



HAL
open science

Statistical cartography: what it is? Version bilingue et commentée

Waldo R Tobler, Françoise Bahoken

► **To cite this version:**

Waldo R Tobler, Françoise Bahoken. Statistical cartography: what it is? Version bilingue et commentée. 1987. hal-04425910

HAL Id: hal-04425910

<https://hal.science/hal-04425910>

Submitted on 30 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Waldo R. Tobler, n.d., Statistical cartography :
what it is ? Note : University of California,
Santa-Barbara, California, 6 p. Poly : 77-81
Version bilingue et commentée

+++++++
TTT
Tribute to **Tobler**
+++++++

Juillet 2022

Version traduite et commentée par Françoise Bahoken (Université Gustave Eiffel), dans le cadre du Projet Tribute to Tobler (Tribut à Tobler), consacré à la valorisation des travaux et apports scientifiques du géographe et cartographe Waldo R. Tobler.

Présentation

Waldo Rudolf Tobler (1930-2018) est un géographe et cartographe américano-suisse formé à l'Université de Washington (Seattle), où il obtient son PhD en 1961. Il y occupe ensuite un poste de d'assistant avant d'intégrer, à partir de 1977, l'Université de Californie à Santa-Barbara comme Professeur de Géographie jusqu'à sa retraite.

Les recherches de Tobler, bien que fondamentalement inscrites en géographie humaine, impliquent toujours des disciplines connexes et transversales, à savoir les mathématiques pour la formalisation et le traitement numérique, l'informatique pour le développement d'outils et la cartographie pour la représentation des résultats obtenus. L'auteur présente ici la « cartographie statistique » ou « cartographie thématique », sachant qu'il s'intéressera par la suite à ce qu'il nomme la « cartographie analytique ».

Les numéros entre crochets dans la version anglaise correspondent à la pagination de la note par l'auteur. Toutes les notes en bleu dans la traduction française sont de la traductrice. Ce qui est en italiques dans le texte et la bibliographie était souligné dans la note originale.

STATISTICAL CARTOGRAPHY : WHAT IT IS ?

Waldo R. Tobler
Professor of Geography
University of California
Santa Barbara, CA 93106

There is a long historical association of statistics and cartography, especially as relates to the theory of the adjustment of observations. Almost all of this history can be evoked by simply mentioning the name of Carl F. Gauss, an inventor of the method of least squares. In this tradition redundant measurements are used to estimate the amount of error contained in empirical observations, and “optimal” estimates are obtained by minimizing the mean square of this error, relative to some model of the phenomena being investigated. The classical theory is applied to the adjustment of surveys but the more modern work, under the name of collocation, also has applicability to interpolation problems, as encountered in the preparation of an isopleth map. These theoretical techniques are widely used in geodesy, but unfortunately are only rarely taught to cartographers or statisticians.

There is also a tradition in which cartography takes the form of graphical illustration of statistical data. Today this is often referred to as “thematic” cartography, sometimes “statistical” cartography. The early roots lie in the work of Playfair, Minard, Quenelet and similar individuals, and are detailed by Funkhouser and several reports of the International Statistical Association, a tradition which continues to this day. By the mid-1800’s choropleth and isopleth maps had been invented, and data were being assembled by rectangular grid cells - a technique lately “rediscovered” in [78] relation to computer assisted cartography. Today thematic cartography is an active area of experimentation, research, and (more recently) psychological testing. One need which I have emphasized for many years is to incorporate our uncertainty into such maps by drawing them in defocused form, fuzzy in proportion to the variance of the data. On a modern CRT one might do this by alternating displays within the refresh cycle. We are now also moving into a period in which color graphics are becoming very inexpensive and it is apparent that experimentation with the tensorial nature of this medium is underway. As our banks of data increase in size the importance of visual summaries can be expected to increase in importance.

In elementary statistics the first descriptive measures learned concern the central tendencies. There are of course the two-dimensional variants of these : the center of gravity, the bivariate median, the point of minimum aggregate travel; and the dispersion measures (the standard distance, bivariate ellipses, Mendeleev’s cartography and its extensions by Bachi, and so on). It is even possible to go further than is usually done, to bivariate

regression, for example - treating the picture of a child's face as a linear (or non-linear) function of the picture of the face of its parents, or regressing a geographical map on its historical precedents. The β -coefficient in such a regression is of course now a tensor rather than a scalar value. We can also apply transformations to the geographical variables. In urban studies logarithmic distances from the center of the city are often used, and cartograms are a form of bivariate uniformization designed to stretch the space so that the effect of some variable, such as the size of the areal units, is eliminated. Unfortunately most useful two dimensional transformation are not separable and require the solution of partial differential equations. Thus they are not easily effected.

The analysis of time series data is often studied only in the more advanced statistical courses. This is mostly because the observations can no longer be considered independent; there is an essential order to the phenomena which the simpler measures (mean, variance, etc.) do not capture. But the geographical case is even more complicated. It would be considered absurd to arrange times series data in *alphabetical* [79] order, by month say, before the analysis. Yet one commonly finds geographical data analyzed in alphabetical order. The spatial dependencies are thus ignored, and *most statistical tests are invalid in such a situation*. It is rare that a statistician applies a geographical "runs" test, or considers geographical adjacency relations when doing a test for the "significance" of some observation, even when the phenomena are as important as the incidence of leukemia. It is necessary to model the spatial spread effects in these data in order to obtain valid inferences. There are also resolution effects which cannot be ignored. When using spatially aggregated data, by county say, we are obtaining a blurred picture of the phenomena being studied. I am continually amazed at how many professional statisticians are unaware of the sampling theorem and its implications for areal measurements. Concommitantly it is obvious that the many techniques which have been developed for the enhancement of spatial data (e.g., edge detection) should be applicable to the bivariate geographical arrangement of health related phenomena. For the comparison of two geographical arrangements (e.g. pollution and cancer) bivariate cross spectral analysis recommends itself, especially as this can now be done using the fast Fourier transform. The interesting and colorful bivariate cross-maps recently introduced by the Bureau of the Census do not seem to me to have quite the sensitivity needed for such important studies. At a more advanced level it is usually appropriate to consider dynamic effects and to go on to time-space series work. Thus this-short essay has only touched on a few facets of the relation between cartography and statistics.

Bibliography

Bennett, R. (1979) *Spatial Time Series*. Cambridge, University Press.

- Cliff, A., et al. (1975) *Elements of Spatial Structure*. Cambridge, University Press.
- Davis, J. and M. McCullagh. (1975) *Display and Analysis of Spatial Data*. New York, J. Wiley. [80]
- Funkhouser, H. (1937) "Historical Development of the Graphical Representation of Statistical Data". *OSIRIS*, vol. 3, pt. 1, pp. 269-405.
- Getis, A. and B. Boots. (1978) *Models of Spatial Processes*. Cambridge, University Press.
- Hagerstrand, T. (1967) *Innovation Diffusion as a Spatial Process*. (Pred Translation), Chicago, University Press.
- Helmert, F.R. (1924) *Die Ausgleichsrechnung nach der Methode der Kleinsten Quadrate*, 3rd ed., Teubner, Berlin.
- Kaula, W. (1967) "Theory of Statistical Analysis of Data Distributed Over a Sphere", *Reviews of Geophysics*, vol. 1, pp. 83-107.
- Larimore, W. (1977) "Statistical Inference on Stationary Random Fields", *Proc. IEEE*, 65, 6, pp. 961-970.
- Mantel, N. (1967) "The Detection of Disease Clustering and a Generalized Regression Approach", *Cancer Research*, 27, 2, pp. 209-220.
- Matheron, G. (1971) *The Theory of Regionalized Variables and its Applications*. Fontainebleau, Ecole Supérieure des Mines.
- Moritz, H. (1970) *Eine Allgemeine Theorie der Verarbeitung von Schwermessungen nach kleinsten Quadraten*. Heft Nr. 67A, Munich, Deutsche Geodätische Kommission.
- Neft, D. (1966) *Statistical Analysis for Areal Distributions*. Philadelphia, Regional Science Assn.
- Robinson, A. (1971) "The Genealogy of the Isopleth", *Cartographic Journal*, pp. 49-53.
- Rosenfeld, A. and A. Kak. (1976) *Digital Picture Processing*. New York, Academic Press.
- Sibert, J. (1975) *Spatial Autocorrelation and the Optimal Prediction of Assessed Values*. Ann Arbor, Department of Geography, University of Michigan. [81]
- Tobler, W. (1969) "Geographical Filters and their Inverses", *Geographical Analysis*, I, 3, pp. 234-253.
- Tobler, W. (1973) "A Continuous Transformation Useful for Districting", *Annals New York Academy of Sciences*, 219, pp. 215-220.
- Tobler, W. (1975) "Linear Operators Applied to Areal Data", in J. Davis and M. McCullagh, *Display and Analysis of Spatial Data*. New York, J. Wiley.
- Tobler, W. (1975) "Mathematical Map Models", *Proceedings*, International Symposium on Computer Aided Cartography, Reston, ACSM, pp. 66-73.
- Tobler, W. (1978) "Comparing Figures by Regression", *Computer Graphics (ACM Siggraph)*, 12, 3, pp. 193-195.

LA CARTOGRAPHIE STATISTIQUE : QU'EST-CE QUE C'EST ?

Waldo R. Tobler
Professor of Geography
University of California
Santa Barbara, CA 93106

Il existe une longue association historique entre les statistiques et la cartographie, notamment en ce qui concerne la théorie de l'ajustement des observations. La quasi-totalité de cette histoire peut être évoquée en mentionnant simplement le nom de Carl F. Gauss, inventeur de la méthode des moindres carrés. Dans cette tradition, des mesures redondantes sont utilisées pour estimer la quantité d'erreur contenue dans les observations empiriques, et des estimations « optimales » sont obtenues en minimisant le carré moyen de cette erreur, par rapport à un certain modèle des phénomènes étudiés. La théorie classique est appliquée à l'ajustement des levés, mais le travail plus moderne, sous le nom de colocalisation, est également applicable aux problèmes d'interpolation, tels que ceux rencontrés dans la préparation d'une carte isoplèthe. Ces techniques théoriques sont largement utilisées en géodésie, mais ne sont malheureusement que rarement enseignées aux cartographes ou aux statisticiens.

Il existe également une tradition dans laquelle la cartographie prend la forme d'une illustration graphique de données statistiques. Aujourd'hui, on parle souvent de cartographie « thématique », parfois de cartographie « statistique ». Les premières racines se trouvent dans le travail de Playfair¹, Minard², Quenelet³ et d'autres personnes similaires, et sont détaillées par Funkhouser⁴ et plusieurs rapports de l'Association Internationale de Statistique, une tradition qui se poursuit à ce jour. Au milieu des années 1800, les cartes choroplèthes et isoplèthes avaient été inventées et les données étaient assemblées par des cellules de grille rectangulaire - une technique récemment « redécouverte » dans le cadre de la cartographie assistée par ordinateur. Aujourd'hui, la cartographie thématique est un domaine actif d'expérimentation, de recherche et (plus récemment) de tests psychologiques. Un besoin

1. William Playfair (1759-1823) était un économiste écossais considéré comme l'inventeur de la statistique graphique pour avoir proposé plusieurs diagrammes issus de données quantitatives.

2. Charles-Joseph Minard (1781-1870) était un ingénieur français des Ponts-et-Chaussées qui consacra sa retraite à la production d'un système de « Cartes figuratives » sur le mouvement des transports.

3. Tobler a vraisemblablement fait une faute de frappe sur le nom de Lambert Adolphe Jacques Quetelet (1796-1874), mathématicien et statisticien franco-belge, précurseur des analyses démographiques et fondateur de plusieurs sociétés savantes et journaux scientifiques.

4. Howard Gray Funkhouser (1898-1984) était un mathématicien américain féru de visualisation de données statistiques, inscrit dans la lignée des travaux de Playfair.

sur lequel j'ai insisté pendant de nombreuses années est d'incorporer notre incertitude dans ces cartes en les dessinant sous une forme dé-focalisée, floue en proportion de la variance des données. Sur un tube cathodique moderne, on peut le faire en alternant les affichages pendant le cycle de rafraîchissement. Nous entrons maintenant dans une période où les graphiques en couleur deviennent très bon marché et il est évident que l'expérimentation de la nature tensorielle de ce support est en cours. Au fur et à mesure que la taille de nos banques de données augmente, on peut s'attendre à ce que l'importance des résumés visuels augmente.

En statistique élémentaire, les premières mesures descriptives apprises concernent les tendances centrales. Il y a bien sûr les variantes bidimensionnelles de celles-ci : le centre de gravité, la médiane bivariée, le point de déplacement minimal de l'agrégat ; et les mesures de dispersion (la distance standard, les ellipses bivariées, la cartographie de Mendeleïev et ses extensions par Bachi, et ainsi de suite). Il est même possible d'aller plus loin que ce qui est fait habituellement, jusqu'à la régression bivariée, par exemple en traitant l'image du visage d'un enfant comme une fonction linéaire (ou non linéaire) de l'image du visage de ses parents ou en faisant régresser une carte géographique sur ses précédents historiques. Le coefficient β d'une telle régression est bien sûr maintenant un vecteur plutôt qu'une valeur scalaire. Nous pouvons également appliquer des transformations aux variables géographiques. Dans les études urbaines, les distances logarithmiques par rapport au centre de la ville sont souvent utilisées et les cartogrammes sont une forme d'uniformisation bivariée conçue pour étirer l'espace de manière à éliminer l'effet d'une certaine variable, telle que la taille des unités aréolaires. Malheureusement, la plupart des transformations bidimensionnelles utiles ne sont pas séparables et nécessitent la résolution d'équations différentielles partielles. Elles ne sont donc pas faciles à réaliser.

L'analyse des séries chronologiques n'est souvent étudiée que dans les cours de statistique les plus avancés. Cela est principalement dû au fait que les observations ne peuvent plus être considérées comme indépendantes ; il existe un ordre essentiel dans les phénomènes que les mesures les plus simples (moyenne, variance, etc.) ne permettent pas de saisir. Mais le cas géographique est encore plus compliqué. Il serait considéré comme absurde de classer les données de séries chronologiques par ordre *alphabétique*, de mois par exemple, avant l'analyse. Pourtant, on trouve couramment des données géographiques analysées par ordre alphabétique. Les dépendances spatiales sont ainsi ignorées et *la plupart des tests statistiques ne sont pas valables dans une telle situation*. Il est rare qu'un statisticien applique un test de « séries » géographiques ou tienne compte des relations de contiguïté géographique lorsqu'il effectue un test de « significativité » d'une observation, même lorsque les phénomènes sont aussi importants que l'incidence de la leucémie. Il est nécessaire de modéliser les effets de dispersion spatiale dans

ces données afin d'obtenir des inférences valides. Il existe également des effets de résolution qui ne peuvent être ignorés. Lorsque l'on utilise des données spatialement agrégées, par comté disons, on obtient une image floue des phénomènes étudiés. Je suis toujours étonné de voir combien de statisticiens professionnels ignorent le théorème de l'échantillonnage et ses implications pour les mesures aréolaires. De la même façon, il est évident que les nombreuses techniques qui ont été développées pour l'amélioration des données spatiales (par exemple, la détection des contours) devraient être applicables à la répartition géographique bivariée des phénomènes liés à la santé. Pour la comparaison de deux arrangements géographiques (par exemple, la pollution et le cancer), l'analyse spectrale croisée bivariée s'impose, d'autant plus qu'elle peut désormais être réalisée à l'aide de la transformation de Fourier rapide. Les cartes croisées bivariées intéressantes et colorées récemment introduites par le Bureau du recensement ne me semblent pas avoir la sensibilité nécessaire pour des études aussi importantes. À un niveau plus avancé, il convient généralement de considérer les effets dynamiques et de passer à des travaux sur les séries spatio-temporelles. Ainsi, ce court essai n'a abordé que quelques facettes de la relation entre la cartographie et les statistiques.

Bibliographie

- Bennett, R. (1979) *Spatial Time Series*. Cambridge, University Press.
- Cliff, A., et al. (1975) *Elements of Spatial Structure*. Cambridge, University Press.
- Davis, J. and M. McCullagh. (1975) *Display and Analysis of Spatial Data*. New York, J. Wiley.
- Funkhouser, H. (1937) « Historical Development of the Graphical Representation of Statistical Data ». *OSIRIS*, vol. 3, pt. 1, pp. 269-405.
- Getis, A. and B. Boots. (1978) *Models of Spatial Processes*. Cambridge, University Press.
- Hagerstrand, T. (1967) *Innovation Diffusion as a Spatial Process*. (Pred Translation), Chicago, University Press.
- Helmert, F.R. (1924) *Die Ausgleichsrechnung nach der Methode der Kleinsten Quadrate*, 3rd ed., Teubner, Berlin.
- Kaula, W. (1967) « Theory of Statistical Analysis of Data Distributed Over a Sphere », *Reviews of Geophysics*, vol. 1, pp. 83-107.
- Larimore, W. (1977) « Statistical Inference on Stationary Random Fields », *Proc. IEEE*, 65, 6, pp. 961-970.
- Mantel, N. (1967) « The Detection of Disease Clustering and a Generalized Regression Approach », *Cancer Research*, 27, 2, pp. 209-220.
- Matheron, G. (1971) *The Theory of Regionalized Variables and its Applications*. Fontainebleau, Ecole Supérieure des Mines.
- Moritz, H. (1970) *Eine Allgemeine Theorie der Verarbeitung von Schwermessungen nach kleinsten Quadraten*. Heft Nr. 67A, Munich, Deutsche Geodatische Kommission.

- Neft, D. (1966) *Statistical Analysis for Areal Distributions*. Philadelphia, Regional Science Assn.
- Robinson, A. (1971) « The Genealogy of the Isopleth », *Cartographic Journal*, pp. 49-53.
- Rosenfeld, A. and A. Kak. (1976) *Digital Picture Processing*. New York, Academic Press.
- Sibert, J. (1975) *Spatial Autocorrelation and the Optimal Prediction of Asses-
sed Values*. Ann Arbor, Department of Geography, University of Michigan.
- Tobler, W. (1969) « Geographical Filters and their Inverses », *Geographical
Analysis*, I, 3, pp. 234-253.
- Tobler, W. (1973) « A Continuous Transformation Useful for Districting », *Annals
New York Academy of Sciences*, 219, pp. 215-220.
- Tobler, W. (1975) « Linear Operators Applied to Areal Data », in J. Davis and M. McCullagh, *Display and Analysis of Spatial Data*. New York, J. Wiley.
- Tobler, W. (1975) « Mathematical Map Models », *Proceedings*, International Symposium on Computer Aided Cartography, Reston, ACSM, pp. 66-73.
- Tobler, W. (1978) « Comparing Figures by Regression », *Computer Graphics (ACM Siggraph)*, 12, 3, pp. 193-195.



La collection « [traductions](#) » du groupe *Tribute to Tobler* propose des rééditions bilingues et commentés d'articles, publiés ou inédits, de Waldo Tobler. Responsable scientifique : Françoise Bahoken (Université Gustave Eiffel) ; éditeur : Laurent Beauguitte (UMR Géographie-cités).

Disponibles en ligne

- F. Bahoken, 2022, « [W. R. Tobler, 1969, Review of Sémiologie graphique: Les Diagrammes – Les réseaux – Les Cartes](#) ».
- F. Bahoken, 2022, « W. R. Tobler, nd, Statistical Cartography : what is it ? ».