of mobile radio channels. The LS solution for quadratic filters relies on accurate estimates of fourth order moments. This limits the use of quadratic filters in a noisy time-varying environment such as a mobile radio channel. Longer blocks, for identification of the prediction parameters, generally improve the performance of the predictors but can not be obtained as the environment does not stay stationary over very long periods of time.

A better approach than to use blocks of data to estimate the coefficients for the predictor, would be to introduce *adaptive predictors*, as in Chapter 5 and 6. The able performance of the linear block predictors points to the use of linear adaptive predictors to obtain reasonable prediction performance at low computational complexity.

The rapid fall off of predictability for the complex taps using the linear filters that can be seen in Figure 4.9 and 4.10, losing 5 dB when going from prediction of 25 samples ahead (2.7 ms) to 50 samples (5.5 ms), indicates that the performance of block-adaptive linear predictors is limited for long-range prediction. However, the small loss in performance when going from the model set to the validation set, as seen in Figure 4.9, indicates good generalization properties for the linear predictors. As the increase of the number of linear parameters hardly improves the performance, we can conclude that even the simple linear predictor exploits most of the available linear dependencies for the complex taps.

4.A Linear Prediction of Signals with Known Basis Functions

When a signal is generated by a known function of time and corrupted by additive noise, is it possible to estimate the parameters of the function and use the function estimate for prediction. We will in the following show that for certain choices of basis functions, as exponentials and polynomials, this can be interpreted as linear prediction using a time invariant linear filter.

Assume that the signal to be predicted can be modeled as a sum of weighted basis functions:

$$y(t) = \sum_{k=1}^{N} \alpha_k f_k(t) + n(t), \tag{4.40}$$

where $f_k(t)$ denote the functions and α_k the parameters. This can be reformulated as a vector multiplication

$$y(t) = [f_1(t) \dots f_N(t)] \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix} + n(t) = \mathbf{f}(t)\boldsymbol{\alpha} + n(t). \tag{4.41}$$

If the parameters $\boldsymbol{\alpha}$ are slowly time varying it can be advantageous to do the identification of the parameters $\boldsymbol{\alpha}$ block-wise within a sliding window of length p. To obtain a unique solution for the parameters the length of the window has to be larger than than the number of basis functions ($p \geq N$). The model (4.40) can, in a matrix formulation, be expressed as

$$\underbrace{\begin{bmatrix} y(t) \\ y(t-1) \\ \vdots \\ y(t-p+1) \end{bmatrix}}_{\mathbf{y}(t)} = \underbrace{\begin{bmatrix} \mathbf{f}(t) \\ \mathbf{f}(t-1) \\ \vdots \\ \mathbf{f}(t-p+1) \end{bmatrix}}_{\mathbf{F}(t)} \boldsymbol{\alpha} + \underbrace{\begin{bmatrix} n(t) \\ n(t-1) \\ \vdots \\ n(t-p+1) \end{bmatrix}}_{\mathbf{n}(t)}. \tag{4.42}$$

A least squares (LS) estimate of $\boldsymbol{\alpha}$ using the p latest samples, stacked in $\mathbf{y}(t)$, is readily obtained as

$$\hat{\boldsymbol{\alpha}} = \mathbf{F}^{\dagger}(t)\mathbf{y}(t). \tag{4.43}$$

A predictor for $y(t+L)$ using the estimated $\hat{\boldsymbol{\alpha}}$ in equation (4.41) with the basis functions evaluated at time $t+L$, would then be

$$\hat{y}(t+L) = \mathbf{f}(t+L)\hat{\boldsymbol{\alpha}} = \mathbf{f}(t+L)\mathbf{F}^{\dagger}(t)\mathbf{y}(t). \tag{4.44}$$

The noise $n(t+L)$ is uncorrelated with $y(t)$ only if $n(t)$ is an MA-process of degree lower than L. Here the unpredictable measurement noise $n(t+L)$ is disregarded.

If the time shifted basis functions $\mathbf{f}(t+L)$ can be described by a linear combination of the basis functions at time L weighted by time dependent coefficients, then the time dependence and the shift dependence can be separated. The LS fit of basis functions to data and extrapolation into the future can then be expressed as filtering with a time invariant linear FIR filter.

Theorem 4.2 *If the basis functions can be decomposed as*

$$\mathbf{f}(t+L) = \mathbf{f}(L)\mathbf{B}(t), \tag{4.45}$$

where $\mathbf{B}(t)$ is an invertible quadratic matrix, then the predictor (4.44) can be expressed as the filter operation

$$\hat{y}(t+L) = \mathbf{f}(t+L)\mathbf{F}^{\dagger}(t)\mathbf{y}(t) = \mathbf{c}\mathbf{y}(t), \tag{4.46}$$

where the time invariant FIR-filter $\mathbf{c}$ is

$$\mathbf{c} = \mathbf{f}(L)\mathbf{F}^{\dagger}(0). \tag{4.47}$$

Proof 4.2 *With the condition (4.45) fulfilled the matrix of basis functions can be decomposed as*

$$\mathbf{F}(t+L) = \mathbf{F}(L)\mathbf{B}(t), \tag{4.48}$$

and the relation (4.42), with $L = 0$, can be written as

$$\mathbf{y}(t) = \mathbf{F}(0)\mathbf{B}(t)\boldsymbol{\alpha} + \mathbf{n}(t). \tag{4.49}$$

The LS estimate (4.43) of $\boldsymbol{\alpha}$ using the p latest samples is then

$$\begin{aligned} \hat{\boldsymbol{\alpha}} &= (\mathbf{F}(0)\mathbf{B}(t))^{\dagger}\mathbf{y}(t) \\ &= (\mathbf{B}^H(t)\mathbf{F}^H(0)\mathbf{F}(0)\mathbf{B}(t))^{-1}\mathbf{B}^H(t)\mathbf{F}^H(0)\mathbf{y}(t) \\ &= \mathbf{B}^{-1}(t)\mathbf{F}^{\dagger}(0)\mathbf{y}(t). \end{aligned} \tag{4.50}$$

Using (4.45) and (4.50) in (4.44) the L step ahead predictor can be written as

$$\hat{y}(t+L) = \mathbf{f}(t+L)\hat{\boldsymbol{\alpha}} = \mathbf{f}(L)\mathbf{B}(t)\mathbf{B}^{-1}(t)\mathbf{F}^{\dagger}(0)\mathbf{y}(t) \tag{4.51}$$

where the time dependent part cancels out so we obtain the time invariant expression

$$\hat{y}(t+L) = \mathbf{f}(L)\mathbf{F}^{\dagger}(0)\mathbf{y}(t) = \mathbf{c}\mathbf{y}(t), \tag{4.52}$$

where $\mathbf{c}$ *is a time invariant linear FIR-filter whose design equation is given by*

$$\mathbf{c} \quad = \quad \mathbf{f}(L)\mathbf{F}^{\dagger}(0).$$

The filter $\mathbf{c}$ is a time invariant linear FIR filter consisting of p taps, designed to predict $\hat{y}(t+L)$ assuming the model (4.40). The FIR filtering prediction, using the described filter, can be interpreted as first making an LS estimate of the weights of the basis functions ($\mathbf{B}^{-1}(t)\mathbf{F}^{\dagger}(0)\mathbf{y}(t)$) and then make a deterministic step in time using these parameters with the basis functions evaluated at time $t+L$ (multiplication with $\mathbf{f}(L)\mathbf{B}(t)$). The time series is not limited to be stationary. The optimization criterion used in the design is not to minimize the prediction error but to minimize the model error for the basis function on a window of data. As the method does not take the color of the noise into consideration, as well as it optimizes the wrong property, it can not be the best linear predictor.

Weighting

Weighting of the N data used by the filter can be performed through a positive definite weighting matrix $\mathbf{Q}$ [20] that enters the filter expression in equation (4.47) as:

$$\mathbf{c} = \mathbf{f}(L)(\mathbf{F}^{H}\mathbf{Q}\mathbf{F})^{-1}\mathbf{F}^{H}\mathbf{Q}. \tag{4.53}$$

The time index 0 for $\mathbf{F}(0)$ is dropped for notational convenience. In the simplest case $\mathbf{Q}$ is just a diagonal matrix with a window function on the diagonal.

Orthogonal basis

In some cases the basis is numerically unsuitable for the filter equation (4.47). It can then be advantageous to design an orthonormal basis. Let the inner product between two basis functions be defined as:

$$< f_k, f_l >= \sum_{j=0}^{p-1} f_k^{H}(-j) f_l(-j) = \mathbf{F}_k^{H}\mathbf{F}_l, \tag{4.54}$$

where $\mathbf{F}_k$ denotes the k:th column of $\mathbf{F}$. With orthonormal basis functions $< f_k, f_l >= \delta_{kl}$ the columns of $\mathbf{F}$ will be orthonormal. Then $\mathbf{F}^{H}\mathbf{F} = \mathbf{I}$ where

$\mathbf{I}$ is the identity matrix. To obtain an orthonormal base, the Gram-Schmidt orthogonalization procedure can be used on the original basis. With an orthonormal basis the filter equation (4.47) reduces to:

$$\mathbf{c} = \mathbf{f}(L)\mathbf{F}^H. \tag{4.55}$$

The resulting filter remains the same but the numerical problems that can occur in the inversion of $\mathbf{F}^H\mathbf{F}$ are avoided.

Noise gain

The sensitivity of the filter to broadband noise is given by the noise gain (NG). For an FIR filter $C(q^{-1}) = \sum_{k=0}^{N-1} c_k q^{-k}$, the noise gain is the sum of the squared amplitude of the filter coefficients:

$$NG = \sum_{l=1}^{N} |c_l|^2 = \mathbf{c}\mathbf{c}^H. \tag{4.56}$$

For the FIR filter in equation (4.47) the noise gain is

$$NG = \mathbf{f}(L)\mathbf{F}^\dagger\mathbf{F}^{\dagger H}\mathbf{f}^H(L) = \mathbf{f}(L)(\mathbf{F}^H\mathbf{F})^{-1}\mathbf{f}^H(L). \tag{4.57}$$

For non-orthogonal basis functions the matrix $\mathbf{F}^H\mathbf{F}$ might be poorly conditioned, resulting in a high noise gain.

4.A.1 Exponential Signal

Assume that the signal to be predicted can be modeled as a sum of weighted exponentials.

$$y(t) = \sum_{k=1}^{N} \alpha_k \rho^{(\sigma_k + i\omega_k)t} + n(t) \tag{4.58}$$

The matrix representation in equation (4.42) becomes

$$\underbrace{\begin{bmatrix} y(t) \\ y(t-1) \\ \vdots \\ y(t-p+1) \end{bmatrix}}_{\mathbf{y}(t)} = \underbrace{\begin{bmatrix} 1 & \cdots & 1 \\ \rho^{-(\sigma_1+i\omega_1)} & & \rho^{-(\sigma_N+i\omega_N)} \\ \vdots & & \vdots \\ \rho^{-(\sigma_1+i\omega_1)(p-1)} & \cdots & \rho^{-(\sigma_N+i\omega_N)(p-1)} \end{bmatrix}}_{\mathbf{F}} \times$$

$$\underbrace{\begin{bmatrix} \rho^{(\sigma_1+i\omega_1)t} & & 0 \\ & \ddots & \\ 0 & & \rho^{(\sigma_N+i\omega_N)t} \end{bmatrix}}_{\mathbf{B}(t)} \boldsymbol{\alpha} + \mathbf{n}(t) \tag{4.59}$$

In this case the basis functions are

$$\mathbf{f}(t) = [\rho^{(\sigma_1+i\omega_1)t} \dots \rho^{(\sigma_N+i\omega_N)t}], \tag{4.60}$$

where the frequencies ω_k and exponents σ_k are assumed to be known. This basis fulfills the conditions (4.45) with $\mathbf{B}(t)$ being a diagonal matrix with $\rho^{(\sigma_k+i\omega_k)t}$ as the (k,k):th element. The linear filter $\mathbf{c}$ is as before given by (4.47) $\mathbf{c} = \mathbf{f}(L)\mathbf{F}^{\dagger}$. In the following we will assume that $\sigma_k = 0, \forall k$, which result in basis functions with oscillating output. Furtermore the basis ρ is taken to be the natural basis e. The filter (4.47) is then similar to that given in [33], where this filter is derived for the real valued case, using another method.

Connection with the DFT

If we constrain the frequencies to only take certain discrete values

$$\omega_n = \frac{2\pi k_n}{p},\ k_n \in [0 \dots p],\ k_n \neq k_m \forall\ n \neq m \tag{4.61}$$

and constrain us to the the case with $\sigma_k = 0, \forall k$, then the basis functions will be orthogonal (but not orthonormal). That means that the matrix $\mathbf{F}^H\mathbf{F} = p\,\mathbf{I}$ is diagonal and $\mathbf{F}^{\dagger} = (\mathbf{F}^H\mathbf{F})^{-1}\mathbf{F}^H = \mathbf{F}^H/p$. The equation (4.52) then reduces to

$$\hat{y}(t+L) = \mathbf{f}(L)\frac{\mathbf{F}^H}{p}\mathbf{y}(t) \tag{4.62}$$

Let us focus on row n of the latter part of the product:

$$\left[\frac{\mathbf{F}^H}{p}\mathbf{y}(t)\right]_n = \frac{1}{p}\sum_{k=0}^{p-1} y(t-k)e^{i\omega_n k}. \tag{4.63}$$

$$\left[\frac{\mathbf{F}^H}{p}\mathbf{y}(t)\right]_n = e^{i\omega_n(p-1)}\frac{1}{p}\sum_{l=0}^{p-1} y(t-p+1+l)e^{-i\omega_n l} = e^{i\omega_n(p-1)}\mathcal{Y}(\omega_n). \tag{4.64}$$

Here $\mathcal{Y}(\omega_n)$ is the p point discrete Fourier transform of $y(\tau)$ at frequency ω_n, where τ goes from $t-p+1$ to t. The discrete Fourier transform gives the amplitude and phase for a complex exponential with frequency ω_n at time $\tau = 0$, that is $t-p+1$. The term $e^{i\omega_n(p-1)}$ phase advances the complex exponential to time t. The corresponding component in $\mathbf{f}(L)$ gives a further phase shift to $t+L$. Thus the linear filtering procedure described in equation (4.62) is equivalent to taking a p samples long discrete Fourier transform of the data and then phase advance the chosen frequencies.

If all the discrete frequencies are used, that is, $k = 0, \ldots, p-1$, then the resulting filter just copies the sample $y(t-p+L)$ to obtain $\hat{y}(t+L)$. This is because the time series is treated as periodic with period p in the Fourier transform. To obtain reasonable results in the prediction, the filter has to be long to give high frequency resolution and if the signal has a low-pass structure then the higher frequency components must be dropped to avoid the copy effect.

Noise gain

For an FIR filter designed for complex exponentials, as described here, the noise gain in equation (4.57) is

$$NG = \mathbf{f}(L)(\mathbf{F}^H\mathbf{F})^{-1}\mathbf{f}^H(L) \approx \frac{2N}{p}, \tag{4.65}$$

where the approximation is valid for $p \gg 1$. For frequencies fulfilling the relationship in equation (4.61) the approximation is exact. The noise gain is thus independent of the prediction horizon L when the filter is long but grows with the numbers of estimated complex exponentials N in the signal.

4.A.2 Polynomial Signal

It is common to locally approximate a function with a polynomial,

$$y(t) = \alpha_1 + \alpha_2 t + \ldots + \alpha_N t^{N-1}. \tag{4.66}$$

The basis functions can then be selected as the powers of the time,

$$\mathbf{f}(t) = [1\ t\ t^2 \ldots t^{N-1}]. \tag{4.67}$$

This basis fulfills the condition (4.45):

$$\mathbf{f}(t+L) = \mathbf{f}(L) \underbrace{\begin{bmatrix} 1 & t & t^2 & t^3 & \ldots & t^{p-1} \\ 0 & 1 & 2t & 3t^2 & \ldots & (N-1)t^{N-1} \\ 0 & 0 & 1 & 3t & \cdot & \cdot \\ \cdot & & & 1 & \cdot & \cdot \\ \cdot & & & & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & 0 & 1 \end{bmatrix}}_{\mathbf{B}(t)} \tag{4.68}$$

The weights in the columns of $\mathbf{B}(t)$ are given by the numbers in Pascal's triangle.

$$\begin{array}{ccccccccc} & & & & 1 & & & & \\ & & & 1 & & 1 & & & \\ & & 1 & & 2 & & 1 & & \\ & 1 & & 3 & & 3 & & 1 & \\ 1 & & 4 & & 6 & & 4 & & 1 \end{array} \tag{4.69}$$

In [38] a similare derivation of the polynomial filter is used to derived a smoothing filter.

An evident problem with this straight forward polynomial basis is the poor condition number of $\mathbf{F}$ in the filter design equation. Having $p = 100$ and $N = 2$ (constant and linear basis function), the two-norm condition number for $\mathbf{F}$ is 114. For $N = 3, 4$ and 5 the condition number increases rapidly to $1.3 \cdot 10^4, 1.4 \cdot 10^6$ and $1.5 \cdot 10^8$. This numerical problem limits the degree of the polynomial that can be predicted and the length of the filter. The problem can be avoided by using orthonormal polynomial basis functions as in [38]. The poor generalization properties of higher order polynomials remain a problem in predictive filtering.

Smoother to detrend data

As polynomial filters of higher order have poor generalization properties, the corresponding filters $\mathbf{c}$ are preferably used as smoothers. This is accomplished by choosing L as a smoothing lag, i.e., a negative number in the filter design equation (4.47). The polynomial smoother can be used to design a detrending smoother $\mathbf{c}_D$, as

$$\mathbf{c}_D = [\ 0 \ldots 0\ 1\ 0 \ldots 0]^T - \mathbf{c}, \tag{4.70}$$

with the 1 in possition $1 - L$ where $L \leq 0$. This smoother locally takes away any polynomial trend of a given order.

4.A.3 Summary

Under the assumption that the data can be described by the model (4.40), where the parameters may be slowly time varying, the following procedure for prediction can be used.

- Take a data window of length p.
- For the data window, estimate the weights of the basis functions using the LS method.

- Use the estimated weights to calculate a prediction of the data at a time $t + L$ as in equation (4.71):

$$\hat{y}(t+L) = \sum_{k=1}^{N} \hat{\alpha}_k f_k(t+L). \tag{4.71}$$

If the basis fulfills the condition (4.45) then this procedure can be expressed as linear filtering of the data,

$$\hat{y}(t+L) = \sum_{k=0}^{p-1} c_k y(t-k) = C(q^{-1})y(t) \tag{4.72}$$

with a time invariant linear filter $C(q^{-1})$. How this filter can be designed directly from the basis functions is described in the previous sections.

Chapter 5

Adaptive Predictors

As the physical environment changes when moving, the parameters of the predictor have to change accordingly to maintain the prediction performance. An adaptive predictor constantly tunes its parameters to achieve the best prediction of the channel coefficients at the current conditions. It thus differs from the block predictor, where it is assumed that the parameters of the predictor are constant over the time span of the block and that estimated prediction parameters are valid also outside this block.

In a time-varying environment the regression model in equation (4.2) should be replaced by the time-varying regression

$$y(t+L) = f_L(t, \mathbf{y}(t)) + \epsilon(t). \tag{5.1}$$

As before the predictor estimates $y(t+L)$ as

$$\hat{y}(t+L) = \hat{f}_L(t, \mathbf{y}(t)), \tag{5.2}$$

where $\hat{f}_L(t, \cdot)$ is the time-varying L-step predictor function, which is an estimate of $f_L(t, \cdot)$ at time t. The adaptive predictor adjusts the parameters of $\hat{f}_L(t, \cdot)$ based on data up to time t and then applies the estimated predictor on new data to predict $\hat{y}(t+L)$. There is thus no need for a separate validation data set as was the case for the block predictors.

In this Chapter we will present an adaptive Kalman predictor based on a simple model for the coefficient variation. Both direct prediction and iterated prediction will be treated. In a direct approach the coefficients for an L-step predictor are adapted and used for prediction. In an indirect approach the coefficients for an m-step predictor, where $m = L/k$ and $k \in [2, 3, \ldots L]$, are adapted. The adapted m-step predictor is then iterated k times to obtain an L-step predictor. The iterated predictor is mainly of interest due to that only one predictor needs to be adapted to cover a large interval of prediction ranges.

5.1 State Space Model

5.1.1 Model for the change of predictor coefficients

The following linear regression will be used for prediction,

$$\hat{y}(t+L) = \hat{f}_L(t, \mathbf{y}(t)) = \boldsymbol{\varphi}(t)^H \mathbf{c}(t+L), \tag{5.3}$$

or

$$y(t) = \underbrace{\boldsymbol{\varphi}(t-L)^H \mathbf{c}(t)}_{\hat{y}(t)} + v(t), \tag{5.4}$$

where $v(t)$ is the estimation error, $\mathbf{c}(t)$ is a column vector containing M time-varying coefficients and where $\boldsymbol{\varphi}(t)$ is the regressor vector consisting of terms of the time series $y(t)$ up to time t. The elements of $\boldsymbol{\varphi}(t)$ may be nonlinear transformations of data. If we for example use the regressor vector

$$\boldsymbol{\varphi}(t) = [y(t)\ y(t-m) \ldots y(t-(M-1)m)]^H, \tag{5.5}$$

then (5.4) is a linear FIR-filter with sub-sampling factor m.

Even though the mobile radio channel shows dramatic changes over short traveled distances the parameterization of the amplitude and phase, see e.g. (2.48), is slowly time varying. We can thus assume that the optimal predictor coefficients change slowly and without abrupt changes. In the following, the M predictor coefficients in the vector $\mathbf{c}(t)$ are assumed to vary, without abrupt changes, on a slower time scale than the variations of the time series $y(t)$.

5.1.2 AR1I models

In the following subsection we shall design an adaptive Kalman filter that tracks the time varying coefficients of the predictor under the assumption that the prediction coefficients behave as *filtered random walks*. The filtered random walk model is used because it is the simplest model that describes smooth changes of the coefficients and that includes integration.[1] Let the predictor coefficients have increments $\Delta\mathbf{c}(t)$

$$\mathbf{c}(t+1) = \mathbf{c}(t) + \Delta\mathbf{c}(t). \tag{5.6}$$

[1]Integration in the model is desired because then the corresponding Kalman filter can estimate time-invariant parameters without bias.

As there should be no abrupt changes, we may model $\Delta\mathbf{c}(t)$ as a stochastic process with low-pass properties. The simplest model of this type is the first order autoregressive model

$$\Delta\mathbf{c}(t) = \frac{\sqrt{1-p^2}}{1-pq^{-1}}\mathbf{e}(t), \quad 0 \leq p < 1 \tag{5.7}$$

where $\mathbf{e}(t)$ is a vector of complex-valued white noises with zero mean and constant or slowly time varying covariance matrix $\mathbf{R}_e$. The factor $\sqrt{1-p^2}$ in the numerator is used to normalize the the variance of $\Delta\mathbf{c}(t+1)$, which becomes independent of p. The equation (5.7) corresponds to the vector difference equation

$$\Delta\mathbf{c}(t+1) = p\Delta\mathbf{c}(t) + \sqrt{1-p^2}\,\mathbf{e}(t+1). \tag{5.8}$$

The predictor coefficients can then be modeled as

$$\begin{aligned} c_i(t) &= \frac{\sqrt{1-p^2}}{(1-q^{-1})(1-pq^{-1})}e_i(t) & (5.9)\\ &= \frac{\sqrt{1-p^2}}{1-(1+p)q^{-1}+pq^{-2}}e_i(t), & (5.10)\end{aligned}$$

where the pole of the model filter $0 \leq p < 1$ is chosen to give good tracking abilities. If the predictor coefficients are expected to have noisy increments then p has to be small whereas if the coefficients vary smoothly then p is chosen close to one. Dependencies between the coefficients can be modeled by introducing non-zero off-diagonal elements in the covariance matrix of $\mathbf{e}(t)$. One can also use individual p values for each coefficient to better model differences between the coefficients.

It is possible to use higher order model filters, like an AR2I-model. In experiments on mobile radio channel data, the AR2I-model have given only minor improvements of the performance as compared to the proposed AR1I-model.

5.1.3 State Space Model

The predictor coefficients and their increments, modeled as the AR1I process (5.10), can conveniently be expressed in a state space form as

$$\begin{pmatrix} c_i(t+1) \\ \Delta c_i(t+1) \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 \\ 0 & p \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} c_i(t) \\ \Delta c_i(t) \end{pmatrix}}_{\mathbf{x}_i(t)} + \underbrace{\begin{pmatrix} 0 \\ \sqrt{1-p^2} \end{pmatrix}}_{\mathbf{B}} e_i(t+1).$$

$$c_i(t) = \underbrace{\begin{pmatrix} 1 & 0 \end{pmatrix}}_{\gamma} \mathbf{x}_i(t). \tag{5.11}$$

The state space models (5.11) for each component of $\mathbf{c}(t)$ are used as building blocks for the pertinent state space model representing the time varying prediction coefficients, with the regression model (5.4) as the measurement equation. Let the states $\mathbf{x}(t) = [\mathbf{x}_1^T(t) \ldots \mathbf{x}_M^T(t)]^T$ represent all predictor coefficients and their increments. The total model for all coefficients is then

$$\underbrace{\begin{pmatrix} \mathbf{x}_1(t+1) \\ \vdots \\ \mathbf{x}_M(t+1) \end{pmatrix}}_{\mathbf{x}(t+1)} = \underbrace{\begin{pmatrix} \mathbf{A} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{A} \end{pmatrix}}_{\mathbf{F}} \underbrace{\begin{pmatrix} \mathbf{x}_1(t) \\ \vdots \\ \mathbf{x}_M(t) \end{pmatrix}}_{\mathbf{x}(t)} + \underbrace{\begin{pmatrix} \mathbf{B} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{B} \end{pmatrix}}_{\mathbf{G}} \underbrace{\begin{pmatrix} e_1(t+1) \\ \vdots \\ e_M(t+1) \end{pmatrix}}_{\mathbf{e}(t+1)} \tag{5.12}$$

$$\mathbf{c}(t) = \underbrace{\begin{pmatrix} \gamma & & 0 \\ & \ddots & \\ 0 & & \gamma \end{pmatrix}}_{\mathbf{H}} \mathbf{x}(t) \tag{5.13}$$

or

$$\mathbf{x}(t+1) = \mathbf{F}\mathbf{x}(t) + \mathbf{G}\mathbf{e}(t+1) \tag{5.14}$$
$$\mathbf{c}(t) = \mathbf{H}\mathbf{x}(t) \tag{5.15}$$
$$y(t) = \boldsymbol{\varphi}(t-L)^H \mathbf{c}(t) + v(t). \tag{5.16}$$

The M elements of the regressor $\boldsymbol{\varphi}(t)$ consist of delayed samples of the time series $y(t)$ up to time t, either in direct or transformed form. The measurement noise term $v(t)$ represents the L-step prediction error. In the Kalman design, $v(t)$ is assumed to be zero mean white noise with variance σ_v^2. The time variability of the predictor coefficients is parameterized by **F**, **G** and **H**.

5.2 The Kalman Filter and Predictor

The best linear state estimator given the state space model (5.14-5.16) is the Kalman filter [39]. It is a state observer with a gain matrix obtained via the solution to a Riccati difference equation [40]. The Kalman filter for the states and the corresponding prediction of the predictor coefficients is given as

$$\begin{aligned}
\epsilon(t) &= y(t) - \boldsymbol{\varphi}(t-L)^H \hat{\mathbf{c}}(t|t-1) && (5.17)\\
\mathbf{K}_f(t) &= \mathbf{K}(t)\boldsymbol{\varphi}(t-L) && (5.18)\\
\hat{\mathbf{x}}(t|t) &= \mathbf{F}\hat{\mathbf{x}}(t-1|t-1) + \mathbf{K}_f\epsilon(t) && (5.19)\\
\hat{\mathbf{c}}(t+k|t) &= \mathbf{H}\mathbf{F}^k\hat{\mathbf{x}}(t|t). && (5.20)
\end{aligned}$$

where $\epsilon(t)$ are the scalar *output innovations*, or the prediction errors, and $\mathbf{K}_f(t)$ is the *Kalman filter gain*. The one step prediction of the predictor coefficients is required in (5.17). The matrix $\mathbf{K}(t)$ depends on the variance of the innovations, denoted $\sigma_\epsilon^2(t)$, and on the state estimation error covariance matrix $\mathbf{S}(t|t-1)$, which are updated by the Riccati difference equation

$$\begin{aligned}
\mathbf{S}(t|t-1) &= \mathbf{F}\mathbf{S}(t-1|t-1)\mathbf{F}^H + \mathbf{G}\mathbf{R}_e\mathbf{G}^H && (5.21)\\
\mathbf{P}(t|t-1) &= \mathbf{H}\mathbf{S}(t|t-1)\mathbf{H}^H && (5.22)\\
\sigma_\epsilon^2(t) &= \sigma_v^2 + \boldsymbol{\varphi}(t-L)^H\mathbf{P}(t|t-1)\boldsymbol{\varphi}(t-L) && (5.23)\\
\mathbf{K}(t) &= \frac{\mathbf{S}(t|t-1)\mathbf{H}^H}{\sigma_\epsilon^2(t)} && (5.24)\\
\mathbf{S}(t|t) &= \mathbf{S}(t|t-1) - \mathbf{K}_f(t)\mathbf{K}_f(t)^H\sigma_\epsilon^2(t) && (5.25)\\
& && (5.26)
\end{aligned}$$

The Riccati equation describes the evolution of the state estimation error covariances $\mathbf{S}(t|t-1)$ and $\mathbf{S}(t|t)$. The matrix $\mathbf{P}(t|t-1)$ is the estimated error covariance matrix for the predictor coefficients $\hat{\mathbf{c}}(t|t-1)$. The measurement noise, $v(t)$, and the process noises, $\mathbf{e}(t)$, are as before assumed to be white and mutually uncorrelated noises with variance σ_v^2 and covariance matrix $\mathbf{R}_e$ respectively.

Initialization

Unless there exists prior knowledge about the states $\mathbf{x}(1|0)$, these are initialized to zero. If the initial states are totally unknown, then the state estimation error covariance matrix must be large. The covariance $\mathbf{S}(0,0)$ can be initialized as a diagonal matrix with large values on the diagonal. This forces the Kalman filter to have a high initial gain, until it obtains reasonable estimates of the states, and thus increases the speed of convergence.

Prediction of coefficients

In an online situation the prediction error $\epsilon(t)$ for $y(t)$, based on the regressors $\varphi(t-L)$ is used at time t to modify $\hat{x}(t|t)$ via (5.19), and to obtain corresponding extrapolated $\hat{c}(t+L|t)$ via (5.20). This estimate of the predictor coefficients is then used together with $\varphi(t)$, which is available at time t, to predict $y(t)$ L steps ahead:

$$\hat{y}(t+L) \quad = \quad \boldsymbol{\varphi}(t)^H \hat{\mathbf{c}}(t+L|t). \tag{5.27}$$

The corresponding prediction error[2] is

$$\bar{\epsilon}(t+L) \quad = \quad y(t+L) - \hat{y}(t+L). \tag{5.28}$$

The matrix $\mathbf{HF}^k$ is built from blocks of $\boldsymbol{\gamma}\mathbf{A}^k$ which is

$$\boldsymbol{\gamma}\mathbf{A}^k = \begin{pmatrix} 1 & \sum_{n=0}^{k-1} p^n \end{pmatrix} = \begin{pmatrix} 1 & \frac{1-p^k}{1-p} \end{pmatrix}. \tag{5.29}$$

The predicted coefficient $\hat{c}_i(t+k|t)$ is thus obtained as the k-step ahead extrapolation of the coefficient at time t, that is $\hat{c}_i(t|t)$, with the sum of the expected changes added as

$$\begin{aligned} \hat{c}_i(t+k|t) \quad &= \quad \hat{c}_i(t|t) + \sum_{n=0}^{k-1} p^n \Delta c_i(t|t) \\ &= \quad \hat{c}_i(t|t) + \frac{1-p^k}{1-p} \Delta c_i(t|t) \end{aligned} \tag{5.30}$$

with the unknown process noise set to zero. When p approaches one (the double integrator) this corresponds to an extrapolation of the present estimate $\hat{c}_i(t+k|t)$ by a linear trend with slope $\Delta c_i(t|t)$.

5.2.1 Tuning parameters

If the process noise elements, $e_i(t)$, are uncorrelated to each other and have the same variance, then the covariance matrix $\mathbf{R}_e$ is a diagonal matrix with entries σ_e^2, that is $\mathbf{R}_e = \sigma_e^2 \mathbf{I}$. The product with $\mathbf{R}_e$ in (5.21) can then be simplified as

$$\mathbf{G}\mathbf{R}_e\mathbf{G}^H = \sigma_e^2 \mathbf{G}\mathbf{G}^H. \tag{5.31}$$

[2]This prediction error, $\bar{\epsilon}(t+L)$, will differ from $\epsilon(t+L)$ obtained from (5.17), since the estimate $\hat{\mathbf{c}}(t+L|t+L)$ is used when computing that error. The error $\epsilon(t)$ is almost white, wheras $\bar{\epsilon}(t)$ often is colored.

The variance of the process noise σ_e^2 and the variance of the measurement noise σ_v^2 can be treated as two tuning parameters. If the predictor coefficients vary fast, then σ_e^2 has to be large. If the measurements are noisy, then σ_v^2 has to be large. These two variances can be combined into one single tuning parameter. To this end the Riccati equations are normalized by $\sigma_e^2 > 0$. This results in a scaling of the covariances $\mathbf{S}(\cdot)$, $\mathbf{P}(\cdot)$ and $\sigma_\epsilon^2(\cdot)$ as

$$\mathbf{S}_s(\cdot) = \frac{\mathbf{S}(\cdot)}{\sigma_e^2} \tag{5.32}$$

$$\mathbf{P}_s(\cdot) = \frac{\mathbf{P}(\cdot)}{\sigma_e^2} \tag{5.33}$$

$$\sigma_{\epsilon,s}^2(\cdot) = \frac{\sigma_\epsilon^2(\cdot)}{\sigma_e^2}. \tag{5.34}$$

The involved equations in the iteration (5.21)-(5.20) are then changed to

$$\mathbf{S}_s(t|t-1) = \mathbf{F}\mathbf{S}_s(t-1|t-1)\mathbf{F}^H + \mathbf{G}\mathbf{G}^H \tag{5.35}$$

$$\mathbf{P}_s(t|t-1) = \mathbf{H}\mathbf{S}_s(t|t-1)\mathbf{H}^H \tag{5.36}$$

$$\sigma_{\epsilon,s}^2(t) = \frac{\sigma_v^2}{\sigma_e^2} + \boldsymbol{\varphi}(t)^H\mathbf{P}_s(t+1|t)\boldsymbol{\varphi}(t) \tag{5.37}$$

$$\mathbf{K}(t) = \frac{\mathbf{S}_s(t|t-1)\mathbf{H}^H}{\sigma_{\epsilon,s}^2(t)} = \frac{\mathbf{S}(t|t-1)\mathbf{H}^H}{\sigma_\epsilon^2(t)} \tag{5.38}$$

$$\mathbf{S}_s(t|t) = \mathbf{S}_s(t|t-1) - \mathbf{K}_f(t)\mathbf{K}_f(t)^H\sigma_{\epsilon,s}^2(t) \tag{5.39}$$

The scaling does not affect the Kalman filter solution as seen when equation (5.24) and (5.38) are compared, as the scaling cancels out. As can be seen in equation (5.37) the parameters σ_v^2 and σ_e^2 do not affect the Kalman gain independently. Only the ratio is important.

We thus have only two design parameters, the noise ratio σ_v^2/σ_e^2 and the position p of the pole in the filter of (5.10). The tuning parameter σ_v^2/σ_e^2, which only appears in equation (5.37), determines how much we trust the measurements versus the model. The adaption becomes slower when σ_v^2/σ_e^2 is increased. The Kalman filter gain $\mathbf{K}_f(t)$ in (5.18) is then reduced, due to an assumed higher noise level in the measurement $y(t)$. If the measurement noise increases then the parameter σ_v^2/σ_e^2 has to increase accordingly. The pole p, appearing in the $\mathbf{F}$ and $\mathbf{G}$ matrixes, determines the assumed smoothness of the predictor coefficients.

Tuning parameters with p close to one

For p close to one, there is actually only one tuning parameter. Instead of using a p dependent $\mathbf{G}$ matrix, the factor $1-p^2$ appearing in $\mathbf{G}\mathbf{G}^H$ could be included in the variance of the process noise. Then the tuning parameter σ_v^2/σ_e^2 is replaced by

$$\frac{\sigma_v^2}{\sigma_e^2} \cdot \frac{1}{1-p^2}. \tag{5.40}$$

The remaining p dependence in matrix $\mathbf{A}$, the building block for $\mathbf{F}$, can for p close to one be dropped

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 0 & p \end{pmatrix} \approx \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}. \tag{5.41}$$

The relative approximation error on the state estimation error covariance matrix is of order $1-p$.

Thus for p close to one there is only one tuning parameter and different choices of p and σ_v^2/σ_e^2 give the same performance of the Kalman filter as long as (5.40) is constant.

Example 5.1

If the variance ratio is increased a factor of ρ then a corresponding change of p^2 from p_1^2 to p_2^2, which keeps the performance constant, is given by

$$p_2^2 = 1 - \rho + \rho p_1^2. \tag{5.42}$$

5.3 Iterative Prediction

There are mainly two reasons for using iterative predictors.

- Complexity reduction. Only one predictor has to be adapted for any prediction range.
- Less sensitivity to errors in the assumptions made in the model filter.

The first reason is obvious. The second will be demonstrated in Example 5.2 and will be explained in the following.

A predictor in linear regression form uses a regressor with samples from a T samples long memory, thus a window that stretches from t back to $t-T+1$, to predict L steps into the future. With a direct approach, the predictor $\hat{f}_L(t, \mathbf{y}(t))$ in equation (5.2) is adapted to minimize the L-step output innovations. For the linear predictor in Section 5.2 this means that we adapt using (5.17)

$$\epsilon(t) = y(t) - \boldsymbol{\varphi}(t-L)^H \hat{\mathbf{c}}(t|t-1)$$

but we predict using (5.27)

$$\hat{y}(t+L) = \boldsymbol{\varphi}(t)^H \hat{\mathbf{c}}(t+L|t).$$

As only samples up to time t are available, the regressor used to obtain the last available adaption of the predictor is $\varphi(t-L)$. Instead of the desired predictor $\hat{\mathbf{c}}(t+L|t+L)$ a prediction of the predictor, $\hat{\mathbf{c}}(t+L|t)$, is used. For a linear predictor with a *correct model* for the coefficient dynamics in the Kalman-filter, this is a not a problem as we will see in Example 5.2.

The extrapolation of coefficients imposes higher demands on the accuracy of the model filter in the adaptive Kalman algorithm than the tracking of the coefficients does. If the model is not fully correct, then the extrapolation of the predictor coefficients will suffer from errors in the L-step prediction (as demonstrated in Example 5.2). A shorter prediction range result in a smaller error. This is especially evident in a situation where the prediction range L is large, as compared to how fast the prediction coefficients should change. Then it is a drawback, if only regressors up to time $t-L$ can be used in the adaption of the predictor.

5.3.1 Iterated FIR-predictor

To decrease the range for prediction of the predictor coefficients, iterations can be used. The L-step predictor can also be obtained as an iterated m-step predictor, where the integer m should be a divisor of L. In the iterative approach the prediction $\hat{y}(t+m)$ from the m-step prediction is used as a substitute for $y(t+m)$. The subsampled memory is thus extended into the future using predicted values. We can define a regressor that contains predicted and measured values as

$$\varphi_p(t+km|t) = [\hat{y}(t+km|t) \dots \hat{y}(t+m|t) \; y(t) \dots y(t+T-km)]^H. \quad (5.43)$$

For $\varphi_p(t+m|t)$ only the first element of the vector is predicted, the rest are measured. For $\varphi_p(t+2m|t)$ the two first are predicted and so on until $\varphi_p(t+T|t)$ where the whole regressor is filled with predicted values.

Time varying FIR predictor

For a time varying system the predictor coefficients will be time varying. As the Kalman-filter produces predictions of the future coefficients, the iterated one-step predictor is

$$\begin{aligned}
\hat{y}(t+m|t) &= \varphi(t)^H \hat{\mathbf{c}}_m(t+m|t) \\
\hat{y}(t+2m|t) &= \varphi_p(t+m|t)^H \hat{\mathbf{c}}_m(t+2m|t) \\
&\vdots \\
\hat{y}(t+L|t) &= \varphi_p(t+L-m|t)^H \hat{\mathbf{c}}_m(t+L|t), \qquad (5.44)
\end{aligned}$$

where $\varphi_p(t+km|t)$ is defined as in (5.43). Here $\hat{\mathbf{c}}_m(t+km|t)$ is an m-step predictor with coeffcients extrapolated km steps ahead.

The coefficients of a predictor for a shorter prediction range vary slower than those for a predictor for a longer prediction range. This is because the changes that are needed are accumulated in the iterations. A very simple model for the coefficient variation, like an AR1I-model, is therefore better suited for short prediction ranges than for long ranges.

Time invariant AR-system

For a linear FIR-predictor the iteration procedure can be interpreted as a new filter, that can be directly calculated from the parameters of the m-step predictor [41]. This is the approach that we will follow in Chapter 6 whenever iterated predictors are used.

If we assume that a time-invariant AR-process is an appropriate description of $y(t)$, which we argued for in Section 4.2, then the time series $y(t)$ is described by

$$A(q^{-1})y(t) = \nu(t), \qquad (5.45)$$

where $A(q^{-1})$ is a polynomial of degree $M+1$. (The filter is derived for the non-subsampled case but it is straight forward to generalize the derivation to the subsampled predictor.) The linear minimum variance one-step predictor is given by

$$\hat{y}(t+1|t) = (1 - A(q^{-1}))y(t) \stackrel{\triangle}{=} C_1(q^{-1})y(t). \qquad (5.46)$$

The minimum variance k-step linear predictor is given by

$$\hat{y}(t+k|t) = C_k(q^{-1})y(t), \qquad (5.47)$$

where the FIR filter $C_k(q^{-1})$ is given as the M degree polynomial solution to

$$1 = A(q^{-1})F(q^{-1}) + q^{-k}C_k(q^{-1}) \qquad (5.48)$$

with $F(q^{-1})$ and $q^{-k}C_k(q^{-1})$ as the quotient and remainder respectively, when dividing 1 (the left hand side of (5.48)) by $A(q^{-1})$ [41]. If we identify the coefficients of the one-step FIR-predictor we obtain the AR-model as $\hat{A}(q^{-1})) = 1 - \hat{C}_1(q^{-1})$ and the k-step predictor $\hat{C}_k(q^{-1})$ is obtained by solving (5.48). Thus, given the minimum-variance one-step FIR predictor of order M, the minimum-variance FIR predictor of order M for *all prediction horizons* can be derived.

To use the minimum variance k-step FIR-predictor $C_k(q^{-1})$ is equivalent to k iterations of the one-step predictor (5.46). The k-step predictor is thus formed by a cascade of k one-step predictors, where predicted values replace the unknown future values in the predictor.

Iteration using $\hat{\mathbf{c}}_m(t+m|t)$, estimated by the Kalman filter, and assuming that the coefficients do not change over the prediction range result in

$$\begin{aligned}
\hat{y}(t+m|t) &= \varphi(t)^H \hat{\mathbf{c}}_m(t+m|t) \\
\hat{y}(t+2m|t) &= \varphi_p(t+m|t)^H \hat{\mathbf{c}}_m(t+m|t) \\
&\vdots \\
\hat{y}(t+L|t) &= \varphi_p(t+L-m|t)^H \hat{\mathbf{c}}_m(t+m|t). \qquad (5.49)
\end{aligned}$$

This differ from (5.44) in that the coefficients are assumed to be constant and that only m-step extrapolation of the coefficients is performed. This is suboptimal but makes the predictor less sensitive to errors in the model filter. The same prediction is obtained if, at each time instance t, the coefficients of $\hat{\mathbf{c}}_m(t+m|t)$ are used in $C_1(q^{-1})$ and the k-step predictor $C_k(q^{-1})$, with $k = L/m$, is calculated as in (5.48). Then the coefficients of $C_k(q^{-1})$ can be used in the L-step predictor $\hat{\mathbf{c}}_{iter}(t)$,

$$\begin{aligned}
\hat{y}(t+L|t) &= \varphi_p(t+L-m|t)^H \hat{\mathbf{c}}_m(t+m|t), \\
&= \varphi(t)^H \hat{\mathbf{c}}_{iter}(t).
\end{aligned}$$

Thus, the iteration process using $\hat{\mathbf{c}}_m(t+m|t)$ k times corresponds to one step with $\hat{\mathbf{c}}_{iter}(t)$. This interpretation makes it possible to perform analysis of the frequency gain and noise gain for the iterated predictor.

5.3.2 Simulation

The following example illustrates the drawback of designing the L-step predictor directly as compared to iterating the m-step predictor.

Example 5.2

In this example we compare the predictor coefficient estimates for a predictor that uses just one delayed sample. We compare a direct ten-step adaptive predictor, a ten times iterated adaptive one-step predictor and the optimal time varying coefficient. The test signal is a single chirp corrupted by an additive white noise

$$y(t) = e^{i\frac{\pi}{2N}(t-\frac{N}{2})^2} + v(t). \tag{5.50}$$

The SNR is 30 dB and the length of the data series is $N = 512$ samples. This chirp spans half of the frequency band, that is from 0.25 Hz to -0.25 Hz (normalized frequency with the sampling frequency set to 1 Hz). The time varying L-step predictor that perfectly predicts $y(t-L)$ from $y(t)$ in a noiseless case is

$$\hat{y}(t+L|t) = c(t)y(t) = e^{i\frac{\pi L}{N}(t+\frac{L-N}{2})}y(t). \tag{5.51}$$

Thus the prediction coefficient is time varying and oscillates with $L/2N$ Hz. Note that the oscilative frequency is proportional to the prediction range L. The time varying prediction coefficient is given by $c(t) = y(t+L)/y(t)$. For the high SNR 30 dB, the optimal coefficient for the L step predictor is approximately given by $c(t)$ in equation (5.51) and the prediction gain will be equal to the SNR.

Two adaptive predictors are used for estimating this coefficient for L=10. Both method use the state space model (5.12)-(5.13) and the adaptive Kalman filter. The parameters p and σ_v^2/σ_e^2 are tuned separately for the direct ten-step predictor and the iterated one-step predictor to optimize their prediction performance. The iterated predictor uses the product of the predicted coefficients $\hat{c}_1(t+1|t)$ up to $\hat{c}_1(t+L|t)$ as in (5.44) for a filter with one coefficient,

$$\hat{y}(t+L|t) = \hat{c}_1(t+1|t)\hat{c}_1(t+2|t)\cdots\hat{c}_1(t+L|t)y(t). \tag{5.52}$$

whereas the direct predictor uses the predicted coefficient as in (5.27)

$$\hat{y}(t+L|t) = \hat{c}_L(t+L|t)y(t), \tag{5.53}$$

The ten times iterated one-step predictor achieves a prediction gain, measured after the initial transient, of 19 dB whereas the direct ten-step predictor only achieves 10 dB. If instead of the predicted coefficients the filtered coefficients $\hat{c}_1(t|t)^L$ and $\hat{c}_L(t|t)$ would have been used for the iterated and direct prediction respectively, then the prediction gain would have been only 9 dB and 4 dB respectively.

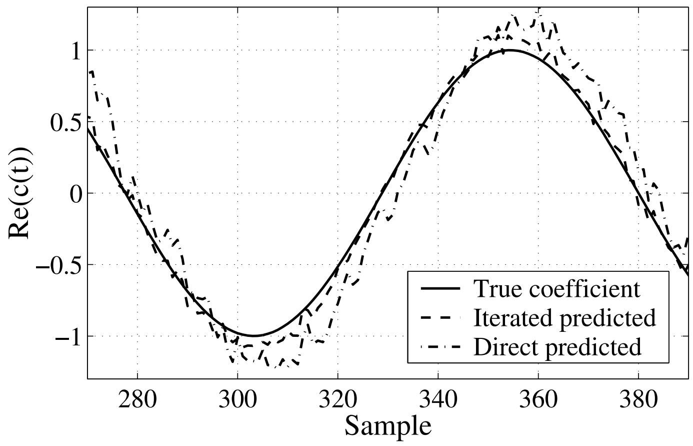

Figure 5.1: Comparison of the optimal predictor coefficient for 10 step prediction, to the coefficients of the iterated adaptive one-step predictor as in (5.52) and the direct adaptive 10-step prediction as in (5.53) of a chirp.

The reason for the relatively poor performance of the direct predictor, even with predicted coefficients, is that the model in the Kalman-filter is wrong. The oscillative behavior of the prediction coefficient is not well modeled by the AR1I model. The iterative predictor is less sensitive to this error in the model, as the one-step predictor coefficient oscillates with 0.98 mHz, ten times slower than the 9.8 mHz of the 10-step predictor coefficient and thus better resembles the trend anticipated by the AR1I model. This can be seen in Figure 5.1 where the direct coefficient predicted L steps ahead results in an overshoot at the peaks, as the coefficient is assumed to continue along the pressent trend.

With an AR2I model, tuned to the frequency of the oscillation of the coefficient, the iterated predictor achieves a prediction gain of 27 dB while 29 dB is obtained for the direct predictor. Thus the use of an AR1I model, as compared to the correct AR2I, results in a loss of performance that is much higher for the direct predictor than for the iterated.

The example above illustrates an important point also valid for mobile radio data. The coefficients for a one-step predictor vary slower than the coefficients of an L-step predictor. The use of an L-step predictor thus places higher demands on the accuracy of the model used in the Kalman-filter.

The iterated predictor is not always to prefer. When the coefficients change slowly it is preferable to use the direct estimate as it minimizes the correct criterion.

In the following example we will see that the iterated approach is competitive, as compared to the direct prediction approach, on a simulated tap. A comparison to the FIR block prediction approach of Chapter 4 is also made. We clearly see that for a scenario with close scatterers the adaptive approach is well motivated.

Example 5.3

For the simple simulation scenario in Section 2.4, with a mobile station driving past three close scatterers at 60 km/h, the performances of the direct and iterated adaptive FIR predictors and an FIR block predictor are compared. The SNR is set to 30 dB, a realistic value also for measured and Wiener-smoothed data. (In Subsection 4.5.1 the SNR is 120 dB.)

All of the predictors have the same FIR structure and use the same regressors. As in Section 4.5.1 we use a sub-sampled regressor with the delays $[0, 25, 50, 75, 100, 125, 150, 175, 200]$. The length of the window is thus 22 ms, or 201 samples. As the sub-sampling factor is m =25, a one-step ahead predictor actually predicts 25 samples ahead. Predictions for longer ranges are obtained either by iterating the predictor for $L = 25$ (iterated) or by tuning a predictor for the desired range. Kalman adaption algorithms with coefficient prediction (5.20) are used. The predictors are tuned, by full search over a grid of tuning parameters, to give their best performance for the different prediction ranges.

The prediction gain of the predictors is evaluated on the noise free data using noisy regressors. The initial transients of the adaptive predictors are not included in the performance evaluation. We thus evaluate the prediction gain on the last 1300 samples only, also for the block predictor.

In Table 5.1 the results are presented. As can be seen in Table 5.1, the block predictor simply can not compete with the adaptive predictors. As p is almost one, similar performance is achieved for a range of settings as was shown by equation (5.42). Thus the tuning parameters $p = 0.999995$, $\sigma_v^2/\sigma_e^2 = 10^2$ can be replaced by $p = 0.9995$, $\sigma_v^2/\sigma_e^2 = 10^4$ with almost the same performance.

Horizon L	Predictor	PG [dB]	NG [dB]	$\frac{\sigma_v^2}{\sigma_e^2}$	p
25	Direct	27.5	0.0	10^2	0.999995
	Block	16.7	11.0		
50	Direct	22.6	-1.9	10^2	0.999995
	Iterated	24.3	-0.8	$3 \cdot 10^2$	0.999995
	Block	9.2	16.0		
100	Direct	15.8	-1.2	10^2	0.999995
	Iterated	16.9	-1.6	10^4	0.9995
	Block	6.7	9.5		

Table 5.1: Table of the results in Example 5.3 for the different predictors on the simulated data. For $L = 25$ the direct and iterated approach are the same.

5.4 Nonlinear Terms in the Regressor

It is possible to extend the regressor $\boldsymbol{\varphi}(t)$ with nonlinear combinations of the samples in the memory. The main idea is to find nonlinear terms that reflect the dynamics of the system and that require almost time invariant predictor coefficients. We can expect the coefficients for the nonlinear terms to be treated successfully only if they are slowly time varying.

If we focus on the chirp, we obtain a candidate for such a nonlinear extension of the regressor. A single chirped signal

$$y(t) = \beta e^{i(\omega t + \zeta t^2)}, \tag{5.54}$$

can be modeled by the following deterministic recursion

$$y(t+1) = e^{i2\zeta} \frac{y^2(t)}{y(t-1)}, \tag{5.55}$$

and for time steps of length L we obtain

$$y(t+L) = e^{i2L^2\zeta} \frac{y^2(t)}{y(t-L)}. \tag{5.56}$$

Thus for an L-step predictor the use of a nonlinearity of the form

$$\varphi_i(t) = \frac{y^2(t)}{y(t-L)}, \tag{5.57}$$

in the regressor might result in improved performance of the predictor.

A drawback with many nonlinear regressors is their amplification of additive noise. In Appendix 5.A it is shown that the nonlinear regressor (5.57) result in a 7 dB noise amplification when used on a chirp, whereas linear terms in the regressor, like $y(t)$, gives 0 dB noise amplification.

In a fading signal the amplitude of $y(t)$ will occasionally be very low. This causes the term $1/y(t-L)$ to grow without bound in the fading dips. To avoid this, the term can be regularized as

$$\varphi_i(t) = \frac{y^2(t)}{y(t-L) + \rho/y^*(t-L)} = \frac{y^2(t)y^*(t-L)}{|y(t-L)|^2 + \rho} \tag{5.58}$$

The regularization constant ρ limits the magnitude and eliminates discontinuities of this regressor. As can be seen in (5.58) the regularization is smaller when $|y(t-L)|$ is far from zero, that is when it is not needed and larger when $|y(t-L)|$ is small. The following bound holds

$$\frac{1}{|y(t-L) + \rho/y^*(t-L)|} \leq \frac{1}{2\sqrt{\rho}}. \tag{5.59}$$

Example 5.4

In Example 5.2 with a single chirp the use of the nonlinear term with $y^2(t)/y(t-L)$ renders a PG of 22.9 dB, almost 7 dB below the limit set by the SNR, due to the noise amplification. When this nonlinearly transformed regressor is introduced, the associated prediction parameter will not be time varying. The performance is better than the iterated adaptive linear predictor using an AR1I model (19 dB) but worse than the direct predictor using the AR2I model (27 dB). Although the nonlinear term is tailored for prediction of chirps, the best linear adaptive predictor is superior. This is mainly due to the lower noise gain of the linear predictor.

Example 5.5

In Example 5.3 the three contributing reflections cause chirps in the resulting channel. In spite of this, the inclusion of two nonlinear terms does not improve the performance of the prediction. The same performance could be achieved without the nonlinear terms in the regressor.

The two nonlinear terms are $y^2(t)y^*(t-m)/(|y(t-m)|^2+\rho)$ and $y^2(t)y^*(t-L)/(|y(t-L)|^2+\rho)$. Here m is the subsampling factor and L is the prediction horizon. The terms had to be regularized with a $\rho > 0.1$ to avoid that the prediction performance was degraded. (Without regularization of the nonlinear term the performance is highly degraded.)

5.5 Conclusion

The Kalman adapted predictor is clearly to prefer to the block predictor in situations with dominant near scatterers, as illustrated in Example 5.3. The AR1I-model for the parameter variation is sufficient to describe the parameter variations in this case. The pole p of the AR1I-model has to be placed close to the unit circle ($z = 1$) to give good tracking and prediction of the predictor coefficients. The iterated predictor is found to give equal or better performance, as compared to direct predictin, on the simulated channel with close scatterers.

Use of the proposed nonlinear term (5.57) does not result in increased performance for prediction of a chirp signal, as compared to a Kalman adapted linear predictor with the proper choice of hyper-model for the coefficient variation.

The examples indicate that the Kalman adapted linear FIR predictor can predict the type of signals that are anticipated in a time varying mobile radio channel.

5.A Noise Amplification of the Proposed Nonlinear Regressor Term

A single chirped signal is given as

$$y(t) = \beta e^{i(\omega t + \zeta t^2)}. \tag{5.60}$$

We want to express $y(t+L)$ in terms of observations up to time t. To this end we look at the phase of the signal for different times,

$$\begin{aligned}
\angle y(t) &= \omega t + \zeta t^2 && (5.61)\\
\angle y(t-L) &= \omega(t-L) + \zeta(t-L)^2 = \angle y(t) - 2\zeta L t + \zeta L^2 && (5.62)\\
\angle y(t+L) &= \omega(t+L) + \zeta(t+L)^2 = \angle y(t) + 2\zeta L t + \zeta L^2 && (5.63)\\
&= 2\angle y(t) - \angle y(t-L) + 2\zeta L^2. && (5.64)
\end{aligned}$$

The phase of $y(t+L)$ can be expressed as a wheighted sum of the phases for $y(t)$, $y(t-L)$ and a time invariant factor. The summation of phases corresponds to multiplication and division of the corresponding signals as

$$y(t+L) = e^{i2\zeta L^2} \frac{y^2(t)}{y(t-L)}, \tag{5.65}$$

which thus is motivates the choise of the nonlinear element of the regressor

$$\varphi_i(t) = \frac{y^2(t)}{y(t-L)}. \tag{5.66}$$

To analyze the sensitivity to noise of the nonlinearity (5.66) we add a white Gaussian noise $n(t)$, with variance σ_n^2, to the time series $y(t)$. The time series $y(t)$ is assumed to be independent of the additive noise $n(t)$. The proposed term (5.66) is then corrupted as

$$\varphi_i(t) = \frac{y^2(t) + 2y(t)n(t) + n^2(t)}{y(t-L) + n(t-L)}. \tag{5.67}$$

The deviation of (5.67) from (5.66) is given by

$$e(t) = \frac{y^2(t)}{y(t-L)} - \frac{y^2(t) + 2y(t)n(t) + n^2(t)}{y(t-L) + n(t-L)}, \tag{5.68}$$

where $e(t)$ is the error term of the nonlinear regressor. If the noise is small relative to the magnitude of the signal we can approximate the inverse of $y(t-L)$ with additive noise as

$$\frac{1}{y(t-L) + n(t-L)} \approx \frac{1}{y(t-L)} \left(1 - \frac{n(t-L)}{y(t-L)}\right). \tag{5.69}$$

In (5.69) we used a Taylor expansion of $1/(1+x)$, with $x = n(t-L)/y(t-L)$. For a single chirp and high SNR this is a valid approximation but for sums of chirps, the signal $y(t)$ will have zero crossings which would further increase the effect of the noise. The present analysis thus describes a best case scenario. If we use the approximation (5.69) in (5.68), the error on the regressor can be approximated as

$$e(t) \approx \frac{y^2(t)n(t-L) - (2y(t)+n(t))(y(t-L)-n(t-L))n(t)}{y^2(t-L)}. \tag{5.70}$$

The power of the error is

$$P_e = E\{|e(t)|^2\} = \frac{|y(t)|^4\sigma_n^2 + 4|y(t)|^2|y(t-L)|^2|\sigma_n^2 + O(\sigma_n^4)}{|y(t-L)|^4}, \tag{5.71}$$

where the average is taken with respect to $n(t)$ and where $O(\sigma_e^4)$ includes all the terms with higher order moments for the noise. These higher order moment are assumed to be comparably small and will be disregarded in the following calculations.

For a single chirp, $|y(t)|$ is constant for all t and we thus obtain

$$P_e \approx 5\sigma_n^2. \tag{5.72}$$

The noise power is thus amplified five times by the nonlinear term. With a second order term included in the approximation of the inverse of $y(t-L)+n(t-L)$ and keeping higher order moments up to order four, the approximate error can be shown to be

$$P_e \approx (5 + 8\sigma_n^2)\sigma_n^2. \tag{5.73}$$

The noise is thus further amplified when the variance of the additive noise is increased.

If we instead assume that there is a change of phase and magnitude between the samples as

$$y(t) = r(t,L)\mathrm{e}^{i\omega(t,L)}y(t-L), \tag{5.74}$$

then we can write the expected power of the error as

$$P_e(t) \approx r^2(t,L)(4 + r^2(t,L))\sigma_n^2. \tag{5.75}$$

As the studied signal is close to band-limited there will be no abrupt changes and for short horizons L, $r(t,L)$ will be close to one. We can thus conclude that using a nonlinear term of the type (5.66) results in a noise amplification of about 5 times, that is 7 dB, in the best cases (no zero crossings). When there occur zero crossings in $y(t)$, $r(t, L)$ can grow without bounds and the noise is highly amplified. *This limits the use of this nonlinear term to situations with high SNR and no zero crossings.*

5.A.1 Regularization

To avoid the large noise amplification and to limit the output at zero crossings we can regularize the term as

$$\varphi_i(t) = \frac{y^2(t)}{y(t-L) + \rho/y^*(t-L)} = \frac{y^2(t)y^*(t-L)}{|y(t-L)|^2 + \rho}. \tag{5.76}$$

The important property of the regularized denominator is that the output becomes bounded when the input is bounded. To demonstrate this we find the maximum value of the magnitude of the regularized term, when $y(t-L)$ change, as a function of the regularization parameter ρ. Set the derivative of the magnitude of the regularized term to zero as

$$\begin{aligned} \frac{d|\varphi_i(t)|}{d|y(t-L)|} &= \frac{d}{d|y(t-L)|}\left|\frac{y^2(t)y^*(t-L)}{|y(t-L)|^2 + \rho}\right| \\ &= |y(t)|^2 \frac{\rho - |y(t-L)|^2}{(|y(t-L)|^2 + \rho)^2} = 0. \end{aligned} \tag{5.77}$$

The solution to equation (5.77) corresponding to the maximum of $|\varphi_i(t)|$ is

$$|y(t-L)| = \sqrt{\rho}. \tag{5.78}$$

The magnitude of the term (5.76) will thus be limited by

$$|\varphi_i(t)| \leq \frac{|y(t)|^2\sqrt{\rho}}{\rho + \rho} = \frac{|y(t)|^2}{2\sqrt{\rho}}. \tag{5.79}$$

Thus, if $y(t)$ is bounded, then the output will be bounded. The quadratic term $y^2(t)$ will never be amplified by more than $1/(2\sqrt{\rho})$.

Chapter 6

Results for Adaptive Predictors

In the following we will see how the adaptive predictors, presented in Chapter 5, perform on the channel coefficient-estimates from the data base of measured broadband data at 1880 MHz, with 6.4 MHz bandwidth, described in Chapter 3. Noise reduction, is performed with Wiener smoothers on each tap, as in Section 4.5.2, for the block prediction evaluation.

As opposed to the block prediction methods the adaptive methods do not need a separate validation set. The adaptive predictors have been tuned so that their initial transients are completed within 78 ms, which is half of the length of the time series. The prediction performance is then measured over the remaining 78 ms intervals.

Even though the prediction is performed on a time series, the change of the channel coefficients is due to movement in space (see Section 2.2). By converting the results from prediction in time to prediction over distance, we obtain a velocity independent measure of the obtained prediction gain. The conversion factor from range measured in samples (time) to range measured in wavelengths (distance) is given by the Doppler frequency, f_D, divided by the channel sampling frequency, which is f_s =9.14 kHz in the measurements treated here. That is, one sample corresponds to f_D/f_s wavelengths of traveled distance. The Doppler frequency is estimated as half the width of the Doppler spectrum, as indecated by the region where the fall-off in dB is highest. This estimate of the Doppler frequency will thus not be affected by any frequency offset in the receiver that may displace the spectrum, as in Figure A.18.

As the time series of a tap is approximately band-limited it can be subsampled, almost without loss of information, as long as aliasing is avoided. To use a subsampled predictor on a time series which is not subsampled, results in a larger available training set as compared to using a non-susbsampled predictor on a subsampled time series. As both predictors would have the same number of coefficients, the subsampled predictor has a faster response to changes of the underlying dynamics of the channel and a faster convergence, as it can use more observations of the time series. We will therefore use subsampled predictors on time series which are not subsampled.

6.1 Prediction of a Tap

In the following the Kalman adapted predictors described in Section 5.1 and the iterated predictor described in Section 5.3 are evaluated on measured channel impulse responses. The predictors are linear FIR-predictors where the variation of the coefficients of the predictor are modeled as AR1I processes.

The adaptive linear filter is quite similar to the time invariant block predictor. The length of the memory should encompass at least one full oscillation of the dominant frequency. The subsampling factor for the memory must be low enough to avoid aliasing, that is the sampling frequency divided by the subsampling factor has to be at lest twice as large as the Doppler frequency. If the dominant frequency is the Doppler frequency, then the memory should correspond to a traveled distance of one wavelength. If most of the energy comes from reflections from the side, then the dominant frequency is lower and a longer memory is needed. This is the case in the example below.

Example 6.1

Concider channel A, presented in Appendix A. The vehicle drives at roughly 90 km/h and there are strong wavefronts coming in from the side, as can be seen in the average Doppler spectrum in Figure A.6. The Doppler frequency is 157 Hz, so the subsampling factor has to be below $f_s/2f_D = 9143/2/157 = 29$. A memory length corresponding to a traveled distance of one wavelength, that is $f_s/f_D = 58$ samples, result in a poor predictor.

Most of the power is collected in one tap as seen in the PDP in Figure A.4 in Appendix A. The strongest tap and the second strongest tap are shown in Figure 6.1. The two taps are very different. The second

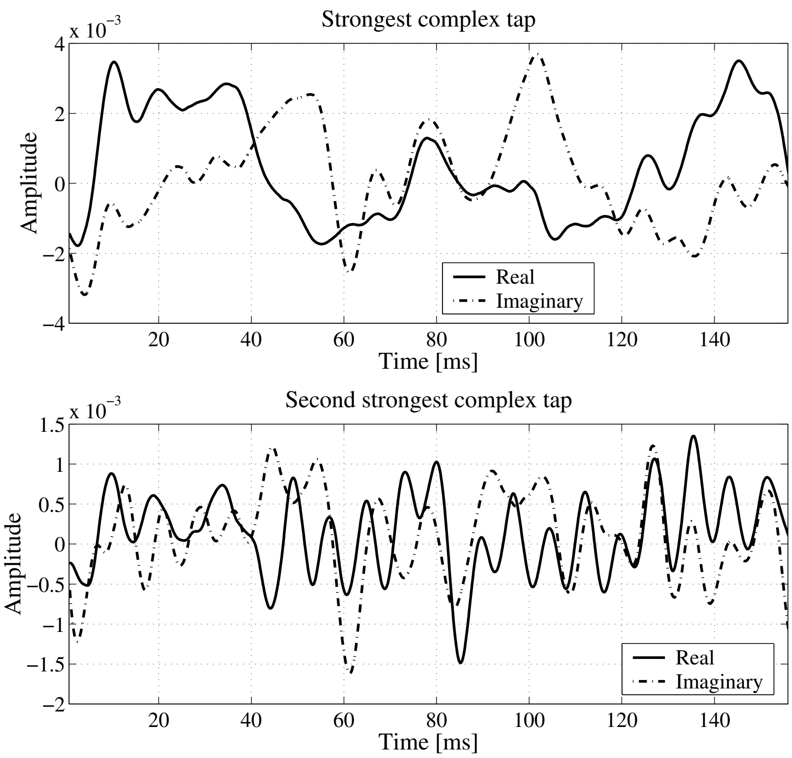

Figure 6.1: The strongest tap (upper plot) and second strongest tap (lower plot) of channel A in Example 6.1. Strong wafronts from the side cause the slow oscilations for the stronger tap. Real part is denoted by solid line and imaginary part by dash-dotted line.

strongest tap does not contain the dominating slow oscillating component, corresponding to wavefronts coming from the side, that is apparent in the strongest tap.

An adaptive predictor with 251 samples in the memory and a subsampling factor of 10, is used for prediction of the strongest tap. The length of the memory is 27.5 ms, which still is shorter than the period for the slow oscilations observed in Figure 6.1. (The memory is chosen short to reduce the number of adapted coefficients for the predictor. There will always be a trade off between the number of coefficients on one hand and the length on the memory on the other.)

To find the best choice for the tuning parameters, σ_v^2/σ_e^2 and p, the prediction gain is evaluated for a grid of different tuning parameters. A search over the tuning parameters reveals that there are two different choices of tuning parameters, resulting in either agile or slow change of the prediction coefficients, that give similar performance. Other choises give lower performance.

Prediction horizon L				PG	NG		
Samp.	[ms]	λ	Predictor	[dB]	[dB]	$\frac{\sigma_v^2}{\sigma_e^2}$	p
10	1.1	0.17	Direct agile	27	-3	10^{-2}	0.99
			Direct slow	24	13	10^{12}	0.0
30	3.3	0.51	Direct agile	8	-3	10^{-2}	0.99
			Direct slow	8	24	10^{12}	0.0
			Iterated agile	8	-5	10^{-2}	0.97
			Iterated slow	9	24	10^{12}	0.0

Table 6.1: Table of the results for the different predictors on the strongest tap. For $L = 10$ the direct and iterated approach are the same.

Prediction horizon L				PG	NG		
Samp.	[ms]	λ	Predictor	[dB]	[dB]	$\frac{\sigma_v^2}{\sigma_e^2}$	p
10	1.1	0.17	Direct agile	21	-3	10^{-2}	0.99
			Direct slow	18	15	10^{12}	0.0
30	3.3	0.51	Direct agile	2	-3	10^{-2}	0.99
			Direct slow	3	23	10^{12}	0.0
			Iterated agile	2	-3	10^{-2}	0.99
			Iterated slow	3	23	10^{12}	0.0

Table 6.2: Table of the results for the different predictors on the second strongest tap. For $L = 10$ the direct and iterated approach are the same. The performance is about 6 dB worse than for the largest tap.

In Tables 6.1 and 6.2 the PG and the corresponding tuning parameters are listed for the strongest and second strongest tap respectively. For prediction 10 samples ahead p is best chosen close to 1 and σ_v^2/σ_e^2 small. This result in an agile high-gain Kalman afilter with large variations for the predictor coefficients. For prediction of 30 samples the difference between the agile and slow predictors are smaller. The best prediction gain is achived with iterated predictor with p chosen as 0 and σ_v^2/σ_e^2 huge, that is the process noise is assumed to be small. This results in a Kalman estimator that behaves like an RLS estimator [42], with a forgetting factor close to one. The prediction coefficients hardly change after the initial transient.

Even though the noise gain becomes high the prediction performance is still good.

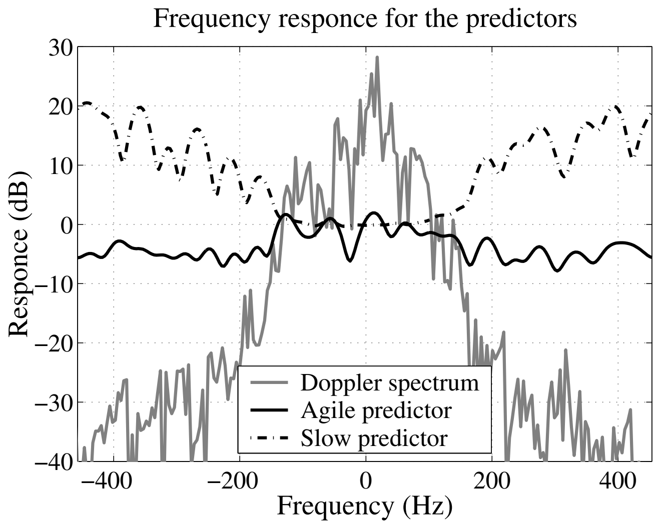

Figure 6.2: Average frequency responses of the direct 10-step ahead predictors on the strongest tap in Example 6.1. The predictor with slowly varying coefficients (dash-dotted) has a higher high-frequency gain than the agile predictor (solid). The gray line is the Doppler spectrum of the strongest tap, with normalized power.

The average frequency response of the agile and slow predictor is shown in Figure 6.2. The slow predictor behaves like a predictor for a time-invariant band-limited processes [43]. It amplifies high frequencies, for which the power in the signal is low, and in the passband the amplification is 0 dB. In contrat, the time varying agile predictor on average suppresses high frequencies, which is reflected in the low noise gain values pressented in Table 6.1.

6.1.1 Evaluation of the PG of the complex taps on the measurements

From each of the 25 measurements, the taps that account for at least 90% of the total power in the channel, are used in a prediction study. The number of used taps from each channel is in the range of 1-15, on average seven taps are used. The predictor is the Kalman adapted iterative FIR predictor, based on AR1I models, pressented in Section 5.2. The same length of the memory, $T = 251$ samples, and subsampling factor, $m = 10$, are used for prediction of all taps. The FIR predictor has 26 coefficients, so the Kalman filter has 52 states. The tuning parameters are chosen to give a slow variation of coefficients, that is p=0 and $\sigma_v^2/\sigma_v^2 = 10^{12}$. This is not an optimal choice of either memory length, subsampling factor or tuning parameters, but it is a set of parameters that works reasonably well on most observed channels.

The predictor is adapted to predict ten samples ahead. Then the predictor is iterated as in Section 5.3.1, in steps of ten samples, up to a range of 100 samples. In the iteration the predictor coefficients are assumed to be time invariant. The performance of the predictor is evaluated at every range, on the last half of the data set (710 samples). Thus, we do not include the trancient behaviour in the first half of the data set. The range, in samples, is then expressed in distance measured in wavelengths. The scatter plot in Figure 6.3 shows all the achieved prediction gains for all the 175 taps and all ranges. The local average of the prediction gain in dB, that is the gray line, is calculated for every 100 distance points. The performance can also be presented as the obtained NMSE, as in Figure 6.4. (The averaging over NMSE values gives a different result as compared to averaging prediction gain on a dB scale. Bad predictions tend to dominate the NMSE average.) There is a large variation in how good the taps can be predicted by the adaptive FIR predictor. However, there is a clear trend that the predictability falls off as the logarithm of the range increases. Better performance can be achieved with individual tuning of the parameters of the adaptive predictor, for each tap.

6.2 Power Prediction

To see how well a whole channel can be predicted we look at predictions of the total received power. The complex taps contain the information about the phase of the complex sinusoids (multipath components) they consist of, whereas the power is just the sum of the squared envelopes. It is the phase that governs when a fading dip occurs. The time-series of received power lacks this extra information. It is thus advantageous to use predictions of the complex taps to obtain predictions of the power as the sum of the squared magnitude of the predictions of the M taps

$$\hat{P}(t+L) = \sum_{k=1}^{M} |\hat{h}_k(t+L)|^2. \tag{6.1}$$

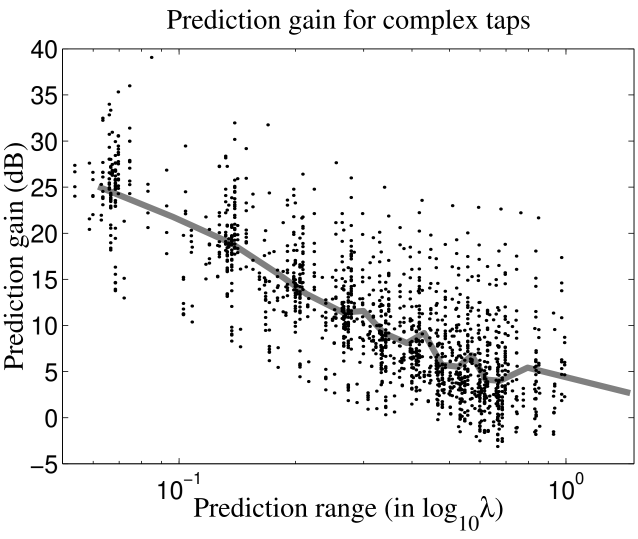

Figure 6.3: Scatter plot of the prediction gain (PG) for the complex taps, as a function of prediction range measured in wavelengths. (Each point corresponds to one tap.) The gray line is the local average PG, taken over 100 points.

It should be noted that it is not always the predictor that also achieves the best PG for the complex channel that achieves the lowest NMSE for the power. For example, a rotation of the predicted complex tap causes a high NMSE for the complex tap but does not affect the power prediction.

The important property in prediction of the power is the relative error. That is the accuracy of the prediction on the dB scale

$$E_{dB}(t) = 10\log_{10} P(t) - 10\log_{10} \hat{P}(t). \tag{6.2}$$

A measure of performance is how large portion of time the prediction is within a certain performance bound relative to the true power. That is, for how large portion of the time is

$$|E_{dB}(t)| \leq B, \tag{6.3}$$

where B is the bound measured in dB. In the following we will investigate both the NMSE for the predicted power for different prediction ranges and the measures of time within performance bounds.

In receivers that use only a subset of the taps, as in many proposed RAKE-receivers, the available power would then be the sum over this subset

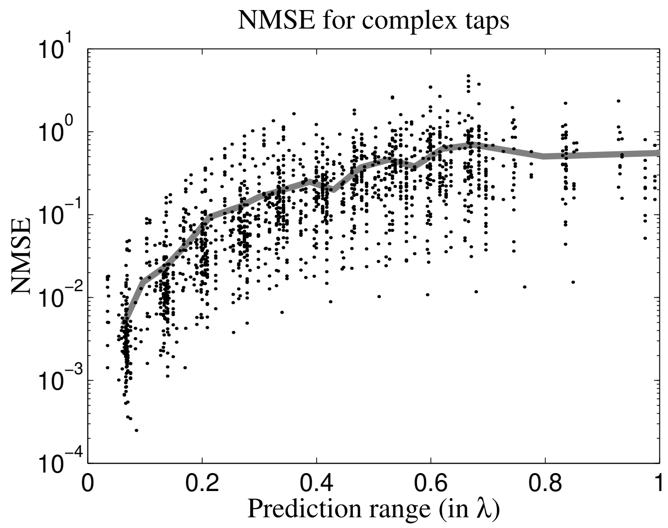

Figure 6.4: Scatter plot of the normalized means square error (NMSE) for the complex taps as a function of prediction range. The gray line is the local average NMSE, taken over 100 points.

of the taps, whereas the other taps act as noise. Here we take the taps accounting for 90% of the received power as our subset. In the following we investigate how well the total power of this subset is predicted. The predictions of the complex taps described in Subsection 6.1.1 are used to obtain estimates of the power for each channel. The adaptive predictor is applied to the data set of noise reduced 1420 samples of the data from each measurement. The performance is evaluated on the last half of the data, to avoid the influence of transient behavior.

The obtained results for the adaptive predictor are compared with the two simplest forms of prediction. The first approach is prediction using the last available sample of the power, thus assuming that the power remains constant after the latest observation. The second approach is prediction using the average power.

In Figure 6.5 the average NMSE for all the 25 channels, in the measurements, for the different methods are compared. The scatter plot is the NMSE for all the channels using the adaptive predictor. As was the case for the individual taps, the NMSE for the power prediction varies from channel to channel. On average the adaptive prediction of the power is better than using the average power as predictor, only up to a prediction range of less than half a wavelength. But as the individual spread is so large, the average

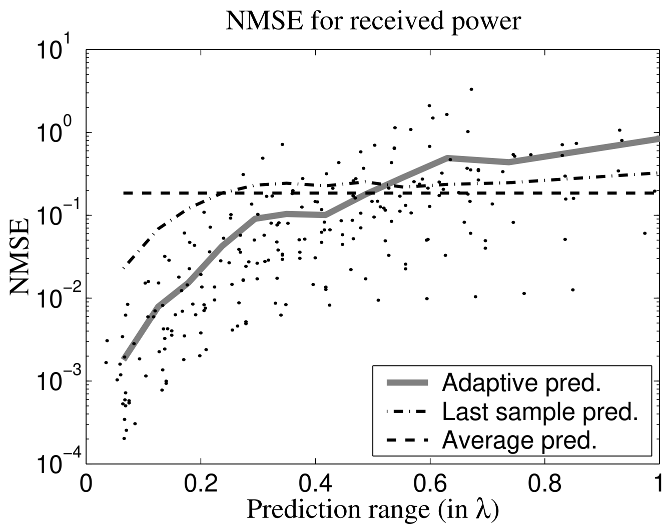

Figure 6.5: The average NMSE. The scatter plot is the NMSE for each channel and range from the adaptive predictor. The solid lines are averages using 30 points.

NMSE is an incomplete indicator.

If we measure the time the prediction is within a certain performance bound (in dB) we obtain a slightly different result. The power predicted by the adaptive predictors is within 1 dB of the power for 95% of the time for prediction ranges below 0.1λ, as seen in Figure 6.6. For half a wavelength it is within the 1 dB bound (that is within 26% from the true power) for 50% of the time. This should be compared to using the average power as predictor, that is within the 1 dB bound less than 40% of the time at all prediction ranges. At a vehicle speed of 90 km/h, 0.1λ corresponds to 0.6 ms whereas 0.5λ is 3.2 m/s. Thus for short range prediction of the power the adaptive predictor does well.

If the bound is expanded to 3 dB, as shown in Figure 6.7 the adaptive prediction is within the bound 95% of the time for ranges up to a quarter of a wavelength.

Due to the large spread in the results among the channels the average can be misleading. Instead, we may look at the number of channels, out of the 25 measurements, for which a given performance of the prediction is achieved. The result is collected in Table 6.3. The power could be predicted with an error of less than 1 dB for 99% of the time up to ranges of a quarter of a wavelength, in 9 out of the 25 measured channels. The corresponding

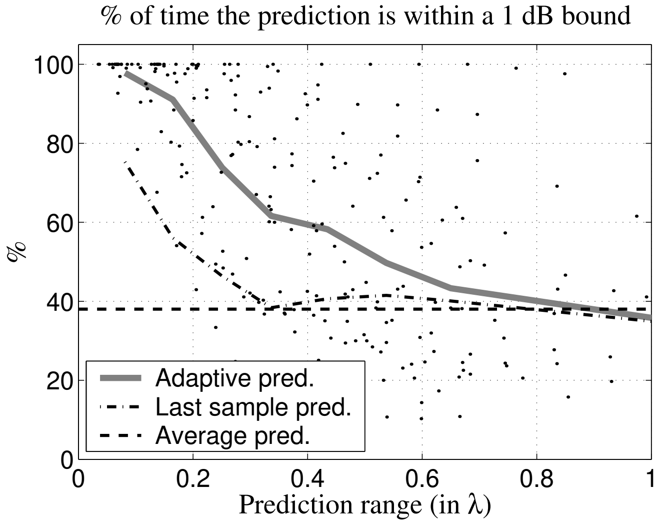

Figure 6.6: Scatter plot for the adaptive prediction of power, showing % of time within 1 dB of the true power. The lines indicates local averages over 30 points for different methods.

number for the 3 dB bound and half a wavelength is 8 channels. Thus, for at least 30% of the measurements at hand, the power can be accurately predicted at least up to a quarter of a wavelength ahead and fairly well up to half a wavelength.

		No. of channels within the bound this amount of time		
Range	Bound	90%	95%	99%
$\lambda/4$	1 dB	17 (68%)	14 (56%)	9 (36%)
$\lambda/2$	3 dB	16 (64%)	15 (60%)	8 (32%)

Table 6.3: Table for the number of channels for which the prediction of the power is within the 1 dB bound for a prediction range up to a quarter of a wavelength and within 3 dB up to half a wavelength, for 90, 95 and 99% of the time. The total number of channels is 25 and the percentage of channels is within parentheses.

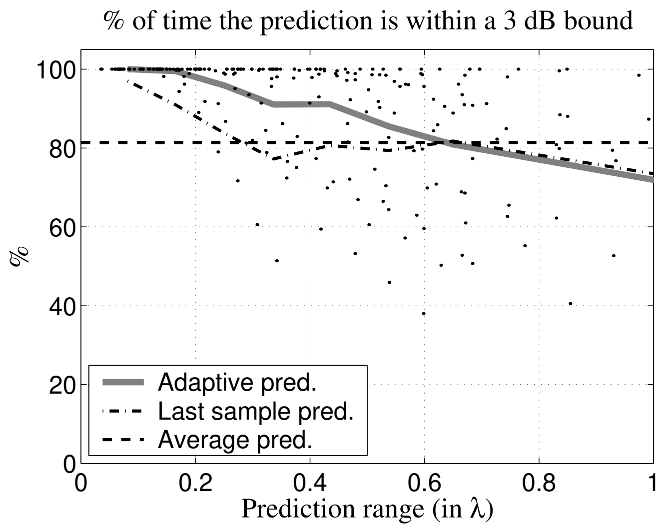

Figure 6.7: Scatter plot for the adaptive prediction of power, showing % of time within 3 dB of the true power. The lines indicates local averages over 30 points for different methods.

6.3 Conclusion

The linear adaptive iterated subsampled FIR predictor generally produces excellent predictions of both complex taps and total power for short ranges, that is up to 0.1 wavelengths. For some measurement the good results are valid up to ranges of at least a quarter of a wavelength. As was the case for the nonadaptive linear FIR predictors in Section 4.5, the prediction gain rapidly falls off when the prediction range increases.

At a carrier frequency of 1880 MHz the wavelength is 16 cm. In most of the measurements, acceptable prediction performance is achieved up to a prediction range of a quarter of a wavelength, which corresponds to 4 cm. At the moderate speed of 10 m/s, that is 36 km/h, this corresponds to a range of 4 ms. The performance of power predictors is reduced markedly at ranges over half a wavelengths, or over 8 ms. The advantage over using just the average power for prediction then becomes small, so we can not claim that the investigated adaptive predictors are efficient for these prediction ranges.

To develop better predictors is a topic for future research.

Chapter 7

Applications

In future packet based wireless communication systems, transmission in the downlink will often dominate the traffic load. High bit-rate applications like WWW-browsing, file transfer, and full motion video will impose strong requirements on the system capacity. An obstacle in this context is the time-variability of the channel: For mobile users, frequently occurring fading dips will cause unnecessary, and capacity degrading, retransmissions.

To achieve a high throughput also over fading channels, adaptive methods for adjustment of e.g. the modulation alphabet, and the coding complexity, can be used. The idea is to make efficient use of the bits: Whenever channel conditions are adequate, transmission of redundant bits should be avoided.

Fading channels confront us with the problem of lost packets and the need for frequent retransmissions. One strategy to combat time-variability is to use averaging: Spread-spectrum signalling can average out variations of the noise and interference level, while coding and interleaving can compensate for the temporary loss of signal strength due to fading dips. Such strategies can combat bad signalling conditions, but are inefficient when conditions are good. A complement to the averaging strategy would be to estimate the time-variations of the channel, due to short-term fading, and adapt the signalling scheme accordingly. We can exploit temporarily good transmission conditions to obtain higher throughput, while reducing the demands on the channel when its condition is bad. Assuming a system making use of either Frequency Division Duplex (FDD) or Time Division Duplex (TDD), with separate (ideal) control channels, the current channel parameters can be estimated and predictions about their future evolutions can be stored for subsequent transmission in the control channel. The bit-rate can be tailored to the current channel conditions by e.g. adjusting the modulation complexity, while keeping the transmitted symbol energy at a constant level. The further into the future the terminal can perform accurate predictions of the channel parameters, the more flexible and efficient the selection of the modulation alphabet will be. Moreover, the traffic on the control channel can be efficiently planned to minimize the signalling overhead. Such approaches have been proposed by Ericsson [14], Sampei, Goldsmith, and their co-writers in [13] and [12].

For a predicted value of the signal to noise ratio (SNR) of the channel, the modulation level is maximized under the constraint of a certain probability of symbol error, for example, $P_M \leq 10^{-5}$. If no modulation level attains the required probability of symbol error, the transmission is deferred until later when the SNR is higher, thus avoiding retransmissions. The reason for using this strategy is that it will stabilize the error probability, thus keeping the retransmission rate at a low and constant level. The averaging strategies mentioned above do not have this feature. On the contrary, they would yield a higher traffic load when conditions are bad, since the increasing error rate would increase the requests for retransmissions.

In a Time Division Multiple Access (TDMA) multi-user scenario it would be possible to schedule the transmissions time-slots to the mobiles with the momentary best receiving conditions. Transmission in deep fades would be avoided and system throughput could be maximized [14].

Chapter 8

Future Work

To achieve accurate prediction over long ranges we will study indirect adaption of the predictors, i.e., adaption of the channel models followed by calculation of the corresponding predictors. A time varying model, e.g. an AR model, of the channel variation can be estimated using an adaptive algorithm, as for the iterated predictor in this work. This model can then be used in a Kalman predictor to predict the channel coefficient. A similar approach, but for equalization of fast fading channels, was presented in [44]. An alternative approach is to use the estimated time varying model to design a Wiener predictor for the channel estimates.

For efficient adaptation of a model for the dynamics of a tap, low-complexity adaptive algorithms for AR processes are needed. To this end we will work on a generalization to AR processes of the tracker for MA processes presented in [42].

Further attention will be given to the problem of choosing the delays in the predictors, the length of the memory and the number of used samples. Possible links between these entities and e.g. Doppler spread and bandwidth will be examined.

The current work is solely focused on prediction of the channel in the time domain. An extension of this work to investigations of prediction of the frequency response of the channel will also be an interesting direction. Predictions of the frequency response can be used in e.g. OFDM systems.

A thorough study of the noise on the estimated channel will be performed. We will study how the time variability of the channel affect the estimation error in least squares estimation of the channel.

Appendix A

Visualized Channels

To further understand the mobile radio channel, visualization of a number of different aspects of the measured impulse responses is of great importance. In the following the estimated channels from three different measurements will be described and following properties will be illustrated.

Received power

The total received power changes as a function of time. When the power decrease rapidly, this is called a fade. The received power depends on the channel as

$$P(t) = \sum_{\tau} |h(\tau, t)|^2. \tag{A.1}$$

Frequency response

The instantaneous value of the time-varying channel has a frequency response. That is a discretized version of the frequency responses in equation (2.23). The fading is thus frequency selective. A fade in the received power corresponds to a situation with many frequencies or taps fading simultaneously.

$$H(\omega, t) = \mathcal{F}_{\tau}\{h(\tau, t)\}. \tag{A.2}$$

Power delay profile

The power delay profile is the power as a function of the delay τ of the impulse response, averaged over the time t for a time-varying response,

$$P(\tau) = \frac{1}{M} \sum_{t} |h(\tau, t)|^2. \tag{A.3}$$

The power delay profile shows how the received power is distributed over the delays. A peak in the PDP corresponds to a path with the corresponding delay τ.

Doppler spectra

The Fourier transform of the complex taps described in equation (2.25) gives the Doppler shift for the different contributing rays.

$$H(\tau, \Omega) = \mathcal{F}_t\{h(\tau, t)\} \tag{A.4}$$

The Doppler spectrum thus shows the frequency distribution of the oscillations of the individual taps. It is plotted as a power spectrum, $|H(\tau, \Omega)|^2$. To increase the resolution, a MASC spectral estimator [23] is used instead of the Fourier transform. The Doppler spectrum gives information about the direction of the incident waves. A peak in the Doppler spectrum close to f_D corresponds to a wave coming from the direction of motion. In the same way a peak close to $-f_D$ is a wave coming from behind the vehicle. Waves from the side correspond to peaks close to 0 Hz.

The key features of the three exemplified channels are listed in Table A.1.

	A	B	C
Velocity [m/s]	25	14	10
Delay spread [μs]	0.4	0.3	0.8
Doppler spread [Hz]	56	43	35
3 dB coherence bandwidth [Hz]	4.9	2.3	0.6

Table A.1: Table of the key features for the displayed channels.

Channel A

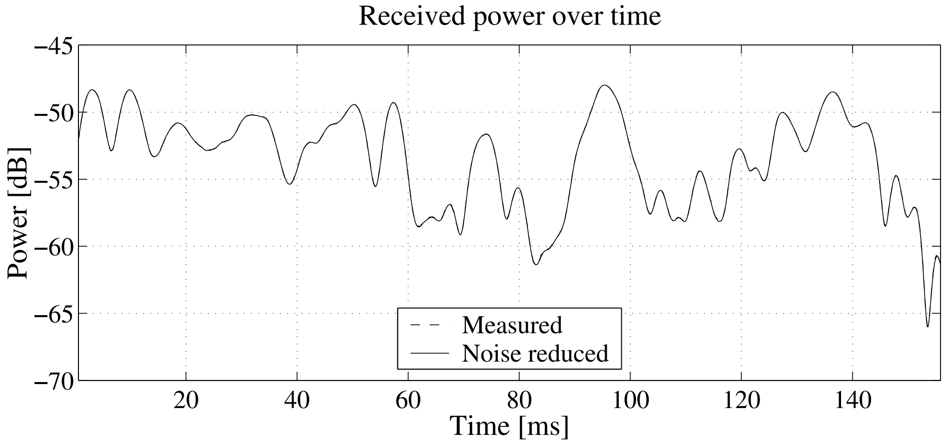

Figure A.1: The fading pattern of channel A changes a lot over time.

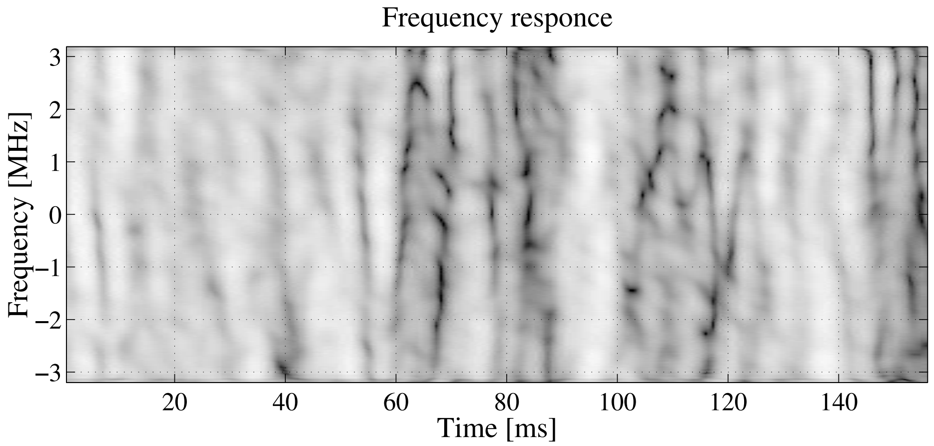

Figure A.2: Frequency response $|H(\omega, t)|$ of the estimated channel A. Dark coloring denotes fades. The dynamic range (black-white) is 40 dB. All frequencies fade simultaneously due to that mainly one tap contributes to the power.

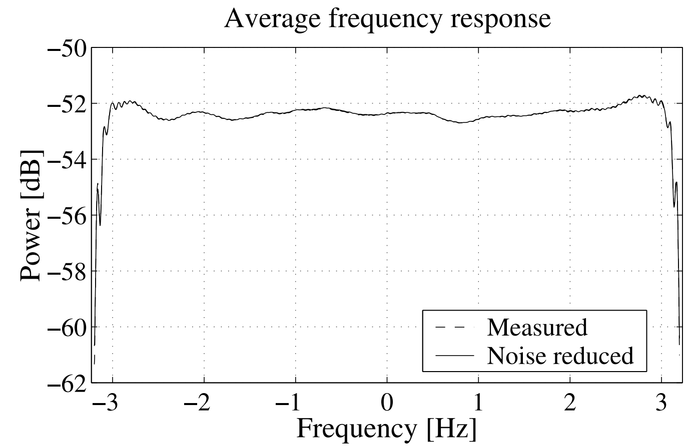

Figure A.3: The average frequency response of channel A is almost flat.

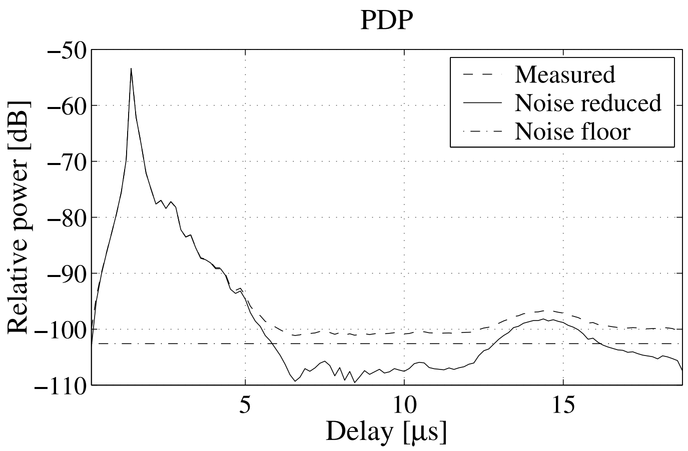

Figure A.4: The PDP of channel A shows that the power comes mainly from one tap. The noise floor is the power of the estimation error.

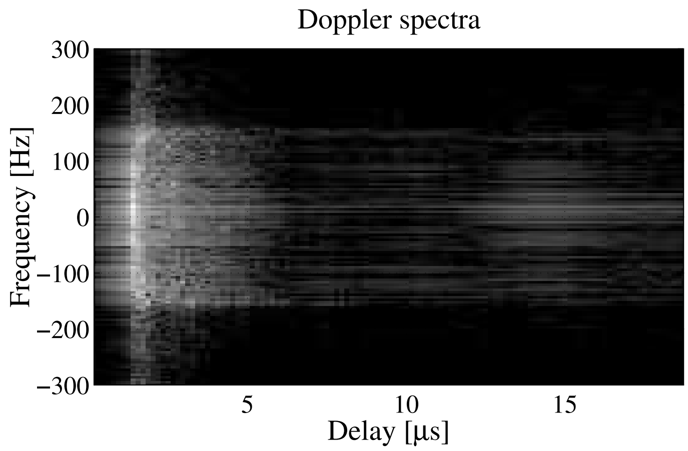

Figure A.5: The Doppler spectra for all the taps of channel A.

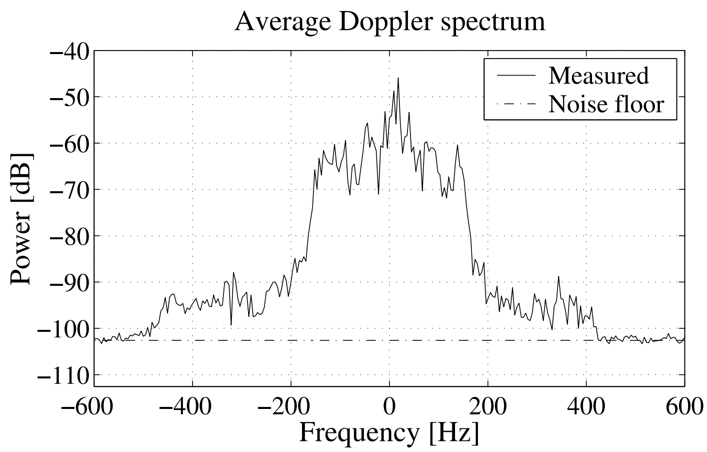

Figure A.6: Average Doppler spectrum of channel A. The strongest wavefronts come from the side as there are peaks close to 0 Hz. Note the small but significant power at frequencies above the Doppler frequency.

Channel B

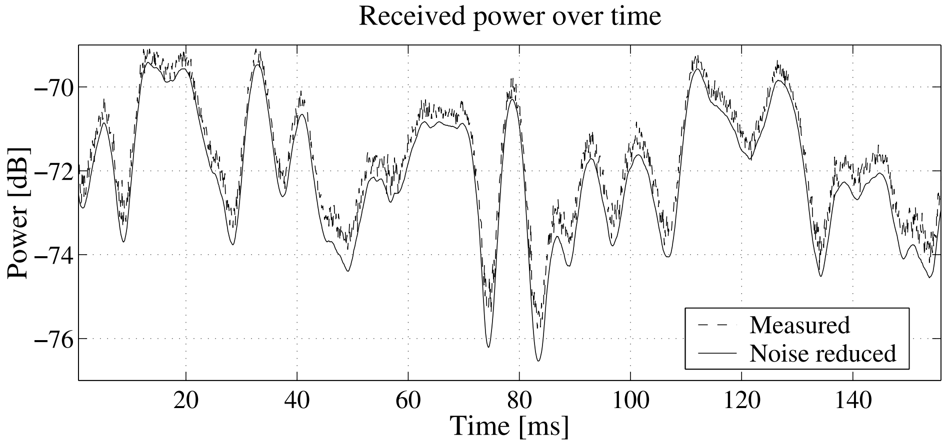

Figure A.7: The the estimated power of channel B. The power decrease when performing noise reduction.

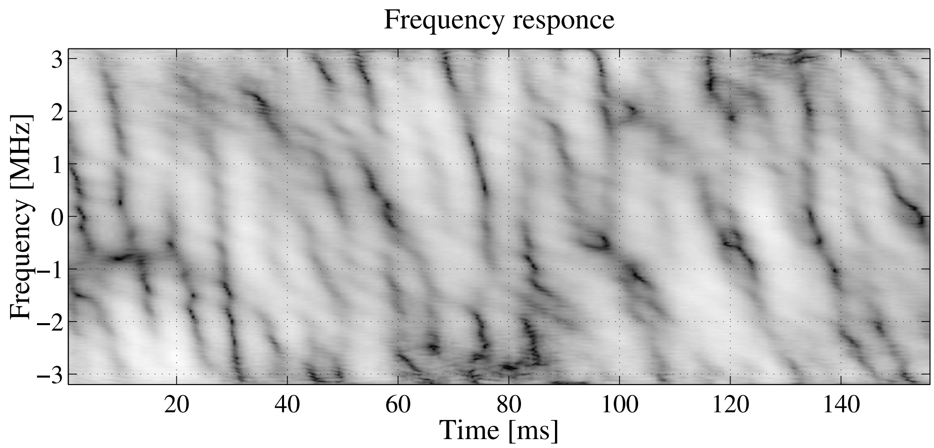

Figure A.8: Frequency response $|H(\omega,t)|$ of the estimated channel B. Dark coloring denotes fades and the dynamic range (black-white) is 40 dB.

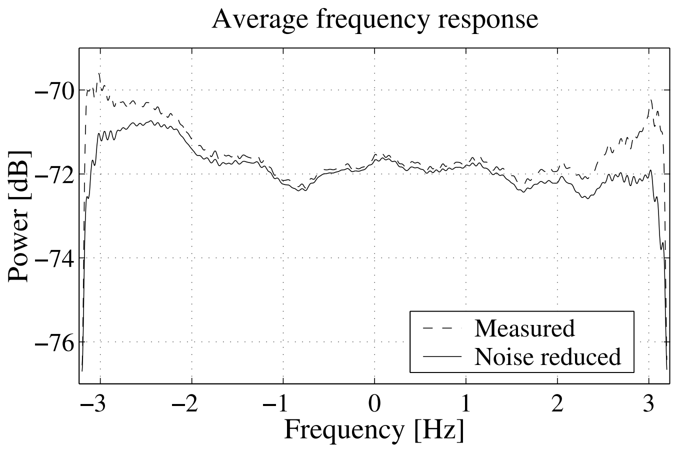

Figure A.9: The average frequency response of channel B is not flat. Noise reduction modifies the estimated frequency response especially at frequencies close to the band-limit.

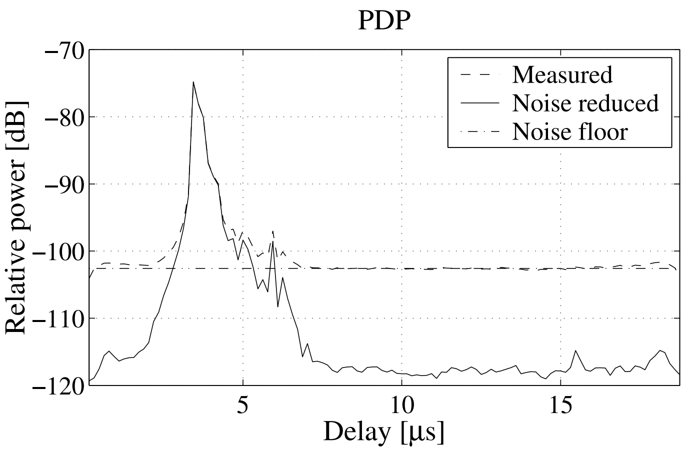

Figure A.10: The PDP of channel B shows that the power mainly comes from a few taps. Noise reduction is very efficient on this channel.

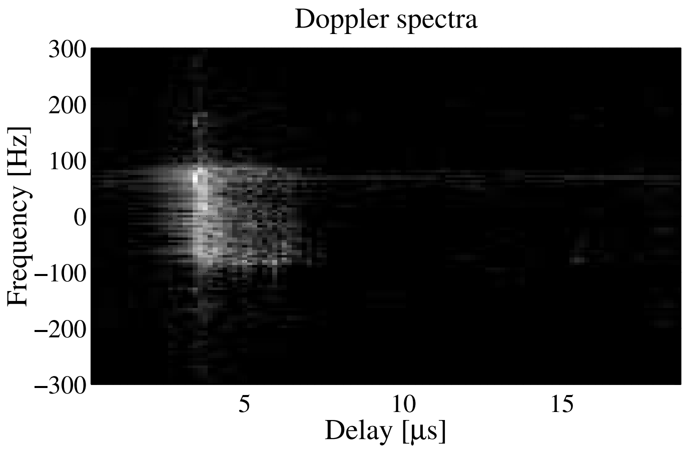

Figure A.11: The Doppler spectra for all the taps of channel B.

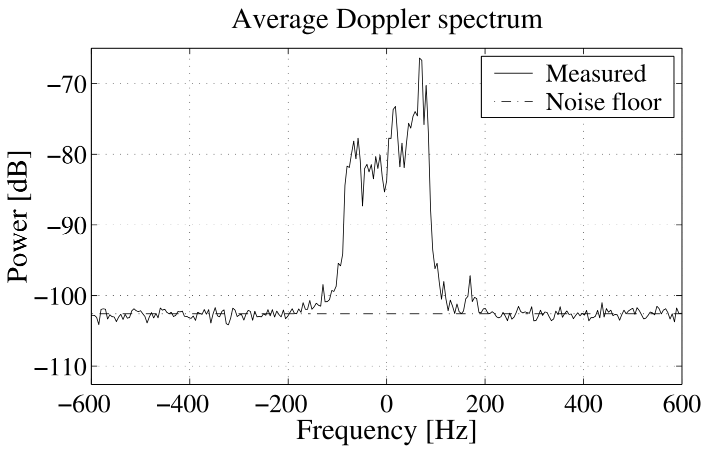

Figure A.12: Average Doppler spectrum of channel B. The strongest wavefronts come from the front.

Channel C

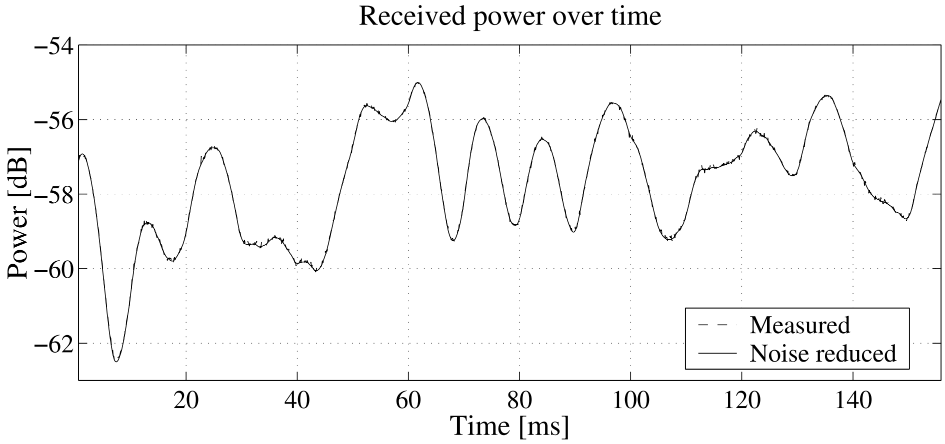

Figure A.13: The fades are not so deep in channel C due to the large delay spread.

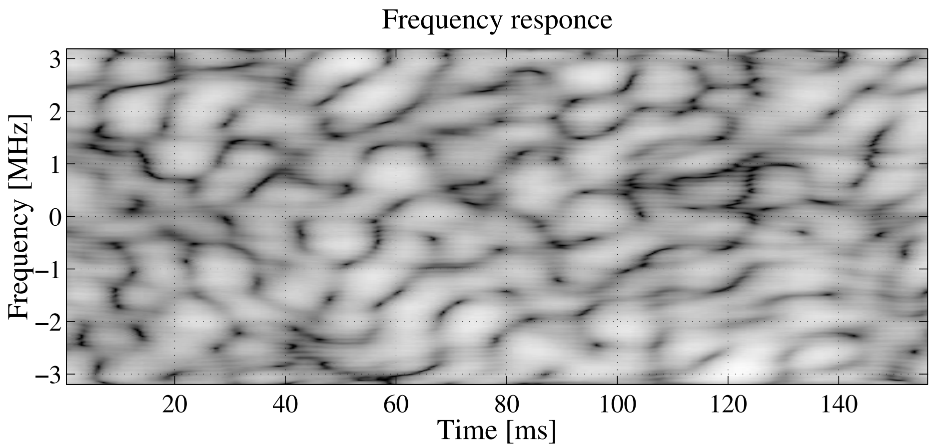

Figure A.14: Frequency response $|H(\omega, t)|$ of the estimated channel. Dark coloring denotes fades and the dynamic range (black-white) is 40 dB. The coherence bandwidth is small as the frequencies fade almost independent of each other.

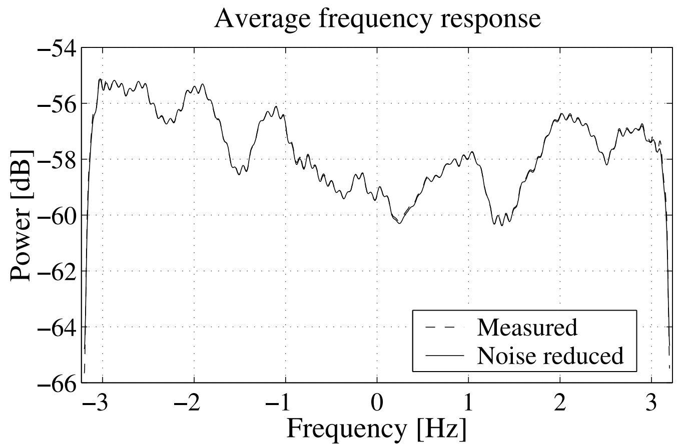

Figure A.15: The average frequency response of channel C.

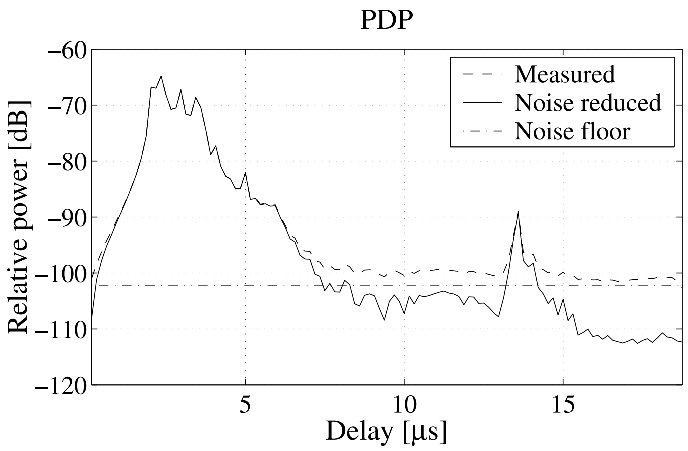

Figure A.16: The PDP of channel C shows that many taps contribute to the total power.

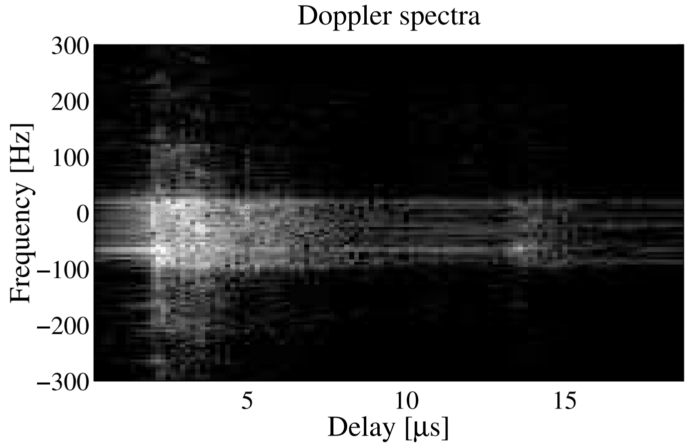

Figure A.17: The Doppler spectra for all the taps of channel C.

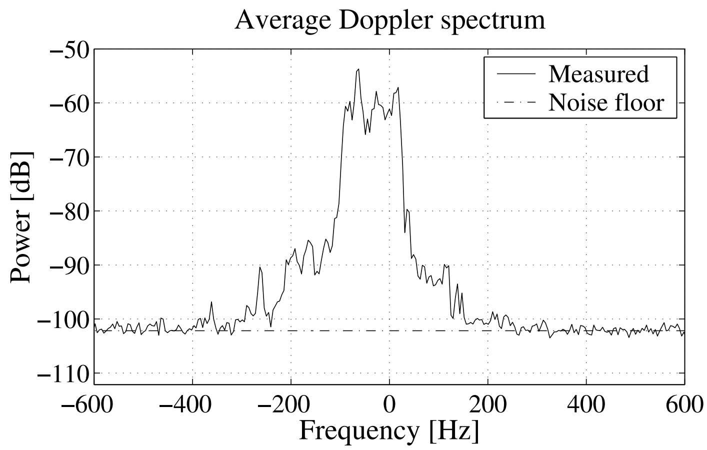

Figure A.18: Average Doppler spectrum of channel C.

Bibliography

[1] Jr W.C. Jakes, Multipath interference, In Jr W.C. Jakes, editor, *Microwave Mobile Communications*. IEEE, Piscataway, NY, 1974.

[2] R.H. Clarke, A statistical theory of mobile-radio reception, *The Bell System Technical Journal*, **47**(6):957–1000, July-August 1968.

[3] J.-K. Hwang and J. H. Winters, Sinusoidal modeling and prediction of fast fading processes, In *Proc. IEEE Globecom'98*, pages 892–897, Sidney, Nov 1998.

[4] J.B. Andersen, Prediction of future fading based on past measurements, In *Proc. IEEE* 50^{th} *Vehic. Tech. Conf. Fall*, pages 151–155, Amsterdam, Sep. 1999.

[5] A. Duel-Hallen T. Eycoz and H. Hallen, Deterministic channel modelling and long range prediction of fast fading radio channels, *IEEE Communications Letters*, **2**(9):254–256, Sept. 1998.

[6] R. Vaughan, P. Teal and R. Raich, Short-term mobile channel prediction using discrete scatterer propagation model and subspace signal processing algorithms, In *Proc. Vehic. Tech. Conf. Fall*, Boston, MA, Sep. 2000.

[7] A. Duel-Hallen, S. Hu and H. Hallen, Long-range prediction of fading signals, *IEEE Signal Processing Magazine*, **17**(3):62–75, May 2000.

[8] X.M. Gao, X.Z. Gao, J.M.A. Tanskanen and S.J. Ovaska, Power control for mobile DS/CDMA systems using a modified Elman neural network, In *Proc. IEEE* 47^{th} *Vehic. Tech. Conf.*, volume 2, pages 750–754, Phoenix, AZ, 1997.

[9] 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.*, pages 61–65, Atlanta, GA, May 1996.

[10] J.M. Friedman, Multivariate adaptive regression splines, *The Annals of Statistics*, **19**(1):1–141, 1991.

[11] T. Ekman and G. Kubin, Nonlinear prediction of mobile radio channels: Measurements and MARS model designs, In *Proc. Int. Conf. Acoust. Speech Sign. Process.*, Phoenix, AZ, March 1999.

[12] S. Chua and A. Goldsmith, Variable-rate variable-power MQAM for fading channels, In *Proc. IEEE* 46^{th} *Vehic. Tech. Conf.*, pages 815–819, May 1996.

[13] S. Sampei T. Ue and N. Morinaga, Adaptive modulation packet radio communication system using NP-CSMA/TDD scheme, In *Proc. IEEE* 46^{th} *Vehic. Tech. Conf.*, pages 416–420, May 1996.

[14] N.C. Ericsson, Adaptive modulation and schedueling of IP-trafic over fading channels, In *Proc. IEEE* 50^{th} *Vehic. Tech. Conf. Fall*, Amsterdam, Sep. 1999.

[15] J.G. Proakis, *Digital Communications*, McGraw–Hill Inc., New York, NY, second edition, 1989.

[16] M. Cedervall B.C. Ng and A. Paulraj, A structured channel estimator for maximum-likelihood sequence estimation, *IEEE Communications Letters*, **1**(2):52–55, March 1997.

[17] G.B. Giannakis and C. Tepedelenlioğlu, Basis expansion of models and diversity techniques for blind identification and equalization of time-varying channels, *Proc. IEEE*, **86**(10):1969–1998, Oct. 1998.

[18] A.F. Molisch J. Fuhl and E. Bonek, Unified channel models for mobile radio systems with smart antennas, *IEE Proc.-Radar, Sonar Navig.*, **145**(1):32–41, Feb. 1998.

[19] D. Kugiumtzis, State space reconstruction parameters in the analysis of chaotic time series - the role of the time window length, *Physica D*, **95**:13–28, 1996.

[20] T. Söderström and P. Stoica, *System Identification*, Prentice Hall International (UK) Ltd, Hertfordshire, UK, 1989.

[21] L. Ljung, *System Identification: Theory for the User*, Prentice Hall, NJ, second edition, 1999.

[22] R.R. Bitmead B. Wahlberg J. Homer, I. Mareels and F. Gustafsson, LMS estimation via structural detection, *IEEE Trans. on Signal Processing*, **46**(10):2651–2663, Oct. 1998.

[23] T. Ekman, A. Jakobsson and P. Stoica, On efficient implementation of the Capon algorithm, In *EUSIPCO2000*, Tampere, Suomi, September 2000.

[24] H. Asplund A.A. Glazunov and J.-E. Berg, Statistical analysis of measured short-term impulse response functions of 1.88Ghz radio channels in Stockholm with corresponding channel model, In *Proc. IEEE* 50^{th} *Vehic. Tech. Conf. Fall*, pages 107–111, Amsterdam, Sept 1999.

[25] A. Ahlén and M. Sternad, Wiener filter design using polynomial equations, *IEEE Trans. Signal Processing*, **39**:2387–2399, Nov. 1991.

[26] R.M. Gray, On the asymptotic eigenvalue distribution of Toeplitz matricse, *IEEE Trans. on Information Theory*, **18**(6):725–730, November 1972.

[27] D.C. Psichogios R.D. De Veaux and L.H. Ungar, A comparison of two nonparametric estimation schemes: Mars and neural networks, *Computers and Chem. Eng.*, **17**(8):819–837, Aug. 1993.

[28] H.-P. Bernhard, A tight upper bound on the gain of linear and nonlinear predictors for stationary stochastic processes, *IEEE Transactions on Signal Processing*, **46**(11):2909–2917, November 1998.

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

[30] H.P. Bernhard and G. Kubin, A fast mutual information calculation algorithm, In M.J.J. Holt et al., editor, *Signal Processing VII: Theories and Applications*, volume 1, pages 50–53. Elsevier, Amsterdam, September 1994.

[31] H-P. Bernhard and G.A. Darbellay, Performance analysis of the mutual information function for nonlinear and linear signal processing, In *Proc. Int. Conf. Acoust. Speech Sign. Process.*, Phoenix, AZ, March 1999.

[32] P. Stoica and R. Moses, *Introduction to Spectral Analysis*, Prentice Hall, Upper Saddle River, NJ, 1997.

[33] P. Händel, Predictive digital filtering of sinusoidal signals, *IEEE Trans. Signal Processing*, **46**:364–374, Feb. 1998.

[34] P.T. Harju, T.I. Laakso and S.J. Ovaska, Applying IIR predictors on Rayleigh fading signal, *Signal Processing*, **48**:91–96, 1996.

[35] Giovanni L. Sicuranza, Quadratic filters for signal processing, *Proc. IEEE*, **80**(8):1263–1285, August 1992.

[36] 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*, **86**(416):864–877, Dec. 1991.

[37] L. Breiman, J,H. Friedman, R.A. Olshen and C.J. Stone, *Classification and Regression Trees*, Wadsworth Inc., 1983.

[38] P. Steffen, On digital smoothing filters: A brief review of closed form solutions and two new filter approaches, *Circuits, Systems and Signal Processing*, **5**(2):187–210, 1986.

[39] B.D.O. Anderson and J.B. Moore, *Optimal Filtering*, Prentice Hall, Englewod Cliffs, NJ, 1979.

[40] A.H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, New York, 1970.

[41] K.J. Åström and B. Wittenmark, *Computer Controlled Systems: Theory and Design*, Prentice Hall, Inc., second edition, 1990.

[42] L. Lindbom, *A Wiener Filtering Approach to the Design of Tracking Algorithms*, PhD thesis, Uppsala University, Uppsala, Sweden, 1995.

[43] A. Papoulis, A note on the predictability of band-limited processes, *Preceedings of the IEEE*, **73**(8):1332–1333, August 1985.

[44] L.M. Davis, I.B. Collings and R.J. Evans, Coupled estimators for equalization of fast-fading mobile channels, *IEEE Trans. on Communications*, **46**(10):1262–1265, Oct. 1998.