

# Validation of Seasonal ARIMA Models on Road Traffic Noise Measurements

Claudio Guarnaccia<sup>1,\*</sup>, Joseph Quartieri<sup>1</sup>, and Carmine Tepedino<sup>1</sup>

<sup>1</sup>Department of Civil Engineering, University of Salerno, via Giovanni Paolo II 132, Fisciano, Italy

**Abstract.** The Time Series Analysis (TSA) technique is largely used in economics and related field, to understand the slope of a given univariate dataset and to predict its future behaviour. The Seasonal AutoRegressive Integrated Moving Average (SARIMA) models are a class of TSA models that, based on the periodicity observed in the series, build a predictive function that can extend the forecast to a given number of future periods. In this paper, these techniques are applied to a dataset of equivalent sound levels, measured in an urban environment. The periodic pattern will evidence a strong influence of human activities (in particular road traffic) on the noise observed. All the three models will exploit the seasonality of the series and will be calibrated on a partial dataset of 800 data. Once the parameters of the models will be evaluated, all the forecasting functions will be tested and validated on a dataset not used before. The performances of all the models will be evaluated in terms of errors values and distributions, such as introducing some error indexes that explain the peculiar features of the models results.

## 1 Introduction

The growth of urban areas, due to the high density of industrial and commercial activities in large and metropolitan cities, is leading to a serious environmental problem. These areas, in fact, are usually affected by several pollutants [1] that can be related primarily to chemical factors influencing air quality, but also to other problems. The main physical agents that affect human health are electromagnetic fields and acoustical noise. In [2] the authors proposed the formulation of a Health Quality Index, based on the composition of several sub-indexes related to different pollutants. Among them, acoustical noise must be considered for both the auditory and non-auditory effects that can cause to any person exposed to high noise doses [3]. In order to assess the noise levels in urban areas, several monitoring campaigns have been promoted by local government. Even though field measurements are the best solution to evaluate the noise exposure, very often it is expensive to monitor acoustical levels all around a large urban area. For this reason, physical and mathematical models are usually built and applied, to assess noise from several sources [4-32].

The authors gave a large contribution to transportation noise assessment and prediction (see for instance [16-32]). In the last years, particular care has been put on models that exploit the information contained in measurements at the receiver: these techniques include analytical models of time series. The authors have developed and improved various deterministic and stochastic models useful for modelling and predicting univariate time series, such as for example [33-39].

In [40], three different Seasonal Auto-Regressive Integrated Moving Average (SARIMA) models of time series analysis have been presented and calibrated on a large dataset of noise measurements in the city of Messina, Italy.

Once the weekly periodicity (lag = 7 days) has been detected, by means of autocorrelation maximization, the calibration of the models parameters has been performed on 800 measured data. Finally, in [40] a comparison between the measured data and the forecasted level has been performed, to highlight the models performances. The comparison has been carried out by means of residuals analysis, both qualitatively and quantitatively, i.e. plotting the results and using different error metrics.

In this paper, the authors present a further step of the application of Time Series Analysis models to road traffic noise, validating the models on data not used in the calibration. This process is necessary to avoid overfitting issues and to verify the real performances of the models on data never used before. In some sense, this test is able to mimic a real application of the models. In fact, in a real application, one can calibrate the chosen model with the data available at the date of the application, and can predict the noise levels in the following days. The goodness of the predictions can be estimated once the real data will be observed, i.e. after the prediction time range is passed. The validation here presented is able to simulate this real scenario, comparing the results of the models with data that are not used for the calibration.

The results of the validation will show that all the models give very good results in one step ahead prediction, with the exception of a small number of outliers.

\* Corresponding author: [cguarnaccia@unisa.it](mailto:cguarnaccia@unisa.it)

## 2 Methods and models

In many scientific fields, it is useful to mathematically describe and predict the evolution over time of a given variable under study. This univariate time series can be modelled, for instance, by a deterministic decomposition model, able to extend the forecast to many periods in the future. Such model typology has been widely adopted by authors for the study of acoustic noise [31-35], concentration of gaseous pollutants [36] and for the evolution of electricity consumption [37].

However, when the studied phenomenon presents rapid fluctuations and a short-term forecast is useful, a stochastic model can be more suitable. In [40], three different stochastic models of the auto-regressive moving-average type have been implemented, also using differentiation operators on an acoustical noise dataset. For a detailed presentation of the models and of the dataset adopted for the calibration, the reader can refer to [40]. In the following subsections, a short description of the models is reported, together with a synthetic resume of the residuals analysis.

### 2.1 Seasonal autoregressive moving average (0,1,1)x(0,1,1)<sub>7</sub> model

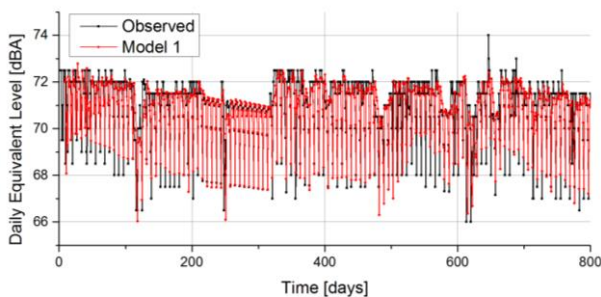
The first model adopted is a Seasonal ARIMA with seasonal lag equal to seven days ( $s = 7$ ): according to the most used notation the model is a SARIMA (0,1,1)x(0,1,1)<sub>7</sub> type. The model can be formulated as follows:

$$\hat{Y}_{t+1} = Y_t + Y_{t-6} - Y_{t-7} - \theta \hat{\epsilon}_t - \theta \hat{\epsilon}_{t-6} + \theta \theta \hat{\epsilon}_{t-7} \quad (1)$$

In table 1, the numerical values of the two moving average coefficients are reported. Figure 1 shows that this model is able to correctly reproduce the general behaviour of the studied seasonal series.

**Table 1.** Estimated value of the coefficients adopted by the SARIMA (0,1,1)x(0,1,1)<sub>7</sub> model. [40]

Coefficients	Estimated Value	Standard Error
MA1( $\theta$ )	0.7078	0.0387
SMA1( $\Theta$ )	0.8966	0.0306
Log likelihood = -974.39; AIC = 1954.78		



**Fig. 1.** Comparison between the observed 800 calibration data and the levels in the same periods predicted by the SARIMA (0,1,1)x(0,1,1)<sub>7</sub> model. [40]

### 2.2 Seasonal autoregressive (7,1,0)x(0,1,0)<sub>7</sub> model

The second model adopted is again a Seasonal ARIMA with seasonal lag equal to seven days ( $s = 7$ ); more precisely the model can be identified as a SARIMA (7,1,0)x(0,1,0)<sub>7</sub> type. In the following formulas,  $y_t$  is the differenced (i.e. “stationarized”) series:

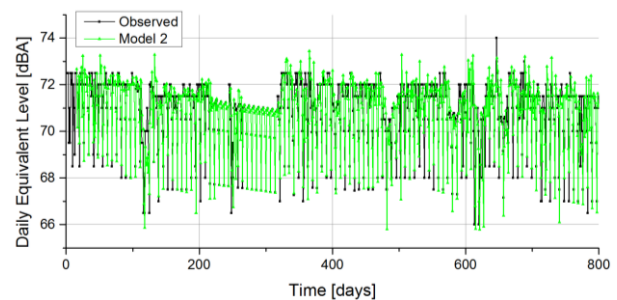
$$y_t = Y_t - Y_{t-1} - Y_{t-7} + Y_{t-8} \quad (2)$$

$$\hat{y}_{t+1} = \phi_1 y_t + \phi_2 y_{t-1} + \phi_3 y_{t-2} + \phi_4 y_{t-3} + \phi_5 y_{t-4} + \phi_6 y_{t-5} + \phi_7 y_{t-6} \quad (3)$$

This second model is based on the assumption that all the useful information is contained in the past seven periods of the series respect to the forecasted day. The model adopt both first and seasonal differentiation and only autoregressive terms. The good predictive performances of this model, in a one step ahead forecasting during the 800 observed days, is shown in Figure 2.

**Table 2.** Estimated values of the coefficients adopted by the SARIMA (7,1,0)x(0,1,0)<sub>7</sub> model. [40]

Coefficients	Estimated Value	Standard Error
AR1( $\phi$ )	-0.3921	0.0330
AR2( $\phi$ )	-0.2703	0.0359
AR3( $\phi$ )	-0.1987	0.0375
AR4( $\phi$ )	-0.1524	0.0377
AR5( $\phi$ )	-0.0442	0.0374
AR6( $\phi$ )	0.0020	0.0362
AR7( $\phi$ )	-0.3763	0.0337
Log likelihood = -1052.27; AIC = 2118.54		



**Fig. 2.** Comparison between the observed 800 calibration data and the levels in the same periods predicted by the SARIMA (7,1,0)x(0,1,0)<sub>7</sub> model. [40]

### 2.3 Seasonal autoregressive moving average (0,0,1)x(1,0,0)<sub>7</sub> model

The third model adopted is a SARIMA (0,0,1)x(1,0,0)<sub>7</sub>, that is slightly different from the previous, since the intercept  $m$  is not null. The basic model and the forecast equation are:

$$Y_t = m + \Phi(Y_{t-7} - m) + e_t - \theta e_{t-1} \quad (4)$$

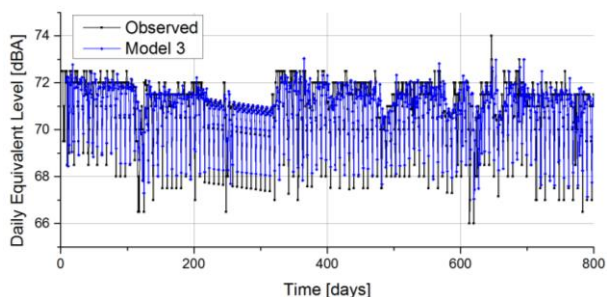
$$\hat{Y}_{t+1} = m + \Phi(Y_{t-6} - m) - \theta \hat{e}_t \quad (5)$$

This third model differs from the others because it does not apply the differentiation operator on the time series under study. Also the third proposed model, with the intercept term, shows good predictive performances, as shown in Figure 3.

**Table 3.** Estimated values of the coefficients adopted by the SARIMA (0,0,1)x(1,0,0)<sub>7</sub> model. [40]

Coefficients	Estimated Value	Standard Error
Intercept	70.7876	0.1924
MA1( $\theta$ )	-0.3535	0.0295
SAR1( $\Phi$ )	0.7833	0.0218

Log likelihood = -1052.88; AIC = 2111.77



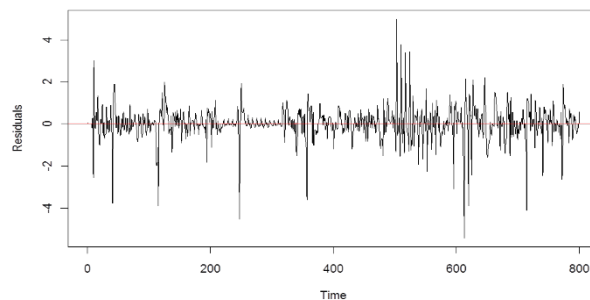
**Fig. 3.** Comparison between the observed 800 calibration data and the levels in the same periods predicted by the SARIMA (0,0,1)x(1,0,0)<sub>7</sub> model. [40]

### 2.4 Residuals analysis of the models in the calibration phase

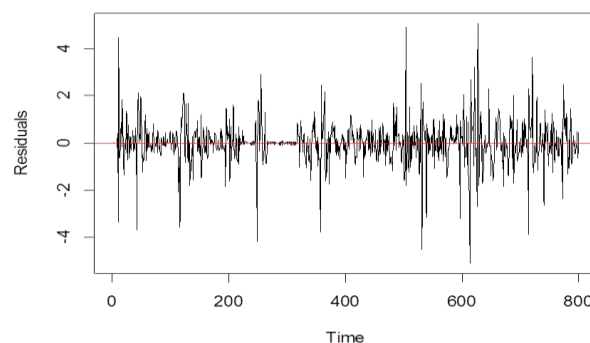
In this section the statistical analysis of the forecasting errors, denoted in the calibration phase with the term “residuals”, is reported. In [40] the measurements of central tendency and dispersion for the three models were calculated, together with the minimum and maximum value of the residuals and skewness and kurtosis indexes. These results are resumed in Table 4. The plots of the residuals for all the models are presented in Figs. 4a, 4b and 4c. It is evident that all the three models are capable of a good forecasting in the calibration dataset: mean and standard deviation are very low and the distributions of the error are almost symmetrical. In some specific periods, the three models are not able to follow drastic fluctuations of the acoustical noise, so the minimum and maximum values of the residuals exceed 5 dBA. An outlier detection and removal analysis could improve these parameters, reducing the value of minimum and maximum residuals.

**Table 4.** Summary statistics of the residuals distribution evaluated on the calibration dataset for the three models.

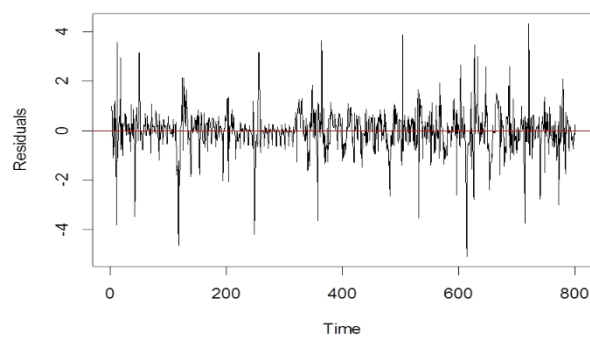
	Mean [dBA]	Std. Dev. [dBA]	Median [dBA]	Min [dBA]	Max [dBA]	Skew	Kurt
Mod1	0.02	0.84	0.03	-5.46	5.00	-0.67	8.44
Mod2	0.00	0.89	0.00	-5.12	5.09	-0.22	6.90
Mod3	-0.01	0.88	0.02	-5.10	4.35	-0.41	6.40



a)



b)



c)

**Fig. 4.** Plot of the residuals in the calibration dataset, obtained using Model 1 (a), Model 2 (b) and Model 3 (c). Horizontal axis scale is in days, vertical axis scale is in dBA.

### 3 Validation results

The dataset adopted for the validation is part of the noise levels time series measured in the city of Messina, Sicily, in the south of Italy, and used for the calibration [40]. The city of Messina has about 240000 inhabitants and, among the various problems of pollution, it shows persistent acoustical noise caused by vehicular traffic. In different areas of the city the local administration has located noise monitoring stations, in which fixed and mobile sound

level meters are installed. Therefore, time series of acoustical levels are available to study the noise pollution phenomenon. In this paper, as well as in [40], the authors have taken into account the equivalent daily levels, weighted with the “A” weighting curve, registered in the site of “Via La Farina”.

The measurement of road traffic noise was carried out by the "Environmental Monitoring Service" of Messina.

The noise levels series used for these studies refers to the daily time period, from 6:00 am to 10:00 pm. 800 data have been used in the calibration phase of the models [40], and the following 132 data are used in the validation phase presented in this paper. The calibration dataset was composed by 628 measured data and 172 data imputed using the technique described in [41].

The summary statistics of the validation dataset are shown in Table 5. These statistics confirm the high noise levels experienced in the urban area under study, considering the high mean, the small standard deviation and the small spread.

In Fig. 5, the observed and simulated noise levels are plotted as a function of the time, in the validation periods. Let us underline that a “one step ahead” prediction is performed, in the sense that each predicted value is obtained giving in input the previous data, according to equations 1-5. The parameters of the models are fixed after the calibration, but the past observed data are needed to produce the prediction. A further forecast, on a larger time range, can be computed considering the seasonal lag. This application will be presented in further papers.

It is evident that all the models furnish good predictions in this time range, even though these data are not used in the calibration phase. Except for few outliers, the observed and predicted data are very close. The “one step ahead” prediction allows to follow short term variations, such in the range 840-860 periods, in which a reduction of the noise levels is observed.

These good results are quantitatively confirmed by the error statistics presented in Table 6. The error is calculated as:

$$\hat{e}_t = Y_t - \hat{Y}_t \tag{6}$$

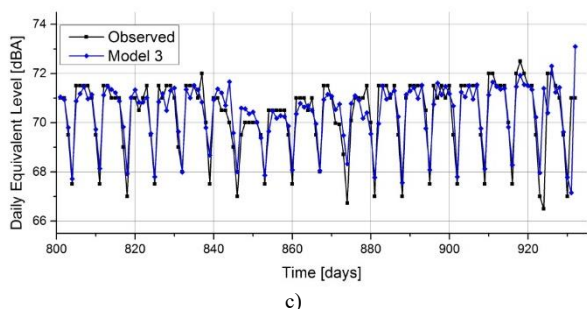
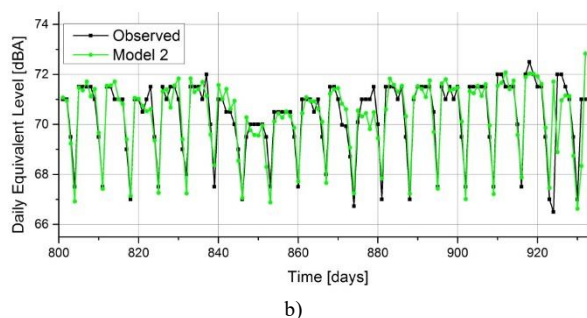
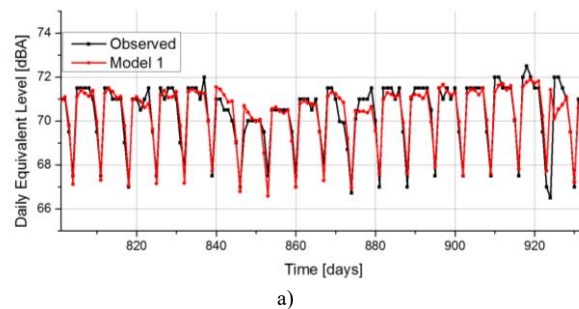
i.e. observed noise level minus predicted value, in each time period of the validation dataset.

Model 1 seems to have the best performances, even though the median is higher than the other models. The minimum and maximum values of the error have larger range in the case of Model 3, in which also the mean and standard deviation are slightly worse.

In Fig. 6 the bisector plots are reported for all the models. The forecasted noise levels are plotted versus the observed data. A good model is expected to present data gathered along the bisector (red solid line). It is possible to see that almost all the predictions are included in the  $\pm 2$  dBA range, highlighted by the red dashed lines. The small number of points that fall outside that range are related to unexpected noise levels that cannot be predicted by this kind of model.

**Table 5.** Summary of statistic main parameters of the 132 days of the validation dataset.

	Mean [dBA]	Std. Dev. [dBA]	Median [dBA]	Min [dBA]	Max [dBA]	Skew	Kurt
Valid. data	70.3	1.49	71.0	66.5	72.5	-1.1	0.04

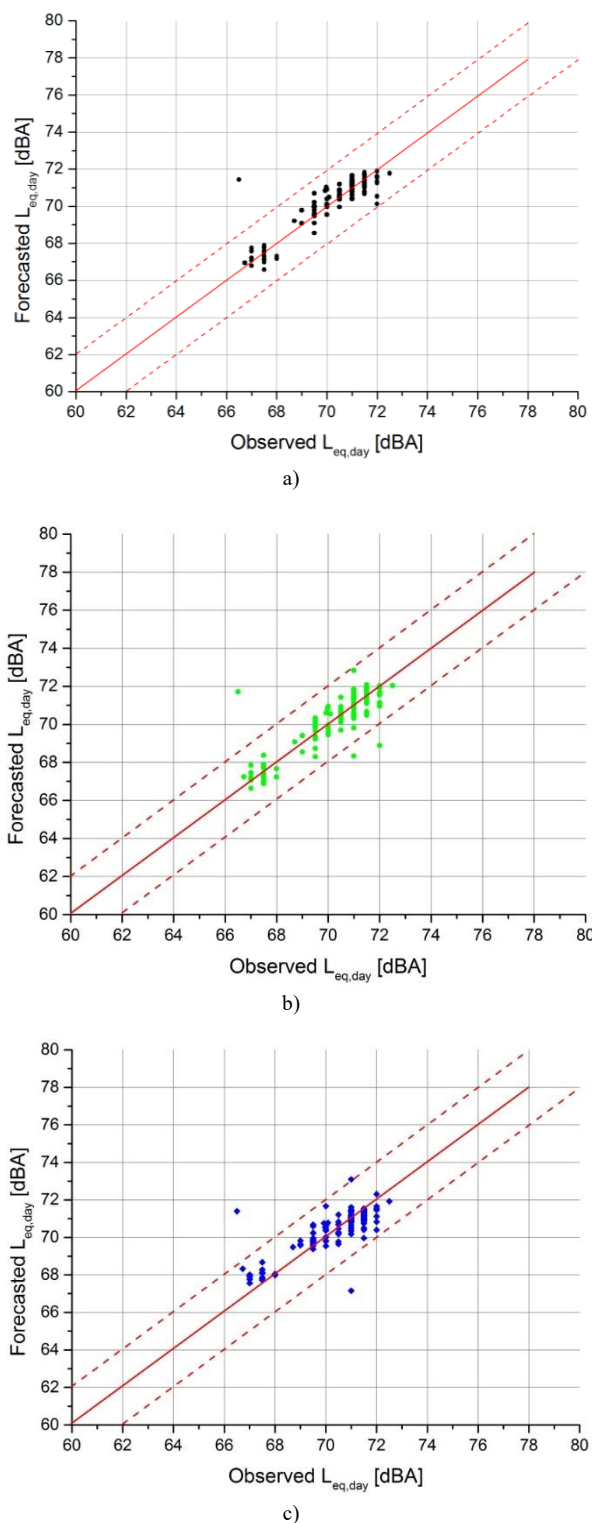


**Fig. 5.** Comparison between the observed 132 validation data and the levels in the same periods forecasted by the one step ahead SARIMA Model 1 (a), Model 2 (b) and Model 3 (c).

**Table 6.** Summary statistics of the errors distribution evaluated on the validation dataset using the one step ahead forecast.

	Mean [dBA]	Std. Dev. [dBA]	Median [dBA]	Min [dBA]	Max [dBA]
Model 1	0.00	0.64	0.03	-4.93	1.88
Model 2	0.01	0.76	-0.01	-5.21	3.11
Model 3	-0.07	0.80	-0.02	-4.90	3.85





**Fig. 6.** Bisector plot with confidence interval between the observed 132 validation data and the levels in the same periods forecasted by the one step ahead SARIMA Model 1 (a), Model 2 (b) and Model 3 (c).

## 7 Conclusions

In this paper, the validation of three SARIMA models calibrated on a long term dataset of urban noise levels was reported.

The models implement the weekly periodicity, with a lag of 7 days, since a large autocorrelation in the noise levels measured in the same day of the week has been observed. Among the three models, two of them embed a differentiation of order 1, to achieve a better stationarity of the series.

The calibration process, on 800 daily noise level measured in the city of Messina, was presented in [40], while in this paper the authors reported the validation on the following 132 data.

The models are able to produce a “one step ahead” prediction with very good results. Both the graphical comparison and the quantitative analysis of the error statistics confirmed that all the models succeeded in following short term variations and in predicting the urban noise levels. In a small number of cases, the predictions are quite far from the observed data, probably due to unexpected variations in the noise source phenomenon.

Let us underline that these models are able to produce very good predictions without the need to input any data about the road traffic volume, mean speed, boundary conditions, geometry of the area, presence of intersections, etc.. Up to the authors opinion, this is the main strength point of the SARIMA approach, i.e. the ability of give good forecasts using in input just a sufficiently large dataset of past measurements.

This paper reports one of the last research topics made by Carmine Tepedino, who passed away on November 30, 2017, at the early age of 34. He implemented the models, performed the analysis of the data and produced the plots. Joseph Quartieri and Claudio Guarnaccia deeply miss his scientific and human great qualities.

## References

1. E. Kahn, *Brookings Institution Press*, Washington, DC, USA, (2006)
2. J. Quartieri, A. Troisi, C. Guarnaccia, P. D’Agostino, S. D’Ambrosio, G. Iannone, Proc. of the WSEAS Int. Conf. EED’09, (Puerto de la Cruz, Tenerife (Spain), 14-16 December 2009)
3. C. Guarnaccia, N. E. Mastorakis, J. Quartieri, *Int. J. of Mech.* Vol. **7**(2), pp 37-44 (2013)
4. A. Can, E. Chevallier, M. Nadji, L. Leclercq, *Acta Acust united Ac* Vol. **96**(3), pp. 482-493 (2010)
5. A. Can, P. Aumond, *Transport Res D-Tr E* Vol. **58**, pp. 155-171 (2018).
6. D. Siano, M. Viscardi, R. Aiello, *Enrgy Proced* Vol **82**, pp. 67-74 (2015).
7. M. Cárdenas Rodríguez, L. Dupont-Courtade, W. Oueslati, *Renewable and Sustainable Energy Rev.* Vol. **53**, Pages 1–9, January (2016)
8. P. E. Kirrian Fiedler, P. H. Trombetta Zannin, *Environmental Impact Assessment Rev.* Vol. **51**, Pages 1–9, February (2015)
9. J. Prezelj, J. Murovec, *Appl. Acou.* Vol. **127**, pp. 160-168, December (2017)
10. Bao-Lin Ye, W. Wu, H. Gao, Y. Lu, Q. Cao and L. Zhu, *Appl. Sci.* Vol. **7**(6), 588, pp. 1-17, (2017)

11. C. Guarnaccia, WSEAS Trans. on Systems Vol. **12**(2), pp. 121-130 (2013)
12. J. Quartieri, N.E. Mastorakis, G. Iannone, C. Guarnaccia, S. D'Ambrosio, A. Troisi, T.L.L. Lenza, Proc. of the 5th WSEAS *Int. Conf. MECHANICS'09*, Puerto de la Cruz, Tenerife, Spain, December 14-16 (2009)
13. C. Guarnaccia, J. Quartieri, A. Ruggiero, Int. J. of Mech. Vol. **8**, pp. 298-306 (2014)
14. F. Bunn, P. H. Trombetta Zannin, Appl. Acou. Vol. **104**, Pages 16–23, (March 2016)
15. A. Ruggiero, J. Quartieri, C. Guarnaccia, S. Hloch, Int. J. of Environmental Research, Vol. **9** (4), pp.1277-1286, (2015)
16. J. Quartieri, A. Troisi, C. Guarnaccia, T.L.L. Lenza, P. D'Agostino, S. D'Ambrosio, G. Iannone, WSEAS Trans. on Systems Vol. **8**(6), pp. 673-682 (2009)
17. J. Quartieri, A. Troisi, C. Guarnaccia, T.L.L. Lenza, P. D'Agostino, S. D'Ambrosio, G. Iannone, WSEAS Trans. on Systems Vol. **8**(4), pp. 481-490 (2009)
18. J. Quartieri, A. Troisi, C. Guarnaccia, T.L.L. Lenza, P. D'Agostino, S. D'Ambrosio, G. Iannone, in Proc. of the 10<sup>th</sup> WSEAS *Int. Conf. AMTA '09*, Prague (Rep.Ceca), 23-25 March (2009)
19. J. Quartieri, A. Troisi, C. Guarnaccia, T.L.L. Lenza, P. D'Agostino, S. D'Ambrosio, G. Iannone, Proc. of the 10<sup>th</sup> WSEAS *Int. conf. AMTA '09*, Prague (Rep.Ceca), 23-25 March (2009)
20. C. Guarnaccia, T.L.L. Lenza, N.E. Mastorakis, J. Quartieri, Int. J. of Mech. Vol. **5**(4), pp. 379-386, (2011)
21. C. Guarnaccia, WSEAS Trans. on Systems Vol. **9**(8), pp.865-874, (2010)
22. G. Iannone, C. Guarnaccia, J. Quartieri, Env. Eng. and Manag. J. Vol. **12**(3), pp. 493-501 (2013)
23. J. Quartieri, G. Iannone, C. Guarnaccia, Proc. of the 11th WSEAS *Int. Conf. AMTA '10*, Iasi, Romania, June 13-15 (2010)
24. C. Guarnaccia, J. Quartieri, Int. J. of Math. Models and Methods in Applied Sciences **6**(8), pp. 926-933, (2012)
25. G. Iannone, C. Guarnaccia, J. Quartieri, in *Recent Researches in Geography, Geology, Energy, Environment and Biomedicine*, Proc. of the 4th WSEAS *Int. Conf. EMESEG '11*, Corfù Island, Greece, July 14-16, pp. 501-507 (2011)
26. J. Quartieri, N.E. Mastorakis, C. Guarnaccia, G. Iannone, Proc. of the 12th *Int. Conf. ACMOS '10*, Catania (Italy), May 29-31, pp. 299-304 (2010)
27. C. Guarnaccia, Proc. of the 11th WSEAS *Int. Conf. AMTA '10*, (Iasi, Romania, June 13-15 2010)
28. J. Quartieri, N.E. Mastorakis, C. Guarnaccia, A. Troisi, S. D'Ambrosio, G. Iannone, Proc. of the 5th WSEAS *Int. Conf. MECHANICS '09*, Puerto de la Cruz, Tenerife, Spain, December 14-16 (2009)
29. C. Guarnaccia, J. Quartieri, N.E. Mastorakis, C. Tepedino, Int. J. of Mech. Vol. **10**, pp. 14-22, (2016)
30. C. Guarnaccia, J. Quartieri, J.M. Barrios, E.R. Rodrigues, J. of the Acou. Society of America Vol. **136**, pp. 1631-1639 (2014)
31. C. Guarnaccia, J. Quartieri, C. Tepedino, E.R. Rodrigues, Appl. Acou. Vol. **91**, pp. 33-39, (2015)
32. J.M. Bandeira, C. Guarnaccia, P. Fernandes, M.C. Coelho, Int. J. of Intelligent Transportation Systems Research Vol. **7**, (2016)
33. C. Guarnaccia, J. Quartieri, C. Tepedino, E. R. Rodrigues, Appl. Acou. Vol. **114**, pp. 203–212 (2016)
34. C. Guarnaccia, J. Quartieri, N.E. Mastorakis, C. Tepedino, WSEAS Trans. on Systems Vol. **13**, pp. 745-756, (2014)
35. C. Guarnaccia, J. Quartieri, E.R. Rodrigues, C. Tepedino, Int. J. of Math. Models and Methods in Applied Sciences Vol. **8**, pp 384-393, (2014)
36. C. Guarnaccia, J. Quartieri and C. Tepedino, in Proc. of the Int. Conf. on Applied Mathematics and Computer Science, AIP Conference Proceedings 1836, pp. 1-7, (2017)
37. C. Guarnaccia, J. Quartieri and C. Tepedino, in Proc. of the Int. Conf. on Applied Mathematics and Computer Science, AIP Conference Proceedings 1836, pp. 1-8, (2017)
38. C. Guarnaccia, J.G. Cerón Bretón, J. Quartieri, C. Tepedino, R.M. Cerón Bretón, Int. J. of Math. Models and Methods in Applied Sciences Vol. **8**, pp 505-515, (2014)
39. C. Tepedino, C. Guarnaccia, S. Iliev, S. Popova and J. Quartieri, Int. J. of Math. Models and Methods in Applied Sciences Vol. **9**, pp 432-445, (2015)
40. C. Guarnaccia, N.E. Mastorakis, J. Quartieri, C. Tepedino, S.D. Kaminaris, S.D., MATEC Web Conf Vol. **125**, 05013 (2017)
41. C. Guarnaccia, J. Quartieri, C. Tepedino, L. Petrovic, Int. J. of Mech. Vol. **9**, pp. 272-278, (2015)
42. P.D. Cryer, K. Chan, Second Edition, Springer, (2008)