A compressed sensing approach for partial differential equations with random input data

L. Mathelin [a,b,*] & K.A. Gallivan [b]

[a] *LIMSI-CNRS, BP 133, 91403 Orsay, France.*

[b] *Mathematics Dpt., 208 Love Building, 1017 Academic Way, Florida State University, Tallahassee FL 32306-4510, USA.*

Abstract

In this paper, a novel approach for quantifying the parametric uncertainty associated with a stochastic problem output is presented. As with Monte-Carlo and stochastic collocation methods, only point-wise evaluations of the stochastic output response surface are required allowing the use of legacy deterministic codes and precluding the need for any dedicated stochastic code to solve the uncertain problem of interest. The new approach differs from these standard methods in that it is based on ideas directly linked to the recently developed compressed sensing theory. The technique allows the retrieval of the modes that contribute most significantly to the approximation of the solution using a minimal amount of information. The generation of this information, *via* many solver calls, is almost always the bottle-neck of an uncertainty quantification procedure. If the stochastic model output has a reasonably compressible representation in the retained approximation basis, the proposed method makes the best use of the available information and retrieves the dominant modes. Uncertainty quantification of the solution of both a 2-D and 8-D stochastic Shallow Water problem is used to demonstrate the significant performance improvement of the new method, requiring up to several orders of magnitude fewer solver calls than the usual sparse grid-based Polynomial Chaos (Smolyak scheme) to achieve comparable approximation accuracy.

Key words: Uncertainty quantification, Compressed sensing, Collocation technique, Stochastic spectral decomposition, Smolyak sparse approximation, Stochastic collocation.

* Corresponding author. Tel: 33-169 85 80 69; fax: 33-169 85 80 88.
Email addresses: mathelin@limsi.fr (L. Mathelin), gallivan@math.fsu.edu (K.A. Gallivan).

1 November 2010

1 Introduction

Uncertainty quantification has become a major concern for a wide range of communities. Indeed, in addition to providing accurate results, many simulation codes are now also expected to account for uncertainty in some of the intrinsic parameters of the problem and to provide confidence intervals and statistics of the outputs. Two basic types of uncertainty can be distinguished. Aleatory uncertainty may arise from the intrinsic variability of a physical quantity, *e.g.*, radioactive disintegration. The second type of uncertainty, referred to as the epistemic uncertainty, arises from a lack of knowledge of the considered quantity. In contrast to the aleatory uncertainty, the epistemic uncertainty may be reduced with additional knowledge on the quantity. The uncertain parameters may be initial or boundary conditions, geometric settings, constitutive material physical properties, etc., and their variability is suitably modeled using random variables. Specific methods must be used to infer the resulting uncertainty of the simulation outputs and provide statistical information such as mean, variance, quantiles, correlations, statistical moments or probability density functions of some quantities of interest, usually a functional of the simulation outputs. The probabilistic approach is a natural framework to achieve these objectives. While the original uncertain problem is sometimes of infinite dimension, reasonably accurate modeling often allows approximating the uncertainty sources with a finite set of real-valued random variables, for instance using a spectral decomposition technique, opening a route for a tractable computational solution method.

Indisputably, the most widely used approach to quantify the uncertainty associated with the solution of an uncertain problem is the Monte-Carlo approach. The probabilistic space is sampled and the associated deterministic problem is solved. From the collection of solutions arising from the N_{MC} samples, statistical information is derived. Several specific features explain the success of the Monte-Carlo approach. The main one is that the method relies only on the solution of deterministic problems, each solved for a given set of deterministic input parameters, avoiding the need for a dedicated uncertainty quantification-oriented code and allowing the use of legacy, well-validated and certified, deterministic codes that are used as a black-box. Further, the samples being drawn independently, it is embarrassingly straightforward to carry the N_{MC} simulations in parallel. The method is very general and robust and does not rely on assumptions on the solution. This robustness and simplicity come with a price that is most apparent in the poor $\mathcal{O}\left(N_{MC}^{-1/2}\right)$ convergence rate. Although numerous variants of the original Monte-Carlo method have been proposed, modifying the functional evaluated (Importance Sampling) or the way independent samples are generated (quasi-Monte-Carlo, Stratified Sampling, etc.), the convergence rate remains unchanged, with only the associated constant improved. This low convergence rate leads to requiring an

unacceptably large number of simulations to compute reasonably converged statistics, precluding the use of Monte-Carlo methods in cases the deterministic simulation computational time is large. However, in contrast with other methods, the $\mathcal{O}\left(N_{MC}^{-1/2}\right)$ convergence rate is insensitive to the stochastic dimension of the uncertainty sources, making the Monte-Carlo approach the method of choice for uncertain problems involving a very large number of independent uncertainty sources. Therefore, unless the uncertain problem of interest involves uncertainty sources of very large stochastic dimensions, the Monte-Carlo method is not usually suitable in practice and alternative uncertainty quantification methods are more appropriate.

Among these alternative approaches to approximate finite variance quantities of interest such as those considered in this paper, the spectral stochastic method approximates on a suitable functional expansion basis. This approach dates back to the pioneering work of Wiener (Wiener, 1938) but has emerged as a widely used tool since the book of Ghanem & Spanos was published, Ghanem and Spanos (1991). Since we restrict ourselves to second order random variables, *i.e.*, finite variance, it is suitable to consider the L^2-space associated with the random variables. The Polynomial Chaos approach exploits any regularity of the solution and consists of deriving a functional representation of a quantity of interest on a stochastic basis spanned by Hermite polynomial functionals. These polynomials are orthogonal w.r.t. the measure associated with a Gaussian random variable and span the stochastic space of finite variance random variables, thus defining a complete basis in the stochastic space. The approximation was proved to converge for any finite variance random variable (Cameron and Martin, 1947). An extension to bases generated by other polynomial functionals has been proposed by Xiu and Karniadakis (2002, 2003). For instance, Legendre polynomials, associated with uniformly distributed random variables, can be used. In Soize and Ghanem (2004), a generalization to bases spanned by functionals associated with random variables of arbitrary measure was proposed. Improvements to the method have taken advantage of the flexibility in choosing trial functionals to approximate the stochastic solution. In particular, several approaches have relied on refinement of the approximation by varying the support and/or the polynomial order of localized bases, in direct relation with the well-known *hp*-spectral scheme in the deterministic discretization framework, see for instance Wan and Karniadakis (2005). Following similar ideas, Wiener-Haar wavelets (Le Maître et al., 2004a) and Multi-Resolution schemes (Le Maître et al., 2004b) have been used while Mathelin and Le Maître (2007) have employed an *a posteriori* error analysis strategy to adaptively refine the approximation in the stochastic space.

Beyond the choice of the trial functions, two classes of methods may be distinguished in the (generalized) Polynomial Chaos approach by the way the deterministic coefficients of the resulting expansion are evaluated. Basically, they may be computed through a direct evaluation, using techniques such as

projection, regression or interpolation, or through a Galerkin procedure. In the Galerkin approach, the residual of the model equation is required to lie in a space orthogonal to the trial basis space. The problem then takes the form of a $P_{\boldsymbol{\xi}}$-coupled-equation problem, $P_{\boldsymbol{\xi}}$ being the number of unknown coefficients in the stochastic spectral expansion. This approach relies on solid mathematical grounds and error estimators as well as proofs of well-posedness exist. However, the coupled character of the resulting problem may constitute a limitation for problems that are large at the deterministic level and requiring a large number $P_{\boldsymbol{\xi}}$ of stochastic modes for an acceptable approximation. Further, deterministic codes may not be used as such and need be deeply reworked for this formulation, hence the term "intrusive" to refer to the approach.

Alternatively, the coefficients may be evaluated by directly computing the integrals involved in their definition. Typical numerical techniques to achieve this are based either on fully tensorized or sparse quadrature rules. As with the Monte-Carlo approach, this allows the use of deterministic solvers only and does not involve coupled problems. Further, aliasing effects are avoided and only the approximation error is present. However, these nice properties are somewhat counterbalanced by the large number of evaluation points required to compute the solution expansion coefficients due to the so-called curse of dimensionality, even when sparse quadrature rules such as the Smolyak scheme (Smolyak, 1963; Novak and Ritter, 1999) are employed. Just as for the Galerkin flavor of the Polynomial Chaos method, different strategies have been proposed to take advantage of anisotropy of the solution, if any. To this end, anisotropic sparse grid schemes have been proposed and shown to potentially significantly reduce the number of required deterministic solver calls, Nobile et al. (2007); Ganapathysubramanian and Zabaras (2007). However, these approaches rely either on *a priori* estimates which are only known in a limited number of specific cases, or on *a posteriori* estimates evaluated through an incremental trial-and-error sequence. Such an incremental procedure sequentially enriches the approximation space along the directions most contributing to the decrease of the error but constitutes a *bottom-up* technique, where the solution is approximated on a basis whose spectrum usually sequentially grows from low towards higher frequencies. If the solution is essentially monochromatic at a high frequency, these solution techniques lead to an unnecessarily large number of evaluation points, significantly increasing the overall computational time.

In this paper, it is proposed that an approximation arising from any inherent compressibility of the solution in the trial basis can be used as the foundation for an effective and efficient method for uncertainty quantification. Relying on the hypothesis that the unknown stochastic solution is reasonably compressible, a method is presented, heavily relying on the compressed sensing theory (Candès and Tao, 2004; Donoho, 2006), that determines the most significant modes for the approximation and discards the others. The resulting

required number of solver calls can be significantly reduced compared to the usual sparse grid techniques.

During the final preparation of this manuscript, the authors came to know about the very recent work by Doostan & Owhadi where similar ideas are developed. Both works were presented at the 2010 SIAM Annual Meeting, Doostan and Owhadi (2010) (and an article submitted to *J. Comput. Phys.*) and Mathelin and Gallivan (2010).

The paper is organized as follows. In Section 2, the stochastic framework is defined and some relevant issues emphasized. The core theoretical ingredients of the compressed sensing (CS) methodology are given in Section 3 and application of conceptually similar ideas to the uncertainty quantification (UQ) framework is discussed and presented in Section 4. Then, the resulting UQ method is demonstrated on a test case based on a Shallow Water problem. Section 5 briefly presents the problem together with the solution method. Results are shown in Section 6 and are discussed for a 2-D and a 8-D stochastic problem formulation, in particular in terms of approximation accuracy for a given number of solver calls. Concluding remarks close the paper in Section 7.

2 Quantifying the parametric uncertainty

2.1 Stochastic problem framework

The essence of the parametric uncertainty propagation and quantification issue is to infer the statistics of the solution $u(\theta)$ of a mathematical model from those of its associated uncertain input parameters $D(\theta)$. The problem is conveniently treated in a probabilistic framework. Specifically, defining a probability space $(\Theta, \mathcal{B}_\Theta, \mu_\Theta)$, where Θ is the space of elementary events θ, $\mathcal{B}_\Theta \subset 2^\Theta$ an associated σ-algebra defined on Θ and μ_Θ a probability measure, the problem can conceptually be expressed in the form of the following equation which holds μ_Θ-almost surely:

$$\mathcal{F}\left(u(\theta); D(\theta)\right) = 0, \quad \mu_\Theta - a.e., \tag{1}$$

where, without loss of generality, the mathematical model $\mathcal{F}$ is deterministic and all involved uncertainty sources have been gathered in the set of input parameters $D(\theta)$.

The original problem is conveniently modeled in terms of random variables in a finite dimensional image probability space $(\Xi, \mathcal{B}_\Xi, \mu_\Xi)$, where $\boldsymbol{\xi}(\theta) \in \Xi \subset \mathbb{R}^{N_{\boldsymbol{\xi}}}$ is a vector-valued random variable, $\mathcal{B}_\Xi$ and μ_Ξ the associated σ-algebra and

probability measure of the image probabilistic space respectively. The problem may be reformulated as:

$$\mathcal{F}\left(u\left(\boldsymbol{\xi}\left(\theta\right)\right); D\left(\boldsymbol{\xi}\left(\theta\right)\right)\right) = 0, \quad \mu_{\Xi} - a.e.. \tag{2}$$

The problem then takes the form of approximating the real-valued functional $u\left(\boldsymbol{\xi}(\theta)\right)$ of interest. Let us consider the problem involving finite variance (*i.e.* second order) real-valued random quantities. This naturally leads us to introduce the corresponding functional space $\mathcal{V}_{\Xi} \equiv L^2\left(\Xi, \mu_{\Xi}\right)$:

$$L^2\left(\Xi, \mu_{\Xi}\right) \equiv \left\{ v : \Xi \ni \boldsymbol{\xi}(\theta) \mapsto v\left(\boldsymbol{\xi}(\theta)\right), \int_{\Xi} v^2\left(\boldsymbol{s}\right)\, d\mu_{\Xi}(\boldsymbol{s}) < +\infty \right\}. \tag{3}$$

Let $\left\langle \cdot, \cdot \right\rangle_{L^2(\Xi,\mu_{\Xi})}$ be the natural inner product associated with the stochastic space:

$$\left\langle v, v' \right\rangle_{L^2(\Xi,\mu_{\Xi})} \equiv \int_{\Xi} v\left(\boldsymbol{s}\right)\, v'\left(\boldsymbol{s}\right)\, d\mu_{\Xi}(\boldsymbol{s}), \quad \forall \left\{v, v'\right\} \in L^2\left(\Xi, \mu_{\Xi}\right), \tag{4}$$

and $\mathcal{V}_{\Xi}$ is thus a Hilbert space so that tools from the approximation theory may be used. In particular, the quantity of interest u may be decomposed as:

$$u\left(\boldsymbol{\xi}(\theta)\right) \approx \sum_{\boldsymbol{\alpha} \in \mathcal{J}} X_{\boldsymbol{\alpha}}\, \psi_{\boldsymbol{\alpha}}\left(\boldsymbol{\xi}(\theta)\right), \tag{5}$$

where $\psi_{\boldsymbol{\alpha}}\left(\boldsymbol{\xi}(\theta)\right)$ belongs to a complete set of orthogonal functions defining an Hilbertian basis:

$$\left\langle \psi_{\boldsymbol{\alpha}}, \psi_{\boldsymbol{\alpha}'} \right\rangle_{L^2(\Xi,\mu_{\Xi})} = \left\langle \psi_{\boldsymbol{\alpha}}, \psi_{\boldsymbol{\alpha}} \right\rangle_{L^2(\Xi,\mu_{\Xi})}\, \delta_{\boldsymbol{\alpha}\boldsymbol{\alpha}'}, \quad \forall \left\{\boldsymbol{\alpha}, \boldsymbol{\alpha}'\right\} \in \mathcal{J} \times \mathcal{J}, \tag{6}$$

$$\left\langle u, \psi_{\boldsymbol{\alpha}'} \right\rangle_{L^2(\Xi,\mu_{\Xi})} = 0,\ \forall u \in L^2\left(\Xi, \mu_{\Xi}\right),\ \forall \boldsymbol{\alpha}' \in \mathcal{J} \quad \Rightarrow \quad u \equiv 0, \tag{7}$$

with $\delta_{\boldsymbol{\alpha}\boldsymbol{\alpha}'}$ the Kronecker delta, $\boldsymbol{\alpha}^T \equiv \left(\alpha_1 \ldots \alpha_{N_{\boldsymbol{\xi}}}\right)$ and $\mathcal{J}$ the set of the $\mathbb{N}^{N_{\boldsymbol{\xi}}}$-valued multi-indexes $\boldsymbol{\alpha}$ such that $|\boldsymbol{\alpha}| \leq N_o$, $|\boldsymbol{\alpha}| \equiv \sum_{n=1}^{N_{\boldsymbol{\xi}}} \alpha_n$. Its cardinality is $|\mathcal{J}| = P_{\boldsymbol{\xi}}$. The set $\{\psi_{\boldsymbol{\alpha}}\}$ is then a family of $N_{\boldsymbol{\xi}}$-D orthogonal polynomials of total degree $\leq N_o$.

To derive an approximated description of the solution, one is left with the unknowns $X_{\boldsymbol{\alpha}}$, $\boldsymbol{\alpha} \in \mathcal{J}$, to estimate. As briefly mentioned in Section 1, a Galerkin technique may be utilized to derive a set of $P_{\boldsymbol{\xi}}$, possibly non-linear, coupled equations to be solved for $X_{\boldsymbol{\alpha}}$. These equations are of the form:

$$\begin{aligned} &\left\langle \mathcal{F}\left(\sum_{\boldsymbol{\alpha} \in \mathcal{J}} X_{\boldsymbol{\alpha}}\, \psi_{\boldsymbol{\alpha}}\left(\boldsymbol{\xi}(\theta)\right); \sum_{\boldsymbol{\alpha}'' \in \mathcal{J}} D_{\boldsymbol{\alpha}''}\, \psi_{\boldsymbol{\alpha}''}\left(\boldsymbol{\xi}(\theta)\right)\right), \psi_{\boldsymbol{\alpha}'} \right\rangle_{L^2(\Xi,\mu_{\Xi})} = 0, \\ &\forall\, \psi_{\boldsymbol{\alpha}'} \in \mathcal{V}_{\boldsymbol{\xi}},\ \boldsymbol{\alpha}' \in \mathcal{J}. \end{aligned} \tag{8}$$

This approach allows us to rely on solid, demonstrated, mathematical results. In particular, both *a priori* and *a posteriori* approximation error results are available, Deb et al. (2001); Frauenfelder et al. (2005); Mathelin and Le Maître (2007). However, it often leads to large, non-linear, systems of equations whose solution method may be a challenge. Further, it necessitates dedicated codes, precluding the direct use of any legacy, well validated and certified, code solving the deterministic problem at hand.

Alternative approaches directly estimate the unknown coefficients $X_{\boldsymbol{\alpha}}$: from Eq. (5) and making use of the basis functions orthonormality:

$$X_{\boldsymbol{\alpha}} = \langle u, \psi_{\boldsymbol{\alpha}} \rangle_{L^2(\Xi,\mu_\Xi)}, \quad \forall\, \psi_{\boldsymbol{\alpha}} \in \mathcal{V}_{\boldsymbol{\xi}},\ \boldsymbol{\alpha} \in \mathcal{J}, \tag{9}$$

where, without loss of generality, the basis functions are hereafter normalized: $\langle \psi_{\boldsymbol{\alpha}}, \psi_{\boldsymbol{\alpha}'} \rangle_{L^2(\Xi,\mu_\Xi)} = \delta_{\boldsymbol{\alpha}\boldsymbol{\alpha}'}$. Since u is not known in closed form, the integral involved in the projection is conveniently estimated through a discrete quadrature strategy:

$$\begin{aligned} X_{\boldsymbol{\alpha}} &= \int_\Xi u\left(\boldsymbol{s}\right)\, \psi_{\boldsymbol{\alpha}}\left(\boldsymbol{s}\right)\, d\mu_\Xi(\boldsymbol{s}), \\ &\approx \sum_{q \in \mathcal{N}_q} u\left(\boldsymbol{\xi}^{(q)}\right)\, \psi_{\boldsymbol{\alpha}}\left(\boldsymbol{\xi}^{(q)}\right)\, w^{(q)}, \end{aligned} \tag{10}$$

where $\mathcal{N}_q$ is the set of quadrature points $\boldsymbol{\xi}^{(q)} \in \Xi$ and $w^{(q)} \in \mathbb{R}$ their associated weights. This approach only requires the evaluation of the model output u at the quadrature points, which is provided by the deterministic solver. However, while appealing, this approach suffers from the lack of rigorous error estimator and from the curse of dimensionality which leads to an intractable number of necessary quadrature points when the stochastic dimension and/or the approximation order increases. While sparse grid schemes may help in reducing the number of required points w.r.t. fully tensorized quadrature rules, the unfavorable scaling with the dimension and the order precludes the use of this so-called *non-intrusive spectral projection* (NISP) approach for large-scale problems requiring a large computational time to evaluate u for a given point $\boldsymbol{\xi}^{(q)}$.

2.2 Motivation for a sparse approach

Let us examine the problem defined by Eq. (10) more closely and consider a sparse grid scheme to evaluate the high-dimensional integral. Classical quadrature rules are known to exactly integrate a polynomial integrand up to a certain maximum degree. For instance, assuming u is a N_o-total degree polynomial and that the test function $\psi_{\boldsymbol{\alpha}}$ is also a N_o-total degree polynomial,

Table 1 shows the maximum N_o that leads to an exact integration in a 2-D ($N_{\boldsymbol{\xi}} = 2$) space using a Smolyak quadrature scheme of varying levels l_S, Smolyak (1963); Petras (2001).

l_S	1	2	3	4	5	6	7	8	9	10	11	12
N_q	5	9	17	33	33	65	97	97	161	161	161	257
N_o	1	2	3	5	5	6	8	8	11	11	11	12

Table 1
Correspondence between Smolyak quadrature rule level l_S, number of evaluation points $N_q = |\mathcal{N}_q|$ and maximum polynomial total degree N_o. 2-D space: $N_{\boldsymbol{\xi}} = 2$.

As expected, the higher the polynomial order of the integrand, the more evaluation points required to compute the exact projection. This has connections with the celebrated Shannon-Nyquist theorem. However, the key is to note that this constitutes a *worst-case scenario*. For a given Smolyak level l_S, the projection computation defined by Eq. (10) is exact for *every* polynomial, provided its total order is lower or equal to N_o. For instance, if one is interested in a projection of u, assumed a given monomial of degree N_o, over a specific mode $\psi_{\boldsymbol{\alpha}}$ of order N_o, properties of the monomial, *e.g.*, (anti-)symmetry, may be exploited to prevent unnecessary solution evaluations. There is thus no need for as many points as expected from Table 1 and fewer evaluation points are sufficient to get the *exact* projection. In a more general framework, if the output at hand has some specific properties (monomial, symmetry, sparsity, ...), they can be exploited to limit the amount of information one needs to approximate it within a particular basis.

The concepts underlying this remark are at the root of the present work. Indeed, most model outputs u of practical interest are not white signals and can be considered to have some degree of sparsity in the orthogonal basis used in the approximation. As a result, the amount of relevant information one must capture to adequately approximate a signal, or model output, is actually often lower than what would be expected from considerations based simply on the cardinality of the trial basis.

A natural desire arises to use a number of evaluation points sufficient to estimate the most significant modes of the approximation, *and only them*. For a sparse output in the given trial basis, this could translate mathematically in looking for the minimal set of modes $\mathcal{J}^\star \subseteq \mathcal{J}$ such that the approximation matches the output at the evaluation points when N_q is sufficiently large:

$$\mathcal{J}^\star \equiv \arg\min_{\mathcal{J}} |\mathcal{J}|, \quad s.t. \quad u\left(\boldsymbol{\xi}^{(q)}\right) = \sum_{\boldsymbol{\alpha}\in\mathcal{J}} X_{\boldsymbol{\alpha}}\, \psi_{\boldsymbol{\alpha}}\left(\boldsymbol{\xi}^{(q)}\right), \quad \forall\, q \in \mathcal{N}_q. \tag{11}$$

The minimal set condition may be conveniently reformulated as a requirement

on the vector of coefficients $\boldsymbol{X} \equiv (X_{\boldsymbol{\alpha}}, |\boldsymbol{\alpha}| \leq N_o)^T \in \mathbb{R}^{P_\xi}$:

$$\boldsymbol{X}^\star \equiv \arg\min_{\boldsymbol{X}} \|\boldsymbol{X}\|_0, \quad s.t. \quad u\left(\boldsymbol{\xi}^{(q)}\right) = \Psi\left(\boldsymbol{\xi}^{(q)}\right) \boldsymbol{X}, \quad \forall\, q \in \mathcal{N}_q, \tag{12}$$

with $\Psi \equiv (\psi_{\boldsymbol{\alpha}}, |\boldsymbol{\alpha}| \leq N_o)$ the retained trial basis and $\|\cdot\|_0$ the "L^0-norm", $\|\boldsymbol{X}\|_0 \equiv \{k : X_k \neq 0\}$. Thanks to the L^0-norm, the above constrained optimization problem tends to determine an approximation with as few non-zero terms as possible in the vector of coefficients $\boldsymbol{X}$.

Formulated as such, this UQ-problem is related to concepts which been extensively studied in the signal processing community, see, for example, Taylor et al. (1979) or Chen et al. (1999), and we will build upon some of their results to place the proposed uncertainty quantification approach in a Compressed Sensing framework.

3 The compressed sensing approach

3.1 From L^0 to L^1

Finding the sparsest, yet accurate, approximation of a signal u from a formulation similar to Eq. (12) has been a subject of interest for decades. Greedy algorithms have been proposed to build-up a K-sparse solution, *e.g.*, Matching Pursuit, Mallat and Zhang (1993). However, problem (12) relies on a non-convex objective function and the underlying optimization problem is NP-hard, requiring exhaustive searches (in fact combinatorial) over all subsets of $\boldsymbol{X}$. Such an approach would suffer from the curse of dimensionality and is not thought to be a computationally tractable route. Fortunately, some recent results have shown that the L^0-norm can be replaced by the (convex) L^1-norm in Eq. (12), the solution of which then still tends to separate the most significant elements of $\boldsymbol{X}$ from those that are negligible. Significant modes are strengthened while the others tend to be discarded and their magnitude set close to zero. An example of a method relying on this approach is a reformulation of the well known LASSO, Tibshirani (1996).

3.2 Core principles

In the last few years, so-called Compressed Sensing theory, Chen et al. (1999); Candès and Tao (2004); Candès et al. (2005); Donoho (2006); Candès and Romberg (2006), has attracted growing interest. It basically builds upon the convexified L^1-norm version of Eq. (12) and considers the projection of the

constraint $u = \Psi \boldsymbol{X}$ onto a basis Φ. Under appropriate choices, the sparsest $\boldsymbol{X}$ giving rise to the observations is retrieved through a computationally tractable algorithm. In a nutshell, the core principles are as follows. Let us consider the redundant dictionary formed by both a $\{\psi_{\boldsymbol{\alpha}}\}$ and $\{\phi_m\}$ set of functions and suppose a discrete signal $f \in \mathbb{R}^N$ is sparse in $\{\psi_{\boldsymbol{\alpha}}\}$ while it is dense in $\{\phi_m\}$. Let this non-zero energy signal have a support $\text{supp}\,(f) = |T|$ in time and $|W|$ in frequency. From the Weyl-Heisenberg principle, $|T|$ and $|W|$ have to follow the constraints:

$$|T|\,|W| \geq N, \quad |T| + |W| \geq 2\sqrt{N}. \tag{13}$$

More generally, it can be shown that

$$|T| + |W| \geq \sqrt{|T|\,|W|} \geq \frac{2}{\mu\,(\Phi, \Psi)}, \tag{14}$$

where $\mu\,(\Phi, \Psi) \equiv \max_{i,j} |\langle \Phi_i, \Psi_j \rangle|$ the coherence between the two considered bases. The weaker the coherence, the stronger the uncertainty relation.

Suppose one now measures M coefficients, y_m, of the unknown signal $f \in \mathbb{R}^N$ in the basis $\{\phi_m\}$. Let K be the cardinality of f in $\{\psi_{\boldsymbol{\alpha}}\}$ and suppose that

$$2\,K\;|W| < N \tag{15}$$

holds. Then, there exists no other signal f' such that the difference $\Delta \equiv f - f'$ is in the nullspace of Φ with support such that Eq. (14) holds. Since $\text{supp}_\Psi\,(f) = K$, then $\text{supp}_\Psi\,(\Delta) \leq 2K$ while $\text{supp}_\Phi\,(\Delta)$ was assumed to be $|W|$. It follows from Eq. (15) that $\text{supp}_\Phi\,(\Delta)\,\text{supp}_\Psi\,(\Delta) \leq 2K\,|W| < N$. However, since Eq. (13) holds for all non-zero signals, Δ must be identically zero, establishing uniqueness of the recovery. Then, there exists an algorithm which allows the stable recovery of the K unknown coefficients from the M measurements. Among them, a L^1-norm minimizing linear program was shown to provide an efficient and tractable computational approach, Donoho and Stark (1989).

The Compressed Sensing (CS) theory considers the projection of the residual $f - \widehat{f}$, $\widehat{f} \equiv \Psi\,\boldsymbol{X}$, onto M randomly chosen elements of the measurement basis $\{\phi_m\}$. Letting

$$y_m \equiv \langle \phi_m, f \rangle\,, \quad \widehat{y}_m \equiv \left\langle \phi_m, \widehat{f} \right\rangle = \langle \phi_m, \Psi\,\boldsymbol{X} \rangle\,, \quad \forall m \in \mathcal{J}^\Phi, \tag{16}$$

with $\mathcal{J}^\Phi \subseteq \{1, \ldots, N\} \subset \mathbb{N}$, $|\mathcal{J}^\Phi| = M$, the CS solution is given by

$$\boldsymbol{X}^\star = \underset{\boldsymbol{X} \in \mathbb{R}^N}{\arg\min} \|\boldsymbol{X}\|_1\,, \quad s.t. \quad \boldsymbol{Y} = A\,\boldsymbol{X}, \tag{17}$$

where $\boldsymbol{Y} \equiv (y_1 \ldots y_M)^T$, $A \equiv \Phi\,\Psi \in \mathbb{R}^{M \times N}$ and ϕ_m the m-th row of Φ. In this form, the problem is called *basis pursuit* and may be efficiently solved by reformulating it as a linear program.

Usually, the N unknowns $\{X_1, \ldots, X_N\}$ call for $M \geq N$ point evaluations. However, if the signal is sparse in the orthonormal trial basis Ψ, the *exact* signal can be recovered with far fewer evaluation points, $M < N$, hence the term *Compressed Sensing* to refer to this technique.

Remark The set $\{\psi_{\boldsymbol{\alpha}}\}$ need not define a basis and it is sufficient it defines a frame for the Compressed Sensing results to apply, Donoho (2006). In particular, a redundant dictionary approach is handled well in the CS framework. This is of interest since it is a popular choice to approximate a signal using an overcomplete set of functions as it may provide both sparser and more accurate representations, taking advantage of a large dictionary of approximating functions.

3.3 Recovery

The conditions under which the problem formulated in Eq. (17) exactly recovers a K-sparse signal have been the subject of numerous papers in the last few years. A popular set of results rely on the so-called Restricted Isometry Property (RIP) that essentially measures the degree of orthonormality of the columns of all submatrices A_T built from $|T|$ randomly selected columns of A. The K-RIP constant δ_K is defined as the smallest quantity such as the following holds:

$$(1 - \delta_K)\ \|\boldsymbol{v}\|_2^2 \leq \|A_T\, \boldsymbol{v}\|_2^2 \leq (1 + \delta_K)\ \|\boldsymbol{v}\|_2^2, \tag{18}$$

for all subsets $T \subseteq \{1, \ldots, N\}$, $|T| \leq K$, and vectors $\boldsymbol{v} \in \mathbb{R}^{|T|}$. This essentially indicates how much every set of $|T|$ arbitrarily chosen columns of A behaves as an orthonormal basis. If

$$\delta_{2K} < \frac{3}{4 + \sqrt{6}} \simeq 0.465, \tag{19}$$

or

$$\delta_K < 0.307, \tag{20}$$

holds, the solution of problem (17) recovers any sparse signal provided its support is such that $|T| \leq K$, Cai et al. (2009a), Foucart (2010).

The recovery property of the algorithm stated in Eq. (17) may alternatively be characterized using the concept of coherence of the sensing matrix A defined as the maximal magnitude of the off-diagonal entries of the Gram matrix $A^T A$ when A is unit-normed, see Donoho et al. (2006), Candès and Plan (2007),

Cai et al. (2009b). More generally, letting

$$\mu\left(A\right) \equiv \max_{1 \leq i \neq j \leq N} \frac{\left|A_i^T \, A_j\right|}{\left\|A_i\right\| \, \left\|A_j\right\|}, \tag{21}$$

it guarantees recovery of K-sparse signals provided the following holds, Donoho et al. (2006): [1]

$$K \leq \frac{1 + \mu\left(A\right)}{4\,\mu\left(A\right)}. \tag{22}$$

Since the most dominant coefficients are retrieved using the CS approach, one may think that it essentially tries to measure them and that an adaptive procedure may improve the recovery performance. For instance, subsequent measurements may take advantage of the information provided by former ones in an adaptive procedure. In fact, this approach would hardly do better than the *non-adaptive* CS technique with the resulting coefficient vector approximation L^2-error being within a factor 2 of that of any adaptive algorithm, see Theorem 2 in Donoho (2006).

3.4 Robustness

In practice, since the approximation basis is truncated to a finite number of functions, the approximation $\widehat{\boldsymbol{Y}} \approx \boldsymbol{Y}$ may not be exact. Further, the measurement vector $\boldsymbol{Y}$ may be subjected to noise. These factors lead the relaxation of the equality in Eq. (17) and reformulate the problem under the well-known *Basis Pursuit Denoising* form:

$$\boldsymbol{X}^\star = \underset{\boldsymbol{X} \in \mathbb{R}^N}{\arg\min} \left\|\boldsymbol{X}\right\|_1, \quad s.t. \quad \left\|\boldsymbol{Y} - A\,\boldsymbol{X}\right\|_2 \leq \epsilon, \tag{23}$$

with ϵ the approximation residual norm.

Further, the signal is rarely exactly sparse but only compressible in the retained approximation basis and it is crucial to have insights about the robustness of the recovery procedure in this framework. The CS theory provides results for this formulation: assume that $\delta_K < 0.307$, then the solution $\boldsymbol{X}^\star$ to

[1] A similar, albeit tighter, result holds for the L^0-norm version of the problem: $K \leq \frac{1+\mu(A)}{2\,\mu(A)}$.

Eq. (23) satisfies, Cai et al. (2009a): [2]

$$\|\boldsymbol{X} - \boldsymbol{X}^\star\|_2 \leq \frac{1}{0.307 - \delta_K} \left(\epsilon + \frac{\|\boldsymbol{X} - \boldsymbol{X}_K\|_1}{\sqrt{K}} \right), \tag{24}$$

where $\boldsymbol{X}_K$ is the K-term approximation of the signal $\boldsymbol{X}$ obtained by retaining the K most significant modes, *i.e.*, it is the K-mode sparsest representation if one was given full knowledge about the unknown signal $\boldsymbol{X}$ by an oracle. This result shows that the L^1-problem solution allows recovery of the K most significant entries, in the L^2-sense, of the unknown signal $\boldsymbol{X}$ from only M measurements and establishes the compressed sensing technique as both a tractable and robust solution method. In particular, it shows that the signal recovery error is simply proportional to the measurement noise ϵ and to the tail of the signal, $\|\boldsymbol{X} - \boldsymbol{X}_K\|_1$.

In practice, most signals exhibit some degree of sparsity. For instance, a very wide class of signals can be encompassed in the set of functions with a power-law decay rate. In particular, smooth and bounded variation signals obey a power-law decay and are then eligible for a compressed-sensing approach. More specifically, consider the class of $\mathbb{R}^N$-supported signals for which the ordered coefficients in the trial basis, $|X_1| \geq |X_2| \geq \ldots \geq |X_N|$, decay as

$$|X_k| \lesssim k^{-1/p}, \quad \forall\, 1 \leq k \leq N,\, 0 < p \leq 1. \tag{25}$$

Suppose the ordered signal belongs to the weak-L^p ball of radius R:

$$|X_k| \leq R\, k^{-1/p}, \tag{26}$$

the best K-term approximation of a signal then obeys (Candès et al., 2005):

$$\|\boldsymbol{X} - \boldsymbol{X}_K\|_1 \leq C_1\, K^{1-1/p}. \tag{27}$$

For compressible signals whose coefficients obey a power law decay, the resulting recovery error, Eq. (24), is then very small if the noise level is low.

4 Towards a CS-uncertainty quantification framework

4.1 Formulation

In this section, we build upon the compressed-sensing philosophy to define a tractable approach to assess the uncertainty associated with the solution of

[2] More precisely, $\|\boldsymbol{X} - \boldsymbol{X}^\star\|_2 \leq \frac{2\sqrt{2}\sqrt{1+\delta_K}}{1 - C_0\,\delta_K} \left(\epsilon + \frac{\|\boldsymbol{X} - \boldsymbol{X}_K\|_1}{\sqrt{K}} \right)$ with $C_0 = 1 + \frac{23}{2\sqrt{26}}$.

problems involving parametric uncertainty.

To characterize the unknown response surface of the quantity of interest u, we rely on a linear information operator $\mathcal{I} : U \mapsto \mathbb{R}^M$ which acts on a class of objects U. It provides the only piece of information one can access about the signal $u \in \mathcal{V}_\Xi \subseteq U$:

$$\mathcal{I}u \equiv (\ll \mathcal{I}_1, u \gg \ \ldots \ \ll \mathcal{I}_M, u \gg)^T, \tag{28}$$

with $\mathcal{I}_m$, $1 \leq m \leq M$, the sampling kernels and $\ll \mathcal{I}_m, \cdot \gg$ a linear functional.

Approximating the unknown response surface, hereafter also termed signal, in a basis $\{\psi_{\boldsymbol{\alpha}}\}$, see Eq. (5), and upon application of the operator $\mathcal{I}$, one writes:

$$\begin{aligned} \mathcal{I}u\left(\boldsymbol{\xi}(\theta)\right) &\approx \mathcal{I} \sum_{\boldsymbol{\alpha}\in\mathcal{J}} X_{\boldsymbol{\alpha}}\, \psi_{\boldsymbol{\alpha}}\left(\boldsymbol{\xi}(\theta)\right), \\ \Leftrightarrow \quad \mathcal{I}u\left(\boldsymbol{\xi}\right) &\approx \sum_{\boldsymbol{\alpha}\in\mathcal{J}} X_{\boldsymbol{\alpha}}\, \mathcal{I}\psi_{\boldsymbol{\alpha}}\left(\boldsymbol{\xi}\right). \end{aligned} \tag{29}$$

Stated this way, the uncertainty quantification problem reduces to the recovering the $\mathbb{R}^{P_{\boldsymbol{\xi}}}$-vector $\boldsymbol{X}$, $P_{\boldsymbol{\xi}} = |\mathcal{J}|$, from a limited number M of measurements. In particular, if u has a reasonably compressible representation in the $\{\psi_{\boldsymbol{\alpha}}\}$ basis, it is expected that $\boldsymbol{X}$ may be recovered with $M \ll P_{\boldsymbol{\xi}}$.

As mentioned above, given a particular choice of $\{\psi_{\boldsymbol{\alpha}}\}$, the trial basis on which the signal is approximated, the choice of $\mathcal{I}$ is critical in order to maximize the recovery property of the resulting sensing matrix A, see criteria in Eqs. (19) or (20). However, we must also consider an important computational issue: the evaluation of the measurement vector $\mathcal{I}u$ involves the unknown quantity u over its whole support, see Eq. (28), while the output of the model is not known over the entire space. Hence, one would like to retain the nice properties of the collocation-like UQ techniques where information on the solution is only required for a given number of realizations of the stochastic germ $\boldsymbol{\xi}$. Then, deterministic codes can be used as such as their output is a point-wise quantity in the stochastic domain. For sake of computational efficiency, it would also be desirable that the information operator be such that it only involves point-wise information from u. One possible choice is to consider the information operator as a series of random linear convolutions with M Dirac distributions δ_m. The resulting operator then requires only M point-wise evaluations $u_q \equiv u\left(\boldsymbol{\xi}^{(q)}\right)$ of u, $\left\{\boldsymbol{\xi}^{(1)} \ldots \boldsymbol{\xi}^{(M)}\right\}$ being chosen at random in Ξ: [3]

[3] with a slight abuse of notation with the indicator function $\mathbf{1}_\Xi$.

$$
\begin{aligned}
\ll \mathcal{I}_m, u \gg &\equiv \sum_{q=1}^{M} \gamma_m^{(q)} \int_{\mathbb{R}^{N_{\boldsymbol{\xi}}}} \delta_{\boldsymbol{\xi}^{(q)}}(-\boldsymbol{s})\ \mathbf{1}_{\Xi}\left[u(\boldsymbol{s})\ \mu_{\Xi}(\boldsymbol{s})\right] d\boldsymbol{s}, \\
&= \sum_{q=1}^{M} \gamma_m^{(q)}\, \mu_{\Xi}\left(\boldsymbol{\xi}^{(q)}\right)\ u_q, \quad 1 \leq m \leq M,
\end{aligned} \tag{30}
$$

so that one may rewrite

$$
\mathcal{I}u = \Phi\, \boldsymbol{u} \equiv \boldsymbol{Y}, \tag{31}
$$

with Φ a $\mathbb{R}^{M\times M}$-matrix, $\Phi_{m,l} = \gamma_m^{(l)}\, \mu_{\Xi}\left(\boldsymbol{\xi}^{(l)}\right)$, $\gamma_m^{(l)} \in \mathcal{N}(0,1)$, $1 \leq l, m \leq M$, and $\boldsymbol{u} = (u_1 \dots u_M)^T$. Similarly:

$$
\ll \mathcal{I}_m, \widehat{u} \gg \equiv \sum_{q=1}^{M} \gamma_m^{(q)} \int_{\mathbb{R}^{N_{\boldsymbol{\xi}}}} \delta_{\boldsymbol{\xi}^{(q)}}(-\boldsymbol{s})\ \mathbf{1}_{\Xi}\left[\sum_{\boldsymbol{\alpha}\in\mathcal{J}} X_{\boldsymbol{\alpha}}\, \psi_{\boldsymbol{\alpha}}(\boldsymbol{s})\ \mu_{\Xi}(\boldsymbol{s})\right] d\boldsymbol{s}, \tag{32}
$$

leading to

$$
\mathcal{I}\widehat{u} = \Phi\, \Psi\, \boldsymbol{X} \equiv \widehat{\boldsymbol{Y}}, \tag{33}
$$

with $\Psi \in \mathbb{R}^{M\times P_{\boldsymbol{\xi}}}$, $\Psi_{l,k} \equiv \psi_k\left(\boldsymbol{\xi}^{(l)}\right)$.

Letting $A \equiv \Phi\,\Psi \in \mathbb{R}^{M\times P_{\boldsymbol{\xi}}}$, the problem may then be reformulated in a form where one looks for the sparsest approximation vector $\boldsymbol{X}^{\star}$ such that the L^2-norm error between the observations $\boldsymbol{Y}$ and the reconstructed solution $\widehat{\boldsymbol{Y}} \equiv A\,\boldsymbol{X}$ is below ϵ:

$$
\boldsymbol{X}^{\star} \equiv \underset{\boldsymbol{X}\in\mathbb{R}^{P_{\boldsymbol{\xi}}}}{\arg\min} \|\boldsymbol{X}\|_1, \quad s.t. \quad \|\boldsymbol{Y} - A\,\boldsymbol{X}\|_2 \leq \epsilon, \tag{34}
$$

where the noise level ϵ characterizes the contribution of both the measurement noise when probing u_q, if any, and the energy of $\widehat{u}^{\top} \notin \mathcal{V}_{\Xi}^{(P_{\boldsymbol{\xi}})}$ with $\mathcal{V}_{\Xi}^{(P_{\boldsymbol{\xi}})} \subset \mathcal{V}_{\Xi}$ the space spanned by $\{\psi_{\boldsymbol{\alpha}}\}$, $\boldsymbol{\alpha} \in \mathcal{J}$.

At first glance, one may think that choosing the sampling kernels such that $\mathcal{I}_m u = \ll \psi_m, u \gg$, $m \in [1, P_{\boldsymbol{\xi}}]$, may be a good strategy, reminiscent of a Galerkin approach in variational methods where one tries to reduce the norm of the residual in the same space $span\{\phi_m\}$ as the space $span\{\psi_{\boldsymbol{\alpha}}\}$ in which the unknown signal is approximated. The resulting formulation has close connections with the standard least-squares interpolation. This choice is actually a bad one and results in very poor performance as it requires $M \geq P_{\boldsymbol{\xi}}$, in general, and achieves a near-unit RIP constant. Letting $\ll f, g \gg = \langle f, g\rangle_{L^2(\Xi,\mu_{\Xi})}$, the kernel of the sensing matrix is then of dimension $P_{\boldsymbol{\xi}} - M$ so that many sparse signals cannot be recovered unless $M \geq P_{\boldsymbol{\xi}}$, a far looser condition than that motivating the present work, $M \ll P_{\boldsymbol{\xi}}$.

We are now in a position of characterizing the solution u from its approximation $\Psi\boldsymbol{X}$. We assume the solution vector exhibits a decaying spectrum so that the uncertainty quantification problem is amenable to a form such that the CS tools and results apply.

Remark Without loss of generality, the variables $\gamma_m^{(l)}$ may be substituted with the Kronecker delta, δ_{ml}. Any element of Φ can then be written $\Phi_{ml} \equiv \mu_\Xi\left(\boldsymbol{\xi}^{(m)}\right)\,\delta_{ml}$, so that Φ is diagonal. Further, in the example treated in Section 6, the stochastic space Ξ is bounded and the measure μ_Ξ is uniform. It results that Φ is then simply within a multiplicative constant of the identity matrix.

4.2 Recovery property

Given that computational efficiency has determined our choice of $\mathcal{I}$, it is desirable to assess the adequacy and performance of the resulting pair $\{\Phi, \Psi\}$ of measurement and representation functions. The recovery property of the sensing matrix A is first investigated in terms of its Restricted Isometry Property (RIP) constant, see Eq. (18). The definition of the RIP is symmetric in the sense that it involves both a lower and upper bound. However, while the largest eigenvalue of $A_T^T\,A_T$ has an impact on the stability of the recovery algorithm, the smallest eigenvalue is of critical importance in the sense that it allows us to distinguish any two K-sparse vectors $\boldsymbol{X}$ and $\boldsymbol{X}'$ from their measurement by A and guarantees that no K-sparse vectors $\boldsymbol{X} \neq \boldsymbol{X}'$ exist such that $A\,\boldsymbol{X} = A\,\boldsymbol{X}'$, Blanchard et al. (2010). The focus will therefore be put on the lower inequality and we now define the RIP constant as

$$\delta_K \equiv \min_{\delta_K \geq 0} \delta_K, \quad s.t. \quad (1-\delta_K)\,\|\boldsymbol{X}\|_2^2 \leq \|A\,\boldsymbol{X}\|_2^2\,, \quad \forall\, K\text{-sparse vectors } \boldsymbol{X} \tag{35}$$

and derive an estimate for our choice of sensing matrix A. The RIP constant achieved by a matrix whose elements are sampled from a zero-mean, $1/M$-variance, Gaussian distribution is also plotted for comparison. This particular sensing matrix is known to be near-optimal in the sense that it allows the best recovery probability with a given number of measurements among all choices of measurement/approximation pairs and thus defines a lower bound to the RIP constant. The evaluation of the RIP constant of a matrix is not trivial. Since δ_K is directly related to the smallest (and largest) eigenvalues of $A_T^T\,A_T$, $|T| \leq K$, it involves computing the minimal (and maximal) eigenvalues of all $\begin{pmatrix} P_{\boldsymbol{\xi}} \\ K \end{pmatrix}$ sub-matrices $A_T^T\,A_T$. This problem is combinatorial in nature and cannot be computed in polynomial time. For the size of matrices considered in this work, the exact numerical computation of the RIP constant is intractable. However, to gain preliminary insights into the efficiency of the present approach, an estimate $\widehat{\delta}$ of the RIP constant is evaluated from over three million randomly selected submatrices A_T, both for the retained measurement/approximation pair and for a Gaussian random matrix. Since not all submatrices A_T can be tested, this estimate only constitutes a lower bound of the actual RIP constant.

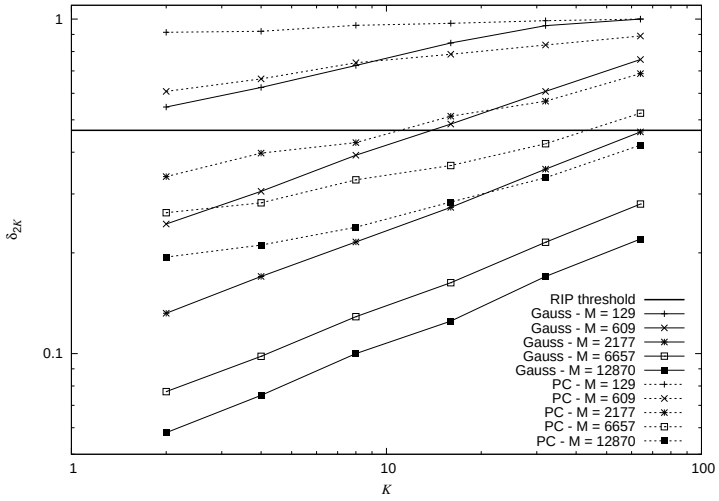

Fig. 1. Indication $\widehat{\delta}$ of the recovery property for various sizes M of the measurement ensemble as a function of the signal sparsity K. The present measurement/approximation pair, labelled 'PC', is compared with the reference Gaussian sensing matrix, labelled 'Gauss'. The recovery threshold $\frac{3}{4+\sqrt{6}}$, see Eq. (19), is also plotted for completeness (solid horizontal line). When K grows, the number of submatrices A_T grows exponentially so that the indicator of the RIP is not expected to be relevant enough and is thus not plotted.

As may be appreciated from Fig. 1, when the number K of non-zero elements of the signal $\boldsymbol{X}$ increases for a given number of measurements M, $\widehat{\delta}_{2K}$ increases to a point where the RIP criterion is not met, *i.e.*, $\delta_{2K} \geq \frac{3}{4+\sqrt{6}}$, consistent with the fact that more measurements are necessary to recover a vector when its sparsity decreases (K increases). As expected from its proven optimality with an overwhelming probability, the Gaussian sensing matrix exhibits better recovery properties than our current pair. For example, it would require about five times less measurements to perfectly recover any 64-sparse signals.

As seen in section 3.3, the recovery of the unknown coefficient vector $\boldsymbol{X}$ from M measurements may alternatively be guaranteed using the incoherence of the resulting matrix A. Both the RIP- and the incoherence-based approaches provide upper bounds on the cardinality of the signal one can recover and constitute *sufficient* conditions. However, these bounds hold with an overwhelming probability for *any* K-sparse signal and hence do not do justice to the recovery ability of the method in practice. As an illustration of this observation, the indicator of the RIP constant plotted in Fig 1 show that the recovery using $M = 2177$ measurements is not guaranteed for a signal which cardinality exceeds about 10 using point-wise samples of the response surface. Since the RIP constant is bounded from below by this indicator, the actual situation may appear even worse. However, results presented in section 6 will show that this bounds is far too pessimistic and that the present technique performs very well in practice. Hence, instead of estimating bounds of limited practical interest, one may instead rely on a k-fold cross-validation approach

to evaluate the number M of required measurements to recover the signal within a given accuracy. An approximation of the response surface is then determined using M measurement points and its accuracy is evaluated in terms of the L^2-norm of the residual computed on a different set of M/k points. The procedure is repeated k times so that each set of M/k points alternatively serves for the approximation and the residual norm estimation. M is subsequently increased whenever the resulting average residual norm is deemed too large.

4.3 Improving the recovery

In the situation where the sample points can be chosen at will, *i.e.*, the unknown response surface can be probed at any point within the domain of interest, an opportunity for improving the efficiency of the present approach arises in the form of a design of experiment. Indeed, once the approximation basis $\{\psi_{\boldsymbol{\alpha}}\}$ is chosen, the set of sampling points is the only degree of freedom one has within the NISP framework to improve on the recovery of the unknown output. By carefully choosing the information operator $\{\mathcal{I}\}$, and hence the associated set of samples, the efficiency of the recovery can be favorably affected and the RIP improved in the sense that criteria stated in Eqs. (19) and (20) are met for a larger K (less compressible signal).

In recent work, Rauhut (2010), Rauhut and Ward (2010) investigate the recovery properties of a strategy based on an approximation basis generated by a system of polynomials $\{\psi_\alpha\}$ orthonormal with respect to a measure $\mu_\xi(\xi)$ satisfying (in 1-D)

$$\left(1-\xi^2\right)^{1/4} \mu_\xi(\xi)^{1/2} \, |\psi_\alpha(\xi)| \leq c_1, \quad \forall\, \alpha \in \mathbb{N},\, \xi \in [-1,1]\,. \tag{36}$$

Within this framework, theoretical results are proven for a CS strategy relying on the orthonormal system $\{\psi_\alpha\}$ and point-wise samples drawn independently according to a Chebyshev probability measure, $d\nu(\xi) = \pi^{-1}\,(1-\xi^2)^{-1/2}\,d\xi$. In particular, it is shown that, provided $M \geq c_2\,\delta^{-2}\,c_\infty^2\,K\,\log^3(K)\,\log(P_{\boldsymbol{\xi}})$, then, with probability at least $1 - P_{\boldsymbol{\xi}}{}^{-c_3\log^3(K)}$, the associated sensing matrix $\widetilde{A}$ obeys the RIP with constant $\delta_K \leq \delta$, with c_2 (depending only on μ_ξ) and c_3 well-behaved constants and c_∞ such that

$$\sup_{k\in[1,P_{\boldsymbol{\xi}}]} \|\psi_k\|_\infty = \sup_{k\in[1,P_{\boldsymbol{\xi}}]} \sup_{\xi\in[-1,1]} |\psi_k\,(\xi)| \leq c_\infty. \tag{37}$$

Further, if $M \geq c_2\,K\,\log^4(P_{\boldsymbol{\xi}})$, consider the following L^1-minimization prob-

lem (in the 1-D case, $N_{\boldsymbol{\xi}} = 1$)

$$\boldsymbol{X}^{\star} \equiv \underset{\boldsymbol{X} \in \mathbb{R}^{P_{\boldsymbol{\xi}}}}{\arg\min} \|\boldsymbol{X}\|_1, \quad s.t. \quad \left\| \widetilde{\boldsymbol{Y}} - \widetilde{A}\, \boldsymbol{X} \right\|_2 \leq \epsilon, \tag{38}$$

where $\widetilde{A} \equiv \Phi\, \Psi$ and $\widetilde{\boldsymbol{Y}} \equiv \Phi\, \boldsymbol{Y}$ with Φ a diagonal $\mathbb{R}^{M \times M}$ matrix with entries $\Phi_{l,l} \sim \left(1 - \left(\xi^{(l)}\right)^2\right)^{1/4} \mu_{\xi}\left(\xi^{(l)}\right) / \sqrt{M}$ and $\Psi_{l,k} \equiv \psi_k\left(\xi^{(l)}\right)$. The solution $\boldsymbol{X}^{\star}$ of this problem satisfies

$$\|\boldsymbol{X} - \boldsymbol{X}^{\star}\|_2 \leq c_3\, \epsilon + c_4 \frac{\|\boldsymbol{X} - \boldsymbol{X}_K\|_1}{\sqrt{K}}, \tag{39}$$

with probability exceeding $1 - {P_{\boldsymbol{\xi}}}^{-c_5 \log^3(P_{\boldsymbol{\xi}})}$. This L^1-minimization problem is similar to that in Eq. (34). The above results constitute a theoretical basis for a provably efficient point-wise orthonormal polynomials-based CS technique and support the approach suggested in this paper. However, a good sample set in the sense of the RIP (or equivalently the mutual coherence) is not necessarily the most pertinent. Indeed, rather than extra precision on the coefficient vector $\boldsymbol{X}$, as guaranteed by a good RIP, one is often interested in retrieving a good approximation of the signal u. From this perspective, a good set is one that minimizes the error in the signal recovery, say:

$$\left\{\boldsymbol{\xi}^{(q)\star}\right\} = \underset{\left\{\boldsymbol{\xi}^{(q)}\right\}}{\arg\min} \|u - \widehat{u}\|_{L^2(\Xi, \mu_{\Xi})}. \tag{40}$$

One thereby favors good recovery of u to the detriment of finely distinguishing between two, weakly contributing, coefficients of $\boldsymbol{X}$. Of course, problem (40) cannot be solved since u is not known. Observe now that the CS-UQ technique proposed in this paper, Eq. (34), leads to finding the minimal L^1-norm vector $\boldsymbol{X}$ so that the reconstructed signal $\widehat{u}$ is close to u at a given set of points $\left\{\boldsymbol{\xi}^{(q)}\right\}$:

$$0 \leq \sum_{q=1}^{M} \left[\left| u\left(\boldsymbol{\xi}^{(q)}\right) - \widehat{u}\left(\boldsymbol{\xi}^{(q)}\right) \right|^2 \mu_{\Xi}\left(\boldsymbol{\xi}^{(q)}\right)^2 \right] \leq \epsilon^2 \ll \sum_{q=1}^{M} \left[\left| u\left(\boldsymbol{\xi}^{(q)}\right) \right|^2 \mu_{\Xi}\left(\boldsymbol{\xi}^{(q)}\right)^2 \right]. \tag{41}$$

To achieve good recovery of u in terms of L^2-norm as stated in Eq. (40), $\left\{\boldsymbol{\xi}^{(q)}\right\}$ could be chosen so that the L^2-norm is well-approximated by the M-term sum, Eq. (41). One is then left with finding a set of points so that a $N_{\boldsymbol{\xi}}$-dimensional integral is well-approximated with a finite sum of integrand evaluations. While deriving the optimal set of points is difficult in general, provably good candidates exist such as low-discrepancy sequences. Among those, a Sobol sequence, Sobol (1967, 1977), presents interesting properties in filling the $\mathbb{R}^{N_{\boldsymbol{\xi}}}$-space at hand and was used in this work.

The efficiency of the NISP CS-UQ technique based on a set of samples issued from both a low-discrepancy Sobol sequence and from a Chebyshev probability measure will be investigated in section 6.3.4.

Remark While not used in this work, the recovery could also be improved by splitting the step of finding the subset of dominant modes from that of evaluating their coefficients. Once the solution $\boldsymbol{X}^\star$ of Eq. (34) has been determined, the subset $\{\psi_\star\}$ of dominant modes is identified. Letting $K_\star \leq M$ be its cardinality, one can then reuse the available information from the measurements, $\boldsymbol{Y}$, to evaluate the coefficients $\boldsymbol{X}_\star$ of the $K_\star$-best term approximation: $\boldsymbol{X}_\star = A_\star^\dagger \boldsymbol{Y}$, $A_\star \equiv \Phi\,\Psi_\star \in \mathbb{R}^{M\times K_\star}$ with $A^\dagger$ the Moore-Penrose pseudo-inverse of A. This second step then concentrates the information from the observations in order to recover the coefficients $\boldsymbol{X}_\star$, *de facto* discarding those not belonging to $\{\psi_\star\}$ found negligible at the first step. Focusing the information on the set of modes found to be dominant then provides superior performance. A similar procedure was proposed in Candès and Plan (2007).

4.4 Solution method

4.4.1 Formulation of the optimization problem

As seen above, the problem takes the form of an inequality constrained optimization problem which is solved for $\boldsymbol{X} \in \mathbb{R}^{P_{\boldsymbol{\xi}}}$. Approximation of the random output u on a Polynomial Chaos requires to determine a number of terms $P_{\boldsymbol{\xi}}$ which grows with the polynomial order N_o and the stochastic space dimension $N_{\boldsymbol{\xi}}$ as:

$$P_{\boldsymbol{\xi}} = \begin{pmatrix} N_{\boldsymbol{\xi}} + N_o \\ N_{\boldsymbol{\xi}} \end{pmatrix} = \frac{(N_{\boldsymbol{\xi}} + N_o)\,!}{N_{\boldsymbol{\xi}}\,!\,N_o\,!}. \tag{42}$$

When the stochastic space dimensionality and/or the polynomial order increases, the number of terms grows exponentially, as a symptom of the curse of dimensionality. For reasonably large stochastic problems, the number of required terms quickly becomes large as well and solving the problem in Eq. (34) may become difficult. A great deal of work has been devoted to solving this class of problem or closely related formulations such as an alternative convex constrained version:

$$\boldsymbol{X}^\star \equiv \arg\min_{\boldsymbol{X}} \|\boldsymbol{Y} - A\,\boldsymbol{X}\|_2, \quad s.t. \quad \|\boldsymbol{X}\|_1 \leq \epsilon_{\boldsymbol{X}}, \tag{43}$$

or a (convex) unconstrained optimization:

$$\boldsymbol{X}^\star \equiv \arg\min_{\boldsymbol{X}} \|\boldsymbol{X}\|_1 + \tau \ \|\boldsymbol{Y} - A\,\boldsymbol{X}\|_2, \tag{44}$$

where τ is a non-negative real parameter.

Algorithms vary depending on the formulation of the problem used. For instance, ideas from the Least Angle Regression (LARS) procedure (Efron et al., 2004) may be used to solve formulation (43) while formulation (34) may be recast as a second order cone program for which efficient algorithms exist, see for instance Becker et al. (2009).

In the present work, the formulation (44) is used together with a memory-limited second-order quasi-Newton approach, Gilbert and Lemaréchal (1989). The maximum size of the problem considered below is about $P_{\boldsymbol{\xi}} \simeq 10^5$ and the optimization step computational time remains much lower than that required to generate the M deterministic solver outputs, even though the deterministic code calls were performed in parallel. Alternative approaches are available to solve Eq. (44) such as projected gradient techniques, Figueiredo et al. (2007); van den Berg and Friedlander (2008), Interior Points methods (IP), Wright (1997), or iterative-shrinkage techniques, see Zibulevsky and Elad (2010). With very large scale problems in mind, where both $P_{\boldsymbol{\xi}}$ and M are large, the dense matrix $A \in \mathbb{R}^{M \times P_{\boldsymbol{\xi}}}$ is not stored and one only computes the $\mathbb{R}^M$-vector resulting from $A\,\boldsymbol{X}$. Further, while not used here, as the CPU cost essentially comes from this matrix-vector multiplication $A\,\boldsymbol{X}$, and since $\boldsymbol{X}$ is compressible, one may make use of sparse multiplication techniques to lower the computational burden.

4.4.2 Choosing τ

As also used for determining the required number of measurements, see section 4.2, a cross-validation technique is used to determine the balance between reconstructed signal norm $\|\boldsymbol{X}\|_1$ and the approximation error $\|\boldsymbol{Y} - A\,\boldsymbol{X}\|_2$ due to the presence of noise and the incompleteness of the approximation basis $\{\psi_{\boldsymbol{\alpha}}\}$. For a given M, the Pareto front is explored by varying τ in Eq. (44). A weakly penalized observation constraints (low τ) would lead to an approximation lying on a too low-dimensional subspace; on the contrary, a large τ may lead to overfitting on the available observations. An example of the Pareto front is given in Fig. (2). The retained $\tau^\star$ is estimated with a k-fold cross-validation technique as that which minimizes the mean reconstruction error over the k folds: $\tau^\star = \arg\min_\tau \sum_{l=1}^{k} \left\| \boldsymbol{u}^{(l)} - \Psi^{(l)}\,\boldsymbol{X} \right\|_2$ with $\boldsymbol{u}^{(l)}$ the l-th set of samplings independent of that used in $\boldsymbol{Y} = \Phi\,\boldsymbol{u}$, $k = 3$ being retained in this work. More sophisticated cross-validation techniques such as the Leave-One-Out or the .632+ bootstrap method (Efron and Tibshirani, 1997) may

lead to lower variance error estimation but are deemed too costly.

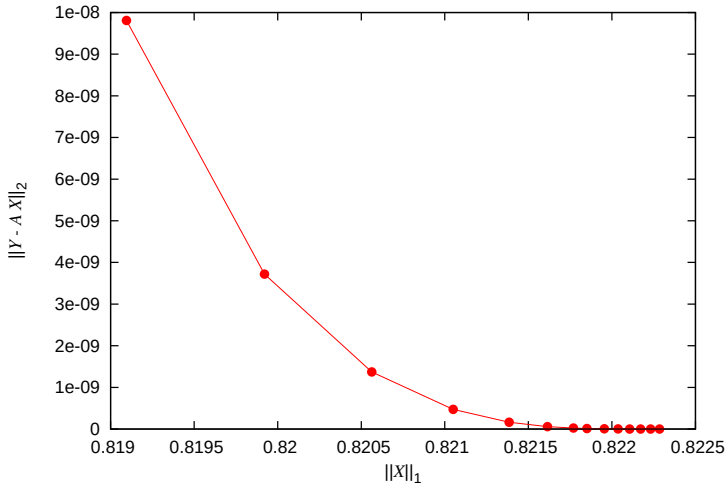

Fig. 2. Example of Pareto front. $M = 2177$.

4.4.3 Modified L^1-norm

Since the L^1-norm is not smooth, a modified L^1-norm was used to ease the optimization procedure. For a real-valued quantity f, the smoothed L^1-norm is taken as, Becker et al. (2009):

$$\begin{aligned} \|f\|_1 &= |f| - \frac{1}{2}\,\epsilon_s \quad \text{if } |f| \geq \epsilon_s, \quad f \in \mathbb{R}, \\ &= \frac{f^2}{2\,\epsilon_s} \quad \text{otherwise,} \end{aligned} \tag{45}$$

with $\epsilon_s \ll 1$ a non-negative smoothing parameter taken as $10^{-7}\,X_{max}$ with X_{max} the estimated maximum magnitude coefficient.

As mentioned above, shifting from a L^0- to a L^1-formulation makes the resulting problem convex and computationally tractable at the expense of a factor 2 in the sparsity bound. From a most efficient recovery perspective, an *ad-hoc* weight is then introduced in the definition to mimic a L^0-norm:

$$\|\boldsymbol{X}\|_1 \longrightarrow \|W\,\boldsymbol{X}\|_1\,, \tag{46}$$

where W is a diagonal matrix which elements are $W_k = \frac{1}{|X_k|+\varepsilon_W}$, $1 \leq k \leq P_{\boldsymbol{\xi}}$, $\varepsilon_W > 0$. The iterative scheme suggested in Candès et al. (2007) is used and the W matrix is updated as a new solution estimate $\boldsymbol{X}$ is available. In this work, $\varepsilon_W = 10^{-7}\ \max_k |X_k|$ was retained as a sparsity threshold.

5 Model problem

5.1 General motivation

To investigate the efficiency and effectiveness of the method presented above, it is applied to the simulation of an underwater seismic event. Uncertainty is assumed in the location, intensity and physical extent of the event as well as in the ocean depth field. The quantity of interest is the maximum height of the resulting ocean surface perturbation at a specific location next to the shore within a given time window after the event has occurred. The length of the time window is related to the time necessary for the seaquake detection, broadcast of the alert and evacuation of the population located close to the shore.

5.2 Governing equations

The shallow water flow is described by the following set of equations

$$\frac{D\, v_x}{D\, t} = f\, v_y - g\, \frac{\partial h}{\partial x} - b\, v_x + S_{v_x}, \tag{47}$$

$$\frac{D\, v_y}{D\, t} = -f\, v_x - g\, \frac{\partial h}{\partial y} - b\, v_y + S_{v_y}, \tag{48}$$

$$\frac{\partial h}{\partial t} = -\frac{\partial \left(v_x\ (H+h)\right)}{\partial x} - \frac{\partial \left(v_y\ (H+h)\right)}{\partial y} + S_h, \tag{49}$$

where it is implicitly assumed that the fluid density and the free surface pressure are constant. Here, f is the term corresponding to the Coriolis force, b the viscous drag coefficient, $\boldsymbol{v} \equiv (v_x\, v_y)^T$ the fluid velocity vector, h the deviation of the ocean surface from its position at rest, g the gravity constant and H the ocean depth field. S_{v_x}, S_{v_y} and S_h are the source terms reflecting the effect of the unknown displacement field. Without loss of generality, it is assumed that the source field acts on the h variable solely $\left(S_{v_x} = 0, S_{v_y} = 0\right)$ and that the drag and the Coriolis forces can be neglected, $f = 0$, $b = 0$. No-slip boundary conditions are prescribed along the edge Γ of the domain $\Omega_{\boldsymbol{x}}$.

5.3 Discretization in the deterministic space

Consider a partition of the 2-D physical domain $\Omega_{\boldsymbol{x}}$ into a set of $N_b = N_x \times N_y$ non-overlapping spectral elements (SE) with respective support $\Omega^l_{\boldsymbol{x}}$ for $l =$

$1, \ldots, N_b$:

$$\Omega_{\boldsymbol{x}} = \bigcup_{l=1}^{N_b} \Omega_{\boldsymbol{x}}^l. \tag{50}$$

The continuous Galerkin spectral elements approximation of the solution over the element $\Omega_{\boldsymbol{x}}^l$ for $v_x(\cdot, t) \in \mathcal{V}_{\boldsymbol{x}}$ is given by:

$$v_x^l\left(\boldsymbol{x} \in \Omega_{\boldsymbol{x}}^l, t; \theta\right) = \sum_{i=1}^{P_x} \sum_{j=1}^{P_y} v_{xi,j}^l(t;\theta)\, \mathcal{L}_i(x)\, \mathcal{L}_j(y), \tag{51}$$

where $\mathcal{L}_i$ are the physical space basis functions, P_x and P_y the spectral orders and $\mathcal{V}_{\boldsymbol{x}}^l$ is a suitable Hilbert space of $\Omega_{\boldsymbol{x}}^l$.

The unknowns are interpolated with Legendre cardinal functions collocated at the Gauss-Lobatto points. The surface height $h^l \in \mathcal{V}_h^l$ is discretized with a lower order polynomial to avoid spurious pressure modes to occur ($Q_N - Q_{N-2}$ scheme), see Iskandarani et al. (1995). At the deterministic level, the discretized quantities are then:

$$v_s^l(\boldsymbol{x} \in \Omega_{\boldsymbol{x}}^l, t; \theta) = \sum_{i=1}^{P_x} \sum_{j=1}^{P_y} v_{si,j}^l(t;\theta)\, \mathcal{L}_i(x)\, \mathcal{L}_j(y), \tag{52}$$

$$h^l(\boldsymbol{x} \in \Omega_{\boldsymbol{x}}^l, t; \theta) = \sum_{i=1}^{P_x-2} \sum_{j=1}^{P_y-2} h_{i,j}^l(t;\theta)\, \mathcal{L}_i(x)\, \mathcal{L}_j(y), \tag{53}$$

where subscript s stands either for x or y.

Integrating the divergence term by parts the governing equations (47–49), yields the variational form of the shallow water equations (SWE), $\forall \varphi_{\boldsymbol{v}}^l \in \mathcal{V}_{\boldsymbol{x}}^l$, $\forall \varphi_h^l \in \mathcal{V}_h^l$:

$$\int_{\Omega_{\boldsymbol{x}}^l} \frac{\partial\, v_x^l}{\partial\, t}\, \varphi_{\boldsymbol{v}}^l\, d\Omega_{\boldsymbol{x}}^l + \int_{\Omega_{\boldsymbol{x}}^l} g\, \frac{\partial h^l}{\partial x}\, \varphi_h^l\, d\Omega_{\boldsymbol{x}}^l = \int_{\Omega_{\boldsymbol{x}}^l} f_x^l\, \varphi_{\boldsymbol{v}}^l\, d\Omega_{\boldsymbol{x}}^l, \tag{54}$$

$$\int_{\Omega_{\boldsymbol{x}}^l} \frac{\partial\, v_y^l}{\partial\, t}\, \varphi_{\boldsymbol{v}}^l\, d\Omega_{\boldsymbol{x}}^l + \int_{\Omega_{\boldsymbol{x}}^l} g\, \frac{\partial h^l}{\partial y}\, \varphi_h^l\, d\Omega_{\boldsymbol{x}}^l = \int_{\Omega_{\boldsymbol{x}}^l} f_y^l\, \varphi_{\boldsymbol{v}}^l\, d\Omega_{\boldsymbol{x}}^l, \tag{55}$$

$$\begin{aligned} \int_{\Omega_{\boldsymbol{x}}^l} \frac{\partial\, h^l}{\partial\, t}\, \varphi_h^l\, d\Omega_{\boldsymbol{x}}^l - \int_{\Omega_{\boldsymbol{x}}^l} \left(\frac{\partial\, \varphi_h^l}{\partial\, x}\, v_x^l + \frac{\partial\, \varphi_h^l}{\partial\, y}\, v_y^l \right) \left(H^l + h^l \right)\, d\Omega_{\boldsymbol{x}}^l \\ = \int_{\Omega_{\boldsymbol{x}}^l} \left(f_h + S_h \right)\, \varphi_h^l\, d\Omega_{\boldsymbol{x}}^l, \end{aligned} \tag{56}$$

where $\varphi_{\boldsymbol{v}}^l$ and φ_h^l respectively denote the velocity and surface height test functions and f_x, f_y and f_h are the generalized forcing terms including the nonlinear advection term. The pressure (since h acts as the pressure) gradient

term $\boldsymbol{\nabla} h$ in the momentum equation and the divergence term in the continuity equation $\boldsymbol{\nabla} \cdot (h\,\boldsymbol{v})$ drive the gravity waves and are integrated implicitly in time. The gravity terms are thus discretized with a Crank-Nicholson time scheme while the remaining terms are treated with a semi-implicit third order Adams-Bashforth scheme. Rearranging, the discretized SWE may be put in the following form:

$$\kappa\, M_{\boldsymbol{v}}\, v_x + \frac{1}{2}\, g\, G_{v_x}\, h = B, \tag{57}$$

$$\kappa\, M_{\boldsymbol{v}}\, v_y + \frac{1}{2}\, g\, G_{v_y}\, h = C, \tag{58}$$

$$-\frac{1}{2}\,\left(E_{v_x}\, v_x + E_{v_y}\, v_y\right) + \kappa\, M_h = D, \tag{59}$$

with $\kappa \equiv 1/\Delta t$, Δt the retained time step. $M_{\boldsymbol{v}}$ and M_h are the mass matrices for the velocity and pressure respectively. The matrices G_{v_x} and G_{v_y} are the discrete gradient operators along x and y for the velocity vector while E_{v_x} and E_{v_y} are the discrete gradient operators for the pressure along x and y. The matrices B, C and D appearing on the right sides of the equations contain the explicit terms together with the sources. Rearranging the system (57–59), a Schur complement formulation is derived and the problem is solved with a matrix-free conjugate gradient method taking advantage of a Schwarz preconditionner. More details about a similar formulation may be found in Douglas et al. (2003).

5.4 Discretization in the stochastic space

A Polynomial Chaos spectral expansion is used to approximate the uncertain output of the model. Without loss of generality, we rely on uniform random *iid* variables $\boldsymbol{\xi} = \left(\xi_1 \ldots \xi_{N_{\boldsymbol{\xi}}}\right)^T$ associated with normalized Legendre polynomials, Abramowitz and Stegun (1970), $\psi_k(\boldsymbol{\xi})$, $k = 1, \ldots, P_{\boldsymbol{\xi}}$, on the stochastic space $L^2\left(\Xi, \mu_\Xi\right)$, $\Xi = [-1, 1]^{N_{\boldsymbol{\xi}}}$.

These polynomials form an orthonormal basis under the measure μ_Ξ:

$$\int_\Xi \psi_i(\boldsymbol{s})\, \psi_j(\boldsymbol{s})\, d\mu_\Xi(\boldsymbol{s}) \equiv \langle \psi_i,\, \psi_j \rangle_{L^2(\Xi, \mu_\Xi)} = \delta_{ij}, \quad \forall\, \{i, j\} \in \{1, \ldots, P_{\boldsymbol{\xi}}\}, \tag{60}$$

and $u\left(\boldsymbol{\xi}\right)$ is approximated by:

$$u\left(\boldsymbol{\xi}\right) \approx \widehat{u}\left(\boldsymbol{\xi}\right) = \sum_{k=1}^{P_{\boldsymbol{\xi}}} X_k\, \psi_k\left(\boldsymbol{\xi}\right). \tag{61}$$

The uncertain depth field $H(\boldsymbol{x})$ is modeled as a N_H-term expansion of the form:

$$H\left(\boldsymbol{x}, \widehat{\boldsymbol{\xi}}(\theta)\right) = \overline{H}\left(\boldsymbol{x}\right) + \sum_{i=1}^{N_H} \sqrt{\lambda_i}\, \widehat{\xi}_i\left(\theta\right)\, \varphi_i^H(\boldsymbol{x}), \tag{62}$$

with $\widehat{\boldsymbol{\xi}} \equiv \left(\widehat{\xi}_1 \ldots \widehat{\xi}_{N_H}\right)^T$, and $\widehat{\xi}_i$, $i = 1, \ldots, N_H$, *iid* uniform random variables. The expansion modes $\varphi_i^H(\boldsymbol{x})$ are eigenvectors of the auto-correlation operator based on the following kernel:

$$C\left(\boldsymbol{x}, \boldsymbol{x}'\right) \equiv \exp^{\left(-\frac{1}{l_C}\left[(\boldsymbol{x}-\boldsymbol{x}')^T(\boldsymbol{x}-\boldsymbol{x}')\right]^{1/2}\right)}, \tag{63}$$

with λ_i the associated eigenvalues and l_C the correlation length. The physical domain extent is $\boldsymbol{x} \in [0, 10^6]^2$ and the correlation length is taken as $l_C = 2 \times 10^5$.

The location of the seaquake source S_h is also unknown by nature. While insights may be gained from past seaquakes and geological considerations, the precise description of the sea bottom displacement field during a seismic event cannot be predicted and it is conveniently modeled as a random field indexed by $\Xi \times T$, with T the time domain. The source model is of the form:

$$\begin{aligned} S_h\left(\boldsymbol{x}, t, \widetilde{\boldsymbol{\xi}}\right) &\equiv A_{S_h}(t)\, \mathcal{N}\left(\boldsymbol{x}; \boldsymbol{x}_{S_h}, \sigma_{S_h}\right), \\ A_{S_h}(t) &= \frac{t^2}{1+t^4}\, \exp^{-20\,t}, \end{aligned} \tag{64}$$

where the temporal envelope $A_{S_h}(t)$ is assumed known. The source physical extent $\mathcal{N}\left(\boldsymbol{x}; \boldsymbol{x}_{S_h}, \sigma_{S_h}\right)$ is assumed isotropic and Gaussian-shaped while its location $\boldsymbol{x}_{S_h}$, physical extent σ_{S_h} and strength $A_{\mathcal{N}}$ are random:

$$\begin{aligned} \boldsymbol{x}_{S_h}\left(\widetilde{\boldsymbol{\xi}}\right) &\equiv \left(\left(0.5 + 0.1\,\widetilde{\xi}_1\right)\overline{x}_{S_h} \quad \overline{y}_{S_h}\right)^T, \\ \mathcal{N}\left(\boldsymbol{x}; \boldsymbol{x}_{S_h}, \sigma_{S_h}\right) &= A_{\mathcal{N}}\left(\widetilde{\xi}_2\right) \exp^{\left[-\frac{1}{\sigma_{S_h}^2\left(\widetilde{\xi}_3\right)}\left(\boldsymbol{x}-\boldsymbol{x}_{S_h}\right)^T\left(\boldsymbol{x}-\boldsymbol{x}_{S_h}\right)\right]}, \end{aligned} \tag{65}$$

with $\left(\overline{x}_{S_h}\, \overline{y}_{S_h}\right)^T$ a reference source location. The amplitude and variance of the source express as

$$\begin{aligned} A_{\mathcal{N}}\left(\widetilde{\xi}_2\right) &= 1.01 + \widetilde{\xi}_2, \\ \sigma_{S_h}^2\left(\widetilde{\xi}_3\right) &= 5\,10^3\,\left(1.25 + \widetilde{\xi}_3\right). \end{aligned} \tag{66}$$

6 Results

6.1 Solution method

Salient results of the SWE UQ problem are presented in this section. First, a low-dimensional stochastic solution is investigated both with the proposed CS-like approach and a sparse grid-based (Smolyak scheme) NISP Polynomial Chaos technique. This low dimensional problem allows relatively high polynomial orders for the approximation of the solution. In a second step, a higher dimensional problem is considered, arising from a more realistic modeling involving additional sources of uncertainty. For this particular case study, both a sparse grid-based Polynomial Chaos (PC) approach and a Stochastic Collocation (SC) approach are considered, together with their CS counterparts (CS-PC and CS-SC). In the PC method, the required coefficients in Eq. (10) are approximated using discrete quadrature. The SC method essentially consists of approximating the output response surface by interpolating Lagrange polynomials and no integrals need be evaluated to determine its coefficients. Both the CS-based PC and CS-based SC use the CS-like approach but they differ in their respective measurement matrices, A, due to the use of different trial bases $\{\psi_{\boldsymbol{\alpha}}\}$.

Throughout the section, the sensing matrix A is never explicitly formed and only its action on a vector $\boldsymbol{X}$ is evaluated, $\widehat{\boldsymbol{Y}} = A\,\boldsymbol{X}$.

6.2 Low dimensional problem

In this section, a 1-term series expansion is considered for the approximation of the ocean depth stochastic field, $N_H = 1$, Eq. (62). Further, it is assumed that there is no uncertainty in the source field extent σ_{S_h} and intensity $A_{\mathcal{N}}$. The problem expresses in terms of $\boldsymbol{\xi}^T = \left(\widehat{\boldsymbol{\xi}}^T\, \widetilde{\boldsymbol{\xi}}^T\right)$ which here reduces to $\boldsymbol{\xi}^T = \left(\widehat{\xi}_1\, \widetilde{\xi}_1\right) = (\xi_1\, \xi_2)$ so that the problem then lies in a 2-D stochastic domain.

Since a major concern of any computational method is the balance between accuracy of the solution and required computational effort, let us first examine the error in the solution as more measurements M are considered. The reconstruction error is defined as:

$$\widehat{\varepsilon}^2_{ex} \equiv \left\| u\left(\boldsymbol{x}^\star, t^\star, \boldsymbol{\xi}\right) - \widehat{u}\left(\boldsymbol{x}^\star, t^\star, \boldsymbol{\xi}\right) \right\|^2_{L^2(\Xi, \mu_\Xi)}, \tag{67}$$

where u is the exact stochastic quantity of interest evaluated at a specific time $t^\star$ and specific location $\boldsymbol{x}^\star$. $\widehat{u}$ is the solution obtained from the uncertainty

l_S	1	2	3	4	5	6	7	8	9	10	11	12
M	5	9	17	33	33	65	97	97	161	161	161	257
N_o	1	2	3	5	5	6	8	8	11	11	11	12
$P_{\boldsymbol{\xi}}$	3	6	10	15	21	28	36	45	55	66	78	91

Table 2
Correspondence between Smolyak level l_S, number of evaluation points $M = |\mathcal{N}_q|$, maximum Polynomial Chaos order N_o and related number of stochastic modes $P_{\boldsymbol{\xi}}$. 2-D stochastic space.

quantification using either the Smolyak scheme-based PC or the CS-like PC strategy.

In the sequel, since the exact solution u is not known, a N_{MC}-sample Monte-Carlo approximation is considered instead, with $N_{MC} = 1.2 \times 10^6$ sufficiently large so that the norm $\|\cdot\|^2_{L^2(\Xi,\mu_\Xi)}$ may be reasonably approximated by its Monte-Carlo estimation:

$$\widetilde{\varepsilon}^2_{ex} \simeq \widetilde{\varepsilon}^2 \equiv \frac{1}{N_{MC}} \sum_{q=1}^{N_{MC}} \left(u\left(\boldsymbol{x}^\star, t^\star, \boldsymbol{\xi}^{(q)}\right) - \sum_{k=1}^{P_{\boldsymbol{\xi}}} X_k\left(\boldsymbol{x}^\star, t^\star\right) \psi_k\left(\boldsymbol{\xi}^{(q)}\right) \right)^2, \quad (68)$$

where the $\boldsymbol{\xi}^{(q)}$ are sampled according to the μ_Ξ measure.

Finally, the relative error norm is defined as:

$$\varepsilon^2 \equiv \widetilde{\varepsilon}^2 \times \left[\frac{1}{N_{MC}} \sum_{q=1}^{N_{MC}} \left(u\left(\boldsymbol{x}^\star, t^\star, \boldsymbol{\xi}^{(q)}\right)\right)^2 \right]^{-1}. \quad (69)$$

The evolution of the approximation error ε is investigated in terms of the number of deterministic solver calls M. The number of solver calls corresponds to the required number of points for the different levels of the Smolyak rule. For a given level l_S of the Smolyak quadrature, the quadrature is exact for a polynomial integrand of a given total order. One wants the integration to be exact if the model output was a polynomial of the same order as the test function. Under this view, the correspondence between the Smolyak level, l_S, number of evaluation points, M, and maximum achievable Polynomial Chaos order is provided in Table 2.

It should be noted that, while the performance of the two approaches (standard PC and CS-PC) is evaluated for a similar number of evaluation points, those points are not the same for the two methods: the Smolyak-based PC approach relies on points issued from a partially tensorized 1-dimensional Gauss-Patterson quadrature (nested) rule while the points for the CS-like strategy are sampled at random in $[-1, 1]^2$ following a uniform joint-probability law.

Remark While the approximation basis can grow (increasing N_o and $P_{\boldsymbol{\xi}}$) when M increases for the PC approach as the Smolyak scheme allows to exactly integrate integrands of growing order, see Table 2, in the CS-PC approach, the approximation basis is fixed for all M, $N_o = 12$, $P_{\boldsymbol{\xi}} = 91$.

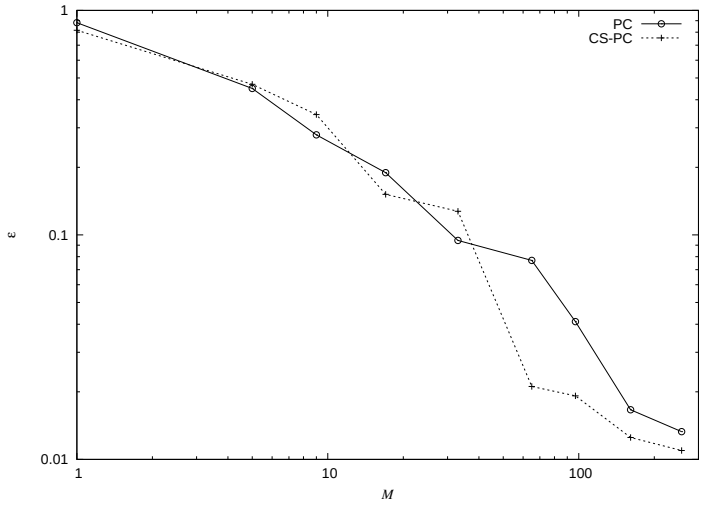

Fig. 3. Convergence analysis of the CS-like (CS-PC) and the Smolyak quadrature (PC) solution strategies in terms of the relative error norm ε.

The evolution of the approximation error ε as a function of the number of deterministic solver calls M is plotted in Fig. 3. The error norm is seen to decrease as M increases, both approaches roughly achieving a 2-order of magnitude error reduction from a 1-point to a 257-point evaluation. The CS-like approach is seen to perform no better than Smolyak-based PC when the number of evaluation points is low. This poor behavior when a limited number of points is available results from the recovery properties as it was seen that a minimum of points was necessary for the RIP constant to drop sufficiently low to achieve recovery of a signal with given sparsity.

Conversely, when the number of evaluation points becomes sufficiently large, the error norm from the CS-like strategy drops dramatically, achieving a more accurate approximation than Smolyak-PC beyond $M \simeq 40$ evaluation points. A remarkable result is that a 65-point CS-like approach achieves almost the same accuracy as a 161-point Smolyak strategy. It must be emphasized that this is achieved without requiring any prior knowledge of the solution, nor making use of a trial-and-error refinement strategy. The underlying compressibility in the solution representation in the given trial basis is intrinsically captured by the procedure which makes the best use of the available M points information in a non-adaptive way. The global behavior is consistent with what was expected from Section 3: relatively poor performance compared to the Smolyak-PC approach for a low number of available measurements but a better convergence rate once the ensemble of solution evaluations gets sufficiently large. It is interesting to note that, when $M = 257$, the Smolyak level

is $l_S = 12$ so that the two methods use the same trial basis ($P_{\boldsymbol{\xi}} = 91$) and their performances are seen to be roughly similar while they rely on the solution of different problems.

The "exact" response surface is plotted in Fig. 4 both from extensive Monte-Carlo simulations and from using the CS-PC approach. The agreement is satisfying but the "exact" response surface (Monte-Carlo) is seen to exhibit a slope discontinuity near $\xi_2 \simeq 0.55$: the solution is not smooth in the stochastic domain and is then poorly approximated with polynomials. This is responsible for the rather slow convergence of the L^2-error observed in Fig. 3. A plot similar to that from the CS-PC strategy is obtained with the standard PC (not shown).

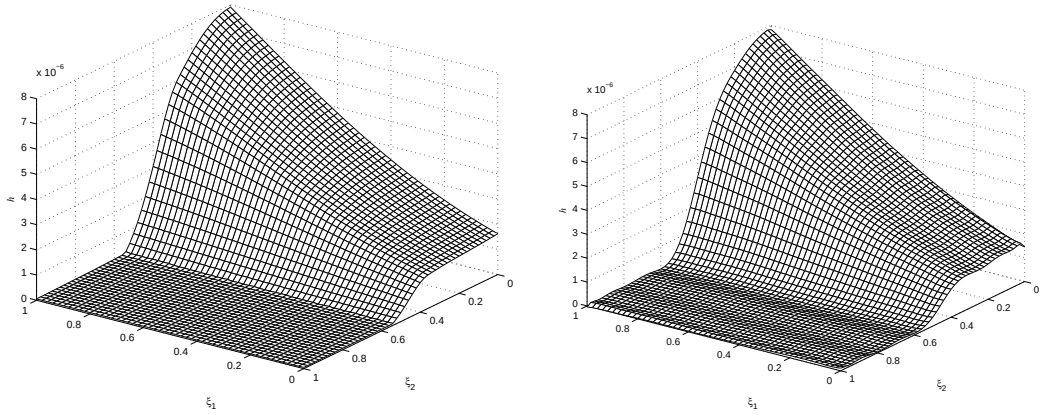

Fig. 4. Exact (left, extensive MC-based simulations) and CS-like approximation (right) of the response surface of the stochastic problem output.

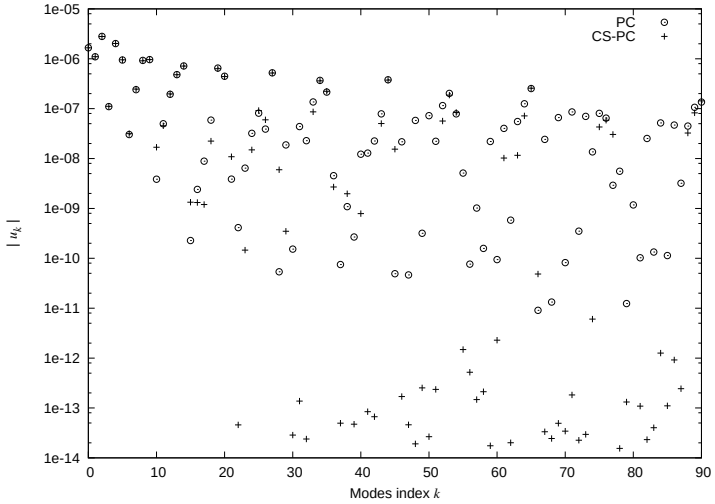

Fig. 5. 257-evaluation point solution approximation spectrum for the PC and CS-PC strategies.

To appreciate the recovery of the coefficients vector $\boldsymbol{X}$, its spectrum from both the Smolyak- and the CS-like approach with $M = 257$ is plotted in Fig. 5. As expected, the major contribution to the solution arises from a small set

of modes while a significant part of the spectrum exhibits a much weaker magnitude. Interestingly, the CS-like spectrum closely matches that of the Smolyak-based solution for the most significant modes, say those with magnitude larger than 10^{-7}. No such match is achieved for the lower magnitude modes and the CS-like solution exhibits more vanishing or negligible modes, in particular in the upper part of the spectrum. This is a clear demonstration of the philosophy of the CS-like approach: concentrate on a few modes that contribute most to the solution approximation and discard or ignore the others unless additional information is provided.

6.3 Higher dimensional stochastic problem

6.3.1 Settings

We now consider a more realistic case where additional sources of uncertainty are present, calling for a higher dimensional Polynomial Chaos basis for a good approximation of the stochastic output and inducing a large computational cost for evaluating the solution at the resulting large number of necessary sampling points.

The problem of interest is basically the same as in the previous section but the sources of uncertainty are more precisely modeled, giving rise to additional stochastic dimensions that must be taken into account. In particular, the depth field is now modeled with a $N_H = 5$-mode expansion instead of $N_H = 1$ as previously considered. This allows not-so-small contributing eigenmodes of the correlation kernel to be taken into account. Further, the seaquake source model is also improved with respect to its intensity $A_{\mathcal{N}}$ as well as its width σ_{S_h} which are now modeled as uncertain quantities in addition to the location $\boldsymbol{x}_{S_h}$ as previously considered. This leads to a source model lying in a 3-dimensional stochastic space: $\widetilde{\boldsymbol{\xi}} \in \Omega_{\widetilde{\Xi}} \subset \mathbb{R}^3$. The resulting uncertain problem therefore is an 8-D stochastic problem: $\boldsymbol{\xi}^T = \left(\widehat{\boldsymbol{\xi}}^T \, \widetilde{\boldsymbol{\xi}}^T\right) \in \Xi \subset \mathbb{R}^8$. This moderately large dimensionality framework is routinely encountered in practice and thus constitutes a test case of stronger practical interest than the previous 2-D case which only serves as a didactic example.

The correspondence between the Smolyak scheme level and the maximum polynomial order that can correspondingly be considered for the output and the test function, within the assumption that the output is a polynomial of the same order as the test function, is given in Table 3. For a polynomial approximation order $N_o = 8$, the required number of quadrature points is $M = |\mathcal{N}_q| = 97153$ and will result in a large computational burden for evaluating the corresponding model outputs.

l_S	1	2	3	4	5	6	7	8
M	17	129	609	2177	6657	17921	43137	97153
N_o	1	2	3	4	5	6	7	8
$P_{\boldsymbol{\xi}}$	9	45	165	495	1287	3003	6435	12870

Table 3
Correspondence between Smolyak level l_S, number of evaluation points $M = |\mathcal{N}_q|$, maximum Polynomial Chaos order N_o and related number of stochastic modes $P_{\boldsymbol{\xi}}$. 8-D stochastic space.

6.3.2 Convergence properties

As mentioned in section 4.4.2, the compromise between overfitting and mismatch with the measurements is essentially driven by the τ parameter and its value is adjusted through a cross-validation approach whenever new observations become available. This step takes advantage of warm-start capabilities of the optimization technique and results in an efficient procedure. An illustration of the evolution of the cross-validation error $\varepsilon^{CV} \equiv \sum_l \frac{1}{card\left(\boldsymbol{u}^{(l)}\right)} \left\| \boldsymbol{u}^{(l)} - \Psi^{(l)}\, \boldsymbol{X} \right\|_2$ is given in Fig. 6 for various values of τ. When τ is low, the constraint of matching the observations $\boldsymbol{Y}$ is weak and a very sparse coefficient vector $\boldsymbol{X}$ is promoted, leading to a large cross-validation error. On the contrary, when τ is large, the recovery algorithm tends to match the measurements disregarding the resulting sparsity of $\boldsymbol{X}$, leading to overfitting and a large cross-validation error as well. There thus exists a compromise $\tau^\star$ between these two extremes leading to a minimal ε^{CV} and the model output approximation is then given by the solution of Eq. (44) with $\tau = \tau^\star$.

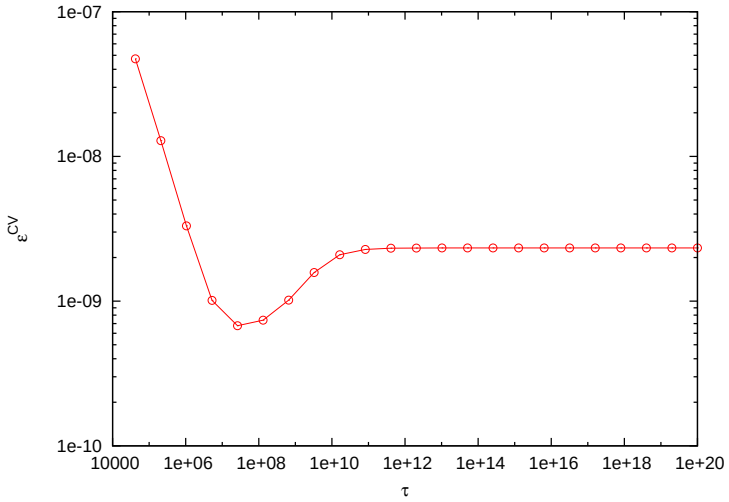

Fig. 6. Evolution of the cross-validation error ε^{CV} when τ varies. 8-dimensional case, $M = 17921$.

A similar investigation of the approximation convergence as for the two-dimensional case is now carried-out. As before, the error norm of the approx-

imation is monitored when the number of the available measurements varies, Fig. 7. Beyond a minimal number of solution evaluation points, $M \gtrsim 30$, the CS-like PC approach is again seen to achieve a better approximation of the stochastic solution than the Smolyak-based PC. In this case, the CS-like approach requires about 10^3 points to approximate the solution within a 10^{-4} relative error L^2-norm while the sparse grid technique needs 10^5 points to reach the same accuracy, roughly achieving a two order of magnitude improvement in terms of computational burden. Note that the physical location $\boldsymbol{x}^\star \in \Omega_{\boldsymbol{x}}$ used in the definition of the output, see Eq. (67), is different in the 8-D case from the 2-D case. The location considered in the 8-D case leads to a smoother response surface, hence a higher convergence rate of the polynomial approximations.

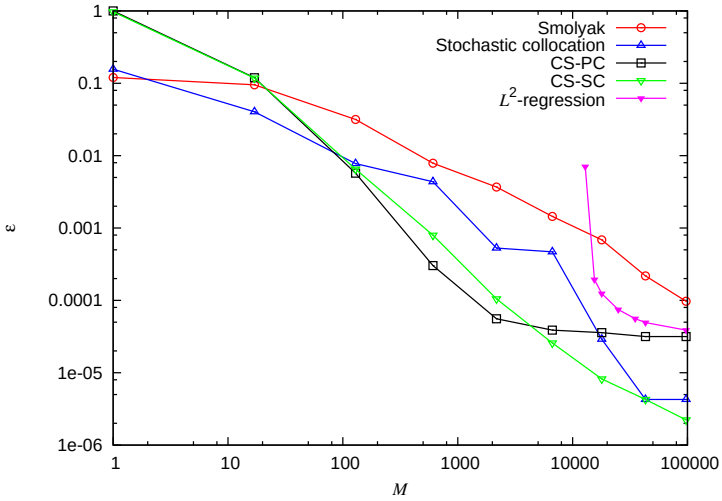

Fig. 7. Convergence analysis of the Polynomial Chaos CS-like (CS-PC) and the Smolyak quadrature (PC) solution strategies together with the Stochastic Collocation (SC) and CS-SC in terms of L^2-error norm. 8-dimensional problem.

Remark The number of samples M reported in the plots corresponds to that used to form the sensing matrix A and the measurement vector $\boldsymbol{Y}$. The actual cost of the CS-UQ method also includes the additional samples used for the cross-validation step. With the 3-fold cross-validation strategy used here, the actual cost is then 33 % higher than M.

In addition to the comparison with the Smolyak scheme-based PC, it is of interest to appreciate the performance of the present CS strategy with a least-squares regression whose resulting approximation error is also plotted in Fig. 7. The regression problem admits a unique solution whenever $M \geq P_{\boldsymbol{\xi}}$ (otherwise the Fisher information matrix is not invertible) so the plot is given for $M \geq 12870$. The approximation error is seen to be large when the number of measurements is only a little larger than the number of unknowns. When M further increases, the approximation gets much better and eventually gives

a similar accuracy as the CS approach when M approaches 10^5. Hence, the L^2-regression approach cannot provide a solution with less than M measurements and, when a solution can be determined ($M \geq P_{\boldsymbol{\xi}}$), its accuracy is quite poor unless M gets really large. This observation illustrates the remark made in section 4.1 about the poor recovery property of choosing $\mathcal{I}$ such that $\mathcal{I}_m u = \ll u, \psi_m \gg$.

To complete the picture, an alternative trial basis is also considered here: the so-called Stochastic Collocation (SC) approach, (Mathelin et al., 2005; Xiu and Hesthaven, 2005; Babuška et al., 2007), essentially consists of approximating the response surface with a linear combination of point values belonging to a given set $\mathcal{S}$. More specifically, the stochastic output is approximated under the form

$$u(\boldsymbol{\xi}) \approx \sum_{q \in \mathcal{S}} u\left(\boldsymbol{\xi}^{(q)}\right) \widehat{\psi}_q(\boldsymbol{\xi}), \tag{70}$$

where the functions $\widehat{\psi}_q$ are typically taken as Lagrange polynomials. Instead of a full tensorization of one-dimensional Lagrange polynomials that would require a prohibitive number of evaluation points $\boldsymbol{\xi}^{(q)}$, the stochastic collocation approach makes use of successive partial tensorizations based on the Smolyak scheme points, here chosen to be nested. It results in linear combinations of hierarchical approximations involving a reasonable number of evaluation points. For a detailed presentation of the method, one can refer to Xiu and Hesthaven (2005) and Nobile et al. (2007).

Since the SC method is essentially an interpolation technique, it does not involve a projection step and, for a given number of point-wise evaluations of the response surface, it allows for approximating the output with a higher polynomial order as compared with the Smolyak-Polynomial Chaos, which is usually used within the assumption that the output surface to approximate is of the same polynomial order, N_o, as the retained trial basis. The SC approach is not affected by such a hypothesis. In the present case, it allows for polynomials of total order 14 as opposed to 8 for the Polynomial Chaos approach.

The hierarchical stochastic collocation method used builds up an approximation by successively adding details to the approximation at the preceding level, see for instance Ganapathysubramanian and Zabaras (2007). The SC-based CS approach used here then tends to select the dominant terms in this normalized hierarchical Lagrange basis.

In Fig. 7, the stochastic collocation approach is seen, in particular, to provide a better approximation than PC for a given number of evaluations M thanks to the higher affordable polynomial order. For a large M, the higher polynomial approximation leads to a dramatic improvement in the approximation quality as the resulting error norm is about 30 times lower with SC than with PC. The CS-counterpart of the stochastic collocation (CS-SC) again achieves a

significant improvement over the regular SC approach for $M \gtrsim 100$, reaching up to an order of magnitude improvement in the approximation error for a given evaluation cost ($\propto M$). Just as with the PC approach, the approximation accuracy of the regular and the CS-like stochastic collocation approaches are similar when M is maximum, *i.e.*, the SC-approach can use the trial basis with maximum order.

For $M \gtrsim 2000$, the approximation error is seen to reach a plateau and further measurements essentially do not improve the recovery accuracy. This comes from the fact that the identified coefficient vector $\boldsymbol{X}$ is already very close to $\left(\langle u, \psi_1 \rangle_{L^2(\Xi,\mu_\Xi)}, \ldots, \left\langle u, \psi_{P_{\boldsymbol{\xi}}} \right\rangle_{L^2(\Xi,\mu_\Xi)}\right)^T$ for $M \simeq 2000$. To further improve the approximation, an enhanced trial basis $\{\psi_{\boldsymbol{\alpha}}\}$ would be necessary, for instance, increasing the polynomial order N_o. However, the recovery property of the sensing matrix A deteriorates as the cardinality $P_{\boldsymbol{\xi}}$ of the approximation basis grows (albeit slowly, not shown). In a nutshell, the maximum correlation of a set $\mathcal{S}$ of vectors defining a frame in a given $\mathbb{R}^K$ increases when $|\mathcal{S}|$ increases. One should then exercise moderation when choosing the trial basis and refrain from using an unnecessary large one.

6.3.3 Solution spectrum

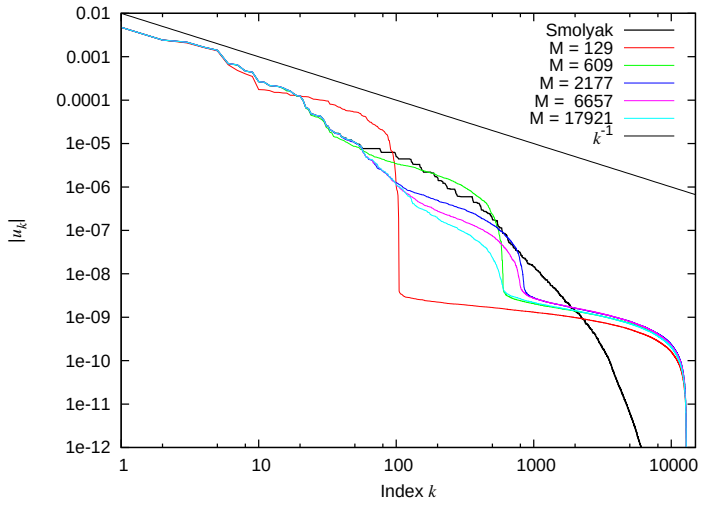

Fig. 8. Sorted spectrum of the approximation. The 97153-point Smolyak-scheme solution is compared with CS-PC approximations determined with various M. The k^{-1} decay is also plotted for comparison.

The spectrum of the approximation coefficient vector $\boldsymbol{X}$, sorted by magnitude, is plotted in Fig. 8 both for the Smolyak-scheme-based approximation (with $M = 97153$) and the present CS-PC approach with various M. While none is ground truth in the sense they all are approximations, their sorted spectra are however seen to exhibit a decay which rate is essentially bounded from above by k^{-1}, therefore *a posteriori* giving confidence that the CS results apply, see

section 3.4. As an illustration of the strongly compressible character of the signal in the $\{\psi_{\boldsymbol{\alpha}}\}$ basis, the top two decades of the sorted spectrum only involves about 20 modes.

While the approximations derived from all the cases plotted in the figure allow accurate retrieval of the dominant modes, the $M = 129$ CS-PC spectrum is seen to "diverge" from the other spectra for $k \gtrsim 7$. Similarly, the $M = 609$ CS-PC spectrum diverges for $k \gtrsim 23$. This is a clear illustration of the CS-UQ approach: focus all the available information, no matter how little, to retrieve the most significant modes and disregard the others. While using a set of $M = 97153$ points, the Smolyak-scheme approximation spectrum is seen to diverge for $k \gtrsim 58$ from the $M = 17921$ CS-PC case, taken here as the reference thanks to its approximation accuracy, see Fig. 7. This difference clearly indicates that the signal to approximate is not a N_o-th order polynomial and that its projection in the $\{\psi_{\boldsymbol{\alpha}}\}$ basis is only approximately evaluated by the quadrature, Eq. (10).

As already mentioned, the evaluation points $\boldsymbol{\xi}^{(q)}$ are not the same for the Smolyak scheme and the CS approach. This brings considerable flexibility in the applicability of the CS-UQ method as it can be applied to situations where one has little control on how the realizations of the uncertain parameters are sampled: as long as the output evaluations at one's disposal are such that the resulting recovery properties defined in Eqs. (19, 20, 22) are satisfied, the CS results are valid. This is a distinguishing feature compared to the standard PC or SC approach where evaluation points are *a priori* defined. Further, it also brings flexibility in the number of points: while the Smolyak-based approach is restricted to a given set of points for each level, *cf.* Table 3, the CS-based UQ can accommodate with any number of points and improves the recovery whenever M increases.

6.3.4 Improved recovery

When the unknown response surface is probed with a numerical solver, one often has the ability of choosing the set of samples. In this last section, the focus is on the alternative sets of samples discussed in section 4.3. The accuracy of the reconstructed surface based on five independent sets of points chosen uniformly at random, as considered so far in this work, and the set resulting from a Sobol sequence are compared in terms of approximation residual L^2-norm. Further, since the Legendre polynomials generating the approximation basis considered in this work belong to the, partially tensorized, Jacobi polynomials family, $P_{N_o}^{(\alpha,\beta)}(\xi)$ with $\alpha = \beta = 0$, theoretical results presented in section 4.3 apply and provide a framework with provable recovery performance. Three sets of samples drawn according to the Chebyshev probability measure on $[-1, 1]^8$ are considered.

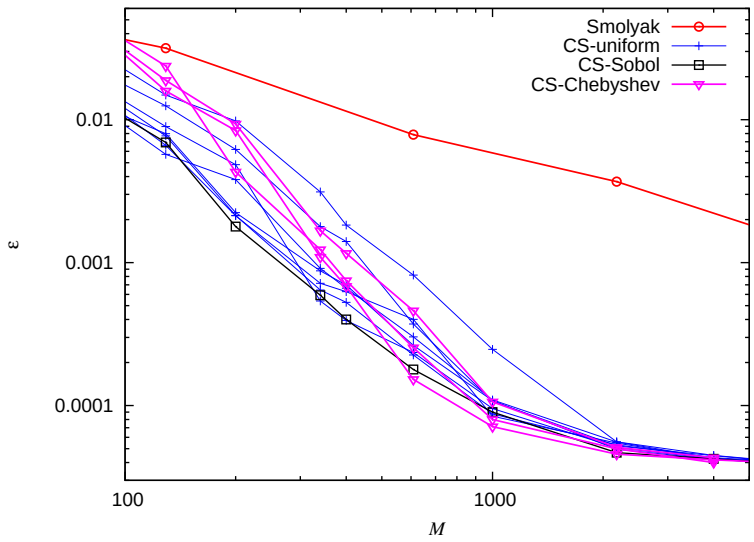

Fig. 9. Convergence of the unknown response function recovery for randomly chosen (uniform and Chebyshev probability measure-) and a Sobol sequence-based samples. 8-dimensional problem.

Figure 9 gathers recovery results for different size M of the sets. As already discussed in previous sections, the recovery is poor for low M but strongly improves when the number of available observations increases. When M gets large enough, $M \gtrsim 4000$, the recovery saturates (in the sense that no better approximation can be derived in the retained approximation basis) and both the uniform and Chebyshev measure-based sampling strategies lead to an excellent recovery. They are seen to exhibit a similar behavior and achieve comparable recovery accuracy for a given M. However, they rely on a random sampling and their performances are subjected to a large variability and hence low reliability. For instance, for $M = 500$, the L^2-norm of the residual with samples drawn uniformly at random varies from 3×10^{-4} to more than 1×10^{-3} for the five sets considered here.

Similarly to the uniform and Chebyshev, the Sobol sampling scheme achieves poor recovery when the number of observations is low. However, when M gets larger, in addition to the direct improvement due to the larger set of available observations, the 8-D integral underlying in the residual L^2-norm is better approximated with the finite sum, see Eq. (41), and the Sobol sequence strategy almost always exhibits a better accuracy in the approximation than that achieved by the randomly sampled points. Again, when M becomes large enough, all strategies achieve excellent recovery. There thus exists a range of sample set size within which the response surface recovery can be improved by carefully chosen samples over a naive random sampling strategy. This range precisely corresponds to the range of interest where the resulting approximation is already decently accurate while not requiring a prohibitive number of response surface probings.

Beyond its brute recovery performance, the crucial point is that a low-discrepancy sequence such as Sobol relies on a deterministic set so that its good recovery properties are not subjected to variability, in contrast to both the uniform- and Chebyshev-measure sampling. Further, it is important to note that these nice properties come at no additional cost for a given M since the L^1-minimization problem remains of the same size, simply relying on a different set of samples.

7 Concluding remarks and perspectives

In this paper, we proposed a novel technique for quantifying the uncertainty associated with the solution of a mathematical model involving stochastic parameters. This approach makes use of a deterministic solver and allows for the direct reuse of any existing legacy code that is run with different sets of input parameters. It heavily relies on concepts borrowed from the technique of compressed sensing and essentially consists of retrieving the most significant modes of the approximated solution from a minimal number of code calls. Since solving the deterministic problem is almost always the bottle-neck of any uncertainty quantification method, reducing the number of required calls to the solver is the route to higher computational efficiency. Rigorous results exist in the literature that prove that this methodology succeeds, with an overwhelming probability, in deriving a good approximation of the solution, provided it has a compressible enough representation in the trial basis considered. The core principles of this approach have been shown to immediately apply to the stochastic framework and ways of achieving good recovery performance were proposed.

The methodology was applied to the uncertainty quantification of an uncertain Shallow Water problem. The response surface of the uncertain surface height at a specific time and location was approximated with Polynomial Chaos. The proposed approach was shown to perform well, both in a 2-D and 8-D stochastic framework as compared to the usual sparse grid projection technique (Smolyak cubature). In particular, the proposed approach may achieve several orders of magnitude improvement in the approximation error L^2-norm over the Smolyak scheme PC for a comparable CPU cost. It was also shown to compare favorably to a hierarchical Stochastic Collocation strategy on the 8-D problem. In fact, this *non-adaptive* approach takes advantage of the weak dependence of the solution on certain modes of the representation basis and uses every bit of available information to estimate the dominant modes, *and only them*. This philosophy is at the root of the method's efficiency.

The benefit of carefully chosen samples over a naive sampling strategy was also shown. A low-discrepancy Sobol sequence was compared to samples drawn from a uniform and a Chebyshev probability measure. The Sobol sequence

achieves performances often at least as good as the random sampling while being deterministic. This, sample-wise, zero-variability brings reliability to the recovery procedure and, based on the present results, an information operator associated to a low discrepancy Sobol sequence is the suggested strategy, whenever possible.

Another key to the method's efficiency is that it allows the use of a large approximation basis even when very little information is available, while more standard methods, like PC, are limited in the sense that the number of available constraints strongly drives the basis one can rigorously use to generate an approximation. In the 8-D example we considered, about 600 deterministic solves where sufficient to lead to good accuracy in the CS-PC case with a $N_o = 8$-basis while a Smolyak-scheme approach was then reasonably limited to an approximation in a $N_o = 3$-basis.

With the CS-UQ approach, one may then want to enhance the trial basis by incorporating as many modes as possible. In particular, in addition to the usual Polynomial Chaos, compact support and/or non-smooth functions may also be included to improve the approximation when the solution is poorly represented by polynomials, *e.g.*, when it exhibits discontinuities in the stochastic space. To some extent, the mild, about $\log^4(P_{\boldsymbol{\xi}})$, dependence of the RIP constant or the mutual coherence of the sensing matrix on the size $P_{\boldsymbol{\xi}}$ of the trial basis seems to encourage such an approach while moderation should however be exercised. In addition to an over-complete dictionary, specific reconstruction properties may be desirable such as minimal total variation of the approximated solution for noisy and/or discontinuous response surface. Finally, the conclusions drawn in this work rely on a specific UQ configuration (8-D uncertain Shallow Water Equations). Supplementary testing with different application problems are desirable to assess their universality. Further, a detailed theoretical analysis of the performance of low-discrepancy sequences for CS-based UQ should be carried-out to reveal crucial properties and help designing better sampling strategies. These developments are the subject of ongoing efforts.

Acknowledgement

This work is supported by the French National Agency for Research (ANR) under projects ASRMEI JC08#375619 and CORMORED ANR-08-BLAN-0115 and by GdR MoMaS. The first author has also benefited from inspiring discussions with André Veragen and the second author from similarly inspiring discussions with Edwin Jimenez. This work started during a visit by the first author to the Computational Science and Engineering group of M.Y. Hussaini at Florida State University in April 2009. Both authors acknowledge and thank him for his support and encouragement over the last several years.

References

Abramowitz, M., Stegun, I., 1970. Handbook of mathematical functions. Dover.

Babuška, I., Nobile, F., Tempone, R., 2007. A stochastic collocation method for elliptic partial differential equations with random input data. SIAM J. Numer. Anal. 45 (3), 1005–1034.

Becker, S., Bobin, J., Candès, E., 2009. Nesta: a fast and accurate first-order method for sparse recovery. Tech. rep., Caltech Institute of Technology.

Blanchard, J., Cartis, C., Tanner, J., 2010. Compressed sensing: how sharp is the restricted isometry property ? SIAM reviewTo appear.

Cai, T., Wang, L., Xu, G., 2009a. New bounds for restricted isometry constants. Tech. rep., Massachusetts Institute of Technology.

Cai, T., Xu, G., Zhang, J., 2009b. On recovery of sparse signals via ℓ_1 minimization. IEEE Trans. Inf. Theory 55, 3388–3397.

Cameron, R., Martin, W., 1947. The orthogonal development of non-linear fonctionals in series of fourier-hermite fonctionals. Ann. Math. 48 (2), 385–392.

Candès, E., Plan, Y., 2007. Near-ideal model selection by ℓ_1 minimization. Annals of Statistics 37, 2145–2177.

Candès, E., Romberg, J., 2006. Quantitative robust uncertainty principles and optimally sparse decompositions. Found. Comput. Math. 6 (2), 227–254.

Candès, E., Romberg, J., Tao, T., 2005. Stable signal recovery from incomplete measurements. Comm. Pure Appl. Math. 59, 1207–1223.

Candès, E., Tao, T., 2004. Near-optimal signal recovery from random projections: universal encoding strategies. IEEE Trans. Inform. Theory 52, 5406–5425.

Candès, E., Wakin, M., Boyd, S., 2007. Enhancing sparsity by reweighted ℓ_1 minimization. J. Fourier Anal. Appl. 17, 877–905.

Chen, S., Donoho, D., Saunders, M., 1999. Atomic decomposition by basis pursuit. SIAM J. Sci. Comput. 20, 33–61.

Deb, M., Babuška, I., J.T.Oden, 2001. Solution of stochastic partial differential equations using galerkin finite element techniques. Comput. Methods Appl. Mech. Eng. 190 (48), 6359–6372.

Donoho, D., 2006. Compressed sensing. IEEE Trans. Infor. Theo. 52 (4), 1289–1306.

Donoho, D., Elad, M., Temlyakov, V., 2006. Stable recovery of sparse overcomplete representations in the presence of noise. IEEE Trans. Infor. Theo. 52 (1), 6–18.

Donoho, D., Stark, P., 1989. Uncertainty principles and signal recovery. SIAM J. Appl. Math. 49 (3), 906–931.

Doostan, A., Owhadi, H., July 2010. A sparse approximation of partial differential equations with random inputs, presented at the SIAM Annual Meeting, Pittsburgh, PA, USA.

Douglas, C., Haase, G., Iskandarani, M., 2003. An additive schwarz precon-

ditioner for the spectral element ocean model formulation of the shallow water equations. Elec. Trans. Numer. Anal. 15, 18–28.

Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., 2004. Least angle regression. Annals of Statistics 32, 407–499.

Efron, B., Tibshirani, R., 1997. Improvements on cross-validation: the .632+ bootstrap method. J. Amer. Stat. Assoc. 92 (438), 548–560.

Figueiredo, M., Nowak, R., Wright, S., 2007. Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems. IEEE Journal of Selected Topics in Signal Processing 1, 586–597.

Foucart, S., 2010. A note on guaranteed sparse recovery via ℓ_1-minimization. Applied and Computational Harmonic Analysis 29 (1), 97–103.

Frauenfelder, P., Schwab, C., Todor, R., 2005. Finite elements for elliptic problems with stochastic coefficients. Comput. Meth. Appl. Mech. Engrg. 194 (2–5), 205–228.

Ganapathysubramanian, B., Zabaras, N., 2007. Sparse grid collocation methods for stochastic natural convection problems. J. Comput. Phys. 225, 652–685.

Ghanem, R., Spanos, P., 1991. Stochastic finite elements. A spectral approach, rev. Edition. Springer Verlag, 222 p.

Gilbert, J., Lemaréchal, C., 1989. Some numerical experiments with variable-storage quasi-newton algorithms. Math. Program. 45, 407–435.

Iskandarani, M., Haidvogel, D., Boyd, J., 1995. A staggered spectral element model with application to the oceanic shallow water equations. Int. J. Num. Meth. Fluids 20 (5), 393–414.

Le Maître, O., Knio, O., Najm, H., Ghanem, R., 2004a. Uncertainty propagation using Wiener-Haar expansions. J. Comput. Phys. 197 (1), 28–57.

Le Maître, O., Najm, H., Ghanem, R., Knio, O., 2004b. Multi-resolution analysis of Wiener-type uncertainty propagation schemes. J. Comput. Phys. 197 (2), 502–531.

Mallat, S., Zhang, Z., 1993. Matching pursuit in a time-frequency dictionary. IEEE Trans. Signal Proc. 41 (12), 3397–3415.

Mathelin, L., Gallivan, K., July 2010. Uncertainty quantification for sparse solution of random pdes, presented at the SIAM Annual Meeting, Pittsburgh, PA, USA.

Mathelin, L., Hussaini, M., Zang, T., 2005. Stochastic approaches to uncertainty quantification in CFD simulations. Num. Algo. 38 (1), 209–239.

Mathelin, L., Le Maître, O., 2007. Dual-based a posteriori error estimate for stochastic finite element methods. Comm. App. Math. Comput. Sci. 2, 83–115.

Nobile, F., Tempone, R., Webster, C., 2007. An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data. SIAM J. Numer. Anal. 46 (5), 2411–2442.

Novak, E., Ritter, K., 1999. Simple cubature formulas with high polynomial exactness. Constructive Approximation 15, 499–522.

Petras, K., 2001. Fast calculation in the smolyak algorithm. Num. Algo. 26,

93–109.
Rauhut, H., 2010. Compressive sensing and structured random matrices. In: Fornasier, M. (Ed.), Theoretical Foundations and Numerical Methods for Sparse Recovery. Vol. 9 of Radon Series Comp. Appl. Math. deGruyter, pp. 1–92.
Rauhut, H., Ward, R., 2010. Sparse legendre expansions via ℓ_1-minimization20 p., preprint.
Smolyak, S., 1963. Quadrature and interpolation formulas for tensor products of certain classes of functions. Dokl. Akad. Nauk. SSSR 4, 240–243.
Sobol, I., 1967. Distribution of points in a cube and approximate evaluation of integrals. USSR Comput. Maths. Math. Phys. 7, 86–112.
Sobol, I., 1977. Uniformly distributed sequences with an additional uniform property. USSR Comput. Maths. Math. Phys. 16, 236–242.
Soize, C., Ghanem, R., 2004. Physical systems with random uncertainties: chaos representations with arbitrary probability measure. SIAM J. Sci. Comput. 26 (2), 395–410.
Taylor, H., Bank, S., McCoy, J., 1979. Deconvolution with the l_1-norm. Geophysics 44, 39–52.
Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. J. Royal Statist. Soc. B 58, 267–288.
van den Berg, E., Friedlander, M., 2008. Probing the pareto frontier for basis pursuit solutions. SIAM J. Sci. Comput. 31 (2), 890–912.
Wan, X., Karniadakis, G., 2005. An adaptive multi-element generalized polynomial chaos method for stochastic diffential equations. J. Comput. Phys. 209, 617–642.
Wiener, N., 1938. The homogeneous chaos. Amer. J. Math. 60 (4), 897–936.
Wright, S., 1997. Primal-Dual Interior-Point Methods. SIAM Publications.
Xiu, D., Hesthaven, J., 2005. High-order collocation methods for differential equations with random inputs. SIAM J. Sci. Comput. 27, 1118–1139.
Xiu, D., Karniadakis, G., 2002. The Wiener-Askey polynomial chaos for stochastic differential equations. SIAM J. Sci. Comput. 24 (2), 619–644.
Xiu, D., Karniadakis, G., 2003. Modeling uncertainty in flow simulations via generalized polynomial chaos. J. Comput. Phys. 187, 137–167.
Zibulevsky, M., Elad, M., 2010. L1-L2 optimization in signal and image processing. IEEE Signal Proc. Mag. 27 (3), 76–88.