

UNIVERSIDAD DE CANTABRIA

**DEPARTAMENTO DE CIENCIAS Y TÉCNICAS
DEL AGUA Y DEL MEDIO AMBIENTE**

TESIS DOCTORAL

**METODOLOGÍAS DE CALIBRACIÓN DE BASES DE
DATOS DE REANÁLISIS DE CLIMA MARÍTIMO**

Presentada por: ANTONIO TOMÁS SAMPEDRO

**Dirigida por: FERNANDO J. MÉNDEZ INCERA
IÑIGO J. LOSADA RODRÍGUEZ**

Mayo, 2009

REFERENCIAS

REFERENCIAS

Referencias.

- AMF, 1991;** Estudio sobre las observaciones visuales de barcos en ruta del N.C.D.C *Publicación del Departamento de Clima Marítimo de Puertos del Estado (Área de Medio Físico del OPPE)*, 46 pp.
- Baldacci, A., Corsini, G., Grasso, R., Manzella, G., Allen, J.T., Cipollini, P., Guymer, T.H., Snaith, H.M., 2001;** A study of the Alboran sea mesoscale system by means of empirical orthogonal function decomposition of satellite data. *Journal of Marine Systems*, 29. pp 293-311. doi: 10.1016/S0924-7963(01)00021-5.
- Barnett, V., Lewis, T., 1994;** Outliers in statistical data. *John Wiley and Son, Chichesler, New York*. 584 pp. ISBN: 0471930946.
- Barstow, S., Haug, O., Krogstad, H., 1998;** Satellite altimetry data in wave energy studies. *Proceedings of the Third International Symposium on Ocean Wave Measurement and Analysis. ASCE*. pp 339-354.
- Bauer, E.S., Hasselmann, S., Hasselmann, K., Graber, H., 1992;** Validation and Assimilation of Seasat Altimeter Wave Heights Using the WAM Wave Model. *Journal of Geophysical Research*, 97(C8). pp 12671-12682. doi: 10.1029/92JC01056.
- Boggs, P.T., Rogers, J.E., 1990;** Orthogonal distance regression. *Contemporary Mathematics. American Mathematical Society*, 112. pp 183-194. ISBN: 0-8218-5117-9.
- Booij, N., Ris, R.C., Holthuijsen, L.H., 1999;** A third-generation wave model for coastal regions, Part I: Model description and validation. *Journal of Geophysical Research*, 104(C4). pp 7649-7666. doi: 10.1029/98JC026222.
- Booij, N., Haagsma, I.J.G., Holthuijsen, L.H., Kieftenburg, A.T.M.M., Ris, R.C., van der Westhuysen, A.J., Zijlema, M., 2004;** User Manual for SWAN, version 40.41. *Delft University of Technology, the Netherlands*. 115 pp.
- Borgman, L.E, 1984;** Directional spectrum estimation for the Sxy gauges. *Technical Report CERC. US Army Corps of Engineers, Waterways Experiment Station, Vicksburg, MS, USA*. 104 pp.

REFERENCIAS

- Camus, P., Vidal, C., Méndez, F.J., Espejo, A., Izaguirre, C., Gutierrez, J.M, Cofiño, A., San-Martin, D., Medina, R., 2007;** A methodology to evaluate wave energy resources in shallow waters. *Proc. 7th European Wave and Tidal Energy Conference, Porto, Portugal.* pp 1-10.
- Caires, S., Sterl, A., Biblot, J.-R., Graham, N., Swail, V., 2002;** Climatological assessment of reanalysis of wave data. *Proc. 7th Int. workshop on Wave Hindcasting and Forecasting, Banff, Canada.* pp 1-12.
- Caires, S., Sterl, A., 2003;** Validation of ocean wind and wave data using triple collocation. *Journal of Geophysical Research, 108(C3), pp 3098.* doi: 10.1029/2002JC001491.
- Caires, S., Sterl, A., 2005;** A new nonparametric method to correct model data: application to significant wave height from the ERA-40 Re-Analysis. *Journal of Atmospheric and Oceanic Technology., 22(4), pp 443-459.* doi: 10.1175/2004JC002631.
- Caires, S., Sterl, A., Gommenginger, C.P., 2005;** Global ocean mean wave period data: Validation and description. *Journal of Geophysical Research, 110(C2).* pp 1-12. doi: 10.1029/2004JC002631.
- Caires, S., Ferreira J.A., 2005;** On the Non-parametric Prediction of Conditionally Stationary Sequences. *Statistical Inference for Stochastic Processes, 8.* pp 151-184. doi: 10.1007/s11203-004-0383-2. Correction, *vol. 9 (2006).* pp 109-110. doi: 10.1007/s11203-006-0005-2.
- Carter, D.J.T., Challenor, P.G., Srokosz, M.A., 1992;** An Assessment of GEOSAT Wave Height and Wind Speed Measurements. *Journal of Geophysical Research, 97(C7), pp 11383-11392.* doi: 10.1029/92JC00465.
- Cartwright, E.D., 1964;** A Comparison of Instrumental and Visually Estimated Wave Heights and Periods Recorded on Ocean Weather Ships. *Report of the Third International Ship Structures Congress, Det Norske Veritas.*
- Castillo, E., 1993;** Introducción a la estadística aplicada con mathematica. *Santander, 374 pp.* ISBN: 8460402991.

Cavaleri, L., Sclavo, M., 2006; The calibration of wind and wave model data in the Mediterranean Sea. *Coastal Engineering*, 53(7), pp 613-627. doi: 10.1016/j.coastaleng.2005.12.006.

CEDEX, 2008; Informe Técnico para Puertos del Estado: Calibración de los temporales sintéticos de las fachadas Noratlántica y Cantábrica y estimación de la distribución bivariada Hm0-Tm02 de los temporales calibrados y registrados en estas fachadas marítimas. Modelos Probabilistas de Fiabilidad de Estructuras. *Centro de Estudios de Puertos y Costas (Ministerio de Fomento)*. Clave CEDEX: 24-308-5-001

Challenor, P.G., Cotton, P.D., 2002; The joint calibration of altimeter and in situ wave heights. *Advances in the Applications of Marine Climatology. The Dynamic Part of the WMO Guide to the Applications of Marine Climatology, WMO/TD-Nº 1081 JCOMM Tech. Rep. Nº 13, World Meteorol. Organ., Geneva.*

Chelton, D.B., 1988; U.S. World Ocean Circulation Experiment Technical Report Nº2. *WOCE/NASA Altimeter Algorithm Workshop. 70 pp.*

Cieslikiewicz, W., Graff, J., 1996; Sea state parameterisation using empirical orthogonal functions. *Proceedings of the twenty fifth international conference. Coastal Engineering, ICCE, 1. pp 703-716.*

Cotton, P.D., Carter, D.J.T., 1994; Cross calibration of TOPEX, ERS-1 and GEOSAT at wave heights. *Journal of Geophysical Research*, 99, pp 25025-25033. doi: 10.1029/94JC02131.

Cotton, P.D., Challenor, P.G., Carter, D.J.T., 1997; An assessment of the accuracy and reliability of GEOSAT, ERS-1, ERS-2 and Topex altimeter measurements of significant wave height and wind speed. *Proceedings CEOS Wind and Wave Valid. Workshop, ESA, ESTEC, The Netherlands, pp 81-93.*

Cotton, P.D., Challenor, P.G., Carter, D.J.T., 1998; Variability in altimetry global wave climate data. *Proceedings of the Third International Symposium on Ocean Wave Measurement and Analysis. ASCE. pp. 809-818.*

Cox, A.T., Swail, V.R., 2001; A global wave hindcast over the period 1958-1997: Validation and climate assessment. *Journal of Geophysical Research*, 106(C2). pp. 2313-2329. doi: 10.1029/2001JC000301.

REFERENCIAS

Crosby, D.S., Breaker, L.C., Gemmill, W.H., 1993; A Proposed Definition for Vector Correlation in Geophysics: Theory and Application. *Journal of Atmospheric and Oceanic Technology*, 10 (3). pp 355-367. doi: 10.1175/1520-0426(1993)010<0355:APDFVC>2.0.CO;2.

Davies, C.G., Challenor, P.G., Cotton, P.D., 1998; Measurements of wave period from radar altimeters. *Proceedings of the Third International Symposium on Ocean wave measurements and analysis. vol 2 ASCE. pp 819-826.*

Dacunha, N.M.C., Hogben, N., Andrews, K.S., 1984; Wave climate synthesis worldwide. *Proc. Int. Symp. On Wave and Wind Climate Worldwide, London, pp 1-11.*

DelBalzo, D.R., Schultz, J.R., Earle, M.D., 2003; Stochastic time-series simulation of wave parameters using ship observations. *Ocean Engineering*, 30. pp 1417-1432. doi: 10.1016/S0029-8018(02)00120-8.

Dinsmore, R.P., 2005; Ocean Weather Ships 1940-1980. *U.S Coast Guard.* http://www.uscg.mil/History/webcutters/rpdinsmore_oceanstations.html.

Draper, N.R., Smith, H., 1998; Applied regression analysis. *Wiley Series in Probability and Statistics.* 736 pp.

Duan, Q., Sorooshian, S., Gupta, V., 1992; Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*, 28(4). pp 1015-1031. ISSN: 0043-1397.

Fu L.-L., Cazedane, A., 2001; Satellite Altimetry Earth Sciences: A Handbook of Techniques. *Academic Press.* 463 pp. ISBN: 0122695453.

GIOC, 1993; Estudio del régimen extremal del oleaje en el archipiélago Canario para el Gobierno de Canarias. *Fundación Leonardo Torres Quevedo, Grupo de Ingeniería Oceanográfica y de Costas (GIOC).*

GIOC, 1999; Documento complementario al documento temático atlas de inundación del litoral peninsular español. Modelo de ayuda a la gestión del litoral. *Dirección General de Costas (Ministerio de Medio Ambiente). Grupo de Ingeniería Oceanográfica y de Costas (Universidad de Cantabria).* 95 pp. <http://www.smc.unican.es>

GIOC, 2001; Manual de referencia del modelo de ayuda a la gestión del litoral español. Modelo de propagación de oleaje espectral (OLUCA-SP 2.0). *Dirección General de Costas (Ministerio de Medio Ambiente). Grupo de Ingeniería Oceanográfica y de Costas (Universidad de Cantabria)*. 170 pp. <http://www.smc.unican.es>

GIOC, 2004; Impactos en la costa española por efecto del cambio climático. *Dirección General de Calidad y Evaluación Ambiental (Ministerio de Medio Ambiente). Grupo de Ingeniería Oceanográfica y de Costas (Universidad de Cantabria)*.

Gommenginger, C.P., Srokosz, M.A, Challenor, P.G., Cotton, P.D., 2003; Measuring ocean wave period with satellite altimeters: A simple empirical model. *Geophysical Research Letters.*, 30(22). pp 2150. doi 10.1029/2003GL017743.

Gower, J.F.R., 1979; The computation of ocean wave heights from GEOS-3 satellite radar altimeter data. *Remote Sensing of Environment*, 8(2). pp 97-114. doi: 10.1016/0034-4257(79)90011-7.

Gower, J.F.R., 1996; Intercalibration of wave and wind data from TOPEX/POSEIDON and moored buoys off the west coast of Canada. *Journal of Geophysical Research*, 101(C2). pp 3817-3829. doi: 10.1029/96JC03281.

Guedes Soares, C., Weisse, R., Carretero, J.C., Álvarez, E., 2002; A 40 years hindcast of wind, sea level and waves in European waters. *Proceedings of the 21st International Conference on Offshore Mechanics and Arctic Engineering (OMAE 2002) ASME, New York. Paper OMAE2002-28604.*

Guedes Soares, C., 2008; Hindcast of Dynamic Processes of the Ocean and Coastal Areas of Europe. *Coastal Engineering. Special Issue* .pp 1-3. doi: 10.1016/j.coastaleng.2008.02.007.

Gulev, S.K., Cotton, D., Sterl, A., 1998; Intercomparison of North Atlantic Wave Climatology from Voluntary Observing Ships, Satellite Data and Modelling. *Physics and Chemistry of the Earth*, 23(5/6), pp 587-592. doi: 10.1016/S0079-1946(98)00075-5.

Gulev, S.K., Hasse, L., 1998; North Atlantic winds waves and wind stress fields from voluntary observing ship data. *Journal of Physical Oceanography*, 28(6), pp 1107-1130. doi: 10.1175/1520-0485(1998)028<1107:NAWWAW>2.0.CO;2.

REFERENCIAS

Gulev, S.K., Grigorieva, V., Sterl, A., Woolf, D., 2003; Assessment of the reliability of wave observations from voluntary observing ships: Insights from the validation of a global wind wave climatology based on voluntary observing ship data. *Journal of Geophysical Research*, 108(C7), pp 3236. doi: 10.1029/2002JC001437.

Günther, H, Rosenthal, W., Stawarz, M., Carretero, J.C., Gómez, M., Lozano, I., Serrano, O., Reistad, M., 1998; The wave climate of the Northeast Atlantic over the period 1955-1994: the WASA wave hindcast. *The Global Atmosphere and Ocean System*, 6(2), pp 121-163. ISSN: 1023-6732.

Gurgel, K.W., Essen, H.H., Schlick, T., 2001; Applications of WERA within the EuroROSE project. *Proceedings of First International Radiowave Oceanography Workshop, ROW 2001.*

Gutiérrez, J.M., Cano, R., Cofiño, A.S., Sordo, C.M., 2004; Redes Probabilísticas y Neuronales en las Ciencias Atmosféricas. *Universidad de Cantabria. Instituto Nacional de Meteorología.* 279 pp. ISBN: 8483202816. <http://grupos.unican.es/ai/meteo/>

Hasselmann, K., Barnett, T.P., Bouws E., Carlson, H., Cartwright, D.E., Enke, K., Ewing, J.A., Gienapp, H., Hasselmann, D.E., Kruseman, P., Meerburg, A., Muller P., Olbers, D.J., Richter, K., Sell, W., Walden, H., 1973; Measurement of wind-wave grown and swell decay during the Joint North Sea Wave Project (JONSWAP). *Report German Hydrographic Institute, Hamburg.*

Hogben, N., Lumb, F.E., 1967; Ocean Wave Statistics. *Her Majesty's Stationery Office. London U.K.* 263 pp.

Holthuijsen, L.H., 2007; Waves in Oceanic and Coastal Waters. *Cambridge University Press.* 387 pp. ISBN: 0521860288.

Hughes, S.A., 1984; The TMA shallow-water spectrum description and applications. *Technical Report CERC-84-7. US Army Corps of Engineers, Waterways Experiment Station, Vicksburg, MS, USA.* 45 pp

Hwang, P.A., Teague, W.J., Jacobs, G. A., Wang, D.W., 1998; A statistical comparison of wind speed, wave height, and wave period derived from satellite altimeters and ocean buoys in the Gulf of Mexico region. *Journal of Geophysical Research* 103(C5). pp 10451-10468. doi: 10.1029/98JC00197.

Janssen, P.A.E.M., Abdalla, S., Hersbach, H., 2003; Error estimation of buoy, satellite and model wave height data. *Technical Memorandum, 402, ECMWF 17 pp.*

Janssen, P.A.E.M., 2004; The Interaction of Ocean Waves and Wind. *Cambridge University Press. 300 pp.*

Jardine, T.P., 1979; The Reliability of Visually Observed Wave Heights. *Coastal Engineering, 3. pp 33-38.*

Luceño, A., 1989; Métodos de estadística Aplicada. *Universidad de Cantabria. 227 pp.* ISBN: 8486928028.

Marsden, R.F., 1999; A Proposal for a Neutral Regression. *Journal of Atmospheric Technology, 16(7). pp 876-883. doi: 10.1175/1520-0426(1999)016<0876:APFANR>2.0.CO;2.*

Massel, S.R., 1996; Ocean Surface Waves: Their Physics and Prediction. *Advanced Series on Ocean Engineering, 11. World Scientific. 491 pp.* ISBN: 981-02-2106-6.

Menéndez, M., 2008; Metodología para el análisis estadístico no estacionario de valores extremos de variables geofísicas. *Tesis Doctoral. Universidad de Cantabria.*

Nordenstrom, N., 1969; Methods for predicting long term distributions of wave loads and probability of failure ships, Appendix II, Relationships between visually estimated and theoretical wave heights and periods. *Research Report 69-22-S, Det Norske Veritas.*

NWSOH N°1, 2004; National Weather Service. Observing Handbook N°1. *Marine Surface Weather Observations, 150 pp.*

Ochi, M.K., 1978; Wave statistics for the design of ships and offshore structures. *Proc. Science in Naval Architecture and Marine Environment, New York, pp 47-74.*

Ochi, M.K., 1998; Ocean Waves: the Stochastic Approach. *Cambridge University Press. 319 pp.* ISBN: 0-521-56378-X.

REFERENCIAS

OPPE, 2003; Reanálisis de 44 años (1958-2001) del clima oceánico y atmosférico en el mar Mediterráneo: informe técnico de la contribución de Puertos del Estado al proyecto Europeo HIPOCAS. *Publicación del Departamento de Clima Marítimo de Puertos del Estado (Área de Medio Físico del OPPE)*, 305 pp.

Queffeuou, P., Cotton, P.D., 2002; Consistency of wind and wave data from pasta and present altimetric missions. *Symposium En Route vers GODAE. Biarritz. France.*

Quilfen, Y., Chapron, B., Collard, F., Serre, M., 2005; Calibration/validation of an altimeter wave period model and application to Topex/Poseidon and Jason-1 altimeters. *Marine Geodesy*, 27. pp 535-550.

Ratsimandresy, A.W., Sotillo, M.G., Carretero Albiach, J.C., Álvarez Fanjul, E., Hajji, H., 2008; A 44-year high-resolution ocean and atmospheric hindcast for the Mediterranean Basin developed within the HIPOCAS Project. *Coastal Engineering. Special Issue .pp 1-3.* doi: 10.1016/j.coastaleng.2008.02.007.

Ray, R.D., Beckley, B.D., 2003; Simultaneous Ocean Wave Measurements by the Jason and Topex Satellites, with Buoy and Model Comparisons. *Marine Geodesy*, 26(3/4). pp 367-382. doi: 10.1080/01490410390251614 (AN 11794491).

Reichert, K., Hessner, K., Dannenberg, J., Tränkmann, I., Björn, L., 2005; X-Band Radar as tool to determine spectral and single wave properties. *Proceedings of the Fifth International Symposium on Ocean Wave Measurement and Analysis. WAVES, Madrid.*

ROM 0.0, 2001; Procedimiento general y Bases de Cálculo. *Recomendaciones Obras Marítimas. Puertos del Estado. 207 pp* ISBN: 84-88975-30-9. http://www.puertos.es/es/programa_rom

ROM 0.3-91, 1992; Acciones Climáticas I: Oleaje. *Recomendaciones Obras Marítimas. Puertos del Estado. 77 pp.* http://www.puertos.es/es/programa_rom

Soares, C.G., 1986a; Assessment of the uncertainty in visual observations of wave height. *Ocean Engineering*, 13(1). pp 37-56.

Soares, C.G., 1986b; Calibration of visual observation of wave period. *Ocean Engineering*, 13(6). pp 539-547.

Soukissian, T., Kechris, C., 2007; About applying linear structural method on ocean data: Adjustment of satellite wave data. *Ocean Engineering*, 34(3/4). pp 371-389. doi: 10.1016/j.oceaneng.2006.04.002.

Sterl, A., Komen, G.J., Cotton, P.D., 1998; Fifteen years of global wave hindcasts using winds from the European Centre for Medium-Range Weather Forecast reanalysis: Validating the reanalyzed winds and assessing the wave climate. *Journal of Geophysical Research*, 103(C3). pp 5477-5492. doi: 10.1029/97JC03431.

Swail V.R., Cox, A.T., 2000; On the use of NCEP/NCAR reanalysis surface marine wind fields for a long-term North Atlantic wave hindcast. *Journal of Atmospheric and Oceanic Technology*, 17(4). pp 532-545. doi: 10.1175/1520-0426(2000)017<0532:OTUONN>2.0.CO;2.

Taylor, K.E., 2001; Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research*, 106(D7). pp 7183-7192. doi: 10.1029/2000JD900719.

Tolman, H.L., 1999; User manual and system documentation of WAVEWATCH-III version 1.18. *NOAA / NWS / NCEP / OMB Technical Note 166*, 110 pp.

Tolman, H.L., 2002; User manual and system documentation of WAVEWATCH-III version 2.22. *NOAA / NWS / NCEP / MMAB Technical Note 222*, 139 pp. (disponible en http://polar.wwb.noaa.gov/mmab/papers/tn222/MMAB_222.pdf)

Tomás, A., Méndez, F.J., López, J.D., Losada, I.J., Rodríguez, I., 2009; Nueva metodología de calibración direccional no paramétrica. *X Jornadas Españolas de Ingeniería de Costas y Puertos, Santander*. Aceptado.

Tomás, A., Méndez, F.J., Losada, I.J., 2008; A method for spatial calibration of wave hindcast data bases. *Continental Shelf Research*, 28(3). pp 391-398. ISSN: 0278-4343. doi: 10.1016/j.crs.2007.09.009.

Tomás, A., Méndez, F.J., Losada, I.J., 2006; Calibración espacial de regímenes medios mensuales de oleaje a partir de datos de reanálisis: aplicación al Mediterráneo. *Ingeniería del Agua*, 13(3). pp 203-218. ISSN: 1134-2196.

REFERENCIAS

Young, I.R., 1999; An intercomparison of GEOSAT, TOPEX and ERS-1 measurements of wind speed and wave height. *Ocean Engineering*, 26. pp 67-81. doi: 10.1016/S0029-8018(97)10016-6.

Uppala, S., 2001; ECMWF ReAnalysis, 1957-2001. *Proc. of the ECMWF workshop on re-analysis, ERA-40 Project Report Series, 3. pp 1-10.*

Von Storch, H., Zwiers, Z.W., 2001; Statistical Analysis in Climate Research. *Cambridge University Press.*

Vrugt, J.A., Gupta, H.V., Bouten, W., Sorooshian, A., 2003; A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resources Research*, 39(8). 1201. doi: 10.1029/2002WR001642.

WAMDI Group: Hasselmann, S., Hasselmann, K., Bauer, E., Janssen, P.A.E.M., Komen, G.J., Bertotti, L., Lionello, P., Guillaume, A., Cardone, V.C., Greenwood, J.A., Reistad, M., Zambresky, L., Ewing, J.A, 1988; The WAM model – a third generation ocean wave prediction model. *Journal of Physical Oceanography, Vol 18. pp 1775-1810.*

Woodruff, S.D., Slutz, R.J., Jenne, R.L., Steurer, P.M., 1987; A comprehensive ocean-atmosphere data set. *Bulletin of the American Meteorological Society*, 68(10). pp1239-1250. doi: 10.1175/1520-0477(1987)068<1239:ACOADS>2.0.CO;2.

Wilkerson, J.C., Earle, M.D., 1990; A study of differences between environmental reports by ships in voluntary observing program and measurements from NOAA buoys. *Journal of Geophysical Research*, 95(C3). pp 3373-3385.

Wong, M.Y., 1989; Likelihood estimation of simple linear regression model when both variables have error. *Biometrika*, 76(1). pp 141-148. <http://www.jstor.org/pss/2336378>.

REFERENCIAS

ANEJO I

DISTRIBUCIONES ESTADÍSTICAS DE RÉGIMEN MEDIO

A.I.1. Introducción.

Para caracterizar el régimen medio (distribución media de probabilidad) de un determinado parámetro de estado de mar, generalmente se ajustan una serie de datos a una función de distribución dada. A continuación se van a enumerar las distintas funciones de distribución que normalmente más se utilizan para ajustar regímenes medios de oleaje, así como la obtención de sus parámetros a través del método de los papeles probabilísticos.

Previamente a la definición de diversas funciones de distribución que sirven para caracterizar el régimen medio de oleaje y las distorsiones necesarias de cada una para representarlas en sus respectivos papeles probabilísticos, se plantea el problema de la representación de la función de distribución empírica. Así, el ajuste de la función de distribución a unos datos, o lo que es lo mismo, la probabilidad de no excedencia que se asigna a cada valor, se puede hacer de dos maneras; la primera es asignando la probabilidad a cada dato individual (ver problema del punteo a continuación), cuando el número de datos es pequeño, pero cuando existe gran cantidad de datos (por ejemplo, los datos de reanálisis) se suele asignar la probabilidad que agrega cada clase o grupo de datos, en el capítulo 3 (apartado 3.2.4.2) se presentan varias formas de seleccionar dichos grupos de datos.

A.I.2 Problema del punteo.

Es bien conocido que la función distribución empírica, $F(x)$, de una variable aleatoria X es una función escalera que toma los valores de probabilidad de no excedencia $1/n, 2/n, \dots, i/n, \dots, 1$ (ver figura A.I.1).

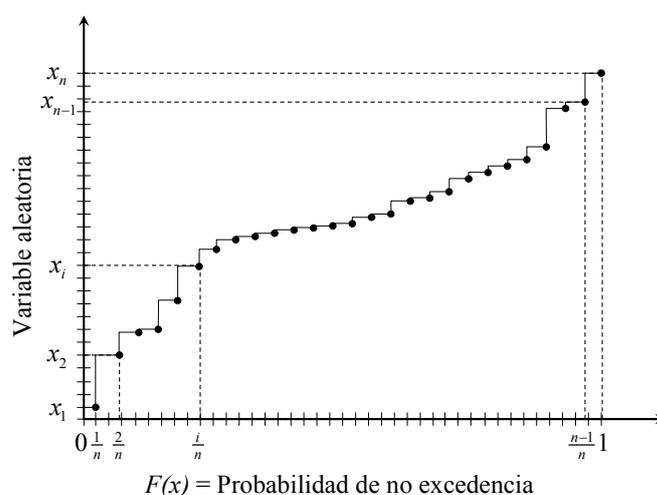


Figura A.I.1. Función distribución empírica de la variable aleatoria X .

Sin embargo los dos extremos de la distribución empírica 0 y 1, en el caso de muchas familias de funciones de distribución, se transforman en $-\infty$ o ∞ respectivamente al dibujarlos en su papel probabilístico correspondiente, este problema es el usualmente denominado problema del punteo.

Una solución de compromiso a este problema, propuesta por Hazen, consiste en usar la aproximación $(i-1/2)/n$ en lugar de i/n , debido a que la probabilidad salta desde $(i-1)/n$ a i/n . Otras posibles aproximaciones clásicas ($F'(x)$) se muestran en la tabla A.I.1 (recopilada de Castillo (1993)):

	Expresión	Autor
$(x_i, F(x_i))$ $F(x) = \frac{i}{n}$	$F'(x) = \frac{i}{n+1}$	-
	$F'(x) = \frac{i - \frac{3}{8}}{n + \frac{1}{4}}$	Blom
	$F'(x) = \frac{i - \frac{1}{2}}{n}$	Hazen
	$F'(x) = \frac{i - 0.44}{n + 0.12}$	Gringorten
	$F'(x) = \frac{3i - 1}{3n + 1}$	-

Tabla A.I.1. Diferentes fórmulas de punteo

A.I.3 Funciones de distribución y papeles probabilísticos.

La función distribución de una variable aleatoria X es una función que asigna a cada valor de x la probabilidad de que la variable aleatoria tome valores iguales o menores que x ($F(x) = \Pr(X \leq x)$; $0 < F(x) < 1$). Existen multitud de funciones de distribución (Normal, Lognormal, Exponencial, Weibull (Fisher Tippett III), Gumbel (Fisher Tippett I), Frechet (Fisher Tippett II),...) con diversas variaciones. Así, cuando la variable aleatoria es un parámetro de estado de mar, la función distribución de dicha variable sirve para caracterizar su régimen medio. A continuación se presentan las funciones de distribución más usuales para definir el régimen medio de oleaje, pero también se va a presentar una función de distribución generalmente usada para caracterizar el régimen extremal, la GEV (*Generalized Extreme Value*), que es la forma de Von Mises para máximos (Castillo, 1993).

Además de definir distintas funciones de distribución para ajustar el régimen medio, así como sus funciones de densidad asociadas ($f(x) = dF(x)/dx$), también se especifican las

transformaciones necesarias de cada una de las distribuciones para representar sus papeles probabilísticos.

El método de los papeles probabilísticos se utiliza para estimar los parámetros de una función de distribución ($F(x; \theta)$). En general, se buscan transformaciones de los ejes ($g(x)$ y $h(y) = h(F(x))$) para que en los nuevos ejes (o variables reducidas) la función de distribución sea una recta, ver el esquema de la figura A.I.2. Finalmente ajustando los parámetros de la recta ($\hat{\Theta} = \{\alpha, \beta\}$) a los datos, por el método OLS (mínimos cuadrados, ver en el apartado 2.3.2.1), se determinan los parámetros de la función de distribución.

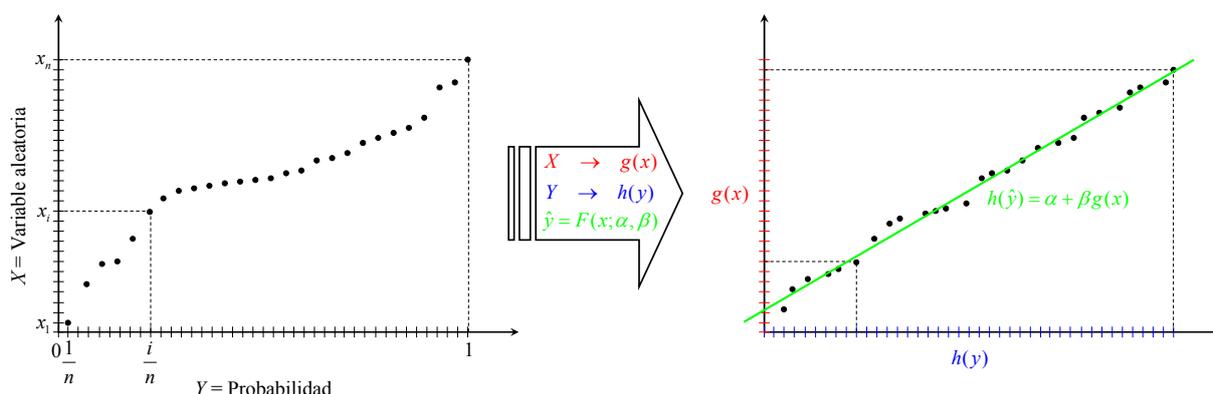


Figura A.I.2. Método gráfico de los papeles probabilísticos para la estimación de la función de distribución de la variable aleatoria X .

A.I.3.1 Distribución Lognormal.

Se dice que una variable aleatoria X es Lognormal si su logaritmo neperiano es Normal. Haciendo uso del cambio de variable, se obtiene fácilmente que la función de distribución de X es:

$$y = F(x; \mu^*, \sigma^*) = \Phi\left(\frac{\log(x) - \mu^*}{\sigma^*}\right) = \frac{1}{\sigma^* \sqrt{2\pi}} \int_0^x \frac{1}{x} \exp\left[-\frac{1}{2}\left(\frac{\log(x) - \mu^*}{\sigma^*}\right)^2\right] dx; \quad 0 < x < \infty$$

(A.I.1)

donde:

$\Phi(z)$ es la función de distribución de la variable z Normal estándar $N(0,1)$

μ^* es la media de la distribución Normal del logaritmo de la variable original (parámetro de localización de la distribución Lognormal)

σ^* es la desviación típica de la distribución Normal del logaritmo de la variable original (parámetro de escala de la distribución Lognormal).

Siendo la función densidad Lognormal para una variable aleatoria X :

$$f(x; \mu^*, \sigma^*) = \frac{dF(x; \mu^*, \sigma^*)}{dx} = \frac{1}{x\sigma^* \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{\log(x) - \mu^*}{\sigma^*} \right)^2 \right]; \quad 0 < x < \infty \quad (\text{A.I.2})$$

Para el papel probabilístico Lognormal las variables reducidas se definen a partir de las transformaciones de la función de distribución siguientes:

$$\begin{aligned} \xi &= g(x) = \log(x) \\ \eta &= h(y) = \Phi^{-1}(y) \end{aligned} \quad (\text{A.I.3})$$

con lo que la familia de rectas ($\Theta = \{\alpha, \beta\}$) queda:

$$\eta = \alpha + \beta\xi = \frac{\xi - \mu^*}{\sigma^*}; \quad \text{con } \alpha = -\frac{\mu^*}{\sigma^*} \text{ y } \beta = \frac{1}{\sigma^*} \quad (\text{A.I.4})$$

A partir de μ^* y σ^* de la distribución Normal original (ecuación A.I.1) se pueden determinar la media μ y la desviación típica σ de la distribución Lognormal:

$$\begin{aligned} \mu &= \exp \left(\mu^* + \frac{\sigma^{*2}}{2} \right) \\ \sigma^2 &= \exp \left(2\mu^* + 2\sigma^{*2} \right) - \exp \left(2\mu^* + \sigma^{*2} \right) \end{aligned} \quad (\text{A.I.5})$$

y despejando μ^* y σ^* de μ y σ de las expresiones A.I.5 se tiene:

$$\begin{aligned} \mu^* &= \log(\mu) - \frac{1}{2} \log \left(\frac{\sigma^2}{\mu^2} + 1 \right) \\ \sigma^{*2} &= \log \left(\frac{\sigma^2}{\mu^2} + 1 \right) \end{aligned} \quad (\text{A.I.6})$$

Cabe señalar, que para el caso de la función de distribución Lognormal, ajustando mediante OLS, se suele utilizar la fórmulas de punteo de Blom, ver tabla A.I.1.

A.I.3.2 Distribución Gumbel de máximos.

La función de distribución Gumbel de máximos de una variable aleatoria X es:

$$y = F(x; \mu, \psi) = \exp\left[-\exp\left(-\frac{x-\mu}{\psi}\right)\right]; \quad -\infty < x < \infty$$

(A.I.7)

donde:

μ es el parámetro de localización (es la moda de la distribución)

ψ es el parámetro de escala (es proporcional a la desviación típica de la distribución, $\sigma^2 = \pi^2\psi^2/6$).

Y la función densidad de la distribución Gumbel de máximos de una variable aleatoria X es:

$$f(x; \mu, \psi) = \frac{dF(x; \mu, \psi)}{dx} = \frac{1}{\psi} \exp\left(-\frac{x-\mu}{\psi}\right) \exp\left[-\exp\left(-\frac{x-\mu}{\psi}\right)\right]; \quad -\infty < x < \infty$$

(A.I.8)

En el papel probabilístico Gumbel de máximos las variables reducidas, definidas a partir de la función de distribución, son:

$$\begin{aligned} \zeta &= g(x) = x \\ \eta &= h(y) = -\log[-\log(y)] \end{aligned}$$

(A.I.9)

y la familia de rectas ($\Theta = \{\alpha, \beta\}$) es:

$$\eta = \alpha + \beta\zeta = \frac{\zeta - \mu}{\psi}; \quad \text{con } \alpha = \frac{-\mu}{\psi} \text{ y } \beta = \frac{1}{\psi}$$

(A.I.10)

Cabe señalar, que para el caso de la función de distribución Gumbel de máximos, ajustando mediante OLS, se suele utilizar la fórmulas de punteo de Gringorten, ver tabla A.I.1.

A.I.3.3 Distribución Weibull de mínimos.

La función de distribución Weibull de mínimos de una variable aleatoria X es:

$$y = F(x; A, B, C) = 1 - \exp\left[-\left(\frac{x-B}{A}\right)^C\right]; \quad B \leq x < \infty$$

(A.I.11)

donde:

A es el parámetro de escala (determina la anchura de la función de distribución)

B es el parámetro de localización (es el mínimo valor posible de x). Para $B = 0$ la función de distribución de Weibull se denomina biparamétrica.

C es el parámetro de forma. Para $C = 1$, se reduce a una función de distribución exponencial y para $C = 2$ se obtiene la función de distribución de Rayleigh¹.

Y la función densidad de la distribución Weibull de mínimos de una variable aleatoria X es:

$$f(x; A, B, C) = \frac{dF(x; A, B, C)}{dx} = \frac{C}{A} \left(\frac{x-B}{A}\right)^{C-1} \exp\left[-\left(\frac{x-B}{A}\right)^C\right]; \quad B \leq x < \infty$$

(A.I.12)

Para el papel probabilístico Weibull de mínimos las variables reducidas se determinan con las transformaciones definidas a partir de la función de distribución:

$$\zeta = g(x) = x$$

$$\eta = h(y) = [-\log(1-y)]^{1/C}$$

(A.I.13)

y la familia de rectas ($\Theta = \{\alpha, \beta\}$) es:

$$\eta = \alpha + \beta\zeta = \frac{\zeta - B}{A}; \quad \text{con } \alpha = \frac{-B}{A} \text{ y } \beta = \frac{1}{A}$$

(A.I.14)

Cabe señalar, por su utilización, que para el caso de la función de distribución Weibull de mínimos, ajustando mediante OLS, se suelen utilizar las fórmulas de punteo que aparecen en la tabla A.I.2, que dependen de los parámetros de la función de distribución.

¹ Función de distribución de Rayleigh: utilizada tradicionalmente para representar la distribución de las alturas de ola dentro de un estado de mar en oleajes gaussianos de banda estrecha (tipo *swell*).

Relación	Autor	
$F'(x) = \frac{i - \lambda_1}{n + \lambda_2}$	$\lambda_1 = 0.30 + 0.18C$ $\lambda_2 = 0.21 + 0.32C$	Petruaskas y Aagaard
	$\lambda_1 = 0.20 + 0.27/\sqrt{C}$ $\lambda_2 = 0.20 + 0.23/\sqrt{C}$	Goda

Tabla A.I.2. Fórmulas de punteo de la distribución Weibull de mínimos.

A.I.3.4 Distribución GEV.

La función distribución GEV (*Generalized Extreme value*) ha sido definida dentro de la teoría de valores extremos (TVE), que describe el comportamiento de los valores extremos (máximos o mínimos) mediante modelos basados en distribuciones de probabilidad. Dicha teoría busca la mejor definición posible de la cola de la distribución. Por ello, la función de distribución de máximos ($Z_n = \max\{X_1, \dots, X_n\}$) de una muestra de tamaño n procedente de una población con función de distribución $G(x)$ es:

$$H_n(x) = \Pr(Z_n \leq x) = [G(x)]^n \tag{A.I.15}$$

La forma de Von Mises que engloban las tres únicas distribuciones límites para máximos (si $G(x)$ está en el máximo dominio de atracción de $H(x)$) es:

$$H(x; \mu, \psi, \xi) = \begin{cases} \exp\left\{-\left[1 + \xi\left(\frac{x - \mu}{\psi}\right)\right]_+^{-1/\xi}\right\} & \xi \neq 0 \\ \exp\left\{-\exp\left[-\left(\frac{x - \mu}{\psi}\right)\right]\right\} & \xi = 0 \end{cases} \tag{A.I.16}$$

donde $[a]_+$ es el máximo de a y 0 ($[a]_+ = \max(a, 0)$), μ es el parámetro de localización, ψ es el parámetro de escala y ξ es el parámetro de forma.

El parámetro de localización especifica la posición de la probabilidad de no excedencia $\exp(-1)$, representando los valores medios. El parámetro de escala controla la dispersión de la unidad de medida dentro del rango de la distribución. Finalmente, el parámetro de forma condiciona el tipo de distribución, la expresiones para $\xi \neq 0$ engloba dos familias de distribución: Frechet para el caso de $\xi > 0$ y Weibull para el caso $\xi < 0$, mientras que para el caso $\xi = 0$ se obtiene la familia de distribuciones tipo Gumbel a partir del limite de la ecuación definida para Frechet/Weibull cuando $\xi \rightarrow 0$.

En el transcurso de la presente tesis, que presenta un gran interés por los valores máximos, se ha identificado con las siglas GEV a la forma de Von Mises de máximos, asumiendo que su función de distribución es la de la ecuación A.I.16 ($F(x) \equiv H(x)$). Nótese como para $\xi = 0$ coincide con la expresión de Gumbel de máximos (ver ecuación A.I.7). Dicha función de distribución presenta la siguiente función de densidad:

$$f(x; \mu, \psi, \xi) = \frac{dF(x; \mu, \psi, \xi)}{dx} = \begin{cases} \frac{1}{\psi} \left[1 + \xi \left(\frac{x - \mu}{\psi} \right) \right]_+^{-1/\xi - 1} \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu}{\psi} \right) \right]_+^{-1/\xi} \right\} & \xi \neq 0 \\ \frac{1}{\psi} \exp \left[- \left(\frac{x - \mu}{\psi} \right) \right] \exp \left\{ - \exp \left[- \left(\frac{x - \mu}{\psi} \right) \right] \right\} & \xi = 0 \end{cases} \quad (\text{A.I.17})$$

Si se realiza el mismo cambio de variable en la ecuación A.I.16 que el que se realiza para el papel probabilístico Gumbel de máximos (ecuación A.I.9), se obtiene para $\xi = 0$ la familia de rectas Gumbel de máximos (igual a la ecuación A.I.10):

$$\eta = \alpha + \beta \zeta = \frac{\zeta - \mu}{\psi}; \quad \text{con } \alpha = \frac{-\mu}{\psi} \text{ y } \beta = \frac{1}{\psi} \quad (\text{A.I.18})$$

En cambio, para $\xi \neq 0$ y con el mismo cambio de variable, se obtiene una familia de curvas logarítmicas (no son rectas) que serán cóncavas para $\xi > 0$ (Frechet) y convexas para $\xi < 0$ (Weibull):

$$\eta = \frac{1}{\xi} \log \left(1 + \xi \frac{\zeta - \mu}{\psi} \right) \quad (\text{A.I.19})$$

ANEJO II

FUNCIONES EMPÍRICAS ORTOGONALES

A.II.1. Introducción.

En el presente Anejo se va a explicar una técnica estadística muy utilizada para estudiar y caracterizar variables geofísicas. Así, cuando se tienen variables de alta dimensionalidad (como las que producen los modelos numéricos en oceanografía), donde los datos pueden estar muy correlacionados en sus distintas componentes, para eliminar la información redundante se pueden utilizar el análisis de componentes principales (PCA, *Principal Component Analysis*). El PCA, también conocido en las ciencias meteorológicas como la descomposición en funciones empíricas ortogonales (EOF, *Empirical Orthogonal Function*) es una técnica estándar para reducir la información redundante con la mínima pérdida de variabilidad.

Dentro de las técnicas EOFs bidimensionales se pueden aplicar, por ejemplo, a variables espacio-temporales o también a variables direccionales (con módulo y argumento) tratándolas como variables complejas. A continuación se va a explicar la técnica EOF aplicada únicamente a una variable espacio-temporal, pero es igualmente aplicable a una variable espacio-direccional o con otras dependencias de sus variables, sin pérdida de generalidad (evitando la utilización de la matemática de los números complejos).

A.II.2 EOF.

La técnica EOF es un método estadístico que descompone una determinada variable numérica espacio-temporal en una combinación lineal de modos o funciones ortogonales numéricas. Esto se logra proyectando el conjunto de datos en un nuevo espacio (de menor dimensión que el original), donde las nuevas variables/dimensiones representan aquellas direcciones del espacio donde los datos tienen mayor varianza.

Así, con sólo unos pocos modos se puede explicar la mayor parte de la variación, tanto espacial como temporal, de la variable $Z(x, \tau)$. Siendo Z la variable numérica, que depende de x , que es la variable espacio, y de τ , que es la variable tiempo. Por lo tanto al elegir los modos que explican la mayor parte de la variabilidad de $Z(x, \tau)$ podemos representar ésta, considerando que el resto de la variabilidad no es significativa para la variación espacio-temporal de la variable Z .

La variable genérica $Z(x, \tau)$, con $x = \{x_1, x_2, \dots, x_i, \dots, x_n\}$ y $\tau = \{\tau_1, \tau_2, \dots, \tau_j, \dots, \tau_m\}$, se ordena en una matriz de $n \times m$ que se denota por $\mathbf{Z}(\mathbf{x}, \boldsymbol{\tau})$; siendo \mathbf{x} un vector que contiene las n posibles posiciones donde está definida la matriz $\mathbf{Z}(\mathbf{x}, \boldsymbol{\tau})$ y $\boldsymbol{\tau}$ un vector que contiene los m

posibles tiempos donde está definida $\mathbf{Z}(\mathbf{x}, \tau)$. La matriz \mathbf{Z} (de n filas y m columnas) tiene en cada elemento $\mathbf{Z}(i, j)$ el valor de la variable $Z(x, \tau)$ para la posición x_i y el tiempo τ_j .

La descomposición EOF de $\mathbf{Z}(i, j)$ se puede realizar de dos maneras: la primera es una descomposición espacial de la variabilidad y la segunda es una descomposición temporal de la variabilidad (Baldacci *et al.*, 2001):

- La descomposición espacial consiste en analizar los datos de la matriz \mathbf{Z} , restándoles previamente la media espacial (\mathbf{Z}_{MS} , sólo depende de τ):

$$\mathbf{Z}_S(i, j) = \mathbf{Z}(i, j) - \mathbf{Z}_{MS}(j) = \mathbf{Z}(i, j) - \frac{1}{n} \sum_{k=1}^n \mathbf{Z}(k, j) \quad (\text{A.II.1})$$

- La descomposición temporal consiste en analizar los datos de la matriz \mathbf{Z} , restándoles previamente la media temporal (\mathbf{Z}_{MT} , sólo depende de \mathbf{x}):

$$\mathbf{Z}_T(i, j) = \mathbf{Z}(i, j) - \mathbf{Z}_{MT}(i) = \mathbf{Z}(i, j) - \frac{1}{m} \sum_{k=1}^m \mathbf{Z}(i, k) \quad (\text{A.II.2})$$

La descomposición espacial destaca los gradientes espaciales, mientras que la descomposición temporal revela las variaciones en el tiempo. Para esta tesis se utiliza la descomposición temporal EOF¹, que se explica seguidamente.

Tras calcular \mathbf{Z}_T (ver ecuación A.II.1), matriz de $n \times m$, ésta se pretende descomponer en una base ortogonal de funciones de la forma:

$$\mathbf{Z}_T = \mathbf{F}\mathbf{G}^T \quad (\text{A.II.3})$$

Donde el superíndice T significa traspuesta; \mathbf{F} es la matriz de $n \times n$ de los modos espaciales, que resulta ser la matriz de autovectores de la matriz espacial de correlación, la columna i -ésima es el modo espacial \mathbf{f}_i (vector columna de $n \times 1$):

$$\mathbf{F} = [\mathbf{f}_1 | \mathbf{f}_2 | \dots | \mathbf{f}_i | \dots | \mathbf{f}_n] \quad (\text{A.II.4})$$

¹ La descomposición temporal en EOF utilizada en esta tesis podrá ser indistintamente temporal o direccional, en función si la variable bidimensional Z es espacio-temporal o espacio-direccional, con lo que Z_{MT} se denotará de manera general como Z_M .

\mathbf{G} es la matriz de $m \times n$ de las amplitudes temporales, la columna i -ésima es la amplitud temporal \mathbf{g}_i (vector columna de $m \times 1$):

$$\mathbf{G} = [\mathbf{g}_1 | \mathbf{g}_2 | \dots | \mathbf{g}_i | \dots | \mathbf{g}_n] \quad (\text{A.II.5})$$

A partir de la definición de la matriz espacial de correlación \mathbf{R}_s (matriz de $n \times n$):

$$\mathbf{R}_s = \mathbf{Z}_T \mathbf{Z}_T^T \quad (\text{A.II.6})$$

se puede relacionar con sus autovectores (\mathbf{f}_i) y autovalores asociados (λ_i), por medio de la ecuación resoluble:

$$\mathbf{R}_s \mathbf{f}_i = \lambda_i \mathbf{f}_i \quad (\text{A.II.7})$$

Con la ecuación A.II.7 se calcula \mathbf{F} (definida en la ecuación A.II.4); una vez calculados los modos espaciales, las amplitudes temporales se despejarán de la ecuación A.II.3:

$$\mathbf{G} = \mathbf{Z}_T^T \mathbf{F} \quad (\text{A.II.8})$$

De manera análoga a la descomposición temporal se hará la descomposición espacial (ver ecuación A.II.2):

$$\mathbf{Z}_s = \mathbf{L} \mathbf{M}^T \quad (\text{A.II.9})$$

donde \mathbf{L} es la matriz de $n \times m$ de amplitudes espaciales y \mathbf{M} es la matriz de $m \times m$ de modos temporales, que se obtiene como los autovectores de la matriz temporal de correlación (\mathbf{R}_T):

$$\mathbf{R}_T = \mathbf{Z}_s^T \mathbf{Z}_s \quad (\text{A.II.10})$$

Las amplitudes espaciales se hallan despejando \mathbf{L} de la ecuación A.II.9:

$$\mathbf{L} = \mathbf{Z}_s \mathbf{M} \quad (\text{A.II.11})$$

ANEJO II

Por lo tanto, y a modo de ejemplo, para reconstruir $\mathbf{Z}(\mathbf{x}, \tau)$ a partir de la descomposición temporal con un número reducido de modos, se hará:

$$\mathbf{Z}(\mathbf{x}, \tau) = \mathbf{Z}_{\text{MT}}(\mathbf{x}) + \mathbf{f}_1(\mathbf{x})\mathbf{g}_1(\tau) + \mathbf{f}_2(\mathbf{x})\mathbf{g}_2(\tau) + \dots + \mathbf{f}_p(\mathbf{x})\mathbf{g}_p(\tau) \quad (\text{A.II.12})$$

siendo p menor que n , que es número total de modos.

La manera de decidir cuantos modos tomar, es evaluando el porcentaje de varianza que explican los p modos. Siendo la varianza que explica cada autovector su autovalor asociado (Cieslikiewicz y Graff, 1996). Por lo tanto el porcentaje de varianza explicada por cada uno (λ_i) es:

$$\frac{\lambda_i}{\sum_{k=1}^n \lambda_k} \quad (\text{A.II.13})$$

Si ordenamos los modos de manera decreciente según sus autovectores ($\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_p, \dots, \mathbf{f}_n$ con sus autovectores ordenados de mayor a menor, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq \dots \geq \lambda_n$), el tanto por ciento (%) que explican los p primeros modos es:

$$100 \frac{\sum_{i=1}^p \lambda_i}{\sum_{k=1}^n \lambda_k} \quad (\text{A.II.14})$$

