



This is a repository copy of *The identification of complex spatiotemporal patterns using Coupled map lattice model*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/74569/>

Monograph:

Pan, Y. and Billings, S.A. (2006) The identification of complex spatiotemporal patterns using Coupled map lattice model. Research Report. ACSE Research Report no. 923 . Automatic Control and Systems Engineering, University of Sheffield

Reuse

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

The Identification of Complex Spatiotemporal Patterns Using Coupled Map Lattice model

Pan, Y. and Billings, S. A.



Department of Automatic Control and Systems Engineering
University of Sheffield
Sheffield, S1 3JD
UK

Research Report No. 923
May 2006

The Identification of Complex Spatiotemporal Patterns Using Coupled Map Lattice Models

Pan, Y. and Billings, S. A.

Department of Automatic Control & Systems Engineering
Sheffield University,
Sheffield, S1 3JD, UK

May, 2006

Abstract

Many complex and interesting spatiotemporal patterns have been observed in a wide range of scientific areas. In this paper, two kinds of spatiotemporal patterns including spot replication and Turing systems are investigated and new identification methods are proposed to obtain Coupled Map Lattice (CML) models for this class of systems. Initially, a new correlation analysis method is introduced to determine an appropriate temporal and spatial data sampling step procedure for the identification of spatiotemporal systems. A new combined Orthogonal Forward Regression and Bayesian Learning algorithm with Laplace priors is introduced to identify sparse and robust CML models for complex spatiotemporal patterns. The final identified CML models are validated using correlation based model validation tests for spatiotemporal systems. Numerical results illustrate the identification procedure and demonstrate the validity of the identified models.

1 Introduction

Spatiotemporal systems can give rise to many complex and interesting phenomena such as self-organized, oscillatory and chaotic patterns. Initially, spatiotemporal phenomena were observed and studied in chemical reactions which can display periodic or chaotic temporal oscillations and pattern formation. By the early 1920s, Lotka had developed a simple model in terms of partial differential equations, based on two sequential autocatalytic reactions, which can produce sustained oscillations.

The Belousov-Zhabotinsky (BZ Model) reaction was demonstrated experimentally in the early 1950's by the Russian biochemist Boris Belousov and later confirmed by Zhabotinsky with similar findings showing travelling waves [49]. Prigogine and his colleagues in Brussels developed a simple model, named the Brusselator model, which was more chemically realistic than Lotka's simple

model and can show a variety of interesting spatial and temporal phenomena. Over the past decade, some scientists discovered that by changing one or two critical parameters nonlinear chemical dynamics can also exhibit fascinating self-organized patterns such as spiral waves, spatial chaos, crystal lattices and Turing structures [24] [32] [46] [48].

Recently, a lot of attention had been focused on the study of spatiotemporal models of interacting populations in nature [3] [25]. Maron and Harrison [35] investigated extensively a host-parasitoid system relating to the dynamics of the western tussock moth and its natural enemies. The low mobility of tussock moths, their high parasitism rates and the high mobility of the parasitoids altogether can form a self-organized pattern as predicted by Turing theory [37].

Turing patterns are a kind of spatiotemporal system that can produce steady state heterogeneous spatial patterns under certain initial and boundary conditions. This type of phenomena, termed diffusion-driven instability, was first investigated by Turing in 1952 in the field of chemistry. In his seminal paper, Turing demonstrated theoretically that a system of reacting and diffusing chemicals could spontaneously evolve to spatially heterogeneous patterns from an initially uniform state in response to infinitesimal perturbations. Turing patterns were first observed in chemical experiments as late as 1990 [7] [10] [40] by using the Lengyel-Epstein model [29]. Turing systems have also been shown to be able to at least qualitatively imitate many biological patterns such as the stripes of a zebra or spots of a cheetah and even more irregular patterns such as those on leopards and giraffes. Also the skin of exotic fish, butterflies or beetles can be imitated by using the Turing model [2] [41].

Another interesting spatiotemporal behavior of spot self-replication was first observed by Pearson [44] during his intensive investigation of the numerical simulation of the diffusive Gray-Scott system. This pattern is composed of a population of chemical spots that divide into daughter spots upon growth and decay with over-population. The pattern formation of self-replicating spots was experimentally confirmed by Lee et al. [27] [28] in chemical systems. It can also be seen that the dynamics of spots are similar to living cells that are undergoing mitosis. The dynamics of spot self-replication has been theoretically studied in the one, two and three dimensional cases [36]. In addition, Nishiura and Ueyama [38] proposed a theoretical mechanism that drives the replication dynamics from a global bifurcation point of view. The study of the dynamics of spot self-replication was extended in the noise-controlled Gray-Scott model [30].

However, most studies of complex spatiotemporal patterns have focused on the analysis of the complex behaviors using given or analytical derived partial differential equation models. Given the complexity and wide applications of spatiotemporal patterns in many scientific subjects, it is of great importance to be able to model the underlying dynamics of spatially extended systems. In this paper, Coupled Map Lattice (CML) models are used to identify spatiotemporal patterns. The Coupled Map Lattice model was initially introduced in the 1980s by Kaneko [21][22]. The CML model is discrete in time and space and, unlike Cellular Automata (CA), has a continuous state value. CML is a d-dimension

lattice where each site evolves in time through a discrete map which describes the influence of the past state and neighboring sites. A CML can be used to model dynamical behaviors that are normally described by partial differential equations. But the computational efficiency is greatly enhanced when using CML models compared with PDEs. It has been shown that CML models can exhibit complex spatiotemporal behaviors, including chaos, intermittency, traveling waves and Turing patterns [23]. Consequently, CML models have been used to study spatiotemporal systems in a wide class of scientific subjects.

The identification and forecasting of CML models from observed spatiotemporal data has recently been studied using various methods [8] [9] [15] [19] [34] [33] [43] [39]. There are two main methods in studying the identification and forecasting problem for spatiotemporal systems. One approach uses embedding to reconstruct the local model describing the underlying dynamics. The other approach uses the orthogonal least square algorithms to identify the CML model. However, most of these methods have only been applied to relatively simple classes of spatiotemporal systems. In this paper, a new and effective approach for the identification of CML models for complex spatiotemporal patterns is proposed. The new identification method is composed of two parts. First, a new correlation analysis method is introduced to select appropriate sampling intervals of the identification data in the time and space domains. This method is based on the nonlinear auto-correlation functions of the output variables of spatiotemporal systems. The identification of CML models from noisy observed data is achieved using a new Orthogonal Forward Regression algorithm with Bayesian learning theory to provide a sparse model to fit the complex spatiotemporal dynamics with good generalization error.

The paper is arranged as follows. Section 2 gives a general description of the CML model and the identification procedure for spatiotemporal systems using regression methods. Section 3 introduces the new correlation based method to determine accurate sampling values of spatiotemporal data for identification. A new identification method using the OFR algorithm and Bayesian Learning theory with Laplace priors is then proposed to obtain CML models for complex spatiotemporal patterns. Section 4 introduces the correlation based model validation tests for spatiotemporal systems to validate the final identified CML models. Finally, numerical examples using spot-replicating patterns and Turing patterns are included in section 5 to illustrate the application of the new identification methods and to demonstrate the accuracy and effectiveness of identified models for complex spatiotemporal patterns.

2 Problem Formulation

Consider the general form of the stochastic input-output CML model for time and spatially invariant lattice dynamical systems[8] [9]

$$y_i(t) = f(q^{n_y} y_i(t), q^{n_u} u_i(t), q^{n_y} s^{m_y} y_i(t), q^{n_u} s^{m_u} u_i(t)) + \varepsilon_i(t) \quad (1)$$

where $i \in I^d$ is the spatial index of a d -dimensional space and $t \in T$ is the temporal index; $y_i(t)$ and $u_i(t)$ are the output and input variables respectively at lattice i and time t , and $\varepsilon_i(t)$ is an independent zero mean random sequence. $q^n(t)$ is a temporal backward shift operator

$$q^n = (q^{-1}, q^{-2}, \dots, q^{-n}) \quad (2)$$

so that

$$\begin{aligned} q^{n_y} y_i(t) &= (y_i(t-1), y_i(t-2), \dots, y_i(t-n_y)) \\ q^{n_u} u_i(t) &= (u_i(t-1), u_i(t-2), \dots, u_i(t-n_u)) \\ q^{n_\varepsilon} \varepsilon_i(t) &= (\varepsilon_i(t-1), \varepsilon_i(t-2), \dots, \varepsilon_i(t-n_\varepsilon)) \end{aligned} \quad (3)$$

where n_y , n_u , n_ε denote the maximum temporal lags corresponding to output y , input u and the residual sequence ε .

In (1), s^m is a multi valued spatial shift operator

$$s^m = (s^{p^1}, s^{p^2}, \dots, s^{p^m}) \quad (4)$$

where $p^j \in I^d$ is the spatial translation multi index, such that

$$\begin{aligned} s^{m_y} y_i &= (y_{i-p^1}, y_{i-p^2}, \dots, y_{i-p^{m_y}}) \\ s^{m_u} u_i &= (u_{i-p^1}, u_{i-p^2}, \dots, u_{i-p^{m_u}}) \\ s^{m_\varepsilon} \varepsilon_i &= (\varepsilon_{i-p^1}, \varepsilon_{i-p^2}, \dots, \varepsilon_{i-p^{m_\varepsilon}}) \end{aligned} \quad (5)$$

The parameters m_y , m_u and m_ε denote the maximum spatial radius corresponding to the output y , input u and the residual sequence ε . The task of system identification for spatiotemporal systems is to obtain the unknown CML model $f(\cdot)$ from observed data.

In practice, the underlying spatiotemporal system for identification evolves continuously in time over a specific spatial domain. Therefore, it is essential to choose an appropriate sampling time T_t and space grid T_s before the spatiotemporal data for identification is collected. There seem to be virtually no results in the literature relating to this important problem. In many situations, the form of the CML model $f(\cdot)$ is unknown, so it is necessary to expand f using a known set of possible candidate model terms. Equation(1) can also be written in the regression format which is constructed as a linear combination of a finite number of model terms.

$$y_i(t) = \sum_k \theta_{i,k} \varphi_{i,k}(t) + \varepsilon_i(t) \quad (6)$$

Here, model terms $\varphi_{i,k}(t)$ are composed of $q^{n_y} y_i(t)$, $q^{n_u} u_i(t)$, $q^{n_y} s^{m_y} y_i(t)$, $q^{n_u} s^{m_u} u_i(t)$ which represent the influence of past inputs and outputs from both the local and neighboring lattices. Therefore, there are two aspects to the problem of model identification from observed data. The first is to select a proper model structure and the second is to determine the parameters that are associated with the model.

Therefore, the identification task for spatiotemporal systems involves the following steps.

- (a) Choose an appropriate sampling time T_s and spatial grid S_s according to the inherent characteristics of the spatiotemporal data.
- (b) Determine the regressor set. This step involves choosing the nonlinear degree, values of maximum temporal lags for inputs and outputs n_u and n_y and maximum spatial radius m_u and m_y .
- (c) Select the proper model terms that can adequately represents the underlying spatiotemporal system.
- (d) Estimate the parameters associated with the model terms.

Each step in the identification procedure will be discussed in detail in the following sections.

3 The Identification of CML models for Complex Pattern Formation

In this section, a practical approach for selecting appropriate sampling intervals of spatiotemporal data in both the space and time domains is investigated. An Orthogonal Forward Regression algorithm and Bayesian learning theory with Laplace priors is then introduced to select appropriate model terms and to give unbiased estimates of the model parameters.

3.1 The Selection of Sampling intervals in the space-time domain

It is well known that the sampling interval of continuously observed purely temporal data for identification can influence the selection of the model structure and parameter estimation during nonlinear temporal system identification [4]. The maximum lags n_u and n_y also vary as the sampling time T_s changes. Furthermore, if the data are over-sampled, the design matrix can become ill-posed due to the high correlation between successive measurements. On the other hand, if the data for identification are under-sampled, important information will be lost. In this situation, the final derived model is likely to be sensitive to new training data or the noise and consequently cannot generalize well.

Similar problems also arise when identifying spatiotemporal systems but the sampling interval in both the time and space domains need to be considered. There are various methods of determining the sampling time for nonlinear temporal system identification. Most of these methods are based on using mutual information to calculate the correlation of adjacent sampling points. The disadvantages of this approach have been reviewed in [45]. Another typical method which has been applied in this area is to use nonlinear correlation functions to select the sampling time by following a simple procedure [4] [6], where both the linear and nonlinear correlation in the data is taken into account. The main idea behind these methods is that although adjacent points in the data

should not be uncorrelated, the sampling time should be long enough so as to avoid over-correlation. Actually, the selection of the sampling time is often a compromise between the over-correlation and the uncorrelated situations. In this paper, the correlation method based on the linear and nonlinear correlation functions applied in the nonlinear temporal system identification is extended to determine the sampling intervals of data in both the time and space domains of spatiotemporal data.

The correlation functions for spatiotemporal systems are composed of two parts with respect to the sampling intervals in the time and space domain respectively. In this method, N_s samples of the original spatiotemporal data from outputs are randomly selected over both the space and time domain. The correlation functions for the sampling time of spatiotemporal data are defined as follows

$$\begin{aligned}\Phi_{yy}^{(t)}(\tau_t) &= \frac{\sum_{(i,t)=S(0)}^{S(N_s-1)} (y_i(t) - \bar{y})(y_i(t - \tau_t) - \bar{y})}{\sum_{(i,t)=S(0)}^{S(N_s-1)} (y_i(t) - \bar{y})^2} \\ \Phi_{y^2y^2}^{(t)}(\tau_t) &= \frac{\sum_{(i,t)=S(0)}^{S(N_s-1)} (y_i^2(t) - \bar{y}^2)(y_i^2(t - \tau_t) - \bar{y}^2)}{\sum_{(i,t)=S(0)}^{S(N_s-1)} (y_i^2(t) - \bar{y}^2)^2}\end{aligned}\quad (7)$$

The correlation functions for the space domain are defined as

$$\begin{aligned}\Phi_{yy}^{(s)}(\tau_s) &= \frac{\sum_{(i,t)=S(0)}^{S(N_s-1)} (y_i(t) - \bar{y})(y_{i-\tau_s}(t) - \bar{y})}{\sum_{(i,t)=S(0)}^{S(N_s-1)} (y_i(t) - \bar{y})^2} \\ \Phi_{y^2y^2}^{(s)}(\tau_s) &= \frac{\sum_{(i,t)=S(0)}^{S(N_s-1)} (y_i^2(t) - \bar{y}^2)(y_{i-\tau_s}^2(t) - \bar{y}^2)}{\sum_{(i,t)=S(0)}^{S(N_s-1)} (y_i^2(t) - \bar{y}^2)^2}\end{aligned}\quad (8)$$

These are extensions to the temporal results in [4]. In (7) and (8), the vector S_s indicates the selection of the random locations (i_k, t_k) in both the time and the space domain, where

$$S_s = ((i_0, t_0), (i_1, t_1), \dots, (i_{N-1}, t_{N-1})), i_k \in I^d, t_k \in T, k = 0, 1, \dots, N-1 \quad (9)$$

and $\bar{\bullet}$ denotes averaging operation over the specific domain defined by the vector S_s . The quantities τ_t and τ_s denote the sampling shift in the time and space domain respectively. Hence, both the linear and nonlinear correlation between adjacent sampling points in the time and space domains can be measured using the correlation functions (7) and (8). On the basis of the correlation functions (7) and (8), the procedure of determining the sampling intervals in the time and space domains for spatiotemporal systems identification can be carried out based on the following two steps.

(a) Define the following quantities with respect to the time and space domains

$$\tau_{m,t} = \min\{\tau_{t,y}, \tau_{t,y^2}\} \quad (10)$$

$$\tau_{m,s} = \min\{\tau_{s,y}, \tau_{s,y^2}\} \quad (11)$$

where $\tau_{t,y}$ and τ_{t,y^2} are time values of the first minimum of $\phi_{yy}^{(t)}(\tau)$ and $\phi_{y^2y^2}^{(t)}(\tau)$ respectively and $\tau_{s,y}$ and τ_{s,y^2} are space values. In practice, the 95% confidence

limits which are equal to $\pm 1.96/\sqrt{N_s}$ are usually used to replace the minimum values of the above correlation functions

(b) Determine the sampling intervals for the spatiotemporal data in both the time [1] and the space domain following the procedure:

$$\frac{\tau_{m,t}}{20} \leq T_t \leq \frac{\tau_{m,t}}{10} \quad (12)$$

$$\frac{\tau_{m,s}}{20} \leq T_s \leq \frac{\tau_{m,s}}{10} \quad (13)$$

It should be pointed out that although the justification of (12) and (13) is unavailable at present, this simple but effective empirical method does work very well and provides useful information for systems identification, which will be shown in the following numerical simulations.

3.2 Model Selection and Identification Algorithm

Consider a set of possible candidate regressors for CML models. The aim of the identification algorithm is to first select the significant terms from this set and then to estimate the corresponding coefficients so that the identified model can explain the underlying spatiotemporal dynamics. In this paper, the Orthogonal Forward Regression (OFR) algorithm [5] is applied to a set of candidate regressors $\{\varphi_i\}_{i=1}^M$ which are defined in (6). The OFR algorithm involves a stepwise orthogonalization of the regressors and a forward selection of the significant terms based on the Error Reduction Ratio (ERR) criterion [5]. Hence, the significant model terms are selected step by step by comparing the ERRs of all possible model terms. This algorithm also computes the optimal least squares estimates of the term coefficients $\Theta = \{\theta_k\}$.

Equation(6) can be written in the compact format

$$\mathbf{Y} = \Phi\Theta + \varepsilon \quad (14)$$

where Φ is the regression matrix or the design matrix, Θ is the coefficient vector and ε is the residual sequence. After orthogonalization, (14) is converted into

$$\mathbf{Y} = \mathbf{W}\mathbf{g} + \varepsilon \quad (15)$$

where

$$\mathbf{A} = \begin{pmatrix} 1 & a_{1,2} & \dots & a_{1,M_s} \\ 0 & 1 & \vdots & a_{2,M_s} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} \quad (16)$$

and

$$\Phi = \mathbf{W}\mathbf{A}, \mathbf{A}\Theta = \mathbf{g} \quad (17)$$

In (15), \mathbf{W} is the corresponding orthogonal matrix and M_s denotes the number of terms in the final model. Detail of the OFR algorithm for iteratively selecting the model terms and computing the estimates of coefficients using the Error Reduction Ratio (ERR) criterion is given in the Appendix.1.

3.3 The Bayesian Learning Method with Laplace Priors

A basic principle in practical system identification is the parsimonious principle of selecting the simplest possible model that explains the underlying dynamics. The OFR algorithm is an effective and practical learning procedure for identifying nonlinear regression models. An important feature of the OFR algorithm is the capability of selecting model terms according to the contribution each term makes to the overall model accuracy and the elimination of redundant terms.

It is well known that under certain conditions the least squares algorithm is theoretically equivalent to the maximum likelihood and prediction error methods and that finding the maximum likelihood parameters may be an ill-posed problem. For example, if the data is noisy, the final model may fit to the noise. If the data is not properly sampled, the design matrix may be ill-conditioned and the identified model can be over-fit. Over-fitting can become more severe for complex models involving high-dimensional real world data such as images, speech and spatiotemporal patterns. A useful technique for solving the problem of over-fitting and improving the robustness of the final model is to exploit regularization methods to reduce the effects due to intrinsic ill-conditioning of the problem or due to noise on the data. Two kinds of regularization methods have been used, one is the L_2 regularization and the other L_1 regularization. In the framework of Bayesian regularization methods, these two kinds of regularisers are also referred to as the Gaussian priors and the Laplace Priors methods.

Almost all regularization algorithms proposed so far for regression model include some hyper parameters associated with the model parameters. Appropriately determining the values of these hyper parameters is crucial for good approximation. The problem of how to determine the optimal values of the hyper parameters has been extensively studied [26]. These techniques include the Discrepancy Principle [13], generalized cross-validation [14] and the L-curve methods [20]. But all these methods are computationally costly. The Bayesian method provides an alternative solution to these problems. The hyper parameters used in the Bayesian method are defined in terms of the noise variance and the measurement of the smoothness of the model fit. The Bayesian method allows the user to objectively assign values to the tuning parameters which are commonly unknown a priori. A typical advantage of the Bayesian learning method is that it can quantitatively rank a whole class of models by evaluating the corresponding evidence, the hyper parameters are consequently tuned to maximize the evidence.

There are two kinds of regularisers in the Bayesian method, the Gaussian prior and the Laplace prior. The main disadvantage of Gaussian priors is that they do not control the structural complexity of the learned function by setting coefficients of irrelevant terms to zero [12] [47]. In this paper, the Bayesian learning method with Laplace priors is adapted with the OFR algorithm for nonlinear system identification. This new method has two main advantages, one is that the important model term selection procedure of the OFR algorithm is maintained and the contribution of the individual regularisers to evidence of the regression model can be evaluated by orthogonalizing the candidate regres-

sors. The other advantage is that the optimal regularisers can be inferred by maximizing the evidence in the Bayesian learning framework.

In the Bayesian framework, the optimal estimates of parameters for the regression model (15) are obtained by maximizing the posterior probability of the parameters $\mathbf{g} = (g_1, \dots, g_{M_s})$ which is given by

$$P(\mathbf{g}|\mathbf{Y}, \mathbf{\Lambda}, \epsilon) = \frac{P(\mathbf{Y}|\mathbf{g}, \mathbf{\Lambda}, \epsilon)P(\mathbf{g}|\mathbf{\Lambda}, \epsilon)}{P(\mathbf{Y}|\mathbf{\Lambda}, \epsilon)} \quad (18)$$

where $P(\mathbf{Y}|\mathbf{g}, \mathbf{\Lambda}, \epsilon)$ is the ikelihood function, $P(\mathbf{g}|\mathbf{\Lambda}, \epsilon)$ is the priori density with regularisers $\mathbf{\Lambda} = (\lambda_1, \dots, \lambda_{M_s})$ and $\epsilon = 1/\sigma_{\mathbf{e}}^2$ which denotes the smoothness of the fitted regression model and the noise model of the data respectively, $P(\mathbf{Y}|\lambda, \epsilon)$ is the density representing evidence of the regression model associated with the regularisers $\mathbf{\Lambda}$ and ϵ .

Here, it is assumed that the residual sequence \mathbf{e} is zero mean gaussian noise with standard deviation $\sigma_{\mathbf{e}}$. Following [31], the likelihood can be therefore described as

$$P(\mathbf{Y}|\mathbf{g}, \mathbf{\Lambda}, \epsilon) = \left(\frac{\epsilon}{2\pi}\right)^{N/2} \exp\left(-\frac{\epsilon \mathbf{e}^T \mathbf{e}}{2}\right) \quad (19)$$

The Laplace prior adopted on the parameters \mathbf{g} can be written as the following function

$$P(\mathbf{g}|\mathbf{\Lambda}, \epsilon) = \prod_{i=1}^{M_s} \frac{\lambda_i}{2} \exp(-\lambda_i |g_i|) \quad (20)$$

maximizing the log posterior probability with Laplace priors with respect to \mathbf{g} is equivalent to minimizing the following cost function.

$$\mathbf{J}_{BL}(\mathbf{g}, \mathbf{\Lambda}, \epsilon) = \epsilon \mathbf{e}^T \mathbf{e} + 2\mathbf{\Lambda}|\mathbf{g}| \quad (21)$$

The optimal values of g_i to maximize the log posterior probability is obtained by setting $\partial \log(P(\mathbf{Y}|\mathbf{g}, \mathbf{\Lambda}, \epsilon)P(\mathbf{g}|\mathbf{\Lambda}, \epsilon))/\partial g_i = 0$, which yields (refer to Appendix.2)

$$g_i = \text{sgn}(w_i^T \mathbf{Y}) \left(|w_i^T \mathbf{Y}| - \frac{\lambda_i}{\epsilon \|w_i\|_2} \right)_+ \quad (22)$$

where $\|\mathbf{v}\| = \sum_i v_i^2$ denotes the squared Euclidean norm, $(\cdot)_+$ is the *positive part operator* (defined as $(a)_+ = a$, if $a \geq 0$, and $(a)_+ = 0$, if $a < 0$), and $\text{sgn}(\cdot)$ is the sign function. Note that when the absolute value of $w_i^T \mathbf{Y}$ is below a threshold, the estimate of g_i is exactly zero; otherwise, the estimate is obtained by subtracting a threshold. This rule is also called a *soft threshold*.

Using the Gaussian approximation method, the log evidence of the model with the Laplace regularisers $\mathbf{\Lambda}$ and ϵ can be approximated as

$$\begin{aligned} \log(P(\mathbf{Y}|\mathbf{\Lambda}, \epsilon)) &\approx \sum_{i=1}^{M_s} \log(\lambda_i/2) - \frac{M_s}{2} \log(\pi) - \frac{N}{2} \log(2\pi) + \frac{N}{2} \log(\epsilon) \\ &\quad - \sum_{i=1}^{M_s} (\lambda_i |g_i|) - \frac{1}{2} \epsilon \mathbf{e}^T \mathbf{e} - \frac{1}{2} \log(\det(\mathbf{B})) + \frac{M_s}{2} \log(2\pi) \end{aligned} \quad (23)$$

where \mathbf{g} is set to be the optimal value of a posterior probability solution, and the Hessian matrix \mathbf{B} is diagonal due to the orthogonalization of the design matrix and is given by

$$\mathbf{B} = \epsilon \mathbf{W}^T \mathbf{W} = \text{diag}\{\epsilon w_1^T w_1, \dots, \epsilon w_{M_s}^T w_{M_s}\} \quad (24)$$

Setting $\partial \log(P(\mathbf{Y}|\mathbf{A}, \epsilon))/\partial \epsilon = 0$ yields the computation formula for the optimal ϵ

$$\epsilon = (N - M_s)/\mathbf{e}^T \mathbf{e} \quad (25)$$

Setting $\partial \log(P(\mathbf{Y}|\mathbf{A}, \epsilon))/\partial \lambda_i = 0$ yields the computation formula for the optimal λ_i

$$\lambda_i = \frac{1}{|g_i|} \quad (26)$$

For a large sample of data, the variance of the optimal estimate of the residual usually changes slightly, so the influence of the noise prior on the parameter \mathbf{g} can be ignored. The optimal estimate of the parameter g_i with the optimal Laplace prior λ_i can be therefore written as

$$g_i^{(BL)} = \text{sgn}(w_i^T \mathbf{Y}) \left(|w_i^T \mathbf{Y}| - \frac{1}{\epsilon \|w_i\|_2 |g_i|} \right)_+ \quad (27)$$

The procedure for the OFR algorithm combined with the Bayesian learning method and Laplace priors can be briefly summarized as follows.

- (a) Use the OFR algorithm described in Appendix.1 to select the significant model terms from the candidate terms and give an initial maximum likelihood estimate of the parameter \mathbf{g} .
- (b) Calculate the value of the Laplace priors \mathbf{A} and the noise prior ϵ using (26) and (25) using the parameter \mathbf{g} .
- (c) Calculate the optimal estimate of the parameter $\mathbf{g}^{(BL)}$ using (27) with the optimal value of the Laplace priors.

3.4 Identification Procedure for Spatiotemporal Patterns

In summary, the new identification approach for identifying CML models of complex spatiotemporal patterns described in the previous sections can be summarized in the following steps.

- (a) Using the new data sampling method for imaged or observed spatiotemporal patterns, the continuously evolving spatiotemporal data should be discretized in both the space and time domains to avoid possible ill-posed identification problems.
- (b) A subset of the spatiotemporal data for identification are randomly chosen over the space and time domains in order to make sure that the selected data fully represent the underlying spatiotemporal dynamics.
- (c) When the data for identification are ready and the candidate terms which include the time lagged input and output variables from both the local lattice and neighboring lattices are determined, the identification of the CML models

for spatiotemporal systems can be converted into the problem of nonlinear system identification.

(d) Using the OFR algorithm with the Bayesian learning method, the significant terms of the final model are selected and their corresponding coefficients can be regulated to approach their optimal values to achieve a best model performance.

4 Model Validation

The purpose of model validation is to validate the correctness of the model structure and the unbiasedness of the estimated model parameters. In this paper, a model validation method for spatiotemporal systems recently proposed in [42] will be used to check the final identified CML models. This model validation approach for spatiotemporal systems is based on the assumption that the model residuals at different spatial sites and/or at different times are randomly distributed with finite variances and are independent of each other. Therefore, when the identified model is correct, the model residuals should not be dependent on past inputs and outputs from both the local site and neighbouring sites. Two higher order correlation functions are introduced to detect possible linear and nonlinear terms in the model residuals. In the context of spatiotemporal systems, the inputs for the correlation tests do not only include the input variables but also the input and output variables from neighbouring sites.

Consider the CML (1) model for spatiotemporal systems, the one step ahead predicted output $\hat{y}_i(t)$ is defined as

$$\begin{aligned} y_i(t) &= f(q^{n_y} y_i(t), q^{n_u} u_i(t), q^{n_y} s^{m_y} y_i(t), q^{n_u} s^{m_u} u_i(t)) + \varepsilon_i(t) \\ &= \hat{y}_i(t) + \varepsilon_i(t) \end{aligned} \quad (28)$$

According to [42], validity tests based on correlation functions about the inputs, one-step-ahead predicted outputs and residuals of a spatiotemporal system can be defined as follows.

$$\begin{aligned} \phi_{\beta \varepsilon^2}(\tau) &= \frac{\sum_{(i,t)=S_v(0)}^{S_v(N_v-1)} \beta_i^0(t) \varepsilon_i^{20}(t-\tau)}{\left[\left(\sum_{(i,t)=S_v(0)}^{S_v(N_v-1)} (\beta_i^0(t))^2 \right) \left(\sum_{(i,t)=S_v(0)}^{S_v(N_v-1)} (\varepsilon_i^{20}(t))^2 \right) \right]^{1/2}} \\ \phi_{\beta u^2}(\tau) &= \frac{\sum_{(i,t)=S_v(0)}^{S_v(N_v-1)} \beta_i^0(t) u_i^{20}(t-\tau)}{\left[\left(\sum_{(i,t)=S_v(0)}^{S_v(N_v-1)} (\beta_i^0(t))^2 \right) \left(\sum_{(i,t)=S_v(0)}^{S_v(N_v-1)} (u_i^{20}(t))^2 \right) \right]^{1/2}} \end{aligned} \quad (29)$$

In (29), N_v is the number of data that are randomly sampled for model validation and the vector S_v indicates the selection of the random locations (i_k, t_k) of the data used for model validation in both the time and space domains. $\beta_i^0(t)$, $\varepsilon_i^{20}(t)$ and $u_i^{20}(t)$ are the normalized variables defined as follows

$$\begin{aligned} \beta_i^0(t) &= \frac{\beta_i(t)}{\left[\frac{1}{N_v} \sum_{(i,t')=S_v(0)}^{S_v(N_v-1)} \beta_i^2(t') \right]^{1/2}} \\ \varepsilon_i^{20}(t) &= \frac{\varepsilon_i^2(t) - \bar{\varepsilon}^2}{\left[\frac{1}{N_v} \sum_{(i,t')=S_v(0)}^{S_v(N_v-1)} (\varepsilon_i^2(t') - \bar{\varepsilon}^2) \right]^{1/2}} \\ u_i^{20}(t) &= \frac{u_i^2(t) - \bar{u}^2}{\left[\frac{1}{N_v} \sum_{(i,t')=S_v(0)}^{S_v(N_v-1)} (u_i^2(t') - \bar{u}^2) \right]^{1/2}} \end{aligned} \quad (30)$$

where $\beta_i(t)$ is a normalized compound variable which is a function of the residual $\varepsilon_i(t)$ and one-step-ahead predicted output $\hat{y}_i(t)$.

$$\beta_i(t) = \frac{\hat{y}_i(t)\varepsilon_i(t) - \overline{\hat{y}\varepsilon}}{\left[\frac{1}{N} \sum_{(i,t')=S_v(0)}^{S(N_v-1)} (\hat{y}_i(t')\varepsilon_i(t') - \overline{\hat{y}\varepsilon})^2\right]^{1/2}} + \frac{\varepsilon_i(t)\varepsilon_i(t) - \overline{\varepsilon^2}}{\left[\frac{1}{N_v} \sum_{(i,t')=S_v(0)}^{S(N_v-1)} (\varepsilon_i(t')\varepsilon_i(t') - \overline{\varepsilon^2})^2\right]^{1/2}} \quad (31)$$

According to the Central Limit Theorem, if the identified model for a spatiotemporal system is correct which means that the model residual $\varepsilon_i(t)$ is unpredictable from all past inputs and outputs, for sufficiently large N_v , the estimates of the correlation functions in (29) should fall within the 95% confidence intervals which are approximately $\pm 1.96/\sqrt{N_v}$.

5 Numerical Examples

5.1 Spot Replicating Patterns

Consider the Gray-Scott equation which describes a cubic autocatalytic chemical reaction in an open spatial reactor [16] [17] [18].

$$\begin{aligned} \dot{u} &= D_u \nabla^2 u - uv^2 + F(1 - u) \\ \dot{v} &= D_v \nabla^2 v + uv^2 - (F + k)v \end{aligned} \quad (32)$$

In (32), u and v represent the dimensionless concentrations of the reactant and autocatalysts, the parameter F denotes the dimensionless feed rate and k the dimensionless rate constant of the second reaction. $\nabla^2(\cdot)$ is the 2nd order spatial Laplace operator and D_u, D_v are the diffusive coefficients for the fast and slow variables respectively. This Gray-Scott model was numerically simulated and studied within a two dimensional spatial lattice by Pearson [44]. It was shown that under finite-amplitude perturbations the Gray-Scott model can show a variety of complex behaviors by changing two critical parameters F and k . In this example, the system size of the Gray-Scott model is 0.5×0.5 with a mesh size 50×50 and other parameters for the spot replicating patterns to appear are set as $F = 0.02$, $k = 0.059$, $D_u = 2 \times 10^{-5}$ and $D_v = 1 \times 10^{-5}$.

The Gray-Scott was simulated with the initial trivial state ($u = 1$ and $v = 0$), with the central mesh points initialized ($u = 1/2$ and $v = 1/4$ corrupted by random noise) as the perturbation. The boundary conditions are zero flux (Neumann) boundary conditions. The system was then numerically integrated for 9000 steps with the time step set to be 0.1 using a 4th order Runge-Kutta method and the spatial derivatives were approximated by the central difference equation. The spatiotemporal data was then down sampled to the order of 10 in the time domain to reduce the data size so the time step was regulated to be 1.

For the reason of simplicity, in this and the following example, only the first subsystem will be investigated. The correlation analysis method introduced in Section 2.1 was first applied to choose the sampling interval in the time and

space domains. The correlation functions defined in (7) and (8) were computed and the numerical results are shown in Fig.(1) and Fig.(2).

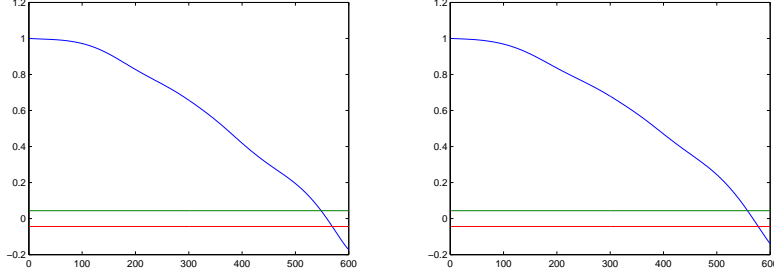


Figure 1: Correlation functions $\phi_{yy}^{(t)}(\tau_t)$ (left) and $\phi_{y^2y^2}^{(t)}(\tau_t)$ (right) in (7) calculated from $N_s = 2000$ random data samples of the u -subsystem of the spot replicating system in Example 1

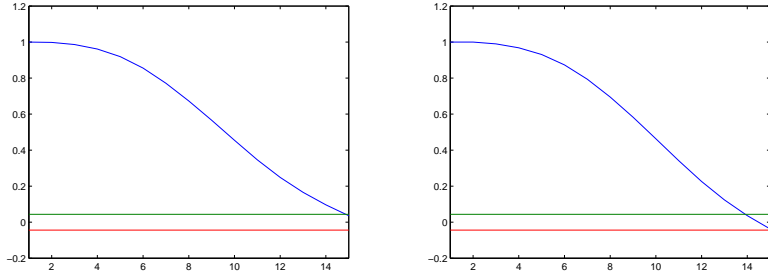


Figure 2: Correlation functions $\phi_{yy}^{(s)}(\tau_s)$ (left) and $\phi_{y^2y^2}^{(s)}(\tau_s)$ (right) in (8) calculated from $N_s = 2000$ random data samples of the u -subsystem of the spot replicating system in Example 1

From the Figs.(1) and (2), the time values when the correlation functions $\phi_{yy}^{(t)}(\tau_t)$ and $\phi_{y^2y^2}^{(t)}(\tau_t)$ first cut the 95% confidence boundary are 550 and 560 time steps. So we can set $\tau_{m,t} = 550$. Similarly, we get $\tau_{m,s} = 14$. According to the selection rule (12) and (13) proposed in Section 2.1, the sampling intervals for the spot replicating system in the time and space domains can be chosen between $\tau_{m,t}/20$ and $\tau_{m,t}/10$. Therefore, T_t is chosen to be 30 which is equal to $\tau_{m,t}/18.3$. Similarly, T_s is chosen to be 1 which is equal to $\tau_{m,s}/14$. So the size of the spot replicating system for identification is regulated as $50 \times 50 \times 30$, where 50×50 is the spatial size that is kept invariant from the original size and 30 is the time which is equivalent to $90/T_t$.

To consider a more realistic situation, the spatiotemporal data of the spot

replicating system were corrupted with normally distributed noise with a standard deviation $\sigma_u = 0.0581$ and $\sigma_v = 0.0058$. The identification method proposed in the previous Section was then applied. The maximum temporal lags and spatial radius were set as $n_u = 1$, $n_v = 1$ and $m_u = 1$, $m_v = 1$. There are three inputs in the regression model (6), $u_{i,j}(t)$, $v_{i,j}(t)$ and $u_{i,j}^*(t)$, where the variable $u_{i,j}^*(t) = u_{i-1,j}(t) + u_{i+1,j}(t) + u_{i,j-1}(t) + u_{i,j+1}(t)$ is a combination of the outputs of neighbouring sites to ensure a symmetric topology and a simpler set of candidate model terms. The initial nonlinearity degree of the candidate model terms for identification was set to be 3. The identification results are shown in Table(1). Snap shots of the system output at different times and the model predicted output of the u -subsystem are displayed in the Fig.(3). Note that in this example only the u -subsystem was investigated, so the output of the v -subsystem is treated as an input during the calculation of the model predicted output of the u -subsystem. The model predicted output of the u -subsystem is defined as

$$u_{i,j}^{(mpo)}(t) = \hat{f}(u_{i,j}^{(mpo)}(t-1), u_{i,j}^{*(mpo)}(t-1), v_{i,j}(t-1), v_{i,j}^*(t-1)) \quad (33)$$

where $\hat{f}(\cdot)$ is the identified CML model. From Fig.(3), it can be seen that the identified CML model predicts very well and the spot replicating process is closely repeated. The model validation tests were applied to validate the final CML model, where $N_v = 1600$ samples of data were randomly selected. The results are shown in the Fig.(4). It can be seen that both the correlation functions of model validation tests $\phi_{\beta\varepsilon^2}(\tau)$ and $\phi_{\beta u^2}(\tau)$ fall within the 95% confidence intervals.

Table 1: Terms and parameters of the identified CML model for u -subsystem in Example 1

Model terms	Estimated parameters without regularisation	Estimated parameters with regularisation	ERR
$u_{i,j}^*(t-1)$	0.0809	0.0830	0.9903
$u_{i,j}(t-1)$	2.0083	1.9907	0.0003
$u_{i,j}^2(t-1)v_{i,j}(t-1)$	-3.6401	-3.6048	0.0002
$u_{i,j}^2(t-1)$	-2.4171	-2.3997	0.0016
$u_{i,j}^3(t-1)$	1.0525	1.0439	0.0002
$u_{i,j}(t-1)v_{i,j}^2(t-1)$	-2.0638	-2.0862	0.0003
$v_{i,j}^3(t-1)$	-1.1913	-1.1671	0.00004

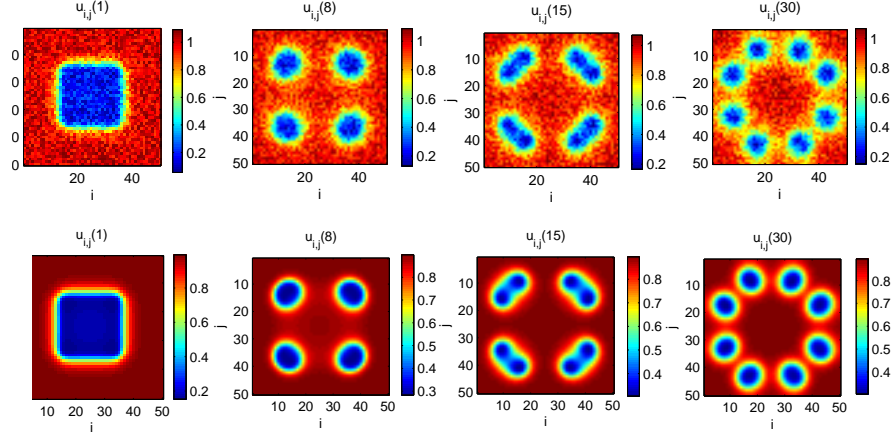


Figure 3: Snap shots of the noisy system output (top) and model predicted output (bottom) at different times for the u -subsystem in Example 1

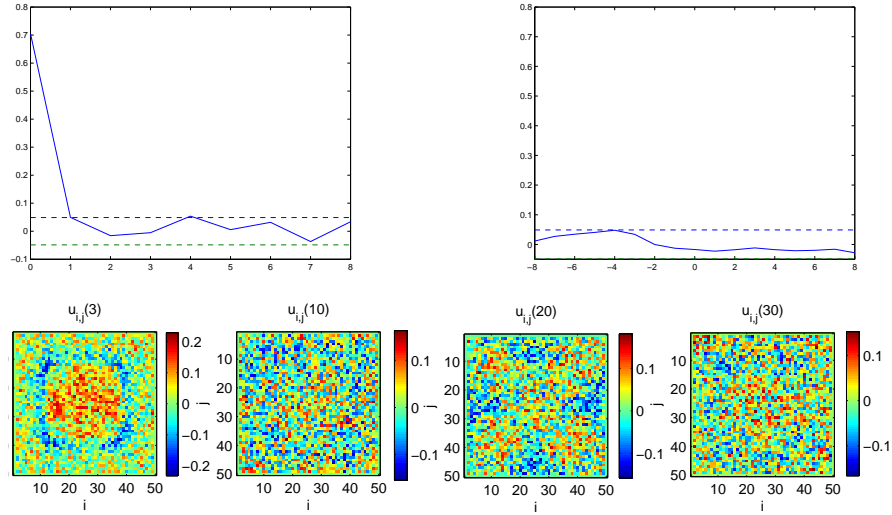


Figure 4: Correlation functions of the validation tests using the final identified model of the u -subsystem of the spot replicating systems in Example 1, (top left) $\phi_{\beta\epsilon^2}(\tau)$, (top right) $\phi_{\beta u^2}(\tau)$ and (bottom) some snap shots of the OSA prediction error at different times.

5.2 Turing Patterns

Consider the two-dimensional BVAM (Barrio-Varea-Aragon-Maini) model that is described by the following coupling partial differential equations[2]

$$\begin{aligned} \dot{u} &= D_u \nabla^2 u - \eta(u + av - Cuv - uv^2) \\ \dot{v} &= D_v \nabla^2 v + \eta(bv + hu + Cuv + uv^2) \end{aligned} \quad (34)$$

The BVAM model was devised as a formal or phenomenological Turing model and it is not based on any real chemical reactions. The terms uv and uv^2 describe the nonlinear inhibition of the activator chemical u by the inhibitor chemical v while the nonlinear term uv^2 is not included since it would describe the reverse behavior. The parameters defining the reaction kinetics were set as $a = 1.112$, $b = -1.01$, $\eta = 0.450$, $h = -1$ and $C = 1.57$. In this example, the diffusion coefficients were set as $D_u = 0.516$, $D_v = 1$ so that $D_u/D_v < 1$ for Turing instability to occur.

The initial concentration distribution corresponds to random perturbations around the trivial stationary state $(u_0, v_0) = (0, 0)$ in the BVAM model with a variance significantly lower than the amplitude of the final patterns. The boundary conditions for (34) were chosen to be zero-flux in this example. The system size of the BVAM model is 32×32 with a mesh size 64×64 . The system was numerically simulated for 192000 steps with the time step set to be 0.05 using the Euler method and the 2nd order spatial derivatives were approximated by the center difference equation. The original data was then down sampled in the time domain from 192000 to 2400 in order to reduce the overhead of computations.

To determine appropriate sampling intervals of the spatiotemporal data for identification, the correlation functions defined in (7) and (8) were calculated. It should be pointed out that the data for the correlation analysis are randomly sampled within the period from $t = 0$ to $t = 1000$ in the time domain since the data in this period are used for identification and the remainder is used for prediction tests. The numerical results are shown in Fig.(5.2) and Fig.(5.2).

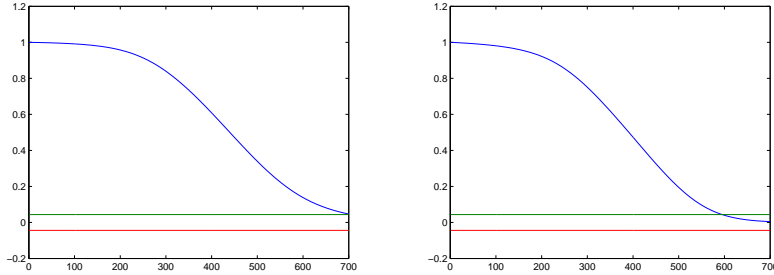


Figure 5: Correlation functions $\phi_{yy}^{(t)}(\tau_t)$ (left) and $\phi_{y^2y^2}^{(t)}(\tau_t)$ (right) in (5.1) calculated from $N_s = 2000$ random data samples of the u -subsystem for the Turing system in Example 2

From the Figs.(5) and (6), the time values when the correlation functions $\phi_{yy}^{(t)}(\tau_t)$ and $\phi_{y^2y^2}^{(t)}(\tau_t)$ cut the 95% confidence boundary are 700 and 595 time steps. Therefore $\tau_{m,t} = 595$. Similarly, $\tau_{m,s} = 14$. According to the selection rule (12) and (13) proposed in Section 2.1, the sampling intervals for the Turing system in the time and space domains can be chosen between $\tau_{m,t}/20$ and

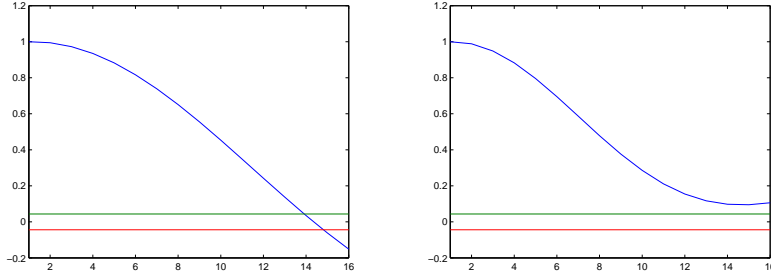


Figure 6: Correlation functions $\phi_{yy}^{(s)}(\tau_s)$ (left) and $\phi_{y^2y^2}^{(s)}(\tau_s)$ (right) in (5.1) calculated from $N_s = 2000$ random data samples of the u -subsystem for the Turing system in Example 2

$\tau_{m,t}/10$. Therefore, T_t is chosen to be 32 which is equal to $\tau_{m,t}/18.6$. Similarly, the sampling interval in the space domain can be chosen between $\tau_{m,s}/20$ and $\tau_{m,s}/10$. And T_s is chosen to be 1 which is equal to $\tau_{m,s}/14$. So the size of the Turing system for identification is regulated to be $64 \times 64 \times 70$, where 64×64 is the spatial size which is kept invariant from the original size and 70 is the time scale which is equivalent to $2240/T_t$.

The data from the Turing system were corrupted with normally distributed noise with standard deviations $\sigma_u = 0.0403$ and $\sigma_v = 0.0058$. The maximum temporal lags and spatial radius of the candidate model terms were set as $n_u = 1$, $n_v = 1$ and $m_u = 1$, $m_v = 1$. Similar to the first example, there are three inputs in the regression model (6), $u_{i,j}(t)$, $v_{i,j}(t)$ and $u_{i,j}^*(t)$, where the variable $u_{i,j}^*(t) = u_{i-1,j}(t) + u_{i+1,j}(t) + u_{i,j-1}(t) + u_{i,j+1}(t)$ is a combination of the outputs of neighbouring sites. The nonlinearity degree of the candidate model terms is set to be 3. The final identified CML model for the u -subsystem is shown in Table(1). Fig.(7) shows some snap shots of the system output and model predicted output of the u -subsystem at different times, while the model predicted output of the u -subsystem is calculated in the same way as (33) where the v -subsystem is treated as an input.

The identification data of the Turing system ($N = 1600$) were sampled from $t = 0$ to $t = 30$ in the time domain and the data after $t = 30$ were used as the new data to test the prediction of the identified model. The identification results are shown in Table(2). Fig.(7) shows some snap shots at different times of the system output and model predicted output of the identified model in Table(2). In this example, only the u -subsystem was investigated, so the output of the v -subsystem was treated as an input during the calculation of the model predicted output of the u -subsystem as defined in (33). From Fig.(7), it can be seen that the identified CML model predicts very well even for the new data after $t = 30$. The model validation tests were applied to validate the final CML model of the Turing system by randomly selecting $N_v = 2500$ samples of data from the whole

time and space domain.. The validation test results are shown in the Fig.(9). It can be seen that both the correlation functions $\phi_{\beta\varepsilon^2}(\tau)$ and $\phi_{\beta u^2}(\tau)$ fall within the 95% confidence intervals. It can also be seen that the correlation tests of the final CML model that was identified with Bayesian regularisation are almost the same as the identified CML using the normal method in this example.

Table 2: Terms and parameters of the identified CML model for the u -subsystem of the Turing system in Example 2

Model terms	Estimated parameters without regularization	Estimated parameters with regularization	ERR
$u_{i,j}^*(t-1)$	0.1229	0.1206	0.9090
$v_{i,j}(t-1)$	-0.6401	-0.6412	0.0081
$u_{i,j}(t-1)$	0.1231	0.1350	0.0014
$v_{i,j}^3(t-1)$	1.8342	1.9426	0.0003
$u_{i,j}(t-1)v_{i,j}(t-1)$	-0.2958	1.0439	0.0001
$u_{i,j}^2(t-1)$	-0.0415	-0.1403	0.00002

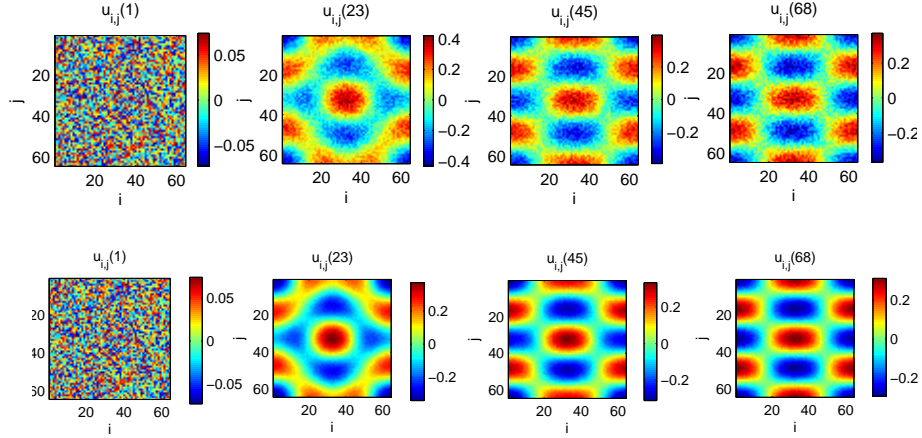


Figure 7: Snap shots of the Turing system noisy output (top) and model predicted output (bottom) at different times for the u -subsystem in Example 2

6 Conclusions

The identification and analysis of complex spatiotemporal patterns has been briefly reviewed and a new approach for identifying CML models of these pat-

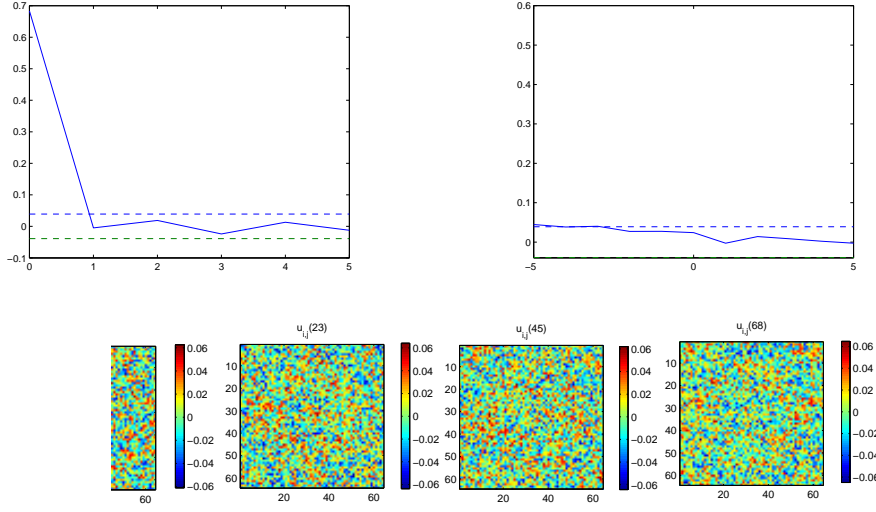


Figure 8: Correlation functions for the model validation using the final identified model for the u -subsystem of the Turing systems in Example 2, (top left) $\phi_{\beta\epsilon^2}(\tau)$, (top right) $\phi_{\beta u^2}(\tau)$ and (bottom) some snap shots of the OSA prediction error at different times.

terns has been proposed. Generally, the difficulties associated with the identification of complex spatiotemporal systems depend on the size of the system dimension and the complexities of the underlying dynamics. All of these effects can result in a ill-posed identification problem.

In this paper, two aspects of the identification problem have been considered to improve the performance of the final models. First, the importance of selecting the sampling interval of spatiotemporal data for identification was stressed and a novel method based on correlation analysis of the observed data has been introduced to choose the sampling intervals in the time and space domains. A combined OFR algorithm and a Bayesian learning method with Laplace priors has also been introduced to select the model terms and to obtain the estimates of the parameters. There are two main advantages of the Bayesian method compared with other traditional methods. One is that in the Bayesian method the noise model is included during the process of inferring the optimal values of the model parameters by setting the inverse of the variance of the residual sequence as a regulariser to avoid the final model fitting to the noise. The other advantage of the Bayesian method is an improvement in the smoothness of the final fitted model and the ability to suppress some irrelevant model terms by introducing the Laplace regularisers with respect to the model parameters.

The new algorithms have been applied to two complex patterns including spot self-replication patterns and Turing patterns. The simulation results demonstrate that the new identification procedure performs very well for these complex

spatiotemporal systems.

7 Appendix

7.1 The OFR Algorithm

For a given candidate set of regressors $G = \{\varphi_i\}_{i=1}^M$ where M is the number of the candidate regressors, the OFR algorithm can be briefly outlined as follows.
Step1: Select the first model term with highest ERR

$$I_1 = I_M = \{1, 2, \dots, M\} \quad (35)$$

$$w_i(t) = \varphi_i(t), \hat{b}_i = \frac{[w_i, Y]}{[w_i, w_i]} \quad (36)$$

$$l_1 = \arg \max_{i \in I_1} \hat{b}_i^2 \frac{[w_i, Y]}{[Y, Y]} = \arg \max_{i \in I_1} (ERR_i) \quad (37)$$

$$w_1^0 = w_{l_1}, c_1^0 = \frac{[w_1^0, Y]}{[w_1^0, w_1^0]}, a_{1,1} = 1 \quad (38)$$

Step $j, j = 2, 3, \dots$: Iteratively orthogonalise the remaining regressors one by one to select the next model term with highest ERR among the remaining candidate terms.

$$I_j = I_{j-1} \setminus l_j - 1 \quad (39)$$

$$w_i(t) = \varphi_i(t) - \sum_{k=1}^{j-1} \frac{[w_k^0, Y]}{[w_k^0, w_k^0]}, \hat{b}_i = \frac{[w_i, Y]}{[w_i, w_i]} \quad (40)$$

$$l_j = \arg \max_{i \in I_j} \hat{b}_i^2 \frac{[w_i, Y]}{[Y, Y]} = \arg \max_{i \in I_j} (ERR_i) \quad (41)$$

$$w_j^0 = w_{l_j}, c_j^0 = \frac{[w_j^0, Y]}{[w_j^0, w_j^0]}, a_{k,j} = \frac{[w_k^0, \varphi_{l_j}]}{[w_k^0, w_k^0]}, k = 1, 2, \dots, j-1 \quad (42)$$

This procedure is terminated at the M_s -th step when a required number of terms in the final model has been selected. The estimated coefficients $\Theta = \{\theta_k\}_{k=1}^{M_s}$ associated with the selected terms $\{\varphi_{l_k}\}_{k=1}^{M_s}$ are computed using

$$\Theta = \mathbf{A}^{-1}C, \quad (43)$$

where \mathbf{A} is upper-triangular matrix which is defined in (16) and $C = (c_1^0, c_2^0, \dots, c_{M_s}^0)$ is the coefficient vector associated with the orthogonalised terms $\{w_k^0\}_{k=1}^{M_s}$.

7.2 Maximizing the posterior probability with Laplace priors

Following [12] [11], when \mathbf{W}^0 is an orthogonal matrix and $(W^0)^T(W^0) = I$, the solution of the parameter g_i^0 maximizing the the posterior probability with Laplace priors should be

$$g_i^0 = \text{sgn}(w_i^{0T} \mathbf{Y}) \left(|w_i^{0T} \mathbf{Y}| - \frac{\lambda_i}{\epsilon} \right)_+ \quad (44)$$

Let $w_i = w_i^0 \|w_i\|$, so that $g_i^0 = g_i \|w_i\|$. Replace the w_i^0 and g_i^0 in (44), gives

$$g_i = \text{sgn}(w_i^T \mathbf{Y}) \left(|w_i^T \mathbf{Y}| - \frac{\lambda_i}{\epsilon \|w_i\|} \right)_+ \quad (45)$$

References

- [1] H. D. I. Abarbanel, R. Brown, and J. B. Kadtko, *Prediction in chaotic nonlinear systems: Methods for time series with broadband fourier spectra*, Phys. Rev. A **41** (1990), 1782–1807.
- [2] R. A. Barrio, C. Varea, J. L. Aragon, and P. K. Maini, *A two-dimensional numerical study of spatial pattern formation in interacting turing systems*, Bull. Math. Bio. **61** (1999), 483–505.
- [3] J. Bascompte and R. V. Sole, *Rethinking complexity: modelling spatiotemporal dynamics in ecology*, Trends Ecol. Evol. **10** (1995), 361–366.
- [4] S. A. Billings and L. A. Aguirre, *Effects of the sampling time of the dynamics and identification of nonlinear models*, Int. J. Bifurcation and Chaos **5** (1995), 1541–1556.
- [5] S. A. Billings, S. Chen, and M. J. Kronenberg, *Identification of mimo non-linear systems using a forward regression orthogonal estimator*, Int. J. Control **49** (1988), 2157–2189.
- [6] S. A. Billings and Q. H. Tao, *Model validity tests for non-linear signal processing applications*, Int. J. Control **54** (1991), 157–194.
- [7] V. Castet, E. Dulos, J. Boissonade, and P. DeKepper, *Experimental evidence of a sustained standing turing type nonequilibrium chemical pattern*, Phys. Rev. Lett. **64** (1990), 2953–2956.
- [8] D. Coca and S. A. Billings, *Identification of coupled map lattice models of complex spatiotemporal patterns*, Phys. Lett. A **287** (2001), 65–73.
- [9] ———, *Identification of finite dimensional models of infinite dimensional dynamical systems*, Automatica **38** (2003), 1851–1856.

- [10] P. DeKepper, V. Castets, E. Dulos, and J. Boissonade, *Turing type chemical patterns in the chlorite-iodide-malonic acid reaction*, Physica D. **49** (1991), 161–169.
- [11] D. Donoho and I. Johnstone, *Ideal spatial adaption via wavelet shrinkage*, Biometrika **81** (1994), 425–455.
- [12] M. A. T. Figueiredo, *Adaptive sparseness for supervised learning*, IEEE Trans. Pattern Analysis and Machine Interlligence **25** (2003), 1150–1159.
- [13] A. Frommer and P. Maass, *Fast cg-based methods for tikhonov-philips regularization*, SIAM J. Sci. Comput. **20** (1999), 1831–1850.
- [14] G. Golub, M. Heath, and G. Wahba, *Generalized cross-validation as a method for choosing a good ridge parameter*, Technometrics **21** (1979), 215–233.
- [15] J. Gradisek, S. Siegert, R. Friedrich, and I. Grabec, *Analysis of time series from stochastic processes*, Phys. Rev. E **62** (2000), 3146–3155.
- [16] P. Gray and S. K. Scott, *Autocatalytic reactions in the isothermal, continuous stirred tank reactor - isolas and other forms of multistability*, Chem. Eng. Sci **38** (1983), 29–43.
- [17] ———, *Autocatalytic reactions in the isothermal, continuous stirred tank reactor - oscillations and instabilities in the system $a+2b \rightarrow 3b$, $b \rightarrow c$* , Chem. Eng. Sci **39** (1984), 1087–1097.
- [18] ———, *Sustained oscillations and other exotic patterns of behavior in isothermal reactions*, J. Phys. Chem. **89** (1985), 22–32.
- [19] L. Z. Guo and S. A. Billings, *Identification of coupled map lattice models of stochastic spatio-temporal dynamics using wavelets*, Dynamical System **19** (2004), 265–278.
- [20] P. C. Hansen, *Analysis of discrete ill-posed problems by means of the l-curve*, SIAM Rev. **34** (1992), 561–580.
- [21] K. Kaneko, *Spatiotemporal intermittency in coupled map lattices*, Prog. Theor. Phys. **74** (1985), 1033–1044.
- [22] ———, *Turbulence in coupled map lattices*, Physica D **18** (1986), 475–476.
- [23] ———, *Pattern dynamics in spatiotemporal chaos pattern selection, diffusion of defect and pattern competition intermittency*, Physica D **34** (1989), 1–41.
- [24] R. Kapral, *Pattern formation in chemical systems*, Physica D **86** (1995), 149–157.

- [25] P. Kareiva and U. Wennergren, *Connecting landscape patterns to ecosystem and population processes*, Nature **373** (1995), 299–302.
- [26] M. E. Kilmer and D. P. O’Leary, *Choosing regularization parameters in iterative methods for ill-posed problems*, SIAM J. Matrix Analysis and Applications **22** (2001), 1204–1221.
- [27] K. J. Lee, W. D. McCormick, Q. Ouyang, and H. L. Swinney, *Pattern formation by interacting chemical fronts*, Science **16** (1993), 192–194.
- [28] K. J. Lee, W. D. McCormick, J. E. Pearson, and H. L. Swinney, *Experimental observation of self-replicating spots in a reaction-diffusion system*, Nature **369** (1994), 215–218.
- [29] I. Lengyel and I. R. Epstein, *Modelling of turing structures in the chlorite-iodide-malonic acid starch reaction system*, Science **251** (1991), 650–652.
- [30] F. Lesmes, D. Hochberg, F. Moran, and J. Perez-Mercader, *Noise-controlled self-replicating patterns*, Phys. Rev. Lett. **91** (2003), 238301–(1–4).
- [31] D. MacKay, *Bayesian interpolation*, Neural Computation **4** (1992), 417–447.
- [32] P. K. Maini, K. J. Painter, and H. N. P. Chau, *Spatial pattern formation in chemical and biological systems*, J. Chem. Soc., Faraday Trans. **93** (1997), 3601–3610.
- [33] S. Mandelj, I. Grabec, and E. Govekar, *Statistical approach to modelling of spatiotemporal dynamics*, Int. J. Bifurcation Chaos **17** (2001), 2731–2738.
- [34] P. Marcos-Nikolaus, J. M. Martin-Gonzalez, and R. V. Sole, *Spatial forecasting detecting determinism from single snapshots*, Int. J. Bifurcation Chaos **12** (2002), 369–376.
- [35] J. L. Maron and S. Harrison, *Spatial pattern formation in an insect host-parasitoid system*, Science **278** (1997), 1619–1621.
- [36] C. B. Muratov and V. V. Osipov, *Static spike autosolitons in the gray-scott model*, J. Phys. A: Math. Gen. **33** (2000), 8893–8916.
- [37] J. D. Murray, *Mathematical biology*, Springer-Verlag, 1989.
- [38] Y. Nishiura and D. Ueyama, *A skeleton structure of self-replicating dynamics*, Physica D **130** (1999), 73–104.
- [39] S. Ostavik and J. Stark, *Reconstruction and cross-prediction in coupled map lattices using spatio-temporal embedding techniques*, Phys. Lett. A **247** (1998), 145–160.
- [40] Q. Ouyang and H. L. Swinney, *Transition from a uniform state to hexagonal and striped turing patterns*, Nature **352** (1991), 610–612.

- [41] K. J. Painter, *Modelling of pigment patterns in fish*, In Mathematical Models for Biological Pattern Formation, IMA Volumes in Mathematics and its Applications **121** (2000), 59–82.
- [42] Y. Pan and S. A. Billings, *Model validation of spatiotemporal systems by using correlation function tests*, Submitted for publication.
- [43] U. Parlitz and C. Merkwirth, *Prediction of spatiotemporal time series based on reconstruction local states*, Phys. Rev. Lett. **84** (2000), 1890–1893.
- [44] J.E. Pearson, *Complex patterns in a simple system*, Science **16** (1993), 189–192.
- [45] M. T. Rosenstein, J. J. Collins, and C. J. De Luca, *Reconstruction expansion as a geometry-based framework for choosing proper delay times*, Physica D **73** (1994), 82–98.
- [46] A. Rovinsky and M. Menzinger, *Interaction of turing and hopf bifurcation in chemical systems*, Phys. Rev. A **46** (1992), 6315–6322.
- [47] R. Tibshirani, *Regression shrinkage and selection via the lasso*, J. Royal Statistical Soc. (B) **58** (1996), 267–288.
- [48] R. D. Vigil, Q. Ouyang, and H. L. Swinney, *Turing patterns in a simple gel reactor*, Physica A **188** (1992), 17–25.
- [49] A. N. Zaikin and A. M. Zhabotinsky, *Concentration wave propagation in 2-dimensional liquid-phase self-oscillating system*, Nature **255** (1970), 535–537.