A Bayesian Analysis of Fluid Flow in Pipelines

Jonathan Rougier* Michael Goldstein

Department of Mathematical Sciences

University of Durham, UK

Abstract

The waterhammer equations are a pair of partial differential equations that describe the behaviour of an incompressible fluid in a pipeline. We generalise these equations to account for uncertainty in (1) the description of the liquid and the pipeline, (2) the behaviour of the pipeline boundaries, and (3) the method of solution. We illustrate applications of our model to pipeline design and to real-time pipeline monitoring, e.g. for leak-detection, and discuss the general features of our approach to the careful sourcing of uncertainty in deterministic models.

Keywords: Random field; Dynamic Linear Model; Real-time Leak Detection; Waterhammer.

1 Introduction

The analysis of fluid flow in pipelines is typically conducted by a deterministic analysis that takes as known the parameters that describe the interaction of the fluid and the pipe. This analysis is used both on-line, to model the behaviour of pipelines in order to detect leaks, and off-line, to examine the response of the pipeline to sudden changes in operating conditions. In either case an incomplete

*Corresponding author: Department of Mathematical Sciences, University of Durham, South Road, Durham DH1 3LE, UK, tel +44 (0)191 374 2361, fax +44 (0)191 374 7388, e-mail J.C.Rougier@durham.ac.uk. This work was supported by EPSRC grant ref. GR/L10031.

or inappropriate analysis raises the probability of pipeline failure with, potentially, human casualties and environmental catastrophe.

What is not known, except through experiment in certain special cases, is the accuracy of the deterministic analysis. There are three sources of uncertainty. First, the coefficients that describe the pipeline are uncertain, and also non-constant in both space and time. Second, the boundaries are uncertain. These boundaries comprise the initial state of the pipeline, and the behaviour of the upstream and downstream terminals. Third, the method of solution introduces truncation errors. In practice these three uncertainties are agglomerated into a margin-for-error. This is far from ideal because a simple margin fails to respect the varying sources of uncertainty, and the way in which they interact, both separately and together, with the physics of fluid flow.

In this paper we show how these uncertainties may be included by generalising the deterministic analysis, and embedding the resulting stochastic formulation within a dynamic model evolving through time. This dyanamic model, suitably linearised, provides a framework for prior analysis of uncertainty, i.e. as required in pipeline design, and also for real-time pipeline monitoring, using Bayesian learning from noisy data collected from the pipeline through time. While our analysis is specific to pipelines, it is suggestive of a more general approach to physical modelling, in which uncertainty is carefully sourced within a system of PDEs, and then carried through into a discrete-time solution amenable to a Bayesian analysis. We discuss this further in the final section of the paper.

2 The waterhammer equations

Wylie and Streeter (1993) gives detailed descriptions of modelling fluid flow in pipelines, while Massey (1989) provides an outline within the context of fluid mechanics. A model of fluid flow in a pipeline relates two state variables, pressure and flow, over two spatial variables, one measuring the distance along the pipeline, the other, time. We use h' for pressure (metres of piezometric head), v' for mean flow (metres per second), x' for distance along the pipeline from the upstream end (metres), and t' for time (seconds). The primes denote quantities on their original scale, as we will shortly be switching to a more convenient 'unitless' scale. These scalings are described in Section 5.

The two equations that describe the relation between pressure and flow in space and time are known as the *waterhammer equations*, and they are applicable for low mach-number flow of any incompressible fluid. The waterhammer equations form a quasi-linear hyperbolic PDE system in h' and v', which we give in their dimensionless form (e.g., Liou and Tian, 1995; Basha and Kassab, 1996), i.e. with state and spatial variables suitably re-scaled and primes dropped,

$$\frac{\partial h}{\partial x} + \frac{\partial v}{\partial t} + \xi\, v|v| = 0, \qquad \frac{\partial h}{\partial t} + \frac{\partial v}{\partial x} = 0. \tag{1}$$

The unitless coefficient ξ, often termed the 'pipeline coefficient', can have both x- and t-extent, and describes the frictional damping due to roughness at each location along the pipeline at any given time.

2.1 Solution for pressure and flow

The waterhammer equations may be given an alternative representation using the method of characteristics (Ames, 1965, sec. 3.1, 7.21), as outlined in Appendix A.1. This method gives rise to the following pair of total differential equations,

$$\pm\frac{dh}{dt} + \frac{dv}{dt} + \xi\, v|v| = 0, \qquad \frac{dx}{dt} = \pm 1 \tag{2}$$

where each equation is valid only in the given (x, t) direction. In this representation, equations are solved on the vertices of a diagonal lattice in space and time, where the edges correspond to characteristics. Figure 1 shows a grid of n reaches (where it is convenient to choose n odd) in which the space- and time-steps are of equal size. The diagonal arrows represent the positive (heading downstream) and negative (heading upstream) characteristics, while the vertical arrows at the upstream end represent boundary conditions (to be discussed below).

The pressure and flow at each of the 'child' node can be solved in terms of the

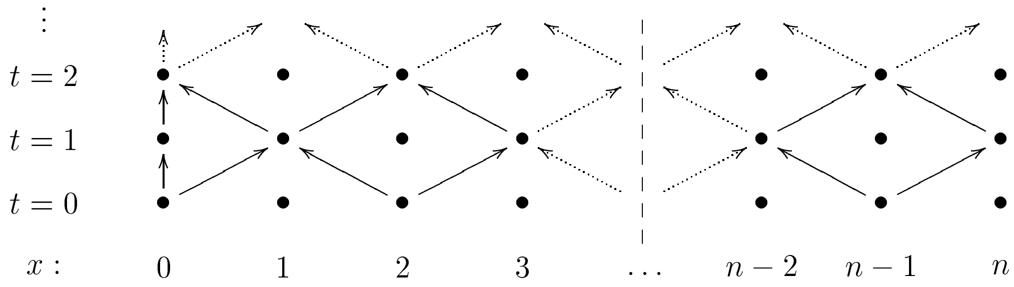


Figure 1: Space and time grid of n reaches, $x = 0$ is upstream.

pressure and flow at the 'parent' nodes. Appendix A.2 shows that the waterhammer equations may be solved for pressure and flow at the interior points of this grid at time t using data available from time $t-1$ as

$$\begin{pmatrix} h_{ti} \\ v_{ti} \end{pmatrix} = \begin{pmatrix} 1 & 1 + |v_{t-1,i-1}|\, z^{+}_{t-1,i-1} \\ -1 & 1 + |v_{t-1,i+1}|\, z^{-}_{t-1,i+1} \end{pmatrix}^{-1} \begin{pmatrix} h_{t-1,i-1} + v_{t-1,i-1} - \epsilon^{+}_{t-1,i-1} \\ -h_{t-1,i+1} + v_{t-1,i+1} - \epsilon^{-}_{t-1,i+1} \end{pmatrix} \tag{3}$$

for t even and $i = 2, 4, \ldots, n-1$; or for t odd and $i = 1, 3, \ldots, n-2$. The '+' and '$-$' superscripts pertain to the positive and negative characteristics, respectively. There are two new terms in this expression. The 'z' terms are line integrals over the pipeline coefficient. They are indexed by the location in space and time of their point of origin, and by their direction, thus

$$z^{+}_{ti} = \int_0^{\Delta t} \xi(i + \tau, t + \tau)\, d\tau,$$

i.e. the value of ξ integrated along the positive characteristic $dx/dt = 1$ from $(i,\, t)$ to $(i + \Delta t,\, t + \Delta t)$. The '$\epsilon$' terms we label 'modelling errors' as they describe the incompleteness of our mathematical representation and solution of the pipeline. A large part of their composition are truncated terms from the line integral approximation, hence they are indexed similarly to the z terms. However, the modelling errors must also capture any imperfections in the waterhammer equations themselves, e.g. the slight compressibility of many liquids that we treat as incompressible (like water).

2.2 Sources of uncertainty

In the standard deterministic formulation the modelling errors, ϵ, are ignored, and the pipeline coefficient ξ is assumed known and, usually, constant. In this case the solution simplifies considerably because

$$z_{ti}^{+} = z_{ti}^{-} = (\Delta t)\, r$$

where $\xi(x,t) = r$ for all x and t. The solution for interior points is now entirely deterministic. In contrast, we retain the ϵ errors and acknowledge uncertainty about the pipeline coefficient ξ. For a pipeline in operation, the value of the pipeline coefficient varies in both space and time in a manner which is generally unobservable and highly dependent upon local conditions, particularly when allowing for the transitory effects induced by fluid-structure interactions (Tijsseling, 1996; Freitas Rachid and Costa Mattos, 1998). While it is possible to provide complex deterministic models of these effects we prefer instead to acknowledge our limited information by modelling the pipeline coefficient ξ as a random field in space and time, with an assigned mean function $r(x,t) = \mathsf{E}\left(\xi(x,t)\right)$ and a covariance kernel

$$\kappa\big((x,t),\, (x',t')\big) = \mathsf{Cov}\left(\xi(x,t),\, \xi(x',t')\right)$$

It follows that z_{it}^{+} and z_{it}^{-} are not known constants, but are themselves random quantities with means, variances and covariances that follow directly from our choice of the functions r and κ:

$$\mathsf{E}\left(z_{ti}^{+}\right) = \int_{0}^{\Delta t} r(i+\tau, t+\tau)\, d\tau$$

$$\mathsf{Cov}\left(z_{ti}^{+},\, z_{t'i'}^{-}\right) = \int_{0}^{\Delta t}\int_{0}^{\Delta t} \kappa\big((i+\tau, t+\tau),\, (i'-\tau', t'+\tau')\big)\, d\tau\, d\tau'$$

(see, e.g., Parzen, 1962).

2.3 Behaviour at the boundary

The waterhammer equations on their own cannot provide solutions for pressure and flow at the two boundaries. This is because the relation between pressure and flow at the boundary will depend upon the equipment that is installed there. For example, the downstream end of a pipeline is often a valve, in which case the pressure and flow at the downstream end are related in the form $v_{tn} \propto \sqrt{h_{tn}}$ where the proportionality depends upon the setting of the valve. This relation can be combined with a positive characteristic from node $n-1$ in order to determine both pressure and flow at node n for alternate time-points (t odd in Figure 1). An alternative is simply to provide a value for one of pressure or flow and then solve for the other, using the available characteristic. For example, at the upstream end of the pipeline there is sometimes a pressure meter. In applications the data from this meter is used directly to compute flow at node 0 (for t even in Figure 1).

Both of these practices introduce errors into the deterministic analysis. The absence of any error in the description of the downstream valve is not consistent with the mainly empirical results that describe the precise relation between pressure and flow at different valve settings, often repesented in graphical form (e.g. Wylie and Streeter, 1993, Appendix C). Likewise, imperfect meters should not have their data fed directly into a deterministic algorithm. Therefore we specify stochastic models at the pipeline boundaries, and leave the assimilation of data to a later stage. At an upstream boundary fed from a reservoir, for example, we might model pressure as a random walk, while at a downstream boundary exiting through a valve we would introduce an error term into the relation between pressure and flow. This is the reason behind the different treatment of the two boundaries in Figure 1. Details of these two models are given in Appendix A.3.

3 Solution as a dynamic model

In this Section we show that the stochastic waterhammer equations described in Section 2 can be represented within a linearised dynamic model. Due to the diagonal lattice, the state equations of the dynamic model must vary according to whether t

is even or odd. The locations of the nodes on which pressure and flow are carried forwards will 'zig-zag' through time, as should be clear from Figure 1.

The state vector itself must contain more then just pressures and flows. The solution of the waterhammer equations, eq. (3), indicates that the z variables, which represent the pipeline friction term, must also be included, as the solution for pressure and flow at time t is conditional upon the values of z_{t-1} which are, in general, unknown. If we assume that ξ is a gaussian random field, the collection $\{z_{t-1},\, z_t\}$ is jointly gaussian and we can use the standard linear updating rule to compute the mean and variance of z_t given z_{t-1}:

$$z_t = a_t + A_t\, z_{t-1} + \zeta_{t-1}$$

where $A_t = \mathsf{Cov}\,(z_t,\, z_{t-1})\,\mathsf{Var}\,(z_{t-1})^{-1}$, $a_t = \mathsf{E}\,(z_t) - A_t\,\mathsf{E}\,(z_{t-1})$, and ζ_{t-1} is a mean zero error term with $\mathsf{Var}\,(\zeta_{t-1}) = \mathsf{Var}\,(z_t) - A_t\,\mathsf{Cov}\,(z_{t-1},\, z_t)$. The error term ζ_{t-1} is written lagged in keeping with our convention that variables defined over time are labelled by their time of origin. As an alternative to modelling ξ as a gaussian random field, the pipeline engineers may prefer to state their beliefs directly as a covariance structure on the collection $\{z_{t-1},\, z_t\}$, in which case the Bayes linear approach (Farrow and Goldstein, 1993; Goldstein, 1999) provides another justification for these relations. Section 4 provides more details about modelling the evolution of z_t.

Technically the ϵ terms should also be included among the state vectors for the same reasons as the z terms. However at this point we make the simplifying assumption that these errors can be treated as independent in time with mean zero and state-invariant variance structure. In Appendix A.2 it is shown that $\epsilon_{t-1,i} \sim O\big((\Delta v_{ti})^3\big)$, where Δv_{ti} is the change in flow along the appropriate characteristic, hence $\epsilon_{t-1,i} \sim O(n^{-3})$. Even in the worst case in which all n components are perfectly correlated the ignored structure in ϵ is only $O(n^{-2})$, and therefore this simplification is reasonable except when n must be small. Where this is unavoidable the ϵ terms can be modelled in the same manner as the z terms.

Collecting the variables h_t and v_t together into the state vector θ_t, the dynamic model for θ may be written

$$\theta_t = g_t(\theta_{t-1},\, z_{t-1},\, \epsilon_{t-1}). \tag{4}$$

Linearising (4) around $\mathsf{E}\,(\theta_{t-1}) = \bar{\theta}_{t-1}$, $\mathsf{E}\,(z_{t-1}) = \bar{z}_{t-1}$ and $\mathsf{E}\,(\epsilon_{t-1}) = \mathbf{0}$ allows us to write the full fluid flow model, i.e. including the evolution of beliefs about ξ, approximately as a Dynamic Linear Model (West and Harrison, 1997, ch. 13),

$$\begin{pmatrix} \theta_t \\ z_t \end{pmatrix} = \begin{pmatrix} f_t \\ a_t \end{pmatrix} + \begin{pmatrix} G_t^\theta & G_t^z \\ \mathbf{0} & A_t \end{pmatrix} \begin{pmatrix} \theta_{t-1} \\ z_{t-1} \end{pmatrix} + \begin{pmatrix} \omega_{t-1} \\ \zeta_{t-1} \end{pmatrix} \tag{5}$$

where $\left(G_t^\theta,\, G_t^z,\, G_t^\epsilon\right)$ is the jacobian matrix of g_t with respect to $(\theta,\, z,\, \epsilon)$ evaluated at $(\bar{\theta}_{t-1}, \bar{z}_{t-1}, \mathbf{0})$,

$$f_t = g_t(\bar{\theta}_{t-1},\, \bar{z}_{t-1},\, \mathbf{0}) - \begin{pmatrix} G_t^\theta & G_t^z \end{pmatrix} \begin{pmatrix} \bar{\theta}_{t-1} \\ \bar{z}_t \end{pmatrix},$$

and $\mathsf{Var}\,(\omega_t) = G_t^\epsilon\, \mathsf{Var}\,(\epsilon_{t-1})\, (G_t^\epsilon)^T$.

In order to assess the impact of linearisation we consider the effect of the non-linear terms of the form $|v_{ti}|z_{ti}$ in (3). These terms will be small, giving rise to an effectively linear system, whenever $\bar{z}_{ti} \approx (\Delta t)\,\bar{\xi}$ is small (where $\bar{\xi}$ is the mean pipeline coefficient, discussed in Section 4). As $\Delta t = 1/n$ we can be confident in our linearisation if we can afford to use a large number of nodes or if we have a smooth pipe. Under these conditions we would make little or no modification to the size of the variance of ω_{t-1}. Otherwise we may choose to increase this variance in order to account for the truncation error of the linear approximation. In extreme circumstances, e.g. a very long and very rough pipeline, it may be necessary to abandon linearisation altogether and switch to a real-time simulation-based approach (as outlined, for example, by Liu and Chen, 1998).

What we achieve in writing the waterhammer equations as a dynamic linear model is to introduce uncertainty explicitly into the modelling of fluid flow in pipelines. While it would be possible simply to add an error term to the deterministic solution, which is in effect what often happens in practice, this ignores the intricate space-time structure that is implied by the physical model. By introducing the different sources of uncertainty at their appropriate points in the development of the model we allow this uncertainty to propogate through time and space in a manner consistent with the physical process. The consequence of this will be most clearly seen in the correlation structure of the state vector. It is a reasonable simplification to say that information passes through the pipeline along the characteristics. Therefore the diagonal lattice in Figure 1 reflects the correlation structure of data at the nodes over space and time. In the absence of any data, the dynamic linear model will update the *mean* of the state vector in exactly the same way as the deterministic approach. Therefore the function of the model is to propagate the *variance* of the state vector in a way that is consistent with the physics of fluid flow in pipelines. This is illustrated in Section 5. In the presence of data, possibly taken with error, the stochastic model allows the updating of beliefs about the state vector using the data values. This includes not just updated values for pressure and flow at the nodes along the pipeline, but also about the pipeline coefficient, ξ, represented in the z terms. This updating follows the standard procedure for dynamic linear models (e.g., West and Harrison, 1997), and is illustrated in Section 6.

4 Modelling the pipeline coefficient

Using the mean function and covariance kernel of ξ, we can compute the conditional mean and variance of $z_t \mid z_{t-1}$ for all t. In general this will require $3n(n+1)/2$ numerical integrations at each time-step (a new mean vector, a new variance matrix and a covariance matrix). To have to repeat this every time-step with large n would defeat the object of being able to compute beliefs about the fluid state in real-time.

We propose a model for the mean and covariance kernel of ξ which is sufficiently general to allow the pipeline engineers some control over the time-extent of the pipeline coefficient, while still being tractable for real-time monitoring:

$$
\begin{aligned}
r(x, t) &= \mu_t \, r(x) \\
\kappa\big((x, t),\, (x', t')\big) &= \sigma_t^2 \, \kappa_1\big(x, x'\big) \, \kappa_2\big(|t - t'|\big),
\end{aligned}
$$

where $\kappa_1(\cdot, \cdot)$ and $\kappa_2(\cdot)$ are known correlation functions, and, for our purposes, t' is either t or $t + 1$. The functions μ_t and σ_t^2 may be taken as known constants for the duration of a single timestep, although in practice they are slowly varying in time. In these two models the x-extent has been favoured at the expense of the t-extent. This is on account of the much greater information available to the pipeline engineer concerning the nature of the pipeline *along* the pipeline. The temporal belief structure $\sigma_t^2 \, \kappa_2(\cdot)$ is taken to be effectively stationary over short periods of time, and therefore the time-evolution of z_t depends only upon whether t is even or odd. Consequently, the necessary numerical calculations need be performed once on initialisation, and then periodically rescaled to account for drift in μ_t and σ_t.

Operationally, it is common to refer to 'the pipeline coefficient' as a single aggregate quantity,

$$
\bar{\xi}_t = \int_0^1 \xi(x, t) \, dx,
$$

where the pipeline length is re-scaled to 1. For example, Liou *et al.* (1992) document the coefficients of 209 pipelines (which typically take values in the range 0 to 10). For this purpose it is helpful to normalize $r(x)$ so that $\int_0^1 r(x) \, dx = 1$, implying $\mathsf{E}\left(\bar{\xi}_t\right) = \mu_t$, the expected pipeline coefficient at time t. Similarly, in order to make the quantity σ_t^2 a direct measure of uncertainty about the pipeline coefficient at time t we also normalize κ_1 so that $\mathsf{Var}\left(\bar{\xi}_t\right) = \sigma_t^2$.

5 Example: Valve transients

A major use of pipeline flow modelling is to calculate the pressure response along the pipeline that is induced by specific boundary conditions, most notably valve closure. Rapid closure of a valve, for example to isolate a faulty component, causes a pressure wave to travel back up the pipeline. These calculations take place during the design of the pipeline, in order to determine the thickness and composition of a pipeline wall of sufficient strength to withstand the rapid changes in pressure, and also in real-time pipeline monitoring. To date the calculations have been performed using a deterministic analysis, perhaps using several alternative pipeline coefficient values. Our approach allows the deterministic analysis to be supplanted by a stochastic analysis in which uncertainties about the pipeline model and the pipeline coefficient translate directly into uncertainty about the pressure and flow at every point in space and time.

5.1 Pipeline description

We reconsider the experimental data of Simpson (1986), discussed in Wylie and Streeter (1993), p. 50. In this experiment a initially-open valve at the end of a copper pipe is rapidly closed, and the resulting pressure behaviour at the valve is monitored. The pipeline is $l = 36$ m long with diameter $d = 19.05 \times 10^{-3}$ m and friction factor $\lambda = 0.0325$. Denoting natural units (as opposed to the re-scaled units used in the differential equations) with primes, a pressure head of $h'_0 = 24.18$ m gives rise to a steady state flow rate of $v'_0 = 0.239$ m/s. The fully-open loss coefficient of the valve is $\beta'_0 = 8,235$. These values taken together satisfy the Darcy equation for steady state flow

$$h'_0 = \left(\frac{\lambda l}{d} + \beta'_0 \right) \frac{(v'_0)^2}{2g},$$

where g is gravitational acceleration, taken to be 9.8 m/s. The rescaling of the state and spatial variables takes the form

$$h = \frac{h' g}{a\, v'_0} \qquad v = \frac{v'}{v'_0} \qquad x = \frac{x'}{l} \qquad t = \frac{a\, t'}{l}$$

where a is the speed of sound in the pipeline, in this case 1280 m/s. On re-arrangment of the Darcy relation in re-scaled terms, the general steady state relation

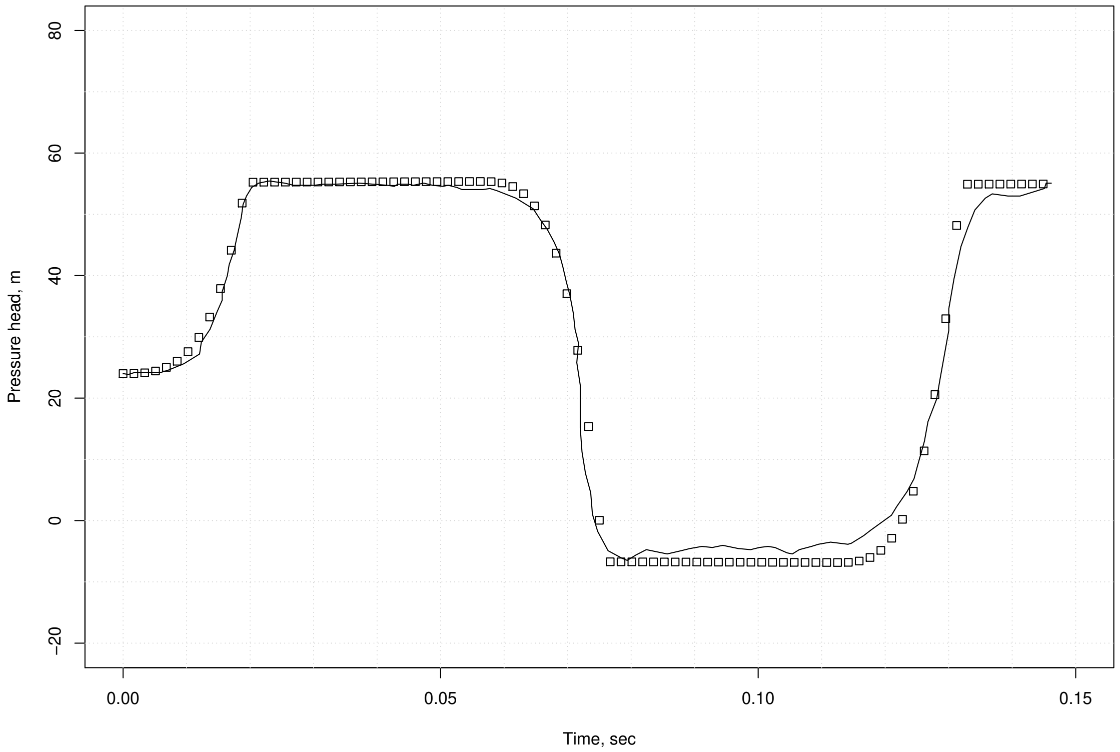


Figure 2: Measured head-time curve (solid line) with deterministic simulator values (squares).

for fixed upstream head h'_0 and variable valve coefficient β'_t is

$$h_0 = (\xi + \beta_t)\, v_t^2 \qquad \text{where} \quad \xi = \frac{\lambda l}{d} \frac{v'_0}{2a}, \quad \beta_t = \beta'_t \frac{v'_0}{2a}.$$

In this expression ξ is the 'pipeline coefficient'. For the experimental pipeline, $\xi = 5.73 \times 10^{-3}$ and $\beta_0 = 0.769$. The fact that $\xi \ll \beta_0$ indicates that, even when fully open, friction losses at the downstream valve dominate friction losses generated by the roughness of the pipe wall, which is not surprising in this very short pipeline. In this example we expect that the major source of uncertainty will be the behaviour of the boundaries, rather than the pipeline coefficient.

Figure 2 shows the pressure (m) just above the downstream valve during and after that valve's closure, as measured and as computed using the deterministic waterhammer equations based on the above parameter values. This Figure corresponds to Figure 3-7 in Wylie and Streeter (1993).

In the following Subsections we introduce the two sources of uncertainty sequentially, in order to contrast the deterministic and the stochastic results. We start with uncertainty in the pipeline coefficient and then we add modelling errors, in each cases introducing an appropriate amount of uncertainty into the prior description of pressure and flow. In Section 6 we use the full structure of the model to update our beliefs about the state vector of pressures, flows and pipeline coefficient, using the data as it becomes available. This will demonstrate the more general application of our approach to real-time pipeline monitoring.

5.2 Uncertainty about the pipeline coefficient

The pipeline coefficient is assumed to be unknown, but constant in space and time, and we assume initially that there are no modelling errors. We set $r(x) = \kappa_1(x, x') = \kappa_2(|t - t'|) = 1$ to make the pipeline coefficient constant in space and time. The updating equation for z_t is simply $z_t = z_{t-1}$, i.e. $a_t = \mathbf{0}$, $A_t = I$ and $\mathsf{Var}\,(\zeta_{t-1}) = \mathbf{0}$. Initial beliefs about the pipeline coefficient are therefore propagated through time with no modification (in the absence of data). These initial beliefs are stated in terms of the 'mean' pipeline coefficient, $\bar{\xi}$, which has mean μ and variance σ^2. This gives initial beliefs $\mathsf{E}\,(z_0) = \mu\,(\Delta t)\,\mathbf{1}$ and $\mathsf{Var}\,(z_0) = \sigma^2\,(\Delta t)^2\,\mathbf{1}\mathbf{1}^T$, where $\Delta t = 1/n$. We use the same discretisation as Simpson, implying $n = 33$. We set $\mu = 5.73 \times 10^{-3}$ and $\sigma = \mu/2$. We do not show the results, as this pipeline is not long enough or rough enough for uncertainty about the effects of friction alone to have any discernable effect on the scale of the mean head-time curve: at time $t = 140$ ms the standard deviation of downstream pressure due solely to coefficient uncertainty is 0.008 m.

5.3 Modelling errors

Now we additionally introduce modelling errors, ϵ_t. These are mainly truncation errors from numerical intregrations, but there are also components that represent boundary model errors (Appendix A.3 gives the details). Our choices for the size of the errors are made on the scale of the mean head-time curve. We set the truncation errors to have standard deviations of 0.1 m and correlate them along the pipeline using a gaussian covariance kernel with a decay rate that implies a correlation of 0.5 between the two ends. The upstream boundary is modelled as a random walk in

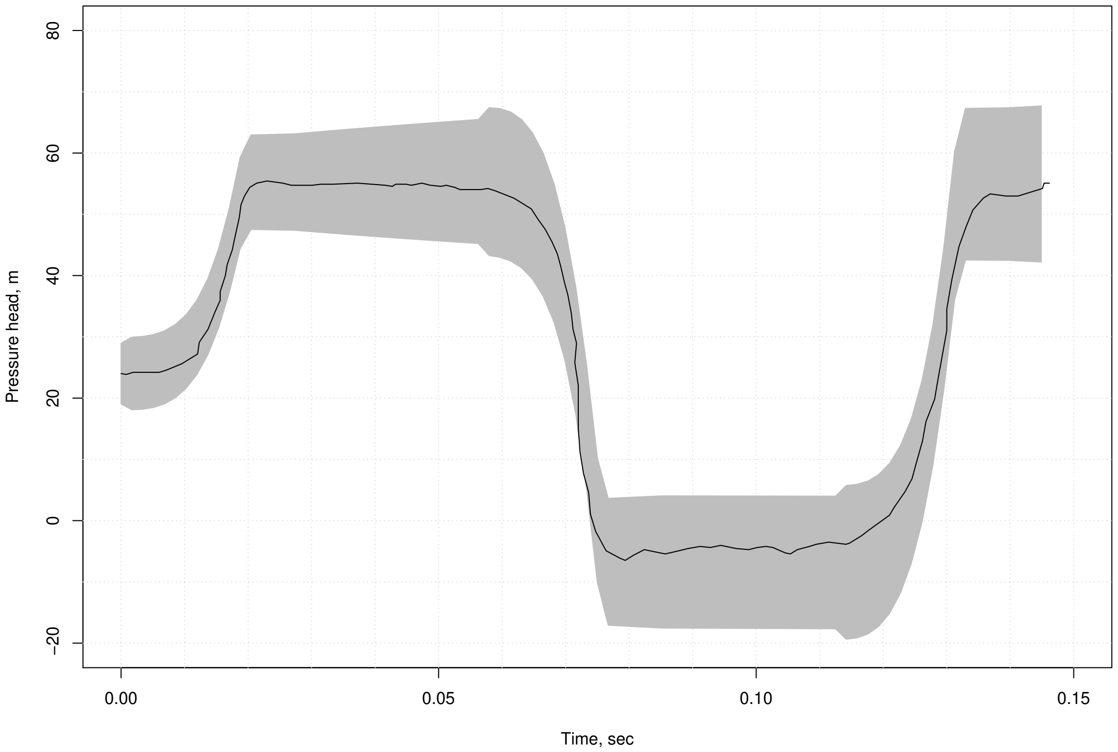


Figure 3: Measured head-time curve (solid line) with ± 2 s.d. forecast interval made at time $t = 0$ (shaded) implied by uncertainty about the pipeline coefficient and the presence of modelling errors.

pressure, and we set the per-period standard deviation so that after one second of random walking from a known starting point the standard deviation of pressure is 1.0 m. The downstream boundary is a valve condition for which we use a standard deviation of 5 m (once the valve has closed this becomes zero). The boundary errors are not correlated with each other or with the truncation errors. In order to set the prior variance for the pressures and flows we iterate forward 100 timesteps from an initial variance of $\mathbf{0}$, with the two boundary error variances set to one third of the above values. This temporary reduction in the boundary errors is made because prior to the experiment the boundaries are quiescent.

The impact of adding modelling errors is shown in Figure 3, while Figure 4 provides more detail about the uncertainty. In Figure 3 the data is shown as a line, with a shaded polygon showing the ± 2 s.d. forecast interval for downstream pressure made at time $t = 0$ (the mean head-time curve is that of the deterministic analysis, as shown in Figure 2). The two 'pulses' in the forecast interval that occur between 50 ms and 60 ms and between 110 ms and 120 ms are an interesting demonstration

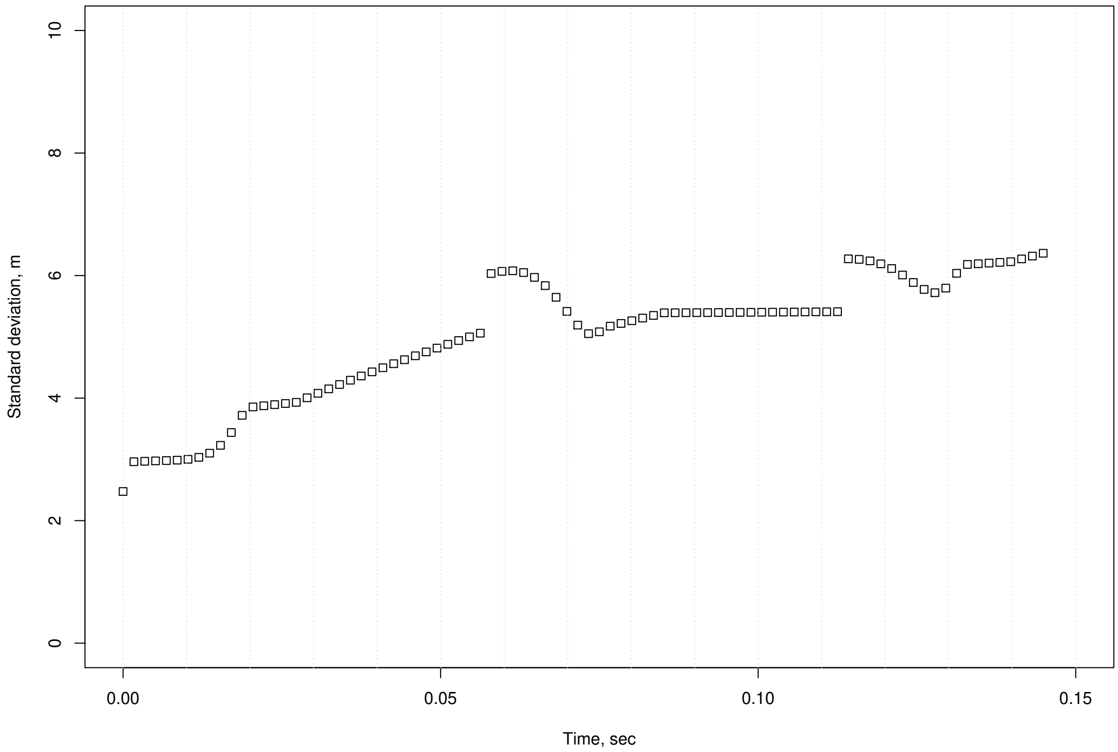


Figure 4: Standard deviation of downstream pressure implied by uncertainty about the pipeline coefficient and the presence of modelling errors, forecast at time $t = 0$.

of the interaction between uncertainty and the physical model. At time $t = 0$ when the valve starts to close we increase the uncertainty on both upstream and downstream boundary models to their chosen values. At just after $t = 50$ ms the extra uncertainty from the upstream boundary arrives at the downstream boundary, while at just after $t = 110$ ms the extra uncertainty from the downstream boundary returns, having been reflected from the upstream boundary. Thus we have 'waves' of uncertainty that behave in many respects like pressure waves. We can see them clearly in this example because the pipeline is so short: with a longer pipeline these effects would be diluted by additional uncertainty accumulated along the way.

6 Bayesian learning

So far our data have been used simply to evaluate the model in the light of its various uncertainties. For issues of design there are no data available, and the pipeline engineer uses the prior uncertainty engendered by an incomplete knowledge of the pipeline to determine likely ranges for pressure and flow under various types of operating condition, as described above. The other use of pipeline models is in real-time monitoring. In this case we would like to use our data, collected at specified locations in space and time, to refine our beliefs about the pressure and flow at all points along the pipeline. We would also like to learn about the pipeline coefficient, as embodied in z_t, a process that is sometimes described as 'tuning' the model to the pipeline. It is not possible to learn very much about the pipeline coefficient in our particular pipeline, because it is very short. Therefore we demonstrate learning about pressure and about flow.

6.1 Downstream pressure

First, we illustrate learning about downstream pressure, which is also the source of our data. We sample the data at an interval of 10 ms. We assume that this time separation is sufficiently large that we may ignore any correlation in measurement errors. The measurement error is assumed to have a standard deviation of 1 m. The results are shown in Figure 5, where the shaded forecast interval, using the latest available data, may be compared with the prior forecast interval made at time $t = 0$, indicated by dashed lines.

Figure 5 provides another illustration of the interaction of uncertainty and the physical model. At the point of data collection the variance of downstream pressure is reduced to at most meter precision. The rate at which uncertainty increases following each datum depends in part on the boundary model. Information passes along the diagonal characteristics: it is specifically *not* the case that information flows 'up' the boundaries through time, except insofar as the boundary models themselves permit. The downstream boundary model imposes a simultaneous relationship on pressure and flow, and so does not directly transmit the information supplied by each datum 'up' the boundary.

A second feature of interest in Figure 5 is the returning wave. In the first low pressure phase a second set of information is superimposed upon the downstream pressure (seen as a second set of 'notches'). This is a result of the early data, which have travelled up the pipeline to be reflected at the upstream end. Were we to

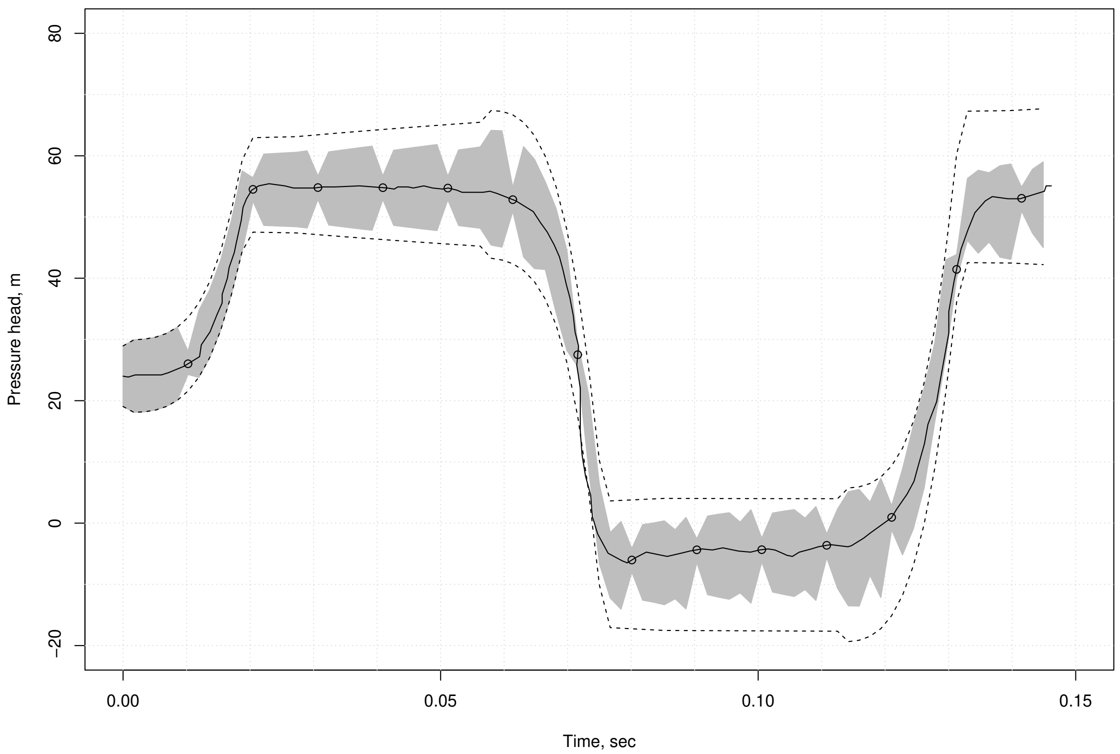


Figure 5: Measured head-time curve (solid line, with sampled data circled) with ± 2 s.d. forecast interval (shaded) using the latest available data. The dashed lines indicate the forecast interval at time $t = 0$ (as shown in Figure 3). The 'v' shape of each notch is an artefact of the discrete time-steps.

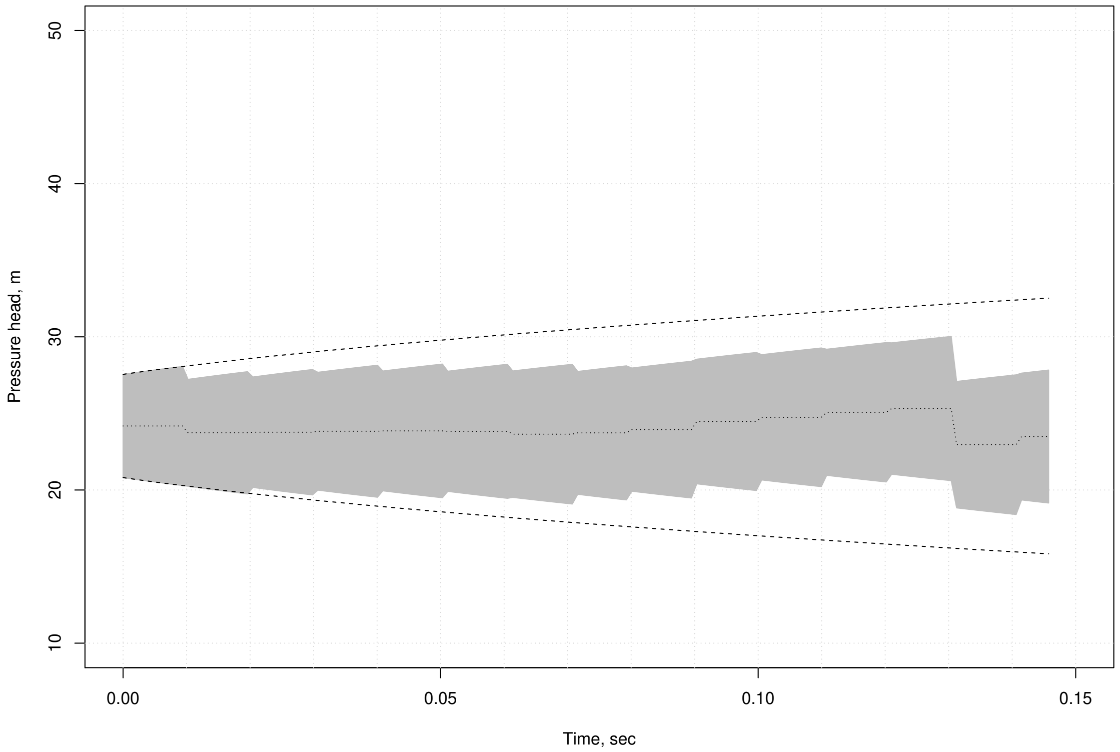


Figure 6: Upstream pressure, as a ± 2 s.d. forecast interval using the latest available data (mean as dotted line). The dashed lines indicate the forecast interval at time $t = 0$.

prolong the simulation we would see further superimpositions of reflected information which, when taken together, would gradually reduce the aggregate uncertainty about downstream pressure.

6.2 Upstream pressure and flow

We also illustrate learning about quantities away from the data. The forecast for upstream pressure is shown in Figure 6. The contemporaneous upstream and downstream pressures are reasonably correlated (0.55 at $t = 0$), and therefore each datum has an immediate impact. As we model the upstream boundary as a random walk in pressure, this information can be passed 'up' the boundary as well as diagonally along the characteristics, unlike at the downstream end. Consequently the reduction in uncertainty that follows the aquisition of new pressure datum is passed forwards to later time-points. This explains the different shape of the intervals in Figures 5 and 6.

The upstream flow rate is shown in Figure 7. There is less contemporaneous

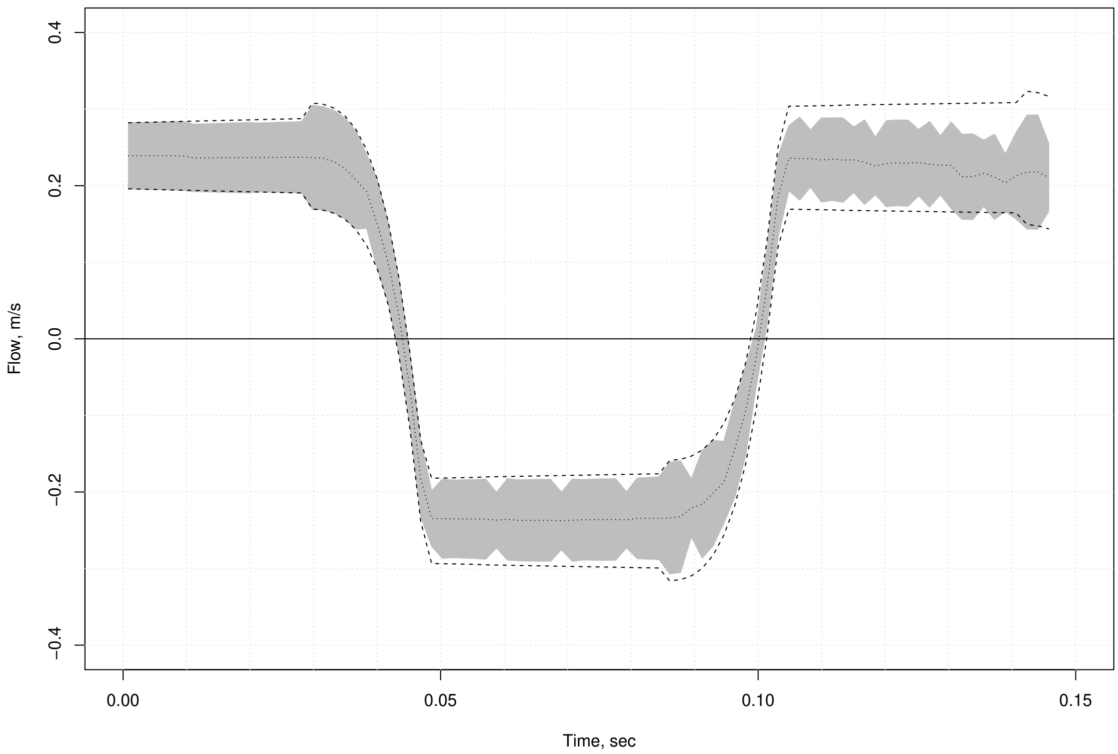


Figure 7: Upstream flow, as a ± 2 s.d. forecast interval using the latest available data (mean as dotted line). The dashed lines indicate the forecast interval at time $t = 0$.

correlation between upstream flow and downstream pressure (0.34 at $t = 0$), with the result that there is little immediate impact from each datum. We tend to learn about upstream flow with a time-lag of fixed length that represents the propogation of information from the downstream valve up the pipeline to the upstream reservoir. Furthermore there is no direct mechanism for propagating flow information 'up' the upstream boundary. As with downstream pressure, the superimposition of the returning wave of uncertainty (having made three trips along the pipeline) is clear in the second high flow phase.

7 Discussion

Although this paper concerns stochastic modelling for pipelines, there are general features of our approach which are of wider applicability. We have turned a deterministic physical model into a stochastic dynamic linear model, in order that we might use this model for real-time Bayesian learning. In so doing, we have taken particular care to source our uncertainty correctly, in the following steps:

1. We represent parameter uncertainty as a random field in the spatial variables, by specifying a mean function and a covariance kernel.

2. We solve the deterministic form of the physical model for a discrete (in space and time) representation in the presence of parameter uncertainty. Thus we create discrete random quantities with mean and covariance structures that follow from our uncertainty about the parameters interacting with the physical process.

3. We provide stochastic models for the evolving boundary conditions. In this way we create a stochastic dynamic model for the physical process.

4. We linearise the dynamic model. The result is a stochastic representation in which the mean of the state vector will evolve in exactly the same way as in the original determinstic model. The difference is that in our model we can also evolve a variance for the state vector from an initial variance.

5. We incorporate the data by updating beliefs about the state vector with due allowance for measurement errors.

In our example we have given several instances of the highly-structured interaction that can exist between parameter and boundary uncertainty and the physical model. Whether it is worth the computational effort required to implement our general approach in specific cases depends upon the importance of obtaining accurate assessments of uncertainty for each case.

We conclude by outlining the benefits of our stochastic approach for the real-time modelling of pipelines in the context of a typical use of pipeline models, namely leak detection. The usual arrangement for such problems is to have one meter providing boundary data for a deterministic model, and other meters providing reference data. If the reference data and the prediction of the deterministic model disagree, then it is possible that a non-simulator event such as a leak has occurred. Unfortunately arrangements of this type provide little guidance about how big a discrepancy is tolerable, resulting in the development of complicated 'voting algorithms' that attempt to pool discrepancies across meters and across time (see, e.g., Mears, 1993; Liou and Tian, 1995).

Our stochastic approach offers the following important advantages. First, there is no distinction between boundary meters and reference meters, and all their data can be used both for diagnosis and for learning, so that, for example, it is possible to have a leak detection system that uses just a single meter. Second, due allowance is made for meter imprecision, avoiding the dubious practice of feeding inaccurate data directly into a deterministic system. Third, our approach is 'self-tuning'. In long pipelines, in which the pipeline coefficient is an important determinant of fluid behaviour, it is possible to learn about its value and to permit this value to evolve spatially. Finally, probability itself can be used as a metric when evaluating the data. The stochastic simulator we have developed here is part of the likelihood function when expressed over a parameter space describing the number of leaks, and, for each leak, the leak's location, size and and time of occurence (generalising the simulator to include leaks is quite straightforward). The role of the simulator as likelihood function paves the way for a fully Bayesian analysis of leak detection, in which it is possible for experts to incorporate detailed beliefs about the 'leakiness' of the pipeline, and, given data from the meters, to derive probabilistic descriptions of the leak or leaks that might have occurred.

A Appendix

A.1 Method of characteristics

The characteristics representation is a standard approach for solving the waterhammer equations. We start with the PDEs

$$h_x + v_t + \xi\, v|v| = 0 \qquad h_t + v_x = 0$$

as given in (1), where, just for this Subsection, we use subscripts to denote partial derivatives. Now consider some scalar quantity γ, and the combined equation

$$\Big(h_x + v_t + \xi\, v|v|\Big) + \gamma\Big(h_t + v_x\Big) = \gamma\Big(\frac{1}{\gamma}\, h_x + h_t\Big) + \Big(\gamma\, v_x + v_t\Big) + \xi\, v|v|$$

which must equal *zero* for all values of γ. We will be able to replace the parenthetical expression in v with a total derivative if and only if

$$\frac{dv}{dt} = v_x \frac{dx}{dt} + v_t = \gamma\, v_x + v_t$$

which implies that we require $\gamma = dx/dt$. Likewise, we will be able to replace the parenthetical expression in h with a total derivative if and only if

$$\frac{dh}{dt} = h_x \frac{dx}{dt} + h_t = \frac{1}{\gamma}\, h_x + h_t$$

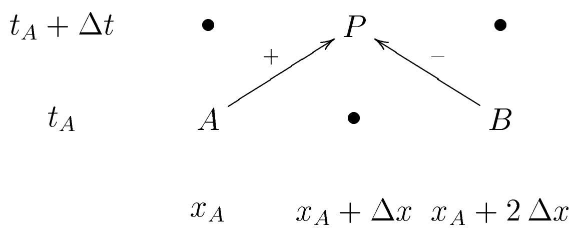
which implies that we require $1/\gamma = dx/dt$. It follows that $\gamma = dx/dt = \pm 1$ is the only solution that allows both expressions to be replaced with total derivatives. By substituting $\gamma = \pm 1$ and simultaneously enforcing the constraint $dx/dt = \pm 1$ we derive the characteristics representation

$$\pm\frac{dh}{dt} + \frac{dv}{dt} + \xi\, v|v| = 0, \qquad \frac{dx}{dt} = \pm 1$$

as given in (2).

A.2 Solving for pressure and flow

The solution method outlined here is a generalisation of Wylie and Streeter (1993), pp. 40-1. Consider the arrangement of the points B and P in (x, t)-space relative to point A so that the three points conform to the positive and negative characteristics (not to scale):



for some time increment Δt, such that $t_P = t_A + \Delta t$, $x_P = x_A + \Delta x$, $x_B = x_A + 2\,\Delta x$ and $\Delta x = \Delta t$. Integrating along the positive characteristic from A to P gives the relation

$$(h_P - h_A) + (v_P - v_A) + \int_A^P \xi\, v|v|\, dt = 0.$$

Evaluate the integral by parts to give, writing the indefinite integral of ξ as $z = \int \xi\, dt$,

$$\begin{aligned}\int_A^P \xi\, v|v|\, dt &= v|v|\, z\Big]_A^P - \int_A^P z\, 2|v|\, \frac{dv}{dt}\, dt \\ &= v_P|v_P|\, z_P - v_A|v_A|\, z_A + \big(v_P - v_A\big)\big(z_A|v_A| + z_P|v_P|\big) + \epsilon \\ &= v_A|v_P|\, z_P - v_P|v_A|\, z_A + \epsilon\end{aligned}$$

where ϵ is the $O\big((v_P - v_A)^3\big)$ integration error that follows from applying the trapezium rule using the two endpoints (see, e.g., Davis and Rabinowitz, 1984, ch. 2). Now v_A and v_P are usually close in value (for a sufficiently fine grid), and so the only time they differ in sign is when they are both close to zero. Consequently $v_A|v_P| \approx v_P|v_A|$, and we simplify the integral further, to give the result

$$\int_A^P \xi\, v|v|\, dt = v_P|v_A|\,(z_P - z_A) + \epsilon = v_P|v_A| \int_A^P \xi\, dt + \epsilon.$$

Writing $z_A^+ = \int_A^P \xi\, dt$ and ϵ_A^+ for the corresponding error term, and using the same notation for the negative characteristic from B to P, the the two characteristics equations together give the solution for $(h_P,\, v_P)^T$ in terms of values available at the previous time-step,

$$\begin{pmatrix} h_P \\ v_P \end{pmatrix} = \begin{pmatrix} 1 & 1 + |v_A|\, z_A^+ \\ -1 & 1 + |v_B|\, z_B^- \end{pmatrix}^{-1} \begin{pmatrix} h_A + v_A - \epsilon_A^+ \\ -h_B + v_B - \epsilon_B^- \end{pmatrix}.$$

A.3 Boundary models

There are many different types of boundary model. Here we illustrate the cases of a reservoir at the upstream end of the pipeline and a valve at the downstream end of the pipeline. In both cases we have to distinguish even and odd time-points, due to the availability of an incoming characteristic (see Figure 1). For simplicity, we retain the A, B and P subscripts on the understanding that on the boundaries A, B and P form a rightangled triangle in space and time with the hypotenuse aligned with the incoming characteristic. Errors on the two boundary models are written as ϵ terms without a '+' or '$-$' superscript.

For the upstream boundary when t_P is even, the two relations are a random walk in pressure

$$h_P - h_A + \epsilon_A = 0$$

as a simple reservoir model, plus the negative characteristic,

$$-h_P + h_B + v_P - v_B + v_P\, |v_B|\, z_B^- + \epsilon_B^- = 0$$

Solving these equations gives

$$\begin{pmatrix} h_P \\ v_P \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1 & 1 + |v_B|\, z_B^- \end{pmatrix}^{-1} \begin{pmatrix} h_A - \epsilon_A \\ -h_B + v_B - \epsilon_B^- \end{pmatrix}.$$

For t_P odd, the reservoir model gives us a solution for h_P but the value of v_P is undefined.

For the downstream boundary when t_P is odd, the two relations are a valve model

$$h_P - \beta_P\, v_P^2 + \epsilon_B = 0$$

where β_P is the valve setting at time t_P, plus the positive characteristic

$$h_P - h_A + v_P - v_A + v_P\, |v_A|\, z_A^+ + \epsilon_A^+ = 0.$$

Solving these equations gives

$$\begin{aligned} h_P &= \beta_P\, v_P^2 - \epsilon_B \\ v_P &= \frac{-(1 + |v_A|\, z_A^+) + \sqrt{(1 + |v_A|\, z_A^+)^2 - 4\,\beta_P\, c}}{2\,\beta_P} \end{aligned}$$

where $c = -\epsilon_B - h_A - v_A + \epsilon_A^+$. For t_P even neither pressure h_P nor flow v_P is defined. In the limit as $\beta_P \to \infty$, which represents complete closure, the valve model switches to the simpler $v_P = 0$, from which the head at the valve solves as $h_P = h_A + v_A - \epsilon_A^+$ for t_P odd, with $\epsilon_B = 0$ in this case.

References

W.F. Ames, 1965. *Nonlinear Partial Differential Equations in Engineering*, volume 1. New York: Academic Press.

H.A. Basha and B.G. Kassab, 1996. A perturbation solution to the transient flow problem. *Journal of Hydraulic Research*, **34**, 633–49.

P.J. Davis and P. Rabinowitz, 1984. *Methods of Numerical Integration*. London: Academic Press, 2nd edition.

M. Farrow and M. Goldstein, 1993. Bayes linear methods for grouped multivariate repeated measurement studies with application to crossover trials. *Biometrika*, **80**, 39–59.

F.B. Freitas Rachid and H.S. Costa Mattos, 1998. Modelling of pipeline integrity takng into account the fluid-structure interaction. *International Journal for Numerical Methods in Fluids*, **28**, 337–55.

M. Goldstein, 1999. Bayes linear analysis. In *Encyclopaedia of Statistical Sciences, update vol. 3*, pages 29–34. London: John Wiley & Sons.

J.C.P. Liou, C.G. Brockway, and R.B. Miller. Pipeline variable uncertainties and their effects on leak detectability. In *Proceedings of the API Pipeline Cybernetics Symposium*, pages 127–49. American Petroleum Institute, Houston, 1992.

J.C.P. Liou and J. Tian, 1995. Leak detection—transient flow simulation approaches. *AMSE Journal of Energy Resources Technology*, **117**, 243–8.

J.S. Liu and R. Chen, 1998. Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association*, **93**, 1032–1044.

B.S. Massey, 1989. *Mechanics of Fluids.* London: Chapman and Hall, sixth edition.

M.N. Mears. Real world applications of pipeline leak detection. In *Proceedings of the International Conference on Pipeline Infrastructure*, 2, pages 189–209. ASCE, 1993.

E. Parzen, 1962. *Stochastic Processes.* San Francisco: Holden-Day, Inc.

A.R. Simpson. *Large Water Hammer Pressures Due to Column Separation in Sloping Pipes.* PhD thesis, Univ. of Michigan, 1986.

A.S. Tijsseling, 1996. Fluid-structure interaction in liquid-filled pipe systems: A review. *Journal of Fluids and Structures*, **10**, 109–46.

M. West and J. Harrison, 1997. *Bayesian Forecasting and Dynamic Models.* New York: Springer-Verlag, second edition.

E.B. Wylie and V.L. Streeter, 1993. *Fluid Transients in Systems.* New Jersey: Prentice-Hall, Inc.

List of captions

Figure 1: Space and time grid of n reaches, $x = 0$ is upstream.

Figure 2: Measured head-time curve (solid line) with deterministic simulator values (squares).

Figure 3: Measured head-time curve (solid line) with ± 2 s.d. forecast interval made at time $t = 0$ (shaded) implied by uncertainty about the pipeline coefficient and the presence of modelling errors.

Figure 4: Standard deviation of downstream pressure implied by uncertainty about the pipeline coefficient and the presence of modelling errors, forecast at time $t = 0$.

Figure 5: Measured head-time curve (solid line, with sampled data circled) with ± 2 s.d. forecast interval (shaded) using the latest available data. The dashed lines indicate the forecast interval at time $t = 0$ (as shown in Figure 3). The 'v' shape of each notch is an artefact of the discrete time-steps.

Figure 6: Upstream pressure, as a ± 2 s.d. forecast interval using the latest available data (mean as dotted line). The dashed lines indicate the forecast interval at time $t = 0$.

Figure 7: Upstream flow, as a ± 2 s.d. forecast interval using the latest available data (mean as dotted line). The dashed lines indicate the forecast interval at time $t = 0$.