

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

R. Theunissen<sup>1,\*</sup>, P. Gjelstrup<sup>2</sup>

1: Dept. of Aerospace Engineering, University of Bristol, BS81TR, UK

2: Dantec Dynamics, DK-2740, Denmark

\* Correspondent author: r.theunissen@bristol.ac.uk

**Keywords:** Adaptive Sampling, Radial Basis Function, scale-dependent sampling, LDA, Design of Experiments, Thin Plate Spline

## 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 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.

---

## 1. Introduction

While 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 are attained with more traditional point-wise methodologies such as e.g. hot-wire anemometry (HWA) and laser Doppler anemometry (LDA) (Lavoie *et al.*, 2007). On the other hand, because measurement volumes are typically hundreds of orders of magnitude smaller compared to the flow region of interest, such techniques require some form of traversing. 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 running costs of the experimental facility yet the more flow detail can be obtained. When faced with the task of characterising an *a-priori* unknown flow phenomenon, every experimentalist is therefore ultimately faced with the same problem: *Where to measure?*. 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: FFS). As the user refines the

sampling grid, the number of measurement points becomes rapidly impractical. It is for this reason that more intelligent sampling approaches are needed, capable of focussing, and thereby minimising, the data accrual to flow regions in which a higher number of measurements are beneficial in the flow field reconstruction and subsequently understanding, of the underlying physics.

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, 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 while rendered more efficient. 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 (DoE). As such, the proposed methodology and associated implementation is expected to benefit a wider range of research.

## 2. Surrogate modelling

In its simplest form the surrogate model in two dimensions consists of an interpolation of available data  $f(\mathbf{x}_k)$ ,  $\mathbf{x}=(x,y)$  onto the evaluation grid. Because of their simplicity and ease of implementation polynomial regression and RBF interpolation are the most common methods (Fasshauer, 2007). Given  $N$  data points ( $k=1,\dots,N$ ), the value of the surrogate model  $s(\mathbf{x}_s)$  at a new random location  $\mathbf{x}_s$  in two dimensional space is formulated through an augmented RBF model with polynomial summands as

$$s(\mathbf{x}_s) = \sum_{k=1}^N \alpha_k \psi(\|\mathbf{x}_k - \mathbf{x}_s\|) + q(\mathbf{x}_s) \quad \mathbf{x} \in \sim^2 \quad (1)$$

where  $\psi(r)$  accounts for the contribution of the  $k^{\text{th}}$  basis function in function of the radial distance  $r = \|\mathbf{x}_k - \mathbf{x}_s\|$  between the new location and the  $k^{\text{th}}$  data point. The polynomial part  $q(\mathbf{x}_s) = \sum_j \beta_j p_j(\mathbf{x}_s)$  is added to guarantee polynomial accuracy up to degree 2 and is given by  $q = \beta_1 + \beta_2 x + \beta_3 y + \beta_4 x^2 + \beta_5 y^2 + \beta_6 xy$ . Scaling coefficients  $\alpha_k, \beta_k \in \sim$  are obtained straightforward by solving Eq.1 for the known data points, i.e. the interpolation conditions  $s(\mathbf{x}_k)=f(\mathbf{x}_k)$  with the 6 additional constraints  $\sum_{k=1}^N \alpha_k p_j(\mathbf{x}_k) = 0, j = 1, \dots, 6$ . This is equivalent to solving the matrix system

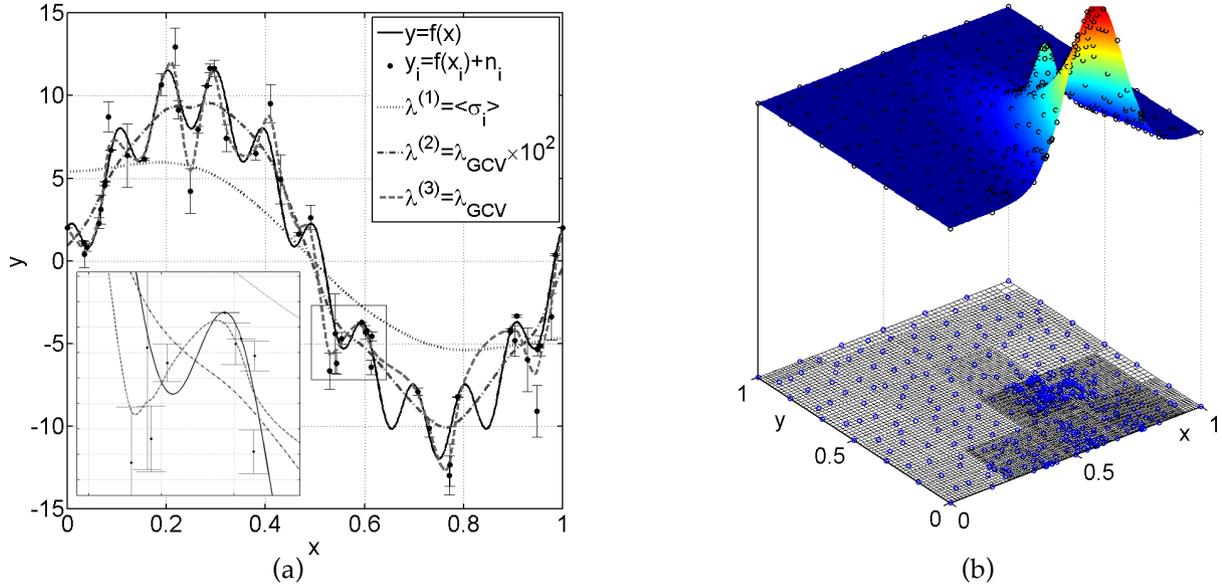
$$\begin{bmatrix} \Psi + \lambda I & P \\ P^T & O \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \mathbf{0} \end{bmatrix} \quad (2)$$

where  $\Psi_{i,j} = \psi(\|\mathbf{x}_i - \mathbf{x}_j\|)$ ,  $i, j = 1, \dots, N$ ,  $P_{il}$  contains the polynomial terms of the  $q_l$ ,  $l=1, \dots, 6$ ,  $O$  is a  $6 \times 6$  zero matrix,  $I$  is the  $N \times N$  identity matrix,  $\alpha = [\alpha_1, \dots, \alpha_N]^T$ ,  $\beta = [\beta_1, \dots, \beta_N]^T$ ,  $\mathbf{f} = [f(\mathbf{x}_1), \dots, f(\mathbf{x}_N)]^T$  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 Eq.1 to solve for the interpolation at any random location.

Thin Plate Splines (TPS) have been chosen as radial basis kernels in the current work since they do not require shape parameters (Fasshauer, 2007). However, they are inhibitive as they are intrinsically not compact (Li *et al.*, 2007) and thus involve solving large matrix systems (Wood, 2003). However, in the experimental applications considered the number of sampling locations will ordinarily lie in the order of  $10^3$  rendering the matrix system (depending on the choice of evaluation grid density) of only low to moderate complexity. More importantly though, considerable work has been done in optimizing the ridge regression parameter  $\lambda$  to accommodate splines for noisy data (Wahba, 1990). Following Duchon (1977) the appropriate degree of the polynomial term when using TPS equals  $m=1$  and  $m=2$  for TPS kernels involving 2<sup>nd</sup> (TPS2) and 3<sup>rd</sup> (TPS4) order basis functions. On uniform grids, the TPS4 has a condition number which increases as  $O(N^7)$  compared to  $O(N^4)$  for TPS2 (Boyd and Gildersleeve, 2011). This translates in ill-conditioned, and subsequently, unstable solutions for the coefficient vectors  $\alpha$  and  $\beta$  in Eq.2. To overcome this problematic, the interpolation matrix is preconditioned. 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. 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.

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 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. A degree of filtering is therefore required. The choice of  $\lambda$  when performing ridge regression must be in relation to the noise level, yet over-smoothing must be avoided as oscillations may genuinely originate from smaller scale phenomena such as e.g. turbulence. To this extent the adaptive sampling routine is nested in a primary process in which the degree of smoothing is progressively reduced Fig.1a. In the first loop, values of  $\lambda^{(0)}$  are obtained on the basis of average data scatter (Forrester *et al.*, 2008) and are updated with each new adaptively placed data sample. Once the adaptive sampling process has converged,

the second loop is initiated in which a smoothed TPS2 spline is estimated recursively using a generalised cross-validation (GCV) scheme  $\lambda_{\text{GCV}}$  as suggested and implemented by Bates *et al.* (1987).



**Fig. 1** (a) Illustration of progressive smoothing using ridge regression. The underlying signal  $f(x) = 10 \sin(2\pi x/L) + 2 \cos(20\pi x/L)$  is sampled with 50 randomly positioned points superimposed with an uncertainty  $\sigma_i = |5\gamma \cos(100\pi x/L) \sin(160\pi x/L)|$  where  $\gamma$  is a random number between 0 and 1 with uniform probability. Measurement noise  $n$  is defined as  $n_i = 3\tau$  with  $\tau$  a random number drawn from a normal probability distribution  $N(0,1)$ . The first iterative value of  $\lambda$  is based on the ensemble average  $\langle \bullet \rangle$  of data scatter. As the value of  $\lambda$  is reduced, the underlying surrogate model reflects smaller scale oscillations. The inset shows a close-up of the graph portion bounded by the rectangle. (b) Adaptive evaluation gridding when sampling Franke's bivariate test function:  $f(5(x-0.6), 5(y-0.3))$ . Hollow circles indicate adaptively placed measurement positions. As points are located more densely in regions where the underlying signal varies, the evaluation grid becomes denser.

### 3. Adaptive sampling objective function

#### 3.1 Curvature dependent sampling

To properly capture the spatial variation in regions of higher signal curvature, the first criterion  $J_c$  ensures an increased number of samples in such regions, similar to mesh-refinement heuristics in Computational Fluid Dynamics (CFD). Curvature has been estimated on the basis of the Laplacian  $\Delta$  of the RBF-based surrogate models, yielding the criterion  $J_L = \Delta s(\mathbf{x}) = \sum_{k=1}^N \alpha_k \Delta \psi(\|\mathbf{x} - \mathbf{x}_k\|) + \Delta q(\mathbf{x})$ . To compensate for the detrimental effect of stochastic noise,  $J_L$  was combined with a measure for local signal change  $J_\Delta = |s - s_{\text{MA}}|$  where  $s_{\text{MA}}$  constituted the moving average filtered response evaluated with a filter of approximately 30% domain width.

### 3.2 Uncertainty-based sampling

The second criterion ensures regions of higher uncertainty are adequately sampled. While each variable is assigned one unique value at each sampling location, these values typically correspond to the average of  $N$  independent measurements taken at each location. The number of readings  $N$  is typically set to 10,000 to retrieve reliable mean statistics or can be locally adjusted according to the local data scatter to ensure a constant relative error across the domain. Subsequently, the objective component pertaining error can be related to the local data scatter  $\sigma$ . The  $\sigma$  values at the data points can also be interpolated using TPS as per section 2 yielding a continuous function  $J_e$ .

### 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 change in the previous underlying approximant merit further exploration. This thought is captured in the improvement objective  $J_i$ , quantifying the relative additive information content,  $\Omega^{(i)} = Q^{(i)}(1 + |s^{(i)}|)^{-1}$  where  $Q^{(i)} = |s^{(i)} - s^{(i-1)}|$  quantifies the model's change at the  $i^{\text{th}}$  iteration. To retain historic model change, the improvement factor combines the four most recent  $\Omega$ 's in a weighted average;  $J_I = \left( \sum_{m=0}^3 w_{i-m} \Omega^{(i-m)} \right)_n$ . Currently all weights are equal to 0.25. Changes in surrogate model are not necessarily linked to the addition of novel data, but can be due to alterations in the ridge-regression parameter ( $\lambda$  is updated every iteration), uncertainty in the TPS due to uncertainty in the acquired data or limitations in interpolation accuracy. Only values of  $Q^{\text{th}}(\mathbf{x})$  exceeding the estimators related to the above three effects are 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  $\Omega^{\text{th}}$ .

### 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. In this work this is achieved through a separation function  $J_s(\mathbf{x})$  attaining zero values at existing measurement locations and tending to unity away from the sample location. The influence of each point should thus be radially symmetric and finite, making the compactly supported Wendland functions ideal kernels (Wendland, 2004). Since increasingly smaller scales are envisaged within the iterative process (Section 2), samples must be allowed to be positioned in closer proximity. This

can be achieved by adjusting the Wendland functions' kernel support in function of the changing fill distance and iteration in ridge regression.

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  $m^{\text{th}}$  sample point and dependent on the iteration number  $i$ ;  $w_{i,m} = 1 - \left(1 - \frac{i - j_m}{N_i}\right)_+^3$  where  $j_m$  designates the iteration number in which the  $m^{\text{th}}$  sample point was introduced. The weighting starts from 0 and gradually reaches a value of 1 after  $N_i$  iterations. Parameter  $N_i$  thus essentially dictates how fast a region can be re-sampled and has been set to  $N_i=15$  in the current study. The cut-off function is defined as  $(X)_+ = X$  for  $X \geq 0$  and 0 else. The interpolated form of  $w_{i,m}$ ,  $J_w(\mathbf{x})$  thus serves as a continuous weighting for the objective function and will avoid recently sampled regions from continuously accruing the majority of the samples in subsequent iterations, yet gradually allows further investigation of already explored domain sections.

### 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 is obtained. Each of the  $N_v$  variables will therefore have an associated  $J_c$ ,  $J_l$  and  $J_e$ . For each variable these elementary functions are combined as

$$J_{o,v}(\mathbf{x}) = (\kappa_c + J_{c,v}(\mathbf{x}))(\kappa_l + J_{l,v}(\mathbf{x}))(\kappa_e + J_{e,v}(\mathbf{x})) \quad (3)$$

The empirical offset constants  $\kappa$  were set as  $\kappa_c=0.5$ ,  $\kappa_l=0.1$  and  $\kappa_e=0.5$ . The definitive objective function is finally given by

$$J_o(\mathbf{x}) = \left( \prod_{v=1}^{N_v} J_{o,v}(\mathbf{x}) \right) \left( 1 - (J_h(\mathbf{x}))_n \right) J_w(\mathbf{x}) \quad (4)$$

## 4. Sample allocation

The evaluation grid defines the underlying nodes at which the surrogate models, and consequently the objective function defined in Eq.4, is computed. Also here adaptivity is implemented by partitioning the global domain into regular blocks and adapting local grid spacing in each block to the current sample spacing. Local grid spacing is based on the smallest nearest neighbour distance between measurement samples such that local spatial fluctuations can be adequately described following the Nyquist sampling criterion. Grid refinement is then applied

only if the the new sample spacing will yield a considerable change in number of evaluation points.

While peaks in the objective function indicate the most suitable sampling locations, such a process would rapidly lead to clustering of measurement stations. To avoid over-sampling samples are allocated based on local objective function peak geometry. Having defined the objective function, a threshold is defined based on the average of all local maxima to select potential regions for further investigation. Regions are sorted on the basis of parameter  $\Gamma$  serving as a heuristic of the local volume and height;  $\Gamma_i = \int_{\Omega_i} J_o(\mathbf{x}) d\mathbf{x} \times \max_{\mathbf{x} \in \Omega_i} (J_o(\mathbf{x}))$ . Having identified regions in which  $J_o$

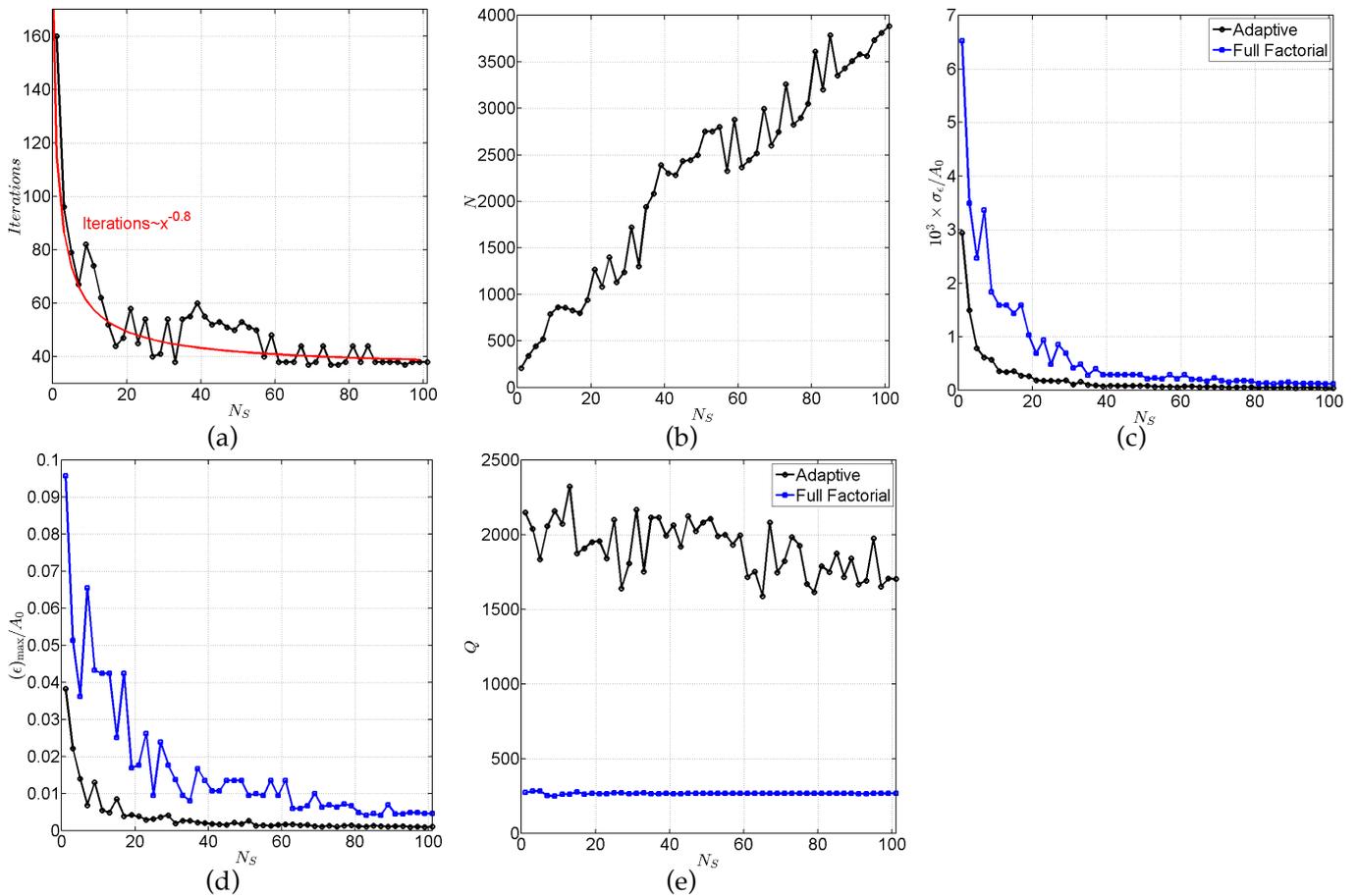
exceeds the threshold value, the  $N_s$  regions associated with the highest  $\Gamma$  values are selected, where  $N_s$  equals the number of new samples to be distributed. Starting with the region of largest  $\Gamma$ , each region is attributed  $N_{s,i} = \max\left(1, \left\lfloor \frac{\Gamma_i}{\bar{\Gamma}} \right\rfloor\right)$  with  $\bar{\Gamma} = \frac{1}{N_s} \sum_i \Gamma_i$ .

## 5. Numerical assessment

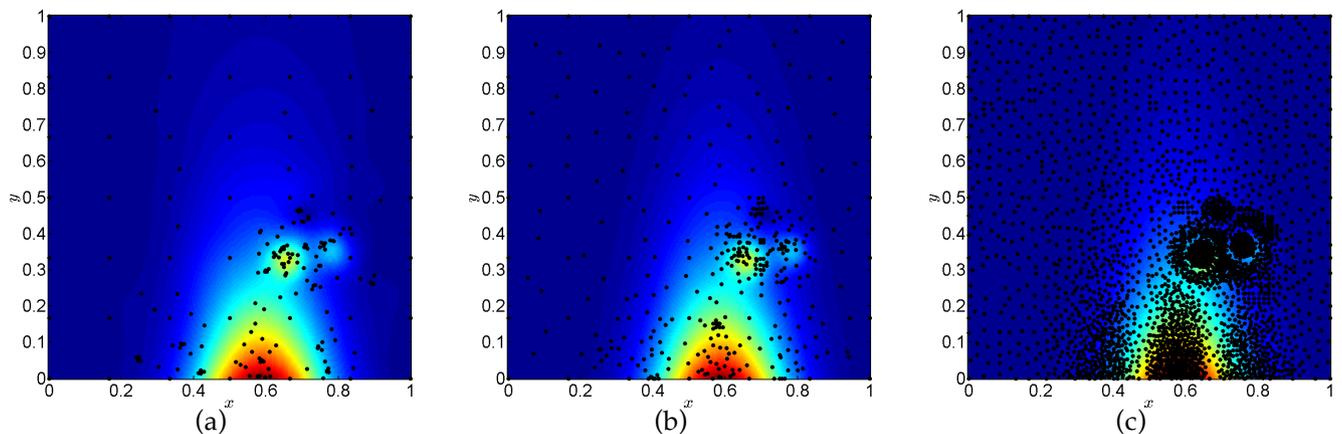
In the numerical assessment two numerical fields were considered; a modified version of Franke's function (Fig.1b) and a lid driven cavity flow at  $Re=100$ . In both cases, the initial samples were positioned on a 7-by-7 structured grid.

Fig. 2 gives an overview of the algorithm performance with respect to the traditional full factorial sampling (FFS) approach in the analysis of the modified Franke function. As the number of samples imposed per iteration  $N_s$  is increased, fewer iterations are needed to reach convergence (Fig.2a). The total number of samples increases however in a near linear fashion (Fig.2b). As a result, the maximum error and standard deviation in the error will initially decrease, reaching asymptotic values equal for both FFS and adaptive sampling. To juxtapose the "goodness" of the adaptive sampling distribution with equispaced full factorial sampling (spacing is given as  $\left[ N^{-0.5} \right]$ , a quality heuristic  $Q$  was introduced defined as  $Q = \frac{1}{N} \sum_{i=1}^N |\Delta f(\mathbf{x}_i)|$ . High values of  $Q$  indicate the sample distribution to be denser in regions of higher signal curvature, in line with curvature dependency criterion. The sampling quality is consistently higher for adaptive sampling as the samples are located in appropriate regions. This can be clearly seen in Fig.3. With  $N_s=71$  the regions of higher curvature are clearly visible through the heightened sample concentrations. However, these regions become oversampled as the interpolation accuracy improves due to the reduced sample spacing and additional samples are placed to further enhance the surrogate model (section 3.3). In view of the time required to perform the experiment, a total number of samples  $O(N) \sim 10^3$  is desirable, in which case the typical number of samples to be added each iteration should be less than 20. Comparing the error distributions (Fig.4) it can be seen that little is gained going from

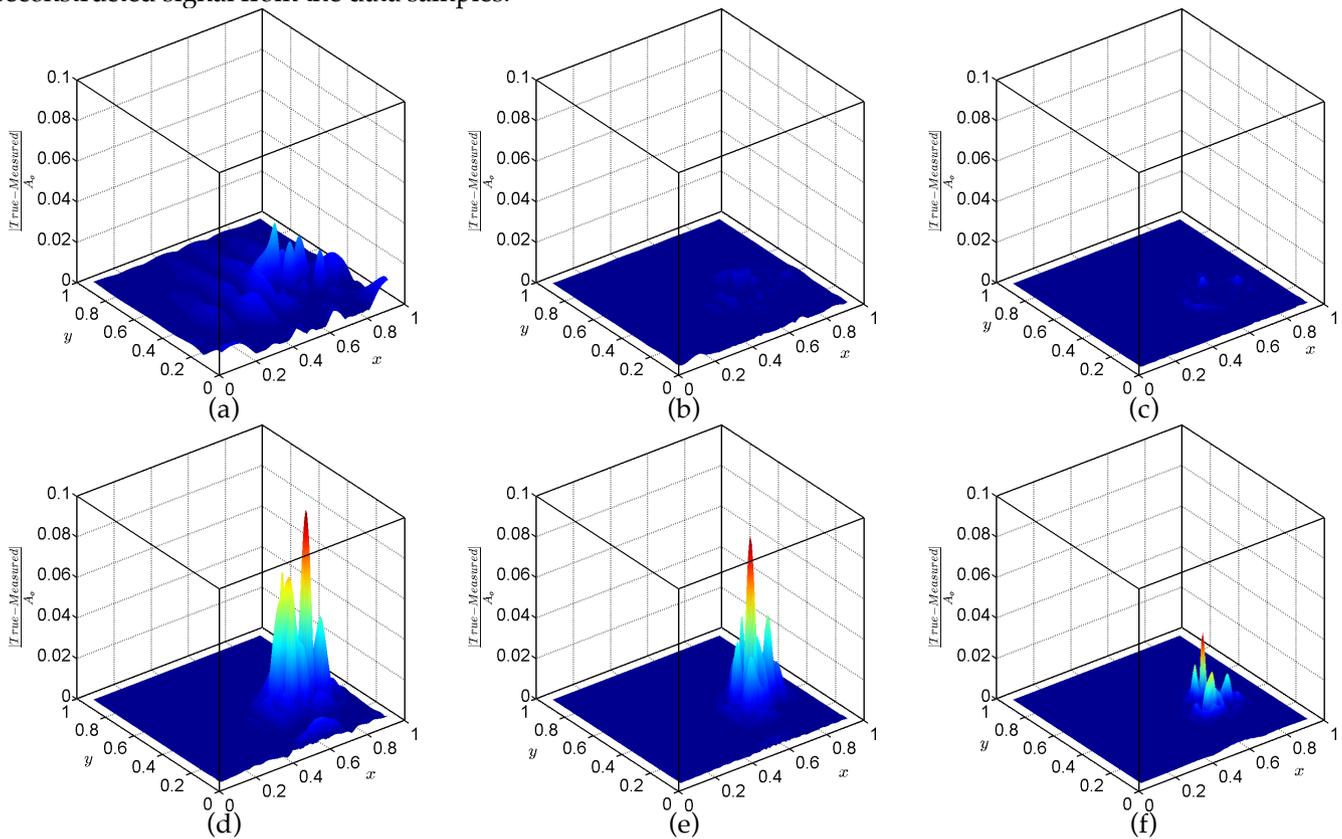
$N_s=5$  to  $N_s=11$ . Both AS and FFS have the largest errors in the regions of largest spatial variance, but already for  $N_s=5$  AS reaches error levels comparable to those achievable with  $N_s=11$ . It is such proposed that  $N_s=5$  is most suitable.



**Fig. 2** Evolution in (a) number of iterations before convergence, (b) total number of samples  $N$ , (c) rms error  $\sigma_\epsilon$  normalised with the maximum value of the underlying signal  $A_0$ , (d) normalised maximum error  $\epsilon_{\max} / A_0$  and (e)  $Q$  heuristic, with imposed samples per iteration  $N_s$  for the case of the modified Franke function (Fig.1b).



**Fig. 3** Sample distributions for the modified Franke function (Fig.1b) (a)  $N_s=1$ , (b)  $N_s=5$ , (c)  $N_s=71$ . Colors indicate the reconstructed signal from the data samples.



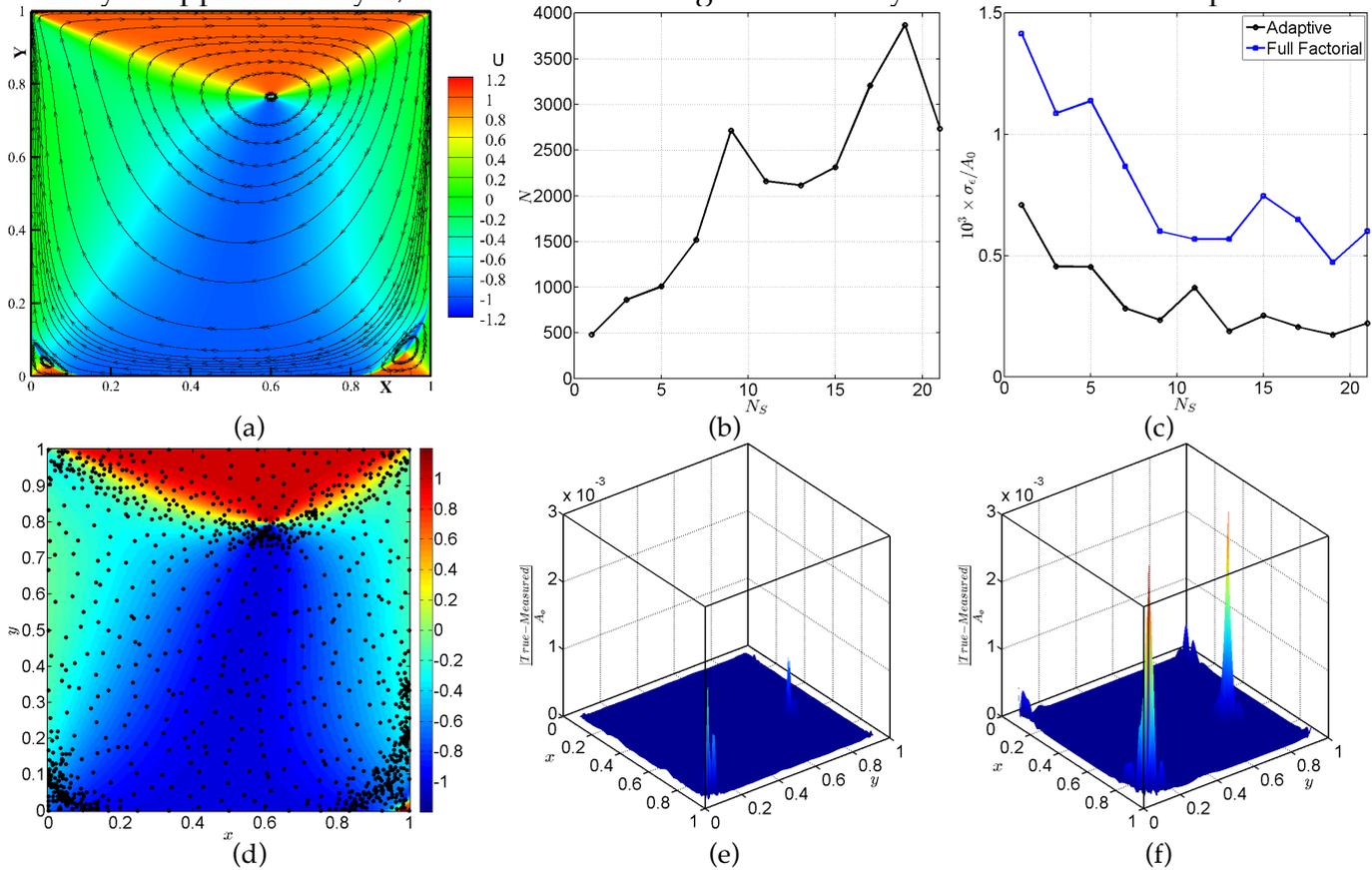
**Fig. 4** Error distribution across the modified Franke function's domain obtained with (top) adaptive sampling (bottom) equispaced sampling. (a),(d)  $N_s=1$ ; (b),(e)  $N_s=5$ ; (c),(f)  $N_s=11$ .

When applied to the lid-driven cavity flow (Fig.5a), the total number of samples rapidly increases with  $N_s$  (Fig.5b). Once again  $N_s=5$  is the most appropriate choice. Samples are distributed in regions of strong gradients and allow proper sampling of the center of the larger central recirculation region as well as the smaller corner ones (Fig.5d). It should be noted that in this test case the sampling is based on surrogate models of both the horizontal (U) and vertical (V) velocity component, although only the U component is depicted herein. Compared to the full factorial approach, the local error in the reconstructed signal thus improves (Fig.5f vs. c) leading to a reduction in global error (Fig.5c).

## 6. 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  $Re_H \sim 37 \times 10^3$ . Experiments were performed in the open-return low-speed wind

tunnel of the University of Bristol at a flow speed of approximately 10.33m/s and a turbulence intensity of approximately 2%. 2D2C Particle Image Velocimetry measurements were performed



**Fig. 5** (a) Illustration of the lid-driven cavity flow field at  $Re=100$ . The colors relate to the horizontal velocity component. Evolution in (b) total number of samples  $N$  and (c) standard deviation in reconstruction error with number of imposed samples per iteration  $N_s$ . (d) Distribution of samples imposing  $N_s=5$ . Colors refer to the reconstructed horizontal velocity component. Spatial distribution in the error in the horizontal displacement component as per (e) adaptive sampling (f) equispaced sampling, imposing  $N_s=5$ , yielding  $N \sim 900$  - see (b).

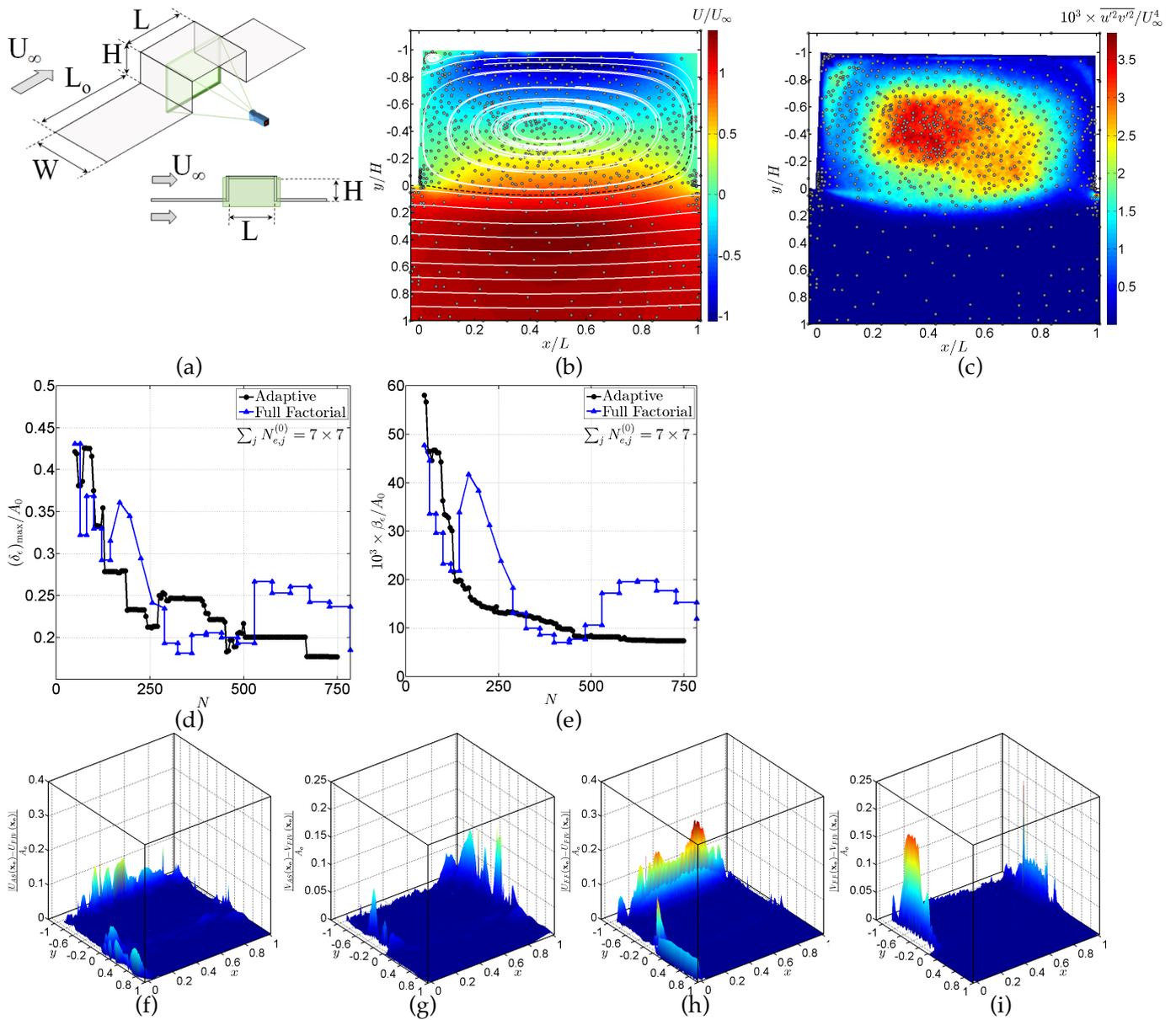
on a cavity 110mm long and 55mm deep using a Dantec Dynamics system consisting of a 200mJ Litron laser and 4MP FlowSense EO camera. A sketch of the setup and exemplary PIV image illustrating the location of the cavity walls is provided in Fig.6a. The 2072 \times 2072 pixels sensor corresponded to a field of view of 115 \times 115 mm<sup>2</sup> and images were analysed using initial interrogation windows of 65 \times 65 pixels<sup>2</sup>, recursively reduced to 17 \times 17 pixels<sup>2</sup>. A mutual window overlap of 75% yielded a final vector spacing of approximately 0.2mm. In total 1000 instantaneous velocity fields were considered in the calculation of the mean and fluctuating velocity components,  $\bar{u}$  and  $\bar{v}$  respectively, where  $\bar{u} = \frac{1}{10^3} \sum_{q=1}^{10^3} u_q(\mathbf{x})$ . 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 0.1mm. In terms of the unit domain, this implies that sample locations must be located on multiples of  $0.1/115$ . 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 (Eq.4) to zero within the corresponding regions as evidenced when overlaying the FOV with the predicted sampling locations (Fig.6b).

The number of consecutive iterations to establish convergence in each of the smoothing phases through ridge regression was automatically set to 16, 15 and 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.  $J_e$ . This is manifested when samples are superimposed onto the product of local Reynolds stresses  $\overline{u^2} \times \overline{v^2}$  (Fig.6c). Whereas the adaptive approach yields a consistent reduction in maximum error with increasing number of samples (Fig.6d), the standard approach shows stronger fluctuations. Not only does AS attain error levels inferior to those of FF (Fig.6d,e), 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 (Fig.6f-i). 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



**Fig. 6** (a) Sketch of the open cavity experimental setup ( $L_c=200\text{mm}$ ,  $H=55\text{mm}$ ,  $L=110\text{mm}$ ,  $W=200\text{mm}$ ) at  $Re_{in}=37\times 10^3$ . The green square indicates the PIV field-of-view. (b) 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  $(x/L, y/H)=(0,0)$ . (c) Contours of the product between the time-averaged fluctuations in the horizontal and vertical velocity component. Evolution in (d) maximum and (e) rms error with total number of samples. Error distribution obtained with adaptive sampling in (f) horizontal and (g) vertical velocity component compared with PIV. Error distribution as per full factorial sampling in (h) U and (i) V component.

## 7. 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.

## Acknowledgement

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

## References

- C. J. Kähler, T. Astarita, P. P. Vlachos, J. Sakakibara, R. Hain, S. Discetti, R. La Foy, C. Cierpka, (2016) Main results of the 4th international PIV challenge, *Experiments in Fluids* 57 97.
- P. Lavoie, G. Avallone, F. De Gregorio, G. P. Romano, R. A. Antonia, Spatial resolution of PIV for the measurement of turbulence, (2007) *Experiments in Fluids* 43 39-51.

- R. Theunissen, J. S. Kadosh, C. B. Allen, Autonomous spatially adaptive sampling in experiments based on curvature, statistical error and sample spacing with applications in LDA measurements, (2015) *Experiments in Fluids* 56.
- G. Fasshauer, *Meshfree Approximation Methods with MATLAB*, (2007) *Interdisciplinary mathematical sciences*, World Scientific.
- J. Li, X. Yang, J. Yu, Compact support thin plate spline algorithm, (2007) *Journal of Electronics (China)* 24 515-522.
- S. N. Wood, Thin plate regression splines, *Journal of the Royal Statistical Society: Series B* (2003) *Statistical Methodology* 65 95-114.
- G. Wahba, *Spline Models for Observational Data*, (1990) *Society for Industrial and Applied Mathematics*, Philadelphia.
- J. Duchon, Splines minimizing rotation-invariant semi-norms in Sobolev spaces, *Springer Berlin Heidelberg*, Berlin, Heidelberg, pp. 85-100.
- J. P. Boyd, K. W. Gildersleeve, Numerical experiments on the condition number of the interpolation matrices for radial basis functions, (2011) *Applied Numerical Mathematics* 61 443-459.
- N. Dyn, D. Levin, S. Rippa, Numerical procedures for surface fitting of scattered data by radial functions, (1986) *SIAM Journal on Scientific and Statistical Computing* 7 639-659.
- A. I. J. Forrester, A. Sbester, A. J. Keane, *Constructing a Surrogate*, John Wiley and Sons Ltd, pp. 33-76.
- D. M. Bates, M. J. Lindstrom, G. Wahba, B. S. Yandell, Gcvpack routines for generalized cross validation, (1987) *Communications in Statistics - Simulation and Computation* 16 263-297.
- H. Wendland, *Scattered Data Approximation*, (2004) *Cambridge Monographs on Applied and Computational Mathematics*, Cambridge University Press.