NONPARAMETRIC MODELING AND SPATIOTEMPORAL DYNAMICAL SYSTEMS

MARKUS ABEL

Universität Potsdam, Institut für Physik, Postfach 601553, D-14415 Potsdam, Germany

This article describes how to use statistical data analysis to obtain models directly from data. The focus is put on finding nonlinearities within a generalized additive model. These models are found by means of backfitting or more general algorithms, like the alternating conditional expectation value one. The method is illustrated by numerically generated data. As an application, the example of vortex ripple dynamics, a highly complex fluid-granular system, is treated.

Keywords: Time series analysis; spatiotemporal modeling; statistical analysis.

1. Introduction

A major goal of spatiotemporal data analysis consists in inferring a model from a given spatiotemporal data set. Conventional ways of modeling are purely theoretical arguments and *a posteriori* evaluation of possible models with suitable measures. In this article, statistical methods and applications to infer generalized additive models from data are presented. Attention is placed on spatiotemporal systems, but there are certainly more possibilities in using the ideas in physics and engineering. Generalized additive models have been introduced to statistical analysis in the 1980's [Hastie & Tibshirani, 1986, 1990]. Solving algorithms and mathematical proofs have been developed contemporarily ([Hastie & Tibshirani, 1990] and references therein).

During the last two decades, natural science has profited considerably by the increase of computer power in modern machines. In nonlinear and statistical sciences, old techniques could be exploited to their full extent and new methods have been developed. The identification of nonlinearities is one of the important topics in data analysis of, e.g. chaotic or pattern forming systems [Kantz & Schreiber, 1997; Ott *et al.*, 1994]. This in turn is feasible by statistical tools.

Careful application of algorithms and thorough interpretation of results lead to a direct way to obtain models from data. Hereafter, the basic ideas of the techniques to solve for a generalized additive model are explained and an application to spatiotemporal data is presented. The intention is not to give a full overview of statistical methods, nor to report an experiment in detail. Rather, it is to show how a consistent procedure to infer models from data can be used as a feedback to theory which in turn can motivate experimental designs to yield new data and so on.

The paper is structured as follows: In Sec. 2 the algorithmic solutions for generalized additive models are pointed out. This is done by the example of ACE, the **A**lternating **C**onditional **E**xpectation value algorithm, which uses the backfitting technique. This presentation is given some room; even though it may not be new to statisticians, technical points need some explanation to be understood. In Sec. 3 more information is provided to be used in the context of spatiotemporal data analysis. In Sec. 4 problems arising from preprocessing of data are discussed along with the example of vortex ripple pattern formation. The paper concludes with a short discussion in Sec. 5.

2. Generalized Additive Models and Backfitting

The standard tool of data analysis for physicists and engineers is the multivariate linear regression, where one fits the model

$$U_0 = C + \sum_{i=1}^{M} \alpha_i U_i + \varepsilon \qquad (1)$$

to experimental data. The symbols ε and U_i (upper case letters), $i = 0, \ldots, M$ denote a random process with given distribution, C is a constant and α_i are model parameters. One measures a realization u_i (lower case letters), $i = 0, \ldots, M$ of the random process and finds estimates $\hat{\alpha}_i$ for the parameters, e.g. by least squares fitting [Teukolsky *et al.*, 1993]. Throughout the text, estimations are denoted by a hat, where ambiguity exists, otherwise additional symbols are omitted.

There are many ways of generalization. One often used way is the generalized linear regression model,

$$U_0 = C + \sum_{i=1}^{M} \alpha_i f_i(U_i) + \varepsilon \, . \qquad (2)$$

In this model, the data analyst chooses beforehand the functions f_i due to some prior knowledge about the system or simply by guessing. Then, a linear regression procedure yields estimates for the parameters α_i. So, the model is expressed in parametric form by the estimates $\hat{\alpha}_i$. For details about linear regression see, e.g. [Teukolsky *et al.*, 1993; Montgomery, 1992].

If no obvious choice for a function is at hand, it is desirable to fit the nonparametric model

$$U_0 = C + \sum_{i=1}^{M} f_i(U_i) + \varepsilon \, . \qquad (3)$$

Given some data, estimates $\hat{f}_i$ must result. With this kind of modeling no prior knowledge about functional dependencies is needed as input; one can find a very general functional form. A disadvantage is the absence of more general terms, like $f_i(U_i, U_j)$, $j \neq i$. The reason for this lack is a practical one: as will become clear below, any implementation finding functions of d variables has to compete with the curse of dimensionality. For the generalized additive model (3), estimates for the functions f_i are found by the *backfitting algorithm* [Hastie & Tibshirani, 1986, 1990; Härdle & Hall, 1993]. This is an iterative procedure, working by the following rules:

(1) Initialization: $C = E(U_0)$, $f_i = f_i^0$, $i = 1, \ldots, M$;
(2) Iteration: for $i = 1, \ldots, M$ calculate

$$f_i = E\left(U_0 - C - \sum_{k \neq i} f_k(U_k) | U_k \right)$$

until convergence,

where $E()$ denotes the expectation value and f_i^0 are some appropriate initial settings. In applications, one has to obtain the functions from data as realizations of the generating process. Then, this operator has to be replaced by its estimator. In each iteration step we estimate the functions by $\hat{f}_i(U_i) = S_i(\cdot|U_i)$, a smoothing operator which returns a function dependent on U_i.

The discussion of properties and choice of the smoother is a subtle issue and lies far beyond the scope of this paper, a detailed discussion is given in [Buja *et al.*, 1989]. The examples presented in this article have been obtained using the simple running mean smoother (moving average). This is not the best choice for many problems, e.g. points at the boundaries have to be considered with care or must be cut. Throughout this article, there have been enough data to afford the luxury of cutting the boundaries, in spatiotemporal systems, a data set often consists of thousands of data points. Smoothing splines are good smoothers in many situations due to their differentiability properties [Hastie & Tibshirani, 1990; Buja *et al.*, 1989]. In any case, a careful check should be undertaken for a concrete set of data.

The iteration works by adjusting only for one function, subtracting all the others from the estimation of U_0, i.e. one smoothes the partial residual $(U_0 - \sum_k f_k)$ against U_i. In spatiotemporal data analysis, one is faced with the task of finding equations of motion from data. These can be coupled map lattices, a set of ordinary coupled differential equations, partial differential equations or even more complicated models. In many of these formulae, the lhs. is linear (e.g. a time derivative) and Eq. (3) constitutes an appropriate additive, nonlinear model. If the measured data are nonlinear transforms through a measuring function, it is desirable to model a nonlinear lhs, too:

$$f_0(U_0) = C + \sum_{i=1}^{M} f_i(U_i) + \varepsilon \, . \qquad (4)$$

The constant can be absorbed w.l.o.g. into the functions f_i. Now, the model is more symmetrical and indeed, in physics, often situations without a clear predictor (lhs) – response (rhs) structure are found. A typical example is pattern formation where one is interested in the nonlinear interaction of stationary states. *A priori* it is not clear which variables constitute the state space nor is it always possible to measure them. Data analysis provides then an estimation of the model as a nonlinear, possibly noninvertible, transformation of the measured data. As an example, take the relation $U_0 = g(U_1 + U_2)$ with g as some nonlinear invertible function. The inverted equation $g^{-1}(U_0) = U_1 + U_2$ cannot be found with (3), but with the model (5). An example for this case is given below.

An algorithm to solve Eq. (4) is ACE, (the Alternating Conditional Expectation value algorithm) [Breiman & Friedman, 1985], it works by minimizing the squared error

$$E\left[f_0(U_0) - \sum_{i=1}^{M} f_i(U_i)\right]^2 . \qquad (5)$$

The method shall be explained from the one-dimensional estimation over the two- to the M-dimensional problem.

(1) $U_0 = f_1(U_1) + \varepsilon$. For this simple model, one obtains from Eq. (5)

$$E[U_0 - f_1(U_1)]^2 = \min , \qquad (6)$$

where minimization is achieved by variation of f_1 in the space of measurable functions [Breiman & Friedman, 1985]. After some elementary steps one finds the solution [Honerkamp, 1994]

$$f_1(U_1) = E(U_0|U_1) . \qquad (7)$$

For applications, again a smoother is used as an estimator of the expectation value operator. The result of the smoothing on a set of numerically produced data is shown for the example $U_0 = U_1^2$ in Fig. 1. The data have been generated by first drawing 10 000 equally distributed random numbers for u_1 in the interval $(-3, 3)$. These numbers have been squared to yield the values for u_0, after the Gaussian noise $N(0, 0.1^2)$ has been added. The bandwidth of the running mean smoother has been set here and below to 200 points. In Fig. 1, the scatterplot of the points (u_1, u_0) is shown (gray) together with the estimate for the function f_1, found by Eq. (7) (black, thick line). As indication for the estimation
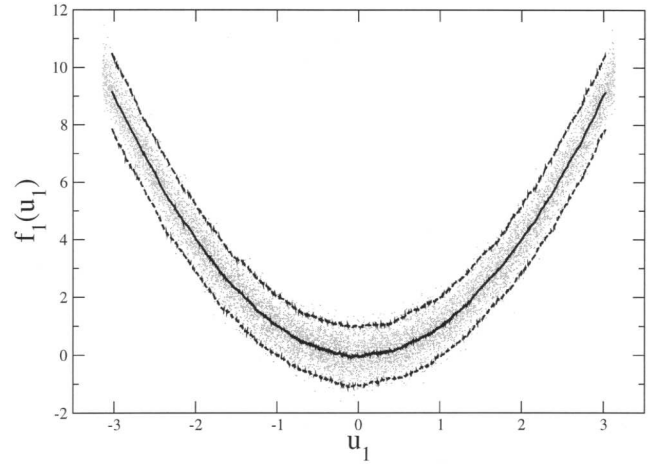

Fig. 1. One-step transformation in the case $f_0 = U_0$, $f_1(U_1) = U_1^2$. The scatterplot of points (u_0, u_1) is shown by gray dots. Applying the running mean smoother yields an estimation for the function f_1 (middle, thick black line), above and below is plotted twice the pointwise standard deviation (dashed black lines).

error the pointwise standard deviation has been calculated from the estimated function and the residuals (upper and lower black lines).

(2) $f_0(U_0) = f_1(U_1) + \varepsilon$. In this case, the solution is found iteratively:

(1) Initialization: $f_0(U_0) = [U_0 - E(U_0)]/\sqrt{\text{VAR}(U_0)}$.
(2) Computation of $f_1(U_1) = E[f_0(U_0)|U_1]$
(3) Computation of $\tilde{f}_0(U_0) = E[f_1(U_1)|U_0]$
(4) Normalization: $f_0(U_0) = \tilde{f}_0(U_0)/\sqrt{\text{VAR}(\tilde{f}_0(U_0))}$.
(5) Iteration of (2)–(4) until convergence.

A smoother with nonzero bandwidth is contracting and thus the trivial solution is excluded by normalization (point 4) with the variance $\text{VAR}(\tilde{f}_0(U_0))$.

In order to show how ACE handles noninvertible relations, data $(A \cos \theta, A \sin \theta)$ have been generated, with $\theta = \varphi + \varepsilon_\varphi$, φ as an equally distributed random number in $[0, 2\pi)$ and $A = 1 + \varepsilon_A$. The noise ε_φ, ε_A is Gaussian distributed, with $N(0, 0.05^2)$. This corresponds to the noisy measurement of an amplitude and a phase with a nonlinear, noninvertible measurement transformation.

Without noise, one has $U_0 = \cos(\varphi)$, $U_1 = \sin(\varphi)$ and thus $U_0^2 = 1 - U_1^2$. With noise, the relation reads to first order $U_0^2 - 2\varepsilon_A U_0 = 1 - U_1^2 + 2\varepsilon_A U_1$. This means that multiplicative noise enters in predictor and response, touching the errors in variables problem which is relevant for many applications. The strong, multiplicative transformation of the noise is mirrored directly in the results.
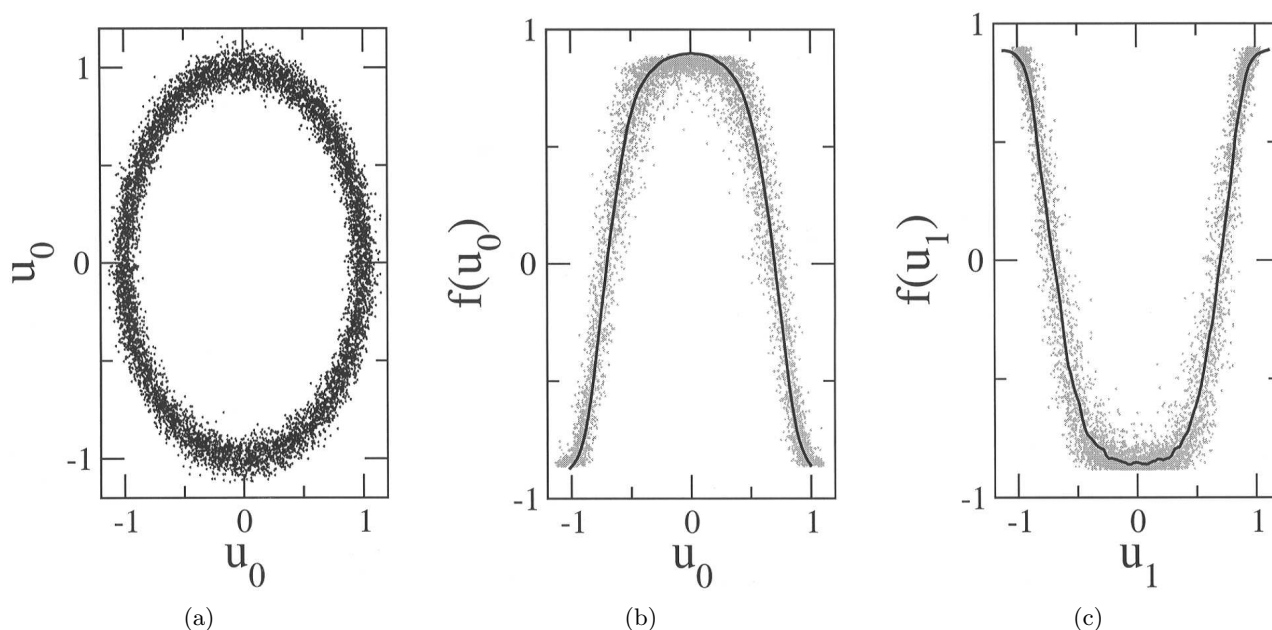
Fig. 2.   Example for $U_0 = A \sin \theta$, $U_1 = A \cos \theta$, where A and θ include measurement noise (a). The expected functional relation is noninvertible: $U_0^2 + U_1^2 = 1$, noise on A, θ is transformed nonlinearly. The gray dots in (b) and (c) show this effect in the residuals. This in turn influences the result in that it induces bending at the ends and in the central part of the results of ACE, where the transformation has its biggest effect f_0 and f_1 [see the middle line in (b) and (c)]. The symmetric pointwise standard deviation is not a good indicator for the asymmetry [upper and lower lines in (b) and (c)], measures should be accordingly calculated. Functions are shifted to zero mean.

In Fig. 2(a) the data are displayed in a scatter-plot, in Figs. 2(b) and 2(c), the resulting functions are shown on top of the respective residuals together with twice the pointwise standard deviation as an error indicator. Due to the nonlinear noise transformation the functions are distorted. The pointwise standard deviation is less significant as error indicator due to the very asymmetric local distribution of values, to be seen in the upper and lower curves of Figs. 2(b) and 2(c). In this case, measures capturing the asymmetry should be calculated. The above example underlines a conceptual problem if one is faced with errors in the measurement variables. The general treatment of this problem is not simple, for more details see [Fan & Truong, 1993; Carroll *et al.*, 1999] and references therein.

Even though the precise estimation of the functional dependence is not possible in the above case, one finds an approximation which is not bad for many experiments. Please note that linear tools are able to find this relation easily, rather one should undergo a trial-and-error procedure until a guess for the right form of the functions is found. Some problems with a non-Gaussian distribution also remain in linear regression.

(3) $f_0(U_0) = \sum_{i=1}^{M} f_i(U_i) + \varepsilon$. This equation defines a (noisy) hypersurface in M-dimensional space (in two dimensions, one finds a line). The functions f_i, fall on a hyper*plane*: the algorithm performs a transformation of the points $\{U_1, \ldots, U_M\}$ to the points $\{f_1, \ldots, f_M\}$ such that the best plane in the mean squared sense is found.

The modification in the algorithm has to be done in Steps 2 and 3 above. They are replaced by

(2) Computation of $f_i(U_i) = E[f_0(U_0) - \sum_{k \neq i} \times f_k(U_k)|U_i]$ by backfitting.
(3) Computation of $\tilde{f}_0(U_0) = E[\sum_i f_i(U_i)|U_0]$.

This shows that the ACE algorithm is based on backfitting. To demonstrate a higher-dimensional example, the relation $U_0 = \exp(U_1 + U_2^2)$ has been chosen. The variables have been generated using equally distributed values for U_1, U_2 in the interval $(-2, 2)$. From the resulting numbers, u_0 has been calculated, thus the smallest possible value of u_0 is $\exp(-2) \simeq 0.14$, the largest is $\exp(6) \simeq 400$. After this procedure (in contrast to the previous example), Gaussian, $N(0, 0.2^2)$, noise has been

added to u_i, $i = 0, 1, 2$. This avoids the problems with transformation of measurement noise, discussed above. But here, another interesting question for the modeler is raised instead: what happens if one tries to include some additionally measured variable in the model?

As an example, one can imagine a measurement of four variables, where it is not clear *a priori* which variables are needed in a minimal model. As a first guess, a generalized additive model $f_0(U_0) = f_1(U_1) + f_2(U_2) + f_2(U_3)$ will be tested. This situation shall now be related to the data generated
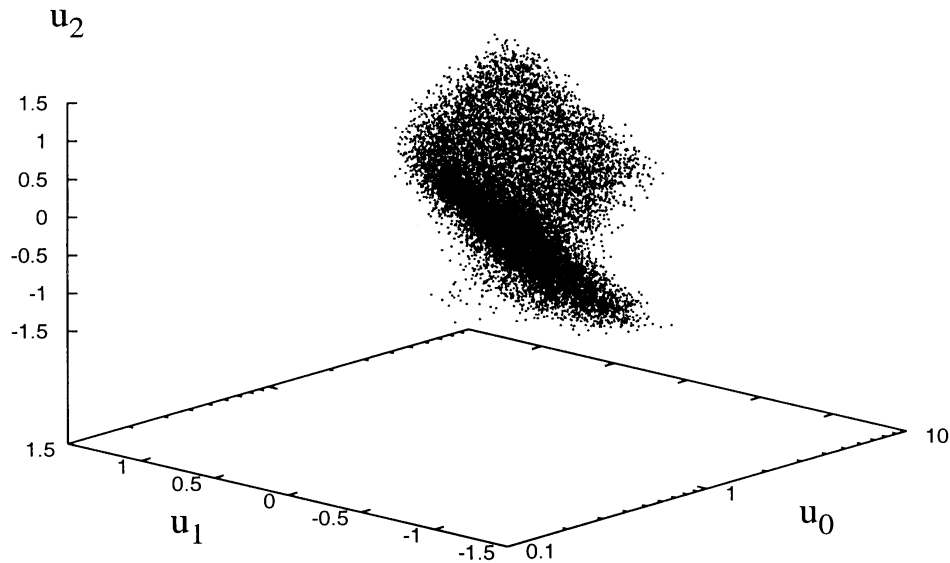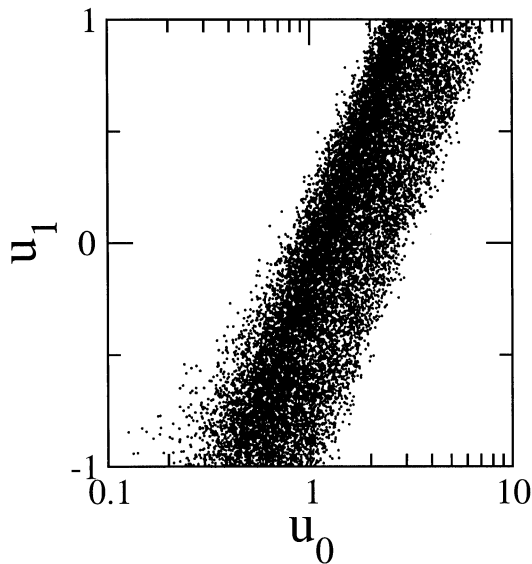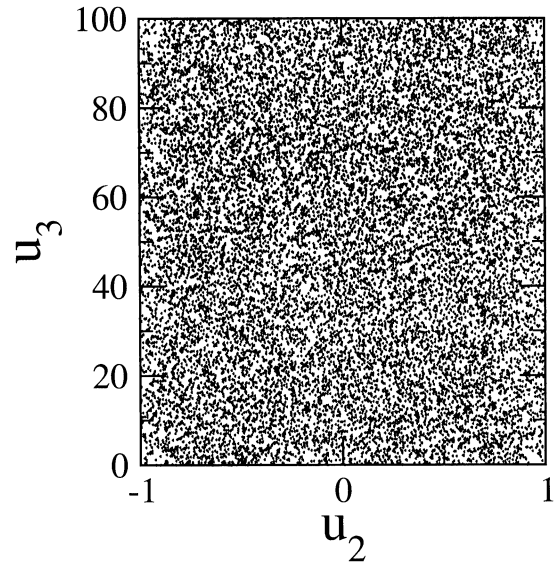

Fig. 3.   Three-dimensional plot of the data (U_0, U_1, U_2) with $U_0 = \exp(U_1 + U_2^2) + \varepsilon$.

(a)

(b)

Fig. 4.   Two-dimensional scatter plots of a part of the data as projection on the plane. The examples u_0, u_1 and u_2, u_3 are shown. Even "guessing" a logarithmic relation for u_0, no clear relation can be read on the left, and on the right, the data are completely random according to the zero correlation of u_3 with the rest.

2032   *M. Abel*

Int. J. Bifurcation Chaos 2004.14:2027-2039. Downloaded from www.worldscientific.com
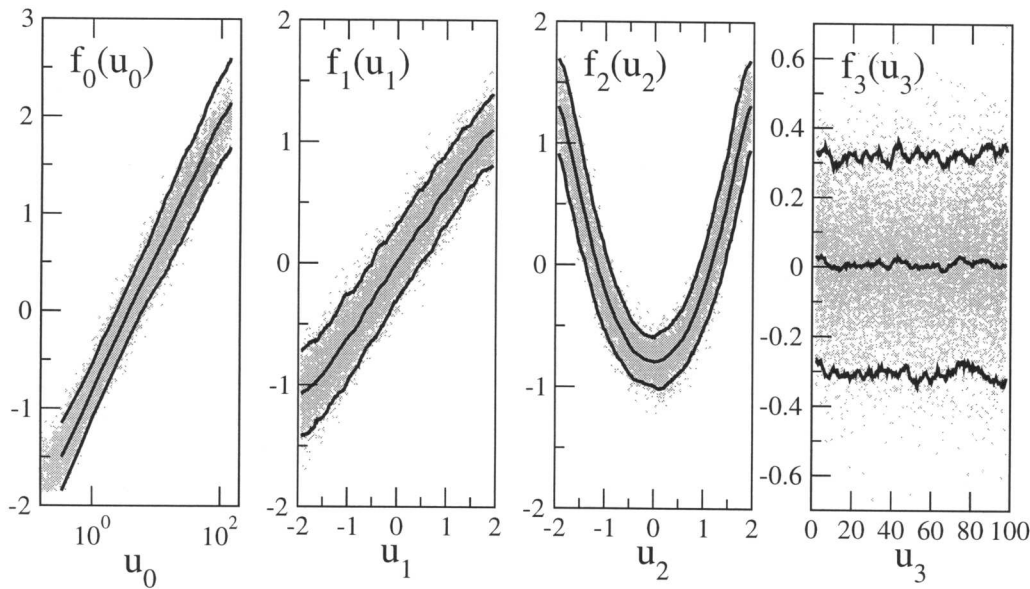by PURDUE UNIVERSITY on 03/20/14. For personal use only.

Fig. 5.   Transformations for the given data with twice the pointwise standard errors (black lines) on top of the residuals (gray). The logarithm $f_0(u_0)$ is found over three decades to hit numerical restrictions at the smallest points. The linear and square functions f_1 and f_2 are found very well. The variable u_3 is found to be "unimportant" with $f_3(u_3) \simeq 0$.

above: an additional, equally distributed variable in the interval (0, 100), uncorrelated to U_0, U_1, U_2, is generated and fed as fourth variable into ACE.

The three-variable additive model is $f_0(U_0) = f_1(U_1) + f_2(U_2)$ and one expects ACE to find the estimates $\hat{f}_0(U_0) = \ln(U_0)$, $\hat{f}_1(U_1) = U_1$, $\hat{f}_2(U_2) = U_2^2$ due to the additive structure. For the four-dimensional model, one expects an unchanged result for $\hat{f}_0$, $\hat{f}_1$, $\hat{f}_2$ plus a zero function $\hat{f}_3(U_3) = 0$ in the additional variable.

In Fig. 3, the 3D scatterplot of the generated relevant data are shown by the points $(\ln(U_0), U_1, U_2)$. A two-dimensional representation is not able to display the full relation, if one plots, e.g. U_0 versus U_1 (Fig. 4), no functional dependence is cognizable. The transformations $(\hat{f}_0, \hat{f}_1, \hat{f}_2, \hat{f}_3)$ found by ACE are shown in Fig. 5 as functions of their arguments. The function f_0 is found to be logarithmic, $f_0 = 0.5 \ln(u_0) + c$, f_1 is linear with a slope of 0.5, and $f_2 = au_2^2 + c$ is quadratic with $a = 0.5$. The factor 0.5 reflects possible ambiguities in the result, as well as the addition of the constant, see Sec. 3.2.

After multiplying the results by 2 and adding a constant to f_0 and f_2 one can speak of a good coincidence of expected and actual results. The additional variable u_3 has no dependence on the others and yields $f_3 = O(10^{-2}) \simeq 0$. Thus, the method is

able to identify uncorrelated terms. A comparison of runs with and without the additional third term on the rhs showed that the functions are virtually unchanged, confirming the stability of the algorithm. Despite the seemingly convincing result it must be noted that results can be misleading if there exist correlations between an additional term and "true" variables. If the task is to find a minimal model, one has to consider this effect. The connection of correlation with ACE is discussed in Sec. 3.3.

3. Some Further Topics

3.1. *Smoothers and the bias-variance dilemma*

As mentioned above, the choice of the smoother can be delicate. Typically, to each smoother belongs a smoothing parameter which characterizes its bandwidth (window width). For the running mean smoother, one obtains for problem (7)

$$\hat{f}_1(U_{1,l}) = \frac{1}{N_l} \sum_{j \in B(U_{0,l})} U_{0,j}, \qquad (8)$$

where $B(U_{0,l})$ is a neighborhood of the point $U_{0,l}$. Increasing bandwidth enlarges the neighborhood and thus the number N_l of data points in B. The

respective expectation and variance are

$$E(\hat{f}_1(U_{1,l})) = \frac{1}{N_l} \sum_{j \in B(U_{0,l})} f(U_{0,j}), \qquad (9)$$

$$\mathrm{VAR}(\hat{f}_1(U_{1,l})) = \frac{\sigma^2}{N_l}, \qquad (10)$$

where $\sigma_l = \sigma$ is assumed for simplicity. Thus, the variance shrinks with increasing bandwidth; the bias, however, increases because more terms different from $f(U_{0,l})$ are involved. If one wants to find the optimum choice of the bandwidth in order to have neither large bias nor large variance, one uses measures like the mean-squared error or the average predictive squared error; these are evaluated using cross-validation and yield a criterion for the bandwidth to be chosen. For more information we refer to the literature [Hastie & Tibshirani, 1990; Honerkamp, 1994].

The iteration procedure should run until convergence. In fact, convergence is not guaranteed for all classes of smoothers, asymptotic properties and convergence is discussed in a rigorous way in [Breiman & Friedman, 1985; Buja *et al.*, 1989; Mammen *et al.*, 1999]. Intuitively, it is clear that the iteration should stop when the errors on the functions fall below the noise ε in the model (4). Then it is a question of how to properly find the respective errors. There are different ways to estimate the error consistently, depending on the smoother and the error models used. Locally, one can choose twice the pointwise standard error as a characterization, global confidence criteria are harder to derive [Hastie & Tibshirani, 1990; Fan & Truong, 1993], an elegant and modern approach is given by the Bayesian formulation of the backfitting algorithm [Hastie & Tibshirani, 2000]. In the case of errors in variables, care has to be taken to calculate the error correctly, a Bayesian approach is appealing in that case. In the examples, the most naive criterion, i.e. twice the pointwise standard error has been used.

3.2. *Peculiarities of ACE*

The ACE algorithm can be related to other algorithms, namely canonical variates and alternating least squares methods. Special properties of all three methods are discussed in [Buja, 1990], peculiarities of ACE are discussed in [Hastie & Tibshirani, 1990; Breiman & Friedman, 1985], here

a few items shall be given without a claim for completeness.

(i) ACE is symmetric in the predictor and the response, which is quite unusual for a regression tool, but suits well several setups in physical applications.

(ii) Transformations of the variables are not reproduced by ACE. Take for example, the case $U_0 = f_1(U_1) + \varepsilon$. If one applies a transformation to each of the sides, ACE does not necessarily find f_1 again. This reflects parts of the nature of the inverse problem, which is in general not uniquely solvable. Especially, adding a constant to either of the sides in the model is ambiguous in ACE.

(iii) Different distributions in the model constituents can "distort" the resulting functions. For example, it may happen that one term in a model with two variables is Gaussian distributed and another one is equally distributed in some interval. Then, the Gaussian variable is bent at the ends (cf. Fig. 5).

3.3. *Maximal correlation*

The ACE algorithm can be regarded as a regression tool, but there exists another possible interpretation. Instead of solving the least squares error minimization problem

$$E\left[f_0(U_0) - \sum_{i=1}^{M} f_i(U_i)\right]^2 = \min \qquad (11)$$

one can reformulate the above to

$$\Psi = \mathrm{corr}\left(f_0(U_0), \sum_{i=1}^{M} f_i(U_i)\right) = \max, \qquad (12)$$

where corr denotes the correlation function. This reflects the principle of the maximal correlation [Rényi, 1959; Bell, 1962; Csaki & Fischer, 1963; Rényi, 1970]. The functions that fulfil Eq. (12) are called optimal transformations (in the sense of correlation), found by the ACE procedure. To obtain a global measure for the importance of a single term $f_i(U_i)$, there are several possibilities, e.g. one can choose the overall variance, $\mathrm{VAR}(f_i)$, but this does not reflect correlation with other terms and a noise term with large variance would be judged as important. Here, the author suggests to use Eq. (12) in a
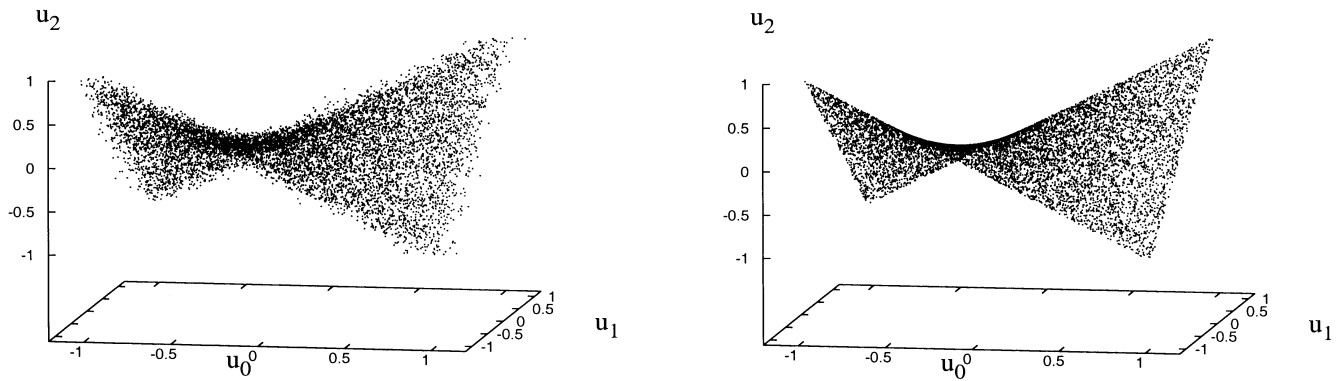
Fig. 6. (Left) Noisy data for the surface $U_0 = U_1 \cdot U_2$, $\sigma = 0.05$ for all U_i. (Right) Optimal two-dimensional transformation. The result is very good, the CEV smoother is able to accurately find the hyperbolic profile.

symmetric way,

$$\Psi_i = \left| \frac{E\left(f_i \cdot \sum_{j \neq i} f_j\right)}{\sqrt{E(f_i^2) \cdot E\left[\left(\sum_{j \neq i} f_j\right)^2\right]}} \right|, \qquad (13)$$

to characterize the importance of a single term. The use of this measure is demonstrated for the data from the example used in Fig. 5. The values $\Psi_0 = 0.9779$, $\Psi_1 = 0.9272$, $\Psi_2 = 0.9690$ and $\Psi_3 = 0.0426$ are found, confirming again the unimportance of the additional uncorrelated variable, this time by the global measure Ψ_3.

If however, the additional variable is correlated, e.g. by passively following another variable, a high correlation will follow, yielding a model with more components. In the case of spatiotemporal modeling one has to be specially careful, because one is interested in minimal (in the number of terms) models, but dynamical dependencies can result easily in correlations, like in the case of slaved variables.

3.4. *Generalizations*

One can generalize the procedure to include more complex, nonadditive couplings, e.g. in the three-dimensional case, one considers the model

$$U_0 = F(U_1, U_2) + \varepsilon, \qquad (14)$$

with F as some function describing a two-dimensional surface in a three-dimensional space spanned by U_0, U_1, U_2. Like for Eq. (7), one finds

$$F(U_1, U_2) = E(U_0|U_1, U_2). \qquad (15)$$

Going to higher dimensions requires in general more data due to the curse of dimensionality, additionally one has to decide again which estimator to use for correct results.

As an example, the relation $U_0 = U_1 \cdot U_2$ [Fig. 6(a)] is used. Ten thousand equally distributed data points in the range $(-1, 1)$ for U_1, U_2 have been generated, from these U_0 is obtained by multiplication. Gaussian, $N(0, 0.1^2)$, noise has been added afterwards to each of the three variables. The estimated function coincides well with the expectation [Fig. 6(b)]. Note that in this case ACE yields as a result $\log U_0 = \log U_1 + \log U_2$ because of the additive structure of the model (5)

For this model, many applications with quasi-two-dimensional setting can be found, e.g. in geological and hydrodynamical systems (like the two-dimensional Euler equations).

4. Application to Spatiotemporal Data

Typical spatiotemporal models of interest for a physicist are coupled map (CML) systems, coupled ordinary differential equations (ODE's) or partial differential equations (PDE's). In all these models, derivatives or finite differences are involved. Here, the measured quantity shall consist of a space-time record, given by a movie, i.e. a series of consecutive pictures. From the time series at a point, one builds

finite differences in time, from the pictures differences in space can be calculated. This procedure is numerically not trivial: addition (or subtraction, respectively) is numerically poor conditioned and can yield extinction of digits, measurement noise is then shifted to the first digits. Basically, one must use the apparatus known from numerical integration [Teukolsky *et al.*, 1993; Lambert, 1973; Mitchell & Griffiths, 1980] to use correct sampling in space and time. In the case of a periodic setup, accuracy loss can be limited by using spectral methods [Teukolsky *et al.*, 1993; Canuto *et al.*, 1988].

As an example, let us take the Complex Ginzburg–Landau Equation. It describes a well patterned formatting system [Cross & Hohenberg, 1993]:

$$\partial_t A = (1 + ic_1)\Delta A + (1 - ic_3)f_1(|A|^2)A. \quad (16)$$

The function f_1 is a complex nonlinear function. The task of data analysis — suppose that A can be measured — is to determine the nonlinearity f_1. To apply the ideas of backfitting we take $U_0(x, t) = \partial_t A/A$, $U_1(x, t) = \Delta A/A$, $U_2(x, t) = |A|^2$. Real and imaginary parts must be treated separately. Using the iteration procedure, one finally obtains estimates for the parameters c_1, c_3 and the full function f_1. In a slightly more complicated setup, this has been performed in [Voss *et al.*, 1999]. In fact, for this example the numerically critical step is rather the correct evaluation of derivatives than the application of backfitting.

If one wants to find a model, but there are no physical arguments about which derivatives should occur or is it even unclear if the measured quantities are state variables or have undergone some measurement transformations, one can still try to find a generalized additive model. In that case, it can be worthwhile to fit a discrete model (CML), that represents the system dynamics, avoiding a good part of the trouble with finite differences (one can choose mapping units big enough to avoid numerical problems).

A recently presented experiment considers underwater vortex ripples [Andersen *et al.*, 1999; Andersen *et al.*, 2002]. The ripple formation is driven by a periodical motion of water over a sand bed with given amplitude and frequency. A picture of the ripples — viewed from the side — is given in Fig. 7(a), the resulting space time plot is shown in Fig. 7(b). For experimental details, see [Andersen *et al.*, 2002] and [Hansen *et al.*, 2001]. Here, the treatment of the data and subsequent results are
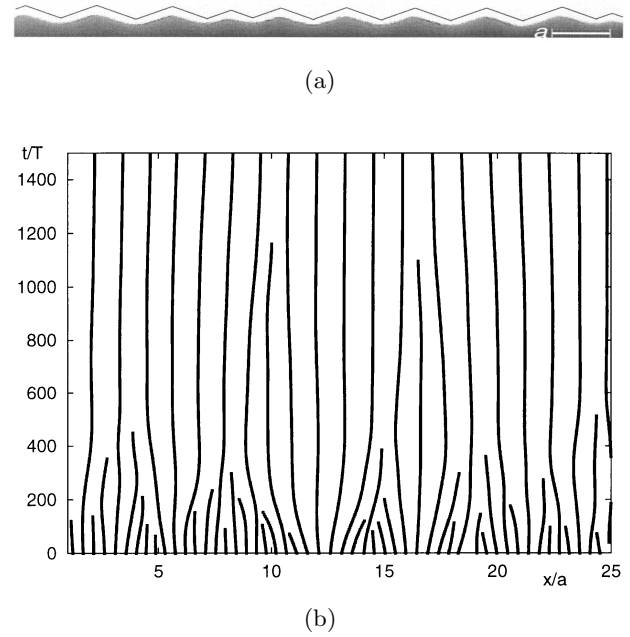

(a)

(b)

Fig. 7. (a) Ripple profile as observed in the experiment. The driving amplitude a is marked. With the bottom at rest, water is oscillating forth and back over the crests. Triangles (top, black line) are fitted to raw picture (lower, dark profile) to determine the crests. (b) The experimental evolution of the position of the ripple crests starting from ripples with a small length yields the shown space time plot.

presented to demonstrate application of the nonparametric regression.

The model is formed by coarse graining of space and time. In time, one uses a mapping from period to period, in space, ripples are characterized by their length λ. The interaction between ripples is modeled by a nonlinear interaction function which must be concave by stability analysis. The model is obtained for ripples close to the final length; without discussing further the physical mechanisms, it reads:

$$\lambda_i(T + 1) - \lambda_i(T) = -f(\lambda_{i-1}(T)) + f(\lambda_i(T))$$
$$- f(\lambda_{i+1}(T)) + \varepsilon. \quad (17)$$

The subscript i denotes the space index. The process of data analysis involves the following steps:

(i) calculating the ripple length; this has been done by fitting triangles to the raw profiles and calculating the distances from crest to crest (cf. Fig. 7). The accuracy amounted to two digits.

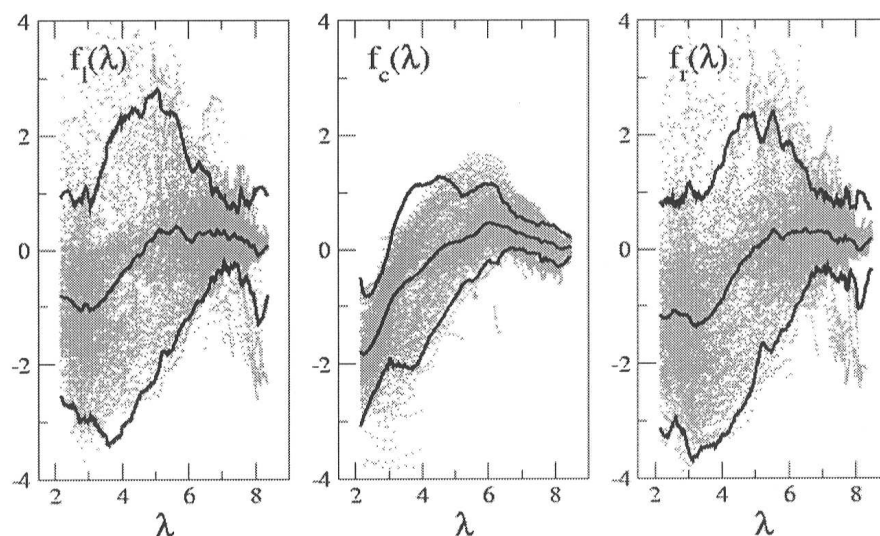(ii) Calculation of the rhs of Eq. (17); this has been done by first calculating the difference,

Fig. 8.   Plot of the functions f_l, f_c, f_r with pointwise standard deviations (black lines) on top of the partial residuals (gray dots). The left and right functions show wide scatter, reflected in the individual correlations (numbers are given in the text). The curves display the expected convex shape. For small ripple length there is a systematic deviation in the left and right functions.

the result has been very noisy due to the numerical extinction of digits. To compensate for this effect, the differences have been smoothed and small numbers have been discarded. Please note that no "unsuitable" points have been excluded (a common, but often wrong technique for outliers), rather we follow a *numerical* requirement. In any case, the model has been developed from considerations of long ripples not too far from the final length and thus very small ripples are not expected to follow the model dynamics.

(iii) To apply a backfitting algorithm, one has to identify the variables U_i as they have been denoted in the previous sections. We formalize as $\Delta_t \lambda = \lambda_i(T+1) - \lambda_i(T) = U_0$, $\lambda_{i-1}(T) = U_1$, $\lambda_i(T) = U_2$, and $\lambda_{i+1}(T) = U_3$. The lhs shall be linear, so the model (3) has been used. For this example, the functions are denoted as $f_1 = f_l$, $f_2 = f_c$, $f_3 = f_r$, for the left, center and right ones, respectively. In the present context, this allows easier interpretation in terms of the spatial structure. The data under consideration have been obtained by nine experimental runs, each between 1000 and 6000 data points.

The experimental runs have been stationary as far as the experimental technique is concerned. Each data point $(\Delta_t \lambda_i(T), \lambda_{i-1}(T), \lambda_i(T), \lambda_{i+1}(T))$ can be considered as an independent realization of

the underlying process. Thus, the data from all nine measurements have been analyzed together. The result is shown in Fig. 8. For each function, the importance criterion is calculated with $\Psi_0 = 0.732$, $\Psi_1 = 0.560$, $\Psi_2 = 0.8604$ and $\Psi_3 = 0.6376$. From these numbers one obtains basically two kinds of information: (1) There is a large part of variance which cannot be modeled. This can have two reasons, the first is measurement noise, which has been reduced as much as possible, the second is that parts of the dynamical equations (higher order interactions or similar) are not included in the model. *A priori* it is not easy (or even impossible) to distinguish which one is the reason, again the error in variables problem has to be treated with care. (2) The growth of a ripple depends mainly on the length of this ripple, but the neighbors cannot be neglected. This is seen by the lower correlations of the neighboring functions.

As a basic test, the resulting functions must show the expected concave shape. This is perfectly true in the range of the model. There are, however differences between the neighbor functions f_l, f_r and the center one, f_c. Mainly one notes a very large standard deviation at certain regions of f_l, f_r and a systematic deviation, a bend, at small ripple lengths. At this point the data analysis result provides feedback for the experiment and the theoretical modeling. It must be clarified if these differences
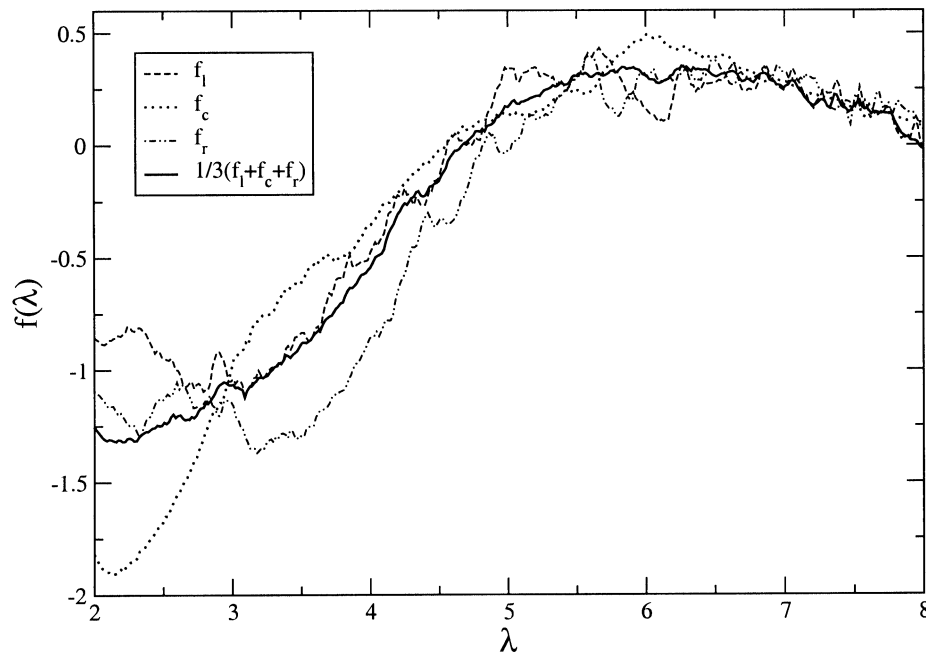
Fig. 9. Results from averaging over different runs (f_l, f_c, f_r). For long ripple lengths — the regime, the model has been developed for — one observes a very good coincidence. Theoretically, a difference between sides and center is not expected; assuming them to be equal, averaging yields the black line as an estimation for the true interaction function.

have a physical background or if they are artificial effects due to the measurement procedure. This is ongoing work.

In Fig. 9 the left, center and right functions are plotted together. Encouragingly, with respect to the established model, the functions fall close together except for small ripple lengths. The model has been developed in the region close to the final length, thus the result confirms the model. For small lengths, the result should be considered as input for further modeling. In the case of small ripple lengths, the deviations of the side functions to the center indicate that in the small ripple length region, a difference between center and sides exists. If one averages over left, center and right functions, assuming that all three functions are equal and differences arise from measurement errors, one obtains the thick black line in Fig. 9.

The model has been obtained as a great simplification of the highly complex granular-fluid system. Given additionally a certain amount of measurement and numerical errors from data processing, one can say that the model is well confirmed by direct, nonparametric data analysis. Further activity aims at a more general modeling in the sense of Eq. (14) to check if there are differences in the interaction functions of the sides and the center

which is theoretically improbable. The direct justification of the model (17) is a success not only for the modeler but for statistical data analysis in physical experiments as well.

5. Discussion and Conclusion

Some of the facts about generalized additive models, backfitting algorithms and ACE have been presented. With a step by step explanation the basic ideas of algorithms and some technical details have been discussed. In the examples, problems with a nonlinear transformation of the data have been pointed out and the effect of additional, uncorrelated variables has been investigated. With data from an example for a complex pattern forming system, it has been shown how to apply the algorithms to spatiotemporal data, and the main difficulties for applications have been explained. With data analysis a simple model has been validated, and developed by physical arguments and coarse graining of the highly complex system of underwater vortex ripples.

Generalized additive models occur often in physical and engineering applications. The idea to obtain models directly from data can be realized with the techniques developed in statistical science.

The focus of the presented methods lies in the possibility to determine nonlinearities. Results from the nonparametric data analysis help the theoretician to improve analytical models, and an applied physicist or engineer who just needs a working model to represent the system dynamics. These aspects can be closed to a circuit — measurement; data analysis; theoretical modeling — which can be iterated until a consistent model for the system under consideration is obtained. It can be expected that the use of nonparametric methods is helpful as well in situations where little is known about the system. There, working models can be inferred without further theoretical justification but as tools.

While in this article additive models have been discussed, there exist techniques to solve for either additive or multiplicative models, e.g. the marginal integration [Linton & Nielsen, 1995] or alternatively Breimans Π-method [Breiman, 1991] (without a claim for completeness). Implementations of the backfitting procedure or the ACE algorithm exist for the S-PLUS language [Hyperlink, 2002a], the GNU R-project [Hyperlink, 2002b], as Matlab routine [Hyperlink, 2002c] or a C++ based GUI [Hyperlink, 2002d].

Statistical approaches are promising tools for the analysis of complex systems, since nonlinearities are automatically fetched. Bayesian formulations promise to yield better error models and another, future application can be stochastic modeling, finding a representation of the residuals in form of a noise term with measured probability distribution.

Acknowledgments

References

Andersen, K., Chabanol, M. & van Hecke, M. [1999] "Dynamical models for sand ripples beneathe surface waves," *Phys. Rev.* **E63**, 66308.

Andersen, K., Abel, M., Krug, J., Ellegaard, C., Soendergaard, L. & Udesen, J. [2002] "Pattern dynamics of vortex ripples in sand: Nonlinear modeling and experimental validation," *Phys. Rev. Lett.* **88**, 4302.

Bell, C. [1962] "Mutual information and maximal correlation as measures of dependence," *Ann. Math. Stat.* **33**, 587–595.

Breiman, L. & Friedman, J. [1985] "Estimating optimal transformations for multiple regression and correlation," *J. Am. Stat. Assoc.* **80**, 580–598.

Breiman, L. [1991] "The π method for estimating multivariate functions from noisy data," *Technometrics* **33**, 125–143.

Buja, A., Hastie, T. & Tibshirani, R. [1989] "Linear smoothers and additve models," *Ann. Stat.* **17**, 453–555.

Buja, A. [1990] "Remarks on functional canonical variates, alternating least squares methods and ACE," *Ann. Stat.* **18**, 1032–1069.

Canuto, C., Hussaini, M., Quarteroni, A. & Zang, T. [1988] *Spectral Methods in Fluid Dynamics* (Springer, NY).

Carroll, R. J., Maca, J. & Ruppert, D. [1999] "Nonparametric regression in the presence of measurement error," *Biometrika* **86**, 541–554.

Cross, M. & Hohenberg, P. [1993] "Pattern formation outside equilibrium," *Rev. Mod. Phys.* **65**, 851–1112.

Csaki, P. & Fischer, J. [1963] "On the general notion of maximal correlation," *Magyar Tud. Akad. Mat. Kutato Int. Kozl.* **8**, 27–51.

Fan, J. & Truong, Y. [1993] "Nonparametric regression with errors in variables," *Ann. Stat.* **21**, 1900–1925.

Hansen, J. L., van Hecke, M., Haaning, A., Ellegaard, C., Andersen, K. H., Bohr, T. & Sams, T. [2001] "Instabilities in sand ripples," *Nature* **410**, 324.

Härdle, W. & Hall, P. [1993] *On the Backfitting Algorithm for Additive Regression Models* (Cambridge University Press, Cambridge).

Hastie, T. & Tibshirani, R. [1986] "Generalized additive models," *Stat. Sci.* **1**, 295–318.

Hastie, T. & Tibshirani, R. [1990] *Generalized Additive Models* (Chapman and Hall, London).

Hastie, T. & Tibshirani, R. [2000] "Bayesian backfitting," *Stat. Sci.* **15**, 196–223.

Honerkamp, J. [1994] *Stochastic Dynamical Systems* (VCH, NY).

Hyperlink [2002a] "The S-PLUS software," http://www.splus.com.

Hyperlink [2002b] "The R-project," http://www.R-project.org/.

Hyperlink [2002c] "Matlab routine," http://www.fdm.uni-freiburg.de/hv/hvsoftware.html.

Hyperlink [2002d] "The ACE algorithms as Qt version," http://www.stat.physik.uni-potsdam.de/~markus/download.

Kantz, H. & Schreiber, T. [1997] *Nonlinear Time Series Analysis* (Cambridge University Press).

Lambert, J. [1973] *Computational Methods in Ordinary Differential Equations* (Springer, NY).

Linton, O. & Nielsen, J. [1995] "A kernel method of estimating structured nonparametric regression based on marginal integration," *Biometrika* **82**, 93–100.

Mammen, E., Linton, O. & Nielsen, J. [1999] "The existence and asymptotic properties of a backfitting projection algorithm under weak conditions," *Ann. Statist.* **27**, 1443–1490.

Mitchell, P. & Griffith, D. [1980] *The Finite Difference Method in Partial Differential Equations* (Wiley, NY).

Montgomery, D. C. [1992] *Introduction to Linear Regression Analysis* (Wiley, NY).

Ott, E., Sauer, T. & Yorke, J. [1994] *Coping with Chaos*, Series in Nonlinear Science (Wiley, NY).

Rényi, A. [1959] "On measures of dependence," *Acta. Math. Acad. Sci. Hungar.* **10**, 441–451.

Rényi, A. [1970] *Probability Theory* (Akadémiai Kiadó, Budapest).

Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. [1993] *Numerical Recipes in C*: *The Art of Scientific Computing*, 2nd edition (Cambridge University Press, Cambridge).

Voss, H. U., Kolodner, P., Abel, M. & Kurths, J. [1999] "Amplitude equations from spatiotemporal binary-fluid convection data," *Phys. Rev. Lett.* **83**, 3422–3425.