

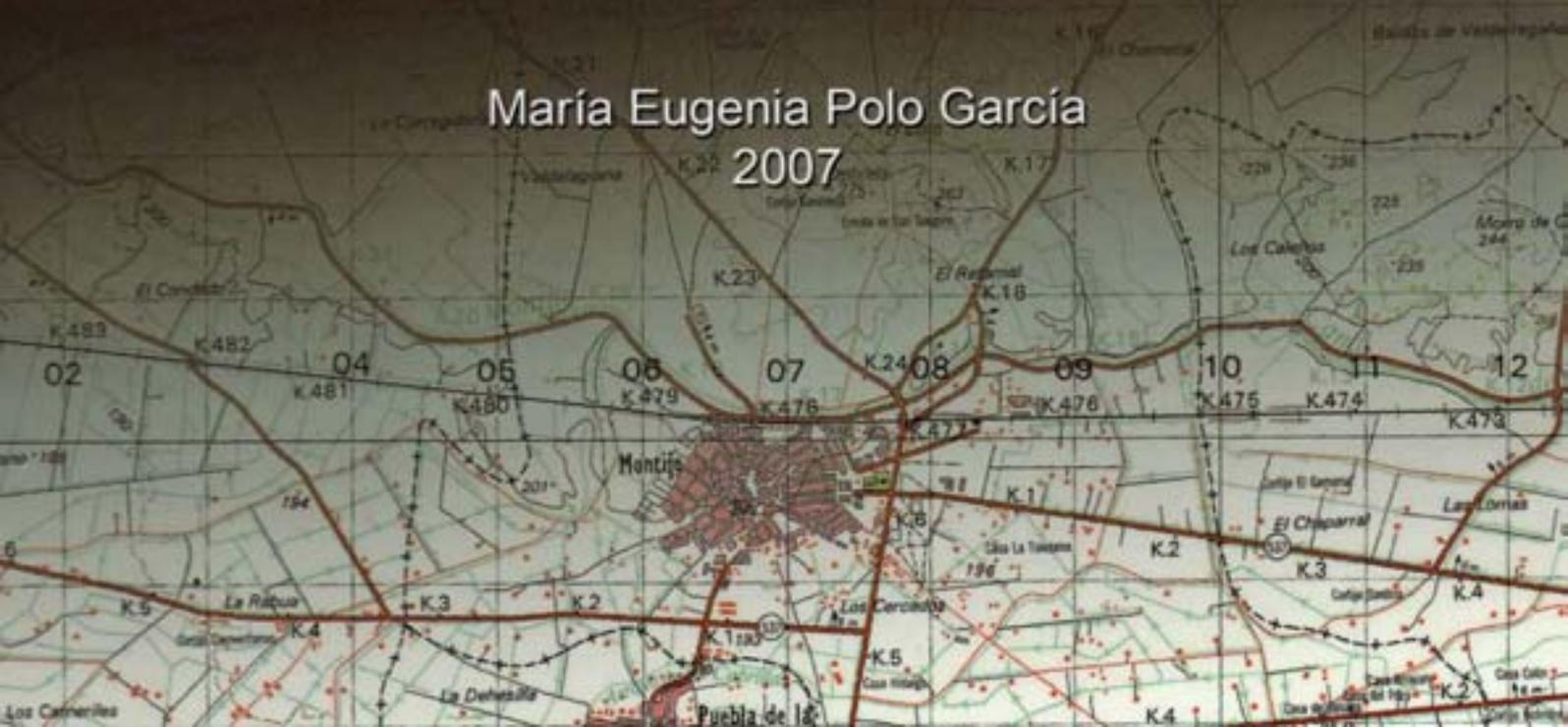
UNIVERSIDAD DE EXTREMADURA
Departamento de Expresión Gráfica



Tesis Doctoral

Gestión integral del error de posición en las bases de datos espaciales

María Eugenia Polo García
2007



***Edita: Universidad de Extremadura
Servicio de Publicaciones***

Caldereros 2. Planta 3^a
Cáceres 10071
Correo e.: publicac@unex.es
<http://www.unex.es/publicaciones>

Departamento de Expresión Gráfica
Centro Universitario de Mérida
UNIVERSIDAD DE EXTREMADURA



“GESTIÓN INTEGRAL DEL ERROR DE POSICIÓN EN LAS
BASES DE DATOS ESPACIALES”

Memoria que presenta

María Eugenia Polo García

Para optar al grado de Doctora por la Universidad de Extremadura

DIRECTOR DE TESIS

Ángel Manuel Felicísimo Pérez

Mérida, septiembre de 2007

Don Ángel Manuel Felicísimo Pérez, Doctor en Biología y Profesor Titular de la Universidad de Extremadura, informa que:

La memoria titulada **Gestión integral del error de posición en las bases de datos espaciales**, que presenta María Eugenia Polo García, Ingeniera en Geodesia y Cartografía, para optar al grado de Doctor ha sido realizada en el Departamento de Expresión Gráfica de la Universidad de Extremadura, bajo mi dirección, reuniendo todas las condiciones exigidas a los trabajos de tesis doctoral.

Mérida, septiembre de 2007

Fdo. Ángel Manuel Felicísimo Pérez

Decía Marco Aurelio que para acometer cualquier empresa hacía falta valentía para comenzarla y perseverancia para terminarla.

Me gustaría añadir que también hace falta ilusión.

AGRADECIMIENTOS

El primer agradecimiento es para mi director de tesis, Ángel, que allá por el 2004, cuando le pedí que me dirigiera la tesis, me dijo que sí. Una vez que entendí dónde me había metido, también comprendí que se necesitaba ayuda para llevar a buen puerto esta empresa. Gracias de todo corazón a todos los que me habéis ayudado; además de buenos amigos, compartimos una misma opinión de lo que debe ser nuestro trabajo.

Gracias Ángel por la atención y el tiempo dedicado y sobre todo por la confianza.

Gracias Emilio Hernández, la estadística resulta ahora menos dura.

Gracias Aurora Cuartero y Alicia Antolín por la ayuda con el funcionamiento del escáner fotogramétrico.

Gracias José Antonio Gutiérrez y Alicia Gómez por escuchar.

Agradecer también a la empresa GTBibérica en la persona de José Ignacio Calvo por la atención técnica prestada en relación al escáner fotogramétrico y a Alberto González Díez, del Departamento de Ciencias de la Tierra y Física de la Materia Condensada de la Universidad de Cantabria, por facilitar datos de calibración del escáner fotogramétrico. A Arthur Pewsey del Departamento de Matemáticas de la Universidad de Extremadura por su atención.

Un agradecimiento especial a Juan José Vales Bravo del Servicio de Información y Evaluación Ambiental de la Consejería de Medio Ambiente de la Junta de Andalucía por la preparación de datos y a Don José Manuel Moreira, Director del Servicio, por su cesión.

Índice

1	INTRODUCCIÓN, CONCEPTOS Y DEFINICIONES	15
1.1	Magnitudes y medidas	16
1.2	Error aleatorio y error sistemático.....	19
1.3	Exactitud, precisión e incertidumbre	21
1.4	Calidad y normalización.....	24
1.4.1	Definiciones.....	24
1.4.2	Elementos de calidad	25
1.4.3	Antecedentes de normalización de datos espaciales	27
1.5	Objetivos del trabajo.....	29
1.6	Organización de la Tesis	30
2	LA NATURALEZA VECTORIAL DEL ERROR DE POSICIÓN	33
2.1	Error aleatorio y error sistemático.....	34
2.2	Exactitud, precisión e incertidumbre	35
2.3	Fuentes y propagación de incertidumbre en los datos espaciales	37
3	MATERIAL Y MÉTODOS.....	41
3.1	Material.....	41
3.1.1	Datos de la calibración de un escáner fotogramétrico	41
3.1.2	Datos de la comparación de cartografía a diferentes escalas y procedencia.....	46
3.2	Metodología	48
3.2.1	Estadísticos básicos.....	48
3.2.1.1	Estadísticos y parámetros básicos para vectores unitarios	50

3.2.1.2	Estadísticos y parámetros básicos de los módulos de los vectores	54
3.2.2	Funciones de distribución de probabilidad	55
3.2.2.1	Funciones de distribución de probabilidad de datos circulares	55
3.2.2.2	Funciones de distribución de probabilidad de datos lineales	59
3.2.2.3	Funciones de distribución conjunta.....	61
3.2.3	Pruebas para la determinación de las características de las distribuciones.....	62
3.2.3.1	Pruebas para la determinación de las características de las distribuciones de los datos circulares	62
3.2.3.2	Pruebas para la determinación de las características de las distribuciones de los datos lineales	71
3.2.3.3	Pruebas para la determinación de las características de las distribuciones conjuntas de datos.....	72
4	CÁLCULOS Y RESULTADOS.....	75
4.1	Cálculos y resultados del tratamiento de datos de la calibración de un escáner fotogramétrico	75
4.1.1	Origen de los datos	75
4.1.2	Características de los datos	75
4.1.3	Descripción de los análisis realizados.....	78
4.1.3.1	Análisis de datos circulares	79
4.1.3.2	Análisis de datos lineales	80
4.1.3.3	Análisis en la distribución conjunta	81
4.1.4	Resultados.....	81
4.1.4.1	Resultados del análisis de datos circulares.....	82
4.1.4.2	Resultados del análisis de datos lineales.....	104
4.1.4.3	Resultados del análisis en la distribución conjunta.....	122
4.1.5	Conclusiones.....	127
4.1.5.1	Conclusiones para datos circulares	127
4.1.5.2	Conclusiones para datos lineales.....	130
4.1.5.3	Conclusiones para la distribución conjunta.....	131

4.2	Cálculos y resultados del tratamiento de datos de la comparación de cartografías a diferentes escalas	132
4.2.1	Origen de los datos	132
4.2.2	Características de los datos	133
4.2.3	Análisis realizados	133
4.2.3.1	Análisis de datos circulares	133
4.2.3.2	Análisis de datos lineales	134
4.2.3.3	Análisis en la distribución conjunta	134
4.2.4	Resultados.....	134
4.2.4.1	Resultados del análisis de datos circulares.....	134
4.2.4.2	Resultados del análisis de datos lineales	138
4.2.4.3	Resultado del análisis de la distribución conjunta	140
4.2.5	Conclusiones.....	144
4.2.5.1	Conclusiones para datos circulares	144
4.2.5.2	Conclusiones para datos lineales.....	145
4.2.5.3	Conclusiones para la distribución conjunta.....	145
5	PROPUESTA DE NORMA	147
6	DISCUSIÓN.....	161
7	LÍNEAS FUTURAS.....	163
8	ANEXO	165
8.1	Estadísticos y parámetros de datos circulares.....	165
8.2	Estándares de datos espaciales.....	168
8.2.1	Federal Geographic Data Committee.....	168
8.2.2	American Society of Photogrammetry and Remote Sensing.....	169
8.2.3	United Status Geological Survey	170
8.2.4	ISO/TC 211.....	171

8.2.5	Instituto Colombiano de Normas Técnicas y Certificación	172
8.3	Instrumental utilizado	173
8.3.1	Escáner fotogramétrico UltraScan 5000 Vexcel.....	173
8.3.2	Calibración geométrica	174
8.4	Aplicaciones informáticas.....	174
8.4.1	Oriana de Kovach	174
8.4.2	Microsoft Office Excel	177
8.4.3	R Project	177
8.4.4	SPSS	177
8.4.5	ArcView de Esri.....	178
8.4.6	gvSIG de la Comunidad Valenciana.....	179
8.5	Acrónimos	180
9	REFERENCIAS BIBLIOGRÁFICAS	183

Índice de Figuras

Fig. 1-1. Curva de Gauss.....	20
Fig. 1-2. Error e incertidumbre.....	23
Fig. 2-1. Vector de error de posición espacial.....	33
Fig. 2-2. Ejemplo de error de posición en dos dimensiones en cartografía.....	34
Fig. 2-3. Concepto de exactitud en datos bidimensionales.	35
Fig. 2-4. Concepto de precisión en datos bidimensionales.	36
Fig. 2-5. Concepto de incertidumbre en datos bidimensionales.....	36
Fig. 3-1. Resultados de los parámetros de transformación en la calibración de un escáner fotogramétrico.	43
Fig. 3-2. Datos del fichero de calibración geométrica del escáner fotogramétrico.	43
Fig. 3-3. Vectores de error en la calibración de un escáner fotogramétrico.....	44
Fig. 3-4. Histograma en fichero de calibración de un escáner fotogramétrico.....	45
Fig. 3-5. Superposición de dos cartografías a diferentes escalas.	47
Fig. 3-6 Representación de datos circulares como vectores y ejes.....	49
Fig. 3-7. Representación de datos brutos de dos muestras de datos.....	51
Fig. 3-8. Ejemplo de distribución de datos con módulo medio pequeño sin ser la distribución uniforme.	52
Fig. 3-9. Distribución uniforme de datos circulares.....	56
Fig. 3-10. Distribución von Mises.....	57
Fig. 3-11. Distribución de probabilidad de la distribución cardioide.....	58
Fig. 3-12. Comparativa de distribuciones von Mises, cardioide y envolvente-normal.....	59
Fig. 3-13. Gráficas de la función de densidad gamma para distintos valores de α y β	60
Fig. 3-14. Gráficas de la función de densidad gamma para distintos valores de α y β	60
Fig. 3-15. Gráfica de la función de densidad de la distribución logarítmico-normal.....	61
Fig. 3-16. Correlación entre acimutes y módulos de una muestra de datos circulares.....	61
Fig. 3-17. Mapa de densidad para el ejemplo a) del escáner fotogramétrico.....	62

Fig. 3-18. Gráfico de valoración de uniformidad sobre los datos de la calibración del escáner.	64
Fig. 3-19. Distribución bimodal de datos circulares.	66
Fig. 3-20. Ajuste gráfico de los datos de acimutes de paleocorrientes a la distribución de von Mises.	69
Fig. 3-21. Histograma de frecuencia de datos del ejemplo a).	71
Fig. 4-1. Parte del resultado gráfico de la calibración.	76
Fig. 4-2. Distribución bimodal de datos circulares.	85
Fig. 4-3. Acimut medio en muestras de diferente tamaño.	92
Fig. 4-4. Módulo medio en muestras de diferente tamaño.	92
Fig. 4-5. Desviación estándar en muestras de diferente tamaño.	93
Fig. 4-6. Parámetro de concentración de von Mises en muestras de diferentes tamaño.	93
Fig. 4-7. Distribución de datos circulares en diferentes tamaños de ventanas.	103
Fig. 4-8. Histograma de módulos no afectados de error.	108
Fig. 4-9. Histograma de módulos afectados de error.	108
Fig. 4-10. Histograma de distribución de datos en muestras de 10x10 elementos afectados de error.	120
Fig. 4-11. Histogramas de distribución de datos lineales para muestras de diferentes tamaños en la calibración '2005-02-21'.	121
Fig. 4-12. Gráficos de correlación entre datos lineales y circulares en las calibraciones a) '2005-04-22e' y b) '2005-02-21'.	123
Fig. 4-13. Mapas de densidad y <i>outliers</i> en 4 calibraciones del escáner fotogramétrico.	125
Fig. 4-14. Modificación de <i>outliers</i> en el fichero '2005-02-21'.	126
Fig. 4-15. Distribución de datos brutos en la comparativa de diferentes cartografías.	137
Fig. 4-16. Histogramas de distribución de valores en los ficheros de la comparativa entre diferentes cartografías.	140
Fig. 4-17. Correlaciones existentes entre módulos y acimutes en los ficheros 'Mérida' y '10-50'.	141
Fig. 4-18. Mapas de densidad y <i>outliers</i> en la comparativa entre cartografías.	143
Fig. 5-1. Relación entre el tamaño de la muestra y fiabilidad en el valor de la desviación típica.	150
Fig. 8-1. Resultados de cálculo de Oriana.	175
Fig. 8-2. Pantalla de resultados de Oriana.	176

Fig. 8-3. Gráfico de Oriana.	176
Fig. 8-4. Pantalla de R.....	177
Fig. 8-5. Histograma de distribución de datos obtenido con SPSS.....	178
Fig. 8-6. Pantalla de resultados en ArcView.....	179
Fig. 8-7. Pantalla de resultados con gvSIG.....	180

Índice de Ecuaciones

Ec. 1-1. Error cuadrático medio.	21
Ec. 2-1. Función de densidad de probabilidad de la distribución uniforme.	34
Ec. 2-2. Función de densidad de probabilidad de la distribución von Mises.	35
Ec. 3-1. Transformación afín bidimensional utilizada en la calibración geométrica.	42
Ec. 3-2. Parámetros de transformación.	43
Ec. 3-3. Cálculo de módulo y acimut del vector de error.	49
Ec. 3-4. Acimut medio.	51
Ec. 3-5. Mediana de la muestra.	51
Ec. 3-6. Módulo del vector resultante y valores de C y S.	52
Ec. 3-7. Módulo medio.	52
Ec. 3-8. Estimación de ρ	52
Ec. 3-9. Varianza circular de la muestra.	53
Ec. 3-10. Varianza circular de la población.	53
Ec. 3-11. Desviación estándar de la muestra.	53
Ec. 3-12. Desviación estándar circular de la población.	53
Ec. 3-13. Dispersión circular.	53
Ec. 3-14. Valor de $\hat{\rho}_2$ y momento coseno.	53
Ec. 3-15. Dispersión circular de la población.	53
Ec. 3-16. Media aritmética.	54
Ec. 3-17. Desviación típica para datos lineales.	55
Ec. 3-18. Función de densidad de probabilidad.	56
Ec. 3-19. Función de distribución.	56
Ec. 3-20. Función de densidad de probabilidad de VM.	56
Ec. 3-21. Función de distribución de VM.	56
Ec. 3-22. Valores de ρ y δ en VM.	57
Ec. 3-23. Función de densidad de probabilidad en distribución cardioide.	57

Ec. 3-24. Función de distribución en distribución cardioide.....	57
Ec. 3-25. Función de densidad de probabilidad de distribución WC.....	58
Ec. 3-26. Función de distribución de WC.....	58
Ec. 3-27. Función de densidad de probabilidad de la distribución normal.....	59
Ec. 3-28. Función de distribución de probabilidad de la distribución gamma.....	60
Ec. 3-29. Función de distribución de probabilidad de la distribución Weibull.....	60
Ec. 3-30. Función de distribución de la probabilidad de la función logarítmico-normal.....	61
Ec. 3-31. Cálculo de x_1	63
Ec. 3-32. Valoración gráfica de uniformidad.....	63
Ec. 3-33. Estadísticos de Kuiper.....	64
Ec. 3-34. Test de Kuiper.....	64
Ec. 3-35. Test de Rayleigh.....	65
Ec. 3-36. Probabilidad de Rayleigh.....	65
Ec. 3-37. Test V.....	65
Ec. 3-38. Test de Rao.....	65
Ec. 3-39. Estadístico de Rao.....	65
Ec. 3-40. Estimación de la mediana.....	67
Ec. 3-41. Error estándar circular.....	67
Ec. 3-42. Intervalo de confianza para acimut medio.....	67
Ec. 3-43. Cálculo de z_i para la valoración gráfica de von Mises.....	68
Ec. 3-44. Cálculo de puntos para la valoración gráfica de von Mises.....	68
Ec. 3-45. Cálculo previo para test de Watson.....	69
Ec. 3-46. Estadístico para el test de Watson.....	69
Ec. 3-47. Intervalo para el acimut medio en von Mises.....	70
Ec. 3-48. Error estándar para distribución von Mises.....	70
Ec. 3-49. Test para detección de <i>outliers</i>	71
Ec. 3-50. Media armónica.....	73
Ec. 5-1. Influencia del tamaño de la muestra en la fiabilidad del valor de la desviación estándar.	149

Ec. 5-2. Determinación del número de puntos de control en función de la fiabilidad de la estimación de la desviación estándar.....	149
Ec. 5-3. Influencia del tamaño de la muestra en la fiabilidad de la exactitud.....	149

Índice de Tablas

Tabla 3-1. Ejemplo de aplicación de estadísticos de datos circulares.....	54
Tabla 3-2. Ejemplos de aplicación de estadísticos de datos lineales.....	55
Tabla 3-3. Ejemplos de aplicación de test de uniformidad.	66
Tabla 3-4. Intervalos de confianza para la mediana y el acimut medio.	68
Tabla 3-5. Ejemplos de aplicación de tests de von Mises.	70
Tabla 3-6. Ejemplo de aplicación de test de Kolmogorov-Smirnov.	72
Tabla 4-1. Ficheros de calibración geométrica que presentan error introducido artificialmente.	76
Tabla 4-2. Tamaño de las ventanas y número de ventanas analizadas en los ficheros de calibración.	77
Tabla 4-3. Variables y unidades utilizadas en los análisis a realizar a los datos procedentes de las calibraciones del escáner fotogramétrico.	77
Tabla 4-4. Resumen de las características de los datos de las diferentes calibraciones del escáner fotogramétrico.	78
Tabla 4-5. Resultados de los estadísticos básicos para vectores unitarios en el análisis general.	82
Tabla 4-6. Resultados de las pruebas de uniformidad de Kuiper, Rao y Watson en el análisis general para 18 de 22 ficheros de datos circulares.	83
Tabla 4-7. Resultados de las pruebas de uniformidad de Kuiper, Rao y Watson en el análisis general que ofrecían resultados diferentes a la mayoría.	83
Tabla 4-8. Resultados del test de Rayleigh para datos circulares en el análisis general.	84
Tabla 4-9. Resultados del error estándar del acimut medio en el análisis general para datos circulares.....	84
Tabla 4-10. Resultados de las pruebas de ajuste a von Mises en el análisis general.....	85
Tabla 4-11. Resultados del cálculo de los estadísticos básicos de datos circulares para la calibración ‘2005-02-21’ con tamaño de muestra de 10x10 elementos.	87
Tabla 4-12. Resultados del cálculo de los estadísticos básicos de datos circulares para la calibración ‘2005-04-18’ con tamaño de muestra de 10x10 elementos.	88
Tabla 4-13. Resultados del cálculo de los estadísticos básicos para la calibración ‘2005-02-21’ con tamaño de muestra de 15x15 elementos.	89
Tabla 4-14. Resultados del cálculo de los estadísticos básicos para la calibración ‘2005-02-21’ con tamaño de muestra de 20x20 elementos.	90

Tabla 4-15. Resultados del cálculo de los estadísticos básicos para la calibración ‘2005-02-21’ con tamaño de muestra de 25x25 elementos.	90
Tabla 4-16. Resultados del cálculo de los estadísticos básicos para la calibración ‘2005-02-21’ con tamaño de muestra de 30x30 elementos.	91
Tabla 4-17. Resultados del cálculo de los estadísticos básicos para la calibración ‘2005-02-21’ con tamaño de muestra de 35x35 elementos.	91
Tabla 4-18. Resultados de test de uniformidad para ficheros con submuestras de 10x10 elementos.	94
Tabla 4-19. Resultados de análisis de submuestras que presentaban evidencias de uniformidad.	94
Tabla 4-20. Resultados de tests de uniformidad para la calibración ‘2005-02-21’ con submuestras de varios tamaños.	94
Tabla 4-21. Resultados del cálculo del error estándar para el acimut medio en la calibración ‘2005-02-21’ con tamaño de muestra de 15x15 elementos.	95
Tabla 4-22. Resultados del cálculo del error estándar para el acimut medio en la calibración ‘2005-02-21’ con tamaño de muestra de 20x20 elementos.	95
Tabla 4-23. Resultados del cálculo del error estándar para el acimut medio en la calibración ‘2005-02-21’ con tamaño de muestra de 25x25 elementos.	96
Tabla 4-24. Resultados del cálculo del error estándar para el acimut medio en la calibración ‘2005-02-21’ con tamaño de muestra de 30x30 elementos.	96
Tabla 4-25. Resultados del cálculo del error estándar para el acimut medio en la calibración ‘2005-02-21’ con tamaño de muestra de 35x35 elementos.	96
Tabla 4-26. Resultados del cálculo del error estándar para el acimut medio para la calibración ‘2005-02-21’ c.	97
Tabla 4-27. Resultados del cálculo del error estándar para el acimut medio para la calibración ‘2005-04-18’ con tamaño de muestra de 10x10 elementos.	97
Tabla 4-28. Resultados del cálculo del error estándar para el acimut medio para la calibración ‘2005-04-21e’ con tamaño de muestra de 10x10 elementos.	98
Tabla 4-29. Resultado del cálculo del error estándar para la calibración ‘2005-04-25’ con tamaño de 10x10.	98
Tabla 4-30. Resultados del cálculo del ajuste a la distribución von Mises en la calibración ‘2005-02-21’ con tamaño de muestra de 15x15 elementos.	99
Tabla 4-31. Resultados del cálculo del ajuste a la distribución von Mises en la calibración ‘2005-02-21’ con tamaños de muestra de 20x20, 25x25, 30x30 elementos y con la muestra total.	100
Tabla 4-32. Resultados del cálculo del ajuste a la distribución von Mises con tamaño de muestra de 10x10 elementos para la calibración 2005-02-21.	101

Tabla 4-33. Resultados del cálculo del ajuste a la distribución von Mises con tamaño de muestra de 10x10 elementos para la calibración 2005-04-18.	102
Tabla 4-34. Resultados de los estadísticos básicos para módulos en el análisis general.	105
Tabla 4-35. Número de <i>outliers</i> resultantes en cada fichero de calibración en análisis generales de datos lineales, considerando como límite la media más el triple de la desviación típica.	106
Tabla 4-36. Número de <i>outliers</i> resultantes en cada fichero de calibración en análisis generales de datos lineales, considerando como límite la media más 5 veces la desviación típica.	106
Tabla 4-37. Resultados de la prueba K-S para la distribución gamma y Weibull a los datos lineales.	107
Tabla 4-38. Resultados del cálculo de los estadísticos básicos para módulos para la calibración ‘2005-02-21’ con tamaño de muestra de 10x10 elementos.	109
Tabla 4-39. Resultados del cálculo de los estadísticos básicos para módulos con tamaño de muestra de 10x10 elementos para la calibración ‘2005-04-18’	110
Tabla 4-40. Resultados del cálculo de los estadísticos básicos para módulos con tamaño de muestra de 15x15 elementos para la calibración ‘2005-02-21’	111
Tabla 4-41. Resultados del cálculo de los estadísticos básicos para módulos con tamaño de muestra de 20x20 elementos para la calibración ‘2005-02-21’	112
Tabla 4-42. Resultados del cálculo de los estadísticos básicos para módulos con tamaño de muestra de 25x25 elementos para la calibración ‘2005-02-21’	112
Tabla 4-43. Resultados del cálculo de los estadísticos básicos para módulos con tamaño de muestra de 30x30 elementos para la calibración ‘2005-02-21’	113
Tabla 4-44. Resultados del cálculo de los estadísticos básicos para módulos con tamaño de muestra de 35x35 elementos para la calibración ‘2005-02-21’	113
Tabla 4-45. Resultados del cálculo de <i>outliers</i> en datos lineales para los 6 ficheros de calibración con submuestras de 100 elementos.	114
Tabla 4-46. Resultados del cálculo de <i>outliers</i> en datos lineales para los ficheros con submuestras de diferente tamaño en el fichero ‘2005-02-21’.	114
Tabla 4-47. Resultados del cálculo de la aplicación del test K-S para diferentes distribuciones en la calibración ‘2005-02-21’ con tamaño de muestra de 10x10 elementos.	115
Tabla 4-48. Resultados del cálculo de la aplicación del test K-S para diferentes distribuciones en la calibración ‘2005-04-18’ con tamaño de muestra de 10x10 elementos.	116
Tabla 4-49. Resultados del cálculo de la aplicación del test K-S para diferentes distribuciones para módulos con tamaño de muestra de 15x15 elementos en la calibración ‘2005-02-21’.	117
Tabla 4-50. Resultados del cálculo de la aplicación del test K-S para diferentes distribuciones para módulos con tamaño de muestra de 20x20 elementos en la calibración ‘2005-02-21’.	118

Tabla 4-51. Resultados del cálculo de la aplicación del test K-S para diferentes distribuciones para módulos con tamaño de muestra de 25x25 elementos en la calibración ‘2005-02-21’.	119
Tabla 4-52. Resultados del cálculo de la aplicación del test K-S para diferentes distribuciones para módulos con tamaño de muestra de 30x30 elementos en la calibración ‘2005-02-21’.	119
Tabla 4-53. Resultados del cálculo de la aplicación del test K-S para diferentes distribuciones para módulos con tamaño de muestra de 35x35 elementos en la calibración ‘2005-02-21’.	120
Tabla 4-54. Resultados de las correlaciones entre los módulos y acimutes.	123
Tabla 4-55. Resultados de los valores de la desviación típica en la dirección SCAN y CCD en las 22 calibraciones.	124
Tabla 4-56. Características de los datos procedentes de la comparación entre cartografías.	133
Tabla 4-57. Resultados de los estadísticos básicos para vectores unitarios de los datos procedentes de la comparación entre diferentes cartografías.	135
Tabla 4-58. Resultados de las pruebas de uniformidad para vectores unitarios de los datos procedentes de la comparación entre diferentes cartografías.	135
Tabla 4-59. Resultados del error estándar de la media para datos circulares en los dos ficheros procedentes de la comparativa entre ambas cartografías.	136
Tabla 4-60. Resultado de las pruebas de ajuste a la distribución von Mises en comparativa entre cartografías.	136
Tabla 4-61. Resultados de los estadísticos básicos para módulos en la comparativa de cartografías de diferentes escalas.	138
Tabla 4-62. Número de <i>outliers</i> y porcentaje en cada fichero en la comparativa de cartografías de diferentes escalas.	139
Tabla 4-63. Resultados de las pruebas de K-S para diferentes distribuciones de datos en la comparativa de cartografías de diferentes escalas.	139
Tabla 4-64. Resultados de las correlaciones en comparativas entre diferentes cartografías.	141
Tabla 4-65. Resultados de los valores de la desviación típica en X e Y para las dos comparativas de cartografías.	141
Tabla 8-1. Principales Normas ISO sobre información geográfica.	172

1 Introducción, conceptos y definiciones

Llamaremos **dato** a un hecho verificable sobre la realidad. Los datos son la base de la mayoría de la actividad científica (con la excepción, posiblemente, de una parte de la matemática) y de la práctica totalidad de la actividad tecnológica. Los datos surgen de la **metrología**, concepto que engloba un amplísimo conjunto de técnicas cuyo objetivo es, precisamente, el conocimiento cuantitativo (medida) de propiedades de los objetos.

Las técnicas de medida evolucionan continuamente con dos objetivos principales: a) aumentar la exactitud y precisión, y b) reducir el tiempo y coste de la captura de datos.

Cuando el volumen de información crece, la localización de un dato concreto se hace cada vez más difícil. En las últimas décadas, la disponibilidad de medios informáticos ha permitido organizar la información en estructuras denominadas **bases de datos**. Una base de datos es, por tanto, un conjunto de información organizado de forma estructurada con dos objetivos principales: a) permitir el acceso selectivo a la información mediante herramientas de consulta y b) garantizar que toda la información contenida en la base de datos está potencialmente disponible.

Las técnicas de medida no son perfectas y es una realidad indiscutible que todo dato contiene un componente de error. Podemos verlo desde el punto de vista de que un dato es una aproximación al valor real de la propiedad del objeto, por lo que el valor medido m_i es la suma del valor real m y el componente de error ε_i , aleatorio o no: $m_i = m + \varepsilon_i$.

Llamaremos **datos espaciales** a aquellos que hacen referencia a la localización de un objeto en el espacio. Los espacios pueden ser de un número arbitrario de dimensiones. Por ejemplo, las propiedades radiométricas de un objeto suelen representarse como localizaciones en un espacio con tantas dimensiones como bandas espectrales se hayan determinado (desde un mínimo de 3 a varios centenares en teledetección hiperespectral).

Los datos espaciales son, consecuentemente, **tuplas** de coordenadas que definen la posición del objeto en un sistema de referencia¹. En un espacio tridimensional y un sistema de referencia cartesiano, las tuplas adoptan convencionalmente la forma (x_i, y_i, z_i) donde cada elemento de la tupla representa la posición respecto a uno de los ejes del sistema de referencia.

Como en el caso general, cada uno de los elementos de una tupla de coordenadas está afectado por un componente de error: $(x_i + \varepsilon_x, y_i + \varepsilon_y, z_i + \varepsilon_z)$. Los valores $(\varepsilon_x, \varepsilon_y, \varepsilon_z)$ representan un vector cuyo origen estaría en la localización real y su final en la locali-

¹ El término tupla no está reconocido en el diccionario de la RAE pero sí es utilizado en disciplinas que van desde la filosofía de la ciencia (Bunge, 1985:27) hasta la programación informática. En este trabajo llamaremos tupla a una serie ordenada de coordenadas que definen una posición; el número de elementos de una tupla depende de la dimensión del espacio.

zación medida por lo que lo llamaremos **vector de error**. Lógicamente, los vectores de error tendrán tantos elementos como dimensiones el espacio de referencia.

La magnitud del error puede tener, obviamente, consecuencias relevantes. Según el trabajo que se realice, unos datos podrán ser utilizados o no, en función del error que los afecte. El problema no es sólo que el dato en sí sea erróneo sino que ese error afecta a todas las operaciones que se realicen con él. Por ejemplo, errores en la determinación de las alturas supondrán errores en la determinación de las pendientes del terreno. Este proceso se denomina **propagación del error**.

Lógicamente, el error que afecta a cada dato no es conocido. Sin embargo, su influencia debe ser estimada para valorar la calidad de los resultados en cualquier operación de análisis espacial. Este problema se resuelve invirtiendo recursos en repetir las medidas de una muestra de datos con mejores métodos. Se supone que estos métodos no son aplicables a la totalidad de los datos por motivos de coste económico, tiempo u otras razones técnicas (accesibilidad, por ejemplo). Por ese motivo, sólo un subconjunto de los datos totales es sometido a medidas de referencia. El objetivo es realizar una **estimación estadística del error** asociable a cada dato asumiendo que estas medidas de referencia están afectadas por un error significativamente menor que las originales. Si esto es cierto, podremos determinar los vectores de error de un conjunto limitado pero suficiente de datos y deducir las distribuciones estadísticas pertinentes, umbrales de confianza, etc. para el conjunto de datos.

El error espacial, hasta el presente, ha sido tratado como una magnitud escalar. La reducción vectorial \rightarrow escalar tiene como consecuencia inmediata que sólo podemos aprovechar una propiedad métrica: la magnitud del error (módulo del vector), perdiendo otras como el acimut (también métrica) y el sentido (topológica). En nuestra opinión, es necesario considerar el error como una magnitud vectorial para aprovechar toda la información de carácter direccional que los vectores representan. Veremos más detalladamente que el principal objetivo de esta Tesis es el tratamiento del error como vector ya que, en nuestra opinión, la información que puede obtenerse es más amplia y relevante que si lo consideramos un escalar.

En esta introducción hemos utilizado ya algunos términos clave en metrología pero sin explicarlos adecuadamente. En el apartado siguiente daremos las definiciones que se usarán en este trabajo, cuestión necesaria no sólo por ser la base conceptual del trabajo sino porque difieren del sentido coloquial que se utiliza en otros ámbitos (incluso en algunas otras ramas de la ciencia). Acotar con claridad términos como error, exactitud, precisión, incertidumbre y calidad es una tarea esencial en este trabajo.

1.1 Magnitudes y medidas

En el año 1978, el Comité Internacional de Pesas y Medidas (CIPM)² reconoce la necesidad de un consenso internacional sobre los términos que expresan la incertidumbre en la medida (CEM 2000a:v). En el año 1984 la Organización Internacional de Normali-

² <http://www.bipm.fr/en/committees/cipm/>

zación (ISO)³, publica la 1ª edición del “Vocabulario Internacional de Términos Básicos y Generales de Metrología” (CEM, 2000:4), mientras que el Centro Español de Metrología (CEM)⁴ dispone una versión en lengua española del citado vocabulario teniendo en cuenta las normas UNE 5 100 87 e ISO 31:1992 (CEM, 2000:4)⁵.

Es el contexto de la metrología y la normalización de términos metrológicos donde se definen los primeros conceptos básicos (Metrología, 2000). Literalmente son los siguientes:

- **Magnitud (mensurable):** “atributo de un fenómeno, cuerpo o sustancia, que es susceptible de ser distinguido cualitativamente y determinado cuantitativamente.”

Se diferencia entre magnitud general (longitud, tiempo, masa...) y magnitud particular, donde las anteriores se aplican a un objeto concreto.

- **Valor (de una magnitud):** “expresión cuantitativa de una magnitud particular, generalmente en forma de una unidad de medida multiplicada por un número.”

La distancia medida entre dos puntos, como magnitud particular, puede tener un valor de 36.23 metros, por ejemplo.

- **Valor verdadero (de una magnitud):** “valor en consistencia con la definición de una magnitud particular dada”.

Se añaden como notas:

“Es un valor que se obtendría por una medición perfecta” y

“Todo valor verdadero es por naturaleza indeterminado”.

- **Valor convencionalmente verdadero (de una magnitud):** “valor atribuido a una magnitud particular y aceptado, algunas veces por convenio, como teniendo una incertidumbre apropiada para un uso dado”.

Este concepto se hace necesario por la indeterminación mencionada en la nota anterior: aceptamos un ‘valor convencionalmente verdadero’ como referente ante la imposibilidad de determinar un ‘valor verdadero’. Este valor convencionalmente verdadero también se conoce, según diferentes bibliografías, como valor asignado, mejor estimación o valor más probable, entre otras expresiones. Obsérvese también que al hablar de valor verdadero se ha utilizado el artículo indefinido ‘un’ en lugar de

³ <http://www.iso.org>

⁴ <http://www.cem.es>

⁵ <http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=3621&ICS1=1&ICS2=60&ICS3=&scopelist=>

'el', pues el valor verdadero puede tener varios valores que se correspondan con la definición de una magnitud particular dada.

- **Medición:** "conjunto de operaciones que tienen por finalidad determinar un valor de una magnitud".

En el texto original vemos que posteriormente se usa indistintamente 'medición' y 'medida'.

- **Método de medida:** "sucesión lógica de las operaciones, descritas de una forma genérica, utilizadas en la ejecución de las mediciones."

Siguiendo con el ejemplo de la medida de una longitud, ésta podría ser observada por métodos mecánicos, ópticos o electrónicos, a cada uno de los cuales correspondería un método (o varios) diferentes.

- **Procedimiento de medida:** "conjunto de operaciones, descritas de forma específica, utilizadas en la ejecución de mediciones particulares según un método dado."

Supongamos que elegimos métodos electrónicos para medir la distancia considerada; el procedimiento de medida incluiría definir exactamente el instrumental utilizado, explicando paso a paso el proceso seguido con el estacionamiento del instrumental, su uso y número de observaciones realizadas para obtener un resultado.

- **Mensurando:** "magnitud particular sometida a medición."

La definición del mensurando puede necesitar indicaciones relativas a otras magnitudes tales como el tiempo, la temperatura o la presión. Así, la medida de la distancia entre dos marcas efectuadas en una barra de metal (magnitud particular) necesitará definir la temperatura ambiente, ya que el resultado depende de ella, en función del coeficiente de dilatación del material.

- **Resultado de una medición:** "valor atribuido a un mensurando, obtenido por medición."

Como nota se indica que una expresión completa del resultado incluirá información sobre la incertidumbre asociada a la medida; este tipo de información la denominaremos metadatos y consideraremos la asociación dato-metadato un factor enormemente relevante para la gestión correcta de las bases de datos espaciales.

La definición de metadatos al ejemplo de medida sobre una barra de metal es bastante intuitivo: dado que el resultado variará con la temperatura, se incluirán como metadatos la temperatura en el momento de hacer la medida y el índice de dilatación del material. Una expresión completa de metadatos podrá incluir también el método y el procedimiento de medida.

Una vez obtenido el valor del mensurando observaremos que no se corresponde con el valor verdadero de la magnitud: Esto es debido a múltiples causas derivadas del propio instrumental de medida, procedimientos, trabajo del observador, etc. Esta diferencia abre un nuevo capítulo de conceptos relacionados con el error. En metrología se admiten a este respecto definiciones diversas y acotaciones, como se indica a continuación.

1.2 Error aleatorio y error sistemático

Como ya se ha indicado en la introducción, el error, ε , es uno de los componentes del valor resultante de una medida m_i : $m_i = m + \varepsilon_i$. De la expresión anterior se deduce que el **error** es la diferencia entre el valor medido y el valor verdadero (o convencionalmente verdadero): $\varepsilon_i = m_i - m$. Esta expresión coincide con la definición oficial: “resultado de una medición menos el valor verdadero del mensurando” (CEM, 2000:24).

Por su importancia en otras definiciones, vamos a considerar los errores divididos en dos categorías: aleatorios y sistemáticos. Las definiciones oficiales son, en nuestra opinión, algo desafortunadas, como comentaremos posteriormente:

- **Error aleatorio:** “resultado de una medición menos la media de un número infinito de mediciones del mismo mensurando, efectuadas bajo condiciones de repetibilidad”.
- **Error sistemático:** “media que resultaría de un número infinito de mediciones del mismo mensurando efectuadas bajo condiciones de repetibilidad, menos un valor verdadero del mensurando”.

En realidad, el error aleatorio no puede definirse como un resultado de una medida sino como una propiedad de las desviaciones entre las medidas y el valor verdadero. En nuestra opinión, ambas categorías de errores se diferencian por propiedades estadísticas simples tal como comentamos a continuación con un ejemplo.

Supongamos que tomamos un número elevado de medidas (réplicas) de temperaturas del agua contenida en un recipiente en equilibrio con hielo mediante un termómetro digital cuya pantalla permite dos dígitos decimales. Por definición física elemental, la temperatura real (valor verdadero) es conocida: $m = 0.00$ °C. Sin embargo, vemos que los resultados en pantalla, m_i , no son siempre 0.00 °C sino que existe una variación; así, la primera lectura es $m_1 = +0.05$ °C, la segunda es $m_2 = -0.02$ °C, etc.

Tomando un número elevado de medidas, podemos representarlas conjuntamente como una distribución de frecuencias. Pueden presentarse dos casos básicos (luego veremos otras posibilidades):

- a. La distribución es una curva normal o de Gauss con media nula.
- b. La distribución es una curva normal o de Gauss con media no nula.

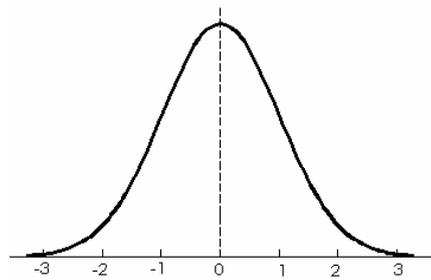


Fig. 1-1. Curva de Gauss.

Curva de Gauss convencional de media nula ($\bar{m} = 0$) y desviación estándar s , que responde a la distribución de los errores aleatorios.

La figura superior muestra el primer caso, una curva de Gauss convencional $N(\bar{m}, s)$, donde la media aritmética es nula ($\bar{m} = 0$), y s la desviación estándar (supongamos $s = 0.08$). En este caso se habla de **error aleatorio**, es decir, todas las desviaciones respecto al valor verdadero están descritas por una distribución gaussiana de media nula.

La otra posibilidad es encontrar una distribución normal pero donde $\bar{m} \neq 0$ (por ejemplo $\bar{m} = -0.32$); en este caso el proceso de medida tiende a dar resultados sistemáticamente más bajos que el valor verdadero por lo que hablamos de que existe un **error sistemático** (que se suma al error aleatorio). Conocido o estimado el error sistemático, las medidas pueden mejorarse globalmente si restamos $0.32 \text{ }^\circ\text{C}$ a los resultados que obtengamos. Normalmente, el error sistemático es el resultado de una mala calibración del instrumento de medida o de un mal diseño del procedimiento.

La definición de error aleatorio propuesta por el Vocabulario Internacional de Metrología (CEM, 2000:25) dice que es el "resultado de una medición menos la media de un número infinito de mediciones del mismo mensurando, realizadas bajo las mismas condiciones de repetibilidad". En realidad, lo que se define así es la desviación de un resultado concreto respecto a la media pero no es la definición del error aleatorio. Nosotros proponemos las siguientes definiciones:

- **Error aleatorio:** categoría de error donde las desviaciones de las medidas respecto al valor verdadero se ajustan a una distribución normal de media nula: $N(\bar{m}, s) \rightarrow \bar{m} = 0$
- **Error sistemático:** categoría de error donde las desviaciones de las medidas respecto al valor verdadero se ajustan a una distribución normal de media no nula: $N(\bar{m}, s) \rightarrow \bar{m} \neq 0$; el valor de la media define la magnitud del error sistemático.

Los parámetros de las distribuciones deberán estimarse mediante un número suficientemente elevado de observaciones que nos darán: a) el valor más probable del mensurando (media) y b) la desviación estándar. Esta última no interviene en estas definiciones de categorías de error pero sí en otras. El conocimiento de los parámetros de la distribución del error permite, entre otras cosas, determinar la probabilidad de que se presenten errores de una magnitud determinada, ya que se deduce de las propiedades de la distribución normal. Una utilidad inmediata de esta posibilidad es la de detectar **errores groseros** o valores aberrantes, ya que su magnitud hace que la probabilidad de que pertenezcan a la población de error aleatorio sea infinitesimal.

El error sistemático puede ser constante o variable, en función de que lo sea la causa que los origina: si una distancia se mide con una cinta cuyo valor nominal no coincide con el real, se producirá un error sistemático. Si esa cinta es metálica, le afectarán los cambios de temperatura, una causa variable, según una ley determinada en función de las diferencias de temperatura y el coeficiente de dilatación de la cinta.

El error sistemático tiene una influencia más relevante que el aleatorio en la propagación del error ya que, en el segundo, las desviaciones positivas y negativas tienden a anularse. Lógicamente, si las causas se determinan, su magnitud puede conocerse y consecuentemente eliminarse; para el ejemplo anterior, existen formulaciones que determinan la cuantía del error sistemático (Wolf and Ghilani, 2002:136).

A tenor de lo dicho anteriormente la expresión general: $m_i = m + \varepsilon_i$ sigue siendo válida pero considerando que ε_i puede desglosarse en una componente aleatoria y otra sistemática.

1.3 Exactitud, precisión e incertidumbre

Según la definición oficial (CEM, 2000:22) la exactitud de medida es el “grado de concordancia entre el resultado de una medición y un valor verdadero del mensurando”, indicando que el concepto de exactitud es cualitativo.

En nuestra opinión, esta definición no es adecuada por dos motivos:

- Es difícil de diferenciar de la de error de medida: “resultado de una medición menos un valor verdadero del mensurando” (CEM, 2000:24) .
- Estamos en desacuerdo con que el ‘concepto de exactitud’ sea cualitativo.

Como ya hemos indicado antes, estamos de acuerdo con la definición de error, que se aplica a resultados de medidas concretos y que es la que adoptamos en este trabajo. Respecto a la exactitud, planteamos lo siguiente:

- La exactitud es una propiedad del procedimiento de medida.
- Denominamos **exactitud** al grado de concordancia de una serie de medidas con el valor verdadero del mensurando.
- La exactitud es una propiedad cuantitativa ya que el ‘grado de concordancia’ se puede expresar mediante un estadístico como el error cuadrático medio (ECM):

$$ECM = \sqrt{\frac{\sum_i (m_i - m)^2}{n}}$$

Ec. 1-1. Error cuadrático medio.

Donde m_i es el resultado de la i -ésima medición; m es el valor verdadero y n es el número de réplicas o medidas.

La exactitud depende del método y del procedimiento de medida, por lo que habrá que poner especial interés en elegir métodos y procedimientos de medida que proporcionen los errores reducidos.

La definición de precisión no figura en el Vocabulario Internacional de Metrología pero a nuestro entender es un término necesario que complementa el de exactitud a la hora de caracterizar los procedimientos de medida. Usaremos en este trabajo los siguientes planteamientos:

- La precisión es una propiedad del procedimiento de medida.
- Denominamos **precisión** al número de cifras significativas con que se obtiene una serie de medidas.
- La precisión es una propiedad cuantitativa y se expresa en función de la unidad de medida.

La precisión está asociada al concepto de **resolución de la escala**, entendiendo que la resolución es la menor diferencia de indicación de un dispositivo visualizador que puede percibirse de forma significativa. Por ejemplo, en un termómetro analógico, donde la escala está graduada a intervalos de 1 °C y la distancia entre marcas es de 1 mm, la resolución efectiva estará, en el mejor de los casos, en $\frac{1}{4}$ °C ya que a simple vista no es posible diferenciar menores distancias. Si el termómetro digital mencionado en un ejemplo anterior posee dos dígitos decimales su precisión está limitada a 0.01 °C. Debe notarse que esta definición no se establece en función del valor verdadero, como ocurre con la exactitud.

Precisión y exactitud suelen estar relacionados pero son conceptos diferentes: una cinta métrica puede ser muy precisa al presentar divisiones hasta milímetros, o incluso acompañarse de dispositivos para aumentar esa resolución de la escala, pero al mismo tiempo ser poco exacta si presenta un error sistemático o si la estabilidad dimensional no es suficiente.

Asociados a la noción de precisión se definen los conceptos de **repetibilidad** y **reproducibilidad**, más concretamente se habla de precisión en condiciones de repetibilidad y precisión en condiciones de reproducibilidad.

La **repetibilidad** es el grado de concordancia entre los resultados de sucesivas mediciones del mismo mensurando efectuadas con aplicación de la totalidad de las mismas condiciones de medida. Estas condiciones de medida comprenden el mismo procedimiento, mismo observador, lugar, instrumento utilizado en las mismas condiciones y la repetición durante un corto periodo de tiempo. La observación de una distancia entre dos puntos varias veces seguidas, con una misma cinta métrica y los mismos operadores y forma de trabajo, proporcionará unos resultados de precisión en condiciones de repetibilidad.

La **reproducibilidad** es el grado de concordancia entre los resultados de las mediciones del mismo mensurando bajo diferentes condiciones de medida. Valdría como ejemplo el mismo del párrafo anterior con sólo variar alguna de las condiciones de medida.

La definición de los conceptos anteriores nos permite definir el de **incertidumbre**, término relativamente nuevo dentro del vocabulario metrológico. Lamentablemente, los

términos error e incertidumbre se han utilizado frecuentemente como sinónimos, cuando su naturaleza es completamente diferente.

La definición oficial (CEM, 2000:24) de incertidumbre de la medida es un “parámetro, asociado al resultado de una medición, que caracteriza la dispersión de los valores que podrían ser razonablemente atribuidos al mensurando”. En este trabajo compartimos plenamente esta definición así como las notas que la acompañan (no literales):

- El parámetro puede ser, por ejemplo, una desviación típica o la semiamplitud de un intervalo con un nivel de confianza determinado.
- Algunos componentes de la incertidumbre pueden ser evaluados a partir de la distribución estadística de los resultados de series de mediciones y pueden caracterizarse por sus desviaciones típicas experimentales.

Así, por ejemplo, a partir de una serie de medidas con el termómetro digital que hemos usado como ejemplo, obtenemos que los errores obedecen a una distribución $N(\bar{m}, s)$ donde $\bar{m} = 0$ y $s = 0.08$. Cualquier nueva medida m_i podrá y deberá ser expresada como, por ejemplo, $m_i \pm 1.96 \cdot 0.08$, donde de acuerdo con las características de la distribución normal, el intervalo responde a un nivel de confianza del 95%. Estamos diciendo, por tanto que la probabilidad de que el valor verdadero se encuentre dentro del intervalo $m_i \pm 0.16$ es del 95%. Lógicamente, la incertidumbre se asocia a los datos (resultados de las medidas) y depende absolutamente del método y procedimiento de medida.

La expresión de la incertidumbre asociada a un dato hace énfasis en que el resultado de una medición no es más que una estimación (en el sentido estadístico) que refleja la imposibilidad de conocer exactamente el valor del mensurando, por la indeterminación de los errores aleatorios y la corrección imperfecta de los errores sistemáticos.

De acuerdo con lo expresado hasta ahora, la diferencia entre error e incertidumbre es clara: el error asociado a un dato es la diferencia entre ese dato y el valor verdadero; como el error no puede ser conocido para la totalidad de los datos, se asocia a cada resultado de una medición un valor de incertidumbre que representa la distribución estimada del error. Esta diferencia se ilustra en la Fig. 1-2 (Perruchet and Priel, 2000:24).

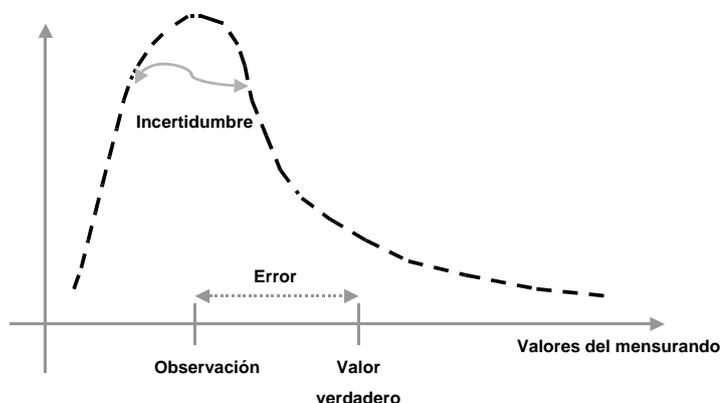


Fig. 1-2. Error e incertidumbre.

Se representa en esta figura el error como diferencia entre el valor verdadero y la observación, mientras que la incertidumbre en la medida aparece por no poderse conocer el error para la totalidad de los datos.

El concepto de incertidumbre se puede entender también como una cuantificación de la calidad del resultado anunciado y sirve como herramienta para la declaración de conformidad con respecto a unas especificaciones o a la comparación de productos

(Perruchet and Priel, 2000:38). Sin la expresión de la incertidumbre de una medida los resultados de diferentes mediciones no pueden compararse entre sí, ni con respecto a otros valores de referencia, lo que sirve para reforzar el concepto de incertidumbre como término genérico para medir la bondad de una medición.

1.4 Calidad y normalización

El creciente aumento en el uso, distribución y tratamiento de datos espaciales facilita a usuarios de diversa procedencia el acceso a todo tipos de datos geográficos. Obviamente esta 'transacción' de datos puede, en algunos casos, no merecer la confianza de los propios usuarios, por lo que es necesario establecer una garantía en la calidad de los datos. Nos encontramos, pues, ante otro concepto de vital importancia en el tratamiento de los datos geográficos: la **calidad**.

1.4.1 Definiciones

La calidad, según definición del Diccionario de la Lengua Española, es la "propiedad o conjunto de propiedades inherentes a algo, que permiten juzgar su valor".

Según (Ariza López, 2002:3) subyace un aspecto de subjetividad en el momento en que la apreciación y la comparación de las propiedades de ese objeto con respecto a otros es función del observador. Será difícil que la ejecución de un trabajo concreto, sea considerada de igual calidad por diferentes personas, sin entrar a matizar los diferentes grados de calidad posibles, según el nivel de exigencia establecido.

Sin embargo, es perfectamente posible (y necesario) que la calidad sea definida en términos objetivos y, consecuentemente, evaluable con independencia del observador.

Una definición de la **calidad** más cercana a este objetivo se refiere a la "totalidad de las características de un producto que le otorgan su aptitud para satisfacer necesidades establecidas e implícitas" (ICONTEC, 2000) o "grado en el que un conjunto de características inherentes cumple con los requisitos (necesidades o expectativas establecidas, generalmente implícitas u obligatorias)" (ISO9000, 2000).

Esta definición plantea una relación directa entre la calidad de un producto y la idoneidad para su uso. Para eliminar la indeterminación que supondrían las apreciaciones subjetivas y llevar el concepto de calidad a un lenguaje científico, ésta se relaciona con **procesos de normalización** con el fin de determinar los requisitos necesarios para un ajuste de conformidad con unas especificaciones dadas. El control de calidad de una cartografía, por ejemplo, necesita un proceso normalizador aceptado y conocido por la comunidad de usuarios.

Según lo indicado en los párrafos anteriores, se considerará, en lo referente a esta Tesis, la definición de **calidad** como "la idoneidad de un producto para satisfacer unas necesidades específicas". Formalmente, esa idoneidad se evaluará mediante el análisis de hasta qué punto se cumplen las especificaciones técnicas de una **norma**, que reflejan las mencionadas 'necesidades específicas'.

Según la Asociación Española de Normalización y Certificación (AENOR, 1991), una **norma** es “un documento de aplicación voluntaria que contiene especificaciones técnicas basadas en los resultados de la experiencia y el desarrollo tecnológico”.

Una norma es fruto del consenso entre las partes interesadas e involucradas en la actividad objeto de la misma. Las normas se convierten en herramientas para mejorar la calidad en la gestión de empresas, fabricación de productos y desarrollo de servicios y permiten hablar un mismo idioma a todos los sectores implicados. En la actualidad, prácticamente todos los aspectos productivos están normalizados.

Los datos espaciales normalmente tienen diferentes orígenes y distintos niveles de calidad, lo que conlleva la necesidad de establecer normas para determinar esos niveles. Estas normas, aunque en principio auspiciadas desde ambientes académicos y de investigación, tienen actualmente una vertiente eminentemente práctica pues tienen que responder a la demanda de calidad de un creciente número de usuarios de todo tipo, sobre todo dentro de un ámbito comercial.

1.4.2 Elementos de calidad

En la elaboración de una norma aplicada a datos espaciales podrán tenerse en cuenta diversos elementos de calidad. Un **elemento de calidad** es un aspecto de la calidad de los datos que puede ser medido cuantitativamente. Su propósito es identificar las inconsistencias de un conjunto de datos con respecto a las especificaciones técnicas (ICONTEC, 2000).

Para describir los aspectos cuantitativos de calidad en datos espaciales se han descrito diferentes elementos. Una síntesis básica de los mismos sería la siguiente (Ariza López, 2002:23; ICONTEC, 2000):

- Exactitud posicional:
 - Describe la cercanía en posición de los objetos en el conjunto de datos, con respecto a su posición verdadera (o la asumida como verdadera) (ICONTEC, 2000). Indica el grado de proximidad entre la posición del objeto en la base de datos y en la realidad, referida al terreno, tanto en su posición planimétrica como altimétrica y se determina según el error medio cuadrático (ICA).

- Exactitud temática:
 - Describe el grado de fidelidad de los valores de los atributos asignados a los elementos en la base de datos con respecto de su verdadera característica presentada en el terreno y la clasificación correcta de los objetos y sus relaciones (ICONTEC, 2000).
 - Definición análoga a la dada para la exactitud posicional pero referida a las propiedades no espaciales. Ejemplo de este parámetro lo tenemos en la medida de la correcta o incorrecta clasificación de las cubiertas de un terreno en la base de datos en relación con la realidad, o en la asignación errónea de un elemento en un mapa, como un vértice geodésico o una edificación (ICA).

- Exactitud temporal:
 - Describe el grado de realidad en la escala del tiempo de los elementos existentes en la base de datos con respecto de las especificaciones del producto (ICONTEC, 2000).
 - Determina el grado de actualización de los datos y debe ser contemplada como el límite temporal dentro del cual el dato es válido (ICA).
- Compleción [completitud]⁶:
 - Describe el nivel de veracidad con el cual los elementos capturados, sus atributos y sus relaciones representan el universo abstracto definido en las especificaciones del producto (ICONTEC, 2000).
 - Representa lo completa que está una base de datos en relación con el mundo real (ICA).
- Consistencia lógica:
 - Describe el grado de certidumbre con el cual un determinado conjunto de datos cumple con las especificaciones en lo que respecta a la estructura interna de los datos y topología (ICONTEC, 2000).
 - Representa el grado de cumplimiento de las relaciones lógicas entre los elementos de la base de datos y sus características. Es la ausencia de contradicciones entre los datos, y no sólo desde un punto de vista espacial, con el cumplimiento de relaciones topológicas, sino también temporal, sólo puede ocurrir un evento en un momento y posición dada, o temática, con inexistencia de contradicciones entre atributos temáticos redundantes.
- Genealogía y linaje: describe los antecedentes y la historia de la base de datos, especificando el origen y la trayectoria de los datos que conforman la base.

Entendemos que no todos estos elementos son aplicables a nuestro caso, referido al error espacial y sus consecuencias, por lo que en este trabajo hemos adaptado y redefinido parcialmente estas últimas directrices de la forma que presentamos a continuación. Nos quedamos con los siguientes elementos y subelementos de la calidad aplicables a bases de datos espaciales:

- **Integridad:** grado de cumplimiento de que todos los datos necesarios para la descripción y comprensión de la realidad están presentes en la base de datos.

Por ejemplo, un modelo digital de elevaciones en formato matricial debe tener una cota asociada a cada celda. Si existen celdas sin cota (vacías) se está rebajando la condición de integridad y deberá calcularse en, por ejemplo, porcentaje de celdas

⁶ “Completitud” o “compleción” son barbarismos derivados del término inglés “completeness” que consideramos innecesarios por existir en español “integridad” palabra que ya aparece en el Diccionario de Autoridades de 1734 con la definición “perfección que constituye las cosas en estado de no faltarles parte alguna...” que se ajusta adecuadamente a nuestro tema de trabajo.

vacías frente a celdas totales. Debemos destacar que estas omisiones no son errores en el sentido que hemos tratado hasta el momento ya que no se introduce un valor erróneo sino que se deja un vacío (que no es equivalente a un cero).

- **Consistencia** (sin.: solidez, estabilidad, firmeza): cumplimiento de condiciones que garantizan la coherencia de la estructura interna de los datos y de las relaciones entre ellos.
 - **Consistencia de dominio:** garantiza que los valores recogidos están en el rango de valores permitidos (aceptables) para el campo.
 - **Consistencia topológica:** garantiza que la estructura de los elementos cumple las propiedades topológicas; por ejemplo, que dos líneas que se cortan tienen su nodo inicial o final en ese punto o que en una secuencia de puntos tomada con GPS en modo cinemático cada punto tiene correctamente asignado el número de orden en la serie.

- **Incertidumbre**
 - **De la posición:** describe el grado de dispersión atribuible a la posición de los objetos y estimada a partir de la medida de la exactitud de una muestra de los mismos.
 - **Temporal:** en su caso, describe el grado de dispersión atribuible al atributo temporal de los objetos; por ejemplo, duración, fecha/hora, etc.
 - **Temática:** describe el grado de ajuste de los valores de los atributos asignados a los elementos en la base de datos con respecto de su clasificación correcta; es una incertidumbre atribuible a campos nominales (clasificaciones) y puede expresarse mediante estadísticos como el índice kappa o el coeficiente de ajuste.

Se observa que hemos sustituido “exactitud” por “incertidumbre”. La idoneidad del término se justifica en las definiciones previas ya que el elemento de calidad se aplica al conjunto de datos. Hemos obviado el elemento “linaje” ya que, aunque su importancia es grande, se trata de información propia de un conjunto de metadatos pero entendemos que no es directamente asociable a una medida de la calidad. Finalmente, hemos agrupado términos esencialmente comunes (las incertidumbres, por ejemplo) y los hemos propuesto como subelementos.

1.4.3 Antecedentes de normalización de datos espaciales

La exactitud posicional de una base de datos espacial se puede contrastar con el cumplimiento de una norma o un estándar predefinido, como se ha indicado anteriormente. Son varios los organismos dependientes de gobiernos u organizaciones de normalización, nacionales e internacionales, que desarrollan normas relativas a los datos espaciales en general (transferencia, intercambio de datos, formatos...) y a su calidad, en particular. Como las normativas sobre exactitud posicional son relativamente escasas, se hace a continuación una revisión de los principales organismos generadores de normas sobre datos espaciales (Caprioli and Tarantino, 2004), invitando al lector al apartado 1 del anexo para una explicación más profunda sobre las normas emitidas.

- **Organizaciones de ámbito nacional**

Uno de los organismos que mejor y más profusamente tiene normalizada la actividad cartográfica es el *Federal Geographic Data Comitee* (FGDC). Este organismo de la administración estadounidense se encarga de promover el desarrollo, uso y distribución de los datos geográficos. Dividido en diferentes subcomités, el FGDC elabora normas que regulan los contenidos y la calidad de los datos geográficos, así como el intercambio de información y la transferencia de datos geográficos para reducir esfuerzos y duplicidades.

El FGDC es la agencia responsable del *National Spatial Data Infrastructure* (NSDI), que pretende crear un marco de trabajo de datos digitales geoespaciales, desarrollando normas que permiten el correcto intercambio de datos digitales geoespaciales. Uno de los trabajos desarrollados por el FGDC es la elaboración del *Geospatial Positioning Accuracy Standards* ⁷, dividido en varias partes, una de las cuales es la norma *National Standard for Spatial Data Accuracy* (NSSDA).

El test NSSDA se desarrolló con la intención de actualizar la norma *National Map Accuracy Standard* y el test ASPRS, que determina la exactitud de mapas a gran escala, de la *American Society of Photogrammetry and Remote Sensing*⁸. El test *Engineering Map Accuracy Standard* (EMAS) se gestó en los años ochenta como alternativa al NMAS para grandes escalas, siguiendo, en general, todas estas normas unas pautas similares al NSSDA para determinar la exactitud posicional.

Otro organismo normalizador dentro del ámbito de los datos espaciales es el **Canadian Geospatial Data Infrastructure**⁹, dentro del cual el gobierno canadiense ha creado la *Inter-Agency Committee on Geomatics*, que tiene entre otras misiones la de desarrollar normas de información geoespacial dentro de la llamada iniciativa Mercator.

El **Canadian General Standards Board**¹⁰, en su comité de Geomática, aprobó en 1993 el *Spatial Archiving and Interchange Format* (SAIF) con la idea de ser un formato de intercambio y almacenamiento de datos.

En el caso concreto del tratamiento de Modelos Digitales de Elevaciones (MDE) el *United States Geological Survey* (USGS)¹¹ de los Estados Unidos ha desarrollado *Standards for Digital Elevation Model*.

En España, **AENOR**¹², es una entidad dedicada al desarrollo de la normalización y certificación en todos los sectores industriales y de servicios. Este organismo tiene reconocida su capacidad normalizadora y de certificación por Orden del Ministerio de Industria y Energía del 26 de febrero de 1986 y con diversos Reales Decretos. AENOR es la responsable de la elaboración de las normas **UNE**.

⁷ <http://fgdc.er.usgs.gov/standards/standards.html>

⁸ <http://www.asprs.org>

⁹ <http://www.cgdi.gc.ca/CGDI.cfm/fuseaction/home.welcome/lang/E/gcs.cfm>

¹⁰ <http://www.pwgsc.gc.ca/cgsb/home/index-e.html>

¹¹ <http://rockyweb.cr.usgs.gov/nmpstds/demstds.html>

¹² <http://www.aenor.es>

Otros organismos nacionales vinculados a la normalización son el *American National Standards Institute* (ANSI)¹³ y la organización alemana *Deutsches Institut für Normung* (DIN)¹⁴.

- **Organizaciones de ámbito internacional**

La organización internacional *Open Geospatial Consortium* (OGC)¹⁵ aglutina a compañías privadas, agencias gubernamentales y a sectores académicos para el desarrollo de datos espaciales y normas para su tratamiento.

La **Organización Internacional para la Estandarización (ISO)**¹⁶, en su Comité Técnico TC211, denominado *Geographic Information/Geomatics*, estableció varios grupos de trabajo para producir especificaciones técnicas que se convertirán en la familia de normas ISO 15046, con la idea de desarrollar modelos de datos geoespaciales y de geoproceto, aunque los detalles de implementación se desarrollen por otras organizaciones como *Open GIS Consortium*.

La *Asociación Cartográfica Internacional*¹⁷ promueve el desarrollo cartográfico en un contexto internacional gracias al trabajo de diferentes comisiones, entre las que destaca la Comisión sobre Estándares de Datos Espaciales¹⁸.

El *Comité Europeo para la Normalización* (CEN)¹⁹, en su comité técnico 287 aborda el tema de las normas geoespaciales para desarrollar un modelo de calidad de la información geográfica²⁰.

Dentro de la *OTAN*, el grupo de trabajo de Información Geográfica Digital (DGIWG) desarrolla una norma de formato de intercambio de datos geoespaciales para aplicaciones militares, la *Digital Geographic Information Exchange Standards* (DIGEST)²¹.

1.5 Objetivos del trabajo

Según lo expuesto hasta ahora, se pueden hacer las siguientes puntualizaciones:

- Existe un cierto grado de confusión en el uso de términos metrológicos, sobre todo respecto a las acepciones relativas a la expresión del error y la incertidumbre.

¹³ <http://www.ansi.org/>

¹⁴ <http://www2.din.de/>

¹⁵ <http://www.opengeospatial.org>

¹⁶ <http://www.isotc211.org/>

¹⁷ <http://www.icaci.org/>

¹⁸ <http://ncl.sbs.ohio-state.edu/ica/>

¹⁹ <http://www.cenorm.be/cenorm/index.htm>

²⁰ <http://www.cenorm.be/CENORM/BusinessDomains/TechnicalCommitteesWorkshops/CENTechnicalCommittees/CENTechnicalCommittees.asp?param=6268&title=CEN%2FTC+287>

²¹ <http://www.digest.org/>

- Es necesario fijar unos estándares de calidad para la exactitud posicional en el uso de los datos espaciales para satisfacer la demanda de calidad de un creciente número de usuarios de este tipo de datos, sobre todo dentro de un ámbito comercial.
- Se debe conocer la incertidumbre de los datos espaciales y cómo se propaga ésta en las diferentes operaciones y tratamientos a los que se someta a los datos.

Algunos de estos problemas han quedado solucionados al tiempo que se ponían en evidencia:

- En los apartados 1.2 y 1.3 se han definido y acotado las diferentes acepciones de error e incertidumbre.
- Se ha realizado una revisión de las principales normas actuales relativas a la exactitud posicional de los datos espaciales en el apartado 1.4.3, con una ampliación de las mismas en el apartado 1 del anexo a esta Tesis.

Pero un problema que no ha sido solventado aún, es el desarrollo de una teoría sobre la **naturaleza vectorial del error de posición**. Tal como se indicó en la introducción a esta Tesis, el error espacial, hasta la fecha, ha sido tratado como una magnitud escalar, con el cálculo del error medio cuadrático como estadístico básico. Si entendemos que el error de posición de un punto es la diferencia entre la posición real y la observada (en 2 ó 3 dimensiones) tenemos un vector (en el plano o en el espacio) que no está definido completamente si sólo señalamos su magnitud (el módulo del vector) siendo necesaria la indicación del acimut y el sentido del vector.

Además, a diferencia de lo que ocurre con el tratamiento conceptual y estadístico de modelos para representar la distribución de magnitudes escalares, la bibliografía para el tratamiento de modelos vectoriales es más bien escasa. De hecho, en la revisión de normativas anteriores, no aparece referencia alguna a la naturaleza vectorial de los errores de posición.

Entendemos, por tanto, que existen una serie de lagunas en cuanto al **tratamiento integral del error de posición en las bases de datos espaciales**, siendo el objetivo de esta Tesis desarrollar un **modelo vectorial para el error de posición** en 2 dimensiones que contemple:

- a) Aplicación de los estadísticos básicos para este tipo de modelos.
- b) Propuesta de modelos espaciales para el control del error.
- c) Propuesta de una normativa de control.

1.6 Organización de la Tesis

Llegados a este punto resulta más fácil explicar al lector cómo se va a organizar la Tesis, pues debe estar claro, tras la lectura de las líneas precedentes, cuál es el problema y cómo se quiere abordar. La Tesis está organizada en seis secciones o partes, tal como se indica a continuación:

- Parte 1. Conceptos básicos y definiciones.
- Parte 2. Planteamiento del error de posición como vector, así como revisión de las fuentes y propagación de incertidumbre en datos espaciales.
- Parte 3. Explicación del material (datos) y métodos estadísticos utilizados para el tratamiento de dichos datos.
- Parte 4. Cálculos y resultados.
- Parte 5. Propuesta de una norma para el control del error de posición en bases de datos espaciales.
- Parte 6. Discusión y conclusiones.
- Parte 7. Líneas futuras.

Se acompaña esta Tesis doctoral con un **anexo** (al final del texto) donde se amplían y explican ciertos conceptos, que aunque deben aparecer en este trabajo, son de importancia secundaria en la lectura del mismo o interferirían en un más cómodo análisis. Nos referimos a todos los datos manejados, al software utilizado o a las características técnicas del instrumental utilizado para la obtención de datos, entre otros.

2 La naturaleza vectorial del error de posición

Las definiciones comentadas y propuestas en el capítulo anterior hacen referencia en todos los casos a magnitudes escalares. En una representación gráfica, sus valores se distribuirían sobre un único eje, por lo que podemos también hablar de magnitudes unidimensionales o lineales. Para la definición, por ejemplo, de la exactitud, hemos utilizado como comparación la medida de la temperatura; asimismo, hemos usado una distribución de Gauss como descriptora de la distribución de los errores.

Sin embargo, si hablamos de la posición en un plano o en un espacio de tres dimensiones, no es suficiente con este enfoque. Como comentamos en la introducción, los datos espaciales se representan mediante tuplas de coordenadas que definen la posición del objeto en un sistema de referencia. En el plano y en el espacio, las tuplas adoptan respectivamente la forma (x_i, y_i) o (x_i, y_i, z_i) . Cada elemento de la tupla está afectado por una componente de error: $(x_i + \varepsilon_x, y_i + \varepsilon_y, z_i + \varepsilon_z)$ de donde se deduce que los valores $(\varepsilon_x, \varepsilon_y, \varepsilon_z)$ representan un **vector de error**, tal como se indica en la figura 2-1.

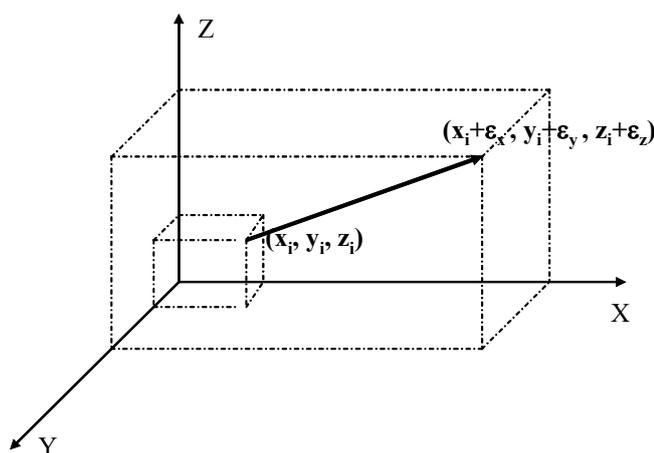


Fig. 2-1. Vector de error de posición espacial. Los datos espaciales se representan mediante tuplas de coordenadas. Las tuplas son de la forma (x_i, y_i, z_i) . Cada elemento de la tupla está afectado por una componente de error $(x_i + \varepsilon_x, y_i + \varepsilon_y, z_i + \varepsilon_z)$, representando un vector de error.

En función de esta forma de abordar el problema, debemos revisar algunas de las expresiones planteadas en el capítulo anterior. Las definiciones propuestas son válidas pero la estadística implicada cambia: debemos utilizar estadística vectorial que, según el caso bi o tridimensional, se denominan también **estadística circular** y **estadística esférica**. Revisamos a continuación algunos conceptos para adaptarlos al nuevo marco de referencia.

La definición general de **error** sigue siendo válida pero en el caso vectorial, la expresión varía: sobre el plano XY, el error tiene la forma $(\varepsilon_x, \varepsilon_y)$, un vector bidimensional con origen en el valor verdadero y final en el resultado de la medida. El aspecto clave de esta definición es que tenemos acceso a las propiedades espaciales de los vectores, que son: **módulo**, **acimut** (métricas) y **sentido** (topológica). Estas propiedades permitirán analizar características, como la isotropía, que no eran determinadas en el caso de datos unidimensionales.

En la figura 2-2 se ilustra la superposición de dos cartografías de una zona de las Vegas Bajas del río Guadiana a escala 1/25 000 y 1/50 000. Se han marcado con vectores las diferencias de posición entre varios puntos homólogos en ambas cartografías. Esta

diferencia de posición implica una **distancia** y una **dirección y sentido**, hecho que habitualmente se obvia en el tratamiento de error y que es lo que se quiere poner de manifiesto en esta Tesis.



Fig. 2-2. Ejemplo de error de posición en dos dimensiones en cartografía. En la figura se representan superpuestas dos cartografías a escala 1/25000 y 1/50000. Los vectores indican la diferencia entre las posiciones de algunos elementos comunes en ambas cartografías. Esta diferencia de posición implica una distancia y una dirección y sentido que habitualmente se obvia en el tratamiento del error.

2.1 Error aleatorio y error sistemático

De acuerdo con esta concepción del error, revisamos las definiciones de error aleatorio y error sistemático:

- **Error aleatorio:** categoría de error donde las desviaciones de las medidas respecto al valor verdadero se ajustan a una distribución uniforme. En esta distribución, la dirección media de la población, μ , no está definida debido a que el vector medio tiene módulo nulo (la longitud media de la población, ρ , es igual a cero) y la dispersión circular δ , es infinita. La función de densidad es, en el caso bidimensional, representando por θ al acimut:

$$f(\theta) = \frac{1}{2\pi} \quad 0 \leq \theta \leq 2\pi$$

Ec. 2-1. Función de densidad de probabilidad de la distribución uniforme.

- **Error sistemático:** categoría de error donde las desviaciones de las medidas respecto al valor verdadero no se ajustan a una distribución uniforme. En un caso estándar los errores se ajustarán a una distribución de von Mises $VM(\mu, \kappa)$ donde μ es la dirección media y κ es un parámetro de concentración (cuando κ tiende a cero, la distribución converge al caso uniforme). La función de densidad de probabilidad es:

$$f(\theta) = [2\pi I_0(\kappa)]^{-1} \exp[\kappa \cos(\theta - \mu)]$$

$$0 \leq \theta \leq 2\pi \quad 0 \leq \kappa \leq \infty$$

Ec. 2-2. Función de densidad de probabilidad de la distribución von Mises.

Siendo $I_0(\kappa)$ la función modificada de Bessel de orden cero (Fisher, 1995:48) explicada al final de la Tesis junto con otros estadísticos y parámetros.

Como en el caso de las magnitudes escalares, los parámetros de las distribuciones deberán estimarse mediante un número suficientemente elevado de observaciones. También será posible la determinación de errores groseros a partir de las funciones de probabilidad, independientemente del tipo de distribución.

La definición de exactitud (grado de concordancia de una serie de medidas con el valor verdadero del mensurando) sigue siendo válida pero la estadística asociada debe cambiar.

2.2 Exactitud, precisión e incertidumbre

De forma análoga a lo indicado para datos unidimensionales, la **exactitud de medida** para **datos bidimensionales** queda definida según las siguientes afirmaciones.

- La exactitud es una propiedad del procedimiento de medida.
- Denominamos **exactitud** al grado de concordancia de una serie de medidas con el valor verdadero del mensurando, con la salvedad que las medidas están realizadas sobre el plano y no sobre un eje. Para ilustrar esta idea, se presentan en la figura 2-3 las punterías realizadas sobre una diana en dos casos. En el primero, los diferentes valores de punterías se consideran más exactos que en el segundo caso por estar muy próximos al centro de la diana, mientras que la segunda imagen muestra valores más separados de la posición correcta.
- La exactitud es una propiedad cuantitativa cuyo grado de concordancia se puede expresar mediante un estadístico como el módulo medio que se explicará posteriormente, junto con otros estadísticos para datos circulares (Apartado 3.2.1.1).

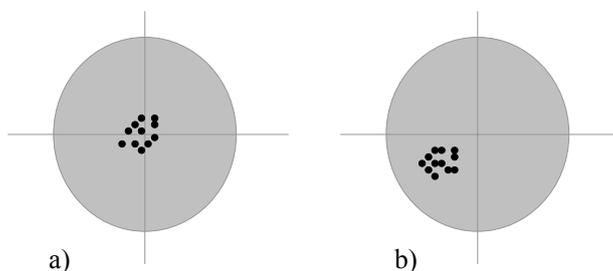


Fig. 2-3. Concepto de exactitud en datos bidimensionales.

En el gráfico a) las punterías a una diana se sitúan en torno al valor central, considerado la posición correcta, mientras que en el caso b) las punterías son menos exactas por estar más separadas de dicha posición.

La precisión de una serie de medidas de datos circulares se define por los siguientes enunciados:

- La precisión es una propiedad del procedimiento de medida.
- Denominamos **precisión** al número de cifras significativas con que se obtiene una serie de medidas. En la figura 2-4, siguiendo con el ejemplo de las punterías a dianas, se muestran punterías de diferente precisión. Para ello utilizamos puntos de mayor diámetro para hacer notar la diferencia de cifras significativas en la posición del mismo, estableciendo como leyenda el mayor diámetro del punto que se relaciona con un menor número de decimales en la posición.
- La precisión es una propiedad cuantitativa.

De igual forma que se indicó anteriormente la precisión está asociada al concepto de **resolución de la escala**, entendiendo que la resolución es la menor diferencia de indicación de un dispositivo visualizador que puede percibirse de forma significativa.

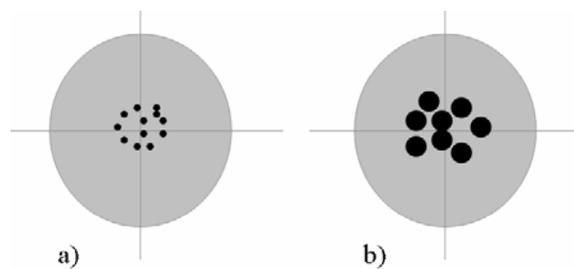


Fig. 2-4. Concepto de precisión en datos bidimensionales. Asociamos el tamaño en la representación del punto con el número de decimales con que está determinado; así, a mayor diámetro de la representación, menor número de cifras significativas.

La **incertidumbre** en la medida de datos bidimensionales dependerá, al igual que en el caso lineal, del método y del procedimiento de trabajo, estando asociada a los datos observados. Si los datos observados son un valor de distancia en X y otro en Y, que posteriormente se transformarán en módulo y acimut, el intervalo de confianza de ambas magnitudes creará un área de indeterminación con forma de cuadrado o rectángulo, según la exactitud de las medidas.

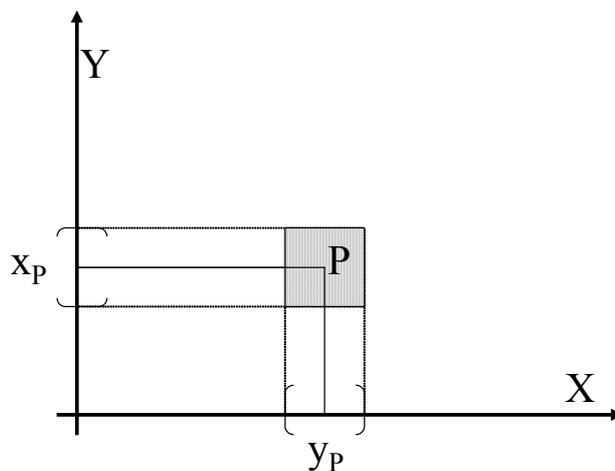


Fig. 2-5. Concepto de incertidumbre en datos bidimensionales.

La zona rayada representa la zona de incertidumbre de posición de datos bidimensionales, que podrá ser acotada aún más a un círculo o una elipse.

En esta tesitura, la posición de un punto P tendrá una indeterminación tanto en la dirección X como en Y que generará la región de incertidumbre, a un nivel de probabilidad, que aparece sombreada en la figura 2-5. Por teoría elemental de errores y probabilidad, sabemos que podemos acotar aún más la región de indeterminación hasta convertirla en un círculo inscrito en el cuadrado con radio la semiamplitud del intervalo de incertidumbre en X ó Y con un nivel de confianza determinado. En el caso de tener incertidumbres diferentes en X ó en Y, la región de indeterminación es un rectángulo y la superficie a tratar una elipse, denominada **elipse de tolerancia**, tal como se considera habitualmente en la aplicación de la indeterminación de posición en Topografía clásica (Ojeda, 1984:35).

2.3 Fuentes y propagación de incertidumbre en los datos espaciales

De todos los elementos de calidad de los datos espaciales (integridad, consistencia, incertidumbre...) es la **incertidumbre de posición** el elemento principal a tratar en esta Tesis. Por este motivo se especifican a continuación las posibles causas que originan incertidumbre en la posición de los datos espaciales (Ariza López, 2002; Pinilla, 1995). Se clasifican estas causas en tres grandes grupos, siguiendo el orden cronológico que sufren los datos al incorporarse a una base de datos espacial: a) la captura de los datos, b) el tratamiento de los mismos y c) el resultado final y la manipulación posterior. En las **causas de incertidumbre** que a continuación se enumeran, se admite que la base de datos se puede presentar tanto en formato digital como en papel.

a) Captura de los datos

- La toma directa de los datos en campo, con material topográfico diverso utilizando métodos y procedimientos de trabajos específicos.
- La digitalización de fuentes analógicas, si la base de datos se crea a partir de la digitalización (manual o escaneo) de dichas fuentes.
- En el caso de utilizar como fuente una cartografía tradicional en papel, ésta presenta generalizaciones, además de posibles deformaciones.
- Las especificaciones de la fotogrametría aérea generan unas limitaciones en las bases de datos creadas por este método.
- Los datos originados a partir de imágenes obtenidas por sensores remotos deben sufrir unas correcciones, que nunca serán perfectas, antes de su incorporación a la base de datos.

b) Tratamiento

- Aunque no es tratamiento de datos propiamente dicho, la insuficiente precisión numérica al almacenar los datos, origina incertidumbres.
- Al georreferenciar los datos, existe una incertidumbre en la definición del sistema de referencia.

- En la operación de interpolación existen incertidumbres inherentes al proceso, al estimar posiciones y atributos desconocidos de puntos a partir de otros conocidos.
- Cuando se cambia de escala se produce una generalización de datos, que supone siempre pérdida de información.
- En general, en la realización de operaciones diversas (álgebra de mapas) que permiten generar o transformar una base de datos nueva a partir de ecuaciones de transformación de la base de datos original, como operaciones aritméticas, lógicas...

c) Resultados y manipulación

- Salidas cartográficas deficientes por deformación del soporte, errores de escala, limitaciones de la impresión, etc.
- Resultado en sí, pues la base de datos se utiliza como modelo de la realidad, que por definición es una representación inexacta de la misma, produciéndose pérdida de información.
- Falta de actualización por variación temporal.
- Falta de información sobre la calidad de los datos (metadatos). El usuario puede someter a la base de datos a unos análisis o transformaciones indebidas por la carencia de información.

Las incertidumbres individuales provenientes de estas fuentes de error suelen ser pequeñas, pero colectivamente afectan de manera significativa a la exactitud de los datos.

Las incertidumbres existentes en una base de datos espacial deben ser consideradas cuando se explota dicha base, pero sobre todo, cuando esa información se transforma y se difunde en otras operaciones, como por ejemplo en un Sistema de Información Geográfica donde los errores en las entradas de datos afectarán a los cálculos y salidas de los mismos. Se debe analizar la forma en que esos errores se propagan y provocan incertidumbres en los resultados de las operaciones, pudiendo hablar de diferentes **modelos en la propagación de incertidumbres**.

Estos modelos de propagación de incertidumbres pueden ser divididos en una primera aproximación en modelos analíticos y modelos estocásticos. El desarrollo de los **modelos analíticos** para estudiar la propagación de errores, presenta en algunos casos serias dificultades de aplicación por la complejidad de las relaciones existentes entre los datos y los resultados. En estos casos la aplicación de **modelos estocásticos** aparece como otra posible solución, con la salvedad de que este tipo de modelos analizan los resultados finales, no el proceso, como los modelos matemáticos. Entre los principales modelos analíticos para estudiar la propagación de incertidumbres en las bases de datos espaciales podemos citar (Ariza López, 2002:20; Heuvelink, 1998:33-39): el método de estimación de la varianza, el modelo de Gauss, los métodos de primer y segundo orden de Taylor o el método de Rosenblueth, entre otros.

Dentro de los métodos estocásticos merecen toda nuestra atención los **métodos de Monte Carlo**, por revelarse como una herramienta efectiva cuando las soluciones analíticas son complejas. En esencia es la creación de un modelo, y por tanto una representación de la realidad, que permita, al analizarlo, la toma de decisiones. La creación de este modelo se apoya necesariamente en métodos informáticos. El método de Monte Carlo (Felicísimo, 1994; Felicísimo and Cuartero, 2005) puede utilizar la generación de diferentes modelos de error para analizar la propagación de errores e incluso para validar soluciones proporcionadas por procedimientos ya establecidos (González, Socorro *et al.*, 2005:3). La idea del método es calcular repetidamente el resultado de una operación sobre los datos originales introduciendo errores distintos en el proceso de modelado. Se obtienen diferentes bases de datos, o muestras de la base de datos, a partir de la original, con diferentes grados de distorsión introducidos, que producirán modelos que serán posteriormente analizados.

3 Material y métodos

Hasta el momento, con lo desarrollado en los puntos 1 y 2 de esta Tesis, hemos consolidado las siguientes ideas:

- **Redefinición de conceptos metrológicos** relativos a la medición de magnitudes y a la incertidumbre creada en la medición.
- **Establecimiento de la completa naturaleza de la incertidumbre de posición**, que debe ser considerada desde un punto de vista vectorial, no escalar.

Por tanto, se procederá en este punto, a desarrollar la metodología necesaria que permita plantear un modelo vectorial para la incertidumbre de posición en dos dimensiones. Para este propósito se necesita: a) un material (datos) y b) una metodología (tratamiento estadístico de datos circulares), tal como se desarrolla a continuación.

3.1 Material

Los datos sobre los que se aplicará el tratamiento estadístico para obtener un modelo vectorial en dos dimensiones, proceden de dos fuentes muy concretas: a) la calibración geométrica de un escáner fotogramétrico, en concreto el modelo Ultrascan 5000 (Vexcel, 2001) con una precisión geométrica de 2 μm sobre un área de barrido de 300 x 450 mm y b) varios grupos de coordenadas de puntos obtenidas de cartografía de una misma zona a diferentes escalas. Ambos conjuntos de datos se describen en los dos apartados siguientes.

3.1.1 Datos de la calibración de un escáner fotogramétrico

Un **escáner** es un instrumento que se utiliza para la digitalización de imágenes analógicas. Estas imágenes suelen provenir de cámaras convencionales pero también es frecuente otro tipo de imágenes como radiografías, termografías, etc. Para realizar los procesos habituales en fotogrametría digital se utilizan **escáneres fotogramétricos** que garantizan la calidad necesaria. Es habitual digitalizar negativos procedentes de cámaras métricas o semimétricas con formatos desde los 6x6 cm a los 23x23 (caso de cámaras aerotransportadas). Lógicamente, como cualquier otra transformación, el proceso de digitalización introduce errores que son esencialmente de tres tipos:

- **error radiométrico:** la mayor o menor opacidad del negativo debe traducirse a un nivel digital, habitualmente entre un valor de 0 (opaco) y 255 (transparente); diversos factores de esta transformación analógico-digital pueden ocasionar que un mismo nivel de opacidad se traduzca a diferentes (aunque similares) valores digitales en lecturas consecutivas.
- **error de generalización:** una imagen digital está formada por píxeles de un tamaño concreto. Un píxel recoge el valor digital medio de una zona cuadrada de

la imagen analógica, lo que implica una mayor generalización o pérdida de resolución espacial cuanto mayor sea el tamaño del píxel. En el proceso fotogramétrico suelen definirse tamaños entre 6 y 24 μm . Lógicamente, el error de generalización depende también de las características de la imagen analógica y muy especialmente del tamaño de grano y del enfoque.

- errores de posición: los píxeles de la imagen digital deben estar correctamente situados, es decir, deben reflejar la posición de los elementos de la imagen original sin distorsión.

Este último aspecto es el que nos interesa en este trabajo y el que va a ser analizado en los apartados siguientes. Para ello introducimos el concepto de calibración geométrica:

- la **calibración geométrica** es un proceso de medida y ajuste que tiene por objeto disminuir los errores espaciales en el proceso de digitalización para obtener la exactitud nominal exigida.

La calibración geométrica no es propia de los escáneres convencionales pero sí de los escáneres fotogramétricos (Baltsavias, 1994; Bethel, 1994) y se realiza escaneando una plantilla de un material dimensionalmente estable, en la que aparecen marcas en forma de cruces o puntos, a distancias regulares, formando una retícula cuya posición se conoce. Las plantillas pueden ser de vidrio (muy estables) o, como ha sido el caso, una plantilla de acetato de menor precio y estabilidad. La plantilla cubre el área de escaneo para poder cuantificar las deformaciones en toda la superficie.

Escaneada la plantilla y determinada la posición de las señales en la imagen digital, se realiza un ajuste entre ésta y las posiciones conocidas de cada punto, obteniéndose las desviaciones entre ambas posiciones. Estas desviaciones van a utilizarse para construir un modelo estadístico simple que persigue mejorar la exactitud geométrica en las sucesivas digitalizaciones de fotogramas.

La mejora de la exactitud se consigue aplicando a la imagen una transformación afín bidimensional calculada a partir de las desviaciones observadas y resolviendo el sistema de ecuaciones mediante un ajuste mínimo cuadrático, lo que permite calcular los residuos y el error cuadrático medio.

La transformación afín a la que se alude en el párrafo anterior es de la forma que se indica en la ecuación siguiente:

$$\begin{aligned} X_C &= a_1x + b_1y + c_1 \\ Y_C &= a_2x + b_2y + c_2 \end{aligned}$$

Ec. 3-1. Transformación afín bidimensional utilizada en la calibración geométrica.

Donde los valores X_C e Y_C , se corresponden con las coordenadas calibradas de los puntos de la plantilla, x e y son las coordenadas medidas de dichos puntos y los valores a_1 , a_2 , b_1 , b_2 , c_1 y c_2 son los parámetros de transformación a determinar tras el ajuste por mínimos cuadrados. Estos últimos parámetros c_1 y c_2 se refieren a las traslaciones en los ejes X e Y , respectivamente, mientras que el resto de los elementos dependen del factor de escala H_X e H_Y en los ejes X e Y , del valor θ , giro entre ambos sistemas de coordenadas y de δ , falta de ortogonalidad de los ejes, tal como aparece en la ecuación siguiente.

$$a_1 = H_x \cos \theta; \quad b_1 = H_y \sin(\delta - \theta)$$

$$a_2 = H_x \sin \theta; \quad b_2 = H_y \cos(\delta - \theta)$$

Ec. 3-2. Parámetros de transformación.

Estos parámetros aparecen resueltos en el ejemplo de la figura 3-1 donde se indica la cabecera del resultado de un fichero de calibración de un escáner fotogramétrico. En este fichero se muestran en primer lugar una serie de datos como la fecha y hora de la calibración, la plantilla utilizada o el número de puntos usados para la calibración. Posteriormente se indican los valores correspondientes a los parámetros de transformación que conforman la matriz de transformación.

```
Geometric Calibration Report for Scanner S3250
calprot.sh -O -r 50 50
Performed on Mon Feb 21 08:18:05 2005 by user Administrator
Calibration with Target 25 from 2005/02/08
6 parameter transformation of: <extract00000> and <500.000 x
500.000 raster>
50 x 50 points
PixelSize: 0.010
by: vxaffine.exe 1.13.1.17

Transformation matrix :
1.00006782    0.00047472  -15699.85839833
-0.00047646    0.99999298  -137.91833246

Decomposed transformation vectors:
Translation : -15699.85839833 -137.91833246
Rotation :    0.02719782 degree
Scaling :    1.00006794    0.99999309
Shearing :   -0.00000177    0.00000000
```

Fig. 3-1. Resultados de los parámetros de transformación en la calibración de un escáner fotogramétrico.

Cabecera del fichero de calibración de un escáner fotogramétrico, donde aparecen en primer lugar una serie de datos como la fecha, hora de calibración y número de puntos utilizados para el proceso y posteriormente el resultado de los parámetros de transformación que forman la matriz de transformación.

Como ya se ha dicho, en este trabajo se ha utilizado un escáner Ultrascan 5000 (Vexcel, 2001) con una precisión geométrica de 2 μm sobre un área de barrido de 300 x 450 mm. Debido a la plantilla disponible, la resolución queda limitada a 6 μm. En este escáner, como resultado de la calibración geométrica se obtiene un fichero (continuación del de la figura 3-1) con la estructura siguiente (figura 3-2, se presentan sólo 3 puntos a modo de muestra).

<i>transferred center</i>		<i>deviation before transf.</i>		<i>deviation after transf.</i>	
<i>Scan</i>	<i>CCD</i>	<i>Scan</i>	<i>CCD</i>	<i>Scan</i>	<i>CCD</i>
[mm]	[mm]	[mm]	[mm]	[um]	[um]
0.001	0.001	156.889	2.278	0.776	0.814
0.002	12.002	156.880	2.279	1.925	1.777
0.001	24.002	156.870	2.279	1.205	1.750

Fig. 3-2. Datos del fichero de calibración geométrica del escáner fotogramétrico.

Cuerpo del fichero de calibración del escáner fotogramétrico con los resultados de sólo 3 puntos a modo de muestra. Las dos primeras columnas se refieren a los puntos medidos en la plantilla de calibración en el eje horizontal (SCAN) y vertical (CCD). La dos columnas siguientes representan los valores antes de la transformación y la dos últimas, los valores después de la misma.

La primera y segunda columna (*transferred center*) representan los puntos medidos en la plantilla de calibración en el eje horizontal (SCAN) y vertical (CCD). La tercera y cuarta columna (*deviation before transformation*) se refieren a las desviaciones antes de la

transformación, es decir, la posición de los puntos medidos al hacer la lectura respecto a la situación en la que se esperaba que estuvieran. Finalmente, las dos últimas columnas (*deviation after transformation*) representan los valores de desviación después de la transformación geométrica, expresados en micras, que se corresponden con la lectura que va a existir en cada uno de los puntos que se muestra con respecto al punto teórico.

En este trabajo usaremos los valores de las dos últimas columnas indicadas en la figura 3-2 para el análisis del error tras la corrección geométrica. Estos valores deben considerarse como incrementos de las coordenadas X e Y que reflejan las diferencias entre la posición correcta y la corregida que, a pesar de las transformaciones, no están exentas de error. Las diferencias se pueden representar como vectores con origen en la posición teórica correcta y fin en la posición final de la imagen. Cada vector queda definido por su módulo (d) y acimut (θ) respecto al eje Y.

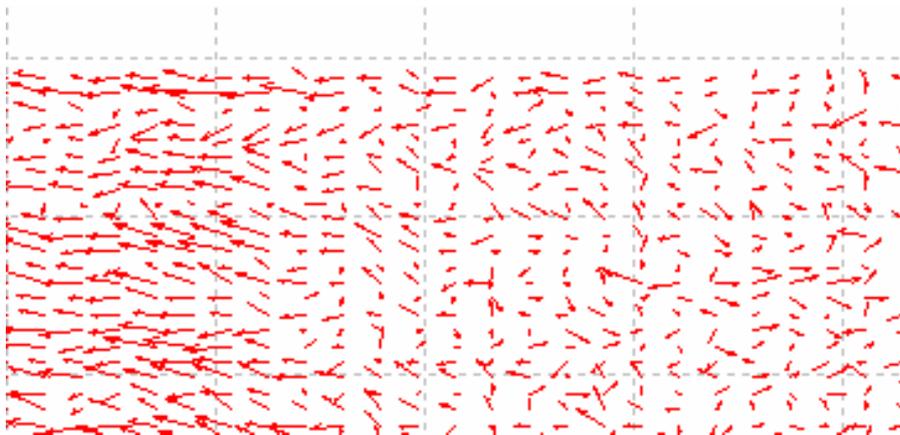


Fig. 3-3. Vectores de error en la calibración de un escáner fotogramétrico. Parte del resultado gráfico de la calibración geométrica del escáner donde se ven los vectores de error, apreciando en la parte izquierda de la imagen una tendencia similar en cuanto a la dirección y sentido de los vectores.

Este proceso de calibración geométrica del escáner fotogramétrico genera, por tanto, una serie de vectores de error en un espacio bidimensional, tal como se ve en la figura 3-3, sobre los que se aplicará un tratamiento estadístico de datos circulares para ver su comportamiento y distribución.

En el final del fichero de calibración (representado en la figura 3-4 y continuación de las figuras 3-1 y 3-2) aparece el histograma de frecuencia de valores de error y el error medio cuadrático y valores máximos y mínimos, antes y después de la transformación, tanto en la dirección SCAN como en la CCD.

```

Histogram:
  1um,  2um,  4um,  8um, 12um, 16um, 32um, more
Scan:  1550, 364, 190, 12,  0,  0,  0,  0
CCD:  1360, 179, 360, 215,  2,  0,  0,  0
Length: 194, 457, 955, 503,  7,  0,  0,  0

Statistics: [mm]
      rms scan  rms ccd  min scan  max scan  min ccd  max ccd  min lng  max lng
before 0.0319   0.0318 -156.9888 -156.8672 -1.5690  -1.4524
after  0.0015   0.0029  -0.0052   0.0044  -0.0081  0.0084  0.0001  0.0091

```

Fig. 3-4. Histograma en fichero de calibración de un escáner fotogramétrico.

Se indica la frecuencia de la magnitud del error en las direcciones SCAN y CCD, así como los estadísticos básicos como el error medio cuadrático y valores máximos y mínimos, antes y después de la transformación.

Con este planteamiento, se realizaron varias calibraciones, en concreto:

- 20 calibraciones a partir de una plantilla con una malla de 46x46 señales, lo que genera un conjunto de 2116 vectores de error por calibración. En realidad, tal como se indica en la figura 3-1, la malla es de 50x50, que tras un proceso de redondeo en el cálculo, se queda en 46x46. De las 20 calibraciones, 19 proceden de un mismo escáner, propiedad de la Universidad de Extremadura (UEX), realizadas entre octubre del 2004 y julio del 2005, lo cual supone al mismo tiempo una comprobación de la posible variación temporal en la estabilidad del instrumento. La última calibración, realizada en junio del 2006, fue proporcionada por el Departamento de Ciencias de la Tierra y Física de la materia condensada de la Universidad de Cantabria (UC), utilizando también un escáner de la misma marca y modelo que el citado, con la idea de tener alguna calibración de otro modelo para validar las posibles conclusiones.
- 2 calibraciones sobre una retícula de 19x19 (en realidad, y por el mismo motivo reseñado, se parte de una malla de 20x20, que producen 316 vectores). Una de las calibraciones procede del escáner de la UEX, utilizada para la exposición de métodos (Apartado 3.2.1) y la otra del departamento ya citado de la Universidad de Cantabria.

Cada fichero de calibración se designa por la fecha en la que fue realizada, siguiendo el criterio marcado por la norma **ISO 8601** ²². Según esta norma, la fecha debe ser consignada de la forma AAAA-MM-DD, correspondiente el valor AAAA, al año dentro del calendario gregoriano; MM, al mes entre el 01 (enero) y 12 (diciembre) y DD, al día entre 01 y 31. De igual forma aparecen las fechas en los resultados de los ficheros de calibración del escáner. Así, el nombre "2005-03-07" se corresponde con la calibración realizada el 7 de marzo del año 2005.

Las características técnicas completas y la procedencia del escáner fotogramétrico utilizado (Vexcel, 2001) aparecen desglosados en el anexo.

²² <http://www.iso.org/iso/en/prods-services/popstds/datesandtime.html>

3.1.2 Datos de la comparación de cartografía a diferentes escalas y procedencia

La cartografía utilizada está en formato digital y se corresponde con las siguientes zonas y formatos:

- Digitalización del Mapa Topográfico Nacional a escala 1/50 000 (MTN 50) en la provincia de Badajoz correspondientes a las hojas: 776 (Montijo), 777 (Mérida), 778 (Don Benito), 779 (Villanueva de la Serena), 780 (Puebla de Alcocer), 781 (Siruela), 803 (Almendralejo), 827 (Alconchel), 828 (Barcarrota) y 829 (Villafranca de los Barros).
- Hoja 776-IV en formato digital a escala 1/25 000 del Mapa Topográfico Nacional (MTN 25), denominada también Montijo.
- Cartografía a escala 1/10 000 de la misma zona abarcada por la hoja 776 del MTN 50, elaborada por la Consejería de Agricultura de la Junta de Extremadura también en formato digital.
- Cartografía digital a escala 1/25 000 de la provincia de Badajoz obtenida a través del Servicio Web de Mapas (*Web Map Server, WMS*) que permite visualizar la Base Cartográfica Numérica (BCN) a 1/25 000 del Instituto Geográfico Nacional (IGN).

La elección de esta zona obedece a varios motivos: a) es una zona amplia con elementos muy dispares desde núcleos de población, embalses, zonas de cultivos diferenciadas, etc. pues corresponde a las Vegas Bajas del río Guadiana, b) existen muchos puntos claramente identificados entre la cartografía y el terreno, pues hay multitud de caminos y acequias pertenecientes a la zona regable de Vegas Bajas, así como varias poblaciones con gran presencia de fincas y cortijos c) en caso de necesitar comprobación en campo es una zona cercana geográficamente al lugar de realización de esta Tesis.

Para el análisis de errores se procedió a identificar puntos comunes entre varias de las cartografías indicadas anteriormente, de tal forma que se obtuvieron datos de la comparación entre:

- La hoja 776-IV en formato digital del MTN 25 y la zona correspondiente digitalizada del MTN 50.
- La cartografía a escala 1/10 000 de la Junta de Extremadura y la correspondiente a la hoja 776 del MTN 50.
- La zona común entre las hojas 776, 777 y 778 del MTN 50 digitalizado y la Base Cartográfica Numérica a 1/25 000 del IGN obtenida a través de WMS.
- La zona común entre las hojas 779, 780 y 781 del MTN 50 digitalizado y la Base Cartográfica Numérica a 1/25 000 del IGN obtenida a través de WMS.
- La zona común entre las hojas 827, 828 y 829 del MTN 50 digitalizado y la Base Cartográfica Numérica a 1/25 000 del IGN obtenida a través de WMS.

- La zona común entre las hojas 777, 803 y 829 del MTN 50 digitalizado y la Base Cartográfica Numérica a 1/25 000 del IGN obtenida a través de WMS.

En la figura siguiente se muestra la superposición de dos cartografías, en concreto las correspondientes a escalas 1/10 000 y 1/50 000, donde se aprecian diferencias de posición entre puntos homólogos.



Fig. 3-5. Superposición de dos cartografías a diferentes escalas.

En esta figura aparecen superpuestas la hoja del MTN 50 y la cartografía a escala 1/10 000 correspondientes a una parte de la localidad pacense de Montijo. Aparte de los evidentes errores de posición, no aparecen los mismos elementos debido a la diferente fecha de elaboración de ambos conjuntos de datos.

Sobre las superposiciones antes citadas y mediante el tratamiento con visor apropiado, se obtuvieron las diferencias de coordenadas en X e Y de puntos definidos en ambas cartografías, que se transformaron en el módulo y el acimut de un vector de error. En total fueron:

- 500 vectores de error bidimensionales obtenidos de la diferencia de posición entre puntos homólogos localizados en la cartografía del MTN 25 y MTN 50.
- 500 vectores de error bidimensionales obtenidos de la diferencia de posición entre puntos homólogos localizados en la cartografía del mapa a 1/10 000 y MTN 50.
- 250 vectores de error bidimensionales obtenidos de la diferencia de posición entre puntos homólogos localizados en la zona común entre las hojas 776, 777 y 778 del MTN 50 digitalizado y la BCN a 1/25 000 del IGN.
- 250 vectores de error bidimensionales obtenidos de la diferencia de posición entre puntos homólogos localizados en la zona común entre las hojas 779, 780 y 781 del MTN 50 digitalizado y la BCN a 1/25 000 del IGN.
- 250 vectores de error bidimensionales obtenidos de la diferencia de posición entre puntos homólogos localizados en la zona común entre las hojas 827, 828 y 829 del MTN 50 digitalizado y la BCN a 1/25 000 del IGN.

- 250 vectores de error bidimensionales obtenidos de la diferencia de posición entre puntos homólogos localizados en la zona común entre las hojas 777, 803 y 829 del MTN 50 digitalizado y la BCN a 1/25 000 del IGN.

3.2 Metodología

Una vez definido el material de trabajo, se plantea ahora cómo obtener un modelo vectorial para la incertidumbre de posición en dos dimensiones. El tratamiento estadístico habitual para datos unidimensionales no es el adecuado para esta tarea por el motivo ya reseñado: tratamos la incertidumbre en la posición en forma de vector, con módulo, acimut y sentido.

La llamada **estadística circular** o el análisis estadístico de datos circulares permite el tratamiento de datos angulares. En muchos ámbitos de la ciencia se realizan mediciones angulares que a veces implican vectores (datos de viento, por ejemplo) o a veces solamente direcciones, sin sentido ni módulo (datos de buzamiento de estratos geológicos o dirección de fallas o fracturas del terreno). Estos valores no pueden ser manejados como datos lineales sin más, pues tienen un componente de dirección que obliga a aplicar este tipo de estadística de datos para un correcto tratamiento de los mismos (Fisher, 1995:xv). Además existen unas diferencias básicas importantes entre los datos circulares y los lineales como son (Batschelet, 1981:3) : a) la dirección cero u origen de los datos circulares es arbitraria (dirección Norte, eje X...) y b) una suma o diferencia de ángulos que exceda del intervalo 0° - 360° (o los valores correspondientes en otras escalas de medidas angulares) deben ser reducidos a este intervalo.

En primer lugar realizaremos una revisión de los estadísticos básicos que se utilizan tanto en el análisis de datos circulares como lineales, junto con la indicación de las funciones de distribución de probabilidad más comunes en ambos tipos de datos, para finalizar con la aplicación de pruebas de control de la incertidumbre a los datos indicados en el apartado de material. Todo esto nos debe llevar a proponer una normativa de control de datos vectoriales, inexistente en la actualidad.

3.2.1 Estadísticos básicos

Aunque el estudio de los datos en esta Tesis es vectorial, se debe diferenciar, tal como se hace en la literatura estadística de datos circulares (Fisher, 1995; Mardia, 1971), entre **vectores**, que serían líneas orientadas con sentido y **ejes**, líneas orientadas pero sin un sentido definido. Así, las direcciones de vientos en una zona determinada vendrán representadas por vectores, mientras que las fracturas de un terreno, lo serán por ejes. Cabe distinguir también los términos **orientación**, aplicado generalmente a vectores y **tendencia**, referido a los ejes.

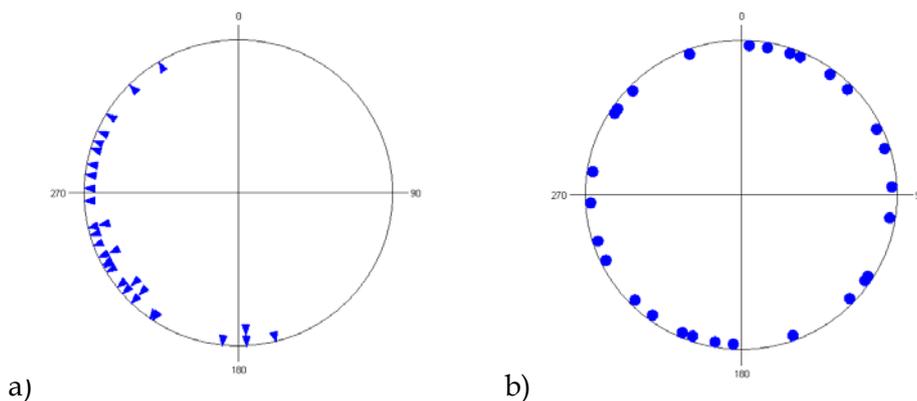


Fig. 3-6 Representación de datos circulares como vectores y ejes.

a) Representación de vectores mediante puntas de flechas en los nodos finales, b) representación de ejes mediante círculos sobre ambos nodos.

En cuanto a la representación gráfica de vectores y ejes, se optará en el primer caso por una línea terminada en flecha (nodo final) apuntando con un acimut determinado, mientras que en el segundo, lo será por una línea con un pequeño círculo en cada extremo (nodos, sin definir inicial y final). Ambas representaciones aparecen en la figura 3-6 sobre una circunferencia de radio unidad, teniendo en cuenta que en el caso de los ejes, cada par de círculos diametralmente situados se corresponde con un único eje.

En lo referente a esta Tesis, el vector de error bidimensional se calcula a partir de las diferencias de coordenadas entre la posición verdadera (X_v, Y_v) y la medida (X_m, Y_m) de los puntos considerados (obtenidos de la calibración del escáner fotogramétrico o de cartografía). Con estos incrementos de coordenadas se obtiene el **módulo** (d) y el **acimut** (θ) de cada vector, según las expresiones indicadas en la ecuación 3-3.

$$d = \sqrt{(X_v - X_m)^2 + (Y_v - Y_m)^2} \quad \theta = \arctg\left(\frac{X_v - X_m}{Y_v - Y_m}\right)$$

Ec. 3-3. Cálculo de módulo y acimut del vector de error.

En relación al cálculo del acimut hay que matizar lo siguiente:

- El acimut es un ángulo plano medido en sentido horario desde la dirección del eje Y en la posición verdadera hasta el vector formado por la posición verdadera y la medida.
- La indicación del sentido del vector está implícita en la afirmación anterior.
- El cálculo del acimut variará según el cuadrante en que se encuentre el punto medido en relación con el verdadero, por lo que la fórmula 3-3 deberá ser particularizada para cada caso (López Cuervo, 1993:174).
- Aún cuando los términos que mejor definen las coordenadas polares son el módulo y el acimut, en el texto que sigue se utilizará también el término de orientación por analogía a cómo se designan estos ángulos en la bibliografía de datos circulares (Fisher, 1995; Mardia, 1971).

- En esta misma bibliografía, la orientación medida crece en sentido contrario a las agujas del reloj a partir del punto (1,0) del círculo unidad. Como en la mayoría de los datos geoespaciales se utiliza una orientación respecto al Norte Geográfico o al Norte de cuadrícula (eje Y) de un sistema de referencia bidimensional ortogonal, se considerará θ creciendo en sentido horario a partir del eje Y, como ya se ha dicho, adaptando las fórmulas de tratamiento estadísticos de datos circulares a este criterio.

En el tratamiento estadístico de datos circulares, se trabajará inicialmente con vectores que tienen su origen en el centro de un **círculo de radio unidad**, pues así está desarrollada hasta la fecha la estadística circular; es decir, los estadísticos y ejemplos que a continuación se desarrollan consideran sólo la orientación del vector, no el módulo, que es la unidad. Posteriormente se analizará tanto la distribución de los módulos como la distribución conjunta de módulos y orientación de las poblaciones de vectores.

3.2.1.1 Estadísticos y parámetros básicos para vectores unitarios

Se citan a continuación los estadísticos y parámetros básicos necesarios para tratar y analizar una muestra de una población de mediciones circulares $(\theta_1, \theta_2, \dots, \theta_n)$. Las expresiones matemáticas de dichos estadísticos, así como la definición de otros y de los correspondientes parámetros de la población, se pueden consultar más exhaustivamente al final de este texto.

Para una mejor comprensión, se acompaña la definición de estos valores con la resolución de los siguientes ejemplos:

a) Tratamiento de 361 vectores de error, con módulos no unitarios expresados en micras, obtenidos del proceso de calibración del escáner fotogramétrico citado en el apartado 3.1.1. Estos datos pueden consultarse en el CD que acompaña a esta Tesis doctoral en la siguiente ubicación:

D://Tesis/Datos/Escáner/2004-10-19

b) Datos correspondientes a 30 mediciones de vectores unitarios expresados en unidades lineales, de estratificación cruzada de paleocorrientes medidos en el anticlinal de Belford (Nueva Gales del Sur) y tomados de los ejemplos de Fisher (1995:242).

La elección de estos ejemplos se debe, aparte del desigual número de observaciones, a la diferencia de distribución, tal como se observa en la figura 3-7, que permitirá contrastar resultados.

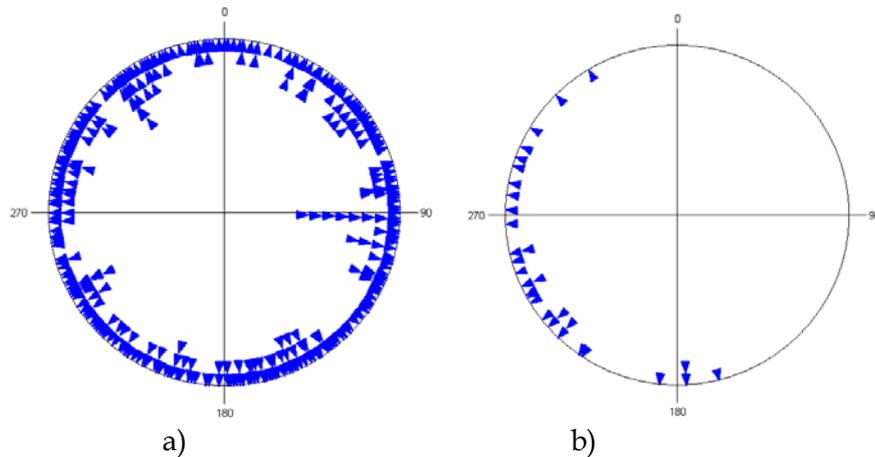


Fig. 3-7. Representación de datos brutos de dos muestras de datos.
a) correspondientes a datos de calibración del escáner fotogramétrico con muestra reducida, b) ejemplo de acimutes de estratificación cruzada en paleocorrientes (tomado de Fisher, 1995:242).

Los principales estadísticos y parámetros básicos para datos circulares son:

- **Estadísticos de tendencia central**

- **Acimut medio ($\bar{\theta}$)**, obtenido mediante la suma vectorial de todos los vectores de la muestra. Representa el acimut medio del vector resultante de la suma vectorial citada y se considera un estimador insesgado del parámetro representado por μ . Es importante indicar que no depende de la dirección cero de origen.

$$\hat{\mu} = \bar{\theta} \quad \text{Ec. 3-4. Acimut medio.}$$

- **Mediana de la muestra ($\tilde{\theta}$)**. En datos unidimensionales la mediana se corresponde con el valor central de una serie de elementos ordenados de forma creciente. Esta definición, por la naturaleza de los datos circulares, no es aplicable a este tipo de datos, dado que el origen de medida de ángulos es arbitrario. La forma de calcular la mediana es minimizando la función siguiente:

$$d(\theta) = \pi - \frac{1}{n} \sum_1^n |\pi - |\theta_i - \theta|| \quad \text{Ec. 3-5. Mediana de la muestra.}$$

Además para distribuciones no unimodales el valor de la mediana puede no ser único, en contra de lo que ocurre en datos unidimensionales.

En el caso de la **mediana de la población**, su representación es $\tilde{\mu}$.

- **Módulo del vector resultante (R)**, calculado a partir de las ecuaciones siguientes, donde C y S son la suma de todos los cosenos y senos de los acimutes de la muestra, respectivamente.

$$R = \sqrt{C^2 + S^2} \quad ; \quad C = \sum_{i=1}^n \cos \theta_i \quad ; \quad S = \sum_{i=1}^n \operatorname{sen} \theta_i$$

Ec. 3-6. Módulo del vector resultante y valores de C y S.

- **Módulo medio** (\bar{R}), obtenido al dividir R entre el número de observaciones. Lógicamente, si trabajamos con vectores unitarios su valor oscila entre 0 y 1. Si es 1 implica que todos los vectores son coincidentes en su dirección; en cambio, un módulo medio nulo puede ser el resultado de una distribución uniforme pero pueden darse situaciones, como la que se aprecia en la figura siguiente, donde el valor de \bar{R} es muy pequeño (en este caso 0.03) y la distribución dista bastante de ser uniforme:

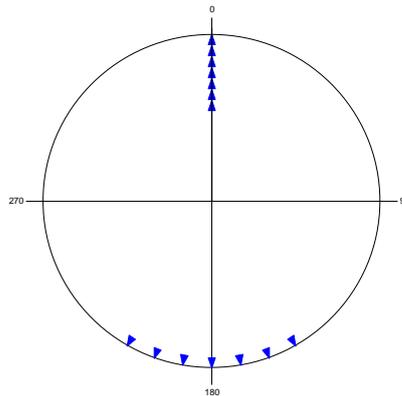


Fig. 3-8. Ejemplo de distribución de datos con módulo medio pequeño sin ser la distribución uniforme.

En esta representación de 14 datos brutos, el valor del módulo medio es muy pequeño (0.03) aunque la distribución dista de ser uniforme, tal como se aprecia en la configuración de estos vectores.

Si n es el número de observaciones se tiene que:

$$\bar{R} = \frac{R}{n}$$

Ec. 3-7. Módulo medio.

Como en el caso del acimut medio, el modulo medio muestral se considera un estimador insesgado del **parámetro poblacional**, ρ :

$$\hat{\rho} = \bar{R}$$

Ec. 3-8. Estimación de ρ .

- **Estadísticos de dispersión**

- **Varianza circular de la muestra** (V). En similitud con la varianza de datos lineales, es un indicador de dispersión angular de forma que cuanto más pequeño sea este valor, más concentrados están los datos. Su valor oscila entre 0 y 1, pero al igual que ocurría con el módulo medio, $V=1$, no implica una distribución uniforme. Su expresión es:

$$V = 1 - \bar{R}$$

Ec. 3-9. Varianza circular de la muestra.

La **varianza circular de la población** (ν), de forma análoga a los valores de la muestra, se calcula por:

$$\nu = 1 - \rho$$

Ec. 3-10. Varianza circular de la población.

- **Desviación estándar circular de la muestra** (ν): Es resultado de una raíz cuadrada por analogía a lo que ocurre en datos lineales aunque en datos circulares queda definida por:

$$\nu = [-2 \log(1 - V)]^{1/2}$$

Ec. 3-11. Desviación estándar de la muestra.

La desviación estándar circular de la población (σ) se obtiene según:

$$\sigma = [-2 \log(1 - \nu)]^{1/2}$$

Ec. 3-12. Desviación estándar circular de la población.

- **Dispersión circular de la muestra** (δ): En el caso de una distribución uniforme este valor es infinito, reduciendo hasta valores próximos a cero en distribuciones muy concentradas. Este estadístico se ha obtenido aplicando la fórmula siguiente:

$$\delta = (1 - \hat{\rho}_2) / (2\bar{R}^2)$$

Ec. 3-13. Dispersión circular.

Donde es necesario definir el valor de $\hat{\rho}_2$ y de los momentos senos y cosenos de segundo orden, tal como aparece en la ecuación siguiente (la expresión del momento seno es análoga a la del coseno).

$$\hat{\rho}_2 = (\bar{C}_2^2 + \bar{S}_2^2)^{1/2} \quad \bar{C}_2 = \frac{1}{n} \sum_{i=1}^n \cos 2\theta_i$$

Ec. 3-14. Valor de $\hat{\rho}_2$ y momento coseno.

La dispersión circular de la población (δ) se obtiene según:

$$\delta = (1 - \rho_2) / (2\rho^2)$$

Ec. 3-15. Dispersión circular de la población.

- **Parámetro de concentración** (κ): La concentración (κ) es un parámetro específico de la distribución von Mises de datos circulares. Mide la variación de la distribución en relación con un círculo perfecto, lo que sería una distribución uniforme. Cuando este parámetro κ tiende a cero, la distribución converge a la distribución uniforme; en caso de tender a infinito, la distribución se concentra en la dirección media del vector resultante. Podemos considerar este parámetro significativo para $\kappa \geq 2$.

Ejemplo 3-1

En la tabla siguiente se indican los resultados de los estadísticos anteriores en los dos ejemplos indicados: a) los datos de calibración de un escáner fotogramétrico y b) los acimutes de estratificación cruzada de paleocorrientes.

Estadístico	Resultado ejemplo a)	Resultado ejemplo b)
$\bar{\theta}$	141.86°	247.62°
$\bar{\theta}$	169.45°	245.00°
R	6.81	23.48
\bar{R}	0.02	0.78
V	0.98	0.22
ν	161.47°	40.09°
$\hat{\delta}$	1317.92	0.51
κ	0.04	2.67

Tabla 3-1. Ejemplo de aplicación de estadísticos de datos circulares.

Los resultados indican que los datos del ejemplo a) están mucho más dispersos que los ejemplo b). Baste comparar los valores bajos de \bar{R} (que conllevan valores altos de V) en el ejemplo a), en contraposición a lo que ocurre en b). La dispersión circular en a) es muy alta, 1317.9, que contrasta con el resultado de 0.5 en el segundo ejemplo. En cuanto al parámetro de concentración, κ , ya se indicó anteriormente que se considera significativo a partir de 2.

3.2.1.2 Estadísticos y parámetros básicos de los módulos de los vectores

Los módulos obtenidos (Ec. 3-3) constituyen muestras o poblaciones de n elementos de datos lineales que representan, en este caso, la magnitud del error de posición. Los estadísticos y parámetros básicos para resumir las características de los datos lineales son más conocidos y usuales que los anteriormente explicados para datos circulares por lo que no se plantea aquí una descripción como la antecedente sino un breve recordatorio.

- **Estadísticos de tendencia central**

- **Media aritmética**

Dado un conjunto de n observaciones, en este caso, módulos de vectores, $d_1, d_2, \dots, d_{n-1}, d_n$, la media aritmética, \bar{d} , se obtiene según la expresión siguiente.

$$\bar{d} = \frac{\sum_{i=1}^n d_i}{n}$$

Ec. 3-16. Media aritmética.

La interpretación de la media como centro de valores se apoya en una propiedad de la media aritmética que indica que la suma de las desviaciones de un conjunto de observaciones con respecto a la media es cero.

- **Estadísticos de dispersión**

- **Desviación típica**

La desviación típica de una muestra de elementos toma valores no negativos y mide, como sabemos, la dispersión de los datos: a mayor desviación típica, mayor dispersión de datos. Se calcula por la expresión siguiente, utilizando la nomenclatura ya indicada para los módulos:

$$s = \sqrt{\frac{\sum_{i=1}^n (d_i - \bar{d})^2}{n - 1}}$$

Ec. 3-17. Desviación típica para datos lineales

El cuadrado de este valor es la **varianza**.

También será necesario el cálculo del rango y los valores máximo y mínimo, así como las frecuencias, absolutas o relativas, para la elaboración de los histogramas. Puede ser necesario utilizar otros estadísticos para describir distribuciones no gaussianas; cuando esto ocurra se comentarán en el lugar adecuado.

Ejemplo 3-2

Los resultados para los estadísticos descriptivos en el ejemplo a) son:

Media	Desv. típica	Valor max.	Valor mín.
1.927	0.990	5.303	0.042

Tabla 3-2. Ejemplos de aplicación de estadísticos de datos lineales.

3.2.2 Funciones de distribución de probabilidad

Del mismo modo que ocurre con los datos lineales, existen **funciones de distribución** de datos circulares. En los dos apartados siguientes se indican las distribuciones más comunes tanto de datos circulares como lineales, siendo esta segunda revisión más somera, por ser las funciones de distribución de datos lineales más conocidas.

3.2.2.1 Funciones de distribución de probabilidad de datos circulares

A continuación se indican las distribuciones de probabilidad en datos circulares más comunes (Fisher, 1995; Jammalamadaka and SenGupta, 2001; Mardia and Jupp, 2000).

- **Distribución uniforme (U_c)**

En esta distribución todas las direcciones entre 0 y 2π radianes son igualmente probables. La función de densidad de probabilidad atiende a la expresión indicada en la ecuación 3-18.

$$f(\theta) = \frac{1}{2\pi} \quad 0 \leq \theta \leq 2\pi$$

Ec. 3-18. Función de densidad de probabilidad.

Mientras que la función de distribución aparece en la ecuación siguiente.

$$F(\theta) = \frac{\theta}{2\pi} \quad 0 \leq \theta \leq 2\pi$$

Ec. 3-19. Función de distribución.

El acimut medio de la población, μ , no está definido, el módulo medio poblacional, ρ , es igual a cero y la dispersión circular δ , es infinita. Un ejemplo de distribución uniforme de datos lo tenemos en la figura siguiente.

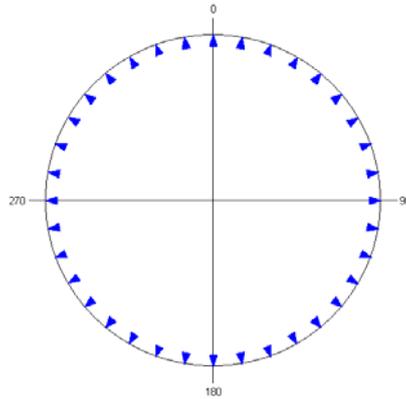


Fig. 3-9. Distribución uniforme de datos circulares. En la distribución uniforme, cualquier dirección es igualmente probable.

Esta distribución será frecuentemente la hipótesis nula sobre la cual se contrastarán otras alternativas de distribución de datos.

- **La distribución von Mises ($VM(\mu, \kappa)$)**

Esta distribución, simétrica y unimodal, es el modelo más habitual de distribución en muestras unimodales de datos circulares, siendo esta distribución en el círculo análoga a la distribución Normal en el caso de variables lineales. La distribución von Mises está definida por dos parámetros: uno que define el acimut medio, μ y otro que cuantifica la dispersión en torno a la localización, κ . La función de probabilidad de distribución aparece en la ecuación 3-20, aunque ya se indicó anteriormente (Ec. 2-2) cuando se propuso que los errores sistemáticos en datos bidimensionales se ajustan a una distribución de von Mises $VM(\mu, \kappa)$, donde se representa por $I_0(\kappa)$ la función modificada de Bessel de orden cero. La expresión genérica para diferentes órdenes aparece al final de este texto.

$$f(\theta) = [2\pi I_0(\kappa)]^{-1} \exp[\kappa \cos(\theta - \mu)]$$

$$0 \leq \theta \leq 2\pi \quad 0 \leq \kappa \leq \infty$$

Ec. 3-20. Función de densidad de probabilidad de VM.

La función de distribución se refleja en la ecuación siguiente:

$$F(\theta) = [2\pi I_0(\kappa)]^{-1} \int_0^\theta \exp[\kappa \cos(\phi - \mu)] d\phi$$

Ec. 3-21. Función de distribución de VM.

El acimut medio de la población, μ , sí está definido, en contra de lo que ocurriría en la distribución uniforme. Los valores de ρ y δ se indican en las ecuaciones siguientes sabiendo que la expresión de $A_1(\kappa)$ se obtiene como cociente entre funciones modificadas de Bessel de diferentes órdenes (Fisher, 1995:50):

$$\rho = A_1(\kappa) \quad ; \quad \delta = [\kappa A_1(\kappa)]^{-1} \quad \text{Ec. 3-22. Valores de } \rho \text{ y } \delta \text{ en VM.}$$

El parámetro de dispersión en la distribución normal sobre la recta, la desviación típica, es próxima a cero cuando la concentración de valores es máxima, mientras que en el caso circular si κ es cero, la distribución es uniforme. En la figura siguiente aparece representada la distribución de probabilidad de von Mises.

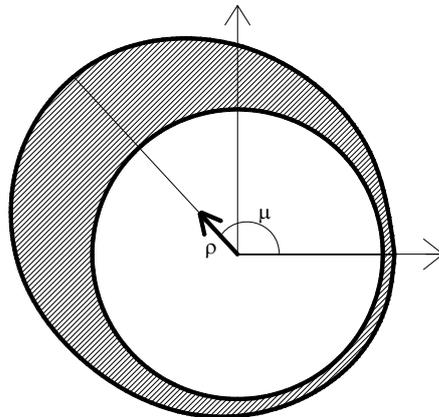


Fig. 3-10. Distribución von Mises.

En esta distribución el acimut medio de la población se representa por μ y el módulo medio poblacional por ρ (Batschelet, 1981:280).

• **Distribución Cardioide ($C(\mu, \rho)$)**

Es una distribución biparamétrica (μ, ρ) , unimodal y simétrica. La función de densidad de probabilidad, $f(\theta)$ aparece en la ecuación siguiente.

$$f(\theta) = \frac{1}{2\pi} \{1 + 2\rho \cos(\theta - \mu)\} \quad 0 \leq \theta \leq 2\pi, 0 \leq \rho \leq 1/2$$

Ec. 3-23. Función de densidad de probabilidad en distribución cardioide.

La función de distribución $F(\theta)$ se representa por:

$$F(\theta) = (\rho / \pi) \text{sen}(\theta - \mu) + \theta / (2\pi), \quad 0 \leq \theta \leq 2\pi$$

Ec. 3-24. Función de distribución en distribución cardioide.

El valor de ρ es igual o menor que 0.5 y la dispersión circular equivale a $1/(2\rho^2)$. Cuando ρ tiende a cero la distribución converge a la distribución uniforme Uc .

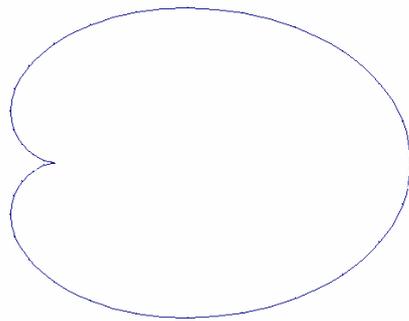


Fig. 3-11. Distribución de probabilidad de la distribución cardioide.

Esta distribución es biparamétrica (μ, ρ) unimodal y simétrica (Batschelet, 1981:283).

- **Distribución Envolvente-Cauchy $(WC(\mu, \rho))$**

Es una distribución modal simétrica obtenida al envolver la distribución de Cauchy lineal alrededor del círculo. La función de densidad de probabilidad se expresa por:

$$f(\theta) = \frac{1}{2\pi} \frac{1 - \rho^2}{1 + \rho^2 - 2\rho \cos(\theta - \mu)}$$

$$0 \leq \theta \leq 2\pi \quad 0 \leq \rho \leq 1$$

Ec. 3-25. Función de densidad de probabilidad de distribución WC.

Mientras que la función de distribución es:

$$F(\theta) = \frac{1}{2\pi} \cos^{-1} \left(\frac{(1 + \rho^2) \cos(\theta - \mu) - 2\rho}{1 + \rho^2 - 2\rho \cos(\theta - \mu)} \right) \quad 0 \leq \theta \leq 2\pi$$

Ec. 3-26. Función de distribución de WC.

Cuando ρ tiende a cero la distribución converge a la distribución uniforme; si tiende a la unidad, la distribución lo hará a la dirección de μ . Esta distribución tiene la propiedad indirecta de ser una alternativa a la distribución von Mises para datos simétricos unimodales. De hecho, para parámetros de dispersión determinados, esta distribución es similar a la envolvente-Normal y a von Mises.

- **Distribución Envolvente-Normal $(WN(\mu, \rho))$**

Es una distribución biparamétrica, unimodal y simétrica, obtenida al envolver la distribución normal o gaussiana en la recta alrededor del círculo. Los valores límites son idénticos a los de la distribución anterior.

Las distribuciones cardioide, envolvente-normal y von Mises son muy similares en apariencia, sobre todo las dos últimas son difíciles de distinguir en aplicaciones prácticas (Mardia, 1971:39). En la figura 3-12 aparece una comparativa de estas distribuciones (Batschelet, 1981:284).

Existen otros modelos para datos circulares, que son de menor aplicación, como otros modelos simétricos, modelos discretos o mezcla de distribuciones (Fisher, 1995:56-57; Jones and Pewsey, 2004; Pewsey, 2002).

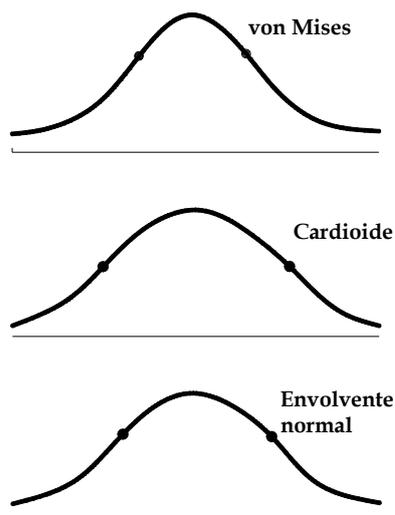


Fig. 3-12. Comparativa de distribuciones von Mises, cardioide y envolvente-normal.

Las distribuciones cardioide, envolvente-normal y von Mises son muy similares en apariencia y difíciles de distinguir en aplicaciones prácticas. Sobre las gráficas aparecen marcados los puntos de inflexión (Batschelet, 1981:284).

3.2.2.2 Funciones de distribución de probabilidad de datos lineales

Las funciones de distribución de probabilidad para datos lineales son más numerosas y conocidas que las de los datos circulares, por lo que entendemos que no tiene sentido una revisión profunda de las mismas, al igual que ocurría con los estadísticos y parámetros básicos. Sí se recuerdan, a continuación, las características generales de las distribuciones de datos lineales más comunes que se usarán a lo largo de este texto (Canavos, 1988). Estas distribuciones de datos para variables continuas corresponden, a excepción de la distribución normal, a distribuciones asimétricas positivas, tal como se verá posteriormente.

- **Distribución normal**

Esta distribución, también conocida como Gaussiana, es una de las de mayor uso dentro de las distribuciones continuas de probabilidad, ya que las distribuciones de muchas estadísticas muestrales tienden hacia la distribución normal conforme crece el tamaño de la muestra. Multitud de estudios son representados por esta distribución como datos meteorológicos, calificaciones obtenidas en un examen o distribución de errores aleatorios. La distribución normal de una variable x depende de dos parámetros: la media \bar{m} , que puede tomar cualquier valor y la desviación estándar s , que es mayor o igual que cero. En el apartado primero de esta Tesis ya se representó la curva de Gauss (Fig. 1-1) de media nula y desviación estándar s . La expresión para la función de densidad de probabilidad para la distribución normal aparece en la ecuación siguiente:

$$f(x, \bar{m}, s) = \frac{1}{\sqrt{2\pi}s} \exp\left[-\frac{1}{2}\left(\frac{x - \bar{m}}{s}\right)^2\right]$$

Ec. 3-27. Función de densidad de probabilidad de la distribución normal.

- **Distribución Gamma**

Otra distribución de gran uso es la gamma, cuya función de distribución de probabilidad está representada en la ecuación siguiente:

$$f(x, \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \exp(-x/\beta) \quad x > 0, \alpha, \beta > 0$$

Ec. 3-28. Función de distribución de probabilidad de la distribución gamma.

Donde α y β son los factores de forma y escala, respectivamente. Como vemos en la figura siguiente, según sean estos factores, se definen diferentes perfiles de la distribución gamma (Canavos, 1988:152). Especial interés para nuestro trabajo reviste la distribución correspondiente a los valores de $\alpha = 2$ y $\beta = 1$, pues corresponde a una distribución asimétrica positiva que se adaptará bastante bien, al igual que la distribución Weibull, a la distribución de los módulos de los vectores de error, como se verá posteriormente.

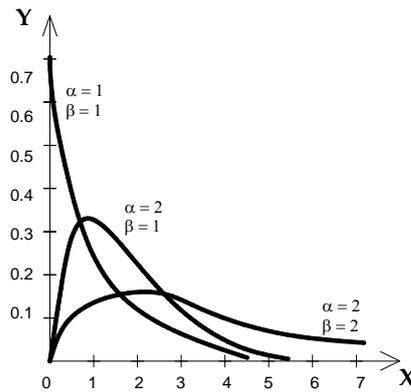


Fig. 3-13. Gráficas de la función de densidad gamma para distintos valores de α y β .

- **Distribución Weibull**

Una variable aleatoria x sigue una distribución Weibull cuando su función de densidad de probabilidad viene dada por:

$$f(x, \alpha, \beta) = \frac{\alpha}{\beta^\alpha} x^{\alpha-1} \exp[-(x/\beta)^\alpha] \quad x > 0, \alpha, \beta > 0$$

Ec. 3-29. Función de distribución de probabilidad de la distribución Weibull.

Donde, al igual que en la distribución anterior, esta función depende de dos parámetros, α , forma y β , escala. En la figura siguiente se representa la función de densidad de Weibull para distintos valores de α y β (Canavos, 1988:160).

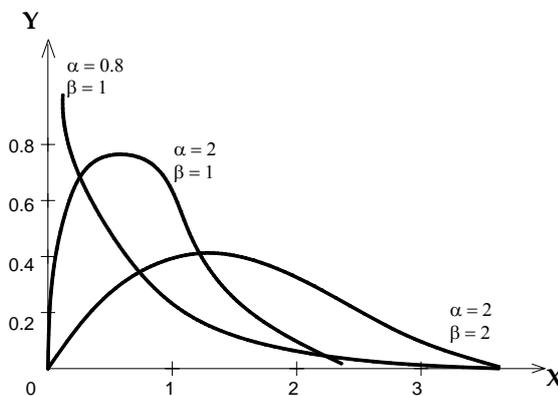


Fig. 3-14. Gráficas de la función de densidad gamma para distintos valores de α y β .

• **Distribución logarítmico-normal**

Esta distribución se suele utilizar como modelo en estudios socio-económicos para variables de ingreso o gasto. El logarítmico de una variable logarítmico-normal sigue una distribución normal. La función de distribución de la probabilidad de la función logarítmico-normal aparece en la siguiente ecuación considerando una media μ y una varianza σ^2 .

$$f(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma x}} \exp\left[-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right] \quad (\mu \in \mathfrak{R}, \sigma \in \mathfrak{R}^+)$$

Ec. 3-30. Función de distribución de la probabilidad de la función logarítmico-normal.

En la figura siguiente aparece la gráfica de esta distribución asimétrica.

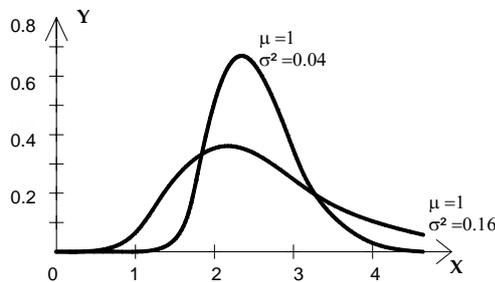


Fig. 3-15. Gráfica de la función de densidad de la distribución logarítmico-normal.

3.2.2.3 Funciones de distribución conjunta

Se han definido por separado las distribuciones circulares y lineales más comunes en las líneas precedentes; el siguiente paso debería ser presentar una distribución conjunta. Este tipo de distribuciones, como tales, no están tipificadas, pues hasta la fecha no se ha contemplado la casuística más compleja de combinar las diferentes distribuciones lineales con las circulares. Un primer paso puede ser calcular las posibles correlaciones existentes entre acimutes y módulos. En el gráfico siguiente se representan los datos correspondientes al ejemplo del escáner fotogramétrico donde aparecen los vectores de error con su módulo y acimut.

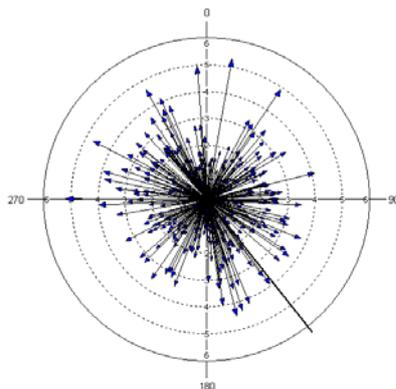


Fig. 3-16. Correlación entre acimutes y módulos de una muestra de datos circulares. Gráfico que muestra la distribución y posible correlación existente entre los acimutes y los módulos en el ejemplo a) de datos del escáner. La recta indica la dirección del acimut medio.

Así, es posible calcular el coeficiente de correlación circular-lineal que determina si existe la correlación entre una variable circular y otra lineal, que debe seguir una distribución normal. Este coeficiente, que no admite valores negativos, se sitúa entre 0 y 1, entendiéndose que el valor nulo indica ausencia de correlación entre ambas variables.

En cualquier caso, parece difícil de esta forma, ofrecer más información de la distribución de ambas variables a la vez, siendo necesario recurrir a **métodos no paramétricos**, sobre todo si la distribución de los datos lineales no es normal, como se verá más adelante. Proponemos, por tanto, transformar los vectores de error en una nube de puntos para encontrar la función de densidad de distribución de dicha nube. En la figura siguiente aparece la nube de puntos correspondiente a los nodos finales de los vectores de error del ejemplo a) del escáner fotogramétrico y superpuesto un mapa de densidad de puntos, calculado según el número de puntos por unidad de área.

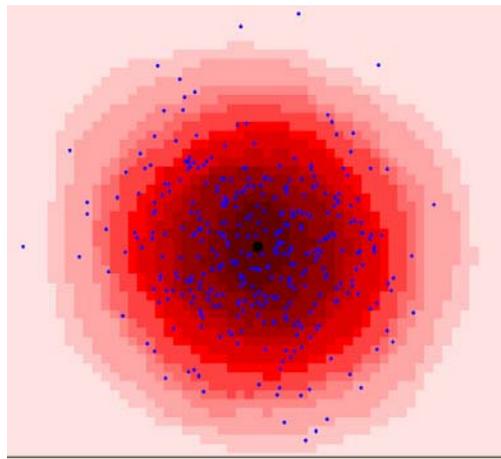


Fig. 3-17. Mapa de densidad para el ejemplo a) del escáner fotogramétrico.

Sobre la nube de puntos correspondiente a los nodos finales de los vectores de error, se ha creado un mapa de densidad calculado a partir del número de puntos por unidad de área.

Estos mapas de densidad nos deben proporcionar un modelo para la función de distribución conjunta, que nos permita realizar simulaciones, estudiar la transmisión de errores y detectar *outliers* con mayor flexibilidad.

3.2.3 Pruebas para la determinación de las características de las distribuciones

La aplicación de diversos tests estadísticos a las mediciones realizadas permitirá analizar la distribución y características de las observaciones e inferir su comportamiento. De igual modo que en apartados anteriores, se seguirá el esquema de analizar por separado los datos circulares (Fisher, 1995) los lineales y de forma conjunta.

3.2.3.1 Pruebas para la determinación de las características de las distribuciones de los datos circulares

Antes de aplicar cualquier test estadístico en busca de las características de la muestra de datos o de la población, es interesante realizar un análisis exploratorio consistente en el dibujo de los datos brutos. Dado que la explicación de las pruebas siguientes se acompaña de la aplicación práctica de las mismas a los dos ejemplos explicados en el apartado 3.2.1.1 (calibración de escáner fotogramétrico y acimutes de paleocorrientes),

se recomienda al lector la vista de la figura de representación de datos brutos (Fig. 3-7). Esta figura nos indica la tendencia de distribución de dichos datos, aunque deba ser completada y corroborada posteriormente de forma analítica.

a) Control de la uniformidad

Denominamos **isotropía** en la distribución de una muestra de valores a la no existencia de una dirección privilegiada, lo que es equivalente a la **aleatoriedad** o **uniformidad** en las direcciones: a cada dirección se le asigna la misma probabilidad que a cualquier otra. Existen diferentes tests para comprobar la uniformidad de una muestra de valores; algunos testean la existencia de cualquier clase de 'no-uniformidad' (test omnibus), pero tienen la desventaja de no ser muy eficientes cuando se está buscando una característica concreta, como por ejemplo la unimodalidad, por lo que sería deseable, en estos casos, trabajar con una prueba específica para ello. Se presentan a continuación una serie de tests para determinar si la muestra ha sido obtenida a partir de una población con distribución uniforme, contrastando con otras alternativas, como la unimodalidad, precedidas de un test gráfico para valorar la uniformidad.

Por la naturaleza de los datos examinados, estos se presentan como continuos (no agrupados) por lo que en los párrafos que siguen se aplican los tests en sus versiones de datos de estas características, remitiendo al lector a la bibliografía indicada para profundizar sobre el tratamiento estadístico de datos agrupados (Fisher, 1995; Mardia and Jupp, 2000).

- **Valoración gráfica de la uniformidad**

Para la valoración gráfica de la uniformidad se calcula en primer lugar los estadísticos de orden lineal $\theta_{(i)}$, al colocar en orden creciente todos los acimutes θ_i de la muestra, procediendo tal como se indica en la ecuación siguiente:

$$x_1 = \theta_{(1)} / 2\pi, \dots, x_n = \theta_{(n)} / 2\pi \quad \text{Ec. 3-31. Cálculo de } x_i.$$

Dibujando posteriormente los siguientes puntos en un sistema de coordenadas rectangulares:

$$\left[\frac{1}{(n+1)}, x_1 \right], \dots, \left[\frac{n}{(n+1)}, x_n \right] \quad \text{Ec. 3-32. Valoración gráfica de uniformidad.}$$

Si los puntos se encuentran dibujados siguiendo una línea aproximada de 45°, pasando por el origen, los datos se adaptan al modelo uniforme, tal como aparece en la figura 3-18 correspondiente a los datos del escáner fotogramétrico.

Otro aspecto a tratar es la simetría de las distribuciones. Las pruebas existentes al respecto son variadas (Pewsey, 2002) aunque la aplicación a los datos tratados no proporcionaron resultados positivos.

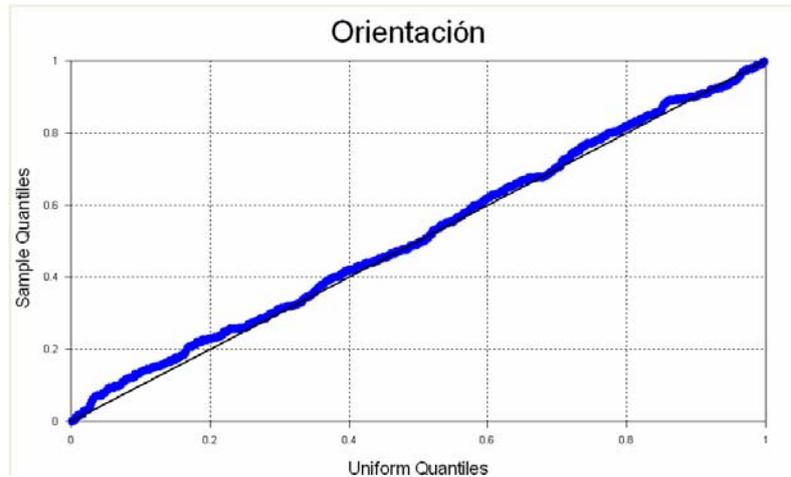


Fig. 3-18. Gráfico de valoración de uniformidad sobre los datos de la calibración del escáner.

El gráfico de valoración de uniformidad se ha aplicado a los datos procedentes de la calibración del escáner fotogramétrico. Los puntos dibujados según la ecuación 3-32 siguen una línea aproximada de 45° pasando por el origen, lo que es un indicio de distribución uniforme de dichos datos.

Al existir una elección arbitraria del origen, es posible que la distribución gráfica de los puntos presente, en algunos casos, anomalías que interfieran la interpretación. Para solucionar este problema, se recurre a sumar la unidad a cada elemento de la pareja de valores del primer 20% de puntos y restar la unidad, también de la misma forma, a las últimas 20% parejas de valores.

- **Test de Kuiper**

La hipótesis nula del test de Kuiper establece que la muestra de valores ha sido obtenida a partir de una distribución concreta, en este caso la uniforme. Se comprueba esta afirmación contra cualquier alternativa de no-uniformidad, lo cual puede llevar a conclusiones diferentes si se analizan los mismos datos con tests más específicos.

A partir de los datos x_1, \dots, x_n (Ec. 3-31), se calculan los estadísticos que aparecen en la ecuación siguiente.

$$D_n^+ = \text{máximo de } \frac{1}{n} - x_1, \frac{2}{n} - x_2, \dots, 1 - x_n$$

$$D_n^- = \text{máximo de } x_1, x_2 - \frac{1}{n}, x_3 - \frac{2}{n}, \dots, x_n - \frac{n-1}{n}$$

Ec. 3-33.
Estadísticos de Kuiper

Con los valores de V_n y V , que se indican a continuación, se acepta o rechaza la hipótesis planteada.

$$V_n = D_n^+ + D_n^-$$

$$V = V_n (n^{1/2} + 0.155 + 0.24/n^{1/2})$$

Ec. 3-34. Test de Kuiper.

En caso de ser V muy grande, según valores críticos, se rechaza la hipótesis nula de que la muestra de valores se ha obtenido a partir de una distribución uniforme.

- **Test de Rayleigh**

En el caso específico de querer detectar una única dirección modal de una muestra de vectores contra la alternativa de aleatoriedad se pueden plantear los casos de tener una dirección media especificada o no especificada.

Supongamos de entrada que no conocemos la dirección media especificada, que será lo más común. La hipótesis de uniformidad se rechaza si la longitud media del vector resultante de la muestra es muy grande. Este test se denomina de **Rayleigh** y el estadístico a calcular es:

$$Z = n\bar{R}^2 \quad \text{Ec. 3-35. Test de Rayleigh.}$$

Siendo la probabilidad P de aceptar la hipótesis de uniformidad, en caso de ser $n \geq 50$:

$$P = \exp(-Z) \quad \text{Ec. 3-36. Probabilidad de Rayleigh.}$$

Este test supone que un valor del módulo medio más largo implica una concentración mayor en torno a la media y por tanto menos probabilidad de que los datos estén uniformemente distribuidos.

En el caso de considerar una dirección media especificada, para probar la alternativa de unimodalidad, esta prueba de Rayleigh recibe el nombre de **Test V** y el estadístico a calcular es:

$$\bar{R}_0 = \bar{R} \cos(\bar{\theta} - \mu_0) \quad \text{Ec. 3-37. Test V.}$$

Donde μ_0 es el valor especificado para la dirección media. La hipótesis nula de uniformidad se rechaza si el valor del estadístico es muy grande.

- **Test de Rao de espaciado de datos**

Esta prueba considera también como hipótesis nula que los datos están distribuidos uniformemente, con la salvedad de buscar si el espaciado entre puntos adyacentes es aproximadamente igual en todo el círculo. Para una distribución uniforme el espaciado entre puntos debería ser $360^\circ/n$. Si el espaciado existente se desvía mucho de este valor, la probabilidad de que los datos pertenezcan a una distribución uniforme se reduce. Se necesita para ello calcular el valor de la ecuación siguiente.

$$L = \frac{1}{2} \sum_{i=1}^n \left| T_i - \frac{2\pi}{n} \right| \quad \text{Ec. 3-38. Test de Rao.}$$

El valor de T_i aparece en la ecuación 3-39:

$$T_i = \theta_{(i)} - \theta_{(i-1)} \quad i = 1, \dots, n-1, \quad T_n = 2\pi - (\theta_{(n)} - \theta_{(1)}) \quad \text{Ec. 3-39. Estadístico de Rao}$$

Este test se presenta como más robusto que el de Rayleigh, pudiendo ser un buen complemento en algunos casos. Supongamos una muestra bimodal, como la que aparece en la figura siguiente, en la que la longitud del vector resultante es muy pequeña, lo que supone, según Rayleigh, que la distribución es uniforme. Evidentemente, y según el espaciado entre puntos, deberá ser rechazada esta hipótesis.

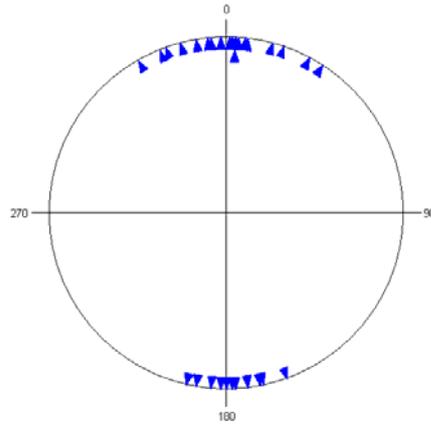


Fig. 3-19. Distribución bimodal de datos circulares. Una muestra bimodal, en la que la longitud del vector resultante es muy pequeña, supondría, según Rayleigh, una distribución uniforme. Según el espaciado entre puntos, en este gráfico deberá rechazarse esta hipótesis.

- **Test de Watson**

El **test de Watson** determina la bondad del ajuste del conjunto de datos a una distribución concreta, como puede ser la uniforme o la von Mises, por ejemplo. Para ello calcula las desviaciones entre los datos y la distribución elegida; si éstas desviaciones son muy grandes, el estadístico a calcular (U^2) también lo será y se rechazará la hipótesis nula, que supone que los datos se adaptan bien a la distribución elegida. Este test se desarrollará cuando se realicen las pruebas de bondad de ajuste para la distribución von Mises.

Ejemplo 3-3

Los resultados de las pruebas de uniformidad explicadas en las líneas precedentes a nuestros dos ejemplos son:

Test	Resultado ejemplo a)	Resultado ejemplo b)
Kuiper	V=1.10	V=3.20
	$\alpha > 0.15$	$\alpha < 0.01$
Rayleigh	Z=0.13	Z=18.38
	P=0.88	P=0.00
V	u=0.51	u=6.06
	P=0.31	P=0.00
Rao	L=123.50	L=205.00
	$0.99 > P > 0.95$	P < 0.01
Watson	$U^2 = 0.06$	$U^2 = 1.02$
	P > 0.50	P < 0.005

Tabla 3-3. Ejemplos de aplicación de test de uniformidad.

Según los datos anteriores podemos hablar de no uniformidad en el segundo apartado y de evidencia de la misma en el primero.

b) Control de la anisotropía

Como la segunda muestra anterior ha dado evidencias de tener una o varias direcciones dominantes, es necesario ahondar más sobre la cuestión, aplicando estadísticos para determinar si existe una dirección preferente en dichos datos, así como la posible existencia de simetría en torno a esta dirección preferente.

Los test que se indican a continuación, versan sobre la determinación de un intervalo de confianza para la mediana y el acimut medio de la población de acimutes a partir de sus correspondientes estimaciones de la muestra. Aunque se aplica un test sobre la mediana, se preferirá el uso del acimut medio sobre aquella por dos motivos: el primero, por la facilidad de cálculo, sobre todo en muestras grandes, como es nuestro caso, y el segundo, porque la media se completa con medidas de dispersión que resumen más convenientemente la distribución de los datos.

- **Estimación de la mediana**

Una vez calculada la mediana de la muestra se puede inferir una región de confianza para la mediana de la población. Para datos no muy dispersos y con $n \geq 16$ (Fisher 1995:73) se calcula m , que se corresponde con el número entero de valores por encima y por debajo de la dirección mediana, que conforman el intervalo de confianza α que depende de m , obteniendo $z_{1/2\alpha}$ de la tabla de valores de distribución normal de datos unidimensionales.

$$m = 1 + INT\left(\frac{1}{2} n^{\frac{1}{2}} z_{\frac{1}{2}\alpha}\right)$$

Ec. 3-40. Estimación de la mediana.

- **Estimación del acimut medio**

Para muestras con $n \geq 25$ (Fisher, 1995:76) se procede a calcular la dispersión circular y el error estándar circular, como se indica en la ecuación siguiente.

$$\hat{\sigma}^2 = \hat{\delta} / n$$

Ec. 3-41. Error estándar circular

El intervalo de confianza para μ viene dado por la ecuación siguiente.

$$\left(\hat{\mu} - \text{sen}^{-1}\left(z_{\frac{1}{2}\alpha} \hat{\sigma}\right), \hat{\mu} + \text{sen}^{-1}\left(z_{\frac{1}{2}\alpha} \hat{\sigma}\right) \right)$$

Ec. 3-42. Intervalo de confianza para acimut medio.

Ejemplo 3-4

Este ejemplo se corresponde con los resultados del control de la anisotropía para los datos del escáner fotogramétrico y los acimutes de las paleocorrientes.

(Con $\alpha=0.05$)	Resultado ejemplo a)	Resultado ejemplo b)
Mediana	169.45°	245.00°
Intervalo mediana	(162.51° , 200.20°)	(299.00° , 272.00°)
Acimut medio	141.86°	247.62°
Intervalo acimut m.	No se puede calcular	(232.71° , 262.52°)

Tabla 3-4. Intervalos de confianza para la mediana y el acimut medio.

No es posible calcular el acimut medio para el ejemplo a) dada la dispersión de valores. Cuando las distribuciones no presentan una dirección preferente clara y los datos tienden a tener una distribución uniforme, no es posible calcular el intervalo de confianza para el acimut medio, pues el elevado valor de la dispersión circular hace que no tenga solución matemática el arco seno de un valor superior a uno, tal como ocurre con los datos del ejemplo del escáner fotogramétrico.

c) Ajuste a distribuciones específicas: la distribución von Mises

Como se indicó en el apartado de modelos de distribución de probabilidad, una de las distribuciones más características dentro de los datos circulares, es la de von Mises. Esta distribución, que es simétrica y unimodal, es el modelo más habitual de distribución en muestras unimodales de datos circulares, siendo análoga a la distribución Normal sobre la recta real. Este es el motivo por el que se presentan a continuación una serie de tests estadísticos para probar si los datos se ajustan a este modelo.

• Método gráfico para la bondad de ajuste del modelo de von Mises

Para la realización práctica de este método, es necesario calcular las estimaciones del acimut medio de la población y del parámetro de concentración (κ), para calcular los valores de z_i , según la ecuación siguiente.

$$z_i = \text{sen} \frac{1}{2} (\theta_i - \hat{\mu}) \quad i = 1, \dots, n$$

Ec. 3-43. Cálculo de z_i para la valoración gráfica de von Mises.

Se ordenan de forma ascendente los valores anteriores obteniendo $z_{(1)} \leq \dots \leq z_{(n)}$, y se dibujan los siguientes puntos:

$$\left(\text{sen} \left(\frac{1}{2} q_i \right), z_{(1)} \right), \dots, \left(\text{sen} \left(\frac{1}{2} q_n \right), z_{(n)} \right)$$

Ec. 3-44. Cálculo de puntos para la valoración gráfica de von Mises.

Siendo q_1, \dots, q_n los cuantiles de la distribución von Mises.

Si los datos proceden de una distribución de von Mises los puntos deberían estar situados a lo largo de una línea de unos 45° de inclinación pasando por el punto (0,0), tal

como se observa en la figura siguiente, correspondiente a los datos procedentes del cálculo del acimut de la estratificación cruzada en las paleocorrientes.

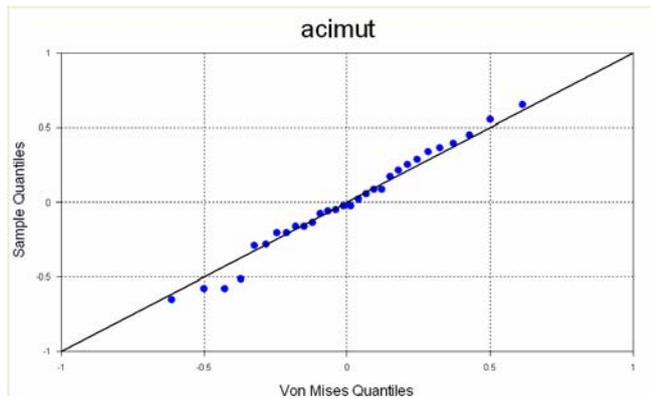


Fig. 3-20. Ajuste gráfico de los datos de acimutes de paleocorrientes a la distribución de von Mises.

A pesar de que los resultados de los estadísticos sugieren un buen ajuste a la distribución von Mises, el ajuste gráfico de los datos no apunta lo mismo.

- **Test de Watson**

El **test de Watson** determina la bondad del ajuste del conjunto de datos a una distribución concreta, en este caso la von Mises. Tal como se dijo en el apartado de tests para comprobar la uniformidad, se calculan las desviaciones entre los datos y la distribución elegida; si éstas desviaciones son muy grandes, el estadístico a calcular (U^2) también lo será y se rechazará la hipótesis nula, que supone que los datos se adaptan bien a dicha distribución.

En caso de no conocer el acimut medio de la población μ ni el parámetro de concentración, κ , se procede calculando las estimaciones de estos parámetros y aplicando la ecuación 3-48, a partir de los valores de frecuencia acumulativa (F_{κ}).

$$z_i = F_{\kappa}(\theta_i - \hat{\mu}), \quad i = 1, \dots, n$$

Ec. 3-45. Cálculo previo para test de Watson.

Se ordenan los valores anteriores en orden creciente y se determina el estadístico de la ecuación siguiente.

$$U^2 = \sum_{i=1}^n [z_{(i)} - (2i-1)/(2n)]^2 - n \left(\bar{z} - \frac{1}{2} \right)^2 + 1/(12n)$$

Ec. 3-46. Estadístico para el test de Watson.

La hipótesis de que la muestra ha sido obtenida a partir de una distribución von Mises se rechaza si el valor del estadístico anterior es muy grande.

En el caso de conocer uno o los dos parámetros anteriores (μ y κ) el proceso sufre ligeras modificaciones (Fisher, 1995:84).

- **Test de Kuiper**

También existe una modalidad del test de Kuiper, explicado en el apartado de uniformidad, para comprobar si la muestra de valores ha sido obtenida a partir de una distribución von Mises.

- **Intervalo de confianza para el acimut medio en la distribución von Mises**

De igual forma que se calculó un intervalo de confianza para la dirección media en el apartado de anisotropía, existe un procedimiento similar para los datos que se adaptan al modelo von Mises. El intervalo de confianza para la estimación del acimut medio viene dado por la ecuación siguiente:

$$\hat{\mu} \pm \text{seno}^{-1} \left(z_{\frac{1-\alpha}{2}} \hat{\sigma}_{VM} \right) \quad \text{Ec. 3-47. Intervalo para el acimut medio en von Mises.}$$

Donde el error estándar para la distribución de von Mises a partir de la estimación del parámetro de concentración, es:

$$\hat{\sigma}_{VM} = 1 / (n \bar{R} \hat{\kappa})^{\frac{1}{2}} \quad \text{Ec. 3-48. Error estándar para distribución von Mises.}$$

De forma similar se puede calcular un intervalo de confianza para el parámetro de concentración o aplicar tests para un acimut medio especificado o un parámetro de concentración concreto. Como en casos anteriores, dado que estos test son de interés menor a nuestros propósitos y siguen un desarrollo similar a otros ya explicados, nos remitimos a la bibliografía ya indicada.

Ejemplo 3-5

Resultados de las pruebas de ajuste a la distribución von Mises.

Test	Resultado ejemplo a)	Resultado ejemplo b)
Kuiper	V= 0.99	V=0.78
	P>0.15	P>0.15
Watson	U ² =0.05	U ² =0.024
	0.25>P>0.15	P>0.5
Acimut medio	141.86°	247.62°
Intervalo acimut medio (α=0.05)	No se puede calcular	(233.33° , 261.94°)

Tabla 3-5. Ejemplos de aplicación de tests de von Mises.

Del mismo modo que ocurría en el ejemplo 3-4, no es posible calcular el intervalo para el acimut medio en el ejemplo a).

d) Detección de outliers

Entendemos por *outlier* a una observación tan alejada del conjunto de datos principal que podemos considerarla como no perteneciente a la población. A continuación se indica un test para considerar si un punto puede considerarse como *outlier* u observación discordante (Fisher, 1995:85).

Sean R y R_k, el módulo del vector resultante, con y sin el punto sospechoso de ser un *outlier*, respectivamente. Calculamos el valor de Mn según la ecuación 3-49 y si este valor es muy grande la observación correspondiente se clasifica como discordante, según la tabla de valores críticos (Fisher, 1995:231).

$$Mn = \frac{R_k - R + 1}{n - R}$$

Ec. 3-49. Test para
detección de *outliers*.

Sin embargo, este test sólo está tabulado hasta $n=20$ por lo que no es adecuado para nuestro trabajo, que implica cientos de datos. Por este motivo y como se verá más adelante, se ha diseñado un método de detección de *outliers* que no presenta estos problemas.

3.2.3.2 Pruebas para la determinación de las características de las distribuciones de los datos lineales

Al igual que en datos circulares partíamos de la representación gráfica de los datos, para posteriormente validar las hipótesis con pruebas estadísticas, en los datos lineales complementaremos el análisis de los mismos con el histograma de valores, tal como aparece en la figura siguiente, donde se representa el histograma de los valores de los módulos en el ejemplo de los datos de calibración del escáner fotogramétrico. Por supuesto, aunque el histograma nos indica una distribución de valores, será necesario validar las afirmaciones iniciales con pruebas de bondad de ajuste a distribuciones concretas.

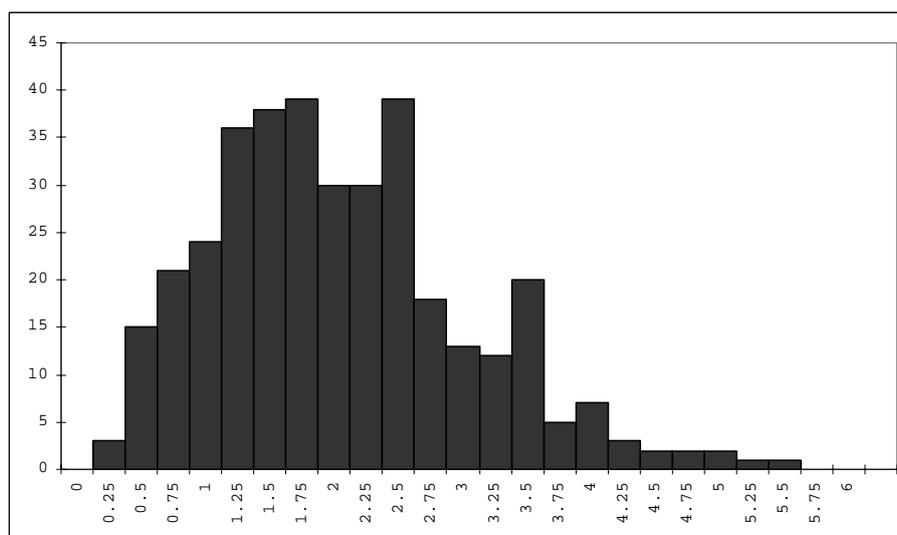


Fig. 3-21. Histograma de frecuencia de datos del ejemplo a).

En el histograma de frecuencias de los datos observamos la distribución de los módulos de los vectores de error que, de entrada, sugiere una distribución asimétrica positiva.

Así, las pruebas para la determinación de las características de las distribuciones en datos lineales se basarán en los siguientes apartados:

a) Cálculo de los estadísticos básicos

Estos estadísticos (media aritmética, desviación típica, valores máximos y mínimos) proporcionan una idea de la dispersión de los valores.

b) Detección de outliers

La metodología para la determinación de observaciones discordantes es mayor en datos lineales que circulares (Ghilani and Wolf, 2006:409-414). Consideraremos *outliers* u observaciones discordantes aquellas que sobrepasen un determinado valor, establecido inicialmente en la media aritmética más tres veces la desviación típica.

c) Pruebas de bondad de ajuste

Una prueba de bondad de ajuste se emplea, como sabemos, para decidir si un conjunto de datos se adapta a una distribución de probabilidad dada. Una de las pruebas de bondad de ajuste más comunes es la de **Kolmogorov-Smirnov** (en adelante K-S) que presenta la ventaja de ser aplicable a muestras pequeñas, aunque no es este nuestro caso, y de ser, en general, superior a otras pruebas similares para datos de variable aleatoria continua como es la prueba de bondad de ajuste de chi-cuadrado. La prueba de K-S compara la función de distribución acumulada de una variable con una distribución teórica determinada como puede ser la normal o la gamma. La Z de K-S se obtiene a partir de la mayor diferencia en valor absoluto entre las funciones de distribuciones acumuladas observadas y teóricas.

d) Histogramas de frecuencias de valores

Tal como se ha indicado al principio de este apartado.

Ejemplo 3-6

Los estadísticos descriptivos para los datos del ejemplo a) del escáner fotogramétrico ya se indicaron en el Ejemplo 3-2. A partir del valor de la desviación típica calculamos el número de observaciones que se consideran discordantes, obteniendo un valor de 3.

Se indican en la tabla siguiente los resultados de la aplicación del test de K-S para comprobar la bondad de ajuste de los datos del ejemplo a) del escáner fotogramétrico a diferentes distribuciones.

Distribución	Normal	Uniforme
Z de K-S	1.176	0.126
p-valor	6.005	0.000

Tabla 3-6. Ejemplo de aplicación de test de Kolmogorov-Smirnov.

3.2.3.3 Pruebas para la determinación de las características de las distribuciones conjuntas de datos

Las pruebas a aplicar para la determinación de las características de una distribución de datos, considerando tanto el módulo como el acimut del vector de error son:

a) Cálculo de las correlaciones entre módulos y acimutes

Se estudiarán las posibles correlaciones entre los módulos y acimutes de una muestra para comprobar si a direcciones preferentes de ángulos le corresponden valores significativamente iguales de módulos, lo que se interpretará como un indicador de la aniso-

tropía de la distribución. Acompañaremos, aquí también, al resultado del test de correlación con el gráfico de distribución correspondiente.

b) Diferencias entre valores en la desviación típica en X e Y y distribuciones de las mismas

Se estudiará la distribución seguida por los valores de X e Y por separado, buscando una normalidad en la distribución. Sabemos que si ambas variables son independientes y siguen una distribución normal, la componente cuadrática, esto es, el módulo, también seguirá esta distribución. Posteriormente se calcularán las diferencias entre los valores de la desviación típica de ambas variables.

c) Determinación de outliers

Los nodos finales de los vectores de error forman una nube de puntos que sigue una distribución que deberemos analizar. En esta nube de puntos estableceremos un porcentaje de elementos que consideraremos como *outliers* y que inicialmente se cifrará en un 2% de los puntos.

El proceso utilizado (Dixon and Chapman, 1980) determina el valor de la media armónica centrando el cálculo en cada punto y eliminando aquellos que presenten un mayor valor. La media armónica (Hm) es el cociente entre el número de observaciones (n) y la suma de las inversas de la distancias entre un punto i y el resto de localizaciones x , tal como se indica en la ecuación siguiente, de tal forma que los puntos más alejados de la nube de puntos tendrán un valor de Hm mayor y se considerarán como *outliers*.

$$Hm = \frac{n}{\sum_{i=1}^n \frac{1}{r_{ix}}}$$

Ec. 3-50. Media armónica.

Una vez eliminados estos puntos podemos repetir el proceso con otros porcentajes, según las características de la distribución.

Este método puede servir para la detección de errores groseros sin asumir requisitos propios de las distribuciones estándar contando, además, con la ventaja de que cada mapa de densidad y gráfico de distribuciones se adapta al caso específico que tratemos en cada momento.

d) Mapas de densidad

A partir de la nube de puntos indicada anteriormente se calcularán los mapas de densidad como el número de puntos que se corresponden por unidad de área. Estos mapas de densidad se configuran a modo de 'curvas de nivel', cuya forma será también indicativa de la distribución de los puntos.

4 Cálculos y resultados

Definidos el material y los métodos de trabajo, corresponde a este apartado desarrollar los cálculos y explicar los resultados de los mismos. Todo este proceso debe permitirnos elaborar una norma de control de datos espaciales en dos dimensiones, como se verá en el punto siguiente.

Procedemos a exponer este apartado analizando por separado los datos procedentes de la calibración del escáner fotogramétrico y los obtenidos por comparación de cartografías a diferentes escalas.

4.1 Cálculos y resultados del tratamiento de datos de la calibración de un escáner fotogramétrico

Recordemos que el material del que partíamos eran los vectores de error proporcionados por una serie de calibraciones geométricas de dos escáneres fotogramétricos de igual marca y modelo.

4.1.1 Origen de los datos

La calibración geométrica de un escáner fotogramétrico se realiza por escaneo de una placa de vidrio o lámina de acetato dimensionalmente estables que tienen marcadas unas señales cuya posición se conoce. Las diferencias entre las posiciones conocidas y las escaneadas permiten calcular una transformación que corrige desviaciones globales y que genera unas desviaciones o residuos no corregidos entre ambas posiciones. Cada desviación se traduce en unos incrementos de la coordenada X e Y entre la posición correcta teórica y la final real afectada de error. Estos incrementos sirven para calcular los vectores de error con módulo (m) y acimut (θ). En total se han realizado 20 calibraciones de 2116 vectores más 2 de 361 vectores.

4.1.2 Características de los datos

Cada fichero de calibración se nombra por la fecha en que se realizó, es decir, el fichero 2005-06-29 contiene los datos correspondientes al tratamiento de los vectores de error obtenidos en la calibración geométrica realizada el 29 de junio del 2005.

Observará el lector que algunos ficheros terminan con la letra 'e', la cual indica la existencia de un error, artificialmente introducido, en la calibración del escáner. Para conseguir este error se taparon uno o varios puntos de la placa de calibración para que el escáner, en su intento de ajustar la posición conocida con la escaneada, generara vectores de error mayores de los reales. De las 22 calibraciones realizadas, 9 tienen errores artificialmente provocados. En algunos casos sólo se tapó una marca y en otros varias, distribuidas bien en filas o en columnas, o agrupadas en esquinas, tal como se explica

en la tabla 4-1. En la figura siguiente se aprecian varios vectores de error con un módulo mucho mayor que el resto. El objetivo de este proceder es comprobar si los tests generales son capaces de detectar errores locales.

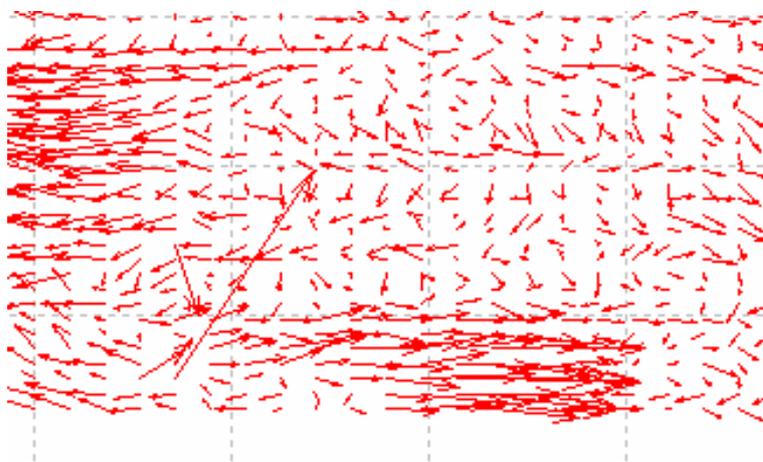


Fig. 4-1. Parte del resultado gráfico de la calibración. En esta parte del resultado gráfico de la calibración se aprecian los vectores de error de mayor tamaño originados artificialmente.

En la tabla siguiente se indican los ficheros afectados de error y la forma en que este error fue introducido.

Fichero	Fuente de error
2005-04-21e	Tapadas 2 cruces centrales.
2005-04-22e	Tapada 1 marca en la esquina inferior izquierda.
2005-05-03e	Tapadas 11 marcas concentradas en la parte superior del cuadrante superior derecho.
2005-05-18e	Tapadas grupos de 3 cruces.
2005-05-23e	Tapadas cruces en vertical cada 4 ó 5 líneas en horizontal.
2005-07-05e	Tapadas dos marcas centrales.
2005-07-11e	Tapada esquina inferior izquierda.
2005-07-12e	Tapada esquina inferior derecha.
2005-07-13e	Tapada esquina inferior izquierda.

Tabla 4-1. Ficheros de calibración geométrica que presentan error introducido artificialmente.

Otros dos ficheros de calibración terminan con la letra 'c' y se corresponden a las calibraciones proporcionadas por la Universidad de Cantabria (UC). El resto procede del escáner fotogramétrico propiedad de la UEX. Los datos de la UC servirán para comparar lo obtenido con los datos de la UEX.

Además del análisis estadístico general de los ficheros de calibración se realizaron análisis mediante ventanas móviles para comprobar si los estadísticos locales daban más información o diferentes resultados que los globales. El método fue superponer un cuadrado o ventana en el origen del fichero de calibración y, moviendo la ventana a intervalos fijos, ir extrayendo los datos cubiertos por ella para ser analizados independientemente.

Para esta prueba se eligieron 6 archivos: 5 de la UEX, tanto con error artificial como sin él, y uno de la Universidad de Cantabria. Se tomaron muestras de 10x10 vectores con solapes tanto en horizontal como en vertical de, al menos, el 50% (en algunos casos superior para mantener el tamaño de ventana). En uno de los ficheros (2005-02-21), se

realizaron además ventanas de otros tamaños: 10x10, 15x15, 20x20, 25x25, 30x30 y 35x35, también con el solape mínimo del 50% indicado.

La forma de numerar las submuestras, teniendo en cuenta que un fichero de 2116 elementos es el resultado de multiplicar 46x46 vectores de error colocados en filas y columnas creando una malla, es: (fila origen-fila final , columna origen-columna final). Por tanto, la primera ventana de tamaño 10x10 será nombrada como (1-10,1-10).

En la tabla 4-2 se resumen los ficheros tratados de esta forma, el tamaño de ventana elegido y el número de ventanas o submuestras que resultan de aplicar el criterio de solape anterior.

Ficheros	Tamaños de ventana (n° de ventanas)
2005-02-21	10x10 (81 ventanas)/ 15x15 (36 ventanas) / 20x20 (16 ventanas) 25x25 (9 ventanas)/ 30x30 (4 ventanas)/35x35 (4 ventanas)
2005-04-18	10x10 (81 ventanas)
2005-04-21e	10x10 (81 ventanas)
2005-04-25	10x10 (81 ventanas)
2005-05-03e	10x10 (81 ventanas)
2006-06-28c	10x10 (81 ventanas)

Tabla 4-2. Tamaño de las ventanas y número de ventanas analizadas en los ficheros de calibración.

En cuanto a las variables utilizadas y sus unidades, con dirección y sentido de lectura, se indican en la tabla 4-3.

Variable	Unidad	Dirección y sentido de lectura
Acimut	Grados sexagesimales	Ángulo plano medido en sentido horario desde la dirección del eje Y.
Módulo	μm	Cualquier dirección y sentido.
Coordenada X	μm	Dirección horizontal del escáner (SCAN).
Coordenada Y	μm	Dirección vertical del escáner (CCD).

Tabla 4-3. Variables y unidades utilizadas en los análisis a realizar a los datos procedentes de las calibraciones del escáner fotogramétrico.

En la tabla siguiente se resume todo lo dicho hasta ahora sobre las características de los datos procedentes de las calibraciones geométricas de un escáner fotogramétrico.

Calibración	n	Origen	Características
2004-10-19	361	UEX	
2005-02-21	2116	UEX	Tamaños de ventana: 10x10/15x15/20x20/25x25/30x30/35x35
2005-03-01	2116	UEX	
2005-03-07	2116	UEX	
2005-03-14	2116	UEX	
2005-04-04	2116	UEX	
2005-04-11	2116	UEX	
2005-04-18	2116	UEX	Tamaño de ventana: 10x10
2005-04-21e	2116	UEX	Tapadas 2 cruces centrales. Tamaño de ventana: 10x10
2005-04-22e	2116	UEX	Tapada 1 marca en la esquina inferior izquierda.
2005-04-25	2116	UEX	Tamaño de ventana: 10x10
2005-05-03e	2116	UEX	Tapadas 11 marcas concentradas en la parte superior del cuadrante superior derecho. Tamaño de ventana: 10x10
2005-05-18e	2116	UEX	Tapadas grupos de 3 cruces.
2005-05-23e	2116	UEX	Tapadas cruces en vertical cada 4 ó 5 líneas en horizontal.
2005-06-09	2116	UEX	
2005-06-29	2116	UEX	
2005-07-05e	2116	UEX	Tapadas dos marcas centrales.
2005-07-11e	2116	UEX	Tapada esquina inferior izquierda.
2005-07-12e	2116	UEX	Tapada esquina inferior derecha.
2005-07-13e	2116	UEX	Tapada esquina inferior izquierda.
2006-05-19c	361	UC	
2006-06-28c	2116	UC	Tamaño de ventana: 10x10

Tabla 4-4. Resumen de las características de los datos de las diferentes calibraciones del escáner fotográfico.

4.1.3 Descripción de los análisis realizados

Con el conjunto de datos explicados en el punto precedente, se procedió a calcular todos los estadísticos básicos para datos circulares y lineales, así como a aplicar las pruebas para determinar las características de las distribuciones citadas en apartados anteriores (Apartado 3.2.3).

Todos los cálculos y resultados se encuentran en el CD que acompaña a esta Tesis, pues el elevado número de datos manejados no aconsejaba su impresión en papel. Lo que sí haremos en estas líneas y en las del anexo, será explicar al lector cómo interpretar esos datos y cálculos y las conclusiones que de ello se derivan. En primer lugar, en el directorio raíz 'Tesis' se accederá a la carpeta 'Datos' y a su vez a 'Escáner'. Dentro de esta última carpeta se encuentran las correspondientes a cada uno de los ficheros de calibración, nombradas, como ya se dijo, por la fecha del evento. La dirección a seguir para consultar los datos es:

D://Tesis/Datos/Escáner/nombre_fichero

Cada una de estas carpetas contiene:

- Fichero de texto original, resultado de la calibración, con extensión .txt y la correspondiente imagen de los vectores de error.
- Hoja de cálculo Excel con el cálculo del módulo y acimut de cada vector de error, así como los estadísticos básicos de datos circulares y la aplicación de los tests estadísticos.
- Cálculo de los estadísticos básicos de datos circulares y aplicación de los tests mediante el programa Oriana (Service, 2004).
- Cálculos con el programa R (Team, 2005), que se explicará en el anexo, para el tratamiento de datos lineales.
- Cálculos con el programa SPSS (SPSS, 2003) que también se explicará en el anexo, para el tratamiento de datos lineales.
- Cálculos de los mapas de densidad.

4.1.3.1 Análisis de datos circulares

Los análisis realizados sobre los datos circulares (sólo acimutes de módulo unidad) procedentes de las diferentes calibraciones del escáner fotogramétrico han sido:

a) Análisis generales

- Cálculos de los estadísticos básicos para vectores unitarios (Apartado 3.2.1.1).

Los valores calculados son: el acimut medio ($\bar{\theta}$), el módulo medio (\bar{R}), la desviación estándar circular de la muestra (v) y el parámetro de concentración de von Mises (κ), pues entendemos que estos valores resumen convenientemente las características principales de la distribución de los datos. Existen, por supuesto, más estadísticos reseñados en el apartado 3.2.1.1 y calculados en los ficheros correspondientes que aportan información redundante. Por ejemplo, la varianza circular de la muestra (V) se obtiene como diferencia entre la unidad y el módulo medio y su significación es, por tanto, complementaria a la del módulo medio.

- Pruebas para la determinación de las características de las distribuciones de los datos circulares (Apartado 3.2.3.1). Estas pruebas se refieren a:
 - Control de la uniformidad de la distribución (tests de Kuiper, Rayleigh, Rao y Watson).
 - Control de la anisotropía (estimación del intervalo de confianza para la dirección media, siempre que sea posible).

- Bondad de ajuste a la distribución von Mises (tests de Watson y Kuiper).
- Gráficos de distribución de datos circulares.

Los gráficos presentados son en realidad gráficos de distribución de datos brutos, que pueden ayudar, en combinación con las pruebas de control de error, a concluir las características de las distribuciones de datos.

b) Análisis locales (ventanas)

Serán los mismos que los realizados en los análisis generales pero aplicados a las ventanas o submuestras indicadas anteriormente (Tabla 4-2).

4.1.3.2 Análisis de datos lineales

Los análisis realizados sobre los módulos de los vectores de error del material que nos ocupa son:

a) Análisis generales

- Cálculos de los estadísticos básicos de los módulos de los vectores (Apartado 3.2.1.2).

En este apartado se incluyen la media aritmética (\bar{d}), la desviación típica (s) y los valores máximos y mínimos de la muestra.

- Determinación de *outliers*.

Con el objetivo de localizar los posibles errores groseros, se localizarán los casos donde el módulo del vector supere la media más 3 veces el valor de la desviación típica.

- Pruebas para la determinación de las características de las distribuciones de los datos lineales (Apartado 3.2.3.2).

Se refiere este apartado a las pruebas de bondad de ajuste a determinadas distribuciones de datos mediante el test de Kolmogorov-Smirnov (K-S). Estas distribuciones, en general, son asimétricas positivas, aunque también se use la distribución normal como elemento de contraste.

- Histogramas de frecuencia de valores.

De igual modo que los gráficos de distribución de datos brutos en datos circulares complementaban las pruebas de error, la inclusión de los histogramas de frecuencias suponen un incremento de la información a la hora de inferir conclusiones.

b) Análisis locales (ventanas)

Se realizan los mismos que para los análisis generales, pero aplicados a las ventanas o submuestras ya indicadas (Tabla 4-2).

4.1.3.3 Análisis en la distribución conjunta

El análisis de la distribución conjunta supone una observación más real de la situación que la realizada a los elementos del vector de error (módulo y acimut) por separado y comprende (Apartado 3.2.3.3):

- Correlaciones existentes entre los módulos y los acimutes en cada muestra.

Si los vectores de error tienen módulo y sentido, que hay que estudiar y acotar, la siguiente pregunta a contestar es si estos módulos y acimutes presentan alguna relación, esto es, si existen direcciones preferentes que se correspondan con un valor de módulo determinado.

- Diferencias entre los valores de la desviación típica en la dirección SCAN (eje X) y CCD (eje Y).

El módulo de los vectores de error tratados se calcula a partir de un incremento de la coordenada X (dirección SCAN) y otro de la coordenada Y (dirección CCD). Si existe una anisotropía en las direcciones puede ser interesante buscar una relación entre los errores cometidos en ambas direcciones.

- Determinación de *outliers*.

En la nube de puntos creada por los nodos finales de los vectores de error estableceremos de entrada como *outlier* un porcentaje de elementos que inicialmente se cifrará en un 2% de los puntos. El método utilizado (Dixon and Chapman, 1980) elimina los puntos según el valor más alto de la media armónica, tal como se explicó en el apartado 3.2.3.3.

- Mapas de densidad.

La nube de puntos indicada anteriormente presenta una forma de la que debemos inferir la distribución. Para ello se calcularán los mapas de densidad como el número de puntos que se corresponden por unidad de área, a modo de curvas de nivel.

4.1.4 Resultados

Los resultados de aplicar las pruebas y análisis anteriormente indicados aparecen a continuación con el mismo esquema utilizado hasta ahora: datos circulares, datos lineales y distribución conjunta.

4.1.4.1 Resultados del análisis de datos circulares

Todos los resultados de los análisis realizados sobre los datos de los 22 ficheros de calibración del escáner fotogramétrico aparecen resumidos en una hoja de cálculo Excel, llamada 'Resumen'. Para acceder a este fichero se debe seguir el siguiente camino en el CD adjunto:

D://Tesis/Datos/Escáner/Resumen

a) Resultados de los análisis generales

- Cálculo de los estadísticos básicos para vectores unitarios.

En la tabla siguiente aparecen diferenciados los valores de los estadísticos (el acimut medio ($\bar{\theta}$), el módulo medio (\bar{R}), la desviación estándar circular de la muestra (ν) y el parámetro de concentración de von Mises (κ)) indicados según el fichero de calibración. Recordemos que las calibraciones '2004-10-19' y '2006-05-19c' tienen solo 361 elementos, mientras que el resto presenta 2116, por lo que a partir de ahora se anotaran estos ficheros en cursiva para diferenciarlos del resto.

Calibración	$\bar{\theta}$	\bar{R}	ν	κ
<i>2004-10-19</i>	141.86°	0.02	161.47°	0.04
2005-02-21	194.35°	0.07	132.04°	0.14
2005-03-01	228.51°	0.02	155.49°	0.05
2005-03-07	179.61°	0.08	128.96°	0.16
2005-03-14	164.23°	0.04	144.11°	0.08
2005-04-04	205.66°	0.03	152.65°	0.06
2005-04-11	2.60°	0.07	133.29°	0.13
2005-04-18	148.89°	0.01	169.79°	0.02
2005-04-25	350.27°	0.03	153.48°	0.05
2005-04-21e	4.05°	0.02	157.65°	0.04
2005-04-22e	174.72°	0.01	174.49°	0.02
2005-05-03e	29.75°	0.07	131.53°	0.14
2005-05-18e	176.82°	0.04	144.56°	0.08
2005-05-23e	52.43°	0.05	141.26°	0.10
2005-06-09	51.22°	0.01	170.53°	0.02
2005-06-29	134.86°	0.02	155.42°	0.05
2005-07-05e	17.71°	0.10	121.70°	0.21
2005-07-11e	209.15°	0.03	149.48°	0.07
2005-07-12e	185.24°	0.04	143.33°	0.09
2005-07-13e	194.22°	0.02	160.91°	0.04
<i>2006-05-19c</i>	184.12°	0.15	112.41°	0.29
2006-06-28c	174.59°	0.09	126.26°	0.18

Tabla 4-5. Resultados de los estadísticos básicos para vectores unitarios en el análisis general.

Observaremos que los resultados que vamos a obtener posteriormente hacen que los valores de acimut medio sean muy poco representativos de las distribuciones, lo cual puede ya verse dados los grandes valores

de desviación circular estándar y los bajos valores de concentración de von Mises.

- Pruebas para la determinación de las características de las distribuciones de los datos circulares.
 - Control de la uniformidad. Se resumen a continuación los resultados de la aplicación de los tests de uniformidad a los datos correspondientes. La mayoría de los ficheros de calibración presentaban unos resultados iguales, en cuanto al valor de probabilidad, a los que aparecen en la tabla siguiente. Se observa que no se indican los resultados de la prueba de Rayleigh, que será objeto de una tabla específica, por presentar valores más dispares.

Prueba	Probabilidad de uniformidad
Kuiper	$P < 0.01$
Rao	$P < 0.01$
Watson	$P < 0.005$

Tabla 4-6. Resultados de las pruebas de uniformidad de Kuiper, Rao y Watson en el análisis general para 18 de 22 ficheros de datos circulares.

Los resultados anteriores muestran que las distribuciones no pueden considerarse uniformes para un nivel de significación de 0.01 lo cual ya empieza a descartar que las distribuciones del error sean aleatorias.

En la tabla que se lista a continuación se indican los resultados de los 4 ficheros que ofrecían probabilidades de uniformidad diferentes a las ya indicadas en la tabla anterior.

Pruebas	2004-10-19	2005-03-01	2005-06-09	2005-06-29
Kuiper	$P > 0.15$	$0.10 > P > 0.05$	$P < 0.01$	$P < 0.01$
Rao	$0.99 > P > 0.95$	$0.90 > P > 0.50$	$0.50 > P > 0.10$	$0.10 > P > 0.05$
Watson	$P > 0.50$	$0.10 > P > 0.05$	$P < 0.005$	$P < 0.01$

Tabla 4-7. Resultados de las pruebas de uniformidad de Kuiper, Rao y Watson en el análisis general que ofrecían resultados diferentes a la mayoría.

De las cuatro pruebas realizadas para comprobar el ajuste de los datos a una distribución uniforme, es la de Rayleigh la que ofrece resultados de probabilidad más dispares, como se observa en la tabla adjunta.

Fichero Calibración	Test de Rayleigh	Fichero Calibración	Test de Rayleigh
2004-10-19	P=0.88	2005-05-03e	P=0.00
2005-02-21	P=0.00	2005-05-18e	P=0.03
2005-03-01	P=0.26	2005-05-23e	P=0.01
2005-03-07	P=0.00	2005-06-09	P=0.74
2005-03-14	P=0.02	2005-06-29	P=0.26
2005-04-04	P=0.17	2005-07-05e	P=0.00
2005-04-11	P=0.00	2005-07-11e	P=0.10
2005-04-18	P=0.72	2005-07-12e	P=0.02
2005-04-25	P=0.20	2005-07-13e	P=0.45
2005-04-21e	P=0.34	2006-05-19c	P=0.00
2005-04-22e	P=0.82	2006-06-28c	P=0.00

Tabla 4-8. Resultados del test de Rayleigh para datos circulares en el análisis general.

Se puede decir que, para estos conjuntos de datos, existen inconsistencias a la hora de evaluar los resultados por lo que será necesario analizar más profundamente las distribuciones para caracterizarlas adecuadamente. Aún así, en una mayoría de análisis, la hipótesis de uniformidad puede descartarse.

- El control de la anisotropía será estudiado en función del error estándar de la media ($\hat{\sigma}$), pues este estadístico se utiliza para determinar el intervalo de confianza para el acimut medio en la distribución de von Mises. En la tabla adjunta aparece el resultado de este estadístico para cada uno de los ficheros de calibración; recordemos que los indicados en letra cursiva tienen 361 elementos mientras que el resto se compone de 2116 vectores de error.

Fichero Calibración	$\hat{\sigma}$	Fichero Calibración	$\hat{\sigma}$
2004-10-19	113.12°	2005-05-03e	12.26°
2005-02-21	12.52°	2005-05-18e	21.24°
2005-03-01	35.00°	2005-05-23e	18.38°
2005-03-07	11.10°	2005-06-09	73.85°
2005-03-14	20.81°	2005-06-29	34.88°
2005-04-04	30.63°	2005-07-05e	8.38°
2005-04-11	13.17°	2005-07-11e	26.47°
2005-04-18	71.08°	2005-07-12e	20.11°
2005-04-25	31.83°	2005-07-13e	45.44°
2005-04-21e	38.79°	2006-05-19c	14.53°
2005-04-22e	91.96°	2006-06-28c	9.96°

Tabla 4-9. Resultados del error estándar del acimut medio en el análisis general para datos circulares.

En general, el valor de este error es muy alto, tal como se aprecia en la tabla anterior, lo que se traduce en algunos casos en intervalos de confianza para el acimut medio carentes de sentido (Ver Fig. 4-2 donde aparece en color rojo un intervalo que prácticamente ocupa la circunferencia entera).

- Las pruebas de bondad de ajuste a la distribución von Mises se resumen a continuación.

Fichero	Test de Watson	Test de Kuiper
2004-10-19	$0.25 > P > 0.15$	$P > 0.15$
2005-03-01	$P < 0.005$	$0.15 > P > 0.10$
Resto de ficheros	$P < 0.005$	$P < 0.01$

Tabla 4-10. Resultados de las pruebas de ajuste a von Mises en el análisis general.

En relación a los errores introducidos artificialmente, no se aprecian diferencias significativas de los datos anteriores entre ficheros que tenían introducido ese error y los que no.

- Gráficos de distribución de datos circulares.

De cada fichero de calibración se realizó un dibujo simple de datos brutos tal como aparece en la figura siguiente, correspondiente a la calibración '2005-04-22e' y muy similar al resto. Se observa también la indeterminación existente a la hora de calcular el intervalo de confianza para el acimut medio y la distribución bimodal de los datos.

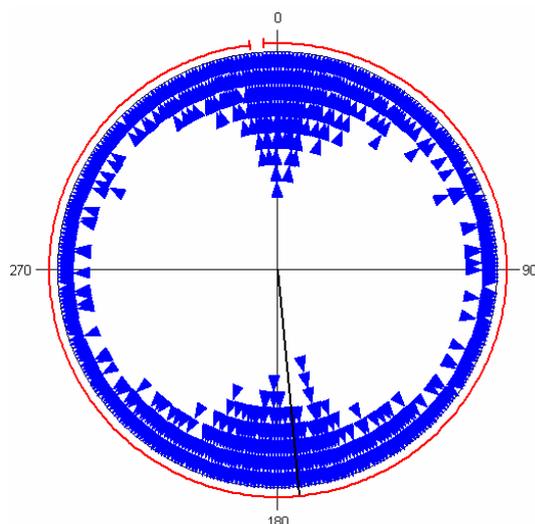


Fig. 4-2. Distribución bimodal de datos circulares.

Se aprecia claramente una distribución bimodal de datos en la línea 0°-180°.

b) Resultados de los análisis locales (ventanas)

Recordemos que en 6 de las 22 calibraciones se analizaron submuestras de vectores, con las características ya indicadas en la tabla correspondiente (Tabla 4-2). En cada una de estas 6 calibraciones se trataron 81 submuestras de 10x10 elementos. Además, en una de las calibraciones, la correspondiente a '2005-02-21', se analizó el fichero varias veces, dividido en ventanas de diferentes tamaños (10x10, 15x15, 20x20, 25x25, 30x30 y 35x35).

Los resultados siguientes, relativos al análisis de estas submuestras, se pueden consultar en la dirección:

D://Tesis/Datos/Escáner/Ventanas

A su vez, dentro de cada carpeta para cada una de las 6 calibraciones objeto de este análisis local, se encuentra una hoja de cálculo Excel, donde aparecen resumidas todas las operaciones realizadas. Los cálculos y resultados de las ventanas de diferentes tamaños se localizan en la dirección adjunta.

D://Tesis/Datos/Escáner/Ventanas/2005-02-21ventana

- Cálculos de los estadísticos básicos para vectores unitarios.

En las tablas que se listan a continuación se indican los valores de los estadísticos básicos de datos circulares para cada una de las 6 calibraciones en las que se realizaron análisis locales con submuestras de 10x10 elementos. En la tabla, al igual que se hizo en el análisis general, aparecen diferenciados los valores del acimut medio ($\bar{\theta}$), el módulo medio (\bar{R}), la desviación estándar circular de la muestra (v) y el parámetro de concentración de von Mises (κ).

Recordemos que los 2116 vectores de cada fichero de calibración se distribuyen en una malla de 46 filas por 46 columnas. El análisis en submuestras o ventanas implica hacer subconjuntos de vectores también en forma de malla de igual número de filas que de columnas. En 6 de las 22 calibraciones estas ventanas tenían 10x10 elementos, construyendo las submuestras con un solape del 50%, tanto en dirección X como en Y. Esto originó 81 ventanas de 100 elementos cada una. La forma de numerar cada ventana es: fila origen-fila final, columna origen-columna final. Para mantener el tamaño de ventana en 100 elementos fue necesario aplicar un solape mayor del 50% a la última serie de elementos. Así, la primera ventana la denominamos (1-10,1-10) y la última (37-46,37-46).

Este criterio de numeración es válido también para el análisis de ventanas de diferentes tamaños con la salvedad correspondiente al distinto tamaño de muestra. También se mantuvo un solape del 50%, siendo necesario aumentar el mismo en algunas ventanas para completar correctamente el análisis.

De las 6 tablas para los estadísticos básicos para vectores unitarios sólo se listan 2 a modo de muestra, dado su gran tamaño y la similitud de resultados. En general, en todos los cálculos que siguen para muestras de tamaño 10x10 elementos y para facilitar la lectura de la Tesis, no aparecen todas las tablas, que, por supuesto, pueden consultarse en la dirección ya indicada del CD adjunto.

	1-10, 1-10	1-10, 5-14	1-10, 11-20	1-10, 15-24	1-10, 21-30	1-10, 25-34	1-10, 31-40	1-10, 35-44	1-10, 37-46
$\bar{\theta}$	25.22°	17.49°	9.74°	10.04°	4.05°	355.42°	349.36°	335.55°	324.28°
\bar{R}	0.86	0.87	0.90	0.90	0.80	0.76	0.75	0.77	0.76
υ	32.10°	30.54°	25.82°	26.54°	37.96°	42.34°	43.02°	41.83°	42.40°
κ	3.75	4.08	5.47	5.21	2.90	2.46	2.40	2.50	2.45
	5-14, 1-10	5-14, 5-14	5-14, 11-20	5-14, 15-24	5-14, 21-30	5-14, 25-34	5-14, 31-40	5-14, 35-44	5-14, 37-46
$\bar{\theta}$	60.26°	48.37°	31.57°	26.47°	26.40°	331.39°	290.27°	288.26°	280.99°
\bar{R}	0.39	0.28	0.35	0.40	0.23	0.09	0.18	0.36	0.49
υ	78.91°	92.10°	83.03°	77.80°	98.66°	127.13°	106.01°	82.37°	68.91°
κ	0.84	0.57	0.75	0.87	0.47	0.17	0.37	0.76	1.11
	11-20, 1-10	11-20, 5-14	11-20, 11-20	11-20, 15-24	11-20, 21-30	11-20, 25-34	11-20, 31-40	11-20, 35-44	11-20, 37-46
$\bar{\theta}$	151.33°	170.61°	164.92°	162.61°	174.15°	179.69°	185.52°	198.93°	203.23°
\bar{R}	0.48	0.72	0.70	0.70	0.82	0.88	0.81	0.74	0.76
υ	69.88°	46.91°	48.79°	48.19°	36.65°	28.76°	37.03°	44.40°	42.44°
κ	1.08	2.10	1.98	2.02	3.06	4.53	3.01	2.29	2.45
	15-24, 1-10	15-24, 5-14	15-24, 11-20	15-24, 15-24	15-24, 21-30	15-24, 25-34	15-24, 31-40	15-24, 35-44	15-24, 37-46
$\bar{\theta}$	161.51°	170.58°	172.00°	171.04°	177.96°	176.01°	178.87°	189.71°	188.91°
\bar{R}	0.45	0.76	0.80	0.86	0.93	0.91	0.80	0.82	0.88
υ	72.91°	42.10°	37.93°	31.84°	21.26°	25.22°	38.80°	36.43°	29.11°
κ	0.99	2.48	2.90	3.81	7.79	5.70	2.80	3.09	4.43
	21-30, 1-10	21-30, 5-14	21-30, 11-20	21-30, 15-24	21-30, 21-30	21-30, 25-34	21-30, 31-40	21-30, 35-44	21-30, 37-46
$\bar{\theta}$	183.00°	182.47°	182.39°	177.78°	179.95°	174.84°	166.46°	174.94°	175.37°
\bar{R}	0.59	0.74	0.87	0.88	0.93	0.82	0.61	0.61	0.71
υ	59.15°	44.54°	29.82°	29.06°	21.43°	35.76°	56.65°	56.60°	47.93°
κ	1.46	2.28	4.25	4.44	7.68	3.17	1.57	1.57	2.04
	25-34, 1-10	25-34, 5-14	25-34, 11-20	25-34, 15-24	25-34, 21-30	25-34, 25-34	25-34, 31-40	25-34, 35-44	25-34, 37-46
$\bar{\theta}$	238.24°	221.76°	194.58°	186.43°	187.06°	181.38°	167.93°	168.60°	161.29°
\bar{R}	0.37	0.53	0.85	0.88	0.92	0.68	0.30	0.27	0.33
υ	81.20°	64.44°	32.37°	29.41°	24.03°	50.01°	89.57°	93.40°	85.87°
κ	0.79	1.26	3.70	4.35	6.22	1.91	0.62	0.55	0.69
	31-40, 1-10	31-40, 5-14	31-40, 11-20	31-40, 15-24	31-40, 21-30	31-40, 25-34	31-40, 31-40	31-40, 35-44	31-40, 37-46
$\bar{\theta}$	283.37°	257.90°	212.79°	208.09°	199.27°	172.06°	63.85°	70.57°	71.34°
\bar{R}	0.68	0.64	0.72	0.61	0.45	0.29	0.12	0.21	0.29
υ	50.52°	53.94°	46.97°	56.91°	72.05°	90.58°	118.13°	101.06°	90.79°
κ	1.88	1.69	2.10	1.55	1.02	0.60	0.24	0.43	0.60
	35-44, 1-10	35-44, 5-14	35-44, 11-20	35-44, 15-24	35-44, 21-30	35-44, 25-34	35-44, 31-40	35-44, 35-44	35-44, 37-46
$\bar{\theta}$	299.73°	284.48°	230.06°	259.36°	6.24°	33.32°	38.59°	46.16°	48.14°
\bar{R}	0.56	0.40	0.29	0.14	0.29	0.37	0.52	0.56	0.59
υ	62.17°	77.57°	89.90°	112.97°	90.65°	81.03°	65.93°	61.30°	58.61°
κ	1.34	0.87	0.61	0.29	0.60	0.79	1.20	1.37	1.48
	37-46, 1-10	37-46, 5-14	37-46, 11-20	37-46, 15-24	37-46, 21-30	37-46, 25-34	37-46, 31-40	37-46, 35-44	37-46, 37-46
$\bar{\theta}$	318.62°	319.21°	291.13°	330.56°	5.82°	20.82°	30.00°	35.86°	38.73°
\bar{R}	0.60	0.44	0.23	0.30	0.65	0.70	0.80	0.81	0.82
υ	58.01°	73.46°	98.94°	88.77°	53.42°	48.14°	38.23°	37.04°	35.95°
κ	1.51	0.98	0.46	0.63	1.72	2.02	2.87	3.01	3.15

Tabla.4-11. Resultados del cálculo de los estadísticos básicos de datos circulares para la calibración '2005-02-21' con tamaño de muestra de 10x10 elementos.

	1-10, 1-10	1-10, 5-14	1-10, 11-20	1-10, 15-24	1-10, 21-30	1-10, 25-34	1-10, 31-40	1-10, 35-44	1-10, 37-46
$\bar{\theta}$	2.17°	355.34°	343.65°	14.36°	38.05°	42.01°	83.38°	142.70°	140.28°
\bar{R}	0.83	0.85	0.48	0.40	0.74	0.74	0.40	0.52	0.71
ν	35.20°	33.12°	69.54°	78.04°	44.47°	44.75°	77.24°	65.97°	47.75°
κ	3.25	3.57	1.09	0.86	2.28	2.26	0.88	1.20	2.05
	5-14, 1-10	5-14, 5-14	5-14, 11-20	5-14, 15-24	5-14, 21-30	5-14, 25-34	5-14, 31-40	5-14, 35-44	5-14, 37-46
$\bar{\theta}$	343.39°	329.88°	268.97°	291.40°	11.65°	10.42°	150.44°	181.40°	176.98°
\bar{R}	0.70	0.73	0.51	0.30	0.74	0.72	0.12	0.71	0.82
ν	48.56°	45.30°	66.37°	89.18°	44.13°	46.74°	117.90°	47.35°	36.55°
κ	2.00	2.22	1.19	0.62	2.31	2.12	0.24	2.08	3.07
	11-20, 1-10	11-20, 5-14	11-20, 11-20	11-20, 15-24	11-20, 21-30	11-20, 25-34	11-20, 31-40	11-20, 35-44	11-20, 37-46
$\bar{\theta}$	291.57°	257.69°	214.49°	214.52°	343.79°	338.11°	207.76°	213.38°	222.72°
\bar{R}	0.31	0.47	0.86	0.54	0.32	0.24	0.42	0.43	0.32
ν	87.62°	70.48°	31.87°	63.99°	86.60°	97.54°	75.24°	74.29°	86.08°
κ	0.65	1.06	3.80	1.27	0.67	0.48	0.93	0.96	0.68
	15-24, 1-10	15-24, 5-14	15-24, 11-20	15-24, 15-24	15-24, 21-30	15-24, 25-34	15-24, 31-40	15-24, 35-44	15-24, 37-46
$\bar{\theta}$	240.14°	220.54°	206.33°	203.80°	236.69°	216.15°	235.54°	287.80°	304.21°
\bar{R}	0.34	0.56	0.93	0.71	0.16	0.20	0.37	0.38	0.41
ν	84.42°	61.93°	22.09°	47.92°	109.08°	102.38°	81.02°	79.66°	77.09°
κ	0.72	1.35	7.26	2.04	0.33	0.41	0.79	0.82	0.88
	21-30, 1-10	21-30, 5-14	21-30, 11-20	21-30, 15-24	21-30, 21-30	21-30, 25-34	21-30, 31-40	21-30, 35-44	21-30, 37-46
$\bar{\theta}$	180.76°	191.24°	196.96°	195.91°	197.84°	194.96°	236.10°	261.02°	247.94°
\bar{R}	0.43	0.79	0.94	0.81	0.53	0.54	0.37	0.40	0.42
ν	74.39°	39.68°	19.79°	36.87°	64.96°	63.32°	80.48°	77.71°	75.04°
κ	0.95	2.71	8.91	3.03	1.23	1.30	0.80	0.87	0.94
	25-34, 1-10	25-34, 5-14	25-34, 11-20	25-34, 15-24	25-34, 21-30	25-34, 25-34	25-34, 31-40	25-34, 35-44	25-34, 37-46
$\bar{\theta}$	143.78°	176.34°	186.01°	182.62°	178.62°	178.95°	212.80°	230.92°	217.79°
\bar{R}	0.48	0.79	0.92	0.79	0.44	0.47	0.41	0.41	0.47
ν	69.96°	39.54°	22.84°	39.75°	73.57°	70.41°	76.62°	76.54°	70.80°
κ	1.08	2.73	6.83	2.70	0.98	1.06	0.90	0.90	1.05
	31-40, 1-10	31-40, 5-14	31-40, 11-20	31-40, 15-24	31-40, 21-30	31-40, 25-34	31-40, 31-40	31-40, 35-44	31-40, 37-46
$\bar{\theta}$	68.17°	108.33°	164.69°	141.33°	54.73°	71.09°	294.80°	346.35°	349.64°
\bar{R}	0.46	0.26	0.41	0.35	0.36	0.19	0.14	0.38	0.36
ν	71.50°	93.87°	76.08°	82.56°	81.96°	104.43°	113.72°	79.90°	81.46°
κ	1.03	0.54	0.91	0.76	0.77	0.39	0.28	0.82	0.78
	35-44, 1-10	35-44, 5-14	35-44, 11-20	35-44, 15-24	35-44, 21-30	35-44, 25-34	35-44, 31-40	35-44, 35-44	35-44, 37-46
$\bar{\theta}$	43.76°	38.52°	100.00°	92.65°	27.45°	28.43°	358.99°	2.63°	4.20°
\bar{R}	0.59	0.47	0.21	0.23	0.60	0.49	0.37	0.70	0.75
ν	58.59°	70.11°	101.63°	98.65°	58.09°	68.38°	80.92°	48.40°	43.67°
κ	1.48	1.07	0.42	0.47	1.50	1.12	0.79	2.01	2.35
	37-46, 1-10	37-46, 5-14	37-46, 11-20	37-46, 15-24	37-46, 21-30	37-46, 25-34	37-46, 31-40	37-46, 35-44	37-46, 37-46
$\bar{\theta}$	40.90°	36.44°	96.36°	125.00°	20.77°	17.01°	352.36°	355.51°	358.00°
\bar{R}	0.63	0.59	0.27	0.21	0.42	0.47	0.47	0.72	0.76
ν	55.55°	59.10°	93.04°	100.68°	75.56°	70.72°	70.91°	46.37°	42.09°
κ	1.62	1.46	0.56	0.44	0.92	1.05	1.05	2.14	2.48

Tabla 4-12. Resultados del cálculo de los estadísticos básicos de datos circulares para la calibración '2005-04-18' con tamaño de muestra de 10x10 elementos.

La calibración correspondiente al fichero '2005-02-21' fue analizada también con diferentes tamaños de ventanas. Los tamaños de muestra analizados fueron 10x10, 15x15, 20x20, 25x25, 30x30 y 35x35. La primera tabla que se muestra es la correspondiente al producto de 15x15 elementos, ya que la de 10x10 se listó anteriormente ().

	1-15, 1-15	1-15, 8-22	1-15, 15-29	1-15, 22-36	1-15, 29-43	1-15, 32-46
$\bar{\theta}$	44.37°	28.94°	20.18°	3.61°	312.37°	304.54°
\bar{R}	0.48	0.45	0.40	0.26	0.30	0.41
ν	69.27°	72.02°	77.58°	94.44°	89.38°	76.69°
κ	1.10	1.02	0.87	0.53	0.62	0.89
	8-22, 1-15	8-22, 8-22	8-22, 15-29	8-22, 22-36	8-22, 29-43	8-22, 32-46
$\bar{\theta}$	133.44°	153.99°	164.02°	174.61°	189.88°	202.28°
\bar{R}	0.40	0.47	0.49	0.56	0.60	0.56
ν	77.48°	70.29°	68.10°	61.73°	57.85°	61.52°
κ	0.88	1.07	1.13	1.36	1.51	1.36
	15-29, 1-15	15-29, 8-22	15-29, 15-29	15-29, 22-36	15-29, 29-43	15-29, 32-46
$\bar{\theta}$	176.09°	172.89°	175.39°	179.13°	178.77°	182.43°
\bar{R}	0.62	0.85	0.89	0.82	0.78	0.74
ν	55.84°	32.58°	27.50°	36.64°	40.17°	44.79°
κ	1.60	3.66	4.89	3.06	2.66	2.26
	22-36, 1-15	22-36, 8-22	22-36, 15-29	22-36, 22-36	22-36, 29-43	22-36, 32-46
$\bar{\theta}$	217.69°	195.43°	185.96°	187.82°	179.74°	181.95°
\bar{R}	0.57	0.76	0.87	0.66	0.52	0.35
ν	60.72°	42.88°	29.88°	52.56°	65.48°	82.74°
κ	1.39	2.41	4.24	1.76	1.21	0.75
	29-43, 1-15	29-43, 8-22	29-43, 15-29	29-43, 22-36	29-43, 29-43	29-43, 32-46
$\bar{\theta}$	270.77°	220.68°	200.67°	255.12°	65.95°	36.28°
\bar{R}	0.47	0.49	0.39	0.02	0.19	0.33
ν	70.91°	68.17°	78.74°	158.58°	104.07°	85.32°
κ	1.05	1.13	0.84	0.04	0.39	0.70
	32-46, 1-15	32-46, 8-22	32-46, 15-29	32-46, 22-36	32-46, 29-43	32-46, 32-46
$\bar{\theta}$	296.81°	258.39°	269.65°	2.38°	36.10°	31.80°
\bar{R}	0.52	0.30	0.14	0.29	0.37	0.50
ν	65.61°	88.41°	114.61°	90.60°	80.50°	67.42°
κ	1.21	0.64	0.27	0.60	0.80	1.15

Tabla 4-13. Resultados del cálculo de los estadísticos básicos para la calibración '2005-02-21' con tamaño de muestra de 15x15 elementos.

Observamos en esta tabla, en relación con la de tamaño 10x10 que el valor medio del módulo medio decrece: 0.60 en el primer caso con 0.49 en el segundo. Lo mismo ocurre con el parámetro de concentración de von Mises (2.18 en tamaño 10x10 y 1.42 en 15x15). La desviación estándar circular de la muestra, por el contrario, aumenta su valor medio, de 58.3° a 70.0°.

La tabla que se muestra a continuación lista los resultados de los estadísticos básicos para vectores unitarios de la calibración '2005-02-21' para un tamaño de muestra de 20x20 elementos.

	1-20, 1-20	1-20, 10-29	1-20, 21-40	1-20, 27-46
$\bar{\theta}$	58.23°	68.05°	225.60°	257.97°
\bar{R}	0.27	0.17	0.03	0.23
ν	92.45°	107.59°	153.84°	98.56°
κ	0.57	0.35	0.05	0.47
	10-29, 1-20	10-29, 10-29	10-29, 21-40	10-29, 27-46
$\bar{\theta}$	168.50°	171.21°	177.36°	186.33°
\bar{R}	0.61	0.76	0.76	0.73
ν	57.13°	42.99°	42.81°	45.81°
κ	1.54	2.40	2.42	2.18
	21-40, 1-20	21-40, 10-29	21-40, 21-40	21-40, 27-46
$\bar{\theta}$	210.34°	192.02°	176.94°	163.36°
\bar{R}	0.55	0.72	0.48	0.35
ν	62.54°	46.49°	69.74°	82.91°
κ	1.32	2.13	1.08	0.75
	27-46, 1-20	27-46, 10-29	27-46, 21-40	27-46, 27-46
$\bar{\theta}$	258.91°	218.05°	48.03°	45.71°
\bar{R}	0.39	0.30	0.07	0.25
ν	62.54°	46.49°	69.74°	82.91°
κ	0.84	0.63	0.14	0.52

Tabla 4-14. Resultados del cálculo de los estadísticos básicos para la calibración '2005-02-21' con tamaño de muestra de 20x20 elementos.

Al igual que ocurría con la tabla anterior al incrementar el tamaño de la muestra se reduce la magnitud media del módulo medio y del parámetro de concentración de von Mises, mientras que aumenta el valor medio de la desviación estándar circular.

	1-25, 1-25	1-25, 12-36	1-25, 22-46
$\bar{\theta}$	102.63°	152.18°	208.36°
\bar{R}	0.21	0.17	0.23
ν	100.99°	108.46°	98.89°
κ	0.43	0.34	0.46
	12-36, 1-25	12-36, 12-36	12-36, 22-46
$\bar{\theta}$	186.22°	182.26°	184.53°
\bar{R}	0.62	0.76	0.67
ν	56.52°	42.31°	51.61°
κ	1.57	2.46	1.82
	22-46, 1-25	22-46, 12-36	22-46, 22-46
$\bar{\theta}$	225.84°	190.51°	123.90°
\bar{R}	0.33	0.26	0.12
ν	84.90°	94.18°	117.68°
κ	0.71	0.54	0.24

Tabla 4-15. Resultados del cálculo de los estadísticos básicos para la calibración '2005-02-21' con tamaño de muestra de 25x25 elementos.

Corresponde ahora desarrollar los resultados para un tamaño de muestras de 30x30 elementos.

	1-30, 1-30	1-30, 17-46
$\bar{\theta}$	151.36°	190.68°
\bar{R}	0.23	0.25
υ	98.02°	94.79°
κ	0.48	0.53
	17-46, 1-30	17-46, 17-46
$\bar{\theta}$	206.78°	164.25°
\bar{R}	0.34	0.27
υ	83.90°	92.81°
κ	0.73	0.56

Tabla 4-16. Resultados del cálculo de los estadísticos básicos para la calibración '2005-02-21' con tamaño de muestra de 30x30 elementos.

Como se comprueba en toda esta serie de tablas, conforme aumentamos el tamaño de la muestra aumenta el valor medio de la desviación estándar circular y disminuye el valor medio del módulo medio y el parámetro de concentración de von Mises. El acimut medio, en general, no presenta variaciones significativas.

	1-35,1-35	1-35,12-46
$\bar{\theta}$	169.58°	190.04°
\bar{R}	0.23	0.28
υ	98.01°	90.86°
κ	0.48	0.59
	12-46,1-35	12-46,12-46
$\bar{\theta}$	189.00°	176.81°
\bar{R}	0.34	0.36
υ	83.66°	82.05°
κ	0.73	0.77

Tabla 4-17. Resultados del cálculo de los estadísticos básicos para la calibración '2005-02-21' con tamaño de muestra de 35x35 elementos.

Sobre estos resultados anteriores realizamos las siguientes observaciones:

Observaciones

- El valor medio del acimut medio es 178.7° en el caso de las submuestras de 10x10 elementos. Este valor es mayor que el medio obtenido en las pruebas generales (143.9°). En la serie de las muestras de diferente tamaño, la tendencia es a aumentar el valor del acimut medio conforme aumenta el tamaño de la muestra, tal como se indica en el gráfico siguiente. Así, la muestra de menor tamaño tiene un valor de 166.3° mientras que la global es de 194.3°.

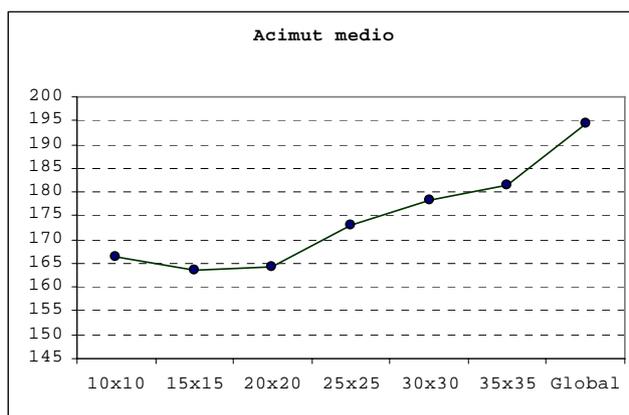


Fig. 4-3. Acimut medio en muestras de diferente tamaño.

En la gráfica se representa el valor medio del acimut medio en el fichero de calibración '2005-02-21' para diferentes tamaños de muestra. Observamos cómo aumenta el valor del acimut medio según aumenta el tamaño de la muestra.

- El valor medio del módulo medio (\bar{R}) es 0.52 con poca diferencia entre las diferentes muestras de 10x10 elementos. Este valor contrasta con el 0.04 del análisis general. En las muestras de diferentes tamaños de ventanas vemos cómo decrece el valor numérico del módulo medio al tiempo que aumenta el tamaño de la muestras, es decir, a mayor tamaño de la muestra mayor indicación de dispersión de valores.

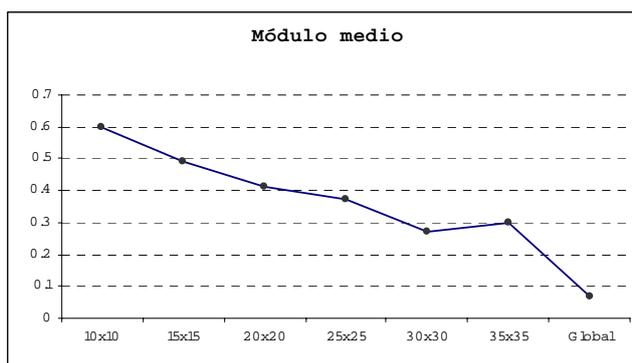


Fig. 4-4. Módulo medio en muestras de diferente tamaño.

Representación del valor del módulo medio para diferentes tamaños de muestra en la calibración '2005-02-21'. Destacamos el valor decreciente de \bar{R} conforme aumenta el tamaño de la muestra.

- Los valores de la desviación estándar circular para las muestras de tamaño 10x10 presentan una media de 67.1° que contrasta claramente con los 147.3° de media del análisis global. Vemos en el gráfico siguiente como también crece la desviación estándar según aumenta el tamaño de la muestra hasta más del doble.

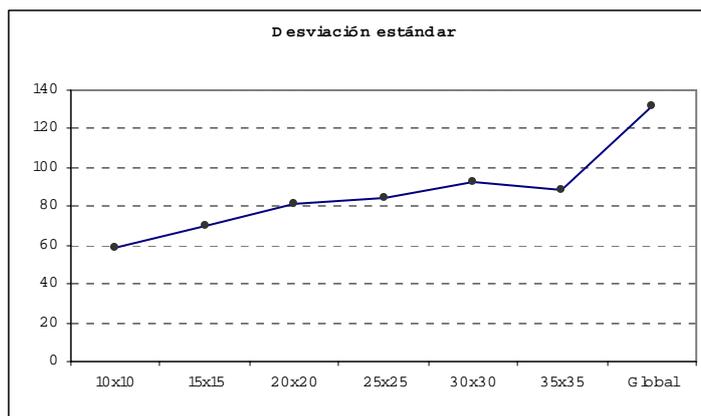


Fig. 4-5. Desviación estándar en muestras de diferente tamaño. Representación del valor de la desviación estándar para diferentes tamaños de muestra del fichero de calibración '2005-02-21'. A mayor tamaño de la muestra le corresponde un valor más alto de desviación.

- El parámetro de concentración de von Mises alcanza un valor medio de 1.67 en las muestras de 10x10 elementos, valor muy alejado de los 0.087 del análisis global. En las submuestras de diferente tamaño, vemos según el gráfico adjunto, que este valor sólo es significativo en muestras pequeñas.

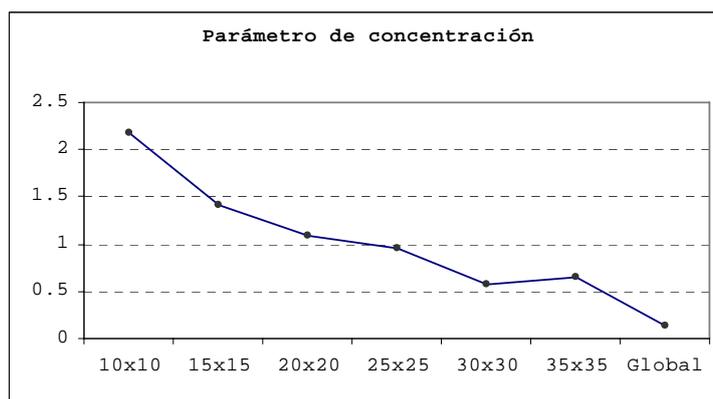


Fig. 4-6. Parámetro de concentración de von Mises en muestras de diferentes tamaño. Representación del valor del parámetro de concentración de von Mises para diferentes tamaños de muestra del fichero de calibración '2005-02-21'. Conforme aumenta el tamaño de la muestra este valor disminuye, por lo que sólo tiene sentido hablar de una dirección preferente en muestras pequeñas.

- Pruebas para la determinación de las características de las distribuciones de los datos circulares.
 - Control de la uniformidad. En la tabla siguiente se indican los resultados obtenidos de la aplicación de las pruebas de uniformidad a la mayoría de las 81 submuestras de 10x10 elementos en que fue dividido cada fichero de calibración.

(81 ventanas)	Kuiper	Rao	Rayleigh	Watson
2005-02-21	P<0.01	P<0.01	P=0	P<0.005
2005-04-18	P<0.01	P<0.01	P=0	P<0.005
2005-04-21e	P<0.01	P<0.01	P=0	P<0.005
2005-04-25	P<0.01	P<0.01	P=0	P<0.005
2005-05-03e	P<0.01	P<0.01	P=0	P<0.005
2006-06-28c	P<0.01	P<0.01	P=0	P<0.005

Tabla 4-18. Resultados de test de uniformidad para ficheros con submuestras de 10x10 elementos.

No todas las submuestras presentaban los valores indicados en la tabla anterior. Se observa que algunas ventanas, entre 3 y 10, presentaban valores del test de Rayleigh y/o Rao que tenían una probabilidad mayor de uniformidad en la distribución.

En la tabla siguiente se indican las submuestras de cada fichero que sí presentaban valores de probabilidad más altos en todas las pruebas, con la forma de nombrar las ventanas ya indicada.

2005-02-21	2005-04-18	2005-04-21e	2005-04-25	2005-05-03e	2006-06-28c
31-40,31-40 35-44,15-24	15-24,21-30 15-24,25-34 31-40,25-34 31-40,31-40	15-24,25-34 21-30,15-24	5-14,15-24 21-30,1-10 21-30,5-14 21-30,11-20 21-30,15-24 21-30,21-30 21-30,31-40 21-30,35-44	37-46,21-30 37-46,25-34	25-34,11 20 25-34,15-24 25-34,21-30 31-40,25-34

Tabla 4-19. Resultados de análisis de submuestras que presentaban evidencias de uniformidad.

El control de la uniformidad para ventanas de diferentes tamaños dentro de la calibración '2005-02-21', arroja los resultados siguientes, con la salvedad hecha en el párrafo anterior para la muestra de tamaño 10x10. En el resto de los tamaños sólo alguna de las ventanas evidenciaba aleatoriedad, principalmente en los tests de Rao y Rayleigh.

2005-02-21	Kuiper	Rao	Rayleigh	Watson
10x10	P<0.01	P<0.01	P=0	P<0.005
15x15	P<0.01	P<0.01	P=0	P<0.005
20x20	P<0.01	P<0.01	P=0	P<0.005
25x25	P<0.01	P<0.01	P=0	P<0.005
30x30	P<0.01	P<0.01	P=0	P<0.005
35x35	P<0.01	P<0.01	P=0	P<0.005
Global	P<0.01	P<0.01	P=0	P<0.005

Tabla 4-20. Resultados de tests de uniformidad para la calibración '2005-02-21' con submuestras de varios tamaños.

- Control de la anisotropía.

El control de la anisotropía se realizará según el valor del error estándar del acimut medio ($\hat{\sigma}$), tal como se indicó para el análisis global. En las tablas siguientes aparecen los resultados correspondientes a este cálculo para 4 de los 6 ficheros de calibración con tamaño de muestra de 10x10 elementos y para la calibración '2005-02-21' en muestras de diferentes tamaños. Se comenzará por estos últimos resultados por comodidad de impresión de las tablas.

	1-15, 1-15	1-15, 8-22	1-15, 15-29	1-15, 22-36	1-15, 29-43	1-15, 32-46
$\hat{\sigma}$	5.26°	5.62°	6.47°	10.33°	8.91°	6.32°
	8-22, 1-15	8-22, 8-22	8-22, 15-29	8-22, 22-36	8-22, 29-43	8-22, 32-46
$\hat{\sigma}$	6.45°	5.39°	5.11°	4.39°	4.01°	4.37°
	15-29, 1-15	15-29, 8-22	15-29, 15-29	15-29, 22-36	15-29, 29-43	15-29, 32-46
$\hat{\sigma}$	3.83°	2.16°	1.83°	2.42°	2.65°	2.96°
	22-36, 1-15	22-36, 8-22	22-36, 15-29	22-36, 22-36	22-36, 29-43	22-36, 32-46
$\hat{\sigma}$	4.28°	2.83°	1.99°	3.55°	4.81°	7.41°
	29-43, 1-15	29-43, 8-22	29-43, 15-29	29-43, 22-36	29-43, 29-43	29-43, 32-46
$\hat{\sigma}$	5.47°	5.12°	6.67°	124.42°	13.93°	7.95°
	32-46, 1-15	32-46, 8-22	32-46, 15-29	32-46, 22-36	32-46, 29-43	32-46, 32-46
$\hat{\sigma}$	4.82°	8.67°	19.88°	9.23°	6.98°	5.03°

Tabla 4-21. Resultados del cálculo del error estándar para el acimut medio en la calibración '2005-02-21' con tamaño de muestra de 15x15 elementos.

El valor medio de los datos de la tabla es 9.21°, valor no muy alejado del obtenido en el análisis global, 12.52°. Los valores máximos y mínimos son 124.42° y 1.83°. Destaca el alto valor máximo correspondiente a la submuestra (29-43, 22-36).

La siguiente tabla muestra el mismo dato tratado, el error estándar para el acimut medio, con tamaño de muestras de 20x20 elementos. La tabla correspondiente al tamaño 10x10 se listará al final junto con las correspondientes a los otros 5 ficheros de calibración citados.

	1-20, 1-20	1-20, 10-29	1-20, 21-40	1-20, 27-46
$\hat{\sigma}$	7.31°	11.72°	74.46°	8.78°
	10-29, 1-20	10-29, 10-29	10-29, 21-40	10-29, 27-46
$\hat{\sigma}$	2.96°	2.13°	2.12°	2.28°
	21-40, 1-20	21-40, 10-29	21-40, 21-40	21-40, 27-46
$\hat{\sigma}$	3.35°	2.31°	3.99°	5.59°
	27-46, 1-20	27-46, 10-29	27-46, 21-40	27-46, 27-46
$\hat{\sigma}$	5.05°	6.55°	29.72°	8.00°

Tabla 4-22. Resultados del cálculo del error estándar para el acimut medio en la calibración '2005-02-21' con tamaño de muestra de 20x20 elementos.

El valor medio de las muestras para esta tabla es 11.02°, con valores máximos de 74.46° y mínimos de 2.12°. El valor máximo en esta tabla, que se alcanza en la ventana (27-46, 21-40) también es alto, en comparación con el resto de valores.

Se muestra a continuación el resultado correspondiente al tamaño de ventana de 25x25 elementos.

	1-25, 1-25	1-25, 12-36	1-25, 22-46
$\hat{\sigma}$	7.57°	9.66°	7.09°
	12-36, 1-25	12-36, 12-36	12-36, 22-46
$\hat{\sigma}$	2.33°	1.68°	2.08°
	22-46, 1-25	22-46, 12-36	22-46, 22-46
$\hat{\sigma}$	4.72°	6.15°	13.31°

Tabla 4-23. Resultados del cálculo del error estándar para el acimut medio en la calibración '2005-02-21' con tamaño de muestra de 25x25 elementos.

Aquí la media es de 6.06° y los valores máximos y mínimos están más próximos que en tablas anteriores (13.31° y 1.68°, respectivamente).

Siguiendo con el análisis, se muestran los resultados del tamaño de muestra 30x30 elementos.

	1-30, 1-30	1-30, 17-46
$\hat{\sigma}$	5.76°	5.22°
	17-46, 1-30	17-46, 17-46
$\hat{\sigma}$	3.83°	4.92°

Tabla 4-24. Resultados del cálculo del error estándar para el acimut medio en la calibración '2005-02-21' con tamaño de muestra de 30x30 elementos.

El valor medio para la tabla anterior es 4.93° y los valores máximos y mínimos están mucho más ajustados (5.76° y 3.83°). La tabla listada a continuación muestra los resultados de las ventanas de tamaño 35x35.

	1-35, 1-35	1-35, 12-46
$\hat{\sigma}$	4.93°	3.99°
	12-46, 1-35	12-46, 12-46
$\hat{\sigma}$	3.26°	3.12°

Tabla 4-25. Resultados del cálculo del error estándar para el acimut medio en la calibración '2005-02-21' con tamaño de muestra de 35x35 elementos.

La media alcanzada es 3.82° y los valores máximos y mínimos siguen estando muy próximos a la media.

En general las diferentes ventanas no muestran diferencias significativas entre sí, ni con el resultado del análisis global, que fue 12.52°. Aparecen ahora los resultados correspondientes al análisis de 10x10 elementos, para 4 de las 6 calibraciones.

	1-10, 1-10	1-10, 5-14	1-10, 11-20	1-10, 15-24	1-10, 21-30	1-10, 25-34	1-10, 31-40	1-10, 35-44	1-10, 37-46
$\hat{\sigma}$	3.20°	3.05°	2.58°	2.65°	3.76°	4.19°	4.26°	4.14°	4.20°
	5-14, 1-10	5-14, 5-14	5-14, 11-20	5-14, 15-24	5-14, 21-30	5-14, 25-34	5-14, 31-40	5-14, 35-44	5-14, 37-46
$\hat{\sigma}$	10.04°	14.46°	11.21°	9.76°	17.61°	47.42°	22.25°	11.01°	7.82°
	11-20, 1-10	11-20, 5-14	11-20, 11-20	11-20, 15-24	11-20, 21-30	11-20, 25-34	11-20, 31-40	11-20, 35-44	11-20, 37-46
$\hat{\sigma}$	8.00°	4.67°	4.88°	4.81°	3.63°	2.87°	3.67°	4.40°	4.20°
	15-24, 1-10	15-24, 5-14	15-24, 11-20	15-24, 15-24	15-24, 21-30	15-24, 25-34	15-24, 31-40	15-24, 35-44	15-24, 37-46
$\hat{\sigma}$	8.62°	4.17°	3.75°	3.17°	2.12°	2.52°	3.84°	3.61°	2.90°
	21-30, 1-10	21-30, 5-14	21-30, 11-20	21-30, 15-24	21-30, 21-30	21-30, 25-34	21-30, 31-40	21-30, 35-44	21-30, 37-46
$\hat{\sigma}$	6.20°	4.42°	2.97°	2.90°	2.14°	3.55°	5.85°	5.84°	4.78°
	25-34, 1-10	25-34, 5-14	25-34, 11-20	25-34, 15-24	25-34, 21-30	25-34, 25-34	25-34, 31-40	25-34, 35-44	25-34, 37-46
$\hat{\sigma}$	10.67°	7.02°	3.23°	2.93°	2.40°	5.02°	13.44°	15.02°	12.11°
	31-40, 1-10	31-40, 5-14	31-40, 11-20	31-40, 15-24	31-40, 21-30	31-40, 25-34	31-40, 31-40	31-40, 35-44	31-40, 37-46
$\hat{\sigma}$	5.08°	5.50°	4.68°	5.88°	8.44°	13.84°	33.82°	18.97°	13.92°
	35-44, 1-10	35-44, 5-14	35-44, 11-20	35-44, 15-24	35-44, 21-30	35-44, 25-34	35-44, 31-40	35-44, 35-44	35-44, 37-46
$\hat{\sigma}$	6.65°	9.70°	13.57°	28.15°	13.86°	10.62°	7.29°	6.51°	6.12°
	37-46, 1-10	37-46, 5-14	37-46, 11-20	37-46, 15-24	37-46, 21-30	37-46, 25-34	37-46, 31-40	37-46, 35-44	37-46, 37-46
$\hat{\sigma}$	6.04°	8.74°	17.76°	13.14°	5.43°	4.81°	3.78°	3.67°	3.56°

Tabla 4-26. Resultados del cálculo del error estándar para el acimut medio para la calibración '2005-02-21' con tamaño de muestra de 10x10 elementos

	1-10, 1-10	1-10, 5-14	1-10, 11-20	1-10, 15-24	1-10, 21-30	1-10, 25-34	1-10, 31-40	1-10, 35-44	1-10, 37-46
$\hat{\sigma}$	3.49°	3.30°	7.94°	9.82°	4.41°	4.44°	9.62°	7.29°	4.76°
	5-14, 1-10	5-14, 5-14	5-14, 11-20	5-14, 15- 24	5-14, 21-30	5-14, 25-34	5-14, 31-40	5-14, 35-44	5-14, 37-46
$\hat{\sigma}$	4.85°	4.50°	7.36°	13.29°	4.37°	4.65°	33.54°	4.72°	3.62°
	11-20, 1-10	11-20, 5-14	11-20, 11-20	11-20, 15-24	11-20, 21-30	11-20, 25-34	11-20, 31-40	11-20, 35-44	11-20, 37-46
$\hat{\sigma}$	12.72°	8.12°	3.18°	6.94°	12.36°	17.01°	9.14°	8.92°	12.18°
	15-24, 1-10	15-24, 5-14	15-24, 11-20	15-24, 15-24	15-24, 21-30	15-24, 25-34	15-24, 31-40	15-24, 35-44	15-24, 37-46
$\hat{\sigma}$	11.64°	6.61°	2.21°	4.78°	24.65°	19.79°	10.62°	10.24°	9.58°
	21-30, 1-10	21-30, 5-14	21-30, 11-20	21-30, 15-24	21-30, 21-30	21-30, 25-34	21-30, 31-40	21-30, 35-44	21-30, 37-46
$\hat{\sigma}$	8.95°	3.92°	1.98°	3.65°	7.12°	6.83°	10.47°	9.74°	9.09°
	25-34, 1-10	25-34, 5-14	25-34, 11-20	25-34, 15-24	25-34, 21-30	25-34, 25-34	25-34, 31-40	25-34, 35-44	25-34, 37-46
$\hat{\sigma}$	8.02°	3.91°	2.28°	3.93°	8.76°	8.11°	9.47°	9.45°	8.18°
	31-40, 1-10	31-40, 5-14	31-40, 11-20	31-40, 15-24	31-40, 21-30	31-40, 25-34	31-40, 31-40	31-40, 35-44	31-40, 37-46
$\hat{\sigma}$	8.33°	15.23°	9.34°	11.06°	10.89°	21.14°	28.90°	10.31°	10.74°
	35-44, 1-10	35-44, 5-14	35-44, 11-20	35-44, 15-24	35-44, 21-30	35-44, 25-34	35-44, 31-40	35-44, 35-44	35-44, 37-46
$\hat{\sigma}$	6.12°	8.05°	19.32°	17.60°	6.05°	7.72°	10.59°	4.84°	4.33°
	37-46, 1-10	37-46, 5-14	37-46, 11-20	37-46, 15-24	37-46, 21-30	37-46, 25-34	37-46, 31-40	37-46, 35-44	37-46, 37-46
$\hat{\sigma}$	5.70°	6.19°	14.86°	18.75°	9.21°	8.17°	8.21°	4.61°	4.16°

Tabla 4-27. Resultados del cálculo del error estándar para el acimut medio para la calibración '2005-04-18' con tamaño de muestra de 10x10 elementos

	1-10, 1-10	1-10, 5-14	1-10, 11-20	1-10, 15-24	1-10, 21-30	1-10, 25-34	1-10, 31-40	1-10, 35-44	1-10, 37-46
$\hat{\sigma}$	7.62°	9.34°	20.40°	15.60°	4.76°	4.48°	7.83°	14.48°	12.65°
	5-14, 1-10	5-14, 5-14	5-14, 11-20	5-14, 15-24	5-14, 21-30	5-14, 25-34	5-14, 31-40	5-14, 35-44	5-14, 37-46
$\hat{\sigma}$	5.51°	10.67°	4.79°	15.23°	2.82°	4.07°	21.76°	4.75°	3.81°
	11-20, 1-10	11-20, 5-14	11-20, 11-20	11-20, 15-24	11-20, 21-30	11-20, 25-34	11-20, 31-40	11-20, 35-44	11-20, 37-46
$\hat{\sigma}$	9.30°	6.04°	2.31°	10.34°	5.93°	20.37°	3.57°	1.61°	3.25°
	15-24, 1-10	15-24, 5-14	15-24, 11-20	15-24, 15-24	15-24, 21-30	15-24, 25-34	15-24, 31-40	15-24, 35-44	15-24, 37-46
$\hat{\sigma}$	6.33°	4.46°	3.00°	11.52°	7.17°	52.44°	4.63°	5.51°	8.49°
	21-30, 1-10	21-30, 5-14	21-30, 11-20	21-30, 15-24	21-30, 21-30	21-30, 25-34	21-30, 31-40	21-30, 35-44	21-30, 37-46
$\hat{\sigma}$	4.96°	3.67°	4.80°	28.96°	5.93°	13.77°	12.95°	7.27°	8.21°
	25-34, 1-10	25-34, 5-14	25-34, 11-20	25-34, 15-24	25-34, 21-30	25-34, 25-34	25-34, 31-40	25-34, 35-44	25-34, 37-46
$\hat{\sigma}$	5.24°	4.90°	8.34°	40.29°	4.95°	7.02°	8.04°	7.55°	10.11°
	31-40, 1-10	31-40, 5-14	31-40, 11-20	31-40, 15-24	31-40, 21-30	31-40, 25-34	31-40, 31-40	31-40, 35-44	31-40, 37-46
$\hat{\sigma}$	7.52°	10.19°	13.01°	15.95°	5.07°	5.51°	8.67°	9.87°	9.91°
	35-44, 1-10	35-44, 5-14	35-44, 11-20	35-44, 15-24	35-44, 21-30	35-44, 25-34	35-44, 31-40	35-44, 35-44	35-44, 37-46
$\hat{\sigma}$	7.08°	8.74°	9.73°	13.93°	5.68°	5.16°	10.90°	11.72°	9.32°
	37-46, 1-10	37-46, 5-14	37-46, 11-20	37-46, 15-24	37-46, 21-30	37-46, 25-34	37-46, 31-40	37-46, 35-44	37-46, 37-46
$\hat{\sigma}$	8.34°	10.55°	7.96°	11.06°	9.60°	5.76°	10.22°	13.62°	11.21°

Tabla 4-28. Resultados del cálculo del error estándar para el acimut medio para la calibración '2005-04-21e' con tamaño de muestra de 10x10 elementos.

	1-10, 1-10	1-10, 5-14	1-10, 11-20	1-10, 15-24	1-10, 21-30	1-10, 25-34	1-10, 31-40	1-10, 35-44	1-10, 37-46
$\hat{\sigma}$	5.01°	4.61°	5.37°	9.89°	6.49°	6.97°	5.46°	4.24°	4.35°
	5-14, 1-10	5-14, 5-14	5-14, 11-20	5-14, 15-24	5-14, 21-30	5-14, 25-34	5-14, 31-40	5-14, 35-44	5-14, 37-46
$\hat{\sigma}$	6.22°	8.44°	14.70°	22.33°	8.13°	10.05°	11.14°	7.02°	6.87°
	11-20, 1-10	11-20, 5-14	11-20, 11-20	11-20, 15-24	11-20, 21-30	11-20, 25-34	11-20, 31-40	11-20, 35-44	11-20, 37-46
$\hat{\sigma}$	3.70°	5.04°	5.62°	3.83°	3.71°	5.67°	10.37°	9.65°	6.42°
	15-24, 1-10	15-24, 5-14	15-24, 11-20	15-24, 15-24	15-24, 21-30	15-24, 25-34	15-24, 31-40	15-24, 35-44	15-24, 37-46
$\hat{\sigma}$	4.82°	5.13°	6.17°	4.64°	4.69°	6.57°	9.37°	8.49°	6.18°
	21-30, 1-10	21-30, 5-14	21-30, 11-20	21-30, 15-24	21-30, 21-30	21-30, 25-34	21-30, 31-40	21-30, 35-44	21-30, 37-46
$\hat{\sigma}$	15.81°	20.90°	20.40°	51.55°	26.45°	68.96°	21.70°	32.17°	17.50°
	25-34, 1-10	25-34, 5-14	25-34, 11-20	25-34, 15-24	25-34, 21-30	25-34, 25-34	25-34, 31-40	25-34, 35-44	25-34, 37-46
$\hat{\sigma}$	7.01°	7.33°	6.67°	12.36°	12.82°	18.82°	21.11°	16.09°	11.04°
	31-40, 1-10	31-40, 5-14	31-40, 11-20	31-40, 15-24	31-40, 21-30	31-40, 25-34	31-40, 31-40	31-40, 35-44	31-40, 37-46
$\hat{\sigma}$	5.06°	5.41°	5.98°	9.08°	11.32°	13.10°	10.10°	8.66°	6.51°
	35-44, 1-10	35-44, 5-14	35-44, 11-20	35-44, 15-24	35-44, 21-30	35-44, 25-34	35-44, 31-40	35-44, 35-44	35-44, 37-46
$\hat{\sigma}$	4.71°	5.13°	7.41°	8.16°	8.21°	8.38°	5.78°	5.26°	4.45°
	37-46, 1-10	37-46, 5-14	37-46, 11-20	37-46, 15-24	37-46, 21-30	37-46, 25-34	37-46, 31-40	37-46, 35-44	37-46, 37-46
$\hat{\sigma}$	4.25°	4.66°	7.35°	8.67°	6.65°	8.36°	6.75°	5.54°	4.66°

Tabla 4-29. Resultado del cálculo del error estándar para la calibración '2005-04-25' con tamaño de 10x10.

- Bondad de ajuste a la distribución von Mises.

La bondad de ajuste a la distribución von Mises se mide según la probabilidad obtenida en la aplicación de las pruebas de Watson y Kuiper a nuestros conjuntos de datos. Estos dos tests, como sabemos, permiten inferir si los datos analizados han sido obtenidos a partir de una distribución von Mises.

Las pruebas se realizaron sobre las 6 calibraciones con tamaño de muestra de 10x10 elementos y sobre las ventanas de diferentes tamaños en la calibración '2005-02-21'. Comenzamos listando estos últimos resultados presentando en la tabla siguiente los resultados para un tamaño de muestra de 15x15 elementos.

	1-15,1-15	1-15,8-22	1-15,15-29	1-15,22-36	1-15,29-43	1-15,32-46
Watson	P< 0.005	P< 0.005	P< 0.005	P< 0.005	P< 0.005	P< 0.005
Kuiper	P< 0.01	P< 0.01	P< 0.01	P< 0.01	P< 0.01	P< 0.01
	8-22,1-15	8-22,8-22	8-22,15-29	8-22,22-36	8-22,29-43	8-22,32-46
Watson	P< 0.005	P< 0.005	P< 0.005	P< 0.005	P< 0.005	0.1>P>0.05
Kuiper	P< 0.01	P< 0.01	P< 0.01	P< 0.01	P< 0.05	P> 0.15
	15-29,1-15	15-29,8-22	15-29,15-29	15-29,22-36	15-29,29-43	15-29,32-46
Watson	P< 0.005	P< 0.05	0.1>P> 0.05	P< 0.005	P< 0.005	0.15>P>0.1
Kuiper	P< 0.01	P> 0.15	0.15>P>0.10	P< 0.025	P< 0.025	P> 0.15
	22-36,1-15	22-36,8-22	22-36,15-29	22-36,22-36	22-36,29-43	22-36,32-46
Watson	P< 0.005	P< 0.005	P> 0.5	P< 0.005	P< 0.005	P< 0.005
Kuiper	P< 0.01	P< 0.025	P> 0.15	P< 0.05	0.15>P>0.1	P> 0.15
	29-43,1-15	29-43,8-22	29-43,15-29	29-43,22-36	29-43,29-43	29-43,32-46
Watson	0.5>P>0.25	P< 0.005	P< 0.005	P< 0.005	P< 0.005	P< 0.005
Kuiper	P> 0.15	0.15>P>0.1	P< 0.01	P< 0.01	P< 0.01	P< 0.01
	32-46,1-15	32-46,8-22	32-46,15-29	32-46,22-36	32-46,29-43	32-46,32-46
Watson	P< 0.005	P< 0.005	P< 0.005	P< 0.005	P< 0.005	P< 0.005
Kuiper	0.15>P>0.1	P< 0.01	P< 0.01	P< 0.01	P< 0.01	P< 0.01

Tabla 4-30. Resultados del cálculo del ajuste a la distribución von Mises en la calibración '2005-02-21' con tamaño de muestra de 15x15 elementos.

Los resultados de la tabla anterior, al igual que ocurre con la de tamaño de ventana de 10x10 que se indican más adelante, son bastante dispares. No sucede lo mismo en el resto de los tamaños de ventanas (20x20, 25x25, 30x30 y 35x35) donde todas las muestras tienen el mismo resultado, como se indica a continuación.

Asimismo los datos que se obtuvieron para la muestra global (Tabla 4-10) eran similares a los obtenidos para tamaño de muestra entre 20x20 y 35x35.

2005-02-21	Watson	Kuiper
20x20	P<0.005	P<0.01
25x25	P<0.005	P<0.01
30x30	P<0.005	P<0.01
35x35	P<0.005	P<0.01
Global	P<0.005	P<0.01

Tabla 4-31. Resultados del cálculo del ajuste a la distribución von Mises en la calibración '2005-02-21' con tamaños de muestra de 20x20, 25x25, 30x30 elementos y con la muestra total.

Las tablas siguientes listan los resultados para 2 de las 6 calibraciones con tamaño de muestra de 10x10 elementos, en concreto para los ficheros '2005-02-21' y '2005-04-18'. El resto de resultados pueden consultarse como ya se ha dicho en el CD adjunto a esta Tesis.

Los resultados de análisis con tamaño de muestra de 10x10 elementos son tan dispares como los indicados para un tamaño de 15x15. Son muchas las submuestras que evidencian ajustarse a la distribución von Mises, dados los resultados de probabilidad obtenidos. Por lo que vemos que conforme aumenta el tamaño de muestra se diluye la existencia de una dirección preferente en los datos.

	1-10, 1-10	1-10, 5-14	1-10, 11-20	1-10, 15-24	1-10, 21-30	1-10, 25-34	1-10, 31-40	1-10, 35-44	1-10, 37-46
W	P< 0.005	0.1 > P > 0.05	P> 0.5	0.5>P> 0.25	P< 0.01	P< 0.05	P> 0.5	P< 0.025	P< 0.025
K	0.1>P>0.05	P> 0.15	P> 0.15	P> 0.15	P> 0.15	0.15> P> 0.1	P> 0.15	0.15 > P > 0.10	0.15 >P> 0.1
	5-14, 1-10	5-14, 5-14	5-14, 11-20	5-14, 15-24	5-14, 21-30	5-14, 25-34	5-14, 31-40	5-14, 35-44	5-14, 37-46
W	P< 0.005	P< 0.005	P< 0.005	P< 0.005	P< 0.005	P< 0.005	P< 0.005	P< 0.025	0.15> P> 0.1
K	0.1>P > 0.05	P< 0.01	P< 0.01	P< 0.01	P< 0.05	P< 0.01	P> 0.15	P> 0.15	P> 0.15
	11-20, 1-10	11-20, 5-14	11-20, 11-20	11-20, 15-24	11-20, 21-30	11-20, 25-34	11-20,31-40	11-20, 35-44	11-20, 37-46
W	P< 0.005	P< 0.005	P< 0.025	P< 0.01	P< 0.005	0.1> P> 0.05	P> 0.5	0.15 > P > 0.1	P> 0.5
K	0.1>P> 0.05	P< 0.01	0.15>P> 0.1	0.10>P>0.05	P< 0.01	P> 0.15	P> 0.15	P> 0.15	P> 0.15
	15-24, 1-10	15-24, 5-14	15-24, 11-20	15-24, 15-24	15-24, 21-30	15-24, 25-34	15-24,31-40	15-24, 35-44	15-24, 37-46
W	P< 0.005	P< 0.005	P> 0.5	P> 0.5	0.1>P > 0.05	P< 0.005	0.5 >P>0.25	P< 0.01	0.25>P>0.15
K	P< 0.05	0.10> P > 0.05	P> 0.15	P> 0.15	0.15>P>0.10	P< 0.05	P> 0.15	0.10 > P > 0.05	P> 0.15
	21-30, 1-10	21-30, 5-14	21-30, 11-20	21-30, 15-24	21-30, 21-30	21-30, 25-34	21-30,31-40	21-30, 35-44	21-30, 37-46
W	P< 0.05	P< 0.01	0.5>P> 0.25	0.5>P > 0.25	P> 0.5	P< 0.025	0.15>P> 0.1	P< 0.005	P< 0.025
K	P> 0.15	0.15 > P > 0.1	P> 0.15	P> 0.15	P> 0.15	0.15 >P> 0.1	P> 0.15	P> 0.15	P> 0.15
	25-34, 1-10	25-34, 5-14	25-34, 11-20	25-34, 15-24	25-34, 21-30	25-34, 25-34	25-34,31-40	25-34, 35-44	25-34, 37-46
W	P< 0.005	P< 0.005	0.5>P > 0.25	P> 0.5	0.5>P > 0.25	P< 0.01	P< 0.005	P< 0.005	P< 0.005
K	P< 0.025	P> 0.15	P> 0.15	P> 0.15	P> 0.15	P< 0.025	0.15>P> 0.1	0.15 > P > 0.10	P> 0.15
	31-40, 1-10	31-40, 5-14	31-40, 11-20	31-40, 15-24	31-40, 21-30	31-40, 25-34	31-40,31-40	31-40, 35-44	31-40, 37-46
W	0.5>P > 0.25	0.15 > P > 0.1	0.1 >P> 0.05	0.15>P > 0.1	P< 0.005	P< 0.005	0.5>P> 0.25	P< 0.025	P< 0.005
K	P> 0.15	P> 0.15	P> 0.15	P> 0.15	P< 0.025	0.1>P > 0.05	P> 0.15	P> 0.15	P> 0.15
	35-44, 1-10	35-44, 5-14	35-44, 11-20	35-44, 15-24	35-44, 21-30	35-44, 25-34	35-44,31-40	35-44, 35-44	35-44, 37-46
W	P> 0.5	P< 0.025	P< 0.05	P< 0.025	P< 0.005	P< 0.005	P< 0.005	P< 0.005	P< 0.005
K	P> 0.15	P> 0.15	P> 0.15	P> 0.15	P< 0.01	P< 0.01	P< 0.01	P< 0.01	P< 0.01
	37-46, 1-10	37-46, 5-14	37-46, 11-20	37-46, 15-24	37-46, 21-30	37-46, 25-34	37-46,31-40	37-46, 35-44	37-46, 37-46
W	P< 0.05	P< 0.005	P< 0.01	P< 0.025	P< 0.005	P< 0.005	P< 0.005	P< 0.005	P< 0.005
K	P> 0.15	P< 0.01	P> 0.15	P> 0.15	P< 0.01	P< 0.01	P< 0.01	P< 0.01	P< 0.01

Tabla 4-32. Resultados del cálculo del ajuste a la distribución von Mises con tamaño de muestra de 10x10 elementos para la calibración 2005-02-21.

	1-10, 1-10	1-10, 5-14	1-10, 11-20	1-10, 15-24	1-10, 21-30	1-10, 25-34	1-10, 31-40	1-10, 35-44	1-10, 37-46
W	P< 0.005	P< 0.005	P< 0.025	P< 0.05	P> 0.5	0.5 >P> 0.25	P< 0.05	0.1 > P > 0.05	< 0.005
K	P< 0.01	P< 0.01	P> 0.15	P> 0.15	P> 0.15	P> 0.15	P> 0.15	P> 0.15	< 0.05
	5-14, 1-10	5-14, 5-14	5-14, 11-20	5-14, 15-24	5-14, 21-30	5-14, 25-34	5-14, 31-40	5-14, 35-44	5-14, 37-46
W	P< 0.005	P< 0.005	P< 0.005	P< 0.005	P> 0.5	0.5>P > 0.25	P< 0.005	P< 0.005	< 0.01
K	P< 0.01	P< 0.01	P< 0.05	P> 0.15	P> 0.15	P> 0.15	P< 0.01	P< 0.025	0.10>P>0.05
	11-20, 1-10	11-20, 5-14	11-20, 11-20	11-20, 15-24	11-20, 21-30	11-20, 25-34	11-20, 31-40	11-20, 35-44	11-20, 37-46
W	P< 0.005	P< 0.005	0.1 > P > 0.05	P< 0.005	P> 0.5	P> 0.5	P< 0.005	P< 0.005	< 0.005
K	P> 0.15	0.10>P>0.05	P> 0.15	P< 0.05	P> 0.15	P> 0.15	0.15>P>0.10	P< 0.025	< 0.01
	15-24, 1-10	15-24, 5-14	15-24, 11-20	15-24, 15-24	15-24, 21-30	15-24, 25-34	15-24, 31-40	15-24, 35-44	15-24, 37-46
W	P< 0.01	P< 0.005	P> 0.5	P< 0.01	P> 0.5	0.1>P > 0.05	P< 0.005	P< 0.005	< 0.005
K	P> 0.15	P< 0.025	P> 0.15	0.10>P>0.05	P> 0.15	P> 0.15	0.15>P>0.10	0.15>P>0.10	< 0.025
	21-30, 1-10	21-30, 5-14	21-30, 11-20	21-30, 15-24	21-30, 21-30	21-30, 25-34	21-30, 31-40	21-30, 35-44	21-30, 37-46
W	0.25>P>0.15	P< 0.005	P< 0.05	P< 0.05	P> 0.5	0.1>P > 0.05	P< 0.025	P< 0.005	< 0.005
K	P> 0.15	P< 0.025	P> 0.15	P> 0.15	P> 0.15	P> 0.15	P> 0.15	P> 0.15	< 0.01
	25-34, 1-10	25-34, 5-14	25-34, 11-20	25-34, 15-24	25-34, 21-30	25-34, 25-34	25-34, 31-40	25-34, 35-44	25-34, 37-46
W	0.5>P > 0.25	P< 0.01	P< 0.005	P< 0.005	0.5>P > 0.25	0.15>P > 0.1	0.1 > P > 0.05	P< 0.005	< 0.005
K	P> 0.15	0.10>P>0.05	P< 0.01	0.10>P>0.05	P> 0.15	P> 0.15	P> 0.15	P> 0.15	0.10 > P > 0.05
	31-40, 1-10	31-40, 5-14	31-40, 11-20	31-40, 15-24	31-40, 21-30	31-40, 25-34	31-40, 31-40	31-40, 35-44	31-40, 37-46
W	0.25>P>0.15	0.1>P>0.05	P< 0.005	P< 0.005	0.25>P>0.15	P> 0.5	P< 0.025	P< 0.005	< 0.005
K	P> 0.15	P> 0.15	P< 0.05	P> 0.15	P> 0.15	P> 0.15	P> 0.15	0.10 > P > 0.05	< 0.01
	35-44, 1-10	35-44, 5-14	35-44, 11-20	35-44, 15-24	35-44, 21-30	35-44, 25-34	35-44, 31-40	35-44, 35-44	35-44, 37-46
W	0.15>P > 0.1	0.5>P>0.25	P< 0.05	P< 0.005	P< 0.005	P> 0.5	0.1 > P > 0.05	P< 0.005	< 0.005
K	P> 0.15	P> 0.15	P> 0.15	P> 0.15	0.10>P>0.05	P> 0.15	P> 0.15	P< 0.01	< 0.01
	37-46, 1-10	37-46, 5-14	37-46, 11-20	37-46, 15-24	37-46, 21-30	37-46, 25-34	37-46, 31-40	37-46, 35-44	37-46, 37-46
W	0.15>P > 0.1	0.5>P>0.25	P< 0.05	P< 0.005	P< 0.005	P> 0.5	0.1 > P > 0.05	P< 0.005	< 0.005
K	P> 0.15	P> 0.15	P> 0.15	P> 0.15	0.10>P>0.05	P> 0.15	P> 0.15	P< 0.01	< 0.01

Tabla 4-33. Resultados del cálculo del ajuste a la distribución von Mises con tamaño de muestra de 10x10 elementos para la calibración 2005-04-18.

Vemos en las tablas anteriores que aunque en la mayoría de las submuestras no se puede hablar de un ajuste razonable a la distribución von Mises, en un porcentaje no desdeñable de datos sí se adaptan a la citada distribución.

- Gráficos de distribución de datos circulares.

Ya se indicó la utilidad de los gráficos de distribución de datos como complemento a las pruebas de determinación de las características de las distribuciones. En la figura siguiente se representa la distribución de los datos circulares para la primera ventana de cada una de las submuestras realizadas a la calibración '2005-02-21' con el tamaño de muestra de 10x10, 15x15, 20x20, 25x25, 30x30 y 35x35, en los apartados a) a f) respectivamente.

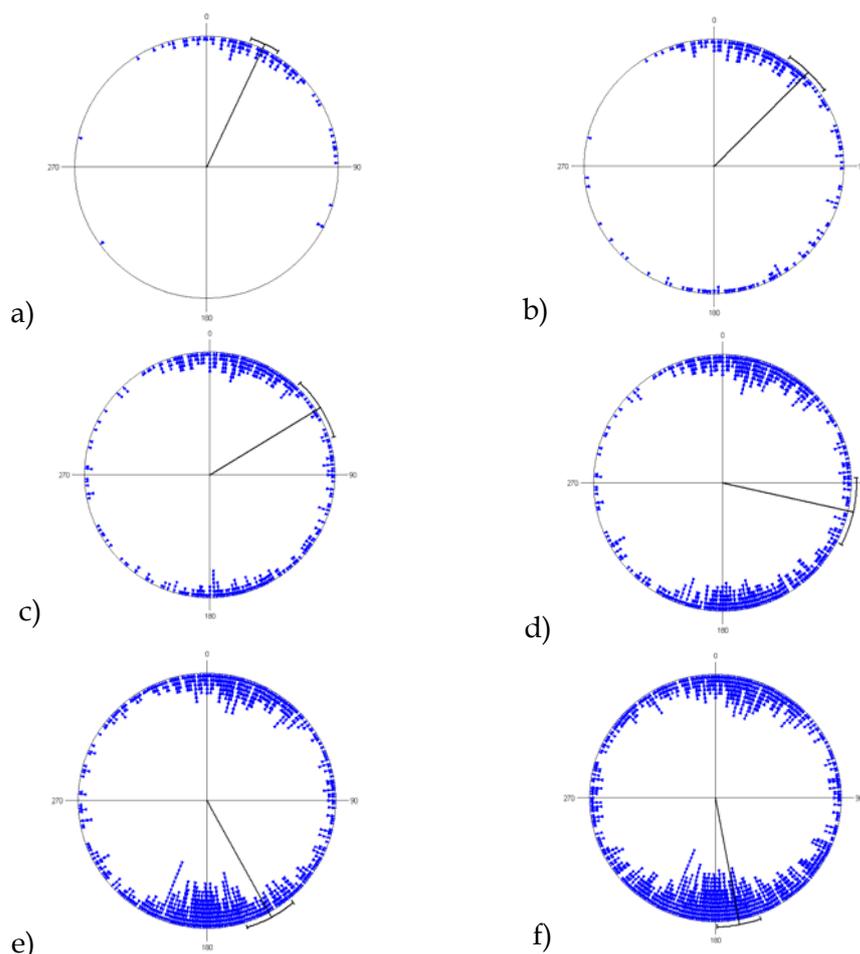


Fig. 4-7. Distribución de datos circulares en diferentes tamaños de ventanas. Todos los gráficos pertenecen a la calibración '2005-02-21' correspondiente a la primera ventana de: a) el análisis de 10x10 elementos, b) el de 15x15 elementos, c) el de 20x20, d) el de 25x25, e) el de 30x30 y f) el de 35x35. Se aprecia un creciente aumento de la bimodalidad junto con el valor del acimut medio.

Es visible en los gráficos anteriores el aumento de la bimodalidad, así como del valor del acimut medio, conforme aumenta el tamaño de la muestra.

Hasta el momento se han listado los resultados del análisis de los datos circulares, sólo acimutes, considerando vectores unitarios. Estos análisis se han centrado en calcular: a) estadísticos básicos (de tendencia central y dispersión), b) pruebas de determinación de las características de las distribuciones de datos (ajustes a determinadas distribuciones y control de la anisotropía) y c) gráficos de distribución de datos circulares.

Los datos objeto de análisis son, como sabemos, 20 muestras de 2116 vectores y 2 de 361 elementos. También 6 de las 20 muestras se analizaron con tamaño de submuestra de 100 elementos y una de éstas 6, se trató con tamaños de muestras variables.

De los resultados se infiere que las muestras de error no son aleatorias siendo esta afirmación más obvia en el análisis por ventanas. También se constata que los tests generales no detectan los errores locales.

4.1.4.2 Resultados del análisis de datos lineales

El acceso a los datos utilizados es el mismo que el indicado para datos circulares. El camino a seguir, por tanto, es:

D://Tesis/Datos/Escáner/nombre_fichero

Para mayor comodidad, los datos lineales están resumidos en ficheros de hojas de cálculo dentro de la siguiente dirección:

D://Tesis/Datos/Escáner/Resumen

a) Resultados de los análisis generales

- Cálculos de los estadísticos básicos para módulos.

En este apartado se incluirán los resultados de la media aritmética (\bar{d}), la desviación típica (s) y los valores máximos y mínimos de los módulos de las 20 calibraciones realizadas con 2116 vectores y las 2 de 361 elementos, que recordemos, seguimos imprimiendo en cursiva para diferenciarlas del resto.

Calibración	\bar{d}	s	$d \max$	$d \min$
2004-10-19	1.927	0.990	5.303	0.042
2005-02-21	2.933	1.550	9.072	0.103
2005-03-01	2.096	1.200	9.903	0.026
2005-03-07	2.624	1.556	9.638	0.05
2005-03-14	2.669	1.657	11.996	0.064
2005-04-04	2.630	1.489	9.808	0.028
2005-04-11	2.686	1.764	10.114	0.093
2005-04-18	2.385	1.329	8.748	0.081
2005-04-21e	2.737	1.995	18.919	0.036
2005-04-22e	3.118	1.948	12.576	0.066
2005-04-25	1.907	0.981	6.317	0.038
2005-05-03e	3.516	1.988	11.224	0.059
2005-05-18e	3.051	1.660	11.652	0.126
2005-05-23e	2.374	1.706	36.099	0.027
2005-06-09	1.797	0.940	5.542	0.019
2005-06-29	1.760	0.944	6.186	0.031
2005-07-05e	2.951	1.907	22.206	0.052
2005-07-11e	2.272	1.184	6.972	0.077
2005-07-12e	2.148	1.187	8.119	0.015
2005-07-13e	2.125	1.158	6.864	0.060
2006-05-19	2.284	1.586	8.907	0.159
2006-06-28	1.975	1.226	8.256	0.032

Tabla 4-34. Resultados de los estadísticos básicos para módulos en el análisis general.

Cabe destacar de la tabla anterior, en la columna de valores máximos para el módulo, que las cifras más altas la alcanzan, en general, aquellas calibraciones en las que se les ha introducido error artificialmente.

- Determinación de *outliers*.

En principio y para las pruebas, estableceremos como observación discordante aquella que sobrepase el valor de la media aritmética más 3 veces la desviación típica. En la tabla que se lista a continuación, aparecen para cada fichero de calibración, el valor de la media aritmética, la desviación típica y el número de módulos que consideramos, según este criterio, observaciones discordantes. Las dos últimas calibraciones se corresponden con los ficheros de 361 elementos. En la última columna se indica el porcentaje de puntos considerados discordantes con respecto al total.

Calibración	\bar{d}	s	n° outlier	%
2005-02-21	2.933	1.550	12	0.6
2005-03-01	2.096	1.200	22	1.0
2005-03-07	2.624	1.556	43	2.0
2005-03-14	2.669	1.657	27	1.3
2005-04-04	2.630	1.489	16	0.8
2005-04-11	2.686	1.764	35	1.7
2005-04-18	2.385	1.329	19	0.9
2005-04-21e	2.737	1.995	53	2.5
2005-04-22e	3.118	1.948	28	1.3
2005-04-25	1.907	0.981	9	0.4
2005-05-03e	3.516	1.988	17	0.8
2005-05-18e	3.051	1.660	10	0.5
2005-05-23e	2.374	1.706	30	1.4
2005-06-09	1.797	0.940	14	0.7
2005-06-29	1.760	0.944	15	0.7
2005-07-05e	2.951	1.907	20	0.9
2005-07-11e	2.272	1.184	14	0.7
2005-07-12e	2.148	1.187	21	1.0
2005-07-13e	2.125	1.158	10	0.5
2006-06-28c	1.975	1.226	49	2.3
2004-10-19	1.927	0.990	3	0.1
2006-05-19c	2.284	1.586	14	0.7

Tabla 4-35. Número de *outliers* resultantes en cada fichero de calibración en análisis generales de datos lineales, considerando como límite la media más el triple de la desviación típica.

Se observa en esta serie de datos que el mayor porcentaje corresponde al 2.5 % de la calibración '2005-04-21e', afectada de error, con un total de 53 observaciones discordantes. El menor porcentaje, con un 0.1% que equivale a 3 *outliers*, lo presenta la calibración '2004-10-19', que recordemos, contaba sólo con 361 observaciones.

Si establecemos ahora el límite para considerar un *outliers* el valor de la media aritmética más 5 veces la desviación típica tenemos los resultados de la tabla siguiente:

Calibración	\bar{d}	s	n° outlier
2005-03-01	2.096	1.200	2
2005-03-14	2.669	1.657	4
2005-04-21e	2.737	1.995	1
2005-05-23e	2.374	1.706	2
2005-07-05e	2.951	1.907	1
2005-07-12e	2.148	1.187	1
2006-06-28c	1.975	1.226	1

Tabla 4-36. Número de *outliers* resultantes en cada fichero de calibración en análisis generales de datos lineales, considerando como límite la media más 5 veces la desviación típica.

Vemos que no todos los ficheros que tienen error inducido presentan observaciones discordantes y, a la inversa, el que mayor *outliers* tiene, 4, no estaba afectado de error extra.

- Pruebas para la determinación de las características de las distribuciones de los datos lineales.

El control para la determinación de las características de las distribuciones de datos lineales se centra en la aplicación de pruebas de bondad de ajuste (Kolmogorov-Smirnov) a diversas distribuciones de datos. Estas distribuciones han sido principalmente la gamma y la distribución Weibull (un caso particular de la anterior), distribuciones asimétricas positivas, que se han adaptado bastante bien a nuestros datos.

Todas las muestras se ajustaban a una de estas dos distribuciones, excepto las correspondientes a los ficheros '2005-04-21e' y '2006-06-28c'. También se utilizó la distribución normal y la logarítmico-normal como contrastes, aunque los resultados de estas pruebas se omiten en la tabla siguiente por obtener un p-valor igual a cero en todos los casos.

Fichero Calibración	p-valor gamma	p-valor Weibull	Fichero Calibración	p-valor gamma	p-valor Weibull
2004-10-19	0.56	0.68	2005-05-03e	0.31	0.13
2005-02-21	0.01	0.87	2005-05-18e	0.06	0.96
2005-03-01	0.05	0.09	2005-05-23e	0.11	0.00
2005-03-07	0.43	0.01	2005-06-09	0.02	0.69
2005-03-14	0.47	0.00	2005-06-29	0.01	0.82
2005-04-04	0.65	0.25	2005-07-05e	0.54	0.01
2005-04-11	0.78	0.01	2005-07-11e	0.01	0.89
2005-04-18	0.01	0.83	2005-07-12e	0.01	0.99
2005-04-21e	0.00	0.00	2005-07-13e	0.01	0.98
2005-04-22e	0.83	0.01	2006-05-19c	0.03	0.002
2005-04-25	0.00	0.94	2006-06-28c	0.00	0.00

Tabla 4-37. Resultados de la prueba K-S para la distribución gamma y Weibull a los datos lineales.

Observamos en la tabla anterior que la mayoría de las calibraciones se adaptan bien a alguna de las dos distribuciones indicadas. En alguna, en cambio, no ocurre así, como la calibración '2005-04-21e'.

- Histogramas de distribución de datos.

A modo de ejemplo se acompañan dos histogramas de frecuencias representativos de los datos lineales. El primero muestra la distribución de módulos de la calibración '2005-02-21', que será la que posteriormente se analice en submuestras de diferentes tamaños. Vemos una distribución asimétrica positiva en la que la mayor frecuencia de datos no se corresponde con los módulos más pequeños.

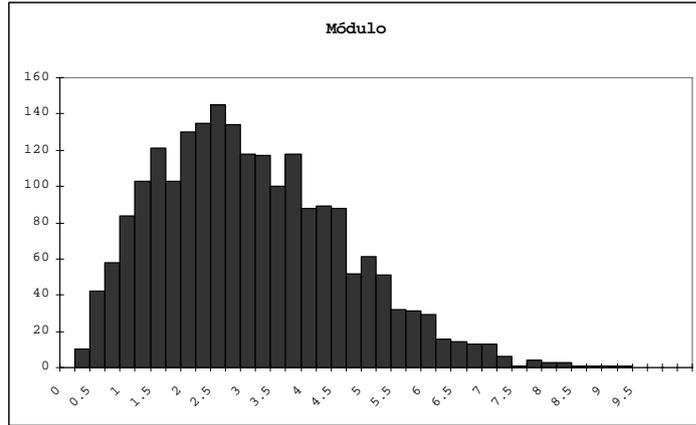


Fig. 4-8. Histograma de módulos no afectados de error. Distribución asimétrica positiva de la calibración '2005-02-21' donde los módulos más pequeños no son los más frecuentes.

El segundo ejemplo se corresponde con la calibración '2005-04-21e' en la que se introdujo un error de forma artificial, como sabemos. La distribución también es asimétrica positiva pero el número de observaciones que podemos catalogar de discordantes es mucho mayor, de las 12 observaciones en el primer caso se pasa a 53 en el segundo.

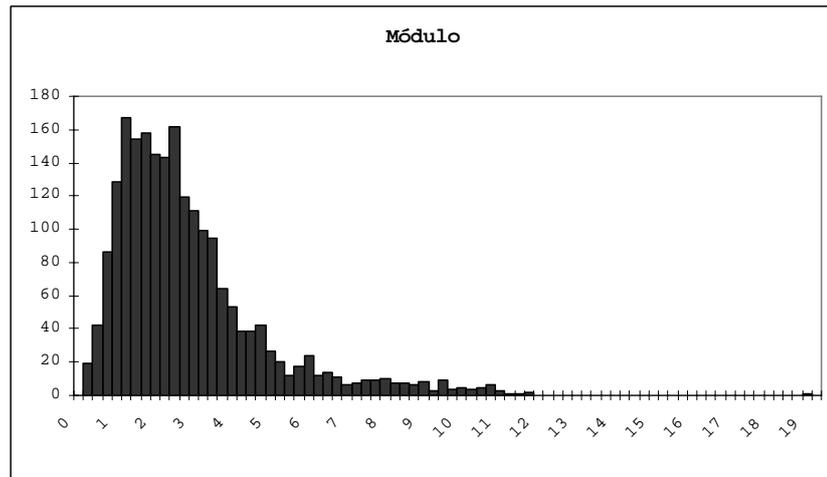


Fig. 4-9. Histograma de módulos afectados de error. Distribución asimétrica positiva donde se aprecia además una serie de observaciones discordantes motivadas por la inclusión de errores artificialmente introducidos.

b) Resultados de los análisis locales (ventanas)

- Cálculos de los estadísticos básicos para módulos.

En las tablas siguientes se indican los valores de los estadísticos básicos para los módulos (media aritmética (\bar{d}), desviación típica (s) y valores máximos y mínimos) para 2 de las 6 calibraciones en las que se analizan submuestras de 10x10 elementos.

	1-10, 1-10	1-10, 5-14	1-10, 11-20	1-10, 15-24	1-10, 21-30	1-10, 25-34	1-10, 31-40	1-10, 35-44	1-10, 37-46
\bar{d}	3.420	3.554	3.510	3.316	3.371	3.119	2.979	3.239	3.061
s	1.273	1.331	1.216	1.294	1.368	1.305	1.274	1.319	1.319
D _{max}	6.096	6.264	6.264	6.545	6.545	5.585	5.918	6.364	6.364
D _{min}	0.584	0.584	0.684	0.684	0.721	0.103	0.103	0.414	0.414
	5-14, 1-10	5-14, 5-14	5-14, 11-20	5-14, 15-24	5-14, 21-30	5-14, 25-34	5-14, 31-40	5-14, 35-44	5-14, 37-46
\bar{d}	2.816	2.965	2.807	2.687	2.922	2.869	2.639	2.664	2.540
s	1.339	1.559	1.525	1.435	1.389	1.285	1.247	1.254	1.214
D _{max}	6.096	6.264	6.264	6.545	6.545	5.585	5.585	6.364	6.364
D _{min}	0.577	0.529	0.181	0.181	0.237	0.103	0.103	0.414	0.399
	11-20, 1-10	11-20, 5-14	11-20, 11-20	11-20, 15-24	11-20, 21-30	11-20, 25-34	11-20, 31-40	11-20, 35-44	11-20, 37-46
\bar{d}	2.269	2.219	2.001	2.567	3.939	3.894	2.967	2.822	2.722
s	1.300	1.293	1.123	1.426	1.963	1.869	1.233	1.267	1.224
D _{max}	5.686	5.686	5.054	6.032	9.072	9.072	6.139	6.139	6.139
D _{min}	0.175	0.175	0.152	0.152	0.237	0.325	0.312	0.227	0.227
	15-24, 1-10	15-24, 5-14	15-24, 11-20	15-24, 15-24	15-24, 21-30	15-24, 25-34	15-24, 31-40	15-24, 35-44	15-24, 37-46
\bar{d}	2.180	2.332	2.279	2.876	4.579	4.340	3.103	3.238	3.180
s	1.263	1.272	1.183	1.351	1.683	1.846	1.262	1.304	1.275
D _{max}	5.263	5.263	5.054	6.032	9.072	9.072	6.139	6.18	6.18
D _{min}	0.175	0.175	0.152	0.152	1.399	1.347	0.312	0.227	0.227
	21-30, 1-10	21-30, 5-14	21-30, 11-20	21-30, 15-24	21-30, 21-30	21-30, 25-34	21-30, 31-40	21-30, 35-44	21-30, 37-46
\bar{d}	2.087	2.489	2.863	2.934	3.799	3.540	2.620	2.863	2.760
s	1.055	1.200	1.269	1.277	1.431	1.614	1.208	1.291	1.350
D _{max}	4.952	5.514	5.514	6.009	8.192	8.192	6.056	6.18	6.18
D _{min}	0.275	0.275	0.273	0.273	1.144	0.434	0.292	0.292	0.362
	25-34, 1-10	25-34, 5-14	25-34, 11-20	25-34, 15-24	25-34, 21-30	25-34, 25-34	25-34, 31-40	25-34, 35-44	25-34, 37-46
\bar{d}	2.074	2.346	2.874	2.983	3.514	3.184	2.137	2.092	2.010
s	0.918	1.114	1.233	1.163	1.261	1.459	1.092	1.059	1.096
D _{max}	4.521	5.514	5.514	6.009	6.684	6.684	4.912	5.068	5.068
D _{min}	0.275	0.275	0.273	0.273	1.144	0.434	0.292	0.292	0.362
	31-40, 1-10	31-40, 5-14	31-40, 11-20	31-40, 15-24	31-40, 21-30	31-40, 25-34	31-40, 31-40	31-40, 35-44	31-40, 37-46
\bar{d}	1.932	1.912	2.041	2.099	2.799	2.965	2.255	2.169	2.200
s	0.970	0.934	0.927	0.896	1.344	1.422	1.304	1.169	1.172
D _{max}	4.369	4.559	4.559	4.076	5.878	5.878	5.865	5.148	5.034
D _{min}	0.108	0.108	0.305	0.305	0.263	0.217	0.217	0.356	0.356
	35-44, 1-10	35-44, 5-14	35-44, 11-20	35-44, 15-24	35-44, 21-30	35-44, 25-34	35-44, 31-40	35-44, 35-44	35-44, 37-46
\bar{d}	1.882	1.791	1.639	1.864	3.420	4.066	3.711	3.563	3.572
s	1.043	0.959	0.753	0.975	1.949	1.939	1.969	1.834	1.773
D _{max}	4.369	4.014	3.291	5.756	8.149	8.149	7.97	7.97	7.582
D _{min}	0.108	0.108	0.275	0.305	0.263	0.217	0.217	0.356	0.356
	37-46, 1-10	37-46, 5-14	37-46, 11-20	37-46, 15-24	37-46, 21-30	37-46, 25-34	37-46, 31-40	37-46, 35-44	37-46, 37-46
\bar{d}	2.064	2.145	1.950	2.336	3.980	4.520	4.349	4.095	4.058
s	1.237	1.361	1.257	1.650	2.154	1.941	1.858	1.741	1.701
D _{max}	5.425	5.853	7.13	7.13	8.655	8.655	7.97	7.97	7.582
D _{min}	0.108	0.108	0.275	0.305	0.263	0.217	0.217	0.533	0.533

Tabla 4-38. Resultados del cálculo de los estadísticos básicos para módulos para la calibración '2005-02-21' con tamaño de muestra de 10x10 elementos.

	1-10, 1-10	1-10, 5-14	1-10, 11-20	1-10, 15-24	1-10, 21-30	1-10, 25-34	1-10, 31-40	1-10, 35-44	1-10, 37-46
\bar{d}	3.029	3.098	1.977	1.886	2.836	3.289	3.045	2.746	2.788
s	1.527	1.654	1.157	0.943	1.204	1.056	1.107	1.104	1.156
d _{max}	6.633	6.633	5.433	4.117	6.505	6.505	5.278	5.278	5.419
d _{min}	0.156	0.364	0.364	0.223	0.223	0.915	0.535	0.535	0.565
	5-14, 1-10	5-14, 5-14	5-14, 11-20	5-14, 15-24	5-14, 21-30	5-14, 25-34	5-14, 31-40	5-14, 35-44	5-14, 37-46
\bar{d}	2.849	3.136	2.052	1.705	2.529	2.929	3.008	3.220	3.171
s	1.698	1.675	1.160	0.884	1.336	1.245	1.259	1.209	1.289
d _{max}	6.633	6.633	5.433	4.434	6.505	6.505	6.351	6.351	6.351
d _{min}	0.156	0.251	0.445	0.171	0.171	0.102	0.102	0.535	0.565
	11-20, 1-10	11-20, 5-14	11-20, 11-20	11-20, 15-24	11-20, 21-30	11-20, 25-34	11-20, 31-40	11-20, 35-44	11-20, 37-46
\bar{d}	1.593	2.084	2.578	2.187	1.844	2.041	2.528	2.938	3.005
s	0.987	1.181	1.002	1.033	0.989	1.018	1.338	1.353	1.404
d _{max}	4.924	4.924	4.434	4.434	4.696	4.696	6.351	6.351	6.351
d _{min}	0.147	0.147	0.445	0.105	0.105	0.102	0.102	0.375	0.375
	15-24, 1-10	15-24, 5-14	15-24, 11-20	15-24, 15-24	15-24, 21-30	15-24, 25-34	15-24, 31-40	15-24, 35-44	15-24, 37-46
\bar{d}	1.497	2.297	3.222	2.606	1.734	1.957	2.380	2.435	2.601
s	0.886	1.268	0.952	1.118	0.872	1.001	1.175	1.205	1.281
d _{max}	5.318	5.318	5.564	5.564	4.657	4.657	5.496	5.496	5.496
d _{min}	0.147	0.147	0.518	0.105	0.081	0.081	0.121	0.375	0.375
	21-30, 1-10	21-30, 5-14	21-30, 11-20	21-30, 15-24	21-30, 21-30	21-30, 25-34	21-30, 31-40	21-30, 35-44	21-30, 37-46
\bar{d}	1.840	2.800	3.743	3.121	1.889	2.083	2.640	2.660	2.971
s	1.088	1.306	1.018	1.273	0.854	1.008	1.074	1.202	1.439
d _{max}	5.638	5.638	6.151	6.151	3.993	5.107	5.107	5.133	6.715
d _{min}	0.164	0.164	1.544	0.764	0.081	0.081	0.402	0.521	0.521
	25-34, 1-10	25-34, 5-14	25-34, 11-20	25-34, 15-24	25-34, 21-30	25-34, 25-34	25-34, 31-40	25-34, 35-44	25-34, 37-46
\bar{d}	1.869	2.481	3.169	2.786	1.877	2.042	2.431	2.408	2.662
s	0.982	1.205	1.249	1.336	0.915	0.963	0.983	1.099	1.443
d _{max}	5.638	5.638	6.151	6.151	4.27	5.107	5.107	5.133	6.715
d _{min}	0.164	0.164	0.736	0.197	0.19	0.19	0.142	0.142	0.142
	31-40, 1-10	31-40, 5-14	31-40, 11-20	31-40, 15-24	31-40, 21-30	31-40, 25-34	31-40, 31-40	31-40, 35-44	31-40, 37-46
\bar{d}	1.816	1.740	1.847	1.880	1.991	2.022	2.145	2.536	2.773
s	0.881	0.870	0.915	0.938	1.014	1.057	1.145	1.418	1.644
d _{max}	4.276	3.776	4.957	4.957	4.595	4.595	5.541	7.074	7.752
d _{min}	0.363	0.129	0.129	0.154	0.197	0.112	0.112	0.142	0.142
	35-44, 1-10	35-44, 5-14	35-44, 11-20	35-44, 15-24	35-44, 21-30	35-44, 25-34	35-44, 31-40	35-44, 35-44	35-44, 37-46
\bar{d}	1.825	1.655	1.530	1.762	2.317	2.207	2.188	3.159	3.708
s	0.905	0.807	0.755	0.856	1.033	1.180	1.290	1.778	2.047
d _{max}	4.276	3.776	4.957	4.957	4.807	4.807	5.541	7.264	8.748
d _{min}	0.144	0.129	0.129	0.154	0.379	0.112	0.112	0.635	0.675
	37-46, 1-10	37-46, 5-14	37-46, 11-20	37-46, 15-24	37-46, 21-30	37-46, 25-34	37-46, 31-40	37-46, 35-44	37-46, 37-46
\bar{d}	1.778	1.674	1.505	1.727	2.292	2.101	2.066	3.199	3.917
s	0.845	0.793	0.681	0.815	1.001	1.129	1.244	1.798	2.037
d _{max}	3.776	3.776	3.023	3.85	4.807	4.807	5.541	7.264	8.748
d _{min}	0.144	0.129	0.108	0.108	0.233	0.284	0.284	0.312	0.312

Tabla 4-39. Resultados del cálculo de los estadísticos básicos para módulos con tamaño de muestra de 10x10 elementos para la calibración '2005-04-18'

Observamos que los ficheros afectados de error presentan valores medios de media aritmética y desviación típica ligeramente más altos que el resto.

A continuación se resumen los mismos estadísticos anteriores pero aplicados a la calibración '2005-02-21' en muestras de distintos tamaños. No se aprecian variaciones significativas entre los diferentes estadísticos conforme aumenta el tamaño de la muestra.

	1-15, 1-15	1-15, 8-22	1-15, 15-29	1-15, 22-36	1-15, 29-43	1-15, 32-46
\bar{d}	3.111	3.068	2.937	3.027	2.965	2.797
S	1.376	1.486	1.359	1.250	1.297	1.264
d_{max}	6.264	6.545	6.545	5.918	6.364	6.364
d_{min}	0.529	0.181	0.181	0.103	0.103	0.103
	8-22, 1-15	8-22, 8-22	8-22, 15-29	8-22, 22-36	8-22, 29-43	8-22, 32-46
\bar{d}	2.192	2.392	3.096	3.396	2.963	2.622
S	1.234	1.257	1.829	1.730	1.392	1.196
d_{max}	5.917	5.917	9.072	9.072	6.844	6.18
d_{min}	0.152	0.152	0.152	0.103	0.103	0.103
	15-29, 1-15	15-29, 8-22	15-29, 15-29	15-29, 22-36	15-29, 29-43	15-29, 32-46
\bar{d}	2.228	2.731	3.600	3.696	3.160	2.666
S	1.196	1.261	1.782	1.730	1.355	1.218
d_{max}	5.514	5.514	9.072	9.072	6.844	6.18
d_{min}	0.152	0.152	0.152	0.292	0.292	0.227
	22-36, 1-15	22-36, 8-22	22-36, 15-29	22-36, 22-36	22-36, 29-43	22-36, 32-46
\bar{d}	2.378	2.691	3.171	3.074	2.567	2.205
S	1.068	1.137	1.322	1.393	1.284	1.204
d_{max}	5.514	5.514	7.078	7.078	6.18	6.18
d_{min}	0.275	0.273	0.273	0.292	0.292	0.292
	29-43, 1-15	29-43, 8-22	29-43, 15-29	29-43, 22-36	29-43, 29-43	29-43, 32-46
\bar{d}	1.892	2.115	2.931	3.337	2.926	2.707
S	1.007	1.150	1.624	1.699	1.681	1.651
d_{max}	5.514	5.514	8.149	8.149	7.97	7.97
d_{min}	0.108	0.275	0.263	0.217	0.217	0.217
	32-46, 1-15	32-46, 8-22	32-46, 15-29	32-46, 22-36	32-46, 29-43	32-46, 32-46
\bar{d}	2.126	2.227	3.027	3.753	3.518	3.369
S	1.163	1.343	1.850	1.897	1.892	1.902
d_{max}	5.853	7.13	8.655	8.655	7.97	7.97
d_{min}	0.108	0.275	0.263	0.217	0.217	0.217

Tabla 4-40. Resultados del cálculo de los estadísticos básicos para módulos con tamaño de muestra de 15x15 elementos para la calibración '2005-02-21'.

La tabla anterior se corresponde al tamaño de muestra de 15x15 elementos, ya que la de tamaño 10x10 se listó anteriormente. En la tabla siguiente se muestran los resultados correspondientes a las ventanas de 20x20 módulos.

	1-20, 1-20	1-20, 10-29	1-20, 21-40	1-20, 27-46
\bar{d}	2.798	3.203	3.314	3.165
s	1.399	1.619	1.535	1.478
d_{max}	6.264	9.072	9.072	9.072
d_{min}	0.152	0.152	0.103	0.103
	10-29, 1-20	10-29, 10-29	10-29, 21-40	10-29, 27-46
\bar{d}	2.273	3.040	3.264	3.092
s	1.218	1.687	1.605	1.606
d_{max}	5.686	9.072	9.072	9.072
d_{min}	0.152	0.152	0.237	0.227
	21-40, 1-20	21-40, 10-29	21-40, 21-40	21-40, 27-46
\bar{d}	2.231	2.848	2.868	2.765
s	1.122	1.410	1.438	1.468
d_{max}	5.514	8.192	8.192	8.192
d_{min}	0.108	0.263	0.217	0.217
	27-46, 1-20	27-46, 10-29	27-46, 21-40	27-46, 27-46
\bar{d}	2.245	3.018	3.517	3.373
s	1.193	1.699	1.850	1.891
d_{max}	7.13	8.655	8.655	8.655
d_{min}	0.108	0.263	0.217	0.217

Tabla 4-41. Resultados del cálculo de los estadísticos básicos para módulos con tamaño de muestra de 20x20 elementos para la calibración '2005-02-21'.

Los resultados de las muestras de 20x20 elementos siguen las mismas pautas anteriores. A continuación se presentan las muestras de tamaño 25x25.

	1-25,1-25	1-25,12-36	1-25,22-46
\bar{d}	2.798	3.143	3.236
s	1.388	1.506	1.464
d_{max}	6.545	9.072	9.072
d_{min}	0.152	0.103	0.103
	12-36,1-25	12-36,12-36	12-36,22-46
\bar{d}	2.490	3.043	3.064
s	1.211	1.489	1.523
d_{max}	6.032	9.072	9.072
d_{min}	0.152	0.152	0.227
	22-46,1-25	22-46,12-36	22-46,22-46
\bar{d}	2.425	3.108	3.330
s	1.287	1.671	1.749
d_{max}	7.130	8.655	8.655
d_{min}	0.108	0.217	0.217

Tabla 4-42. Resultados del cálculo de los estadísticos básicos para módulos con tamaño de muestra de 25x25 elementos para la calibración '2005-02-21'.

Se finaliza este apartado con las tablas de tamaño 30x30 y 35x35.

	1-30,1-30	1-30,17-46
\bar{d}	3.028	3.123
S	1.524	1.463
d_{max}	9.072	9.072
d_{min}	0.152	0.103
	17-46,1-30	17-46,17-46
\bar{d}	2.819	3.320
S	1.634	1.723
d_{max}	9.072	9.072
d_{min}	0.108	0.217

Tabla 4-43. Resultados del cálculo de los estadísticos básicos para módulos con tamaño de muestra de 30x30 elementos para la calibración '2005-02-21'.

La última tabla, para tamaño de muestra de 35x35 elementos es similar a la anterior.

	1-35, 1-35	1-35, 12-46
\bar{d}	2.941	2.990
S	1.439	1.426
d_{max}	9.072	9.072
d_{min}	0.103	0.103
	12-46, 1-35	12-46, 12-46
\bar{d}	2.825	3.093
S	1.599	1.652
d_{max}	9.072	9.072
d_{min}	0.108	0.152

Tabla 4-44. Resultados del cálculo de los estadísticos básicos para módulos con tamaño de muestra de 35x35 elementos para la calibración '2005-02-21'.

En general, el análisis entre los diferentes tamaños de ventanas no ofrece grandes discrepancias en estos estadísticos.

- Determinación de *outliers*.

Análogamente a como se hizo en el análisis general, consideramos *outliers* en datos lineales a aquellos módulos que sobrepasen la media aritmética más 3 veces el valor de la desviación típica. En esta primera tabla se indican los resultados del cálculo de *outliers* en los 6 ficheros en los que se analizaron las 81 ventanas de tamaño de 10x10 elementos, de tal forma que bajo el título de 'Totales', aparece el número de observaciones discordantes encontradas en todas las submuestras. Cabe decir que como las submuestras aparecen con solapes mínimos del 50% es seguro que un mismo *outlier* se contabilice en análisis contiguos. Para complementar esta eventualidad, el dato de 'Porcentaje' se corresponde con el tanto por ciento de observaciones discordantes por número de valores analizados, esto es 8100 (81 ventanas de 100 elementos).

2005-02-21 (10x10) n=81		2005-04-18 (10x10) n=81	
Totales	Porcentaje	Totales	Porcentaje
4	0.05%	15	0.18%
2005-04-21e (10x10) n=81		2005-04-25 (10x10) n=81	
Totales	Porcentaje	Totales	Porcentaje
22	0.27%	25	0.31%
2005-05-03e (10x10) n=81		2006-06-28c (10x10) n=81	
Totales	Porcentaje	Totales	Porcentaje
15	0.18%	17	0.21%

Tabla 4-45. Resultados del cálculo de *outliers* en datos lineales para los 6 ficheros de calibración con submuestras de 100 elementos.

En esta segunda tabla y con las mismas consideraciones anteriores se anotan los resultados correspondientes al cálculo de *outliers* en ventanas de diferentes tamaños en el fichero de calibración '2005-02-21'. El cálculo del porcentaje se va ajustando al tamaño de la ventana.

2005-02-21 (10x10) n=81		2005-02-21 (15x15) n=36	
Totales	Porcentaje	Totales	Porcentaje
4	0.05%	20	0.24%
2005-02-21 (20x20) n=16		2005-02-21 (25x25) n=9	
Totales	Porcentaje	Totales	Porcentaje
33	0.5%	34	0.6%
2005-02-21 (30x30) n=4		2005-02-21 (35x35) n=4	
Totales	Porcentaje	Totales	Porcentaje
25	0.69%	28	0.57%

Tabla 4-46. Resultados del cálculo de *outliers* en datos lineales para los ficheros con submuestras de diferente tamaño en el fichero '2005-02-21'.

- Pruebas para la determinación de las características de las distribuciones de los datos lineales.

La aplicación de pruebas de control de ajuste como la de Kolmogorov-Smirnov a todo el conjunto de datos nos permite asignar una distribución como más probable a nuestras observaciones. Las distribuciones objeto de contraste son, como sabemos, la normal, la lognormal, la gamma y la Weibull. Esta última, aunque un caso particular de la distribución gamma, proporcionó buenos ajustes en algunos de los análisis generales. De igual forma que en párrafos precedentes indicamos en primer lugar los resultados obtenidos para 2 de las 6 calibraciones con muestras de 10x10 elementos y posteriormente el análisis en muestras de diferente tamaño en la calibración '2005-02-21'. Lo que se muestra en las tablas siguientes es el p-valor del ajuste a las citadas distribuciones.

	1-10, 1-10	1-10, 5-14	1-10, 11-20	1-10, 15-24	1-10, 21-30	1-10, 25-34	1-10, 31-40	1-10, 35-44	1-10, 37-46
G	0.12	0.25	0.03	0.03	0.40	0.73	0.30	0.01	0.04
W	0.67	0.96	0.29	0.21	0.97	0.86	0.93	0.21	0.40
N	0.64	0.99	0.31	0.32	0.92	0.76	0.99	0.51	0.83
Ln	0.02	0.05	0.01	0.01	0.14	0.17	0.04	0.00	0.01
	5-14, 1-10	5-14, 5-14	5-14, 11-20	5-14, 15-24	5-14, 21-30	5-14, 25-34	5-14, 31-40	5-14, 35-44	5-14, 37-46
G	0.45	0.39	0.34	0.67	0.55	0.51	0.44	0.26	0.35
W	0.95	0.57	0.41	0.95	0.66	0.87	0.63	0.59	0.87
N	0.86	0.33	0.38	0.34	0.20	0.69	0.66	0.78	0.92
Ln	0.16	0.34	0.25	0.16	0.31	0.17	0.06	0.06	0.08
	11-20, 1-10	11-20, 5-14	11-20, 11-20	11-20, 15-24	11-20, 21-30	11-20, 25-34	11-20, 31-40	11-20, 35-44	11-20, 37-46
G	0.69	0.89	0.69	0.30	0.15	0.81	0.46	0.34	0.42
W	0.96	0.85	0.88	0.77	0.53	0.92	0.95	0.51	0.93
N	0.34	0.15	0.30	0.81	0.40	0.46	0.92	0.73	0.55
Ln	0.15	0.83	0.46	0.05	0.04	0.37	0.10	0.16	0.11
	15-24, 1-10	15-24, 5-14	15-24, 11-20	15-24, 15-24	15-24, 21-30	15-24, 25-34	15-24, 31-40	15-24, 35-44	15-24, 37-46
G	0.88	0.75	0.32	0.16	0.96	0.84	0.88	0.28	0.73
W	0.69	0.91	0.57	0.71	0.66	0.95	0.61	0.83	0.96
N	0.31	0.49	0.51	0.85	0.59	0.78	0.29	0.98	0.66
Ln	0.33	0.38	0.25	0.03	0.61	0.41	0.48	0.15	0.27
	21-30, 1-10	21-30, 5-14	21-30, 11-20	21-30, 15-24	21-30, 21-30	21-30, 25-34	21-30, 31-40	21-30, 35-44	21-30, 37-46
G	0.43	0.59	0.39	0.29	1.00	0.85	0.95	0.68	0.81
W	0.93	0.99	0.81	0.80	0.79	1.00	0.45	0.81	0.98
N	0.75	0.75	0.71	0.46	0.78	0.88	0.13	0.37	0.48
Ln	0.10	0.15	0.15	0.09	0.96	0.42	0.63	0.38	0.55
	25-34, 1-10	25-34, 5-14	25-34, 11-20	25-34, 15-24	25-34, 21-30	25-34, 25-34	25-34, 31-40	25-34, 35-44	25-34, 37-46
G	0.03	0.45	0.67	0.77	0.99	0.58	0.82	0.86	0.89
W	0.21	0.78	0.98	0.97	0.51	0.99	0.87	0.90	0.76
N	0.60	0.33	0.92	0.71	0.43	0.93	0.53	0.40	0.17
Ln	0.00	0.10	0.25	0.35	1.00	0.15	0.58	0.74	0.92
	31-40, 1-10	31-40, 5-14	31-40, 11-20	31-40, 15-24	31-40, 21-30	31-40, 25-34	31-40, 31-40	31-40, 35-44	31-40, 37-46
G	0.08	0.37	0.62	0.30	0.60	0.23	0.38	0.90	0.88
W	0.33	0.89	0.91	0.83	0.99	0.70	0.70	0.98	0.86
N	0.92	0.84	0.87	0.98	0.99	0.62	0.13	0.38	0.26
Ln	0.02	0.11	0.28	0.08	0.16	0.04	0.18	0.49	0.72
	35-44, 1-10	35-44, 5-14	35-44, 11-20	35-44, 15-24	35-44, 21-30	35-44, 25-34	35-44, 31-40	35-44, 35-44	35-44, 37-46
G	0.45	0.65	0.57	0.89	0.69	0.05	0.25	0.54	0.63
W	0.60	0.94	0.98	0.93	0.80	0.38	0.55	0.90	0.91
N	0.51	0.54	0.72	0.31	0.16	0.48	0.59	0.86	0.74
Ln	0.10	0.12	0.15	0.68	0.41	0.01	0.11	0.29	0.18
	37-46, 1-10	37-46, 5-14	37-46, 11-20	37-46, 15-24	37-46, 21-30	37-46, 25-34	37-46, 31-40	37-46, 35-44	37-46, 37-46
G	0.35	0.83	0.44	0.52	0.03	0.00	0.05	0.45	0.67
W	0.62	0.91	0.18	0.43	0.07	0.03	0.40	0.99	0.92
N	0.37	0.09	0.01	0.02	0.24	0.12	0.54	0.94	0.83
Ln	0.19	0.46	0.58	0.93	0.02	0.00	0.01	0.15	0.29

Tabla 4-47. Resultados del cálculo de la aplicación del test K-S para diferentes distribuciones en la calibración '2005-02-21' con tamaño de muestra de 10x10 elementos.

	1-10, 1-10	1-10, 5-14	1-10, 11-20	1-10, 15-24	1-10, 21-30	1-10, 25-34	1-10, 31-40	1-10, 35-44	1-10, 37-46
G	0.33	0.35	0.95	0.99	0.21	0.98	0.61	0.78	0.44
W	0.75	0.85	0.58	0.92	0.80	0.91	0.51	0.97	0.99
N	0.36	0.56	0.08	0.39	0.93	0.96	0.49	0.93	0.98
Ln	0.05	0.09	0.88	0.55	0.04	0.75	0.47	0.36	0.14
	5-14, 1-10	5-14, 5-14	5-14, 11-20	5-14, 15-24	5-14, 21-30	5-14, 25-34	5-14, 31-40	5-14, 35-44	5-14, 37-46
G	0.81	0.20	0.87	0.74	0.21	0.15	0.34	0.57	0.05
W	0.99	0.66	0.51	0.66	0.63	0.69	0.95	0.84	0.50
N	0.64	0.55	0.08	0.17	0.56	0.96	0.96	0.88	0.70
Ln	0.22	0.04	0.88	0.43	0.04	0.03	0.07	0.18	0.01
	11-20, 1-10	11-20, 5-14	11-20, 11-20	11-20, 15-24	11-20, 21-30	11-20, 25-34	11-20, 31-40	11-20, 35-44	11-20, 37-46
G	0.55	0.32	0.04	0.15	0.95	0.29	0.31	0.40	0.10
W	0.75	0.80	0.43	0.47	0.98	0.89	0.87	0.52	0.37
N	0.30	0.84	0.67	0.60	0.31	0.99	0.47	0.63	0.59
Ln	0.12	0.06	0.01	0.07	0.39	0.04	0.04	0.10	0.01
	15-24, 1-10	15-24, 5-14	15-24, 11-20	15-24, 15-24	15-24, 21-30	15-24, 25-34	15-24, 31-40	15-24, 35-44	15-24, 37-46
G	0.81	0.44	0.65	0.49	0.29	0.37	0.40	0.92	0.85
W	0.82	0.89	0.88	0.98	0.68	0.82	0.81	0.84	0.92
N	0.13	0.68	0.86	0.91	0.39	0.93	0.92	0.57	0.68
Ln	0.26	0.16	0.27	0.20	0.03	0.03	0.08	0.51	0.50
	21-30, 1-10	21-30, 5-14	21-30, 11-20	21-30, 15-24	21-30, 21-30	21-30, 25-34	21-30, 31-40	21-30, 35-44	21-30, 37-46
G	0.35	0.50	0.20	0.47	0.40	0.58	0.33	0.53	0.60
W	0.13	0.70	0.72	0.69	0.94	0.89	0.96	0.80	0.90
N	0.01	0.46	0.52	0.56	0.83	0.96	0.98	0.83	0.63
Ln	0.72	0.18	0.11	0.20	0.14	0.13	0.07	0.19	0.21
	25-34, 1-10	25-34, 5-14	25-34, 11-20	25-34, 15-24	25-34, 21-30	25-34, 25-34	25-34, 31-40	25-34, 35-44	25-34, 37-46
G	0.55	0.94	0.34	0.49	0.44	0.77	0.19	0.44	0.71
W	0.22	0.57	0.63	0.91	0.88	0.83	0.82	0.93	0.70
N	0.03	0.22	0.66	0.58	0.42	0.50	0.81	0.98	0.21
Ln	0.59	0.44	0.24	0.17	0.16	0.29	0.03	0.18	0.25
	31-40, 1-10	31-40, 5-14	31-40, 11-20	31-40, 15-24	31-40, 21-30	31-40, 25-34	31-40, 31-40	31-40, 35-44	31-40, 37-46
G	0.18	0.83	0.10	0.74	0.91	0.34	0.59	0.98	1.00
W	0.49	0.98	0.33	0.99	0.93	0.35	0.86	0.80	0.75
N	0.73	0.42	0.38	0.55	0.45	0.51	0.35	0.23	0.11
Ln	0.09	0.25	0.01	0.21	0.36	0.18	0.17	0.46	0.67
	35-44, 1-10	35-44, 5-14	35-44, 11-20	35-44, 15-24	35-44, 21-30	35-44, 25-34	35-44, 31-40	35-44, 35-44	35-44, 37-46
G	0.08	0.57	0.08	0.36	0.77	0.45	0.91	0.93	0.54
W	0.29	0.88	0.32	0.77	0.60	0.62	0.94	0.96	0.77
N	0.64	0.87	0.72	0.54	0.25	0.61	0.42	0.29	0.66
Ln	0.03	0.13	0.01	0.07	0.48	0.22	0.27	0.70	0.18
	37-46, 1-10	37-46, 5-14	37-46, 11-20	37-46, 15-24	37-46, 21-30	37-46, 25-34	37-46, 31-40	37-46, 35-44	37-46, 37-46
G	0.21	0.65	0.12	0.43	0.55	0.55	0.66	0.81	0.25
W	0.58	1.00	0.62	0.98	0.82	0.38	0.49	0.81	0.77
N	0.59	0.87	0.96	0.81	0.76	0.12	0.12	0.16	0.73
Ln	0.11	0.14	0.02	0.08	0.26	0.28	0.78	0.53	0.06

Tabla 4-48. Resultados del cálculo de la aplicación del test K-S para diferentes distribuciones en la calibración '2005-04-18' con tamaño de muestra de 10x10 elementos.

Los análisis realizados en las ventanas de las 6 muestras anteriores proporcionan ajustes que difieren de los realizados al fichero global correspondiente. Si en los ajustes globales los p-valores de las distribuciones normal y lognormal son nulos, en los análisis por ventanas se alcanzan valores altos en bastantes casos. La distribución Weibull se ciñe mejor a los datos globales que la gamma.

En general, el análisis en cuanto al ajuste de los datos por ventanas a las citadas distribuciones difiere notablemente del realizado de forma global.

Se indica a continuación los resultados de la aplicación de la prueba K-S a los datos de la calibración '2005-02-21' con diferentes tamaños de ventanas. Las pruebas objeto de contraste son como sabemos, la gamma (G), la Weibull (W), la normal (N) y la logarítmico normal (Logn) y el resultado mostrado es el p-valor.

	1-15, 1-15	1-15, 8-22	1-15, 15-29	1-15, 22-36	1-15, 29-43	1-15, 32-46
G	0.06	0.02	0.09	0.23	0.05	0.07
W	0.72	0.26	0.42	0.73	0.40	0.78
N	0.77	0.35	0.31	0.34	0.77	0.62
Logn	0.00	0.00	0.03	0.08	0.00	0.00
	8-22, 1-15	8-22, 8-22	8-22, 15-29	8-22, 22-36	8-22, 29-43	8-22, 32-46
G	0.84	0.30	0.80	0.58	0.25	0.05
W	0.85	0.78	0.99	0.82	0.69	0.66
N	0.08	0.21	0.13	0.11	0.57	0.40
Logn	0.10	0.03	0.07	0.19	0.04	0.00
	15-29, 1-15	15-29, 8-22	15-29, 15-29	15-29, 22-36	15-29, 29-43	15-29, 32-46
G	0.18	0.21	0.36	0.58	0.37	0.33
W	0.83	0.51	0.98	0.91	0.91	0.81
N	0.31	0.56	0.57	0.35	0.32	0.19
Logn	0.01	0.03	0.02	0.18	0.17	0.02
	22-36, 1-15	22-36, 8-22	22-36, 15-29	22-36, 22-36	22-36, 29-43	22-36, 32-46
G	0.19	0.58	0.68	0.37	0.22	0.89
W	0.68	0.90	0.78	1.00	0.69	0.51
N	0.21	0.35	0.29	0.68	0.11	0.03
Logn	0.01	0.08	0.11	0.04	0.07	0.47
	29-43, 1-15	29-43, 8-22	29-43, 15-29	29-43, 22-36	29-43, 29-43	29-43, 32-46
Gam	0.33	0.79	0.73	0.00	0.89	0.82
W	0.96	0.94	0.98	0.88	0.77	0.45
N	0.27	0.11	0.14	0.41	0.08	0.01
Logn	0.04	0.16	0.07	0.04	0.15	0.32
	32-46, 1-15	32-46, 8-22	32-46, 15-29	32-46, 22-36	32-46, 29-43	32-46, 32-46
G	0.32	0.99	0.78	0.06	0.11	0.22
W	0.94	0.92	0.43	0.48	0.36	0.42
N	0.13	0.07	0.00	0.48	0.12	0.07
Log	0.02	0.28	0.53	0.00	0.02	0.03

Tabla 4-49. Resultados del cálculo de la aplicación del test K-S para diferentes distribuciones para módulos con tamaño de muestra de 15x15 elementos en la calibración '2005-02-21'.

Los resultados de la tabla anterior son similares a los proporcionados en el análisis de 10x10 elementos. El ajuste de cada muestra a las diferentes distribuciones ofrece valores satisfactorios en muchos casos. Los mejores ajustes son para las distribuciones gamma y Weibull, pero también se dan p-valores altos en algunas muestras para la distribución normal y la lognormal. Estos valores difieren del resultado global que sólo presenta un ajuste razonable a la distribución Weibull. Observamos, por tanto, diferencias de resultados de las ventanas entre sí y con el fichero global.

En la tabla siguiente se indican los resultados del análisis con tamaño de muestra de 20x20 elementos.

	1-20, 1-20	1-20, 10-29	1-20, 21-40	1-20, 27-46
G	0.01	0.02	0.16	0.09
W	0.30	0.36	0.96	0.57
N	0.22	0.18	0.19	0.64
Logn	0.00	0.00	0.01	0.00
	10-29, 1-20	10-29, 10-29	10-29, 21-40	10-29, 27-46
G	0.24	0.47	0.46	0.73
W	0.66	0.99	0.71	0.67
N	0.07	0.04	0.03	0.19
Logn	0.00	0.02	0.06	0.05
	21-40, 1-20	21-40, 10-29	21-40, 21-40	21-40, 27-46
G	0.04	0.20	0.24	0.39
W	0.00	1.00	1.00	0.85
N	0.41	0.29	0.18	0.04
Logn	0.00	0.01	0.01	0.03
	27-46, 1-20	27-46, 10-29	27-46, 21-40	27-46, 27-46
G	0.12	0.85	0.10	0.10
W	0.92	0.90	0.50	0.34
N	0.09	0.01	0.20	0.04
Logn	0.00	0.03	0.00	0.00

Tabla 4-50. Resultados del cálculo de la aplicación del test K-S para diferentes distribuciones para módulos con tamaño de muestra de 20x20 elementos en la calibración '2005-02-21'.

Las tabla anterior (tamaño 20x20) y posterior (25x25) ofrece unos resultados que se van aproximando más a los ofrecidos por el análisis global.

	1-25,1-25	1-25,12-36	1-25,22-46
G	0.01	0.10	0.10
W	0.00	0.00	0.80
N	0.10	0.05	0.08
Logn	0.00	0.00	0.00
	12-36,1-25	12-36,12-36	12-36,22-46
G	0.03	0.27	0.35
W	0.80	0.78	0.97
N	0.07	0.05	0.04
Logn	0.00	0.00	0.00
	22-46,1-25	22-46,12-36	22-46,22-46
G	0.09	0.45	0.10
W	0.89	0.80	0.01
N	0.04	0.01	0.01
Logn	0.00	0.01	0.00

Tabla 4-51. Resultados del cálculo de la aplicación del test K-S para diferentes distribuciones para módulos con tamaño de muestra de 25x25 elementos en la calibración '2005-02-21'.

En las dos últimas tablas se hace patente la similitud de resultados del análisis global y con tamaño de muestra mayor.

	1-30, 1-30	1-30, 17-46
G	0.01	0.02
W	0.38	0.82
N	0.13	0.00
Logn	0.00	0.00
	17-46, 1-30	17-46, 17-46
G	0.34	0.18
W	0.81	0.98
N	0.01	0.01
Logn	0.00	0.00

Tabla 4-52. Resultados del cálculo de la aplicación del test K-S para diferentes distribuciones para módulos con tamaño de muestra de 30x30 elementos en la calibración '2005-02-21'.

Concluimos este apartado con la tabla de tamaño 35x35 de la calibración '2005-02-21' incidiendo en la idea de que en tamaños de muestra pequeños existen diferencias de ajuste por zonas y con el resultado general y que al aumentar el tamaño de la muestra, ésta presenta resultados parecidos al análisis global.

	1-35, 1-35	1-35, 12-46
G	0.02	0.01
W	0.68	0.83
N	0.00	0.02
Logn	0.00	0.00
	12-46, 1-35	12-46, 12-46
G	0.28	0.27
W	0.80	0.75
N	0.00	0.00
Logn	0.00	0.00

Tabla 4-53. Resultados del cálculo de la aplicación del test K-S para diferentes distribuciones para módulos con tamaño de muestra de 35x35 elementos en la calibración '2005-02-21'.

- Histogramas de distribución de datos.

En la figura siguiente aparece el histograma correspondiente a la submuestra de tamaño 10x10 denominada (25-34,21-30) del archivo '2005-04-21e' por ser la ventana que contiene la observación discordante mayor. En el apartado de histograma de distribución de datos en análisis generales, se acompaña (Fig. 4-9) el ejemplo correspondiente a esta misma calibración con todas las observaciones.

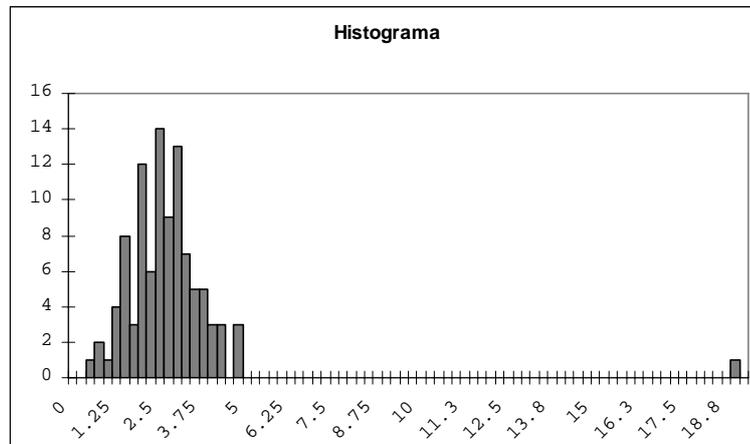


Fig. 4-10. Histograma de distribución de datos en muestras de 10x10 elementos afectados de error.

Histograma correspondiente a la submuestra de tamaño 10x10 denominada (25-34,21-30) del archivo '2005-04-21e' por ser la submuestra que contiene la observación discordante mayor.

En el gráfico siguiente se muestra el histograma correspondiente a la primera ventana del análisis de 10x10 elementos (100 valores), 15x15 elementos (225 valores), 20x20 elementos (400 valores), 25x25 elementos (625 valores), 30x30 elementos (900 valores) y 35x35 elementos (1225 valores) de la calibración '2005-02-21'.

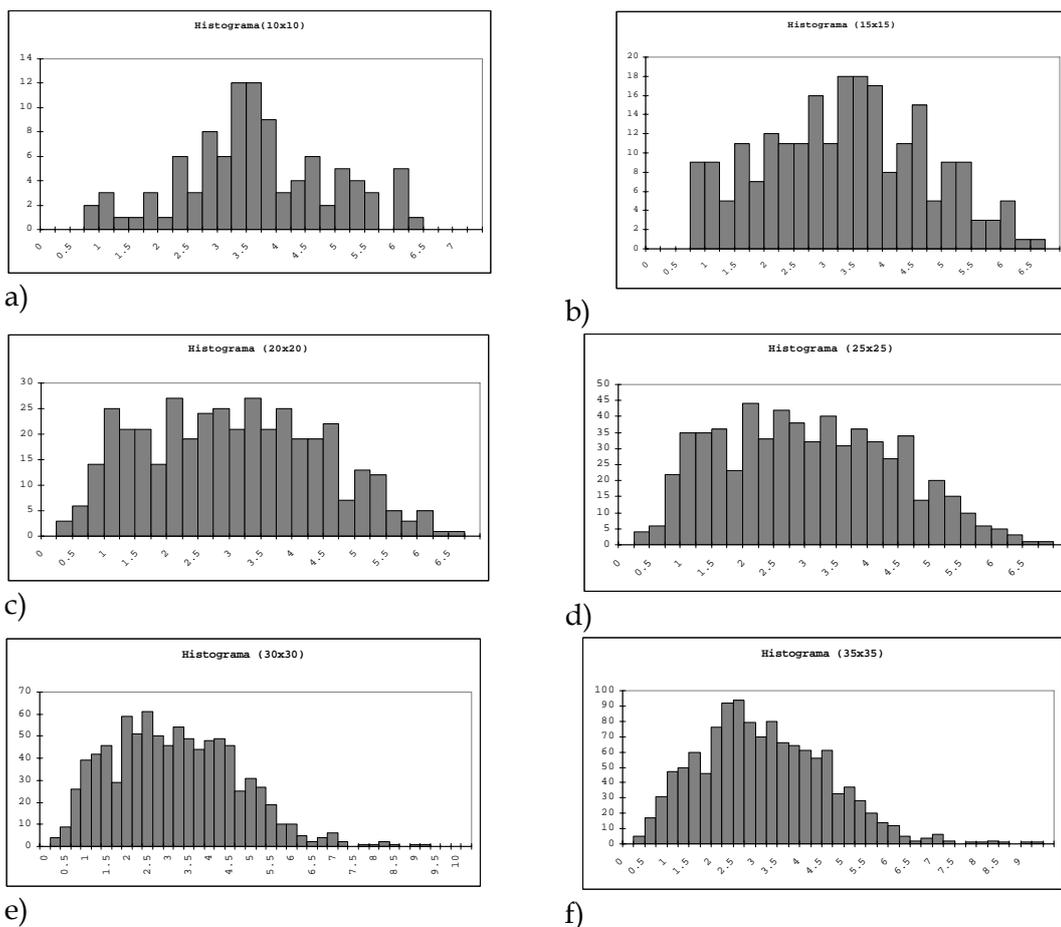


Fig. 4-11. Histogramas de distribución de datos lineales para muestras de diferentes tamaños en la calibración '2005-02-21'.

En el gráfico siguiente se muestra el histograma correspondiente a la primera ventana del análisis de: a) 10x10 elementos (100 valores), b) 15x15 elementos (225 valores), c) 20x20 elementos (400 valores), d) 25x25 elementos (625 valores), e) 30x30 elementos (900 valores) y f) 35x35 elementos (1225 valores) de la calibración '2005-02-21'. Vemos que conforme aumenta el tamaño de la muestra las distribuciones tienden a una asimetría positiva.

Observamos en la figura anterior que conforme aumentamos el tamaño de la muestra aparece el carácter asimétrico de la distribución; hecho que habíamos corroborado de forma analítica anteriormente al comprobar que en muestras de menor tamaño sí existían ajustes satisfactorios a la distribución normal. A medida que aumenta el volumen de datos tratados el mejor ajuste corresponde a distribuciones asimétricas positivas como la Weibull.

Según lo visto hasta el momento, tanto las distribuciones lineales como las circulares varían según el tamaño de la muestra analizada. Este hecho debe justificar la aplicación de pruebas no paramétricas para el estudio de los vectores de error, como se verá en el apartado siguiente de distribución conjunta.

Es el momento de resumir los resultados del análisis de los datos lineales de los vectores de error. Estos análisis se han centrado en calcular: a) estadísticos básicos (de tendencia central y dispersión), b) determinación de *outliers*, c) pruebas de determinación de las características de las distribuciones de datos (ajustes a determinadas distribuciones) y d) histogramas de frecuencia de valores.

Los datos objeto de análisis son, como sabemos, 20 muestras de 2116 vectores y 2 de 361 elementos. También 6 de las 20 muestras se analizaron con tamaño de submuestra de 100 elementos y una de éstas 6, se trató con tamaños de muestras variables.

De los resultados se infiere que la distribución de los datos depende del número de elementos de la muestra y que no todos los errores se detectan con los métodos aplicados, lo que sugiere que los métodos paramétricos para el estudio de las características de la muestra deben ser complementados con la aplicación de métodos no paramétricos.

4.1.4.3 Resultados del análisis en la distribución conjunta

Los resultados de este apartado suponen un tratamiento conjunto de los ángulos y los módulos una vez que se ha analizado el comportamiento de ambos por separado. La forma de acceder a estos datos es consultando la siguiente dirección:

D://Tesis/Datos/Escáner/nombre_fichero

También aparecen resumidos en la dirección siguiente los resultados de las correlaciones entre módulos y acimutes, así como la diferencia entre los valores de la desviación típica entre la dirección SCAN y CCD.

D://Tesis/Datos/Escáner/Resumen

- Correlaciones existentes entre módulos y acimutes.

Sabemos a estas alturas de la Tesis de la importancia de analizar los errores de posición no sólo en módulo sino también en orientación. Cabe preguntarse si a errores grandes en una magnitud le corresponden, de forma directa o inversa, errores en la otra. El estudio de las correlaciones existentes entre magnitudes lineales y angulares cobra, por tanto, importancia y se realiza analizando el coeficiente de correlación, que representamos por r , entre ambos elementos. Según los datos de la tabla adjunta observamos que, en la mayoría de los casos, no existe correlación entre ambas variables.

Fichero Calibración	Valor r correlación	Fichero Calibración	Valor r correlación
2004-10-19	0.39	2005-05-03e	0.00
2005-02-21	0.00	2005-05-18e	0.00
2005-03-01	0.00	2005-05-23e	0.00
2005-03-07	0.00	2005-06-09	0.13
2005-03-14	0.00	2005-06-29	0.00
2005-04-04	0.00	2005-07-05e	0.00
2005-04-11	0.00	2005-07-11e	0.00
2005-04-18	0.12	2005-07-12e	0.00
2005-04-21e	0.04	2005-07-13e	0.01
2005-04-22e	0.48	2006-05-19c	0.00
2005-04-25	0.00	2006-06-28c	0.00

Tabla 4-54. Resultados de las correlaciones entre los módulos y acimutes.

En la figura siguiente se muestran dos gráficos a) y b) correspondientes a la correlación existente entre datos lineales y angulares en las calibraciones '2005-04-22e' y '2005-02-21', respectivamente. Se exponen estos dos ficheros de datos por presentar el primero el valor más alto de correlación de todos los analizados ($r = 0.48$) mientras que el apartado b) tiene un valor nulo.

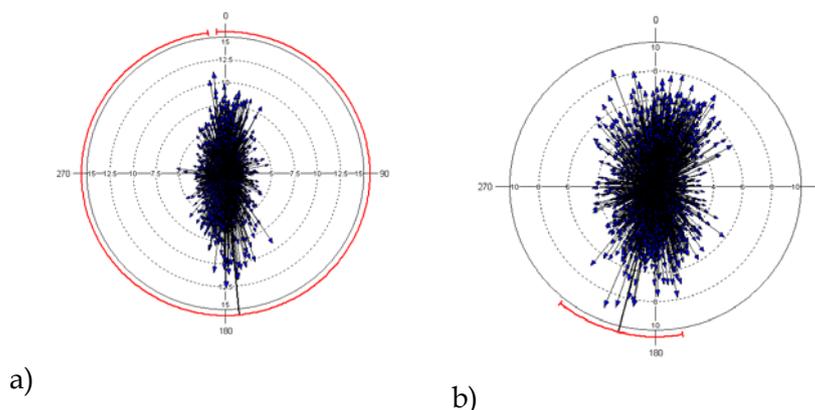


Fig. 4-12. Gráficos de correlación entre datos lineales y circulares en las calibraciones a) '2005-04-22e' y b) '2005-02-21'. Los datos del gráfico a) tienen una correlación de 0.48 mientras que los del b) presentan un valor nulo.

Aunque el apartado a) del gráfico anterior corrobora el resultado analítico, debemos recordar que 16 de los 22 ficheros ofrecían un valor de correlación nulo y de los 6 que tenían un resultado diferente de 0, sólo 4 podían ser considerados significativos.

- Diferencias entre los valores de la desviación típica en la dirección SCAN (eje X) y CCD (eje Y).

En la tabla siguiente se indica el valor de la desviación típica en cada fichero de calibración para la dirección SCAN, que recordemos era la considerada como eje de abscisas, y la CCD o eje de ordenadas. Se observa mayor valor de la desviación en la columna CCD que la SCAN. Como se

ha visto anteriormente (Fig. 4-2) existe una distribución bimodal angular en la dirección 0°-180°, que se refleja también en un mayor error lineal en la dirección del considerado eje Y (CCD).

Fichero Calibración	Desv.tip SCAN	Desv.tip CCD	Fichero Calibración	Desv.tip SCAN	Desv.tip CCD
2004-10-19	1.553	1.513	2005-05-03e	1.528	3.739
2005-02-21	1.548	2.934	2005-05-18e	1.601	3.083
2005-03-01	1.540	1.862	2005-05-23e	1.617	2.435
2005-03-07	1.454	2.683	2005-06-09	1.344	1.519
2005-03-14	1.436	2.794	2005-06-29	1.537	1.275
2005-04-04	1.615	2.555	2005-07-05e	1.547	3.155
2005-04-11	1.579	2.798	2005-07-11e	1.457	2.107
2005-04-18	1.485	2.291	2005-07-12e	1.325	2.066
2005-04-21e	1.465	3.055	2005-07-13e	1.469	1.923
2005-04-22e	1.337	3.425	2006-05-19c	1.183	2.520
2005-04-25	1.607	1.420	2006-06-28c	1.430	1.830

Tabla 4-55. Resultados de los valores de la desviación típica en la dirección SCAN y CCD en las 22 calibraciones.

Cabría replantearse ahora por qué los valores de correlación anteriores no corroboran estos datos. Entendemos que este tipo de distribución con tendencia bimodal, con un elevado número de elementos, no permite obtener un estadístico de correlación fiable.

- *Outliers.*

El método usado para determinar los valores discordantes en la distribución conjunta de datos era la identificación de un porcentaje de elementos en función del cálculo de la media armónica. Recordemos que la media armónica es el cociente entre el número de observaciones y la suma de las inversas de las distancias entre un punto *i* y el resto de localizaciones. Centrando el cálculo en cada punto catalogamos como *outliers* aquellos que se alejen más de la masa central de puntos. El porcentaje calculado se cifró de entrada en un 2%, lo que suponía aislar como posibles *outliers* a 42 puntos en los ficheros de 2116 elementos y a 7 en los de 316 puntos.

Las consideraciones sobre los *outliers* se realizan sobre la figura 4-13, al tiempo que se estudian los mapas de densidad.

- Mapas de densidades.

Dado que la distribución de datos lineales no es normal y tal como se apuntó en el resumen del resultado de análisis de datos lineales, es conveniente recurrir a métodos no paramétricos para estudiar la distribución conjunta de datos. Para ello transformamos los vectores de error en una nube de puntos correspondientes a los nodos finales de los citados vectores y les superponemos un mapa de densidad de puntos según el número de puntos por unidad de área. En el gráfico siguiente aparece la

nube de puntos y los mapas de densidad en varias calibraciones. Se hacen notar los *outliers* con un color diferente y de mayor tamaño.

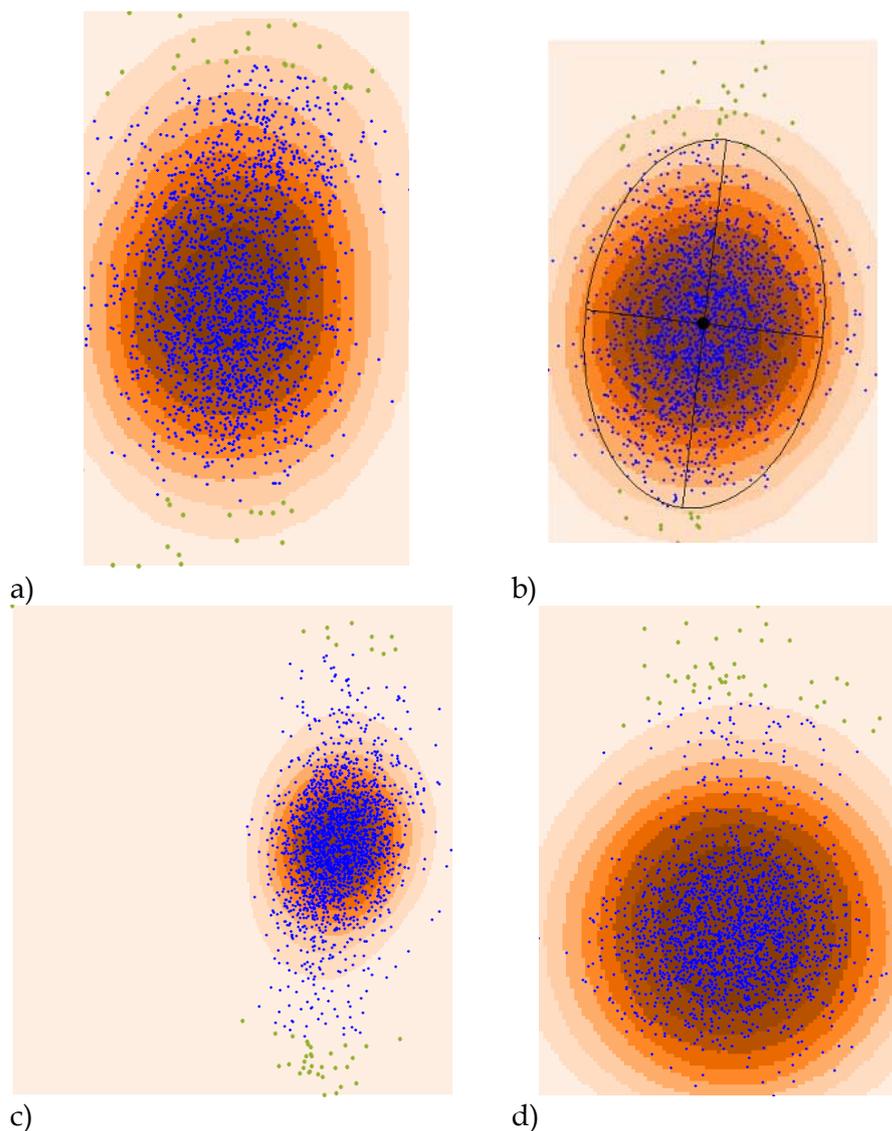


Fig. 4-13. Mapas de densidad y *outliers* en 4 calibraciones del escáner fotogramétrico. Esta figura se corresponde con los mapas de densidad e indicación de *outliers* (en color verde) de las calibraciones: a) '2005-02-21', b) '2005-04-18', c) '2005-04-21e' y d) '2006-06-28'. En general, la distribución de puntos sigue una línea en torno a los 10° - 190° aproximadamente.

La figura anterior se corresponde con los mapas de densidad e indicación de *outliers* (en color verde y mayor tamaño) de las calibraciones: a) '2005-02-21', b) '2005-04-18', c) '2005-04-21e' y d) '2006-06-28' a modo de muestra. En general, todos los mapas realizados mostraban una distribución de puntos que originaban mapas de densidad de forma alargada, excepto algún caso como el mostrado en el apartado d) de la figura anterior correspondiente a la calibración procedente de la Universidad de Cantabria. Esta forma alargada de distribución sigue una dirección aproximada en torno a los 10° - 190° en 15 de las 22 calibraciones, otras 3 siguen una dirección 0° - 180° y el resto, sólo 4, variable. La figura b) en la que se ha superpuesto una elipse de probabilidad al 95%, ilustra bastante bien la idea indicada.

Los *outliers* están en los extremos de la distribución hacia la parte superior e inferior, rara vez a izquierda o derecha. En el apartado c) de la figura anterior, correspondiente a la calibración '2005-04-21e', se aprecia un *outlier*, en este caso provocado artificialmente, en la esquina superior izquierda, muy alejado de la masa de puntos.

En la figura siguiente sobre la calibración '2005-02-21' se modificaron una serie de puntos para que fueran catalogados como *outliers* y estudiar así como variaban los mapas de densidad, y por tanto la distribución.

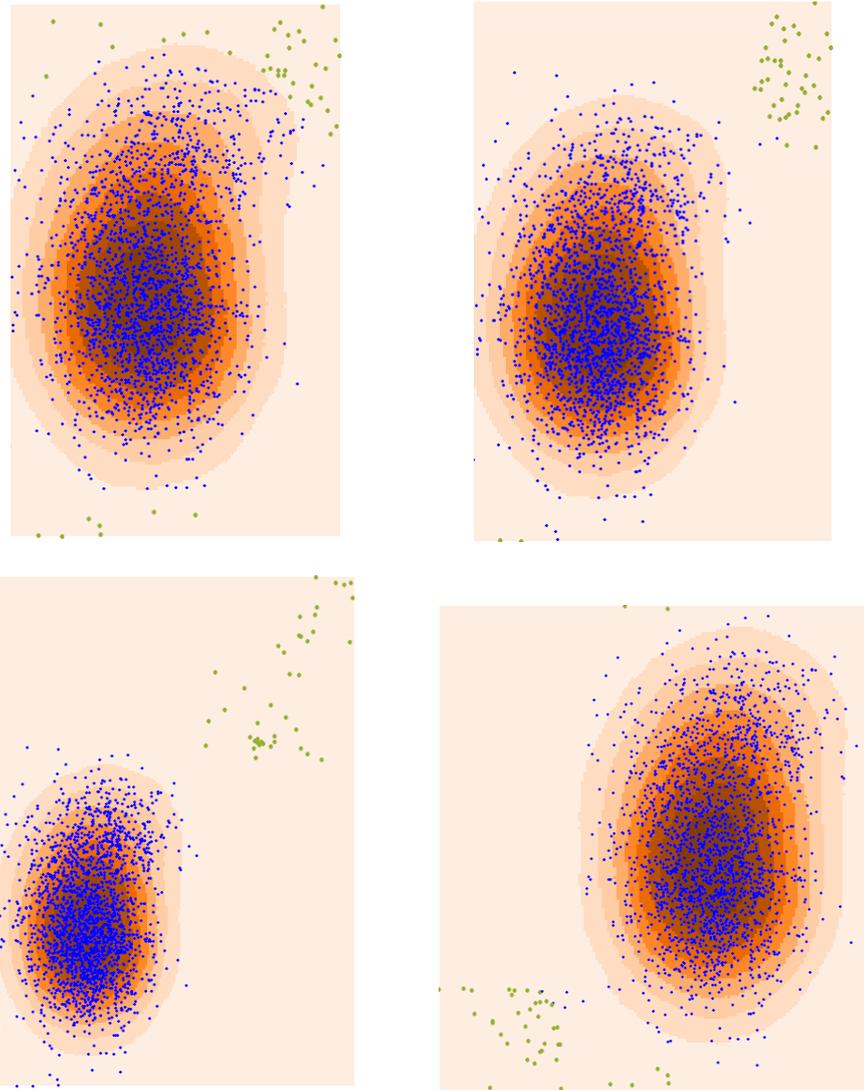


Fig. 4-14. Modificación de *outliers* en el fichero '2005-02-21'.

Sobre la calibración '2005-02-21' se modificaron un 2% de los puntos en diferentes sentidos para estudiar la variación de los mapas de densidad.

La utilidad de estos resultados es determinar dónde pueden estar concentrados los problemas de mayores errores en nuestro conjunto de datos.

Resumen

El análisis de la distribución conjunta de datos se ha centrado en calcular: a) correlaciones entre los módulos y los acimutes, b) diferencias entre los valores de la desviación típica en la dirección X e Y, c) determinación de *outliers* y d) elaboración de mapas de densidad.

Los datos objeto de análisis vuelven a ser las 20 muestras de 2116 vectores y 2 de 361 elementos.

De los resultados destacamos que las distribuciones de puntos siguen unas direcciones preferentes, producto probablemente de una influencia de tipo mecánico en la obtención de los datos, siendo necesario recurrir a métodos no paramétricos para modelizar la distribución.

La determinación de *outliers* mediante mapas de densidad permite localizar dónde se producen los mayores errores y pueden sugerir soluciones para corregirlos.

4.1.5 Conclusiones

Los análisis y resultados obtenidos anteriormente nos inducen a establecer unas conclusiones que se exponen siguiendo las mismas pautas anteriores.

4.1.5.1 Conclusiones para datos circulares

a) Conclusiones de los análisis generales

- Estadísticos básicos para vectores unitarios
 - El acimut medio tiene el mismo sentido que el acimut topográfico, esto es creciendo desde la dirección del eje Y en sentido horario. El valor medio para los dos grupos de datos manejados (n= 2116 y 316) se corresponde con 143.95° y 162.99°, respectivamente. Este valor por sí solo, no permite extraer más conclusiones que la obvia.
 - El valor medio del módulo medio (\bar{R}) tiene un valor, para los 20 ficheros de 2116 vectores, de 0.04 y de 0.08 para los de 361 elementos. Recordemos que este valor oscila entre 0 y 1. Cuando el resultado sea la unidad todas las observaciones deben coincidir, aunque valores próximos a cero no implican necesariamente una distribución uniforme. Sí podemos descartar con estos resultados una concentración alta de valores en una dirección concreta.

- Los valores de desviación estándar circular de las muestras son muy altos (147.34° y 136.94°) lo que implica una dispersión de datos grande.
- El parámetro de concentración de von Mises podía ser considerado significativo a partir de 2, valor muy lejano a las medias obtenidas (0.087 y 0.16) que indican distribuciones más próximas a la uniforme que a una en la que exista una dirección preferente de datos.
- Pruebas para la determinación de las características de las distribuciones de datos circulares.
 - Con los resultados de los test de uniformidad se concluye que no existe evidencia de este tipo de distribución en los datos, excepto en los ficheros correspondientes a las 361 observaciones.
 - El control de la anisotropía se realiza en función del error estándar de la media, que es muy alto en la mayoría de los casos, lo que se traduce en algunas muestras, en intervalos de confianza para el acimut medio que prácticamente ocupan la circunferencia entera. Descartamos también la existencia de una sola dirección preferente en el conjunto de los datos.
 - Las pruebas de bondad de ajuste a la distribución von Mises descartan también la posibilidad de que los datos se ajusten a esta distribución, hecho que intuíamos al analizar el parámetro de concentración de von Mises.
 - En los resultados anteriores tampoco se aprecian diferencias significativas entre los ficheros que están afectados de error artificial y los que no. Por lo que, de entrada, entendemos que los test aplicados de forma general a toda la población no son capaces de detectar errores locales. Se debe proceder, por tanto, a analizar muestras o ventanas de diferente tamaño para comparar datos globales y locales.
- Gráficos de distribución de datos circulares.

En el dibujo de los datos circulares, en la práctica totalidad de las calibraciones, se aprecia en general una bimodalidad en cuanto a la distribución de la dirección de los vectores en torno a una dirección que puede situarse entre los 0° - 10° y los 180° - 190° . Esta dirección 0° - 180° se corresponde con la dirección Y o CCD del escáner, lo que comporta una influencia del movimiento del sensor en el escaneo y se supone motivada por la componente mecánica en el movimiento del sensor en el escaneo de la placa de calibración. En los ficheros de 361 vectores esta bimodalidad es menor. También aparece en la calibración procedente de la Universidad de Cantabria, aunque no tan acusada, lo que incita a comprobar si es este un problema genérico en este tipo de escáneres.

b) Conclusiones de los análisis locales

- Estadísticos básicos para vectores unitarios.

En el caso de las 6 calibraciones en las que se analizaron muestras de 10×10 elementos, extraemos las siguientes conclusiones:

- El valor medio del acimut medio oscila entre 166° y 184° , y no se concluye nada en relación con los existentes para la calibración completa y para las submuestras.
- El módulo medio presenta valores más altos, en torno a 0.50, que los vistos para el análisis general, que era de 0.04, lo que implica una concentración mayor de datos.
- Los valores de la desviación estándar circular son muchos más pequeños, casi una tercera parte en algunos casos, que los del análisis general corroborando una mayor concentración de datos.
- El parámetro de concentración de von Mises aumenta en las submuestras llegando a ser claramente significativo en muchas ventanas con valores superiores a 2, teniendo algún caso extremo que alcanza hasta 20.77, lo que implica una distribución con una clara dirección preferente.

En la calibración '2005-02-21' al analizar muestras de diferentes tamaños observamos que:

- El acimut medio aumenta al tiempo que lo hace el tamaño de la muestra, mientras que el valor de la desviación típica para el acimut medio disminuye.
 - El valor medio del módulo medio disminuye desde 0.6 en tamaño de muestras de 10×10 a 0.04 en el fichero completo. Esto supone que aumenta el grado de dispersión de valores conforme consideramos la muestra completa.
 - La desviación estándar circular de la muestra aumenta junto con el tamaño de la muestra, implicando, al igual que el estadístico anterior, un aumento de la dispersión de valores según se incrementa el volumen de datos manejados.
 - El parámetro de concentración de von Mises, en la misma línea anterior, disminuye claramente al aumentar el tamaño de la muestra.
- Pruebas para la determinación de las características de las distribuciones de datos circulares.

Las conclusiones extraídas en el caso de las 6 calibraciones en las que se analizaron muestras de 10×10 elementos, son:

- Sólo 22 de las 486 (81 muestras \times 6 calibraciones) submuestras analizadas evidenciaban probabilidad de uniformidad, por lo que concluimos que las ventanas de tamaño 10×10 no tenían, al igual que las muestras globales, esta distribución.
- El control de la anisotropía, por el error estándar del acimut medio, mostraba valores mucho más bajos en el análisis por ventanas (10.3°

de media) que en el global (29.4°) lo que implica una tendencia a una concentración de valores mayor.

- El ajuste a la distribución von Mises ofrecía resultados diversos, pero un número significativo de muestras presentaban esta distribución, hecho que no se observaba en el análisis global.

En la calibración '2005-02-21', al analizar muestras de diferentes tamaños, concluimos que:

- Exceptuando alguna muestra aislada no existe isotropía en la distribución.
 - El control de la anisotropía indica una concentración de valores mayor que en el caso del análisis global.
 - Sólo algunas de las ventanas de tamaño 10x10 se ajustaban a la distribución von Mises. Todos los demás tamaños, al igual que el global, no presentaban evidencias de esta distribución.
- Gráficos de distribución de datos circulares.

El análisis de la figura correspondiente (Fig. 4-7) permite corroborar como aumenta el valor del acimut medio de la muestra y el carácter bimodal de la distribución, al tiempo que tratamos con muestras de mayor tamaño.

4.1.5.2 Conclusiones para datos lineales

Tras el análisis realizado a los datos lineales inferimos las siguientes conclusiones:

a) Conclusiones de los análisis generales

- Cálculos de los estadísticos básicos de los módulos de los vectores.

La comparación entre los valores de los estadísticos básicos entre los ficheros afectados de error y los que no lo estaban, no ofrecía diferencias considerables. Entendemos que los puntos afectados de error quedan diluidos entre el elevado número de vectores de la muestra. Los valores de desviación típica son relativamente altos con respecto a la media lo que implica errores aleatorios elevados.

- Determinación de *outliers*.

El porcentaje de observaciones discordantes oscila entre el 0.4% y el 2.5%. No necesariamente los ficheros afectados de error son los que mayor porcentaje de *outliers* presentan.

- Pruebas para la determinación de las características de las distribuciones de datos lineales.

Todas las calibraciones analizadas se adaptaban bien a la distribución gamma o a la Weibull, lo que permite concluir que los módulos siguen una distribución asimétrica positiva. Las excepciones a esta regla son los ficheros '2005-04-21e', '2005-05-23e' y '2006-05-19c', teniendo los dos primeros error provocado que puede dificultar un correcto ajuste.

- Histograma de frecuencia de valores

Corroboran en mayor o menor medida la distribución asimétrica positiva de valores.

b) Conclusiones de los análisis locales

- Cálculos de los estadísticos básicos de los módulos de los vectores.

De los 6 ficheros analizados con tamaño de muestra de 10x10 elementos los valores más altos de media aritmética y desviación típica se corresponden con las submuestras afectadas de error, como era de esperar.

- Determinación de *outliers*.

El porcentaje de *outliers* es inferior en el análisis local que en el global.

- Pruebas para la determinación de las características de las distribuciones de datos lineales.

Mientras que las muestras del análisis global presentaban p-valores nulos para el ajuste a la distribución normal, observamos que esta distribución se adapta mejor en muestras de menor tamaño. Conforme aumenta el tamaño de la muestra se evidencia la asimetría. Existe por tanto diferencias ente el análisis local y global y entre las distintas zonas del local.

- Histograma de frecuencia de valores

En tamaños de muestras de 10x10 elementos la distribución predominante es la normal (ver Fig. 4-10) mientras que en la comparativa de histogramas para muestras de diferentes tamaños (Fig. 4-11) se observa como los datos tienden a un carácter asimétrico según aumenta el tamaño de la muestra, partiendo de una distribución más próxima a la normal.

4.1.5.3 Conclusiones para la distribución conjunta

Corresponde a este punto desarrollar las conclusiones para la distribución conjunta de datos a tenor de los cálculos realizados anteriormente.

- Correlaciones existentes entre los módulos y los acimutes en cada muestra.

En los ficheros globales no existe correlación en general entre acimutes y módulos. Entendemos que la distribución bimodal de datos enmascara una posible correlación.

- Diferencias entre los valores de la desviación típica en la dirección SCAN (eje X) y CCD (eje Y).

Los valores de la desviación típica en la dirección considerada como eje Y (CCD) son superiores a los determinados en la dirección del eje de abscisas (SCAN), lo que subraya la distribución bimodal ya indicada anteriormente en la dirección aproximada 0°-180°.

- Determinación de *outliers*.

La determinación de *outliers* a partir de parámetros como la media armónica se revela como un método efectivo y complementario a los ya utilizados.

- Mapas de densidad.

Los mapas de densidad de puntos permiten determinar las distribuciones reales de los datos en dos dimensiones lo que libera de “usar” círculos o elipses de probabilidad, ya que la forma y densidad de las distribuciones se deciden localmente, no globalmente.

4.2 Cálculos y resultados del tratamiento de datos de la comparación de cartografías a diferentes escalas

Se repite el esquema del apartado anterior, esto es, la explicación del origen de los datos, características, análisis..., pero referido ahora a los datos obtenidos de la comparación de cartografías a diferentes escalas.

4.2.1 Origen de los datos

Como ya se explicó en el apartado de material (3.1.2) disponemos de varios conjuntos de vectores de error de posición procedentes de la comparación de puntos comunes entre diferentes bases de datos espaciales. En concreto se dispone de:

- 500 vectores de error de la comparativa del MTN 25 y MTN 50.
- 500 vectores de error obtenidos de la comparativa entre cartografía a escala 1/10 000 y el MTN 50.
- 4 grupos de 250 vectores de error extraídos de la superposición de la cartografía a escala 1/25 000 del IGN obtenida desde un WMS y el MTN 50.

Estos vectores de error se han calculado a partir de las diferencias en X e Y entre la posición considerada como verdadera, la de mayor escala, y la de menor escala. Con cada incremento de coordenada en X e Y se calcula un módulo y un acimut que genera un vector de error.

Del mismo modo que se indicó para los datos del escáner fotogramétrico, los cálculos afectarán tanto a los incrementos de coordenadas como a los vectores de error y estos últimos serán también analizados en ángulo y módulo, por separado, y conjuntamente.

4.2.2 Características de los datos

En esencia las características de estos datos son similares a las del apartado anterior (4.1.2) pues tratamos igualmente con incrementos de coordenadas planas que se transforman en acimut y módulo. Las diferencias existentes (tamaños de la muestra y unidades) se reflejan en la tabla siguiente. Los ficheros de datos se denominan por '10-50' y '25-50', que, como es lógico, contienen los datos de la comparativa entre la cartografía a escala 1/10 000 y el MTN 50 y la correspondiente a la relación entre el MTN 25 con el MTN 50, respectivamente. En el caso de la cartografía obtenida de un servidor de mapas Web en relación con hojas del 50 000, el fichero se denomina por el nombre de la localidad más importante de la zona analizada.

Fichero	n	Unidad lineal	Superficie
10-50	500	metro	48 793 ha
25-50	500	metro	12 198 ha
Mérida	250	metro	146 381 ha
Villafranca	250	metro	146 381 ha
Villanueva	250	metro	146 381 ha
Almendralejo	250	metro	146 381 ha

Tabla 4-56. Características de los datos procedentes de la comparación entre cartografías.

4.2.3 Análisis realizados

Los análisis realizados comienzan con los datos circulares, después los lineales y finalmente la distribución conjunta. En este apartado no se ha introducido ningún error de forma artificial ni se han analizado submuestras, como en los datos procedentes del escáner fotogramétrico, por lo que el estudio será más simple que el ya descrito.

4.2.3.1 Análisis de datos circulares

Una vez determinados los vectores de error de ambos ficheros, se le aplicarán a los datos circulares los siguientes análisis:

- Cálculos de los estadísticos básicos para vectores unitarios (Apartado 3.2.1.1)
- Pruebas para la determinación de las características de las distribuciones de datos circulares (Apartado 3.2.3.1)
- Gráficos de distribución de datos circulares.

Obviamos una explicación más detallada, por ser los mismos análisis que ya han sido desarrollados y calculados en párrafos precedentes.

4.2.3.2 Análisis de datos lineales

En el análisis de los módulos de error procederemos a examinar lo siguiente:

- Cálculos de los estadísticos básicos de los módulos de los vectores (Apartado 3.2.1.2).
- Determinación de *outliers*.
- Pruebas para la determinación de las características de las distribuciones de datos lineales (Apartado 3.2.3.2).
- Histogramas de frecuencia de valores.

Prescindimos aquí también de una explicación que ya ha sido dada y remitimos al lector a los apartados indicados.

4.2.3.3 Análisis en la distribución conjunta

Volvemos a repetir el mismo esquema que comprende los siguientes cálculos para la distribución conjunta de los datos procedentes de la comparativa entre cartografías de diferentes escalas:

- Correlaciones existentes entre los módulos y los acimutes en cada muestra.
- Diferencias entre los valores de la desviación típica en la dirección del eje X e Y.
- Determinación de *outliers*.
- Elaboración de mapas de densidad.

4.2.4 Resultados

En las líneas siguientes se muestran los resultados obtenidos de realizar los análisis anteriormente citados a todos los ficheros de vectores de error generados al comparar las diferentes bases de datos espaciales.

4.2.4.1 Resultados del análisis de datos circulares

Los datos sobre los que realizaremos este primer análisis se localizan en la dirección siguiente donde encontraremos carpetas denominadas: '10-50', '25-50' y 'gvsig'. En esta última carpeta, a su vez, se encuentran, los resultados de las 4 comparativas ya citadas correspondientes a las muestras de 250 elementos. Se presenta también una tercera carpeta llamada 'Resumen', que obviamente es un extracto más cómodo de consultar de los análisis que a continuación se explican.

D://Tesis/Datos/Cartografía/nombre_fichero

Procedemos en el mismo orden que el análisis efectuado a los datos relativos a la calibración del escáner fotogramétrico, esto es: cálculo de los estadísticos básicos para vectores unitarios, pruebas de control del error y gráfico de distribución de datos circulares.

- Cálculo de los estadísticos básicos para vectores unitarios.

Los resultados de los estadísticos básicos para vectores unitarios (el acimut medio ($\bar{\theta}$), el módulo medio (\bar{R}), la desviación estándar circular de la muestra (ν) y el parámetro de concentración de von Mises (κ)) se distribuyen en la tabla adjunta para las comparativas tratadas.

	10-50 (n=500)	25-50 (n=500)	Mérida(n=250)
$\bar{\theta}$	277.59°	307.48°	324.92°
\bar{R}	0.15	0.43	0.25
ν	111.82°	74.14°	95.54°
κ	0.30	0.96	0.51
	Villafranca (n=250)	Villanueva (n=250)	Almendra.(n=250)
$\bar{\theta}$	98.57°	130.73°	276.23°
\bar{R}	0.07	0.10	0.19
ν	132.55°	122.01°	103.85°
κ	0.14	0.21	0.39

Tabla 4-57. Resultados de los estadísticos básicos para vectores unitarios de los datos procedentes de la comparación entre diferentes cartografías.

Con los valores bajos del módulo medio y el parámetro de concentración de von Mises y la elevada desviación estándar circular de la muestra, descartamos concentraciones de datos en una dirección preferente.

- Pruebas para la determinación de las características de las distribuciones de datos circulares.
 - Control de la uniformidad. Los tests aplicados son ya conocidos por cálculos anteriores tal como se observa en la tabla siguiente.

Tests	10-50 (n=500)	25-50 (n=500)	Mérida(n=250)
Kuiper	P< 0.01	P< 0.01	P< 0.01
Rao	P< 0.01	P< 0.01	P< 0.01
Rayleigh	0	0	0
Watson	P< 0.005	P< 0.005	P< 0.005
Tests	Villafranca(n=250)	Villanueva (n=250)	Almendr.(n=250)
Kuiper	P< 0.01	P< 0.01	P< 0.01
Rao	P< 0.01	P< 0.01	P< 0.01
Rayleigh	0.31	0.07	0
Watson	P< 0.005	P< 0.025	P< 0.005

Tabla 4-58. Resultados de las pruebas de uniformidad para vectores unitarios de los datos procedentes de la comparación entre diferentes cartografías.

Los resultados anteriores reflejan la inexistencia de una distribución uniforme de los datos.

- Control de la anisotropía. La anisotropía es medida según el error estándar de la media, pues este estadístico se utiliza para determinar el intervalo de confianza del acimut medio para la distribución de von Mises.

Error	10-50 (n=500)	25-50 (n=500)	Mérida(n=250)
$\hat{\sigma}$	12.10°	3.97°	10.13°
Error	Villafranca (n=250)	Villanueva (n=250)	Almen.(n=250)
$\hat{\sigma}$	37.17°	24.67°	13.12°

Tabla 4-59. Resultados del error estándar de la media para datos circulares en los dos ficheros procedentes de la comparativa entre ambas cartografías.

Los resultados no son elevados, proporcionando algún fichero ('25-50') valores pequeños.

- Pruebas de bondad de ajuste a la distribución von Mises. La aplicación de los tests de Kuiper y Watson proporciona una probabilidad de ajuste a la distribución von Mises que se refleja a continuación.

Tests	10-50 (n=500)	25-50 (n=500)	Mérida(n=250)
Kuiper	0.10 > P > 0.05	0.15 > P > 0.10	P<0.01
Watson	P< 0.005	0.1 > P > 0.05	P<0.005
Tests	Villafranca (n=250)	Villanueva (n=250)	Almen.(n=250)
Kuiper	P<0.025	P>0.15	0.10 > P > 0.05
Watson	P<0.005	P<0.005	P<0.005

Tabla 4-60. Resultado de las pruebas de ajuste a la distribución von Mises en comparativa entre cartografías.

Tampoco puede hablarse de ajuste a la citada distribución, como cabía esperar, según los valores del parámetro de concentración de von Mises visto en la tabla 4-57.

- Gráficos de distribución de datos circulares.

La distribución de los datos circulares de las comparativas entre cartografías aparece en la figura siguiente. Los resultados correspondientes a la comparación entre la cartografía a 1/10 000 y 1/50 000 en el apartado a) y la existente entre la 1/25 000 y 1/50 000 en el b), ambos con 500 elementos. Los apartados c) a f) con n=250, representan la comparativa entre hojas del 1/25 000 del IGN a partir de un WMS y las correspondientes a 1/50 000 en diferentes zonas de la provincia de Badajoz denominados respectivamente Mérida, Villafranca, Villanueva y Almendralejo.

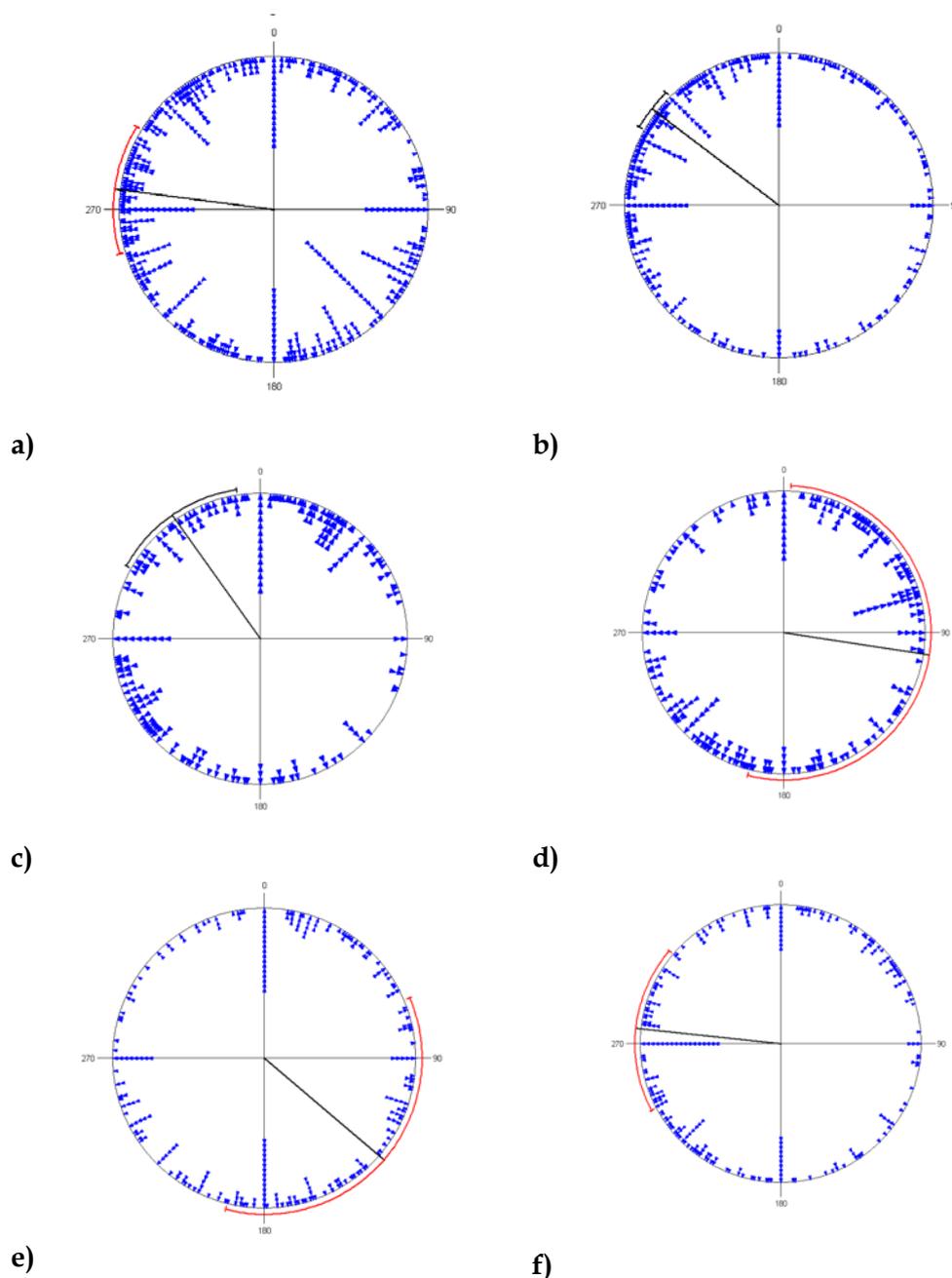


Fig. 4-15. Distribución de datos brutos en la comparativa de diferentes cartografías. Las figuras se corresponden a la distribución de datos circulares de vectores de error al comparar a) la cartografía 1/10 000 y 1/50 000, b) la 1/ 25 000 y 1/50 000, c) a f) la comparativa entre 1/25 000 del IGN a partir de un WMS y el 1/50 000 escaneado en diferentes hojas de la provincia de Badajoz denominados respectivamente Mérida, Villafranca, Villanueva y Al-mendralejo. Obsérvese cómo determinados vectores de error se concentran en direcciones concretas.

Lo más característico de los gráficos anteriores es ver cómo determinados vectores de error se acumulan en direcciones concretas aunque el resto ocupe en mayor o menor medida la circunferencia completa. Según los cálculos anteriores no podemos hablar de uniformidad ni ajuste a una dirección preferente pero sí es destacable cómo en todos los gráficos existen varias direcciones (en la mayoría eje X e Y) donde se acumulan un porcentaje no desdeñable de datos.

4.2.4.2 Resultados del análisis de datos lineales

El acceso a los cálculos del análisis de datos lineales es el mismo que el indicado para datos circulares, esto es:

D://Tesis/Datos/Cartografía/nombre_fichero

Para mayor comodidad se recomienda la consulta de los cálculos resumidos que aparecen en la dirección siguiente:

D://Tesis/Datos/Cartografía/Resumen

- Cálculos de los estadísticos básicos para módulos.

En la tabla siguiente se listan los resultados de la media aritmética (\bar{d}), la desviación típica (s) y los valores máximos y mínimos de los módulos de los ficheros ya indicados.

Fichero	\bar{d}	s	\bar{d} max	\bar{d} min
10-50	17.53	11.35	85.43	0.00
25-50	23.18	13.57	93.20	0.00
Mérida	52.62	23.05	114.77	0.00
Villafranca	46.36	20.51	88.78	0.00
Villanueva	41.12	18.16	79.82	0.00
Almendralejo	49.31	20.75	120.99	0.00

Tabla 4-61. Resultados de los estadísticos básicos para módulos en la comparativa de cartografías de diferentes escalas.

Las desviaciones son relativamente elevadas y los valores máximos también. Habrá que estudiar si son diferencias sistemáticas u *outliers*.

- Determinación de *outliers*.

Junto con la indicación de *outliers* en la tabla siguiente, se anota el porcentaje de observaciones discordantes. Recordemos que el límite para determinar si una observación era discordante o no, se establecía si sobrepasaba el valor de la media aritmética más tres veces la desviación típica. Como con este límite no se acotaban observaciones discordantes en algunos ficheros, se decidió rebajar el condicionante a la media más dos veces el valor de la desviación típica, tal como aparece en la tabla siguiente.

Fichero	\bar{d}	s	outlier ($\bar{d}+3s$)	%	outlier ($\bar{d}+2s$)	%
10-50	17.53	11.35	5	1%	23	4.6%
25-50	23.18	13.57	8	1.6%	16	3.2%
Mérida	52.62	23.05	0	0%	5	2%
Villafranca	46.36	20.51	0	0%	1	0.4%
Villanueva	41.12	18.16	0	0%	8	3.2%
Almendralejo	49.31	20.75	1	0.4%	7	2.8%

Tabla 4-62. Número de *outliers* y porcentaje en cada fichero en la comparativa de cartografías de diferentes escalas.

Los porcentajes de *outliers* para la primera acotación son muy bajos, en especial para las comparativas obtenidas del *WMS*, de ahí que se optara por realizar un segundo cálculo flexibilizando el condicionante. Ahora sí se obtienen porcentajes similares a los conseguidos para el escáner fotogramétrico oscilando entre el 0.4% y el 4.6%.

- Pruebas para la determinación de las características de las distribuciones de datos lineales

La comprobación del ajuste de los datos a diferentes distribuciones se ha realizado por aplicación de las pruebas de bondad de ajuste de Kolmogorov-Smirnov. Los resultados aparecen en la tabla siguiente.

Fichero	p-valor gamma	p-valor Weibull	p-valor normal	p-valor log-normal
10-50	0.13	0.55	0	0.01
25-50	0.10	0.49	0.01	0.01
Mérida	0	0	0.04	0.08
Villafranca	0	0.17	0.13	0.03
Villanueva	0	0.01	0.07	0.14
Almendralejo	0.10	0.51	0.16	0.40

Tabla 4-63. Resultados de las pruebas de K-S para diferentes distribuciones de datos en la comparativa de cartografías de diferentes escalas.

Los resultados son como vemos variables siendo la distribución Weibull a la que mejor se adaptan los datos, tal como ocurría con los módulos de los ficheros del escáner fotogramétrico. La vista de los histogramas de distribución del apartado siguiente ayuda a comprender la disparidad de resultados de éste.

- Histograma de distribución de datos.

Los histogramas de distribución de datos aparecen en el gráfico adjunto siguiendo el mismo orden que en pruebas precedentes.

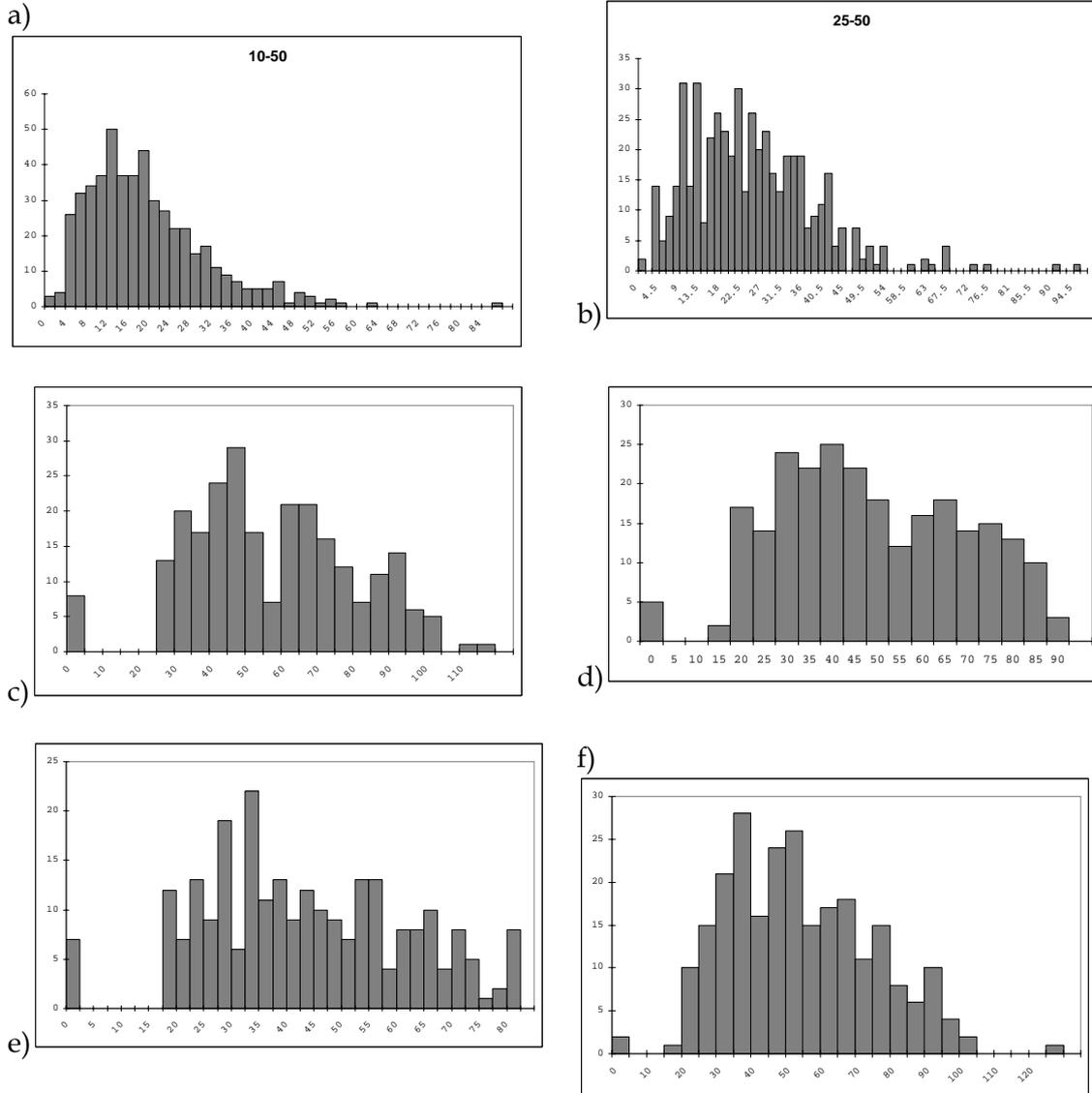


Fig. 4-16. Histogramas de distribución de valores en los ficheros de la comparativa entre diferentes cartografías.

En la figura anterior se representa en los apartados a) y b) los histogramas de distribución de valores para los ficheros '10-50' y '25-50', respectivamente. En los apartados c) a f) la comparativa entre el 1/25 000 del IGN a partir de un WMS y el 1/50 000 escaneado en diferentes hojas de la provincia de Badajoz denominadas respectivamente Mérida, Villafranca, Villanueva y Almendralejo.

En los histogramas anteriores se percibe un carácter asimétrico a la derecha menos obvio que el contrastado en los datos procedentes de la calibración del escáner fotogramétrico (pag. 108). Se aprecia también discontinuidades en los datos de tal forma que, en la mayoría de los ficheros, no existen módulos que tengan una dimensión entre 5 y 15 metros.

4.2.4.3 Resultado del análisis de la distribución conjunta

Los resultados del análisis de la distribución conjunta siguen el mismo orden que el visto en apartado precedentes:

- Correlaciones existentes entre los módulos y los acimutes de la muestra.

En la tabla siguiente sólo se indican los resultados distintos de cero.

Fichero	Valor r correlación
Mérida	0.14
Villafranca	0.12
Almendralejo	0.03

Tabla 4-64. Resultados de las correlaciones en comparativas entre diferentes cartografías.

A tenor de estos resultados no se observa una correlación clara entre los módulos y los acimutes de los vectores de error de posición.

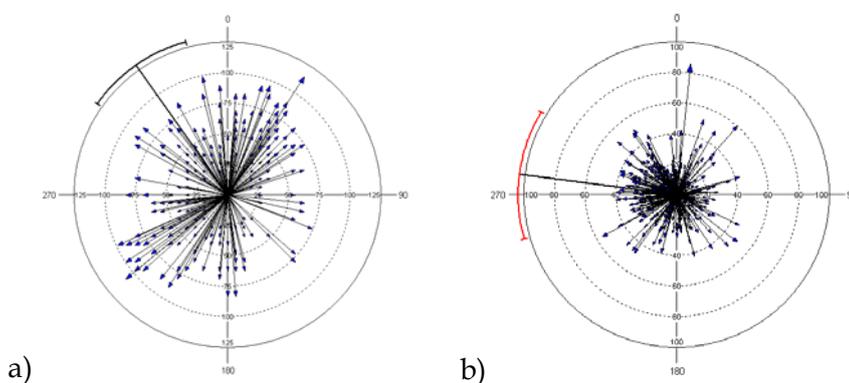


Fig. 4-17. Correlaciones existentes entre módulos y acimutes en los ficheros 'Mérida' y '10-50'. Se han elegido estas dos figuras por presentar la primera el máximo valor de r (0.14) mientras que en la segunda este estadístico es nulo.

En la figura anterior se han representado los ficheros 'Mérida' y '10-50' en los apartados a) y b) respectivamente, por presentar los valores máximo (0.14) y mínimo (0) de correlación entre variables angulares y lineales.

- Diferencias entre los valores de la desviación típica en la dirección X e Y.

Calculada la desviación típica para los valores de X e Y en ambas comparativas se obtienen los resultados que aparecen en la tabla siguiente expresada en metros. Observamos diferencias poco significativas entre los resultados.

σ	X	Y
10-50	14.41	14.64
25-50	17.39	16.54
Mérida	37.05	42.23
Villafranca	33.70	37.73
Villanueva	29.13	33.47
Almendralejo	37.18	36.80

Tabla 4-65. Resultados de los valores de la desviación típica en X e Y para las dos comparativas de cartografías.

Por contra a lo que ocurría en los datos procedentes del escáner fotogramétrico, en los que el error en la dirección Y era mayor que en la X, aquí no hay diferencias acusadas. Recordemos, de los gráficos de distribución de datos circulares (pag. 137), que muchos vectores se acumulaban en las direcciones X e Y al mismo tiempo, por lo que el cálculo de la desviación típica en ambos ejes puede tender a igualarse.

- Determinación de *outliers*.

Al igual que ocurría en los datos procedentes de la calibración del escáner fotogramétrico, se ha establecido un porcentaje del 2% de puntos a considerar como elementos discordantes, obtenidos en función del valor más alto de la media armónica. Dado que contamos con 500 observaciones en dos de las comparativas tratadas, tenemos un total de 10 *outliers* determinados para cada fichero. En el caso de los ficheros de 250 elementos se han acotado 5 *outliers*.

- Mapas de densidad

Existen diferencias entre las distribuciones de los ficheros '10-50' y '25-50' con respecto a los obtenidos a partir de mapas web, independientemente del número de puntos analizados. Los mapas de densidad de los dos primeros siguen una distribución más próxima al círculo. Se ha superpuesto una elipse de probabilidad al 95% sobre los datos del apartado b) para ilustrar mejor la dirección de los datos.

Las distribuciones obtenidas a partir de las comparaciones con cartografía procedente de un WMS siguen una pauta diferente. La distribución es alargada, no circular, y el eje principal de la masa de datos tiene una dirección en torno a la línea 10°-190° aproximadamente.

En la figura siguiente se ilustran estos resultados para los 6 casos tratados. Los *outliers* están indicados en color verde.

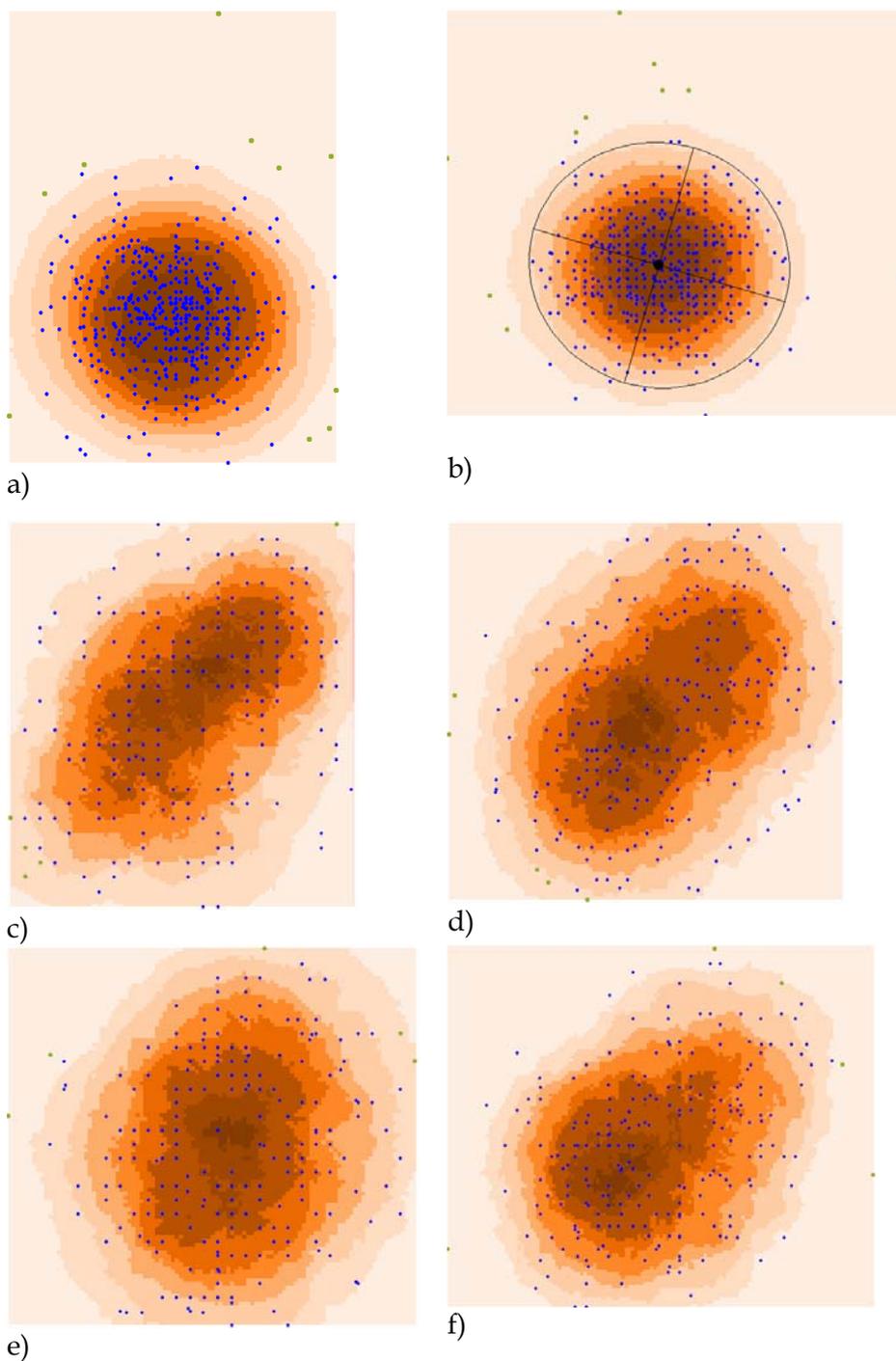


Fig. 4-18. Mapas de densidad y *outliers* en la comparativa entre cartografías. Se indican los mapas de densidad con indicación de *outliers* de los ficheros '10-50' y '25-50' en los apartados a) y b) respectivamente. Los apartados c) a f) corresponden a la comparativa entre el 1/25 000 del IGN a partir de un WMS y el 1/50 000 escaneado en diferentes hojas de la provincia de Badajoz denominadas respectivamente Mérida, Villafranca, Villanueva y Almendralejo.

4.2.5 Conclusiones

Tras la realización de los cálculos correspondientes a la comparativa de cartografías a diferentes escalas, inferimos las siguientes conclusiones para datos circulares, lineales y en el tratamiento conjunto:

4.2.5.1 Conclusiones para datos circulares

El análisis de datos circulares nos lleva a las conclusiones siguientes.

- Estadísticos básicos para vectores unitarios.
 - El acimut medio de las muestras no ofrece valores especialmente significativos para cada una de las 6 comparativas.
 - El módulo medio ofrece un valor bajo en todos los ficheros excepto en el '25-50' ($\bar{R} = 0.43$) lo que implica una mayor concentración de datos en este último conjunto.
 - La desviación estándar circular de la muestra presenta valores altos en todas las muestras lo que supone una dispersión de valores importantes.
 - El parámetro de concentración de von Mises no tiene valores significativos y no sugiere aglutinación de datos en una dirección determinada.
- Pruebas para la determinación de las características de las distribuciones de datos circulares.
 - Las pruebas relativas al control de la uniformidad revelan ausencia de la misma en todos los conjuntos de datos.
 - La anisotropía, medida por el error estándar del acimut medio, tiene valores bajos, lo que al menos, descarta una distribución homogénea de valores.
 - El ajuste a la distribución von Mises ofrece resultados algo heterogéneos pero en ningún caso nos permite afirmar que los datos se adaptan a la distribución citada.
- Gráficos de distribución de datos circulares.

Lo más llamativos de los gráficos consultados (Fig. 4-15) y después de haber descartado numéricamente distribuciones aleatorias o con una dirección claramente preferente, es la existencia de vectores de error acumulados en varias direcciones claramente definidas.

4.2.5.2 Conclusiones para datos lineales

Realizadas las conclusiones para datos circulares exponemos las análogas para datos lineales.

- Estadísticos básicos de los módulos de los vectores.

Las desviaciones típicas de las muestras son relativamente elevadas, así como los valores máximos, que llegan hasta los 121 metros en algún caso. Como existen valores mínimos igual a cero, descartamos los errores sistemáticos en general, aunque pueden afectar a algún conjunto de puntos en concreto.

- Determinación de *outliers*.

Rebajando el límite para determinar observaciones discordantes de la media más 3 veces la desviación típica a sólo 2 veces, el número de *outliers* oscila entre el 0.4% y el 4.6%.

- Pruebas para la determinación de las características de las distribuciones de datos lineales.

Los ajustes de los datos a determinadas distribuciones ofrecen resultados variables, aunque sí se constata un mejor ajuste a distribuciones asimétricas positiva.

- Histogramas de frecuencia de valores.

Ratifican lo indicado en el párrafo anterior sobre las distribuciones asimétricas positivas aunque con intervalos de ausencia de datos en los histogramas (entre los 5 y 15 metros).

4.2.5.3 Conclusiones para la distribución conjunta

Finalizamos ya esta exposición de conclusiones con las obtenidas por el tratamiento conjunto de datos.

- Correlaciones entre los módulos y los acimutes.

No existe una correlación clara entre los módulos y los acimutes en las diferentes muestras tratadas.

- Diferencias entre los valores de desviación típica en la dirección X e Y.

No existen diferencias significativas entre los valores de desviación típica en la dirección X e Y.

- Determinación de *outliers*.

No se aprecian características especiales en la ubicación de las observaciones discordantes.

- Mapas de densidad.

Difieren notablemente los mapas de densidad según la fuente de los datos. En el caso de la comparativa directa entre las cartografías MTN25, MTN50 y a escala 1/10 000, los mapas de densidad tienen forma circular. Los obtenidos por comparación con cartografía del IGN obtenida de un WMS tienen forma alargada con el eje principal de esta distribución en dirección 10° - 190° aproximadamente.

5 Propuesta de norma

La mayoría de los estándares para determinar la exactitud posicional de los datos espaciales requieren una norma para la componente horizontal y otra para la vertical, así como una descripción del método usado para evaluar la exactitud: comparación con una fuente de mayor exactitud, ajuste por mínimos cuadrados, pruebas estadísticas... (NCHRP, 2003:16). Son varias las pruebas existentes en la actualidad al respecto pudiendo citar como más conocidos los tests NMAS (USBB, 1947), EMAS (ASCI, 1983), ASPRS (ASPRS, 1989) o NSSDA (FGDC, 1998), cuyas aplicaciones están relativamente generalizadas (Ariza López, García Balboa *et al.* 2004).

Gracias a herramientas como los receptores GPS y al desarrollo de cálculos computacionales, las pruebas para la determinación de errores de posición están migrando desde el uso exclusivo de elementos puntuales como control al de los elementos lineales, entendiendo que la mayor parte de información en una base de datos espacial es de carácter lineal (Ariza López, 2002; Atkinson Gordo and Ariza López 2002). Se introduce aquí el concepto de **banda de indeterminación** para representar la incertidumbre existente en torno a un segmento del que se conocen las elipses de error de sus extremos.

Sorprende, que pese a la profusión de normas sobre datos geográficos, las referentes a la exactitud posicional sean limitadas. En algunos casos se plantean incluso las llamadas **estimaciones de experto**, tal como se indica en el informe del Grupo de Trabajo del Núcleo Español de Metadatos (NEM):

“No siempre es posible obtener una exactitud posicional, en esos casos se acepta una estimación intuitiva del valor esperado basado en la experiencia previa o unos valores máximos del error. Incluso si esto no es posible se recomienda utilizar los valores por defecto: no aplicable, no disponible.” (NEM, 2005:26)

Otro aspecto que entendemos merece ser resaltado es cómo se contempla el control de la exactitud posicional desde algunos estamentos receptores de trabajos cartográficos en nuestro país, principalmente la administración (Atkinson Gordo, 2005). Así, en la mayoría de pliegos de condiciones técnicas consultados, y en el mejor de los casos, la valoración de la exactitud posicional se limita a comprobar que, en al menos un determinado porcentaje de puntos, la discrepancia de posición con otra fuente de datos de mayor exactitud no sea superior a un determinado valor en un nivel de confianza que se suele situar en el 95% (Fomento, 2007).

En general la mayoría de los tests para el control del error de posición siguen una serie de directrices como son:

- El método para evaluar la calidad de la base de datos es la comparación con fuentes de mayor exactitud.
- Se hereda la filosofía del tratamiento de bases de datos analógicas, siendo muy difícil aglutinar en una sola norma los diferentes tipos de elementos que conforman las bases de datos digitales actuales (JRC 2005; USGS, 1996; USGS, 1999; USGS, 2001).

- Los elementos sobre los que se realiza la comparación son puntos.
- Se analiza la componente horizontal separada de la vertical.
- En el caso del control por puntos se discriminan al menos 20 de fácil identificación. No se cuestiona el número de puntos por superficie o si la distribución y calidad de los mismos es la adecuada, aspecto que dista de estar zanjado.
- Tras determinar la diferencia de posición entre los puntos de la base de datos en cuestión y la fuente de mayor exactitud se obtienen unas diferencias de coordenadas.
- Con estas diferencias de coordenadas se calculan el RMSE o la desviación típica y se aplica algún test estadístico asumiendo la distribución normal de los datos.
- Con un nivel de confianza determinado se acepta o rechaza el control de exactitud.

Como se ha indicado anteriormente, un aspecto relevante a considerar es la elección de los **puntos de control**. La elección de estos puntos en la verificación de la calidad de este tipo de trabajos, repercutirá claramente en el resultado de dicho control, más concretamente las características de estos puntos de control tiene efectos en la **fiabilidad** de las cifras finales de exactitud del control (Li, 1991). Entendemos aquí el concepto de fiabilidad (reliability) propuesto por Li para modelos digitales del terreno (MDT) como 'la probabilidad con la que se ha estimado como correcto el valor de exactitud'. La fiabilidad está acompañada de una incertidumbre que se define como el grado de corrección con el cual el valor de exactitud está definido.

Aunque este trabajo de Li se refiere exclusivamente a los MDT, lo tomaremos como base para ahondar más en el complicado tema de definir las características del número de puntos de control para determinar la exactitud posicional de una base de datos.

Dentro de las **características de los puntos de control** cabe indicar tres aspectos relevantes: a) la exactitud, b) el número y c) la distribución de los puntos. El primer apartado hace referencia a si los puntos están libres de error. En relación al número de puntos es obvio indicar que se trata de conseguir los valores máximos de exactitud con el menor número de puntos de control posible o, en su caso, determinar cuál es el número mínimo de puntos de control requeridos para un grado de fiabilidad determinado para las estimaciones de exactitud. Sobre este difícil equilibrio de máximos y mínimos se ha escrito bastante, entendiéndolo la autora de esta Tesis que no puede considerarse un tema zanjado (Cuartero and Felicísimo, 2003). Así, en el caso concreto de los MDE, los estándares del USGS hablan de tomar al menos 28 puntos, 20 interiores y los 8 restantes en los bordes (USGS, 1997). Ley (Ley, 1986) indica que son necesarios 150 puntos para garantizar que la exactitud tenga una desviación estándar del 10%. Más tarde Li (Li, 1991), profundizando en el tema, desarrolla la siguiente expresión, partiendo de que los puntos de control están libres de error:

$$R(SD) = \frac{1}{\sqrt{2(n-1)}} 100\%$$

Ec. 5-1. Influencia del tamaño de la muestra en la fiabilidad del valor de la desviación estándar.

Donde $R(SD)$ es la influencia del tamaño de la muestra en la fiabilidad de la estimación de la desviación estándar y n el número de puntos utilizados. Despejando el valor de n tenemos que:

$$n = \frac{1}{2R^2(SD)} + 1$$

Ec. 5-2. Determinación del número de puntos de control en función de la fiabilidad de la estimación de la desviación estándar.

Así, si se pretende una fiabilidad del 10%, el número de puntos de control necesarios es 51.

Otra expresión obtenida por Li, para estudiar el tamaño de la muestra en la fiabilidad de la exactitud, es la que aparece en la ecuación siguiente:

$$n = \frac{Z_r^2 \cdot SD^2}{S}$$

Ec. 5-3. Influencia del tamaño de la muestra en la fiabilidad de la exactitud.

Donde Z_r representa el valor límite dentro del cuál los valores de la variable estudiada, en este caso diferencia de altura, entran con una probabilidad r , siendo S el grado de exactitud para la estimación de la media aritmética. Así, suponiendo un conjunto de datos con una desviación estándar de 0.59 m, con un valor de S de ± 0.05 metros y con un nivel de confianza del 95% ($Z_r = 1.96$ en la tabla de distribución normal) se necesitarían 535 puntos de control (Li, 1991:1336).

Ahondando en el tema, proponemos realizar el siguiente cálculo, que aparece reflejado en la figura 5-1. En el fichero procedente de los datos cartográficos denominado 'Mérida' consideramos la muestra total de 250 módulos de los vectores de error ya explicados. Se han escogidos de forma aleatoria grupos de módulos con diferentes tamaños de muestra (5, 10, 25, 50, 75, 100, 150 y 200 elementos) para calcular la media y la desviación típica de la media. Cada operación anterior se repitió 10 veces con la finalidad de obtener la media aritmética y la desviación típica de estas 10 reiteraciones para diferentes tamaños de muestra. Observamos en el gráfico siguiente como a medida que aumenta el tamaño de la muestra tiende a disminuir el valor medio, al tiempo que también disminuye la desviación típica, representada en la gráfica como un intervalo de confianza de $\pm\sigma$.

Es obvia la relación entre el tamaño de la muestra y la probabilidad con la que se ha estimado como correcto el valor de exactitud. Entendemos que el número de puntos de control debe venir en función del grado de fiabilidad requerido, y no como sucede en la mayoría de las normas sobre exactitud espacial que definen un valor fijo que generalmente se establece en 20 puntos.

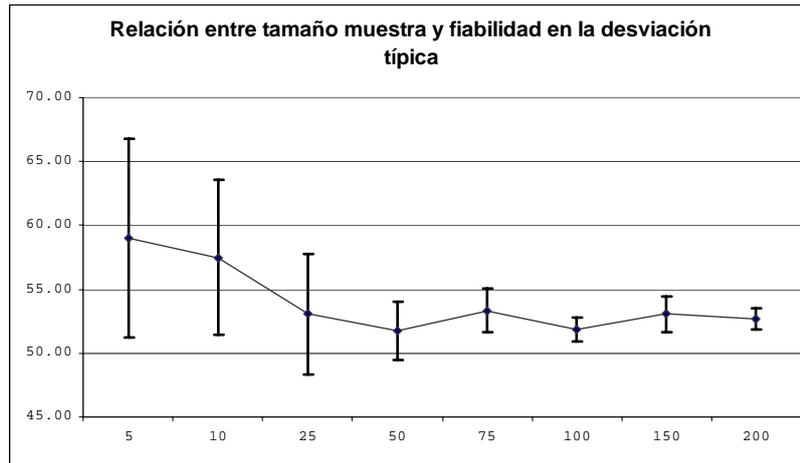


Fig. 5-1. Relación entre el tamaño de la muestra y fiabilidad en el valor de la desviación típica.

A medida que aumenta el tamaño de la muestra disminuye el intervalo de confianza en torno a la media de valor $\pm\sigma$.

Por último la distribución y situación de los puntos de control es tema no sólo de controversia sino también de errores de bulto importantes: elegir como puntos de control los utilizados para las transformaciones, tomar exclusivamente los vértices geodésicos que, como sabemos, se encuentran en lugares elevados o elegir puntos de forma aleatoria sin tener en cuenta que la metodología de toma de datos pueda hacer que determinadas direcciones tengan un mayor error. Tradicionalmente los puntos de control se han elegido o bien formando una retícula o de forma aleatoria, entendiendo este último método que cualquier punto tenía la misma probabilidad de ser elegido como punto de control. En ninguno de los dos casos se tiene en cuenta las particularidades del terreno y la distribución de datos espaciales.

A tenor de lo dicho anteriormente en relación a las **características de los puntos de control** entendemos que dichas características afectan claramente a la verificación de la calidad del control de exactitud posicional. Recordemos los cálculos realizados en esta Tesis con diferentes tamaños de muestra donde veíamos cómo variaban los estadísticos básicos y las pruebas para determinar las características de las distribuciones según lo hacía el tamaño de muestra utilizado.

Proponemos, por tanto, definir las características de los puntos de control en los siguientes términos:

- Exactitud. Supone conocer si los puntos están exentos de error. Obviamente esto no será posible, aunque estemos tratando con fuentes de mayor exactitud y habrá que indicar el grado de fiabilidad determinado entendiendo este término, según la RAE, como 'la probabilidad de buen funcionamiento de algo'. Matizamos mejor la idea adaptando la expresión propuesta por Li (1991) diciendo que la fiabilidad es 'la probabilidad con la que se ha estimado como correcto el valor de exactitud'. Será necesario previamente establecer el nivel de probabilidad requerido.
- Número de puntos de control. Será establecido también según un grado de fiabilidad determinado, al igual que la exactitud. Parece conveniente

indicar varios niveles de fiabilidad según objetivos de calidad para dar flexibilidad a la norma y hacerla apta a diferentes situaciones. El gráfico 5-1 es bastante ilustrativo al respecto pues muestra la relación entre el número de elementos de la muestra y la fiabilidad en la determinación del valor de la desviación típica para la media aritmética de los errores de posición.

- **Distribución.** Se debe tener en cuenta las características del terreno, situando la mayoría de los puntos en las zonas de mayor densidad de elementos, además deben estar representadas todas las estructuras de la zona a comprobar. La distribución de puntos no debe seguir ni una pauta sistemática ni una distribución aleatoria.

En cualquier caso, y en relación al control del error de posición, entendemos que el problema dista de estar solucionado de forma plena, pues como se ha dicho a lo largo de la Tesis, la consideración del error de posición hasta la fecha es sobre todo escalar, no vectorial. Son pocas las referencias a la aplicación de datos circulares en la determinación de la exactitud posicional (Ariza López, 2002:143; DMA,1991:36) y menos aún las aplicaciones directas (Felicísimo, Cuartero *et al.*, 2006) . Por todo lo dicho hasta el momento, sugerimos como documento de discusión el siguiente borrador de norma para el control del error de posición en dos dimensiones en bases de datos espaciales:

Control integral del error de posición en 2D para bases de datos espaciales

1 Ámbito

La presente norma especifica el procedimiento a seguir para controlar el error de posición en dos dimensiones en bases de datos espaciales. Esta norma se utilizará para evaluar y presentar informes sobre la exactitud de los datos espaciales.

2 Campo de aplicación

Esta norma evalúa y presenta informes sobre la exactitud de datos geoespaciales georeferenciados en dos dimensiones obtenidos a partir de fotografías aéreas, imágenes de sensores aerotransportados y levantamientos topográficos y geodésicos por métodos clásicos o mediante técnicas de navegación global por satélite (GNSS).

Se establece en términos de exactitud relativa entendiendo este valor como un estadístico que da la incertidumbre de la posición de dos puntos después de haber eliminado los efectos de los errores comunes. Por el contrario, la exactitud absoluta es un estadístico que da la incertidumbre de un punto respecto a un datum. Implica considerar los efectos de todas las fuentes de error, sistemáticas y aleatorias y se establece en términos de dos componentes: horizontal y vertical.

3 Referencias

Effects of check points on the reliability of DTM accuracy estimates obtained from experimental tests. Zhilin Li (1991).

Estándares de precisión para el geoposicionamiento. Instituto Colombiano de Normas Técnicas y Certificación (2004).

National Standard for Spatial Data Accuracy. Federal Geographic Data Comitee (1998).

Standars for digital raster graphic. U. S. Geological Survey (2001).

Vocabulario Internacional de Metrología. Ministerio de Fomento (2000).

Norma ISO 19115: 2003 Información geográfica. Metadatos.

4 Definiciones

Las definiciones utilizadas en esta norma son las ya indicadas en los apartados 1 y 2 de esta tesis relativas a los siguientes términos:

Dato, base de datos, datos espaciales, vector de error, magnitud, valor de una magnitud, valor verdadero, valor convencionalmente verdadero, medición, método de medida, procedimiento de medida, mensurando, resultado de la medición.

Error aleatorio, error sistemático, exactitud, precisión e incertidumbre se establecerán tanto para datos unidimensionales como bidimensionales.

A las definiciones ya mencionadas se une el término fiabilidad.

Fiabilidad: probabilidad con la que se ha estimado como correcto el valor de exactitud.

5 Procedimiento

5.1 General

El método empleado para desarrollar la norma es la comparación con fuentes de mayor exactitud y posterior análisis estadístico de las diferencias. Esta fuente de mayor exactitud debe ser obtenida de forma separada de la base de datos que se desea comprobar y sus características indicadas en el apartado de metadatos.

Esta norma no definirá umbrales de exactitud de los valores. Cada entidad establecerá este umbral de acuerdo con las necesidades de sus productos y aplicaciones.

5.2 Puntos de control

Los puntos de control cumplirán una serie de características como son:

- Exactitud

Los puntos de control medidos en la fuente de mayor exactitud no estarán nunca exentos de error. En cualquier caso se indicará un grado de fiabilidad determinado que repercutirá en el número de puntos elegidos.

- Número

El número de puntos de control vendrá definido por la fiabilidad requerida, al igual que la exactitud. Es necesario indicar varios niveles de fiabilidad según objetivos de calidad para dar flexibilidad a la norma y hacerla apta a diferentes situaciones.

- Distribución

Serán representativos de los elementos presentes en la zona y claramente identificados. No se considerarán como puntos de control los vértices geodésicos ni los puntos que se utilicen para transformaciones. El número de los puntos de control será mayor en zonas con más densidad de elementos y menor en áreas de poco interés, evitando en cualquier caso una distribución de puntos claramente aleatoria o sistemática.

5.3 Niveles de calidad y tipo de datos

Los niveles de calidad lo establecerá cada entidad de acuerdo con las necesidades de sus productos y aplicaciones y el tipo de dato utilizado. A cada nivel le corresponde un objetivo de calidad.

En cuanto al tipo de datos, establecemos los siguientes:

- a) Datos vectoriales
- b) Datos raster. Imagen rectificadas, georreferenciadas de un mapa escaneado o de salidas de procesos de revisión de mapas digitales. Imágenes obtenidas por teledetección.
- c) Ortofotos.
- d) Bases de datos analógicas.
- e) Modelos digitales de elevaciones. (Aunque la norma es para datos en 2D, se introduce este tipo de datos por su importancia y por que esta norma debe formar parte del caso más general que trate los datos en 3 dimensiones).

5.4 Métodos de comprobación

Una vez definidos los puntos de control se obtienen las coordenadas X e Y de los citados puntos en la base de datos tratada y en la fuente de mayor exactitud. Con esas coordenadas se calculan los siguientes valores para cada pareja de coordenadas de puntos homólogos:

- a) La diferencia de coordenada en X
- b) La diferencia de coordenada en Y
- c) El acimut
- d) La distancia reducida.

La diferencia de posición entre cada punto de control medido en la base de datos y en la fuente de mayor exactitud genera un vector de error. El control de exactitud posicional debe tratar el comportamiento de los vectores de error tanto en su componente angular como en módulo.

Los datos se someterán a la siguientes pruebas:

a) Pruebas para datos circulares.

La llamada estadística circular permite el tratamiento de datos angulares realizando el análisis estadístico de ángulos. En estas pruebas estos ángulos vienen definidos por el **acimut** (θ_i) que es el ángulo plano medido en sentido horario desde la dirección del eje Y en la posición verdadera hasta el vector formado por la posición verdadera y la medida. Cada valor θ_i representa un vector de módulo unidad. Si denominamos por (Xv, Yv) la posición verdadera de un punto sobre la base de datos de mayor exactitud y (Xm, Ym) la posición medida del mismo punto sobre la base de datos a comprobar tenemos que:

$$\theta = \text{arctg} \left(\frac{Xv - Xm}{Yv - Ym} \right)$$

Tendremos así una muestra de n acimutes θ_i sobre los que realizaremos los siguientes cálculos. En primer lugar calcularemos el **vector resultante** que se obtiene al sumar vectorialmente todos los vectores de la muestra. La dirección media del vector resultante es el **acimut medio** ($\bar{\theta}$).

El **módulo del vector resultante** (R) es la longitud del vector resultante. Se calcula a partir de las ecuaciones siguientes donde C y S son la suma de todos los cosenos y senos de los acimutes de la muestra, respectivamente.

$$C = \sum_{i=1}^n \cos \theta_i \quad ; \quad S = \sum_{i=1}^n \text{seno} \theta_i \quad ; \quad R = \sqrt{C^2 + S^2}$$

El **módulo medio** (\bar{R}) es la longitud media del módulo del vector resultante. Su valor oscila entre 0 y 1. Si es 1 implica que todos los vectores son coincidentes, pero si es 0 no supone necesariamente una distribución uniforme. Si n es el número de vectores se tiene que:

$$\bar{R} = \frac{R}{n}$$

La **varianza circular de la muestra** (V), en similitud con la varianza de datos lineales, es un indicador de dispersión angular de forma que cuanto más pequeño sea este valor, más concentrados están los datos. Su valor oscila entre 0 y 1, pero al igual que ocurría con el módulo medio, $V=1$, no implica una distribución uniforme. Su expresión es:

$$V = 1 - \bar{R}$$

La **desviación estándar circular de la muestra** (ν) es resultado de una raíz cuadrada por analogía a lo que ocurre en datos lineales aunque en datos circulares queda definida por:

$$\nu = [-2 \log(1 - V)]^{1/2}$$

El primer indicador de la distribución de datos circulares será una medida de tendencia central, el acimut medio, más una de dispersión, la desviación estándar de la muestra.

$$\bar{\theta} \pm \nu$$

La distribución de datos circulares se presupone de entrada isotrópica, es decir, que no existen direcciones preferentes para este tipo de datos. Para comprobar esta casuística se aplicará un test de uniformidad o isotropía (test de Rayleigh) y se calculará un parámetro de concentración de datos conocido como von Mises.

El test de **Rayleigh** establece que la hipótesis de uniformidad se rechaza si la longitud media del vector resultante de la muestra es muy grande. El estadístico a calcular es:

$$Z = n\bar{R}^2$$

La probabilidad P de aceptar la hipótesis de uniformidad, en caso de ser $n \geq 50$:

$$P = \exp(-Z)$$

Este test supone que un valor del módulo medio más largo implica una concentración mayor en torno a la media y por tanto menos probabilidad de que los datos estén uniformemente distribuidos.

El parámetro de concentración de von Mises (κ) mide la variación de la distribución en relación con un círculo perfecto, lo que sería una distribución uniforme. Cuando este parámetro κ , tiende a cero, la distribución converge a la distribución uniforme; en caso de tender a infinito, la distribución se concentra en la dirección media del vector

resultante. La expresión para su valor estimado es:

$$\hat{\kappa} = \begin{cases} 2\bar{R} + \bar{R}^3 + 5\bar{R}^5 / 6 & \bar{R} < 0.53 \\ -0.4 + 1.39\bar{R} + 0.43/(1 - \bar{R}) & 0.53 \leq \bar{R} \leq 0.85 \\ 1/(\bar{R}^3 - 4\bar{R}^2 + 3\bar{R}) & \bar{R} \geq 0.85 \end{cases}$$

Las pruebas anteriores para datos circulares se proponen sean aplicadas a la muestra total de puntos de control y a, al menos, 3 submuestras de diferentes tamaño, elegidos en las zonas de mayor densidad de elementos topográficos o en las zonas más conflictivas. Obviamente lo que se pretende es contrastar un análisis global con varios locales, pero este hecho queda también en función de los niveles de calidad establecidos o exigidos por las entidades implicadas.

b) Pruebas para datos lineales

Dado un conjunto de n observaciones, en este caso módulos de vectores, $d_1, d_2, \dots, d_{n-1}, d_n$, la media aritmética, \bar{d} , se obtiene según la expresión siguiente.

$$\bar{d} = \frac{\sum_{i=1}^n d_i}{n}$$

La interpretación de la media como centro de valores se apoya en una propiedad de la media aritmética que indica que la suma de las desviaciones de un conjunto de observaciones con respecto a la media es cero.

La desviación típica de una muestra de elementos toma valores no negativos y mide, como sabemos, la dispersión de los datos: a mayor desviación típica, mayor dispersión de datos. Se calcula por la expresión siguiente, utilizando la nomenclatura ya indicada para los módulos:

$$s = \sqrt{\frac{\sum_{i=1}^n (d_i - \bar{d})^2}{n - 1}}$$

Así el indicador de la distribución de datos lineales será una medida de tendencia central, la media aritmética, más una de dispersión, la desviación típica de la muestra.

$$\bar{d} \pm s$$

Estimaremos también los *outliers* en datos lineales, entendiendo que son aquellos que sobrepasan el valor de la media aritmética más 3 veces la desviación típica.

$$outliers \geq \bar{d} \pm 3s$$

De nuevo se sugiere que las pruebas anteriores para datos lineales sean aplicadas a la muestra total de puntos de control y a, al menos, 3 submuestras de diferentes tamaños elegidos en las zonas de mayor densidad de elementos topográficos o en las zonas

más conflictivas. De la misma forma que en datos circulares el número de submuestras queda en función de los niveles de calidad establecidos o exigidos por las entidades implicadas.

c) Pruebas conjunta

Consistirán en crear una nube de puntos formados por los nodos finales de los vectores de error. A partir de esta nube de puntos se calcularán mapas de densidad como el número de puntos por unidad de área que permiten localizar zonas donde se producen los mayores errores.

Estos mapas de densidad permitirán catalogar como *outliers* a las observaciones más alejadas de la masa principal de puntos.

Al igual que en los apartados anteriores se recomienda realizar junto con el análisis global varios locales.

5.5 Cálculos

Los cálculos a realizar indicados en apartados anteriores son:

En primer lugar las diferencias de coordenadas en X, e Y, el acimut y el módulo de cada vector de error.

Para el análisis circular:

El acimut medio ($\bar{\theta}$), el módulo medio (\bar{R}), la varianza circular de la muestra (V), la desviación estándar circular de la muestra (v). Aplicación del un test de uniformidad de datos (test de Rayleigh) y cálculo del parámetro de concentración de von Mises (κ).

Para el análisis lineal, los cálculos son:

La media aritmética (\bar{d}), la desviación típica (s) y el cálculo de *outliers*.

Para el análisis conjunto, los cálculos a realizar son:

Elaboración de mapas de densidad y determinación de *outliers*.

Las fórmulas son las explicadas en el apartado anterior.

5.6 Análisis estadístico

El análisis estadístico se centrará en determinar hasta dónde se aceptan los valores de dispersión, isotropía, etc según los niveles y tipos de datos utilizados, tal como se indica en el apartado siguiente.

5.7 Errores máximos

Como ya se indicó anteriormente esta norma no definirá umbrales de exactitud de los valores, siendo en cada caso la entidad receptora del control de calidad quien establecerá este umbral de acuerdo con las necesidades de sus productos y aplicaciones. En cualquier caso, se hacen las siguientes recomendaciones:

- Especificar el tipo de dato utilizado (vectorial, raster, ortofoto, base de datos analógica o MDE).
- Asociar unos niveles de calidad a cada tipo de datos según aplicaciones. Por ejemplo, bajo, medio y alto.
- Establecer valores máximos para el análisis circular, lineal y conjuntamente.
- Es posible que un determinado control de calidad no supere todas las pruebas a la vez. Esto debe venir contemplado en los niveles de calidad. Por ejemplo, una base de datos puede cumplir los requisitos para el control circular pero no para el lineal por lo que esa base será apta para un determinado nivel de calidad previamente establecido.

6 Metadatos

La información facilitada concierne tanto a la base de datos a testear como a la base de mayor exactitud. Reunir información clara y fiable sobre el conjunto de datos es en la mayoría de los casos caro y complejo. Existen normativas muy claras en cuanto a la estructura y composición de los metadatos (ISO 19115, principalmente) cuya aplicación puede exceder notablemente en determinados controles de calidad cartográficos, por lo que sugerimos que, al menos, aparezca información relativa a los siguientes aspectos tanto para la base de datos a tratar como para la fuente de mayor precisión:

- Fecha de creación y periodo de actualización
- Creador del conjunto de datos
- Extensión geográfica de los datos
- Modo de obtención de la información
- Formato de la información.
- Exactitud posicional (por ser este un control de posición) en especial de la fuente de datos de mayor exactitud.

7 Informe final

La base de datos comprobada, una vez superadas las pruebas anteriores, se acompañará del siguiente informe:

La base de datos espaciales ' _____ ' formada por datos de tipo ' _____ ' ha sido comprobada según la "Norma para el control integral del error de posición en 2D en bases de datos espaciales" en el nivel de calidad ' _____ ' considerando apta en el control del error posicional en datos circulares y lineales.

Se debe hacer constar los umbrales de exactitud definidos para cada nivel de calidad.

8 Anexo

Se debe incluir un ejemplo práctico de aplicación donde se observen todas las fórmulas y pasos seguidos para un nivel de calidad dado.

9 Información complementaria

Entendemos como información complementaria:

- Los acrónimos utilizados a lo largo de la norma
- Normas complementarias a las que se haga referencia
- Previsiones de cambio y/o actualización en la base de datos, que sugieran un control en fechas determinadas.

6 Discusión

La exactitud de los datos espaciales puede abordarse desde perspectivas diferentes. Una de ellas es la mejora tecnológica, lo que permite el desarrollo de métodos de captura más exactos y más precisos. Otro enfoque complementario del anterior es el control de las medidas, para lo cual es necesario desarrollar pruebas eficaces y fiables.

En este trabajo se han realizado pruebas sobre conjuntos de datos espaciales que han dado como resultado la convicción de que, mientras las tecnologías de captura de datos han avanzado, las de control de exactitud apenas se han desarrollado en las últimas décadas. Esta situación de estancamiento se refleja en las normas existentes y en su manera de abordar los controles estadísticos.

Actualmente los datos espaciales son accesibles en gran medida a través de las llamadas Infraestructuras de Datos Espaciales y su difusión refuerza la importancia de conocer su exactitud ya que pueden ser utilizadas por mucha gente para fines muy diferentes. Los metadatos de calidad son escasos incluso en el mejor de los casos previstos en los estándares pero hemos visto en este trabajo que, además, no podemos confiar demasiado en que las pruebas que normalmente se aplican detecten los problemas existentes en los datos. Los mínimos tamaños de las muestras de control reducen la fiabilidad de los resultados y los tests paramétricos y globales no son capaces de discernir variaciones dentro de la información.

Los ejemplos usados en este trabajo no son más que una fracción mínima de los datos reales pero nos ha sido imposible obtener datos de control en cantidad suficiente y de casos diversos, lo cual parece ser síntoma de que no hay una excesiva preocupación por este tipo de controles. Probablemente se confía en la tecnología y en que los resultados serán suficientemente buenos por defecto y tal vez ese sea el motivo de la llamativa ausencia de metadatos sobre este tema, casi completa.

Hemos discutido a lo largo de la realización de esta Tesis sobre lo evidente que es que la exactitud de los datos espaciales debe tratarse como un problema vectorial. El que sea evidente ha llevado a veces a recapacitar sobre si lo que se estaba haciendo no era una obviedad. Pero parece que, si lo es, es una obviedad necesaria porque en la bibliografía puede verse la escasez de trabajos que utilicen estas técnicas en su desarrollo.

De nuevo aquí podríamos establecer dos campos de actuación diferentes. El primero es el técnico y el segundo el científico. ¿Por qué no se ha adoptado una norma más completa en la tarea diaria de empresas y administraciones? Suponemos que este tipo de tareas está en equilibrio entre la mayor corrección del trabajo y el coste económico. Se ha llegado a una solución de compromiso aunque sin demasiada reflexión ya que los pliegos de condiciones se copian frecuentemente entre unos y otros sin cambios en su redacción con lo que las administraciones asumen que lo que hay es de calidad suficiente y no es necesario profundizar más en el asunto.

Por la parte científica parece que el mayor problema es el desconocimiento de la estadística circular (o esférica) por parte de los especialistas en la materia. Los estadísticos no tratan el tema cartográfico y los cartógrafos en sentido amplio no se plantean los cambios porque nunca han tenido contacto con esta rama de la estadística.

Creemos que lo desarrollado en esta tesis apoya la idea de que algún cambio es necesario si consideramos que el control de la exactitud merece algo de atención.

7 Líneas futuras

Parece obvio continuar y terminar el tratamiento del error de posición en bases de datos espaciales con el análisis en tres dimensiones, toda vez que se ha expuesto y acotado la problemática en el plano.

La principal dificultad con la que topamos es la de encontrar una población de errores lo suficientemente amplia y fiable como para poder hacer los análisis pertinentes. Ya se constató este problema en el análisis 2D y se resolvió, como sabemos, utilizando dos fuentes de datos: los vectores de error procedentes de la calibración de un escáner fotogramétrico y la comparativa entre cartografías a diferentes escalas.

Los datos utilizados para el análisis de error en 3D han sido facilitados por la Secretaría General de Políticas Ambientales de la Consejería de Medio Ambiente de la Junta de Andalucía. Los datos aportados proceden de vuelos fotogramétricos para aerotriangulación, en concreto, son las coordenadas GPS²³ correspondientes a la posición del avión en el momento de la toma de cada fotograma realizado en el vuelo. Se analizarán las diferencias en X, Y y Z de las coordenadas GPS de partida corregidas y las coordenadas finales ajustadas, obtenidas en la toma de cada foto. En total son 75 ficheros en formato Excel, en general un fichero por bloque de aerotriangulación, con un número variable de fotograma por bloque: 63 en el menor de los casos y 713 en el mayor. Esto genera 27558 fotogramas que se traduce en el mismo número de vectores de error en tres dimensiones.

La metodología de trabajo debe ser similar a la ya descrita con las salvedades propias de introducir una nueva dimensión. Esto implica introducir el concepto de **dato esférico** (Fisher.N.I., Lewis *et al.*, 1987) entendiendo este dato como la orientación de una línea en el espacio. En algunos contextos hablamos de **vectores** como líneas orientadas con sentido y en otros de **ejes**, que tendrían orientación pero sin un sentido definido.

Al igual que ocurría con los datos circulares, los datos esféricos se plantean con módulo unitario. Habrá que analizar, por tanto, la componente angular y lineal por separado y realizar un tratamiento conjunto, de la misma forma que se hizo con los datos en 2D.

Análogamente se debe finalizar el estudio planteando una norma para el tratamiento de errores de posición en 3D. Esta norma puede ser el caso genérico del que se derive la norma en 2D o bien la proposición de un estándar diferente al establecido en 2D para aplicar exclusivamente a este tipo de datos.

Otro aspecto que no ha sido tratado en esta Tesis y que se considera pendiente de definir es la forma en cómo se propaga la incertidumbre de los datos en operaciones espaciales. Ya se comentó que esta información debe formar parte de los metadatos, siendo la asociación dato-metadato un factor relevante para la gestión de las bases de datos espaciales.

²³ Se entiende coordenadas obtenidas con receptores GPS en su sistema de referencia.

8 Anexo

En este punto se acompaña a la Tesis de una serie de información que, aunque debe aparecer en el trabajo, puede considerarse de importancia secundaria o que interfiere en un más cómodo análisis del texto.

Como también se ha indicado a los largo de la Tesis en el CD adjunto a la misma aparecen todos los datos y cálculos mencionados.

La información del anexo se centra en los siguientes apartados:

- Revisión de estadísticos y parámetros de datos circulares
- Estándares de datos espaciales
- Material e instrumental utilizado
- Aplicaciones informáticas
- Acrónimos utilizados

8.1 Estadísticos y parámetros de datos circulares

Estadísticos de la muestra:

- **Acimut** (θ_i): ángulo plano medido en sentido horario desde la dirección del eje Y en la posición verdadera hasta el vector formado por la posición verdadera y la medida. Cada valor θ_i representa un vector de módulo unidad.
- **Acimut medio** ($\bar{\theta}$): El vector resultante se obtiene al sumar vectorialmente todos los vectores de la muestra. La dirección media del vector resultante es el acimut medio.
- **Módulo del vector resultante** (R): Es la longitud del vector resultante. Se calcula a partir de las ecuaciones siguientes donde C y S son la suma de todos los cosenos y senos de los acimutes de la muestra, respectivamente.

$$R^2 = C^2 + S^2 \quad ; \quad C = \sum_{i=1}^n \cos \theta_i \quad ; \quad S = \sum_{i=1}^n \text{seno} \theta_i$$

- **Módulo medio** (\bar{R}): Es la longitud media del módulo del vector resultante. Su valor oscila entre 0 y 1. Si es uno implica que todos los vectores son coincidentes, pero si es cero no supone necesariamente una distribución uniforme. Si n es el número de vectores se tiene que:

$$\bar{R} = \frac{R}{n}$$

- **Varianza circular de la muestra (V):** Tiene su similitud a la varianza de datos lineales, de tal forma que cuanto más pequeño sea este valor, más concentrados están los datos. Su valor oscila entre 0 y 1, pero al igual que ocurría con la longitud media del vector resultante, $V=1$, no implica una distribución uniforme. Su expresión es:

$$V = 1 - \bar{R}$$

- **Desviación estándar circular (ν):** Se define por

$$\nu = \{-2 \log(1 - V)\}^{1/2}$$

- **Coseno medio y seno medio (\bar{C}, \bar{S}):** Obtenidos a partir de las expresiones siguientes son necesarios para el cálculo de otros estadísticos importantes.

$$\bar{C} = C/n \quad \bar{S} = S/n$$

En relación a estos valores se cumple que:

$$\bar{R} = (\bar{C}^2 + \bar{S}^2)^{1/2}$$

- **Momentos coseno y seno p-ésimos (\bar{C}_p, \bar{S}_p):** Se indica a continuación la expresión para el coseno, siendo análoga la del seno.

$$\bar{C}_p = \frac{1}{n} \sum_{i=1}^n \cos p\theta_i$$

- **Momento trigonométrico descentrado (m'_p):** Los momentos trigonométricos describen características de la muestra o de la población. La expresión para el momento p-ésimo, siendo i la raíz de -1 , es:

$$m'_p = \bar{C}_p + i\bar{S}_p$$

- **Momento trigonométrico centrado (m_p):** Relativo a la dirección media resultante.

$$m_p = \frac{1}{n} \sum_{i=1}^n \cos p(\theta_i - \bar{\theta}) + i \frac{1}{n} \sum_{i=1}^n \sin p(\theta_i - \bar{\theta})$$

- **Dispersión circular de la muestra ($\hat{\delta}$):** En el caso de una distribución uniforme este valor es infinito.

$$\hat{\delta} = (1 - \hat{\rho}_2)/(2\bar{R}^2) \quad \hat{\rho}_2 = (\bar{C}_2^2 + \bar{S}_2^2)^{1/2}$$

- **Asimetría (\hat{s}):** Su expresión es

$$\hat{s} = [\hat{\rho}_2 \sin(\hat{\mu}_2 - 2\hat{\mu})]/(1 - \bar{R})^{3/2}$$

Teniendo en cuenta que, al denotar por $\hat{\cdot}$ a un valor de estimación, tenemos que:

$$\hat{\mu} = \bar{\theta} \quad \hat{\rho} = \bar{R}$$

- **Curtosis o aplastamiento (\hat{k})**: Se obtiene a partir de

$$\hat{k} = [\hat{\rho}_2 \cos(\hat{\mu}_2 - 2\hat{\mu}) - \bar{R}^4] / (1 - \bar{R})^2$$

- **Mediana de la muestra ($\tilde{\theta}$)**: Se calcula minimizando la función siguiente

$$d(\theta) = \pi - \frac{1}{n} \sum_1^n |\pi - |\theta_i - \theta||$$

- **Dirección modal ($\tilde{\theta}$)**: Es la dirección que presenta una frecuencia más alta de valores.

Parámetros de la población

- **Acimut medio (μ)**. La estimación de este valor es igual a la dirección media del vector resultante de la muestra, como ya se indicó.

$$\mu = \bar{\theta}$$

- **Módulo medio (ρ)**. La estimación de este valor es igual a la longitud media del vector resultante de la muestra.

$$\hat{\rho} = \bar{R}$$

- **Estimación p-ésima de ρ ($\hat{\rho}_p$)**. Se calcula según

$$\hat{\rho}_p^2 = \bar{C}_p^2 + \bar{S}_p^2$$

- **Varianza circular (ν)**. De forma análoga a los valores de la muestra, tenemos que:

$$\nu = 1 - \rho$$

- **Desviación estándar circular (σ)**. Su expresión es:

$$\sigma = [-2 \log(1 - \nu)]^{1/2}$$

- **Dispersión circular (δ)**. Se obtiene según:

$$\delta = (1 - \rho_2) / (2\rho^2)$$

- **Asimetría (s)**

$$s = \beta_2 / (1 - \rho)^{3/2}$$

- **Curtosis o aplastamiento (K)**

$$K = (\alpha_2 - \rho^4)/(1 - \rho)^2$$

Otras expresiones

- **Parámetro de concentración (κ):** Mide la variación de la distribución en relación con un círculo perfecto, lo que sería una distribución uniforme. Cuando este parámetro κ , tiende a cero, la distribución converge a la distribución uniforme; en caso de tender a infinito, la distribución se concentra en la dirección media del vector resultante. La expresión para su valor estimado es:

$$\hat{\kappa} = \begin{cases} 2\bar{R} + \bar{R}^3 + 5\bar{R}^5 / 6 & \bar{R} < 0.53 \\ -0.4 + 1.39\bar{R} + 0.43/(1 - \bar{R}) & 0.53 \leq \bar{R} \leq 0.85 \\ 1/(\bar{R}^3 - 4\bar{R}^2 + 3\bar{R}) & \bar{R} \geq 0.85 \end{cases}$$

- **Funciones de Bessel modificadas ($I_p(\kappa)$).** Atendiendo a la siguiente expresión:

$$I_p(\kappa) = \sum_{r=0}^{\infty} [(r+p)!r!]^{-1} \left(\frac{1}{2}\kappa\right)^{2r+p} \quad p = 1, 2, \dots$$

8.2 Estándares de datos espaciales

Bajo este epígrafe se aglutinan reseñas de los principales estándares existentes relativos a los datos espaciales, con especial incidencia a los referentes a la **incertidumbre de posición**, agrupados según el organismo emisor.

8.2.1 Federal Geographic Data Committee

Uno de los trabajos desarrollados por el Federal Geographic Data Committee de Estados Unidos (FGDC) es la elaboración de la *Geospatial Positioning Accuracy Standards*²⁴ que se compone de las siguientes partes:

- Parte 1. Reporting Methodology.
- Parte 2. Standard for Geodetic Control Networks. Geospatial Positioning Accuracy Standards

²⁴ <http://fgdc.er.usgs.gov/standards/standards.html>

- Parte 3. National Standard for Spatial Data Accuracy (NSSDA)

El NSSDA ofrece una metodología estadística para determinar la exactitud posicional de puntos de productos cartográficos y unificar criterios de exactitud proporcionando un lenguaje común para los usuarios de datos geográficos. Este test analiza la componente horizontal separada de la vertical. Para ello se selecciona una muestra de 20 puntos como mínimo, perfectamente definidos sobre el producto cartográfico y sobre una fuente de mayor exactitud. Se calculan los errores (EMC) de la muestra para las componentes X e Y y posteriormente para la Z, definiendo un coeficiente de exactitud posicional con un nivel de confianza generalmente del 95%. Como resultado de la prueba, junto al producto cartográfico, aparece la siguiente leyenda:

“Comprobado/compilado para X metros de exactitud horizontal/vertical al 95% de nivel de confianza”.

Este texto indica que, con un nivel de confianza del 95%, la posición de los puntos de la base de datos espacial tendrá un error con respecto al valor convencionalmente verdadero del terreno igual o menor que el valor X indicado.

Esta norma no define un umbral de exactitud, si no que en cada caso, según el producto a comprobar, es el usuario quien establece el límite de exactitud permitido.

Parte 4. Architecture, Engineering, Construction and Facilities Management.

8.2.2 American Society of Photogrammetry and Remote Sensing

El **test EMAS** (Engineering Map Accuracy Standard) fue desarrollado por la ASPRS junto con la American Society of Civil Engineers y el American Congress on Surveying and Mapping con la idea de proporcionar la exactitud de los mapas a gran escala.

Esta prueba, al igual que el test NMAS compara puntos homólogos en una cartografía origen y en otra de mayor exactitud para analizar la componente horizontal y vertical. El procedimiento a seguir para ejecutar la prueba es el siguiente:

- Se identifican al menos 20 puntos sobre la cartografía origen.
- Se calcula el error en la componente X para cada punto, como diferencia entre la coordenada X del punto en cuestión sobre la cartografía origen y la fuente de mayor exactitud.
- Se calcula el error medio para todos los puntos que componen la muestra así como la desviación.
- El valor de error y de desviación obtenido debe pasar pruebas de aceptación para errores sistemáticos y accidentales.
- Se repite el proceso para la componente Y y Z.

Se considera que la cartografía supera el test si todas las componentes pasan las pruebas citadas sobre errores accidentales y sistemáticos. La principal ventaja de este test sobre el NMAS es que el límite de error aceptable lo puede fijar el usuario

Una alternativa al test EMAS es la **fórmula de Koppe** que analiza la componente vertical de puntos comparándola con fuentes de mayor exactitud. El procedimiento, en esencia, es el siguiente:

- Se obtiene el valor de Z en la cartografía origen y en una fuente de mayor exactitud.
- Se calcula el error vertical en cada punto y el valor de la pendiente.

Calcular un error vertical y horizontal medio en función de la pendiente.

8.2.3 United Status Geological Survey

El **test NMAS** (Nacional Map Accuracy Standard) es utilizado por el USGS desde 1947 y se desarrolla comparando con fuentes de mayor exactitud. Los elementos analizados son puntos, tanto su componente horizontal como vertical; la X y la Y se analizan de forma conjunta. En esencia el procedimiento de trabajo es el siguiente:

- Sobre mapa se seleccionan un mínimo de 20 puntos fácilmente identificables.
- Se determinan las coordenadas de dichos puntos en el mapa y en otra fuente de mayor exactitud y se calcula su diferencia.
- Se determina si se alcanza el estándar de exactitud horizontal y/o vertical predefinido a partir de la diferencia entre las coordenadas del mapa origen y el de mayor exactitud. Así, para el estándar de exactitud horizontal, en el caso del USGS, se establece que como máximo, el 10% de los puntos muestreados tiene un error horizontal mayor de 1/30 de pulgada (0.846 mm) para cartografías con escalas mayores de 1/20 000 o de 1/50 de pulgada para cartografías con escalas menores de 1/20 000 (0.508 mm).

Este test se caracteriza por su simplicidad. El resultado de la aplicación sólo valora si se cumple o no dicho test, pero sin indicación de exactitud, siendo esta su principal desventaja. También recordemos que los puntos elegidos para el control deben ser fácilmente localizables en ambas cartografías, lo que propicia un muestreo un tanto arbitrario.

El USGS desarrolló unas **normas para modelos digitales de elevaciones (MDE)**²⁵ formadas por los siguientes apartados:

²⁵ <http://rockyweb.cr.usgs.gov/nmpstds/demstds.html>

- Parte 1: General.

Describe las características generales de cinco tipos de productos diferentes de MDE. Indica, además, el sistema de coordenadas horizontales, unidades utilizadas, espacio entre perfiles y tipo de datos del MDE.

- Parte 2: Especificaciones.

Define las especificaciones para la exactitud de los datos, niveles de clasificación, formatos, geometrías, áreas de elevación constante, datums horizontales y verticales y sistemas de referencia para productos MDE.

- Parte 3: Control de calidad.

Describe varios procedimientos para determinar la exactitud y el proceso de control de calidad de los DEM.

También se hace mención en este apartado a la norma *Spatial Data Transfer Standard* (SDTS). La SDTS es una norma de distribución de datos del gobierno de Estados Unidos de América. En la actualidad el USGS es el depositario y custodio de esta norma que requiere de un informe de calidad de los datos.

8.2.4 ISO/TC 211

Como ya se indicó, el comité técnico TC 211 dentro de la organización ISO, tiene como objeto la redacción de normas dentro del campo de la información geográfica digital. A continuación se indican las principales normas e informes técnicos de ISO/TC 211 referidos a la información geográfica (López Romero, Rodríguez Pascual *et al.*, 2004):

Norma	Título
ISO 6709:1983	Representación estándar de latitud, longitud y altitud para la localización de puntos geográficos.
ISO 19101:2002	Información geográfica-Modelo de Referencia.
ISO 19103:2005	Información geográfica- Lenguaje de esquema conceptual.
ISO 19105:2000	Información geográfica-Conformidad y evaluación.
ISO 19109:2005	Información geográfica- Normas para esquemas de aplicación.
ISO 19106:2004	Información geográfica-Perfiles.
ISO 19107:2003	Información geográfica-Esquemas para información espacial.

ISO 19108:2002	Información geográfica-Esquemas para información temporal.
ISO 19110:2005	Información geográfica- Metodología para la catalogación de elementos.
ISO 19111:2003	Información geográfica-Sistemas de Referencia de coordenadas.
ISO 19112:2003	Información geográfica-Referencia espacial para identificadores geográficos.
ISO 19113:2002	Información geográfica-Principios de calidad.
ISO 19114:2003	Información geográfica-Procedimientos de evaluación de calidad.
ISO 19115:2003	Información geográfica-Metadatos.
ISO 19116:2004	Información geográfica-Servicios de posicionamiento.
ISO/TR 19120:2001	Información geográfica-Estándares sobre funcionalidad.
ISO/TR 19121:2000	Información geográfica-Imágenes y datos de cobertura.
ISO 19123:2005	Información geográfica-Esquema para cobertura y funciones
ISO 19124	Información geográfica-Imágenes y componentes de datos raster.
ISO 19125-1:2004	Información geográfica-Acceso a entidades simples. Parte 1.
ISO 19125-2:2004	Información geográfica-Acceso a entidades simples. Parte 2.
ISO/TS 19127:2005	Información geográfica – Códigos geodésicos y parámetros.
ISO 19131:2007	Información geográfica-Especificaciones para productos de datos.
ISO/TS 19138:2006	Información geográfica –Medida de calidad de los datos.

Tabla 8-1. Principales Normas ISO sobre información geográfica.

8.2.5 Instituto Colombiano de Normas Técnicas y Certificación

El Instituto Colombiano de Normas Técnicas y Certificación (ICONTEC)²⁶ ha desarrollado una norma relativa a la precisión en el geoposicionamiento. Se pretende aplicar

²⁶ <http://www.icontec.org.co/homemember.asp>

una metodología estadística y de prueba para estimar la precisión en la posición de los datos espaciales. No define umbrales de precisión, sino que lo hace cada entidad de acuerdo con las especificaciones de sus productos. El estadístico utilizado es el error raíz media cuadrática (RMS), obtenido al comparar las coordenadas de unos 20 puntos con las correspondientes de una fuente de mayor exactitud para lograr un nivel de confianza de 95%.

La precisión se indicará al 95% del nivel de confianza para datos probados en su posición horizontal o vertical de la forma siguiente:

Precisión horizontal probada al 95% del nivel de confianza _____(metros)

Precisión horizontal probada al 95% del nivel de confianza _____(metros)

Se indica otro tipo de informe similar en el caso de que las pruebas se hayan realizado sin disponer de una fuente de mayor exactitud y con un método alternativo de precisión, debiendo indicar la leyenda "compilados para cumplir con_____".

8.3 Instrumental utilizado

Se indican a continuación las características del material utilizado. En primer lugar se adjuntan las características técnicas del **escáner fotogramétrico UltraScan 5000 Vexcel**, (Vexcel, 2001) propiedad de la Escuela Politécnica de Cáceres, (Universidad de Extremadura), así como las líneas generales del procedimiento de calibración Recordemos que también se utilizó, a modo de comparación, dos calibraciones realizadas en la Universidad de Cantabria por el mismo modelo de escáner.

8.3.1 Escáner fotogramétrico UltraScan 5000 Vexcel

Las características técnicas del escáner fotogramétrico citado de la UEX son:

- o Escáner fotogramétrico plano.
- o CCD Trilineal con 3 x 6000 elementos.
- o Tecnología x-y con movimiento del array de CCD.
- o Exactitud geométrica de $\pm 2\mu\text{m}$ de error medio cuadrático.
- o Resolución óptima máxima y mínima de $5\mu\text{m}$ y $29\mu\text{m}$, respectivamente.
- o Especificación radiométrica: Densidad de 0 a 3.4 logD.
- o Sin unidad de rollo de película, pero admite imagen en color o blanco y negro, transparente u opaca, positivo o negativo.

Número de referencia 3250. Modelo 1318.

8.3.2 Calibración geométrica

La **calibración geométrica** de un escáner fotogramétrico tiene por objeto disminuir los errores en el proceso de digitalización para obtener la precisión nominal. Para ello se escanea una plantilla de un material lo más estable posible, en el que aparecen marcas a distancias regulares, formando una retícula, cuya posición se conoce. Estas plantillas pueden ser de vidrio muy estables o, como ha sido el caso, una plantilla de acetato de menor precio y estabilidad modelo "Target 25". La plantilla cubre el área de escaneo para poder cuantificar las deformaciones en toda la superficie.

El proceso de calibración consiste en medir las coordenadas de los puntos sobre la imagen digitalizada de la plantilla para aplicar una transformación que relacione las coordenadas medidas y las calibradas. Las precisiones conseguidas dependen del número de parámetros de transformación y de los puntos.

También es necesaria la realización de una calibración radiométrica del escáner, de la que no se hace mención, por no ser objetivo ni necesidad del presente trabajo.

8.4 Aplicaciones informáticas

Para la realización de esta Tesis han sido necesarias varias aplicaciones informáticas que se explican a continuación.

8.4.1 Oriana de Kovach

Oriana²⁷ es un programa para Microsoft Windows que permite analizar estadísticamente datos circulares. Los datos pueden ser vectores o ejes y venir presentados, tanto en grados, como en direcciones geográficas, o bien como horas del día, días de la semana o meses del año.

Oriana (Service, 2004) calcula los estadísticos básicos para datos circulares como la mediana, la media y las medidas de dispersión correspondientes, con intervalos de confianza para la dirección media. También aplica los principales tests estadísticos a muestras o multimuestras de datos circulares y analiza si los datos introducidos se adaptan a modelos como la distribución uniforme o la de von Mises. Parte del fichero generado al calcular los estadísticos básicos de una muestra de 361 vectores introducidos en grados sexagesimales, aparece en la figura siguiente.

²⁷ <http://www.kovcomp.co.uk/oriana/index.html>

BASIC STATISTICS

**Data file - D:\Mis Documentos\Tesis\Cálculos\1-191004\estadist1.ori
191004**

Analysis begun: lunes, 23 de mayo de 2005 15:40:13

Variable	Orientación
Data Type	Angles
Number of Observations	361
Data Grouped?	No
Group Width (& Number of Groups)	
Mean Vector (μ)	141.857°
Length of Mean Vector (r)	0.019
Median	169.45°
Concentration	0.038
Circular Variance	0.981
Circular Standard Deviation	161.475°
Standard Error of Mean	113.119°
95% Confidence Interval (-/+) for μ	290.099° 3.615°

Fig. 8-1. Resultados de cálculo de Oriana.

Resultado del análisis estadístico de una muestra de vectores procedentes de la calibración de un escáner fotogramétrico.

Sobre datos y los resultados hallados se pueden realizar las operaciones básicas de todo programa bajo este formato, como copiar, pegar, ordenar, imprimir, importar, guardar, etc.

En la figura siguiente aparece el resultado del cálculo estadístico proporcionado por Oriana donde tras introducir los correspondientes acimutes en graduación sexagesimal (recordemos que el módulo se considera la unidad) se calculan los estadísticos básicos: dirección media, longitud media resultante, mediana, parámetro de concentración y varianza circular. También aparecen los resultados del test de Rayleigh, test de Rao, de Watson y Kuiper, todos explicados en el apartado 3.2.1.1 de la Tesis y además, los dos últimos contrastando tanto la posibilidad de que la muestra provenga de una población con distribución uniforme como de von Mises.

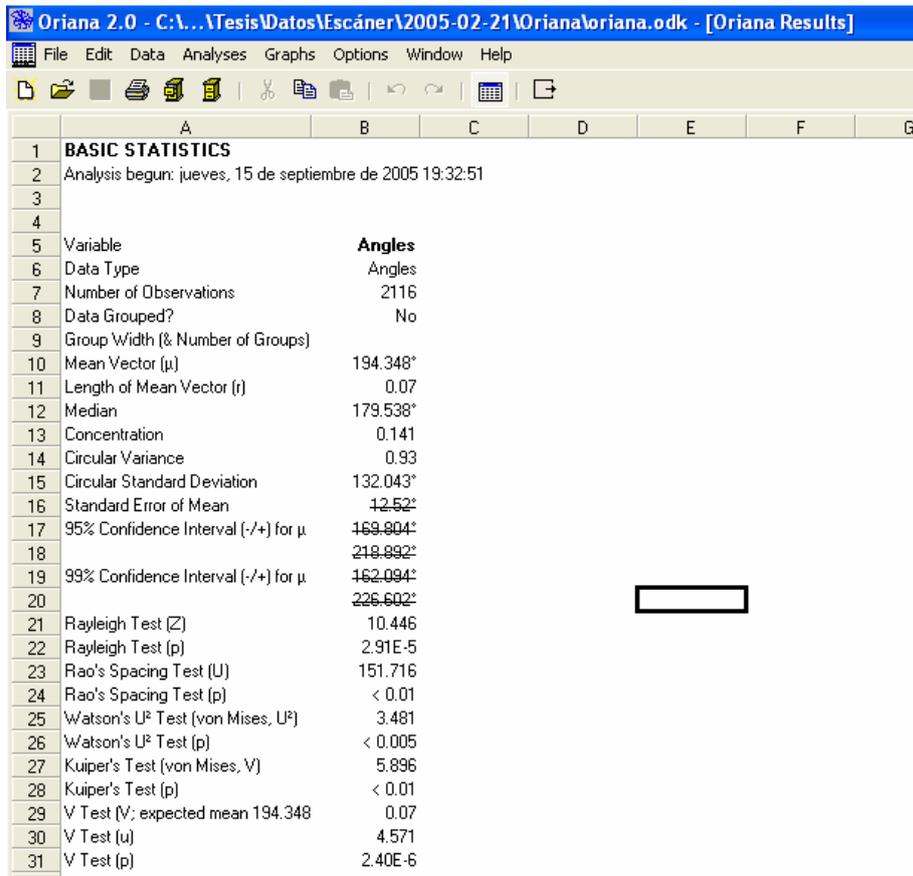


Fig. 8-2. Pantalla de resultados de Oriana.

Se presentan los resultados del análisis estadístico de datos circulares a una muestra de 2116 elementos. Los datos tachados no son fiables debido a lo dispersa de la muestra.

El programa calcula y dibuja los gráficos de bondad de ajuste de los datos a las distribuciones uniforme y von Mises y el ploteo de datos brutos, como el que aparece en la figura siguiente.

Existe la posibilidad de realizar gráficos bajo diferentes premisas. Desde el simple ploteo de datos brutos como los que aparecen en la figura a estudios de correlación en análisis multimuestra o de valoración gráfica en la bondad de ajuste para la distribución uniforme o la de von Mises.

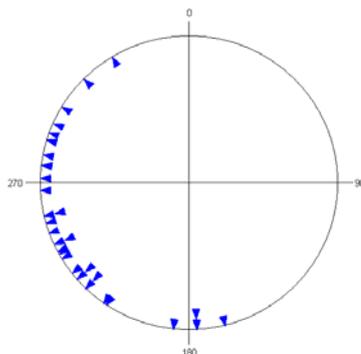


Fig. 8-3. Gráfico de Oriana.

Gráfico de datos brutos correspondientes a 30 observaciones vectoriales.

8.4.2 Microsoft Office Excel

Muchos de los cálculos fueron realizados con la hoja de datos Microsoft Office Excel, parte de Microsoft Office Profesional Edition 2003. Dado que este es un programa muy extendido en aplicaciones ofimáticas obviamos una explicación detallada.

8.4.3 R Project

R²⁸ es un software de código abierto y libre para análisis estadístico y gráficos creado por Ross Ihaka y Robert Gentleman. R tiene una naturaleza doble de programa y lenguaje de programación, se considera un dialecto del lenguaje S creado por los laboratorios AT&T Bell. R es un programa gratuito bajo los términos GNU General Public Licence.

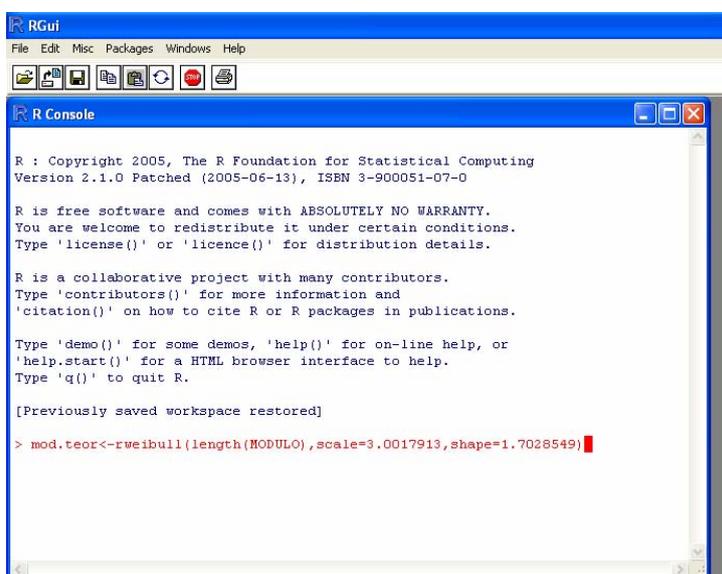


Fig. 8-4. Pantalla de R.

R es un software abierto para análisis estadístico y creación de gráficos.

R es un lenguaje orientado a objetos por lo que es flexible en cuanto a sus aplicaciones. Esto implica que los datos, funciones y resultados se almacenan en la memoria activa del ordenador en forma de objetos con nombre específico.

Los datos de la versión utilizada son: Copyright 2005, The R Foundation for Statistical Computing. Version 2.1.0 Patched (2005-06-13), ISBN 3-900051-07-0.

8.4.4 SPSS

El SPSS (*Statistical Package for Social Sciences*)²⁹ es un programa estadístico informático usado sobre todo en ciencias sociales. El programa consta de un módulo base y módulos anexos que se van actualizando constantemente. SPSS tiene un sistema de

²⁸ <http://cran.r-project.org>

²⁹ <http://www.spss.com>

ficheros siendo los principales los archivos de datos (extensión .SAV), los archivos de salida (extensión .SPO) y los de sintaxis (.SPS). Al abrir un fichero de datos con el programa, obtenemos la vista de datos, una tabla en la que las filas indican los casos y las columnas las variables. Cada celda corresponde al valor que una determinada variable adopta en un cierto caso.

Se ha contado para esta Tesis con una licencia campus educativa SPSS 12.0 para Windows (Versión 12.0.0).

Gráfico

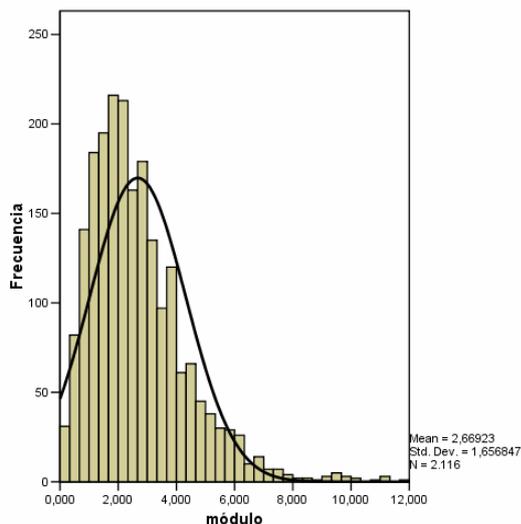


Fig. 8-5. Histograma de distribución de datos obtenido con SPSS.

El histograma de distribución de datos lleva superpuesta la curva de distribución normal.

8.4.5 ArcView de Esri

ArcView GIS 3.2³⁰, desarrollada por la empresa estadounidense ESRI (ESRI, 1999) y con licencia educativa permite tratar y representar datos georreferenciados. Este programa está diseñado de forma modular y es posible añadirle diferentes **extensiones** o programas complementarios que proporcionan funciones especializadas de GIS. Posee un lenguaje de programación propio llamado **Avenue**.

Los formatos de datos compatibles con Arcview son:

- Archivos shapefile, propio de ArcView
- Coberturas de ArcInfo
- Archivos de Intercambio de ArcInfo

Asimismo es posible trabajar con dibujos CAD, imágenes de satélites y fotos aéreas y datos en forma de tablas (DBase) o texto (ASCII).

³⁰ <http://www.esri.com>

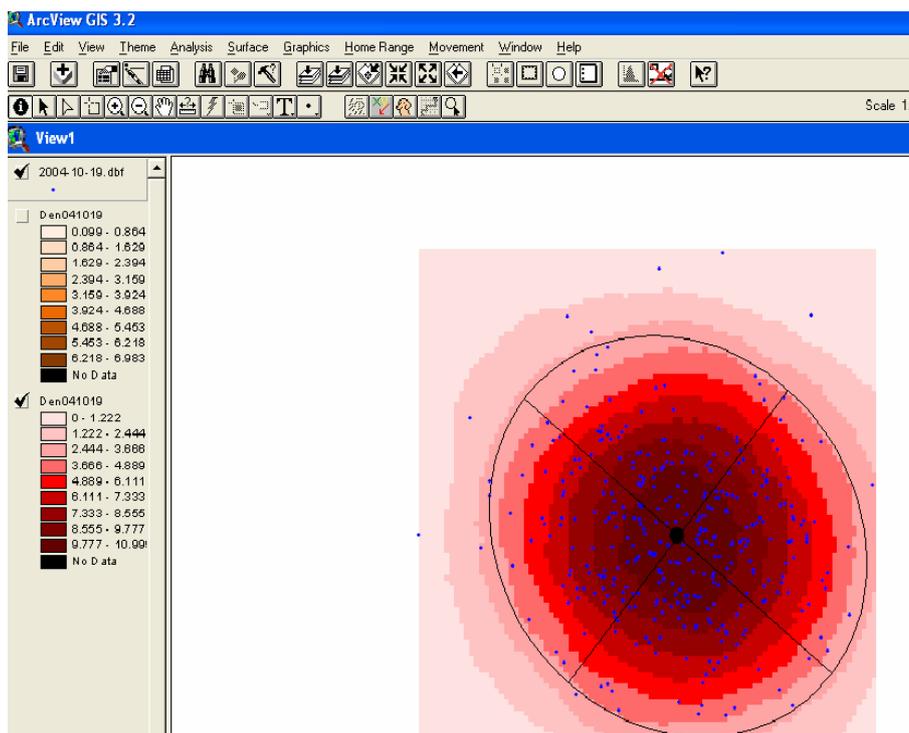


Fig. 8-6. Pantalla de resultados en ArcView.

Esta imagen se corresponde con el tratamiento que permite ArcView sobre una nube de puntos para determinar los mapas de densidad, los outliers y el ajuste de una elipse de probabilidad al 95% sobre los datos.

8.4.6 gvSIG de la Comunidad Valenciana

GvSIG³¹ es una herramienta orientada al manejo de información geográfica elaborada por la Consejería de Infraestructura y Transporte de la Comunidad Valenciana.

Este software libre (open source) permite ampliar funcionalidades desarrollando aplicaciones nuevas a partir de las librerías usadas en gvSIG. Es capaz de acceder a los formatos más usuales, ráster y vectorial, y de integrar datos locales como remotos a través de un origen WMS, WCS o WFS.

³¹ <http://www.gvsig.gva.es>

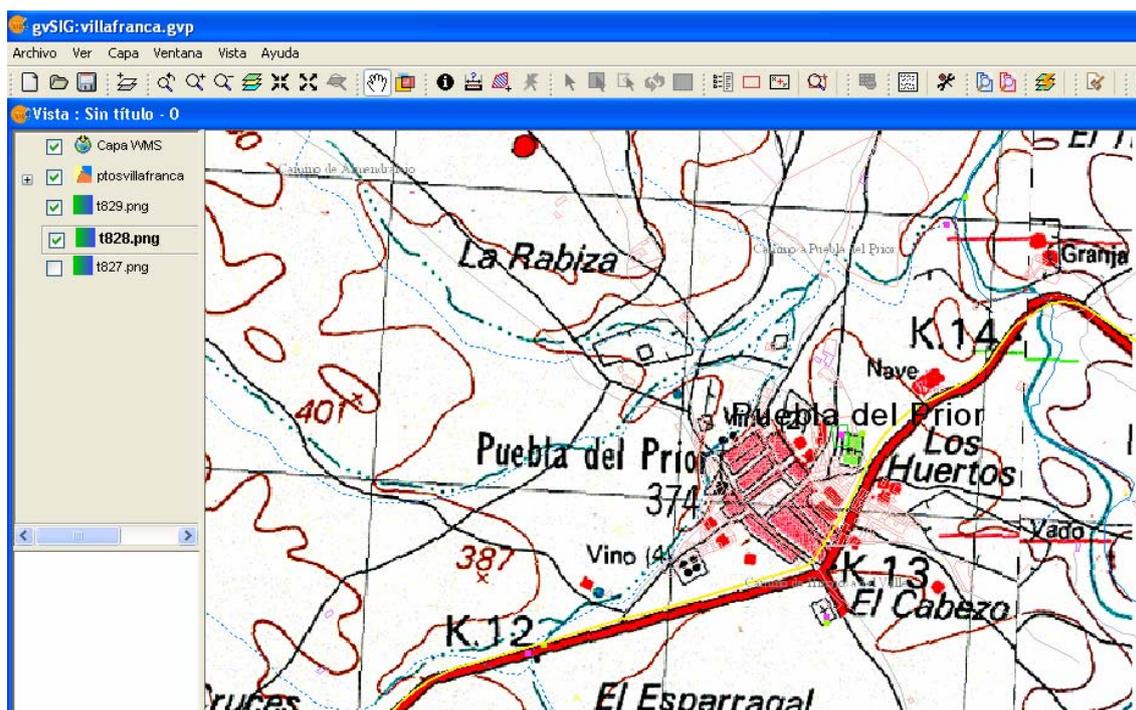


Fig. 8-7. Pantalla de resultados con gvSIG.

GvSIG permite conectarse a un servidor de mapas web (WMS) o trabajar con cartografía procedente de archivos y realizar las tareas propias de un sistema de información geográfica.

8.5 Acrónimos

Se indican a continuación los acrónimos utilizados a lo largo del texto.

AENOR	Asociación Española de Normalización y Certificación.
ANSI	American National Standards Institute
ASPRS	American Society of Photogrammetry and Remote Sensing
BCN	Base Cartográfica Numérica
CEM	Centro Español de Metrología
CEN	Comité Europeo de Normalización
CIPM	Conferencia Internacional de Pesas y Medidas
DIGEST	Digital Geographic Information Exchange Standards
DIN	Deutsches Institut für Normung
DMA	Defense Mapping Agency

EMAS	Engineering Map Accuracy Standard
EMC	Error Medio Cuadrático
FGDC	Federal Geographic Data Comité
GIS	Geographic Information System
GPS	Global Position System
ICA	Asociación Cartográfica Internacional
ICONTEC	Instituto Colombiano de Normas Técnicas y Certificación
IGN	Instituto Geográfico Nacional
ISO	Asociación Internacional de Normalización
MDE	Modelo Digital de Elevaciones
MDT	Modelo digital del terreno
MTN	Mapa Topográfico Nacional
NCHRP	National Cooperative Highway Research Program
NEM	Núcleo Español de Metadatos
NMAS	National Map Accuracy Standard
NSDI	National Spatial Data Infraestructura
NSSDA	National Standard for Spatial Data Accuracy
NTC	Norma Técnica Colombiana
OGC	Open GIS Consortium
RAE	Real Academia Española
SAIF	Spatial Archiving and Interchange Format
UEX	Universidad de Extremadura
UNE	Una Norma Española
USBB	United Status Bureau of the Budget
USGS	United States Geological Survey
VIM	Vocabulario Internacional de Metrología
VM	Von Mises

WC	Wrapped Cauchy
WMS	Web Map Server

9 Referencias bibliográficas

- AENOR (1991) 'Normalización y certificación. Conceptos básicos.' (AENOR: Madrid)
- Ariza López FJ (2002) 'Calidad en la Producción Cartográfica.' (RA-MA)
- Ariza López FJ, García Balboa JL, Amor Pulido R (2004) 'Casos prácticos de calidad en la producción cartográfica.' (Universidad de Jaén: Jaén)
- ASCI (1983) 'Map Uses, scales and accuracies for engineering and associated purposes.' American Society of Civil Engineers, Committee on Cartographic Surveying, Surveying and Mapping Division, New York.
- ASPRS (1989) Accuracy standards for large scale maps. *Photogrammetry and Remote Sensing* **56**.
- Atkinson Gordo A (2005) Control de calidad posicional en cartografía: análisis de los principales estándares y propuesta de mejora. Jaén.
- Atkinson Gordo A, Ariza López FJ (2002) Nuevo enfoque para la calidad posicional en cartografía mediante estudios basados en la geometría lineal. In 'XIV Congreso Internacional de Ingeniería Gráfica'. Santander p. 10. (INGEGRAF)
- Baltsavias EP (1994) Test and calibration procedures for image scanners. *Internacional Archives of Photogrammetry and Remote Sensing* **30**, 163-170.
- Batschelet E (1981) 'Circular Statistics in Biology.' (Academic Press: London)
- Bethel J (1994) Calibration of a Photogrammetry Image Scanner. *Technical papers of ASPRS/ACSM Annual Convention* **1**, 81-88.
- Bunge M (1985) 'Seudociencia e ideología.' (Alianza Editorial: Madrid)
- Canavos GC (1988) 'Probabilidad y estadística. Aplicaciones y métodos.' (McGraw-Hill: Mexico)
- Caprioli M, Tarantino E (2004) La Normativa y la Calidad dentro del contexto de los SIG. *Topografía y Cartografía* **XXI**, 9.
- CEM (2000a) 'Guía para la expresión de la incertidumbre (2ª edn).' (Ministerio de Fomento: Madrid)
- CEM (2000b) 'Vocabulario Internacional de Metrología (2ª edn).' (Ministerio de Fomento: Madrid)
- Cuartero A, Felicísimo AM (2003) Rectificación y ortorrectificación de imágenes de satélite: análisis comparativo y discusión. *Geofocus*, 45-47.

- Dixon KR, Chapman JA (1980) Harmonic mean measure of animal activity areas. *Ecology* **61**, 1040-1044.
- DMA (1991) 'Error Theory as applied to Mapping, Charting and Geodesy ' Defense Mapping Agency, Fairfax.
- ESRI (1999) ArcView. In.
- Felicísimo AM (1994) Parametric statistical method for error detection in digital elevation models. *ISPRS* **49**, 29-33.
- Felicísimo AM, Cuartero A (2005) Uncertainty raster-map estimation by means of stochastic simulation:a new spatial metadata approach. In 'XXII International Cartographic Conference'. A Coruña p. 5. (International Cartographic Association)
- Felicísimo AM, Cuartero A, Polo ME (2006) Analysis of homogeneity and isotropy of spatial uncertainty by means of GPS kinematic check lines and circular statistics. In '7th International Symposium on Spatial Accuracy Assessment in Natural resources and Environmental Sciences'. Lisbon. (Ed. MP Caetano, Marco) pp. 85-90
- FGDC (1998) National Standars for Spatial Data Accuracy. In. pp. 25. (Federal Geographic Data Committee)
- Fisher NI (1995) 'Statistical analysis of circular data.' (Cambridge University Press: Cambridge)
- Fisher.N.I., Lewis T, Embleton BJJ (1987) 'Statistical analysis of spherical data.' (Cambridge University Press)
- Fomento Md (2007) Actualización de la BCN200 en 16 provincias. In. (Ed. DGdIG Nacional) pp. 14. (Ministerio de Fomento)
- Ghilani CD, Wolf PR (2006) 'Adjustment Computations Spatial Data Analysis.' (Wiley: New Yersey)
- González JJ, Socorro P, Romero J, Rodríguez JI (2005) Estimación de la incertidumbre con técnicas de muestreo. In 'III Congreso Español de Metrología'. Zaragoza p. 8
- Heuvelink GBM (1998) 'Error Propagation in environmental modelling.' (University of Amsterdam)
- ICONTEC (2000) Información Geográfica.Conceptos básicos de calidad. In. pp. 39. (Instituto Colombiano de Normas y Técnicas de Certificación: Santafé de Bogotá)
- ISO9000 (2000) 'Norma ISO 9000:2000.'
- Jammalamadaka SR, SenGupta A (2001) 'Topics in circular statistics.' (World Scientific Publishing: Singapore)

- Jones MC, Pewsey A (2004) 'A family of simmetryc distributions on the circle.' The Open University, 04/07, Tenerife.
- JRC (2005) 'Guidelines for Best Practice and Quality Checking of Ortho Imagery.' Joint Research Centre. Institute for the Protection and Security of the Citizen, Ispra.
- Ley R (1986) Accuracy assesment of digital terrain models. In 'Auto Carto London'. London pp. 455-464
- Li Z (1991) Effects of check points on the reliability of DTM accuracy estimates obtained from experimental tests. *Photogrammetric engineering and remote sensing* **57**, 1333-1340.
- López Cuervo S (1993) 'Topografía.' (Mundi-Prensa: Madrid)
- López Romero E, Rodríguez Pascual A, Abad Power P, Sánchez Maganto A (2004) Normas y estándares en el entorno de la IDEE. In 'VIII Congreso Nacional de Topografía y Cartografía'. Madrid p. 6
- Mardia KV (1971) 'Statistics of direccional data.' (Academic Press Inc: London)
- Mardia KV, Jupp PE (2000) 'Directional Statistics.' (Wiley)
- Metrología CED 'Vocabulario Internacional de Metrología (2ª edn).' (Ministerio de Fomento: Madrid)
- NCHRP (2003) 'Quality and accuracy of positional data in transportation.' National Cooperative Highway Research Program, 506, Washington.
- NEM (2005) 'Nucleo Español de Metadatos ', Madrid.
- Ojeda JL (1984) 'Métodos Topográficos y Oficina Técnica.' (Madrid)
- Perruchet C, Priel M (2000) 'Estimación de la incertidumbre. Medidas y ensayos.' (AENOR: Madrid)
- Pewsey A (2002) Testing circular symmetry. *The Canadian Journal of Statistics* **30**, 591-600.
- Pinilla C (1995) 'Elementos de teledetección.' (RA-MA: Madrid)
- Service KC (2004) Oriana. In.
- SPSS (2003) SPSS. In.
- Team RDC (2005) R: A language and environment for statistical computing. In. (R Foundation for Statistical Computing: Vienna, Austria)
- USBB (1947) 'United States National Map Accuracy Standards.' US Bureau of the Budget.

USGS (1996) Digital orthophoto standars. *National Mapping Program Technical Instructions*.

USGS (1997) Standards for Digital Elevation Models.

USGS (1999) Standars for digital line graphis. In 'National Mapping Program Technical Instructions'. (U. S. Geological Survey)

USGS (2001) 'Standars for digital raster graphics.' U. S. Geological Survey.

Vexcel (2001) UltraScan 5000 User Guide. 63.

Wolf PR, Ghilani CD (2002) 'Elementary Surveying. An introductions to Geomatics.' (Prentice Hall: New Jersey)

