

Validación de datos procedentes de inventarios a escala monte

I.J. DÍAZ-MAROTO HIDALGO¹
P. VILA LAMEIRO¹
O. GARCÍA²

¹ Universidade de Santiago de Compostela. Escola Politécnica Superior de Lugo. Departamento de Enxeñería Agroforestal

² University of Northern British Columbia. Forest Growth and Yield. PhD (Georgia). Canadá

RESUMEN

Uno de los trabajos menos agradecidos en la investigación forestal consiste en la realización de mediciones y obtención de datos en monte. A esta tarea, que frecuentemente concluye con la obtención de información tanto útil como complementaria, le sigue el laborioso trabajo de digitalizar esos datos en gabinete.

Ante la imposibilidad material de realizar mediciones digitales directas, la intervención del hombre siempre está presente en la medición, anotación y digitalización de los datos obtenidos en el monte, por lo que el error humano siempre puede intervenir en cualquiera de estos tres aspectos, pudiendo desvirtuar las mediciones y el posterior análisis estadístico.

Resulta imposible eliminar completamente el error humano, pero no así su atenuación. En el presente trabajo se analizan los errores cometidos en el proceso de toma de datos en diferentes inventarios a escala monte. Los datos obtenidos se comprueban con el programa informático GNUPLOT RELEASE 4.3, mediante el cual se resaltan fácilmente cualquiera de los errores analizados anteriormente, con total independencia de su origen.

PALABRAS CLAVE: Inventario, errores, análisis, estandarización

ABSTRACT

One of the less thanked projects in forest investigation is the measurement and mountain data attainment. This task frequently supposes the obtainment of useful as well as complementary information, and continues with the cabinet digitizing laborious work of those data.

Direct digital measurements results complicated to obtain; then human intervention in measurement, annotation and digitization of mountain data is usually present. Consistently human mistake is present in anyone of these three aspects every time, being able to spoil the measurements and the subsequent statistic analysis.

Human mistake results impossible to eliminate, but it's possible its lessening. This paper analyzes the error made in several mountain scale inventories taking data process. The data obtained are runned under the computer program GNUPLOT RELEASE 4.3, which emphasizes easily everyone of the previously analyzed mistakes, wherever its origin were.

KEY WORDS: Inventory, error, análisis, standarization

MATERIAL Y MÉTODOS

El objetivo final del presente trabajo nos llevó, en primer lugar, a investigar cuales eran las fuentes de datos de las que se disponía de partida. Era necesario disponer de una fuente más o menos homogénea de información, y, a poder ser, referente a la misma especie. Esos requerimientos, y las disponibilidades de información existentes, justifican el hecho de que la especie con la que se afrontó este trabajo fue *Pinus pinaster*.

En este caso el cómputo total de datos se resumía a cuatro bases de datos (b.d.d.) distintas, procedentes de inventarios requeridos y utilizados con anterioridad en diversos trabajos, así como en Tesis Doctorales.

En primer lugar se dispone de los datos de una serie de parcelas permanentes de árboles tipo. En este caso no constan los datos de la vegetación acompañante, conociendo únicamente estimaciones de la densidad por hectárea en cada momento de inventario, a parte de los datos

individuales. Estas parcelas serán las que a partir de ahora se citen como parcelas del Grupo I.

En segundo lugar, se dispone de los datos de 23 parcelas permanentes repartidas por las cuatro provincias administrativas gallegas, y sin existir repeticiones de parcelas. En adelante se citan como parcelas del Grupo II.

A continuación, y en tercer lugar se pueden manejar también los datos correspondientes a otras 37 parcelas permanentes repartidas por toda la geografía gallega, de las que se hacen tres inventarios sucesivos, disponiendo además de tres repeticiones de cada parcela. Serán las parcelas del Grupo III.

Como última fuente de información se contó con los datos de otro trabajo localizado en montes de la provincia de Ourense. Concretamente, son los datos del Grupo IV y que hacen referencia a 22 parcelas temporales instaladas en tres puntos distintos de la citada provincia.

Algo común a todo proceso de tratamiento de datos procedentes del inventario en monte son las operaciones llevadas a cabo una vez que los datos se encuentren en nuestro poder. Todas estas operaciones se pueden resumir bajo la operación global de homogeneización y desarrollo de un formato común a todos ellos.

En este caso concreto se transformaron los formatos iniciales en que fue recibida la documentación (dBase IV: *.dbf) a un formato común más actual y manejable (Access 97: *.mdb). Acompañando esta transformación se procedió a otra elemental y simultánea, consistente en agrupar toda la información por autores y provincias.

Como fácilmente se puede pensar, esta operación no se realiza de este modo en todos los trabajos forestales, sin embargo, se ha querido analizar uno de los casos más desfavorables, incluyendo el caso de integración de diferentes fuentes de información. Así mismo se dotó a cada una de las b.d.d. de una estructura lo más parecida posible, de modo que fuese posible intercambiar información y relacionarlas a través de campos comunes.

La primera toma de contacto con los datos permitió definir la idea de lo que se deseaba que fuese la b.d.d. final. Esto, junto con el estudio más en detalle de la herramienta informática disponible (Access 97) y sus posibilidades, permitió elaborar el boceto de lo que sería el formato de la definitiva b.d.d.

Se estimó necesario diferenciar dentro de la propia b.d.d. entre la información que se refiere a las parcelas y la que se refiere a los pies individuales. En consecuencia, de partida, se plantearon como obligatorias cuatro tablas: una con datos generales de parcelas, otra con datos dasométricos también a nivel de parcela, una tercera con los resultados inventariales de los pies individuales y por último una tabla de árboles tipo.

El primer paso, fundamental en los futuros manejos de la b.d.d., fue establecer el campo común que posteriormente serviría de código a la hora de hacer consultas. Para ello se estimó conveniente elegir como tal una referencia al nombre de la parcela. Luego se estandarizaron estas referencias en un código de cuatro dígitos de significado distinto según el caso de las mediciones.

En las del Grupo I, la primera letra hace referencia a la provincia (C, Coruña; L, Lugo; O, Ourense; P, Pontevedra), el segundo dígito, se refiere a la numeración de parcela dentro de cada provincia en concreto. En este grupo de parcelas se han contemplado diferentes niveles de actuación selvícola, y es el tercer dígito, segunda letra, el que cuantifica la intensidad de la actuación selvícola que se llevó a cabo en la parcela (desde una actuación nula o mínima -A- a otra muy intensa -R- en la que se reserva menos del 50 % de f.c.c.). Por último, el segundo número identifica a qué repetición se refiere (1, 2 o 3).

En las parcelas de Grupo II, la primera letra coincide en significación con la letra del código del grupo anterior. Los dos siguientes dígitos se corresponden con la numeración que el propio autor asignó a sus parcelas en el momento de su medición. El último número sólo toma valores 0 o 1 significando la ausencia (0) o presencia (1) de fertilización en la parcela.

En las parcelas del Grupo III y IV, se siguió el mismo modelo, usando las cuatro primeras letras del nombre del lugar en el que se replanteó la parcela. Las excepciones fueron las parcelas dobles, en las que se usaron solamente las tres primeras letras, reservando el último dígito para diferenciar entre las dos repeticiones de cada parcela: parcela A y B.

Finalmente, en las parcelas del Grupo IV sólo existieron complicaciones al diferenciar Pereira y Pereiro, y se solucionó denominándolas PERA y PERO respectivamente.

Otra excepción la constituyeron las parcelas levantadas en el lugar de Pedreira, y

pertenecientes a este tercer grupo. Las parcelas no suelen registrar datos de más de tres inventarios consecutivos, independientemente del autor. Sólo este caso concreto incluye una serie de 10 años de datos. Para no desaprovecharlos y conservar la información que ofrece esa continuidad en el tiempo, pero dentro de la estructura de parcelas de tres inventarios, se optó por el formato siguiente:

Parcela	IDENTIFICACIÓN	Inventarios reales
Pedreira A1	PEA1	1, 2, 3
Pedreira A2	PEA2	3, 4, 5 etc

Junto con ese campo (nombrado como IDENTIFICACIÓN), se repite en todas las tablas de la b.d.d. otros dos en los que se cita el autor (AUTOR) -por lo tanto, el origen de los datos- y el número de inventario (Nº DE INV). El fin de estos dos campos es más facilitar la organización de la información que el ofrecer información por sí mismos.

Una vez estandarizados los datos y las claves de las parcelas se procedió a introducir la información en cada una de las cuatro tablas antes definidas, con los tres campos comunes -ID, Nº DE INV y AUTOR-, más los considerados convenientes en cada caso y que fueron los siguientes:

• **TABLA 1**

		Nombre: Parcelas: Datos	
Campos:	Clave		ID
	Año de plantación de la masa		AÑO PLANT
	Superficie		SUPF
	Lugar		LUGAR
	Número de inventarios disponibles		INVENT.
	Fecha de realización de los inventarios		FECHA INV
	Autor		AUTOR
	Observaciones		OBSERV.

• **TABLA 2**

		Nombre: Parcelas: Resultados	
Campos:	Clave		ID
	Número de inventario		Nº DE INV
	Fecha de realización del inventario		FECHA
	Diámetro normal medio antes de la actuación		DI a
	Idem después		DI d
	Altura total media antes de la actuación		Ha a
	Idem después		Ha d
	Altura dominante		Ho
	Volumen por Ha antes de la actuación		V a
	Idem después		V d
	Crecimiento por Ha antes de la actuación		CR a
	Idem después		CR d
	Coefficiente mórfico antes de la actuación		CM a
	Idem después		CM d
	Autor		AUTOR

• **TABLA 3**

		Nombre: Pies individuales	
Campos:	Clave		ID
	Número del árbol		PIE
	Número del inventario		Nº DE INV
	Diámetro normal		D
	Altura total		H
	Autor		AUTOR

• **TABLA 4**

		Nombre: Pies: Arboles tipo	
Campos:	Clave		ID
	Número del árbol		PIE
	Número del inventario		Nº DE INV
	Diámetro normal		D
	Altura total		H
	Autor		AUTOR

Algunos de estos datos tuvieron que ser tecleados debido a que no se encontraban en las b.d.d. manejadas que se citaron en un principio. En concreto la casi totalidad de la tabla 2, en la que, además, fue necesario calcular datos por hectárea al disponer de ellos solamente por parcela.

Con estas operaciones se concluye el proceso de introducción de información en una b.d.d. El siguiente paso es su depuración, para lo cual se eliminan registros con valor cero en campos de

diámetros o alturas, se eliminan los ceros de algunos registros y que alterarían su manipulación (altura dominante, altura media, diámetro normal, densidad,... antes y/o después de una actuación). Además se deben eliminar los registros en los que se desconoce la numeración del pie en cuestión, carece de datos de diámetro, no se conoce la altura si es un árbol tipo, etc.

Este proceso de análisis se debe llevar a cabo con el fin de detectar errores de medición, transcripción o interpretación. Para ello se grafican tanto los datos de parcelas como los de pies individuales y árboles tipo. Se aplica entonces el programa informático GnuPlot for Windows Release 3.4.

El citado programa necesita disponer de los datos en un formato particular. Deben transformarse a un formato TXT, ASCII, DAT, ... Dentro de cada uno de esos archivos, cada unidad de datos (parcela, pie individual, ...) debe ir separada de la siguiente y la precedente por una línea en blanco. De igual forma, los datos de esa unidad se distribuirán en líneas (cada línea una unidad temporal de datos, normalmente años) y en columnas (en una las unidades temporales y en otra los valores en concreto).

Con ese fin se exportan los datos de Access a Word en donde se les da el formato descrito anteriormente y se exportan los archivos como archivos de texto TXT. Durante este proceso es posible desechar aquellos datos con referencias en un solo año y que no muestran una evolución temporal.

Luego, desde el mencionado GnuPlot se importan los datos, se grafican y se imprimen, obteniendo así una representación gráfica de la evolución de las masas.

Con los datos impresos se revisan señalando aquellos que pudiesen resultar sospechosos. Estas sospechas se centran fundamentalmente en inversión de tendencias, crecimiento excesivo de una variable,... Debido a la subjetividad de las observaciones en muchos de los casos planteados, se debe optar por señalar todas las dudas por mínimas que estas fuesen.

Detectados los datos que se consideran como erróneos se procede a su corrección usando para ello los estadillos originales.

RESULTADOS

En el momento en que se exportaron los datos de Access a Word para darles el formato descrito anteriormente y luego volverlos a exportar, ahora como archivos de texto TXT, fue la primera ocasión en la que se detectaron algunas anomalías en los mismos. Así, durante este proceso fue posible desechar aquellos datos con referencias en un solo año y que no mostraban una evolución temporal.

A esta operación le siguió el proceso de estandarización ya descrito, y que se llevó a cabo, a excepción de los datos del Grupo III y Grupo IV, en los cuales no fue posible esa estandarización de campos de la b.d.d.. En el primer caso porque los datos de los que se disponían eran de árboles tipo, y no de inventario pie a pie dentro de cada parcela.

En el segundo caso el problema surgió porque la estructura era totalmente distinta de la deseada y que era bastante próxima a la que presentaban las b.d.d. del Grupo II y Grupo I. En estos casos cada pie individual era un registro, cada parcela un grupo de registros y cada campo una variable distinta medida o estimada. En el caso de los datos del Grupo IV, cada parcela era un registro y cada variable en cada pie un campo distinto.

En los datos del Grupo I y Grupo II se optó por mantener los campos: Nombre clave con el que se hace referencia a la parcela en cuestión, calidad de estación (evaluada en Grupo II), existencia o no de fertilización (Idem), número de pie y resultados de las mediciones diametrales y de altura en los sucesivos inventarios (hasta cuatro en parcelas del Grupo II y tres en las del Grupo I). Cada registro representa un pie de cada parcela y/o repetición de parcela. Finalmente se consideró oportuno elaborar una tabla propia en la que se pudiesen recoger todas las referencias no selvícolas que interesase conocer de cada parcela. A saber: Nombre del monte, ayuntamiento al que pertenece y superficie.

Al introducir los datos de los árboles tipo del Grupo III, surgieron multitud de inconvenientes, todos ellos relacionados con irregularidades de los inventarios o pérdida de datos de los mismos. En concreto, y particularizando ejemplos de este grupo, algunos problemas surgidos siguieron los siguientes patrones:

Parcela	ID	Problema
Geve	GEVE	Hay tres inventarios y falta el tercero. En el segundo no existen datos de los pies número 41 y 125, y aparecen el 32 y el 70, que se podría corresponder con el 41.
Morrillo A	MORA	Falta el pie 15 en el tercer inventario y aparecen otros: 23, 143, 147, 188, 211.
Morrillo B	MORB	Falta el pie 45 en el segundo y tercer inventarios. En el tercero aparecen además los pies 9, 96 y 102.
Pedreira A1	PEA1	Faltan los datos del primer inventario y en los del segundo (1936) no se numeraron los pies.
Pedreira A2	PEA2	Faltan los datos de los pies 39, 95 y 152 del tercer inventario (5° general de la parcela).
Pedreira A4	PEA4	Desde el segundo inventario (9° general de la parcela) desapareció el pie 54.
Ginde	GIND	No se encontraron los datos de los árboles tipo.
Buxel A	BUXA	Desde el tercer inventario aparecen más árboles tipo: 16, 60, 109, 123, 127, 142 y 143.
Buxel B	BUXB	Los datos de árboles tipo corresponden a mediciones de 1943 y 1944. Sin embargo, los datos del inventario diamétrico pie a pie de la Tabla 3 -Pies individuales-, son de 1940 y 1941.
Castro B	CASB	Desde el segundo inventario, desaparece el pie 146.

Superada esa fase inicial de introducción de los datos, los problemas detectados fueron múltiples, y sus causas también. Los casos más sencillos fueron aquellos en los que los datos estaban mal tecleados y que se corrigieron (por ejemplo el pie 101 de la parcela L1A1 del Grupo I que aparecía con 253 mm cuando le corresponden 226 mm). En ocasiones la diferencia es bastante pequeña, sin embargo sí destaca la variación en la tendencia de los demás pies de la parcela, algo que siempre suele ocurrir.

En otros casos se confundía la medición con el número de árbol o encontramos datos con errores manifiestos de medición (se eliminaron). Este fue, por ejemplo, el caso de pies de la parcela L1C1 del Grupo I o de la parcela O131 del Grupo II.

Otras situaciones más comprometidas fueron aquellas en que el dato era dudosamente erróneo, los datos de distintas parcelas estaban intercambiados, o, simplemente, los registros tecleados no se correspondían con ninguna realidad (procediendo a su retecleo). Ese fue el caso de algunas parcelas del Grupo I.

En otros casos se detecta una importante disminución en la altura; debido a que desde el anterior inventario, su guía terminal se ha roto. Igualmente, hubo manifiestas disminuciones de densidad en parcelas del Grupo II. La causa posibles cortas no registradas. Incluso en la tabla 4 de la b.d.d. -Pies: árboles tipo- fue necesario introducir todos los datos de las parcelas del Grupo III, así como completar los existentes informatizados de otros autores.

Por otro lado, en muchas de las ocasiones en las que un valor resultaba sospechoso, al comprobarlo en los estadillos se veía que era correcto. En ese caso, se admitía su validez, de no ser, como hemos dicho, su error manifiesto.

En otro orden de cosas, se asumieron como válidos algunos datos que denotan claramente errores de medición. Este es el caso de parcelas con un bajo crecimiento y en las que la segunda medición resulta ligeramente menor a la primera. Para corregir esto tomaron la segunda medición igual a la primera, lo cual se mantuvo como aceptable. El origen de tal desviación es un error de apreciación humano. Se trata en todos los casos de parcelas del Grupo II. En dicho grupo se realizó un inventario piloto en el mes de Septiembre para calcular la intensidad del inventario. Luego ese inventario piloto se tomó como inventario número 1, y el 2 se realizó a partir del mes de Febrero del año siguiente. En consecuencia el crecimiento detectado era mínimo, y, en algunos casos, una variación en la apreciación propició el que ciertas mediciones indicasen una tendencia decreciente.

Al final de las comprobaciones y en la provincia de A Coruña, en los datos del Grupo I se detectaron parcelas inexistentes o con datos totalmente distintos a los de los estadillos. Para su resolución fue necesario reteclear toda la información de esas parcelas. Resultó totalmente imposible poder identificar el origen de tan manifiesto error. El problema sólo afectó a catorce parcelas, pero

fue suficiente para hacer surgir la duda de si las demás parcelas no comprobadas (porque sus datos no mostraban tendencias sospechosas) se encontrarían en esta misma situación. Para saberlo se procedió a una rápida comprobación de coincidencia entre información en papel e informática. Al final se desechó ese posible vicio en el resto de parcelas.

Llegados a este punto, y dado ese primer paso de estandarizado y corregido de los datos de todas las parcelas, comenzó el proceso de elaboración para la obtención de resultados. Después de cada operación o resultado intermedio alcanzado, se repetían los trabajos de revisión. De este modo, conocidos los datos de altura dominante, se comprobaron errores y resultados incoherentes. Para esto se recurrió de nuevo a la graficación de los mismos. Las mayores incongruencias se detectaron en las parcelas de Grupo III. En ellas sólo se disponía de unos pocos pies por parcela tecleados en la b.d.d. El resultado era que, por escasa información, los valores obtenidos de altura dominante iban desde los negativos a los exageradamente altos. Se resolvió buscando información sin teclear de Grupo III y procesándola.

En los demás casos fueron problemas puntuales resueltos con una nueva consulta de los datos y su corrección en caso de que fuese necesario. Finalmente, se repitieron los cálculos y se obtuvieron los definitivos valores de altura dominante por parcela. Son al final 551 parcelas con los correspondientes datos de altura dominante y edad.

La siguiente operación realizada se llevó a cabo mediante el análisis de troncos para definir algún patrón de crecimiento. Esto supuso la consecución de los valores de edad y altura para un árbol tipo ideal para cada parcela. Como siempre, el paso final fue el graficado de los resultados para buscar posibles errores; Sin embargo, y por primera vez durante los análisis, estos errores fueron nulos debido a su corrección a estas alturas de refinado.

En consecuencia el número de registros de la b.d.d. evolucionó del siguiente modo:

Nº DE REGISTROS		
TABLA	SIN REVISIÓN	REVISADA
PARCELAS: DATOS	201	201
PARCELAS: RESULTADOS	407	407
PIES INDIVIDUALES	71849	52207
PIES: ARBOLES TIPO	19887	19643
TOTAL	92344	72458

CONCLUSIONES

1. Los procesos de recogida de información en campo son fundamentales en cualquier trabajo científico, y muy especialmente en trabajos forestales.
2. Resulta fundamental asegurar al máximo la veracidad de la información con la que se trabaja, para así poder afirmar con rotundidad los resultados finales obtenidos.
3. Los errores de medición son difíciles de eliminar en la mayor parte de los casos. Frecuentemente lo único que podemos hacer es detectarlos y eliminar el dato, para que, cuando menos, no desvirtúe el análisis.
4. Los errores de manejo y transcripción de los datos pueden ser eliminados fácilmente, mediante una mínima dedicación al proceso de análisis y estandarización de los datos manejados.
5. El programa informático GnuPlot for Windows Release 3.4 se muestra como una herramienta de manejo muy sencilla y práctica a la hora de graficar y detectar errores tanto de medición, como de transcripción como de manejo de los datos.
6. La práctica en los procesos de revisión de datos permite descubrir ciertos indicadores que facilitan el identificar una medición o valor como correcto o dudoso. Tal es el caso de analizar tendencias de parcela y no valores cuantitativos, conocer condiciones en que se desarrolló el trabajo de inventariación y parámetros que condicionaron la evolución de la masa durante su desarrollo.
7. La naturaleza puede deparar ciertas “sorpresas”, y determinados valores que pueden parecerse erróneos o fuera de toda lógica, son perfectamente válidos y consecuencia de la actividad del medio.
8. Finalmente, las conclusiones todo proceso de análisis de datos, si no ha ido precedido de otro

de estandarización y corrección de los datos del inventario, debe ser tachado, cuando menos, de dudoso, pues está partiendo de unos datos que no pueden ser certificados como mínimamente fiables o carentes de errores de bulto.

BIBLIOGRAFÍA

M.O.P.T. (1992): Guía para la elaboración de estudios del medio físico; Ministerio de Obras Públicas y Transportes. Madrid.

PÉREZ MOREIRA, R. (1993): Ecoloxía forestal. Xunta de Galicia. Santiago de Compostela.

PITA CARPENTER, P.A. (1973): El inventario en la ordenación de montes. Ministerio de Agricultura. Madrid.