Package 'carbondate'

July 22, 2025

Title Calibration and Summarisation of Radiocarbon Dates

Version 1.1.0

Description Performs Bayesian non-parametric calibration of multiple related radiocarbon determinations, and summarises the calendar age information to plot their joint calendar age density (see Heaton (2022) <doi:10.1111/rssc.12599>).
Also models the occurrence of radiocarbon samples as a variable-rate (inhomogeneous) Poisson process, plotting the posterior estimate for the occurrence rate of the samples over calendar time, and providing information about potential change points.

License GPL (>= 3)

Encoding UTF-8

Suggests knitr, rmarkdown, testthat (>= 3.0.0)

Config/testthat/edition 3

Depends R (>= 3.5.0)

LazyData true

Imports graphics, grDevices, stats, utils

URL <https://github.com/TJHeaton/carbondate>,
<https://tjheaton.github.io/carbondate/>

BugReports <https://github.com/TJHeaton/carbondate/issues>

VignetteBuilder knitr

LinkingTo cpp11

RoxygenNote 7.3.2

NeedsCompilation yes

Author Timothy J Heaton [aut, cre, cph] (ORCID:
<https://orcid.org/0000-0002-9994-142X>),
Sara Al-assam [aut, cph] (ORCID:
<https://orcid.org/0000-0001-7280-6802>)

Maintainer Timothy J Heaton <T.Heaton@leeds.ac.uk>

Repository CRAN

Date/Publication 2025-04-08 02:20:02 UTC

Contents

alces	*Example real-life data - Alces in Yukon and Alaska*

Description

58 radiocarbon determinations collated by Dale Guthrie, R. related to Alces (moose) in Yukon and Alaska. Samples are restricted to those between 25,000–6000 ^{14}C yrs BP.

Reference:
Dale Guthrie, R. New carbon dates link climatic change with human colonization and Pleistocene extinctions. *Nature* **441**, 207–209 (2006). https://doi.org/10.1038/nature04604

Usage

    alces

Format

`alces`:

A data frame with 58 rows and 7 columns:

lab_code The sample code for the ^{14}C laboratory

site_code The site/museum code

location The location of the sample

c14_age The observed ^{14}C ages of the samples (in ^{14}C yr BP)

c14_sig The uncertainty in the observed ^{14}C ages reported by the radiocarbon laboratory

f14c The observed F^{14}C concentrations

f14c_sig The uncertainty in the observed F^{14}C concentrations reported by the radiocarbon laboratory

Source

https://doi.org/10.1038/nature04604

armit	*Example real-life data - Population Decline in Iron Age Ireland*

Description

2021 radiocarbon determinations collated by Armit et al. from archaeological groups operating in Ireland, to investigate whether a wetter environment around 2700 cal yr BP led to a population collapse.

Reference:
Armit, I., Swindles, G.T., Becker, K., Plunkett, G., Blaauw, M., 2014. Rapid climate change did not cause population collapse at the end of the European Bronze Age. *Proceedings of the National Academy of Sciences* **111**, 17045–17049.

Usage

```
armit
```

Format

`armit:`

A data frame with 2021 rows and 4 columns:

c14_age The observed ^{14}C ages of the samples (in ^{14}C yr BP)

c14_sig The uncertainty in the observed ^{14}C ages reported by the radiocarbon laboratory

f14c The observed F^{14}C concentrations

f14c_sig The uncertainty in the observed F^{14}C concentrations reported by the radiocarbon laboratory

Source

http://doi.org/10.1073/pnas.1408028111

---

bison   *Example real-life data - Bison in Yukon and Alaska*

Description

64 radiocarbon determinations collated by Dale Guthrie, R. related to Bison in Yukon and Alaska. Samples are restricted to those between 25,000–6000 ^{14}C yrs BP.

Reference:

Dale Guthrie, R. New carbon dates link climatic change with human colonization and Pleistocene extinctions. *Nature* **441**, 207–209 (2006). https://doi.org/10.1038/nature04604

Usage

```
bison
```

Format

`bison:`

A data frame with 64 rows and 7 columns:

lab_code The sample code for the ^{14}C laboratory

site_code The site/museum code

location The location of the sample

c14_age The observed ^{14}C ages of the samples (in ^{14}C yr BP)

c14_sig The uncertainty in the observed ^{14}C ages reported by the radiocarbon laboratory

f14c The observed F^{14}C concentrations

f14c_sig The uncertainty in the observed F^{14}C concentrations reported by the radiocarbon laboratory

Source

https://doi.org/10.1038/nature04604

---

buchanan   *Example real-life data - Palaeo-Indian demography*

Description

628 radiocarbon determinations collated by Buchanan et al. representing the ages of distinct archaeological sites found across Canada and North America during the time of the palaeoindians.

Reference:
Buchanan, B., Collard, M., Edinborough, K., 2008. Paleoindian demography and the extraterrestrial impact hypothesis. *Proceedings of the National Academy of Sciences* **105**, 11651–11654.

Usage

buchanan

Format

buchanan**:**

A data frame with 628 rows and 4 columns:

c14_age The observed ^{14}C ages of the samples (in ^{14}C yr BP)

c14_sig The uncertainty in the observed ^{14}C ages reported by the radiocarbon laboratory

f14c The observed F^{14}C concentrations

f14c_sig The uncertainty in the observed F^{14}C concentrations reported by the radiocarbon laboratory

Source

http://doi.org/10.1073/pnas.0803762105

CalibrateSingleDetermination
Calibrate a Single Radiocarbon Determination

Description

Uses the supplied calibration curve to calibrate a single radiocarbon determination and uncertainty (expressed either in terms of radiocarbon age, or as an F^{14}C concentration) and obtain its calendar age probability density estimate.

Usage

```
CalibrateSingleDetermination(
  rc_determination,
  rc_sigma,
  calibration_curve,
  F14C_inputs = FALSE,
  resolution = 1,
  plot_output = FALSE,
  plot_cal_age_scale = "BP",
  interval_width = "2sigma",
  bespoke_probability = NA,
  denscale = 3,
  plot_pretty = TRUE
)
```

Arguments

rc_determination

A single observed radiocarbon determination provided either as the radiocarbon age (in ^{14}C yr BP) or the F^{14}C concentration.

rc_sigma   The corresponding measurement uncertainty of the radiocarbon determination (must be in the same units as above, i.e., reported as ^{14}C age or F^{14}C)

calibration_curve

A dataframe which must contain one column calendar_age_BP, and also columns c14_age and c14_sig or f14c and f14c_sig (or both sets). This format matches the curves supplied with this package, e.g., intcal20, intcal13, which contain all 5 columns.

F14C_inputs   TRUE if the provided rc_determination is an F^{14}C concentration and FALSE if it is a radiocarbon age. Defaults to FALSE.

resolution   The distance between the calendar ages at which to calculate the calendar age probability. Default is 1.

plot_output   TRUE if you wish to plot the determination, the calibration curve, and the posterior calibrated age estimate on the same plot. Defaults to FALSE

plot_cal_age_scale

Only for usage when plot_output = TRUE. The calendar scale to use for the x-axis. Allowed values are "BP", "AD" and "BC". The default is "BP", corresponding to plotting in cal yr BP.

interval_width   Only for usage when plot_output = TRUE. The confidence intervals to show for the calibration curve and for the highest posterior density ranges. Choose from one of "1sigma" (68.3%), "2sigma" (95.4%) and "bespoke". Default is "2sigma".

bespoke_probability

The probability to use for the confidence interval if "bespoke" is chosen above. E.g. if 0.95 is chosen, then the 95% confidence interval is calculated. Ignored if "bespoke" is not chosen.

denscale   Whether to scale the vertical range of the calendar age density plot relative to the calibration curve plot (optional). Default is 3 which means that the maximum calendar age density will be at 1/3 of the height of the plot.

plot_pretty   logical, defaulting to TRUE. If set TRUE then will select pretty plotting margins (that create sufficient space for axis titles and rotates y-axis labels). If FALSE will implement current user values.

Value

A data frame with one column calendar_age_BP containing the calendar ages, and the other column probability containing the probability at that calendar age.

Examples

```
# Calibration of a single determination expressed as 14C age BP
calib <- CalibrateSingleDetermination(860, 35, intcal20)
plot(calib, type = "l", xlim = c(1000, 600))

# Incorporating an automated plot to visualise the calibration
CalibrateSingleDetermination(860, 35, intcal20, plot_output = TRUE)

# Calibration of a single (old) determination expressed as 14C age BP
calib <- CalibrateSingleDetermination(31020, 100, intcal20)
plot(calib, type = "l", xlim = c(36500, 34500))

# Calibration of a single (old) determination expressed as F14C concentration
calib <- CalibrateSingleDetermination(
    0.02103493, 0.0002618564, intcal20, F14C_inputs = TRUE)
plot(calib, type = "l", xlim = c(36500, 34500))

# Calibration of a single determination expressed as 14C age BP
# against SHCal20 (and creating an automated plot)
CalibrateSingleDetermination(1413, 25, shcal20, plot_output = TRUE)

# Implementing a bespoke confidence interval level and plot in AD
CalibrateSingleDetermination(
    1413,
    25,
    shcal20,
    plot_output = TRUE,
    plot_cal_age_scale = "AD",
    interval_width = "bespoke",
    bespoke_probability = 0.8)

# Changing denscale (so the calendar age density takes up less space)
CalibrateSingleDetermination(
    1413,
    25,
    shcal20,
    plot_output = TRUE,
    interval_width = "bespoke",
    bespoke_probability = 0.8,
```

```
    denscale = 5)
```

---

cervus   *Example real-life data - Cervus in Yukon and Alaska*

Description

63 radiocarbon determinations collated by Dale Guthrie, R. related to Cervus (wapiti) in Yukon and Alaska. Samples are restricted to those between 25,000–6000 ^{14}C yrs BP.

Reference:

Dale Guthrie, R. New carbon dates link climatic change with human colonization and Pleistocene extinctions. *Nature* **441**, 207–209 (2006). https://doi.org/10.1038/nature04604

Usage

```
cervus
```

Format

cervus:

A data frame with 63 rows and 7 columns:

lab_code The sample code for the ^{14}C laboratory

site_code The site/museum code

location The location of the sample

c14_age The observed ^{14}C ages of the samples (in ^{14}C yr BP)

c14_sig The uncertainty in the observed ^{14}C ages reported by the radiocarbon laboratory

f14c The observed $F^{14}C$ concentrations

f14c_sig The uncertainty in the observed $F^{14}C$ concentrations reported by the radiocarbon laboratory

Source

https://doi.org/10.1038/nature04604

---

equus   *Example real-life data - Equus in Yukon and Alaska*

Description

84 radiocarbon determinations collated by Dale Guthrie, R. related to Equus (horse) in Yukon and Alaska. Samples are restricted to those between 25,000–6000 ^{14}C yrs BP.

Reference:
Dale Guthrie, R. New carbon dates link climatic change with human colonization and Pleistocene extinctions. *Nature* **441**, 207–209 (2006). https://doi.org/10.1038/nature04604

Usage

```
equus
```

Format

`equus`:

A data frame with 84 rows and 7 columns:

lab_code The sample code for the ^{14}C laboratory

site_code The site/museum code

location The location of the sample

c14_age The observed ^{14}C ages of the samples (in ^{14}C yr BP)

c14_sig The uncertainty in the observed ^{14}C ages reported by the radiocarbon laboratory

f14c The observed F^{14}C concentrations

f14c_sig The uncertainty in the observed F^{14}C concentrations reported by the radiocarbon laboratory

Source

https://doi.org/10.1038/nature04604

FindPosteriorMeanRate *Find Posterior Mean Rate of Sample Occurrence for Poisson Process Model*

Description

Given output from the Poisson process fitting function [PPcalibrate](#) calculate the posterior mean rate of sample occurrence (i.e., the underlying Poisson process rate $\lambda(t)$) together with specified probability intervals, on a given calendar age grid (provided in cal yr BP).

An option is also provided to calculate the posterior mean rate *conditional* upon the number of internal changepoints within the period under study (if this is specified as an input to the function).

Note: If you want to calculate and plot the result, use [PlotPosteriorMeanRate](#) instead.

For more information read the vignette:
`vignette("Poisson-process-modelling", package = "carbondate")`

Usage

```
FindPosteriorMeanRate(
  output_data,
  calendar_age_sequence,
  n_posterior_samples = 5000,
  n_changes = NULL,
  interval_width = "2sigma",
  bespoke_probability = NA,
  n_burn = NA,
  n_end = NA
)
```

Arguments

output_data	The return value from the updating function [PPcalibrate](#). Optionally, the output data can have an extra list item named label which is used to set the label on the plot legend.
calendar_age_sequence	
	A vector containing the calendar age grid (in cal yr BP) on which to calculate the posterior mean rate.
n_posterior_samples	
	Number of samples it will draw, after having removed n_burn, from the (thinned) MCMC realisations stored in output_data to estimate the rate $\lambda(t)$. These samples may be repeats if the number of, post burn-in, realisations is less than n_posterior_samples. If not given, 5000 is used.
n_changes	(Optional) If wish to condition calculation of the posterior mean on the number of internal changepoints. In this function, if specified, n_changes must be a single integer.
interval_width	The confidence intervals to show for both the calibration curve and the predictive density. Choose from one of "1sigma" (68.3%), "2sigma" (95.4%) and "bespoke". Default is "2sigma".
bespoke_probability	
	The probability to use for the confidence interval if "bespoke" is chosen above. E.g., if 0.95 is chosen, then the 95% confidence interval is calculated. Ignored if "bespoke" is not chosen.

n_burn	The number of MCMC iterations that should be discarded as burn-in (i.e., considered to be occurring before the MCMC has converged). This relates to the number of iterations (n_iter) when running the original update functions (not the thinned output_data). Any MCMC iterations before this are not used in the calculations. If not given, the first half of the MCMC chain is discarded. Note: The maximum value that the function will allow is n_iter - 100 * n_thin (where n_iter and n_thin are the arguments that were given to [PPcalibrate](#)) which would leave only 100 of the (thinned) values in output_data.
n_end	The last iteration in the original MCMC chain to use in the calculations. Assumed to be the total number of iterations performed, i.e. n_iter, if not given.

Value

A list, each item containing a data frame of the calendar_age_BP, the rate_mean and the confidence intervals for the rate - rate_ci_lower and rate_ci_upper.

See Also

[PlotPosteriorMeanRate](#)

Examples

```
# NOTE: All these examples are shown with a small n_iter and n_posterior_samples
# to speed up execution.
# Try n_iter and n_posterior_samples as the function defaults.

pp_output <- PPcalibrate(
    pp_uniform_phase$c14_age,
    pp_uniform_phase$c14_sig,
    intcal20,
    n_iter = 1000,
    show_progress = FALSE)

# Default plot with 2 sigma interval
FindPosteriorMeanRate(pp_output, seq(450, 640, length=10), n_posterior_samples = 100)
```

FindPredictiveCalendarAgeDensity

Find Predictive Estimate of Shared Calendar Age Density from Bayesian Non-Parametric DPMM Output

---

Description

Given output from one of the Bayesian non-parametric summarisation functions (either [PolyaUrn-BivarDirichlet](#) or [WalkerBivarDirichlet](#)) calculate the predictive (summarised/shared) calendar age density and probability intervals on a given calendar age grid (provided in cal yr BP).

Note: If you want to calculate and plot the result, use [PlotPredictiveCalendarAgeDensity](#) instead.

Usage

```
FindPredictiveCalendarAgeDensity(
  output_data,
  calendar_age_sequence,
  n_posterior_samples = 5000,
  interval_width = "2sigma",
  bespoke_probability = NA,
  n_burn = NA,
  n_end = NA
)
```

Arguments

output_data   The return value from one of the Bayesian non-parametric DPMM functions,
                    e.g. PolyaUrnBivarDirichlet or WalkerBivarDirichlet, or a list, each item con-
                    taining one of these return values. Optionally, the output data can have an extra
                    list item named label which is used to set the label on the plot legend.

calendar_age_sequence

                    A vector containing the calendar age grid (in cal yr BP) on which to calculate
                    the predictive (summarised/shared) density.

n_posterior_samples

                    Number of samples it will draw, after having removed n_burn, from the (thinned)
                    realisations stored in the DPMM outputs to estimate the predictive calendar age
                    density. These samples may be repeats if the number of, post burn-in, realisa-
                    tions is less than n_posterior_samples. If not given, 5000 is used.

interval_width   The confidence intervals to show for both the calibration curve and the predic-
                    tive density. Choose from one of "1sigma" (68.3%), "2sigma" (95.4%) and
                    "bespoke". Default is "2sigma".

bespoke_probability

                    The probability to use for the confidence interval if "bespoke" is chosen above.
                    E.g., if 0.95 is chosen, then the 95% confidence interval is calculated. Ignored
                    if "bespoke" is not chosen.

n_burn   The number of MCMC iterations that should be discarded as burn-in (i.e., con-
                    sidered to be occurring before the MCMC has converged). This relates to the
                    number of iterations (n_iter) when running the original update functions (not
                    the thinned output_data). Any MCMC iterations before this are not used
                    in the calculations. If not given, the first half of the MCMC chain is dis-
                    carded. Note: The maximum value that the function will allow is n_iter –
                    100 * n_thin (where n_iter and n_thin are the arguments given to PolyaU-
                    rnBivarDirichlet or WalkerBivarDirichlet) which would leave only 100 of the
                    (thinned) values in output_data.

n_end   The last iteration in the original MCMC chain to use in the calculations. As-
                    sumed to be the total number of iterations performed, i.e. n_iter, if not given.

Value

A data frame of the calendar_age_BP, the density_mean and the confidence intervals for the
density density_ci_lower and density_ci_upper.

See Also

[PlotPredictiveCalendarAgeDensity](#)

Examples

```
# NOTE: All these examples are shown with a small n_iter and n_posterior_samples
# to speed up execution.
# Try n_iter and n_posterior_samples as the function defaults.

# First generate output data
polya_urn_output <- PolyaUrnBivarDirichlet(
    two_normals$c14_age,
    two_normals$c14_sig,
    intcal20,
    n_iter = 100,
    show_progress = FALSE)

# Find results for example output, 2-sigma confidence interval (default)
FindPredictiveCalendarAgeDensity(
    polya_urn_output, seq(3600, 4700, length=12), n_posterior_samples = 500)
```

FindSummedProbabilityDistribution

Find the summed probability distribution (SPD) for a set of radiocarbon observations

---

Description

Takes a set of radiocarbon determinations and uncertainties, independently calibrates each one, and then averages the resultant calendar age estimates to give the SPD estimate.

Important: This function should not be used for inference as SPDs are not statistically rigorous. Instead use either of:

- the Bayesian non-parametric summarisation approaches [PolyaUrnBivarDirichlet](#) or [Walker-BivarDirichlet](#);
- or the Poisson process rate approach [PPcalibrate](#)

The SPD function provided here is only intended for comparison. We provide a inbuilt plotting option to show the SPD alongside the determinations and the calibration curve.

Usage

```
FindSummedProbabilityDistribution(
  calendar_age_range_BP,
  rc_determinations,
```

```
  rc_sigmas,
  calibration_curve,
  F14C_inputs = FALSE,
  resolution = 1,
  plot_output = FALSE,
  plot_cal_age_scale = "BP",
  interval_width = "2sigma",
  bespoke_probability = NA,
  denscale = 3,
  plot_pretty = TRUE
)
```

Arguments

calendar_age_range_BP

> A vector of length 2 with the start and end calendar age BP to calculate the SPD over. These two values must be in order with start value less than end value, e.g., (600, 1700) cal yr BP. The code will extend this range (by 400 cal yrs each side) for conservativeness

rc_determinations

> A vector of observed radiocarbon determinations. Can be provided either as ^{14}C ages (in ^{14}C yr BP) or as F^{14}C concentrations.

rc_sigmas   A vector of the (1-sigma) measurement uncertainties for the radiocarbon determinations. Must be the same length as rc_determinations and given in the same units.

calibration_curve

> A dataframe which must contain one column calendar_age_BP, and also columns c14_age and c14_sig or f14c and f14c_sig (or both sets). This format matches the curves supplied with this package, e.g., [intcal20](), [intcal13](), which contain all 5 columns.

F14C_inputs   TRUE if the provided rc_determinations are F^{14}C concentrations and FALSE if they are radiocarbon ages. Defaults to FALSE.

resolution   The distance between the calendar ages at which to calculate the calendar age probability. Default is 1.

plot_output   TRUE if you wish to plot the determinations, the calibration curve, and the SPD on the same plot. Defaults to FALSE

plot_cal_age_scale

> Only for usage when plot_output = TRUE. The calendar scale to use for the x-axis. Allowed values are "BP", "AD" and "BC". The default is "BP", corresponding to plotting in cal yr BP.

interval_width   Only for usage when plot_output = TRUE. The confidence intervals to show for the calibration curve. Choose from one of "1sigma" (68.3%), "2sigma" (95.4%) and "bespoke". Default is "2sigma".

bespoke_probability

> The probability to use for the confidence interval if "bespoke" is chosen above. E.g. if 0.95 is chosen, then the 95% confidence interval is calculated. Ignored if "bespoke" is not chosen.

denscale	Whether to scale the vertical range of the calendar age density plot relative to the calibration curve plot (optional). Default is 3 which means that the maximum SPD will be at 1/3 of the height of the plot.
plot_pretty	logical, defaulting to TRUE. If set TRUE then will select pretty plotting margins (that create sufficient space for axis titles and rotates y-axis labels). If FALSE will implement current user values.

Value

A data frame with one column `calendar_age_BP` containing the calendar ages, and the other column `probability` containing the probability at that calendar age

See Also

[PolyaUrnBivarDirichlet,](#) [WalkerBivarDirichlet](#) for rigorous non-parametric Bayesian alternatives; and [PPcalibrate](#) for a rigorous variable-rate Poisson process alternative.

Examples

```
# An example using 14C age BP and the IntCal 20 curve
SPD <- FindSummedProbabilityDistribution(
   calendar_age_range_BP=c(400, 1700),
   rc_determinations=c(602, 805, 1554),
   rc_sigmas=c(35, 34, 45),
   calibration_curve=intcal20)
plot(SPD, type = "l",
   xlim = rev(range(SPD$calendar_age_BP)),
   xlab = "Calendar Age (cal yr BP)")

# Using the inbuilt plotting features
SPD <- FindSummedProbabilityDistribution(
   calendar_age_range_BP=c(400, 1700),
   rc_determinations=c(602, 805, 1554),
   rc_sigmas=c(35, 34, 45),
   calibration_curve=intcal20,
   plot_output = TRUE,
   interval_width = "bespoke",
   bespoke_probability = 0.8)

# An different example using F14C concentrations and the IntCal 13 curve
SPD <- FindSummedProbabilityDistribution(
   calendar_age_range_BP=c(400, 2100),
   rc_determinations=c(0.8, 0.85, 0.9),
   rc_sigmas=c(0.01, 0.015, 0.012),
   F14C_inputs=TRUE,
   calibration_curve=intcal13)
plot(SPD, type = "l",
   xlim = rev(range(SPD$calendar_age_BP)),
   xlab = "Calendar Age (cal yr BP)")
```

GenerateOxcalCode   *Outputs code suitable for running in OxCal from a series of radiocarbon determinations*

Description

Outputs code suitable for running in OxCal from a series of radiocarbon determinations that can be given as either ^{14}C age or F^{14}C.

Usage

```
GenerateOxcalCode(
  model_name,
  rc_determinations,
  rc_sigmas,
  rc_names = NULL,
  F14C_inputs = FALSE,
  outfile_path = NULL
)
```

Arguments

model_name	The name given to the model in the OxCal code.
rc_determinations	
	A vector of observed radiocarbon determinations. Can be provided either as ^{14}C ages (in ^{14}C yr BP) or as F^{14}C concentrations.
rc_sigmas	A vector of the (1-sigma) measurement uncertainties for the radiocarbon determinations. Must be the same length as rc_determinations and given in the same units.
rc_names	Optional. The name of each data point - if given it must be the same length of rc_determinations.
F14C_inputs	TRUE if the provided rc_determinations are F^{14}C concentrations and FALSE if they are radiocarbon ages. Defaults to FALSE.
outfile_path	Optional. If given the OxCal code will be output to the file at the path given, otherwise it will be output to the terminal.

Value

None

Examples

```
# Generate names automatically and outputs to the screen for 14C ages
GenerateOxcalCode("My_data", c(1123, 1128, 1135), c(32, 24, 25))

# Provide name automatically and outputs to the screen for F14C concentrations
```

```
GenerateOxcalCode(
  "My_data",
  c(0.832, 0.850, 0.846),
  c(0.004, 0.003, 0.009),
  c("P-1", "P-2", "P-3"),
  F14C_inputs=TRUE)
```

---

human   *Example real-life data - Humans in Yukon and Alaska*

Description

46 radiocarbon determinations collated by Dale Guthrie, R. related to archaeological sites (i.e., evidence of human existence) in Alaska. Samples are restricted to those between 25,000–6000 ^{14}C yrs BP.

Reference:
Dale Guthrie, R. New carbon dates link climatic change with human colonization and Pleistocene extinctions. *Nature* **441**, 207–209 (2006). https://doi.org/10.1038/nature04604

Usage

human

Format

human:

A data frame with 46 rows and 7 columns:

lab_code The sample code for the ^{14}C laboratory

site_code The site/museum code

location The location of the sample

c14_age The observed ^{14}C ages of the samples (in ^{14}C yr BP)

c14_sig The uncertainty in the observed ^{14}C ages reported by the radiocarbon laboratory

f14c The observed F^{14}C concentrations

f14c_sig The uncertainty in the observed F^{14}C concentrations reported by the radiocarbon laboratory

Source

https://doi.org/10.1038/nature04604

| intcal04 | *IntCal04 calibration curve* |

Description

The IntCal04 Northern Hemisphere radiocarbon age calibration curve on a calendar grid spanning from 26,000–0 cal yr BP (Before Present, 0 cal yr BP corresponds to 1950 CE).

Note: This dataset provides ^{14}C ages and F^{14}C values on a calendar age grid. This is different from the ^{14}C ages and Δ^{14}C values provided in oxcal .14c files.

Reference:
PJ Reimer, MGL Baillie, E Bard, A Bayliss, JW Beck, C Bertrand, PG Blackwell, CE Buck, G Burr, KB Cutler, PE Damon, RL Edwards, RG Fairbanks, M Friedrich, TP Guilderson, KA Hughen, B Kromer, FG McCormac, S Manning, C Bronk Ramsey, RW Reimer, S Remmele, JR Southon, M Stuiver, S Talamo, FW Taylor, J van der Plicht, and CE Weyhenmeyer. 2004. Intcal04 Terrestrial Radiocarbon Age Calibration, 0–26 Cal Kyr BP. *Radiocarbon* **46**(3):1029-1058 https://doi.org/10.1017/S0033822200032999.

Usage

```
intcal04
```

Format

`intcal04`:

A data frame with 3,301 rows and 5 columns providing the IntCal04 radiocarbon age calibration curve on a calendar grid spanning from 26,000–0 cal yr BP:

calendar_age The calendar age (in cal yr BP)

c14_age The ^{14}C age (in ^{14}C yr BP)

c14_sig The (1-σ) uncertainty in the ^{14}C age

f14c The ^{14}C age expressed as F^{14}C concentration

f14c_sig The (1-σ) uncertainty in the F^{14}C concentration

Source

https://doi.org/10.1017/S0033822200032999

---

intcal09   *IntCal09 calibration curve*

Description

The IntCal09 Northern Hemisphere radiocarbon age calibration curve on a calendar grid spanning from 50,000–0 cal yr BP (Before Present, 0 cal yr BP corresponds to 1950 CE).

Note: This dataset provides ^{14}C ages and F^{14}C values on a calendar age grid. This is different from the ^{14}C ages and Δ^{14}C values provided in oxcal .14c files.

Reference:

PJ Reimer, MGL Baillie, E Bard, A Bayliss, JW Beck, PG Blackwell, C Bronk Ramsey, CE Buck, GS Burr, RL Edwards, M Friedrich, PM Grootes, TP Guilderson, I Hajdas, TJ Heaton, AG Hogg, KA Hughen, KF Kaiser, B Kromer, FG McCormac, SW Manning, RW Reimer, DA Richards, JR Southon, S Talamo, CSM Turney, J van der Plicht, CE Weyhenmeyer. 2009. IntCal09 and Marine09 Radiocarbon Age Calibration Curves, 0–50,000 Years cal BP *Radiocarbon* **51**(4):1111-1150 https://doi.org/10.1017/S0033822200034202.

Usage

```
intcal09
```

Format

intcal09:

A data frame with 3,521 rows and 5 columns providing the IntCal09 radiocarbon age calibration curve on a calendar grid spanning from 50,000–0 cal yr BP:

calendar_age The calendar age (in cal yr BP)

c14_age The ^{14}C age (in ^{14}C yr BP)

c14_sig The (1-σ) uncertainty in the ^{14}C age

f14c The ^{14}C age expressed as F^{14}C concentration

f14c_sig The (1-σ) uncertainty in the F^{14}C concentration

Source

http://doi.org/10.1017/S0033822200034202

---

intcal13   *IntCal13 calibration curve*

Description

The IntCal13 Northern Hemisphere radiocarbon age calibration curve on a calendar grid spanning from 50,000–0 cal yr BP (Before Present, 0 cal yr BP corresponds to 1950 CE).

Note: This dataset provides ^{14}C ages and F^{14}C values on a calendar age grid. This is different from the ^{14}C ages and Δ^{14}C values provided in oxcal .14c files.

Reference:
Reimer PJ, Bard E, Bayliss A, Beck JW, Blackwell PG, Bronk Ramsey C, Buck CE, Cheng H, Edwards RL, Friedrich M, Grootes PM, Guilderson TP, Haflidason H, Hajdas I, Hatt? C, Heaton TJ, Hogg AG, Hughen KA, Kaiser KF, Kromer B, Manning SW, Niu M, Reimer RW, Richards DA, Scott EM, Southon JR, Turney CSM, van der Plicht J. 2013. IntCal13 and Marine13 radiocarbon age calibration curves 0–50000 years calBP. *Radiocarbon* **55**(4) https://doi.org/10.2458/azu_js_rc.55.16947.

Usage

```
intcal13
```

Format

intcal13:

A data frame with 5,141 rows and 5 columns providing the IntCal13 radiocarbon age calibration curve on a calendar grid spanning from 50,000–0 cal yr BP:

calendar_age The calendar age (in cal yr BP)

c14_age The ^{14}C age (in ^{14}C yr BP)

c14_sig The (1-σ) uncertainty in the ^{14}C age

f14c The ^{14}C age expressed as F^{14}C concentration

f14c_sig The (1-σ) uncertainty in the F^{14}C concentration

Source

http://doi.org/10.2458/azu_js_rc.55.16947

intcal20   *IntCal20 calibration curve*

Description

The IntCal20 Northern Hemisphere radiocarbon age calibration curve on a calendar grid spanning from 55,000–0 cal yr BP (Before Present, 0 cal yr BP corresponds to 1950 CE).

Note: This dataset provides ^{14}C ages and F^{14}C values on a calendar age grid. This is different from the ^{14}C ages and Δ^{14}C values provided in oxcal .14c files.

Reference:
Reimer PJ, Austin WEN, Bard E, Bayliss A, Blackwell PG, Bronk Ramsey C, Butzin M, Cheng H, Edwards RL, Friedrich M, Grootes PM, Guilderson TP, Hajdas I, Heaton TJ, Hogg AG, Hughen KA, Kromer B, Manning SW, Muscheler R, Palmer JG, Pearson C, van der Plicht J, Reimer RW, Richards DA, Scott EM, Southon JR, Turney CSM, Wacker L, Adolphi F, Büntgen U, Capano M, Fahrni S, Fogtmann-Schulz A, Friedrich R, Köhler P, Kudsk S, Miyake F, Olsen J, Reinig F, Sakamoto M, Sookdeo A, Talamo S. 2020. The IntCal20 Northern Hemisphere radiocarbon age calibration curve (0-55 cal kBP). *Radiocarbon* **62** https://doi.org/10.1017/RDC.2020.41.

Usage

```
intcal20
```

Format

`intcal20`:

A data frame with 9,501 rows and 5 columns providing the IntCal20 radiocarbon age calibration curve on a calendar grid spanning from 55,000–0 cal yr BP:

calendar_age The calendar age (in cal yr BP)

c14_age The ^{14}C age (in ^{14}C yr BP)

c14_sig The (1-σ) uncertainty in the ^{14}C age

f14c The ^{14}C age expressed as F^{14}C concentration

f14c_sig The (1-σ) uncertainty in the F^{14}C concentration

Source

http://doi.org/10.1017/RDC.2020.41

---

intcal98   *IntCal98 calibration curve*

Description

The IntCal98 Northern Hemisphere radiocarbon age calibration curve on a calendar grid spanning from 24,000–0 cal yr BP (Before Present, 0 cal yr BP corresponds to 1950 CE).

Note: This dataset provides ^{14}C ages and F^{14}C values on a calendar age grid. This is different from the ^{14}C ages and Δ^{14}C values provided in oxcal .14c files.

Reference:

M. Stuiver, P. J. Reimer, E. Bard, J. W. Beck, G. S. Burr, K. A. Hughen, B. Kromer, F. G. McCormac, J. v. d. Plicht and M. Spurk. 1998. INTCAL98 Radiocarbon Age Calibration, 24,000–0 cal BP. *Radiocarbon* **40**(3):1041-1083 https://doi.org/10.1017/S0033822200019123.

Usage

    intcal98

Format

intcal98:

A data frame with 1,538 rows and 5 columns providing the IntCal98 radiocarbon age calibration curve on a calendar grid spanning from 24,000–0 cal yr BP:

calendar_age The calendar age (in cal yr BP)

c14_age The ^{14}C age (in ^{14}C yr BP)

c14_sig The (1-σ) uncertainty in the ^{14}C age

f14c The ^{14}C age expressed as F^{14}C concentration

f14c_sig The (1-σ) uncertainty in the F^{14}C concentration

Source

https://doi.org/10.1017/S0033822200019123

InterpolateCalibrationCurve
                    Interpolate a calibration curve at a set of calendar ages

---

Description

Interpolate a calibration curve at a set of calendar ages

Usage

```
InterpolateCalibrationCurve(
  new_calendar_ages_BP,
  calibration_curve,
  F14C_outputs = NA
)
```

Arguments

new_calendar_ages_BP

A scalar or vector containing calendar ages (in cal yr BP) at which to interpolate the values (both the means and uncertainties) of the given calibration curve. If not provided (and NA is given), will use the range from the minimum calendar age to the maximum calendar age of the original calibration curve spaced by 1.

calibration_curve

A dataframe which must contain one column calendar_age_BP, and also columns c14_age and c14_sig or f14c and f14c_sig (or both sets). This format matches the curves supplied with this package, e.g., intcal20, intcal13, which contain all 5 columns.

F14C_outputs   TRUE if only $F^{14}C$ concentrations are required, FALSE if only the radiocarbon ages (in ^{14}C yrs BP) are required and NA if both are required for the new curve.

Value

A new dataframe with entries for the interpolated c14_age, and c14_sig, f14c and f14c_sig values at the calendar_age_BP values that were given in new_calendar_ages_BP.

Examples

```
# Interpolate intcal20 at a single calendar age. Generates both 14C ages and F14C scales.
InterpolateCalibrationCurve(51020, intcal20)

# Interpolate intcal20 at two calendar ages. Generates F14C estimates only.
InterpolateCalibrationCurve(c(51017, 51021), intcal20, TRUE)

# Interpolate intcal20 at two calendar ages. Generate 14C age estimates (cal yr BP) only.
InterpolateCalibrationCurve(c(51017, 51021), intcal20, FALSE)

# Interpolate intcal20 at every integer calendar age within the range of dates
# (for intcal20 this is 0 to 55000 cal yr BP), and create estimates for both radiocarbon scales.
cal_curve <- InterpolateCalibrationCurve(NA, intcal20)
```

---

kerr   *Example real-life data - Irish Rath*

Description

255 radiocarbon determinations collated by Kerr and McCormick related to the building and use of raths in Ireland in the early-medieval period.

Reference:
Kerr, T., McCormick, F., 2014. Statistics, sunspots and settlement: influences on sum of probability curves. *Journal of Archaeological Science* **41**, 493–501.

Usage

    kerr

Format

    kerr:

A data frame with 255 rows and 4 columns:

c14_age The observed ^{14}C ages of the samples (in ^{14}C yr BP)

c14_sig The uncertainty in the observed ^{14}C ages reported by the radiocarbon laboratory

f14c The observed F^{14}C concentrations

f14c_sig The uncertainty in the observed F^{14}C concentrations reported by the radiocarbon laboratory

Source

http://doi.org/10.1016/j.jas.2013.09.002

---

mammuthus   *Example real-life data - Mammuthus in Yukon and Alaska*

Description

117 radiocarbon determinations collated by Dale Guthrie, R. related to Mammuthus (mammoth) in Yukon and Alaska. Samples are restricted to those between 25,000–6000 ^{14}C yrs BP.

Reference:
Dale Guthrie, R. New carbon dates link climatic change with human colonization and Pleistocene extinctions. *Nature* **441**, 207–209 (2006). https://doi.org/10.1038/nature04604

Usage

    mammuthus

Format

mammuthus**:**

A data frame with 117 rows and 7 columns:

lab_code The sample code for the ^{14}C laboratory

site_code The site/museum code

location The location of the sample

c14_age The observed ^{14}C ages of the samples (in ^{14}C yr BP)

c14_sig The uncertainty in the observed ^{14}C ages reported by the radiocarbon laboratory

f14c The observed F^{14}C concentrations

f14c_sig The uncertainty in the observed F^{14}C concentrations reported by the radiocarbon laboratory

Source

https://doi.org/10.1038/nature04604

PlotCalendarAgeDensityIndividualSample

Plot Posterior Calendar Age Estimate for an Individual Determination after Joint Calibration

---

Description

Once a joint calibration function (any of [PolyaUrnBivarDirichlet](), [WalkerBivarDirichlet]() or [PPcalibrate]()) has been run to calibrate a set of related radiocarbon determinations, this function plots the posterior calendar age estimate for a given single determination. Shown are a (direct) histogram of the posterior calendar ages generated by the MCMC chain and also a (smoothed) kernel density estimate obtained using a Gaussian kernel. The highest posterior density (HPD) interval is also shown for the interval width specified (default 2σ).

For more information read the vignettes:
vignette("Non-parametric-summed-density", package = "carbondate")
vignette("Poisson-process-modelling", package = "carbondate")

Note: The output of this function will provide different results from independent calibration of the determination. By jointly, and simultaneously, calibrating all the related ^{14}C determinations using the library functions we are able to share the available calendar information between the samples. This should result in improved individual calibration.

Usage

```
PlotCalendarAgeDensityIndividualSample(
  ident,
  output_data,
  calibration_curve = NULL,
  plot_14C_age = TRUE,
```

```
    plot_cal_age_scale = "BP",
    hist_resolution = 5,
    density_resolution = 1,
    interval_width = "2sigma",
    bespoke_probability = NA,
    n_burn = NA,
    n_end = NA,
    show_hpd_ranges = FALSE,
    show_unmodelled_density = FALSE,
    plot_pretty = TRUE
)
```

Arguments

ident
: The individual determination for which you want to plot the posterior density estimate of the calendar age.

output_data
: The return value either from one of the Bayesian non-parametric DPMM functions ([PolyaUrnBivarDirichlet](#) or [WalkerBivarDirichlet](#)); or from the Poisson process modelling function ([PPcalibrate](#)).

calibration_curve
:
: This is usually not required since the name of the calibration curve variable is saved in the output data. However, if the variable with this name is no longer in your environment then you should pass the calibration curve here. If provided, this should be a dataframe which should contain at least 3 columns entitled `calendar_age`, `c14_age` and `c14_sig`. This format matches [intcal20](#).

plot_14C_age
: Whether to use the radiocarbon age (^{14}C yr BP) as the units of the y-axis in the plot. Defaults to `TRUE`. If `FALSE` uses F^{14}C concentration instead.

plot_cal_age_scale
:
: The calendar scale to use for the x-axis. Allowed values are "BP", "AD" and "BC". The default is "BP" corresponding to plotting in cal yr BP.

hist_resolution
:
: The distance between histogram breaks when plotting the individual posterior calendar age density. Default is 5.

density_resolution
:
: The distance between calendar ages for the returned smoothed calendar age probability. Default is 1.

interval_width
: The confidence intervals to show for the calibration curve and for the highest posterior density ranges. Choose from one of "1sigma" (68.3%), "2sigma" (95.4%) and "bespoke". Default is "2sigma".

bespoke_probability
:
: The probability to use for the confidence interval if "bespoke" is chosen above. E.g., if 0.95 is chosen, then the 95% confidence interval is calculated. Ignored if "bespoke" is not chosen.

n_burn
: The number of MCMC iterations that should be discarded as burn-in (i.e., considered to be occurring before the MCMC has converged). This relates to the number of iterations (n_iter) when running the original update functions (not the thinned output_data). Any MCMC iterations before this are not used

in the calculations. If not given, the first half of the MCMC chain is discarded. Note: The maximum value that the function will allow is n_iter – 100 * n_thin (where n_iter and n_thin are the arguments given to [PolyaUrnBivarDirichlet](#) or [WalkerBivarDirichlet](#)) which would leave only 100 of the (thinned) values in output_data.

n_end   The iteration number of the last sample in output_data to use in the calculations. Assumed to be the total number of (thinned) realisations stored if not given.

show_hpd_ranges

Set to TRUE to also show the highest posterior density (HPD) range on the plot.

show_unmodelled_density

Set to TRUE to also show the unmodelled density (i.e., the result of independent calibration using [CalibrateSingleDetermination](#)) on the plot. Default is FALSE.

plot_pretty   logical, defaulting to TRUE. If set TRUE then will select pretty plotting margins (that create sufficient space for axis titles and rotates y-axis labels). If FALSE will implement current user values.

Value

A data frame with one column calendar_age_BP containing the calendar ages, and the other column probability containing the (smoothed) kernel density estimate of the probability at that calendar age.

See Also

[CalibrateSingleDetermination](#) for independent calibration of a sample against a calibration curve.

Examples

```
# NOTE 1: These examples are shown with a small n_iter to speed up execution.
# When you run ensure n_iter gives convergence (try function default).

# NOTE 2: The examples only show application to  PolyaUrnBivarDirichlet output.
# The function can also be used with WalkerBivarDirichlet and PPcalibrate output.

polya_urn_output <- PolyaUrnBivarDirichlet(
    two_normals$c14_age,
    two_normals$c14_sig,
    intcal20,
    n_iter = 500,
    n_thin = 2,
    show_progress = FALSE)

# Result for 15th determination
PlotCalendarAgeDensityIndividualSample(15, polya_urn_output)

# Now change to show 1 sigma interval for HPD range and calibration curve
# and plot in yr AD
PlotCalendarAgeDensityIndividualSample(
    15,
```

```
    polya_urn_output,
    plot_cal_age_scale = "AD",
    interval_width = "1sigma",
    show_hpd_ranges = TRUE,
    show_unmodelled_density = TRUE)

# Plot and then assign the returned probability
posterior_dens <- PlotCalendarAgeDensityIndividualSample(15, polya_urn_output)
# Use this to find the mean posterior calendar age
weighted.mean(posterior_dens$calendar_age_BP, posterior_dens$probability)
```

PlotConvergenceData   *Plot KL Divergence of Predictive Density to Assess Convergence of*
                             Bayesian Non-Parametric DPMM Sampler

---

Description

This plots the Kullback-Leibler (KL) divergence between a fixed (initial/baseline) predictive density and the predictive density calculated from later individual realisations in the MCMC run of one of the Bayesian non-parametric summarisation approach. The divergence from the initial predictive density is plotted as a function of the realisation/iteration number.

This aims to identify when the divergence, from the initial estimate of the shared $f(\theta)$ to the current estimate, has begun to stabilise. Hence, to (informally) assess when the MCMC chain has converged to equilibrium for the shared, underlying, predictive $f(\theta)$.

For more information read the vignette:
vignette("determining-convergence", package = "carbondate")

Usage

```
PlotConvergenceData(output_data, n_initial = NA)
```

Arguments

output_data	The return value from one of the Bayesian non-parametric DPMM summarisation functions, i.e., PolyaUrnBivarDirichlet or WalkerBivarDirichlet.
n_initial	The number of (thinned) realisations to use for the 'initial' predictive shared density. This predictive density is then compared with the predictive obtained at each subsequent realisation in the (thinned) DPMM output. If not specified, then the minimum of 1000 realisations, or 1 / 10 of the total number of realisations, will be used.

Value

None

Examples

```
# Plot results for the example two_normal data
# NOTE: This does not show meaningful results as n_iter
# is too small. Try increasing n_iter to 1e5.
polya_urn_output <- PolyaUrnBivarDirichlet(
    two_normals$c14_age,
    two_normals$c14_sig,
    intcal20,
    n_iter = 500,
    show_progress = FALSE)
PlotConvergenceData(polya_urn_output)
```

PlotGelmanRubinDiagnosticMultiChain

> *Plot Histogram of the Gelman-Rubin Convergence Diagnostic for Multiple Independent MCMC Chains*

Description

This plots a histogram of the potential scale reduction factors (PSRF) for each of the individual posterior calendar age estimates for multiple independent MCMC chains. Achieved by comparing the *within-chain* variance with the *between-chains* variance after n_burn iterations. The PSRF of each sample's posterior calendar age is calculated. If the chain have converged to the target posterior distribution, then PSRF should be close to 1 for all of the samples (a stringent condition is that all values are less than 1.1).

For more information read the vignette:
```
vignette("determining-convergence", package = "carbondate")
```

Usage

```
PlotGelmanRubinDiagnosticMultiChain(output_data_list, n_burn = NA)
```

Arguments

output_data_list

> A list, each item containing the return value from one of the updating functions e.g. PolyaUrnBivarDirichlet, WalkerBivarDirichlet or PPcalibrate. The minimum number of elements in the list is 2.

n_burn
> The number of MCMC iterations that should be discarded for burn-in. This relates to the total number of iterations n_iter when running the original update functions (not the thinned output_data). Any MCMC iterations before this are not used in the calculations of the PSRF. If not given, the first half of the MCMC chain is discarded. Note: The maximum value that the function will allow is n_iter - 100 * n_thin (where n_iter and n_thin are the arguments that were given to PPcalibrate) which would leave only 100 of the (thinned) values in output_data.

Value

None

Examples

```
# Plot results for the example data - n_iter is too small for convergence
# Try increasing n_iter to see the values of the PSRF decrease
po <- list()
for (i in 1:3) {
    set.seed(i)
    po[[i]] <- PolyaUrnBivarDirichlet(
        two_normals$c14_age,
        two_normals$c14_sig,
        intcal20,
        n_iter=400,
        show_progress = FALSE)
}
PlotGelmanRubinDiagnosticMultiChain(po)
```

PlotGelmanRubinDiagnosticSingleChain

> *Plot Histogram of the Gelman-Rubin Convergence Diagnostic for a Single MCMC Chain*

Description

This plots a histogram of the potential scale reduction factors (PSRF) for each of the individual posterior calendar age estimates within a single MCMC chain. Achieved by splitting the chain into segments after n_burn and comparing the *within-chain* variance with the *between-chains* variance of the segments. The PSRF of each sample's posterior calendar age is calculated. If the chain have converged to the target posterior distribution, then PSRF should be close to 1 for all of the samples (a stringent condition is that all values are less than 1.1).

For more information read the vignette:
vignette("determining-convergence", package = "carbondate")

Usage

```
PlotGelmanRubinDiagnosticSingleChain(output_data, n_burn = NA, n_segments = 3)
```

Arguments

output_data   The return value from one of the updating functions, e.g., PolyaUrnBivarDirich-
               let, WalkerBivarDirichlet or PPcalibrate.

n_burn   The number of MCMC iterations that should be discarded for burn-in. This
               relates to the total number of iterations n_iter when running the original update
               functions (not the thinned output_data). Any MCMC iterations before this are
               not used in the calculations of the PSRF. If not given, the first half of the MCMC

chain is discarded. Note: The maximum value that the function will allow is n_iter - 100 * n_thin (where n_iter and n_thin are the arguments that were given to [PPcalibrate](#)) which would leave only 100 of the (thinned) values in output_data.

n_segments   The number of segments to split the chain into. Default is 3, must be a number between 2 and 10.

Value

None

Examples

```
# Plot results for the example data - n_iter is too small for convergence
# Try increasing n_iter to see the values of the PSRF decrease
polya_urn_output <- PolyaUrnBivarDirichlet(
    two_normals$c14_age,
    two_normals$c14_sig,
    intcal20,
    n_iter = 500,
    show_progress = FALSE)
PlotGelmanRubinDiagnosticSingleChain(polya_urn_output)
```

PlotNumberOfClusters   *Plot Number of Calendar Age Clusters Estimated in Bayesian Non-Parametric DPMM Output*

Description

Given output from one of the Bayesian non-parametric summarisation functions (either [PolyaUrn-BivarDirichlet](#) or [WalkerBivarDirichlet](#)) plot the estimated number of calendar age clusters represented by the ^{14}C samples.

For more information read the vignette:
vignette("Non-parametric-summed-density", package = "carbondate")

Usage

```
PlotNumberOfClusters(output_data, n_burn = NA, n_end = NA)
```

Arguments

output_data   The return value from one of the Bayesian non-parametric DPMM functions, e.g. [PolyaUrnBivarDirichlet](#) or [WalkerBivarDirichlet](#), or a list, each item containing one of these return values. Optionally, the output data can have an extra list item named label which is used to set the label on the plot legend.

n_burn   The number of MCMC iterations that should be discarded as burn-in (i.e., con-
                sidered to be occurring before the MCMC has converged). This relates to the
                number of iterations (n_iter) when running the original update functions (not
                the thinned output_data). Any MCMC iterations before this are not used
                in the calculations. If not given, the first half of the MCMC chain is dis-
                carded. Note: The maximum value that the function will allow is n_iter –
                100 * n_thin (where n_iter and n_thin are the arguments given to PolyaU-
                rnBivarDirichlet or WalkerBivarDirichlet) which would leave only 100 of the
                (thinned) values in output_data.

n_end   The last iteration in the original MCMC chain to use in the calculations. As-
                sumed to be the total number of iterations performed, i.e. n_iter, if not given.

Value

None

See Also

PlotPredictiveCalendarAgeDensity and PlotCalendarAgeDensityIndividualSample for more plot-
ting functions using DPMM output.

Examples

```
# NOTE: these examples are shown with a small n_iter to speed up execution.
# When you run ensure n_iter gives convergence (try function default).

polya_urn_output <- PolyaUrnBivarDirichlet(
    two_normals$c14_age,
    two_normals$c14_sig,
    intcal20,
    n_iter = 500,
    n_thin = 2,
    show_progress = FALSE)

PlotNumberOfClusters(polya_urn_output)
```

PlotNumberOfInternalChanges
                        Plot Number of Changepoints in Rate of Sample Occurrence for Pois-
                        son Process Model

Description

Given output from the Poisson process fitting function PPcalibrate, plot the posterior distribution
for the number of internal changepoints in the underlying rate of sample occurrence (i.e., in $\lambda(t)$)
over the period under study.

For more information read the vignette:
vignette("Poisson-process-modelling", package = "carbondate")

Usage

```
PlotNumberOfInternalChanges(output_data, n_burn = NA, n_end = NA)
```

Arguments

output_data   The return value from the updating function [PPcalibrate]. Optionally, the output
               data can have an extra list item named label which is used to set the label on
               the plot legend.

n_burn   The number of MCMC iterations that should be discarded as burn-in (i.e., con-
               sidered to be occurring before the MCMC has converged). This relates to the
               number of iterations (n_iter) when running the original update functions (not
               the thinned output_data). Any MCMC iterations before this are not used
               in the calculations. If not given, the first half of the MCMC chain is dis-
               carded. Note: The maximum value that the function will allow is n_iter - 100
               * n_thin (where n_iter and n_thin are the arguments that were given to [PP-
               calibrate]) which would leave only 100 of the (thinned) values in output_data.

n_end   The last iteration in the original MCMC chain to use in the calculations. As-
               sumed to be the total number of iterations performed, i.e. n_iter, if not given.

Value

None

Examples

```
# NOTE: This example is shown with a small n_iter to speed up execution.
# Try n_iter and n_posterior_samples as the function defaults.

pp_output <- PPcalibrate(
    pp_uniform_phase$c14_age,
    pp_uniform_phase$c14_sig,
    intcal20,
    n_iter = 1000,
    show_progress = FALSE)

PlotNumberOfInternalChanges(pp_output)
```

PlotPosteriorChangePoints

Plot Calendar Ages of Changes in Rate of Sample Occurrence for Poisson Process Model

---

Description

Given output from the Poisson process fitting function [PPcalibrate](#), plot the posterior density estimates for the calendar ages at which there are internal changepoints in the rate of sample occurrence $\lambda(t)$. These density estimates are calculated **conditional** upon the number of internal changepoints within the period under study (which is specified as an input to the function).

Having conditioned on the number of changes, n_change, the code will extract all realisations from the the posterior of the MCMC sampler which have that number of internal changepoints in the estimate of $\lambda(t)$. It will then provide density estimates for the (ordered) calendar ages of those internal changepoints. These density estimates are obtained using a Gaussian kernel.

Note: These graphs will become harder to interpret as the specified number of changepoints increases

For more information read the vignette:
```
vignette("Poisson-process-modelling", package = "carbondate")
```

Usage

```
PlotPosteriorChangePoints(
  output_data,
  n_changes = c(1, 2, 3),
  plot_cal_age_scale = "BP",
  n_burn = NA,
  n_end = NA,
  kernel_bandwidth = NA,
  plot_lwd = 2
)
```

Arguments

output_data   The return value from the updating function [PPcalibrate](#). Optionally, the output
               data can have an extra list item named label which is used to set the label on
               the plot legend.

n_changes   Number of internal changepoints to condition on, and plot for. A vector which
               can contain at most 4 elements, with values in the range 1 to 6. If not given, then
               c(1, 2, 3) will be used.

plot_cal_age_scale
               (Optional) The calendar scale to use for the x-axis. Allowed values are "BP",
               "AD" and "BC". The default is "BP" corresponding to plotting in cal yr BP.

n_burn   The number of MCMC iterations that should be discarded as burn-in (i.e., con-
               sidered to be occurring before the MCMC has converged). This relates to the
               number of iterations (n_iter) when running the original update functions (not
               the thinned output_data). Any MCMC iterations before this are not used
               in the calculations. If not given, the first half of the MCMC chain is dis-
               carded. Note: The maximum value that the function will allow is n_iter - 100
               * n_thin (where n_iter and n_thin are the arguments that were given to [PP-
               calibrate](#)) which would leave only 100 of the (thinned) values in output_data.

n_end   The last iteration in the original MCMC chain to use in the calculations. As-
               sumed to be the total number of iterations performed, i.e. n_iter, if not given.

kernel_bandwidth

    (Optional) The bandwidth used for the (Gaussian) kernel smoothing of the calendar age densities. If not given, then 1/50th of the overall calendar age range will be used.

plot_lwd   The line width to use when plotting the posterior mean (and confidence intervals). Default is 2 (to add emphasis).

Value

None

Examples

```
# NOTE: This example is shown with a small n_iter to speed up execution.
# Try n_iter and n_posterior_samples as the function defaults.

pp_output <- PPcalibrate(
    pp_uniform_phase$c14_age,
    pp_uniform_phase$c14_sig,
    intcal20,
    n_iter = 1000,
    show_progress = FALSE)

# Plot the posterior change points for only 2 or 3 internal changes
PlotPosteriorChangePoints(pp_output, n_changes = c(2, 3))

# Changing the calendar age plotting scale to cal AD
PlotPosteriorChangePoints(pp_output, n_changes = c(2, 3),
    plot_cal_age_scale = "AD")
```

PlotPosteriorHeights   *Plot Heights of Segments in Rate of Sample Occurrence for Poisson Process Model*

Description

Given output from the Poisson process fitting function PPcalibrate, plot the posterior density estimates for the heights (i.e., values) of the piecewise-constant rate $\lambda(t)$ used to model sample occurrence. These density estimates are calculated **conditional** upon the number of internal changepoints within the period under study (which is specified as an input to the function).

Having conditioned on the number of changes, n_change, the code will extract all realisations from the the posterior of the MCMC sampler which have that number of internal changepoints in the estimate of $\lambda(t)$. It will then provide density estimates for the heights (i.e., the value) of the rate function between each of the determined (ordered) changepoints. These density estimates are obtained using a Gaussian kernel.

Note: These graphs will become harder to interpret as the specified number of changepoints increases

For more information read the vignette:
vignette("Poisson-process-modelling", package = "carbondate")

Usage

```
PlotPosteriorHeights(
  output_data,
  n_changes = c(1, 2, 3),
  n_burn = NA,
  n_end = NA,
  kernel_bandwidth = NA,
  plot_lwd = 2
)
```

Arguments

output_data   The return value from the updating function PPcalibrate. Optionally, the output
               data can have an extra list item named label which is used to set the label on
               the plot legend.

n_changes   Number of internal changepoints to condition on, and plot for. A vector which
               can contain at most 4 elements, with values in the range 1 to 6. If not given, then
               c(1, 2, 3) will be used.

n_burn   The number of MCMC iterations that should be discarded as burn-in (i.e., con-
               sidered to be occurring before the MCMC has converged). This relates to the
               number of iterations (n_iter) when running the original update functions (not
               the thinned output_data). Any MCMC iterations before this are not used
               in the calculations. If not given, the first half of the MCMC chain is dis-
               carded. Note: The maximum value that the function will allow is n_iter - 100
               * n_thin (where n_iter and n_thin are the arguments that were given to PP-
               calibrate) which would leave only 100 of the (thinned) values in output_data.

n_end   The last iteration in the original MCMC chain to use in the calculations. As-
               sumed to be the total number of iterations performed, i.e. n_iter, if not given.

kernel_bandwidth
               (Optional) The bandwidth used for the (Gaussian) kernel smoothing of the cal-
               endar age densities. If not given, 1/50th of the maximum height will be used.

plot_lwd   The line width to use when plotting the posterior mean (and confidence inter-
               vals). Default is 2 (to add emphasis).

Value

None

Examples

```
# NOTE: This example is shown with a small n_iter to speed up execution.
# Try n_iter and n_posterior_samples as the function defaults.

pp_output <- PPcalibrate(
    pp_uniform_phase$c14_age,
    pp_uniform_phase$c14_sig,
    intcal20,
    n_iter = 1000,
```

```
      show_progress = FALSE)

  # Plot the posterior heights for only 2 or 3 internal changes
  PlotPosteriorHeights(pp_output, n_changes = c(2, 3))
```

PlotPosteriorMeanRate   *Plot Posterior Mean Rate of Sample Occurrence for Poisson Process*
                         Model

Description

Given output from the Poisson process fitting function PPcalibrate calculate and plot the posterior mean rate of sample occurrence (i.e., the underlying Poisson process rate $\lambda(t)$) together with specified probability intervals, on a given calendar age grid (provided in cal yr BP).

Will show the original set of radiocarbon determinations (those you are modelling/summarising), the chosen calibration curve, and the estimated posterior rate of occurrence $\lambda(t)$ on the same plot. Can also optionally show the posterior mean of each individual sample's calendar age estimate.

An option is also provided to calculate the posterior mean rate *conditional* upon the number of internal changepoints within the period under study (if this is specified as an input to the function).

Note: If all you are interested in is the value of the posterior mean rate on a grid, without an accompanying plot, you can use FindPosteriorMeanRate instead.

For more information read the vignette:
vignette("Poisson-process-modelling", package = "carbondate")

Usage

```
PlotPosteriorMeanRate(
  output_data,
  n_posterior_samples = 5000,
  n_changes = NULL,
  calibration_curve = NULL,
  plot_14C_age = TRUE,
  plot_cal_age_scale = "BP",
  show_individual_means = TRUE,
  show_confidence_intervals = TRUE,
  interval_width = "2sigma",
  bespoke_probability = NA,
  denscale = 3,
  resolution = 1,
  n_burn = NA,
  n_end = NA,
  plot_pretty = TRUE,
  plot_lwd = 2
)
```

Arguments

output_data	The return value from the updating function [PPcalibrate](). Optionally, the output data can have an extra list item named label which is used to set the label on the plot legend.
n_posterior_samples	
	Number of samples it will draw, after having removed n_burn, from the (thinned) MCMC realisations stored in output_data to estimate the rate $\lambda(t)$. These samples may be repeats if the number of, post burn-in, realisations is less than n_posterior_samples. If not given, 5000 is used.
n_changes	(Optional) If wish to condition calculation of the posterior mean on the number of internal changepoints. In this function, if specified, n_changes must be a single integer.
calibration_curve	
	This is usually not required since the name of the calibration curve variable is saved in the output data. However, if the variable with this name is no longer in your environment then you should pass the calibration curve here. If provided, this should be a dataframe which should contain at least 3 columns entitled calendar_age, c14_age and c14_sig. This format matches [intcal20]().
plot_14C_age	Whether to use the radiocarbon age (^{14}C yr BP) as the units of the y-axis in the plot. Defaults to TRUE. If FALSE uses $F^{14}C$ concentration instead.
plot_cal_age_scale	
	(Optional) The calendar scale to use for the x-axis. Allowed values are "BP", "AD" and "BC". The default is "BP" corresponding to plotting in cal yr BP.
show_individual_means	
	(Optional) Whether to calculate and show the mean posterior calendar age estimated for each individual ^{14}C sample on the plot as a rug on the x-axis. Default is TRUE.
show_confidence_intervals	
	Whether to show the pointwise confidence intervals (at chosen probability level) on the plot. Default is TRUE.
interval_width	The confidence intervals to show for both the calibration curve and the predictive density. Choose from one of "1sigma" (68.3%), "2sigma" (95.4%) and "bespoke". Default is "2sigma".
bespoke_probability	
	The probability to use for the confidence interval if "bespoke" is chosen above. E.g., if 0.95 is chosen, then the 95% confidence interval is calculated. Ignored if "bespoke" is not chosen.
denscale	(Optional) Whether to scale the vertical range of the Poisson process mean rate plot relative to the calibration curve plot. Default is 3 which means that the maximum of the mean rate will be at 1/3 of the height of the plot.
resolution	The distance between calendar ages at which to calculate the value of the rate $\lambda(t)$. These ages will be created on a regular grid that automatically covers the calendar period specified in output_data. Default is 1.
n_burn	The number of MCMC iterations that should be discarded as burn-in (i.e., considered to be occurring before the MCMC has converged). This relates to the

number of iterations (n_iter) when running the original update functions (not the thinned output_data). Any MCMC iterations before this are not used in the calculations. If not given, the first half of the MCMC chain is discarded. Note: The maximum value that the function will allow is n_iter - 100 * n_thin (where n_iter and n_thin are the arguments that were given to [PP-calibrate](#)) which would leave only 100 of the (thinned) values in output_data.

n_end   The last iteration in the original MCMC chain to use in the calculations. Assumed to be the total number of iterations performed, i.e. n_iter, if not given.

plot_pretty   logical, defaulting to TRUE. If set TRUE then will select pretty plotting margins (that create sufficient space for axis titles and rotates y-axis labels). If FALSE will implement current user values.

plot_lwd   The line width to use when plotting the posterior mean (and confidence intervals). Default is 2 (to add emphasis).

Value

A list, each item containing a data frame of the calendar_age_BP, the rate_mean and the confidence intervals for the rate - rate_ci_lower and rate_ci_upper.

Examples

```
# NOTE: All these examples are shown with a small n_iter and n_posterior_samples
# to speed up execution.
# Try n_iter and n_posterior_samples as the function defaults.

pp_output <- PPcalibrate(
    pp_uniform_phase$c14_age,
    pp_uniform_phase$c14_sig,
    intcal20,
    n_iter = 1000,
    show_progress = FALSE)

# Default plot with 2 sigma interval
PlotPosteriorMeanRate(pp_output, n_posterior_samples = 100)

# Decrease the line width of the posterior mean
PlotPosteriorMeanRate(pp_output, n_posterior_samples = 100, plot_lwd = 1)

# Specify an 80% confidence interval
PlotPosteriorMeanRate(
    pp_output,
    interval_width = "bespoke",
    bespoke_probability = 0.8,
    n_posterior_samples = 100)

# Plot the posterior rate conditional on 2 internal changes
PlotPosteriorMeanRate(
    pp_output,
    n_changes = 2,
    interval_width = "bespoke",
    bespoke_probability = 0.8,
```

```
    n_posterior_samples = 100)
```

PlotPredictiveCalendarAgeDensity

> *Plot Predictive Estimate of Shared Calendar Age Density from Bayesian Non-Parametric DPMM Output*

Description

Given output from one of the Bayesian non-parametric summarisation functions (either PolyaUrnBivarDirichlet or WalkerBivarDirichlet) calculate and plot the predictive (summarised/shared) calendar age density and probability intervals on a given calendar age grid (provided in cal yr BP).

Will show the original set of radiocarbon determinations (those you are summarising), the chosen calibration curve, and the summarised predictive calendar age density on the same plot. Can also optionally show the SPD estimate.

Note: If all you are interested in is the estimated value of the predictive density on a grid, without an accompanying plot, you can use FindPredictiveCalendarAgeDensity instead.

For more information read the vignette:
`vignette("Non-parametric-summed-density", package = "carbondate")`

Usage

```
PlotPredictiveCalendarAgeDensity(
  output_data,
  n_posterior_samples = 5000,
  calibration_curve = NULL,
  plot_14C_age = TRUE,
  plot_cal_age_scale = "BP",
  show_SPD = FALSE,
  show_confidence_intervals = TRUE,
  interval_width = "2sigma",
  bespoke_probability = NA,
  denscale = 3,
  resolution = 1,
  n_burn = NA,
  n_end = NA,
  plot_pretty = TRUE,
  plot_lwd = 2
)
```

Arguments

output_data   The return value from one of the Bayesian non-parametric DPMM functions, e.g. PolyaUrnBivarDirichlet or WalkerBivarDirichlet, or a list, each item containing one of these return values. Optionally, the output data can have an extra list item named label which is used to set the label on the plot legend.

n_posterior_samples

    Number of samples it will draw, after having removed n_burn, from the (thinned) realisations stored in the DPMM outputs to estimate the predictive calendar age density. These samples may be repeats if the number of, post burn-in, realisations is less than n_posterior_samples. If not given, 5000 is used.

calibration_curve

    This is usually not required since the name of the calibration curve variable is saved in the output data. However, if the variable with this name is no longer in your environment then you should pass the calibration curve here. If provided, this should be a dataframe which should contain at least 3 columns entitled calendar_age, c14_age and c14_sig. This format matches intcal20.

plot_14C_age   Whether to use the radiocarbon age (^{14}C yr BP) as the units of the y-axis in the plot. Defaults to TRUE. If FALSE uses $F^{14}C$ concentration instead.

plot_cal_age_scale

    The calendar scale to use for the x-axis. Allowed values are "BP", "AD" and "BC". The default is "BP" corresponding to plotting in cal yr BP.

show_SPD   Whether to calculate and show the summed probability distribution on the plot (optional). Default is FALSE.

show_confidence_intervals

    Whether to show the pointwise confidence intervals (at the chosen probability level) on the plot. Default is TRUE.

interval_width   The confidence intervals to show for both the calibration curve and the predictive density. Choose from one of "1sigma" (68.3%), "2sigma" (95.4%) and "bespoke". Default is "2sigma".

bespoke_probability

    The probability to use for the confidence interval if "bespoke" is chosen above. E.g., if 0.95 is chosen, then the 95% confidence interval is calculated. Ignored if "bespoke" is not chosen.

denscale   Whether to scale the vertical range of the summarised calendar age density plot relative to the calibration curve plot (optional). Default is 3 which means that the maximum predictive density will be at 1/3 of the height of the plot.

resolution   The distance between calendar ages at which to calculate the predictive shared density. These ages will be created on a regular grid that automatically covers the calendar period of the given set of ^{14}C samples. Default is 1.

n_burn   The number of MCMC iterations that should be discarded as burn-in (i.e., considered to be occurring before the MCMC has converged). This relates to the number of iterations (n_iter) when running the original update functions (not the thinned output_data). Any MCMC iterations before this are not used in the calculations. If not given, the first half of the MCMC chain is discarded. Note: The maximum value that the function will allow is n_iter – 100 * n_thin (where n_iter and n_thin are the arguments given to [PolyaU-rnBivarDirichlet](PolyaUrnBivarDirichlet) or WalkerBivarDirichlet) which would leave only 100 of the (thinned) values in output_data.

n_end   The last iteration in the original MCMC chain to use in the calculations. Assumed to be the total number of iterations performed, i.e. n_iter, if not given.

plot_pretty   logical, defaulting to TRUE. If set TRUE then will select pretty plotting margins
               (that create sufficient space for axis titles and rotates y-axis labels). If FALSE
               will implement current user values.

plot_lwd   The line width to use when plotting the posterior mean (and confidence inter-
               vals). Default is 2 (to add emphasis).

Value

A list, each item containing a data frame of the calendar_age_BP, the density_mean and the
confidence intervals for the density density_ci_lower and density_ci_upper for each set of
output data.

See Also

FindPredictiveCalendarAgeDensity if only interested in the estimated value of the predictive density
on a grid; PlotNumberOfClusters and PlotCalendarAgeDensityIndividualSample for more plotting
functions using DPMM output.

Examples

```
# NOTE: All these examples are shown with a small n_iter and n_posterior_samples
# to speed up execution.
# Try n_iter and n_posterior_samples as the function defaults.

polya_urn_output <- PolyaUrnBivarDirichlet(
    two_normals$c14_age,
    two_normals$c14_sig,
    intcal20,
    n_iter = 500,
    show_progress = FALSE)
walker_output <- WalkerBivarDirichlet(
    two_normals$c14_age,
    two_normals$c14_sig,
    intcal20,
    n_iter = 500,
    show_progress = FALSE)

# Plot results for a single calibration
PlotPredictiveCalendarAgeDensity(polya_urn_output, n_posterior_samples = 50)

# Plot results from two calibrations on the same plot
PlotPredictiveCalendarAgeDensity(
    list(walker_output, polya_urn_output), n_posterior_samples = 50)

# Plot and show the 80% confidence interval, show the SPD, add a custom label
polya_urn_output$label = "My plot"
PlotPredictiveCalendarAgeDensity(
    polya_urn_output,
    n_posterior_samples = 50,
    interval_width = "bespoke",
    bespoke_probability = 0.8,
    show_SPD = TRUE)
```

PlotRateIndividualRealisation

Plot Individual Realisations of Posterior Rate of Sample Occurrence for Poisson Process Model

Description

Given output from the Poisson process fitting function PPcalibrate plot individual realisations from the MCMC for the rate of sample occurrence (i.e., realisations of the underlying Poisson process rate $\lambda(t)$), on a given calendar age grid (provided in cal yr BP). Specify either n_realisations if you want to select a random set of realisations, or realisations if you want to provide a vector of specific realisations.

Usage

```
PlotRateIndividualRealisation(
  output_data,
  n_realisations = 10,
  plot_realisations_colour = NULL,
  realisations = NULL,
  calibration_curve = NULL,
  plot_14C_age = TRUE,
  plot_cal_age_scale = "BP",
  interval_width = "2sigma",
  bespoke_probability = NA,
  denscale = 3,
  resolution = 1,
  n_burn = NA,
  n_end = NA,
  plot_pretty = TRUE,
  plot_lwd = 2
)
```

Arguments

output_data   The return value from the updating function PPcalibrate. Optionally, the output data can have an extra list item named label which is used to set the label on the plot legend.

n_realisations Number of randomly sampled realisations to be drawn from MCMC posterior and plotted. Default is 10.

plot_realisations_colour

The colours to be used to plot the individual realisations. Default is greyscale (otherwise should have same length as number of realisations).

realisations   Specific indices of realisations (in thinned version) to plot if user does not want to sample realisations randomly). If specified will override n_realisations.

calibration_curve

> This is usually not required since the name of the calibration curve variable is saved in the output data. However, if the variable with this name is no longer in your environment then you should pass the calibration curve here. If provided, this should be a dataframe which should contain at least 3 columns entitled `calendar_age`, `c14_age` and `c14_sig`. This format matches intcal20.

plot_14C_age
> Whether to use the radiocarbon age (^{14}C yr BP) as the units of the y-axis in the plot. Defaults to `TRUE`. If `FALSE` uses $F^{14}C$ concentration instead.

plot_cal_age_scale

> (Optional) The calendar scale to use for the x-axis. Allowed values are "BP", "AD" and "BC". The default is "BP" corresponding to plotting in cal yr BP.

interval_width
> The confidence intervals to show for the calibration curve. Choose from one of `"1sigma"` (68.3%), `"2sigma"` (95.4%) and `"bespoke"`. Default is `"2sigma"`.

bespoke_probability

> The probability to use for the confidence interval if `"bespoke"` is chosen above. E.g., if 0.95 is chosen, then the 95% confidence interval is calculated. Ignored if `"bespoke"` is not chosen.

denscale
> (Optional) Whether to scale the vertical range of the Poisson process mean rate plot relative to the calibration curve plot. Default is 3 which means that the maximum of the mean rate will be at 1/3 of the height of the plot.

resolution
> The distance between calendar ages at which to calculate the value of the rate $\lambda(t)$. These ages will be created on a regular grid that automatically covers the calendar period specified in `output_data`. Default is 1.

n_burn
> The number of MCMC iterations that should be discarded as burn-in (i.e., considered to be occurring before the MCMC has converged). This relates to the number of iterations (`n_iter`) when running the original update functions (not the thinned `output_data`). Any MCMC iterations before this are not used in the calculations. If not given, the first half of the MCMC chain is discarded. Note: The maximum value that the function will allow is `n_iter - 100 * n_thin` (where `n_iter` and `n_thin` are the arguments that were given to [PPcalibrate](PP-calibrate)) which would leave only 100 of the (thinned) values in `output_data`.

n_end
> The last iteration in the original MCMC chain to use in the calculations. Assumed to be the total number of iterations performed, i.e. `n_iter`, if not given.

plot_pretty
> logical, defaulting to `TRUE`. If set `TRUE` then will select pretty plotting margins (that create sufficient space for axis titles and rotates y-axis labels). If `FALSE` will implement current user values.

plot_lwd
> The line width to use when plotting the posterior mean (and confidence intervals). Default is 2 (to add emphasis).

Value

None

Examples

```
#' # NOTE: All these examples are shown with a small n_iter and n_posterior_samples
```

```
# to speed up execution.
# Try n_iter and n_posterior_samples as the function defaults.

pp_output <- PPcalibrate(
    pp_uniform_phase$c14_age,
    pp_uniform_phase$c14_sig,
    intcal20,
    n_iter = 1000,
    show_progress = FALSE)

# Plot 10 random realisations in greyscale
PlotRateIndividualRealisation(
    pp_output,
    n_realisations = 10)

# Plot three random realisations with specific colours
PlotRateIndividualRealisation(
    pp_output,
    n_realisations = 3,
    plot_realisations_colour = c("red", "green", "purple"))

# Plot some specific realisations
PlotRateIndividualRealisation(
    pp_output,
    realisations = c(60, 73, 92),
    plot_realisations_colour = c("red", "green", "purple"))
```

PolyaUrnBivarDirichlet

Calibrate and Summarise Multiple Radiocarbon Samples via a Bayesian Non-Parametric DPMM (with Polya Urn Updating)

Description

This function calibrates sets of multiple radiocarbon (^{14}C) determinations, and simultaneously summarises the resultant calendar age information. This is achieved using Bayesian non-parametric density estimation, providing a statistically-rigorous alternative to summed probability distributions (SPDs).

It takes as an input a set of ^{14}C determinations and associated 1σ uncertainties, as well as the radiocarbon age calibration curve to be used. The samples are assumed to arise from an (unknown) shared calendar age distribution $f(\theta)$ that we would like to estimate, alongside performing calibration of each sample.

The function models the underlying distribution $f(\theta)$ as a Dirichlet process mixture model (DPMM), whereby the samples are considered to arise from an unknown number of distinct clusters. Fitting is achieved via MCMC.

It returns estimates for the calendar age of each individual radiocarbon sample; and broader output (including the means and variances of the underpinning calendar age clusters) that can be used by other library functions to provide a predictive estimate of the shared calendar age density $f(\theta)$.

For more information read the vignette:
`vignette("Non-parametric-summed-density", package = "carbondate")`

Note: The library provides two slightly-different update schemes for the MCMC. In this particular function, updating of the DPMM is achieved by a Polya Urn approach (Neal 2000) This is our recommended updating approach based on testing. The alternative, slice-sampled, approach can be found at [WalkerBivarDirichlet](#)

References:
Heaton, TJ. 2022. "Non-parametric Calibration of Multiple Related Radiocarbon Determinations and their Calendar Age Summarisation." *Journal of the Royal Statistical Society Series C: Applied Statistics* **71** (5):1918-56. https://doi.org/10.1111/rssc.12599.
Neal, RM. 2000. "Markov Chain Sampling Methods for Dirichlet Process Mixture Models." *Journal of Computational and Graphical Statistics* **9** (2):249 https://doi.org/10.2307/1390653.

Usage

```
PolyaUrnBivarDirichlet(
  rc_determinations,
  rc_sigmas,
  calibration_curve,
  F14C_inputs = FALSE,
  n_iter = 1e+05,
  n_thin = 10,
  use_F14C_space = TRUE,
  slice_width = NA,
  slice_multiplier = 10,
  n_clust = min(10, length(rc_determinations)),
  show_progress = TRUE,
  sensible_initialisation = TRUE,
  lambda = NA,
  nu1 = NA,
  nu2 = NA,
  A = NA,
  B = NA,
  alpha_shape = NA,
  alpha_rate = NA,
  mu_phi = NA,
  calendar_ages = NA
)
```

Arguments

rc_determinations

> A vector of observed radiocarbon determinations. Can be provided either as ^{14}C ages (in ^{14}C yr BP) or as F^{14}C concentrations.

rc_sigmas   A vector of the (1-sigma) measurement uncertainties for the radiocarbon determinations. Must be the same length as `rc_determinations` and given in the same units.

calibration_curve

A dataframe which must contain one column `calendar_age_BP`, and also columns `c14_age` and `c14_sig` or `f14c` and `f14c_sig` (or both sets). This format matches the curves supplied with this package, e.g., [intcal20](#), [intcal13](#), which contain all 5 columns.

F14C_inputs   `TRUE` if the provided `rc_determinations` are $F^{14}C$ concentrations and `FALSE` if they are radiocarbon ages. Defaults to `FALSE`.

n_iter   The number of MCMC iterations (optional). Default is 100,000.

n_thin   How much to thin the MCMC output (optional). Will store every `n_thin`[th] iteration. 1 is no thinning, while a larger number will result in more thinning. Default is 10. Must choose an integer greater than 1. Overall number of MCMC realisations stored will be $n_{\text{out}} = \text{floor}(n_{\text{iter}}/n_{\text{thin}}) + 1$ so do not choose `n_thin` too large to ensure there are enough samples from the posterior to use for later inference.

use_F14C_space   If `TRUE` (default) the calculations within the function are carried out in $F^{14}C$ space. If `FALSE` they are carried out in ^{14}C age space. We recommend selecting `TRUE` as, for very old samples, calibrating in $F^{14}C$ space removes the potential affect of asymmetry in the radiocarbon age uncertainty. *Note:* This flag can be set independently of the format/scale on which `rc_determinations` were originally provided.

slice_width   Parameter for slice sampling (optional). If not given a value is chosen intelligently based on the spread of the initial calendar ages. Must be given if `sensible_initialisation` is `FALSE`.

slice_multiplier

Integer parameter for slice sampling (optional). Default is 10. Limits the slice size to `slice_multiplier * slice_width`.

n_clust   The number of clusters with which to initialise the sampler (optional). Must be less than the length of `rc_determinations`. Default is 10 or the length of `rc_determinations` if that is less than 10.

show_progress   Whether to show a progress bar in the console during execution. Default is `TRUE`.

sensible_initialisation

Whether to use sensible values to initialise the sampler and an automated (adaptive) prior on μ_ϕ and (A, B) that is informed by the observed `rc_determinations`. If this is `TRUE` (the recommended default), then all the remaining arguments below are ignored.

lambda, nu1, nu2   Hyperparameters for the prior on the mean ϕ_j and precision τ_j of each individual calendar age cluster j:

$$(\phi_j, \tau_j)|\mu_\phi \sim \text{NormalGamma}(\mu_\phi, \lambda, \nu_1, \nu_2)$$

where μ_ϕ is the overall cluster centering. Required if `sensible_initialisation` is `FALSE`.

A, B   Prior on μ_ϕ giving the mean and precision of the overall centering $\mu_\phi \sim N(A, B^{-1})$. Required if `sensible_initialisation` is `FALSE`.

alpha_shape, alpha_rate

Shape and rate hyperparameters that specify the prior for the Dirichlet Process (DP) concentration, α. This concentration α determines the number of clusters

we expect to observe among our n sampled objects. The model places a prior on $\alpha \sim \Gamma(\eta_1, \eta_2)$, where η_1, η_2 are the alpha_shape and alpha_rate. A small α means the DPMM is more concentrated (i.e. we expect fewer calendar age clusters) while a large alpha means it is less less concentrated (i.e. many clusters). Required if sensible_initialisation is FALSE.

mu_phi   Initial value of the overall cluster centering μ_ϕ. Required if sensible_initialisation is FALSE.

calendar_ages   The initial estimate for the underlying calendar ages (optional). If supplied, it must be a vector with the same length as rc_determinations. Required if sensible_initialisation is FALSE.

Value

A list with 10 items. The first 8 items contain output of the model, each of which has one dimension of size $n_{\mathrm{out}} = \mathrm{floor}(n_{\mathrm{iter}}/n_{\mathrm{thin}}) + 1$. The rows in these items store the state of the MCMC from every $n_{\mathrm{thin}}^{\mathrm{th}}$ iteration:

cluster_identifiers   A list of length n_{out} each entry gives the cluster allocation (an integer between 1 and n_clust) for each observation on the relevant MCMC iteration. Information on the state of these calendar age clusters (means and precisions) can be found in the other output items.

alpha   A double vector of length n_{out} giving the Dirichlet Process concentration parameter α.

n_clust   An integer vector of length n_{out} giving the current number of clusters in the model.

phi   A list of length n_{out} each entry giving a vector of length n_clust of the means of the current calendar age clusters ϕ_j.

tau   A list of length n_{out} each entry giving a vector of length n_clust of the precisions of the current calenadar age cluster τ_j.

observations_per_cluster   A list of length n_{out} each entry giving a vector of length n_clust of the number of observations for that cluster.

calendar_ages   An n_{out} by n_{obs} integer matrix. Gives the current estimate for the calendar age of each individual observation.

mu_phi   A vector of length n_{out} giving the overall centering μ_ϕ of the calendar age clusters.

where n_{obs} is the number of radiocarbon observations i.e. the length of rc_determinations.

The remaining items give information about the input data, input parameters (or those calculated using sensible_initialisation) and the update_type

update_type   A string that always has the value "Polya Urn".

input_data   A list containing the ^{14}C data used, and the name of the calibration curve used.

input_parameters   A list containing the values of the fixed hyperparameters lambda, nu1, nu2, A, B, alpha_shape, alpha_rate and mu_phi, and the slice parameters slice_width and slice_multiplier.

See Also

WalkerBivarDirichlet for our less-preferred MCMC method to update the Bayesian DPMM (otherwise an identical model); and PlotCalendarAgeDensityIndividualSample, PlotPredictiveCalendarAgeDensity and PlotNumberOfClusters to access the model output and estimate the calendar age information.

See also PPcalibrate for an an alternative (similarly rigorous) approach to calibration and summarisation of related radiocarbon determinations using a variable-rate Poisson process

Examples

```
# Note these examples are shown with a small n_iter to speed up execution.
# When you run ensure n_iter gives convergence (try function default).

# Basic usage making use of sensible initialisation to set most values and
# using a saved example data set and the IntCal20 curve.
polya_urn_output <- PolyaUrnBivarDirichlet(
    two_normals$c14_age,
    two_normals$c14_sig,
    intcal20,
    n_iter = 100,
    show_progress = FALSE)

# The radiocarbon determinations can be given as F14C concentrations
polya_urn_output <- PolyaUrnBivarDirichlet(
    two_normals$f14c,
    two_normals$f14c_sig,
    intcal20,
    F14C_inputs = TRUE,
    n_iter = 100,
    show_progress = FALSE)
```

PPcalibrate	*Model Occurrence of Multiple Radiocarbon Samples as a Variable-Rate Poisson Process*

Description

This function calibrates a set of radiocarbon (^{14}C) samples, and provides an estimate of how the underlying rate at which the samples occurred varies over calendar time (including any specific changepoints in the rate). The method can be used as an alternative approach to summarise calendar age information contained in a set of related ^{14}C samples, enabling inference on the latent *activity* rate which led to their creation.

It takes as an input a set of ^{14}C determinations and associated 1σ uncertainties, as well as the radiocarbon age calibration curve to be used. The (calendar age) occurrence of these radiocarbon (^{14}C) samples is modelled as a Poisson process. The underlying rate of this Poisson process $\lambda(t)$, which represents the rate at which the samples occurred, is considered unknown and assumed to vary over calendar time.

Specifically, the sample occurrence rate $\lambda(t)$ is modelled as piecewise constant, but with an un-known number of changepoints, which can occur at unknown times. The value of $\lambda(t)$ between any set of changepoints is also unknown. The function jointly calibrates the given ^{14}C samples under this model, and simultaneously provides an estimate of $\lambda(t)$. Fitting is performed using reversible-jump MCMC within Gibbs.

It returns estimates for the calendar age of each individual radiocarbon sample in the input set; and broader output on the estimated value of $\lambda(t)$ which can be used by other library functions. Analysis of the estimated changepoints in the piecewise $\lambda(t)$ permits wider inference on whether the occurrence rate of samples changed significantly at any particular calendar time and, if so, when and by how much.

For more information read the vignette:
vignette("Poisson-process-modelling", package = "carbondate")

Usage

```
PPcalibrate(
  rc_determinations,
  rc_sigmas,
  calibration_curve,
  F14C_inputs = FALSE,
  n_iter = 1e+05,
  n_thin = 10,
  use_F14C_space = TRUE,
  show_progress = TRUE,
  calendar_age_range = NA,
  calendar_grid_resolution = 1,
  prior_h_shape = NA,
  prior_h_rate = NA,
  prior_n_internal_changepoints_lambda = 3,
  k_max_internal_changepoints = 30,
  rescale_factor_rev_jump = 0.9,
  bounding_range_prob_cutoff = 0.001,
  initial_n_internal_changepoints = 10,
  grid_extension_factor = 0.1,
  use_fast = TRUE,
  fast_approx_prob_cutoff = 0.001
)
```

Arguments

rc_determinations

A vector of observed radiocarbon determinations. Can be provided either as ^{14}C ages (in ^{14}C yr BP) or as F^{14}C concentrations.

rc_sigmas   A vector of the (1-sigma) measurement uncertainties for the radiocarbon deter-minations. Must be the same length as rc_determinations and given in the same units.

calibration_curve

A dataframe which must contain one column calendar_age_BP, and also columns

c14_age and c14_sig or f14c and f14c_sig (or both sets). This format matches the curves supplied with this package, e.g., intcal20, intcal13, which contain all 5 columns.

F14C_inputs   TRUE if the provided rc_determinations are $F^{14}C$ concentrations and FALSE if they are radiocarbon ages. Defaults to FALSE.

n_iter   The number of MCMC iterations (optional). Default is 100,000.

n_thin   How much to thin the MCMC output (optional). Will store every n_thinth iteration. 1 is no thinning, while a larger number will result in more thinning. Default is 10. Must choose an integer greater than 1. Overall number of MCMC realisations stored will be $n_{out} = \text{floor}(n_{iter}/n_{thin}) + 1$ so do not choose n_thin too large to ensure there are enough samples from the posterior to use for later inference.

use_F14C_space   If TRUE (default) the calculations within the function are carried out in $F^{14}C$ space. If FALSE they are carried out in ^{14}C age space. We recommend selecting TRUE as, for very old samples, calibrating in $F^{14}C$ space removes the potential affect of asymmetry in the radiocarbon age uncertainty. *Note:* This flag can be set independently of the format/scale on which rc_determinations were originally provided.

show_progress   Whether to show a progress bar in the console during execution. Default is TRUE.

calendar_age_range

(Optional) Overall minimum and maximum calendar ages (in cal yr BP) permitted for the set of radiocarbon samples, i.e., calendar_age_range[1] $< \theta_1, \ldots, \theta_n$ < calendar_age_range[1]. This is used to bound the start and end of the Poisson process (so no events will be permitted to occur outside this interval). If not selected then automated selection will be made based on given rc_determinations and value of bounding_range_prob_cutoff

calendar_grid_resolution

The spacing of the calendar age grid on which to restrict the potential calendar ages of the samples, e.g., calendar ages of samples are limited to being one of t, t + resolution, t + 2 * resolution, ... Default is 1 (i.e., all calendar years in the overall calendar range are considered). Primarily used to speed-up code if have large range, when may select larger resolution.

prior_h_shape, prior_h_rate

(Optional) Prior for the value of the Poisson Process rate (the height rate_h) in any specific interval:

$$\text{rate\_h} \sim \text{Gamma}(\text{shape} = \text{prior\_h\_shape}, \text{rate} = \text{prior\_h\_rate}).$$

If they are both NA then prior_h_shape is selected to be 1 (so rate_h follows an Exponential distribution) and prior_h_rate chosen adaptively (internally) to match n_observations.

prior_n_internal_changepoints_lambda

Prior mean for the number of internal changepoints in the rate $\lambda(t)$.

$$\text{n\_internal\_changepoints} \sim \text{Po}(\text{prior\_n\_internal\_changepoints\_lambda})$$

k_max_internal_changepoints

Maximum permitted number of internal changepoints

rescale_factor_rev_jump

> Factor weighting probability of dimension change in the reversible jump update step for Poisson process rate_h and rate_s

bounding_range_prob_cutoff

> Probability cut-off for choosing the bounds for the potential calendar ages for the observations

initial_n_internal_changepoints

> Number of internal changepoints to initialise MCMC sampler with. The default is 10 (so initialise with diffuse state). Will place these initial changepoints uniformly at random within overall calendar age range.

grid_extension_factor

> If you adaptively select the calendar_age_range from rc_determinations, how far you wish to extend the grid beyond this adaptive minimum and maximum. The final range will be extended (equally on both sides) to cover (1 + grid_extension_factor) * (calendar_age_range)

use_fast, fast_approx_prob_cutoff

> A flag to allow trimming the calendar age likelihood (i.e., reducing the range of calendar ages) for each individual sample to speed up the sampler. If TRUE (default), for each individual sample, those tail calendar ages (in the overall calendar_age_grid) with very small likelihoods will be trimmed (speeding up the updating of the calendar ages). If TRUE the probability cut-off to remove the tails is fast_approx_prob_cutoff.

Value

A list with 7 items. The first 4 items contain output of the model, each of which has one dimension of size $n_{\text{out}} = \text{floor}(n_{\text{iter}}/n_{\text{thin}}) + 1$. The rows in these items store the state of the MCMC from every n_{thin}th iteration:

rate_s A list of length n_{out} each entry giving the current set of (calendar age) changepoint locations in the piecewise-constant rate $\lambda(t)$.

rate_h A list of length n_{out} each entry giving the current set of heights (values for the rate) in each piecewise-constant section of $\lambda(t)$.

calendar_ages An n_{out} by n_{obs} matrix. Gives the current estimate for the calendar age of each individual observation.

n_internal_changes A vector of length n_{out} giving the current number of internal changes in the value of $\lambda(t)$.

where n_{obs} is the number of radiocarbon observations, i.e., the length of rc_determinations.

The remaining items give information about input data, input parameters (or those calculated) and update_type

update_type A string that always has the value "RJPP".

input_data A list containing the ^{14}C data used, and the name of the calibration curve used.

input_parameters A list containing the values of the fixed parameters pp_cal_age_range, prior_n_internal_changepo
k_max_internal_changepoints, prior_h_shape, prior_h_rate, rescale_factor_rev_jump,
calendar_age_grid, calendar_grid_resolution, n_iter and n_thin.

Examples

```
# NOTE: This example is shown with a small n_iter to speed up execution.
# When you run ensure n_iter gives convergence (try function default).

pp_output <- PPcalibrate(
    pp_uniform_phase$c14_age,
    pp_uniform_phase$c14_sig,
    intcal20,
    n_iter = 100,
    show_progress = FALSE)
```

---

pp_uniform_phase   *Example artificial data - Uniform Phase*

Description

40 simulated radiocarbon determinations for which the underlying calendar ages are drawn (uniformly at random) from the period 550–500 cal yr BP.

$$f(\theta) = U[550, 500]$$

The observational uncertainty of each determination is set to be 15 ^{14}C yrs.

The corresponding ^{14}C ages are then simulated based upon the IntCal20 calibration curve (convolved with the 15 ^{14}C yr measurement uncertainty):

$$X_i|\theta_i \sim N(m(\theta_i), \rho(\theta_i)^2 + 15^2),$$

where $m(\theta_i)$ and $\rho(\theta_i)$ are the IntCal20 pointwise means and uncertainties.

This dataset matches that used in the package vignette to illustrate the Poisson process modelling.

Usage

```
pp_uniform_phase
```

Format

pp_uniform_phase:

A data frame with 40 rows and 4 columns:

c14_age The simulated ^{14}C age (in ^{14}C yr BP)

c14_sig The (fixed) ^{14}C age measurement uncertainty used in the simulation (set at 15 ^{14}C yrs)

f14c The corresponding simulated values of F^{14}C concentration

f14c_sig The (fixed) corresponding F^{14}C measurement uncertainty used in the simulation

---

shcal04   *SHCal04 calibration curve*

Description

The SHCal04 Southern Hemisphere radiocarbon age calibration curve on a calendar grid spanning from 11,000–0 cal yr BP (Before Present, 0 cal yr BP corresponds to 1950 CE).

Note: This dataset provides ^{14}C ages and F^{14}C values on a calendar age grid. This is different from the ^{14}C ages and Δ^{14}C values provided in oxcal .14c files.

Reference:
FG McCormac, AG Hogg, PG Blackwell, CE Buck, TFG Higham, and PJ Reimer 2004. SHCal04 Southern Hemisphere Calibration 0–11.0 cal kyr BP. *Radiocarbon* **46**(3):1087–1092 https://doi.org/10.1017/S003382220003

Usage

    shcal04

Format

    shcal04:

A data frame with 2,202 rows and 5 columns providing the SHCal04 radiocarbon age calibration curve on a calendar grid spanning from 11,000–0 cal yr BP:

calendar_age The calendar age (in cal yr BP)

c14_age The ^{14}C age (in ^{14}C yr BP)

c14_sig The (1-σ) uncertainty in the ^{14}C age

f14c The ^{14}C age expressed as F^{14}C concentration

f14c_sig The (1-σ) uncertainty in the F^{14}C concentration

Source

http://doi.org/10.1017/S0033822200033014

---

shcal13   *SHCal13 calibration curve*

Description

The SHCal13 Southern Hemisphere radiocarbon age calibration curve on a calendar grid spanning from 50,000–0 cal yr BP (Before Present, 0 cal yr BP corresponds to 1950 CE).

Note: This dataset provides ^{14}C ages and F^{14}C values on a calendar age grid. This is different from the ^{14}C ages and Δ^{14}C values provided in oxcal .14c files.

Reference:

Alan G Hogg, Quan Hua, Paul G Blackwell, Caitlin E Buck, Thomas P Guilderson, Timothy J Heaton, Mu Niu, Jonathan G Palmer, Paula J Reimer, Ron W Reimer, Christian S M Turney, Susan R H Zimmerman. 2013. SHCal13 Southern Hemisphere Calibration, 0-50,000 Years cal BP. *Radiocarbon* **55**(4):1889–1903 https://doi.org/10.2458/azu_js_rc.55.16783.

Usage

```
shcal13
```

Format

`shcal13`:

A data frame with 5,141 rows and 5 columns providing the SHCal13 radiocarbon age calibration curve on a calendar grid spanning from 50,000–0 cal yr BP:

calendar_age The calendar age (in cal yr BP)

c14_age The ^{14}C age (in ^{14}C yr BP)

c14_sig The (1-σ) uncertainty in the ^{14}C age

f14c The ^{14}C age expressed as F^{14}C concentration

f14c_sig The (1-σ) uncertainty in the F^{14}C concentration

Source

http://doi.org/10.2458/azu_js_rc.55.16783

---

shcal20 *SHCal20 calibration curve*

Description

The SHCal20 Southern Hemisphere radiocarbon age calibration curve on a calendar grid spanning from 55,000–0 cal yr BP (Before Present, 0 cal yr BP corresponds to 1950 CE).

Note: This dataset provides ^{14}C ages and F^{14}C values on a calendar age grid. This is different from the ^{14}C ages and Δ^{14}C values provided in oxcal .14c files.

Reference:

Hogg AG, Heaton TJ, Hua Q, Palmer JG, Turney CSM, Southon J, Bayliss A, Blackwell PG, Boswijk G, Bronk Ramsey C, Pearson C, Petchey F, Reimer P, Reimer R, Wacker L. 2020. SHCal20 Southern Hemisphere calibration, 0–55,000 years cal BP. *Radiocarbon* **62** https://doi.org/10.1017/RDC.2020.59.

Usage

```
shcal20
```

Format

shcal20:

A data frame with 9,501 rows and 5 columns providing the SHCal20 radiocarbon age calibration curve on a calendar grid spanning from 55,000–0 cal yr BP:

calendar_age The calendar age (in cal yr BP)

c14_age The ^{14}C age (in ^{14}C yr BP)

c14_sig The (1-σ) uncertainty in the ^{14}C age

f14c The ^{14}C age expressed as F^{14}C concentration

f14c_sig The (1-σ) uncertainty in the F^{14}C concentration

Source

http://doi.org/10.1017/RDC.2020.59

---

two_normals   *Example artificial data - Mixture of Normal Phases*

Description

50 simulated radiocarbon determinations for which the underlying calendar ages are drawn from a mixture of two normals:

$$f(\theta) = 0.45N(3500, 200^2) + 0.55N(5000, 100^2)$$

i.e., a mixture of a normal centred around 3500 cal yr BP; and another (slightly more concentrated/narrower) normal centred around 5000 cal yr BP.

The corresponding 50 radiocarbon ages were then simulated using the IntCal20 calibration curve incorporating both the uncertainty in the calibration curve and a hypothetical measurement uncertainty:

$$X_i|\theta_i \sim N(m(\theta_i), \rho(\theta_i)^2 + \sigma_{i,\text{lab}}^2),$$

where $m(\theta_i)$ and $\rho(\theta_i)$ are the IntCal20 pointwise means and uncertainties; and $\sigma_{i,\text{lab}}$, the simulated laboratory measurement uncertainty, was fixed at a common value of 25 ^{14}C yrs.

This dataset is included simply to give some quick-to-run examples.

Usage

```
two_normals
```

Format

two_normals:

A data frame with 50 rows and 4 columns:

c14_age The simulated ^{14}C age (in ^{14}C yr BP)

c14_sig The (fixed) ^{14}C age measurement uncertainty used in the simulation (set at 25 ^{14}C yrs)

f14c The corresponding simulated values of F^{14}C concentration

f14c_sig The (fixed) corresponding F^{14}C measurement uncertainty used in the simulation

Examples

```
# Plotting calendar age density underlying two_normals
# Useful for comparisons against estimation techniques
weights_true <- c(0.45, 0.55)
cluster_means_true_calBP <- c(3500, 5000)
cluster_precisions_true <- 1 / c(200, 100)^2

# Create mixture density
truedens <- function(t, w, truemean, trueprec) {
  dens <- 0
  for(i in 1:length(w)) {
    dens <- dens + w[i] * dnorm(t, mean = truemean[i], sd = 1/sqrt(trueprec[i]))
  }
  dens
}

# Visualise mixture
curve(truedens(
  x,
  w = weights_true,
  truemean = cluster_means_true_calBP,
  trueprec = cluster_precisions_true),
  from = 2500, to = 7000, n = 401,
  xlim = c(7000, 2500),
  xlab = "Calendar Age (cal yr BP)",
  ylab = "Density",
  col = "red"
)
```

WalkerBivarDirichlet   *Calibrate and Summarise Multiple Radiocarbon Samples via a Bayesian Non-Parametric DPMM (with Walker Updating)*

Description

This function calibrates sets of multiple radiocarbon (^{14}C) determinations, and simultaneously summarises the resultant calendar age information. This is achieved using Bayesian non-parametric density estimation, providing a statistically-rigorous alternative to summed probability distributions (SPDs).

It takes as an input a set of ^{14}C determinations and associated 1σ uncertainties, as well as the radiocarbon age calibration curve to be used. The samples are assumed to arise from an (unknown) shared calendar age distribution $f(\theta)$ that we would like to estimate, alongside performing calibration of each sample.

The function models the underlying distribution $f(\theta)$ as a Dirichlet process mixture model (DPMM), whereby the samples are considered to arise from an unknown number of distinct clusters. Fitting is achieved via MCMC.

It returns estimates for the calendar age of each individual radiocarbon sample; and broader output (the weights, means and variances of the underpinning calendar age clusters) that can be used by other library functions to provide a predictive estimate of the shared calendar age density $f(\theta)$.

For more information read the vignette:
`vignette("Non-parametric-summed-density", package = "carbondate")`

Note: The library provides two slightly-different update schemes for the MCMC. In this particular function, updating of the DPMM is achieved by slice sampling (Walker 2007). We recommend use of the alternative to this, implemented at PolyaUrnBivarDirichlet

Reference:
Heaton, TJ. 2022. "Non-parametric Calibration of Multiple Related Radiocarbon Determinations and their Calendar Age Summarisation." *Journal of the Royal Statistical Society Series C: Applied Statistics* **71** (5):1918-56. https://doi.org/10.1111/rssc.12599.
Walker, SG. 2007. "Sampling the Dirichlet Mixture Model with Slices." *Communications in Statistics - Simulation and Computation* **36** (1):45-54. https://doi.org/10.1080/03610910601096262.

Usage

```
WalkerBivarDirichlet(
  rc_determinations,
  rc_sigmas,
  calibration_curve,
  F14C_inputs = FALSE,
  n_iter = 1e+05,
  n_thin = 10,
  use_F14C_space = TRUE,
  slice_width = NA,
  slice_multiplier = 10,
  show_progress = TRUE,
  sensible_initialisation = TRUE,
  lambda = NA,
  nu1 = NA,
  nu2 = NA,
  A = NA,
  B = NA,
```

```
    alpha_shape = NA,
    alpha_rate = NA,
    mu_phi = NA,
    calendar_ages = NA,
    n_clust = min(10, length(rc_determinations))
)
```

Arguments

rc_determinations

> A vector of observed radiocarbon determinations. Can be provided either as ^{14}C ages (in ^{14}C yr BP) or as F^{14}C concentrations.

rc_sigmas   A vector of the (1-sigma) measurement uncertainties for the radiocarbon determinations. Must be the same length as rc_determinations and given in the same units.

calibration_curve

> A dataframe which must contain one column calendar_age_BP, and also columns c14_age and c14_sig or f14c and f14c_sig (or both sets). This format matches the curves supplied with this package, e.g., intcal20, intcal13, which contain all 5 columns.

F14C_inputs   TRUE if the provided rc_determinations are F^{14}C concentrations and FALSE if they are radiocarbon ages. Defaults to FALSE.

n_iter   The number of MCMC iterations (optional). Default is 100,000.

n_thin   How much to thin the MCMC output (optional). Will store every n_thin[th] iteration. 1 is no thinning, while a larger number will result in more thinning. Default is 10. Must choose an integer greater than 1. Overall number of MCMC realisations stored will be $n_{out} = \text{floor}(n_{iter}/n_{thin}) + 1$ so do not choose n_thin too large to ensure there are enough samples from the posterior to use for later inference.

use_F14C_space   If TRUE (default) the calculations within the function are carried out in F^{14}C space. If FALSE they are carried out in ^{14}C age space. We recommend selecting TRUE as, for very old samples, calibrating in F^{14}C space removes the potential affect of asymmetry in the radiocarbon age uncertainty. *Note:* This flag can be set independently of the format/scale on which rc_determinations were originally provided.

slice_width   Parameter for slice sampling (optional). If not given a value is chosen intelligently based on the spread of the initial calendar ages. Must be given if sensible_initialisation is FALSE.

slice_multiplier

> Integer parameter for slice sampling (optional). Default is 10. Limits the slice size to slice_multiplier * slice_width.

show_progress   Whether to show a progress bar in the console during execution. Default is TRUE.

sensible_initialisation

> Whether to use sensible values to initialise the sampler and an automated (adaptive) prior on μ_ϕ and (A, B) that is informed by the observed rc_determinations. If this is TRUE (the recommended default), then all the remaining arguments below are ignored.

lambda, nu1, nu2 Hyperparameters for the prior on the mean ϕ_j and precision τ_j of each individual calendar age cluster j:

$$(\phi_j, \tau_j)|\mu_\phi \sim \text{NormalGamma}(\mu_\phi, \lambda, \nu_1, \nu_2)$$

where μ_ϕ is the overall cluster centering. Required if `sensible_initialisation` is `FALSE`.

A, B Prior on μ_ϕ giving the mean and precision of the overall centering $\mu_\phi \sim N(A, B^{-1})$. Required if `sensible_initialisation` is `FALSE`.

alpha_shape, alpha_rate

Shape and rate hyperparameters that specify the prior for the Dirichlet Process (DP) concentration, α. This concentration α determines the number of clusters we expect to observe among our n sampled objects. The model places a prior on $\alpha \sim \Gamma(\eta_1, \eta_2)$, where η_1, η_2 are the `alpha_shape` and `alpha_rate`. A small α means the DPMM is more concentrated (i.e. we expect fewer calendar age clusters) while a large alpha means it is less less concentrated (i.e. many clusters). Required if `sensible_initialisation` is `FALSE`.

mu_phi Initial value of the overall cluster centering μ_ϕ. Required if `sensible_initialisation` is `FALSE`.

calendar_ages The initial estimate for the underlying calendar ages (optional). If supplied, it must be a vector with the same length as `rc_determinations`. Required if `sensible_initialisation` is `FALSE`.

n_clust The number of clusters with which to initialise the sampler (optional). Must be less than the length of `rc_determinations`. Default is 10 or the length of `rc_determinations` if that is less than 10.

Value

A list with 11 items. The first 8 items contain output of the model, each of which has one dimension of size $n_{\text{out}} = \text{floor}(n_{\text{iter}}/n_{\text{thin}}) + 1$. The rows in these items store the state of the MCMC from every $n_{\text{thin}}{}^{\text{th}}$ iteration:

cluster_identifiers An n_{out} by n_{obs} integer matrix. Provides the cluster allocation (an integer between 1 and n_clust) for each observation on the relevant MCMC iteration. Information on the state of these calendar age clusters (means, precisions, and weights) can be found in the other output items.

alpha A double vector of length n_{out} giving the Dirichlet Process concentration parameter α.

n_clust An integer vector of length n_{out} giving the current number of clusters in the model.

phi A list of length n_{out} each entry giving a vector of the means of the current calendar age clusters ϕ_j.

tau A list of length n_{out} each entry giving a vector of the precisions of the current calendar age clusters τ_j.

weight A list of length n_{out} each entry giving the mixing weights of each calendar age cluster.

calendar_ages An n_{out} by n_{obs} integer matrix. Gives the current estimate for the calendar age of each individual observation.

mu_phi A vector of length n_{out} giving the overall centering μ_ϕ of the calendar age clusters.

where n_{obs} is the number of radiocarbon observations, i.e., the length of `rc_determinations`.

The remaining items give information about the input data, input parameters (or those calculated using `sensible_initialisation`) and the update_type

update_type A string that always has the value "Walker".

input_data A list containing the ^{14}C data used, and the name of the calibration curve used.

input_parameters A list containing the values of the fixed hyperparameters `lambda`, `nu1`, `nu2`, `A`, `B`, `alpha_shape`, and `alpha_rate`, and the slice parameters `slice_width` and `slice_multiplier`.

See Also

PolyaUrnBivarDirichlet for our preferred MCMC method to update the Bayesian DPMM (otherwise an identical model); and PlotCalendarAgeDensityIndividualSample, PlotPredictiveCalendarAgeDensity and PlotNumberOfClusters to access the model output and estimate the calendar age information.

See also PPcalibrate for an an alternative (similarly rigorous) approach to calibration and summarisation of related radiocarbon determinations using a variable-rate Poisson process

Examples

```
# NOTE: These examples are shown with a small n_iter to speed up execution.
# When you run ensure n_iter gives convergence (try function default).

walker_output <- WalkerBivarDirichlet(
    two_normals$c14_age,
    two_normals$c14_sig,
    intcal20,
    n_iter = 100,
    show_progress = FALSE)

# The radiocarbon determinations can be given as F14C concentrations
walker_output <- WalkerBivarDirichlet(
    two_normals$f14c,
    two_normals$f14c_sig,
    intcal20,
    F14C_inputs = TRUE,
    n_iter = 100,
    show_progress = FALSE)
```

Index