**Paper submitted on Sensor Topic for: International Test and Evaluation Association (ITEA)/ White Sands Missile Range**
**The T&E SYSTEMS-OF-Systems Conference January 24-27 EL Paso, Texas**

# Sensor Validation using Linear Regression for Error Detection between prediction data behavior and acquired data applied on sensing vertical Ground Reaction forces on Human Gait Analysis.

**Author:** Jorge Garza-Ulloa, IEEE member, UTEP Electrical and Computer Eng. Graduate student

**Abstract:**

The significance of sensor validation involves knowledge of common factors that cause low reliability of the acquired data. Such data includes power instability, temperature changes, out-of-range data, internal and external noises, and also synchronization problems that occur when there is the integration of multiple sensor systems. This paper proposes a novel software validation based on Model-driven technique for the prediction of the behavior of the acquired data using Autoregressive Moving Averages (ARMA), and the Data-Driven technique for make a final decision based on the predicted signals. This methodology facilitate the process for determining the validity of the acquired sensor signals, evaluating the levels of noise and providing a timely warning from the expected signals. Experimental results show that this model applied as an example for sensor data validation of the vertical Ground Reaction Forces for human gait is a robust and efficient method to make a correct decision based on groups signal validation results that can be applied on others kind of sensor with trustable results.

**Keywords:** Sensors, Sensor validation, ARMA equation, Regression analysis, Human gait analysis

## 1. Introduction

Today, sensors are everywhere. A lot of them can be found around us like Portable phones, Smoke detection, Fire Alarm, just to mention only some of them. They can be found on different places as: Hospitals, Offices, Industrial Control, and almost every electronic product. There are thousands of different kinds of sensors for: Temperature, Pressure, Velocity, Acceleration, etc. Sensors are so important that can be used since Bio-medical to monitor our health up to Defense and Homeland Security departments that could use hundreds of tiny, wireless sensors packed with computing power to help secure U.S. borders, bridges, power plants, and ships by detecting suspicious movements or dangerous cargo and radioing warnings back to a command center; then the importance of having sensors uncertainties is of urgent priority to resolve. The significance of sensor validation involves knowledge of common factors that cause low reliability of the acquired data as: power instability, temperature changes, out-of-range data, internal and external noises, and also synchronization problems on the integration of multiple sensor systems. As an example to resolve this complex issued on this paper a mathematical model is presented for the validation of sensor data relating to the measured ground reaction forces on the human Gait Cycle, acquired from the instrumented treadmill (Bertec®, Boston, USA). The instrumented treadmill has two separate force plates mounted beneath the dual belts for left and right legs, so the human gait cycle is detected as a walking force and being identified as vertical Ground Reaction Forces (vGRF). The human gait (way locomotion is achieved using human limbs) is indentified on two phases: Stance and Swing. The stance is subdivided on five sub-divisions: Initial Contact (first contact with the right foot), Loading Response, Mid Stance, Terminal Stance and Pre-swing); and the Swing is subdivided on three: Initial Swing (begin when the right foot is not making contact with the floor), Mid Swing and Terminal Swing [Sarkodie-Gyan, 11]. The instrumented treadmill, with typical vGRF chart and the two phases and are shown on Figure 1.

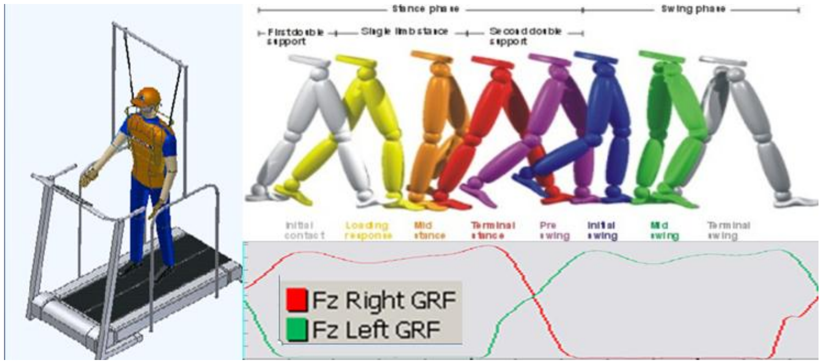



**Figure 1 Instrumented Treadmill, Human Gait phases and a typical vGRF Chart**

On general, there are two main techniques for sensor validation: Physical and Analytical redundancy methods. The physical method known as Hardware Validation is based on measuring variables from more than one sensor for comparison, and the Analytical known as Software validation is based on model-driven or data-driven methods. Regularly on Sensor Data validated by software use only one driven method, on this paper Software validation using the combination of the two, where the model-driven methods are applied for the observation and the estimation of the likelihoods using the Autoregressive Moving Average (ARMA) to presented data and then compared with the Data-Driven Method using the actual data, and from the difference detected take the decision if the data is between acceptable ranges and give a final decision. It is our goal for this software analysis using Model-Data Driven that each data value obtained from the sensors is evaluated on real time as possible for fast response and desirable quality precision decision.

As examples of others similar researches on data validation are: Sensor Data Validation in Aeroengine Vibration Tests [Ai Yanting ,1] whereas the Autoregressive (AR) model and Empirical Mode Decomposition (EMD) were used for diagnosing failures in aero-engine vibration tests, the Sensor validation and fusion of the Nadaraya-Watson statistical estimator an integrated sensor validation and fusion scheme [S. J. Wellington, 2], the Neural Networks for Sensor Validation used for NASA/TM [Duane L. Mattern, 3] and others using Artificial Intelligent techniques and the Fuzzy Logic [Sujit Nath Pant,4][ Huiying Yu, 5].

## 2. Model-driven method: ARMA Equation

Autoregressive Moving Average (ARMA) linear model consists of two parts: Autoregressive (AR) model and Moving Average (MA) model, respectively. The AR model is a regression model in which a time series is regressed on its previous value, and the MA model is a form in which the time series is a moving average of white noise (a random signal with a flat power spectral density: equal power and a fixed bandwidth at any center frequency). ARMA is a statistical forecasting model that is computed based only on the past values of a time-series data [ Brian Borcher,6][ Zhang Shufang ,7][ R. Murray-Smith,8].

The basic ARMA (p,q) equation can be expressed as:

$$r_n = \underbrace{\sum_{i=1}^{p} \varphi_i r_{n-i}}_{AR-Model} + \underbrace{\varepsilon_n - \sum_{i=1}^{q} \theta_i \varepsilon_{n-i}}_{MA-Model} \qquad \text{Eq. (1)}$$

Where:

$r_n$ is the autoregressive time series data;

AR-Model is:

$p$ is the order of the autoregressive data;

$\varphi_i$ (i=1,2,…,p) is the coefficient of AR model;

$r_{n-1}$ is the time series raw data in the previous state;

MA-Model is:

$\varepsilon_n$ is the predicted white noise;

$q$ is the order of the moving average (MA) model of the white noise;

$\theta_i$ (i=1,2,…,q) is the coefficient of the MA model of the white noise.

$\varepsilon_{n-1}$ is the white noise in the previous state

To resolve the model of equation (1) three steps are necessary:

1) Get the coefficients $\{\varphi\}$ of the AR-Model and $\{\theta\}$ for the MA-Model
2) Estimate the order $p$ of the $AR_{(p)}$ and the order $q$ of the $MA_{(q)}$
3) Verification of the convergence of the equation obtained.

The get the coefficients $\{\varphi\}$ of the AR-Model different methods are known, the Yule Walker method is used in this paper [Gidon Eshel, 9], these are calculates using matrices as:

$$\varphi = R^{-1} r \qquad \text{Eq. 2)}$$

Where:

$$\boldsymbol{R} = \begin{pmatrix} 1 & r_1 & r_2 & \cdots & r_{p-2} & r_{p-1} \\ r_2 & 1 & r_1 & \cdots & r_{p-3} & r_{p-2} \\ & \vdots & & & \vdots & \\ r_{p-2} & r_{p-3} & r_{p-4} & \cdots & 1 & r_1 \\ r_{p-1} & r_{p-2} & r_{p-3} & \cdots & r_1 & 1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{r} = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_{p-1} \\ r_p \end{pmatrix}$$

After getting AR-Model coefficients $\{\varphi\}$ the coefficients $\{\theta\}$ for the MA-Model From equation (1) are calculated translating the AR-model to the left side of the equation (1) as shown on the next equation:

$$r_n - \sum_{i=1}^{p} \varphi_i r_{n-i} = \varepsilon_n - \sum_{i=1}^{q} \theta_i \varepsilon_{n-i} \qquad (3)$$

The Estimation of the order $p$ of the $AR_{(p)}$ and the order $q$ of the $MA_{(q)}$ it is very critical because higher order of them will result in data that require more computing time, whereas much lower order may exhibit a reduction in the precision of the model estimation. On this paper specially for the sensor vertical Grounded Reaction Forces (vGRF) several combination of the order $p$ and $q$ were tested, and the results suggested that there very small differences in the errors (less than 0.05% of the normalized value error) for parameters greater than (5,5). Therefore, the order ***p***=5 and the order ***q***=5 were estimated as ARMA (5, 5), and this was verified with the Eigenvalues of the Covariance Matrix method [Gang Liang, 10].

For the verification of the convergence of the results shows the accumulation of errors doesn't take the results to infinity, and then the next transfer function [Brian Borcher,6][ of was used for coefficient verification:

$$\omega(B) = \frac{\theta(B)}{\varphi(B)} \quad \text{Where } \omega(B) \text{ must converges for every } |B| \le 1 \quad (4)$$

Where:

$B$ as backward operator (note that the $B$ is the notation rather than a number)

$\theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$

$\varphi(B) = 1 - \varphi_1 B - \cdots - \varphi_p B^p$

## 3. Data-Driven Experimental Procedure and Model-Data Driven algorithm

Ten physically healthy male were recruited in this study to test the algorithm for different real situations, and different kind of noise some of them induced externally as vibrations. The range of Body Mass Index (BMI) for all subjects is between 20 and 30 kg/m2 and Age from 19 to 40 years old. Each subject was asked to walk with their self-selected speed (from .9 to 1.1 m/s) on the instrumented treadmill (Bertec®, Corp. Boston, USA) continuously for three minutes, each sample every 0.1 sec for a total of 18,000 samples [Huiying Yu, 5]. The vertical ground reaction force (vGRF) data obtained from the instrumented treadmill were extracted for signal processing. Only the right side of the signal was illustrated in this study for easy understanding of this methodology. The Institutional Review Board (IRB) was approved from the University of Texas at El Paso and subjects signed the inform consent prior to the experiment.

The Model-Data driven algorithm for the Sensor Validation using Linear Regression for Error Detection between prediction data behavior and acquired data applied for vGRF on the Human Gait Analysis can be can be summarized on three general steps (Figure.2):

1) **Data Acquisition & Feature Extraction** _ Data representing the vGRF right data with natural frequency of 100 Hz were extracted with total 18000 data samples (three minutes) are obtained and the main features calculated based on Stance and Swings phases: Stride time (one human cycle time) and velocity.
2) **Data Validation using ARMA Model** _ARMA equation is generated and predicted are compared with the actual data, and then flags (Alarms indications) are generated to classify the amount of error obtained on the data validation as: 'Good', 'Medium', 'Warning', and 'Stop' Alarms. On the Alarms (flags) the limit for each alarm band was defined based on the maximum error estimated from the experimental results and can be adjusted accordingly with the type of sensor and its precision application needed. On this case the signal quality is 'Good' if Alarm has a maximum error less than 0.05; 'Medium' if Alarm has a maximum error greater or equal to 0.05 and less than 0.15; it is 'Warning' if the Alarm has a maximum error greater than or equal to 0.15 and less than 0.20, which means the sensor operator has to discard this particular data; and " Stop" Alarm where the operator can manually or automatically stop and check the system when the signal a error greater than or equal to 0.20.
3) **Decision Making based on Validation Results_** Verification are effectuated and validated with the Alarms obtained to give a diagnostic recommendation or action about the data validated.

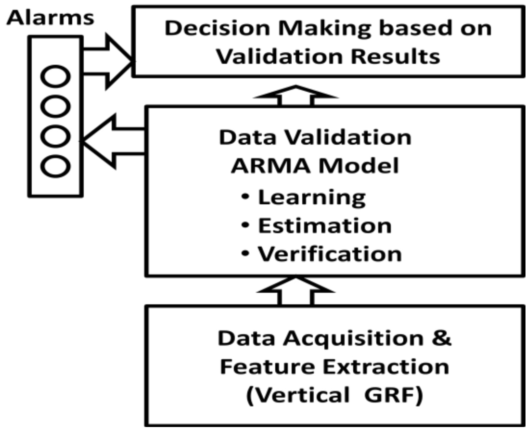
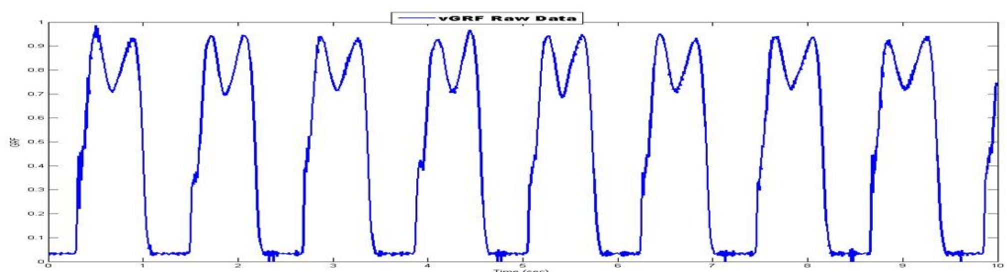



Figure 2 Steps for Sensor Validation using Model and data driven methods

## 4. Results of Data Validation using Model-Data Driven algorithm: ARMA

To test the algorithm one each subject, the right signal from the sensor known as raw data vGRF (without any filtering) to be used later as Data-Driven, then applying the data validation Algorithm using the ARMA process for the Model-Driven, the equation is generated and using this equation the expected signals are obtained, then we obtain the difference from the original signal based on raw-data and the expected signal to apply for the Data-Model driven as shown on Figure 3 (based on data of subject one).

a)




b)

**ARMA (5,5) for Equation (1)**

$AR(5)\quad \varphi\ Model\ Coefficients\ n{=}5$

$r_n = 1r_{n-1} + 5.13r_{n-2} + 4.41r_{n-3} - 8.14r_{n-4} - 2.40r_{n-5}$

$MA(5)\ \theta\ Model\ Coefficients\ n{=}5$

$\varepsilon_n = -1\varepsilon - (5.59 + 0.01i)\varepsilon_{\ n-2} - (7.23 + 0.07i)\varepsilon_{n-3} - (-3.33 + 0.09i)\varepsilon_{n-4} - (0.21 - 0.041i)\varepsilon_{n-5}$

c)

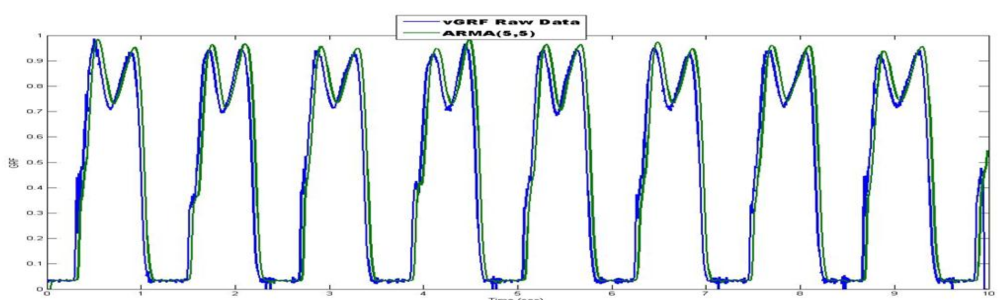



Figure 3 Subject 1: a) Show Signal vGRF Raw Data as Data-Driven b) ARMA (5, 5) Equation calculate for Model Driven, c) Both charts compared for the Data-Model Driven (The first from left to right is raw data and the second the ARMA)

The Data-Model Driven represents the two data: the original Raw Data and the generated by ARMA (5, 5) as the predicted signal. The difference between them is calculated and the Alarms are calculated based on Maximum errors values detected, as indicated on Table 1, where the data of subject 1 is validated with a result of maximum error of 0.13 of normalized signal values.

| **Soft Alarm** (Error < 0.05) | **Medium Alarm** (Error > 0.05 and <0.15) | **Urgent Alarm** (Error > 0.15 and <0.20) | **Stop Alarm** (Error >=0.20) |
|---|---|---|---|
| **0.00** | **0.13** | **0.00** | **0.00** |
| **Accepted Data Validated as Medium Alarm with error <=.15** | | | |

**Table 1 Alarms generated for Subject 1 based on the Model-Data Driven algorithm proposed**

For Each subject walking to a constant speed (from .9 to 1.1 m/s) on the instrumented treadmill continuously for three minutes, taking sample every 0.1 sec for a total of 18,000 samples. To facilitate the location of where data has error the total samples was divided on 18 groups each one with 1000 values and the Alarms are generated now for groups as indicated on Table 2. On this subject one all the 18 groups of 1000 samples where each was accepted gave as a result a maximum error of 0.13 of normalized data values.

| **Segment** ( 1000 samples) | **Soft Alarm** (Error < 0.05) | **Medium Alarm** (Error > 0.05 and <0.15) | **Urgent Alarm** (Error > 0.15 and <0.20) | **Stop Alarm** (Error >=0.20) |
|---|---|---|---|---|
| **1** | **0.00** | **0.11** | **0.00** | **0.00** |
| **2** | **0.00** | **0.13** | **0.00** | **0.00** |
| **3** | **0.00** | **0.13** | **0.00** | **0.00** |
| **4** | **0.00** | **0.11** | **0.00** | **0.00** |
| **5** | **0.00** | **0.12** | **0.00** | **0.00** |
| **6** | **0.00** | **0.12** | **0.00** | **0.00** |
| **7** | **0.00** | **0.12** | **0.00** | **0.00** |
| **8** | **0.00** | **0.13** | **0.00** | **0.00** |
| **9** | **0.00** | **0.13** | **0.00** | **0.00** |
| **10** | **0.00** | **0.12** | **0.00** | **0.00** |
| **11** | **0.00** | **0.12** | **0.00** | **0.00** |
| **12** | **0.00** | **0.12** | **0.00** | **0.00** |
| **13** | **0.00** | **0.13** | **0.00** | **0.00** |
| **14** | **0.00** | **0.12** | **0.00** | **0.00** |
| **15** | **0.00** | **0.11** | **0.00** | **0.00** |
| **16** | **0.00** | **0.12** | **0.00** | **0.00** |
| **17** | **0.00** | **0.12** | **0.00** | **0.00** |
| **18** | **0.00** | **0.12** | **0.00** | **0.00** |
| **Maximum** | | **0.13** | | |
| | **Accepted all Data groups Validated as Medium Alarm with error <=.15** | | | |

**Table 2 The total 18,000 samples (3 min with sample every 0.01 sec) are divided on 18 groups of 1000 samples each for easy error location.**

The data validation results of all ten subjects are summarized on Table 3, on each row only the maximum Alarms Values are indicated with their Maximum Error Normalized values, specifying Group number where that error was detected with the final Data Validation result.

| Subject | Alarm Maximum | Error Maximum | Groups | Data Validation Result |
|---|---|---|---|---|
| **1** | **Medium** | **0.13** | **2, 3, 8, 9, 13** | **Accepted** |
| **2** | **Medium** | **0.12** | **3, 5, 8, 16** | **Accepted** |
| **3** | **Stop** | **0.35** | **3** | **Rejected** |
| **4** | **Medium** | **0.08** | **11, 13, 17, 18** | **Accepted** |
| **5** | **Medium** | **0.10** | **2,5,6,10** | **Accepted** |
| **6** | **Urgent** | **0.15** | **17** | **Accepted?** |
| **7** | **Medium** | **0.11** | **2, 14, 18** | **Accepted** |
| **8** | **Medium** | **0.08** | **4,5,6,10,11,14** | **Accepted** |
| **9** | **Medium** | **0.05** | **17** | **Accepted** |
| **10** | **Medium** | **0.15** | **1,12,17** | **Accepted** |
| **Not Accepted Data Subject 3 Verify group 3** | | | | |

**Table 3 Summarize all ten subjects indicating on each row the Maximum Alarms detected, Maximum Error, the Group numbers with that error and the Final Data Validation results.**

## 5. Discussion

Our experimental results show on table 3 indicates that the data of subject three has a big error when the Model-Data Driven based on ARMA is calculated emitting a "Stop Alarm on Group 3". On Figure 4 the data of this subject including Group one and three are illustrated. Figure 4 a) Group 1: the Raw data (Data-driven) and the expected signal (Model Driven) with acceptable difference, and b) Group 3 shows a cycle interruption on the Raw Data and the expected signal it is not as expected; that why the "Stop Alarm" was generated invalidating the data for this Subject.

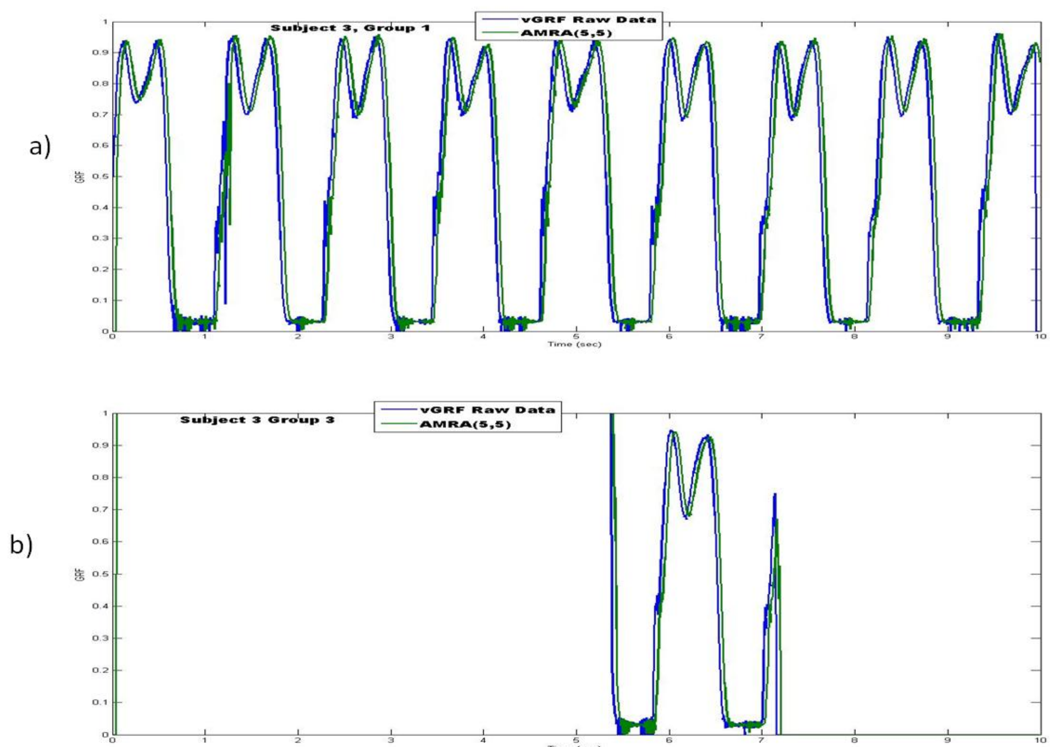



**Figure 4 Subject Three a) Group 1 showing aceptable signal and b) Group 3 with an interruption as Data Corruption**

As another example subject 6 was validated as "Warning Alarm with 0.15 normalized error detected on group 17". Three charts are shown on Figure 5: a) Group 1, b) Group 17 and c) zooming on the same Group 17 where the error was noticeable indicated as Warning Alarm. Because of the Warning Alarm this is a data example that could be validated as acceptable or not, depending of the precision criteria establish for the target application on this case for the vGRF of the human gait cycle.

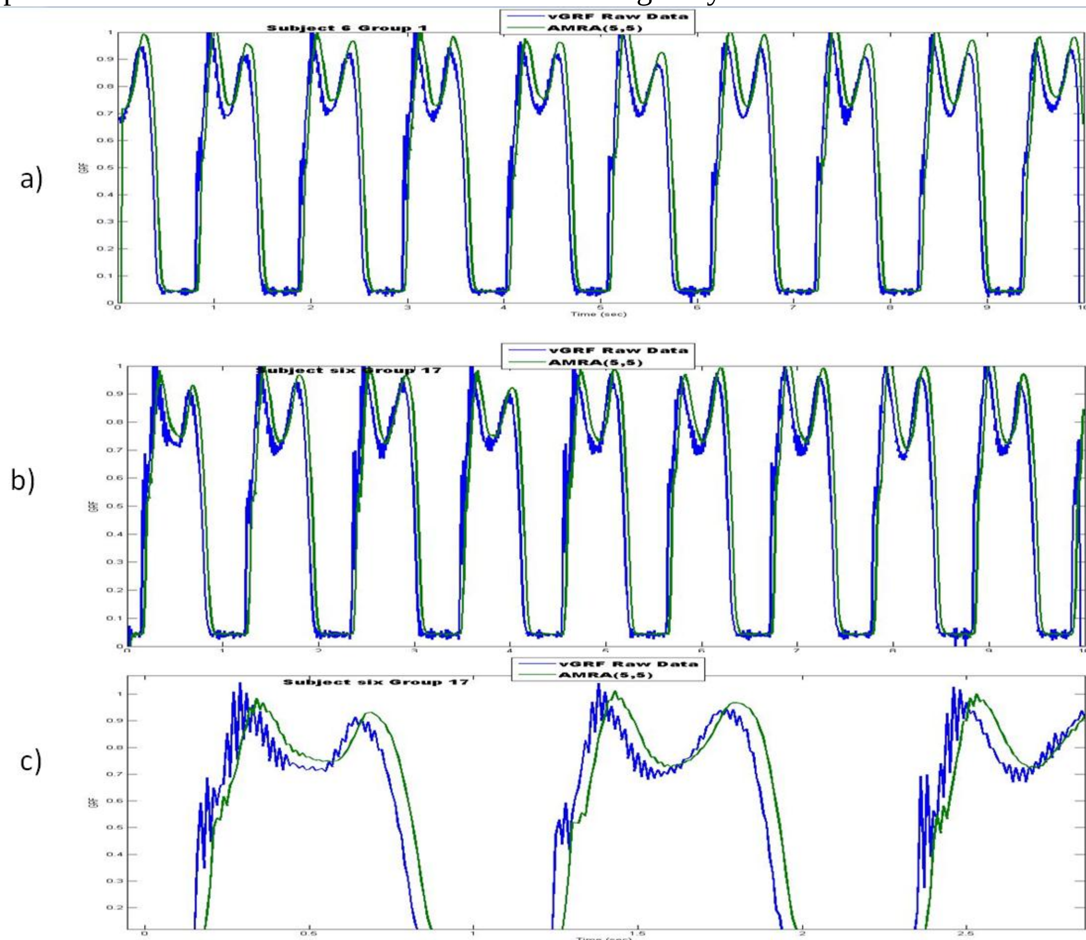



**Figure 5 On subject 6 a) Group 1 is acceptable, b) Group 17 a bigger noise is detected and a warning signal emitted. c) Group 17 with a zoom where the excess of noise was detected.**

## 6. Conclusions

On this paper this mathematical Model-Data driven algorithm for sensor signal validation with respect to ground reaction force analysis of the human Gait cycle prove that is a robust and efficient method to detect when the signals behaviors are correct or between a well defined range of error. This specific Model-Data driven algorithm can be applied on different kind of sensors when the normal behavior is known; it is fast, with sensitivity adjustable to specific needs , allowing easy reproducibility on different areas where sensor validation is a must.

## 7. Future Research

The real benefits of this Model-Data driven algorithm for sensor signal validation will be on the networking smart sensors (wireless, optical and Ethernet) applied as validation on multisensor data fusion on different application as Integrated System Health Management (ISHM), Automotive Monitor systems, Military systems, Aeronautics, Smart Grid, Industrial Control and many more.

**References:**


[1] Ai Yanting, Sun Xiaoqian, Zhang Cheng, Wang Boping: Research on Sensor Data Validation in Aeroengine Vibration Tests, International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), 2010; 3:162-166.

[2] S. J. Wellington, J. K. Atkinson, R. P. Sion: Sensor validation and fusion using the Nadaraya-Watson statistical estimator, Prodeedings of the Fifth International Conference on Information Fusion, 2002; 1:321-326.

[3] Duane L. Mattern, Link C. Jaw, Ten-Huei Guo, Ronald Graham, William McCoy: Using Neural Networks for Sensor Validation, NASA/TM, 1998. AIAA-98-3547.

[4] Sujit Nath Pant, Keith E. Holbert: Fuzzy Logic in Decision Making and Signal Processing, from: http://enpub.fulton.asu.edu/powerzone/FuzzyLogic/

[5] Huiying Yu, J. Riskowski, R. Brower, T. Sarkodie-Gyan: Gait Variability while Walking with Three Different Speeds, 2009 IEEE 11th International Conference on Rehabilitation Robotics Kyoto International Conference Center, Japan, June 23-26, 2009.

[6] Brian Borcher, Rick Aster: Notes on ARMA Processes, November 11, 2008, from http://www.ce.memphis.edu/7137/PDFs/signal%20Processing/arma.pdf.

[7] Zhang Shufang, Liu Renjie: Rapid Algorithm for On-line and Real-time ARMA Modeling, WCCC-ICSP 2000 5th International Conference on Signal Processing, 2000; 1: 230-233.

[8] R. Murray-Smith, A. Girard (2001): *Gaussian process priors with ARMA noise models.* Proceedings of the Irish Signals and Systems Conference, 25th - 27th, June, 2001, NUI Maynooth, Ireland, pp. 147-152

[9] Gidon Eshel, Teaching note on: The Yule Walker Equations for the AR Coefficients, from http://www-stat.wharton.upenn.edu/~steele/Courses/956/ResourceDetails/YWSourceFiles/YW-Eshel.pdf.

[10] Gang Liang, Mitchell Wiles, James Cadzow: ARMA Model Order Estimation Based on the Eigenvalues of the Covariance Matrix, IEEE Transactions on Signal Processing, Vol. 41, No. 10, October, 1993.

[11] Thompson Sarkodie-Gyan (2005): Neurorehabilitation Devices,McGraw-Hill.

[12] Simulation of ARMA-processes, Lund University, Mathematical Statistics, Centre for Mathematical Sciences, Time Series Analysis, Fall 2010, from http://www.maths.lth.se/matstat/kurser/fms051mas216/2010ht/lab0.pdf