

# Load field estimation based on Principal Component Analysis

D. Dessi

INSEAN

National Research Council, Rome, Italy

## ABSTRACT

In this paper, a new technique for obtaining a close spatial sampling of a load field on the basis of a few global measurements is presented. This technique is based on a combination of Principal Component Analysis (PCA) and polynomial spline approximation under integral constraints and is herein illustrated with respect to the experimental prediction of the wave load distribution of a ship. For this case, preliminary numerical results are reported in the paper. The input data are the time-histories of the vertical forces acting on the segments of a virtual segmented-hull model and are obtained by spatially integrating the sectional forces provided with a strip-theory approach over the length of each segment. The relative motion of the segmented-hull is that effectively experienced in seakeeping tests by the corresponding physical model. After application of the Principal Component Analysis on the global loads, spatial basis functions are identified and their values provide the constraints for the splines approximating the spatial basis functions in the continuous case, *i.e.*, the longitudinal distribution of sectional forces.

## INTRODUCTION

Recently, Mariani and Dessi [1] have proposed an application of the Proper Orthogonal Decomposition, a technique derived from the fundamental work of Karhunen and Loeve ([2]-[3]), to investigate the modal parameter of the wet-modes of floating bodies. In this paper, the same technique is used to investigate a different phenomenon which the concept of modes - as it is known in structural mechanics - does not apply to. For this reason, to avoid any ambiguity with the previous problem, the term Principal Component Analysis has been preferred to label this technique as it is done in statistics, where a set of data comes with no reference with the underlying physical, economical or social phenomenon.

The representation of physical phenomena in spatial manifolds through the knowledge of the values of the associated physical variables in a limited number of points is a problem pertaining to approximation theory. This problem is appealing in the case of floating bodies as well, where for instance a pressure field  $p(\mathbf{x})$ , with  $\mathbf{x} \in \mathcal{R}^3$  the coordinate vector, is measured in a few points but one is interested to reconstruct the entire scalar field  $f(\mathbf{x})$ . The problem becomes computationally expensive if the considered field is also time-dependent. In this case, a set of approximation problems, one for each time step, say  $f(\mathbf{x}, t_i)$ , has to be solved. However, if the following decomposition of the field  $f(\mathbf{x}, t)$  holds:

$$f(\mathbf{x}, t) = \alpha_1(t) \phi_1(\mathbf{x}) + \alpha_2(t) \phi_2(\mathbf{x}) + \dots, \quad (1)$$

and if the number of terms to represent correctly the field is small, the problem can be greatly simplified. In fact, in this case, it would be sufficient to solve the approximation problem only for  $\phi_1(\mathbf{x})$ ,  $\phi_2(\mathbf{x})$  and so on.

In the present paper, a more cumbersome problem is raised by experimental investigations. Let us suppose that some vertical load measurements  $F_i$  are available via a segmented model, where  $i = 1, \dots, N_s$ , with  $N_s$  the number of the seg-

ments. They are represented as vectors applied to the connecting points of the  $N_s$  segments of the backbone beam in Fig. 1. We are interested to know the sectional force distribution  $f(x, t)$ ,  $x \in \mathcal{R}$  the longitudinal coordinate, depicted as a continuous line along the ship length in Fig. 1. The intensity  $F_i(t)$  of each vertical force is given by the integration of the field  $f(x, t)$  over the length of the segments and defines a vector process  $\mathbf{F}(t) = \{F_1(t), F_2(t), \dots, F_{N_s}(t)\}^T$ . If Eq. 1 holds for the field  $f(x, t)$ , it applies also to the discrete force distribution  $\mathbf{F}(t)$ , *i.e.*,

$$\mathbf{F}(t) = \beta_1(t) \mathbf{p}^1 + \beta_2(t) \mathbf{p}^2 + \dots, \quad (2)$$

where  $\mathbf{p}^1$ ,  $\mathbf{p}^2$  and so on are vectors to be determined. Objective of this paper is to show that, exploiting the relationship between the shape functions  $\phi_j(x)$  and the vectors  $\mathbf{p}^i$ , an approximation problem for  $\phi_j(x)$  can be conveniently stated and solved with spline approximation and, finally, the field  $f(x, t)$  can be reconstructed. The proposed procedure is based massively on Principal Component Analysis (the discrete form of the Karhunen-Loeve decomposition, also known as Proper Orthogonal Decomposition), that is exploited to determine these invariants or basis functions of the load process. This new technique is applied to a virtual experiment where the ship motion is given from seakeeping tests and the resulting sectional load distribution  $f(x, t)$  is calculated with a strip-theory approach and integrated to provide the time-histories of the segment forces  $F_i(t)$ . The reason, in this preliminary application, to use a realistic but numerical experiment is to avoid experimental uncertainties and to have a reference continuous distribution to assess the outlined procedure. In fact, performing a decomposition of both  $f(x, t)$  and  $\mathbf{F}(t)$ , the spline approximations  $S^{(j)}(x)$  of the basis functions can be compared with the original  $\phi_j(x)$ .

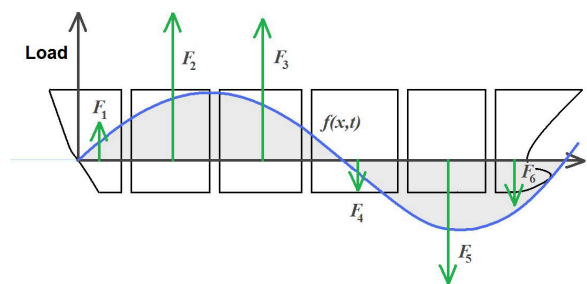


Fig. 1 Segment forces (vectors) and force per unit length (continuous line).

## TECHNIQUE OUTLINE

Let us suppose that  $f(x, t)$  is a scalar, real-valued, time-independent field, with  $x \in \Omega \subset \mathcal{R}$ , which has to be approximated on the basis of its integral values  $F_1(t), F_2(t), \dots, F_{N_s}(t)$ , defined in the subdomains  $\Omega_j$ , such as  $\Omega = \Omega_1 \cup \Omega_2 \cup \dots \cup \Omega_{N_s}$ . In the present case,  $f(x, t)$  is the force per unit length representing the vertical sectional load acting on a ship, being  $x$  the coordinate which denotes the section position along the ship length, ranging in the interval  $\Omega = [0, L_{pp}]$ . The time-histories  $F_i(t)$  are experimental

or numerical global load estimations that represent the vertical forces acting on the segments of a segmented-hull, and the intervals  $\Omega_i = [x_{i-1}, x_i]$  define the wetted length of each segment (assumed to be constant), where the coordinates  $x_i$  denote the positions of the cuts between adjacent segments. Thus, the global loads  $F_i(t)$  and the local load function  $f(x, t)$  satisfy the following relationship:

$$F_i(t) = \int_{x_{i-1}}^{x_i} f(x, t) dx \quad (3)$$

for  $i = 1, \dots, N_s$ . The fundamental assumption is that the field  $f(x, t)$  admits the following decomposition:

$$f(x, t) = \sum_{j=1}^L \alpha_j(t) \phi_j(x), \quad (4)$$

where  $\phi_j(x)$  are suitable basis functions and  $L$  is the number of basis functions, with  $L \leq N_s$ . If Eq. 4 is substituted into Eq. 3, it yields

$$F_i(t) = \int_{x_{i-1}}^{x_i} \sum_{j=1}^L \alpha_j(t) \phi_j(x) dx. \quad (5)$$

Exchanging the positions of the integral operator and the summation symbol, and moving outside the integral the time functions  $\alpha_j(t)$ , one obtains:

$$F_i(t) = \sum_{j=1}^L \alpha_j(t) \int_{x_{i-1}}^{x_i} \phi_j(x) dx. \quad (6)$$

For sake of conciseness, let us define:

$$\Phi_{ij} = \int_{x_{i-1}}^{x_i} \phi_j(x) dx \quad (7)$$

that represents the integration over the  $i$ -th interval,  $\Omega_i = [x_{i-1}, x_i]$ , of the  $j$ -th basis function  $\phi_j(x)$ . In this way, Eq. 6 can be recast as:

$$F_i(t) = \sum_{j=1}^L \Phi_{ij} \alpha_j(t). \quad (8)$$

Introducing vector and matrix notation, with  $\Phi = [\Phi_{ij}]$ ,  $\mathbf{F}(t) = \{F_1(t), F_2(t), \dots, F_{N_s}(t)\}^T$  and  $\mathbf{a}(t) = \{\alpha_1(t), \alpha_2(t), \dots, \alpha_{N_s}(t)\}^T$ , Eq. 8 becomes

$$\Phi \mathbf{a}(t) = \mathbf{F}(t), \quad (9)$$

that, once the time  $t$  is fixed, represents a system of equations in the unknowns  $\alpha_j$ .

At each time instant  $t_k$ , depending on the sampling of the load time-histories  $F_i(t)$ , it is possible to solve the linear system 9 and obtain the time coordinates  $\alpha_j(t_k)$ . Therefore, the local load distribution  $f(x, t)$ , satisfying the experimental constraints given by Eq. 3, can be reconstructed, provided that the shape functions  $\phi_j(x)$  have been determined.

Indeed, the main question at this point concerns the calculation of the basis function  $\phi_j(x)$ . From a mathematical point of view, any set of independent functions may provide a regular matrix of coefficients required to solve the system. Nevertheless, such a choice does not indicate how many terms in Eq. 4 are necessary to represent with sufficient accuracy the load function  $f(x, t)$  and, for this reasons, it does not indicate how many global load evaluations or measurements are required to solve the system. Thus, it is convenient that the shape functions are defined in order to be physically consistent with the problem under investigation.

## GENERAL ISSUES ON PCA

Let us consider an  $N$  dimensional state-space with a generic process  $t \mapsto \mathbf{v}(t)$ . The Principal Component Analysis (also known as discrete Karhunen-Loeve decomposition or Proper Orthogonal Decomposition) defines in such a space a set of  $L \leq N$  vectors  $\{\mathbf{p}^h\}$ ,  $h = 1, \dots, L$  which determine an *optimal* basis for the linear representation of the process  $\mathbf{v} : \mathcal{R}^+ \mapsto \mathcal{R}^N$

$$\mathbf{v}(t) = \sum_{h=1}^L \zeta(t) \mathbf{p}^h \quad (10)$$

where the  $\zeta : \mathcal{R}^+ \mapsto \mathcal{R}$  are the dynamic components of the process referred to the POD basis.

The optimality is defined by searching the unknown direction  $\mathbf{p}$  such that the projection of  $\mathbf{v}(t)$  on  $\mathbf{p}$  be maximum in average on the time interval  $\bar{T} = [0, T]$ ,  $T \in \mathcal{R}^+$ . Thus, imposing as constraint that the vector  $\mathbf{p}$  has a unit magnitude, one has the constrained maximum problem

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T [\mathbf{v}(t) \cdot \mathbf{p}]^2 dt - \sigma(\mathbf{p} \cdot \mathbf{p} - 1) = \max_{\mathbf{p} \in \mathcal{R}^N} \quad (11)$$

with

$$\mathbf{p} \cdot \mathbf{p} = 1 \quad (12)$$

The steady condition for this quadratic form yields (for all  $r$  and considering the Einstein convention):

$$0 = \frac{\partial}{\partial p_r} \left[ \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T [v_i(t) p_i]^2 dt - \sigma p_i p_i \right] \quad (13)$$

Developing Eq. 13, one obtains

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T v_r(t) v_i(t) dt p_i = \sigma p_r \quad (14)$$

or

$$\mathbf{R} \mathbf{p} = \sigma \mathbf{p} \quad (15)$$

where

$$\mathbf{R} := \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \mathbf{v}(t) \otimes \mathbf{v}(t) dt \quad (16)$$

is the correlation matrix associated to the process  $\mathbf{v}(t)$ .

Equation 15 states that the PCA-basis vectors  $\mathbf{p}^h$  are the eigenvectors associated to the matrix  $\mathbf{R}$ , whereas the eigenvalues  $\sigma_i$  represent the energy associated to the projection of the process  $\mathbf{v}$  along the direction  $\mathbf{p}^h$ .

Let us consider the generic quadratic form associated to  $\mathbf{R}$ , i.e.,  $\forall \mathbf{x} \in \mathcal{R}^N$ :

$$\begin{aligned} r^2(\mathbf{x}) &:= \mathbf{x} \cdot \mathbf{R} \cdot \mathbf{x} = \mathbf{x} \cdot \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \mathbf{v}(t) \otimes \mathbf{v}(t) dt \cdot \mathbf{x} = \\ &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \mathbf{x} \cdot \mathbf{v}(t) \otimes \mathbf{v}(t) \cdot \mathbf{x} dt = \\ &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T [\mathbf{v}(t) \cdot \mathbf{x}] [\mathbf{v}(t) \cdot \mathbf{x}] dt = \\ &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T [\mathbf{v}(t) \cdot \mathbf{x}]^2 dt > 0 \end{aligned} \quad (17)$$

From Equations 17 the positiveness of matrix  $\mathbf{R}$  can be stated. Then,  $\mathbf{R}$  is then symmetric (by definition) and positive definite; resulting in real-valued and positive eigenvalues (*Principal Component Values*, PCVs) and real and orthogonal eigenvectors (*Principal Component Functions*, PCFs).

## PCA OF THE GLOBAL LOAD PROCESS

The Principal Component Analysis of the load process  $F(t)$  leads to define some integral constraints on the field  $f(x, t)$  through the consideration of their corresponding shape functions. Provided that  $\{p^1, p^2, \dots, p^L\}$  are the computed Principal Component Functions, it yields

$$F(t) = \sum_{h=1}^L \beta_h(t) p^h. \quad (18)$$

The physical meaning of the principal component vectors can be explained as follows. Equation 18 can be re-written as

$$F_i(t) = \sum_{h=1}^L \beta_h(t) p_i^h \quad (19)$$

Substituting Eq. 8 into the left-hand side of Eq. 19, one obtains:

$$\sum_{j=1}^L \alpha_j(t) \Phi_{ij} = \sum_{h=1}^L \beta_h(t) p_i^h. \quad (20)$$

Once  $j = h$  is set, the following fundamental assumptions are made:

$$\Phi_{ih} = p_i^h \quad \text{and} \quad L = N_s. \quad (21)$$

The second condition is necessary in order to let  $\Phi$  be a square matrix. Equations 21 imply, recalling Eq. 7, that

$$\int_{x_{i-1}}^{x_i} \phi_h(x) dx = p_i^h, \quad i = 1, \dots, N_s. \quad (22)$$

Equation 22 defines  $N_s$  integral constraints for the  $N_s$  unknown functions  $\phi_h(x)$ . Further conditions are provided by the continuity of the shape functions and its space derivatives throughout the domain  $\Omega$ . This observation suggests using splines for the solution of the approximation of the unknown functions  $\phi_h(x)$ . Let us introduce  $N_s$  spline functions  $S^{(h)}(x)$ , where the index  $h$  indicates the targeted basis function. Let us define with  $s_i^{(h)}(x)$  the expression of the  $h$ -th spline function over the interval  $\Omega_i$ . Therefore, the following set of conditions is assumed:

$$\begin{aligned} s_i^{(h)}(x_{i-1}) &= s_i^{(h)}(x_i) \\ ds_i^{(h)}(x_{i-1})/dx &= ds_i^{(h)}(x_i)/dx \end{aligned} \quad (23)$$

The introduction of further conditions is subjected to the choice of the degree of the spline polynomial, an aspect that will be discussed later in this paper.

## EXPERIMENTAL SET-UP

The model tests, providing the relative motion data, were carried out in the INSEAN towing tank basin. The linear basin is 220 m long, 9 m wide and 3.8 m deep. It is equipped with a single-flap wave-maker capable to generate regular and irregular wave patterns. As in standard seakeeping tests, the physical model is free to heave, to pitch and to surge. It is a fast-ferry with wedge-shaped sections. In every run the following physical quantities were measured: (i) the rigid-body degrees of freedom, (ii) the vertical forces at the legs connecting the segments to the backbone beam, (iii) the vertical bending moment on several beam sections and (iv) the wave height and the model speed. The vertical force is measured in 6 points by using load cells placed between the segments and the beam. These load cells then performs as supporting and connecting elements as well. The bending moment acting upon the beam is measured in 12 points by using strain gauges glued on the top and bottom faces of the beam. The calibration of the strain gauges was performed loading statically the

beam and comparing the theoretical bending moments with the voltage values. The acquisition system based on a National Instruments SCXI module recorded globally 28 signals at a 500 Hz sampling rate.

## LOAD DECOMPOSITION AND COMPARISON WITH A STRIP-THEORY APPROACH

In order to explore the usefulness of this technique, a virtual experiment is carried out to provide data similar to those that can be obtained in towing-tank experiments. From seakeeping tests the motion of a fast ferry sailing in regular waves is recorded. From the measurement of heave  $w_G$ , pitch  $\theta$  and absolute wave elevation  $h$ , the relative vertical motion  $w_r$  with respect to the free-surface and its derivatives is given as:

$$w_r(x, t) = w(x, t) + (x - x_G)\theta(t) - h(x, t). \quad (24)$$

Using Eq. 24 within a strip-theory approach (see Dessi and Mariani [5]), the vertical force can be roughly calculated for each transversal section. Thus, the Froude-Kryloff forces can be calculated, providing the vector of sectional vertical loads  $f(t) = \{f_1(t), f_2(t), \dots, f_M(t)\}^T$ , where each component is given as:

$$f_i(t) = f(x_i, t), \quad i = 1, \dots, M \quad (25)$$

with  $M = 38$  in the present calculations. It is worth to underline that  $f(t)$  is a vector of forces per unit length  $[N/m]$  whereas  $F(t)$  is a vector of vertical forces  $[N]$ . Applying the PCA on the vector process  $f(t)$ , the following PCVs  $\sigma_i$  are obtained:

PCVs	$\sigma_1$	$\sigma_2$	$\sigma_3$	$\sigma_4$	$\sigma_5$	$\sigma_6$
$f(t)$	89.58	7.56	1.65	0.62	0.49	0.12
$F(t)$	88.99	7.95	2.08	0.59	0.30	0.07

Table 1 Principal component values expressed as percentage values of the total energy.

The PCVs reported in Table 1 indicate that the first three PCFs are sufficient to collect the 98% of the ‘energy’ related to the vertical load. Therefore, it is reasonable to state:

$$f(t) \simeq a_1(t) q^1 + a_2(t) q^2 + a_3(t) q^3 \quad (26)$$

The global force acting on each segment can be obtained using the simple mid-point integration scheme, providing the following expression:

$$F_i = \sum_{l=H_{i-1}}^{H_i} f_l \Delta x_l. \quad (27)$$

The symbols  $H_i$  indicate the initial and final indexes (ranging between 1 and  $M$ ) of the strips belonging to the  $i$ -th segment and the intervals  $\Delta x_l$ , where the sectional load is supposed to be nearly constant, are obtained dividing the length of each segment by the corresponding number of subdivisions.

The PCA is then carried out also on the vector process  $F(t)$ , providing the PCVs reported in the second row of Table 1 and the PCFs  $q^h$  depicted in Fig. 2. The PCVs are quite close to those obtained for the sampled continuous distribution, *i.e.*,  $f(t)$ . This is due to the fact that the ‘energy’ of the signal is spread over a smaller number of PC functions, determining a different distribution especially at higher-orders. Thus, it holds again:

$$F(t) \simeq \beta_1(t) p^1 + \beta_2(t) p^2 + \beta_3(t) p^3 \quad (28)$$

Cubic splines are subjected to undesirable oscillations for the problem under investigation [4]. Therefore, quadratic splines

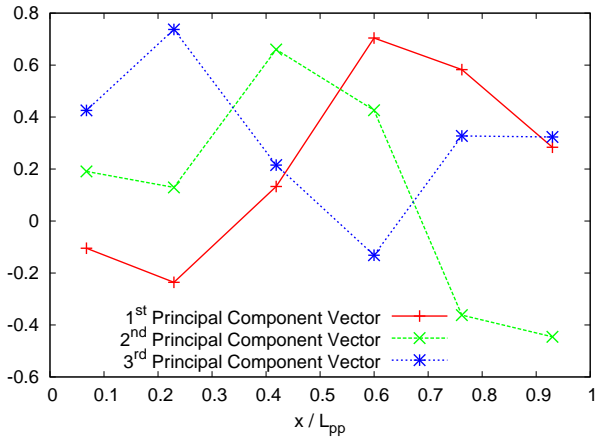


Fig. 2 Normalized Principal Component functions for the forces acting on the segments.

will be preferred for this specific problem. These polynomials are defined with  $3 N_s$  coefficients that require the same number of conditions. Thus, considering Eqs. 22 and 23,  $3 N_s - 2$  conditions can be set. The remaining two conditions can be established as boundary conditions at the contour of the domain  $\Omega$ , *i.e.*, in this case at the ends of the interval  $[0, L_{pp}]$ . Therefore, three quadratic splines (one for each identified shape function), given as a  $3 \times 6$  matrix of quadratic polynomials  $s_i^{(h)}(x)$ , are sought after, under the following constraints:

$$\int_{x_{i-1}}^{x_i} s_i^{(h)}(x) dx = p_i^h, \quad s_{N_s}^{(h)}(L_{pp}) = 0, \quad s_1^{(h)}(0) = \hat{s}^{(h)} \quad (29)$$

The condition  $s_{N_s}^{(h)}(L_{pp}) = 0$  is due to the decrease of the wetted area toward the ship bow. The value  $\hat{s}^{(h)}$  is set as  $\hat{s}^{(h)} = \phi_h(0)$ . The shape functions obtained from the global loads  $F_i(t)$  are then compared with those extracted directly from the local load distribution  $f(x, t)$  in Figs. 3, 4 and 5. The spline approximation is fine for the first two curves, whereas some differences appear for the third curve toward the transom stern. It can be shown that this fluctuations derive by the fact that Eq. 22 is less satisfied as long as the order of the PC functions increases. On the other hand, these errors have a minor influence on the load reconstruction because their weight on the solution is small (see again Tab. 1). Finally, recalling Eq. 4, the load field  $f(x, t)$  can then be reconstructed.

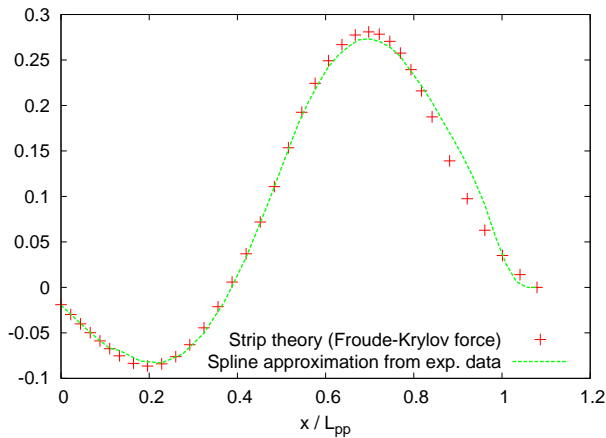


Fig. 3 Comparison between  $\phi_1(x)$  and its spline approximation  $S^{(1)}(x)$ .

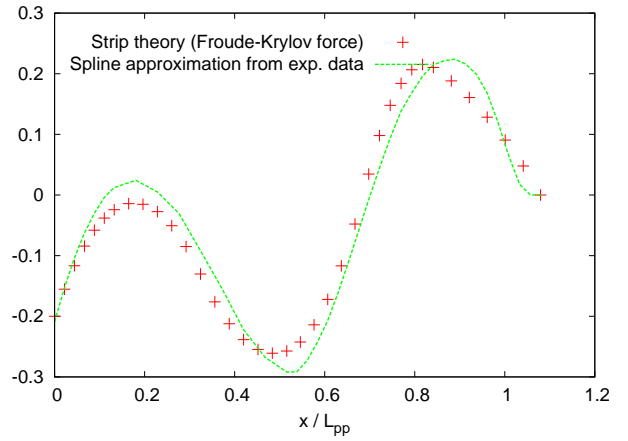


Fig. 4 Comparison between  $\phi_2(x)$  and its spline approximation  $S^{(2)}(x)$ .

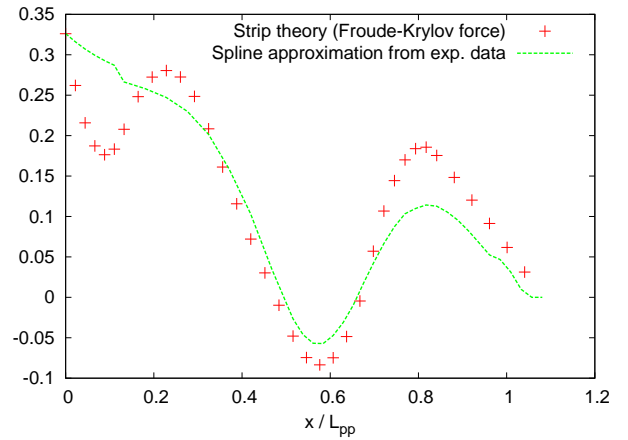


Fig. 5 Comparison between  $\phi_3(x)$  and its spline approximation  $S^{(3)}(x)$ .

## FINAL REMARKS

Preliminary results shows the possibility to exploit global load evaluation to provide the sectional forces on a floating body. Indeed, this technique appears to be more fruitful when applied to experimental data, as those provide by a segmented-hull, for instance. In this paper, the need to clarify the technique and its underlying idea has prevented from providing more results, that will be considered for oral presentation.

## REFERENCES

- [1] Mariani, R. and Dessi, D., 2011. "Analysis of the global bending modes of a floating structure using the proper orthogonal decomposition", *Journal of Fluids and Structures*, in press, doi: 10.1016/j.jfluidstructs.2011.11.009.
- [2] Karhunen, K., 1946. Under lineare methoden in der wahrscheinlichkeitsrechnung. *Annals of Academic Science Fennicae Series, A 1*, Math. Phys., 37(2).
- [3] Loeve, M. M., 1955. Probability theory, Von Nostrand, Princeton, New Jersey.
- [4] Prenter, P. M., 1975. Splines and variational methods, Pure and applied mathematics series, Wiley.
- [5] Dessi, D. and Mariani, R., 2008. "Analysis and prediction of slamming induced loads of an high-speed monohull in regular waves", *Journal of Ship Research*, Vol. 52, No. 1, 2008, pp. 71-86.