Academic Year 2010/2011

UNIVERSITY "FEDERICO II" - NAPLES

PhD In Computational biology and Bioinformatics

XXIV CYCLE

MODELLING THE MOVEMENT OF TRANSFORMED CELL POPULATIONS

Coordinator:
Prof. Sergio Cocozza

Student:
Francesca Fioretti

Tutor:
Prof. Giovanni Paolella

Co-tutor:
Prof.ssa Luisa D'Amore

Index

Abstract

The aim of this work was to analyze migrating cell populations, in search of motion models able to accurately describe the features of normal and transformed fibroblast movement, under different experimental conditions. Diverse *diffusive* models were evaluated, by fitting them to experimental datasets obtained by collecting quantitative data on cell direction and step length under standard culture condition as well as after introduction of a wound in the cell layer. The analysis showed that fibroblast movement has the features of a superdiffusion in all the examined cases, as the squared displacements scale superlinearly with time. The *persistent random walk* model fits all data from fibroblast populations, although a varying degree of persistence was found under different experimental conditions. For NIH3T3 fibroblasts, high persistence was always observed, while NIHRas fibroblasts showed lower persistence in non wound stimulated cultures compared to stimulated ones. In presence of a wound stimulus, directional analysis allowed to associate persistence with a marked ability of fibroblasts to move, all together, towards a specific direction defined by the wound stimulus. Another superdiffusive model, the *Lévy walk,* was not successful in describing NIH3T3 and NIHRas fibroblast populations.

The procedures have been made available within MotoCell, a web application previously developed to quantitatively study cell movement, in an effort to generate an effective tool, that can easily perform all the steps of a cell motility study, ranging from cell tracking, to descriptive statistics and model fit analysis.

1. Introduction

1.1. Cell migration

Cells are dynamic entities that continuously change their shape and position, also in response to stimuli coming from the surrounding environment. The resulting transfer of individual cells or cell groups from one location to another is often referred to as cell migration, a phenomenon easily observed in cultured cells, as well as *in vivo*. Cell migration is involved at various extents in many fundamental processes, such as development, organogenesis, growth and survival, and may also be related to pathological situations such as inflammation or atherosclerosis, as well as tissue invasion by cancer cells and formation of tumour metastasis.

1.1.1. ***Embryogenesis***

Migration is central in the developing embryo, where cells from different areas move, over short or long distance paths, in order to reach their final destination; this movement is associated with morphogenesis and correct positioning of the developing tissues and organs. Different forms of migration are involved in embryogenesis: single cells can individually migrate through the extracellular matrix (ECM) or be passively transported by the blood circle, while continuous epithelial sheets may fold, invaginate or expand. At all stages of development, migration is coordinated and highly regulated, and defective migration can results in severe embryonic malformations often incompatible with life (Kurosaka and Kashina 2008). For example, during gastrulation cells migrate within the blastocyst as large populations to form embryonic layers (endoderm, mesoderm and ectoderm); these cells start to differentiate by evolving into precursors that need to migrate again to their final destination in order to undergo terminal differentiation and generate the different organs. Muscle cell precursors migrate from somites to the emerging limbs, while, in the developing brain, neuronal precursors migrate out of the neural tube and take residence in the layers that will form the brain. Migration of cells from the neural crest is among the best studied embryonic migrations: these cells arise from

the top of the neural tube and migrate to a plethora of locations including bone, cartilage, peripheral nerves, and skin (Locascio and Nieto 2001).

In the embryo, different stimuli are involved in starting the migratory behaviour, including chemotactic or haptotactic molecules, contact guidance, contact inhibition of movement, and population pressure.

1.1.2. *Response to injury*

Human bodies are constantly exposed to environmental stress of different origin, such as thermal changes, chemicals or infectious agents as bacteria and viruses. In many cases, response to these stimuli involves recruitment of specific cell types to the site of injury, to accomplish the inflammatory or immunological response, aimed to combat infection or a toxic agent. In these examples, cells orderly migrate towards the damaged site with different mechanisms: going through the extracellular matrix, being transported by the blood stream or literally rolling along the vascular wall in small vessels before extravasating at the site of inflammation. This rolling movement is mediated by reversible adhesive interactions between the leukocytes primary involved in injury, monocytes and T lymphocytes, and the vascular surface (Crane and Liversidge 2008).

Cell migration is also involved in tissue repair, where the skin or another organ tissue heals itself after injury. Cutaneous wound healing includes a complex series of events, in which epidermal repair is initiated by the progressive extension of a tongue, i.e. a stratified sheet of epidermal cells from peripheral epidermis across the wound. In this process, migration acts in concert with mitosis. Although mitotic activity is not specifically located at or close to the margin of the epidermal sheet, where cellular migration plays a very important role in filling the gap, mitosis plays an important role in replacing cells that have moved away to keep overall epidermal continuity (Krawczyk 1971).

1.1.3. *Cell migration in disease*

In pathological processes, abnormal external signals or altered cell response may result in migration of the wrong cell type and/or to the wrong place, and these events may have effects on tissue homeostasis and, more generally, on health state. Examples include autoimmune syndromes in which immune cells selectively move

to specific locations, such as the joints in rheumatoid arthritis or central nervous system in multiple sclerosis, where they destroy the supporting tissue, causing severe damage (Prat et al. 2002); dissemination of tumour cells leading to metastatization. Tumour invasion is a complex multistep process, in which cells leave the primary tumour to migrate towards distant tissues where they eventually will generate secondary foci. This process happens for cells naturally able to migrate, but may also involve non migrating cell types that acquire this ability after transformation (O'Hayre et al. 2008). The mechanisms of tissue invasion by primary or metastatizing cells are well studied in many cancer types. For example, in metastasis formation by many adenocarcinomas, cell migration is followed by invasion of adjacent tissues and intravasation into blood/lymphatic vessels. In these tumors, the migratory phenotype of invasive carcinoma cells is often associated with increased expression or modifications of genes known to affect cell motility. Among these, the genes coding for proteins of the Ras superfamily, small GTPases involved in cellular signalling pathways responsible for growth, migration, adhesion and more, are often overexpressed or constitutively activated (Bos 1989). Deregulation of Ras oncogene has been related to almost every step of multi-step tumorigenesis and contributes to many different cancer cell features. For example, expression of dominant negative H-Ras (S17N) has been shown to inhibit motility in many carcinoma cells, including serum-stimulated chemotaxis of T24 bladder carcinoma cells (Gildea et al. 2000), hepatocyte growth factor (HGF) stimulated chemotaxis of various ovarian carcinoma cells (Ueoka et al. 2000) and chemotaxis of MCF-7 breast cancer cells toward urokinase-type plasminogen activator, uPA (Nguyen et al. 1999). The effects of H-Ras on cell motility in many cases appear to be mediated by the activation of MAP kinase cascade, although other pathways may be involved; for example in EGF-stimulated MTLn3 carcinoma cells, Ras is required for PI3K activation, PIP3 production and lamellipodial extension, thus suggesting a role for Ras in forming protrusions, a membrane structure required for cell motility (Yip et al. 2007).

1.1.4. *Dynamics of cell movement*

Cell displacement is obtained through a highly coordinated process beginning with the creation of a spatial asymmetry and polarization of the cell body according to an axis oriented along the displacement direction. The overall translocation process may

be described as a cycle, where four steps may be distinguished: extension of a protrusion at the leading edge, establishment of new sites of adhesion to the substratum at the front, cell body contraction, and detachment of adhesions at the cell rear. These steps involve assembly, disassembly or reorganization of actin cytoskeleton, and must be coordinated both in space and time to generate productive movement (Pollard and Borisy 2003). An essential role of the actin cytoskeleton is to stabilize the asymmetric distribution of key components of the directional response apparatus. An early event in polarization, in fact, is a change in the distribution of filamentous F-actin which loses the circular symmetry, around the cell rim, to concentrate in specific regions in preparation of the extension of protrusive structures. These structures at the leading edge of motile cells are highly dynamic and contain dense arrays of actin filaments. These filaments are organized with their barbed ends (fast growing, or plus, ends) preferentially oriented in the direction of the protrusion. The simplest such structures are filopodia, thin cylinders that can extend tens of microns. Lamellipodia, instead, are thin protrusive sheets that dominate the leading edges of cultured fibroblasts and other motile cells. The characteristic rufflings at the fibroblast leading edges is due to lamellipodia that lift up off the substrate and move backwards (Mitchison 1996).

The response to environmental cues in terms of cell migration involves the actin cytoskeleton but also requires a crosstalk between these structures and the focal adhesion machinery. Focal adhesions are dot-like adhesion structures, located underneath the lamellipodium with the aim to sense the extracellular environment and thus affect their behaviour (Riveline 2001). At the focal adhesion level, actin filaments can be linked to the cytoplasmic domains of integrin subunits through numerous anchoring proteins, such as talin, that connect two integrin dimers with actin filaments (Geiger 2009). These contacts are highly dynamic, in fact, during cell migration, both composition and morphology of the focal adhesion change: when a cell proceeds along its chosen path, a focal adhesion becomes progressively closer to the trailing edge and eventually must be dissolved to complete cell body translocation (Huttenlocher 1997).

1.2. Methods and tools for studying cell migration

1.2.1. *Experimental setup*

Cell cultures are often studied as model systems for movement, as the same complex membrane machinery supports cell movement on the culture surface *in vitro,* as well as cell migration *in vivo*. For this reason eukaryotic cell cultures are a good experimental system where cells may move according to random patterns, but may also be induced to coordinated migration by means of specific stimuli, such as a chemical attractant or a mechanical injury.

A variety of methods are used to study cell migration. Many of them evaluate movement by comparing the initial state with the final one, as for example the commonly used transwell assay, in which the movement of cells to the opposite side of a porous membrane is evaluated. The assay uses cell culture devices consisting of an upper and a lower chamber, separated by a microporous, polycarbonate membrane. Cells are placed in the upper chamber and their response to a chemotactic gradient, created by placing a chemoattractant in the lower chamber, is measured by quantifying the number of cells that adhere to the underside of the membrane or that appear in the lower chamber, in suspension or even attached to the bottom. Other methods use underagarose migration assay, where a layer of agarose gel is placed between a cell population and a chemoattractant. As a concentration gradient develops from the diffusion of the chemoattractant into the gel, cells migrate through the agarose and may be visualized over time by using microphotography, as they move upward through the gel along the gradient. Originally developed in 1975 by Nelson et al. (Nelson et al. 1975), this technique has been used for a variety of purposes, including quantifying chemotaxis (Krauss et al. 1994), determining the effects of disease on leukocyte chemotaxis (Flo et al. 1994), and measuring neutrophil responses to multiple chemotactic gradients (Heit et al. 2002; Heit and Kubes 2003).

Wound healing

The *in vitro* scratch assay is an easy, low-cost and well-developed method to measure cell migration *in vitro*. Confluent or subconfluent cell monolayers are scraped to create a “wound”, i.e. an acellular area, which is then monitored to see

how the remaining cells move into it and fill the empty space. This method is based on the observation that, upon creation of an artificial gap on a confluent cell monolayer, the cells at the edge of the newly created gap move toward the opening space to close the "wound" (Liang et al. 2007). Capturing images before and after migration and comparing them, to evaluate the ability of cells to close the gap, is the typical end point wound healing assay (Fenteany et al. 2000). By using this assay, for example, the characteristic collective migration of L1 fibroblasts has been modeled as a sheet migration, in which fibroblasts spread and migrate as a group, in order to close the empty space (Bindschadler and McGrath 2007).

1.2.2. *Video time lapse*

Many of the described methods may take advantage of dynamic microscopy, as the progressive acquisition of subsequent frames during the experiment, that allows to have a detailed record of the phenomenon as a function of time. Random movements of cells plated on the plastic surface of commonly used Petri culture dishes, may be acquired and analyzed to quantitatively evaluate migration. In addition, morphological changes of the membrane and cell shape may be visually evaluated, to study the behaviour of membrane structures involved in cell rearrangements in response to changes in extracellular environment.

Wound healing experiments may also be evaluated in this way, by following the advancement of the edge in the course of the experiment. The collection of recorded images allows to observe the modality of filling the gap and to evaluate the extension of the closed area, together with the time used to complete the process.

Time-lapse microscopy techniques provide a complete picture of complex migration phases that occurs over time. Different types of mammary epithelial cells have been observed to follow a bimodal behaviour during random migration. A bimodal correlated model of migration, as a continuum alternation of directional and re-orientation phase, allowed to describe the behaviour of epithelial cells during time (Potdar et al. 2009 and 2010).

1.2.3. *Motion analysis tools*

Various programs are available for tracking the movement of cells and to analyze the path covered by the cells. Some of these tools are freely available for the analysis of

the movement within large applications, such as ImageJ plugins, Particle Tracker (ParTrac) and MtrackJ (MtrackJ). ImageJ (Abramoff et al. 2004) is a tool for analysis and processing of images widely used, in which were included plugins, such as the Particle Tracker (website: ParTrac). It allows to view and analyze later the selected cells in several ways: by saving trajectories and analyzing them at a later time, by displaying all the trajectories in a non-progressive way or progressively focusing on all of them, displaying the trajectories present in a selected area. These trajectories can also be kept as the file containing the coordinates of the cell positions in order to obtain a graph of the shifts as a function of time. MtrackJ (website: MtrackJ), however, has been developed to facilitate manual tracking of moving objects in two-dimensional or multidimensional images and may be used to evaluate cell displacement and velocity during the period of analysis.

Cell_motility (Martens et al. 2006) is an open source Java application that allows a clear and concise analysis for a large number of data related to cell movement, starting from a data file containing the coordinates of individual cells. This is able to reprocess these data in order to obtain three outputs: a picture of tracks followed by each cell; a graph that represents the relation between mean square displacement and time; and a table summarizing the data and parameters used for calculations.

In order to analyze the movement of cells, in particular experimental conditions, some software tools were developed. For example, the freely available TScratch tool (website: TScratch) was designed to automate the measurement of area progressively occupied by cells in wound healing assay (Geback et al. 2009).

Commercial programs, such as MetaMorph (websites: Metam), however, provide data on the motility and morphology of an entire cell population, such as the length of the trajectories, the directionality index, the index of migration, and in case of wound healing experiments, the closing speed of wound.

Similarly, the DIAS (Dynamic image analysis system, Solltech inc.) software, starting from the dynamics of the cell centroid, computes parameters such as velocity, direction change and persistence; and by considering the dynamics changes in contours, it provides parameters, such as width, length area and perimeter (Wessels et al. 2006 and 2009).

1.3. MotoCell

A freely accessible web application, MotoCell, was previously developed in our laboratory to evaluate migration of cell populations, maintained in various experimental conditions. Methods for cell tracking, parameter evaluation and statistical analysis procedures have been integrated in the software package (Cantarella et al. 2009, website: MotoCell).

1.3.1. ***Processing of experimental results***

MotoCell has been designed to track and evaluate paths followed by cultured cells moving on a surface. Cell tracking results in a set of x-y coordinates that are written to a table. The final destiny of each tracked cell is also recorded: paths may last for the whole observation time, but may also prematurely end because of the death of a moving cell, or a mitotic event, or following the movement of the cell beyond the limits of the observation field. The paths may be shown superimposed upon the images of the movie, and their coordinates may be edited, in order to correct eventual tracking errors. The system may be used to associate cells into subsets, which may be separately evaluated. Cell coordinates for each time steps are stored in text files and further processed through the web interface to evaluate movement statistics.

1.3.2. ***Statistical analysis***

Cell coordinates are used by MotoCell to calculate statistical parameters as speed, direction and linearity, able to describe the movement of single cells along their path or of whole cell populations. Analysis of directional movement may be carried out by using circular statistics, where the single displacements are treated as circularly varying quantities. In this way, directional movement of the cells may be studied by eliminating the *modulo* effect; the calculated linear dispersion coefficient (R), and the corresponding angular dispersion (S), describe the dispersion of cell directions around the average direction. The Rayleigh test may be used to assess the significance of R linear dispersion coefficient.

The results are quickly made available as a number of tables and plots, designed according to specific experimental situations, which may be downloaded as text or pdf files. The software allows to quantitatively evaluate the behaviour of single cells

as well as entire populations or subgroups of cells identified on the basis of shared features. For example, in wound healing assays, the tool allows to group cells on the basis of distance from the wound, thus creating subpopulations that can be analyzed separately or all together (Cantarella et al. 2009).

1.4. Motion models for cell migration

Cell migration may be studied by breaking it down to a discrete number of path segments, where the cell moves in a straight line producing cell displacements joined by turning angles, in which a "decision making" process possibly results in a change of path direction. These decisions are taken on the basis of the current cell condition and/or information, or lack thereof, from the surrounding medium. This information may come in various forms, for example composition of the extracellular matrix or concentration of soluble factors, and affect cell migration.

Movement can be modelled in various ways, the simplest possibly being standard diffusion. More complex models assume a non-uniform distribution of turning angles, as in the case of a correlated random walk, where the walker, once taken a decision, tends to persist in it for a certain time, or in the case of a directional bias, where, when challenged with a decision, some directions are taken more often than others. Another alternative is the so-called *Lévy walk*, where no persistence or directional bias is assumed, but the distance covered in the various time units follows a well defined power law distribution model.

1.4.1. ***Simple random walk***

A *simple random walk* represents a spatial-temporal trajectory, in which the probability of turning angles is uniform at each step and step displacements are short. These features mean that, in a *simple random walk*, the walker randomly selects, at each time point, an angle from a uniform distribution and a displacement from a distribution following a fast decay function, such as Gaussian or exponential.

The simplest model of movement using random walks is uncorrelated and unbiased. Uncorrelated means that, at each step, direction of movement is completely independent of the previous steps, and unbiased means that there is no preferential direction: position at the end of each step is only dependent on the starting position

and the followed direction is completely random. In this context, the mean squared displacement, defined as the squared distance averaged over many random walk paths all starting from the origin, is related with time by the expression

$$MSD \sim t^{\lambda} \quad (1)$$

where λ is equal to 1. Brownian motion, that well describes the diffusion of particles within a fluid, does not perfectly describe the movements of living organisms, which deviate from a typical Brownian motion. Migration of living cells may in fact be better described by the standard diffusion equation:

$$p(x,t) = \frac{1}{\sqrt{4\pi D t}} e^{\left(-\frac{x^2}{4Dt}\right)} \quad (2)$$

where p(x,t) is the probability that the walker be at distance x at time t, x is the walker position and D is the diffusion coefficient. Assuming that movement in any direction is allowed, the *simple random walk* model can be well described by this equation (Codling et al. 2008).

This model is the basis of most of the theory of diffusive process. By using this model it was possible to describe the purely random movement of animals and cells, and, at the same time, by searching for the features of this model in different experimental conditions, the *simple random walk* model could be observed as a component of more complex models of movement. The first evidence for *simple random walk* behaviour in bacterial cell migration was on the soil bacteria *P. putida*, where cells propel themselves through their surrounding media by rotating flagella that form a tuft at one end of their body (Harwood et al. 1989). In the absence of a chemical gradient, a single cell traces a path consisting of a series of straight runs interrupted by changes in direction, initiated by a reversal in rotational direction of the flagellar motors of the bacteria. This swimming pattern resembles a three-dimensional random walk and is shared by other bacterial species.

The *simple random walk* model has been evaluated also in higher organisms, by searching for these features in eukaryotic cell motion. By performing time-lapse imaging of medial ganglionic eminence (MGE)-derived cortical interneurons,

tangentially migrating in the marginal zone (MZ) in flat-mount cortices, Tanaka et al. found that they exhibit a range of diverse types of behaviour in terms of the rate and direction of migration. They distinguish three different interneuron populations, based on their motility behaviour: stationary, directed and wandering cells. By computing the mean squared displacement of these cells during time, the predominant population, the wandering one, could be effectively modeled by a random walk. The authors note that the wandering population exhibits this motion model while waiting for the arrival of signals, such as a guidance cue which activated a directed movement (Tanaka et al. 2009)

In some cases the *simple random walk* model has been found in the study of cell migration in concomitance with loss of directionality associated with extracellular factors, such as in the formation of new blood vessels during angiogenesis, a process mediated by directed migration and adhesion of endothelial cells (ECs). Individual cells follow this model when they lose their sense of direction upon addition of arachidonic acid (AA), an amphiphilic compound incorporated into the cellular membrane, and expected to alter its dynamics by causing changes in its composition. Though instantaneous speed of individual cells is not affected by higher AA concentration, cells lose their ability for directed migration and the mean squared displacement is found to be linearly related to time with a diffusion coefficient λ close to 1, as in a *simple random walk* model (Rossen et al. 2011).

Computational models have been used to simulate cell dispersal and predict the diffusion coefficient λ of the migration process. T cell movement within lymph nodes, both in the absence and presence of antigens, has been studied by using computer simulation of idealized random walk (Beauchemin et al. 2007). It was found that λ may be correctly predicted starting from three parameters: time that a cell spends crawling, time required for cytoskeleton reorganization, and average cell speed. In this way, the proposed random walk model allowed to well describe the behaviour of T cells in two phases: the first, in which cells perform a random walk for displacements over long times; the second, in which they follow straight trajectories over short periods of times (Beauchemin et al. 2007).

As discussed, the mean squared displacement, defined in (1), for a typical diffusive random walk is linearly related to time. However, there are situations where dispersal is not diffusive, and the MSD is not linear in time but, instead, has some other power-law relationship. Such situations are known as anomalous diffusion. The key

is the value of λ parameter in the relation (1): for λ < 1 the situation is subdiffusion, for λ > 1 is superdiffusion, for λ=2 is a ballistic or wavelike diffusion.

1.4.2. *Correlated random walk*

It has been observed that cell migration does not always occur in a completely random manner, but a tendency often exists to maintain the same direction for shorter or longer time intervals. This behaviour means that, although the *simple random walk* model is unable to describe this motion process, a more complex one, which implies a correlation between subsequent steps, might do. This led to formulate a new mathematical model, the *correlated random walk* (Gail and Boone 1970). Several scientists attempted to define this model, with different assumptions, as the common result was to correlate the mean squared displacements, $d^2(t)$, with time by means of two parameters, speed (S) and time persistence (P) (Dunn 1983). This last parameter defines the correlation between subsequent movements as the time in which a cell persists in keeping the same direction. According to this model, the mean-squared displacement, $d^2(t)$, can be obtained from (Alt and Hoffmann 1990):

$$d^2(t)=2S^2P\left[t-P\left(1-e^{-\frac{t}{P}}\right)\right] \quad (3)$$

For $t >> P$, $d^2(t)\sim 2S^2P$, where movement is described by a normal diffusion; whereas for $t<<P$ the equation becomes $d^2(t)\sim S^2t^2$, where motion is essentially unidirectional, without changes in direction (Dickinson and Tranquillo 1993).

Fürth found this random walk model for the first time in his study of protozoa motility. This observation was the basis of later studies, where a search strategy was hypothesized for both *Dictyostelium* and *Polysphondylium* cells, characterized by a zig-zag motion manner, where a left turn is followed by a right turn, while maintaining a directional persistence (Li et al. 2008; Maeda et al. 2008). These cells bias their motion by apparently remembering the last turn and moving in the same direction as the previous one for almost 10 minutes, turning eventually away from this direction. Once oriented by reorganizing their cytoskeletral complex in order to elongate a protrusion, they take advantage of this time persistence to move away

from their position in a more or less straight line (Li et al. 2008). The *persistent random walk* behaviour of *Dictyostelium* has been confirmed by using a stochastic model with a detailed characterization of pseudopodia extension associated to migration (Van Haarster 2010).

Gail and Boone used the *correlated random walk* model to study mammalian cell motility in a study on mouse fibroblast migration. Observations over successive short time intervals revealed a tendency for cells to persist in their motion direction for 2.5 hours (Gail and Boone 1970). Years later, Selmeczi et al. saw that, by evaluating the parameters obtained from the *persistent random walk* model for both fibroblasts and keratinocytes, it is possible to show that these cells maintain memory of the past movement (Selmeczi et al. 2005).

How the biochemical signals are related to physical processes underlying locomotion of fibroblasts has been studied on complex human matrix substratum (Ware et al. 1998). It was observed that EGF induced motility could only be found over a certain range of matrix concentrations, because different amounts of both permissive and inhibitory signals modulate motility in response to the stimulus. In these conditions, cell speed and directional persistence were respectively found to be directly and indirectly related to membrane extension rate induced by EGF treatment during sustained migration (Ware et al. 1998). This effect of EGF on cell migration has also been observed in glioblastoma cell migration in three-dimensional matrices (Kim et al. 2008).

The *persistent random walk* model was also found to be associated to the movement of pre-plasma cell (pre-PCs) during *in vivo* differentiation of mouse lymph nodes (Fooksman et al. 2010). *In vivo* movement of pre-PCs between germinal center compartment and medullary cords in lymph nodes is typically guided by chemokine gradients. Although cell tracks obtained in presence of this stimulus were initially expected to follow a random walk, a more in-depth analysis revealed that the pre-PCs displayed a greater velocity with very long steps between turns. This motion has been well described by the *persistent random walk* model (Fooksman et al. 2010).

1.4.3. ***Biased random walk***

Another well-known specialized random walk model is the *biased random walk* (BRW). This model is used to describe motion processes in which a directional bias

affects the walker in choosing which direction to take. When cell migration is affected by the presence of environmental factors, such as nutrient availability, molecular gradients, electrical fields and so on, this model is possibly able to better describe motion under the experimental conditions, compared to the *simple random walk* model. The assumption is that, if there is a directional bias, the walker moves with higher probability following the preferred direction or towards a given target. In some cases motion is affected by a directional bias and, at the same time, shows directional persistence in time, producing a model called *biased and correlated random walk* (Codling et al. 2008). The main feature of these models is the presence of a directional bias that may be quantified by analyzing the probability distribution of movement directions in a circular plot (Codling and Hill 2005; Codling et al. 2008). When the distribution is uniform, net displacements are distributed with equal probability F(θ) around the unit circle: in this case

$$f(\theta)=\frac{1}{2\pi} \qquad (4)$$

The concentration of directions around a preferred one is well described by the von Mises distribution, which relates probability F(θ) to μ and κ as in

$$f(\theta;\mu,k)=\frac{1}{2\pi I_0(k)}e^{k\cdot\cos(\theta-\mu)} \qquad (5)$$

where I_0 denotes the modified Bessel function of the first kind and zero-order, which is defined by

$$I_0(k)=\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{k\cdot\cos\theta}d\theta \qquad (6)$$

The parameter μ is the mean angle and the parameter k is known as the concentration parameter, where $k\geq 0$. The distribution is unimodal and is symmetrical around $\theta=\mu$ (Bentley 2006).

The uncorrelated random walk model has often been used to model bacterial movement, as an alternation of run and tumble where, if there is a preferred direction, the probability of making long straight runs towards this direction becomes more likely. Keller and Segel (1971) modelled as a random walk the movement of *E. coli* when the concentration of chemical attractants is able to affect the turning rate. Bacterial chemotaxis, in this way, is determined by diffusion, with the addition of a directional bias depending on the concentration of a chemical substance. This particular movement was also studied by means of a random walk foraging algorithm, developed by Nicolau et al. (2009). Chemotaxis is a directional movement driven by a concentration gradient of signal molecules. Similarly, haptotaxis is a form of directional movement, occurring in response to extracellular matrix (ECM) molecules, such as fibronectin or hyaluronic acid (Petrie et al. 2009). The movement of different eukaryotic cells has been studied over the years by using different type of chemotaxis assays. For example, polymorphonuclear leukocyte behaviour, in response to different concentration of chemotactic peptide N-formylnorleucylleucyphenylalanine (FNLLP), has been modeled by taking into account the kinetic fluctuation in chemoattractant-receptor binding (Tranquillo et al. 1988). In this study it was found that, in environments where chemoattractants are uniformly distributed, paths follow a *persistent random walk* model, whereas in presence of chemoattractant gradients they switch to a *biased random walk*, characterized by a directional movement most likely related to an increased number of receptors at the leading edge (Tranquillo et al. 1988). *The biased random walk* model was widely used to describe the movement of chemosentitive cells like bacteria or leukocytes (Hill and Hader 1997; Hartman et al. 1994; Ambravaneswaran et al. 2010), but this directed motion has been observed also in physiological processes, such as differentiation (Wong et al. 2010) and wound healing (Schneider et al. 2010). The neural stem cell mode of migration *in vitro* has been studied, both in absence of directional cues and in presence of applied electric fields. By generating a model for biased migration of neural stem cell, it was observed that pharmacological treatments, such as the LY294002 phosphatidylinositol-3-OH kinase inhibitor, negatively affect directional migration (Arocena et al. 2010). Similarly, Al-Shanti et al. demostrated that both velocity and directionality of murine myoblasts during wound healing are strongly affected by pharmacological inhibitor MAPK and PI3K

pathways, involved in the processes of cell protrusion extension in the direction of cellular migration (Al-Shanti et al. 2011).

Various studies focus on cell migration with directional cues determined by external stimuli such as mithogenic factors and contact guidance, associated with extracellular matrix modifications (Fang et al. 1999; Groh and Louis 2010). The overall conclusion is that, by means of different intracellular pathways activated by the ligand-receptor interaction, these factors regulate cell directionality and induce biased migration .

In mouse fibroblasts, *biased random walk* may be strongly affected by primary ciliary formation. The authors in fact demonstrated that ORPK mouse fibroblasts, a line defective in ciliary assembly, respond to a wound stimulus by moving at higher speed but with randomly distributed directions, unlike the parental MEF line which show a strongly directional behaviour. This happens even in presence of PDGF, a key regulator of directional cell migration in wound healing (Scheneider et al. 2010).

1.4.4. ***Lévy walk***

The *Lévy walk* model is a type of random walk in which the movement of the walker is described by a uniform distribution of step directions and a power-law distribution of step lengths. This means that the step lengths follow the distribution function:

$$P(l) \sim l^{-\mu} \quad (7)$$

where l are the lengths, $P(l)$ their probability and μ is an exponent that, for in *Lévy walk*, is included in the interval $1<\mu<3$, given that for $\mu>3$, owing to the central limit theorem, the distribution becomes close to the *Gaussian,* thus returning to the simple random model (Bartumeus 2007). Breaking down a motion process described by a *Lévy walk* model in time intervals of different width, the power law distribution of step lengths can still be observed, because of the scale invariance properties of this model. The step length distribution results in a superdiffusive process, characterized by the presence of a number of very long steps and the mean squared displacement of the walker scales superlinearly with time, unlike normal diffusion where this relation is linear (Selmeczi et al. 2008).

Lévy walk models have been widely used in different fields in order to characterize dispersal of animals in ecology, but also migration of microrganisms and eukaryotic cells. It is generally assumed that the advantage for walker in selecting step lengths in this way, compared with *simple random walk* model, is that *Lévy walk* represents an optimal solution to the biological problem of searching for food when it is sparsely and randomly distributed outside the walker's sensory detection range.

Matthaus et al. observed that, in the absence of chemical gradients, the length of individual step is commonly distributed exponentially, but in presence of fluctuations in the level of CheR, an enzyme of the chemotaxis-signalling pathway of *E. coli*, the distribution follows a power-law, assuming a *Lévy walk* behaviour (Matthaus et al. 2009).

This motion strategy was also found to describe the spontaneous movement of *Dictyostelium discoideum* (Takagi et al. 2008). Quantitative analysis at different stages showed that velocity distribution closely follows a power-law tail in all conditions, although in later stages increased persistence is observed. A possible explanation of this behaviour is that motion characteristics of the amoeba's cells have a physiological origin, in that in the first phases of development, *Lévy walk* is an optimal strategy for foraging, while in subsequent phases, the need to form the multicellular organism, pushes the cells to make a more targeted walk (Takagi et al. 2008).

This Lévy like behaviour of *Dictyostelium* cells was hypothesized by Van Haastert, in his study of extended pseudopodia direction while cells searching for food. In this case, however, they did not observe *Lévy walk* behaviour for *Dictyostelium* cells movement. They motivated this observation with the suppression of *de novo* pseudopodia formation in favor of splitting the existing ones, so causing an increase in long displacements in starvation conditions (Van Haastert and Bosgraaf 2009).

Recent studies on cell movement associated this model also to mammalian cells. It was observed that the movements of the rat's microglia were well described by a *Lévy walk* model (Grinberg et al. 2011). In this study by using control slice cultures, the authors found that microglial cells move according to the *Lévy walk* model. Furthermore, they showed that the number of cells moving at longer distances, such as described by a *Lévy walk*, is inversely proportional to synaptic activity: it was reduced by increasing neuronal activity with lipopolyssacharide; in contrast, it was increased by blocking synaptic activity via tetrodotoxin (Grinberg et al. 2011).

2. Scope

Main target of this thesis was to analyze migrating cell populations, by using motion models able to describe different specializations of the *simple random walk* model, also in relation to environmental conditions. Starting from cell migration features, four random walk models were selected: the simple normal diffusion, and three superdiffusive models, i.e. *Lévy, persistent* and *biased random walk*s. One section of the work was focused on the development of computational procedures able to estimate the parameters involved in the equations describing the selected models. Another aim of this project was to introduce these procedures in MotoCell, a migration analysis tool previously developed in our laboratory, through a reorganization of the software object architecture. Finally, the developed procedures for modelling cell migration were applied to the study of mouse fibroblast motion. To this aim NIH3T3 and transformed NIHRas fibroblasts were analyzed in experiments in which cells move randomly on the culture surface and in wound healing assays, which allow to study the effects of a directional stimulus on cell movement.

3. Results

3.1. Modelling cell migration

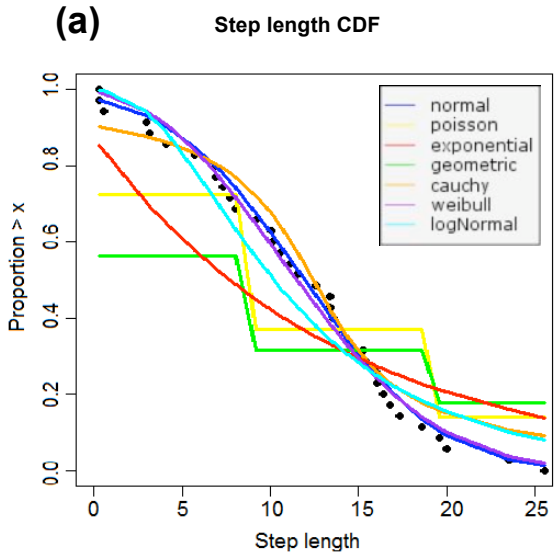
3.1.1. *Method for motion analysis through simple random walk model*

Simple random walk is probably the simplest commonly used motion model and is characterized by the absence of time correlation and directional bias. This means that the direction of movement is completely independent of the previous directions, and there is no preferred direction of migration.

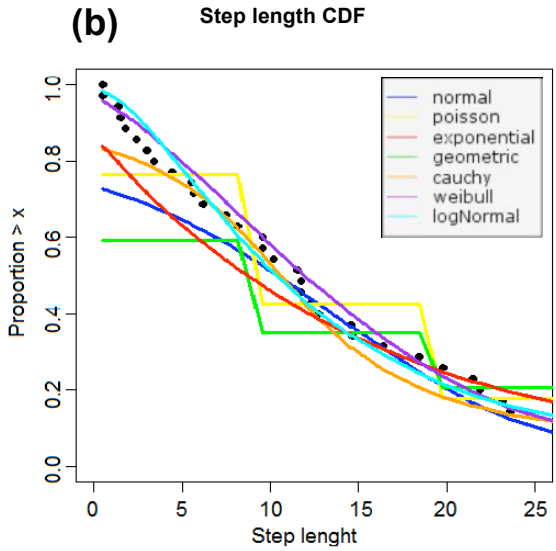
In a discrete space, the *simple random walk* model well describes movements in which the walker moves in all possible directions making steps of variable length, normally distributed around an average value. This model is therefore characterized by a *Gaussian* distribution of step lengths and a *uniform* distribution of directions. This means that the *simple random walk* model results in a normal diffusion process, in which the mean squared displacement is linearly correlated with time. Starting from these considerations, the applicability of the model can be analyzed by focusing on its different features: *Gaussian* distribution of step lengths, *uniform* distribution of directions and normal diffusion during time.

Evaluation of the distribution of step lengths

One feature of the random walk model, as said before, is that the step lengths are expected to follow a *Gaussian* distribution. As in our experiments the step length data assume the form of a discrete variable, a procedure was developed aimed to test this hypothesis, by fitting a *Gaussian* model to the inverse cumulative distribution taken from a dataset. In this way, the problem of bins with zero step length values was largely reduced. In this procedure *Gaussian,* as well as a number of non *Gaussian* well-known distributions, were fitted with a collection of step lengths generated from a simulation of *simple random walk*, as shown in figure 1.1a. The distributions used in this analysis were, in addition to *Gaussian*: *Poisson*, *exponential*, *Weibull*, *log-normal*, *Cauchy* and *geometric*. For each tested distribution, parameter estimation was obtained by using a maximum likelihood method (see ‘Methods’). Goodness of fit of the estimated distributions to the



DistrAnalysis			
Distribution	R^2	Kolmog. test (D)	Cr. Value (0.01)
Normal	0.99	0.07	0.09
Weibull	0.98	0.10	0.09
Log-normal	0.94	0.10	0.09
Cauchy	0.96	0.10	0.09
Exponential	0.73	0.30	0.09
Poisson	0.72	0.30	0.09
Geometric	0.44	0.40	0.09



DistrAnalysis			
Distribution	R^2	Kolmog. test (D)	Cr. Value (0.01)
Normal	0.87	0.13	0.07
Weibull	0.98	0.07	0.07
Log-normal	0.99	0.09	0.07
Cauchy	0.95	0.13	0.07
Exponential	0.90	0.17	0.07
Poisson	0.88	0.23	0.07
Geometric	0.63	0.40	0.07

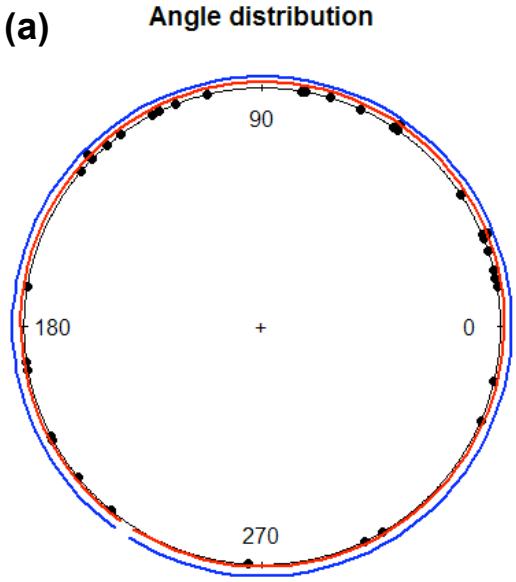
Fig. 1.1. Cumulative step length distribution of two dataset generated from a simulation of *simple random walk* **(a)** and *correlated random walk* **(b)**. Data are plotted together with different estimated cumulative distribution functions: *normal*, *Poisson*, *exponential*, *geometric*, *Cauchy*, *Weibull* and l*og-normal*. Tables on the right side report the R^2 coefficients and the results of the Kolmogovor-Smirnov test used to assess the goodness of fit (99% confidence level).

simulated datasets was qualitatively evaluated by calculating the R^2 determination coefficient and by applying the *Kolmogorov-Smirnov* test. As expected, the *Gaussian* distribution is the one that best fits the simulated data, as also confirmed by the high value of 0.99 for the R^2 coefficient, together with the result of *Kolmogorov-Smirnov* test, although other distributions show very good levels of fit. As a control, another dataset was generated from a *correlated random walk* simulation, a motion model where a *Gaussian* distribution of the steps is not expected (figure 1.1b). In this case, other distributions, such as the *Weibull* and the *log-normal*, fit much better to the simulated step length data, while the *Gaussian* distribution shows a worse fit, actually the worst after the geometric one. The goodness of fit for these statistical distributions is also confirmed by the *Kolmogorov-Sminov* test, positive for $P<0.01$.

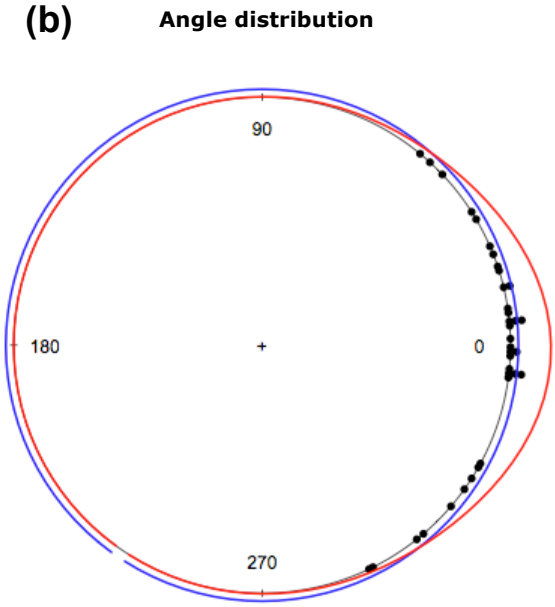
Directional analysis: circular distribution models

As said before, the *simple random walk* model implies that, at each time point, a moving object randomly chooses one of all possible directions. This means that a *circular uniform* distribution must be expected. In order to evaluate this aspect, an assay was set up, where the *circular* distribution was tested, together with *von Mises*, another *circular* distribution which well describes direction angles more or less tightly concentrated around a mean value. The maximum likelihood estimation method was used to estimate the *von Mises* distribution parameters μ and κ. The fit of the predicted models to the original datasets was assessed by using the *Watson* test for both *uniform* and *von Mises* distributions.

The method was tested by using different datasets, generated from a *simple* and a *biased random walk* simulations; the results for one dataset were reported in figure 1.2. The estimated curves for *uniform* distribution, in blue, and for *von Mises* distribution, in red, are reported around a circular plot, together with the simulated data (black points). As seen in panel (a) for the *simple random walk* simulation, the *uniform* distribution is confirmed by the result of *Watson* test, clearly away from the critical value for a confidence interval of 99%. Viceversa, in the case of the *biased random walk* simulation, fig. 1.2b, the *uniform* distribution hypothesis should be rejected (admitted error of 1%), while the *von Mises* distribution should still be accepted, as shown from the results of the *Watson* test. The different fit to the data is clearly visible in the circular plots, where the red curve representing the estimated



	Test vM.	Cr. value vM	Test Circ.	Cr. value Circ.
Dataset 1	0.03	0.09	0.18	0.27



	Test vM.	Cr. value vM	Test Circ.	Cr. value Circ.
Dataset 2	0.11	0.16	1.66	0.27

Fig. 1.2. Distribution of directions of two datasets simulating a *simple random walk* **(a)** and a *biased random walk* **(b)**. The circular *uniform* distribution, in blue, and the estimated curve of *von Mises* distribution, in red, are plotted together with the black points representing directions (•). The results of Watson test for goodness of fit have been reported, for both von Mises (test vM) and uniform (test Circ), in the tables below the charts (confidence level 99%).

von Mises distribution is in much better agreement with the data than the blue one, representing the *uniform* distribution.

Evaluating diffusion behaviour

Normal diffusion is the third important aspect of the *simple random walk* model. To verify whether an empirical dataset resembles a normal or an anomalous diffusion process, the trend of mean squared displacement in time is commonly analyzed. In fact, in normal diffusion this relationship is linear, unlike other models where the displacement may grow faster or more slowly. Basically the expected distance from the starting point may be described according to the formula:

$$MSD \sim t^{\lambda} \quad (8)$$

where MSD is the mean squared displacement, and λ is the random motility exponential coefficient, that for simple diffusion is equal to 1. For other diffusion models, the lambda coefficient may be lower than 1 (subdiffusion) or higher, between 1 and 2 (superdiffusion). Values higher than 2 may only be reached by walkers not moving at constant speed, but accelerating in time.
The estimation of the unknown λ coefficient for a given dataset was performed with the commonly used *Gauss-Newton* algorithm for nonlinear equation solution. This analysis, as seen in the figure 1.3 for different datasets generated from simulations of superdiffusion, normal diffusion and subdiffusion, allows to clearly discriminate the different diffusion types.

By using the three described approaches together, it is therefore possible to study the different features of a random walk process and to easily discriminate movement described well by a *simple random walk* model; these same features can be used as a hint for investigating different motion models other than simple diffusion.

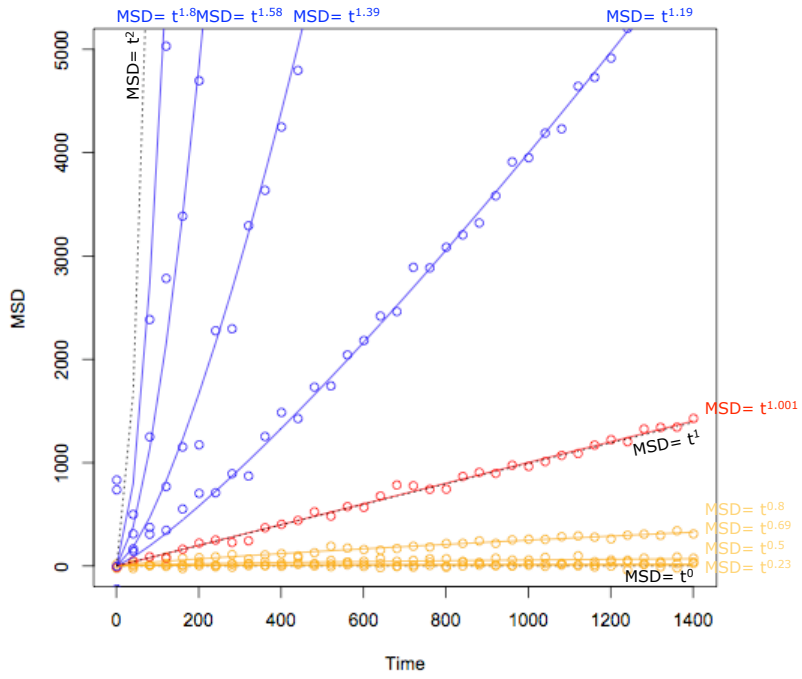


Fig. 1.3. Graphic representation of diffusion types. Mean squared displacements (MSD) are plotted vs time for different datasets respectively simulating subdiffusion (orange), normal diffusion (red) and superdiffusion (blue). Dotted black curves represent the trend of diffusion with fixed λ values of: 0,1,2.

3.1.2. ***Step length distribution model: "Lévy walk"***

The *Lévy walk* model is a specialized random walk, characterized by a typical distribution of the step length frequencies following a power law tail. Such a frequency distribution is described by:

$$P(l) \sim l^{-\mu} \quad (9)$$

where l is the step length and μ the exponential decay index.

In the case of *Lévy walk* model, the necessary condition for the power law, describing the distribution of movement lengths, is that the exponent μ falls within the range $1 < \mu \leq 3$. So, to detect if a movement follows the model of *Lévy walk,* this exponent should be estimated, in order to verify if it is in the Lévy index interval definition. To perform the power law exponent estimation and test the *Lévy walk* model, there are some considerations to do on the power law equation and estimation methods.

Exponent estimation methods

The probability density function (PDF) decays following a power law, as implied by the fact that $P(l)$ is proportional to $l^{-\mu}$, as shown in equation 9. But this function does not precisely define the proportionality term that appears between $P(l)$ and $l^{-\mu}$. This implies that, to use this function for parameter estimation, it is necessary to explicitly define the proportionality constant that links these two variables. Since the proportional hypothesis is concerned with the tail of the PDF, it requires the definition of a value at which the tail starts. Referring to the analyzed data, i.e. step lengths of moving cells, it is possible to assume that all data fall in the tail of the power law distribution. By naming a the beginning of the tail , it was defined as the shortest measured or measurable movement which allows to include in the analysis all movements observed in the dataset. Considering movements that are $\geq a$, the PDF for x is:

$$f(x) = Cx^{-\mu}, x \geq a \quad (10)$$

where C is the proportionality constant given by $C=(\mu-1)a^{\mu-1}$. The *f(x)* now defines a valid probability density function to work with.

Starting from this equation, the estimation of the μ exponent has been attempt by using three parameter estimation methods: the maximum likelihood estimation, the cumulative distribution function and the logarithmic binning with normalization.

To test whether differences in the selected methods result in inaccurate estimation of μ exponent, different sets of random numbers (n=500) were generated from a probability density distribution $P(l)\sim l^{-\mu}$, with μ=2. A sample size of 500 elements with exponent equal to 2, well represents a set of step lengths in a movement describing a theoretically optimal *Lévy walk*, and it is comparable with the minimum size of cell population experimentally produced.

Maximum likelihood estimation (MLE) is one of the preferred approaches for estimating frequency distribution parameters. The MLE method first requires determination of the likelihood function. Given known data x, and selected *a* as the tail start, the log-likelihood function for the power law tail is:

$$\log[L(\mu \mid x)] = n \bullet \log(\mu - 1) + n \bullet (\mu - 1) \bullet \log(a) - \mu \sum_{j=1}^{n} \log(x_j) \qquad (11)$$

where L(μ|x) is the likelihood function of a particular value of the unknown parameter μ given the know data x (n is the sample size). By applying the MLE method, the results for the simulated datasets yielded an average μ of -2.1 (± 0.5), with a mean determination coefficient $R^2 = 0.64$ (table 1.1). This estimation result for one set of data is described in figure 1.4a, in which the simulated step lengths are represented as the logarithmic value of frequency distribution. As shown, the red line of estimated data doesn't match well with the blue line representing the simulated set.

The second method of work is the cumulative distribution function (CDF): in this case

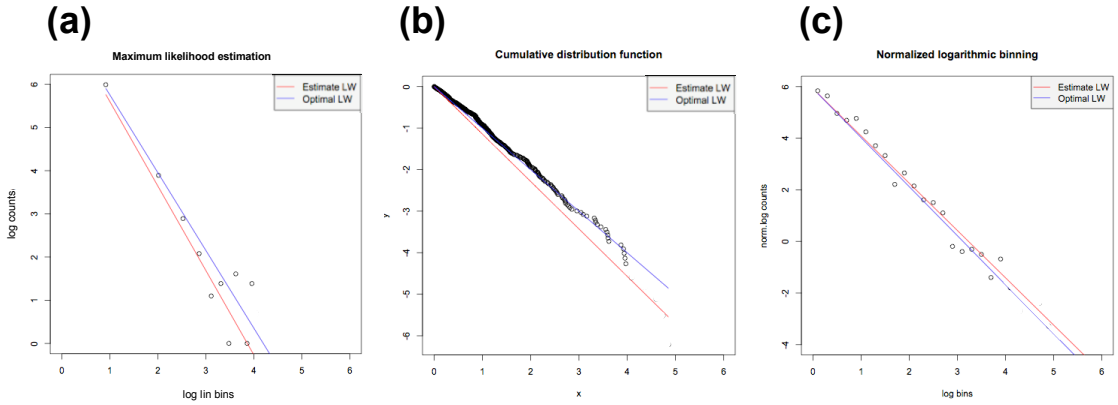


Fig. 1.4. Methods for estimating power law exponent (μ) for a *Lévy walk* simulated dataset 1 with μ=2. Logarithm of probability density function **(a)**, cumulative distribution function **(b)** and observations frequency, reported for each bin as divided by the width of logarithmic bin **(c)** were plotted. The line interpolating the simulated data is blue, while the estimated lines obtained with the maximum likelihood (MLE, in a), the cumulative distribution function (CDF, in b) and the logarithmic binning with normalization (LBN, in c) methods are red.

Estimating method	Estimated μ	R^2
MLE	-2.1 ± 0.50	0.64 ± 0.260
CDF	-2.1 ± 0.30	0.93 ± 0.130
LBN	-1.9 ± 0.06	0.98 ± 0.006

Table 1.1. Estimation of μ exponent for 10 sample datasets tested for *Lévy walk* model. Three estimating methods are used: CDF = cumulative distribution function; MLE = maximum likelihood estimation; LBN = normalized logarithmic binning. Average μ coefficient and R^2 obtained are reported together whit standard deviation values.

$$F(x_j) = (\frac{x_j}{b})^{(\mu+1)} \quad (12)$$

where *b* is the maximum value of the data. The CDF is progressively constructed for a set of observed data, by ranking the n observed x_j values from smallest to largest. The probability that an observation is less than or equal to x_j is then estimated as j/n. Having determined the CDF for a power-law distribution, the exponent μ of the probability density function can be estimated by using regression. The traditional approach is to transform the equation for the CDF in order to render the slope directly linked with μ:

$$\log(F(x_j)) = -(\mu + 1)\log(b) + (\mu + 1)\log(x_j) \quad (13)$$

In this way the slope of the regression is equal to μ+1, making necessary to subtract 1 to obtain μ. In figure 1.4b the results for one simulated dataset are showed: for each value of x_j, the logarithm of the rank is plotted against $\log(x_j)$. In this plot it is possible to see that the estimated data, in red, again seem to have a margin of error respect to the empirical data, in blue. Making the CDF estimation for all the simulated datasets with the known exponent μ=2, the average value was -2.1 (± 0.3), with an R^2 value of 0.93 (table 1.1).

The third way to estimate the exponent of a power-law distribution is logarithmic binning with normalization method (LBN). This method (Fig. 1.4c) involves setting bin breaks such that each bin is twice the width of the preceding bin. This approach is similar to the linear binning method, but uses bins of logarithmically rather than linearly increasing width. This causes a reduction in the number of empty and low-count bins at large values of x_j, because the linear width of a bin increases linearly with x. The count of each bin is then divided by the width of that bin to have a frequency density for each bin. The frequency density for each bin is then plotted, as logarithmic values, against the geometric midpoint of that bin, as seen in figure 1.4c. A linear regression of the plotted points, in red, gives an estimate of μ, as the negative of the slope, for one of the simulated dataset. As shown, in this case the red line of the estimated data seems to be a very good estimation of the simulated data, draw in blue. The μ values obtained by applying the LBN method showed an average

of 1.9, with a very low standard deviation (0.06). Furthermore, the goodness of the estimation was confirmed by a high determination coefficient of 0.98 (table 1.1).

Comparative result evaluation

The average value of the exponents estimated with the three techniques, for the simulated datasets, was reported in table 1.1, together with their standard deviation and the mean determination coefficient R^2, used as index of goodness of fit. As seen in the table, although all the three estimation methods seem to estimate quite well the power law exponent of 2, they don't have the same precision. In fact, the MLE and the CDF methods show a high standard deviation for the exponent values (0.5 for MLE, and 0.3 for CDF) and quite low R^2 index of 0.64 (MLE) and 0.93 (CDF). This poor parameter estimation, for the two methods above, was explained by the selected dataset characteristics. In fact, even if the sample size of 500 elements seems to be larger, at large step length values their frequency is still very low. So, these two methods estimate the μ index with lower precision, because they were strongly affected by the low number of large values in the tail.
On the other hand, the LBN method seems to solve this problem, in fact it estimates very well the μ exponent with a low standard deviation of 0.06 and a high R^2 index of 0.98. In fact, by setting a different weight for the long steps bins, this method smoothes the effect of dataset size, performing, in presented experimental condition, the better exponent estimation than the other two methods, MLE and CDF.

Selected method: LBN

The LBN provides an accurate method to estimate the parameter of a power law distribution, for a set of data that simulate a real step lengths distribution, directly from the slope of the linear regression; for this reason, this method was used to perform Lévy index estimation. The procedure to evaluate the best fit for the *Lévy walk* model can be resumed in these following steps: compute the step lengths frequency distribution from the coordinates of moving cells; set the starting limit, *a*, of the power law tail; estimate the exponent of the distribution function by using the logarithmic binning with normalization method; assess the estimation with the R^2 coefficient; verify if the estimated exponent falls in the range $1 < \mu \leq 3$.

3.1.3. ***Correlated random motion: "Persistent random walk model"***

Correlated random walk, unlike *simple random walk*, is a motion model that implies a correlation between orientated successive steps. This feature produces a local directional bias. In fact, the assumption is that each step tends to point towards the same direction as the previous one, although the influence of the initial direction of motion progressively diminishes over time and step directions are uniformly distributed in the long time scale.

The *persistent random walk* is a form of *correlated random walk*, in which the correlation between successive steps is determined by the persistence parameter. This parameter is the time that, on average, a cell spends in moving in the same direction as the previous step. In the *persistent random walk model*, mean squared displacements (MSD or d^2) are related to time by means of persistence and speed parameters:

$$d^2(t)=2S^2P\left[t-P\left(1-e^{-\frac{t}{P}}\right)\right] \quad (14)$$

where d^2 is the mean squared displacement (μm^2), S the speed ($\mu m/min$), P the persistence (min) and t the observation time (min). The mean squared displacement can be determined from experimental data, for a time interval $t_i = n_i \Delta t$ (where n_i is the number of steps at t_i), tracked for the total observation time and calculated by:

$$\langle d^2(t_i)\rangle = \frac{1}{N-n+1}\sum_{i=0}^{N-n}\left[(x_{(n+i)\Delta t}-x_{i\Delta t})(x_{(n+i)\Delta t}-x_{i\Delta t})\right] \quad (15)$$

where $\bar{x}_i$ is the position vector at time i. The d^2 values are calculated for all time intervals during the total motion observation period. In *persistent random walk* model, given the observation time intervals, the mean square displacement of the analyzed dataset is only dependent on parameters S (speed) and P (persistence).

Nonlinear estimation problem

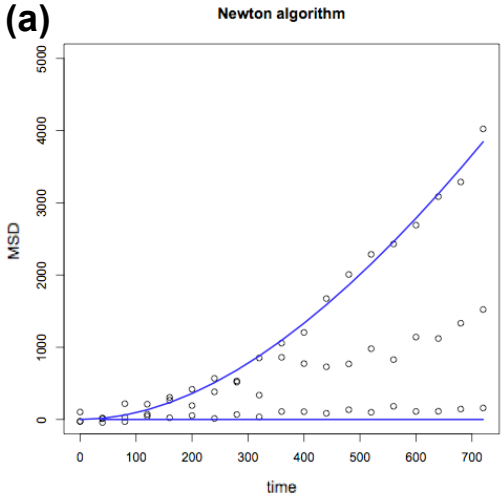
As seen above, the expression describing the *persistent random walk* model is a nonlinear function; therefore, in order to estimate the unknown parameters, P and S, a fit through a nonlinear regression method is required. As in other cases involving nonlinear regression, persistence and speed values are computed by a procedure requiring an iterative approach, directed to minimize the sum of the squared residuals, as indicated by:

$$RSS=\sum_{i=1}^{N}\left(d_r^2-d_p^2\right)^2 \quad (16)$$

where the subscripts *r* and *p* respectively indicate real data values and values estimated from the model. The basic idea of iterative method involves these steps: 1) calculate the sum of squares for the curve defined by the initial set of parameters, 2) adjust the parameters to make the curve closer to the experimental data, 3) stop the calculation when the sum of squared residuals (RSS) reaches its minimum, 4) report the best fit result. The only difficult step in this process is adjusting the equation parameters to decrease the distance between empirical and estimated data. There are several algorithms used to solve this problem, whose effectiveness depends on the data used to calculate the curve from the model function.

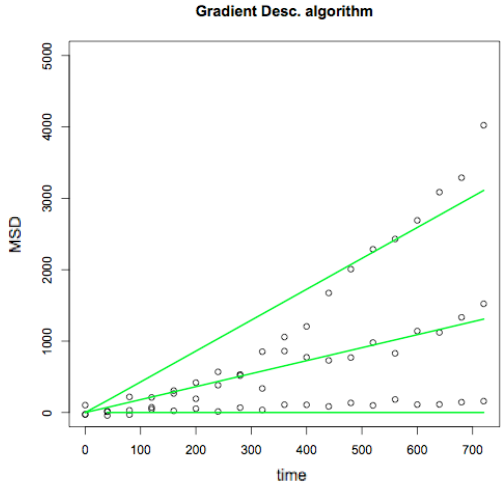
In this study, to obtain the best parameter estimation, different optimization algorithms have been tested. Three different datasets were generated, according to different motion models, which simulate an average speed of 0.1 μm/min and different persistence values, for an observation time of 24 hours, in which the step interval was fixed at 40 minutes. *Dataset 1* was built by using a random walk with no persistence (P=10 min), the *dataset 2* with a moderate persistence (P=120 min), and *dataset 3* with a strong persistence (P=720 min). These sets of parameters include most of the experimental situations used in this study, where moving cell populations were observed during 24 hours, and tracked every 40 minutes. In figure 1.5 the mean squared displacements computed for these datasets were plotted as a function of time, together with the results obtained for the PRW estimation by using three

(a)



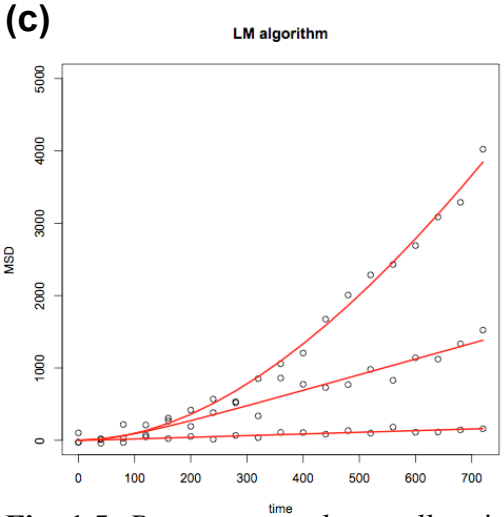
	P (min)	S (µm/min)	R^2
Dataset 1	0	0	NaN
Dataset 2	-	-	-
Dataset 3	277.86	0.12	0.98

(b)



	P (min)	S (µm/min)	R^2
Dataset 1	0	0	NaN
Dataset 2	5.83	0.40	0.97
Dataset 3	1.00	1.45	0.88

(c)



	P (min)	S (µm/min)	R^2
Dataset 1	7.68	0.11	0.73
Dataset 2	108.17	0.1	0.99
Dataset 3	797.84	0.09	0.99

Fig. 1.5. *Persistent random walk* estimation methods. Three methods were selected to estimate parameters of *persistent random walk:* Newton algorithm (**a**), gradient descendent algorithm (**b**), and Levenberg-Marquardt algorithm (**c**). Performances were compared by applying them on three samples having the same speed of 0.1 µm/min and different persistence, respectively of 10 min, 120 min and 720 min [added noise N(0,100)]. Mean square displacements were plotted as a function of time together with the estimated curves (respectively in blue, green and red).

different estimation algorithms: “Newton” (a), “Gradient descendent” (b) and Levenberg-Marquardt (c). As shown in panel a, the first approach of “Newton” algorithm appears to reach a good fit only for *dataset 3* with maximum persistence value, whereas it completely fails the estimation for *dataset 2*, and gets a wrong fit for the *dataset 1*. Also by using different simulation datasets, the “Newton” method failed to make the estimation. This happens because in the iterative search for convergence, this method executes the calculation of the inverse of a matrix. If this matrix is singular, as for some of used simulations, the iteration cannot go on and the algorithm fails. This problem, together with the high computational complexity of the calculation for each iteration, makes this method poorly suited for the presented study.

Similarly, the second approach of “Gradient descendent” algorithm is able to make the estimation but, for all three datasets, this is not well approximated to the simulated data (figure 1.5b). In fact, as reported in the table on the right, it estimates speed and persistence parameters that are far from those used for all simulated datasets. The “Gradient descendent” approach was tested by using different sets of PRW model simulations with known parameters, P and S. The estimation results were very discordant, because the estimated parameter values were far from the real ones. This method, again, did not provide a good solution to the *persistent random walk model*, starting from the data presented in this study.

Finally, the “Levenberg-Marquardt” algorithm (LM) was tried to solve the *persistent random walk* model equation. As shown in figure 1.5c, unlike the first two approaches, the LM method is able to reach a better fit. By applying the LM algorithm on the three datasets generated from the PRW model, the method iteratively computes the minimum of sum of squared residuals and gets better values for the estimating parameters speed and persistence, reported in the table on the right. This method is a blend of “Gradient descendent” and “Gauss-Newton” algorithms. The “Levenberg-Marquardt” method uses the “Gradient descent” in the early iterations, performing long steps to faster reach the convergence area of RSS, then decreases the step size by gradually switching to a modified “Newton” approach, the “Gauss-Newton”, that works far better in estimating the approximation of the local minimum value. In this way the LM algorithm solves the estimation problem found by using the other two methods, by taking advantage in term of computational quickness. The LM algorithm works very well in estimating the model

parameters and representing the best fit to the empirical data for all cases of persistent motion (P=10 min., P=120 min. and P=720 min.). In fact, it yielded estimated values of speed and persistence that are very close to the real ones used in the simulations, still in the case of very low persistent motion (dataset 1). It is possible to note that the red line representing the estimated values closely matches the empirical data and allows to identify, through the slope of the curve, the different persistence behaviour.

Being the LM an iterative approach, it is possible to improve the estimation performance by making some considerations on the experimental dataset. By performing a preliminary analysis on the experimental data, it is possible to do some assumptions about the parameters, speed and persistence, to be estimated. These two measures are obviously defined as finite positive values and, therefore, the persistence value is defined within the interval from zero to the total observation time, and speed is determined by starting from a short interval around the average experimental value. In the present study, during a 24 hour observation time, the persistence ranges between $0 > P > 1400$ minutes. However, this is still a large interval, that implies an ideal movement hypothesis, where cells move along a straight line during all the observation time. For this reason, a further narrowing of the parameter intervals may be done, according to the experimental situation and by taking advantage from the information available in literature on the behaviour of similar cell types. Taking all this into consideration, the initial persistence interval was adjusted depending on the performed analysis. The expected interval for the speed parameter was selected in a range around the expected mean values. By providing the iterative LM algorithm with these hints, it was possible obtain a quicker and more accurate estimate of the unknown parameters.

3.1.4. ***Biased random walk model***

A directional bias is introduced when the probability of moving in a specific direction is greater than that of going in other directions. Paths showing a consistent bias towards a preferred direction or towards a given target are defined as *biased random walks*, or biased and correlated random walks if persistence is also observed. In order to assess the relevance of the different aspects of this model, correlation between the direction of subsequent steps and directional bias may be separately

analyzed. The *persistent random walk* model is used to study directional correlation in time, as reported in the previous section, while directional analysis may be used to assess the directional bias.

Directional analysis

As described in previous section, analysis of directionality was carried on by fitting the *circular uniform* distribution, that describes movements in all possible directions, and the *von Mises* distribution, that describes movements concentrated around a mean direction, to experimental data. In presence of a directional bias, the distribution of net displacements is not well described by the *circular uniform* distribution but a good fit may still be observed with a *von Mises* distribution. As shown in figure 1.2b, when a high concentration of net displacements is present, as in the case of a dataset generated from a *biased random walk* simulation, the data are described well by the *von Mises* distribution (red curve), but the fit with the circular uniform is lost, as also confirmed by the Watson test. This combination allows to highlight a motion pattern characterized by a preferred direction. On the other hand, as shown in figure 1.2a, by testing a dataset generated from a *simple random walk* simulation, the high dispersion of net displacement angles is well described by both the uniform distribution and the *von Mises* one. The effectiveness of this method in distinguishing biased and non-biased migration was confirmed by testing it on different datasets, from simulated experiments or from the observation of cells migrating under random conditions and in presence of a directional stimulus,
Taken together, the *persistent random walk* model, analyzed by means of the Levenberg-Marquardt algorithm for nonlinear regression, and the directional analysis, performed by using circular distributions, allowed the analysis of the *biased and correlated random walk* model.

3.2. MotoCell

In the course of this study, different computational procedures were developed, as shown before, in order to model the process of cell migration: these include methods to evaluate fitting of *simple random walk*, *Lévy walk* and *persistent random walk* models to empirical data. In order to quickly apply these methods to the study of

experimental data, MotoCell, a tool previously developed in our laboratory (see Introduction), was expanded to add new procedures. To this aim, different architectural modifications were introduced, directed to the use of new statistical tests, necessary to analyze and model movement in cell populations.

3.2.1. ***Software reorganization***

Figure 2.1 shows the new MotoCell object architecture, where newly inserted objects are identified with the "new" label, whereas modified old objects are marked with an asterisk ("*"). Five object groups are present: objects used to create the web interface, in purple; objects used to define the physical entities involved in cell migration, in blue; objects responsible for text and graphical output, in green; analysis objects, in red or orange (R based analyzers, see below). A *motocalc* object was also created as the link between the interface generator objects and the new analyzer objects.

Analysis objects are all derived from the *Analyzer* object, developed to perform generic data analysis. This was then extended by creating several children objects, built to execute specific analysis procedures. These include all statistical analyses, already available in MotoCell and discussed in "Introduction", used to determine descriptive parameters for linear and circular statistics, calculated for the whole population as well as for each individual cell and/or step of migration. As many computational procedures were based on R, a complex environment for statistical computing (see Methods), an *Ranalyzer* object was also created, able to directly execute R scripts in MotoCell. This new structure allowed to easily generate new groups of statistical tests, that include, among others, migration analysis through *Lévy* and *persistent random walk* models.

As shown in figure 2.2, the idea was to design an independent object for each analysis group. The objects have the form of a plugin, i.e. they may be added without introducing any changes to the existing code, by just dropping an extra file in a plugin directory. During the construction of the main interface page, the software queries the Analyzer plug-ins to know the implemented tests and their class; this information is used to construct the interface and to present the various analysis tools in a consistently organized way. If selected, the plug-in is then called to perform the calculations necessary to the requested analysis and later on to generate the output.

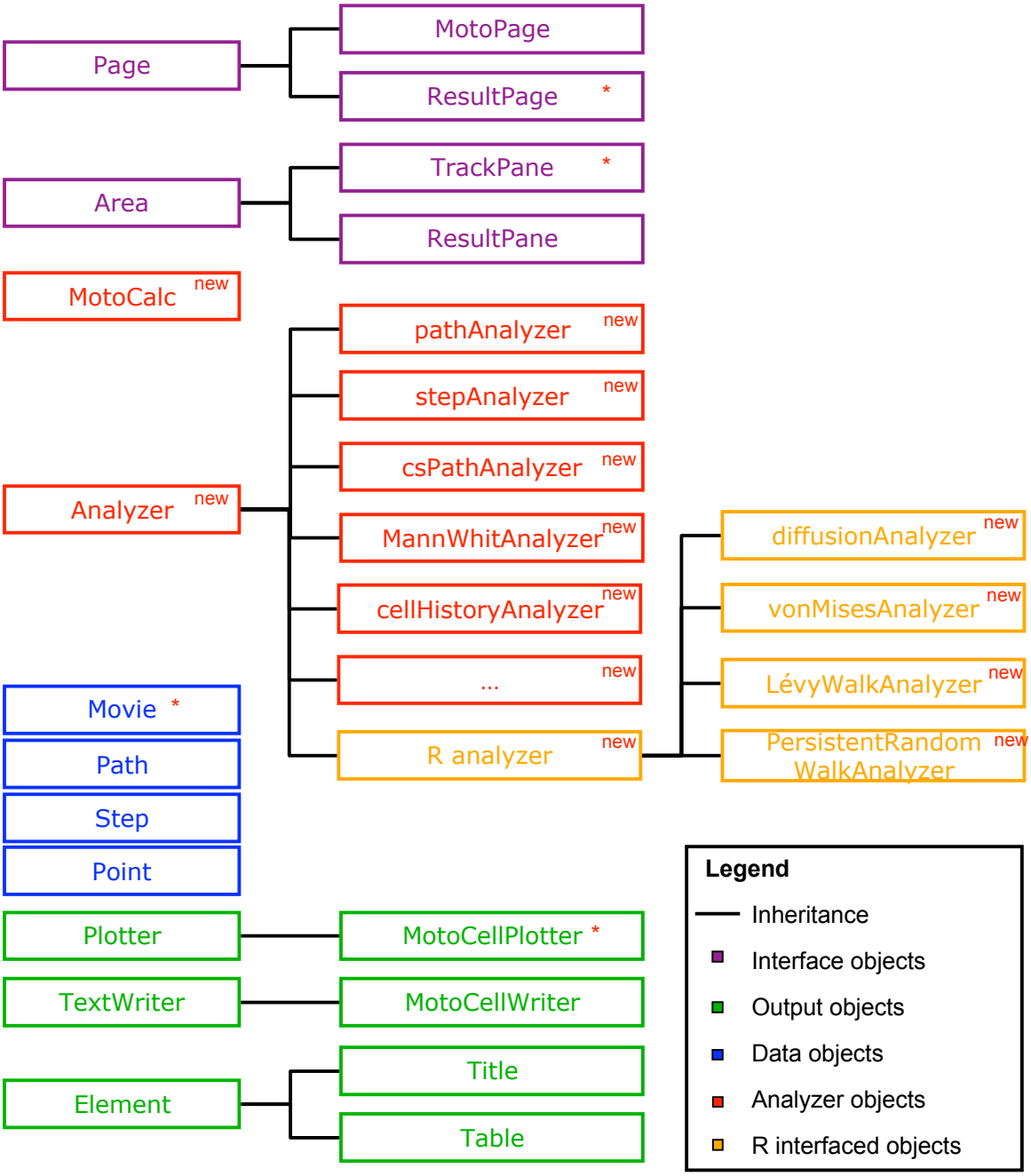


Fig. 2.1. New object architecture of MotoCell: the objects are grouped in order to easily recognize those responsible for building the interface (purple), the output objects (green) and the data objects (blue). The new analyzer objects (red) and R interface objects (orange) are marked as "new" while the "*" identifies the objects modified within this work.

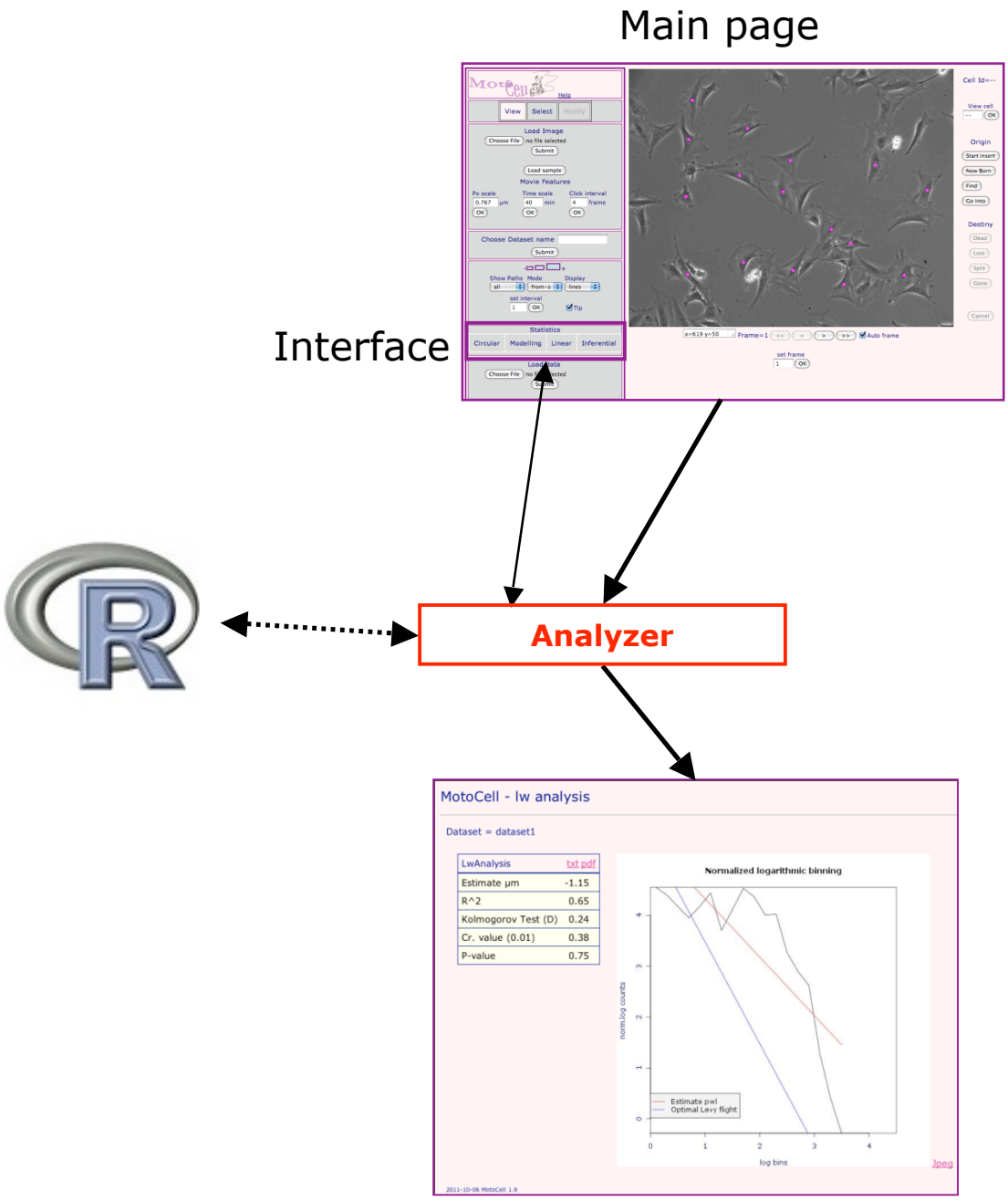


Fig. 2.2. Functioning of analysis plug-in. MotoCell queries an Analyzer object to build the main page interface; the Analyzer performs all computing operations also by using R scripts and generates the output results.

A new class includes all the objects involved in migration analysis by means of motion models, such as *simple random walk*, *persistent* and *biased random walk*, *Lévy walk*.

3.2.2. ***New tracking procedure***

When migration analysis is performed for long time periods, cell observation may be affected by a number of different events: cells may split following a mitosis event, die by apoptosis or other ways, or simply go out of the observation area. Of course cells may also increase, as splitting produces newborn cells and new cells may enter the field of view. These events affect the statistical analysis by introducing two types of bias. First of all, cell split and death cause a progressive reduction in cell number during the observation time, meaning that, although a significant quantitative evaluation may be done in the first part of the analysis, at the end of the observation time the same procedures are often no longer representative of the behaviour of the whole cell population. Furthermore, paths followed by cells that go against these events have a reduced duration compared to the whole duration of the analysis. This means that, by performing a quantitative evaluation of motion during time, the behaviour of cells that move for long time periods will have a higher weight in the analysis compared to cells followed for a short time. This is especially strange, as during the observation time cells tend instead to grow, rather than reduce in number.
In order to counter this effect, a new tracking method was introduced in MotoCell to track cells appearing later in the course of the observation time. The cell origin is recorded, for example as “newborn” in the case of cells derived from a mitosis, or “find” for cell that suddenly appear within the field of view. By using this new tracking procedure, daughter cells can be directly tracked, by also keeping a link to their parent cell, soon after the birth, or later on. The continuous introduction of new cells was shown to improve the analysis of migration, by including a larger number of paths thus reducing or solving the problems deriving from the small number of cell in the final stages.
The tracking of all cells that appear during the analysis provided us with an easy method to perform a measure of the state of the cell culture during time. In figure 2.3 the output of this analysis is shown, where the whole cell population is represented by lines, one for the lifetime of each cell: the symbols at the beginning and end of the

Origins	txt pdf
Earlier	18 cells
Find	0 cells
New born	66 cells
Go into	0 cells

Destinies	txt pdf
Full path	36 cells
Dead	2 cells
Lost	2 cells
Split	34 cells
Gone	10 cells

Proliferation	txt pdf
Average life span	12.0 hour
Included cell	19.0

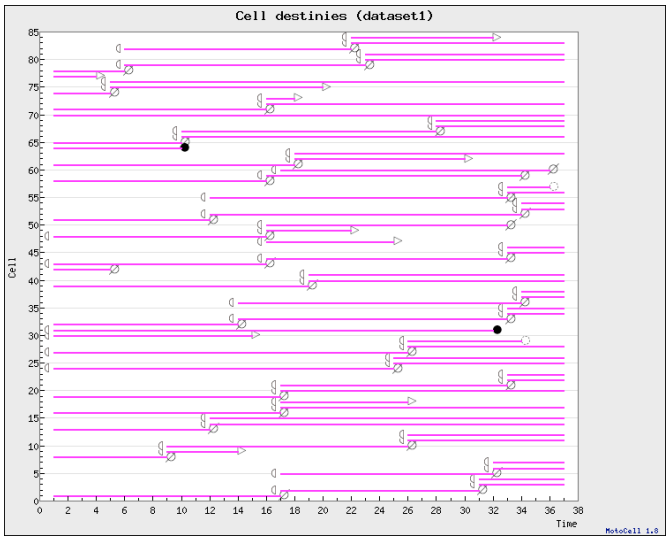


Fig. 2.3. Population viability analysis. The plot shows the result of cell life span analysis. Each segment represents the life span of a single cell, and ends with a symbol representing its final destiny. The table reports statistics on cell origin, destiny and average life span during the observation time.

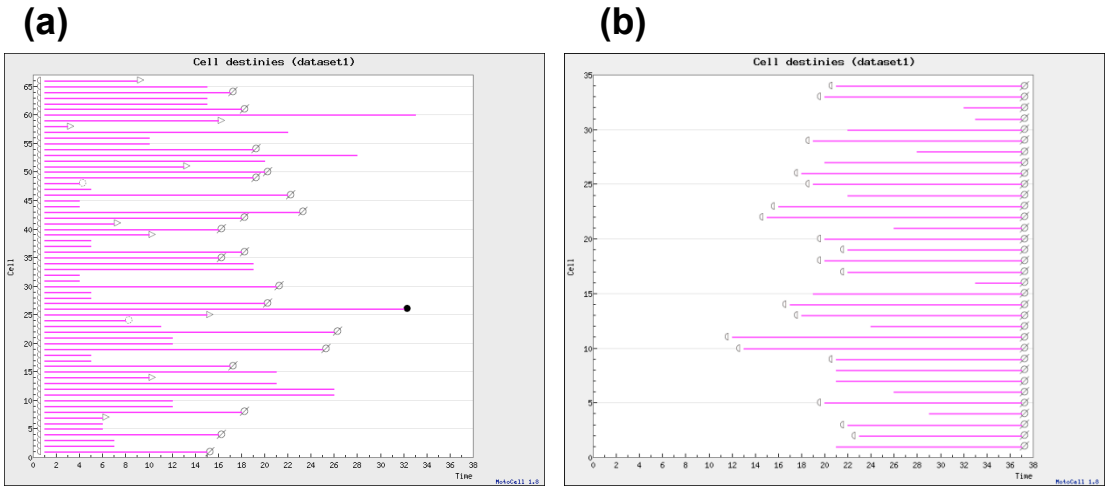


Fig. 2.4. Virtual cell synchronization. The plots show two different modes of synchronizing cells: **(a)** synchronized on the cell birth event, **(b)** synchronized on the cell division event.

line respectively represent its origin and final destiny. A viability analysis of the cell population is also included, obtained by combining the information on the lifespan of each cell: average life span is calculated as the mean time between cell birth and the next mitosis event, whereas proliferation rate, i.e. the time needed for the population's duplication is calculated as follows:

$$PR=\frac{((e+n+gi+fi)-(d+l+go+fp)+n)}{fp} \quad (17)$$

where PR is the proliferation rate, *e* the cell number at start, *n* the newborn cells, *gi* the cells that appear into the field of view, *fi* the cells found during the analysis, *d* the dead cells, *l* the lost cells, *go* the cells that go out the field of view, *fp* the cells that cover all time of the analysis.

Virtual synchronization of cell population allows to study the different migration behaviour of cells in relation to their phase of life, as reported in figure 2.4, where cell paths are aligned according to the birth (a) or split (b) events.

3.3. Fibroblast migration models

3.3.1. ***Searching for simple random walk features***

Quantitative analysis of cell motility was carried out by observing cells under the microscope and producing time lapse series by acquiring images at regular intervals. The paths followed by each cell have been sampled at 10 minute intervals (steps) for 24 hours, although quantitative analysis and statistical evaluation have usually been done at 40 minute intervals in order to avoid the noise introduced by an exceedingly large number of very small steps of just 1 or two pixels. As said, MotoCell software was used according to the described procedures. NIH3T3 and Ras transformed fibroblasts were studied in order to evaluate the effect of the oncogene on their migration features both when moving randomly and when responding to a directional stimulus in wound healing experiments.

The assumption that cells move according to a *simple random walk*, was evaluated by the previously described tests, checking for features such as *Gaussian* distribution of step lengths, *uniform* distribution of cell directions and normal diffusion during time. These tests were carried on for both NIH3T3 and NIHRas populations and

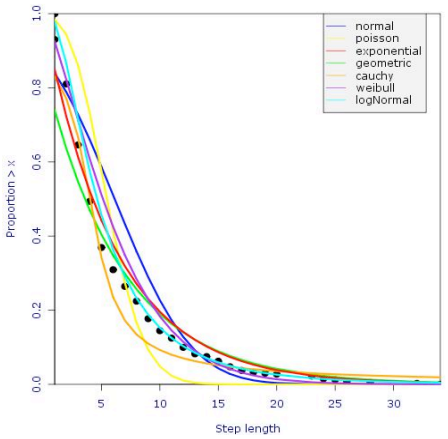
under random and stimulated motion conditions. In the following sections the aspects of step length distribution and diffusion will be analyzed, while the distribution of cell directions will be discussed later on.

Step length distribution

The distribution of the step lengths was tested by evaluating, for each cell population, the inverse cumulative distribution of the step lengths, i.e. the distance travelled in a 40 minute interval. Cumulative distribution of the *Gaussian* that best fit the experimental data was estimated. The results have been reported in figure 3.1a-d. For normal and transformed fibroblasts, under both experimental conditions, step length distribution, appears to be well described by a *Gaussian* distribution, as shown by the blue curves. This observation was also confirmed by the values of the R^2 coefficient, reported in the corresponding tables, close to 1 in almost all cases. By quantitatively testing the goodness of fit by means of the Kolmogorov-Smirnov test, it was observed that the *Gaussian* distribution hypothesis of step length might be accepted for almost all fibroblast population, although in some cases with a very low margin, as for example for the NIH3T3 fibroblast population under stimulated condition (fig 3.1c).

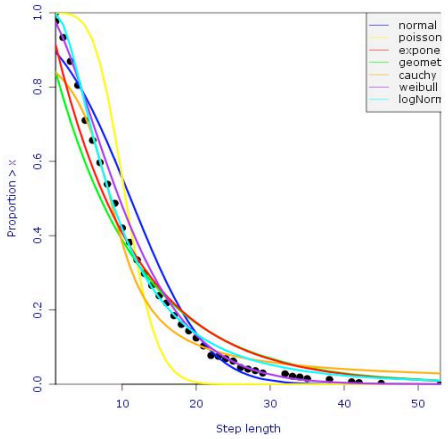
The data have also been tested against other statistical distributions, represented in figure 3.1. Six statistical distributions have been selected to evaluate their fit to the experimental data: the *Poisson*, the *exponential*, the *geometric*, the *Cauchy*, the *Weibull* and the *log-normal*. For each statistical distribution, the estimated curves are reported in figure 3.1, together with the cumulative distribution of the step lengths for randomly moving (a-b) and wound stimulated (c-d) NIH3T3 and NIHRas fibroblasts. A shown, the step length distribution is often well described by these other distributions: specifically distributions of the exponential family, such as the *log-normal* (cyan curve) or the *Weibull* (purple curve) often fit to the data as well as the Gaussian one. The goodness of fit was quantitatively assessed, again, by the high values of R^2 coefficient, and by the Kolmogorov-Smirnov test, as reported in the tables. The test was repeated for several experimental data sets, and the computed R^2 coefficients for the different distributions have been reported in table 3.1. Also in these cases, a good fit of *Gaussian* and other exponential distributions to the experimental data was often found, indicating that these results cannot be used to

(a) NIH3T3



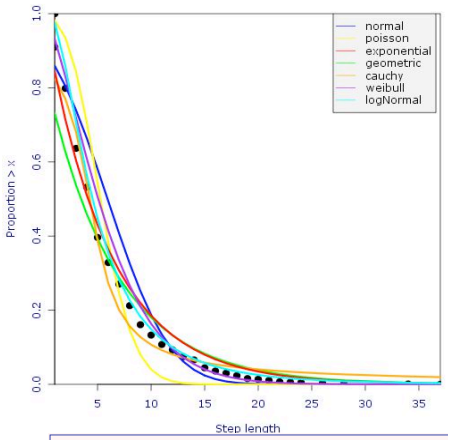
DistrAnalysis			
Distribution	R^2	Kolmogorov Test (D)	Cr. value (0.01)
Normal	0.90	0.22	0.20
Poisson	0.90	0.24	0.20
Exponential	0.97	0.21	0.20
Geometric	0.93	0.29	0.20
Cauchy	0.97	0.17	0.20
Weibull	0.97	0.14	0.20
Log-Normal	0.99	0.10	0.20

(b) NIHRas



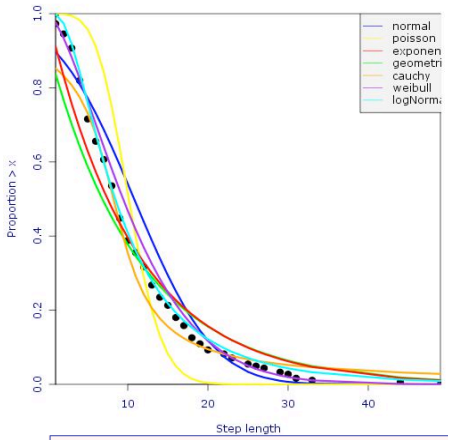
DistrAnalysis			
Distribution	R^2	Kolmogorov Test (D)	Cr. value (0.01)
Normal	0.96	0.14	0.17
Poisson	0.85	0.27	0.17
Exponential	0.97	0.17	0.17
Geometric	0.95	0.22	0.17
Cauchy	0.97	0.16	0.17
Weibull	0.99	0.06	0.17
Log-Normal	1.00	0.07	0.17

(c)



DistrAnalysis			
Distribution	R^2	Kolmogorov Test (D)	Cr. value (0.01)
Normal	0.93	0.18	0.19
Poisson	0.93	0.21	0.19
Exponential	0.97	0.20	0.19
Geometric	0.93	0.28	0.19
Cauchy	0.97	0.17	0.19
Weibull	0.98	0.11	0.19
Log-Normal	0.99	0.09	0.19

(d)



DistrAnalysis			
Distribution	R^2	Kolmogorov Test (D)	Cr. value (0.01)
Normal	0.95	0.15	0.18
Poisson	0.87	0.26	0.18
Exponential	0.96	0.22	0.18
Geometric	0.93	0.26	0.18
Cauchy	0.97	0.15	0.18
Weibull	0.99	0.08	0.18
Log-Normal	1.00	0.08	0.18

Fig. 3.1. Analysis of step length distribution for NIH3T3 and NIHRas fibroblasts randomly moving or stimulated by a wound. Different statistical distributions were fitted to four experimental datasets. Results of fitting are reported as cumulative distributions together with the data from each population: **(a)** NIH3T3 and **(b)** NIHRas, both randomly moving, **(c)** NIH3T3 and **(d)** NIHRas, wound stimulated. Goodness of fit, evaluated by using R^2 coefficient and Kolmogorov-Smirnov test, are reported in the tables.

NIH3T3							
random							
Distribution	**Normal**	**Weibull**	**Log-Normal**	**Cauchy**	**Exponential**	**Poisson**	**Geometric**
Dataset 1	0.90	0.97	0.99	0.97	0.97	0.90	0.93
Dataset 2	0.95	0.98	0.99	0.98	0.95	0.97	0.89
Dataset 3	0.97	0.99	0.99	0.98	0.96	0.97	0.91
Dataset 4	0.91	0.98	0.99	0.98	0.97	0.89	0.93
wound							
Distribution	**Normal**	**Weibull**	**Log-Normal**	**Cauchy**	**Exponential**	**Poisson**	**Geometric**
Dataset 5	0.95	0.98	0.99	0.97	0.96	0.92	0.91
Dataset 6	0.94	0.98	0.99	0.97	0.96	0.95	0.88
Dataset 7	0.95	0.98	0.99	0.98	0.95	0.96	0.88
Dataset 8	0.97	0.99	0.99	0.87	0.93	0.97	0.87
Dataset 9	0.95	0.97	0.99	0.98	0.93	0.98	0.84
Dataset 10	0.94	0.97	0.99	0.98	0.93	0.98	0.86
Dataset 11	0.93	0.98	0.99	0.97	0.97	0.93	0.93
NIHRas							
random							
Distribution	**Normal**	**Weibull**	**Log-Normal**	**Cauchy**	**Exponential**	**Poisson**	**Geometric**
Dataset 12	0.96	0.99	1.00	0.97	0.97	0.85	0.95
Dataset 13	0.94	0.99	1.00	0.98	0.97	0.90	0.94
wound							
Distribution	**Normal**	**Weibull**	**Log-Normal**	**Cauchy**	**Exponential**	**Poisson**	**Geometric**
Dataset 14	0.95	0.99	1.00	0.98	0.97	0.90	0.94
Dataset 15	0.98	0.99	0.98	0.96	0.94	0.86	0.90
Dataset 16	0.89	0.97	0.99	0.95	0.98	0.84	0.94
Dataset 17	0.97	0.99	0.99	0.98	0.95	0.89	0.91
Dataset 18	0.96	0.99	1.00	0.98	0.94	0.94	0.89
Dataset 19	0.95	0.99	1.00	0.97	0.96	0.87	0.93

Table 3.1. Summary of R^2 values obtained by fitting different distributions to several datasets. NIH3T3 and NIHRas fibroblasts were assayed both as randomly moving and under wound stimulus.

conclusively define the *Gaussian* as the best fitting model for NIH3T3, but also for NIHRas, in any of the two migration conditions. This ambiguous result led to investigate other features of the *simple random walk* model.

Diffusion analysis

A *simple random walk*, as described before, typically results in a normal diffusion process, in which the mean squared displacements are linearly correlated with time. As the relation between mean squared displacements and time may be used to distinguish normal diffusion from other diffusion types, diffusive behaviour was studied in NIH3T3 and NIHRas populations under unstimulated and stimulated motion conditions. In figure 3.2 the mean squared displacements were plotted as a function of time for a 12 hours observation time. For the two populations in both unstimulated (a-b) and stimulated (c-d) conditions, the resulting curves do not follow the trend expected for normal diffusion. In fact, the calculated MSD is not linearly correlated with the time, but the points representing the experimental data fall into the expected interval of a superdiffusive process, characterized by a λ exponent included in the interval between 1 and 2 (black dotted lines in figure). Estimation of the λ exponent for different data sets (see table 3.2) confirmed that the observed migration has the features of a superdiffusive process, as shown by the lambda values falling within the interval between 1-2. In general, a different degree of superdiffusion was observed between the two populations: superdiffusion is more pronounced for populations of NIHRas transformed fibroblasts (b-d) than for the parental NIH3T3 fibroblasts (a-c), in both experimental conditions.
This superdiffusive behaviour is typical of fast moving cells, i.e. cells increasing their distance from the starting point faster than if moving according to a normal diffusion model (bottom black line in fig. 3.2). This result led to investigate other random walk models, able to better describe motion types with these features.

3.3.2. ***Superdiffusive random walk models***

The *simple random walk* is a probabilistic model that implies strong simplifications, and, as such, it is not necessarily fit to describe real cell movement. Superdiffusive random walk models are more complex and involve additional parameters, such as

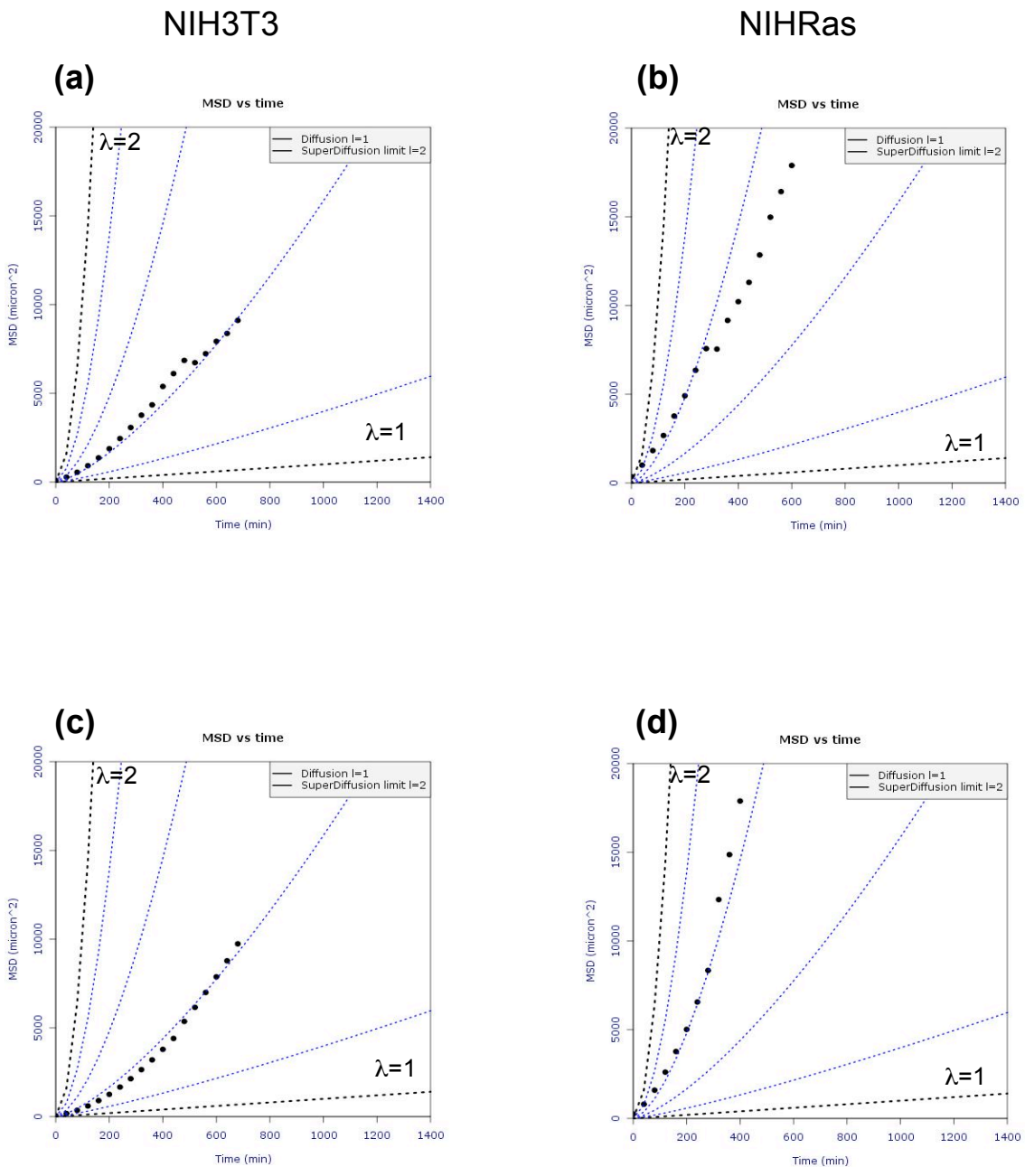


Fig. 3.2. Diffusion analysis of mouse fibroblasts. Mean squared displacements obtained in absence of directional stimulus **(a** and **b)** and in presence of it **(c** and **d)**. Results for NIH3T3 and NIHRas cell populations are reported on the right and on the left side respectively.

NIH3T3	
random	
Dataset 1	1.41
Dataset 2	1.23
Dataset 3	1.39
Dataset 4	1.40
wound	
Dataset 5	1.46
Dataset 6	1.39
Dataset 7	1.48
Dataset 8	1.50
Dataset 9	1.41
Dataset 10	1.40

NIHRas	
random	
Dataset 12	1.51
Dataset 13	1.42
wound	
Dataset 14	1.55
Dataset 15	1.64
Dataset 16	1.39
Dataset 17	1.50
Dataset 18	1.51
Dataset 19	1.63

Table 3.2. Summary of results obtained by estimating λ diffusion exponent for several datasets. NIH3T3 and NIHRas fibroblasts were assayed both as randomly moving (random) and under wound stimulus (wound).

speed, persistence and others, introduced to better describe the migration phenomena. As observed (fig. 3.1) when analysing step length distribution, in many cases, the Gaussian function is not always able to properly describe the step length distributions observed for moving NIH3T3 and NIHRas populations, while other long tail distributions are sometimes as good as if not better than the *Gaussian* one. By taking into account this information two models were selected among the superdiffusive ones: the *Lévy walk* and the *persistent random walk*.

Lévy walk

Lévy walk is a well known model of superdiffusion, characterized by step lengths following a power law distribution function as in:

$$P(l) \sim l^{-\mu},\ 1 < \mu < 3 \qquad (18)$$

where l is the step length and μ is the Lévy index. Estimates for the Lévy μ index were obtained from the step lengths of a cell population by using the procedure developed, described under 2.1.2, which uses normalized logarithmic binning to describe the step length distribution as a line whose slope corresponds to the Lévy μ index.

The results obtained for NIH3T3 fibroblasts are shown in figure 3.3. The tested data points, although falling into the *Lévy walk* interval delimited by the black dotted lines, are poorly described by the model, as shown by the nonlinear shape of the distribution function and by the R^2 coefficients, not very high in both experimental conditions (a-b). However, there is a difference between the two experimental conditions and in the case of the population moving after the wound stimulus, the calculated function appears to better follow the experimental data, with an estimated μ exponent of 1.89, i.e. within the interval expected of the Lévy model.

NIHRas fibroblast populations, randomly plated and stimulated by a directional stimulus, were also analyzed and results are shown in figure 3.4. The experimental data do not result in a straight line in both experimental conditions (a-b) and are therefore again not described by the *Lévy walk* model, with lower values of R^2 coefficient. The test was repeated for several data sets for each cell line and for each experimental condition: R^2 and estimated μ index, reported in table 3.3, confirm the

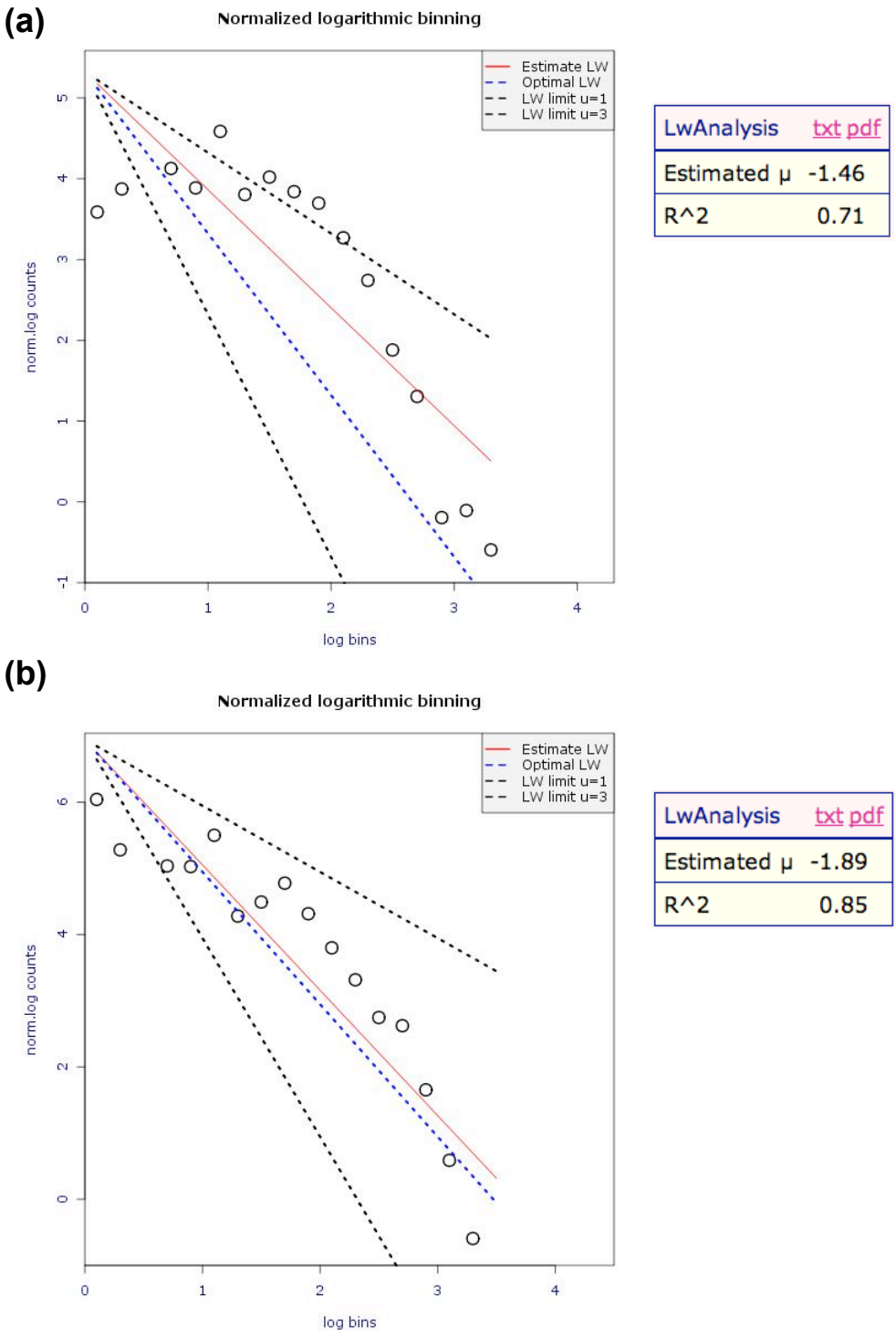


Fig. 3.3. Migration analysis of NIH3T3 fibroblasts by using the *Lévy walk* model. Normalized logarithmic bins of the step lengths obtained in absence of directional stimulus **(a)** and under wound effect **(b)**, were reported on the charts together with the line of the best fit of the *Lévy walk* to the experimental data (red line). The two dotted black lines delimit the μ exponent range of *Lévy walk* (μ=1-3) and, between these, the blue dotted line indicates the trend when μ is fixed to 2. In the tables on the right, the estimated μ exponents, calculated for each experimental dataset, were reported together with the coefficient R^2.

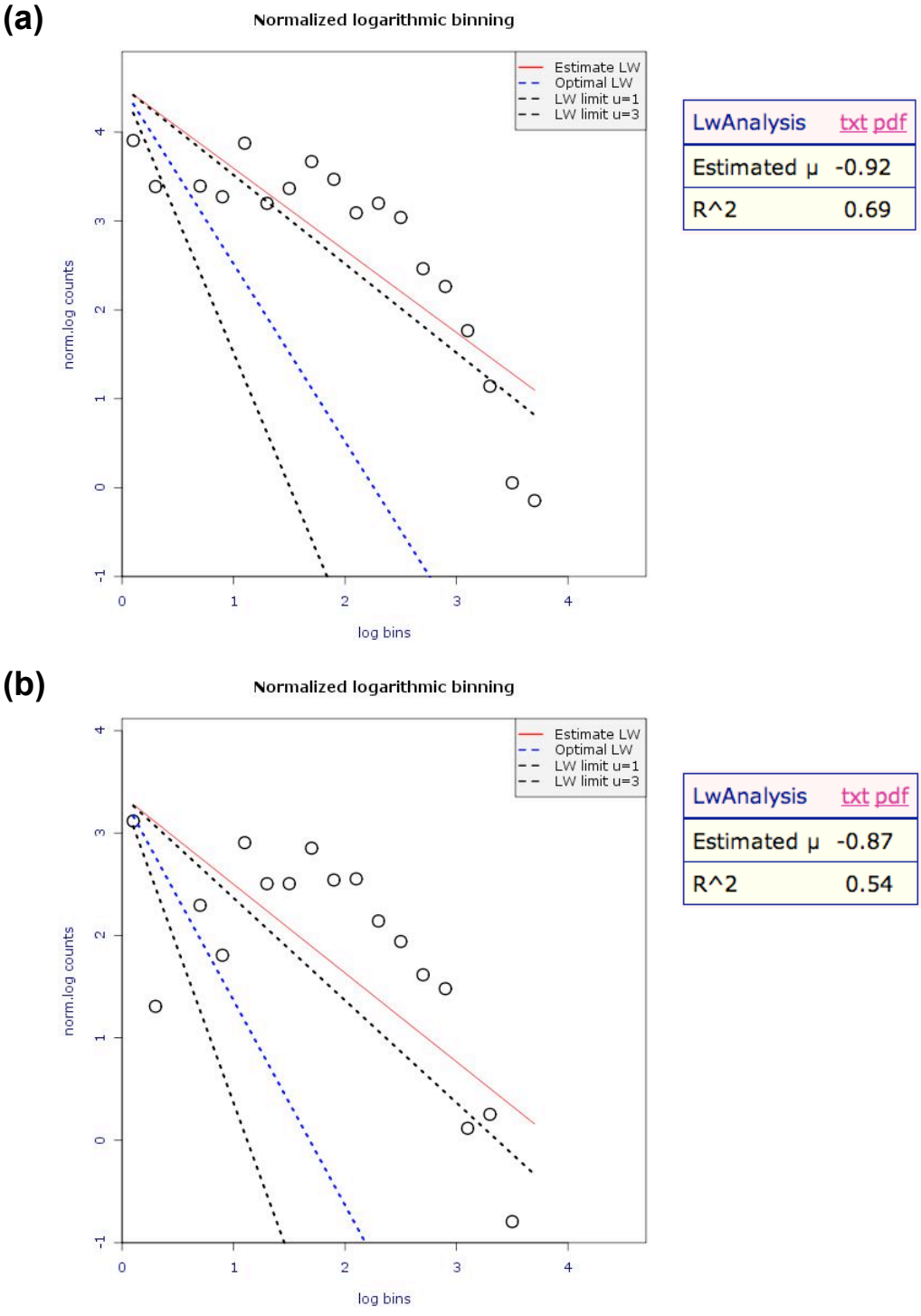


Fig. 3.4. Migration analysis of NIHRas fibroblasts by using the *Lévy walk* model. Normalized logarithmic bins of the step lengths obtained in absence of directional stimulus **(a)** and under wound effect **(b)**, were reported on the charts together with the line representing the best fit of *Lévy walk* to the data (red line). The two dotted black lines delimit the μ exponent range of *Lévy walk* (μ=1-3) and, between these, the blue dotted line indicates the trend when μ is fixed to 2. In the tables on the right, the estimated μ exponents, calculated for each experimental dataset, were reported together with the coefficient R^2.

NIH3T3		
	Estimated μ	**R^2**
random		
Dataset 1	1.53	0.9
Dataset 2	1.46	0.71
Dataset 3	1.89	0.79
Dataset 4	1.22	0.75
wound		
Dataset 5	1.66	0.75
Dataset 6	1.09	0.8
Dataset 7	1.89	0.85
Dataset 8	1.11	0.64
Dataset 9	0.63	0.57
Dataset 10	0.83	0.55
NIHRas		
	Estimated μ	**R^2**
random		
Dataset 12	0.92	0.69
Dataset 13	1.24	0.77
wound		
Dataset 14	0.19	0.19
Dataset 15	0.63	0.35
Dataset 16	0.64	0.35
Dataset 17	0.96	0.66
Dataset 18	0.87	0.54
Dataset 19	0.78	0.56
Dataset 20	1.19	0.63

Table 3.3. Summary of results obtained by estimating μ parameters for several datasets. NIH3T3 and NIHRas fibroblasts were assayed both as randomly moving and under wound stimulus.

results shown before for both NIH3T3 and NIHRas fibroblasts, with the *Lévy walk* model poorly describing the migration, with the only possible exception of stimulated NIH3T3 fibroblasts.

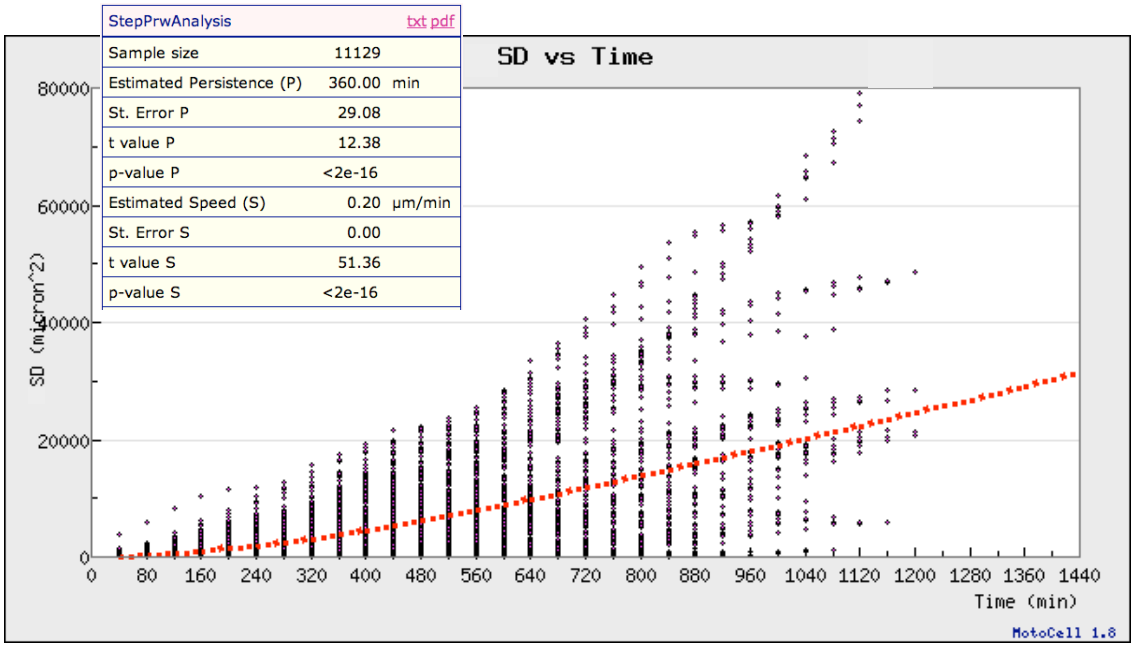
Persistent random walk

In order to analyze fibroblast motion by using the *persistent random walk* model, speed and persistence parameters were estimated by using the developed procedure based on the Levenberg-Marquardt algorithm for nonlinear regression (see Results 2.1.3).

The results obtained for NIH3T3 populations, in random motion and stimulated conditions, were shown in figure 3.5, where the plotted points represent the squared displacements of each cell at time intervals ranging from 40 min to 24 hours. The red dotted lines represent the mean squared displacements estimated by using the LM algorithm for the *persistent random walk* model estimation. The model appears to fit NIH3T3 migration data in both motion conditions (fig. 3.5a-b), with the reliability of the estimated parameters confirmed by using the Student's t-test. A similar curvature, corresponding to a similar degree of persistence (200-300 minutes) was observed for unstimulated (figure 3.5a) and stimulated conditions (fig. 3.5b). The speed estimates also appear similar between the two populations and are close to the values obtained by averaging the step lengths for the two fibroblast populations.

Motion analysis through the *persistent random walk* model for NIHRas fibroblasts, in both random and stimulated culture conditions is reported in figure 3.6. As for the parental NIH3T3fibroblasts, the migration is well described by the *persistent random walk* model. However in this case, a consistent increase of the squared displacements after directional stimulus can be observed in wounded compared to unstimulated fibroblasts, associated with a different curvature of the estimated curves (red), indicative of a different degree of persistence, depending on motion conditions. Persistence for the NIHRas fibroblasts raises from 100 min under unstimulated conditions to over 600 min after wound, corresponding to a change from 2 to about 16 steps. Regarding to the estimated speed, increased values were always found for NIHRas populations compared to parental ones, but they are apparently not dependent on the directional stimulus. Also in this case the estimated speed is comparable to that calculated by averaging the step lengths.

(a)



(b)

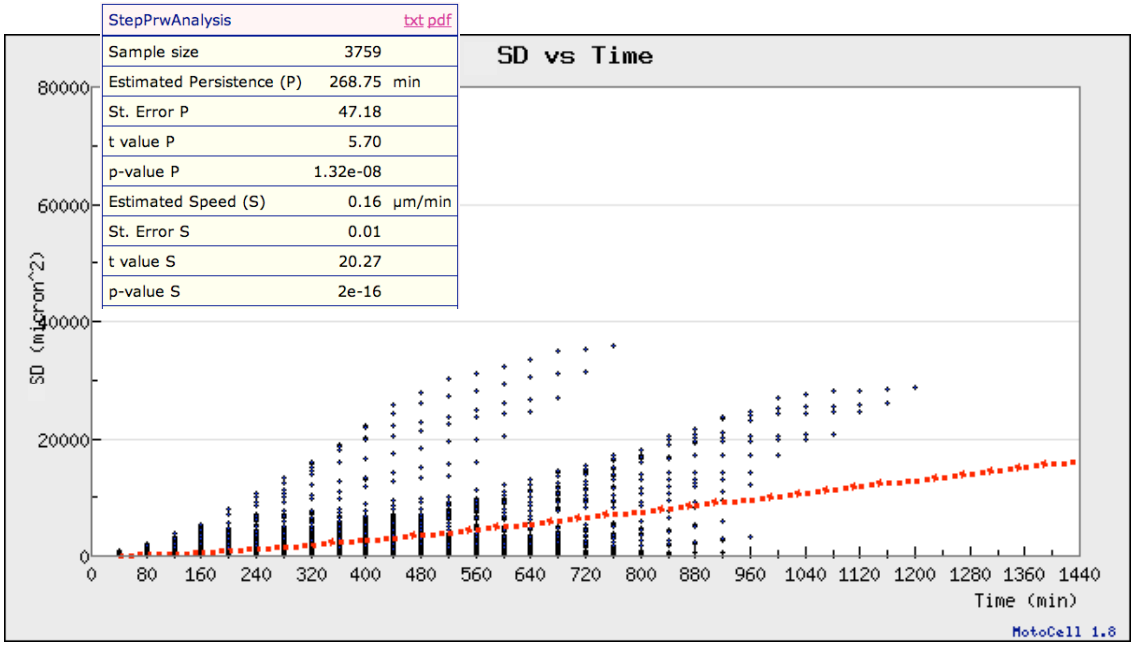
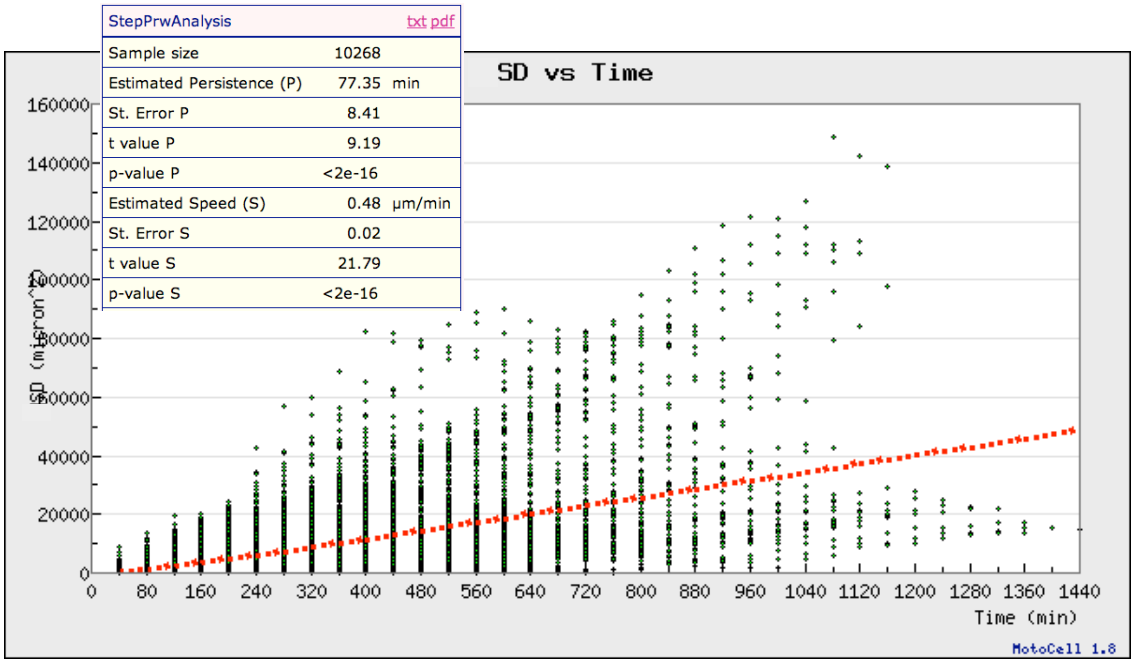


Fig. 3.5. Migration analysis of NIH3T3 fibroblasts by using the *persistent random walk* model. Squared displacements of NIH3T3 cells randomly moving **(a)** and under directional stimulus **(b)** have been plotted as a function of the time. The better estimation obtained by using the *persistent random walk* model has been superimposed to the points as a red dotted curve. In order to evaluate the goodness of fit, the R^2 coefficients have been reported in the tables together with estimated persistence and speed, and with the corresponding Student's t-test for significance.

(a)



(b)

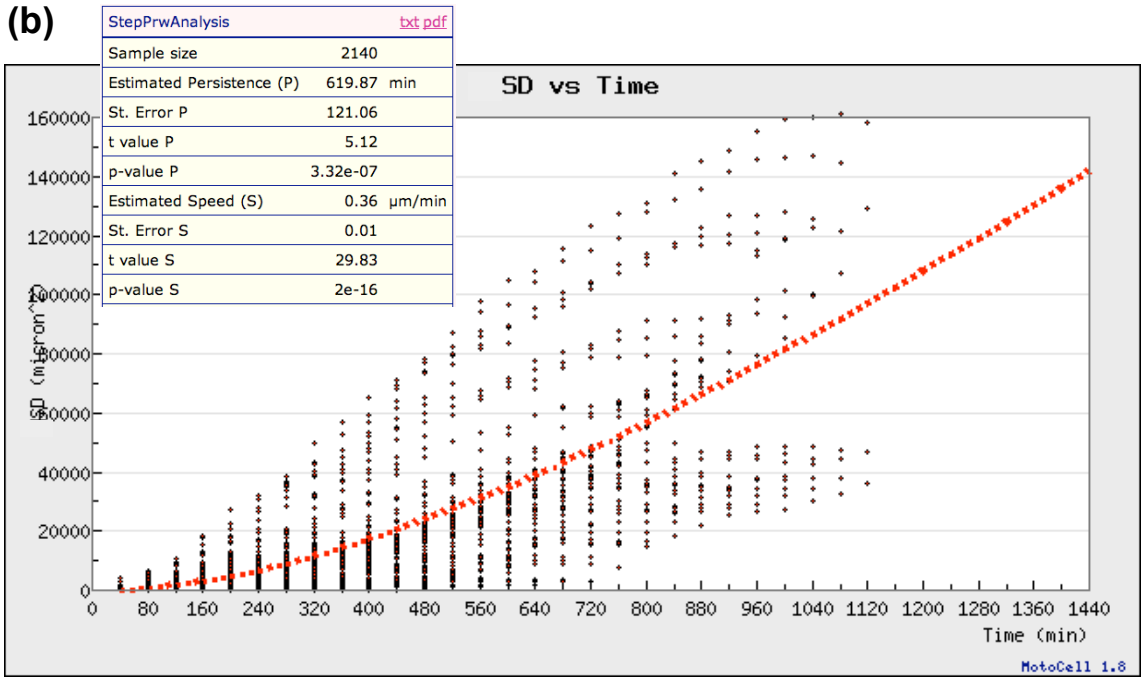


Fig. 3.6. Migration analysis of NIHRas fibroblasts by using the *persistent random walk* model. Squared displacements of NIHRas cells randomly moving **(a)** and under directional stimulus **(b)** have been plotted as a function of the time. The better estimation obtained by using the *persistent random walk* model has been superimposed to the points as a red dotted curve. In order to evaluate the goodness of fit, the R^2 coefficients have been reported in the tables together with estimated persistence and speed, and with the corresponding Student's t-test for significance.

The different degree of persistence observed in parental and transformed fibroblast populations under both random and wound conditions, was confirmed for several datasets analyzed by using the *persistent random walk* model, as reported in table 3.4. As shown, the estimated speed values for both populations do not change much between the different experimental conditions. On the other hand, persistence, which for NIH3T3 cells appears similar in unstimulated and stimulated conditions, for NIHRas is strongly dependent on the experimental conditions, low in random but very high under the effect of directional stimulus.

Taken together, these results allow to conclude that NIHRas fibroblasts respond to the presence of a directional stimulus by increasing directional persistence, defined as the time spent moving in the same direction, while showing a speed similar to that observed in random motility assays.

3.3.3. ***Directionally biased walk***

Biased and correlated random walk also belongs to the superdiffusive family. This model takes into account a directional bias such as that provided by an external stimulus in addition to the correlation between subsequent movement directions, studied by the *persistent random walk* model. In this study, directional analysis was performed by using circular distribution models.

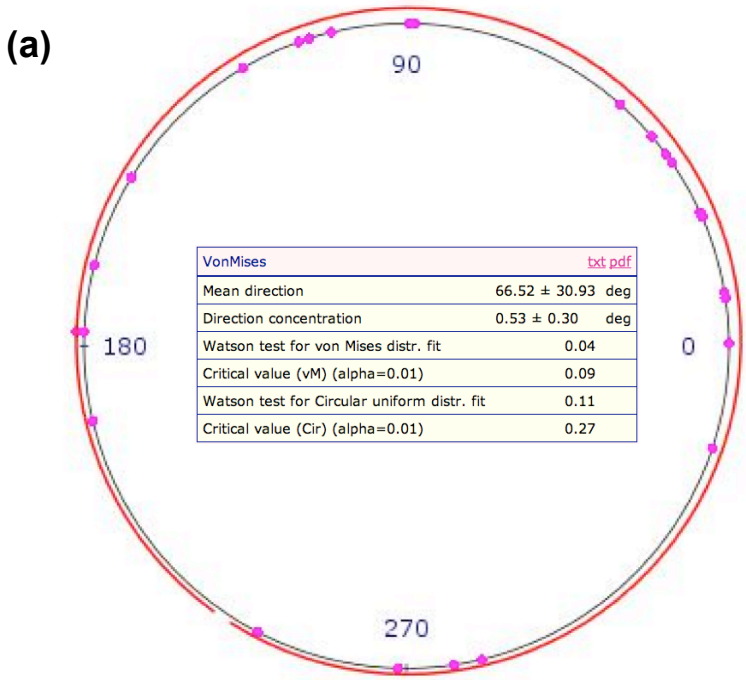
Directional analysis

The effects of a directional bias on migration were evaluated by using circular distributions, in order to distinguish uniform from unidirectional and other non uniform distributions of directions. The von Mises distribution, commonly used as a model for points tightly concentrated around a mean direction, was used for evaluating the unidirectional movement.

In order to attempt to model the behaviour of cell populations according to von Mises distribution, the net displacements of each cell following a path have been used to compute the maximum likelihood estimate of the parameters of the von Mises distribution. Overlays of the theoretical curve with the experimental data obtained for randomly moving NIH3T3 and NIHRas populations, are reported in figure 3.7, and for both populations under stimulated conditions in fig. 3.8. The distribution of net cell displacements in the course of random movement clearly

NIH3T3		
random		
	P (min)	**S (μm/min)**
Dataset 3	360.00	0.20
Dataset 21	291.40	0.27
Dataset 22	256.18	0.25
Dataset 23	220.69	0.32
wound		
Dataset 24	190.14	0.21
Dataset 25	68.77	0.27
Dataset 26	187.17	0.37
Dataset 27	145.44	0.31
Dataset 28	268.75	0.16
NIHRas		
random		
	P (min)	**S (μm/min)**
Dataset 13	86.56	0.29
Dataset 29	67.49	0.56
Dataset 30	60.04	0.58
Dataset 31	77.35	0.48
wound		
Dataset 14	463.90	0.16
Dataset 15	220.34	0.45
Dataset 16	340.30	0.24
Dataset 17	720.00	0.26
Dataset 18	619.87	0.36
Dataset 19	345.70	0.28

Table 3.4. Persistence (P) and speed (S) estimation for several populations of NIH3T3 and NIHRas fibroblasts. Cells were assayed both as randomly moving and under wound stimulus.



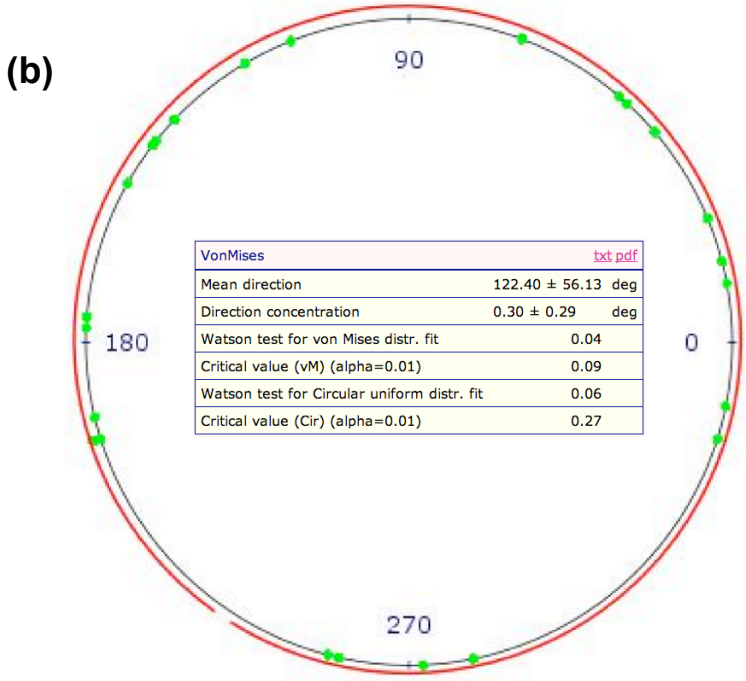


Fig. 3.7. Directions analysis of mouse fibroblasts. The distributions of cell directions for NIH3T3 in **(a)** and for NIHRas in **(b)** in random motion conditions, has been reported together with the theoretical von Mises curve (in red). The Watson test was used to evaluate the goodness of fit of the von Mises and the Circular uniform distributions to the experimental data, as shown in tables.

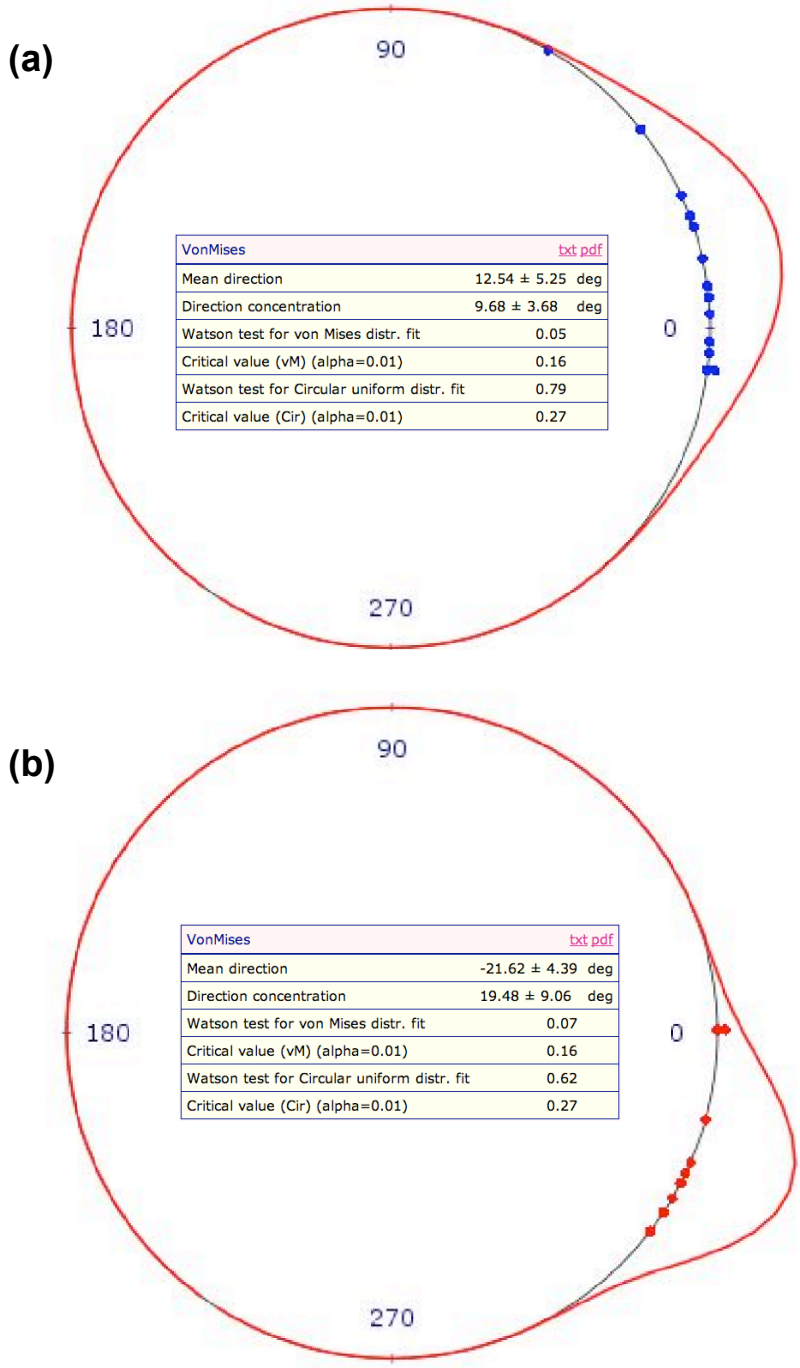


Fig. 3.8. Directions analysis of mouse fibroblasts under a directional stimulus. The distributions of cell directions for NIH3T3 in **(a)** and for NIHRas in **(b)** during wound healing has been reported together with the theoretical von Mises curve (in red). Von Mises parameters, mean direction and directions concentration, have been estimated and reported in the tables. The Watson test was used to evaluate the goodness of fit of the von Mises and the Circular uniform distributions to the experimental data, as shown in tables.

follows a circular uniform distribution (figure 3.7a-b), as also confirmed by the Watson test results reported in the table. As expected for these two populations, the hypothesis of von Mises distribution may also be accepted. A different behaviour was observed for the two fibroblast populations, NIH3T3 and NIHRas, stimulated by the presence of the wound. As shown in figure 3.8, a strong concentration of cell net displacements can be observed, confirmed by the poor fitting shown by the uniform circular distribution (Watson test values well above the threshold for $P>0.01$). This gathering of cell directions is well described by the von Mises distribution, as shown from the good fit of the estimated von Mises distribution, in red, to the experimental data. This goodness of fit was confirmed by the Watson test that allows to accept the von Mises distribution hypothesis. The directional behaviour of both parental and transformed fibroblasts in presence of the wound stimulus was also found in the several analyzed datasets. The estimated parameters for the von Mises distribution that best fits to each different dataset are reported in table 3.5.

In conclusion, by taking into account the results obtained in directional analysis of parental and transformed mouse fibroblasts, it was possible to conclude that a gap introduced into a cell monolayer leads to a strong directional bias of migration. In fact, by observing the distributions of fibroblast net displacements, they were strongly concentrated around a mean direction that closely follows the direction expected because of the stimulus (fig 3.8). As shown before, this directional bias is associated with increased persistence in the same direction and therefore the *biased and correlated random walk* model is the one which best describes the migration behaviour of both NIH3T3 and NIHRas fibroblast populations in presence of a wound.

NIH3T3		
	Mean Direction	**Dir. Concentration**
random		
Dataset 1	66.52 ± 30.93	0.53 ± 0.30
Dataset 2	-17.43 ± 60.61	0.24 ± 0.25
Dataset 3	-77.45 ± 21.12	0.42 ± 0.16
Dataset 4	113.60 ± 21.81	0.95 ± 0.40
wound		
Dataset 5	24.38 ± 41.21	0.64 ± 0.48
Dataset 6	-162.00 ± 19.56	1.56 ± 0.69
Dataset 7	149.70 ± 10.21	3.70 ± 1.50
Dataset 8	12.54 ± 5.25	9.68 ± 3.68
Dataset 9	-3.72 ± 12.91	2.76 ± 1.12
Dataset 10	-21.58 ± 3.52	26.96 ± 11.94
Dataset 11	-28.04 ± 6.42	1.93 ± 0.31
NIHRas		
	Mean Direction	**Dir. Concentration**
random		
Dataset 12	122.40 ± 56.13	0.29 ± 0.29
Dataset 13	89.31 ± 29.21	0.37 ± 0.19
wound		
Dataset 14	97.53 ± 12.21	0.93 ± 0.22
Dataset 15	-21.62 ± 4.39	19.48 ± 9.06
Dataset 16	-9.45 ± 11.05	3.01 ± 1.12
Dataset 17	1.68 ± 9.83	3.64 ± 1.38
Dataset 18	7.68 ± 15.12	1.81 ± 0.66
Dataset 19	-21.08 ± 15.38	2.00 ± 0.78

Table 3.5. Summary of data obtained by estimating mean direction and direction concentration by using the von Mises distribution. Different populations of NIH3T3 and NIHRas fibroblasts were assayed as randomly moving and under wound stimulus.

4. Discussion

Focus of this work was to propose models for the description of normal and transformed fibroblast migration. The starting idea was that, because of the stochastic nature of the migration process, it would be possible to formalize cell motion, by using well-known models of random walk, that could be used as a predictive tool. Random walk models have been applied in the past to a wide range of different problems in the biological sciences, such as the study of animal dispersal in ecology (Bartumeus et al. 2005; Smouse et al. 2010) and, including cell migration, to derive information about the process underlying cell motility (Gail and Bone 1970; Tranquillo et al. 1988; Hartman et al. 1994). In the currently available software tools for analysis of cell migration, the modelling approach has been often neglected, with a few exceptions as in the case of "Cell_motility" by Martens et al. (2006), which allows to study migration by using the *persistent random walk* model. In the context of this thesis, MotoCell, a web application previously developed in our laboratory, was enhanced by implementing the described model fitting procedures. Special care was taken towards statistical significance of the analysis, by introducing various tests to evaluate goodness of fit and by attempting to increase the number of cell movements under evaluation through the use of partial paths as those followed by cells that are lost or appear later within the observation field. The practice of selecting full paths, that usually excludes shorter paths from quantitative analysis, is commonly used by many scientists (Selmeczi et al. 2005; Potdar et al. 2010). However, while in principle producing more homogeneous data, in practice this may result in migration analysis with low statistical significance, because of a reduced number of cells. In its current form, MotoCell allows to study the whole observed population by maintaining a more stable number of cells during time, and without the need for path selection.

Within this work, a first approach was the evaluation of the diffusive features of the migration process. Diffusion analysis allows to distinguish cell populations moving according to a *simple random walk model* from those following other super- or subdiffusive models. Observations about these anomalous diffusion processes have been often associated to cancer cell migration *in vitro* (Dieterich et al. 2007). In the current work, a superdiffusive behaviour was found for both normal and transformed mouse fibroblasts under different experimental conditions, i.e. in random movement, when cells are exploring the surrounding space, as well as in wound healing assays, where migration

is directed to the invasion of the empty space created by scratching the cell layer. The superdiffusive behaviour may have different origins, among which the presence of short displacements alternated with very long displacements as in the *Lévy walk* model, or strong correlation between subsequent motion directions as in the *persistent random walk* model (Campos et al. 2010).

The analysis of fibroblast migration was carried out by fitting these theoretical models to experimental data, through the estimation of equation parameters by regression. The problem of *Lévy walk* model estimation has been long studied in literature (Edwards et al. 2007; White et al. 2008; Sims et al. 2007), with controversial findings about the correct estimation method. For the experimental data discussed in this study, it was found that the best working estimation method is logarithmic binning with normalization: it allows to reduce the error derived from dataset variability by distributing data within bins of different width, that end up by containing a more homogeneous number of data per bin. In the case of *persistent random walk* model, the problem of parameter estimation was dealt with by using the Levenberg-Marquardt algorithm for nonlinear regression, a method that, as in other cases (Heglung et al. 2007), was proved successful also with our data.

Because of the intrinsic features of cells migrating on a substrate, which typically spend some time testing the surroundings by extending protrusions before translocating the cell body to the new position in a short time, the *Lévy walk* model, which typically alternates many short steps with a few longer ones, appeared to be a reasonable candidate. This model was found to work in other physiological and pathological situations, such as movement of cells within the epidermis to form new tissue (Reynolds 2010), or microglial cells moving within brain sections (Grinberg et al. 2011). However, in our tests, the *Lévy walk* model was generally not found to produce good fit, with the only possible exception of wound stimulated NIH3T3.

The *persistent random walk* model was tested as an alternative superdiffusive migration model. This has been used in other studies on the behaviour of cells in standard culture conditions, often in cases where directional persistence over time may be expected (Ware et al. 1998; Kim et al. 2008). Heglund et al. (2007) found that a behaviour compatible with a *persistent random walk* could be observed in diabetic but not in non-diabetic fibroblast migration. In the present work, normal and transformed fibroblast populations appear to closely follow the *persistent random walk* model, although at different degrees of persistence, and also in relation with the experimental conditions. NIH3T3 fibroblasts

in random movement conditions show a relatively high persistence as well as cells moving under the directional stimulus produced by the wound. This behaviour indicates that in these cells, persistence is an intrinsic feature, not affected by the directional stimulus. A possible explanation of this behaviour lies in the time that cells spend to rearrange their cytoskeletal structures, which keeps cells moving for a while along the direction indicated by a previously extended protrusion. On the other hand, a change in the degree of persistence was seen for transformed fibroblasts: they show lower persistence in random motion than in presence of a directional stimulus. This effect is associated with a different path linearity: typically more tortuous paths may be observed in these cells when freely moving in random motility assays. In cells induced to move by the wound stimulus, these paths are more linear and result in higher persistence. This suggests that, in these cells, the presence of the wound strongly affects cell behaviour, by increasing cell persistence well over the values found for the parental NIH3T3 cell line. It is expected that a role is played by cell transformation by the Ras oncogene, which has otherwise been associated to deregulated cell migration induced by constitutive Ras activation which upregulates different intracellular pathways involved in cell adhesion and migration.

This directional persistence observed in subsequent steps, may be associated with the presence of a directional bias determining a motion process following a model defined *biased* and *correlated random walk*. This combined random walk model has been observed for example in microorganism chemotaxis (Hill and Hader 1997; Codling and Hill 2005). In the analyzed populations moving under the wound stimulus, the two aspects of persistence and directionality have been often found associated. However, these two features may also result in separated motion models described by the *persistent random walk*, and in the other case, by a *biased random walk*. In this study, a high persistent behaviour without directional bias was observed for mouse fibroblast populations in random movement conditions. A similar behaviour has been described for polymorphonuclear neutrophils that follow a *persistent* or a *biased random walk* model, when cells are plated respectively in presence of a uniform concentration or a gradient of chemoattractants (Tranquillo et al. 1988).

In the present study the superdiffusive migration models were found to best describe the behaviour of the observed fibroblast populations. In other cases, where migration was studied after introducing variations within the environmental conditions, such as changes in soluble factors or extracellular matrix components, this superdiffusive behaviour was

shown to be lost or modified in favor of a diffusive process (Ware et al. 1998; Rossen et al. 2011). These environmental changes allow to investigate cell migration under conditions that mimic physiological and pathological situations occurring *in vivo*, and to test the effects of different pharmacological treatments. Transformation of cells with the Ras oncogene reveals the influence that constitutive Ras activation exerts on persistence of motion. By introducing a directional bias, as in wound healing assays, a strong directional and persistent migration was seen for both normal and transformed fibroblasts. This may be changed by treatment of transformed cells with PD, an inhibitor of the MAPK pathway, which revealed that, under stimulated condition, these cells lose directionality, despite maintaining time persistence (data not shown). These results allow to conclude that the presented random walk model approach is an effective tool, that may be further developed to generate computational schemes tuned for modelling cell migration under experimental conditions that reproduce specific *in vivo* events.

5. Methods

5.1. Cell propagation

Murine fibroblasts with normal and transformed phenotype, respectively NIH3T3 and NIHRas, were grown in DMEM (Dulbecco's Modified Eagle's Medium) culture medium supplemented with 10% FBS (foetal bovine serum). Penicillin (10U/ml) and streptomycin (10 ng/ml) antiobiotics and L-Glutamine 2mM were added to the culture medium. Each operation was conducted in conditions of sterility under a vertical laminar flow hood. Cells were mantained in incubator at 37 ° C in atmosphere made of 95% air and 5% CO2. Culture conditions and cell growth were followed by optical microscope observations.

Cell propagation was performed by detaching cells with a solution of trypsin/EDTA (trypsin 0.05% and 0.53 mM EDTA) and collecting them with complete culture medium. After centrifugation at 1200 rpm for10 minutes, pellets were suspended in fresh medium, properly diluted and plated again. In order to plate cells at specific densities, the Burker chamber was used to count them.

5.2. Random motility assay

To investigate the random movement ability, cells were seeded (25000/well) in 12 well plates and maintained in complete medium at 37°C in an incubator with 5% CO_2. After 16-18 h, the multiwell was placed in the incubator chamber of the microscope. Phase contrast images at 10x magnification were recorded every 10 minutes for 24 hours with a video camera. The pixel/micron ratio of the acquired images was determined by acquiring images of a square with fixed known dimensions.

5.3. Wound healing assay

The dynamics of wound healing was studied by seeding 25000 cells/cm^2 for NIH3T3, and 50000 cell/cm^2 for smaller NIHRas, in complete medium; 24 h after

plating, the cell layer was scratched with a tip and the closure process was followed for further 24 hours by acquiring 1 digital frame every 10 minutes.

5.4. Microscopy and image acquisition

Images of different samples have been acquired by using the Zeiss Cell Observer system. The system is composed by an inverted microscope (Axiovert 200M), an incubator chamber for observation of living cells and a digital camera (Axiocam H/R). The microscope is connected to an Intel personal computer running Window XP, through the Zeiss acquisition software (Axiovision 6.0) that manages the microscope and captures images. The microscope is equipped with phase contrast optics, while a motorized stage along three axes (x, y and z) permits prolonged automatic acquisition at different positions or levels. An incubator chamber controls the temperature (maintained at 37 °C), the CO_2 (at 5%) and the humidity for the long observation of living cells. Within this work, digital frames were acquired as 14 bit images of 650x514 pixels.

5.5. Cell tracking and quantitative analysis

The acquired movies were analyzed with MotoCell (an application previously developed and revised during this work as described in the Results section). With this tool, cell paths were semi-automatically tracked by inputting and storing x and y cell coordinates at each time step. Cells were typically followed over a 24 hour period and paths were tracked every 40 minutes.

Concerning the quantitative evaluation of cell migration, some definitions are necessary in order to understand the terms used:

- Step is the elementary displacement between two subsequent time points in which the observation period has been divided;
- Step length is the distance covered at each step;
- Squared displacement is the square of distance moved by cell over different time interval;
- Mean squared displacement is obtained by averaging the squared displacements over all tracked cells at each time interval;
- Direction is the angle of a vector describing the net displacement of the cell.

In the present study, the fit of various model functions to experimental data was tested. The results of the fitting analysis was associated with different tests of inferential statistics in order to evaluate the significance of parameters estimated from the models, and to assess the goodness of fit of the estimated values to the experimental data. Inferential statistics tools, integrated in MotoCell, are: the Student's t-test used to test the reliability of the estimated parameters; R^2 determination coefficient together with the Kolmogorov-Smirnov (Stephens 1974) and the Watson tests (Watson 1961; Lockhart 1985) used to quantify the goodness of fit between estimated and experimental data.

5.6. Generation of simulated datasets

In order to test the procedures, simulated dataset were generated in order to assay the procedures with samples controlled in number, type and variability. The "adehabitat" package available within R environment, consisting of a collection of tools developed for the analysis of habitat selection of animals (Calenge 2007), was used for this purpose. By using the simulations function group, dataset were generated for *simple random walk* model ("simm.brown"), *correlated random walk* model ("simm.crw"), *Lévy walk* ("simm.levy") and *biased random walk* ("simm.mou"). Simulated datasets for the *persistent random walk* model were generated by using model equation with different values for the two parameters speed and persistence, and by adding a *Gaussian* noise N (0,100).

5.7. Parameter Estimation

R based methods were used to obtain the estimation of the different parameters involved in the discussed models. Within step length distribution analysis, the method "fitdistr" (MASS package) was used to estimate the parameters of the different statistical distributions, by means of the maximum likelihood estimation method; the cumulative generator functions of each statistical distribution were used to compute the corresponding estimated curves (pnorm, pweibull, plnorm, pcauchy, pexp, ppois, pgeom). For the maximum likelihood estimation of parameters κ and μ, relative to the circular *von Mises* distribution, the "mle.vonmises" function of the "circular" R package was used. In the study of regression methods for the analysis of

cell migration through the *Lévy walk* model, the following function were used: ”mle” (package “stats4”) for the maximum likelihood estimation method, “lm” for linear regression in both cumulative distribution and logarithmic binning with normalization method, after logarithmic conversion of data. In order to achieve the nonlinear regression analysis, both in the study of diffusion and in the analysis of movement through the *persistent random walk* model (PRW), “modFit” (package “FME”) was used with different input options. The algorithm used for regression analysis was specified for this function as an option; the Newton algorithm was used to determine the λ exponent of diffusion relation; this, together with the Gradient and Levenberg-Marquardt algorithm (Moré 1978), were used for the estimation of speed and persistence parameters of PRW equation.

5.8. Software development

The procedures to study cell motion with random walk models were developed by using R (website: R), an open source environment for statistical computations and graphics, which runs on a wide variety of platforms. R is a combination of different statistics packages and programming language. For the applications described in the present study, R scripts were used for each of the performed modelling analysis. In the R environment, there are several packages supplied for statistical techniques, which can be extended via packages available through the CRAN family. In this study the additional packages used were: “MASS” and “matlab”, for implementation of methods derived from the S and the MATLAB languages; “FME” and “stats4”, for nonlinear regression analysis; “circular”, for circular data analysis; “Hmisc”, for linear data analysis and high-level graphics; “adehabitat”, for dataset simulations. The software version used was R 2.12.1.

The revision of MotoCell and the implementation of new features described in the Results section, have been reported in Cantarella et al. 2009. Software development was carried out by using PHP version 5.2.

References

Abramoff, M. D., P. J. Magalhães, et al. (2004). "Image processing with ImageJ." Biophotonics international **Volume: 11**(Issue: 7 (2004)): pp. 36-42.

Al-Shanti, N., S. H. Faulkner, et al. (2011). "A semi-automated programme for tracking myoblast migration following mechanical damage: manipulation by chemical inhibitors." Cell Physiol Biochem **27**(6): 625-36.

Alt, W. and G. Hoffmann (1990). "Correlation analysis of two-dimensional locomotion paths." Biological Motion **89**: 254.

Ambravaneswaran, V., I. Y. Wong, et al. (2010). "Directional decisions during neutrophil chemotaxis inside bifurcating channels." Integr Biol (Camb) **2**(11-12): 639-47.

Arocena, M., M. Zhao, et al. (2010). "A time-lapse and quantitative modelling analysis of neural stem cell motion in the absence of directional cues and in electric fields." J Neurosci Res **88**(15): 3267-74.

Bartumeus (2007). "Lévy processes in animal movement: an evolutionary hypothesis." Fractals **15**(2): 151-162.

Bartumeus, F., M. G. E. Da Luz, et al. (2005). "Animal search strategies: A quantitative. random-walk analysis." Ecology **86**(11): 3078-3087.

Beauchemin, C., N. M. Dixit, et al. (2007). "Characterizing T cell movement within lymph nodes in the absence of antigen." J Immunol **178**(9): 5505-12.

Bentley (2006). Modelling circular data using a mixture of von mises and uniform distributions, SIMON FRASER UNIVERSITY.

Bindschadler, M. and J. L. McGrath (2007). "Sheet migration by wounded monolayers as an emergent property of single-cell dynamics." J Cell Sci **120**(Pt 5): 876-84.

Bos, J. L. (1989). "ras oncogenes in human cancer: a review." Cancer Res **49**(17): 4682-9.

Calenge (2007). "Exploring Habitat Selection by Wildlife with adehabitat." Journal of Statistical Software **22**(6).

Campos, D., V. Mendez, et al. (2010). "Persistent random motion: uncovering cell migration dynamics." J Theor Biol **267**(4): 526-34.

Cantarella, C., L. Sepe, et al. (2009). "Analysis and modelling of motility of cell populations with MotoCell." BMC Bioinformatics **10 Suppl 12**: S12.

Codling, E. A. and N. A. Hill (2005). "Sampling rate effects on measurements of correlated and biased random walks." J Theor Biol **233**(4): 573-88.

Codling, E. A., M. J. Plank, et al. (2008). "Random walk models in biology." J R Soc Interface **5**(25): 813-34.

Crane, I. J. and J. Liversidge (2008). "Mechanisms of leukocyte migration across the blood-retina barrier." Semin Immunopathol **30**(2): 165-77.

Dickinson, R. B. and R. T. Tranquillo (1993). "Optimal estimation of cell movement indices from the statistical analysis of cell tracking data." AIChE J **39**(12): 1995-2010.

Dieterich, P., R. Klages, et al. (2008). "Anomalous dynamics of cell migration." Proceedings of the National Academy of Sciences of the United States of America **105**(2): 459-463.

Dunn, G. A. (1983). "Characterising a kinesis response: time averaged measures of cell speed and directional persistence." Agents Actions Suppl **12**: 14-33.

Edwards, A. M., R. A. Phillips, et al. (2007). "Revisiting Levy flight search patterns of wandering albatrosses, bumblebees and deer." Nature **449**(7165): 1044-U5.

Fang, K. S., E. Ionides, et al. (1999). "Epidermal growth factor receptor relocalization and kinase activity are necessary for directional migration of keratinocytes in DC electric fields." J Cell Sci **112 (Pt 12)**: 1967-78.

Fenteany, G., P. A. Janmey, et al. (2000). "Signaling pathways and cell mechanics involved in wound closure by epithelial cell sheets." Curr Biol **10**(14): 831-8.

Flo, R. W., A. Naess, et al. (1994). "A longitudinal study of phagocyte function in HIV-infected patients." AIDS **8**(6): 771-7.

Fooksman, D. R., T. A. Schwickert, et al. (2010). "Development and migration of plasma cells in the mouse lymph node." Immunity **33**(1): 118-27.

Gail, M. H. and C. W. Boone (1970). "The locomotion of mouse fibroblasts in tissue culture." Biophys J **10**(10): 980-93.

Geback, T., M. M. Schulz, et al. (2009). "TScratch: a novel and simple software tool for automated analysis of monolayer wound healing assays." Biotechniques **46**(4): 265-74.

Geiger, B., J. P. Spatz, et al. (2009). "Environmental sensing through focal adhesions." Nat Rev Mol Cell Biol **10**(1): 21-33.

Gildea, J. J., W. L. Golden, et al. (2000). "Genetic and phenotypic changes associated with the acquisition of tumorigenicity in human bladder cancer." Genes Chromosomes Cancer **27**(3): 252-63.

Grinberg, Y. Y., J. G. Milton, et al. (2011). "Spreading depression sends microglia on Levy flights." PLoS One **6**(4): e19294.

Groh, A. and A. K. Louis (2010). "Stochastic modelling of biased cell migration and collagen matrix modification." J Math Biol **61**(5): 617-47.

Hartman, R. S., K. Lau, et al. (1994). "The fundamental motor of the human neutrophil is not random: evidence for local non-Markov movement in neutrophils." Biophys J **67**(6): 2535-45.

Harwood, C. S., K. Fosnaugh, et al. (1989). "Flagellation of Pseudomonas putida and analysis of its motile behavior." J Bacteriol **171**(7): 4063-6.

Heglund, R., B. Inouye, et al. (2007). "Methods for the quantitative analysis of diabetic fibroblast migration." Journal of Undergraduate Research in Bioengineering: 5357.

Heit, B. and P. Kubes (2003). "Measuring chemotaxis and chemokinesis: the under-agarose cell migration assay." Sci STKE **2003**(170): PL5.

Heit, B., S. Tavener, et al. (2002). "An intracellular signaling hierarchy determines direction of migration in opposing chemotactic gradients." J Cell Biol **159**(1): 91-102.

Hill, N. A. and D. P. Hader (1997). "A biased random walk model for the trajectories of swimming micro-organisms." J Theor Biol **186**(4): 503-26.

Huttenlocher, A., S. P. Palecek, et al. (1997). "Regulation of cell migration by the calcium-dependent protease calpain." J Biol Chem **272**(52): 32719-22.

Keller, E. F. and L. A. Segel (1971). "Traveling bands of chemotactic bacteria: a theoretical analysis." J Theor Biol **30**(2): 235-48.

Kim, H. D., T. W. Guo, et al. (2008). "Epidermal growth factor-induced enhancement of glioblastoma cell migration in 3D arises from an intrinsic increase in speed but an extrinsic matrix- and proteolysis-dependent increase in persistence." Mol Biol Cell **19**(10): 4249-59.

Krauss, A. H., A. L. Nieves, et al. (1994). "Determination of leukotriene effects on human neutrophil chemotaxis in vitro by differential assessment of cell motility and polarity." J Leukoc Biol **55**(2): 201-8.

Krawczyk, W. S. (1971). "A pattern of epidermal cell migration during wound healing." J Cell Biol **49**(2): 247-63.

Kurosaka, S. and A. Kashina (2008). "Cell biology of embryonic migration." Birth Defects Res C Embryo Today **84**(2): 102-22.

Li, L., S. F. Norrelykke, et al. (2008). "Persistent cell motion in the absence of external signals: a search strategy for eukaryotic cells." PLoS One **3**(5): e2093.

Liang, C. C., A. Y. Park, et al. (2007). "In vitro scratch assay: a convenient and inexpensive method for analysis of cell migration in vitro." Nat Protoc **2**(2): 329-33.

Locascio, A. and M. A. Nieto (2001). "Cell movements during vertebrate development: integrated tissue behaviour versus individual cell migration." Current Opinion in Genetics & Development **11**(4): 464-469.

Lockhart (1985). "Tests of fit for the von Mises distribution " Biometrika **72** (3): 6.

Maeda, Y. T., J. Inose, et al. (2008). "Ordered patterns of cell shape and orientational correlation during spontaneous cell migration." PLoS One **3**(11): e3734.

Martens, L., G. Monsieur, et al. (2006). "Cell_motility: a cross-platform, open source application for the study of cell motion paths." BMC Bioinformatics **7**: 289.

Matthaus, F., M. Jagodic, et al. (2009). "E. coli superdiffusion and chemotaxis-search strategy, precision, and motility." Biophys J **97**(4): 946-57.

Mitchison, T. J. and L. P. Cramer (1996). "Actin-based cell motility and cell locomotion." Cell **84**(3): 371-9.

Moré (1978). "The Levenberg-Marquardt algorithm: Implementation and theory." NUMERICAL ANALYSIS Lecture Notes in Mathematics **630**: 17.

Nelson, R. D., P. G. Quie, et al. (1975). "Chemotaxis under agarose: a new and simple method for measuring chemotaxis and spontaneous migration of human polymorphonuclear leukocytes and monocytes." J Immunol **115**(6): 1650-6.

Nguyen, D. H., A. D. Catling, et al. (1999). "Myosin light chain kinase functions downstream of Ras/ERK to promote migration of urokinase-type plasminogen activator-stimulated cells in an integrin-selective manner." J Cell Biol **146**(1): 149-64.

Nicolau, D. V., Jr., J. P. Armitage, et al. (2009). "Directional persistence and the optimality of run-and-tumble chemotaxis." Comput Biol Chem **33**(4): 269-74.

O'Hayre, M., C. L. Salanga, et al. (2008). "Chemokines and cancer: migration, intracellular signalling and intercellular communication in the microenvironment." Biochem J **409**(3): 635-49.

Petrie, R. J., A. D. Doyle, et al. (2009). "Random versus directionally persistent cell migration." Nat Rev Mol Cell Biol **10**(8): 538-49.

Pollard, T. D. and G. G. Borisy (2003). "Cellular motility driven by assembly and disassembly of actin filaments." Cell **112**(4): 453-65.

Potdar, A. A., J. Jeon, et al. (2010). "Human mammary epithelial cells exhibit a bimodal correlated random walk pattern." PLoS One **5**(3): e9636.

Potdar, A. A., J. Lu, et al. (2009). "Bimodal analysis of mammary epithelial cell migration in two dimensions." Ann Biomed Eng **37**(1): 230-45.

Prat, A., K. Biernacki, et al. (2002). "Migration of multiple sclerosis lymphocytes through brain endothelium." Arch Neurol **59**(3): 391-7.

Reynolds, A. M. (2010). "Can spontaneous cell movements be modelled as Lévy walks?" Physica A **389**: 273-277.

Riveline, D., E. Zamir, et al. (2001). "Focal contacts as mechanosensors: externally applied local mechanical force induces growth of focal contacts by an mDia1-dependent and ROCK-independent mechanism." J Cell Biol **153**(6): 1175-86.

Rossen, N. S., A. J. Hansen, et al. (2011). "Arachidonic acid randomizes endothelial cell motion and regulates adhesion and migration." PLoS One **6**(9): e25196.

Schneider, L., M. Cammer, et al. (2010). "Directional cell migration and chemotaxis in wound healing response to PDGF-AA are coordinated by the primary cilium in fibroblasts." Cell Physiol Biochem **25**(2-3): 279-92.

Selmeczi, D., S. Mosler, et al. (2005). "Cell motility as persistent random motion: theories from experiments." Biophys J **89**(2): 912-31.

Selmeczi D., L. L., L. I.I. Pedersen, S. F. Nrrelykke, P. H. Hagedorn, S. Mosler, N. B. Larsen, E. C. Cox and H. Flyvbjerg (2008). "Cell motility as random motion: A review." Eur. Phys. J. Special Topics **157**: 1-15.

Sims, D. W., D. Righton, et al. (2007). "Minimizing errors in identifying Levy flight behaviour of organisms." J Anim Ecol **76**(2): 222-9.

Smouse, P. E., S. Focardi, et al. (2010). "Stochastic modelling of animal movement." Philos Trans R Soc Lond B Biol Sci **365**(1550): 2201-11.

Stephens (1974). "EDF Statistics for Goodness of Fit and Some Comparisons." Journal of the American Statistical Association **69**(347).

Takagi, H., M. J. Sato, et al. (2008). "Functional analysis of spontaneous cell movement under different physiological conditions." PLoS One **3**(7): e2648.

Tanaka, D. H., M. Yanagida, et al. (2009). "Random walk behavior of migrating cortical interneurons in the marginal zone: time-lapse analysis in flat-mount cortex." J Neurosci **29**(5): 1300-11.

Tranquillo, R. T., D. A. Lauffenburger, et al. (1988). "A stochastic model for leukocyte random motility and chemotaxis based on receptor binding fluctuations." J Cell Biol **106**(2): 303-9.

Ueoka, Y., K. Kato, et al. (2000). "Hepatocyte growth factor modulates motility and invasiveness of ovarian carcinomas via Ras-mediated pathway." Br J Cancer **82**(4): 891-9.

Van Haastert, P. J. (2010). "A model for a correlated random walk based on the ordered extension of pseudopodia." PLoS Comput Biol **6**(8).

Van Haastert, P. J. and L. Bosgraaf (2009). "Food searching strategy of amoeboid cells by starvation induced run length extension." PLoS One **4**(8): e6814.

Ware, M. F., A. Wells, et al. (1998). "Epidermal growth factor alters fibroblast migration speed and directional persistence reciprocally and in a matrix-dependent manner." J Cell Sci **111 (Pt 16)**: 2423-32.

Watson (1961). "Goodness-Of-Fit Tests on a Circle " Biometrika **48** (1 and 2).

Wessels, D., S. Kuhl, et al. (2006). "Application of 2D and 3D DIAS to motion analysis of live cells in transmission and confocal microscopy imaging." Methods Mol Biol **346**: 261-79.

Wessels, D., S. Kuhl, et al. (2009). "2D and 3D quantitative analysis of cell motility and cytoskeletal dynamics." Methods Mol Biol **586**: 315-35.

White, E. P., B. J. Enquist, et al. (2008). "On estimating the exponent of power-law frequency distributions." Ecology **89**(4): 905-12.

Wong, S. Y., K. H. Chiam, et al. (2010). "Computational model of cell positioning: directed and collective migration in the intestinal crypt epithelium." J R Soc Interface **7 Suppl 3**: S351-63.

Wright, A., Y. H. Li, et al. (2008). "The differential effect of endothelial cell factors on in vitro motility of malignant and non-malignant cells." Ann Biomed Eng **36**(6): 958-69.

Yip, S. C., M. El-Sibai, et al. (2007). "The distinct roles of Ras and Rac in PI 3-kinase-dependent protrusion during EGF-stimulated cell migration." J Cell Sci **120**(Pt 17): 3138-46.

Web sites

Metam - http://www.moleculardevices.com/pages/software/metamorph.html

MotoCell - http://motocell.ceinge.unina.it

MtrackJ - http://www.imagescience.org/meijering/software/mtrackj/

ParTrac - http://weeman.inf.ethz.ch/particletracker/

R – http://www.r-project.org

TScratch

http://www.cse-lab.ethz.ch/index.php?&option=com_content&view=article&id=363

Zeiss - http://www.zeiss.com

List of tables and figures