A System Dynamics Perspective on Applying a New Energy Efficiency Metric for Data Centers

James Simonelli
Schneider Electric – IT Business
85 Rangeway Rd.
Billerica MA 01862
Jim.Simonelli@Schneider-Electric.com

Abstract

This paper explores the dynamics of setting energy policies for data centers in the face of new metrics and regulations. With these new metrics there is a potential for management to establish policies that achieve the specific metric target while sub-optimizing the total energy reduction opportunity. This paper will address the question of whether new insight can be gained by using a system dynamics approach versus static return on investment forecasting. The first part of this paper describes the unique dynamics of energy consumption in a data center and the application of one particular metric used to indicate energy efficiency. The second part of this paper proposes a model that represents the interconnected behavior of data center energy consumption and metric based policy implementation. This approach is compared to static methods and the insights gained from taking a systems approach.

Introduction

Data Centers represent one of the most electro intensive elements in a non-industrial type building (EPA, 2007). By applying lessons from the US Department of Energy's EnergySTAR, LEED and other similar programs for energy reduction, a new group of energy metrics are emerging specifically targeting Data Centers(EPA, 2010). These metrics are an attempt to provide tangible and quantified measures to guide data center energy reduction programs in the commercial and public sectors. However, using these metrics for goal setting without understanding of the structure and dynamics of data center energy consumption can lead to sub-optimized initiatives, potentially leaving millions of kilowatt-hours of energy reduction efforts unaddressed. When establishing policies to reduce energy consumption in a data center, simple return on investment tools can be used to determine policies that guide the best activity for achieving a particular goal (Lawrence Berkley National Lab, Ancis Inc., 2010). However, these tools generally look out into the future (desired state), compare against the present condition (current state) and analyze different investment scenarios that lead from the current state to the future state. The minimal total investment to reach the desired state is typically put forward as the driver for the new policy. This approach, which this paper will call a static return on investment (static ROI) fails to take into account the dynamics of key indicators as the policy is implemented. If these key indicators have dynamics that temporarily trend in the wrong direction, unwarranted intervention in the program could result, which can lead to sub-optimal overall program success.

This paper asks the question of whether the insight gained from modeling a data center energy reduction program can help avoid unwarranted program intervention and give enhanced insight in to policy implementation as compared to static ROI methods.

The paper is organized by first describing the energy dynamics and growth patterns of data centers and describes the problem that can arise when an emphasis on instantaneous metrics drives program investment. A model is then proposed that represents the aggregate structure and dynamics of data center energy consumption and an energy reduction investment program. A sensitivity analysis of key policy parameters is done and the results are compared against static ROI methodology results. The insights gained from the dynamic model are explained and strengths and weaknesses of implementing this simplified model approach are discussed.

Data Center Energy Consumption

Since 2006, the energy consumed by data centers has risen to the point where it is catching the attention of government and business leadership. In one study commissioned by the Department of Energy, the total energy consumed by data centers in the United States was estimated at 1.5% of the total electricity available from the grid. The report also went on to state that at the projected rates of increase, this percentage was anticipated to rise to 3% by the year 2011 (EPA, 2007)(Koomey, 2008). The electro intensity of data centers coupled with its fast rate of increase pose electrical supply concerns for the country, given the time and environmental concerns associated with bringing new electrical capacity to the grid. Businesses also are taking notice as the cost to power a data center typically exceeds to cost of the information and communication technology (ICT) equipment it houses (Mitchell-Jackson, 2001) (Belady C. L., 2007) (EPA, 2007). This rise in energy consumption, coupled with the associated costs and environmental impacts have driven governments and businesses to develop energy efficiency programs and policies for data centers.

Data centers present an interesting efficiency challenge because “efficiency” is often used in a generalized manner. In a classical definition efficiency would be measured by calculating a ratio between the useful energy and the total energy. In this classical definition useful energy is typically “heat” measured in BTU/hr or kWh and total energy has the same units. However, for a data center, what is the useful energy? Or conversely what is the wasted energy? By definition a data center produces “data” that has a business utility and this utility is typically not measurable in kWh. As such, in a data center context, it becomes difficult to say that a data center consuming 10MWh is any less “efficient” than a data center consuming 100MWh if the business utility they serve is different (# of user, applications, and types of calculations…). A more detailed explanation of these challenges can be found in (The Green Grid, 2007)
The data center industry and governmental agencies have devised metrics to try to segment this problem by looking at two areas of energy consumption.

1. Energy consumed (lost) in the infrastructure (INF) delivering power and cooling to ICT equipment

2. Energy consumed by the ICT equipment

The first area is generally a problem of the infrastructure of the data center and its architecture for delivering reliable and efficient power and cooling to ICT equipment. The second area is broader reaching and is impacted by types of ICT equipment (servers, storage, network), amount of equipment (# of servers, storage architecture…), technology of equipment (low power processors, efficient power supplies…) and software architecture (processor utilization, virtualization, power savings mode…).

A typical ratio of energy consumption between these two components is one to one – for every watt consumed by ICT equipment an additional watt is used to power and cool the equipment (EPA, 2007)(Patterson, 2006). If we assume that all energy consumed by ICT equipment is "useful" and the energy used to distribute power to and remove heat from the ICT equipment as "waste," then the efficiency of a data center could be measure as:

$$DCiE = \frac{ICT\ Power}{ICT\ Power + INF\ Powerloss}$$

Where

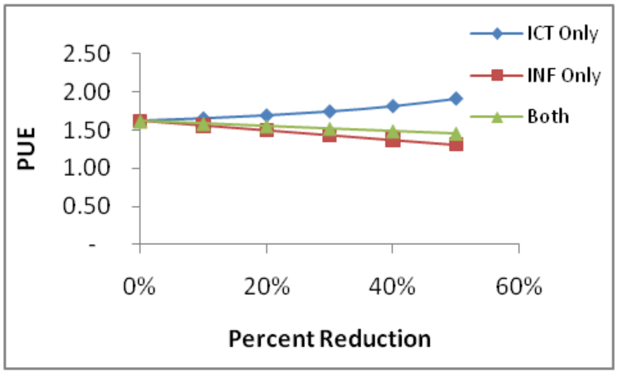
$$INF\ Powerloss = \text{Electriccal cooling, distribution and conversion losses}$$
$$ICT\ Power = \text{Power consumed by ICT equipment}$$
$$DCiE = \text{Data Center infrastrcture efficiency}$$

The intent of DCiE is to measure the infrastructure efficiency and not to measure the true productivity of the data center (which would need to include some form of useful output in terms of data such as MIPS/kW or some other ratio of useful work per kW consumed).
Even with this narrower intent, this measure is so often used, that a related metric is now widely used in the industry to gain comparative sense of "efficiency" between data centers. The measure is called Power Utilization Effectiveness (Belady, 2007) or PUE for short and is actually the inverse of DCiE. PUE tries to give a sense for how much of the power entering a data center is actually used to power the ICT equipment. A ratio of 1.0 means that all of the power entering the data center is used to power and cooling ICT, while a ratio of 2.0 means that for every watt consumed by ICT equipment a watt is "lost" in delivery.

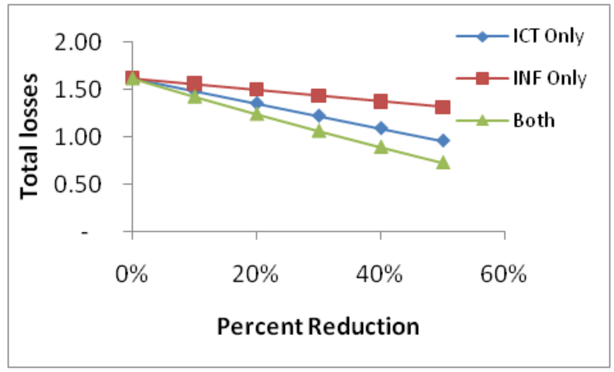
$$PUE = \frac{1}{DCiE}$$

$$PUE = \frac{ICT\ Power + INF\ Powerloss}{ICT\ Power}$$

An interesting attribute of the PUE metric is that the ratio has ICT Power in the denominator. This means that if ICT Power is decreased (generally a good thing) the metric will increase (higher is bad). The non intuitive implications are that a reduction in energy consumption (in this case ICT) drives an indicator for "energy efficiency" in the wrong direction. A policy strongly guided by this single metric could end up falling short of its potential for overall energy reduction.



(a)



(b)

Figure 1 Graph showing impact on various power reduction schemes on PUE (a) and Total losses (b) in per unit where 1 equal 100% ICT power design capacity. ICT Only scheme reduces ICT power source, INF Only scheme reduces INF loss source only and Both reduces both ICT and INF loss sources.

Figure 1 shows the impact of independently and jointly reducing the energy waste on the PUE metric and the Total losses. Notice that if the goal was to minimize PUE, then implementing the scheme that reduces the greatest total losses (Both) would not achieve this goal. if the goal was narrowly focused on the PUE metric, effort placed solely on the infrastructure reduction would yield the best PUE results (1.3) but would not minimize total power losses (1.3 pu vs. 0.73 pu) as shown in 1b.

One subtly, in the above analysis is that the INF losses are also related to the ICT losses because the power for ICT equipment must be delivered through the infrastructure. This relationship is fully described in reference (Rasmussen, 2006) and is a 2nd order curve fit described by:

$$P_{total}(P_{ICT}, P_{INF}) = P_{ICT} + P_{INF}(P_{ICT})$$

where,

$$P_{INF}(P_{ICT}) = a + b * P_{ICT} + c * {P_{ICT}}^2$$

and,

a = **Fixed coefficient**: Represent infrastructure losses that exist even with ICT load equal to zero
b = **Proportional coefficient**: Represents infrastructure losses that scale proportional to ICT load
c = **Square loss coefficient**: Represents infrastructure losses that scale as the square of the ICT load (usually representing copper losses)

Therefore the mechanism for reducing INF loss contribution only would be to reduce the coefficients *a*, *b* and *c*. The implications are that reducing ICT losses decreases BOTH P_{ICT} and P_{INF} (and thus total power) whereas impacting the INF loss contributions by changing *a*, *b* and *c* only impacts P_{INF}.

Given this non intuitive relationship, why is PUE used as a metric anyway? One factor is that it is possible for efficiency to be compared from site to site, much like miles per gallon can be used to compare automobile "efficiency" between similar vehicles. Another factor is that efficiency is deterministic and does not rely on comparisons to a baseline (differential metrics). This deterministic nature and normalization for any site, has made efficiency metrics popular industry benchmarks (EPA, 2010). However, differential metrics are often more meaningful, but have the challenge of requiring a baseline to be known and not meaningful when comparing different sites. For example, take the statement: "my data center uses 30% less energy than it did before." Without knowing the baseline or the opportunity to reduce energy (maybe it could have been 80%), it is a challenge to create industry benchmarks.

Framing the specific problem

To examine this challenge more deeply we will use a hypothetical single data center as the baseline for the forthcoming analysis. The current state of this data center is shown in Table 1.

Table 1 Current state parameters

Parameter	Value	Units	Notes
Max ICT load	1	Per Unit (pu)	
Current ICT load	0.5	Per Unit (pu)	50% of the available ICT load is currently being used
Current INF losses	0.46	Per Unit (pu)	Fixed loss = 0.3 Prop. loss = 0.3 Square loss = 0.02
Current PUE	1.9		
Max estimated ICT power reduction opportunity	60%		Goal for final ICT power is 0.2 Per Unit (pu)
ICT growth rate	0%		No new equipment will be added. This constraint is added to simplify the analysis and does not change the methodology. However, ICT growth will have a positive impact on PUE but a negative one on energy consumption.
Max estimated INF reduction opportunity	40%		Goal for final INF loss is 0.25 Per Unit (pu)
Average ICT reduction sensitivity	0.01	kW/USD	For every 1 dollar invested, 10 watts (.01 kW) is removed
Average INF reduction sensitivity	0.01	kW/USD	For every 1 dollar invested, 10 watts (.01 kW) is removed
Cost of electricity	0.1	USD/kW-h	Assume all energy comes from electricity

The above situation describes a data center that has excess capacity (50% currently loaded) and has not recently had an energy efficiency program. The implications are that investment actions are likely to yield results approaching the maximum reduction opportunities. Examples of some of these actions are shown in Table 2 and are fully described in (US DOE, 2008).

Table 2 Power reduction opportunities

ICT Reduction possibilities	INF Reduction opportunities
Shutting off un-used servers	Increase set point temperature of data center
Increase utilization of existing servers through virtualization	Orient IT racks in hot-cold aisle arrangement and add blanking panels to open spaces
Replace existing servers with newer energy efficient servers (but no net add)	Ensure proper floor tile vent size and location (if raised floor is used)
	Replace Air Conditioners with newer models with fan speed controls. Or add variable frequency drives to existing air conditioners (if possible)
	Leverage cool outside temperature when possible to minimize mechanical work of chilled water system (ie install economizers)

The problem: What is the best investment policy to implement that minimizes the initial investment, but yields savings that are at least 50% of the estimated maximum achievable (where "estimated maximum achievable" equals 60% reduction in INF losses, 40% reduction in ICT losses in this example). A non-stated, but relevant management expectation is that the current PUE will be reduced by an amount consistent with the formulaic impact yielded by these savings (1.72 in this example). The problem is straightforward, but attention must be paid to the fraction of investments made to reduce ICT losses, those used to reduce INF losses and the sequence/mix of each if both the savings and PUE expectations are to be achieved.

Literature review

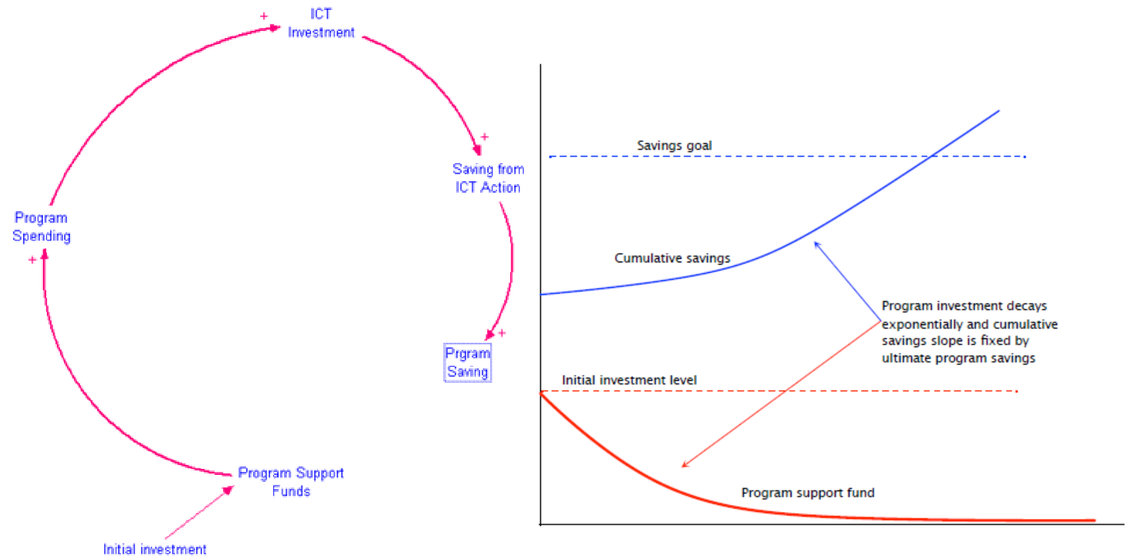
While there is literature on specific topics around subsystem level energy reduction such as heat density and efficient cooling (Belady, 2007) (Zhang, VanGilder, & Healey, 2009) (Lawrence Berkeley National Lab, 2007) (Shehabi, 2009), there is a scarcity of peer reviewed literature on the specific topic of data center energy reduction methodologies. However, there are a number of public sources on the topic from credible organizations such as The Green Grid (The Green Grid, 2007), US Department of Energy(US DOE, 2008) (EPA, 2007)(EPA, 2010)(EPA, 2010), Lawrence Berkley National Labs (Lawrence Berkley National Lab, Ancis Inc., 2010) as well as many white papers, case studies and software tools available (most for free) from corporate enterprises (Rasmussen, 2006). The governmental sponsored and The Green Grid materials are known to be publicly debated before reports are issued, so where possible this material should be more heavily weighted. There also exists literature on optimizing server compute load dynamically based on server demand to minimize energy consumption (Parolini, Sinopoli, & Krough, 2008) but the focus of this literature is on temporary, dynamic reductions and not base line reductions.

The case studies reported by the EPA show the potential impact of holistic energy reduction approaches (Lucasfilm, 2008) (Sybase, 2009) (Verizon, 2008). These studies indicate where sources of energy reduction exist, the costs that when into the reduction and report the improvements achieved. However, they do not address the policy/program implementation nor discuss alternative policy approaches. Additionally, there are governmental funded programs to establish Data Center assessment and benchmarking processes (EPA, 2010). These provide insight as to a prescriptive approach of what process should be used to reduce energy, but implementation strategies and alternatives are not handled.

From a System Dynamics perspective, the work performed by Hsueh (Hsueh, 2010) on gaining insight from modeling on qualitative systems thinking provides a good background for how to approach the modeling implementation of highly aggregate system archetypes. This methodology was applied to the shifting the burden archetype (Senge, 1990), and helped guide the approach used in this paper which implements the limits to growth archetype (Senge, 1990). Additionally, when identifying potential limits and how they possibly could be managed, the analysis by Ceresia (Ceresia, 2009) on balanced scorecard dynamics and (Vittorio, 2002) on critical drivers for management, provide insight on how some of the concerns discussed in this paper can be addressed.

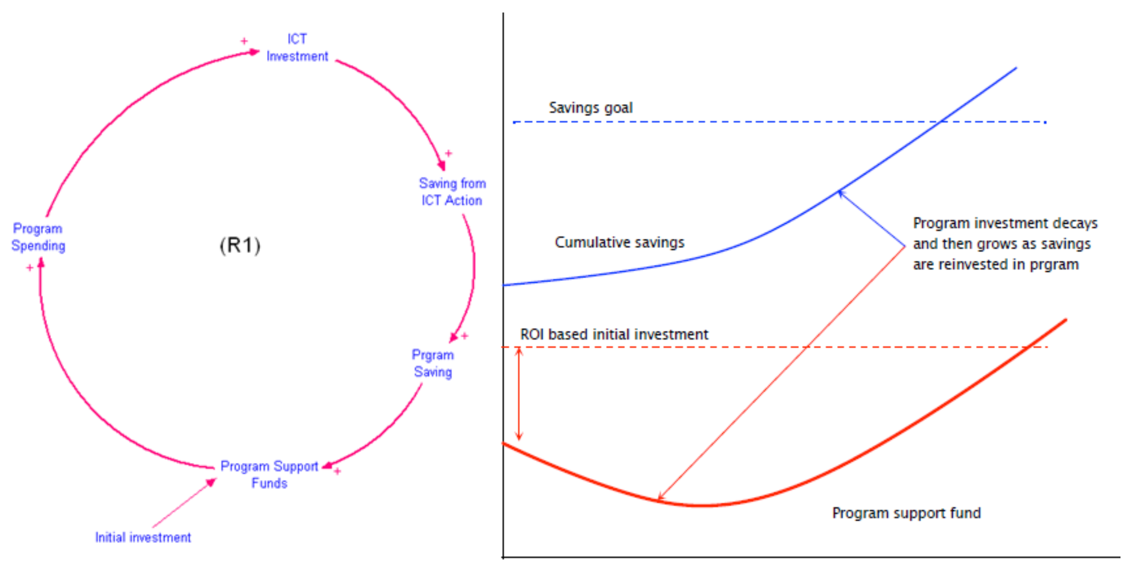
Problem approaches

This paper will explore three methodologies for solving this problem - one utilizing simple static ROI analysis and two utilizing insight from system dynamics. The system dynamics insight is gained from recognizing that understanding and leveraging the compounding structure and the limits to growth archetype (Senge, 1990) can inform the selection of policies that can yield greater success when compared to ROI analysis. The three approaches and reference modes are shown in Figure 2, Figure 3 and Figure 4.



(a) (b)

Figure 2 Approach 1, Static ROI.



(a) (b)

Figure 3 Approach 2, Reinforcing loop

(a)

(b)

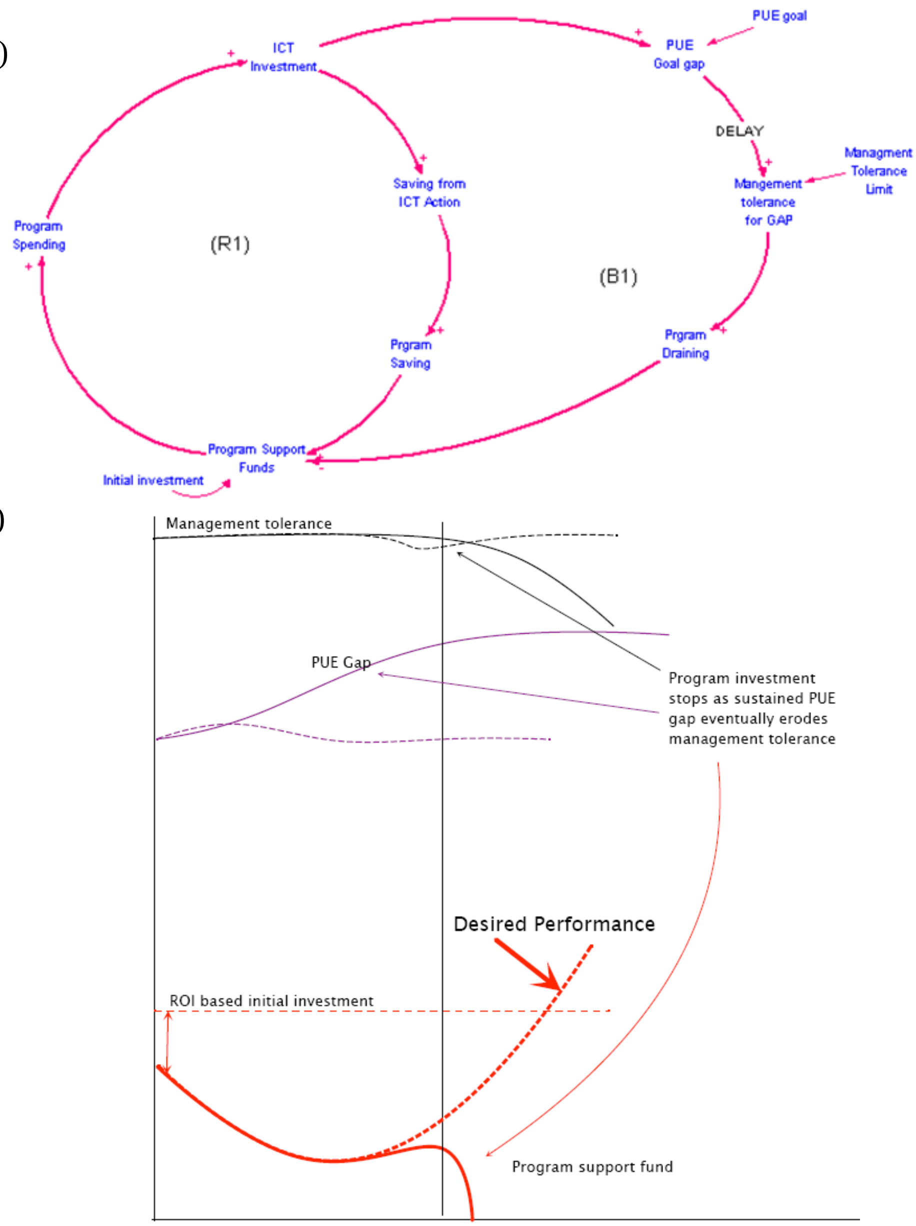


Figure 4 Reinforcing loop with limiting balancing loop

The above structures are shown with only the energy reduction associated with ICT actions shown. A more complete view is shown by the causal loop diagram of depicted Figure 5 and the associated dynamic hypothesis of Figure 6.

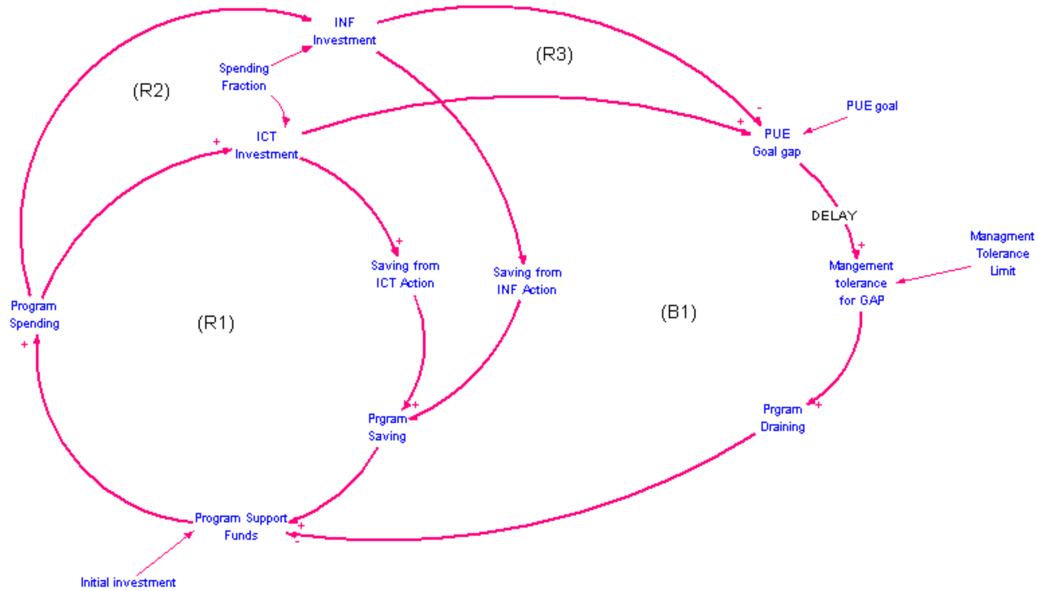


Figure 5 Causal loop diagram showing addition of actions associated with infrastructure. Reference mode is similar to figure 4b

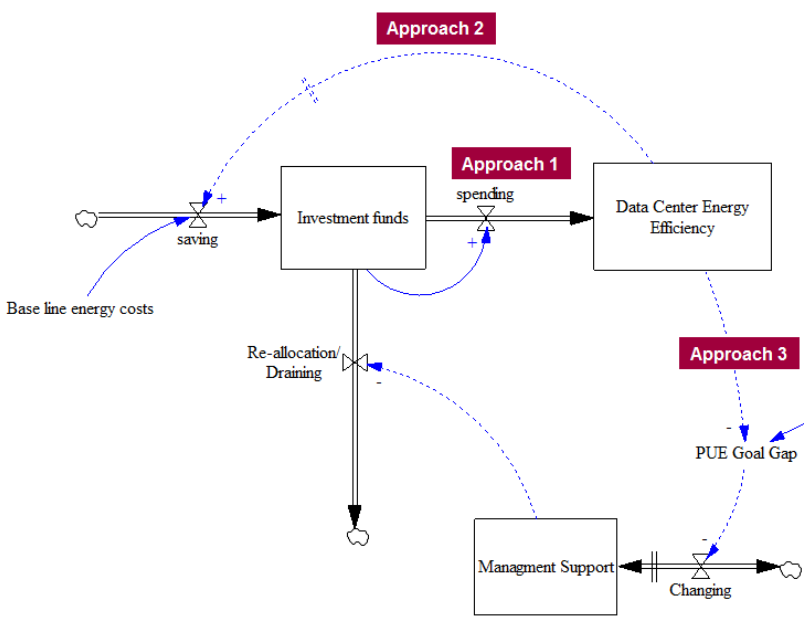


Figure 6 An aggregate dynamic hypothesis showing the paths that become active as the buildup occurs from Approach 1 to 3.

Each of these structures are modeled and three questions are explored to help inform policy decision.

Question 1: What is the impact of investing in infrastructure activities versus ICT activities?

Question 2: What are the implications changing the initial investment?

Question 3: What are the implications of investment sequence when a limiting loop exists?

Open loop, Static ROI approach

The open loop causal loop diagram is shown in Figure 5 where the investment in ICT activities and INF activities both contribute to energy reduction losses. However, as described in table 1, there is a limitation that no more than a 60% reduction can occur in the infrastructure and no more than 40% reduction can occur in ICT no matter how much is invested. This establishes an idealized split of 60/40 between investments in INF versus ICT activities. To explore the limits of this investment split, the split is varied while holding all else constant. The model is setup such that the maximum investment amount is capped such that unnecessary funds are not allocated that cannot be spent (ie allocating 1 pu funding when only INF investments are being made which has a max of 0.6 pu.

Figure 7 shows results of the impact on investment fund, PUE, cumulative saving and total achieved energy reduction as the investment between INF and ICT spending is varied. These results address question 1 above.

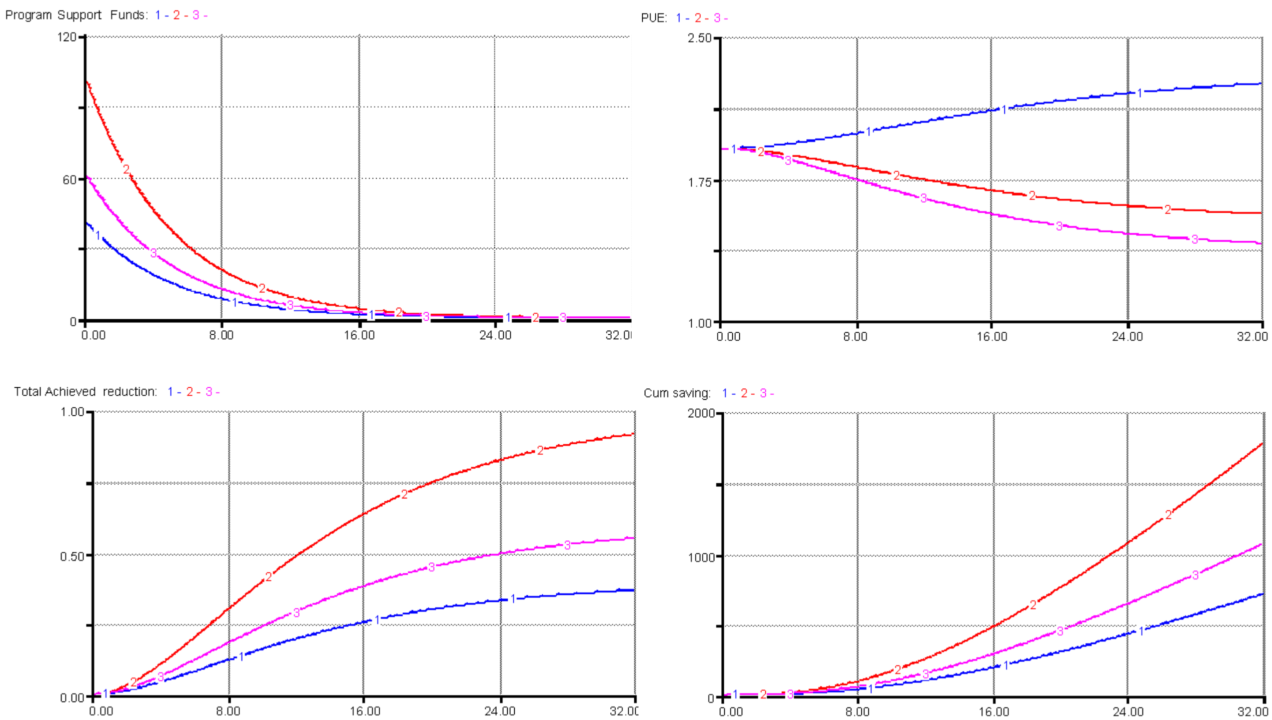


Figure 7 Plot line 1 (Blue) INF = 0, ICT =0.4 reduction; Plot line 2 (Red) INF = 0.6, ICT = 0.4 reduction; Plot line 3(Magenta) INF = 1, ICT = 0 reductions.

The results indicate that for this specific example set of inputs (table 1), that the lowest PUE does not correspond to the investment profile that targets the greatest savings. Additionally, an investment profile that only works on ICT reduction will increase the PUE rating.

Insight to Question 1: An investment profile that maximizes energy loss opportunity will not minimize PUE (unless energy loss opportunity only exists in infrastructure).

To address question 2, the same model was run, but with INF and ICT spending split set to its optimum and the initial investment is varied from its maximum of 1 to a low of 0.2. The results are shown in Figure 8.

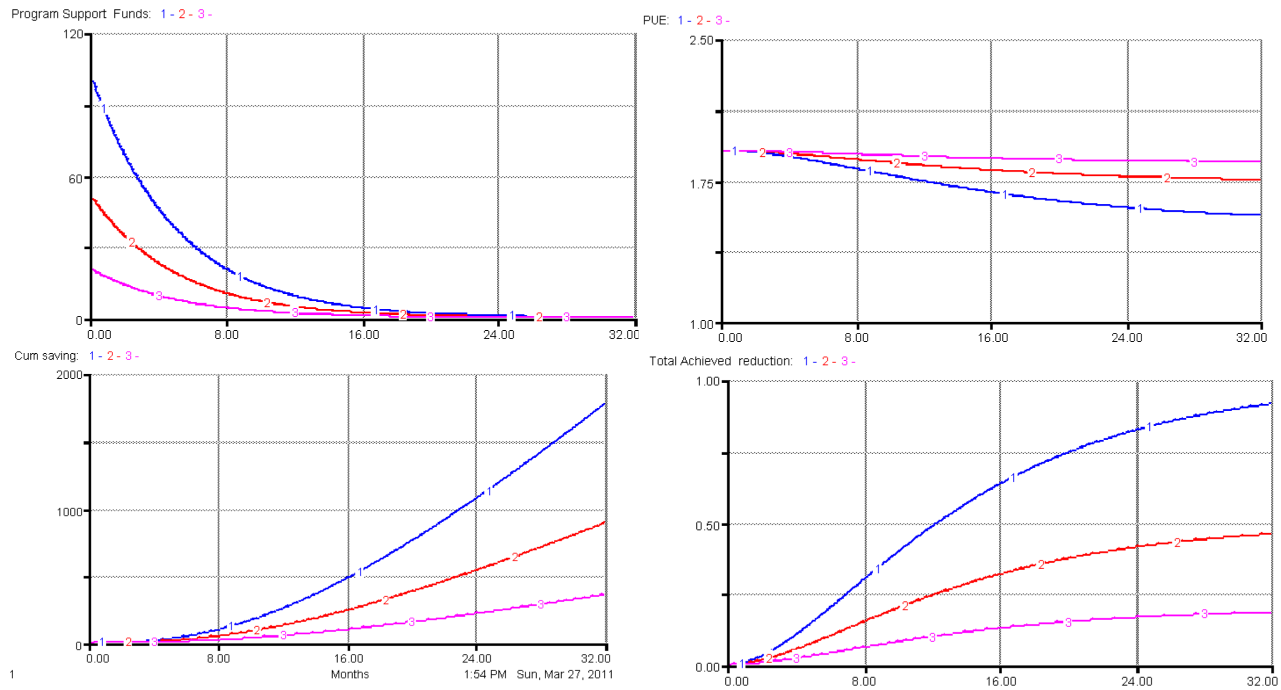


Figure 8 Plot line 1 (Blue) Initial investment = 1; Plot line 2 (Red) Initial investment = 0.5; Plot line 3(Magenta) Initial investment = 0.2

These results show that as the initial investment is decreased, the overall performance of PUE, savings and opportunity addressed all decline.

Insight to Question 2: The lower the initial investment, the lower the gains.

Question 3 cannot be answered with this simple model because there is no loop structure involved.

System dynamic approach (Reinforcing)

The performance result to the open loop approach is completely dependent on the amount of initial investment as answered in question 2. With the open loop approach, the savings generated are returned to an entity exogenous to the model. The implications are that the request for initial investment (also exogenous), and the application of this investment toward infrastructure and/or ICT activities must be made up front. This can lead to management request that may be too large to support or too low to satisfy expectations for improvements. One way to mitigate these risks is the "bootstrap" the investment by allowing the savings from the program to be re-invested and thus self fund the program until the improvement limits are reached.

The funds to re-invest come from savings in both infrastructure and ICT investments as shown in the causal loop diagram of Figure 9.

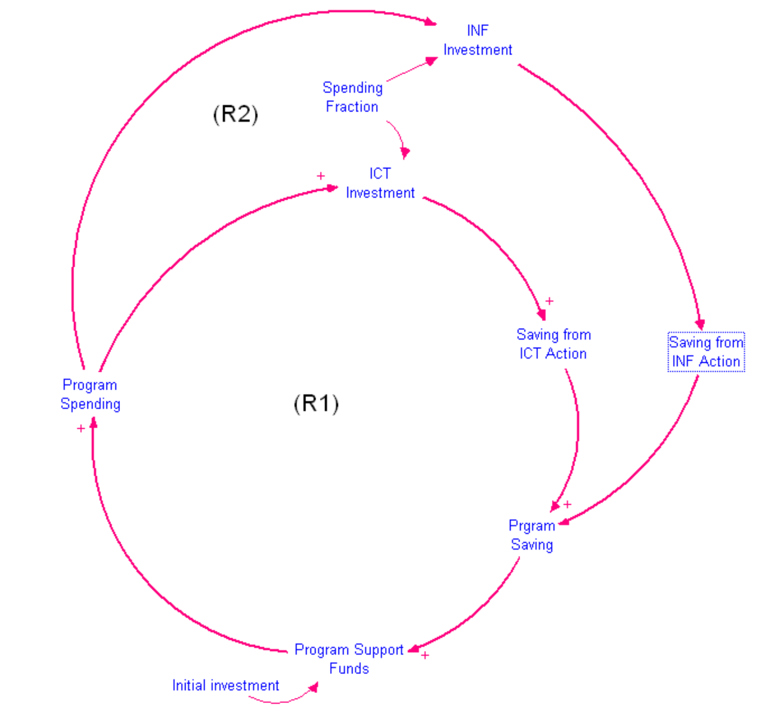


Figure 9 Reinforcing loop causal loop diagram

An exogenous input is used to load the initial Program Support Funds. This structure has two reinforcing loops R1 and R2 and the both operate in the same manner. An increase in the Program Support Funds, increases the Program spending which increases the investment in ICT (and/or INF) which increases the savings from ICT (and/or INF) action increasing the overall program savings which in turn further increase the Program support funds. There are delays related to how fast improvement activities can be implemented to yield savings and other real world factors that are more expressly shown in the stock/flow structure of Figure 10.

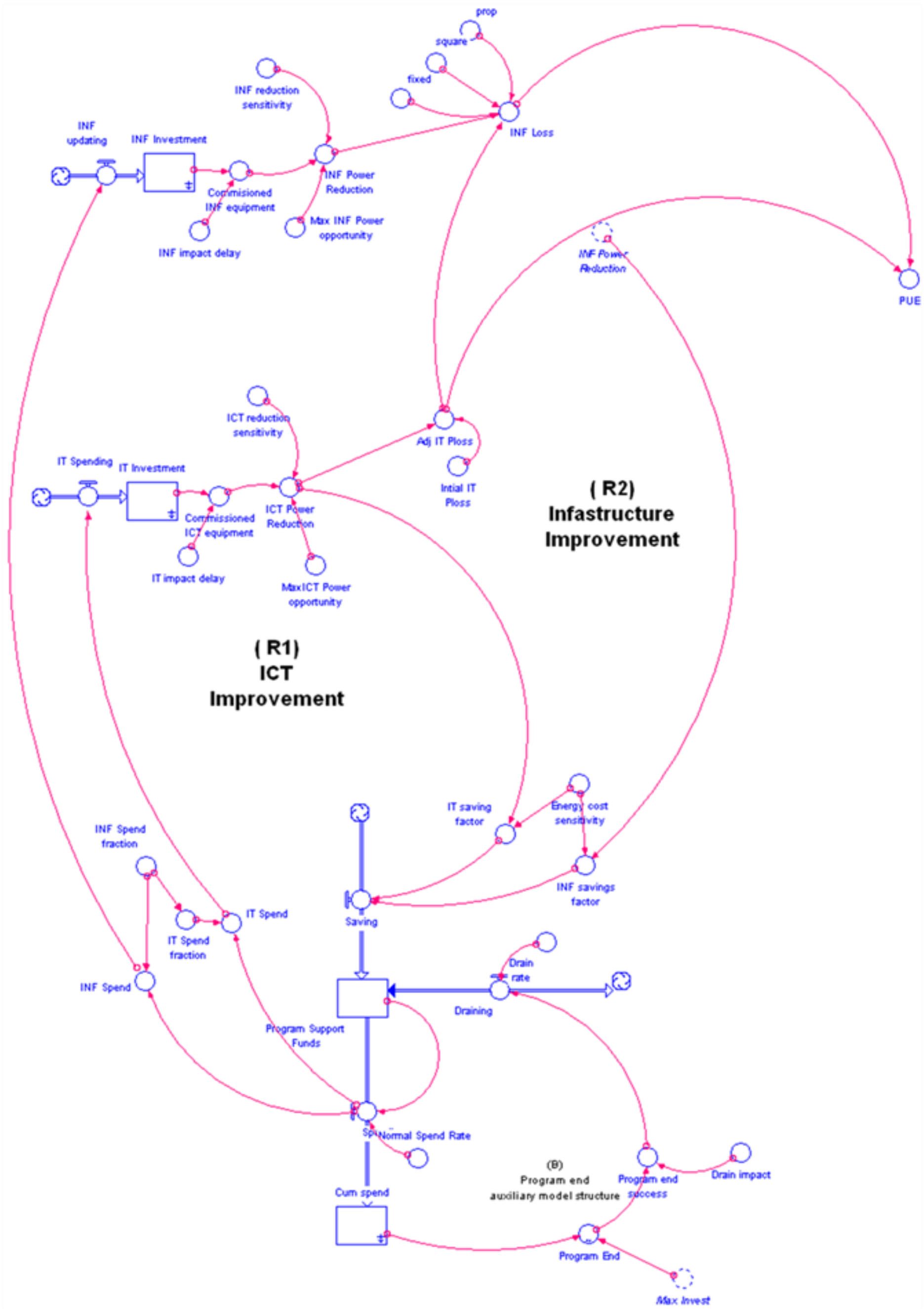


Figure 10 Dual reinforcing loops to grow initial investment

Figure 10 represents a highly aggregated view of the investment dynamics in a self funding scheme for energy improvement in a data center. Because this structure is self funding and will continue to grow the Program Support Funds as long as there is a savings rate, a mechanism needed to be added address the problem of when to terminate the investment program. To facilitate this and auxiliary loop was added to drain the funds once the cumulative spending had reached the point where no further improvements in infrastructure or ICT can be made. This loop is designed to not interfere with the dynamics when the fund is naturally growing or shrinking.

A first order built-in smooth function is used (Commissioned ICT (or INF) equipment) along with and associated impact delay to model the delay between spending and operational changes. Also the loop impacts are both capped by a MIN function that won't allow the power reductions to exceed either the maximum INF or ICT power reduction. This also has the impact of limiting the maximum savings rate. With this model we can now test Question 1 and 2.

Figure 11 shows the results of the model as the fraction of investment in INF vs ICT is made with an initial investment equal to the maximum possible (same approach as with the Open loop). Figure 11 shows similar results to the open loop model in that the lowest PUE does not occur with the investment ratio that maximizes energy reduction (plot line 2). However, one difference is in the time scale. In the reinforcing loop structure after 32 units of time (months), the system has reached steady state and cumulative savings have grown from approximately 1500 to over 2000 for the closed loop case. This is due to the reinvesting structure associated with the Program Support Fund stock. After slack is exogenously added to the stock the fund initially begins to decay as spending increases and the delays to implement the savings have yet to take place. But as the savings come on line, the slack is replenished by the savings flow. If

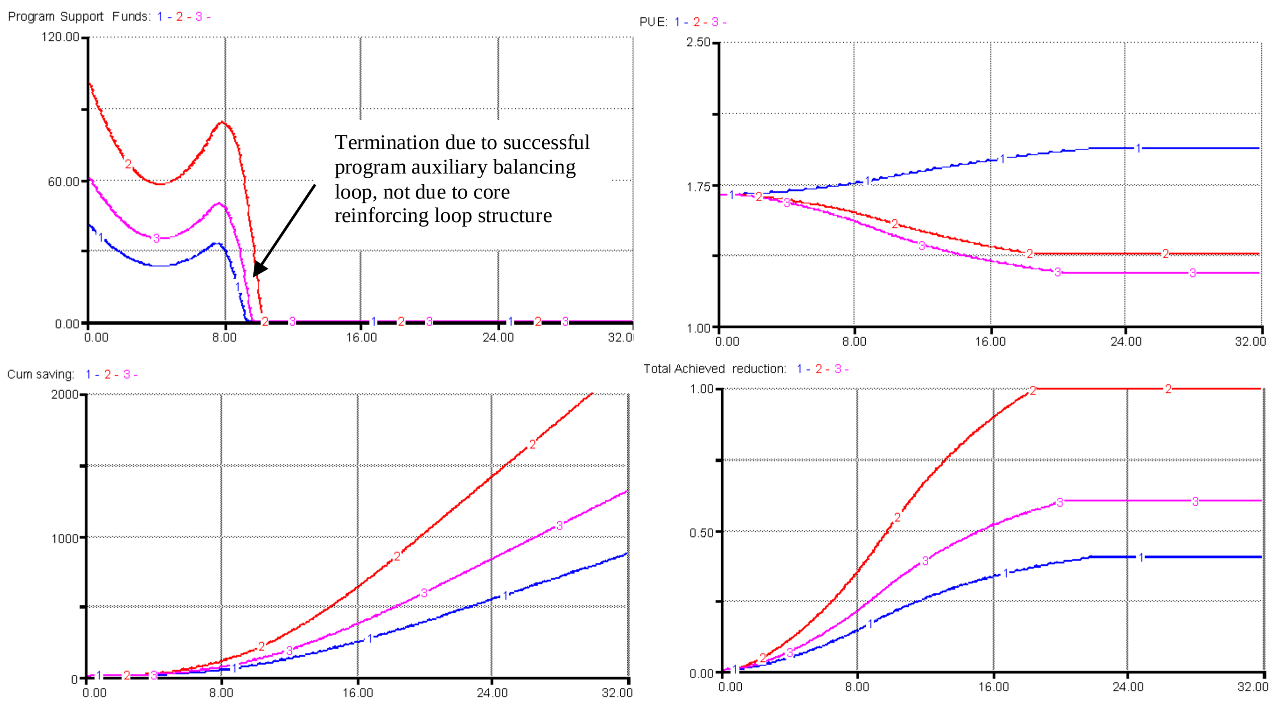


Figure 11 Plot line 1 (Blue) INF = 0, ICT =0.4 reduction; Plot line 2 (Red) INF = 0.6, ICT = 0.4 reduction; Plot line 3(Magenta) INF = 1, ICT = 0 reductions.

the savings program was not terminated by the successful end auxiliary structure, the stock would reach a steady state value when the savings flow is equal to the spend flow. This occurs when the stock equal to the spend rate times the normal spend rate or equivalently in equilibrium the savings rate times the normal spend rate. This steady state slack would represent the maximum return on investment (initial slack). However, for this model, the maximum return is not allowed to occur because after break-even, the additional savings are drained off to other funds through the increase of the draining flow.

Insight to Question 1: As with an open loop investment approach, an investment profile that maximizes energy loss opportunity will not minimize PUE (unless energy loss opportunity only exists in infrastructure). However, because funds are directed back into the energy reduction fund, the time for the investment to return the initial investment is shortened (for a comparative spend rate as a function of the fund size).

To address question 2, again, the same model was run, but with INF and ICT spending split set 0.6 (optimal) and the initial investment was varied from 1 to a low of 0.2. The results are shown in Figure 12.

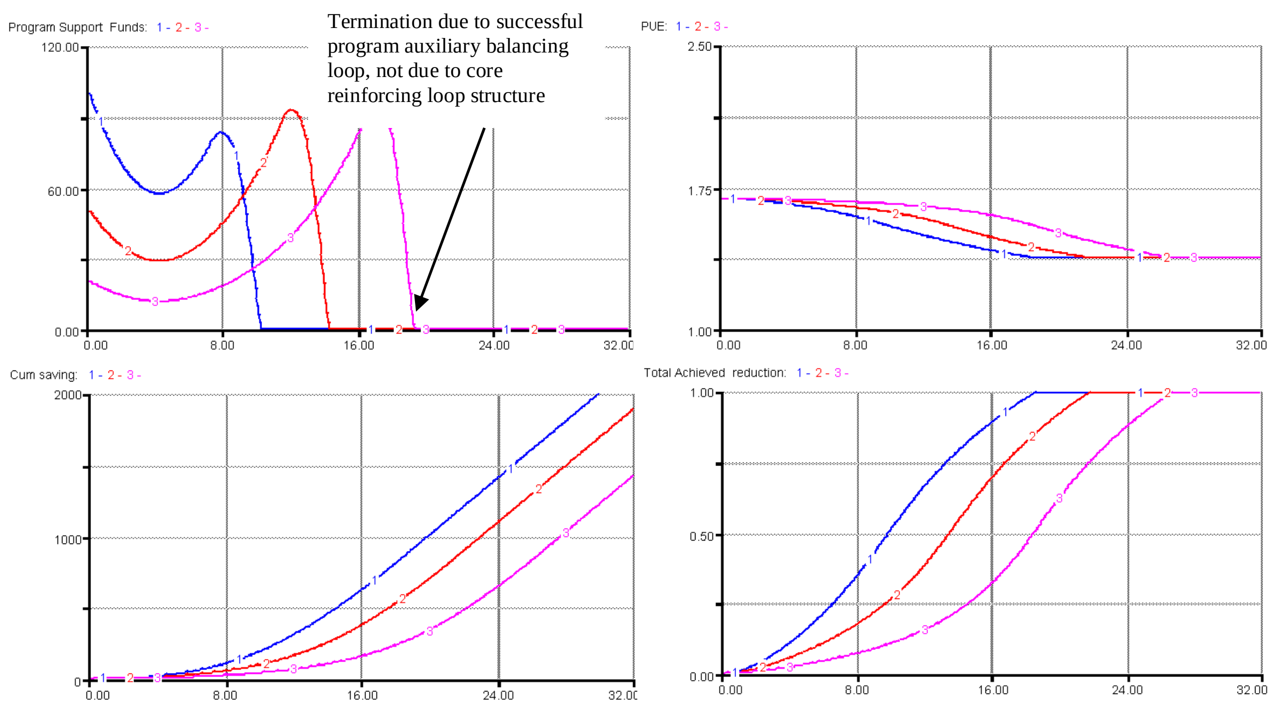


Figure 12 Plot line 1 (Blue) Initial investment = 1; Plot line 2 (Red) Initial investment = 0.5; Plot line 3(Magenta) Initial investment = 0.2

Figure 12 shows very different results from the open loop model. As indicated in the plots showing the Program Support Fund output, we notice no matter how small the initial investment is, the fund is grown to a point where all possible investments are fully funded. The only distinction that occurs is the time it takes to reach this point. This is driven by the fact that the lower the slack, the lower the spend rate (because it is proportional the slack size), this means that initially, the spend rate is low driving a slower saving rate after a commissioning delay. While the savings rate may be lower for lower initial investment, it eventually grows as the slack grows ultimately fully funding the program. Because all of the investments are fully funded (due to the optimum INF/ICT funding split) the final PUE is the same for all runs. Because of the slower saving rate, the cumulative savings at any point in time will be lower for as the initial investment is lowered.

Insight to Question 2: Because of the reinforcing loop structure any initial investment will lead to a fully funded program due to the reinvestment of savings. Practically speaking, the initial investment needs to be large enough to cover startup costs, but no greater.

System dynamic approach (Limits to growth)

Finally, the model is expanded to include the behavior that occurs when an energy reduction program fails to meet management expectations. In this case, we model management expectations associated with the PUE metric to demonstrate the potential complexities that arise due to path dependencies of the investment strategies. The causal loop diagram showing this new structure is shown previously in Figure 5. The additional structure used to model this behavior is shown in Figure 13.

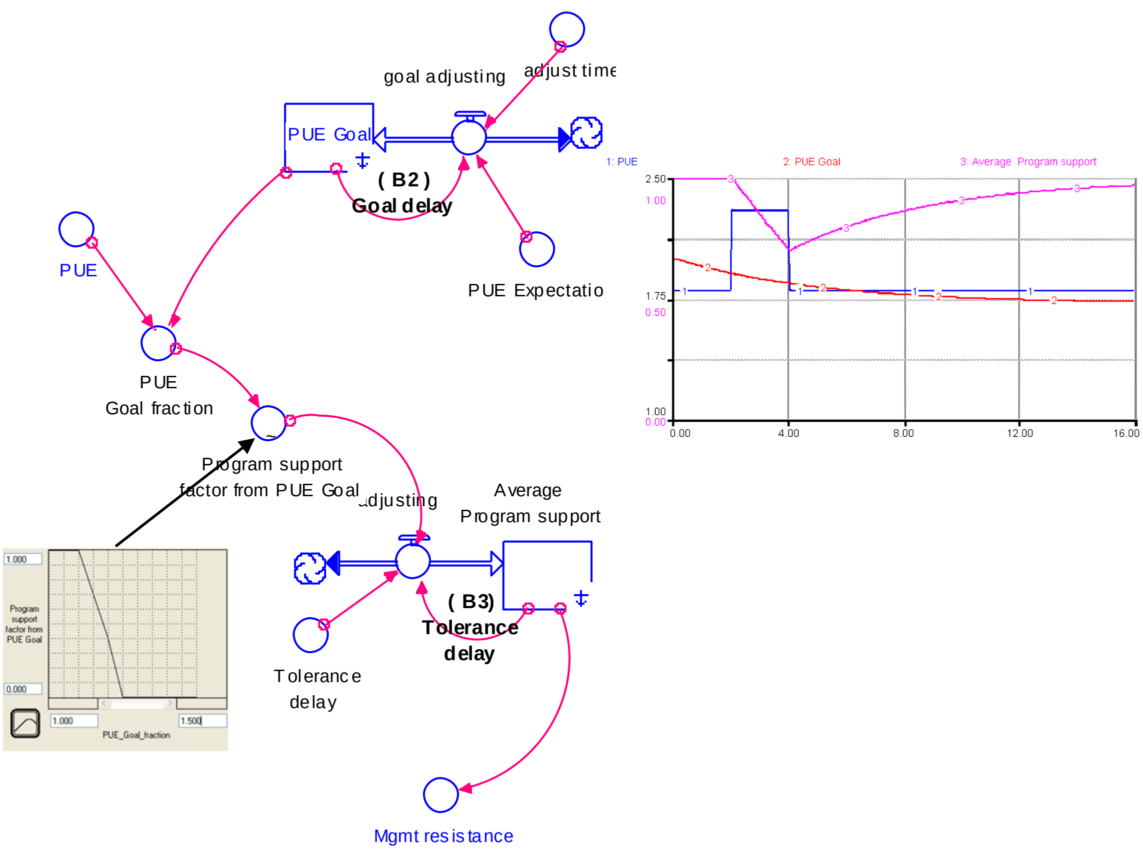


Figure 13 Structure depicting the delay in establishing the new goal and the tolerance delay. Inset shows the delayed response to a step function in the PUE allowing for short duration gaps to exist before management can lose support for the program.

Because the reinforcing loops that drive the saving and spending rates are the same as previously analyzed, the insight for this model to answer question 1 and question 2 are the same. However, this model can be used to explicitly test question 3. Given the fact that investment in ICT savings alone will increase PUE, there is a possibility that if the investment policy was to first invest in ICT reductions followed by INF reduction efforts, the risk exists that the increased PUE will exceed the tolerance level management has for the program and the program will prematurely end. To test this hypothesis, two sequences are shown.

Sequence 1: Invest exclusively in infrastructure reduction for 4 months followed by investment in ICT.

Sequence 2: Invest exclusively in ICT reduction for four months followed by investment in INF reduction.

The results of these two sequences are shown in Figure 14.

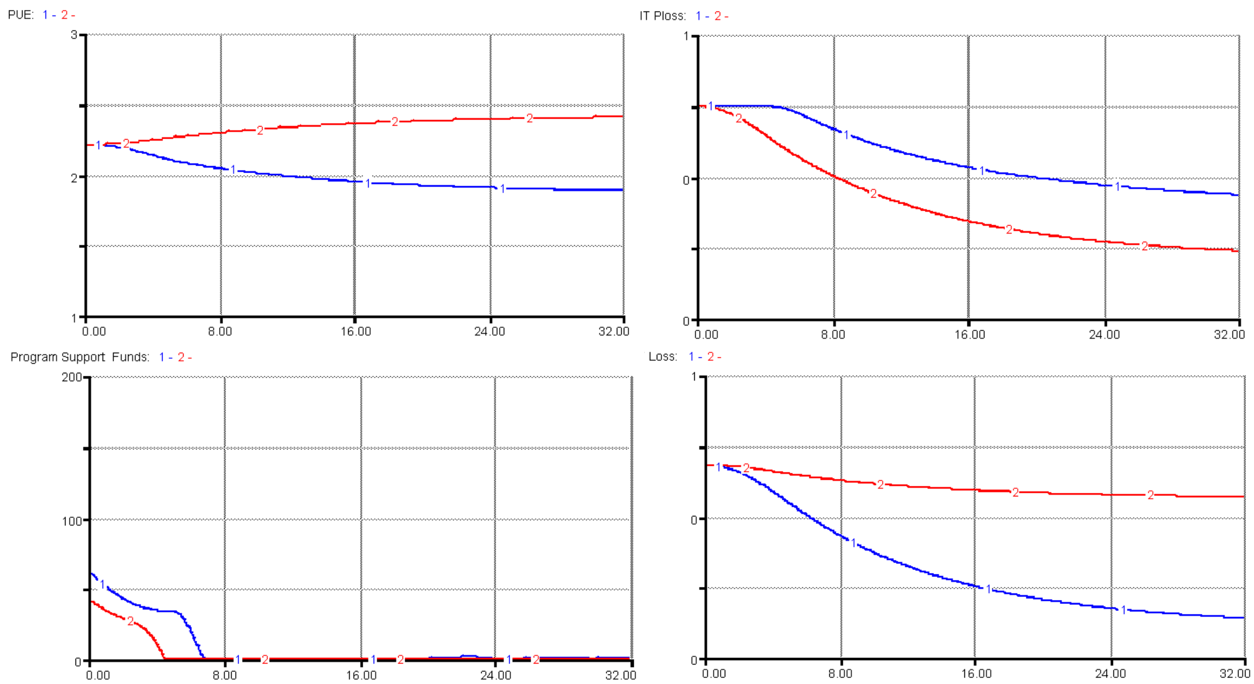


Figure 14 Plot line 1 (Blue) Sequence 1; Plot line 2 (Red) Sequence 2. PUE goal is 1.72.

For sequence 1, where the investment is first biased toward infrastructure investment, the INF losses fall, but the ICT losses remain. This action causes the PUE to fall. When the investment swaps from full INF reduction to full ICT investment, both the ICT and INF losses fall and the decrease in PUE slows. Because the PUE is below or fast approaching its goal, the program can fully fund all investments and successfully terminate. In this sequence, the limit of the balance loop is never reached.

For sequence 2, where the investment is first biased toward ICT investment, both the ICT and INF losses fall. This action causes the PUE to rise. The rise in the PUE reaches the limit of management tolerance for being above the program goal and the program is drained of its fund before any further investments are made.

Insight to Question 3: Because of the limit to growth structure, path of investment is important in ensuring the limiting action is never reached and the program will support full energy reduction efforts. An order of INF investment followed by ICT investment is more likely to succeed given a limiting structure governed by the PUE indicator.

Summary

The characteristics of energy flow and a key efficiency metric for data centers have many connecting parameters linking the overall performance of an energy reduction program. By comparing a program operating in an open loop manner to one that has its policy guided by recognition of a loop based program additional benefits and insight can be gained to support an energy efficiency program. These insights are characterized in Table 3.

Table 3 Summary of insights gained

	Open Loop	**Reinforcing loop**	**Limits to growth**
Question 1: What is the impact of investing in infrastructure activities versus ICT activities?	An investment profile that maximizes energy loss opportunity will not minimize PUE	Same as Open Loop PLUS because funds are directed back into the energy reduction fund, the time for the investment to return the initial investment is shortened	Same as Reinforcing loop
Question 2: What are the implications changing the initial investment?	The lower the initial investment, the lower the gains.	Because of the reinforcing loop structure any initial investment will lead to a fully funded program due to the reinvestment of savings.	Same as Reinforcing loop
Question 3: What are the implications of investment sequence when a limiting loop exists?	NA	NA	Because of the limit to growth structure, path of investment is important in ensuring the limiting action is never reached and the program will support full energy reduction efforts. An order of INF investment followed by ICT investment is more likely to succeed given a limiting structure governed by the PUE indicator.

Conclusion

This paper demonstrates how the thought process of system dynamics can be applied to a data center energy reduction program. A popular metric being adopted to benchmark data centers from different enterprises, complicates the program implementation by having a characteristic that is not always monotonically improving with a program that has sustained energy reduction. This characteristic is shown to potentially cause a reference mode between program funding and management perception of success that mimics the limits to growth archetype. By understanding the behavior of the parameters of this structure in the data center context, it was shown that a policy implementation that biases initial action toward first acting on parameters to improve PUE and energy savings are more likely to succeed.

This paper only covers an aggregate perspective of data center behavior and answer only three specific questions. Additional research should go into behavior of the model as different energy parameters such as, maximum power saving limits, ICT exogenous growth and other high impact parameters are sensitivity tested. Also, research based on interviews with data center energy reduction programs would be useful in identifying real examples of program behavior over time which would lead to more insightful models. Finally, a more general study of what are the dynamic structures that tend to drive a bias toward the more tangible and thus comparable measures as applied to any domain warranted.

References

Belady, C. L. (2007, February). In the data center, power and cooling costs more than the IT equipment it supports. *Electronics Cooling* , pp. 24-27.

Ceresia, F. (2009). A Model of Goal Dynamics in Organizations: Goal Setting, Goal Commitment,Training and Management by Objectives. *Proceedings of the 2009 Conference on System Sciences, Management Sciences & System Dynamics.* Tongji University, Shanghai, China.

EPA. (2010). Energy Star for Data Centers. *The Green Grid Technical Forum.* US Environmental Protection Agency, Energy Star.

EPA. (2007). *Report to Congress on Server and Data Center Energy Efficiency.* US Environmental Protection Agency, Energy Star Program Public Laaw 109-431 August 2.

Hsueh, J. (2010). What Can Formal Modeling add to Qualitative Systems Thinking. *The 28th International System Dynamics Conference.* Seoul, Korea.

Koomey, J. (2008). Worldwide electricity used in data centers. *Environmental Research Letters vol 3, no 034008. September 23* .

Lawrence Berkley National Lab, Ancis Inc. (2010). *DCEP Program Energy Training-Assessment Process Manual.* US Department of Energy.

Lucasfilm. (2008). *Save Energy Now, Data Center Assessment Summary - Lucasfilm.* US Department of Energy, Energy Efficiency and Renewable Energy,Industrial Technologies Program.

Patterson, M. A. (2006). From UPS to Silicon: an End-to-End Evaluation of Data Center Efficiency. *Enterprise Server and Data Centers: Opportunities for Energy Savings Conference.* Santa Clara, CA, January 31.

Rasmussen, N. (2006). *Electrical Efficiency Modeling of Data Centers.* APC, White paper #113.

Senge. (1990). *The Fifth Discipline.*

Sybase. (2009). *Save Energy Now, Data Center Energy Efficiency - Sybase.* US Department of Energy, Energy Efficiency and Renewable Energy, Industrial Technologies Program.

The Green Grid. (2007). *Green Grid Metrics: Describing Data Center Power Efficiency.* The Green Grid, Technical Committee. Feruary 20.

US DOE. (2008). *Data Center Assesment Tool DC Pro, Master List of Energy Efficiency Actions.* US Department of Energy, Save Energy Now.

Verizon. (2008). *Save Energy Now, Data Centter Assessment Summary - Verizon.* US Department of Energy, Energy Efficiency and Renewable Energy, Industrial Technologies Program.