

# Effect of step size on probabilistic streamlines: implications for the interpretation of connectivity analyses.

J.-D. Tournier<sup>1,2</sup>, F. Calamante<sup>1,2</sup>, and A. Connelly<sup>1,2</sup>

<sup>1</sup>Brain Research Institute, Florey Neuroscience Institutes, Melbourne, Victoria, Australia, <sup>2</sup>Department of Medicine, University of Melbourne, Melbourne, Victoria, Australia

**Introduction:** To perform robust tractography, it is now clear that higher order models are needed to resolve crossing fibres [e.g. 1], and that probabilistic algorithms are required that take into account the inherently noisy nature of DWI data [e.g. 2]. It is now common therefore to characterise fibre orientations as a distribution (often referred to as the fibre orientation distribution (FOD) or the fibre orientation density function (FODF)) rather than as a single discrete orientation (or a small set of such orientations). In such a framework, random samples can be generated from the distribution, making it ideal for use in probabilistic algorithms. Many methods are now available to estimate these distributions (e.g. spherical deconvolution [3]), or to generate a set of samples from this distribution (e.g. Bayesian inference via MCMC sampling [2,4], or ‘bootstrapping’ [5,6]). However, the way this information is handled by the tractography algorithm (typically an adaptation of the simple streamlines algorithm) can also potentially influence the results. In this study, we show that the step size used by probabilistic streamlines tractography algorithms can indeed have a dramatic effect on the estimated uncertainty of the tracks, in addition to the known effect of curvature [7].

**Theory:** The effect of step size on probabilistic tractography results can be estimated using perturbation theory, as in [8]. In a straight, coherent fibre bundle such as that shown in Figure 1, each orientation sample will effectively be drawn from the same distribution with variance  $\sigma_0^2$ . The variance of the track position in the plane orthogonal to the fibre direction will be approximately  $\delta x^2 \times \sigma_0^2$ , where  $\delta x$  is the step size (this holds for small angles). After  $N$  steps, this variance will scale as  $N \delta x^2 \sigma_0^2$ . Since the distance  $d$  travelled after  $N$  steps is  $N \delta x$ , the variance in the dispersion of the tracking for a fixed distance  $d$  scales as  $d \delta x \sigma_0^2$ . The standard deviation of the track end-points after tracking a distance  $d$  is therefore dependent on the square root of the step size, according to  $\sigma_y = \sigma_0 \sqrt{d \delta x}$ .

**Methods:** Simulated FOD images were generated representing a single coherent fibre bundle, assuming a 2mm isotropic voxel size, as shown in Figure 1. FODs were represented using the spherical harmonics (SH) basis truncated at maximum harmonic degree  $l_{max} = 12$ . Probabilistic streamlines tractography was performed using the following algorithm. At each step, the SH coefficients of the FOD at the current point were obtained by trilinear interpolation. A random sample was generated from the FOD using rejection sampling, constrained to lie within  $30^\circ$  of the incoming direction of tracking, with a minimum FOD amplitude of 0.1. The algorithm then stepped along this direction by the user-specified step size, until the track exited the data set boundary, or if no suitable sample could be found after 1,000 attempts (this normally occurs if the FOD amplitude falls below the threshold of 0.1 for all candidate orientations). 1,000 such streamlines were generated over a range of step sizes, and the distribution of end-points in the plane transverse to the direction of tracking a distance of 40 voxels (80mm) away from the seed point was recorded.

In vivo data were acquired from a healthy volunteer on a 3T Siemens Trio system using a DW twice-refocused echo-planar imaging (EPI) sequence (FOV=240x240mm, matrix size=104x104, 54 contiguous slices, 2.3mm slice thickness, 150 DW directions,  $b=3000$  s/mm<sup>2</sup>). FOD were computed using constrained spherical deconvolution (CSD) at  $l_{max} = 10$  using MRtrix [9]. 1,000 tracks were generated using the algorithm described above, seeding from a point within the cortico-spinal tract at the level of the medulla oblongata. This was repeated for a range of step sizes.

**Results:** The dispersion in the tracking results generated from the simulated data is clearly dependent on the step size, as can be seen in Figure 2. In this synthetic example, the standard deviation of the track end-points does indeed scale with the square root of the step size, as predicted by the simple theoretical example above. This step size dependence was also observed with the in vivo data, with a marked reduction in the spread of the tracks generated using the smaller step sizes (Figure 3).

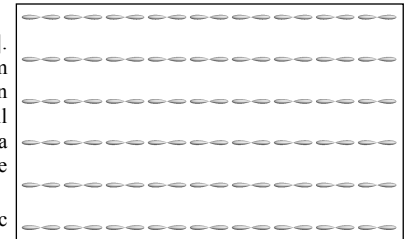
**Discussion:** The dependence of the spread of probabilistic tracking results on the step size clearly has implications for their interpretation. As shown here, the use of a small step size will tend to reduce the spread of the results and give a grossly misleading representation of the probability of a connection.

On the other hand, deviation errors in deterministic streamlines algorithms have previously been shown to be independent of step-size [8]. A more robust measure of the uncertainty would therefore be obtained by running an otherwise equivalent deterministic algorithm on multiple independent repeats of the same acquisition. This can be achieved in practice using ‘bootstrap’-based methods [e.g. 5,6], whereby repeated samples of the entire dataset are generated with distinct noise realisations. Such ‘bootstrap’ approaches can therefore be made to be immune to step size effects, provided that for the each track, a single bootstrap realisation is used for every voxel, no matter how many steps are taken per voxel. Note that while some approaches seem to satisfy the criterion above, the use of a probabilistic trilinear interpolation method [2] will unfortunately re-introduce the problem [e.g. 5].

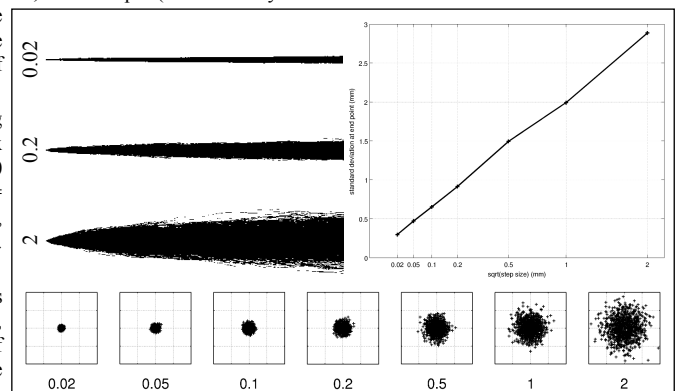
Probabilistic streamlines approaches that generate a fresh sample at each step will suffer from the limitations described here. For these methods, a more accurate measure of the spread may only be obtained by using a step size similar to the voxel size, which will unfortunately introduce ‘overshoot’ errors in regions of significant curvature. These errors can however be minimised using a second-order method, such as the recently proposed iFOD2 algorithm [10] that would allow tracking with the appropriate step size of one voxel without large errors due to curvature.

**Conclusion:** Probabilistic streamlines tractography algorithms are prone to underestimation of the uncertainty of the generated tracks if a small step size is used. While some approaches can be made robust to these effects, many algorithms in current use will be affected.

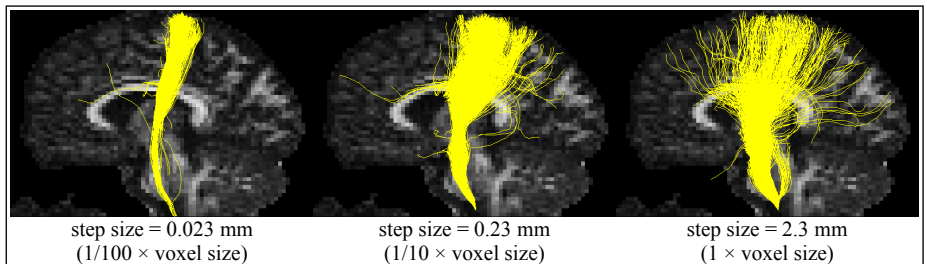
**References:** [1] Jeurissen et al., Proc ISRM 18, #573 (2010). [2] Behrens et al., MRM 50:1077-88 (2003). [3] Tournier et al., NeuroImage 23:1176-85 (2004). [4] Hosey et al., MRM 54:1480-9 (2005). [5] Haroon et al., IEEE-TMI 28: 535-50 (2009). [6] Jeurissen et al., HBM doi: 10.1002/hbm.21032 (2010). [7] Tournier et al., MRM 47: 701-8 (2002). [8] Anderson, MRM 46:1174-88 (2001). [9] <http://www.brain.org.au/software> [10] Tournier et al., Proc ISMRM 18, #1670 (2010).



**Figure 1:** a section of the simulated FOD image for a straight, coherent fibre bundle running left-right. The FOD is represented by a glyph centred on its corresponding voxel.



**Figure 2:** the effect of step size in a simulated FOD image. Left column: example tracks for 3 step sizes. Right: the standard deviation of the dispersion of the tracks at the end-points (y-axis), plotted against the square root of the step size (x-axis). Bottom: the actual track end-points plotted in the plane orthogonal to the fibre direction, for all step sizes used.



**Figure 3:** the effect of step size on in vivo data. For each step size, 1,000 probabilistic streamlines were seeded from a single point in the cortico-spinal tract. These are displayed as a sagittal projection, overlaid on the corresponding FA map.