Air Quality

Nelson Oliveira

March 2019

1 Research design

1.1 Features

1. Month
2. Day
3. Year
4. (Hour)
5. Temperature in C
6. Relative Humidity (%)
7. AH Absolute Humidity
8. Monday (boolean)
9. Tuesday (boolean)
10. Wednesday (boolean)
11. Thursday (boolean)
12. Friday (boolean)
13. Saturday (boolean)
14. Sunday (boolean)

1.1.1 Algorithms

1. Support Vector Machines (SVM)
2. Random Forest (RF)
3. Linear Regression (LR)
4. Gaussian Processes (GP)
5. Multi Layer Perceptron (MLP)

2 Predictions

In the figure 2 we present the predictions of the used algorithms to predict the concentration of the Benzene gas. We observe that the RF algorithm has a better perception of the variations of the concentration comparing to the other algorithms when comparing to the real values of the gas concentration. We can also see that the in the first and the last few hours of the day the GP predicts closer gas concentrations.

Figure 1: Legend

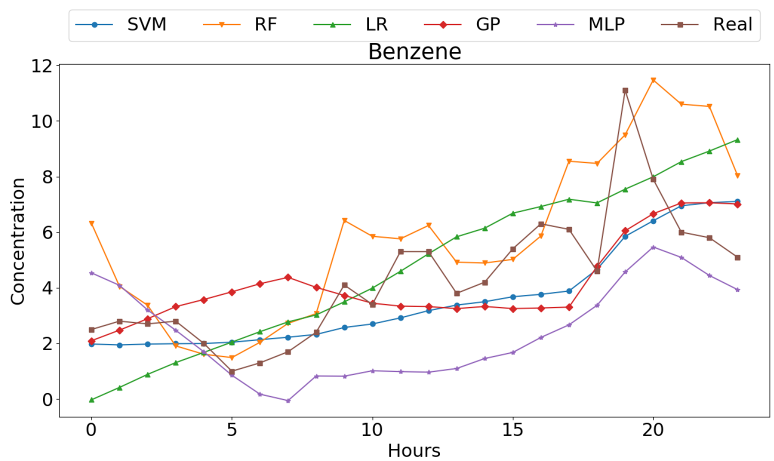


Figure 2: Benzene

In the figure 3 we present the predictions of the used algorithms to predict the concentration of the CO gas. We observe that the RF algorithm has a better perception of the variations of the concentration comparing to the other algorithms when comparing to the real values of the gas concentration. We can also see that the in the first few hours of the day the SVM predicts closer gas concentrations, and the MLP algorithm predicts closer gas concentrations in the last few hours of the day.

In the figure 4 we present the predictions of the used algorithms to predict the concentration of the Indium Oxide (PT08.S5) gas. We observe that the RF algorithm has a better perception of the variations of the concentration comparing to the other algorithms when comparing to the real values of the gas concentration.

In the figure 5 we present the predictions of the used algorithms to predict the concentration of the NO2 gas. We observe that the RF algorithm has a

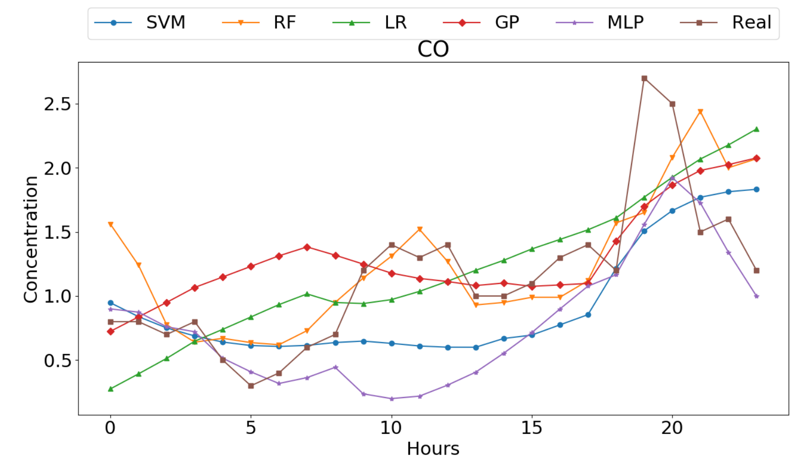


Figure 3: CO

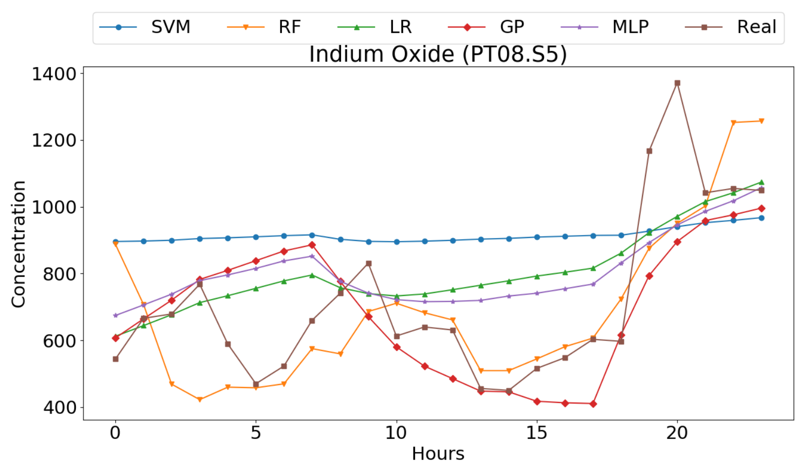


Figure 4: Indium Oxide (PT08.S5)

better perception of the variations of the concentration comparing to the other algorithms when comparing to the real values of the gas concentration.

In the figure 6 we present the predictions of the used algorithms to predict the concentration of the Non-Metanic HydroCarbons gas. We observe that the RF algorithm has a better perception of the variations of the concentration comparing to the other algorithms when comparing to the real values of the gas concentration.

In the figure 7 we present the predictions of the used algorithms to predict the concentration of the NOx gas. We observe that the RF algorithm has a better perception of the variations of the concentration comparing to the other

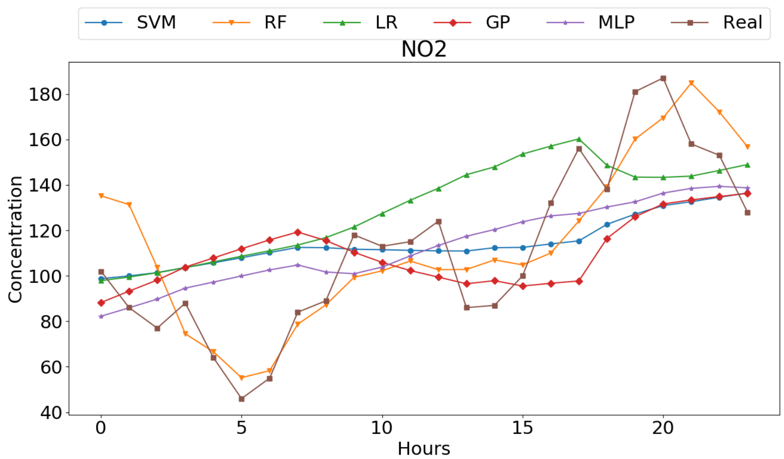


Figure 5: NO2

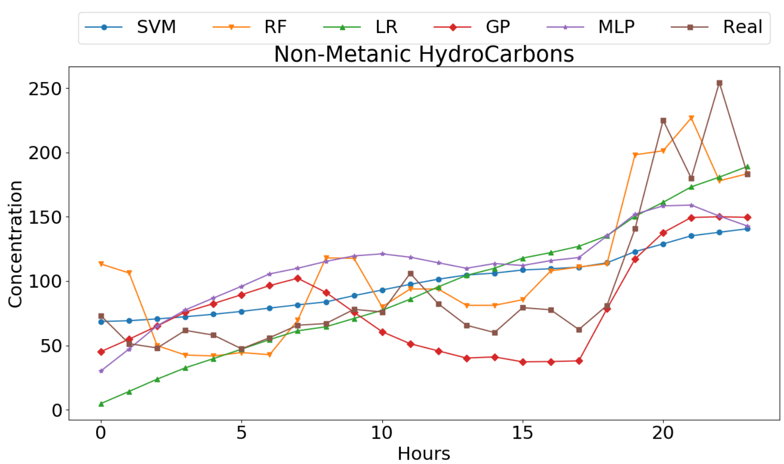


Figure 6: Non-Metanic HydroCarbons

algorithms when comparing to the real values of the gas concentration.

In the figure 8 we present the predictions of the used algorithms to predict the concentration of the Tin Oxide (PT08.S1) gas. We observe that the RF algorithm has a better perception of the variations of the concentration comparing to the other algorithms when comparing to the real values of the gas concentration.

In the figure 9 we present the predictions of the used algorithms to predict the concentration of the Titania (PT08.S2) gas. We observe that the RF algorithm has a better perception of the variations of the concentration comparing to the other algorithms when comparing to the real values of the gas concentration.

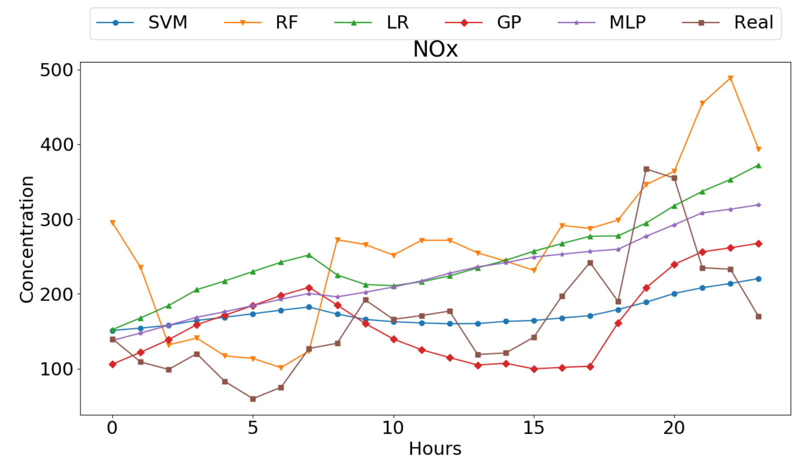


Figure 7: NOx

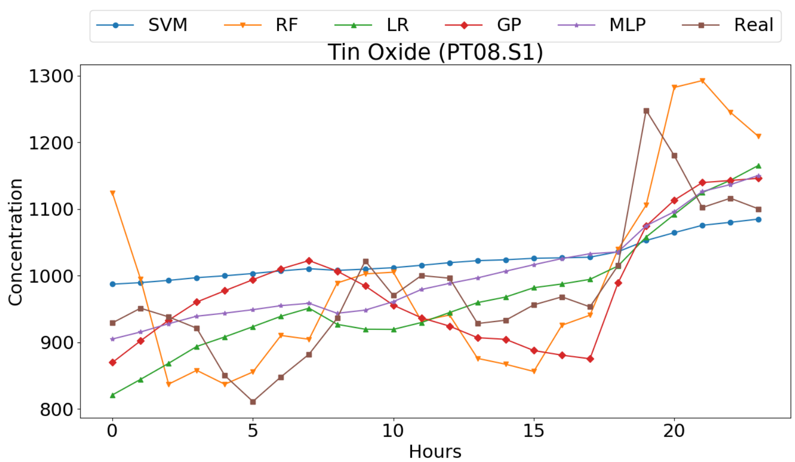


Figure 8: Tin Oxide (PT08.S1)

In the figure 10 we present the predictions of the used algorithms to predict the concentration of the Tungsten Oxide (PT08.S3) gas. We observe that the RF algorithm has a better perception of the variations of the concentration comparing to the other algorithms when comparing to the real values of the gas concentration.

In the figure 11 we present the predictions of the used algorithms to predict the concentration of the Tungsten Oxide (PT08.S4) gas. We observe that the RF algorithm has a better perception of the variations of the concentration comparing to the other algorithms when comparing to the real values of the gas concentration. We can also see that the in the first and the last few hours of

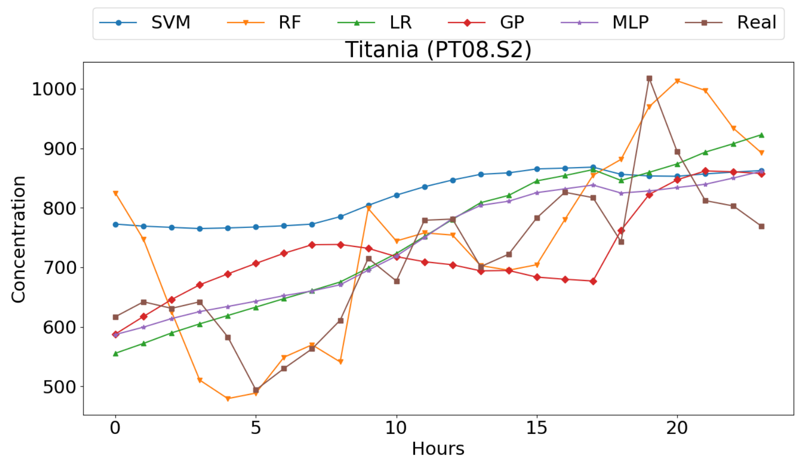


Figure 9: Titania (PT08.S2)

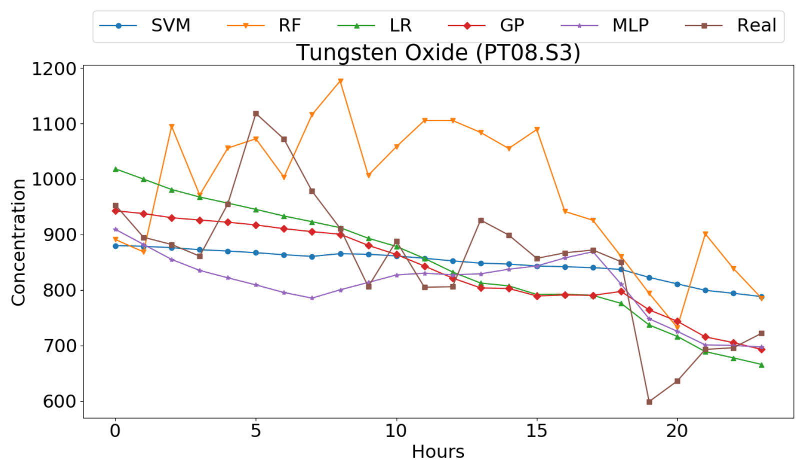


Figure 10: Tungsten Oxide (PT08.S3)

the day the MLP predicts closer gas concentrations.

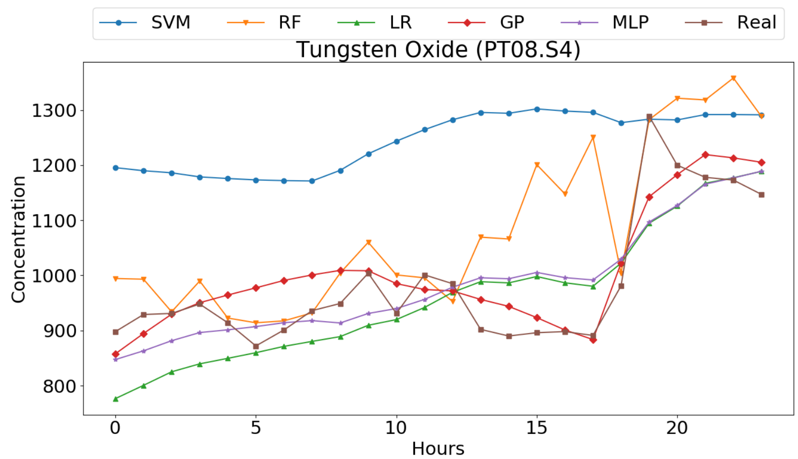
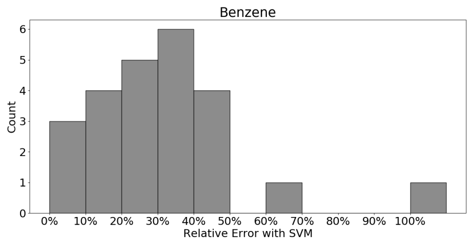


Figure 11: Tungsten Oxide (PT08.S4)

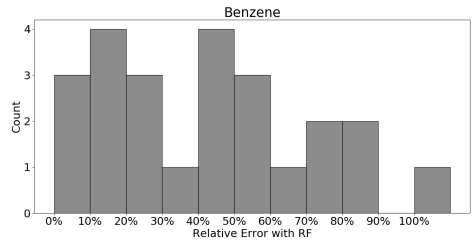
3 Relative Error Distribution

In figure 12 we present the number of hours that have between 0% to 10%, 10% to 20%, 20% to 30%, 30% to 40%, 40% to 50%, 50% to 60%, 60% to 70%, 80% to 90%, 90% to 100% and more than 100% of relative error for each one of the used algorithms, SVM, RF. LR, GP and MLP of the Benzene gas. We observe that the algorithm that has less relative errors over 50% is the SVM, and the one that has the most is the MLP.

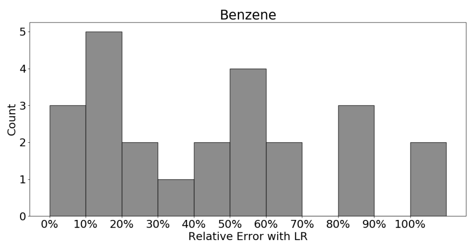
3.0.1 Benzene



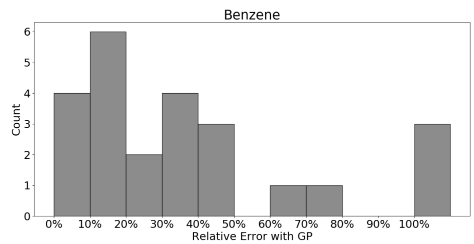
(a) SVM



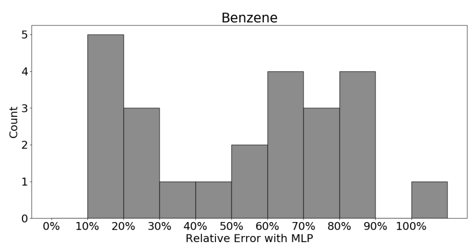
(b) RF



(c) LR



(d) GP

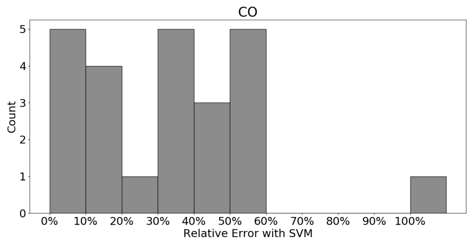


(e) MLP

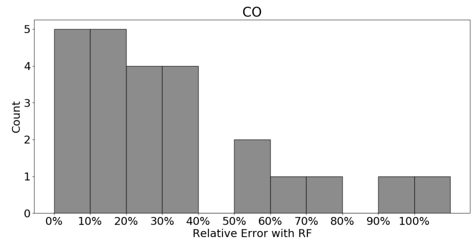
Figure 12: Benzene

In figure 13 we present the number of hours that have between 0% to 10%, 10% to 20%, 20% to 30%, 30% to 40%, 40% to 50%, 50% to 60%, 60% to 70%, 80% to 90%, 90% to 100% and more than 100% of relative error for each one of the used algorithms, SVM, RF. LR, GP and MLP of the CO gas. We observe that the algorithm that has the most relative errors over 50% is the RF but at the same time is the one that has the most hours within a relative error less than 10%.

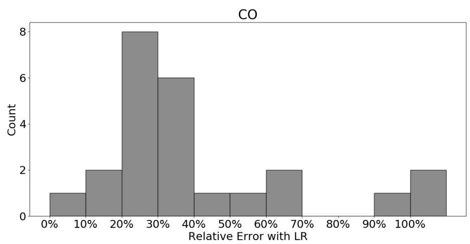
3.0.2 CO



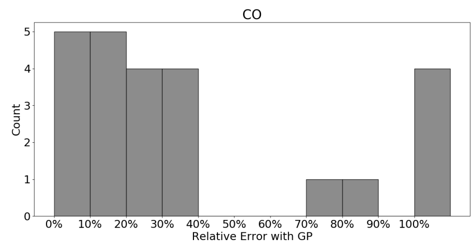
(a) SVM



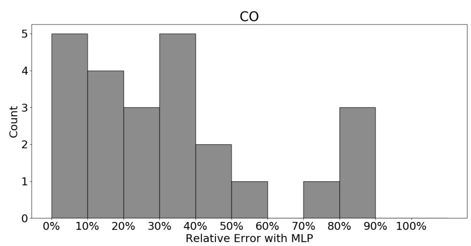
(b) RF



(c) LR



(d) GP

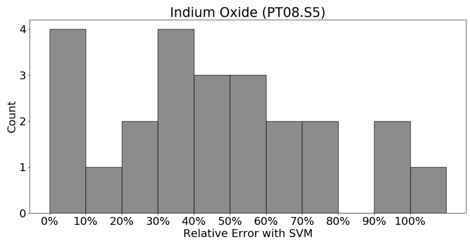


(e) MLP

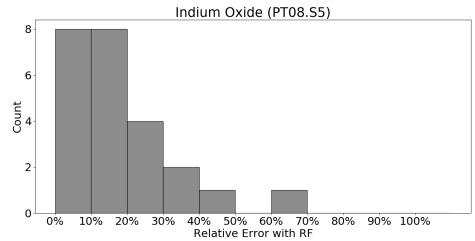
Figure 13: CO

3.0.3 Indium Oxide (PT08.S5)

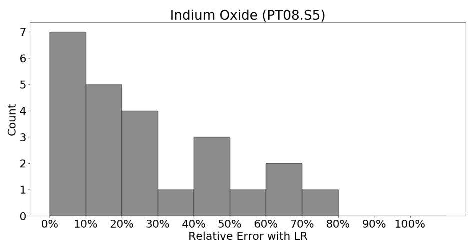
In figure 14 we present the number of hours that have between 0% to 10%, 10% to 20%, 20% to 30%, 30% to 40%, 40% to 50%, 50% to 60%, 60% to 70%, 80% to 90%, 90% to 100% and more than 100% of relative error for each one of the used algorithms, SVM, RF. LR, GP and MLP of the Indium Oxide (PT08.S5) gas. We observe that the algorithms that have less relative errors over 50% are the RF and GP, and the one that has the most is the SVM.



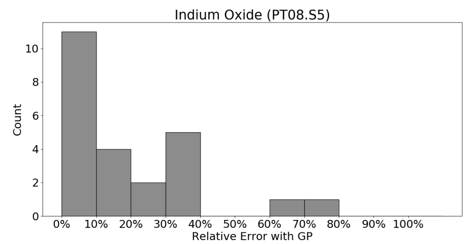
(a) SVM



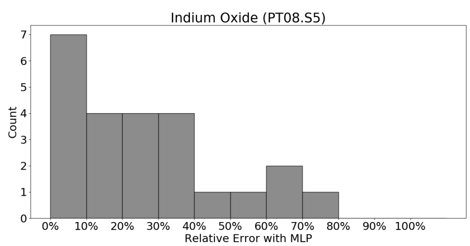
RF]
(b)



(c) LR



(d) GP

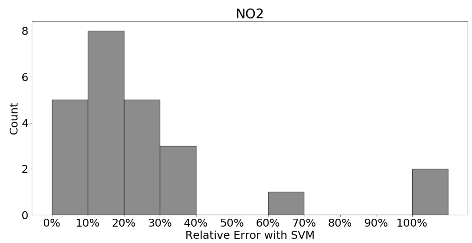


(e) MLP

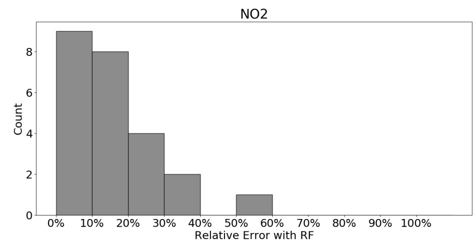
Figure 14: Indium Oxide (PT08.S5)

3.0.4 NO2

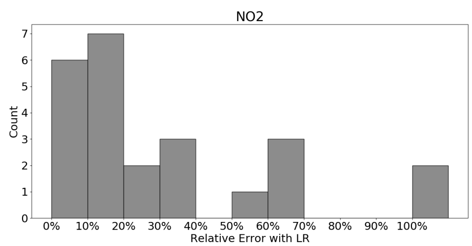
In figure 15 we present the number of hours that have between 0% to 10%, 10% to 20%, 20% to 30%, 30% to 40%, 40% to 50%, 50% to 60%, 60% to 70%, 80% to 90%, 90% to 100% and more than 100% of relative error for each one of the used algorithms, SVM, RF. LR, GP and MLP of the NO2 gas. We observe that the algorithm that has less relative errors over 50% is the RF which in this case has none, and the one that has the most is the LR.



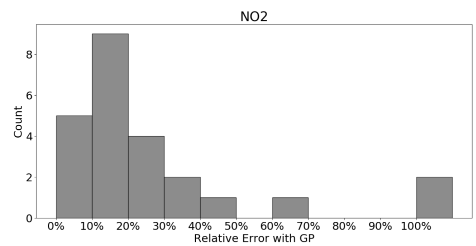
(a) SVM



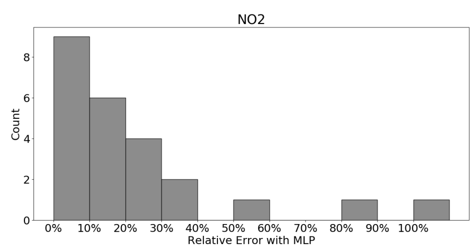
(b) RF



(c) LR



(d) GP

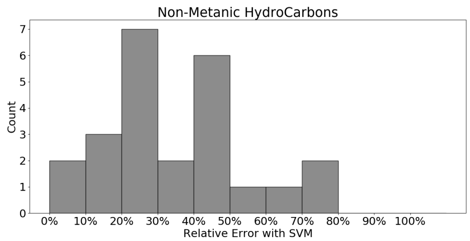


(e) MLP

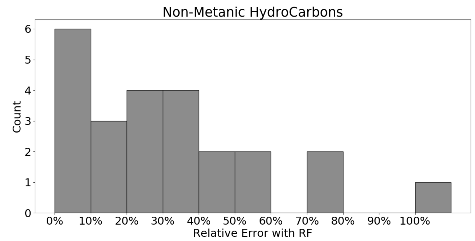
Figure 15: NO2

3.0.5 Non-Metanic HydroCarbons

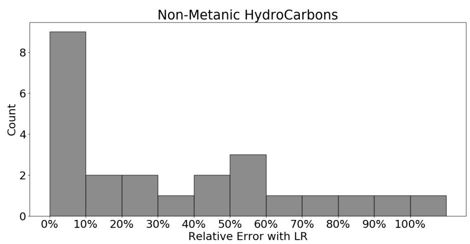
In figure 16 we present the number of hours that have between 0% to 10%, 10% to 20%, 20% to 30%, 30% to 40%, 40% to 50%, 50% to 60%, 60% to 70%, 80% to 90%, 90% to 100% and more than 100% of relative error for each one of the used algorithms, SVM, RF. LR, GP and MLP of the Non-Metanic HydroCarbon gas. We observe that the algorithm that has less relative errors over 50% is the SVM, and the one that has the most is the MLP.



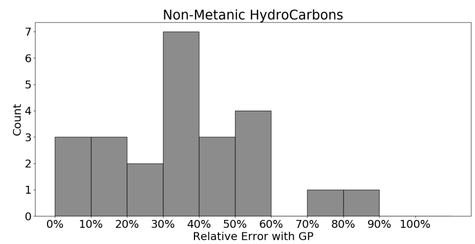
(a) SVM



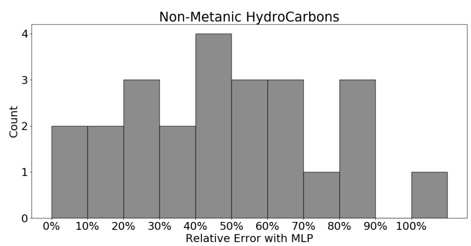
(b) RF



(c) LR



(d) GP

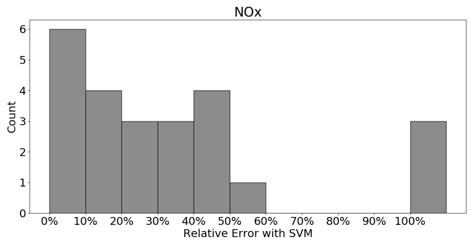


(e) MLP

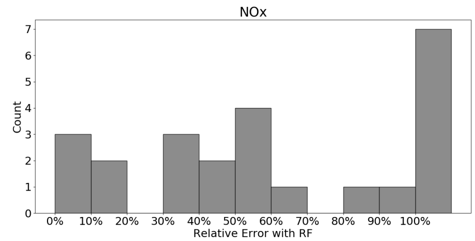
Figure 16: Non-Metanic HydroCarbons

3.0.6 NOx

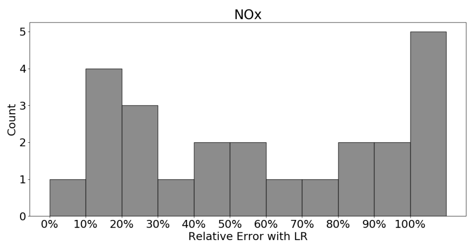
In figure 17 we present the number of hours that have between 0% to 10%, 10% to 20%, 20% to 30%, 30% to 40%, 40% to 50%, 50% to 60%, 60% to 70%, 80% to 90%, 90% to 100% and more than 100% of relative error for each one of the used algorithms, SVM, RF. LR, GP and MLP of the NOx gas. We observe that the algorithm that has less relative errors over 50% is the SVM, and the one that has the most is the LR.



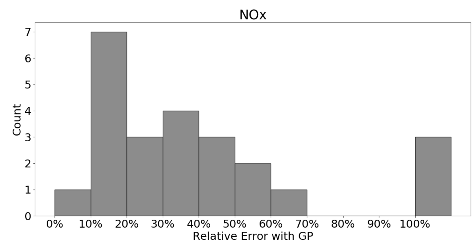
(a) SVM



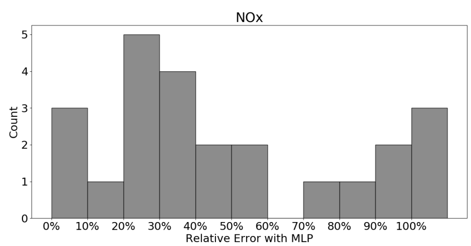
(b) RF



(c) LR



(d) GP

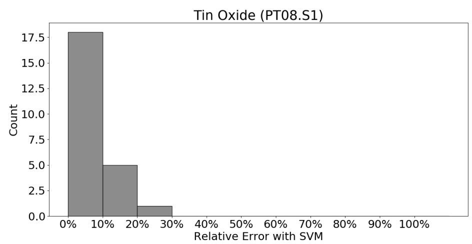


(e) MLP

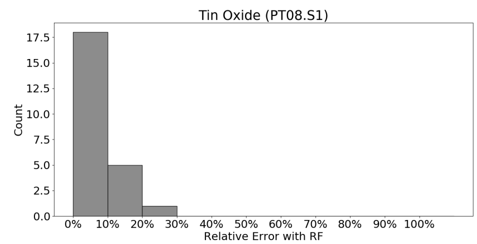
Figure 17: NOx

3.0.7 Tin Oxide (PT08.S1)

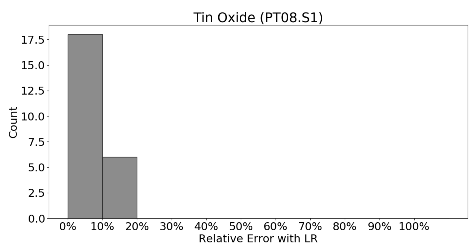
In figure 18 we present the number of hours that have between 0% to 10%, 10% to 20%, 20% to 30%, 30% to 40%, 40% to 50%, 50% to 60%, 60% to 70%, 80% to 90%, 90% to 100% and more than 100% of relative error for each one of the used algorithms, SVM, RF. LR, GP and MLP of the Tin Oxide (PT08.S1) gas. We observe that the algorithm that all the relative errors are between 0 and 30%, in this case the algorithm that has the most number of hours with a relative error less than 10% is the MLP.



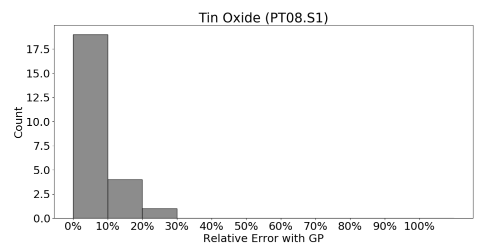
(a) SVM



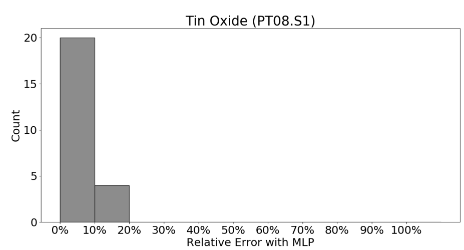
(b) RF



(c) LR



(d) GP

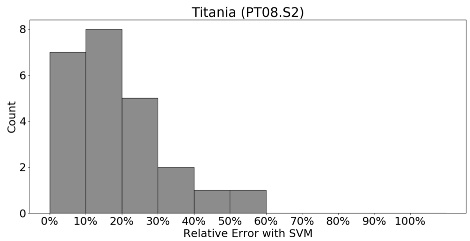


(e) MLP

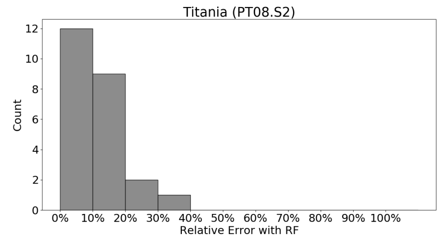
Figure 18: Tin Oxide (PT08.S1)

3.0.8 Titania (PT08.S2)

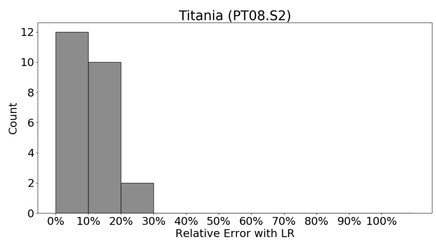
In figure 18 we present the number of hours that have between 0% to 10%, 10% to 20%, 20% to 30%, 30% to 40%, 40% to 50%, 50% to 60%, 60% to 70%, 80% to 90%, 90% to 100% and more than 100% of relative error for each one of the used algorithms, SVM, RF. LR, GP and MLP of the Titania (PT08.S2) gas. We observe that the algorithm that all the relative errors are between 0 and 30%, in this case the algorithm that has the most number of hours with a relative error less than 10% is the MLP.



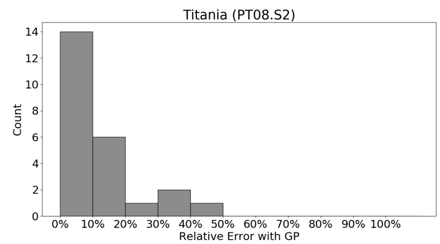
(a) SVM



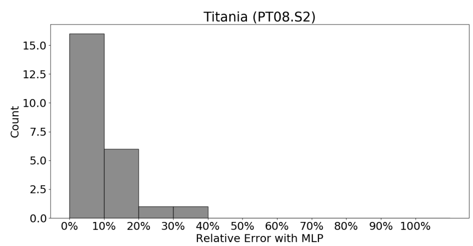
(b) RF



(c) LR



(d) GP

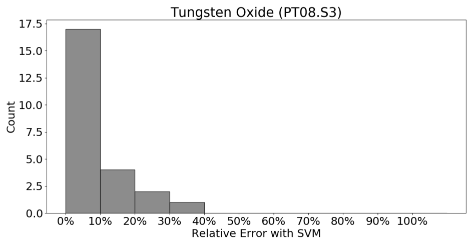


(e) MLP

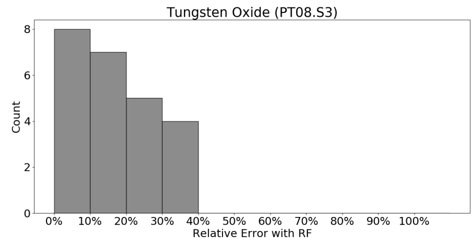
Figure 19: Titania (PT08.S2)

3.0.9 Tungsten Oxide (PT08.S3)

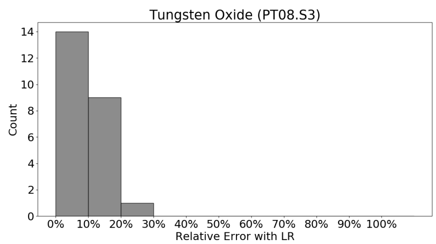
In figure 20 we present the number of hours that have between 0% to 10%, 10% to 20%, 20% to 30%, 30% to 40%, 40% to 50%, 50% to 60%, 60% to 70%, 80% to 90%, 90% to 100% and more than 100% of relative error for each one of the used algorithms, SVM, RF. LR, GP and MLP of the Tungsten Oxide (PT08.S3) gas. We observe that the algorithm that all the relative errors are between 0 and 30%, in this case the algorithm that has the most number of hours with a relative error less than 10% is the GP.



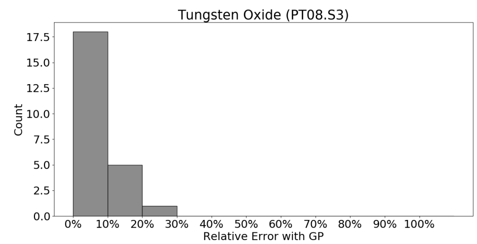
(a) SVM



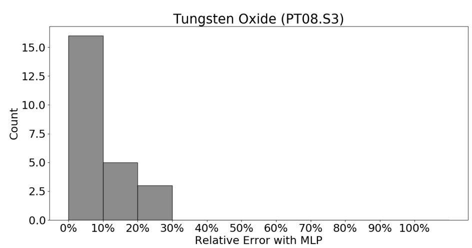
(b) RF



(c) LR



(d) GP

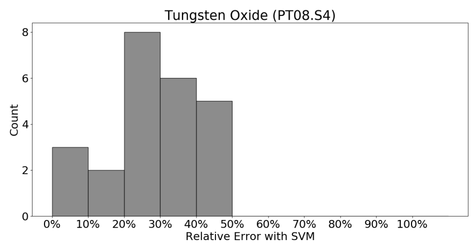


(e) MLP

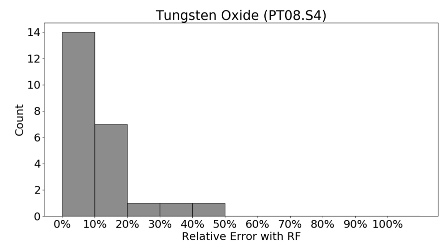
Figure 20: Tungsten Oxide (PT08.S3)

3.0.10 Tungsten Oxide (PT08.S4)

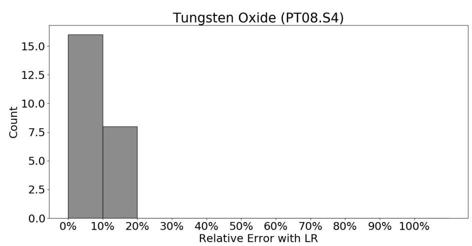
In figure 21 we present the number of hours that have between 0% to 10%, 10% to 20%, 20% to 30%, 30% to 40%, 40% to 50%, 50% to 60%, 60% to 70%, 80% to 90%, 90% to 100% and more than 100% of relative error for each one of the used algorithms, SVM, RF. LR, GP and MLP of the Tungsten Oxide (PT08.S4) gas. We observe that the algorithm that all the relative errors are between 0 and 30%, in this case the algorithm that has the most number of hours with a relative error less than 10% is the GP.



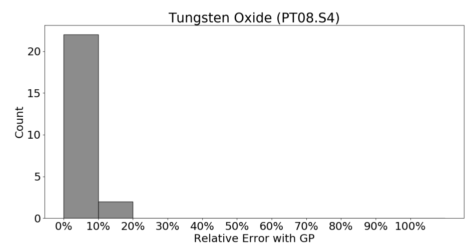
(a) SVM



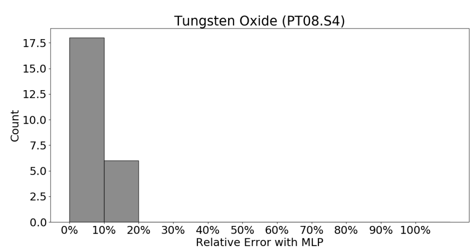
(b) RF



(c) LR



(d) GP



(e) MLP

Figure 21: Tungsten Oxide (PT08.S4)

4 Mean Relative Error

In Table 4 we present the mean relative error for each algorithm for each gas of the training data. We can observe that the RF algorithm performed better that the others for all of the gases.

Algorithm / Gas	SVM	RF	LR	GP	MLP
Benzene	58.97% ± 94.35%	**13.41% ± 28.64%**	94.04% ± 205.52%	77.59% ± 148.78%	64.96% ± 121.47%
CO	48.03% ± 94.52%	**13.19% ± 34.12%**	75.77% ± 142.74%	71.01% ± 133.3%	45.88% ± 86.43%
Indium Oxide (PT08.S5)	34.12% ± 34.86%	**6.25% ± 8.06%**	33.86% ± 34.07%	29.79% ± 29.57%	31.32% ± 31.04%
NO2	33.51% ± 80.07%	**7.04% ± 24.8%**	35.89% ± 79.33%	33.51% ± 80.55%	31.26% ± 72.57%
Non-Metanic HydroCarbons	62.74% ± 66.16%	**15.88% ± 22.21%**	94.31% ± 113.85%	78.59% ± 96.43%	83.29% ± 98.34%
NOx	75.91% ± 140.82%	**11.73% ± 18.91%**	87.43% ± 138.77%	72.89% ± 125.3%	72.49% ± 111.43%
Tin Oxide (PT08.S1)	13.54% ± 9.67%	**2.62% ± 2.73%**	13.81% ± 9.7%	12.93% ± 9.46%	12.9% ± 9.38%
Titania (PT08.S2)	20.33% ± 16.62%	**3.81% ± 16.62%**	20.24% ± 16.52%	18.71% ± 15.35%	15.35% ± 15.49%
Tungsten Oxide (PT08.S3)	21.16% ± 18.28%	**3.96% ± 4.44%**	20.09% ± 17.49%	19.12% ± 17.16%	18.31% ± 16.12%
Tungsten Oxide (PT08.S4)	14.8% ± 13.49%	**2.35% ± 2.5%**	11.26% ± 8.34%	10.38% ± 7.73%	10.97% ± 8.31%

Table 1: Mean Relative error of each algorithm for each gas for the training data

Algorithm / Gas	SVM	RF	LR	GP	MLP
Benzene	**30.55% ± 21.51%**	45.08% ± 32.52%	44.42% ± 30.84%	52.33% ± 68.88%	52.51% ± 27.73%
CO	33.48% ± 23.23%	**32.91% ± 28.31%**	44.52% ± 39.11%	53.95% ± 74.7%	33.85% ± 25.57%
Indium Oxide (PT08.S5)	45.56% ± 28.13%	17.44% ± 14.22%	25.93% ± 22.15%	**19.81% ± 19.88%**	26.24% ± 20.94%
NO2	27.74% ± 30.72%	**15.91% ± 11.53%**	32.38% ± 33.28%	29.49% ± 32.94%	24.38% ± 26.67%
Non-Metanic HydroCarbons	34.77% ± 19.32%	38.81% ± 26.33%	35.18% ± 31.49%	36.13% ± 20.38%	49.5% ± 27.13%
NOx	**39.86% ± 43.86%**	62.41% ± 39.38%	73.62% ± 66.92%	46.31% ± 47.75%	56.07% ± 48.87%
Tin Oxide (PT08.S1)	7.98% ± 5.99%	7.23% ± 4.74%	6.26% ± 4.05%	7.38% ± 5.69%	**5.75% ± 4.44%**
Titania (PT08.S2)	19.02% ± 12.87%	10.42% ± 8.37%	10.46% ± 6.79%	12.35% ± 11.08%	**8.83% ± 7.33%**
Tungsten Oxide (PT08.S3)	9.55% ± 9.0%	17.23% ± 10.75%	8.4% ± 5.35%	**8.0% ± 6.27%**	8.09% ± 8.44%
Tungsten Oxide (PT08.S4)	27.62% ± 12.44%	10.28% ± 10.71%	7.27% ± 4.41%	**4.37% ± 3.31%**	5.7% ± 4.13%

Table 2: Mean Relative error of each algorithm for each gas for the testing data

Algorithm / Gas	RF	RF Tunned
Benzene	45.08% ± 32.52%	**35.83% ± 30.39%**
CO	32.91% ± 28.31%	**30.3% ± 24.7%**
Indium Oxide (PT08.S5)	17.44% ± 14.22%	**16.56% ± 13.11%**
NO2	15.91% ± 11.53%	**18.09% ± 13.63%**
Non-Metanic HydroCarbons	31.24% ± 26.33%	**35.75% ± 26.33%**
NOx	62.41% ± 39.38%	**29.91% ± 25.15%**
Tin Oxide (PT08.S1)	7.23% ± 4.74%	**4.1% ± 3.43%**
Titania (PT08.S2)	10.42% ± 8.37%	**9.09% ± 6.19%**
Tungsten Oxide (PT08.S3)	17.23% ± 10.75%	**13.44% ± 8.87%**
Tungsten Oxide (PT08.S4)	10.28% ± 10.71%	**9.51% ± 4.54%**

Table 3: Mean Relative error of each algorithm with the best algorithm of the training data Tunned