

Package: dhcurve (via r-universe)

August 22, 2024

Title Automated Modelling of Diameter Height Curves for Trees

Version 0.2

Description Model diameter height curves for individual tree species and forests.

License GPL-3

URL <https://inbo.github.io/dhcurve>, <https://github.com/inbo/dhcurve>

BugReports <https://github.com/inbo/dhcurve/issues>

Depends R (>= 3.3.0)

Imports assertthat, dplyr, DT, ggplot2, nlme, plotly, plyr, purrr, rlang, rmarkdown, stats, tidyr

Suggests knitr, stringr, testthat, tibble, xml2

VignetteBuilder knitr

Encoding UTF-8

Language nl-BE

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 7.2.3

Repository <https://inbo.r-universe.dev>

RemoteUrl <https://github.com/inbo/dhcurve>

RemoteRef HEAD

RemoteSha a7458d9460f1be33726e69047f80c40a869487b0

Contents

| | |
|---------------------------------|---|
| afwijkendeCurves | 2 |
| afwijkendeMetingen | 3 |
| curvekarakteristieken | 4 |
| dhcurvesrapport | 5 |
| fit.afgeleid | 6 |

| | |
|------------------------------------|-----------|
| fit.basis | 7 |
| fit.lokaal | 7 |
| hoogteschatting.afgeleid | 8 |
| hoogteschatting.basis | 9 |
| initiatie | 11 |
| modelparameters | 13 |
| outputIVANHO | 13 |
| resultaat | 15 |
| rmse.basis | 16 |
| rmse.verschuiving | 18 |
| testdata1domein | 19 |
| testdataset | 20 |
| validatie.afgeleid | 20 |
| validatie.basis | 22 |
| validatie.lokaal | 23 |
| validatie.uitbreiding | 25 |
| validatierapport | 27 |
| Index | 29 |

afwijkendeCurves *Lijst de afwijkende curves op*

Description

Functie die op basis van de curvekarakteristieken van de modellen (extremen en buigpunten) de mogelijk slechtste modellen oplijst met de reden. Deze functie kan ook gebruikt worden voor lokale modellen, in dat geval moet ook de dataset als parameter meegegeven worden.

Deze functie roept voor de berekening van de extremen en buigpunten de functie curvekarakteristieken() op.

Usage

```
afwijkendeCurves(Basismodel, Data = NULL)
```

Arguments

| | |
|------------|--|
| Basismodel | Model per boomsoort zoals teruggegeven door de functie fit.basis() of model per boomsoort-domeincombinatie zoals teruggegeven door de functie fit.lokaal() |
| Data | Dataset op basis waarvan het lokaal model berekend is (enkel nodig voor lokaal model) |

Value

Dataframe dat de mogelijk afwijkende curves oplijst, met volgende velden:

- BMS: Boomsort
- DOMEIN_ID
- Reden: reden waarom de curve afwijkend is
- Omtrek_Buigpunt.d: midden van omtrekklassse waarin het buigpunt van de curve van het domeinmodel ligt
- Omtrek_Extr_Hoogte.d: midden van omtrekklassse waarin het maximum van de curve van het domeinmodel ligt

| | |
|--------------------|---|
| afwijkendeMetingen | <i>Geeft de afwijkende metingen uit een gegeven model</i> |
|--------------------|---|

Description

Berekent afwijkende metingen, dit zijn metingen met een afwijking $> 2,5 * RMSE$, en geeft deze weer volgens dalende afwijking. Voor de 20 domeinen met de hoogste RMSE, waarbij de "foutenmarge" breder is, worden minstens 10 meetresultaten geselecteerd als "afwijkend" (nl. de 10 meetresultaten met de hoogste error).

Usage

```
afwijkendeMetingen(Dataset, AantalDomHogeRMSE = 20)
```

Arguments

| | |
|-------------------|---|
| Dataset | Dataframe met meetresultaten, geschatte waarden voor het domeinmodel en het Vlaamse model en de RMSE voor het domeinmodel. Deze dataframe moet dezelfde velden bevatten als de dataframes in de list teruggegeven in de functie <code>initiatie()</code> , en daarnaast de velden <code>H_D_finaal</code> , <code>rmseD</code> en <code>maxResid</code> . |
| AantalDomHogeRMSE | Standaard worden de 20 domeinen met de hoogste RMSE geselecteerd en voor elk van deze domeinen worden minstens 10 afwijkende metingen geselecteerd. <code>AantalDomHogeRMSE</code> laat toe om dit aantal van 20 domeinen aan te passen. |

Value

Lijst met afwijkende metingen ($> 2,5 * RMSE$), inclusief vlag uit databank

curvekarakteristieken *Berekent curvekarakteristieken van de domeinmodellen*

Description

Berekent de curvekarakteristieken die toelaten om mogelijke afwijkingen in de curvevorm te detecteren: extremen (min en max), buigpunt,...

Usage

```
curvekarakteristieken(Basismodel, Data = NULL)
```

Arguments

| | |
|------------|--|
| Basismodel | Model per boomsoort zoals teruggegeven door de functie <code>fit.basis()</code> of model per boomsoort-domeincombinatie zoals teruggegeven door de functie <code>fit.lokaal()</code> |
| Data | Dataset op basis waarvan het lokaal model berekend is (enkel nodig voor lokaal model) |

Value

Dataframe met curvekarakteristieken per domein en boomsoort met de velden:

- parameters Ad, Bd en Cd van het model en enkele generieke velden (DOMEIN_ID, BMS,...)
- `Omtrek_Extr_Hoogte.d`: midden van omtrekklassen waarin een extreem (minimum of maximum) van de curve van het domeinmodel ligt
- `Extr_Hoogte.d`: hoogteschatting die aansluit bij het vorige veld
- `Hoogteverschil.d`: verschil tussen de hoogteschattingen van het extreem van de curve en de bovengrens van het bruikbaar interval. Deze maat is relevant als het extreem een maximum is (enkel dan is deze waarde positief).
- `Omtrek_Buigpunt.d`: midden van omtrekklassen waarin een buigpunt van de curve van het domeinmodel ligt (deze en volgende variabelen zijn relevante maten bij het voorkomen van een minimum)
- `Verschil_rico_BP_Q5.d`: het verschil in de waarde van de richtingscoëfficiënt tussen het buigpunt en de ondergrens van het bruikbaar interval.
- `Verschil_rico_BP_Q5_per_omtrek.d`: de voorgaande variabele gedeeld door het verschil in omtrek tussen het buigpunt en de ondergrens van het bruikbaar interval.

Ingeval van een basismodel worden deze variabelen aangevuld met dezelfde variabelen voor het Vlaams model. Hierbij is de 'd' op het einde van de variabelenaam vervangen door 'vl'.

dhcurvesrapport *Stelt de finale diameter-hoogtecurves grafisch voor in een rapport*

Description

De functie genereert een rapport (.html-bestand) in de working directory (of opgegeven directory) met informatie en grafieken van de finale modellen.

Deze functie rendert het bestand dhcurverapport.Rmd, dat voor elke boomsoort-domeincombinatie child dhcurve.Rmd toevoegt.

Usage

```
dhcurvesrapport(  
  OutputIVANHO,  
  Datalijst,  
  Bestandsnaam = "dhcurves.html",  
  KleurUitbreiding = FALSE,  
  verbose = TRUE,  
  PathWD = getwd()  
)
```

Arguments

| | |
|------------------|---|
| OutputIVANHO | Lijst met geschatte hoogte per omtrekklassse (dataframe zoals gegenereerd door de functie outputIVANHO()). |
| Datalijst | Datasets met gemeten waarden die gebruikt zijn voor de modellen, zoals gegenereerd door de functie initiatie(). |
| Bestandsnaam | Een naam voor het rapport (.html-bestand) dat gegenereerd wordt, bestaande uit een string die eindigt op .html |
| KleurUitbreiding | Moeten de metingen van de uitbreiding in een andere kleur weergegeven worden? Default is FALSE, waarbij (eventuele) metingen van een uitbreiding niet in een andere kleur weergegeven worden. |
| verbose | Dit geeft de toestand van het systeem aan en zorgt ervoor dat boodschappen niet onnodig gegeven worden. (Defaultwaarde behouden.) |
| PathWD | Het path van de working directory, dus het path waarin het validatierapport opgeslagen moet worden (default wordt het in de op dat moment actieve working directory opgeslagen). |

Value

De functie genereert in de working directory (of opgegeven directory) een rapport (html) met de te controleren modellen. Hierin wordt per model (boomsoort-domeincombinatie) de volgende algemene informatie vermeld: boomsoort, domein (en ID), modeltype, aantal metingen, RMSE en bruikbaar interval.

Daaronder wordt telkens grafisch volgende info weergegeven:

- een puntenwolk die de metingen voorstelt (geen individuele metingen, maar een jitter)
- curve van het domeinmodel, aangeduid met zwarte ruiten
- grenzen van het bruikbaar interval (curves eindigen bij de klassenmiddens die overeenkomen met deze grenzen)

| | |
|--------------|---|
| fit.afgeleid | <i>Fit afgeleid model op basis van het basismodel en de opgegeven dataset</i> |
|--------------|---|

Description

Functie die op basis van de opgegeven dataset een model opstelt per boomsoort-domeincombinatie door verschuiving van het Vlaams model voor die boomsoort (afgeleid uit het basismodel). Dit komt overeen met het fitten van het lineair model $Hoogte_{afgeleid} \sim \delta A + Hoogte_{Vlaams}$ op basis van de opgegeven dataset, waarbij $Hoogte_{Vlaams}$ de hoogteschatting is volgens het Vlaams model.

Usage

```
fit.afgeleid(Data.afgeleid, Basismodel)
```

Arguments

| | |
|---------------|--|
| Data.afgeleid | De dataframe "Afgeleid" uit de list die teruggegeven wordt bij de functie <code>initiatie()</code> . (Een dataframe met minimaal de velden BMS (boomsoort), IDbms (identificatienummer van de boomsoort), DOMEIN_ID (identificatienummer van het domein), BOS_BHI (domeinnaam), HOOGTE, Omtrek, logOmtrek, logOmtrek2, Status (van data: "niet gecontroleerd", "gecontroleerd", ...), nBomen (aantal bomen per boomsoort-domeincombinatie), nBomenOmtrek05 (aantal bomen per boomsoort-domeincombinatie met omtrek > 0.5 m), nBomenInterval (aantal bomen per boomsoort-domeincombinatie binnen bruikbaar interval), nBomenIntervalOmtrek05 (aantal bomen per boomsoort-domeincombinatie binnen bruikbaar interval met omtrek > 0.5 m), Q5k en Q95k ("bruikbaar interval").) |
| Basismodel | Model per boomsoort zoals teruggegeven door de functie <code>fit.basis()</code> : tibble met de velden BMS (boomsoort) en Model (lme-object met het gefit mixed model voor die boomsoort) |

Value

List met 2 tibbles:

- tibble met velden BMS (boomsoort), DOMEIN_ID en Model (lme-object van model per boomsoort-domeincombinatie)
- tibble met de ingevoerde dataset, waaraan het veld `H_VL_finaal` (de hoogteschatting volgens het Vlaams model) toegevoegd is

fit.basis

Fit basismodel op basis van de opgegeven dataset

Description

Functie die het lineair mixed model $Hoogte_{basis} \sim A_{Vlaams} + A_{domein} + (B_{Vlaams} + B_{domein}) \log(Omtrek) + (C_{Vlaams} + C_{domein}) \log(Omtrek)^2$ fit op basis van de opgegeven dataset.

Usage

```
fit.basis(Data.basis)
```

Arguments

`Data.basis` De dataframe `Basis` uit de list die teruggegeven wordt bij de functie `initiatie()`. (Een dataframe met minimaal de velden `BMS` (boomsoort), `IDbms` (identificatienummer van de boomsoort), `DOMEIN_ID` (identificatienummer van het domein), `BOS_BHI` (domeinnaam), `HOOGTE`, `Omtrek`, `logOmtrek`, `logOmtrek2`, `Status` (van data: "niet gecontroleerd", "gecontroleerd",...), `nBomen` (aantal bomen per boomsoort-domeincombinatie), `nBomenOmtrek05` (aantal bomen per boomsoort-domeincombinatie met omtrek > 0.5 m), `nBomenInterval` (aantal bomen per boomsoort-domeincombinatie binnen bruikbaar interval), `nBomenIntervalOmtrek05` (aantal bomen per boomsoort-domeincombinatie binnen bruikbaar interval met omtrek > 0.5 m), `Q5k` en `Q95k` ("bruikbaar interval").)

Value

Dataframe (tibble) met de velden `BMS` (boomsoort) en `Model` (lme-object met het gefit mixed model voor die boomsoort)

fit.lokaal

Lokaal model fitten op basis van de opgegeven dataset

Description

Functie die het lineair (fixed) model $Hoogte_{lokaal} \sim A_{domein} + B_{domein} \log(Omtrek) + C_{domein} \log(Omtrek)^2$ fit op basis van de opgegeven dataset.

Usage

```
fit.lokaal(Data.lokaal)
```

Arguments

`Data.lokaal` De dataframe Lokaal uit de list die teruggegeven wordt bij de functie `initiatie()`. (Een dataframe met minimaal de velden `BMS` (boomsoort), `IDbms` (identificatienummer van de boomsoort), `DOMEIN_ID` (identificatienummer van het domein), `BOS_BHI` (domeinnaam), `HOOGTE`, `Omtrek`, `logOmtrek`, `logOmtrek2`, `Status` (van data: "niet gecontroleerd", "gecontroleerd",...), `nBomen` (aantal bomen per boomsoort-domeincombinatie), `nBomenOmtrek05` (aantal bomen per boomsoort-domeincombinatie met omtrek > 0.5 m), `nBomenInterval` (aantal bomen per boomsoort-domeincombinatie binnen bruikbaar interval), `nBomenIntervalOmtrek05` (aantal bomen per boomsoort-domeincombinatie binnen bruikbaar interval met omtrek > 0.5 m), `Q5k` en `Q95k` ("bruikbaar interval").)

Value

Dataframe (tibble) met de velden `BMS` (boomsoort), `DOMEIN_ID` en `Model` (lm-object met het gefit lineair model voor die boomsoort-domeincombinatie).

hoogteschatting.afgeleid

Hoogteschatting op basis van opgegeven afgeleid model

Description

Functie die de gemiddelde hoogte per omtrekklassie schat voor de domeincurves en Vlaamse curves van het opgegeven afgeleid model. De teruggegeven dataframe kan gebruikt worden om grafieken te maken of afwijkende metingen te bestuderen. Opgelet! In tegenstelling tot de meeste functies van dit package werkt deze functie op basis van 1 model en de bijhorende meetgegevens. Zie voorbeeld voor een methode om deze functie te kunnen toepassen op de volledige dataset `Afgeleidmodel`.

Usage

```
hoogteschatting.afgeleid(Domeinsoortmodel, Domeinsoortdata)
```

Arguments

`Domeinsoortmodel`
 verschoven Vlaams model voor 1 boomsoort-domeincombinatie

`Domeinsoortdata`
 de gegevens die hierbij horen: meetresultaten voor 1 boomsoort-domeincombinatie

Value

dataframe met de meetresultaten en de schattingen van de hoogtes voor het domeinmodel en de Vlaamse model

Examples

```

library(dplyr)

#Datasets inladen en het basismodel en afgeleid model berekenen
Data <- testdataset()
Datalijst <- initiatie(Data)
Data.basis <- Datalijst[["Basis"]]
Basismodel <- fit.basis(Data.basis)
Data.afgeleid <- Datalijst[["Afgeleid"]]
Afgeleidmodel <- fit.afgeleid(Data.afgeleid, Basismodel)

#De hoogteschatting voor een afgeleid model
Afgeleidmodel[[1]] %>%
  inner_join(
    Afgeleidmodel[[2]],
    by = c("BMS", "DOMEIN_ID")
  ) %>%
  group_by(
    .data$BMS,
    .data$DOMEIN_ID
  ) %>%
  do(
    hoogteschatting.afgeleid(.$Model[[1]],
                             select(., -Model))
  ) %>%
  ungroup()

```

hoogteschatting.basis *Hoogteschatting op basis van opgegeven basismodel*

Description

Functie die de gemiddelde hoogte per omtrekklasse schat voor de domeincurves en Vlaamse curves van het opgegeven basismodel. De teruggegeven dataframe kan gebruikt worden om grafieken te maken of afwijkende metingen te bestuderen. *Opgelet!* In tegenstelling tot de meeste functies van dit package werkt deze functie op basis van 1 model en de bijhorende meetgegevens. Zie voorbeelden voor een methode om deze functie te kunnen toepassen op het Basismodel of het Lokaalmodel en Data.lokaal.

Usage

```

hoogteschatting.basis(
  Soortmodel,
  Soortdata,
  Typemodel,
  BMS,
  Uitbreiding = FALSE
)

```

Arguments

| | |
|-------------|---|
| Soortmodel | model voor één boomsoort (basis) of één boomsoort-domeincombinatie (lokaal) |
| Soortdata | meetgegevens van boomsoort (basis) of boomsoort-domeincombinatie (lokaal) |
| Typemodel | "Basis" of "Lokaal"? |
| BMS | Boomsoort |
| Uitbreiding | Gaat het hier over berekening voor een uitbreiding? (Facultatief argument om met hogere hoogteklassen te kunnen omgaan.) Default is FALSE, wat betekent dat het niet over een uitbreiding gaat. |

Value

dataframe met de meetresultaten en de schattingen van de hoogtes voor het domeinmodel en de Vlaamse model

Examples

```
library(dplyr)

#Dataset inladen en het basismodel berekenen
Data <- testdataset()
Datalijst <- initiatie(Data)
Data.basis <- Datalijst[["Basis"]]
Basismodel <- fit.basis(Data.basis)

#De hoogteschatting voor een basismodel
Basismodel %>%
  rowwise() %>%
  do(
    hoogteschatting.basis(.$Model, .$Model$data, "Basis", .$BMS)
  ) %>%
  ungroup()

#Dataset inladen en het lokaal model berekenen
Data.lokaal <- Data.basis %>%
  filter(DOMEIN_ID == "A")
Lokaalmodel <- fit.lokaal(Data.lokaal)

#De hoogteschatting voor een lokaal model
Lokaalmodel %>%
  inner_join(
    Data.lokaal,
    by = c("BMS", "DOMEIN_ID")
  ) %>%
  group_by(
    BMS,
    DOMEIN_ID
  ) %>%
  do(
    hoogteschatting.basis(.$Model[[1]],
                          select(., -Model),
```

```

                                "Lokaal", unique(.$BMS))
) %>%
ungroup()

```

initiatie

Initiatiestap met opsplitsen van dataset volgens modeltypen

Description

De functie `initiatie()` maakt de dataset klaar voor de verdere stappen van het opstellen van diameter-hoogtecurves: ze berekent de noodzakelijke variabelen en splitst de dataset op basis van het aantal metingen per boomsoort-domeincombinatie in 4 delen: 3 delen die als parameter meegegeven kunnen worden voor de `fit.xxx()`-functies per modeltype (basismodel, afgeleid model en lokaal model), en een vierde deel met de resterende gegevens.

Usage

```

initiatie(
  Data,
  Uitzonderingen = data.frame(DOMEIN_ID = "", BMS = "", min_basis = NA_integer_,
    min_afgeleid = NA_integer_, stringsAsFactors = FALSE),
  Bestandsnaam = "VerwijderdeGegevensInitiatie.html",
  verbose = TRUE,
  PathWD = getwd()
)

```

Arguments

- | | |
|----------------|--|
| Data | Dataframe met metingen van bomen die als basis moeten dienen om diameter-hoogtecurves op te stellen. De dataframe moet de velden <code>DOMEIN_ID</code> (identificatienummer voor domein), <code>BOS_BHI</code> (domeinnaam), <code>IDbms</code> (identificatienummer voor boomsoort), <code>BMS</code> (boomsoort), <code>C13</code> (omtrek in centimeter, gemeten op 1,3 m hoogte), <code>HOOGTE</code> (in meter) en <code>Status</code> bevatten en mag eventueel velden <code>TYPE_METING</code> en <code>JAAR</code> bevatten (die worden bij <code>rmse.basis()</code> als groeperende variabelen gebruikt). <code>Status</code> mag enkel gegevens met status "Niet gecontroleerd", "Te controleren" of "Goedgekeurd" bevatten, dus gegevens met status "Afgekeurd" of "Negeren" moeten vooraf verwijderd worden. |
| Uitzonderingen | Lijst met boomsoort-domeincombinaties waarvoor uitzonderingen gelden voor de limieten van minimum 50 en minimum 10 bomen. De dataframe moet de velden <code>DOMEIN_ID</code> , <code>BMS</code> , <code>min_basis</code> (= vervangende waarde voor 50) en <code>min_afgeleid</code> (= vervangende waarde voor 10) bevatten. Daarnaast kan in een extra veld <code>min_uitbreiden_model</code> aangegeven worden vanaf hoeveel metingen een uitbreiding naar hogere omtrekklassen opnieuw bekeken moet worden (= vervangende waarde voor 10). De waarde <code>NA</code> geeft telkens aan dat de default gebruikt mag worden. De default voor dit argument is dat er geen uitzonderingen zijn (wat sowieso het geval is als de curves nog niet gevalideerd zijn). |

| | |
|--------------|--|
| Bestandsnaam | Een naam voor het validatierapport (.html-bestand) dat gegenereerd wordt, bestaande uit een string die eindigt op .html |
| verbose | Dit geeft de toestand van het systeem aan en zorgt ervoor dat boodschappen niet onnodig gegeven worden. (Defaultwaarde behouden.) |
| PathWD | Het path van de working directory, dus het path waarin het validatierapport opgeslagen moet worden (default wordt het in de op dat moment actieve working directory opgeslagen). |

Value

Als er gegevens verwijderd worden, genereert de functie een validatierapport (.html-bestand) waarin een overzicht gegeven wordt van de verwijderde gegevens, dit zijn gegevens met omtrek > 3 m en omtrek < 0.2 m.

De functie geeft een list van dataframes terug, met in elke dataframe behalve de variabelen uit Data de berekende variabelen Omtrek (= omtrekklassen), logOmtrek, logOmtrek2, Q5k en Q95k (de grenzen van het bruikbaar interval), nBomen (= aantal metingen behalve de verwijderde gegevens), nBomenOmtrek05 (aantal metingen met omtrek > 0.5 m en < 2.8 m), nBomenInterval (= aantal metingen binnen het bruikbaar interval), nBomenIntervalOmtrek05 (aantal metingen binnen het bruikbaar interval met omtrek > 0.5 m) en nExtra (aantal metingen boven het bruikbaar interval op basis waarvan een uitbreiding van de curve naar hogere omtrekklassen gevalideerd zou kunnen worden)

De 4 dataframes die achtereenvolgens in de list zitten, zijn:

- [["Basis"]] gegevens van boomsoorten waarvoor meer dan 50 metingen (binnen het bruikbaar interval met omtrek > 0.5 m) beschikbaar zijn in minimum 6 domeinen, waarbij enkel gegevens worden opgenomen van de domeinen waarvoor minimum 50 metingen beschikbaar zijn. Op basis van deze dataset kan een basismodel berekend worden, bestaande uit een Vlaams model per boomsoort en domeinspecifieke modellen.
- [["Afgeleid"]] gegevens van domeinen met minder metingen (< 50 metingen binnen het bruikbaar interval en > 10 metingen boven 0.5 m) van boomsoorten waarvoor een Vlaams model berekend kan worden (dus boomsoorten die in dataset "Basis" voorkomen), op basis waarvan een afgeleid model berekend kan worden.
- [["Lokaal"]] gegevens van domeinen met veel metingen voor een boomsoort (> 50 metingen binnen het bruikbaar interval met omtrek > 0.5 m) waarvan er te weinig domeinen (< 6) zijn met voldoende metingen om een Vlaams model op te stellen. Voor deze boomsoort-domeincombinaties kan een lokaal model berekend worden.
- [["Rest"]] metingen van de boomsoort-domeincombinaties die niet tot de 3 voorgaande categorieën behoren en waar dus geen model voor opgesteld kan worden.

Voor de eerste en derde dataframe worden metingen binnen het bruikbaar interval gemarkeerd als VoorModelFit en ook metingen boven dit interval tot een omtrek van 3 m worden bijgehouden voor een eventuele uitbreiding van het model (tot maximaal 10 omtrekklassen); voor het afgeleid model (2de dataframe) worden de metingen met omtrek tussen 0,5 m en 2,8 m bijgehouden.

In geval er gegevens verwijderd zijn, wordt aan de list een extra dataframe [["VerwijderdeGegevens"]] toegevoegd met de gegevens uit het validatierapport.

| | |
|-----------------|--|
| modelparameters | <i>Berekent modelparameters voor opgegeven model</i> |
|-----------------|--|

Description

Functie die de modelparameters berekent op basis van een opgegeven basismodel, lokaal model of afgeleid model. Ze berekent de parameters voor het domeinmodel en ingeval van het basismodel ook voor het Vlaams model, en geeft ook de grenzen van het bruikbaar interval.

(Deze functie verwijst naar de interne functies `modelparameters.basis()`, `modelparameters.lokaal()` of `modelparameters.afgeleid()`, afhankelijk van de situatie)

Usage

```
modelparameters(Basismodel, Data = NULL, Afgeleidmodel = NULL)
```

Arguments

| | |
|---------------|---|
| Basismodel | model per boomsoort (basismodel) of model per boomsoort-domeincombinatie (lokaalmodel) |
| Data | meetgegevens (enkel nodig voor lokaal model) |
| Afgeleidmodel | voor de berekening van de modelparameters van het afgeleid model moeten zowel het basismodel als het afgeleid model gegeven worden, dus in dit geval wordt hier het afgeleid model meegegeven. Voor de andere modellen mag dit argument niet toegevoegd worden. |

Value

Dataframe met parameters voor domeinmodel (Ad, Bd en Cd) en ingeval van het basismodel de parameters voor Vlaams model (Av1, Bv1 en Cv1). Ingeval van een afgeleid model worden de parameters voor het Vlaams model gegeven (Av1, Bv1 en Cv1), en een parameter Ad die de verschuiving van het Vlaams model naar het afgeleide domeinmodel weergeeft (dus een extra intercept)

| | |
|--------------|---|
| outputIVANHO | <i>Berekent de geschatte hoogtes per omtrekklasse</i> |
|--------------|---|

Description

De functie `outputIVANHO()` voert de berekeningen uit die leiden tot de gevraagde tabel om in IVANHO te importeren: een tabel met per boomsoort, domein en omtrekklasse een schatting van de hoogte. Als de curve een maximum hoogte vertoont binnen het bestudeerde interval, wordt deze maximumwaarde als hoogte meegegeven aan alle omtrekclasses hoger dan de omtrekklasse van dit maximum. (Dus verschillend van de validatierapporten daalt de hoogte hier niet terug na het maximum, maar de hoogste waarde wordt aangehouden.) Als de curve een minimum hoogte vertoont binnen het bestudeerde interval, wordt deze minimumwaarde als hoogte meegegeven aan

alle omtrekklassen lager dan de omtrekklassse van dit minimum. Als voor een bepaalde omtrekklassse de hoogteschatting lager is dan 2,5 m, dan wordt deze omtrekklassse weggelaten.

Voor deze functie worden volgende hulpfuncties aangeroepen: `hoogteschatting.basis()`, `hoogteschatting.afgeleid()` en `curvekarakteristieken()`.

Usage

```
outputIVANHO(
  Basismodel = NULL,
  Afgeleidmodel = NULL,
  Lokaalmodel = NULL,
  Data.lokaal = NULL,
  Uitbreiding = NULL
)
```

Arguments

| | |
|---------------|---|
| Basismodel | Model per boomsoort zoals teruggegeven door de functie <code>fit.basis()</code> : tibble met de velden BMS (boomsoort) en Model (lme-object met het gefit mixed model voor die boomsoort). |
| Afgeleidmodel | Model per boomsoort-domeincombinatie zoals teruggegeven door de functie <code>fit.afgeleid()</code> : list met 2 tibbles. |
| Lokaalmodel | Model per boomsoort-domeincombinatie zoals teruggegeven door de functie <code>fit.lokaal()</code> : tibble met de velden BMS (boomsoort), DOMEIN_ID en Model (lm-object met het gefit lineair model voor die boomsoort-domeincombinatie). |
| Data.lokaal | Dataset op basis waarvan het opgegeven lokaal model berekend is. |
| Uitbreiding | Dataset met velden BMS, DOMEIN_ID en MaxOmtrek die teruggegeven wordt door functie <code>validatie.uitbreiding()</code> . Als zowel een Basismodel en als Lokaal model opgegeven worden, voeg dan de uitvoer van <code>validatie.uitbreiding()</code> van beide modeltypen samen. |

Value

Dataframe met geschatte hoogtes per domein en per boomsoort met velden:

- DOMEIN_ID: domeincode
- BOS_BHI: domeinnaam
- IDbms (identificatienummer van de boomsoort)
- BMS: boomsoort
- Modeltype ("basismodel"(= eigen model op basis van mixed model) of "afgeleid model"(= verschoven Vlaams model, afgeleid van fixed factor uit basismodel) of "Vlaams model"(= fixed factor uit basismodel, niet toegevoegd omdat niet relevant) of "lokaal model"(= apart model voor 1 boomsoort-domeincombinatie) of "geen model"(= boomsoort-domeincombinatie waarvoor geen model berekend kan worden))
- Omtrek: klassenmidden van omtrekklassse
- OmtrekklassetypeID: het overeenkomstige identificatienummer van de omtrekklassse

- Omtrekklassse
- Hoogte: de geschatte hoogte
- RMSE (root mean square error, zie vignet voor meer info: `vignette("Handleiding", package = "dhcurve")`)

 resultaat

Berekent de modelparameters per domein

Description

De functie resultaat berekent op basis van de opgegeven modellen voor elke boomsoort-domeincombinatie de modelparameters A, B en C voor een model van de vorm $Hoogte \sim A + B \log(Omtrek) + C \log(Omtrek)^2$.

Voor deze functie worden volgende hulpfuncties aangeroepen: `modelparameters()`, `rsme.basis()` en `rsme.afgeleid()`

Verder worden parameters A, B en C uit het model gehaald en een aantal in de functie `initiatie()` berekende gegevens toegevoegd.

Usage

```
resultaat(
  Basismodel = NULL,
  Afgeleidmodel = NULL,
  Lokaalmodel = NULL,
  Data.lokaal = NULL,
  Data.onbruikbaar = NULL
)
```

Arguments

- | | |
|------------------|---|
| Basismodel | Model per boomsoort zoals teruggegeven door de functie <code>fit.basis()</code> : tibble met de velden BMS (boomsoort) en Model (lme-object met het gefit mixed model voor die boomsoort). |
| Afgeleidmodel | Model per boomsoort-domeincombinatie zoals teruggegeven door de functie <code>fit.afgeleid()</code> : list met 2 tibbles. |
| Lokaalmodel | Model per boomsoort-domeincombinatie zoals teruggegeven door de functie <code>fit.lokaal()</code> : tibble met de velden BMS (boomsoort), DOMEIN_ID en Model (lm-object met het gefit lineair model voor die boomsoort-domeincombinatie). |
| Data.lokaal | Dataset op basis waarvan het opgegeven lokaal model berekend is. |
| Data.onbruikbaar | Evt. lijst met meetresultaten van boomsoort-domeincombinaties waarvoor geen model opgesteld kan worden, in de vorm van een dataframe zoals de dataframe Rest uit de list die door de functie <code>initiatie()</code> teruggegeven wordt. |

Value

Dataframe met modellen per domein en per boomsoort met velden:

- `DomeinID` (identificatienummer voor domein)
- `BMS` (boomsoort)
- `Modeltype` ("basismodel" (= eigen model op basis van mixed model) of "afgeleid model" (= verschoven Vlaams model, afgeleid van fixed factor uit basismodel) of "Vlaams model" (= fixed factor uit basismodel, niet toegevoegd omdat niet relevant) of "lokaal model" (= eigen model voor 1 boomsoort-domeincombinatie) of "geen model" (= boomsoort-domeincombinatie waarvoor minstens 1 boom opgemeten is maar geen model berekend kan worden))
- parameters A, B en C (zie 'Description' bovenaan)
- bruikbaar interval (Q5k en Q95k, zie vignet voor meer info)
- `RMSE` (root mean square error, zie vignet voor meer info)
- `nBomen` (totaal aantal opgemeten bomen met omtrek tussen 0,2 en 3,0 m)
- `nBomenOmtrek05` (aantal metingen met omtrek tussen 0.5 en 2,8 m, dus waarop afgeleid model gebaseerd is)
- `nBomenInterval` (aantal metingen binnen bruikbaar interval, dus waarop basismodel of lokaal model gebaseerd is)
- `nBomenIntervalOmtrek05` (aantal metingen binnen bruikbaar interval met omtrek > 0.5 m, dus waarop RMSE-berekening gebaseerd is)
- `nExtra` (aantal metingen boven bruikbaar interval waarop een eventuele uitbreiding gebaseerd is)

evt. kan een tweede dataframe toegevoegd worden met Vlaamse modellen per boomsoort, of deze kan toegevoegd worden aan de vorige dataframe, waarbij `DomeinID` leeg gelaten wordt of een specifieke waarde "Vlaams model" krijgt

rmse.basis

Berekent RMSE van basismodel

Description

Deze functie berekent de RMSE door cross-validatie op basis van 6 subsets. Deze functie kan ook gebruikt worden voor het lokaal model (ze bepaalt het verschil tussen de datasets op basis van het al dan niet aanwezig zijn van een veld `DOMEIN_ID` in de dataset). **Opgelet!** In tegenstelling tot de meeste functies van dit package werkt deze functie op basis van de meetgegevens van 1 model. Zie voorbeelden voor een methode om deze functie te kunnen toepassen vertrekkend van meetgegevens (bv. `Data.lokaal`) of vertrekkend van een model waar de meetgegevens uit gehaald kunnen worden (bv. `Basismodel`).

Deze functie berekent de RMSE op basis van testgroepen en omvat de volgende deelstappen:

- metingen opdelen in 6 testgroepen (veld testgroep)
- modellen fitten voor testgroepen, waarbij ze de functie `fit.basis()` 6 keer oproept
- RMSE berekenen voor domeinmodellen en Vlaams model op basis van de gemeten waarden en schattingen voor de testgroepen

Usage

```
rmse.basis(Data, Typemodel, BMS)
```

Arguments

| | |
|-----------|--|
| Data | Meetgegevens van één boomsoort-domeincombinatie (dataframe zoals de dataframes die in de list teruggegeven worden door de functie <code>initiatie()</code>) |
| Typemodel | "Basis" of "Lokaal"? |
| BMS | Boomsoort |

Value

Dataframe met `RMSE_domein` (`rmseD`), `RMSE_Vlaams` (`rmseVL`, niet voor lokaal model) en `maxResid`

Examples

```
library(dplyr)

#Dataset inladen voor het basismodel
Data <- testdataset()
Datalijst <- initiatie(Data)
Data.basis <- Datalijst[["Basis"]]

#De RMSE berekenen voor een basismodel op basis van de dataset
Data.basis %>%
  group_by(
    BMS
  ) %>%
  do(
    rmse.basis(., "Basis", .data$BMS)
  ) %>%
  ungroup()

#Dataset inladen voor het lokaal model
Data.lokaal <- Data.basis %>%
  filter(DOMEIN_ID == "A")

#De rmse berekenen voor een lokaal model
Data.lokaal %>%
  group_by(
    BMS,
    DOMEIN_ID
  ) %>%
  do(
    rmse.basis(., "Lokaal", .data$BMS)
  ) %>%
  ungroup()
```

| | |
|-------------------|---|
| rmse.verschuiving | <i>Berekent RMSE van afgeleid model</i> |
|-------------------|---|

Description

Deze functie berekent de RMSE op basis van verschil tussen de hoogteschatting volgens het domeinmodel en de gemeten hoogte (voor omtrekklassen > 0.5 m). Dit is slechts een deel van de totale RMSE van het afgeleid model, want behalve deze RMSE van de verschuiving moet hier ook de RMSE van het Vlaams model (basismodel) in rekening gebracht worden. Opgelet! In tegenstelling tot de meeste functies van dit package werkt deze functie op basis van de meetgegevens van 1 model. Zie voorbeeld voor een methode om deze functie te kunnen toepassen vertrekkend van het berekende model.

Usage

```
rmse.verschuiving(Verschovenmodel, Boomsoort, Domein)
```

Arguments

| | |
|-----------------|--|
| Verschovenmodel | Afgeleid model voor één boomsoort-domeincombinatie (lm-object) |
| Boomsoort | BMS |
| Domein | DOMEIN_ID |

Value

Dataframe met BMS (boomsoort), DOMEIN_ID, RmseVerschuiving

Examples

```
library(dplyr)

# Datasets inladen en het basismodel en afgeleid model berekenen
Data <- testdataset()
Datalijst <- initiatie(Data)
Data.basis <- Datalijst[["Basis"]]
Basismodel <- fit.basis(Data.basis)
Data.afgeleid <- Datalijst[["Afgeleid"]]
Afgeleidmodel <- fit.afgeleid(Data.afgeleid, Basismodel)

# De RMSE berekenen voor de verschuiving van het Vlaams model naar een
# afgeleid model
Afgeleidmodel[[1]] %>%
  rowwise() %>%
  do(
    rmse.verschuiving(.$Model, .$BMS, .$DOMEIN_ID)
  ) %>%
  ungroup()
```

```
#Nota: voor een berekening van de volledige RMSE van een afgeleid model  
#moet ook de RMSE van het basismodel in rekening gebracht worden
```

| | |
|-----------------|---|
| testdata1domein | <i>Genereert testdata voor 1 boomsoort-domeincombinatie</i> |
|-----------------|---|

Description

Deze functie genereert een testdataset voor 1 domein die voldoet aan de opgegeven vereisten.

Usage

```
testdata1domein(  
  nBomen = 100,  
  minOmtrek = 20,  
  maxOmtrek = 239,  
  A = 30,  
  B = 15,  
  C = 1,  
  sd = 2  
)
```

Arguments

| | |
|-----------|---|
| nBomen | Aantal gemeten bomen voor het domein |
| minOmtrek | Minimale omtrek voor de data |
| maxOmtrek | Maximale omtrek voor de data |
| A | Parameter voor het intercept van de functie |
| B | Parameter voor de eerstegraadsterm |
| C | Parameter voor de tweedegraadsterm |
| sd | foutenmarge t.o.v. model voor de data |

Value

Dataframe met testdata voor 1 boomsoort-domeincombinatie met velden C13 en H00GTE.

| | |
|-------------|------------------------------|
| testdataset | <i>Genereert testdataset</i> |
|-------------|------------------------------|

Description

Deze functie genereert een testdataset die voldoet aan de opgegeven vereisten. De testdataset kan gebruikt worden om het package te testen, ze voldoet als invoer voor functie `initiatie()`

Usage

```
testdataset(
  nBomenDomein = c(200, 100, 500, 100, 100, 100, 50, 20, 8),
  BMS = "testboom",
  IDbms = 1,
  minOmtrek = 20,
  maxOmtrek = 239
)
```

Arguments

| | |
|---------------------------|--|
| <code>nBomenDomein</code> | Vector met aantal gemeten bomen per domein |
| <code>BMS</code> | Naam van boomsoort |
| <code>IDbms</code> | Identificatienummer van boomsoort |
| <code>minOmtrek</code> | Minimale omtrek voor de data |
| <code>maxOmtrek</code> | Maximale omtrek voor de data |

Value

Dataframe met testdata voor 1 boomsoort met velden `BMS`, `IDbms`, `DOMEIN_ID`, `BOS_BHI`, `C13`, `HOOGTE`, `Status` en `ID`.

| | |
|--------------------|---|
| validatie.afgeleid | <i>Validatie van het afgeleid model</i> |
|--------------------|---|

Description

Functie die de validatie uitvoert op het verschoven Vlaams model en een overzicht geeft van de afwijkende metingen (zodat de gebruiker deze kan valideren).

`validatie.afgeleid()` roept meerdere hulpfuncties op:

- `rmse.basis()` en `rmse.verschuiving()`
- `afwijkendeMetingen()`
- `validatierapport()`

Voorafgaand aan het uitvoeren van deze laatste functie worden eerst de slechtste modellen opgelijst (op basis van RMSE en afwijkende metingen).

Usage

```

validatie.afgeleid(
  Basismodel,
  Afgeleidmodel,
  AantalDomHogeRMSE = 20,
  ExtraCurvesRapport = NULL,
  Bestandsnaam = "Validatie_Afgeleid.html",
  TypeRapport = "Dynamisch",
  PathWD = getwd()
)

```

Arguments

| | |
|--------------------|---|
| Basismodel | Model per boomsoort zoals teruggegeven door de functie <code>fit.basis()</code> : tibble met de velden BMS (boomsoort) en Model (lme-object met het gefit mixed model voor die boomsoort) |
| Afgeleidmodel | Model per boomsoort-domeincombinatie zoals teruggegeven door de functie <code>fit.afgeleid()</code> : list met 2 tibbles. |
| AantalDomHogeRMSE | Standaard worden de 20 domeinen met de hoogste RMSE geselecteerd en voor elk van deze domeinen worden minstens 10 afwijkende metingen geselecteerd. AantalDomHogeRMSE laat toe om dit aantal van 20 domeinen aan te passen. |
| ExtraCurvesRapport | Optie om een lijst met extra boomsoort-domeincombinaties op te geven waardoor de curve getoond moet worden in het validatierapport. Deze moeten opgegeven worden als een dataframe met velden DOMEIN_ID en BMS, met benamingen die overeenkomen met deze in de opgegeven dataset. |
| Bestandsnaam | Een naam voor het validatierapport (.html-bestand) dat gegenereerd wordt, bestaande uit een string die eindigt op .html |
| TypeRapport | Default is "Dynamisch", waarbij de figuren in het .html-bestand kunnen worden aangepast (meetgegevens weergegeven door muis erover te bewegen (inclusief ID als deze in de dataset meegegeven is), items uit legende wegklikken, grafiek inzoomen,...). Een andere optie is "Statisch", waarbij de figuren vast zijn. |
| PathWD | Het path van de working directory, dus het path waarin het validatierapport opgeslagen moet worden (default wordt het in de op dat moment actieve working directory opgeslagen). |

Value

De functie genereert een validatierapport (.html-bestand) in de working directory met informatie en grafieken van de te controleren metingen. De afwijkende metingen zijn in rood aangeduid (zie `?validatierapport` of `vignet` voor meer informatie).

De functie geeft een dataframe terug met de te controleren metingen, met behalve de informatie uit de databank een aantal berekende waarden:

- `H_D_finaal`: een geschatte hoogte voor de omtrekklasse volgens het domeinmodel
- `rsmeD`: de foutenschatting voor het domeinmodel

- `H_VL_finaal`: een geschatte hoogte voor de omtrekklasse volgens het Vlaams model waarvan het domeinmodel afgeleid is
- `rmseVL`: de foutenschatting voor dit Vlaams model
- `HogeRmse`: TRUE als het domeinmodel een hoge RMSE heeft, anders NA

 validatie.basis

Validatie van het basismodel

Description

Functie die alle nodige validaties uitvoert op het opgegeven (volledige) model en een overzicht geeft van de afwijkende metingen en slechte curves (zodat de gebruiker deze kan valideren).

De functie roept meerdere hulpfuncties op:

- `rmse.basis()`
- `afwijkendeMetingen()`
- `afwijkendeCurves()`
- `validatierapport()`

Voorafgaand aan het uitvoeren van deze laatste functie worden eerst de slechtste modellen opgelijst (op basis van RMSE, afwijkende metingen en afwijkende curves).

Usage

```
validatie.basis(
  Basismodel,
  AantalDomHogeRMSE = 20,
  ExtraCurvesRapport = NULL,
  GoedgekeurdeAfwijkendeCurves = NULL,
  Bestandsnaam = "Default",
  TypeRapport = "Dynamisch",
  PathWD = getwd()
)
```

Arguments

- | | |
|--------------------------------|--|
| <code>Basismodel</code> | Model per boomsoort zoals teruggegeven door de functie <code>fit.basis()</code> : tibble met de velden <code>BMS</code> (boomsoort) en <code>Model</code> (lme-object met het gefit mixed model voor die boomsoort). |
| <code>AantalDomHogeRMSE</code> | Standaard worden de 20 domeinen met de hoogste RMSE geselecteerd en voor elk van deze domeinen worden minstens 10 afwijkende metingen geselecteerd. <code>AantalDomHogeRMSE</code> laat toe om dit aantal van 20 domeinen aan te passen. |

| | |
|------------------------------|--|
| ExtraCurvesRapport | Optie om een lijst met extra boomsoort-domeincombinaties op te geven waardoor de curve getoond moet worden in het validatierapport. Deze moeten opgegeven worden als een dataframe met velden DOMEIN_ID en BMS, met benamingen die overeenkomen met deze in de opgegeven dataset. |
| GoedgekeurdeAfwijkendeCurves | Optie om goedgekeurde afwijkende curves niet meer te tonen in het validatierapport zolang er geen extra metingen toegevoegd zijn voor de boomsoort-domeincombinatie (om als gebruiker enkel de te keuren curves in het rapport over te houden). De goedgekeurde en dus te negeren curves moeten opgegeven worden in een dataframe met velden DOMEIN_ID, BMS en nBomenTerugTonen, met nBomenTerugTonen het aantal bomen (nBomenInterval) in de boomsoort-domeincombinatie vanaf wanneer de curve terug getoond moet worden. (In dit geval wordt de curve uiteraard enkel terug getoond als ze nog steeds afwijkend is.) |
| Bestandsnaam | Een naam voor het validatierapport (.html-bestand) dat gegenereerd wordt, bestaande uit een string die eindigt op .html |
| TypeRapport | Default is "Dynamisch", waarbij de figuren in het .html-bestand kunnen worden aangepast (meetgegevens weergegeven door muis erover te bewegen (inclusief ID als deze in de dataset meegegeven is), items uit legende weggklikken, grafiek inzoomen,...). Een andere optie is "Statisch", waarbij de figuren vast zijn. |
| PathWD | Het path van de working directory, dus het path waarin het validatierapport opgeslagen moet worden (default wordt het in de op dat moment actieve working directory opgeslagen). |

Value

De functie genereert een validatierapport (.html-bestand) in de working directory met informatie en grafieken van de te controleren modellen. De afwijkende metingen en curvedelen zijn in rood aangeduid; boven de curve is het probleem ook woordelijk beschreven (zie ?validatierapport of vignet voor meer informatie).

De functie geeft een dataframe terug met de te controleren metingen, met behalve de informatie uit de databank een aantal berekende waarden:

- `H_D_finaal`: een geschatte hoogte voor de omtrekklassse volgens het domeinmodel (dus specifiek voor de boomsoort-domeincombinatie)
- `H_VL_finaal`: een geschatte hoogte voor de omtrek volgens het Vlaams model (dus voor het overkoepelend boomsoort-model)
- `rmseD`: de foutenschatting voor het domeinmodel
- `HogeRmse`: TRUE als het domeinmodel een hoge RMSE heeft, anders NA

Description

Functie die alle nodige validaties uitvoert op het opgegeven lokaal model en een overzicht geeft van de afwijkende metingen en slechte curves (zodat de gebruiker deze kan valideren).

De functie roept meerdere hulpfuncties op:

- `rmse.basis()`
- `afwijkendeMetingen()`
- `afwijkendeCurves()`
- `validatierapport()`

Voorafgaand aan het uitvoeren van deze laatste functie worden eerst de slechtste modellen opgelijst (op basis van RMSE, afwijkende metingen en afwijkende curves).

Usage

```
validatie.lokaal(
  Lokaalmodel,
  Data,
  AantalDomHogeRMSE = 20,
  ExtraCurvesRapport = NULL,
  GoedgekeurdeAfwijkendeCurves = NULL,
  Bestandsnaam = "Default",
  TypeRapport = "Dynamisch",
  PathWD = getwd()
)
```

Arguments

| | |
|------------------------------|---|
| Lokaalmodel | Model per boomsoort-domeincombinatie zoals teruggegeven door de functie <code>fit.lokaal()</code> : tibble met de velden BMS (boomsoort), DOMEIN_ID en Model (lm-object met het gefit lineair model voor die boomsoort-domeincombinatie). |
| Data | Dataset op basis waarvan het opgegeven lokaal model berekend is. |
| AantalDomHogeRMSE | Standaard worden de 20 domeinen met de hoogste RMSE geselecteerd en voor elk van deze domeinen worden minstens 10 afwijkende metingen geselecteerd. AantalDomHogeRMSE laat toe om dit aantal van 20 domeinen aan te passen. |
| ExtraCurvesRapport | Optie om een lijst met extra boomsoort-domeincombinaties op te geven waardoor de curve getoond moet worden in het validatierapport. Deze moeten opgegeven worden als een dataframe met velden DOMEIN_ID en BMS, met benamingen die overeenkomen met deze in de opgegeven dataset. |
| GoedgekeurdeAfwijkendeCurves | Optie om goedgekeurde afwijkende curves niet meer te tonen in het validatierapport zolang er geen extra metingen toegevoegd zijn voor de boomsoort-domeincombinatie (om als gebruiker enkel de te keuren curves in het rapport over te houden). De goedgekeurde en dus te negeren curves moeten opgegeven worden in een dataframe met velden DOMEIN_ID, BMS en nBomenTerugTonen, met nBomenTerugTonen |

| | |
|--------------|---|
| | het aantal bomen (nBomenInterval) in de boomsoort-domeincombinatie vanaf wanneer de curve terug getoond moet worden. (In dit geval wordt de curve uiteraard enkel terug getoond als ze nog steeds afwijkend is.) |
| Bestandsnaam | Een naam voor het validatierapport (.html-bestand) dat gegenereerd wordt, bestaande uit een string die eindigt op .html |
| TypeRapport | Default is "Dynamisch", waarbij de figuren in het .html-bestand kunnen worden aangepast (meetgegevens weergegeven door muis erover te bewegen (inclusief ID als deze in de dataset meegegeven is), items uit legende wegklikken, grafiek inzoomen,...). Een andere optie is "Statisch", waarbij de figuren vast zijn. |
| PathWD | Het path van de working directory, dus het path waarin het validatierapport opgeslagen moet worden (default wordt het in de op dat moment actieve working directory opgeslagen). |

Value

De functie genereert een validatierapport (.html-bestand) in de working directory met informatie en grafieken van de te controleren modellen. De afwijkende metingen en curvedelen zijn in rood aangeduid; boven de curve is het probleem ook woordelijk beschreven (zie ?validatierapport of vignet voor meer informatie).

De functie geeft een dataframe terug met de te controleren metingen, met behalve de informatie uit de databank een aantal berekende waarden:

- H_D_finaal: een geschatte hoogte voor de omtrekklasser volgens het domeinmodel
- rmseD: de foutenschatting voor het domeinmodel
- HogeRmse: TRUE als het domeinmodel een hoge RMSE heeft, anders NA

validatie.uitbreiding *Validatie van een uitbreiding van een model naar hogere omtrekklassen*

Description

Functie die toelaat om de opgegeven basismodellen of lokale modellen te valideren ter hoogte van omtrekklassen boven het bruikbaar interval ingeval hiervoor voldoende gegevens beschikbaar zijn. De functie berekent de waarden `diff_min`, `diff_mediaan` en `diff_max` en geeft een overzichtsrappport met de 20 slechtste niet eerder afgekeurde curves (= curves met hoogste absolute waarde voor `diff_mediaan`), waarvan het de bedoeling is dat de gebruiker ze valideert.

Usage

```
validatie.uitbreiding(
  Model,
  Dataset,
  AantalDomValidatie = 20,
  GoedgekeurdeUitbreidingen = NULL,
  Bestandsnaam = "Default",
```

```

    TypeRapport = "Dynamisch",
    PathWD = getwd()
)

```

Arguments

| | |
|---------------------------|---|
| Model | Basismodel per boomsoort of lokaal model per boomsoort-domeincombinatie zoals teruggegeven door de functies <code>fit.basis()</code> of <code>fit.lokaal()</code> |
| Dataset | Dataset zoals teruggegeven door functie <code>initiatie()</code> , op basis waarvan het opgegeven model (basismodel of lokaal model) berekend is. |
| AantalDomValidatie | Standaard worden de 20 domeinen met de hoogste gemiddelde afwijking (<code>DiffMediaan</code>) geselecteerd om te valideren. (Hierbij worden <code>GoedgekeurdeUitbreidingen</code> niet meer getoond.) <code>AantalDomValidatie</code> laat toe om dit aantal van 20 domeinen aan te passen. |
| GoedgekeurdeUitbreidingen | Optie om goedgekeurde uitbreidingen niet meer te tonen in het validatierapport zolang er geen extra metingen toegevoegd zijn voor de boomsoort-domeincombinatie (om als gebruiker enkel de te keuren curves in het rapport over te houden). De goedgekeurde en dus te negeren curves moeten opgegeven worden in een dataframe met velden <code>DOMEIN_ID</code> , <code>BMS</code> en <code>nBomenTerugTonen</code> , met <code>nBomenTerugTonen</code> het aantal bomen boven het bruikbaar interval (<code>nExtra</code>) in de boomsoort-domeincombinatie vanaf wanneer de curve terug getoond moet worden. (In dit geval wordt de curve uiteraard enkel terug getoond als ze bij de 20 slechtste curves hoort, met hoogste <code>diff_mediaan</code> .) |
| Bestandsnaam | Een naam voor het validatierapport (<code>.html</code> -bestand) dat gegenereerd wordt, bestaande uit een string die eindigt op <code>.html</code> |
| TypeRapport | Default is "Dynamisch", waarbij de figuren in het <code>.html</code> -bestand kunnen worden aangepast (meetgegevens weergegeven door muis erover te bewegen (inclusief ID als deze in de dataset meegegeven is), items uit legende wegklikken, grafiek inzoomen,...). Een andere optie is "Statisch", waarbij de figuren vast zijn. |
| PathWD | Het path van de working directory, dus het path waarin het validatierapport opgeslagen moet worden (default wordt het in de op dat moment actieve working directory opgeslagen). |

Value

De functie genereert een validatierapport (`.html`-bestand) in de working directory met informatie en grafieken van de te controleren modellen. De weergegeven grafieken zijn gerangschikt van slechtste (= mediaan heeft grootste afwijking) naar minder slecht. Idee is om in dit rapport het blauwe deel van de grafiek te valideren.

De functie geeft een dataframe terug met per boomsoort-domeincombinatie de omtrekklasse tot waar het model uitgebreid kan worden (met velden `BMS`, `DOMEIN_ID` en `MaxOmtrek`). Deze tabel kan als argument `Uitbreidingen` in functie `outputIVANHO()` toegevoegd worden om deze uitbreidingen mee te nemen voor het eindresultaat. Belangrijk is hierbij dat afgekeurde curves eerst verwijderd zijn door bij `initiatie()` via `Uitzonderingen` de variabele `min_uitbreiden_model` op te trekken en daarna `validatie.uitbreiding()` opnieuw gerund is met de aangepaste dataset.

| | |
|------------------|--|
| validatierapport | <i>Stelt de slechtste curves en afwijkende metingen grafisch voor in een rapport</i> |
|------------------|--|

Description

De functie genereert een validatierapport (.html-bestand) in de working directory (of opgegeven directory) met informatie en grafieken van de te controleren modellen. De afwijkende metingen en curvedelen zijn in rood aangeduid; boven de curve is het probleem ook woordelijk beschreven.

Deze functie rendert het bestand `Validatierapport.Rmd`, dat afhankelijk van de opgegeven variabele `TypeRapport` voor elke boomsoort-domeincombinatie als child `DomeincurveDynamisch.Rmd` of `Domeincurve.Rmd` toevoegt.

Usage

```
validatierapport(
  SlechtsteModellen,
  AfwijkendeMetingen,
  Dataset,
  Bestandsnaam = "Validatie.html",
  TypeRapport = c("Dynamisch", "Statisch"),
  Uitbreidingsrapport = FALSE,
  verbose = TRUE,
  PathWD = getwd()
)
```

Arguments

| | |
|---------------------------------|---|
| <code>SlechtsteModellen</code> | Lijst van de slechtste modellen, dit zijn modellen met hoge RMSE, afwijkende vorm (op basis van extremen en buigpunten) en/of modellen met afwijkende metingen. Deze dataframe moet volgende velden bevatten: BMS (boomsoort), DOMEIN_ID en Reden (= tekstuele opsomming van afwijkingen, om weer te geven boven grafiek). |
| <code>AfwijkendeMetingen</code> | Lijst met afwijkende metingen (dataframe zoals gegenereerd door de functie <code>afwijkendeMetingen()</code>). |
| <code>Dataset</code> | Dataset met gemeten waarden en geschatte waarde voor domeinmodel en Vlaams model (inclusief RMSE) |
| <code>Bestandsnaam</code> | Een naam voor het validatierapport (.html-bestand) dat gegenereerd wordt, bestaande uit een string die eindigt op <code>.html</code> |
| <code>TypeRapport</code> | Default is "Dynamisch", waarbij de figuren in het <code>.html</code> -bestand kunnen worden aangepast (meetgegevens weergegeven door muis erover te bewegen (inclusief ID als deze in de dataset meegegeven is), items uit legende wegklikken, grafiek inzoomen,...). Een andere optie is "Statisch", waarbij de figuren vast zijn. |

Uitbreidingsrapport

| | |
|----------------------|---|
| | <p>Gaat het over een validatierapport voor <code>validatie.uitbreiding()</code>? Default is <code>FALSE</code>, wat betekent dat het validatierapport de standaard layout heeft voor de validatierapporten, met afwijkingen in rood en standaard informatie die getoond wordt. In het validatierapport voor <code>validatie.uitbreiding()</code> is de uitbreiding die gevalideerd moet worden, in blauw weergegeven. Ook wijkt de weergegeven informatie iets af van de standaard.</p> |
| <code>verbose</code> | <p>Dit geeft de toestand van het systeem aan en zorgt ervoor dat boodschappen niet onnodig gegeven worden. (Defaultwaarde behouden.)</p> |
| <code>PathWD</code> | <p>Het path van de working directory, dus het path waarin het validatierapport opgeslagen moet worden (default wordt het in de op dat moment actieve working directory opgeslagen).</p> |

Value

De functie genereert in de working directory (of opgegeven directory) een rapport (.html) met de te controleren modellen. Hierin wordt per model (boomsoort-domeincombinatie) de volgende algemene informatie vermeld: boomsoort, domein (en ID), aantal metingen, RMSE, bruikbaar interval en de mogelijke problemen die bij het model optreden.

Daaronder wordt telkens grafisch volgende info weergegeven:

- een puntenwolk die de metingen voorstelt (geen individuele metingen, maar een jitter)
- curve van het Vlaams model (als beschikbaar, dus niet voor het lokaal model)
- curve van het domeinmodel
- grenzen van het bruikbaar interval (curves eindigen bij de klassenmiddens die overeenkomen met deze grenzen)
- afwijkende metingen: in rood (andere metingen in zwart)
- afwijkende deel van een curve in rood (rest van curve in zwart)

Bij de keuze voor een uitbreidingsrapport (argument `Uitbreidingsrapport = TRUE`) wijkt het rapport af:

- algemene informatie: de RMSE is vervangen door de variabelen `DiffMediaan`, `DiffMin` en `DiffMax`
- kleur van grafiek: geen afwijkingen in rood, wel is de uitbreiding weergegeven in blauw (zowel jitter van metingen als curve)

Index

afwijkendeCurves, [2](#)
afwijkendeMetingen, [3](#)

curvekarakteristieken, [4](#)

dhcurvesrapport, [5](#)

fit.afgeleid, [6](#)
fit.basis, [7](#)
fit.lokaal, [7](#)

hoogteschatting.afgeleid, [8](#)
hoogteschatting.basis, [9](#)

initiatie, [11](#)

modelparameters, [13](#)

outputIVANHO, [13](#)

resultaat, [15](#)
rmse.basis, [16](#)
rmse.verschuiving, [18](#)

testdata1domein, [19](#)
testdataset, [20](#)

validatie.afgeleid, [20](#)
validatie.basis, [22](#)
validatie.lokaal, [23](#)
validatie.uitbreiding, [25](#)
validatierapport, [27](#)