Improved air valve design using evolutionary polynomial regression

G. Balacco and D. Laucelli

ABSTRACT

Air valves are usually sized by heuristic methods or, sometimes, even oversized. Although the technical literature has long focused on the correct sizing of air valves to reduce the overpressure generated by the filling of a pipe, the phenomenon is complex and does not seem to be representable by physically based equations in an easy way, to be of practical use for technicians and designers. In this paper, air valve design is approached through an alternative data-modelling approach, based on evolutionary polynomial regression, with the aim to provide symbolic formulas of variable complexity and accuracy, suitable for physical interpretation, and at the same time easy to be used and applied for design purposes. The present investigation suggests a design formula that, given the geometric parameters of the pipeline system where the air valve is installed, provides the maximum tolerable overpressure, thus allowing the optimal air valve orifice size to be identified.

Key words | air valve, entrapped air, evolutionary polynomial regression, hydraulic transients

INTRODUCTION

Air valves are usually installed at high points of a Water Distribution System (WDS) to allow air to exit or enter during, respectively, pipe filling or emptying and to release air accumulated in the pipeline during normal operations. The importance of these valves is undeniable and their use is essential, but the general belief is that their sizing is simple. Nevertheless, very often the technical information provided by the manufacturers deliberately chooses larger valves (Romer 2017).

AWWA (2001) and even earlier Lescovich (1972) identified three kinds of air valve: (1) air/vacuum valves; (2) air-release valves; and (3) combination air valves. The first one, also called large orifice valves, is designed for large quantities of air during pipeline filling/emptying and to allow the entry of large quantities of air when the pressure drops below the atmospheric pressure. The second one, smaller than the former, is dedicated to smaller air flows, such as due to pressure changing during normal operations. The latter is a combination of the previous two. From this point onwards, all the knowledge and study carried out will refer to air/vacuum valves, generally indicated here as air valves.

Air exit from an air valve plays a key role in the transient following the filling of a pipeline because of the liquid column that follows the air during the process. The liquid column that follows the air stops abruptly due to the difference in density between air and water, when it arrives near the air valve orifice. In order to limit the consequent overpressure, technical practice suggests a very low filling rate.

The problem of the filling process in transient flows into water distribution networks has been widely addressed by De Marchis et al. (2010) through numerical models. This very complex condition to be studied and represented, especially if compared with the problem at stake, even if, in any case, it is important to predict the pressure trend during the filling process. Balacco et al. (2018) experimentally investigated pressure trends by varying orifice diameter, supply pressure and volume of entrapped air in
the descending pipe, during the filling of an initially empty undulating pipeline, showing that such a task is more complicated than it may appear in other similar studies (e.g., Lingireddy et al. 2004).

For small orifices the peak pressure is achieved during the mass oscillation phase; conversely, in the case of larger orifices, a steady-state air-water interface is generated at the high point, then the air is progressively expelled from the orifice, and no significant peak pressure originates after the initial mass oscillation (Balacco et al. 2017). Despite several studies and research (e.g., Albertson & Andrews 1971; Meunier 1980; De Martino et al. 2008; Zhou et al. 2011; Tran 2017), the dynamic behavior of an air valve during its functioning is not clear and there is no definitive method for its optimal sizing.

Many studies have considered an experimental setup consisting of a rising pipe with an air valve at its end (De Martino et al. 2008; Carlos et al. 2011; Zhou et al. 2011) or at the end of a horizontal pipe (Zhou et al. 2002; Lee 2005; Vasconcelos & Wright 2008). Nevertheless, in field installations the pipeline profile is a sequence of ascending and descending sections with air valves at high points. Several practical formulas are suggested by the technical literature and manufacturers (AWWA 2001; Bianchi et al. 2007; Val-Matic 2015), most of them permitting the evaluation of air flow rate and, consequently, the sizing of the air valve orifice, often relying on manufacturers’ tables that relate water flow to maximum overpressure accepted into the pipeline system.

Figure 1 contains charts for the choice of an air valve provided by two different manufacturers; it can easily be verified that with the same air flow rate to evacuate and maximum allowable pressure during the filling (right side of each chart), the suggested air valve diameter to be used for a vent is different. For example, for a flow rate equal to 1,000 m³/h and a maximum overpressure in the filling phase equal to 0.2 bar, the chart of Manufacturer 1 provides an air valve diameter (DN) included between 50–60–65 mm, while the chart of Manufacturer 2 returns a suggested DN 80 mm. This example shows how considering two manufacturers it is impossible to obtain the same air vent diameter and nowadays air vent design is still entrusted to the sensibility of the designer or to the chosen manufacturer.

Based on data coming from some experimental studies (Balacco et al. 2015, 2018; Apollonio et al. 2016) this paper describes the analysis of a wide asset database containing information on overpressure generated during the filling of an initially empty pipeline with varying boundary conditions. The procedure aims to overcome the above-mentioned concerns about the physical modelling of air exit/entry at air valves during transient conditions in

**Figure 1** | Charts for the choice of air valves.
pipelines. For this reason, a well-known data-driven technique, evolutionary polynomial regression (EPR) (Giustolisi & Savic 2009), is here considered to find a symbolic model for supporting the air valve sizing, starting from the available database. The main aim of this study is to provide a general law that, given the geometry of a pipeline system and fixed maximum acceptable overpressure, suggests the optimal air valve orifice size.

**MULTI-OBJECTIVE GENETIC ALGORITHM EVOLUTIONARY POLYNOMIAL REGRESSION (EPR-MOGA)**

EPR-MOGA (Giustolisi & Savic 2009) is a data-modelling technique based on evolutionary algorithms that has been widely used in various fields of engineering (Ugarelli et al. 2009; Doglioni et al. 2012; Kornelsen & Coulibaly 2014; Yin et al. 2016).

It is a combined search for symbolic polynomial structures by genetic algorithm (GA) and estimation of coefficients of polynomials by least squares (LS) optimization, thus assuming a biunique relationship between a structure and its parameters (Giustolisi & Savic 2009). The EPR framework assumes as base model structure the following pseudo-polynomial expression structure:

\[
Y = a_0 + \sum_{j=1}^{m} a_j \cdot (X_1)^{ES_{(j,1)}} \cdot \ldots \cdot (X_k)^{ES_{(j,k)}} \cdot f\left(\left(X_1\right)^{ES_{(j,k)},2}\right) \ldots \cdot f\left(\left(X_k\right)^{ES_{(j,k),2}}\right)
\]

where \(a_j\) = adjustable parameter for the \(j^{th}\) term, \(a_0\) = optional bias, \(m\) = number of terms in the polynomial expression, \(X_k\) = the \(k^{th}\) column of the matrix of inputs \(\mathbf{X}\), \(\mathbf{ES}\) = matrix of candidate exponents, \(f(\cdot)\) is a function defined by the user among a set of available functions (logarithm, exponential, hyperbolic tangent, hyperbolic secant), \(Y\) = vector of outputs. More details on EPR can be found in Giustolisi & Savic (2009).

Assuming \(m\) pseudo-polynomial terms, MOGA-EPR explores the space of \(m\)-term model expressions by means of a MOGA strategy (Goldberg 1989), using the following three (conflicting) objectives:

1. maximization of model accuracy;
2. minimization of the number of polynomial coefficients;
3. minimization of the number of inputs.

The used MOGA approach is based on the Pareto dominance criterion and is named OPTIMOGA (Laucelli & Giustolisi 2011). It can determine the Pareto front including the best model expressions found optimizing parsimony (number of polynomial coefficients and inputs) and accuracy. This allows the easy interpretation of EPR results that are ranked according to the parsimony and accuracy objectives.

The Pareto front of optimal solutions, together with the symbolic nature of the EPR expressions, offers the possibility to analyze the importance of the available inputs for the problem at stake, for example, in terms of their presence in the Paretian expressions, the expressions with the lowest value of terms (i.e. \(m = 1\) or \(2\)). In the following investigation, the MS-Excel version of EPR-MOGA will be used (Laucelli et al. 2012).

**CASE STUDY**

In WDSs there are some important pipelines (e.g., those connecting water sources and urban reservoirs) that usually are filled or emptied by pipe segment between two isolation valves, rather than doing the same operation over the whole pipe, aiming to minimize the water volumes involved for both economic and time reasons. It is noteworthy that pipe systems are normally characterized by undulating profiles. Thus, this study adopts the experimental setup of Balacco et al. (2018) characterized by an undulating profile with a free orifice at the highest point to simulate an air valve. The upstream pipe was about 7.00 m long, having a slope of 11°, measured between the upstream butterfly valve and the orifice; the downstream pipe was about 7.40 m long, measured between the orifice and the downstream butterfly valve (Figure 2).

Previous experiments (Balacco et al. 2015; Apollonio et al. 2016) highlighted how the influencing parameters on the overpressure due to pipeline filling are the orifice size \(d\), upstream butterfly valve opening degree \(\psi_u\) and volume of air pocket \(V_{\text{air pocket}}\), as already confirmed by several studies (e.g. Lee & Martin 1999; Zhou 2000; Lee 2005).
The pipeline filling started opening the upstream valve, and its opening degree $\psi_u$ varied to reduce the supply pressure and thus the filling velocity. The orifice size was set to be like commercial air valves, to be sure the ratio $d/D$ fell within the range suggested by manufacturers. Finally, the volume of the air pocket ($V_{air\ \text{pocket}}$) was varied thanks to four outlets fitted along the descending pipe, which permitted the varying of the entrapped air volume downstream of the orifice (Figure 2).

The range of investigated values is summarized in Table 1. A whole asset database of 525 data records, obtained by the above-mentioned test rig, was adopted in this study. The presented procedure aims at identifying the functional relationships between the three possible candidate variables ($d^\circ$, $V^\circ$, $\psi_u$) and one output ($P^\circ$). Therefore, aiming to generalize the returned formulations and avoid problems related to dimensionality, the dimensional input data were modified in the following dimensionless parameters:

$$d^\circ = d/D \quad V^\circ = V_{air}/V_{pipe} \quad P^\circ = P_{max}/P_0$$

where $d$ is the orifice diameter of the air valve (mm), $D$ the pipe diameter (mm), $V_{air}$ the pocket air volume (m$^3$) and $V_{pipe}$ the volume of the sectioned pipe (m$^3$), $P_{max}$ the peak observed overpressure (bar) and $P_0$ the steady-state pressure (bar).

### EPR APPLICATION: SETTING AND MODEL SELECTION

To determine a relationship between the peak pressure $P^\circ$ and the candidate input data represented respectively by the orifice diameter $d^\circ$, the upstream valve opening degree $\psi_u$ and the air pocket volume $V^\circ$, the EPR model structure shown in Equation (3) was used without the inner function $f$:

$$Y = \sum_{j=1}^{m} a_j \cdot (X_1)^{ES(j,1)} \cdot \ldots \cdot (X_k)^{ES(j,k)} \cdot f((X_1)^{ES(j,k+1)} \cdot \ldots \cdot (X_k)^{ES(j,2k)})$$  \hspace{1cm} (3)

Aiming to limit the dimension of the search space and to obtain equations easier to be physically interpreted, the following candidate exponents ranged from $-3$ to $3$. In particular, the exponent 0 is crucial to deselect an input during the search, while the other exponents are related to an attenuation or an amplification effect on the model’s variables.

The model size $m$ was set equal to three terms according to the experimental results already obtained by Balacco et al. (2018) and Apollonio et al. (2016), and the bias term was assumed to be equal to zero. Finally, LS parameter estimation was constrained to search for positive polynomial coefficient values ($a_j > 0$).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\psi_u$ (%)</td>
<td>25, 50, 100</td>
</tr>
<tr>
<td>$d$ (mm)</td>
<td>6.5, 9.0, 12.0, 17.0, 19.0, 24.0, 33.0</td>
</tr>
<tr>
<td>$d/D$ (-)</td>
<td>0.096, 0.133, 0.177, 0.251, 0.280, 0.354, 0.487</td>
</tr>
<tr>
<td>$V_{air\ \text{pocket}}$ (m$^3$)</td>
<td>0.033, 0.037, 0.040, 0.044, 0.054</td>
</tr>
</tbody>
</table>
The data records are split into the training set, containing 80% of the total available data, and a test set. The former is used by EPR-MOGA to identify the models, while the latter is used to validate the models and test their generalization abilities on data not used (i.e., ‘unseen’) for model construction.

The adopted MOGA optimization strategy uses three conflicting objective functions: the minimization of the number of inputs ($X_i$), the minimization of the number of terms ($m$) (both representative of the minimization of model parsimony) and the maximization of the model accuracy (Sum of Squared Errors – SSE). The evolutionary construction of models by EPR-MOGA is based on a genetic algorithm, therefore, a number of generations is assumed for the optimization of model structures. According to the criterion reported by Giustolisi & Savic (2009) the maximum number of generations is assumed to be 810, which value depends on the length of the training set, number of candidate inputs, exponents and maximum number of monomial terms of each model. The fitness of predictions based on the EPR-MOGA models is here evaluated in terms of the Coef fi cient of Determination (CoD) which is based on the SSE as follows:

$$CoD = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_{\text{exp}})^2}{\sum_{i=1}^{n}(y_{\text{exp}} - \text{avg}(y_{\text{exp}}))^2} = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - \text{avg}(\hat{y}_{\text{exp}}))^2}{\sum_{i=1}^{n}(y_{\text{exp}} - \text{avg}(y_{\text{exp}}))^2}^{SSE}$$

where $n$ is the number of samples, $\hat{y}$ is the value predicted by the model and $\text{avg}(y_{\text{exp}})$ is the average value of the corresponding observations.

The EPR-MOGA run returned a number of optimal prediction models, as trade-offs between model parsimony and fitting to the experimental data. Table 2 lists the obtained formulas, while Figure 3(a) and 3(b) show respectively the Pareto front in terms on fitness vs number of terms and fitness vs number of selected inputs.

![Figure 3](https://iwaponline.com/ws/article-pdf/19/7/2036/607400/ws019072036.pdf)

**Table 2 | Formulas returned by EPR-MOGA**

<table>
<thead>
<tr>
<th>EPR formula</th>
<th>CoD</th>
<th>No. of $a_j$</th>
<th>No. of $X_h$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P^* = 0.818(1/d^{0.5})$</td>
<td>0.261</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$P^* = 0.522(1/d^{0.50}) + 0.938 \psi_{u}^{0.5}$</td>
<td>0.659</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>$P^* = 0.522(1/d^{0.50}) + 0.725 (\psi_{u}^{0.5}/V^{*})$</td>
<td>0.715</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>$P^* = 0.551(V^{<em>}/d^{0.5})^{0.5} + 0.805(\psi_{u}^{0.5}/V^{</em>})$</td>
<td>0.721</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>$P^* = 0.046(V^{<em>3}/d^{</em>}) + 0.422(1/d^{0.50}) + 0.825(\psi_{u}^{0.5}/V^{*})$</td>
<td>0.733</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>$P^* = 0.043(V^{<em>2.5}/d^{</em>} V^{<em>0.5}) + 0.339(1/d^{0.50}) + 0.949(\psi_{u}^{0.5}/V^{</em>})$</td>
<td>0.742</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>$P^* = 0.136(V^{<em>2}/d^{</em>}) + 0.274(1/d^{0.50} V^{<em>}) + 1.332(\psi_{u}^{0.5}/V^{</em>})$</td>
<td>0.745</td>
<td>3</td>
<td>7</td>
</tr>
</tbody>
</table>

It can be observed that the most important influencing input is $d^{0.5}$, even if the presence of $\psi_{u}$ is really important since model 2 it is better than model 1. In particular, the maximum overpressure $P^*$ in every case is an inverse function of the air valve orifice size $d^*$, while it is a direct function of the upstream valve opening degree $\psi_{u}$; finally,
the air pocket volume $V^*$ appears in some cases as numerator or denominator. Moreover, only the first two formulas do not have all three available input variables and these are characterized by the lower CoD values in the Pareto set. It is evident how the introduction of the $V^*$ improves the model performance, but at the same time it is clear how the model structure characterized by three polynomial terms does not increase significantly the accuracy of the model with respect to those having $m = 2$.

Starting from these considerations and given a consistent physical meaning for every term, the following expression (the third model in Table 2) is selected:

$$P_{\text{max}} = P_0 \left( 0.522 \frac{\sqrt{D}}{\sqrt{d}} + 0.725 \frac{V_{\text{pipe}}}{V_{\text{air}}^{0.5}} \right)$$

as obtained reporting the formulations of dimensionless variables.

The chosen model highlights that overpressure due to the filling pipeline is directly proportional to the square root of the upstream valve opening degree and inversely, respectively, to the square root of the air valve orifice size and the air pocket volume.

The selected model fits the training set with $\text{CoD} = 0.715$, while on the test data it shows a value of 0.731; this indicates good generalization skills, making it suitable also for other layouts with different ranges of input parameters. The chosen model has a very simple and intuitive mathematical structure, easily applied in technical contexts. Figure 4 shows respectively (a) the scatter plot of EPR results vs the training data and (b) the scatter plot of EPR results vs test data. Both figures confirm the good accuracy of the selected formula.

The model in Equation (5) permits the definition of the optimal air valve orifice size to adopt into the assumed pipeline, given the pipe diameter, the pipeline volume to fill, the upstream valve opening degree (a few percentage points, usually), the steady-state pressure and assuming the maximum overpressure allowable for the system.

Obviously, it is recommended to adopt a filling velocity limited to about 0.4 m/s with the aim to limit the overpressure due to the water column impact when the water front reaches the orifice. However, especially during the filling process, velocities can be very high due to the high piezometric gradient, but using valves (preferably needle valves) can be useful for controlling filling velocity (Fontana et al. 2016).

For instance, assuming a steady-state pressure of 4 bar, a maximum overpressure of 0.2 bar, a pipe diameter of 300 mm, a pipe length 200 m, the condition of being totally empty, and given an opening degree of the upstream valve of 20%, as a function of the filling velocity, the minimum optimal diameter for the air valve is about 12 mm. Instead, in the hypothesis of a rapid filling ($\psi_u = 100\%$), it can be verified how the minimum diameter to be adopted to limit the overpressure at the indicated value is 70 mm.

Figure 4 | (a) EPR vs training data; (b) EPR vs test data.
CONCLUSIONS

The MOGA-EPR has been used here to investigate the existence of a reliable formulation for air valve sizing with the aim to minimize overpressure due to the pipeline filling process. The EPR technique identified a set of optimal formulas starting from a set of training data coming from a wide experimental campaign conducted by Balacco et al. (2015, 2018) and Apollonio et al. (2016). These studies highlighted how the pressure trend due to the filling of an initially empty undulating pipeline is more complicated compared with what has been described in the literature so far, but above all has brought greater clarity to the parameters that mostly affect the phenomena. The same parameters were used as candidate variables for the EPR and available data input were used as training and test data.

Starting from these few variables, usually easily retrievable in real applications, the returned models in this paper are quite simple and understandable, allowing the analyst to interpret their physical consistency, thus being more confident in their application to real-life cases.

The suggested formula defines the maximum tolerable overpressure and, knowing the geometric parameters of the pipeline system to fill, permits the optimal air valve orifice size to be quickly identified.

REFERENCES


First received 19 February 2019; accepted in revised form 16 May 2019. Available online 3 June 2019.