Math Geosci (2009) 41: 509–533
DOI 10.1007/s11004-009-9220-x

Application of FFT-based Algorithms for Large-Scale Universal Kriging Problems

J. Fritz · I. Neuweiler · W. Nowak

Received: 11 September 2007 / Accepted: 1 September 2008 / Published online: 1 April 2009

Abstract Looking at kriging problems with huge numbers of estimation points and measurements, computational power and storage capacities often pose heavy limitations to the maximum manageable problem size. In the past, a list of FFT-based algorithms for matrix operations have been developed. They allow extremely fast convolution, superposition and inversion of covariance matrices under certain conditions. If adequately used in kriging problems, these algorithms lead to drastic speedup and reductions in storage requirements without changing the kriging estimator. However, they require second-order stationary covariance functions, estimation on regular grids, and the measurements must also form a regular grid. In this study, we show how to alleviate these rather heavy and many times unrealistic restrictions. Stationarity can be generalized to intrinsicity and beyond, if decomposing kriging problems into the sum of a stationary problem and a formally decoupled regression task. We use universal kriging, because it covers arbitrary forms of unknown drift and all cases of generalized covariance functions. Even more general, we use an extension to uncertain rather than unknown drift coefficients. The sampling locations may now be irregular, but must form a subset of the estimation grid. Finally, we present asymptotically exact but fast approximations to the estimation variance and point out application to conditional simulation, cokriging and sequential kriging. The drastic gain in

J. Fritz (✉) · W. Nowak
Institute of Hydraulic Engineering, Department of Hydromechanics and Modeling of Hydrosystems, Universität Stuttgart, Stuttgart, Germany
e-mail: jochen.fritz@iws.uni-stuttgart.de

I. Neuweiler
Institute for Fluid Mechanics, Leibniz Universität Hannover, Hannover, Germany
e-mail: Neuweiler@hydromech.uni-hannover.de

W. Nowak
Civil and Environmental Engineering, University of California, Berkeley, CA, USA
e-mail: Wolfgang.Nowak@iws.uni-stuttgart.de

computational and storage efficiency is demonstrated in test cases. Especially high-resolution and data-rich fields such as rainfall interpolation from radar measurements or seismic or other geophysical inversion can benefit from these improvements.

Keywords Fast Fourier transform · Efficient geostatistical estimation · Spectral methods · Superfast Toeplitz solver

1 Introduction

Spatially distributed quantities such as rainfall intensities, contaminant concentrations, or hydraulic conductivities are frequently interpolated between scattered measurements by kriging. Large data sets and finely resolved estimation grids (Wesson and Pegram 2004) can easily exceeds storage capacities and computational power of desktop computers. Our goal is to make kriging low on memory and computational costs, in order to evaluate vast kriging problems without reverting to supercomputers. In most forms of kriging, the three computationally most demanding tasks are: obtaining the kriging weights from an $m \times m$ system of equations with the auto-covariance matrix of m measurements, obtaining the estimate from superposition of the kriging weights with the cross-covariance between measurements and unknowns, and repeating this procedure m times to obtain the estimation variance. If the covariance function is second-order stationary and the n points of estimation form a regular and equispaced grid, the auto-covariance matrix of the unknowns has symmetric Toeplitz structure. In this case, only its first column has to be stored, reducing storage from n^2 to n matrix elements (Zimmerman 1989). FFT-based algorithms allow to evaluate the product of Toeplitz matrices with vectors (identical to discrete convolution) quickly in $\mathcal{O}(n \log n)$ operations (van Loan 1992), while standard evaluation requires $\mathcal{O}(n^2)$ operations.

The present study will generalize the applicability of FFT-based methods from stationary to intrinsic cases, i.e., to universal kriging (Kitanidis 1997) extended towards Bayesian geostatistics (Kitanidis 1986). Universal kriging covers arbitrary forms of unknown drift: polynomial and other trend functions, zonation models, external drifts, and any other set of explanatory regression variables. It also covers all cases described by generalized covariance functions (Kitanidis 1993). For estimation on fine grids, we choose the dual formulation (function estimate form in Kitanidis 1996, 1997). Coming from cokriging-like cases (Nowak and Cirpka 2004), extended universal kriging for uncertain prior knowledge on the drift coefficients. We will adopt this extension because it is most general in treating the drift. A specific case of Bayesian geostatistics (Kitanidis 1986) accounts for the uncertainty in the definition of the mean value. It is closely related to the Bayesian kriging idea by Omre (1987), which allows to include qualified guesses and expert knowledge into the kriging procedure without introducing artificial data points. The key for FFT-based methods in intrinsic problems is a decomposition of universal kriging into a stationary problem and a coupled regression problem. In the end, we merely evaluate the stationary parts via FFT, leaving the kriging estimate itself untouched and free of approximations. Splitting off a stationary part from intrinsic problems bears much resemblance to the

global neighborhood kriging technique by Davis and Grivet (1984). In contrast to the current study, however, they did not use the dual formulation. Therefore, they could not draw on the power of superposition via FFT. Since FFT-PCG solvers had not been invented at that time, they were limited to much smaller data sets than those targeted here. Their alternative solver for banded covariance matrices required a regular grid and a defined range of the covariance function. Finally, they did not generalize towards prior knowledge on the drift coefficients.

Many of our ideas extend the work of Pegram (2004). For the case of simple and ordinary kriging (known and unknown constant mean, respectively), he traced the superposition in task (2) back to convolution and executed it via FFT. This reduces computational costs from $\mathcal{O}(mn)$ to $\mathcal{O}(n \log n)$ and is a great advantage for medium and large numbers of measurements ($m > \log_2 n$). Contrary to the methods suggested in the current study, his method does not allow for arbitrary intrinsic cases. A more fundamental problem is to solving the kriging system of equations. Nearest neighborhood kriging avoids large systems by only considering measurements within a certain radius around each point of estimation. This requires setting up and solving as many as n individual systems, which may be very time consuming (Davis and Grivet 1984; Kitanidis 1997), and implicitly assumes a moving-average type of mean. The largest drawbacks of working with neighborhoods are spurious discontinuities in the estimate (Davis and Culhane 1984; Davis and Grivet 1984) and the inability to account for boundary conditions. A more sophisticated reduction of the kriging system is sequential kriging (Vargas-Guzmán and Yeh 1999). Splitting the data into convenient subsets, this approach is quite promising if the main problem is the size m of the data set. Unfortunately, in the current form, it occupies memory of $\mathcal{O}(n^2)$ for intermediate conditional covariance matrices, impossible for large n. We reduce this to $\mathcal{O}(mn)$ and show how to use FFT-based methods for the sequential case.

If the measurements form a regular grid, the universal kriging system contains a symmetric and positive-definite Toeplitz matrix. Solving such systems has been studied extensively in the signal processing community (Gallivan et al. 1996; Kailath and Sayed 1999; Van Barel et al. 2001). The iterative Toeplitz solver that we find the most promising within the kriging context is the FFT-based Preconditioned Conjugate Gradient (FFT-PCG) method with circulant preconditioners (Chan and Ng 1996). It solves the kriging system with computational effort of $\mathcal{O}(m \log m)$ instead of $\mathcal{O}(m^3)$. If the measurements are irregularly spaced, the Toeplitz structure within the kriging system is lost, and the existing FFT-PCG is not applicable. Pegram (2004) solves the $m \times m$ kriging equations using FFT-based iterative constrained deconvolution (ICD) in $\mathcal{O}(n \log n)$ while storing $\mathcal{O}(n)$ elements instead of $\mathcal{O}(m^2)$, a crucial advantage for large data sets. ICD allows the measurements to be irregularly scattered if they lie on the regular grid estimation. We will upgrade this by combining it with the FFT-PCG algorithm. Also, Pegram (2004) neglects correlation among the data when estimating the global mean parameter, and we will avoid this approximation. The remaining problem is that both tasks need to be performed m times in order to obtain the estimation variance. We show how to improve this by approximations in several levels of trade-off versus accuracy. These approximations are asymptotically exact for certain special cases. Again, one of our suggestions is an improved version of an approximation already used by Pegram (2004), assuming measurements to be almost

uncorrelated. Finally, we demonstrate how these methods allow for handling huge kriging problems even on ordinary desktop computers in a series of performance tests. Application to related problems such as generating conditional realizations, sequential kriging and cokriging is pointed out. Ensembles of conditional realizations can be used to evaluate the estimation variance in case none of our approximations apply.

2 Extended Universal Kriging

The following briefly summarizes universal kriging, extended to prior knowledge on the drift coefficients by Nowak and Cirpka (2004). The extension helps to address the drift in a most generalized fashion. The choice of universal kriging is the key step for transferring FFT-based methods to intrinsic cases.

Let $\mathbf{s}$ be an $n \times 1$ multi-Gaussian vector of unknowns (also called target point values in regression-like problems). It has expectation $E[\mathbf{s}|\boldsymbol{\beta}] = \mathbf{X}\boldsymbol{\beta}$ and second-order stationary covariance $\mathrm{Cov}[\mathbf{s}|\boldsymbol{\beta}] = \mathbf{Q_{ss}}$. $\mathbf{X}$ denotes the $[n \times p]$ matrix of discrete drift functions with its $[p \times 1]$ vector $\boldsymbol{\beta}$ of drift coefficients. Both $\mathbf{Q_{ss}}$ and $\mathbf{X}$ must be known a priori and constitute the geostatistical model assumption. By adequate choice of $\mathbf{X}$, all intrinsic cases can be addressed and decomposed into a second-order stationary part plus a regression-like drift problem (Kitanidis 1997, p. 125 ff.). In the extended case, $\boldsymbol{\beta}$ is again Gaussian with prior mean $\boldsymbol{\beta}^*$ and covariance $\mathbf{Q}_{\boldsymbol{\beta}\boldsymbol{\beta}}$ instead of being entirely unknown. The special case of unknown drift is recovered by setting $\mathbf{Q}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{-1} = 0$. The covariance of $\mathbf{s}$ for uncertain $\boldsymbol{\beta}$ is $\mathbf{G_{ss}} = \mathbf{Q_{ss}} + \mathbf{X}\mathbf{Q}_{\boldsymbol{\beta}\boldsymbol{\beta}}\mathbf{X}^T$. For polynomial trends with unknown coefficients, $\mathbf{G_{ss}}$ is known as generalized covariance (Kitanidis 1993). Consider further $\mathbf{Y}$ an $m \times 1$ vector of measurements (control point values in regression-like problems). The corresponding cross- and auto-covariance matrices are denoted by $\mathbf{Q_{sy}}$ and $\mathbf{Q_{yy}}$, respectively, sized $n \times m$ and $m \times m$. If measurements are subject to error, an error covariance matrix $\mathbf{R}$ is added to $\mathbf{Q_{yy}}$, typically a scalar matrix (i.e., a diagonal matrix with a constant term on the diagonal) when assuming homoscedasticity. In this notation, the kriging estimate $\hat{\mathbf{s}}$ is given by

$$\hat{\mathbf{s}} = \begin{bmatrix} \mathbf{Q_{ys}} \\ \mathbf{X}^T \end{bmatrix}^T \begin{bmatrix} \boldsymbol{\xi} \\ \widehat{\boldsymbol{\beta}} \end{bmatrix}. \tag{1}$$

The $m \times 1$ vector of kriging weights $\boldsymbol{\xi}$ (reciprocal data in Pegram 2004 and the $p \times 1$ vector of trend coefficients $\hat{\boldsymbol{\beta}}$ are taken from the solution of the following kriging system, Nowak and Cirpka 2004)

$$\begin{bmatrix} \mathbf{Q_{yy}} & \mathbf{x} \\ \mathbf{x}^T & -\mathbf{Q}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\xi} \\ \widehat{\boldsymbol{\beta}} \end{bmatrix} = \begin{bmatrix} \mathbf{Y} \\ -\mathbf{Q}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{-1}\boldsymbol{\beta}^* \end{bmatrix}, \tag{2}$$

where the drift functions $\mathbf{X}$ evaluated at the locations of measurements are denoted by the smaller $m \times p$ matrix $\mathbf{x}$. The associated estimation variance $\hat{\boldsymbol{\sigma}}$ is the $n \times 1$ vector on the diagonal of the conditional covariance matrix

$$\mathbf{Q_{ss|y}} = \mathbf{Q_{ss}} - \begin{bmatrix} \mathbf{Q_{ys}} \\ \mathbf{X}^T \end{bmatrix}^T \begin{bmatrix} \mathbf{Q_{yy}} & \mathbf{x} \\ \mathbf{x}^T & -\mathbf{Q}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{Q_{ys}} \\ \mathbf{X}^T \end{bmatrix}. \tag{3}$$

The conditional distribution of $\boldsymbol{\beta}$ is again Gaussian with mean $E[\boldsymbol{\beta}|\mathbf{s}] = \hat{\boldsymbol{\beta}}$ from (2) and covariance $\mathbf{Q}_{\boldsymbol{\beta}\boldsymbol{\beta}|\mathbf{s}}$ from (15). Equations (1) and (2) constitute a best linear unbiased estimator and were originally derived from Bayesian principles. $\mathbf{Q}_{\mathbf{ss}}$ can be seen as the covariance matrix of regression error, and $\mathbf{Q}_{\mathbf{ss}|\mathbf{y}}$ as that of the kriging error. The expensive tasks mentioned in the introduction are solving (2) to obtain the kriging weights, performing the superposition $\mathbf{Q}_{\mathbf{sy}}\boldsymbol{\xi}$ in (1) to evaluate the estimate, and evaluating $m + p$ equivalents to task one and two to obtain the estimation variance from the diagonal of (3).

The latter fact is easily demonstrated by a small but insightful rearrangement of (3)

$$\mathbf{Q}_{\mathbf{ss}|\mathbf{y}} = \mathbf{Q}_{\mathbf{ss}} - \underbrace{\left(\begin{bmatrix} \mathbf{Q}_{\mathbf{ys}} \\ \mathbf{X}^T \end{bmatrix}^T \underbrace{\begin{bmatrix} \mathbf{Q}_{\mathbf{yy}} & \mathbf{x} \\ \mathbf{x}^T & -\mathbf{Q}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{-1} \end{bmatrix}^{-1} \mathbf{I}_{m+p}}_{\boldsymbol{\Xi}} \right)}_{\mathbf{S}} \begin{bmatrix} \mathbf{Q}_{\mathbf{ys}} \\ \mathbf{X}^T \end{bmatrix}. \tag{4}$$

Here, we inserted the $(m + p) \times (m + p)$ identity matrix $\mathbf{I}_{m+p}$ without changing the equation. With analogy to (2), $\boldsymbol{\Xi}$ can be seen as a set of $(m + p)$ kriging weight vectors $[\boldsymbol{\xi}_i; \boldsymbol{\beta}_i]$ obtained from the $(m + p)$ unit vectors $\mathbf{e}_i$ that constitute $\mathbf{I}_{(m+p)}$. By analogy to (1), this makes the matrix $\mathbf{S}$ a set of $(m + p)$ kriging estimates $\mathbf{s}_i$, each corresponding to a unit data vector $\mathbf{e}_i$.

3 Interface to FFT-based Methods

The targeted FFT-based methods handle $\mathbf{Q}_{\mathbf{ss}}$ and $\mathbf{Q}_{\mathbf{yy}}$ extremely efficiently in terms both of storage and computational effort, if $\mathbf{Q}_{\mathbf{ss}}$ is stationary. Handling intrinsic situations while assuring stationary $\mathbf{Q}_{\mathbf{ss}}$ has been discussed above. Now we will relate as many terms as possible to $\mathbf{Q}_{\mathbf{ss}}$ and decouple the regression-like drift terms from the stationary terms in each individual step. To install a formal notational framework, consider a $m \times n$ sampling matrix $\mathbf{H}$ defined as

$$\mathbf{H}_{i,j} = \begin{cases} 1 & \text{if } \mathbf{x}_i = \mathbf{x}_j, \\ 0 & \text{otherwise,} \end{cases} \tag{5}$$

where $\mathbf{x}_i$ are the coordinates of the ith measurement location and $\mathbf{x}_j$ the coordinates of the jth estimation point. $\mathbf{H}$ resembles the constraint operator used by Pegram (2004) or the missing data indicator in Fuentes (2007). These two studies, however, target a regular grid of data with a few missing values. In quasi-linear geostatistical estimation techniques with similarity to cokriging, $\mathbf{H}$ has widely been used as sensitivity matrix (Kitanidis and Vomvoris 1983; Kitanidis 1995). In fact, the definitions of $\mathbf{H}$ as indicator, discrete sampling array or sensitivity matrix are identical in our case.

Written as formal matrix operations for arbitrary vectors $\mathbf{a}$ and $\mathbf{A}$ (sized $m \times 1$ and $n \times 1$), sampling and its reverse operation, i.e., injection, are

$$\text{sampling:} \quad \mathbf{a}_{m\times 1} = \mathbf{H}\mathbf{A}_{n\times 1}, \tag{6}$$

$$\text{injection:} \quad \mathbf{A}_{n\times 1} = \mathbf{H}^T\mathbf{a}_{m\times 1}. \tag{7}$$

In practice, sampling is picking specific values $\mathbf{a}$ from the larger vector $\mathbf{A}$. An example at hand is that $\mathbf{x}$ is formally given by $\mathbf{HX}$, even if it would not be evaluated this way in practice. Injection is writing $\mathbf{a}$ into a larger vector of initial zeros $\mathbf{A}$ at the sampling positions. When interpreting $\mathbf{H}$ as a sensitivity matrix for linear error propagation (Schweppe 1973), the following identities hold

$$\mathbf{Q}_{\mathbf{ys}} = \mathbf{H}\mathbf{Q}_{\mathbf{ss}}, \tag{8}$$

$$\mathbf{Q}_{\mathbf{sy}} = \mathbf{Q}_{\mathbf{ss}}\mathbf{H}^T, \tag{9}$$

$$\mathbf{Q}_{\mathbf{yy}} = \mathbf{H}\mathbf{Q}_{\mathbf{ss}}\mathbf{H}^T + \mathbf{R}, \tag{10}$$

which express all appearing covariances via $\mathbf{Q}_{\mathbf{ss}}$. This notation is common practice in linearized geostatistical inversion. Note that the appearance of $\mathbf{Q}_{\mathbf{ss}}$ in the above equations is only necessary in derivations. In practice, the sampling/injection operation will be applied to vectors, whenever a vector occurs next to $\mathbf{H}$ in equations. When $\mathbf{H}$ appears next to $\mathbf{Q}_{\mathbf{ss}}$, this requires to simply evaluate a certain column of $\mathbf{Q}_{\mathbf{ss}}$. As explained in the following sections, all columns of $\mathbf{Q}_{\mathbf{ss}}$ are always derived from its first column.

3.1 Evaluating the Estimator

It is well-known that the term $\mathbf{Q}_{\mathbf{sy}}\boldsymbol{\xi}$ in (1) is a superposition of $\boldsymbol{\xi}$ with the covariance function. Within the current notation

$$\mathbf{Q}_{\mathbf{sy}}\boldsymbol{\xi} = \mathbf{Q}_{\mathbf{ss}}\mathbf{H}^T\boldsymbol{\xi} = \mathbf{Q}_{\mathbf{ss}}\underbrace{\left(\mathbf{H}^T\boldsymbol{\xi}\right)}_{\boldsymbol{\Xi}}. \tag{11}$$

If first injecting the $m \times 1$ vector $\boldsymbol{\xi}$ into an $n \times 1$ vector $\boldsymbol{\Xi}$, the remaining matrix–vector product $\mathbf{Q}_{\mathbf{ss}}\boldsymbol{\Xi}$ resembles a convolution of $\boldsymbol{\Xi}$ and $\mathbf{q}_{\mathbf{ss},e}$ (Pegram 2004) and is readily evaluated via FFT in $\mathcal{O}(n \log n)$ operations instead of $\mathcal{O}(mn)$. The computational advantage even for small data sets will be demonstrated later on.

If $\mathbf{Q}_{\mathbf{yy}}$ is itself a Toeplitz matrix because the measurement locations form a regular grid, (8) is not required, but FFT-based techniques can be applied to $\mathbf{Q}_{\mathbf{yy}}$ directly. For irregularly scattered data, however, (8) provides the basis for the ICD algorithm used by Pegram (2004), and for the much faster FFT-based PCG for irregularly spaced data which we will introduce later on.

Assuming for the moment that solving a system with $\mathbf{Q}_{\mathbf{yy}}$ is fast, it still needs to be decoupled from the remaining blocks in (2). First, we introduce two auxiliary variables

$$\mathbf{y} = {\mathbf{Q}_{\mathbf{yy}}}^{-1}\mathbf{Y}, \tag{12}$$

$$\mathbf{z} = {\mathbf{Q}_{\mathbf{yy}}}^{-1}\mathbf{x}. \tag{13}$$

Note that this is identical to the decoupling used by Davis and Grivet (1984) for similar purposes. Then, we partition the inverse of the kriging matrix in (2) as follows (Kitanidis 1996; Nowak and Cirpka 2004),

$$\begin{bmatrix} \mathbf{Q_{yy}} & \mathbf{HX} \\ \mathbf{X}^T\mathbf{H}^T & -\mathbf{Q}_{\beta\beta}^{-1} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{P_{yy}} & \mathbf{P}_{\mathbf{y}\beta} \\ \mathbf{P}_{\beta\mathbf{y}} & \mathbf{P}_{\beta\beta} \end{bmatrix}, \tag{14}$$

express the sub-matrices $\mathbf{P}$ according to Schweppe (1973), and immediately simplify using our auxiliary quantities

$$\mathbf{P}_{\beta\beta} = -\left(\mathbf{x}^T\mathbf{z} + \mathbf{Q}_{\beta\beta}^{-1}\right)^{-1} = -\mathbf{Q}_{\beta\beta|\mathbf{s}}, \tag{15}$$

$$\mathbf{P}_{\beta\mathbf{y}} = \mathbf{P}_{\mathbf{y}\beta}^T = -\mathbf{P}_{\beta\beta}\mathbf{z}^T, \tag{16}$$

$$\mathbf{P_{yy}} = \mathbf{Q_{yy}}^{-1} + \mathbf{z}\mathbf{P}_{\beta\beta}\mathbf{z}^T. \tag{17}$$

This yields a partitioned form of the coefficient vector, similar to the form used by Nowak and Cirpka (2004)

$$\widehat{\boldsymbol{\beta}} = -\mathbf{P}_{\beta\beta}\left(\mathbf{z}^T\mathbf{Y} + \mathbf{Q}_{\beta\beta}^{-1}\boldsymbol{\beta}*\right), \tag{18}$$

$$\boldsymbol{\xi} = \mathbf{y} - \mathbf{z}\widehat{\boldsymbol{\beta}}. \tag{19}$$

This formulation decouples the stationary problem (19) from the previously coupled regression task (18).

In this decoupled form, the entire estimator requires to

1. Compute $\mathbf{y}$ and $\mathbf{z}$ ((12) and (13)) using the most appropriate solver.
2. Evaluate the partitioned solution vector (18).
3. Evaluate the estimate (1), using superposition via FFT (11).

Step 1 requires $1 + p$ solutions of a system with coefficient matrix $\mathbf{Q_{yy}}$. Computational costs will be as low as $\mathcal{O}(m \log m)$ or $\mathcal{O}(n \log n)$, depending on whether the measurements lie on a regular grid or not. Step 2 involves only smaller operations of $\mathcal{O}(mp)$ and $\mathcal{O}(p^3)$ to treat the rank-p perturbations in the structure of the kriging matrix. Step 3 requires one superposition via FFT and a simple $n \times p$ product, with computational costs as low as $\mathcal{O}(n \log n + np)$.

3.2 Evaluating the Estimation Variance

An $n \times n$ dyadic matrix with rank m is defined by the matrix product $\mathbf{A}\mathbf{B}^T$, where $\mathbf{A}$ and $\mathbf{B}$ are $n \times m$ matrices. Its diagonal can be efficiently evaluated using

$$\operatorname{diag}\left(\mathbf{A}_{(n\times m)}\mathbf{B}_{(n\times m)}^T\right) = \sum_{i=1}^{m}(\mathbf{a}_i \circ \mathbf{b}_i), \tag{20}$$

where $\mathbf{a}_i$ and $\mathbf{b}_i$ are the ith columns of $\mathbf{A}$ and $\mathbf{B}$, respectively, and $[] \circ []$ denotes the element-wise vector product. Since (4) contains such a dyadic product, our suggestion

to efficiently evaluate the estimation variance $\hat{\boldsymbol{\sigma}}^2$ is

$$\hat{\boldsymbol{\sigma}}^2 = \operatorname{diag}(\mathbf{Q}_{\mathbf{ss}|\mathbf{y}}) = \boldsymbol{\sigma}^2 - \sum_{i=1}^{m+p} \mathbf{s}_i \circ [\mathbf{Q}_{\mathbf{sy}} \quad \mathbf{X}]_i \tag{21}$$

with the following steps:

1. Initialize $\hat{\boldsymbol{\sigma}}^2 = \boldsymbol{\sigma}^2$, where $\boldsymbol{\sigma}^2 = \sigma^2\mathbf{I}$ and σ^2 is the field variance of $\mathbf{s}$ that populates the diagonal of $\mathbf{Q}_{\mathbf{ss}}$.
2. Evaluate the $(m + p)$ unit estimators $\mathbf{s}_i$ by using the $(m + p)$ unit vectors $\mathbf{e}_i$ as data vector for kriging.
3. Perform the Hadamard product of each unit estimator $\mathbf{s}_i$ with the ith column of $[\mathbf{Q}_{\mathbf{sy}}, \mathbf{X}]$ and subtract from $\hat{\boldsymbol{\sigma}}^2$.

In total, this constitutes an effort of $(m + p)$ kriging estimators, $(m + p)$ Hadamard products and summation of $(m + p)$ vectors, each sized $n \times 1$, resulting in asymptotic complexity of $\mathcal{O}(mn \log_2 n + m^2 \log_2 m)$ or $\mathcal{O}(2mn \log_2 n)$ for $p \ll m$, again depending on whether the measurements lie on a regular grid or not.

4 A Review and Extension of FFT-based Algorithms

In this section, we summarize an existing collection of FFT-based methods to perform matrix operations for Toeplitz matrices at speed, and extend it by an FFT-based solver for irregular-grid data. For finely resolved kriging problems in larger domains, the unknowns $\mathbf{s}$ are typically discretized on a regular and equispaced grid. Equispaced discretization and second order stationarity are sufficient to make $\mathbf{Q}_{\mathbf{ss}}$ a symmetric Toeplitz matrix. In the d-dimensional case, the structure is called symmetric level-d block-Toeplitz (Golub and van Loan 1996). Within Toeplitz matrices, the entries along each diagonal have the same value. For level-d block-Toeplitz matrices, the same pattern applies to nested block structures. In the following, symmetric level-d block-structure is always included implicitly.

Regardless of the dimensionality d, the first column $\mathbf{q}_{\mathbf{ss}}$ of the Toeplitz matrix $\mathbf{Q}_{\mathbf{ss}}$ contains all values that ever appear, reducing storage from n^2 to n elements (Zimmerman 1989). Many efficient algorithms have been found that work on $\mathbf{q}_{\mathbf{ss}}$ only (Kailath and Sayed 1995). An important property is that a Toeplitz-vector product is a discrete convolution of the vector with the first row (or column in the symmetric case) of the matrix. Any Toeplitz matrix $\mathbf{Q}_{\mathbf{ss}}$ can be embedded in a larger circulant matrix $\mathbf{Q}_{\mathbf{ss},e}$, again with extension to level-d block structures. Circulant matrices are Toeplitz matrices with a periodic sequence forming the first row $\mathbf{q}_{\mathbf{ss},e}$. They arise, for example, as covariance matrices in periodic domains. The diagonalization theorem states that their eigenvalues are the Fourier transform of $\mathbf{q}_{\mathbf{ss},e}$ (Varga 1954; Barnett 1990, pp. 350–354), paving the way for a plethora of powerful FFT-based methods. These methods become applicable to Toeplitz matrices when first embedding $\mathbf{q}_{\mathbf{ss}}$ in a larger vector $\mathbf{q}_{\mathbf{ss},e}$ for the circulant counterpart, performing the required operations with the FFT of $\mathbf{q}_{\mathbf{ss},e}$, and then extracting (i.e., the reverse of embedding) the result.

4.1 Embedding and Extraction

The easiest embedding technique is to append the second through the last but one elements of $\mathbf{q}_{\mathbf{ss}}$ to the end of $\mathbf{q}_{\mathbf{ss}}$ in reverse order. For the d-dimensional case, this procedure has to be performed on the entire nested block structure. Intuitively spoken, complement a finite-domain covariance function to a larger periodic-domain covariance function with perfect correlation over the period length, as illustrated by Kozintsev (1999) and by Nowak et al. (2003). Specific applications may allow or require smaller or larger embedding sizes, with details provided by Newsam and Dietrich (1994), Dietrich and Newsam (1997) and Nowak et al. (2003). In brief, FFT-based matrix decomposition requires positive definite circulants, FFT-based PCG solvers require non-negativity, and convolution via FFT is flexible.

Cirpka and Nowak (2004) use a matrix notation for embedding and extraction, based on the $n_e \times n$ mapping matrix $\mathbf{M}$

$$\mathbf{M} = \begin{bmatrix} \mathbf{I}_{n\times n} \\ \mathbf{0}_{(n_e-n)\times n} \end{bmatrix}. \tag{22}$$

Embedding an $n \times 1$ vector $\mathbf{x}$ is denoted by $\mathbf{x}_e = \mathbf{M}\mathbf{x}$, and extracting from an $n_e \times 1$ vector $\mathbf{x}_e$ is denoted by $\mathbf{x} = \mathbf{M}^T\mathbf{x}_e$. Extracting a Toeplitz matrix from the embedding circulant is denoted by

$$\mathbf{Q}_{\mathbf{ss}} = \mathbf{M}^T\mathbf{Q}_{\mathbf{ss},e}\mathbf{M}. \tag{23}$$

In practice, these operations are achieved via zero-padding or disregarding excess elements in adjacent vectors, and multiplication with $\mathbf{Q}_{\mathbf{ss},e}$ is a simple convolution that only requires the first row or column $\mathbf{q}_{\mathbf{ss},e}$ to be stored. Throughout the remaining paper, the subscript e denotes embedded vectors and matrices.

4.2 Convolution and Superposition via FFT

Consider $\mathbf{a}$ an arbitrary $n \times 1$ vector, $\mathbf{Q}_{\mathbf{ss}}$ a $n \times n$ Toeplitz matrix, $\mathbf{q}_{\mathbf{ss},e}$ the first column of the embedding circulant matrix, and let $\mathcal{F}[\cdot]$ and $\mathcal{F}^{-1}[\cdot]$ denote the Fourier transform and its inverse, respectively. Then, according to van Loan (1992), in the current notation

$$\mathbf{Q}_{\mathbf{ss}}\mathbf{a} = \mathbf{M}^T\mathbf{Q}_{\mathbf{ss},e}\mathbf{M}\mathbf{a} = \mathbf{M}^T\mathcal{F}^{-1}\big[\mathcal{F}[\mathbf{q}_{\mathbf{ss},e}] \circ \mathcal{F}[\mathbf{M}\mathbf{a}]\big]. \tag{24}$$

The Fourier transform is evaluated by the FFT or the FFTW, its extension to arbitrary vector length (Cooley and Tukey 1965; Frigo and Johnson 1998). This reduces computational complexity from $\mathcal{O}(n^2)$ to $\mathcal{O}(n \log n)$, and storage requirements from $\mathcal{O}(n^2)$ to $\mathcal{O}(n)$.

Equation (11) traces superposition back to convolution. Combining with (24)

$$\mathbf{Q}_{\mathbf{sy}}\boldsymbol{\xi} = \mathbf{M}^T\mathcal{F}^{-1}\big[\mathcal{F}[\mathbf{q}_{\mathbf{ss},e}] \circ \mathcal{F}\big[\big(\mathbf{M}\mathbf{H}^T\big)\boldsymbol{\xi}\big]\big]. \tag{25}$$

Here, injection ($\mathbf{H}^T$)and embedding ($\mathbf{M}$) form one joint operation ($\mathbf{M}\mathbf{H}^T$). Storage requirements are $\mathcal{O}(n)$ instead of $\mathcal{O}(nm)$. At a computational complexity of

$\mathcal{O}(n \log n)$, this scheme is obviously faster than direct superposition for $m > \log n$. Later performance analysis will show a break-even point at very small data sets.

4.3 FFT-based PCG Solver for Toeplitz Systems

Equations (2) and (3) require to solve systems with the $m \times m$ matrix $\mathbf{Q_{yy}}$. For regular-grid data, $\mathbf{Q_{yy}}$ is Toeplitz just like $\mathbf{Q_{ss}}$, and we suggest the FFT-based Preconditioned Conjugate Gradient (FFT-PCG) solver described below. Other Toeplitz solvers, less compatible with the current FFT-based framework but almost equal in speed, include look-ahead Schur algorithms and algorithms based on generalized displacement structures (Gallivan et al. 1996; Kailath and Sayed 1999; Van Barel et al. 2001). The preconditioned conjugate gradient (PCG) method (Shewchuk 1994) is an iterative solver for linear systems $\mathbf{Ax} = \mathbf{b}$. Its convergence depends on the number of distinct eigenvalues of $\mathbf{A}$. Given a preconditioner $\mathbf{V}$ that clusters the eigenvalues of $\mathbf{V^{-1}A}$ around unity, the PCG method converges in only a few steps. In the FFT-PCG algorithm, $\mathbf{A} = \mathbf{Q_{yy}}$ is Toeplitz and $\mathbf{V}$ is circulant (Chan and Ng 1996). By virtue of the diagonalization theorem, applying the circulant preconditioner $\mathbf{V}$ is a deconvolution via FFT (Good 1950; Rino 1970). Since all matrix operations are accomplished via FFT, the FFT-PCG takes $\mathcal{O}(n \log_2 n)$ overall operations while storing $\mathcal{O}(2n)$ matrix elements. Chan and Ng (1996) compare different circulant preconditioners. We suggest and use an enlarged version of the preconditioner by Strang (1986) that uses $\mathbf{V} = \mathbf{Q_{yy},}_e$ already used for the superposition task, cutting storage to $\mathcal{O}(n)$. The resulting algorithm is provided in the Appendix as special case.

4.4 Regularization of Poorly Conditioned Kriging Systems

The poor condition number of covariance matrices sampled on fine grids has been investigated in depth by Ababou et al. (1994). Poor conditions lead to numerical artifacts in the kriging estimate. Dietrich and Newsam (1989) argue that the measurement error matrix $\mathbf{R}$ regularizes $\mathbf{Q_{yy}}$ by amplifying the diagonal, but that it will induce a loss of information if merely added for regularization. Their idea is essentially identical to Tikhonov regularization or ridge regression. Wesson and Pegram (2004) successfully used a singular value decomposition of $\mathbf{Q_{yy}}$ to suppress numerical artifacts. Similar problems arise in the choice of preconditioners in iterative solvers. For poor-conditioned Toeplitz systems, the preconditioner $\mathbf{V}$ itself is poorly conditioned, leading to divergence or numerical artifacts. Trapp (1973) discussed generalized inverses of circulants which could replace $\mathbf{V}^{-1}$, but we encountered poor convergence with this idea in preliminary studies. Nowak (2005) added a regularization term to the diagonal of $\mathbf{V}$ (Appendix), obtaining reliable and fast convergence. Because only the preconditioner is modified, the original system remains unaffected, and no loss of information occurs. We suggest to use this form of regularization whenever necessary.

4.5 Extended FFT-PCG for Irregular Grids

For irregularly scattered data, $\mathbf{Q_{yy}}$ has no specific structure. For this case, Pegram (2004) solved (2) by Iterative Constrained Deconvolution (ICD). ICD resembles a

Steepest Descent (SD) algorithm with $\boldsymbol{\xi}_{i+1} = \boldsymbol{\xi}_i + \alpha(\mathbf{Y} - \mathbf{Q}_{\mathbf{yy}}\boldsymbol{\xi}_i)$ and a heuristic step size coefficient α. $\mathbf{Q}_{\mathbf{yy}}\boldsymbol{\xi}_i$ is expressed by sampling/injection

$$\mathbf{Q}_{\mathbf{yy}}\boldsymbol{\xi}_i = \mathbf{H}\big[\mathbf{Q}_{\mathbf{ss}}\big(\mathbf{H}^T\boldsymbol{\xi}_i\big)\big] \tag{26}$$

in which $\mathbf{Q}_{\mathbf{ss}}\mathbf{H}^T\boldsymbol{\xi}$ is evaluated by superposition via FFT. Basically, this relates $\mathbf{Q}_{\mathbf{yy}}$ to the Toeplitz matrix $\mathbf{Q}_{\mathbf{ss}}$ via (8), and then links it to the embedding circulant $\mathbf{Q}_{\mathbf{ss},e}$

$$\mathbf{Q}_{\mathbf{yy}} = \big(\mathbf{H}\mathbf{M}^T\big)(\mathbf{Q}_{\mathbf{ss},e} + \mathbf{R}_e)\big(\mathbf{M}\mathbf{H}^T\big), \tag{27}$$

where $\mathbf{R}_e$ is a diagonal matrix of homoscedastic measurement error, fitting in size with $\mathbf{Q}_{\mathbf{ss},e}$. Again, only the first column of $(\mathbf{Q}_{\mathbf{ss},e} + \mathbf{R}_e)$ needs to be stored and processed.

We seized this idea and placed it within the FFT-PCG framework. This leads to a new FFT-PCG solver for irregularly spaced data (Appendix). Later performance analysis will prove it to be much more efficient than ICD. In contrast to the spectral technique by Fuentes (2007) to approximate the likelihood of irregularly spaced data sets, our extended FFT-PCG solver is exact.

5 Efficient Approximations to the Estimation Variance

Equation (4) demonstrated that the exact estimation variance requires $(m + p)$ evaluations of the Kriging estimator, making it a strong limitation on the side of computational costs. In this section, we provide several asymptotically exact approximations to alleviate this restriction. If none of these approximation seems adequate, the estimation variance may be evaluated from conditional simulation, as will be discussed in a later section. For the following analysis, we again decouple the stationary random part and the regression in (3) by inserting (12) to (17), then apply (20) and obtain after some rearrangement

$$\begin{aligned}\hat{\sigma}^2 = \sigma^2 &- \sum_{i=1}^{m}\big[\mathbf{Q}_{\mathbf{sy}}\mathbf{Q}_{\mathbf{yy}}{}^{-1}\big]_i \circ \mathbf{Q}_{\mathbf{sy},i} - \sum_{i=1}^{p}[\mathbf{Q}_{\mathbf{sy}}\mathbf{z}\mathbf{P}_{\boldsymbol{\beta}\boldsymbol{\beta}}]_i \circ [\mathbf{Q}_{\mathbf{sy}}\mathbf{z}]_i \\ &+ 2\sum_{i=1}^{p}[\mathbf{Q}_{\mathbf{sy}}\mathbf{z}\mathbf{P}_{\boldsymbol{\beta}\boldsymbol{\beta}}]_i \circ \mathbf{X}_i - \sum_{i=1}^{p}[\mathbf{X}\mathbf{P}_{\boldsymbol{\beta}\boldsymbol{\beta}}]_i \circ \mathbf{X}_i .\end{aligned} \tag{28}$$

The first term originates from kriging with known mean, and involves m simple kriging unit estimators similar to those defined in (4). Hence, it has well-known properties. The following approximations will focus on the asymptotic behavior of this first term under specific conditions. All other terms relate to the uncertainty in estimating the drift coefficients and have negligible computational complexity over that of the first term. Therefore, we will not simplify them any further. Evaluating the $n \times p$ matrix $\boldsymbol{\zeta} = \mathbf{Q}_{\mathbf{sy}}\mathbf{z}$ requires p superpositions via FFT of $\mathcal{O}(n \log_2 n)$ each. The remaining steps are $\mathcal{O}(np^2)$ to evaluate $\boldsymbol{\zeta}\mathbf{P}_{\boldsymbol{\beta}\boldsymbol{\beta}}$ and $3p$ Hadamard products.

5.1 Single-Point Approximation

Pegram (2004) suggested a single-point approximation to the estimation variance, restricted to zero measurement error and to either simple or ordinary kriging. Correlation between measurements is neglected to reduce the computational complexity by one order in m. In the current study, the context with (28) allows transfer to universal kriging, and we allow for non-zero $\mathbf{R} = \sigma_{\text{err}}^2\mathbf{I}$:

1. For vanishing correlation among the measurements, $\mathbf{Q_{yy}}$ approaches a scalar matrix $\mathbf{Q_{yy}} \approx (\sigma^2 + \sigma_{\text{err}}^2)\mathbf{I}$. Then

$$\sum_{i=1}^{m}\left[\mathbf{Q_{sy}}\mathbf{Q_{yy}}^{-1}\right]_i \circ \mathbf{Q_{sy},}_i \approx \frac{1}{\sigma^2 + \sigma_{\text{err}}^2}\sum_{i=1}^{m}\mathbf{Q_{sy},}_i \circ \mathbf{Q_{sy},}_i . \tag{29}$$

2. After decoupling the regression terms, the estimation variance of simple kriging is supposed to equal σ_{err}^2 at the measurement locations for vanishing mutual correlation. If the correlation among measurements is too large for the above simplification, this condition may still be enforced by requiring the first term to equal $\sigma^2 - \sigma_{\text{err}}^2$ at the locations of measurements. This is achieved by solving the subsidiary kriging problem

$$\sum_{i=1}^{m}\left[\mathbf{Q_{sy}}\mathbf{Q_{yy}}^{-1}\right]_i \circ \mathbf{Q_{sy},}_i \approx \mathbf{Q}_{\mathbf{sy}}^{*}(\mathbf{Q}_{\mathbf{yy}}^{*})^{-1}\mathbf{u}_m\left(\sigma^2 - \sigma_{\text{err}}^2\right), \tag{30}$$

   where $\mathbf{u}_m$ is a $m \times 1$ vector of ones used as data vector in the subsidiary problem, $\mathbf{Q_{ss}}^{*} = \sigma^{-2}\mathbf{Q_{ss}} \circ \mathbf{Q_{ss}}$, $\mathbf{Q}_{\mathbf{sy}}^{*} = \mathbf{Q_{ss}}^{*}\mathbf{H}^T$, and $\mathbf{Q_{yy}}^{*} = \mathbf{H}\mathbf{Q_{ss}}^{*}\mathbf{H}^T$.

5.2 Infinite Regular Grid Approximation

If the grid of measurements is regular and equispaced and its extent is much larger than the range of correlation, the number of mutually correlated measurements is mostly constant, and only differs within a thin boundary area. The inner section of the measurement behaves statistically stationary. One may neglect these boundary effects and apply the correlation pattern of one single representative measurement point in the center of the domain to all others by shifting

$$\mathbf{Q_{sy}}\mathbf{Q_{yy}}^{-1} \approx \text{shift}_m[\mathbf{Q_{sy}}\boldsymbol{\xi}_r]. \tag{31}$$

The term inside the square brackets is a $n \times 1$ zero-mean unit estimator $\mathbf{s}_r$ evaluated for a unit measurement vector $\mathbf{e}_r$ as defined below (4), with $\boldsymbol{\xi}_r$ being the known-mean reciprocal data for $\mathbf{e}_r$. The unit data vector $\mathbf{e}_r = \mathbf{e}_i$ is chosen such that the measurement location $\mathbf{x}_i$ is a representative one in the inner section of the domain. $\text{Shift}_m[\cdot]$ denotes an $n \times m$ matrix obtained from concatenating this representative unit estimator (RUE) $\mathbf{s}_i$ for m times to obtain a $n \times m$ matrix from individual $n \times 1$ columns, each time shifted to the respective measurement location $\mathbf{x}_i$, $i = 1, \ldots, m$. The shifting operation is the same one would perform on an integral kernel in superposition. The RUE needs to be larger than the actual domain so that it covers the entire domain

wherever it is shifted to. The enlarged $\mathbf{s}_r$ is obtained by evaluating $\mathbf{Q}_{\mathbf{sy}}\boldsymbol{\xi}_c$ via (25) and omitting the final extraction step denoted by $\mathbf{M}$. Since measurements in the interior convey less independent information than those at the boundaries, this approximation yields a conservative upper bound of the estimation variance. It is exact for the case of an infinite (or periodic) grid of measurements. The remaining complexity is again reduced by one order in m.

5.3 Hybrid Regular Grid Approximation

In the above approximation, the largest errors occur along the boundaries of the measurement grid. The corresponding unit estimators $\mathbf{s}_i$ may be evaluated separately in their exact form, whereas the bulk inner part of the unit estimators are approximated by shifting the RUE. A user-defined break criterion for the observed difference between RUE and exact unit estimator can be used to determine the affected boundary zone.

5.4 Infinite Regular Grid Average

Another simplifying option is to consider the spatial average of the estimation variance. This is acceptable for relatively fine and regular measurement grids. Although a seemingly crude approximation, its high relevance lies in equivalence to the A-criterion of optimal design (Pukelsheim 2006; Müller 2007). It is achieved by averaging the diagonal of $\mathbf{Q}_{\mathbf{ss}|\mathbf{y}}$. A related semi-analytical solution for the averaged conditional covariance of ordinary kriging has been developed by Cirpka and Nowak (2003). We will translate the formalism into a current, more general context. Due to averaging, the diagonals $\mathrm{diag}(\cdot)$ in (20) and (21) become traces $\mathrm{Tr}(\cdot)$ divided by n. Traces are invariant with respect to cyclic permutations, so that the averaged equivalent of (28) is

$$\hat{\sigma}^2 = \sigma^2 - \frac{1}{n}\mathrm{Tr}\big(\mathbf{Q}_{\mathbf{yy}}{}^{-1}\mathbf{Q}_{\mathbf{ys}}\mathbf{Q}_{\mathbf{sy}}\big) - \frac{1}{n}\mathrm{Tr}\big(\mathbf{P}_{\boldsymbol{\beta}\boldsymbol{\beta}}\mathbf{z}^T\mathbf{Q}_{\mathbf{ys}}\mathbf{Q}_{\mathbf{sy}}\mathbf{z}\big) + \frac{2}{n}\mathrm{Tr}\big(\mathbf{P}_{\boldsymbol{\beta}\boldsymbol{\beta}}\mathbf{X}^T\mathbf{Q}_{\mathbf{sy}}\mathbf{z}\big) - \frac{1}{n}\mathrm{Tr}\big(\mathbf{P}_{\boldsymbol{\beta}\boldsymbol{\beta}}\mathbf{X}^T\mathbf{X}\big). \tag{32}$$

In this form, all traces but the first have boiled down to traces of $p \times p$ matrices. If we again evaluate $\boldsymbol{\zeta} = \mathbf{Q}_{\mathbf{sy}}\mathbf{z}$ via FFT, these small matrices are obtained from $3p$ scalar products and are computationally cheap. For small numbers of measurements, the first trace is best evaluated as

$$\frac{1}{n}\mathrm{Tr}\big(\mathbf{Q}_{\mathbf{yy}}{}^{-1}\mathbf{Q}_{\mathbf{ys}}\mathbf{Q}_{\mathbf{sy}}\big) = \frac{1}{n}\sum_{i=1}^{m}\big[\mathbf{M}\mathbf{H}^T\mathbf{y}_i\big]^T\mathbf{Q}_{\mathbf{ss},e}\mathbf{M}\mathbf{M}^T\mathbf{Q}_{\mathbf{ss},e}\big[\mathbf{M}\mathbf{h}_i^T\big], \tag{33}$$

where $\mathbf{y}_i$ is the solution of $\mathbf{Q}_{\mathbf{yy}}\mathbf{y}_i = \mathbf{e}_i$ and $\mathbf{h}_i$ is the ith row of $\mathbf{H}$, the square brackets indicate the optimal sequence of injections and embeddings, and a total of two convolutions with intermediate deletion of excessive elements ($\mathbf{M}\mathbf{M}^T$) is necessary for each i. This procedure may quickly explode in computational effort for larger m. If the measurement grid is sufficiently large or even periodic, we may approximate the

terms for all i by one representative one with a unit measurement $\mathbf{e}_i$ in the center of the domain and multiply the result by m. This is equivalent to approximation of all matrices as circulants, where trace of the resulting circulant is quite trivially given by m times the first element (Davis 1979).

6 Performance Tests

In this section, we demonstrate the performance of the described FFT-based methods in comparison to standard methods. Our performance analysis was carried out on a contemporary desktop computer (i386, 2.8 GHz Intel Xeon dual-core, 2 GB RAM, Suse Linux 9.2) purchased in 2004. All methods were implemented in MATLAB (Release 2006b). MATLAB includes a standard implementation of the FFTW (Frigo and Johnson 1998) and the Basic Linear Algebra Subprograms (BLAS). As all significant operations are performed by these libraries, the performance compares to an implementation in C or C++. For solving dense systems conventionally, we use the MATLAB built-in (C/C++) Gaussian elimination. In order to keep the FFTW algorithm efficient for arbitrary domain sizes, we implemented a small algorithm that chooses embedding sizes with prime factors of 2, 3, 5, and 7 only. The required relative error norm for all iterative solvers is set to 10^{-10}.

In all performance tests, we used random measurement data and varied both the number n of estimation points and the number m of measurements. We assumed an uncertain constant mean value, so that the number p of trend coefficients is one and $\mathbf{X}$ is a $n \times 1$ vector of ones. The individual test series are composed of $n = 2^k$, $k = 2, \ldots, 24$ estimation points and ratios between n and the number m of estimation points given by $n/m = 2^\ell$, $\ell = 2, \ldots, 14$. All computations above $10^5 s$ (approx. 1 day) were stopped and their CPU times estimated by extrapolation of fitted complexity models.

6.1 Individual Operations

Before demonstrating the overall performance of kriging evaluated with FFT-based algorithms, we provide a summary of performance tests for the individual FFT-based algorithms that have been published elsewhere, e.g., in Nowak et al. (2003) and Nowak (2005), plus tests for our extended FFT-PCG solver for irregular-grid data. Convolution via FFT ((24), solid line in Fig. 1) is faster than conventional evaluation of $\mathbf{Q_{ss}x}$ (bold dash-dotted line) by three orders of magnitude over large ranges of problem size n. For conventional evaluation, n^2 elements need to be stored, leading to very early memory overflow (X-marks). A reduction in storage to $\mathcal{O}(n)$ may also be achieved when only storing the integral kernel represented by $\mathbf{q_{ss},}_e$ which is then shifted and successively added n times. This technique remedies the memory problem, but remains at a computational complexity of $\mathcal{O}(n^2)$, represented by the straight-line extension of the standard method.

Superposition via FFT is compared to conventional superposition for different values of n/m (also Fig. 1). In spite of its overhead and independence of m, the FFT-based algorithm is faster than the standard method in the entire explored range. The

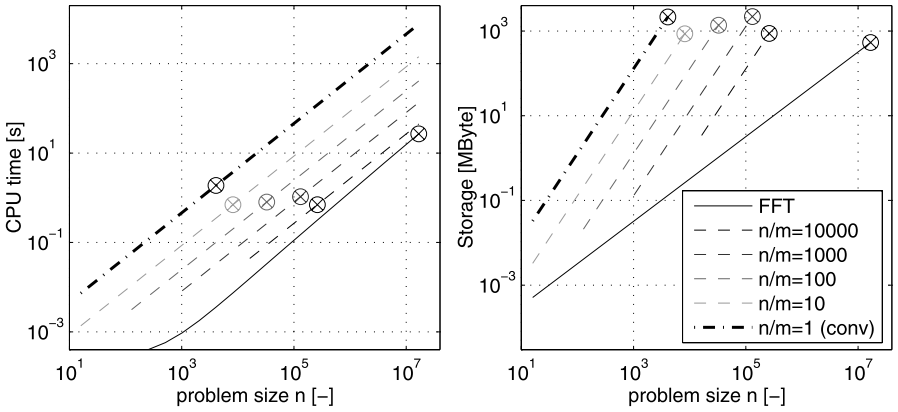

Fig. 1 Convolution/Superposition: CPU time (*left*) and storage requirements (*right*) for conventional discrete superposition versus superposition via FFT for different problem sizes n and numbers of superimposed terms m. Lines: fitted complexity models. *X-Circles*: memory overflow. *Solid line*: superposition/convolution via FFT (same for both tasks). *Dashed lines*: superposition via brute-force matrix product at different ratios n/m. *Dash-dotted line*: convolution via matrix product ($n = m$). Lines exceeding point of memory overflow: superposition/convolution via successive addition of shifted kernel function. Lower line ends: minimum of $m = 1$ superimposed term

largest speedup of nearly three orders of magnitude is achieved for high numbers of m relative to n.

For regular grids of measurements, comparing standard Gaussian elimination versus solution of Toeplitz systems by FFT-PCG (Fig. 2) shows a break-even point at $m \approx 300$. Still below $m = 1000$, FFT-PCG is faster by more than an order of magnitude. At $m = 10000$, the standard solver experienced memory overflow, whereas FFT-PCG worked up to $m = 1.6 \times 10^7$ at CPU times below ten minutes.

Exceeding the analyses published elsewhere, we also implemented other iterative solvers that can use FFT-based convolution to handle the Toeplitz matrix (also Fig. 2). The ICD algorithm by Pegram (2004) can outrun steepest descent (e.g., Press et al. 1992) if the empirical α is chosen well. The largest individual speedup is achieved by preconditioning, which requires more effort per iteration step but drastically reduces the number of iterations. The conjugation of gradients increases the effort per step once more, but the net effect of even fewer iteration steps prevails.

Testing our new extension of FFT-PCG to irregular grids for different problem size m and sizes of the finer regular grid n reveals that using the finer underlying grid can introduce a substantial overhead (Fig. 3). Therefore, the standard solver is faster than the FFT-PCG in the majority of cases. This affects the methods if a high spatial accuracy of measurement locations on the finer grid is desired, but is insignificant if used to fill in missing data on a grid which is otherwise regular (e.g., Wesson and Pegram 2004). The overwhelming advantage of our extended FFT-PCG are the reduced storage requirements in solving much larger systems of equations. On our reference computer, the limiting size of the underlying regular grid was $n = 1.6 \times 10^7$. Our extended FFT-PCG solver outperforms the ICD algorithm by a factor of roughly 10.

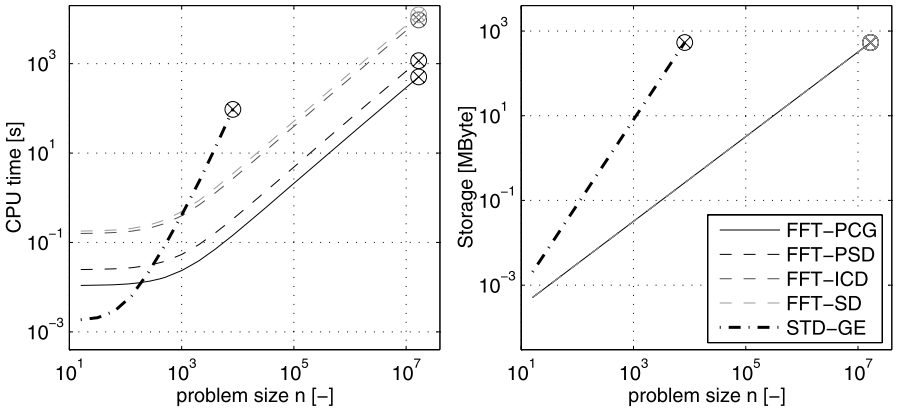

Fig. 2 Solvers for regular measurement grids: CPU time (*left*) and storage requirements (*right*) of different solution techniques for Toeplitz systems for different problem sizes m. Lines: fitted complexity models. *X-Circles*: memory overflow. FFT-PCG: FFT-based PCG solver. FFT-PSD: FFT-based Preconditioned Steepest Descent. FFT-ICD: Iterative Constrained Deconvolution. FFT-SD: FFT-based Steepest Descent. STD-GE: Standard Gaussian Elimination

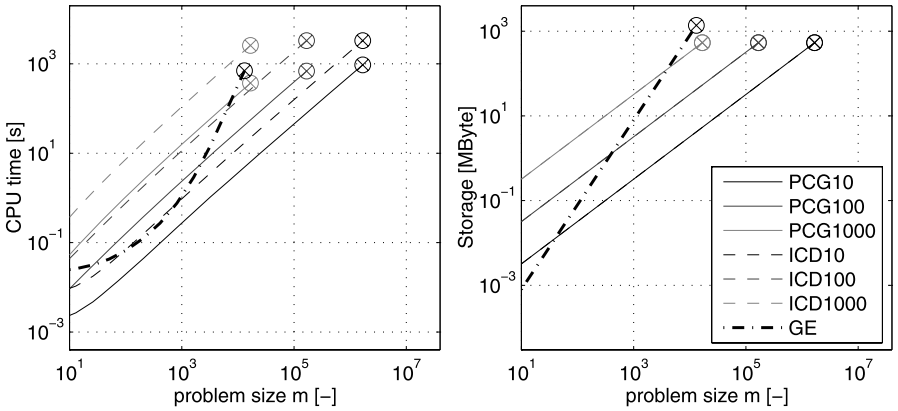

Fig. 3 Extension to irregular measurement grids: CPU time (*left*) and storage requirements (*right*) of different solution techniques for nearly Toeplitz systems for different problem sizes m and different sizes n of the embedding Toeplitz matrix. *Lines*: fitted complexity models. *X-Circles*: memory overflow. PCG#: FFT-based PCG solver with # times larger regular grid. ICD#: FFT-based Iterative Constrained Deconvolution with # times larger regular grid. GE: Standard Gaussian Elimination

6.2 Kriging with Conventional Solver

Our base case for comparison is ordinary kriging with uncertain mean on a regular estimation grid with irregularly scattered measurements (Fig. 4, solid lines). The standard implementation includes Gaussian elimination for solving (2), conventional superposition via successive summation to evaluate $\mathbf{Q}_{\mathbf{sy}}\boldsymbol{\xi}$, and the exact estimation variance. Storage of the kriging matrix (mainly $\mathbf{Q}_{\mathbf{yy}}$) with $\mathcal{O}(m^2)$ is the effective limitation to the admissible problem size (X-marks), so that CPU time never rose above

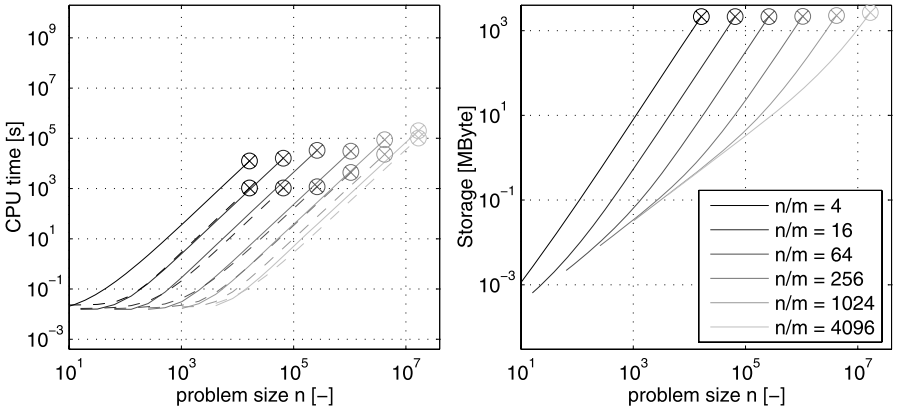

Fig. 4 Performance of Kriging with conventional solver: CPU time (*left*) and storage requirements (*right*) for different numbers n of estimation points and different numbers m of measurements. *Lines*: fitted complexity models. *X-Circles*: memory overflow. *Solid lines*: using conventional solver and conventional superposition. *Dashed lines*: using conventional solver and superposition via FFT

one day. The one-point approximation of the estimation variance (not shown here) would have reduced CPU times by a factor of roughly m. When plugging in superposition via FFT, the asymptotic order of complexity is still dominated by the solution of the Kriging system, but speedup factors of up to 50 occur for large m (small n/m in Fig. 4, dashed lines). Superposition by successive summation, as used in the base case, has storage requirements for $\mathbf{Q_{sy}}$ of only $\mathcal{O}(n)$. Therefore, the storage requirements still coincide with the base case. For the brute-force approach of superposition, with storage requirements of $\mathcal{O}(nm)$, limitations by memory would have been much more severe.

6.3 Kriging with FFT-based Solvers

Using the FFT-PCG solver instead of Gaussian elimination, combined with superposition via FFT radically cancels the restriction by storage requirements related to $\mathcal{O}(m^2)$ (Fig. 5, solid lines). Now, only the grid of estimation limits the admissible problem size to $n = 1.6 \times 10^7$ on our machine. For small m, it is more efficient to revert to conventional Gaussian elimination (Fig. 4, dashed lines). Approximations to the estimation variance can drastically reduce the overall computational effort by an approximate factor of m (Fig. 5, dashed lines for the one-point approximation), with speedups of up to five orders of magnitude for large numbers of measurements (small n/m). The greatest advantage can be made if the measurements are on a regular grid (Fig. 6, solid lines), because the FFT-PCG solver for regular-grid data outruns Gaussian elimination for any problem size and permits vast numbers of measurements. The CPU times are smaller by about one order of magnitude over a large range of problem sizes than for the same number of irregular-grid data. Especially for large problems, the infinite-grid approximation of the estimation variance (Fig. 6, dashed lines) is likely to be sufficiently accurate for many purposes, associated with computational speedup of about 5 orders of magnitude compared to evaluating the exact estimation variance.

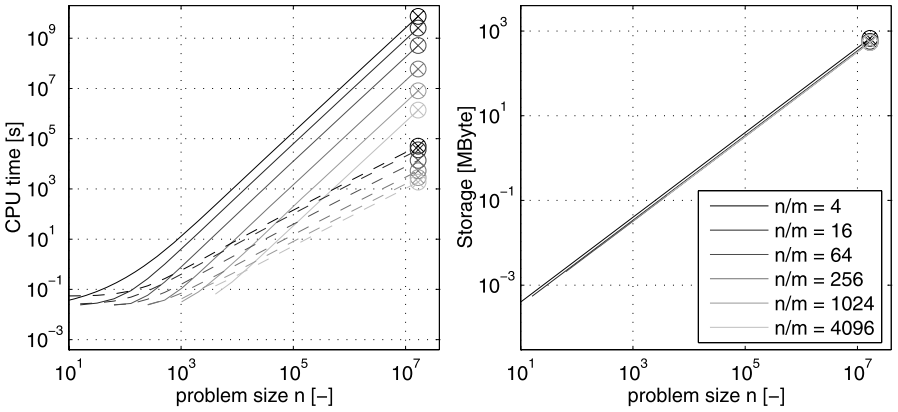

Fig. 5 Performance of Kriging with FFT-based PCG solver for irregularly scattered measurements: CPU time (*left*) and storage requirements (*right*) for different numbers n of estimation points and different numbers m of measurements. *Lines*: fitted complexity models. *X-Circles*: memory overflow. *Solid lines*: exact estimation variance. *Dashed lines*: one-point approximation

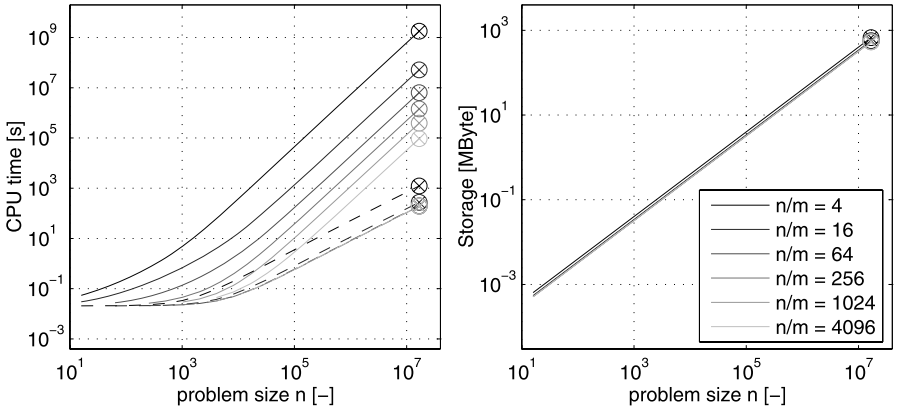

Fig. 6 Performance of Kriging with FFT-based PCG-solver with measurements on regular grids: CPU time (*left*) and storage requirements (*right*) for different numbers n of estimation points and different numbers m of measurements. *Lines*: fitted complexity models. *X-Circles*: memory overflow. *Solid lines*: exact estimation variance. *Dashed lines*: large-grid approximation

7 Extensions to Further Applications

7.1 Generation of Conditional Random Fields

Past studies have presented fast and exact generation of Gaussian random fields via FFT (Dykaar and Kitanidis 1992; Dietrich and Newsam 1993). The overall computational effort per realization is as low as $\mathcal{O}(n \log_2 n)$ at storage requirements of, once again, only $\mathcal{O}(n)$. Conditioning to direct or linearly dependent measurements has been dealt with by Dietrich and Newsam (1996), but the conditioning procedure does not make use of FFT-based methods, yet. The basic well-known step is to correct

unconditional realizations to conditional ones through kriging. Without going into further details, the FFT-based algorithms described in this study can be applied in the corrective kriging step to improve the computational efficiency of generating conditional realizations. The classical path of Monte Carlo simulation is an alternative to evaluate the estimation variance. Once all steps towards conditional realizations are FFT-based, the computational costs are $\mathcal{O}(n \log_2 n + m \log_2 m)$ per conditional realization. Monte Carlo approximation of the variance converges with a rate of $\sqrt{k/2}$, so that a number of realizations in the order of $k = 1000$ is sufficient for most applications. If none of our approximations apply and $m \gg 1000$, Monte Carlo simulation can be considered a helpful alternative.

7.2 Multivariate Cases

If several types of stationary auto- and cross-covariance functions are used to define the matrix $\mathbf{Q_{ss}}$ for multivariate estimation problems, there will be individual blocks that are each accessible to the FFT-techniques demonstrated in this study. Basically, this block structure has to be disassembled to obtain individual additive stationary sub-problems. Then, convolution via FFT is straightforward, and superposition via FFT and the FFT-PCG solver for regular and irregular data grids follow trivially. If the individual univariate cases are intrinsic rather than stationary, each case first has to be decomposed into a stationary part and a coupled regression problem as outlined in (1) to (21). If the multivariate data depend quasi-linearly on $\mathbf{s}$, one can obtain the corresponding auto- and cross-covariance matrices again using FFT-based methods developed by Nowak et al. (2003).

7.3 Sequential Kriging

Sequential Kriging (Vargas-Guzmán and Yeh 1999) considers individual subsets of data one by one, identical to sequential Bayesian updating. Mutual correlation among the subsets is accounted for by always using a covariance matrix $\mathbf{Q_{ss}}$ that is conditional on all previous subsets. For notational convenience and consistency with the original publication, we use simple kriging here. Initially, set $\hat{\mathbf{s}}_0 = \mathbf{0}$, $\mathbf{Q}_{\mathbf{ss},0} = \mathbf{Q_{ss}}$ and $k = 0$. Given a new subset of data $\mathbf{Y}_{k+1}$ with sampling matrix $\mathbf{H}_{k+1}$ and error matrix $\mathbf{R}_{k+1}$, evaluate the matrices

$$\mathbf{Q}_{\mathbf{sy},k+1} = \mathbf{Q}_{\mathbf{ss},k}\mathbf{H}_{k+1}^T = \mathbf{Q}_{\mathbf{ys},k+1}^T, \tag{34}$$

$$\mathbf{Q}_{\mathbf{yy},k+1} = \mathbf{H}_{k+1}\mathbf{Q}_{\mathbf{sy},k+1} + \mathbf{R}_{k+1} \tag{35}$$

and update

$$\hat{\mathbf{s}}_{k+1} = \hat{\mathbf{s}}_k + \mathbf{Q}_{\mathbf{sy},k+1}\mathbf{Q}_{\mathbf{yy},k+1}^{-1}\mathbf{Y}_{k+1}, \tag{36}$$

$$\mathbf{Q}_{\mathbf{ss},k+1} = \mathbf{Q}_{\mathbf{ss},k} - \mathbf{Q}_{\mathbf{sy},k+1}\mathbf{Q}_{\mathbf{yy},k+1}^{-1}\mathbf{Q}_{\mathbf{ys},k+1} \tag{37}$$

$$= \mathbf{Q}_{\mathbf{ss},0} - \sum_{\ell=1}^{k} \mathbf{Q}_{\mathbf{sy},\ell}\mathbf{Q}_{\mathbf{yy},\ell}^{-1}\mathbf{Q}_{\mathbf{ys},\ell}. \tag{38}$$

Vargas-Guzmán and Yeh (1999) suggested to explicitly store $\mathbf{Q}_{\mathbf{ss},k+1}$ at each step, which is impossible for large n. The implicit nature of (34) to (37) leads to a quadratically increasing number of correction terms in (37), of which it is impossible to keep track. Equation (38) only requires to handle the initial stationary covariance matrix and individual, explicitly defined terms which can be accumulated and stored during the sequential procedure. The full-sized matrix never has to be assembled when choosing an adequate order of evaluation for (34)

$$\mathbf{Q}_{\mathbf{sy},k+1} = \mathbf{Q}_{\mathbf{ss},0}\mathbf{H}_{k+1}^T - \sum_{\ell=1}^{k} \mathbf{Q}_{\mathbf{sy},\ell}\mathbf{Q}_{\mathbf{yy},\ell}^{-1}\left(\mathbf{Q}_{\mathbf{ys},\ell}\mathbf{H}_{k+1}^T\right). \tag{39}$$

The latter has already been used on conditional covariance matrices by Nowak and Cirpka (2004) aside from the sequential context. Our suggested modification reduces storage from $\mathcal{O}(n^2)$ to $\mathcal{O}(mn)$. When using convolution via FFT for the first matrix product in (39) and choosing subsets so small that the time to invert each $\mathbf{Q}_{\mathbf{yy},k}$ is insignificant, the overall computation never exceeds $\mathcal{O}(n \log n + mn)$ per updating step. Applications of sequential kriging and cokriging in geostatistical inverse modeling, such as the successive linear estimator (Vargas-Guzmán and Yeh 2002), will greatly benefit from this improvement. Other non-sequential geostatistical inverse modeling techniques, such as the quasi-linear approach by Kitanidis (1995), have been applied for up to one million of unknown parameters (Nowak and Cirpka 2006) on a 2004 desktop computer when using FFT-based methods.

8 Summary and Conclusions

In this study, we were able to evaluate the kriging estimator for millions of estimation points and thousands of measurements in no more than seconds up to a few minutes on a contemporary desktop computer purchased in 2004 (2.8 GHz processor). To achieve this, we compiled a toolbox of existing and extended FFT-based methods that includes FFT-based convolution, FFT-based superposition and FFT-based PCG solvers. All methods apply to estimation on regular and equispaced grids. The measurements may either lie on regular and equispaced grids or be irregularly scattered, if they remain a subset of the finer grid of estimation. Originally, these methods were applicable only to stationary problems such as simple kriging. We decomposed the equations of universal kriging into a stationary part handled by the above FFT-based methods, and a decoupled regression-like part. Thanks to this decoupling, all intrinsic cases and cases with any form of unknown or uncertain drift (arbitrary trend functions, zonation models, external drift or any set of explanatory variables as used in linear regression, with or without prior knowledge on the coefficients) can now be addressed by FFT-based methods. This constitutes a major advantage, since few real-world cases obey stationary models with uniform mean values.

We demonstrated the efficiency of all methods in a series of performance tests on an ordinary desktop computer. The first eye-catching advantage was that now only one column of the auto-covariance matrix of unknowns needs to be stored (resembling a stationary covariance function). This makes the storage requirements shrink

from $\mathcal{O}(nm)$ to only $\mathcal{O}(n)$, where n is the number of estimation points and m is the number of measurements. FFT-based superposition has computational costs in the order of $n \log_2 n$ instead of mn. This turned out to be highly favorable even for very small data sets. For regular and equispaced measurement grids, the FFT-based PCG solver with its complexity of $\mathcal{O}(m \log_2 m)$ outruns standard solvers for dense systems by orders of magnitude. For irregularly scattered measurements, our extended FFT-PCG solver is less efficient than the standard solver for low and medium numbers of measurements. The break-even point is at roughly $m = 1000$ measurements, depending on the resolution of the finer regular grid. The great advantage, however, is that the FFT-based solver does not require to store the $m \times m$ coefficient matrix, so it is applicable for arbitrarily high m. Any solver that requires to explicitly store the dense $m \times m$ kriging matrix broke down due to memory restrictions at a maximum of $m = 10{,}000$ on our reference computer. Combining these methods, we could go as far as 16 million estimation points and the same numbers of measurements on our 2.8 GHz, 4 GB RAM desktop computer. The maximum problem size was solved in less than a day.

Evaluating the estimation variance is computationally much more demanding. For the exact estimation variance, an equivalent of m estimation procedures has to be performed. This is strictly inhibiting for larger data sets, even in spite of the speedup achieved by FFT-based methods. We alleviated this situation by reviewing, extending and proposing several fast approximations which are asymptotically exact in specific cases. These cases include negligible correlation among the measurements and very large regular grids of measurements. The approximations offer an additional speedup of up to five orders of magnitude for large m, so that the largest admissible problem including the estimation variance could still be completed in less than one day. For regular measurement grids, the same tasks were completed within a few minutes. We also pointed out how to transfer FFT-based methods to related applications that will greatly benefit of the results of this study. This includes the generation of conditional realizations, multivariate estimation problems, and sequential kriging.

We expect further advances in non-uniform FFT algorithms (NUFFT), also called the generalized FFT (Duijndam and Schonewille 1999; Fessler and Sutton 2003; Fourmont 2003; Greengard and Lee 2004; Liu and Ngyuen 1998). The kriging context requires NUFFT algorithms that work efficiently in both transform directions, featuring frequency-space data on a regular grid with real-space data on an irregular grid. We assume that highly efficient algorithms for these requirements will be readily available within a few years, further increasing the efficiency for irregularly scattered measurements. Further improvement of storage efficiency may be achievable with low-rank representations of covariance matrices, or by hierarchical approximation (e.g., Börm et al. 2003).

Acknowledgements We are indebted to and thank Prof. Pegram at the University of KwaZulu-Natal, South Africa, for his valuable comments on the manuscript, and to the anonymous reviewers for their helpful suggestions. This study was funded in part by the German Research Foundation (DFG) under the grants No. 824/2-2, No. 824/3 and No. 805/1-1 and by the international research training group NUPUS, financed by the German Research Foundation (DFG) (GRK 1398) and the Netherlands Organisation for Scientific Research (NWO) (DN 81-754).

Appendix: FFT-based PCG for Irregular and Regular Grids

The Conjugate Gradients Method is attributed to Hestenes and Stiefel (1952). The following is the preconditioned version taken from Shewchuk (1994), combined with convolution via FFT (e.g., van Loan 1992) and the regularized circulant preconditioner by Nowak (2005).

The eigenvalues of a circulant matrix are given by the Fourier transform of its first column (e.g., Trapp 1973; Barnett 1990, pp. 350–354). If $\mathbf{C}$ is a (level-d) real circulant matrix with first column $\mathbf{c}$ and $\tilde{\mathbf{c}} = \mathcal{F}_d(\mathbf{c})$ is the (d-dimensional) Fourier transform of the first column, then the condition c of $\mathbf{C}$ is the ratio of the largest value $c_{\max}$ and the smallest value $c_{\min}$ in $\tilde{\mathbf{c}}$. The regularized preconditioner by Nowak (2005) installs a maximum condition c^*(e.g., 10^5) of the preconditioner through a diagonal regularization $\mathbf{C}^* = \mathbf{C} + \varepsilon^* \mathbf{I}$, where ε is chosen according to

$$\varepsilon* = \frac{c_{\max} - c_{\min} c^*}{c^* - 1}. \tag{40}$$

Automatically, the Fourier transform $\tilde{\mathbf{c}}^* = \mathcal{F}_d(\mathbf{c}^*)$ of the first column of $\mathbf{C}^*$ is given by $\tilde{\mathbf{c}}^* = \tilde{\mathbf{c}} + \varepsilon^*$.

In the following, $\circ$ denotes the element-wise (Hadamard) product and $\div$ denotes element-wise division.

Algorithm 1 (Preconditioned Conjugate Gradients with circulant preconditioning for nearly-Toeplitz system) The linear system $\mathbf{Q_{yy}y} = \mathbf{Y}$ is to be solved for a real symmetric positive-definite $m \times m$ matrix $\mathbf{Q_{yy}} = \mathbf{HQ_{ss}H}^T$, where $\mathbf{H}$ is a sampling matrix as defined in (5) and $\mathbf{Q_{ss}}$ is a (level-d) symmetric positive-definite Toeplitz matrix sized $n \times n$. $\mathbf{Q_{ss}}$ has a symmetric positive-definite embedding circulant matrix $\mathbf{C}$ with $\mathbf{Q_{ss}} = \mathbf{M}^T\mathbf{CM}$, where $\mathbf{M}$ is a mapping matrix as defined in (22). $\mathbf{c}$ is the first column of $\mathbf{C}$ with (d-dimensional) Fourier transform $\tilde{\mathbf{c}} = \mathcal{F}_d(\mathbf{c})$. An initial guess $\mathbf{y}_0$, an error tolerance $\varepsilon < 1$ and a maximum allowable condition c^* are provided. Initialize the algorithm with counter $k = 1$, error vector $\mathbf{r} = \mathbf{Y} - \mathbf{Q_{yy}y}_0$, the preconditioned conjugate gradient $\mathbf{d} \approx \mathbf{Q}_{\mathbf{yy}}^{-1}\mathbf{r}$, the residual $\delta_1 = \mathbf{r}^T\mathbf{d}$, the initial residual $\delta_0 = \delta_1$ and evaluate ε^* according to (40). Then

1. Update the trial solution $\mathbf{x}$ using

$$\begin{aligned} \mathbf{q} &= \mathbf{HM}^T \mathcal{F}_d^{-1}\left[\mathcal{F}_d\left[\mathbf{MH}^T\mathbf{d}\right] \circ \tilde{\mathbf{c}}\right] \quad (= \mathbf{Q_{yy}d}), \\ \alpha &= \frac{\delta_k}{\mathbf{d}^T\mathbf{q}}, \\ \mathbf{y} &= \mathbf{y} + \alpha\mathbf{d}. \end{aligned}$$

2. Update the error vector and residual

$$\begin{aligned} \mathbf{r} &= \mathbf{r} - \alpha\mathbf{q}, \\ \mathbf{s} &= \mathbf{HM}^T \mathcal{F}_d^{-1}\left[\mathcal{F}_d\left[\mathbf{MH}^T\mathbf{r}\right] \div \left(\tilde{\mathbf{c}} + \varepsilon^*\right)\right] \quad \left(\approx \mathbf{Q}_{\mathbf{yy}}^{-1}\mathbf{r}\right), \\ \delta_{k+1} &= \mathbf{r}^T\mathbf{s}. \end{aligned}$$

3. Update the preconditioned conjugate gradient

$$\mathbf{d} = \mathbf{s} + \frac{\delta_{k+1}}{\delta_k}\mathbf{d}.$$

4. Increase k by 1 and repeat until $k > k_{\max}$ or $\delta_{k+1} < \varepsilon^2\delta_0$.

The variables $\mathbf{q}$, α and $\mathbf{s}$ are auxiliaries to reduce the computational costs. All products with $\mathbf{M}$ and $\mathbf{H}$ are evaluated by simple embedding/extraction and injection/sampling or using sparse representations of the matrix $\mathbf{N} = \mathbf{M}\mathbf{H}^T$. The FFT-PCG algorithm requires only one matrix–vector product per iteration step evaluated via FFT, which has an asymptotic cost estimate of $\mathcal{O}(n \log_2 n)$ and another operation of $\mathcal{O}(n \log_2 n)$ to apply the preconditioner. The corresponding steps in the initialization are evaluated accordingly, resulting in an overall complexity of $\mathcal{O}(n \log_2 n)$. Here, the finer regular grid with n nodes is not necessarily as large (as fine) as the grid of estimation, subject to the desired accuracy of discretizing the locations of the measurements.

In case the measurements themselves lie on a regular grid, matrix $\mathbf{Q_{yy}}$ in the above algorithm is a Toeplitz matrix itself, substituted for $\mathbf{Q_{ss}}$, and $\mathbf{H} = \mathbf{H}^T = \mathbf{I}$ may be omitted in the entire algorithm. In that case, the computational complexity drops to $\mathcal{O}(m \log_2 m)$.

References

Ababou R, Bagtzoglou AC, Wood EF (1994) On the condition number of covariance matrices in kriging, estimation, and simulation of random fields. Math Geol 26(1):99–133

Barnett S (1990) Matrices methods and applications. Oxford applied mathematics and computing science series. Clarendon, Oxford

Börm S, Grasedyck L, Hackbusch W (2003) Introduction to hierarchical matrices with applications. Eng Anal Bound Elem 27:405–422. doi:10.1016/S0955–7997(02)00152–2

Chan RH, Ng MK (1996) Conjugate gradient methods for Toeplitz systems. SIAM Rev 38(3):427–482

Cirpka OA, Nowak W (2003) Dispersion on kriged hydraulic conductivity fields. Water Resour Res. doi:10.1029/2001WR000598

Cirpka OA, Nowak W (2004) First-order variance of travel time in non-stationary formations. Water Resour Res 40:W03507. doi:10.1029/2003WR002851

Cooley JW, Tukey JW (1965) An algorithm for the machine calculation of complex Fourier series. Math Comput 19:297–301

Davis PJ (1979) Circulant matrices. Pure and applied mathematics. Wiley, New York

Davis MW, Culhane PG (1984) Contouring very large data sets using kriging. In: Verly G et al (eds) Geostatistics for natural resources characterization, Part 2. Reidel, Dordrecht

Davis MW, Grivet C (1984) Kriging in a global neighborhood. Math Geol 16(3):249–265

Dietrich CR, Newsam GN (1989) A stability analysis of the geostatistical approach to aquifer transmissivity identification. Stoch Hydrol Hydraul 3:293–316

Dietrich CR, Newsam GN (1993) A fast and exact method for multidimensional Gaussian stochastic simulations. Water Resour Res 29(8):2861–2869

Dietrich CR, Newsam GN (1996) A fast and exact method for multidimensional Gaussian stochastic simulations: Extension to realizations conditioned on direct and indirect measurements. Water Resour Res 32(6):1643–1652

Dietrich CR, Newsam GN (1997) Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix. SIAM J Sci Comput 18(4):1088–1107

Duijndam AJW, Schonewille MA (1999) Nonuniform fast Fourier transform. Geophysics 64(2):539–551

Dykaar BB, Kitanidis PK (1992) Determination of the effective hydraulic conductivity for heterogeneous porous media using a numerical spectral approach. 1. Method. Water Resour Res 28(4):1155–1166
Fessler JA, Sutton BP (2003) Nonuniform fast Fourier transform using min-max interpolation. IEEE Trans Signal Process 51(2):560–574
Fourmont K (2003) Non-equispaced fast Fourier transforms with applications to tomography. J Fourier Anal Appl 9(5):431–450
Frigo M, Johnson SG (1998) FFTW: An adaptive software architecture for the FFT. In: Proc ICASSP, vol 3. IEEE Press, New York, pp 1381–1384. http://www.fftw.org
Fuentes M (2007) Approximate likelihood for large irregularly spaced spatial data. J Am Stat Assoc 102(477) 321–331. doi:10.1198/016214506000000852
Gallivan K, Thirumalai S, Dooren PV, Vermaut V (1996) High performance algorithms for Toeplitz and block Toeplitz matrices. Linear Algebra Appl 241–243(13):343–388
Golub GH, van Loan CF (1996) Matrix computations, 3rd edn. John Hopkins University Press, Baltimore
Good IJ (1950) On the inversion of circulant matrices. Biometrika 37:185–186
Greengard L, Lee J-Y (2004) Accelerating the nonuniform fast Fourier transform. SIAM Rev 46(3):443–454
Hestenes MR, Stiefel E (1952) Methods of conjugate gradients for solving linear systems. J Res Nat Bur Stand 49:409–436
Kailath T, Sayed AH (1995) Displacement structure: Theory and applications. SIAM Rev 37(3):297–386
Kailath T, Sayed AH (1999) Fast reliable algorithms of matrices with structure. SIAM, Philadelphia
Kitanidis PK (1986) Parameter uncertainty in estimation of spatial functions: Bayesian analysis. Water Resour Res 22(4):499–507
Kitanidis PK (1993) Generalized covariance functions in estimation. Math Geol 25(5):525–540
Kitanidis PK (1995) Quasi-linear geostatistical theory for inversing. Water Resour Res 31(10):2411–2419
Kitanidis PK (1996) Analytical expressions of conditional mean, covariance, and sample functions in geostatistics. Stoch Hydrol Hydraul 12:279–294
Kitanidis PK (1997) Introduction to geostatistics. Cambridge University Press, Cambridge
Kitanidis PK, Vomvoris EG (1983) A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations. Water Resour Res 19(3):677–690
Kozintsev B (1999) Computations with Gaussian random fields. PhD thesis, Institute for Systems Research, University of Maryland
Liu QH, Ngyuen N (1998) An accurate algorithm for nonuniform fast Fourier transforms (NUFFT's). IEEE Microw Guided Wave Lett 8(1):18–20
Müller WG (2007) Collecting spatial data. Optimum design of experiments for random fields, 3rd edn. Springer, Berlin
Newsam GN, Dietrich CR (1994) Bounds on the size of nonnegative definite circulant embeddings of positive definite Toeplitz matrices. IEEE Trans Inf Theory 40(4):1218–1220
Nowak W (2005) Geostatistical methods for the identification of flow and transport parameters in subsurface flow. Ph.D. thesis, Institut für Wasserbau, Universität Stuttgart, http://elib.uni-stuttgart.de/opus/frontdoor.php?source_opus=2275
Nowak W, Cirpka OA (2004) A modified Levenberg–Marquardt algorithm for quasi-linear geostatistical inversing. Adv Water Resour 27(7):737–750
Nowak W, Cirpka OA (2006) Geostatistical inference of conductivity and dispersion coefficients from hydraulic heads and tracer data. Water Resour Res 42:W08416. doi:10.1029/2005WR004832
Nowak W, Tenkleve S, Cirpka OA (2003) Efficient computation of linearized cross-covariance and auto-covariance matrices of interdependent quantities. Math Geol 35(1):53–66
Omre H (1987) Bayesian kriging-merging observations and qualified guesses in kriging. Math Geol 19(1):25–39
Pegram GGS (2004) Spatial interpolation and mapping of rainfall (SIMAR). Data merging for rainfall map production, vol 3. Water Research Commission Report (1153/1/04)
Press WH, Teukolsky BPFSA, Vetterling WT (1992) Numerical recipes: the art of scientific computing, 2nd edn. Cambridge University Press, Cambridge
Pukelsheim F (2006) Optimal design of experiments. Classics in applied mathematics. SIAM, Philadelphia
Rino CL (1970) The inversion of covariance matrices by finite Fourier transforms. IEEE Trans Inf Theory 16:230–232
Schweppe FC (1973) Uncertain dynamic systems. Prentice-Hall, Englewood Cliffs
Shewchuk JR (1994). An introduction to the conjugate gradient method without the agonizing pain. http://www.cs.berkeley.edu/~jrs/

Strang G (1986) A proposal for Toeplitz matrix calculations. Stud Appl Math 74:171–176
Trapp GE (1973) Inverses of circulant matrices and block circulant matrices. Kyungpook Math J 13(1): 11–20
Van Barel M, Heinig G, Kravanja P (2001) A stabilized superfast solver for nonsymmetric Toeplitz systems. SIAM J Matrix Anal A 23(2):494–510
van Loan CF (1992) Computational frameworks for the fast Fourier transform. SIAM, Philadelphia
Varga RS (1954) Eigenvalues of circulant matrices. Pac J Math 4:151–160
Vargas-Guzmán JA, Yeh T-CJ (1999) Sequential kriging and cokriging: Two powerful geostatistical approaches. Stoch Environ Res Risk Assess 13:416–435
Vargas-Guzmán JA, Yeh T-CJ (2002) The successive linear estimator: a revisit. Adv Water Resour 25:773–781
Wesson SM, Pegram GGS (2004) Radar rainfall image repair techniques. Hydrol Earth Syst Sci 8(2):8220–8234
Zimmerman DL (1989) Computationally exploitable structure of covariance matrices and generalized covariance matrices in spatial models. J Stat Comput Simul 32(1/2):1–15