Journal of Hydraulic Research Vol. 0, No. 0 (2006), pp. 1–16

On inducing equations for vegetation resistance

M.J. BAPTIST, *Delft University of Technology, Faculty of Civil Engineering and Geosciences, Water Resources Section, Stevinweg 1, 2628 CN Delft, The Netherlands and WL | Delft Hydraulics, Rotterdamseweg 185, 2629 HD Delft, The Netherlands*

V. BABOVIC, *National University of Singapore, 1 Engineering Drive 2, Singapore 117 576 and WL | Delft Hydraulics, Rotterdamseweg 185, 2629 HD Delft, The Netherlands*

J. RODRÍGUEZ UTHURBURU, *UNESCO-IHE, Westvest 7, 2611 AX, Delft, The Netherlands*

M. KEIJZER, *WL | Delft Hydraulics, Rotterdamseweg 185, 2629 HD Delft, The Netherlands*

R.E. UITTENBOGAARD, *WL | Delft Hydraulics, Rotterdamseweg 185, 2629 HD Delft, The Netherlands*

A. MYNETT, *WL | Delft Hydraulics, Rotterdamseweg 185, 2629 HD Delft, The Netherlands and UNESCO-IHE, Westvest 7, 2611 AX, Delft, The Netherlands*

A. VERWEY, *WL | Delft Hydraulics, Rotterdamseweg 185, 2629 HD Delft, The Netherlands and UNESCO-IHE, Westvest 7, 2611 AX, Delft, The Netherlands*

ABSTRACT

The paper describes the process of induction of equations for the description of vegetation-induced roughness from several angles. Firstly, it describes two approaches for obtaining theoretically well-founded analytical expressions for vegetation resistance. The first of the two is based on simplified assumptions for the vertical flow profile through and over vegetation, whereas the second is based on an analytical solution to the momentum balance for flow through and over vegetation. In addition to analytical expressions the paper also outlines a numerical 1-DV k–ε turbulence model which includes several important features related to the influence plants exhibit on the flow. Last but not least, the paper presents a novel way of applying genetic programming to the results of the 1-DV model, in order to obtain an expression for roughness based on synthetic data. The resulting expressions are evaluated and compared with an independent data set of flume experiments.

RÉSUMÉ

Keywords: Vegetation, roughness, resistance, genetic programming, knowledge discovery.

1 Introduction

Proper modelling of the flow resistance of wetlands and vegetated floodplains is of great practical importance. Many research initiatives have been undertaken in order to improve on the description of the relationship between flow resistance and the presence and spatial distribution of vegetation. Both analytical and experimental studies of vegetation-related resistance to flow and the equivalent resistance coefficients have shown that the resistance coefficients are water depth dependent. Consequently, the traditional approach of using a single resistance coefficient fails to correctly describe the physics of the phenomenon. One way

Revision received xxx. Open for discussion till xxx.

of improving upon this description is updating the equivalent resistance coefficient based on the computed water depth. In order to do so, a relation between vegetation characteristics, bed resistance, water depth and equivalent resistance coefficient is needed.

In this paper, four different methods to obtain such a relation are elaborated. The first is based on a simple analytical description of the division in discharge through and over submerged vegetation. The second is based on an analytical solution to the momentum balance of flow through and over vegetation. The third is based on a numerical 1-DV model and finally, the fourth is based on data mining.

Data mining is concerned with extracting useful information from data. However, mining the data *alone* is not the entire story. Scientific theories encourage the acquisition of new data and this data in turn leads to the generation of new theories. Traditionally, the emphasis is on a theory, which demands that appropriate data be obtained through observation or experiment. In such an approach, the discovery process is what we may refer to as *theory-driven*. Especially when a theory is expressed in mathematical form, *theory-driven* discovery may make extensive use of strong methods associated with mathematics or with the subject matter of the theory itself. The converse view takes a body of data as its starting point and searches for a set of generalizations, or a theory, to describe the data parsimoniously or even to explain it. Usually such a theory takes the form of a precise mathematical statement of the relations existing among the data. This is the *data-driven* discovery process.

We believe that the most appropriate way for scientific applications of data mining is to combine the better of the two approaches: *theory-driven*, understanding-rich with *data-driven* discovery process. Clearly, there is an enormous amount of knowledge and understanding of physical processes that should be incorporated in the discovery process. The work described here is part of a research effort aiming at providing new hypotheses built from data. The ultimate objective is to build models which can be interpreted and further manipulated by the domain experts. Once a model is interpreted, it can be used with confidence. It is only in this way that one can take full advantage of knowledge discovery and advance our understanding of physical processes, of, in this case, vegetation resistance.

2 Background

The effect of vegetation on flow is generally expressed as an effect on the hydraulic roughness. Early measurements (18th century) of flow velocities in channels revealed that the depth-averaged flow velocity (m/s) was a function of the water level slope i (m/m) and the hydraulic radius R (m). In 1776 Antoine de Chézy published a simple equation that includes an engineering factor C, the Chézy value, which was at first thought to be a constant (Vernon-Harcourt, 1896). The well-known Chézy formula is:

$$\bar{u} = C\sqrt{Ri} \tag{1}$$

In this equation C is a parameter that expresses the hydraulic roughness of the bed and banks of a channel ($\mathrm{m}^{1/2}/\mathrm{s}$). Further investigations, by Nikuradse (1930), revealed that the roughness of the bed affects the roughness length z_0 (m) in the logarithmic velocity profile for a fully rough bed derived by, among others, Prandtl and Von Kármán:

$$u(z) = \frac{u_*}{\kappa} \ln\left(\frac{z}{z_0}\right) \tag{2}$$

Nikuradse showed that for hydraulically rough walls, the roughness length of the logarithmic velocity profile can be expressed as $k_\mathrm{N}/30$, where k_N is the Nikuradse equivalent roughness (Nikuradse, 1930). Calculating the depth-averaged velocity, and applying the Chézy formula and the Nikuradse roughness height yields the White–Colebrook formula for the Chézy value:

$$C = 18\log\left(\frac{12R}{k_\mathrm{N}}\right) \tag{3}$$

With an increasing roughness height the value for C decreases. Various alternative expressions for flow resistance exist, for example those of Strickler or Manning, which can all be mutually converted. Essentially, in using a single roughness equation such as the White–Colebrook formula, vegetation is treated as large bed structures with a logarithmic flow profile above them. In reality, however, there is flow over and through submerged vegetation, and the vertical flow profile deviates from the logarithmic one. A typical time-averaged velocity profile for submerged vegetation is shown in Fig. 1. Four distinct zones can be identified in the time-averaged vertical velocity profile for flow through and above submerged vegetation:

1. In the first zone, near the bed, the velocity is highly influenced by the bed, and its vertical profile joins the logarithmic boundary layer profile.
2. In the second zone, which corresponds to the zone inside the vegetation sufficiently away from the bed and from the top of the vegetation, the velocity is uniform.
3. In the third zone, near the top of the vegetation, there is a transitional profile between the uniform velocity inside the

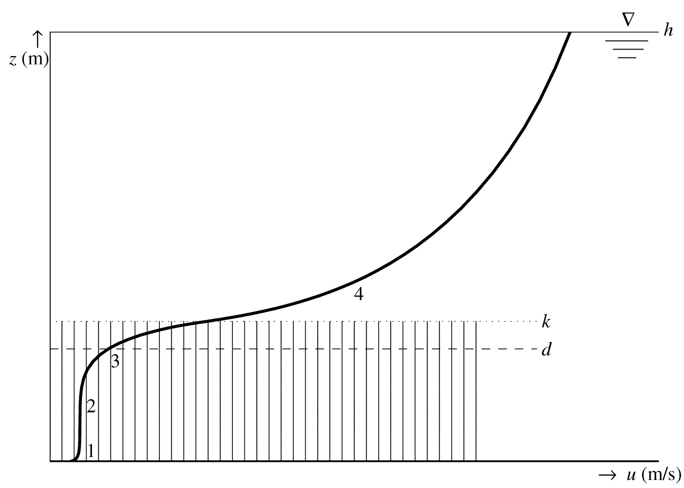

Figure 1 Four zones in the vertical profile for horizontal velocity, $u(z)$, through and over vegetation, h = water depth (m), k = vegetation height (m), d = zero-plane displacement (m).

vegetation and the logarithmic profile above it. The profile in this zone can be approximated by an exponential function.

4. Finally, the fourth zone corresponds to the zone above the vegetation, where a logarithmic profile is observed, which has a zero-plane displacement below the top of the vegetation layer.

The White–Colebrook formula fails here and another type of resistance formula should be sought for. A considerable amount of research has been carried out on the effects of vegetation on the hydraulic resistance, extending the basic ideas of Nikuradse (1930). Early work includes Einstein and Banks (1950), Kouwen *et al*. (1969), Kouwen and Unny (1973), Klaassen and Van der Zwaard (1974) and Petryck and Bosmajian (1975). In a study by Dawson and Charlton (1988), a literature search has been carried out on the hydraulic resistance of vegetation, resulting in over 360 publications. Since then, many more publications have followed. A limited overview of recent research includes studies on the improvement of flow resistance formulae (Darby, 1999; Hasegawa *et al*., 1999; Meijer and Van Velzen, 1999; Stephan and Gutknecht, 2002; Järvelä, 2002, 2004; Mason *et al*., 2003; James *et al*., 2004), on analytical approaches for the vertical profile of horizontal velocity (Klopstra *et al*., 1997; Carollo *et al*., 2002; Katul *et al*., 2002), on biomechanics and streamlining of vegetation (Fathi-Maghadam and Kouwen, 1997) and on turbulence characterisation for submerged rods and vegetation (Shimizu and Tsujimoto, 1994; Ikeda and Kanazawa, 1996; Nezu and Naot, 1999; Nepf and Vivoni, 2000; Ikeda *et al*., 2001; Fisher-Antze *et al*., 2001; López and García, 2001; Ghisalberti and Nepf, 2002; Righetti and Armanini, 2002; Wilson *et al*., 2003; Ghisalberti and Nepf, 2004). However, no acceptable formulation for roughness induced by submerged vegetation, valid for a wide range of vegetation properties and water depths has been found as of yet. This is the main inspiration for the present work.

3 Theoretical, analytical formulations of resistance due to vegetation

In this section, approaches for obtaining vegetation-related resistance using theory-based formulae are described. The formulae were derived from basic physical concepts on flow through and above vegetation.

Modelling flow through a porous medium, such as vegetation, in principle involves a correction for the presence of vegetation within the volume of water. A common way to deal with this is to introduce the dimensionless parameter A_p, the solidity, which is defined as the fraction of horizontal area taken by the cylinders (Li and Shen, 1973; Taylor *et al*., 1985; Stone and Shen, 2002; Hoffmann, 2004):

$$A_p = \frac{1}{4}\pi D^2 m \tag{4}$$

The solidity can be introduced to calculate the pore, or microscopic velocity in between the vegetation, which determines the resistance force of the vegetation. In addition, the solidity can be used to correct for the available volume, or available horizontal area in the calculation of the fluid shear stress or the bed shear stress, respectively. However, various authors report different theoretical approaches to determine the pore velocity, the drag coefficient associated with this pore velocity, or the correction for available volume or area. None of the approaches are underpinned in a satisfactory manner with experimental evidence. More importantly, experimental evidence has shown that this correction term can be neglected to calculate vegetation resistance in natural circumstances with no significant loss of accuracy (James *et al*., 2004). Consequently it can be concluded that the solidity can be disregarded in simple analytical expressions for flow through and over vegetation, especially in the light of the uncertainties introduced by describing vegetation properties in terms of stem density, height and diameter.

3.1 Case of non-submerged vegetation

Non-submerged flow conditions can be successfully treated analytically. The balance of horizontal momentum in uniform steady flow conditions through non-submerged vegetation expressed as cylinders dictates that total shear stress is equal to the sum of the bed shear stress and the equivalent shear stress due to vegetation drag:

$$\tau_t = \tau_b + \tau_v \tag{5}$$

where τ_t denotes the total fluid shear stress:

$$\tau_t = \rho_0 g h i \tag{6}$$

τ_b denotes the bed shear stress:

$$\tau_b = \rho_0 g \frac{u^2}{C_b^2} \tag{7}$$

and τ_v is the vegetation resistance force per unit horizontal area:

$$\tau_v = \frac{1}{2}\rho_0 C_D m D h u^2 \tag{8}$$

The vegetation resistance force is thus modelled as the drag force on a random or staggered array of rigid cylinders with uniform properties. The uniform flow velocity through non-submerged vegetation follows from the momentum balance and is given by:

$$u_{cb} = \sqrt{\frac{hi}{1/C_b^2 + (C_D m D h)/(2g)}} \tag{9}$$

The discharge per unit width through the vegetation is given by:

$$q = h u_{cb} \tag{10}$$

From the discharge through the vegetation the representative Chézy value for non-submerged vegetation is calculated as:

$$C = \frac{q}{h\sqrt{hi}} \tag{11}$$

Therefore, the representative Chézy value for non-submerged vegetation becomes:

$$C_k = \sqrt{\frac{1}{1/C_b^2 + (C_D m D h)/(2g)}} \tag{12}$$

When the bed resistance is negligible with respect to the drag force of the vegetation, the resistance coefficient can be reduced to:

$$C_k = \sqrt{\frac{2g}{C_{\mathrm{D}} m D h}} \tag{13}$$

3.2 *Case of submerged vegetation*

Two different methods to derive theoretical resistance formulae for submerged vegetation were applied by Baptist (2005), i.e. the "method of effective water depth" and a method utilizing an analytical derivation of the momentum balance for flow through vegetation.

3.2.1 *Submerged vegetation — method of effective water depth*

The "method of effective water depth" is inspired by earlier approaches by Hong (1995), Campana (1999) and Stone and Shen (2002). In this method, the four zones in the velocity profile of flow through and above vegetation are represented in a simplified way and reduced to two flow zones:

1. A uniform flow velocity, u_{c}, inside the vegetation, and
2. A logarithmic flow profile, u_{u}, above the vegetation, with zero-plane displacement at height k and with an additional slip velocity of size u_{c}.

This simplified velocity profile is presented in Fig. 2. The flow velocity in the vegetated section follows from the momentum balance for flow through vegetation. The flow velocity in the uniform part of the profile is, therefore, given by the formula for flow through fully submerged vegetation, which resembles Eq. (9):

$$u_{\mathrm{c}} = \sqrt{\frac{hi}{1/C_{\mathrm{b}}^2 + (C_{\mathrm{D}} m D k)/(2g)}} \tag{14}$$

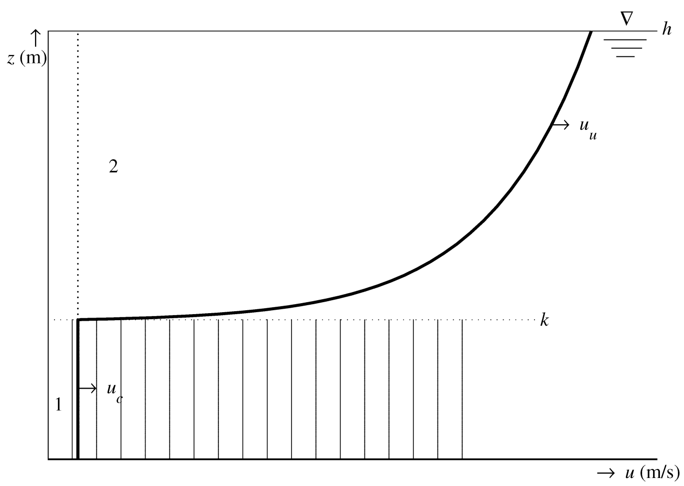

Figure 2 Representation of the vertical velocity profile in two zones for the method of effective water depth, h = water depth (m), k = vegetation height (m), u_{c} = uniform flow velocity profile (m/s), u_{u} = logarithmic flow velocity profile (m/s).

The logarithmic velocity profile above the vegetation (u_{u}) is given by:

$$u_{\mathrm{u}}(z) = \frac{u_*}{\kappa} \ln\left(\frac{z-k}{z_0}\right) + u_{\mathrm{c}} \tag{15}$$

Note that the uniform flow velocity through the vegetation, u_{c}, is added to the logarithmic velocity as a constant "slip velocity" to match both profiles at $z = k$.

The height-averaged velocity above the vegetation then becomes:

$$\begin{aligned}\bar{u}_{\mathrm{u}} &= \frac{1}{h-k}\int_k^h u_{\mathrm{u}}(z)\mathrm{d}z \\ &= \frac{u_*}{\kappa}\ln\left(\frac{h-k}{z_0} - 1\right) + u_{\mathrm{c}} = \frac{u_*}{\kappa}\ln\left(\frac{h-k}{ez_0}\right) + u_{\mathrm{c}}\end{aligned} \tag{16}$$

where e is the base of the natural logarithm.

In the method of effective water depth, the discharge per unit width through and over the vegetation is distributed by their respective water depths. The addition of the slip velocity to the logarithmic flow profile results in a distinction in two flow parts (see Fig. 2). Part 1 contains the uniform flow velocity over the entire water depth h and Part 2 contains a logarithmic flow velocity over the height h–k:

$$q = ku_{\mathrm{c}} + (h-k)\bar{u}_{\mathrm{u}} = hu_{\mathrm{c}} + (h-k)\bar{u}_{\mathrm{u}} - (h-k)u_{\mathrm{c}} \tag{17}$$

Therefore:

$$\begin{aligned} q = {} & h\sqrt{\frac{hi}{1/C_{\mathrm{b}}^2 + (C_{\mathrm{D}} m D k)/(2g)}} \\ & + (h-k)\frac{\sqrt{g(h-k)i}}{\kappa}\ln\left(\frac{h-k}{ez_0}\right)\end{aligned} \tag{18}$$

From the discharge relationship, Eq. (11), the representative Chézy value follows as:

$$\begin{aligned} C_{\mathrm{r}} = {} & \sqrt{\frac{1}{1/C_{\mathrm{b}}^2 + (C_{\mathrm{D}} m D k)/(2g)}} \\ & + \frac{(h-k)^{3/2}(\sqrt{g}/\kappa)\ln((h-k)/ez_0)}{h^{3/2}}\end{aligned} \tag{19}$$

Note that the first term on the right-hand side equals the representative roughness of non-submerged vegetation for $h = k$, Eq. (12). Note further that part of the second term on the right-hand side can be approximated by the White–Colebrook formula:

$$\frac{\sqrt{g}}{\kappa}\ln\left(\frac{h-k}{ez_0}\right) \equiv 18\log\left(\frac{12(h-k)}{k_{\mathrm{N}}}\right) \tag{20}$$

where k_{N}, the equivalent Nikuradse roughness height of the top of the vegetation layer (m), is taken equal to $30z_0$ (Nikuradse, 1930). This approach presents one unknown variable: z_0 being the roughness height of the top of the vegetation layer. In Sec. 3.2.2 an analytical expression for z_0 will be derived.

3.2.2 Submerged vegetation — analytical method

In the analytical method the velocity profile inside the vegetation layer is obtained by analytically solving the momentum equation:

$$\frac{\partial \tau_{xz}}{\partial z} - \rho_0 g \frac{\partial h}{\partial x} - \frac{1}{2}\rho_0 C_\mathrm{D} m D u^2(z) = 0 \tag{21}$$

Using Boussinesq's eddy viscosity approach, the Reynolds stress τ_{xz} is:

$$\tau_{xz} = \rho_0 \nu_\mathrm{T} \frac{\partial u}{\partial z} \tag{22}$$

Klopstra *et al.* (1997) derived an analytical method approximating the eddy viscosity, ν_T, by the formula:

$$\nu_\mathrm{T}(z) = \alpha \cdot u(z) \tag{23}$$

where α is a characteristic length scale to be determined. Baptist (2005) derived a similar analytical method, but applying the mixing-length theory for the eddy viscosity:

$$\nu_\mathrm{T}(z) = \ell \sqrt{k_\mathrm{T}} \tag{24}$$

Assuming that the turbulence inside the vegetation is governed by the stem turbulence, the mixing length is equal to the available length scale for eddies between the vegetation L_p:

$$\ell = L_\mathrm{p} = c_\ell \left(\frac{1 - A_\mathrm{p}}{m}\right)^{1/2} \tag{25}$$

where c_ℓ is a coefficient that affects the geometrical length scale and can be used for calibration, but is taken to be 1, and A_p is the solidity, which is included here for completeness. The turbulent kinetic energy, k_T, is determined by the local mean velocity, it then follows from Eq. (24) that the eddy viscosity becomes:

$$\nu_\mathrm{T}(z) = c_\mathrm{p} \ell u(z) \tag{26}$$

where the coefficient c_p is the turbulence intensity, height-averaged over the vegetation height k:

$$c_\mathrm{p} = \frac{(1/k)\int_0^k \sqrt{k(z)}\,\mathrm{d}z}{(1/k)\int_0^k u(z)\,\mathrm{d}z} \tag{27}$$

The mixing length ℓ and the coefficient c_p are assumed constant over the vegetation height in our analytical approach resulting in the following ordinary differential equation:

$$\frac{1}{2} c_\mathrm{p} \ell \frac{\mathrm{d}^2 u^2}{\mathrm{d}z^2} - \frac{1}{2} C_\mathrm{D} m D u^2 = gi \tag{28}$$

This differential equation has an analytical solution:

$$u(z) = \sqrt{u_{s0}^2 + a \exp\left(\frac{z}{L}\right) + b \exp\left(-\frac{z}{L}\right)} \tag{29}$$

with uniform flow over part of the cylinder height u_{s0} (m/s) and length scale L(m):

$$u_{s0} = \sqrt{\frac{2gi}{C_\mathrm{D} m D}} \tag{30}$$

$$L = \sqrt{\frac{c_\mathrm{p} \ell}{C_\mathrm{D} m D}} \tag{31}$$

The integration constants a and b should be determined by imposing boundary conditions. The flow profile as depicted by Eq. (29) describes zones 1–3 of Fig. 1. However, for the purpose of obtaining an expression for the equivalent resistance coefficient, Baptist (2005) simplified Eq. (29) to:

$$u_\mathrm{v}(z) = \sqrt{u_{s0}^2 + a \exp\left(\frac{z}{L}\right)} \tag{32}$$

This simplified profile only describes zones 2 and 3 of Fig. 1, where zone 2 reaches down to the bed. In this way, this expression cannot represent properly the velocity profile near the bed, where u must vanish, but this expression allows for the calculation of the resistance C, since it can be integrated over the depth. Above the vegetation, in zone 4 of Fig. 1, the following logarithmic velocity profile is assumed:

$$u_\mathrm{o}(z) = \frac{u_*}{\kappa} \ln\left(\frac{z - d}{z_0}\right) \tag{33}$$

where d is the zero-plane displacement (m), which is located at a distance from the bed inside the vegetation. Note that we now disregard a slip velocity. The shear velocity is given by:

$$u_* = \sqrt{g(h - k)i} \tag{34}$$

Note further that the definition for a logarithmic velocity profile in a hydraulically rough turbulent boundary layer is strictly followed. In this definition, the level of the shear velocity is determined at that height above which the flow is not affected directly by individual roughness elements (Jackson, 1981). This equals the level $z = k$, the average height of the roughness forming elements.

The expression for the integration constant a follows from the boundary condition that at the top of the vegetation the shear stress of the overlying flow must equal the shear stress of the flow inside the vegetation layer. The shear stress at height k from the profile inside the vegetation layer is given by:

$$\tau_{xz}(k) = \rho_0 c_\mathrm{p} \ell u_\mathrm{v}(k) \frac{\partial u_\mathrm{v}}{\partial z}(k) = \frac{\rho_0 c_\mathrm{p} \ell a \exp(k/L)}{2L} \tag{35}$$

The shear stress at height k from the profile above the vegetation layer is given by:

$$\tau_{xz}(k) = \rho_0 g(h - k)i \tag{36}$$

By equalling Eqs. (35) and (36) a is obtained as:

$$a = \frac{2Lg(h - k)i}{c_\mathrm{p} \ell \exp(k/L)} \tag{37}$$

Since both u_{s0} and a include the water level slope in their formulations, and the water level slope is an unknown variable closely related to the resistance coefficient, Eq. (32) is rewritten as:

$$u_\mathrm{v}(z) = \sqrt{i\left(u_{v0}^2 + a_\mathrm{v} \exp\left(\frac{z}{L}\right)\right)} \tag{38}$$

with:

$$u_{v0} = \sqrt{\frac{2g}{C_\mathrm{D} m D}} \tag{39}$$

and Eq. (37) is rewritten as:

$$a_\mathrm{v} = \frac{2Lg(h - k)}{c_\mathrm{p} \ell \exp(k/L)} \tag{40}$$

Now the equivalent Chézy resistance coefficient C_r can be calculated as:

$$C_\mathrm{r} = \frac{k\,\bar{u}_\mathrm{v} + (h-k)\bar{u}_\mathrm{u}}{h\sqrt{hi}} \tag{41}$$

where $\bar{u}_\mathrm{v}$ is the height-averaged velocity inside the vegetation layer, and $\bar{u}_\mathrm{o}$ is the height-averaged velocity above the vegetation layer. These height-averaged velocities are calculated analytically as:

$$\bar{u}_\mathrm{v} = \frac{1}{k}\int_0^k u_\mathrm{v}(z)\mathrm{d}z$$

$$\bar{u}_\mathrm{v} = \frac{L\sqrt{i}}{k}\left[2\left(u_{\mathrm{v}k} - \sqrt{a_\mathrm{v} + u_{\mathrm{v}0}^2}\right) + u_{\mathrm{v}0}\ln\left(\frac{(u_{\mathrm{v}k} - u_{\mathrm{v}0})\left(\sqrt{a_\mathrm{v}+u_{\mathrm{v}0}^2} + u_{\mathrm{v}0}\right)}{(u_{\mathrm{v}k} + u_{\mathrm{v}0})\left(\sqrt{a_\mathrm{v}+u_{\mathrm{v}0}^2} - u_{\mathrm{v}0}\right)}\right)\right] \tag{42}$$

$$\bar{u}_\mathrm{o} = \frac{1}{h-k}\int_k^h u_\mathrm{o}(z)\mathrm{d}z$$

$$\bar{u}_o = \frac{\sqrt{g(h-k)i}}{\kappa(h-k)}\left[(h-d)\ln\left(\frac{h-d}{z_0}\right) - (k-d)\ln\left(\frac{k-d}{z_0}\right) - (h-k)\right] \tag{43}$$

where $u_{\mathrm{v}k}$ is the flow velocity at the top of the vegetation, height k, following from the flow profile inside the vegetation. The formula for the equivalent Chézy resistance coefficient follows from Eq. (41) and becomes:

$$C_\mathrm{r} = \frac{1}{h^{3/2}}\left\{\begin{aligned} &L\left[2\left(u_{\mathrm{v}k} - \sqrt{a_\mathrm{v}+u_{\mathrm{v}0}^2}\right) + u_{\mathrm{v}0}\ln\left(\frac{(u_{\mathrm{v}k} - u_{\mathrm{v}0})\left(\sqrt{a_\mathrm{v}+u_{\mathrm{v}0}^2} + u_{\mathrm{v}0}\right)}{(u_{\mathrm{v}k} + u_{\mathrm{v}0})\left(\sqrt{a_\mathrm{v}+u_{\mathrm{v}0}^2} - u_{\mathrm{v}0}\right)}\right)\right] \\ &+ \frac{\sqrt{g(h-k)}}{\kappa(h-k)}\left[(h-d)\ln\left(\frac{h-d}{z_0}\right) - (k-d)\ln\left(\frac{k-d}{z_0}\right) - (h-k)\right]\end{aligned}\right. \tag{44}$$

In order to describe the vertical velocity profile given by the combination of Eqs (32) and (33), and therefore, to determine the resistance of the submerged vegetation, three unknown parameters need to be determined, the zero-plane displacement d, the roughness height z_0 and the closure coefficient c_p for the mean turbulence intensity.

The expression for d follows from the definition of Thom (1971):

$$d = \frac{\int_0^k (\mathrm{d}\tau_{xz}(z)/\mathrm{d}z)z\mathrm{d}z}{\int_0^k (\mathrm{d}\tau_{xz}(z)/\mathrm{d}z)\mathrm{d}z} \tag{45}$$

which has been further extended by Jackson (1981) and can be written as (Baptist, 2005):

$$d = k - \int_0^k \frac{\tau_{xz}(z)}{\tau_{xz}(k)}\mathrm{d}z \tag{46}$$

Substitution of Eq. (35) yields:

$$d = k - \int_0^k \frac{\exp(z/L)}{\exp(k/L)}\mathrm{d}z = k - L\left(1 - \exp\left(-\frac{k}{L}\right)\right) \tag{47}$$

The expression for z_0 is found by the boundary condition that at the top of the vegetation the flow velocity of the vegetation profile, $u_\mathrm{v}(k)$, equals the flow velocity of the overlying logarithmic profile, $u_\mathrm{o}(k)$:

$$\sqrt{u_{\mathrm{s}0}^2 + a\exp\left(\frac{k}{L}\right)} = \frac{\sqrt{g(h-k)i}}{\kappa}\ln\left(\frac{k-d}{z_0}\right) \tag{48}$$

Substituting Eq. (30) for $u_{\mathrm{s}0}$ and Eq. (37) for a, and rewriting using Eq. (31) for L, yields:

$$z_0 = (k-d)\exp\left(-\kappa\sqrt{\frac{2L}{c_\mathrm{p}\ell}\left(1 + \frac{L}{h-k}\right)}\right) \tag{49}$$

This expression for z_0 can be applied in the method of effective water depth, Eq. (19), as well yielding an analytical estimate for the roughness height of the top of the vegetation, dependent on known vegetation properties, water depth and the closure coefficient c_p for the mean turbulence intensity.

The formulae for the vegetation-related resistance coefficient for flow through submerged vegetation can be applied for known vegetation characteristics: diameter D, density m, bulk drag coefficient C_D and the water depth h. The bulk drag coefficient for flow through vegetation is a parameter that is difficult to determine, and many researchers have been working on it, for instance Li and Shen (1973) and Neary (2003). In this study we take the theoretical viewpoint that vegetation can be modelled as rigid cylinders and we simply apply a drag coefficient of 1.0, disregarding the placement of the cylinders or the flow Reynolds number.

The formulae for submerged vegetation include one additional parameter that cannot be measured directly in the field, nor easily estimated: in both methods the mean turbulence intensity c_p is needed. Van Velzen *et al.* (2003) compared experimental flume data on submerged reed with the results of the analytical equation of Klopstra *et al.* (1997) and found an adequate expression for the characteristic length scale of turbulent eddies inside the vegetation, α (which equals to $c_\mathrm{p}\ell$). The turbulence intensity c_p is given by:

$$c_\mathrm{p} = \frac{0.015\sqrt{hk}}{\ell} \tag{50}$$

The value for c_p affects the length scale L and, therefore, the zero-plane displacement d given by Eq. (47). To validate this expression for c_p, a comparison is made with flume data from Nepf and Vivoni (2000). In their experiment with flexible plastic plants they carefully measured the vertical profiles of Reynolds stress and calculated the zero-plane displacement by applying Eq. (45). The vegetation characteristics are: $D = 0.0167\,\mathrm{m}$, $m = 330\,\mathrm{m}^{-2}$ (yielding $mD = 5.5\,\mathrm{m}^{-1}$), $\mathrm{C_D} = 1.0$ and $k = 0.16\,\mathrm{m}$. For increasing depth ratios of h/k, Nepf and Vivoni

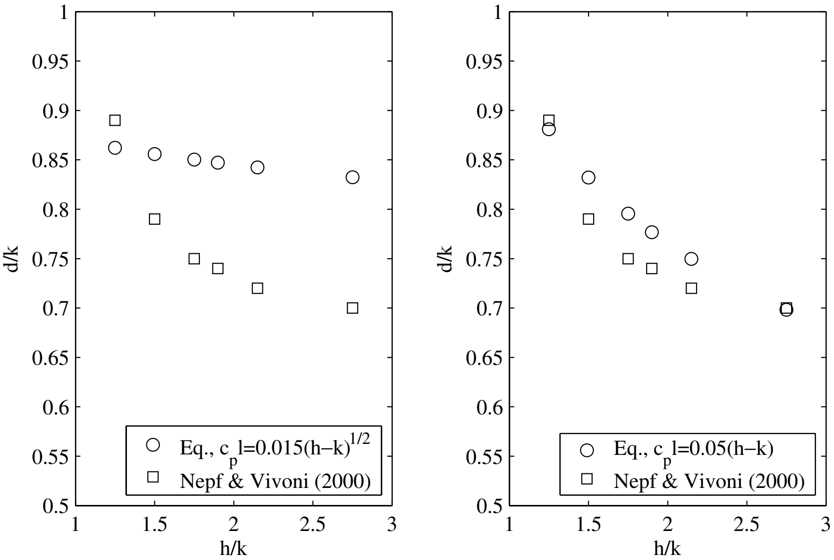

Figure 3 Comparison between measured and modelled dimensionless zero-plane displacement, d/k.

(2000) found decreasing ratios for d/k. Figure 3 presents the comparison of the measurements with the results of the analytical equation for d. The left panel shows that by applying Eq. (47) in combination with Eq. (50) for the c_p-value, the fit is not very good. Subsequently, various dimensionally correct formulae for c_p were tried and as a result the following formula for c_p gave a reasonable fit:

$$c_p = \frac{1}{20}\frac{h-k}{\ell} \tag{51}$$

This comparison with data demonstrates that Eq. (47) seems a valid approximation for the zero-plane displacement, but to simulate the zero-plane displacement, and therefore, the physical processes, accurately, the closure coefficient for the turbulence intensity, c_p, needs to be fitted. The c_p-coefficient is dependent on the submerged water depth (h–k) and mixing length ℓ, and may be different for flexible vegetation than for rigid vegetation.

Finally, Fig. 4 presents the equations for the zero-plane displacement and the roughness height for the top of the vegetation in graphical form, as a function of the plant characteristics C_DmD, for an arbitrary vegetation height and water depth. The value for c_p is given by Eq. (51). It can be seen that with increasing vegetation density the zero-plane displacement increases, in other words, the penetration of overlying eddies diminishes. The roughness height of the vegetation shows a more complex relationship with the vegetation density, showing a maximum at a relatively low density. Typical ranges for mD for natural vegetation are 0.1–1.0 m^{-1} for open herbaceous and marsh types of vegetation, and 10–15 m^{-1} for natural grasslands. Furthermore, Fig. 4 shows that the often used estimate $d = (2/3)k$ (Garratt, 1992) is within the valid range, but the exact value for d is dependent on the vegetation properties.

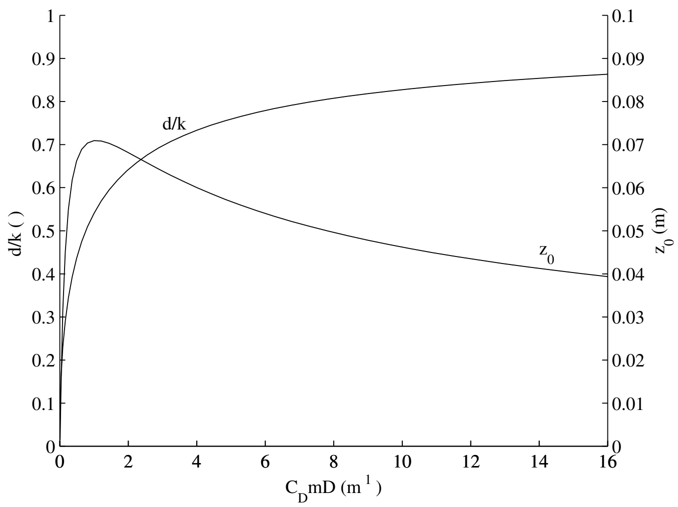

Figure 4 The dimensionless zero-plane displacement, d/k (left axis), and the roughness height for the top of the vegetation, z_0 (right axis), as function of C_DmD. $h = 2$ m, $k = 0.5$ m, $C_D = 1$.

4 1 DV turbulence model

Another way of obtaining a detailed description of resistance of flow through and above vegetation, is to perform detailed numerical simulations based on a 1-DV k–ε turbulence model that has been developed by Uittenbogaard (2003). This model is a simplification of the full 3-D Navier–Stokes equations in order to account for horizontal flow conditions only. The model describes vegetation as rigid cylinders, similar to the previous approaches. At the same time and in order to include the effects of vegetation into the k–ε turbulence closure, the following modifications have been included: (i) the decrease of the available cross-section for the vertical exchange of momentum, turbulence kinetic energy and turbulent dissipation, (ii) the drag force exerted by the plants in the horizontal direction, (iii) an additional turbulence production term due to vegetation, and (iv) an additional turbulence dissipation term due to vegetation. The 1-DV model assumes uniform flow in horizontal direction. The momentum equation reads:

$$\rho_0 \frac{\partial u(z)}{\partial t} + \frac{\partial p}{\partial x} = \frac{\rho_0}{1 - A_\mathrm{p}(z)} \frac{\partial}{\partial z} \left((1 - A_\mathrm{p}(z))(\nu + \nu_\mathrm{T}(z)) \frac{\partial u(z)}{\partial z} \right) - \frac{F(z)}{1 - A_\mathrm{p}(z)} \tag{52}$$

where F is the drag force of the vegetation per unit volume (Nm^{-3}):

$$F(z) = \frac{1}{2} \rho_0 C_\mathrm{D}(z) D(z) m(z) u(z) |u(z)| \tag{53}$$

The pressure gradient is constant along the water depth, according to the hydrostatic pressure assumption, and is either provided as input to the model, or numerically adjusted to satisfy a given depth-averaged bulk velocity U. The bulk velocity is defined by the volume flux of water divided by the channel's wetted cross-section and relates to the pore velocity by:

$$U = \frac{1}{h} \int_0^h (1 - A_\mathrm{p}) u(z) \mathrm{d}z \tag{54}$$

The continuity equation is given by:

$$\frac{\partial u}{\partial x} = 0 \tag{55}$$

The k–ε turbulence model provides a closure for the eddy viscosity, relating it to the turbulent kinetic energy, k_T ($\mathrm{m}^2/\mathrm{s}^2$) and its dissipation rate, ε ($\mathrm{m}^2/\mathrm{s}^3$):

$$\nu_\mathrm{T} = c_\mu \frac{k_\mathrm{T}^2}{\varepsilon} \tag{56}$$

where c_μ is a constant (with the standard value of 0.09). The k_T-equation in the k–ε turbulence model is modified to take into account the effect of vegetation on the additional production of turbulence and on the vertical diffusion of turbulent kinetic energy:

$$\frac{\partial k_\mathrm{T}}{\partial t} = \frac{1}{1 - A_\mathrm{p}} \frac{\partial}{\partial z} \left((1 - A_\mathrm{p}) \left(\nu + \frac{\nu_\mathrm{T}}{\sigma_k} \right) \frac{\partial k_\mathrm{T}}{\partial z} \right) + T + P_k - B_k - \varepsilon \tag{57}$$

The first term in the right-hand side of Eq. (57) is the vertical diffusion of turbulent kinetic energy by its own mixing action, corrected for the specific area, $(1 - A_\mathrm{p})$. A closure coefficient of $\sigma_k = 1.0$ was applied. The second term, T, denotes the additional turbulence generated by the vegetation. Considering fully turbulent flow, all the work done by the fluid against the plants drag force is converted into turbulent kinetic energy, making the expression for T:

$$T(z) = F(z) u(z) \tag{58}$$

For transient or laminar flow, part or all of this power would be transferred into heat by work against viscous forces and correction terms depending on Reynolds number would be needed. The third term in the right-hand side, P_k, is the turbulence production in shear flows:

$$P_k = \nu_\mathrm{T} \left(\frac{\partial u}{\partial z} \right)^2 \tag{59}$$

The fourth term, B_k, is the buoyancy term which represents the conversion of turbulent kinetic energy into gravitational energy according to:

$$B_k = -\frac{\nu_\mathrm{T}}{\sigma_k} \frac{g}{\rho_0} \frac{\partial \rho}{\partial z} \tag{60}$$

Finally, the last term, ε, corresponds to the dissipation rate of the turbulent kinetic energy. This is modelled by the ε-equation:

$$\frac{\partial \varepsilon}{\partial t} = \frac{1}{1 - A_\mathrm{p}} \frac{\partial}{\partial z} \left((1 - A_p) \left(\nu + \frac{\nu_\mathrm{T}}{\sigma_k} \right) \frac{\partial \varepsilon}{\partial z} \right) + P_\varepsilon - B_\varepsilon - \varepsilon_\varepsilon + c_{2\varepsilon} \frac{T}{\tau_\mathrm{eff}} \tag{61}$$

The first term in the right-hand side of Eq. (61) represents the vertical diffusion of ε by the turbulent eddies. For the closure coefficient, $\sigma_\varepsilon = 1.3$ is applied. The next three terms P_ε, B_ε and ε_ε, correspond to the production, buoyancy and dissipation of ε, respectively, and are related to the production, buoyancy and dissipation of k_T by the expressions:

$$P_\varepsilon = c_{1\varepsilon} \frac{\varepsilon}{k_\mathrm{T}} P_k \tag{62}$$

$$B_\varepsilon = c_{1\varepsilon} (1 - c_{3\varepsilon}) \frac{\varepsilon}{k_\mathrm{T}} B_k \tag{63}$$

$$\varepsilon_\varepsilon = c_{2\varepsilon} \frac{\varepsilon^2}{k_\mathrm{T}} \tag{64}$$

where $c_{1\varepsilon} = 1.44$, $c_{2\varepsilon} = 1.92$ and $c_{3\varepsilon} = 0$ or 1 (depending on stratification). Universal values for closure coefficients, as derived by Launder and Spalding (1974), have been applied.

The important part is in the last term in the right-hand side, which corresponds to the dissipation rate of turbulence produced by vegetation. This dissipation rate depends on the effective turbulence dissipation time scale (τ_eff) and it is affected by the closure coefficient $c_{2\varepsilon}$. To obtain τ_eff, Uittenbogaard (2003) related it to the different length scales that control turbulence inside vegetation. First of all, at sufficient distance from the bed as well as from the top of the vegetation, the length scale of internally generated turbulence is smaller than the available fluid space inside the vegetation, and therefore the relevant time scale of this small scale turbulence corresponds to the intrinsic turbulence time scale:

$$\tau_\mathrm{int} = \frac{k_\mathrm{T}}{\varepsilon} \tag{65}$$

This time scale is adopted as an effective time scale by Shimizu and Tsujimoto (1994) and López and García (2001). It is here where the turbulence model of Uittenbogaard (2003) differs from that of previous authors. Uittenbogaard includes the penetration of shear turbulence from above the vegetation into the top layer of the vegetation. Above the vegetation a shear layer is formed by the vertical exchange of horizontal momentum with the retarded flow inside the vegetation. The large eddies that are advected from above the vegetation have to be squeezed into smaller-scale eddies of the length scale of the available pore space inside the vegetation. In this way, the relevant time scale for the dissipation is determined by the geometrical properties of the vegetation, according to:

$$\tau_{\text{geom}} = \left(\frac{L_{\text{p}}^2}{c_\mu^2 T} \right)^{1/3} \tag{66}$$

where L_{p} is the available length scale for eddies between the vegetation:

$$L_{\text{p}}(z) = c_\ell \left(\frac{1 - A_p(z)}{m(z)} \right)^{1/2} \tag{67}$$

in which c_ℓ is a coefficient that affects the geometrical length scale. For cylinders, it is assumed that $c_\ell = 1$, but it is noted that for real vegetation with twigs and foliage c_ℓ might be smaller (Baptist, 2005). The 1-DV model computes both time scales τ_{int} and τ_{geom} over the vertical and evaluates the effective time scale by a MAX-operator:

$$\tau_{\text{eff}}^{-1} = \max\left(\tau_{\text{int}}^{-1}, \tau_{\text{geom}}^{-1}\right) \tag{68}$$

Finally, the representative roughness is obtained from the 1-DV model by calculating the energy slope, i:

$$i = \frac{\partial p / \partial x}{\rho_0 g} \tag{69}$$

and subsequently applying Chézy's relationship. The validity of the 1-DV model has been demonstrated by comparing the outcome with measurements of vertical profiles for flow and turbulence by Meijer and Van Velzen (1999), Nepf and Vivoni (2000) and López and García (2001), and will be demonstrated in a later section of this paper.

5 Equation building with genetic programming

When refining a model of a physical process, a scientist focuses on the agreement of theoretically predicted and experimentally observed behaviour. If these agree in some accepted sense, then the model is "correct" within that context. In the process of making sense of experimental data it is generally desirable to express the relation between the variables in a symbolic form: an equation. In this work we consider the problem inverse to verification of theoretical models: how can we obtain the governing equations directly from measurements? To do this, we will apply genetic programming.

5.1 Dimensionally aware genetic programming

Genetic programming (GP) is a technique that can be used to find the symbolic form of an equation, including a set of coefficients. One of the advantages of genetic programming over other methods for regression is the symbolic nature of the solutions that are produced. This is especially pronounced in empirical modelling of unknown phenomena where an underlying theoretical model is not known. For a detailed description of genetic programming from a water resources perspective the interested reader is referred to Babovic and Abbott (1997) and Babovic and Keijzer (2000). Inspired by Koza's (1992) pioneering work on GP and in order to improve performance of his standard approach, an augmented version of GP has been proposed — dimensionally aware GP (Keijzer and Babovic, 1999) — which is arguably more useful in the process of scientific discovery. Dimensionally aware genetic programming (Keijzer and Babovic, 1999) differs from the standard approach in that the raw observations are used together with their units of measurement. The system of units of measurement can be viewed as a typing scheme and as such can be used in some form of typed genetic programming. The dimensionally aware approach proposes what can be called a *weakly typed* or *implicit casting* approach. Here dimensional correctness is not enforced, but promoted. An extra objective for selection, goodness-of-dimension, is introduced that is used next to a goodness-of-fit objective. These two objectives are then used in a multi-objective optimization routine using the concepts of dominance and Pareto optimality. Goodness-of-dimension is measured by calculating how many constants with appropriate units should be introduced to render an equation dimensionally correct. The result of a single run of such unit typed genetic programming is a number of equations — a so-called Pareto front of non-dominated solutions — that balance dimensional correctness (goodness-of-dimension) with goodness-of-fit.

5.2 Determination of a vegetation-related resistance formula for submerged vegetation using genetic programming

Dimensionally aware genetic programming was applied to a set of 990 results of the 1-DV model for submerged vegetation. These 990 cases were chosen to represent a wide variety of vegetation properties and water depths for cylinders that represent pioneer species, (stiff) grasses, herbaceous vegetation, reed and bushes. The input variables are presented in Table 1.

For purposes of dimensional consistence a slightly adapted Chézy coefficient was used:

$$C' = \frac{C}{\sqrt{g}} \tag{70}$$

By virtue of using such a dimensionless coefficient, time-related units are avoided and the resistance coefficient became solely a function of the geometry of the system. GP was employed in a multi-objective sense, so that the following three objectives were simultaneously optimized: (i) root mean square error (RMSE): measure of overall accuracy of the formula, (ii) coefficient of determination (CoD): measure of the goodness of the shape of

Table 1 Inputs to the 1-DV k–ε model

Input	Dimension	Description
D	L	Diameter of the stems
m	L^{-2}	Number of stems per square metre
k	L	Vegetation height
C_D	—	Bulk drag coefficient
C_b	$L^{0.5}/T$	Bed Chézy resistance coefficient
h	L	Water depth

the formula and (iii) dimensional error: measure of dimensional consistency of the formulae.

The GP induction system was run many times with different parameters using various subsets of the 990 cases. Subsequently the front of non-dominated solutions for all the runs was examined to find a suitable formula. Genetic programming results provided a dimensionally consistent formula that had both smallest RMSE and the highest CoD:

$$\frac{C_r}{\sqrt{g}} = \sqrt{\frac{2}{c_D m D k}} + \ln\left\{\left(\frac{h}{k}\right)^2\right\} \tag{71}$$

This formula can be rearranged to:

$$C_r = \sqrt{\frac{2g}{c_D m D k}} + 2\sqrt{g}\ln\left(\frac{h}{k}\right) \tag{72}$$

Clearly, the relationship of the resistance coefficient C_r and the water depth h consists of a h-independent term plus a h-dependent logarithmic term. For $h = k$, the logarithmic term reduces to zero, making the resistance equivalent to the resistance for flow through non-submerged vegetation, for maximally flooded conditions, $h = k$, see Eq. (13). Having found this, it would be better to have an expression compatible with the more accurate expression for resistance for non-submerged conditions that includes the bed resistance, as well, Eq. (12) for fully submerged conditions:

$$C_k = \sqrt{\frac{1}{(1/C_b^2) + (C_D m D k/2g)}} \tag{73}$$

Finally, replacing the constant value of 2 with the more theoretically founded Von Kármán's constant $1/\kappa \approx 2.5$, the final expression was determined to be:

$$C_r = \sqrt{\frac{1}{(1/C_b^2) + (C_D m D k/2g)}} + \frac{\sqrt{g}}{\kappa}\ln\left(\frac{h}{k}\right) \tag{74}$$

Although fractionally more complicated than the original formulation, expression (74) was found to be in good agreement with the data, especially in regions with higher Chézy values (Fig. 5d). Furthermore, it is theoretically well founded, combining the resistance for the flow inside the vegetation with the observed logarithmic profile above the vegetation.

Subsequent research on the formulation revealed an agreement with the work by Kouwen *et al*. (1969), where a general formula for resistance induced by vegetation was proposed as:

$$C_r = C_1 + \frac{\sqrt{g}}{\kappa}\ln\left(\frac{h}{k}\right) \tag{75}$$

Kouwen *et al.* proposed several relationships for C_1, but no definitive conclusions were drawn. Equation (74) gives an exact formulation for C_1 and as such presents a step forward in the modelling of resistance. Note that the final formulation is a combination of a computer induced expression that fits the data well, and theoretically based modifications to fit the theory. Further analysis reveals that the equation is equivalent to the equation produced by the method of effective water depth, Eq. (19) if, (i) the depth balance is ignored (i.e., the factors k and $(h - k)$ are considered equal to h), and (ii) the roughness length z_0 is set to k/e where e is the base of the natural logarithm. This leads to the assumption that the resistance is mainly governed by a logarithmic flow profile over the submerged depth, with a boundary condition at the level of zero intercept, which is at the top of the vegetation. Note further that Eq. (74) is obtained by integrating the differential equation:

$$\frac{\partial C_r}{\partial h} = \frac{\sqrt{g}}{\kappa h} \tag{76}$$

with the boundary condition of $C = C_{ref}$ at $h = h_{ref}$. In this case the logical boundary condition is $h_{ref} = k$.

6 Comparison of the formulations

There are several ways to evaluate the formulations. First and foremost is the ability to model the data under study. This is evaluated by comparing the results of the analytical formulations with those of the 1-DV numerical model. Evaluation parameters are the RMSE:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_{Oi} - x_{Mi})^2} \tag{77}$$

where N is the number of observations, and x_{Oi} and x_{Mi} are the observed and modelled values, respectively, and the CoD, which in this linear regression case equals R^2, where R is the correlation coefficient. Results are presented in Table 2. It can be seen that the expressions based on the genetic programming results are in better agreement with the synthetic dataset than the

Table 2 Root mean squared errors (RMSE) and coefficients of determination (CoD) for the four formulations of the resistance coefficients, including the 1-DV model which was used for generation of training data, compared with the 1-DV model results.

Equation	RMSE 1-DV data ($m^{1/2}/s$)	CoD 1-DV data (–)
Method of effective water depth, Eq. (19)	1.3048	0.97485
Analytical solution method, Eq. (44)	1.4705	0.93724
Original GP-formula, Eq. (71)	0.9746	0.97065
Modified GP-formula, Eq. (74)	1.2061	0.97579
1-DV numerical model	0	1.00

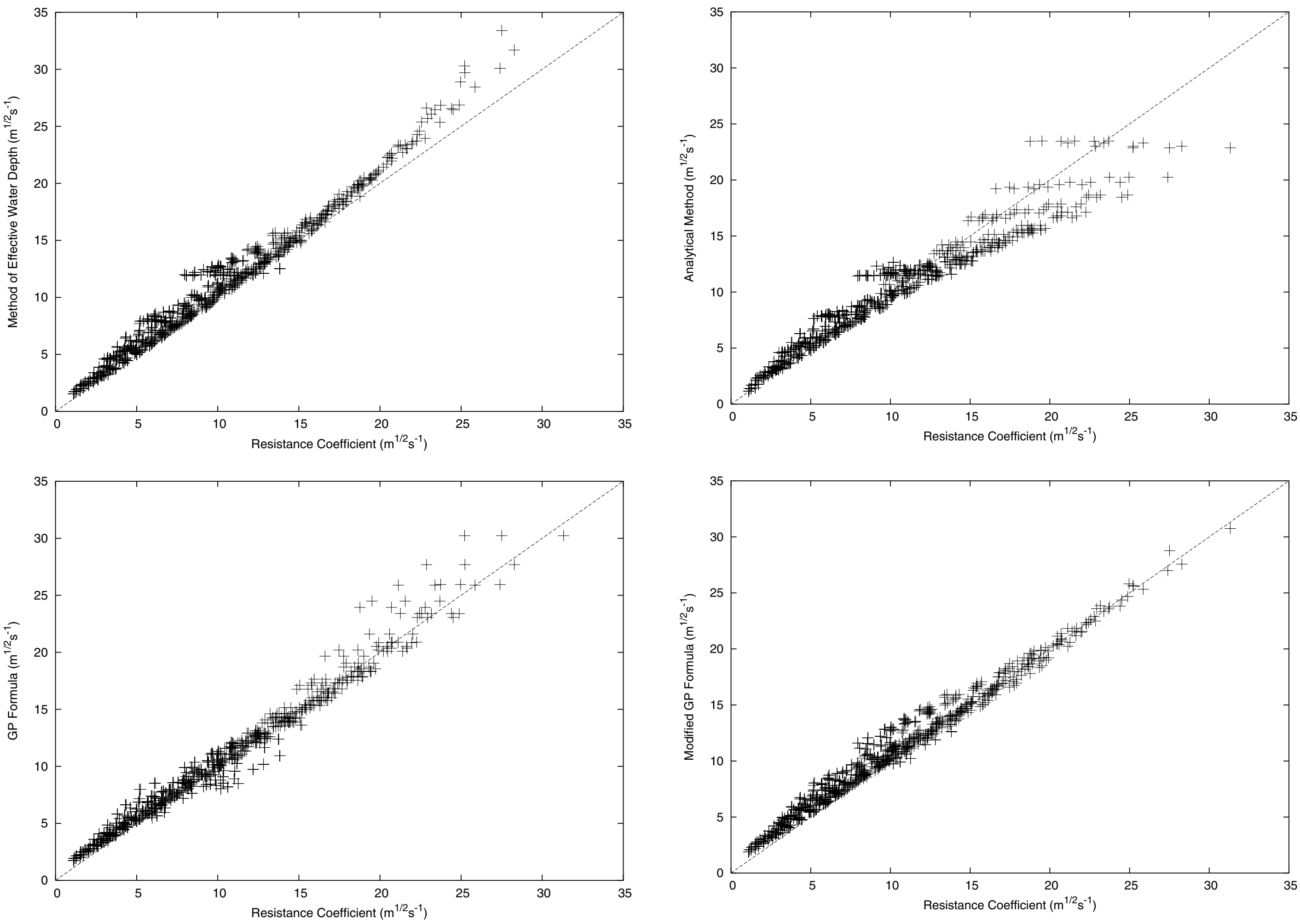

Figure 5 Scatter plots for the four equations on 1-DV data. (a) upper left: method of effective water depth, Eq. (19), (b) upper right: analytical solution method, Eq. (44), (c) lower left: original GP-formula, Eq. (71) and (d) lower right: modified GP-formula, Eq. (74).

manually induced formulations. There is a small level of systematic error, but the GP-formulae model the variance admirably. Even though the modified GP-formula has a higher RMSE than the original GP-formula, the scatter plot in Fig. 5 reveals that the improved formulation removes the large residuals associated with high Chézy values, and thus is more reliable over the entire domain than the original formulation.

Up to this point, both training and comparisons were performed on synthetic data, generated by the 1-DV model. To ultimately test our approaches, 177 experimental runs based on laboratory flume experiments were collected from 10 independent studies. These data were not used in the equation building process, but kept aside to validate the equations induced.

These studies include rigid and flexible, artificial and natural vegetation types from Kouwen *et al*. (1969), Ree and Crow (1977), Murota *et al*. (1984), Tsujimoto and Kitamura (1990), Tsujimoto *et al*. (1993), Ikeda and Kanazawa (1996), López and García (1997, 2001), Meijer (1998a, b) and Järvelä (2003). The data contained all input information, except the bed roughness C_b, which was assumed to be negligible in the experiments for smooth flume beds, whereupon it was set to 60 $m^{1/2}$/s. In cases where the drag coefficient was not defined it was assumed to be 1.0. For flexible vegetation experiments, the deflected height was applied in the formulations. An overview of the data of the 177 experiments is given in Appendix A of Baptist (2005). It is noted that although the range in parameter values in these experiments is quite large, in comparison with natural vegetation types, flume experimental data are not covering the full range of existing types.

For this particular dataset it was possible to also test the performance of the 1-DV model itself. Results for the comparison can be found in Table 3 and Fig. 6, and it can be seen that the genetic programming induced equations give a highly competitive agreement with the data. What is particular startling is that their performance is even competitive with the 1-DV model the equations are based upon.

Finally, Fig. 7 presents a comparison between the improved genetic programming equation and the original 1-DV model, both applied to the validation set of flume experiments. No serious discrepancies between the dynamical model and the simple equation are observed.

From the perspective of simplicity of the equations, it might be enlightening to compare the apparatus of expressions (Eqs (5)–(49)) that leads to the definition of Eqs (19) and (44), with the conciseness and elegancy of the genetic programming induced Eq. (71) and its human-manipulated variant, Eq. (74). The genetic programming induced equations are based purely on the 1-DV

Table 3 Root mean squared errors (RMSE) and coefficients of determination (CoD) for the four formulations of the resistance coefficients, including the 1-DV model which was used for generation of training data, compared with flume experiments

Equation	RMSE flume experiments ($m^{1/2}/s$)	CoD flume experiments (–)
Method of effective water depth, Eq. (19)	2.7187	0.83594
Analytical solution method, Eq. (44)	2.2796	0.81325
Original GP-formula, Eq. (71)	2.1093	0.83737
Modified GP-formula, Eq. (74)	2.1826	0.87418
1-DV numerical model	1.8600	0.87300

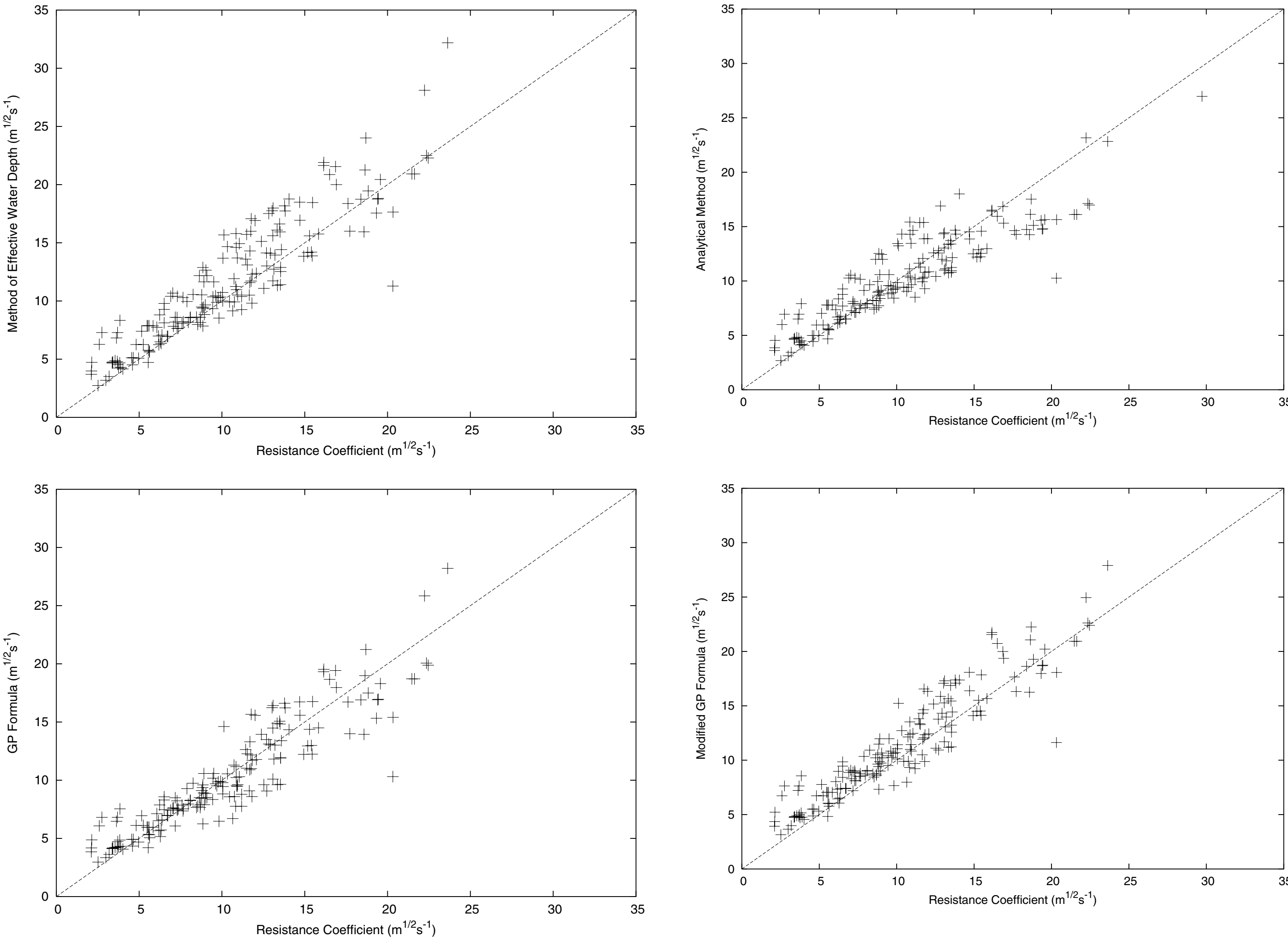

Figure 6 Scatter plots for the four equations for experimental flume data. (a) upper left: method of effective water depth, Eq. (19), (b) upper right: analytical solution method, Eq. (44), (c) lower left: original GP-formula, Eq. (71) and (d) lower right: modified GP-formula, Eq. (74).

model data, and yet ignore considerations about turbulence length scales and turbulence intensity. They focus primarily on obtaining good agreement with the data, and in this case, the problematic interplay between the vegetation and the turbulence induced by the vegetation apparently are of secondary importance to the simple logarithm on the ratio of water depth over plant height.

7 Discussion

This paper reports the comparison of four different methods for calculating flow resistance due to vegetation. In particular, a GP algorithm showed to induce an appropriate equation for vegetation resistance from the output of a 1-DV numerical model. In comparison to a variety of flume data, a modified version of this equation is shown to provide the best fit with minimized residuals over the whole data range. The induced equation is sufficiently simple to be applied to wide-area depth-averaged models, to calculate, for example, flow over vegetated floodplains.

The main limitation of the study is the assumption that vegetation can be represented as rigid cylinders. In vegetated floodplains, several vegetation types, such as herbs and grasses will bend due to the force of the flow. Bending of vegetation decreases the frontal area for drag, introduces lift forces

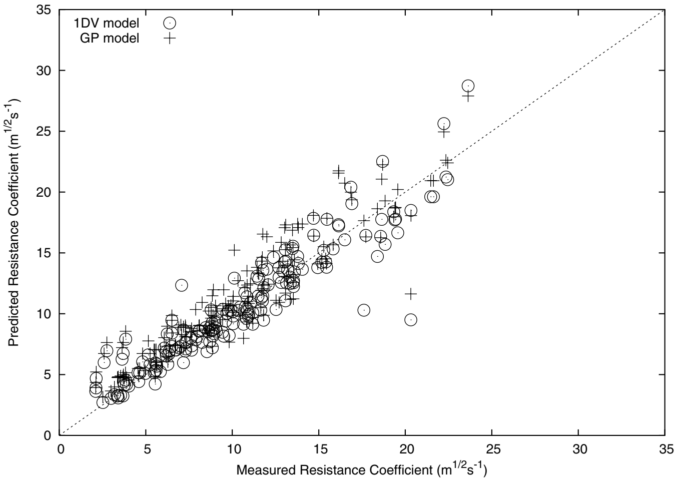

Figure 7 Direct comparison between the 1-DV model (circles) and the modified GP-formula, Eq. (74) (plusses), using experimental flume data (i.e. out of sample validation).

and furthermore, swaying vegetation may introduce additional resistance. It has been found that for rigid vegetation the linear relation between force and the square of velocity holds, as depicted in this paper, but for flexible plants a linear increase of drag force with flow velocity is observed (Armanini *et al.*, 2005). This implies that the equation for resistance found is less suitable for flexible vegetation, although the comparison with flume data, which included flexible vegetation as well, showed a reasonable fit.

Testing expressions for their ability to model a phenomenon such as resistance induced by vegetation requires experimental data. Particularly when using data driven methods, data are needed for steering the error minimization process. For equations induced by scientists such data is needed to test the proposed equation for its capacity to model the phenomenon under study.

Even though the variables for this relationship should be measurable in wetlands and vegetated floodplains, this cannot yet be accomplished for resistance coefficients measured at a fine scale, at different water depths and with different types of vegetation. Obtaining data at such a fine scale in realistic circumstances is prohibitive in terms of effort and cost. Instead, we here chose to use a fine-scaled numerical 1-DV model, describing turbulence and flow properties to generate data. Such a model employs available knowledge about characteristics of vegetation, turbulence induced by vegetation, and resistance caused by the drag forces on the plants. Determining values of the resistance coefficients from such a microscopic model is trivial, and can be understood as a noise free approximation of the phenomenon under study. However, a compact expression describing the phenomenon is not readily available from a numerical 1-DV model.

It could be argued that using synthetic data defeats the purpose of finding an equation. If a dynamic model exists, why not simply use that one instead of going the laborious route of finding a compact equation. The purpose of finding a compact equation lies in the type of modelling that it enables. Vegetation resistance is a typical three-dimensional problem due to the spatial heterogeneity of vegetation and the water depth dependency of submerged or non-submerged vegetation. A full dynamical model thus operates on a 3-D grid, which is computationally expensive. An analytical solution to the problem of resistance induced by vegetation, which includes water depth dependency, makes two-dimensional, depth-averaged modelling possible, allowing for faster model computations and the possibility to apply the model to larger areas. In addition, an analytical expression can also be used in 1-DH computations, or even in spreadsheet models.

A difficulty of the approach, even in simple models, is that floodplains have inhomogeneous vegetation types with a complex structure (Baptist *et al.*, 2005). This shifts the modelling problem from estimating an equivalent bed roughness value to estimating plant properties for density, diameter, height and drag coefficient. In principle it is possible, although not easy, to obtain the complete horizontal and vertical structure of grasses, herbaceous vegetation, bushes, floodplain forests and other vegetation types. The drag coefficient, however, is a property that cannot be measured directly in the field. In principle, the drag coefficient for smooth cylinders is known from experimental studies and theory and is dependent on the value for the Reynolds number of the flow and the spatial arrangement of the cylinders. In applications on vegetation roughness this coefficient is usually estimated or used as a calibration parameter, for example in an effort to account for rough surfaces or for the foliage. Given Eq. (74), and calibrated on the drag coefficient, its main advantage over just calibrating a value for C_r directly, is that it gives a relationship with water depth.

8 Conclusions

Four formulations for water depth-related resistance induced by vegetation were studied and compared. Two of the equations studied here were created through analysis and a process of derivation by a scientist. One equation was derived by dimensionally aware genetic programming, and finally one expression was created by manually analyzing and improving the genetic programming equation. It was found that the genetic programming equations were superior to the manually derived equations, both on their performance on synthetic training and laboratory testing data, and in the economy of detail that needs to be modelled. The manually improved expression was found to be in good agreement with previous expressions found in the literature, and performed competitive on experimental data with the detailed numerical model that was used to generate the training data.

This paper presented a case study in the use of genetic programming as a hypothesis generator for use in scientific discovery. Not only does it show that genetic programming is capable of producing equations that are comparable or perhaps better than their human derived competitors, it produces expressions that are amenable to further analysis and manual improvement. This is potentially a much more useful result, as it shows that the symbolic nature of genetic programming can be used to build up knowledge in a problem domain. In contrast with many machine learning algorithms, where the trained model is the end result of a problem statement, the genetic programming induced expressions can be used to start a new cycle of inquiry.

The elegant Eq. (74) that was induced with genetic programming is both theoretically and experimentally justified, and can be used to estimate the resistance coefficient of submerged vegetation.

Notation

a = Integration constant (m/s^2)
A_p = Solidity; the fraction of horizontal area taken by the cylinders
a_v = Integration constant a divided by i (m/s^2)
b = Integration constant (m/s^2)
B_k = Buoyancy of turbulence (m^2/s^3)
B_ε = Buoyancy of turbulence dissipation (m^2/s^4)
C = Chézy coefficient (m$^{1/2}$/s)
c_μ = Constant
$c_{1\varepsilon}$ = Closure coefficient
$c_{2\varepsilon}$ = Closure coefficient
$c_{3\varepsilon}$ = Closure coefficient
C_b = Chézy coefficient of the bed (m$^{1/2}$/s)
C_D = Bulk drag coefficient
C_k = Representative Chézy value for non-submerged vegetation (m$^{1/2}$/s)
c_ℓ = Coefficient for the geometrical length scale
c_p = Turbulence intensity, height-averaged over the vegetation height
C_r = Representative Chézy value for vegetation (m$^{1/2}$/s)
D = Cylinder diameter (m)
d = Zero-plane displacement in the logarithmic velocity profile (m)
F = Drag force of the vegetation per unit volume (N/m^3)
g = Gravitational acceleration (m/s^2)
h = Water depth (m)
i = Energy gradient or water level slope
k = Vegetation height (m)
k_N = Nikuradse equivalent roughness (m)
k_T = Turbulent kinetic energy per unit mass (m^2/s^2)
L = Length scale (m)
ℓ = Mixing length (m)
L_p = Available length scale for eddies between the vegetation (m)
m = Number of cylinders per m^2 horizontal area (m^{-2})
p = Pressure (N/m^2)
P_k = Turbulence production in shear flows (m^2/s^3)
P_ε = Production of turbulence dissipation (m^2/s^4)
q = Discharge per unit width (m^2/s)
R = Hydraulic radius (m)
t = Time (s)
T = Turbulence generated by vegetation (m^2/s^3)
u = Velocity (m/s)
U = Depth-averaged bulk velocity (m/s)
u_* = Shear velocity (m/s)
u_c = Uniform flow velocity through fully submerged vegetation (m/s)
u_{cb} = Uniform flow velocity through non-submerged vegetation (m/s)
u_o = Velocity above the vegetation layer (without slip velocity) (m/s)
$\bar{u}$ = Depth-averaged velocity (m/s)
$\bar{u}_o$ = Height-averaged velocity above the vegetation layer (m/s)
u_{s0} = Uniform flow velocity over part of the cylinder height (m/s)
u_u = Velocity above the vegetation layer (including slip velocity) (m/s)
u_v = Velocity inside the vegetation layer (m/s)
$\bar{u}_v$ = Height-averaged velocity inside the vegetation layer (m/s)
u_{v0} = Uniform flow velocity over part of the cylinder height divided by i (m/s)
u_{vk} = Flow velocity at the top of the vegetation, $u_v(k)$ (m/s)
x = Horizontal coordinate (m)
z = Vertical coordinate (m)
z_0 = Roughness height in the logarithmic velocity profile for a fully rough bed (m)

Greek symbols

α = Characteristic length scale (m)
ε = Turbulence dissipation rate per unit mass (m^2/s^3)
ε_ε = Dissipation rate of turbulence dissipation (m^2/s^4)
κ = Von Kármán constant
ν = Kinematic viscosity of water (m^2/s)
ν_T = Eddy viscosity (m^2/s)
ρ_0 = Fluid density (kg/m^3)
σ_k = Closure coefficient
σ_ε = Closure coefficient
τ_b = Bed shear stress (N/m^2)
τ_{eff} = Effective turbulence dissipation time scale (s)
τ_{geom} = Geometric turbulence dissipation time scale (s)
τ_{int} = Intrinsic turbulence dissipation time scale (s)
τ_t = Total fluid shear stress (N/m^2)
τ_v = Vegetation resistance force per unit horizontal area (N/m^2)
τ_{xz} = Reynolds stress or shear stress (N/m^2)

References

1. Armanini, A., Righetti, M. and Grisenti, P. (2005). "Direct measurement of Vegetation Resistance in Prototype Scale". *J. Hydraul. Res.* 43(5), 481–487.

2. BABOVIC, V. and ABBOTT, M.B. (1997). "The Evolution of Equation from Hydraulic Data, Part I: Theory and Part II: Application", *J. Hydraul. Res.* 35(3), 397–430.
3. BABOVIC, V. and KEIJZER, M. (2000). "Genetic Programming as a Model Induction Engine". *J. Hydroinformat.* 2(1), 35–60.
4. BAPTIST, M.J. (2005). "Modelling Floodplain Biogeomorphology". PhD thesis, Delft University of Technology, Faculty of Civil Engineering and Geosciences, Section Hydraulic Engineering, Delft, ISBN 90-407-2582-9, 213 pp.
5. BAPTIST, M.J., VAN DEN BOSCH, L.V., DIJKSTRA, J.T. and KAPINGA, S. (2005). "Modelling the Effects of Vegetation on Flow and Morphology in Rivers". *Archiv. Hydrobiol. Suppl.* 155(1–4), *Large Rivers* 15(1–4), 339–357.
6. CAMPANA, T.R. (1999). "Hydraulic Resistance of Submerged Floodplain Vegetation". MSc Thesis, IHE-Delft, Delft, H.E.043.
7. CAROLLO, F.G., FERRO, V. and TERMINI, D. (2002). "Flow Velocity Measurements in Vegetated Channels". *J. Hydraul. Engng.* 128(7), 664–673.
8. DAWSON, F.H. and CHARLTON, F.G. (1988). "Bibliography on the Hydraulic Resistance of Vegetated Watercourses." Technical Report, Freshwater Biological Association, Occasional Publication No. 25, ISNN 0308-6739, 25 pp.
9. DARBY, S.E. (1999). "Effect of Riparian Vegetation on Flow Resistance and Flood Potential". *J. Hydraul. Engng.*, 125(5), 443-445.
10. EINSTEIN, H.A. and BANKS, R.B. (1950). "Fluid Resistance of Composite Roughness". *Trans. Am. Geophy. Union* 31(4), 603–610.
11. FATHI-MAGHADAM, M. and KOUWEN, N. (1997). "Nonrigid, Nonsubmerged, Vegetative Roughness on Floodplains". *J. Hydraul. Engng.* 123(1), 51–57.
12. FISCHER-ANTZE, T., STOESSER, T., BATES, P. and OLSEN, N.R.B. (2001). "3D Numerical Modelling of Open-channel Flow with Submerged Vegetation". *J. Hydraul. Res.* 39(3), 303–310.
13. GARRATT, J. (1992). *The Atmospheric Boundary Layer*. Cambridge University Press, New York, NY, 316 pp.
14. GHISALBERTI, M. and NEPF, H.M. (2002). "Mixing Layers and Coherent Structures in Vegetated Aquatic Flows". *J. Geophys. Res.* 107(C2), 3-1–3-11.
15. GHISALBERTI, M. and NEPF, H.M. (2004). "The Limited Growth of Vegetated Shear Layers". *Water Resour. Res.* 40(doi:10.1029/2003WR002776, 2004).
16. HASEGAWA, K., ASAI, S., KANETAKA, S. and BABA, H. (1999). "Flow Properties of a Deep Open Experimental Channel with a Dense Vegetation Bank". *J. Hydrosci. Hydraul. Engng.* 17(2), 59–70.
17. HOFFMANN, M. (2004). "Application of a Simple Space-time Averaged Porous Media Model to Flow in Densely Vegetated Channels". *J. Porous Media* 7(3), 183–191.
18. HONG, H.T.M. (1995). "Hydraulic Resistance of Flexible Roughness". MSc Thesis, IHE Delft, H.H.237.
19. IKEDA, S. and KANAZAWA, M. (1996). "Three-dimensional Organized Vortices above Flexible Water Plants". *J. Hydraul. Engng.* 122(11), 634–640.
20. IKEDA, S., YAMADA, T. and TODA, Y. (2001). "Numerical Study on Turbulent Flow and Honami in and above Flexible Plant Canopy". *Int. J. Heat Fluid Flow* 22, 252–258.
21. JACKSON, P.S. (1981). "On the Displacement Height in the Logarithmic Velocity Profile". *J. Fluid Mechs.* 111, 15–25.
22. JAMES, C.S., BIRKHEAD, A.L., JORDANOVA, A.A. and O'SULLIVAN, J.J. (2004). "Flow Resistance of Emergent Vegetation". *J. Hydraul. Res.* 42(4), 390–398.
23. JÄRVELÄ, J. (2002). "Flow Resistance of Flexible and Stiff Vegetation: A Flume Study with Natural Plants". *J. Hydrol.* 269(1–2), 44–54.
24. JÄRVELÄ, J. (2003). "Influence of Vegetation on Flow Structure in Floodplains and Wetlands". *River Coastal and Estuarine Morphodynamics 2003*, Barcelona, Spain, pp. 845–856.
25. JÄRVELÄ, J. (2004). "Determination of Flow Resistance Caused by Non-submerged Woody Vegetation". *Int. J. River Basin Manage.* 2(1), 1–10.
26. KATUL, G., WIBERG, P., ALBERTSON, J. and HORNBERGER, G. (2002). "A Mixing Layer Theory for Flow Resistance in Shallow Streams". *Water Resour. Res.* 38(11), 1250, doi:10.1029/2001WR000817, 2002.
27. KEIJZER, M. and BABOVIC, V. (1999). "Dimensionally Aware Genetic Programming". In BANZHAF, W., DAIDA, J., EIBEN, A.E., GARZON, M.H., HONAVAR, V., JAKIELA, M. and SMITH, R.E. (eds), *Proceedings of the Genetic and Evolutionary Computation Conference*, Vol. 2, pp. 1069–1076, Orlando, Florida, USA, 13–17 July 1999. Morgan Kaufmann.
28. KLAASSEN, G.J. and VAN DER ZWAARD, J.J. (1974). "Roughness Coefficients of Vegetated Floodplains". *J. Hydraul. Res.* 12(1), 43–63.
29. KLOPSTRA, D., BARNEVELD, H.J., VAN NOORTWIJK, J.M. and VAN VELZEN, E.H. (1997). "Analytical Model for Hydraulic Roughness of Submerged Vegetation". *27th International IAHR Conference*, San Fransico, CA, pp. 775–780.
30. KOUWEN, N., UNNY, T.E. and HILL, H.M. (1969). "Flow Retardance in Vegetated Channels". *J. Irrig. Drain. Div.* 95(IR2), 329–342.
31. KOUWEN, N. and UNNY, T.E. (1973). "Flexible Roughness in Open Channels". *J. Hydraul. Div.* 99(HY5), 713–728.
32. KOZA, J.R. (1992). *Genetic Programming: On the Programming of the Computes by Means of Natural Selection*. MIT Press.
33. LAUNDER, B. and SPALDING, D. (1974). "The Numerical Computation of Turbulent Flows". *Computer Methods Appl. Mech. Engng.* 3, 269–289.
34. LI, R.M. and SHEN, H.W. (1973). "Effect of Tall Vegetations on Flow and Sediment". *J. Hydraul. Div.* 99(HY6), 1085–1103.
35. LÓPEZ, F. and GARCÍA, M.H. (1997). "Open-channel Flow Through Simulated Vegetation: Turbulence Modelling

and Sediment Transport". Technical Report WRP-CP-10, Hydrosystems Laboratory, University of Illinois, Urbana, IL.

36. LÓPEZ, F. and GARCÍA, M.H. (2001). "Mean Flow and Turbulence Structure of Open-Channel Flow through Non-Emergent Vegetation". *J. Hydraul. Engng*. 127(5), 392–402.
37. MASON, D.C., COBBY, D.M., HORRITT, M.S. and BATES, P.D. (2003). "Floodplain Friction Parameterization in Two-dimensional River Flood Models using Vegetation Heights Derived from Airborne Scanning Laser Altimetry". *Hydrol. Process.* 17, 1711–1732.
38. MEIJER, D.M. (1998a). "Flume Studies of Submerged Vegetation". *PR121.10*, HKV, Lelystad (in Dutch).
39. MEIJER, D.M. (1998b). "Flume Studies of Submerged Reed". *PR177*, HKV, Lelystad (in Dutch).
40. MEIJER D.G. and VAN VELZEN, E.H. (1999). "Prototype-scale Flume Experiments on Hydraulic Roughness of Submerged Vegetation". *28th International IAHR Conference*, Graz.
41. MUROTA, A., FUKUHARA, T. and SATO, M. (1984). "Turbulence Structure in Vegetated Open Channel Flows". *J. Hydrosci. Hydraul. Engng.* 2(1), 47–61.
42. NEARY, V.S. (2003). "Numerical Solution of Fully Developed Flow with Vegetative Resistance". *J. Engng. Mech.* 129(5), 558–563.
43. NEPF, H.M. and VIVONI, E.R. (2000). "Flow Structure in Depth-limited, Vegetated Flow". *J. Geophys. Res.* 105(C12), 28, 547–28, 557.
44. NEZU, I. and NAOT, D. (1999). "Partly Vegetated Open Channels; New Experimental Evidence". *28th International IAHR Conference*, Graz.
45. NIKURADSE, J. (1930). "Turbulente Strömung in nichtkreisförmigen Rohren". *Ing.-Arch.* 1(306).
46. PETRYCK, S. and BOSMAJIAN, G. III, (1975). "Analysis of Flow Through Vegetation". *J. Hydraul. Div.* 101(HY7), 871–884.
47. REE, W.O. and CROW, F.R. (1977). "Friction Factors for Vegetated Waterways of Small Slope". *ARS-S-151*, Agricultural Research Service, U.S. Department of Agriculture, Stillwater, Oklahoma, USA.
48. RIGHETTI, M. and ARMANINI, A. (2002). "Flow Resistance in Open Channel Flows with Sparsely Distributed Bushes". *J. Hydrol.* 269(1–2), 55–64.
49. SHIMIZU, Y. and TSUJIMOTO, T. (1994). "Numerical Analysis of Turbulent Open-channel Flow over Vegetation Layer using a k–ε Turbulence Model". *J. Hydrosci. Hydraul. Engng.* 11(2), 57–67.
50. STEPHAN, U. and GUTKNECHT, D. (2002). "Hydraulic Resistance of Submerged Flexible Vegetation". *J. Hydrol.* 269(1–2), 27–43.
51. STONE, B.M. and SHEN, H.T. (2002). "Hydraulic Resistance of Flow in Channels with Cylindrical Roughness". *J. Hydraul. Engng.* 128(5), 500–506.
52. TAYLOR, R.P., COLEMAN, H.W. and HODGE, B.K. (1985). "Prediction of Turbulent Rough-wall Skin Friction using a Discrete Element Approach". *J. Fluids Engng.* 107, 251–257.
53. THOM, A.S. (1971). "Momentum Absorption by Vegetation". *Q.J.R. Meteorol. Soc.* 97, 414–428.
54. TSUJIMOTO, T. and KITAMURA, T. (1990). "Velocity Profile of Flow in Vegetated-bed Channels". KHL Progressive Report 1, Hydraulic Laboratory, Kanazawa University.
55. TSUJIMOTO, T., OKADA, T. and KONTANI, K. (1993). "Turbulent Structure of Open-channel Flow Over Flexible Vegetation". KHL Progressive Report 4, Hydraulic Laboratory, Kanazawa University.
56. UITTENBOGAARD, R. (2003). "Modelling Turbulence in Vegetated Aquatic Flows". *International Workshop on RIParian FORest Vegetated Channels: Hydraulic, Morphological and Ecological Aspects*, Trento, Italy, 20–22 February 2003.
57. VAN VELZEN, E.H., JESSE, P., CORNELISSEN, P. and COOPS, H. (2003). "*Hydraulic Resistance of Vegetation in Floodplains, Part 2: Background Document Version 1-2003*". Ministry of Transport, Public Works and Water Management, Institute for Inland Water Management and Waste Water Treatment, report 2003.029, Lelystad, The Netherlands (in Dutch).
58. VERNON-HARCOURT, L.F. (1896). *Rivers and Canals, Vol. 1 Rivers*. Clarendon Press, Oxford, 312 pp.
59. WILSON, C.A.M.E., STOESSER, T., BATES, P.D. and BATEMANN PINZEN, A. (2003). "Open Channel Flow through Different Forms of Submerged Flexible Vegetation". *J. Hydraul. Engng*. 129(11), 847–853.