

Article

## An online calculator for spatial data and its applications

**Kalle Remm, Tiiu Kelviste**

Institute of Ecology and Earth Sciences, University of Tartu, 46 Vanemuise St., 51014 Tartu, Estonia

E-mail: kalle.remm@ut.ee

*Received 6 December 2013; Accepted 10 January 2014; Published online 1 March 2014*



### Abstract

An online calculator (<http://digiarhiiv.ut.ee/kalkulaator/>) for statistical analysis of spatial data is introduced. The calculator is applicable in a wide range of spatial research and for courses involving spatial data analysis. The present version of the calculator contains 35 web pages for statistical functions with several options and settings. The input data for most functions are pure Cartesian coordinates and variable values, which should be copied to the input cell on the page of a particular spatial operation. The source code for the computational part of all functions is freely available in C# programming language. Examples are given for thinning spatially dense observation points to a predefined minimum distance, for calculating spatial autocorrelations, for creating habitat suitability maps and for generalising movement data into spatio-temporal clusters.

**Keywords** spatial statistics; online tool; habitat suitability; autocorrelation; spatio-temporal clustering.

<p><b>Computational Ecology and Software</b> ISSN 2220-721X URL: <a href="http://www.iaees.org/publications/journals/ces/online-version.asp">http://www.iaees.org/publications/journals/ces/online-version.asp</a> RSS: <a href="http://www.iaees.org/publications/journals/ces/rss.xml">http://www.iaees.org/publications/journals/ces/rss.xml</a> E-mail: <a href="mailto:ces@iaees.org">ces@iaees.org</a> Editor-in-Chief: Wenjun Zhang Publisher: International Academy of Ecology and Environmental Sciences</p>
---

### 1 Introduction

Statistical problems in ecology, earth and environmental sciences, human geography, and other fields of science and technology are often related to location or distance between observations. Spatial statistics deals with functions that involve location in one way or another. For example, calculating mean precipitation level at weather stations is not spatial statistics, as the location of stations is not involved in the summarising. Finding out how the difference in precipitation values is related to the distance between stations is spatial analysis since location is directly involved.

Scientific studies can be divided into descriptive (exploratory) and inferential (confirmatory) approach. Exploratory spatial statistics deals with finding generalised values, relationships, spatial patterns, spatial clustering and segmentation (Zhang, 2010). The first goal of inferential spatial analysis is usually to prove the non-randomness of the pattern, followed by attempts to find possible reasons and to model the processes which created the non-randomness.

Spatial statistical functions are included into major commercial GIS (Geographical Information Science) software packages. Tools for spatial statistical analysis in freeware packages like QGIS are much more limited.

There are ample online calculators for mathematical functions and ordinary statistical tests (see, e.g. <http://www.martindalecenter.com/Calculators2.html>), but an online calculator for spatial analysis is hard to find. Existing online calculators for spatial data mostly transform geodetic coordinates only.

The goal of this publication is to introduce the spatial data calculator at <http://digiarhiiv.ut.ee/kalkulaator/> that is applicable in a wide range of spatial research and supports several courses on data management and programming at the University of Tartu. The idea was formed, the original user interface designed and the coding and testing project mainly developed by the first author. From the technological side, the calculator was developed as Microsoft ASP.NET project, functions written in the C# programming language. The calculator was created primarily for teaching and learning purposes, although has already been used in research (Kotta et al., 2013). Computations for the following examples are made using the online calculator.

## 2 Main Characteristics of the Calculator

The calculator currently involves a home page and 35 web pages for statistical functions; 28 of these are directly for spatial analysis. Most pages have several options to set the initial parameters for calculations. The calculator does not demand any client-side installation other than a web browser. Input data and the parameters set by the user are transferred to the server, which responds with computation results to the client's browser. The calculator has no extra demands on the memory or processing unit of the client's device except for data transfer and browsing ability. The amount of input data and the choice of functions are restricted, as online applications are expected corresponding within a reasonable period of time. The restrictions depend on function. Generally, for larger data sets and more complicated tasks, special GIS and/or statistical software should be used instead.

Data input to most functions must be pure Cartesian coordinates and variable values. The input cell of a web page in the calculator can be filled with example data by clicking the "Example data" button.

The measurement units of output distances and input parameters are the same as for coordinates, e.g. if the coordinates are given in metres, then distance intervals and search radii must also be in metres. The decimal separator must be a point, and the column separator a space, tab or semicolon. Empty cells are not allowed – these should be filled or removed during data mounting. It is easy to prepare the input matrix as a spreadsheet (Excel, Access) and to copy the values without column headers to the input cell of the online calculator. The introductory text at every statistical function, as well as example data, indicate the necessary format of input data. The point data can be in one data set or alternatively as separate samples of source locations and destination points. In the last case, distances are measured only between two types of points. The web pages of the calculator functions also include buttons for viewing a scatter plot of input data in a pop-up window, a button that initiates calculation, and a button that opens the source code window (Fig. 1).

The meaning of X and Y coordinates as input columns can be switched with regard to which is the west-east direction and which is south-north (Fig. 2). The X axis of example data in the calculator is always directed to the east. For the European cartographic system, where the first coordinate axis is directed to the north, the north direction must be set to the first position (select option "N direction first").

Suitability surface as an Idrisi unpacked headerless rst format raster can be added in several functions. It enables the user, for example, to delimit the location of generated random points, and to define an irregularly shaped, uneven and/or patchy study surface.

The location of input data points can be visualised in a pop-up window as an XY scatterplot or a bubble chart, if the data points have values. Area borders set by the user determine the extent of the scatterplot, the borders of the suitability surface (if used), and delimit random locations for the null model. A map background from the Web Map Service (WMS) server of the Estonian Land Board can be added to scatterplots of source

data and to output maps, given the coordinates are acceptable for this WMS service.

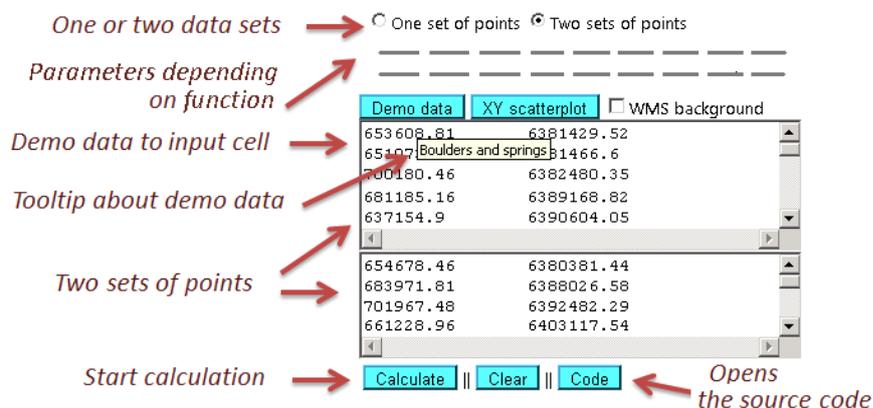


Fig. 1 Input cell of the calculator and the main buttons around it.

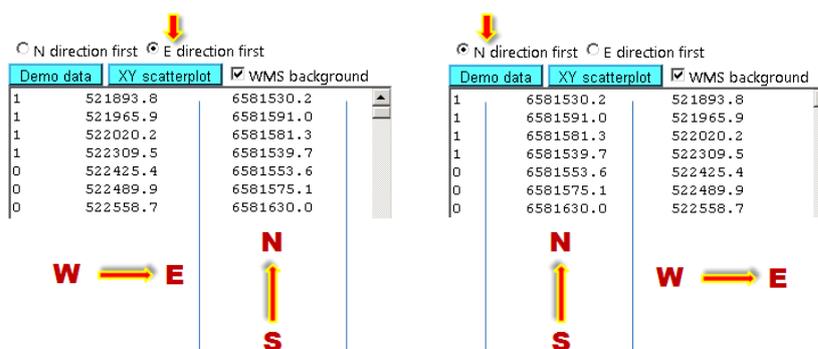


Fig. 2 Alternative meaning of X and Y coordinates in input data.

Pressing the button labelled "Calculate" initiates calculation on the server. Every click on the "Calculate" button is counted to see the usage intensity and dynamics of every function. The calculation result is output as a tab-delimited text that can be easily imported to other software solutions. The results include the time spent on calculations on the server. The time spent on transmission between the server and the client's computer is to be added to the pure calculation time.

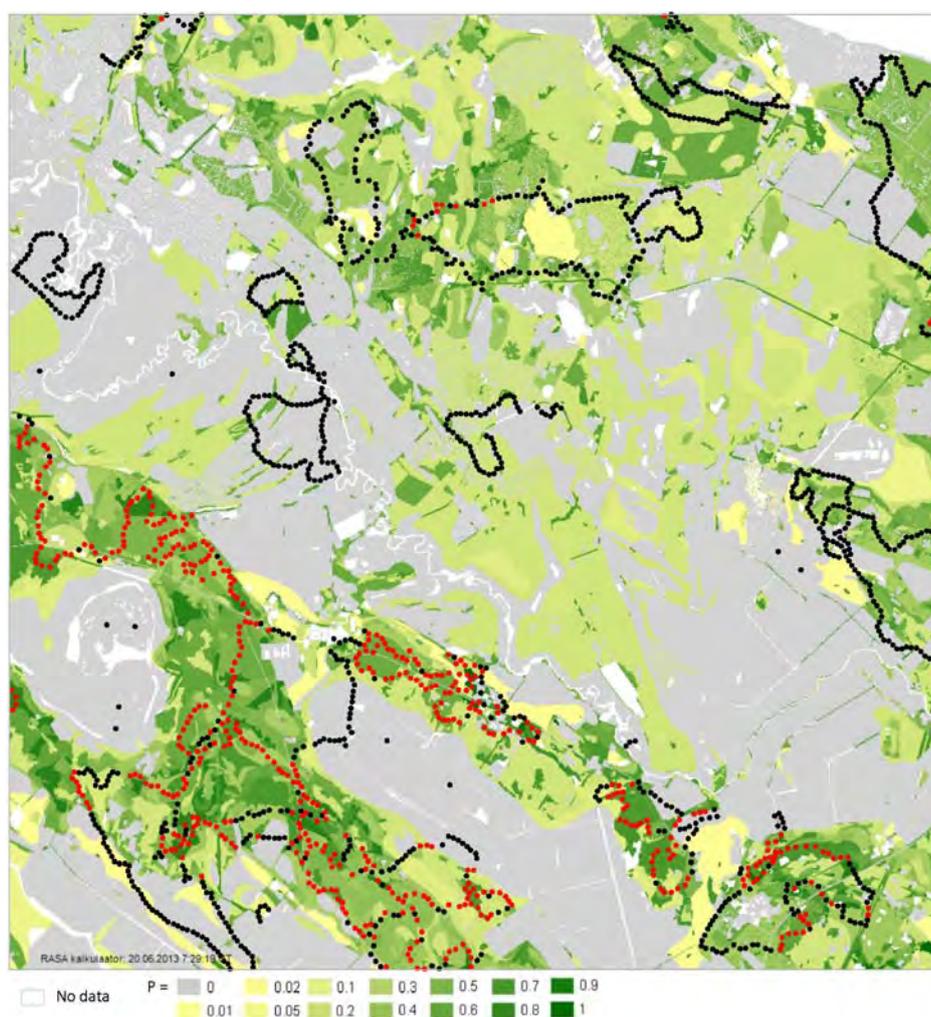
The source code for the computational part of the calculator web pages is public – advanced users can see in detail how the result is calculated. The authors do not charge a fee for the use of this online calculator, and do not take responsibility for any incorrect or unexpected results obtained by using this calculator. The calculator is still in development, the code is continuously improved according to the developers' skill, and additional functions will be added if useful in teaching or for research. Users can send comments and proposals for advancing this application directly from the home page.

### 3 Application Examples

#### 3.1 Habitat suitability of *Potentilla fruticosa* and the patchiness of found sites

*Potentilla fruticosa* is as a flowering shrub mainly known as an ornamental cultivar. *P. fruticosa* subsp. *fruticosa* grows naturally in a few locations in Sweden, Great Britain, Estonia and Latvia (Elkington, 1969; Leht and Reier, 1999).

For the following example, we used points on GPS-recorded field observation tracks and found locations of this species from a study area of 10 km × 10 km square (sheet 6382 of the Estonian 1: 20 000 base map) in the middle of the natural distribution area of *P. fruticosa* in north-west Estonia. Coordinates of observation tracks and *P. fruticosa* observation locations were recorded using a Garmin Vista HC+ GPS recorder during field trips on foot by the first author in summers 2008–2013. The initial data contained 12,299 track points and 1469 observation locations. Firstly, all automatically recorded track points closer than 50 m from active observation points were removed using the online calculator's thinning function. Then, both recorded track points and observation locations were thinned to have at least 50 m between each accepted site. The thinned data contained 499 find sites and 1122 absent sites (332 actively and 790 automatically recorded) (Fig. 3).



**Fig. 3** *Potentilla fruticosa* observation sites within map sheet 6382 of the Estonian 1: 20 000 base map. Black dots – absent sites, red dots – find sites, background green intensity – habitat suitability representing the probability to find *P. fruticosa* according to soil and land use category.

The find sites of *P. fruticosa* are spatially clustered, lying on the tracked observation routes and aggregating at more suitable habitats. *P. fruticosa* occurrence may also be spatially clustered because of the limited dispersal ability of the plant. To assess the effect of habitat on the apparent spatial clustering of the *P. fruticosa* find sites, we need a habitat suitability map covering the study area.

The online calculator offers several options for suitability mapping according to landscape characteristics

by calculating: 1) combinations of presence and absence frequency; 2) probability of occurrence, and 3) expected presence or absence according to similarity of sites using either the  $k$  nearest neighbours ( $kNN$ ),  $d$  nearest neighbours ( $dNN$ ) or *sumsim* algorithm. For the  $kNN$  algorithm, the number of the most similar exemplars included has to be fixed;  $dNN$  algorithm sets a limit on the acceptable similarity level; the *sumsim* algorithm, which is applied in the Constud software system (Remm and Kelviste, 2011; Remm and Remm, 2008), needs the sum of similarity of the most similar accepted exemplars to be given.

For this study, habitat suitability of *P. fruticosa* was calculated using probability mapping within a map sheet. The resulting map indicates a higher probability to find the plant in alvar grasslands, where the land cover category is either bush land or so-called other (unmanaged) open land, and the favourite soil categories are Gleisoiil, Endogleyic Luvisol, and Rendzic Leptosol. The bushes in these alvar grasslands are mainly junipers.

The second part of this example is analysing the spatial pattern of *P. fruticosa* find sites. A spatial pattern of points: its clustering, regularity or randomness can be described in a generalised form using several formal statistics presented in the online calculator: the nearest neighbour distances, distribution of all distances between points, the mean squared distance,  $K(t)$ ,  $L(t)-t$ ,  $O(r)$ ,  $G(r)$ ,  $F(r)$ , and  $J(r)$  functions. Clustering of *P. fruticosa* find sites relative to total area, to estimated habitat suitability, to observation sites, and to suitability weighted observation sites according to  $K(t)$ ,  $L(t)-t$  and  $O(r)$  statistics with 95% confidence envelope, calculated using the online calculator, is presented here as an example.

The  $K(t)$  statistic was introduced by B.D. Ripley (Ripley, 1976, 1977, 1981) as the mean number of neighbouring objects within radius  $t$  from a source point divided by the mean density of objects. Letters  $d$  and  $r$  often stand for radius in the  $K(t)$  function instead of Ripley's original sign,  $t$ . In the following,  $r$  is preferred to make the notation of radius uniform. The expected number of randomly located neighbours within radius  $r$  is  $\lambda K(r)$ , where  $\lambda$  is the mean density of objects. The  $K(r)$  statistic is widely accepted in science, since it does not depend on the density of points (He and Duncan, 2000).

Ripley's  $K(r)$  is often transformed to  $L(r)-r$  statistic (1), for which the expected value in case of spatial randomness is zero, not depending on the radius.

$$L(r) - r = \sqrt{\frac{K(r)}{\pi}} - r \quad (1)$$

The  $K(r)$  and  $L(r)-r$  statistics are cumulative functions of radius, contrary to differential statistics measuring the frequency of neighbours in distance intervals. The cumulative function is more stable, but pattern properties close to the source affect values of the statistic at larger distances. Differential functions are calculated separately for every distance zone and the result becomes unstable as the number of objects per interval diminishes.

The function characterising neighbour density according to distance is called radial distribution in physics. In ecology it has different names: pair correlation (Law et al., 2009), O-ring statistic (Wiegand et al., 1999; Wiegand and Moloney, 2004), relative neighbourhood density (Condit et al., 2000), and neighbour density distribution (Remm and Luud, 2003). O-ring statistic and neighbour density is not normalised by the mean density; pair correlation and radial distribution function are normalised to have the expected density equal to one. All point pattern statistics that use distances between objects can be used to characterise a pattern of uniform point objects or for describing relationship (spatial association, segregation) between different category point objects. In the last case, source and target objects are different.

In case of ecological data, habitat properties can seldom be considered equal across the study area. More

suitable regions have more resources per surface unit and support a higher density of living objects. In order to estimate relative location preferences of objects not caused by unequal distribution of resources, the density of objects should be counted not relative to geometrical area, but to the suitability field (Malanson, 1985; Remm et al., 2006).

A common issue affecting both observed and predicted values of neighbourhood statistics is the edge effect – no target objects are counted outside the study area, although distance zones around source objects closer to the boundary than the counting radius are partly outside the study area. Therefore, the local density of neighbours may be underestimated near boundaries of the study area. The algorithm used in the online calculator includes an edge effect correction option for  $K(r)$ ,  $L(r)-r$  and  $O(r)$  functions. Instead of using area within a radius, the area of arbitrary grid units (pixels) within the study area is summed. The grid units have suitability values if the suitability field is included; otherwise, all pixels have the same value. When computing boundary and suitability corrected  $L(r)-r$  statistic in the calculator, the explicit radius  $r$  in the right side of the formula (1) is replaced by a variable derived from the suitability corrected area ( $A$ ) within the boundaries of the study plot and within the given radius  $r$  (2).

$$L(r)-r = \sqrt{\frac{K(r)}{\pi}} - \sqrt{\frac{A(r)}{\pi}} \quad (2)$$

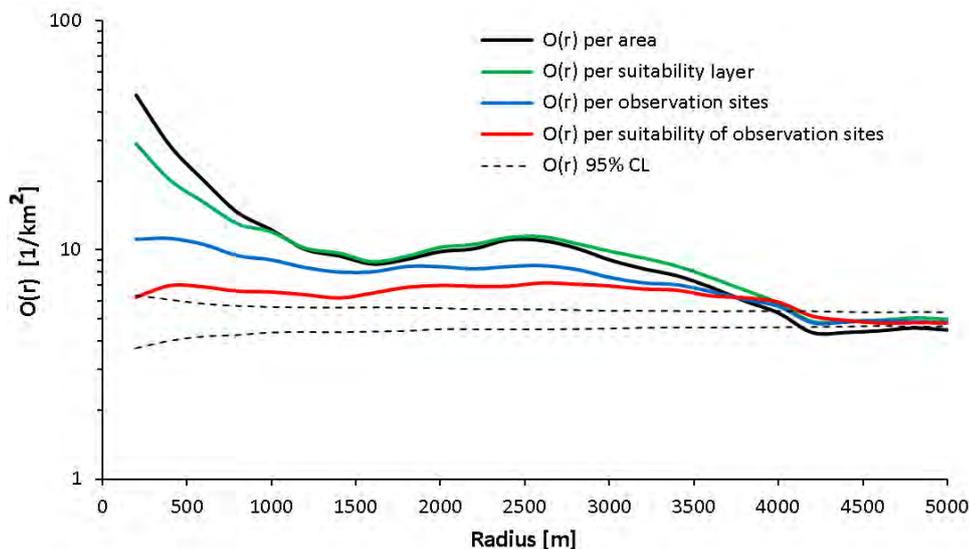
To describe statistical properties of the spatial pattern of *P. fruticosa* find locations,  $K(r)$ ,  $L(r)-r$  and  $O(r)$  functions were calculated within 200-m-wide intervals up to 5 km distance using the online calculator. These functions were calculated relative to untransformed area, relative to the mapped habitat suitability, relative to observation sites and relative to suitability weighted observation sites to reveal the effect of habitat suitability and uneven location of observations. Suitability field was calculated based on proportion of finds in combinations of land use and soil categories as described above (Fig. 3). A raster layer where pixels of observation sites were assigned a value of one and all other study area of zero was used to calculate the functions relative to observed sites. To count suitability at observed sites, the observed pixels received suitability value from the suitability coverage; all other pixels remained equal to zero and were excluded from the calculation of zone area. The 95% confidence envelope for the null model of random location of neighbours was obtained from 1000 iterations.

The density of neighbours calculated per suitability of observation tracks should remove the effect of habitat patchiness, clustering of observations retaining the effect of the species' limited spreading ability, and probably some uncontrolled effects as well. Relating the  $O(r)$  statistic only to observed sites is justified, since the data on presence/absence of the species at unobserved sites is not known and therefore should not affect the results. The density of neighbouring finds should be much higher if the density of observations was higher. The area divisor available in the calculator for correcting spatial units and densities was used to set the total study area to 100 units, regardless of the sum of suitability values and the area of observed sites.

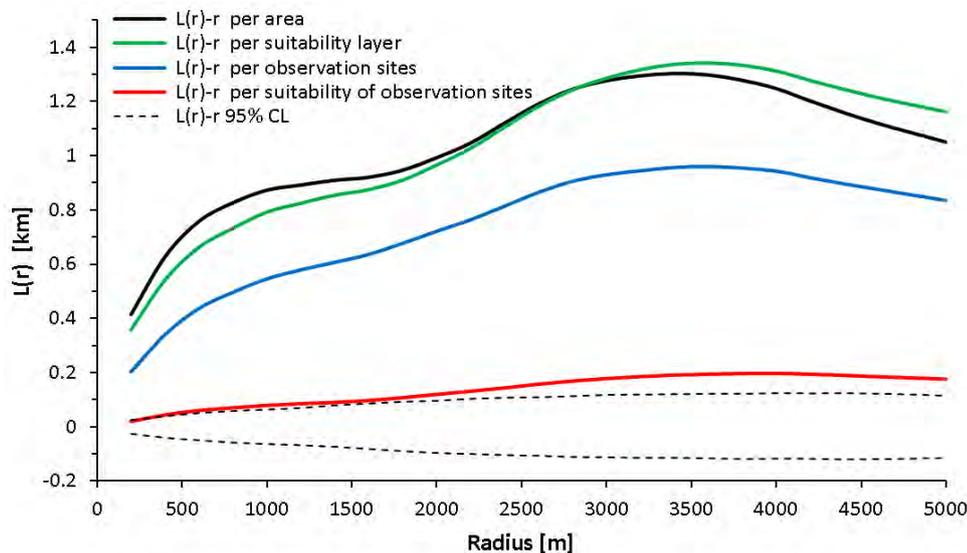
As expected, the  $O(r)$  indicates how clustering depends on distance in more detail than the cumulative  $L(r)-r$  function (Fig. 4, 5). *P. fruticosa* find sites are remarkably spatially clustered, with the density of neighbours about 20 times higher within 200 m from a find site than the average density of finds in the study area. This clustering is partly caused by the general patchiness of the habitat suitability in the study area and by clustering of observation sites on moving tracks. The effect of habitat suitability in the neighbourhood is evident only at close distances up to about 700 m, where per-suitability calculated density of neighbours (Fig. 4, green line) is less than per-equal-properties-area calculated density of neighbours (Fig. 4, black line).

The clustering of observations has a more evident effect on clustering of find locations. The density of

neighbours calculated relative to observation sites (Fig. 4, blue line) and per suitability of observation sites is higher than expected in case of randomness up to 4 km of distance (Fig. 4, red line). The  $O(r)$  statistic calculated per suitability weighted observation sites is relatively stable up to this distance, indicating the existence of larger patches of this species in the study area, although most of the visible clustering is caused by patchiness of the observation sites and habitat properties.



**Fig. 4** Neighbour density function [ $O(r)$ ] of *Potentilla fruticosa* find sites relative to the edge-corrected area within the given radius, relative to habitat suitability corrected area, relative to observation sites, and relative to suitability-weighted observation sites. Dashed lines indicate 95% confidence limits for the null model of random location for neighbouring sites within a suitable area. The vertical axis has a logarithmic scale.



**Fig. 5** The  $L(r)-r$  function of *Potentilla fruticosa* find sites relative to the edge-corrected area within the given radius, relative to habitat suitability corrected area, relative to observation sites, and relative to suitability-weighted observation sites. Dashed lines indicate 95% confidence limits for the null model of random location for neighbouring sites within a suitable area.

### 3.2 Spatial autocorrelation of precipitation in the Baltic states

Spatial relationships depend on the location of data points and can be related to the distance and/or direction between observations but also to region and scale. Autocorrelation indicates the relationship between values of

the same variable. It can be both positive, meaning a higher distance-related similarity than expected from random placement of values, and negative, meaning dissimilarity compared to spatial randomness. As an analogue to spatial relationships, temporal autocorrelation and correlation depend on difference and direction in time.

As a rule, close observations tend to be similar because they are affected by the same spatially continuous factors. If spatial phenomena are recorded, interrelated close data points add less new knowledge than an independent observation. Measuring autocorrelation is essential to estimate the effect of spatial continuity and to estimate at which distance the observations can be considered to be independent, as classical statistical methods presume independence of observations. Ordinary statistical tests overestimate the degrees of freedom in the spatially autocorrelated data, yielding a higher probability of type one statistical errors, i.e. false meaningful conclusions (Legendre, 1993; Malanson, 1985). Neglecting the effect of autocorrelation when comparing the explanatory variables leads to overestimation of positively autocorrelated factors (Lennon, 2000). An overview of species and habitat distribution models including spatial relationships is given in (Miller et al., 2007).

The online spatial data calculator offers calculation of classical correlation coefficients between numerical variables, both Pearson linear correlation and Spearman rank correlation. For autocorrelation, the present version of the calculator includes two options for the Moran's  $I$  statistic. The Moran's  $I$  is a quotient expressing the ratio between autocovariation and variance of data values with a predefined relative location (distance and/or direction) (Moran, 1950). The two options are: 1) a general Moran's  $I$  where the global variance is used for standardising; and 2)  $I(d)$  that is calculated separately for distance zones and standardised by the variance of data values belonging to the same distance interval only. The formula for Moran's  $I$  (3) differs from the algorithm for Pearson  $R$  by including relative location (usually distance) dependent weights.

$$I = \frac{N \sum_{i=1}^N \sum_{j=1}^N w_{ij} (z_i - \bar{z})(z_j - \bar{z})}{\sum_{i \neq j} w_{ij} \sum_{i=1}^N (z_i - \bar{z})^2}, \quad (3)$$

where  $N$  is the number of observations,  $i$  and  $j$  are site indices of the members of observation pairs ( $i \neq j$ ),  $w_{ij}$  is the weight depending on distance between sites  $i$  and  $j$  ( $w_{ij} = 1$  if the distance between sites  $i$  and  $j$  is within a given value, otherwise  $w_{ij} = 0$ );  $\bar{z}$  is the mean of  $z$  values.

The expected value  $E(I)$  in case of spatial randomness of data values approaches zero if the number of observations is large:

$$E(I) = \frac{-1}{N-1}. \quad (4)$$

The Moran's  $I$  assumes spatial stability of the variance (homoscedasticity), otherwise the global variance does not represent the variance at a particular distance interval and the range of  $I$  values exceeds  $-1 \dots +1$ . In case of calculating  $I(d)$  separately for each distance interval, the variance among values not belonging to the interval does not affect the result, and therefore the range of Moran's  $I(d)$  (5) is within  $-1 \dots +1$ .

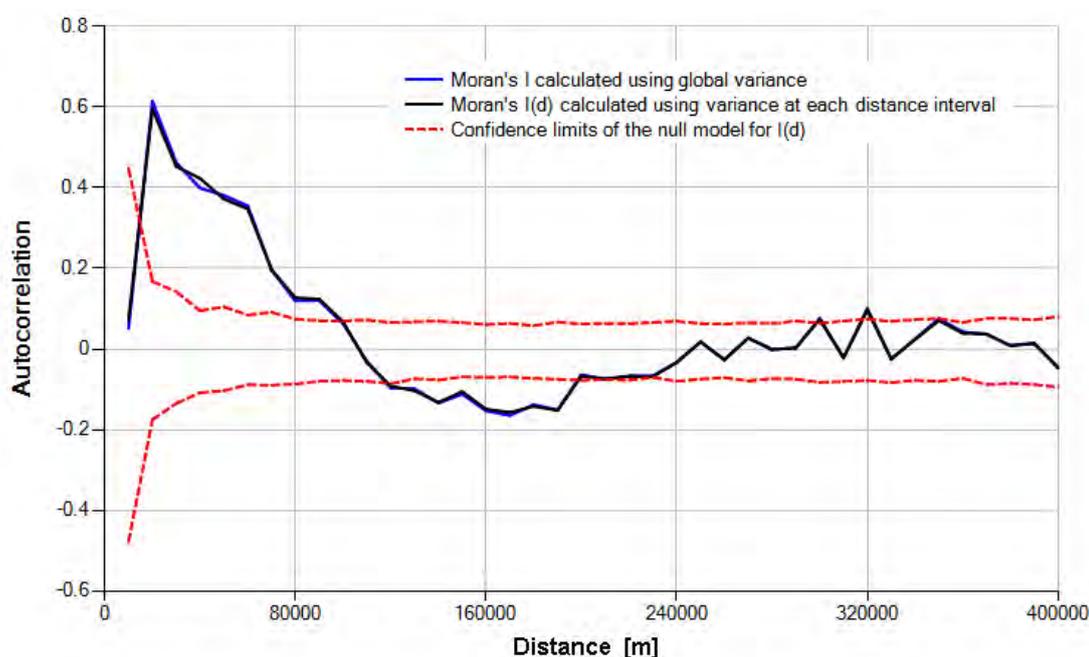
$$I(d) = \frac{N(d) \sum_{i=1}^N \sum_{j=1}^N (z_i - \bar{z})(z_j - \bar{z})}{W(d) \sum_{i=1}^N (z_i - \bar{z})^2}, \quad (5)$$

where  $N(d)$  is the number of observation points satisfying the distance condition  $d$  and  $W(d)$  is the number of observation pairs at distance  $d$ .

The online calculator has options to calculate both cumulative and discrete spatial (auto)correlation. In the first case,  $d$  means a radius, in the second case a distance interval. An analysis in distance intervals gives a more detailed picture of the relationship, but may yield unstable and noisy results if the number of observations per interval is not sufficient. The dependence of (auto)correlation on distance is graphically depicted as an (auto)correlogram (Fig. 6).

Long-term mean annual precipitation values at 245 meteorological stations in the Baltic states (Estonia, Latvia and Lithuania) are used in the following autocorrelation examples. More details on these data can be found in (Jaagus et al., 2010; Remm et al., 2011).

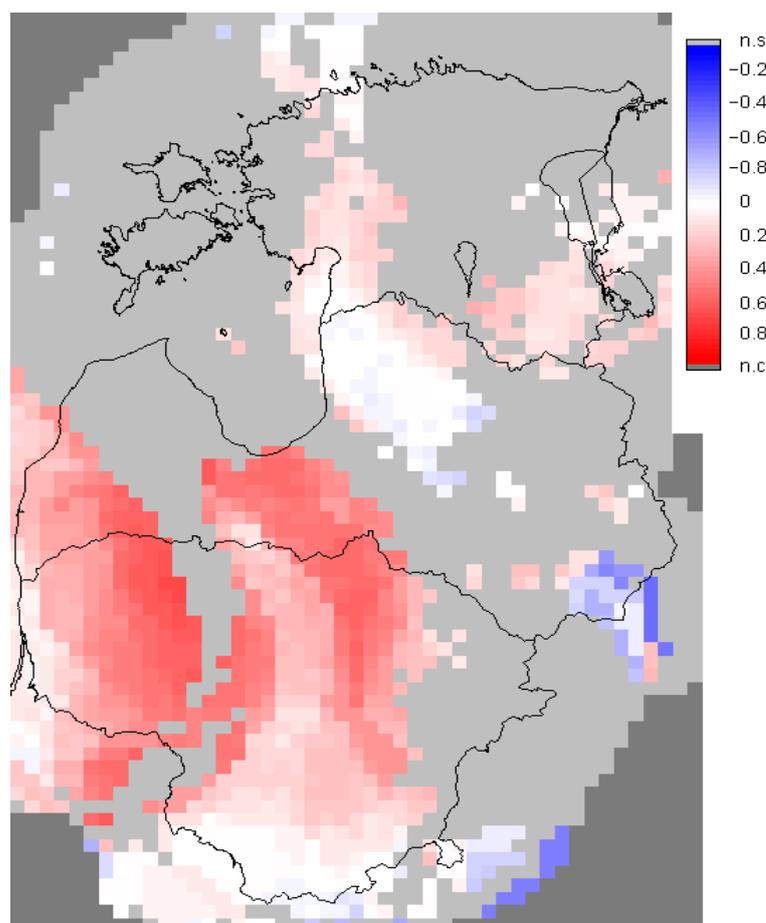
The spatial autocorrelation of the annual amount of precipitation is significantly positive up to a distance of 90 km between stations (Fig. 6). This can be considered the typical extent of the maritime rainy coastal belt, especially along the coast of Lithuania and western Latvia, and the more continental part of the study area. Spatial autocorrelation of the mean annual precipitation is negative at distances of 120–190 km. Autocorrelation is statistically insignificant ( $p > 0.05$ ) at a distance of 10 or fewer kilometres, since the number of station pairs that are this close is only 16 in these data.



**Fig. 6** Distance-dependent autocorrelation of the long-term mean annual precipitation in the Baltic countries, along with the 95% probability envelopes for the null model (red dashed lines). Data from Remm et al. (2011).

The strength of spatial relationships depends on location and can be mapped if calculated in multiple locations. The online calculator enables users to calculate local spatial autocorrelation and correlation at nodes

of a regular grid covering the entire study area. In addition to the interval between grid nodes, local search radius around grid nodes and the distance between included observations must be set in the calculator. The local spatial autocorrelation map of the long-term mean annual precipitation highlights regions in western Lithuania and Latvia where similar values are spatially aggregated, that is are located closer than expected from randomness (Fig. 7). Here, the contrast between maritime and continental climate is most expressed. Negative autocorrelation values in eastern Latvia are related to a small number of stations attended by relatively variable measured mean annual precipitation amount in neighbouring stations.



**Fig. 7** Statistically significant local spatial autocorrelation values of the long-term mean annual precipitation in the Baltic countries. Interval between grid points of 10 km, search radius around grid points of 100 km, and maximum distance between included stations of 50 km. Red marks positive local Moran's  $I$  values, and blue – negative;  $n. s-p > 0.05$ ,  $n. c$  – not calculated. Data from Remm et al. (2011).

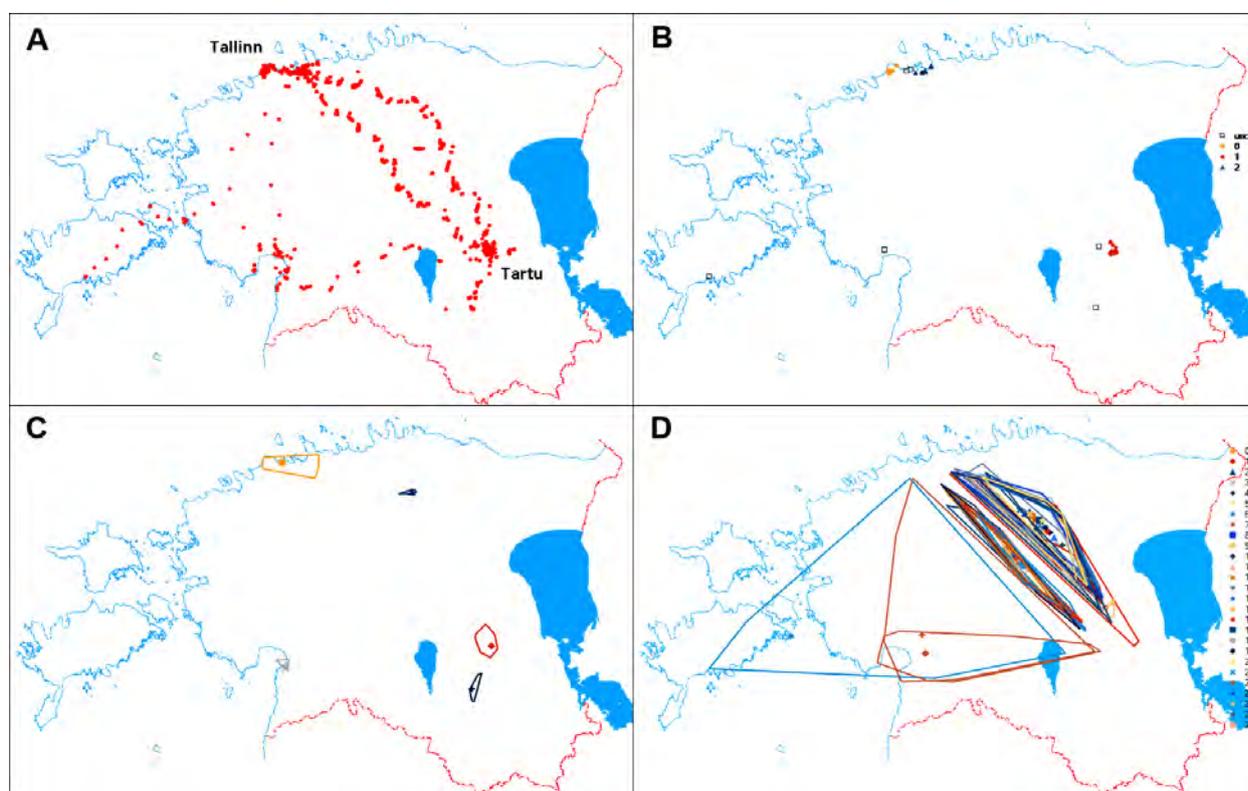
### 3.3 Thinning and spatial clusters of movement data

Currently, movement data are of major interest in many domains, e.g. GPS tracking of wildlife, pets and vehicles, studies of human migration and tourist movement patterns. The raw data for a movement study are usually tracking data consisting of millions of location coordinates, which are senseless without some sort of generalisation and summarisation. To distinguish different aspects of movement (coverage, anchor points, temporal pattern, etc.) several mathematical operations are needed. The choice of specific method largely depends on the research subject and aim.

The present version of the online calculator includes  $k$ -means clustering,  $QT$  (Quality Threshold) and

*DBSCAN* (Density-Based Spatial Clustering for Applications with Noise) clustering. The *k*-means algorithm demands a prefixed number of clusters, which is not desirable while identifying recurring places from movement data, as the number of favourite places is *a priori* not known. The *QT* algorithm needs only a user-decided search radius to delimit the extent of clusters. The input parameters for *DBSCAN* clustering are the search radius and the minimum number of objects in a cluster, but not the given number of clusters. The level of generalisation of the resulting clusters can easily and flexibly be controlled by the two parameters. A special advantage of the *DBSCAN* algorithm is the acceptance of noise in input – clusters highlight regions where the density of points is higher, leaving some points and some parts of the study area un-clustered.

To illustrate the online calculator functions, the active mobile phone positioning data of a volunteer living in Tartu, Estonia, was used for one year (January 2, 2010 to December 31, 2010). Active positioning data in this case means that the location of the user's mobile phone was registered continuously after every 15 minutes. The average positioning error is expected to be about 3 km. The number of raw tracking points was 32,169 (Fig. 8A).



**Fig. 8** A – Location of track points in Estonia, B – Main overnight locations of the study object, C – On-stay DBSCAN clusters depicted by a convex hull polygon and a dot at the cluster centre, D – DBSCAN clusters of on-move track points (transport connections between Tartu and Tallinn: the railway is at the north-east, motorway in the centre).

Traditionally, the movement studies involve only two spatial dimensions (Zhou et al., 2007), although movement is related to time. Therefore, several functions in the online calculator include a spatio-temporal option where the input includes *X* and *Y* coordinates, plus *Z* dimension, which can be time in the general Date Time format. For example, thinning a set of raw spatio-temporal data starting at 3 a.m. to a one-day clearance interval enables places to be extracted where the object was most frequently located at 3 a.m.; that is, where he/she most frequently overnights. Thinning of the example data resulted in 339 retained separate night locations. Spatial clustering of these points by the *DBSCAN* module, using a search radius of 6000 m and

cluster size of 30 resulted in three frequent overnight places for this person: at home in Tartu, somewhere in the capital, Tallinn, and one about 20 km west of Tallinn (Fig. 8B). The other locations can be considered occasional.

A second exercise was separating two kinds of tracking points and generalising these into two types of clusters: on-move clusters that reflect the trajectories of movement, and on-stay clusters that indicate locations where the study object has spent more time. To get on-stay clusters, the search radius along XY plane should be delimited to a shorter value (6000 m in this case) and the time difference should be at least as long as the study period. The minimum number of objects in a cluster was set to 30. Most of the track points were classified as on-stay category at this stage. The rest (515) were considered on-move points, which were generalised into clusters using a 12-hour maximum time difference and not limiting search radius. The minimum number of objects in an on-move cluster was set to 5, since in some cases only 5 track points were recorded during Tallinn–Tartu train travel.

As a result, five on-stay clusters and two on-way clusters were generalised for this person. The largest on-stay cluster is around the Tartu residence; the on-way clusters represent one-day travel routes (49 times between the hometown and capital, Tallinn), one tour near Tartu, and three longer trips to western Estonia (Fig. 8C, D).

#### 4 Conclusions

The online calculator introduced above supports a wide range of spatial analysis operations, although, the application is still merely a calculator primarily developed for teaching purposes. For larger datasets and more comprehensive studies, special software is needed or should be developed. The code for all functions used in the online calculator is freely open for assessment, criticism, and development of other applications.

The presented examples of spatial analysis were all calculated using the online calculator. The example cases yielded the following results.

- (1) Spatial analysis of *Potentilla fruticosa* find locations reveals that the main reasons for the obvious clustering of finds are the patchiness of suitable habitats and clustering of observation sites. Although the seeds of this plant are not spreading far, the inherent dispersal-related patchiness has a weaker effect than soil properties and land use type combined.
- (2) Spatial autocorrelation of the annual amount of precipitation at the Baltic meteorological stations is significantly positive at a distance of 10–90 km between stations.
- (3) Spatio-temporal clustering of movement data supports insights into large sets of tracking data, generalisation and visualisation of movement.

#### Acknowledgments

The authors thank Markus Unt for participating in code development, colleagues Jaanus Remm for developing an earlier solution for directional analysis and Ain Kull for advice supporting the present module of directional analysis in the calculator; and also the proof reader Dirk Lloyd. The development of this calculator was financially supported by the Estonian Ministry of Education and Research (Project IUT 2-16).

#### References

- Condit R, Ashton PS, Baker P, et al. 2000. Spatial patterns in the distribution of tropical tree species. *Science*, 288(5470): 1414-1418
- Elkington TT. 1969. Cytotaxonomic variation in *Potentilla fruticosa* L. *New Phytologist*, 68(1): 151-160

- He F, Duncan RP. 2000. Density-dependent effects on tree survival in an old-growth Douglas fir forest. *Journal of Ecology*, 88(4): 676-688
- Jaagus J, Briede A, Rimkus E, Remm K. 2010. Precipitation pattern in the Baltic countries under the influence of large-scale atmospheric circulation and local landscape factors. *International Journal of Climatology*, 30(5): 705-720
- Kotta J, Remm K, Vahtmäe E, et al. 2013. In-air spectral signatures of the Baltic Sea macrophytes and their statistical separability. *Journal of Applied Remote Sensing* (submitted manuscript).
- Law R, Illian J, Burslem DFRP, et al. 2009. Ecological information from spatial patterns of plants: insights from point process theory. *Journal of Ecology*, 97(4): 616-628
- Legendre P. 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology*, 74(6): 1659-1673
- Leht M, Reier Ü. 1999. Origin, chromosome number and reproduction biology of *Potentilla fruticosa* (Rosaceae) in Estonia and Latvia. In: *Chorological Problems in the European Flora*. (Uotila P, ed). Proceedings of the VIII meeting of the Committee for Mapping the Flora of Europe. 191-196, Helsinki, Finland
- Lennon JJ. 2000. Red-shifts and red herrings in geographical ecology. *Ecography*, 23(1): 101-113
- Malanson GP. 1985. Spatial autocorrelation and distributions of plant species on environmental gradients. *Oikos*, 45(2): 278-280
- Miller J, Franklin J, Aspinall R. 2007. Incorporating spatial dependence in predictive vegetation models. *Ecological Modelling*, 202(3-4): 225-242
- Moran PAP. 1950. Notes on continuous stochastic phenomena. *Biometrika*, 37(1-2): 17-23
- Remm J, Lõhmus A, Remm K. 2006. Tree cavities in riverine forests: what determines their occurrence and use by hole-nesting passerines? *Forest Ecology and Management*, 221(1-3): 267-277
- Remm K, Jaagus J, Briede A, et al. 2011. Interpolative mapping of mean precipitation in the Baltic countries by using landscape characteristics. *Estonian Journal of Earth Sciences*, 60(3): 172-190
- Remm K, Kelviste T. 2011. Constud Tutorial. University of Tartu, Chair of Geoinformatics and Cartography, Tartu, Estonia
- Remm K, Luud A. 2003. Regression and point pattern models of moose distribution in relation to habitat distribution and human influence in Ida-Viru county, Estonia. *Journal for Nature Conservation*, 11(3): 197-211
- Remm M, Remm K. 2008. Case-based estimation of the risk of enterobiasis. *Artificial Intelligence in Medicine*, 43(3): 167-177
- Ripley BD. 1976. The second-order analysis of stationary point processes. *Journal of Applied Probability*, 13(2): 225-266
- Ripley BD. 1977. Modelling spatial patterns. *Journal of the Royal Statistical Society, B* 39(2): 172-212
- Ripley BD. 1981. *Spatial Statistics*. Wiley, New York, USA
- Wiegand T, Moloney KA. 2004. Rings, circles and null-models for point pattern analysis in ecology. *Oikos*, 104(2): 209-229
- Wiegand T, Moloney KA, Naves J, Knauer F. 1999. Finding the missing link between landscape structure and population dynamics: a spatially explicit perspective. *The American Naturalist*, 154(6): 605-627
- Zhang WJ. 2010. *Computational Ecology: Artificial Neural Networks and Their Applications*. World Scientific, Singapore
- Zhou C, Bhatnagar N, Shekhar S, et al. 2007. Mining Personally Important Places from GPS Tracks. *Data Engineering Workshop, IEEE 23rd International Conference*, 517-526