
Preface

This book emanated from many discussions about collaborative research among the editors. The discussions have focussed on using signal processing methods for knowledge extraction and information fusion in a number of applications from telecommunications to renewable energy and biomedical engineering. They have led to several successful collaborative efforts in organizing special sessions for international conferences and special issues of international journals. With the growing interest from researchers in different disciplines and encouragement from Springer editors Alex Greene and Katie Stanne, we were spurred to produce this book.

Knowledge extraction and information fusion have long been studied in various areas of computer science and engineering, and the number of applications for this class of techniques has been steadily growing. Features and other parameters that describe a process under consideration may be extracted directly from the data, and so it is natural to ask whether we can exploit digital signal processing (DSP) techniques for this purpose. Problems where noise, uncertainty, and complexity play major roles are naturally matched to DSP. This synergy of knowledge extraction and DSP is still under-explored, but has tremendous potential. It is the underlying theme of this book, which brings together the latest research in DSP-based knowledge extraction and information fusion, and proposes new directions for future research and applications. It is fitting, then, that this book touches on globally important applications, including sustainability (renewable energy), health care (understanding and interpreting biomedical signals) and communications (extraction and fusing of information from sensor networks).

The use of signal processing in data and sensor fusion is a rapidly growing research area, and we believe it will benefit from a work such as this, in which both background material and novel applications are presented. Some of the chapters come from extended papers originally presented at the special sessions in ICANN 2005 and KES 2006. We also asked active researchers in signal processing with specializations in machine learning and multimodal signal processing to make contributions to augment the scope of the book.

This book is divided in four parts with four chapters each.

Collaborative Signal Processing Algorithms

Chapter 1 by Jelfs et al. addresses hybrid adaptive filtering for signal modality characterization of real-world processes. This is achieved within a collaborative signal processing framework which quantifies in real-time, the presence of linearity and nonlinearity within a signal, with applications to the analysis of EEG data. This approach is then extended to the complex domain and the degree of nonlinearity in real-world wind measurements is assessed.

In Chap. 2, Hirata et al. extend the wind modelling approaches to address the control of wind farms. They provide an analysis of the wind features which are most relevant to the local forecasting of the wind profile. These are used as prior knowledge to enhance the forecasting model, which is then applied to the yaw control of a wind turbine.

A collaborative signal processing framework by means of hierarchical adaptive filters for the detection of sparseness in a system identification setting is presented in Chap. 3, by Boukis and Constantinides. This is supported by a thorough analysis with an emphasis on unbiasedness. It is shown that the unbiased solution corresponds to existence of a sparse sub-channel, and applications of this property are highlighted.

Chapter 4 by Zhang and Chambers addresses the estimation of the reverberation time, a difficult and important problem in room acoustics. This is achieved by blind source separation and adaptive noise cancellation, which in combination with the maximum likelihood principle yields excellent results in a simulated high noise environment. Applications and further developments of this strategy are discussed.

Signal Processing for Source Localization

Kuh and Zhu address the problem of sensor network localization in Chap. 5. Kernel methods are used to store signal strength information, and complex least squares kernel regression methods are employed to train the parameters for the support vector machine (SVM). The SVM is then used to estimate locations of sensors, and to track positions of mobile sensors. The chapter concludes by discussing distributed kernel regression methods to perform localization while saving on communication and energy costs.

Chapter 6, by Lenz et al., considers adaptive localization in wireless networks. They introduce an adaptive approach for simultaneous localization and learning based on theoretical propagation models and self-organizing maps, to demonstrate that it is possible to realize a self-calibrating positioning system with high accuracies. Results on real-world DECT and WLAN groups support the approach.

In Chap. 7, Host-Madsen et al. address signal processing methods for Doppler radar heart rate monitoring. This provides unobtrusive and ubiquitous detection of heart and respiration activity from distance. By leveraging recent advances in signal processing and wireless communication technologies, the authors explore robust radar monitoring techniques through MIMO signal processing. The applications of this method include health monitoring and surveillance.

Obradovic et al. present the fusion of onboard sensors and GPS for real-world car navigation in Chap. 8. The system is based on the position estimate obtained by Kalman filtering and GPS, and is aided by corrections provided by candidate trajectories on a digital map. In addition, fuzzy logic is applied to enhance guidance. This system is in operation in a number of car manufacturers.

Information Fusion in Imaging

In Chap. 9, Chumerin and Van Hulle consider the detection of independently moving objects as a component of the obstacle detection problem. They show that the fusion of information obtained from multiple heterogeneous sensors has the potential to outperform the vision-only description of driving scenes. In addition, the authors provide a high-level sensor fusion model for detection, classification, and tracking in this context.

Aghajan, Wu, and Kleihorst address distributed vision networks for human pose analysis in Chap. 10. This is achieved by collaborative processing and data fusion mechanisms, and under a low bandwidth communication constraint. The authors employ a 3D human body model as the convergence point of the spatiotemporal and feature fusion. This model also allows the cameras to interact and helps the evaluation of the relative values of the derived features.

The application of information fusion in E-cosmetics is addressed by Tsumura et al. in Chap. 11. The authors develop a practical skin color analysis and synthesis (fusion) technique which builds upon both the physical background and physiological understanding. The appearance of the reproduced skin features is analysed with respect to a number of practical constraints, including the imaging devices, illuminants, and environments.

Calhoun and Adalı consider the fusion of brain imaging data in Chap. 12. They utilize multiple image types to take advantage of the cross information. Unlike the standard approaches, where cross information is not taken into account, this approach is capable of detecting changes in functional magnetic resonance imaging (fMRI) activation maps. The benefits of the information fusion strategy are illustrated by real-world examples from neurophysiology.

Knowledge Extraction in Brain Science

Chapter 13, by Mandic et al. considers the “data fusion via fission” approach realized by empirical mode decomposition (EMD). Extension to the complex

domain also helps to extract knowledge from processes which are strongly dependent on synchronization and phase alignment. Applications in real-world brain computer interfaces, e.g., in brain prosthetics and EEG artifact removal, illustrate the usefulness of this approach.

In Chap. 14, Rutkowski et al. consider some perceptual aspects of the fusion of information from multichannel EEG recordings. Time–frequency EMD features, together with the use of music theory, allow for a convenient and unique audio feedback in brain computer and brain machine (BCI/BMI) interfaces. This helps to ease the understanding of the notoriously difficult to analyse EEG data.

Cao and Chen consider the usefulness of knowledge extraction in brain death monitoring applications in Chap. 15. They combine robust principal factor analysis with independent component analysis to evaluate the statistical significance of the differences in EEG responses between quasi-brain-death and coma patients. The knowledge extraction principles here help to make a binary decision on the state of the consciousness of the patients.

Chapter 16, by Golz and Sommer, addresses a multimodal approach to the detection of extreme fatigue in car drivers. The signal processing framework is based on the fusion of linear (power spectrum) and nonlinear (delay vector variance) features, and knowledge extraction is performed via automatic input variable relevance detection. The analysis is supported by results from comprehensive experiments with a range of subjects.

London,
October 2007

Danilo Mandic
Martin Golz
Anthony Kuh
Dragan Obradovic
Toshihisa Tanaka

Wind Modelling and its Possible Application to Control of Wind Farms

Yoshito Hirata, Hideyuki Suzuki, and Kazuyuki Aihara

Because of global warming and oil depletion, the number of wind turbines is increasing. Wind turbines commonly have a horizontal axis, but some wind turbines have a vertical axis. A problem of wind turbines with horizontal axis is that we need to face them towards the wind for maximising energy production. If the geography around a wind turbine is complicated, then the wind direction keeps changing. Thus, if one can predict the wind direction to some extent, then one may generate more electricity by controlling wind turbines according to the prediction.

In this chapter, we discuss how to model the wind. First, we formulate a problem and clarify which properties of the wind we need to predict. Second, we discuss the characteristics for time series of the wind. Third, we model the wind based on the knowledge and predict it. We prepare different models for predicting wind direction and absolute wind speed. Finally, we apply the prediction and simulate control of a wind turbine. Since we integrate predictions for wind direction and absolute wind speed to obtain an optimal control, the whole scheme can be regarded as heterogeneous fusion.

2.1 Formulating Yaw Control for a Wind Turbine

There are three purposes for predicting the wind in applications of wind turbine. The first purpose is to estimate energy production by wind turbines in the future. By estimating the energy production, we can prepare other sources of energy production to meet the demand. The time scale of the prediction for this purpose ranges from 30 min to some days and the prediction can be done by either a medium-range weather forecasts [20] or time series prediction [2–4]. The second purpose is to avoid gusts. For this purpose, the time scale of prediction is on the order of seconds. A Markov chain [12–14, 18] is effective for the prediction of this purpose. The third purpose is to generate more electricity. On this chapter, we focus on this third purpose.

There are two types of wind turbines. The major one has a horizontal axis, while the minor one has a vertical axis. The horizontal axis is more popular because it can attain better efficacy.

There are several factors one may control in a wind turbine with a horizontal axis. The examples include the direction of wind turbine (yaw angle), pitch angles, and speeds of rotations. Although mechanical details could be different, the problems for controlling these factors can be written mathematically in a similar way. As an example, we formulate a problem of yaw control.

In yaw control, the net benefits we can get are divided into two parts: the first part is benefits we may obtain; the second part is costs for controls. The unit of the net benefits is a currency: yen, US dollars, UK pounds, or other currencies. The benefits we may obtain from the energy production correlates with the amount of energy production. Thus, they are shown by using a power curve, a function $f(u)$ that returns the amount of energy produced given a wind speed u . Here the wind speed is perpendicular to the face of the rotor. Let v_t be the absolute wind speed at time t , θ_t , the wind direction, and $\varphi(t)$, the direction of the wind turbine. We also denote a set of v_t , θ_t , $\varphi(t)$ by v , θ and φ , respectively. Then, the benefits $b(\varphi, v, \theta)$ from time k to $k + K$ can be written as

$$\sum_{t=k+1}^{k+K} f(v_t \cos(\theta_t - \varphi(t))). \quad (2.1)$$

Since we have freedom for choosing the scale of the currency, we scale the currency in a way that the maximum electricity generated within 2s yields a unit of the currency.

The costs for the yaw control include the electricity we need, replacements of devices due to stress, fatigue, and friction, and others coming from unknown sources. The term of the costs is a function of the orbit φ , wind speed v , and wind direction θ . Therefore, we denote it by $c(\varphi, v, \theta)$. In the simulations shown later, we assume that $c(\varphi, v, \theta)$ can be written as

$$\begin{aligned} & C_1 \sum_{t=k+1}^{k+K} |\varphi(t) - \varphi(t-1)| \\ & + C_2 \sum_{t=k+1}^{k+K} |(\varphi(t) - \varphi(t-1)) - (\varphi(t-1) - \varphi(t-2))| \\ & + C_3 \sum_{t=k+1}^{k+K} |\varphi(t) - \theta_t|. \end{aligned} \quad (2.2)$$

The first term depends on how much we turn a wind turbine. The second term depends on how much we change the angular velocity of wind turbine. The third term depends on the difference between the directions of the wind and the wind turbine.

The net benefits can be written as $b(\varphi, v, \theta) - c(\varphi, v, \theta)$. By replacing v_t and θ_t with their predictions \hat{v}_t and $\hat{\theta}_t$, a problem of yaw control can be written as

$$\min_{\varphi} b(\varphi, \hat{v}, \hat{\theta}) - c(\varphi, \hat{v}, \hat{\theta}). \quad (2.3)$$

The above problem means that for yaw control, we need to predict the wind direction and absolute wind speed for the duration of K .

2.2 Characteristics for Time Series of the Wind

To predict the future wind, we need to know its properties. In this section, we discuss the characteristics of the wind.

2.2.1 Surrogate Data

To characterise the data sets, we used surrogate data analysis [21–25]. Surrogate data analysis is hypothesis testing. At the beginning of the analysis, one decides a null-hypothesis. Then we generate a set of random data that are consistent with the null-hypothesis. These random data are used for obtaining a confidence interval of a test statistic. If the value of the test statistic obtained from the original data is within the confidence interval, then one cannot reject the null-hypothesis. Otherwise, one rejects the null-hypothesis.

Common null-hypotheses are non-serial dependence [21], linearity [22, 25], and periodic orbits with uncorrelated noise [24]. In this chapter, we use the first two since the third surrogate has meaning when a time series is pseudo-periodic.

2.2.2 Results

In this section, we used a data set of the wind measured on 25 August 2004 at the top of a building in Institute of Industrial Science, The University of Tokyo. The measurement was of 50 Hz and it has three components, namely, the east wind, the north wind, and the upward wind. Part of data set is shown in Fig. 2.1.

First we preprocessed the data set. We divided a data set of the east and north winds into 36 segments of scalar time series of length 10,000. Then we matched the beginning and the end of time series [23] to avoid artefacts their mismatch may cause. By applying the method of Kennel [15] with four-dimensional embedding space with a uniform delay of 2 s, we confirmed the stationarity for 10 segments out of 36 segments. We used the delay of 2 s because this is a time scale we are interested in the later part of this chapter. We used the four-dimensional embedding space because in most cases we tested, the false nearest neighbour method [16] showed that with four-dimensional embedding space, the ratio of the false nearest neighbours is less

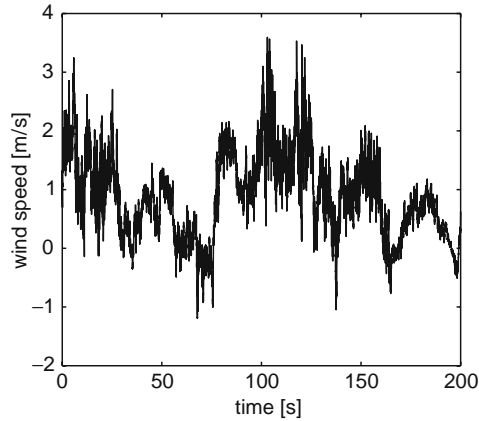


Fig. 2.1. Part of data set used in surrogate data analysis

than 1%. We had to exclude nonstationary segments since nonstationarity might cause a spurious rejection [26].

Second, we used random shuffle surrogates [21] for testing serial dependence. We used the Wayland statistic [27] for the test statistic. We used the embedding dimensions between 1 and 12. We generated 39 random shuffle surrogates and thus the significant level for each embedding dimension was 5%. To make the significant level of the multiple tests 1%, we need to have more than or equal to three rejections out of 12 embedding dimensions. We found that all the ten stationary segments have serial dependence. But a peculiar feature is that the Wayland statistic for the original data is greater than that for the random shuffle surrogates [5] (see Fig. 2.2). This may be interpreted

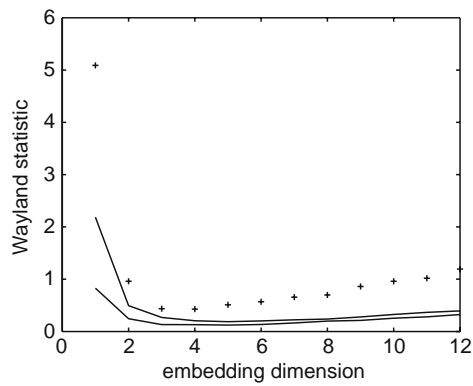


Fig. 2.2. Surrogate test for serial dependence. The *crosses* show the values obtained from the actual data set and the *solid lines*, the minimum and maximum for 39 random shuffle surrogates

that the original data are less deterministic than the random shuffle surrogates while this interpretation should be wrong. After testing some models, we found that this strange phenomenon could happen if an original data set is contaminated with either observational noise or dynamical noise [5]. The common characteristic between these two possible causes is that a time series has small fluctuations around trends. Thus, the wind may be considered as small fluctuations with long-term trends.

Third, we tested the nonlinearity. We generated iterative amplitude adjusted Fourier transform surrogates [22] and compared the original data sets and their surrogates with the Wayland statistic of the original data. We found a rejection in a multiple test for three segments out of 10 stationary segments. The number of rejections is significantly higher than the chance level. Thus the wind may be assumed to be a nonlinear process.

2.3 Modelling and Predicting the Wind

In the previous section, we discussed that a nonlinear model is appropriate for modelling the wind. In this section, we build a model and evaluate its performance by a prediction error.

2.3.1 Multivariate Embedding

In most cases, nonlinear time series analysis starts with delay embedding. Let $x_i(t)$ be the i th observable at time t . Let $x(t)$ represent whole the state at time t . Then, we choose

$$(x_i(t), x_i(t - \tau), x_i(t - 2\tau), \dots, x_i(t - (\kappa - 1)\tau)). \quad (2.4)$$

If κ is sufficiently large, then this vector gives an embedding, meaning that the vector and $x(t)$ is one-to-one and the associated tangent space at each point is also one-to-one. For the comparison later, we call this vector uniform embedding.

In practice, Judd and Mees [10] extended the notion of delay embedding and defined the following vector:

$$(x_i(t - \tau_1), x_i(t - \tau_2), \dots, x_i(t - \tau_\kappa)). \quad (2.5)$$

They call this vector non-uniform embedding. The method of Judd and Mees [10] fits a linear model and chooses the vector that minimises description length, an information criterion.

The non-uniform embedding was extended to multivariate time series by Garcia and Almeida [1]. Namely, one tries to construct a vector which looks like

$$(x_{i_1}(t - \tau_1), x_{i_2}(t - \tau_2), \dots, x_{i_\kappa}(t - \tau_\kappa)). \quad (2.6)$$

Their method is based on the extension of false nearest neighbour method [16], a standard method for estimating the embedding dimension. In this chapter, we used the extension [8] of Judd and Mees [10], another method for obtaining a non-uniform embedding from a multivariate time series, since the method gives, within a shorter time, a non-uniform embedding which is as reasonable as the method of Garcia and Almeida.

2.3.2 Radial Basis Functions

The model we built was a radial basis function model since it can approximate any continuous function relatively well. We use $x^i(t)$ for showing a reconstruction for the i th coordinate at time t . Then a radial basis function model can be written as

$$x_i(t+1) = \delta_i + \alpha_i \cdot x^i(t) + \sum_l \beta_{i,l} \exp \left[-\frac{\|x^i(t) - \gamma_{i,l}\|^2}{2\nu^2} \right], \quad (2.7)$$

where δ_i , α_i , $\beta_{i,l}$, $\gamma_{i,l}$, and ν are parameters for the model. The centres of the radial basis functions were chosen using the technique of chaperons [9]. In this technique, one chooses a set of points from observations and adds Gaussian noise to them, whose standard deviation is 30% of that of the data. We prepared 100 radial basis functions using this technique. We also set ν to the standard deviation of the time series.

Since we have already decided $\gamma_{i,k}$ and ν , now our model is a pseudo-linear model and the remaining parameters may be decided by the least squares solution. But a simple application of the least squares solution will face the problem of overfitting. To avoid overfitting, we decided the remaining parameters for the radial basis function model as prescribed by Judd and Mees [9] using the Normalised Maximum Likelihood [19] for a model selection criterion. To explain the details, first let us define the Normalised Maximum Likelihood.

First we need to define a set of basis functions. In this chapter, the set is made of the 100 radial basis functions and linear terms. Suppose that now we have a number P of basis functions $\{\zeta_j : j \in F\}$, where $F = \{1, 2, \dots, P\}$ is the set of indices for all the basis functions. Let D be the biggest delay among $\{\tau_i\}$. The length of given time series is denoted by N . Define $\tilde{N} = N - D$. Let V be the $\tilde{N} \times P$ matrix whose (q, r) component is the value for the r th basis function at time q , i.e.,

$$V = \begin{pmatrix} \zeta_1(x^i(D)) & \zeta_2(x^i(D)) & \cdots & \zeta_P(x^i(D)) \\ \vdots & & & \vdots \\ \zeta_1(x^i(t)) & \zeta_2(x^i(t)) & \cdots & \zeta_P(x^i(t)) \\ \vdots & & & \vdots \\ \zeta_1(x^i(N-1)) & \zeta_2(x^i(N-1)) & \cdots & \zeta_P(x^i(N-1)) \end{pmatrix}. \quad (2.8)$$

Denote a set of indices for basis functions by B . Thus, it holds that $B \subset F$. Let us denote by V_B the matrix that is formed from the columns of V with indices in $B = \{j_1, \dots, j_J\}$, i.e.,

$$V_B = \begin{pmatrix} \zeta_{j_1}(x^i(D)) & \zeta_{j_2}(x^i(D)) & \cdots & \zeta_{j_J}(x^i(D)) \\ \vdots & \vdots & & \vdots \\ \zeta_{j_1}(x^i(t)) & \zeta_{j_2}(x^i(t)) & \cdots & \zeta_{j_J}(x^i(t)) \\ \vdots & \vdots & & \vdots \\ \zeta_{j_1}(x^i(N-1)) & \zeta_{j_2}(x^i(N-1)) & \cdots & \zeta_{j_J}(x^i(N-1)) \end{pmatrix}. \quad (2.9)$$

Let us also define ξ as

$$\xi = (x_i(D+1), \dots, x_i(t+1), \dots, x_i(N))^T. \quad (2.10)$$

Here ‘T’ shows the transposition of matrix.

Let λ_B be the least squares solution for $\xi \approx V_B \lambda_B$. Then the prediction error e_B can be written as $e_B = \xi - V_B \lambda_B$. When minimising the squared error $e_B^T e_B$, we can find the solution by

$$\lambda_B = (V_B^T V_B)^{-1} V_B^T \xi. \quad (2.11)$$

Define R by

$$R = \frac{1}{\tilde{N}} (V_B \lambda_B)^T V_B \lambda_B. \quad (2.12)$$

Then the Normalised Maximum Likelihood [19] for a set of p basis functions B is given by

$$\mathcal{L}(B) = \frac{\tilde{N} - p}{2} \ln \frac{e_B^T e_B}{\tilde{N}} + \frac{p}{2} \ln R - \ln \Gamma \left(\frac{\tilde{N} - p}{2} \right) - \ln \Gamma \left(\frac{p}{2} \right) - \ln p. \quad (2.13)$$

Here Γ shows the Gamma function.

Using the Normalised Maximum Likelihood and these notations, we can write down the algorithm proposed in [9] for selecting an optimal set of basis functions in the following way:

1. Normalise V so that each column has unit length.
2. Let B and B' be empty sets.
3. Obtain the prediction error. If B' is empty, then $e_{B'} = \xi$. If not, then $e_{B'} = \xi - V_{B'} \lambda_{B'}$. Here $\lambda_{B'}$ can be obtained using the formula of (2.11).
4. Find the basis function that matches the prediction error best. Let $\mu = V^T e_{B'}$. Then the biggest component of μ corresponds to the best basis function matching to the prediction error. Let g be the index for the best matching basis function. Let $B' \leftarrow B' \cup \{g\}$.
5. Find the basis function in B' that least contributes to making the error small. Let h be the index for the basis function whose corresponding $\lambda_{B'}$ is the smallest. If $g \neq h$, then $B' \leftarrow B' \setminus \{h\}$ and go to Step 3.

6. If B is empty or $\mathcal{L}(B') < \mathcal{L}(B)$, then set $B \leftarrow B'$ and go back to Step 3.
7. The chosen B is the optimal set of basis functions.

We used the above algorithm for deciding δ_i , α_i , and $\beta_{i,l}$. Namely, for basis functions which are not selected, we set the corresponding components to 0. For selected basis functions, we use values obtained from the above fitting.

2.3.3 Possible Coordinate Systems

There are two possible coordinate systems for representing a wind vector. The first one is the polar coordinate system. The second one is the rectangular coordinate system. The problem of the polar coordinate system is how to represent the wind direction. The wind direction has a value between 0° and 360° , which is on a ring. If one tries to show the wind direction using a value in an interval, then one has to split the ring somewhere and it causes discontinuity. As a result, the wind direction cannot be predicted well [6]. Thus, the rectangular coordinate system is better than the polar coordinate system for predicting the wind direction.

2.3.4 Direct vs. Iterative Methods

The model now at our hands predicts a step ahead. However, as discussed in Sect. 2.1, we need to predict the wind direction and absolute wind speed for a certain range of time. There are two options for it: The first option is to prepare several models each of which has a different prediction step. This approach is called a direct method. The second option is to use a model for predicting a step ahead several times and realise a multiple step prediction. This approach is called an iterative method.

There are some discussions on which model is better on a certain occasion [11, 17]. Generally speaking, when we have a good one-step prediction, then an iterative method is better. Otherwise, a direct method might be better. Judd and Small [11] proposed a mixture of direct and iterative methods. In their method, one first uses an iterative method and predicts the future values. This step is called ϕ -step since the one-step model is shown using ϕ . Then one uses a corrector ψ that makes the prediction more accurate. This step is called ψ -step due to the notation of the corrector. As a total, the method is called the $\psi\phi$ method.

2.3.5 Measurements of the Wind

The data sets we used were measured at about 1 m high on the campus of the Institute of Industrial Science, The University of Tokyo. We used two identical anemometers which record the east, the north, and the upward winds with 50 Hz. We measured the wind for nine different days between September 2005 and May 2006. On some days, we placed the two anemometers 5 m apart in

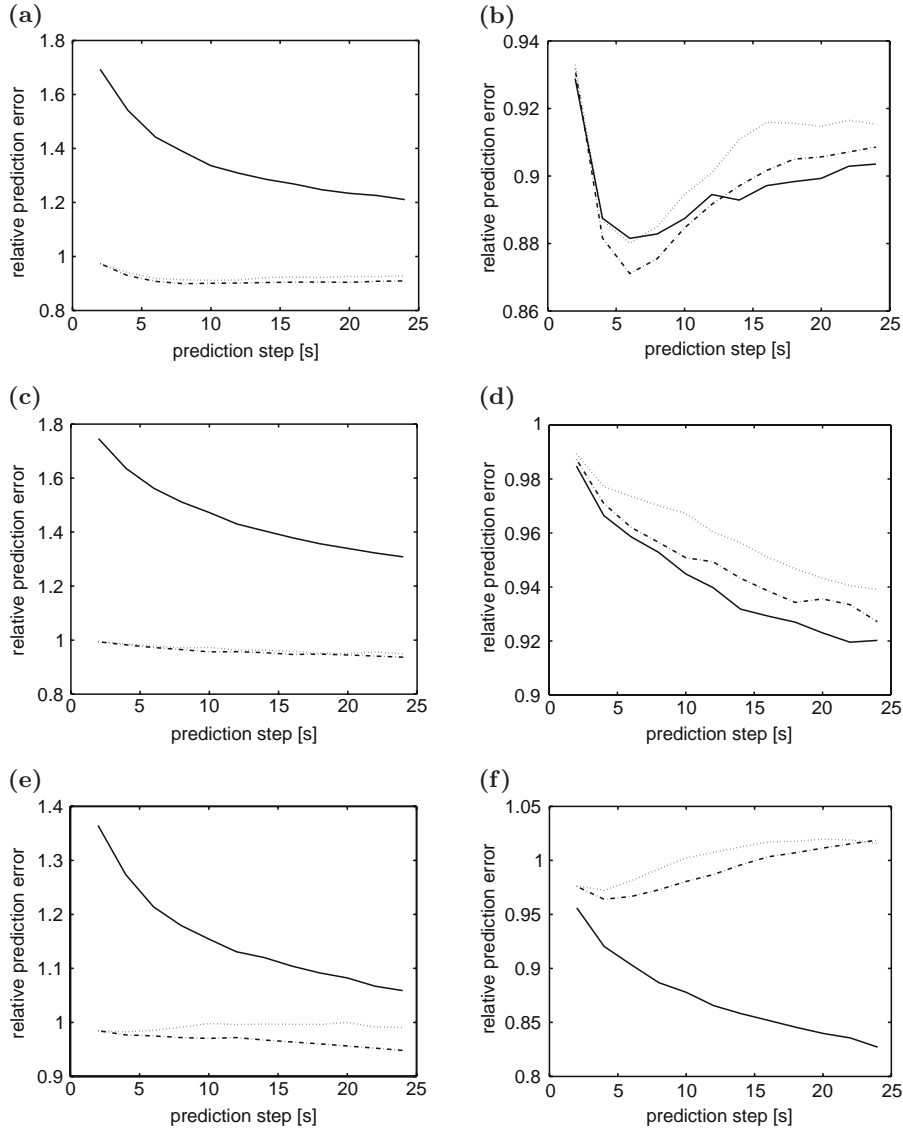


Fig. 2.3. Means of root mean square errors for the compared combinations of methods relative to those of persistent prediction. (a) and (b) are cases where the additional observation was at upstream. (c) and (d) are cases where the additional observation was at downstream. (e) and (f) are cases where the additional observation was nearby but at neither downstream nor upstream. (a), (c), and (e) show root mean square errors for the wind direction, and (b), (d), and (f), those for the absolute wind speed. In each part, the *solid line* shows the case of polar coordinates with the direct method, the *dotted line*, that of rectangular coordinates with the direct method, and the *broken line*, that of rectangular coordinates with the $\psi\phi$ method

the north–south direction. On the other days, we placed them 5 m apart in the east–west direction. Since we are interested in the dynamics of 2 s order, we took the moving average of 2 s and resampled it by 2 s.

We used measurements taken from the two points for predicting the wind at one of them as we can expect better prediction by doing this [7].

2.3.6 Results

In the previous sections, we argued that there are two coordinate systems and two possible methods for the prediction. Thus, theoretically there are four possible combinations. In this section, however, we compare three of them, namely, the polar coordinates with the direct method, the rectangular coordinates with the direct method, and the rectangular coordinates with the $\psi\phi$ method, since the polar coordinates with an iterative method does not provide good prediction since they have to use the prediction of the wind direction several times.

In Fig. 2.3, we compared the performances of the three combinations of the methods by the prediction error for each method relative to that for the persistent prediction, where we used the value of 2 s before as a prediction.

When we predicted the wind direction, the rectangular coordinates with the $\psi\phi$ method achieved smaller mean prediction errors than the other two combinations (Fig. 2.3a, c, e). We interpret this results as follows: Because the wind direction takes a value on the ring, the polar coordinates are not a good idea for representing the wind direction. By using the $\psi\phi$ method, we can make the prediction errors smaller compared to the direct prediction.

When we predicted the absolute wind speed (Fig. 2.3b, d, f), the polar coordinates with the direct method gave the smallest mean prediction error for all cases except for short prediction steps of the case where the additional observation was at upstream. We think that the polar coordinates were better

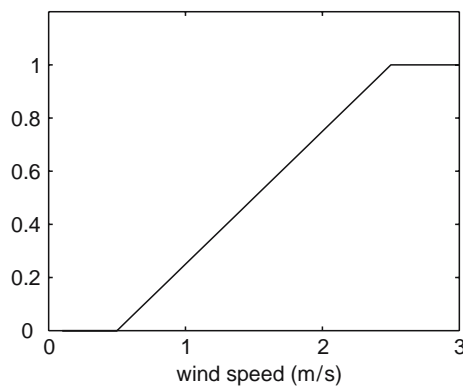


Fig. 2.4. Power curve used for the simulation

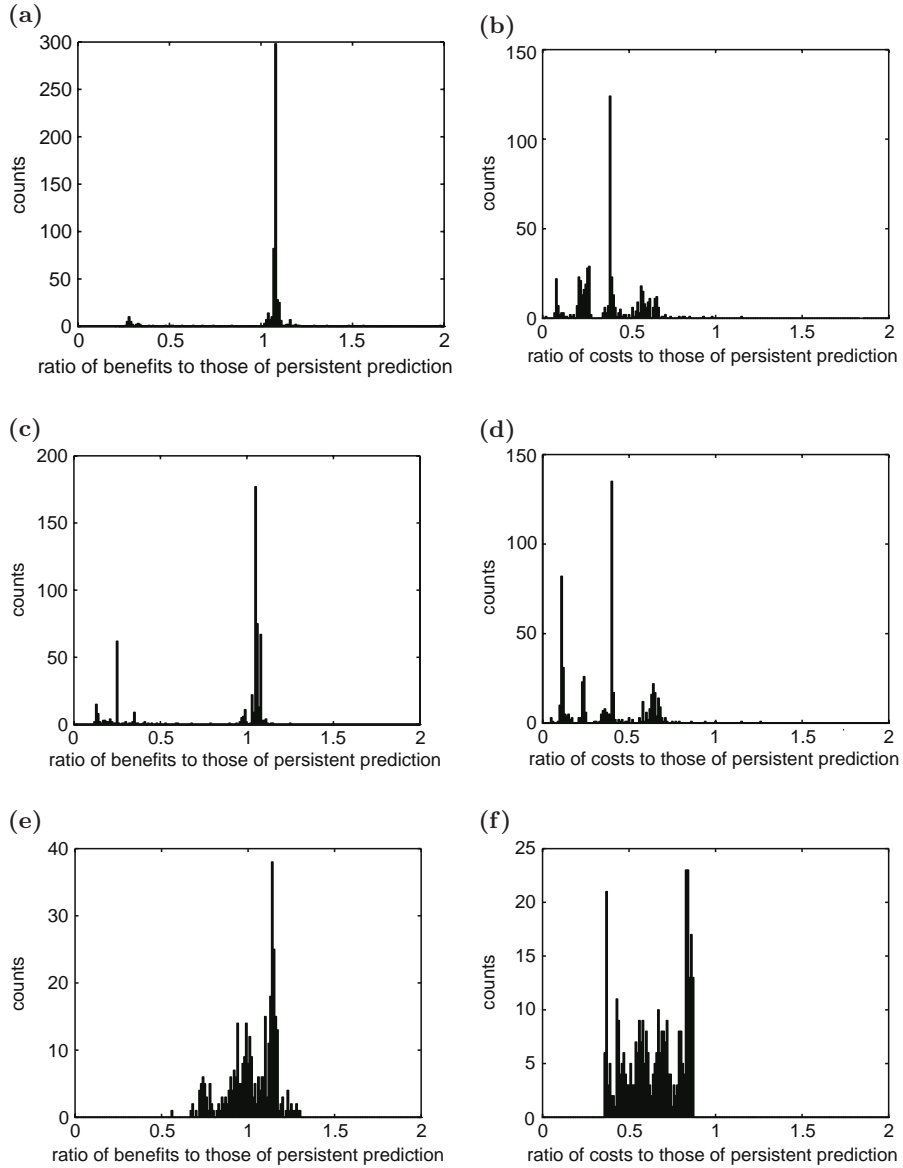


Fig. 2.5. Results of simulations for energy production. (a), (c), and (e) show the histograms of ratios of expected net benefits for the combination of the proposed prediction methods to those for the persistent prediction. (b), (d), and (f) show the histograms of ratios of expected costs for controlling a wind turbine based on the proposed prediction to those based on the persistent prediction. (a) and (b) are cases where the additional observation was at upstream. (c) and (d) are cases where the additional observation was at downstream. (e) and (f) are cases where the additional observation was nearby but at neither downstream nor upstream

for predicting the absolute wind speed because the polar coordinates represent what we want to predict directly.

2.4 Applying the Wind Prediction to the Yaw Control

In Sect. 2.3, we observed that the wind direction was predicted well when the rectangular coordinates with the $\psi\phi$ method was used, while the absolute wind speed was predicted well when the polar coordinates and the direct method were used. In this section, we compared the combination of these methods with the persistent prediction in expected energy.

From the actual data, we predicted the wind directions and absolute wind speeds for prediction steps between 2 and 24 s. Then we fed the predictions to the model of wind turbine presented in Sect. 2.1 and decided the control policy. For estimating energy production, we used the actual wind data.

We used a power curve shown in Fig. 2.4 for the simulation. We set $C_1 = C_2 = C_3 = 0.01$.

We used the polar coordinates with the direct method and the rectangular coordinates with the $\psi\phi$ method for predicting the absolute wind speed and the wind direction, respectively. This combination is the best, based on the results of Sect. 2.3. We compared the results with the case where we used the persistent prediction for both the absolute wind speed and the wind direction.

The results were shown in Fig. 2.5. The median of ratio of the expected net benefits for the proposed prediction to those for the persistent prediction is greater than 1. The median of ratio of the expected control costs for the proposed prediction to those for the persistent prediction is smaller than 1. These results mean that the proposed prediction works well for controlling a wind turbine. The above observations were always true and do not depend on the place of the additional observation point much.

2.5 Conclusions

In this chapter, we showed it is possible that wind modelling may help to increase energy production. To show this, we first formulated the problem for controlling the yaw angle of a wind turbine and found that we need to predict the wind direction and the absolute wind speed for a certain range of the future. Next, we investigated the properties of time series of the wind using surrogate data analysis. We showed that the time series always have serial dependence and they sometimes have nonlinearity. We also argued that time series of the wind may be characterised by small fluctuations with trends. These findings are positive signs for the prediction. Then, we modelled the wind and found that the wind direction is predicted well by the rectangular coordinates and the $\psi\phi$ method, and the absolute wind speed is predicted well if one uses the polar coordinates and the direct method. Finally, we applied

the prediction of the wind to the simulation of wind turbine control. We found that the nonlinear prediction can potentially increase the energy production. An important point is that although our gain of information is small in terms of the root mean square error, it is big enough for controlling a wind turbine optimally.

We feel that the methods discussed here could be useful for controlling a small wind turbine since a small wind turbine can be directed with a small force very quickly, and in the scale of a small wind turbine, the wind at the edges of the blades is not so much different from that at the centre.

The focus of this chapter has been that a heterogeneous data fusion strategy is useful for controlling a wind turbine, and therefore the models of turbine, stress, fatigue, and friction used are synthetic. Although these models are not guaranteed to provide a feasible and safe solution for a real wind turbine, we believe that they are sufficiently realistic to illustrate the proposed framework.

Acknowledgement

This study was partially supported by the Industrial Technology Research Grant Program in 2003, from the New Energy and Industrial Technology Development Organization (NEDO) of Japan.

References

1. Garcia, S.P., Almeida, J.S.: Multivariate phase space reconstruction by nearest neighbor embedding with different time delays. *Physical Review E* **72**(2), 027205 (2005)
2. Goh, S.L., Chen, M., Popvic, H., Aihara, K., Obradovic, D., Mandic, D.P.: Complex-valued forecasting of wind profile. *Renewable Energy* **31**(11), 1733–1750 (2006)
3. Goh, S.L., Mandic, D.P.: A complex-valued RTRL algorithm for recurrent neural networks. *Neural Computation* **16**(12), 2699–2713 (2004)
4. Goh, S.L., Mandic, D.P.: Nonlinear adaptive prediction of complex-valued signals by complex-valued PRNN. *IEEE Transactions on Signal Processing* **53**(5), 1827–1836 (2005)
5. Hirata, Y., Horai, S., Suzuki, H., Aihara, K.: Testing serial dependence by Random-shuffle surrogates and the wayland method. *Physics Letters A* (2007). DOI 10.1016/j.physleta.2007.05.061
6. Hirata, Y., Mandic, D.P., Suzuki, H., Aihara, K.: Wind direction modelling using multiple observation points. *Philosophical Transactions of the Royal Society A* (2007). DOI 10.1098/rsta.2007.2112
7. Hirata, Y., Suzuki, H., Aihara, K.: Predicting the wind using spatial correlation. In: *Proceedings of 2005 International Symposium on Nonlinear Theory and its Applications (NOLTA 2005)* (2005)

8. Hirata, Y., Suzuki, H., Aihara, K.: Reconstructing state spaces from multivariate data using variable delays. *Physical Review E* **74**(2), 026202 (2006)
9. Judd, K., Mees, A.: On selecting models for nonlinear time-series. *Physica D* **82**(4), 426–444 (1995)
10. Judd, K., Mees, A.: Embedding as a modeling problem. *Physica D* **120**(3–4), 273–286 (1998)
11. Judd, K., Small, M.: Towards long-term prediction. *Physica D* **136**(1–2), 31–44 (2000)
12. Kantz, H., Holstein, D., Ragwitz, M., Vitanov, N.K.: Extreme events in surface wind: Predicting turbulent gusts. In: S. Bocalletti, B.J. Gluckman, J. Kurths, L.M. Pecora, R. Meucci, O. Yordanov (eds.) *Proceedings of the 8th Experimental Chaos Conference*, no. 742 in *AIP Conference Proceedings*. American Institute of Physics, New York (2004)
13. Kantz, H., Holstein, D., Ragwitz, M., Vitanov, N.K.: Markov chain model for turbulent wind speed data. *Physica A* **342**(1–2), 315–321 (2004)
14. Kantz, H., Holstein, D., Ragwitz, M., Vitanov, N.K.: Short time prediction of wind speeds from local measurements. In: J. Peinke, P. Schaumann, S. Barth (eds.) *Wind Energy: Proceedings of the EUROMECH Colloquium*. Springer, Berlin Heidelberg New York (2006)
15. Kennel, M.B.: Statistical test for dynamical nonstationarity in observed time-series data. *Physical Review E* **56**(1), 316–321 (1997)
16. Kennel, M.B., Brown, R., Abarbanel, H.D.I.: Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical Review A* **45**(6), 3403–3411 (1992)
17. McNames, J.: A nearest trajectory strategy for time series prediction. In: *Proceedings of the International Workshop on Advanced Black-Box Techniques for Nonlinear Modeling* (1998)
18. Ragwitz, M., Kantz, H.: Detecting non-linear structure and predicting turbulent gusts in surface wind velocities. *Europhysics Letters* **51**(6), 595–601 (2000)
19. Rissanen, J.: MDL denoising. *IEEE Transactions on Information Theory* **46**(7), 2537–2543 (2000)
20. Roulston, M.S., Kaplan, D.T., Hardenberg, J., Smith, L.A.: Using medium-range weather forecasts to improve the value of wind energy production. *Renewable Energy* **28**(4), 585–602 (2003)
21. Scheinkman, J.A., LeBaron, B.: Nonlinear dynamics and stock returns. *Journal of Business* **62**(3), 311–337 (1989)
22. Schreiber, T., Schmitz, A.: Improved surrogate data for nonlinearity tests. *Physical Review Letters* **77**(4), 635–638 (1996)
23. Schreiber, T., Schmitz, A.: Surrogate time series. *Physica D* **142**(3–4), 346–382 (2000)
24. Small, M., Yu, D., Harrison, R.G.: Surrogate test for pseudoperiodic time series. *Physical Review Letters* **87**(18), 188101 (2001)
25. Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., Farmer, J.D.: Testing for nonlinearity in time-series: the method of surrogate data. *Physica D* **58**(1–4), 77–94 (1992)
26. Timmer, J.: Power of surrogate data testing with respect to nonstationarity. *Physical Review E* **58**(4), 5153–5156 (1998)
27. Wayland, R., Bromley, D., Pickett, D., Passamante, A.: Recognizing determinism in a time-series. *Physical Review Letters* **70**(5), 580–582 (1993)