Determining intervention thresholds that change output behavior patterns

Citation for published version (APA):

Document license:
CC BY

DOI:
10.1002/sdr.1564

Document status and date:
Published: 01/07/2016

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher’s website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.
Determining intervention thresholds that change output behavior patterns

Bob Walrave*

Abstract
This paper details a semi-automated method that can calculate intervention thresholds—that is, the minimum required intervention sizes, over a given timeframe, that result in a desired change in a system’s output behavior pattern. The method exploits key differences in atomic behavior profiles that exist between classifiable pre- and post-intervention behavior patterns. An automated process of systematic adjustment of the intervention variable, while monitoring the key difference, identifies the intervention thresholds. The results, in turn, can be studied and presented in intervention threshold graphs in combination with final runtime graphs. Overall, this method allows modelers to move beyond ad hoc experimentation and develop a better understanding of intervention dynamics. This article presents an application of the method to the well-known World 3 model, which helps demonstrate both the procedure and its benefits.

Copyright © 2017 System Dynamics Society


Additional Supporting Information may be found online in the supporting information tab for this article.

Introduction

Because so many systems and problems are characterized by dynamic complexity, the number of studies that apply system dynamics (SD) has increased accordingly (e.g., Repenning, 2001; Romme et al., 2010; Van Oorschot et al., 2013). Applications of SD range from global-level analyses (Meadows et al., 2004) to studies on the firm (Walrave et al., 2015) or individual (Repenning, 2001) levels. Of particular interest are intervention studies, which explore “the degree of change in model behavior as a result of alternative policies or scenarios” (Yücel and Barlas, 2015, p. 173). In other words, intervention studies pertain to how an issue or problem can be corrected (Forrester, 1961) and rely on model-based experimentation. Such explorations, often referred to as “what-if” experiments (Morecroft, 1988), typically are conducted through ad hoc adjustments of key model parameters (e.g., Repenning, 2001; Walrave et al., 2011). Yet such a manually conducted approach implies that most modelers work with a very limited number of experiments and evaluations, simply due to time constraints, which in turn limits the policy formulation and analysis phase of SD (Sterman, 2000).

* Correspondence to: Bob Walrave, School of Industrial Engineering, Eindhoven University of Technology. E-mail: b.walrave@tue.nl

Accepted by Markus Schwaninger, Received 9 November 2015; Revised 24 May 2016, 7 December 2016 and 1 February 2017; Accepted 3 February 2017

System Dynamics Review

Published online in Wiley Online Library (wileyonlinelibrary.com) DOI: 10.1002/sdr.1564
Although scholars have made significant progress with automating various parts of the SD modeling process—including advances in automated sensitivity analyses (e.g., Ford, 1990; Pruyt and Islam, 2016), the inclusion of different statistical approaches for rigorous parameter estimation (e.g., Oliva, 2003; Peterson, 1980), and parameter specification based on automated behavior pattern feature recognition (e.g., Yücel and Barlas, 2011)—modelers still lack a focused method to automate what-if experiments. In particular, no specifically designed approach exists to determine intervention thresholds, defined here as the minimum required intervention sizes, over a given time span, to achieve a change in output behavior pattern that “corrects the problem”. Some methods potentially could be customized to determine such intervention thresholds (e.g., Kwakkel and Pruyt, 2015; Yücel and Barlas, 2011), but it would require complex manipulations. Perhaps, as a result, many system dynamicists refrain from moving beyond ad hoc experimentation, which in turn limits the development of our understanding of intervention dynamics.

In response, this article presents a semi-automated method designed specifically to calculate intervention thresholds, by monitoring a key difference between classifiable pre- and post-intervention behavior patterns, in terms of their atomic behavior, while systematically adjusting the intervention variable of interest. The method is of value to modelers who want to go beyond ad hoc experimentation and conduct systematic analyses of intervention thresholds and how they change over time. The latter question has long been subject to calls for increased attention, at least in organization science settings (e.g., Hannan and Freeman, 1984). In addition, in proposing an intervention thresholds graph, in combination with final runtime graphs, this article suggests a means to illustrate and study intervention dynamics.

The next section provides the building blocks for the development of the method, including a brief review of behavior patterns and key characteristics of atomic behavior. The steps detailed thereafter specify the process that results in intervention thresholds graphs. To illustrate the method, this process is applied to the well-known World 3 model (Meadows et al., 2004, 2008). This article concludes with a discussion of some benefits and limitations of the method, including suggestions for further research.

**Toward intervention thresholds analyses in intervention studies**

Even the simplest SD models can exhibit complex nonlinear behavior, due to combinations of feedback loops, delays, and shifts in loop dominance. As a result, the SD community started to explore automated model configuration and analyses techniques, to better cope with the dynamic complexity exhibited by many SD models. Perhaps the best-known contributions are automated sensitivity analyses methods, such as those that rely on random univariate sampling or multivariate Monte Carlo sampling, which are now
widely incorporated into SD software packages (e.g., Ford et al., 1983; Ford, 1990). Barlas and Kanar (1999) and Yücel and Barlas (2011, 2015) advance a method for the automatic recognition of behavior pattern features that allows for, among other things, the automatic specification of model parameters. To make the parameter estimation more rigorous, scholars have also suggested integrating various statistical approaches into the SD modeling process for calibration (e.g., full-information maximum likelihood via optimal filtering, model reference optimization; Oliva, 2003; Peterson, 1980).

Beyond model calibration and validation, system dynamicists frequently seek to determine the effect that parameter changes have on system behavior, through what-if experiments. Such input manipulations often appear in the context of intervention studies (e.g., Romme et al., 2010; Repenning, 2001; Walrave et al., 2011). For example, explorations might address which intervention size, at which moment in time, can break a reinforcing behavior that has manifested itself as an unanticipated side effect, as exemplified by “fixes that fail” structures (Senge, 1990). In this respect, interventions often aim to result in some particular change in output behavior patterns, such as inducing a shift from exponential decline to goal-seeking growth in firm performance. For such efforts, the intervention thresholds underlying such pattern change represent highly pertinent information. Walrave et al. (2015) calculate the intervention thresholds (i.e., months of managerial commitment) required to counteract an unanticipated self-reinforcing phenomenon (i.e., success trap) for all possible intervention moments (all $t$ in the model). When an intervention size at a given moment is smaller than the intervention threshold, the outcome behavior is reinforced decline, but when the intervention size increases above the threshold the outcome behavior shifts to goal-seeking growth. Therefore, the method proposed herein considers the intervention threshold size, relative to its timing, that is required to achieve an anticipated change in the output behavior pattern.

Such an approach requires many unique simulation runs (possible intervention sizes $\times$ permissible timeframe), so resource constraints likely prevent the manual discovery of intervention thresholds. Intervention thresholds also can rarely be deduced analytically (cf. Rudolph and Repenning, 2002). Instead, an exploratory approach is necessary to assess all (theoretically) possible intervention sizes over a permissible timeframe, as might be achieved by customizing existing methods. For example, the exploratory modeling and analysis workbench (Kwakkel et al., 2013; Kwakkel and Pruyt, 2015) incorporates pattern classification and clustering features, which can be used, among other things, to automatically determine output behavior patterns. The pattern-oriented parameter specifier discussed by Yücel and Barlas (2011, 2015) also can be applied to determine a parameter value that yields a specific output behavior pattern. Yet the adaptation of these methods requires rather complex manipulations. Instead, this article details a specifically designed, semi-automated process, building on work by Barlas
Automated pattern recognition refers to “the automatic discovery of regularities [in datasets] through the use of computer algorithms and the use of these regularities to take actions” (Yücel and Barlas, 2015, p. 176). Such an approach has been successfully applied in various research domains, such as economics, medicine, marketing, and biology (Angstenberger, 2001; Cordus and Piccolo, 2008).

The proposed method exploits the tendency for dynamics to reflect a limited set of behavior patterns. Based on Sterman (2000), Barlas and Kanar (1999), and Yücel and Barlas (2011, 2015), I recognize seven main modes of behavior, as outlined in Table 1: (1) zero/constant behavior; (2) linear growth/decline; (3) exponential growth/decline; (4) goal seeking growth/decline; (5) S-shaped growth/decline; (6) growth and decline or decline and growth; and (7) oscillation with/without growth/decline. These main modes can be further subdivided into 15 behavior patterns that possess distinctive (sequences of) first derivatives (i.e., slope), second derivatives (i.e., curvature), and means. In other words, every behavior pattern has a distinctive atomic behavior profile. These atomic behavior profiles effectively identify output behavior patterns (Yücel and Barlas, 2015). To develop a parsimonious approach to identify intervention thresholds, the current study proposes a custom approach for systems that show classifiable pre- and post-intervention output behavior patterns, such that there is only a need to distinguish between two behavior profiles, rather than identify them, which can be achieved by comparing a key difference in their atomic behavior.

For example, a comparison of S-shaped growth against growth and decline (see No. 5a and No. 6a in Table 1) reveals that the latter, at some point, displays a negative slope. The two behavior patterns can thus be distinguished by monitoring the first derivative of the output variable: the first derivative of S-shaped growth will always be positive, but the first derivative of growth and decline will become negative at some particular moment in time. This difference can be monitored by making the derivative an indicator variable (in the model), with a cut-off value of zero. The sign change in this indicator variable, in turn, points to a change in the output behavior pattern. By monitoring it, while systematically adjusting the intervention size between a lower and an upper bound and over a permissible timeframe, it is possible to identify the intervention thresholds. The first intervention size—for every moment in the timeframe—that causes the indicator variable to switch sign is the intervention threshold.

While some changes in behavior patterns can be captured by simply observing the first or second derivative, identifying other pattern changes may require a more sophisticated approach. Consider, for example, a change from growth and decline to oscillation. Table 1 shows an identical atomic behavior profile for these two behavior patterns, yet only in the case of oscillation does this
Table 1. Output behavior patterns and key characteristics of atomic behavior

<table>
<thead>
<tr>
<th>No.</th>
<th>Output behavior pattern</th>
<th>First derivative (slope)</th>
<th>Second derivative (curvature)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>Zero¹</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1b</td>
<td>Constant¹</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2a</td>
<td>Linear growth</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>2b</td>
<td>Linear decline</td>
<td>−</td>
<td>0</td>
</tr>
<tr>
<td>3a</td>
<td>Exponential growth</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>3b</td>
<td>Exponential decline</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>4a</td>
<td>Goal seeking growth</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>4b</td>
<td>Goal seeking decline</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>5a</td>
<td>S-Shaped growth (exp. gr. → goal seek. gr.)</td>
<td>+ → +</td>
<td>+ → −</td>
</tr>
<tr>
<td>5b</td>
<td>S-Shaped decline (exp. decl. → goal seek. decl.)</td>
<td>− → −</td>
<td>− → +</td>
</tr>
<tr>
<td>6a</td>
<td>Growth and decline (exp. gr. → goal seek. gr. → exp. decl. → goal seek. Decl.)</td>
<td>+ → + → − → − → + → − → − → +</td>
<td></td>
</tr>
<tr>
<td>6b</td>
<td>Decline and growth (exp. decl. → goal seek. Decl. → exp. gr. → goal seek. gr.)</td>
<td>− → − → + → + → − → + → + → −</td>
<td></td>
</tr>
<tr>
<td>7a</td>
<td>Oscillation²,³</td>
<td>Multiple episodes of</td>
<td>Multiple episodes of</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+ → + → − → −</td>
<td>+ → − → − → +</td>
</tr>
<tr>
<td>7b</td>
<td>Oscillation with growth²,³</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7c</td>
<td>Oscillation with decline²,³</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This table reveals key differences among output behavior patterns. Modelers should first assess any difference in slope; if no difference in slope exists (e.g., linear vs. exponential growth), they should check for any difference in curvature; if no such difference exists (e.g., oscillation vs. oscillation with growth), they should evaluate differences in the mean.

¹The mean effectively discriminates between the zero and the constant output behavior patterns.
²The atomic behavior profile for growth and decline or decline and growth and oscillation (with or without growth/decline) might be identical. Yet only in the case of oscillation does the profile unfold more than once.
³The different types of oscillation are best discriminated by the first derivative of their moving averages.

behavior profile unfold more than once. As such, one should introduce an indicator variable that counts the number of switches between a positive and negative first derivative over the full model run. While the first derivative of growth and decline only switches once (from positive to negative), the first
derivative of oscillation switches more than once, implying that the correct cut-off value for the aforementioned indicator variable equals 2.

Table 1 suggests means to select an appropriate indicator variable and cut-off value for distinguishing between various behavior patterns. Note that this approach requires the modeler to be able to anticipate the nature of the output behavior pattern change, due to an intervention, because two specific patterns need to be compared. The proposed method is thus not (yet) fit to accommodate unpredictable output behavior.

**Intervention thresholds analysis**

Figure 1 outlines the workflow for the intervention thresholds analysis, which consists of five main steps, such that a modeler should:

1. *Determine pre- and post-intervention output behavior patterns.* Table 1 serves to identify these output behavior patterns.

2. *Determine the indicator variable and its cut-off value.* Table 1, columns 3 and 4, aid in selecting an indicator variable and the correct cut-off value on the basis of a key difference in atomic behavior that best discriminates between two output behavior patterns.

3. *Determine the boundaries for intervention size and timing.* The modeler should decide on a (theoretically informed) lower and upper bound for intervention size and a permissible intervention timeframe. It is the responsibility of the modeler, who should be familiar with the model’s structure and dynamic behavior, to make these decisions with great care. Carefully designed experiments to uncover the post-intervention output behavior pattern can assist modelers in this decision-making process.

4. *Run the automated intervention thresholds analysis.* This automated fourth step involves systematic IF-THEN experiments, as shown in Figure 1. A script is instructed to start a FOR loop\(^1\) (Loop 1), which iterates through all possible intervention moments. A second FOR loop (Loop 2) then starts, which operates within loop 1 and is directed to iterate through all possible intervention sizes until it either identifies an intervention threshold or reaches the upper bound intervention size (i.e., no intervention threshold found). Specifically, the script instructs the model to run a simulation with the two inherited parameters (i.e., size and timing), after which the simulation output is saved. The script then assesses the indicator variable for an intervention threshold. If the cut-off value is not exceeded, no intervention threshold is identified. The script then determines whether all possible intervention sizes were assessed. If not, the intervention size is adjusted by one increment, and the analysis repeats. If an intervention

\(^1\)This FOR loop refers to a conditional loop used in programming, not to the traditional feedback loops used by system dynamicists.
If a threshold is identified (i.e., cut-off value is exceeded) or all possible intervention sizes are assessed, the script exits the second loop and determines whether the entire timeframe was considered. If not, the script adjusts the intervention timing by one increment; otherwise, the script ends.

Alternatively, modelers may generate the required data through sensitivity analyses. By modeling a STEP function on the intervention variable, where the step size denotes the intervention size and the step time indicates the...
intervention timing, then conducting a sensitivity analysis on these two inputs, a modeler can generate the required raw data for the next step. When applying this approach, the modeler must identify the actual intervention thresholds from the raw data (i.e., by inspecting the indicator variable and cut-off value in relation to intervention size and timing, perhaps in a spreadsheet program). This approach circumvents the need for external macros and may decrease the computational load, but it also limits the potential for extensions and/or modifications (e.g., investigating two-stage interventions by including a third FOR loop).

5. **Draw an intervention thresholds graph.** Using the results of step 4, the modeler creates an intervention thresholds graph, with the intervention threshold size on the y-axis and timing on the x-axis. Figure 2 displays an example. For every analyzed $t$, the graph shows the intervention threshold. Rather than depicting a continuous line that unfolds over time, the intervention thresholds graph presents the minimum intervention size required to establish an anticipated shift in the output behavior pattern (x-axis) at every analyzed intervention time during the permissible timeframe (y-axis). In Figure 2, for example, an intervention at $t = 5$ requires a minimum intervention size of 12 to prompt the anticipated output behavior pattern. Any intervention at any particular moment in time that is equal to or larger than the value in the intervention thresholds graph thus results in the classified post-intervention behavior pattern. Any intervention smaller than this value does not. By studying the graph, it is possible to observe shifts in the model’s resistance to change, as a function of the intervention timing.
Applying the intervention thresholds analysis to World 3

To illustrate the method, I turn to seminal work by the Club of Rome (Meadows et al., 1972) and more recent updates (Meadows et al., 2004, 2008). The World 3–03 model (World3_03_Scenario.vmf, revision date 14 August 2008) features a dynamic system, including population, industrial growth, food production, and limits to the Earth’s ecosystems, resources, and pollution. The model also describes various scenarios. Scenario 10, which serves as the starting point for this example, postulates that if a particular policy package were to have been implemented by society in 1982 (i.e., intervention timing), we would have been able to “maintain our standard of living and support its improving technologies with no problems” (Meadows et al., 2008, model tab Table of Scenarios). In this scenario, the Earth does not experience any significant decline in human population, due to the more sustainable interplay between its population and its carrying capacity.

An important feature of the previously mentioned policy package, which to some extent drives model behavior, is the industrial output per capita desired (IOPCD), which represents the desired wealth per capita (in U.S. dollars per person per year). The higher the IOPCD, the higher the population’s desired living standard and the faster the depletion of non-renewable resources needed to achieve this level and the higher the likelihood of population overshoot and subsequent decline will be. The following example applies an intervention thresholds analysis to World 3, with the size of the IOPCD as the focal intervention.

1. **Determine pre- and post-intervention output behavior patterns.** The description of Scenario 10 suggests the population should follow an S-shaped growth pattern. If interventions were introduced sometime after 1982, behavior instead is increasingly likely to follow the growth and decline pattern. Thus, depending on the size and timing of the intervention, a change in output behavior pattern can be expected, from S-shaped growth to growth and decline. Systematic experimentation reveals that this observation is not strictly true, though. Figure 3 shows the behavior of Population in eight experiments in which only the intervention timing varied—from 1980 to 2050, at 10-year increments. That is, rather than introducing the policy package in 1982, different intervention years were chosen, with a constant intervention size (i.e., at 500). As Figure 3 illustrates, all runs show some amount of overshoot, but whereas some runs overshoot only marginally and then stabilize (runs 1980, 1990, and 2000), practically approaching S-shaped growth, others oscillate strongly (runs 2010 and onward), clearly following a growth and decline pattern.

2. **Determine the indicator variable and its cut-off value.** The categorized output behavior patterns aid decision making related to the appropriate indicator variable and its cut-off value. Table 1 indicates that the main difference between S-shaped growth and growth and decline pertains to their slopes: always positive for the former; initially positive but then
negative for the latter. The appropriate indicator variable in this case therefore must relate to the first derivative of Population, with a cut-off value of zero. As Figure 3 illustrates, though, this heuristic might not be applicable in a strict manner in this particular example. Table 2 confirms that the slope of the Population variable becomes negative at least once during each run. This implies that a cut-off value of zero is not effective in determining a change in output behavior patterns. However, as Table 2 shows, the most drastic change in the steepest negative slope observed (over the full model run) occurs between 2000 and 2010, which corresponds to a change in output behavior pattern. Further inspection of the values in Table 2 suggests setting the cut-off value at approximately $-0.001$ to differentiate effectively between the two behavior patterns.

From a purely technical point of view, this cut-off value does not perfectly correspond to the prescriptions from Table 1. That is, we can speak of S-shaped growth only if the first derivative is never negative. Yet, as this example serves to illustrate, the approach fits even if the model runs do not strictly correspond to the fundamental output behavior patterns in Table 1. These patterns likely are sufficient for many studies, and Table 1 can serve

Table 2. Steepest negative slope observed in Population (over the full model run)

<table>
<thead>
<tr>
<th>Intervention timing</th>
<th>Steepest negative slope (until year 2100)</th>
<th>Behavior pattern (approximated)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980</td>
<td>$-0.00060$</td>
<td>S-shaped growth</td>
</tr>
<tr>
<td>1990</td>
<td>$-0.00053$</td>
<td>S-shaped growth</td>
</tr>
<tr>
<td>2000</td>
<td>$-0.00077$</td>
<td>S-shaped growth</td>
</tr>
<tr>
<td>2010</td>
<td>$-0.00743$</td>
<td>Growth and decline</td>
</tr>
<tr>
<td>2020</td>
<td>$-0.01835$</td>
<td>Growth and decline</td>
</tr>
<tr>
<td>2030</td>
<td>$-0.02096$</td>
<td>Growth and decline</td>
</tr>
<tr>
<td>2040</td>
<td>$-0.02314$</td>
<td>Growth and decline</td>
</tr>
<tr>
<td>2050</td>
<td>$-0.02521$</td>
<td>Growth and decline</td>
</tr>
</tbody>
</table>
as an inspiration for choosing indicator variables. Yet, in practical terms, the preceding numbers clearly indicate a difference between the two sets of runs. In cases characterized by uncertainty regarding the correct cut-off value, modelers are advised to conduct a step 2b, which involves a limited exploration for the purposes of indicator evaluation. First of all, a range of cut-off values can be selected to assess cut-off value sensitivity; in the current example, the modeler could choose a set of cut-off values ranging between \(-0.001\) and \(-0.005\). Furthermore, modelers should visually check the effectiveness of both indicator variable and cut-off value—see Figure 5 in the results section—and adjust the indicator variable/cut-off value if necessary.

3. Determining the boundaries for intervention size and timing. For this example, the value of IOPCD was assigned a lower limit of $200/(person \times year)$ and an upper limit of $1000/(person \times year)$; values outside of this range are theoretically unlikely. The model default for Scenario 10 was $350/(person \times year)$. The period 1980–2050 functions as the permissible intervention timeframe, and the original model runtime (1900–2100) was maintained.

4. Run the automated intervention thresholds analysis. To automate the intervention thresholds analysis for the World 3–03 example, a custom script is required. The pseudo-code in Box 1 provides building blocks to automate the intervention thresholds analysis according to the workflow outlined in Figure 1. In online supplementary materials, I provide the code for Microsoft’s Visual Basic for Applications in combination with Microsoft Excel and Ventana Vensim DSS. Running such a script results in a table that denotes the minimum required intervention sizes to produce the anticipated output behavior pattern, in relation to the intervention timing. This output serves as the input for the next step.

Box 1. Pseudo-code for intervention thresholds analysis

```
Iterate over all possible intervention times (For $t_1$ to $t_k$) { 
    Iterate over all possible intervention sizes (For $S_1$ to $S_n$) { 
        Determine value of indicator variable for $S_i$ at $t_i$
        IF size of indicator variable $<$ cut-off value {
            Increase intervention size
        } ELSE {
            Save intervention size (and other data of interest) for $t_i$
            Break loop 2
        }
    }
}
```
5. **Draw an intervention thresholds graph.** Finally, as explained, the intervention thresholds graph displays the intervention threshold sizes as a function of intervention timing. Figure 4(a) depicts the graph for this example, based on the output of step 4. The graph should not be read as a continuous line unfolding over time; rather, the “lines” in Figure 4(a) depict intervention threshold sizes (y-axis, in absolute values of IOPCD) that result in the anticipated output behavior pattern, at the indicated intervention time (x-axis). That is, an IOPCD value lower than the intervention threshold size (at a particular moment in time) results in S-shaped growth; a higher IOPCD results in growth and decline.

Building on this graph, it is possible to draw final runtime graphs (at \( t = 2100 \)) for the Human Welfare Index (Figure 4b) and Population (Figure 4c). That is, Figures 4(b, c) denotes the final runtime values (values at \( t = 2100 \)) for the Human Welfare Index and Population at each intervention threshold. According to Figure 4(a), the model that contains an indicator variable with a cut-off value of \( /C0.001 \) has an intervention threshold at an IOPCD of 363 (intervention size) at \( t = 2010 \) (intervention timing). Then the final runtime values, at \( t = 2100 \), for the Population and the Human Welfare Index associated with this intervention threshold equal approximately 8 billion and 82 percent respectively, as displayed in Figure 4(b, c).

**Results: Validation and interpretation**

Figure 4 contains the results for five cut-off values, with the same indicator variable, to illustrate the sensitivity of the analysis. If the patterns changed significantly across different cut-off values, further investigation would be warranted, such as by choosing or constructing a different, more robust indicator variable and cut-off value. The results in this example instead illustrate that, though the intervention threshold sizes are higher for higher cut-off values, the general trend of the results remains constant, which is a sign of robustness.

To illustrate the dynamic behavior of Population that results from different interventions, Figure 5 presents three model runs. Keeping both the intervention timing and the cut-off value constant (at 2010 and \(-0.001\), respectively), three scenarios depicted an intervention size that was (a) 25 percent lower than the intervention threshold, (b) equal to the intervention threshold, and (c) 25 percent higher than the intervention threshold. The output of the first run clearly shows S-shaped growth, whereas the output of the third distinctly exhibits growth and decline. However, the second run appears to be on the border between S-shaped growth and growth and decline. As such, Figure 5 visually validates the indicator variable and its cut-off value, as well as the anticipated pre- and post-intervention output behavior patterns.
Fig. 4. Intervention thresholds graph (a) and final runtime graphs (b, c) for five cut-off values.
Going into detail about the implications of the results is beyond the scope of this paper, but further inspection of the dynamics in Figure 4(a), for a cut-off value of $-0.001$, underscores some important observations. As Figure 4(a) shows, the initial high levels of IOPCD are sustainable, in that they do not result in an undesired change in output behavior pattern. A healthy balance between resource demand and availability can be maintained. Yet around the year 1990, high levels of IOPCD are no longer sustainable, and the intervention threshold size drops very quickly up to 2018. Thereafter, a limit is reached; that is, a “threshold” seems to appear within the intervention thresholds graph. From this point on, even the smallest IOPCD cannot prevent the undesired change in output behavior patterns; Population always overshoots a sustainable balance between resource demand and availability, followed by a significant decline. This point crops up abruptly and is associated with a big negative step in the final Human Welfare Index. This finding also serves to illustrate the importance of investigating intervention dynamics.

It is important to note, with respect to the former observation, that the different interventions, over the permissible timeframe, had dissimilar incubation times, due to the fixed final runtime. Behavior characteristics thus may be pushed beyond the simulation horizon. For example, a particular parameter change might slow down the growth part of a growth and decline output behavior pattern, thereby pushing the decline part beyond the simulation horizon. To counteract this potential bias, a relatively long minimum delay of 50 years was maintained between the intervention and final run. Nevertheless, the observed threshold does not necessarily exist in the behavior space of the model. The results, however, represent a tipping point in a policy context: the moment in time when an intervention is still able to trigger a particular output behavior pattern, before a given deadline.

More results could also be distilled from this analysis. For example, the steep decline in the IOPCD, required to prevent the system from overshooting,
clearly illustrates the importance of the timing of the intervention. Further analysis of these findings—and other results that can be developed with intervention thresholds analyses and graphs—represent interesting avenues for research.

**Discussion, outlook, and conclusion**

Studies that apply SD have steadily increased, largely due to the “increasingly complex nature of [systems and] common problems faced” by researchers, policy makers, and practitioners alike (Rahmandad et al., 2015, p. 1). Human intuition falls short in navigating such situations, and formal models become indispensable for learning and decision making (Oliva, 2003). As models and their dynamics become increasingly complex, modelers turn to automated model configuration and analysis techniques (Ford, 1990; Peterson, 1980; Pruyt and Islam, 2016; Yücel and Barlas, 2011). Following in this tradition, this article details a dedicated, semi-automated method to uncover intervention thresholds that result in changes in output behavior patterns. In the context of intervention studies, the proposed intervention thresholds analysis can discover the set of minimally required intervention sizes, over a permissible timeframe, to address an issue or problem.

The method exploits differences between classifiable pre- and post-intervention behavior patterns, in terms of a key difference in atomic behavior profiles (Barlas and Kanar, 1999). Through an automated process of systematic adjustments of the intervention variable, while simultaneously monitoring the key difference (i.e., the indicator variable), intervention thresholds can be calculated. The results of this analysis then can be presented in an intervention thresholds graph, which denotes the minimum intervention size (y-axis) in relation to intervention timing (x-axis). This graph illustrates the change in the system’s resistance to interventions as a function of intervention timing. The method differs from existing frameworks (e.g., Kwakkel and Pruyt, 2015; Yücel and Barlas, 2011), in that it is designed specifically to calculate intervention thresholds (graphs) and is more lightweight as a result. In turn, system dynamicists can go beyond manually conducted, ad hoc experiments—as are commonly presented in management and organization science (e.g., Romme et al., 2010; Walrave et al., 2011)—which should stimulate new studies of intervention dynamics.

This method aims to uncover a change in output behavior patterns, to solve an issue or problem, but potentially it could be used in research settings in which no such change is expected. For example, Van Oorschot et al. (2011) describe the effectiveness of different interventions (decision-making heuristics) for a particular performance indicator (new product sales) but do not anticipate any change in output behavior pattern as a result of this intervention; the performance indicator is always characterized by S-shaped
growth. The method proposed herein could be customized, however, to identify automatically which intervention, at which moment in time, results in maximum performance (e.g., based on the magnitude of sales).

Furthermore, this method could be extended to determine intervention thresholds that underlie tipping points in the behavior space of a model. A tipping point is an important property of a dynamic system that indicates the critical size of a variable at which a change in loop dominance occurs, at a particular moment in time. Some of the system communities’ most influential contributions build on tipping point analyses to support their arguments and insights. For example, Rudolph and Repenning’s (2002) disaster dynamics study demonstrates how the accumulation of interruptions can drive an organizational system from a self-regulating system to a “fragile, self-escalating regime” (p. 1). In this context, the tipping point reflects the particular, critical setting that causes the system to undergo a fundamental change in behavior—that is, a shift in loop dominance. Some researchers have been successful at deducing tipping points analytically (typically, because their models can be described using first- and second-order differential equations), but it remains a challenge for researchers facing more complex models. The proposed method and further developments of this approach could thus prove very valuable in efforts to probe for tipping points. Some challenges still need to be overcome, though. First, when there is no clear-cut intervention variable, modelers must identify a parameter that drives the tipping point or else adapt the method to facilitate multiple parameters. Second, detecting a shift in loop dominance is more complicated than identifying an anticipated change in output behavior patterns. Even if a loop is (and remains) dominant, system behavior might change significantly. Further research should extend the presented method to address these challenges.

Every method is subject to limitations; the one presented herein is attuned to systems with low to intermediate behavior complexity, because it requires classifiable pre- and postintervention output behavior patterns. In its current form, the method is not applicable to extremely complex models with unpredictable output behavior, despite being sufficient for many cases. Therefore, further work could extend this method to deal with increasing complexity, such as by means of incorporating automatic pattern recognition to distinguish more than two output behavior patterns (see Yücel and Barlas, 2015).

Furthermore, as noted, the World 3 example maintains a fixed final runtime (at \( t = 2100 \)), while varying intervention timing over a set timeframe. As a result, the different interventions had dissimilar incubation times, which could push the behavior characteristics beyond the simulation horizon. If a modeler is interested in uncovering tipping points in a policy context, this potential limitation is not really a problem, but in other cases a dynamic final runtime may be required.

The computational load involved with this method also might be problematic in some cases, such as analyses that include many possible
intervention sizes and a large permissible intervention window and that are subject to small increment sizes for both intervention size and timing. To manage the computational load, modelers might increase or decrease increment sizes, depending on the computing power available. I recommend modelers start their intervention thresholds analysis with relatively large increment sizes, which will decrease the time required to run the analysis (and perhaps adjust some settings, such as lower and upper bound intervention sizes). The increment sizes can then be set to smaller values to render smooth intervention thresholds and final runtime graphs.

**Note**

i. Yücel and Barlas (2015) recognize seven main modes but 25 different behavior patterns, rather than the 15 in Table 1. This difference arises because Yücel and Barlas describe more variations of the growth-and-decline and decline-and-growth patterns. For ease of understandability, I omit these variations and stay true(er) to the fundamental modes described by Sterman (2000).

**Acknowledgments**

I thank the three anonymous reviewers and the journal editors for their insightful comments and suggestions. Furthermore, I would like to acknowledge Sharon Dolmans, Georges Romme, and Kim van Oorschot for their valuable input. I also gratefully acknowledge the members of the System Dynamics Community for the feedback provided during the 2014 annual meeting in Delft, the Netherlands.

**References**


