International Journal of Scientific & Engineering Research, Volume 7, Issue 3, March-2016
ISSN 2229-5518

An Introductory Approach to Virtual Sensors and Its Modelling Techniques

Durvesh Pathak [1], Vishal P Halale [2]

[1] BE- Electronics Engineering, University of Mumbai   Electronics Engineer: Cummins EBU- Controls
2 Mtech- Mechatronics, VIT University, Vellore, TN   Electronics Engineer: Cummins EBU-Controls

Abstract:- In Control System Industry with the advent of electronics, new control strategies and regulatory requirements have led to the use of numerous kind of physical sensors for advanced diagnostics requirement. The critical inputs from the physical sensors controls the performance of the entire control system. Though these physical sensors are of great importance, they are expensive and failure of these components can lead to devastating effects. So this approach covers the virtual modelling of these physical sensors to avoid the inventible failures.

Index Terms— Virtual Sensors; Empirical Modelling; Control Systems; Multi variate Regression analysis, MATLAB Simulation, Modelling.

1. Introduction

Almost all of the Control System today uses control module for controlling the performance of the system and also for important real time decision making, the performance / decision made is based on some inputs from various sensors which are present in the Control system. As we know electronic control systems comprises of mesh of sensors and its subsystems; these sensors along with the subsystems play a critical role on the performance of the control system. System performance is all the way controlled through output values from these sensors, and these values are critical to the safety and stability of the system. Installing such sensors in the system adds the price of the system which is also one of the important factor in today's competitive market. Hence development of a virtual sensor via data driven modelling and physics based modelling will play an important role in the coming great future of System engineering.

2. Fundamentals of Virtual Sensor Design and Its Modelling Techniques

As the term suggests virtual sensor is not a physical sensor but behaves like one. These breed of sensors can be used, to act as backup sensors or as main sensors where placing a physical sensor is not possible due to parochial nature (e.g. Mechanical restriction, cost restriction, etc.). Virtual sensor modelling can be either Empirical based, where historical data can be used to derive a correlation between the response and the predictors or it can be analytical based where physics and formulas are used to derive the value of the sensors based on certain actual sensor inputs. Selection of these modelling technique depends upon the design of the sensor, its environment and mathematical calculation approach.

These two different types of modelling techniques have their own pros and cons. Analytical or physics based modelling is based on the physics of the system and sometimes fails to incorporate any noise that might be introduced in the system, which does not provide an accurate model. Whereas Empirical based modelling is dependent on the historical data that we have, regarding the system. This data containing the noise factor can be used to find the correlation between response and predictors using various techniques like Linear regression, Multi variate Regression, Weighted least square regression, etc.

These virtual sensors as discussed before can be used for replacing an actual physical sensor through modelling it which estimates correct output values or it can be used as a backup for the actual physical sensor. The proposed methodology is based on data driven modelling and it can be fine-tuned further to get accurate output values. The most important part for such model development is the accurate inputs i.e. getting a robust data set which incorporates the entire operating range of the system and differentiating between good and bad data. For implementing this one has to have knowledge about the system dynamics and its performance.

Now let us understand how these systems are represented / modelled,

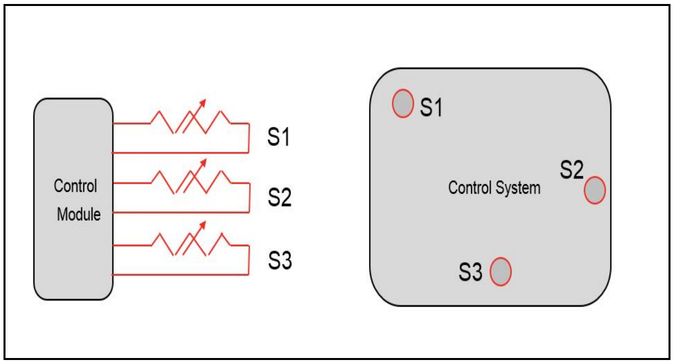


Fig 1 – Sensors in the control system

In figure 1, S1, S2 & S3 are the three sensors placed at different positions in the control system. Sensor 1 &2 are less in cost as compared to Sensor 3. Now Sensor 3 can be modelled virtually by using the data of sensors 1 &2 to reduce the cost of the system and also to make the system compact. E.g. output of Sensor 3 can be expressed as a function of outputs of Sensors 1 & 2.

$$S3=f(S1, S2) \tag{1}$$

3. Modelling technique / Design of Experiment

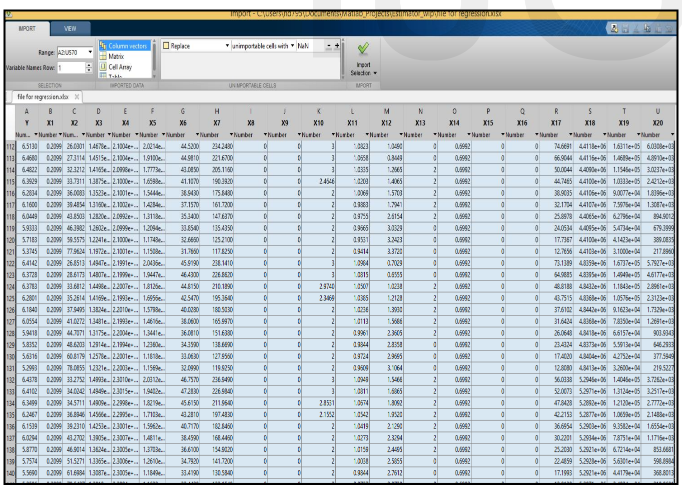
Our idea involves modelling a virtual sensor using empirical modelling technique. Once we have the historical data for entire operating range and all the influencing parameter (Predictors), we can model the sensor virtually. This type of modelling techniques are also called data driven modelling. The modelling and data analysis done further in this paper is done using MATLAB. It follows the sequence of data collection, its analysis, sorting, regression / curve fitting, finding the coefficients, mapping them in the system and see their results.

A. Data collection

It covers the data collection for the entire operating range of the “predictors” that might directly or indirectly influence the “response” and keeping in mind the variation that occurs due to mechanical parts, due to various setups etc. so that we can get a robust data set. We need to collect the data in regards to “response” (Dependent variable) which is our reference value and a set of “predictors” (Independent Values). The response have to be correlated with the predictors and an equation based on the data collected has to be formulated. While doing all this we must be sure that we are filtering out the “bad data” as it may give inevitable results. Then we regress the data set of “predictors” w.r.t “actual response” (Reference Value).

Some of the predictors are actual data from physical sensors placed on the systems and others are based on derivative data which have been derived from the actual data from the actual sensor. (E.g. see in the figure 2 the output X3 = X2 * X1), these relationship and assumptions is based on the control system architecture. This type of modelling gives good results when there is more number of data points.

$$Y= f (X1, X2, X3, \dots , Xn)$$



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20
	Num...	Number	Num...	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number
112	6.5130	0.2099	26.0301	1.4678e...	2.1004e+...	2.0214e...	44.5200	234.2480	0	0	3	1.0823	1.0490	0	0.6992	0	0	74.6691	4.4118e+06	1.6311e+05	6.0308e+03
113	6.4680	0.2099	27.3114	1.4515e...	2.1004e+...	1.9100e...	44.9810	221.6700	0	0	3	1.0658	0.8449	0	0.6992	0	0	66.9044	4.4116e+06	1.4689e+05	4.8910e+03
114	6.4822	0.2099	32.3212	1.4165e...	2.0998e+...	1.7773e...	43.0850	205.1160	0	0	3	1.0335	1.2665	2	0.6992	0	0	50.0044	4.4090e+06	1.1546e+05	3.0237e+03
115	6.3929	0.2099	33.7311	1.3875e...	2.1000e+...	1.6598e...	41.1070	190.3920	0	0	2.4646	1.0203	1.4065	2	0.6992	0	0	44.7465	4.4100e+06	1.0333e+05	2.4212e+03
116	6.2834	0.2099	36.0083	1.3523e...	2.1001e+...	1.5444e...	38.9430	175.8480	0	0	2	1.0069	1.5703	2	0.6992	0	0	38.9035	4.4106e+06	9.0077e+04	1.8396e+03
117	6.1600	0.2099	39.4854	1.3160e...	2.1002e+...	1.4284e...	37.1570	161.7200	0	0	2	0.9883	1.7941	2	0.6992	0	0	32.1704	4.4107e+06	7.5976e+04	1.3087e+03
118	6.0449	0.2099	43.8503	1.2820e...	2.0992e+...	1.3118e...	35.3400	147.6370	0	0	2	0.9755	2.6154	2	0.6992	0	0	25.8978	4.4065e+06	6.2796e+04	894.9012
119	5.9333	0.2099	46.3982	1.2602e...	2.0999e+...	1.2094e...	33.8540	135.4350	0	0	2	0.9665	3.0329	0	0.6992	0	0	24.0534	4.4095e+06	5.4734e+04	679.3999
120	5.7183	0.2099	59.5575	1.2241e...	2.1000e+...	1.1748e...	32.6660	125.2100	0	0	2	0.9531	3.2423	0	0.6992	0	0	17.7367	4.4100e+06	4.1423e+04	389.0835
121	5.3745	0.2099	77.9624	1.1972e...	2.1001e+...	1.1508e...	31.7660	117.8250	0	0	2	0.9414	3.3720	0	0.6992	0	0	12.7656	4.4103e+06	3.1000e+04	217.8960
122	6.4142	0.2099	26.8513	1.4947e...	2.1991e+...	2.0436e...	45.9190	238.1410	0	0	3	1.0984	0.7029	0	0.6992	0	0	73.1389	4.8359e+06	1.6737e+05	5.7927e+03
123	6.3728	0.2099	28.6173	1.4807e...	2.1999e+...	1.9447e...	46.4300	226.8620	0	0	3	1.0815	0.6555	0	0.6992	0	0	64.9885	4.8395e+06	1.4949e+05	4.6177e+03
124	6.3783	0.2099	33.6812	1.4498e...	2.2007e+...	1.8126e...	44.8150	210.1890	0	0	2.9740	1.0507	1.0238	2	0.6992	0	0	48.8188	4.8432e+06	1.1843e+05	2.8961e+03
125	6.2801	0.2099	35.2614	1.4169e...	2.1993e+...	1.6956e...	42.5470	195.3640	0	0	2.3469	1.0385	1.2128	2	0.6992	0	0	43.7515	4.8368e+06	1.0576e+05	2.3123e+03
126	6.1840	0.2099	37.9495	1.3824e...	2.2010e+...	1.5798e...	40.0280	180.5030	0	0	2	1.0236	1.3930	2	0.6992	0	0	37.6102	4.8442e+06	9.1623e+04	1.7329e+03
127	6.0554	0.2099	41.0272	1.3481e...	2.1993e+...	1.4616e...	38.0600	165.9970	0	0	2	1.0113	1.5686	2	0.6992	0	0	31.6424	4.8368e+06	7.8350e+04	1.2691e+03
128	5.9418	0.2099	44.7071	1.3175e...	2.2004e+...	1.3441e...	36.0810	151.6380	0	0	2	0.9961	2.3605	2	0.6992	0	0	26.0648	4.8418e+06	6.6157e+04	903.9343
129	5.8352	0.2099	48.6203	1.2914e...	2.1994e+...	1.2360e...	34.3590	138.6690	0	0	2	0.9844	2.8358	0	0.6992	0	0	23.4324	4.8373e+06	5.5913e+04	646.2933
130	5.6316	0.2099	60.8179	1.2578e...	2.2001e+...	1.1818e...	33.0630	127.9560	0	0	2	0.9724	2.9695	0	0.6992	0	0	17.4020	4.8404e+06	4.2752e+04	377.5949
131	5.2993	0.2099	78.0855	1.2321e...	2.2003e+...	1.1569e...	32.0990	119.9250	0	0	2	0.9609	3.1064	0	0.6992	0	0	12.8080	4.8413e+06	3.2600e+04	219.5227
132	6.4378	0.2099	33.2752	1.4993e...	2.3010e+...	2.0312e...	46.7570	236.9490	0	0	3	1.0949	1.5466	2	0.6992	0	0	56.0338	5.2946e+06	1.4046e+05	3.7262e+03
133	6.4102	0.2099	34.0242	1.4949e...	2.3015e+...	1.9402e...	47.2830	226.9840	0	0	3	1.0811	1.6865	2	0.6992	0	0	52.0073	5.2971e+06	1.3124e+05	3.2517e+03
134	6.3499	0.2099	34.5711	1.4909e...	2.2998e+...	1.8219e...	45.6150	211.9640	0	0	2.8531	1.0674	1.8092	2	0.6992	0	0	47.8428	5.2892e+06	1.2120e+05	2.7772e+03
135	6.2467	0.2099	36.8946	1.4566e...	2.2995e+...	1.7103e...	43.2810	197.4830	0	0	2.1552	1.0542	1.9520	2	0.6992	0	0	42.2153	5.2877e+06	1.0659e+05	2.1488e+03
136	6.1539	0.2099	39.2310	1.4253e...	2.3001e+...	1.5962e...	40.7170	182.8460	0	0	2	1.0419	2.1290	2	0.6992	0	0	36.6954	5.2903e+06	9.3582e+04	1.6554e+03
137	6.0294	0.2099	43.2702	1.3905e...	2.3007e+...	1.4811e...	38.4590	168.4460	0	0	2	1.0273	2.3294	2	0.6992	0	0	30.2201	5.2934e+06	7.8751e+04	1.1716e+03
138	5.8770	0.2099	46.9014	1.3624e...	2.3005e+...	1.3703e...	36.6100	154.9020	0	0	2	1.0159	2.4495	2	0.6992	0	0	25.2030	5.2921e+06	6.7214e+04	853.6681
139	5.7574	0.2099	51.5271	1.3365e...	2.3006e+...	1.2610e...	34.7920	141.7200	0	0	2	1.0038	2.5855	0	0.6992	0	0	22.4859	5.2928e+06	5.6301e+04	598.8984
140	5.5690	0.2099	61.6984	1.3087e...	2.3005e+...	1.1849e...	33.4190	130.5840	0	0	2	0.9844	2.7612	0	0.6992	0	0	17.1993	5.2921e+06	4.4179e+04	368.8013

Fig 2 –Entire data set including Y (response) and X1, X2….Xn (Predictors)

B. Data analysis

Data once collected for the entire operating range, should undergo sanity checks. In this we observed trends of ‘Response’ w.r.t ‘Predictors’. Then we filtered the data by removing the outliers or garbage value if any. The more we filter the data, better was the fit. Filtering of data is based on the knowledge about the system. Ideally all the outlier data points should be removed. The behavior of the “Response” can be studied by plotting them and filtering them manually with knowledge. We did the same thing. Below graphs shows the response data plotted and how we filtered it in reference to predictors. After couple of iterations and filtering the data sets we got the final fine and filtered data set.

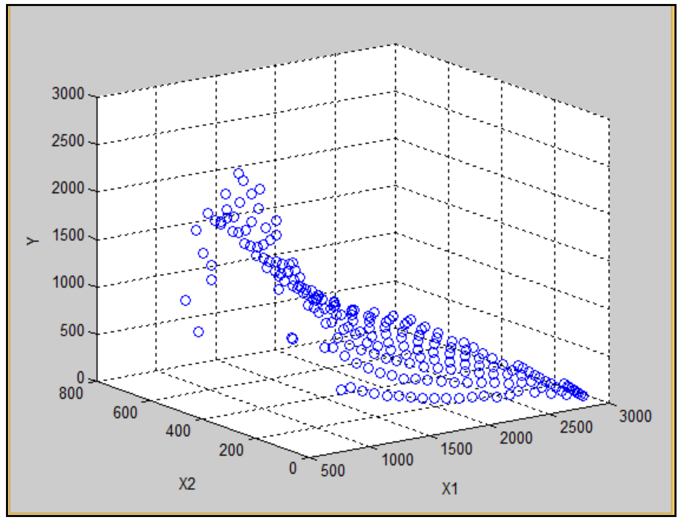


Fig 3 – 3D Plot for Y (Response) w.r.t X (Predictors)

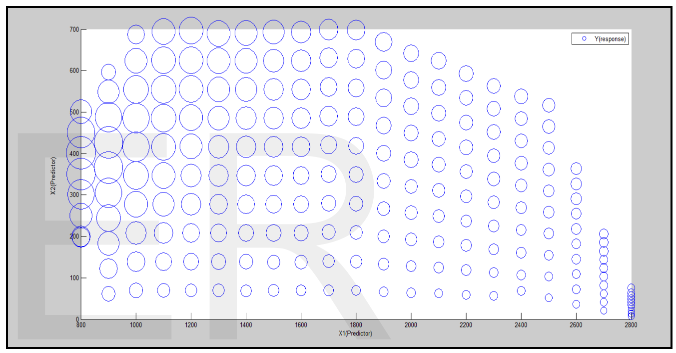


Fig 4 – 2D Plot for Y (Response) w.r.t X (Predictors)

These data plots helped us to filter “Bad Data’’. Now as per our control system considered, the trend of the data should always be increasing with increase in X2. So we filtered the data in such a way that where ever the trend is not being followed, we removed those data points and then considered them for regression. Figure 5 & Figure 6

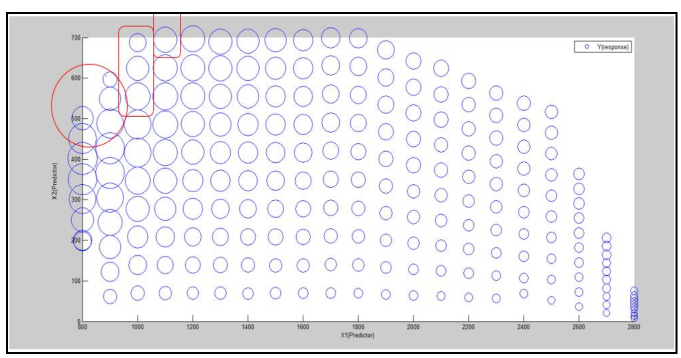


Fig 5 – The data points enclosed in red circle do not follow the trend.

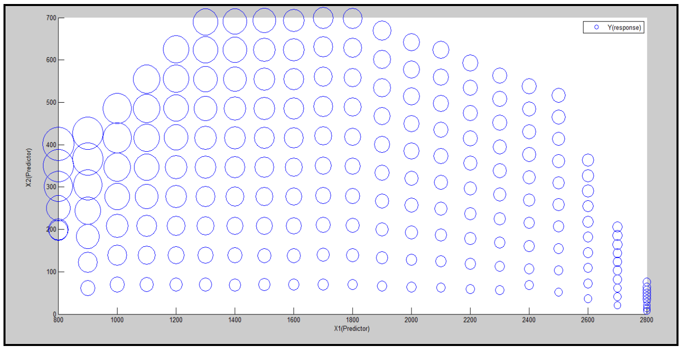


Fig 6 – Filtered data set with outliers removed

C. Regression analysis

With the initial input data for regression, we started working on it which was aimed at generating an equation that can be modelled in MATLAB Simulink.

Regression analysis is a statistical process of correlating the independent variable (Predictors) to the dependent variable (Response). It estimates the conditional expectation of the dependent variable given that the independent variable are held fixed. Regression analysis is widely used in prediction and forecasting. It also depends on the data set being used and the data generating process.

In more detail the general multiple regression model, contains n independent variables:

$$Y_i = a_1 * X_{i1} + a_2 * X_{i2} + \ldots + a_n * X_{in}$$

Where Xin is the ith observation on the nth independent variable, and where the first independent variable takes the value 1 for all i (so "a0" is the regression intercept).

Then the least squares parameter estimates are obtained from n normal equations. The residual can be written as,

$$a_0 = y_i - a_1 * X_{i1} - \ldots - a_n * X_{in}$$

The normal equations are

$$\sum_{i=1}^{n}\sum_{k=1}^{p} X_{ij}X_{ik}\hat{\beta}_k = \sum_{i=1}^{n} X_{ij}y_i,\ j = 1,\ldots,p.$$

Once these outlier data points were removed we generated regression equation using MATLAB or Minitab and reiterate the process until we get a good fit and least possible P-Value and highest percentage of R-square (predictor) value.

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
89.7908	96.05%	95.83%	95.44%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	498	725	0.69	0.493	
X1	0.067	0.140	0.48	0.635	143.23
X2	5.803	0.388	14.94	0.000	115.32
X3	0.2073	0.0752	2.76	0.007	1.75
X4	-7.22	2.53	-2.85	0.005	50.82
X5	193.5	17.3	11.17	0.000	22.28
X6	224.5	42.4	5.30	0.000	433.95
X7	-29.62	3.08	-9.62	0.000	250.90
X8	16.97	7.31	2.32	0.022	6.99
X9	0.44	7.55	0.06	0.953	1.22

Regression Equation

Y = 498 + 0.067 X1 + 5.803 X2 + 0.2073 X3 - 7.22 X4 + 193.5 X5 + 224.5 X6 - 29.62 X7 + 16.97 X8 + 0.44 X9

Fig 7 – Data fit regression using Minitab 17

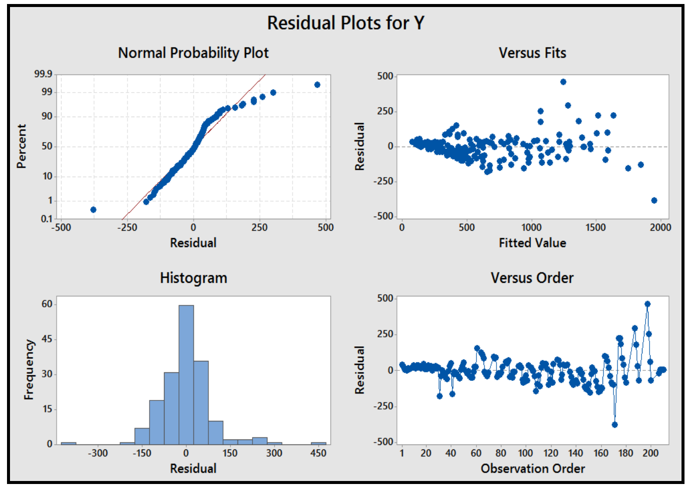


Fig 8 – Residual plots of regression done: Minitab 17

And once we perform multi variate linear regression of the data set we get a form

$Y = a_0 + \sum a_n * X_n$ …………..n ={0,1,2,3…n}

Where,

Y is the response.

a0 is the Regression constant

an is the Coefficients of predictors

Xn is the predictor value obtained from data set

Modelling the regression equation in MATLAB Simulink

This form is a very common form that is encountered during Multivariate Linear regression. Once this equation was generated it was modelled in MATLAB Simulink to generate a data driven model for virtual sensor (Figure 9).

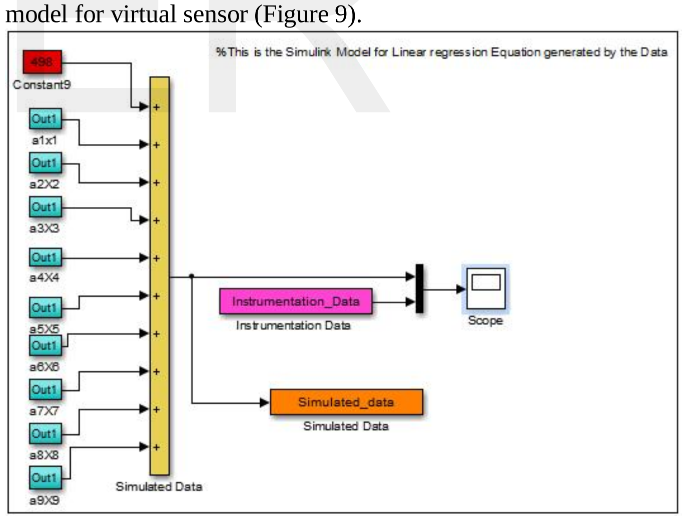


Fig 9 – Modelling the regression equation in MATLAB Simulink (*Instrumentation data is the actual sensor data)*

Above figure shows the regression modelling done using MATLAB Simulink. a1-a9 are the regression coefficients, X1-X4 are the predictors, 498 is the regression constant and the output of the sum block (simulated data block) is the response / output, which is given to the scope to see the results.

This model was later tested and validated using actual test data and plots were compared for Instrumentation data and Simulated data.

4. Results

When the simulation was run for the first time the following trend was observed (Fig 10). The yellow line is the simulated data by the model developed by Simulink MATLAB and the magenta line is the actual physical sensor / instrumentation data. Note that here the simulated data and the instrumentation data almost matches but there are some differences at the peaks of the instrumentation data. This is may be due to incomplete / lack of fine tuning the regression coefficients. We will again fine tune these coefficients further using MATLAB.

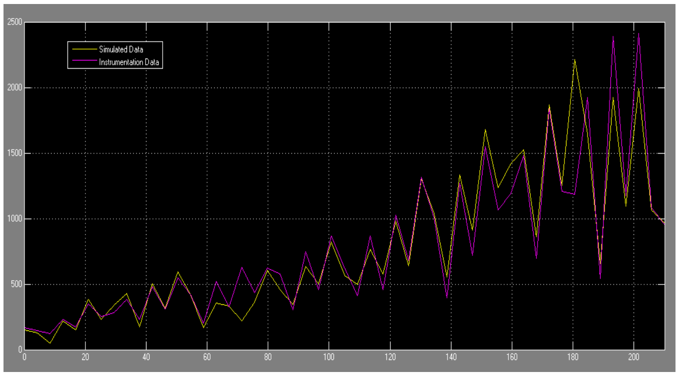


Fig 10 – Simulation results without fine tuning the regression coefficients

After fine tuning the regression coefficients by filtering of data in MATLAB we got the following results (Fig 11).

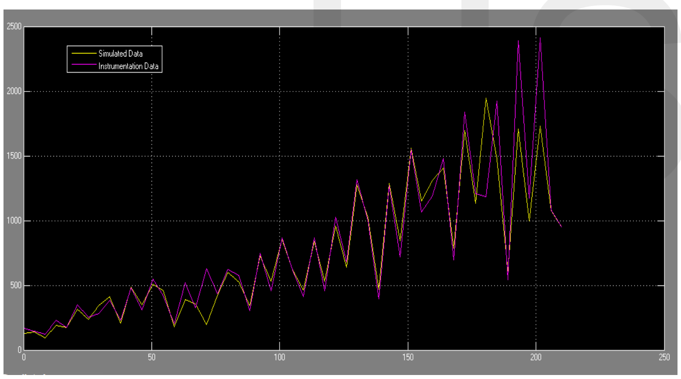


Fig 11 – Simulation results after fine tuning the regression coefficients

Note that here the peaks of the instrumentation data matches almost with the simulated data after fine tuning the regression coefficients. Also there is match in data trends for both of the data.

5. Validation Results

After fine tuning the regression coefficients we got the accuracy level within ± 10 % (Figure 12). Now the instrumentation output (blue line) and the virtual sensor output (red line) matches in values as well as trends. The green line indicates the percentage deviation of the virtual sensor output with respect to the instrumentation output. The % deviation is within ± 10 % as per figure.

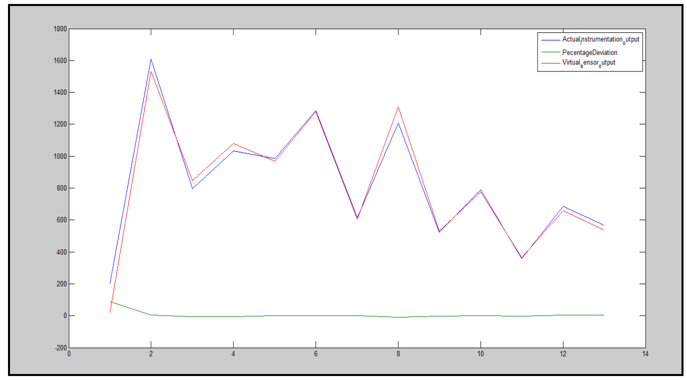


Fig 12 – Validation results with the actual test data

6. Conclusion

We can model virtual sensors to back them up during failure or replace them by developing accurate models which estimates the parameter values based on actual sensors which are present. This proposed Methodology is based on data driven modelling and it can be fine-tuned to get accurate values. The most important part for such model development is getting the correct data and differentiate between good and bad data. This technique has a very wide scope of application in Control Systems for diagnostics to develop models for virtual sensors for fast reactive systems (System where response time required is high).

Acknowledgment

We would like to extend our gratitude to Cummins India Ltd to provide us with the resources, which enabled us to develop this approach towards modelling. We would also like to thank our coaches Nikhil A Pundle and Bipin J Hingu for their constant support.

References

[1] S.Kabadayi, A. Pridgen, and C. Julien, "*Virtual Sensors: Abstracting data from physical sensors,*" in proc. Int.symp. world of wireless, Mobile and Multimedia Networks,2006,pp. 587-592

[2] Patton, R,P Frank, and R Clark, Fault Diagnostics in Dynamic systems: *Theory and application..*

[3] Wunenberg, J and P.M. Frank " *Model Based Residual generation for Dynamics systems with unknown inputs.*" Proc. 12th IMAC world congress on Scientific computation. Paris July 1988.

[4] Virtual sensing techniques and their applications *Lichuan Liu; Kuo, S. M.; Zhou, M. (2009). "Virtual sensing techniques and their applications". 2009 International Conference on Networking, Sensing and Control.*

[5] A systematic approach for soft sensor development Lin, B.; Recke, B.; Knudsen, J. R. K. H.; Jørgensen, S. B. (2007). "A systematic approach for soft sensor development".

[6] David A. Freedman *(2009). Statistical Models: Theory and Practice.* Cambridge University Press*. p. 26.*

[7] *Yan, Xin (2009),* Linear Regression Analysis: Theory and Computing, *World Scientific, pp. 1–2*

[8] Matthew West and Julian Fowler (1999). Developing High Quality Data Models.

[9] Hidalgo, Bertha, and Melody Goodman. "Multivariate or multivariable regression." American journal of public health 103.1