

**Draft**

**Appendix B**

**Stochastic Emission Inventories of Continuous Emissions  
(Mort Webster, University of North Carolina)**

## The Stochastic Emissions Inventory Generator: Methodology and Assumptions

### 1. Overview

This Appendix details the methods and assumptions for constructing the stochastic emissions inventory generator. This tool has been developed as a part of project H13 under the Houston Advanced Research Consortium (HARC). Project H13 develops tools and methods to support the mid-course correction of the ozone state implementation plan (SIP) for the Houston-Galveston, Texas region. One of the primary goals of project H13 is to improve the models' ability to represent the large amount of temporal variability observed in VOC emissions, to test the effect of variability in emissions on ozone exceedences, and to test potential regulatory designs for reducing the frequency and/or magnitude of ozone exceedences. This document describes the tool developed to simulate the variability in VOC emissions from industrial point sources.

The layout of this Appendix is as follows. In Section 2, we describe the different types and relative contributions of industrial point sources of VOC emissions. In section 3, we describe the point sources for which we currently have observations available, and show the variability that occurs in individual sources. Section 4 outlines the approach used to model and simulate the stochastically varying emissions from each sample source. Section 5 describes the method for simulating emissions from the entire Houston-Galveston point source emissions inventory. The estimated parameters and probability distributions for each observation set are given in detail in the attachments.

### 2. Industrial VOC Point Sources

Figure 1 shows a rough breakdown of industrial sources of VOC emissions in Houston by source type. Half of the emissions are considered "fugitives", a blanket term for multiple, small leaks within an industrial facility. About a third come from flares, emission points that can be fed from a variety of processes in the facility. Under ideal conditions, the flare combusts up to 99% of the outgoing VOCs. Nevertheless, these flares exhibit extremely wide variability, as shown in the next section. The other two large categories are cooling towers and vents, each contributing about 8% of the annual total VOC emissions.

Not all flares have emissions of the same magnitude. In fact the largest 8 flares account for nearly a third of annual emissions (Figure 2). The top 19 account for 50% of annual emissions. And of course the speciation, the percentage of emissions that are the highly reactive VOCs, varies among sources, and for each source, varies over time.

Figure 1: Relative Contribution of VOC Emissions by Point-Source Type in Houston

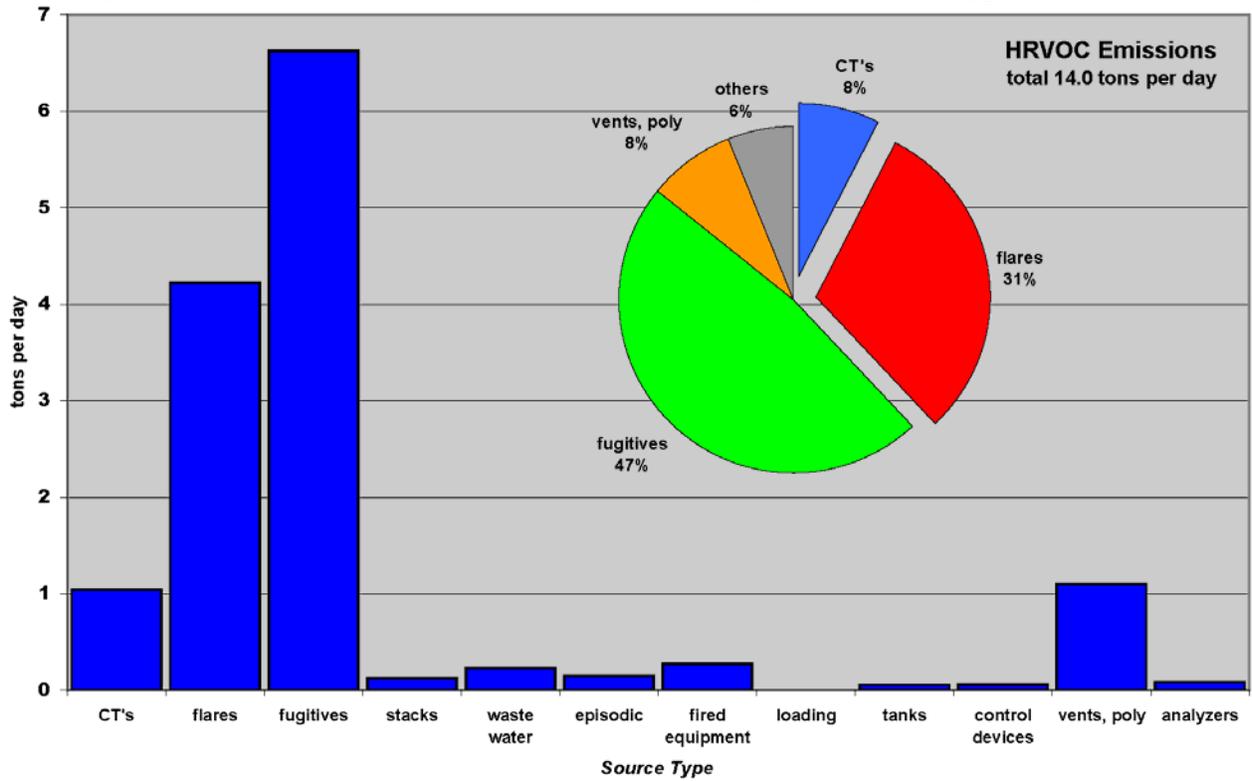
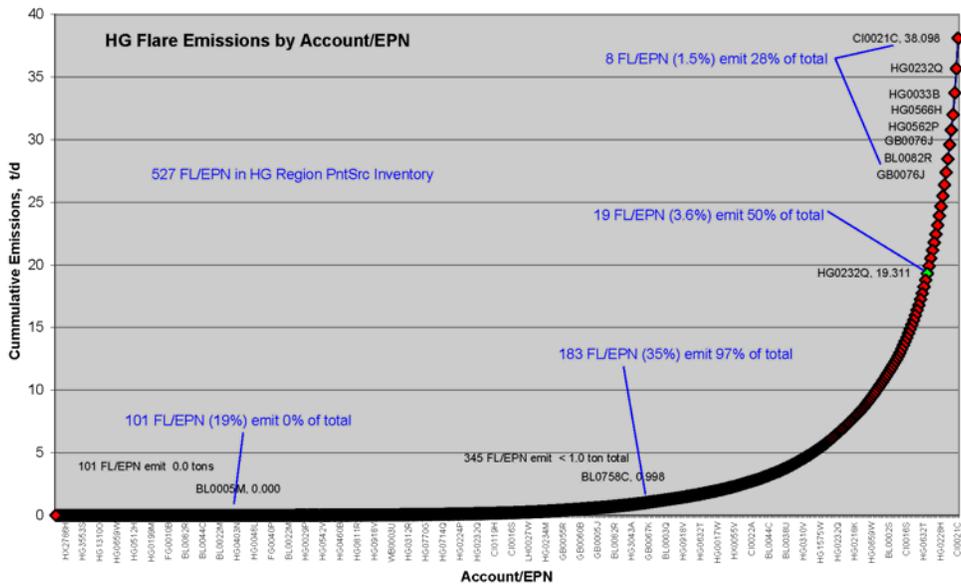


Figure 2: Cumulative Distribution of VOC Emissions from Flares in Houston



### 3. Observed Variability in Individual Point Sources

Figure 3 shows the hourly measurements of VOC emissions over the course of a year from a typical flare at an industrial facility. Although the annual mean emissions are in fact lower than the permitted level, this allows occasional high spikes of emissions. All but two of these spikes were above the daily permitted level. We have obtained similar sample sets for 16 sources from different facilities, mostly flares and a few cooling towers (Table 1). All exhibit significant variability, although the details of the pattern also vary from one flare to another. Figure 5 shows the emissions from four different sources; note that each one has a different pattern of variability. The temporal pattern of emissions from each source is given in the attachments (along with simulations of each source).

Figure 3: VOC mass flow from a typical flare

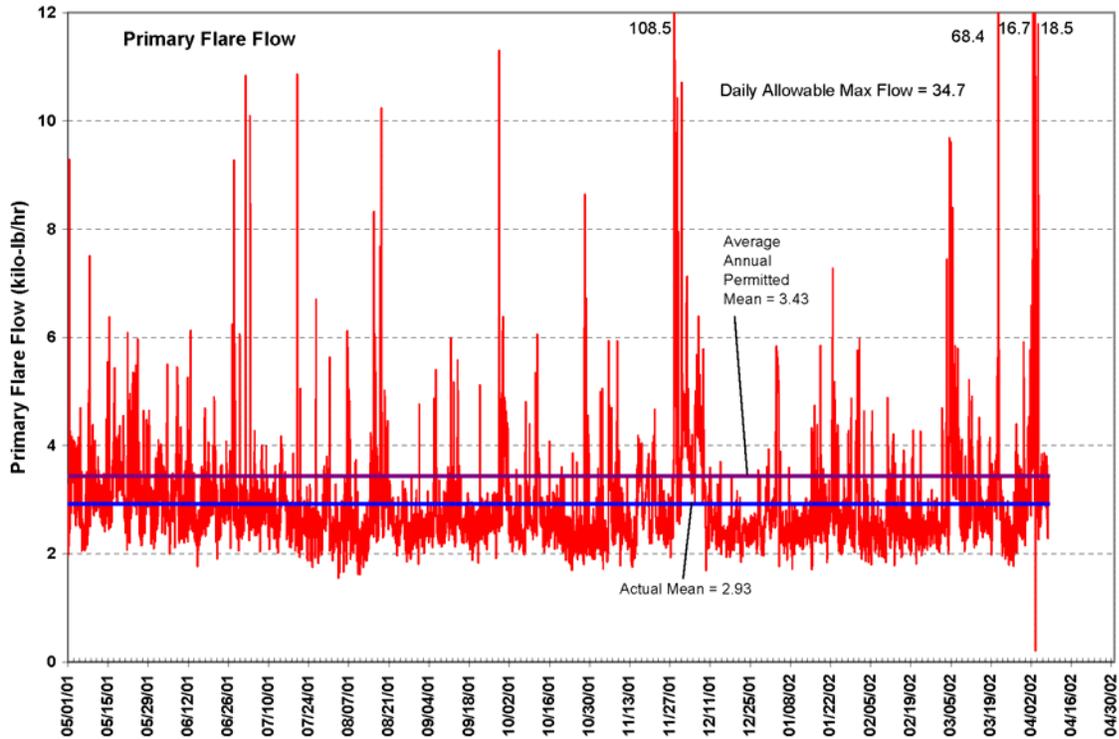
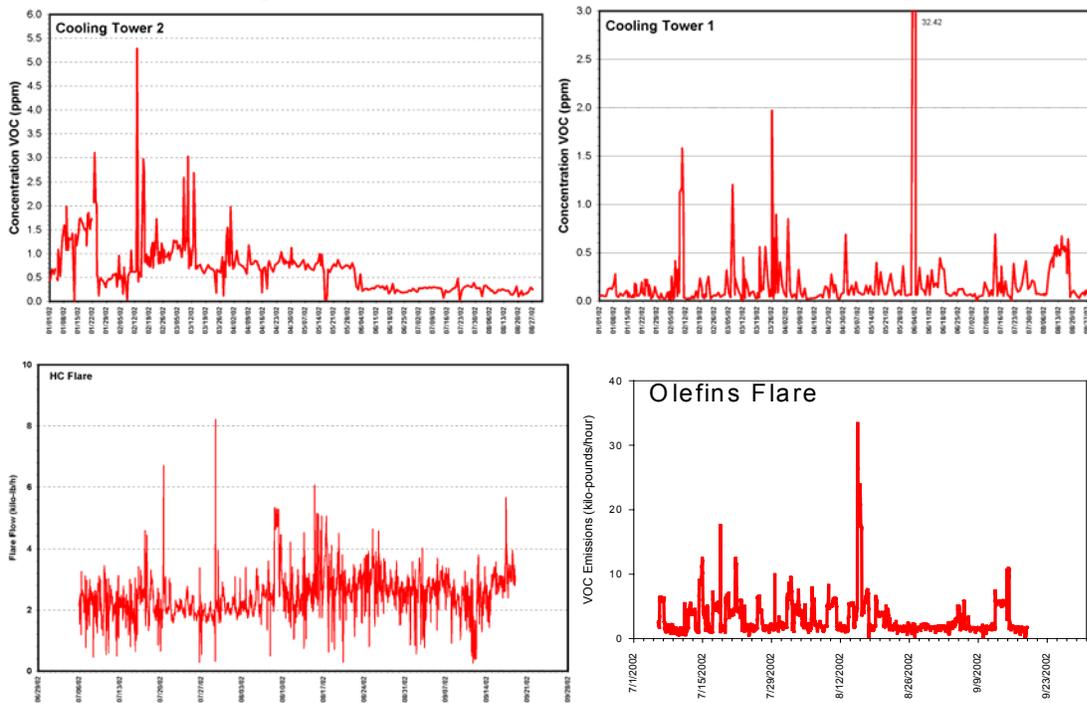


Table 1: Sample Sets for Emission Sources

Name	Type	# of Observations
Flare 1	Flare	8208
Flare 2	Flare	720
Flare 5	Flare	3624
HC Flare	Flare	1800
Olefins Flare	Flare	1800
FCCU	Flare	17533
SRU Flare	Flare	17543
Mercox Flare	Flare	17543
Low Pressure Flare	Flare	17543
General Service #1	Flare	17543
General Service #2	Flare	17543
Cooling Tower 1	Cooling Tower	314
Cooling Tower 2	Cooling Tower	340
Olefins OP3 Total	Flare	10799
ALKY Flare	Flare	10799
ESO Flare	Flare	10799

Figure 4: Emissions from two Flares and two Cooling Towers



#### 4. Simulating Individual Point Sources

The simplest approach is to fit the emissions to a probability distribution, and generate random samples from that distribution. For example, the emissions of Flare 1 (Figure 3) can be well approximated by a lognormal distribution (Figure 5a). But random generation from that lognormal produces an emissions pattern (Figure 5b) with no resemblance to the actual behavior of Flare 1.

To develop a reasonable model of emissions, need to explicitly represent more detail about the emissions behavior. Upon closer examination of Flare 1 emissions (Figure 6), we can see that there are several different distinct “modes” of variability. There is one component, which we will refer to as “nearly constant”, where both the mean and the variance are lowest. This may in fact correspond to some “base” operation level for the plant processes. The second component we label as “routinely variable”. This mode will have higher mean and higher variability in emissions than the “nearly constant” mode, and include moderate emission spikes that are still within the legal permitted level. The third mode we call “allowable episodic”, which consists of shorter periods of much higher emissions spikes, and larger variability. These are also within permitted levels, but can release significantly large amounts of VOCs within an hour. This mode usually corresponds to minor mechanical failures within the process, which can sometime take hours or days before it is corrected. Finally, the highest mode, “emission events” or “upsets”, will be treated separately. These emissions do exceed permitted levels, and there are currently rules in place to address them. We focus on the other three modes in order to test whether the legally permitted variability in emissions is contributing to ozone exceedences.

To model these different components, we apply statistical mixture theory. In other words, we identify each hourly emission as belonging to one of the components, fit probability distributions to each component separately, and model the probability of switching from one component to another.

Figure 5: Simulation from Simple Univariate Distribution

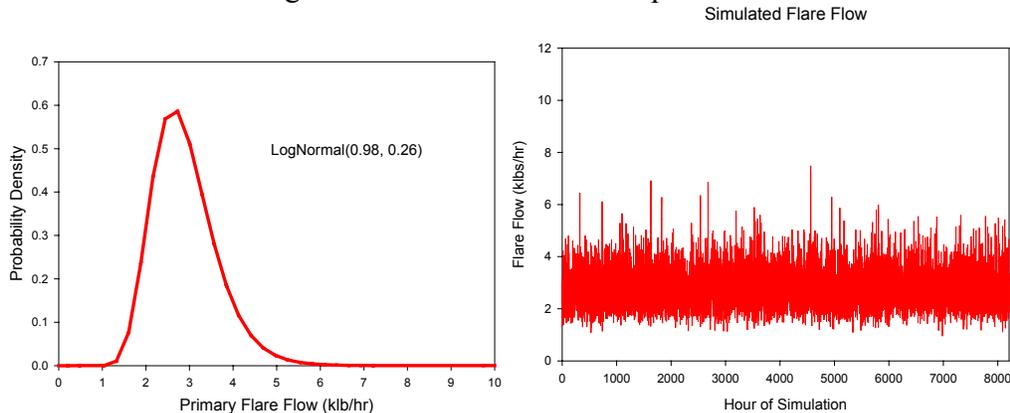
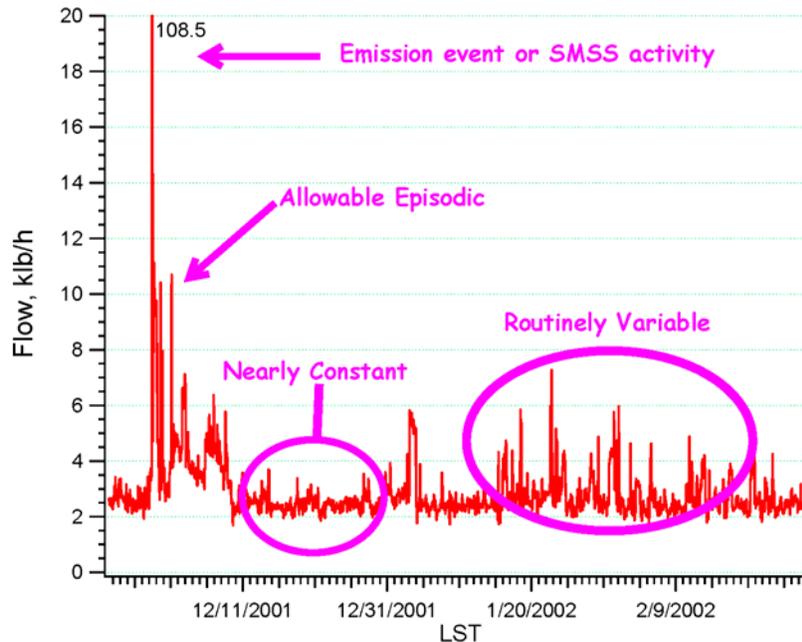
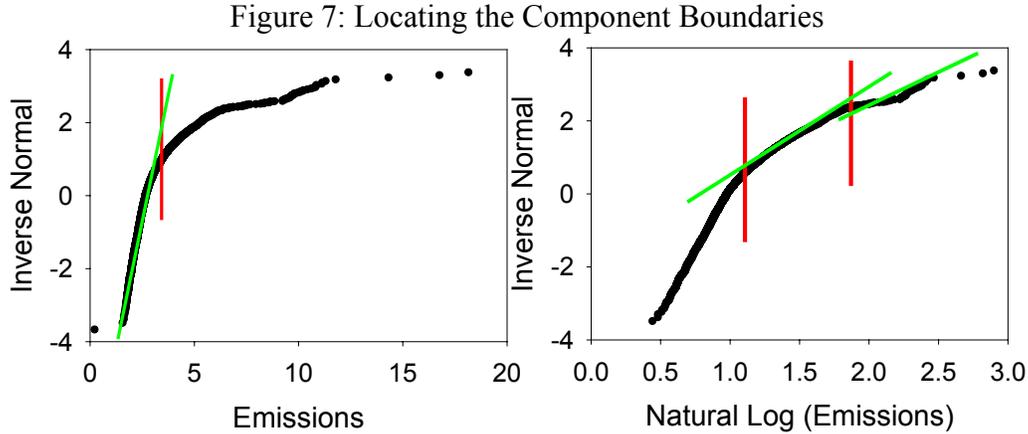


Figure 6: Different “Modes” or “Components” of Emissions Variability



The first step to building a model of this behavior is to assign each hour’s emissions as belonging to one of the components. Once we can subdivide the observations, we can then estimate the parameters for each component distribution from its observations. The current version uses a simple graphical statistical technique, as illustrated in Figure 7. The observations from the source (e.g., Flare 1) are sorted in order of size, rather than by time. Each size-ordered emission value is then assigned its number in the order (e.g., 1 to 8208). The inverse normal is then calculated this rank number divided by the total number ( $n/8208$ ). We then graph the emissions against the inverse normal of the rank of the emissions, as shown in Figure 7a. If any segment of this curve is a straight line, it is reasonable to assume that that range of emissions are normally distributed. Furthermore, the mean and standard deviation can be estimated by the range midpoint and the slope, respectively. We also repeat this procedure with the natural log of the emissions, and graph again versus the inverse normal (Figure 7b). A straight line segment on this graph indicates a lognormal distribution. Looking at the graphs for Flare 1 in the figure, it appears that the first component (“nearly constant”) is normally distributed, and the second and third components (“routinely variable” and “allowable episodic”) are lognormally distributed. A least-squares regression line is fit to each line segment to estimate the slope, and thereby the standard deviation. The fitted regression equation is given in the table in the attachments for each component for every source.



Once the boundaries between components have been identified, we can fit probability distributions to each component, either normal or lognormal. Figure 8a shows probability distributions fit to each of the three components of Flare 1 emissions. Since the process is in each component some fraction of the time, we can show the resulting total emissions uncertainty by scaling each pdf by its proportion of total emissions (Figure 8b).

Finally, we model the temporal behavior by combining three elements: using state transition probabilities, probabilistic time within one mode, and imposing autocorrelation during emissions sampling. In any of the observed emission examples, one can see that the process often tends to remain in one mode for some period of time, the length of which also varies. Having identified which component each emission value “belongs to”, we can resort by time and obtain the number of hours the process remains in one state before switching to another. Using this data, we fit exponential probability distributions to the number of hours a process will remain in each mode. Figure 9 shows these distributions for the time within each component for Flare 1. The transition probabilities are not derived for a full Markov model for this version. We simply use the relative proportion of hours in each component as the probabilities of moving to that state at the next transition time. Finally, the emissions are highly autocorrelated, both within and across components. Since they result from a continuous industrial process, this should not be surprising. Based on the samples obtained, we impose an autocorrelation of 0.99 (with the previous hour’s emissions) on each emissions sample generated.

Figure 8: Component Distributions for Flare 1

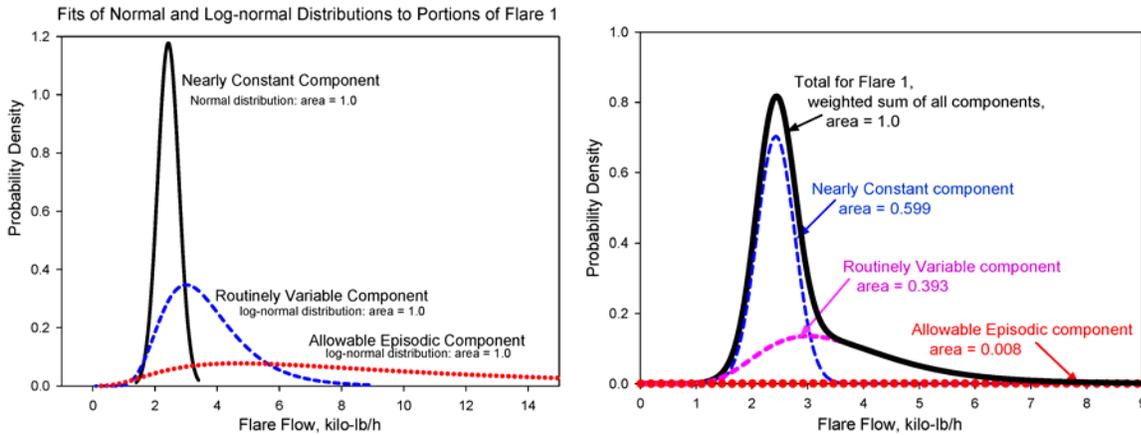
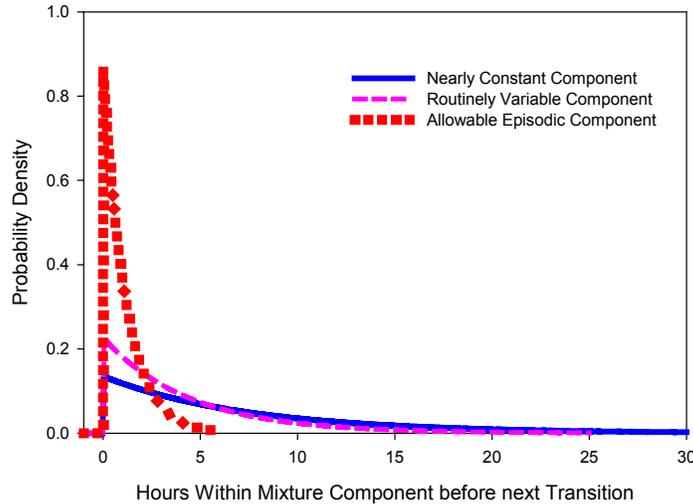


Figure 9: Time within each Component before next Transition



To summarize, the representation of any single point source includes the following information:

- Three probability distributions of emissions, one for each component, either Normal or Lognormal, with mean and standard deviation
- Three exponential distributions of the time in hours spent within each component
- The proportion of emissions associated with each component.

The algorithm for generating emission samples is:

1. Use proportions to randomly select which component is current.
2. Randomly draw the number of hours to remain in this component
3. Randomly draw emissions from the distribution for this component, imposing a correlation of 0.99 with the previous hour's emissions
4. If number of hours to remain here are zero, repeat from Step 1), otherwise repeat from step 3).

This procedure is drawn schematically in Figure 10 as a flow chart.

Figure 10: Procedure for Simulation Emissions from a Point Source

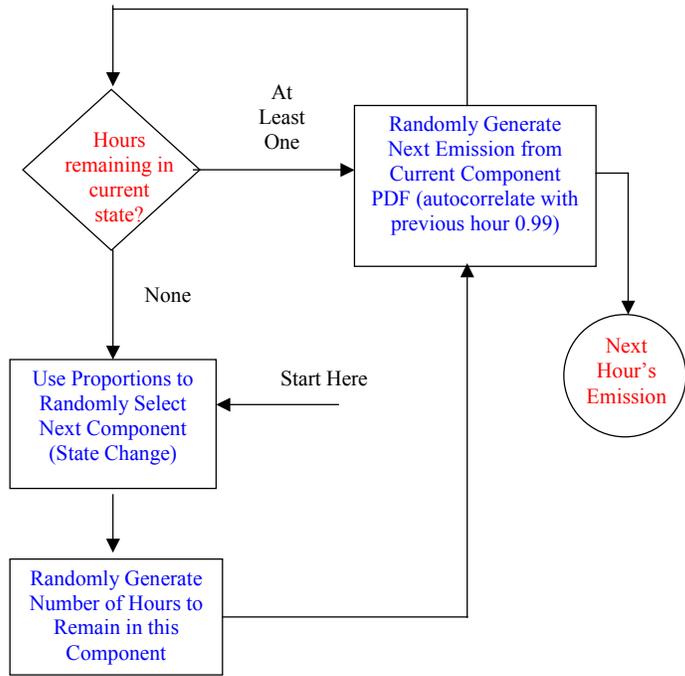
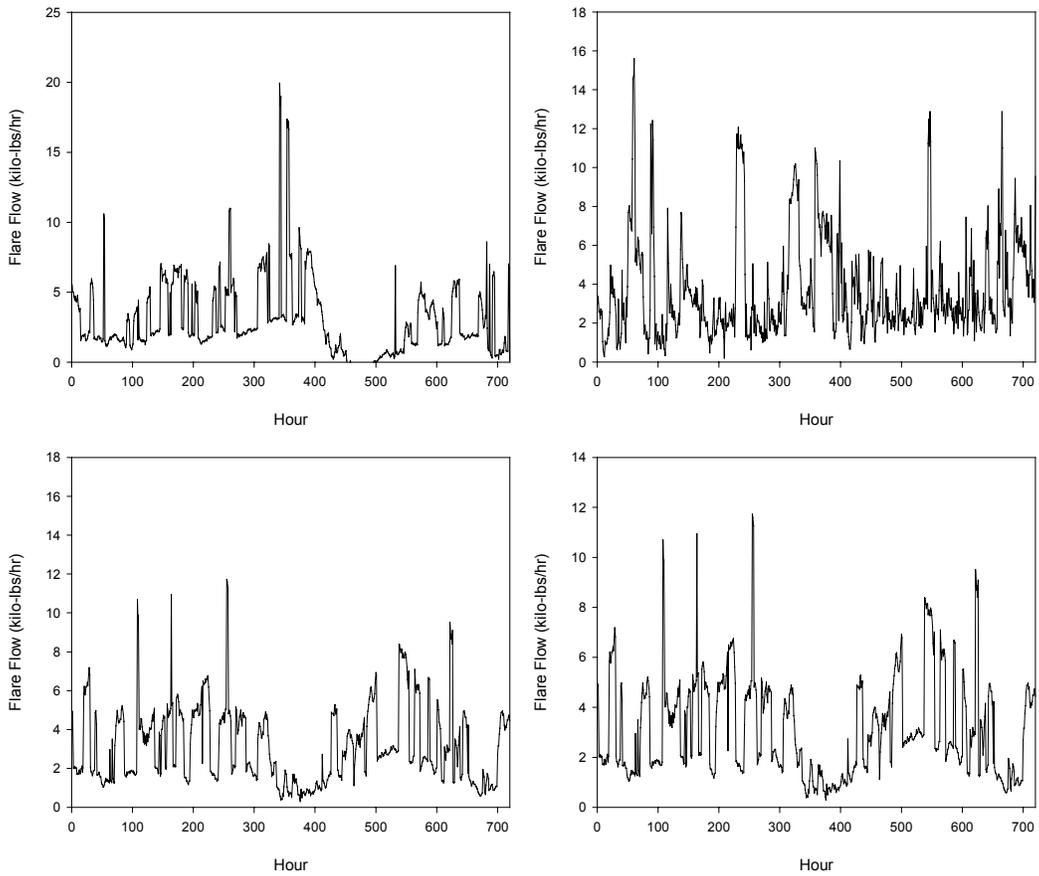


Figure 11: Actual Emissions and Three Simulations for Flare 2



## Draft

Figure 11 shows an example of the results obtained from this simulation method. One of the panels is the actual emissions measured at “Flare 2”. The other three are simulated from the estimated stochastic model. Can you guess which one is the real Flare 2 (answer given in the attachments)?

### 5. Simulating Houston-Galveston Emissions Inventory

The final step in the procedure is to apply these models to the full point source emission inventory for Houston-Galveston. The Stochastic Emission Generator (SEG) is designed to read in the standard emission inventory database used by TCEQ. For all flares and cooling towers, we then assign one of the known source models, just described. Currently, we make this assignment randomly. In future versions, we hope to combine knowledge of process and facility types to make deliberative assignments for each point source in the inventory. The assigned mixture model is then scaled so that the mean will be the annual average emissions from the inventory, and preserves the relative variance (the ratio of the standard deviation to the mean is preserved in the scaling). This allows SEG to simulate time-varying VOC emissions from all point sources in Houston. These results will then be used as inputs to air quality models to explore the effects on ozone production.

In this section we present the variability in aggregate VOC emissions that results from imposing variability in each individual source. Figure 12 shows one possible hourly profile of total VOC emissions for 200 days. This is an “instance” or random sample for the aggregate of all VOC emissions over all of Houston. Many other instances are possible. One way to describe the variability in total VOC emissions is with a probability density function, as shown in Figure 13.

Figure 12:

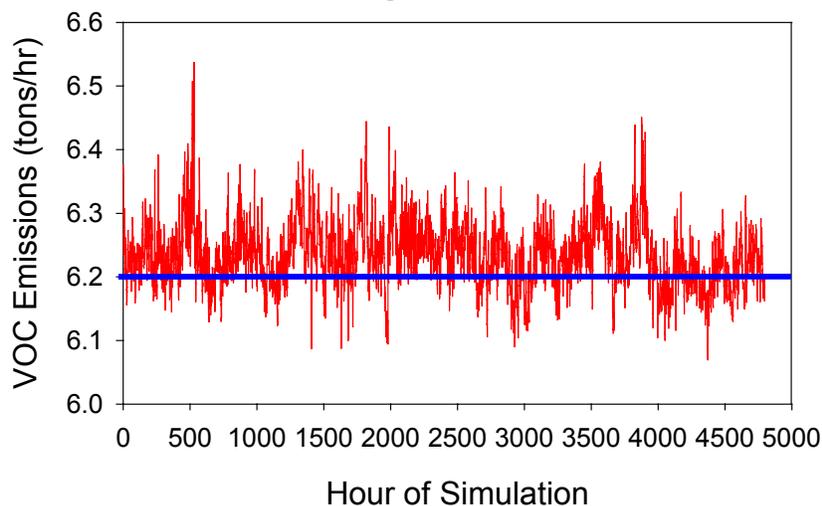
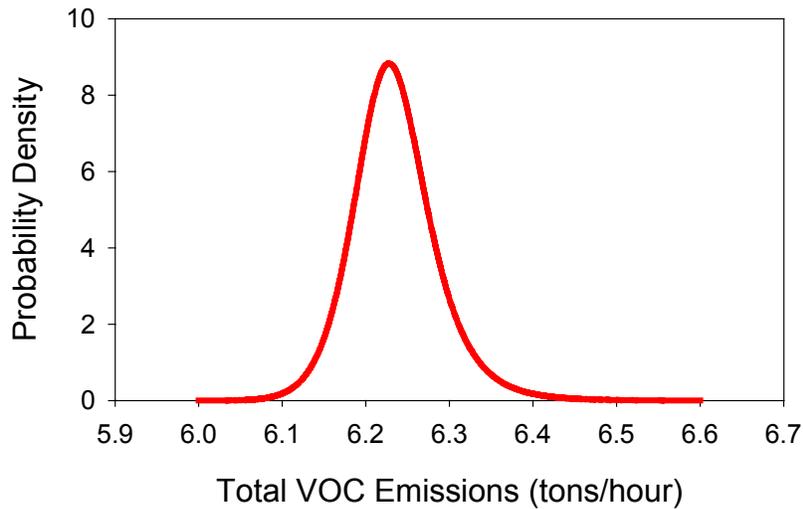


Figure 13:



The variability in total VOC emissions for the Houston area is an initial way to get a sense of the variation in drivers of ozone formation. However, because ozone formation at any location will depend on the local concentrations of NO<sub>x</sub> and VOC, the Houston-wide aggregate will probably underestimate the local variability. We can focus instead on a specific geographic sub-region of interest. As an example, we present here the results for a region south of the ship channel and including Deer Park (Figure 14). Ignoring transport for simplicity, we can extract the VOC emissions within these latitude-longitude boundaries. Figure 15a shows an instance of total VOC emissions for this region. We can also examine specific VOC species of interest, particularly the highly reactive species. Figure 15b shows the ethylene emissions for the region during the same instance.

Figure 14:

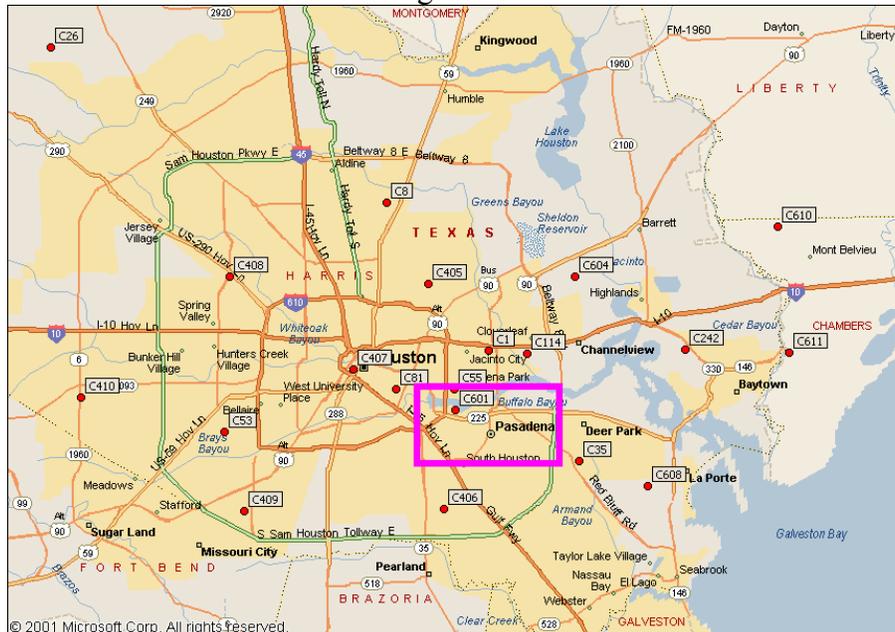
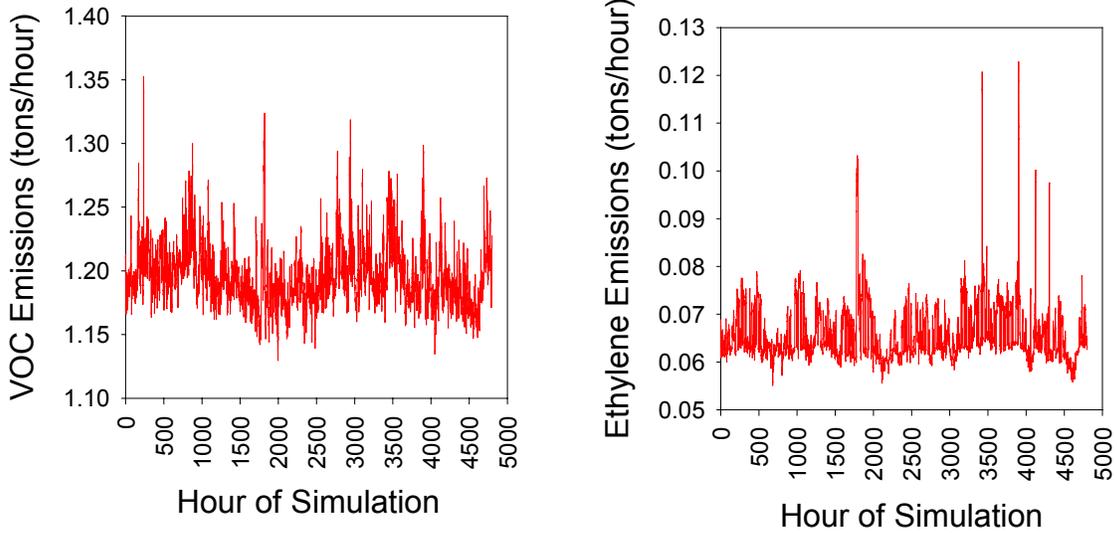
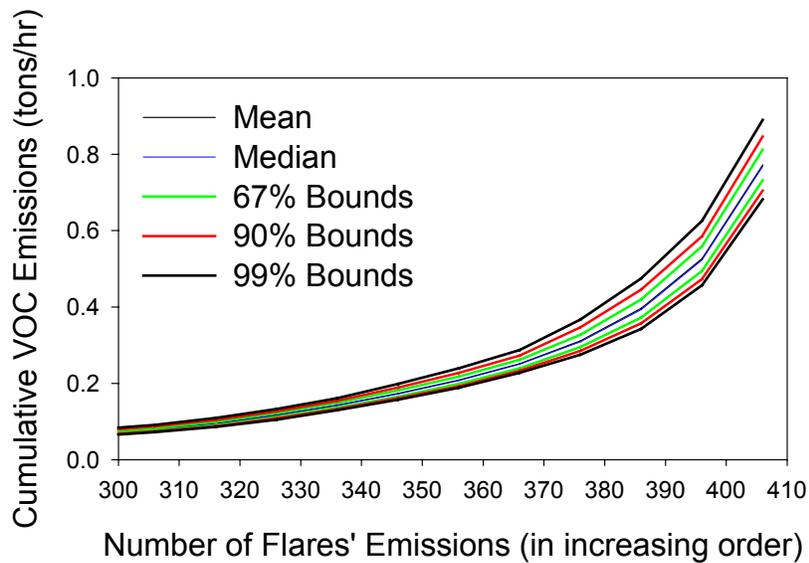


Figure 15:



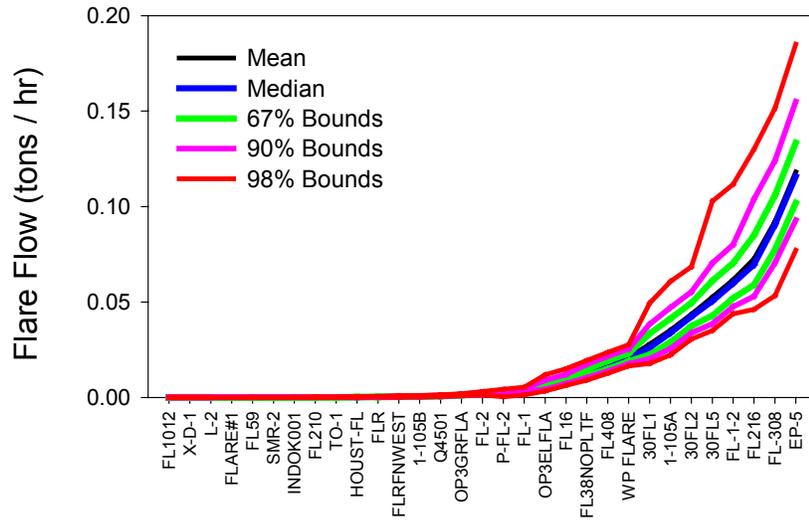
Finally, we can use these initial results to look at which point sources, in terms of relative size, are driving the variability in total emissions. Figure 16 shows the cumulative VOC emissions from flares in order of increasing size of annual average emissions. The largest 50 flares, out of 410 flares total in the emissions inventory, are driving the majority of the variability. Again, even more relevant for ozone formation is the variability within a small region. Figure 17 shows the same cumulative emissions graph for the 32 flares within the “deer park” subregion. Of these, the largest 8 flares cause most of the observed variation.

Figure 16:



Draft

Figure 17:



Draft

Attachments

Appendix 1: Table of Fitted Parameter Values for Sample Sources

Source Name	Total Mean	Component	Interval start	Interval finish	Standard Deviation	Mean	Mean (LN value)	Equation	Number of Observations	Proportion (%)	Normalized Proportion (%)	Exponential mean time
FCCU Obs. 17533	21.87	1	10.13	23.95	1.30	20.00		$y = 3.1284x + 21.012$	12,383	70.6	0.7000	195.31
		2	23.96	45.69	14.53	29.38		$y = 11.929x + 14.798$	4,270	24.3	0.2990	17.85
		3	45.96	47.71	1.20	54.02	3.97	$y = 1.1989x + 0.9081$	12	0.1	0.0010	12.60
unused								868	5	1.0000		
Olefins Flare Obs. 1800	3.09	1	0.26	2.13	0.53	1.60		$y = 0.5317x + 1.94$	1,078	59.9	0.599	10.18
		2	2.14	12.46	3.10	4.91		$y = 3.7624x + 1.4663$	700	38.9	0.389	14.58
		3	17.02	26.31	0.42	19.25	2.95	$y = 0.4231x + 1.8602$	21	1.2	0.012	6.50
unused								1	0.0			
Cooling Tower 1 Obs. 314	0.30	1	0.01	0.23	0.10	0.08	-2.69	$y = 0.8666x - 2.3558$	243	77.4	0.779	5.72
		2	0.24	1.97	0.30	0.54	-0.73	$y = 0.9988x - 2.025$	64	22	0.221	0.80
									7	0.6		
unused												
SRU Flare Obs. 17543	9.33	1	1.13	11.29	0.50	8.30		$y = 1.1922x + 8.426$	16,728	95.4	0.850	103.20
		2	11.32	14.80	1.20	13.22	2.58	$y = 1.1485x + 0.5054$	365	2.1	0.135	6.00
		3	14.84	73.52	3.00	27.34	3.17	$y = 3.5354x - 4.3525$	1.5	0.015	27.00	
unused								450	1	1.000		
General Service #1 Obs. 17543	21.70	1	13.45	24.46	1.66	21.45		$y = 1.6552x + 21.59$	16,892	96.3	0.610	155.47
		2	24.47	28.76	0.38	25.97	3.26	$y = 0.3834x + 2.5017$	405	2.3	0.320	2.00
		3	28.87	47.05	0.43	32.19	3.47	$y =$	233	1.3	0.070	5.78

Draft

unused							$0.427x + 2.3918$					
<b>General Service #2</b>	17.85	1	11.05	20.76	1.31	17.80		13	0.1			
Obs. 17543		2	22.77	21.42	0.08	21.03	3.05			17.322	98.7	0.650
		3	21.43	36.97	2.90	26.26	3.26			141	0.8	0.250
unused								10	0.1	70	0.4	0.100
<b>Merox Flare</b>	500.21	1	0.33	247.84	874.78	4.63				307	1.8	0.056
Obs. 17543		2	248.00	699.90	105.19	494.09				15,705	89.5	0.914
		3	700.13	819.55	0.001	754.96	6.63			1,152	6.6	0.030
unused								379	2.1			10.65
<b>Low Pressure Flare</b>	25.54	1	18.93	22.80	0.70	20.98	3.04			200	1.1	0.082
Obs. 17543		2	22.81	29.06	0.04	25.43	3.23			16,830	95.9	0.900
		3	29.08	35.69	0.50	31.87	3.46			314	1.8	0.018
unused								199	1.2	1,000		38.37
<b>Flare 1</b>	2.92	1	1.56	2.79	0.34	2.43				4,913	59.9	0.599
Obs. 8208		2	2.80	6.44	0.36	3.49	1.23			3,223	39.3	0.393
		3	6.58	18.15	0.80	9.18	2.19			68	0.8	0.008
unused								4	0.0			4.41
<b>Cooling Tower 2</b>	0.75	1	0.007	0.336	6.30	0.23				98	28.8	0.291
Obs. 340		2	0.356	0.873	0.39	0.67				148	43.5	0.439
		3	0.892	3.099	0.68	1.39	0.28			91	26.8	0.270
unused								3	0.9			2.29
<b>Flare 5</b>	1391.8	1	224.43	1021.9	230.52	758.83				1,757	48.5	0.512
	1			7								2.09

Draft

Obs. 3624		2	1022.1 0	2738.4 3	1250.5 0	1621.7 2		$y = 1197.5x + 945.63$	1,706	47.1	0.447	5.36
(linear)		3	2741.8 2	5993.5 2	2641.0 0	3303.6 8		$y = 2641x - 2065.3$	147	4.1	0.041	1.40
(log)		3		0.61		8.08	$y = 0.7108x + 6.6388$					
unused									14	0.3	1.000	
<b>Flare 2</b>	3.71	1	0.200	2.806	0.80	1.89		$y = 0.9902x + 2.6739$	360	50.0	0.501	2.90
Obs. 720		2	2.810	10.376	2.00	4.99		$y = 4.1013x + 2.1949$	331	46.0	0.460	3.50
		3	10.439	14.704	0.33	11.65	2.45	$y = 0.2643x + 1.8876$	28	3.9	0.039	2.63
unused									1	0.1		
<b>HC Flare</b>	2.47	1	0.26	1.58	1.20	1.15	0.07	$y = 1.1226x + 2.2007$	128	7.1	0.071	0.10
Obs. 1800		2	1.59	4.12	0.30	2.51	0.90	$y = 0.2567x + 0.8758$	1,625	90.3	0.903	2.28
		3	4.21	6.71	0.30	4.86	1.58	$y = 0.2957x + 0.893$	46	2.6	0.026	0.35
unused									1	0.0		
<b>FU5967H2_OP3</b>	4.57	1	0.008	8.57	3.60	4.17		$y = 3.5955x + 4.1001$	8,379	77.6	0.887	41.36
Obs. 10799	without t0:	2	8.571	21.28	0.74	11.86	2.45	$y = 0.7365x + 1.2253$	1,069	9.9	0.113	3.42
	5.18	3	21.29	32.58	0.30	23.60	3.15	$y = 0.2981x + 2.3192$	71	0.01	0.00008	1.77
unused									1,280	12.5	1.000	
<b>FU5852_OP3</b>	4.99	1	0.40	8.40	2.58	5.41		$y = 2.6766x + 5.5055$	7062	65.4	0.843	35.76
Obs. 10799	without t0:	2	8.40	17.19	3.35	10.94		$y = 5.8455x + 1.5412$	1316	12.2	0.157	4.18
	5.19	3	17.20	19.765	0.15	18.47	2.92	$y = 0.1507x + 2.4634$	36	0.003	0.00004	0.32
unused									2,385	22.4	1.000	
<b>FU5854H2_OP3</b>	2.11	1	0.04	0.83	10.34	0.50		$y = 1.5378x + 2.7886$	833	7.7	0.080	18.83
Obs. 10799	without t0:	2	0.83	4.72	0.54	1.97	0.60	$y = 0.5434x +$	9184	85.0	0.879	30.03

Draft

unused	2.19	3	4.73	81.02	1.44	9.77	2.10	0.5344 y = 1.467x - 1.0577	430	4.0	0.041	10.14
<b>Olefins Total</b>	11.67	1	0.12	16.26	4.76	9.86		y = 4.764x + 11.132	9015	83.5	0.841	
Obs. 10799		2	16.27	29.14	9.45	19.85		y = 9.4542x + 6.3886	1559	14.4	0.145	
		3	29.16	112.80	0.81	39.82	3.65	y = 0.8101x + 1.5941	154	1.4	0.014	
unused								71	0.7	1.000		
<b>F632ABSL _ALKY</b>	0.05	1	0.00	0.05	0.02	0.02		y = 0.0203x + 0.0236	8252	76.4	0.830	33.78
Obs. 10799	withou t 0:	2	0.05	0.76	0.34	0.17		y = 0.3443x - 0.3476	1637	15.2	0.165	5.13
		3	0.05	0.76	1.00	0.21	0.83	-0.19 y = 0.2128x - 0.8047	50	0.5	0.005	9.40
unused								860	7.9	1.000		
<b>FI30418_E SO</b>	8.50	1	0.03	4.44	6.45	1.42		y = 6.454x + 6.9654	2,105	19.5	0.227	9.17
Obs. 10799	withou t 0:	2	4.45	15.06	6.02	9.22		y = 6.0187x + 7.7011	6,300	58.3	0.678	11.21
		3	9.47	15.06	30.16	1.35	20.95	3.02 y = 1.3489x + 1.0776	880	8.2	0.095	4.62
unused								1,514	14.1	1.000		

not used:

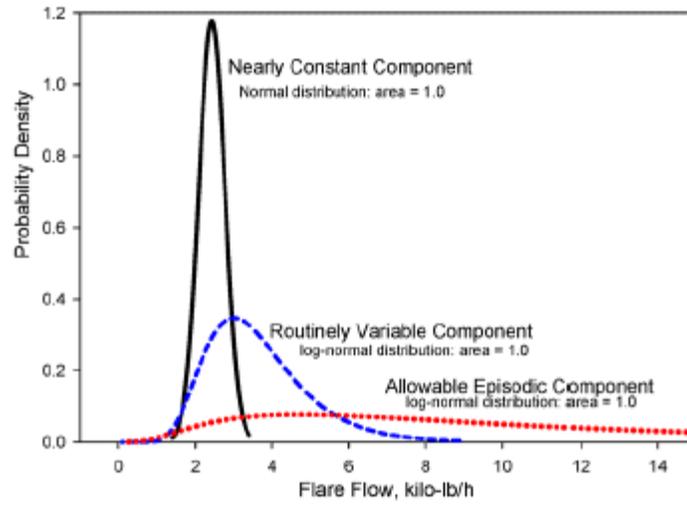
Olefins Flare System Total = FU5967H2\_OP3 + FU5852\_OP3 + FU5854H2\_OP3

WestPropFlareFlow - only has 82 observations scattered throughout the sample

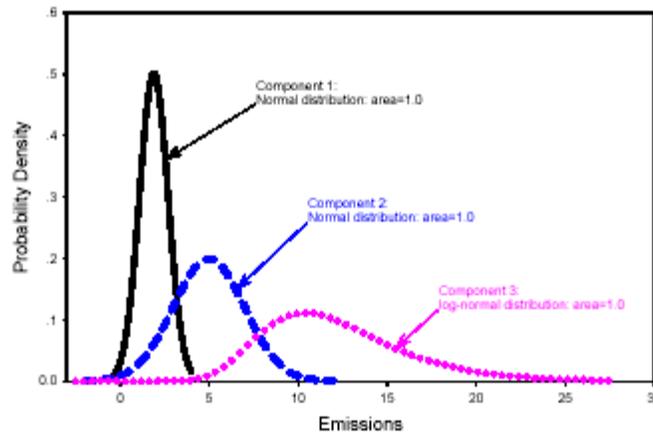
# Draft

## Appendix 2: Graphs of Fitted Component Distributions

Fits of Normal and Log-normal Distributions to Portions of Flare 1

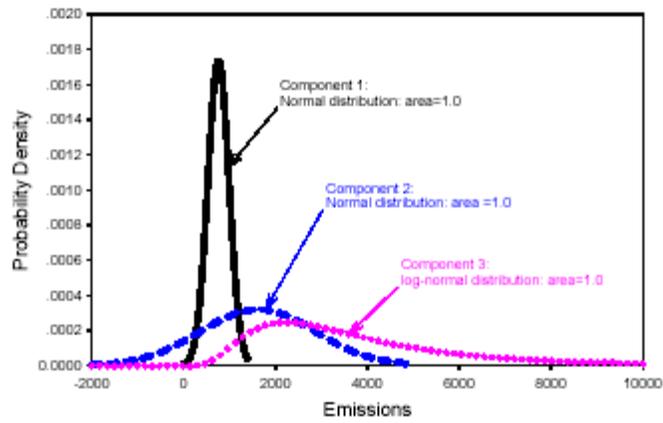


Fits of Normal and Log-normal Distributions to Portions of Flare 2

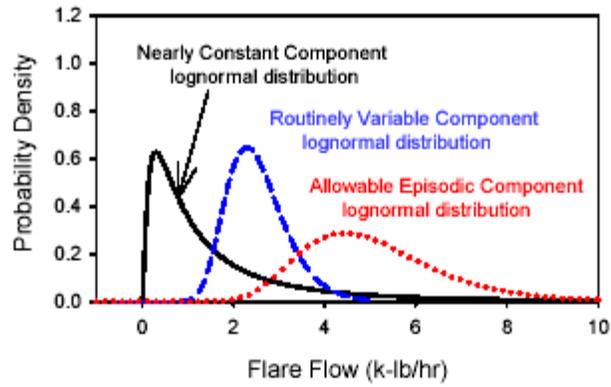


# Draft

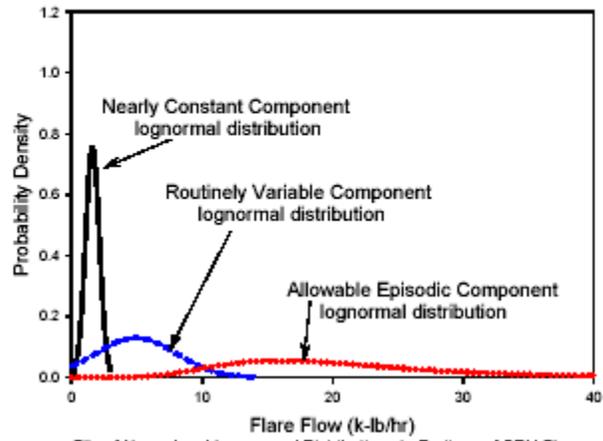
### Fits of Normal and Log-normal Distributions to Portions of Flare 5



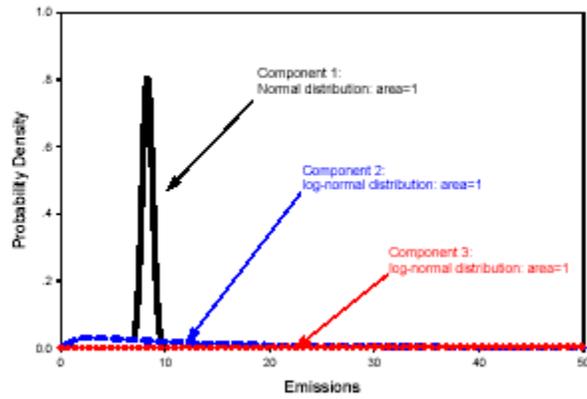
### Fits of Log-Normal Distributions to Portions of HC Flare



Fits of Distributions to Portions of Olefins Flare

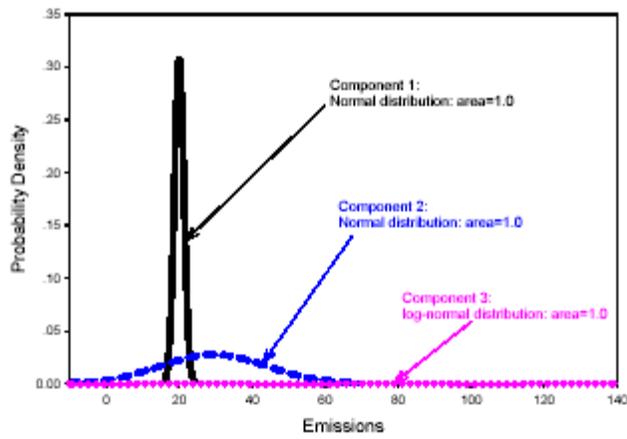


Fits of Normal and Log-normal Distributions to Portions of SRU Flare

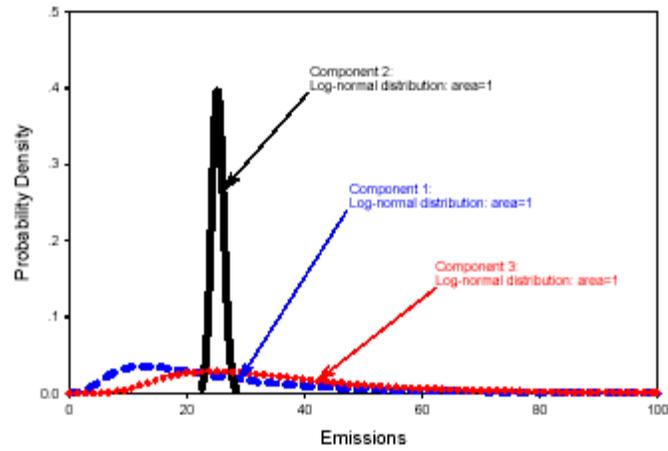


# Draft

Fits of Normal and Log-normal Distributions to Portions of FCCU

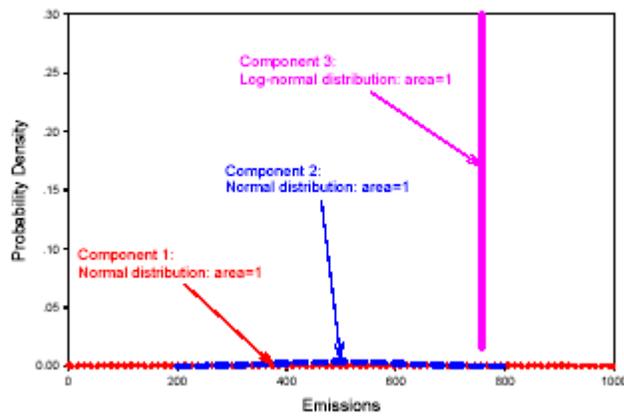


Fits of Normal and Log-normal Distributions to Portions of Low Pressure Flare

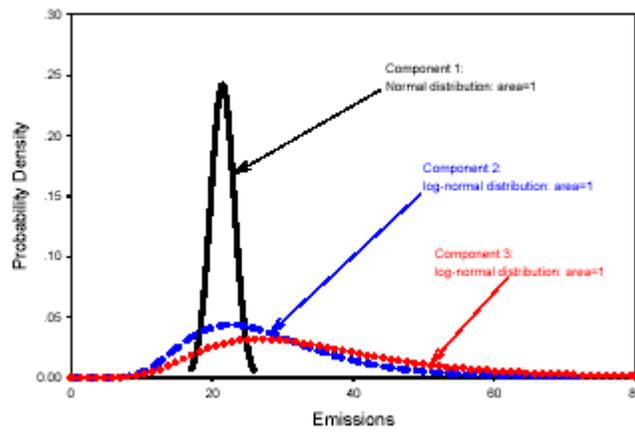


# Draft

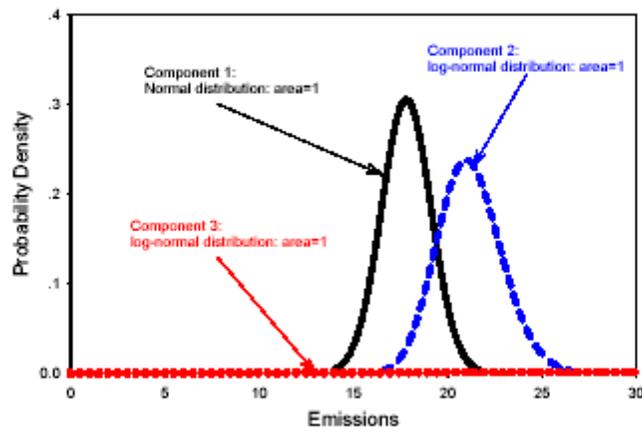
Fits of Normal and Log-normal Distributions to Portions of Merox Flare



Fits of Normal and Log-normal Distributions to Portions of GS #1

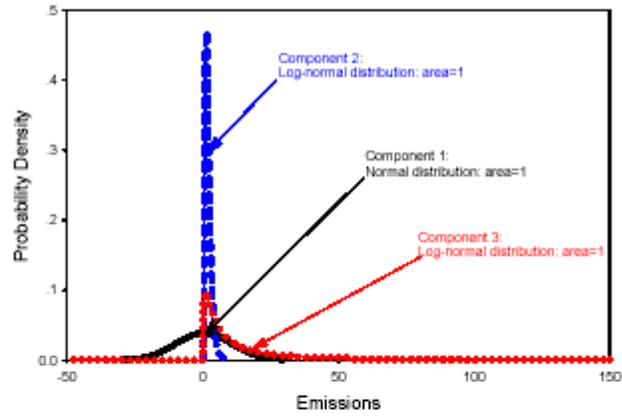


Fits of Normal and Log-normal Distributions to Portions of GS #2

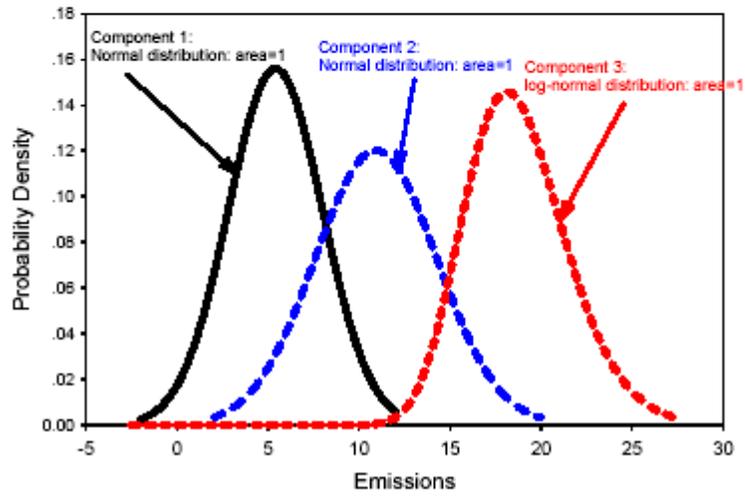


Draft

Fits of Normal and Log-normal Distributions to Portions of FU5854H2\_OP3

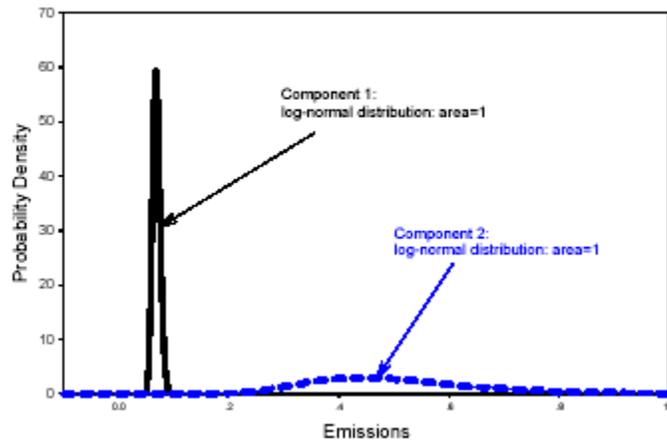


Fits of Normal and Log-normal Distributions to Portions of FU5852\_OP3

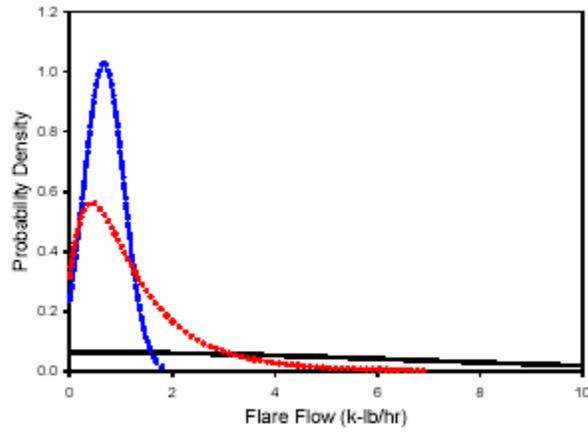


# Draft

Fits of Normal and Log-normal Distributions to Portions of CT 1

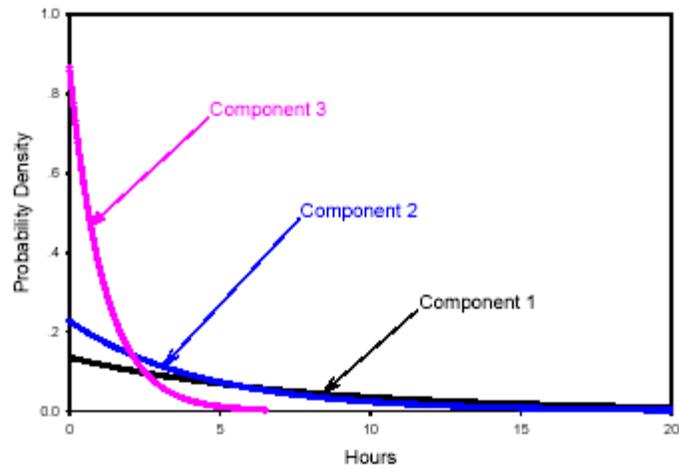


Fits of Log-Normal Distributions to Portions of Cooling Tower 2

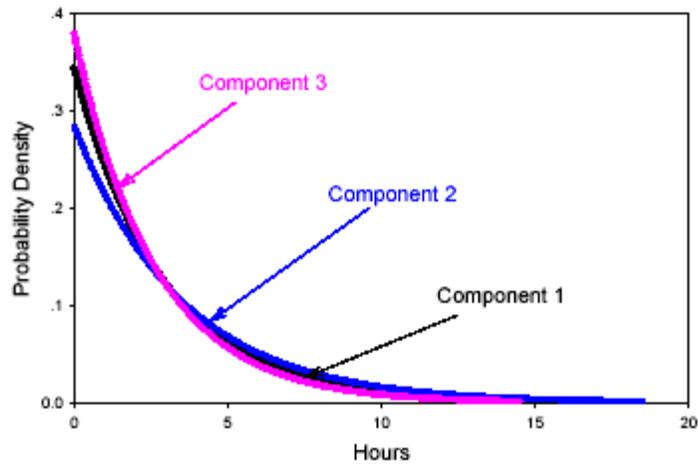


# Draft

## Appendix 3: Graphs of Distributions of Time Within Component PDF of Time within Each Component Flare 1

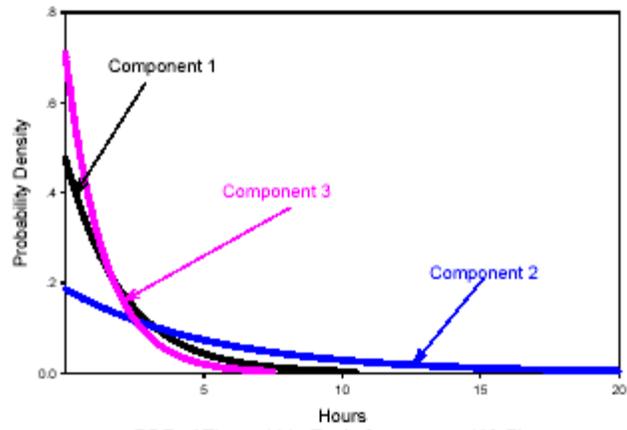


PDF of Time within Each Component Flare 2

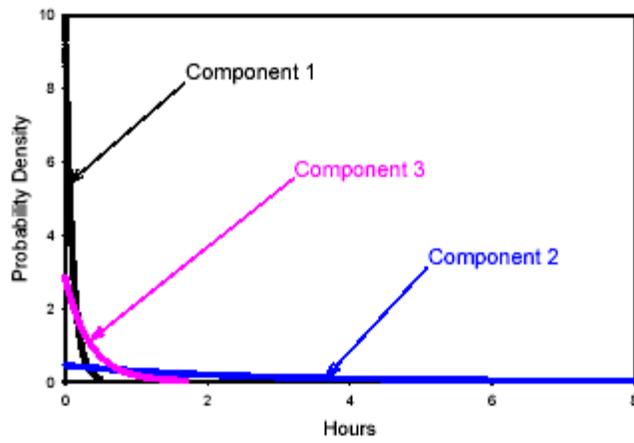


# Draft

PDF of Time within Each Component Flare 5

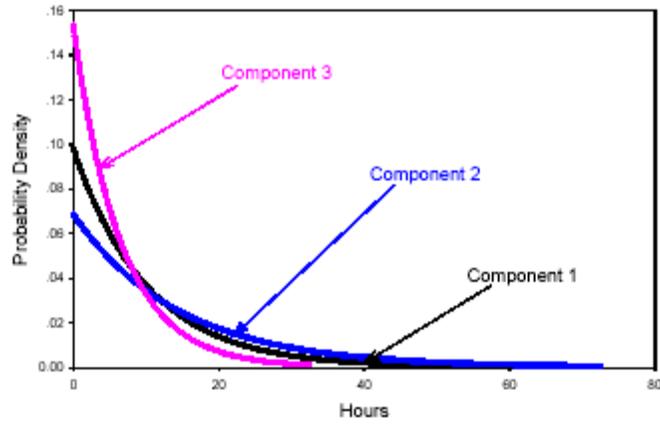


PDF of Time within Each Component HC Flare

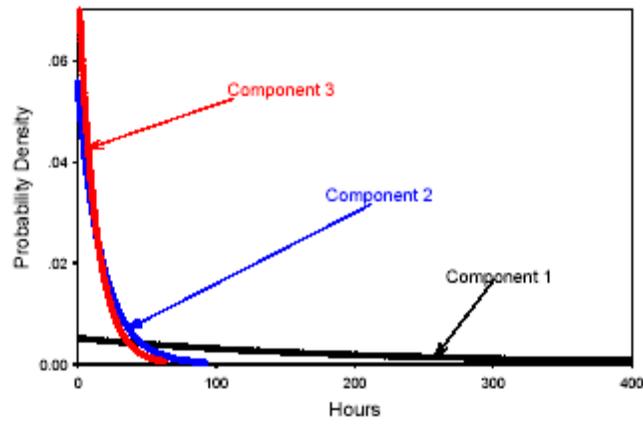


# Draft

PDF of Time within Each Component Olefins Flare

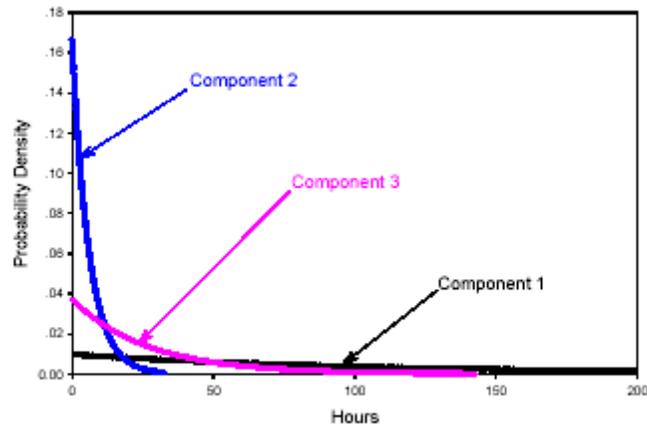


PDF of Time within Each Component FCCU

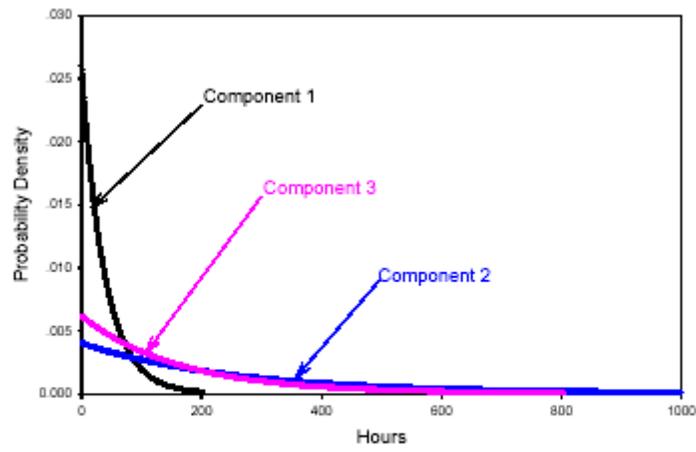


# Draft

PDF of Time within Each Component SRU Flare

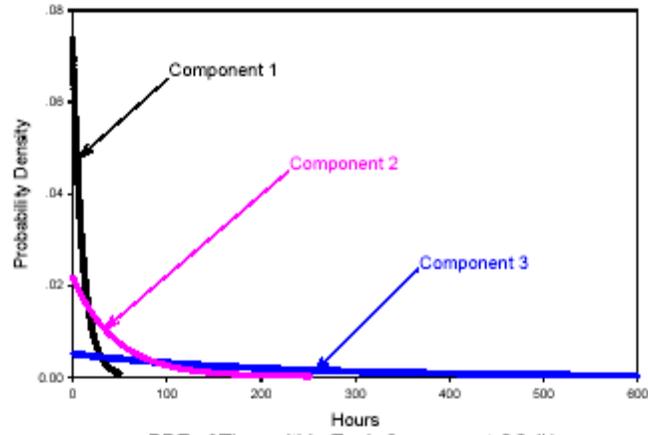


PDF of Time within Each Component Low Pressure Flare

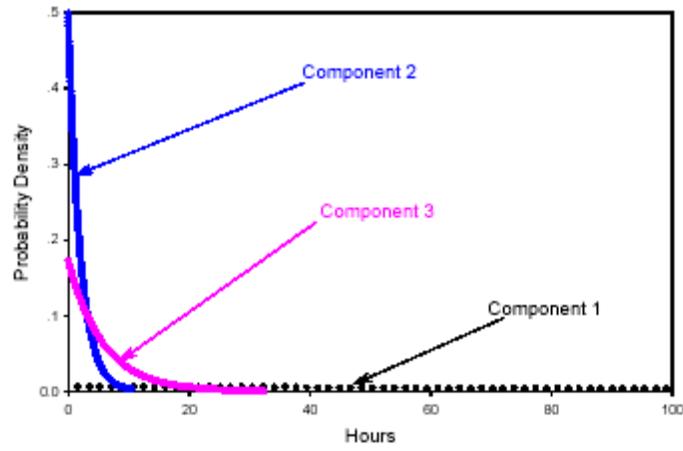


# Draft

PDF of Time within Each Component Merox Flare

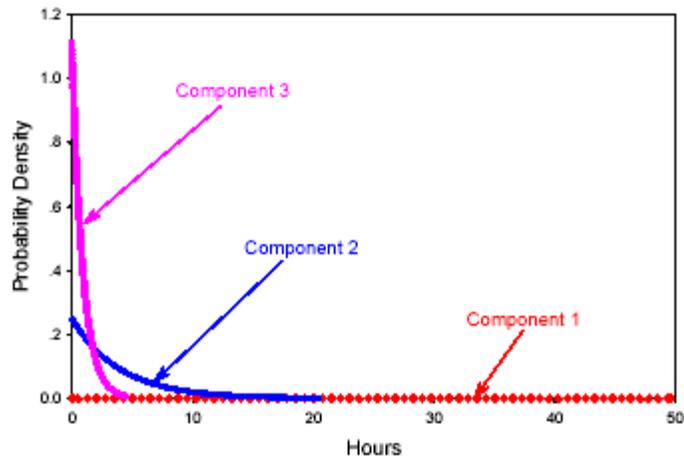


PDF of Time within Each Component GS #1

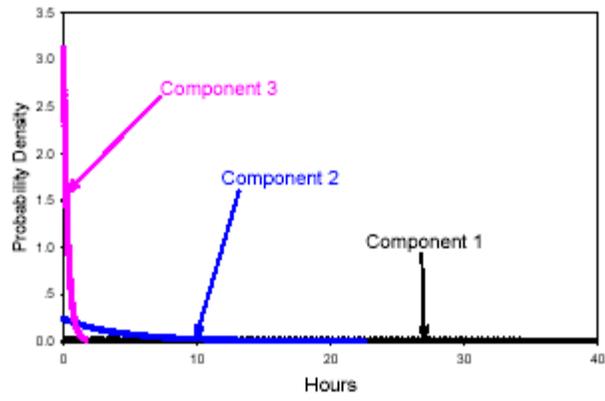


Draft

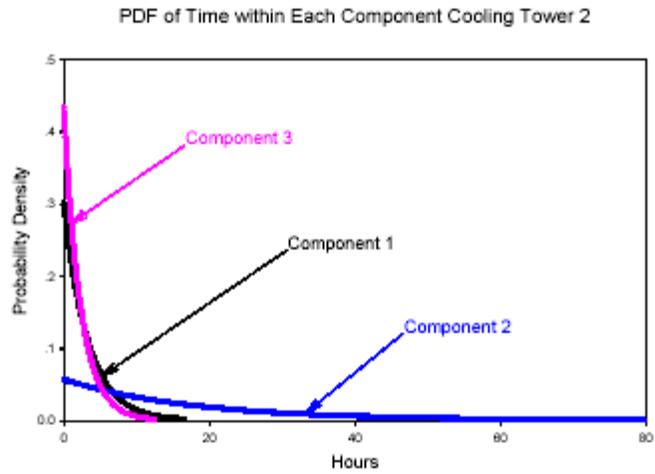
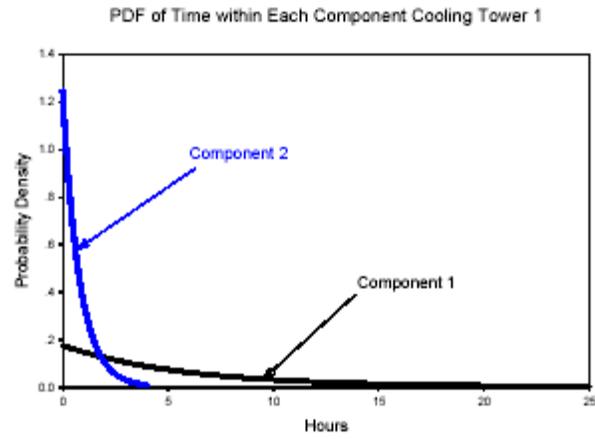
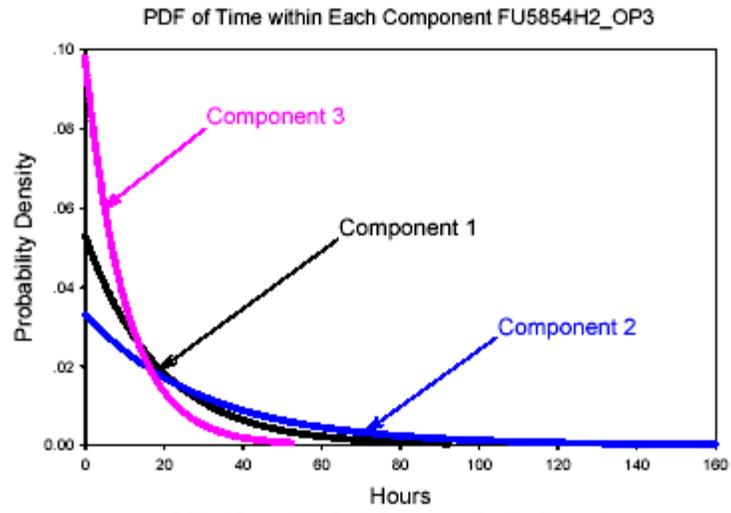
PDF of Time within Each Component GS #2



PDF of Time within Each Component FU5852\_OP3



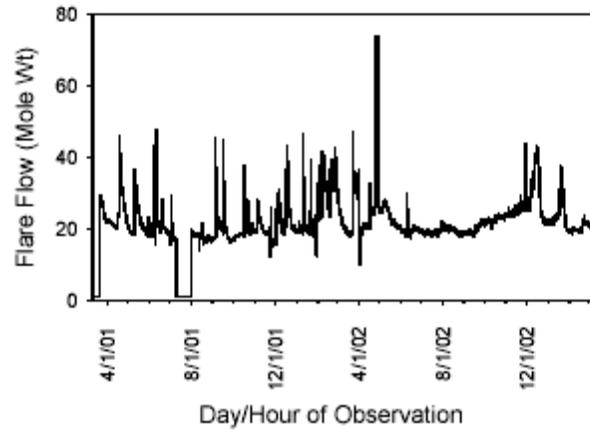
Draft



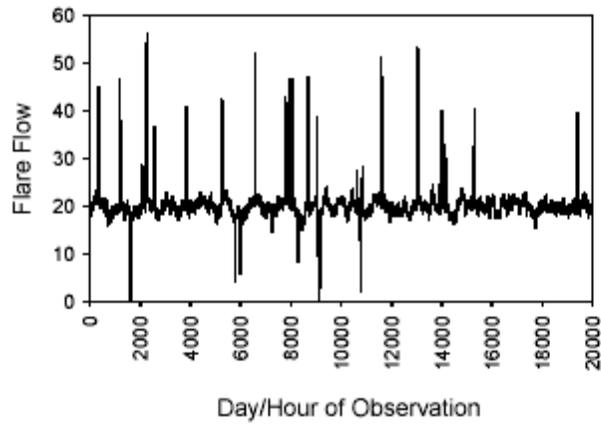
Draft

Appendix 4: Comparisons of Actual vs. Simulated Emissions

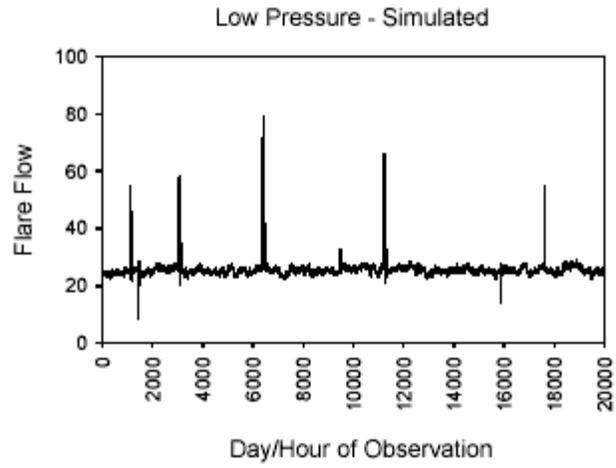
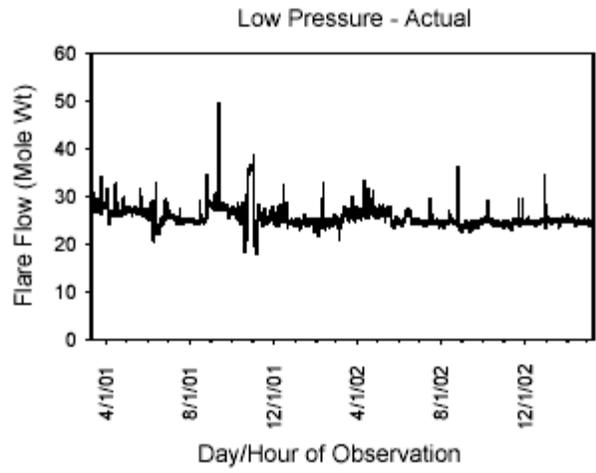
FCCU Actual



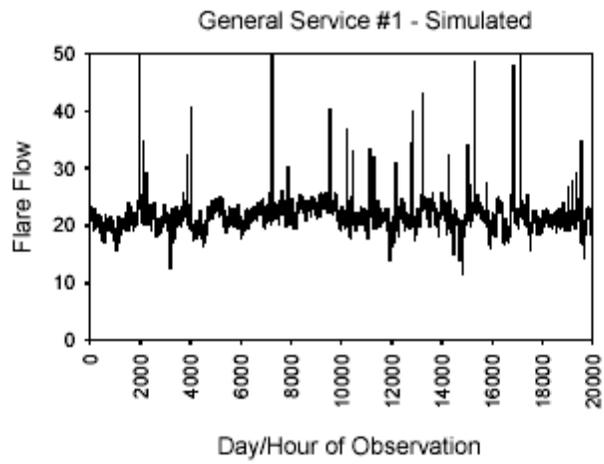
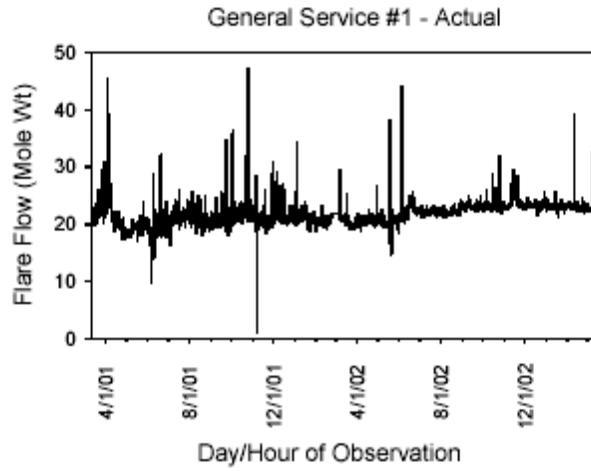
FCCU Simulated



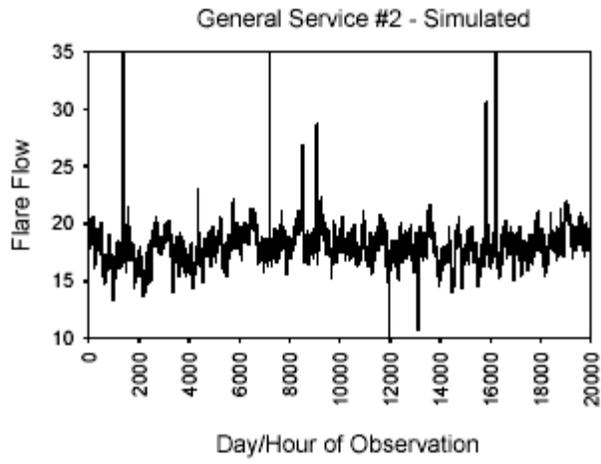
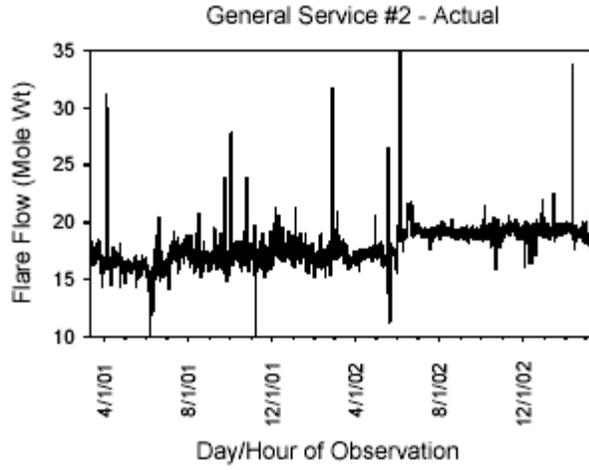
Draft



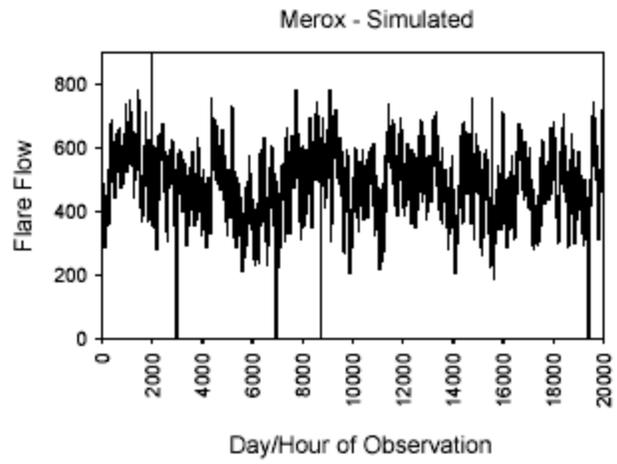
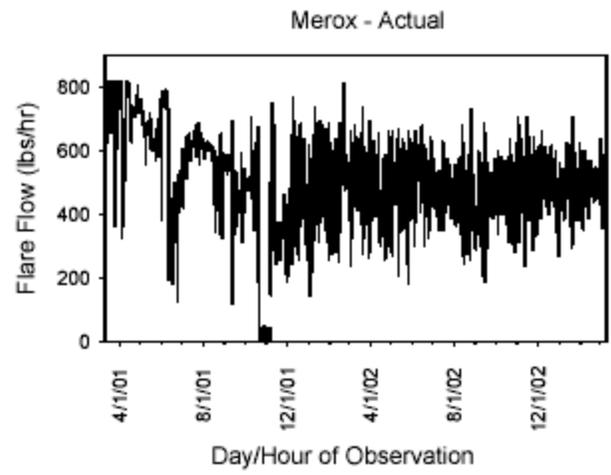
Draft



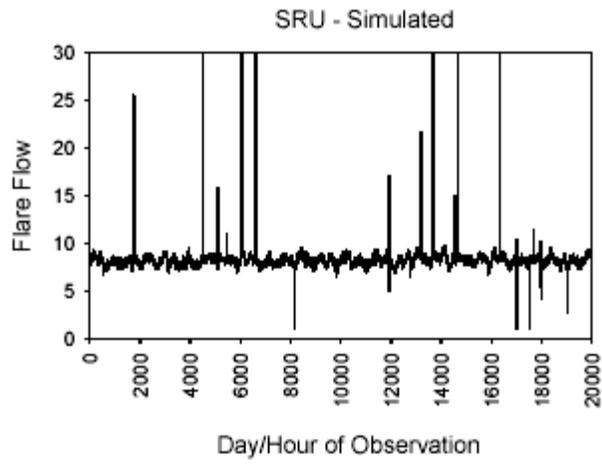
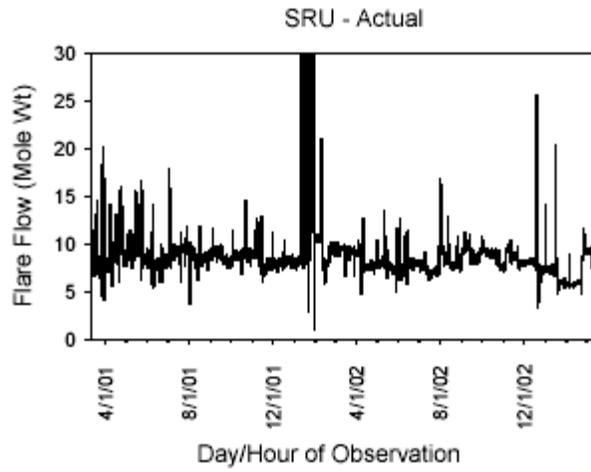
Draft



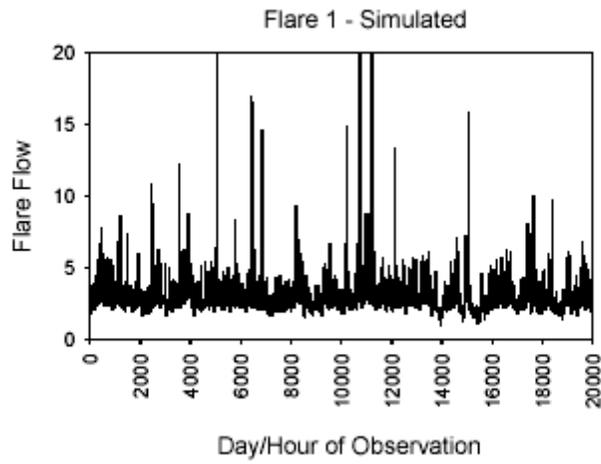
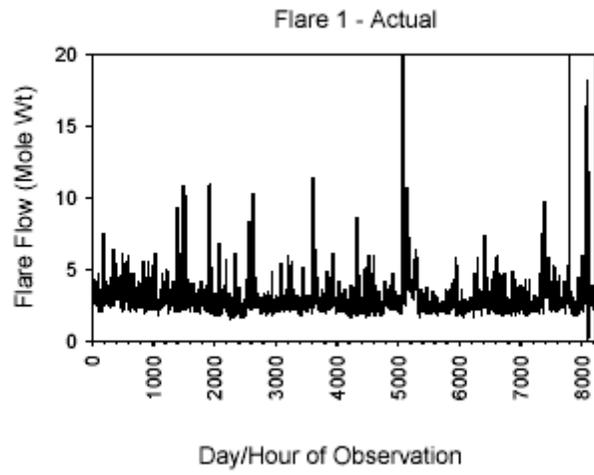
Draft



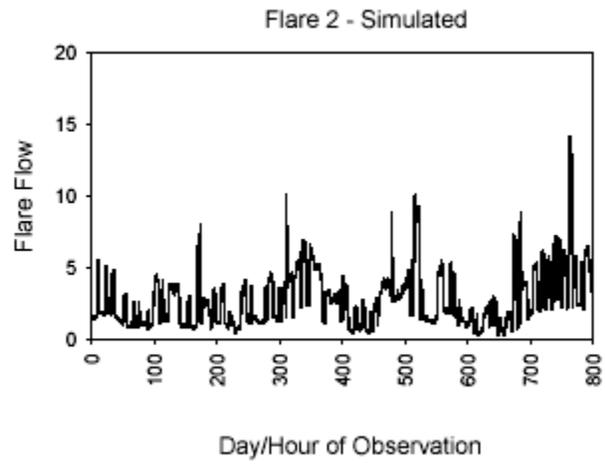
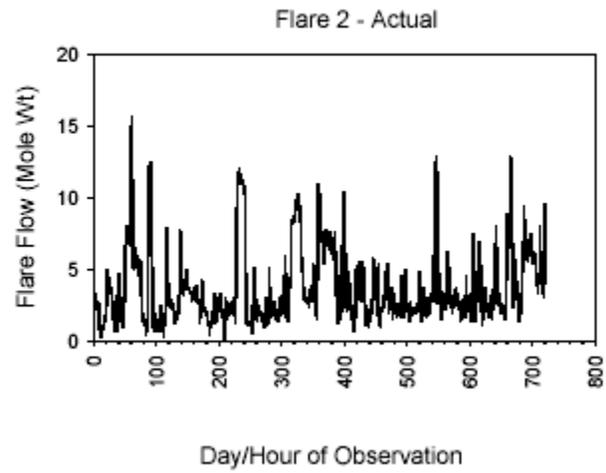
Draft



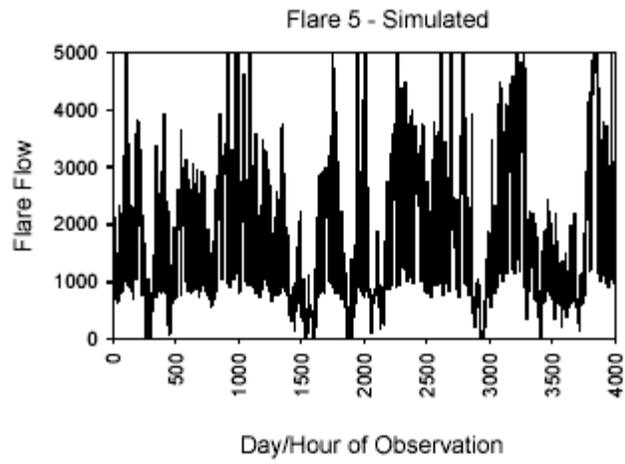
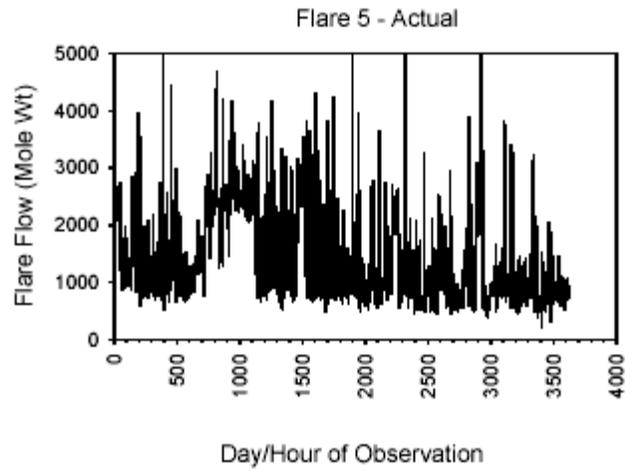
# Draft



# Draft

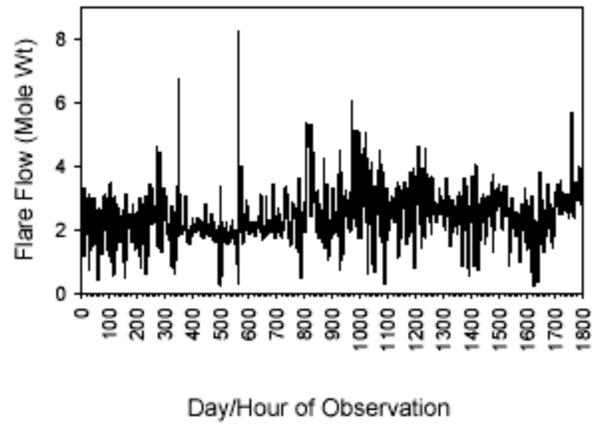


# Draft

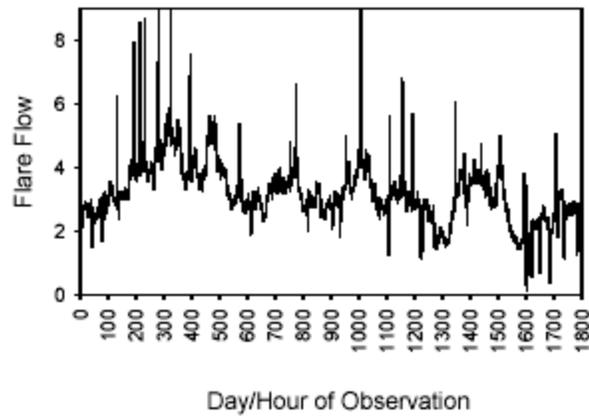


# Draft

## HC Flare - Actual

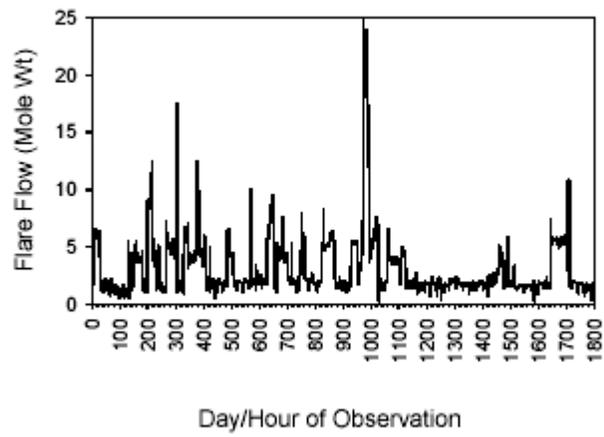


## HC Flare - Simulated

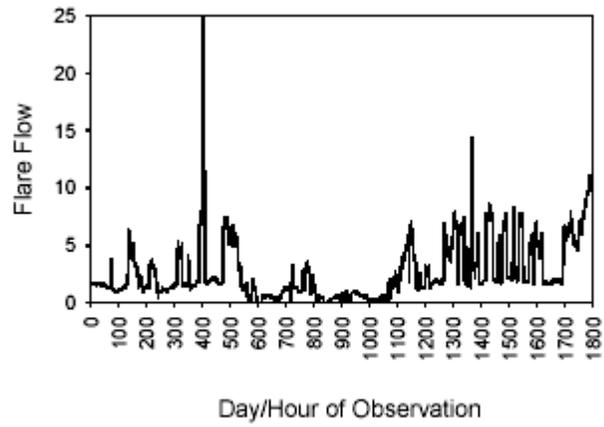


# Draft

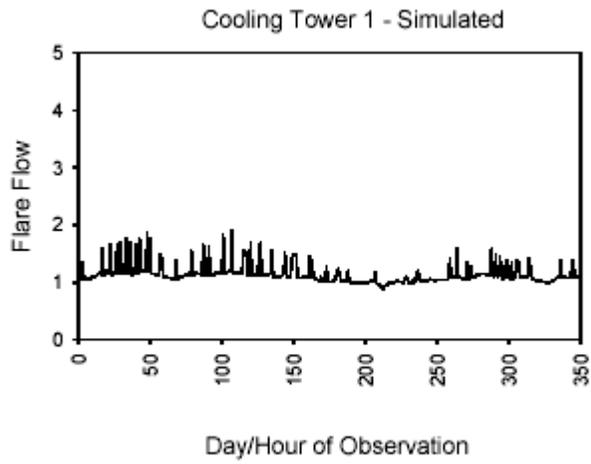
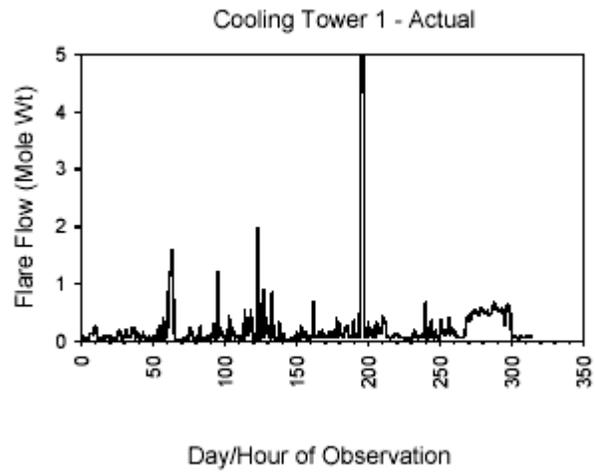
Olefins Flare - Actual



Olefins Flare - Simulated

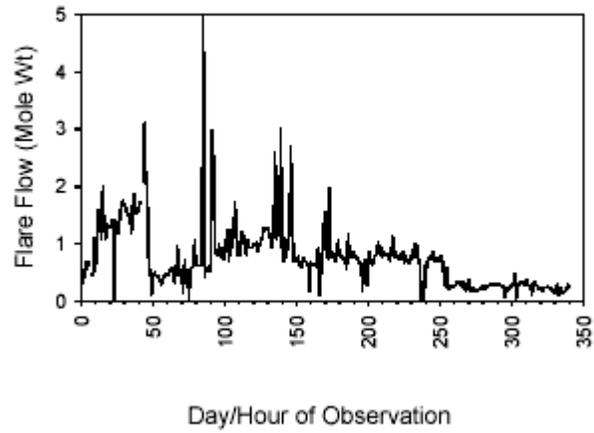


# Draft

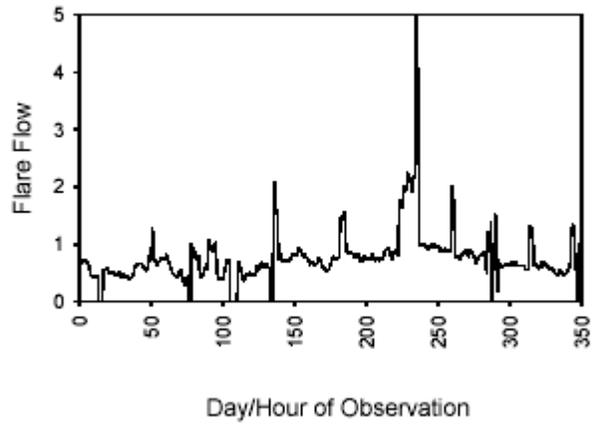


# Draft

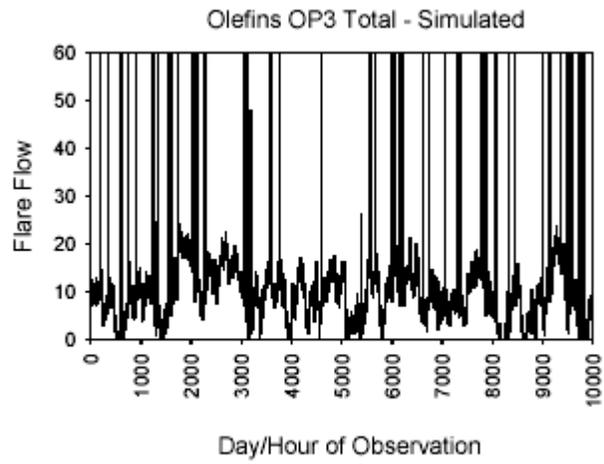
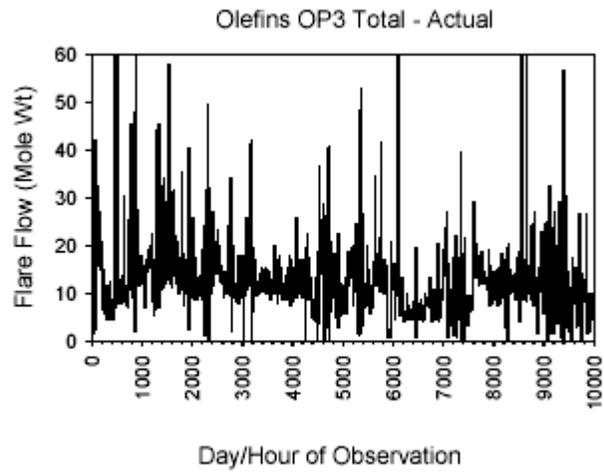
Cooling Tower 2 - Actual



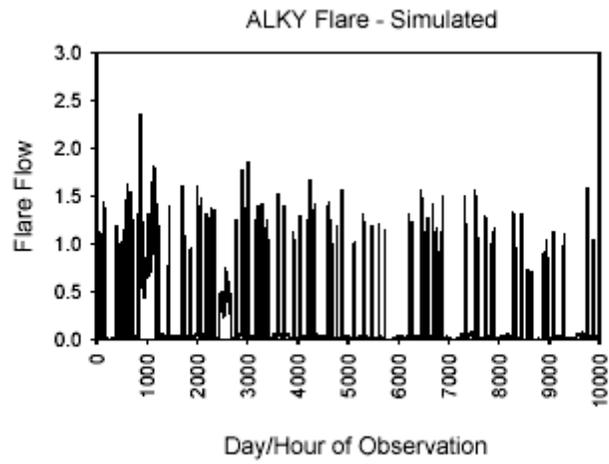
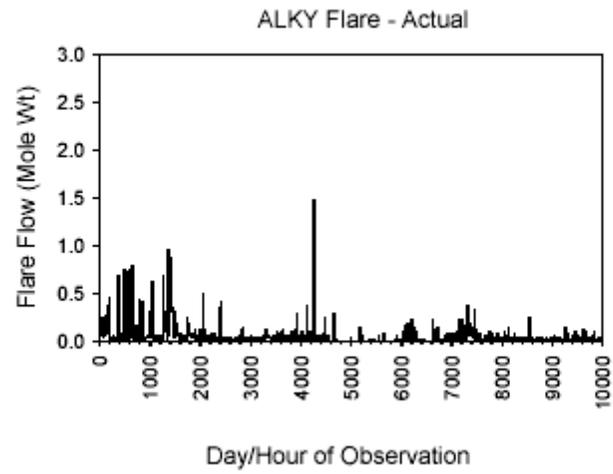
Cooling Tower 2 - Simulated



# Draft



# Draft



# Draft

