Shift factors in ERCOT congestion pricing

Ross Baldick*
March 5, 2003

1. Introduction

In this paper, I discuss the ERCOT zonal and local congestion management scheme and analyze various economic and engineering aspects of its implementation. I investigate the economic inefficiencies that are due to the simplified ERCOT commercial transmission model and will demonstrate that short-term efficiency is compromised by the ERCOT commercial transmission model. I will also comment on the effects on long-term efficiency.

In section 2, I begin with a brief discussion of the role of congestion pricing. In section 3, I first summarize the ERCOT commercial transmission model, introducing and explaining "shift factors" and their role in the ERCOT commercial model. I then present two important assumptions underlying the ERCOT commercial model and then discuss these assumptions in more detail.

Section 3 provides the background material for the development in section 4, where I show that the ERCOT implementation substantively violates the assumptions underlying the ERCOT commercial model. In section 5, I discuss another implicit assumption in the model that is also violated in the ERCOT implementation. In section 6, I discuss remedies to the model and conclude.

2. Objectives of Congestion Pricing

The objectives of congestion pricing include (Oren, 2002):[1]

1. efficient use of scarce transmission resources,
2. provision of economic signals for dispatch,
3. provision of economic signals for location of generation and demand, and
4. provision of economic signals for transmission expansion.

A theoretically ideal congestion pricing approach would provide incentives for efficient use of transmission, which means efficient rationing when the demand for transmission services exceeds the available transmission capability. To the extent that the prices for

* The author is with the Department of Electrical and Computer Engineering at the University of Texas at Austin, baldick@ece.utexas.edu. A previous version of this paper was prepared on behalf on Competitive Power Advocates for presentation to the Texas Public Utility Commission on January 14, 2003. The author would like to thank Krishnan Dixit of the University of Texas for help preparing figures 7-14 and for the data in section 4.4.

[1] Shmuel S. Oren, "Congestion Pricing and Transmission Rights," Presented at the PUCT Public Workshop on Elements of Market Design, Austin, Texas, November 1, 2002, page 3.

congestion differ from the efficient level, these incentives will be distorted and poor economic decisions will be made.

In the short-run, poor economic decisions related to non-ideal congestion pricing involve inefficient dispatch. Two particular examples of inefficient dispatch related to transmission resources are:

1. solving transmission congestion using generation that is more costly than necessary, and
2. operating a transmission line below its capacity when there was cheaper generation available at the sending end that could displace more expensive generation at the receiving end.

In the long-run, such poor economic decisions can also encourage bad choices for the location of new generation, load, and transmission.

In practice, every congestion pricing approach will differ from the ideal. From a practical perspective, it is desirable that any deviation from efficient pricing be relatively small so that the distortion of the economic decisions will be relatively minor. The desire to minimize the distortion of the economic signal must be balanced against the costs of implementation and operation of the congestion pricing approach.

One axis of this trade-off is the complexity of the commercial model. In ERCOT, for clearing of transmission congestion in the daily scheduling process, a simplified commercial transmission model is adopted that approximates the multi-thousand bus and multi-thousand line ERCOT system by a four zone model with three important transmission corridors. That is, in ERCOT the implicit trade-off has been towards simplicity of implementation. I will discuss this commercial transmission model in more detail in section 3.

3. Description of ERCOT commercial transmission model

The ERCOT commercial transmission model is determined in a several step process. The first step is the annual identification of a (hopefully) relatively small number of "commercially significant" transmission constraints (CSCs) in ERCOT. These CSCs are chosen to represent the limitations on moving power within ERCOT. For 2003, ERCOT has determined three CSCs. In sections 3 and 4, I will take as given the process for determining the CSCs; however, I will revisit the definition of CSCs in section 5.

Whenever patterns of generation or demand vary, the flows on the CSCs are affected according to Kirchhoff's laws.[2] If the flow on any CSC is at its limit, no further increase on the flow can be allowed. If Qualified Scheduling Entities (QSEs) in ERCOT schedule more flow on a CSC than its limit then the schedules must be adjusted to bring the flows to within the CSC limit. That is, the ability to flow power on the CSC can be a scarce resource. Moreover, each transaction from a generator to a demand affects the CSCs differently.

Determination of the appropriate amount of capacity needed to support a transaction necessitates an evaluation of how each transaction contributes to flow on the CSCs. Consequently, the second step in developing the ERCOT commercial model is the

[2] Kirchhoff's laws are laws of physics specialized to electrical circuits. See, for example, Arthur R. Bergen and Vijay Vittal, *Power Systems Analysis*, Second Edition, 2000, Prentice-Hall, Upper Saddle River, New Jersey.

evaluation of these contributions, which are termed "power transfer distribution factors" (PTDFs) or "shift factors."[3] Section 3.1 provides the details about calculations of the shift factors.

The most direct approach to using the shift factors would be for each transaction from generation to demand to be charged based on its shift factor and the cost of congestion on the CSCs. Assuming that the cost of congestion on the CSC were priced correctly with an appropriate "shadow price" determined by an auction or other mechanism, then the product of the shadow price times the shift factor would be the efficient or ideal price for the transaction.[4]

However, instead of using the actual shift factors, ERCOT takes a third step where busses having similar shift factors are clustered into a "zone," the shift factors are averaged across the zone, and the zonal average shift factor is used for each generator in a given zone. (The average shift factors are calculated as *generation-weighted* averages across all generation in the zone. For convenience in the following discussion, I will refer to "zonal generation-weighted average" as "zonal average." The "zonal average shift factor" will be in contrast to the "actual shift factor" for each bus and CSC.) Naturally, using the zonal average shift factors introduces an error compared to using the actual shift factors.

In section 3.2, I present two assumptions that would, if satisfied, justify the use of the zonal average shift factors instead of the actual shift factors. I will refer to these as ERCOT assumptions 1 and 2, respectively. (I will introduce and discuss a third assumption in the ERCOT model in section 5.)

In section 3.3, I then describe the characteristics of an *ideal* system where ERCOT assumptions 1 and 2 are exactly satisfied. This system will be unrealistic, but will serve as a basis to understand and evaluate the economic and engineering implications of practical systems where the assumptions are not exactly satisfied.

I then observe in section 3.4 that all *practical* systems violate ERCOT assumptions 1 and 2 and identify that the key issue is whether the assumptions are satisfied sufficiently and accurately enough that the implications of violation are negligible. I describe qualitatively the case for a realistic system.

In section 4, I will then focus on whether or not ERCOT assumptions 1 and 2 are satisfied sufficiently and accurately enough in the ERCOT system so that the error

---

[3] Sometimes, the terms PTDF and shift factor are used in slightly different senses; however, as will be discussed in section 3.1, the terms always refer to the ratio of a change in flow on a line or lines to the change in injection and withdrawal at a pair or multiple pairs of buses. I will use the terms PTDF and shift factor interchangeably in this paper.

[4] For more detailed discussion and analysis of the efficiency of such prices, see, for example, Hung-po Chao and Stephen Peck, "A Market Mechanism for Electric Power Transmission," *Journal of Regulatory Economics*, 10(1):25-59, July 1996 and Hung-po Chao, Stephen Peck, Shmuel Oren, and Robert Wilson, "Flow-based Transmission Rights and Congestion Management," *The Electricity Journal*, 13(8):38-58, 2000. Outside of Texas, the term "flowgate" is more commonly used instead of CSC and the literature reflects this more common usage. It should also be noted that there is some disagreement about whether the CSC approach can, even in principle, provide efficient price signals. See, for example Larry E. Ruff, "Flowgates, Contingency-Constrained Dispatch, and Transmission Rights," *The Electricity Journal*, 14(1):34-55, January/February 2001 and William Hogan, "Flowgate Rights and Wrongs," Manuscript, John F. Kennedy School of Government, Harvard University, August 2000. However, except for the discussion in section 5, I will take it as given that the CSC approach can, at least in principle, provide efficient price signals.

committed results in negligibly small effects on the economic efficiency of the ERCOT market.

3.1 PTDFs / Shift Factors

To calculate a PTDF (or shift factor), we consider an injection of power at one bus and a withdrawal at another bus. A shift factor to a particular line is the effect of this injection and withdrawal on the flow on the line. That is, the shift factor is the ratio of the change in the flow on a line to a change in the injection and a corresponding change in the withdrawal.

For convenience in calculations, the "reference bus" is often chosen as the notional point of withdrawal corresponding to any injection. In ERCOT, this bus is Norwood in the North Zone and all of the shift factors are calculated for withdrawal at Norwood. For a transaction that involves injection at a point A and withdrawal at a point B, neither of which are Norwood, we can calculate the appropriate shift factor by subtracting the shift factor for injection at B from the shift factor for injection at A. Effectively, we are considering the combination of two transactions:

1. injection at A and withdrawal at Norwood, and
2. injection at Norwood and withdrawal at B.

The simultaneous withdrawal and injection at Norwood of the same power results in no net injection or withdrawal at Norwood. That is, we have obtained the shift factor for injection at A and withdrawal at B. For this reason, in the calculations that follow, we will always be subtracting two shift factors having Norwood as point of withdrawal to obtain the shift factor for the transaction of interest.[5]

As mentioned above, a cluster analysis is performed on the shift factors to identify "zones" where the shift factors are similar. A zonal average shift factor is then calculated for each zone and these are then used in the ERCOT commercial model as the shift factor for all busses in the zone independent of the actual shift factor. The assumptions behind this averaging are introduced in the next section.

3.2 Assumptions in the ERCOT commercial model

There are two implicit assumptions in the use of zonal average shift factors in the ERCOT commercial transmission model (Oren 2002).[6] These assumptions are (Oren 2002):

1. "All generators in a zone have the same PTDFs with respect to CSC[s],"
2. Given a fixed total generation and demand in one zone, the sharing of generation within that zone "does not impact…congestion in other zones."

---

[5] Strictly speaking, using the difference between the shift factors is only correct in the ideal case that there are no losses; however, the error is typically small in the realistic case of losses. Moreover, the shift factors used in ERCOT are so-called "DC" shift factors that are calculated ignoring the effect of the operating point on the shift factors. This introduces further errors, but these are typically also only small. See, for example, Ross Baldick, "Variation of Distribution Factors with Loading," to appear in *IEEE Transactions on Power Systems,* 2003.

[6] Shmuel S. Oren, "Congestion Pricing and Transmission Rights," presented at the Public Utility Commission of Texas Public Workshop on Elements of Market Design, Austin, Texas, November 1, 2002, page 57.

Under these assumptions the actual shift factor is equal to the zonal average shift factor for each bus and no error is committed in using the average shift factor instead of the actual shift factor.

For convenience, I will refer to these assumptions as ERCOT assumptions 1 and 2, respectively, and focus on these assumptions in the next two sections. (In section 5, I will discuss one additional assumption that is implicit in the ERCOT model.)

3.3 Idealized System

To illustrate an idealized system where ERCOT assumptions 1 and 2 hold true, consider figure 1 in Appendix A, which shows a hypothetical system but having a general geographical layout similar to ERCOT. There are various transmission lines represented in figure 1 as solid lines. Busses in the system are located at the ends of each line, but are not explicitly shown. Similarly, generators and demand are not explicitly shown, but should be thought of as being connected to the busses.

The thickest lines in figure 1 represent the CSCs. The intermediate thickness lines represent *hypothetical* lines that have zero electrical impedance.[7] In each of four zones there are six such lines forming a ring around the zone. The thinnest lines are other lines in the system. The zero impedance lines are an idealization since no line can have an impedance of zero. Nevertheless, this idealization will be helpful in understanding practical systems.

The zero impedance ring around each zone has a very specific effect on the shift factors. In particular, it effectively removes the influence of the *location* of generation inside a zone on flows on lines outside the zone. For any given line connecting *between* zones, the effect on the line flow due to an injection in one particular zone and withdrawal in any other particular zone will be independent of the location of the point of injection in the zone of injection and independent of the point of withdrawal in the zone of withdrawal. That is, the shift factors to lines between zones are identical throughout a zone and ERCOT assumption 1 is satisfied for the system shown in figure 1 having zero impedance rings around each zone.

Moreover, within a zone, if generators change their individual levels of generation but the same total generation is maintained overall in the zone then there is no change in the line flows for lines outside the zone. That is, ERCOT assumption 2 is satisfied for the system shown in figure 1 having zero impedance rings around each zone.

In summary, for an idealized system with zero impedance rings around each zone and for a given level of total generation in each zone, the exact pattern of dispatch within one zone will not affect the line flows outside that zone. For this ideal system, ERCOT assumptions 1 and 2 would be exactly satisfied.

In the next section, I will discuss more realistic systems.

3.4 Practical system

In reality, there are no zero impedance lines and no zone has a zero impedance ring around it. A more realistic system is illustrated in figure 2 in Appendix A. In this figure, the CSCs are still illustrated as the thick lines. All the rest of the lines are

---

[7] Impedance is an electrical property of transmission lines. See, for example, Arthur R. Bergen and Vijay Vittal, *Power Systems Analysis*, Second Edition, 2000, Prentice-Hall, Upper Saddle River, New Jersey.

illustrated as thin lines. In this case, the location of injection in a zone and withdrawal in another zone *does* affect the flows on the lines joining the zones. Moreover, changes in shares of generation within a zone will affect flows on lines in other zones.

In summary, ERCOT assumptions 1 and 2 are not satisfied, at least not exactly, in real systems. However, the presence of a "mesh" of transmission lines within a zone having relatively small electrical impedances would serve to keep the variation of shift factors relatively small in a zone. That is, the use of a zonal average shift factor would not, in that case, introduce significant error. The important practical questions then are:

1. how large is the variation of the shift factors across a zone and
2. how large are the economic implications of the variation of the shift factors.

The first question depends on the electrical characteristics of the ERCOT system, while the second depends on both the electrical characteristics and the economics of electricity generation in ERCOT.

In section 4, I investigate these questions by considering the actual shift factors in ERCOT as posted on the ERCOT website, comparing them to the zonal averages that are used in the commercial model. I then use typical costs and prices to evaluate the economic implications.

4 Analysis of ERCOT commercial model Based on Actual ERCOT Shift Factors

In this section, I apply the discussion from section 3 to analyze the economic implications of using zonal average shift factors instead of the actual shift factors. In section 4.1, I describe data posted on the ERCOT website that I use for the analysis, then focus in sections 4.2 to 4.5 on the STP-Dow CSC, discussing the effect on the short- and long-term incentives of using the zonal instead of the actual shift factors. Then, in section 4.6, I discuss the implications for local congestion management. In section 4.7, I briefly indicate that similar considerations apply for the other CSCs. The results in sections 4.2 to 4.7 use shift factors that are calculated by ERCOT on an annual basis. In section 4.8, I show that using shift factors calculated by ERCOT for each month would lead to essentially the same conclusions. I summarize the conclusions in section 4.9.

4.1 Source data

To illustrate the variation of shift factors, consider the shift factors specified in the spreadsheet P39_D237.XLS that is available from the ERCOT website.[8] The spreadsheet summarizes determination of annual shift factors and zones for the 2003 CSCs. For reference, an extract of the data from this spreadsheet appears as Appendix B in this paper. I will use this data in sections 4.2 to 4.7.

The shift factors can vary somewhat month by month due to changes in the configuration of the transmission system. Monthly shift factors for January, February, and March 2003 are also available from the ERCOT website and are specified in the spreadsheets P46_D251.XLS, P48_D265.XLS, and P51_D275.XLS.[9] I will use this data in section 4.8.

---

[8] The Excel Spreadsheet P39_D237.XLS is available from the ERCOT website, www.ercot.com, under "ERCOT TCR Auctions," "Related TCR Postings / News," "Zonal Average Weighted Shift Factors."

[9] These Excel Spreadsheets are available from the ERCOT website, www.ercot.com, under "ERCOT TCR Auctions," "Related TCR Postings / News."

Also on the ERCOT website are a series of Balancing Energy Summaries that provide data on the congestion prices. The spreadsheets are named 2002-01_BES.XLS, 2002-02_BES.XLS, etc.[10] The data for March 2002 through February 2003 from these spreadsheets will be used to calculate an annual average congestion price.[11] I will use this data in section 4.4.

4.2 Comparison of STP-Dow CSC Zonal Average and Actual Shift Factors

As an example of the discrepancy between zonal average and actual shift factors, consider the STP-Dow CSC. I refer to the spreadsheet P37_D237.XLS (an extract of which is reproduced in Appendix B.1) for actual shift factors. For bus numbers 48621 and 48622, which are the locations of the Dow generation facilities, the actual shift factor to Dow-STP is approximately –0.4212. For bus numbers 48671 through 48674, which are the locations of Frontier generation, the actual shift factor to Dow-STP is approximately –0.0755.

There is considerable generation located at these six busses and they are all in the Houston zone but they clearly do not have the same actual shift factors with respect to the Dow-STP CSC. In particular, ERCOT assumption 1 is violated since Dow and Frontier are in the same zone but the Dow shift factor of –0.4212 is not equal to the Frontier shift factor of –0.0755. These generators are in the same zone but actually have very different effects on the STP-Dow CSC. Injection of power at Dow causes a relatively greater reduction in power flow on STP-Dow than does injection at Frontier.

The ERCOT commercial congestion model uses the zonal average shift factor for clearing congestion instead of using the actual shift factors. According to the spreadsheet P37_D237.XLS (and as reproduced in Appendix B.2), the zonal average shift factor for the Houston zone to the STP-Dow CSC is approximately –0.1790. The zonal average shift factor for the South zone to the STP-Dow CSC is approximately 0.1892.

4.3 Short-Term Incentives

To assess the economic significance of using the zonal average shift factor instead of the actual shift factors, suppose that the submitted generation schedules from the ERCOT QSEs would result in congestion on the STP-Dow CSC. For simplicity, we assume that the QSEs have submitted balanced schedules and that STP-Dow is the only CSC that is congested; however, the discussion below is similar if other CSCs are congested or if multiple CSCs are congested simultaneously.

To clear the zonal congestion, ERCOT procures “balancing energy.” That is, it solicits bids for energy on a zonal basis and solves what is effectively a four-node transmission-constrained bid-based economic dispatch using the CSCs as constraints and assuming zonal average shift factors in each zone. Clearing prices are determined in each zone and these set the price for the balancing energy purchased and sold. To understand

---

[10] The Excel Spreadsheets 2002-01_BES.XLS, 2002-02_BES.XLS, etc are available from the ERCOT website, www.ercot.com, under “Market Participants and Stakeholders,” “Public Market Information,” “Market Information,” “Market Papers,” “BES Papers,” collected together in 2002_BES.zip. The Spreadsheets 2003-01_BES.XLS and 2003-02_BES.XLS are available on the same page.

[11] Prior to February 15, 2002, congestion costs were “uplifted.” I will base calculations on the congestion prices subsequent to “direct assignment” of congestion costs. The zonal definitions in 2002 and 2003 are different; however, I will assume that March to December 2002 congestion prices are representative of expectations for March to December 2003 and beyond.

the incentives for clearing congestion, we will consider the "shadow price" on the STP-Dow CSC, which implicitly determines the energy price differences between zones. In particular, the product of:

- the shadow price on the STP-Dow CSC and
- the difference between the South Zone and Houston Zone average shift factors of 0.1892-(-0.1790) = 0.3682,

provides the effective price for transmitting power from the South Zone to the Houston Zone (assuming that no other CSCs are congested.)

Suppose that both Frontier and Dow can increase their generation and have bid to provide zonal balancing energy. (Similar arguments to the following would also apply for generators that are electrically "nearby" to Frontier and Dow, respectively.) For simplicity in the argument also assume that there is a generator available in the South zone that is prepared to back down its generation and that the actual shift factor at this South zone generation is approximately equal to the South zonal average shift factor.[12]

Let us suppose that either or both of Frontier and Dow are chosen by ERCOT to increase generation and that the South zone resource is chosen to decrease generation. The combined effect of generation backing down in the South Zone and either or both of Frontier and Dow increasing generation in the Houston Zone is that the flow on the STP-Dow CSC would be decreased to help relieve the congestion.

As mentioned above, the zonal average shift factor for the South Zone to the STP-Dow CSC is approximately 0.1892. Consider a pricing interval when the shadow price on STP-Dow congestion was $50/MWh. Then the net incentive in the ERCOT market due to the congestion price to decrease generation in the South Zone and increase generation in the Houston zone would be approximately (0.1892 –(-0.1790)) times $50/MWh or $18.41/MWh. Generators in the Houston Zone should be willing to increase generation and generators in the South should be willing to decrease generation so long as the difference between their (marginal) production costs is not more than $18.41/MWh.

However, based on the actual shift factor from Dow and assuming the same shadow price on the STP-Dow CSC, the correct incentive should be approximately (0.1892-(-0.4212)) times $50/MWh or approximately $30.52/MWh, reflecting the fact that Dow causes a relatively greater reduction in power flow on STP-Dow than implied by the zonal average shift factor.[13] The correct incentive is significantly more than the incentive based on the zonal average shift factors. In fact, the ERCOT pricing signal is only approximately 60% of the value of the correct pricing signal. The correct incentive for Dow to increase its generation and for the South Zone resource to decrease its generation would be more than $12/MWh higher than the incentive they receive under the ERCOT congestion pricing for a shadow price on STP-Dow of $50/MWh. This discrepancy between the ERCOT commercial model and the correct incentive is due to the difference of (-0.1790) – (-0.4212) = 0.2422 between the zonal average and the actual shift factors for Dow to the STP-Dow CSC.

---

[12] I will refine this calculation in section 4.7 and show that with the actual shift factors in the South Zone the situation can be even worse than as calculated here using the zonal average shift factor for the South Zone.

[13] Using the correct shift factors to clear congestion could result in a different shadow price, but I will ignore this issue here.

It is not hard to imagine situations where a $12/MWh difference in the incentive to Dow could be significant to Dow's decision to change output level at its facility (and could be significant for making the corresponding change in the South Zone.) The correct incentive for Dow would encourage more generation from Dow when there was congestion on the STP-Dow CSC. That is, the short-term incentives of the ERCOT pricing signal are too low at Dow for the congestion condition described.

The short-term incentives due to the ERCOT pricing signal can also be too high compared to the correct signal. For example, using the actual shift factor from Frontier, the correct incentive for generation at Frontier should be approximately (0.1892-(–0.0755)) times $50/MWh or approximately $13.24/MWh, reflecting the fact that Frontier causes a relatively smaller reduction in power flow on STP-Dow than implied by the zonal average shift factor. While the correct incentive deviates by only about $5/MWh from the value of $18.41/MWh calculated using the Houston zonal average shift factor, it illustrates that the incentive provided by the ERCOT congestion pricing scheme can also be too high compared to the correct signal.

4.4 Long-Term Incentives

At all times when there is congestion on STP-Dow, the ERCOT congestion price is only approximately 60% of what it should be between the South Zone and Dow and it is approximately 140% of what it should be between the South Zone and Frontier. These incorrect pricing signals provided by the ERCOT congestion pricing scheme contribute to expectations about future prices and consequently also distort forward price curves. In turn, distortion of forward prices will produce distorted incentives for new construction. That is, in the long-term, the distortion of pricing signals will cause new facilities to be located inappropriately by not encouraging enough generation to be located in the vicinity of Dow or by encouraging too much to be located at Frontier, for example. This is an extremely serious problem since the construction of new generation at poor locations will ultimately necessitate more expensive transmission infrastructure.

To estimate the effect on long-term incentives for generation construction, consider the time average shadow price on the STP-Dow CSC for the year from March 2002 to February 2003. Based on the ERCOT documents 2002-03_BES.XLS, 2002-04_BES.XLS, etc, the time average shadow price for the year for STP-Dow is on the order of about $2/MWh.

Consider the incentive "at the margin" to invest in 1 kW of hypothetical new generation capacity at the Dow location. I have observed that the ERCOT signal is too low, with a discrepancy between the zonal average and actual shift factors of 0.2422. Assuming a time average shadow price of $2/MWh for STP-Dow, the ERCOT signal for 1 kW of potential new generation is too low by about $4/(kW.year). To assess whether or not this is significant, we must estimate the capital carrying cost for new generation capacity.

Let us assume that the yearly capital carrying cost for a new generation investment is on the order of $80/(kW.year). As calculated above, the ERCOT incentive at the Dow location is too low by about $4/(kW.year), which is approximately 5% of the capital carrying cost for new capacity. That is, the error in the ERCOT pricing signal at Dow is a not insignificant fraction of the cost of new capacity and will likely have an effect on location decisions.

Conversely, the long-term incentives at Frontier are to encourage too much generation at Frontier. From a long-term perspective, distortion of prices at Frontier is particularly serious because it encourages facilities to locate in places that are not, in reality beneficial to the system. The error in the ERCOT pricing signal is likely to have a significant and bad effect on the location of new generation capacity in ERCOT. Ultimately this will require the construction of new transmission infrastructure to support inappropriately located generation.

4.5 Summary of Implications of Violation of ERCOT Assumption 1

In summary, ERCOT assumption 1 is violated for Dow and Frontier with respect to the STP-Dow CSC. The ERCOT pricing signal for congestion between the South Zone and Dow is approximately 60% of the value it should be. The ERCOT pricing signal for congestion between the South Zone and Frontier is approximately 140% of the value it should be. The price differences between South and Houston should range from around \$13/MWh to around \$30/MWh, for a shadow price on STP-Dow of \$50/MWh; however, the ERCOT pricing signal blunts this signal, sending a price difference of \$18.41 between busses in the South Zone and busses in the Houston Zone for a shadow price on STP-Dow of \$50/MWh.

The violation of ERCOT Assumption 1 has both short-term and long-term implications for generation, as discussed in the examples above. In addition, Loads Acting as Resources will also be exposed to these incorrect incentives and transmission planning is likely to be distorted.

4.6 Local Congestion Management

The difference in shift factors for generators within a zone also has implications for local congestion management. As an example, suppose that there is local congestion in the Houston Zone and suppose that Frontier and Dow were situated with respect to the local constraint such that by reducing generation at Dow and increasing it at Frontier the local congestion could be relieved. (We will also consider the converse case where the local congestion required shift of generation in the opposite direction from Frontier to Dow.)

If a re-dispatch of generation from Dow to Frontier occurs, it will affect not just flows on local lines within the Houston Zone but will also affect flow on the STP-Dow CSC. In particular, because the actual Frontier and Dow shift factors to the STP-Dow CSC differ by (-0.0755)-(-0.4212) = 0.3457, then for each MW of generation that is shifted from Dow to Frontier, the flow on the Dow-STP CSC would increase by nearly 0.35 MW, whereas if ERCOT assumption 1 were even approximately true then the change would be much less than 1 MW. This means that if these or nearby generators were to participate in local congestion management, then the flows on Dow-STP would change significantly. This would not only affect the flow on Dow-STP but would also affect the flows in the South and other zones, complicating the congestion management process. That is, ERCOT assumption 2 is violated.

To see an example of the implications, suppose that STP-Dow were already at its limiting flow prior to the clearing of local congestion. Shifting generation from Dow to Frontier would increase the flow on STP-Dow. Consequently, shifting of generation from Dow to Frontier to clear local congestion would then necessitate further "last-minute" instructions from ERCOT to adjust generation to relieve the congestion on STP-Dow. This adjustment will conceivably necessitate expensive generation such as combustion turbines to be committed and run because of the lack of advance planning. That is, the additional transmission congestion on Dow-STP created by the local congestion management protocol could ultimately be relieved using generation that is more costly than necessary because last-minute commitment of extra capacity will likely require additional start-up and minimum loading costs to be incurred.

Moreover, the local congestion in the Houston Zone is also affected by patterns of injections and withdrawals throughout ERCOT. Under the ERCOT protocol, only generation in the Houston Zone can be used to relieve local congestion in the Houston Zone (unless the local congestion cannot be relieved with Houston Zone resources), even if the congestion could be relieved at lower cost by re-dispatching generation in another zone. Again, the local congestion management protocol may lead to more costly relief of congestion than necessary.

As another, related, example suppose that the local congestion were such that it could be relieved by shifting generation in the opposite direction, from Frontier to Dow. Again suppose that Dow-STP CSC were at its limit prior to this shift, with cheaper capacity available in the South Zone and more expensive capacity being used in the Houston Zone. Because Frontier and Dow have different shift factors to Dow-STP, the shift of generation from Frontier to Dow will decrease the flow on the Dow-STP CSC. That is, after the adjustment to relieve local congestion there will be additional unused capability available on Dow-STP; however, there is no mechanism for this capability to be used to import the cheaper power from the South Zone to displace expensive generation within the Houston Zone. This means that Dow-STP would be operated at below its capacity when there was cheaper generation available at the sending end that could displace more expensive generation at the receiving end. That is, the transmission capability is not being used efficiently.

In summary, because of the variation of shift factors across zones, the solution of local congestion leads to changes in flows on the CSCs. This either overloads the CSCs, necessitating further costly re-dispatch, or leaves CSCs operating below capacity when there was cheaper generation capacity available at the sending end. Both of these outcomes are economically inefficient and both are specifically due to the ERCOT local congestion management protocol.

4.7 Shift Factors on Other CSCs.

The examples described in the previous sections are not unique to the Dow-STP CSC. In fact, similar considerations apply for each CSC due to the variation of actual shift factors across each zone. The extract from the spreadsheet P39_D237.XLS in Appendix B.1 provides actual shift factors at various generator busses that I use to compare to the zonal average shift factor differences.

For the Graham-Parker CSC, the difference between the West and North zonal average shift factors is 0.4208. Together with the Graham-Parker shadow price, this shift factor difference determines the incentive for relieving congestion on the Graham-Parker CSC using West Zone and North Zone resources. However, considering the difference between actual shift factors at generation busses in the West and North Zones, the actual shift factor differences range from 0.5438 to 0.2101. That is, the ERCOT zonal average shift factors provide an incentive that can be as little as 77% of or as much as 200% of the correct signal.

These incorrect signals provided by the ERCOT zonal average shift factors for the Graham-Parker CSC are not just isolated instances. Figure 3 in Appendix C shows actual shift factors at each bus to the Graham-Parker CSC, separated by zone, ordered from smallest to largest shift factor, and shown versus the cumulative generation at these busses. Also shown are the zonal average shift factors for each zone, drawn as the horizontal lines on the figure. As shown in figure 3, there is roughly six-hundred megawatts of generation in the North Zone and hundreds of megawatts of generation in the West Zone that are exposed to zonal average shift factors that are wrong, compared to the actual shift factor, by approximately 0.1. Since the shift factor differences between zones are only at most on the order of 0.5, an error of 0.1 or more in the shift factor contributes an error to the congestion incentive of 20% or more.

For the Sandow-Temple CSC, the difference between the South and North zonal average shift factors is 0.3920. Together with the Sandow-Temple shadow price, this shift factor difference determines the incentive for relieving congestion on the Sandow-Temple CSC using South Zone and North Zone resources. However, considering the difference between actual shift factors at generation busses in the South and North Zones, the actual shift factor differences range from 0.5493 to 0.1686. That is, the ERCOT zonal average shift factors provide an incentive that can be as little as 71% of or as much as 230% of the correct signal.

Figure 4 shows actual shift factors at each bus to the Sandow-Temple CSC, separated by zone, ordered from smallest to largest shift factor, and shown versus the cumulative generation at these busses. Also shown are the zonal average shift factors for each zone. As shown in figure 4, more than 5000 MW of generation in the South Zone is exposed to shift factors that are wrong by roughly 0.05 or more. Thousands of megawatts more in the Houston and North Zones are also exposed to shift factors that are wrong by 0.05 or more.

Finally, for the STP-Dow CSC, the difference between the South and Houston zonal average shift factors is 0.3682. However, considering the difference between actual shift factors at generation busses in the South and Houston Zones, the actual shift factor differences range from 0.7392 to 0.1696. That is, the ERCOT zonal average shift factors provide an incentive that can be as little as 50% of or as much as 217% of the correct signal. (This range is worse than as calculated in section 4.3 because here I have used the actual variation of shift factors in the South Zone and not just the zonal average shift factor for the South Zone.)

Figure 5 shows actual shift factors at each bus to the STP-Dow CSC, separated by zone, ordered from smallest to largest shift factor, and shown versus the cumulative generation at these busses. Also shown are the zonal average shift factors for each zone. As shown in figure 5, there is approximately 5000 MW of generation exposed to shift factors that are wrong by approximately 0.1 or more.

4.8 Monthly Shift Factors

The calculations in sections 4.2 to 4.7 and figures 3 to 5 were all based on the 2003 yearly shift factors, both actual and zonal average. In this section, I consider the monthly shift factors, both actual and zonal average.

Figures 6-8 show shift factors versus cumulative generation for the three CSCs for January 2003. Figures 9-11 show shift factors versus cumulative generation for the three CSCs for February 2003. Figures 12-14 show shift factors versus cumulative generation for the three CSCs for March 2003.

Figures 6, 9, and 12, which show monthly shift factors for Graham-Parker, are qualitatively similar to figure 4, which shows annual shift factors for Graham-Parker. Figures 7, 10, and 13, which show monthly shift factors for Sandow-Temple, are qualitatively similar to figure 5, which shows annual shift factors for Sandow-Temple. Figure 8, 11, and 14, which show monthly shift factors for STP-Dow, are qualitatively similar to figure 6, which shows annual shift factors for STP-Dow.

Figures 6 through 14 show that the deviation of the shift factors from the average shift factor in each zone is significant for the monthly shift factors. In particular, it is not the case that the actual shift factors for each month cluster more closely around the average shift factor than do the shift factors for the annual study. That is, the analysis in sections 4.2 to 4.7, which was based on annual shift factors, also holds qualitatively true for the monthly shift factors. This observation is not surprising since the dispersion of the shift factors is due to the underlying electrical properties of the ERCOT system and does not depend on demand or generation conditions.[14]

4.9 Summary

The discrepancies between the incentives due to the zonal average shift factors and the incentives from the actual shift factors can be significant in each zone and for each CSC. The error due to using zonal average shift factors is large, sometimes providing an incentive that is too small by a factor of two and sometimes providing an incentive that is too large by a factor of more than two. The zonal average shift factors provide the wrong incentives and the error in the incentive is greatest when the congestion costs are greatest. It is important to realize that it is precisely under the conditions of high congestion costs that it is most important for the incentives to be correct. The ERCOT congestion management protocol fails to provide correct incentives. Clustering all of the ERCOT busses into just four, large zones produces a variation in shift factors across each zone that substantively violates ERCOT assumptions 1 and 2.

5 Outage Shift Factors

There is a further assumption implicit in the ERCOT congestion model. It is that:

3. The capacity of the lines in the CSCs are the limiting constraints on flows between the zones.

I will refer to this as ERCOT assumption 3. Perhaps surprisingly, ERCOT assumption 3 is violated for each of the three CSCs in ERCOT. The reason for this is that in electricity networks the limiting constraints are very often on the flows that would result in the event of a *contingency* and, moreover, the worst contingency (that produces the limiting constraint) is very often an outage of a large line, such as an outage of the lines associated with the CSCs. To operate the system so that it is N-1 *secure,* these contingency constraints must be respected. This is true for each of the ERCOT CSCs.

[14] As mentioned above, ERCOT uses the DC shift factors that ignore the effect of the operating point on the shift factors.

Failure to respect N-1 security would result in cascading outages and widespread blackouts in the event of any single failure of a transmission element. Since failures of transmission elements can occur at any time due to lightning or other causes, this constraint on operation is important and significant. All transmission systems in North America, including ERCOT, are operated so as to respect N-1 security.

In section 5.1, I will describe an example system that will be used to explain and illustrate N-1 security. In section 5.2, I check that the base-case flows on the example system are N-1 secure. I then describe the mathematical representation of the security constraints in section 5.3, introducing "outage shift factors." In section 5.4, I describe the procedure used for determining the ERCOT CSCs. In section 5.5, I show that the ERCOT CSC procedure generally represents the security constraints *incorrectly.* In section 5.6, I describe conditions under which the ERCOT CSC procedure would happen to represent the constraint correctly.

Then, in section 5.7, I turn again to the STP-Dow CSC describing several closely related transmission elements. I evaluate the outage shift factors for the security constraints associated with STP-Dow CSC in section 5.8. In section 5.9, I compare the correct representation of these security constraints with the ERCOT STP-Dow CSC representation, showing that the ERCOT CSC representation is incorrect. In section 5.10, I make a similar comparison for the other CSCs. I summarize the issues in section 5.11.

5.1 Example system

Consider figure 15, which shows a system with three nodes: S, D, and W. Unlike in figures 1 and 2, figure 15 represents busses explicitly as thick vertical lines. Transmission lines are represented as thinner lines.

We will consider W to be the reference bus for the shift factor calculations and use DC power flows for all calculations. The line S-D represents the CSC, with the other lines:

- S-W,
- D-W, circuit 1, and
- D-W circuit 2,

representing other lines in the system that are not CSCs. We assume that the two circuits of D-W are on different rights-of-way so that we do not have to consider a double contingency of these lines.[15] The S-D CSC has impedance 0.5 unit and capacity of 2 units, whereas each other line in the system has impedance of 1 unit and capacity of 1 unit.

5.2 Security analysis for base-case

Let us suppose that we have a base-case dispatch with 1 unit of injection at S and 0.5 unit of injection at D, with withdrawal of 1.5 units at the reference bus W. With the system as shown in the figure, the flow on S-D is 0.375 units, the flow on S-W is 0.625 units, and the flow on each circuit of D-W is 0.4375 units. For each line, the corresponding base-case flow is within the line rating.

However, the base-case flows are not the only issues limiting the operation of the system. We must check that, in the event of any outage, the flows on the system would

[15] We might interpret the two circuits of D-W to be a network equivalent for various local lines between D and W in the ERCOT system.

still be within acceptable limits. According to Kirchhoff's laws, in the event of such an outage, the flows would be re-distributed throughout the system.[16] We consider all of the contingency constraints for this system:

- On outage of S-W, the flow on S-D would be 1 and the flow on each circuit of D-W would be 0.75. This is within limits.[17]
- On outage of S-D, the flow on S-W would be 1 and the flow on each circuit of D-W would be 0.25. This is within limits, but the capacity constraint on S-W is just binding.
- On outage of either circuit of W-D, the flow on S-W would be 0.8, the flow on S-D would be 0.2, and the flow on the remaining circuit of D-W would be 0.7. This is within limits.

The binding contingency for these injections is the flow on S-W on outage of S-D. Moreover, the base-case is just at the limit.

For simplicity in the following sections, by "base-case" I mean the base-case pre-contingency as illustrated in figure 15. By "outage-case" I mean the flows on the system with the *same* pre-contingency injections at each bus as in the base-case but with the S-D line out-of-service.

5.3 Outage shift factors and security constraints

Analysis of the outage-case showed that the flow on S-W in the event of outage of S-D was the binding limit on operation of the system. We must ensure that the outage case flow on S-W remains within limits. To predict the flow on S-W in the event of outage of S-D for conditions other than the base-case injections and withdrawals, we calculate "outage shift factors." An outage factor is the ratio of the change in the outage-case flow on a line to a change in the base-case injection and a corresponding change in the base-case withdrawal. Such outage shift factors are calculated routinely and used for dispatch in PJM, for example.[18]

In particular, we calculate the outage shift factors to flow on S-W for the outage of S-D. This, together with the capacity of S-W, will characterize the conditions for feasibility with respect to the contingency constraint in terms of the base-case injections.

To calculate the outage shift factors, observe that in the event of an outage of S-D:

- the pre-contingency injection at D has *no* effect on the post-contingency flow on S-W in the outage-case (that is, the outage shift factor from D to S-W is 0.0);
- all of the pre-contingency injection at S flows on S-W in the outage-case (that is, the outage shift factor from S to S-W is 1.0); and,
- since W is the reference node, the outage shift factor from W to S-W is 0.

To prevent S-W being overloaded in the outage-case, we must ensure that the pre-contingency injection at S is such that the outage-case flow on S-W is below the S-W limit of 1.0. The pre-contingency injection at D is irrelevant to the flow on S-W in the

[16] For a discussion of contingency or outage analysis, see, for example, Allen J. Wood and Bruce F. Wollenberg, Power Generation, Operation, and Control, Second Edition, 1996, Wiley, New York.

[17] For simplicity, I am assuming that, for each line, its "normal" and "emergency" ratings are the same. If emergency ratings differ from normal ratings then the emergency rating should be used in the evaluation of feasibility under outage conditions.

[18] See, for example, Andrew L. Ott, "Experience with PJM Market Operation, System Design and Implementation," Sections III.B.3 and III.C.3, To Appear in *IEEE Transactions on Power Systems,* 2003.

outage-case. That is, the correct representation of the contingency constraint is that $P_S \leq 1.0$, where P_S is the pre-contingency injection at S. The constraint $P_S \leq 1.0$, if imposed on the base-case pre-contingency injections, will ensure that the post-contingency flow on S-W is limited to 1.0 and (together with the other contingency constraints) ensures *N*-1 security for the system.

5.4 ERCOT CSC rating and shift factors

As noted above, the base-case dispatch is within limits but the contingency constraint on outage of S-D is just binding. The base-case flow on S-D is 0.375. For this system, the ERCOT CSC procedure would use a flow limit of 0.375 units on S-D in the base-case *pre-contingency* as a *proxy* to the contingency limit on S-W. That is, under the ERCOT congestion management protocol, the *pre-contingency* flow on S-D would be limited to 0.375. We evaluate the shift factors to maintain the pre-contingency flow on S-D to 0.375.[19]

Since W is the reference bus, the shift factor for W is 0. The shift factors for the base-case are:

- For S-D, the shift factor from D is –0.25 and from S is 0.5.
- For S-W, the shift factor from D is 0.25 and from S is 0.5.
- For each circuit of W-D, the shift factor from D is 0.375 and from S is 0.25.

Using the S-D proxy limit of 0.375 and the base-case shift factors for S-D imposes a proxy constraint on operations of the form $0.5P_S - 0.25P_D \leq 0.375$, where P_S and P_D are the pre-contingency injections at S and D, respectively.

5.5 Comparison of correct representation of contingency constraint and ERCOT CSC proxy constraint

The proxy constraint specified by the ERCOT procedure of $0.5P_S - 0.25P_D \leq 0.375$ is very different to the correct contingency constraint of $P_S \leq 1.0$. The two constraints describe different regions and constrain the allowed injections in different ways. In particular, the boundaries of the regions are determined by the slopes of the lines $0.5P_S - 0.25P_D = 0.375$ and $P_S = 1.0$. The first line is oblique, whereas the second line is parallel to the P_D-axis. The lines intersect at the base-case operating point but are not coincident and therefore the regions described by the constraints are different. The proxy constraint used by ERCOT only provides the correct limit on operations if the actual injections and withdrawals turn out to be the same as the base-case.

As a concrete example, suppose that the demand at W were 1.62 and that the proposed generation schedules to meet this demand involved $P_S = 1.02$ and $P_D = 0.6$. These injections satisfy the ERCOT proxy constraint since $0.5(1.02) - 0.25(0.6) = 0.36$, which is less than 0.375. Consequently, the shadow price on this CSC would be zero. However, these injections do not satisfy *N*-1 security because if there were an outage of S-D then the flow on S-W would be 1.02, which exceeds the capacity of S-W.

In fact, ERCOT would not allow these proposed schedules to be implemented and would presumably have to reduce the flow limit on S-D below 0.375 in order that the

[19] All shift factors presented in this section will be "actual" rather than "zonal average." Consequently, I will drop the adjective actual with the understanding that all shift factors quoted here are actual rather than average.

post-contingency flow on S-W were within limits. Because the ERCOT proxy constraint provides the wrong approximation for the contingency constraint, the proxy limit has to be adjusted to guarantee security. Moreover, the incentives provided by the shift factors associated with the ERCOT proxy constraint are incorrect. In particular, if the shadow price on the CSC were non-zero then the zonal clearing price at D would be increased relative to W, encouraging too much generation at D.

If, instead, the contingency constraint of $P_S \leq 1.0$ had been used, the proposed schedules of $P_S = 1.02$ and $P_D = 0.6$ would have immediately been identified as infeasible and bids for zonal balancing energy would have been used to clear the constraint. A non-zero shadow price on the contingency constraint $P_S \leq 1.0$ would directly encourage generation at S to be decreased. The zonal clearing prices at D and W would be equal, as they should be for this example and contingency constraint.

In summary, the ERCOT proxy constraint allows as feasible some schedules that do not satisfy *N*-1 security. This presumably necessitates revision of the CSC limit even when no physical change has actually occurred in the transmission system.

The ERCOT proxy constraint also rules out some schedules that are, in fact, *N*-1 secure. For example, suppose that the demand at W were 1.38 and that the proposed schedules were $P_S = 0.98$ and $P_D = 0.4$. These injections do not satisfy the ERCOT proxy constraint since $0.5(0.98) - 0.25(0.4) = 0.39$, which is greater than 0.375. Consequently, zonal balancing energy would be sought to satisfy the constraint. However, this schedule is in fact *N*-1 secure and there is no need to acquire zonal balancing energy. In order to satisfy the ERCOT proxy constraint, total generation at S would decrease and total generation at D would increase, presumably increasing the total cost of dispatch.

Because the regions specified by *N*-1 security analysis and by the ERCOT proxy constraint are different, there are some injections that satisfy the ERCOT CSC constraint but do not satisfy the correct contingency constraint. Other injections satisfy the correct contingency constraint but do not satisfy the ERCOT CSC constraint. That is, using the ERCOT CSC constraint will erroneously:

1. allow some dispatches that fail to be *N*-1 secure and
2. rule out some dispatches that are *N*-1 secure.

Since security must in fact be respected, the first error will necessitate some further process such as adjustment of the CSC limit, while the second means that the transmission capability is not being fully utilized so that dispatch will be more expensive than necessary.

5.6 Conditions under which ERCOT CSC limit and contingency limit would be the same

The regions described by the ERCOT CSC limit and the contingency limit could only be the same if the boundary of each region were the same; that is, if the lines describing the boundaries were the same. By elementary geometry, this would only be true if the corresponding coefficients of the equations describing the boundary were all in the same ratio. The corresponding coefficients are shift factors to the CSC and outage shift factors to the binding element, respectively.

For the example system, the ratio of the shift factor for S to the outage shift factor for S is (0.5)/(1.0) = 0.5, while the ratio of the shift factor for D to the outage shift factor for D is (-0.25)/(0.0)) or infinite. These two ratios are not the same and so the regions described by the ERCOT CSC limit and the contingency constraint are not the same.

The ERCOT CSC procedure yields a proxy constraint on operation that is different from the true contingency limit. The degree of difference between the correct contingency limit and the ERCOT CSC limit is characterized by the variation, over busses in the system, of the ratio of:

- the shift factor to the CSC, divided by
- the outage shift factor to the limiting element.

I will use this observation in sections 5.7 through 5.10 to evaluate the ERCOT CSC procedure for the ERCOT system.

5.7 Example Using STP-Dow

I will use the STP-Dow CSC to illustrate the ERCOT CSC procedure for the ERCOT system. Figure 16 shows a portion of the ERCOT transmission system, showing several “closely related elements” for the STP-Dow CSC, including the STP-Dow double circuit line itself and the STP-WA Parish line. As in figure 15, figure 16 represents busses explicitly as thick vertical lines. Transmission lines are represented as thinner lines. The three arrows represent connections to the rest of the ERCOT system, which are not shown explicitly. The dashed line shows the notional border between the South Zone and the Houston Zone. Both the STP-Dow double circuit and the STP-WA Parish line run from the South Zone to the Houston zone.

Under normal circumstances, the STP-Dow double circuit, the STP-WA Parish line, (and other lines not shown) are all in service. If a transaction in ERCOT involves injection in the South Zone and withdrawal in the Houston Zone, then some of this power flows on STP-Dow, some power flows on STP-WA Parish, and the rest flows on other lines in the ERCOT system not shown in figure 16.

Naturally, the flows on all these lines must be kept within the respective ratings of the lines. However, as discussed in section 5.2, this is not the only factor limiting the ability to move power from the South Zone to the Houston Zone (and limiting the allowable flows between other zones.) We must also consider *N*-1 security. The ERCOT system is operated to ensure that even in the event of an outage of any transmission element, the resulting flows on the lines in the remaining system will not cause overloads on any remaining line.

For the STP-Dow CSC, the conditions that limit flow involve a double-circuit outage of the STP-Dow lines. A double-circuit outage of the STP-Dow lines is shown in figure 17. In the event of this outage, all the power that was flowing on STP-Dow would then flow on other lines. In particular, the flow on the STP-WA Parish line would be increased significantly in the event of an outage of STP-Dow because considerable of the power that was flowing on STP-Dow would then be flowing on STP-WA Parish.

Re-distribution of flow in the event of an outage of STP-Dow can lead to an overload on STP-WA Parish. This would violate *N*-1 security. To prevent this, the system must be operated so that in the event of an outage of the STP-Dow double circuit, the resulting flow on STP-WA Parish is within its ratings. It is not the capacity of STP-Dow but rather the capacity of STP-WA Parish that limits the ability to move power from the South Zone to the Houston Zone (and between other zones.)

That is, STP-WA Parish is the limiting element, *not* the STP-Dow CSC itself. Flows in the system must be limited so that the flow on STP-WA Parish, in the event of an outage of STP-Dow, would be at no more than the (emergency) rating of STP-WA

Parish. The implications of this observation for congestion pricing will be explored in the next section.

5.8 Congestion pricing and *N*-1 security

Recall that the goal of congestion pricing is to provide signals for efficient use of transmission. As discussed in section 5.3, for the STP-Dow CSC the correct signal would involve predicting the effect of pre-contingency injections and withdrawals in the system on the flow that would result on STP-WA Parish in the event of a contingency of STP-Dow. This outage shift factor, together with the capacity of STP-WA Parish, could, in principle, be used in the congestion pricing scheme as the CSC. For example, the outage shift factor to STP-WA Parish for injection at Frontier is –0.0916. The outage shift factor to STP-WA Parish for injection at Dow is –0.1556.[20]

If congestion prices were based on the outage shift factors and STP-WA Parish as the CSC then ERCOT assumption 3 would be satisfied. This would provide an economically efficient congestion pricing signal for rationing scarce transmission capacity between the South Zone and the Houston Zone.

5.9 ERCOT commercial model

Instead of using the outage shift factor and the capacity of STP-WA Parish for the CSC, the ERCOT commercial model uses the approach described in sections 5.2 and 5.4. First, the conditions of injection and withdrawal under which STP-WA Parish reaches its capacity under outage of STP-Dow are determined using a power flow model and some assumed base-case injections and withdrawals. Then, STP-Dow is returned to service in the power flow model and the flow on STP-Dow is then calculated for the same set of injections and withdrawals. To summarize, the maximum value of flow on STP-Dow pre-contingency is found such that post-contingency there would be no overload on STP-WA Parish.

For convenience in the following description, I will use "outage shift factor" to specifically mean the shift factor to STP-WA Parish in the event of an outage of STP-Dow. "Shift factor" (without the adjective "outage") will continue to mean a shift factor calculated without any outage considered.

In the ERCOT congestion model, the maximum flow on STP-Dow is used as a *proxy,* referred to STP-Dow, for the effect on STP-WA Parish. The shift factors to STP-Dow are used with this proxy limit. However, as discussed in section 5.5, the shift factors and proxy limit only provides the correct limit on operations for one particular set of base-case injections and withdrawals. Under other conditions, the proxy limit is incorrect because, for any given generator, the effect of its generation on STP-WA Parish, as reflected in its outage shift factor, will be different to the effect of its generation on STP-Dow, as reflected in its shift factor. (In addition, in the ERCOT

[20] These and all subsequent outage shift factors were calculated on my behalf by consultants at PowerWorld corporation based on ERCOT power flow data available from the ERCOT website www.ercot.com. As a cross-check, PowerWorld also verified some of the ERCOT shift factor calculations reported in the spreadsheet P39_D237.XLS from the ERCOT website. There were minor discrepancies between the ERCOT shift factors reported in P39_D237.XLS and the PowerWorld calculations. In sections 5.8 to 5.10, I rely upon the ERCOT calculations for shift factors and the PowerWorld calculations for the outage shift factors. Similar conclusions would be reached if the PowerWorld calculations had been used for both the shift factors and outage shift factors.

congestion pricing scheme, there is a further approximation due to the averaging of the shift factors across each zone. This issue was discussed in section 3 and will be neglected here, where I will concentrate on the effect of using a proxy limit instead of the actual limit.)

To consider the effect of the approximation, we must compare the incentives due to:

1. using the proxy flow limit on STP-Dow and the shift factors to STP-Dow, and
2. using the actual flow limit on STP-WA Parish and the outage shift factors to STP-WA Parish.

As discussed in section 4.2 and tabulated in Appendix B.1, the shift factor to STP-Dow for injection at Frontier is -0.0755, while the shift factor to STP-Dow for injection at Dow is -0.4212. As mentioned above, the outage shift factor to STP-WA Parish for injection at Frontier is –0.0916. The outage shift factor to STP-WA Parish for injection at Dow is –0.1556.

As discussed in section 5.6, in order for the proxy constraint to provide the correct signal, the ratio of the shift factor to the outage shift factor would have to be the same for each bus. If this were the case and the correct shadow price was assigned to the STP-Dow CSC then the proxy limit on STP-Dow and the shift factor would provide the same incentive as the limit on STP-WA Parish and the outage shift factors.

However, the ratio of the shift factor to the outage shift factor for Frontier is approximately 0.8, while the ratio of the shift factor to the outage shift factor for Dow is approximately 2.7. This means that the incentive provided by the proxy limit on STP-Dow and the shift factors, on the one hand, is very different from the incentive provided by the actual flow limit on STP-WA Parish and the outage shift factors. If the "correct" incentive were provided to Frontier for congestion on STP-WA Parish then the incentive at Dow would be too high. This result is not surprising since Dow is electrically nearby to the outaged line but relatively far from the limiting element, so that injections at Dow have a very different effect on STP-WA Parish than does Frontier. Moreover, their effects on STP-WA Parish are both different to their effects on STP-Dow.

In summary, the ERCOT congestion pricing scheme provides the wrong incentives to generators because it does not use the outage shift factor to the limiting element.

5.10 Outage Shift Factors for Other CSCs and Other Busses

The example described in the previous section is not unique to the Dow-STP CSC. In fact, similar considerations apply for each CSC because the binding constraint associated with each CSC is the flow on an associated line in the event of an outage of the CSC.

For example, for the Graham-Parker CSC, the limiting element is the Jacksbro-Willow Creek line on outage of Graham-Parker and Graham-Benbrook, not the Graham-Parker CSC itself. We consider shift factors from the Jacksbro (1429) and the Graham (1430) busses, which are both in the West Zone.

According to the spreadsheet P39_D237.XLS, the shift factor from Jacksbro to the Graham-Parker CSC is approximately 0.3489. The outage shift factor from Jacksbro to the Jacksbro-Willow Creek line is approximately 0.3805. The ratio of the shift factor to the outage shift factor for Jacksbro is approximately 0.9.

According to the spreadsheet P39_D237.XLS, the shift factor from Graham to the Graham-Parker CSC is approximately 0.5371. The outage shift factor from Graham to the Jacksbro-Willow Creek line is approximately 0.3220. The ratio of the shift factor to the outage shift factor for Graham is approximately 1.7.

The ratios of the shift factors to the outage shift factors are different for the two buses. This is because injection at or nearby to Jacksbro and Graham, respectively, have different relative effects on the limiting element compared to their relative effects on the Graham-Parker CSC.

This disparity between the ratios of shift factors to the Graham-Parker CSC divided by the outage shift factors to the Jacksbro-Willow Creek line are not just isolated instances. Figure 18 shows the ratios of shift factors to the Graham-Parker CSC divided by the outage shift factors to the Jacksbro-Willow Creek line for sixteen selected busses. (The bus names and numbers are detailed in table 1.) For many of the selected busses, the ratio is relatively close to 1.4. Arguably, if these particular busses were the only ones involved in congestion relief on Graham-Parker then the ERCOT CSC procedure would provide appropriate congestion incentives.

However, for several of the busses shown in figure 18 there is significant deviation of the ratios from the value of 1.4. For two of the busses the ratio is *negative,* which means that the shift factor and shadow price on the CSC encourage a decrease in generation at the corresponding bus when in fact the contingency constraint would be relieved by an increase in generation at the bus. That is, the ERCOT incentive is incorrect in sign at these two busses and provides the wrong congestion incentives at these busses.

As another example, for the Sandow-Temple CSC, the limiting element is Bellso-Peters on outage of Sandow-Temple, not the Sandow-Temple CSC itself. We consider shift factors from the Sandow (3429) and the Bellso (7270) busses, which are both in the South Zone.

According to the spreadsheet P39_D237.XLS, the shift factor from Sandow to the Sandow-Temple CSC is approximately 0.6250. The outage shift factor from Sandow to Bellso-Peters is approximately 0.0794. The ratio of the shift factor to the outage shift factor for Sandow is approximately 8.

According to the spreadsheet P39_D237.XLS, the shift factor from Bellso to the Sandow-Temple CSC is approximately 0.3528. The outage shift factor from Bellso to Bellso-Peters is approximately 0.4570. The ratio of the shift factor to the outage shift factor for Bellso is approximately 0.8.

The ratios of the shift factors to the outage shift factors are different by a factor of ten. This is because injection at or nearby to Sandow and Bellso, respectively, have very different relative effects on the limiting element compared to their relative effects on the Sandow-Temple CSC.

This disparity between the ratios of shift factors to the Sandow-Temple CSC divided by the outage shift factors to Bellso-Peters are not just isolated instances. Figure 19 shows the ratios of shift factors to the Sandow-Temple CSC divided by the outage shift factors to Bellso-Peters for the same sixteen busses as shown in figure 18. In this case, the ratios vary widely with many of the ratios negative. The incentive provided by the Sandow-Temple CSC deviates significantly from the correct incentive. The ERCOT CSC procedure provides incorrect incentives in this case.

Finally, the disparity between the ratios of shift factors to the STP-Dow CSC divided by the outage shift factors to STP-WA Parish presented in section 5.9 are also not just isolated instances. Figure 20 shows the ratios of shift factors to the STP-Dow CSC divided by the outage shift factors for STP-WA Parish for the same sixteen busses as shown in figure 18. In this case, the ratios for many of the busses are indeed close to 0.8; however, there are several busses with ratios that are significantly different from 0.8.

5.11 Summary

To summarize, the ERCOT congestion management protocol uses the *wrong* shift factors, since the binding constraints on flows are not the flows on the CSCs themselves but rather the flows on other limiting elements in the event of a contingency on the CSCs. The correct shift factors are *outage* shift factors. This compounds the error due to using zonal average shift factors and provides incorrect incentives to generators to relieve congestion.

These incorrect incentives complicate the task of ERCOT, which must ultimately clear congestion on the limiting elements in order to maintain security. By maintaining the fiction that the CSCs are the binding constraints, ERCOT is forced to adjust the proxy limits on the CSCs to track the actual binding constraints. This presumably contributes to the need to modify the proxy limits.

The use of proxy limits and shift factors on the CSCs provides the wrong incentives. In some cases, the ERCOT CSC congestion prices encourage a decrease in generation at a bus when, in fact, an increase in generation at the bus would help to relieve the constraint. This is a fundamental market design flaw in the ERCOT congestion management system.

6 Remedies and Conclusion

One basic problem with the ERCOT zonal approach is that the variation of the shift factors (and of outage shift factors) is significant across each zone. Put simply, the zones are too big to support ERCOT assumption 1 that the shift factors are constant in a zone and too big to support ERCOT assumption 2 that re-dispatch of generation within one zone does not affect congestion elsewhere. This problem could be remedied by creating more zones. The implication of creating more zones is that portfolios of generation would have to be distinguished geographically in finer detail than they currently are for zonal congestion management.

It is an open empirical question as to how many zones would be necessary to, say, reduce the error between zonal average and actual shift factors so that the effect on incentives were less than, say, 1% of the typical price of energy. However, one of the main arguments for a zonal model over, for example, a nodal model is the apparent simplicity of a zonal model. If twenty or thirty zones were required to achieve sufficient accuracy then this simplicity of having a small number of zones would be lost. In the absence of a compelling argument for the simplicity of a zonal model using zonal average shift factors, a nodal market (with many more than four nodes) or the use of actual shift factors for each bus might be preferable.

Implementation of a large number of zones, of a nodal market, or the use of actual shift factors would have significant implications for portfolio bidding. In particular, portfolio bidding would no longer be possible since the dis-aggregation into smaller zones or at the nodal level would mean that injections and bids would also have to be specified at this level of geographical dis-aggregation. A move to use the correct economic signals for congestion management is incompatible with portfolio-based bidding.

A second basic problem with the ERCOT commercial congestion model is the emphasis on designated CSCs that are not the binding constraints on power flow. This could be remedied by considering the binding elements and the outage shift factors; however, it is likely that there would be a large number of potentially binding constraints.[21] Alternatively, or additionally, point-to-point financial transmission rights could be used based on an *N*-1 security-constrained transmission model. This would avoid the need to determine proxy limits and would enable ERCOT to manage congestion efficiently.

[21] See the Excel Spreadsheet P8_D18.XLS from the ERCOT website, www.ercot.com, for a list of closely related elements for each CSC. Each such element could be a limiting element under some loading condition.

Appendix A: Figures1, 2, 15 to 20, and Table 1. (See separate document for figures 3 through 14.)

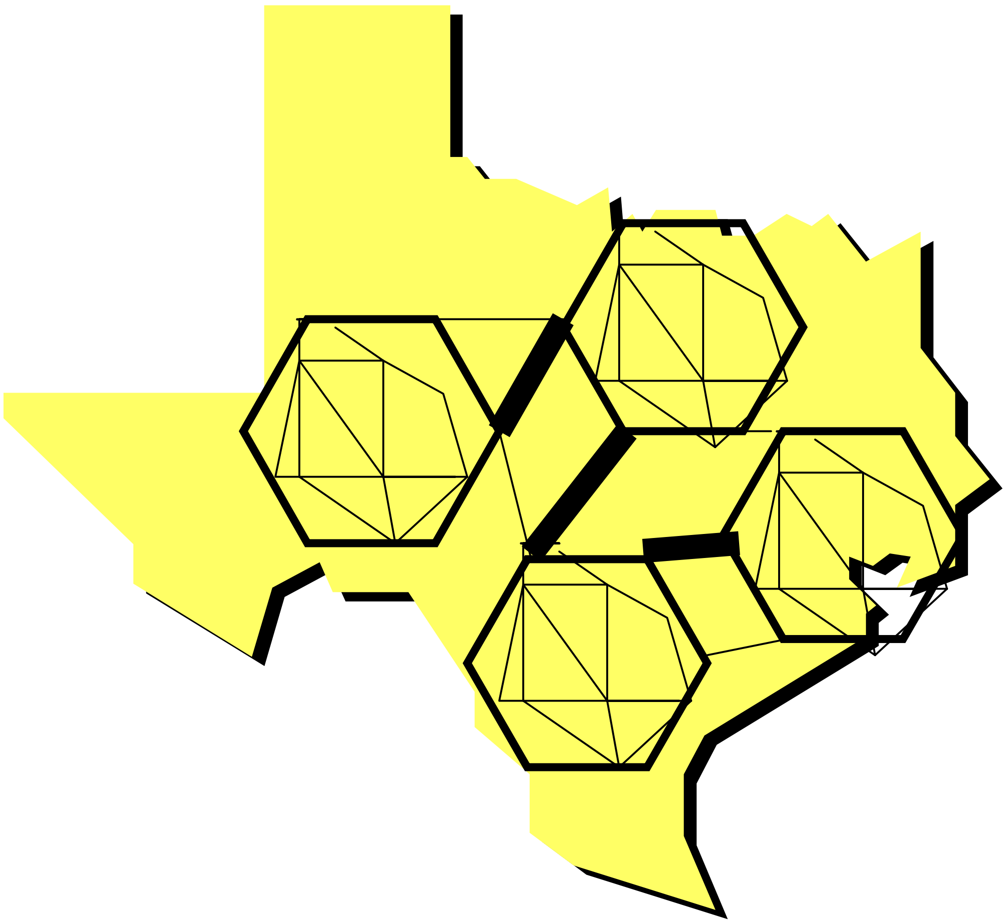


Figure 1. Ideal transmission system where ERCOT assumptions 1 and 2 are satisfied. Busses, generation, and demands are not shown explicitly in this figure. The thickest lines represent the CSCs. The intermediate thickness lines represent *hypothetical* lines that have zero electrical impedance. In each of four zones there are six such lines forming a ring around the zone. The thinnest lines are other lines in the system.

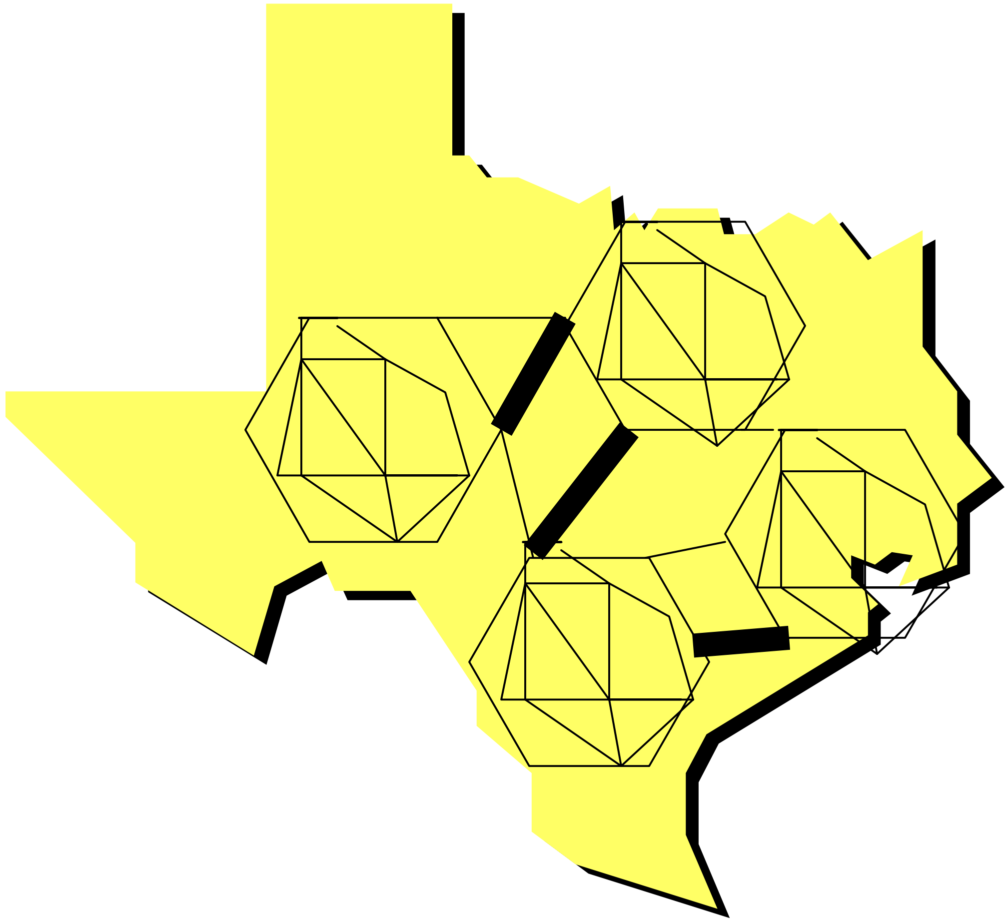


Figure 2. A more realistic transmission system. The CSCs are illustrated as the thick lines. All the rest of the lines are illustrated as thin lines.

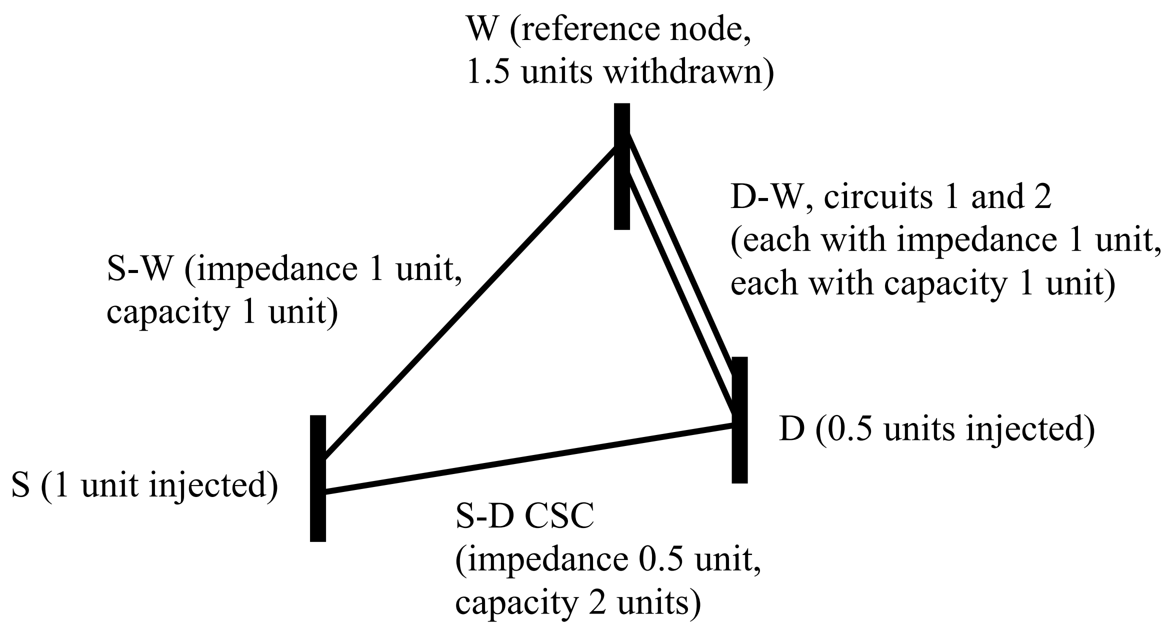


Figure 15. Three node example system to illustrate security constraints. Unlike previous figures, busses are shown explicitly as thick vertical lines. Transmission lines are represented as thinner lines.

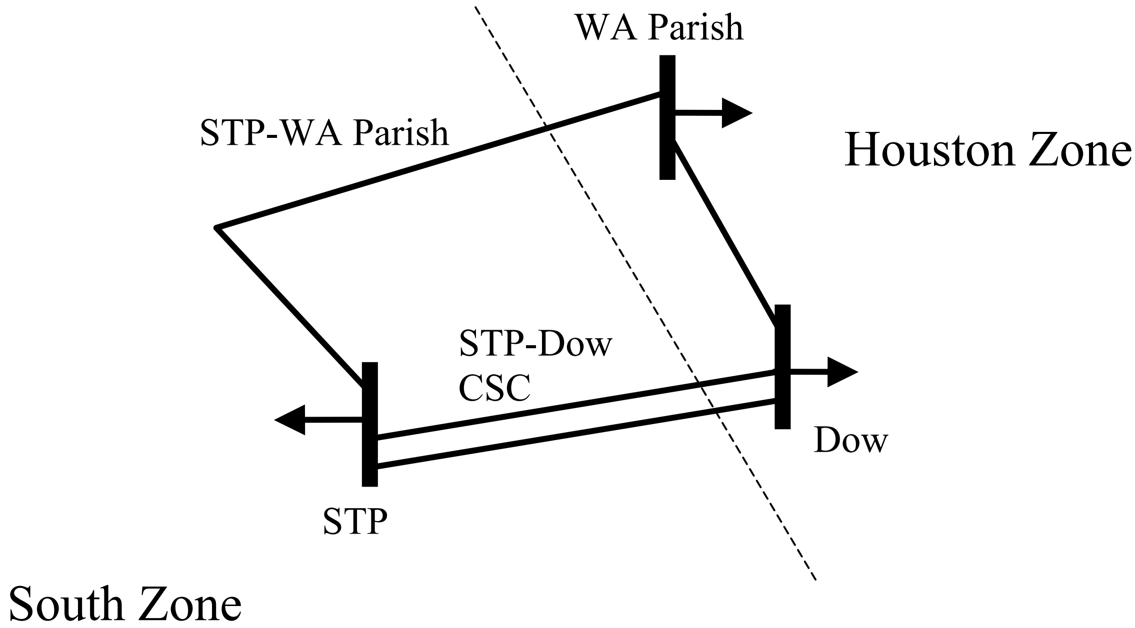


Figure 16. Closely related elements for STP-Dow CSC. The three arrows represent connections to the rest of the ERCOT system, which are not shown explicitly. The dashed line shows the notional border between the South Zone and the Houston Zone.

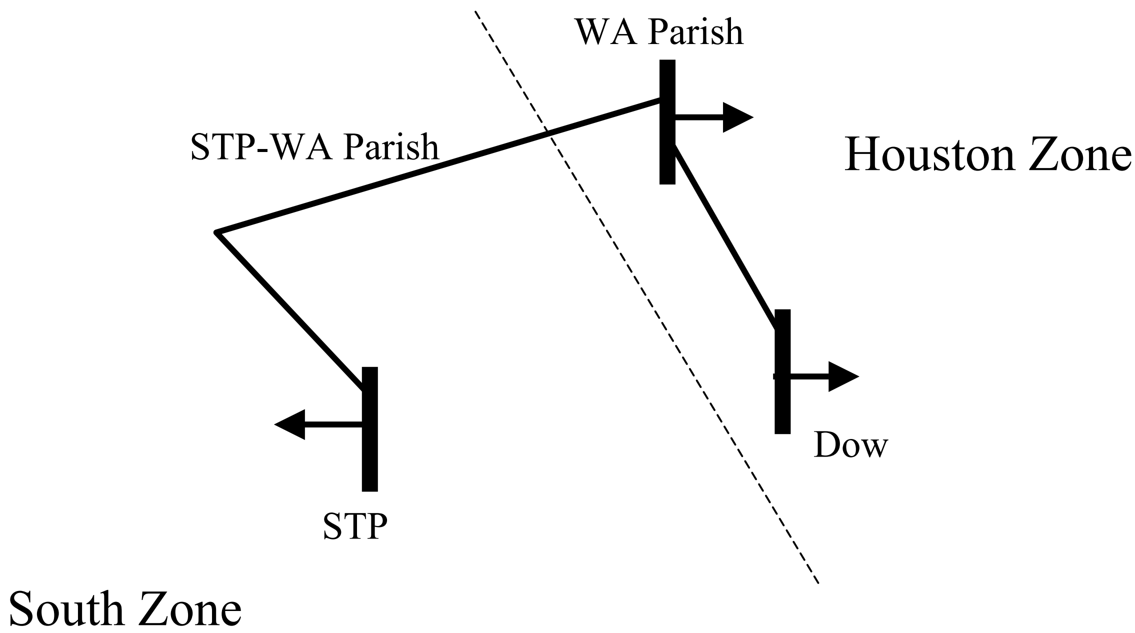


Figure 17. Outage of STP-Dow Double-Circuit.

Figure 18: Ratios of shift factors to Graham-Parker divided by outage shift factors to Jacksbro-Willow Creek for selected busses.

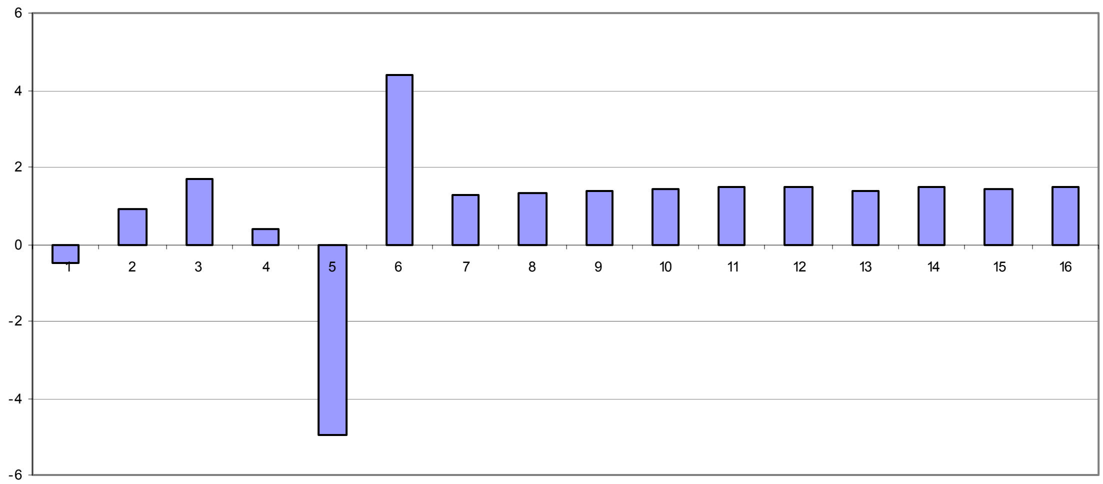


Table 1: Names and bus numbers for busses in used in figures 18 to 20. The number in the first column corresponds to the horizontal axis labels in figures 18 to 20. The second column shows the ERCOT bus name and (in parentheses) the ERCOT bus number.

Numbers used in figures 18 to 20	ERCOT bus names (and numbers)
1	WILLOWCK (1421)
2	JXBRO SS (1429)
3	GRAHAM (1430)
4	PARKER (1436)
5	MOSES (1695)
6	BENBRK (1873)
7	LAKE CRK (3409)
8	LAKE CRK (3410)
9	TEMPSSLT (3413)
10	SANDOW (3429)
11	SO TEX 5 (5915)
12	BELLSO13 (7270)
13	FRONTR (40600)
14	DOW345 5 (42500)
15	W_A_P_ 5 (44000)
16	PETERS 8 (46220)

Figure 19: Ratios of shift factors to Sandow-Temple divided by outage shift factors to Bellso-Peters for selected busses.

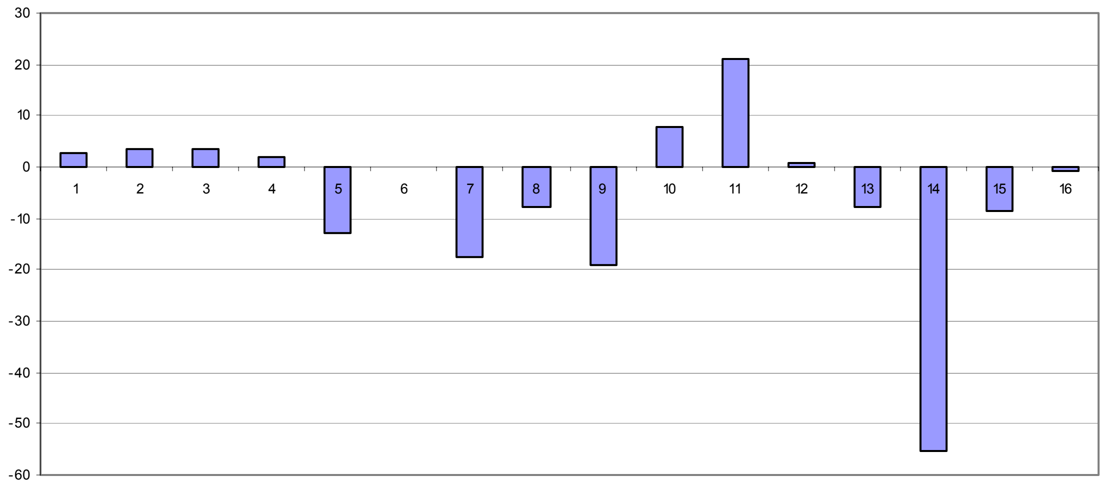
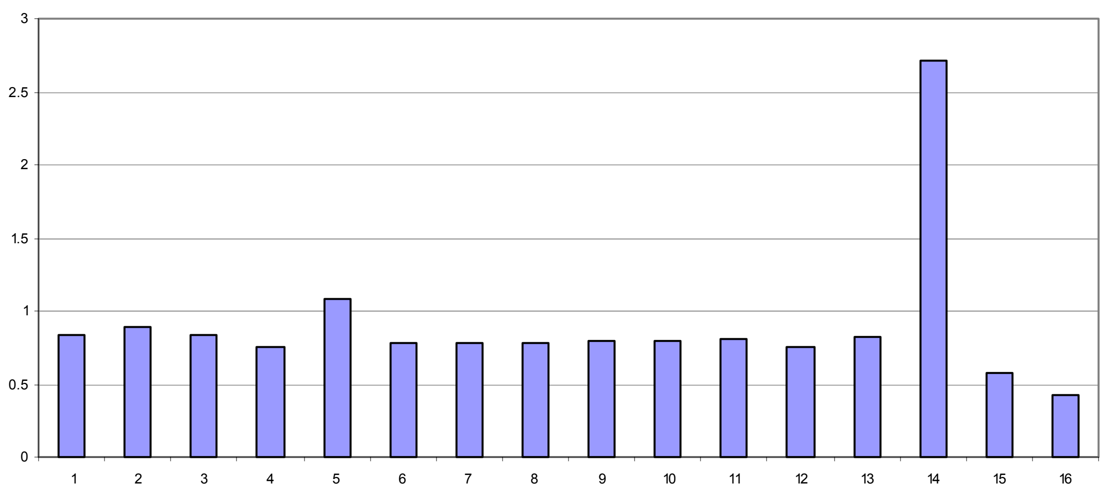


Figure 20: Ratios of shift factors to STP-Dow divided by outage shift factors to STP-WA Parish for selected busses.



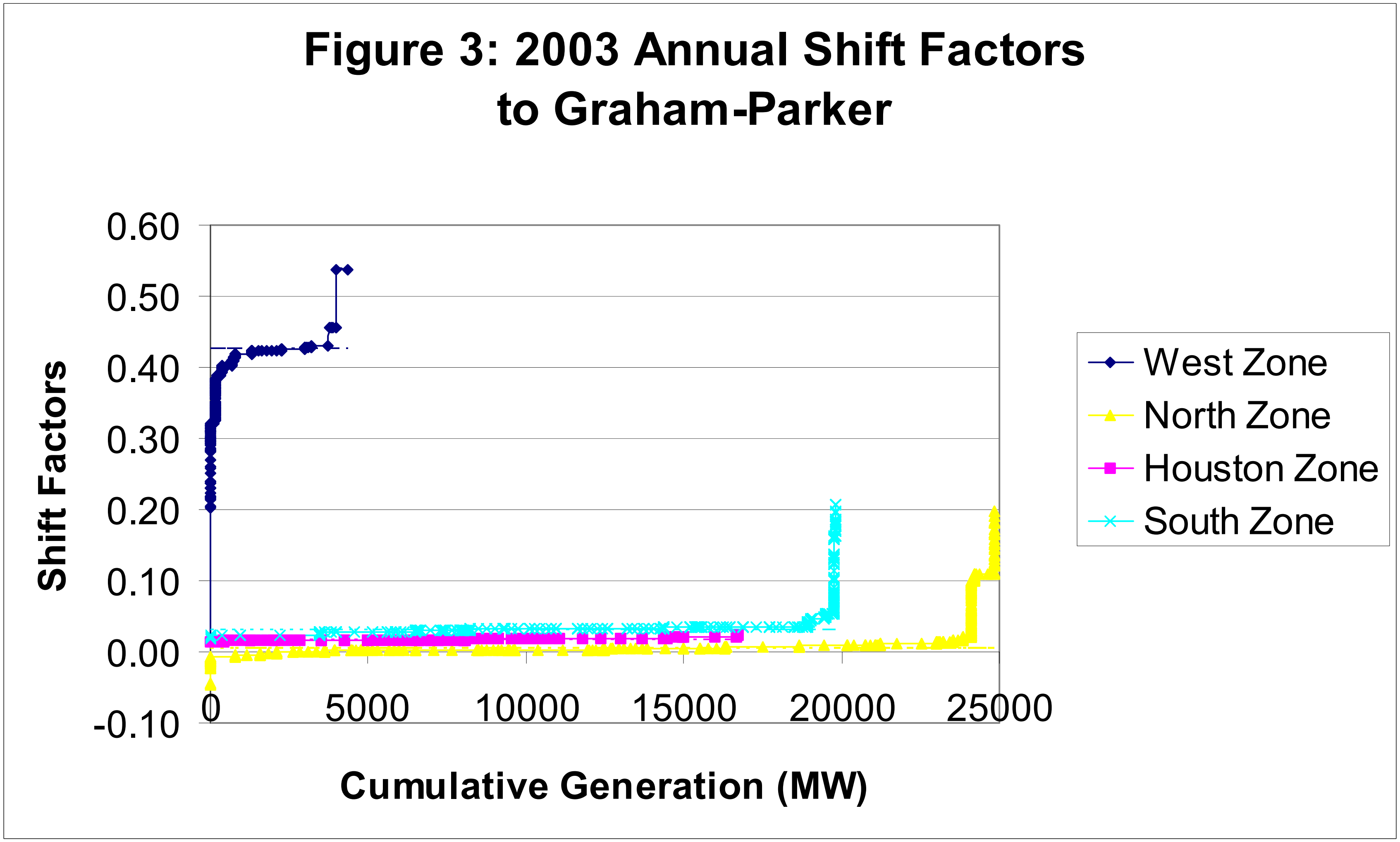
Appendix B: Extract from Spreadsheet P39_D237.XLS, available in entirety from the ERCOT website www.ercot.com.
B.1 Actual shift factors at Selected Busses

BusId	BName	PI	Pgen	CSCzone	Graham - Parker SF	Sandow - Temple SF	STP-DOW SF	Grahparker WSF	SandTemple WSF	StpDow WSF
6763	SAPS 2G		96.9	1	0.320445687	0.068812862	0.033903904	31.051187070	6.667966328	3.285288298
1429	JXBRO SS			1	0.348914415	0.008149936	0.007653684			
1430	GRAHAM			1	0.537107468	0.010247589	0.008798921			
1432	GRAM 2 G		370.5	1	0.537107468	0.010247589	0.008798921	198.998316894	3.796731725	3.260000231
1889	DEC 1 G		777.1	2	-0.006747274	-0.001647472	0.005210041	-5.243306625	-1.280250491	4.048722861
332	MILLER1		70	2	0.110317267	0.000036494	0.008883230	7.722208690	.002554580	.621826100
333	MILLER2		131.1	2	0.110317267	0.000036494	0.008883230	14.462593704	.004784363	1.164591453
334	MILLER3		228	2	0.110317267	0.000036494	0.008883230	25.152336876	.008320632	2.025376440
335	MILLER4		104	2	0.110317267	0.000036494	0.008883230	11.472995768	.003795376	.923855920
336	MILLER5		104	2	0.110317267	0.000036494	0.008883230	11.472995768	.003795376	.923855920
3403	THSE 1 G		536.75	2	0.012092412	-0.086336195	0.029871013	6.490602141	-46.340952666	16.033266228
3404	THSE 2 G		777.1	2	0.012092412	-0.086336195	0.029871013	9.397013365	-67.091857135	23.212764202
892	DANSBY14		110	2	0.013255968	0.133173063	-0.037619341	1.458156480	14.649036930	-4.138127510
40600	FRONTR			3	0.012829033	0.121142708	-0.075485729			
48671	FTR1		75	3	0.012829033	0.121142708	-0.075485729	.962177475	9.085703100	-5.661429675
48672	FTR2		75	3	0.012829033	0.121142708	-0.075485729	.962177475	9.085703100	-5.661429675
48673	FTR3		75	3	0.012829033	0.121142708	-0.075485729	.962177475	9.085703100	-5.661429675
48674	FTR4		190	3	0.012829033	0.121142708	-0.075485729	2.437516270	23.017114520	-14.342288510
42500	DOW345 5			3	0.021463448	0.257407248	-0.421174377			
48621	DOW1		960	3	0.021463448	0.257407248	-0.421174377	20.604910080	247.110958080	-404.327401920
48622	DOW2		693.4	3	0.021463448	0.257407248	-0.421174377	14.882754843	178.486185763	-292.042313012
3429	SANDOW			4	0.023297865	0.624988377	0.100061588			
3430	SANDOW	311.26		4	0.023723345	0.462974548	0.094139501			
3431	SANDOW	210.54	360	4	0.023723345	0.462974548	0.094139501	8.540404200	166.670837280	33.890220360
3432	SAND 4 G		583	4	0.023723345	0.462974548	0.094139501	13.830710135	269.914161484	54.883329083
5911	STP1		1250	4	0.024276076	0.301846743	0.318001181	30.345095000	377.308428750	397.501476250
5912	STP2		1250	4	0.024276076	0.301846743	0.318001181	30.345095000	377.308428750	397.501476250
7270	BELLSO13	14.2		4	0.024279812	0.352822453	0.052189611			

B.2 Zonal Average Shift Factors

CM Zone names	CM Zone	Average Weighted Shift Factor on Commercially Significant Constraint #1 - Graham to Parker	Average Weighted Shift Factor on Commercially Significant Constraint #2 - Sandow to Temple	Average Weighted Shift Factor on Commercially Significant Constraint #3 - Stp to Dow
West2002	Zone1	0.428484450	0.024988360	0.015584690
North2002	Zone2	0.007655688	0.005107820	-0.003440607
Houston2002	Zone3	0.017975560	0.202391595	-0.179045116
South2002	Zone4	0.031717639	0.397075141	0.189228026

Appendix C: Figures 3 through 14.



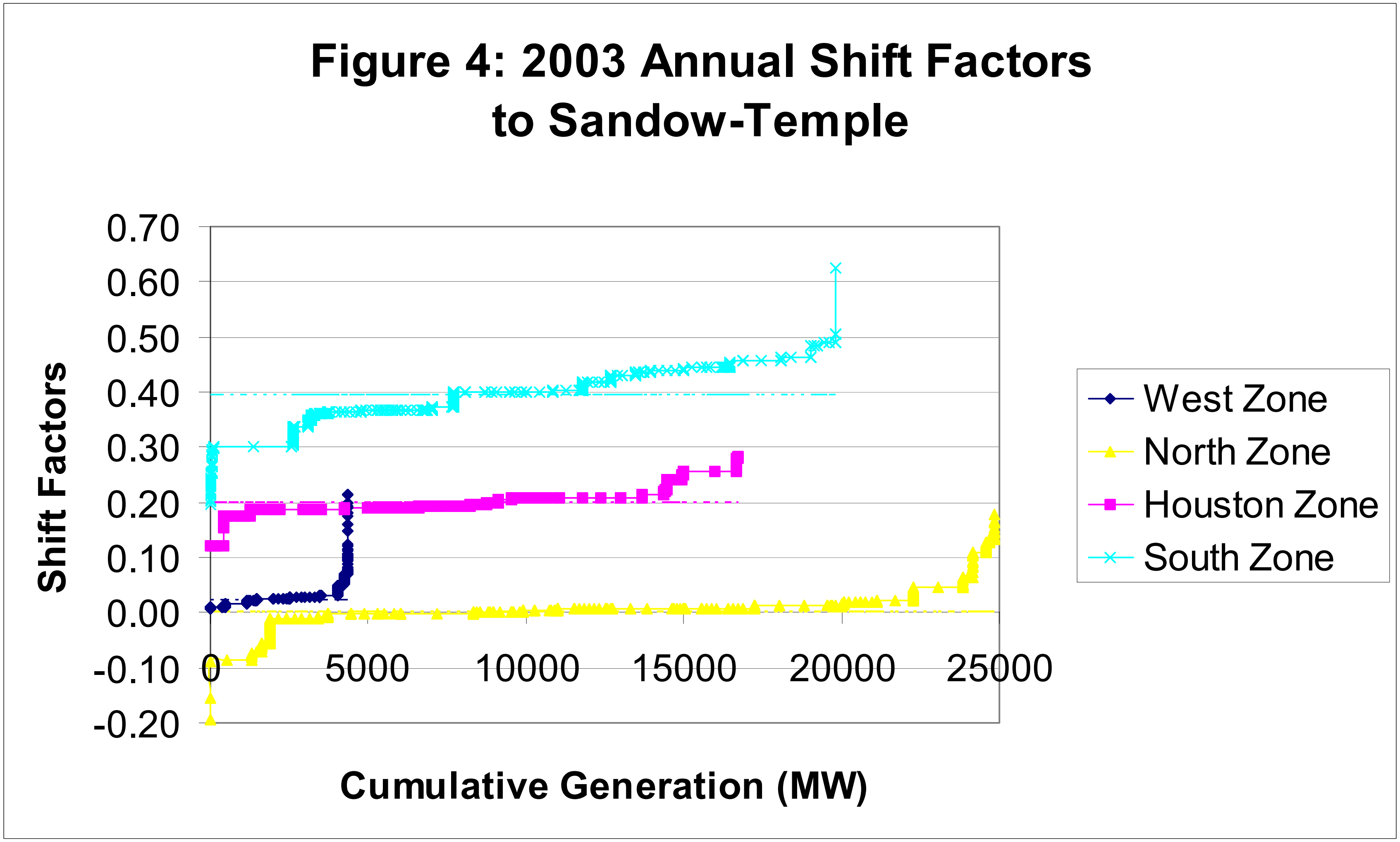


Figure 4: 2003 Annual Shift Factors
to Sandow-Temple
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
-0.10
-0.20
Shift Factors
0
5000
10000
15000
20000
25000
Cumulative Generation (MW)
West Zone
North Zone
Houston Zone
South Zone

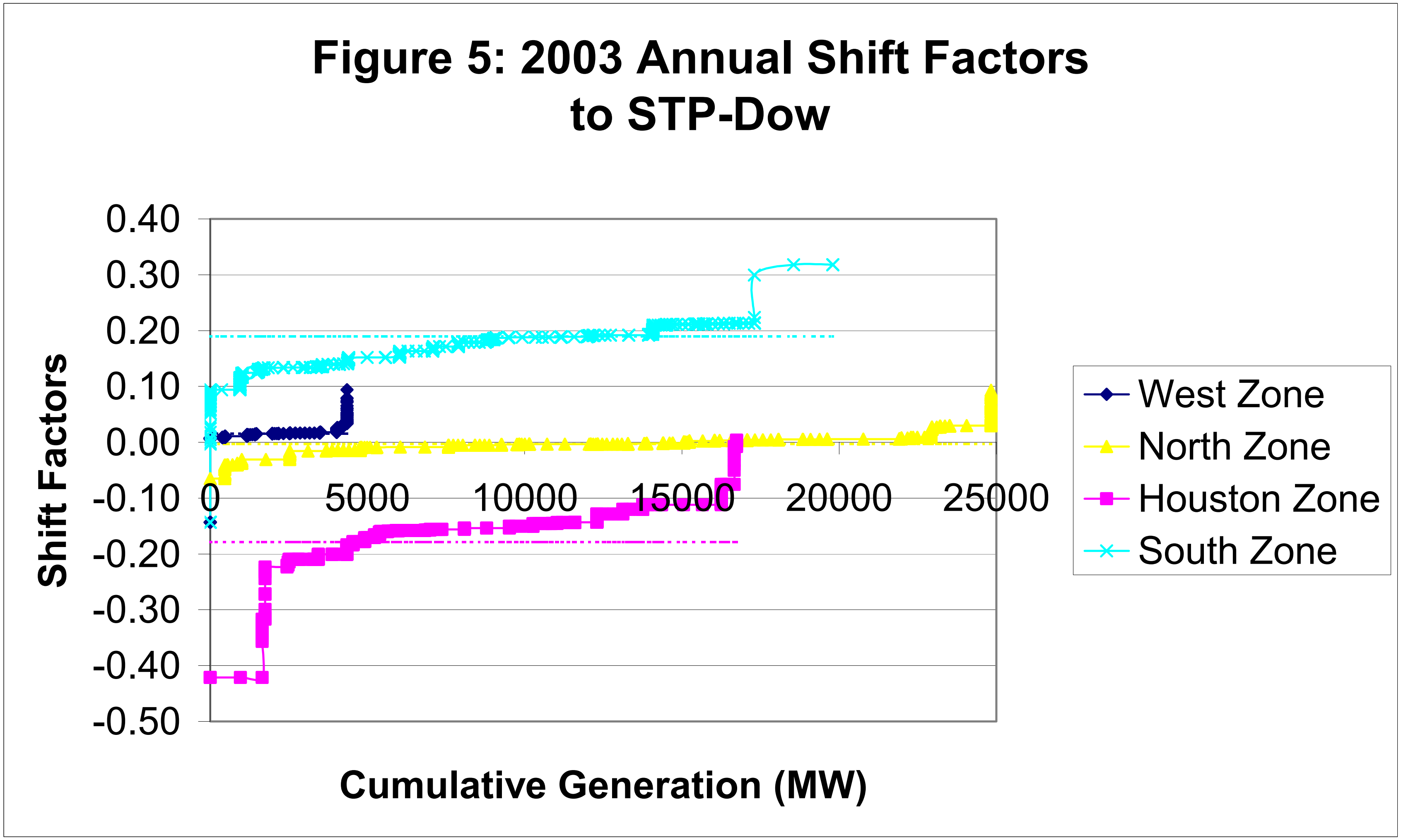


Figure 5: 2003 Annual Shift Factors
to STP-Dow
0.40
0.30
0.20
0.10
0.00
-0.10
-0.20
-0.30
-0.40
-0.50
Shift Factors
0
5000
10000
15000
20000
25000
Cumulative Generation (MW)
West Zone
North Zone
Houston Zone
South Zone

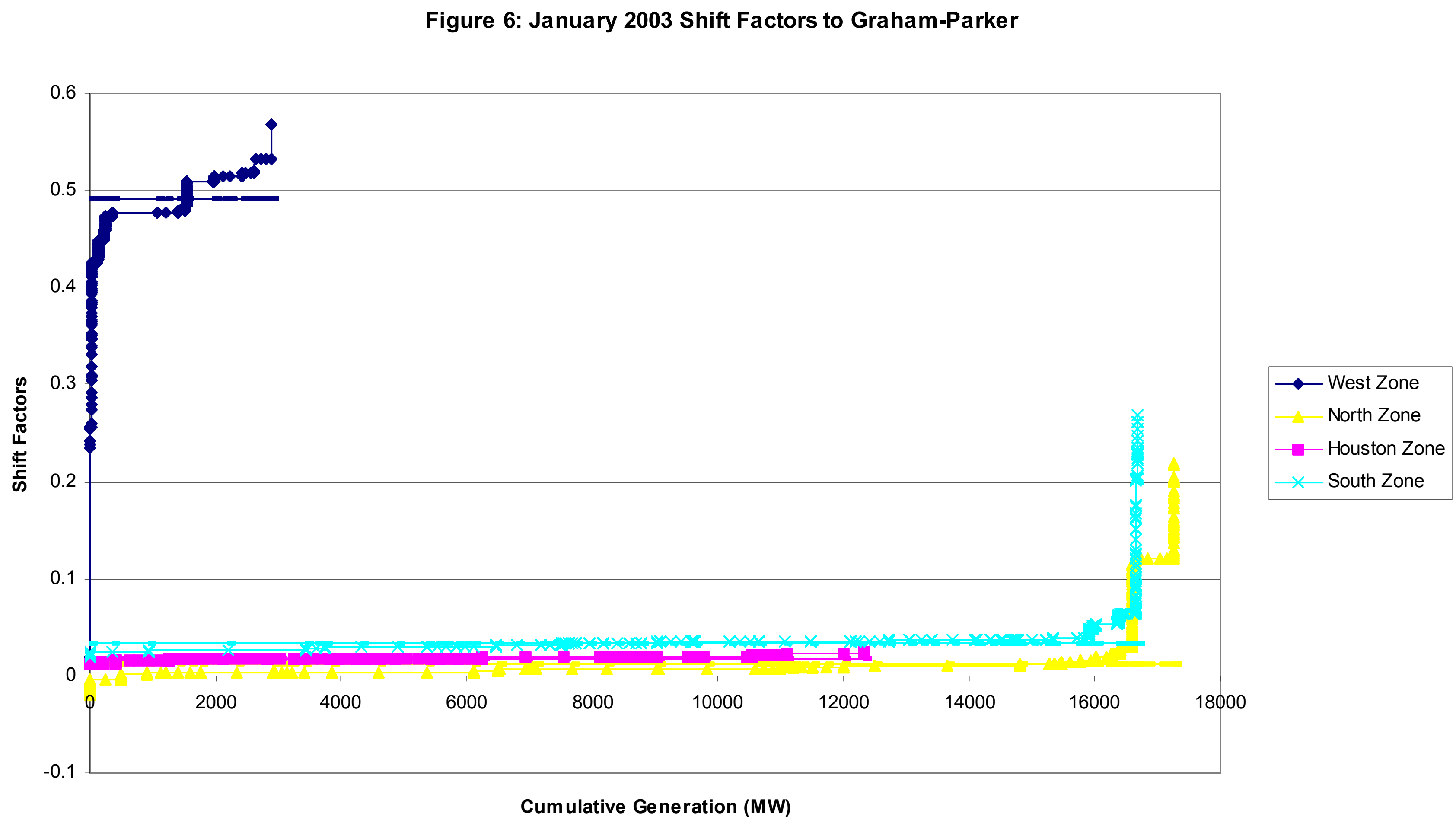


Figure 6: January 2003 Shift Factors to Graham-Parker
Shift Factors
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
0
2000
4000
6000
8000
10000
12000
14000
16000
18000
Cumulative Generation (MW)
West Zone
North Zone
Houston Zone
South Zone

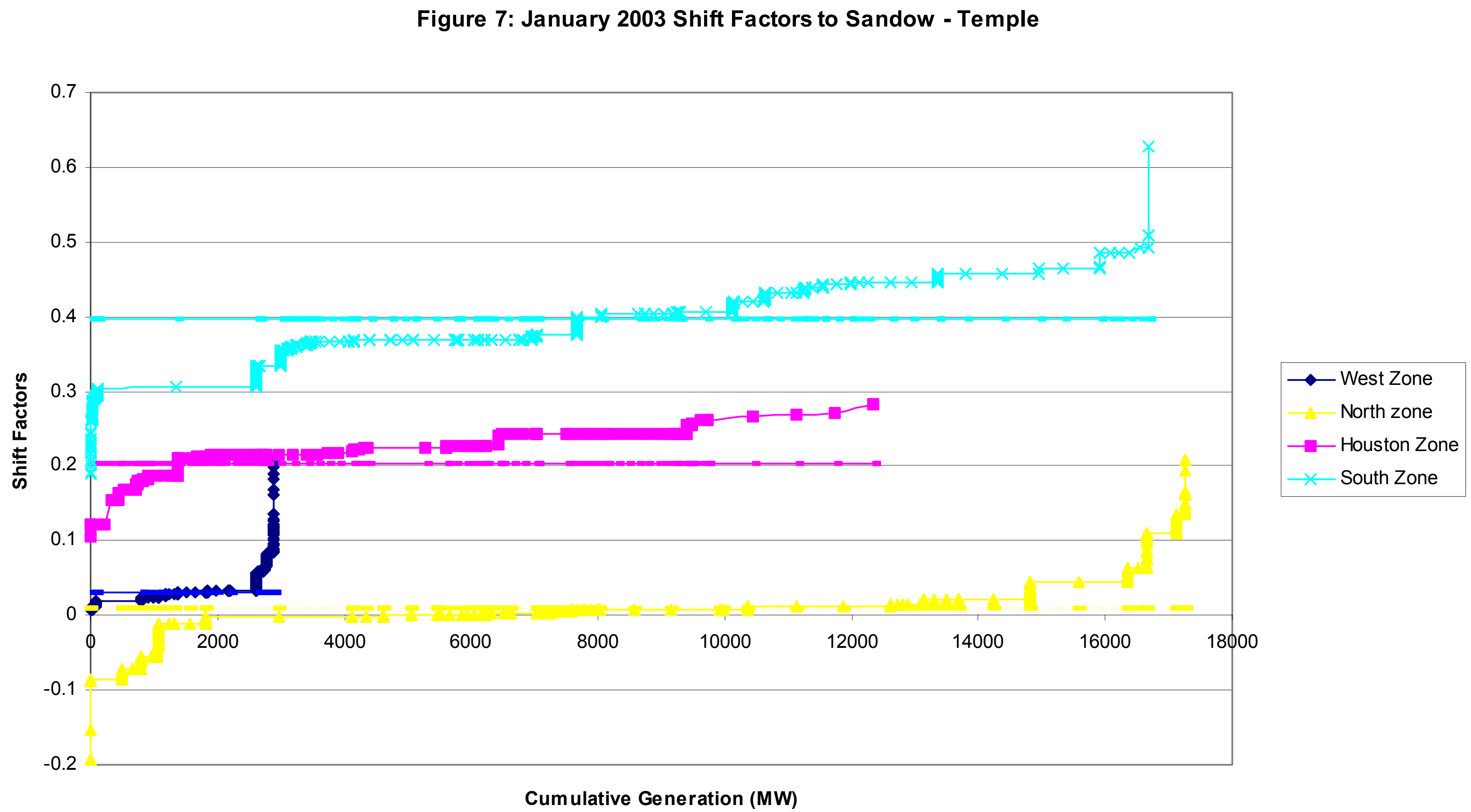


Figure 7: January 2003 Shift Factors to Sandow - Temple
Shift Factors
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
0
2000
4000
6000
8000
10000
12000
14000
16000
18000
Cumulative Generation (MW)
West Zone
North zone
Houston Zone
South Zone

Figure 8: January 2003 Shift Factors to STP-Dow

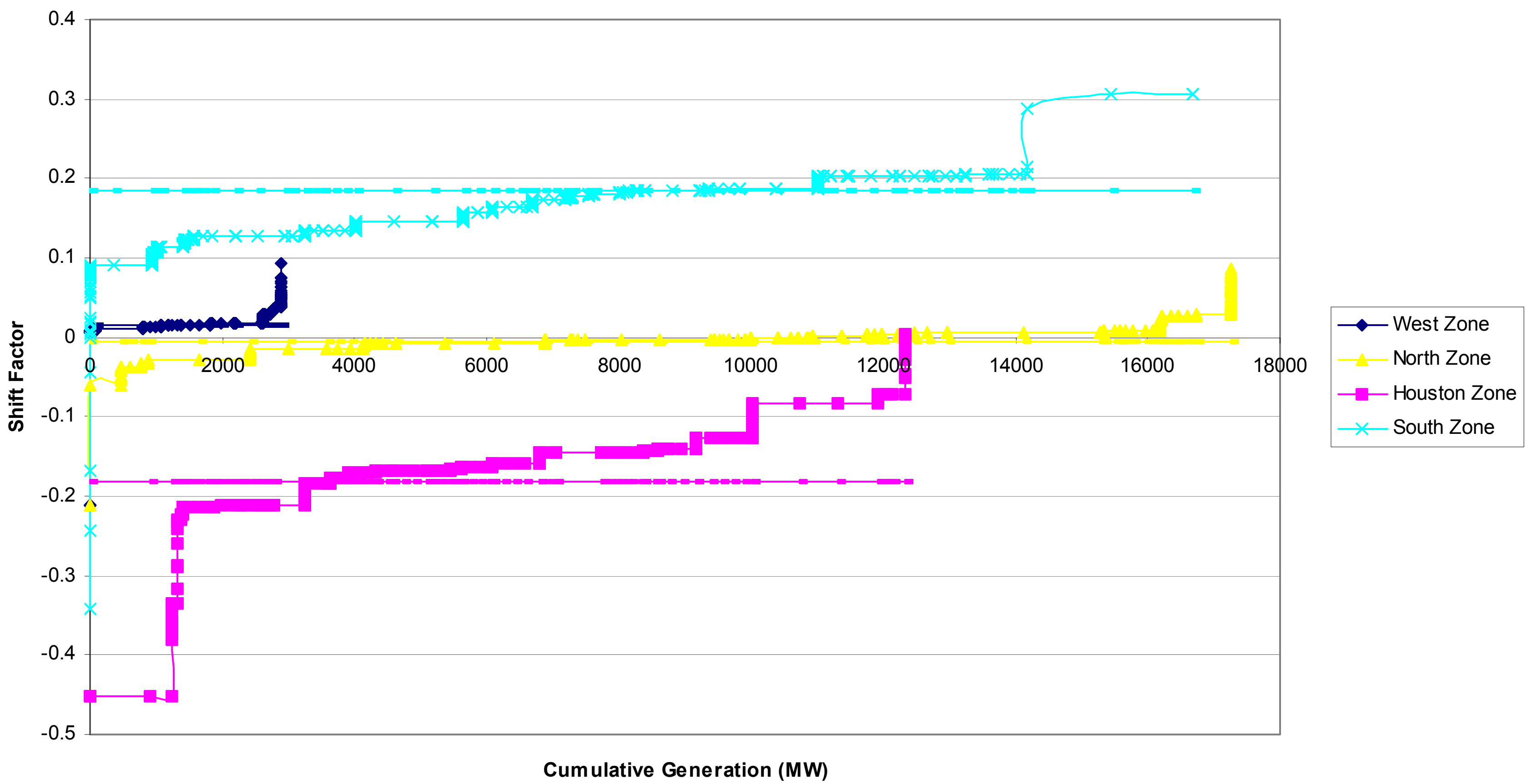


Figure 9: February 2002 Shift Factors to Graham-Parker

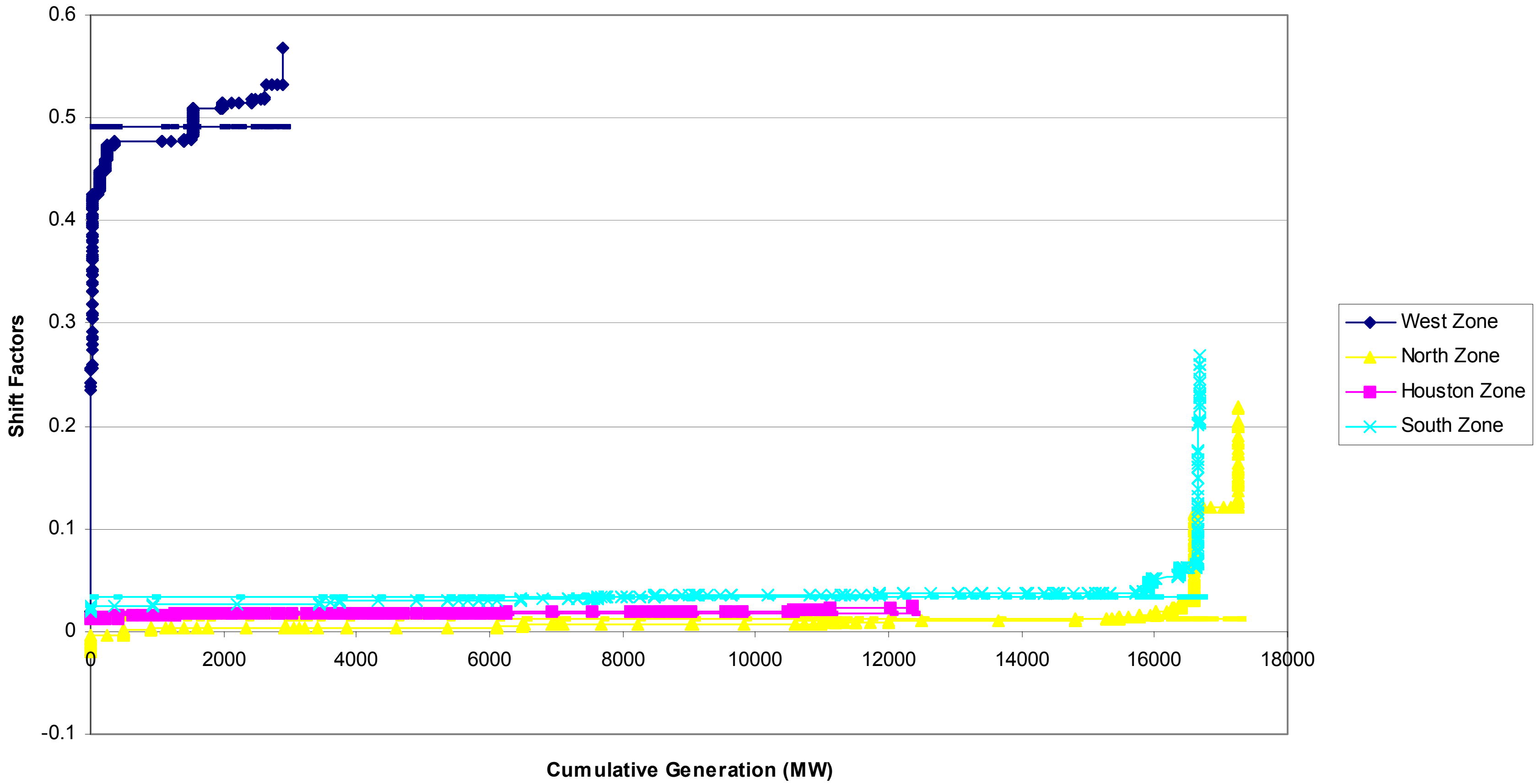


Figure 10: February 2003 Shift Factors to Sandow - Temple

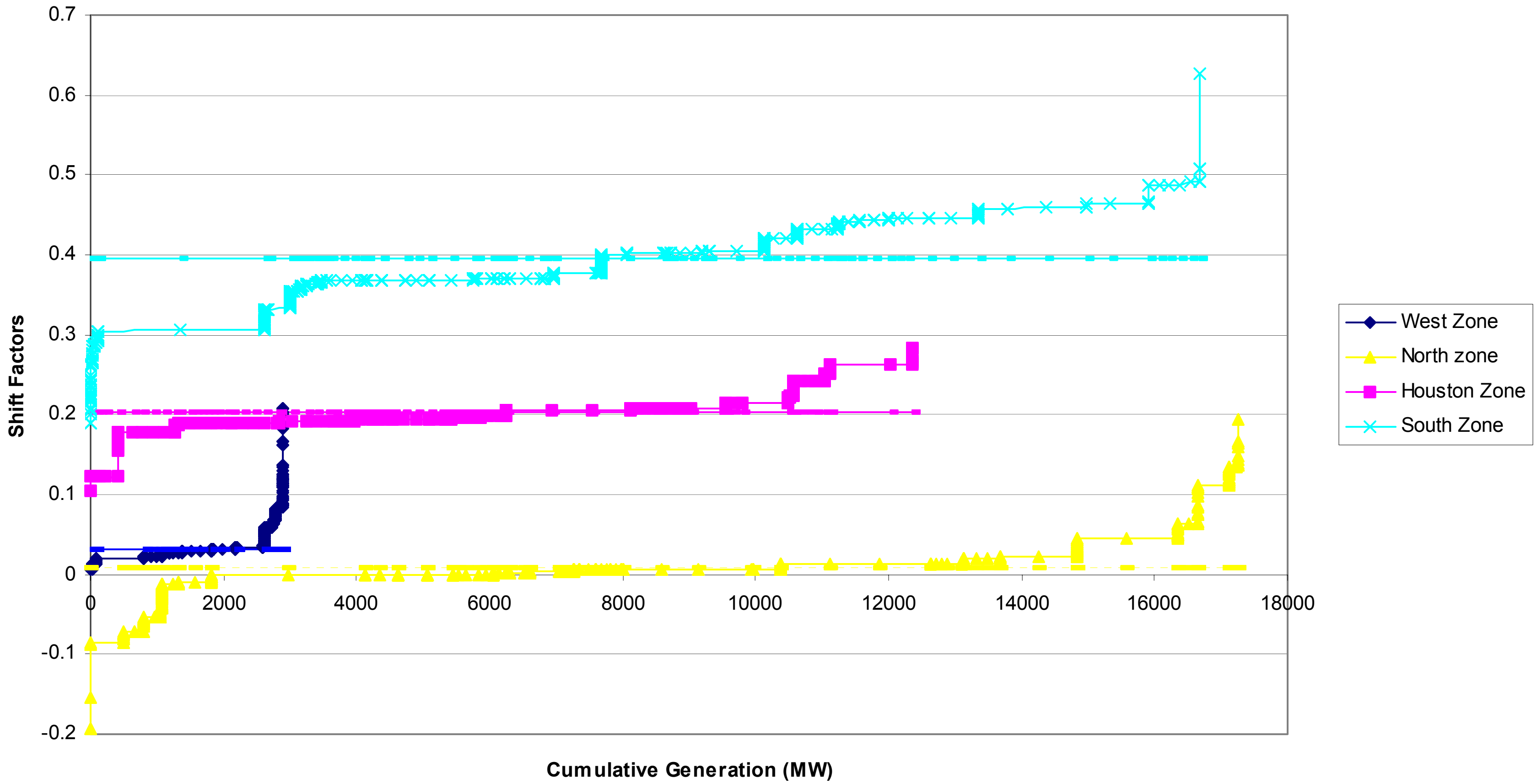


Figure 11: February 2003 Shift Factors to STP-Dow

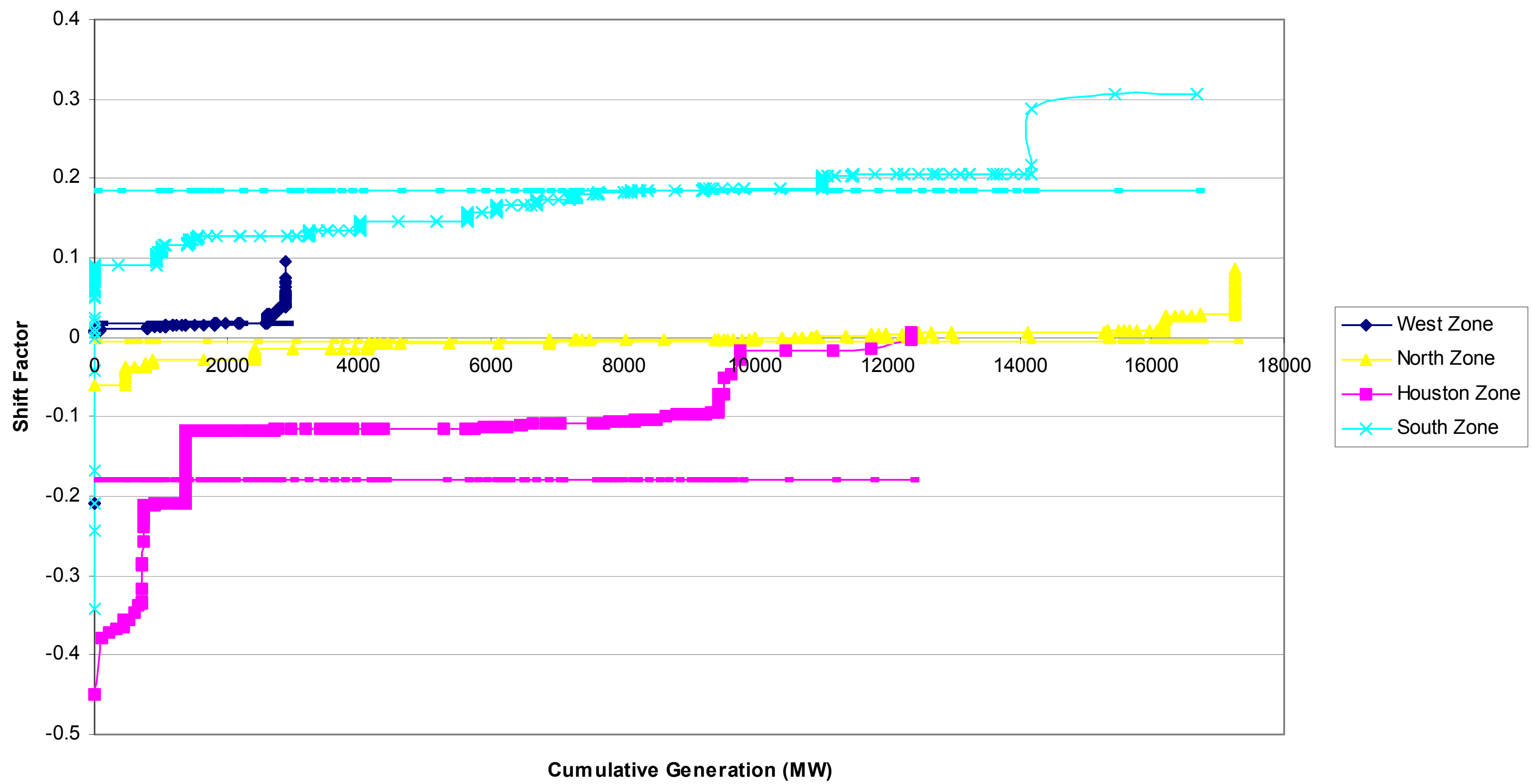


Figure 12: March 2003 Shift Factors to Graham-Parker

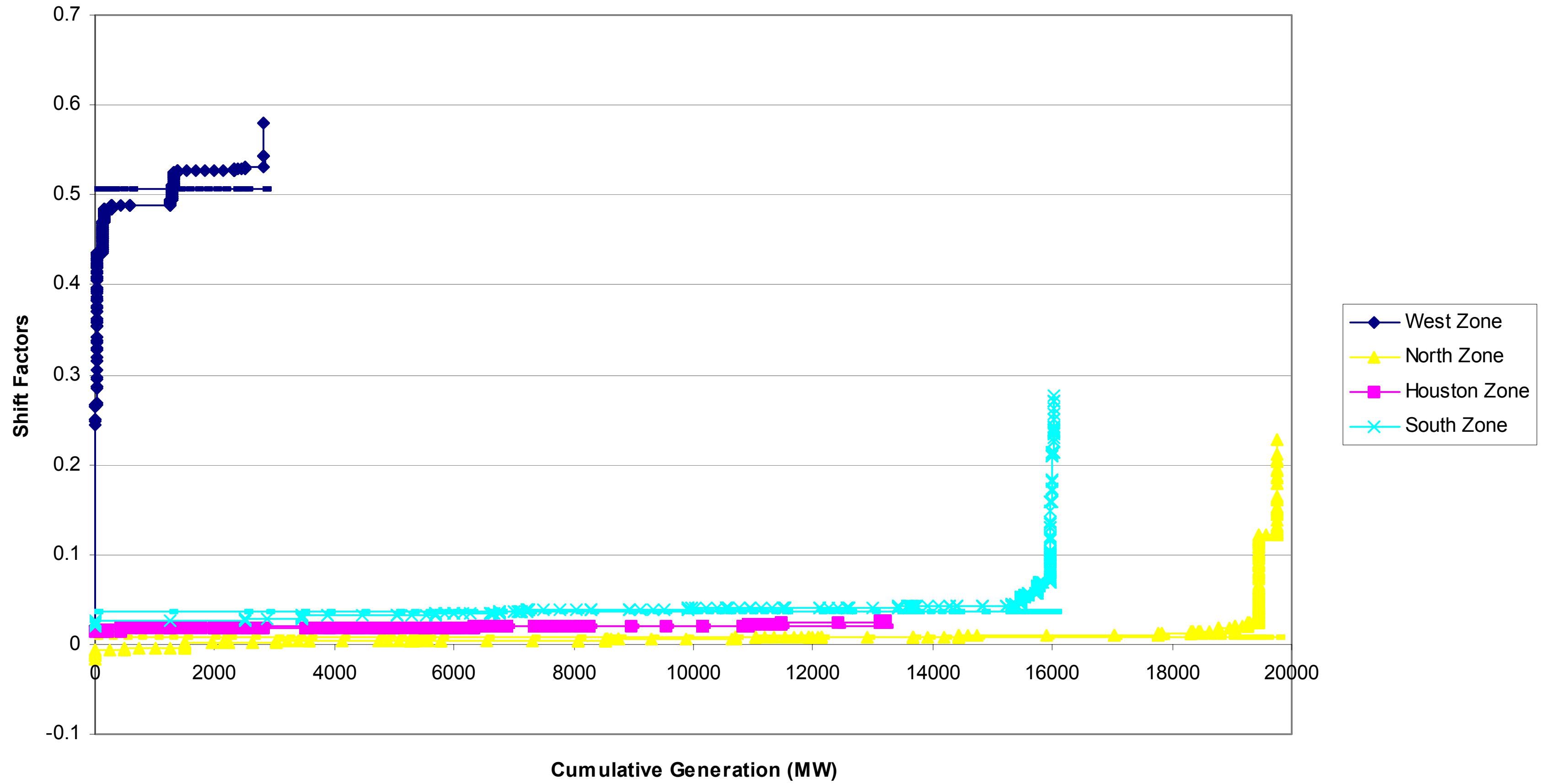


Figure 13: March 2003 Shift Factors to Sandow - Temple

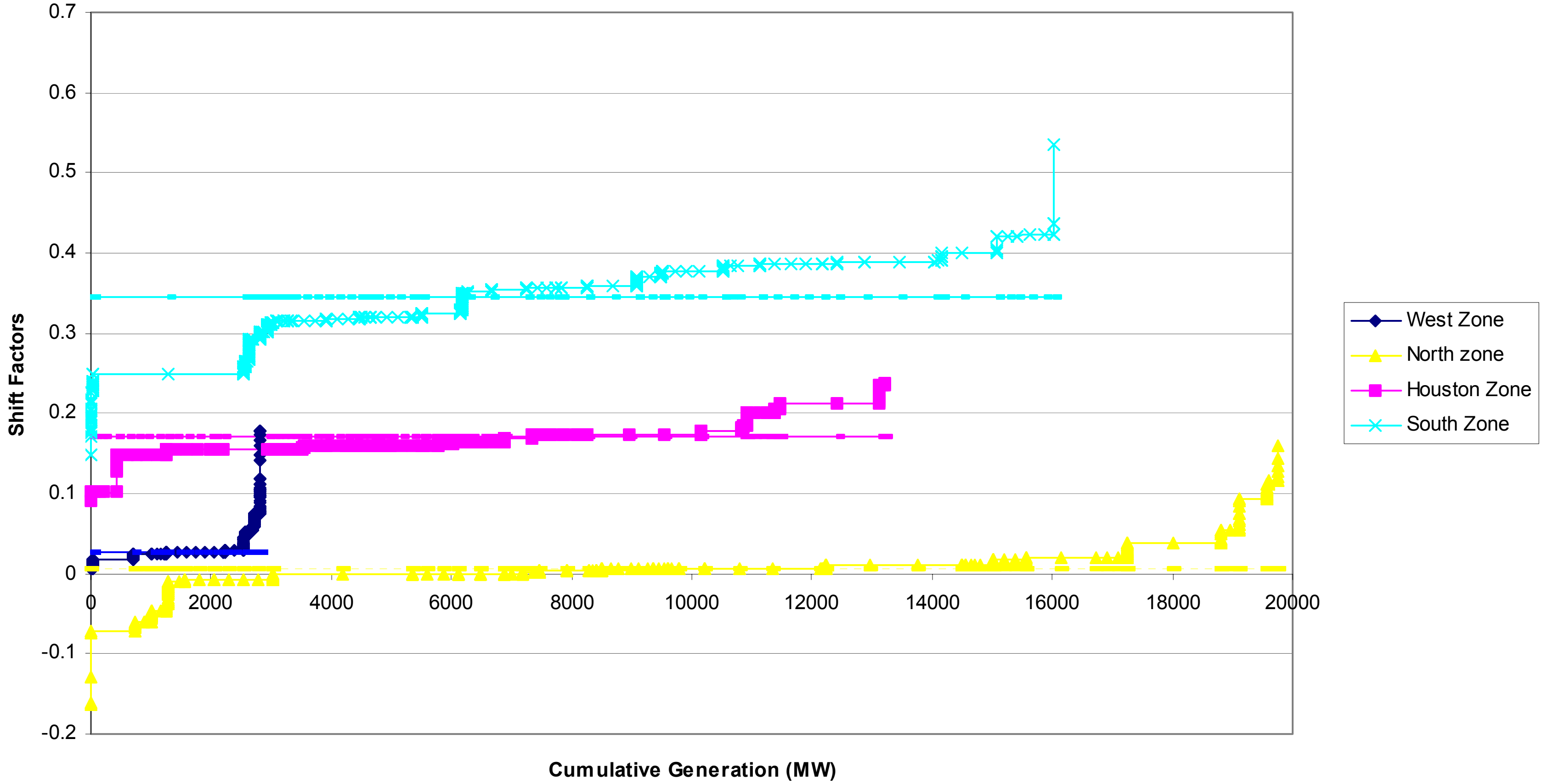


Figure 14: March 2003 Shift Factors to STP-Dow