

Vizumap: an R package for visualising uncertainty in spatial data

Lydia R. Lucchesi^{1, 2}, Petra M. Kuhnert², and Christopher K. Wikle³

¹ Australian National University, Canberra, Australia ² CSIRO Data61, Canberra, Australia ³ University of Missouri, Columbia, USA

DOI: [10.21105/joss.02409](https://doi.org/10.21105/joss.02409)

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Editor: [Ben Stabler](#) ↗

Reviewers:

- [@nuest](#)
- [@GISerDaiShaoqing](#)

Submitted: 05 June 2020

Published: 22 February 2021

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

Summary

To make a sound data-driven decision, it is necessary to consider the quality and strength of the evidence. Therefore, it is important that the uncertainty of statistical estimates be effectively communicated to decision makers to ensure it is properly considered in the decision-making process. Generally, this uncertainty information is shared through visualisation (e.g., error bars). However, in spatial applications, finding methods that can communicate additional information, about the spatial estimates, in an understandable and meaningful way can be challenging. To address this visualisation shortcoming in spatial statistics, we developed the Vizumap R package. It is a toolkit designed for statisticians, scientists, data journalists, etc., discussing uncertainty in spatial data.

Vizumap contains a series of straightforward functions that can be used to create four different types of maps, which are discussed in detail in [Lucchesi & Wikle \(2017\)](#) and [Kuhnert et al. \(2018\)](#). The visualisation methods include the bivariate choropleth map, pixel map, glyph map, and exceedance probability map. Following is a brief description of each method.

- For the bivariate choropleth map, two colour scales are merged to create a bivariate grid that can encode the estimates and errors at once, making it possible to fill each region with a single colour that represents both values.
- The pixel map is generated by dividing each region into small pixels, which are then filled with a colour value sampled from the colour gradient range. The method for assigning colours to pixels depends on the type of distribution specified. Areas of greater uncertainty appear pixelated while areas of less uncertainty appear smooth, as though filled by only a single colour. These pixel maps can be animated so that areas of high uncertainty flicker while areas of low uncertainty appear static to the eye.
- The glyph map is intended to address the issue of unequal region sizes on a choropleth map. Glyphs of equal size are placed at the centroid of each region, filled with a colour that represents the estimate, and rotated depending on the degree of uncertainty.
- Exceedance probability maps show the probability of exceeding a nominated threshold of concern. These probabilities can be pre-calculated and passed to the exceedance probability map function or calculated within the function given a prescribed probability distribution function, estimate, and error. For example, let X denote a random variable, and x is a possible value of that random variable, X . $F_x(x)$ is the cumulative distribution function and equals the probability that the value of X is less than or equal to a specific value or threshold, x , so

$$F_x(x) = Pr[X \leq x]. \quad (1)$$

The probability that the value X is greater than the specific threshold x can then be written as

$$Pr[X > x] = 1 - F_x(x) = 1 - Pr[X \leq x]. \quad (2)$$

In R, assuming an exponential distribution, this can be evaluated through the following expressions: `1 - pexp(q, rate)` or `pexp(q, rate, lower.tail = FALSE)`.

Other related R packages include `pixelate` and `biscale`. The `pixelate` package explores the use of pixelation for uncertainty visualisation on isopleth maps (Taylor et al., 2020), and the `biscale` package can be used to generate bivariate maps for two variables (Prener et al., 2020).

A comprehensive vignette that demonstrates how to use `Vizumap` is available after package installation. The functions are divided into three categories: formatting, building, and viewing. Formatting functions prepare data for use in the building functions, which are used to build the colour palettes, maps, and map keys. Viewing functions are used to check and combine the different graphical objects designed with the building functions. Previous applications of the `Vizumap` R package include visualising American Community Survey estimates with their corresponding margins of error. We believe `Vizumap` is useful in a wide range of applications, and we will continue to improve the toolkit to enhance its utility.

As an illustration below, we use `Vizumap` to visualise estimated pollutant loads of sediment from the upper Burdekin catchment in Queensland, Australia, into the Great Barrier Reef (GBR). The predictions and uncertainties, published in Kuhnert et al. (2018), were developed from a Bayesian Hierarchical Model (BHM) that assimilated estimates of sediment concentration and flow with modelled output from a catchment model developed on the upper Burdekin catchment. Development of this modelling strategy is discussed in Gladish et al. (2016). Here we just focus on total suspended sediment (TSS). The export of pollutants from coastal catchments within Australia has important implications for the health of the GBR lagoon, and `Vizumap` offers a variety of methods for communicating these predictions and uncertainties to catchment managers and policy makers.

Bivariate map

Figure 1 is a bivariate map of the upper Burdekin catchment in Queensland, Australia. It visualises predicted total suspended sediment (TSS) concentrations and prediction uncertainty. In this example, a custom colour palette is created and then used to build the bivariate map and colour key. The bivariate bins are defined using terciles.

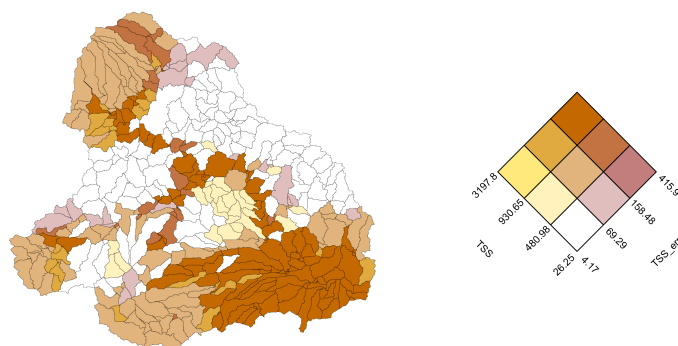


Figure 1: Bivariate map and bivariate key for the upper Burdekin catchment in Queensland, Australia, showing the total suspended sediment (TSS) concentration in mg/L.

Pixel map

Figure 2 visualises the uncertainty of TSS predictions, while giving a general idea of estimated TSS concentrations. For a closer look at the pixelation, a subset of regions is included to the right of the map (Figure 2(B)). The colours filling the pixels in each region were sampled from the estimate's relative frequency distribution. This pixel map can be animated so that the pixels flicker between sampled values. If the map below were to be animated, the areas that appear most pixelated in the static map would, correspondingly, have the most visible movement among pixels in the animated map. Movement among pixels in areas of low uncertainty would be hard to detect due to the minimal differences between the sampled values of orange. An example of an animated pixel map can be found in [Lucchesi & Wikle \(2017\)](#).

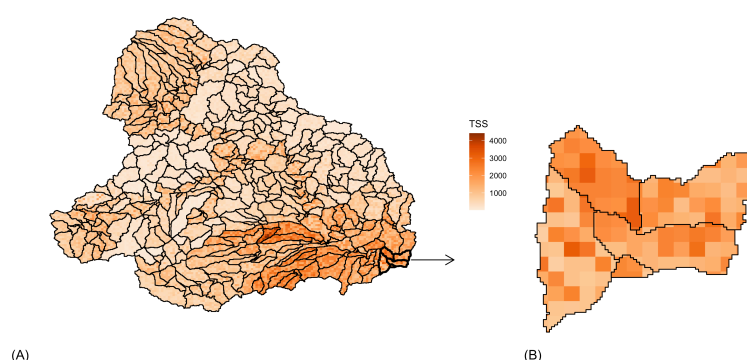


Figure 2: Pixel map showing the TSS concentrations for the upper Burdekin catchment in Queensland, Australia. Figure 2(B) provides a closer look at the pixelation in five regions from Figure 2(A).

Glyph map

Figure 3 depicts estimated TSS concentrations and the uncertainty of these predictions. The colour filling each glyph (located at the region centroid) represents the estimate, and the rotation of the glyph represents the uncertainty.

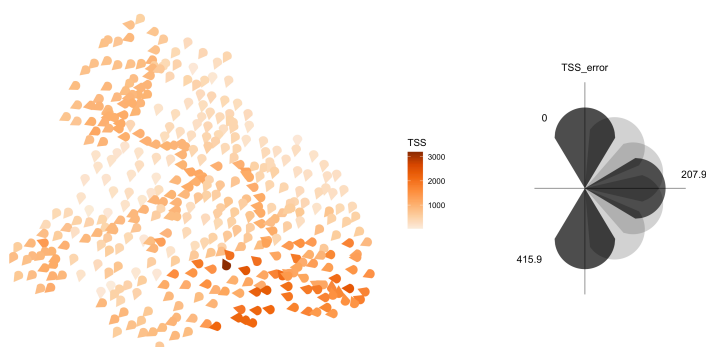


Figure 3: Glyph map and glyph key showing the TSS predictions and uncertainties for the upper Burdekin catchment in Queensland, Australia.

Exceedance probability map

In Figure 4, the calculated probability of exceeding a sediment concentration greater than 837 mg/L (a threshold of concern discussed in Kuhnert et al. (2018)) is plotted in order to draw attention to the high-risk regions on the map. These probabilities were calculated from the posterior distributions of a BHM as outlined in Kuhnert et al. (2018).

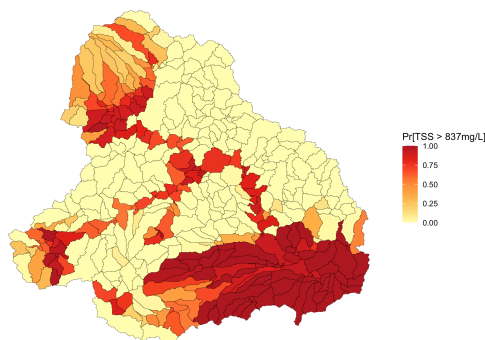


Figure 4: Exceedance probability map showing the probability of exceeding the nominated guideline of 837 mg/L for TSS in the upper Burdekin catchment in Queensland, Australia.

Acknowledgements

We would like to acknowledge the support and funding from the CSIRO Digiscape Future Science Platform for the second author.

Vizumap was built using the following R packages: ggplot2, animation, broom, dplyr, geoaxe, ggmap, grDevices, gridExtra, maps, maptools, plyr, reshape2, rgdal, rgeos, roxygen2, sp, spbabel, testthat, usethis, and utils.

References

- Gladish, D. W., Kuhnert, P. M., Pagendam, D. E., Wikle, C. K., Bartley, R., Searle, R. D., Ellis, R. J., Dougall, C., Turner, R. D. R., Lewis, S. E., Bainbridge, Z. T., & Brodie, J. E. (2016). Spatio-temporal assimilation of modelled catchment loads with monitoring data in the great barrier reef. *The Annals of Applied Statistics*, 10(3), 1590–1618. <https://doi.org/10.1214/16-AOAS950>
- Kuhnert, P., Pagendam, D., Bartley, R., Gladish, D., Lewis, S., & Bainbridge, Z. (2018). Making management decisions in the face of uncertainty: A case study using the burdekin catchment in the great barrier reef. *Marine and Freshwater Research*, 69(8), 1187–1200. <https://doi.org/10.1071/MF17237>
- Lucchesi, L. R., & Wikle, C. K. (2017). Visualizing uncertainty in areal data with bivariate choropleth maps, map pixelation and glyph rotation. *Stat*, 6(1), 292–302. <https://doi.org/10.1002/sta4.150>
- Prener, C., Grossenbacher, T., & Zehr, A. (2020). *Biscale: Tools and palettes for bivariate thematic mapping*. <https://CRAN.R-project.org/package=biscale>
- Taylor, A. R., Watson, J. A., & Buckee, C. O. (2020). *Pixelate: Pixelate spatial predictions as per their average uncertainty*. <https://github.com/artaylor85/pixelate>