9.
Referências Bibliográficas

ABBEólica (2015). Boletim de Dados Jan. 2015. Disponível em: <http://www.abeeolica.org.br/pdf/Boletim-de-Dados-ABEEolica-Janeiro-2015-Publico.pdf> Acesso em Janeiro de 2014.

ABBEólica (2014). Boletim de Dados Nov. 2014. Disponível em: <http://www.abeeolica.org.br/pdf/Boletim-de-Dados-ABEEolica-Novembro-2014-Publico.pdf> Acesso em Dezembro de 2014.

ABEEólica (2012). Boletim Anual De Geração Eólica – 2012. Disponível em: <http://www.abeeolica.org.br/novo-site/index.php/dados.html > Acesso em: Agosto, 2013.

Agrawal, R., Gehrke, J., Gunopulos, D., & RAGHAVAN, P. (1998). Automatic Subspace Clustering of High Dimensional Data for Data Mining Applications. *IBM Almaden Research Center*.

Ailliot, P., Monbet, V., & Prevosto, M. (2006). An autoregressive model with time-varying coefficients for wind fields. *Environmetrics*, *17*(2), 107–117.

Ailliot, P., & Monbet, V. (2012). Markov-switching autoregressive models for wind time series. *Environmental Modelling & Software*, *30*, 92–101.

Al-Buhairi, M. H., & Al-Haydari, A. (2012). Monthly and Seasonal Investigation of Wind Characteristics and Assessment of Wind Energy Potential in Al-Mokha, Yemen. *Energy and Power Engineering*, *04* (May), 125–131.

Al-Yahyai, S., Gastli, A., & Charabi, Y. (2012). Probabilistic wind speed forecast for wind power prediction using pseudo ensemble approach. In *2012 IEEE International Conference on Power and Energy (PECon)*. Kota Kinabalu Sabah, Malaysia, (pp. 127-132).

Alexiadis, M. C., Dokopoulos, P. S., Sahsamanoglou, H. S., & Manousaridis, I. (1998). Short-term forecasting of wind speed and related electrical power. *Solar Energy, 63*, 61-68

Alexiadis, M. C., Dokopoulos, P. S., & Sahsamanoglou, H. S. (1999). Wind speed and power forecasting based on spatial correlation models. *IEEE, Transactions on Energy Conversion, 14*(3), 836–842.

Amarante, O. A. C., Brower, M., Zack, J., & Sá, A. L. (2001). Atlas do Potencial Eólico Brasileiro. Centro de Pesquisas de Energia Elétrica, Ed. CEPEL, Brasília.

Anastasiades, G & McSharry, P. (2013). Quantile Forecasting of Wind Power Using Variability Indices. *Energies, 6*, 662-695.

ANEEL (2005). Atlas de energia elétrica do Brasil. 2ª Edição- Brasília, 243p. ISBN: 85-87491-09-1.

Azzalini, A., & Genton, M.G. (2008). Robust likelihood methods based on the skew-t and related distributions. *International Statistical Review*, *76*(1), 106–129.

Balouktsis, A., Tsanakas, D., & Vachtsevanos, G. (1986). Stochastic simulation of hourly and daily average wind speed sequences. *Wind Engineering*, *10*(1), 1-11.

Barbounis, T. G., Theocharis, J. B., Alexiadis, M. C., & Dokopoulos, P. S. (2006). Long-term wind speed and power forecasting using local recurrent neural network models. *IEEE, Transactions on Energy Conversion*, *21*(1), 273–284.

Bashtannyk, D. M., & Hyndman, R. J. (2001). Bandwidth selection for kernel conditional density estimation. *Computational Statistics and Data Analysis*, 36, 279-298.

Beneki, C., Eeckels, B., & Leon, C. (2009). Signal extraction and forecasting of the UK tourism income time series. A singular spectrum analysis approach, MPRA paper no. 18354. Online at http://mpra.ub.uni-muenchen.de/18354/1/MPRA_paper_18354.pdf

Bessa, R. J., Sumaili, J., Miranda, V., Botterud, A., Wang, J., & Constantinescu, E. (2011a). Time-adaptive kernel density forecast: A new method for wind power

uncertainty modeling. In *17th Power Systems Computation Conference*. Stockholm, Sweden.

Bessa, R. J., Mendes, J., Miranda, V., Botterud, A., Wang, J., & Zhou, Z. (2011b). Quantile-copula density forecast for wind power uncertainty modeling. In *2011 IEEE Trondheim PowerTech*, Trondheim, Norway, (pp. 1–8).

Bessa, R. J., Miranda, V., Botterud, A., Zhou, Z., & Wang, J. (2012a). Time-adaptive quantile-copula for wind power probabilistic forecasting. *Renewable* Energy, *40*(1), 29–39.

Bessa, R. J., Miranda, V., Botterud, A., Wang, J., & Constantinescu, E. M. (2012b). Time adaptive conditional kernel density estimation for wind power forecasting. *IEEE Transactions on Sustainable Energy*, *3*(4), 660–669.

Beyer, H.G., Degner, T., Hausmann, J., Hoffmann, M., & Rujan P. (1994). Short-term prediction of wind speed and power output of a wind turbine with neural networks. In *Proceedings of the 2nd European Congress on Intelligent Techniques and Soft Computing*. Aachen, Germany, (pp. 349–352).

Blanchard, M. & Desrochers G. (1984). Generation of autocorrelated wind speeds for wind energy conversion system studies. *Solar Energy*, *33*, 571-579.

BNDES (2012). Apoio do BNDES a concessões e PPPS em infraestrutura. Disponível em: <http://www.sinaenco.com.br/downloads/BNDES%20Zurli.pdf> Acesso em Outubro de 2014.

Boccard, N. (2009). Capacity factor of wind power realized values vs. estimates. *Energy Policy*, *37*, 2679–2688.

Bossanyi, E. A (1985). Short-term wind prediction using Kalman filters. *Wind Engineering*, *9*(1), 1-8.

Botterud, a, Wang, J., Bessa, R. J., Keko, H., & Miranda, V. (2010). Risk management and optimal bidding for a wind power producer. *IEEE PES General Meeting*. Minneapolis, MN, USA, (pp. 1–8).

Botterud, A., Zhou, Z., Wang, J., Valenzuela, J., Sumaili, J., Bessa, R. J., Keko, H., & Miranda, V. (2011a). Unit commitment and operating reserves with probabilistic wind power forecasts. *PowerTech, 2011 IEEE Trondheim*. Trondheim, (pp. 1–7).

Botterud, A., Zhou, Z., Wang, J., Bessa, R. J., Keko, H., Mendes, J., Sumaili, J., & Miranda, V. (2011b). Use of Wind Power Forecasting in Operational Decisions. *Argonne National Laboratory* ANL/DIS-11-8.

Botterud, A., Sumaili, J., Keko, H., Mendes, J., Bessa, R., & Miranda, V. (2013). Demand dispatch and probabilistic wind power forecasting in unit commitment and economic dispatch: A case study of Illinois. *2013 IEEE Power & Energy Society General Meeting*, *4*(1), 1–1.

Bourlis, D., & Bleijs, J. (2010). A wind speed estimation method using adaptive Kalman filtering for a variable speed stall regulated wind turbine. In *2010 IEEE 11th International Conference on Probabilistic Methods Applied to Power Systems*. Singapore, (pp. 89–94).

Box, G. E. P., & Jenkins, G. M. (1970). Time Series Analysis: Forecasting and Control, Holden–Day Inc., San Francisco.

Bremnes, J.B. (2004). Probabilistic Wind Power Forecasts using Local Quantile Regression. *Wind Energy*, *7*(1), 47-54.

Bremnes, J.B. (2006). A Comparison of a Few Statistical Models for Making Quantile Wind Power Forecasts. *Wind Energy*, *9*(1-2), 3-11.

Broomhead, D. S., & King, G. P. (1986a). Extracting qualitative dynamics from experimental data. *Physica D*, *20*, 217-236.

Broomhead, D. S., & King, G. P. (1986b). On the qualitative analysis of experimental dynamical systems. *Nonlinear Phenomena and Chaos*, 113-144.

Broomhead, D. S., Jones, R., King, G. P., & Pike, E. R. (1987). Singular system analysis with application to dynamical systems. In E.R. Pike and L.A. Lugaito, editors, *Chaos, Noise and Fractals* (pp. 15–27). IOP Publishing, Bristol,

Brown, B. G., Katz, R. W. & Murphy A. H. (1984). Time Series Models to Simulate and Forecast Wind Speed and Wind Power. *Journal of Climate and Applied Meteorology*, *23*(8), pp. 1184-1195.

Cacoullos, T. (1966). Estimation of a multivariate density. *Annals of the Institute of Statistical Mathematics*, *18*(1), 179-189.

Cadenas, E., & Rivera, W. (2007). Wind speed forecasting in the South Coast of Oaxaca, México. *Renewable Energy*, *32*(12), 2116–2128.

Cadenas, E., & Rivera, W. (2009). Short term wind speed forecasting in La Venta, Oaxaca, México, using artificial neural networks. *Renewable Energy*, *34*(1), 274–278.

Cadenas, E., & Rivera, W. (2010). Wind speed forecasting in three different regions of Mexico, using a hybrid ARIMA–ANN model. *Renewable Energy*, *35*(12), 2732–2738.

Caporin, M., & Preś, J. (2012). Modelling and forecasting wind speed intensity for weather risk management. *Computational Statistics and Data Analysis*, *56*(11), 3459–3476.

Carpenter, G. A., & Grossberg, S. (1988). The ART of adaptive pattern recognition by a self-organizing neural network. *Computer*, *21*(3), 77–88.

Carpinone, A., Langella, R., Testa, A., & Giorgio, M. (2010). Very short-term probabilistic wind power forecasting based on Markov chain models. *Probabilistic Methods Applied to Power Systems (PMAPS), IEEE 11th International Conference on*. Singapore, (pp. 107-112).

Carta, J. A., & Ramírez, P. (2007). Use of finite mixture distribution models in the analysis of wind energy in the Canarian Archipelago. *Energy Conversion and Management*, *48*(1), 281–291.

Carta, J. A., Ramírez, P., & Velazquez, S. (2008). Influence of the level of fit of a density probability function to wind-speed data on the WECS mean power output estimation. *Energy Conversion and Management*, *49*(10), 2647–2655.

Carta, J. A., Ramírez, P., & Velazquez, S. (2009). A review of wind speed probability distributions used in wind energy analysis: case studies in the Canary Islands. *Renewable and Sustainable Energy Reviews*. *13*(5), 933–955.

Carvalho, M., Rodrigues, P.C., & Rua, A. (2012). Tracking the U.S. business cycle with a singular spectrum analysis. *Economics Letters*, *114*, 32–35.

Castino, F., Festa, R., & Ratto, C.F. (1998). Stochastic modelling of wind velocities time series. *Journal of Wind Engineering and Industrial Aerodynamics*, *74–76*, 141–151

Cassiano, K. M. (2014). Análise de séries temporais usando Análise Espectral Singular (SSA) e clusterização de suas componentes baseada em densidade. *Tese de doutorado do Programa de Pós Graduação em Engenharia Elétrica*, PUC-Rio, Rio de Janeiro, 172 p.

Catalão, J. P. S., Pousinho, H. M. I., & Mendes, V. M. F. (2011). Short-term wind power forecasting in Portugal by neural networks and wavelet transform. *Renewable Energy*, *36*(4), 1245–1251.

Cencov, N. N. (1962). Evaluation of an unknown density from observations. *Soviet Mathematics*, *3*, 1559-1562.

Cleveland, W. S. (1979). Robust LocallyWeighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association*, *74*, 829-836.

Cole, R. M. (1998). *Clustering with genetic algorithms*. M. Sc., thesis, Department of Computer Science, University of Western Australia, Perth, Australia.

Costa, A., Crespo, A., Navarro, J., Lizcano, G., Madsen, H., & Feitosa, E. (2008). A review on the young history of the wind power short-term prediction, *Renewable and Sustainable Energy Reviews*, *12*(6), 1725-1744.

Crochet, P. (2004). Adaptive Kalman filtering of 2-metre temperature and 10-metre wind-speed forecasts in Iceland. *Meteorological Applications*, *11*(2), 173–187.

Custódio , R. S. (2009). Energia Eólica para Produção de Energia Elétrica. Eletrobras. Rio de Janeiro.

Damousis, I.G. & Dokopoulos, P. (2001). A fuzzy model expert system for the forecasting of wind speed and power generation in wind farms. In *Proceedings of the IEEE International Conference on Power Industry Computer Applications PICA*. Sydney, NSW, (pp. 63- 69).

Damousis, I.G., Alexiadis, M.C., Theocharis, J.B. & Dokopoulos, P. (2004). A fuzzy model for wind speed prediction and power generation in wind farms using spatial correlation. *IEEE, Transactions on Energy Conversion*, 19, 352–361.

Daniel, A. & Chen, A. (1991). Stochastic simulation and forecasting of hourly wind speed sequences in Jamaica. *Solar Energy*, *46*, 1–11.

Danilov, D. L. (1997a). Principal components in time series forecast. *Journal of Computational and Graphical Statistics*, *6*, 112-121.

Danilov, D. L. (1997b). The Caterpillar method for time series forecasting. *Em Principal Components of Time Series: The Caterpillar Method* (Editado por D. Danilov e A. Zhigljavsky), 73-104. Saint Petersburg State University, Saint Petersburg, Russia.

De Giorgi, M. G., Ficarella, A., & Tarantino, M. (2011). Error analysis of short term wind power prediction models. *Applied Energy*, *88*(4), 1298–1311. doi:10.1016/j.apenergy.2010.10.035

De Luna, X., & Genton, M.G. (2005). Predictive spatio-temporal models for spatially sparse environmental data. *Statistica Sinica*, *15*, 547–568.

Dutra, R. (2008). Tutorial de Energia Eólica: Princípios e Tecnologia. Centro de Referência para Energia Solar e Eólica Sérgio de Salvo Brito (CRESESB) Cepel. Disponível em: <http://www.cresesb.cepel.br/download/tutorial/tutorial_eolica_2008_e-book.pdf > Acesso em: Julho de 2013.

EIA (2013). International Energy Outlook 2013. Disponível em: <http://www.eia.gov/forecasts/ieo/pdf/0484(2013).pdf> Acesso em Julho de 2013.

Ecopolítica (2013). China produz mais eólica que nuclear. Disponível em: <http://www.ecopolitica.com.br/2013/07/02/china-ja-produz-mais-energia-eolica-que-nuclear/> Acesso em: Agosto, 2013.

EERE (2013). Wind and Water Program: How Wind Turbines Work. Disponível em: <http://www1.eere.energy.gov/wind/wind_how.html> Acesso em: Fevereiro 2013.

El-Fouly, T. H. M., El-Saadany, E. F., & ASalama, M. M. (2006). One day ahead prediction of wind speed using annual trends. In *Proceedings of the Power Engineering Society General Meeting* (pp. 1–7).

ELSAM (1993). Wind power prediction tool in central dispatch centers. Final report, EU Project JOU2-CT92-0083.

Epanechnikov, V. A. (1969). Non-parametric estimation of a multivariate probability density. *Theory of Probability and its Applications*, *14*(1), 153-158.

EPE (2012). Plano Decenal de Expansão de Energia 2021. Disponível em: <http://www.epe.gov.br/PDEE/20130326_1.pdf > Acesso em: Agosto, 2013.

Erdem, E., & Shi, J. (2011). ARMA based approaches for forecasting the tuple of wind speed and direction. *Applied Energy*, *88*(4), 1405–1414.

Esquivel, R. M. (2012). Análise Espectral Singular: modelagem de séries temporais através de estudos comparativos usando diferentes estratégias de previsão. 2012. 174 f. Dissertação (Mestrado em Modelagem Computacional e Tecnologia Industrial) - SENAI CIMATEC, Salvador.

Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)* (pp. 226–231).

Ewing, B. T., Kruse, J. B., Schroeder, J. L., & Smith, D. A. (2007). Time series analysis of wind speed using VAR and the generalized impulse response technique. *Journal of Wind Engineering and Industrial Aerodynamics*, *95*(3), 209–219.

Fan, J., Yao, Q., & Tong, H. (1996). Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika*, *83*(1), 189-206.

Fan, J., & Yim, T. H. (2004). A cross-validation method for estimating conditional densities. *Biometrika, 91*(4), 819-834.

Fausett, L. V. (1994). Fundamentals of neural networks: architectures, algorithms, and applications (Vol. 40). Englewood Cliffs: Prentice-Hall.

Fix, E., & Hodges, J. L. Jr. (1951). Nonparametric discrimination: consistency properties. Report Number 4, USAF School of Aviation Medicine, RandoIph Field, Texas.

Foley, A. M., Leahy, P. G., Marvuglia, A., & McKeogh, E. J. (2012). Current methods and advances in forecasting of wind power generation. *Renewable Energy*, *37*(1), 1–8.

Geerts, H. (1984). Short range prediction of wind speeds: a system-theoretic approach. In *Proceedings of European wind energy conference*. Hamburg, Germany, (pp. 594–599).

Giebel, G. (2001). On the benefits of distributed generation of wind energy in Europe. Ph.D. thesis from the Carl von Ossietzky Universität Oldenburg, Düsseldorf.

Giebel, G., Landberg, L., Kariniotakis, G. & Brownsword R. (2003). State-of-the-Art on methods and software tools for short-term prediction of wind energy. In *Proceedings of the European wind energy conference & exhibition, EWEC2003,* Madrid, Spain.

Giebel, G., Brownsword, R., Kariniotakis, G., Denhard, M., & Draxl, C. (2011). The State of the Art in Short Term Prediction of Wind Power A Literature Overview, 2ed. Project report for the Anemos.plus and Safe Wind projects. Risø, Roskilde, Denmark.

Gneiting, T., Larson, K., Westrick, K., Genton, M., Aldrich, E. (2006). Calibrated probabilistic forecasting at the Stateline Wind Energy Center: The regime- switching space-time method. *Journal of the American Statistical Association*, *101* (475), 968–979.

Gneiting, T. (2011a). Quantiles as Optimal Point Forecasts. *International Journal of Forecasting*, *27*, 197-207.

Gneiting, T. (2011b). Making and Evaluating Point Forecasts. *Journal of the American Statistical Association, 106*, 746-762.

Golyandina, N., Nekrutkin, V., & Zhigljavsky, A. (2001). Analysis of Time Series Structure: SSA and Related Techniques. Chapman & Hall/CRC, New York.

Golyandina, N., & Zhigljavsky, A. (2013). Singular Spectrum Analysis for Time Series. New York: Springer.

GWEC (2011). Analysis of the regulatory framework in Brazil 2011. Disponível em: <http://gwec.net/wpcontent/uploads/2012/06/2ANALISE_DO_MARCO_REGULATORIO_
PARA_GERACAO_EOLICA_NO_BRASIL.pdf> Acesso em Abril de 2013.

GWEC (2012a). Global Wind Energy Outlook - 2012. Disponível em: <http://www.gwec.net/wp-content/uploads/2012/11/GWEO_2012_lowRes.pdf> Acesso em Abril de 2013.

GWEC (2012b). Global Wind Report 2012 - Annual market update 2012. Disponível em: <http://www.gwec.net/wp-content/uploads/2012/06/Annual_report_2012_LowRes.pdf> Acesso em Abril de 2013.

GWEC (2013). Global Wind Report 2012 - Annual market update 2013. Disponível em: <http://www.gwec.net/wp-content/uploads/2014/04/GWEC-Global-Wind-Report_9-April-2014.pdf> Acesso em Julho de 2014.

GWEC (2014). Global Wind Energy Outlook 2014. Disponível em: <http://www.gwec.net/wp-content/uploads/2014/10/GWEO2014_WEB.pdf> Acesso em Novembro de 2014.

Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). On clustering validation techniques. *Journal of Intelligent Information Systems*, *17*(2-3), 107-145.

Hall, P., Wolff, R. C. L., & Yao, Q. (1999). Methods for estimating a conditional distribution function. *Journal of the American Statistical Association*, *94*(445), 154-163.

Hall, P., Racine, J., & Li, Q. (2004). Cross-validation and the estimation of conditional probability densities. *Journal of the American Statistical Association*, *99*(468), 1015-1026.

Han, J., & Kamber, M. (2001). Cluster Analysis. In: Morgan Publishers (eds.), *Data Mining: Concepts and Techniques*, 1 ed., chapter 8, NewYork, USA, Academic Press.

Hassani, H. (2007). Singular spectrum analysis: methodology and comparison. *Journal of Data Science*, *5*, 239–257.

Hassani, H., Heravi, S., & Zhigljavsky, A. (2009). Forecasting European industrial production with singular spectrum analysis. *International Journal of Forecasting*, 25, 103–118.

Hassani , H., & Thomakos, D. (2010). A review on singular spectrum analysis for economic and financial time series. *Statistics and Its Interface*, *3*, 377–397.

Hering, A. S., & Genton, M. G. (2010). Powering up with space-time wind forecasting. *Journal of the American Statistical Association*, *105*(489), 92–104.

Huang, Z., & Chalabi, Z. S. (1995). Use of time-series analysis to model and forecast wind speed. *Journal of Wind Engineering and Industrial Aerodynamics*, *56*, 311–322.

Hussain, S., Elbergali, A., Al-Masri, A., & Shukur, G. (2004). Parsimonious modeling, testing and forecasting of long-range dependence in wind speed. *Environmetrics*, *15*(2), 155–171.

Hyndman, R. J., Bashtannyk, D. M., & Grunwald, G. K. (1996). Estimating and visualizing conditional densities. *Journal of Computational and Graphical Statistics*, *5*(4), 315-336.

Hyndman, R. J., & Yao, Q. (2002). Nonparametric estimation and symmetry tests forconditional density functions. *Nonparametric Statistics*, *14*, 259-278.

Jangamshetti, S. H., & Rau, V. G. (1999). Site matching of wind turbine generators: A case study. *IEEE Transactions on Energy Conversion*, *14*(4), 1537–1543.

Jeon, J., & Taylor, J. W. (2012). Using Conditional Kernel Density Estimation for Wind Power Density Forecasting. *Journal of the American Statistical Association*, *107*(497), 66–79.

Ji, G.R., Han, P., & Zhai, Y. (2007). Wind speed forecasting based on support vector machine. In *Proceedings of the Sixth International Conference on Machine Learning and Cybernetics*. Hong Kong, (pp. 2735-2739).

Jiang, W., Yan, Z., Feng, D. H., & Hu, Z. (2012). Wind speed forecasting using autoregressive moving average / generalized autoregressive conditional heteroscedasticity model. *European Transactions On Electrical Power*, *22*, 662-673.

Johnson, P., Negnevitsky, M., & Muttaqi, K. M. (2007). Short term wind power forecasting using adaptive neuro-fuzzy inference systems. In *Power Engineering Conference, 2007. AUPEC 2007. Australasian Universities*. Perth, WA, (pp. 1–6).

Johnson, R., A. & Wichern, D. W. (2002). Applied Multivariate Statistical Analysis (5th Ed.). Upper Saddle River, N.J., Prentice-Hall.

Jónsson, T., Pinson P., Madsen, H., & Nielsen, H. A. (2013). Predictive Densities for Day-Ahead Electricity Prices Using Time-Adaptive Quantile Regression. Working Paper, Preprint submitted to Applied Energy.

Juban, J., Siebert, N., & Kariniotakis, G. N. (2007a). Probabilistic Short-term Wind Power Forecasting for the Optimal Management of Wind Generation. In *2007 IEEE Lausanne Power Tech* (pp. 683–688).

Juban, J., Fugon, L., & Kariniotakis, G. (2007b). Probabilistic short-term wind power forecasting based on kernel density estimators. *In Probabilistic wind power forecasting - European Wind Energy Conference*. Milan, Italy, (pp. 1–11).

Kamal, L., & Jafri, Y. Z. (1997). Time series models to simulate and forecast hourly averaged wind speed in Quetta, Pakistan. *Solar Energy*, *61*(1), 23–32.

Kariniotakis, G., Stavrakakis, G., & Nogaret, E. (1996a). Wind power forecasting using advanced neural network models. *IEEE, Transactions on Energy Conversion*, *11*(4), 762-767.

Kariniotakis, G., Stavrakakis, G., & Nogaret, E (1996b). A fuzzy logic and neural network based wind power model. In *Proceedings of the 1996 European Wind Energy Conference*. Goteborg, Sweden, (pp. 596-599).

Kariniotakis, G. N., Nogaret, E., & Stavrakis, G. (1997). Advanced Short-Term Forecasting of Wind Power Production. In *Proceedings of the European Wind Energy Conference*. Dublin, Ireland, (pp. 751-754).

Kariniotakis, G. N., Nogaret, E., Dutton, A.G., Halliday, J.A., & Androutsos, A. (1999). Evaluation of Advanced Wind Power and Load Forecasting Meghods for the Optimal Management of Isolated Power Systems. In *Proceedings of the European Wind Energy Conference*. Nice, France, (pp. 1082-1085).

Kariniotakis G.N., Marti I., Nielsen T.S., Giebel G., Tambke J., Waldl I., Usaola J., Brownsword, R., Kallos, G., Focken, U., Sanchez, I., Hatziargyriou,, N., Palomares, Frayssinet, A. M., & Frayssinet P. (2006a). Advanced Short-term Forecasting of Wind Generation – Anemos. *IEEE Trans, On Power Systems*. Contract No ENK5-CT-2002-00665, 2002-2006, 24 partners, 7 countries, European Commission under the FP5 R&D Project ANEMOS.

Kariniotakis, G., Pinson, P., Siebert, N., Giebel, G., & Barthelmie R. (2006b). State of the art in short-term prediction of wind power – From an offshore perspective. In *European Wind Energy Conference*, Atenas, Grécia.

Karunamuni, R.J., & Alberts, T. (2005). On boundary correction in kernel density estimation, *Statistical Methodology*, *2*(3), 191–212.

Kavasseri, R. G., & Seetharaman, K. (2009). Day-ahead wind speed forecasting using f-ARIMA models. *Renewable Energy*, *34*(5), 1388–1393.

Kennedy, S., & Rogers, P. (2003). A Probabilistic Model for Simulating Long-Term Wind-Power Output. *Wind Engineering*, *27*(3), 167–181.

Kohonen, T. (1989). Self-organization and associative memory. 3rd Edition. Springer-Verlag.

Kohonen, T. (1997). Self-organizing maps. 2nd Ed. Heidelberg: Springer.

Kollu, R., Rayapudi, S. R., Narasimham, SVL, & Pakkurthi, K. M. (2012). Mixture probability distribution functions to model wind speed distributions. *International Journal of Energy and Environmental Engineering*, *3*(1), 27.

Kou, P., Gao, F., & Guan, X. (2013). Sparse online warped Gaussian process for wind power probabilistic forecasting. *Applied Energy*, *108*, 410–428.

Kou, P., Liang, D., Gao, F., & Gao, L. (2014). Probabilistic wind power forecasting with online model selection and warped gaussian process. *Energy Conversion and Management*, *84*, 649–663.

Kretzschmar, R., Eckert, P., Cattani, D., & Eggimann, F. (2004). Neural Network Classifiers for Local Wind Prediction. *Journal of Applied Meteorology*, *43*(5), 727-738.

Kaufman L., Rousseeuw, P. J. (1990). Finding Groups in Data: an Introduction to Cluster Analysis, Wiley Inter-science, Canadá.

Lance, G. N., & Williams, W. T. (1967). A general theory of classificatory sorting strategies: 1. Hierarchical systems. *Computer Journal*, *9*(4), 373 – 380.

Landberg, L. (2001). Short-term prediction of local wind conditions. *Journal of Wind Engineering and Industrial Aerodynamics*, *89*(3-4), 235-245.

Lange, M. & Focken, U. (2008). New developments in wind energy forecasting. In *Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century, 2008 IEEE* (pp.1-8).

Lau, A. & McSharry, P. (2010). Approaches for multi-step density forecasts with application to aggregated wind power. *The Annals of Applied Statistics*, *4*(3), 1311–1341.

Lei, M., Shiyan, L, Chuanwen, J., Hongling, L., & Yan, Z. (2009). A review on the forecasting of wind speed and generated power. *Renewable and Sustainable Energy Reviews*, *13*(4), 915-920.

Lin, L., Eriksson, J.T., Vihriala, H., & Soderlund, L. (1996). Predicting wind behavior with neural networks. In *Proceedings of the 1996 European Wind Energy Conference*. Goteborg, Sweden, (pp. 655-658).

Linden, R. (2009). Técnicas de Agrupamento. *Revista de Sistemas de Informação da FSMA*, 4, 18 – 36.

Liu, D., Niu, D., Wang, H., & Fan, L. (2014). Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm. *Renewable Energy*, *62*, 592–597.

Liu, H., Erdem, E., & Shi, J. (2011). Comprehensive evaluation of ARMA–GARCH(-M) approaches for modeling the mean and volatility of wind speed. *Applied Energy*, *88*(3), 724–732.

Liu, H., Tian, H., & Li, Y. (2012). Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction. *Applied Energy*, *98*(0), 415–424.

Liu, Y., Yan, J., Han, S., & Peng, Y. (2012a). Uncertainty Analysis of Wind Power Prediction Based on Quantile Regression. In *Power and Energy Engineering Conference (APPEEC), 2012 Asia-Pacific* (pp. 1–4). Shanghai.

Loftsgaarden, D. O., & Quesenberry, C. P. (1965). A nonparametric estimate of a multivariate density function. *Annals of Mathematical Statistics*, *36*(3), 1049-1051.

Lojowska, A., Kurowicka, D., Papaefthymiou, G., & Sluis, L. (2010). Advantages of ARMA-GARCH wind speed time series modeling. In *2010 IEEE 11th International Conference on Probabilistic Methods Applied to Power Systems*. Singapore, (pp. 83–88).

Louka, P., Galanis, G., Siebert, N., Kariniotakis, G., Katsafados, P., Pytharoulis, I., & Kallos, G. (2008). Improvements in wind speed forecasts for wind power prediction purposes using Kalman filtering. *Journal of Wind Engineering and Industrial Aerodynamics*, *96*(12), 2348–2362.

Madsen, H. (1995). Wind Power Prediction Tool in Control Dispatch Centres. ELSAM, Skaerbaek, Denmark, ISBN 87-87090-25-2.

Mahmoudvand, R., Alehosseini, F., & Zokaei, M. (2013). Feasibility of singular spectrum analysis in the field of forecasting mortality rate. *Journal of Data Science, 11*, 851-866.

Malmberg, A., Holst, U. & Holst, J. (2005). Forecasting near-surface ocean winds with Kalman filter techniques. *Ocean Engineering*, *32*, 273–291.

Mathaba, T., Xia, X., & Zhang, J. (2012). Short-term Wind Power Prediction using Least-Square Support Vector Machines. In *Power Engineering Society Conference and Exposition in Africa (PowerAfrica), 2012 IEEE*. Johannesburg, (pp. 1-6).

Matos, M. A., & Bessa, R. J. (2011). Setting the operating reserve using probabilistic wind power forecasts. *IEEE, Transactions on Power Systems*, *26*(2), 594–603.

Methaprayoon, K., Yingvivatanapong, C., Lee, W.-J., & Liao, J. R. (2007). An Integration of ANN Wind Power Estimation Into Unit Commitment Considering the Forecasting Uncertainty. *IEEE, Transactions on Industry Applications*, *43*(6), 1441–1448.

MME (2014). Boletim: Energia Eólica no Brasil e Mundo. Disponível em: <http://www.mme.gov.br/documents/10584/1256600/Folder+Energia+Eolica.pdf/b1a3e78c-7920-4ae5-b6e8-7ba1798c5961> Acesso em Dezembro de 2014.

Mohandes, M. A., Rehman, S., & Halawani, T. O. (1998). A neural networks approach for wind speed prediction. *Renewable Energy*, *13*(3), 345–354.

Mohandes, M. a., Halawani, T. O., Rehman, S., & Hussain, A. a. (2004). Support vector machines for wind speed prediction. *Renewable Energy*, *29*(6), 939–947.

Monteiro, C., Bessa, R., Miranda, V., Botterud, A., Wang, J., & Conzelmann G. (2009). Wind Power Forecasting: State-of-the-Art 2009. *Argonne National Laboratory* ANL/DIS-10-1.

More, A., & Deo, M. C. (2003). Forecasting wind with neural networks. *Marine Structures*, *16*(1), 35–49.

Morettin, P. A., & Toloi, L.M.C. (2006). Análise de Séries Temporais, 2ª Ed. ABE. Projeto Fisher. Editora: Edgard Blucher.

Møller, J.K., Nielsen, H.A., & Madsen, H. (2008). Time-adaptive quantile regression. *Computational Statistics and Data Analysis*, *52*(3), 1292–1303.

Morgan, E. C., Lackner, M., Vogel, R. M., & Baise, L. G. (2011). Probability distributions for offshore wind speeds. *Energy Conversion and Management*, *52*(1), 15–26.

Ng, R. T., & Han, J. (1994). Efficient and Effective Clustering Methods for Spatial Data Mining. In *Proceedings of the 20th VLDB Conference*, Santiago-Chile, (pp. 144 – 155).

Nielsen, T.S., & Madsen, H. (1996). Using Meteorological Forecasts in On-line Predictions of Wind Power. ELSAM, Skaerbaek, Denmark.

Nielsen, H.A., Madsen, H., & Nielsen, T.S. (2006). Using Quantile Regression to Extend an Existing Wind Power Forecasting System with Probabilistic Forecasts. *Wind Energy*, *9*(1-2), 95-108.

ONS (2014). Boletim Mensal de Geração Eólica Novembro/2014. Disponível em: <http://www.ons.org.br/download/resultados_operacao/boletim_mensal_geracao_eolica/Boletim_Eolica_nov-2014.pdf> Acesso em Janeiro de 2015.

Ortiz-García, E. G., Salcedo-Sanz, S., Pérez-Bellido, Á. M., Gascón-Moreno, J., Portilla-Figueras, J. A. & Prieto, L. (2011). Short-term wind speed prediction in wind farms based on banks of support vector machines. *Wind Energy*, *14*(2), 193–207.

Parzen, E. (1962). On estimation of probability density function and mode. *Annals of Mathematical Statistics*, *33*(3), 1065-1076.

PER (2013). Energia Eólica. Disponível em: <http://www.energiasrenovaveis.com/images/upload/flash/anima_como_funciona/eolica29.swf. > Acesso em: Junio 2013.

Pessanha, J. F. M., Barcelos, G. F. B., Faria, A. V. C., & Ferreira, V. M. F. (2010). Análise Estatística de Registros Anemométricos e Seleção de Turbinas Eólicas: Um Estudo de Caso. *XLII Simpósio Brasileiro de Pesquisa Operacional*, Bento Gonçalves – RS.

Philippopoulos, K., & Deligiorgi, D. (2009). Statistical simulation of wind speed in Athens, Greece based on Weibull and ARMA models. *International Journal of Energy and Environment*, *3*(4), 151-158.

Pinson, P., Juban, J., & Kariniotakis, G. N. (2006). On the Quality and Value of Probabilistic Forecasts of Wind Generation. In *2006 International Conference on Probabilistic Methods Applied to Power Systems* (pp.1–7).

Pinson, P., Chevallier, C., & Kariniotakis, G. N. (2007a). Trading Wind Generation From Short-Term Probabilistic Forecasts of Wind Power. *IEEE, Transactions on Power Systems*, *22*(3), 1148–1156.

Pinson, P., Nielsen, H., Møller, J., Madsen, H., & Kariniotakis, G. (2007b). Nonparametric probabilistic forecasts of wind power: Required properties and evaluation. *Wind Energy*, *10*(6), 497–516.

Pinson, P., Christensen, L. E. A., Madsen, H. Sørensen, P., Donovan, M. H. & Jensen, L.E. (2008). Regime-switching modelling of the fluctuations of offshore wind generation. *Journalof Wind Engineering & Industrial Aerodynamics*, *96*(12), 2327–2347.

Pinson, P., & Madsen, H. (2009), Ensemble-based probabilistic forecasting at Horns Rev. *Wind Energy*, *12*(2), 137-155.

Pinson, P., Madsen, H., Nielsen, H. A., Papaefthymiou, G. & Klöckl, B. (2009a). From probabilistic forecasts to statistical scenarios of short-term wind power production. *Wind Energy*, *12*(1), 51-62.

Pinson, P., Nielsen, H. A., Madsen, H., & Kariniotakis, G. (2009b). Skill forecasting from ensemble predictions of wind power. *Applied Energy*, *86*(7–8), 1326–1334.

Pinson, P. (2012). Very-short-term probabilistic forecasting of wind power with generalized logit–normal distributions. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *61*(4), 555–576.

Poggi, P., Muselli, M., Notton, G., Cristofari, C., & Louche, A. (2003). Forecasting and simulating wind speed in Corsica by using an autoregressive model. *Energy Conversion and Management*, *44*(20), 3177–3196.

Polonik, W., & Yao, Q. (2000). Conditional minimum volume preditive regions for stochastic processes. *Journal of the American Statistical Association*, *95*, 509-519.

Potter, C. W., & Negnevitsky, M. (2006). Very short-term wind forecasting for Tasmanian power generation. *IEEE, Transactions on Power Systems*, *21*(2), 965–972.

Pritchard, G. (2011). Short-term variations in wind power Some quantile-type models for probabilistic forecasting. *Wind Energy*, *14*(2), 255-269.

Qadrdan, M., Ghodsi, M., & Wu, J. (2013). Probabilistic wind power forecasting using a single forecast. *International Journal of Energy and Statistics*, *01*(02), 99–111. doi:10.1142/S2335680413500075

Ramírez, P., & Carta, J. A. (2005). Influence of the data sampling interval in the estimation of the parameters of the Weibull wind speed probability density distribution: a case study. *Energy Conversion and Management*, *46*(15-16), 2419–2438.

Ramirez-Rosado, I. J., Ferandez-Jimenez, L. A., Monteiro, C., Sousa, J., & Bessa, R. (2009). Comparison of two new short-term wind power forecasting systems. *Renewable Energy*, *34*, 1848–54

Reikard, G. (2010). Regime-switching models and multiple causal factors in forecasting wind speed. *Wind Energy*, *13*(5), 407 - 408.

Robinson, P. M. (1991). Consistent nonparametric entropy-based testing. *Review of Economic Studies*, *58*, 437-453.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, *27*(3), 832-837.

Rosenblatt, M. (1969). Conditional probability density and regression estimates. Multivariate Analysis II, Ed. P.R. Krishnaiah, pp. 25-31. New York: Academic Press.

Roulston, M., Kaplan, D., Hardenberg, J., & Smith, L. (2003). Using medium-range weather forecasts to improve the value of wind energy production. *Renewable Energy*, *28*(4), 585–602.

Sánchez, I. (2006). Short-Term Prediction of Wind Energy Production. *International Journal of Forecasting*, 22, 43-56.

Sanei, S., Ghodsi, M., & Hassani, H. (2011). An adaptive singular spectrum analysis approach to murmur detection from heart sounds. *Medical Engineering & Physics*, *33*, 362–367.

Sangita, B. P., & Deshmukh, S. R. (2011). Use of support vector machine for wind speed prediction. In *2011 International Conference on Power and Energy Systems* (pp.1–8).

Schmid, J., & Klein, H. P. (1991). Performance of European wind turbines, Commission of the European communities, Elsevier Applied Science, London and New York.

Schwartz, S. C. (1967). Estimation of a Probability Density by an Orthogonal Series. *Annals of Mathematical Statistics*, *38*(4), 1262-1265.

Schwartz, M., & Milligan M. (2002). Statistical Wind Forecasting at the U.S. National Renewable Energy Laboratory. In *Proceedings of the First IEA Joint Action Symposium on Wind Forecasting Techniques*, Norrköping, Sweden, (pp. 115-124B). Published by FOI - Swedish Defence Research Agency.

Scott, D. W., Gotto, A. M., Cole, J. S., & Gorry, G. A. (1978). Plasma Lipids as Collateral Risk Factors in Coronary Artery Disease: A Study of 371 Males with Chest Pain. *JournaI of Chronic Diseases*, *31*(5), 337-345.

Scott, D. W. (1992). Multivariate density estimation. Probability and mathematical statistics. Wiley, New York.

Sfetsos, A. (2000). A comparison of various forecasting techniques applied to mean hourly wind speed time series. *Renewable Energy*, *21*(1), 23–35.

Sfetsos, A. (2002). A novel approach for the forecasting of mean hourly wind speed time series. *Renewable Energy*, *27*(2), 163–174.

Sideratos, G., & Hatziargyriou, N. D. (2007). An Advanced Statistical Method for Wind Power Forecasting. *IEEE, Transactions on Power Systems*, *22*(1), 258–265.

Sideratos, G., & Hatziargyriou, N. D. (2012). Probabilistic wind power forecasting using radial basis function neural networks. *IEEE, Transactions on Power Systems*, *27*(4), 1788–1796.

Silva, G. X., Fonte, P. M., & Quadrado, J. C. (2006). Radial basis function networks for wind speed prediction. In *Proceedings of the 5th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases (AIKED'06)*. Wisconsin, USA, (pp. 286-290).

Silverman, B. W. (1978). Density Ratios, Empirical Likelihood and Cot Death. *Applied Statistics*, *27*(1). 26-33.

Silverman, B. W. (1986). Density Estimation for Statistics and Data Analysis. Chapman & Hall, New York.

Sloughter, J. M., Gneiting, T., & Raftery, A. E. (2010). Probabilistic Wind Speed Forecasting using Ensembles and Bayesian Model Averaging. *Journal of the American Statistical Association, 105*(489), 25-35.

Soman, S. S., Zareipour, H., Malik, O. & Mandal, P. (2010). A review of wind power and wind speed forecasting methods with different time horizons. In *North American Power Symposium (NAPS)*. Arlington, TX, (pp. 1-8).

Späth, H. (1980). Cluster Analysis Algorithms. Chichester, UK: Ellis Horwood.

Tantareanu, C. (1992). Wind Prediction in Short Term: A first step for a better wind turbine control. Nordvestjysk Folkecenter for Vedvarende Energi, ISBN 87-7778-005-1.

Tarter, M. E., & Kronmal, R. A. (1970). On multivariate density estimates based on orthogonal expansions. *Annals of Mathematical Statistics*, *41*(2), 718-722.

Taylor, J. W. (2003). Short-Term Electricity Demand Forecasting Using Double Seasonal Exponential Smoothing. *Journal of Operational Research Society*, *54*, 799–805.

Taylor, J. W., McSharry, P. E., & Buizza, R. (2009). Wind Power Density Forecasting using Ensemble Predictions and Time Series Models. *IEEE Transactions on Energy Conversion*, *24*, 775-782.

Tjøstheim, D. (1994). Non-linear time series: A selective review. *Scandinavian Journal of Statistics*, 21, 97-130

Thorarinsdottir, T. L., & Gneiting, T. (2010). Probabilistic forecasts of wind speed: ensemble model output statistics by using heteroscedastic censored regression. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *173*(2), 371–388.

Torres, J. L., García, A., De Blas, M., & De Francisco, A. (2005). Forecast of hourly average wind speed with ARMA models in Navarre (Spain). *Solar Energy*, *79*(1), 65–77.

Trombe, P.J., Pinson, P., & Madsen, H. A. (2012). General probabilistic forecasting framework for offshore wind power fluctuations. *Energies*, *5*, 621–657.

Vasconcelos, S. (2011). Análise de Componentes Principais. Disponível em: <http://www.ic.uff.br/~aconci/PCA-ACP.pdf>. Acesso em Setembro de 2014.

Wahba, G. (1971). A Polynomial Algorithm for Density Estimation. *Annals of Mathematical Statistics*, *42*(6), 1870-1886.

Wand, M. P., & Jones, M. C. (1995). Kernel Smoothing. Chapman & Hall, London.

Wang, J., Botterud, A., Bessa, R., Keko, H., Carvalho, L., Issicaba, D., Sumaili, J., & Miranda, V. (2011). Wind power forecasting uncertainty and unit commitment. *Applied Energy*, *88*(11), 4014–4023.

Wang, X., Guo, P. & Huang, X. (2011). A Review of Wind Power Forecasting Models. *Energy Procedia*, 12, 770–778.

Watson, G. S., & Leadbetter, M. R. (1963). On the Estimation of the Probability Density I. *Annals of Mathematical Statistics*, *34*(2), 480-491.

Wikle, C.K., & Cressie, N. (1999). A dimension-reduced approach to space-time Kalman filtering. *Biometrika*, *86*, 815–829.

Wu, G., & Dou, Z. (1995). Non linear wind prediction using a fuzzy modular temporal neural network. In *Windpower '95, American Wind Energy Association Conference* (pp. 285-94).

WWEA (2015). New record in worldwide wind installations. Disponível em: <http://www.wwindea.org/new-record-in-worldwide-wind-installations/> Acesso em: Fevereiro 2015.

WWEA (2014). The World Wind Energy Association. Half-year Report 2014. Disponível em: <http://www.wwindea.org/webimages/WWEA_half_year_report_2014.pdf> Acesso em: Dezembro 2014.

WWEA (2012), World Wind Energy Report 2010. Disponível em: <http://www.wwindea.org/webimages/WorldWindEnergyReport2012_final.pdf> Acesso em: Julho 2013.

Xiaojuan, H., Xilin, Z., & Bo, G. (2012). Short-Term Wind Speed Prediction Model of LS-SVM Based on Genetic *Algorithm. Advances in Intelligent and Soft Computing*, *116*, 221 - 229.

Yang, M., Fan, S., & Lee, W.-J. (2012). Probabilistic short-term wind power forecast using componential Sparse Bayesian Learning. *Industrial & Commercial Power Systems Technical Conference (I&CPS), 2012 IEEE/IAS 48th*. Louisville, KY, (pp. 1–8).

Zaïane, O. R., Foss, A., Lee, C. H., & Wang, W. (2002). On data clustering analysis: Scalability, constraints, and validation. In *Advances in Knowledge Discovery and Data Mining* (pp. 28-39). Springer Berlin Heidelberg.

Zhang, N., Kang, C., Member, S., Xia, Q., Member, S., & Liang, J. (2014). Modeling Conditional Forecast Error for Wind Power in Generation Scheduling. *IEEE TRANSACTIONS ON POWER SYSTEMS*, *29*(3), 1316–1324.

Zhang, T., Ramakrishnan, R., & Livny, M. (1996). BIRCH: An Efficient Data Clustering Method for Very Large Databases. In *Proceedings of the ACM SIGMOD Conference on Management of Data*. Montreal, Canada (pp. 103-114).

Zhao, X., Wang, S., & Li, T. (2011). Review of Evaluation Criteria and Main Methods of Wind Power Forecasting. *Energy Procedia*, *12*, 761-769.

Zheng, Z. W., Chen, Y. Y., Huo, M. M., & Zhao, B. (2011). An Overview: the Development of Prediction Technology of Wind and Photovoltaic Power Generation. *Energy Procedia*, *12*, 601–608.

Zhou, J., Shi, J., & Li, G. (2011). Fine tuning support vector machines for short-term wind speed forecasting. *Energy Conversion and Management*, *52*(4), 1990–1998.

Zhou, Z., Botterud, A., Wang, J., Bessa, R.J., Keko, H., Sumaili, J. and Miranda, V. (2013), Application of probabilistic wind power forecasting in electricity markets. *Wind Energy*, *16*(3), 321–338.

10. Apêndices

10.1.
Apêndice A – Análise de Clusters Usando Métodos Hierárquicos Aglomerativos

A *Análise de Clusters* é um nome dado para o grupo de técnicas computacionais cujo propósito consiste em separar indivíduos/objetos em classes, baseando-se nas características que estes indivíduos possuem. A ideia básica consiste em colocar em um mesmo grupo indivíduos que sejam similares de acordo com algum critério pré-determinado (Linden, 2009). Formalmente, esta metodologia pretende que os n indivíduos de um conjunto de dados sejam agrupados em K classes, ou subgrupos mutuamente *excludentes* denominados *cluster*s, de forma tal que cada *cluster* seja internamente homogêneo, isto é, que esteja conformado por indivíduos *similares* entre si, mas *dissimilares* em relação aos indivíduos pertencentes aos outros *clusters*, fazendo com que os *clusters* sejam heterogêneos entre si.

No entanto, outra metodologia para agrupamento é *Análise Discriminante*, a qual parte do pressuposto que é conhecida uma subdivisão específica dentro do conjunto de dados disponível, e visa procurar direções no espaço que evidenciem a separação desses subgrupos ou que permita determinar uma regra para futuras classificações. Atualmente não existe um agrupamento disponível desse tipo, e o problema consiste em identificar quais (e quantas) são as diferentes classes de indivíduos existentes nos dados disponíveis.

Os métodos que permitem determinar ditas classes ou subgrupos são aqueles pertencentes à *Análise de Cluster*. Nos dois extremos de possíveis classificações encontram-se a classificação de todos os indivíduos em uma única classe, e a classificação de cada indivíduo como uma classe separada. Estes dois grandes grupos de métodos de *Análise de Clusters* são: métodos hierárquicos e os métodos não-hierárquicos.

A *Análise de Clusters* é uma ferramenta útil para a análise de dados em muitas situações diferentes. Por exemplo, esta técnica pode ser usada para reduzir a dimensão de um conjunto de dados, eliminando uma ampla gama de indivíduos; e por tanto, restringindo-se à informação do centro do seu conjunto de dados (Linden, 2009). Outro caso no qual dita técnica poder ser usada é no agrupamento dos vetores singulares resultantes da SVD, já que eles podem ser tratados como um conjunto de componentes a ser clusterizado em grupos disjuntos com características específicas.

Existem diversos métodos para o agrupamento através da *Análise de Clusters*, no entanto, salientam-se os *métodos hierárquicos* como técnica de agrupamento, embora a

literatura técnica de SSA não reporte que eles tenham sido empregados para este propósito, mas que para o objetivo e necessidade de agrupamento de séries de alta frequência contidas no desenvolvimento desta tese, pois eles fornecem bons resultados. Isto é suportado no fato de que sabendo sobre a dificuldade de examinar todas as combinações de grupos possíveis em um grande volume de dados, eles conseguem auxiliar na formação dos agrupamentos de forma eficiente.

Uma *Análise de Cluster* criteriosa exige métodos que apresentem as seguintes características (Zaïane et al., 2002):

- Ser capaz de lidar com dados com alta dimensionalidade
- Ser "escalável" com o número de dimensões e com a quantidade de elementos a serem agrupados
- Habilidade para lidar com diferentes tipos de dados
- Capacidade de definir agrupamentos de diferentes tamanhos e formas
- Exigir o mínimo de conhecimento para determinação dos parâmetros de entrada
- Ser robusto à presença de ruído
- Apresentar resultado consistente independente da ordem em que os dados são apresentados

Em termos gerais, nem todos os algoritmos atendem estes requisitos, por isso, é importante entender as características de cada algoritmo para a escolha do método mais adequado que atenda a cada tipo de dado ou problema (Haldiki, 2001).

10.1.1. Medidas de Similaridade

A grande maioria dos métodos de *Análise de Clusters* requer uma medida de similaridade entre os elementos a serem agrupados, usualmente esta medida pode ser expressa como uma função distância ou métrica.

Desta forma, as dissimilaridades d_{ij} entre os indivíduos i e j são medidas de distância que refletem as diferenças (maiores ou menores) entre os valores que esses indivíduos registaram em um conjunto de p atributos/variáveis. No entanto, não é obrigatório que existam observações subjacentes de p variáveis, sendo até possível que as medidas de dissimilaridade sejam atribuídas de forma subjetiva pelo pesquisador. Muitos métodos de *Análise de Clusters* trabalham com dados armazenados na memória principal, normalmente, usam uma das seguintes estruturas de dados no seu processamento:

- ***Matriz de dados***: as linhas representam cada um dos indivíduos a serem agrupados e as colunas, os atributos ou características de cada indivíduo. Considerando n indivíduos com p atributos, tem-se a matriz $n \times p$ seguinte:

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1p} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2p} \\ x_{31} & x_{32} & x_{33} & \cdots & x_{3p} \\ x_{41} & x_{42} & x_{43} & \cdots & x_{4p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & x_{n3} & \cdots & x_{np} \end{bmatrix} \tag{10.1}$$

- ***Matriz de similaridades***: cada elemento da matriz representa a distância entre pares de indivíduos. Dado que a distância entre o indivíduo i e o indivíduo j é igual à distância entre o indivíduo j e o indivíduo i (essa, inclusive, é uma das propriedades inerentes às medidas de similaridade como será visto mais adiante), não é necessário armazenar todas as distâncias entre os indivíduos. Ao considerar n indivíduos a serem agrupados obtêm-se a seguinte matriz quadrada de distâncias $D_{n \times n}$:

$$\begin{bmatrix} 0 & d_{12} & d_{13} & & & d_{1n} \\ d_{21} & 0 & d_{23} & & \ldots & d_{2n} \\ d_{31} & d_{32} & 0 & & & d_{3n} \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & d_{n3} & & \cdots & 0 \end{bmatrix} = D \tag{10.2}$$

Sendo d_{ij} a distância ou similaridade entre o indivíduo i e o indivíduo j como foi dito anteriormente.

As medidas de similaridade são números positivos que representam a distância entre dois indivíduos. Quanto menor o valor de d_{ij}, mais semelhantes serão os indivíduos e, de acordo com um critério, ficarão no mesmo grupo. Pelo contrario, quanto maior a distância, menos similares serão os indivíduos e, em consequência, eles deverão estar em grupos distintos. Quando um método que trabalha com matrizes de dissimilaridade recebe uma matriz de dados, ele primeiro transforma-a em uma matriz de dissimilaridade antes de iniciar suas etapas de clusterização (Han & Kamber, 2001). Uma medida de similaridade é uma função de distância $D: n \times n \rightarrow \mathbb{R}$ tal que para quaisquer $i, j, h \in D$, ela é definida de forma tal que obedeça as seguintes propriedades;

1. Positividade: $d_{ij} \geq 0, \quad \forall\, i,j = 1{:}\,n$;
2. Simetria: $d_{ij} = d_{ji}, \quad \forall\, i,j = 1{:}\,n$;
3. Reflexiva: $d_{ij} = 0 \Leftrightarrow i = j, \quad \forall\, i = 1{:}\,n$;
4. $Desigualdade\ Triangular\ d_{ij} \leq d_{ih} + d_{hj}$.

Não existe uma medida de similaridade que sirva para todos os tipos de variáveis que podem existir em uma base de dados. As medidas que são geralmente utilizadas para calcular as dissimilaridades de indivíduos descritos por variáveis contínuas são: a distância *Euclidiana*, distância *Euclidiana Generalizada*, distância *de Mohalanobis*, distância *de Minkowski e* distância *Manhattan*, distância *Canberra* e separação angular entre outros.

A unidade da medida usada pode afetar a análise. Para evitar isso, sugere-se normalizar os dados antes da aplicar a *Análise de Cluster*. A normalização é efetuada para cada variável *f* (cada atributo dos indivíduos) da seguinte forma:

1. Calcular a média do desvio absoluto, s_f:

$$s_f = \frac{1}{n}\left(\left|x_{1f} - m_f\right| + \left|x_{2f} - m_f\right| + \cdots + \left|x_{nf} - m_f\right|\right) \tag{10.3}$$

Os valores x_{1f} a x_{nf} são os valores do atributo f para os n indivíduos a serem agrupados. Já m_f, é o valor médio do atributo *f*, isto é:

$$m_f = \frac{1}{n}\left(x_{1f} + x_{2f} + \cdots + x_{nf}\right) \tag{10.4}$$

O valor do *i*-ésimo indivíduo da variável f normalizado será dado por:

$$z_{if} = \frac{x_{if} - m_f}{s_f} \tag{10.5}$$

As medidas de similaridade mais usadas são:

- **Distância Euclidiana:** é a distância geométrica no espaço multidimensional $\mathbb{R}^p$. Assim que para dois indivíduos com p atributos a distância está definida por:

$$d_{i,j} = \sqrt{\left(x_{i1} - x_{j1}\right)^2 + \left(x_{i2} - x_{j2}\right)^2 + \cdots + \left(x_{ip} - x_{jp}\right)^2} \tag{10.6}$$

$$d_{i,j} = \sqrt{\sum_{k=1}^{p}\left(x_{ik} - x_{jk}\right)^2} \tag{10.7}$$

- **Distância Euclidiana Generalizada**: Esta distância está presentada pela Equação seguinte:

$$d_{i,j} = \sqrt{\sum_{k=1}^{p}\sum_{l=1}^{p} w_{kl}\left(x_{ik} - x_{jk}\right)\left(x_{il} - x_{jl}\right)} \tag{10.8}$$

Sendo w_{kl} um elemento da matriz $\boldsymbol{W}$ definida positiva. Alguns casos especiais resultam de escolher:

- ✓ $\boldsymbol{W = D^{-2}}$, onde $\boldsymbol{D^{-2}}$ é a matriz diagonal dos recíprocos das variâncias, e que corresponde a tomar a distância *Euclidiana* usual entre os dados normalizados.
- ✓ $\boldsymbol{W} = \Sigma^{-1}$, onde Σ é a matriz de variâncias-covariâncias das variáveis, calculada com todos os indivíduos. Esta escolha gera a chamada distância de *Mahalanobis*, que é invariante a mudanças de escala nas variáveis, ou seja, invariante a qualquer transformação linear não singular e tende a formar *clusters* hiperelípticos.

- **Distância de Minkowski**: esta distância de *Minkowski* é a generalização das distâncias anteriormente descritas e cuja expressão está dada por:

$$d_{i,j} = \left(\sum_{k=1}^{p}\left|x_{ik} - x_{jk}\right|^{\lambda}\right)^{1/\lambda} \tag{10.9}$$

- ✓ Com $\boldsymbol{\lambda = 1}$, obtêm-se a **distância de *Manhattan***, **distância $\boldsymbol{\ell_1}$** ou "*City-block" metric*. Esta medida de similaridade é mais facilmente calculada do que a *Euclidiana*, mas ela pode não ser adequada se os atributos estão correlacionados, pois não há garantia da qualidade dos resultados obtidos (Cole, 1998).

- ✓ Com $\boldsymbol{\lambda = 2}$, tem-se a **distância *Euclidiana*** habitual ou **distância $\boldsymbol{\ell_2}$** definida na Subseção 4.1.1.
- ✓ No limite $\boldsymbol{\lambda \to \infty}$, obtêm-se a **distância de *Chebychev***, $d_{ij} = \max_k |x_{ik} - x_{jk}|$ que se denomina também **distância do *Máximo*** ou **distância $\boldsymbol{\ell_\infty}$**.

Em geral, quanto maior for o valor de $\boldsymbol{\lambda}$, maior será o peso relativo de indivíduos muito dissimilares dos restantes.

- **Distância de Canberra**: esta distância está desenhada para variáveis que unicamente possam tomar valores não-negativos variando no intervalo [0,1], e está definida:

$$d_{i,j} = \sum_{k=1}^{p} \frac{|x_{ik} - x_{jk}|}{(x_{ik} + x_{jk})} \tag{10.10}$$

e d_{ij} indica a máxima semelhança que ocorre apenas quando os indivíduos i e j são idênticos.

Criou-se esta medida com o propósito de obter uma métrica invariante a transformações de escala diferenciadas em cada variável. Nela é relativizada cada diferença através do denominador (cujas unidades de medida são iguais às do numerador).

A escolha do critério de distância entre indivíduos a ser utilizado não tem que se limitar às opções acima indicadas. Algumas destas opções são desaconselhadas para certos tipos de dados (por exemplo, as distâncias de *Minkowski* podem não ser aconselháveis para o caso em que as p variáveis tiverem diferentes unidades de medida).

- **Coeficiente de Correlação de Pearson**: dados dois indivíduos i e j, este coeficiente está definido por:

$$c_{i,j} = \frac{\sum_{k=1}^{p}(x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)}{\left[\sum_{k=1}^{p}(x_{ik} - \bar{x}_i)^2 \sum_{k=1}^{p}(x_{jk} - \bar{x}_j)^2\right]^{1/2}} \tag{10.11}$$

Sendo

$$\bar{x}_s = \frac{\sum_{k=1}^{p} x_{sk}}{p}$$

Este coeficiente varia no intervalo $[-1,1]$, em que $c_{ij} = 1$ indica a similaridade máxima, mas não necessariamente identidade entre as características dos indivíduos i e j. Se $c_{ij} = -1$ indica o máximo de dissimilaridade.

- **Coeficiente de Separação Angular (ou cosseno)**: entre os indivíduos i e j, está definido por:

$$c_{ij} = \frac{\sum_{k=1}^{p} x_{ik} \cdot x_{jk}}{\left(\sum_{k=1}^{p} x_{ik}^2 \sum_{k=1}^{p} x_{jk}^2\right)^{1/2}} \tag{10.12}$$

  Este coeficiente de similaridade define-se no intervalo $[-1,1]$ onde:

  - $c_{ij} = 1$ indica que a similaridade é máxima (dissimilaridade mínima),
  - $c_{ij} = -1$ indica que a similaridade é mínima (dissimilaridade máxima),
  - c_{ij} é o cosseno do ângulo formado pelas duas semi-retas que unem a origem aos respectivos indivíduos, representados como pontos no espaço.

Tanto o coeficiente de correlação de Pearson como o coeficiente de separação angular (ou cosseno) podem ser usados para quantificar semelhanças entre os objetos observados i e j, em um espaço de dimensão p, ainda que eles não consigam satisfazer a propriedades das medidas de similaridade.

10.1.2. Similaridades e Dissimilaridades entre Indivíduos

Na situação em que o ponto de inicio seja um conjunto de medidas de similaridade entre cada par de indivíduos, procede-se a operar diretamente sobre estas medidas de similaridade ou, opcionalmente, convertem-se as medidas de similaridade em medidas de dissimilaridade. Partindo do fato que estão sendo usadas medidas de similaridade, são frequentemente utilizadas as seguintes formas de converter essas similaridades em dissimilaridades:

1. $d_{ij} = 1 - s_{ij}$,
2. $d_{ij} = 1 - s_{ij}^2$,
3. $d_{ij} = \sqrt{1 - s_{ij}}$,

4. $d_{ij} = \sqrt{1 - s_{ij}^2}$.

Posterior à obtenção das medidas de dissimilaridade (seja porque foram quantificadas da base de dados disponibilizada ou porque um conjunto de medidas foi fornecido para trabalhar baseado nele), continua-se com a aplicação do método de agrupamento escolhido levando em consideração o tipo de variáveis e medidas calculadas.

10.1.3. Métodos Hierárquicos

Os algoritmos de *Análise de Clusters* fundamentados no método hierárquico organizam um grupo de dados em uma estrutura hierárquica de acordo com a proximidade entre os indivíduos. Logo, dado um conjunto de n indivíduos, o ponto de partida para os métodos de classificação hierárquicos será, em geral, uma matriz $n \times n$ cujo elemento genérico (i, j) é uma *medida de similaridade* (ou *dissimilaridade*) entre o indivíduo i e o indivíduo j. Portanto, podem ser considerados inicialmente os n indivíduos como constituindo n classes diferentes. Procedendo por etapas, vai-se fundindo um par de classes em cada etapa. A fusão a efetuar em uma dada etapa é a fusão dos dois subgrupos *clusters* considerados mais "semelhantes". Este processo pode ser feito de forma iterativa até que todos os indivíduos sejam fusionados em uma única classe.

Convencionalmente, os resultados do método hierárquico são exibidos de forma gráfica através de uma árvore bi-dimensional denominada dendrograma (Figura 10.1), a qual representa as sucessivas fusões dos *clusters*, ou seja, é uma árvore que iterativamente divide a base de dados em subconjuntos menores. A raiz do dendrograma representa o agrupamento de todos os indivíduos e os nós folhas representam os indivíduos. O resultado da clusterização pode ser obtido cortando-se o dendrograma em diferentes níveis de acordo com o número de *cluster*s *K* desejados. Esta forma de representação fornece descrições informativas e visualização para as estruturas de grupos potenciais. Em tais hierarquias, cada nó da árvore representa um *cluster* da base de dados.

Na criação dos dendrogramas os métodos hierárquicos estabelecem duas tipologias de abordagens: os métodos aglomerativos ou ascendentes (*bottom-up*) e os métodos divisivos ou descendentes (*top-down*). Em ambos os métodos as divisões ou fusões, uma vez realizadas, não podem ser revertidas, o que significa que uma vez a abordagem

aglomerativa une dois indivíduos posteriormente não pode separá-los; enquanto que na abordagem divisiva uma vez separados dos dois indivíduos, não podem ser unidos novamente. Segundo Kaufman & Rousseuw (1990), os métodos hierárquicos têm o defeito de nunca poderem reparar o que foi feito em conjuntos prévios.

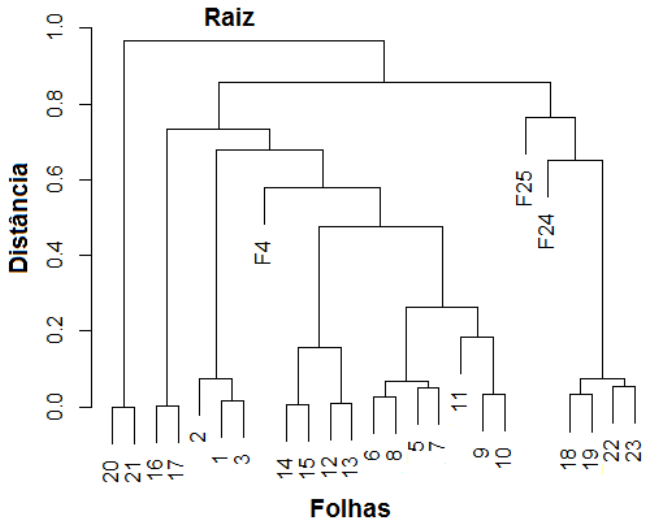

Figura 10.1 – Exemplificação de um dendrograma. (Fonte: Elaboração do autor)

Fazendo uso do dendrograma e com base no conhecimento sobre a estrutura dos dados, deve-se estabelecer uma distância de corte para determinar quais serão os grupos formados. Essa decisão é subjetiva, e deve ser feita de acordo o objetivo da análise e o número de grupos desejados.

10.1.3.1. Métodos Aglomerativos

Dentro dos métodos hierárquicos, eles são considerados os mais populares por serem amplamente usados. A abordagem inicializa-se considerando cada indivíduo como sendo um *cluster*, totalizando *n clusters* (cada *cluster* está conformado por um indivíduo), partindo das folhas para a raiz. Em cada etapa, calcula-se a distância entre cada par de *clusters*. Estas distâncias são geralmente, armazenadas em uma matriz de dissimilaridade simétrica. Posteriormente, escolhem-se dois *clusters* com a distância mínima e logo são unidos. A seguir, atualiza-se a matriz de distâncias. Este processo continua iterativamente até que todos os indivíduos estejam em um único *cluster* (o nível mais alto da hierarquia), ou até que uma condição de término ocorra (Agrawal et al., 1998; Ng & Han, 1994; Han

& Kamber, 2001). Os métodos aglomerativos podem ser descritos em quatro passos (Johnson & Wichern, 2002):

1. Considerar os indivíduos iniciais como sendo *clusters* (grupos) singulares e calcular matriz de matriz de dissimilaridade, entre o indivíduo i e indivíduo j, com $i = 1,\ldots,n$; $j = 1,\ldots,n$ e $D_{n\times n} = [d_{ij}]$, como foi definido anteriormente.
2. Procurar na matriz $D_{n\times n}$ os pares de indivíduos i e j mais semelhantes, ou seja, com menor dissimilaridade. Caso existam vários valores iguais se optará pelo indivíduo com menor valor alfanumérico. Partindo do fato que cada indivíduo, inicialmente, constitui um grupo, descreve-se este procedimento da seguinte forma: detectar os dois grupos mais semelhantes, por exemplo, G e H; a sua dissimilaridade está representada por d_{GH}.
3. Fusionar os pontos G e H à distância crítica – distância entre os dois grupos, d_{GH}. Actualizar a matriz $D_{n\times n}$, eliminando as linhas e as colunas correspondentes aos grupos G e H e introduzindo uma nova linha e coluna com as dissimilaridades calculadas entre o novo grupo (GH) e cada um dos grupos restantes. Com esta operação a ordem da matriz baixa uma unidade.
4. Repetir os passos 2 e 3 um total de $n-1$ vezes até obter um único grupo que, desta forma, inclua todos os indivíduos.

Na Seção 10.1.4 serão definidas as metodologias mais usadas para o calculo da distância entre dois grupos no processo de clusterização. Existem diversos enfoques, e cada um deles proporciona um método hierárquico aglomerativo diferente.

10.1.3.2. Métodos Divisivos

Os métodos divisivos funcionam de forma inversa que os métodos aglomerativos, isto é, partem da raiz para as folhas; assim que ao ser um processo inverso, começa-se com todos os indivíduos em um único *cluster* contendo todos os indivíduos. Este *cluster* é dividido em dois grupos distintos com base em algum critério de dissimilaridade. Em cada etapa, um *cluster* é escolhido e dividido em dois *clusters* menores. Este processo continua até que se tenham n *clustersk* (n é o número total de indivíduos) ou até que uma condição de finalização, por exemplo, que o número de *clusters K* desejado seja atingido (Ng & Han, 1994; Han & Kamber, 2001).

Computacionalmente é exigido que em cada etapa sejam calculadas $2^K - 1$ dissimilaridades correspondentes à divisão dos K indivíduos em dois grupos distintos. Na primeira etapa são $2^n - 1$ dissimilaridades, uma vez que inicialmente existe um único cluster com os n indivíduos.

A nível gráfico, movimenta-se da raiz do dendrograma para as folhas, ao contrário do que acontece nos métodos aglomerativos, que se movem das folhas para a raiz. Os métodos divisivos têm a vantagem de que ao culminar as primeiras etapas do processo são fornecidos grandes grupos, informando, portanto, sobre a estrutura principal dos dados, que é o geralmente é o interesse do investigador. No entanto, são menos utilizados que os métodos aglomerativos por seu custo computacional.

Os métodos aglomerativos são mais populares do que os métodos divisivos. Segundo Zhang et al. (1996) os métodos hierárquicos não tentam encontrar os melhores *clusters*, mas manter junto o par mais próximo (ou separar o par mais distante) de indivíduos para formar *clusters*.

10.1.4.
Critérios de (Des)Agregação de Clusters

As medidas de dissimilaridade entre indivíduos não são o único processo para realizar uma *Análise de Cluster*. Para isto, é preciso também determinar o critério para medir a dissimilaridade entre os grupos conhecido como o método de agregação de classes. Suponha que foram dadas duas classes, G e H, para medir a similaridade/dissimilaridade entre elas, de forma tal que se possam realizar fusões entre as classes, é necessário escolher alguma métrica de dissimilaridade. Assim que a continuação serão apresentados os métodos mais comuns.

10.1.4.1.
Método do Vizinho mais Próximo

Método conhecido também como: *nearest neighbour*, *single-link* ou *connected*. Desde o ponto de vista computacional é o método mais "económico" e consiste em considerar que a distância entre dois classes ou grupos é a menor distância entre um indivíduo de um grupo e um indivíduo do outro grupo:

$$D_{GH} = \min_{\substack{k \in G \\ l \in H}} d_{kl} \tag{10.13}$$

O método do vizinho mais próximo tende a produzir grupos "alongados", com indivíduos que podem estar muito distantes entre si, mas pertencendo a uma mesma classe. Esta situação é conhecida pelo nome de *encadeamento*, a qual se apresenta quando existe um indivíduo de um grupo "próximo" de um único indivíduo de outro grupo para que estes sejam fusionados, independente de que outros indivíduos dos grupos que estejam muito distantes entre si. Este *encadeamento* normalmente se reflete no dendrograma numa árvore com grupos mal definidos, onde as fusões acontecem rapidamente.

10.1.4.2.
Método do Vizinho mais Distante

Este método por sua vez é conhecido também como: *furthest neighbour*, *complete-link* ou *compact*). Consiste em considerar que a distância entre dois grupos é a maior distância entre um indivíduo de um grupo e um indivíduo do outro grupo:

$$D_{GH} = \max_{\substack{k \in G \\ l \in H}} d_{kl} \tag{10.14}$$

Os métodos do vizinho mais próximo e do vizinho mais distante são os únicos dois que são invariantes a transformações monótonas do conceito de dissimilaridade, o implica que eles produzem a mesma classificação antes e depois de uma transformação crescente ou decrescente das dissimilaridades.

10.1.4.3.
Método das Distâncias Médias entre Grupos

Método também denominado: *group average* ou *average link*. Aqui é considerada que a distância entre duas classes é a média de todas as distâncias entre pares de indivíduos (um em cada classe):

$$D_{GH} = \frac{1}{n_G \cdot n_H} \sum_{k=1}^{n_G} \sum_{l=1}^{n_H} d_{kl} \tag{10.15}$$

Por outra parte, considerem-se dois métodos adicionais de dissimilaridade entre classes. Eles são válidos quando subjacente às dissimilaridades entre indivíduos existe, não somente uma matriz $n \times n$ de dissimilaridades, mas uma matriz $X_{n \times p}$ de dados multivariados.

10.1.4.4. Método da Inércia Mínima (Método de Ward)

O método de *Ward*, denominado também: *minimum variance method*. Estabelece que a inércia de uma classe G é a soma de quadrados das diferenças entre cada indivíduo e o "indivíduo médio" dessa classe (dado pelo centro de gravidade da nuvem de n pontos em $\mathbb{R}^p$):

$$I_G = \sum_{l=1}^{p} \left[\sum_{k \in G} \left(x_{kl} - \overline{x_l^G} \right)^2 \right] \tag{10.16}$$

sendo $\overline{x_l^G}$ a média dos valores da variável l para os indivíduos da classe G. Tome-se agora a distância entre as classes G e H como sendo o aumento na soma total das inércias provocado pela fusão dos grupos G e H. Em outras palavras, seja I_G a inércia da classe G, I_H a inércia da classe H e $I_{G \cup H}$ a inércia da classe resultante de fundir as classes G e H, então (uma vez que a fusão de G e H não afecta as inércias das restantes classes):

$$D_{GH} = I_{G \cup H} - (I_G + I_H) \tag{10.17}$$

Caso $\overline{x_G}$ e $\overline{x_H}$ sejam os vectores centro de gravidade das classes G e H, respectivamente, tem-se então:

$$D_{GH} = \frac{n_G \cdot n_H}{n_G + n_H} \| \overline{x_G} - \overline{x_H} \|^2 \tag{10.18}$$

onde $\|\cdot\|$é a habitual norma euclidiana.

O método da inércia mínima (Método de *Ward*) tende a formar grupos com um número de indivíduos praticamente igual.

10.1.4.5. Método dos Centroides

Neste método considera-se a distância entre duas classes como sendo a distância entre os centros de gravidade, ou outros pontos considerados "representativos" (centroides), das classes:

$$D_{GH} = \|\overline{x_G} - \overline{x_H}\| \tag{10.19}$$

É importante destacar que os métodos do vizinho mais distante, da distância média entre grupos e dos centroides tendem a produzir grupos "esféricos", ou seja, grupos onde não existem grandes diferenças nas distâncias entre os pares de indivíduos mais distantes, ao serem avaliados em várias direções.

Qualquer uma das definições de distância entre classes acima definida pode ser vista como um caso particular de uma única expressão para a distância entre a classe K e a fusão das classes G e H (Lance, 1967). De facto, representando a distância entre duas classes K e G por D_{KG} e a distância entre qualquer classe K e a fusão das classes G e H por $D_{K\cdot GH}$, tem-se, a forma geral seguinte:

$$D_{K\cdot GH} = \alpha_G D_{KG} + \alpha_H D_{KH} + \beta D_{GH} + \gamma |D_{KG} - D_{KH}|$$

Desta forma, tem-se que para:

- **Vizinho mais próximo**:

$$\alpha_G = \alpha_H = \frac{1}{2}; \quad \beta = 0 \quad \gamma = -\frac{1}{2}$$

- **Vizinho mais distante**:

$$\alpha_G = \alpha_H = \gamma = \frac{1}{2}; \quad \beta = 0$$

- **Distâncias médias entre grupos:**

$$\alpha_G = \frac{n_G}{n_G + n_H}; \quad \alpha_H = \frac{n_H}{n_G + n_H}; \quad \beta = \gamma = 0$$

- **Inércia mínima:**

$$\alpha_G = \frac{n_G + n_K}{n_G + n_H + n_K}; \quad \alpha_H = \frac{n_H + n_K}{n_G + n_H + n_K}; \quad \beta = -\frac{n_K}{n_G + n_H + n_K}; \quad \gamma = 0$$

- **Centroides:**

$$\alpha_G = \frac{n_G}{n_G + n_H}; \quad \alpha_H = \frac{n_H}{n_G + n_H}; \quad \beta = -\frac{n_G n_H}{(n_G + n_H)^2}; \quad \gamma = 0$$

A escolha do método de agregação, isto é, as distâncias D_{GH}, condicionará o agrupamento resultante. A escolha de um método específico deve, portanto, ser justificável com base na natureza dos dados e no objetivo da análise. O facto de que os agrupamentos realizados possam discrepar segundo os critérios de distâncias entre indivíduos e/ou entre classes, sugere que pode ser útil experimentar vários métodos de distâncias, com o propósito de verificar se há robustez (consistência) na clusterização geral.

Embora o agrupamento (para uma dada escolha de distâncias entre indivíduos e entre grupos) seja único, a representação no dendrograma não é única, uma vez que a ordem dos indivíduos (das folhas da árvore) é arbitrária. Em certas ocasiões, reordenamento das folhas produz árvores com um aspecto diferente (ainda que a informação contida nelas seja idêntica). Em geral, os programas computacionais de *Análise de Clusters* visam escolher um ordenamento que evite que as folhas da árvore se cruzem, além de gerar uma representação gráfica na qual não apenas a hierarquia do agrupamento é visível, senão também a dissimilaridade entre os grupos que se vão fusionando. Esta informação é refletida na altura (admitindo que os indivíduos estejam dispostos no eixo horizontal) das barras horizontais que unem os grupos fusionados.

Uma classificação apropriada deverá corresponder a um corte claro na árvore, ou seja, a um agrupamento que resulte de cortar o dendrograma em uma zona onde as separações entre grupos correspondam a grandes distâncias (o que se traduz em uma união de barras de grupos relativamente compridas). Este situação refletirá uma heterogeneidade entre grupos, que como foi referido no início, é um dos objetivos da classificação. O outro objetivo, o da homogeneidade interna dos grupos, será melhor atingido quanto mais próximo seja feito o corte das folhas da árvore.

10.1.5.
Matriz de Correlação Ponderada e o Coeficiente de Correlação Angular

A matriz de *correlação ponderada* é interpretada como sendo os cossenos dos *ângulos formados entre os vetores (componentes)* correspondentes. Analogamente, o *coeficiente de separação angular* é o *cosseno do ângulo formado pelas duas semi-retas* que unem a origem aos respectivos indivíduos, representados como pontos no espaço. Por conseguinte, ao rearranjar os termos da Equação (4.15) que permitem calcular a matriz de *correlações ponderadas*, e substituindo o valor de 1 para o parâmetro de ponderação ω_k na equação de *correlação ponderada* tem-se:

$$\rho_{12}^{(\omega)} = \frac{\left(\mathrm{Y}_{\mathrm{T}}^{(1)}, \mathrm{Y}_{\mathrm{T}}^{(2)}\right)_{\omega}}{\left\|\mathrm{Y}_{\mathrm{T}}^{(1)}\right\|_{\omega} \left\|\mathrm{Y}_{\mathrm{T}}^{(2)}\right\|_{\omega}} = \frac{\sum_{k=1}^{\mathrm{T}} \omega_k \mathrm{Y}_k^{(1)} \mathrm{Y}_k^{(2)}}{\sqrt{\left(\mathrm{Y}_{\mathrm{T}}^{(1)}, \mathrm{Y}_{\mathrm{T}}^{(1)}\right)_{\omega}} \sqrt{\left(\mathrm{Y}_{\mathrm{T}}^{(2)}, \mathrm{Y}_{\mathrm{T}}^{(2)}\right)_{\omega}}}$$

$$= \frac{\sum_{k=1}^{\mathrm{T}} 1 \cdot \mathrm{Y}_k^{(1)} \mathrm{Y}_k^{(2)}}{\sqrt{\sum_{k=1}^{\mathrm{T}} 1 \cdot \mathrm{Y}_k^{(1)} \mathrm{Y}_k^{(1)}} \sqrt{\sum_{k=1}^{\mathrm{T}} 1 \cdot \mathrm{Y}_k^{(2)} \mathrm{Y}_k^{(2)}}}$$

$$= \frac{\sum_{k=1}^{\mathrm{T}} \mathrm{Y}_k^{(1)} \mathrm{Y}_k^{(2)}}{\sqrt{\sum_{k=1}^{\mathrm{T}} \left(\mathrm{Y}_k^{(1)}\right)^2 \sum_{k=1}^{\mathrm{T}} \left(\mathrm{Y}_k^{(2)}\right)^2}}. \quad (10.20)$$

Generalizando para qualquer componente i e j, tem-se:

$$\rho_{ij}^{(\omega)} = \frac{\sum_{k=1}^{\mathrm{T}} \mathrm{Y}_k^{(i)} \mathrm{Y}_k^{(j)}}{\left[\sum_{k=1}^{\mathrm{T}} \left(\mathrm{Y}_k^{(i)}\right)^2 \sum_{k=1}^{\mathrm{T}} \left(\mathrm{Y}_k^{(j)}\right)^2\right]^{1/2}}. \quad (10.21)$$

Comparando esta última expressão com a Equação (10.12) do *coeficiente de separação angular* pode-se verificar que:

$$\rho_{ij}^{(\omega)} = \frac{\sum_{k=1}^{\mathrm{T}} \mathrm{Y}_k^{(i)} \mathrm{Y}_k^{(j)}}{\left[\sum_{k=1}^{\mathrm{T}} \left(\mathrm{Y}_k^{(i)}\right)^2 \sum_{k=1}^{\mathrm{T}} \left(\mathrm{Y}_k^{(j)}\right)^2\right]^{1/2}} \cong \frac{\sum_{k=1}^{p} x_{ik} \cdot x_{jk}}{\left(\sum_{k=1}^{p} x_{ik}^2 \sum_{k=1}^{p} x_{jk}^2\right)^{1/2}} = c_{i,j} \tag{10.22}$$

Com o intuito de exemplificar a equivalência entre o *coeficiente de correlação angular* e a matriz de *correlação ponderada* quando o parâmetro de ponderação ω_k é igual a 1, desenvolveu-se o processo de construção da matriz de similaridades *D* empregando o coeficiente de correlação angular levando em consideração a informação da matriz de dados da Tabela 10.1.

Tabela 10.1 – Matriz de dados para construir a matriz de similaridades.

Indivíduos\Atributos	**X**	**Y**
1	4	3
2	2	7
3	4	7
4	2	3

Observe que a matriz contem 4 indivíduos ($n = 4$) e 2 atributos ($p = 2$), portanto, a dimensão da matriz de similaridade será de 4×4. Usando a Equação (10.12), obtêm-se os seguintes resultados:

$$c_{ij} = \frac{\sum_{k=1}^{2} x_{ik} \cdot x_{jk}}{\left(\sum_{k=1}^{2} x_{ik}^2 \sum_{k=1}^{2} x_{jk}^2\right)^{1/2}} = \frac{\left(x_{i1} \cdot x_{j1}\right) + \left(x_{i2} \cdot x_{j2}\right)}{\left[\left(x_{i1}^2 + x_{i2}^2\right) \cdot \left(x_{j1}^2 + x_{j2}^2\right)\right]^{1/2}} \quad com \quad i,j = 1,2,3,4$$

- Cálculo de c_{11}:

$$c_{11} = \frac{(x_{11} \cdot x_{11}) + (x_{12} \cdot x_{12})}{[(x_{11}^2 + x_{12}^2) \cdot (x_{11}^2 + x_{12}^2)]^{1/2}} = \frac{(x_{11}^2 + x_{12}^2)}{[(x_{11}^2 + x_{12}^2)^2]^{1/2}} = 1$$

- Cálculo de c_{12}:

$$c_{12} = \frac{(x_{11} \cdot x_{21}) + (x_{12} \cdot x_{22})}{[(x_{11}^2 + x_{12}^2) \cdot (x_{21}^2 + x_{22}^2)]^{1/2}} = \frac{(4 \cdot 2) + (3 \cdot 7)}{[(16 + 9) \cdot (4 + 49)]^{1/2}} = 0.7967$$

- Cálculo de c_{13}:

$$c_{13} = \frac{(x_{11} \cdot x_{31}) + (x_{12} \cdot x_{32})}{[(x_{11}^2 + x_{12}^2) \cdot (x_{31}^2 + x_{32}^2)]^{1/2}} = \frac{(4 \cdot 4) + (3 \cdot 7)}{[(16 + 9) \cdot (16 + 49)]^{1/2}} = 0.9179$$

- Cálculo de c_{14}:

$$c_{14} = \frac{(x_{11} \cdot x_{41}) + (x_{12} \cdot x_{42})}{[(x_{11}^2 + x_{12}^2) \cdot (x_{41}^2 + x_{42}^2)]^{1/2}} = \frac{(4 \cdot 2) + (3 \cdot 3)}{[(16 + 9) \cdot (4 + 9)]^{1/2}} = 0.9430$$

- Cálculo de c_{23}:

$$c_{23} = \frac{(x_{21} \cdot x_{31}) + (x_{22} \cdot x_{32})}{[(x_{21}^2 + x_{22}^2) \cdot (x_{31}^2 + x_{32}^2)]^{1/2}} = \frac{(2 \cdot 4) + (7 \cdot 7)}{[(4 + 49) \cdot (16 + 49)]^{1/2}} = 0.9711$$

- Cálculo de c_{24}:

$$c_{24} = \frac{(x_{21} \cdot x_{41}) + (x_{22} \cdot x_{42})}{[(x_{21}^2 + x_{22}^2) \cdot (x_{41}^2 + x_{42}^2)]^{1/2}} = \frac{(2 \cdot 2) + (7 \cdot 3)}{[(4 + 49) \cdot (4 + 9)]^{1/2}} = 0.9524$$

- Cálculo de c_{24}:

$$c_{34} = \frac{(x_{31} \cdot x_{41}) + (x_{32} \cdot x_{42})}{[(x_{31}^2 + x_{32}^2) \cdot (x_{41}^2 + x_{42}^2)]^{1/2}} = \frac{(4 \cdot 2) + (7 \cdot 3)}{[(16 + 49) \cdot (4 + 9)]^{1/2}} = 0.9976$$

Desta forma:

$$D_{4\times4} = \begin{bmatrix} 1 & 0.7967 & 0.9179 & 0.9430 \\ 0.7967 & 1 & 0.9711 & 0.9524 \\ 0.9179 & 0.9711 & 1 & 0.9976 \\ 0.9430 & 0.9524 & 0.9976 & 1 \end{bmatrix}$$

A matriz de *correlações ponderadas* foi calculada fixando o parâmetro de ponderação $\omega_k = 1$. Os resultados foram idênticos aos obtidos com a medida de similaridade do *coeficiente de correlação angular*. Portanto, numericamente se estabelece a equivalência entre a matriz de *correlações ponderadas*, calculada na metodologia de SSA e a medida de similaridade *coeficiente de correlação angular*, pertencente ao método de clusterização hierárquico aglomerativo. Consequentemente este resultado permite justificar a escolha da matriz de *correlações ponderadas* como medida de similaridade, com a finalidade de sintetizar procedimentos e cálculos matemáticos no processo de agrupamento.

10.2.
Apêndice B – Dendrogramas

A continuação apresentam-se as Figuras dos dendrogramas restantes do processo de agrupamento das autotriplas da segunda fase da metodologia SSA *sequencial*, exposta na Seção 7.3.

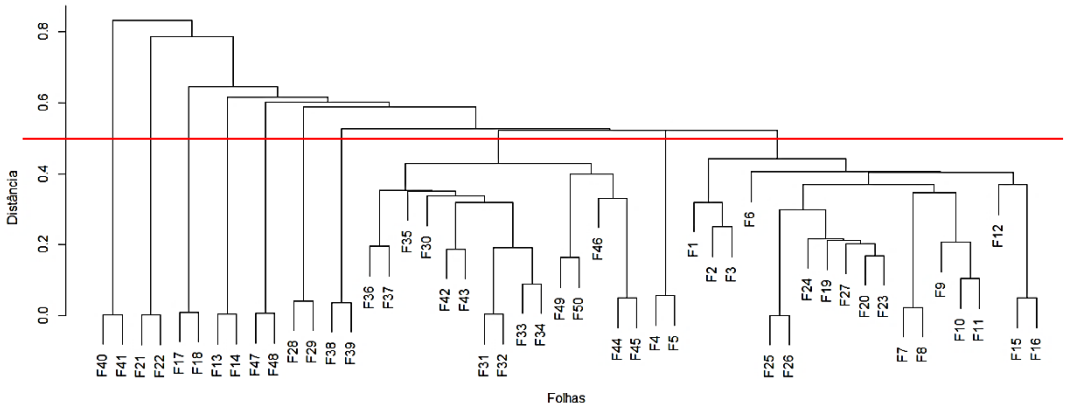

Figura 10.2 – Dendrograma das autotriplas 51 até 100.

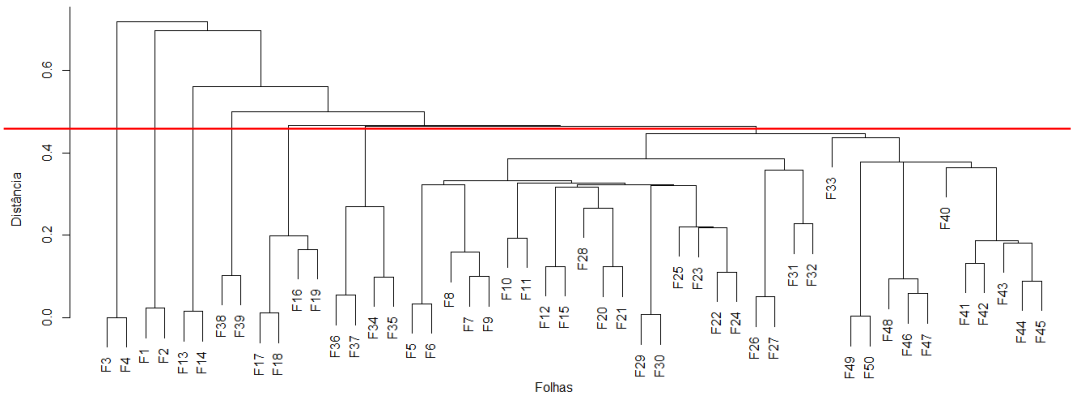

Figura 10.3 – Dendrograma das autotriplas 101 até 150.

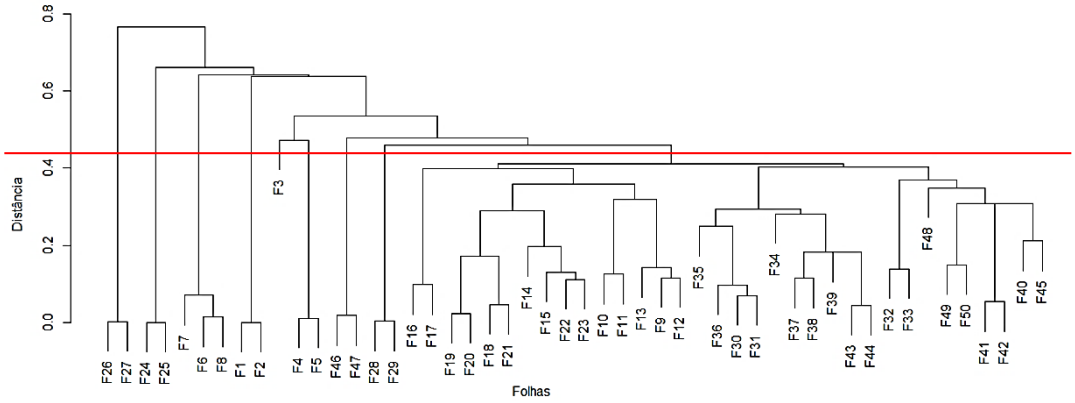

Figura 10.4 – Dendrograma das autotriplas 151 até 200.

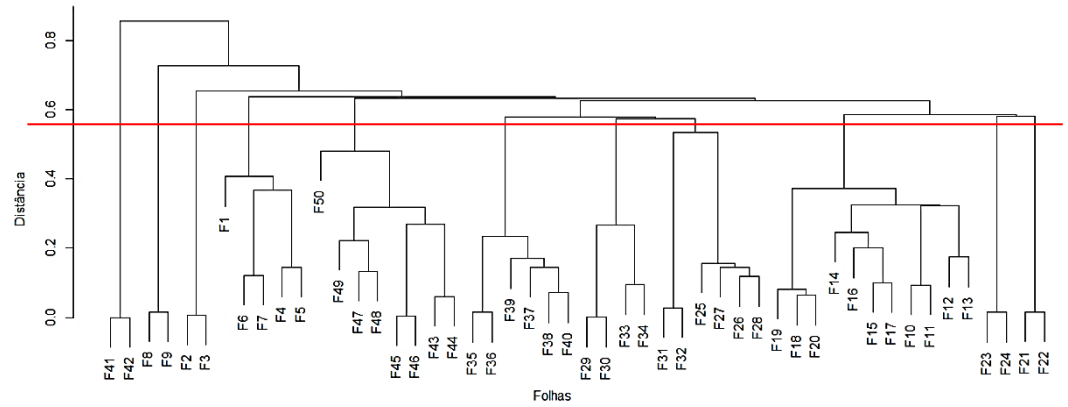

Figura 10.5– Dendrograma das autotriplas 201 até 250.

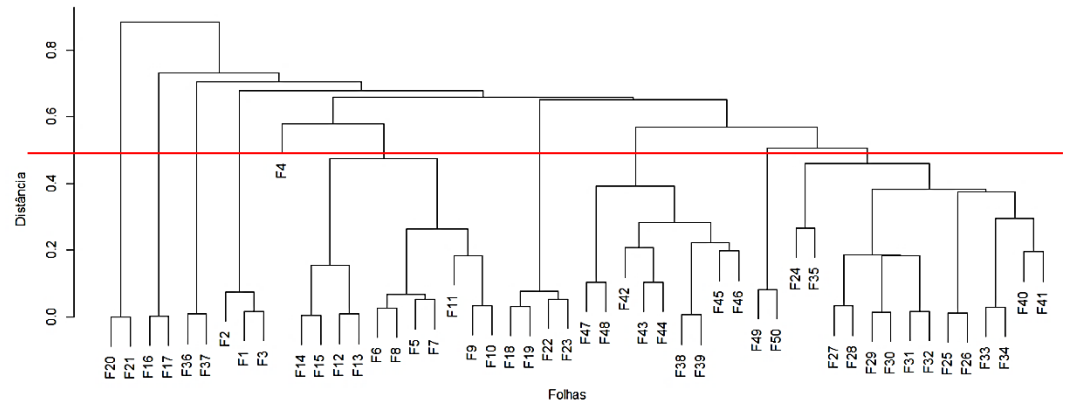

Figura 10.6 – Dendrograma das autotriplas 251 até 300.

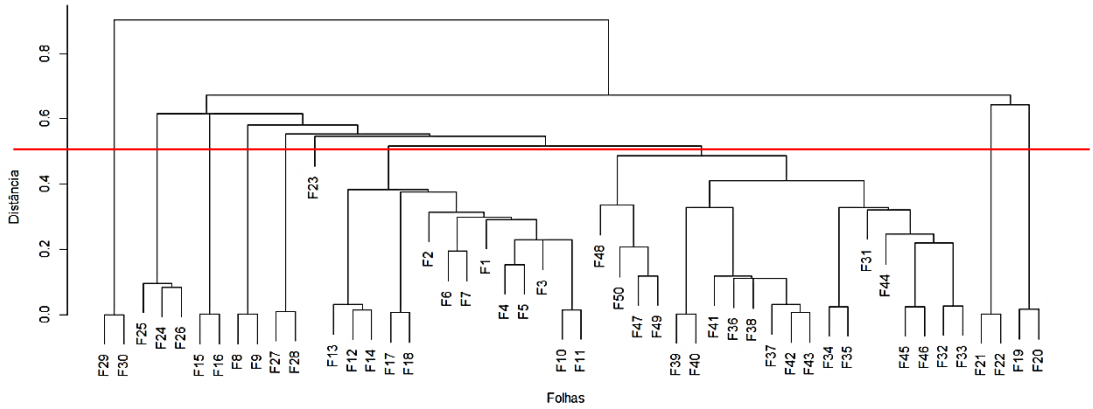

Figura 10.7 – Dendrograma das autotriplas 301 até 350.

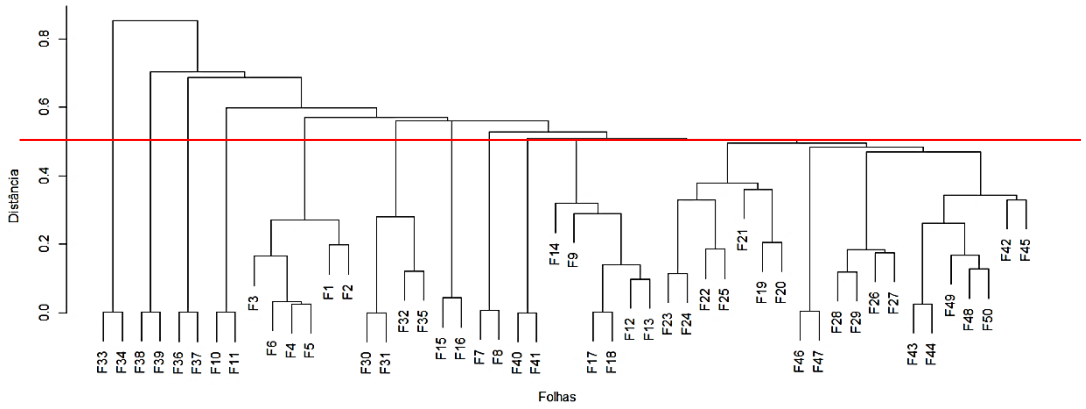

Figura 10.8 – Dendrograma das autotriplas 351 até 400.

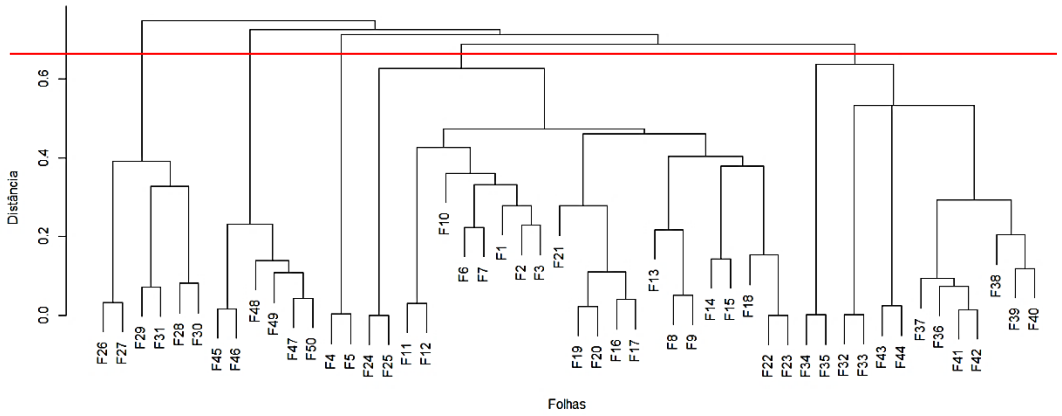

Figura 10.9 – Dendrograma das autotriplas 401 até 450.

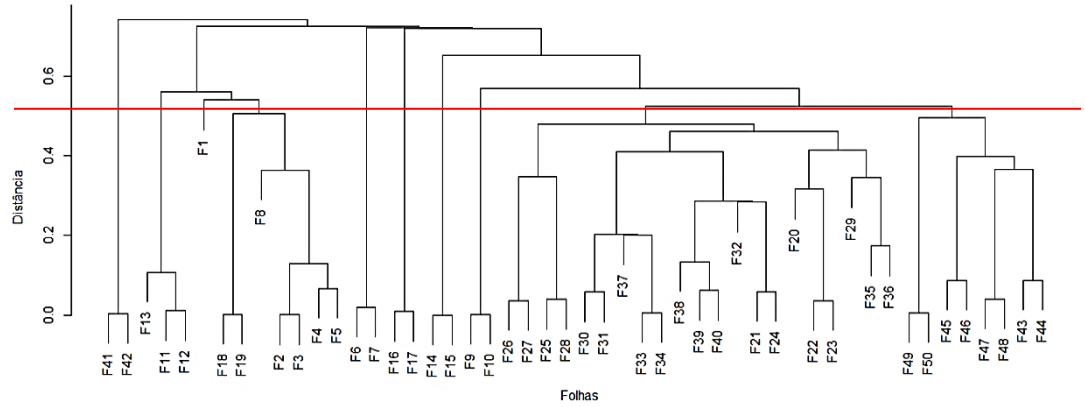

Figura 10.10 – Dendrograma das autotriplas 451 até 500.

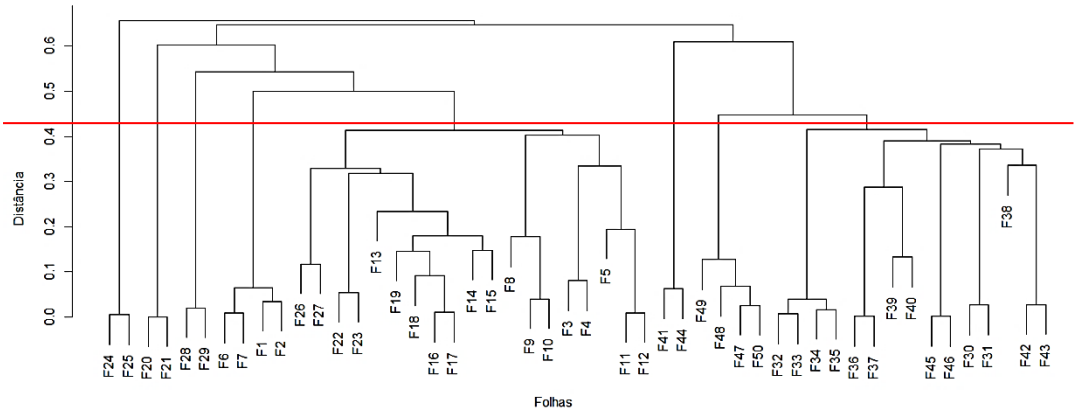

Figura 10.11 – Dendrograma das autotriplas 501 até 550.

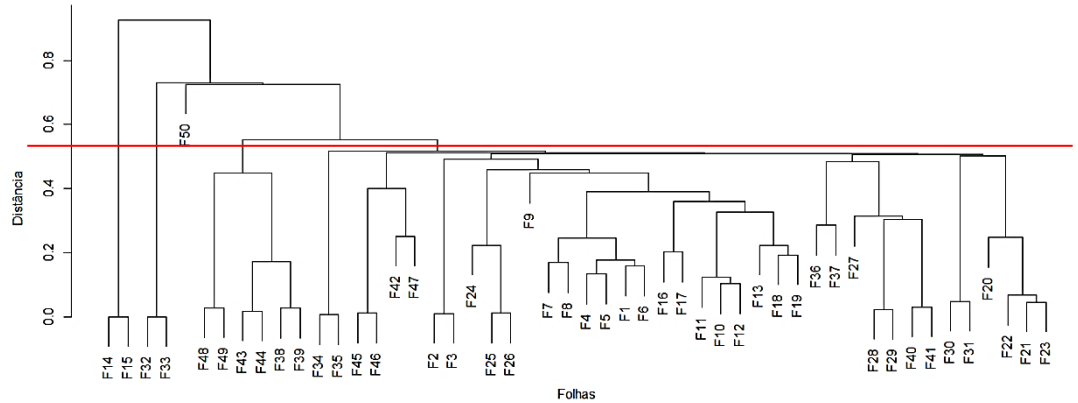

Figura 10.12 – Dendrograma das autotriplas 551 até 600.

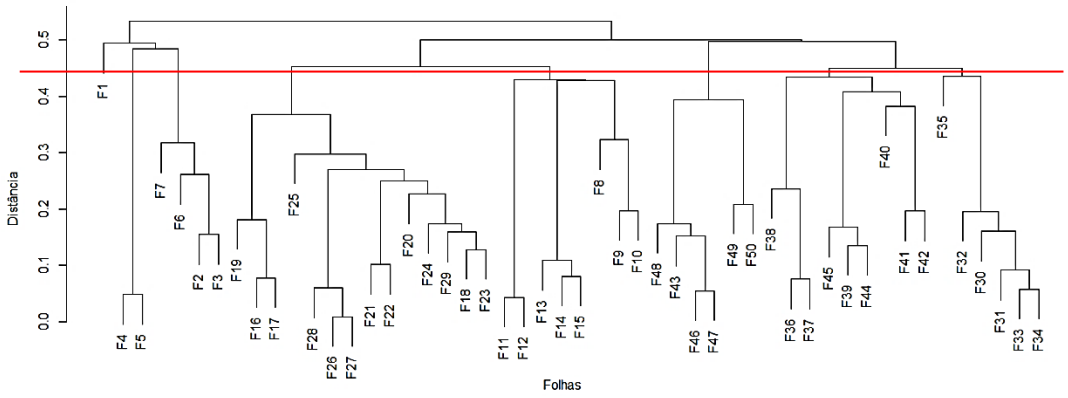

Figura 10.13 – Dendrograma das autotriplas 601 até 650.

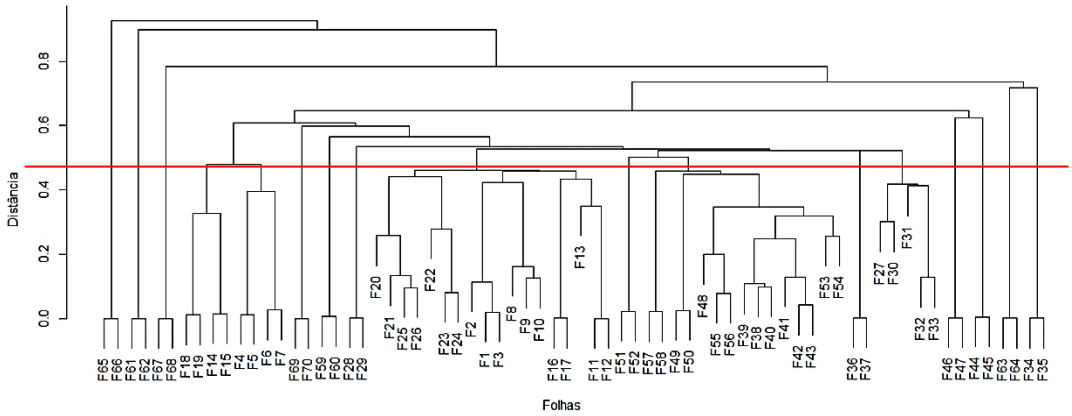

Figura 10.14 – Dendrograma das autotriplas 651 até 720.

10.3.
Apêndice C – Modelos de Box & Jenkins

O modelo ARIMA (*AutoRegressive Integrate Moving Average*) pertence à família de modelos proposto por Box & Jenkins (1970), o qual é o resultado da procura de um modelo que represente o processo estocástico gerador da série temporal. Nestes modelos a série temporal y_t é explicada pelos seus valores passados defasados e o termo do erro estocástico. O modelo ARMA é um submodelo do modelo ARIMA, uma vez que só é aplicável a séries estacionárias, isto é, a série não possui raízes unitárias; enquanto que o modelo ARIMA permite a modelagem de séries que apresentam uma média não-estacionária. A maioria das séries de temporais, especialmente as séries físicas tais como as séries hidrológicas são não-estacionárias, o que quer dizer que a média (e as vezes a variância) da série muda ao longo do tempo. A diferenciação muitas das vezes é usada para remover o componente de tendência de uma série temporal antes da construção do modelo para descrever a série. No entanto, séries temporais como da velocidade do vento são estacionárias, não precisando de diferenciação na parte autorregresiva, este tipo de séries também possui a presença de componentes sazonais, por este motivo os modelos SARIMA (Box e Jenkins, 1976) foram desenvolvidos, os quais incluem a correlação serial dentro e entre os períodos sazonais. Neste caso, a equação do modelo segue:

Geralmente, um processo multiplicativo SARIMA (Box & Jenkins, 1970) pode ser representado pela equação:

$$\phi_p(B)\Phi_P(B^s)\Delta^d\Delta_s^D y_t = c + \theta_q(B)\Theta_Q(B^s)\varepsilon_t \qquad (10.20)$$

sendo $\phi_p(B)$ e $\theta_q(B)$ são os parâmetros não-sazonais do processo Autorregressivo e de Médias Móveis de ordem p e q, respectivamente. $\Phi_P(B^S)$ e $\Theta_Q(B^S)$ são os parâmetros sazonais do processo Autorregressivo e de Médias Móveis de ordem P e Q, respectivamente. B é um operador de defasagem. $\Delta^d = (1 - B)^d$ é o operador de diferença não sazonal de ordem d, enquanto que $\Delta_s^D = (1 - B^s)^D$ é o operador diferença sazonal de ordem D e ε_t corresponde ao termo de erro de eventos aleatórios que não podem ser explicados pelo modelo; c é uma constante, e y_t é a séries de temporal.

O modelo com essa estrutura é denominado SARIMA$(p, d, q) \times (P, D, Q)_s$ e o procedimento para a obtenção desse modelo segue os mesmos passos que a modelagem ARIMA, usada para séries não sazonais.

A continuação são apresentados os resultados do modelo SARIMA multiplicativo para velocidade do vento como passo intermediário para a previsão da geração de energia

eólica. Na Figura 10.2 é ilustrada a função de autorrelação (FAC) e autocorrelação parcial (FACP), a qual indica que esta série temporal possui um padrão sazonal.

Por outra parte, o ajuste do modelo é apresentado na Tabela 10.1. No processo de estimação destaca-se que várias configurações de modelos foram testadas previamente, e escolheu-se o modelo que minimiza os critérios de informação de Akaike, Hannan-Quinn e de Schwarz. O melhor modelo foi o SARIMA(2,0,0)x(1,1,1)$_{24.}$ Uma vez estimado o modelo, este foi utilizado para prever a velocidade do vento 24 horas à frente, isto é, uma previsão horária para o dia 01 de Janeiro de 2008 e 72 horas à frente, os resultados destas previsões são apresentados nas Figuras 7.15 e 7.16 respectivamente para ser comparado em conjunto com os resultados dos outros modelos para previsão da velocidade do vento. As previsões obtidas foram utilizadas para calcular a previsão da geração eólica. Para um melhor detalhamento e análise dos resultados obtidos veja a Seção 7.2.1.

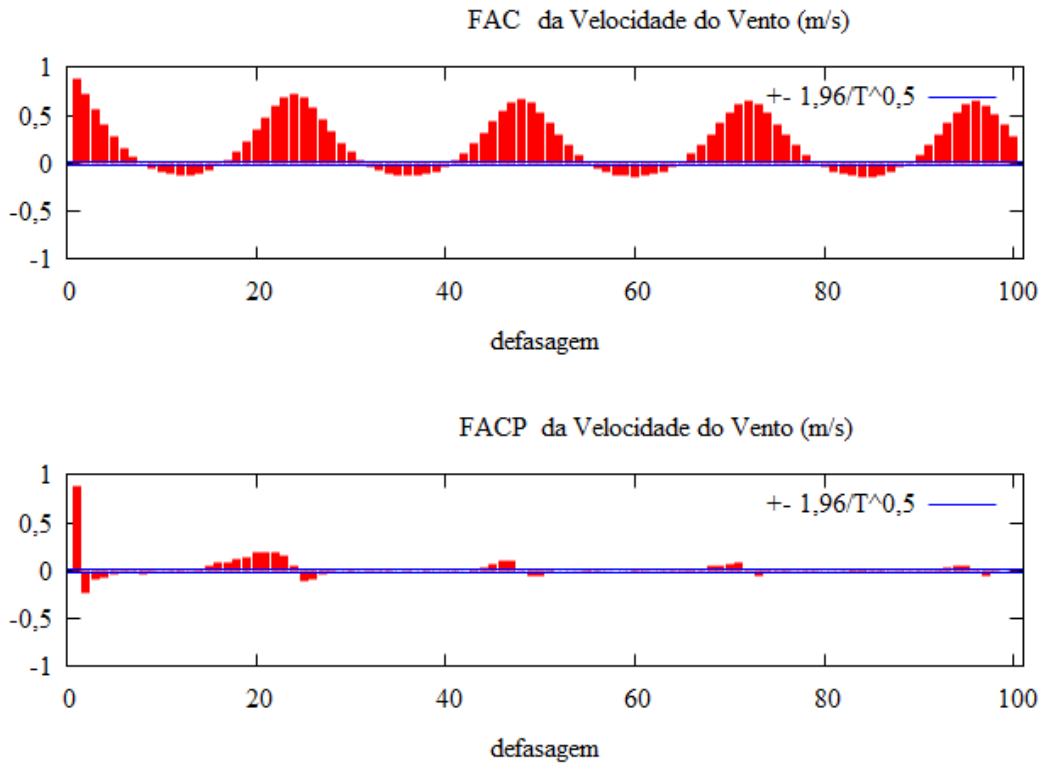

Figura 10.15 – Função de Autocorrelação (FAC) e função de autocorrelação parcial (FACP) da série temporal da velocidade do vento.

Tabela 10.2 – Modelo ajustado SARIMA(2, 0, 0)x$(1,1,1)_{24}$

	coeficiente	erro padrão	z	p-valor	
ϕ1	0,799451	0,0108263	73,84	0,0000	***
ϕ2	-0,0256117	0,0106182	-2,412	0,0159	**
Φ1	0,0679778	0,0116126	5,854	4,80e-09	***
Θ1	-0,941079	0,00428697	-219,5	0,0000	***

Loq da verossimilhança	-16706,96	Critério de Akaike	33423,92
Critério de Schwarz	33459,30	Critério Hannan-Quinn	33435,97

10.4.
Apêndice D – Métodos de alisamento exponencial: Double Seasonal Holt-Winters

O método básico Holt-Winters trabalha com séries que possuem nível, tendência e um único padrão sazonal. A estrutura da sazonalidade do método em sua forma básica é representada pelas expressões (10.21) – (10.24). Existem duas principais variações deste método que diferem na sua componente sazonal, gerando desta forma dois tipologias de métodos: aditivo e multiplicativo. O método aditivo é usado quando as variações sazonais são praticamente constantes através da série, enquanto que o método multiplicativo é empregado quando as variações sazonais estão mudando proporcional ao nível da série.

Nível: $$S_t = \alpha\left(\frac{X_t}{I_{t-s}}\right) + (1-\alpha)(S_{t-1} + T_{t-1}) \tag{10.21}$$

Tendência: $$T_t = \gamma(S_t - S_{t-1}) + (1-\gamma)T_{t-1} \tag{10.22}$$

Sazonalidade: $$I_t = \delta\left(\frac{X_t}{S_t}\right) + (1-\delta)I_{t-s} \tag{10.23}$$

Previsão: $$\hat{X}_t(k) = (S_t + kT_t)I_{t-s+k} \tag{10.24}$$

Sendo X_t, um valor observado, I_t é o índice sazonal local do período S, α, γ e δ são os parâmetros de alisamento e $\hat{x}_t(k)$ é a previsão k passos à frente. Apesar deste método ser

amplamente usado, ele somente pode comportar um único padrão sazonal na sua modelagem. No entanto, existem problemáticas as quais apresentam dois ciclos e, portanto, requerem uma modelagem que inclua esta característica. Por esta razão, Taylor (2003) estendeu o modelo Holt-Winters padrão para que se adequasse com esta situação. As expressões (10.25) – (10.29) apresentam a extensão das equações anteriores, levando em consideração um padrão sazonal adicional (Taylor, 2003) e são denominadas equação de atualização.

$$\text{Nível: } S_t = \alpha\left(\frac{X_t}{D_{t-s_1}W_{t-s_2}}\right) + (1-\alpha)(S_{t-1} + T_{t-1}) \tag{10.25}$$

$$\text{Tendência: } T_t = \gamma(S_t - S_{t-1}) + (1-\gamma)T_{t-1} \tag{10.26}$$

$$\text{Sazonalidade 1: } D_t = \delta\left(\frac{X_t}{S_t W_{t-s_2}}\right) + (1-\delta)D_{t-s_1} \tag{10.27}$$

$$\text{Sazonalidade 2: } W_t = \omega\left(\frac{X_t}{S_t D_{t-s_1}}\right) + (1-\omega)W_{t-s_2} \tag{10.28}$$

$$\text{Previsão: } \hat{X}_t(k) = (S_t + kT_t)D_{t-s_1+k} * W_{t-s_2+k} \tag{10.29}$$

Sendo D_t e W_t os dois tipos de fatores sazonais: diários e intradiários. O modelo aditivo Double Seasonal (D.S.) Holt-Winters pode ser obtido de forma semelhante, a partir do modelo aditivo básico Holt-Winters, através da adição dos correspondentes efeitos sazonais nas equações de atualização.

É importante lembrar que, para a utilização das equações de atualização, devem ser conhecidos os valores dos parâmetros no instante anterior, isto é, em $t - 1$. Devido a isto, é necessário que haja um modo para inicialização desses parâmetros em t_0. Também, é necessário conhecer os valores das constantes de amortecimento.

Desta forma, o modelo D.S. Holt-Winters é usado para fazer previsão da velocidade do vento. Os períodos sazonais diários e intradiário são, respectivamente, $S_1 = 12$ e $S_2 = 24$. Também existem períodos sazonais semanais (S =168), porém como está-se trabalhando um horizonte de previsão de curto prazo, de no máximo 72 horas, este terceiro ciclo não será considerado. Denote-se X_t como sendo a série temporal das

velocidades médias horárias observadas. Para o processo de inicialização dos parâmetros e de previsão foi usada a função "***dshw***" do software **R Project**. Os resultados correspondentes a esta previsão da são apresentados na Seção 7.2.2 e 7.2.3. No entanto na Tabela 10.2 são apresentado valores das constantes de alisamento.

Tabela 10.3 – Valores das constantes de alisamento

Coeficiente	**Valor**
$\boldsymbol{\alpha}$	0.01795
$\boldsymbol{\gamma}$	0.00145
$\boldsymbol{\delta}$	1.8891e-07
$\boldsymbol{\omega}$	0.23845