Modelling boundary shear stress distribution in open channels using a face recognition technique

Martinez-Vazquez, Pedro; Sharifi, Soroosh

DOI:
10.2166/hydro.2016.068

Document Version
Peer reviewed version

Citation for published version (Harvard):
Martinez-Vazquez, P & Sharifi, S 2017, 'Modelling boundary shear stress distribution in open channels using a face recognition technique', *Journal of Hydroinformatics*, vol. 19, no. 2, pp. 157-172.
https://doi.org/10.2166/hydro.2016.068

Modelling Boundary Shear Stress Distribution in Open Channels Using a Face Recognition Technique

Short Title: Modelling Shear Stress in Open Channels Using a Face Recognition Technique

Pedro Martinez-Vazquez, Ph.D. *Lecturer, Department of Civil Engineering, University of Birmingham, Edgbaston, B15 2TT, UK, +44(0)121 414 5059,*

Soroosh Sharifi *, **Ph.D.** *Lecturer, Department of Civil Engineering, University of Birmingham, Edgbaston, B15 2TT, UK, +44(0)121 414 5100, S.Sharifi@bham.ac.uk*

* Corresponding Author

Abstract

This paper describes a novel application of a pattern recognition technique for predicting boundary shear stress distribution in open channels. In this approach, a synthetic database of images representing normalized shear stress distributions is formed from a training data set using recurrence plot analysis. The face recognition algorithm is then employed to synthesize the recurrence plots and transform the original database into short-dimension vectors containing similarity weights proportional to the principal components of the distribution of images. These vectors capture the intrinsic properties of the boundary shear stress distribution of the cases in the training set, and are sensitive to variations of the corresponding hydraulic parameters. The process of transforming one-dimensional data series into vectors of weights is invertible, and therefore, shear stress distributions for unseen cases can be predicted. The developed method is applied to predict boundary shear stress distributions in smooth trapezoidal and circular channels. The results show a cross correlation coefficient above 92%, mean square errors within 0.04% and 4.48%, and average shear stress fluctuations within 2% and 5%, thus, indicating that the proposed method is capable of providing accurate estimations of the boundary shear stress distribution in open channels.

Keywords

Boundary Shear Stress; Data Modelling; Face Recognition; Open Channel; Recurrence Plot Analysis

Introduction

Boundary shear stress is the result of the tangential component of the hydraulic forces that act in the direction parallel to the channel's boundaries and transfer momentum to its bed and walls (Chow, 1959). Excessive shear stress can undermine channel stability by eroding bank sides and cause changes in the river morphology by affecting the transport and deposition of sediments (Julien, 1995). Erosion often results in higher levels of turbidity and lower water quality levels. Furthermore, an increase in sediment movement and deposition can cause a decrease in channel capacity, and consequently, higher flood risk. Computation of flow resistance, side-wall correction, sediment discharge, channel erosion or deposition, cavitation problems, and design of stable channels are among the problems which require accurate estimates of the boundary shear stress distribution (Yang and Lim, 1997; Guo and Julien, 2005; Blanckaert et al., 2010).

The distribution of boundary shear stress over the wetted perimeter of a channel cross-section is non-uniform. This is true even for steady flows in straight prismatic channels with a simple cross-sectional geometry. This non-uniformity is mainly due to the anisotropy of the turbulence which produces transverse gradients of Reynolds stresses and secondary circulations (Gessner, 1973). Tominaga *et al*. (1989) and Knight and Demetriou (1983) showed that the boundary shear stress increases where the secondary currents flow towards the wall, and decreases when they flow away from the wall. Other factors that govern the distribution of shear stress are the geometry of the cross-section, lateral and longitudinal boundary roughness distributions (Blanckaert et al., 2010) and sediment concentration (Khodashenas et al., 2008).

To date, numerous investigations have been conducted and various mechanistic and empirical methods have been developed for understanding and estimating the magnitude and distribution of boundary shear stress. However, due to the complexities involved, boundary shear stress has proven to be one of the most challenging parameters to quantify and measure, even for simple smooth prismatic channels with uniform flow.

For steady uniform open channel flow, an approximation of the average boundary shear stress can be found by applying Newton's second law on a free body, and balancing the downslope component of the fluid weight by the frictional force exerted by the boundary:

$$\bar{\tau} PL = \gamma AL \sin\alpha \qquad (1)$$

where $\bar{\tau}$ is the average boundary shear stress (Nm^{-2}), A is the channel's cross-section (m^2), P is the channel's wetted perimeter (m), L is the reach length (m), γ is water's specific weight (kgm^{-3}) and α is the slope angle of the channel bed plane. Rearranging Eq. (1) gives:

$$\bar{\tau} = \gamma \sin\alpha \frac{A}{P} = \gamma \sin\alpha R \qquad (2)$$

where R is the hydraulic radius of the channel (m). This simple equation, often referred to as the slope method, is valid for both laminar and turbulent flow regimes, but only provides the average boundary shear stress.

The logarithmic "law of the wall" (Patel, 1965) is another popular and simple method for indirect estimation of the boundary shear stress in rivers and channels. This law, for a two dimensional turbulent flow is given by:

$$\frac{u(z)}{u_*} = \frac{1}{\kappa}\ln\left(\frac{u_* z}{\upsilon}\right) + C \qquad (3)$$

where u is the time-averaged (mean) streamwise velocity profile (ms^{-1}), z is the vertical coordinate (m), u^* is the shear velocity (ms^{-1}) given by $u_* = \left(\tau / \rho\right)^{1/2}$, ν is the kinematic viscosity (m^2s^{-1}), κ is the von Karman's constant =0.41 and C is a dimensionless integration constant related to the thickness of the viscous sub-layer. Although the log law is strictly valid for the turbulent sublayer (approximately the lower 20% of the depth), it is commonly extended over the entire flow depth in rivers and channels (Petrie and Diplas, 2015). If the mean velocity profile, $u(z)$, is known, then a simple linear regression (e.g. least squares) can be applied to fit the velocity profile to Eq. (3) and calculate the log law parameters, the shear velocity, and consequently the shear stress. The advantage of this approach is that it does not need detailed information about bed roughness, however it requires measurements of the streamwise velocity profile, and making assumptions for the viscous sublayer thickness, which to some extent limits its applicability and accuracy.

Preston's (1954) method is the most widely practiced technique for measuring boundary shear stress in smooth channels. In this method, a Preston tube is used to infer the velocity of the water flow by recording the difference between static and total pressures. A non-dimensional calibration function is then established based on the "law of the wall", Eq. (3), and used to determine the boundary shear stress from the differential pressures. The simplicity of the experimental setup and its operation are the main reasons behind the popularity of this method. However, for rough boundaries, application of the technique is substantially more complicated, due to the absence of a viscous sublayer. A number of studies (Hwang and Laursen, 1963; Ghosh and Roy, 1970; Hollick, 1976, Hollingshead and Rajratnam, 1980) have attempted to extend the use of this technique to rough surfaces, and have calibrated curves for the Preston tube by using Nikuradse's (1933) model of velocity distribution over rough boundaries. Although promising, these methods can only be applied when the sand equivalent roughness height of the surface is known, which makes them unsuitable for application to a variety of open channels. Other methods based on fitting the log law of the wall such as Clauser's method (1956) and the boundary characteristics method (Hinze 1975, Papanicolaou et al., 2012) have been developed and applied to gradually (Afzalimehr and Anctil, 2000) and rapidly varying flows over spatially varying boundaries (Papanicolaou et al., 2012).

Geometrical methods for estimating shear stress distribution (Leighly, 1932, Einstein, 1942, Lundgren and Johnson, 1964; Khodashenas and Paquier, 1999; Yang and Lim, 1997; 2005, Yu and Tan, 2007; and Abderrezzak et al., 2008) consist of splitting the channel cross-section into sub-regions where the shear force along each segment of the boundary is calculated by balancing the forces against the weight of fluid in the corresponding sub-region. In these approximations, mapping and discretising the wetted perimeter is often a complicated and sensitive task, however, they have the advantage of requiring relatively low computational effort.

Where abundant experimental data existed, researchers (e.g. Knight, 1981; Knight et al., 1984, 1994; Flintham and Carling, 1988; Pizzuto, 1991; Olivero et al., 1999) have used regression and correlation analysis to derive empirical and semi-empirical equations for boundary shear stress. These equations are capable of only calculating mean, maximum and percentages of shear stress carried on the channel’s walls and beds with relatively good accuracy, but are unable to provide the distribution of shear stress along the entire wetted perimeter. Some other researchers (e.g. Zheng and Jin, 1998, Jin et al., 2004; Guo and Julien, 2005 and Bilgil, 2005) have solved the governing energy transport, continuity, and momentum equations to formulate analytical and semi-analytical solutions for calculating the boundary shear stress. These methods often rely on a number of subjective and controversial assumptions and require a large amount of computing resources which make them impractical. With the advent of more powerful computers, Computational Fluid Dynamic (CFD) techniques have been also used (e.g. Christensen and Fredsoe, 1998; De Cacqueray *et al*., 2009) to solve the referred set of equations and calculate the boundary shear stress distribution. Nonetheless, these methods are computationally expensive and the model outputs are extremely sensitive to mesh size, the turbulence closure model, and other internal parameters which are defined by the user.

Recently, information theory and machine learning techniques have been used to tackle this problem. For instance, the principle of maximum entropy has been used (e.g. Sterling and Knight, 2002; Li and Zhang, 2008; Bonakdari et al., 2015) to establish relationships for the boundary shear stress. A comparison with experimental data has shown that these approximations provide relatively flat shear stress distributions which make them unreliable. The divergence between the numerical and experimental results increases at the regions around the corners of the sections where secondary flow structures are more pronounced. Cobaner et al. (2010) used a neural network with 4 hidden layers to predict the percentage of the shear force acting on the walls of smooth rectangular channels and ducts. The study concluded that the ANN predictions were less biased and slightly more accurate than the classic empirical models suggested by Knight et al. (1984) and Knight and Patel (1985).

Measuring the actual local shear stress along the channel’s boundaries is difficult and costly owing to the complexity of the turbulent velocity field, presence of flow structures, and the small magnitude of the stress. Shear stress also represents a difficult parameter to calculate due to the variability of channel slope, geometry and flow structures, which are the main influencing factors in the complex flow process. To date, all the developed methods are inherently based on some sort of simplifying assumption, and therefore, the problem of accurately estimating these stresses has only been partially resolved (Zheng and Jin, 1998).

The recent relative abundance in available experimental data has offered an opportunity to test and validate novel techniques for describing the boundary shear stress distribution. In this paper, an advanced pattern recognition technique is employed to predict the distribution of boundary shear stress in open channels. This technique, which results from merging two existing algorithms (Recurrence Plots and Eigenfaces for Recognition), is combined with a standard regression model for the prediction of data series representing shear stress distribution of flows with known attributes (i.e. Froude numbers, flow depths and channel slopes).

In the following Sections the Recurrence Plot analysis and its adaptation to the Eigenfaces for Recognition is explained. This is followed by a description of the experimental data used in the study and details of the proposed methodology. Next, the prediction of boundary shear stress distributions in trapezoidal and circular channels are presented and critically discussed. The paper concludes with a discussion on the advantages of the method and suggestions for improvement.

Background

The proposed approach for predicting boundary shear stress distribution combines Recurrence Plot (RP) analysis (Eckmann et al., 1987) and Eigenfaces for Recognition (Turk and Pentland, 1991). RP is used to transform one-dimensional data series into two-dimensional arrays which can be graphically represented. Eigenfaces for Recognition is then used as a means of identifying patterns in the arrays and to transform these into short-dimension vectors which can then be used to predict boundary shear stress distributions. It is to note that despite using a method that was originally developed for the recognition of human faces using 2D still images; no “recognition” is involved in the proposed methodology. Instead, the technique is used to filter the original data and reduce its dimensionality whilst preserving intrinsic qualities. These reduced databases are then used to produce shear stress distribution for unseen cases.

Recurrence Plots

Recurrence Plots are visualization tools that can be used to picture the recurrence behaviors, hidden patterns and nonlinearities in data sets (Marwan et al., 2007). In this technique, starting from the first point of a data series, d-dimensional vectors are formed by taking a sample of $\boldsymbol{d}$ consecutive points in the data series:

$$\begin{aligned} \vec{\boldsymbol{q}}_{\boldsymbol{j}} &= \{q_j,\ q_{j+1}.\ .,\ q_{j+d-1}\} \\ &\vdots \\ \vec{\boldsymbol{q}}_{\boldsymbol{k}} &= \{q_k,\ q_{k+1}.\ .,\ q_{k+d-1}\} \end{aligned} \tag{4}$$

where subscripts j and k represent the j^{th} and k^{th} data points in the data series. The d-dimensional vectors are then correlated by calculating the Euclidean distance between them. This parameter can then be used to form the RP matrix:

$$RP = \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1N} \\ e_{21} & e_{jk} & & \vdots \\ \vdots & & \ddots & \vdots \\ e_{N1} & \cdots & \cdots & e_{NN} \end{bmatrix} \tag{5}$$

where e_{jk} is the distance between vectors $\vec{q}_j$ and $\vec{q}_k$ and N is the total number of vectors, which define the number of data points in each column and row of the matrix. Note that the values of e_{jk} in the RP matrix vary with d whilst any value of d would result in a matrix that could be used in the recognition method as described further below. However by letting $d = 1$ each row in the RP matrix effectively becomes a normalized version of the original one-dimensional data series which has the benefit of maintaining its basic structure throughout the pattern recognition process.

By projecting the RP matrix on a Cartesian space, a map of the data can be generated. In this case each pixel on the map has the coordinates $\{j,k\}$ with j, k = 1, 2, …, N, as well as a numerical value that is proportional to its associated distance, e_{jk}. In an 8-bit grayscale image representation, the values of e_{jk} can take values within the range 0 to 255, where the brightness intensity of each pixel indicates a larger e_{jk}. Through the calculation of the *RP* matrix, the correlation between all data points within the data series is established whilst preserving the basic structure of the database. Such transformation and visualization helps to make explicit features of data which otherwise would be difficult to observe in the original series.

The graphical representations of *RP*s helps to visualize characteristic patterns of the data, although for numerical analysis, its elements need to be sorted in a N^2 dimension vector where all rows of the *RP* matrix are assembled in sequence:

$$\Gamma^T = \{e_{11},\ e_{12}, \ldots,\ e_{NN}\} \tag{6}$$

The arrangement of columns of the RP matrix into the Γ vector is schematically shown in Figure 1. The unique configuration of RPs makes them particularly suitable for machine learning. That is because pattern recognition methods are capable of identifying data sets with unique features such as those made explicit through the RP method. In the following sections the post-processing of Γ vectors and their relationship with the hydraulic parameters that control shear stress distributions is discussed in detail.

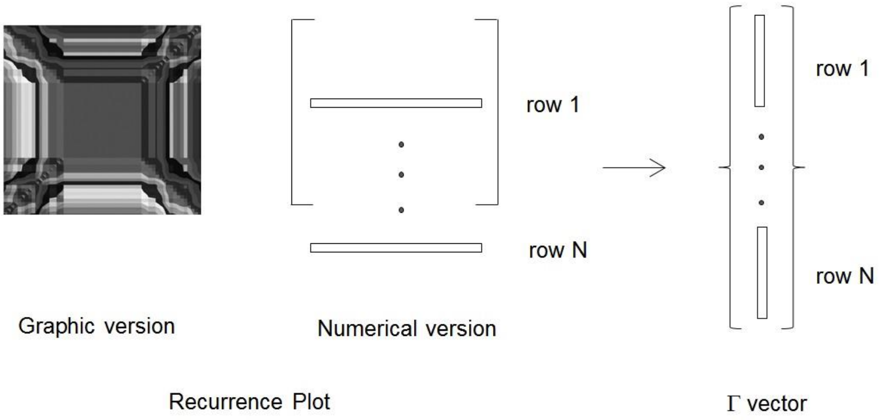


Figure 1. Relationship between the RP matrix and the Γ vector.

Eigenfaces for Recognition

The Eigenfaces for Recognition is based on the premise that any 2D image of resolution *N* x *N* can be represented by an N^2 size vector Γ, where each element is a real number that represents an individual pixel in the image. If the training set consists of *M* images, then the average face of the training set, ϑ , is defined by:

$$\vartheta = \frac{1}{M}\sum_{i=1}^{i=M}\Gamma_i \qquad (7)$$

and hence, the difference between each image, Γ_i, and the average face, ϑ, is given by:

$$\phi_i = \Gamma_i - \vartheta; \quad i = 1, 2, ..., M \qquad (8)$$

Performing principal component analysis (PCA) on the collection of all ϕ_i, would result in a set of *M* orthonormal vectors which best describe the distribution of data. PCA is a statistical procedure that is able to identify orthogonal modes or degrees of freedom within a numerical array, and transform a number of possibly correlated variables into a smaller number of uncorrelated variables, which are called the principal components. The eigenvectors and eigenvalues of these principal components can be determined from the covariance matrix:

$$C = AA^T \qquad (9)$$

where

$$\mathrm{A} = [\phi_1, \phi_2 \ldots, \phi_M] \qquad (10)$$

Matrix *C* is of size N^2, and finding its eigenvectors and eigenvalues is computationally expensive. If the number of training images, *M*, is less than the dimension of the space, N^2, then there will only be *M*-1 meaningful eigenvectors (Turk and Petland, 1991), and hence, to reduce the calculations, an *M* by *M* matrix *L* can be constructed to find the meaningful eigenvectors:

$$L = A^T A \qquad (11)$$

where

$$L_{ij} = \phi_i^T \phi_j \qquad (12)$$

The principal components of the distribution of images are called the eigenfaces, u_l, which can be calculated from a linear combination of the images and eigenvectors:

$$u_l = \sum_{i=1}^{M} v_{li} \phi_i \quad l = 1, \ldots, M \qquad (13)$$

where v_{li} is the i^{th} eigenvector of the covariance matrix. The collection of eigenface vectors defines a subspace of training images which is called the "*face space*". Any input image expressed in vector form Γ, can be projected into the *face space* through the following operation:

$$\omega_i = u_i^T (\Gamma - \vartheta), \quad i = 1, \ldots, M' \qquad (14)$$

where ω_i is a weight factor that describes the contribution of the i^{th} eigenface in representing the image, and M' is the number of significant eigenvectors, associated with the M largest eigenvalues i.e. $M' \leq M$. Furthermore, the set of weights ordered in a short-dimension vector $\Omega^T = \{\omega_1, \omega_2 \ldots, \omega_{M'}\}$ can be used to project any new image, Γ', into the *face space* by:

$$\Omega = U^T (\Gamma' - \vartheta), \quad i = 1, \ldots, M' \qquad (15)$$

where $U=\{u_i\}$ is the collection of eigenfaces. Eq. (15) suggests that the process of encoding data into Ω vectors can be inverted for prediction purposes. If a reliable estimation of weights factors is available to conform a new vector Ω', then a prediction of its associated image can be made through:

$$\Gamma' = U\Omega' + \vartheta \qquad (16)$$

As will be further explained in the following sections, a reliable estimation of the weight factors, i.e. vector Ω', would be based on the known vectors obtained through Eq. (15) together with the parameters that characterize the original data sets. These vectors and parameters can be typically related with the aid of a simple regression model or more complicated method such as an artificial neural network, and the output Ω' can be considered to be reliable if the modeling error is less than 5%. Note that the validity of Eq. (16) is provided in Appendix A.

Experimental datasets

In this study, laboratory measurements of flow velocity and boundary shear stress in trapezoidal and circular open channels were taken directly from the University of Birmingham's Flow Database (www.flowdata.bham.ac.uk).

Trapezoidal datasets

Two sets of experimental data relating to uniform flow in trapezoidal channels were used in this study: Yuen (1989) and Yuen and Knight (1990). The data included local boundary shear stress measurements in trapezoidal channels (Figure 2) using a Preston tube, made in fully developed flow under uniform flow conditions in a 22 m long titling flume. Two different base widths (2b) of 0.15m and 0.45m were considered, and the bed slope was varied from $1x10^{-3}$ to $2.337x10^{-2}$ in order to observe shear stress distribution for Froude (Fr) and Reynolds (Re) numbers within ranges of 0.58 $\leq$ Fr $\leq$ 3.59 and $0.46x10^5 \leq$ Re $\leq 6.18\ x10^5$, respectively, which derives from flow velocities (*V*) between 0.39 ms^{-1} and 2.69 ms^{-1}. Measurements of velocity and shear stress were taken on average every 20 mm along the wetted perimeter (i.e. between 16 and 32 measurement points for each case), and measurement accuracy was estimated to be within +/-5% (Yuen, 1989).

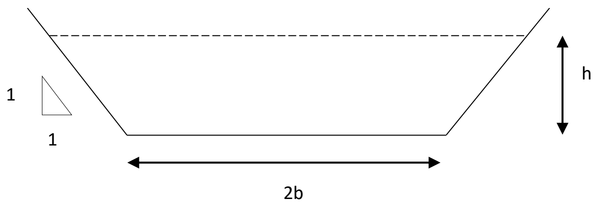


Figure 2. Trapezoidal channel cross section.

To obtain homogenous subsets suitable for pattern recognition, and to test the sensitivity of the approach to the size of the training set, k-nearest neighbors (k-nn) clustering analysis (Fix and Hodges, 1951) was first performed. The fundamental idea of the k-nn algorithm is to simply separate the data based on the assumed similarities between various clusters. Here, the Euclidean distance metric was used to measure the similarity between clusters, and shear stress data was non-dimensionalized by the average shear stress to eliminate the scale effects. K-nn was run with different *k* values, and consequently, three clusters (subsets) were identified by investigating the resultant dendrograms, i.e. graphical tree-structures that show the hierarchical relationships among clusters, ensuring highest similarity within each cluster (homogeneity) and lowest similarity between clusters were achieved.

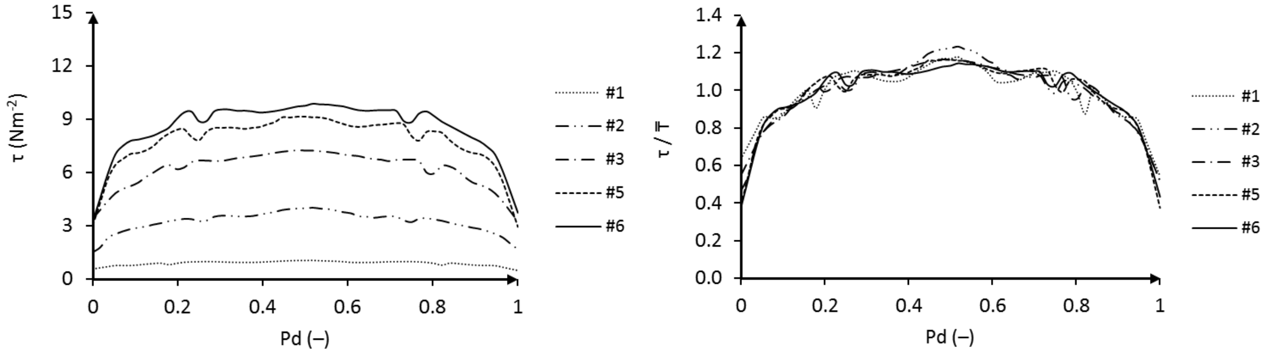
Table (1) lists the geometric and hydraulic parameters of all the experiments. In each subset, one experiment (highlighted in Table 1) was randomly selected and excluded to be used for validation whilst the remaining were considered for training. Since the method requires all data series in the set to have the same number of measurements taken at relatively even distances, for each experiment, the horizontal coordinates of all data points were normalized using a perimetric distance defined as *Pd* = *s*/*p*, where *s* is the distance along the wetted perimeter starting at the left bank at the free surface, moving around the wetted perimeter, and *p* is the total length of the wetted perimeter. The shear stress measurements in each experiment were also non-dimensionalized by the average shear stress. Where required, linear interpolation was used to obtain shear stress values from adjacent neighboring points. It is to note that at regions where shear stress varied at higher rates, experimental measurements were taken at smaller increments, thus resulting in the standardization of the degree of accuracy across the wetted perimeter.

Figure (3) shows the distribution of the measured and non-dimensionalized shear stress for each of the three sets. It can be seen that the data series corresponding to each set share patterns such as the location of peak values and inflection points, which are attributed to the secondary flow structures, and the range of shear stress fluctuations across the wetted perimeter.

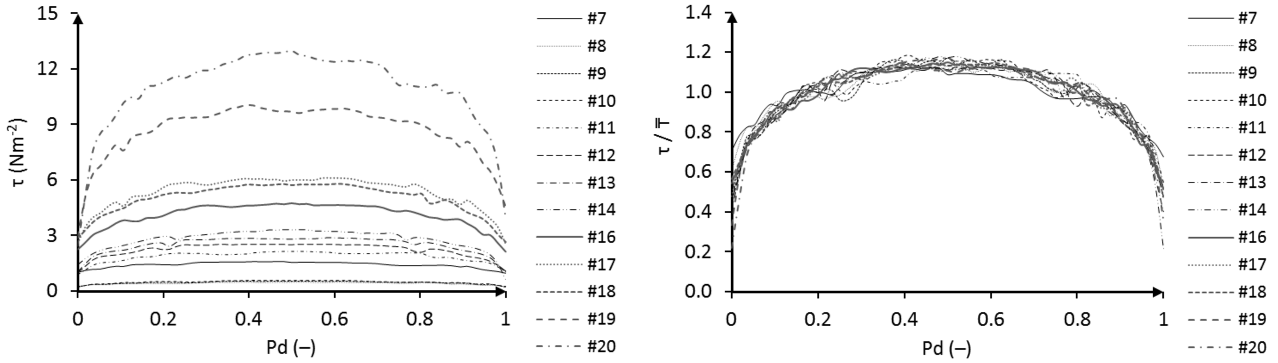
Table 1. Geometric and hydraulic parameters of trapezoidal experiments.

		1	2	3	4	5	6	7	8	9	10	11
	ID	**2*b* (m)**	***h* (m)**	***A* (m²)**	***S₀***	***P* (m)**	***R***	***V* (ms⁻¹)**	***Q* (m³s⁻¹)**	**$\bar{\tau}$ (Nm⁻²)**	**Fr**	**Re**
Set 1	#1	0.15	0.030	0.005	0.0040	0.235	0.023	0.565	0.003	0.894	1.12	11392
	#2	0.15	0.058	0.012	0.0087	0.313	0.038	1.308	0.016	3.256	1.97	43656
	#3	0.15	0.037	0.007	0.0234	0.255	0.027	1.843	0.013	6.223	3.35	43800
(TVc-1)	**#4**	**0.15**	**0.042**	**0.008**	**0.0234**	**0.269**	**0.030**	**1.901**	**0.015**	**6.871**	**3.27**	**50251**
	#5	0.15	0.050	0.010	0.0234	0.291	0.034	2.080	0.021	7.859	3.32	62609
	#6	0.15	0.057	0.012	0.0234	0.310	0.038	2.190	0.026	8.625	3.32	72546
Set 2	#7	0.45	0.044	0.022	0.0040	0.574	0.038	0.893	0.019	1.472	1.42	29574
	#8	0.45	0.050	0.025	0.0010	0.591	0.042	0.398	0.010	0.414	0.60	14743
	#9	0.45	0.056	0.029	0.0010	0.609	0.047	0.428	0.012	0.459	0.61	17563
	#10	0.45	0.060	0.031	0.0010	0.620	0.049	0.439	0.013	0.484	0.60	18998
	#11	0.15	0.029	0.005	0.0087	0.231	0.022	0.924	0.005	1.882	1.88	17923
	#12	0.15	0.036	0.007	0.0087	0.250	0.026	1.010	0.007	2.244	1.87	23347
	#13	0.15	0.041	0.008	0.0088	0.266	0.029	1.113	0.009	2.521	1.94	28663
	#14	0.15	0.048	0.009	0.0087	0.284	0.033	1.231	0.012	2.815	2.01	35702
(TVc-2)	**#15**	**0.45**	**0.044**	**0.022**	**0.0087**	**0.574**	**0.038**	**1.381**	**0.030**	**3.228**	**2.19**	**45751**
	#16	0.45	0.059	0.030	0.0087	0.615	0.048	1.553	0.046	4.124	2.17	65743
	#17	0.45	0.044	0.022	0.0145	0.574	0.038	1.822	0.040	5.384	2.89	60371
	#18	0.15	0.029	0.005	0.0234	0.231	0.022	1.592	0.008	5.052	3.24	30888
	#19	0.45	0.045	0.022	0.0234	0.576	0.038	2.272	0.050	8.752	3.59	76145
	#20	0.45	0.059	0.030	0.0234	0.615	0.048	2.427	0.072	11.067	3.38	102739
Set 3	#21	0.15	0.075	0.017	0.0040	0.362	0.047	0.960	0.016	1.812	1.29	39299
	#22	0.15	0.107	0.028	0.0010	0.453	0.061	0.497	0.014	0.596	0.58	26583
(TVc-3)	**#23**	**0.15**	**0.125**	**0.034**	**0.0010**	**0.504**	**0.068**	**0.544**	**0.019**	**0.669**	**0.59**	**32599**
	#24	0.15	0.150	0.045	0.0010	0.574	0.078	0.584	0.026	0.768	0.59	40170
	#25	0.45	0.075	0.039	0.0010	0.662	0.060	0.514	0.020	0.583	0.64	26845
	#26	0.15	0.073	0.016	0.0087	0.356	0.046	1.468	0.024	3.896	2.00	58886
	#27	0.15	0.099	0.025	0.0087	0.430	0.057	1.667	0.041	4.891	2.00	84008
	#28	0.15	0.030	0.005	0.0145	0.235	0.023	1.296	0.007	3.272	2.58	26146
	#29	0.15	0.075	0.017	0.0145	0.361	0.046	1.943	0.033	6.598	2.62	78915
	#30	0.15	0.099	0.025	0.0234	0.430	0.057	2.690	0.066	13.129	3.23	135513
	#31	0.45	0.044	0.004	0.0234	0.574	0.007	1.679	0.007	5.264	3.13	9997

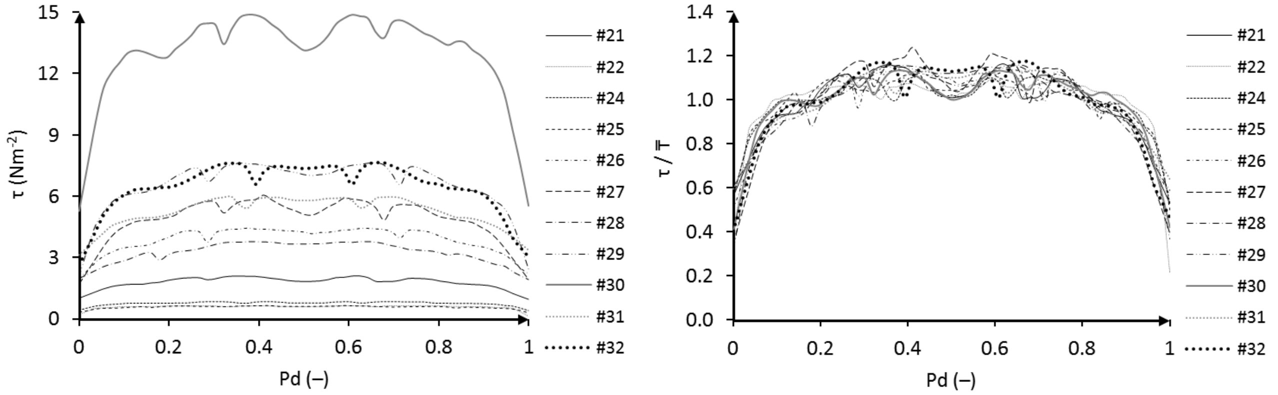
#32	0.45	0.058	0.006	0.0234	0.614	0.010	1.785	0.011	6.513	2.96	15045



a) Set 1



b) Set 2



c) Set 3

Figure 3. Measured and non-dimensionalized shear stress distribution of trapezoidal channels.

Circular Channels

A separate set of experimental data containing local boundary shear stress measurements in a circular channel, with and without a flat bed, running partially full was also used in this study (Figure 4). This data has been described and analyzed in detail by Sterling (1998), Knight and Sterling (2000) and Sterling and Knight (2000).

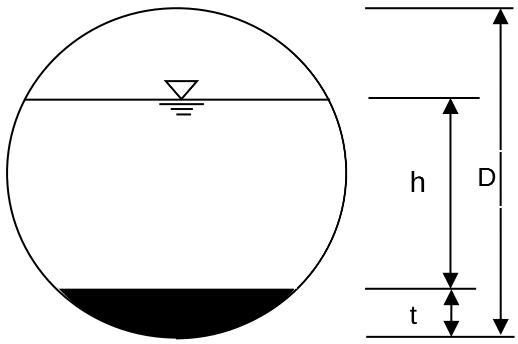


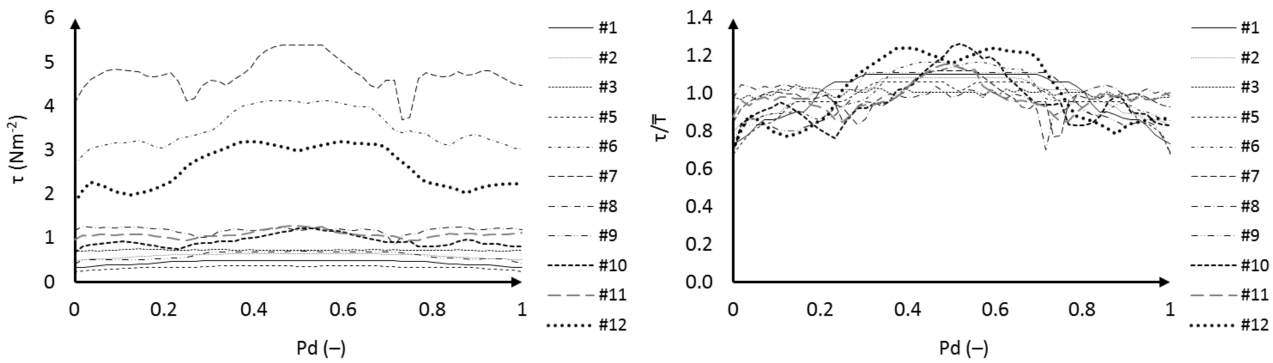
Figure 4. Circular channel cross section.

Similar to the trapezoidal data, k-nn clustering analysis was performed and only one major cluster was identified. It is to note that for this particular data set, even if the clustering was performed in a different way (e.g. with a different distance measure or clustering technique) and more than one cluster was obtained, the quality of the prediction would increase, given enough data is available for training the model. The only major difference would be the increased computation required for the extra clusters. Hence, in this context, one can see clustering as the process to find the optimum data sets on which modelling can be applied to without loosing accuracy. Table (2) lists the geometric and hydraulic parameters of all the circular channel test cases. Two validation cases, labelled CVc-1 and CVc-2, were randomly chosen to validate the method and were excluded from the training set. The local shear stresses were originally measured at 10mm intervals around the wetted perimeter using a Preston tube. Hence, the difference in water depth between experiments resulted in different number of measurements in each data series, ranging from 30 to 60 point measurements. Similar to what was done for the trapezoidal case studies, a perimetric distance, *Pd*, was used to uniformize the number of measurements along the wetted perimeter. Where data points did not exist in the original series, linear interpolation was used to infer local boundary shear stress from adjacent neighboring points. Figure (5) depicts the boundary shear stress distribution of the cases in the training set.

Table 2. Case studies for circular section.

		1	2	3	4	5	6	7	8	9	10	11	12
	ID	***D***	***t/D***	***h/D***	**A**	***S_0***	***P***	***R***	**V**	**Q**	$\bar{\tau}$	**Fr**	**Re**
		m			**(m^2)**	**x 10^{-2}**	**(m)**		**(ms^{-1})**	**(m^3s^{-1})**	**(Nm^{-2})**		
	#1	0.24	0	0.33	0.014	0.1	0.300	0.045	0.394	0.005	0.441	0.52	15687
	#2	0.24	0	0.51	0.024	0.1	0.386	0.061	0.493	0.012	0.597	0.51	26580
	#3	0.24	0	0.83	0.041	0.1	0.557	0.074	0.554	0.023	0.721	0.38	36068
CVc-1	**#4**	0.24	**0.25**	**0.15**	0.008	0.196	**0.078**	**0.106**	**0.403**	**0.003**	**0.545**	**0.70**	37377
	#5	0.24	0.25	0.08	0.004	0.196	0.044	0.100	0.294	0.001	0.337	0.67	25865
	#6	0.24	0.25	0.25	0.014	0.862	0.127	0.111	1.283	0.018	3.538	1.70	125381
	#7	0.24	0.25	0.42	0.024	0.862	0.210	0.114	1.625	0.039	4.804	1.59	162219
	#8	0.24	0.25	0.55	0.031	0.196	0.282	0.109	0.775	0.024	1.198	0.63	74130
	#9	0.24	0.33	0.17	0.010	0.2	0.083	0.117	0.449	0.004	0.612	0.72	46203

	#10	0.24	0.33	0.33	0.020	0.2	0.166	0.117	0.625	0.012	0.967	0.69	64359
	#11	0.24	0.33	0.47	0.027	0.2	0.241	0.110	0.833	0.022	1.106	0.72	80571
	#12	0.24	0.5	0.16	0.009	0.9	0.081	0.117	0.886	0.008	2.571	1.40	91194
CVc-2	**#13**	**0.24**	**0.5**	**0.25**	**0.014**	**0.88**	**0.126**	**0.111**	**1.143**	**0.016**	**3.341**	**1.42**	**111576**



a) measured shear stress b) non-dimensionalized shear stress

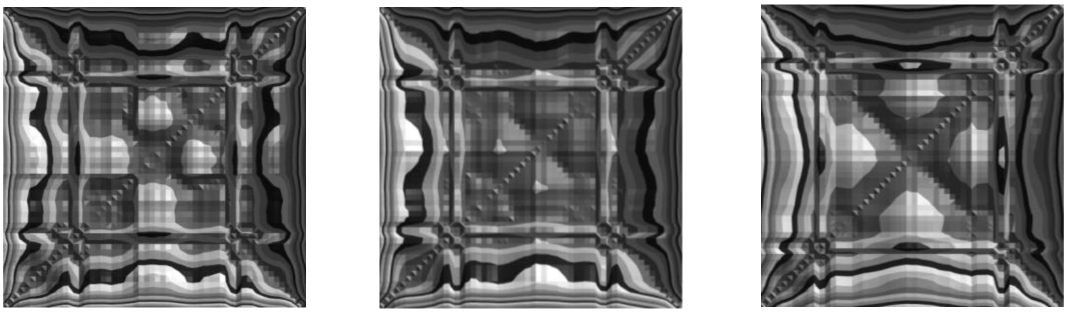
Figure 5. Measured and non-dimensionalized shear stress distribution of circular channels.

Methodology

The initial step in using Recurrence Plots and Eigenface Recognition for predicting boundary shear stress distribution is forming a training set from available experimental data. As mentioned in the previous section, the raw experimental data sets typically consist of a number of local boundary sheer stress measurements, taken along the wetted perimeter of a channel, and presented as a data series. If measurements are not taken at the same relative locations across the different channels, then, to make the raw data suitable for use in the data mining algorithm, interpolation is carried out to find shear stresses at the same relative distances along the wetted perimeter, for all data series.

As the Eigenfaces for Recognition algorithm accepts two-dimensional arrays of numbers, the original (one-dimensional) data series in the training set has to be pre-processed. This can be done through the Recurrence Plot algorithm. To this end, the dimension of the $\vec{q}$ vectors identified in Eq. (4) is set to be equal to 1, so that each vector would contain numerical differences of shear stress between consecutive points in the data series, effectively resulting in each row of the RP matrix to become a normalized version of the original data series. In line with the image recognition algorithm, once the RP matrix is formed, Eq. (6) is used to construct a unique Γ_i vector representing the *i*-th experimental data set.

Figure (6) shows the recurrence plots for the shear stress distribution of selected trapezoidal and circular data sets. It can be seen that the patterns of RPs associated to trapezoidal sections are fairly consistent. There is a square region at the center of most images whose silhouette stretches along the diagonals more than it does towards the sides. This does not appear to be the case for circular cases (Figure 6b) where the RPs follow at least three types of patterns. The consistency of the datasets is also present in the original series shown in Figures 3 and 4 although the use of the RP technique has made those intrinsic properties more explicit. That is the main reason for pre-processing the data prior to applying the full recognition method. As will be shown in the Results Section, the apparent constraint found in data from circular channels did not have a significant impact on the prediction of shear stress distributions.



ID #6 ID #5 ID #3

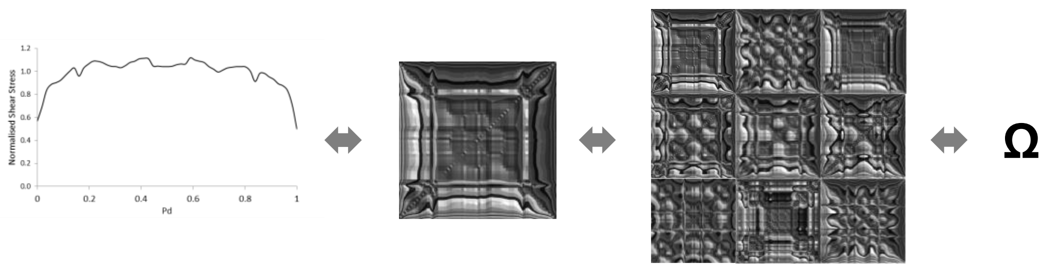
(a) Trapezoidal cross-section Set 1

ID #1 ID #3 ID #6

(b) Circular cross-section

Figure 6. Recurrence Plots of selected a) Trapezoidal and b) Circular channels.

Once the Recurrence Plot vector representation for all members of the training set are obtained, Eq. (7) is used to calculate the average face, ϑ, and consequently, the difference between each image in the training set and the average face, ϕ_i, are found by using Eq. (8). Then, performing Principal Component Analysis, the eigenvectors that characterize the face space are computed, and consequently, the set of weights, Ω, are determined by using Eq. (14). Figure (6) outlines the steps involved for encoding the training data sets and obtaining the vectors of weights, Ω.



Original data series RP eigenspace Encoded Data

Figure 7. Process of encoding data series of shear stress.

As mentioned in the background Section, to use this approach for predicting the boundary shear stress distribution of an unseen case, i.e. experimental data not included in the training set, also hereafter referred to as a validation case, a set of weights must be obtained to be used in Eq. (16). For the sake of generality, the weights associated to any experimental case are related to non-dimensional parameters which represent the major characteristics of channel's geometry and flow. In this research we have used the following non-dimensional attributes:

- Trapezoidal channels: $2b/h$, $2bh/A$, Fr, Re
- Circular channels: $(h+t)/D$, $Q/(VD^2)$, Fr, Re

Where h is the water depth, V is the mean velocity, Q represents discharge, and Fr and Re are the Froude and Reynolds numbers, respectively. The bottom width for trapezoidal channels is represented by 2b whilst t/D is the base height to diameter ratio for circular channels. In order to relate weights and non-dimensional attributes, a simple regression model can be established:

$$\omega_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_n x_{in} \qquad (17)$$

where ω_i represents a regression estimation of the i-th weighting factor, x_{ij} is the j-th non-dimensional hydraulic/geometric parameter of the *i-th* training experiment and β_j are regression parameters. The solution to Eq. (17) is given by:

$$\beta = \left(\boldsymbol{X}^T \boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \Omega \qquad (18)$$

where $\hat{\beta}$ is the best estimator vector of the target β factors $\boldsymbol{X}$ represents the matrix of hydraulic/geometric non-dimensional parameters, and Ω is the vector of target weighting factors. Once the set of β factors is determined, the prediction model can be established. After investigating a number of non-dimensional attributes, the ones listed above were found to strongly influence the established relationship with the target weighting factors.

Figure (6) shows the steps involved in the process of encoding the original data series. This process can be reversed to obtain a new set of weighting values, ω_i, e.g. for test cases not included in the training set. Eq. (17) enables to find those weighting factors which can then be stored in the short-dimension vector *(Ω')*. Following, The vector *Γ'* can be predicted by applying Eq.(16).

To help the reader better understand the entire modeling process, a simple step-by-step guide to using the proposed approach is presented in Appendix B. Furthermore, a copy of the code written in C++ is available at: http://shear-stress-using-face-recog.sourceforge.net

Results

Trapezoidal Channels

The proposed methodology was first applied to the three trapezoidal data sets presented in Table (1). In summary, for each training set, the process shown in Figure (6) was first followed to obtain the weight factors for the training set, Ω. Then, Eq. (17) was used to establish a linear regression between the weights associated with each eigenface and the non-dimensional attributes of the geometric and hydraulic parameters of the experiments. The corresponding β regression parameters were then obtained using Eq. (18) and the vector of estimated weights for the unseen test case, Ω' was constructed. All weighting and β factors are provided in Appendix C. Subsequently, Eq. (16) was used to obtain the Γ' vector, and its elements were transformed to find non-dimensionalised shear stress values across the channel. It should be noted that the first component of the Γ' vector, i.e. e_{11}, which corresponds to the boundary shear stress value at $Pd = 0$ is always zero. The boundary shear stress at this point was obtained through a separate linear regression between the predicted values of local shear stresses:

$$\tau_{0i} = \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_n x_{in} \qquad (19)$$

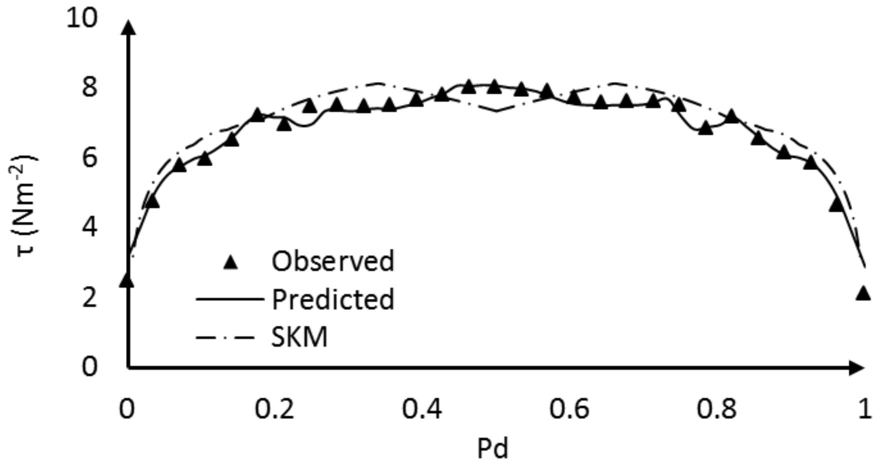
where τ_{0i} represents a regression estimation of the i-th shear stress at $Pd = 0$, x_{ij} is the j-th non-dimensional hydraulic/geometric parameter of the *i-th* training experiment and β_j are regression parameters.

Finally, the predicted series were rescaled by multiplying their ordinates by the estimated average shear stress ($\bar{\tau}$) obtained by the Slope method (Eq. 1). For practical purposes, this parameter could be estimated using any other prediction model, such as the ones suggested by Knight (1981), Knight et al., (1984&1994) and Flintham and Carling, (1988).

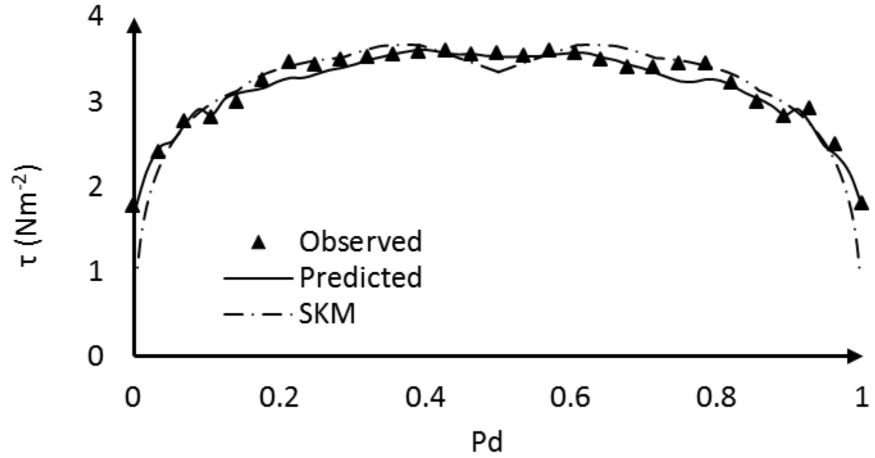
Figure (7) shows the predicted vs. observed boundary shear stress distributions for the validation cases in each of the two sets along with the predictions of the well-established Shiono and Knight model (SKM) (Shiono & Knight, 1988; 1990). As it can be seen, although there is some difference between the observed and predicted values particularly at the edges of the wetted perimeter, the shape and amplitude of the predicted curves accurately follow the observed distributions. Moreover, the proposed method outperforms SKM in all three cases, particularly at the edges. It is also inferred that the relatively small number of experiments in training Set 1 did not have a significant impact on prediction accuracy. The mean square error (MSE) between observed and predicted ordinates averaged over all data points along the wetted perimeter was found to be of 4.4%, 0.88%, and 0.04% for TVc-1, 2, and 3, respectively. Table (3) shows a more detailed comparison between observed and simulated data series for the validation cases. In this table $\left|\Delta\bar{\tau}\right|$ is the average absolute difference between the ordinates of the observed and predicted data series, $\left|\Delta\tau_{\max}\right|$ is the largest estimated difference and MSE_{SKM} is the mean square error for the SKM model. The table also provides the ratio of those divergence parameters with respect to the average stress, in addition to the relative location of $\Delta\tau_{peak}$ along the wetted perimeter covering an interval [0-1], with 0 and 1 corresponding to the utmost left and right edges of the wetted perimeter, respectively. Furthermore, the cross correlation, ρ_τ , between the observed and predicted time series is presented for each case:

$$\rho_\tau = \frac{1}{N}\left[\sum_{i=1}^{N}(\tau_{\text{obs}} - \overline{\tau}_{\text{obs}})(\tau_{\text{model}} - \overline{\tau}_{\text{model}}) / \sigma_{\text{obs}}\sigma_{\text{model}}\right] \quad (21)$$

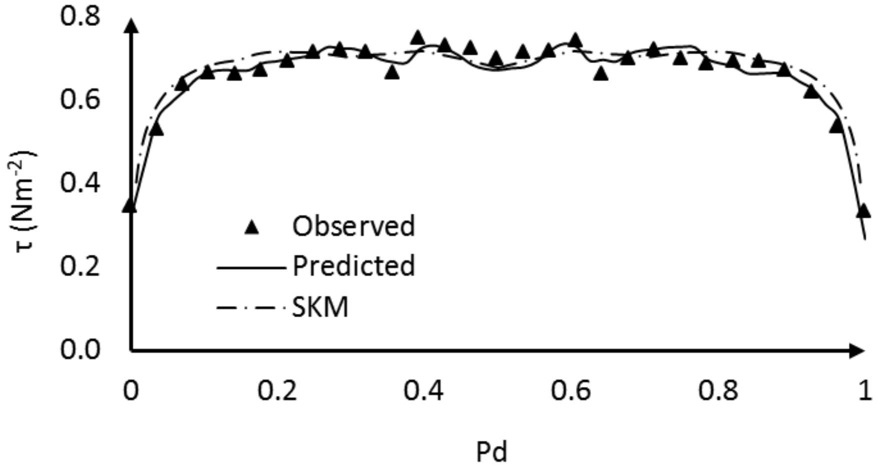
where $\bar{\tau}$ is the average shear stress and σ is the corresponding standard deviation.



a) TVc-1 (Fr=3.27)



b) TVc-2 (Fr=2.19)



c) TVc-3 (Fr=0.59)

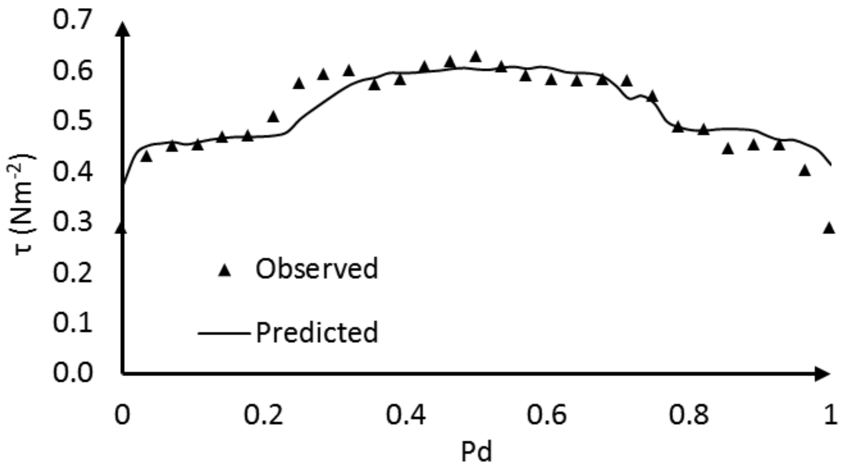
Figure 8. Modelled vs measured shear stress distributions for trapezoidal validation cases.

Table 3. Overview of predicted boundary shear stress for trapezoidal channels.

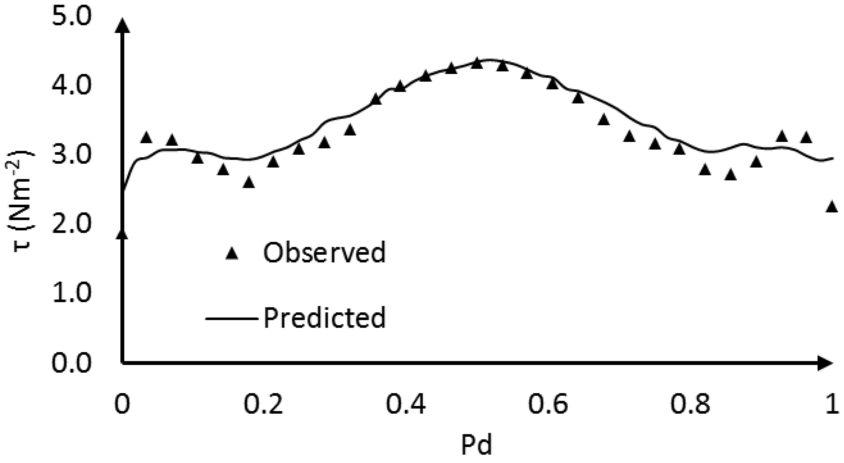
Validation case	Fr	*MSE*	$\lvert\Delta\bar{\tau}\rvert$ **(Nm^{-2})**	$\frac{\lvert\Delta\bar{\tau}\rvert}{\bar{\tau}}$	$\lvert\Delta\tau_{\max}\rvert$ **(Nm^{-2})**	$\frac{\lvert\Delta\tau_{max}\rvert}{\bar{\tau}}$	Relative location of $\lvert\Delta\tau_{\max}\rvert$	ρ_τ	MSE_{SKM}
TVc-1	3.2701	0.0448	0.1404	0.0204	0.7682	0.1118	1	0.994	0.6589
TVc-2	2.1934	0.0088	0.0694	0.0215	0.2340	0.0725	0.946	0.984	0.1107
TVc-3	0.5926	0.0004	0.0167	0.0249	0.0641	0.0958	1	0.975	0.0008

Circular Channels

The methodology was also applied to the circular channel data set introduced in the Experimental Datasets Section. Figure (8) shows the modelled vs. measured shear stress distribution for the validation test cases. The MSE between observed and predicted values were found to be 0.12% in the first case, CVc-1, and 4.9 % in the second validation case, CVc-2, which are of similar order than those found for the trapezoidal cases. The largest divergence was obtained at the channel edges (Pd =0, 1) which seems to be a reflection of the scatter of the input data shown in Figure (4). It is also noted from Figure (8) that the predicted curves tend to be smoother than the observed ones. This can be due to a *group* effect which causes the predicted shear distributions to tend to the average face value established in the face space. This effect is to some extent implicit in Eq. (8). Nonetheless, the cross correlation parameter, which is well above 0.9 in both cases, together with the magnitude of the differences indicate that the shear stress contours have been captured with excellent accuracy. Table (4) provides an extended comparison between observed and predicted data for circular validation test cases.



a) CVc-1 (Fr=0.7)



b) CVc-2 (Fr=1.42)

Figure 9. Modelled vs measured shear stress distributions for circular validation cases.

Table 4. Overview of predicted boundary shear stress for circular channels.

Validation case	Fr	*MSE*	$\left\|\Delta\bar{\tau}\right\|$ **(Nm^{-2})**	$\frac{\|\Delta\bar{\tau}\|}{\bar{\tau}}$	$\left\|\Delta\tau_{\max}\right\|$ **(Nm^{-2})**	$\frac{\|\Delta\tau_{max}\|}{\bar{\tau}}$	Relative location of $\left\|\Delta\tau_{\max}\right\|$	ρ_τ
CVc-1	0.696	0.0012	0.0245	0.0448	0.1239	0.2264	1	0.922
CVc-2	1.420	0.0494	0.1710	0.0512	0.7048	0.2110	1	0.944

Summary and Conclusions

Recurrence Plot analysis and Eigenface for Recognition were used to predict the distribution of boundary shear stress in trapezoidal and circular channels. In this approach, first, the RPs of all training set members are constructed and the differences between them and the average RPs are computed. Principal component analysis is then performed and weight factors proportional to the eigenvectors are obtained. To obtain predictions of boundary shear stress, a simple regression equation is established to relate the weight factors to non-dimensional attributes of the training set's hydraulic and geometric characteristics. For each validation case, corresponding weights are obtained by the regression equations, and the reverse of the process is performed to obtain the distribution of the boundary shear stress.

The method was applied to two trapezoidal data sets and one circular data set. The results showed that:

- The technique is capable of capturing the intrinsic patterns of the RPs which makes it suitable for the prediction of shear stress distributions.
- The method is valid for both sub and supercritical flow conditions.
- The average error obtained across all predicted series is of 2.09% and the cross correlation is within 92% of accuracy for all trapezoidal and circular verification cases.

The accuracy of the predictions was found to be somewhat higher for trapezoidal channels compared to circular channels. This can be a reflection of the consistency of the input information which in the case of circular channels is less, i.e. the distributions are less uniform. The variation of the shape of the wetted section with the increase of the water level appears to be the reason of such variability.

The present investigation was based on a database formed by a limited number of experimental test cases, particularly for the case of circular sections. Nevertheless, the prediction results were satisfactory. The robustness of the methodology should be further tested with a larger training database containing further combinations of hydraulic parameters and section dimensions, and additional validation cases. This would help to ensure the generality of the weighting factors, and therefore, the overall accuracy of the prediction models. Based on the analysis presented here it is clear that the method works for relatively low number of input data series which in this research ranged between 5 and 13 data series in the clusters. Furthermore, the linear regression models were demonstrated to be adequate estimators for the relatively smooth bed shear stresses studied, as they were able to capture the rates of shear stress variation with accuracy. It is to note that such simple estimators might not be accurate when modelling more complex configurations and patterns, and a more robust estimator (e.g. artificial neural network) may be more suitable depending on the degree of non-linearity observed in the input data.

References

Abderrezzak, K. E. K., Camenen, B., & Paquier, A. (2008). Estimation of the boundary shear stress distribution in open channel using flownet. *Journal of Hydraulic Research*, *46*(5), 716-720.

Bilgil, A. (2005). Correlation and distribution of shear stress for turbulent flow in a smooth rectangular open channel. *Journal of Hydraulic Research*, *43*(2), 165-173.

Blanckaert, K., Duarte, A., & Schleiss, A. J. (2010). Influence of shallowness, bank inclination and bank roughness on the variability of flow patterns and boundary shear stress due to secondary currents in straight open-channels. *Advances in Water Resources*, *33*(9), 1062-1074.

Bonakdari, H., & Moazamnia, M. (2015). Modeling of velocity fields by the entropy concept in narrow open channels. *KSCE Journal of Civil Engineering*,*19*(3), 779-789.

Chow, V. T. (1959). Open-channel hydraulics. In Open-channel hydraulics. McGraw-Hill, New York, United States

Christensen, B. and Fredsoe, J. (1998) Bed shear stress distribution in straight channels with arbitrary cross section. Lyngby, Denmark, Institute of Hydrodynamics and Hydraulic Engineering, Technical University of Denmark.

Clauser, F. H. (1956). The turbulent boundary layer. Advances in applied mechanics, 4, 1-51.

Cobaner, M., Seckin, G., Seckin, N., & Yurtal, R. (2010). Boundary shear stress analysis in smooth rectangular channels and ducts using neural networks. *Water and Environment Journal*, *24*(2), 133-139.

De Cacqueray, N., Hargreaves, D. M., & Morvan, H. P. (2009). A computational study of shear stress in smooth rectangular channels. *Journal of Hydraulic Research*, *47*(1), 50-57.

Eckmann, J. P., Kamphorst, S. O., & Ruelle, D. (1987). Recurrence plots of dynamical systems. *Europhys. Lett*, 4(9), 973-977.

Einstein, H.A. (1942) Formulas for transportation of bed load. *Transactions of the American Society of Civil Engineers,* ASCE*,* 68 (8, Part 2): 561-577.

Fix, E., & Hodges Jr, J. L. (1951). Discriminatory analysis-nonparametric discrimination: consistency properties. California Univ Berkeley.

Flintham, T. P., & Carling, P. A. (1988). Prediction of Mean Bed and Wall Boundary Shear in Uniform and Compositely Rough Channels. In International Conference on River Regime. Hydraulics Research Limited, Wallingford, Oxon UK. 1988. p 267-287

Gessner, F. B. (1973). The origin of secondary flow in turbulent flow along a corner. *Journal of Fluid Mechanics*, 58(01), 1-25.

Ghosh, S. N., & Roy, N. (1970). Boundary shear distribution in open channel flow. *Journal of the Hydraulics Division*.

Guo, J. and Julien, P.Y. (2005) Shear stress in smooth rectangular open-channel flows. *Journal of Hydraulic Engineering,* ASCE*,* 131 (1): 30-37.

Hollick, M. (1976). Boundary Shear Stress Measurement by Preston Tube. *Journal of the Hydraulics Division*, *102*(7), 1053-1057.

Hollingshead, A. B., & Rajaratnam, N. (1980). A calibration chart for the Preston tube. *Journal of hydraulic research*, *18*(4), 313-326.

Hwang, L. S., & Laursen, E. M. (1963). Shear Measurement Techniques for Rough Surfaces. *Journal of the Hydraulics Division*, *89*(2), 19-37.

Jin, Y. C., Zarrati, A. R., & Zheng, Y. (2004). Boundary shear distribution in straight ducts and open channels. *Journal of hydraulic engineering*, *130*(9), 924-928.

Julien, P.Y. (1995) Erosion and sedimentation. Cambridge University Press, Cambridge

Khodashenas, S.R. and Paquier, A. (1999) Geometrical method for computing the distribution of boundary shear stress across irregular straight open channels. *Journal of Hydraulic Research*, IAHR, 37 (3): 381-388.

Khodashenas, S.R., El Kadi, A.K. and Paquier, A. (2008) Boundary shear stress in open channel flow: A comparison among six methods. *Journal of Hydraulic Research,* IAHR*,* 46 (5): 598-609.

Knight, D.W. and Sterling, M (2000) Boundary shear in circular pipes running partially full. *Journal of Hydraulic Engineering, ASCE*, 126(4): 263-275.

Knight, D.W. (1981) Boundary shear stress in smooth and rough channels. *Journal of the Hydraulics Division,* ASCE*,* 107 (7): 839–851.

Knight, D.W. and Demetriou, J.D. (1983) Flood plain and main channel flow interaction. *Journal of Hydraulic Engineering,* ASCE*,* 109 (8): 1073-1092.

Knight, D.W. and Patel, H.S. (1985) Boundary shear in smooth rectangular ducts. *Journal of Hydraulic Engineering,* ASCE*,* 111 (1): 29-47.

Knight, D.W., Demetriou, J.D. and Hamed, M.E. (1984) Boundary shear in smooth rectangular channels. *Journal of Hydraulic Engineering,* ASCE*,* 110 (4): 405-422.

Knight, D.W., Yuen, K.W.H. and Al-Hamid, A.A.I. (1994) Boundary shear stress distributions in open channel flow. In Beven, K., Chatwin, P. & Millbank, J. (Eds.) *Physical Mechanisms of Mixing and Transport in the Environment.* John Wiley & Sons, New York, 51-87.

Leighly, J. B. (1932) Toward a theory of the morphologic significance of turbulence in the flow of water in streams. *Geography,* University of California Publications, 6(1): 1-22.

Li, Z., & Zhang, Y. K. (2008). Multi-scale entropy analysis of Mississippi River flow. *Stochastic Environmental Research and Risk Assessment*, *22*(4), 507-512.

Lundgren, H. and Jonsson, I.G. (1964) Shear and velocity distribution in shallow channels. *Journal of the Hydraulics Division*, ASCE, 90 (1): 1-21.

Marwan, N., Romano, M. C., Thiel, M., & Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Physics Reports*, *438*(5), 237-329.

Nikuradse J. (1933) Translates to: Laws of flow in rough pipes, Verein deutscher Ingenieure, *Forschubgsheft*, No. 361, Berlin.

Olivero, M., Aguirre-Pey, J. and Moncada, A. (1999) Shear stress distribution in rectangular channels. Proceeding of XXVIII IAHR Congress, Graz, Austria, papers on CD-ROM, pp. 6.

Patel, V.C. (1965) Calibration of the Preston tube and limitations on its use in pressure gradients. *Journal of Fluid Mechanics*, 23:185-208.

Petrie, J., & Diplas, P. (2015). Evaluation of the Logarithmic Law of the Wall for River Flows. *River Research and Applications*.

Pizzuto, J.E. (1991) A numerical model for calculating the distributions of velocity and boundary shear stress across irregular straight open channels. *Water Resources Research*, 27 (9): 2457-2466.

Prandtl, L. (1926) Uber die ausgebildete Turbulenz, Proc. 2nd Int. Cong. Appl. Mech., Zurich, 62-74; translated as Turbulent flow, Nut. Adv. Comm. Aero., Wash., Tech. Mem. no. 435, 1927.

Preston, J.H. (1954) The Determination of Turbulent Skin Friction by Means of Pitot Tubes. *Journal of the Royal Aeronautical Society,* 58 109-121.

Shiono, K. and Knight, D.W. (1988) Two-dimensional analytical solution for a compound channel. Proceedings of the 3 rd International Symposium on Refined Flow Modelling and Turbulence Measurements, Tokyo, Japan, 503-510.

Shiono, K. and Knight, D.W. (1990) Mathematical models of flow in two or multi stage straight channels. Proceedings of the Conference on River Flood Hydraulics. [Ed. White, W.R.], Wallingford, 229-238

Sterling, M. (1998) A study of boundary shear stress, flow resistance and the free overfall in open channels with a circular cross-section, *PhD thesis*, University of Birmingham

Sterling, M. and Knight, D.W. (2000) Resistance and boundary shear in circular conduits with flat beds running part full. *Proceedings of the Institution of Civil Engineers - Water and Maritime Engineering.* 142, Dec., 229-240.

Sterling, M., & Knight, D. (2002). An attempt at using the entropy approach to predict the transverse distribution of boundary shear stress in open channel flow. *Stochastic Environmental Research and Risk Assessment*, *16*(2), 127-142.

Tominaga, A., Nezu, I., Ezaki, K. and Nakagawa, H. (1989) Three-dimensional turbulent structure in straight open channel flows. *Journal of Hydraulic Research,* IAHR*,* 27 (1): 149-173.

Turk, M. and Pentland, A. (1991) Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, 3(1).

Yang, S.Q. and Lim, S.Y. (1997) Mechanism of energy transportation and turbulent flow in a 3D channel. *Journal of Hydraulic Engineering*, ASCE, 123 (8): 684-692.

Yang, S.Q. and Lim, S.Y. (2005) Boundary shear stress distributions in trapezoidal channels. *Journal of Hydraulic Research*, IAHR, 43 (1): 98-102.

Yu, G., & Tan, S. K. (2007). Estimation of boundary shear stress distribution in open channels using flownet. *Journal of Hydraulic Research*, *45*(4), 486-496.

Yuen W (1989). A study of boundary shear stress, flow resistance, and momentum transfer in open channel with simple and compound trapezoidal cross section. *PhD Thesis,* University of Birmingham.

Yuen, W. and Knight, D.W. (1990) Critical Flow in a two stage channel. *Proc. Int. Conf. on River Flood Hydraulics*, G4:267-276

Zheng, Y. and Jin, Y.C. (1998) Boundary shear in rectangular ducts and channels. *Journal of Hydraulic Engineering,* ASCE*,* 124 (1): 86-89.

Appendix A: Derivation of Eq. (16)

From (15) we have:

$$\Omega = U^T(\Gamma - \vartheta) \tag{A1}$$

Expanding gives:

$$\Omega = U^T\Gamma - U^T\vartheta \tag{A2}$$

Therefore,

$$U^T\Gamma - \Omega = U^T\vartheta \tag{A3}$$

Since the matrix *U* is orthogonal $U^T = U^{-1}$, and therefore:

$$UU^{-1}\Gamma = UU^{-1}\vartheta + U\Omega \tag{A4}$$

Finally leading to,

$$\Gamma = \vartheta + U\Omega \tag{A5}$$

Appendix B: Step-by-step guide to using the proposed approach

Step No.	Action	Detail	Comment
1	Integrate database	It will contain M data series	
2	Take sample vectors of dimension *d*	$\vec{\boldsymbol{q}}_{\boldsymbol{j}} = \{q_j, q_{j+1} .., q_{j+d-1}\}$ $\vdots$ $\vec{\boldsymbol{q}}_{\boldsymbol{k}} = \{q_k, q_{k+1} .., q_{k+d-1}\}$	
3	Calculate Euclidean distance amongst ***q***-vectors	e_{jk}: Euclidean Distance between $\boldsymbol{q_j}$ and $\boldsymbol{q_k}$	
4	Assemble RP Matrix There will be one RP Matrix per data series.	$RP = \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1N} \\ e_{21} & e_{jk} & & \vdots \\ \vdots & & \ddots & \vdots \\ e_{N1} & \cdots & \cdots & e_{NN} \end{bmatrix}$	Project RP Matrix on Cartesian space to plot RP if required
5	Form M vectors of dimension N^2	$\Gamma^T = \{e_{11}, e_{12}, \ldots, e_{NN}\}$	
6	Estimate the average Γ vector (ϑ) and its fluctuating components (ϕ_i)	$\vartheta = \frac{1}{M}\sum_{i=1}^{i=M}\Gamma_i; \ \phi_i = \Gamma_i - \vartheta; \quad i = 1, 2, ..., M$	
7	Assemble A Matrix and related covariance matrix	$A = [\phi_1, \phi_2 ...., \phi_M]; \ C = AA^T$	
8	Perform eigenvalue and eigenvector analysis on matrix C	$\mathbf{u}_l$: *l-th* eigenvector of matrix C if the reduced matrix *L* were used then, $u_l = \sum_{i=1}^{M} v_{li}\phi_i \quad l = 1, ..., M$	Alternatively, calculate eigenvectors from a reduced matrix L $L = A^T A$
9	Determine M weighting factors (*ω*) to assemble the short-dimension vectors Ω	$\omega_i = u_i^T(\Gamma - \vartheta), \quad i = 1, ..., M'$ $\Omega^T = \{\omega_1, \omega_2, .. \omega_{M'}\}$ where M' is the number of meaningful eigenvectors	There will be one Ω vector per original data series
Prediction Process			
10	Put together estimators for predicting new sets of Ω vectors for cases not included in the original database as well as for predicting the initial value of the data series (τ_0)	Given the estimation of Ω' the corresponding Γ' vector ca be calculated, $\Gamma' = U\Omega' + \vartheta$ τ_0: estimated value of shear stress at Pd=0	Estimators could be based on linear regression models or machine learning techniques
11	Infer the RP' Matrix	See Step 4	
12	Infer predicted data series	Any row in the RP' Matrix represents a normalized version of the predicted data series	Select Row 1
13	Denormalize the predicted series	Rescale the selected row in the RP' Matrix by using the estimated value of τ_0	End of prediction process

Appendix C: Weighting and β factors

Table C1. Weighting factors for trapezoidal channels - Set 1

# \ ω	ω_1	ω_2	ω_3	ω_4	ω_5
1	-1.1900	0.1480	-0.5080	0.7390	0.8080
2	-0.1180	0.1780	0.0928	-0.0506	-0.1020
3	0.0592	0.1080	-0.1420	-0.0327	0.0075
4	-0.0141	0.0015	0.0265	-0.1130	0.0994
5	0.0000	0.0000	0.0000	0.0000	0.0000

The weighting factors represent the components of the vector Ω^T in step 9 of the algorithm - as described in Appendix B, and are the base to setup the estimator for predicting new sets of Ω vectors not included in the original data set as outlined in step 10 of the same procedure.

Table C2. β factors for trapezoidal channels - Set 1

# \ β	β_1	β_2	β_3	β_4
1	0.9556	7.6825	-18.9163	0.4420
2	-1.2443	-5.2398	15.1313	-0.1711
3	0.1478	1.1778	-2.4132	-0.0767
4	0.5686	2.5755	-7.2138	0.0755
5	0.0000	0.0000	0.0000	0.0000

Table C3. Weighting factors for trapezoidal channels - Set 2

# \ λ	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8	ω_9	ω_{10}	ω_{11}	ω_{12}	ω_{13}
1	-1.730	-0.970	-0.080	0.238	0.048	0.258	0.547	-0.187	-0.482	-0.016	-0.650	1.080	1.940
2	0.178	-0.004	-0.063	-0.024	0.149	0.015	-0.061	-0.337	-0.017	0.030	-0.052	0.156	0.028
3	-0.054	-0.095	-0.176	-0.207	0.160	0.077	0.094	0.070	0.065	0.091	0.051	0.006	-0.081
4	-0.071	-0.067	0.064	0.070	0.136	-0.137	-0.098	0.015	-0.001	0.066	0.059	0.009	-0.044
5	0.004	0.110	0.047	-0.107	0.077	-0.063	-0.004	0.044	-0.054	-0.016	-0.078	-0.011	0.051
6	-0.075	0.064	0.046	-0.008	-0.032	-0.003	0.074	-0.053	-0.035	0.054	0.006	0.055	-0.093
7	0.073	-0.047	-0.019	-0.005	-0.047	-0.029	-0.005	0.067	-0.088	0.067	-0.024	0.070	-0.014
8	0.008	-0.017	0.009	0.045	0.065	0.059	0.033	0.006	-0.102	-0.048	0.006	-0.036	-0.028
9	0.006	-0.015	0.032	-0.047	-0.016	-0.004	-0.004	-0.010	-0.022	-0.027	0.075	0.014	0.018
10	-0.007	0.035	-0.052	0.018	-0.001	-0.023	0.003	-0.001	-0.018	-0.001	0.031	0.001	0.014
11	-0.018	0.018	-0.003	-0.005	-0.003	0.043	-0.055	0.007	-0.012	0.016	0.004	0.008	-0.001
12	0.002	-0.003	0.005	-0.003	-0.004	0.002	0.005	-0.012	-0.009	0.030	0.004	-0.029	0.014
13	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

The weighting factors represent the components of the vector Ω^T in step 9 of the algorithm - as described in Appendix B, and are the base to setup the estimator for predicting new sets of Ω vectors not included in the original data set as outlined in step 10 of the same procedure.

Table C4. β factors for trapezoidal channels - Set 2

# \ β	β_1	β_2	β_3	β_4
1	-0.3051	-4.0757	6.9595	0.4486
2	-0.0256	-0.9874	-1.4916	0.0220
3	0.0108	0.4969	-0.8857	0.0539
4	-0.0296	-0.5460	0.9375	0.0056
5	0.0212	0.2969	-0.5460	-0.0122
6	0.0136	0.1696	-0.3224	-0.0080
7	0.0634	0.8683	-1.7137	0.0087
8	-0.0127	-0.0032	0.1463	-0.0163
9	-0.0190	-0.1555	0.2696	0.0126
10	-0.0020	-0.0359	0.0540	0.0048
11	0.0004	-0.0260	0.0341	0.0017
12	-0.0021	-0.0428	0.0743	0.0004
13	0.0000	0.0000	0.0000	0.0000

Table C5. Weighting factors for trapezoidal channels - Set 3

# \ ω	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8	ω_9	ω_{10}	ω_{11}
1	-0.681	0.806	-0.639	-0.732	-0.220	1.040	-0.987	1.240	0.556	-0.935	0.556
2	0.029	-0.470	-0.255	-0.244	0.129	0.262	0.253	0.074	-0.121	0.101	0.242
3	0.232	-0.049	-0.103	0.094	0.051	0.155	-0.028	0.011	0.018	-0.141	-0.239
4	0.007	-0.007	0.078	-0.053	-0.171	0.174	0.004	-0.185	0.111	0.032	0.011
5	-0.003	0.059	-0.104	0.084	-0.064	-0.022	0.120	-0.043	-0.005	-0.095	0.073
6	-0.118	0.021	-0.012	0.011	-0.001	0.045	0.089	0.034	0.005	0.040	-0.113
7	0.037	-0.004	0.013	-0.096	-0.001	-0.069	0.066	0.022	0.099	-0.039	-0.028
8	-0.044	-0.075	0.007	0.067	0.010	-0.013	-0.022	0.006	0.076	-0.024	0.012
9	-0.013	0.026	-0.043	-0.012	0.064	0.000	-0.014	-0.055	0.029	0.017	0.001
10	0.016	0.000	-0.040	0.007	-0.032	-0.014	-0.013	0.017	0.019	0.047	-0.005
11	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

The weighting factors represent the components of the vector Ω^T in step 9 of the algorithm - as described in Appendix B, and are the base to setup the estimator for predicting new sets of Ω vectors not included in the original data set as outlined in step 10 of the same procedure.

Table C6. β factors for trapezoidal channels - Set 3

# \ β	β_1	β_2	β_3	β_4
1	-0.2908	-0.0080	0.2281	0.4332
2	0.0096	-0.2285	-0.1630	0.2357
3	0.0180	0.0150	-0.0858	0.0193
4	-0.0164	0.0021	0.0256	0.0143
5	0.0267	-0.0275	-0.0603	0.0119
6	0.0157	-0.0156	-0.0478	0.0183
7	-0.0082	-0.0063	-0.0115	0.0318
8	0.0075	-0.0131	-0.0240	0.0136
9	-0.0026	0.0055	0.0128	-0.0093
10	0.0022	-0.0148	-0.0063	0.0116
11	0.0000	0.0000	0.0000	0.0000

Table C7. Weighting factors for circular channels

# \ ω	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8	ω_9	ω_{10}	ω_{11}
1	0.575	0.045	-1.390	0.645	0.382	-0.580	-1.590	1.010	0.587	-0.815	1.130
2	0.342	0.236	0.298	0.153	-0.130	-0.243	0.015	0.089	-0.548	-0.264	0.052
3	-0.113	-0.081	0.116	-0.329	0.094	-0.148	0.041	0.025	-0.079	0.067	0.408
4	0.062	0.002	0.029	-0.065	-0.056	0.181	-0.186	-0.008	-0.086	0.107	0.020
5	0.019	-0.010	0.150	-0.055	-0.076	-0.116	-0.093	0.099	0.130	0.037	-0.086
6	0.094	0.065	0.019	-0.060	0.078	-0.023	-0.037	-0.153	0.065	-0.040	-0.009
7	-0.071	-0.025	0.063	0.118	0.016	-0.036	-0.061	-0.079	-0.006	0.040	0.039
8	-0.025	-0.047	0.065	-0.001	0.020	0.070	-0.009	0.013	0.019	-0.113	0.008
9	-0.010	-0.016	0.005	-0.009	0.109	-0.013	-0.023	0.035	-0.042	0.018	-0.055
10	0.041	-0.058	-0.000	0.008	-0.000	-0.006	0.011	-0.007	0.000	0.009	0.002
11	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

The weighting factors represent the components of the vector Ω^T in step 9 of the algorithm - as described in Appendix B, and are the base to setup the estimator for predicting new sets of Ω vectors not included in the original data set as outlined in step 10 of the same procedure.

Table C8. β factors for circular channels

# \ β	β_1	β_2	β_3	β_4
1	17.021	3.8897	-91.674	0.2276
2	6.0396	0.2841	-28.183	-0.0307
3	0.1384	0.3188	-0.7383	-0.0455
4	0.1545	-0.2010	-0.8489	0.0717
5	0.7182	0.2648	-2.7239	-0.1242
6	0.4712	-0.2330	-2.1671	0.0612
7	0.0934	0.0730	-0.4932	-0.0118
8	-0.3949	-0.1719	1.8068	0.0465
9	-0.3392	-0.1728	1.4864	0.0504
10	0.0485	0.0327	-0.2667	-0.0047
11	0.0000	0.0000	0.0000	-0.0001