



2025 | **16-20**
GIJÓN | JUNIO

9º CONGRESO **FORESTAL** ESPAÑOL

9CFE-1360

Organiza





forestables: un paquete de R para la armonización de datos de Inventarios Forestales Nacionales

TOVAR-RODRÍGUEZ, A. (1), GRANDA-GARCÍA, V. (1) y DE CÁCERES-AINSA, M. (1)
(1) CREAM, Cerdanyola del Vallés 08193, España

Resumen

forestables es un paquete de R diseñado para armonizar datos abiertos de los Inventarios Forestales Nacionales (IFN) de EE. UU. (FIA), Francia (FFI) y España (IFN). Este paquete estandariza información sobre el estrato arbóreo, el regenerado, el sotobosque y las características topográficas de la parcela. Los datos se organizan en tablas, donde cada fila representa una combinación única de parcela y año. Las primeras columnas contienen los metadatos, mientras que las últimas incluyen tablas anidadas con datos detallados sobre árboles, sotobosque y regeneración.

forestables incluye funciones para descargar y obtener datos de fuentes oficiales, filtrar regiones específicas, limpiar los resultados de valores faltantes y convertir los objetos a formato espacial (*sf*). Aunque pueden servir para muchos otros propósitos, los *outputs* generados por el paquete están especialmente diseñados para ser utilizados en simulaciones de dinámica forestal, ya sea mediante modelos empíricos o modelos basados en procesos. El paquete está disponible en GitHub (<https://github.com/emf-creaf/forestables>), sigue los principios del *tidyverse* y ofrece documentación completa en formato de ayuda y viñetas con ejemplos prácticos para su implementación.

Palabras clave

Armonización de datos, Inventarios Forestales Nacionales, Datos abiertos, Modelización ecológica

1. Introducción

Los Inventarios Forestales Nacionales (IFNs) son herramientas esenciales para el monitoreo, la gestión y la planificación forestal a nivel nacional e internacional (VIDAL et al., 2016). Aunque muchos países ofrecen acceso público a estos datos, la falta de armonización entre formatos y metodologías, incluso dentro de regiones como España, representa un desafío significativo para el personal investigador y técnico que trabaja con datos provenientes de distintas provincias o países. Hasta ahora, la armonización de estos datos se ha llevado a cabo de manera independiente y según la necesidad de cada grupo, lo que dificulta la integración y comparación de datos de distintas regiones. Estas limitaciones reducen la efectividad de los análisis y restringen el alcance de las decisiones basadas en ellos (GSCHWANTNER et al., 2024).

Para abordar este problema, presentamos *forestables* (TOVAR & GRANDA, 2024), un paquete de R diseñado para armonizar datos forestales de tres inventarios nacionales: el *Forest Inventory and Analysis* (FIA) de los Estados Unidos, el *Inventario Forestal Nacional de Francia* (en adelante FFI) y el *Inventario Forestal Nacional* (IFN) de España. *forestables* estandariza información sobre árboles, regenerado, arbustos y características de la parcela, organizándola en formato de tabla. Estos *outputs* permiten estudiar y comparar ecosistemas forestales y sus trayectorias a lo largo del tiempo. Además, están diseñados específicamente para su uso en simulaciones de dinámica forestal, ya sea mediante modelos empíricos o modelos basados en procesos, como *medfate* o *medfateland* (DE CÁCERES et al,

2023a; DE CÁCERES et al, 2024b).

2. **Objetivos** El objetivo general ha sido desarrollar un paquete de R que realizara las siguientes funciones:
 - **Automatizar la descarga y armonización de datos de inventarios forestales internacionales:** Desarrollar un paquete en R que permita la descarga automática y la conversión de datos de los inventarios forestales de EE. UU., Francia y España a un formato estandarizado, reduciendo la necesidad de preprocesamiento manual.
 - **Estandarizar variables clave:** Unificar las variables comunes entre los inventarios forestales y asegurar un formato consistente que facilite el análisis comparativo entre países o versiones del IFN.
 - **Facilitar la generación de datos de entrada para modelos de dinámica forestal:** Proporcionar una herramienta para generar los datos de entrada requeridos por modelos de dinámica forestal.
3. **Metodología**

El funcionamiento general del paquete se ilustra en *Proceso esquemático de funcionamiento del paquete forestables (Figura 1.)*

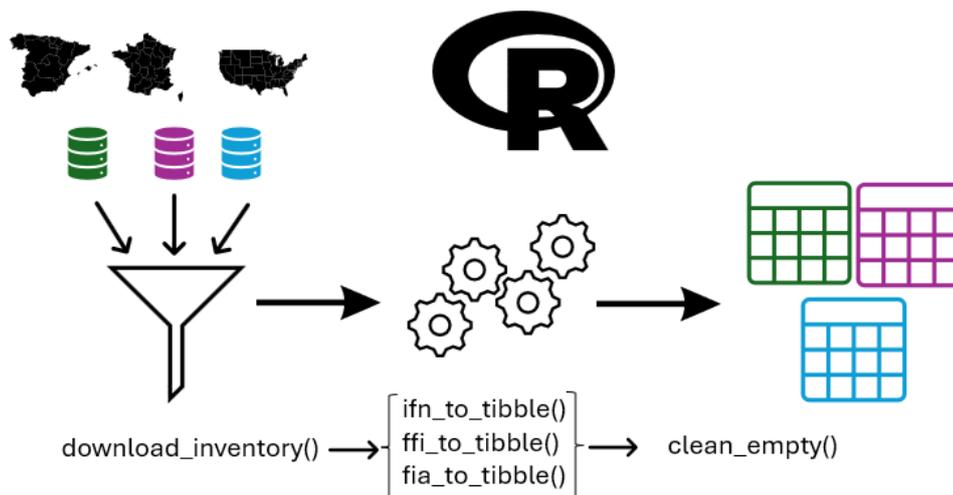


Figura 1. Proceso esquemático de funcionamiento del paquete forestables.

Datos

El paquete emplea los **datos brutos de los inventarios de los portales oficiales**. Estos datos tienen un formato y una estructura que es **específica de cada inventario**. Cada inventario consta de varias **tablas** donde se recogen de manera separada los datos de los componentes del muestreo tales como características de la parcela, estrato arbóreo, regenerado y matorral o sotobosque.

Procedimiento de armonización

El **proceso de armonización** se realiza de manera interna a través de funciones específicas que no están visibles al usuario. Estas funciones acceden a los datos de inventarios forestales de España, Francia y Estados Unidos y seleccionan un **conjunto de variables comunes** entre los inventarios. Debido a que no es posible una armonización total, se incorporan variables específicas de cada país, las cuales



se detallan en la documentación del paquete.

Algunas de esta variable se **transforman** para asegurar que hay una coherencia en las unidades empleadas. Paralelamente, se realizan algunas **aproximaciones** a variables que no están directamente muestreadas pero que pueden derivarse a partir de otras variables (e.g., caso de la altura matorral en el FIA estimada a partir de punto medio del estrato de muestreo), o bien se asignan **valores por defecto** (e.g., caso del diámetro y altura del regenerado del IFN). Estos procedimientos implican una **toma de decisiones** que es útil para los objetivos de simulación. Conociendo las reglas de transformación, el usuario puede cambiar estos valores a otros de su interés. Todas las transformaciones están explicadas en una **viñeta** del paquete llamada inventory data ingestion process.

Ejemplo de transformación: el caso del regenerado del IFN

El regenerado se recoge de manera distinta entre IFN2, IFN3 e IFN4. Para poder **armonizar la información** entre versiones de IFN se han transformado las variables originales a otras más intuitivas: densidad, altura, diámetro y especie. Se han unificado las distintas categorías de desarrollo: a cada una se le asignan una altura y un diámetro por defecto (esto permite distinguirlas). También se ha calculado la densidad en individuos por hectárea, multiplicado el factor de expansión de radio 5 m por el número de individuos. Cuando el número de individuos se recoge como intervalo, se selecciona el punto medio o el mínimo +5 (para la última clase “>15”).

Después de obtener y transformar las variables se **ordena la información** en un formato de salida específico, un *tibble* (estructuras de datos propias de R). **Cada fila representa una combinación de parcela y año y cada columna es una variable armonizada**. Las primeras columnas corresponden con los metadatos de las parcelas mientras que las columnas finales se corresponden con los datos anidados (celdas que contienen tablas) de árboles, sotobosque y regenerado respectivamente. (Figura 2).

```
# A tibble: 5,207 × 9
  id_unique_code year plot coordx coordy coord_sys tree understory regen
  <chr> <chr> <dbl> <dbl> <chr> <list> <list> <list>
1 24_0191_NN_A1_A1 92 0191 675000 4748000 ED50 <tibble [17 × 8]> <tibble [1 × 2]> <tibble [4 × 6]>
2 24_0192_NN_A1_xx 92 0192 675000 4747000 ED50 <tibble [32 × 8]> <tibble [1 × 2]> <tibble [3 × 6]>
3 24_0193_NN_xx_xx 92 0193 700000 4753000 ED50 <tibble [0 × 0]> <tibble [1 × 2]> <tibble [0 × 0]>
4 24_0194_NN_xx_xx 92 0194 687000 4752000 ED50 <tibble [8 × 8]> <tibble [1 × 2]> <tibble [4 × 6]>
5 24_0195_NN_xx_xx 92 0195 700000 4752000 ED50 <tibble [0 × 0]> <tibble [1 × 2]> <tibble [1 × 6]>
6 24_0196_NN_xx_xx 92 0196 683000 4750000 ED50 <tibble [0 × 0]> <tibble [1 × 2]> <tibble [0 × 0]>
7 24_0197_NN_xx_xx 92 0197 682000 4749000 ED50 <tibble [0 × 0]> <tibble [1 × 2]> <tibble [2 × 6]>
8 24_0198_NN_A1_A1 92 0198 684000 4749000 ED50 <tibble [27 × 8]> <tibble [1 × 2]> <tibble [3 × 6]>
9 24_0199_NN_A3C_xx 92 0199 692000 4749000 ED50 <tibble [0 × 0]> <tibble [1 × 2]> <tibble [0 × 0]>
10 24_0200_NN_xx_xx 92 0200 693000 4749000 ED50 <tibble [3 × 8]> <tibble [1 × 2]> <tibble [1 × 6]>
# i 5,197 more rows
# i Use `print(n = ...)` to see more rows
```

Figura 2. Ejemplo de output de forestables.

Funciones de la interfaz de usuario

A continuación, se describen las funciones **accesibles al usuario**. Puede encontrarse información más detallada en la documentación del paquete: (<https://emf-creaf.github.io/forestables/reference/index.html>).

- **Descarga** de los datos de los inventarios FIA, FFI o IFN en bruto: `download_inventory()`.
- **Visualización preliminar** de los datos de las parcelas: `show_plots_from()`.
- **Creación de listas de filtrado** de datos de regiones específicas: `create_filter_list()`.

- **Procesamiento** de los datos de cada inventario y generación de *tibbles*: `fia_to_tibble()`, `ffi_to_tibble()`, `ifn_to_tibble()`.
- **Limpieza** de las filas con resultados vacíos en las columnas anidadas especificadas: `clean_empty()`.
- **Transformación** de los *tibbles* en objetos *sf* - *simple features* (PEBESMA, 2018; PEBESMA, 2023): `inventory_as_sf()`.

Además, en el paquete se incluyen ejemplos de salida (`fia_output_example`; `ffi_output_example`; `ifn_output_example`) que sirven como referencia para verificar la estructura y contenido de los datos tras el procesamiento.

En el apartado de resultados se desarrolla un ejemplo de uso con varias de las funciones descritas.

Acceso a las columnas anidadas

Dado que las columnas de árboles, arbustos y regenerado están anidadas (es decir, contienen tablas dentro de cada celda), la forma más sencilla para acceder a estos datos es mediante la función `unnest()` del paquete *tidyr* (WICKHAM et al, 2023; WICKHAM et al, 2024). Esta función expande las columnas anidadas y las integra en la tabla en forma de filas.

Uso de *forestables* en el contexto de *medfateverse*

El paquete *forestables* ha sido diseñado específicamente **para facilitar la generación de datos de entrada** de los modelos *medfate* y *medfateland*, que forman parte del entorno *medfateverse* (DE CÁCERES et al., 2023a; DE CÁCERES et al., 2024a). En particular, ya se ha desarrollado una función que procesa el objeto generado por *forestables* para convertirlo en un nuevo objeto en formato “*forest*” (un *input* requerido por el modelo *medfate* con *información forestal*). Este nuevo objeto se crea a partir de las columnas *tree*, *understory* y *regen*, manteniendo la estructura general del objeto *sf*. La Figura 3 ilustra las relaciones entre paquetes del *medfateverse*.

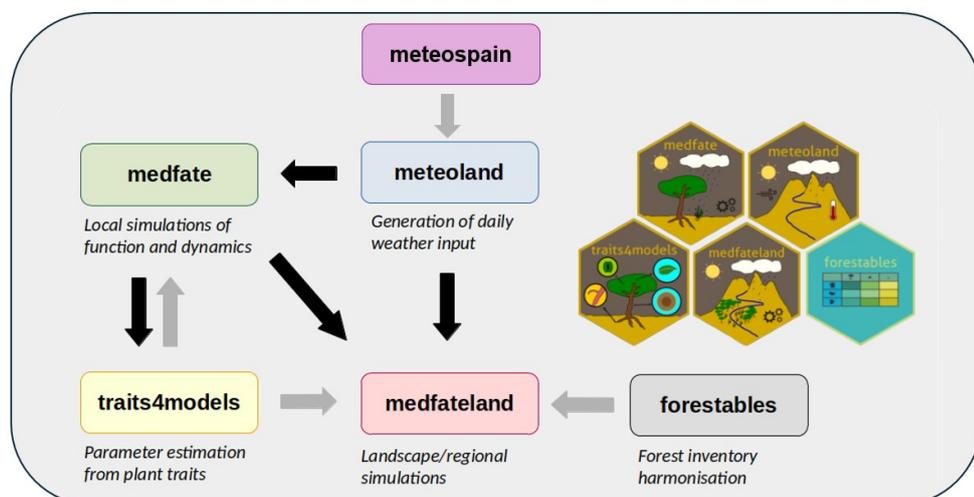




Figura 3. Marco de trabajo interconectado de medfateverse.

4. Resultados

Para ejemplificar el uso del paquete se desarrolla un ejemplo con datos del IFN. Se ejecutan en el entorno rstudio, software R versión 4.4.2 . Todas las funciones siguen la filosofía de *tidyverse* (WICKHAM et al,2024) , es decir, se pueden concatenar con *pipes* (`|>` o `%>%`) , lo que facilita el manejo y la integración en flujos de trabajo complejos con varios pasos.

En la *Figura 4* se ejemplifica el **flujo de trabajo** para instalar el paquete, descargar los datos del IFN, definir las provincias de interés para el análisis, obtener los datos procesados en formato *tibble*, limpiar datos faltantes, convertir a *sf* y desanidar los datos de árboles. Todo ello usando pipes en solo 10 líneas de código.

```
#1 Instalar paquete desde GitHub
install.packages("devtools")
library(devtools)
devtools::install_github("emf-creaf/forestables")

#2 Descargar datos en un directorio temporal
ifn_folder<- tempdir()
forestables::download_inventory("IFN", ifn_folder)

#3 Definir las provincias de interés
north_spain_provinces <- c("33", "15", "27", "39", "48", "20")

#4 Obtener los datos
ifn_data <- forestables::ifn_to_tibble(
  provinces = north_spain_provinces,
  versions = "ifn3",
  folder = ifn_folder )

#5 Limpiar y convertir en formato sf y desanidar
ifn_data |>
forestables::clean_empty(c("tree")) |>
forestables::inventory_as_sf() |>
tidyr::unnest("tree")
```

Figura 4. Ejemplo de flujo de trabajo para obtener datos del IFN.

En la siguiente figura se recoge un ejemplo de salida de datos desanidados para la columna de árboles, *Ejemplo de datos desanidados y filtrados de árboles (Figura 5)*. Para simplificar el entendimiento se han omitido algunas columnas de metadatos que también estarían presentes.



```
# A tibble: 79,084 × 7
  id_unique_code year plot tree sp_name dia height
  <chr> <chr> <chr> <chr> <chr> <dbl> <dbl>
1 24_0191_NN_A1_A1 92 0191 1 Quercus petraea 43.0 21
2 24_0191_NN_A1_A1 92 0191 2 Quercus petraea 15.4 18.5
3 24_0191_NN_A1_A1 92 0191 3 Quercus petraea 16.5 17.5
4 24_0191_NN_A1_A1 92 0191 4 Quercus petraea 34.5 17.5
5 24_0191_NN_A1_A1 92 0191 5 Quercus petraea 45.2 20
6 24_0191_NN_A1_A1 92 0191 6 Quercus petraea 28.3 9.5
7 24_0191_NN_A1_A1 92 0191 7 Quercus petraea 45.9 22.5
8 24_0191_NN_A1_A1 92 0191 8 Quercus petraea 47.2 19.5
9 24_0191_NN_A1_A1 92 0191 9 Quercus petraea 49.6 21
10 24_0191_NN_A1_A1 92 0191 10 Quercus petraea 84.8 20.5
# i 79,074 more rows
# i Use `print(n = ...)` to see more rows
```

Figura 5. Ejemplo de datos desanidados y filtrados de árboles.

Finalmente, se visualiza un gráfico que puede obtenerse a partir de los datos provenientes de *forestables* (Fig. 6) En ella se muestran la distribución de las especies más ampliamente presentes en el norte de la península y los promedios por parcela de los diámetros (dbh) encontrados.

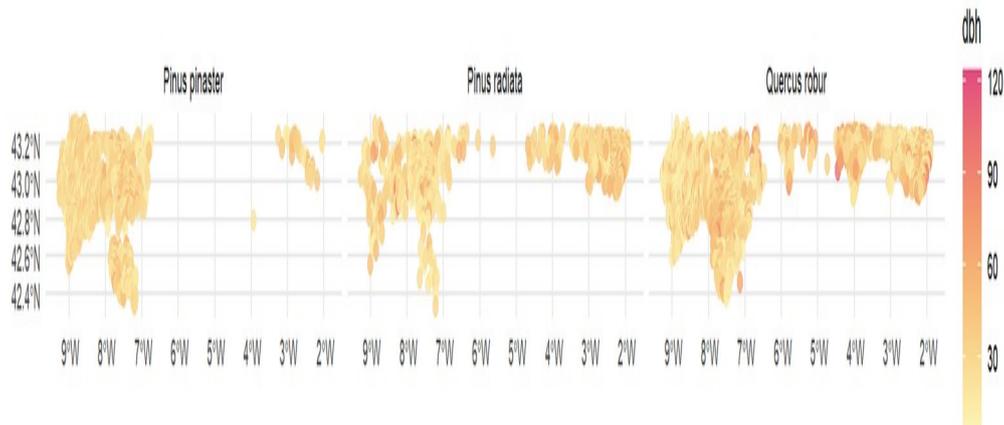


Figura 6. Ejemplo de visualización de datos del IFN obtenidos con *forestables*

5. Discusión

El paquete desarrollado tiene como objetivo principal **facilitar** la obtención y el manejo de datos de inventarios forestales, con miras a su uso en modelos de dinámica forestal.

Existen otros paquetes de R que sirven para descargar o procesar datos de inventarios individuales. Por ejemplo, el paquete *FIESTA* para el inventario americano (FIA) (<https://github.com/USDAForestService/FIESTA>), el paquete *basifoR* para el inventario español

(<https://cran.r-project.org/web/packages/basifoR/index.html>), y el paquete *FrenchNFIfindeR* para el inventario francés, (<https://github.com/Jeremy-borderieux/FrenchNFIfindeR>). Los paquetes *FIESTA* y *basifoR* se centran en el procesado de datos del inventario para calcular otras variables mientras que *FrenchNFIfindeR* se emplea exclusivamente para la descarga de datos. La



diferencia de estos paquetes respecto a *forestables* es que ninguno de ellos hace esfuerzos por armonizar y estandarizar datos de varios inventarios. *forestables* **no corrige errores ni inconsistencias** en los inventarios. Un ejemplo destacado es el del IFN, donde las diferencias en los listados de especies entre versiones pueden generar conflictos. Para abordar esto, el paquete usa internamente un listado unificador, pero no se garantiza la resolución de todas las situaciones problemáticas. Asimismo, los cambios en la asignación de códigos de especie o la desaparición de árboles entre versiones del IFN son aspectos que el usuario debe verificar por sí mismo.

Adicionalmente, las personas usuarias deben ser conscientes de que las **coordenadas geográficas proporcionadas no son precisas** debido a restricciones de seguridad y privacidad. En casos donde se necesite precisión, se recomienda contactar directamente con los organismos correspondientes. También es importante considerar que no todas las variables están incluidas en el paquete.

Por lo tanto, es necesario que las personas usuarias **consulten las fuentes oficiales** de los inventarios, especialmente si no están familiarizados con ellos, así como la **documentación del paquete** para poder interpretar y trabajar correctamente con los datos. Del mismo modo, el usuario debe asumir la **responsabilidad** de comprender las limitaciones de los inventarios y de la herramienta, y aplicar los procedimientos adecuados para tratar los datos satisfactoriamente.

Por último, el paquete se ha desarrollado con licencia de **código abierto** en GitHub, esto quiere decir que cualquier persona con los conocimientos técnicos adecuados puede colaborar para añadir nuevas funcionalidades al paquete. Sería interesante añadir inventarios de otros países en el futuro. Esto facilitaría la realización de análisis o simulaciones a escala europea, por ejemplo.

En resumen, el paquete constituye una **herramienta valiosa** para la estandarización y manejo de datos forestales por parte de técnicos, estudiantes e investigadores, pero exige un **uso informado y responsable** para garantizar la fiabilidad de los datos.

6. Conclusiones

El paquete de R *forestables* constituye una **herramienta accesible y de utilidad** para ecólogos, modeladores y otros investigadores interesados en el **estudio de los ecosistemas forestales**. Las funciones optimizan el acceso, la estandarización y la visualización de los datos, facilitando los ejercicios de comparación entre inventarios nacionales y/o versiones del IFN. Los *outputs* generados por *forestables* también pueden ser empleados como *inputs* de **modelos de dinámica forestal** tales como *medfate* o *medfateland*. Además, la filosofía *tidyverse* inherente al paquete permite **simplificar flujos de trabajo** complejos y optimizar la manipulación de grandes volúmenes de datos. Aunque actualmente solo soporta inventarios de EE. UU., Francia y España, su desarrollo en **código abierto** favorece la **colaboración** para incluir nuevos inventarios nacionales, lo que sin duda fortalecería la realización de estudios y simulaciones a nivel regional o global.

7. Agradecimientos

Este trabajo ha sido realizado en el marco del proyecto [PID2021-126679OB-I00], financiado por MCIN/AEI/10.13039/501100011033/ y cofinanciado por FEDER, 'Una



manera de hacer Europa'.

8. Bibliografía

DE CACERES, M., MARTIN-STPAUL, N., TURCO, M., CABON, A., & GRANDA, V. (2018). Estimating daily meteorological data and downscaling climate models over landscapes. *Environ. Model. and Softw.*, 108, 186-196. <https://doi.org/10.1016/j.envsoft.2018.08.003>.

DE CÁCERES, M., MOLOWNY-HORAS, R., CABON, A., MARTÍNEZ-VILALTA, J., MENCUCCINI, M., GARCÍA-VALDÉS, R., NADAL-SALA, D., SABATÉ, S., MARTIN-STPAUL, N., MORIN, X., D'ADAMO, F., BATLLORI, E., & AMÉZTEGUI, A. (2023a). MEDFATE 2.9.3: A trait-enabled model to simulate Mediterranean forest function and dynamics at regional scales. *Geosci. Model Dev.*, 16, 3165-3201. <https://doi.org/10.5194/gmd-16-3165-2023>

DE CÁCERES, M., GRANDA, V., CABON, A. & AMEZTEGUI, A. (2023b). emf-creaf/medfate: MEDFATE v2.9.3 (v2.9.3). Zenodo. <https://doi.org/10.5281/zenodo.7695331>

DE CÁCERES, M., AMÉZTEGUI, A., GONZÁLEZ, M., AQUILUÉ, N., CAVIEDES-VOULLIÈME, D., & MORALES-HERNÁNDEZ, M. (2024a). *medfateland: Mediterranean Landscape Simulation*. R package version 2.5.0, <https://github.com/emf-creaf/medfateland>, <https://emf-creaf.github.io/medfateland/>.

DE CÁCERES, M., MARTIN-STPAUL, N., TOVAR, A., & GRANDA, V. (2024b). *traits4models: From plant traits to model parameters*. R package version 0.2.0, <https://emf-creaf.github.io/traits4models/>.

GRANDA, V. (2024c). *meteospain: Access to Spanish Meteorological Stations Services*. R package version 0.2.0, <https://github.com/emf-creaf/meteospain>, <https://emf-creaf.github.io/meteospain/>.

GSCHWANTNER, T., RIEDEL, T., HENNING, L., ADAME, P., ADOLT, R., AGUIRRE, A., ALBERDI, I., ET AL. (2024). Improved large-area forest increment information in Europe through harmonisation of National Forest Inventories. *For. Ecol. Manage.*, 562, 121913. <https://doi.org/10.1016/j.foreco.2024.121913>

FRESCINO, T.S., MOISEN G.G., PATTERSON, P.L., TONEY, C., WHITE, G.W. (2023). *FIESTA: A forest inventory estimation and analysis R package*. USDA Forest Service, Rocky Mountain Research Station, 4919 S 1500 W, Ste. 200, Riverdale, UT, USA. tracey.frescino@usda.gov.

PEBESMA, E., & BIVAND, R. (2023). *Spatial Data Science: With Applications in R*. Chapman and Hall/CRC. <https://doi.org/10.1201/9780429459016>

PEBESMA, E., (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *R J.* 10 (1), 439-446, <https://doi.org/10.32614/RJ-2018-009>

TOVAR, A., GRANDA, V. (2024). *forestables: Making Forest Inventory Data Tidy*. R package version 0.0.6, <https://emf-creaf.github.io/forestables/>

VIDAL, C., ALBERDI, I., REDMOND, J., ET AL. (2016). The role of European National Forest Inventories for international forestry reporting. *Ann. For. Sci.*, 73, 793-806. <https://doi.org/10.1007/s13595-016-0545-6>

WICKHAM, H., FRANÇOIS, R., HENRY, L., MÜLLER, K., & VAUGHAN, D., (2023). *dplyr: A Grammar of Data Manipulation*. R package version 1.1.4. <https://CRAN.R->



[project.org/package=dplyr](https://CRAN.R-project.org/package=dplyr)

WICKHAM, H., VAUGHAN, D., & GIRLICH, M. (2024). *tidyr: Tidy Messy Data*. R package version 1.3.1. <https://CRAN.R-project.org/package=tidyr>