A nonparametric test for dependency between estimated local bulk movement patterns

Thomas R. Honnor, Adam M. Johansen, Julia A. Brettschneider

March 17, 2017

Abstract

In many applied problems for which the spatial distribution of physical quantities is recorded over time, the temporal evolution of this spatial distribution is of interest. Further, comparison of the temporal evolution of different samples in a test of dependence may determine the existence of a relationship between the physical quantities represented in each sample. One such problem is the interaction between biomolecular species. We introduce the methods currently applied to this particular problem, including their drawbacks, before formulating an alternative methodology that both estimates local bulk movement patterns and compares between two such patterns in a test of dependence. Application to both simulated and real data sets indicate that our methodology may be effectively applied to the problem of interacting biomolecular species and a wider class of more general problems.

1 Introduction

For many phenomena in which recordings of a physical process are made across various times and locations, a question of interest is the evolution in location over time and the comparison of this evolution between observations. In cases of small numbers of clearly defined individual units, estimation of movement patterns may be obtained via object tracking algorithms. Such procedures may be used for application to the movement of a species of animal (McFarlane and Schofield, 1995) or the movement of specific subcellular structures (Chenouard et al., 2014).

Our interest extends further to physical processes for which individual units are not resolvable or for which a very large number of individual units precludes use of more standard object tracking procedures. The example considered in this report is the location of biomolecules of a particular species within cells, as observed by microscope imaging. In this case each biomolecule is a discrete unit, but their large number and limits on microscope resolution make it impossible to identify and attempt to track each biomolecule. Observations are therefore interpreted as a density of biomolecules across space which evolves in time, from which we propose an estimation of movement patterns. An additional example may be the distribution of water molecules, observed via depth of water within a tank. The problem of determining movement patterns using a sequence of observations is an ill-posed inverse problem and as a result we provide an estimate of movement patterns averaged over local regions, which we term local bulk movement patterns.

Once movement patterns have been estimated, it may be desirable to compare between observations to determine whether the degree of dependence between movement patterns is statistically

Paper No. 17-03, www.warwick.ac.uk/go/crism

significant. In the case of this report the purpose is to investigate whether two biomolecular species of interest interact, a process which typically requires biomolecules from each species to become conjoined, at which times they are similarly located and undergo similar movements. Comparison of movement patterns may also be of interest in other settings, for example for animal species to understand predator-prey relationships (Mitchell and Lima, 2002). We introduce a methodology which summarises and compares the high dimensional data resulting from estimation of movement patterns in a manner which is both meaningful and takes into account the spatial nature of the observations.

Proposed methodologies are applied to independent simulated data, the results of which support the theoretical validity of the permutation testing procedure. Further testing on simulations with varying degrees of dependence indicate that the testing procedure is able to identify such scenarios with good power.

The methodology is then applied to a data set comprised of fluorescence microscopy images of TACC3, Transforming Acidic Coiled-Coil Containing Protein 3, and EB3, End-Binding protein 3. As an end binding protein, EB3 is known to locate at the growing end of microtubules during mitosis (Mimori-Kiyosue et al., 2000). It is further believed that TACC3 influences microtubule structure during mitosis (Booth et al., 2011), with our investigation considering whether EB3 and TACC3 undergo dependent movements in order to make inference on the location of TACC3.

Following this introduction, required statistical background information is included in Section 2. Section 3 describes in detail the methodology for estimating movement patterns and their comparison, with Section 4 then introducing a selection of simulations for which movement patterns and dependence is known. Analysis of a data set arising from fluorescence microscope images of EB3 and TACC3 during mitosis is then carried out in Section 5, after which conclusions are presented in Section 6.

2 Statistical Background

2.1 Existing measures of colocalisation

Colocalisation analysis is a widely used technique for the analysis of fluorescence microscopy images (Adler and Parmryd, 2012). A number of colocalisation statistics have been proposed, formulated to quantify the degree to which biomolecules are deemed to interact based upon similarities in their location. Although a commonly used term, colocalisation is poorly defined and may be used by different authors to refer to both co-occupation and correlation (Adler and Parmryd, 2012). Co-occupation is deemed to occur when light of sufficiently high intensity is observed in the same places for both channels, while correlation occurs when there is a linear relationship between intensity values paired at the same locations.

Given pixel intensity values $m^0(x)$ and $m^1(x)$ across locations x within a region of interest χ^*, a subset of the image space $\chi^* \subseteq \chi = \{1, 2, \ldots, n_1\} \times \{1, 2, \ldots, n_2\}$, a number of the most commonly used colocalisation statistics are as follows:

Definition 1. Pearson's correlation coefficient (Pearson, 1895) is given by

$$r_\rho = \frac{\sum_{x\in\chi^*}(m^0(x)-\bar{m}^0)(m^1(x)-\bar{m}^1)}{\sqrt{\sum_{x\in\chi*}(m^0(x)-\bar{m}^0)^2}\sqrt{\sum_{x\in\chi^*}(m^1(x)-\bar{m}^1)^2}} \in [-1,1],$$

where

$$\bar{m}^0 = \frac{1}{|\chi^*|}\sum_{x\in\chi^*} m^0(x) \qquad\qquad \bar{m}^1 = \frac{1}{|\chi^*|}\sum_{x\in\chi^*} m^1(x).$$

As a measure of colocalisation, Pearson's correlation coefficient is clearly a measure of correlation. A variant on Pearson's correlation coefficient, formulated to highlight the implication of values of $m(x) = 0$ and $m(x) > 0$, is Manders' overlap coefficient.

Definition 2. Manders' overlap coefficient (Manders et al., 1993) is given by

$$r = \frac{\sum_{x\in\chi*} m^0(x)m^1(x)}{\sqrt{\left(\sum_{x\in\chi^*} m^0(x)^2\right)\left(\sum_{x\in\chi^*} m^1(x)^2\right)}} \in [0,1],$$

which in turn leads to the specification of the split overlap coefficients

$$k_1 = \frac{\sum_{x\in\chi^*} m^0(x)m^1(x)}{\sum_{x\in\chi^*} m^0(x)^2} \qquad\qquad k_2 = \frac{\sum_{x\in\chi^*} m^0(x)m^1(x)}{\sum_{x\in\chi^2} m^1(x)^2},$$

such that $r = \sqrt{k_1 k_2}$.

Manders' overlap coefficient quantifies a combination of correlation and co-occupation in unclear proportions, leading some to recommend against its use in favour of the alternatives (Adler and Parmryd, 2010). A further alternative, Manders' colocalisation coefficients quantify colocalisation solely through co-occupation.

Definition 3. Manders' colocalisation coefficients (Manders et al., 1993) are given by

$$M_0 = \frac{\sum_{x\in\chi^*} m^0(x)\mathbb{1}\{m^1(x)>0\}}{\sum_{x\in\chi^*} m^0(x)} \in [0,1] \qquad M_1 = \frac{\sum_{x\in\chi^*} m^1(x)\mathbb{1}\{m^0(x)>0\}}{\sum_{x\in\chi^*} m^1(x)} \in [0,1].$$

A development of Manders' colocalisation coefficients which sets an automatic threshold, $t > 0$, to reduce the impact of background noise is given by Costes approach.

Definition 4. Costes' approach (Costes et al., 2004) suggests coefficients

$$\tilde{M}_0 = \frac{\sum_{x\in\chi^*} m^0(x)\mathbb{1}\{m^0(x)>t, m^1(x)>at+b\}}{\sum_{x\in\chi^*} m^0(x)} \in [0,1]$$
$$\tilde{M}_1 = \frac{\sum_{x\in\chi^*} m^1(x)\mathbb{1}\{m^0(x)>t, m_1(x)>at+b\}}{\sum_{x\in\chi^*} m^1(x)} \in [0,1],$$

based upon a threshold value of t. The values of a and b are determined as the intercept and slope respectively of the orthogonal regression of $m^1(x)$ on $m^0(x)$, $x \in \chi^*$. The threshold is reduced from $\max\{m^0(x), (m^1(x)-b)/a\}$ to a critical value t at which the correlation of $\{(m^0(x), m^1(x)), x : m^0(x) < t \text{ or } m^1(x) < at+b\}$ is zero.

Differences in the quantity being measured between colocalisation statistics make them difficult to interpret and compare between experiments where different statistics have been used. Co-occupation based measures are typically easier to interpret as the proportion of each biomolecular species observed at shared locations, but ignore the fact that interacting biomolecules require a fixed number of biomolecules of each species and thus a linear relationship between intensities. Correlation based measures do take into account the relationship between intensity values. In the commonly expected presence of background noise co-occupation may be recorded at every pixel location, an issue which Costes' approach attempts to resolve through thresholding. Background noise may also impact correlation statistics, for which specification of a region of interest χ^* containing large numbers of pixels with intensity levels consistent with noise alone may mask the strength of any linear relationship between intensity values.

The majority of the presented colocalisation statistics take values in the range [0,1], with zero indicating absence of colocalisation and one indicating complete colocalisation. Pearson's correlation coefficient differs, taking values in the range [-1,1], with zero indicating absence of colocalisation and one indicating strong colocalisation. Negative values of Pearson's correlation coefficient are difficult to interpret in the biological context although well understood in a statistical context. Fixed ranges of values provide some ability to interpret the strength of colocalisation, but there is no convincing method of analysing the significance of obtained colocalisation statistics. Instead, colocalisation statistic values may be classified into one of five categories from very weak to very strong based upon crude threshold values (Zinchuk and Zinchuk, 2008).

The biggest criticism of each of the proposed and often used colocalisation statistics is their ignorance of the spatial nature of the data within the region of interest χ^*. Statistical fields such as geostatistics and disease modelling analyse spatial data via methodologies which take the spatial nature into account, tools from which may be adapted to the study of colocalisation. Of particular interest may be common component models (Fanshawe and Diggle, 2011; Knorr-Held and Best, 2001), for which each observation is modelled as the combination of an observation-specific component and a scaled shared component. Comparison of the common component to observation-specific components could quantify the expected degree of interaction between biomolecules, with a plot of the common component illustrating where any interactions might be taking place. A difficulty of applying such techniques to biological images or more general data sets is their reliance upon modelling assumptions which may not be satisfied.

In cases where a single pair of images are compared to analyse colocalisation, it is difficult to distinguish between coincidental co-occupation of biomolecular species and true interaction. When a sequence of images is collected over time, colocalisation may be quantified at each time point to provide a more reliable indicator of interaction. We approach the problem of determining interaction by estimating and comparing movement patterns between consecutive time points, on the basis that chance similar movements are less likely to occur than chance similar localisations.

On an experimental level FRET, Fluorescence (or Förster) resonance energy transfer (Clegg, 1995), is an alternative methodology which may more accurately determine interaction between biomolecular species. However, false negatives may be recorded by FRET due to the requirement that fluorophores be very closely separated, which may not be the case even for interacting biomolecules, and false positives may be recorded as a result of cross-talk or bleed-through between fluorophore colours (Piston and Kremers, 2007).

2.2 Earth mover's distance

In cases where observed spatio-temporal processes are non-negative and finite, observations may be normalised to be considered as probability densities over space which evolve over time

$$\mu^0(x) = \frac{m^0(x)}{\sum_{y\in\chi} m^0(y)} \qquad\qquad \mu^1(x) = \frac{m^1(x)}{\sum_{y\in\chi} m^1(y)}.$$

Similar to colocalisation statistics presented in the previous section, there are a number of methods for quantifying the distance between probability distributions including total variational distance and Hellinger distance, see for example the summary by Gibbs and Su (2002). A distance of particular interest is the Wasserstein metric (Givens and Shortt, 1984).

Definition 5. Let (χ, d) be a metric space. The Wasserstein metric between μ^0 and μ^1 on (χ, d) is then $W(\mu^0, \mu^1) = \inf \mathbb{E}[d(X, Y)]$, taken over joint distributions of X and Y with marginals μ^0 and μ^1 respectively.

Importantly for our analysis, the Wasserstein metric takes into account the space on which the probability measures are defined, through d, in a manner that alternatives such as the total variation distance and Hellinger distance do not.

In application, it may undesirable to normalise observations m^0 and m^1 into distributions μ^0 and μ^1 as the relative total mass of each observation is informative. In such cases an alternative, but closely related, measure of the distance between m^0 and m^1 is provided by the earth mover's distance (Rubner et al., 2000).

Definition 6. The earth mover's distance between two non-negative spatial processes m^0 and m^1 over discrete spaces χ^0 and χ^1 is given by

$$\begin{aligned}\mathrm{EMD}(m^0, m^1) &= \frac{\sum_{x\in\chi^0, y\in\chi^1} \hat{f}(x,y) d(x,y)}{\sum_{x\in\chi^0, y\in\chi^1} \hat{f}(x,y)} \\ \hat{f} &= \underset{f\in\eta(m^0,m^1)}{\arg\min} \; f(x,y) d(x,y),\end{aligned}$$

for $d(x, y)$ a cost function and $\eta(m^0, m^1)$ the set of f for which

$$\begin{aligned} f(x,y) &\geq 0 && \forall x \in \chi^0, y \in \chi^1 \\ \sum_{x\in\chi^0} f(x,y) &\leq m^1(y) && \forall y \in \chi^1 \\ \sum_{y\in\chi^1} f(x,y) &\leq m^0(x) && \forall x \in \chi^0 \\ \sum_{x\in\chi^0, y\in\chi^1} f(x,y) &= \min\left(\sum_{x\in\chi^0} m^0(x), \sum_{y\in\chi^1} m^1(y)\right). \end{aligned}$$

If m^0 and m^1 are interpreted as spatial distributions of mass and the cost of moving unit mass from $x \in \chi^0$ to $y \in \chi^1$ is $d(x, y)$ then the earth mover's distance is the minimal total cost of rearranging m^0 into m^1 normalised by the total mass moved. The conditions on $\eta(m^0, m^1)$ ensure that only positive quantities of mass are moved, the total mass moved into $y \in \chi^1$ is no more than $m^1(y)$, the total mass moved out of $x \in \chi^0$ is no more than $m^0(x)$ and that the total amount of mass moved is the minimum of the total mass in m^0 and the total mass in m^1.

In the case where the total mass of m^0 and m^1 and the spaces χ^0 and χ^1 are identical, the earth mover's distance is equivalent to the Wasserstein distance (Levina and Bickel, 2001). The earth mover's distance is typically defined over a discrete space, but may be alternatively expressed for continuous spaces. In the case where $\sum_{x\in\chi^0} m^0(x) > \sum_{y\in\chi^1} m^1(y)$ the earth mover's distance may be equated to the Wasserstein distance by augmenting χ^1 with the location z such that $d(x,z) = 0\ \forall x \in \chi^0$ and specifying $m^1(z) = \sum_{x\in\chi^0} m^0(x) - \sum_{y\in\chi^1} m^1(y)$. A similar argument may be given for $\sum_{y\in\chi^1} m^1(y) > \sum_{x\in\chi^0} m^0(x)$.

Specification of the earth mover's distance and its use in practice has been motivated by problems in image analysis and in particular image comparison and retrieval (Peleg et al., 1989; Rubner et al., 2000). Given a colour image, a histogram may be produced which summarises the number of pixels satisfying intensity constraints in each of the blue, red and green channels according to a proposed binning. Related images are expected to result in histograms resulting in small values of the earth mover's distance when compared, allowing the single or multiple best matching images within a collection to be retrieved.

Calculation of the earth mover's distance is an optimal transportation problem, which with the augmentation of χ^0 or χ^1 to equalise total mass is an assignment problem (Munkres, 1957). The computational complexity of solving such a problem is $O(n^3 \log n)$ for $n = |\chi^0| + |\chi^1|$, the total number of locations (Rubner et al., 2000). An implementation of the earth mover's distance is available in R (R Core Team, 2016) via the emdist package (Urbanek and Rubner, 2012), which returns the value of the earth mover's distance and, important in later considerations, the optimal set of flows $\hat{f}$.

2.3 Permutation testing

Within the framework of statistical hypothesis testing, a null hypothesis H_0 is tested in the presence of data x using an observed test statistic $t(x)$. In the case where the distribution of the corresponding random statistic $T(X)$ under H_0 is known analytically, the value of $t(x)$ may be compared to this distribution to obtain a p-value, the probability of observing values of the test statistic as or more extreme than that calculated for the observed data. Alternatively, the distribution of $T(X)$ under H_0 may not be known analytically. In such cases, if under H_0 there exists exchangeability of X under a set of operations $\Lambda = \{\lambda_0, \lambda_1, \ldots, \lambda_m\}$, that is the joint distribution of λX is identical to that of X for all $\lambda \in \Lambda$, then H_0 may be tested using what is known as a permutation test.

Denoting by $\lambda_0 \in \Lambda$ the identity operation, that is $\lambda_0 X = X$, an exact permutation test (Edgington, 1964) requires calculation of the test statistic under all exchangeability operations $\lambda_i \in \Lambda$, producing $\{t(\lambda_0 x), t(\lambda_1 x), \ldots, t(\lambda_m x)\}$. The reported p-value in the case of a two-sided test of H_0 is then given by

$$p = \frac{1}{m+1} \sum_{\lambda\in\Lambda} \mathbb{1}[|t(\lambda x)| \geq |t(\lambda_0 x)|],$$

the proportion of the calculated $t(\lambda_i x)$ under exchangeability determined by H_0 which are as or more extreme than the observed value of the test statistic.

In the case where it is not feasible to calculate every test statistic $t(\lambda x)$ under H_0, potentially because calculation of $t(\lambda x)$ is computationally expensive or the number of possibilities $m + 1$ is

too large, the null hypothesis may be tested using what is known as an approximate permutation test (Edgington, 1969). Of the $m+1$ possible exchangeability operations, a simple random sample Λ', including λ_0, of size $m' < m+1$ is taken from Λ. The reported approximate p-value in the case of a two-sided test of H_0 is then given by

$$p = \frac{1}{m'} \sum_{\lambda \in \Lambda'} \mathbb{1}[|t(\lambda x)| \geq |t(\lambda_0 x)|],$$

the proportion of test statistics under exchangeability determined by H_0 which are as or more extreme than the observed value of the test statistic.

For both exact and approximate permutation tests the null hypothesis is then rejected at significance level α if $p < \alpha$. Both the exact and approximate permutation tests are valid as exchangeability under permutations λ under the null hypothesis implies that the $t(\lambda X)$ are identically distributed. Thus, the probability that the observed value $t(\lambda_0 x)$ lies in the most extreme α proportion of the set of all considered $t(\lambda x)$, i.e. $p < \alpha$, is α. In the case of the approximate permutation test power increases with the size of the random sample, m'. When testing at the $\alpha = 0.05$ significance level with $m' = 1000$ the power of the approximate permutation test is at least 94.5 percent of the exact test (Jöckel, 1986), rising to at least 98.3 percent when $m' = 10000$.

3 Estimating movement patterns and a test for dependence

3.1 Modelling data mathematically

We wish to analyse the spatio-temporal process M, that is $M_t(x) \in \mathbb{Q}$ for locations $x \in \chi$ and times $t \in \tau$, under a minimum of modelling assumptions. We require that the process M is non-negative across all locations and times, such that M may be likened to the distribution of a collection of basic units or particles. In situations where the number of basic units is very large, M may be likened to a density and tracking individuals becomes theoretically and computationally very difficult. We therefore propose a methodology to investigate bulk movement patterns on a scale greater than that of individual units for application in such cases.

The data available for the analysis of M is a collection of observation values $m(x,t) \in \mathbb{Q}^+$ across locations $x \in \Psi$ and times $t \in \Upsilon$. Observed values may be obtained from a realisation of the process M via projection or averaging, with observations also subject to noise. Motivating examples have been provided in the introduction, with our focus in this report on M the distribution of biomolecules of a single species and $m_t(x)$ an average of the light intensity emitted by biomolecules in a neighbourhood of pixel locations $x \in \Psi = \{1, 2, \ldots, n_1\} \times \{1, 2, \ldots, n_2\}$ at discrete time points $t \in \Upsilon = \{t_1, t_2, \ldots, t_{n_3}\}$ as recorded in the presence of background noise by a digital camera attached to a confocal microscope.

In future it will be necessary to refer to the collection of values across a set of locations $\psi \subseteq \Psi$ and times $\upsilon \in \Upsilon$ which will be denoted by $m_\upsilon(\psi)$. Our aim is to make inference on the dependency between two processes M^0 and M^1 via the comparison of two collections of observations, denoted by $m^0_\Upsilon(\Psi)$ and $m^1_\Upsilon(\Psi)$, over identical location, Ψ, and time, Υ, spaces.

3.2 Approximation of movement

Under the specification in the previous section as a scaling of the number of basic units at each location at each time point, M may be interpreted as a spatial distribution of mass which evolves over time. The process describing the quantity of such mass moving from location x at time s to location y at time t may then be denoted by $F_{s,t}(x,y)$. We specifically consider direct dependency in movement patterns F^0 and F^1, corresponding to processes M^0 and M^1, such that $F^0_{s,t}(x,y)$ is positively associated with $F^1_{s,t}(x,y)$. That is, the quantity of mass moving from locations x to y between time points s and t is positively associated between processes M^0 and M^1, across all pairs of locations and times. Although $F_{s,t}$ is described as a movement pattern, it also includes a description of those masses which remain fixed in place through $F_{s,t}(x,x)$, considered as movements which both start and end at the same location.

Determination of $F_{s,t}(x,y)$ using the information available in $m_{\{s,t\}}(\psi)$ may be formulated as a solution of the inverse problem of minimising

$$\delta^{\psi}_{s,t} = \left|\left| m_t(x) - \left(m_s(x) - \sum_{y\in\psi^*} F_{s,t}(x,y) + \sum_{y\in\psi^*} F_{s,t}(y,x) \right) \right|\right|_2,$$

subject to the constraints

$$\begin{aligned} F_{s,t}(x,y) &\geq 0 && \forall x,y, \in \psi^* \\ \sum_{y\in\psi^*} F_{s,t}(x,y) &= m_s(x) && \forall x \in \psi^* \\ \sum_{x\in\psi^*} F_{s,t}(x,y) &= m_t(y) && \forall y \in \psi^*, \end{aligned}$$

where ψ^* is the augmentation of ψ with the additional location z such that

$$\begin{aligned} m_s(z) &= \min\left\{ \sum_{x\in\psi} m_t(x) - m_s(x), 0 \right\} \\ m_t(z) &= \min\left\{ \sum_{x\in\psi} m_s(x) - m_t(x), 0 \right\}, \end{aligned}$$

to account for differences in the total mass at times s and t. The first constraint ensures that only positive masses are moved, the second ensures that the total mass moving out of each location x at s is $m_s(x)$ and the third ensures that the total mass moving into each location y at t is $m_t(y)$.

There exist solutions to this problem for which $\delta^{\psi}_{s,t} = 0$. However, the problem is ill-posed because the solution is not unique. We therefore consider $\hat{F}^{\psi}_{s,t} = \cup_{x,y\in\psi}\hat{F}_{s,t}(x,y)$, the solution to the regularised problem of minimising

$$\hat{\delta}^{\psi}_{s,t} = \left|\left| m_t(x) - \left(m_s(x) - \sum_{y\in\psi^*} F_{s,t}(x,y) + \sum_{y\in\psi^*} F_{s,t}(y,x) \right) \right|\right|_2 + ||F_{s,t}||_c,$$

according to the same set of constraints as previously and where

$$||F_{s,t}||_c = \sum_{x,y\in\psi} F_{s,t}(x,y)c(x,y),$$

for some cost function c. The resulting $\hat{F}^{\psi}_{s,t}$ is that which obtains $\delta^{\psi}_{s,t} = 0$ and minimises $||\hat{F}_{s,t}||_c$. The minimiser of $\hat{\delta}^{\psi}_{s,t}$ is not necessarily unique (see further discussion in the following section), but the problem is closer to being well-posed and later steps in the methodology are designed to result in identical outcomes for a majority of the solutions.

The resulting estimator, $\hat{F}^{\psi}_{s,t}(x, y)$, may be interpreted as the set of movements which minimises the total cost of rearranging the distribution of mass $m_s(\psi)$ into $m_t(\psi)$ according to a cost function $c(x, y)$. This can be calculated using the earth mover's distance with

$$\hat{F}^{\psi}_{s,t} = \underset{f \in \eta^{\psi}_{s,t}}{\arg\min} \sum_{x,y \in \psi} f_{s,t}(x, y)c(x, y)$$

where $\hat{F}^{\psi}_{s,t} = \cup_{x,y\in\psi}\hat{F}^{\psi}_{s,t}(x, y)$ and $\eta^{\psi}_{s,t}$ is the set of all functions f satisfying

$$f_{s,t}(x, y) \geq 0 \qquad \forall x, y, \in \psi \tag{1}$$

$$\sum_{y\in\psi} f_{s,t}(x, y) \leq m_s(x) \qquad \forall x \in \psi \tag{2}$$

$$\sum_{x\in\psi} f_{s,t}(x, y) \leq m_t(y) \qquad \forall y \in \psi \tag{3}$$

$$\sum_{x,y\in\psi} f_{s,t}(x, y) = \min\left\{\sum_{x\in\psi} m_s(x), \sum_{y\in\psi} m_t(y)\right\}. \tag{4}$$

For the analyses proposed we focus on a single cost function, $c(x, y) = ||x - y||$, equal to Euclidean distance. This cost function is chosen to be homogeneous and isotropic across $x, y \in \psi$ and penalises proposed movements only according to the distance moved, as we wish to avoid further assumptions. Returning to the interpretation of m as a distribution of mass, if we assume that applied forces are constant between s and t then $c(x, y) = ||x - y||$ is exactly the energy required to move unit mass from x to y. The estimated collection of movements $\hat{F}^{\psi}_{s,t}$ is then that which minimises the total energy required to rearrange $m_s(\Psi)$ into $m_t(\Psi)$. Some further discussion of the choice of cost function follows in Sections 3.3 and 3.7.

3.3 Movement summary statistic

The collection of movements $\hat{F}^{\psi}_{s,t}$ is a straightforward estimator of F^{ψ}, taking values in the high dimensional set $(\mathbb{Q}^+)^{\psi\times\psi}$. To facilitate comparison between $\hat{F}^{\psi,0}$ and $\hat{F}^{\psi,1}$ resulting from observations m^0 and m^1 we first summarise the information in $\hat{F}^{\psi}$ via $\hat{S}^{\psi} \in (\mathbb{R}^+)^8$.

Defining $a(v) \in (0, 2\pi]$ to be the anticlockwise angle between the vector $(1, 0)$ and the vector $v \in \mathbb{R}^2$, $\theta_1 = (15\pi/8, 2\pi]\cup(0, \pi/8]$ and $\theta_j = ((2j-3)\pi/8, (2j-1)\pi/8]$ for $j \in \{2, 3, \ldots, 8\}$, elements of the summary statistic are given by

$$(\hat{S}^{\psi}_{s,t})_j = \sum_{x,y\in\psi: a(y-x)\in\theta_j} \hat{F}^{\psi}_{s,t}(x, y)||x - y||_2 \qquad\qquad j \in \{1, 2, \ldots, 8\}.$$

An illustration of this summary is presented in Figure 1. To elaborate on the formulation of the summary statistic, $(\hat{S}^{\psi}_{s,t})_j$ is equal to the total momentum in directions within an angle of $\pi/8$ of the cardinal or ordinal direction (E,NE,N,NW,W,SW,S,SE) corresponding to j, multiplied by $t - s$. This formulation is specific to $\psi \subseteq \mathbb{R}^2$ as is the focus of this report, however, generalisation may be made to spaces of dimension other than two.

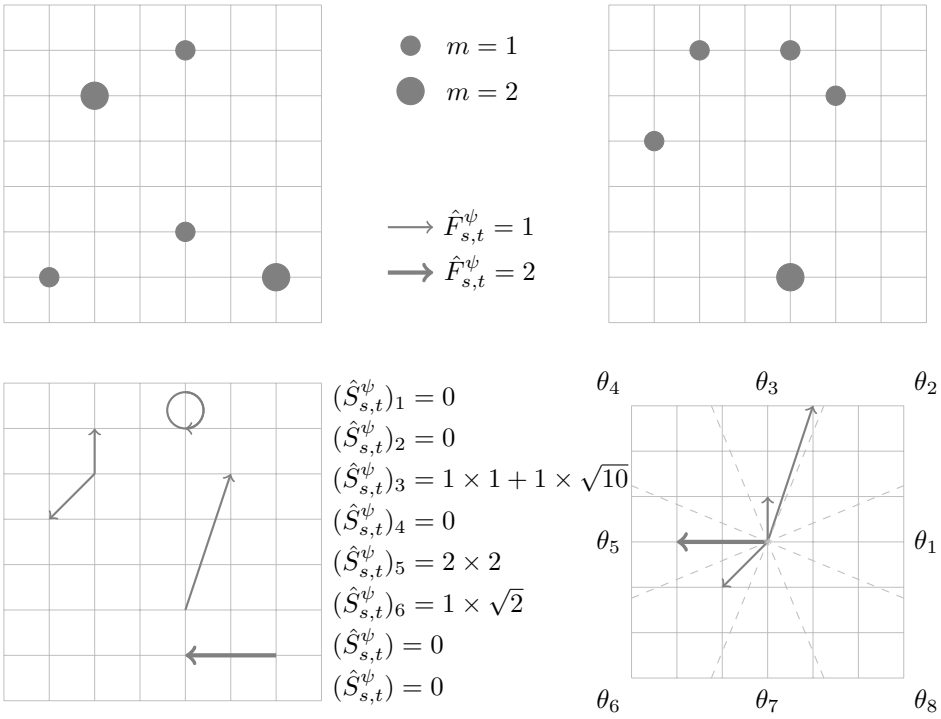


Figure 1: Distribution of $m_s(\psi)$, top left, and $m_t(\psi)$, top right, estimated $\hat{F}^{\psi}_{s,t}$, bottom left, and contributions to the calculation of $\hat{S}_{s,t}$, bottom right.

The division of movements into eight different directions provides more detail than if they were for instance separated into the four cardinal directions. In the particular instance where ψ is a regular grid, $\psi = \{1, 2, \ldots, n_1\} \times \{1, 2, \ldots, n_2\}$, the proposed division also avoids peculiarities at the boundaries of sets θ_j as there exists no direction $y - x$ lying exactly on the border of any of the sets θ_j. To establish this, note that $a(y - x) = \pi/8$ if and only if $(y - x)_2/(y - x)_1 = \tan(\pi/8) = 1 + \sqrt{2} \in \mathbb{R} \setminus \mathbb{Q}$, whereas ψ restricts $(y - x)_2/(y - x)_1 \in \mathbb{Q}$. A similar argument holds for all boundaries between θ_i and θ_{i+1}, each of which occurs at angles which are odd multiples of $\pi/8$.

Particular distributions m_s and m_t alongside the specification of $c(x, y) = ||x - y||_2$ may lead to non-uniqueness of $\hat{F}^{\psi}_{s,t}$. For example, consider three colinear locations $\psi = \{x, y, z\}$ with $||x-y|| + ||y-z|| = ||x-z||$ and $\Upsilon = \{0, 1\}$ with $m_0(\psi) = (1, 1, 0)$ and $m_1(\psi) = (0, 1, 1)$. In such a scenario $\hat{F}^{\psi}_{0,1}(a, b) = \mathbb{1}\{a = x, b = z\}$ and $\hat{F}^{\psi}_{0,1}(a, b) = \mathbb{1}\{a = x, b = y \text{ or } a = y, b = z\}$ both satisfy the requirements in the previous section. However, in both cases $\sum_{a,b \in \psi} \hat{F}^{\psi}_{0,1}(a, b)||a - b||_2 = 2$ is identical. The specification of $\hat{S}^{\psi}$ in terms of momentum, moving masses multiplied by distances moved, therefore resolves the most common scenario of non-unique $\hat{F}^{\psi}$ into a consistent value of $\hat{S}^{\psi}$. Alternative scenarios in which $\hat{F}^{\psi}_{s,t}$ is not unique and which result in different values of $\hat{S}_{s,t}$ do exist, but require particular values of m_s and m_t at more than three locations and complex interactions with values of m_s and m_t at the remaining locations in ψ, which are expected to be unlikely and increasingly so for larger spaces ψ.

3.4 Summary statistic comparison

Our interest is in the degree to which movement patterns differ between observations m^0 and m^1 and the significance of any difference. In order to quantify this we require a method of comparing summary statistics $\hat{S}^{\psi,0}_{s,t}$ and $\hat{S}^{\nu,1}_{s,t}$ between subsets $\psi, \nu \in \Psi$. We propose to make this comparison using the function $\hat{G}^{\psi,\nu}_{s,t}$.

Before specifying $\hat{G}^{\psi,\nu}_{s,t}$ we first define the intermediate quantity

$$\mu^{\psi,0}_{s,t}(x) = \sum_{j=1}^{8} \mathbb{1}\left\{x = e_j \left(\hat{S}^{\psi,0}_{s,t}\right)_j\right\} \qquad\qquad x \in \mathbb{R}^2,$$

where e_j is the unit length vector such that $a(e_j) = (j-1)\pi/4$. The intermediary $\mu^{\psi,0}_{s,t}$ may then be seen as the spatial distribution of eight unit masses, each of which is at a distance $(\hat{S}^{\psi,0}_{s,t})_j$ from the origin in the direction e_j.

The value of the comparison is then given by

$$\hat{G}^{\psi,\nu}_{s,t} = \text{EMD}(\mu^{\psi,0}_{s,t}, \mu^{\nu,1}_{s,t}),$$

the minimal cost required to rearrange the eight unit masses with cost function Euclidean distance. Small values of $\hat{G}$ imply close agreement of bulk movement patterns, with large values of $\hat{G}$ indicating differences.

Comparison of summarised movement patterns $\hat{S}$ is proposed using this method rather than a more straightforward alternative because the $\hat{S}$ summarise spatial information. The first element of $\hat{S}$ quantifies momentum in easterly directions (between east north east and east south east), with the second in north-easterly directions and the fifth in westerly directions. Intuitively, $\hat{S} = (0,1,0,0,0,0,0,0)$ is therefore more similar to $\hat{S} = (1,0,0,0,0,0,0,0)$ than to $\hat{S} = (0,0,0,0,1,0,0,0)$. The earth mover's distance takes this into account, while approaches which treat $\hat{S}$ as a vector generally do not.

In the case where M is the distribution of a physical quantity the elements of $\hat{S}$ are proportional to the momentum of the movements, with $\hat{G}^{\psi,\nu}$ then proportional to the impulse required to transform $\hat{S}^{\psi,0}$ into $\hat{S}^{\nu,1}$.

3.5 Combination of summary statistic comparisons

As previously stated, the high dimensionality of estimated movement patterns, $\hat{F}$, makes them difficult to compare. We have therefore proposed summary statistics $\hat{S}$ and a method to compare them between observations m^0 and m^1 via $\hat{G}$.

A comparison of bulk movement patterns across the entire space Ψ between two consecutive time points is given by $\hat{G}^{\Psi,\Psi}_{s,t}$. However, aggregating movements over Ψ for large image spaces may result in unintuitive values of $\hat{G}^{\Psi,\Psi}_{s,t}$. For example, if movement in the upper half of Ψ is in easterly directions and in the lower half of Ψ is in westerly directions for m^0 and vice versa for m^1 then $\hat{S}^{\Psi,0}_{s,t}$ and $\hat{S}^{\Psi,1}_{s,t}$ are expected to be very similar and the value of $\hat{G}^{\Psi,\Psi}_{s,t}$ very small.

A more effective comparison may be obtained by comparing bulk movement patterns over smaller subregions of Ψ and then combining these values. We therefore consider the partition of

Ψ into w subregions denoted by $\Psi^w = \{\psi_1, \psi_2, \ldots, \psi_w\}$ and the comparison statistic

$$\hat{H}_{s,t}^{\Psi^w,\Psi^w} = \sum_{j=1}^{w} \hat{G}_{s,t}^{\psi_j,\psi_j}.$$

Choice of the number, sizes and organisation of subregions clearly has an effect on the value of $\hat{H}$. In future examples we consider the regular grid location space $\Psi = \{1, 2, \ldots, n_1\} \times \{1, 2, \ldots, n_2\}$ which may be partitioned into identically sized square subregions which tessellate Ψ, an illustration which may be seen in Figure 2. This restricts the number of subregions to $w = u_1 u_2$ for u_1 dividing n_1 and u_2 dividing n_2. Selection of a large value of w results in comparisons over smaller subregions, resulting in a value of $\hat{H}^{\Psi^w,\Psi^w}$ which is sensitive to differences in bulk movement patterns on a smaller local scale.

On the other hand, $\hat{H}_{s,t}^{\Psi^w,\Psi^w}$ is dependent upon estimated movement patterns $\hat{F}_{s,t}^{\psi_j}$ which are calculated using only the information in $m_{\{s,t\}}(\psi_j)$. Movements estimated using all available data, $\hat{F}^{\Psi}$, are expected to be the closest estimator of the true F, with the combination of regional estimates $\cup_{i=1}^{w} \hat{F}^{\psi_j}$ expected to decrease in accuracy with w as movements between subregions are not taken into consideration. There is therefore a balance in choosing w large enough that the local comparisons are sensitive, but small enough that the local movement patterns are accurate.

Partition of Ψ into Ψ^w has computational benefits, discussed in more detail in Section 3.7.

Assuming mass is evenly distributed over the space and moves at a consistently distributed speed, the magnitude of $\hat{S}^{\psi_j}$ is approximately proportional to subregion size which is in turn inversely proportional to the number of subregions. Values of $\hat{H}^{\Psi^w,\Psi^w}$ may therefore be considered to be on the same scale regardless of the choice of w. Despite this, we do not compare $\hat{H}^{\Psi^w,\Psi^w}$ for different values of w as we are more interested in tests of the significance of $\hat{H}^{\Psi^w,\Psi^w}$ through comparison to $\hat{H}^{\Psi^w,\lambda\Psi^w}$ for some rearrangement operator λ on the collection of subregions Ψ^w.

3.6 Significance quantification

Our aim is to determine whether or not the evolutions of M^0 and M^1 over time are dependent. The formulation of $\hat{H}^{\Psi^w,\Psi^w}$ gives an insight into this, with small values indicating similarity in local bulk movement patterns and evidence for dependence and the converse for large values. However, without making further assumptions on the evolution of processes M^0 and M^1 it is not possible specify a parametric distribution of $\hat{H}$ under which its significance may be quantified. We therefore consider nonparametric testing of the significance of the observed value $\hat{H}_{s,t}^{\Psi^w,\Psi^w}$.

In particular, we consider a permutation test under the action of $\lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_w\} \in \Lambda$ on the collection of subregions $\Psi^w = \{\psi_1, \psi_2, \ldots, \psi_w\}$ with $\lambda\Psi^w = \{\lambda_1\psi_1, \lambda_2\psi_2, \ldots, \lambda_w\psi_w\}$, chosen such that under a specified null hypothesis H_0 there exists exchangeability of the set $\{S_{s,t}^{\psi_1}, S_{s,t}^{\psi_2} \ldots, S_{s,t}^{\psi_w}\}$ under the action of any $\lambda \in \Lambda$, that is

$$\{S_{s,t}^{\psi_1}, S_{s,t}^{\psi_2} \ldots, S_{s,t}^{\psi_w}\} \stackrel{\mathrm{d}}{=} \{S_{s,t}^{\lambda_1\psi_1}, S_{s,t}^{\lambda_2\psi_2} \ldots, S_{s,t}^{\lambda_w\psi_w}\},$$

with $\stackrel{\mathrm{d}}{=}$ denoting equality in distribution. In practise, the set Λ is too large to calculate $\hat{H}_{s,t}^{\Psi^w,\lambda\Psi^w}$ for all $\lambda \in \Lambda$ and so an approximate permutation test is carried out using a random subset Λ' of Λ which includes the identity operator $\lambda^0 : \lambda^0\Psi^w = \Psi^w$. The resulting p-value is then given by

$$p = \frac{1}{|\Lambda'|} \sum_{\lambda \in \Lambda'} \mathbb{1}\{\hat{H}_{s,t}^{\Psi^w,\lambda\Psi^w} \leq \hat{H}_{s,t}^{\Psi^w,\Psi^w}\}.$$

Each null hypothesis tested according to this method is comprised of three statements. The first statement is that there is between-sample independence of local bulk movement patterns, that is $\{S_{s,t}^{\psi_1,0}, S_{s,t}^{\psi_2,0}, \ldots, S_{s,t}^{\psi_w,0}\}$ is independent of $\{S_{s,t}^{\psi_1,1}, S_{s,t}^{\psi_2,1}, \ldots, S_{s,t}^{\psi_w,1}\}$. The second statement is required to specify the set of operations Λ under which permutation testing is carried out, in particular those operations for which the marginal distributions are identical, $S_{s,t}^{\psi_j} \overset{\mathrm{d}}{=} S_{s,t}^{\lambda_j \psi_j}$, examples of which are provided in Section 3.6.2. The final statement is that there is within-sample independence of local bulk movement patterns. That is $S_{s,t}^{\psi_j}$ is independent of $S_{s,t}^{\psi_k}$ for $j \neq k$, required to ensure exchangeability under the action of $\lambda \in \Lambda$.

Within-sample dependence between $S_{s,t}^{\psi_j}$ and $S_{s,t}^{\psi_k}$ is caused by movements between subregions, $F_{s,t}(x, y) > 0$ for $x \in \psi_j, y \in \psi_k$. In cases where the interval between time points, $t - s$, is small in comparison to the speed at which mass moves in M, $F_{s,t}(x, y)$ is expected be non-zero only for those locations x and y close to the border between adjacent subregions ψ_j and ψ_k. In the rest of this report we consider observations from designed experiments for which the time points are chosen close together to satisfy this. The proportion of movements which are inter-subregion may be further limited by reducing the number of subregions w, limiting the proportion of the space Ψ which is adjacent to a boundary between subregions.

Variation in the second statement of the null hypothesis allows for the testing of independence under differing assumptions. In Section 3.6.2 four example null hypotheses are given with corresponding sets of operators Λ. However, the methodology is by no means restricted to these four null hypotheses, but rather is more broadly applicable to any null hypothesis for which a suitable set of operations Λ may be determined under which the S^{ψ_j} are exchangeable. Sets of operations Λ are typically created based upon null hypothesis statements that permit reflection or rotational symmetry.

3.6.1 Operator definitions

In the following sections we will repeatedly refer to particular rearrangements of the sets ψ_j which we therefore define here.

Let R_e define a set of rearrangements of ψ_j based upon reflections. Firstly, $\rho_1\psi_j$ is the rearrangement based upon the reflection of locations $x \in \psi_j$ across the line passing through the centre of ψ_j in the direction $(1, 0)$. Similarly, $\rho_2\psi_j$ across the line in direction $(1, 1)$, $\rho_3\psi_j$ across the line in direction $(0, 1)$ and $\rho_4\psi_j$ across the line in direction $(-1, 1)$. Further, let $\rho_0\psi_j = \psi_j$ be the identity rearrangement.

Let R_o define a set of rearrangements of ψ_j based upon rotations. Firstly, $\varrho_1\psi_1$ is the rearrangement based upon the rotation of locations $x \in \psi_j$ anticlockwise about the centre of ψ_j by angle $\pi/2$. Similarly, $\varrho_2\psi_j$ by an angle of π and $\varrho_3\psi_j$ by an angle of $3\pi/2$. Further, let $\varrho_0\psi_j = \psi_j$ be the identity rearrangement.

A diagram illustrating $\rho \in R_e$ and $\varrho \in R_o$ for Ψ may be seen in Figure 2.

Applied in combination $\rho\varrho$ or $\varrho\rho$ with $\rho \in R_e = \{\rho_0, \rho_1, \rho_2, \rho_3, \rho_4\}$ and $\varrho \in R_0 = \{\varrho_0, \varrho_1, \varrho_2, \varrho_3\}$ a number of permutations are identical. For example, $\rho_3\varrho_1\psi_j = \rho_2\varrho_0\psi_j$. There are a total of eight unique transformations of ψ_j of this form, one listing of which is $\{\rho_0\varrho_0, \rho_1\varrho_0, \rho_2\varrho_0, \rho_3\varrho_0, \rho_4\varrho_0, \rho_0\varrho_1, \rho_0\varrho_2, \rho_0\varrho_3\}$.

The value of $\hat{S}_{s,t}^{\rho\varrho\psi_j}$ for any of the $\rho \in \{\rho_0, \rho_1, \rho_2, \rho_3, \rho_4\}$ and $\varrho \in \{\varrho_0, \varrho_1, \varrho_2, \varrho_3\}$ is obtained by a straightforward rearrangement of $\hat{S}_{s,t}^{\psi_j}$. For example, $\hat{S}_{s,t}^{\rho_3\varrho_1\psi_j} = ((\hat{S}_{s,t}^{\psi_j})_3, (\hat{S}_{s,t}^{\psi_j})_2, (\hat{S}_{s,t}^{\psi_j})_1, (\hat{S}_{s,t}^{\psi_j})_8,$

$(\hat{S}^{\psi_j}_{s,t})_7, (\hat{S}^{\psi_j}_{s,t})_6, (\hat{S}^{\psi_j}_{s,t})_5, (\hat{S}^{\psi_j}_{s,t})_4)$.

3.6.2 Example hypotheses

Definition 7. The isotropic null hypothesis, H_0^I, specifies on a scale according to w between-sample independence of local bulk movement patterns, that local bulk movement patterns, S^{ψ_j}, are homogeneous and isotropic across the whole space Ψ and within-sample independence of local bulk movement patterns. The corresponding set of operators for which there exists exchangeability under H_0^I is denoted by Λ^I.

Under the assumption of isotropy, local bulk movement patterns S^{ψ_j} are identically distributed under rearrangement of ψ_j by rotation and reflection by the action of $\rho\varrho, \rho \in R_e, \varrho \in R_o$ on ψ_j. Further, under the assumption of homogeneity, local bulk movement patterns S^{ψ_j} are identically distributed under any rearrangement of subregions, realised as a reordering of $\Psi^w = \{\psi_1, \psi_2, \ldots, \psi_w\}$. The set of operators Λ^I is therefore

$$\Lambda^{H,I} = \{(\lambda_1, \lambda_2, \ldots, \lambda_w) : \lambda_j \psi_j = \rho\varrho\psi_{\sigma(j)}, \rho \in R_e, \varrho \in R_o, \sigma \in \mathcal{S}_w\},$$

where $\mathcal{S}_w$ is the symmetric group of size w. The total number of unique operators is $|\Lambda^I| = 8^w \times w!$.

Definition 8. The homogeneous null hypothesis, H_0^H, specifies on a scale according to w between-sample independence of local bulk movement patterns, that local bulk movement patterns, S^{ψ_j}, are homogeneous across the whole space Ψ and within-sample independence of local bulk movement patterns. The corresponding set of operators for which there exists exchangeability under H_0^H is denoted by Λ^H.

Under the sole assumption of homogeneity the set of operators Λ^H is given by

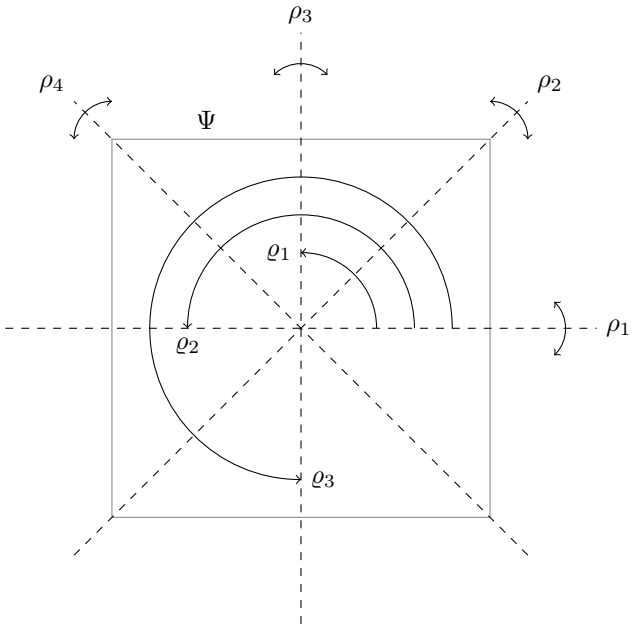
$$\Lambda^H = \{(\lambda_1, \lambda_2, \ldots, \lambda_w) : \lambda_j \psi_j = \psi_{\sigma(j)}, \sigma \in \mathcal{S}_w\}.$$

The total number of unique operators is $|\Lambda^H| = w!$. Note that $\Lambda^H \subset \Lambda^I$.

Definition 9. The symmetric null hypothesis, H_0^S, specifies on a scale according to w between-sample independence of local bulk movement patterns, that local bulk movement patterns, S^{ψ_j}, are symmetric across the whole space under the application of $\rho\varrho, \rho \in R_e, \varrho \in R_o$ to Ψ and within-sample independence of local bulk movement patterns. The corresponding set of operators for which there exists exchangeability under H_0^S is denoted by Λ^S.

Symmetry in the distribution of S^{ψ_j} under a limited set of rotations and reflection of Ψ allows us to partition Ψ^w into classes for which ψ_j and ψ_k are members of the same class if and only if $S^{\psi_j} \stackrel{\mathrm{d}}{=} S^{\rho\varrho\psi_k}$ for some suitable choice of ρ and ϱ. The specification of symmetry under the application of all rotations and reflections $\rho\varrho, \rho \in R_e, \varrho \in R_o$ to Ψ requires Ψ to be square.

An example in the case of $w = 25$ is illustrated in Figure 2, for which the exchangeability classes are $\{\psi_{13}\}$, $\{\psi_8, \psi_{12}, \psi_{14}, \psi_{18}\}$, $\{\psi_7, \psi_9, \psi_{17}, \psi_{19}\}$, $\{\psi_3, \psi_{11}, \psi_{15}, \psi_{23}\}$, $\{\psi_2, \psi_4, \psi_6, \psi_{10}, \psi_{16}, \psi_{20}, \psi_{22}, \psi_{24}\}$ and $\{\psi_1, \psi_5, \psi_{21}, \psi_{25}\}$. Examples of allowed transformations are $S^{\rho\varrho\psi_{13}} \stackrel{\mathrm{d}}{=} S^{\psi_{13}}$ $\forall \rho \in R_e, \varrho \in R_o$, $S^{\rho_4\psi_8} \stackrel{\mathrm{d}}{=} S^{\varrho_1\psi_8} \stackrel{\mathrm{d}}{=} S^{\psi_{14}}$ and $S^{\rho_3\psi_2} \stackrel{\mathrm{d}}{=} S^{\psi_4}$.



Ψ

ψ_{21}	ψ_{22}	ψ_{23}	ψ_{24}	ψ_{25}
ψ_{16}	ψ_{17}	ψ_{18}	ψ_{19}	ψ_{20}
ψ_{11}	ψ_{12}	ψ_{13}	ψ_{14}	ψ_{15}
ψ_{6}	ψ_{7}	ψ_{8}	ψ_{9}	ψ_{10}
ψ_{1}	ψ_{2}	ψ_{3}	ψ_{4}	ψ_{5}

Figure 2: Illustration of rotations and reflections of Ψ, left, and the division of Ψ into subregions $\psi_j \in \Psi^{25}$ with shading according to the exchangeability classes, right.

The total number of unique operators is

$$|\Lambda^S| = \begin{cases} 8 & \text{if } w = 1 \\ 2^4 \times 4! & \text{if } w = 4 \\ 8 \times (2^4 \times 4!) & \text{if } w = 9 \\ (k \times 2^4 \times 4!) \times \left(\frac{k(k-1)}{2} \times 8!\right) & \text{if } w = (2k)^2, k = 2, 3, \ldots \\ 8 \times (2k \times 2^4 \times 4!) \times \left(\frac{k(k-1)}{2} \times 8!\right) & \text{if } w = (2k+1)^2, k = 2, 3, \ldots. \end{cases}$$

Note that $\Lambda^S \subset \Lambda^I$ as Λ^I includes all rearrangements, rotations and reflections and Λ^S includes only those which preserve the classes of Ψ^w. Further, $\Lambda^S \nsubseteq \Lambda^H$ as Λ^H includes only rearrangements of subregions and Λ^S requires rotations and reflections of subregions alongside rearrangements, and $\Lambda^H \nsubseteq \Lambda^S$ as Λ^H includes all rearrangements of subregions and Λ^S restricts rearrangements within the classes partitioning Ψ^w.

Definition 10. The horizontal reflection null hypothesis, H_0^R, specifies on a scale according to w between-sample independence of local bulk movement patterns, that local bulk movement patterns, S^{ψ_j}, are horizontally symmetric across the whole space under the application of ρ_3 to Ψ and within-sample independence of local bulk movement patterns. The corresponding set of operators for which there exists exchangeability under H_0^R is denoted by Λ^R.

Under H_0^R subregions ψ_j are each paired with the corresponding subregion located at the same position as $(\rho_3 \Psi^w)_j$ for which rearrangement is allowed under the application of ρ_3 to both ψ_j and $(\rho_3 \Psi^w)_j$.

In the case where Ψ^w is u_1 subregions wide and u_2 subregions high, with $w = u_1 u_2$, the total number of unique operators is

$$|\Lambda^R| = 2^{\lceil \frac{u_1}{2} \rceil u_2},$$

where $\lceil u \rceil$ is the value of u rounded up to the nearest integer.

3.7 Computational considerations

Estimation of local bulk movement patterns $\hat{S}^{\psi}_{s,t}$ first requires estimation of movement patterns $\hat{F}^{\psi}_{s,t}$ resulting from calculation of the earth mover's distance between $m_s(\psi)$ and $m_t(\psi)$. Calculation of the earth mover's distance is based upon solution of an assignment problem, the computational cost of which is superlinear in the number of origin and destination locations $x, y \in \psi$, requiring $O(|\psi|^3 \log |\psi|)$ operations (Rubner et al., 2000). As a result, partitioning Ψ into w subregions ψ_j results in a collection of $\hat{S}^{\psi_j}_{s,t}$ which may be determined at less overall computational cost than $\hat{S}^{\Psi}_{s,t}$, by a factor of $1/w^2$ as

$$\begin{aligned} w \times O(|\psi_j|^3 \log |\psi_j|) &= O(w|\psi_j|^3 \log |\psi_j|) \\ &= O\left(w \left(\frac{|\Psi|}{w} \right)^3 \log \frac{|\Psi|}{w} \right) \\ &= O\left(\frac{1}{w^2} |\Psi|^3 \log \Psi - \frac{1}{w^2} |\Psi|^3 \log w \right) \\ &= \frac{1}{w^2} O(|\Psi|^3 \log |\Psi|). \end{aligned}$$

Furthermore, as calculation of $\hat{S}^{\psi_j}_{s,t}$ is independent of calculation of $\hat{S}^{\psi_k}_{s,t}$ for disjoint sets ψ_j and ψ_k, partitioning of Ψ into subregions ψ_j permits parallelisation of earth mover's distance calculations.

The cost function for estimating movements, $\hat{F}^{\psi}_{s,t}(x, y)$, is specified as $c(x, y) = ||x - y||_2$, the Euclidean distance between locations x and y. This cost function satisfies the triangle inequality, that is for locations x, y and z we have $c(x, z) \leq c(x, y) + c(y, z)$. An interpretation of the triangle inequality for our application is that it is always as or more expensive to move mass from x into y and equal mass from y to z than it is to move mass directly from x to z. As a result, provided c satisfies the triangle inequality, we can state before calculation of the earth mover's distance that $\hat{F}^{\psi}_{s,t}(x, x) = \min\{m_s(x), m_t(x)\}$. Calculation of $\hat{F}^{\psi}_{s,t}(x, y)$ may therefore be based upon the collection of data $\{m_{s,+}(\psi), m_{t,-}(\psi)\} = \{m_s(x) - \min\{m_s(x), m_t(x)\}, m_t(x) - \min\{m_s(x), m_t(x)\}, x \in \psi\}$. However, for every pair $m_{s,+}(x), m_{t,-}(x)$ at least one is zero. Therefore, the total number of origin and destination locations is reduced by at least half, resulting in a large computational saving when c satisfies the triangle inequality.

As stated in Section 3.6.1, the value of $\hat{S}^{\rho\varrho\psi_j}_{s,t}$ may be obtained without further calculation from $\hat{S}^{\psi_j}_{s,t}$ by rearrangement if $\rho \in R_e$ and $\varrho \in R_o$. In cases where the null hypothesis permits operations of rearrangement, reflection and rotation dramatic savings can therefore by made by avoiding repeated recalculation of the earth mover's distance.

4 Validation study

4.1 Simulation description

A general simulation may be made up of a total of k objects, where the centre of object i at time t is denoted by $c_i(t)$ and its intensity by b_i. The evolution over time of the process is determined by the relationship between $c_i(s)$ and $c_i(t)$ for time points s and t. The shape of object i is specified by $d_i(x)$ the set of points belonging to shape i whose centre is located at x. In this manner the simulations may be thought of as sequences of germ-grain models, with point locations or germs $c_i(t)$ associated with sets corresponding to object shapes or grains $d_i(x)$. Simulations may include an observation error term $\epsilon(x,t)$ at locations x at times t. Under such a formulation we may specify

$$m(x,t) = \sum_{i=1}^{k} b_i \mathbb{1}\{x \in d_i(c_i(t))\} + \epsilon(x,t)$$
$$F_{s,t}(x,y) = \sum_{i=1}^{k} b_i \mathbb{1}\{x \in d_i(c_i(t)), y - x = c_i(t) - c_i(s)\}.$$

For the validation study investigated in this section we consider a discrete location space $\Psi = \{1,2,\ldots,60\} \times \{1,2,\ldots,60\}$ and a discrete time space $\Upsilon = \{1,2,3,4,5\}$. The number of objects, k, differs between simulation classes but objects are all of the same intensity $b_i = 30$. Objects are further all the same size and shape, that of a Greek cross of the four locations directly adjacent to $x = (x_1, x_2)$ and x itself

$$d_i(x) = \{x, (x_1, x_2 - 1), (x_1, x_2 + 1), (x_1 - 1, x_2), (x_1 + 1, x_2)\},$$

and initially independently distributed uniformly over Ψ, $c_i(1) \sim \text{Uniform}\{\Psi\}$. The evolution of object centres over time also differs between simulations and is therefore described separately. The observation error is Poisson distributed white noise with mean three, that is $\epsilon(x,t) \sim \text{Poisson}(3)$ independently for all times $t \in \Upsilon$ and locations $x \in \Psi$.

For each class of simulation 15 replicates are produced, allowing 105 unique comparisons between simulations of the same class. Dependence between movement patterns in observations of the same class is introduced by the inclusion of a proportion of identical objects in each simulation.

Definition 11. Noise simulations contain no objects, $k = 0$, and therefore represent only white noise.

There is no dependence between noise simulations.

Definition 12. Isotropic simulations contain 100 objects, $k = 100$, for which

$$c_i(t+1) = c_i(t) + 3(\cos(\theta_{i,t}), \sin(\theta_{i,t})) \qquad \text{mod } 60,$$

for $\theta_{i,t} \sim \text{Uniform}[0, 2\pi)$ simulated independently for each object and each time point. Centre locations are calculated under the specification that $60 \bmod 60 = 60$ and with rounding of $c_i(t)$ to the nearest location in Ψ.

The evolution of object centres for isotropic simulations is independent of location and uniform across all directions, resulting in movement patterns which are both isotropic, homogeneous and reflection and rotationally symmetric under combinations $\rho\varrho, \rho \in R_e, \varrho \in R_o$ applied to Ψ. Taking

object centres modulo 60 ensures that all objects remain within the location space Ψ, appearing to wrap around from each edge to the opposite edge.

Isotropic 10 simulations each contain the same 10 objects and 90 objects simulated independently for each observation. Similarly, isotropic 30 simulations each contain the same 30 objects and 70 objects simulated independently for each observation.

Definition 13. Homogeneous simulations contain 100 objects, $k = 100$, for which

$$c_i(t+1) = c_i(t) + 3\left(\cos\left(\frac{\pi}{4}\right), \sin\left(\frac{\pi}{4}\right)\right) \qquad \text{mod } 60,$$

again under the specification that $60 \bmod 60 = 60$ and with rounding of $c_i(t)$ to the nearest location in Ψ.

The evolution of object centres for homogeneous simulations is independent of location, resulting in movement patterns which are homogeneous. Evolutions are also symmetric under reflection of Ψ by ρ_2 as centres move in the direction of the vector $(1, 1)$. Movement patterns in homogeneous simulations are neither isotropic nor reflection or rotationally symmetric in any other way. Taking object centres modulo 60 again ensures that all objects remain within the location space Ψ.

Homogeneous 10 simulations each contain the same 10 objects and 90 objects simulated independently for each observation. Similarly, homogeneous 30 simulations each contain the same 30 objects and 70 objects simulated independently for each observation.

Definition 14. Symmetric simulations contain 100 objects, $k = 100$, for which

$$c_i(t+1) = c_i(t) + 3\frac{z - c_i(t)}{||z - c_i(t)||},$$

where $z = (30.5, 30.5)$ is the point at the centre of Ψ and locations $c_i(t)$ are rounded to the nearest location in Ψ.

The evolution of object centres for symmetric simulations produces movement patterns which are reflection and rotationally symmetric under combinations $\rho\varrho, \rho \in R_e, \varrho \in R_o$ applied to Ψ. Movement patterns in symmetric simulations are neither homogeneous nor isotropic, additionally differing from all other described simulations in that the distribution of the collection of objects is expected to vary over time as they aggregate towards the centre of Ψ.

Symmetric 10 simulations each contain the same 10 objects and 90 objects simulated independently for each observation. Similarly, symmetric 30 simulations each contain the same 30 objects and 70 objects simulated independently for each observation.

Object speeds have been fixed at three as a compromise, with speeds smaller than three resulting in movements which may be accurately determined by eye alone and speeds larger than three expected to limit accuracy of $\hat{F}$ as an estimator of F. The following section provides the results of testing the described classes of simulations under a variety of hypotheses for a range of subregion sizes and thus a range of values of w. With square subregions of width 10, 12, 15 and 20 (corresponding to $w = 36$, 25, 16 and 9), object speeds of three ensures that the proportion of objects moving between subregions is small.

Illustrative examples of each class of simulation are presented in Figure 3.

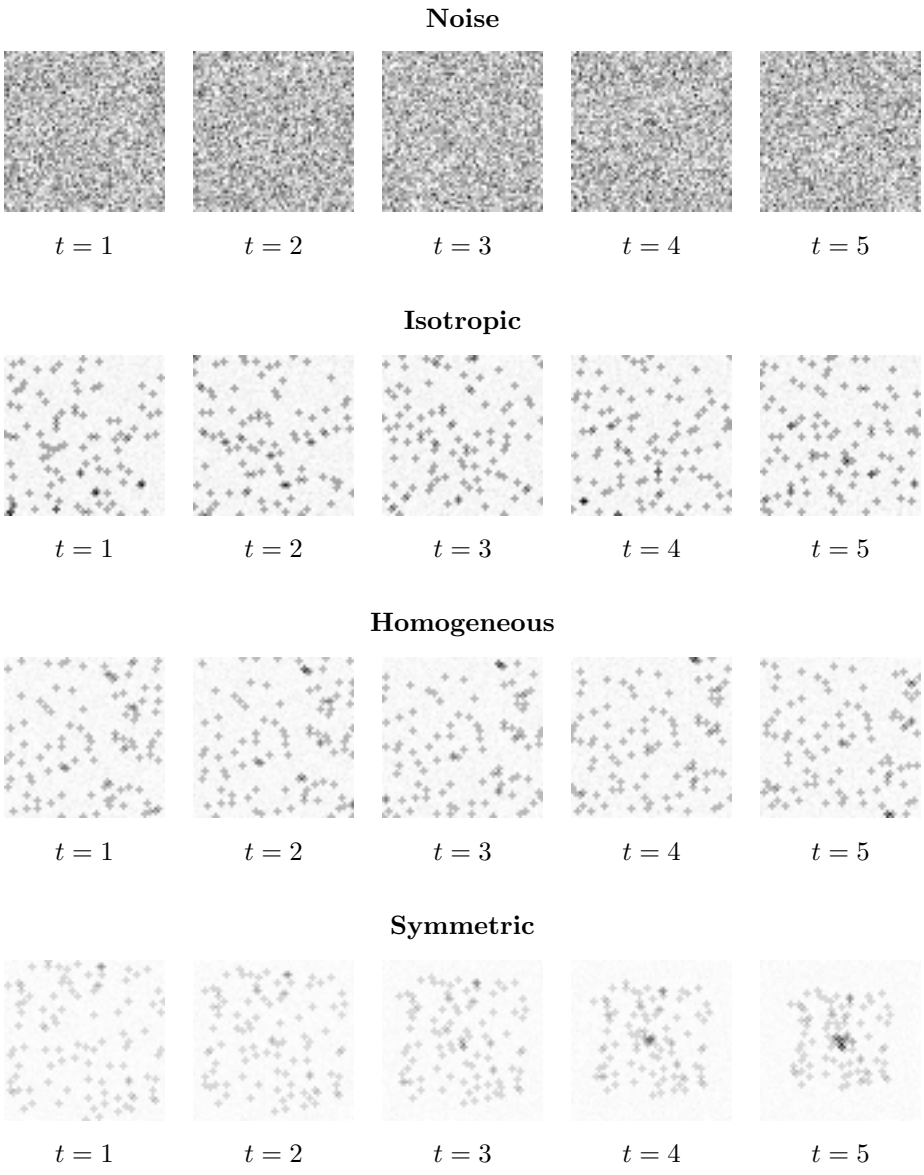


Figure 3: Examples of simulated data. Pixel intensities correspond to values of m, linearly scaled such that the maximum value of m across all time points is black and the value $m = 0$ is white.

4.2 Study Results

4.2.1 Movement pattern estimation

Figure 4 displays a single example of the movements obtained via estimation of $\hat{F}_{s,t}$ and the true movements given by $F_{s,t}$. Data is a single isotropic simulation of 10 objects on a region of size 20 by 20, with the additional condition that objects must remain within the observation window.

Generally good agreement is observed between $\hat{F}_{s,t}$ and $F_{s,t}$, in particular when objects are well separated. In cases where objects overlap and trajectories cross there is a reduction in accuracy which is to be expected based upon the formulation of the estimator $\hat{F}_{s,t}$. Investigation of this example and others suggest that the proposed method of determining $\hat{F}_{s,t}$ results in a reasonable, but by no means perfect, estimator of $F_{s,t}$.

4.2.2 Independent simulations

The first set of tests carried out are for each of the independent simulations under each of the hypotheses H_0^I, H_0^H and H_0^S. These tests are carried out across all four pairs of consecutive time points $1 \to 2$, $2 \to 3$, $3 \to 4$ and $4 \to 5$, and across each of the four considered subregion sizes 20, 15, 12 and 10 corresponding to $w = 9$, 16, 25 and 36. Table 1 summarises the results of these approximate permutation tests under 10 000 permutations, presenting the results as the proportion of the 105 returned between-sample p-values which are less than 0.05. Furthermore, † is used to denote those collections of between-sample p-values which reject the omnibus null hypothesis of Uniform[0,1] distribution of between-sample p-values according to the Kolmogorov-Smirnov test (Massey, 1951) at the five percent level. The issue of multiple testing is discussed in Section 4.2.4.

Noise simulations contain no objects and therefore no structured movements. As a result, local bulk movement patterns, $\hat{S}^{\psi_j}$, are expected to be identically distributed under the action of all operators $\lambda \in \Lambda^I, \Lambda^H$ and Λ^S. Between-sample independence of simulations and within-sample independence of local bulk movement patterns $\hat{S}^{\psi_j}$ by construction therefore suggests that the null hypotheses should be satisfied, reflected in Table 1 by values close to 0.05 and a general absence of markers †. This is the observed behaviour, with the two marked occasions rejecting the omnibus null hypothesis doing so with p-values of 0.013 and 0.009.

Isotropic simulations are also independent and expected to produce local bulk movement patterns $\hat{S}^{\psi_j}$ which are isotropic, homogeneous and symmetric. A difference between noise and isotropic simulations is the movement of objects between subregions, which introduce a within-sample dependence between local bulk movement patterns. The results in Table 1 indicate that there is not enough evidence to reject within-sample independence as the proportion of between-sample p-values less than 0.05 remains approximately 0.05 and the single rejection of the omnibus null hypothesis at the five percent level occurs with p-value 0.024. We therefore conclude that within simulations local bulk movement patterns are approximately independent, an important result for the interpretation of future test results and one that has been shown to hold across all considered subregion sizes.

Homogeneous simulations are independent, producing local bulk movement patterns which are homogeneous but neither isotropic nor symmetric. The result of this is rejection of null hypotheses H_0^I and H_0^S as local bulk movement patterns $\hat{S}^{\psi_j}$ are not identically distributed under the action of operators $\lambda \in \Lambda^I$ and Λ^S. Under H_0^I and H_0^S a generally greater proportion of between-sample

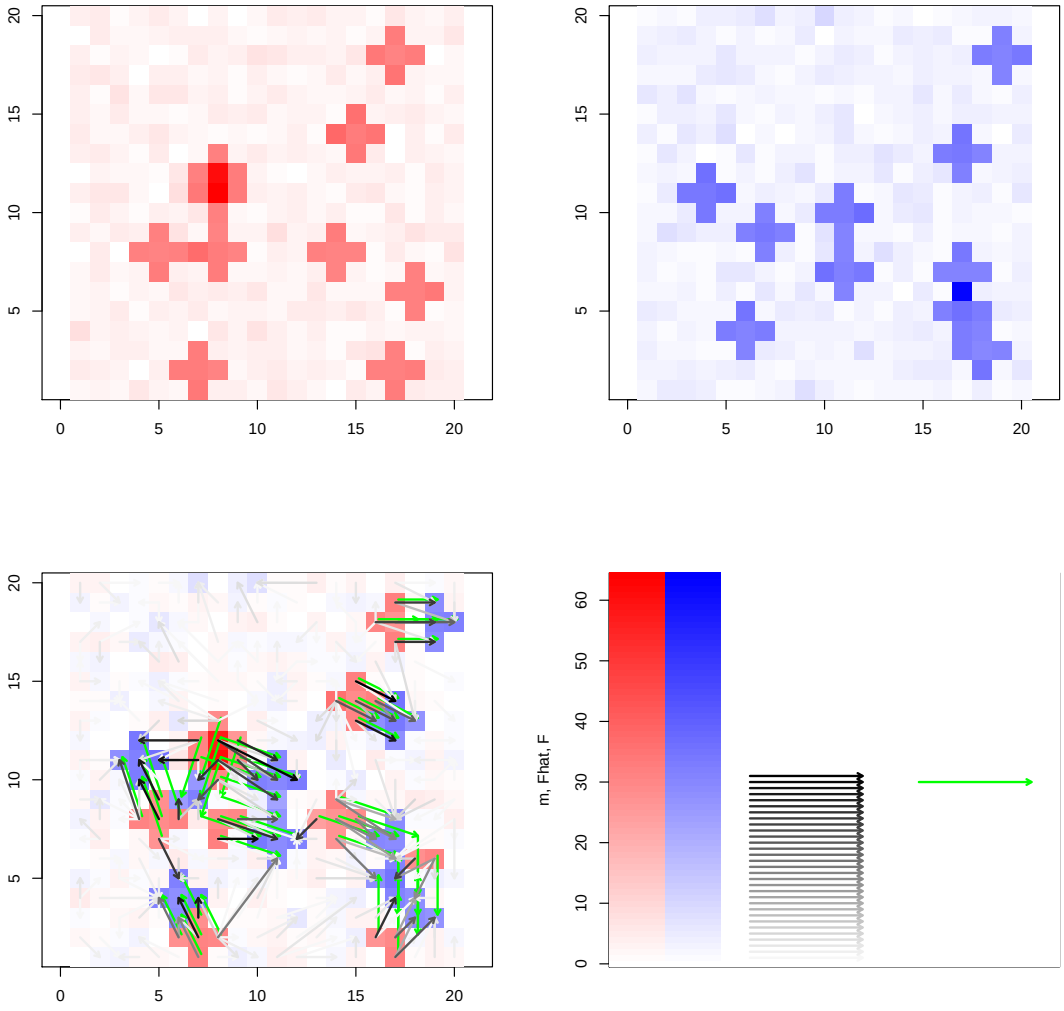


Figure 4: Illustration of observations m_s, top left, m_t, top right, estimated $\hat{F}_{s,t}$ and true $F_{s,t}$, bottom left, for a single simulation. A scale is provided in the bottom right. Simulated movements are represented by green arrows, with grey arrows used to represent estimated movements of varying quantities of mass and mass distributions displayed as shades of red for m_s and blue for m_t. The background image of the bottom left plot displays $m_s - \min\{m_s, m_t\}$ in red and $m_t - \min\{m_s, m_t\}$ in blue, the information on which movements are estimated.

p-values are less than 0.05 for larger subregion sizes, potentially because movement patterns $\hat{F}^{\psi_j}$ are more accurately estimated for larger subregions. Homogeneous simulations are particularly susceptible to inaccuracy in the estimation of $\hat{F}^{\psi_j}$ for small subregion sizes as an object exiting the northeast of the subregion may be replaced by an object entering the southwest of the subregion, resulting in an estimated movement in the opposite direction to that which actually occurred. An increase in significant between-sample p-values for increasing subregion size is counterintuitive in terms of rejection of the null hypothesis due to within-sample dependence, as smaller subregions are expected to increase the proportion of objects crossing boundaries, increasing the within-sample dependence. Homogeneous simulations typically fail to reject the omnibus null hypothesis under H_0^H at the five percent level as expected, with the one exception doing so with p-value 0.027.

Symmetric simulations produce similarly expected results, consistently rejecting H_0^I and H_0^H as local bulk movement patterns $\hat{S}^{\psi_j}$ are not identically distributed under the action of $\lambda \in \Lambda^I$ and Λ^H. When testing against H_0^S, failure to reject the omnibus null hypothesis at the five percent level is also commonly seen, with the two exceptions rejecting with p-values 0.032 and 0.022.

The results of testing independent simulations across a number of hypotheses indicate that if a null hypothesis is chosen which is suitable for the movement patterns of the process, the testing procedure generally returns between-sample p-values whose distribution is indistinguishable from Uniform[0,1]. That is to say, local bulk movement patterns are between-sample independent (by construction), equal in distribution under the action of $\lambda \in \Lambda$ for an appropriate specification of Λ (also by construction) and within-sample independent. This suggests that the proposed testing procedure is valid under considered specification of the null hypothesis, following which rejection of future tests may be taken as evidence against between-sample independence of local bulk movement patterns.

4.2.3 Dependent simulations

The power of the proposed testing procedure is assessed by testing dependent simulations under the most appropriate choice of null hypothesis. That is isotropic 10 and isotropic 30 simulations under H_0^I, homogeneous 10 and homogeneous 30 simulations under H_0^H and symmetric 10 and symmetric 30 simulations under H_0^S, where each test is an approximate permutation test using 10 000 random samples from the corresponding set of operators Λ. Table 2 summarises the results of these tests, presenting the proportion of the 105 between-sample p-values which are less than 0.05 and marking with ‡ those which fail to reject the omnibus null hypothesis of Uniform[0,1] distribution of between-sample p-values under the Kolmogorov-Smirnov test at the five percent level. Tests are again carried out over four pairs of consecutive time points and four subregion sizes varying between 10 and 20.

Over all comparisons the testing procedure is generally able to detect dependence when it exists, observed as generally consistent rejection of the omnibus null hypothesis at the five percent level. This is always the case for the simulations sharing 30 out of the 100 total objects, but there are some failures to reject the omnibus null hypothesis at the five percent level when the degree of dependence is weaker and only 10 objects are shared. For isotropic 10 simulations the failure to reject the omnibus null hypothesis occurs with a p-value of 0.739. For homogeneous 10 simulations there are more cases, failing to reject the omnibus null hypothesis with p-values 0.201, 0.142, 0.145 and 0.439. For symmetric 10 simulations we again see a few failures to reject the omnibus null

		H_0^I				H_0^H				H_0^S			
		20	15	12	10	20	15	12	10	20	15	12	10
noise	$1 \to 2$	0.03	0.06	0.06	$0.04^\dagger$	0.02	0.08	0.06	$0.02^\dagger$	0.03	0.06	0.05	0.05
	$2 \to 3$	0.06	0.03	0.08	0.06	0.06	0.02	0.06	0.06	0.07	0.02	0.06	0.03
	$3 \to 4$	0.04	0.04	0.04	0.10	0.06	0.04	0.05	0.09	0.05	0.07	0.04	0.05
	$4 \to 5$	0.04	0.04	0.03	0.05	0.04	0.04	0.04	0.04	0.02	0.02	0.04	0.04
isotropic	$1 \to 2$	0.06	0.10	0.07	0.03	0.08	0.09	0.07	0.03	0.07	0.04	0.03	0.03
	$2 \to 3$	0.05	0.04	0.03	0.07	0.05	0.01	0.04	0.06	0.08	0.02	0.04	0.06
	$3 \to 4$	0.06	0.07	0.10	0.08	0.05	0.06	$0.08^\dagger$	0.06	0.06	0.04	0.08	0.05
	$4 \to 5$	0.05	0.03	0.06	0.05	0.04	0.03	0.06	0.05	0.05	0.02	0.04	0.05
homogeneous	$1 \to 2$	$0.88^\dagger$	$0.73^\dagger$	$0.62^\dagger$	$0.44^\dagger$	0.05	0.02	0.05	0.06	$0.90^\dagger$	$0.71^\dagger$	$0.67^\dagger$	$0.52^\dagger$
	$2 \to 3$	$0.53^\dagger$	$0.40^\dagger$	$0.39^\dagger$	$0.41^\dagger$	0.04	0.03	0.09	0.05	$0.6^\dagger$	$0.40^\dagger$	$0.36^\dagger$	$0.39^\dagger$
	$3 \to 4$	$0.80^\dagger$	$0.70^\dagger$	$0.56^\dagger$	$0.55^\dagger$	0.03	0.06	0.03	0.03	$0.79^\dagger$	$0.70^\dagger$	$0.65^\dagger$	$0.60^\dagger$
	$4 \to 5$	$0.70^\dagger$	$0.57^\dagger$	$0.57^\dagger$	$0.33^\dagger$	$0.04^\dagger$	0.03	0.05	0.03	$0.70^\dagger$	$0.57^\dagger$	$0.54^\dagger$	$0.33^\dagger$
symmetric	$1 \to 2$	$.00^\dagger$	$1.00^\dagger$	$0.98^\dagger$	$0.94^\dagger$	$1.00^\dagger$	$1.00^\dagger$	$0.98^\dagger$	$0.94^\dagger$	0.02	0.08	0.07	0.06
	$2 \to 3$	$1.00^\dagger$	$1.00^\dagger$	$0.91^\dagger$	$0.98^\dagger$	$1.00^\dagger$	$1.00^\dagger$	$0.93^\dagger$	$0.98^\dagger$	0.08	$0.07^\dagger$	0.07	0.05
	$3 \to 4$	$0.99^\dagger$	$0.96^\dagger$	$0.99^\dagger$	$1.00^\dagger$	$1.00^\dagger$	$0.96^\dagger$	$0.99^\dagger$	$1.00^\dagger$	0.08	0.03	0.06	0.06
	$4 \to 5$	$0.98^\dagger$	$1.00^\dagger$	$1.00^\dagger$	$1.00^\dagger$	$1.00^\dagger$	$1.00^\dagger$	$1.00^\dagger$	$1.00^\dagger$	0.09	0.07	0.10	$0.05^\dagger$

Table 1: The results of testing each class of independent simulations under each of the three hypotheses H_0^I, H_0^H and H_0^S for a range of consecutive time points and subregion sizes. Comparison between the 15 simulations in each class produces 105 between-sample p-values for each test, with the table presenting the proportion of p-values less than 0.05 and † used to denote sets of p-values which reject the omnibus hypothesis of Uniform[0,1] distribution of between-sample p-values under the Kolmogorov-Smirnov test at the five percent level.

hypothesis, with p-values 0.935, 0.627, 0.081, 0.273 and 0.116.

Although the omnibus null hypothesis is typically rejected at the five percent level, it is by no means the case that every between-sample test rejects the respective null hypothesis at the five percent level, with the proportions of tests that do so varying between 0.04 and 0.90. This is to be expected due to the difficulty of the problem and the general solution that we propose, requiring only minor assumptions on movement patterns to determine exchangeability of local bulk movement patterns. Despite this, when the optimal choice of subregion size is made and the dependence is 30 percent of all objects the proportion of between-sample tests reported significant at the five percent level is on average 0.77, which we take as an indicator that the proposed testing procedure has an acceptable level of power.

There is almost unanimous improvement in detection ability, measured by an increase in the proportion of between-sample p-values which are less than 0.05, with decreasing subregion size. As subregion size reduces, the expected number of independent objects in each subregion is reduced. The relative contribution to $\hat{S}^{\psi_j}$ by shared objects in the subregions in which they are located is therefore increased, making their existence more easily detectable under the testing procedure. Despite this, there is expected to be a limit below which further reduction of subregion sizes will result in a reduction in power, as the accuracy of the estimator $\hat{F}^{\psi_j}$ is reduced as the probability of objects moving between subregions increases.

In general the proportion of between-sample p-values less than 0.05 is greater for isotropic simulations than for both homogeneous and symmetric simulations, for which the proportions are broadly comparable. A possible explanation for this may be that for homogeneous and symmetric simulations movements are determined solely by the location of objects, with closely separated objects undergoing very similar movements. This could result in greater numbers of coincidentally similar movements, making dependencies more difficult to identify for homogeneous and symmetric simulations in comparison to isotropic simulations for which movement direction is independent of object location and coincident similarities are as a result rarer.

There is expected to be some dependence between tests at different pairs of consecutive time points, but the form of this dependence is difficult to quantify. For isotropic and homogeneous simulations local bulk movement patterns are expected to be identically distributed in time and observed results are similarly consistent over time. This is not the case for symmetric simulations, as objects aggregate at the centre of Ψ over time. Despite this, there is no discernible difference in results for symmetric simulations across different pairs of consecutive time points, suggesting that whether movements are spread over Ψ or congregated closer to its centre has no impact on the performance of the testing procedure.

The power of the proposed testing procedure is further assessed by testing dependent isotropic simulations under the three null hypotheses H_0^I, H_0^H and H_0^S. Rejection of any of these hypotheses may be taken as evidence of between-sample dependence because local bulk movement patterns for isotropic simulations are exchangeable under operators $\lambda \in \Lambda^I$ and both sets of operators $\Lambda^H \subset \Lambda^I$ and $\Lambda^S \subset \Lambda^I$ by construction. Table 3 summarises the results of these tests, presenting the proportion of the 105 between-sample p-values which are less than 0.05 and marking with ‡ those which fail to reject the omnibus null hypothesis of Uniform[0,1] distribution of between-sample p-values under the Kolmogorov-Smirnov test at the five percent level. These test are again carried out over the four pairs of consecutive time points and four subregion sizes varying between

		20	15	12	10
isotropic 10	$1 \to 2$	0.19	0.11	0.23	0.26
	$2 \to 3$	0.10	0.12	0.10	0.17
	$3 \to 4$	$0.07^{\ddagger}$	0.20	0.10	0.23
	$4 \to 5$	0.10	0.11	0.18	0.23
isotropic 30	$1 \to 2$	0.37	0.47	0.84	0.90
	$2 \to 3$	0.47	0.40	0.83	0.88
	$3 \to 4$	0.49	0.51	0.93	0.90
	$4 \to 5$	0.57	0.33	0.84	0.88
homogeneous 10	$1 \to 2$	$0.03^{\ddagger}$	$0.12^{\ddagger}$	0.13	0.13
	$2 \to 3$	0.15	0.16	0.22	0.30
	$3 \to 4$	$0.05^{\ddagger}$	0.06	$0.10^{\ddagger}$	0.13
	$4 \to 5$	0.10	0.11	0.09	0.21
homogeneous 30	$1 \to 2$	0.37	0.40	0.62	0.80
	$2 \to 3$	0.24	0.63	0.50	0.90
	$3 \to 4$	0.16	0.45	0.50	0.73
	$4 \to 5$	0.20	0.51	0.67	0.69
symmetric 10	$1 \to 2$	$0.05^{\ddagger}$	0.11	0.09	0.16
	$2 \to 3$	$0.07^{\ddagger}$	0.10	0.16	0.17
	$3 \to 4$	$0.06^{\ddagger}$	0.10	0.10	0.09
	$4 \to 5$	$0.04^{\ddagger}$	0.12	$0.10^{\ddagger}$	0.06
symmetric 30	$1 \to 2$	0.21	0.39	0.53	0.60
	$2 \to 3$	0.24	0.48	0.37	0.53
	$3 \to 4$	0.19	0.26	0.77	0.70
	$4 \to 5$	0.27	0.30	0.40	0.69

Table 2: The results of testing each class of dependent simulations against the most appropriate hypothesis for a range of consecutive time points and subregion sizes. Comparison between the 15 simulations in each class produces 105 between-sample p-values for each test, with the table presenting the proportion of p-values less than 0.05 and ‡ used to denote sets of p-values which fail to reject the omnibus null hypothesis of Uniform[0,1] distribution of between-sample p-values under the Kolmogorov-Smirnov test at the five percent level.

10 and 20.

Across all null hypotheses there is consistent rejection of omnibus null hypothesis at the five percent level. The exceptions to this fail to reject the omnibus null hypothesis with p-values of 0.739 for H_0^I, 0.824 for H_0^H and 0.624 for H_0^S. The proportion of between-sample tests reporting significance at the five percent significance level is greatest under H_0^I, but only mildly greater than under H_0^H which is in turn mildly greater than under H_0^S. These two properties support the effectiveness of the testing procedure using any suitably large valid set of operations Λ under which local bulk movement patterns are exchangeable, with a minor reduction in power in comparison to the test carried out under the maximal set of operations Λ under which local bulk movement patterns are exchangeable.

4.2.4 Multiple testing

Both the testing of simulations in this section and the testing of real data in the following section raise questions of multiple comparisons. Focusing on a single example, the results presented in Table 1 for the testing of noise simulations under H_0^I are based upon tests of 105 between-sample comparisons across four subregion sizes and four pairs of consecutive time points, for a total of 1 680 statistical tests. Considering the testing of the 16 omnibus null hypotheses via the Kolmogorov-Smirnov test at the five percent level, we observe one rejection of the null hypothesis. However, this individual result must be considered in the wider picture of all sixteen omnibus tests.

The family wise error rate of a collection of statistical tests is the probability of making one or more type I errors, rejecting the null hypothesis when it is true, when performing multiple hypothesis tests (Shaffer, 1995). Assuming the previously described 16 Kolmogorov-Smirnov tests at the five percent level are independent, the family wise error rate is then $1-(1-0.05)^{16}=0.56$. That is, even if between-sample p-values are Uniform[0,1] distributed under H_0^I we have a greater than 50 percent chance of rejecting at least one of the 16 tests. Our conclusions must be aware of this fact and appropriately take it into account.

A method for the control of the family wise error rate across a total of n tests is the Bonferroni correction, for which each individual test is carried out at the reduced significance level of α/n (Dunn, 1961). The family wise error rate in the case of independent tests, $1-(1-\alpha/n)^n \leq \alpha$ by Boole's inequality, is controlled by the Bonferroni correction to be no more than the significance level α. However, for large numbers of comparisons the Bonferroni correction can result in tests which lack power and are conservative when the results of individual tests are positively correlated (Simes, 1986).

In all tests of the omnibus null hypothesis we expect strong positive correlation between test results, the exact form of which is difficult to quantify. This is because tests across different subregion sizes are applied to the same simulation data each time, and simulation data is identically distributed across all time points, indicating that tests between $s=1,t=2$ and $s=2,t=3$ are expected to produce very similar results. The Bonferroni correction and alternative corrections are therefore not applied, with tests carried out at the nominal five percent level but resulting conclusions made in light of the presence of multiple testing.

		isotropic				homogeneous				symmetric			
		20	15	12	10	20	15	12	10	20	15	12	10
isotropic 10	$1 \to 2$	0.19	0.11	0.23	0.26	0.20	0.11	0.24	0.26	0.17	0.15	0.18	0.22
	$2 \to 3$	0.10	0.12	0.10	0.17	0.13	0.14	0.10	0.16	0.08	0.11	0.06	0.15
	$3 \to 4$	$0.07^{\ddagger}$	0.20	0.10	0.23	$0.06^{\ddagger}$	0.19	0.10	0.22	$0.05^{\ddagger}$	0.20	0.09	0.19
	$4 \to 5$	0.10	0.11	0.18	0.23	0.10	0.13	0.22	0.26	0.10	0.10	0.15	0.17
isotropic 30	$1 \to 2$	0.37	0.47	0.84	0.90	0.27	0.48	0.81	0.89	0.30	0.46	0.76	0.90
	$2 \to 3$	0.47	0.40	0.83	0.88	0.47	0.43	0.84	0.89	0.41	0.39	0.83	0.76
	$3 \to 4$	0.49	0.51	0.93	0.90	0.30	0.54	0.95	0.84	0.44	0.50	0.75	0.79
	$4 \to 5$	0.57	0.33	0.84	0.88	0.41	0.30	0.83	0.79	0.56	0.32	0.73	0.78

Table 3: The results of testing dependent isotropic simulations against the three appropriate hypothesis for a range of consecutive time points and subregion sizes. Comparison between the 15 simulations in each class produces 105 p-values for each test, with the table presenting the proportion of p-values less than 0.05 and ‡ used to denote sets of p-values which fail to reject the omnibus null hypothesis of Uniform[0,1] distribution under the Kolmogorov-Smirnov test at the five percent level.

5 Investigation of EB3 and TACC3 data

5.1 EB3 and TACC3 background

As stated in the introduction, we aim to make inference on dependence between the local bulk movement patterns of biomolecular species TACC3, Transforming Acidic Coiled-Coil Containing Protein 3, and EB3, End-Binding protein 3. The protein EB3 is known to localise at the tip of growing microtubules during mitosis (Mimori-Kiyosue et al., 2000) and the biological question of interest is whether TACC3 is similarly located so as to shed light on its potential impact on the process of mitosis.

The available data is comprised of confocal fluorescence microscopy images collected across seven samples at a total number of between 47 and 57 time points. Images are collected of live cells during mitosis with TACC3 tagged with a green fluorescing protein and EB3 tagged with a red fluorescing protein. Microscope resolution is such that each pixel is 68.9nm square and images are collected at a rate of one per second. Green and red intensities are recorded at the same time by two different digital cameras, resulting in two greyscale images for each sample at each time point, one corresponding to EB3 and one to TACC3. Digital cameras are used for image acquisition, meaning that the data contained in each image is accessible as a matrix of integer valued intensity levels, one for each pixel location. We treat intensity levels as a surrogate measure for quantity of the relevant biomolecular species at each pixel location. Example images from five consecutive time points for three of the samples may be seen in Figure 5, illustrating the structure seen, variability within samples between time points and between samples more generally.

The proposed testing methodology requires an assumption under which the subregions ψ^j may be rearranged and remain identically distributed. As EB3 is located at the end of microtubules which grow in a spindle structure during mitosis, we make the assumption that movement patterns are symmetric across the line connecting microtubule organising centres, the poles of the approximately ellipsoid spindle structure. Original image sequences are therefore rotated and cropped to focus on only the spindle region, resulting in consistent image sizes of 180 pixels wide and 240 pixels high across all samples, permitting exact tessellation by square subregions ψ^j of side lengths 20, 15 and 12. The miotic spindle is a three dimensional structure, with the resulting two dimensional images a projection into a single plane.

Collection of images every second is expected to capture the location of biomolecules at a great enough time resolution that their movement patterns may be estimated. In investigation of the same data (Gutierrez-Caballero et al., 2015) biomolecule locations were specified and tracked using an automatic object tracking process requiring specification of a number of parameters tuned based upon the observed images and prior beliefs about the biology underpinning the process (Applegate et al., 2011). Results indicated that both biomolecular species are located at the end of growing microtubules, separated by an average distance of 229nm, with TACC3 closest to the growing tip. We propose to investigate the same data set without specification of parameters other than the assumption of reflection symmetry across a single line and further statistically test for dependence between TACC3 and EB3 movement patterns.

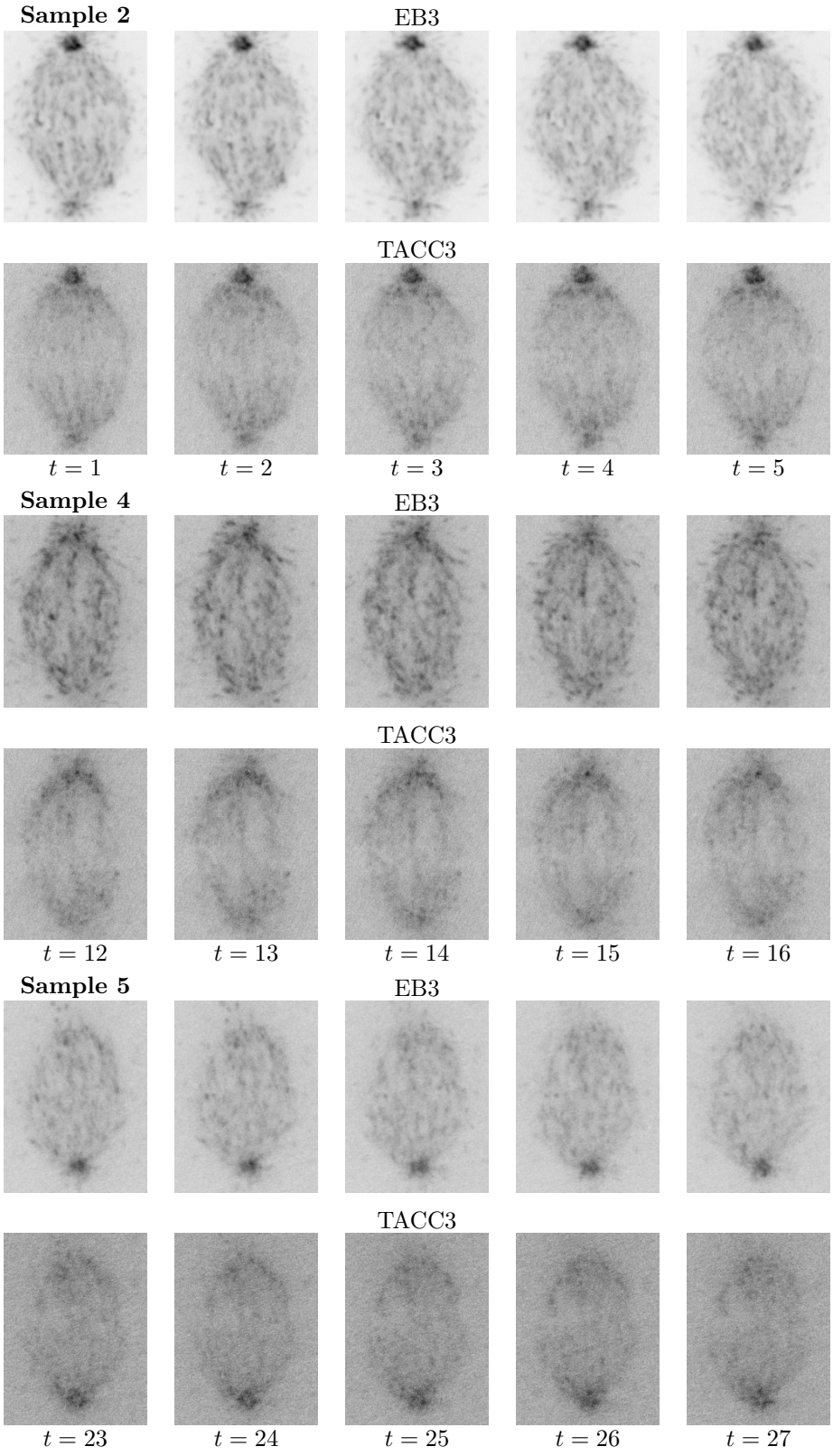


Figure 5: EB3 and TACC3 images from three samples across five consecutive time points. Pixel intensities correspond to values of m with the maximum value of m across all time points black and $m = 0$ white.

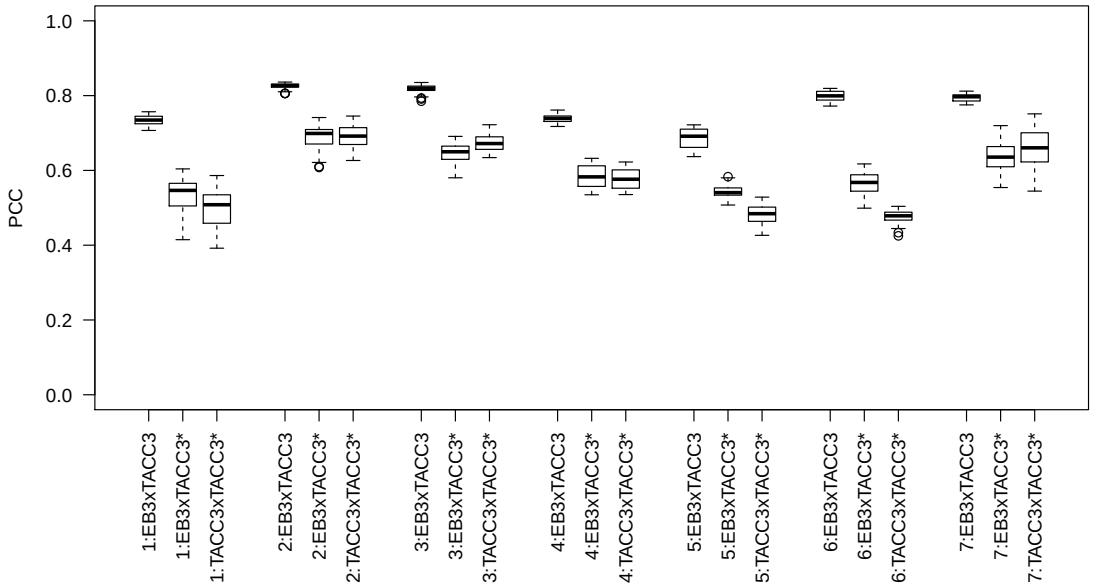


Figure 6: Boxplots to display the distribution of values of Pearson's correlation coefficient across all pairs of consecutive time points for each of the sample comparisons considered.

5.2 Exploratory data analysis

Intensities related to EB3 are in general greater than that of the TACC3 channel due to expression of TACC3 at a lower level. Greater expression of TACC3 is avoided, as it results in a brighter image but also aggregation of TACC3 away from microtubule tips, and consequently away from EB3, which could obscure the ability to investigate interaction between TACC3 and EB3. Estimation of movement patterns for TACC3 is therefore expected to be more challenging as intensities are in some cases on a similar scale to background noise. Consistent scaling of intensity in a subregion, $m(\psi)$, by a positive constant across time points s and t results in an identical scaling of the summary statistic, $\hat{S}^{\psi}_{s,t}$, and a change to the comparison score $\hat{G}^{\psi,\psi}_{s,t}$. However, if the scaling is consistent across all subregions and all time points then the impact on test results should be minimal.

As an exploratory investigation, colocalisation between image pairs may be quantified using Person's correlation coefficient, the results of which are displayed in Figure 6. Calculations are made for comparison pairs EB3 $\times$ TACC3, for which we are interested in the degree of similarity, and EB3 $\times$ TACC3* and TACC3 $\times$ TACC3*, where TACC3* is the vertical reflection of TACC3 and we expect to see only coincidental similarity. For each of the seven samples correlation values are greater for EB3 $\times$ TACC3 than the alternatives, indicating that there is more than coincidental similarity between EB3 and TACC3. However, without a methodology for quantifying the significance of obtained correlation values, it is impossible to conclude that there is dependence between the distribution of EB3 and TACC3.

A major foundation of the estimator $\hat{F}^{\psi}_{s,t}$ is that $m_s(\psi)$ and $m_t(\psi)$ represent distributions of approximately the same masses at two different time points. In the context of the EB3 and TACC3 image data this equates to minimal changes in intensity on a subregion by subregion basis, caused

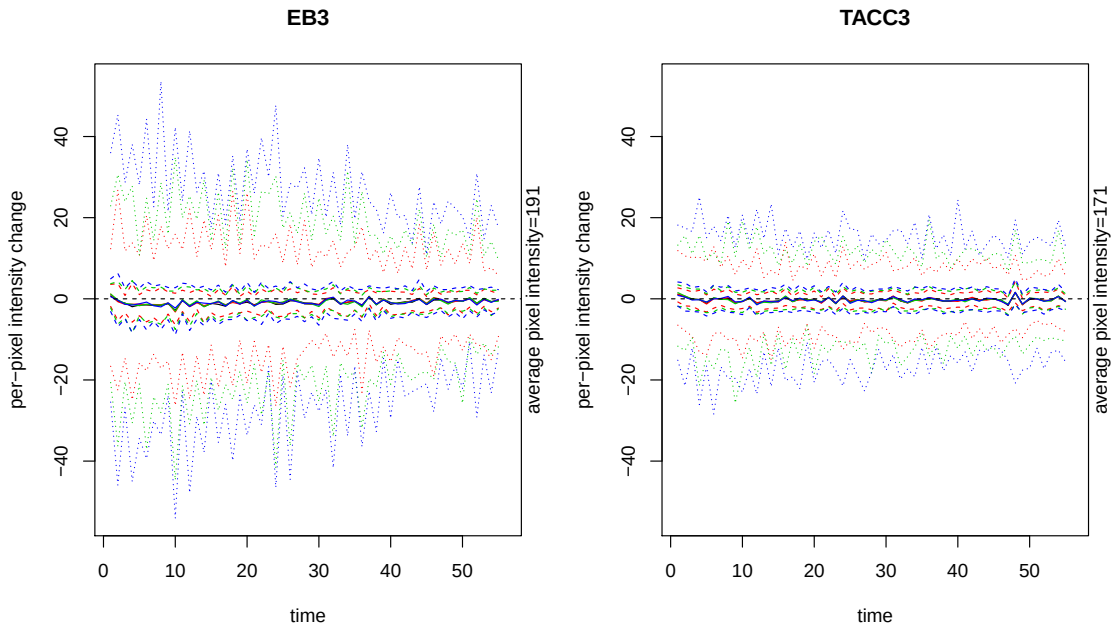


Figure 7: Variation in region intensity over time on a per-pixel scale for EB3, left, and TACC3, right. Solid black line denotes the mean intensity change over Ψ, with solid red, green and blue lines corresponding to medians over subregions ψ_j of size 20, 15 and 12 respectively. Coloured dashed lines denote upper and lower quartiles over the collection of subregions ψ_j, with dotted lines denoting maximum and minimum differences. Horizontal black dashed line denotes no change in average intensity.

by biomolecules moving between subregions, and on a whole image basis, caused by photobleaching over time. To investigate the validity of these assumptions we plot changes in pixel average intensity between consecutive time points

$$\frac{1}{|\psi|}\sum_{x\in\psi} m_{t+1}(x) - m_t(x),$$

for subregions ψ_j of sizes 20, 15 and 12 and for the whole image space Ψ, the results of which may be seen for sample four in Figure 7.

Average intensity plots indicate that there is a consistent but minor reduction in intensity across the whole space between consecutive time points, consistent with photobleaching, but not believed to be significant enough to render the estimation of $\hat{F}$ as unreliable. There are sometimes large changes in subregion intensity between consecutive time points, corresponding to particularly large concentrations of biomolecules moving between subregions, but the vast majority of changes on an individual pixel scale are small in comparison to average pixel intensity, indicating that movements are largely contained within subregions. Predictably, larger intensity changes are observed for smaller subregion sizes as the proportion of locations $x \in \psi_j$ which are close to the subregion boundary increases, meaning biomolecules are more likely to move between subregions.

5.3 Application of proposed methodology

As described in the previous section, we make three comparisons between EB3, TACC3 and TACC3*, the vertical reflection of TACC3 across the horizontal line through the centre of Ψ. Testing is carried out as an approximate permutation test of 10 000 random permutations under H_0^H, which specifies on a scale determined by the number of subregions, w, between-sample independence of local bulk movement patterns, identical distribution of the collection of $S_{s,t}^{\psi_j}$ under the action of any $\lambda \in \Lambda^H$ and within-sample independence of local bulk movement patterns.

Test results are presented in Table 4 as the proportion of between sample comparisons across all consecutive time points reporting p-values of less than 0.05, with † used to denote those collections of p-values which reject the omnibus null hypothesis of Uniform[0,1] distribution of between-sample comparisons under the Kolmogorov-Smirnov test at the five percent level. Note the difference between presentation of EB3 and TACC3 results in comparison to simulation results - here collections of p-values tested against the omnibus hypothesis are combined across the between 46 and 56 pairs of consecutive time points, while for simulation data they were collected across the 105 pairs of between-sample comparisons at a single pair of consecutive time points.

Rejection of the null hypothesis is expected to occur if any of the three components of H_0^H are not met, however, the biological question of interest relates only to the first condition of between-sample independence of local bulk movement patterns. It is for this reason that the comparison between EB3 and TACC3* is made, for which between-sample independence is expected by construction and rejection of H_0^H may be attributed to improper specification of Λ^H or within-sample dependence of local bulk movement patterns. The results indicate that the omnibus null hypothesis is never rejected at the five percent level for comparison between EB3 and TACC3, suggesting that there is insufficient evidence to refute the assumption of horizontally symmetric bulk movement patterns and within-sample independence. The omnibus null hypothesis is rejected in some instances of comparison between TACC3 and TACC3*, but these may be attributed to between-sample dependencies in particular for subregions along the vertical centre of Ψ which may be compared to reflections of themselves under $\lambda \in \Lambda^H$.

Across all seven samples the omnibus null hypothesis comparing EB3 and TACC3 is rejected at the five percent level, taken as strong evidence of between-sample dependence in local bulk movement patterns between EB3 and TACC3. The proportion of between-sample tests reporting p-values less than 0.05 varies between 0.36 and 1.00, rising to between 0.58 and 1.00 when subregions of the smallest size, 12, are considered. This is evidence of regular rejection of H_0^H, improving with reduction in subregion size but not only detectable at a single subregion size which must be accurately specified. The effect of varying subregion sizes is similar to that observed for simulated data sets.

Investigation of the distribution of between-sample p-values across the range of consecutive time points shows that non-significant values are interspersed within significant values at the five percent level. This suggests that rather than periods of dependence and periods of independence of local bulk movement patterns, between some pairs of time points movement estimation may be particularly inaccurate as objects of considerable intensity move between subregions and resulting in an insignificant test result.

		20	15	12
Sample 1	EB3 × TACC3	0.48†	0.55†	0.68†
	EB3 × TACC3*	0.05	0.07	0.04
	TACC3 × TACC3*	0.09	0.14	0.12†
Sample 2	EB3 × TACC3	0.85†	0.98†	1.00†
	EB3 × TACC3*	0.11	0.07	0.07
	TACC3 × TACC3*	0.11	0.20	0.22†
Sample 3	EB3 × TACC3	0.53†	0.60†	0.69†
	EB3 × TACC3*	0.02	0.05	0.04
	TACC3 × TACC3*	0.18	0.25†	0.25†
Sample 4	EB3 × TACC3	0.36†	0.67†	0.71†
	EB3 × TACC3*	0.05	0.07	0.07
	TACC3 × TACC3*	0.11	0.09	0.22†
Sample 5	EB3 × TACC3	0.47†	0.51†	0.58†
	EB3 × TACC3*	0.02	0.07	0.05
	TACC3 × TACC3*	0.11†	0.13	0.11
Sample 6	EB3 × TACC3	0.78†	0.85†	0.93†
	EB3 × TACC3*	0.05	0.07	0.07
	TACC3 × TACC3*	0.09	0.18	0.16
Sample 7	EB3 × TACC3	0.47†	0.53†	0.67†
	EB3 × TACC3*	0.00	0.05	0.11
	TACC3 × TACC3*	0.09†	0.11	0.09†

Table 4: Comparison of seven samples of biological data, each of which comprises two sets of images representing the locations of EB3 and TACC3 over between 47 and 57 time points and a resulting third set, TACC3*, which is the vertical reflection of the TACC3 data. Table values are the proportion of between-sample p-values under H_0^H which are significant at the five percent level for subregion sizes varying between 20 and 10. † is used to indicate the collections of between-sample p-values which reject the omnibus null hypothesis of Uniform[0,1] distribution at the five percent level according to a Kolmogorov-Smirnov test. Testing is carried out using approximate permutation tests with 10 000 permutations.

6 Conclusions

We have proposed a method for the estimation of local bulk movement patterns within two samples and a testing procedure for quantifying the significance of the dependence between these patterns. The procedure relies upon a minimal set of assumptions, namely consistency of total mass within the whole space Ψ, the ability to define identically sized and shaped subregions $\psi_j \subseteq \Psi$ for which movements of mass between subregions is minimal in comparison to movements within subregions and symmetry or otherwise of subregion movement patterns such that they may be transformed and rearranged and remain identically distributed. These assumptions are generally easier to satisfy for observations derived from physical phenomena where observations have been collected at a sufficiently high time resolution.

Application of the proposed methodology to simulated data for which movement patterns are independent results in tests which reject correctly specified null hypotheses at a rate consistent with that of the size of the test, supporting the theoretical validity of the testing procedures. In cases of incorrectly specified null hypotheses, tests may reject the null hypothesis when movement patterns are independent as the null hypothesis is a composite of three statements and failure to satisfy any one is sufficient for rejection of the null hypothesis.

Application of the proposed methodology to simulated data where movement patterns are partially dependent often results in rejection of correctly specified null hypotheses, but due to the complexities of the problem and the minimal set of assumptions is not perfect. In the case of 10 percent dependence between movement patterns, correctly specified null hypotheses are rejected at the five percent level typically between 10 and 30 percent of the time. In the case of 30 percent dependence between movement patterns, correctly specified null hypotheses are rejected at the five percent level between 20 and 90 percent of the time. In almost all cases, the collection of between-sample p-values rejects the omnibus null hypothesis of Uniform[0,1] distribution according to the Kolmogorov-Smirnov test at the five percent level, indicating that the omnibus hypothesis may be useful to detect dependence when comparing over multiple samples or time points.

The statistical testing procedure compares local bulk movement patterns, where local is on a scale defined by the size of subregions which are inversely proportional to the number of subregions. In application to simulated data the test is more powerful for smaller subregion sizes, indicating that dependencies between movement patterns may be insignificant when summaries are made over larger areas.

Analysis of a biological data set to compare movement patterns of EB3 and TACC3 biomolecule species during mitosis reports consistent rejection of the omnibus null hypothesis at the five percent level. Further, there is a general failure to reject the omnibus hypothesis when EB3 observations are compared to a transformation of TACC3 data by vertical reflection. Taken in combination, these results suggest that null hypotheses are in this instance rejected on the basis of dependence between movement patterns rather than due to a lack of exchangeability within each observation under the specified set of transformations and rearrangement of subregions. As in the case of simulated data, an increase in the proportion of significant between-sample p-values is seen as subregion size decreases. Our analysis therefore supports the previous work that the movement patterns of EB3 and TACC3 are dependent, potentially through their localisation on the tips of growing microtubules.

7 Acknowledgements

We gratefully acknowledge Dr. Stephen Royle and his research group within the Centre for Mechanochemical Cell Biology at the University of Warwick for making the EB3 and TACC3 data available to us and for discussions on its biological background.

References

J. Adler and I. Parmryd. Quantifying colocalization by correlation: The Pearson correlation coefficient is superior to the Mander's overlap coefficient. *Cytometry Part A*, 77A(8):733–742, March 2010. doi: 10.1002/cyto.a.20896.

J. Adler and I. Parmryd. Colocalization analysis in fluorescence microscopy. In *Methods in Molecular Biology*, pages 97–109. Springer Nature, 2012. doi: 10.1007/978-1-62703-056-4_5.

K. T. Applegate, S. Besson, A. Matov, M. H. Bagonis, K. Jaqaman, and G. Danuser. plusTipTracker: Quantitative image analysis software for the measurement of microtubule dynamics. *Journal of Structural Biology*, 176(2):168–184, November 2011. doi: 10.1016/j.jsb.2011.07.009.

D. G. Booth, F. E. Hood, I. A. Prior, and S. J. Royle. A TACC3/ch-TOG/clathrin complex stabilises kinetochore fibres by inter-microtubule bridging. *The EMBO Journal*, 30(5):906–919, February 2011. doi: 10.1038/emboj.2011.15.

N. Chenouard, I. Smal, F. de Chaumont, M. Maška, I. F. Sbalzarini, Y. Gong, J. Cardinale, C. Carthel, S. Coraluppi, M. Winter, A. R. Cohen, W. J. Godinez, K. Rohr, Y. Kalaidzidis, L. Liang, J. Duncan, H. Shen, Y. Xu, K. E. G. Magnusson, J. Jaldén, H. M. Blau, P. Paul-Gilloteaux, P. Roudot, C. Kervrann, F. Waharte, J.-Y. Tinevez, S. L. Shorte, J. Willemse, K. Celler, G. P. van Wezel, H.-W. Dan, Y.-S. Tsai, C. O. de Solórzano, J.-C. Olivo-Marin, and E. Meijering. Objective comparison of particle tracking methods. *Nature Methods*, 11(3): 281–289, January 2014. doi: 10.1038/nmeth.2808.

R. M. Clegg. Fluorescence resonance energy transfer. *Current Opinion in Biotechnology*, 6(1): 103–110, January 1995. doi: 10.1016/0958-1669(95)80016-6.

S. V. Costes, D. Daelemans, E. H. Cho, Z. Dobbin, G. Pavlakis, and S. Lockett. Automatic and quantitative measurement of protein-protein colocalization in live cells. *Biophysical Journal*, 86 (6):3993–4003, June 2004. doi: 10.1529/biophysj.103.038422.

O. J. Dunn. Multiple comparisons among means. *Journal of the American Statistical Association*, 56(293):52–64, March 1961. doi: 10.1080/01621459.1961.10482090.

E. S. Edgington. Randomization tests. *The Journal of Psychology*, 57(2):445–449, April 1964. doi: 10.1080/00223980.1964.9916711.

E. S. Edgington. Approximate randomization tests. *The Journal of Psychology*, 72(2):143–149, July 1969. doi: 10.1080/00223980.1969.10543491. URL `http://dx.doi.org/10.1080/00223980.1969.10543491`.

T. R. Fanshawe and P. J. Diggle. Bivariate geostatistical modelling: a review and an application to spatial variation in radon concentrations. *Environmental and Ecological Statistics*, 19(2): 139–160, August 2011. doi: 10.1007/s10651-011-0179-7.

A. L. Gibbs and F. E. Su. On choosing and bounding probability metrics. *International Statistical Review*, 70(3):419–435, December 2002. doi: 10.1111/j.1751-5823.2002.tb00178.x.

C. R. Givens and R. M. Shortt. A class of Wasserstein metrics for probability distributions. *The Michigan Mathematical Journal*, 31(2):231–240, 1984. doi: 10.1307/mmj/1029003026.

C. Gutierrez-Caballero, S. G. Burgess, R. Bayliss, and S. J. Royle. TACC3-ch-TOG track the growing tips of microtubules independently of clathrin and aurora-a phosphorylation. *Biology Open*, 4(2):170–179, January 2015. doi: 10.1242/bio.201410843.

K.-H. Jöckel. Finite sample properties and asymptotic efficiency of Monte Carlo tests. *The Annals of Statistics*, 14(1):336–347, 1986. ISSN 00905364. URL `http://www.jstor.org/stable/2241285`.

L. Knorr-Held and N. G. Best. A shared component model for detecting joint and selective clustering of two diseases. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 164(1):73–85, February 2001. doi: 10.1111/1467-985x.00187.

E. Levina and P. Bickel. The earth mover's distance is the mallows distance: some insights from statistics. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 2, pages 251–256. IEEE, Institute of Electrical and Electronics Engineers (IEEE), 2001. doi: 10.1109/iccv.2001.937632.

E. M. M. Manders, F. J. Verbeek, and J. A. Aten. Measurement of co-localization of objects in dual-colour confocal images. *Journal of Microscopy*, 169(3):375–382, March 1993. doi: 10.1111/j.1365-2818.1993.tb03313.x.

F. J. Massey. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46(253):68–78, March 1951. doi: 10.1080/01621459.1951.10500769. URL `http://dx.doi.org/10.1080/01621459.1951.10500769`.

N. J. B. McFarlane and C. P. Schofield. Segmentation and tracking of piglets in images. *Machine Vision and Applications*, 8(3):187–193, May 1995. doi: 10.1007/bf01215814.

Y. Mimori-Kiyosue, N. Shiina, and S. Tsukita. The dynamic behavior of the APC-binding protein EB1 on the distal ends of microtubules. *Current Biology*, 10(14):865–868, July 2000. doi: 10.1016/s0960-9822(00)00600-x.

W. A. Mitchell and S. L. Lima. Predator-prey shell games: large-scale movement and its implications for decision-making by prey. *Oikos*, 99(2):249–259, November 2002. doi: 10.1034/j.1600-0706.2002.990205.x.

J. Munkres. Algorithms for the assignment and transportation problems. *Journal of the Society for Industrial and Applied Mathematics*, 5(1):32–38, 1957. ISSN 03684245. URL `http://0-www.jstor.org.pugwash.lib.warwick.ac.uk/stable/2098689`.

K. Pearson. Note on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, 58:240–242, 1895. ISSN 03701662. URL `http://0-www.jstor.org.pugwash.lib.warwick.ac.uk/stable/115794`.

S. Peleg, M. Werman, and H. Rom. A unified approach to the change of resolution: space and gray-level. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 11(7):739–742, July 1989. doi: 10.1109/34.192468.

D. W. Piston and G.-J. Kremers. Fluorescent protein FRET: the good, the bad and the ugly. *Trends in Biochemical Sciences*, 32(9):407–414, September 2007. doi: 10.1016/j.tibs.2007.08.003.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016. URL `https://www.R-project.org/`.

Y. Rubner, C. Tomasi, and L. J. Guibas. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, 40(2):99–121, 2000. doi: 10.1023/a:1026543900054.

J. P. Shaffer. Multiple hypothesis testing. *Annual Review of Psychology*, 46(1):561–584, January 1995. doi: 10.1146/annurev.ps.46.020195.003021.

R. J. Simes. An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, 73 (3):751–754, 1986. doi: 10.1093/biomet/73.3.751.

S. Urbanek and Y. Rubner. *emdist: Earth Mover's Distance*, 2012. URL `https://CRAN.R-project.org/package=emdist`. R package version 0.3-1.

V. Zinchuk and O. Zinchuk. *Quantitative Colocalization Analysis of Confocal Fluorescence Microscopy Images*, chapter 4.19, pages 1–16. John Wiley & Sons, Inc., 2008. ISBN 9780471143031. doi: 10.1002/0471143030.cb0419s39. URL `http://dx.doi.org/10.1002/0471143030.cb0419s39`.