Gradient-based estimation of local parameters for flow and transport in heterogeneous porous media

M. Stöhr[1] and K. Roth

Institute for Environmental Physics, Heidelberg, Germany

Received 27 October 2004; revised 3 May 2005; accepted 13 May 2005; published 2 August 2005.

[1] We present the application of the gradient-based total least squares (TLS) method to the local estimation of parameters for flow and transport in porous media. The concept is based on the evaluation of partial derivatives of spatially and temporally resolved data using TLS as a maximum likelihood estimator. While ordinary inverse modeling approaches are often complicated by the spatially varying properties of porous media, the present approach can directly localize the estimation to an arbitrary range in space and time. The estimation of the local parameters can be achieved without requiring any explicit solution of the respective transport equation. First the basic ideas and the formalism of TLS are introduced with a simple example of a straight line fit. Then the ideas of the gradient-based approach and its application to the parameter estimation for a large class of dynamic processes are presented. We further discuss relevant computational issues such as the calculation of the derivatives, choice of the local neighborhood and the determination of a measure of confidence. The performance of the method is then exemplified by the estimation of local velocities and dispersion coefficients from numerical solutions of the convection-dispersion equation.

Citation: Stöhr, M., and K. Roth (2005), Gradient-based estimation of local parameters for flow and transport in heterogeneous porous media, *Water Resour. Res.*, *41*, W08401, doi:10.1029/2004WR003768.

1. Introduction

[2] The estimation of model parameters like velocity or dispersion coefficient from experimental data is one of the crucial steps on the way toward a successful model of flow and transport in a porous medium. Field tracer experiments are performed in order to infer an adequate transport model for an aquifer, and to determine the respective model parameters (for a recent overview, see *Ptak et al.* [2004]). In addition, laboratory tracer experiments are employed to study small-scale processes in porous media and their relation to effective macroscopic parameters.

[3] The most commonly used transport model is the convection-dispersion equation (CDE)

$$\frac{\partial c}{\partial t} + \mathbf{v} \cdot \nabla c - \nabla(\mathsf{D}\nabla c) = 0, \tag{1}$$

where the velocity $\mathbf{v}$ and the dispersion tensor D are regarded as macroscopic, so-called effective parameters which therefore are constant in space and time. Since this model is based on the central limit theorem, the CDE (1) is only valid if the system is ergodic and has reached equilibrium. To fulfil the latter constraint the tracer must have covered a sufficiently long travel distance. Only then its dynamics can be described by the effective CDE (1). Numerous theoretical studies [e.g., *Dagan*, 1990; *Dentz et al.*, 2000a, 2000b] focus on the approach to the asymptotic state for various idealized media and initial conditions. Numerical simulations indicate that for the highly localized initial distributions that are preferred for tracer experiments, this distance is of the order of 10^2 correlation lengths of the underlying hydraulic structure.

[4] A major hurdle in applying such results for prediction but also for the analysis of experimental data is the fact that the presumed idealized media are hardly ever encountered in nature. In the majority of field experiments the asymptotic state is often not reached even at the end of the distance traveled due to the multiscale heterogeneity of soils and aquifers. This may be deduced from the observation that apparent dispersivities increase with scale [*Gelhar et al.*, 1992]. Consequently, the CDE (1) does not provide a valid description of the dynamics. Such a situation is illustrated by the 1D synthetic data set shown in Figure 1a, where the heterogeneity of v and D prevents the concentration distribution from retaining its initially gaussian shape. For stationary flow, a valid description of the temporal evolution of $c(\mathsf{x}, t)$ in such a situation is given by the CDE

$$\frac{\partial c}{\partial t} + \mathbf{v}(\mathbf{x}) \cdot \nabla c - \nabla(\mathsf{D}(\mathbf{x})\nabla c) = 0, \tag{2}$$

with the local velocity $\mathbf{v}(\mathbf{x})$ and dispersion tensor $\mathsf{D}(\mathbf{x})$. As a consequence, an adequate model for flow and transport in such a medium requires reliable information about the local values $\mathbf{v}(\mathbf{x})$ and $\mathsf{D}(\mathbf{x})$ instead of one single macroscopic value for $\mathbf{v}$ and D. This implies that the concentrations are measured with a sufficiently high spatial and temporal

[1]Now at Max Planck Institute for Marine Microbiology, Bremen, Germany.

Copyright 2005 by the American Geophysical Union.
0043-1397/05/2004WR003768

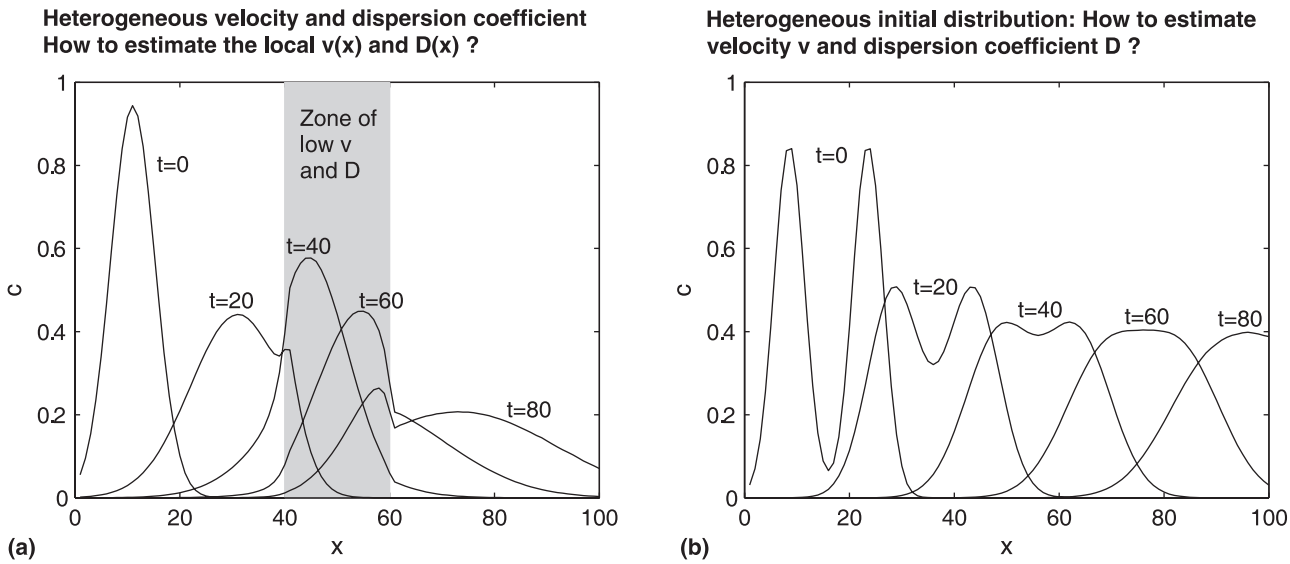

Figure 1. 1D particle tracking simulation of the CDE (2) (a) with a low-velocity/dispersion zone at $40 < x < 60$ (b) with constant v and D where the initial distribution $c(x, 0)$ is a superposition of two Gaussians at $x = 7.5$ and $x = 22.5$. The concentration distributions $c(x, t)$ are plotted for $t = 0$, 20, 40, 60 and 80. The TLS estimation of velocities and dispersion coefficients from this data is shown in Figures 6 and 7.

resolution. A growing number of techniques for the laboratory, like, e.g., light transmission with Hele-Shaw cells [*Berest et al.*, 1999; *Detwiler and Rajaram*, 2000], 3D planar laser-induced fluorescence [*Stöhr et al.*, 2003], positron emission projection imaging [*Loggia et al.*, 2004] and X-ray computed tomography [*Keller et al.*, 1999] are available to provide such data. Although the task is much more challenging on the field scale, high-resolution measurements of tracer migration have been obtained recently using 3D electrical resistivity tomography [*Kemna et al.*, 2002; *Singha et al.*, 2003].

[5] The commonly used approach for the estimation of the set of model parameters, denoted by $\mathbf{p}$, like, for example, $\mathbf{p} = (v, D)^T$, is the method known as inverse modeling (see, e.g., [*Abbaspour et al.*, 1997] and [*Zimmerman et al.*, 1998]). Here a numerical solution $c_{\mathrm{mod}}(\mathbf{x}, t; \mathbf{p})$ of the model, as, e.g., the CDE (1) or (2), is calculated for properly chosen initial and boundary conditions and an initial guess of the sought parameters $\mathbf{p}$. Then the value of $\mathbf{p}$ is estimated by an iterative minimization of the objective function

$$\chi^2 = \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \left(c(\mathbf{x}_i, t_i) - c_{\mathrm{mod}}(\mathbf{x}_i, t_i; \mathbf{p})\right)^2 \tag{3}$$

with respect to $\mathbf{p}$. This so-called least squares approach assumes that the differences between the data $c(\mathbf{x}_i, t_i)$ and the model $c_{\mathrm{mod}}(\mathbf{x}_i, t_i; \mathbf{p})$ are independent and normally distributed with variance σ_i^2. The so estimated $\mathbf{p}$ is then the maximum likelihood estimate.

[6] However, this prevalent approach has a number of limitations and shortcomings.

[7] 1. Estimation of local parameters with the inverse modeling approach, like $\mathbf{v}(\mathbf{x})$ and $D(\mathbf{x})$ in the CDE (2) as illustrated in Figure 1a, is virtually impossible for all but the most simple heterogeneities. The reason is that the number of estimated parameters will strongly increase and the objective function (3) will possess local minima.

[8] 2. The accurate knowledge of initial and boundary conditions is mandatory for a reliable estimation. This situation is illustrated in the 1D synthetic data shown in Figure 1b. Even though the system is in equilibrium, the tracer distribution has not yet reached a gaussian shape due to the heterogeneity of the initial distribution $c(\mathbf{x}, 0)$. If the actual initial distribution is not correctly taken into account for the numerical solution $c_{\mathrm{mod}}(\mathbf{x}, t; \mathbf{p})$, and e.g. the solution for a pulse-like injection (i.e. the evolution of a single gaussian) is used instead, the estimated parameters will be biased.

[9] 3. An informed guess about the initial values of $\mathbf{p}$ is required for the iterative minimization of (3) in order to avoid running into local minima.

[10] 4. The iterative minimization of (3) is computationally expensive due to the repeated computation of $c(\mathbf{x}, t; \mathbf{p})$.

[11] 5. If the problem is underdetermined or ill-posed, e.g. due to dependencies between parameters, its solution will diverge. The identification and treatment of such a problem requires an additional analysis and, if necessary, a repetition of the estimation.

[12] In the present paper, a method will be presented which allows to overcome the above mentioned limitations for any dynamic process, provided that its differential equation is linear in the estimated parameters. For the 1D examples shown in Figure 1, it allows e.g. the estimation of the local velocity $\mathbf{v}(\mathbf{x})$ and dispersion tensor $\mathsf{D}(\mathbf{x})$ for arbitrarily shaped initial concentration distributions.

[13] The basis of the method is the so-called total least squares (TLS) technique, which is introduced and opposed to the commonly used ordinary least squares (OLS) approach in section 2. Section 3 will then describe the application of the TLS technique to the gradient-based, local parameter estimation for linear dynamic processes, like e.g.

a convection-dispersion. Several computational issues, which are essential for the practical application of the method, are addressed in section 4. In section 5 the performance of the method will be demonstrated and discussed with the analysis of the synthetic 1D data sets shown in Figure 1, and a numerical solution of a 2D convection-dispersion in a heterogeneous medium. Finally, section 6 will give a summary and conclusions.

2. Total Least Squares (TLS) Estimator

[14] Generally, a major challenge for the estimation of model parameters from a set of measured data is the separation of the relevant information from superimposed noise. A widely-used approach is the framework of least squares methods. The parameter estimation technique presented in this paper relies on the total least squares method, which itself belongs to the class of subspace methods. These methods, which have found several applications in digital signal processing and computer vision over the last decade [see, e.g., *Van Huffel and Lemmerling*, 2002; *Mühlich and Mester*, 1999; *Haußecker et al.*, 1999], achieve this separation through dividing the vector space of an overdetermined set of equations into a data space and an error space. In the following the ideas and the formalism of the TLS method will be introduced and opposed to the commonly used ordinary least squares (OLS) technique by the example of a straight line fit. For a more comprehensive treatment, see *Björck* [1996] and *Van Huffel and Vandewalle* [1991].

2.1. Ordinary Least Squares (OLS)

[15] One of the simplest parameter estimation problems is the fitting of a straight line

$$y = ax + b \tag{4}$$

to a set of noisy measurements

$$y_i = y_{i0} + \varepsilon_i, \langle \varepsilon_i \rangle = 0 \qquad \langle \varepsilon_i^2 \rangle = \sigma_i^2 \qquad i = 1..n \tag{5}$$

at the positions x_i, where the y_i are assumed to be a superposition of undisturbed values y_{i0} and independent and normally distributed noise ε_i with zero mean and variance σ_i^2. For $n > 2$ this results in the overdetermined set of equations

$$y_i = ax_i + b \qquad i = 1..n. \tag{6}$$

If the σ_i are equal, the maximum likelihood solution for the sought parameters a and b is given by those values that yield the minimum sum of squared differences

$$\sum_{i=1}^{n} (ax_i + b - y_i)^2 \stackrel{!}{=} \min. \tag{7}$$

With

$$\mathsf{A} = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix} \quad \mathbf{p} = \begin{pmatrix} a \\ b \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \tag{8}$$

the minimization constraint (7) can be written as

$$\| \mathsf{A}\mathbf{p} - \mathbf{b} \|^2 \stackrel{!}{=} \min \tag{9}$$

and the minimum can be found by setting the partial derivatives to zero:

$$\frac{\partial}{\partial a} \| \mathsf{A}\mathbf{p} - \mathbf{b} \|^2 = 0 \qquad \frac{\partial}{\partial b} \| \mathsf{A}\mathbf{p} - \mathbf{b} \|^2 = 0 \tag{10}$$

$$\mathsf{A}^T(\mathsf{A}\mathbf{p} - \mathbf{b}) = 0 \tag{11}$$

$$\mathbf{p} = \left(\mathsf{A}^T\mathsf{A}\right)^{-1}\mathsf{A}^T\mathbf{b} \tag{12}$$

$(\mathsf{A}^T\,\mathsf{A})^{-1}\,\mathsf{A}^T$ is the so-called pseudoinverse. If the σ_i are not equal, the solution is

$$\mathbf{p} = \left(\mathsf{A}^T\mathsf{W}^T\mathsf{W}\mathsf{A}\right)^{-1}\mathsf{A}^T\mathsf{W}^T\mathsf{W}\mathbf{b} \tag{13}$$

$$\text{with} \qquad \mathsf{W} = \begin{pmatrix} 1/\sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1/\sigma_n \end{pmatrix}. \tag{14}$$

This solution for $\mathbf{p}$ is optimal in the sense of minimal quadratic error and maximum likelihood.

2.2. Total Least Squares (TLS)

[16] The OLS estimator is based on the assumption that the x_i are error-free and only the y_i contain noise. In contrast, the TLS estimator accepts both, x_i and y_i, to be noisy:

$$x_i = x_{i0} + \eta_i \qquad \langle \eta_i \rangle = 0 \qquad \langle \eta_i^2 \rangle = \sigma_x^2 \tag{15}$$

$$y_i = y_{i0} + \varepsilon_i \qquad \langle \varepsilon_i \rangle = 0 \qquad \langle \varepsilon_i^2 \rangle = \sigma_y^2 \tag{16}$$

Again, the approach to the maximum likelihood estimate is the minimization of the sum of squared errors ([*Van Huffel and Vandewalle*, 1991]):

$$\sum_{i=1}^{n} (ax_i + b - y_i)^2 \stackrel{!}{=} \min. \tag{17}$$

With

$$\mathsf{A} = \begin{pmatrix} -y_1 & x_1 & 1 \\ -y_2 & x_2 & 1 \\ \vdots & \vdots & \vdots \\ -y_n & x_n & 1 \end{pmatrix} \quad \mathbf{p} = \begin{pmatrix} 1 \\ a \\ b \end{pmatrix} \tag{18}$$

the set of equations (6) can be written as

$$\mathsf{A}\mathbf{p} = 0 \tag{19}$$

and the least sum of squared differences (17) is defined as

$$\| \mathsf{A}\mathbf{p} \|^2 \stackrel{!}{=} \min. \tag{20}$$

In order to avoid the trivial solution $\mathbf{p} = 0$, the norm of $\mathbf{p}$ is required to be an arbitrary constant c, say $c = 1$, by the additional constraint

$$\| \mathbf{p} \|^2 = 1. \tag{21}$$

Here, the difference between this so-called total least squares (TLS) approach and the ordinary least squares (OLS) estimation described above becomes apparent: In OLS the minimum is found in the 2D vector space of the 2D vector $\mathbf{p}$, while in TLS $\mathbf{p}$ is a 3D vector and the minimum is lying on the unit sphere of the according 3D vector space. The TLS minimization (20) with the constraint (21) is then carried out with the method of Lagrange multipliers:

$$\| \mathsf{A}\mathbf{p} \|^2 + \lambda\left(1 - \| \mathbf{p} \|^2\right) \stackrel{!}{=} \min \tag{22}$$

or with $\mathsf{J} = \mathsf{A}^T \mathsf{A}$

$$\mathbf{p}^T \mathsf{J}\mathbf{p} + \lambda\left(1 - \mathbf{p}^T\mathbf{p}\right) \stackrel{!}{=} \min. \tag{23}$$

The minimum is found by setting the partial derivatives with respect to the elements of $\mathbf{p}$ to zero, which leads to the eigenvalue equation

$$\mathsf{J}\mathbf{p} = \lambda\mathbf{p}. \tag{24}$$

From the properties of J, (1) symmetric $\rightarrow$ real eigenvalues; (2) non-negative definite $\rightarrow$ non-negative eigenvalues, it follows that the minimum is given by the eigenvector to the smallest eigenvalue $\lambda_{\min}$ of J:

$$\mathsf{J}\mathbf{p}_{\min} = \lambda_{\min}\mathbf{p}_{\min}. \tag{25}$$

The values for a and b are then obtained by scaling $\mathbf{p}_{\min}$ to satisfy $p_1 = 1$ in (18):

$$a = \frac{p_{\min 2}}{p_{\min 1}} \qquad b = \frac{p_{\min 3}}{p_{\min 1}}. \tag{26}$$

2.3. Equilibration

[17] It can be shown [*Van Huffel and Vandewalle*, 1991] that $\mathbf{p}_{\min}$ given by (25) is the maximum likelihood estimator if the errors of the elements of A are independent random variables with zero mean and equal standard deviation σ_{A}:

$$\mathsf{A} = \mathsf{A}_0 + \Delta\mathsf{A} \quad \langle \Delta\mathsf{A}_{ij} \rangle = 0 \quad \langle \Delta\mathsf{A}_{ij}^2 \rangle = \sigma_{\mathsf{A}}^2 \tag{27}$$

A typical situation is that each column of A has a different error variance:

$$\mathsf{A} = \begin{pmatrix} -y_1 & x_1 & 1 \\ -y_2 & x_2 & 1 \\ \vdots & \vdots & \vdots \\ -y_n & x_n & 1 \end{pmatrix} \tag{28}$$
$$\begin{array}{ccc} \uparrow & \uparrow & \uparrow \\ \sigma_y^2 & \sigma_x^2 & \sigma_1^2 \end{array}$$

[18] In the example shown in Figure 2 for instance these values are $\sigma_x^2 = 1$, $\sigma_y^2 = 2$ and $\sigma_1^2 = 0$. Therefore the requirement (27) is violated, which explains the erroneous result of the TLS estimation in Figure 2b.

[19] To fulfil the requirement (27) the columns of A have to be scaled with a weight matrix W:

$$\mathsf{A}\mathsf{W}\mathsf{W}^{-1}\mathbf{p} = 0 \tag{29}$$

$$\mathsf{W} = \begin{pmatrix} 1/\sigma_y & 0 & 0 \\ 0 & 1/\sigma_x & 0 \\ 0 & 0 & 1/\sigma_1 \end{pmatrix} \tag{30}$$

$$\mathsf{A}_{\mathrm{eq}}\mathbf{p}_{\mathrm{eq}} = 0 \quad \mathsf{A}_{\mathrm{eq}} = \mathsf{A}\mathsf{W} \quad \mathbf{p}_{\mathrm{eq}} = \mathsf{W}^{-1}\mathbf{p} \tag{31}$$

with A_{eq} fulfilling the condition (27). This transformation toward a matrix A_{eq} with equal standard deviations of the elements is called *equilibration*.

[20] For the final optimal solution $\mathbf{p}_{\min}$ the eigenvector $\mathbf{p}_{\mathrm{eqmin}}$ to the smallest eigenvalue λ_{eqmin} of $\mathsf{J}_{\mathrm{eq}} = \mathsf{A}_{\mathrm{eq}}^T\mathsf{A}_{\mathrm{eq}}$ has then to be rescaled:

$$\mathbf{p}_{\min} = \mathsf{W}\mathbf{p}_{\mathrm{eqmin}} \qquad a = \frac{p_{\min 2}}{p_{\min 1}} \qquad b = \frac{p_{\min 3}}{p_{\min 1}}. \tag{32}$$

As expected the result for the equilibrated TLS shown in Figure 2d proves to be better than OLS and non-equilibrated TLS.

[21] If one column of A is assumed to be error-free, like in the present example $\sigma_1 = 0$, the corresponding element in W becomes infinite and therefore A_{eq} is undefined. This problem can be circumvented by approximating σ_1 by the smallest possible value with respect to the machine precision or by using the mixed OLS-TLS method given by *Van Huffel and Vandewalle* [1991].

2.4. Concluding Remarks

[22] In the previous paragraphs it has been shown that TLS is a generalization of OLS in the sense that it allows errors in all variables. This feature, which qualifies it for the local parameter estimation of dynamic processes described in the next section, involves some essential differences in the estimation procedure. Compared to OLS, the TLS parameter vector $\mathbf{p}$ has one additional element and is subject to the auxiliary constraint (21). Instead of the pseudoinverse (12) for OLS, the maximum likelihood estimator for TLS is given by the solution of the eigenvalue problem (25). The eigenvectors corresponding to the set of eigenvalues $\lambda_1 > \lambda_2 > \lambda_3$ represent the separation of the data matrix A into the data space spanned by $\mathbf{e}_1$ and $\mathbf{e}_2$ and the error space spanned by $\mathbf{e}_3$. An intuitive geometric interpretation of the difference between OLS and TLS for the straight line fit is shown in Figure 2: Whereas OLS minimizes the vertical distances, TLS minimizes the euclidean distances to the straight line.

3. Estimation of Parameters for Linear Dynamic Processes

[23] We consider a solute distribution as an object whose movement and change of shape shall be estimated. In a first step, we assume that the shape and brightness (i.e. concentration) of the object are invariant and its velocity is

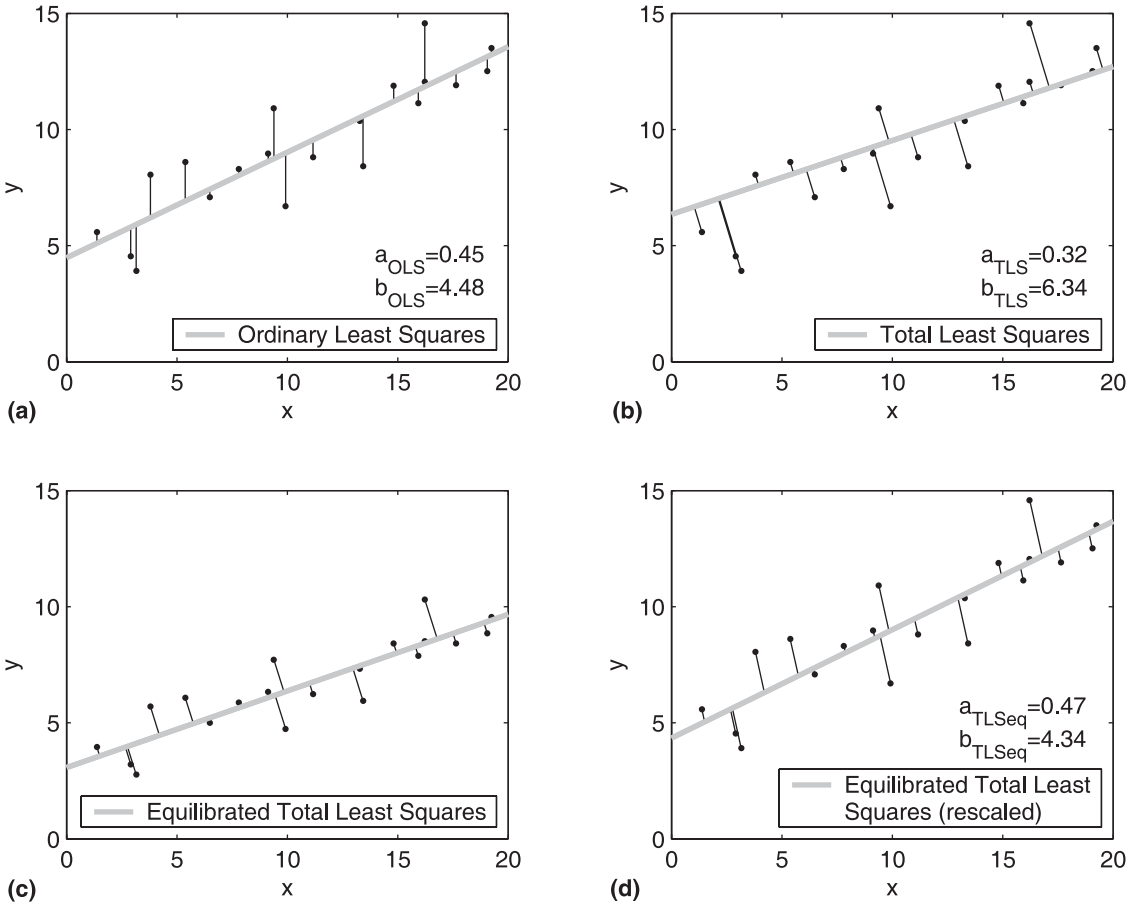

Figure 2. Geometric interpretation of different least squares estimators used for a straight line fit $y = ax + b$ to data with $a = 0.5$, $b = 4$ and errors in x and y ($\sigma_x^2 = 1$ and $\sigma_y^2 = 2$): (a) OLS minimizes the deviations in y-direction to the estimated line, while (b) TLS minimizes the Euclidean distances. This is only correct for $\sigma_x = \sigma_y$, which is achieved by an equilibration, i.e. scaling of y with a factor $1/\sqrt{2}$ (c). The so-obtained solution has then to be rescaled for the correct result (d). As indicated by the estimated values for a and b, equilibrated TLS is the maximum likelihood estimator in this situation and therefore provides better results than OLS. The results of TLS without equilibration are, however, typically worse than OLS.

constant. We then extend the description to the realistic case of an object that changes its shape and brightness, as e.g. according to the dispersion equation.

3.1. Gradient-Based Motion Estimation

[24] Suppose you want to estimate the velocity of an object from a time series of 2D images, represented by an array of intensity values $g(\mathbf{x}, t)$ at discrete positions $(\mathbf{x}, t)$, as illustrated in Figure 3a. If the two assumptions

[25] 1. the object is moving with constant velocity

[26] 2. the brightness of the object is constant are fulfilled, the temporal evolution of $g(\mathbf{x}, t)$ can be described by the brightness constancy constraint equation (BCCE)

$$\frac{\mathrm{d}g}{\mathrm{d}t} = \frac{\partial g}{\partial t} + \frac{\partial g}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial g}{\partial y}\frac{\partial y}{\partial t} = \frac{\partial g}{\partial t} + \mathbf{u}\nabla_{\mathbf{x}} g = 0 \tag{33}$$

$$\text{with} \qquad \mathbf{u} = \begin{pmatrix} u_x \\ u_y \end{pmatrix} = \begin{pmatrix} \frac{\partial x}{\partial t} \\ \frac{\partial y}{\partial t} \end{pmatrix}, \nabla_{\mathbf{x}} g = \begin{pmatrix} \frac{\partial g}{\partial x} \\ \frac{\partial g}{\partial y} \end{pmatrix}. \tag{34}$$

Since the BCCE gives only one constraint for the two unknown velocities u_x and u_y for every $(\mathbf{x}, t)$-position, additional constraint are required. A commonly used approach is to extend the estimation to a local neighborhood, assuming that $\mathbf{u}$ is constant in this neighborhood. Since $g(\mathbf{x}, t)$ virtually always contains noise, the BCCE (33) will not exactly equal zero, and $\mathbf{u}$ may be found by the minimization of the sum of squared residuals

$$E = \sum_{\mathbf{x}',t'} w(\mathbf{x} - \mathbf{x}', t - t') \cdot \left(\left.\frac{\partial g}{\partial t}\right|_{(\mathbf{x}',t')} + \mathbf{u}\nabla_{\mathbf{x}} g|_{(\mathbf{x}',t')} \right)^2 \stackrel{!}{=} \min. \tag{35}$$

The window function w, e.g. a box function or a Gaussian, defines the shape and extent of the local neighborhood at each point $(\mathbf{x}, t)$, for which the velocities are estimated. The possibility to specify an arbitrary range in space and time for each estimate is an important feature of gradient-based techniques. For a given noise level, a larger neighborhood will lead to a more accurate estimation of u_x and u_y. However it should not be chosen too large in order to satisfy

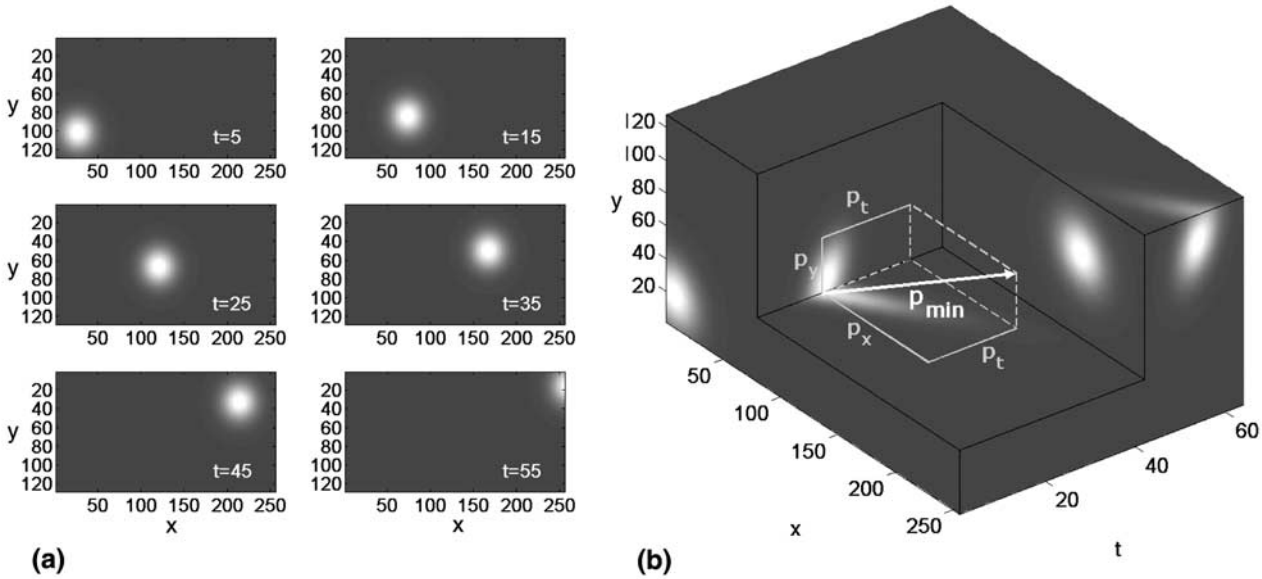

Figure 3. Time series $g(x, y, t)$ (t = 1..64) of 2D intensity images (x = 1..256, y = 1..128) with a Gaussian distribution moving with $\mathbf{u} = (4.7, -1.7)^T$ (a) shown for six different times and (b) represented by a 3D $\mathbf{x} - t$ visualization. From the 3D spatiotemporal representation the velocities u_x and u_y can be calculated from the direction $\mathbf{p}_{\min}$ of minimal intensity change according to (40).

the requirement of constant $\mathbf{u}$ throughout the neighborhood (see section 4.4).

[27] In principle, the parameters u_x and u_y could be found by setting the partial derivatives of E with respect to u_x and u_y to zero. This approach would be analogous to the OLS straight line fit given by (10)–(12) and leads to a corresponding pseudoinverse solution. However, this OLS estimation is not optimal since it is based on the assumption that only $\frac{\partial g}{\partial t}$ is noisy. Because in (35) all data, namely $\frac{\partial g}{\partial t}$, $\frac{\partial g}{\partial x}$ and $\frac{\partial g}{\partial y}$ are contaminated with noise, the optimal solution in the sense of maximum likelihood for u_x and u_y is instead given by the TLS minimization of

$$E = \sum_{\mathbf{x}',t'} w(\mathbf{x} - \mathbf{x}', t - t') \left(\nabla g|^T_{(\mathbf{x}',t')} \mathbf{p} \right)^2 \stackrel{!}{=} \min \tag{36}$$

$$\text{with} \qquad \mathbf{p} = \begin{pmatrix} p_x \\ p_y \\ p_t \end{pmatrix} = \begin{pmatrix} u_x \\ u_y \\ 1 \end{pmatrix} \qquad \nabla g = \begin{pmatrix} \frac{\partial g}{\partial x} \\ \frac{\partial g}{\partial y} \\ \frac{\partial g}{\partial t} \end{pmatrix} \tag{37}$$

and the additional constraint $\mathbf{p}^T \mathbf{p} = 1$. With

$$\mathsf{J} = \sum_{\mathbf{x}',t'} w(\mathbf{x} - \mathbf{x}', t - t') \left(\nabla g \nabla g^T |_{(\mathbf{x}',t')} \right) \tag{38}$$

$$= \begin{pmatrix} \overline{\frac{\partial g}{\partial x}\frac{\partial g}{\partial x}} & \overline{\frac{\partial g}{\partial x}\frac{\partial g}{\partial y}} & \overline{\frac{\partial g}{\partial x}\frac{\partial g}{\partial t}} \\ \overline{\frac{\partial g}{\partial y}\frac{\partial g}{\partial x}} & \overline{\frac{\partial g}{\partial y}\frac{\partial g}{\partial y}} & \overline{\frac{\partial g}{\partial y}\frac{\partial g}{\partial t}} \\ \overline{\frac{\partial g}{\partial t}\frac{\partial g}{\partial x}} & \overline{\frac{\partial g}{\partial t}\frac{\partial g}{\partial y}} & \overline{\frac{\partial g}{\partial t}\frac{\partial g}{\partial t}} \end{pmatrix} \tag{39}$$

the minimum of (36) is given according to (23)–(25) by the eigenvector $\mathbf{p}_{\min}$ to the smallest eigenvalue of J. The maximum likelihood solution for $\mathbf{u}$ is then given by

$$\mathbf{u} = \begin{pmatrix} u_x \\ u_y \end{pmatrix} = \begin{pmatrix} \frac{p_{\min x}}{p_{\min t}} \\ \frac{p_{\min y}}{p_{\min t}} \end{pmatrix}. \tag{40}$$

[28] The 3D $\mathbf{x} - t$ representation in Figure 3b gives an intuitive geometric interpretation of the above described motion estimation: The motion of the Gaussian at a certain ($\mathbf{x}$, t)-position is characterized by the direction of minimal intensity change in the local neighborhood. Since at one point ($\mathbf{x}$, t) this direction is perpendicular to the vector of maximal intensity change $\nabla g|_{\mathbf{x},t}$ and therefore $\nabla g|^T_{\mathbf{x},t}\, \mathbf{p} = 0$, the integration over the local neighborhood directly leads to (36). Thus it appears that the TLS estimate $\mathbf{p}_{\min}$ is equivalent to the direction of minimal intensity change in the 3D local $\mathbf{x} - t$ neighborhood defined by the window function w.

3.2. Extended Models

[29] Until now it was assumed that the temporal change of intensity $\frac{\partial g}{\partial t}$ was caused only by motion. This led to the BCCE (33) and to the further conclusion that the velocity corresponds to the direction of minimal intensity changes defined by (36). However this assumption is often violated by additional dynamic processes causing intensity changes, like e.g. diffusion, exponential decay or changes of illumination. Taking into account such additional processes leads to an extension of the BCCE (33) and consequently to a larger number of parameters for estimation. In the following this extension will be described for the example of 2D motion with isotropic diffusion shown in Figure 4.

[30] The temporal change of intensity due to 2D isotropic diffusion is described by the differential equation

$$\frac{\partial g}{\partial t} = D \left(\frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2} \right) \tag{41}$$

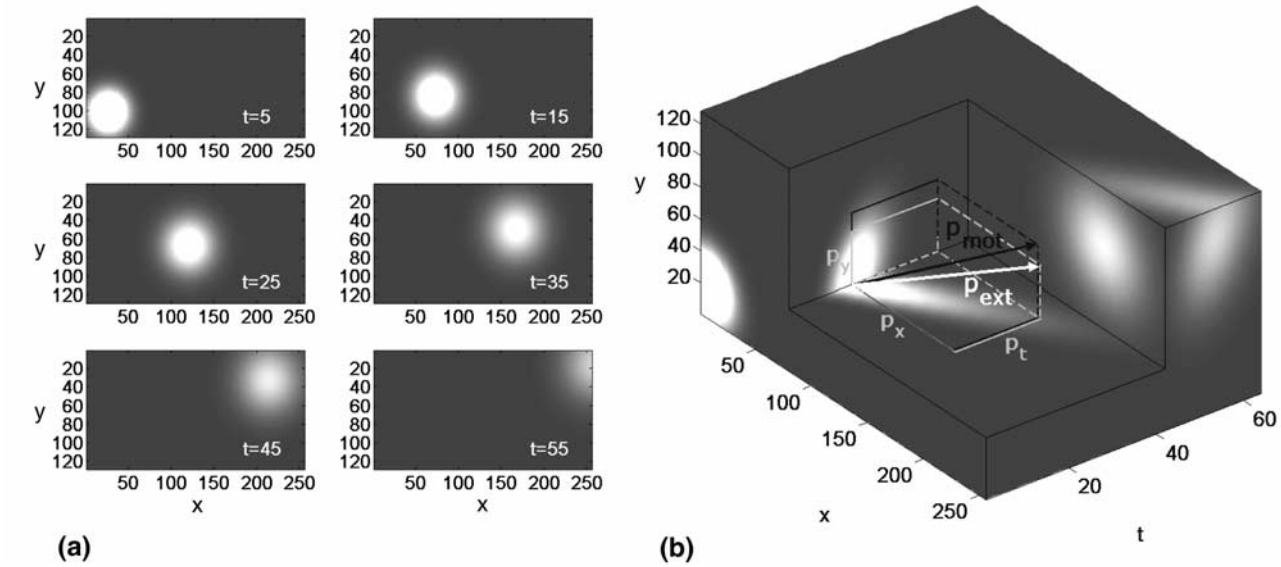

Figure 4. Time series $g(x, y, t)$ ($t = 1..64$) of 2D intensity images ($x = 1..256$, $y = 1..128$) with a gaussian distribution moving with $\mathbf{u} = (4.7, -1.7)^T$ and additional isotropic diffusion (a) shown for six different times and (b) represented by a 3D $\mathbf{x} - t$ visualization. Now the parameters u_x, u_y and D are not obtained from the direction of minimal intensity change $\mathbf{p}_{\mathrm{mot}}$, but from the eigenvector $\mathbf{p}_{\mathrm{ext}}$ of the extended tensor (45) according to (46).

with D denoting the diffusion coefficient. Together with a constant movement with velocity $\mathbf{u}$ the extended version of the BCCE (33) is given by

$$\frac{\mathrm{d}g}{\mathrm{d}t} = \frac{\partial g}{\partial t} + \frac{\partial g}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial g}{\partial y}\frac{\partial y}{\partial t} - D\left(\frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2}\right) = 0 \tag{42}$$

$$= \frac{\partial g}{\partial t} + \mathbf{u}\nabla_{\mathbf{x}} g - D\Delta g = 0. \tag{43}$$

In order to estimate the extended set of parameters u_x, u_y and D, the minimization constraint (36) is reformulated in the generalized form

$$E = \sum_{\mathbf{x}',t'} w(\mathbf{x} - \mathbf{x}', t - t')\left(\mathbf{d}^T\mathbf{p}\right)^2 \overset{!}{=} \min. \tag{44}$$

For the present example of 2D motion with isotropic diffusion the data vector $\mathbf{d}$ is given by $\mathbf{d} = (\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y}, \frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2}, \frac{\partial g}{\partial t})^T$ and the parameter vector $\mathbf{p}$ is $\mathbf{p} = (u_x, u_y, -D, 1)^T$. The further procedure of parameter estimation is analogous to the estimation of velocities in (38)–(40): With the extended tensor J defined as

$$\mathsf{J} = \sum_{\mathbf{x}',t'} w(\mathbf{x} - \mathbf{x}', t - t')\left(\mathbf{d}\mathbf{d}^T\big|_{(\mathbf{x}',t')}\right) \tag{45}$$

the maximum likelihood estimate $\mathbf{p}_{\min}$ is given by the eigenvector to the smallest eigenvalue of J, and the resulting parameters u_x, u_y and D are

$$u_x = \frac{p_{\min 1}}{p_{\min 4}} \quad u_y = \frac{p_{\min 2}}{p_{\min 4}} \quad D = \frac{-p_{\min 3}}{p_{\min 4}}. \tag{46}$$

This general framework, consisting of the minimization constraint (44) and the eigen analysis of the tensor J, provides a methodology for the estimation of local parameters of a large class of dynamic processes. This way of parameter estimation does not require a solution of the respective differential equation. The implementation of a certain model is simply accomplished by an adequate selection of a data vector $\mathbf{d}$ for the computation of J and a parameter vector $\mathbf{p}$ which will result from the eigen analysis of J. For that purpose the differential equation of the model to be implemented must be a sum of products between data and parameters. Each of these products becomes one element of $\mathbf{d}$ and $\mathbf{p}$. Table 1 shows a compilation of several common models and their differential equations together with the respective data vector $\mathbf{d}$ and parameter vector $\mathbf{p}$.

3.3. Equilibration

[31] The computation of the elements of $\mathbf{d}$, i.e. the derivatives $\frac{\partial g}{\partial t}$, $\frac{\partial g}{\partial x}$, $\frac{\partial g}{\partial y}$, $\frac{\partial^2 g}{\partial x^2}$ etc., from an image sequence like the example shown in Figure 4 is typically accomplished through a convolution of $g(\mathbf{x}, t)$ with an appropriate filter mask as discussed in section 4.2. The employment of different filter masks gives rise to different error variances for each element of $\mathbf{d}$, which can be calculated by error propagation from the respective filter mask and the noise spectrum of g (see section 4.3). These different variances of each element of $\mathbf{d}$ must then be equilibrated as described in section 2.3,

$$\mathbf{d}_{\mathrm{eq}} = \mathsf{W}\mathbf{d} \quad \mathsf{J}_{\mathrm{eq}} = \mathsf{W}\mathsf{J}\mathsf{W} \quad \mathsf{W} = \begin{pmatrix} 1/\sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1/\sigma_n \end{pmatrix} \tag{47}$$

Table 1. Examples of Models for Which the TLS Parameter Estimation Technique Can Be Employed[a]

Model	Differential Equation	Parameters	Data Vector **d**	Parameter Vector **p**
1D motion	$\frac{\partial g}{\partial t} + u\frac{\partial g}{\partial x} = 0$	Velocity u	$\left(\frac{\partial g}{\partial x}, \frac{\partial g}{\partial t}\right)^T$	$(u_x, 1)^T$
1D motion + diffusion	$\frac{\partial g}{\partial t} + u\frac{\partial g}{\partial x} - D\frac{\partial^2 g}{\partial x^2} = 0$	Velocity u Diffusion Coefficient D	$\left(\frac{\partial g}{\partial x}, \frac{\partial^2 g}{\partial x^2}, \frac{\partial g}{\partial t}\right)^T$	$(u, -D, 1)^T$
2D motion	$\frac{\partial g}{\partial t} + \mathbf{u}\nabla g = \frac{\partial g}{\partial t} + u_x\frac{\partial g}{\partial x} + u_y\frac{\partial g}{\partial y} = 0$	Velocity $\mathbf{u} = \begin{pmatrix} u_x \\ u_y \end{pmatrix}$	$\left(\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y}, \frac{\partial g}{\partial t}\right)^T$	$(u_x, u_y, 1)^T$
2D motion + isotropic diffusion	$\frac{\partial g}{\partial t} + \mathbf{u}\nabla g - D\Delta g = \frac{\partial g}{\partial t} + u_x\frac{\partial g}{\partial x} + u_y\frac{\partial g}{\partial y} - D\left(\frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2}\right) = 0$	Velocity $\mathbf{u} = \begin{pmatrix} u_x \\ u_y \end{pmatrix}$ Diffusion Coefficient D	$\left(\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y}, \frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2}, \frac{\partial g}{\partial t}\right)^T$	$(u_x, u_y, -D, 1)^T$
2D motion + anisotropic diffusion	$\frac{\partial g}{\partial t} + \mathbf{u}\nabla g - \nabla(D\nabla g) = \frac{\partial g}{\partial t} + u_x\frac{\partial g}{\partial x} + u_y\frac{\partial g}{\partial y} - D_{xx}\frac{\partial^2 g}{\partial x^2} - 2D_{xy}\frac{\partial^2 g}{\partial x \partial y} - D_{yy}\frac{\partial^2 g}{\partial y^2} = 0$	Velocity $\mathbf{u} = \begin{pmatrix} u_x \\ u_y \end{pmatrix}$ Diffusion Tensor $D = \begin{pmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{pmatrix}$	$\left(\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y}, \frac{\partial^2 g}{\partial x^2}, \frac{\partial^2 g}{\partial x \partial y}, \frac{\partial^2 g}{\partial y^2}, \frac{\partial g}{\partial t}\right)^T$	$(u_x, u_y, -D_{xx}, -2D_{xy}, -D_{yy}, 1)^T$
2D motion + exponential decay	$\frac{\partial g}{\partial t} + \mathbf{u}\nabla g + \kappa g = \frac{\partial g}{\partial t} + u_x\frac{\partial g}{\partial x} + u_y\frac{\partial g}{\partial y} + \kappa g = 0$	Velocity $\mathbf{u} = \begin{pmatrix} u_x \\ u_y \end{pmatrix}$ Decay Constant κ	$\left(\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y}, g, \frac{\partial g}{\partial t}\right)^T$	$(u_x, u_y, \kappa, 1)^T$
2D motion + linear change	$\frac{\partial g}{\partial t} + \mathbf{u}\nabla g - q = \frac{\partial g}{\partial t} + u_x\frac{\partial g}{\partial x} + u_y\frac{\partial g}{\partial y} - q = 0$	Velocity $\mathbf{u} = \begin{pmatrix} u_x \\ u_y \end{pmatrix}$ Source strength q	$\left(\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y}, 1, \frac{\partial g}{\partial t}\right)^T$	$(u_x, u_y, -q, 1)^T$
3D motion + anisotropic diffusion	$\frac{\partial g}{\partial t} + \mathbf{u}\nabla g - \nabla(D\nabla g) = \frac{\partial g}{\partial t} + u_x\frac{\partial g}{\partial x} + u_y\frac{\partial g}{\partial y} + u_z\frac{\partial g}{\partial z} - D_{xx}\frac{\partial^2 g}{\partial x^2} - D_{yy}\frac{\partial^2 g}{\partial y^2} - D_{zz}\frac{\partial^2 g}{\partial z^2} - 2D_{xy}\frac{\partial^2 g}{\partial x \partial y} - 2D_{xz}\frac{\partial^2 g}{\partial x \partial z} - 2D_{yz}\frac{\partial^2 g}{\partial y \partial z} = 0$	Velocity $\mathbf{u} = \begin{pmatrix} u_x \\ u_y \\ u_z \end{pmatrix}$ Diffusion Tensor $D = \begin{pmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{xy} & D_{yy} & D_{yz} \\ D_{xz} & D_{yz} & D_{zz} \end{pmatrix}$	$\left(\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y}, \frac{\partial g}{\partial z}, \frac{\partial^2 g}{\partial x^2}, \frac{\partial^2 g}{\partial y^2}, \frac{\partial^2 g}{\partial z^2}, \frac{\partial^2 g}{\partial x \partial y}, \frac{\partial^2 g}{\partial x \partial z}, \frac{\partial^2 g}{\partial y \partial z}, \frac{\partial g}{\partial t}\right)^T$	$(u_x, u_y, u_z, -D_{xx}, -D_{yy}, -D_{zz}, -2D_{xy}, -2D_{xz}, -2D_{yz}, 1)^T$

[a]**d** and **p** are the respective data and parameter vectors as defined in section 3.2.

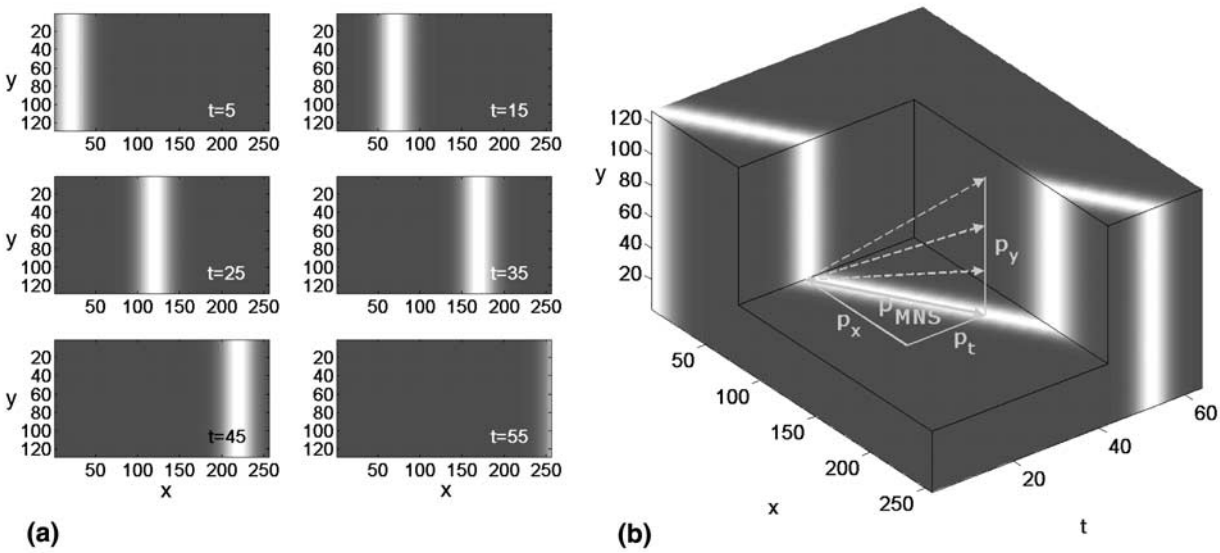

Figure 5. (a) 2D image sequence (x = 1..256, y = 1..128, t = 1..64) with a so-called aperture problem: Because the moving vertical line has structure only in x-direction, its velocity can obviously also be estimated only in x-direction. (b) In the corresponding 3D representation, this problem is reflected by the nonuniqueness of the direction of minimal intensity change. A unique solution, which represents the velocity vector in x-direction, is then given by the minimum norm solution $\mathbf{p}_{\mathrm{MNS}}$.

and the maximum likelihood estimate of $\mathbf{p}_{\min}$ is given by the rescaled eigenvector to the smallest eigenvalue $\lambda_{\min\ \mathrm{eq}}$ of J_{eq} according to $\mathbf{p}_{\min} = \mathsf{W}\ \mathbf{p}_{\mathrm{mineq}}$.

3.4. Aperture Problem

[32] In practical applications it often occurs that the data allows only for the estimation of a subset of the parameters. Figure 5 illustrates such a situation for the 2D motion estimation discussed in section 3.1. Since the moving object has structure only in one direction, its velocity can obviously be estimated only in the perpendicular direction. The best possible description of motion in this case of a so-called aperture problem is given by

$$\mathbf{u} = \begin{pmatrix} u_x \\ 0 \end{pmatrix} + k \begin{pmatrix} 0 \\ u_y \end{pmatrix}, \tag{48}$$

where the one degree of freedom is represented by the unknown variable k. For the TLS parameter estimation this involves a rank reduction of J by one and consequently an additional second zero eigenvalue. Accordingly the estimate of $\mathbf{p}$ is given by any linear combination of the two eigenvectors spanning the nullspace of J. Due to the additional constraint $p_3 = 1$ the estimate of $\mathbf{u}$ then has one degree of freedom and can be described by (48). It is one of the main features of the TLS method that the eigenvalue spectrum of J provides an accurate characterization of the available information. For the example of 2D motion the different cases can be classified as follows:

[33] 1. $\lambda_1 = \lambda_2 = \lambda_3 = 0$: constant gray value → no further analysis.

[34] 2. $\lambda_1 > 0$, $\lambda_2 = \lambda_3 = 0$: spatial orientation and constant motion → only one velocity component can be estimated.

[35] 3. $\lambda_1, \lambda_2 > 0$, $\lambda_3 = 0$: distributed spatial structure and constant motion → both velocity components can be estimated.

[36] 4. $\lambda_1, \lambda_2, \lambda_3 > 0$: distributed spatial structure and non-constant motion → no further analysis.

[37] If more than one eigenvalue of J is zero and therefore the estimated parameters have one or more degrees of freedom, a unique solution can be obtained through the application of additional regularization constraints. A common approach is the so-called minimum norm solution [*Spies*, 2001], defined by $\|\mathbf{p}\|^2 = \min$. The corresponding unique solution for the case of an ambigious 2D motion estimate represented by (48) is $k = 0$ or $\mathbf{u} = (u_x, 0)^T$. For models of processes with anisotropic diffusion however, this constraint leads to unphysical results and a modified criterion based on physical constraints as presented by [*Stöhr*, 2003] has to be applied instead.

[38] In practical calculations an eigenvalue will hardly be exactly zero and therefore the numerical distinction between $\lambda = 0$ and $\lambda > 0$ requires some further considerations which are given in section 4.5.

4. Computational Issues

[39] The gradient-based technique described in the previous section estimates parameters from partial derivatives of a scalar quantity as e.g. concentration or intensity. In the course of the practical calculation of the final estimates from the initial data a series of critical computational issues arises. Appropriate choices for the respective numerical transformations are mandatory for an adequate interpretation and optimal accuracy of the resulting parameters. TLS methods are a large and active topic of research mainly in the field of computer vision. In the following section the most important aspects for the implementation of the

gradient-based technique are addressed and references for further details are given. For compactness the discussion is restricted to a single spatial dimension x. Extensions to higher dimensions are straightforward.

4.1. Suppression of Noise

[40] The initially available data typically consists of a spatially and temporally resolved set of the scalar quantity $g(x_i, t_j)$ at discrete positions x_i and times t_j ($i = 1..N_x$, $j = 1..N_t$) on a regular grid with size (N_x, N_t). If x_i or t_j are not regularly spaced, an appropriate interpolation onto a regular grid is necessary for the application of the subsequent calculations.

[41] The accuracy of the finally estimated parameters is strongly affected by noise in the initial data. As a first step it is therefore essential to suppress noise as far as possible. In general, the optimal choice of a method for reducing the noise without affecting the underlying signal is a challenging issue. The most common approach is the convolution of g with a suitable filter mask $h(x_k, t_l)$:

$$g'(x_i, t_j) = h * g = \sum_{k=-R_x}^{R_x} \sum_{l=-R_t}^{R_t} h(x_k, t_l) g(x_i - x_k, t_j - t_k) \quad (49)$$

The shape of h is usually a gaussian with an adequately chosen number of coefficients n and variance σ^2 for each dimension, as e.g. (0.05, 0.24, 0.4, 0.24, 0.05) for a 1D filter with $n = 5$ and $\sigma^2 = 1$. Several more sophisticated techniques based on wavelets, splines, anisotropic diffusion, simulated annealing etc. are available.

4.2. Computation of Partial Derivatives

[42] The partial derivatives of g, which are the basis for the calculation of J according to (45), are then computed by the convolution of g with an adequate filter mask h as defined in (49). The actual choice of h is commonly a tradeoff between the size of the filter mask and the desired properties of its transfer function [*Jähne et al.*, 1999]. As a first order derivative with 5 coefficients, the 1D filter mask $(-1, 6, 0, -6, 1)/8$ is optimal relative to deviations from the ideal transfer function [*Jähne et al.*, 1999]. Higher order derivatives are then obtained by repeated convolutions of g with the chosen 1D filter mask directed in the respective dimensions.

4.3. Calculation of the Equilibration Weight Matrix

[43] As discussed in sections 2.3 and 3.3, the data vector $\mathbf{d}$ must be equilibrated with a weight matrix W in order to obtain a maximum likelihood estimate for $\mathbf{p}$. The elements of W are related to the uncertainties of the elements of $\mathbf{d}$, which are therefore required to be known a priori. According to the discussion above, the respective elements of $\mathbf{d}$ are given by convolutions of g with particular filter masks h. If the noise in g is independent and identically distributed with zero mean and variance σ_g^2, the variance of the noise in $h * g$ is given by

$$\sigma_{h*g}^2 = \sigma_g^2 \sum_{k=-R_x}^{R_x} \sum_{l=-R_t}^{R_t} h^2(x_k, t_l). \quad (50)$$

According to section 2.3, only the relative values of the elements of W are required for a maximum likelihood estimation. The value of σ_g^2, which is usually not known a priori, can therefore be set to $\sigma_g^2 = 1$.

[44] As an example, the variance for the first order derivative filter $h = (-1, 6, 0, -6, 1)/8$ is $\sigma_h^2 = 1.16$. The variance for the corresponding second order derivative $h^2 = h * h = (1, -12, 36, 12, -74, 12, 36, -12, 1)/64$ is $\sigma_{h*h}^2 = 2.11$. This demonstrates the higher noise sensitivity of the second order derivative. The respective values of $1/\sigma$ then form the elements of W according to (47). For further details about equilibration, e.g., if the above assumption for noise does not hold, see *Mühlich and Mester* [1999].

4.4. Choice of a Local Neighborhood

[45] An crucial point is the proper choice of the spatio-temporal local neighborhood represented by the window function w defined in (35), since this decides if the method can be applied to an experimental data set with a given heterogeneity and spatiotemporal resolution. In general, the neighborhood must include a number of linear independent data vectors $\mathbf{d}$ which is equal or higher than the number of unknowns, i.e. parameters. If the number is lower, a subset of the parameters may be estimated as explained in section 3.4. Since the data vectors $\mathbf{d}$ are virtually always contaminated with noise, an increased size of the neighborhood will reduce the uncertainty of the final estimates. If it is however chosen too large so that the basic requirement of constant parameters throughout the neighborhood is violated, the uncertainty will again strongly increase (see examples in section 5).

[46] There are no general restrictions regarding the shape of the neighborhood: it may extend exclusively in the spatial or temporal domain, or in both. If the parameters can be assumed to be stationary, it is reasonably extended over the full temporal domain. All elements may be weighted equally (w = const.) or with individual values $w(\mathbf{x}, t)$.

[47] The above discussion is strongly coupled to the question how many samples are needed to estimate a set of parameters. For a single point in the neighborhood, a number of surrounding samples is needed for the computation of the partial derivatives at that point. This number is given by the sizes of the respective filter masks. An analogous set of additional surrounding samples is required for the full neighborhood.

4.5. Measure of Confidence

[48] As a final step, we wish to determine the confidence of the parameter estimates. This depends on the noise in the data and on the degree to which the chosen model corresponds to reality.

[49] The determination of suitable measures of confidence for TLS problems is an object of ongoing research [*Nestares et al.*, 2000]. Because of the nonlinearity of the involved eigenvalue problem, the errors of the TLS estimates are, in contrast to OLS, nonlinear functions of the data and the noise level. Whereas TLS is relatively insensitive against low noise levels, higher levels can change the order of the eigenvalues and thereby strongly affect the solution [*Mühlich and Mester*, 1999]. Most of the approaches are based on the analysis of the eigenvalue spectrum of J [*Spies*, 2001; *Stöhr*, 2003]. When the data is contaminated with noise, the smallest eigenvalue of J is no longer zero. In the case of additive, independent noise with

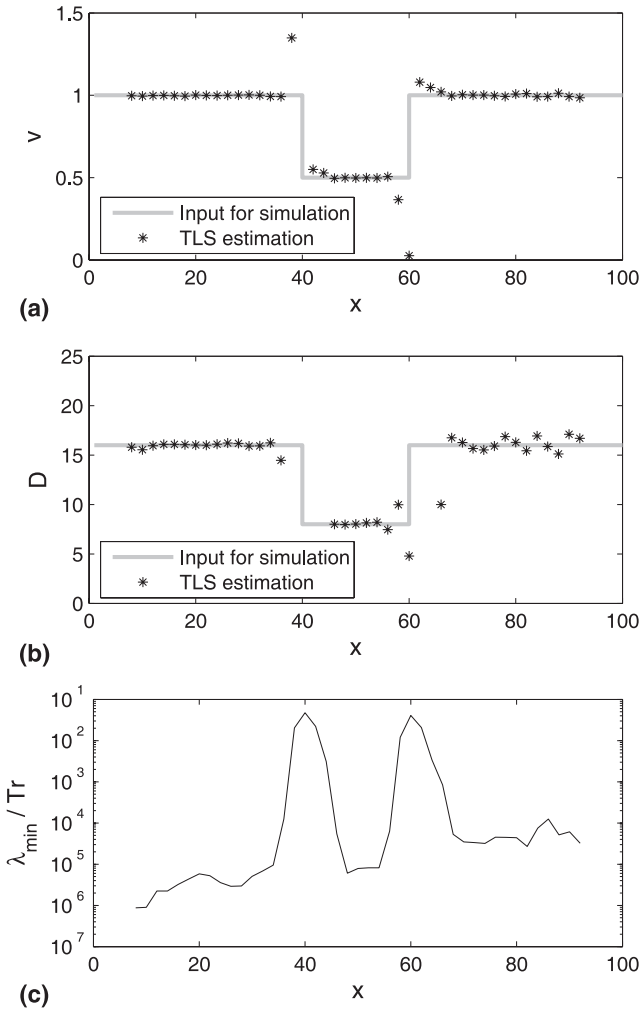

Figure 6. Results of the TLS parameter estimation from the numerical solution of a 1D-CDE with a low velocity/dispersion zone at $40 < x < 60$ shown in Figure 1a. Comparison of local estimates of (a) velocity $v(x)$ and (b) dispersion coefficient $D(x)$ with the values used for simulation. (c) Confidence measure $\frac{\lambda_{\min}}{\mathrm{Tr}(\mathsf{J})}$.

variance σ^2 in the equilibrated data matrix $\mathbf{d}_{\mathrm{eq}}$ defined in (47), the eigenvalues λ_i of J are shifted according to

$$\lambda_i \rightarrow \lambda_i + k\sigma^2, \tag{51}$$

where k is determined by the window function w as $k = \sqrt{\sum_{\mathbf{x},t} w^2(\mathbf{x},t)}$ [*Spies*, 2001]. Consequently, the smallest eigenvalue provides an estimate of the noise in the data. The consideration of this shift is mandatory for the classification of the eigenvalue spectrum as described in section 3.4. Aside from that, the smallest eigenvalue will also be shifted if the chosen model cannot describe the dynamic process in the data.

[50] Another relevant quantity is the trace of J, which represents the signal strength of the data ($\mathrm{Tr}(\mathsf{J}) = \sum_i \mathsf{J}_{ii}$). It has been shown [*Stöhr*, 2003] that the ratio of the trace and the smallest eigenvalue of J,

$$\omega_c = \frac{\mathrm{Tr}(\mathsf{J})}{\lambda_{\min}}, \tag{52}$$

can be interpreted as a signal-to-noise ratio and serves as a measure of confidence.

5. Application to Simulated Data

[51] In this section the application of the gradient-based TLS method are demonstrated by means of the local parameter estimation from simulated data sets. The reason for the usage of simulated data is the availability of the true parameters for comparison with the results of the estimation. For the simulations the spatial distribution of velocities and dispersion coefficients and an initial concentration distribution have been a priori chosen and then the concentration distributions for the following times were calculated using a particle tracking method.

5.1. One-Dimensional Convection-Dispersion

[52] At first we use the method for the estimation of local velocities and dispersion coefficients from the solutions $c(x, t)$ of the 1D CDE (2) for the region $x \in [1, 100]$, $t \in [1, 100]$ and the problems shown in Figure 1. The concentrations are calculated on a discrete grid with $\Delta x = 1$ and $\Delta t = 1$.

[53] Since concentrations obtained from particle tracking simulations include fluctuations owing to the limited number of particles, the initial step is the reduction of noise. Therefore the data was smoothed by convolution with a 1D Gaussian kernel ($n = 5$, $\sigma = 1.4$) in both x- and t-direction as described in section 4.1.

[54] The next step is then the computation of the derivatives $\frac{\partial c}{\partial x}$, $\frac{\partial^2 c}{\partial x^2}$ and $\frac{\partial c}{\partial t}$, the elements of $\mathbf{d}$ for this model, from the smoothed data. According to section 4.2, the first and second order derivatives were computed by convolutions with the filter masks $(-1, 6, 0, -6, 1)/8$ and $(1, -12, 36, 12, -74, 12, 36, -12, 1)/64$ respectively.

[55] The size of the local neighborhoods (see section 4.4) in x-direction is $\Delta x = 3$, and the parameters were calculated at the central positions $x_c = 8, 10, 12, \ldots, 92$. Under the assumption of stationary parameters the whole timeframe, excluding the boundaries, was chosen as the temporal neighborhood, i.e. $t = 7..94$. All elements were weighted equally, so that the corresponding window function $w(x - x', t)$ is a box function with $w = 1/(3*88)$ for $-1 \leq x' \leq 1 \vee 7 \leq t \leq 94$ and else $w = 0$.

[56] The tensor J was then calculated for every x according to (45) and the parameters $v(x_c)$ and $D(x_c)$ were obtained from the eigen analysis of the equilibrated tensor $\mathsf{J}_{\mathrm{eq}}(x_c)$ in analogy to (46). The results presented in Figures 6 and 7 show a good agreement with the a priori chosen values used for the simulation in the areas where $v(x)$ and $D(x)$ are constant. For increasing x the errors of the estimates increase due to the decreasing signal-to-noise ratio of the data. Because the susceptibility to noise is higher for the second than for the first order derivative, the relative errors are higher for D than for v. At the locations where $v(x)$ and $D(x)$ are discontinuous ($x = 40$ and 60 in Figure 1a) large errors occur because here the basic requirement for the BCCE (33), i.e. the constancy of the parameters, is violated.

[57] A measure of uncertainty, given by the inverse of the confidence measure ω_c defined in (52) is plotted in Figures 6c and 7c. The comparison with the corresponding estimates of $v(x)$ and $D(x)$ shows that this measure correlates well with the incorrect results at the discontinuities and

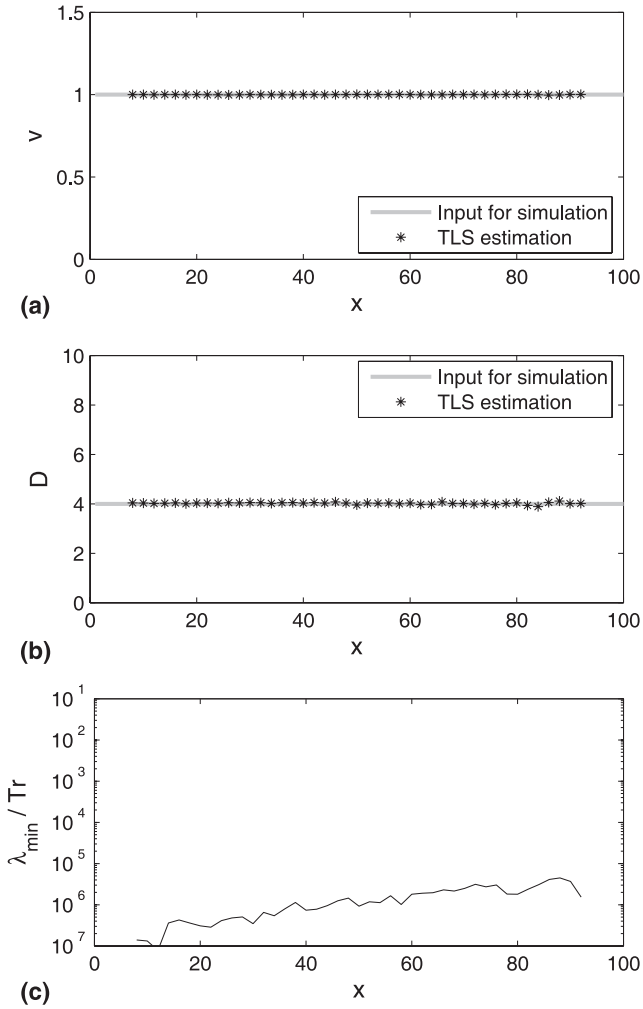

Figure 7. Results of the TLS parameter estimation from the numerical solution of a 1D-CDE with constant $v(x)$ and $D(x)$ and a heterogeneous initial distribution shown in Figure 1b. Comparison of local estimates of (a) velocity $v(x)$ and (b) dispersion coefficient $D(x)$ with the values used for simulation. (c) Confidence measure $\frac{\lambda_{\min}}{\mathrm{Tr}(\mathsf{J})}$.

also with the increasing errors due to the decreasing signal-to-noise ratio at $x > 60$.

5.2. Two-Dimensional Convection-Diffusion

[58] In the following the method is demonstrated for the model of 2D motion with isotropic diffusion, which has been discussed in section 3.2.

[59] An array of local velocities $\mathbf{v}(\mathbf{x})$ (x = 1..500, y = 1..500) has been obtained from a numerical solution of the Darcy equation for a heterogeneous medium with a zone in the center where the permeability is half the value of the outer region. Then a solution $c(\mathbf{x}, t)$ (t = 1..200) of the CDE (42)–(43) with homogeneous D was computed using a particle tracking simulation. From this solution, which is shown in Figure 8, the local velocities and diffusion coefficients have been estimated with the present method.

[60] Analogously to the previous 1D example the data was first smoothed and then the derivatives $\frac{\partial c}{\partial t}$, $\frac{\partial c}{\partial x}$, $\frac{\partial c}{\partial y}$, $\frac{\partial^2 c}{\partial x^2}$, $\frac{\partial^2 c}{\partial y^2}$, $\frac{\partial^2 c}{\partial x \partial y}$, which are the elements of the data vector for the corresponding model (see Table 1), were calculated. The tensors $\mathsf{J}_{\mathrm{eq}}(\mathbf{x}_c)$ were computed at the central positions $x_c = y_c$ = 26, 32, 38,..,470 with a box-shaped local neighborhood of $\Delta x = \Delta y = 6$. Assuming stationary parameters, the neighborhood extends over the whole timeframe except the boundaries (Δt = 180). The resulting TLS estimates of $\mathbf{v}(\mathbf{x}_c)$ and $D(\mathbf{x}_c)$ are shown in Figure 9. As expected, the velocities in the central low permeability zone are lower than those in the outer region, whereas the diffusion coefficients are consistent in both regions. The results are not shown for those locations where the inverse confidence measure ω_c^{-1} is higher than an empiric threshold $\tau = 0.62 \cdot 10^{-5}$, which indicates that the estimation has failed. As expected this happens in two different situations: First, at the boundary between high and low permeability, where the basic requirement that $\mathbf{v}(\mathbf{x})$ and $D(\mathbf{x})$ are constant in the local neighborhood (see section 3.1) is violated and consequently no meaningful result can be obtained. As a consequence of the applied convolutions, the effect of the boundary propagates over a certain distance and leads to the observable distortions of the velocity vectors along the boundary. Second, at the outer locations, where only small

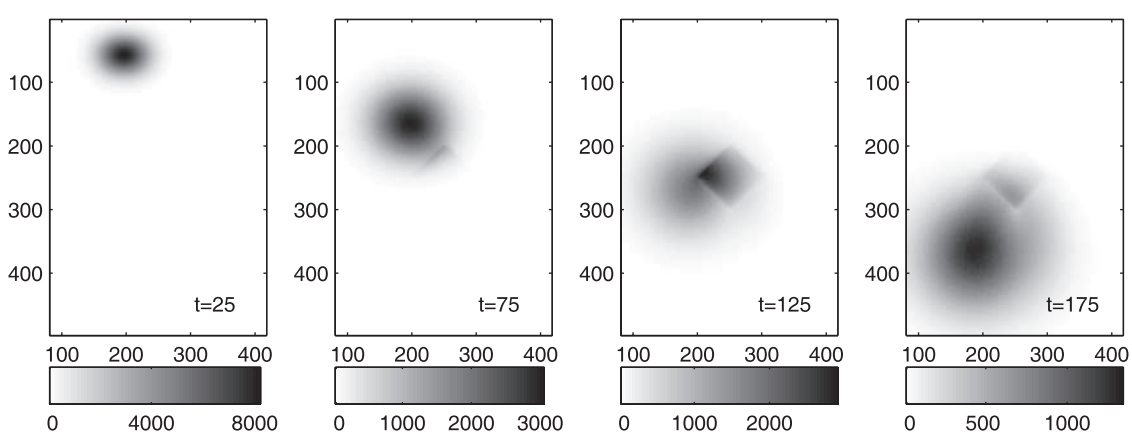

Figure 8. Numerical solution $c(\mathbf{x}, t)$ (x = 1..500, y = 1..500, t = 1..200) of the CDE (2) for a medium with a low permeability zone in the center.

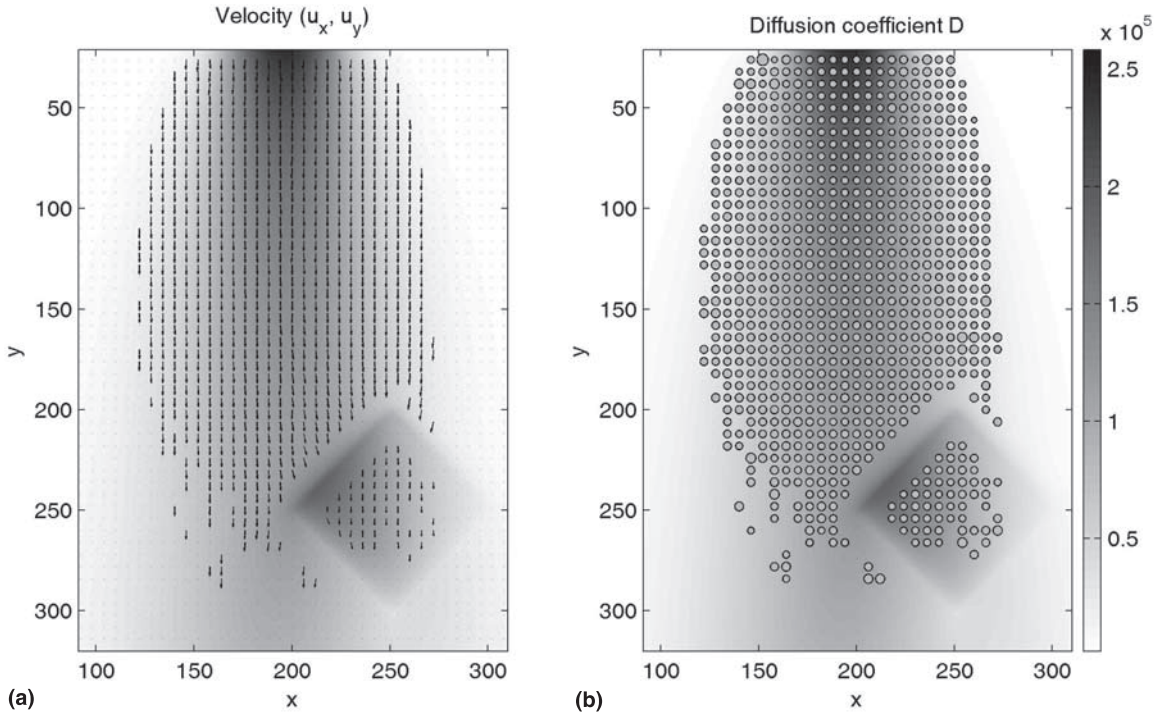

Figure 9. (a) Local velocities $\mathbf{v}(\mathbf{x})$ and (b) diffusion coefficients $D(\mathbf{x})$ (represented by the radii of the circles) estimated from the simulated time series $c(\mathbf{x}, t)$ shown in Figure 8. The gray values represent the cumulated concentrations $\sum_{t=1}^{200} c(\mathbf{x}, t)$ as a proxy for the local information content. Estimates are only shown for $\omega_c^{-1} < 0.85 \cdot 10^{-5}$ (see Figure 10).

or no concentration is present, the available information is not sufficient to infer any parameters.

[61] In Figure 10, the values of the inverse confidence measure ω_c^{-1} are compared to the relative error $|\mathbf{p}_{\text{est}} - \mathbf{p}_{\text{true}}|/|\mathbf{p}_{\text{true}}|$ of the estimated parameters. Obviously ω_c^{-1} reflects both types of failed estimates: High values of ω_c^{-1} occur both at the boundary between the permeability zones and in the regions of low concentration. Figure 10b shows that the relative errors of the estimates plotted in Figure 9 are mostly below 10^{-2} and correlate well with the distribution of ω_c^{-1}.

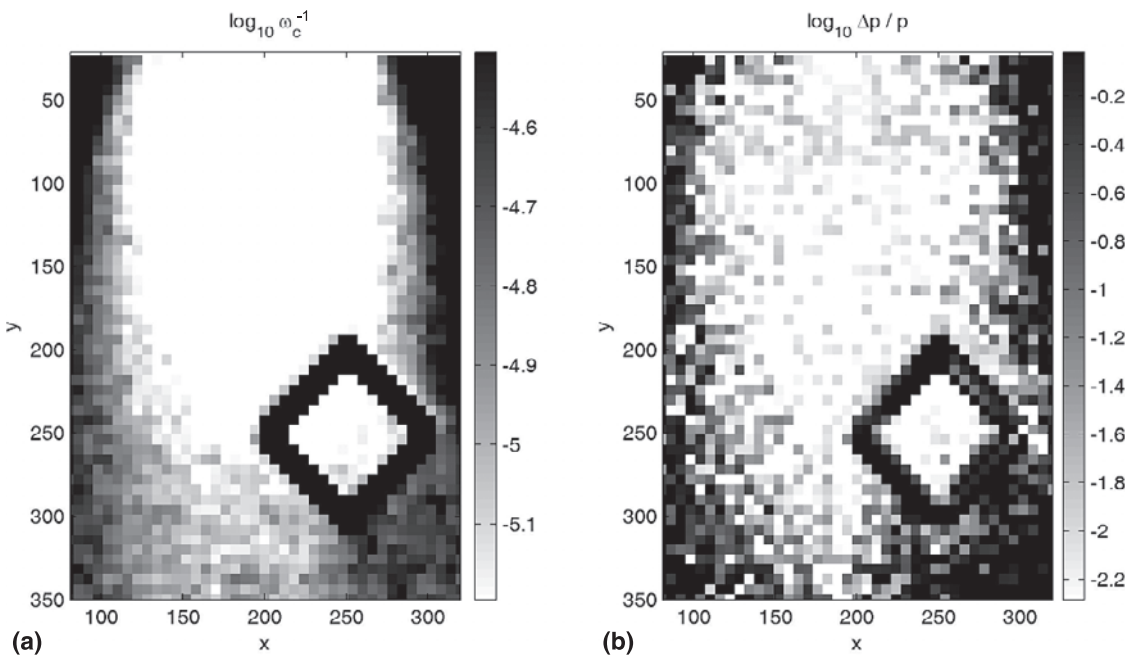

Figure 10. (a) Inverse measure of confidence $\omega_c^{-1} = \dfrac{\lambda_{\min}}{\mathrm{Tr}(\mathrm{J})}$ and (b) relative error of the estimates shown in Figure 9.

6. Summary and Conclusions

[62] Because commonly used approaches for the estimation of model parameters from experimental data, like, e.g., inverse modeling, require ergodicity and rely on the knowledge of initial and boundary conditions, they often cannot be employed since experimental conditions like heterogeneity in the porous medium do not meet these requirements. As an alternative approach aimed to overcome these difficulties a gradient-based method, which is already used in different fields of computer vision, has been introduced. The method is based on the formalism of TLS as a maximum likelihood estimator for the model parameters, which has therefore been described and distinguished from the common OLS. The issues of equilibration, measures of confidence and the identification and treatment of aperture problems have been addressed and the essential computational aspects of an implementation of the method have been explicated.

[63] On this basis, the employment of TLS to the gradient-based parameter estimation for dynamic processes provides several beneficial features.

[64] 1. The dynamic process which is to be estimated can be directly integrated into the estimation procedure in the form of a differential equation. A solution to this equation is not necessary.

[65] 2. The formalism allows for a direct specification of the location and size of the estimate in space and time. Any assumption about heterogeneities, as defined e.g. by a REV, can be included in the estimation.

[66] 3. The method yields a direct non-iterative maximum likelihood estimate with no need of any initial values.

[67] 4. The cause of a failure of the estimation, like e.g. a low signal-to-noise ratio or an aperture problem, can be readily detected from an eigenvalue spectrum and potentially be corrected by adapted additional constraints.

[68] 5. The essential algorithms are limited to a series of convolutions and eigenvalue calculations, which can be easily implemented and have a reasonable computational demand.

[69] The requirement of high-resolution data, which is inherent to any approach to local parameter estimation, is fulfilled by a growing number of experimental techniques for the laboratory and few promising results from field measurements.

[70] The method has been applied to numerical solutions of the CDE with spatially varying parameters $\mathbf{v}(x)$ and $\mathsf{D}(x)$ and heterogeneous initial conditions. Whereas the evaluation of $\mathbf{v}(x)$ and $\mathsf{D}(x)$ in such situations would hardly be feasible with commonly used methods, it has been demonstrated that the present method yields direct and accurate local estimates.

[71] **Acknowledgments.** We gratefully acknowledge the financial support from the graduate program "Modeling and Scientific Computing in Mathematics and Natural Sciences" of the Interdisciplinary Center for Scientific Computing (IWR) of the University of Heidelberg. We further appreciate the beneficial collaboration with the image processing research group of Bernd Jähne at the IWR.

References

Abbaspour, K. C., M. T. van Genuchten, R. Schulin, and E. Schläppi (1997), A sequential uncertainty domain inverse procedure for estimating subsurface flow and transport parameters, *Water Resour. Res.*, *33*, 1879–1892.

Berest, P., N. Rakotomalala, J. Hulin, and D. Salin (1999), Experimental and numerical tools for miscible fluid displacement studies in porous media with large heterogeneities, *Eur. Phys. J. Appl. Phys.*, *6*, 309–321.

Björck, A. (1996), *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, Pa.

Dagan, G. (1990), Transport in heterogeneous porous formations: Spatial moments, ergodicity, and effective dispersion, *Water Resour. Res.*, *26*, 1281–1290.

Dentz, M., H. Kinzelbach, S. Attinger, and W. Kinzelbach (2000a), Temporal behaviour of a solute cloud in a heterogeneous porous medium: 1. point-like injection, *Water Resour. Res.*, *36*, 3591–3604.

Dentz, M., H. Kinzelbach, S. Attinger, and W. Kinzelbach (2000b), Temporal behaviour of a solute cloud in a heterogeneous porous medium: 2. spatially extended injection, *Water Resour. Res.*, *36*, 3605–3614.

Detwiler, R., and H. Rajaram (2000), Solute transport in variable-aperture fractures: An investigation of the relative importance of Taylor dispersion and macrodispersion, *Water Resour. Res.*, *36*, 1611–1625.

Gelhar, L. W., C. Welty, and K. R. Rehfeldt (1992), A critical review of data on field-scale dispersion in aquifers, *Water Resour. Res.*, *28*(7), 1955–1974.

Haußecker, H., C. Garbe, H. Spies, and B. Jähne (1999), A total least squares framework for low-level analysis of dynamic scenes and processes, in *DAGM Symposium*, pp. 240–249, Springer, New York.

Jähne, B., H. Scharr, and S. Körkel (1999), Principles of filter design, in *Handbook of Computer Vision and Applications*, vol. 2, edited by B. Jähne, H. Haußecker, and P. Geißler, pp. 125–152, Elsevier, New York.

Keller, A., P. Roberts, and M. Blunt (1999), Effect of fracture aperture variations on the dispersion of contaminants, *Water Resour. Res.*, *35*, 55–63.

Kemna, A., J. Vanderborght, B. Kulessa, and H. Vereecken (2002), Imaging and characterisation of subsurface solute transport using electrical resistivity tomography (ert) and equivalent transport models, *J. Hydrol.*, *267*, 125–146.

Loggia, D., P. Gouze, R. Greswell, and D. Parker (2004), Investigation of the geometrical dispersion regime in a single fracture using positron emission projection imaging, *Transp. in Porous Media*, *55*, 1–20.

Mühlich, M., and R. Mester (1999), Subspace methods and equilibration in computer vision, *Tech. Rep. XP-TR-C-21*, Computer vision group, Institute for Applied Physics, Goethe-Univ., Frankfurt am Main, Germany.

Nestares, O., D. J. Fleet, and D. J. Heeger (2000), Likelihood functions and confidence bounds for total-least-squares problems, in *IEEE Conference on Computer Vision and Pattern Recognition, Hilton Head, South Carolina, vol. 1*, pp. 523–530, IEEE Press, Piscataway, N. J.

Ptak, T., M. Piepenbrink, and E. Martak (2004), Tracer tests for the investigation of heterogeneous porous media and stochastic modelling of flow and transport—A review of some recent developments, *J. Hydrol.*, *294*, 122–163.

Singha, K., A. Binley, J. Lane, J. W., and S. Gorelick (2003), Electrical imaging of tracer migration at the massachusetts military reservation, cape cod, in *Symposium on the Application of Geophysics to Engineering and Environmental Problems (SAGEEP), San Antonio, Texas, Proceedings: Denver, Colorado, Environmental and Engineering Geophysical Society*, pp. 1–11, IEEE Press, Piscataway, N. J.

Spies, H. (2001), Analysing dynamic processes in range data sequences, Ph.D. thesis, University of Heidelberg, Heidelberg, Germany.

Stöhr, M. (2003), Analysis of flow and transport in refractive index matched porous media, Ph.D. thesis, University of Heidelberg, Heidelberg, Germany.

Stöhr, M., K. Roth, and B. Jähne (2003), Measurement of 3d pore-scale flow in index-matched porous media, *Exp. Fluids*, *35*, 159–166.

Van Huffel, S., and P. Lemmerling (Eds.) (2002), *Total Least Squares and Errors-in-Variables Modeling: Analysis, Algorithms and Applications*, Springer, New York.

Van Huffel, S., and J. Vandewalle (1991), *The Total Least Squares Problem - Computational Aspects and Analysis*, Soc. for Ind. and App. Math, Philadelphia, Pa.

Zimmerman, D. A., et al. (1998), A comparison of seven geostatistically based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow, *Water Resour. Res.*, *34*, 1373–1413.

K. Roth, Institute for Environmental Physics, Im Neuenheimer Feld 229, D-69120 Heidelberg, Germany. (kurt.roth@iup.uni-heidelberg.de)

M. Stöhr, Max Planck Institute for Marine Microbiology, Celsiusstrasse 1, D-28359 Bremen, Germany. (mstoehr@mpi-bremen.de)