Towards a better understanding of the meaning of diffusion kurtosis in the human brain. Insights from theoretical models and simulations.

Norbert A'Campo
Master thesis
Biomedical image sciences
University of Utrecht
n.c.acampo@students.uu.nl

✦

Abstract—DKI is an emerging technique based on non-Gaussian water diffusion analysis. A better explanation of the meaning of kurtosis in terms of features of microstructural anatomy would make DKI more concrete and accessible for clinicians. This work aims at a better understanding on the meaning of kurtosis in the human brain. A literature study is performed on possible causes of changes in kurtosis according to some popular diffusion models. First we review some of these models and their validity. Second we give an overview of anatomical links between diffusion models and kurtosis metrics that are implied by the reviewed models. Tortuosity differences, exchange, axon/neurite fraction, axon radius, neurite distribution orientation, intrinsic axon/neurite diffusivity, axon bending, bulging, breaking and crimping are in this way found to affect kurtosis metrics. Axon orientation dispersion, CSF contamination, astrocytes, trapped water and axon undulation could possibly also affect kurtosis metrics. Several rules of thumb are stated which provides anatomical explanations for a positively or negatively observed change in a kurtosis metric.

Index Terms—DKI; Diffusion kurtosis imaging; Non-Gaussian diffusion

CONTENTS

1 INTRODUCTION

Due to its unique sensitivity to the directional origin of tissue microstructure, diffusion weighted magnetic resonance imaging (MRI) has found many applications in clinical and fundamental science [1]. Water molecules move randomly, and hence follow macroscopically a probability density function (pdf) like in figure 1. This pdf is Gaussian in a glass of water, but deviates in the brain because of hindrance and restriction from tissue microstructure. Measuring this deviation from Gaussianity could obtain extra information about the anatomical status of this tissue. Diffusion kurtosis imaging (DKI) is a model-independent MRI method that provides an estimate for the kurtosis of the diffusion pdf, which is a statistic of the deviation from a Gaussian distribution [2].

As DKI is model-independent and kurtosis is just a statistic which measures the degree of non-Gaussianity of the diffusion distribution, a certain kurtosis value measured by DKI does nothing less and nothing more than quantifying how non-Gaussian the diffusion distribution is. However, if a pathology would change the diffusion kurtosis significantly, then one could possibly detect and monitor this pathology with DKI. According to Steven (2014) [3], kurtosis has many potential clinical applications in being such a pathologic biomarker, but more studies are required to link changes in kurtosis to changes in anatomical parameters such as membrane and myelin integrity. These links between kurtosis and anatomical parameters can be investigated empirically [4], but also theoretically. Since diffusion models are essentially Fourier transforms of a diffusion distribution, they imply by definition a certain kurtosis. Jensen (2010) [5] consulted the literature for such relations between diffusion models and kurtosis metrics. This resulted in an overview of some anatomical features that theoretically affects kurtosis: tissue heterogeneity, complexity and diffusional exchange. However, recent preclinical studies demand a more specific explanation of kurtosis, e.g. cerebrospinal fluid (CSF) contamination Yang (2013) [6] and astrogliosis Zhuo (2012) [7].

This work builds further on the theoretical and literature-based approach of Jensen (2010) [5] to give an answer to the question: what can cause changes in kurtosis in the human brain? This answer should ideally be specific enough to answer the demands from the mentioned preclinical studies whereas it should be at least more specific than "tissue heterogeneity, tissue complexity and exchange", since we want to extend rather than repeat the conclusions of Jensen (2010) [5]. The 'causes of kurtosis' are in this work ultimately formulated as 'rules of thumb', such that, after being investigated and validated more extensively, these rules could be used in the (pre)clinic. However, one is encouraged to follow the whole reasoning behind these 'rules of thumb' to get a feeling for which and how anatomical parameters affect kurtosis.

This work is structured as follows. We start with a general introduction in DKI. The second section is a literature review on diffusion models and the kurtosis metrics they imply. Since not all models are possibly equally sufficient also the models itself are reviewed in this section, since the validity of the kurtosis properties depends critically on the validity of the models from which they are derived. In the third section, the possible causes of kurtosis will be discussed, resulting in a comprehensive overview of 'rules of thumb' on kurtosis. In the fourth resp. fifth section the limitations of this work and some directions for future work will be discussed. Finally we end with a conclusion. A list of abbreviations can be found at the end of this work.

2 INTRODUCTION TO DIFFUSION MRI

2.1 Diffusion kurtosis imaging

In what follows some very basic background in probability and statistics is required to follow the whole story. Those readers who are not familiar with concepts such as random variable, probability density function, expectation value, variance and covariance, are referred to appendix A or to any introductory book on probability and statistics like [8]. In this section we first start with some underlying mathematics of diffusion. Then we will study some physics of diffusion, whereafter we will see how the kurtosis of a diffusion process can be measured by MRI. This theoretical background on DKI is also well described in Veraart (2013) [9], however, possibly with slightly different notations.

2.1.1 Mathematics of diffusion

To understand diffusion kurtosis imaging well, one should first aim to understand some underlying mathematical and physical concepts. In this mathematical section, V denotes a random variable.

2.1.1.1 **Kurtosis**: Water diffusion in white and gray matter of the brain is not Gaussian distributed [10]. One quantity that can give information about the deviation from the Gaussian distribution is the kurtosis, which we define for the scope of this work as[2]:

$$K[V] = \frac{E[V^4]}{E[V^2]^2} - 3 \tag{1}$$

The subtraction of 3 is a convention such that the Gaussian distribution has a kurtosis of 0. In the DKI literature, a diffusion distribution with positive kurtosis is often associated with a higher peak or with heavier tails [5].

2.1.1.2 **The function $E[e^{-isV}]$** : There are a few facts to consider about the function $E[e^{-isV}]$ of the pdf $f_V(s)$ of V. 1. $\frac{d^n E[e^{-isV}]}{ds^n}|_{s=0} = (-i)^n E[V^n]$, provided that the n-th derivative exists. 2. If V has a symmetric pdf, then $E[e^{-isV}] = E[cos(sV)]$ 3. if $V \sim N(0, \sigma^2)$, then

$$E[e^{-isV}] = e^{-\frac{\sigma^2}{2}s^2}, \tag{2}$$

for which a derivation can be find in appendix D.

2.1.1.3 **Cumulant expansion**: Suppose V follows a symmetric pdf. Then $E[e^{isV}] = E[cos(sV)]$. For small angles θ the cosine can be approximated by $cos(\theta) \approx 1 - \frac{\theta^2}{2} + \frac{\theta^4}{4!}$. Hence for a symmetrically distributed random variable, since the expected value operator is linear, one can make the approximation

$$E[e^{-isV}] = E[cos(sV)] = 1 - \frac{E[V^2]}{2!}s^2 + \frac{E[V^4]}{4!}s^4 + O(s^6). \tag{3}$$

By taking the log of $E[cos(sV)]$ and using the Taylor approximation $log(1 + x) = x - \frac{x^2}{2} + O(x^3)$, this gives the variance

---

2. This is actually the definition of *excess* kurtosis. Because in the rest of the DKI literature the term kurtosis is used instead, we will also speak about kurtosis without the excess prefix.

and the kurtosis of V:

$$\log(E[\cos(sV)]) = \log(1 - \frac{E[V^2]}{2!}s^2 + \frac{E[V^4]}{4!}s^4 + O(t^6))$$

$$= -\frac{1}{2}E[V^2]s^2 + \frac{1}{24}(E[V^4] - 3E[V^2]E[V^2])s^4$$

$$+ O(s^6) \tag{4}$$

$$= -\frac{1}{2}\kappa_2 s^2 + \frac{1}{24}\kappa_4 s^4 + O(s^6) \tag{5}$$

with $\kappa_2 = E[V^2] = Var[V]$ and $\kappa_4 = E[V^4] - 3E[V^2]E[V^2] = K[V] * Var[V]^2$. This expansion is called the **(quadratic) cumulant expansion** where κ_2 and κ_4 are called the **cumulants**. An appealing property of cumulants is that they cumulate, i.e. for $W = V_1 + \ldots + V_n$, provided that the V_i are independent and identical distributed, $\kappa_2(W) = n\kappa_2(V_1)$ and $\kappa_4(W) = n\kappa_4(V_1)$. A consequence of this is that

$$K(W) = \frac{\kappa_4(W)}{\kappa_2(W)^2} = \frac{n\kappa_4(V_1)}{n^2\kappa_2(V_1)^2} = K(V_1)/n. \tag{6}$$

Hence, by summing up independent and identical distributed random variables, the kurtosis decreases by a factor $\frac{1}{n}$ whereas the variance increases by a factor n. The cumulants show a remarkable relation with the **moments** $E[V^k]$ for symmetric and centralized distributions (see appendix M),

$$\kappa_2(V) = E[V^2] \tag{7}$$

$$\kappa_4(V) = E[V^4] - 3E[V^2]E[V^2] \tag{8}$$

$$\kappa_6(V) = E[V^6] - 15E[V^4]E[V^2] + 30E[V^2]^3 \tag{9}$$

2.1.1.4 Brownian motion:

Suppose every second, a molecule takes a step of size V, where V is a random variable with expected value $E[V] = 0$ and $E[V^2] = \sigma^2$. After n seconds, the molecule has made an accumulated step of size $V_1 + \ldots + V_n$. The central limit theorem (CLT) guarantees that the distribution of the accumulated step sizes of the molecule will approximate a Gaussian distribution after sufficient steps [8]. Let V_t be the position of a particle at time t. Suppose the particle takes a random step every Δt seconds and let $\Delta t \to 0$. This kind of random motion is called **Brownian motion**, and can be modeled by a generalized **Wiener process** [9], for which we take the following properties for granted:

1) $V_0 = 0$
2) The function $t \to V_t$ is continuous for $t > 0$
3) V_t has independent increments such that $V_t - V_s \sim N(0, \sigma^2(t - s))$

From these properties it follows immediately that $Var[V_t] = \sigma^2 t$ and hence $V_t = \sqrt{t}V_1$.

It also follows that the probability density function (pdf) p of V_t at time t is the centralized Gaussian distribution

$$p(V_t = s) = \frac{1}{\sigma t\sqrt{2\pi}}e^{-\frac{s^2}{2\sigma^2 t}} \tag{10}$$

2.1.2 Physics of Gaussian diffusion

From this point, we consider 'real' physical diffusion. Let $\vec{R}(t) = (X_t, Y_t, Z_t)^\top$ be the displacement at time t due to diffusion. Let $\vec{R} = (X, Y, Z)^\top$, be the diffusional displacement in 'unit time', i.e. $\vec{R} = (\frac{X_t}{\sqrt{t}}, \frac{Y_t}{\sqrt{t}}, \frac{Z_t}{\sqrt{t}})^\top$. Notice that, by the third property of the Wiener process, $\vec{R}$ is a quantity that is time-independent.
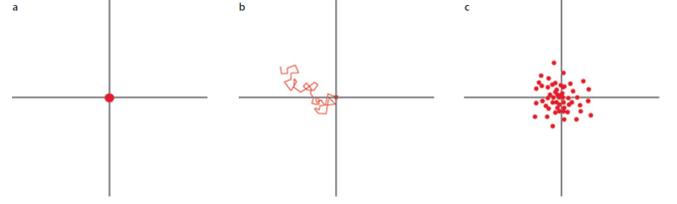
Figure 1. a. Particles start at the origin. b Each particle performs a random walk in space. c This leads to the distribution of the particles in space [11].

2.1.2.1 Isotropic Gaussian diffusion:

In the scope of this work, **water diffusion** is defined as the random motion of water molecules due to thermal energy. Water diffusion haves two basic characteristics. 1. It is Gaussian or non-Gaussian. 2. It is isotropic or anisotropic. Let us start with the simplest case of isotropic Gaussian diffusion. In this case the coordinates X_t, Y_t and Z_t of $\vec{R}(t)$ each follow a Wiener process such that $\sigma_X^2 = \sigma_Y^2 = \sigma_Z^2 = \sigma^2$ and such that X_t, Y_t, Z_t are mutually independent.

Then it is easy to see that

$$p(\vec{R}(t) = \vec{r}) = p(X_t = x)p(Y_t = y)p(Z_t = z) \tag{11}$$

where $\vec{r} = (x, y, z)^\top$. Hence the pdf of the diffusion displacement distribution $\vec{R}(t)$ follows a multivariate Gaussian distribution

$$p(\vec{R}(t) = \vec{r}) = \frac{1}{\sqrt{(4\pi)^3 t^3 |\mathbf{D}|}}e^{\frac{\vec{r}^\top \mathbf{D}^{-1} \vec{r}}{4t}} \tag{12}$$

where $|\cdot|$ is the matrix determinant, $\vec{r} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ and $\mathbf{D} = \begin{pmatrix} \frac{\sigma_X^2}{2} & 0 & 0 \\ 0 & \frac{\sigma_Y^2}{2} & 0 \\ 0 & 0 & \frac{\sigma_Z^2}{2} \end{pmatrix}$. The *matrix* $\mathbf{D}$ is called the **diffusion tensor**. Because $\sigma_X^2 = \sigma_Y^2 = \sigma_Z^2$,

$$\vec{r}^\top \mathbf{D}^{-1} \vec{r} = \frac{2}{\sigma^2}(x^2 + y^2 + z^2) = \vec{r}^\top \frac{2}{\sigma^2}\vec{r} \tag{13}$$

Hence in the case of isotropic diffusion the displacement pdf becomes

$$p(\vec{R}(t) = \vec{r}) = \frac{1}{\sqrt{(4\pi)^3 (Dt)^3}}e^{-\frac{|\vec{r}|^2}{4Dt}} \tag{14}$$

Where $D = \frac{\sigma^2}{2}$. We call D the **diffusion coefficient** or shortly the **diffusivity**. D has units of $\mu m^2/ms$. D_{free}, that is diffusion in free water, i.e. water without substances that hinder the diffusion process, is found to be around 2.5 $\mu m^2/ms$ at 37 C° [12].

2.1.2.2 Anisotropic Gaussian diffusion:

Diffusion can be anisotropic, which means that the diffusivities vary over the directions. In this case the diffusion pdf is still a multivariate Gaussian but the diffusion tensor $\mathbf{D}$ also has off-diagonal entries $\begin{pmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{pmatrix}$. The off diagonal elements are the mutual covariances of the coordinates, e.g. $D_{xy} = \frac{E[XY]}{2}$. $\mathbf{D}$ is always positive definite and symmetric [13]. Therefore $\mathbf{D}$ can be diagonalized, which means that a basis can be chosen
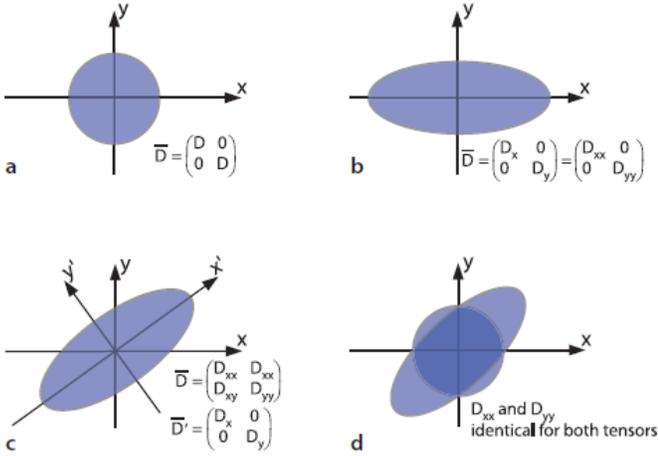
Figure 2. The diffusion tensor can be diagonalized by a coordinate change [11].

such that

$$\mathbf{D} = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix},$$

see figure 2. Suppose $\vec{v}_1$ is the first basis vector for which the diffusivity is maximal, for example when a bundle of axons (fiber) points in this direction. Then λ_1 is the diffusivity in the direction $\vec{v}_1$, whereas $\vec{v}_2$ and $\vec{v}_3$ correspond to the directions perpendicular to $\vec{v}_1$.

2.1.3 Kurtosis measured by MRI

After having reviewed the mathematical and physical basics of diffusion, we can consider how we can measure diffusion by MRI, especially diffusion kurtosis in DKI experiments.

2.1.3.1 **From diffusion to signal attenuation**: In the most elementary way the measurement of diffusion by MRI goes as follows: First a $90°$ excitation pulse is applied. Suppose a proton is at position $\vec{R} = 0$ at time $t = 0$ while a gradient vector $\vec{g}$ is applied during a very short time interval $[0, \delta]$. At time T, the proton is at another position $\vec{R}(T)$ due to diffusion. Suppose during the time interval $[T - \delta, T]$ there is an opposite gradient vector $-\vec{g}$ applied (This gradient scheme is drawn in figure 3). This yields at time T

$$\Phi(T) = -\left(\gamma \int_{t=0}^{t=\delta} \vec{g} \cdot \vec{R}(t) dt + \gamma \int_{t=T-\delta}^{t=T} -\vec{g} \cdot \vec{R}(t) dt, \right) \quad (15)$$

where γ is the gyromagnetic ratio of a proton [14]. Notice that we use a capital Φ because the phase of a particle is a random variable since it depends on the random position at time T. Now we make the 'short pulse approximation', by assuming that all the gradient strength is added during a single time point. This approximation yields $\Phi(T) \approx \gamma\delta\vec{g} \cdot (\vec{R}(0) - \vec{R}(T)) = -\gamma\delta\vec{g} \cdot \vec{R}(T)$. Let $M(t)$ be an isochromat of spins that are all in phase at time 0. After some time, the spins will get different positions due to diffusion and therefore different phases. This causes the isochromat to dephase, resulting in the attenuation of the signal because spins will partially cancel each other out, as drawn in figure 3. We know that the signal arising from an individual spin is simply proportional to the complex exponent of the phase $e^{i\phi(t)}$. Hence, the signal intensity (normalized to its 'starting value', i.e. the signal intensity without any diffusion
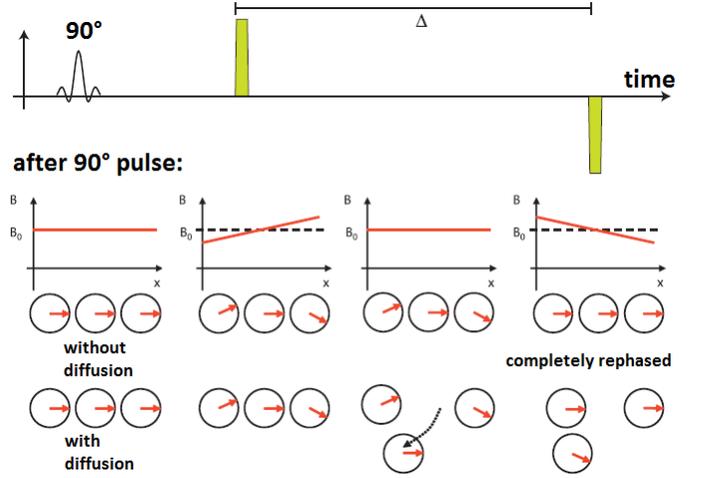
Figure 3. **Basic principle of diffusion weighting in MRI**. The red arrows represent the spins in the rotating frame of reference. The signal can in some sense be regarded as the sum of the red arrows [11].

gradients applied) of the isochromat is the average of the signal contributions of all the individual spins:

$$\frac{S_{M(t)}}{S_{M(0)}} = E[e^{i\Phi(t)}] = E[e^{-i\gamma\delta\vec{g} \cdot \vec{R}(t)}], \quad (16)$$

where $\vec{R}(t)$ is the displacement of a particle in the isochromat. The voxel signal intensity S at time t is the sum over the signal intensities of all the isochromats and is hence equal to

$$\frac{S}{S(0)} = E[e^{-i\gamma\delta\vec{g} \cdot \vec{R}(t)}], \quad (17)$$

where $\vec{R}(t)$ is in this case the displacement of a particle in the voxel.

2.1.3.2 **Gaussian diffusion measured by MRI**: Let $\vec{n}$ be the direction of $\vec{g}$, i.e. $\vec{n} = \frac{\vec{g}}{|g|}$. By equation (2), it is easy to see that for isotropic Gaussian diffusion the signal attenuates by a factor

$$E[e^{-i\gamma\delta\vec{g} \cdot \vec{R}(t_d)}] = E[e^{-i\gamma\delta|\vec{g}|(n_x X + n_y Y + n_z Z)\sqrt{t_d}}] = e^{-\gamma^2\delta^2|\vec{g}|^2 t_d D} \quad (18)$$

By calling $b = \gamma^2\delta^2|\vec{g}|^2 t_d$ the **diffusion sensitivity (b-value)**, the signal attenuation reduces to the simple expression

$$e^{-bD} \quad (19)$$

For anisotropic Gaussian attenuation the phase is equal to

$$\Phi(b) = -(n_x X + n_y Y + n_z Z)\sqrt{b} \quad (20)$$

which is a Gaussian distributed random variable such that

$$E[\Phi(b)^2] = 2b(n_x^2 D_{xx} + n_y^2 D_{yy} + n_z^2 D_{zz} + n_x n_y D_{xy} + n_x n_z D_{xz} + n_y n_z D_{yz}) \quad (21)$$

Hence

$$E[e^{i\Phi(b)}] = e^{-\vec{n}^\top \mathbf{D}\vec{n} b} \quad (22)$$

Because the diffusion displacement is weighted by the gradient strength, field inhomogeneities could make Gaussian diffusion appear non-Gaussian in a MR experiment [15].
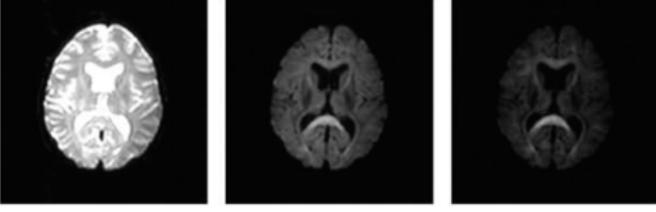
Figure 4. Signal measured at different b-values [16].

2.1.3.3 **A simple DKI experiment**: Suppose the diffusion in a voxel is isotropic, but non-Gaussian. The kurtosis value K is then clearly non-zero, and we would like to know its value. From the acquisition of 3 MRI images we can estimate the parameters S_0, D and the desired K. First we make a reference image without diffusion gradients applied to obtain S_0. Thereafter we make two images by applying two times a sequence with two different gradient strengths, i.e. one with strength b_1 and the other with strength b_2. These three acquired images are shown in figure 4. Suppose one voxel has an intensity 150 in the second image and an intensity 100 in the third image. By fitting these intensities to the quadratic cumulant expansion we obtain

$$150 = S_0 e^{-Db_1 + \frac{1}{24}\kappa_4 b_1^2}$$
$$100 = S_0 e^{-Db_2 + \frac{1}{24}\kappa_4 b_2^2}$$

Taking the log yields

$$log(150) - log(S_0) = -Db_1 + \frac{\kappa_4}{24}b_1^2$$
$$log(100) - log(S_0) = -Db_2 + \frac{\kappa_4}{24}b_2^2$$

These equations can be solved for D and $K = \frac{\kappa_4}{\kappa_2^2} = \frac{\kappa_4}{4D^2}$. If diffusion is anisotropic, or the gradients are not very short applied, by solving the equations we still get a value for D and K. Because a gradient is of course never infinitely short applied and we don't know beforehand if diffusion is isotropic, it is more correct to speak about the *apparent* diffusivity and kurtosis D_{app} and K_{app}. In real experiments, multiple images are acquired and more sophisticated parameter estimation methods and gradient schemes are used.

2.1.3.4 **Anisotropic non-Gaussian diffusion measured by MRI**: When diffusion is anisotropic one can naturally consider D_{app} and K_{app} in all directions. Like D_{app} becomes a second order tensor $\boldsymbol{D}^{app}$ by going from isotropic to anisotropic diffusion, we will see that $\boldsymbol{K}^{app}$ becomes a fourth order tensor. The random displacement of water molecules as probed by MRI is a symmetric random variable, i.e. $E[X^{2k+1}] = 0$ [13]. Therefore, when a gradient in the $\vec{n}$ direction is applied, the log of the anisotropic signal attenuation yields by equation (5)

$$log\left(\frac{S_b(\vec{n})}{S(0)}\right) = -\frac{1}{2}\kappa_2(n_xX + n_yY + n_zZ)b$$
$$+ \frac{1}{24}\kappa_4(n_xX + n_yY + n_zZ)b^2 + O(b^3)$$

where by equation (7)

$$\frac{1}{2}\kappa_2(n_xX + n_yY + n_zZ) = \frac{1}{2}E[(n_xX + n_yY + n_zZ)^2]$$

and by equation (8)

$$\kappa_4(n_xX + n_yY + n_zZ) = E[(n_xX + n_yY + n_zZ)^4]$$
$$-3E[(n_xX + n_yY + n_zZ)^2]^2$$

with

$$E[(n_xX + n_yY + n_zZ)^4] =$$
$$n_x^4E[X^4] + n_y^4E[Y^4] + n_z^4E[Z^4]$$
$$+ 4(n_x^3n_yE[X^3Y] + n_x^3n_zE[X^3Z] + n_y^3n_xE[Y^3X]$$
$$+ n_y^3n_zE[Y^3Z] + n_z^3n_xE[Z^3X] + n_z^3n_yE[Z^3Y])$$
$$+ 6(n_x^2n_y^2E[X^2Y^2] + n_x^2n_z^2E[X^2Z^2] + n_y^2n_z^2E[Y^2Z^2])$$
$$+ 12(n_x^2n_yn_zE[X^2YZ] + n_y^2n_xn_zE[Y^2XZ] + n_z^2n_xn_yE[Z^2XY])$$

By using the notation

$$W_{ijkl} = 9 * \frac{E[R_iR_jR_kR_l] - E[R_iR_j]E[R_kR_l]}{E[(\vec{R}\cdot\vec{R})^2]}$$
$$+ 9 * \frac{-E[R_iR_k]E[R_jR_l] - E[R_iR_l]E[R_jR_k]}{E[(\vec{R}\cdot\vec{R})^2]},$$

where $R_1 = X$, $R_2 = Y$ and $R_3 = Z$, a more convenient expression can be obtained [2]

$$log[\frac{S_b(\vec{n})}{S_0}] = -b\sum_{i=1}^{3}\sum_{j=1}^{3}n_in_jD_{ij}$$
$$+ \frac{1}{6}b^2\left(\frac{1}{3}\sum_{i=1}^{3}D_{ii}^{app}\right)^2\sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3}\sum_{l=1}^{3}n_in_jn_kn_lW_{ijkl}$$
$$+ O(b^3)$$

We will call $\boldsymbol{W}$ the **kurtosis tensor**. $\boldsymbol{W}$ has $3^4 = 81$ components. However, because of symmetry, only 15 components are independent. We then can solve for the 22 independent parameters S_0, $(D_{xx}\ D_{yy}\ D_{zz}\ D_{xy}\ D_{xz}\ D_{yz})$ and $(W_{xxxx}\ W_{yyyy}$ $W_{zzzz}\ W_{xxyy}\ W_{xxzz}\ W_{yyzz}\ W_{xxyz}\ W_{yyxz}\ W_{zzxy}\ W_{xxxy}\ W_{xxxz}$ $W_{yyyz}\ W_{yyyx}\ W_{zzzx}\ W_{zzzy})$. With these parameters in hand, the diffusion and kurtosis can be estimated in every direction.

2.1.3.5 **Diffusion metrics**: There are several metrics of interest to derive from the diffusion and kurtosis tensor that are summarized in table 1. The mean diffusivity (MD) and mean kurtosis (MK) say something about the overall diffusion (notice that for isotropic diffusion, $MK \approx K_{app}$). The radial and axial diffusivity and kurtosis (RD, AD, RK and AK) are especially useful to assess the diffusion perpendicular and parallel to a fiber. The diffusion and kurtosis fractional anisotropy (FA resp. KA) measures the amount of anisotropy of the diffusion distribution in a voxel. Since the kurtosis tensor contains more information about the diffusion process than the diffusion tensor alone it should provide extra information about tissue microstructure. Empirically, however, a strong correlation has been found between the radial kurtosis and the radial diffusion and also a correlation between FA and MK (see figure 5).

2.1.3.6 **Eigenvalues of the kurtosis tensor**: Since a kurtosis tensor consists of 15 independent parameters, there are more possibilities to extract relevant metrics out of the kurtosis tensor. For example, 'eigenvalues' of the fourth order kurtosis tensor can be found, in a manner analogues to the eigenvalues of a diffusion tensor [18].

Directional Diffusion [5]	$D(\vec{n}) = \vec{n}^\top \mathbf{D}^{-1} \vec{n}$
Mean Diffusivity [12]	$MD = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3}$
Diffusivity in main direction [12]	$D_\parallel = \lambda_1$
Radial Diffusion [12]	$D_\perp = \frac{\lambda_2 + \lambda_3}{2}$
Diffusion Fractional Anisotropy [12]	$FA = \sqrt{\frac{3}{2} \frac{(\lambda_1 - MD)^2 + (\lambda_2 - MD)^2 + (\lambda_3 - MD)^2}{(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)}}$
Directional Kurtosis [5]	$K(\vec{n}) = \frac{MD^2}{D(\vec{n})^2} \sum n_i n_j n_k n_l W_{ijkl}$
Parallel Kurtosis [5]	$K_\perp = K(\vec{v}_1)$
Radial Kurtosis [5]	$K_\perp = \frac{1}{2\pi} \int K(\vec{n}) \delta(\vec{n} \cdot \vec{v_2}) d\Omega_n$
Mean Kurtosis [5]	$MK = \frac{1}{4\pi} \int_{S^2} K(\vec{n}) d\vec{n}$
Kurtosis Fractional Anisotropy [19]	$AK = \sqrt{\frac{3}{2} \frac{(K_1 - MK)^2 + (K_2 - MK)^2 + (K_3 - MK)^2}{K_1^2 + K_2^2 + K_3^2}}$

Table 1
DTI and DKI metrics . Other authors may use (slightly) different definitions. $\delta(x)$ is the Dirac delta function.
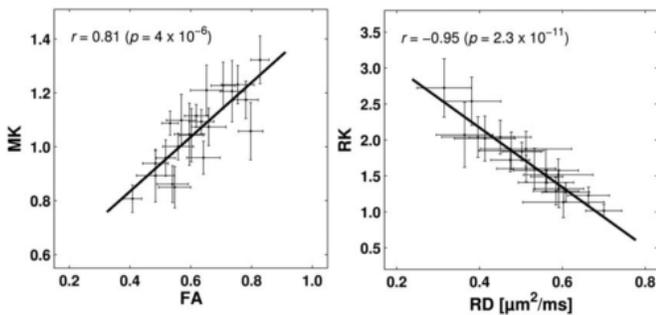
Figure 5. Correlation between diffusion metrics in healthy white matter, means and standard deviations [17].
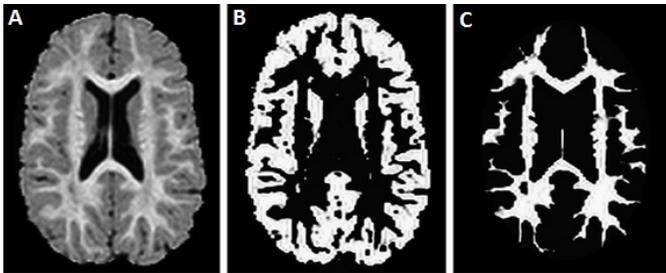
Figure 6. Segmentation of (B) GM, and (C) WM [20].

2.2 Diffusion in the human brain

2.2.1 Anatomy

The brain can be divided into gray matter, white matter and CSF, as shown in figure 6. A drawing of various cell components that appear in both WM and GM is given in figure 7.
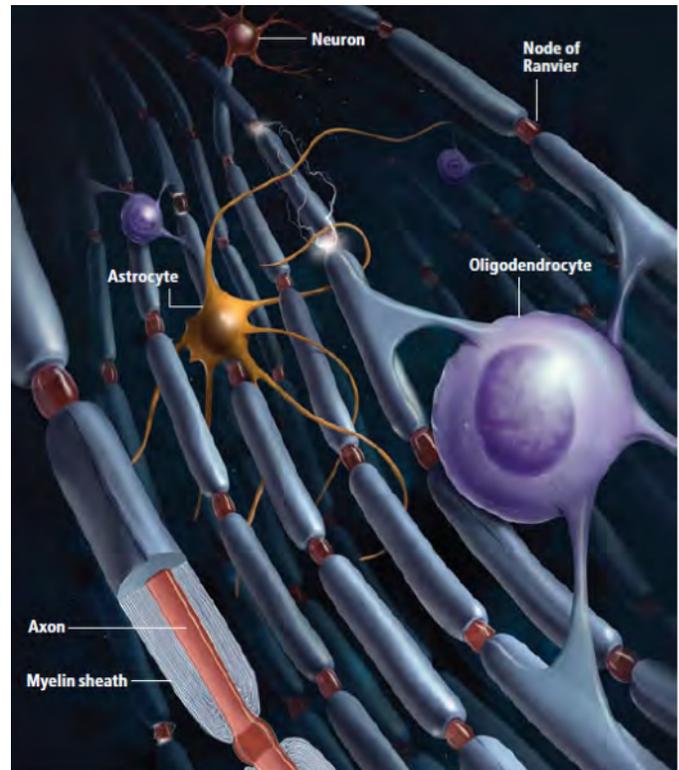
Figure 7. Drawing of the cell components in neural tissue [21].
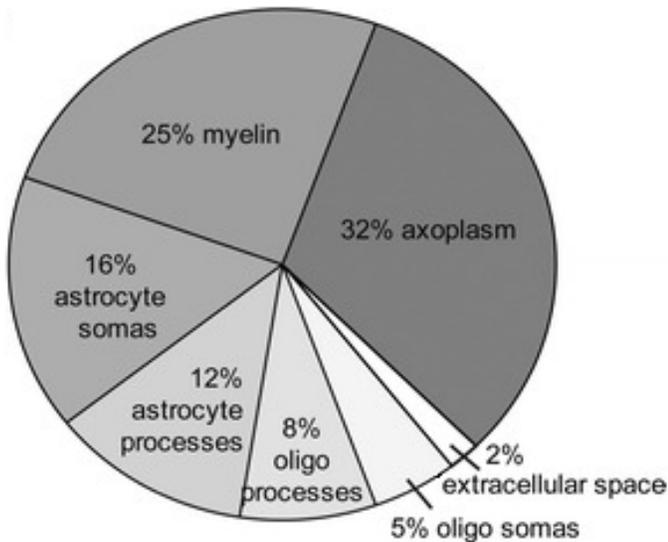
Figure 8. Allocation of space in the optic nerve [22]. These percentages are not always consistent among different published reports [15].
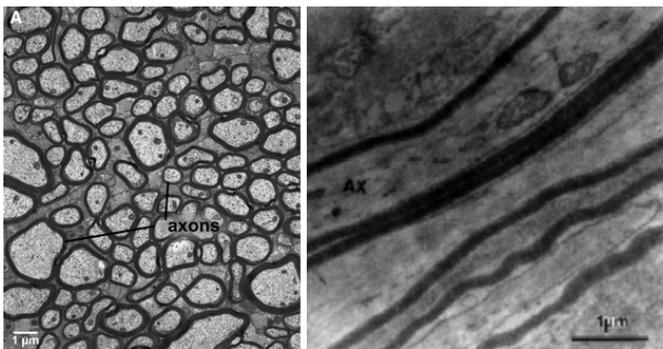
Figure 9. Myelinated axons in the optic nerve (electron micrograph) a. transverse view b. longitudinal view [23].
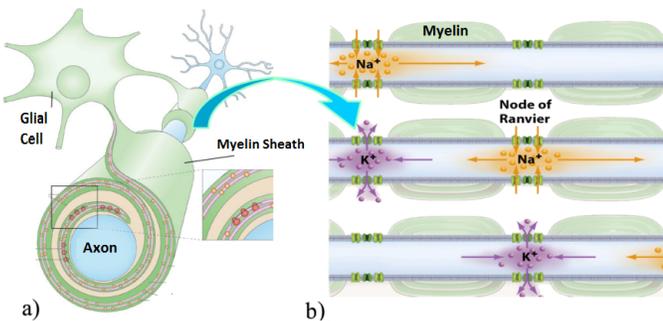
Figure 10. a) Myelinated axon diagram [24]. b) Representation of the axon showing the nodes of Ranvier [25].

2.2.1.1 **White matter**: The white matter of the central nervous system contains axons, their myelin sheats and glial cells. The relative fractions of each type of microtissue are illustrated in figure 8. The axons carry the chemical and electrical signals to and from neurons concentrated in the gray matter. They are grouped into bundles forming tracts. Most axons are arranged in parallel bundles, such as in the optic nerve. As can be seen in figure 9, axons often have an insulating sheath called myelin to increase the speed at which impulses propagate along the axon. The myelin sheath does not cover the entire axon; it leaves small exposed sections uncovered which are called nodes of Ranvier (see figure 10). As shown in figure 11, the extra-axonal space consists of glial cells. The glial cells are named after their structure: oligodendrocytes because they have few branches, astrocytes due to their star shape and microglia because of their small size. The primary function of oligodendrocytes is to produce myelin. Both oligodendrocytes and astrocytes provide physical support to the axons. Also capillary bloodvessels are present in WM. The mean fraction v of myelinated axons in a voxel is found to be around 0.33 and is higher in denser regions such as the corpus callosum. Fractions up to $v = 0.7$ have been reported for the ex vivo commisura anterior of rat brain [26]. The mean radius of an axon is found to be around $5\ \mu m$ [27]. The diffusivity in the extra-axonal space is found to be only about 15% less than D_{free} and much higher in the directions parallel to the axons than perpendicular to the axons [26].

2.2.1.2 **Gray matter**: The gray matter of the central nervous system contains neuronal cell bodies, dendrites and axons. These dendrites and axons are, unlike white matter, mostly not myelinated and are oriented in a three structure around the cell body (see figure 12). Gray matter also contains capillary blood vessels and glial cells.

2.2.1.3 **Diffusion metrics measured in the brain**: In table 2 some values of diffusion kurtosis metrics are given. Everywhere in the GM and the WM a positive mean kurtosis between 0.6 and 1.32 is measured. In WM a higher amount of isotropy is measured than in GM. In general, also a higher mean, radial and axial kurtosis is measured in WM. In figure 13 a kurtosis map of the human brain is given, which clearly shows the contrast between GM and WM. There can be a large difference between in vivo and ex vivo measurements [19].

2.2.2 Potential clinical applications of DKI

After considering the theoretical foundations of DKI the reader may wonder about the clinical importance of this technique. In this section, a brief overview of potential applications of DKI will be given.

2.2.2.1 **Aging**: Diffusion kurtosis in the prefrontal lobe is found to be age dependent (n=24) [30]. The GM MK shows an increase with age. The WM MK showed an increase until age 18, with a shift to lower values with aging. The WM FA appears to be highly correlated with the WM MK in this study.

2.2.2.2 **Attention-Deficit Hyperactivity Disorder (ADHD)**: A study on prefrontal WM (n=17) [31] suggested that the process of changing mean kurtosis with age is slower for children with ADHD. This study does not mention if the FA shows a similar pattern.

2.2.2.3 **Multiple sclerosis (MS)**: A significant lower mean kurtosis has been observed in MS patients (n=11). Also a less significant lower FA value has been found [32].

2.2.2.4 **Schizophrenia**: A WM decrease in MK and FA is observed in schizophrenia patients (n=18) [33]. The decrease

ROI	MK	RK	MD	RD	FA
External Capsule	0.81 ± 0.05	1.02 ± 0.09	0.9 ± 0.05	0.70 ± 0.04	0.41 ± 0.03
ALIC	1.04 ± 0.10	1.6 ± 0.28	0.87 ± 0.05	0.53 ± 0.05	0.60 ± 0.04
PLIC	1.23 ± 0.09	2.04 ± 0.23	0.89 ± 0.09	0.45 ± 0.07	0.71 ± 0.04
CC, body	1.17 ± 0.07	2.07 ± 0.41	0.92 ± 0.07	0.38 ± 0.07	0.78 ± 0.04
CC, genu	1.06 ± 0.11	2.07 ± 0.45	0.93 ± 0.06	0.36 ± 0.07	0.80 ± 0.04
CC, splenium	1.32 ± 0.09	2.72 ± 0.41	0.89 ± 0.09	0.31 ± 0.07	0.83 ± 0.03
Centrum semiovale	1.09 ± 0.04	1.72 ± 0.16	0.80 ± 0.04	0.47 ± 0.05	0.63 ± 0.04
Cingulate, body	1.07 ± 0.07	1.85 ± 0.26	0.86 ± 0.07	0.48 ± 0.08	0.66 ± 0.06
Cingulate, temporal	0.85 ± 0.08	1.13 ± 0.21	0.92 ± 0.12	0.60 ± 0.10	0.55 ± 0.05
Corona Radiata	1.09 ± 0.04	1.49 ± 0.09	0.84 ± 0.05	0.56 ± 0.04	0.53 ± 0.03
CST, cerebral crus	1.23 ± 0.07	2.04 ± 0.28	0.88 ± 0.08	0.40 ± 0.09	0.75 ± 0.05
IFO, anterior basal	0.86 ± 0.07	1.29 ± 0.19	0.89 ± 0.05	0.58 ± 0.05	0.54 ± 0.05
ILF, posterior	0.96 ± 0.06	1.60 ± 0.18	0.90 ± 0.06	0.51 ± 0.07	0.64 ± 0.05
SLF, posteroir	1.11 ± 0.04	1.84 ± 0.13	0.83 ± 0.04	0.50 ± 0.05	0.62 ± 0.04
Frontal sWM	0.94 ± 0.05	1.23 ± 0.12	0.91 ± 0.05	0.66 ± 0.05	0.48 ± 0.04
Parietal sWM	1.00 ± 0.05	1.41 ± 0.12	0.86 ± 0.06	0.56 ± 0.07	0.56 ± 0.05
Temporal sWM	0.96 ± 0.07	1.27 ± 0.13	0.88 ± 0.08	0.61 ± 0.06	0.52 ± 0.03
Caudate head	0.61 ± 0.08	0.59 ± 0.07	0.87 ± 0.05	0.80 ± 0.04	0.14 ± 0.03
Globus pallidus	1.06 ± 0.08	1.05 ± 0.10	0.86 ± 0.08	0.74 ± 0.06	0.27 ± 0.04
Putamen	0.67 ± 0.08	0.61 ± 0.08	0.79 ± 0.03	0.73 ± 0.03	0.15 ± 0.02
Thalamus	0.86 ± 0.07	0.92 ± 0.09	0.87 ± 0.10	0.73 ± 0.09	0.32 ± 0.03

Table 2

Regional values of diffusion metrics in healthy volunteers (n=36), mean and standard deviation. $\Delta = 37.4$ ms [17]. ALIC Anterior limb of the internal capsule PLIC Posterior limb of the internal capsule CC Corpus callosum sWM Superficial white matter CST Corticospinal tract IFO Inferior fronto-occipital fasciculus ILF Inferior longitudinal fasciculus SLF Superior longitudinal fasciculus
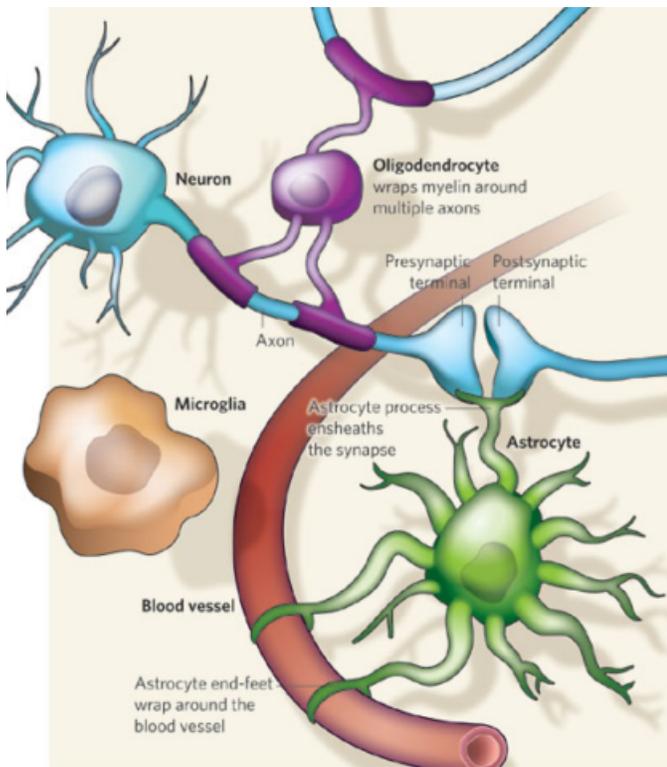
Figure 11. Schematic representation of different types of glial cells [28].

Figure 12. A neuron with the neurites uniform oriented in a three structure around the cell body [29].
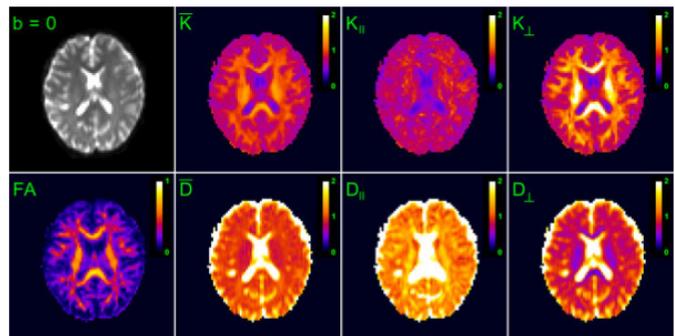
Figure 13. Map of diffusivity and kurtosis metrics in a human brain [5].

in MK was much more significant than the decrease in FA.

2.2.2.5 **Alzheimer**: A frontal WM MK decrease has been found in Alzheimer patients, accompanied by a MD increase and a FA decrease in one study (n=8) [34]. Another study [35] found that DKI may be more sensitive to Alzheimer than DTI.

2.2.2.6 **Mild cognitive impairment (MCI)**: The mean kurtosis and radial kurtosis in the anterior corona are found to be the best discriminators between MCI patients and healthy controls ($n = 13$) [36].

2.2.2.7 **Parkinson**: In a study on Parkinson's disease, the mean kurtosis is found to be higher in the basal ganglia. MK has a significant higher diagnostic accuracy than FA [37].

2.2.2.8 **Cerebral infarct**: Kurtosis metrics show significantly higher percent changes than complementary diffusion tensor metrics during a cerebral infarct (n=44) [38].

2.2.2.9 **Epilepsy**: A relation between MK, FA and epilepsy has been found in some parts of the brain [39].

2.2.2.10 **Grading of brain tumors**: Higher graded brain tumors are found (n=28) to increase the kurtosis metrics MK, RK, and KA whereas the DTI metrics MD and FA do not change significantly [40].

2.2.2.11 **Hydrocephalus**: In hydrocephalus patients, the DTI metrics FA, MD, λ_1 are found to be significantly higher whereas the DKI metrics MK and $K_\parallel$ are found to be significantly lower (n=17) [41].

2.2.2.12 **Huntington disease**: In a rat study on Huntington disease, the kurtosis anisotropy has been found to be the only metric that changes significant [42].

To obtain significant results in a study, a minimal group size is needed. By acknowledging that the minimal group size needed in DKI studies is usually around 30 but can be as large as 199 [43], it is clear that most studies may not comply to this minimal size. However, all these studies together demonstrate that DKI metrics could be used as a biomarker to detect pathologies. But when DTI metrics can detect the same pathologies just as good, DTI is preferred since it needs a lower group size and shorter echo times. However, DKI may still be preferable to DTI in tissue where the DTI model is invalid, for example, in regions with complex fiber organization [43].

2.3 The central limit theorem

Before considering any paper from the literature about diffusion kurtosis, we can already say a lot about it by only using the central limit theorem. This theorem says that under some conditions every infinite sum of random values is a Gaussian random variable.

Theorem. *Let $X_1, \ldots, X_n$ be independent and identical distributed random variables*[3] *such that $E[X] = 0$ and $E[X_i^2] = \sigma^2 < \infty$. Then $\sum_{i=1}^n X_i / \sqrt{n}$ converges to a Gaussian distribution with mean 0 and standard deviation σ. [8]*

So when diffusion is *not* Gaussian, clearly at least *one* of these conditions could be violated. Let us take X_i as the diffusion displacement in an arbitrary time interval [i,i+1]. Where can the central limit theorem fail?

2.3.1.1 1. The X_i are not independent: Over a short period of time X_0 and X_δ are correlated (hence not independent), because the diffusion displacement obviously depends on the starting position. In the long term, however, X_i might not be dependent on the starting position anymore. The time it takes that X_i becomes uncorrelated with X_0 will be called the **correlation time** and will be only a few milliseconds because of the micrometer-scale structure that particles encounter. Hindered diffusion is non-Gaussian, but becomes Gaussian when the diffusion time is larger than the correlation time [2]. If diffusion

---

3. The identical distribution assumption can be weakened, which is beyond the scope of this work.
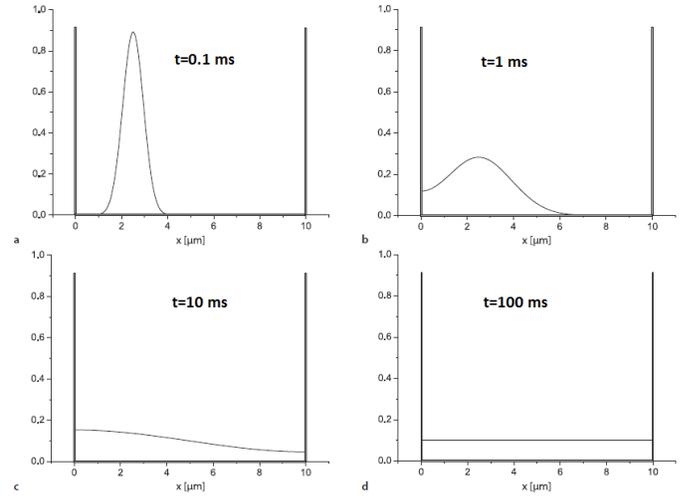
Figure 14. Diffusion distribution of a diffusing particle starting at $x = 2.5 \ \mu m$ for diffusion times ranging from 0.1 ms to 100 ms. The motion is restricted by two plates at $x = 0 \ \mu m$ and at $x = 10 \ \mu m$ [11].

is totally restricted by a non-permeable membrane, then after sufficient time the diffusion distribution becomes non-Gaussian but uniform (see figure 14), which results in a negative kurtosis of around -0.5 [2].

2.3.1.2 2. $E[X_i^2]$ is infinite: This might appear as an odd violation of the CLT because in real life diffusing particles in the brain should always have a begin and endpoint somewhere in the brain. Some mathematical models may, however, (implicitly) assume that $E[X_i^2]$ is infinite, such as the stretched exponential model, in which we will go into detail in section 3.4.

To summarize, the central limit theorem teaches us that non-Gaussian diffusion due to hindrance may become Gaussian after sufficient time. However, non-Gaussian diffusion due to restriction or an infinite $E[X_i^2]$, remains non-Gaussian with increasing diffusion time.

2.4 Propagation of the compartmental kurtosis in the voxel kurtosis

Suppose there are several compartments in a voxel each having relative fractions p_i, diffusivity D_i and apparent kurtosis K_i. The apparent kurtosis of the voxel in this case is

$$K_{app} = \frac{3 * \sum_{i=1}^n p_i (D_i - \bar{D})^2 + \sum_{i=1}^n p_i D_i^2 K_i}{\bar{D}^2} \quad (23)$$

Hence, if a compartment has a low diffusivity, the kurtosis of that compartment barely propagates in the overall kurtosis of the voxel. This can be seen easily by considering that the tails of a distribution with a low standard deviation are generally the shoulders of a distribution with a high standard deviation. This formula turns out to be very helpful, therefore it will be given a name, i.e. the **partial volume kurtosis** formula. I did not find this formula in the diffusion imaging literature, therefore I will derive it in appendix F.

3 LITERATURE REVIEW ON NON-GAUSSIAN DIFFUSION MODELS AND THEIR RELATION TO KURTOSIS METRICS.

This section is a critical review of some popular human brain diffusion models and their relations to kurtosis metrics. There

are numerous diffusion models available in the literature of varying complexity; one can model diffusion by a simple biexponential, but also by advanced graduate mathematics, see [44] for example. Since deriving the kurtosis by a cumulant expansion of a simple model already can be a tedious exercise and arguably one first should aim at the most simple explanations of kurtosis, this work will generally restrict itself to the more simple but popular diffusion models. For every reviewed model in this work, we will state the general ideas and assumptions, possible validation results from experiments and simulations, and possible relations with kurtosis. At the end of every model or class of models a discussion follows.

3.0.1 Model assessment criteria

To perform a critical review on diffusion models, one would want to have some general criteria for the quality of a diffusion model. The BIC and AIC are popular measures of how well a model explains the experimental data by quantifying the trade-off between goodness of fit and complexity. A lower value should indicate that a model is more predictive [27]. It is nice to have a model that haves a low BIC or AIC, but one also wants a model that explains the measured signal in more than a statistical way. Therefore in diffusion modeling arguably the current emphasis lays on microstructural models which for example model the signal as being dependent on axon diameter or membrane permeability. However, a disadvantage of assessing a model by how it can estimate anatomical parameters, is that a ground truth from histology is not always available, since histology has it artifacts such as shrinkage, tissue disintegration and non-uniform staining [45]. In model validation experiments, the use of higher b-values gives the validation more (statistical) 'power' [15].

3.1 Biexponential model

3.1.1.1 **Idea and assumptions**: The biexponential model is a 'classic' diffusion model, proposed by Niendorf (1996) [10]. It consists of two compartments where diffusion is Gaussian but with a different diffusion coefficient, i.e. a fast and a slow compartment D_{fast} and D_{slow}, which each obey a fraction of p_{fast} resp. p_{slow} of the voxel.

$$\frac{S(b)}{S(0)} = p_{slow}e^{-bD_{slow}} + (1 - p_{slow})e^{-bD_{fast}}. \qquad (24)$$

These compartments were originally interpreted as the intra- and extracellular space .

3.1.1.2 **Experimental and simulation results**: The fractions of intra- and extracellular space implied by the biexponential model do not match with the physiological values in the literature [46]. Biexponential diffusion is observed within the intracellular space of a *single* cell [47].

3.1.1.3 **Relation to kurtosis**: For the apparent kurtosis of the biexponential model, we have [26]

$$K_{app} = 3 * \frac{p_{slow} * (1 - p_{slow}) * (D_{slow} - D_{fast})^2}{D_{app}^2} \qquad (25)$$

3.1.1.4 **Discussion**: In [48], it is suggested that the two compartments correspond to structured water around the membrane and free water in the cytoplasm, which allows biexponential signal decay also in single cells. According to [22], the compartmental mismatch can be due to the fact that some intracellular water, such as myelin water, is MR-invisible at clinical echo times. To date however, none of the explanations

of the biexponential model has gained acceptance [49]. Some authors suggest therefore to accept that the biexponetial model is "largely insufficient" [46]. However, WM can be modeled in the radial direction by a biexponential model, which is for example stated in Fieremans (2011) [26]. In this case, the two compartments correspond to the intra- and extra-axonal space (IAS and EAS). I derived in appendix G that the kurtosis is maximal whenever

$$v = \frac{D_{EAS}}{D_{IAS} + D_{EAS}}, \qquad (26)$$

where v is the axonal fraction, i.e. the fraction of water that belongs to the IAS. Because axon diffusion is highly restricted by the axon membrane in the radial direction, we have that $D_{IAS} << D_{EAS}$, and hence this formula reveals that generally a high axonal fraction would cause a high radial kurtosis.

3.2 The Kärger model

3.2.1.1 **Idea**: The (simple) Kärger model originates from chemical physics and is investigated in the context of diffusion MRI for example by Fieremans (2010) [50]. The idea of the Kärger model is that compartments exchange particles with each other, at a rate linearly proportional to the relative distribution of the spins over the compartments. This introduces the parameter τ, which describes the mean time that a particle resides in one compartment. For myelinated axons, the in vivo estimated τ ranges from 300 ms to 2500 ms [22]. For non-myelinated membranes, data about τ is scarce, but for erythrocytes it is found to be between 6 and 17 ms [26]. The mean exchange time in glial cells is estimated to be around 50 ms [51].

3.2.1.2 **Assumptions**: The following assumptions are made in the simple Kärger model: 1. The voxel consists of two compartments, with diffusivities D_1 and D_2, such that $D_1 > D_2$. 2. The membrane permeability is assumed to be low, otherwise the 'fast' component would vanish leading to Gaussian diffusion. 3. The diffusion time is longer than the correlation time. 4. Diffusion is isotropic Gaussian in a single compartment. 5. Every particle has an equal chance to pass the membrane, independent of the location.

3.2.1.3 **Analytical properties**: Suppose at $t = 0$ there are two compartments with relative fractions p_1 and p_2 and diffusion coefficients D_1 and D_2. This yields the following system of ordinary differential equations for the signal intensity:

$$\frac{\delta S_1}{\delta t} = -q^2 D_1 - \frac{S_1}{\tau_1} + \frac{S_2}{\tau_2}$$
$$\frac{\delta S_2}{\delta t} = -q^2 D_2 - \frac{S_2}{\tau_2} + \frac{S_1}{\tau_1}$$

with initial value $\frac{S_1(0)}{S_2(0)} = \frac{p_1}{p_2}$ and $\frac{p_1}{\tau_1} = \frac{p_2}{\tau_2}$. Here $q^2 = \frac{b}{t}$. Solving this equation gives the following expression for the voxel signal

$$\frac{S(t)}{S(0)} = p_1' e^{-D_1' q^2 t} + p_2' e^{-D_2' q^2 t} \qquad (27)$$

where the diffusion coefficients are

$$D_{1,2}' = \frac{1}{2}\left[D_1 + D_2 + \frac{1}{q^2}\left(\frac{1}{\tau_1} + \frac{1}{\tau_2}\right)\right]$$
$$\pm \sqrt{\left(D_2 - D_1 + \frac{1}{q^2}\left(\frac{1}{\tau_2} - \frac{1}{\tau_1}\right)\right)^2 + \frac{4}{q^4 \tau_1 \tau_2}}$$

with $p_1' = 1 - p_2'$ and $p_2' = \frac{p_1 D_1 + p_2 D_2 - D_1'}{D_2' - D_1'}$. From these two equations, it is straightforward to show that in this model

exchange does not change the apparent diffusivity [50]

$$\bar{D} = p_1 D_1 + p_2 D_2. \tag{28}$$

3.2.1.4 **Relation to kurtosis**: At time $t = 0$, the kurtosis is the same as in the non-exchange case [50]

$$K_{app} = 3 \frac{Var[D]}{\bar{D}^2}$$

but when $t > 0$ [50] ,

$$K_{app} = 3 \frac{Var[D]}{\bar{D}^2} \frac{2}{\bar{t}} \left(1 - \frac{1}{\bar{t}}(1 - e^{-\bar{t}}) \right) \tag{29}$$

where $\tau_{ex} = p_1 \tau_2 = p_2 \tau_1$ and $\bar{t} = \frac{t}{\tau_{ex}}$. Hence, K_{app} is strictly decreasing in time. By filling in the numbers with a realistic exchange time of $\tau_{ex} = 20 \ ms$, at $t_d = 30 \ ms$ the apparent kurtosis will have been decreased to 0.64 times its initial value. At $t_d = 10 \ ms$, the apparent kurtosis will have been decreased to 0.46 times its initial value.

3.2.1.5 **Extension of the Kärger model**: The simple Kärger model is extended by Meier (2003) [52] to model exchange in WM between a restricted intra-axonal compartment with zero radius and a Gaussian extra-axonal space.

3.2.1.6 **Results from experiments**: Experiments give ambiguous outcomes on the dependence of the water signal on diffusion time. In in vivo rat brain gray matter, the kurtosis is found to be strictly decreasing after a certain time, as predicted by the Kärger model [46] (see figure 15). In contrast, in [53] a positive correlation between mean kurtosis and time in gray matter is found. There are, however, at least 4 other experiments known where the in vivo signal-vs.-b curve is independent of the diffusion time [22].

3.2.1.7 **Results from simulations**: Fieremans (2010) [50] performed a simulation on a tissue model consisting of a set of parallel cylinders. When the diffusion time t_d is longer than the correlation time of the restricted diffusion in the cylinders, the kurtosis matched well with the Kärger model for a range of permeabilities that apply to most 'leaky' cells, except myelinated axons. The lowest permeability for which a simulation was performed had an exchange time of $\tau = 20 \ ms$, considerably shorter then the range of supposed exchange times in myelinated axons.

3.2.1.8 **Discussion**: The Kärger model is interesting since it predicts Gaussianity after sufficient time from voxels consisting of multiple compartments if they allow exchange between each other. This justifies models which for example assume that the extra-axonal space consists of just one Gaussian compartment. The idea behind the formulas of the Kärger model is that due to exchange, less diffusion heterogeneity arises between compartments, i.e. the fast compartment becomes effectively more slow and the slow compartment becomes more fast. A rule of thumb from the Kärger model is therefore that a higher permeability results in a lower kurtosis. This could arguably explain why GM has a lower MK than WM, since GM tissue is generally less myelinated and hence more permeable.
Fieremans (2010) [50] should have considered higher axon exchange times than 20 ms, since these exchange times may be as long as 2500 ms in axons [22].

3.3 Statistical model

Yablonskiy (2003) [54] proposes to model the intravoxel signal intensity caused by diffusion as a continuous spectra of isotropic 'compartments', because spins at different positions
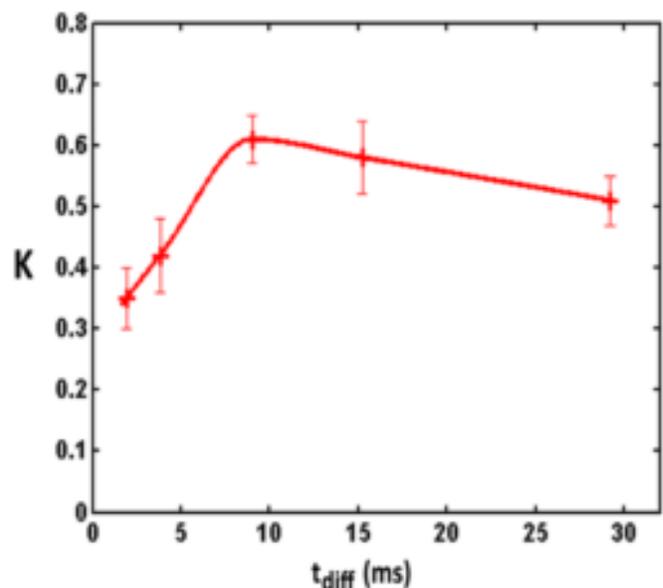
Figure 15. Variation of the apparent kurtosis over diffusion time in rat brain cortex[46].

encounter different environments. This motivates the following model

$$\frac{S(b)}{S(0)} = \int P(D) e^{-bD} dD \tag{30}$$

where $P(D)$ can be any probability density function. This model is called the **statistical model**. Yablonskiy (2003) [54] proposes to use for $P(D)$ a Gaussian that is truncated at zero to not allow a negative diffusivity

$$P(D) = \begin{cases} A * e^{-\frac{(D - ADC)^2}{2\sigma^2}} & \text{if } D \geq 0 \\ 0 & \text{if } D < 0, \end{cases} \tag{31}$$

where A is a normalization constant.
Other distributions for $P(D)$ that are suggested in the literature are the gamma [5], log-normal [55], and beta [15] distribution. The gamma and log-normal statistical models are improvements over the truncated-Gaussian, since both only allow positive values of D, and allow more variation in shape of $P(D)$, in particular in skewness. The beta model is an improvement over the gamma model, since the beta statistical model bounds the diffusion coefficient by D_{free}.

3.3.1.1 **Results from experiments**: One study considering b-values up to $5000 \ s/mm^2$ showed that a truncated-Gaussian statistical model does not give a bad fit but performs significantly worse than models with comparable amounts of parameters [16].

3.3.1.2 **Anisotropic statistical models**: The 'statistical model' can be made anisotropic by not imposing a distribution of diffusivities $P(D)$, but a distribution of diffusion tensors $P(\mathbf{D})$

$$\frac{S(\vec{n}, b)}{S(0)} = \int P(\mathbf{D}) e^{-b\vec{n}^\top \mathbf{D} \vec{n}} d\mathbf{D}. \tag{32}$$

Jian (2007) [56] chooses a Wishart probability distribution for $P(\mathbf{D})$. One could roughly say that the Wishart distribution is the multivariate equivalent of the chi-square distribution.
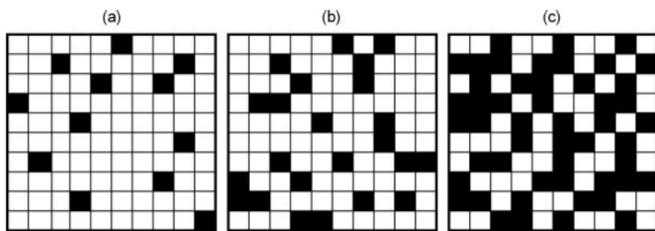
Figure 16. Images of increasing complexity. Fractal dimensions are (a) 0.31, (b) 0.477 and (c) 1.041 [58].

Basser (2003) [57] proposes a 'Gaussian' tensor distribution

$$P(\boldsymbol{D}) = C * e^{-\frac{1}{2} \sum D_{ij} W_{ijkl} D_{kl}}, \qquad (33)$$

where C is a constant and $\boldsymbol{W}$ a fourth order tensor.

3.3.1.3 **Relation to kurtosis**: When D is distributed as a truncated-Gaussian distribution, the kurtosis equals [15]

$$K_{app} = \frac{3}{D_{app}^2}(\sigma^2 - D_{app}^2 + D_{app} * ADC), \qquad (34)$$

where ADC and σ^2 are the mean and variance of $P(D)$. The model of Jian (2007) should be an anisotropic extension of the 'statistical' models. However, Jensen (2010) [5] showed that the kurtosis implied by this model is isotropic.

3.3.1.4 **Discussion**: The statistical model assumes that the individual compartments follow a Gaussian distribution. This might even be a reasonable assumption for non-Gaussian models with a relatively low diffusivity, according to the partial volume kurtosis formula (equation 23).

A disadvantage is that the statistical models lack a biophysical basis. However, as we will see in later sections, parts of other 'physical' models are in fact statistical models, such as a model with a random gamma distributed cylinder radius in section 3.5.3.2 and an axonal dispersion model in section 3.5.2.3.

The Wishart distribution of Jian (2007) [56] may be a good distribution to model anisotropic diffusion that is non-Gaussian, but for example has a small isotropic apparent kurtosis, like CSF [53].

Basser (2003) [57] gives an interesting interpretation of the kurtosis tensor; the kurtosis tensor can in this framework be seen as a measure of how likely it is to observe a certain diffusion *tensor* in a voxel, equivalent to how a diffusion tensor can be seen as a measure of the likeliness of observing a certain diffusion displacement *vector* in a voxel.

3.4 Stretched exponential

3.4.1.1 **Idea**: Hall (2008) [58] proposes the following GM model called the stretched exponential model

$$\frac{S(0)}{S(b)} = e^{-Cb^{\alpha}}, \qquad (35)$$

where C is a constant and where α is a measure of the fractal dimension of the microstructure [59]. Examples of several geometries with different fractal dimensions are given in figure 16.

3.4.1.2 **Relation to kurtosis**: The underlying distribution of the stretched exponential haves the property that for independent and identical distributed $X_1, \ldots, X_n$, the sum $\sum X_i$ is up to scaling identical distributed as a single X_i [60]. The Gaussian distribution is an example of such a distribution, in fact when
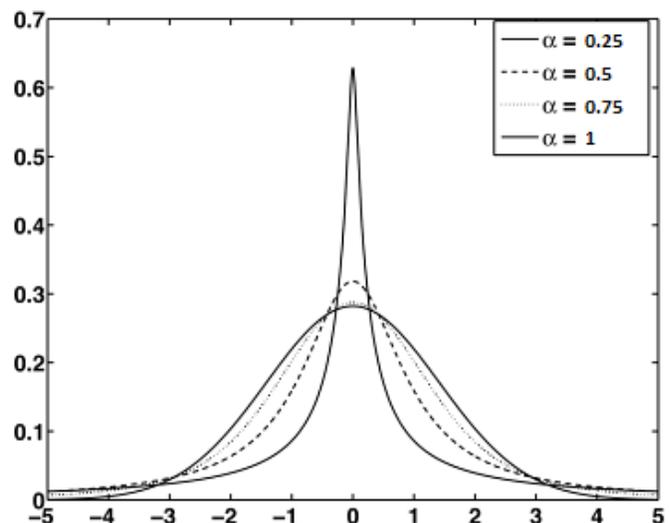
Figure 17. The symmetric alpha-stable distribution for different values of α. Decreasing α results in more peaked and heavier tailed distributions [60].

$\alpha = 1$ the stretched exponential corresponds to a Gaussian. For other values of α, the variance is infinite because such a distribution has too heavy tails (see figure 17). In that case the kurtosis is undefined [5]. However, as can be acknowledged from figure 17, C and α describe the width and the amount of heavy-tailedness of the stable distributions in a manner equivalent to the diffusivity and the kurtosis.

3.4.1.3 **Results from experiments**: One study [61] compared the goodness of fit of the stretched exponential and the biexponential model in rat cortical ribbon (n=6) up to a b-value of $6500\ s/mm^2$ and found that in 20% of the voxels the stretched exponential (which has one parameter less) gives a better fit. In [58] tissue contrast is observed from the parameters of the stretched exponential. In [16] a good fit to the signal for b-values up to $5000\ s/mm^2$ is observed in GM, whereas in [62] a significant difference is observed between α in human high-grade gliomas and healthy tissue for b-values up to $4000\ s/mm^2$.

3.4.1.4 **Discussion**: GM neurites are arguably oriented in a 'fractal' three structure around the cell body (see figure 12) and may therefore be well characterized by their fractal dimension.

There are a few experimental studies supporting the stretched exponential up to b-values of $6500\ s/mm^2$. It would be interesting to investigate if the stretched exponential still gives good results at higher b-values, because a minor change in the signal curve can have a considerable effect on the implied distribution [63]. For this reason, according to [63], the use of a stretched exponential to model diffusion is more likely due to its convenient mathematical representation than due to the validity of the underlying distribution.

The stretched exponential can describe the deviation from Gaussian diffusion in one parameter, while there are in theory an infinite amount of cumulants $\kappa_4, \kappa_6, \kappa_8, \ldots$ possibly bearing information on the deviation from Gaussianity. For this reason α might be a better measure of *non-Gaussianity*, of course provided that the stretched exponential model is valid.

3.5 Microstructural compartmental models without exchange

A class of models that are popular in modeling white matter are what is called in this work **microstructural compartmental models without exchange**. These models consist generally of two compartments, i.e. an intra-axonal (IAS) and an extra-axonal (EAS) compartment, but can also contain a third compartment, for example CSF. Because myelin water is invisible at clinical scan times [22], we can only probe water diffusivity within the axon membrane, not in the myelin sheath. Exchange between the IAS and the EAS is neglected, for reasons of a 'negligible' axon permeability due to myelination [26] [64]. These models can be split in three subgroups by how they model the axon radial diffusion. The first subgroup of models that is reviewed assumes that the radial diffusivity is Gaussian. The second assumes that the axon radius is zero (hereafter referred to as a stick models). The third subgroup that is reviewed assumes that the axon is a totally impermeable cylinder (hereafter referred to as cylinder models). An overview of each presented microstructural compartmental model with the underlying assumptions can be found in table 3.

3.5.1 biTensor models

3.5.1.1 **biTensor model**: Tuch (2002) [65] proposes a biTensor model consisting of two tensors that corresponds to the IAS and the EAS

$$\begin{pmatrix} D_{a,11} & D_{a,21} & D_{a,31} \\ D_{a,21} & D_{a,22} & D_{a,32} \\ D_{a,31} & D_{a,32} & D_{a,33} \end{pmatrix} \quad \begin{pmatrix} D_{e,11} & D_{e,21} & D_{e,31} \\ D_{e,21} & D_{e,22} & D_{e,32} \\ D_{e,31} & D_{e,32} & D_{e,33} \end{pmatrix}$$

This model uses 14 parameters. How could we reduce the amount of parameters? One can hypothesize that the diffusion distribution is cylinder symmetric around the axon axis in both the IAS and the EAS. This gives an IAS and EAS tensor [66]

$$\begin{pmatrix} D_{a,\perp} & & \\ 0 & D_{a,\parallel} & \\ 0 & 0 & D_{a,\parallel} \end{pmatrix} \quad \begin{pmatrix} D_{e,\perp} & & \\ 0 & D_{e,\parallel} & \\ 0 & 0 & D_{e,\parallel} \end{pmatrix}$$

We call these two tensors a **zeppelin** tensor because the radial diffusivities are equal, which gives the 3-dimensional diffusivity distribution a zeppelin shape. This biZeppelin model uses 9 parameters.

3.5.2 Stick models

3.5.2.1 **Tensor & stick:** The tensor & stick model studied by Fieremans (2011) [26] is a restriction of the biTensor model. The axonal perpendicular diffusivity is in this model assumed to be zero, which should be a reasonable assumption whenever $t_d >> R^2/D_a$ [26]. This gives two tensors

$$\begin{pmatrix} D_{a,11} & D_{a,12} & D_{a,13} \\ D_{a,21} & D_{a,22} & D_{a,23} \\ D_{a,31} & D_{a,32} & D_{a,33} \end{pmatrix} \quad \begin{pmatrix} D_{\parallel,e}cos\theta & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Relation to kurtosis The tensor & stick model implies a radial kurtosis of [26]

$$K_\perp = 3 * \frac{v}{1-v}, \tag{36}$$

where v is the axonal fraction. Therefore the axonal fraction can be estimated by the radial kurtosis

$$v = \frac{K_\perp}{K_\perp + 3}. \tag{37}$$

Hence, according to this model, kurtosis solely depends on the axon fraction v.

3.5.2.2 **Stick models with orientation dispersion**: Axons are, however, almost never totally coherent aligned. The full-width at half-maximum (FWHM) of the dispersion to the main axon direction is found to be around $34°$ in the corpus callosum [22]. This motivates the extension of the stick component S_c with a dispersion component, i.e.

$$S_c(\vec{n}) = \int_\nu p(\vec{\nu})e^{-bD_{e,\parallel}(\vec{n}\cdot\vec{\nu})^2}d\vec{\nu}, \tag{38}$$

for which the axons in directions $\vec{\nu}$ are dispersed around the central axon direction $\vec{\mu}$. The (continuous) relative fraction of axons with direction $\vec{\nu}$ is denoted by $p(\vec{\nu})$. This is called the **orientation distribution function** of $\vec{\nu}$.

3.5.2.3 **Noddi** : The NODDI model [67] is a stick model extended by a Watson distribution, which orientation distribution function is given by

$$p(\vec{\nu}) = [1F1(\frac{1}{2}, \frac{3}{2}, \kappa)]^{-1}e^{\kappa(\vec{\mu}\cdot\vec{\nu})^2} \tag{39}$$

where $\vec{\mu}$ is the mean axon direction, κ is a dispersion parameter, and where $1F1$ is the confluent hypergeometric function of the first kind [67]. The Watson distribution should be the simplest spherical pdf that can account for orientation dispersion.

The EAS is modeled by a distribution of zeppelins, with an extra tortuousity restriction $D_{e,\perp} = D_{e,\parallel} * (1 - v)$. Every zeppelin in this distribution corresponds to a particular axon orientation. Hence if a fraction of the axons point in $\vec{\nu}$, a fraction of the EAS is tortuous in the direction perpendicular to $\vec{\nu}$ because of hindrance from this fraction of axons. Also an isotropic Gaussian CSF compartment is included.

3.5.2.4 **Validation of stick models with orientation dispersion**: In a preliminary work Ferizi (2013) [66] compares stick models with and without orientation dispersion in the human brain in vivo on relatively coherently oriented fibers in the corpus callosum (CC). Stick models with an orientation dispersion component give considerably better results than those without orientation dispersion. For regions with a minor orientation dispersion a CSF compartment is favored over a zero radius sphere (dot) as a third compartment. When, however, the orientation dispersion increases, a dot model is favored. A zeppelin is found to be the best choice to model the EAS.

3.5.2.5 **Discussion**: It makes sense to accommodate for orientation dispersion in WM diffusion models. The remarkable behavior of dot and sphere compartments when the dispersion is varied is an indication that the inclusion of a third EAS compartment based on microstructure should not be undisputed. The orientation dispersion models are in fact 'statistical' models in the sense of section 3.3. Because the cylinders make different angles with the gradient vector, for every neurite direction a different effective diffusivity is measured [15]. This gives variance in diffusivity and hence kurtosis.

3.5.3 Cylinder models

The cylinder models presented in this section essentially differ from the previously reviewed stick models by just one aspect, that is the axon radius R, which is non-zero in these models.

3.5.3.1 **Composite hindered and restricted model of diffusion (CHARMED)** : The CHARMED model of Assaf (2004) [68] assumes that the radial axon diffusion is totally restricted. A formula for restricted diffusion within impermeable cylinders from Neuman (1974) [69] is used to model the signal from

the IAS in the perpendicular direction[4]

$$\frac{S_{a,\perp}(TE,|\vec{g}|)}{S(0)} = e^{-\frac{R^4\gamma^2|\vec{g}|^2}{D_a}\frac{7}{96}\left(TE-\frac{99}{112}\frac{R^2}{D_a}\right)} \qquad (40)$$

where D_a is the intrinsic intra-axonal diffusivity and TE is the echo time. This formula tacitly assumes that $t_d >> R^2/D_a$. Diffusion in the EAS is modeled by a zeppelin component.

Relation to kurtosis DeSantis (2012) [70] investigates numerically the relations between $D_{a\parallel}$, v and $K_\perp$ that are implied by the CHARMED model. As can be seen in figure 18 a positive relation is found between v and $K_\perp$ and also between $D_{a\parallel}$ and $K_\perp$, for $v < 0.8$. When $v > 0.8$, for $D_a \leq 1\ \mu m^2/ms$ the relation between $K_\perp$ and v is negative, whereas for other values of D_a this relation is positive. These trends are also tested in vivo in the same study. The relation between $K_\perp$ and v is found to be significant, whereas the relation between $K_\perp$ and the intrinsic diffusivity $D_{a\parallel}$ was found to be insignificant.

3.5.3.2 **AxCaliber** : The CHARMED model is extended to the AxCaliber model by Assaf (2008) [45] to accommodate for variation in axon diameters. The axon diameter is not fixed like in the CHARMED model, but randomly distributed following a gamma distribution.

Results from experiments High correlations of 0.98 and 0.86 between the axon radii measured by AxCaliber and by light microscopy are reported in [45] (see figure 20). However, small axon diameters are difficult to probe, since the apparent diffusivity goes rapidly to zero over time in those axons [45].

3.5.3.3 **Minimal model of white matter diffusion (MMWMD)**: Alexander (2010) [64] proposes a 'minimal' white matter model consisting of 4 compartments. This MMWMD model is essentially the CHARMED model extended with a CSF and a trapped water compartment (such that $D \approx 0$). This trapped water component should correspond to water bound to membranes and water inside glial cells and other small structures. Also the restriction $D_{a,\parallel} = D_{e,\parallel}$ holds. The model only uses the parameters S_0, $D_{a\parallel}$, D_{CSF}, R, $\vec{\mu}$ and the relative fractions $p_{trapped}$, p_{IAS} and p_{EAS}.

3.5.3.4 **Discussion**: Modeling a restricted axon component by solving differential equations like Neuman (1974) [69] seems to explain the measured signal reasonable and seems to estimate the axon radius well. However, for those who are not familiar with the tedious derivation of differential equations like in Neuman (1974) the signal formulas for the restricted axon are a black box. Because the CHARMED model assumes that $t_d >> R^2/D_a$, for which the radial axon diffusion distribution should be approximately uniform [2], it should be possible to derive a simpler expression for the radial axon diffusivity without trading in too much accuracy.

In figure 19, I plotted $K_\perp$ of an 'ordinary' biexponential function to the same values of v as DeSantis (2012) [70] did for the CHARMED model. One can appreciate from this figure that it captures for $v < 0.8$ the same trends (notice that the axon fraction v is in practice never larger than 0.8) as the CHARMED model. It would have been arguably far more interesting if DeSantis (2012) [70] has investigated the theoretical time dependence of the radial kurtosis implied by the CHARMED model.

A totally restricted axon in the radial direction leads to time dependence of the radial diffusivity and kurtosis, whereas

4. In the original article of Neuman (1974) [69] 7/296 instead of 7/96 appears in the signal formula. It is to the best of my knowledge not clear where this discrepancy comes from.
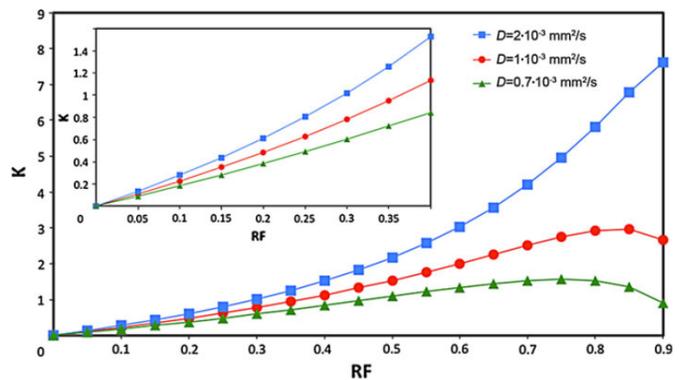
Figure 18. $K_\perp$ vs axonal fraction RF for 3 different $D_{a\parallel}$ values according to the CHARMED model [70].
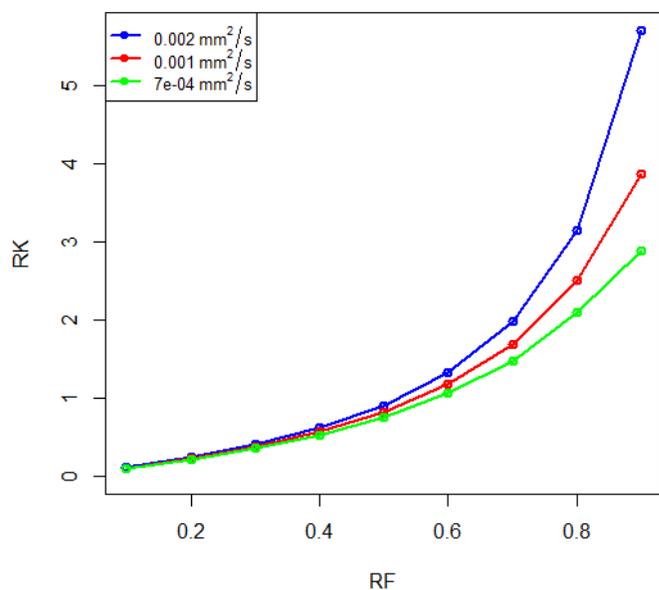
Figure 19. $K_\perp$ vs axonal fraction RF for the same 3 different $D_{a\parallel}$ values according to the biexponential model, assuming $D_{a,\perp} = 0.05$.
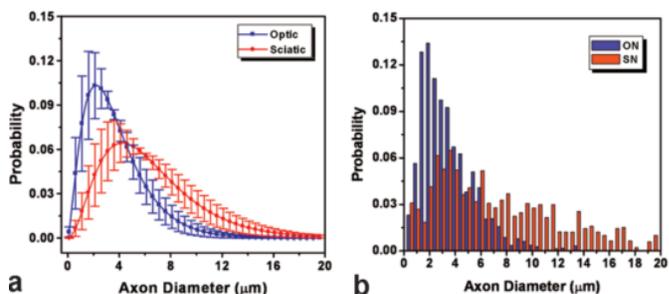
Figure 20. a: Extracted AxCaliber axon diameter distribution b: Axon diameter distribution derived from electron microscopy. ON and SN: porcine optic and sciatic nerve [45].

	BiZ	TS	NOD	MMW	CHD	AxC
# Parameters	9	12	9	10	9	10
No exchange	X	X	X	X	X	X
Straight & parallel axons	X	X		X	X	X
Gaussian parallel axonal diffusion	X	X	X	X	X	X
$t >> \frac{R^2}{D_a}$		X	X	X	X	X
Restricted axonal diffusion		X	X	X	X	X
Zero axon radius		X	X			
Random distributed axon radius						X
EAS dimension	2	3	2	2	2	2
$D_{a\parallel} = D_{e\parallel}$			X	X		
Trapped water compartment				X		
CSF compartment			X	X		

Table 3

Overview of microstructural compartment models that have been reviewed in this work.

experiments show no dependence between the kurtosis and the diffusion time in white matter [71], [22]. It is therefore questionable if the high correlations between the AxCaliber model and histological parameters that are measured would be consistent over a broader range of diffusion times. Moreover, true restricted diffusion effects have not been really observed for water in the brain, according to [48]. It seems more logical that axons are not totally restricted, but allow limited exchange between the IAS and EAS. In section 6.1 I will show that the apparently negligible difference between total restriction and almost total restriction can have a significant effect on the radial kurtosis.

3.6 Comparison of microstructural multi-compartment models without orientation dispersion

Ferizi (2013) [27] performed an systematic experimental comparison of all types of multi-compartment models without orientation dispersion. These models are a combination of the following compartments:

- Restricted extra-axonal diffusion possibly consisting of two compartments. The first compartment can be chosen to be

1) Isotropic (ball).
2) Anisotropic with cylinder symmetry (zeppelin).
3) Fully anisotropic (tensor).

   A possible second EAS compartment can be chosen from

1) Astrosticks; uniform orientation distributed cylinders with zero radius.
2) Astrocylinders; uniform orientation distributed cylinders with non-zero radius.
3) Isotropic spheres; diffusion inside a sphere. This can be interpreted as CSF contamination.
4) Dot; spheres with zero radius, which implies that particles inside do not move.

- An intra-axonal compartment modeled as a cylinder pointing at $\vec{\mu}$, where the cylinder radius can be zero, fixed, or gamma distributed.

All combinations of those compartments were tested in vivo on relatively straight fibers in the corpus callosum of a human brain and evaluated to the BIC criterion. B-values ranged from 218 to 10.308 s/mm^2 in 45 directions. The study draws some interesting conclusions:

1) Even the simplest two-compartment models, such as the ball & stick model, explain the signal intensity much better than the diffusion tensor model.
2) The tensor and zeppelin models consistently perform better than the isotropic ball in the EAS, supporting the need to accommodate for anisotropy in the EAS. The radial eigenvalues of the tensor are found to be similar, therefore a zeppelin is the best model for the EAS.
3) Three compartment models outperform two compartment models. "Sphere" and "dot" are favored over "astrosticks" and "astrocylinders".
4) "Sphere" and "dot" three compartment models underestimate the axon fraction by almost a half, probably caused by some restricted diffusion from the cylinders that is captured by the third restricted component. "Astrosticks" models overestimate the mean axonal radius roughly four times.
5) The fitting procedure estimates the radius of the astrocylinders and sphere to be almost zero (0.1 μm), therefore the BIC prefers the simpler astrosticks and dot compartments.

3.6.1 Discussion

According to the comparison study, white matter models should accommodate for cylindrical symmetric diffusion in both the IAS and the EAS. On top of that, the addition of a third restricted compartment to a model gives a better BIC, but a worse estimation of parameters. In this light it is interesting that another experiment did not find any suggestion for an extra-axonal isotropic restricted compartment [72].

Irrespective of whether such a third restricted compartment should be included or not, one can ask: what should be the third compartment? A trapped water compartment is favored by the BIC, but it can be challenging to distinguish a water signal arising from environments with highly restricted diffusion ($D \approx 0$) from noise [22]. The second best choice according to the BIC is a spherical compartment, which can be interpreted as CSF contamination. It should be kept in mind, however, that diffusion in the CSF is not truly Gaussian, because of flow effects, choroid plexus and membranes [53]. The worst BIC belongs to an astrocyte compartment. The reader should not be surprised: at clinical diffusion times, a substantial part of water molecules may already been leaked out, whereas the astrocytes are modeled as *impermeable* cylinders. What the results of the comparison study show but what is not further addressed in Ferizi (2013) [27], is that a biZeppelin model gives better results than a zeppelin & stick or a zeppelin & cylinder model. Therefore the radial axonal diffusion may be better modeled as Gaussian diffusion than by a restricted stick or cylinder.

By assuming diffusion to be equal in all radial directions, one reduces the diffusion process in fact to 2 dimensions. Therefore the kurtosis tensor reduces to the independent parameters $E[X^4]$, $E[X^2Y^2]$ and $E[Y^4]$. Because independence implies that $E[X^2Y^2] = E[X^2]E[Y^2] = 4D_\parallel D_\perp$, the only independent kurtosis parameters in the kurtosis tensor are essentially $K_\perp$ and $K_\parallel$. Hence one can see, by making a cumulant expansion of the model signal intensity, that this signal does not depend on so many parameters as one might a priori expect. Therefore one -regarding the many parameter that are involved in those microstructural models- should be cautious for overfitting in studies like Ferizi (2013) [27].

3.7 Modeling neurite orientation and compartmentalization in GM

Since WM can be modeled well by intra- and extracellular compartmentalization and orientation distributed axons, one may ask if we could use this approach to model GM neurites as well. The literature provides far less such microstructural models for GM, but one such model seems to explain GM diffusion particular well. This is the neurite orientation model of Jespersen (2007) [73] which can, by some minor adjustments, also model WM diffusion. This model is included because it has an interesting relation with kurtosis, but also to give a taste of more advanced diffusion modeling. Because within the framework of Ferizi (2013) [27] simpler microstructural WM models are already well modeled and validated, this section will only focus on the application of the neurite orientation model in GM.

3.7.1 Neurite orientation model

The in this work called **neurite orientation model** of Jespersen (2007) [73], contains an isotropic Gaussian compartment and a compartment that exists of a distribution of neurites which point in different directions. This leads to the the signal intensity

$$\frac{S(b, \vec{n})}{S(0)} = (1-v)e^{-bD_{eff}} + v\int_{S^2} p(\vec{v})e^{-b(D_L(\vec{v}\cdot\vec{n})^2 + D_T(1-\vec{v}\cdot\vec{n})^2)}d\vec{v} \tag{41}$$

where D_{eff} is the diffusivity outside the neurite compartment, D_L and D_T are the parallel resp. radial diffusivity of the neurites, and $p(\vec{v})$ is the fraction of neurites that point in direction $\vec{v}$. A visual example of such an distribution is given in figure 21. Since $\vec{v}$ is a vector on the sphere, we can also write $p(\vec{v})$ as a function of spherical coordinates

$$p(\vec{v}) = p(\theta, \phi) \tag{42}$$

where θ and ϕ are the spherical coordinates of $\vec{v}$. It is well known from basic anatomy that $p(\theta, \phi)$ should generally be more uniform for GM neurites than for WM axons, that often have only one or two main orientations.

Jespersen (2007) [73] expands the spherical distribution function $p(\theta, \phi)$ into certain basis functions called spherical harmonics. This expansion is based on the fact that any well behaved function on the sphere can be written as a sum of some basis functions, like a Fourier expansion does for periodic functions. In appendix N we will provide a rigorous derivation, especially for the reader who is not yet convinced of this striking mathematical fact. The spherical harmonics expansion yields the following signal intensity

$$\frac{S(b, \theta, \phi)}{S(0)} = (1 - v)e^{-bD_{eff}}$$
$$+ v\pi e^{-bD_T} \cdot \sum_{l,m} f_{lm}C_l(b(D_L - D_T))Y_{lm}(\theta, \phi) \tag{43}$$

where C_l and $Y_{lm}(\theta, \phi)$ are functions related to the spherical harmonic expansion. The harmonic series is truncated at $l = 2$ or $l = 4$, which results in an amount of parameters between 10 and 18 depending on the truncation.

3.7.1.1 **Results from experiments**: Jespersen (2007) [73] compares the truncation of the harmonic series ex vivo in a newborn baboon brain at $l = 2$ and at $l = 4$, with $\Delta = 50$ ms and with b-values up to $15000 \ s/mm^2$ in 153 directions.
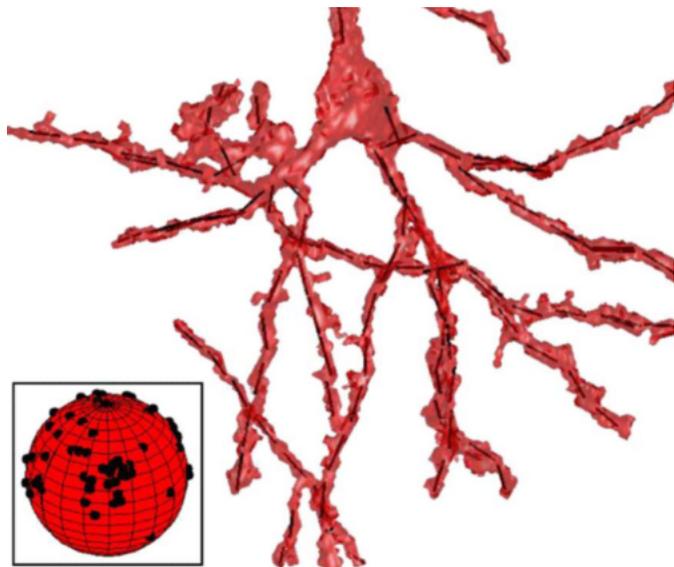
Figure 21. Orientation distribution of neuronal processes [74].
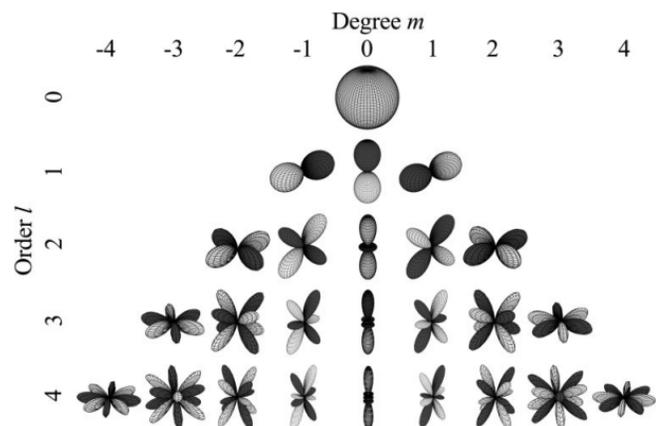
Figure 22. Plots of the real combinations of the spherical harmonics. Black corresponds to negative values [75].

In GM regions the $l = 2$ truncation is sufficient to explain the measured signal.

The truncation at $l = 4$ has been studied on ex vivo rat brain by [1], with b-values up to $15000 \ s/mm^2$ in 144 directions. Because dendrites have a low exchange time of $\tau = 10$ ms, Δ is set to 8 ms. Significant Pearson correlations in GM were observed between the measured neurite fractions and the neurite fraction measured by light microscopical staining.

3.7.1.2 **Relation to kurtosis**: In general, we have for the apparent kurtosis tensor of the neurite orientation model, according to Jespersen (2012) [74]

$$\boldsymbol{K}_{app} = \frac{[vD_A^2\boldsymbol{W} + (1-v)v((D_{eff} - D_T)\boldsymbol{I} - D_A\boldsymbol{T})^{\otimes 2})}{Tr(\boldsymbol{D}_{app}/3)^2} \tag{44}$$

with $\boldsymbol{D_{app}} = (1 - v)D_{eff} + v(D_T\boldsymbol{I} + D_A\boldsymbol{T})$, $D_A = D_L - D_T$, $\otimes 2$ the tensor product of a tensor with itself (see appendix C), and $\boldsymbol{T}$ resp. $\boldsymbol{W}$ the second resp. fourth order tensor of the

orientation distribution function $p(\vec{\nu})$, i.e. $T_{ij} = E[\nu_i \cdot \nu_j]$ and

$$W_{ijkl} = 3E[\nu_i\nu_j\nu_k\nu_l] - E[\nu_i\nu_j][\nu_k\nu_l]$$
$$- E[\nu_i\nu_k]E[\nu_j\nu_l] - E[\nu_i\nu_l]E[\nu_k\nu_j]. \quad (45)$$

What this formula especially reveals is that the kurtosis tensor metrics -according to this model- depend on the kurtosis tensor of the orientation distribution.

If the orientation distribution of the neurites is isotropic, then the mean kurtosis is equal to (see Mulkern (2009) [76] and appendix J for a derivation)

$$3\frac{(1-v)D_{eff}^2 + v\left(D_T^2 + \frac{2*D_T(D_L - D_T)}{3} + \frac{(D_L - D_T)^2}{5}\right)}{\left((1-v)D_{eff} + v(\frac{2}{3}D_T + \frac{1}{3}D_L)\right)^2} - 3$$
$$(46)$$

This formula is especially useful to calculate the kurtosis of an astrocyte compartment, since we already saw that this can be modeled as an uniform distribution of impermeable cylinders. Because the directional kurtosis $K(\vec{n})$ is, in all generality, a function on the sphere, it can also be expanded into spherical harmonics [44]. Notice, however, that this is something different than expanding an orientation distribution function of neurites or axons in spherical harmonics.

3.7.1.3 **Discussion**: The neurite orientation model can make a good fit to the signal and seems to estimate the neurite fraction reasonably in ex vivo rat brains. The imaging protocol that is used in the validation experiments is, however, clinical impractical [67]. Therefore the model is not yet validated in the human brain in vivo. This model uses almost 5 times as much parameters as alternative GM models like the stretched exponential or the Kärger model. Therefore one should be cautious for overfitting.

3.8 Axonal morphology

Axonal morphology, e.g. bulging, bending and undulation could have an effect on the kurtosis metrics in WM. Monte Carlo simulations can reliable estimate diffusion distributions [77] and can be used as a tool to investigate diffusion processes in axons with different morphologies.

3.8.1 Axonal damage

Landman (2010) [77] investigates the effect of axon damage on the kurtosis with a Monte Carlo simulation. The kurtosis metrics of healthy, bulged, crimped and broken axons resulting from the simulation are shown in figure 23. The largest effect from axon damage is observed for the axon kurtosis in the parallel direction in case of a bulging axon, in which case the radial axon kurtosis is also affected. A crimped axon causes in both principal directions a very small increase in axon kurtosis. A broken axon only changes the parallel axon kurtosis.

3.8.2 Axonal bending

Landman (2010) [77] also performed a simulation on the effect of axonal bending on the radial and perpendicular diffusion probability distribution. As shown in figure 24, the effective radial diffusivity of the axon increases when the axon bends.

3.8.3 Axonal undulation

Axonal courses change due to tension and compression in a manner similar to that of the bellows of an accordion. This axonal undulation is considered to have a protective function, and is present in those parts of the central nervous system that
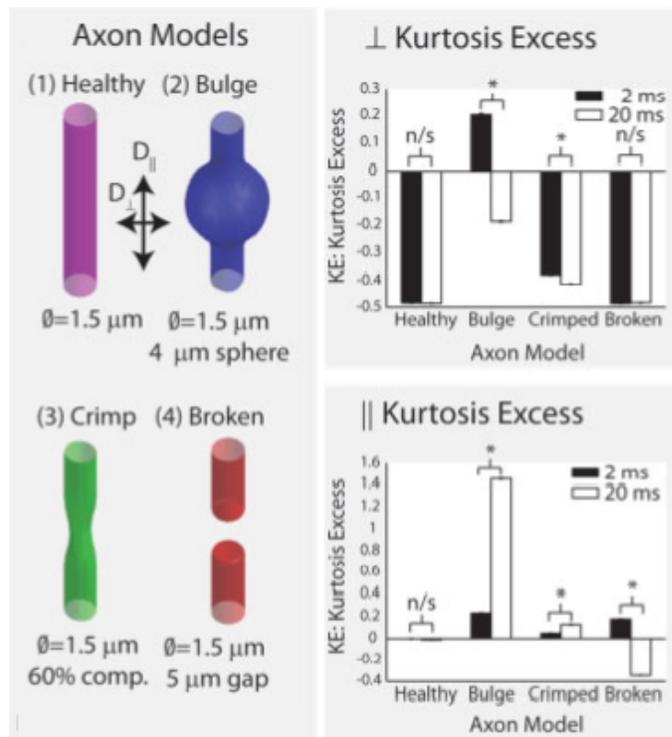
Figure 23. The simulated effect of axon damage on kurtosis [77].
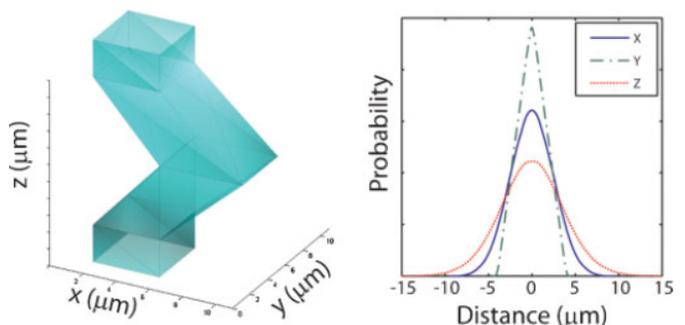
Figure 24. Effects of axonal bending on the axonal diffusion distribution [77].

are subjected to strain during locomotion such as the optic nerve. The undulation pattern is observed to be sinusoidal, see figure 25, or helical [78]. Nilsson (2012) [78] investigated the effect of axonal undulation on the diffusion distribution analytical and by simulations. In case of axon undulation the effective radial axon diffusivity will be higher, see figure 26.

4 DISCUSSION

After we obtained in the previous section some feeling for how to model diffusion in the human brain, this part of the text will summarize the causes of kurtosis in the human brain which are implied by the reviewed diffusion models in the literature. The discussion is structured as follows. First, we will give an answer to the question: how to model diffusion in the human brain appropriately with simple models? Then we
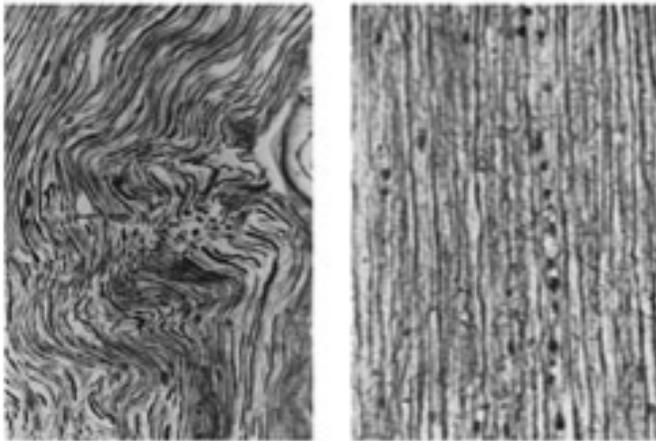
Figure 25. Undulation in sagittal sections through the cervical cord at the level of C6 [78].
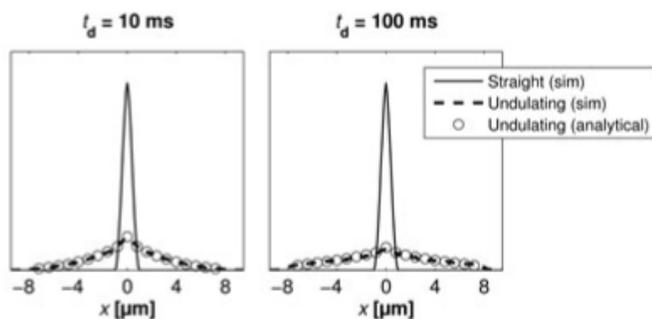
Figure 26. Radial diffusion displacement distribution, from straight and undulated axons obtained analytical (circles) and by simulation (broken line) [78].

will sum up all the relations that could be established between diffusion model parameters and kurtosis metrics. After these considerations, a conclusion will be given on which parameters could be expected to have an effect on kurtosis metrics.

4.1 How to model diffusion in the human brain with relatively simple models?

In WM, relatively simple signal intensity models are successfully applied in vivo to regions of high fiber coherence; models which account only for orientation dispersion in the IAS and diffusion anisotropy in the EAS already can explain the measured signal well and estimate parameters reasonably. The inclusion of an exchange, CSF, trapped water or astrocyte component may also improve a WM diffusion model, but this has yet not been totally clarified. Making new diffusion models by combining components from different existing models has proven to be successful in WM. In table 4 an overview is given of the WM models reviewed in this work together with their relative performance.

The literature shows less advancement in the modeling of GM. In the neurite orientation model, GM is modeled in a same fashion as WM where the axons are replaced by neurites. In the stretched exponential model, the signal intensity arising from GM is related to the microstructural fractal dimension. Both

WM models	Realistic parameters	Signal explanation
BiZeppelin	+/-	+
Tensor & stick	+/-	+/-
MMWMD	+/-	+
CHARMED	+/-	+/-
AxCaliber	+[1]	-
NODDI	+/-	++

Table 4
Relative in vivo performance of WM models in regions of coherent fibers, based on [27] [66]

1. Ex vivo

models provide some good results, but are sparsely validated and compared to other models.

4.2 Relations between biophysical parameters models and kurtosis metrics

This section discussed the causes of kurtosis and ultimately leads to some DKI rules of thumb. Jensen (2010) [5], Jespersen (2010) [1], DeSantis (2012) [70], Mulkern (2009) [76], Fieremans (2010) [50], Fieremans (2011) [26], and Yablonskiy (2010) [15] are examples of articles in the literature which relate diffusion models to kurtosis metrics or at least to a cumulant expansion. From these articles I obtained some relations between anatomical parameters and kurtosis. I will show, however, that it is possible to derive even more relations by combining for example several sources from the literature.

4.2.1 Relations which can be directly obtained from the literature

This section is about relations between kurtosis and diffusion models which can be straight obtained from the articles just named.

4.2.1.1 **Difference between directional tortuousity between compartments**: When a compartment is more tortuous, i.e. particles have to travel a longer path to reach the same point because of hindrance, it has effectively a lower diffusivity. Therefore, a difference in tortuousity between compartments in a certain direction causes variance in diffusivity and therefore kurtosis in that direction [26].

4.2.1.2 **Exchange**: According to the Kärger model studied by Fieremans (2010) [50], exchange reduces the kurtosis metrics.

4.2.1.3 **Axon/neurite fraction**: According to the tensor & stick model of Fieremans (2011) [26], axon fraction is the only parameter that makes sense in WM radial kurtosis. A higher axon fraction leads to a higher kurtosis. This is confirmed by the numerical calculations of DeSantis (2012) [70] as shown in figure 18 which where performed on the CHARMED model of Assaf (2004) [68]. The kurtosis tensor of the GM neurite orientation model from Jespersen (2010) [1] reveals us that neurite fraction has an effect on the kurtosis metrics.

4.2.1.4 **Axon radius**: According to DeSantis (2012) [70], the radial kurtosis is dependent on the axon radius.

4.2.1.5 **Intrinsic axon/neurite diffusivity**: The numerical calculations of DeSantis (2012) [70] show that a higher axonal intrinsic diffusivity causes a higher radial kurtosis, according to the CHARMED model of Assaf (2004) [68]. Jespersen (2012) [74] derived that the GM voxel kurtosis metrics are depended on the intrinsic neurite diffusivity according to the neurite orientation model.

4.2.1.6 **Second and fourth order tensor of the orientation distribution**: Jespersen (2012) [74] derived that the kurtosis metrics of a voxel which includes a compartment of orientation distributed neurites depends on the second and fourth order tensor $\boldsymbol{T}$ and $\boldsymbol{W}$ of the orientation distribution. One can conclude from Jespersen (2012) [74] that in general higher values in the kurtosis tensor $\boldsymbol{W}$ of the axon orientation distribution causes a higher kurtosis in the voxel.

4.2.2 Relations which can be indirectly obtained from the literature

In this section relations between the kurtosis and the reviewed diffusion models are stated for which one has to combine several sources from the literature and possibly perform some calculations.

4.2.2.1 **Axon radius**: DeSantis (2012) [70] states that kurtosis is dependent on the axon diameter but does not specify in which direction. However, since a larger radius causes a higher radial diffusivity $D_{a,\perp}$ and hence less diffusion heterogeneity in the radial direction, a higher axon radius causes a lower radial voxel kurtosis in WM.

4.2.2.2 **Glial cells and other small restricted compartments**: According to the MMWMD model of Alexander (2010) [64], water in glial cells and other small restricted compartments is trapped, i.e. for these compartments it holds that $D \approx 0$ $\mu m^2/ms$. Since these trapped water particles add up to the peak of the diffusion displacement distribution, the trapped water fraction can almost surely be expected to increase the kurtosis metrics. After all, in the context of diffusion MR, kurtosis is a measure of peakedness [70].

4.2.2.3 **Water that is bound to membranes**: Water that is bound to membranes has, because $D \approx 0$ $\mu m^2/ms$, following an equivalent reasoning a similar effect on kurtosis metrics as glial cells.

4.2.2.4 **Axonal radial kurtosis** : Because an axon has a small radial diffusivity of around $D_{a,\perp} = 0.05$ $\mu m^2/ms$, for which the exact value depends on the axon diameter, again the partial volume kurtosis formula (equation 23) is of use. With this formula we can derive that assuming restricted diffusion, which typically gives an axonal kurtosis $K_{a,\perp}$ of around -0.5, the effect of $K_{a,\perp}$ on the voxel RK is generally negligible since $K_{a,\perp}$ gets weighted by the relatively very small $D_{a,\perp}^2$.

4.2.2.5 **Axon crimping**: Landman (2010) [77] performed a simulation to asses the effect of crimping on axons, as in figure 23, on $K_{a,\perp}$ and $K_{a,\parallel}$. From the axon diffusion distribution (not shown), no change in diffusivity was apparent. An effect on $K_{a,\perp}$ was measured, but this will only contribute marginally to the voxel radial kurtosis as discussed in the previous paragraph. Hence the crimping of axons has a negligible effect on the voxel RK and a small effect on the voxel AK.

4.2.2.6 **Bulging and broken axon**: A bulging and a broken axon are found by Landman (2010) [77] to have a $K_{a,\parallel}$ of 1.4 resp -0.4. If the fiber points straight in one direction, $K_{a,\parallel}$ of the bulged or broken axon compartment has a major impact on the voxel AK, because of the relatively large diffusivity of the axon, by the partial volume kurtosis formula (equation 23).

4.2.2.7 **Orientation dispersion, axon undulation** & **axon bending**: Everything that makes the width of the radial displacement distribution of an axon wider causes less diffusion heterogeneity in the radial direction and hence less kurtosis. Hence one could argue that more orientation dispersion within a fiber, axon bending and undulation all causes a decrease of the voxel RK.

4.2.2.8 **Astrocyte fraction**: Because, according to Ferizi (2013) [66], one can model astrocytes as cylinders or sticks that are uniformly distributed over the sphere, one can use equation (46) to obtain, assuming an intrinsic astrocyte diffusivity of 0.8 $\mu m^2/ms$, an apparent kurtosis of 0.46 for a stick, and an apparent kurtosis of 0.45 for a cylinder. Hence for an uniform distribution of cylinders, the kurtosis seems to be independent on the cylinder radius. Now consider the genu of the CC. In table 2 one can find the in this region RK and RD to equal 2.07 resp. 0.38. Suppose, for example due to astrogliosis, that 10% of the voxel fraction becomes occupied by astrocytes. The new radial kurtosis will, according to equation (23), decrease from 2.07 to 2.01. Hence in this example a slightly lower RK is caused by a higher astrocyte fraction. This may be somewhat counter intuitive since one may think that more variety in voxel composition may automatically lead to a higher kurtosis. In general, the effect of the astrocyte compartment depends on the exact value of a handful of parameters: the DKI metrics of the voxel without an astrocyte fraction, the astrocyte diffusivity and the relative fraction of the astrocyte compartment.

4.2.2.9 **Astrocyte radius**: As derived in the latter paragraph, the radius of cylinders in the uniform orientation model which is applicable to astrocytes does not have a significant effect on the apparent kurtosis.

4.2.2.10 **CSF fraction**: Yang (2013) [6] experimentally investigated the effect of CSF contamination on GM and WM mean kurtosis in vivo. According to Yang (2013) the direction of the effect of CSF contamination on kurtosis metrics is not a priori obvious because experiments have found that CSF has a low kurtosis which would reduce the voxel kurtosis, whereas on the other hand diffusion heterogeneity caused by CSF contamination would raise the voxel kurtosis. Because I derived formula (23) which states how kurtosis behaves under partial volume effects and how the two contributions that Yang (2013) describes are weighted against each other, i.e.

$$\frac{3v_{csf}(1 - v_{csf})(D_{csf} - D)^2 + v_{csf}D_{csf}^2 K_{csf} + (1 - v_{csf})D^2 K}{D_{app}^2}$$

where D an K are the voxel diffusivity and kurtosis before the addition of an (extra) CSF compartment. My hypothesis is therefore that one a priori could have known how CSF changes the kurtosis, by considering this formula. Yang (2013) also founds that GM kurtosis is more susceptible for CSF contamination than WM kurtosis. The reason should be the location of GM and WM relative to each other and to the CSF. Again, my hypothesis is that this can be explained by the partial volume formula (equation 23).

Consider an isotropic voxel such that $D = 0.8$ $\mu m^2/ms$ and such that $K = 0.7$. Suppose this voxel gets for 5% contaminated by CSF, for which D and K are observed to be around 2 $\mu m^2/ms$ resp. 0.5 [53]. The kurtosis of the voxel contaminated by CSF can be calculated by formula (23) to increase from 0.7 to 0.99. Hence, we can generally expect the mean kurtosis to be increased by CSF contamination. It depends however, on a handful of parameters: the DKI metrics of the voxel without CSF contamination, the diffusivity of the CSF compartment, and the relative fraction of the CSF in the voxel.

4.2.2.11 **Shape of distribution of axon radii**: According to the AxCaliber model, axons in the same voxel do not have a fixed radius, but a whole distribution of radii. This distribution is originally assumed to be a gamma distribution, but according to the comparison study of Ferizi (2013) [66], it can be also

modeled by a Gaussian distribution. Because the diffusivity is proportional with the square of the radius, the distribution of the *diffusivities* is again a gamma distribution [8], for which the shape parameter α should be at most 1.5 to be physically plausible. Hence, by formula (53) the axonal radial kurtosis $K_{a\perp}$ is due to the diffusivity variance at most 2. By applying the partial volume kurtosis formula (equation 23), the effect of the axon radii distribution on the voxel RK can be found to be very small (< 0.05). However in an extreme situation where the mean radius and axon fraction are relatively high, the shape of the axon distribution has a heavy tail and the extracellular space is very tortuous, one can calculate that the voxel RK could be increased by as much as 1 due to the shape of the axon radii distribution.

4.2.3 "What causes kurtosis in the human brain?"

In this work we try to formulate an answer to the question: "What causes kurtosis in the human brain?", from an theoretical and literature-based approach. Ideally, we would like to have encountered in the literature microstructural models that are thoroughly validated and from which all relations between parameters and kurtosis metrics are established. However, studies that compare several models according to an uniform criterion such as the AIC and BIC are sparse. Also studies that derive kurtosis metrics analytical from diffusion models are sparse, given the numerous *possible* relations between kurtosis and model parameters. It is difficult to compare histological parameters with parameters derived from a MR signal. Moreover, not necessarily every parameter of a model could have a significant effect on the kurtosis.

Therefore it is not obvious how to make a distinction between parameters that affect kurtosis and parameters that do not affect kurtosis, based on this literature study. I, however, ordered the parameters that a priori, i.e. before performing experiments, could reasonably be expected to have an effect on the kurtosis metrics according to two criteria: 1. A parameter should be included in a reasonable diffusion model, that is reviewed in this work 2. Theoretical results in the literature should turn out that the contribution of the parameter to a kurtosis metric is generally significant.

Parameter that meet both of these criteria, and therefore one can say affect kurtosis theoretically, are :

- Tortuousity differences between compartments
- Axon/neurite fraction
- Axon radius
- Neurite distribution orientation
- Intrinsic axon/neurite diffusivity
- Exchange
- Axon bulging, breaking and crimping (in the parallel direction)

Parameters that are suspected to have a relation with kurtosis, but for which these possible relations have not yet been derived in the literature, are:

- Axon orientation dispersion
- CSF fraction
- Astrocyte fraction
- Water that is bound to membranes
- Glial cells and other small restricted compartments
- Axon undulation & axon bending

Parameters that, generally, cannot meet criterion 2, and hence in theory generally do not affect kurtosis are:

- Radial kurtosis of a restricted axon compartment.

- Distribution of axon radii.
- Axon bulging, breaking and crimping (in the radial direction)
- Astrocyte radius

This list may answer the main question of this work satisfactory. However, one has to keep in mind that we are not talking about empirical relations between kurtosis and tissue microstructure, but that we have limited ourself to a more theoretical question of what effect parameters would have when the voxel was exactly composed as in the tissue models. In table 5 and 6, some rules of thumb are stated about kurtosis metrics according to some of the reviewed diffusion models. A warning should be made again that the kurtosis metrics are practically never a linear function of the model parameters, and are almost never only depended on just one single parameter. Therefore the precise magnitude, but possibly also the direction, of the effect of a parameter on a kurtosis metric depends simultaneously on the exact values of all relevant parameters.

5 LIMITATIONS

In this work, we investigated the sources of non-Gaussian diffusion measured with DKI by a literature study on theoretical diffusion models. There are some limitations to this work, which should be noted.

5.1.1.1 **Crossing fibers:** The value of the voxel $K_\parallel$ and $K_\perp$ makes sense because they correspond to the voxel kurtosis in the radial and axial direction of the fiber. However this logically does not hold anymore if there are two or more major bundles of fibers with different directions in the voxel. In this case one could of course consider other kurtosis metrics, but in this work most results on WM kurtosis are about $K_\perp$ and $K_\parallel$ and are hence only valid in voxels with one major fiber direction.

5.1.1.2 **Model simplicity**: As shown in the latter section, many parameters can be linked to the kurtosis metrics. These relations are derived by using models and simulations, which treats water molecules movements as drunken man walks in a hindered environment. With this approach, someone with a minimal knowledge of anatomy and physics, and virtually without any knowledge about chemistry, can infer much information about the diffusion behavior of the chemical compound dihydrogen monoxide, i.e. water, in tissue microstructure. However, one of the pioneers of diffusion MRI, Dennis Le Bihan, points out that cell water largely differs from bulk water and is not just a structureless, space-filling background medium, but a chemical that is not homogeneous at the nanoscopic level and that makes all kind of interactions with the environment. Therefore the non-Gaussian diffusion in brain tissue could well result from strong interactions between water, proteins, phospholipids, etc within the cytoplasm and at the interface with membranes LeBihan (2007) [48]. It could therefore be that the models reviewed in this work are too simple. I believe, however, that at a micro scale all the special properties of water may play a role, but on a macro scale one can simplify the diffusion process of water as a hindered random walk, regarding all the promising validation studies done on some of the reviewed models.

5.1.1.3 **Does kurtosis really measure the degree of non-Gaussianity?**: It is widely stated in the literature [5] [2] [70] [79] [80] that a kurtosis of < 0 is more sharply peaked than a Gaussian distribution, that kurtosis is a measure of deviation from a Gaussian distribution, or a measure of peakedness. All

WM		
Lower $K_\perp$	• Higher EAS tortuousity	Tensor & stick [26]
	• Lower axonal fraction	Tensor & stick [26]
	• Lower intrinsic axonal diffusivity	DeSantis (2012)/CHARMED [70]
	• "Higher mean axon diameter"	biTensor [26]
	• "More IAS EAS exchange"	Kärger model [50]
	• "Axon undulation"	Simulations [78]
	• "Axon bending"	Simulations [77]
	• "More axon orientation dispersion"	NODDI [67]
	• "Less water that is bound to membranes"	MMWMD [64]
	• "Less glial cells"	MMWMD [64]
	• "Less CSF contamination"	MMWMD [64]
	• "Higher astrocyte fraction"	Ferizi (2013) [27]
Higher $K_\perp$	• Lower EAS tortuousity	Tensor & stick [26]
	• Higher axonal fraction	Tensor & stick [26]
	• Higher intrinsic axonal diffusivity	DeSantis (2012)/CHARMED [70]
	• "Lower mean axon diameter"	biTensor [26]
	• "Less IAS EAS exchange"	Kärger model [50]
	• "Axon 'straightening'"	Simulations [77]
	• "Less axon orientation dispersion"	NODDI [67]
	• "More water that is bound to membranes"	MMWMD [64]
	• "More glial cells"	MMWMD [64]
	• "More CSF contamination"	MMWMD [64]
	• "Lower astrocyte fraction"	Ferizi (2013) [27]
Lower $K_\parallel$	• Less diffusion heterogeneity between EAS and IAS	Tensor & stick [26]
	• "More exchange"	Kärger model [50]
	• "Less water that is bound to membranes"	MMWMD [64]
	• "Less glial cells (trapped water)"	MMWMD [64]
	• "Less CSF contamination"	MMWMD [64]
	• "Broken axons"	Simulations [77]
Higher $K_\parallel$	• More diffusion heterogeneity between EAS and IAS	Tensor & stick [26]
	• "Less exchange"	Kärger model [50]
	• "More water that is bound to membranes"	MMWMD [64]
	• "More glial cells" (trapped water)	MMWMD [64]
	• "More CSF contamination"	MMWMD [64]
	• "Bulging and crimping axons"	Simulations [77]

Table 5

Rules of thumb about possible causes of changing kurtosis metrics in WM, based on reviewed diffusion models and simulations. Relations between inverted commas "" are not literally stated in the literature, but are in this work derived from the literature.

GM		
Lower MK	• More exchange	Kärger model [26]
	• Less variance in tortuousity	Biexponential model [10]
	• 'Lower' orientation kurtosis tensor	Neurite orientation model [73]
	• "Less CSF contamination"	MMWMD [64]
	• "Less water that is bound to membranes"	MMWMD [64]
	• "Less glial cells (trapped water)"	MMWMD [64]
Higher MK	• Less exchange	Kärger model [26]
	• More variance in tortuousity	Biexponential model [10]
	• 'Higher' orientation kurtosis tensor	Neurite orientation model [73]
	• "More CSF contamination"	MMWMD [64]
	• "More water that is bound to membranes"	MMWMD [64]
	• "More glial cells (trapped water)"	MMWMD [64]

Table 6

Rules of thumb about possible causes of changing kurtosis metrics in GM, based on reviewed diffusion models and simulations. Relations between inverted commas "" are not literally stated in the literature, but are in this work derived from the literature.

	$K(\vec{n})(t)$	$K_\perp$	$K_\parallel$	Ref
Restricted diffusion	-3/7			[2]
Biexponential	$3 * \frac{f_{slow}*(1-f_{slow})*(D_{slow}-D_{fast})^2}{D_{app}^2}$			[2]
Kärger model	$3 * \frac{Var[D]}{\bar{D}^2} \frac{2}{t}\left[1 - \frac{1}{t}(1-e^{-\bar{t}})\right]$			[2]
Statistical truncated-Gaussian	$\frac{3}{\bar{D}^2}(\sigma^2 - \bar{D}^2 + \bar{D}*ADC)$			[15]
Statistical log-normal[1]	$3(e^{\sigma^2}-1)$			(55)
Statistical gamma	$\frac{3}{\alpha}$			[2]
Statistical beta[1]	$3 * \frac{\nu}{\mu(\nu+\mu+1)}$			(54)
Statistical Wishart	$\frac{3}{p}$			[2]
Stretched exponential[1]	undefined			[5]
Tensor & stick		$3 * \frac{v}{1-v}$		[26]
CHARMED		$\sim D_{e,\parallel}, \sim -R, \sim v \sim t$		[70]
Neurite orientation[3]	$\frac{[vD_a^2 \boldsymbol{W}+(1-v)*v(\boldsymbol{D}-\boldsymbol{I}D_T-D_A\boldsymbol{T})]^{\otimes 2}}{Tr(\boldsymbol{D_{app}}/3)^2}$			[74]
Isotropic neurite orientation	$3 * \frac{(1-v)*D_{eff}^2+v*\left(D_T^2+\frac{2*D_T(D_L-D_T)}{3}+\frac{(D_L-D_T)^2}{5}\right)}{((1-v)*D_{eff}+v*(\frac{2}{3}D_T+\frac{1}{3}D_L))^2} - 3$			[76]
Bulged axon		-0.2 ↓[2]	1.4 ↑[2]	[77]
Crimped axon		-0.4 ↑[2]	0.1 ↑[2]	[77]
Broken axon		-0.5[2]	-0.4 ↓[2]	[77]

Table 7

Kurtosis metrics implied by various diffusion models. 1. Obtained by the author. 2. Kurtosis of the IAS 3. Kurtosis tensor
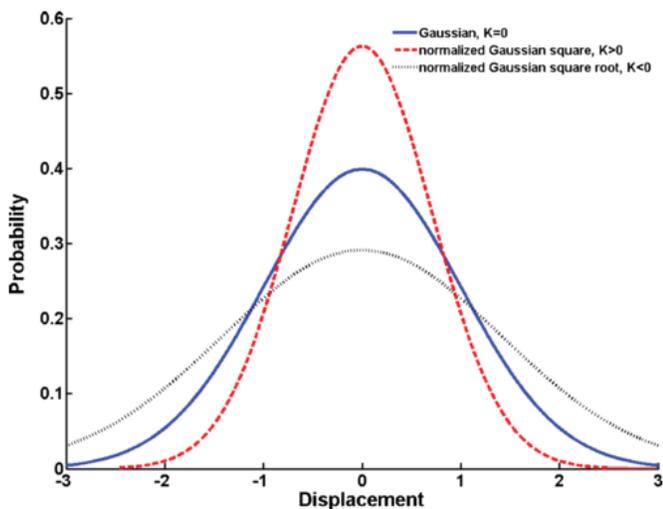
Figure 27. Graph of a Gaussian, a less peaked ($K < 0$) and a more peaked ($K > 0$) distribution [79]. When distributions get more complicated, the relation between kurtosis and non-Gaussianity becomes less straightforward.

this does not have to be necessarily true; Balanda (1988) [81] wrote a review about kurtosis showing an example of a non-Gaussian distribution which has kurtosis 0 whereas being more peaked than a Gaussian. The conclusion of the review was: "It is best to define kurtosis vaguely as the location- and scale-free movement of probability mass from the shoulders of a distribution into its center and tails." Figures appearing in the DKI literature showing a Gaussian distribution accompanied with similar but less and more peaked distributions which have a negative and positive kurtosis, such as figure 27, are in a sense misleading, because distributions can have other shapes than a bell curve. They can have for example finite support, or be multimodal. In appendix B we will give a physical more or less plausible situation which has a kurtosis of zero but a non-Gaussian distribution. However, it seems that the distributions one encounters in diffusion imaging experiments are bell curve like and are indeed more peaked and tailed whereas their kurtosis is > 0. Therefore the statement "kurtosis is a measure of tailedness and peakedness in comparison to a Gaussian" may be a good explanation for what kurtosis measures in DKI experiments.

6 FUTURE WORK

After we have completed the literature study on the causes of kurtosis in the human brain, I would like to present some ideas for further original, non-literature based, work which I gained from this study on non-Gaussian diffusion and diffusion kurtosis. These ideas are partitioned in the following three sections

- Modeling of axonal exchange.
- Use of the partial volume kurtosis formula.
- The cumulant expansion.

6.1 Modeling of axonal exchange

The microstructural models that are compared by Ferizi (2013) [27] model the radial part of axons as being impermeable and hence neglect exchange. This should be justified by the fact that the myelin sheath makes the axon permeability very low. However, Nilsson (2013) [22] performed a literature study on the role of exchange in WM, and concluded that exchange should be included in WM models. I performed a simulation to asses the necessity of an exchange component in diffusion models at very low axon permeabilities from a kurtosis perspective. Motivated by the outcomes of this simulation I will propose a modification to existing WM models.

6.1.1.1 **Monte Carlo simulation** : Fieremans (2010) [50] performed a simulation on the RK of WM when axonal exchange was allowed. However, the highest axonal exchange time τ used was 20 ms. I performed a similar simulation[5], but limited to 1D, on $K_\perp$ of WM using more plausible, higher axonal τ values ranging from $500 - 2500$ ms. This gave a few promising preliminary results, shown in figure 28, which could be investigated further:

- For the range of τ used, $K_\perp$ was increasing with t_d, whereas in the simulation of Fieremans (2010) [50], $K_\perp$ decreases with t_d.
- Even very low axon permeabilities had an effect on $K_\perp$ (and hence on $S_\perp$). Hence, a low axon permeability might be anything but negligible. This supports the claim of Nilsson (2013) [22] to include exchange in WM models.
- By allowing limited exchange, only few particles pass the membrane, whereas the bulk of them stays in the axon, resulting in a highly peaked and tailed distribution. This simulation therefore reported an effect on the axon radial kurtosis $K_{a,\perp}$ up to 25 due to axonal exchange, see figure 29. This is a major effect on $K_{a,\perp}$, for example in comparison with the effect on $K_{a,\perp}$ reported in Landman (2010) [77] due to axon damage. However, one has to acknowledge that $K_{a,\perp}$ gets weighted by $D_{a,\perp}^2$ in the voxel RK, which dampens the effect of $K_{a,\perp}$ on the voxel RK.

6.1.1.2 **Hypothesis about the axonal exchange time**: It is not yet clarified what the axonal exchange time is, other than somewhere in the range of $300 - 2500$ ms [22]. According to Nilsson (2013) [22], in most studies $K_\perp$ is found to be independent of diffusion time t_d. In the simulation that I performed, $K_\perp$ was almost independent of t_d for $\tau = 500$ ms, but was increasing with t_d for higher exchange times. This leads to an interesting hypothesis. If τ is (much) below 500 ms, according to the Kärger model and simulation of Fieremans (2010) [50] the kurtosis decreases with t_d. If τ is (much) higher than 500 ms, $K_\perp$ increases with t_d. It would hence follow logically that the axonal exchange time must be in the order of 500 ms.

6.1.1.3 **Modeling axon permeability analytically** : To account for exchange in WM models, I would add an extra compartment to WM models, which represents the particles that pass the axon membrane during the diffusion pulse sequence. Because myelinated axons have a non-zero but very low permeability, and the diffusion time is not very long in MR experiments, the amount of particles that cross the membrane more than once is arguably negligible and the rate of exchange over time can be assumed to be constant. For these reasons a

5. I used some different parameters than Fieremans (2010) [50], i.e. $v = 0.7$, $R = 2.5$ μm, $D_a = 0.46$ $\mu m^2/ms$ and $D_e = 0.99$ $\mu m^2/ms$. The qualitative results should, however, be generally valid [50].
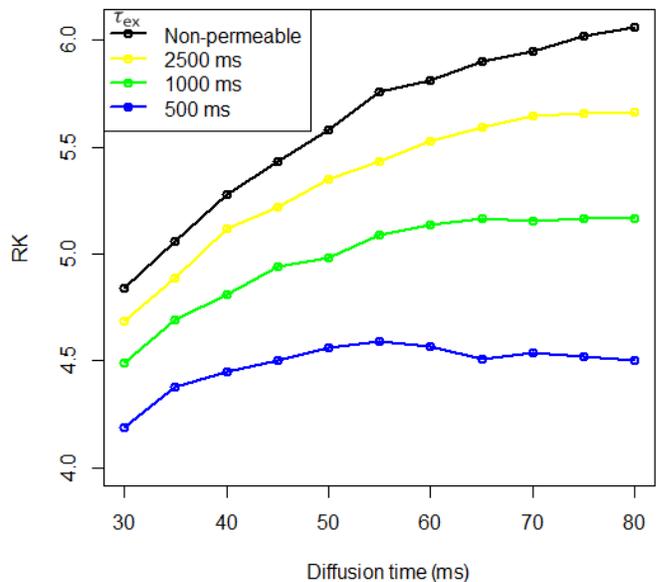
Figure 28. Radial voxel kurtosis plotted against diffusion time for several exchange times.
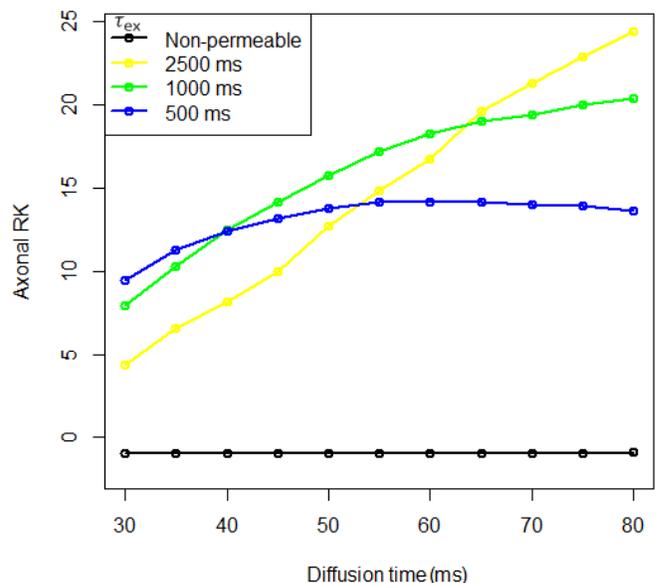
Figure 29. Radial kurtosis of the axon compartment plotted against diffusion time for several exchange times.

differential equation, which is proposed by Meier [52], should not be necessary to model exchange in WM.

Let us model the IAS by a permeable stick. One can observe that the diffusivity of an individual particle that starts in the IAS is $D * p * t_d$, where p is the proportion of time that the particle resides in the EAS. Because it is assumed that the exchange rate is constant over time, and that particles pass the axonal membrane not more than once, $p = t_{pass}/t_d$ for particles originating from the EAS, and $p = (t_{pass} - t_d)/t_d$ for particles originating from the IAS, where t_{pass} is the time when the particle makes the 'jump' from between the IAS and the

EAS. Therefore an uniform distribution of Gaussian tensors can be used to model the particles that cross the axon membrane during the pulse sequence, where the distribution $P(\boldsymbol{D})$ is essentially the distribution of t_{pass} that is uniform because we assumed a constant rate of exchange. This gives the following compartmental signal intensity

$$\frac{S_{ex}(\vec{n}, b)}{S_0} = \int_{t=0}^{t_d} \xi e^{-b\vec{n}^\top \boldsymbol{D}_e \vec{n} \frac{t}{t_d}} dt \qquad (47)$$

where ξ is the fraction of water particles that cross the axon membrane per millisecond and $\boldsymbol{D}_e$ is the EAS diffusion tensor. Notice that this expression incorporates particles that start in the IAS and end in the EAS, but also the other way around, since both result in Gaussian diffusion with limited diffusivity. The exchange tensor distribution could be easily extended to account for orientation dispersion. In this case, for a given t_{pass}, one considers the random variable

$$R = \sqrt{t_{pass}} X cos\Theta + \sqrt{t_d - t_{pass}} Y \qquad (48)$$

where X and Y are standard normal distributed, and Θ is distributed according to some orientation distribution. R is then Gaussian distributed, but for every combination of t_{pass} and Θ, it has a different effective diffusivity. This gives a signal intensity of

$$\int_\theta \int_{t=0}^{t_d} P(\theta) e^{-b(D_a t cos\theta + D_e(t_d - t))} dt d\theta. \qquad (49)$$

6.1.1.4 Relation to dot and sphere compartments: If we assume the exchange compartment from the previous section to be isotropic and if we evaluate the integral in equation (47) we get the signal expression

$$\frac{S(b)}{S(0)} = \int_0^{t_d} \xi e^{-bD_e * t/t_d} dt, \qquad (50)$$

we get the signal expression

$$\frac{S(b)}{S(0)} = \frac{t_d \xi}{bD_e} (1 - e^{-bD_e}). \qquad (51)$$

Hence the extra exchange compartment that I just proposed, is actually a dot compartment subtracted by a sphere compartment. This could explain the remarkable shift in preference of the BIC criterion between the dot model and the sphere model after a change of orientation dispersion in the work of Ferizi (2013) [66].

6.2 Use of the partial volume kurtosis formula.

The partial volume kurtosis formula (equation 23) is not yet apparent in the diffusion literature. For two compartments, this formula yields

$$K_{app} = \frac{3 * v(1-v)(D_1 - D_2)^2 + v * D_1^2 K_1 + (1-v) * D_2^2 K_2}{D_{app}^2} \qquad (52)$$

and hence exactly tells what happens with the kurtosis when an extra compartment is added to a voxel. This is useful when for example the effect of a CSF or astrocytes compartment is studied. It reveals also that the effect of the radial kurtosis on the voxel kurtosis is generally negligible. Combined with the argument that axons are arguably the only really non-Gaussian compartments in a voxel, it justifies the statistical model to assume that a voxel is composed of a variety of *Gaussian* compartments. Furthermore the formula shows that

generally the shape of the distribution of radii has no effect on the voxel kurtosis.

Future work should incorporate and exploit this formula in investigating the effects of diffusion compartmentalization on the kurtosis metrics.

6.3 Cumulant expansion

The cumulant expansion is the mathematical 'trick' to infer the DTI and DKI parameters from a model or an empirical MRI signal. In this section I will discuss some topics where the cumulant expansion plays a central role in further investigations on non-Gaussian diffusion.

6.3.1.1 A general relation between the 'statistical' model and the kurtosis: The kurtosis of the gamma, log-normal and beta 'statistical' model can be derived by a cumulant expansion to equal (see appendix E)

$$K_{gamma} = \frac{3}{\alpha} \qquad D_{gamma} = \frac{\alpha}{\beta} \qquad (53)$$

$$K_{beta} = 3 * \frac{\nu}{\mu(\nu + \mu + 1)} \qquad D_{beta} = \frac{\nu}{\mu + \nu} \qquad (54)$$

$$K_{lognormal} = 3 * (e^{\sigma^2} - 1) \qquad D_{lognormal} = e^{\mu + \sigma^2} \qquad (55)$$

For example, one can easily derive that in the gamma model $\alpha = \frac{3}{K}$ and $\beta = \frac{3}{DK}$. We see hence that the parameters of these 3 statistical models are functionally related to the apparent kurtosis and diffusivity. Therefore one should be cautious that the parameters of the 'statistical' model are not superfluous to the DKI metrics, as in figure 30. This was also noted by Jensen (2010) [5], but only for the gamma model at small b-values. Future work should further investigate this possible false dichotomy between DKI and 'statistical' parameters.

6.3.1.2 Interpretation of higher order cumulants: The signal intensity of the 'statistical' model is $E[e^{-bD}]$, which is a moment generating function of D, i.e. a function which derivatives in b at zero are the moments[6] of D. Therefore the 'statistical' model gives an interpretation of the higher order cumulants κ_6 and κ_8 as being the skewness and 'kurtosis' of the distribution of diffusivities $P(D)$. When the SNR would allow it, future work could explore the use of these parameters as pathological biomarkers on top of DKI and DTI metrics.

6.3.1.3 Cumulant expansion of diffusion models: If axons are directionally distributed over the sphere, I derived in appendix K that $K_{a\parallel}$ gets multiplied by the 'kurtosis' of the orientation distribution function

$$K_{a,\parallel}(\vec{n}) = \frac{E[X_{a\parallel}^4] E[(\vec{\nu} \cdot \vec{n})^4]}{E[X_{a\parallel}^2]^2 E[(\vec{\nu} \cdot \vec{n})^2]^2} - 3 \qquad (56)$$

where $\vec{\nu}$ and $\vec{n}$ are the orientation of the axon resp. diffusion gradient. This 'kurtosis' of the orientation distribution function

$$\frac{E[(\vec{\nu} \cdot \vec{n})^4]}{E[(\vec{\nu} \cdot \vec{n})^2]^2} \qquad (57)$$

could be investigated further. How should the axons be distributed on the sphere to give a maximal or minimal (mean) kurtosis, and what are these extreme values like? Does this spherical 'kurtosis' increase or decrease the voxel kurtosis, and for which kind of distributions does this happen? And how is the dispersion parameter κ of the Watson distribution related to this?

6. Up to a minus sign.

The fourth moment of the orientation distribution $E[(\vec{\mu} \cdot \vec{n})^4]$ can be split in the fourth moments of the spherical harmonics basis functions Y_{lm}

$$E_H[(\vec{\mu} \cdot \vec{n})^4] = \sum f_{lm} E_{Y_{lm}}[(\vec{\mu} \cdot \vec{n})^4]. \tag{58}$$

Hence the 'kurtosis' of an orientation function can be expanded in the 'kurtosis' of the spherical harmonics. Therefore it would be interesting to derive the 4-th moment $E[(\vec{\mu} \cdot \vec{n})^4]$ of some of the spherical harmonic basis functions.

Similar to the expression that Fieremans (2011) [26] has derived for the radial kurtosis of a tensor & stick model of

$$K_\perp = 3 * \frac{v}{1-v}, \tag{59}$$

one could also derive insightful expressions for radial kurtoses of other models in future work, such as a biZeppelin model with the restrictions $D_{e,\perp} = (1-v)D_{e,\parallel}$ and $D_{a,\parallel} = D_{e,\parallel}$. I derived for example, that when $v \approx 0.5$ (see appendix I)

$$K_\perp \approx 3 * \frac{v}{1-v} \left(1 - \frac{1}{(1-v)^2} \frac{D_{a\perp}}{D_{a\parallel}}\right)^2 \tag{60}$$

Furthermore, the (radial kurtosis) of the WM exchange model of Meier (2003) [52] and of the CHARMED model of Assaf (2004) [68] are interesting to derive analytically in future work, to see how those kurtoses depend on the diffusion time.

6.3.1.4 **Model comparison criteria** : Suppose diffusion within a voxel is non-Gaussian but isotropic, and free of noise. At very low b-values, $log[S(b)/S(0)]$ is almost linear, and therefore already 2 parameters may cause overfitting. At some higher b-values, $log[S(b)/S(0)]$ is almost quadratic, and therefore 3 or more parameters may cause overfitting. It may for example be investigated that the biexponential model fits the measured signal better than the quadratic cumulant expansion, but as [82] noted the biexponential model can be casted into the quadratic cumulant expansion as

$$logS = -b\bar{D} + \frac{1}{2}b^2 \left[p_1(D_1 - \bar{D})^2 + p_2(D_2 - \bar{D})^2\right] + O(b^3). \tag{61}$$

One can thus not draw conclusions about how good an analytic model explains the signal by not using b-values beyond the range where a cumulant expansion of that model with equal or less parameters fits the data equally well [82]. Criteria that quantify the tradeoff between goodness of fit and complexity should therefore acknowledge the specific signal intensity function arising from diffusion MRI experiments, which is approximately a polynomial of a certain degree, depending on the range of b-values. Future work could, possibly heuristically, tailor criteria such as the AIC and BIC to penalize for example the amount of parameters in comparison to the minimal amount of parameters that a cumulant expansion polynomial needs to fit the same signal curve well.

6.3.1.5 **Convergence of the cumulant expansion**: This last piece of future work is a bit more technical. In this section let $q = \gamma \delta g \sqrt{t}$, such that $b = q^2$. The cumulant expansion of a 1D random diffusion displacement X, is in fact a Taylor series. This Taylor series only converges to the actual function within a certain range of b-values, bounded by the radius of convergence (which can be infinite). In [82], an example is given of a function $log(2 - exp(-x))$ for which the radius of convergence is finite, whereafter it is derived that by assuming a certain amount of noise the cumulant expansion of a biexponential model only converges for b-values such that $b\bar{D} < 7.44$.
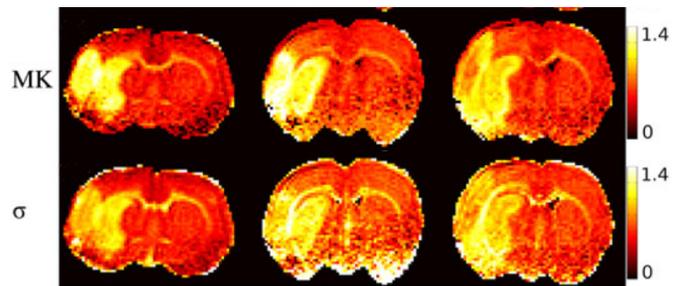
Figure 30. The σ parameter of the log normal 'statistical' model and the MK of the DKI model give very similar contrast in ischaemic lesions in three animals. Since MK is a function of σ 55, this may not be surprising. [55]
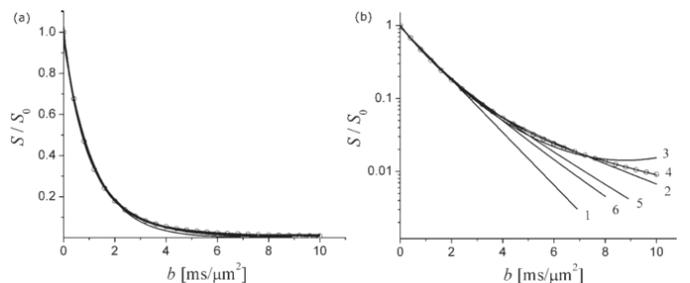
Figure 31. Fit of different models to a signal generated by the statistical model on a (a) normal scale, (b) log scale, (1) mono-exponential model, (2) biexponential model 3.1, (3) cumulant expansion truncated at the first three terms, (4) cumulant expansion truncated at four first terms, (5) cylindrical model [15], (6) stretched-exponential model 3.4 [15]. At low b-values, these signal curves are hardly distinguishable.

The problem is essentially on the level of high school mathematics: if $E[e^{-iqX}]$ is zero, the log of $E[e^{-iqX}]$ cannot be taken. However, $E[e^{-iqX}]$ has an infinite radius of convergence (see appendix L). One could make the Taylor expansion of $E[e^{-iqX}]$ directly, without taking the log first. One can obtain in this way the second and fourth order moments, whereafter the second and fourth order cumulants can be obtained by equation (8). By this moment expansion, one could possibly obtain the kurtosis metrics even outside the supposed radius of convergence of the cumulant expansion.

7 CONCLUSION

In summary, this work demonstrate that it is possible to derive from the literature a list of causes of changing kurtosis metrics due to microstructural changes beyond tissue heterogeneity and complexity, based on various diffusion models. This work shows that deriving kurtosis metrics from diffusion models shows promising insights in the microstructural basis of kurtosis. However, not all possible relations between models and kurtosis metrics are investigated yet. Future studies should therefore not only focus on improving and validating diffusion models, but should also examine relations between these diffusion models and kurtosis metrics.

ACKNOWLEDGMENT

ABBREVIATIONS

Medical abbreviations

WM	White matter
GM	Gray matter
IAS	Intra-axonal space
EAS	Extra-axonal space
CSF	Cerebrospinal fluid
CC	Corpus callosum
MS	Multiple sclerosis
MCI	Mild cognitive impairment

Biophysical parameters

R	Axon radius
v	Neurite fraction
D_e	Extra-axonal diffusivity
D_a	Intra-axonal diffusivity
D_{free}	Diffusivity in free water
τ	Mean exchange time
$\vec{\mu}$	Main fiber orientation

MRI symbols

γ	Gyromagnetic ratio of a proton (267.51 rad/s)
$\vec{g}$	Gradient vector
$\vec{n}$	Gradient direction
Δ	Time between the onsets of the two gradients
δ	Time that the diffusion weighting gradient is applied
t_d	Diffusion time $(\Delta - \frac{\delta}{3})$
b	B-value $(\delta^2 \|\vec{g}\|^2 t_d)$
Φ	Phase of a particle
$M(t)$	Isochromat of spins at time t [14]
S	Voxel signal intensity
$S_{M(t)}$	Signal intensity of an isochromat $M(t)$ at time t
$S(0)$	Voxel signal intensity when $b = 0$.
$S(b)$	Voxel isotropic signal intensity when b is applied
$S(\vec{n}, b)$	Voxel intensity when b is applied in direction $\vec{n}$
$\parallel$	Parallel to the main diffusion direction
$\perp$	Perpendicular to the main diffusion direction

Mathematical symbols

1D	One dimensional
$\|\vec{v}\|$	Eucledian length of a vector $\vec{v}$
$\| \cdot \|$	Determinant of a matrix
tr	Trace of a matrix
$\top$	Transpose
$\boldsymbol{M}$	Tensor
$\otimes$	Tensor product
S^2	Sphere
$O(x^k)$	Error term of order k [83]

Probability theoretic abbreviations and symbols

IID	Independent and identically distributed
pdf	Probability density function
$E[X]$	Expected value
$\vec{R}_t$	Molecule position at time t, $R_t = (X_t, Y_t, Z_t)^{\top}$
$\vec{R}(t)$	Idem
$f_{\vec{R}_t}(p)$	Pdf of $\vec{R}_t$ at point p
$Var[X]$	Variance
$X \sim N(\mu, \sigma^2)$	X follows a Gaussian with mean μ and sd σ
σ	Standard deviation (symbol)
$\kappa_2(X)$	Second cumulant of a random variable X
$\kappa_4(X)$	Fourth cumulant of a random variable X

Diffusion MRI terminology, abbreviations and symbols

DKI	Diffusion Kurtosis Imaging
DTI	Diffusion Tensor Imaging
λ	Eigenvalue of the diffusion tensor
D	Diffusivity (when diffusion is isotropic)
D_{app}	Apparent diffusivity
MD	Mean diffusion (abbreviation)
$\bar{D}$	Mean diffusion (symbol)
$D(\vec{n})$	Diffusivity in direction $\vec{n}$
$\boldsymbol{D}$	Diffusion tensor
FA	Fractional Anisotropy
K	Kurtosis (when diffusion is isotropic)
K_{app}	Apparent kurtosis
MK	Mean Kurtosis (abbreviation)
$\bar{K}$	Mean Kurtosis (symbol)
$K(\vec{n})$	Kurtosis in direction $\vec{n}$
$\boldsymbol{W}$	Kurtosis tensor

REFERENCES

[1] Jespersen SN, Bjarkam CR, Nyengaard JR, Chakravarty MM, Hansen B, Vosegaard T, et al., 2010. Neurite density from magnetic resonance diffusion measurements at ultrahigh field: comparison with light microscopy and electron microscopy. *NeuroImage*, 49(1):205–16.

[2] **Jensen JH, Helpern J, Ramani A, Lu H, Kaczynski K, 2005. Diffusional kurtosis imaging: the quantification of non-Gaussian water diffusion by means of magnetic resonance imaging. *Magnetic Resonance in Medicine*, 53(6):1432–40.

[3] Steven AJ, Zhuo J, Melhem ER, 2014. Diffusion kurtosis imaging: an emerging technique for evaluating the microstructural environment of the brain. *AJR. American journal of roentgenology*, 202(1):W26–33.

[4] Bar-Shir A, Duncan ID, Cohen Y, 2009. QSI and DTI of excised brains of the myelin-deficient rat. *NeuroImage*, 48(1):109–16.

[5] **Jensen JH, Helpern JA, 2010. MRI quantification of non-Gaussian water diffusion by kurtosis analysis. *NMR in Biomedicine*, 23(7):698–710.

[6] Yang AW, Jensen JH, Hu CC, Tabesh A, Falangola MF, Helpern Ja, 2013. Effect of cerebral spinal fluid suppression for diffusional kurtosis imaging. *Journal of Magnetic Resonance Imaging*, 37(2):365–71.

[7] Zhuo J, Xu S, Proctor JL, Mullins RJ, Simon JZ, Fiskum G, et al., 2012. Diffusion kurtosis as an in vivo imaging marker for reactive astrogliosis in traumatic brain injury. *NeuroImage*, 59(1):467–477.

[8] Engelhardt M, Bain LJ, 1987. *Introduction to Probability and Mathematical Statistics*. Duxbury Press.

[9] **Veraart J, 2013. Optimal estimation of diffusion MRI parameters. Phd thesis, University of Antwerp, Belgium.

[10] Niendorf T, Dijkhuizen RM, Norris DG, van Lookeren Campagne M, Nicolay K, 1996. Biexponential diffusion attenuation in various states of brain tissue: implications for diffusion-weighted imaging. *Magnetic Resonance in Medicine*, 36(6):847–57.

[11] Laun FB, Fritzsche KH, Kuder TA, Stieltjes B, 2011. Einführung in die Grundlagen und Techniken der Diffusionsbildgebung. *Der Radiologe*, 51(3):170–179.

[12] Leemans A, 2006. Modeling and processing of diffusion tensor magnetic resonance images for improved analysis of brain connectivity. Ph.D. thesis, University of Antwerp, Belgium.

[13] Basser P, Mattiello J, LeBihan D, 1994. Estimation of the effective self-diffusion tensor from the NMR spin echo. *Journal of Magnetic Resonance, Series B*, 103(3):247–254.

[14] Haacke E, Brown R, Thomsom M, Venkatesan R, 1999. *Magnetic Resonance Imaging. Physical Principles and Sequence Design*. Wiley-Liss (John Wiley Sons), New York.

[15] **Yablonskiy D, Sukstanskii A, 2010. Theoretical models of the diffusion weighted MR signal. *NMR in Biomedicine*, 23(7):661–681.

[16] Kristoffersen A, 2011. Statistical assessment of non-Gaussian diffusion models. *Magnetic Resonance in Medicine*, 66(6):1639–48.

[17] Lätt J, Nilsson M, Wirestam R, Ståhlberg F, Karlsson N, Johansson M, et al., 2013. Regional values of diffusional kurtosis estimates in the healthy brain. *Journal of Magnetic Resonance Imaging*, 37(3):610–8.

[18] Qi L, Wang Y, Wu EX, 2008. D-Eigenvalues of Diffusion Kurtosis Tensors. *Journal of Computational and Applied Mathematics*, 221(1):150–157.

[19] Hui ES, Cheung MM, Qi L, Wu EX, Theory A, 2008. Towards better MR characterization of neural tissues using directional diffusion kurtosis analysis. *NeuroImage*, 42(1):122–34.

[20] Aljabar P, Bhatia KK, Murgasova M, Hajnal JV, Boardman JP, Srinivasan L, et al., 2008. Assessment of brain growth in early childhood using deformation-based morphometry. *NeuroImage*, 39(1):348–58.

[21] Fields R, 2008. White matter matters. *Scientific American*, 298(3):54–61.

[22] Nilsson M, Westen DV, 2013. The role of tissue microstructure and water exchange in biophysical modelling of diffusion in white matter. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 26(4):345–370.

[23] Beaulieu C, 2002. The basis of anisotropic water diffusion in the nervous system - a technical review. *NMR in Biomedicine*, 15(7-8):435–55.

[24] Sherman D, Brophy P, 2005. Mechanisms of axon ensheathment and myelin growth. *Nature Reviews Neuroscience*, 6(9):683–690.

[25] Freeman S, 2005. *Biological Science and CW+ Grade Tracker Acces Card 2/E*. Benjamin Cummings.

[26] **Fieremans E, Jensen JH, Helpern JA, 2011. White matter characterization with diffusional kurtosis imaging. *NeuroImage*, 58(1):177–188.

[27] **Ferizi U, Schneider T, Panagiotaki E, Nedjati-Gilani G, Zhang H, Wheeler-Kingshott C, et al., 2013. A ranking of diffusion MRI compartment models with in vivo human brain data. *Magnetic Resonance in Medicine*.

[28] Allen NJ, Barres BA, 2009. Neuroscience: Glia - more than just brain glue. *Nature*, 457(7230):675–677.

[29] Caserta F, Eldred WD, Fernandez E, Hausman RE, Stanford LR, Bulderev SV, et al., 1995. Determination of fractal dimension of physiologically characterized neurons in two and three dimensions. *Journal of Neuroscience Methods*, 56(2):133–44.

[30] Falangola MF, Jensen JH, Babb JS, Hu C, Castellanos FX, Di Martino A, et al., 2008. Age-related non-Gaussian diffusion patterns in the prefrontal brain. *Journal of Magnetic Resonance Imaging*, 28(6):1345–50.

[31] Helpern J, Falangola M, Martino AD, 2007. Alterations in brain microstructure in ADHD by diffusional kurtosis imaging. In *Proceedings of the 15th Annual Meeting of ISMRM*, 2006, p.1580, Berlin, Germany.

[32] Yoshida M, Hori M, Yokoyama K, Fukunaga I, Suzuki M, Kamagata K, et al., 2013. Diffusional kurtosis imaging of normal-appearing white matter in multiple sclerosis: preliminary clinical experience. *Japanese Journal of Radiology*, 31(1):50–5.

[33] Ramani A, Jensen J, Szulc K, Ali O, 2007. Assessment of abnormalities in the cerebral microstructure of schizophrenia patients: a diffusional kurtosis imaging study. In *Proceedings of the 15th Annual Meeting of ISMRM*, p.648, Berlin, Germany.

[34] Lu H, Jensen JH, Ramani A, Helpern J, 2006. Three-dimensional characterization of non-gaussian water diffusion in humans using diffusion kurtosis imaging. *NMR in Biomedicine*, 19(2):236–47.

[35] Gong NJ, Chan CC, Leung LM, Wong CS, 2013. Axonal loss or demyelination? Decreased integrity of white matter tracts revealed by diffusional kurtosis imaging in Alzheimer's disease and mild cognitive impairment. *Alzheimer's Dementia*, 9(4):88–89.

[36] Falangola MF, Jensen JH, Tabesh A, Hu C, Deardorff RL, Babb JS, et al., 2013. Non-Gaussian diffusion MRI assessment of brain microstructure in mild cognitive impairment and Alzheimer's disease. *Magnetic Resonance Imaging*, 31(6):840–6.

[37] Wang J, Lin W, Lu C, Weng Y, 2011. Parkinson disease: diagnostic utility of diffusion kurtosis imaging. *Radiology*, 261(1):210–17.

[38] Hui ES, Fieremans E, Jensen JH, Tabesh A, Feng W, Bonilha L, et al., 2012. Stroke assessment with diffusional kurtosis imaging. *Stroke*, 43(11):2968–73.

[39] Zhang Y, Yan X, Gao Y, Xu D, Wu J, Li Y, 2013. A preliminary study of epilepsy in children using diffusional kurtosis imaging. *Clinical Neuroradiology*, 23(4):293–300.

[40] Cauter SV, Veraart J, Sijbers J, 2012. Gliomas: diffusion kurtosis MR imaging in grading. *Radiology*, 263(2):492–501.

[41] Nakanishi A, Fukunaga I, Hori M, Masutani Y, Takaaki H, Miyajima M, et al., 2013. Microstructural changes of the corticospinal tract in idiopathic normal pressure hydrocephalus: a comparison of diffusion tensor and diffusional kurtosis imaging. *Neuroradiology*, 55:971–976.

[42] Blockx I, Verhoye M, Audekerke JV, Bergwerf I, Kane JX, Delgado R, et al., 2012. Identification and characterization of Huntington related pathology : An in vivo DKI imaging study. *NeuroImage*, 63(2):653–662.

[43] Szczepankiewicz F, Lätt J, Wirestam R, Leemans A, Sundgren P, van Westen D, et al., 2013. Variability in diffusion kurtosis imaging: impact on study design, statistical power and interpretation. *NeuroImage*, 76:145–54.

[44] Schultz T, Fuster A, Ghosh A, Deriche R, Florack L. Higher-Order Tensors in Diffusion Imaging. 1–34.

[45] **Assaf Y, Blumenfeld-Katzir T, Yovel Y, Basser PJ, 2008. AxCaliber: a method for measuring axon diameter distribution from diffusion MRI. *Magnetic Resonance in Medicine*, 59(6):1347–54.

[46] Pyatigorskaya N, Le Bihan D, Reynaud O, Ciobanu L, 2013. Relationship between the diffusion time and the diffusion MRI signal observed at 17.2 tesla in the healthy rat brain cortex. *Magnetic Resonance in Medicine*.

[47] Sehy JV, Ackerman JJH, Neil JJ, 2002. Evidence that both fast and slow water ADC components arise from intracellular space. *Magnetic Resonance in Medicine*, 48(5):765–70.

[48] Le Bihan D, 2007. The 'wet mind': water and functional neuroimaging. *Physics in Medicine and Biology*, 52(7):R57–R90.

[49] Jones DK, Knösche TR, Turner R, 2013. White matter integrity, fiber count, and other fallacies: the do's and don'ts of diffusion MRI. *NeuroImage*, 73:239–54.

[50] **Fieremans E, Novikov D, 2010. Monte Carlo study of a two compartment exchange model of diffusion. *NMR in Biomedicine*, 23(7):711–24.

[51] Pfeuffer J, Flo U, Dreher W, Leibfritz D, 1998. Restricted diffusion and exchange of intracellular water : theoretical modelling and diffusion time dependence of 1 H NMR measurements on perfused glial cells. 11:19–31.

[52] Meier C, Dreher W, Leibfritz D, 2003. Diffusion in compartmental systems. I. A comparison of an analytical model with simulations. *Magnetic Resonance in Medicine*, 50(3):500–9.

[53] Fukunaga I, Hori M, Masutani Y, Hamasaki N, Sato S, Suzuki Y, et al., 2013. Effects of diffusional kurtosis imaging parameters on diffusion quantification. *Radiological Physics and Technology*, 6(2):343–48.

[54] Yablonskiy DA, Bretthorst GL, Ackerman JJH, 2003. Statistical model for diffusion attenuated MR signal. *Magnetic Resonance in Medicine*, 50(4):664–9.

[55] Grinberg F, Ciobanu L, Farrher E, Shah NJ, 2012. Diffusion kurtosis imaging and log-normal distribution function imaging enhance the visualisation of lesions in animal stroke models. *NMR in Biomedicine*, 25(11):1295–304.

[56] Jian B, Vemuri BC, Ozarslan E, Carney PR, Mareci TH, 2007. A novel tensor distribution model for the diffusion-weighted MR signal. *NeuroImage*, 37(1):164–76.

[57] Basser PJ, Pajevic S, 2003. A normal distribution for tensor-valued random variables: applications to diffusion tensor MRI. *IEEE transactions on medical imaging*, 22(7):785–94.

[58] Hall MG, Barrick TR, 2008. From diffusion-weighted MRI to anomalous diffusion imaging. *Magnetic Resonance in Medicine*, 59(3):447–55.

[59] Mandelbrot B, 1967. How long is the coast of britain? Statistical self-similarity and fractional dimension. *Science (New York, N.Y.)*, 156(3775):636–8.

[60] Rachev S, 2007. *Financial Econometrics: From Basics to Advanced Modeling Techniques*. John Wiley Sons, Ltd., New York.

[61] Bennett KM, Schmainda KM, Bennett RT, Rowe DB, Lu H, Hyde JS, 2003. Characterization of continuously distributed cortical water diffusion rates with a stretched-exponential model. *Magnetic Resonance in Medicine*, 50(4):727–34.

[62] Kwee TC, Galbán CJ, Tsien C, Junck L, Sundgren PC, Ivancevic MK, et al., 2010. Intravoxel water diffusion heterogeneity imaging of human high-grade gliomas. *NMR in Biomedicine*, 23(2):179–87.

[63] Blackledge J, Blackledge MD, 2011. Magnetic Resonance Image Processing using Levy. In *10th International Conference on Environment and Electrical Engineering EEEIC*, Rome.

[64] **Alexander DC, Hubbard PL, Hall MG, Moore E, Ptito M, Parker GJM, et al., 2010. Orientationally invariant indices of axon diameter and density from diffusion MRI. *NeuroImage*, 52(4):1374–89.

[65] Tuch DS, Reese TG, Wiegell MR, Makris N, Belliveau JW, Wedeen VJ, 2002. High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. *Magnetic Resonance in Medicine*, 48(4):577–82.

[66] Ferizi U, Schneider T, Panagiotaki E. The Importance of Being Dispersed: A Ranking of Diffusion MRI Models for Fibre Dispersion Using In Vivo Human Brain Data. In *Medical Image Computing and Computer-Assisted Intervention 2013*, 74–81, Springer Berlin Heidelberg.

[67] **Zhang H, Schneider T, Wheeler-Kingshott CA, Alexander DC, 2012. NODDI: practical in vivo neurite orientation dispersion and density imaging of the human brain. *NeuroImage*, 61(4):1000–16.

[68] Assaf Y, Freidlin R, 2004. New modeling and experimental framework to characterize hindered and restricted water diffusion in brain white matter. *Magnetic Resonance in Medicine*, 52(5):965–978.

[69] Neuman CH, 1974. Spin echo of spins diffusing in a bounded medium. *The Journal of Chemical Physics*, 60:4508–4511.

[70] **Santis SD, Assaf Y, Jones D, 2012. Using the biophysical CHARMED model to elucidate the underpinnings of contrast in diffusional kurtosis analysis of diffusion-weighted MRI. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 25(4):267–276.

[71] Clark CA, Hedehus M, Moseley ME, 2001. Diffusion time dependence of the apparent diffusion tensor in healthy human brain and white matter disease. *Magnetic Resonance in Medicine*, 45(6):1126–9.

[72] Assaf Y, Basser PJ, 2005. Composite hindered and restricted model of diffusion (CHARMED) MR imaging of the human brain. *NeuroImage*, 27(1):48–58.

[73] Jespersen SN, Kroenke CD, Ostergaard L, Ackerman JJH, Yablonskiy DA, 2007. Modeling dendrite density from magnetic resonance diffusion measurements. *NeuroImage*, 34(4):1473–86.

[74] **Jespersen S, Leigland L, 2012. Determination of axonal and dendritic orientation distributions within the developing cerebral cortex by diffusion tensor imaging. *IEEE transactions on medical imaging*, 31(1):16–32.

[75] Minati L, Weglarz W, 2007. Physical foundations, models, and methods of diffusion magnetic resonance imaging of the brain: A review. *Concepts in Magnetic Resonance Part A*, 30(5):278–307.

[76] **Mulkern RV, Haker SJ, Maier SE, 2009. On high b diffusion imaging in the human brain: ruminations and experimental insights. *Magnetic Resonance Imaging*, 27(8):1151–62.

[77] **Landman BA, Farrell JAD, Smith SA, Reich DS, Calabresi PA, van Zijl PCM, 2010. Complex geometric models of diffusion and relaxation in healthy and damaged white matter. *NMR in biomedicine*, 23(2):152–62.

[78] **Nilsson M, Lätt J, Ståhlberg F, van Westen D, Hagslätt H, 2012. The importance of axonal undulation in diffusion MR measurements: a Monte Carlo simulation study. *NMR in Biomedicine*, 25(5):795–805.

[79] Wu EX, Cheung MM, 2010. MR diffusion kurtosis imaging for neural tissue characterization. *NMR in Biomedicine*, 23(7):836–48.

[80] Lätt J, Nilsson M, Wirestam R, Johansson E, Larsson EM, Ståhlberg F, et al., 2008. In vivo visualization of displacement-distribution-derived parameters in q-space imaging. *Magnetic resonance imaging*, 26(1):77–87.

[81] Balanda KP, Macgillivray HL, 2013. Kurtosis : A Critical Review. *The American Statistician*, 42(2):111–119.

[82] Frøhlich AF, Ostergaard L, Kiselev VG, 2006. Effect of impermeable boundaries on diffusion-attenuated MR signal. *Journal of Magnetic Resonance*, 179(2):223–233.

[83] Hundsdorfer W, 2013. Ordinary Differential Equations. Course Notes.

[84] Whittle P, 2000. *Probability via expectation*. 6th edition, Springer, New York.

[85] Gallier J, 2009, Notes on Spherical Harmonics and Linear Representations of Lie Groups. Course Notes.

Key references are denoted with **

APPENDIX A

SOME BASIC CONCEPTS OF PROBABILITY AND STATISTICS

Suppose we denote the outcome of throwing a dice by X. Then X is a random number between 1 and 6, for which each outcome has an equal probability (assuming that we have a fair dice, of course). Therefore we call X, in all generality, a **random**
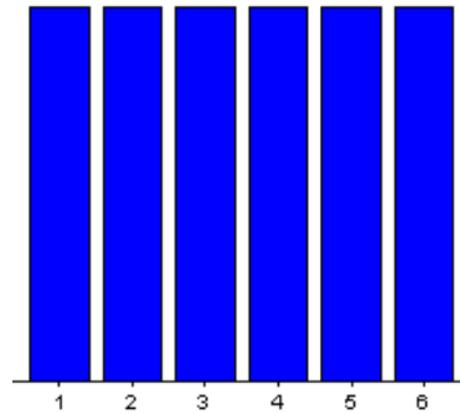
Figure 32. The histogram of X, i.e. throwing a dice once.
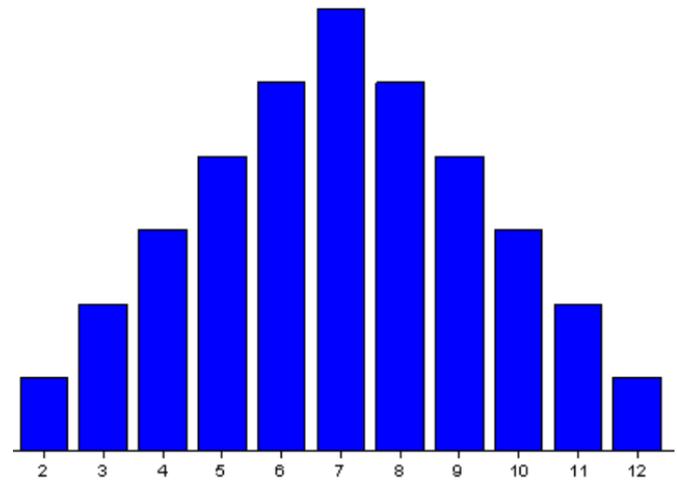
Figure 33. The histogram of $X + Y$, i.e. throwing a dice twice. One can appreciate that this is the convolution of the histogram of a single dice.

variable. Suppose we make a histogram of the outcomes of the experiments. This histogram would have 6 bins which are all equally filled, see figure 32. Suppose we throw another dice Y and we denote the outcome of the sum of the two dices by $X + Y$. Clearly, we throw minimal a 2 and maximal a 12. What are the chances of a particular outcome?

For a 2, we would need to have $X = 1$ and $Y = 1$.

For a 3, we would need to have $X = 1$ and $Y = 2$ or $X = 2$ and $Y = 1$.

For a 4, „ „ $X = 3$ and $Y = 1$, or $X = 2$ and $Y = 2$, or $X = 1$ and $Y = 3$.

......

For a 12 „ „ $X = 6$ and $Y = 6$.

When we draw a histogram of $X + Y$ in figure 33, we see that this is essentially a convolution of X and Y. We also see that $X + Y$ is more bell shaped than the flat histograms of X and Y. The more dices one throws, the more bell shaped the histogram of the sum will become (this effect is called the central limit theorem).

What is the mean value of throwing a dice, say 100000 times? This is of course 3.5 and is denoted by $E[X]$. This is called the **expected value** or **expectation value**, because this is the value

one could expect from the experiment. If we only knew the histogram of X we could also have just sum all the bins times their value, divided by the total amount of bins:

$$E[X] = \frac{\sum i * bin_X(i)}{\sum bin_X(i)}. \tag{62}$$

If we now smooth and normalize the histogram such that we get a continuous function f, which is called a **probability density function**, then

$$E[X] = \int_t f_X(t)t\,dt. \tag{63}$$

What is $E[X + Y]$? Of course we can look what we can expect from X and what from Y and add both up. Hence $E[X + Y]$ should be $E[X] + E[Y]$. This also follows from the fact that integrals can be split up, so formally

$$E[X + Y] = \int t(f_X(t) + f_Y(t)) \tag{64}$$

$$= \int tf_X(t) + \int tf_Y(t) \tag{65}$$

$$= E[X] + E[Y] \tag{66}$$

If we predict that 3.5 is the mean outcome of throwing, say 10 times, a dice, then we also would like to know how accurate this prediction is. Therefore we take the 'mean square error': $E[(E[X] - X)^2]$, this is also called the **variance** because it says how variable the outcomes of an experiment are. The square root of the variance is called the **standard deviation**. In histogram terms, this can be seen for now as the full width at half maximum. Another special expectation is the covariance, which is denoted by $E[XY] - E[X]E[Y]$. This can be seen as a measure how related the outcomes of two experiments are. For example if $X_1, \ldots, X_{1600}$ are pixel values in image X and $Y_1, \ldots, Y_{1600}$ are the corresponding pixel values in image Y, one can register the images by trying to reach a maximum **covariance** between the X_i and $T(Y_i)$ for a transformation T. We can generally apply any function h on X which would give an expectation of

$$E[h(X)] = \int_s f_X(s)h(s)\,ds.$$

Hence we can also ask our self what the outcome of for example cX, $cos(X)$, X^2 and e^{iXt} is. The expected value of the latter is actually the inverse Fourier transform of the probability density function f of X

$$E[e^{iXt}] = \int_s f_X(s)e^{its}\,ds.$$

We already know that convolution means multiplication in Fourier space, therefore we can investigate sums of random variables like $X + Y$ better in Fourier space. Suppose now for example that X and Y not dependent on each other and are standard normal distributed. Taken for granted that the **characteristic function** of a standard Gaussian distribution is just $e^{-\frac{1}{2}}$, it is easy to calculate that the characteristic function of $X + Y$, since

$$E[e^{i(X+Y)t}] = E[e^{iXt}]E[e^{iYt}] \tag{67}$$

is equal to

$$e^{-\frac{1}{2}}e^{-\frac{1}{2}} = e^{-1}. \tag{68}$$

Hence $X + Y$ is in this case also Gaussian distributed, but then more wider, one could say more blurred. This would be considerably more elaborate to calculate outside Fourier space.

APPENDIX B
EXAMPLE OF A DIFFUSION DISTRIBUTION THAT IS NOT A GAUSSIAN DISTRIBUTION BUT CAN HAVE KURTOSIS 0, > 0 OR < 0

Consider a voxel consisting of an extra-axonal and a intra-axonal part with a axonal fraction of $v = 0.82$. Assume that diffusion in the axonal part is restricted. Then the intra-axonal diffusion is after sufficient time nearly uniform distributed and could have therefore a radial kurtosis of -1.2 [2]. Suppose that because of the high axonal density the extra-axonal diffusivity is somewhat lower than normal, say 0.31 $\mu m^2/ms$. Suppose the intra-axonal radial diffusivity is 0.13 $\mu m^2/ms$. Then by the partial volume kurtosis formula (equation 23), the voxel radial kurtosis is equal to

$$\frac{3 * 0.82 * 0.18(0.31 - 0.13)^2 - 0.82 * 0.13^2 * 1.2}{(0.82 * 0.31 + 0.18 * 0.13)^2} = -0.09$$

Suppose now the extra-axonal diffusivity is 0.325 $\mu m^2/ms$. Then

$$\frac{3 * 0.82 * 0.18(0.325 - 0.13)^2 - 0.82 * 0.13^2 * 1.2}{(0.82 * 0.325 + 0.18 * 0.13)^2} = 0.002$$

For some extra-axonal diffusivity in between the kurtosis is zero, for a non-Gaussian diffusion distribution. These values are all stretched to their maximum or minimum level to make the kurtosis zero, but are on the other hand still not totally physical *unplausible*. A negative kurtosis is, however, to the best of my knowledge never observed in the human brain, probably because total restricted diffusion is not apparent in the human brain [48].

APPENDIX C
TENSOR PRODUCT

Suppose V and W are $n \times n$ matrices. Then we call the tensor product the product of all combinations of elements of V and W

$$(V \otimes W)_{ijkl} = V_{ij}W_{kl}$$

One can also take the tensor product of V with V, i.e.

$$(V \otimes V)_{ijkl} = V_{ij}V_{kl}$$

which can also be denoted by $V^{\otimes 2}$.

APPENDIX D
THE CHARACTERISTIC FUNCTION OF THE GAUSSIAN DISTRIBUTION 2

Lemma. *Suppose $X \sim \mathcal{N}(0, \sigma^2)$. Then $E[e^{-itX}] = e^{\frac{-\sigma^2}{2}t^2}$*

Proof.

$$E[e^{-ixb}] = \frac{1}{\sigma\sqrt{2\pi}} \int_\infty^\infty e^{-\frac{x^2}{2\sigma^2}} e^{-itx}\,dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-\frac{u^2}{2}} e^{-it\sigma u}\,du$$

$$= e^{-\frac{\sigma^2 b^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-\frac{v^2}{2}}\,dv$$

$$= e^{-\frac{\sigma^2 b^2}{2}}$$

Where the substitutions $u = \frac{x}{\sigma}$ and $v = u + it\sigma$ and the fact that $(u + i\sigma t)^2 = u^2 + 2it\sigma u - \sigma^2 t^2$ are used.

$\square$

APPENDIX E
HOW TO DERIVE THE KURTOSIS FROM A MODEL SIGNAL INTENSITY?

The kurtosis can be derived from the model signal intensity in two different ways:

1) Taylor expand the log of the model signal intensity $S(\vec{n}, b)/S(0)$. This leads to

$$log[\frac{S(\vec{n}, b)}{S(0)}] = -\frac{\kappa_2}{2}b + \frac{\kappa_4}{24}b^2 + O(b^3) \qquad (69)$$

.

2) Obtain κ_2 and κ_4
3) Derive the kurtosis in direction $\vec{n}$ by

$$K(\vec{n}) = \frac{\kappa_4}{\kappa_2^2} \qquad (70)$$

Moment expansion

1) Taylor expand the model signal intensity $S(\vec{n}, b)/S(0)$. This leads to

$$S(\vec{n}, b)/S(0) = 1 - \frac{1}{2}E[(\vec{R}\cdot\vec{n})^2]b + \frac{1}{24}E[(\vec{R}\cdot\vec{n})^4]b^2 + O(b^3) \qquad (71)$$

2) Obtain $E[(\vec{R}\cdot\vec{n})^2]$ and $E[(\vec{R}\cdot\vec{n})^4]$.
3) Derive the kurtosis in direction $\vec{n}$ by

$$K(\vec{n}) = \frac{E[(\vec{R}\cdot\vec{n})^4] - 3E[(\vec{R}\cdot\vec{n})^2]^2}{E[(\vec{R}\cdot\vec{n})^2]^2}, \qquad (72)$$

which can be simplified to

$$\frac{E[(\vec{R}\cdot\vec{n})^4]}{E[(\vec{R}\cdot\vec{n})^2]^2} - 3 \qquad (73)$$

In general generally both methods are equivalent, although the moment expansion is easier to make but gives less insightful kurtosis expressions.

APPENDIX F
KURTOSIS OF A NON-GAUSSIAN ISOTROPIC MIXTURE

Consider a mixture having n compartments with relative fractions, diffusivities and kurtosis $p_1, \ldots, p_n$, D_1, …, D_n, $K_1, \ldots, K_n$. We already know how to compute the kurtosis of a Gaussian mixture from formula 25. Therefore we will only look at which part should be added to that formula.
We have a signal intensity of

$$\frac{S(b)}{S(0)} = c * b + \frac{1}{24}\sum_{i=1}^{n}p_i\kappa_4^i b^2 + O(b^3) \qquad (74)$$

Hence we have that, by using $log[x] = x - \frac{x^2}{2} + O(x^3)$

$$log[\frac{S(b)}{S(0)}] = c * b + \frac{1}{24}\sum_{i=1}^{n}p_i\kappa_4^i b^2 + e * b^2 + O(b^3) \qquad (75)$$

Where c and e are constants independent of K_i.
Because

$$\kappa_4^i = (\kappa_2^i)^2 * K_i = 4D_i^2 K_i,$$

we have

$$log[\frac{S(b)}{S(0)}] = c * b + \frac{1}{24}\sum_{i=1}^{n}4p_i D_i^2 K_i b^2 + e * b^2 + O(b^3),$$

therefore

$$\frac{\sum_{i=1}^{n}p_i\kappa_4^i}{\kappa_2} = \frac{\sum_{i=1}^{n}4p_i D_i^2 K_i}{4\bar{D}^2}$$

$$= \frac{\sum_{i=1}^{n}p_i D_i^2 K_i}{\bar{D}^2}$$

hence the part that should be added to the formula 25 is

$$\frac{\sum_{i=1}^{n}p_i D_i^2 K_i}{\bar{D}^2}$$

where $\bar{D} = \sum_{i=1}^{n}p_i D_i$.
Hence the voxel kurtosis is equal to

$$K = \frac{3 * Var[D] + \sum_{i=1}^{n}p_i D_i^2 K_i}{\bar{D}^2}$$

where $Var[D] = \sum_{i=1}^{n}p_i(D_i - \bar{D})^2$

APPENDIX G
MAXIMUM KURTOSIS OF THE BIEXPONENTIAL MODEL 26

Suppose there are two Gaussian compartments with relative fractions p_1 and p_2, and diffusion coefficients D_1 and D_2. Then, because for a Gaussian $E[X^4] = 12\bar{D}^2$ and $E[X^2] = 2\bar{D}$ we have for the kurtosis:

$$K = 3 * \frac{p_1 D_1^2 + p_2 D_2^2}{\bar{D}^2} - 3 \qquad (76)$$

To find the maximal kurtosis, the derivative of the logarithm with respect to p_1 of the latter expression is solved to zero. This leads (since $p_1 D_1 + p_2 D_2 = \bar{D}$) to the following equality

$$\frac{\delta log(p_1 D_1^2 + p_2 D_2^2)}{\delta p_1} = \frac{\delta log((p_1 D_1 + p_2 D_2)^2)}{\delta p_1} \qquad (77)$$

which is equivalent to

$$\frac{D_1 + D_2}{p_1 D_1^2 + p_2 D_2^2} = \frac{2}{p_1 D_1 + p_2 D_2}$$

which leads to

$$p_1 D_1 D_2 + p_2 D_1 D_2 = p_1 D_1^2 + p_2 D_2^2$$

which yields

$$p_1 = \frac{D_1}{D_1 + D_2}$$

APPENDIX H
DERIVATION OF 3.5.2

In a particular direction, we have for the diffusion coefficient D_i

$$D_i = fD_{a,i} + (1-f)D_{e,i}, \qquad (78)$$

and the kurtosis in a particular direction is

$$\begin{aligned}
K_i &= 3 * \frac{f(D_i - D_{i,a})^2 + (1-f)*(D_i - D_{i,e})^2}{D_i^2}\\
&= 3 * \frac{f(D_{i,a} - (1-f)D_{i,e} - fD_{i,a})^2}{D_i^2}\\
&\quad + 3 * \frac{(1-f)(D_{i,e} - fD_{i,a} - (1-f)D_{i,e})^2}{D_i^2}\\
&= 3 * \frac{f(1-f)^2(D_{i,e} - D_{i,a})^2 + (1-f)f^2(D_{i,e} - D_{i,a})^2}{D_i^2}\\
&= 3 * f(1-f)\frac{(D_{i,e} - D_{i,a})^2}{D_i^2} \qquad (79)
\end{aligned}$$

APPENDIX I
RADIAL KURTOSIS IMPLIED BY A BIZEPPELIN MODEL WITH TORTUOUSITY RESTRICTIONS

Under the assumption that $D_{e,\parallel} = D_{a,\parallel}$ and that $D_{e,\perp} = (1 - v) * D_{e,\parallel}$, we have for the radial kurtosis

$$
\begin{aligned}
K_\perp &= 3 * \frac{(1-v)v(D_{e\perp} - D_{a\perp})^2}{((1-v)D_{e\perp} + vD_{a\perp})^2} \\
&= 3 * \frac{(1-v)v(D_{a\parallel}(1-v) - D_{a\perp})^2}{((1-v)^2 D_{a\parallel} + vD_{a\perp})^2} \\
&= 3 * \frac{(1-v)^3 v(D_{a\parallel} - \frac{D_{a\perp}}{1-v})^2}{(1-v)^4(D_{a\parallel} + \frac{v}{(1-v)^2}D_{a\perp})^2} \\
&= 3 * \frac{v(D_{a\parallel} - \frac{D_{a\perp}}{1-v})^2}{(1-v)(D_{a\parallel} + \frac{v}{(1-v)^2}D_{a\perp})^2} \\
&= 3 * \frac{v}{1-v} \left(\frac{1 - \frac{1-v}{(1-v)^2}\frac{D_{a\perp}}{D_{a\parallel}}}{1 + \frac{v}{(1-v)^2}\frac{D_{a\perp}}{D_{a\parallel}}} \right)^2 \\
&\approx 3 * \frac{v}{1-v} \left(1 - \frac{1}{(1-v)^2}\frac{D_{a\perp}}{D_{a\parallel}} \right)^2 \quad (80)
\end{aligned}
$$

where we used the approximation $\frac{1-\delta_1}{1+\delta_2} \approx x - \delta_1 - \delta_2$ when $\delta_1 + \delta_2 << 1$, which is reasonable since $D_{a,\perp} << D_{a,\parallel}$ and since we assumed that $v \approx 0$.

APPENDIX J
DERIVATION OF THE KURTOSIS OF AN ISOTROPIC CYLINDER MODEL

(Part of this derivation is adapted from Mulkern (2009) [76]) Every axon direction vector starts in the origin and has a length of 1 and thus lies on the surface of the unit sphere. If we cut the sphere in circles around the gradient axis, every direction endpoint lies on such a circle. When the orientation is isotropic, the amount of directions that make a certain angle θ with the gradient axis is proportional to the circumference of the circle around the gradient axis which is touched by a direction of angle θ. This circumference approaches zero for angles of zero, and approaches 2π when the angle approaches $\pi/2$. It can be simply shown that the circumference and thus the amount of spins that make a certain angle θ varies as $\sin\theta$ between 0 and $\pi/2$. Hence we can write for the signal intensity when the distribution is uniform

$$
S_c = v * \int_0^{\frac{\pi}{2}} \sin\theta e^{-b(D_L \cos^2\theta + D_T \sin^2\theta)} d\theta
$$

By using that

$$
\sin^2\theta = 1 - \cos^2\theta,
$$

we obtain

$$
S_c = v * \int_0^{\frac{\pi}{2}} \sin\theta e^{-bD_T} e^{-b(D_L - D_T)\cos^2\theta} d\theta
$$

By Taylor expanding the exponentials, neglecting $O(b^3)$ and higher terms and performing the integrals we obtain

$$
\begin{aligned}
S_c &= v * (1 - D_T b + \frac{D_T^2 b^2}{2}) * \\
&\quad \int_0^{\pi/2} \sin(\theta) - b(D_L - D_T)\cos^2\theta\sin\theta \\
&\quad + \frac{b^2(D_L - D_T)^2}{2}\cos^4\theta\sin\theta d\theta \\
&= v * (1 - D_T b + \frac{D_T^2 b^2}{2}) * \\
&\quad (1 - \frac{b}{3}(D_L - D_T) + \frac{b^2}{10}(D_L - D_T)^2) \\
&= v * (1 - (1/3 D_L + 2/3 D_T)b) + \\
&\quad v * \left(\frac{D_T^2}{2} + \frac{D_T(D_L - D_T)}{3} + \frac{(D_L - D_T)^2}{10} \right) b^2
\end{aligned}
$$

Where we use the fact that $\int_0^{\frac{\pi}{2}} \cos^n\theta\sin\theta d\theta = \frac{1}{n+1}$
Because of the moment expansion

$$
\frac{S(b)}{S(0)} = 1 - \frac{E[X^2]}{2}b + \frac{E[X^4]}{24}b^2 + O(b^3)
$$

Hence for the cylindrical compartment we have

$$
\bar{D}_c = \frac{1}{3}D_L + \frac{2}{3}D_T
$$

and

$$
E[X_c^4] = 12 * \left(D_T^2 + \frac{2 * D_T(D_L - D_T)}{3} + \frac{(D_L - D_T)^2}{5} \right)
$$

The total signal is equal to

$$
S(b) = (1 - v) * e^{-bD_{eff}} + v * S_C
$$

Hence, by using $K = \frac{E[X^4]}{E[X^2]^2} - 3$ we obtain for the kurtosis

$$
3 * \frac{(1-v) * D_{eff}^2 + v * \left(D_T^2 + \frac{2*D_T(D_L-D_T)}{3} + \frac{(D_L-D_T)^2}{5} \right)}{\left((1-v) * D_{eff} + v * (\frac{2}{3}D_T + \frac{1}{3}D_L) \right)^2} - 3
$$

APPENDIX K
KURTOSIS OF THE BALL & RACKETS MODEL.

Let X be the radial displacement of an axon. Then

$$
\begin{aligned}
S_c(\vec{n}) &= \int_x \int_{S^2} f(x)h(\vec{\mu})e^{-iqx\vec{\mu}\cdot\vec{n}} d\vec{\mu}dx \\
&= \int_x \int_{S^2} f(x)h(\vec{\mu})(1 - \frac{x^2(\vec{\mu}\cdot\vec{n})^2}{2}b + \frac{x^4(\vec{\mu}\cdot\vec{n})^4}{24}b^2 + O(b^3)) \\
&= 1 - \frac{1}{2}\int_x \int_{S^2} x^2(\vec{\mu}\cdot\vec{n})^2 bd\vec{\mu}dx + O(b^3) \\
&\quad + \frac{1}{24}\int_x \int_{S^2} x^4(\vec{\mu}\cdot\vec{n})^4 b^2 d\vec{\mu}dx + O(b^3) \\
&= 1 - \frac{1}{2}E[X^2(\vec{\mu}\cdot\vec{n})^2]b + \frac{1}{24}E[X^4(\vec{\mu}\cdot\vec{n})^4]b^2 + O(b^3) \\
&= 1 - \frac{1}{2}E[X^2]E[(\vec{\mu}\cdot\vec{n})^2]b + \frac{1}{24}E[X^4]E[(\vec{\mu}\cdot\vec{n})^4]b^2
\end{aligned}
$$

The latter step is justified if the radial displacement X is independent of the angle it makes with the gradient vector. Hence the kurtosis in direction $\vec{n}$ for the axonal compartment is

$$
K(\vec{n}) = \frac{E[X^4]E[(\vec{\mu}\cdot\vec{n})^4]}{E[X^2]^2 E[(\vec{\mu}\cdot\vec{n})^2]^2} - 3 \quad (81)
$$

For a centralized Gaussian distribution, it is well known that $E[X^2] = \sigma^2$ and $E[X^4] = 3\sigma^4$. This yields

$$K(\vec{n}) = 3 * \left(\frac{E[(\vec{\mu} \cdot \vec{n})^4]}{E[(\vec{\mu} \cdot \vec{n})^2]^2} - 1 \right)$$

and for the cylindrical diffusivity we have

$$D_{a\parallel}(\vec{n}) = D_{a\parallel} E[(\vec{\mu} \cdot \vec{n})^2]$$

Thus

$$v * D_{a\parallel}(\vec{n})^2 * K(\vec{n}) = 3D_{a,\parallel}^2 (E[(\vec{\mu} \cdot \vec{n})^4] - E[(\vec{\mu} \cdot \vec{n})^2]^2)$$

Because of the partial volume kurtosis formula (equation 23) we have that the total voxel kurtosis in direction $\vec{n}$ is independent of $D_\parallel$ and equal to

$$3 \frac{v(1-v)(1 - E[(\vec{\mu} \cdot \vec{n})^2])^2 + v(E[\vec{\mu} \cdot \vec{n}]^4 - E[(\vec{\mu} \cdot \vec{n})^2]^2)}{((1-v) + vE[(\vec{\mu} \cdot \vec{n})^2])^2}$$

APPENDIX L
$E[e^{iqX}]$ HAS AN INFINITE RADIUS OF CONVERGENCE

If for any q

$$lim_{n\to\infty} \frac{|q^n|E[|X^n|]}{n!} \to 0, \tag{82}$$

the radius of convergence of the Taylor expansion of $E[e^{-iqX}]$ is infinite [84]. The particle displacement X during a pulse sequence is of course bounded[6]. Therefore $q^n E[|X^n|]$ is bounded by $(qM)^n$ for some $M > 0$, and hence the $n!$ term dominates for sufficiently large n in equation (82), which therefore goes to zero.

APPENDIX M
CUMULANTS (2.1.1.3)

In this section, the coefficients of the cumulant expansion will be obtained by plain counting. As often is the case with counting problems, one may find it easier to read the derivation and use it as a hint to count the coefficients by himself rather then trying to keep track of all the counting and notation in somebody else's derivation.

Lemma. *Let $n \in \mathbb{N}$. Let the collection (set of sets) $\mathcal{N}_n$ be defined as*

$$\mathcal{N}_n := \{\{r_1, \ldots, r_k\} | r_i \in \mathbb{N}_{>0}, \sum_{i=1}^{k} r_i = n\}$$

Let $\mathcal{N}_n$ be all permutations (in the high school meaning of the word) one can make out of elements in $\mathcal{N}_n$. For example, if $n = 4$, then $\mathcal{N}_4 = \{\{1,1,1,1\}, \{1,2,1\}, \{1,3\}, \{4\}\}$ and $\mathcal{N}_4 = \{(1,1,1,1), (1,1,2), (2,1,1), (1,2,1), (1,3), (3,1), (4)\}$.
Let $\boldsymbol{r} \in \mathcal{N}_n$. Than $a_{\boldsymbol{r}}$ is defined as the coefficient belonging to $\prod_{r_i \in \boldsymbol{r}} E[X^{r_i}]$ in the n-th cumulant κ_n, i.e.

$$\kappa_n = \sum_{\boldsymbol{r} \in \mathcal{N}_n} a_{\boldsymbol{r}} \prod_{r_i \in \boldsymbol{r}} E[X^{r_i}]$$

Let M^* be the set of all matrices such that $M_{ij} = 0$ if $i < j$. Let us define the set of matrices $\mathcal{M}_n$ as all matrices M^* where the row sums are equal to $(r_1, \ldots, r_k)$ for a $r \in \mathcal{N}_n$.

$$\mathcal{M}_n := \{M \in M^* | \exists \boldsymbol{r} \in \mathcal{N} : \forall r_i : \sum_{j=1}^{n} M_{ij} = r_i - 1\} \tag{83}$$

6. By the skull of the brain or at least by the thick walls of the MR room.

Then

$$a_{\boldsymbol{n}} = (-1)^{(|\boldsymbol{n}-1|)}(|\boldsymbol{n}| - 1)! \sum_{M \in \mathcal{M}_n} \prod_{j \le n} \left(\frac{\sum_i M_{ij}}{\prod_i M_{ij}!} \right)$$

Proof. The first derivative of $E[e^{itX}]$ is $i \frac{E[Xe^{itX}]}{E[e^{itX}]}$. The second derivative of $E[e^{itX}]$ is according to the product rule of differentiation equal to

$$i^2 \left(\frac{E[X^2 e^{itX}]}{E[e^{itX}]} - \frac{E[Xe^{itX}]E[Xe^{itX}]}{E[e^{2itX}]} \right).$$

Hence, every time a derivative is taken, the product rule doubles the amount of terms by deriving the denominator or the numerator. A term $E[X^2]E[X]$ could hence be established by deriving the numerator, numerator and denominator in any order. However, when a term like $E[e^{tX}]E[e^{tX}]$ appears in the numerator, the first term or the second term can be derived in the product rule, which makes the coefficients less straightforward to count than just looking at Pascals triangle. Consider $E[X^2]E[X]E[X]$. Let us distinguish the different X by the order in which they where 'created' by the product rule. Hence $E[X_1^2]E[X_2]E[X_3]$ means that the first X that was 'created' gets a power of two at the end, the second and third created gets a power of one. We can code the combinations $E[X_1^2]E[X_2]E[X_3]$ as $(2,1,1)$. The coefficient of $E[X^2]E[X]E[X]$ is the amount of possibilities to start at $(1,0,0)$, which codes for $E[X]$, and to arrive at $(2,1,1)$, but also at $(1,2,1)$ and at $(1,1,2)$. Now let us consider a higher order example. $E[X^3]E[X^3]E[X^3]$ is coded as $(3,3,3)$. How to arrive from $(1,0,0)$ at $(3,3,3)$? It is obvious that this path crosses through $(1,0,0)$, through $(a,1,0)$ and through $(b,c,1)$. We can easily count those sub paths, e.g. from $(a,1,0)$ to $(b,c,1)$ yields $\binom{b-a+c-1}{c-1}$ different paths. By considering all intermediate a, b and c, we obtain all paths and thus the coefficient of the corresponding moment expression if we also multiply it by $(-1)^{n-1}(n-1)!$ where n is the amount of moments in the expression, i.e. every time one wants to go from $E[X]$ to $E[X]E[X]$, the denominator needs to be derived, which multiplies the expression by an extra constant, e.g. the derivative of $\frac{1}{E[e^{2tX}]}$ is $-\frac{2E[Xe^{2tX}]}{E[e^{3tX}]}$.
Now lets consider a more advanced example. How to go from $(1,0,0)$ to $(5,6,4)$ via $(x_1,1,0)$ and $(x_2,y_2,1)$? Call $X_1 = x_1 - 1$, $X_2 = x_2 - x_1$, $X_3 = 5 - x_2$, $Y_2 = y_2 - 1$, $Y_3 = 6 - y_2$, $Z_3 = 4 - 1 = 3$. The amount of different paths is equal to

$$\binom{X_2 + Y_2}{X_2} * \binom{X_3 + Y_3 + Z_3}{Y_3 + X_3} * \binom{X_3 + Y_3}{X_3}$$

By writing out the binomial coefficients, this can be simplified to

$$\frac{X_2 + Y_2}{X_2! Y_2!} * \frac{X_3 + Y_3 + Z_3}{X_3! Y_3! Z_3!}$$

For a convenient notation we can make a matrix M where we store the numbers $X_1, \ldots Z_3$.

$$\begin{pmatrix} X_1 & X_2 & X_3 \\ 0 & Y_2 & Y_3 \\ 0 & 0 & Z_3 \end{pmatrix} \tag{84}$$

Every such matrix codes for all paths going trough $(x_1, 1, 0)$

and $(x_2, y_2, 1)$ and must therefore obey the relations

$$X_1 + X_2 + X_3 = 5$$
$$Y_1 + Y_2 = 6$$
$$Z_1 = 4$$

By summing over all these intermediate points, i.e. the matrices $\mathcal{M}_n$, we obtain the lemma. $\quad\square$

Lemma. *Suppose*

$$\mathcal{M}^*_n := \{M \in \mathcal{M}_n | M \text{ is a diagonal matrix}\}$$

then

$$a_{\boldsymbol{n}} = (-1)^{(|\boldsymbol{n}-1|)}(|\boldsymbol{n}|-1)! \left(\sum_{M \in \mathcal{M}_n} \prod_{j \leq n} (\frac{\sum_i M_{ij}}{\prod_i M_{ij}!}) + \sum_{M \in \mathcal{M}^*_n/\mathcal{M}_n} 1 \right)$$

Proof. When M is a diagonal matrix, this means that first the first coordinate is totally 'filled', then the second, then the third... There is obviously only one way to do this. $\quad\square$

Example: The coefficient of $E[X^2]^2 E[X]^2$ in κ_6 is equal to 270.

The set $r = \{2, 2, 1, 1\} \in \mathcal{N}_6$ corresponds to $E[X^2]^2 E[X]^2$. The list of all permutations of r has the following elements:

$$(2, 2, 1, 1)$$
$$(2, 1, 2, 1)$$
$$(2, 1, 1, 2)$$
$$(1, 2, 2, 1)$$
$$(1, 1, 2, 2)$$
$$(1, 2, 1, 2)$$

Consider the first permutation $(2, 2, 1, 1)$. To make a matrix out of this permutation that is contained in $\mathcal{M}$ each sum of the first two rows has to be 1 and each sum of the last two rows has to be 0. This is an example of such a matrix.

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

As we can see there are 15 ways to make a matrix $M \in \mathcal{M}$ out of $(2, 2, 1, 1)$, and 3 ways to make such a matrix that $M_i \cdot M_j = 1$, i.e, $M \in \mathcal{M}^*$. Thus the first permutation gives an contribution of $3 * \binom{2}{1} + (12 - 3) = 15$. In general we can see that the contribution of a permutation is $a*b+b$ where a is the index of the first 2, b is the index of the second 2 (for example $(2, 1, 2, 1)$ gives $a = 4, b = 2$)

$$\begin{array}{ll} (2, 2, 1, 1) & 4 * 3 + 3 = 15 \\ (2, 1, 2, 1) & 4 * 2 + 2 = 10 \\ (2, 1, 1, 2) & 4 * 1 + 1 = 5 \\ (1, 2, 2, 1) & 3 * 2 + 2 = 8 \\ (1, 1, 2, 2) & 2 * 1 + 1 = 3 \\ (1, 2, 1, 2) & 3 * 1 + 1 = 4 \end{array}$$

Together this sums up to 45, multiplied with $(|r| - 1)!(= 6)$ gives us the desired 270.

APPENDIX N
SPHERICAL HARMONICS (3.7.1)

I modified this section from [85], by reorganizing the arguments and filling in some gaps.
Let S^2 be a sphere of radius 1:

$$S^2 := \{(x, y, z) \in \mathbb{R}^3 | x^2 + y^2 + z^2 = 1\} \quad (85)$$

The Laplacian is given by

$$\Delta f = \frac{\delta^2 f}{\delta x^2} + \frac{\delta^2 f}{\delta y^2} + \frac{\delta^2 f}{\delta z^2} \quad (86)$$

The idea is essentially that every function on the sphere is a solution of the Laplacian, whereas every solution of the Laplacian is a sum of some 'basis' functions.
The spherical coordinates are

$$x = r sin\theta cos\phi \quad (87)$$
$$y = r sin\theta sin\phi \quad (88)$$
$$z = r cos\theta \quad (89)$$

And the chain rule for partial differentiation is

$$\frac{\delta}{\delta x} = \frac{\delta r}{\delta x}\frac{\delta}{\delta r} + \frac{\delta \theta}{\delta x}\frac{\delta}{\delta \theta} + \frac{\delta \phi}{\delta x}\frac{\delta}{\delta \phi} \quad (90)$$

With these facts in hand, we can derive the Laplacian in spherical coordinates.

$$\Delta f = \frac{1}{r^2}\frac{\delta}{\delta r}(r^2 \frac{\delta f}{\delta r}) + \frac{1}{r^2}\Delta_{S^2} f \quad (91)$$

where

$$\Delta_{S^2} f = \frac{1}{sin\theta}\frac{\delta}{\delta\theta}\left(sin\theta\frac{\delta f}{\delta\theta}\right) + \frac{1}{sin^2\theta}\frac{\delta^2 f}{\delta\phi^2} \quad (92)$$

Let us look for functions $f(r, \theta, \phi) = r^k g(\theta, \phi)$ such that $\Delta f = 0$
We get

$$\Delta f = \frac{1}{r^2}\frac{\delta}{\delta r}\left(r^2\frac{\delta(r^k g)}{\delta r}\right) + \frac{1}{r^2}\Delta_{S^2}(r^k g) \quad (93)$$
$$= \frac{1}{r^2}\frac{\delta}{\delta r}(k r^{k+1} g) + r^{k-2}\Delta_{S^2} g \quad (94)$$
$$= r^{k-2}k(k+1)g + r^{k-2}\Delta_{S^2} g \quad (95)$$
$$= r^{k-2}(k(k+1)g + \Delta_{S^2} g) \quad (96)$$

Therefore,

$$\Delta f = 0 \Leftrightarrow \Delta_{S^2} g = -k(k+1)g \quad (97)$$

that is, g is an eigenfunction of Δ_{S^2} for the eigenvalue $-k(k+1)$.
If we use the separation method $g(\theta, \phi) = \Theta(\theta)\Phi(\phi)$, then we get the equation

$$\frac{\Phi}{sin\theta}\frac{\delta}{\delta\theta}(sin\theta\frac{\delta\Theta}{\delta\theta}) + \frac{\Theta}{sin^2\theta}\frac{\delta^2\Phi}{\delta\phi^2} = -k(k+1)\Theta\Phi \quad (98)$$

Dividing by $\Theta\Phi$ and multiplying by $sin^2\theta$,

$$\frac{sin\theta}{\Theta}\frac{\delta}{\delta\theta}(sin\theta\frac{\delta\Theta}{\delta\theta}) + k(k+1)sin^2\theta = -\frac{1}{\Phi}\frac{\delta^2\Phi}{\delta\phi^2} \quad (99)$$

Since Θ and Φ are independent functions, it follows that both sides are equal to a constant, say μ. This leads to two equations

$$\frac{\delta^2 \Phi}{\delta \phi^2} + \mu \Phi = 0 \qquad (100)$$

$$\frac{sin\theta}{\Theta} \frac{\delta}{\delta\theta}(sin\theta \frac{\delta\Theta}{\delta\theta}) + k(k+1)sin^2\theta - \mu = 0 \qquad (101)$$

However, we want Φ to be periodic in ϕ since we are considering functions on the sphere, so μ must be of the form $\mu = m^2$. Then, we know that solutions of the equation

$$\frac{\delta^2 \Phi}{\delta \phi^2} + m^2 \Phi = 0 \qquad (102)$$

are

$$\Phi = \lambda_1 cos(m\phi) + \lambda_2 sin(m\phi). \qquad (103)$$

We also have to solve the equation

$$sin\theta \frac{\delta}{\delta\theta}(sin\theta\frac{\delta\Theta}{\delta\theta}) + (k(k+1)sin^2\theta - m^2)\Theta = 0 \qquad (104)$$

which is equivalent to

$$sin^2\theta\Theta^{''} + sin\theta cos\theta\Theta^{'} + (k(k+1)sin^2\theta - m^2)\Theta = 0. \qquad (105)$$

For this, by using the change of variable, $\Theta(\theta) = y(cos\theta)$ and, with $x = cos\theta$, we get:

$$(1-x^2)y^{''} - 2xy^{'} + \left(k(k+1) - \frac{m^2}{1-x^2}\right)y = 0 \qquad (106)$$

where $y^{'} = \frac{\delta y}{\delta x}$.
To solve this equation, we can make the substitution

$$y(x) = (1-x^2)^{\frac{m}{2}}u(x). \qquad (107)$$

Notice that by subsequently applying the chain rule and the product rule, we have

$$y^{'} = (1-x^2)^{\frac{m}{2}}u^{'}(x) - mx(1-x^2)^{\frac{m}{2}-1}u(x) \qquad (108)$$

$$y^{''} = (1-x^2)^{\frac{m}{2}}u^{''}(x) - 2xm(1-x^2)^{\frac{m}{2}-1}u^{'}(x) \qquad (109)$$

$$- m(1-x^2)^{\frac{m}{2}-1}u(x) + m(m-2)x^2(1-x^2)^{\frac{m}{2}-2}u(x) \qquad (110)$$

We get the equation:

$$au^{''} + bu^{'} + cu = 0 \qquad (111)$$

with

$$a = (1-x^2)(1-x^2)^{\frac{m}{2}}$$
$$b = -2x[(1-x^2)^{\frac{m}{2}}] - 2xm(1-x^2)^{\frac{m}{2}}$$
$$= -2(m+1)x(1-x^2)^{\frac{m}{2}}$$
$$c = [k(k+1) - \frac{m^2}{1-x^2}][(1-x^2)^{\frac{m}{2}} - (2mx^2(1-x^2)^{\frac{m}{2}-1})$$
$$- m(1-x^2)^{\frac{m}{2}} + m(m-2)x^2(1-x^2)^{\frac{m}{2}-1}]$$
$$= [k(k+1)[(1-x^2)^{\frac{m}{2}}] - m^2(1-x^2)(1-x^2)^{\frac{m}{2}-1}$$
$$- m(1-x^2)^{\frac{m}{2}}$$
$$= (k(k+1) - m(m+1))(1-x^2)^{\frac{m}{2}} \qquad (112)$$

Dividing by $(1-x^2)^{\frac{m}{2}}$ we get

$$(1-x^2)u^{''} - 2(m+1)xu^{'} + (k(k+1) - m(m+1))u = 0 \qquad (113)$$

When $m = 0$, we get the Legendre equation

$$(1-x^2)u^{''} - 2xu^{'} + k(k+1)u = 0 \qquad (114)$$

We propose that

$$P_k(x) = \frac{1}{2^k k!}\frac{d^k}{dx^k}(x^2-1)^k \qquad (115)$$

is a solution of the Legendre equation.
Let

$$v = (x^2-1)^l$$

Then

$$\frac{dv}{dx} = l(x^2-1)^{l-1}2x \qquad (116)$$

and

$$(x^2-1)\frac{dv}{dx} = l(x^2-1)^l2x = 2xlv \qquad (117)$$

Differentiating the left hand side $l+1$ times wrt x gives

$$\frac{d^{l+1}}{dx^{l+1}}(x^2-1)\frac{dv}{dx} = (x^2-1)\frac{d^{l+2}v}{dx^{l+2}}+(l+1)2x\frac{d^{l+1}v}{dx^{l+1}}+\frac{(l+1)l}{2!}2\frac{d^lv}{dx^l} \qquad (118)$$

Differentiating the right hand side $l+1$ times wrt x gives

$$2lx\frac{d^{l+1}v}{dx^{l+1}} + (l+1)2l\frac{d^lv}{dx^l} \qquad (119)$$

and thus

$$(x^2-1)\frac{d^{l+2}v}{dx^{l+2}} + (l+1)2x\frac{d^{l+1}v}{dx^{l+1}} + \frac{(l+1)l}{2!}2\frac{d^lv}{dx^l} = \qquad (120)$$

$$2lx\frac{d^{l+1}v}{dx^{l+1}} + (l+1)2l\frac{d^lv}{dx^l} \qquad (121)$$

$$\qquad (122)$$

which simplifies to

$$(x^2-1)\frac{d^{l+2}v}{dx^{l+2}} + 2x\frac{d^{l+1}v}{dx^{l+1}} - l(l+1)\frac{d^lv}{dx^l} = 0. \qquad (123)$$

Hence $P_k(x)$ is a solution of the Legendre equation.
Let us now return to our differential equation

$$(1-x^2)u^{''} - 2(m+1)xu^{'} + (k(k+1) - m(m+1))u = 0 \qquad (124)$$

If we differentiate with respect to x we get the equation

$$(1-x^2)u^{'''} - 2(m+2)xu^{''} + (k(k+1)) - (m+1)(m+2))u^{'} = 0. \qquad (125)$$

This shows that if $P_k(x)$ is a solution of (124) for a given k and $m = 0$, then $P_k^{'}(x)$ solves (124) for the same k and $m = 1$ and $P_k^{''}(x)$ solves (124) for the same k and $m = 2$.
Therefore the original equation

$$(1-x^2)y^{''} - 2xy^{'} + (k(k+1) - \frac{m^2}{1-x^2})y = 0 \qquad (126)$$

has the solution

$$y(x) = (1-x^2)^{\frac{m}{2}}\frac{d^m}{dx^m}(P_k(x)) \qquad (127)$$

The function $y(x)$ is denoted $P_k^m(x)$ and is called the associated Legendre function.
Recall that the original function was a solution of the equation

$$sin\theta\frac{\delta}{\delta\theta}\left(sin\theta\frac{\delta\Theta}{\delta\theta}\right) + (k(k+1)sin^2\theta - m^2)\Theta = 0 \qquad (128)$$

By making the reverse substitution $y(cos\theta) = \Theta(\theta)$ to get the function Θ back, we have that

$$\Theta(\theta) = P_k^m(cos\theta) \tag{129}$$

is a solution of the above equation. Putting everything together, as

$$f(r, \theta, \phi) = r^k \Theta(\theta) \Phi(\phi) \tag{130}$$

we proved that the functions

$$c_k^m = cos m\phi P_k^m(cos\theta) \tag{131}$$

$$s_k^m = sin m\phi P_k^m(cos\theta) \tag{132}$$

are eigenfunctions of the Laplacian Δ_{s^2}, on the sphere of radius 1 for the eigenvalue $-k(k+1)$.

If we now define the inner product on $L^2(S^2)$ as

$$< f, g > = \int_0^{2\pi} \int_0^{\pi} f(\theta, \phi) g(\theta, \phi) sin\theta d\theta d\phi \tag{133}$$

Then, because the functions 131 and 132 are eigenfunctions belonging to the eigenvalue $k(k-1)$, they form -with the proper scaling- an orthonormal system.

The Laplacian is self-adjoint on the sphere, that is,

$$< \Delta_{S^2} v, w > = < v, \Delta_{S^2} w > \tag{134}$$

The latter fact combined with the fact that c_k^m is an eigenfunction of Δ_{S_2} with eigenvalue $k(k+1)$, gives

$$k(k+1) < c_k^m, c_l^m > = < \Delta_{S_2} c_k^m, c_l^m >$$
$$= < c_k^m, \Delta_{S_2} c_l^m >$$
$$= l(l+1) < c_k^m, c_l^m > \tag{135}$$

and therefore $< c_k^m, c_l^m > = 0$ if $l \neq k$. This also holds for $< s_k^m, s_l^m >$ and $< s_k^m, c_l^m >$.

It turns out that every sufficiently smooth function on the sphere $f \in L^2(S^2)$ can be written as a linear combination of c_k^m and s_k^m

$$f = \sum_{k=0}^{\infty} \sum_{m=0}^{\infty} a_k^m c_k^m + b_k^m s_k^m \tag{136}$$