Use First-Principles Techniques to get FOPDT Coefficient Values

R. Russell Rhinehart

A Feature Article for CONTROL Magazine
Vol. 35, No. 2, February 2022, pp 37-40

Many control techniques are based on the ideal First-Order Plus Deadtime (FOPDT) model of the process. Conventionally, we obtain model coefficient values by open-loop step-testing of the process; but this creates undesirable process upsets, and only reveals the local process behavior. This article suggests that simple phenomenological (first-principles mechanistic) modeling can provide coefficient values, without upsetting the process; and it also reveals how the values change with operating conditions.

Control techniques that use FOPDT models include PID tuning, gain scheduling, Internal Model Control, and structuring feedforward and decouplers. In addition, appropriate pairing for SISO loops in a multivariable control strategy, for instance using the Relative Gain Array (RGA) technique, are based on the gains in the FOPDT models. As well FOPDT descriptions are commonly used to describe process dynamic responses. FOPDT modeling is important to control practice.

The FOPDT model coefficients relate to the process gain (response sensitivity to an input), the first-order lag time-constant, and the delay. Alternate names and acronyms include First-Order Lag Plus Delay (FOLPD). If there is a step-and-hold pattern in an input, the model pretends that the process responds with a first-order lag, after a delay. See Figure 1 for this ideal relation.

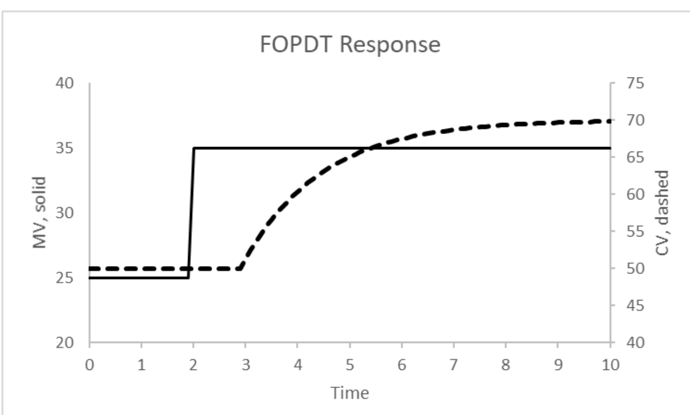

Figure 1 – Step-and-hold influence and ideal FOPDT response.

Often a Laplace notation is used to convey a FOPDT model. Sorry, but here it is:

$$\hat{y} = \frac{K_p e^{-\theta_p s}}{\tau_p s + 1} \hat{u}$$

$\hat{u}$ is the Laplace transformed influence variable, and $\hat{y}$ is the Laplace transformed response variable. K_p, τ_p, and θ_p are the modeled idealizations for the process gain, time-constant and delay. A first-order model has one time-constant. It does not matter whether you prefer to show the values in a Laplace transfer function, list them in a table, show them in a differential equation, or simply describe them with text; the same three coefficients are in any FOPDT representation.

The objective is to obtain values for K_p, τ_p, and θ_p.

Empirical Approaches

The classic approach to determine coefficient values in FOPDT models is the reaction curve method: Make a step-and-hold in the process influence from an initial steady state value, observe the process response, wait until a final steady state, then use one of many techniques to fit a FOPDT model to the reaction curve. One of the original pre-computer techniques is to draw a line on the strip chart to extrapolate the point of most rapid CV change to the initial and final values. Then read the delay and time-constant at the intersection points of the line and initial and final steady state values. More recently, parametric methods are preferred, such as using the times for the CV to move ¼ and ¾ of the total CV change. Rather than this two-point fit, regression could be used to provide an overall best fit. All these methods return similar, but different, τ_p and θ_p values.

However, noise, disturbances, and nonlinearity will often confound your ability to get a good model; and a one-sided step creates a deviation from a nominally desired controlled variable value. To address these issues, make 4 steps in an up-down-down-up pattern. This provides 4 process responses. The average of the 4 responses will diminish the confounding characteristics, and the above and below aspects keep the process averaging about the nominal value. Even though the sequence of step-and-hold inputs tempers such effects, the τ_p and θ_p values still depend on the method to best fit the model to the data.

Unfortunately, the 4-step sequence might take too long a time to be practicable. First, one must wait until an initial steady state, then wait for 4 subsequent steady states. Using a rough measure that steady state takes 5 time-constants, this means that the experimental period requires about 25 time-constants. If a time-constant is 10 minutes, this is over a 4-hr experimental duration; and it may not be possible to prevent external changes to operation or to permit such long periods away from setpoint, which may upset other human stakeholders. It also means that it requires frequent attention, and subjectivity to determine when the process is at each steady state within the vagaries of noise and uncontrolled disturbances.

One method to accelerate the experimental process and to avoid the other undesirable aspects just mentioned, is to use a skyline influence (so called because it resembles a city skyline) to replace the step-and-hold patterns, then best fit the FOPDT model to the data in a least-squares regression. Figure 2 illustrates a skyline input, which is a series of random steps within a range, held for random durations up to about two time-constants. It also illustrates the response and a best fit model. See Rhinehart, R. R., "FOPDT Modeling", Develop Your Potential Series in

CONTROL magazine, November 2016, Vol. XXIX, No. 11, pages 46-48. I offer an Excel VBA program to generate the skyline pattern and best fit models to data on my web site www.r3eda.com.

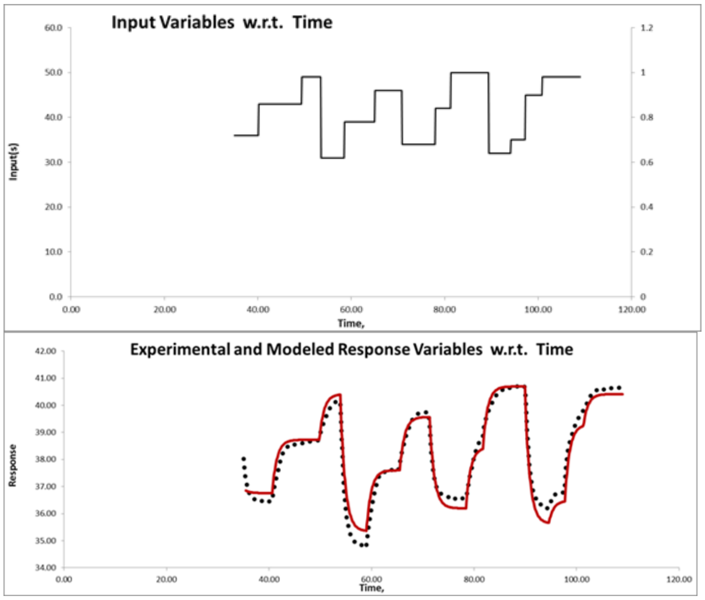

Figure 2 – (a) skyline influence, (b) response (solid, red) and model (black markers).

The Problem with Empirical Approaches

Empirical approaches get a locally valid model. However, if the process is nonlinear, the values for the gain, delay, and time-constant will likely change with new operating conditions. Changes in external disturbances during the test will shift the response during the test. You can see from Figure 2b that the model approximately represents the data, but not exactly.

Further, as the process changes in time, perhaps due to tank levels changing, piping rerouting, raw material property changes, ambient influences, again, the values for the gain, delay, and time-constant will likely change over time. Using empirical approaches to keep FOPDT models representative of the process as it continually changes may be onerous.

Things to accept

FOPDT models are not the process. Even if we term them as process models and use the subscript "p" on model coefficients, they are not. They are ideal approximations of reality. For example, if the process is high order without any delay, the FOPDT model (from a ¼ ¾ parametric fit) pretends that it has a first-order response after a pure delay. See Figure 3. The ideal FOPDT model is not the process.

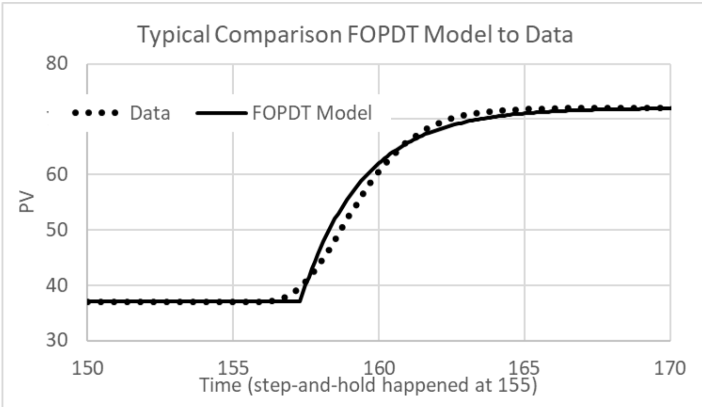

Figure 3 – A FOPDT model representation of a high-order process with zero delay.

Empirical testing includes process nonlinearities, measurement noise, valve sticktion, and continuing unmeasured disturbances. Even if a process was truly FOPDT, getting coefficient values from empirical testing would not return the true values.

The process behavior changes with operating level and due to continual changes in equipment, environmental influences, and raw material. The FOPDT models need to change over time.

The method you use for getting the τ_p and θ_p values (one point, two point, regression) returns different values.

Although imperfect, FOPDT models are usually good enough for the control application. They provide a balance of sufficiency with perfection. Continuing that acceptable imperfection balance, use first-principles approaches to determine FOPDT model coefficient values.

The Digital Twin Approach

If you have a reasonable simulator of your process, perhaps just a simulator, or perhaps a digital twin, then do the step-testing on your simulator, and fit the FOPDT model to its response. A digital twin is a simulator that is adjusted to fit the process better. See Rhinehart, R. R., "Understanding the Digital Twin", Part I, CONTROL for the process industries, Vol. 34, No. 10, October, 2021, pp. 61-65, Part II, CONTROL for the process industries, Vol. 34, No. 11, November, 2021, pp. 35-39, Part III, CONTROL for the process industries, Vol. 34, No. 12, December, 2021, pp. xx-xx.

But perhaps you can do your own phenomenological modeling.

First-Principles Approach

First-principles means to consider the major mechanisms. This contrasts with rigorous models, which seek to include even the minor mechanisms and perfectly model the constitutive relations.

First-principles models are characteristic of what they used in college courses and introductory textbooks to explain heat exchangers, reactions, distillation, filters, etc.

To start, consider what accumulates (the inventory) that relates to the response of interest. For instance, a process does not inventory liquid level. Contents within a tank are added or removed, and the inventory is the volume of material. Level does not flow in or out, but it is a measure of that inventory. Similarly, a process does not inventory temperature. The inventory is thermal energy (heat) that is added or removed. Temperature is a measure of the heat that has accumulated. If you want to model a temperature response, consider mechanisms for heat addition or removal. If you want to model pressure response, consider mechanisms for addition or removal of gas molecules.

The quantity being inventoried is often labeled the conserved quantity. It does not disappear. Heat can be transferred, material can be transported or converted into reaction products, but neither disappear in chemical processes. Commonly mass, momentum, and energy are the conserved quantities.

Constitutive relations relate the quantity of the inventory to the response. The ideal gas law, rearranged as $P = nRT/V$, could relate number of moles of gas (the conserved quantity) within a volume at a particular temperature to the resulting pressure (the response) that might be measured. A tank geometry could relate level to liquid volume, and liquid density could relate volume to mass. For a right circular cylinder $V = \pi r^2 h$ and $V = m/\rho$, so the constitutive relation that relates level to the inventory of mass is $h = m/(\pi r^2 \rho)$.

Next, identify the point (or volume) in which the quantity being inventoried is accumulated. This could be a tank, or vessel, volume of liquid in a heat exchanger tube, or a moving fluid.

Developing a First-Principles Dynamic Model

I'll use a simple example of mixing hot and cold fluid in a tank to illustrate the procedure. See Figure 3.

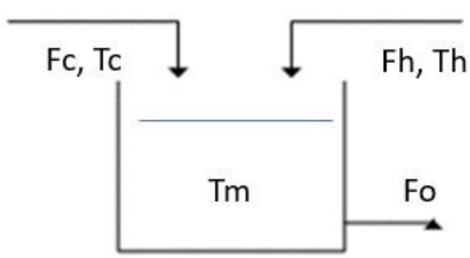

Figure 3 - Mixing hot and cold fluid in a tank

- Define a control volume: The tank contents.
- Choose a conserved quantity: Thermal energy, because we are interested in the temperature response.
- Assume: The water entering the tank is instantly and perfectly mixed. The contents are always spatially uniform.
- Apply IOGA: The acronym refers to the inventory going IN, less the inventory going OUT, plus the inventory internally GENERATED, which is equal to the inventory ACCUMULATED in a small increment of time, Δt.

$$\dot{E}_{inflowing}\Delta t - \dot{E}_{outflowing}\Delta t + \dot{E}_{generated}\Delta t = E_{inventory}\Big|_{t+\Delta t} - E_{inventory}\Big|_{t}$$

Here, $\dot{E}_i$represents the rate of energy transferred or produced, and $\dot{E}_i\Delta t$ is the quantity in the time interval.

Insert the elementary constitutive relations between measurable variables and the IOGA quantities. ie: $\dot{E}_{out} = F_o\rho c_p T_o$ which assumes the reference temperature is 0 and the coefficients are constants.

There are two sources of heat inflow (both the cold and hot water), one heat outflow (the exit stream) and no internal reaction to generate heat. The thermal energy IOGA becomes

$$F_c\rho c_p T_c\Delta t + F_h\rho c_p T_h\Delta t - F_o\rho c_p T_o\Delta t + 0 = V\rho c_p T_o\Big|_{t+\Delta t} - V\rho c_p T_o\Big|_{t}$$

Since ρ, c_p, and V are considered constants, then $F_{out} = F_c + F_h$

Divide by ρc_p to simplify, and divide by Δt

$$F_c T_c + F_h T_h - (F_c + F_h)T_o = V\frac{T_o|_{t+\Delta t} - T_o|_t}{\Delta t}$$

In the limit of very small Δt, the finite difference is the derivative

$$F_c T_c + F_h T_h - F_o T_o = V\frac{dT_o}{dt}$$

In standard ODE form

$$\frac{V}{(F_c + F_h)}\frac{dT_o}{dt} + T_o = \frac{F_c T_c + F_h T_h}{(F_c + F_h)}$$

Noteworthy regarding the time-constant:

- $\frac{V}{(F_c+F_h)}$ has the units of time. It is a time-constant. It is the value of τ_p needed for the FOPDT model. $\tau_p = \frac{V}{(F_c+F_h)}$.

- Here, the time-constant is the volume for inventory accumulation divided by the flow rate through the volume. In general, τ_p is the capacity for the control volume to accumulate the relevant inventory divided by the rate through it.
- If you know the value at one operating condition, then the equation shows how it scales with alternate operating conditions. It provides the gain scheduling rules. $\tau_2 = \tau_1 \frac{V_2}{V_1}\frac{F_{T1}}{F_{T2}}$. Here the subscripts indicate two conditions, and the F_T represents total flow rate.
- If you don't trust your ability to get the models right, trust your ability to use the models to scale time-constants to other operating conditions.

Regarding the process order:

- The one volume of inventory accumulation led to a first-order differential equation – a first-order response.
- Two volumes of inventory, such as tanks in series, lead to two ODEs – a second-order response.

Regarding the gains:

- At steady state, $\frac{dT_o}{dt} = 0$, so the equation shows that $T_{o,SS} = \frac{F_c T_c + F_h T_h}{(F_c + F_h)}$ is the steady state temperature (if $F_c,\ T_c,\ F_h,\ or\ T_h$ all remain constant).
- The first-principles model indicates what are the influence variables. On the right-hand side these are F_c, T_c, F_h, and T_h.
- If T_c is considered a disturbance, then the derivative of the steady state temperature w.r.t. T_c is the gain of the disturbance. Here calculus is very easy, $K_d = \frac{F_c}{(F_c + F_h)}$. If F_h is considered the controller influence, then the derivative of the steady state temperature w.r.t. F_h is the gain. Here calculus is not so easy, $K_{F_h} = \frac{(T_h - T_c)F_c}{(F_c + F_h)^2}$. If you are rusty on analytical calculus, use a finite difference numerical approach. Calculate $T_{o,SS}$ for two similar F_h values then calculate the gain by the ratio of the difference in $T_{o,SS}$ values divided by the difference in the F_h values. The gain from the first-principles model is the gain needed in the FOPDT model.
- The equation reveals how gain changes with operating conditions, which is a part of a gain scheduling rule.
- If F_c, for instance is considered a wild flow, and F_h is to be controlled as a ratio then the nominal ratio value is $r = \frac{T_{SS} - T_c}{T_h - T_{SS}}$. The actual value will likely need to be adjusted by feedback to compensate for sensor errors or influences not included in the model (such as ambient heat losses).
- If, again, for instance, F_c and F_h are both being manipulated to manage two controlled variables, such as total flow rate and temperature, then the gains from the steady state equation can be used in a relative gain array procedure for loop pairing.
- Your controller outputs are likely to be a % signal sent to a control valve, not individual F_c or F_h values. If you know the valve gain K_v (Rhinehart, R. R., "Understanding valve

characteristics", Develop Your Potential Series in CONTROL magazine, Vol. 33, No. 8, August, 2020, pp 41-42), then the gain experienced by the controller is the $K_{F_h} K_v$ product.

- If you don't trust your ability to get the models right, trust your ability to use the models to scale gains to other operating conditions.

If you have the steady state model $y_{ss} = f(u, d)$, which you might from equipment design, process optimization, or process analysis, then:

- If there is one major volume of inventory, the model is first order, which means $\tau \frac{dy}{dt} + y = y_{ss} = f(u, d)$.
- With a modicum of process experience, you probably know a reasonable value for the process time-constant.
- The steady state model reveals the gains. Again if you are rusty with calculus, use a finite difference approach.

Including a Delay

Deadtime is a delay from the event occurrence to its observation. Common examples are that thunder follows lightening by about 5 sec per mile distance. The batter hears the "crack" immediately, but the spectator, high in right field, hears it 2 seconds after seeing the batter swing.

The constitutive model for transport delay in a pipeline is $\theta = \frac{L}{v} = \frac{L\pi D^2/4}{F}$ (if ideal plug flow). In the mixing tank example, if the mixed fluid temperature is measured downstream of the tank, then there would be a transport delay. There might also be delays due to sample transport to either the lab or to the on-line analytical instrument, and then there is the lab analysis time or on-line analytical instrument cycle time. If there is a dominant lag, but there are several smaller lags, the time-constants for the smaller lags can be considered as pseudo delays. For example, if the thermowell and the control valve have 3 sec and 1 sec time-constants, which are small compared to a 1-min time-constant for the tank, these can be modeled as delays. Add all the delays to get the total. $\theta_{total} = \theta_1 + \theta_2 + \theta_3 + \cdots$ This θ_{total} is the value needed in the FOPDT model.

Typically, delays from sampling interval or communication devices are inconsequential.

Noteworthy Regarding the Delay:

- The model indicates how delay scales with operating conditions. For instance, if only flow transport $\theta_2 = \theta_1 \frac{L_2}{L_1} \frac{F_1}{F_2}$.
- Even, if a bit more complicated, the models can be used for gain scheduling rules.
- If you don't trust your ability to get the models right, trust your ability to use the models to scale delays to other operating conditions.

Takeaway

Even though the FOPDT reaction curve concept is presented as a response to a step-and-hold influence, if you are seeking to get process data to generate model coefficient values, use a skyline input pattern to generate empirical data.

However, consider using first-principles model approaches. First-principles models can:

1. Provide values for FOPDT model coefficients, considerably saving time and money, and avoiding process upsets.
2. Provide insight for control structure, such as which variables should be considered as MVs and which are disturbances, assigning loop pairing, should you choose ratio or feedforward, etc.
3. Reinforce a mechanistic process understanding, which has many auxiliary benefits in training, troubleshooting, and process improvement.
4. Reveal the FOPDT model coefficient values. Don't use calculus derivatives to get the gains. Use a numerical finite difference approximation. And certainly, even though the model may be a differential equation, don't solve dynamic models. Use the information to solve a process problem.
5. Generate adequate FOPDT coefficient values. The FOPDT idealization is not the process. Even with a best possible FOPDT model, the process is not FOPDT. Further, the coefficients are corrupted by noise and disturbances and the method to determine the τ_p and θ_p values. Classic modeling approaches are imperfect, as are first-principles approaches. As long as the FOPDT coefficient values are reasonable, the model will be useful.

\- - - - - - -

Russ Rhinehart started his career in the process industry. After 13 years and rising to engineering supervision, he transferred to a 31-year academic career. Now "retired", he enjoys coaching professionals through books, articles, short courses, and postings on his web site www.r3eda.com.