Package 'invacost'

February 5, 2022

Type Package

Title Analyse Biological Invasion Costs with the 'InvaCost' Database

Version 1.1-4

Date 2022-02-05

Author Boris Leroy [cre, aut], Andrew Kramer [aut], Anne-Charlotte Vaissière [ctb], Christophe Diagne [ctb]

Maintainer Boris Leroy <leroy.boris@gmail.com>

Description Provides an up-to-date version of the 'InvaCost' database (<doi:10.6084/m9.figshare.12668570>) in R, and several functions to analyse the costs of invasive alien species.

License GPL (>= 2.0)

Encoding UTF-8

Depends R (>= 2.10), ggplot2

Imports dplyr, earth, grDevices, lmtest, mgcv, quantreg, robustbase, sandwich, scales, stats, utils

RoxygenNote 7.1.2

NeedsCompilation no

Repository CRAN

Date/Publication 2022-02-05 11:00:02 UTC

R topics documented:

---

computeAvgTotCost	*Calculate the cumulative and average annual cost over a single period of time*

Description

This simple function calculates the cumulative cost average annual cost of invasive alien species over a single period of time. It is used internally by summarizeCosts.

Usage

```
computeAvgTotCost(
  costdb,
  cost.column = "Cost_estimate_per_year_2017_USD_exchange_rate",
  year.column = "Impact_year",
  min.year = NULL,
  max.year = NULL
)
```

Arguments

costdb	The **expanded 'InvaCost' database** output from expandYearlyCosts, where annual costs occurring over several years are repeated for each year.
cost.column	Name of the cost column to use in costdb (usually, choose between the exchange rate (default) or Purchase Power Parity cost per year)
year.column	Name of the year column to use in costdb.
min.year	The minimum year of the period (specify only if different from the range of data)
max.year	The minimum year of the period (specify only if different from the range of data)

Value

a named list with 5 elements

- initial_year: first year in the data
- final_year: last year in the data
- time_span: the difference between initial and final years.
- total_cost: total cost.
- annual.cost: cost per year
- number_estimates: the number of cost estimates before expansion via expandYearlyCosts
  number_year_values: the number of costs per year included

Note

Arguments `min.year` and `max.year` do not filter the data. Only specify them if you wish to change the interval over which averages are calculated. For example, if your data have values from 1960 to 1964 but you want to calculated the average value from 1960 to 1969, set `min.year = 1960` and `max.year = 1969`.

However, if you want to calculate values for an interval narrower than your data, filter the data BEFORE running this function.

Author(s)

Boris Leroy <leroy.boris@gmail.com>, Andrew Kramer, Anne-Charlotte Vaissière, Christophe Diagne

References

<https://github.com/Farewe/invacost>

Leroy Boris, Kramer Andrew M, Vaissière Anne-Charlotte, Courchamp Franck & Diagne Christophe (2020). Analysing global economic costs of invasive alien species with the invacost R package. biorXiv. doi: 10.1101/2020.12.10.419432

See Also

`expandYearlyCosts` to get the database in appropriate format.

Examples

```
data(invacost)

### Cleaning steps
# Eliminating data with no information on starting and ending years
invacost <- invacost[-which(is.na(invacost$Probable_starting_year_adjusted)), ]
invacost <- invacost[-which(is.na(invacost$Probable_ending_year_adjusted)), ]
# Keeping only observed and reliable costs
invacost <- invacost[invacost$Implementation == "Observed", ]
invacost <- invacost[which(invacost$Method_reliability == "High"), ]
# Eliminating data with no usable cost value
invacost <- invacost[-which(is.na(invacost$Cost_estimate_per_year_2017_USD_exchange_rate)), ]

### Expansion

db.over.time <- expandYearlyCosts(invacost,
                                  startcolumn = "Probable_starting_year_adjusted",
                                  endcolumn = "Probable_ending_year_adjusted")

### Analysis
res <- computeAvgTotCost(db.over.time,
                         min.year = 1960,
                     max.year = 2020) # Excluding data after 2020 (e.g. planned budgets)
res
```

---

expandYearlyCosts | *Expand cost values per year of the database over relevant periods of impact*

Description

For costs that occur over several years, this function repeats the cost value per year over each year of the relevant period of impact.

Usage

```
expandYearlyCosts(costdb, startcolumn, endcolumn)
```

Arguments

costdb   The **'InvaCost' database**, usually obtained with data([invacost](#))

startcolumn   Name of the column containing starting years

endcolumn   Name of the column containing ending years

Value

A data.frame containing the 'InvaCost' database where all costs occurring over several years are repeated for each year.

Note

Information on the beginning and ending years was not directly provided in literature sources of economic costs for a substantial part of entries in the database (Probable_starting_year and Probable_ending_year columns). Therefore, for papers for which this information was not available, educated guesses were made by the 'InvaCost' team on the probable starting and ending years. These educated guesses were designed to be conservative, and make no assumption as of whether the economic impacts have been continued after the publication year of the material where the cost was collected. Therefore, we used the publication year as the probable ending year. For costs repeated over several years but for which no information with respect to the exact periods of impact was available, we counted only a single year. These educated guesses are included in the columns (columns Probable_starting_year_adjusted and Probable_ending_year_adjusted), and we recommend using them the base scenario, as its conservative assumptions limit over-estimations of the costs.

Author(s)

Boris Leroy <leroy.boris@gmail.com>, Andrew Kramer, Anne-Charlotte Vaissière, Christophe Diagne

References

https://github.com/Farewe/invacost

Leroy Boris, Kramer Andrew M, Vaissière Anne-Charlotte, Courchamp Franck & Diagne Christophe (2020). Analysing global economic costs of invasive alien species with the invacost R package. bioRxiv. doi: 10.1101/2020.12.10.419432

Examples

```
data(invacost)

### Cleaning steps
# Eliminating data with no information on starting and ending years
invacost <- invacost[-which(is.na(invacost$Probable_starting_year_adjusted)), ]
invacost <- invacost[-which(is.na(invacost$Probable_ending_year_adjusted)), ]
# Keeping only observed and reliable costs
invacost <- invacost[invacost$Implementation == "Observed", ]
invacost <- invacost[which(invacost$Method_reliability == "High"), ]
# Eliminating data with no usable cost value
invacost <- invacost[-which(is.na(invacost$Cost_estimate_per_year_2017_USD_exchange_rate)), ]

### Expansion

db.over.time <- expandYearlyCosts(invacost,
                                  startcolumn = "Probable_starting_year_adjusted",
                                  endcolumn = "Probable_ending_year_adjusted")
```

---

getInvaCostVersion   *Download and read a specific version of 'InvaCost'*

Description

This function will download the requested major release of 'InvaCost' to the specified file, then read it into R, such that analyses done on older releases of the database can be reproduced.

Usage

```
getInvaCostVersion(version = "4.1", destination_file = NULL)
```

Arguments

version   character indicating the major release to download. Current versions include: "1.0", "2.0", "2.1", "3.0", "4.0" and "4.1",

destination_file
                 character indicating the name of the saved file

Details

The public archive for 'InvaCost' releases is available here: https://figshare.com/articles/
dataset/InvaCost_References_and_description_of_economic_cost_estimates_associated_
with_biological_invasions_worldwide_/12668570

The files used in this function correspond to official releases by the 'InvaCost' team and are down-
loaded in CSV (sep = ";") from a dedicated GitHub repository: https://github.com/Farewe/
invacost_versions

Value

a data.frame with dimensions variable depending on the chosen version.

Author(s)

Boris Leroy <leroy.boris@gmail.com>, Andrew Kramer, Anne-Charlotte Vaissière, Christophe
Diagne

References

https://github.com/Farewe/invacost

Leroy Boris, Kramer Andrew M, Vaissière Anne-Charlotte, Courchamp Franck & Diagne Christophe
(2020). Analysing global economic costs of invasive alien species with the invacost R package.
biorXiv. doi: [10.1101/2020.12.10.419432](10.1101/2020.12.10.419432)

Examples

```
invacost <- getInvaCostVersion("1.0")
dim(invacost)
```

---

  invacost   *'InvaCost' database*

Description

The 'InvaCost' database compiling published values of economic costs of Invasive Alien Species.
Version 4.1

Usage

```
data(invacost)
```

Format

A data frame with 13553 rows and 66 variables

Details

InvaCost_ID Public unique identifier of the cost entry, which is used since the version 4.1 of the database; this is formulated as follows: 'IC_x_y' with 'IC' meaning InvaCost, 'x' being the version number of the database, and 'y' being the sequential number attributed following the order of integration of costs in the database

Cost_ID Original identifier of the cost entry, as proposed by the contributors and then confirmed/modified by the database managers to ensure consistency within the whole database. This code is used to designate the cost entry for (i) internal tracing of the cost estimate by the database managers and (ii) identifying the cost entry when it is mentioned in the Overlap column

Repository Literature engine (Web of Science (WoS), Google Scholar (GS), Google search engine (Go), Pubmed, Scielo) or original source (Targeted collection (TC)) from which the reference was collected (see Diagne et al. 2020 Scientific Data for further details); cells are left empty when no repository was specified or the reference was shared by external users that did not give any information about this

Reference_ID Identifier for the reference where the cost entry is reported; note that this field is currently being improved internally to have a consistent terminology across references within the database

Reference_title Title of the reference where the cost entry is reported. As much as possible, this is the original source where the cost was first provided

Authors Authors of the reference where the cost entry is reported

Abstract If existing/accessible, the abstract of the reference where the cost entry is reported

Publication_year Year of publication of the reference where the cost entry is reported

Language Main language used in the original reference reporting the cost entry; 22 languages are currently recorded in the database: Catalan, Chinese, Croatian, Czech, Danish, Dutch, English, Finnish, French, Galician, German, Greek, Hindi, Icelandic, Italian, Japanese, Norwegian, Portuguese, Russian, Spanish, Swedish, Ukrainian

Type_of_material Type of reference analyzed (i.e. scientific peer-reviewed article or grey literature); for grey literature, the exact nature of the reference was indicated (e.g., official report, press release)

Previous_materials If any, the list of successive materials checked before reaching the original reference providing the cost entry

Availability The accessibility of the original reference as a searchable document (yes/no)

Kingdom Taxonomic kingdom of the invasive species associated with the cost entry

Phylum Taxonomic phylum of the invasive species associated with the cost entry

Class Taxonomic class of the invasive species associated with the cost entry

Order Taxonomic order of the invasive species associated with the cost entry

Family Taxonomic family of the invasive species associated with the cost entry

Genus Taxonomic genus of the invasive species associated with the cost entry

Species Taxonomic species of the invasive species associated with the cost entry

Subspecies Taxonomic sub-species of the invasive species associated with the cost entry

Common_name Non-scientific (or vernacular) name(s) provided in the original reference, or by the International Union for Conservation of Nature (IUCN) when not provided

Environment Type of environment (aquatic, terrestrial, semi-aquatic, diverse/unspecified) where the cost estimate occurred

Environment_IAS Type of environment where the invasive species lives, independently of where the cost occurred: aquatic (species with a close association with aquatic systems at any life stage, including for reproduction, development and/or foraging), semi-aquatic (species with a looser association with aquatic systems) or terrestrial (otherwise); diverse/unspecified is used when there are multiple IAS pertaining to different environments.

HabitatVerbatim Copy from the original reference of the sentence/paragraph indicating the habitat typology of the studied area

Habitat The type of habitat where the cost occurred (1.Forests: closed vegetation dominated by deciduous or evergreen trees; 2.Open forests: woodland vegetation with canopy openings created by environmental stress or disturbance; 3.Scrub: shrublands maintained by environmental stress (aridity) or disturbance; 4.Grasslands: open graminoid-dominated habitats maintained either by climate (steppes, prairies, savannas) or land-use (grazing, mowing) or combination of both – if possible, specified if it corresponds with 4a.Natural grasslands or 4b.Human-maintained grasslands; 5.Sandy: dunes and other habitats on unstable sandy substrate, stressed by low nutrients, drought and disturbed by sand movement; 6.Rocky: cliffs and rock outcrops with very shallow or no soil; 7.Dryland: habitats in which drought stress limits vegetation development; 8.Saline: habitats stressed by high soil salinity; 9.Riparian: a mosaic of wetlands, grasslands, tall-forb stands, scrub and woodlands in stream corridors; 10.Wetland: sites with permanent or seasonal influence of moisture, ranging from oligotrophic to eutrophic; 11.Aquatic: water bodies and streams with submerged and floating plant species; 12.Man-made: habitats created by humans or where human factor is the main shaping force - if possible, specified if it corresponds with 12a.Ruderal (= urban) or 12b.Agricultural habitats

urbanArea Assessment of the geographical area where the cost occurred as being strictly urban if the cost occurred in an urban area only; other if the cost occurred in a non-urban area or in a large area where urban and non-urban areas cannot be distinguished; both if the study compares the cost between urban and non-urban areas (i.e. urban vs. suburban, or suburban vs. non-urban; the comparison must be explicit) or presents the cost across contrasted areas regarding their level of urbanization (e.g. urbanization gradient). Note that we consider the urban nature of study sites purely based on an 'habitat' perspective (i.e. city versus rural areas) rather than a demographic one (e.g. population size or human density)

protectedArea Assessment of the geographical area where the cost occurred as a protected area (Y) or not (N); NA is used if the area comprises both protected and unprotected areas, or if the protection status of the place is unknown

Island Assessment of the geographical area where the cost occurred as an island (Y) or not (N); NA is used when the cost information is not clearly provided, unknown, or comprises both island and mainland together

Geographic_region Geographical region(s) where the cost occurred (Africa, Antarctic-Subantarctic, Asia, Central America, Europe, North America, Oceania, Pacific islands, South America); Diverse/unspecified was used when the cost estimate occurred over several regions simultaneously

Official_country Country where the cost occurred; sometimes, this is not congruent with the geographic region as some territories (e.g., overseas areas) are located in another continent than their official country of attachment; we used information from www.naturalearthdata.com/ as a reference for country's name

State|Province|Administrative_area The second level of geographic division (state, province or territorial administrative area) for the official country where the cost occurred

Location When provided, the precise location (e.g., city, area) where the cost estimate occurred. Otherwise, we put (i) NA when such information was not provided and the cost entry was provided at a unit or site Spatial scale, or (ii) diverse/unspecified when such information was not provided and the cost entry was provided at global, intercontinental, continental, regional or country Spatial scale

Spatial_scale Order of magnitude of the extent, size of the land/water area where the costs incurred. Options include: global (worldwide-scale), intercontinental (areas from two or more geographic regions), continental ('Geographic region' level), regional (several countries within a single 'Geographic region'), country, site (area at intra-country level, including USA states) and unit (well-defined surface area or entity)

Period_of_estimation If provided, the exact period of time covered by the cost, otherwise the raw formulation provided in the reference analyzed (e.g. late 90s, during 5 years)

Time_range Two options: period if the cost is given for a period exceeding a year; or year if the cost is given yearly or for a period up to one year. Alternatively, we put Unspecified if no information is given or guessable from the source

Probable_starting_year Year in which the cost is known or assumed to have started to occurred. When not explicitly provided by the authors, we mentioned unspecified; in the case of a cost estimate provided for a one-year period straddling two calendar years, we mentioned the latest year of the cost occurrence.

Probable_ending_year Year in which the cost is known or assumed to have ended. When not explicitly provided by the authors, we mentioned unspecified; in the case of a cost estimate provided for a one-year period straddling two calendar years, we mentioned the latest year of the cost occurrence.

Probable_starting_year_adjusted Probable starting year and Probable ending year columns where the cells with unspecified information are replaced, as much as possible, by a specific year from educated guesses based on the duration time provided in the original reference (see Period of estimation column). When relevant (e.g. the authors provide a cost that occurs "since/for a well-defined number of years"), we considered the Publication year as a reference for the probable starting/ending year from which we added or subtracted the number of years provided; when vague formulations were used (e.g., early 90s), we still translated them in probable ending/starting year (e.g., 1990–1995); when annual costs were provided, but without clear information on the temporal range, we conservatively considered the year of the cost occurrence (or the Publication year, if not provided) in both columns. When no relevant approximation is feasible, we leave blank in one and/or the other column(s). These columns are those used for obtaining the number of years by which the raw cost estimates are divided to get the cot estimates per year

Probable_ending_year_adjusted Probable starting year and Probable ending year columns where the cells with unspecified information are replaced, as much as possible, by a specific year from educated guesses based on the duration time provided in the original reference (see Period of estimation column). When relevant (e.g. the authors provide a cost that occurs "since/for a well-defined number of years"), we considered the Publication year as a reference for the probable starting/ending year from which we added or subtracted the number of years provided; when vague formulations were used (e.g., early 90s), we still translated them in probable ending/starting year (e.g., 1990–1995); when annual costs were provided, but without clear information on the temporal range, we conservatively considered the year of the cost

occurrence (or the Publication year, if not provided) in both columns. When no relevant approximation is feasible, we leave blank in one and/or the other column(s). These columns are those used for obtaining the number of years by which the raw cost estimates are divided to get the cot estimates per year

Occurrence Status of the cost estimate as potentially ongoing (if the cost can be expected to continue over time) or one-time (if the cost was deemed as unlikely to continue)

Raw_cost_estimate_original_currency Cost estimate directly retrieved from the analyzed reference

Min_Raw_cost_estimate_original_currency Lower boundary of the Raw cost estimate original currency (if a range of estimates was provided by the authors)

Max_Raw_cost_estimate_original_currency Higher boundary of the Raw cost estimate original currency (if a range of estimates was provided by the authors)

Raw_cost_estimate_2017_USD_exchange_rate Raw cost estimate original currency standardised from original Currency and Applicable year to 2017 US$ based on the official market exchange rate (original currency unit per US$) provided by the World Bank Open Data (available at https://data.worldbank.org/indicator/PA.NUS.FCRF?end=2017&start=1960)

Raw_cost_estimate_2017_USD_PPP Raw cost estimate original currency standardized to 2017 US$ based on the official Purchase Power Parity (PPP; original currency unit per US$) provided by the World Bank Open Data (available at https://data.worldbank.org/indicator/PA.NUS.PPP?end=2017&start=1 and the Organisation for Economic Cooperation and Development (available at https://data.oecd.org/conversion/purchas powerparities-ppp.htm)

Cost_estimate_per_year_original_currency Raw cost estimate original currency transformed to a cost estimate per year of the Period of estimation, which was obtained by dividing the raw cost estimate by the number of years between the Probable starting year_completed and Probable ending year_completed. Blank cells are those that have no information in at least one of these two previous columns

Cost_estimate_per_year_2017_USD_exchange_rate Cost estimate per year original currency standardized from original Currency and Applicable year to 2017 USD based on the official market exchange rate (original currency unit per US$) provided by the World Bank Open Data (available at https://data.worldbank.org/indicator/PA.NUS.FCRF?end=2017&start=1960). Blank cells are those that have no information in at least one of these two previous columns

Cost_estimate_per_year_2017_USD_PPP Cost estimate per year original currency standardized from original Currency and Applicable year to 2017 USD based on Purchase Power Parity (PPP; original currency unit per US$) provided by the World Bank Open Data (available at https://data.worldbank.org/indicator/PA.NUS.PPP?end=2017&start=1990) and the Organisation for Economic Cooperation and Development (available at https://data.oecd.org/conversion/purchasing-powerparities-ppp.htm). Blank cells are those that have no information in at least one of these two previous column

Currency Currency of the Raw cost estimate original currency as extracted in the original reference and standardised (when possible) using internationally recognized ISO 4217 codes (https://www.iso.org/fr/iso-4217-currency-codes.html); 50 currencies are currently recorded in InvaCost: ARS (Argentine peso), AUD (Australian Dollars), CAD (Canadian Dollars), CHF (Swiss Fanc), CLP (Chilean Peso), COL (Colombian Peso), CUP (Cuban Peso), CZK (Czech Crown), DEM (Deutsche Mark), DKK (Danish Krone), ESP (Spanish Peseta), EUR (Euro), FJD (Fijian Dollar), FRF (French Franc), GBP (Pound Sterling), HKD (Hong Kong Dollars), HRK (Croatian Kuna), INR (Indian Rupee), ISK (Icelandic Crown), JPY (Yen),

KES (Kenyan Shilling), LKR (Sri Lankan Rupee), LTL (Lithuanian Litas), MAD (Moroccan Dirhams), MGA (Malagasy Ariary), MXN (Mexican Peso), NGN (Naira), NLG (Dutch Guilder), NOK (Norwegian Krone), NZD (New Zealand Dollar), PKR (Pakistani Rupee), PLN (Polish Zoty), REA (Brazilian Real), CNY (Yuan Renminbi), RUB (Russian Ruble), SEK (Swedish Krona), SGD (Singapore Dollar), SOL (Peruvian Sol), TND (Tunisian Dinar), TRY (Turkish Lira), TWD (New Taiwan Dollar), UAH (Ukrainian Hryvnia), USD (United States Dollars), UYU (Uruguay New Peso), XAF (CFA Franc BEAC), XOF (CFA franc BCEAO), XPF (Pacific Franc), ZAR (South African Rand), ZWL (Zimbabwean Dollar)

Applicable_year Year of the Currency value (not the year of the cost occurrence) considered for the conversion/standardization of the cost estimate

Type_of_applicable_year Assessment of the applicable year as effective if explicitly stated by the authors or publication year if no explicit information was provided in the reference analyzed

Implementation This states — at the time of the estimation — whether the reported cost was actually observed (i.e., cost actually incurred) or potential (i.e., not incurred but expected cost)

Acquisition_method Method used to obtain the cost estimate: report/estimation directly obtained or derived (using inference methods) from field-based information or extrapolation (cost predicted beyond the original spatial and/or temporal observation range from computational modelling)

Impacted_sector Sector impacted by the cost estimate in our socio-ecosystems: Agriculture (considered at its broadest sense, food and other useful products produced by human activities through using natural and/plant resources from their ecosystems such as crop growing, livestock breeding, beekeeping, land management); Authorities-Stakeholders (governmental services and/or official organisations such as conservation agencies, forest services, associations) that allocate efforts for the management sensu lato of biological invasions (e.g. control programs, eradication campaigns, research funding) ; Environment (impacts on natural resources, ecological processes and/or ecosystem services that have been valued by authors such as disruption of native habitats or degradation of local habitats); Fishery (fish-based activities and services such as fishing and aquaculture); Forestry (forest-based activities and services such as timber production/industries and private forests); Health (every item directly or indirectly related to the sanitary state of people such as vector control, medical care and other derived damage on human productivity and well-being); Public and social welfare (activities, goods or services contributing - directly or indirectly - to the human well-being and safety in our societies, including local infrastructures such as electric system, quality of life (e.g. income, recreational activities), personal goods (e.g. private properties, lands), public services (e.g. transports, water regulation), and market activities (e.g. tourism, trade)); Unspecified if no information is given in the source.

Type_of_cost A variety of terms/categories pertianing to damage and losses incurred by an invasion (e.g. damage-loss, damage repair, medical care, crop losses) or means dedicated to understand or predict (e.g., research), prevent (e.g., education, biosecurity), detect (e.g., monitoring, surveillance) and/or manage (e.g., control, eradication) the invasions; Unspecified is mentionned if no information is given or guessable from the source

Type_of_cost_merged Categories of the Type of cost column reassigned into damage (economic losses due to direct and/or indirect impacts of invaders, such as yield loss, health injury, land alteration, infrastructure damage, or income reduction), management (monetary resources allocated to mitigate the spread or impacts of invaders, such as prevention, control, research, long-term management, eradication) or mixed (when when costs included both 'damage' and

'management' components); every cost for which the exact nature of cost was not clearly
defined was assigned to unspecified

Management_type Pre-invasion_management (monetary investments for preventing successful
invasions in an area - including quarantine or border inspection, risk analyses, biosecurity
management, etc.), post-invasion_management (money spent for managing invasions in in-
vaded areas - including control, eradication, containment), knowledge funding (money al-
located to all actions and operations that could be of interest at all steps of management at
pre- and post-invasion stages - including administration, communication, education, research,
etc.), or mixed was assigned when costs include at least (and without possibility to disentangle
the specific proportion of) two of the previous categories; every management cost (in the Type
of cost column) for which the exact nature of cost was not clearly defined was assigned to
unspecified. Every entry that has partly or fully associated with damage costs was assigned to
NA

Method_reliability Assessment of the methodological approach used for cost estimation as of (i)
high reliability if either provided by officially pre-assessed materials (peer-reviewed articles
and official reports) or the estimation method was documented, repeatable and/or traceable if
provided by other grey literature, or (ii) low reliability if not

Method_reliability_refined Assessment of the methodological approach used for cost estimation
as of high or low reliability based on the evaluation of the estimation methodology by expert
contributors

Method_reliability_Explanation Detailed explanation why a particular methodological approach
used for cost estimation was deemed as of high or low reliability based on expert opinion

Method_reliability_Expert_Name Complete name and contact details of the expert had deemed
the reliability of the cost entry

Overlap List of cost entries (using Cost_ID) that might overlap with other cost entry(s) – this is
noted as follows: Z(V/W) when the overlap is known and one or more cost ID (V,W) are
included in another cost ID (Z); Z/V/W: when the overlap is not clear; and Z(V); Z/W if there
are more than one unrelated overlaps

Benefit_value(s) Mention (if any) of the benefit value in the analyzed material (yes/no); 'benefit'
refers here to a monetary estimate associated with profitable activities based on IAS. This
definition therefore excludes any economic gain based on avoided or mitigated effects of the
IAS (for instance, due to efficient management actions over time that may result in apparent
benefits from avoidance of damages) - the latter is considered as an "avoided cost" in the
InvaCost database. Note that the benefit figure was not recorded or described as being out of
the scope of InvaCost

Details When necessary, narrative elements deemed important either to understand the cost esti-
mate or to support choices made for completing the database; this column was left unchanged
from the original entries in order to allow trace-back investigations

Initial contributors_names Name of contributor(s) having collated the cost entry

Double-checking Assessment of cost information collated (by at least) two contributors; yes if it
has been double checked, no if it has not. The names/contacts of each contributor are provided
in the Initial contributors_names column

Source

References

Diagne, C., Leroy, B., Gozlan, R. E., Vaissière, A. C., Assailly, C., Nuninger, L., Roiz, D., Jourdain, F., Jaric, I., & Courchamp, F. (2020). InvaCost, a public database of the economic costs of biological invasions worldwide. Scientific Data, 7(1), 1–12. doi: 10.1038/s4159702000586z

https://github.com/Farewe/invacost

Leroy Boris, Kramer Andrew M, Vaissière Anne-Charlotte, Courchamp Franck & Diagne Christophe (2020). Analysing global economic costs of invasive alien species with the invacost R package. bioRxiv. doi: 10.1101/2020.12.10.419432

History of database releases: doi: 10.6084/m9.figshare.12668570

---

modelCosts	*Model the trend of invasive alien species costs over time*

Description

This function fits different models on 'InvaCost' data expressed per year in order to estimate and predict the trend of invasive alien Species costs over time.

Usage

```
modelCosts(
  costdb,
  cost.column = "Cost_estimate_per_year_2017_USD_exchange_rate",
  year.column = "Impact_year",
  cost.transf = "log10",
  in.millions = TRUE,
  confidence.interval = 0.95,
  minimum.year = 1960,
  maximum.year = max(costdb[, year.column]),
  final.year = max(costdb[, year.column]),
  incomplete.year.threshold = NULL,
  incomplete.year.weights = NULL,
  gam.k = -1,
  mars.nprune = NULL,
  ...
)
```

Arguments

costdb	The **expanded 'InvaCost' database** output from expandYearlyCosts, where costs occurring over several years are expanded over each year of impact
cost.column	Name of the cost column to use in costdb (usually, choose between the exchange rate (default) or Purchase Power Parity cost per year)
year.column	Name of the year column to use in costdb

cost.transf   Type of transformation you want to apply on cost values. The default is a log10
                 transformation, which is commonly applied in economics, allows to fit linear re-
                 gression with a normal distribution of residuals, and makes plots easier to read.
                 You can apply another transformation by specifying the name of the transforma-
                 tion function (e.g., natural logarithm, ″log″). Specify NA or NULL to avoid any
                 transformation

in.millions   If TRUE, cost values are transformed in millions (to make graphs easier to read),
                 else if , cost values are not transformed

confidence.interval
                 A numeric value between 0 and 1, corresponding to the desired confidence in-
                 terval around model predictions

minimum.year   The starting year of this analysis. By default, 1960 was chosen because it marks
                 the period from which world bank data is available for exchange rates and infla-
                 tion values

maximum.year   The ending year for this analysis. By default, the last year of costdb is chosen

final.year   The year for which the costs predicted by models is printed in the output. De-
                 fault is the last year of costdb. Note that this is only for convenience, since
                 predictions for all years are available in the estimated.annual.costs element
                 of the output object

incomplete.year.threshold
                 Estimated threshold for incomplete cost data. All years above or equal to this
                 threshold will be excluded from model calibration, because of the time-lag be-
                 tween economic impacts of invasive alien species and the documentation and
                 publication of these impacts

incomplete.year.weights
                 A named vector containing weights of years for the regressions. Useful to de-
                 crease the weights of incomplete years in regressions. Names of this vector must
                 correspond to years

gam.k   The smoothing factor of GAM; default value of -1 lets the GAM find the smooth-
                 ing factor automatically. Provide a manual value if you have expectations about
                 the shape of the curve and want to avoid overfitting because of interannual vari-
                 ations

mars.nprune   The maximum number of model terms in the MARS model. Lowering this value
                 reduces the number of terms in the MARS model, which can be useful if you
                 have expectations about the shape of the curve and want to reduce the impact of
                 interannual variations

...   Other arguments (you do not need them!)

Value

A list with 3 to 6 elements (only the first three are provided if you selected a cost transformation
different from log10):

   • input.data: the input cost data, for reproducibility of analyses

   • cost.data: the costs of invasions per year, as sums of all costs for each year

- parameters: parameters used to run the function. The `minimum.year` and `maximum.year` are based on the input data (i.e., the user may specify `minimum.year = 1960` but the input data may only have data starting from 1970, hence the `minimum.year` will be 1970)

- `fitted.models`: a list of objects containing the fitted models. They can be extracted individually for refining analyses or making new predictions

- `estimated.annual.costs`: a data.frame containing the predicted cost values for each year for all the fitted models

- `RMSE`: an array containing RMSE of models for the calibration data and for all data. **NOTE: the RMSE for Quantile Regressions is not a relevant metric, IGNORE it unless you know what you are doing!**

- `final.year.cost`: a vector containing the estimated annual costs of invasive alien species based on all models for `final.year`.

The structure of this object can be seen using `str()`.

Author(s)

Boris Leroy <leroy.boris@gmail.com>, Andrew Kramer, Anne-Charlotte Vaissière, Christophe Diagne

References

<https://github.com/Farewe/invacost>

Leroy Boris, Kramer Andrew M, Vaissière Anne-Charlotte, Courchamp Franck & Diagne Christophe (2020). Analysing global economic costs of invasive alien species with the invacost R package. bioRxiv. doi: 10.1101/2020.12.10.419432

See Also

`expandYearlyCosts` to get the database in appropriate format.

Examples

```
data(invacost)

### Cleaning steps
# Eliminating data with no information on starting and ending years
invacost <- invacost[-which(is.na(invacost$Probable_starting_year_adjusted)), ]
invacost <- invacost[-which(is.na(invacost$Probable_ending_year_adjusted)), ]
# Keeping only observed and reliable costs
invacost <- invacost[invacost$Implementation == "Observed", ]
invacost <- invacost[which(invacost$Method_reliability == "High"), ]
# Eliminating data with no usable cost value
invacost <- invacost[-which(is.na(invacost$Cost_estimate_per_year_2017_USD_exchange_rate)), ]

### Expansion

db.over.time <- expandYearlyCosts(invacost,
                                  startcolumn = "Probable_starting_year_adjusted",
```

```
                                    endcolumn = "Probable_ending_year_adjusted")

### Analysis
res <- modelCosts(db.over.time)
res
```

---

plot.invacost.costmodel

Plot model predictions of cost trends over time

Description

This function provides different plotting methods for the estimated annual cost of invasive species based on the temporal trend of costs.

Usage

```
## S3 method for class 'invacost.costmodel'
plot(
  x,
  plot.breaks = 10^(-15:15),
  plot.type = "facets",
 models = c("ols.linear", "ols.quadratic", "robust.linear", "robust.quadratic", "gam",
    "mars", "quantile"),
  evaluation.metric = FALSE,
  graphical.parameters = NULL,
  ...
)
```

Arguments

x	The output object from [modelCosts](#)
plot.breaks	a vector of numeric values indicating the plot breaks for the Y axis (cost values)
plot.type	"single" or "facets". Defines the type of plot you want to make: a single facet with all models ("single"), or a facet per category of model ("facets")
models	the models the user would like to appear in the plots. Can be any subset of the models included in 'modelCosts'. Default is all models.
evaluation.metric	
	TRUE or FALSE. If TRUE, the Root Mean Square Error evaluation metric will be displayed on bottom right of the graph (except for quantile regressions, for which it not relevant). The displayed RMSE is the one based on calibration data only (see the slot RMSE in your [modelCosts](#) object)
graphical.parameters	
	set this to "manual" if you want to customise ggplot2 parameters. By default, the following layers are configured: ylab, xlab, scale_x_continuous, theme_bw and, if cost.transf = "log10", scale_y_log10 and annotation_logticks. If you specify grahical.parameters = "manual", all defaults will be ignored.
...	additional arguments, none implemented for now

Note

Error bands represent 95 regression, GAM and quantile regression. We cannot construct confidence intervals around the mean for MARS techniques. However, we can estimate prediction intervals by fitting a variance model to MARS residuals. Hence, the error bands for MARS model represent 95 by fitting a linear model to the residuals of the MARS model. To learn more about this, see varmod

If the legend appears empty (no colours) on your computer screen, try to zoom in the plot, or to write to a file. There is a rare bug where under certain conditions you cannot see the colours in the legend, because of their transparency; zooming in or writing to a file are the best workarounds.

References

https://github.com/Farewe/invacost

Leroy Boris, Kramer Andrew M, Vaissière Anne-Charlotte, Courchamp Franck & Diagne Christophe (2020). Analysing global economic costs of invasive alien species with the invacost R package. bioRxiv. doi: 10.1101/2020.12.10.419432

Examples

```
data(invacost)

### Cleaning steps
# Eliminating data with no information on starting and ending years
invacost <- invacost[-which(is.na(invacost$Probable_starting_year_adjusted)), ]
invacost <- invacost[-which(is.na(invacost$Probable_ending_year_adjusted)), ]
# Keeping only observed and reliable costs
invacost <- invacost[invacost$Implementation == "Observed", ]
invacost <- invacost[which(invacost$Method_reliability == "High"), ]
# Eliminating data with no usable cost value
invacost <- invacost[-which(is.na(invacost$Cost_estimate_per_year_2017_USD_exchange_rate)), ]

### Expansion

db.over.time <- expandYearlyCosts(invacost,
                                  startcolumn = "Probable_starting_year_adjusted",
                                  endcolumn = "Probable_ending_year_adjusted")

### Analysis
res <- modelCosts(db.over.time,
                  minimum.year = 1970,
                  maximum.year = 2020)

### Visualisation
plot(res)
plot(res, plot.type = "single")
```

plot.invacost.costsummary

Plot raw cumulated cost of invasive species over different periods of time

Description

This function provides different plotting methods for the raw average annual cost of invasive species over different periods of time

Usage

```
## S3 method for class 'invacost.costsummary'
plot(
  x,
  plot.breaks = 10^(-15:15),
  plot.type = "points",
  average.annual.values = TRUE,
  cost.transf = "log10",
  graphical.parameters = NULL,
  ...
)
```

Arguments

x	The output object from [summarizeCosts](#)
plot.breaks	a vector of numeric values indicating the plot breaks for the Y axis (cost values)
plot.type	"points" or "bars". Defines the type of plot you want to make; bars are not advised in log scale because the base value (0) is infinite in log-scale.
average.annual.values	
	if TRUE, the plot will represent average annual values rather than cumulative values over the entire period
cost.transf	Type of transformation you want to apply on cost values. Specify NULL to avoid any transformation. Only useful for graphical representation.
graphical.parameters	
	set this to "manual" if you want to customise ggplot2 parameters. By default, the following layers are configured: ylab, xlab, scale_x_continuous, theme_bw and, if cost.transf = "log10", scale_y_log10 and annotation_logticks. If you specify grahical.parameters = "manual", all defaults will be ignored.
...	additional arguments, none implemented for now

References

<https://github.com/Farewe/invacost>

Leroy Boris, Kramer Andrew M, Vaissière Anne-Charlotte, Courchamp Franck & Diagne Christophe (2020). Analysing global economic costs of invasive alien species with the invacost R package. bioRxiv. doi: [10.1101/2020.12.10.419432](#)

Examples

```
data(invacost)

### Cleaning steps
# Eliminating data with no information on starting and ending years
invacost <- invacost[-which(is.na(invacost$Probable_starting_year_adjusted)), ]
invacost <- invacost[-which(is.na(invacost$Probable_ending_year_adjusted)), ]
# Keeping only observed and reliable costs
invacost <- invacost[invacost$Implementation == "Observed", ]
invacost <- invacost[which(invacost$Method_reliability == "High"), ]
# Eliminating data with no usable cost value
invacost <- invacost[-which(is.na(invacost$Cost_estimate_per_year_2017_USD_exchange_rate)), ]

### Expansion

db.over.time <- expandYearlyCosts(invacost,
                                  startcolumn = "Probable_starting_year_adjusted",
                                  endcolumn = "Probable_ending_year_adjusted")

### Analysis
res <- summarizeCosts(db.over.time,
                      minimum.year = 1970,
                      maximum.year = 2020)

### Visualisation
plot(res)
plot(res, plot.type = "bars")
```

---

prettySummary   *Provide a pretty summary of model results*

Description

This function is useful for presenting the main results (coefficients, tests, etc.) of models in a paper

Usage

```
prettySummary(x)
```

Arguments

x   output object from [modelCosts](#)

Author(s)

Boris Leroy <leroy.boris@gmail.com>, Andrew Kramer, Anne-Charlotte Vaissière, Christophe Diagne

References

https://github.com/Farewe/invacost

Leroy Boris, Kramer Andrew M, Vaissière Anne-Charlotte, Courchamp Franck & Diagne Christophe (2020). Analysing global economic costs of invasive alien species with the invacost R package. biorXiv. doi: 10.1101/2020.12.10.419432

Examples

```
data(invacost)

### Cleaning steps
# Eliminating data with no information on starting and ending years
invacost <- invacost[-which(is.na(invacost$Probable_starting_year_adjusted)), ]
invacost <- invacost[-which(is.na(invacost$Probable_ending_year_adjusted)), ]
# Keeping only observed and reliable costs
invacost <- invacost[invacost$Implementation == "Observed", ]
invacost <- invacost[which(invacost$Method_reliability == "High"), ]
# Eliminating data with no usable cost value
invacost <- invacost[-which(is.na(invacost$Cost_estimate_per_year_2017_USD_exchange_rate)), ]

### Expansion

db.over.time <- expandYearlyCosts(invacost,
                                  startcolumn = "Probable_starting_year_adjusted",
                                  endcolumn = "Probable_ending_year_adjusted")

### Analysis
res <- modelCosts(db.over.time)
prettySummary(res)
```

---

summarizeCosts   *Summarize costs of invasions over periods of time*

Description

This function summarizes the cumulative costs and average annual costs of invasive alien species and breaks it down into regular periods of time, on the basis of cost estimates as they appeared in the provided in the source references collected in the 'InvaCost' database

Usage

```
summarizeCosts(
  costdb,
  cost.column = "Cost_estimate_per_year_2017_USD_exchange_rate",
  year.column = "Impact_year",
  in.millions = TRUE,
  minimum.year = 1960,
  maximum.year = max(costdb[, year.column]),
```

```
    year.breaks = seq(minimum.year, maximum.year, by = 10),
    include.last.year = TRUE
)
```

Arguments

costdb	The **expanded 'InvaCost' database** output from [expandYearlyCosts](#), where costs occurring over several years are expanded over each year of impact.
cost.column	Name of the cost column to use in costdb (usually, choose between the exchange rate (default) or Purchase Power parity cost per year)
year.column	Name of the year column to use in costdb(usually, "Impact_year" from [expandYearlyCosts](#)
in.millions	If TRUE, cost values are transformed in millions (to make graphs easier to read), otherwise if FALSE, cost values are not transformed.
minimum.year	The starting year of the analysis. By default, 1960 was chosen because it marks the period from which world bank data is available for exchange rates and inflation values.
maximum.year	The ending year for the analysis. By default, the last year of costdb is chosen.
year.breaks	A vector of breaks for the year intervals over which you want to summarize cost values
include.last.year	
	TRUE or FALSE. Defines if the last year of the dataset is included in the last interval (TRUE) or is considered as an interval of its own (FALSE). Generally only useful if the last year is at the limit of an interval.

Details

Missing data are ignored. However, note that the average for each interval is always calculated on the basis of the full temporal range. For example, if there is only data for 1968 for the 1960-1969 interval, then the total cost for the interval will be equal to the cost of 1968, and the average annual cost for 1960-1969 will be the cost of 1968 / 10.

Value

A list with 6 elements:

- cost.data: the input data
- parameters: parameters used to run the function. The minimum.year and maximum.year are based on the input data (i.e., the user may specify minimum.year = 1960 but the input data may only have data starting from 1970, hence the minimum.year will be 1970)
- year.breaks: the years used to define year intervals over which costs were calculated.
- cost.per.year: the costs of invasions expressed per year, as sums of all costs for each year
- average.total.cost: the average annual cost of invasive Alien species calculated over the entire time period
- average.cost.per.period: a data.frame containing the average annual cost of invasive alien species calculated over each time interval

The structure of this object can be seen using str().

Author(s)

Boris Leroy <leroy.boris@gmail.com>, Andrew Kramer, Anne-Charlotte Vaissière, Christophe Diagne

References

<https://github.com/Farewe/invacost>

Leroy Boris, Kramer Andrew M, Vaissière Anne-Charlotte, Courchamp Franck & Diagne Christophe (2020). Analysing global economic costs of invasive alien species with the invacost R package. biorXiv. doi: 10.1101/2020.12.10.419432

See Also

expandYearlyCosts to get the database in appropriate format.

Examples

```
data(invacost)

### Cleaning steps
# Eliminating data with no information on starting and ending years
invacost <- invacost[-which(is.na(invacost$Probable_starting_year_adjusted)), ]
invacost <- invacost[-which(is.na(invacost$Probable_ending_year_adjusted)), ]
# Keeping only observed and reliable costs
invacost <- invacost[invacost$Implementation == "Observed", ]
invacost <- invacost[which(invacost$Method_reliability == "High"), ]
# Eliminating data with no usable cost value
invacost <- invacost[-which(is.na(invacost$Cost_estimate_per_year_2017_USD_exchange_rate)), ]

### Expansion

db.over.time <- expandYearlyCosts(invacost,
                                  startcolumn = "Probable_starting_year_adjusted",
                                  endcolumn = "Probable_ending_year_adjusted")

### Analysis
res <- summarizeCosts(db.over.time,
                maximum.year = 2020) # Excluding data after 2020 (e.g. planned budgets)
res
```

Index