Environmental Modelling & Software 26 (2011) 1526–1539

Contents lists available at SciVerse ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Uncertainty and sensitivity analysis of a depth-averaged water quality model for evaluation of *Escherichia Coli* concentration in shallow estuaries

L. Cea [a,*], M. Bermúdez [a], J. Puertas [b]

[a] E.T.S. Ingenieros de Caminos Canales y Puertos, Universidad de A Coruña, 15071, A Coruña, Spain
[b] Environmental and Water Engineering Group (GEAMA), Civil Engineering School, University of A Coruña, Spain

ARTICLE INFO

Article history:
Received 21 August 2010
Received in revised form
2 August 2011
Accepted 3 August 2011
Available online 14 September 2011

Keywords:
Water quality modelling
Depth-averaged modelling
Estuary
Sensitivity
Uncertainty
Escherichia coli
Monte carlo simulations
ANOVA

ABSTRACT

Sensitivity and uncertainty analysis investigate the robustness of numerical model predictions and provide information about the factors that contribute most to the variability of model output, identifying the most important parameters for model calibration. This paper presents a sensitivity and uncertainty analysis of a 2D depth-averaged water quality model applied to a shallow estuary. The model solves the mass transport equation for *Escherichia Coli*, including the effects of water temperature, salinity, solar radiation, turbulent diffusion and short wave dispersion. The sensitivity of the concentration of *E. Coli* in the estuary to input parameters and the different sources of uncertainty are studied using Global Sensitivity Analysis based on Monte Carlo simulation methods and sensitivity measures based on linear and non-linear regression analysis, in order to aid modellers in the calibration process and in the interpretation of model output. The extinction coefficient of light in water and the depth of the vertical layer over which the *E. coli* spread were found to be the most relevant parameters of the model. In the shallowest regions of the estuary errors in the bathymetry are also an important source of uncertainty on model output. Globally, the combination of these three parameters was found to be very effective for calibration purposes in the whole estuary.

1. Introduction

The flow in complex estuaries is commonly studied with the aid of sophisticated hydraulic and water quality models. Model predictions are often treated as deterministic data, even though they are subject to uncertainties which originate from a number of sources as model formulation, model parameters, input data, and calibration data. The stochastic component of 2D shallow water models is often overlooked, and uncertainty analysis of model results is rarely done in everyday applications. However, uncertainties cannot be eliminated from a numerical model and therefore, it is necessary to understand how they propagate through the model. To this regard it is important to determine the relevance of the different processes being modelled, and to identify the main sources of uncertainty. Sensitivity and uncertainty analysis are useful tools for model calibration, to determine the model prediction capability, and to establish the reliability of model results (Saltelli et al., 2008). Identifying the most relevant parameters of a model allows us to identify the physical processes responsible for model output and therefore, aids modellers in the interpretation of the numerical results, e.g. if solar radiation is identified as the most relevant parameter in a model, any pattern in the results might be explained to a great extent by the solar radiation input data. It is generally accepted that identification of uncertainties in environmental models is indispensable in providing decision makers with realistic information about the model outcomes (Warmink et al., 2010).

Increasing the conceptual complexity of a numerical model has the potential to improve the accuracy of the results as new processes are accounted for, but at the same time the uncertainty in model output may increase due to the uncertainty in the new input data required. Several studies show that increasing the structural complexity of a water quality model does not necessarily increase model accuracy (Lees et al., 2000; der Perk, 1997). However, increased model complexity hinders the formal evaluation of uncertainty, due to the large number of model components to be simultaneously analysed. A huge effort in model development and calibration might produce little improvement in the results if the model output is not sensitive to the calibration parameters or if the uncertainty in the input parameters is large. Moreover, the inclusion of processes that cannot be identified from the data might lead to unjustified modelling effort and ill-founded conclusions,

* Corresponding author. Tel.: +34 981167000x1492.
E-mail address: lcea@udc.es (L. Cea).

1364-8152/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.envsoft.2011.08.001

increasing the danger of busting data uncertainty (McIntyre, 2004). Hence, increasing the complexity of a numerical model is justified as long as the prediction capability of the model increases while the uncertainty of model output remains within acceptable limits. Parameters associated with processes not relevant might not be included in the model formulation to simplify the computations, while those associated with relevant processes should be carefully calibrated.

Two-dimensional depth-averaged models are the most commonly used to study the tidal hydrodynamics in non-stratified shallow estuaries (Cea et al., 2006; French, 2010; Kashefipour et al., 2006). Previous sensitivity analysis of 2D hydrodynamic models of rivers and estuaries have helped to improve the interpretation of numerical results and the calibration of model parameters (Cea et al., 2006; French, 2008; Guinot and Cappelaere, 2009). Compared to hydrodynamic models, water quality models depend on a larger set of input parameters and formulations. Several sensitivity analysis of zero and 1D water quality models have been published recently (Estrada and Diaz, 2010; Hellweger, 2007; Kneis et al., 2009; Manache and Melching, 2008; Marsili-Libelli and Giusti, 2008; Vandenberghe et al., 2007). However, only a few sensitivity and uncertainty studies of two- and three-dimensional estuarine water quality models are found in the literature (Pastres and Ciavatta, 2005; French, 2010).

This paper presents a sensitivity and uncertainty analysis of a 2D water quality model for evaluation of *Escherichia Coli* concentration in shallow estuaries. The water quality model used solves the mass transport equations for *E. coli*, including the effects of water temperature, salinity, solar radiation, turbulent diffusion, and short wave dispersion. The model is applied to the estuary of Ferrol, in the North West of Spain. This is a rather narrow and shallow estuary well-mixed in winter, but where some stratification events due to temperature gradients might occur in the summer season (Bode et al., 2005). The sensitivity of *E. coli* concentration in the estuary to a number of input parameters and data such as turbulent diffusion, reaction constants, water temperature, salinity, solar radiation, wave field, and bathymetry offsets is analysed. A local sensitivity analysis is performed first to determine the relevance of turbulence modelling and seasonal variations of water depth, solar radiation and wave field. A global sensitivity analysis based on Monte Carlo simulations is done and variance-based sensitivity measures are evaluated using linear as well as non-linear non-parametric regression methods. Global sensitivity analysis enables the identification of the most relevant parameters whose uncertainty most affects the output, and can be used to rank variables, account for interactions between variables and fix unessential variables (Kucherenko et al., 2011). The most relevant parameters are identified to aid modellers in the interpretation of the numerical results and in the calibration of this kind of models.

2. Numerical models

The computation of water quality in estuaries requires at least a 2D model to obtain the spatial and temporal distribution of *E. coli* in the estuary. A 1D model cannot represent correctly the spatial distribution of the concentration. A 3D model can produce a more detailed description of the *E. coli* concentration and might be able to take into account the effects of stratification. However, 3D water-quality models are usually reserved for deep and wide estuaries where vertical mixing is of special importance (Cox, 2003), which is not the case of the application studied in this paper.

The water quality model used in this work includes a 2D shallow water hydrodynamic model, a turbulence model, a wave field model and an *E. coli* transport model which includes the effects of turbulent diffusion and wave dispersion. A free version of the hydrodynamic model is available for download at www.iberaula.es. The water quality model is available upon request to the authors. Experimental validation of the hydrodynamic model used in this work can be found in (Cea et al., 2006, 2007). The equations solved by these models are briefly described in the following.

2.1. Hydrodynamics

The hydrodynamic model solves the 2D unsteady depth-averaged turbulent shallow water equations to compute the water depth and the two horizontal components of the depth-averaged velocity. The hydrodynamic equations are solved with an unstructured finite volume solver based on a second order extension of the upwind scheme of Roe. The hydrodynamic equations are solved with an explicit scheme and therefore, there is a link between the maximum time step which can be used in the simulations and the size of the numerical mesh, given by the Courant-Friedrichs-Lewy condition (Toro, 2001). A detailed description of the numerical schemes implemented in the solver is not relevant to the aim of this paper and can be found in the following references (Cea et al., 2006; Cea and Vázquez-Cendón, 2010; Toro, 2001).

2.2. Turbulence models

Three depth-averaged eddy viscosity turbulence models have been considered to compute the turbulent diffusion in the hydrodynamic and *E. coli* equations: a parabolic model, a mixing-length model and a $k-\varepsilon$ model. These are the most commonly used models for shallow water flows, the parabolic model being the simplest and the $k-\varepsilon$ the most complex of the three. The parabolic model considers only the production of turbulence by bed friction, and computes the eddy viscosity as:

$$\nu_t = C_s \frac{1}{6} \kappa u_f h \tag{1}$$

where $\kappa = 0.41$ is the von Karman's constant, u_f is the bed friction velocity, and C_s is a calibration parameter, with $C_s = 1$ for uniform channel flow. The value of C_s is often increased to account for the production of turbulence due to horizontal shear and other processes not considered in Equation (1). Usual values of C_s used in real applications range from 1 to 10 (French, 2008). The mixing-length model is an algebraic model which accounts for the production of turbulence due to bed friction and horizontal velocity gradients (Cea et al. (2007). The $k-\varepsilon$ model solves transport equations for the turbulent kinetic energy and for its dissipation rate. The depth-averaged version of the model used in this work is described in detail in (Cea et al. (2007). The $k-\varepsilon$ model is computationally much more expensive than the mixing-length and parabolic models, since it requires the numerical solution of two partial differential equations to compute the eddy viscosity. In the model used in this work, the $k-\varepsilon$ equations are solved with a non-structured finite volume solver described in detail in (Cea et al. (2006, 2007).

2.3. Wave field model

As detailed in the next section, the computation of *E. coli* dispersion due to wave action requires as an input parameter the wave field in the estuary (wave height, period, and direction of propagation). The spatial distribution of these wave parameters was computed with the spectral wave model SWAN, which solves the spectral action balance equation, considering the effects of wave propagation, refraction, shoaling, generation by wind,

dissipation by bottom friction and wave breaking, and non-linear wave–wave interactions. A complete description of the SWAN model can be found in (Holthuijsen, 2007).

2.4. Escherichia Coli model

The following transport equation is solved to compute the concentration of *E. coli*:

$$\frac{\partial hC}{\partial t}+\frac{\partial q_x C}{\partial x}+\frac{\partial q_y C}{\partial y}=\frac{\partial F_{d,x}}{\partial x}+\frac{\partial F_{d,y}}{\partial y}-K_d C h \tag{2}$$

where C is the concentration of *E. coli*, usually expressed as the number of colony forming units per 100 mL of water (CFU/100 ml), and K_d is the degradation coefficient of *E. coli*. The terms ($F_{d,x}$, $F_{d,y}$) are the two components of the diffusive/dispersive flux, and they include turbulent diffusion and wave dispersion of *E. coli*. The same numerical schemes used to solve the $k-\varepsilon$ equations are used to solve Equation (2).

The effect of wave dispersion is not considered explicitly in most *E. coli* models. Instead of modelling waves in order to account for their dispersive effects, *E. coli* models often include a global non-isotropic diffusion/dispersion coefficient calibrated from field data when available, as done in (Harris et al., 2004; Kashefipour et al., 2006, 2002). These global dispersion/diffusion coefficients vary within a wide range of values in different coastal and estuarine basins (Fischer et al. (1979) and therefore require site calibration. In this work both effects are modelled with physical based formulations, since no calibration data is available.

Turbulent diffusion is assumed to be isotropic in the two horizontal directions and proportional to the eddy viscosity coefficient ν_t computed by the turbulence model. On the other hand, dispersion due to wave action occurs in the direction in which waves propagate and therefore, the dispersion coefficient is non-isotropic. The two components of the diffusive/dispersive flux are computed as:

$$F_{d,x}=h\frac{\nu_t}{S_{c,t}}\frac{\partial C}{\partial x}+hD_{xx}\frac{\partial C}{\partial x}+hD_{xy}\frac{\partial C}{\partial y} \tag{3}$$

$$F_{d,y}=h\frac{\nu_t}{S_{c,t}}\frac{\partial C}{\partial y}+hD_{yx}\frac{\partial C}{\partial x}+hD_{yy}\frac{\partial C}{\partial y}$$

$$D_{xx}=D_l\cos^2\alpha\; D_{xy}=D_{yx}=D_l\cos\alpha\sin\alpha\; D_{xx}=D_l\sin^2\alpha$$

where $S_{c,t}$ is the turbulent Schmidt number, D_l is the longitudinal wave dispersion coefficient, and α is the angle between the wave propagation direction and the x axis. The first addend in the right hand side of Equation (3) accounts for the isotropic turbulent diffusion, while the second and third addends account for the non-isotropic dispersion due to wave propagation. The turbulent Schmidt number relates the turbulent diffusivities of momentum and mass. Its value has been taken as constant, with a value of $S_{c,t}=1.7$. As shown in section 4, horizontal turbulent diffusion does not have a significant effect in model output and therefore, neither does the turbulent Schmidt number.

The longitudinal wave dispersion coefficient takes into account the dispersion of solutes due to the wave action, which is mainly caused by two factors: the wave-induced oscillatory motion of particles with the wave's period, and the time-averaged drift profile generated under surface waves. The relative significance of these two factors was studied numerically in (Law, 2000), concluding that the dispersion due to the oscillatory motion is much smaller than the dispersion introduced by the time-averaged drift, since the wave period is much shorter than the time which the solute needs to spread over the water depth (Law, 2000). The formulation proposed in (Law, 2000), which considers a time-averaged drift profile with viscous effects, has been used to compute the longitudinal wave dispersion coefficient as:

$$D_l=\frac{H^4}{T^2\Gamma_e}f_v(\varepsilon_w)\varepsilon_w=2\pi h/L \tag{4}$$

where H is the wave height, h is the water depth, L is the wave length, T is the wave period, and Γ_e is the effective depth-averaged vertical diffusivity, which is equal to $\Gamma_e=\nu_t/S_{c,t}$. The function f_v in Equation (4) depends on the ratio between the water depth and the wave length, and its mathematical expression and derivation can be found in (Law, 2000). The longitudinal wave dispersion coefficient given by Equation (4) is evaluated at each spatial point from the fields of significant wave height, peak period and main direction of wave propagation computed by SWAN.

Regarding the degradation of *E. coli*, several levels of refinement can be used to model microbial decay in coastal environments. The two main possibilities are the assumption of a constant decay rate or the inclusion of a time dependent decay rate. In this latter case, the decay rate can be obtained as a function of different parameters such as sunlight intensity, temperature or salinity. An intermediate approach is the assumption of constant day- and night-time decay rates in order to consider the effect of sunlight radiation on bacterial decline. Examples of the different approaches can be found in works such as (Garcia-Barcina et al. (2002) for a constant decay rate, in (Kashefipour et al. (2006) for constant day/night decay rates, or (Kashefipour et al., 2006; Harris et al., 2004; Canteras et al., 1995) for non-constant decay rates. In this work we use a non-constant decay rate in which the degradation coefficient K_d is computed with the following formulation proposed by Mancini (Mancini, 1978), which takes into account the effects of water temperature, salinity and solar radiation, and is considered one of the most complete models of the fecal coliform decay process (Manache et al., 2007):

$$K_d=(0.8+0.02S)1.07^{(T-293)}+0.086\frac{I_s}{k_e h_c}(1-\exp(-k_e h_c)) \tag{5}$$

where K_d is the depth-averaged degradation coefficient of *E. coli* in d^{-1}, S is the water salinity in Kg/m^3, T is the water temperature in K, I_s is the total solar radiation at the free surface in W/m^2, k_e is the absorption coefficient of light in water in m^{-1}, and h_c is the depth of the vertical layer over which the *E. coli* spread, computed as:

$$h_c=\min(h,L_c) \tag{6}$$

where L_c is a parameter specified by the user and h is the water depth. If the *E. coli* spread over the whole water column $h_c=h$. On the other hand, if the *E. coli* spread over a vertical layer thinner than the water depth, the value of h_c will be limited by the depth of this layer ($h_c=L_c$). Typical values of L_c for urban sewage discharges can be found in (Ludwig, 1988). It usually takes higher values in winter, when the estuary is well-mixed over the water depth, than in summer, when vertical stratification might occur. The models used in this work are 2D depth-averaged models which assume complete vertical mixing, and which cannot account for stratification processes and therefore, the value of L_c must be fixed by the user as an input parameter. The numerical modelling of L_c would imply 3D modelling of the velocity, turbulence, salinity, temperature and *E. coli* fields, including the effects of short wave propagation and wind. This would require an accurate modelling of vertical gradients which can only be achieved if the model includes detailed information of meteorological and oceanographic data,

such as wind stresses, short waves, rainfall, solar radiation, and coastal upwelling events.

The degradation rate of *E. coli* depends strongly on water turbidity, which diminishes the average radiation intensity in the water column. This is considered in the model via the absorption coefficient of light k_e in Equation (5). Values of k_e in coastal waters range from around 0.15 m^{-1} for clear offshore water to around 21 m^{-1} for very turbid coastal water (Emery and Thomson, 2001). After dilution of wastewater, a lag phase in which a negligible decrease in the bacterial population is observed, often precedes the bacterial decay phase (Gonzalez, 1995). The model of *E. coli* used in this paper does not consider this phenomenon.

3. Methodology

3.1. Study site

The water quality model presented in section 2 was used to compute the concentration of *E. coli* in the estuary of Ferrol (Fig. 1). The aim of the study is to assess the sensitivity of the numerical model to its various input parameters and data, in order to evaluate which are the main sources of model output variability, and to determine on which parameters a calibration process should focus.

The estuary of Ferrol is a coastal inlet originated by a deep fluvial valley invaded by the sea (Bode et al., (2005). The water of the estuary is mainly of marine origin, as the local rivers inflow is very low. The total surface of the estuary is approximately 32 Km^2, with a length of 17 Km and a maximum water depth of approximately 25 m at the mouth. The minimum, average and maximum tidal ranges are respectively 0.86 m, 2.59 m and 4.58 m. The main characteristic of the estuary is a narrow channel near the mouth, which is approximately 400 m wide, 4 Km long, and 20 m deep. Inside this channel, high current velocities occur at mean flood and ebb tides, with values in the range 0.4–0.6 m/s. In its inner part, the estuary gets wider and shallower, with water depths ranging from 5 to 15 m and with much lower velocities. In such a narrow estuary the flow is predominantly tidal and the wind has little influence in the generation of currents (Fischer et al. (1979). Therefore, wind generated currents have not been considered in the model. The discharge of local rivers is very low and it is not significant compared to the tidal flow, the water in the estuary being mainly of marine origin (Bode et al., (2005). Regarding wave action, the inner part of the estuary is very well protected against short waves by the narrow channel near the mouth. Inundation and drying of intertidal areas are not relevant in this estuary.

Both thermal and haline stratification are very weak in winter. Winter conditions are characterised by a relatively low water temperature (13 °C) and a nearly homogeneous vertical distribution of salinity, temperature and nutrients through the water column in the whole estuary (Bode et al., (2005). Differences in water temperature over the water depth remain usually lower than 0.5 °C in winter, while salinity values vary between 33 g/l and 35 g/l (Casas et al. (1997). These conditions reflect the complete mixing of the water column in winter.

During the summer, thermal and haline stratification appear in the upper part of the water column as a consequence of the increase in solar heating. The limit of the upper wind-mixing layer ranges between 2 and 8 m (Bode et al., (2005), with an average value of 5 m. Differences in water temperature over the water depth remain usually lower than 3 °C in summer, while the salinity oscillates between 30 g/l and 35 g/l (Bode et al., (2005).

3.2. Input data, boundary conditions and numerical mesh

Since the concentration of *E. coli* in the estuary depends on a combination of variables which vary through the year, as solar radiation, water temperature, wave conditions and tidal range, a whole year has been considered in order to have statistically representative results. Therefore, to capture possible seasonal differences in model sensitivities each numerical simulation extends over a period of 1 year.

The water depth and velocity fields in the estuary are computed using a tidal level time series obtained from the main 25 tidal constituents. The harmonic constituents were evaluated from the tide levels registered in the tide gauge of A Coruña over a 7 year period, and are available at www.puertos.es (Fig. 2). We have considered a period of 15 days which represents properly both spring and neap tides in the estuary (from the 1st to the 15th of January in Fig. 2). The velocity and water depth fields computed during that period were stored at regular intervals of 30 min (720 fields for the 15 days) and used afterwards to compute the convective term in Equation (2) for the one year long water quality computations.

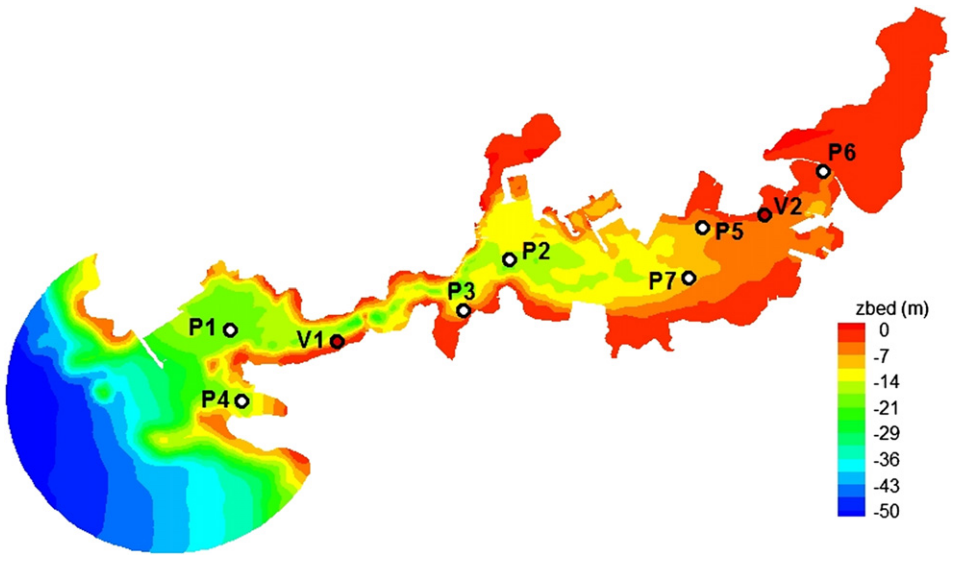

Fig. 1. Bathymetry of the Ferrol estuary used in the numerical model, relative to mean sea level at the mouth. Location of sewage spills (V1 and V2) and control points (P1-P7).

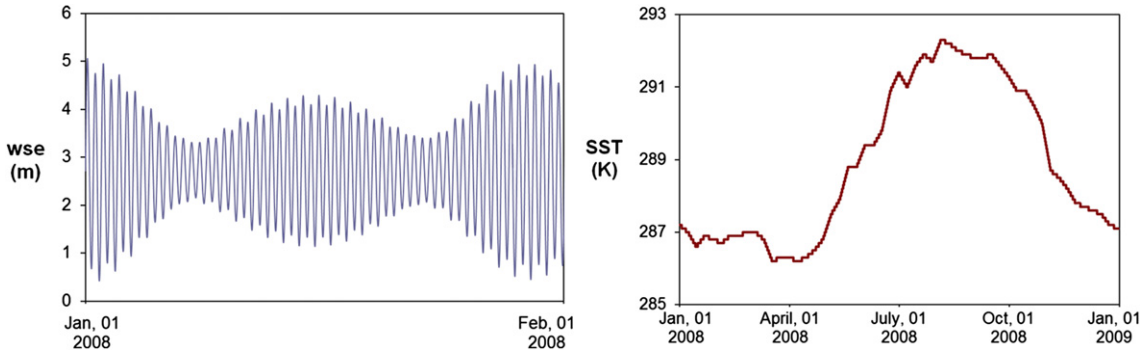

Fig. 2. Extract of 1 month of tidal time series at the mouth of the estuary (left) and 1 year long time series of sea surface temperature (right).

The water temperature in Equation (5) is imposed from weekly measurements of the sea surface temperature (SST) obtained from the NOAA/OAR/ESRL PSD (Boulder, Colorado, USA) web site at www.esrl.noaa.gov/psd (Fig. 2). The solar radiation is imposed from measurements at time intervals of 10 min obtained from the meteorological station CIS Ferrol (www.meteogalicia.es) (Fig. 3).

The longitudinal wave dispersion coefficient in Equation (4) is computed at each spatial point inside the estuary as a function of the significant wave height and peak period computed by SWAN. The wind fetch inside the estuary is rather small and therefore, wind has been neglected in the computation of the wave field. The significant wave height, peak period and direction of propagation at open sea were imposed in SWAN at the open boundary. The wave height and period were imposed as a one year long time series (Fig. 3). These data were obtained from the scalar wave buoy of A Coruña, located 5 km away from the mouth of the estuary at a water depth of 50 m. Wave directional information was obtained from the directional wave buoy of Langosteira, located 15 km away from the mouth of the estuary, and from the WANA point 1046074. Wave direction is mainly from the North West (Fig. 3) and thus, it has been considered as a constant input parameter in the simulations.

Two sewage spills were located at two different points in the estuary, V_1 and V_2 in Fig. 1. Spill V_1 is located in a region with high current velocities and will therefore disperse rapidly, reaching a large area of the estuary. Spill V_2 is located in the inner part of the estuary, in a region with lower velocities where dispersion of contaminants is much slower and therefore, higher concentrations of *E. coli* are reached (Fig. 4). The continuous spills simplify the identification of the parameters which are more relevant in the modelling, and allows us to compare easily the model behaviour in different seasons of the year.

All the numerical computations were performed in a non-structured finite volume mesh with 22761 control volumes, the average mesh size being approximately 35 m (Fig. 5), with a time step of 5 s. Spatial resolution of the numerical mesh was determined after a mesh convergence analysis, searching a balance between computational cost and numerical accuracy. The same grid was used to solve the hydrodynamic, turbulence, *E. coli*, and wave spectral action equations.

3.3. Performance and sensitivity measures

Calibration and sensitivity analysis of 2D shallow water models is generally made through the comparison of modelled and observed data sets at a number of control points, using different performance measures to quantify the agreement between model output and observed data. In complex 2D estuarine models, different parameter sets can produce similar model results (French, 2010), specially when model output is evaluated using a single performance measure at single control point. It is therefore desirable to analyse model sensitivity at multiple control points using different performance measures. In this work sensitivity is analysed based on the time series of *E. coli* concentrations at seven control points distributed over the whole estuary (Fig. 1).

The annual average concentration of *E. coli* (C_m), as well as the concentrations exceeded 10% and 90% of the time (C_{10},C_{90}), are considered to compare the output results for different model runs. In addition, the Nash-Sutcliffe Efficiency (NSE), defined as the ratio of the error variance to the variance of the observed time series (Nash and Sutcliffe, 1970), is also used to compare different model runs. The NSE is normally used to evaluate model performance by comparing numerical results against measured data, being defined as:

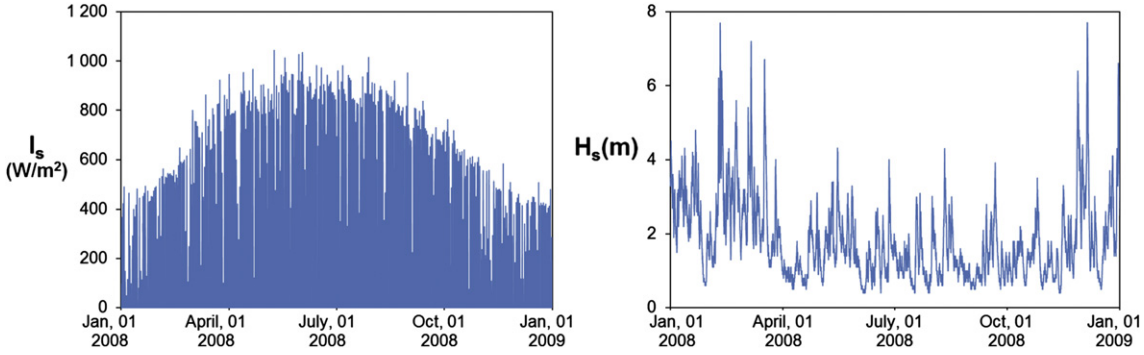

Fig. 3. Annual time series of solar radiation (left) and significant wave height data outside the estuary from bouy of A Coruña (right).

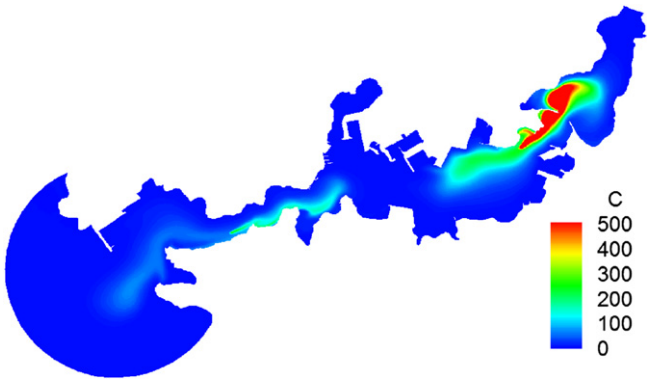

Fig. 4. *E. coli* concentration in the estuary at low tide (reference simulation).

$$NSE = 1 - \frac{\sum_{i=1}^{N} (\hat{C}_i - C_i)^2}{\sum_{i=1}^{N} (\hat{C}_i - \hat{C}_m)^2} \quad (7)$$

where $\hat{C}_i$ is the observed concentration at time t_i, $\hat{C}_m$ is the time average of the observed concentration, and C_i is the computed concentration at time t_i. Since no experimental data is available, in this work the NSE is used to evaluate model sensitivity by comparing the results of a reference simulation with the results obtained with different input parameter sets. In that case the values of the observed concentration, $\hat{C}_i$ and $\hat{C}_m$ in Equation (7), are replaced by the values of the concentration computed in the reference simulation. In this context, a value of NSE = 1 means that the model output is not sensitive to the input parameter considered and therefore, the process represented by that parameter is not relevant in the modelling (Mayer and Butler, 1993). The lower is the value of NSE, the higher is the sensitivity to that parameter.

The parameters of the reference simulation are fixed according to the physical characteristics of the estuary, as described in section 3.1. Each input parameter is given an average value defined from our knowledge of the estuary and from the meteorological data described in section 3.2. To perform a global sensitivity analysis of the model, in section 4 the input data and parameters of the reference simulation are modified according to their uncertainty. The uncertainty level assumed for each parameter is justified in section 4.4. The average value of the input parameters used in the reference simulation and the range of variation assumed in the global sensitivity analysis is shown in Table 5. The values of C_m, C_{10} and C_{90} computed in the reference simulation are shown in Table 1.

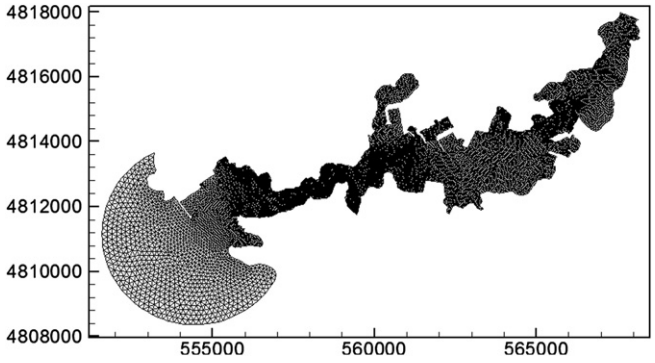

Fig. 5. Finite volume mesh used in the numerical computations.

Table 1
Mean water depth h_m (m), annual average concentration C_m (CFU/100 ml), and percentiles C_{10}, C_{90} at the 7 control points defined in Fig. 1.

	P1	P2	P3	P4	P5	P6	P7
h_m	22.6	19.8	9.5	19.3	10.3	4.7	13.0
C_m	36.1	25.1	78.0	3.62	142.0	278.0	18.5
C_{10}	73.2	66.0	195.0	11.6	391.0	586.0	61.8
C_{90}	4.58	0.191	3.86	0.027	7.72	36.6	0.107

Monte Carlo simulations have been used to study global sensitivity of the model outputs to the input parameters. In non-linear models, as the one analysed in this paper, global sensitivity measures are generally preferred to local measures, since the former ones provide information over the total range of variation of the input parameters and are able to identify interactions between parameters (Saltelli et al. (2008). A recent revision on the drawbacks of local sensitivity analysis in comparison to global sensitivity analysis is given in (Saltelli and Annonis, 2010). Practical applications of global sensitivity analysis to complex environmental models can be found in (Estrada and Diaz, 2010; Manache and Melching, 2008; Pastres and Ciavatta, 2005).

The following procedure has been followed. First, a sensible range for each input parameter was defined. Then, a number of input parameter sets was randomly generated, and the numerical model was run for each parameter set. We have used a Sobol quasi-Monte Carlo low-discrepancy sequence (Sobol, 1998; Saltelli et al., 2008) to generate 150 random sets of 8 parameters each. Low-discrepancy sequences provide very good convergence properties and are therefore very adequate sampling techniques for computationally demanding models, because they allow for the extraction of a large amount of sensitivity information with a relatively small sampling size (Sobol, 1976; Saltelli et al., 2008). The results of all the Monte Carlo computations were analysed using Sobol sensitivity indices (Saltelli et al. (2008) based on both linear and non-linear non-parametric regression of model output, which provide unambiguous information on the importance of different subsets of input variables to the output variance (Kucherenko et al. (2011).

Sensitivity measures based on a linear regression model are valid when the coefficient of determination (R^2) of the linear model is close to one. For that purpose, it can be assumed that the minimum R^2 that defines a good fit of model output is somewhere in the range [0.75, 0.90] (Storlie et al. (2009). In that case, if the correlation between input parameters is low, the Standardized Regression Coefficients (Storlie and Helton, 2008) are a very effective measure to establish a ranking of the input parameters according to the uncertainty they generate on model output (Manache and Melching, 2008).

If the model is non-linear, variance-based methods can be used to compute sensitivity indices which describe the non-linear main effects of each parameter as well as the interactions between parameters. Variance-based methods are described in detail in (Saltelli et al. (2008) and can be defined by referring to a decomposition of the model output y into terms of increasing dimensionality (High Dimensional Model Representation, HDMR) as:

Table 2
Sensitivity to weekly variations of water temperature with $T_m = 288.9$K. Reference simulation computed with water temperature time series from Fig. 2.

	P1	P2	P3	P4	P5	P6	P7
Error in C_m (%)	−2.84	−1.94	−1.00	−4.33	−1.40	−1.09	−4.81
NSE	0.99	0.99	1.00	0.99	1.00	1.00	0.99

$$y = f(x_1, x_2, \ldots, x_k) = f_0 + \sum_i f_i + \sum_i \sum_{j>i} f_{ij} + \ldots + f_{12\ldots k} \quad (8)$$

where $f_0 = E(y)$ is a constant equal to the average of y, $f_i = f(x_i)$, $f_{ij} = f(x_i,x_j)$, $f_{ijk} = f(x_i,x_j,x_k)$, and so on. The main effect of the parameter x_i on model output is given by the function f_i, and its first order variance-based sensitivity index is defined as:

$$S_i = V(f_i)/V(y) \quad (9)$$

where $V(f_i)$ is the variance of f_i and $V(y)$ is the unconditional variance of y. The functions f_{ij} are known as the second order effects, and measure the interaction between pairs of parameters. The second order sensitivity indices are defined as:

$$S_{ij} = V\left(f_{ij}\right)/V(y) \quad (10)$$

If each term in the HDMR given by Equation (8) has zero mean, and all the terms are orthogonal in pairs, the following relation (ANOVA-HDMR decomposition) applies (Saltelli et al. (2008)):

$$\sum_i S_i + \sum_i \sum_{j>1} S_{ij} + \ldots + S_{12\ldots k} = 1 \quad (11)$$

If there is no interaction between parameters $\sum S_i = 1$ (additive model). In that case the second and higher order effects are zero $S_{ij} = \ldots = S_{12\ldots k} = 0$. The total effect of a parameter x_i accounts for the total contribution of parameter x_i to model output (first order plus high order contributions due to interactions with other parameters) and is measured as the sum of all the sensitivity indices in Equation (11) which include the parameter x_i. If the total effect of a parameter is nearly zero, that parameter can be fixed anywhere over its range of variability without appreciably affecting the value of the output variance (Saltelli et al. (2008). The total effect can be also explained as the percentage of the unconditional variance $V(y)$ that would be left on average if the value of all the other parameters are fixed.

To compute the previous sensitivity measures in an efficient way with a relatively low number of model evaluations, a non-linear non-parametric metamodel of the original model is estimated and the sensitivity measures are based on this metamodel (Saltelli et al. (2008). In this work, the terms in a truncated ANOVA-HDMR decomposition which considers only the main effects and the second order interaction terms in Equation (8) have been estimated using the non-parametric approach presented in (Ratto et al., 2007; Ratto and Pagano, 2010). The method applies State Dependent Parameter modelling to build an approximation of the computational model under analysis and to estimate the first and second order variance based sensitivity indices.

4. Results and discussion

The results obtained from the sensitivity analysis presented in this section are valid only for 2D depth-averaged models coupled to an *E. coli* model with a similar structure to the one described in section 2. It is not valid neither for more sophisticated models (3D models which compute stratification), nor for simpler models (1D models or 2D models using other water quality formulations). Moreover, the specific results presented in this section are valid for estuaries with the same characteristics as the estuary of Ferrol, i.e. slightly or non-stratified shallow estuaries, with a low freshwater river inflow, and with an elongated deep channel morphology.

Table 3
Sensitivity to daily oscillation of solar radiation with $I_{s,m} = 37$ W/m^2. Reference simulation computed with solar radiation time series from Fig. 3.

	P1	P2	P3	P4	P5	P6	P7
Lc = 5 m							
Error in C_m (%)	−32.58	−25.57	−20.69	−41.44	−22.41	−22.18	−42.28
NSE	0.40	0.65	0.71	0.55	0.76	0.70	0.45
Well-mixed							
Error in C_m (%)	−6.33	−6.23	−6.44	−10.43	−10.76	−17.57	−21.22
NSE	0.86	0.93	0.92	0.90	0.87	0.76	0.72

Table 4
Sensitivity to wave field. Reference simulation computed with significant wave height time series from Fig. 3.

	P1	P2	P3	P4	P5	P6	P7
Error in C_m (%)	7.59	5.88	−7.75	14.85	−0.63	−1.75	4.55
NSE	0.98	0.99	0.99	0.98	1.00	1.00	1.00

4.1. Sensitivity to the turbulence model

In order to analyse the relative contribution of horizontal turbulent diffusion to model output, and which is the improvement in model performance when a complex turbulence model is used rather than a simple one, the $k-\varepsilon$ model, the mixing-length model and the parabolic model were used to compute the eddy viscosity, and the results were compared with those of a simulation in which the turbulent diffusion was neglected ($\nu_t = 0$). The numerical results show that the *E. coli* concentration is not sensitive to the formulation used to compute the eddy viscosity. If the $k-\varepsilon$ model is used to compute the eddy viscosity, the NSE of the model output and the annual average concentration C_m differ in less than 1% with the case in which the horizontal turbulent stresses are neglected ($\nu_t = 0$). This indicates that the role of horizontal turbulence is less relevant than other processes involved in the transport of *E. coli*. Therefore, it seems sensible to run the model with a simple turbulence closure, rather than to use a complex model which does not have a significant impact on the results. The fact that most water quality studies in estuarine waters either use a simple model to compute horizontal turbulent diffusion (Hogdins et al., 1998; Kashefipour et al., 2006) or neglect its contribution to model output (Garcia-Barcina et al. (2002) is consistent with these results. Velocity and water depth fields not sensitive to the turbulence model have been found previously in several shallow estuarine models (Ali et al., 2009; Cea et al., 2006; Lane, 2004).

4.2. Sensitivity to weekly variations of water temperature and daily oscillations of solar radiation

The value of the pathogen decay rate K_d depends on the water temperature and the sunlight intensity, which vary with time. However, the use of constant values for the die-off rates seems to be a common approach in the applied model studies reported in the literature (Garcia-Barcina et al., 2002; Kashefipour et al., 2006). Nevertheless, the use of a constant K_d in continuous modelling may result in problems due to the considerable changes

Table 5
Parameter ranges considered in the Monte Carlo simulations.

Parameter	Min	Max	$\bar{x}$	σ_x
ΔZ_b	−0.5	0.5	0.0	0.288
L_c	5	21	12.7	4.89
k_e	1	3	2.0	0.577
I_{s0}	0.8	1.2	1.0	0.115
ΔT	−2	2	0.0	1.15
S	30	35	32.5	1.44
C_s	1	10	5.35	2.71
H_{s0}	0.75	1.25	1.0	0.144

Table 6
Mean and standard deviation (SD) values of C_m, C_{10} and C_{90} computed from the 150 Monte Carlo simulations.

	P1	P2	P3	P4	P5	P6	P7
Mean C_m	3.45E + 01	2.42E + 01	7.37E + 01	3.80E + 00	1.40E + 02	2.70E + 02	1.77E + 01
SD C_m	6.44E + 00	3.72E + 00	7.91E + 00	9.80E − 01	1.91E + 01	2.85E + 01	3.78E + 00
Mean C_{10}	7.35E + 01	6.33E + 01	1.80E + 02	1.04E + 01	3.90E + 02	5.80E + 02	5.93E + 01
SD C_{10}	1.10E + 01	7.59E + 00	1.50E + 01	2.50E + 00	4.00E + 01	4.55E + 01	1.17E + 01
Mean C_{90}	4.53E + 00	1.95E − 01	3.70E + 00	4.14E − 02	7.40E + 00	3.35E + 01	1.06E − 01
SD C_{90}	1.91E + 00	1.20E − 01	1.31E + 00	3.00E − 02	3.22E + 00	1.55E + 01	7.00E − 02

with time of solar radiation and water temperature (Manache et al., 2007)). The use of a variable decay rate for day/night, for wet/dry weather and for summer/winter conditions, has been found to be important for predicting the *E. coli* concentrations accurately in several models (Kashefipour et al., 2006, 2002; Harris et al., 2004; Pommepuy et al., 2006). In these studies, decay rate values in seawater varying from 0.02 h^{-1}–4 h^{-1} are reported, depending mainly on the meteorological and water turbidity conditions. Therefore, its value must either be calibrated from field data or evaluated from a radiation dependent formulation, as Mancini formulation. In our case study, with the meteorological and physical conditions introduced in the model, the decay rate given by the formulation of Mancini varies over the year within 0.04 h^{-1} and 0.62 h^{-1} for the reference simulation, and between 0.03 h^{-1} and 2 h^{-1} if we consider the 150 Monte Carlo simulations.

To check the relevance of weekly variations of temperature in model output, a simulation was done with a constant water temperature equal to its annual average (T_m = 288.9K), and the results were compared with those obtained using the weekly measurements of SST defined in Fig. 2. The results shown in Table 2 indicate that differences in model output are in general low, although larger than the influence of the turbulence model.

The sensitivity to daily and seasonal oscillations of solar radiation was analysed by comparing a simulation done with a constant value of solar radiation equal to its annual average ($I_{s,m} = 37$ W/m^2) with a simulation in which the solar radiation was imposed from measurements at time intervals of 10 min (Fig. 3). *E. coli* concentrations are much more sensitive to daily oscillations in solar radiation under stratified conditions than under well-mixed conditions, as shown in Table 3. This is because solar radiation is higher near the water surface, and diminishes quickly with the water depth. When the estuary is under stratified conditions, most part of *E. coli* are near the free surface receiving a strong radiation intensity, and the degradation rate is higher. In any case, even under well-mixed conditions, sensitivity to daily oscillations in solar radiation is much higher than sensitivity to weekly oscillations in water temperature. For practical purposes, in long-term computations it might be adequate to consider a constant water temperature equal to its annual average, but not a constant solar radiation. Based on these results, the use of a constant value for the die-off rate is not recommended. The use of different day–night constants, or time varying decay rates dependent upon solar radiation can achieve better predictions, as pointed out in (Kashefipour et al. (2006).

4.3. Sensitivity to the wave field

Short waves produce a dispersive effect in the transport of *E. coli*. This effect is taken into account in the model with the wave dispersion coefficient given by Equation (4). To analyse the importance of considering wave dispersion in the model, the wave field in the estuary was computed imposing as wave boundary condition at the mouth a one year long time series of significant wave height with a time resolution of 4 h (Fig. 3). The comparison of the results obtained with and without considering the wave field in the estuary show that, even in a narrow estuary well-protected against wave action, if wave dispersion is not considered in the model, significant differences in the annual average concentration of *E. coli* occur in the outer and middle estuary, specially at the control points P1, P3 and P4, but also at P2 and P7 (Table 4). On the other hand, in the most inner and shallow part of the estuary (control points P5 and P6) the sensitivity to wave dispersion is much lower.

4.4. Global sensitivity analysis. monte carlo simulations

A global sensitivity analysis was carried out based on Monte Carlo simulations. The 8 following model inputs were considered in the analysis: bathymetry offsets, depth of the vertical layer over which *E. coli* spread, absorption coefficient of light in water, solar radiation, water temperature, water salinity, coefficient used in the computation of the eddy viscosity, and significant wave height at the seaward boundary. For each input parameter a nominal value and a sensible range of variation were defined according to the properties of the estuary, as described in section 3.

Bathymetric errors are rarely considered in the numerical modelling of estuaries. However, when different data sources are combined to build up the bathymetry of the model, systematic errors in form of datum offsets might be introduced in the model. Offsets in the bathymetry are also likely to occur when defining the boundary condition at the mouth, due to the multiple reference levels used to measure the tidal elevation. Other sources of bathymetric errors in estuaries are the spatial resolution of the model bathymetry, and the post-survey evolution of the bed elevation due to sediment transport (French, 2008). A global uncertainty of $\Delta Z_b = \pm 0.5$ m was assumed to test the sensitivity of model output to bathymetric offsets.

Incomplete vertical mixing must be considered as a source of uncertainty in water quality models (Kneis et al. (2009). In shallow

Table 7
Coefficient of determination R^2 for the linear regression analysis performed on NSE, C_m, C_{10} and C_{90}.

	P1	P2	P3	P4	P5	P6	P7
NSE	0.258	0.267	0.322	0.095	0.056	0.046	0.141
C_m	0.949	0.944	0.937	0.942	0.957	0.981	0.934
C_{10}	0.953	0.944	0.939	0.953	0.973	0.977	0.935
C_{90}	0.944	0.900	0.935	0.838	0.897	0.981	0.846

Table 8
Coefficient of determination R^2 for the first order non-linear regression analysis performed on NSE, C_m, C_{10} and C_{90}.

	P1	P2	P3	P4	P5	P6	P7
NSE	0.446	0.456	0.543	0.287	0.461	0.688	0.464
C_m	0.990	0.991	0.995	0.987	0.996	0.996	0.991
C_{10}	0.991	0.993	0.996	0.992	0.998	0.996	0.993
C_{90}	0.977	0.934	0.980	0.861	0.973	0.986	0.915

Table 9
Coefficient of determination R^2 for the second order non-linear regression analysis performed on NSE, C_m, C_{10} and C_{90}.

	P1	P2	P3	P4	P5	P6	P7
NSE	0.995	0.995	0.995	0.966	0.995	0.997	0.988
C_m	0.998	0.991	0.995	0.985	0.997	1.000	0.991
C_{10}	0.998	0.996	0.996	0.991	0.999	0.999	0.993
C_{90}	0.976	0.914	0.980	0.840	0.974	0.999	0.876

water bodies a mixing depth larger than 3 m is very common (Padisak and Reynolds, 2003), in agreement with what was pointed out in section 3 for the estuary of Ferrol. The average depth of the mixing layer in the estuary of Ferrol in summer is about 5 m, while in winter the estuary is usually completely mixed over the water depth, being the average water depth in the estuary 21 m. Therefore, random values of L_c have been generated within the range $L_c = 5–21$ m.

The parameter k_e was assumed to vary in the range $k_e = 1–3\ m^{-1}$, based on measurements of the absorption coefficient of light in similar estuaries located in the same region as the estuary of Ferrol. These are rather high values, although values of k_e in coastal waters ranging from 0.15 m^{-1} to 21 m^{-1} have been reported in the literature (Emery and Thomson, 2001). The solar radiation is parameterized with the coefficient I_{s0}, which multiplies the solar radiation time series shown in Fig. 3. This coefficient accounts for the uncertainty introduced by the annual variability in solar radiation. From the annual average of the solar radiation registered during several years in the estuary of Ferrol (available at www.meteogalicia.es), the coefficient I_{s0} was varied within the range $I_{s0} = 0.8–1.2$.

Water temperature is imposed from weekly measurements of the sea surface temperature (SST). However, under certain conditions the depth-averaged temperature might be different from the SST. The water temperature is therefore defined from the time series shown in Fig. 2 plus a constant temperature offset ΔT which accounts for the uncertainty in water temperature. Based on the observed oscillations of salinity in the estuary throughout the year (Casas et al. (1997), salinity was varied within the range $S = 30–35$ g/l.

The sensitivity to the eddy viscosity was quantified using the parabolic model given by Equation (1). The parabolic model is commonly used with a coefficient C_s larger than one, in order to account for the contribution to the vertical turbulence of processes as wind stresses or wave breaking. In the present simulations the parabolic model was used with values of C_s varying within the range $C_s = 1–10$, which covers the usual values found in the literature (French, 2008).

To account for its annual variability, the wave height at the seaward boundary is parameterized with the coefficient $H_{s\,0}$, which multiplies the wave height time series shown in Fig. 3. From the annual average of significant wave height registered during several years (available at www.puertos.es), the coefficient H_{s0} was varied within the range $H_{s0} = 0.75–1.25$.

Considering the previous ranges of variation of the input parameters, a Sobol quasi-Monte Carlo low-discrepancy sequence (Sobol, 1998; Saltelli et al., 2008) was used to generate 150 random parameter sets. The input parameters were assumed to be statistically independent and uniformly distributed within the ranges defined in Table 5. At each one of the control points defined in Fig. 1, the variables C_m, C_{10} and C_{90} computed for each Monte Carlo run were used to analyse uncertainty in model output. The average values and standard deviation of these variables computed for the 150 Monte Carlo simulations are shown in Table 6. At the same time, the NSE of each simulation was evaluated at each control point by comparing the *E. coli* time series with those of a reference simulation computed with the average value of the input parameters (Table 5).

At each control point both linear regression analysis as well as the non-linear HDMR given by Equation (8) were used to fit the variables C_m, C_{10}, C_{90} and NSE as a function of the input parameters. The non-parametric non-linear regression of these variables was performed with a matlab toolbox for Global Sensitivity Analysis developed by the Joint Research Center of the European Commission, and freely available to download from the web site http://eemc.jrc.ec.europa.eu/EEMCArchive. While a linear regression model is able to capture most of the C_m, C_{90} and C_{10} variance, representation of NSE requires at least a second order HDMR, such as the truncated ANOVA-HDMR decomposition used in (Ratto et al., 2007; Ratto and Pagano, 2010), which includes the main effects and the second order interaction terms (Tables 7, 8 and 9). As shown in Table 9, the truncated ANOVA-HDMR decomposition considering the main effects and the second order interaction terms fits almost perfectly all model outputs. Therefore, sensitivity measurements based on linear regression (like Standardized Regression Coefficients) or based on first order non-linear regression are only justified if we want to fit C_m, C_{90} or C_{10}, but not if we are interested in the analysis of NSE.

The Standardized Regression Coefficients (SRC) obtained from the linear fit of C_m are shown in Fig. 6. The sign of the coefficient indicates if an increase in the parameter implies an increase or decrease in the concentration of *E. coli*. According to this measure, in global the most relevant parameter is k_e followed by L_c. In the inner estuary the parameter k_e by itself is able to represent more than 50% of the variance of C_m, while sensitivity to L_c is very low. On

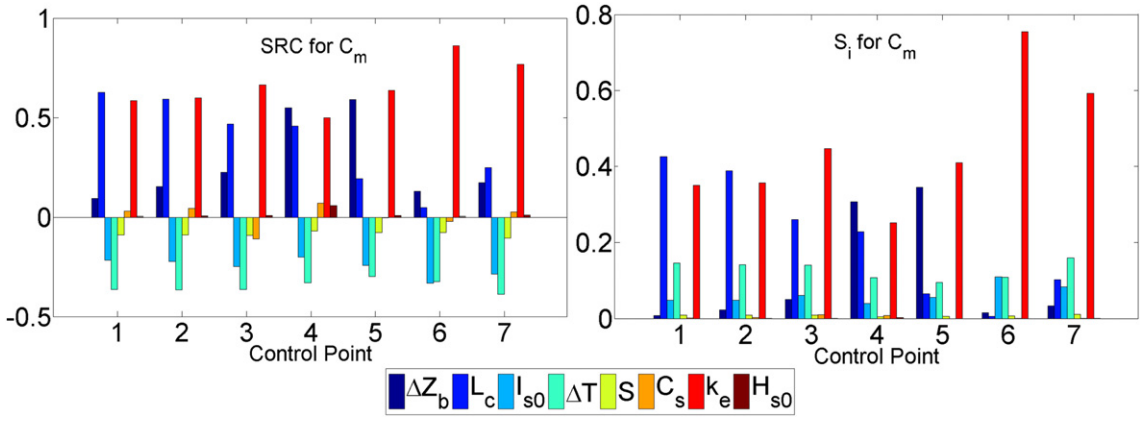

Fig. 6. Standardized regression coefficient (*SRC*) and first order sensitivity index (S_i) for C_m.

Table 10
Different sensitivity measurements for NSE calculated at control points P1 and P5. Squared standardized regression coefficient (SRC^2), first order effects (S_i) and total effects ($S_{T,i}$).

	P1			P5		
ΔZ_b	0.001	0.000	0.008	0.001	0.211	0.571
L_c	0.168	0.271	0.667	0.041	0.031	0.122
I_{s0}	0.017	0.004	0.075	0.000	0.003	0.126
ΔT	0.005	0.005	0.126	0.000	0.002	0.145
S	0.001	0.000	0.009	0.000	0.000	0.015
C_s	0.006	0.000	0.007	0.019	0.000	0.002
k_e	0.100	0.136	0.509	0.047	0.214	0.605
H_{s0}	0.000	0.000	0.001	0.000	0.000	0.000
Sum	0.298	0.416		0.108	0.461	

Table 11
Most relevant second order sensitivity indices for NSE (interactions between parameters). Second order interactions higher than 0.05 are marked with.*

	P1	P2	P3	P4	P5	P6	P7
$\Delta Z_b - L_c$	0.006	0.008	0.006	0.109*	0.027	0.000	0.003
$\Delta Z_b - \Delta T$	0.001	0.004	0.008	0.081*	0.080*	0.004	0.008
$\Delta Z_b - k_e$	0.001	0.005	0.013	0.135*	0.205*	0.004	0.009
$Lc - ke$	0.289*	0.284*	0.218*	0.177*	0.058*	0.006	0.116

the other hand, in the outer estuary, sensitivity to L_c increases significantly. In the deepest regions of the estuary (control point P1 and P2), sensitivity to L_c and k_e is similar. Although not as influential as k_e and L_c, solar radiation and water temperature are also relevant sources of uncertainty in the whole estuary. The magnitude of the variability introduced by the offsets in the bathymetry depends on the region of the estuary. The sensitivity to ΔZ_b is much higher in the shallow regions of the estuary, while in deep regions the sensitivity to ΔZ_b is almost zero. The uncertainty introduced by the water salinity, the eddy viscosity and the wave field is not significant in any region of the estuary. Similar conclusion can be derived from the first order variance based sensitivity indices obtained from the ANOVA decomposition (Fig. 6).

Sensitivity measures based on the NSE are more adequate for model calibration purposes, since it is necessary to have a perfect fit between model output and the reference time series to have NSE = 1, while different output timeseries with NSE<1 can fit the average concentration C_m of the reference simulation. The variance based sensitivity indices obtained with the ANOVA-HDMR decomposition of the NSE represent the contribution of each parameter to the total variance of NSE. The main effect of a specific parameter is given by its first order sensitivity index S_i, while its total effect on model output ($S_{T,i}$) is given by all the first and second order terms of the ANOVA-HDMR decomposition in which the parameter appears. Table 10 shows that the measurements based on the main effects of each parameter are not sufficient to adequately represent the variance on NSE, since the sum of these indices for all the parameters is much lower than one. Also, the difference between the main and total effects

Table 12
Total effects for NSE. The two most relevant parameters at each control point according to this measure are marked with.*

	P1	P2	P3	P4	P5	P6	P7
ΔZ_b	0.008	0.025	0.044	0.439	0.571*	0.293*	0.23
L_c	0.667*	0.645*	0.463*	0.475*	0.0122	0.006	0.192
I_{s0}	0.075	0.078	0.091	0.088	0.126	0.207	0.189
S	0.009	0.008	0.006	0.009	0.015	0.011	0.023
C_s	0.007	0.005	0.009	0.002	0.002	0.000	0.002
k_e	0.509*	0.522*	0.577*	0.455*	0.605*	0.699*	0.804*
H_{s0}	0.001	0.001	0.000	0.002	0.000	0.000	0.002

Table 13
Total effects for NSE. Combinations of 2 parameters which explain more than 80% of model output variance at least at one control point. The most relevant combination of 2 parameters at each control point is marked with.*

	P1	P2	P3	P4	P5	P6	P7
$L_c + k_e$	0.886*	0.883*	0.821*	0.753	0.669	0.700	0.879
$k_e + \Delta Z_b$	0.516	0.542	0.607	0.759	0.970*	0.987*	0.818
$k_e + \Delta T$	0.585	0.589	0.619	0.586	0.698	0.726	0.874
$k_e + I_{s0}$	0.554	0.569	0.618	0.518	0.659	0.721	0.847
$L_c + \Delta Z_b$	0.669	0.662	0.501	0.805*	0.666	0.299	0.212

reveals non-linear interactions between input parameters (Tables 11, 12, 13 and 14).

The non-linear effects in NSE are related mainly with L_c and k_e, but also with ΔZ_b at some control points (Figs. 7 and 8). The sensitivity to L_c at control point P5, located in a shallow region of the estuary, is low because the depth of the water layer over which the *E. coli* spread is strongly limited by the bed topography rather than by L_c. For the same reason, sensitivity to offsets in the bathymetry, as well as non-linearities related to this parameter, are much larger at this control point. On the other hand, sensitivity to k_e is important in the whole estuary. Notice also in Figs. 7 and 8 how sensitivity to k_e and L_c decreases as the value of these parameters increases, to the point that sensitivity to L_c at P5 is almost zero for values of L_c higher than the water depth at high tide. These non-linear effects are also reflected in the comparison of the sensitivity indices based on linear and non-linear regression (Table 10).

Table 11 shows the most relevant second order terms of the ANOVA-HDMR decomposition for the NSE, which represent the interaction between the input parameters. The strongest interaction appears between the parameters L_c and k_e, although at some control points there is a remarkable interaction between ΔZ_b and the parameters k_e, L_c and T. This is related with the fact that the depth-averaged degradation coefficient K_d, which depends on k_e, L_c and ΔT, diminishes as the water depth increases.

Since the second order interactions between parameters are very relevant, the main effect is not the most adequate sensitivity measure to establish the influence of each parameter on model output. Alternatively, the total effect of each parameter (Table 12) tells us the contribution that this parameter alone has, on average, on output variance. According to this measure, k_e and L_c are the most relevant parameters for model calibration. As pointed out in the interpretation of the SRC for C_m, the parameter L_c losses significance in the control points located in shallow areas of the estuary. The total effect of S, C_s and H_{s0} is always lower than 0.02, which again identifies these parameters as the less relevant of the model.

If we consider the output variance which is explained by combinations of two and three input parameters (Tables 13 and 14), we find that the combination of L_c and k_e can explain at several control points more than 80% of NSE variance. At control points P5 and P6, located in the shallowest region of the estuary, the combination of k_e and ΔZ_b explains more than 97% of NSE variance. These results indicate that globally, the most efficient parameters

Table 14
Total effects for NSE. Combinations of 3 parameters which explain more than 90% of model output variance at least at one control point. The most relevant combination of 3 parameters at each control point is marked with.*

	P1	P2	P3	P4	P5	P6	P7
$L_c + k_e + \Delta Z_b$	0.89	0.90	0.85*	0.95*	1.00*	0.99	0.89
$L_c + k_e + \Delta T$	0.90*	0.90*	0.85	0.84	0.76	0.73	0.94*
$L_c + ke + I_{s0}$	0.90	0.89	0.84	0.79	0.72	0.72	0.91
$k_e + \Delta Z_b + I_{s0}$	0.56	0.59	0.65	0.79	0.98	1.00	0.86
$K_e + \Delta Z_b + \Delta T$	0.59	0.60	0.64	0.81	0.98	1.00*	0.88

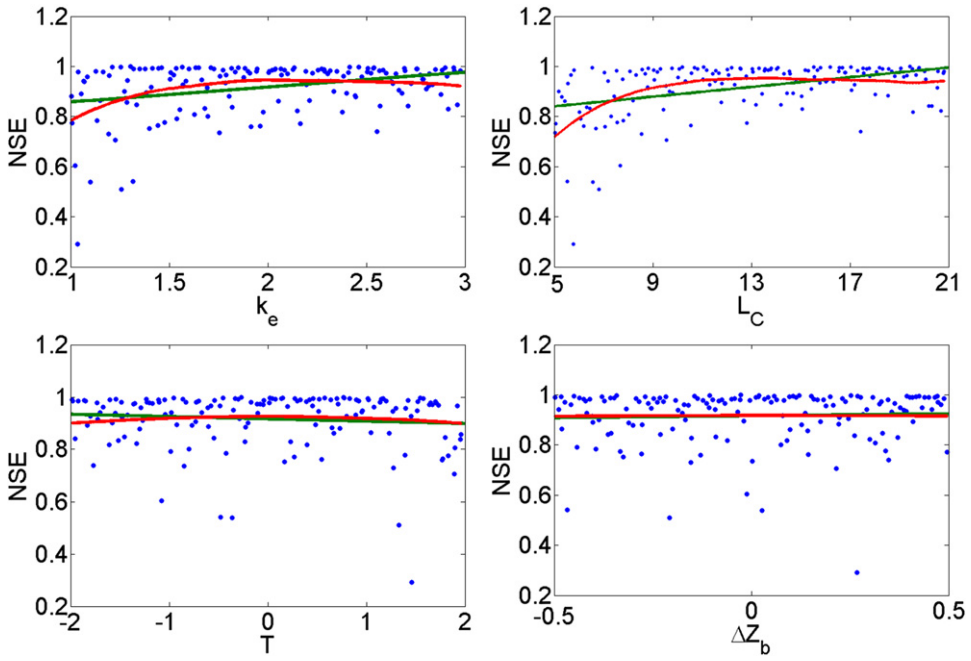

Fig. 7. Linear and non-linear regressions on *NSE* at control point P1.

for model calibration purposes are L_c, k_e and ΔZ_b. In fact, the combination of these 3 parameters explains more than 85% of NSE variance at all the control points, as shown in Table 14. It is also interesting to notice that although the water temperature and the solar radiation have not been pointed out as the most influential parameters, the combination L_c–k_e–ΔT and L_c–k_e–I_{s0} are very effective regarding model calibration, since they are able to explain at least 85% of NSE variance at 4 control points. At the control points P5 and P6 it is necessary to include ΔZ_b as a calibration parameter, otherwise a significant amount of variance is not represented in the model. At these points any combination not including both ΔZ_b and k_e is not efficient for calibration.

For calibration purposes, the optimum value of the parameters k_e, L_c and ΔZ_b can be clearly identified in the non-linear regression of model output shown in Figs. 7 and 8. As it could be expected, the optimum values match those of the reference simulation. It is interesting to notice that at control point P5 it is not possible to obtain a good fit (NSE = 1) with offsets in the bathymetry of the order of 0.5 m, which stresses the relevance of this parameter in the shallow regions of the estuary. This result should be remarked,

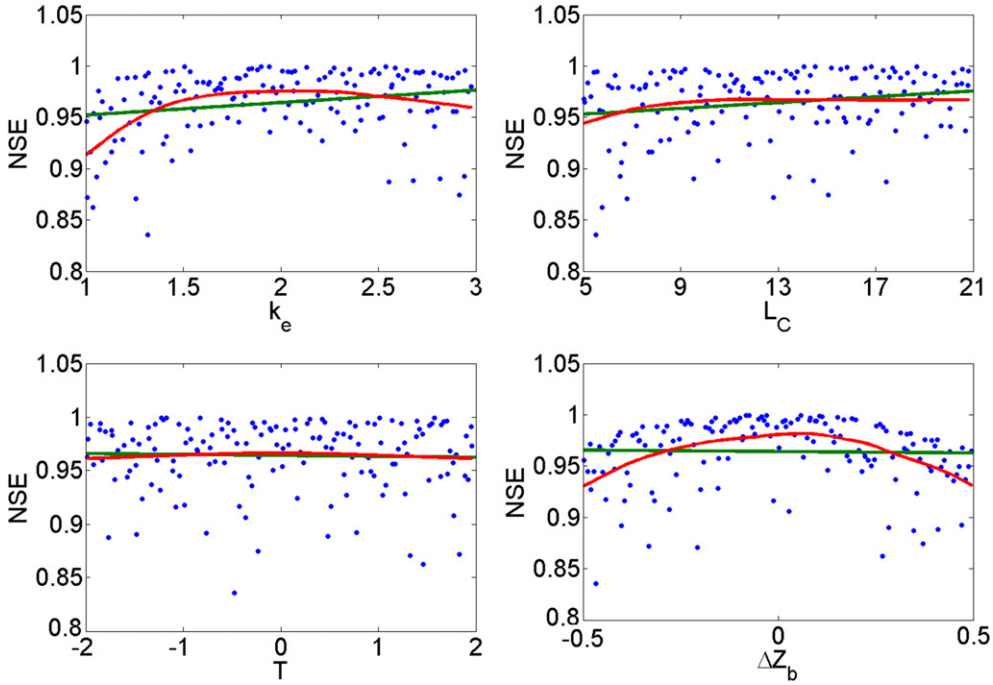

Fig. 8. Linear and non-linear regressions on *NSE* at control point P5.

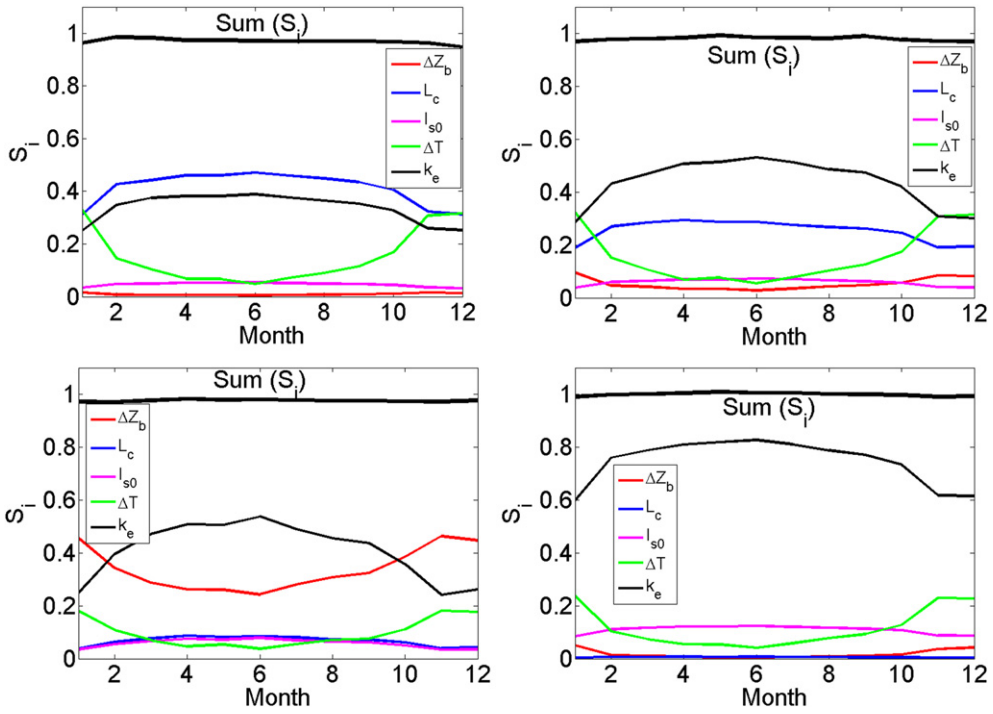

Fig. 9. Seasonal evolution of the most relevant sensitivity indices during the year at control points P1 (upper-left), P3 (upper-right), P5 (lower-left) and P6 (lower-right).

together with the high total sensitivity index obtained for this parameter (Table 12), since bathymetric errors are not usually considered as a source of uncertainty, neither as a model calibration parameter (French, 2010). Identification of the optimum value of all the other parameters is much more uncertain since their sensitivity is much lower. This is clearly shown in the temperature plot in Figs. 7 and 8, and is reflected by the fact that its first order sensitivity index is very low, as shown in Table 10.

Seasonal variation of the sensitivity indices of the most relevant parameters is shown in Fig. 9. It is interesting to notice that the relative importance of the parameters k_e, L_c and I_{s0} decreases considerably during the winter season and increases in the summer. This is because these three parameters determine the contribution of solar radiation to the value of the degradation coefficient K_d. During the summer solar radiation in high and therefore the concentration of *E. coli* in the estuary is more sensitive to these parameters.

To analyse the evolution of uncertainty during a spring-neap tidal cycle, the time evolution of the uncertainty on model output ε_c defined as the average difference between the 150 Monte Carlo runs and the reference simulation computed with the average value of the input parameters is plot in Fig. 10, where ε_c is computed at each control point as:

$$\varepsilon_c(t) = \frac{1}{150}\sum_{j=1}^{150}\frac{|C_j(t) - \hat{C}(t)|}{C_m} \tag{12}$$

where $C_j(t)$ is the concentration of *E. coli* at time t computed from the Monte Carlo run j, $\hat{C}(t)$ is the concentration at time t computed from the reference simulation (using the average value of the input parameters), and C_m is the annual average concentration of *E. coli* (Table 1). At control point P1 there is a higher uncertainty during neap tides. In that region of the estuary, peaks in uncertainty occur at low tide, since P1 is located seaward from spill V1. It is also interesting to see the coupling between uncertainty, solar radiation and tidal level. Peaks in uncertainty are higher when the low tide occurs at dusk (line BB in Fig. 10) and lower when the low tide occurs at dawn (line AA in Fig. 10). This is because the uncertainty in

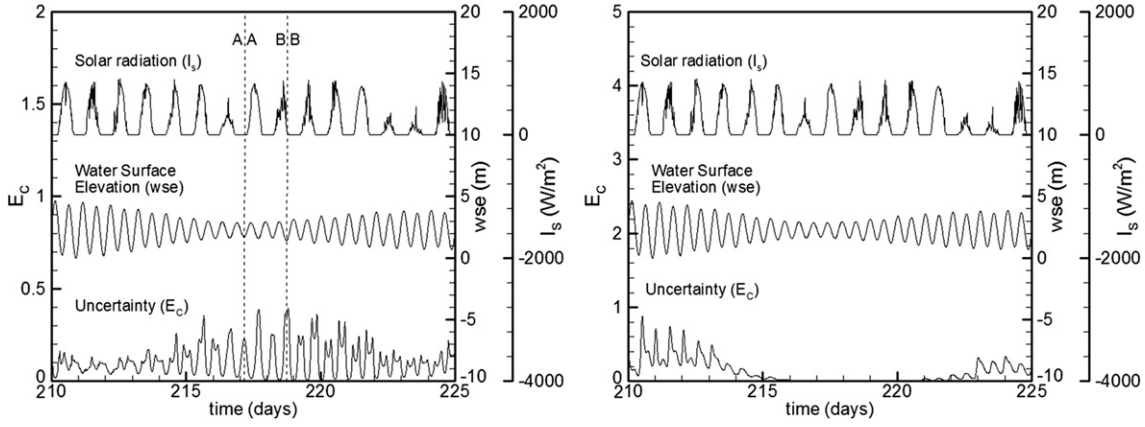

Fig. 10. Uncertainty evolution during a 15 days spring-neap tidal cycle at control points P1 (left) and P4 (right).

the degradation coefficient of *E. coli* is higher during the day time, since at night the solar radiation is zero and therefore, the parameters k_e and L_c do not introduce any uncertainty in the model. A similar behaviour occurs at control points P3 and P6, except that in these cases peaks in uncertainty occur at high tide, because they are located inward from spills V1 and V2 respectively. At control points P4 and P5 the behaviour in the short term is completely different, the uncertainty being higher during spring tides, because during neap tides the spills do not reach these regions of the estuary.

The input parameters considered in this study can be grouped in those which determine the *E. coli* degradation rate (I_{s0},k_e,L_c,S,ΔT and ΔZ_b in shallow regions), and those which influence *E. coli* dispersion (C_s,H_s,ΔZ_b). According to the previous analysis, the former ones are the most influential on model output. Taking into account all the sensitivity measures analysed, the absorption coefficient of light in water k_e is clearly the most relevant parameter of the model. The parameter L_c is one of the most important sources of uncertainty in deep regions of the outer and middle estuary, while its relative importance diminishes in the inner part of the estuary and in shallower regions. On the other hand, the uncertainty on model output introduced by the solar radiation and water temperature is similar throughout the whole estuary. Offsets in the bathymetry have a minor importance except at control points P4, P5 and P6, where it is the most influential parameter. Regarding *E. coli* dispersion, the influence of short waves is more significant than turbulence. Nevertheless, our results indicate that dispersion is much less influential on model output than degradation and therefore, a higher effort should be placed in the calibration of K_d.

The approach used in some models which consider a constant value of K_d might be appropriate if it is used to compute the evolution of *E. coli* concentration over a few hours. Otherwise, the daily and seasonal variations of K_d are so relevant on model output, that a constant value of K_d (even if calibrated with experimental data) will not be able to capture properly the *E. coli* concentration. This effect has also been reported in (Kashefipour et al. (2006), where a large improvement on model results was found by just using a different value of K_d during the day and night. If long-term conditions need to be computed, in addition to daily variations, seasonal variations of K_d need to be included in the model, since the relevance of the input parameters which determine its value varies considerably over the year, as shown in Fig. 9.

5. Conclusions

A 2D water quality model linked to a depth-averaged hydrodynamic model for evaluation of *E. coli* concentration in shallow estuaries has been presented. The model includes the effects of turbulent diffusion, wave dispersion, solar radiation, water temperature and salinity in the transport of *E. coli*. Model output uncertainty and sensitivity to the input parameters and data have been analysed in a shallow estuary using variance based measures computed from a non-linear non-parametric model representation. The extinction coefficient of light in water was found to be the most relevant parameter of the model in the whole estuary and therefore, special care should be placed in its determination and calibration to reduce the uncertainty on model output. The second most relevant parameter is the depth of the vertical layer over which the *E. coli* spread. The relative importance of this parameter is lower in the shallower regions of the estuary. In deep regions its value should be carefully calibrated. Errors in the bathymetry are an important source of uncertainty on model output in the shallowest regions of the estuary. Even if bathymetric errors are common in estuarine models, uncertainty on the bathymetry is rarely considered in numerical models of estuaries, although the sensitivity analysis shows that it might be an efficient calibration parameter in certain estuaries. Time series of water temperature and solar radiation should also be carefully imposed on the model from accurate field measurements, since these are also relevant sources of uncertainty. These results indicate that in long-term computations it is of major importance to consider daily and seasonal variations of K_d in the calibration of the model.

The combination of the parameters k_e, L_c and ΔZ_b is very effective for calibration purposes in the whole estuary. The combination $L_c-k_e-\Delta T$ and $L_c-k_e-I_{s0}$ are also very efficient for model calibration. The uncertainty introduced by the annual variability of wave height and water salinity, as well as by turbulent diffusion, is low and therefore, these parameters might be ignored for model calibration, since altogether they just contribute to approximately 2% of NSE variance.

Acknowledgements

The authors would like to acknowledge Puertos del Estado (Spanish Ministry of Public Works), for providing the wave data from the buoys of A Coruña and Langosteira, and the Spanish Ministry of Education (FPU grant reference AP2009-2070).

References

Ali, A., Zhang, H., Lemckert, C., 2009. Numerical study of the hydrodynamics of a very shallow estuarine system: Coombabah Lake, Gold Coast Australia. Journal of Coastal Research Special Issue 56, 922–926.

Bode, A., González, N., Rodríguez, C., Varela, M., Varela, M.M., 2005. Seasonal variability of plankton blooms in the Ria de Ferrol (NW Spain): I. Nutrient concentrations and nitrogen uptake rates. Estuarine, Coastal and Shelf Science 63, 269–284.

Canteras, J.C., Juanes, J.J., Pérez, L., Koev, K., 1995. Modelling the coliforms inactivation rates in the cantabrian sea (Gulf of Biscay) from "in situ" and laboratory determinations of T90. Water Science and Technology 32, 37–44.

Casas, B., Varela, M., Canle, M., González, N., Bode, A., 1997. Seasonal variations of nutrients, seston and phytoplankton, and upwelling intensity of La Coruña (NW Spain). Estuarine. Coastal and Shelf Science 44, 767–778.

Cea, L., French, J., Vázquez-Cendón, M.E., 2006. Numerical modelling of tidal flows in complex estuaries including turbulence: an unstructured finite volume solver and experimental validation. International Journal for Numerical Methods in Engineering 67, 1909–1932.

Cea, L., Puertas, J., Vázquez-Cendón, M.E., 2007. Depth averaged modelling of turbulent shallow water flow with wet-dry fronts. Archives of Computational Methods in Engineering (ARCME) 14.

Cea, L., Vázquez-Cendón, M.E., 2010. Unstructured finite volume discretisation of two-dimensional depth averaged shallow water equations with porosity. International Journal for Numerical Methods in Fluids 63, 903–930.

Cox, B.A., 2003. A review of currently available in-stream water-quality models and their applicability for simulating dissolved oxygen in lowland rivers. The Science of the Total Environment 314-316, 335–377.

Emery, W.J., Thomson, R.E., 2001. Data Analysis Methods in Physical Oceanography. Elsevier.

Estrada, V., Diaz, M.S., 2010. Global sensitivity analysis in the development of first principle-based eutrophication models. Environmental Modelling & Software 25, 1539–1551.

Fischer, H.B., List, E.J., Koh, R.C.Y., Imberger, J., Brooks, J., 1979. Mixing in inland and coastal waters. Academic Press, San Diego.

French, J.R., 2008. Hydrodynamic modelling of estuarine flood defence realignment as an adaptive management response to sea-level rise. Journal of Coastal Research 24, 1–12.

French, J.R., 2010. Critical perspectives on the evaluation and optimisation of complex numerical models of estuary hydrodynamics and sediment dynamics. Earth Surface Processes and Landforms 35, 174–189.

Garcia-Barcina, J.M., Oteiza, M., Sota, A.D.L., 2002. Modelling the faecal coliform concentrations in the Bilbao estuary. Hydrobiologia 475-476, 213–219.

Gonzalez, J., 1995. Modelling enteric bacteria survival in aquatic systems. Hydrobiology 316, 109–116.

Guinot, V., Cappelaere, B., 2009. Sensitivity analysis of 2D steady-state shallow water flow. Application to free surface flow model calibration. Advances in Water Resources 32, 540–560.

Harris, E.L., Falconer, R.A., Lin, B., 2004. Modelling hydroenvironmental and health risk assessment parameters along the south wales coast. Journal of Environmental Management 73, 61–70.

Hellweger, F.L., 2007. Ensemble modeling of *E. Coli* in the Charles River, Boston, Massachusetts, USA. Water Science & Technology 56, 39–46.
Hogdins, D.O., Tinis, S.W., Taylor, L.A., 1998. Marine sewage outfall assessment for the capital regional district, british colombia, using nested three-dimensional models. Water Science Technol 38, 301–308.
Holthuijsen, L.H., 2007. Waves in Oceanic and Coastal Waters. Cambridge University Press, Cambridge, United Kingdom.
Kashefipour, S.M., Lin, B., Falconer, R.A., 2006. Modelling the fate of faecal indicators in a coastal basin. Water Research 40, 1413–1425.
Kashefipour, S.M., Lin, B., Harris, E., Falconer, R., 2002. Hydro-environmental modelling for bathing water compliance of an estuarine basin. Water Research 36, 1854–1868.
Kneis, D., Forster, S., Bronstert, A., 2009. Simulation of water quality in a flood detention area using models of different spatial discretization. Ecological Modelling 220, 1631–1642.
Kucherenko, S., Feil, B., Shah, N., Mauntz, W., 2011. The identification of model effective dimensions using global sensitivity analysis. Reliability Engineering & System Safety 96, 440–449.
Lane, A., 2004. Bathymetric evolution of the mersey estuary, UK, 1906-1997 causes and effects. Estuarine. Coastal and Shelf Science 59, 249–263.
Law, A.W.K., 2000. Taylor dispersion of contaminants due to surface waves. Journal of Hydraulic Research 38, 41–48.
Lees, M.L., Camacho, L.A., Chapra, S.C., 2000. On the relationship of transient storage and aggregated dead zone models of longitudinal transport in streams. Water Resource Research 36, 213–224.
Ludwig, R.G., 1988. Environmental impact assessment: siting and design of submarine outfalls. Monitoring and Assessment Research Centre, London (MARC Report, 43), Kings College London.
Manache, G., Melching, C.S., 2008. Identification of reliable regression- and correlation-based sensitivity measures for importance ranking of water-quality model parameters. Environmental Modelling & Software 23, 549–562.
Manache, G., Melching, C.S., Lanyon, R., 2007. Calibration of a continuous simulation fecal coliform model based on historical data analysis. Journal of Environmental Engineering 133, 681–691.
Mancini, J.L., 1978. Numerical estimates of coliform mortality rates under various conditions. Journal (Water Pollution Control Federation) 50, 2477–2484.
Marsili-Libelli, S., Giusti, E., 2008. Water quality modelling for small river basins. Environmental Modelling & Software 23, 451–463.
Mayer, D.G., Butler, D.G., 1993. Statistical validation. Ecological Modelling 68, 21–32.
McIntyre, N.R., 2004. Analysis of uncertainty in river water quality modelling. PhD Dissertation. Department of Civil and Environmental Engineering, Imperial College London.
Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models. Part I - A discussion of principles. Journal of Hydrology 10, 282–290.
Padisak, J., Reynolds, C.S., 2003. Shallow lakes: the absolute, the relative, the functional and the pragmatic. Hydrobiologia 506-509, 1–11.
Pastres, R., Ciavatta, S., 2005. A comparison between the uncertainties in model parameters and in forcing functions: its application to a 3D water-quality model. Environmental Modelling & Software 20, 981–989.
der Perk, M.V., 1997. Effect of model structure on the accuracy and uncertainty of results from water quality models. Hydrological Processes 11, 227–239.
Pommepuy, M., Hervio, D., Caprais, M.P., Gourmelon, M., Saux, J.C.L., Guyader, S.L., 2006. Fecal contamination in coastal areas: an engineering approach. chapter oceans and health: pathogens in the marine environment (Book chapter). Open access version: http://archimer.ifremer.fr/doc/00000/1207/ edition331–359.
Ratto, M., Pagano, A., 2010. Using recursive algorithms for the efficient identification of smoothing spline ANOVA models. Advances in Statistical Analysis 94, 367–388.
Ratto, M., Pagano, A., Young, P., 2007. State dependent parameter metamodelling and sensitivity analysis. Computer Physics Communications 177, 863–876.
Saltelli, A., Annonis, P., 2010. How to avoid a perfunctory sensitivity analysis. Environmental Modelling & Software 25, 1508–1517.
Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., 2008. Global Sensitivity Analysis. The Primer. John Wiley & Sons.
Sobol, I.M., 1976. Uniformly distributed sequences with additional uniformity properties. USSR Comput. Math. Math. Phys. 16, 236–242.
Sobol, I.M., 1998. On quasi-monte carlo integrations. Mathematics and Computers in Simulation 47, 103–112.
Storlie, C.B., Helton, J.C., 2008. Multiple predictor smoothing methods for sensitivity analysis: description of techniques. Reliability Engineering and System Safety 93, 28–54.
Storlie, C.B., Swiler, L.P., Helton, J.C., Sallaberry, C.J., 2009. Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models. Reliability Engineering and System Safety 94, 1735–1763.
Toro, E.F., 2001. Shock-capturing Methods for Free-Surface Shallow Flows. Wiley, Chichester, West Sussex PO19 1UD, England.
Vandenberghe, V., Bauwens, W., Vanrolleghem, P.A., 2007. Evaluation of uncertainty propagation into river water quality predictions to guide future monitoring campaigns. Environmental Modelling & Software 22, 725–732.
Warmink, J.J., Janssen, J.A.E.B., Booij, M.J., Krol, M.S., 2010. Identification and classification of uncertainties in the application of environmental models. Environmental Modelling & Software 25, 1518–1527.