ISSN 1980-7694 ON-LINE

THE ESTIMATION OF THE HEIGHT OF SEEDLINGS OF *Tibouchina granulosa* COGN. (MELASTOMATACEAE) IN ARTIFICIAL NEURAL NETWORKS

Jéssica Cristina Barbosa Ferreira [1]; Bruno Oliveira Lafetá [2]; Tamires Mousslech Andrade Penido[3]; Paulo Modesto de Campos[4]; Pedro Mendes Castro[5] [6]

ABSTRACT

This study aimed to evaluate the estimation efficiency for seedlings height of *Tibouchina granulosa* in terms of root collar diameter (RCD), under different substrate compositions, using Artificial Neural Networks (ANN). We selected 72 seedlings produced in plastic tubes for transplanting in 25-L buckets. The experiment was established in a completely randomized design with three repetitions, constituted of four substrate compositions. Each experimental unit was composed of six seedlings. At 13 months of age, the RCD and total height (H) of all seedlings were measured. We trained 200 ANN to estimate H, being 100 Multilayer Perceptron (MLP) and 100 Radial Basis Function (RBF). The variables used as input to ANN for estimating seedlings height numerical (RCD and H) and categorical (T: Substrate 1 - T1; Substrate 2 - T2; Substrate 3 – T3, and Substrate 4 – T4). The ANN modeling using MLP architecture is appropriate and accurate to estimate seedlings height of *Tibouchina granulosa.*

Key-words: Artificial intelligence; Hypsometric Relationships; Urban vegetation.

ALTURA DE MUDAS DA *Tibouchina granulosa* COGN. (MELASTOMATACEAE) ESTIMADA POR REDES NEURAIS ARTIFICIAIS

RESUMO

O objetivo do presente trabalho foi avaliar a eficiência da estimação da altura de mudas da *Tibouchina granulosa* em função do diâmetro do coleto, sob diferentes composições de substrato, empregando Redes Neurais Artificiais (RNA). Foram selecionadas 72 mudas produzidas via tubetes para a repicagem em baldes de 25 litros. Adotou-se delineamento experimental inteiramente casualizado, com três repetições, sendo os tratamentos constituídos por quatro composições de substrato. Cada unidade experimental foi composta por seis mudas. Aos 13 meses de idade foram mensurados o Diâmetro à Altura do Coleto (DAC) e a altura total (H) de todas as mudas. Foram treinadas 200 RNA para estimar a H, sendo 100 *Multilayer Perceptron* (MLP) e 100 *Radial Basis Function* (RBF). As variáveis utilizadas como entrada das RNA para estimação da altura das mudas foram numéricas (DAC e H) e categórica (T: Substrato 1 – T1; Substrato 2 – T2; Substrato 3 – T3 e Substrato 4 – T4). Conclui-se, assim, que a modelagem por RNA utilizando arquitetura MLP é adequada e precisa para estimar a altura de mudas da *Tibouchina granulosa*.

Palavras-chave: Inteligência artificial; Relações hipsométricas; Vegetação urbana.

[1] Graduanda de Tecnologia em Silvicultura, IFMG, bruno.lafeta@ifmg.edu.br; jessicacbf.ifmg@gmail.com
[2] Mestre em Ciência Florestal, Professor de Tecnologia em Silvicultura, IFMG, bruno.lafeta@ifmg.edu.br
[3] Graduanda de Engenharia Florestal, UFVJM - penidotma@gmail.com
[4] Especialista em Fruticultura Comercial, IFMG - paulo.campos@ifmg.edu.br
[5] Engenheiro Ambiental, Arborista Especialista em Sistemas Elétricos, CEMIG - pedro@cemig.com.br
[6] recebido em 20/01/2014 e aceito para publicação em 15/03/2014

INTRODUCTION

Urban afforestation provides social, ecological, and aesthetic benefits to populations (ROCHA et al., 2004). It requires a prior planning to avoid damage to sidewalks, underground pipelines, buildings, and aerial electrical networks (GONÇALVES et al., 2004). The use of plants adequate to urban afforestation is essential, and the seedling size can influence the urban design.

The production of seedlings for landscaping is a sustainable activity that seeks to increase green areas in cities, avoiding heat islands and providing more pleasant environmental conditions for the urban residents (ROSSATO et al., 2008).

The characterization of the height and diameter of seedlings is an important step to determining the tree volume, which assists the dimensioning of groves and best places for tree planting in urban areas.

Tibouchina granulosa Cogn. (Melastomataceae) is widely used in urban landscapes in Brazilian cities. It is popularly known as "quaresmeira" that has tolerance to direct light and hardiness (CÉZAR, et al., 2009). In addition, quaresmeira may be indicated for afforestation of narrow streets and under aerial electrical networks, improving air quality due to the ability of their crowns in reducing particulate matter (LOPES et al, 2005; ZAMPIERI et al., 2013).

Generally, measurements of the tree trunk diameters are usually easy and quick to obtain, however, the tree height is more time consuming and therefore more taxing for data collection. The relationship between the diameter and the height of plants is called hypsometry (CAMPOS; LEITE, 2009). The use of hypsometric equations is frequent and in the search for new options for efficient and safe estimation of height, artificial intelligence stands out. Artificial Neural Networks (ANN) represent a new approach for the development of predictive models able to learn complex patterns and data trends, even if they are not normal (HAYKIN, 2001; SCRINZI et al., 2007).

The RNA are computer models that simulate the structure and processing of the human nervous system, establishing mathematical relationships between dependent and independent variables (GORGENS et al., 2009; MAEDA et al., 2009). The network architecture influences the processing of input data, allowing linear approximations or not in the medium layer (LAFETÁ et al., 2012). The most common architectures have the basis of radial (Radial Basis Function or RBF) and the multiple layers (Multilayer Perceptron or MLP).

The performance of RNA has been superior to regression models vis-à-vis its own characteristics of adaptability of the weights of the connections to environmental modifications, fault tolerance and noises, neurobiological analogy, complex problem solving, and learning skills (BINOTI et al., 2010; BINOTI, 2010).

A report of the RNA can be evaluated by coordination of a type of cross-validation ratio-emotion designated as validation method (HAYKIN, 2001). It is based on the non-portioning of the dataset into subsets mutually exclusive, where part of the data is used in training and the rest, in test and validation. This procedure is important to test the capacity to classify correctly patterns not included without training, and, therefore, in the definition of the network with better generalization capacity (FERNANDES et al., 2004).

Studies on hypsometric relationships and advanced statistical techniques can assist in establishing practical criteria for urban afforestation. Thus, this work aimed to evaluate the efficiency of estimating the height of seedlings of *Tibouchina granulosa* in terms of diameter of the samples in different substrate compositions, using RNA.

MATERIALS AND METHODS

The experiment was conducted in an open environment in the nursery of seedlings of Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais – Campus São João Evangelista, Minas Gerais State, Brazil (IFMG/SJE) at 22°16'13" S lat; and 54°48'20"W long. The climate of the region is temperate rainy-mesothermal and classified as Cwa (Köppen system), with dry winters and rainy summers. The annual average rainfall is 1,400 mm and the average annual temperature of 21°C (BRAGA et al., 1999).

We selected 72 seedlings of *Tibouchina granulosa* produced in tubes for transplantation to 25-L buckets. The seedlings were healthy, free of injuries caused by phytopathogens and with a length of 20-25 cm.

The experimental design was completely randomized with three repetitions, and treatments were T1 – Substrate 1 (79.2% of subsoil earth, 10.0% of coal waste, 10.0% of commercial substrate, 0.13% of osmocote 19:6:10 and 0.67% of NPK 6:30:6). T2 – Substrate 2 (73.54% of subsoil earth, 7.33% of coal waste, 18.33% of commercial substrate, 0.13% of osmocote 19:6:10 and 0.67% of NPK 6:30:6). T3 – Substrate 3 (49.87% of subsoil earth, 49.33% of aged cow manure, 0.13% of osmocote 19:6:10 and 0.67% of NPK 6:30:6). T4 – Substrate 4 (49.60% of subsoil earth, 24.80% of commercial substrate, 24.80% of aged cow manure, 0.13% of osmocote 19:6:10 and 0.67% of NPK 6:30:6). Each experimental unit was comprised of six seedlings. At 13 months of age, the samples were measured at root collar diameter (RCD) and total height (H) of all seedlings.

The training of an artificial neural network, also called learning, consists of iterative adjustment of network parameters (weights and bias) through a learning algorithm (MAEDA et al., 2009). In this process, the training data (set of examples) are presented to a pre-established architecture, that is, a certain number of arrangements of neurons in layers. The training algorithm extracts characteristics to represent the data provided and perform a certain task. The variables used as input of RNA for estimating the height of seedlings were numeric (DAC ($_{mm}$) and H ($_{cm}$)) and categorical (T: 1 – T1; 2 – T2; 3 – T3 and 4 – T4).

Feedforward networks were used, trained through backpropagation algorithm of error, that is, during the network training, we performed calculations from the input layer of the network to the output and the error propagated to previous layers. In all pre-processing, we performed on standardization and equalization of the data. The data were divided into training groups (80.0% of the samples), test (10.0% of the samples) and validation (10.0% of samples) using the random sampling method.

We trained 200 RNA to estimate H, being 100 Multilayer Perceptron (MLP) and 100 Radial Basis Function (RBF). We used the heuristic model backward elimination, as described by Cerqueira et al. (2001). From these RNA, we selected one of each type based on deviations of the estimated and observed values. One of the most common problems encountered in training RNA is the overfitting. Seeking to avoid it, the training of the networks was interrupted when the error started to increase, in accordance to BRADSHAW et al. (2002) and MAEDA et al. (2009).

The optimal number of medium layers and neurons per layer, generally, is not known *a priori*. Once defined architecture and learning parameters, the ANN is trained in an iterative manner (BLACKARD; DEAN, 1999). Therefore, the definition of architecture was optimized networks in the tool Intelligent Problem Solver (IPS) of the software Statistica 7.0 (STATSOFT, 2007).

The points that extrapolated the general trend of each treatment were not eliminated to check the capacity of ANN to deal with outliers or noises. The assessment of accuracy and precision and the comparison of networks training was carried out according to the criteria defined by Lafetá et al. (2012). The criteria were based on relative error, root of the mean square error (RMSE), bias (Bias), the paired *t*-test at 5.0% of statistical significance, and analysis of scatter charts and distribution percentage frequency of residual percentages.

The training, the test and validation of the networks were carried out with the aid of software Statistica 7.0 (STATSOFT, 2007).

RESULTS

The nets showed non-linear activation functions in medium layers (Table 1). The sigmoidal behavior on output layer was observed only in MLP. The training phase of networks with MLP and RBF architecture had 20 and 182 cycles, respectively.

Table 1. Characterization of Artificial Neural Networks (ANN) for seedlings height estimation of the *Tibouchina granulosa* for urban landscape.

RNA	n	Architecture	Activation Function	
			Medium	Output
MLP	72	5-11-1	Exponential	Logistic
RBF	72	5-12-1	Gaussian	Identity

MLP = *Multilayer perceptron.* RBF = *Radial Basis Function*. n = number of observations.

The average values of RMSE for MLP and RBF were of 9.01% and 12.52%, respectively (Table 2). In general, during the training we observed low RMSE and Bias. We observed larger ranges of relative errors in RBF at all stages of processing. The observed and estimated data were statistically similar to each other by *t*-test in both networks built.

Table 2. Accuracy of Artificial Neural Networks (ANN) for seedlings height estimation of the *Tibouchina granulosa* for urban landscape

RNA	Phases	$RMSE_{\%}$	$Bias_{\%}$	Relative Errors (%)			*t* test
				Maximum	Mean	Minimum	*p*
MLP	Training	8.21	-0.15	29.46	0.85	-15.14	0.8902
	Test	9.94	3.20	12.80	-2.16	-11.17	0.3989
	Validation	8.89	-0.52	29.15	2.33	-11.40	0.8818
RBF	Training	9.08	0.00	43.45	0.96	-16.74	0.9993
	Test	11.93	0.55	24.78	1.07	-12.72	0.9060
	Validation	16.55	-5.17	73.60	9.96	-3.76	0.4141

MLP = *Multilayer perceptron*. RBF = *Radial Basis Function*.

The networks had few noises as they assumed as outliers data lines that, after processing, presented estimate of H higher than 2.0 units of standard deviation relative to the corresponding observed data (Fig. 1). The MLP network showed homogeneity of variances. The estimated and observed data for the selected networks are shown in Figure 2. The networks generated statistically similar estimates. Although we observed a disadvantage with the loss in accuracy as the RCD of the seedlings decreased.

Figure 1. Representation of the residual dispersion percentage due to root collar diameter (RCD) and error classes for Artificial Neural Networks (ANN) constructed to estimate seedlings height of *Tibouchina granulosa* for urban landscape. MLP = Multilayer perceptron. RBF = Radial Basis Function

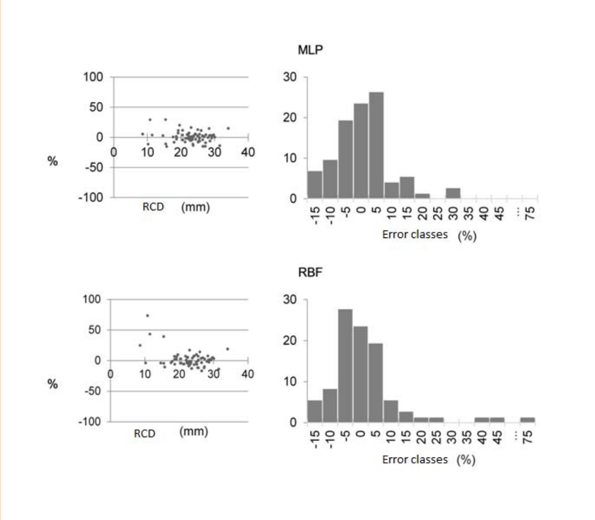
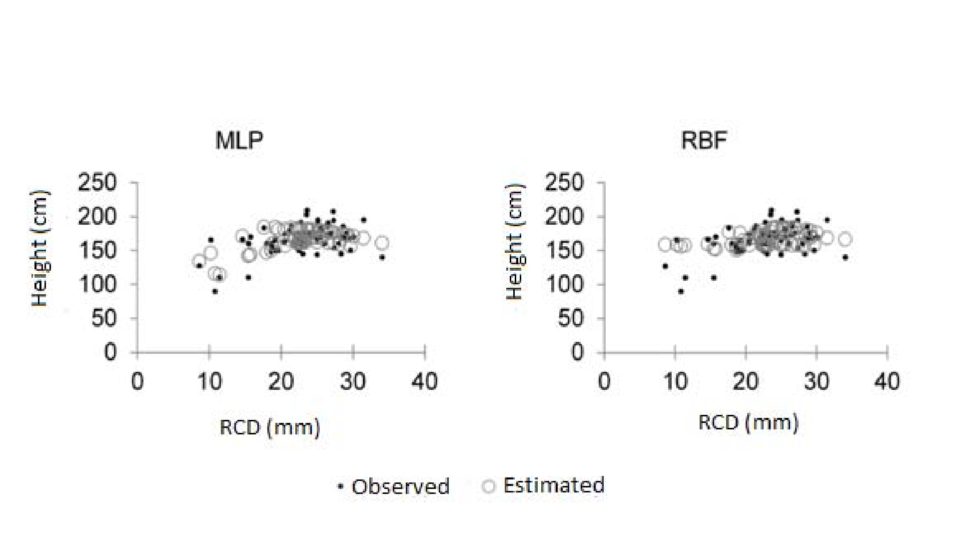


Figure 2. Seedlings height estimation of *Tibouchina granulosa* for urban landscape by MLP network. MLP (Multilayer perceptron) and RBF (Radial Basis Function)



DISCUSSION

The observed non-linear trend in both networks (Table 1) provided capacity to the successive layers to solve higher-order problems in input space (BRAGA et al., 2007). The MLP and RBF networks generated global and local receptive fields, respectively. The RBF network expressed greater complexity vis-à-vis the largest number of neurons in the medium layer. The definition of the architecture of networks, that is, the number of layers and the number of neurons per layer are optimized in the IPS tool.

Due to complexity to relate the diameter and height of seedlings under different substrate compositions, the precision of the networks was considered satisfactory. Although the MLP network showed a good training and a worse validation, the RMSE and Bias varied slightly between the processing phases and the relative range errors were low (Table 2). Therefore, this performance was not necessarily generated by an excessive memorization of training data. This corroborates with the lack of statistical significance in the paired *t*-test and the use of a limited number of cycles in the network. We applied to early interruption technique and standardization of data as heuristics. The heuristic allows an approximation of the optimal solution (SOARES et al., 2011; STATISTICA, 2007). However, the small number of cycles may have compromised the performance and computational capacity of the RBF network, featuring an underfitting (HAYKIN, 2001), because RBF showed greater variation between the processing phases (training, test, and validation).

The MLP network presented a simpler architecture (fewer neurons in the medium layer), fewer deviations and bias, expressing best learnability (Fig. 1). The MLP architecture builds global bonds, being composed of an input layer, one or more medium layers and an output layer (SOARES et al., 2011). Although few noises were observed, the capacity to deal with outliers in the adjusting process of its weights through a learning algorithm was proven (Fig. 2). These deviations were evenly distributed between the observations.

A disadvantage observed was the loss in accuracy as the RCD of the seedlings decreased, mainly in the RBF network. The total variance of experimental trials is attributed in part to factors controlled by known and independent causes. Besides, other uncontrolled factors of random nature, therefore they are not flawless (LAFETÁ et al., 2012).

The network with MLP architecture showed capacity to capture the realism of biological data, learn and generalize the knowledge assimilated to a set of unknown data non-used during training, or test and validation sets (Fig. 2).

The capacity of generalization and connectivity of the MLP architecture allowed to use only one network to perform simultaneous estimation of the height of seedlings of *Tibouchina granulosa* regardless of substrate composition studied. On the contrary, the use of traditional methods would involve carrying out analysis of individual regression of the height for each substrate composition.

The height of seedlings is essential for planning urban afforestation and the statistical method proposed in this study provided estimates. This method, in addition to making the obtaining these related forest biometry results faster, is less costly and less taxing because of the possibility of using diametric data regardless of the substrate composition used.

The use of MLP architecture in RNA can be recommended to estimate the height of seedlings of *Tibouchina granulosa* from RCD, regardless of the substrate composition the evaluated in this research. This methodology can provide important benefits for the estimation of seedling heights of other forest species, native or exotic, applying RNA.

CONCLUSION

Artificial neural networks modeling with Multilayer Perceptron (MLP) architecture proved to be appropriate and accurate to estimate the height of seedlings of *Tibouchina granulosa* from the diameter of the collar, regardless of the substrate composition.

ACKNOWLEDGMENTS

The authors wish to thank the Instituto Federal de Educação, Ciência e Tecnologia (IFMG) - *Campus* de São João Evangelista-MG and the Companhia Energética de Minas Gerais (CEMIG) for all the logistic and financial support to carry out the current study.

REFERENCES

BINOTI, D. H. B.; LEITE, H. G.; NOGUEIRA, G. S.; SILVA, M. L. M.; GARCIA, S. L. R.; CRUZ, J. P. Uso da função Weibull de três parâmetros em um modelo de distribuição diamétrica para plantios de eucalipto submetidos a desbaste. **Revista Árvore**, Viçosa, v. 34, n.1, p. 147-156, 2010.

BINOTI, M. L. M. S. **Redes neurais artificiais para prognose da produção de povoamentos não desbastados de eucalipto**. Viçosa, 2010. 54f. Dissertação (Mestrado em Ciência Florestal) – Universidade Federal de Viçosa, Viçosa, 2010.

BLACKARD, J. A.; DEAN, D. J. Comparative accuracies of artificial neural networks and discriminant analysis in predicting forest cover types from cartographic variables. **Computers and Electronics in Agriculture**, Amsterdam, v. 24, p. 131-151, 1999.

BRADSHAW, C. J. A.; DAVIS, L. S.; PURVIS, M.; ZHOU, Q.; BENWELL, G. L. Using artificial neural networks to model the suitability of coastline for breeding by New Zealand fur seals (*Arctocephalus forsteri*). **Ecological Modelling**, Amsterdam, v. 148, p. 111-131, 2002

BRAGA, A. P.; CARVALHO, A. C. P. L. F.; LUDERMIR, T. B. **Redes neurais artificiais**: Teoria e aplicações. 2 ed. Rio de Janeiro: LTC, 2007. 226 p.

BRAGA, F. A.; BARROS, N. F.; SOUZA, A. L.; COSTA, L. M. Características ambientais determinantes da capacidade produtiva de sítios cultivados com eucalipto. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 23, n.2 p.291-298, 1999.

CAMPOS, J. C. C.; LEITE, H. G. **Mensuração Florestal**. 3 ed., Viçosa: Editora UFV, 261 p., 2009.

CERQUEIRA, E. O.; ANDRADE, J. C.; POPPI, R. J. Redes neurais e suas aplicações em calibração multivariada. **Química Nova**, São Paulo, v. 24, n. 6, p. 864-873, 2001.

CÉZAR, T. M.; SOUZA, F. C.; MACIEL, R. T.; DEMBISKI, W.; ZUFFELLATO-RIBAS, K. C. RIBAS; L. L. F.; KOEHLER, H. S. Estaquia e alporquia de *Tibouchina fothergillae* (D. C.) Cogn. (Melastomataceae) com a aplicação de ácido naftaleno acético. **Revista Scientia Agraria**, Curitiba, v. 10, n. 6, p. 463-468, 2009.

FERNANDES, A. M.; UTKIN, A. B.; LAVROV, A. V.; VILAR, R. M. Development of neural network committee machines for automatic forest fire detection using lidar. **Pattern Recognition**, Amsterdam, v. 37, p. 2039-2047, 2004.

GONÇALVES, E. O.; PAIVA, H. N.; GONÇALVES, W.; JACOVINE, L. A. G. Avaliação qualitativa de mudas destinadas à arborização urbana no estado de Minas Gerais. **Revista Árvore**, Viçosa, v. 28, n. 4, p. 479-486, 2004.

GORGENS, E. B.; LEITE, H. G.; SANTOS, H. N.; GLERIANI, J. M. Estimação do volume de árvores utilizando redes neurais artificiais. **Revista Árvore**, Viçosa, v. 33, n. 6, p. 1141-1147, 2009.

HAYKIN, S. **Redes Neurais**: Princípios e prática. 2 ed. Porto Alegre: Bookman, 2001; 900 p.

LAFETÁ, B. O.; SANTANA, R. C.; NOGUEIRA, G. S.; PENIDO, T. M. A.; COUTO, L. Área foliar específica e clorofila em eucalipto estimadas por redes neurais artificiais. **Biomassa & Energia**, Viçosa, v. 5, n. 1, p. 67-75, 2012.

LOPES, J. C.; DIAS, P. C.; PEREIRA, M. D. Maturação fisiológica de sementes de quaresmeira. **Pesquisa Agropecuária Brasileira**, Brasília, v. 40, n. 8, p. 811-816, 2005.

MAEDA, E. E.; FORMAGGIO, A. R.; SHIMABUKURO, Y. E.; ARCOVERDE, G. F. B.; HANSEN, M. C. Predicting forest fire in the Brazilian Amazon using MODIS imagery and artificial neural networks. **International Journal of Applied Earth Observation and Geoinformation**, Amsterdam, v. 11, p. 265-272, 2009.

ROCHA, R. T.; LELES, P. S. S.; OLIVEIRA NETO, S. N. Arborização de vias públicas em nova iguaçu, RJ: o caso dos bairros rancho novo e centro. **Revista Árvore**, Viçosa, 28, n. 4, p. 599-607, 2004.

ROSSATTO, D. R.; TSUBOY, M. S. F.; FREI, F. Arborização urbana na cidade de Assis-sp: uma abordagem quantitativa. **Revista Sociedade Brasileira de Arborização Urbana**, Piracicaba, v. 3, n. 3, p. 1-16, 2008.

SCRINZI, G.; MARZULLO, L.; GALVAGNI, D. Development of a neural network model to update forest distribution data for managed alpine stands. **Ecological Modelling**, Amsterdam, v. 206, p. 331-346, 2007.

SOARES, F. A. A. M. N.; FLÔRES, E. L.; CABACINHA, C. D.; CARRIJO, G. A.; VEIGA, A. C. P. Recursive diameter prediction and volume calculation of eucalyptus trees using Multilayer Perceptron Networks. **Computers and Electronics in Agriculture**, Amsterdam, v. 78, p. 19–27, 2011.

STATSOFT, INC. Statistica (data analysis software system), version 7. Disponívelem: <http://www.statsoft.com>, 2007.

ZAMPIERI, M. C. T.; SARKIS, J. E. S.; PESTANA, R. C. B.; TAVARES, A. R.; MELO-DE-PINNA, G. F. A. Characterization of *Tibouchina granulosa* (Desr.) Cong. (Melastomataceae) as a biomonitor or air pollution and quantification of particulate matter adsorbed by leaves. **Ecological Engineering**, Amsterdam, v. 61, p. 316-327, 2013.