

PAPER • OPEN ACCESS

## Adaptive sampling in higher dimensions for point-wise experimental measurement techniques

To cite this article: R Theunissen and P Gjelstrup 2018 *Meas. Sci. Technol.* **29** 085007

View the [article online](#) for updates and enhancements.

### You may also like

- [Surface form inspection with contact coordinate measurement: a review](#)  
Yijun Shen, Jieji Ren, Nuodi Huang et al.
- [Efficient data acquisition and training of collisional-radiative model artificial neural network surrogates through adaptive parameter space sampling](#)  
Nathan A Garland, Romit Maulik, Qi Tang et al.
- [Machine learning and excited-state molecular dynamics](#)  
Julia Westermayr and Philipp Marquetand

# Adaptive sampling in higher dimensions for point-wise experimental measurement techniques

R Theunissen<sup>1</sup>  and P Gjelstrup<sup>2</sup>

<sup>1</sup> Department of Aerospace Engineering, University of Bristol, Bristol, BS8 1TR, United Kingdom

<sup>2</sup> Dantec Dynamics, Skovlunde, DK-2740, Denmark

E-mail: [r.theunissen@bristol.ac.uk](mailto:r.theunissen@bristol.ac.uk)

Received 9 March 2018, revised 17 May 2018

Accepted for publication 4 June 2018

Published 5 July 2018



## Abstract

When performing point-wise measurements within a pre-defined domain, the experimentalist is faced with the problem of defining the spatial locations where to collect data based on an *a priori* unknown underlying signal. While structured sampling grids are most common, these are rarely optimal from a time-efficiency perspective. In this work an adaptive process is presented for point-wise measurement techniques to guide the spatial distribution of sampling locations in two-dimensions. Thin plate splines of varying degree of smoothing are utilised to iteratively obtain surrogate models of the spatial distribution of the quantities of interest. Adaptive sampling criteria reflecting typical human decisions are combined into a unique objective function, allowing the derivation of suitable measurement positions. The total number of sampling locations thus no longer requires to be pre-defined, as the sampling process is terminated automatically once the addition of measurements is deemed not to provide additional information. The overall implementation is elaborated and the methodology is assessed on the basis of numerical simulations and an experimental test case. Results support the new method's advocated superiority compared to traditional full-factorial sampling in terms of accuracy and reliability.

Keywords: adaptive sampling, radial basis function, scale-dependent sampling, LDA, design of experiments, thin plate spline

(Some figures may appear in colour only in the online journal)

## 1. Introduction

Whereas full-field image-based measurement techniques such as particle image velocimetry (PIV) are capable of providing velocity data within planar or even volumetric sections of the scrutinised flow (Kähler *et al* 2016), highest spatial resolutions (excluding advanced PIV alternatives e.g. micro-PIV (Meinhart *et al* 1999)) continue to be attainable with more traditional point-wise methodologies e.g. hot-wire anemometry (HWA) or laser Doppler anemometry (LDA) (Lavoie *et al*

2007). Contrary to PIV where data rates are defined by the laser's pulse repetition and camera frame rate (and therefore associated costs), metrologies such as HWA and LDA also offer cheaper alternatives with high frequency responses. This makes such techniques extremely suitable for spectral analyses in turbulent flow research. Although the limited measurement volume of the point-wise techniques enables the detection of smaller flow scales, because measurement volumes are typically hundreds of orders of magnitude smaller compared to the flow region of interest, some form of traversing is required when performing a flow survey. Typical HWA wire diameters for example are in the order of 5  $\mu\text{m}$  compared to flow-related length scales in the order of centimeters (Gao *et al* 2015). Unless flow-related time scales are much larger than the

Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



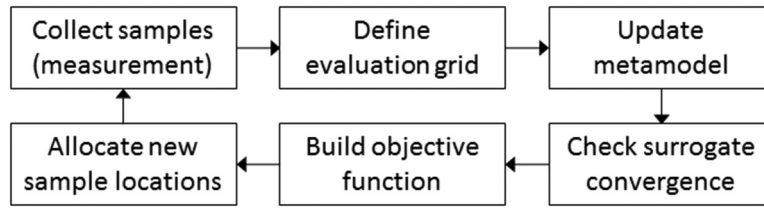


Figure 1. Procedural steps in adaptive sampling.

overall measurement time, full-field data analyses with point-wise measurement techniques are therefore often limited to temporal averages, i.e. time-average flow fields.

Point-wise probes are commonly mounted on traverse mechanisms and are therefore able to be, in principle, positioned continuously in space. Ignoring limitations imposed by the finite probe dimensions, the level of flow detail captured is thus directly defined by the spacing of spatial sampling locations considered through the well-known Nyquist sampling criterion (Petersen and Middleton 1962). This criterion dictates the smallest measurable length scale to be at least twice the minimum sample spacing. The experimentalist is thus forced to compromise between data content and time efficiency: the more samples of the flow, the longer the measurement time, the higher the running costs of the experimental facility yet the more flow detail can be obtained. In turn, when faced with the task of characterising an unknown flow phenomenon, every experimentalist will be ultimately faced with the same conundrum: where to measure and how many samples to take.

Of the various sampling strategies, space-filling is the most common and includes, among others, the distribution of equidistant samples across the measurement domain (full-factorial sampling: FF), Latin Hypercube (LH) sampling, Poisson disc sampling, etc. The number of samples adopted in such methods will always vary as a power of the ratio between domain extent  $L$  and user-defined minimum sample spacing  $h$ . For full factorial and Poisson disk sampling in  $d$ -dimensional space, the number of samples will for example be in the order of  $\mathcal{O}((L/h)^d)$ , whereas, because of the permutations, LH typically results in  $\mathcal{O}(L/h)$  samples. As the user refines the sampling grid, the number of measurement points clearly becomes rapidly impractical. Assuming the data collection spans ten seconds at each measurement location (common in LDA), characterisation of a three-dimensional velocity field with only ten samples along each coordinate axis will require a total run-time of approximately three hours for  $10^3$  samples while yielding a still under-resolved flow field. It is for this reason that more intelligent sampling approaches are needed, capable of focusing, and thereby minimising, the data accrual to flow regions in which a higher number of measurements are beneficial in the reconstruction and subsequently understanding, of the underlying *a priori* unknown flow field.

All space filling methods blindly adopt the identical sampling strategy irrelevant of the scrutinised process (LH does involve the probability density function of each variable in theory, yet for spatial sampling each coordinate has a uniform probability, eliminating any probability dependency). Instead, adaptive processes whereby the sampling strategy is adjusted *in situ* on the basis of collected observations, have the potential

to yield results more efficiently if the stopping criteria are quality-defined, or with higher accuracy, when limiting the permitted number of accrued data points. It is therefore not surprising that adaptive sampling has been widely applied in e.g. surface reconstruction (Chen and Peng 2017), survey studies (Yu *et al* 2012), computer vision (Bolin and Meyer 1998), computer graphics (Chen *et al* 2011), computer aided design (Li 1995), optimisation (Mackman *et al* 2013) and in the prediction of suitable measurement locations for PIV (Theunissen *et al* 2007) or one-dimensional velocity surveys (Theunissen *et al* 2015).

The typical adaptive process outline is depicted in figure 1 and starts with collecting data at initial sampling locations, defining the domain outline in which an evaluation grid is specified. This grid defines the locations at which the surrogate model is evaluated and must subsequently be sufficiently small to capture the model's behaviour. The surrogate model itself is constructed from the available data and needs to be updated each time new data is accrued. Various routes exist to construct surrogate or meta-models such as for example neural networks (Holeňa *et al* 2010), support vector machines (Cicczazzo *et al* 2014), splines (Grimstad *et al* 2015), polynomial regression (Gano *et al* 2006), Kriging or Gaussian process regression (Denimal *et al* 2016) and radial basis functions (RBF) (Durantin *et al* 2017).

Adaptivity is finally based on some form of criterion, the so-called objective, enabling potential samples to be ranked according to their quantified suitability. In numerical grid refinement for example, this criterion can be based on a second derivative error norm of the surrogate model (Li 2010), local field curvature (Mackman and Allen 2010) or a combination of various functionals. When dealing with sampling optimisation, such cost functions are generally based on estimates of the spatial distributions of the scrutinized parameters (Forrester *et al* 2008).

In the work of Theunissen *et al* (2015), which forms the basis of the current study, a procedure was presented whereby sequential sample locations were selected adaptively thus eliminating user-dependency and/or *a priori* knowledge of the investigated flow. Weighted criteria involving local measurement error estimates, space filling heuristics, local surrogate improvement and curvature were combined into a unique objective function such that its peak location indicated the most suitable sampling location, replicating the experimentalist's decision processes. A  $C^2$ -continuous Wendland function, extended with third-order polynomial interpolation, was adopted because of its ability to provide second order derivative estimates with relative computational ease. The autonomous methodology was shown to be

more efficient and able to capture flow features of interest more accurately, although restricted to one-dimensional, single component velocity profile extraction. In the current work such concepts are extended to higher dimensions with a focus on implementation.

The aim of the current work is to devise an autonomous routine capable of guiding the spatial sampling process in a typical experiment in which point-wise measurement techniques are utilized to survey a prescribed domain in terms of scalar quantities such as e.g. pressure, velocity, temperature, etc. At each sampling location the statistical average of multiple readings serves as the quantity of interest in the characterisation of the underlying phenomenon. Given the exponential increase of the number of sampling positions with dimensionality of the problem (and consequently the experiment run time), irrespective of adaptation, considered applications are limited to two-dimensions. The positioning of the data extraction locations within the 2D domain is performed through an adaptive process similar to Theunissen *et al* (2015) but extended into higher dimensions. To this extent the current study starts with a review of the objective function followed by a discussion on the choice of RBF kernel for interpolation. Additional computational procedures are presented to ensure viability and reduce computational effort in each of the different procedural steps. Preceding the conclusions, a numerical assessment as well as an experimental application with varying degree of data uncertainty corroborate the advocated advantages.

It is important to note that the importance of efficient collection of meaningful data spreads beyond the boundaries of experimental fluid mechanics and has a broader application referred to as the design of experiments. As such, the proposed methodology and associated implementation is expected to benefit a wider range of research.

## 2. Surrogate modelling

### 2.1. RBF interpolation

In its simplest form the surrogate model in  $d$ -dimensions consists of an interpolation of available data  $f(\mathbf{x}_t)$  with  $\mathbf{x} = (x_1, \dots, x_d)$  onto the evaluation grid. In view of required adaptivity to local curvature, this interpolant must be at least second order continuous ( $C^2$ ). Because of their simplicity and ease of implementation, polynomial regression and RBF interpolation are the most common methods (Fasshauer 2007). Moreover, RBF functions have been shown to substantially reduce effects of measurement noise when interpolating (Casa and Krueger 2013). Given  $N^{(i)}$  data points ( $t = 1, \dots, N^{(i)}$ ), the value of the surrogate model  $s^{(i)}(\mathbf{x}_s)$  in the  $i$ th iteration at a random location  $\mathbf{x}_s$  in  $d$ -dimensional space is formulated through an augmented RBF model with polynomial summands as

$$s^{(i)}(\mathbf{x}_s) = \sum_{t=1}^{N^{(i)}} \alpha_t \psi(\|\mathbf{x}_t - \mathbf{x}_s\|) + q(\mathbf{x}_s) \quad \text{with} \quad \mathbf{x} \in \mathbb{R}^d \quad (1)$$

where  $\psi(r)$  accounts for the contribution of the  $t$ th basis function in function of the radial distance  $r = \|\mathbf{x}_t - \mathbf{x}_s\|$  between the new location and the  $t$ th data point. It is evident that with every iteration the number of samples  $N^{(i)}$  can change and the metamodel  $s^{(i)}(\mathbf{x})$  needs to be re-evaluated. Without loss of clarity, the superscript ( $i$ ) will be omitted for brevity hereafter.

The polynomial part  $q(\mathbf{x}_s)$  is added to guarantee polynomial accuracy up to degree  $m$  and is given by the summation of  $M = \binom{m+d}{d} = \frac{(m+d)!}{m!d!}$  weighted terms  $p_1, \dots, p_M$ ;  $q(\mathbf{x}) = \sum_{j=1}^M \beta_j p_j(\mathbf{x})$  where a possible choice of polynomial basis functions is given by  $p_1(\mathbf{x}) = 1, p_2(\mathbf{x}) = x_1, \dots, p_{d+1}(\mathbf{x}) = x_d, p_{d+2}(\mathbf{x}) = x_1 x_2, p_{d+3}(\mathbf{x}) = x_1 x_3, \dots, p_{M-d+1}(\mathbf{x}) = x_1^m, \dots, p_M(\mathbf{x}) = x_d^m$ . Scaling coefficients  $\alpha_t, \beta_l \in \mathbb{R}$  are obtained in a straightforward manner by solving equation (1) for the known data points, i.e. the interpolation conditions  $s(\mathbf{x}_t) = f(\mathbf{x}_t)$  with the  $M$  additional constraints  $\sum_{t=1}^N \alpha_t p_j(\mathbf{x}_t) = 0, j = 1, \dots, M$ . This is equivalent to solving the matrix system

$$\begin{bmatrix} \Psi + \lambda I & P \\ P^T & O \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \mathbf{0} \end{bmatrix} \quad (2)$$

where  $\Psi_{ij} = \psi(\|\mathbf{x}_i - \mathbf{x}_j\|)$ ,  $i, j = 1, \dots, N$ ,  $P_{il} = p_l(\mathbf{x}_i)$ ,  $l = 1, \dots, M$ ,  $O$  is a  $M \times M$  zero matrix,  $I$  is the  $N \times N$  identity matrix,  $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_N]^T$ ,  $\boldsymbol{\beta} = [\beta_1, \dots, \beta_M]^T$ ,  $\mathbf{f} = [f(\mathbf{x}_1), \dots, f(\mathbf{x}_N)]^T$  and  $\mathbf{0}$  is a  $1 \times N$  vector. The additional term  $\lambda I$  is referred to as ridge-regression and introduces a degree of smoothing, which can be controlled through  $\lambda$ . Having calculated the coefficients for each data point, these are then substituted into equation (1) to solve for the interpolation at any random location.

Especially the Wendland kernels have received great attention as choices for  $\psi(\mathbf{r})$  because these can be easily extended to higher dimensions and any degree of smoothness (Wendland 2004). More importantly, these kernels introduce sparsity in the matrix systems to solve (see equation (2)) because of their compactness (Schaback and Wendland 1999), thus reducing computational effort. However, RBF functions require a well-defined support radius since enlarging the RBF kernel generally leads to enhanced interpolation accuracy at the expense of denser and ill-conditioned matrix systems (Boyd and Gildersleeve 2011). Too large values of the support cause the RBF interpolant to exhibit spurious overshoots reminiscent of the Gibbs phenomenon for polynomial interpolants. To this extent (Schaback and Wendland 1999) suggested the support radius to be chosen such that each RBF support contains approximately the same number of samples, which in turn depends on the sample density (i.e. the fill distance defined as the largest of the nearest neighbor sample distances). Especially given the expected non-uniformity in sample density, a unique choice of support radius may prove troublesome and much continuing research is devoted to the proper choice of a potentially adaptive scale parameter (Rippa 1999, Scheuerer 2011, Uddin 2017).

Thin Plate Splines (TPS) on the other hand do not require shape parameters (Fasshauer 2007) and are more intuitive as they provide an interpolation minimising the overall ‘bending’

**Table 1.** 2D TPS basis functions of order  $n$ ,  $\psi(r) = r^{2n-2} \log(r)$ , with associated derivatives.

$r = \ \mathbf{x}_k - \mathbf{x}_s\ _2$	Order $n = 2$ (TPS2)	Order $n = 3$ (TPS4)
$\psi(r)$	$r^2 \log(r)$	$r^4 \log(r)$
$\frac{d}{dr} \psi(r)$	$r(2 \log(r) + 1)$	$r^3(4 \log(r) + 1)$
$\frac{d^2}{dr^2} \psi(r)$	$2 \log(r) + 3$	$r^2(12 \log(r) + 7)$
$\Delta \psi(r) = \frac{d^2}{dr^2} \psi(r) + \frac{1}{r} \frac{d}{dr} \psi(r)$	$4 \log(r) + 4$	$8r^2(2 \log(r) + 1)$

energy. They are inhibitive though as they are intrinsically not compact (Li *et al* 2007) and thus involve solving large matrix systems (Wood 2003). However, experimental applications targeted in the current work will concern a number of sampling locations ordinarily in the order of  $10^3$ – $10^4$ , rendering the matrix system (depending on the choice of evaluation grid density) of only low to moderate complexity. Considerable work has also been done in optimising the ridge regression parameter  $\lambda$  (see equation (2)) to accommodate splines for noisy data (Wahba 1990). This will prove to be advantageous as it allows the sampling process to consider recursively refined spatial scales of the underlying function through iterative reduction of  $\lambda$  (see section 2.3). Moreover, the TPS kernel has a well-defined integral, contrary to the Wendland functions, which will be shown to reduce the computational complexity of spatial filtering based on moving averages (see section 3.3). For this reason, in the following TPS radial basis functions of second ( $n = 2$ ) and third order ( $n = 3$ ) (see table 1) will be adopted.

In terms of polynomial order  $m$ , the inclusion of high order polynomials has been reported to worsen the conditioning problem as well as increasing the cost of the matrix solution (Costin and Allen 2013). It is therefore advisable to retain the polynomial order to a minimum. Following (Duchon, 1977) the order of the polynomial will be at most  $m = n - 1$  ( $m = 1$  for TPS2 and  $m = 2$  for TPS4).

### 2.2. Preconditioning

The definition of TPS basis functions in two dimensions ( $d = 2$ ) and associated derivatives are tabulated in table 1. The most common TPS is that of order two (TPS2), despite its second derivative exhibiting an unavoidable singularity at the origin. In view of the requirement to calculate the surrogate model’s curvature (see section 3.1), the order three variant (TPS4) is more suitable. On uniform grids, the TPS4 has a condition number which increases as  $\mathcal{O}(N^7)$  compared to  $\mathcal{O}(N^4)$  for TPS2 (Boyd and Gildersleeve 2011). This translates in ill-conditioned, and subsequently, unstable solutions for the coefficient vectors  $\alpha$  and  $\beta$  in equation (2). To overcome this problematic, interpolation matrices are preconditioned and TPS2 kernels are adopted to construct the surrogate models whereas the TPS4 kernels are applied in the calculation of curvature heuristics. The latter will be elaborated in section 3.1.

The preconditioned equivalent of matrix system  $\mathbf{Ax} = \mathbf{y}$  is given as  $\mathbf{CAx} = \mathbf{Cy}$  such that  $\mathbf{CA}$  is better conditioned. For two dimensional problems the preconditioning strategy proposed by Dyn *et al* (1986) is specifically tailored to TPS2 and has been extended here for TPS4. Contrary to equation (2), the system considered in Dyn did not involve ridge regression. However, the addition of  $\lambda$  along the diagonal elements of  $\Psi$  does not invalidate the presumptions that the iterated Laplacian of the radial basis function (including the TPS kernels) tends towards zero as the radial distance increases and approaches infinity in the limit of zero radial distance. Preconditioning also intends to increase the diagonal elements relative to the off-diagonal ones in  $\Psi$ . Ridge regression enforces such differences and as such, the described preconditioning remains valid.

The preconditioning matrix  $C$  in Dyn is obtained through the identity  $\sum_{t=1}^N (\Delta^n \mathbf{u})_t v_t = \mathbf{v}^T \mathbf{Cu}$ . The  $n$ th iterated Laplacian is given by  $n$  consecutive Laplacians  $\Delta^n(f) = \Delta(\Delta(\dots \Delta(f)))$  with  $\Delta(f) = \partial^2 f / \partial x^2 + \partial^2 f / \partial y^2$ . The summation can be represented on an analogue (Delaunay) triangulated domain containing  $N_\Delta$  triangles on the basis of the  $N$  datapoints;

$$\sum_{j=1}^{N_\Delta} \sum_{i=0}^n \binom{n}{i} \left( \frac{\partial^n u}{\partial x^i \partial y^{n-i}} \right)_j \left( \frac{\partial^n v}{\partial x^i \partial y^{n-i}} \right)_j T_j \equiv \mathbf{v}^T \mathbf{Cu}. \quad (3)$$

Functions  $u$  and  $v$  can be approximated on the  $j$ th triangle, which has an area  $T_j$ , as  $n$ th order polynomials  $\tilde{u}_j = \sum_{t=0}^n a_{t,j} x^t \sum_{l=0}^n b_{l,j} y^l = \sum_{p+q \leq n} \xi_{pq,j} x^p y^q = X_{pq} \xi_j$  and similarly  $\tilde{v}_j = X_{pq} \eta_j$ . Vectors  $\xi_j$  and  $\eta_j$  contain the  $N_\xi = \frac{1}{2}(n+2)(n+1)$  polynomial coefficients for  $\tilde{u}$  and  $\tilde{v}$  respectively, which can be solved for in a least-squares manner  $\xi_j = (X_{pq}^T X_{pq})^{-1} X_{pq}^T \mathbf{u}_j = S_j \mathbf{u}_j$  and alike  $\eta_j = S_j \mathbf{v}_j$ . Vectors  $\mathbf{u}_j$  and  $\mathbf{v}_j$  are  $N_{v,j} \times 1$  in size and contain the  $u$  and  $v$ -values respectively of the  $N_v$  vertices neighbouring the  $j$ th triangle, i.e.  $N_{v,j}$ . These matrix systems can be rewritten such that all  $N$  vertices are included;  $S_j \mathbf{u}_j = \mathbf{B} \mathbf{u}$  and  $S_j \mathbf{v}_j = \mathbf{B} \mathbf{v}$ . Matrix  $\mathbf{B}$  is then a  $N_\xi \times N$  sparse matrix with copied entries of the  $N_\xi \times N_{v,j}$  matrix  $S_j$  and vectors  $\mathbf{u}$  and  $\mathbf{v}$  contain the  $u$  and  $v$  values of all the  $N$  vertices. Using the polynomial representations, each of the derivative terms in equation (3) can now be expressed in terms of  $\xi_j$  and  $\eta_j$  and consequently with reference to discrete values  $\mathbf{u}$  and  $\mathbf{v}$ . Matrix  $C$  is constructed by summing up the different terms pertaining the contributing polynomial coefficients for all triangles. The process is clarified for the case of  $n = 3$  in appendix A.1. It was shown in Dyn *et al* (1986) that this type of preconditioner reduces the condition number by several orders of magnitude. Solutions for the TPS coefficients  $\alpha$  are retrieved through an iterative conjugate gradient scheme, which is presented in appendix A.2 for completeness.

### 2.3. Scale-adaptive TPS-based modelling

Measurement data will unavoidably contain some level of uncertainty due to noise originating from temporal fluctuations in the data, limited spatial resolution of the measurement

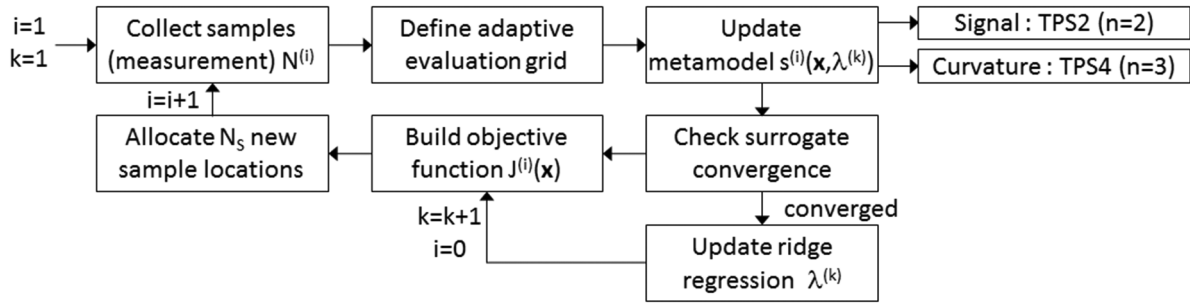


Figure 2. Procedural steps in scale-adaptive sampling.

system, electronic noise, etc. In view of adjusting the sampling process to signal curvature, such regions of jitter will unnecessarily attract samples and continue to do so, as with each new sample the noise component will alter the surrogate model and amplify higher order spatial derivatives. A degree of filtering is therefore required. Data can be low-pass filtered prior to the interpolation process, although this requires an appropriate choice of smoothing kernel and extent (Carr *et al* 2003). Both are dependent on the data landscape *a priori* unknown to the user. Alternatively, ridge regression presented in equation (2) renders the RBF interpolation an approximant.

With increasing  $\lambda$ , the surrogate model equation (1) becomes a smoother representation of the underlying data, effectively modulating smaller scale oscillations. To this extent the choice of  $\lambda$  must be in relation to the noise level. However, over-smoothing must be avoided as oscillations (in space) may genuinely originate from smaller scale phenomena such as e.g. turbulence. For this reason the adaptive sampling routine is nested in a primary process in which the degree of smoothing,  $\lambda^{(k)}$ , is progressively reduced over three iterations (figure 2).

In the first iteration, values of  $\lambda^{(1)}$  are obtained on the basis of ensemble average data scatter  $\langle \sigma_t \rangle = \frac{1}{N} \sum_{t=1}^N \sigma_t$  (Forrester *et al* 2008) and are updated with each new adaptively placed data sample. Multiple readings are taken at every sampling location enabling the estimation of local signal scatter as the signal standard deviation  $\sigma_t$  (see also section 3.2). Once the adaptive sampling process has converged, the second iteration is initiated in which a smoothed TPS2 spline is estimated recursively using a generalised cross-validation (GCV) scheme as suggested and implemented by Bates *et al* (1987). The procedure to obtain the coefficients  $\alpha_\theta$  and  $\beta_\theta$  of the smoothed spline (with polynomial precision) is incorporated in appendix B for completeness. While the smoothed spline is not obtained through ridge regression, it can be expressed in an equivalent form, which will prove useful when evaluating relative improvements in the underlying surrogate model (section 3.3); representing the vector containing the values of the smoothed spline evaluated at the sampling locations by  $\mathbf{f}_\theta$ , the approximation utilising the coefficients of the smoothed spline is given by  $\Psi \alpha_\theta + P \beta_\theta = \mathbf{f}_\theta$ . This can be written analogously in terms of ridge regression (equation (2)),  $(\Psi + \lambda I) \alpha + P \beta = \mathbf{f}$ , such that  $\Psi \alpha + P \beta = \mathbf{f}_\theta$ . Equating the two expressions implies  $\alpha = \alpha_\theta$  and  $\beta = \beta_\theta$  and the corresponding ridge parameter is obtained as

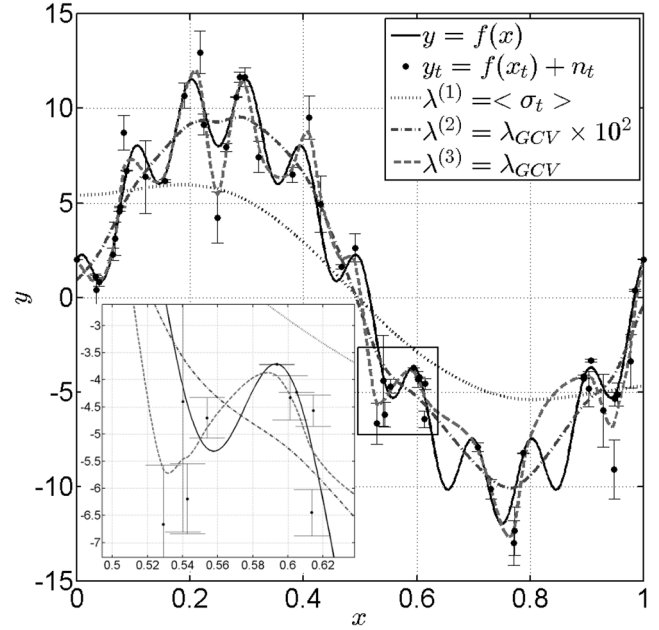


Figure 3. Illustration of progressive smoothing using ridge regression. The underlying signal  $f(x) = 10 \sin(2\pi x) + 2 \cos(20\pi x)$  is sampled with 50 randomly positioned points ( $N = 50$ ), each superimposed with a local uncertainty (error bar width)  $\sigma_t = |5\gamma_t \cos(100\pi x_t) \sin(160\pi x_t)|$  where  $\gamma_t$  are random numbers between 0 and 1 with uniform probability. Measurement noise  $n_t$  is defined as  $n_t = 3\tau_t$  with  $\tau_t \in \mathcal{N}(0, 1)$ . The inset shows a close-up of the graph portion bounded by the rectangle.

$$\begin{aligned}
 (\Psi + \lambda_{GCV} I) \alpha_\theta + P \beta_\theta &= \mathbf{f} \Leftrightarrow \lambda_{GCV} I \alpha_\theta \\
 &= \mathbf{f} - \mathbf{f}_\theta \Leftrightarrow \lambda_{GCV}^2 = \frac{(\mathbf{f} - \mathbf{f}_\theta)^T (\mathbf{f} - \mathbf{f}_\theta)}{\alpha_\theta^T \alpha_\theta}.
 \end{aligned} \quad (4)$$

To ensure a gradual decrease in smoothing and allow for sampling of smaller scales in case of low data uncertainty,  $\lambda^{(k)}$ -values are iteratively scaled down with the following intermediate conditional parameter value

$$\begin{aligned}
 \lambda^{(1)} &= \langle \sigma_t \rangle, \lambda^{(2)} = \lambda_{GCV} \times 10^{\frac{\kappa}{2}}, \lambda^{(3)} = \lambda_{GCV} \quad \text{if } \kappa = \left\lceil \log_{10} \left( \frac{\lambda^{(1)}}{\lambda_{GCV}} \right) \right\rceil > 2 \\
 \lambda^{(1)} &= \langle \sigma_t \rangle, \lambda^{(2)} = \lambda_{GCV}, \lambda^{(3)} = \lambda_{GCV} \times 10^{-2} \quad \text{else.}
 \end{aligned} \quad (5)$$

An example of the influence of  $\lambda$  is illustrated in figure 3. In the first iteration  $\lambda^{(1)}$  is based on the ensemble averaged data scatter  $\langle \sigma_t \rangle$  and the surrogate model reflects the overall signal behaviour. As the value of  $\lambda^{(k)}$  is gradually reduced

towards  $\lambda_{\text{GCv}}$  (in the current example  $\lambda_{\text{GCv}} = 3 \times 10^{-4}$ ), the underlying surrogate model tends to reflect smaller scale oscillations, though a degree of smoothing is retained.

### 3. Adaptive sampling objective function

The objective function  $J^{(i)}(\mathbf{x}_s)$  defined in Theunissen *et al* (2015) consisted of various individual objective factors converting the typical sampling criteria used by experimentalists in analytical formats. With new information added every iteration the objective function is iteratively updated (also here the subscripts ( $i$ ) will be omitted for brevity) and maxima in the function then indicate new suitable sampling locations.

#### 3.1. Curvature dependent sampling

The first criterion  $J_C$  ensures an increased number of samples in regions of higher curvature, similar to mesh-refinement heuristics in computational fluid dynamics e.g. Mavriplis (1995). Curvature is estimated on the basis of the Laplacian  $\Delta$  of the RBF-based surrogate model  $s$ , yielding the criterion  $J_L$ :

$$J_L(\mathbf{x}_s) = \Delta s(\mathbf{x}_s) = \sum_{t=1}^N \alpha_t \Delta \psi(\|\mathbf{x}_s - \mathbf{x}_t\|) + \Delta q(\mathbf{x}_s). \quad (6)$$

The Laplacian of the TPS kernels is tabulated in table 1. For TPS2 kernels the Laplacian of the polynomial term vanishes. Whereas the surrogate model is constructed using the 2nd order TPS, its singularity in the second-order derivative is avoided by basing the curvature on the 3rd order TPS ( $\Delta q(\mathbf{x}_s) = 2(\beta_5 + \beta_6)$ ). The objective function aims to indicate regions of higher curvature, not to be an accurate representation. As such, the use of multiple kernels is permitted.

To compensate for the detrimental effect of stochastic noise,  $J_L$  is combined with a measure for local signal change

$$J_\Delta(\mathbf{x}_s) = |s(\mathbf{x}_s) - s_{\text{MA}}(\mathbf{x}_s)| \quad (7)$$

where  $s_{\text{MA}}(\mathbf{x}_s)$  constitutes the moving average filtered response evaluated with a filter of approximately 30% domain width. While a moving average operation is computationally relatively inexpensive in one dimension when selecting neighbouring data points through a kd-tree, the process becomes too time consuming when extended into two dimensions. Advantage can be taken though of the existence of explicit analytical expressions for the integral of the TPS functions over a  $2a \times 2b$  rectangular domain;

$$\begin{aligned} s_{\text{MA}}(\mathbf{x}_s) &= \frac{1}{4ab} \int_{x_s-a}^{x_s+a} \int_{y_s-b}^{y_s+b} s(\mathbf{x}) \, d\mathbf{x} \\ &= \frac{1}{4ab} \int_{x_s-a}^{x_s+a} \int_{y_s-b}^{y_s+b} \left( \sum_{t=1}^N \alpha_t \psi(\|\mathbf{x} - \mathbf{x}_t\|) + q(\mathbf{x}) \right) \, d\mathbf{x} \\ \xi = (\xi, \eta) = \mathbf{x} - \mathbf{x}_t & \frac{1}{4ab} \sum_{t=1}^N \alpha_t \int_{\xi_1=x_s-a-x_t}^{\xi_2=x_s+a-x_t} \int_{\eta_1=y_s-b-y_t}^{\eta_2=y_s+b-y_t} \psi(\xi) \, d\xi + q_{\text{MA}}(\mathbf{x}_s) \\ &= \frac{1}{4ab} \sum_{t=1}^N \alpha_t \psi_{\text{MA}}(\mathbf{x}_s, \mathbf{x}_t) + q_{\text{MA}}(\mathbf{x}_s) \end{aligned} \quad (8)$$

which for the case of the 2nd order TPS ( $n = 2$ ) gives

$$\begin{aligned} \psi_{\text{MA}}(\mathbf{x}_s, \mathbf{x}_t) &= \frac{1}{3} \log(\xi_2^2 + \eta_2^2)(\xi_2^3 \eta_2 + \eta_2^3 \xi_2) - \frac{1}{3} \log(\xi_1^2 + \eta_1^2)(\xi_1^3 \eta_1 + \eta_1^3 \xi_1) \\ &\quad + \frac{1}{3} \log(\xi_1^2 + \eta_1^2)(\xi_1^3 \eta_1 + \eta_1^3 \xi_1) - \frac{1}{3} \log(\xi_1^2 + \eta_2^2)(\xi_1^3 \eta_2 + \eta_1^3 \xi_2) \\ &\quad - \frac{5}{9}(\xi_2^3 - \xi_1^3)(\eta_2 - \eta_1) - \frac{5}{9}(\eta_2^3 - \eta_1^3)(\xi_2 - \xi_1) \\ &\quad + \frac{1}{3} \xi_2^4 \arctan\left(\frac{(\eta_2 - \eta_1)\xi_2}{\xi_2^2 + \eta_1 \eta_2}\right) - \frac{1}{3} \xi_1^4 \arctan\left(\frac{(\eta_2 - \eta_1)\xi_1}{\xi_1^2 + \eta_1 \eta_2}\right) \\ &\quad + \frac{1}{3} \eta_2^4 \arctan\left(\frac{(\xi_2 - \xi_1)\eta_2}{\eta_2^2 + \xi_1 \xi_2}\right) - \frac{1}{3} \eta_1^4 \arctan\left(\frac{(\xi_2 - \xi_1)\eta_1}{\eta_1^2 + \xi_1 \xi_2}\right) \\ q_{\text{MA}}(\mathbf{x}_s) &= q(\mathbf{x}_s) + \beta_2 a + \beta_3 b. \end{aligned} \quad (9)$$

In one dimension the smoothed kernels reduce to ( $\xi_{1,2} = x_s - x_t \pm a = \xi \pm a$ )

$$\begin{aligned} \psi_{\text{MA}}(x_s, x_t) &= \int_{\xi_1}^{\xi_2} \xi^2 \log(|\xi|) \, d\xi \\ &= (\log(\xi_2) - \log(\xi_1)) + \frac{\xi_2^3}{9} (3 \log(|\xi_2|) - 1) \\ &\quad - \frac{\xi_1^3}{9} (3 \log(|\xi_1|) - 1) \\ q_{\text{MA}}(x_s) &= q(x_s) + \beta_2 a. \end{aligned} \quad (10)$$

The evaluation of the spatially smoothed surrogate hence reverts to a re-evaluation of equation (1), with a novel kernel and polynomial but readily available pre-determined coefficients  $\alpha$  and  $\beta$  (recall that the TPS2 kernel is used to generate the surrogate model while the TPS4 kernel is only used in the calculation of curvature). Following a unity-based normalisation, referenced with subscript  $n$ ,  $(\cdot)_n = [(\cdot) - (\cdot)_{\min}] / [(\cdot)_{\max} - (\cdot)_{\min}]$ , both objectives are combined into the curvature objective  $J_C$ :

$$J_C(\mathbf{x}_s) = ((J_L)_n \times (0.5 + (J_\Delta)_n))^{0.5}. \quad (11)$$

#### 3.2. Uncertainty-based sampling

The second criterion ensures regions of higher uncertainty are adequately sampled by introducing the local level of data scatter as a component in the objective function. Regions in which accrued data shows larger scatter will undoubtedly demand additional samples not only from the perspective of data validation but also from a flow physics perspective. Additional samples will thus verify whether the increased scatter is due to an isolated outlier event or to a coherent region of higher turbulence intensity.

While each variable is assigned one unique value at each sampling location, these values typically correspond to the average of  $N_i$  independent measurements taken at each sampling station. Accordingly, the relative measurement error at the  $i$ th sampling location for 95% confidence level in variable  $(\cdot)$  is given by

$$\epsilon_{(\cdot),i} = 1.96 \times \frac{\sigma_{(\cdot),i}}{\sqrt{N_i} \times (\cdot)_i} \quad \rightarrow \quad J_E(\mathbf{x}_t) = (\epsilon_{(\cdot),i})_n \quad (12)$$

where  $(\cdot)_i$  refers to the mean statistic (time-average) of  $N_i$  values and  $\sigma_{(\cdot),i}$  the standard deviation in the data. The number

of readings  $N_i$  is typically set to 10 000 to retrieve reliable mean statistics or can be locally adjusted according the local data scatter to ensure a constant relative error across the domain. Subsequently, the objective component pertaining error can be expressed as  $J_E(\mathbf{x}_t) = (\epsilon_{(\cdot),t})_n$ . The discrete  $J_E$  values at the data points can also be interpolated using thin plate splines as per section 2.1 yielding a continuous function  $J_E(\mathbf{x}_s)$  for each of the considered variables.

### 3.3. Improvement-driven sampling

Sampling locations whose addition does not alter the surrogate model are evidently of low importance. Conversely, regions where the addition of new data yields a significant change in the previous underlying approximant merit further exploration. This thought is captured in the improvement objective  $J_I^{(i)}$ , incorporating the relative additive information content,  $\Omega^{(i)}(\mathbf{x}_s) = \frac{Q^{(i)}(\mathbf{x}_s)}{1+|s^{(i)}(\mathbf{x}_s)|}$ , where  $Q^{(i)}(\mathbf{x}_s)$  quantifies the model's change in the  $i$ th iteration  $Q^{(i)}(\mathbf{x}_s) = |s^{(i)}(\mathbf{x}_s) - s^{(i-1)}(\mathbf{x}_s)|$ . To retain historic model change, the improvement objective combines the four most recent  $\Omega$ 's in a weighted average, with all weights  $w_{i-m}$  currently equal to 1/4;

$$J_I(\mathbf{x}_s) = \left( \sum_{m=0}^3 w_{i-m} \Omega^{(i-m)}(\mathbf{x}_s) \right)_n. \quad (13)$$

Changes in surrogate model are however not necessarily linked to the addition of novel data, but can be due to alterations in the ridge-regression parameter, uncertainty in the model due to uncertainty in the acquired data and/or limitations in interpolation accuracy. This has been accounted for through three parameters quantifying each of these sources of variation in surrogate model;  $E_\lambda(\mathbf{x}_s)$ ,  $E_\sigma(\mathbf{x}_s)$  and  $E_{\text{TPS}}$ .

1. With each  $i$ th iteration, the ridge regression parameter  $\lambda^{(k)}$  is updated as new data is added. As such, the surrogate model can show variations which do not necessarily relate to an enhanced representation of the underlying signal, but are merely due to a different degree in smoothing, i.e. changes in ridge regression parameter. To quantify these changes, the matrix system in equation (2) is first rewritten as  $[K + \lambda I^*] [\alpha; \beta] = \mathbf{f}_0$  where  $K = [\Psi \ P; P^T \ O]$ ,  $\mathbf{f}_0 = [\mathbf{f}; \mathbf{0}]$  and  $I^*$  is an  $(N + M) \times (N + M)$  matrix. The first  $N$  rows and columns in  $I^*$  constitute the identity matrix and the trailing  $M$  rows and columns are zero. With  $L = [\Psi \ P]$  evaluated at a location  $\mathbf{x}_s$ , the surrogate model value in the  $i$ th iteration is given as  $s^{(i)}(\mathbf{x}_s) = L \left( K + \lambda_{(i)}^{(k)} I^* \right)^{-1} \mathbf{f}_0$ . The regression parameter in the  $(i - 1)$ th iteration is  $\lambda_{(i-1)}^{(k)}$  and the change in surrogate model due to this variation in  $\lambda$ -value,  $\delta_\lambda = \lambda_{(i-1)}^{(k)} - \lambda_{(i)}^{(k)}$ , can then be evaluated as

$$\begin{aligned} E_\lambda(\mathbf{x}_s) &= L \left[ (K + \lambda_{(i)}^{(k)} I^* + \delta_\lambda I^*)^{-1} - (K + \lambda_{(i)}^{(k)} I^*)^{-1} \right] \mathbf{f}_0 \\ &\stackrel{\tilde{K} = K + \lambda_{(i)}^{(k)} I^*}{=} L \left[ \{ \tilde{K}^{-1} - \tilde{K}^{-1} (I + \delta_\lambda I^* \tilde{K}^{-1})^{-1} \right. \\ &\quad \left. \times \tilde{K}^{-1} I^* \tilde{K}^{-1} \} - \tilde{K}^{-1} \right] \mathbf{f}_0 \\ &= -\delta_\lambda L \tilde{K}^{-1} (I + \delta_\lambda I^* \tilde{K}^{-1})^{-1} I^* \tilde{K}^{-1} \mathbf{f}_0 \\ &= L \left[ -\delta_\lambda \tilde{K}^{-1} (I + \delta_\lambda I^* \tilde{K}^{-1})^{-1} \right] \alpha^{(i)}. \end{aligned} \quad (14)$$

In the second step of the derivation the Woodbury matrix identity was applied. It should be noted that  $E_\lambda$  is not simply equivalent to  $s^{(i-1)}(\mathbf{x}_s) - s^{(i)}(\mathbf{x}_s)$  due to the addition of new sampling locations with every iteration. The difference between  $s^{(i-1)}$  and  $s^{(i)}$  thus encompasses a change in ridge regression as well as new data and would require re-evaluation of  $s^{(i-1)}$  on the  $i$ th evaluation grid as this can change between iterations (section 4.2). Instead, in the final expression of equation (14),  $\tilde{K}^{-1}$  has already been calculated in the  $i$ th iteration to evaluate the TPS interpolation coefficients  $\alpha^{(i)}$  for the surrogate model  $s^{(i)}(\mathbf{x}_s)$ . With each of the terms readily available, the change in surrogate model  $E_\lambda(\mathbf{x}_s)$  caused purely by a change in ridge regression parameter is easily obtained.

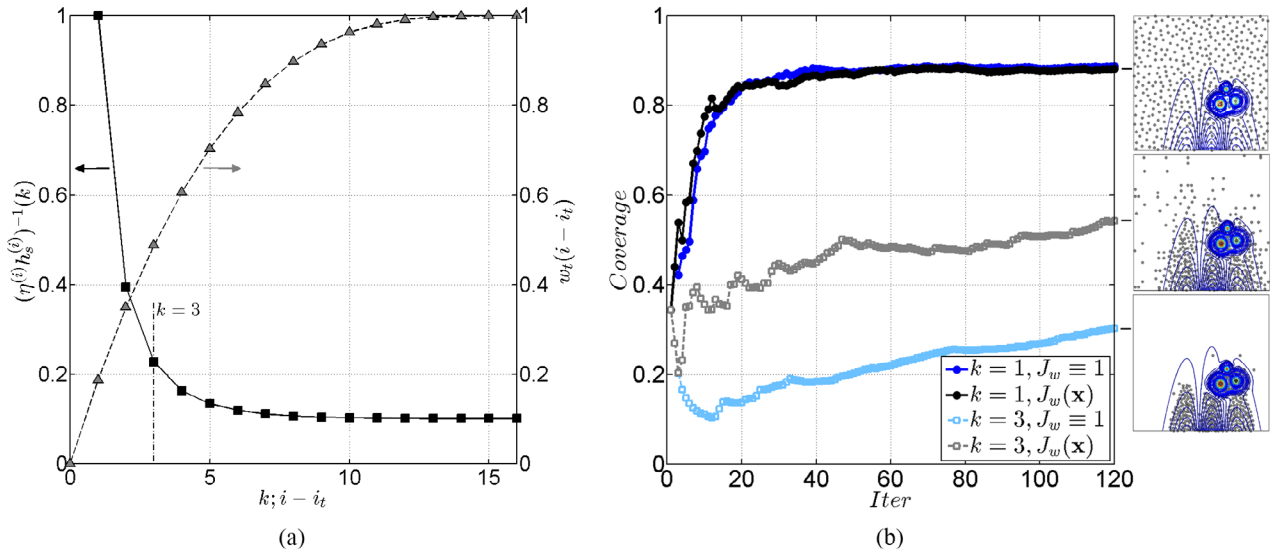
2. When performing point-wise measurements, multiple readings are taken at each station and local data scatter can be suitably modelled through a normal distribution  $\mathcal{N}(0, \sigma)$ . The local average of the readings is subsequently used to fit the surrogate model. When imposing a regression parameter to smoothen too rapid changes in underlying signal merely due to noise, the surrogate model becomes sensitive to the fitted data values as these might vary as a result of the underlying scatter. This subsequently introduces an uncertainty in the thin plate spline approximant, which has been dealt with in Wahba (1990). In this case the variation in surrogate model is caused by data uncertainty. Let  $\sigma_t$  symbolise the data scatter in the  $t$ th sampling point and  $\langle \sigma_t \rangle$  the ensemble average of all  $\sigma_{t=1 \dots N}$  values. The suggested Bayesian confidence interval at 95% for the  $t$ th sample is given as

$$E_\sigma(\mathbf{x}_t) = 1.96 \times \langle \sigma_t \rangle \times \sqrt{H_{tt}} \quad (15)$$

where  $H_{tt}$  is the  $t$ th entry on the diagonal of the influence matrix  $H$ . The influence matrix changes with  $\lambda$  and is expressed in appendix B for completeness. The sporadic values can be subsequently interpolated yielding the function  $E_\sigma(\mathbf{x}_s)$ , quantifying the possible variation in surrogate model, for a given ridge regression parameter, due to uncertainty in the underlying data.

3. Finally, a variation in surrogate model can be due to the inherent interpolation error. Based on Fasshauer (2007) an upper error bound for TPS interpolation is proposed involving the fill distance  $h_s = \sup_{i=1, \dots, N} (\min_{\substack{j=1, \dots, N \\ j \neq i}} (\|\mathbf{x}_i - \mathbf{x}_j\|))$ , i.e. the largest nearest-neighbour distance based on the avail-





**Figure 4.** (a) Evolution of  $(\eta^{(i)} h_s^{(i)})^{-1}$  and  $w_t$  with  $k$  and  $i - i_t$  as per equations (17) and (18) respectively setting  $N_\kappa = 15$  (equation (18)). (b) Illustration of the influence of the iterative weighting factor on the Franke function (figure 8(a)). Insets depict the obtained sampling distribution corresponding to the curves in the graphs aside, superimposed on the contour levels of the underlying Laplacian of the signal.

able data points within the domain, and a scaled native space norm  $F$ ;

$$E_{\text{TPS}} = 10^{\lceil \log_{10}(\chi) \rceil} \times h_s \times F \quad \text{where } \chi = (s(\mathbf{x}_s))_{\max} - (s(\mathbf{x}_s))_{\min} \text{ and } F = \frac{1}{\chi} \boldsymbol{\alpha}^T \Psi \boldsymbol{\alpha}. \quad (16)$$

The authors would like to stress that they do not claim  $E_{\text{TPS}}$  to be a proven or mathematically underpinned error bound. The obtained error bound only provides a conservative *guesstimate* above which changes in surrogate model can no longer be ascribed to interpolation errors.

Values of  $Q^{(i)}(\mathbf{x}_s)$  exceeding the three thresholds;  $E_\lambda(\mathbf{x}_s)$ ,  $E_\sigma(\mathbf{x}_s)$  and  $E_{\text{TPS}}$ , are finally considered sufficiently high to be changes in the surrogate model, genuinely caused by the addition of newly accrued data and are taken into account in the calculation of  $\Omega^{(i)}$ .

### 3.4. Iteration dependent exploratory sampling

New measurement locations will be prescribed by the objective function. However, each region within the domain should ideally be sampled at least once to ensure a representative objective function. To this extent oversampling of local extrema in objective function should be avoided, which in this work is achieved through a separation function  $J_h(\mathbf{x}_s)$ . This function attains zero values at each existing measurement location and tends to unity away from the sample location. The objective function surrounding the points is thus given less importance, promoting exploration of the domain. The influence of each point should be radially symmetric and finite, making the compactly supported Wendland functions ideal kernels (Wendland 2004). Function  $J_h(\mathbf{x}_s)$  is therefore obtained through RBF interpolation as per section 2.1 with  $\psi(r) = (1 - \eta r)_+^6 (35(\eta r)^2 + 18\eta r + 3)$ ,  $\mathbf{f} = \mathbf{1}$  and

omitting the polynomial term. The cutoff function is defined as  $(\zeta)_+ = \zeta$  for  $\zeta \geq 0$  and  $(\zeta)_+ = 0$  else. The radial extent of the kernel is controlled through  $\eta$ . Since increasingly smaller scales are envisaged within the iterative process (section 2.3), samples must be allowed to be positioned in closer proximity to each other as the iterations continue. This can be achieved by adjusting the Wendland kernel's support in function of the changing fill distance  $h_s$  and iteration  $k$  in ridge regression (figure 4(a));

$$(\eta^{(i)})^{-1} = 10^{0.7 + 1.4 \times (k-1)} \times h_s^{(i)}. \quad (17)$$

Starting with a radial extent equal to the fill distance, the kernel reduces asymptotically to  $h_s^{(i)}/10$  (figure 4(a)). The authors would like to draw the reader's attention to the definition of  $\eta$  as the scaling factor preceding  $h_s^{(i)}$  is altered with every ridge-regression iteration  $k$  while the fill-distance itself is updated with every sample addition (iteration  $i$ ).

In line with the concept behind domain exploration, it would not be efficient to sample the domain in sequential regions. Instead, areas of interest must receive a higher number of samples, while avoiding global under-sampling. This is accounted for by introducing a secondary weighting to the objective function, particular to the  $i$ th sample point and dependent on the iteration number  $i$ ;

$$w_t(i) = 1 - \left(1 - \frac{(i - i_t)}{N_\kappa}\right)_+^3 \quad (18)$$

where  $i_t$  designates the iteration number the  $i$ th sample point was introduced. The weighting starts from 0 and gradually reaches a value of 1 after  $N_\kappa$  iterations (figure 4(a)). Parameter  $N_\kappa$  essentially dictates how fast a region can be re-sampled and has been set to  $N_\kappa = 15$  in the current study. The  $w_t$  value at each of the sampling locations is subsequently interpolated again using the TPS2 kernel producing a continuous

weighting  $J_w(\mathbf{x}_s)$  for the objective function.  $J_w$  will avoid recently sampled regions from being re-sampled, enforces domain exploration, yet gradually allows further investigation of already explored domain sections.

The effect of the new weighting term is demonstrated in figure 4(b) for a translated Franke function (see section 7), considering the first ( $k = 1$ ) and last ( $k = 3$ ) data smoothing iteration (see section 2.3). The objective function was based solely on curvature and separation as per equations (11) and (17) respectively. Depicted is the coverage with increasing number of iterations, whereby five samples ( $N_S = 5$ ) are added each iteration. A circular area of  $\pi/4N$  was centered on each sample. Considering  $N$  congruent, non-overlapping circles of diameter  $D$  in a square  $L \times L$  domain such that  $\sqrt{ND} = L$  (in the current approach the domain will be normalised, i.e.  $L \equiv 1$ —section 4.1), the ratio between the total area of the circles and domain is given by  $N \frac{\pi D^2}{4} \frac{1}{L^2} = \frac{\pi}{4}$ . The coverage was subsequently calculated as the ratio between the unified domain area covered by all sampling points and the ideal maximum ( $\pi/4$ ). In the first scale-adaptive smoothing iteration ( $k = 1$ ) homogeneity in the sampling distribution (coverage tending to unity) is driven by  $J_h(\mathbf{x})$  and  $J_w$  is of lesser importance. Indeed, with increasing number of samples the omission of  $J_w$  (or equivalently  $J_w \equiv 1$ ) yields a distribution quasi identical to that obtained with active  $J_w$ . When  $\eta^{(i)}$  is decreased in the third smoothing step ( $k = 3$ ), the effect of  $J_w$  becomes more prominent. Sampling locations become clustered in regions of high curvature when  $J_w \equiv 1$ . A spatially varying  $J_w$  on the other hand returns a distribution more populated in those regions, while simultaneously exploring their vicinity, which is reflected in the higher coverage attained.

### 3.5. Combined objective

At each sampling station multiple variables can be measured. In the case of two-dimensional LDA for example, the velocity magnitudes in two directions are obtained. Each of the  $N_{\text{var}}$  variables will therefore have an associated  $J_C(\mathbf{x}_s)$ ,  $J_E(\mathbf{x}_s)$  and  $J_I(\mathbf{x}_s)$ . For each variable these elementary functions are combined as

$$J_{o,v}(\mathbf{x}_s) = (\kappa_C + J_{C,v}(\mathbf{x}_s)) \times (\kappa_I + J_{I,v}(\mathbf{x}_s)) \times (\kappa_E + J_{E,v}(\mathbf{x}_s)). \quad (19)$$

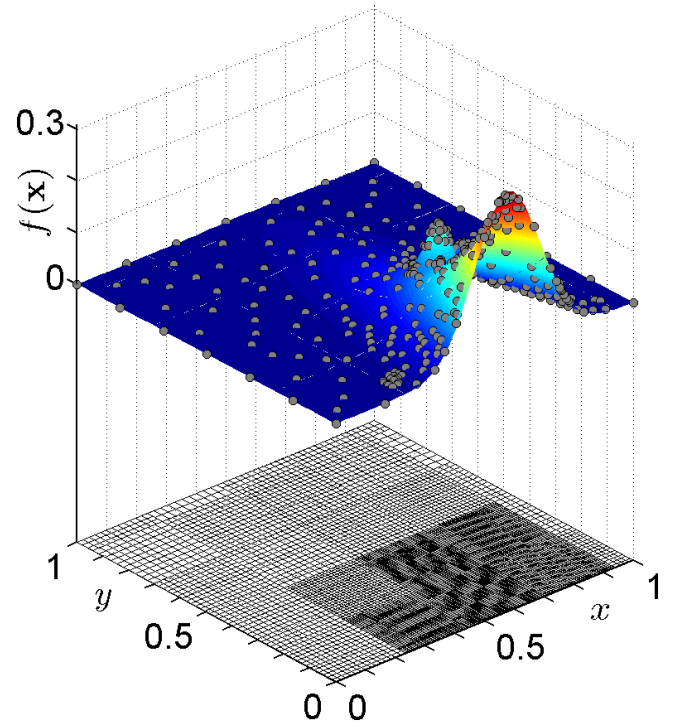
The empirical offset constants  $\kappa$  were set as  $\kappa_C = 0.5$ ,  $\kappa_I = 0.1$  and  $\kappa_E = 0.5$ . The definitive encompassing objective function is finally given by

$$J_O(\mathbf{x}_s) = \left( \prod_{v=1}^{N_{\text{var}}} J_{o,v}(\mathbf{x}_s) \right) \times (1 - (J_h(\mathbf{x}_s))_n) \times J_w(\mathbf{x}_s). \quad (20)$$

## 4. Sample allocation

### 4.1. Domain normalisation

While the above procedures are generally valid, a domain normalisation is introduced for simplicity, i.e. data locations are rescaled to the unit domain such that  $\mathbf{x}_s \in [0, 1]^d$ . This is



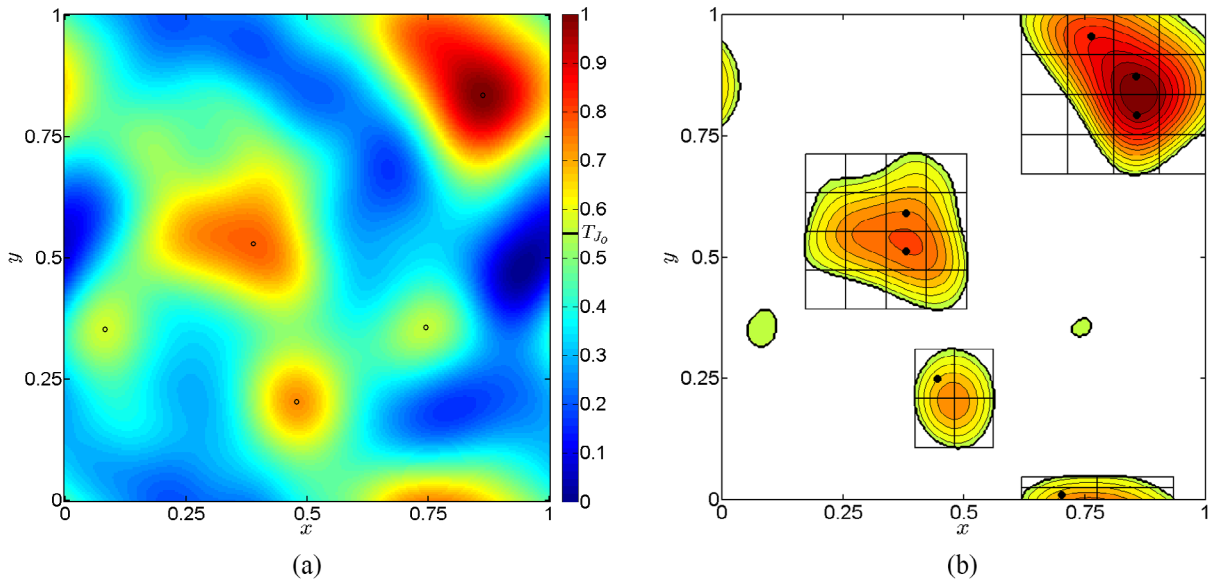
**Figure 5.** Illustration of adaptive evaluation gridding when sampling Franke's bivariate test function:  $f(5(x - 0.6), 5(y - 0.3))$ . Gray dots indicate adaptively placed measurement positions. As points are located more densely in regions where the underlying signal varies, the evaluation grid clearly becomes denser.

achieved by initiating the adaptive process by covering the  $d$ -dimensional domain with equispaced samples along each dimension. The minimum and maximum along each dimension thus define the extent of the observed flow region.

The choice of initial number of samples is left user dependent. However, it should be noted that the total number of samples increases as a power of the dimension. Selecting for example 20 samples in each direction of a two-dimensional domain quickly yields a total of 400 initial samples, while considering a typical total of 1000 measurement locations. Moreover, the numerical assessments presented hereafter will show that the performances of adaptive sampling are relatively insensitive to the choice of initial sampling density. To that extent a modest 7-by-7 sampling grid is deemed a good compromise between effort and information content.

### 4.2. Adaptive evaluation grid

The evaluation grid defines the underlying nodes at which the surrogate models, and consequently the objective function defined in equation (20), is computed. While coarse grids hamper the detection of smaller structures in the objective function, a too fine grid spacing will needlessly augment the computation time. The authors circumvented this problem by partitioning the global domain into regular blocks and adapting the local evaluation grid spacing in each block in response to the local sample spacing. This concept is illustrated in figure 5.



**Figure 6.** Illustration of sample allocation. (a) Local peaks ( $\circ$ ) of  $J_O(\mathbf{x}_s)$  are used to define the threshold  $T_{J_O}$ . (b) Regions of  $J_O(\mathbf{x}_s)$  above the threshold value are attributed a certain number of samples in proportion to their value of  $\Gamma$ . Starting with the region of largest  $\Gamma$ , samples ( $\bullet$ ) are progressively allocated in the c.g. of the smaller sub-portions until the budget of samples (set to  $N_S = 7$  in the current figure) has been depleted.

The Nyquist sampling criterion implies that an oscillation in a signal can only be reconstructed when sampled at least twice. The smallest nearest neighbour distance between samples is obtained as  $h_t = \inf_{i=1,\dots,N} (\min_{i \neq j} \|\mathbf{x}_i - \mathbf{x}_j\|)$  and this will be representative of the smallest measurable spatial fluctuation. Proper characterisation of any fluctuation at this scale will necessitate at least two evaluation grid points and for this reason the spacing is further reduced to fit five grid points. The local grid spacing of the equispaced evaluation grid is thus defined as  $\delta = \frac{1}{5}(h_t)_{MA}$  where  $(h_t)_{MA}$  is the smoothed distribution of  $h_t$  obtained by means of the moving averaging filtering described in section 3.1. Representing, in the  $i$ th iteration, the number of nodes in the  $j$ th block, which has a square size  $l \times l$ , as  $N_{e_j}^{(i)} = \left\lceil \frac{l^2}{(\delta_j^{(i)})^2} \right\rceil$ , the evaluation grid is only updated provided it is sufficiently different from the previous iteration i.e.  $N_{e_j}^{(i)} > 1.25 \times N_{e_j}^{(i-1)}$ .

#### 4.3. Objective function guided sample allocation

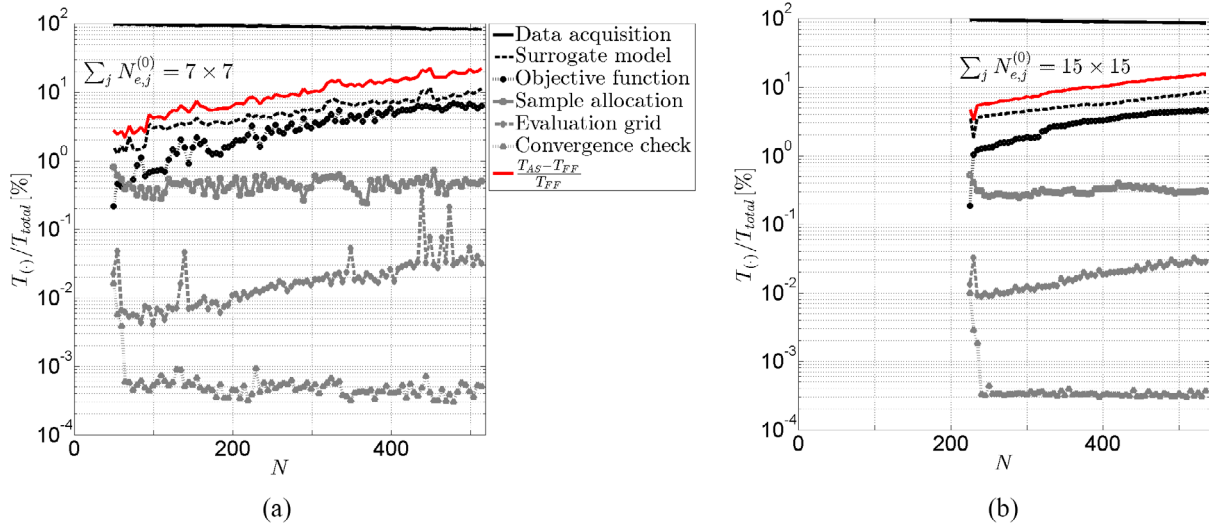
While peaks in the objective function  $J_O(\mathbf{x}_s)$  (equation (20)) indicate the most suitable sampling locations, such a process would rapidly favour local extrema and hence clustering of the newly to-be-added measurement stations within an iteration. To avoid over-sampling, an alternative process has been implemented in this work, basing the sample locations no longer on simply the highest  $J_O$  value, but on local peak geometry. Having defined the objective function, a threshold  $T_{J_O}$  is calculated based on the average of all detected local maxima (figure 6(a)). Potential regions for further investigation are identified where  $J_O(\mathbf{x}_s)$  exceeds the threshold and sorted on the basis of their parameter  $\Gamma$ . This heuristic is evaluated for each region  $V_j$  and quantifies the local extent of the objective

function (in case of two dimensional domains this equals the volume underneath the objective function) and height (maximum value);

$$\Gamma_j = \int_{V_j} J_O(\mathbf{x}) \, d\mathbf{x} \times \max_{\mathbf{x}_s \in V_j} (J_O(\mathbf{x}_s)) \quad \text{with} \quad \forall \mathbf{x}_s \in V_j : J_O(\mathbf{x}_s) \geq T_{J_O}. \quad (21)$$

Each region is then attributed  $n_j = \max\left(1, \left\lfloor \frac{\Gamma_j}{\langle \Gamma \rangle} \right\rfloor\right)$  samples with  $\langle \Gamma \rangle = \frac{1}{N_S} \sum_j \Gamma_j$  and  $N_S$  equals the number of new samples to be distributed each iteration. Of all the identified regions, only the  $N_S$  regions with highest  $\Gamma$  are retained and each is subsequently sub-divided into  $2\lceil\sqrt{n_j}\rceil \times 2\lceil\sqrt{n_j}\rceil$  patches. For each region, patches with the highest  $J_O$  average are then assigned one of the  $n_j$  samples in their corresponding local centers of gravity (based on  $J_O(\mathbf{x}_s)$ ). This process tends to deliver an evenly-spread distribution of points across each region, while ensuring samples are spatially assigned in accordance with the objective function as illustrated in figure 6(b). The minimum sample spacing is thus also limited, allowing for a well-conditioned interpolation matrix.

The number of new sampling points  $N_S$  to allocate each iteration is highly dependent on the underlying function. Generally, adding too many new samples each iteration will automatically enforce space exploration and negate any benefit of adaptive sampling. On the other hand, incrementing the number of samples by only one will yield a lengthy and computationally intense sampling process since the surrogate model needs re-evaluation with each newly accrued data point. The evolution in error and iteration number with  $N_S$  is presented in section 7.1 and the authors found  $N_S = 5$  to provide a reasonable compromise while for one-dimensional adaptivity  $N_S = 1$  is more conducive (Theunissen *et al* 2015).



**Figure 7.** Partition in computational effort with total number of samples  $N$  for the case of the shifted Franke function with  $N_s = 5$  and (a)  $\sum_j N_{e,j}^{(0)} = 7 \times 7$ , (b)  $\sum_j N_{e,j}^{(0)} = 15 \times 15$ .

Once the sampling locations have been identified, the order in which they are visited is optimised analogous to the Traveling Salesman Problem to minimise traverse time.

## 5. Surrogate model convergence

While the number of repeated measurement at a certain sampling location for a Gaussian process can be related to confidence level, local fluctuation level in the data and a user-imposed relative error (equation (12)), the required number of sampling locations to retrieve a reliable estimate of the underlying spatial data distribution is less evident and *a priori* unknown. Too few samples will yield a poor representation of the underlying data landscape, while oversampling is an inefficient use of computational and experimental effort. For this reason the adaptive sampling process should be considered converged when the addition of new data points yields a minimal change in surrogate model over a sufficient number of iterations. The experimentalist is therefore left with the problem of deciding for how many iterations the improvement must remain below a user-defined threshold. In this work a modified version of the adaptive stopping procedure of Theunissen *et al* (2015) has been employed.

While convergence in the surrogate model would imply a near-zero discrepancy, small model changes due to interpolation inaccuracies or modifications caused by uncertainty in the data should actually be discarded. Convergence in the surrogate model is thus essentially equivalent to demanding the previously introduced improvement to be below a certain level, which has been discussed already in section 3.3. The number of consecutive converged iterations  $N_{CCI}$  in which the maximum improvement in surrogate model of each of the measured variables is negligible, i.e.  $Q^{(i)}(\mathbf{x}_s) \leq \max(E_\lambda(\mathbf{x}_s), E_\sigma(\mathbf{x}_s), E_{TPS}) \forall \mathbf{x}_s$ , has been defined as

$$N_{CCI} = 1 + \left\lceil 20 \left( \frac{\mu_1 + \mu_2}{2} \right) \right\rceil. \quad (22)$$

Parameter  $\mu_1$  quantifies the relative importance of the measurement uncertainty with respect to the overall surrogate;  $\mu_1 = \frac{\langle \epsilon \rangle^{(i)}}{|s_{\max}^{(i)} - s_{\min}^{(i)}|}$  with  $\langle \epsilon \rangle^{(i)} = \frac{1}{N^{(i)}} \sum_{t=1}^{N^{(i)}} \epsilon_t$  as per equation (12). The contrast in surrogate model is essentially quantified by  $\mu_2$ ;  $\mu_2 = \frac{|(J_\Delta^{(i)})_{\max} - (J_\Delta^{(i)})_{\min}|}{|(J_\Delta^{(i)})_{\max} + (J_\Delta^{(i)})_{\min}|}$ , adopting the previous definition of  $J_\Delta$  (equation (7)). An overall large data scatter  $\langle \epsilon \rangle$  with respect to the underlying signal will result in high values of  $\mu_1$  and subsequently, an increase in  $N_{CCI}$ . Such a more stringent convergence criterion will thus lead to an increase in number of samples placed and concomitantly an improved confidence in the obtained surrogate model's representativeness of the underlying signal. High values of  $\mu_2$  imply the surrogate landscape to contain few, yet very distinct (high amplitude) peaks. In the presence of such isolated features, a higher number of samples is required to encourage further domain exploration, which is analogous to a higher  $N_{CCI}$ . Conversely, a relatively flat signal would correspond to low values of  $\mu_2$  and requires only few samples to be appropriately modelled. Note that  $N_{CCI}$  is adjusted with every addition of new samples. In the ideal situation of no scatter in the data ( $\mu_1 = 0$ ) and uniform signal (hence  $J_\Delta^{(i)} = 0$  and  $\mu_2 = 0$ ) the minimum number of consecutively converged iterations will equal  $N_{CCI} = 11$  (equation (22)), which was empirically found to be a suitable number.

## 6. Additional remarks

### 6.1. Computational overhead

It is clear from the foregoing that adaptive sampling will be accompanied by an additional overhead compared to

traditional sampling. In view of the numerical performance assessment in the following paragraph, figure 7 depicts the contribution of the different stages in adaptive sampling to the overall time ( $T_{\text{total}}$ ) for the case where the underlying signal is defined by Franke's function. All processes pertaining adaptive sampling were implemented in the commercial software MATLAB (2014) on a 64-bit standard PC (Intel Core i7-4790, 3.6GHz CPU) with 16GB RAM. Regarding data accrual, at each sampling location  $N_i = 10^4$  samples are assumed to be collected at a typical data rate  $f_s$  of 1 kHz to retrieve meaningful statistical moments.

Neither the allocation of samples (section 4.3), nor the establishment of an adaptive evaluation grid (section 4.2), nor convergence checks (section 5) signify a high computational load as seen in figure 7(a). The total time for adaptive sampling is dominated by data acquisition since in each iteration 50 seconds ( $= \frac{N_s \times N_i}{f_s}$ ) are allocated to data collection. With increasing number of samples the construction of the surrogate model and objective function start to take up a significant portion of the total processing time. The establishment of the surrogate model involves the evaluation of the interpolation coefficients  $\alpha$  and  $\beta$  by means of an iterative conjugate gradient method (appendix A.2), the RBF matrix  $\Phi$ , as well as a preconditioning and generalised cross-validation (section 2). The objective function involves the interpolation of the model onto an adaptively refined evaluation grid (section 4.2). As the number of samples grows, so do the involved matrices and hence the computational effort in their manipulation. Exactly these two subroutines constitute the additional overhead compared to the full factorial approach as illustrated by the red lines in figure 7. At approximately 600 samples, adaptive sampling takes approximately 20% longer than FF. However, as will be shown hereafter, with the benefit of obtaining consistent data quality and lower error levels compared to the structured sampling approach.

In figure 7(b) the initial sampling grid consists of 15 nodes in each direction compared to the  $7 \times 7$  grid in figure 7(a). The evaluation grid thus contains a larger number of blocks (section 4.2) and, consequently, each block will accommodate fewer evaluation grid points. Indeed, for a given number of total samples, comparison between figures 7(a) and (b) indicates a reduction in additional computational effort. Moreover, the establishment of the surrogate model, which is independent of the evaluation grid, also shows to take slightly less effort, implying that optimised internal memory management may bring the computational cost down further. This is however beyond the scope of the current work.

## 6.2. Extension to higher dimensions

In describing the adaptive routines proposed, no limitations were imposed in terms of dimensionality of the problem and as such, the concept underlying the proposed adaptive process remains valid in higher dimensions. Some implementation aspects can be expected to require more attention. The preconditioning described in section 2.2 for example was tailored for surface fitting and hence no longer suitable when sampling in

a volumetric domain. In that case, alternative RBF functions, such as the Wendland functions mentioned afore, may prove to be more appropriate. Especially given the rapid increase in data points, the sparsity promoted by these RBF kernels will be beneficial in reducing the computational overhead and negating the need for preconditioning. However, these kernels will require a proper definition of the inherent support radius. Also the optimal ridge-regression parameter  $\lambda^{(k)}$  will necessitate a review of its estimation when changing RBF kernel (Schaback 2007) as this further influences the improvement-driven sampling. Given the vast application of RBF interpolation, the majority of these problematics has already been solved, making the adaptation of the presented strategy for volumetric sampling less cumbersome.

## 7. Performance study

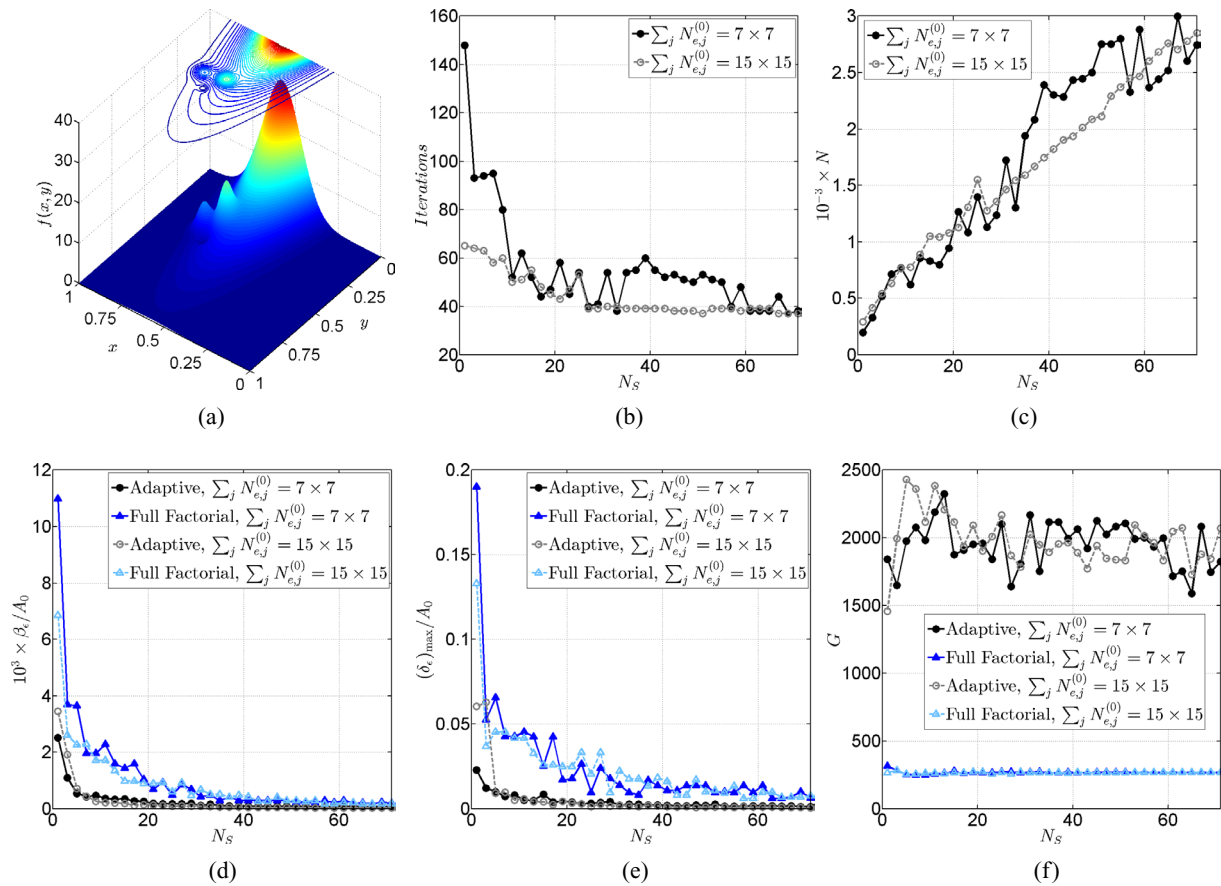
### 7.1. Numerical assessment

To test the performances of the proposed methodology, the adaptive sampling strategy has been tested on two analytical test functions; Franke's bivariate function (Franke 1979) and the numerically simulated flow field underlying a 2D lid driven cavity at a cavity-height-based Reynolds number of 100.

Franke's function was translated by 0.6 and 0.3 in  $x$  and  $y$  direction respectively and coordinates were rescaled by a factor five;  $f(5(x - 0.6), 5(y - 0.3))$ . The surface consists of localised peaks and a trough superimposed on a global exponential (figure 8(a)). The lid driven cavity is chosen as it involves adaptivity to multiple velocity components and presents a complex flow field with a central vortex and two additional regions of recirculating flow in the cavity corners (figure 11(a)).

With no uncertainty in the data, the  $\mu_1$  component in equation (22) equaled zero for all cases while  $\mu_2 = 1$  as the minimum in  $J_D$  was 0. Accordingly, all adaptive sampling processes, each consisting of three iterations in smoothing parameter  $\lambda^{(k)}$ , were considered converged after  $N_{\text{CCI}} = 11$  consecutive iterations in which the improvement in surrogate model was negligible.

Two initial structured grids were considered to construct the very first estimate of the surrogate model and initiate the adaptive process; a  $7 \times 7$  and  $15 \times 15$  grid. Once the adaptive approach converged, the uniform sample spacing in the FF was defined as  $h_{\text{FF}} = \lceil 1/\sqrt{N} \rceil$  at which the function values were evaluated. Note that the number of samples with FF will always satisfy  $N_{\text{FF}} = h_{\text{FF}}^{-2} \geq N$ . To juxtapose the 'goodness' of the adaptive sampling (AS) distribution with equispaced sampling, a Laplacian-based quality heuristic  $G$  was introduced defined as  $G = \frac{1}{N_{(\cdot)}} \sum_{t=1}^{N_{(\cdot)}} |\Delta f(\mathbf{x}_t)|$  where  $N_{(\cdot)} = N$  or  $N_{(\cdot)} = N_{\text{FF}}$  depending on the approach selected. High values of  $G$  indicate the sample distribution to be more dense in regions of higher signal curvature, in line with the curvature dependency criterion. Errors were quantified by defining a common  $200 \times 200$  evaluation grid ( $\mathbf{x}_e$ ) with a fixed spacing of  $h_e = 0.005$  onto which the FF and AS values



**Figure 8.** (a) Illustration of the modified Franke function  $f(5(x - 0.6), 5(y - 0.3))$  with superimposed contours. Considering Adaptive and Full Factorial sampling with two initial sampling densities; evolution in (b) number of iterations before convergence, (c) total number of samples  $N$ , (d) rms error ( $\beta_\epsilon$ ) normalised with the maximum value of the underlying signal  $A_0$ , (e) normalised maximum error  $(\delta_\epsilon)_{\max} / A_0$  and (f)  $G$  heuristic, with imposed samples per iteration  $N_S$ .

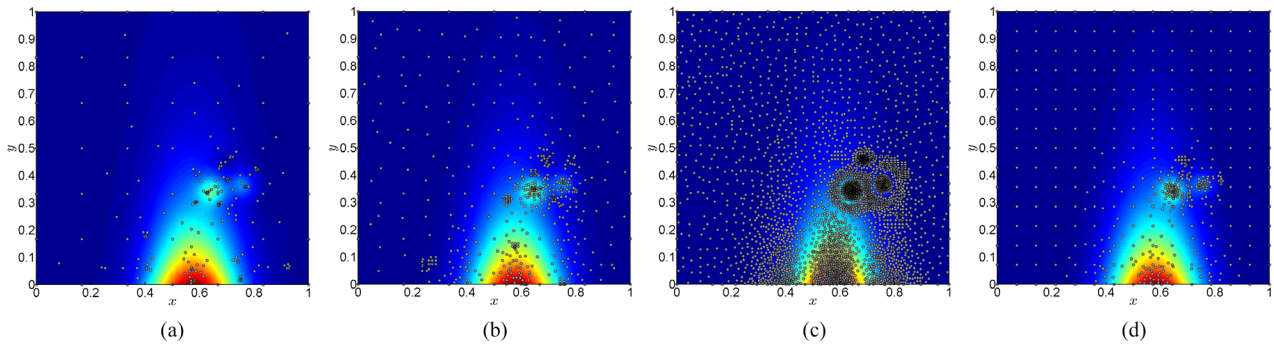
were re-interpolated using a 2nd order Thin Plate Spline augmented with a linear polynomial (see section 2.1). To minimise the influence of interpolation errors on the domain edge, the evaluation grid span between 0.05 and 0.95 in both directions. At each of the  $N_e = 4 \times 10^4$  grid nodes the absolute difference between imposed and interpolated value constituted the local error  $\delta_\epsilon(\mathbf{x}_e) = |s(\mathbf{x}_e) - f(\mathbf{x}_e)|$  and the single-valued root mean square (rms) error was defined as

$$\beta_\epsilon = \frac{1}{N_e} \sum_{q=1}^{N_e} \delta_{\epsilon,q}^2 - \left( \frac{1}{N_e} \sum_{q=1}^{N_e} \delta_{\epsilon,q} \right)^2.$$

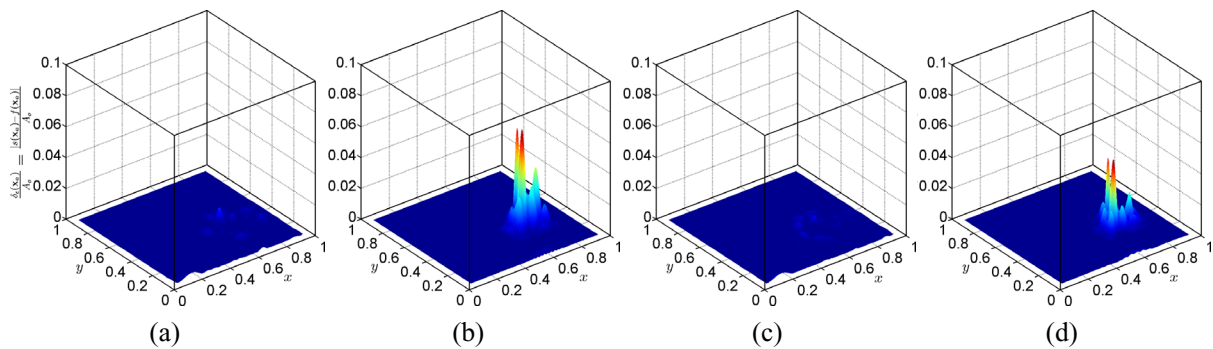
The authors would like to re-iterate the fact that the total number of samples was not pre-defined, but automatically determined by the algorithm based on the stopping criteria discussed in section 5.

Figure 8 gives an overview of the algorithm's performance with respect to the traditional full factorial sampling (FF) approach in the analysis of the shifted Franke function. As the number of samples imposed per iteration  $N_S$  is increased, fewer iterations are needed to reach convergence (figure 8(b)). Augmenting the initial number of samples from 49 to 225 shows the process to converge faster because of the better initial estimate of the underlying signal. However, the total number of samples continues to increase in a near linear fashion (figure 8(c)) and is nearly equal for both initial grids. The standard deviation in error (figure 8(d)) and

maximum error (figure 8(e)) initially decrease, reaching asymptotic values equal for both FF and adaptive sampling. Up to  $N_S \approx 30$  AS reduces both errors by a factor 3 compared to FF, independent of the initial sampling density used. As such, the adaptive approach shows to be independent of the initial sampling grid used, provided of course the initial surrogate model is at least to some degree representative of the underlying function. Also the sampling quality is consistently higher for adaptive sampling as the samples are located in appropriate regions (figure 8(f)). This can be clearly seen in figure 9. With  $N_S = 71$  the regions of higher curvature are clearly visible through the heightened sample concentrations. However, these regions become increasingly oversampled when  $N_S$  is enlarged as the interpolation accuracy improves due to the reduced sample spacing. Driven by the objective  $J_I(\mathbf{x}_s)$  (section 3.3) additional samples are placed to further enhance the surrogate model. For  $N_S = 5$  the denser initial (structured) grid provides an overall, better initial estimate of the underlying function (figure 9(d)), yet, the function's extremities are sampled to a lesser amount compared to the sparser initial grid, yielding a comparable maximum error (figure 8(e)). For both initial grids tested, the addition of 5 new samples ( $N_S = 5$ ) with each iteration shows to result in an appropriate sampling. In view of the time required to perform the experiment, a total number of samples  $N \sim \mathcal{O}(10^3)$



**Figure 9.** Adaptive sample distributions for the modified Franke function (figure 8(a)). Initial grids  $\sum_j N_{e_j}^{(0)} = 7 \times 7$ ; (a)  $N_S = 1$ , (b)  $N_S = 5$ , (c)  $N_S = 71$ . (d)  $\sum_j N_{e_j}^{(0)} = 15 \times 15$   $N_S = 5$ . The colors indicate the reconstructed signal from the data samples.



**Figure 10.** Error distribution across the modified Franke function obtained with an initial grid of  $7 \times 7$  samples and (a) and (c) adaptive sampling and (b) and (d) equispaced sampling. (a) and (b)  $N_S = 5$ ; (c) and (d)  $N_S = 11$ .

is desirable, in which case the typical number of samples to be added each iteration should be less than 20 according to figure 8(c). Comparing the error distributions (figure 10) it can be seen that little is gained going from  $N_S = 5$  to  $N_S = 11$  despite a near doubling of the total number of samples. Both AS and FF have the largest errors in the regions of largest spatial variance, but already for  $N_S = 5$  AS reaches error levels comparable to those achievable with  $N_S = 11$ . It is as such proposed that  $N_S = 5$  is most suitable.

For the lid driven cavity figures 11(b) and (c) depict the evolution in maximum and random error with respect to the total number of samples adding with each iteration five new samples ( $N_S = 5$ ). It should be noted that the sampling is based on surrogate models of both the horizontal ( $U$ ) and vertical ( $V$ ) velocity component. All heuristics, calculated for each of the components, are therefore arithmetically averaged and displayed in figure 11. Initially the error plots for both AS and FF approaches are very similar. However, structured sampling yields a higher variation in maximum error and increasing the number of samples does not guarantee a reduction in maximum error (figure 11(b)). The adaptive process on the other hand yields a typical reduction in  $(\delta_\epsilon)_{\max}$  and random error by a factor two compared to FF. In other words, for FF to achieve the same error level as AS, at least double the number of samples are required. Even then, based on figure 11(b) this still does not ensure FF to attain the same error level.

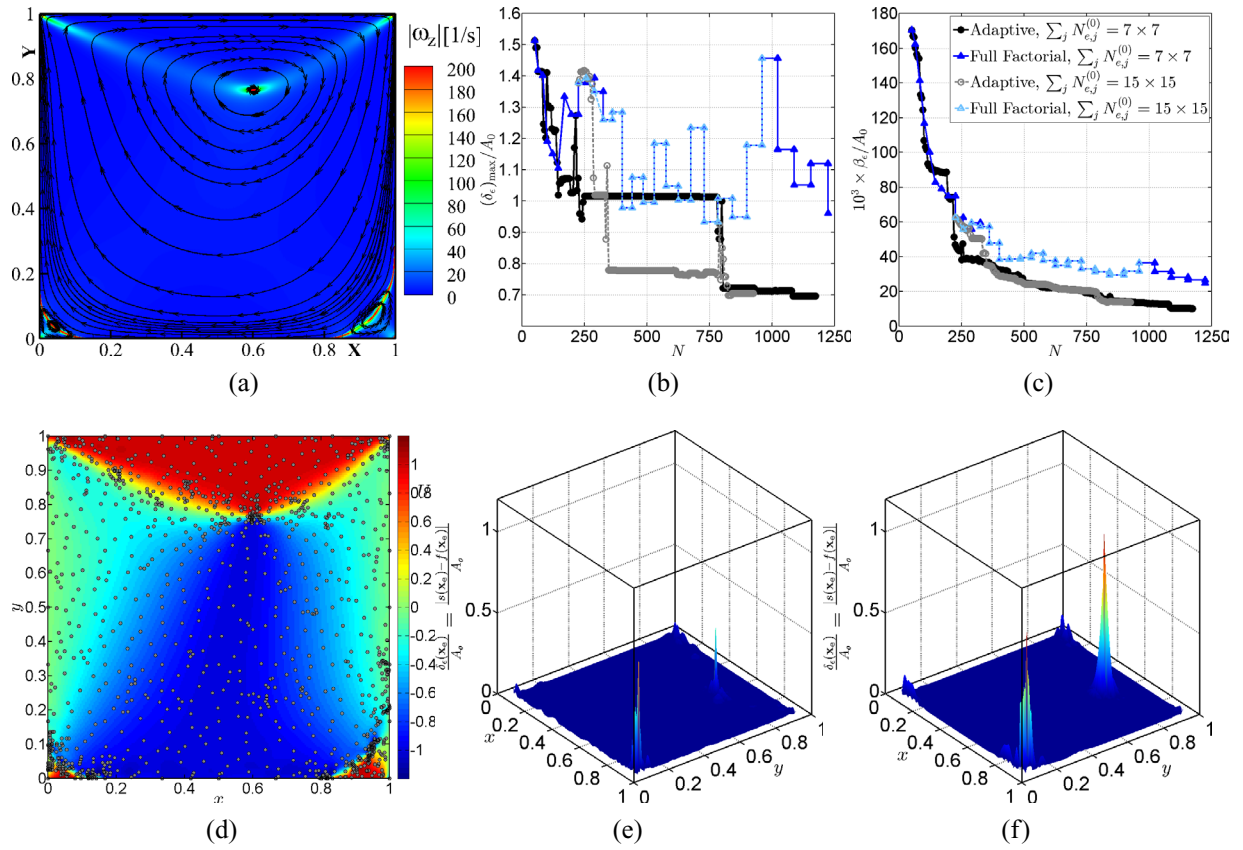
The smoothing iterations can be clearly identified in figures 11(b) and (c); for an initial grid of  $7 \times 7$  samples

the value  $\lambda^{(1)}$ , based on data scatter, is used for the first 38 iterations ( $N = 7 \times 7 + 38 \times 5 = 239$ ). Although the maximum error for  $\lambda^{(2)}$  remains largely constant, this tendency does not reflect the local improvements in surrogate model. As can be seen in figure 11(c), the random error continues to decrease. In approximately the 144th iteration ( $N = 7 \times 7 + 144 \times 5 = 769$ ) the final smoothing is applied. In this stage smaller flow scales are considered, which, especially for  $\sum_j N_{e_j}^{(0)} = 7 \times 7$  yields a sudden reduction in maximum error.

Increasing the initial sampling density allows AS to reach lower maximum errors during the converging process. However, as can be seen from figures 11(b) and (c), once the process converges, final error levels are identical. Higher sampling densities with AS are observed in regions of strong gradients, notably the corner vortices, the top of the cavity and the central inflection point ( $X \approx 0.6, Y \approx 0.75$ ). This clearly leads to a drastic reduction in local error (figures 11(e) versus (f)). With FF exactly these regions consistently present the highest errors.

## 7.2. Experimental assessment

The final assessment consisted of the application of the adaptive methodology to an experimental test. The case of an open cavity flow with a length-to-depth ( $L/H$ ) ratio of 2 was considered at a Reynolds number  $Re_H \approx 37 \times 10^3$ . Experiments were performed in the open-return low-speed wind tunnel



**Figure 11.** (a) Illustration of the 2D lid driven cavity at  $Re = 100$  where the streamline pattern is superimposed onto the contours of the absolute out-of-plane vorticity component  $\omega_z = \partial V/\partial X - \partial U/\partial Y$ . Evolution in (b) normalised maximum error ( $(\delta_\epsilon)_{\max}/A_0$ ) and (c) normalised rms error ( $\beta_\epsilon/A_0$ ) with total number of samples  $N$  for a varying initial sampling grid and adaptive or full factorial sampling strategy. (d) Sampling locations for  $N_s = 5$  superimposed on contours of the horizontal ( $U$ ) velocity component for an initial sampling grid of  $7 \times 7$  nodes. Corresponding normalised error distribution in  $U$  with (e) adaptive and (f) full factorial sampling.

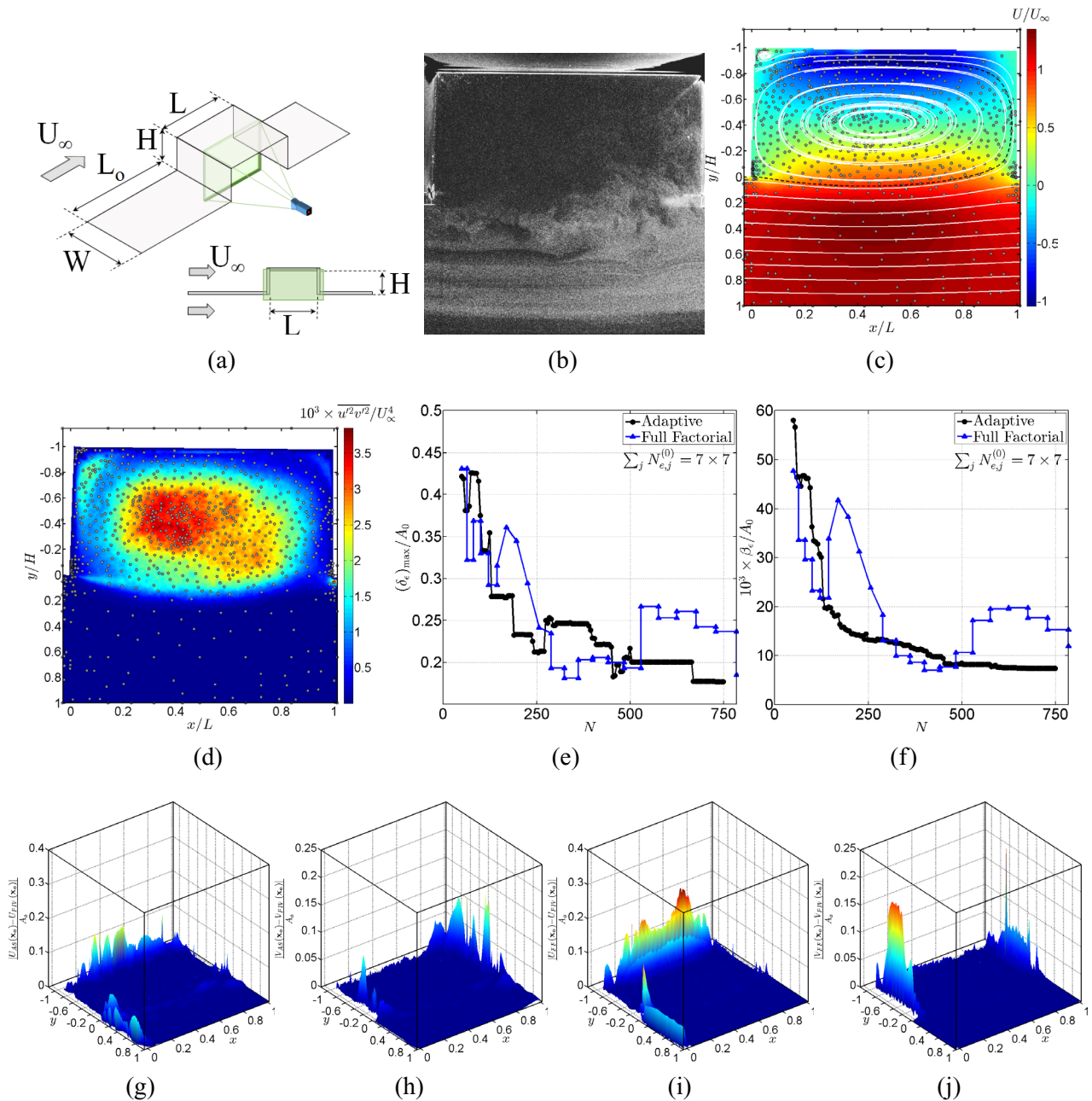
of the University of Bristol at a flow speed of approximately  $10.33 \text{ m s}^{-1}$  and a turbulence intensity of approximately 2%. 2D2C Particle Image Velocimetry measurements were performed on a cavity 110 mm long and 55 mm deep using a Dantec Dynamics system consisting of a 200 mJ Litron laser and 4MP FlowSense EO camera. A sketch of the setup and exemplary PIV image illustrating the location of the cavity walls is provided in figure 12. The  $2072 \times 2072 \text{ px}^2$  sensor corresponded to a field of view of  $115 \times 115 \text{ mm}^2$  and images were analysed using initial interrogation windows of  $65 \times 65 \text{ px}^2$ , recursively reduced to  $17 \times 17 \text{ px}^2$ . A mutual window overlap of 75% yielded a final vector spacing of approximately 0.2 mm.

In total 1000 instantaneous velocity fields were considered ( $N_i = 10^3$ ) in the calculation of the mean and fluctuating velocity components,  $(\bar{u}, \bar{v})$  and  $(\overline{u^2}, \overline{v^2})$  respectively, where  $\bar{u}(\mathbf{x}) = \frac{1}{N_i} \sum_{q=1}^{N_i=10^3} u_q(\mathbf{x})$  and  $\overline{u^2}(\mathbf{x}) = \frac{1}{N_i} \sum_{q=1}^{N_i} u_q^2(\mathbf{x}) - \bar{u}(\mathbf{x})^2$ . The latter stresses then enabled the determination of the local standard deviations, e.g.  $\sigma_u = \sqrt{\frac{N_i}{N_i-1} \overline{u^2}}$ , to establish relative measurement estimates as per equations (12) and (15). The extracted PIV data subsequently served as input into the adaptive sampling process adopting linear data interpolation from the structured PIV data to the intermediate locations.

The adaptive process was initiated by taking measurements at 49 ( $7 \times 7$ ) equispaced samples. Each iteration five new samples ( $N_s = 5$ ) were added. Typical LDA traverse increments are in the order of 0.1 mm. In terms of the unit domain, this implies that sample locations must be located on multiples of  $\frac{0.1}{115}$ . This was taken into account when attributing new data positions. Useful data can only be extracted within the flow region. To this extent cavity walls are to be excluded when attributing new sampling locations. This was accounted for by setting the objective function (equation (20)) to zero within the corresponding regions as evidenced when overlaying the FOV with the predicted sampling locations (figure 12(c)).

The number of consecutive iterations to establish convergence in each of the smoothing phases through ridge regression was automatically set to  $N_{\text{CCI}} = 16$ ,  $N_{\text{CCI}} = 15$  and  $N_{\text{CCI}} = 15$  respectively. A higher concentration of samples is allocated within the cavity compared to the outer flow region. A zone of recirculation can be noted in the upstream corner of the cavity as visualised by the streamlines and this region, as well as the upstream and downstream cavity stagnation points are adaptively allocated more samples. Contrary to the numerical cases, which are void of any fluctuation in signal, the distribution of new samples is now influenced by the local data uncertainty, i.e.  $J_E$ .





**Figure 12.** (a) Sketch of the open cavity experimental setup ( $L_o = 200\text{ mm}$ ,  $H = 55\text{ mm}$ ,  $L = 110\text{ mm}$ ,  $W = 200\text{ mm}$ ) at  $\text{Re}_H = 37 \times 10^3$ . The green square indicates the PIV field-of-view. (b) Exemplary PIV image (flow from left to right) in which the cavity walls can be seen in the top part. The shear layer shed from the upstream stagnation point and impinging on the far cavity wall, is rendered visible through the inhomogeneous seeding density as a result of inherent flow turbulence. (c) Contours of the longitudinal velocity component with superimposed streamlines (white lines) and adaptively placed sampling positions (grey dots). The dashed black line represents the dividing streamline issued from  $(x/L, y/H) = (0, 0)$ . (d) Contours of the product between the time-averaged fluctuations in the horizontal and vertical velocity component. Evolution in (e) maximum and (f) rms error with total number of samples. Error distribution obtained with adaptive sampling in (g) horizontal and (h) vertical velocity component compared with PIV. Error distribution as per full factorial sampling in (i)  $U$  and (j)  $V$  component.

This is manifested when samples are superimposed onto the product of local Reynolds stresses  $\overline{u^2} \times \overline{v^2}$  (figure 12(d)). Whereas the adaptive approach yields a consistent reduction in maximum error with increasing number of samples (figure 12(e)), the standard approach shows stronger fluctuations. Not only does AS attain error levels inferior to those of FF (figures 12(e) and (f)), but, as per the numerical cavity flow, adding more data does not ensure either the maximum, or the

rms error of FF to reduce. On the contrary, adding samples can actually increase the maximum error as samples are not optimally placed in flow regions of spatially rapid changes such as the cavity corners. This leads to poor representations of the flow as manifested in the error distribution plots (figures 12(g)–(j)). In the bulk of the flow where spatial gradients in time-averaged velocity are low, differences between full factorial and adaptive sampling are marginal. However,

in the close vicinity of the cavity walls, adaptive sampling clearly yields a representation in better agreement with the PIV results.

Both this test case and foregoing numerical assessments corroborate AS to be more reliable and efficient in the measurement of an *a priori* unknown flow field;

1. Error levels in the reconstructed signal with data sampled through AS are lower than those sampled by means of FF.
2. Without knowing the underlying signal, it is impossible to preset the required number of samples to collect. In the FF approach  $N$  is thus rendered a critical user-defined parameter. The adaptive routine on the other hand automatically stops when sufficient convergence is reached and does not require the experimentalist to make this decision.
3. The observed instability in error with FF, from an experimental perspective, is a highly unwanted anomaly. When using the traditional equispaced sampling, the assessments have indicated that the experimentalist cannot be certain that the obtained result would be more accurate if more data points were added. He/she is therefore required to gradually increase the number of samples and make decisions based on observed changes in the reconstructed flow field. Such a process however intrinsically replicates the concept behind adaptive sampling.
4. Adaptive sampling shows a more consistent decrease in error (both maximum and rms) with total number of samples. Aborting the adaptive process before reaching convergence will subsequently still yield an overall result more reliable than FF.

## 8. Conclusions

With the aim of improving experimental efficiency when surveying e.g. flow regions by means of point-wise metrologies, an adaptive strategy has been proposed to properly sample 2D areas of interest in an automated manner. The methodology incorporates sampling criteria based on local signal curvature, improvement, uncertainty and domain exploration. In addition, initial smoothing is applied and progressively reduced as to mitigate the effect of local data jitter due to measurement noise. The different processes are elaborated within this paper. Thin Plate Splines finally allow the reconstruction of the surrogate model onto a self-adapting evaluation grid. Automated stopping criteria have been implemented to imply convergence is reached and the numerical and computational implications are discussed within the document.

The performance of the adaptive sampling strategy is assessed and juxtaposed with results obtained by traditional, equispaced sampling (full factorial sampling) based on Franke's function and the lid driven cavity flow. In addition,

both approaches are applied to an open cavity experiment. All test cases attest regions of interest to be consistently attributed a higher number of samples through adaptive sampling. The optimum number of samples to add with each iteration was found to be 5. This yields a drastic reduction (typically factor 2–3) in error compared to traditional full factorial sampling, independent of the initial number of samples. To reach the same level of error as the adaptive approach, traditional factorial sampling thus requires at least twice the number of samples. More importantly, whereas adaptive sampling yields lower error levels with increasing number of samples, full factorial sampling does not guarantee such a tendency. However, without any *a priori* knowledge of the flow field, the experimentalist is unable to determine an appropriate number of samples when adopting equispaced sampling. Traditional sampling thus demands the experimentalist to frequently re-evaluate the reconstructed signal and augment the number of samples until presumed convergence is reached. Such a process intrinsically replicates the concept behind adaptive sampling and, combined with the fact that the automatic stopping criteria implemented have shown to be effective in the assessments provided, corroborates the adaptive approach to be more reliable and efficient in the measurement of an unknown signal/flow field.

## Acknowledgments

This work was supported by the Engineering and Physical Sciences Research Council Impact Acceleration Account (Grant number EP/K503824/1).

## Data access statement

All underlying data are described and provided in full within this paper.

## Appendix A. Surrogate model construction

### A.1. TPS preconditioning

The preconditioner for the case of  $d = 2, n = 3$  (TPS4) discussed in section 2.2 is provided below for clarity. A third-order polynomial is locally fitted to the available measurement data, enabling a straightforward evaluation of the spatial derivatives defining the preconditioning matrix (equation (3));

$$\begin{aligned} \tilde{u}_j = & \xi_{30}x^3 + \xi_{03}y^3 + \xi_{21}x^2y + \xi_{12}xy^2 + \xi_{20}x^2 + \xi_{02}y^2 \\ & + \xi_{11}xy + \xi_{10}x + \xi_{01}y + \xi_{00}. \end{aligned} \quad (\text{A.1})$$

Following the procedure presented in section 2.2, the polynomial coefficients are obtained by means of least-squares fitting and can be expressed in terms of all available data points;

$\xi_j = \mathbf{B}\mathbf{u}$  and  $\eta_j = \mathbf{B}\mathbf{v}$ . The preconditioning matrix  $C$  then follows from approximating the Laplacian in terms of the derivatives of the polynomials;

$$\begin{aligned} & \sum_{j=0}^{N_\Delta} \sum_{i=0}^3 \binom{3}{i} \left( \frac{\delta^3 \mathbf{u}}{\delta x^i \delta y^{3-i}} \right)_j \left( \frac{\delta^3 \mathbf{v}}{\delta x^i \delta y^{3-i}} \right)_j T_j \\ & \approx \sum_{i=0}^{N_\Delta} (36\xi_{30}\eta_{30} + 12\xi_{12}\eta_{12} + 12\xi_{21}\eta_{21} + 36\xi_{03}\eta_{03})_j T_j \\ & \approx \sum_{j=0}^{N_\Delta} (36B_{1,*}\mathbf{u}B_{1,*}\mathbf{v} + 12B_{4,*}\mathbf{u}B_{4,*}\mathbf{v} + 12B_{3,*}\mathbf{u}B_{3,*}\mathbf{v} + 36B_{2,*}\mathbf{u}B_{2,*}\mathbf{v})_j T_j \\ & \approx \mathbf{v}^T \sum_{j=0}^{N_\Delta} \begin{bmatrix} 6B_{1,*}^T & 6B_{2,*}^T & \sqrt{12}B_{3,*}^T & \sqrt{12}B_{4,*}^T \end{bmatrix}_j T_j \mathbf{u} \\ & \approx \mathbf{v}^T \sum_{j=0}^{N_\Delta} C_j \mathbf{u} \equiv \mathbf{v}^T C \mathbf{u}. \end{aligned} \tag{A.2}$$

The notation  $B_{i,*}$  refers to the  $i$ th row of matrix  $B$  including all columns.

### A.2. Iterative conjugate gradient

The solution process to the preconditioned system of equations for TPS interpolation is provided as per (Dyn *et al* 1986) for completeness. For further details regarding derivation the reader is referred to the related reference.

$$\begin{aligned} \boldsymbol{\alpha}^{(-1)} &= \boldsymbol{\alpha}^{(0)} = \mathbf{0}, \quad \mathbf{r}^{(0)} = \mathbf{f}, \quad \rho^{(1)} = 1, \\ \mathbf{x}^{(q)} &= C\mathbf{r}^{(q)}, \quad \mathbf{y}^{(q)} = (\Psi + \lambda I)\mathbf{x}^{(q)}, \quad \eta^{(q)} = (\mathbf{r}^{(q)})^T \mathbf{x}^{(q)}, \\ w^{(q)} &= \frac{\eta^{(q)}}{(\mathbf{x}^{(q)})^T \mathbf{y}^{(q)}}, \quad \rho^{(q+1)} = \left( 1 - \frac{w^{(q)}\eta^{(q)}}{w^{(q-1)}\eta^{(q-1)}\rho^{(q)}} \right)^{-1}, \\ \mathbf{r}^{(q+1)} &= \rho^{(q+1)} \left( \mathbf{r}^{(q)} + w^{(q)}\mathbf{y}^{(q)} \right) + \left( 1 - \rho^{(q+1)} \right) \mathbf{r}^{(q-1)}, \\ \boldsymbol{\alpha}^{(q+1)} &= \rho^{(q+1)} \left( \boldsymbol{\alpha}^{(q)} + w^{(q)}\mathbf{x}^{(q)} \right) + \left( 1 - \rho^{(q+1)} \right) \boldsymbol{\alpha}^{(q-1)}. \end{aligned} \tag{A.3}$$

Superscript ( $q$ ) indicates the iteration number. The process is stopped when convergence is reached in  $\boldsymbol{\alpha}^{(q)}$  or the number of iterations exceeds a maximum.

The preconditioner was constructed by imposing the condition that  $C$  eliminates the polynomial contribution in the RBF interpolation (equation (2));  $CP = \mathbf{0}$ . The polynomial coefficients can then be recovered through  $P\boldsymbol{\beta} = \mathbf{f} - (\Psi + \lambda I)\boldsymbol{\alpha}$ , which can again be solved for  $\boldsymbol{\beta}$  in a least-squares manner

$$\boldsymbol{\beta} = (P^T P)^{-1} P^T (\mathbf{f} - (\Psi + \lambda I)\boldsymbol{\alpha}). \tag{A.4}$$

## Appendix B. Smoothed spline estimation using GCV

The procedure to obtain a smoothed spline approximation using Generalised Cross Validation (GCV) is repeated from Bates *et al* (1987) for completeness and extended to include the regression parameter  $\lambda$ . The routine starts by taking a QR decomposition of the matrix containing the  $M = \frac{(n-1+d)!}{(n-1)!d!}$  polynomial basis functions evaluated at the  $N$  sampling

locations ( $k = 1, \dots, N$ );  $P = FG$  with  $P_{k,*} = [x_k \ y_k \ 1]$ . For the case of a 2nd order spline ( $n = 2$ ) in two dimensions ( $d = 2$ ),  $M = 3$ . Matrices  $F_1$  and  $F_2$  respectively contain the first  $M$  and trailing  $N - M$  columns of  $F$  ( $F = [F_1 \ F_2]$ ) while  $G_1$  consists of the first  $M$  rows of  $G$  ( $G = [G_1; G_2]$ ). An additional matrix  $L$  is obtained through a Cholesky decomposition of the matrix product  $L^T L = F_2^T \Psi F_2$ , followed by the singular value decomposition of  $L^T$ ;  $USV^T = L^T$ . Matrix  $S$  will be a  $N \times N$  diagonal matrix with elements  $s_i$  along the diagonal. Introducing the vector  $\mathbf{z} = U^T F_2^T \mathbf{f}$  and the  $(N - M) \times (N - M)$  identity matrix  $I$ , the optimal values for the coefficients are given as

$$\boldsymbol{\alpha}_\theta = F_2 U (S^2 + N\theta I)^{-1} \mathbf{z} \tag{B.1}$$

$$\boldsymbol{\beta}_\theta = F_1^T \mathbf{f} - F_1^T \Psi \boldsymbol{\alpha}_\theta. \tag{B.2}$$

The most suitable value of  $\theta$  corresponds to the location where  $V(\theta)$  attains a minimum, which can be solved by means of e.g. a golden section search.

$$V(\theta) = \frac{\sum_{j=1}^{N-M} (\kappa_j z_j)^2}{\left( \sum_{j=1}^{N-M} \kappa_j \right)^2} \quad \text{with} \quad \kappa_j = \frac{N\theta}{s_j^2 + N\theta}. \tag{B.3}$$

The influence matrix is given by

$$H(\theta) = F_1 F_1^T + F_2 U S^2 (S^2 + N\theta I)^{-1} U^T F_2^T.$$

Using equation (B.1) to obtain an expression for  $(S^2 + N\theta I)^{-1}$  in terms of  $\boldsymbol{\alpha}_\theta = \boldsymbol{\alpha}$ , substituting the definition of  $\mathbf{z}$  and straightforward algebra, the ‘hat’ matrix  $H$  can be expressed in terms of the interpolation coefficients, which are obtained either through smoothed spline fitting—hence a function of  $\theta$ —or equivalent ridge regression (see equation (4)) in which case  $H$  becomes a function of  $\lambda$ ;

$$H(\boldsymbol{\alpha}) = H(\lambda) = F_1 F_1^T + F_2 U S^2 \underbrace{(U^T U)^{-1} U^T (F_2^T F_2)^{-1} F_2^T \boldsymbol{\alpha} \boldsymbol{\alpha}^T (\mathbf{f} \mathbf{f}^T)^{-1}}_{=(S^2 + N\theta I)^{-1} U^T F_2^T}. \tag{B.4}$$

## ORCID iDs

R Theunissen  <https://orcid.org/0000-0001-5844-5158>

## References

Bates D M, Lindstrom M J, Wahba G and Yandell B S 1987 Gcvpack-routines for generalized cross validation *Commun. Stat.—Simul. Comput.* **16** 263–97

Bolin M R and Meyer G W 1998 A perceptually based adaptive sampling algorithm *Proc. 25th Annual Conf. on Computer Graphics and Interactive Techniques* (New York: ACM) pp 299–309

Boyd J P and Gildersleeve K W 2011 Numerical experiments on the condition number of the interpolation matrices for radial basis functions *Appl. Numer. Math.* **61** 443–59

Carr J C, Beatson R K, McCallum B C, Fright W R, McLennan T J and Mitchell T J 2003 Smooth surface reconstruction from noisy range data *Proc. 1st Int. Conf. on Computer Graphics and Interactive Techniques in Australasia and South East Asia* (New York: ACM) p 119

- Casa L and Krueger P 2013 Radial basis function interpolation of unstructured, three-dimensional, volumetric particle tracking velocimetry data *Meas. Sci. Technol.* **24** 065304
- Chen J, Wang Y, Overbeck R S, Yong J H and Wang W 2011 Efficient depth-of-field rendering with adaptive sampling and multiscale reconstruction *Comput. Graph. Forum* **30** 1667–80
- Chen Y and Peng C 2017 Intelligent adaptive sampling guided by Gaussian process inference *Meas. Sci. Technol.* **28** 105005
- Ciccazzo A, Pillo G D and Latorre V 2014 Support vector machines for surrogate modeling of electronic circuits *Neural Comput. Appl.* **24** 69–76
- Costin W J and Allen C B 2013 Numerical study of radial basis function interpolation for data transfer across discontinuous mesh interfaces *Int. J. Numer. Methods Fluids* **72** 1076–95
- Denimal E, Nechak L, Sinou J J and Nacivet S 2016 Kriging surrogate models for predicting the complex eigenvalues of mechanical systems subjected to friction-induced vibration *Shock Vib.* **2016** 22
- Duchon J 1977 Splines minimizing rotation-invariant semi-norms in Sobolev spaces *Constructive Theory of Functions of Several Variables: Proc. of a Conf. held at Oberwolfach (25 April–1 May 1976)* (Berlin: Springer) pp 85–100
- Durantin C, Rouxel J, Désidéri J A and Glière A 2017 Multifidelity surrogate modeling based on radial basis functions *Struct. Multidiscip. Optim.* **56** 1061–75
- Dyn N, Levin D and Rippa S 1986 Numerical procedures for surface fitting of scattered data by radial functions *SIAM J. Sci. Stat. Comput.* **7** 639–59
- Fasshauer G 2007 *Meshfree Approximation Methods with Matlab* (Singapore: World Scientific)
- Forrester A I J, Sobester A and Keane A J 2008 Constructing a surrogate *Engineering Design via Surrogate Modelling—a Practical Guide* (New York: Wiley)
- Franke R 1979 *A Critical Comparison of Some Methods for Interpolation of Scattered Data* (Monterey, CA: Defense Technical Information Center)
- Gano S E, Kim H and Brown D E 2006 Comparison of three surrogate modelling techniques: Datascape, Kriging and second order regression *Proc. 11th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conf. (Portsmouth, Virginia)*
- Gao N, Li Y and Bai H 2015 Measuring large scale flow structures behind a bluff body using a hot-wire rake *Proc. Eng.* **126** 532–6
- Grimstad B, Robertson P M and Foss B 2015 Virtual flow metering using B-spline surrogate models *IFAC-PapersOnline* **48** 292–7
- Holeňa M, Linke D, Rodemerck U and Bajer L 2010 Neural networks as surrogate models for measurements in optimization algorithms *Proc. 17th Int. Conf. on Analytical and Stochastic Modeling Techniques and Applications* (Berlin: Springer) pp 351–66
- Kähler C J, Astarita T, Vlachos P P, Sakakibara J, Hain R, Discetti S, La Foy R and Cierpka C 2016 Main results of the 4th International PIV Challenge *Exp. Fluids* **57** 97
- Lavoie P, Avallone G, De Gregorio F, Romano G P and Antonia R A 2007 Spatial resolution of PIV for the measurement of turbulence *Exp. Fluids* **43** 39–51
- Li J, Yang X and Yu J 2007 Compact support thin plate spline algorithm *J. Electron.* **24** 515–22
- Li S 1995 Adaptive sampling and mesh generation *Comput. Aided Des.* **27** 235–40
- Li S 2010 Comparison of refinement criteria for structured adaptive mesh refinement *J. Comput. Appl. Math.* **233** 3139–47
- Mackman T J, Allen C B, Ghoreishi M and Badcock K 2013 Comparison of adaptive sampling methods for generation of surrogate aerodynamic models *Am. Inst. Aeronaut. Astronaut. J.* **51** 797–808
- Mackman T J and Allen C B 2010 Investigation of an adaptive sampling method for data interpolation using radial basis functions *Int. J. Numer. Methods Eng.* **83** 915–38
- MATLAB 2014 The MathWorks, Inc., Natick, Massachusetts, United States
- Mavriplis D J 1995 Unstructured mesh generation and adaptivity *26th Computational Fluid Dynamics Lecture Series* (Belgium: von Karman Institute for Fluid Dynamics)
- Meinhart C D, Wereley S T and Santiago J G 1999 PIV measurements of a microchannel flow *Exp. Fluids* **27** 414–9
- Petersen D P and Middleton D 1962 Sampling and reconstruction of wave-number-limited functions in N-dimensional euclidean spaces *Inf. Control* **5** 279–323
- Rippa S 1999 An algorithm for selecting a good value for the parameter c in radial basis function interpolation *Adv. Comput. Math.* **11** 193–210
- Schaback R 2007 *A Practical Guide to Radial Basis Functions* (Electronic Resource)
- Schaback R and Wendland H 1999 Using compactly supported radial basis functions to solve partial differential equations *Boundary Element Technology XIII* (Southampton: WitPress)
- Scheuerer M 2011 An alternative procedure for selecting a good value for the parameter C in RBF-interpolation *Adv. Comput. Math.* **34** 105–26
- Theunissen R, Kadosh J S and Allen C B 2015 Autonomous spatially adaptive sampling in experiments based on curvature, statistical error and sample spacing with applications in LDA measurements *Exp. Fluids* **56** 116
- Theunissen R, Scarano F and Riehtmuller M L 2007 An adaptive sampling and windowing interrogation method in PIV *Meas. Sci. Technol.* **18** 275
- Uddin M 2017 On the selection of a good value of shape parameter in solving time-dependent partial differential equations using RBF approximation method *Appl. Math. Modelling* **38** 135–44
- Wahba G 1990 *Spline Models for Observational Data* (Philadelphia, PA: Society for Industrial and Applied Mathematics)
- Wendland H 2004 *Scattered Data Approximation (Cambridge Monographs on Applied and Computational Mathematics)* (Cambridge: Cambridge University Press)
- Wood S N 2003 Thin plate regression splines *J. R. Stat. Soc. B* **65** 95–114
- Yu H, Jiao Y, Su Z and Reid K 2012 Performance comparison of traditional sampling designs and adaptive sampling designs for fishery-independent surveys: a simulation study *Fisheries Res.* **113** 173–81