Instantaneous helical axis estimation via natural, cross-validated splines

INSTANTANEOUS HELICAL AXIS ESTIMATION VIA NATURAL, CROSS-VALIDATED SPLINES+

H.J. Woltring*, A. de Lange**, J.M.G. Kauer** and R. Huiskes**

Philips Medical Systems Division*, Best, The Netherlands and
Lab. Exper. Orthop.**, University of Nijmegen, The Netherlands

1. INTRODUCTION

In studies of biological joint motion, quantification of translations and rotations by means of a reference point and attitude angles does not provide a clear insight in the relation between kinematics and joint geometry. Because of its geometric simplicity, a better picture can be obtained by means of the **Instantaneous Helical Axis** (IHA), also known as the instantaneous screw axis, twist axis, or axis of rotation. At each moment in time, joint motion is seen as the movement of one body segment with respect to an adjacent segment (usually distal with respect to proximal), with a translation component along, and a rotation component about a directed line in space which is uniquely determined as long as the rotatory component does not vanish: see Figure 1. The total amounts of translation and rotation along the path of motion can be defined as the time integrals of the instantaneous translation and rotation velocities at the IHA from a given reference time.

The position and direction of the IHA, and the translation and rotation velocities at the IHA can be estimated by smoothing and interpolating raw measurements obtained through stereophotogrammetry or electrogoniometry, followed by conventional rigid-body calculus. In this way, the high errors incurred by direct estimation of the **Finite Helical Axis** (FHA) as an approximation for the IHA can be avoided: see Woltring et al. (1985).

Smoothing and interpolation are possible with optimally **regularized,** natural splines of sufficiently high order; this is shown to be equivalent to classical Butterworth filtering, with data-driven determination of the smallest bandwidth for which signal loss is negligible, under the assumption of a low-pass signal with additive, white measurement noise.

2. INSTANTANEOUS RIGID-BODY KINEMATICS

Given a body-fixed co-ordinate system E^x, and a global co-ordinate system E^y, the position vectors $\underline{x}$ in E^x and $\underline{y}(t)$ in E^y of some point X on the moving body can be related as

$$\underline{y}(t) = R(t).\underline{x} + \underline{p}(t) \tag{1}$$

where $\underline{p}(t)$ is the instantaneous position of the origin of E^x in E^y, and $R(t)$ the instantaneous attitude matrix of E^x in E^y. The matrix $R(t)$ is **orthonormal**, i.e.

$$R(t)'R(t) = R(t).R(t)' = I \tag{2}$$

+ From the Laboratory of Experimental Orthopaedics, University of Nijmegen The Netherlands

The IHA can be defined as the locus of all points $\{x\}$ with minimal velocity $|\dot{\underline{y}}(t)|$, since the rotation component in $\dot{\underline{y}}(t)$ vanishes for all points on the IHA (the translation and rotation components are perpendicular to each other). Differentiation of (1) and elimination of $\underline{x}$ via re-substitution of (1) results in

$$\dot{\underline{y}}(t) = \dot{R}(t).R(t)'.\{\underline{y}(t) - \underline{p}(t)\} + \dot{\underline{p}}(t) \quad (3)$$

The product $\dot{R}(t).R(t)'$ is **skew-symmetric** (cf. Woltring **et al.,** 1985) as can be shown by differentiating the orthonormality constraint (2). The axial vector $\underline{\omega}(t)$ of this skewed matrix is known as the instantaneous rotation velocity vector, with amplitude $\omega(t)$ and unit direction vector $\underline{n}(t)$,

$$\omega(t) = \{\underline{\omega}(t)'\underline{\omega}(t)\}^{\frac{1}{2}}, \qquad \underline{n}(t) = \underline{\omega}(t)/\omega(t) \quad (4)$$

and relation (3) can be expressed in vector-product form as

$$\dot{y}(t) = \underline{\omega}(t) * \{\underline{y}(t) - \underline{p}(t)\} + \dot{\underline{p}}(t) \quad (5)$$

The locus of the IHA follows by minimizing $|\dot{\underline{y}}(t)|^2$ as a function of $\underline{y}(t)$. Taking partial derivatives results in

$$Q.\{\underline{y}(t) - \underline{p}(t)\} = \underline{\omega}(t) * \dot{\underline{p}}(t), \qquad Q = \omega^2(t).I - \underline{\omega}(t).\underline{\omega}(t)' \quad (6)$$

which yields a **class** of solutions since Q is singular. For $\omega(t) \neq 0$, Q has rank 2, with $\underline{\omega}(t)$ as its null vector. Thus, for some point $\underline{y}(t) = \underline{s}(t)$ meeting (6), the IHA can be parametrically represented by the line $\underline{s}(t) + \lambda.\underline{n}(t)$. A useful choice for $\underline{s}(t)$ is the projection of $\underline{p}(t)$ onto the IHA, and this yields the additional condition

$$\underline{\omega}(t)'\{\underline{s}(t) - \underline{p}(t)\} = 0 \quad (7)$$

Combination of (6) and (7) results in the following explicit relation for $\underline{s}(t)$,

$$\underline{s}(t) = \underline{p}(t) + \underline{\omega}(t) * \dot{\underline{p}}(t)/\omega^2(t) \quad (8)$$

while the instantaneous translation or **shift** speed $v(t)$ along the IHA follows by projecting $\dot{\underline{y}}(t)$ onto $\underline{n}(t)$,

$$v(t) = \underline{n}(t)'\dot{\underline{p}}(t) \quad (9)$$

Given the position $\underline{p}(t)$, its velocity $\dot{\underline{p}}(t)$, and the rotation velocity $\underline{\omega}(t)$, the IHA is completely determined by its position $\underline{s}(t)$ and direction $\underline{n}(t)$, while the amount of instantaneous motion follows from the translation velocity $v(t)$ and rotation velocity $\omega(t)$. However, the IHA becomes <u>undefined</u> for vanishing $\omega(t)$. (NB: it is often said that the IHA moves to infinity for vanishing $\omega(t)$, but this is strictly incorrect). For reasons of continuity, problems should be expected if the rotation velocity becomes "small", especially if the IHA is to be estimated from noisy position or angle data.

The occurrence of derivatives, especially in the denominators of most IHA parameters renders the IHA estimation problem **ill-posed** or **incorrectly posed** (Morozov, 1984). Unless special precautions are taken, small measurement errors have an inordinately large influence on the unknowns in such problems.

3. DERIVATIVE ESTIMATION VIA OPTIMAL REGULARIZATION

Many practical problems including derivative estimation from noisy data are ill-posed in the above sense, and additional constraints must be imposed in order to arrive at manageable solutions. A common approach is based on **regularization theory** (Morozov, 1984). Here, a balance is sought between two conflicting goals: goodness of fit to the noisy data, and smoothness of the fitted signal in terms of a functional of the signal and its derivatives. An early example of this approach is the procedure of Anderssen & Bloomfield (1974) for equidistantly sampled, uniformly weighted data which are processed via an FFT-algorithm (cf. Hatze, 1981).

Spline functions

More recently, the use of optimally regularized spline functions (see Silverman, 1985, for a review) has been found to offer additional advantages, since they can also accomodate non-equidistantly sampled, non-uniformly weighted data, while exhibiting less boundary artefacts in the required derivatives if the **order** of the spline is choosen sufficiently high (Woltring 1985, 1986a).

Spline functions can be defined in terms of regularization and variation calculus. Given a set of N noisy measurements $\{y_n\}$ at known times $\{t_n\}$, positive weight factors $\{w_n\}$ which should be inversely proportional to the local noise variance of the measurements, and an as yet unspecified, positive **regularization** factor α, the function $s_\alpha(t)$ is sought which minimizes the quadratic criterion function

$$C_\alpha = \sum_{n=1}^{N} w_n\{y_n - s_\alpha(t_n)\}^2 + \alpha \int_{t_1}^{t_N} \left| s_\alpha^{(m)}(t) \right|^2 dt \qquad (10)$$

from the class of functions $\{s(t)\}$ which are continuous up to the (m-1)st derivative, and squared integrable in the m-th derivative on the interval $[t_1,t_N]$. One can prove that these functions are completely defined by the class of piecewise polynomials of order 2m which are continuous up to and including the (2m-2)nd derivative at selected **knot positions** in the measurement interval $[t_1,t_N]$. For the purposes of the present paper, these knots can be taken equal to the measurement times $\{t_n\}$; furthermore, certain boundary constraints must be imposed at the record ends. Depending on these boundary conditions, one can distinguish **periodic splines** with $s_\alpha^{(k)}(t_1) = s_\alpha^{(k)}(t_N)$ for k=0,2m-2, **complete splines** where the boundary derivatives for k=1,m-1 are included as measured constraints with prior given weights (Hu & Schumaker, 1986), and **natural splines** where the derivatives for k=m,2m-2 are set to zero. If the true data underlying the noisy measurements deviate strongly from the assumed boundary conditions, problems may occur **throughout the measurement interval,** especially from the lowest derivative with conflicting boundary conditions, and higher. However, the lower derivatives and the smoothing spline proper may be well-behaved.

Since the splines are completely defined in terms of piecewise polynomials of order 2m, they are called **splines of the order 2m.** It appears that they are continuous at the chosen knots at higher derivatives than m-1; however, these derivatives may fluctuate rather wildly. It is, in fact, advisable to select m higher than the highest derivative required.

It appears that the solution of (10) can be expressed in terms of a linear equation system. Denoting the raw measurements as a vector $\underline{y}$, and their spline-predicted values as a vector $\hat{\underline{y}}_\alpha$, they are related as

$$\hat{\underline{y}}_\alpha = H_\alpha \underline{y}, \qquad H_\alpha = B(B + \alpha W^{-1}E)^{-1} \qquad (11)$$

where H_α is the so-called **influence matrix** (in statistics also known as the "hat" matrix), B and E are certain design matrices defined by the measurement times $\{t_n\}$ and the spline order 2m, and $W = DIAG\{w_n\}$ is the matrix of weight factors; see Woltring (1986a). Furthermore, the spline model allows to evaluate $s_\alpha^{(k)}(t)$ for k=0,2m-1 and $t\epsilon[t_1,t_N]$.

For complete and natural splines, B and E can be constructed as band-limited matrices, and this renders the estimation problem numerically efficient. For periodic splines, at least for the equidistantly sampled, uniformly weighted case, the spline model allows an (approximate) treatment in the frequency domain using the Fast Fourier Transform which renders the matrices B and E purely diagonal as discussed below.

The question now arises how to select a suitable value for the regularization parameter α. For the uniformly weighted case, Craven & Wahba (1979) introduced the **Generalized Cross-Validation Criterion** (GCV) which can be defined as

$$GCV_\alpha = \frac{1}{N}\sum_{n=1}^{N}\{y_n - s_\alpha(t_n)\}^2 \,/\, TRACE^2\{\frac{1}{N}(I - H_\alpha)\} \tag{12}$$

In essence, cross-validation is based on fitting splines $s_{\alpha,n}(t)$ to all datapoints except for the n-th, calculating the predicted residuals at each omitted point, and finding the value $\alpha = \alpha_{OCV}$ minimizing the rms value of these residuals. It appears that this Ordinary Cross-Validation Criterion (OCV) is ill-behaved if the influence matrix is (nearly) diagonal; GCV corresponds to making H_α maximally non-diagonal via a prior orthonormal transformation in $\underline{y}$, $\underline{\hat{y}}_\alpha$, and H_α. Generalization to the weighted case is possible (Silverman, 1985).

The denominator of (12) can be interpreted as the square of the relative number of degrees of freedom Q_α in the smoothing process, and the numerator as the average value of the squared residuals. Thus, the GCV-criterion is seen to strike a balance between the residual variance $\hat{\sigma}_\alpha^2$ and Q_α,

$$\hat{\sigma}_\alpha^2 = \sum_{n-1}^{N}\{y_n - s_\alpha(t_n)\}^2 \,/\, Trace(I - H_\alpha), \quad Q_\alpha = \frac{1}{N}Trace(I - H_\alpha) \tag{13}$$

by minimizing the **ratio** of these quantities.

The residuals $\{y_n - s_\alpha(t_n)\}$ at the optimal value $\alpha = \alpha_{GCV}$ should be largely caused by the stochastic errors in the measurements $\{y_n\}$, rather than by systematic biases in the smoothing process. Thus, the GCV-function should be largely stochastic in its minimun, and the number of residual degrees of freedom at $\alpha = \alpha_{GCV}$ must be sufficiently large to ensure a stable, meaningful (i.e., smooth!) minimum. For multiple datasets as are typically encountered in the present rigid-body context, this stability can be enhanced by simultaneous processing of multiple datasets with identical influence matrices, possibly with additional weight factors per trajectory to accomodate different noise levels for identical signal characteristics (Woltring, 1986a). For example, the depth co-ordinate in photogrammetric movement reconstruction can be much noisier than the two other co-ordinates (Woltring **et al.**, 1985). The mean value of (12) for all these datasets will exhibit a larger number of degrees of freedom.

In Woltring (1986a), a general package for natural, smoothing splines has been presented. With Q_α the smoothing redundancy, Trace (H_α) can be seen as the number N_α of effectively estimated spline parameters, by analogy to conventional, parametric least-squares. For natural splines, N_α ranges between N (interpoliation, $\alpha=0$, with $H_\alpha=I$) and m ($\alpha=\infty$, i.e., zero derivative constraint in (10), when $s_\alpha(t)$ is an m-th order polynomial with m characterizing parameters). N_α is a useful quantity in the frequency-domain context discussed in the next section.

4. BUTTERWORTH-FILTER EQUIVALENCE OF THE SMOOTHING SPLINE

If the data are periodic and sufficiently oversampled, the class of functions $\{s(t)\}$ can be reasonable approximated by means of sinusoidal basis signals with frequencies well below the (average) Nyquist frequency, rather than by piecewise polynomials. For equidistantly sampled, uniformly weighted data, B and E in (11) become cyclical, and W becomes the identity matrix. Thus, (11) reduced to a convolution, with Discrete Fourier Transform (DFT),

$$\hat{Y}_\alpha(k) = H_\alpha(k).Y(k), \qquad k = 0,N-1 \tag{14}$$

with Y(k) and $\hat{Y}_\alpha(k)$ the DFT's of $\underline{y}$ and $\hat{\underline{y}}_\alpha$, respectively, and $H_\alpha(k)$ the DFT of the first row of H_α. Using Parceval's theorem, the criterion function (10) can be transformed to the discrete frequency domain as

$$C_\alpha = \frac{1}{N}\sum_{k=0}^{N-1} \left|\{1 - H_\alpha(k)\}Y(k)\right|^2 + \alpha\tau.\omega_k^{2m}\left|H_\alpha(k).Y(k)\right|^2 \tag{15}$$

where τ is the sampling interval, and $\omega_k = (2\pi/N\tau)(k-\frac{1}{2}N)$ the discrete radial frequency. Bij means of a variational argument, the optimal $H_\alpha(k)$ minimizing (15) can be derived as

$$H_\alpha(k) = \frac{1}{1 + (\omega_k/\omega_\alpha)^{2m}}, \qquad \omega_\alpha = (\alpha\tau)^{-1/2m} \tag{16}$$

This relation can be recognized as the discrete transfer function of two cascaded, m-th order Butterworth filters with cut-off frequency ω_α, without phase and frequency distortion (cf. Oppenheim & Schafer, 1975). The effective number of spline parameters N_α can be evaluated as

$$N_\alpha = \sum_{k=0}^{N-1} H_\alpha(k) \approx k_m.\omega_\alpha.\tau.N \qquad \text{if} \quad \omega_\alpha\tau << \pi \tag{17a}$$

with

$$k_m = \frac{1}{\pi}\int_0^\infty (1 + x^{2m})^{-1}dx \downarrow \frac{1}{\pi} \qquad \text{if } m \to \infty \tag{17b}$$

For example, $k_1 = 1/2$, $k_2 = 1/\sqrt{8}$, and $k_3 = 1/3$. A similar relation has been provided by Craven & Wahba (1979). Empirically, the following relation has been found to hold through experiments with the spline package of Woltring (1986a),

$$N_\alpha \approx m/2 + k_m.\omega_\alpha.\tau.N, \qquad m < N_\alpha << N \tag{18}$$

both for equidistantly and for slightly non-equidistantly sampled data. This suggests that m/2 datapoints are used by the natural spline to accomodate boundary effects. For the remaining datapoints, the natural spline behaves approximately as a periodic spline.

The m-th order, double Butterworth filter is often applied in a non-periodic context. First, the data are processed with two cascaded, recursive time-domain implementations with the second filter in reverse time in order to cancel phase distortion. Subsequently, the derivatives are estimated from finite differences. The equivalence with non-periodic splines explains why m-th derivatives estimated in this way are very sensitive to the selected initialization conditions of the recursive filters, and it is advisable to select m higher than the highest derivative being sought (cf. Woltring, 1985).

For IHA estimation, no higher derivatives than the first are required, so cubic, natural splines and 2nd-order Butterworth filters are appropriate. however, the situation is different if also the **2nd-order IHA** (Suh & Radcliffe, 1978) is required, for assessing the pivoting behaviour of the IHA: in a finite displacement context, Fischer (1907) and Chao & An (1982) have advocated the utility of the 3-D instantaneous pivot or **central point** about which the IHA changes its direction, and of the trajectory formed by this point.

Under a frequency domain interpretation, the residual variance $\hat{\sigma}_{\alpha}^{2}$ can be viewed as the mean residual power per frequency component stopped by the filter, and Q_{α} as the relative stopping bandwidth. If the measurements exist of a low-pass signal with additive, white noise, the effect is that GCV tries to find the lowest frequency for which the residual noise is white; this criterion is slightly more conservative than the optimal cut-off frequencies defined by some authors in Biomechanics (Jackson, 1979; Wells & Winter, 1980). Thus, the GCV-spline can be seen as a familiar tool in a new guise, with the additional advantage of also accomodating interpolation and non-equidistantly sampled data. However, large datagaps should be avoided since they are interpolated by local polynomials of order 2m or less, and their amplitudes may become inordinately large.

5. IHA ESTIMATION RESULTS FROM A KNOWN MOVEMENT

The spline smoothing procedure discussed above has been applied to a set of marker co-ordinates collected via Röntgenstereophotogrammetry from a known movement (de Lange **et al.**, 1986). An isotropic distribution of 8 landmarks ($\rho \approx 7$ mm, cf. Woltring et al, 1985) was moved around a fixed rotation axis, in 180 steps of about 1° each. The rotation axis was oriented into the x-direction, and the photogrammetric reconstruction errors of the landmarks had standard deviations of about 100 µm in the x-direction, and of about 40 µm in the y- and z-directions.

For the smoothing procedure (Woltring, 1986a), the independent "time" variable was defined as the cumulated, finite helical rotations estimated from the unsmoothed data by means of the procedures outlined in Woltring **et al.** (1985). Subsequently, the raw landmark co-ordinates were processsed with the spline package, as a function of the effective number of spline parameters N_{α}. A quintic spline (m=3) was choosen in view of future use of the 2nd-order IHA, and the data were processed in the "simultaneous" mode, with uniform weighting over time, and weight factors per co-ordinate inversely proportional to the position variances. In a pilot study, the signal-bearing co-ordinates (y and z) were found to possess very similar α_{GCV} values, while the x-coordinates contained almost pure noise plus a constant offset.

The 1° rotation increments were sufficiently small to render FHA's as calculated from the smoothed data suitable estimators for a discrete set of IHA's; therefore, the smoothed data were directly passed through the FHA procedure, and sample statistics were assessed for both sets of

FHA's. In other situations, the smoothed data should be densified between the measurement times by interpolation. Note that the small-angle noise efffect of the FHA (Woltring **et al.**, 1985) presupposes uncorrelated noise between times; this assumption is no longer met after smoothing and interpolating the raw data.

Since the movement was about a fixed but numerically unknown axis, dispersion measures for the positions and directions of the IHA's with respect to a **Mean Helical Axis** (MHA, cf. Woltring, 1986b) provide an indication of the total precision of the IHA estimation procedure. In essence, the MHA follows by finding a mean rotation "pivot" of all IHA's, defined as the point with the smallest rms distance SD_s to the IHA's. Subsequently, an optimal direction vector is defined through this pivot wich has minimal rms distance of an arbitrary point on the direction vector to the IHA's. This is equivalent to minimizing the rms value of the sinuses of the angles between the MHA and the IHA's, and the angular dispersion SD_n is defined as the arcsine of this rms value. The difference with a truely angular rms value is negligible, and the solution follows from an eigenvalue problem which plays also a role in the calculation of the mean pivot.

In Figure 2, the results of this study are shown. The equivalent Butterworth cut-off frequency BWfre exhibits the behaviour of relation (18) except for the case of maximal smoothing ($\alpha \rightarrow \infty$, i.e..,$N_\alpha \rightarrow$ m=3) where the data are fitted by parabolas. The GCV-function exhibits at its minimum the smooth behaviour to be expected for a sufficiently overdetermined problem. The optimal value is located at $N_\alpha = 8.5$, and the strong increase in the GCV-function for smaller values of N_α follows from the inappropriateness of least-squares fitting parabolas to the 180° sinusoidal arcs subtended by the y- and z-coordinates as a function of time. The white-noise nature of the measurement errors is borne out by the slowly increasing character of the GCV-function for $N_\alpha > 8.5$, and this corresponds with a horizontal path of the residual variance function (not plotted).

The position and direction dispersions were found to be 34 µm and 0.078° at the GCV-minimum. In contrast, their values were 2.75 mm and 12° in the unsmoothed case, i.e., 81 and 154 times as large, respectively. Thus, the small-angle noise effect discussed by Woltring **et al.**, (1985) could be effectively abolished.

It appears that some oversmoothing would be acceptable, since SD_s and SD_n attain minimal values at N_α=4.5. Apparently, reducing the cut-off frequency has a beneficial effect in the derivative domain which largely determines the error sensitivity of the IHA. This is understandable since high-frequency noise has a strong effect, while low frequency signal components are strongly attenuated during differentiation. Thus, the GCV-criterion can be said to smooth as little as possible, at least if the data are sufficiently redundant and if they meet the spline's model properties of a stationary, low-pass signal with additive, white noise. In biological joint motion studies, this may be a reasonable assumption since the opposite might entail large, potentially dangerous inertial forces.

6. ACKNOWLEDGEMENTS

The authors are indebted to mrs. I. de Bruin and mr. J. de Vocht for digitizing 180 Röntgen stereophotographs.

7. REFERENCES

1. Anderssen RS, Bloomfield P: Numerische Mathematik 22, 157-182, 1974.
2. Chao EYS and An KN: Biomechanics: Principles and Applications (R. Huiskes **et al.**, Eds). Nijhoff, The Hague (The Netherlands), pp. 1-18, 1982.

3. Craven P & Wahba G: Numerische Mathematik 31, 377-403, 1979.
4. Fischer O: Kinematik Organischer Gelenke. Vieweg, Braunschweig (Germany), 1907.
5. Hatze H: Journal of Biomechanics 14(1), 13-18, 1981.
6. Hu CL & Schumaker LL: Numerische Mathematik 49(1), 1-10, 1981.
7. Jackson KM: IEEE Transactions on Biomedical Engineering BME-26(2), 122-124, 1979.
8. Lange A de, Huiskes R, Woltring HJ & Kauer JMG: Effects of data smoothing on the reconstruction of kinematic parameters (in preparation).
9. Morozov VA: Methods for solving incorrectly posed problems. Springer, New York, 1984.
10. Oppenheim AV & Schafer RW: Digital Signal Processing. Prentice-Hall, Englewood Cliffs (NJ, USA) 1975.
11. Silverman BW: Journal of the Royal Statistical Society B 47(1), 1-52, 1985.
12 Suh CH & Radcliffe CW: Kinematics and Mechanisms Design. Wiley, New York, 1978.
13. Wells RP & Winter DA: Proceedings Human Locomotion I. Can. Soc. of Biomechanics (Ottawa), pp. 92-93, 1980.
14. Woltring HJ: Human Movement Science 4(3), 229-245, 1985.
15. Woltring HJ: Advances in Engineering Software 8(2), 104-113, 1986a.
16. Woltring HJ: Model and measurement error influences in kinematics. In: Proc. Study Institute and Conference on Biomechanics of Human Movement. Formia, (Italy); 1986b (in print).
17. Woltring HJ, Lange A de, Huiskes R & Veldpaus FE: Journal of Biomechanics 18(5), 379-389, 1985.

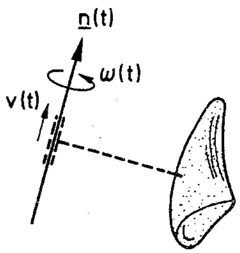

FIGURE 1:
Instantaneous Helical Axis.

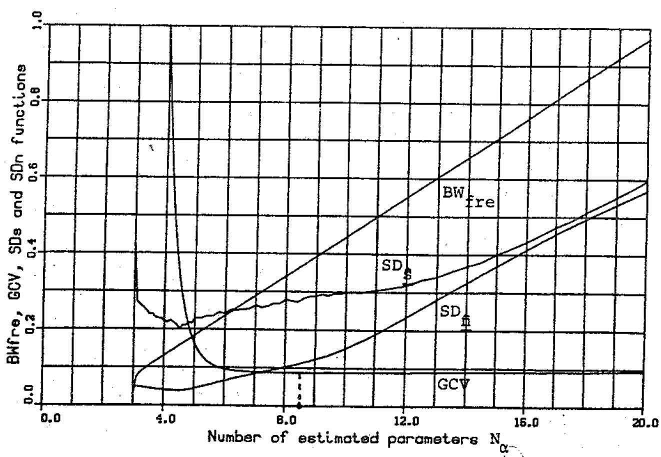

FIGURE 2:
IHA smoothing results for a quintic, natural spline, with position dispersion SD_s, direction dispersion SD_n, GCV-function GCV, and equivalent Butterworth cut-off frequency BW_{fre}, as a function of N_α. SD_s, SD_n, and BW_{fre} were normalized w.r.t. arbitrary scaling values.