Plantwide control: Towards a systematic procedure

Sigurd Skogestad
Norwegian University of Science and Technology (NTNU)
Department of Chemical Engineering, 7491 Trondheim, Norway

Abstract

Plantwide control deals with the structural decisions of the control system, including what to control and how to pair the variables to form control loops. Although these are very important issues, these decisions are in most cases made in an ad-hoc fashion, based on experience and engineering insight, without considering the details of each problem. In the paper, a systematic procedure towards plantwide control is presented. It starts with carefully defining the operational and economic objectives, and the degrees of freedom available to fulfill them. Other issues, discussed in the paper, include inventory control, decentralized versus multivariable control, loss in performance by bottom-up design, and recycle systems including the snowball effect.

1. Introduction

A chemical plant may have thousands of measurements and control loops. In practice, the control system is usually divided into several layers:

- scheduling (weeks),
- site-wide optimization (day),
- local optimization (hour),
- supervisory/predictive control (minutes)
- regulatory control (seconds)

We here consider the three lowest layers. The local optimization layer typically recomputes new setpoints only once an hour or so, whereas the feedback layer operates continuously. The layers are linked by the controlled variables, whereby the setpoints are computed by the upper layer and implemented by the lower layer. An important issue is the selection of these variables.

By the term plantwide control it is not meant the tuning and behavior of each control loop, but rather the control philosophy of the overall plant with emphasis on the
structural decisions (Foss, 1973); (Morari, 1982); (Skogestad and Postlethwaite, 1996). These involve the following tasks:

1. Selection of controlled variables c ("outputs"; variables with setpoints)
2. Selection of manipulated variables m ("inputs")
3. Selection of (extra) measurements v (for control purposes including stabilization)
4. Selection of control configuration (the structure of the overall controller $K$ that interconnects the variables $c_s$ and $v$ (controller inputs) with the variables $m$)
5. Selection of controller type (control law specification, e.g., PID, decoupler, LQG, etc.).

A recent review of the literature on plantwide control can be found in Larsson and Skogestad (2000). In practice, the problem is usually solved without the use of existing theoretical tools. In fact, the industrial approach to plantwide control is still very much along the lines described by Page Buckley in his book from 1964. The realization that the field of control structure design is underdeveloped is not new. Foss (1973) made the observation that in many areas application was ahead of theory, and he stated that

The central issue to be resolved by the new theories is the determination of the control system structure. Which variables should be measured, which inputs should be manipulated and which links should be made between the two sets... The gap is present indeed, but contrary to the views of many, it is the theoretician who must close it.

This paper is organized as follows. First, an expanded version of the plantwide control design procedure of Larsson and Skogestad (2000) is presented. In reminder of the paper some issues related to this procedure are discussed in more detail:

Degree of freedom analysis
Selection of controlled variables
Inventory control

Finally, we discuss recycle systems and the so-called snowball effect.

2. A procedure for plantwide control

The proposed design procedure is summarized in Table 1. In the table we also give the purpose and typical model requirements for each layer, along with a short discussion on when to use decentralized (single-loop) control or multivariable control (e.g. MPC) in the supervisory control layer. The procedure is divided in two main parts:

I. Top-down analysis, including definition of operational objectives and consideration of degrees of freedom available to meet these (tasks 1 and 2)
II. Bottom-up design of the control system, starting with the stabilizing control layer (tasks 3, 4 and 5 above)
<table>
<thead>
<tr>
<th>STEP</th>
<th>Comments, analysis tools and model requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>I. TOP-DOWN ANALYSIS:</strong></td>
<td></td>
</tr>
</tbody>
</table>
| 1. PRIMARY CONTROLLED VARIABLES: | Steady-state economic analysis:  
Control active constraints  
Remaining DOFs: Control variables for which constant setpoints give small (economic) loss when disturbances occur.  
Steady-state economic analysis:  
Define cost and constraints  
Degree of freedom (DOF) analysis.  
Optimization w.r.t. DOFs for various disturbances (gives active constraints) |
| Which (primary) variables $c$ should we control?  
Control active constraints  
Remaining DOFs: Control variables for which constant setpoints give small (economic) loss when disturbances occur. | |
| 2. MANIPULATED VARIABLES | Usually given by design, but check that there are enough DOFs to meet operational objectives, both at steady state (step 1) and dynamically. If not, may need extra equipment. |
| Select manipulated variables $m$ (valves and actuators) for control. | |
| 3. PRODUCTION RATE: | Optimal location follows from steady-state optimization (step 1), but may move depending on operating conditions. |
| Where should the production rate be set?  
(Very important choice as it determines the structure of remaining inventory control system.) | |
| **II. BOTTOM-UP DESIGN:** | Controllability analysis: Compute zeros, poles, pole vectors, relative gain array, minimum singular values, etc. |
| (With given controlled and manipulated variables) | |
| 4. REGULATORY CONTROL LAYER. | 4.1 Pole vector analysis (Havre and Skogestad, 1997) Select measured variables and manipulated inputs corresponding to large elements in pole vector to minimize input usage caused by measurement noise.  
4.2 Partially controlled plant analysis. Control secondary measurements ($v$) so that the effect of disturbances on the primary outputs ($c$) can be handled by the layer above (or the operators).  
Model: Tuning may be done with local linear models or on-line with no model. Analysis requires linear multivariable dynamic model (generic model sufficient). |
| 4.1 Stabilization  
4.2 Local disturbance rejection | |
| *Purpose:* “Stabilize” the plant using single-loop PID controllers to enable manual operation (by the operators)  
*Main issue:* What more should we control?  
Select secondary controlled variables (measurements) $v$  
Pair these with manipulated variables $m$, avoiding $m$’s that saturate (reach constraints) | |
5. SUPERVISORY CONTROL LAYER.

**Purpose:** Keep (primary) controlled outputs \( c \) at optimal setpoints \( c_s \), using unused manipulated variables and setpoints \( v_s \) for regulatory layer as degrees of freedom (inputs).

**Main structural issue:** Decentralized or multivariable control?

### 5a. Decentralized (single-loop) control
Possibly with addition of feed-forward and ratio control.
- May use simple PI or PID controllers.
- Structural issue: choose input-output pairing

### 5b. Multivariable control
Usually with explicit handling of constraints (MPC)
- Structural issue: Choose input-output sets for each multivariable controller

5. REAL-TIME OPTIMIZATION LAYER

**Purpose:** Identify active constraints and compute optimal setpoints \( c_s \) for controlled variables

**Main structural issue:** What should \( c \) be (see step 1)

6. VALIDATION

**Nonlinear dynamic simulation of critical parts**

The procedure is generally iterative and may require several loops through the steps, before converging at a proposed control structure.

**Additional comment:**

(i) “Stabilization” (step 4). The objective of the regulatory control layer is to “stabilize” the plant. We have here put stabilize in quotes because we use the word in an extended meaning, and include both modes which are mathematically unstable as well as slow modes (“drift”) that need to be “stabilized” from an operator point of view.

(ii) Model requirements: In the control layers (step 4 and 5) we control variables at given setpoints, and it is usually sufficient with linear dynamic models (local for each loop) with emphasis on the time scale corresponding to the desired closed-loop response time (of each loop). The steady-state part of the model is not important, except for cases with pure feedforward control. For analysis it is usually sufficient with a generic model (which does not match exactly the specific plant), but for controller design model
identification is usually required. In the *optimization layer* (steps 1 and 6) a nonlinear steady-state model is required. Dynamics are usually not needed, except for batch processes and cases with frequent grade changes.

(iii) **Decentralized versus multivariable control (step 5).** First note that there is usually some decentralization, that is, there is usually a combination of several multivariable and single-loop controllers. An important reason for using multivariable constraint control (MPC) is usually to avoid the logic of reconfiguring single loops as active constraint move. The optimization in step 1 with various disturbances may provide useful information in this respect, and may be used to set up a table of possible combinations of active constraints. MPC should be used if a structure with single-loop controllers will require excessive reconfiguration of loops.

(iv) **Why not a single big multivariable controller?** Most of the steps in Table 1 could be avoided by designing a single optimizing controller that stabilizes the process and at the same time perfectly coordinates all the manipulated variables based on dynamic on-line optimization. There are fundamental reasons why such a solution is not the best, even with tomorrows computing power. One fundamental reason is the cost of modeling and tuning this controller, which must be balanced against the fact that the hierarchical structuring proposed in this paper, without much need for models, is used effectively to control most chemical plants.

3. **Degree of freedom analysis (step 1)**

The first step in a systematic approach to plantwide control is to formulate the operational objectives. This is done by defining a cost function $J$ that should be minimized with respect to the $N_{opt}$ optimization degrees of freedom, subject to a given set of constraints. A degree of freedom analysis is a key element in this first step. We start with the number of operational or control degrees of freedom, $N_m$ (here denotes manipulated). $N_m$ is usually easily obtained by process insight as the number of independent variables that can be manipulated by external means (typically, the number of number of adjustable valves plus other adjustable electrical and mechanical variables). Note that the original manipulated variables are always extensive variables.

To obtain the number of degrees of freedom for optimization, $N_{opt}$, we need to subtract from $N_m$

- the number of manipulated (input) variables with no effect on the cost $J$ (typically, these are “extra” manipulated variables used to improve the dynamic response, e.g. an extra bypass on a heat exchanger), and
- the number of (output) variables that need to be controlled, but which have no effect on the cost $J$ (typically, these are liquid levels in holdup tanks).

In most cases the cost depends on the steady state only, and $N_{opt}$ equals the number of steady-state degrees of freedom. The typical number of (operational) steady-state degrees of freedom for some process units are (with given pressure)
For example, for the reactor-distillation-recycle process with given pressure (Figure 4), there are four degrees of freedom at steady state (fresh feedrate, reactor holdup, boilup and reflux in distillation column). The optimization is generally subject to constraints, and at the optimum many of these are usually "active". The number of "free" (unconstrained) degrees of freedom that are left to optimize the operation is then $N_{\text{opt}} - N_{\text{active}}$. This is an important number, since it is generally for the unconstrained degrees of freedom that the selection of controlled variables (task 1 and step 1) is a critical issue.

4. What should we control? (steps 1 and 4)

A question that puzzled me for many years was: Why do we control all these variables in a chemical plant, like internal temperatures, pressures or compositions, when there are no *a priori* specifications on many of them? The answer to this question is that we first need to control the variables directly related to ensuing optimal economic operation (these are the primary controlled variables, see step 1):

- Control active constraints
- Select unconstrained controlled variables so that with constant setpoints the process is kept close to its optimum in spite of disturbances. These are the less intuitive ones, for which the idea of self-optimizing control (see below) is very useful.

In addition, we need to need to control variables in order to achieve satisfactory regulatory control (these are the secondary controlled variables, see step 4):

Self-optimizing control (step 1).

The basic idea of self-optimizing control was formulated about twenty years ago by Morari et al. (1980) who write that "we want to find a function $c$ of the process variables which when held constant, leads automatically to the optimal adjustments of the manipulated variables." To quantify this more precisely, we define the (economic) loss $L$ as the difference between the actual value of the cost function and the truly optimal value, i.e. $L = J(u; d) - J_{\text{opt}}(d)$. Self-optimizing control (Skogestad, 2000) is achieved if a constant setpoint policy results in an acceptable loss $L$ (without the need to reoptimize...
The main issue here is not to find the optimal setpoints, but rather to find the right variables to keep constant. To select controlled variables for self-optimizing control, one may use the stepwise procedure of Skogestad (2000):

**Step 1.1** Degree of freedom analysis
**Step 1.2** Definition of optimal operation (cost and constraints)
**Step 1.3** Identification of important disturbances
**Step 1.4** Optimization (nominally and with disturbances)
**Step 1.5** Identification of candidate controlled variables
**Step 1.6** Evaluation of loss with constant setpoints for alternative combinations of controlled variables (when there are disturbances or implementation errors)
**Step 1.7** Final evaluation and selection (including controllability analysis)

This procedure has been applied to several applications, including distillation column control (Skogestad, 2000), the Tennessee-Eastman process (Larsson et al., 2001) and a reactor-recycle process (Larsson et al., 2002).

5. **Production rate and inventory control (step 3)**

The liquid inventory (holdup, level) in a processing unit usually has no or little on steady-state effect, but it needs to be controlled to maintain stable operation. The bottom-up design of the control system (step 4) therefore usually starts with the design of the liquid level control loops. However, one needs to be a bit careful here because the design of the level loops has a very large effect of the remaining control system design, as it consumes steady-state degrees of freedom, and determines the initial effect of feedrate disturbances. Also, because the level loops link together the transport of mass from the input to the output of the plant, the level loops are very much dependent on each other. There are many possible ways of pairing the inventory loops, and the basic issue is whether to control the inventory (level) using the inflow or outflow? A little thought reveals that the answer to this question is mainly determined by where in the plant the production rate is set, and that we should control inventory (a) using the outflow downstream of the location where the production rate is set, and (b) using the inflow upstream of this location. This justifies why there in Table 1 is a separate step called "Production rate", because the decision here provides a natural transition from step 1 (top-down economic considerations and identification of active constraints) to step 4 (bottom-up design, usually starting with the level loops).

The production rate is most commonly assumed to be set at the inlet to the plant, so outflows are used for level control. One important reason for this is probably that most of the control structure decisions are done at the design stage (before the plant is build) where we usually fix the feedrate. However, during operation the feedrate is usually a degree of freedom, and very often the economic conditions are such that it is optimal to maximize production. As we increase the feedrate we reach a point where some flow variable E internally in the plant reaches its constraint E_max and becomes a bottleneck for further increase in production. In addition, as we reach the constraint we loose a degree of freedom for control, and to compensate for this we have several options:
1) Reduce the feedrate and “back off” from the constraint on E (gives economic loss).
2) Use the feedrate as a manipulated variable to take over the lost control task (but this usually gives a very “slow” loop dynamically because of long physical distance). To avoid this slow loop one may either:
3) Install a surge tank upstream of the bottleneck, and reassign its outflow to take over the lost control task, and use the feedrate to reset the level of the surge tank, or:
4) Reassign all level control loops upstream of the bottleneck from outflow to inflow (which may involve many loops).

All of these options are undesirable. A better solution is probably to permanently reassign the level loops, and we have the following rule: 

Identify the main dynamic (control) bottleneck (see definition below) in the plant by optimizing the operation with the feedrate as a degree of freedom (steady state, step 1). Set the production rate at this location. The justification for this rule is that the economic benefits of increasing the production are usually very large (when the market conditions are such), so that it is important to maximize flow at the bottleneck. On the other hand, if market conditions are such that we are operating with a given feed rate or given product rate, then the economic loss imposed by adjusting the production rate somewhere inside the plant is usually zero, as deviations from the desired feed or production rate can be averaged out over time, provided we have storage tanks for feeds or products. However, one should be careful when applying this rule, as also other considerations may be important, such as the control of the individual units (e.g. distillation column) which may be effected by whether inflow or outflow is used for level control.

We have here assumed that the bottleneck is always in the same unit. If it moves to another unit, then reassignment of level loops is probably unavoidable if we want to maintain optimal operation.

Note that we here have only considered changes in operating conditions that may lead to bottlenecks and thus to the need to reassign inventory (level) loops. Of course, other active constraints may move and the best unconstrained controlled variable (with the best self-optimizing properties) may change, but the reconfiguration of these loops are usually easier to handle locally, as they. This may also require configuration of loops, but usually may done locally.

MPC in regulatory control layer

The above discussion assumes that we use single-loop controllers in the regulatory control layer (which includes level control), and that we want to minimize the logic needed for reassigning loops. An alternative approach, which overcomes most of the above problems, is to use a multivariable model-based controller with constraints handling (MPC), which automatically tracks the moving constraints and reassigns control tasks in an optimal manner. This is many ways a more straightforward approach, but such controllers are more complex, and its sensitivity to errors and failures is quite unpredictable, so such controllers are usually avoided at the bottom of the control hierarchy.
Another alternative, which is more failure tolerant, is to implement a MPC system on top of a fixed single-loop regulatory control layer (which includes level control). As shown in Theorem 1 (below) this gives no performance loss provided we let the multivariable have access also to the setpoints of the lower-layer regulatory controllers (including the ability to dynamically manipulate the level setpoints). The regulatory layer then provides a back-up if the MPC controller fails, but under normal conditions does not affect control performance.

**Definition of bottleneck**

Consider a given objective function, given parameters, given equipment (including given degrees of freedom) and given constraints (including quality constraints on the products). A unit (or more precisely, an extensive variable $E$ within this unit) is a bottleneck (with respect to the flow $F$) if

1. With the flow $F$ as a degree of freedom, the variable $E$ is optimally at its maximum constraint (i.e., $E = E_{\text{max}}$ at the optimum)
2. The flow $F$ is increased by increasing this constraint (i.e., $\frac{dF}{dE_{\text{max}}} > 0$ at the optimum).

A variable $E$ is a dynamic (control) bottleneck if in addition

3. The optimal value of $E$ is unconstrained when $F$ is fixed at a sufficiently low value

Otherwise $E$ is a steady-state (design) bottleneck.

**Remarks on definition:**

1. Typically, $F$ is the flowrate of the main feed or main product.
2. Most of the information required to identify bottlenecks follow from the optimization with various disturbances in step 1.
3. The fact that an extensive variable is at its maximum constraint does not necessarily imply that it is a bottleneck, because we may have that $\frac{dF}{dE_{\text{max}}} = 0$ (e.g., this may happen if the variable $E$ is a cheap utility).
4. In many cases $F$ is also the objective function to be maximized, and the values of $\frac{dF}{dE_{\text{max}}}$ are then directly given by the corresponding Lagrange multipliers.
5. We may in some cases have several bottlenecks at the same time, and the main bottleneck is then the variable with the largest value of $E_{\text{max}} \frac{dF}{dE_{\text{max}}}$. (i.e. with the largest relative effect on the flowrate).
6. The location of the bottleneck may move with time (i.e., as a function of other parameters)
7. The concept of “bottleneck” is clearly of importance when redesigning a plant to increase capacity. It is also important in terms of operation and control, because the main bottleneck is the variable that should be operated closest to its constraint.
8. Steady-state bottlenecks may be important in terms of design, but need normally not be considered any further when it comes to deciding on a control structure (as they should always be kept at their maximum). Examples of possible steady-state bottleneck variables are reactor volumes and heat exchangers areas.
9. A control policy based on fixing intensive variables is not steady-state optimal for systems with bottlenecks.

6. Application: Recycle systems and the snowball effect

Figure 1. Reactor with recycle process with control of recycle ratio \(L/F\), \(M_r\) (maximum reactor holdup), and \(x_B\) (given product composition).

Luyben (1993) introduced the term “snow-ball effect” to describe what can happen to the recycle flow in response to an increase in fresh feedrate \(F_0\) for processes with recycle of unreacted feed (see Figure). Although this term has been useful in pointing out the importance of taking a plantwide perspective, it has lead to quite a lot of confusion.

To understand the problem, let us first consider the “default” way of dealing with a feedrate increase, which is to keep all the intensive variables (compositions) in the process constant, by increasing all flows and other extensive variables with a steady-state effect in proportion to \(F_0\). This is similar to how one scales the production rate when doing process simulation, and is the idea behind the “balanced” control structures of Wu and Yu (1996). Specifically, this requires that we keep the residence time \(M_r/F\) constant, that is, we need to increase the reactor holdup \(M_r\) in proportion to the reactor feedrate \(F\).

However, changing the reactor holdup (volume) during operation is usually not possible (gas phase reactor), or at least not desirable since for most reactions it is economically optimal to use a fixed maximum reactor volume in order to maximize per pass conversion (i.e., reactor holdup is a steady-state bottleneck, see above). To increase conversion (in response too an increase in feedrate \(F_0\)) one may instead increase the
concentration of reactant by recycling unreacted feed. However, the effect of this has limitations, and the snowball effect occurs because even with infinite recycle the reactor concentration cannot exceed that of pure component. In practice, because of constraints, the flow rates do not go to infinity. Most likely, the liquid or vapor rate in the column will reach its maximum value, and the result of the snowballing will be a breakthrough of component A in the bottom product, that is, we will find that we are no longer able to maintain the product purity specification ($x_B$).

To avoid snowballing Luyben et al. (1993, 1994) and Wu and Yu (1996) propose, to use the “default” approach with a varying reactor holdup, rather than a `conventional" control structure with constant holdup. Their simulations show that a variable holdup policy works better, but these simulations are strongly misleading, because in the “conventional” structure they fix the reactor holdup at a value well below the maximum values used in the varying holdup structures. In fact, the lowest value of the recycle D for a given value of $F_0$ is when the reactor holdup $M_r$ is at its maximum, so the conventional structure with maximum holdup is actually better in terms of avoiding snowballing.

A more careful analysis of the reactor with recycle process shows that there are four degrees of freedom at steady-state, including the fresh feedrate $F_0$. With a fixed feedrate $F_0$, there are three degrees of freedom. If the economic objective is to minimize the energy usage (i.e., minimize boilup V), then optimization with respect to the three degrees of freedom, give that $M_r$ should be kept at its maximum (to maximize conversion), and that the product composition $x_B$ be kept at its specification (overpurifying costs energy). These two variables should then be controlled (active constraint control). This makes the Luyben structure and the two balanced structures of Wu and Yu (1996) economically unattractive. There is one unconstrained degree of freedom left, and the issue is next to decide which variable we should select to keep constant. Alternatives are, for example, the amount of recycle D or F (“Luyben rule”), composition $x_D$ (conventional structure), reflux L, reflux ratios L/D or L/F, etc. Larsson et al. (2002) evaluated the energy loss imposed by keeping these constant when there are disturbances in $F_0$ and recommended for the case with a given feedrate $F_0$ to use the recycle ratio structure shown in the Figure. Keeping D or F constant (Luyben rule) yields infeasible operation for small disturbances. This confirms the results of Wu and Yu (1996). This is easily explained: As the feedrate $F_0$ is increased, we must with constant $F=F_0+D$ reduce the recycle $D$ to the reactor. Therefore light component A will accumulate in the distillation column and operation becomes infeasible.

For this plant the reactor holdup is a steady-state (design) bottleneck, whereas the column capacity ($V_{max}$) is the dynamic (control) bottleneck. Thus, if it is likely that the plant will be operated under conditions where we want to maximize production, then we should probably use a control structure where the production rate is set at the column bottleneck (V), and inventory control should use inflow upstream of this location. In Figure 4, this would involve using the feedrate $F_0$ to control the reactor level, using the column feed F to control bottom composition, and using the boilup V to reset the feedrate $F_0$ to its given value (note that $F_0$ is both an input and output in this case).
In summary, the "snowball effect" is a real operational problem if the reactor is "too small", such that we may encounter or get close to cases where the feedrate is larger than the reactor can handle. The "snowball effect" makes control more difficult and critical, but it not a control problem in the sense that it can be avoided by use of control. Rather, it is a design problem that can be easily avoided by installing a sufficiently large reactor to begin with. The Luyben rule of fixing a flow in the recycle loop seems to have little basis, as it leads to control structures that can only handle very small feedrate changes.

7. Conclusion

We have here presented a first step towards a systematic approach to plantwide control. There are many outstanding research issues related to filling in out more detailed procedures in Table 1 on what to do in each step of the procedure. For example, more work is needed in order to understand how to decompose and coordinate the layers of the control system.

References


