STATISTICS FOR PATCH OBSERVATIONS

K. L. Hingee\textsuperscript{a,b,∗}

\textsuperscript{a}CSIRO, Underwood Ave, Floreat, Perth, Australia
\textsuperscript{b}School of Mathematics & Statistics, University of Western Australia, Stirling Highway, Perth, Australia - kassel.hingee@research.uwa.edu.au

ABSTRACT:

In the application of remote sensing it is common to investigate processes that generate patches of material. This is especially true when using categorical land cover or land use maps. Here we view some existing tools, landscape pattern indices (LPI), as non-parametric estimators of random closed sets (RACS). This RACS framework enables LPIs to be studied rigorously. A RACS is any random process that generates a closed set, which encompasses any processes that result in binary (two-class) land cover maps. RACS theory, and methods in the underlying field of stochastic geometry, are particularly well suited to high-resolution remote sensing where objects extend across tens of pixels, and the shapes and orientations of patches are symptomatic of underlying processes. For some LPI this field already contains variance information and border correction techniques. After introducing RACS theory we discuss the core area LPI in detail. It is closely related to the spherical contact distribution leading to conditional variants, a new version of contagion, variance information and multiple border-corrected estimators. We demonstrate some of these findings on high resolution tree canopy data.

1. INTRODUCTION

Statistical analysis of images can be grouped into two main branches (Molchanov, 1997): (a) describing/classifying an observed scene or (b) considering the scene to be generated by a random process and inferring properties of this process. We are concerned mostly with the latter, and especially those processes observed in remotely sensed maps of categorical variables. Such analysis occurs when comparing different regions, comparing the same region at different times, gaining understanding of random processes (e.g. spatial dependence), or model fitting.

A random closed set (RACS) is a generic framework for modelling randomness in processes that generate spatial patterns of patches. It encompasses common models in remote sensing, such as those derived from Markov random fields and Gaussian random fields, and a wide range of other models (e.g. germ-grain models, birth-growth models, fibre processes, and tessellations). Markov random fields excel at contextual investigations, but have difficulty describing geometrical properties (Descombes, 2012). Gaussian random fields, completely determined by their mean and covariance (Chiu et al., 2013), can only capture first and second order characteristics of a process. Other RACS models can describe complex geometrical shapes and infinite-order characteristics (Descombes, 2012). These models reveal new methods for describing the geometry of random scenes such as contact distributions (Baddeley and Gill, 1994) and tangent point analysis (Barbour and Schmidt, 2001).

These tools have been combined with remote sensing to model uncertain object boundaries (Zhao et al., 2009), meteorological features (Cressie et al., 2012), fine-scale grass patterns (Sadler, 2006) and urban tree locations (Rossi et al., 2015). The RACS framework has also been used to test the spatial dependence of beetle infestation (Kautz et al., 2011) and the relationship between tree deaths and bore locations (Chang et al., 2013). Numerous other examples exist (Descombes, 2012).

We examine parallels between non-parametric RACS summary functions and landscape pattern indices (LPI). To the author’s knowledge this is the first published discussion on this topic. A RACS framework leads to more rigorous treatment of landscape pattern data and improved understanding of LPI behaviour, including border corrections, resolution robustness and conditional probabilities. In the next section we introduce RACS theory in more detail, with a focus on concepts relevant to remote sensing applications. Subsequently Section 3 discusses RACS and LPI. Finally Section 4 describes an application of non-parametric RACS estimators to maps of tree canopy.

2. RANDOM CLOSED SETS AND REMOTE SENSING

For brevity we omit many technical details and the full generality of RACS. These details can be found in a number of texts (Molchanov, 2005; Chiu et al., 2013). A RACS is any random process that generates a closed set. A closed set in Euclidean space is any set for which all points on the edge of the set are also in the set. A RACS, \( X \), is then any process that maps from some state space \( \Omega \) to closed subsets of Euclidean space

\[
P(X(\omega) \text{is closed}) = 1. \tag{1}
\]

This is similar to the familiar definition of a random variable, except each realisation is a closed subset of space. Figure 1 contains two example realisations of RACS. A RACS
produces categorical maps if its realisations are closed 2D regions. This definition allows only two classes, inside $X$ or not inside $X$, but can be generalised to multiclass realisations (Molchanov, 1984; Ayala and Simó, 1995; Kautz et al., 2011).

As an example the pattern of mould on the surface of an old slice of bread can be modelled as a RACS. The mould pattern may have some systematic properties (e.g. patch size) but the precise location, size and shape of the mould is unpredictable. For this RACS a single realisation is the pattern of mould on a single slice of bread.

The mathematical field underlying the statistical investigation of random scenes is known as Stochastic Geometry (Molchanov, 1997). It is the study of statistical methods for geometrical patterns (Matheron, 1975; Chiu et al., 2013). Historically an important application has been the inference of 2D or 3D properties (e.g. properties of rocks) from lower dimensional samples (Baddeley and Jensen, 2004). This application area, known as stereology, has developed tools for inference from 0D samples (e.g. pixels) and from images (2D samples).

Often it is the case in remote sensing of the Earth that only one realisation of the process is available. For example the generation of a native forest, including all its past disturbances, is usually seen only once at each location. We can mitigate this issue by assuming that the statistical properties of the process are similar at different locations and that the dependence between distant locations is small. Thus observations of different regions act like multiple realisations. These sort of assumptions are common in spatial analysis.

The strongest similarity assumption is stationarity which assumes the statistical properties of a process are translation invariant. In other words the probability of any closed, bounded set $K$ intersecting $X$, written $P(K \cap X \neq \emptyset)$, is independent of translations of $K$; the probability depends only on the shape and size of $K$. As an example consider the centre of the mouldy bread slice at high magnification so that we may ignore the border of the bread and assume that the process is stationary. This stationarity assumption implies that, before looking at the bread, the probability that any particular location is mouldy $P(x \in X)$, known as the coverage probability, is the same for all $x$. It also means that the probability that a circular region contains mould depends only on the radius, which leads to the spherical contact distribution (SCD) discussed later.

These examples correspond to intersecting with a set $\{x\}$ and a disc respectively. The stationary assumption has been common in landscape ecology (Fortin et al., 2003).

3. AN APPLICATION TO LANDSCAPE PATTERN INDICES

Landscape ecologists study both the effect of spatial pattern on ecological processes wherein spatial pattern is considered a covariate and the effect of ecological processes on spatial pattern wherein the spatial pattern is considered a response or symptom of the ecological process (Fortin and Agarwal, 2005). Frequently this requires comparing patterns observed in remotely sensed maps of land cover type.

LPIs are numerical descriptions of spatial configuration that are commonly used in landscape ecology (Lustig et al., 2015). Current LPIs have proved sensitive to resolution and boundary effects (Kupfer, 2012). Many LPIs are also difficult to interpret (Schröder and Seppelt, 2006) and highly correlated with other LPIs (Cushman et al., 2008; Schindler et al., 2008; Turner, 2005). There have been numerous calls for more rigorous statistical interpretation of LPIs (Lustig et al., 2015; Dramstad, 2009; Wang and Cumming, 2011) and some calls for more process-based metrics (Fortin et al., 2003; Remmel and Csillag, 2003). RACS provide a generic probabilistic framework that allows statistical interpretation, process-based descriptions and rigorous study of LPI behaviour.

The most popular software package for the calculation of LPIs is FRAGSTATS (McGarigal, 2015) and many other packages incorporate LPIs that are conceptually derived...
from FRAGSTATS (VanDerWal et al., 2015). Within the collection of metrics available in FRAGSTATS we have noticed that:

1. The percentage of landscape index is an estimator of coverage probability
2. The edge density index is a (potentially biased) estimator of an identically named concept in RACS theory.
3. The percentage of core area is an estimator of the core probability which is related to the spherical contact distribution for a RACS.
4. The description of the radius of gyration given by (Keitt et al., 1997) is related to the mean star of intersection of a RACS (Molchanov, 1997).
5. Variants of the contagion index can be constructed that express second-order and geometric properties of RACS.

A crucial contribution here is that LPIs can be viewed as estimators of underlying random processes. To the author’s knowledge this is the first time non-parametric RACS summary functions have been explicitly correlated with LPI concepts. Kautz et al. (Kautz et al., 2011) provide an example of the power of non-parametric RACS very similar to LPI without directly referring to LPIs. Otherwise previous use of RACS in relation to LPIs have been for simulation studies (Hargis et al., 1998) and model fitting (Diggle, 1981; Sadler, 2006).

The remainder of this section discusses the core area, the closely related SCD, and a new contagion index that uses the SCD. The new SCD-based contagion is less resolution dependent than the classic pixel-adjacency contagion. For core area, the RACS perspective provides variance information, border corrections and natural conditioning on events. We will also make use of the coverage probability which is easily estimated by the percentage of landscape. The details for other indices will be discussed in a forthcoming paper.

In the following suppose that X is a stationary, mixing RACS process, and that we have observed a single realisation, $X_{obs}$, in a window $W$.

### 3.1 Percentage of Core Area

For a user chosen buffer distance, $r$, core area is the area within a class that is more than $r$ distance from the edge of the class (McGarigal, 2015; Didham and Ewers, 2012). The percentage of core area in a window is an estimate of the probability that a point will be further than $r$ distance from the exterior of $X$. This is also the probability of placing a disc of radius $r$ entirely within $X$.

\[
\text{Percentage of core area} \approx P(B_r(o) \subseteq X), \tag{2}
\]

where $B_r(o)$ is a disc of radius $r$ about the origin, $o$. The origin is used here arbitrarily because the stationarity of $X$ requires that the probability is the same regardless of the disc’s centre. Thus the analogous concept of core area for a random process could be described as Core Probability and is the probability that a point will be further than $r$ distance from the exterior of $X$. In the next section (Section 3.1.1) we show that core probability is closely related to the SCD.

### 3.1.1 Core Probability and the SCD

The SCD (also known as the empty space function) is a popular tool for exploration and inference of random point processes (Baddeley et al., 2015) and sets (Diggle, 1981; Molchanov, 1997; Heinrich, 1993). The unconditional version is the probability of $X$ intersecting an arbitrarily located disc of radius $r$,

\[
SCD_X^{unc}(r) = P(B_r(o) \cap X \neq \emptyset). \tag{3}
\]

where $\emptyset$ is the empty set so $\emptyset \neq \emptyset$ can be read as ‘not empty’ and we have arbitrarily used the origin as the centre of the disc.

The conditional version of the SCD is the probability of $X$ intersecting a disc given that the centre of the disc is not in $X$,

\[
SCD_X^{cond}(r) = P(B_r(o) \cap X \neq \emptyset | o \notin X). \tag{4}
\]

If $o \in X$ then $X$ intersects the disc so the two versions are related by

\[
SCD_X^{cond}(r) = 1 - \frac{1 - SCD_X^{unc}(r)}{1 - p}, \tag{5}
\]

where $p$ is the coverage probability. Recall that the coverage probability is the probability of an arbitrary point being in $X$, it can be estimated by the percentage of $W$ covered by $X_{obs}$ (Baddeley and Jensen, 2004).

The SCD describes the sizes of space outside $X$; a RACS with a large conditional SCD at radius $r$ is less likely to contain gaps in which a disc of radius $r$ can fit.

The space that is not interior to $X$ is also a RACS, we denote it by $X^c$. The superscript ‘c’ denotes the set of locations not in $X$ and the overline represents the inclusion of the edges. If $X$ is stationary then so is $X^c$ and thus the SCD of $X^c$ is well defined. Furthermore the unconditional SCD of $X^c$ is the probability that an arbitrarily located disc of radius $r$ intersects $X^c$ which is the negation of the core probability. In other words

\[
SCD_{X^c}^{unc}(r) = P(B_r(o) \cap X^c \neq \emptyset) = P(B_r(o) \notin X) \tag{6}
\]

$= 1 - $ core probability.

The conditional version is the probability that a point in $X$ is within distance $r$ of the outside of $X$

\[
SCD_{X^c}^{cond}(r) = P(B_r(o) \cap X^c \neq \emptyset | o \notin X^c). \tag{7}
\]

which suggests a conditional core probability (CCP), the probability that a point in $X$ is in the core of $X$, the exterior of $X$.

\[
\text{conditional core probability} = 1 - \frac{SCD_{X^c}^{cond}(r)}{1 - p} = P(B_r(0) \subseteq X | o \in X). \tag{8}
\]

Thus the non-parametric properties of the unconditional and conditional SCD are identical to those of the core probability and the CCP respectively.

### 3.1.2 Border Correction

Estimation of the core probability using percentage of core area risks making an implicit assumption about $X$ outside the observation window. See Figure 2. There are points which appear further than $r$ distance from the edge of $X_{obs}$ but may not be in the core of $X$. The same is true for the SCD. The effect is...
larger with larger buffer distance because the area within r distance of the boundary is a larger proportion of the window area. We mention three different border corrected estimators for the unconditional SCD that are available in the spatstat package within the R statistical computing environment (Baddeley et al., 2015). For brevity we describe them only in terms of the core probability. Despite theoretical differences the three estimators perform similarly well in simulation experiments for a variety of random patterns of points (Stoyan, 2006; Baddeley et al., 2015), which suggests that they will also perform with similar quality on RACS. Note that the CCP can be estimated from the ratio of a core probability estimate and a coverage probability estimate.

The reduced sample estimate uses only those points further than r distance from the window’s boundary (Heinrich, 1993). See Figure 2. Due to the different sample size for each r the estimated core probability can increase as the buffer distance increases, however an increase in the true core probability is not possible because a point not in the core of X for buffer distance r cannot be in the core for larger buffer distances.

The Chiu-Stoyan correction uses the length of the set of points exactly k distance from the boundary of X(obs) and integrates k between 0 and the desired radius r. It produces estimates for each radius r that are unbiased and can not increase with r, but there is a chance that the estimate will pass below 0 (Chiu and Stoyan, 1998).

Alternatively the Kaplan-Meier correction, which uses methods for the analysis of censored survival times (Baddeley and Gill, 1994), also provides non-increasing estimates of the core probability.

3.1.3 Variance. For each radius, r, the reduced sample estimator is similar to the coverage probability estimator. Molchanov (Molchanov, 1997, §4.3) provides a variance if certain properties of the process are known a priori. It may also be possible to estimate the variance using spectral density (Mase, 1982; Böhm et al., 2004), however this method requires a suitable choice of smoothing bandwidth.

3.2 Disc-State Contagion

Contagion is a popular entropy-inspired LPI for describing aggregation of classes. The unnormalised version is defined as (O’Neill et al., 1988; Li and Reynolds, 1993)

\[ \text{Contagion} := \sum_{i=1}^{m} \sum_{j=1}^{m} P_{ij} \ln(P_{ij}), \]  

where m is the number of classes and P_{ij} is the probability of randomly selected adjacent pixels being in class i and class j respectively. Although it was initially designed for multiple categories we restrict our focus to two-class maps.

Contagion as it is defined above describes the aggregation of classes within a distance of double the ground sample distance (double the ground sample distance because pixels are typically an average or weighted integral of the corresponding sample region). Contagions calculated at different resolutions thus describe aggregation at different scales making contagion very sensitive to resolution changes. Moreover because the definition depends on the resolution of an observation technique there is no canonical definition for the contagion of a real landscape. A number of other metrics that use the same pixel adjacency concept have similar issues.

Ramezani and Holm (Ramezani and Holm, 2011) encountered this issue when they tried to apply contagion to polygonal data. Instead they considered the adjacency probabilities P_{ij} to be the probability of class j intersecting a circle of radius r around a point in class i. Thus contagion became a functional metric (a function of the radius r).

We present another variant of contagion that describes the mixing of classes within a disc, termed disc-state contagion. Let 1 denote inside X and 0 denote outside X. We define P_{00}(r) as the probability that a point is in X and not in the core of X,

\[ P_{00}(r) = P(o \in X, B_r(o) \not\subseteq X), \]

and P_{01}(r) as the probability that a point is in X and not in the core of X\[,\]

\[ P_{01}(r) = P(o \in X, B_r(o) \not\subseteq X^c). \]

We define the remaining elements, P_{10}(r) as the core probability of X,

\[ P_{10}(r) := P(B_r(o) \subseteq X), \]

and similarly P_{00} as the core probability of X\[.\]

For each radius r, this version of contagion quantifies the
disorder of a system with four states (1) the entire disc is in $X$, (2) the entire disc is outside $X$, (3) the centre is within $X$ and some of the disc is outside $X$, and (4) the centre is outside $X$ and some of the disc is inside $X$. A diagram of each state is in Figure 3. The lowest contagion (and highest entropy) occurs when the probability of each of these states is $\frac{1}{4}$.

This disc-state contagion can be estimated using estimates of the SCD and core probability. It is well defined for any stationary RACS, and because the disc size is not linked to resolution it is more robust to resolution changes than the classic pixel-adjacency contagion (9).

![Figure 3: Examples of each disc state in disc-state contagion. Hatched regions denote locations inside the realisation of $X$.](image)

4. EXPLORATION OF TREE CANOPY PATTERN PROCESSES AND FEED QUALITY

We explored decimetre resolution maps of tree canopy in Perth, Australia, using conditional SCD, CCP, disc-state contagion and coverage probability. The canopy height maps were derived from stereo photography through height estimates (stereo matching) and spectral values (Caccetta et al., 2015). Two-class categorical maps were then obtained by only keeping canopy that was higher than 4 m.

We compared the RACS estimators with field-based analysis of feed quality of Banskia woodlands for an endangered bird (van Dongen et al., 2016). For each location of the field-based assessment we used a circular window of 30 m radius to estimate RACS properties. We randomly selected 50% of these field-assessed locations for initial exploration, reserving the remaining 50% for validation.

In making these estimates it was assumed that each window observed a stationary and mixing tree canopy process. Given little information on covariates such as soil, moisture and wind this is a tolerable representation for a first analysis.

The best gradation of feed quality was obtained by the CCP. A small cluster of very high feed quality processes appeared at core buffer distances of 0.75 m and above (Figure 4). In comparison the coverage probability could not separate very high feed locations from moderate feed locations (Figure 5). The SCD and disc-state contagion did not show any obvious association to feed quality (Figure 6).

Unfortunately the above discrimination of high feed locations did not generalise to the validation data (Figure 7). Further exploration of other RACS summary functions and different canopy heights might uncover real associations. Discrimination might also be achieved through multi-type RACS (with each class corresponding to a different height range), 3D stochastic models, or non-stationary RACS.

![Figure 4: Conditional core probability for tree canopy above 4 m. Each location/observation is a different curve. The colours correspond to feed quality (see colour bar on right). The canopy processes with the three highest feed quality scores (in red) are clustered together for buffer distances from 0.75 m to 2 m (black polygon). Reduced sample boundary correction was used here but the effect was minor at these distances (see Section 4.1 and Figure 8). Note that the steps in the functions were due to the map resolution of 0.2 m, and the slight upward directions of some of these jumps was caused by the reduced sample correction.](image)
6. CONCLUSION

RACS provide a powerful generic framework for modelling the processes underlying categorical maps. They are especially useful at high resolutions where the geometries created by patches of many pixels are important. The RACS framework provides a conceptual tool to guide LPI use and design including the treatment of sensing artefacts such as resolution and map extent. Here we focused on non-parametric RACS tools under the assumptions of stationarity and mixing. These tools are stochastic, process-centric versions, of LPI and we discussed the core area LPI in detail.

Core area is closely related to the spherical contact distribution. We first linked core area to a probabilistic concept, the core probability, and then showed that core probability was the opposite of the spherical contact distribution evaluated at a particular radius. This lead to functional versions of the core probability, a conditional core probability, and border corrected estimators.

The well-defined spherical contact distribution also suggested a resolution-free version of the contagion LPI. This new version of contagion describes the entropy of the state of a disc and its centre. The aggregation scale that this new contagion expresses is chosen by the user and is independent of the imaging resolution. In comparison the original pixel-adjacency contagion describes interaction over the width of two pixels and is thus much more sensitive to resolution change.

A preliminary exploration of tree canopy processes briefly demonstrated core probability, the spherical contact distribution, disc-state contagion and border corrections. Due to time restrictions the demonstrations did not include hypothesis tests.
Figure 9: The SCD estimate for one location using reduced sample correction (solid line) and no border correction (dashed line). For context the reduced sample estimates of some other observations are shown in grey. Note that, similar to the uncorrected estimates of CCP, the uncorrected estimate of SCD is lower because it has implicitly assumed that no tree canopy existed outside the observation window. For some observations, such as the one shown here, the difference between border corrected and uncorrected estimates can be quite large. Thus using the uncorrected estimates could have serious implications for applications of the SCD.

7. ACKNOWLEDGEMENTS

Many thanks to Australia’s Commonwealth Scientific and Industrial Research Organisation for supplying the tree canopy maps, and to Ricky van Dongen and Geoff Barrett of Western Australia’s Dept. of Parks and Wildlife for providing the feed data.

I would also like to thank my PhD supervisors Adrian Baddeley, Peter Caccetta, and Gopalan Nair.

References


Sadler, R., 2006. Image-based modelling of pattern dynamics in a semiarid grassland of the Pilbara, Australia. PhD, School Of Plant Biology, University of Western Australia, Perth, Australia. (Thesis).


