

Package: pixelate.fork (via r-universe)

June 3, 2024

Title Pixelate spatial predictions as per their average uncertainty

Version 0.0.1

Description Pixelate spatially continuous predictions as per their average uncertainty. The package pixelate centres around a single function also called pixelate. The function pixelate groups predictions into a specified number of large pixels; computes the average uncertainty within each large pixel; then, for each large pixel, depending on its average uncertainty, either averages the predictions across it or across smaller pixels nested within it. The averaged predictions can then be plotted. The resulting plot of averaged predictions is selectively pixelated, similar to a photo that is deliberately pixelated to disguise a person's identity. Areas of high average uncertainty in the pixelated plot are unresolved, while areas with high average certainty are resolved, similar to information poor versus rich regions of a satellite map.

License MIT + file LICENSE

URL <https://github.com/artaylor85/pixelate>

BugReports <https://github.com/artaylor85/pixelate/issues>

Encoding UTF-8

LazyData true

Imports dplyr (>= 0.8.3), sp

Suggests ggplot2 (>= 3.2.1), markdown, knitr (>= 1.24), rmarkdown (>= 1.15)

RoxygenNote 7.2.3

VignetteBuilder knitr

LazyDataCompression xz

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

RemoteUrl <https://github.com/bobverity/pixelate>

RemoteRef HEAD

RemoteSha d58d242178581f2538b6e0f9585323b627a8e2ac

Contents

CentralAfrica_Pf_incidence	2
CentralAfrica_shp	3
pixelate	4
SubSaharanAfrica_Pf_incidence	7
SubSaharanAfrica_shp	8
Index	9

CentralAfrica_Pf_incidence

Plasmodium falciparum predicted incidence

Description

P. falciparum predicted all-age incidence (clinical cases per 1,000 population per annum) in 2017 for central Africa at 2.5 arcminute (approximately 5km) resolution [1].

Usage

CentralAfrica_Pf_incidence

Format

An data frame with 270083 observations and four variables:

x Longitude in decimal degrees

y Latitude in decimal degrees

z Median predicted incidence at location x y

u Width of the 95% predicted incidence credible interval at location x y

Details

The median and credible interval were computed using samples from a posterior predictive simulation that approximated the joint posterior predictive distribution thereby accounting for spatial covariance [1,2].

Source

These data are available at the Malaria Atlas Project (MAP) website <https://map.ox.ac.uk/>. Specifically, they were obtained by selecting 'ANNUAL MEAN OF PF INCIDENCE' at <https://map.ox.ac.uk/malaria-burden-data-download/>.

References

- [1] Weiss DJ, Lucas TCD, Nguyen M, et al. Mapping the global prevalence, incidence, and mortality of Plasmodium falciparum, 2000–17: a spatial and temporal modelling study. *Lancet* 2019.
- [2] Gething PW, Patil AP, and Hay SI. Quantifying aggregated uncertainty in Plasmodium falciparum malaria prevalence and populations at risk via efficient space-time geostatistical joint simulation. *PLoS Computational Biology* 2010.

Examples

```
str(CentralAfrica_Pf_incidence)
head(CentralAfrica_Pf_incidence)
```

CentralAfrica_shp	<i>Shape files for central Africa</i>
-------------------	---------------------------------------

Description

An object of class `SpatialPolygonsDataFrame` from the R package `sp` v1.3-1 (see reference) containing shape file data for central Africa.

Usage

```
CentralAfrica_shp
```

Format

An object of class `SpatialPolygonsDataFrame` with 20 rows and 16 columns.

Source

Obtained using `malariaAtlas::getShp`; see https://github.com/artaylor85/pixelate/blob/master/data-raw/get_shape_files.R.

References

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

See Also

<https://www.rdocumentation.org/packages/sp/versions/1.3-1/topics/SpatialPolygonsDataFrame-class>

 pixelate

Pixelate as per average uncertainty

Description

Pixelate spatially continuous predictions according to the uncertainty that surrounds them.

Usage

```
pixelate(
  obs_df,
  num_bigk_pix = c(15, 15),
  bigk = 6,
  scale = "imult",
  scale_factor = 1,
  square_pix = TRUE
)
```

Arguments

obs_df	Data frame. Contains a row per observation with four variables: longitude, x; latitude, y; prediction, z; and uncertainty measure u.
num_bigk_pix	Integer vector length two. Specifies a lower bound on the number of complete large pixels (pixels of the bigk-th size) in the x and y directions i.e. pixelate try to fit at least num_bigk_pix[1] large pixels in the x direction and at least num_bigk_pix[2] large pixels in the y direction.
bigk	Integer. Specifies the number of average uncertainty quantile intervals and thus different pixel sizes.
scale	Character equal to either "imult" or "iexpn". Specifies whether to scale pixel sizes (in units of observations) from class k = 3,...,bigk by iterative multiplication or iterative exponentiation (see Details).
scale_factor	Integer. Specifies a factor (in units of observations) that features in either iterative multiplication or iterative exponentiation (see Details)
square_pix	A logical value indicating whether pixels are square or not (in which case they are rectangular).

Details

This is a wrapper function which, given a data frame of observations and several arguments, pixelates as follows.

Let a single observation denote a set containing a prediction, its coordinates, and its uncertainty represented by a single value, e.g. 95% credible interval width. Let a pixel refer to a square or rectangle comprising one or more observations and thus predictions. By default, pixels are square.

Uncertainties are averaged over a limited number of large pixels (pixels of the bigk-th size). We specify a lower bound on the number of large pixels. The function pixelate internally calculates

the smallest number of large pixels greater than or equal to the specified lower bound, while also accounting for other specified arguments. The lower bound can either be an integer or integer vector length two. If a single integer is specified, the number of pixel is calculated relative to the lower bound in the smallest dimension. This is the default. If an integer vector of length two is specified, pixels are rectangular and the number of them is calculated relative to the lower bounds in both directions x and y.

Average uncertainties are classified as high, intermediate (with $bigk-2$ subdivisions), or low, according to the quantile interval they fall into, where the number of quantile intervals is equal to a specified number of different pixel sizes ($k = 1, \dots, bigk$) and the quantiles are based on the empirical distribution of average uncertainties.

The k -th pixel size is defined by a count of observations per pixel (opp) in the x and y direction. We do not specify opps directly; they are calculated internally to best match the specified parameters. Arguments `scale` and `scale_factor` determine the rate at which opps scale. There are two scales, `imult` and `iexpn`. Both scale over $k = 3, \dots, bigk$ for $bigk > 2$, because $opp_1 = 1$ always, and opp_2 is calculated internally to best match the specified parameters. `imult` specifies scaling by iterative multiplication (i.e. a geometric series):

$$opp_k = opp_2 * (2 * scale_factor)^{(bigk - 2)}.$$

. `iexpn` specifies scaling by iterative exponentiation:

$$opp_k = opp_2^{(2 * scale_factor)^{(bigk - 2)}}.$$

The factor 2 is necessary to ensure pixels nest within one another.

If the average uncertainty is high (falls within the top quantile interval), predictions within the large pixel are averaged. If the average uncertainty is intermediate (falls with an intermediate quantile interval), predictions are averaged across smaller pixels nested within the large pixel. If the average uncertainty is low (falls within the bottom quantile interval), predictions are not averaged ($opp_1 = 1$).

Importantly, observations containing missing predictions and predictions that are zero with certainty are excluded from the entire pixelation process (i.e. computation and classification of average uncertainty, and computation of average prediction across large or nested pixel sizes).

Value

`pixelate` returns a list.

`pix_df` The original observation data frame with additional variables: average uncertainty, `u_bigk`; the average uncertainty quantile interval allocation, `bins`; and averaged predictions, `pix_z`.

`pix_df_expanded` A spatially expanded observation data frame with additional variables: the average uncertainty, `u_bigk`; average uncertainty quantile interval allocation, `bins`; and averaged predictions, `pix_z`. All variables besides x and y are NA in spatially expanded observations.

`uncertainty_breaks` The values of average uncertainty at the $bigk+1$ quantiles of the empirical distribution of average uncertainties.

`opp` The observations per pixel (opp) for $k = 1, \dots, bigk$ pixel sizes in the x and y direction.

`obs_df_dim` The dimensions (in units of observations) of the original observation data frame.

`obs_mem` A data frame of observation memberships, where each membership specifies the quantile interval that the large pixel containing the specified observation falls into.

`arguments` The arguments passed to `pixelate` when it was called.

Examples

```

=====
# Use pixelate and inspect its output
=====
# Pixelate using default parameters
px_def <- pixelate(SubSaharanAfrica_Pf_incidence)

# Inspect list returned by pixelate
str(px_def)

# Inspect a sample of uncertain pixelated predictions
uncertain_ind = which(px_def$pix_df$u > 0)
head(px_def$pix_df[uncertain_ind, ])

# Pixelate using alternative parameters
px_alt <- pixelate(SubSaharanAfrica_Pf_incidence,
                  num_bigk_pix = c(25,25), bigk = 5)

# Pixelate as little as possible by allowing
# rectangular pixels and by using only two
# pixels sizes
px_min <- pixelate(SubSaharanAfrica_Pf_incidence,
                  num_bigk_pix = c(2,2), bigk = 2)

# Inspect the observations per pixel
px_min$opp
=====
# Plotting pixelate's output
=====
# Load and attach ggplot2
if (!require("ggplot2")){
  stop("Package ggplot2 needed for the following code. Please install it.")
}

# Define a plotting function
plot_sp_pred <- function(sp_pred){

  ggplot(sp_pred) +

    # Add raster surface
    geom_raster(mapping = aes(x = x, y = y, fill = pix_z)) +

    # Add gradient
    scale_fill_gradientn(name = "Median incidence rate",
                        colors = c("seashell", "tomato", "darkred"),
                        na.value = 'lightblue') +

    # Add axis labels
    ylab('Latitude (degrees)') +
    xlab('Longitude (degrees)') +

    # Ensure the plotting space is not expanded

```

```
coord_fixed(expand = FALSE) +

# Modify the legend and add a plot border:
theme(legend.justification = c(0, 0),
      legend.position = c(0.02, 0.01),
      legend.background = element_rect(fill = NA),
      legend.title = element_text(size = 8),
      legend.text = element_text(size = 8),
      panel.border = element_rect(fill = NA))

}

# Plot
plot_sp_pred(px_def$pix_df)
plot_sp_pred(px_alt$pix_df)
plot_sp_pred(px_min$pix_df)
```

SubSaharanAfrica_Pf_incidence

Plasmodium falciparum predicted incidence

Description

P. falciparum predicted all-age incidence (clinical cases per 1,000 population per annum) in 2017 for sub-Saharan Africa at 2.5 arcminute (approximately 5km) resolution [1].

Usage

```
SubSaharanAfrica_Pf_incidence
```

Format

A data frame with 1794240 observations and four variables:

x Longitude in decimal degrees

y Latitude in decimal degrees

z Median predicted incidence at location x y

u Width of the 95% predicted incidence credible interval at location x y

Details

The median and credible interval were computed using samples from a posterior predictive simulation that approximated the joint posterior predictive distribution thereby accounting for spatial covariance [1,2].

Source

These data are available at the Malaria Atlas Project (MAP) website <https://map.ox.ac.uk/>. Specifically, they were obtained by selecting 'ANNUAL MEAN OF PF INCIDENCE' at <https://map.ox.ac.uk/malaria-burden-data-download/>.

References

- [1] Weiss DJ, Lucas TCD, Nguyen M, et al. Mapping the global prevalence, incidence, and mortality of Plasmodium falciparum, 2000–17: a spatial and temporal modelling study. *Lancet* 2019.
- [2] Gething PW, Patil AP, and Hay SI. Quantifying aggregated uncertainty in Plasmodium falciparum malaria prevalence and populations at risk via efficient space-time geostatistical joint simulation. *PLoS Computational Biology* 2010.

Examples

```
str(SubSaharanAfrica_Pf_incidence)
head(SubSaharanAfrica_Pf_incidence)
```

SubSaharanAfrica_shp *Shape files for sub-Saharan Africa*

Description

An object of class SpatialPolygonsDataFrame from the R package **sp** v1.3-1 (see reference) containing shape file data for sub-Saharan Africa.

Usage

```
SubSaharanAfrica_shp
```

Format

An object of class SpatialPolygonsDataFrame with 55 rows and 16 columns.

Source

Obtained using malariaAtlas::getShp; see https://github.com/artaylor85/pixelate/blob/master/data-raw/get_shape_files.R.

References

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

See Also

<https://www.rdocumentation.org/packages/sp/versions/1.3-1/topics/SpatialPolygonsDataFrame-class>

Index

* datasets

- CentralAfrica_Pf_incidence, [2](#)
- CentralAfrica_shp, [3](#)
- SubSaharanAfrica_Pf_incidence, [7](#)
- SubSaharanAfrica_shp, [8](#)

- CentralAfrica_Pf_incidence, [2](#)
- CentralAfrica_shp, [3](#)

- pixelate, [4](#)

- SubSaharanAfrica_Pf_incidence, [7](#)
- SubSaharanAfrica_shp, [8](#)