ELSEVIER

Physica D 171 (2002) 1–7

PHYSICA D

www.elsevier.com/locate/physd

Fitting parameters in partial differential equations from partially observed noisy data

T.G. Müller [a,b], J. Timmer [a,b,*]

[a] *Freiburg Center for Data Analysis and Modelling (FDM), Eckerstr. 1, D-79106 Freiburg, Germany*
[b] *Fakultät für Physik, Albert-Ludwigs-Universität, Hermann-Herder-Str. 3, D-79104 Freiburg, Germany*

Received 12 October 2001; received in revised form 18 March 2002; accepted 17 May 2002
Communicated by E. Kostelich

Abstract

We describe a new method for parameter estimation in systems with nonlinear spatiotemporal dynamics. The technique is able to reliably determine parameter estimates from noisy data and is even applicable in cases with unobserved components. In a simulation study we investigate its performance in comparison to standard methods and show its superiority. As numerical example we analyze in detail a complex Ginzburg–Landau equation under realistic conditions.

PACS: 05.45.−a; 05.45.Tp

Keywords: Parameter estimation; Partial differential equations; Spatiotemporal systems; Nonlinear dynamics; Complex Ginzburg–Landau equation

1. Introduction

The study of nonlinear pattern-forming systems that can be mathematically described with partial differential equations (PDEs) has drawn much attention in recent years. In most cases, the focus has been on the derivation of amplitude equations based on symmetry considerations, separation-of-scales arguments and other approximations, and the analysis of bifurcation points and chaos [1,2]. To connect the theoretical work to experimental studies, it is necessary to adapt the mathematical model to experiments, since for most systems, not all dynamical parameters are known with sufficient precision. This leads to the problem of estimating dynamical parameters and variables based on data.

Unfortunately, currently available methods require low-noise data where all dynamical variables can be observed [3–7]. These approaches rely on estimating temporal and/or spatial derivatives from the data, sometimes even of higher order, which is critical when the data are corrupted by noise. Bär et al. [5] explicitly state that “noise remains a crucial problem.” To circumvent these problems, we introduce a novel approach based on the multiple shooting technique, which is already a valuable tool in parameter estimation in ordinary differential equations (ODEs) [8–10]. No derivatives have to be estimated from the data, and the method is reliable even when little is known about the parameters and dynamical variables and when the

* Corresponding author. Tel.: +49-761-203-5829; fax: +49-761-203-5967.
E-mail address: jeti@fdm.uni-freiburg.de (J. Timmer).

0167-2789/02/$ – see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0167-2789(02)00546-8

system is only partially observed. Since this approach is developed from the maximum likelihood estimator (MLE), it is possible to compute confidence intervals and use the results for further statistical evaluation [11].

The main purpose of this manuscript is to develop the principal idea of this new technique and apply the approach to an example of the complex Ginzburg–Landau equation (CGLE), which captures the dynamics of an extended system near a codimension-2 bifurcation [1,5,12]. In a simulation study, we compare this technique to standard methods used in the field.

2. Mathematical analysis and algorithmic implementation

For ease of notation, we use only one spatial variable and assume periodic boundary conditions. In more general situations the following derivation also holds.

Let the dynamics of the system be described by the PDE

$$\dot{\mathbf{z}}(t,x) = \mathbf{f}(\mathbf{z}, \mathbf{p}, t, x, \partial_x, \partial_{xx}, \dots)$$

with observed data

$$\mathbf{y}^{\mathrm{D}}(t,x) = \mathbf{g}(\mathbf{z}, t, x) + \eta(\mathbf{z}, t, x)$$

with dynamical variable $\mathbf{z} \in \mathbb{R}^Q$, parameter vector $\mathbf{p} \in \mathbb{R}^P$, time $t \in [t_0, t_{\mathrm{f}}]$ with initial and final time t_0 and t_{f}, spatial variable $x \in [x_{\mathrm{lb}}, x_{\mathrm{rb}})$ with left and right boundary x_{lb} and x_{rb}, observation $\mathbf{y}^{\mathrm{D}}(t,x) \in \mathbb{R}^K$ and noise η. In the following, we assume that η follows a Gaussian distribution with zero mean and standard deviation $\sigma_{\mathrm{noise}} = \sigma_0\sigma_{\mathrm{signal}}$. σ_0 is the noise level, which is a measure of the strength of the noise.

In most cases, some of the dynamical variables can be observed directly, while the others are not observable. Additionally, observed data are sampled in time $(t_i, i = 1, \dots, I)$ and space $(x_j, j = 1, \dots, J)$, so that the observation function $\mathbf{g}$ simplifies to

$$\mathbf{y}_k^{\mathrm{D}}(t_i, x_j) = \mathbf{z}_k(t_i, x_j) + \eta(t_i, x_j),$$
$$k = 1, \dots, K \le Q.$$

To implement the algorithm, it is necessary to parameterize the initial condition $\mathbf{z}(t_0, x)$, e.g., with a suitable parametric function. In the following, we will use L values $\mathbf{z}^l, l = 1, \dots, L$, characterizing the dynamical variables at the initial time t_0 and at discrete locations x_l equally distributed in $[x_{\mathrm{lb}}, x_{\mathrm{rb}})$. L has to be chosen carefully to make sure that the initial condition and the system state in general are approximated sufficiently well to capture the entire dynamics. The parameterization of the system state at time t_0 is denoted by $\mathbf{Z}^0 = \mathbf{Z}(t_0) = (\mathbf{z}^1, \dots, \mathbf{z}^L) \in \mathbb{R}^{Q\times L}$. The general initial condition $\mathbf{z}(t_0, x)$, which is needed to accurately integrate the PDE, can be easily computed from $\mathbf{Z}^0$ with help of spline functions or polynomials. It is noteworthy that the system state at any time t, $\mathbf{z}(t, x)$, can now be approximated by the corresponding parameterized state $\mathbf{Z}(t)$.

Using the MLE to estimate dynamical parameters and initial conditions, an objective function χ^2 has to be minimized, given by

$$\chi^2(\mathbf{p}, \mathbf{Z}^0) = \sum_{i,j,k} \left(\frac{\mathbf{y}_k^{\mathrm{D}}(t_i, x_j) - \mathbf{y}_k^{\mathrm{M}}(\mathbf{Z}^0, \mathbf{p}, t_i, x_j)}{\sigma_{\mathrm{noise}}} \right)^2,$$

whose numerator denotes the distance between the theoretical model trajectory $\mathbf{y}^{\mathrm{M}}$ and the data $\mathbf{y}^{\mathrm{D}}$. Here $\mathbf{y}_k^{\mathrm{M}}(\mathbf{Z}^0, \mathbf{p}, t_i, x_j)$ is the kth component of the solution of the PDE $\mathbf{z}^{\mathrm{M}}(\mathbf{Z}^0, \mathbf{p}, t_i, x_j)$ at time t_i and location x_j with parameterized initial condition $\mathbf{Z}^0$.

Since $\chi^2(\mathbf{p}, \mathbf{Z}^0)$ is highly nonlinearly dependent on $\mathbf{p}$ and $\mathbf{Z}^0$, χ^2 has numerous local minima apart from the global one that corresponds to the true parameters. Therefore, the direct minimization of χ^2, the so-called initial value approach (IVA) that couples the integration of the PDE with standard optimization techniques, generally fails.

To circumvent these difficulties, the problem is formulated in a different way. By introducing a set of time points $(\tau_h, h = 1, \dots, H, t_0 = \tau_0 < \tau_1 < \cdots < \tau_H < \tau_{H+1} = t_{\mathrm{f}})$ and additional variables $\mathbf{Z}^1, \dots, \mathbf{Z}^H, \mathbf{Z}^h \in \mathbb{R}^{Q\times L}$ which are the parameterized estimates of the dynamical variable $\mathbf{z}$ at time points τ_h, $\mathbf{z}(\tau_h, x)$, the problem can be reformulated as a constrained, over-determined multi-point boundary value problem which is solved with a generalized Gauss–Newton method [13]. Now, the aim is to minimize

$$\chi^2(\mathbf{p}, \mathbf{Z}^0, \mathbf{Z}^1, \dots, \mathbf{Z}^H) = \sum_{h,j,k} \sum_{\tau_h \le t < \tau_{h+1}} \left(\frac{\mathbf{y}_k^{\mathrm{D}}(t_i, x_j) - \mathbf{y}_k^{\mathrm{M}}(\mathbf{Z}^h, \mathbf{p}, t_i, x_j)}{\sigma_{\mathrm{noise}}} \right)^2 \quad (1)$$

with the additional constraints

$$\mathbf{Z}^h = \mathbf{Z}^{\mathrm{M}}(\mathbf{Z}^{h-1}, t = \tau_h), \quad h = 1, \dots, H.$$

$\mathbf{Z}^h$ is the parameterized system state at time τ_h and $\mathbf{Z}^{\mathrm{M}}(\mathbf{Z}^{h-1}, t = \tau_h)$ is the solution of the PDE at time τ_h with the parameterized initial condition $\mathbf{Z}^{h-1}$.

$\mathbf{Z}^0$, the parameterized system state at time t_0, is the only initial condition that has to be estimated, since

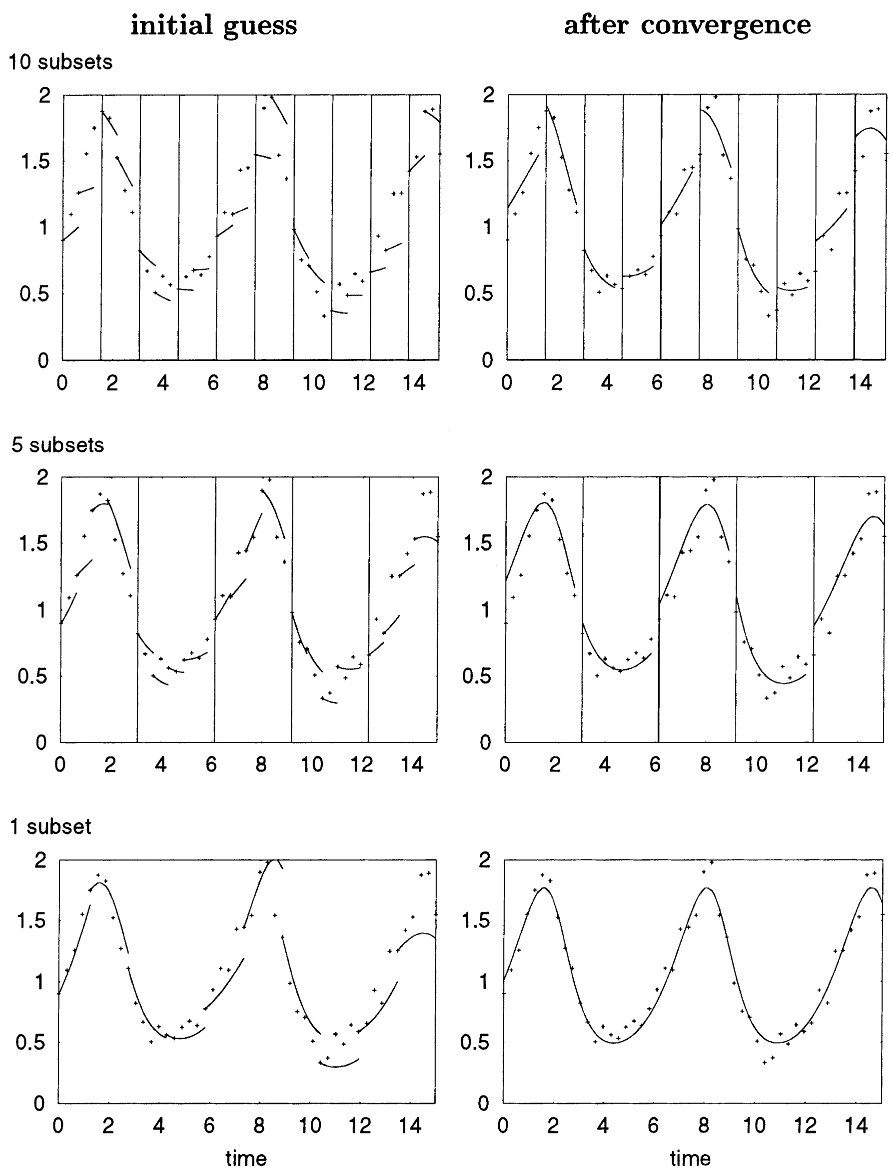

Fig. 1. Demonstration of the fitting method for an ODE: for a decreasing number of subsets the MSM is applied in each subset leading to piecewise continuous trajectories after convergence. By reducing the number of subsets and thus the number of degrees of freedom during the fitting procedure, we obtain a continuous trajectory for the whole data set.

all other unknowns $\mathbf{Z}^h$, $h = 1, \dots, H$, are determined by the continuity constraints, which are applied in a linearized form to allow discontinuous trajectories that help circumvent local minima during the optimization procedure. After convergence, the applied linearized continuity constraints are equivalent to the general continuity constraints [8].

The advantage of this approach is that much more a priori information from the data can be utilized, since initial estimates for the $\mathbf{Z}^h$ can be directly drawn from the observed data $\mathbf{y}^{\mathrm{D}}$. Despite potentially poor initial guesses for the parameters, the model trajectory stays close to the data due to the initially discontinuous trajectory.

Although this approach works in many cases, we found that local minima are still a major problem due to the fact that the continuity constraints are applied too strictly in the multiple shooting method (MSM) (see also [9]). This led us to the generalization of the method in the sense that the data set is divided into several subsets. In each subset, the MSM is applied separately while dynamical parameters are estimated simultaneously. These estimated parameters are then used as initial guesses for a reduced number of subsets. By further reducing the number of subsets until only one data set remains, this method leads to a continuous trajectory.

In principle, we generalize the MSM in the sense that we introduce a multiple shooting hierarchy. Continuity constraints are applied successively to different time scales: in the beginning, independent for every subset and for a small number of time points, and in the end for all data points on a much larger time scale. Each subset requires the estimation of the respective starting values and a high number of subsets is connected to a high number of degrees of freedom. Hence, this procedure slowly reduces the number of degrees of freedom during the parameter estimation process.

The idea of the method is exemplified for the Lotka–Volterra system, a simple ODE [14] (see Fig. 1). Starting with a subdivision of the data set into 10 subsets using two multiple shooting intervals in each subset, the MSM leads to a continuous trajectory for each subset with a discontinuous trajectory for the whole data set. Using the estimated dynamical parameters from the 10 subset situation, the number of degrees of freedom is reduced and parameters are estimated anew for five subsets using four multiple shooting intervals. At the last stage, with one data set and 10 multiple shooting intervals, the algorithm leads to a continuous trajectory.

3. Results

To demonstrate the power of this new technique for PDEs, we analyzed simulated data from a dynamical system of the CGLE-type [5,12]. The PDE of this system reads

$$\dot{\mathbf{z}}_1 = \mathbf{z}_2,$$
$$\dot{\mathbf{z}}_2 = (\mu - \mathbf{z}_1^2)\mathbf{z}_2 - \mathbf{z}_1 - a\mathbf{z}_1^2 - \mathbf{z}_1^3 + \partial_{xx}\mathbf{z}_1 + \kappa\partial_{xx}\mathbf{z}_2$$

with parameter values $\mu = 0.2$, $a = 2.08$ and $\kappa = 1$. Data were simulated with random initial conditions with the method of lines [15] with sufficiently accurate temporal and spatial discretizations so that the a posteriori error was below 0.01 [16]. The system length is 100.0 ($x_{\mathrm{lb}} = 0.0$, $x_{\mathrm{rb}} = 100.0$) while the integration time is 60.0 ($t_0 = 0.0$, $t_{\mathrm{f}} = 60.0$). The number of observed spatial data points per time point was 32 and overall 50 time points were recorded as data. Different levels of observational noise were added and the parameter estimation was carried out. Initial guesses for the starting conditions for every subset were estimated directly from the noisy data with the help of smoothing filters, and initial guesses for the parameters were $\mu = 1.0$, $a = 0.5$ and $\kappa = 5.0$. To make results independent from individual data sets, we used 10 different trajectories all computed with random initial conditions.

In the first simulation study, we investigated the performance of the algorithm when the data (see Fig. 2(a)) are corrupted with noise (Fig. 2(b)). In Figs. 2(c)–(f), the output from the algorithm is shown for the different stages of the minimization procedure. The algorithm starts with a discontinuous trajectory for eight subsets and is close to the data despite poor initial guesses for the parameters (Fig. 2(c)). After estimating dynamical parameters simultaneously in all eight subsets, the number of degrees of freedom is reduced and parameters are estimated with four

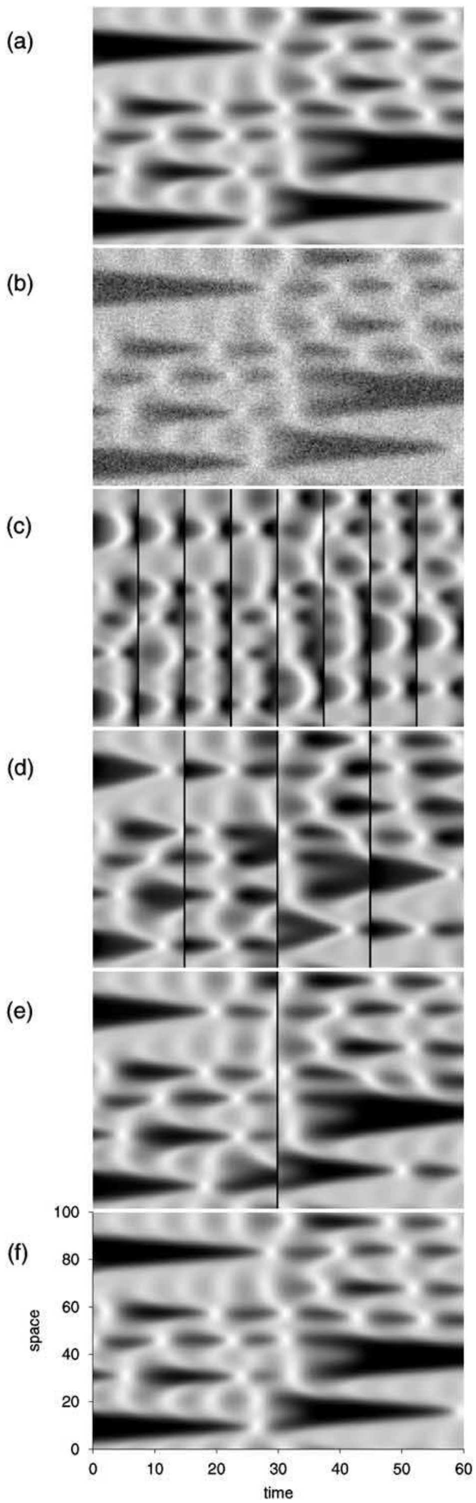

Fig. 2. Different stages of the fitting procedure: (a) true data set; (b) observed data set; (c) initial state with eight subsets; (d) initial state with four subsets; (e) initial state with two subsets; (f) final estimated trajectory.

subsets (Fig. 2(d)) and subsequently with two subsets (Fig. 2(e)). In both the cases, estimated parameters from the previous fitting procedure are used as initial guesses. In the end, Fig. 2(f), the algorithm leads to a continuous trajectory that is very close to the true one.

To make the illustration more concise, multiple shooting intervals are not shown, and for better visualization, the data set is presented in Fig. 2 consists of 400 time points and 128 spatial points. Note that for the estimation procedure, a much smaller number of data points has been used. The results of this study are given in Table 1(b) demonstrating that the approach works well even for high noise levels.

In a second study, to investigate the performance of the algorithm in the case of unobserved components for several noise levels, the dynamical variable $\mathbf{z}_2$ was not observed (see Table 1(c)). As expected, in the case with a high noise level, the algorithm is not able to estimate parameters as reliably as in the first case, since the amount of information is approximately halved, but the results are still reasonably good.

In a third study, we compared the eMSM with the IVA and the MSM without hierarchical shooting. We investigated the performance of all three methods with respect to different noise levels and different initial parameter guesses with both components observed.

We chose five sets of initial guesses for the parameter vector $\mathbf{p}_i = (\mu, a, \kappa)$: $\mathbf{p}_1 = (0.20, 2.08, 1.0)$, being the true parameter vector, $\mathbf{p}_2 = (0.22, 2.00, 1.2)$, $\mathbf{p}_3 = (0.24, 1.92, 1.4)$, $\mathbf{p}_4 = (0.30, 1.80, 1.6)$, and $\mathbf{p}_5 = (0.40, 1.50, 2.0)$. Note that, as in the previous study, the initial parameter guess was $\mathbf{p} = (1.0, 0.5, 5.0)$. As before, the noise levels cover a wide range: $\sigma_0 = 0.01, 0.05, 0.10, 0.25, 0.40$.

To make the results independent from individual data sets, we used 20 different trajectories, all computed with random initial conditions. To present results in an easily accessible way, we do not display the mean and standard deviation of the estimated parameters as before, but rather define the following statistical acceptance measure: based on the covariance matrix at the estimated parameter vector, we calculated 95% confidence intervals. Only if the true parameter value was within this interval was the estimate rated acceptable, since then we expect the estimate to be close

Table 1
(a) True parameters and (b, c) mean value and standard deviation for parameter estimates from 10 different data sets for both/one observed dynamical variables with different noise levels σ_0 (initial guesses for the parameters are $\mu = 1.0$, $a = 0.5$ and $\kappa = 5.0$)

σ_0	μ	a	κ
(a) *True parameters*			
	0.2	2.08	1.0
(b) *Estimated parameters with noise level* σ_0			
0.01	0.2007 ± 0.0053	2.088 ± 0.009	1.001 ± 0.080
0.03	0.1996 ± 0.0051	2.082 ± 0.008	1.035 ± 0.062
0.05	0.2008 ± 0.0040	2.081 ± 0.014	1.018 ± 0.066
0.10	0.1990 ± 0.0065	2.092 ± 0.009	1.033 ± 0.094
0.25	0.2012 ± 0.0125	2.094 ± 0.019	1.088 ± 0.107
0.40	0.1880 ± 0.0472	2.094 ± 0.026	1.140 ± 0.715
(c) *Estimated parameters with one observed component*			
0.01	0.2022 ± 0.0178	2.082 ± 0.007	1.084 ± 0.306
0.03	0.2120 ± 0.0130	2.087 ± 0.010	1.041 ± 0.278
0.05	0.2137 ± 0.0190	2.089 ± 0.018	1.090 ± 0.244
0.10	0.1813 ± 0.0155	2.166 ± 0.025	1.167 ± 0.290
0.25	0.2343 ± 0.0440	2.074 ± 0.022	1.265 ± 0.528
0.40	0.2505 ± 0.0571	2.103 ± 0.037	1.855 ± 1.536

to the global minimum. All other values were rated as local minima. Additionally, if the confidence intervals were too large (> 20% of the parameter value), the estimate was also discarded to avoid coincidental accepted estimates. This was only the case for a few simulations.

The results are presented in Table 2. The IVA clearly has the worst performance of all three techniques, and if initial guesses for the parameter values are above a critical level, it fails completely. The MSM is also not as good as the eMSM since it also often converges into local minima. Higher noise levels lead for all three methods to a decreasing performance.

All computations were performed on 700–866 MHz Pentium III machines. The average computation time for one estimation was 70 min for the IVA, 40 min for the MSM, and 120 min for the eMSM.

Table 2
Results of the performance study using 20 data sets investigating the number of accepted estimates for different noise levels σ_0 and different initial guesses of dynamical parameters $\mathbf{p}_i$ for all three methods (the numbers in each triple refer to the results of IVA/MSM/eMSM; initial parameter guesses $\mathbf{p}_1$ to $\mathbf{p}_5$ are described in the text)

σ_0	Initial parameter guess				
	$\mathbf{p}_1$	$\mathbf{p}_2$	$\mathbf{p}_3$	$\mathbf{p}_4$	$\mathbf{p}_5$
0.01	19/20/20	13/20/20	1/18/19	0/4/20	0/0/17
0.05	11/19/19	12/20/20	1/19/20	0/3/19	0/0/16
0.10	11/20/19	8/19/18	0/17/20	0/2/18	0/0/14
0.25	9/19/18	4/19/19	0/17/18	0/2/14	0/0/13
0.40	6/18/18	1/17/17	0/9/14	0/0/13	0/0/10

4. Conclusions

To summarize our results, we have derived a new technique for estimating parameters and dynamical variables in systems described by PDEs. The high mathematical and computational effort of this new method is necessary to circumvent the problem of accurately estimating spatial and temporal derivatives from noisy data. This new approach is able to deal with cases with noise and unobserved components even with a small number of data points. In comparison to standard methods, it is much more reliable, especially when initial guesses for dynamical parameters are far from the true ones.

We expect this approach to be useful for adapting models of spatiotemporal dynamics to experimental data and for quantitative detection of deviations from theory. Future work will concentrate on applying this new approach to experimental data measured from traveling waves, weak and full spatiotemporal chaos.

Acknowledgements

TGM received financial support from the Deutsche Forschungsgemeinschaft (DFG).

References

[1] P.C. Cross, M.C. Hohenberg, Pattern formation outside of equilibrium, Rev. Mod. Phys. 65 (3) (1993) 853–1088.
[2] I. Aronson, L. Kramer, The world of the complex Ginzburg–Landau equation, Rev. Mod. Phys. 74 (2002) 99–143.
[3] D.P. Vallette, G. Jacobs, J. Gollub, Oscillations and spatiotemporal chaos of one-dimensional fluid fronts, Phys. Rev. E 55 (4) (1997) 4274–4287.
[4] H. Voss, M.J. Bünner, M. Abel, Identification of continuous, spatiotemporal systems, Phys. Rev. E 57 (1998) 2820–2823.
[5] M. Bär, R. Hegger, H. Kantz, Fitting differential equations to space–time dynamics, Phys. Rev. E 59 (1) (1999) 337–342.
[6] H.U. Voss, P. Kolodner, M. Abel, J. Kurths, Amplitude equations from spatiotemporal binary-fluid convection data, Phys. Rev. Lett. 83 (1999) 3422–3425.
[7] U. Parlitz, C. Merkwirth, Prediction of spatiotemporal time series based on reconstructed local states, Phys. Rev. Lett. 84 (9) (2000) 1890–1893.
[8] H.G. Bock, Recent advances in parameter identification techniques for ODE, in: P. Deuflhard, E. Hairer (Eds.), Numerical Treatment of Inverse Problems in Differential and Integral Equations, Vol. 2, Progress in Scientific Computing, Birkhäuser, Basel, 1983.
[9] E. Baake, M. Baake, H. Bock, K. Briggs, Fitting ordinary differential equations to chaotic data, Phys. Rev. A 45 (8) (1992) 5524–5529.
[10] J. Timmer, H. Rust, W. Horbelt, H. Voss, Parametric, nonparametric and parametric modeling of a chaotic circuit time series, Phys. Lett. A 274 (2000) 123–134.
[11] D. Cox, C. Hinkley, Theoretical Statistics, Chapman & Hall, London, 1974.
[12] M. Argentina, P. Coullet, Chaotic nucleation of metastable domains, Phys. Rev. E 56 (3) (1997) 2359–2362.
[13] H. Bock, Numerical treatment of inverse problems in chemical reaction kinetics, in: K. Ebert, P. Deuflhard, W. Jäger (Eds.), Modelling of Chemical Reaction Systems, Springer, Berlin, 1981, pp. 102–125.
[14] J. Murray, Mathematical Biology, Springer, Heidelberg, 1993.
[15] W.F. Ames, Numerical Methods for Partial Differential Equations, Academic Press, San Diego, 1992.
[16] S. Adjerid, J. Flaherty, I. Babuska, A posteriori error estimation for the finite element method-of-lines solution of parabolic problems, Math. Mod. Meth. Appl. 9 (2) (1999) 261–286.