Flexible and reliable profile estimation using exponential splines

R. Fischer* and V. Dose*

*Max-Planck-Institut für Plasmaphysik, EURATOM Association,
Boltzmannstr. 2, D-85748 Garching, Germany

Abstract. Flexible and reliable non-parametric distribution estimation is achieved by using exponential splines. In Bayesian function estimation the number of spline knots as well as the parameters for knot position, amplitude and stiffness are marginalized. The resulting marginal posterior probability distribution allows to estimate profiles, profile gradients and their uncertainties in a natural way.

Key Words: Exponential spline, non-parametric distribution estimation, Integrated Data Analysis (IDA), Occam's razor, Markov Chain Monte Carlo (MCMC) sampling

INTRODUCTION

Reliable profile and profile gradient estimates are of utmost importance for many different physical models in fusion science, e.g. transport modeling. The results often crucially depend on the functional representation of the profile. The estimation uncertainty of the profile and, in particular, the estimation of the profile gradient and its uncertainty is closely coupled with the provided profile flexibility. Flexibility is frequently obtained by using non-parametric profile functionals, e.g. linear interpolation between pointwise estimations or cubic or B-splines. Profile flexibility to allow for a form-free description of the data often competes with profile reliability. As the number of degree-of-freedom (DOF) increases the estimation reliability decreases. Reliability is frequently obtained in plasma physics profile estimation by either providing a family of tailored parametric functionals or piecewise polynomial functions combined with modified hyperbolic tangent functions (tanh). The aim is to have a robust technique to allow for a reasonable balance between flexibility and reliability in order to achieve balance between modeling the significant information content in the data and avoiding noise fitting.

EXPONENTIAL SPLINES

Consider a set of function values f_i given at E support points ξ_i (knots). The exponential spline function $S_i(x)$ in the interval $\xi_i \leq x \leq \xi_{i+1}$ is then given by [1]

$$S_i(x) = \alpha_i + \beta_i(x-\xi_i) + \gamma_i \psi_i(x-\xi_i) + \delta_i \phi_i(x-\xi_i). \qquad (1)$$

The auxiliary functions ψ_i and ϕ_i contain a stiffness parameter λ_i on the support $[\xi_i, \xi_{i+1}]$ and are given by the hyperbolic functions

$$\psi_i(x-\xi_i) = 2\{\cosh[\lambda_i(x-\xi_i)]-1\}/\lambda_i^2 \quad (2)$$
$$\phi_i(x-\xi_i) = 6\{\sinh[\lambda_i(x-\xi_i)]-\lambda_i(x-\xi_i)\}/\lambda_i^3 \quad (3)$$

From the series expansions of the hyperbolic functions we obtain the two limiting cases of a cubic spline ($\lambda \to 0$) and a linear interpolation ($\lambda \to \infty$) [1]. Since the stiffness parameters λ_i are allowed to vary over the intervals $[\xi_i, \xi_{i+1}]$ the character of the exponential spline function might vary from linear to third order polynomial on adjacent support intervals which provides extremely high flexibility.

The so far unknown coefficients $\alpha, \beta, \gamma, \delta$ are determined from the requirement of continuity of function, first and second derivatives at the knot positions ξ_i. Continuity of function and second derivative yields already an explicit representation of the exponential spline function in terms of function values $\{f_i\}$ and second derivatives $\{M_i\}$ at the knot positions $\{\xi_i\}$. Introducing the definitions $h_i = \xi_{i+1} - \xi_i$, $z_i = \lambda_i(x - \xi_i)$, and $\mu_i = \lambda_i h_i$, we obtain

$$\begin{aligned} S_i(x) &= \frac{\xi_{i+1}-x}{h_i} f_i + \frac{x-\xi_i}{h_i} f_{i+1} \\ &+ \frac{M_i}{\lambda_i^2}\left\{\frac{\sinh(\mu_i - z_i)}{\sinh(\mu_i)} + \frac{z_i}{\mu_i} - 1\right\} + \frac{M_{i+1}}{\lambda_i^2}\left\{\frac{\sinh(z_i)}{\sinh(\mu_i)} - \frac{z_i}{\mu_i}\right\} \end{aligned} \quad (4)$$

The terms involving the function values f_i and f_{i+1} represent the linear interpolation part of $S_i(x)$. The terms involving the second derivatives M_i and M_{i+1} introduce the curvature. In order to determine the so far unknown second derivatives $\{M_i\}$ in terms of the function values $\{f_i\}$ we use finally the continuity requirement for the first derivative. This yields the system of equations

$$\begin{aligned} & M_{i-1} h_{i-1} \frac{\sinh(\mu_{i-1}) - \mu_{i-1}}{\mu_{i-1}^2 \sinh(\mu_{i-1})} + \\ & M_i \left\{ h_{i-1} \frac{\mu_{i-1}\cosh\mu_{i-1} - \sinh\mu_{i-1}}{\mu_{i-1}^2 \sinh\mu_{i-1}} + h_i \frac{\mu_i \cosh\mu_i - \sinh\mu_i}{\mu_i^2 \sinh\mu_i} \right\} \\ & M_{i+1} h_i \frac{\sinh\mu_i - \mu_i}{\mu_i^2 \sinh\mu_i} = \frac{f_{i+1}-f_i}{h_i} - \frac{f_i - f_{i-1}}{h_{i-1}} \end{aligned} \quad (5)$$

For E knots this is a system of $E-2$ equations. The system can be closed by putting $M_1 = M_E = 0$ or by given values of the first derivative at the end points.

To estimate profiles and profile gradients from noisy data $d_i = d(x_i)$ it is useful to have the linear representation of the exponential spline as a function of the stiffness parameters $\vec{\lambda}$ at N_p positions $\vec{x}$, e.g. at the data abscissae,

$$S(\vec{x}) = W(\vec{x}, \vec{\lambda}, \vec{\xi}) \vec{f} \quad (6)$$

The $(N_P \times E)$ matrix W can be separated into two parts W_1 and W_2. W_1 represents the coefficients of f in (4). W_2 is obtained by multiplying the coefficients of M with the solution of the system (5) including the two additional constraints chosen. For numerical stability approximations have to be applied for large as well as for small values of λ. The profile gradient is straightforwardly calculated from analytical derivatives of W with respect to x, $S'(\vec{x}) = W'(\vec{x}, \vec{\lambda}, \vec{\xi})\vec{f}$.

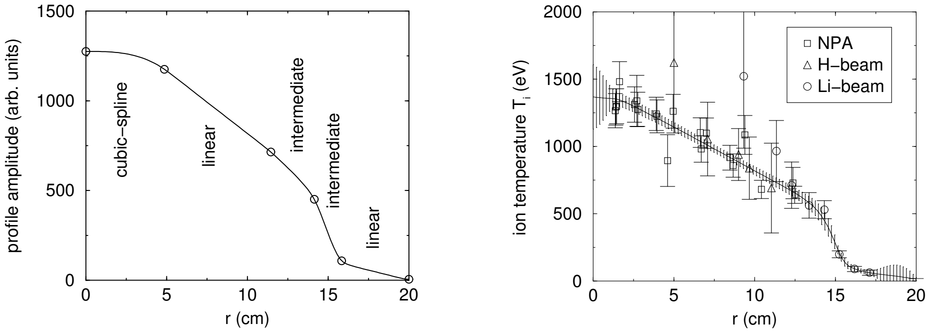

FIGURE 1. Left: Sample of an exponential spline with 6 knots. The stiffness parameters $\lambda_i, i = 1-5$, determine if the exponential spline segments are similar to a cubic spline, to a linear curve or if it has intermediate properties. Right: Ion temperature profile and profile gradient marginalized over all number of knots.

The left panel of figure 1 depicts a typical exponential spline with heterogeneous properties in its segments. The 5 segmental stiffness parameters between 6 spline knots determine if the exponential spline is similar to a cubic spline, to a linear segment or if it has intermediate properties.

THE BAYESIAN FRAMEWORK

In our Bayesian approach we focus on the probability of the profile having a value S_j at any position x_j represented by $p(S_j|\vec{d}, \mathcal{M}, \mathcal{I})$. This posterior probability depends on the full data set $\vec{d}$, a model $\mathcal{M}$ for the profile functional to be used and all relevant information $\mathcal{I}$ concerning the nature of the physical situation and knowledge of the experiment. $\mathcal{I}$ includes knowledge about the noise level of the experimental measurements, additional knowledge about the profile or profile gradient, e.g. positivity constraints, physical constraints resulting in strictly monotonic profiles or maximum gradient values from stability criteria. All these specifications might play a crucial role since they provide information that restricts the profiles to physically sound solutions.

Equation (6) allows us to focus on ($\vec{f}$, $\vec{\lambda}$, $\vec{\xi}$) as the fundamental set of parameters to be estimated. According to Bayes theorem the posterior probability for ($\vec{f}$, $\vec{\lambda}$, $\vec{\xi}$) is

$$p(\vec{f}, \vec{\lambda}, \vec{\xi}|\vec{d}, \vec{\sigma}, E, \mathcal{I}) = \frac{p(\vec{d}|\vec{f}, \vec{\lambda}, \vec{\xi}, E, \mathcal{I})\, p(\vec{f}, \vec{\lambda}, \vec{\xi}|E, \mathcal{I})}{p(\vec{d}|E, \mathcal{I})} \quad . \tag{7}$$

The number of knots E are given explicitely since it is a model parameter effecting the fitting properties. The denominator (evidence of the data)

$$p(\vec{d}|E, \mathcal{I}) = \int d^E f\; d^{E-1}\lambda\; d^{E-2}\xi\; p(\vec{d}|\vec{f}, \vec{\lambda}, \vec{\xi}, E, \mathcal{I})\, p(\vec{f}, \vec{\lambda}, \vec{\xi}|E, \mathcal{I}) \tag{8}$$

guarantees that the posterior is normalized. In our adaptive model the evidence plays a central role in determining the number of spline knots E.

The Likelihood

The likelihood of the experimental data, $p(\vec{d}|\vec{\sigma},\vec{f},\vec{\lambda},\vec{\xi},E)$, quantifies the probability of measuring the data set $\vec{d}$, given their uncertainties $\vec{\sigma}$ and given the profile parameters $\vec{f},\vec{\lambda},\vec{\xi}$ of E spline knots. The data analyzed in this work are given by spatially resolved profile measurements from various diagnostics [2]. Since the underlying level of uncertainty of the data is frequently difficult to estimate in plasma physics, relative uncertainties are often reasonably described but the absolute value might be subject of discussion. To allow for flexibility in the absolute scale of the uncertainties a factor s_k is introduced which scales the uncertainties of data set $\vec{d_k}$ measured/derived from diagnostic k. Within a diagnostic the scaling factor of the errors are assumed to be unique whereas they might differ between different diagnostics. A value of $s_k < 1$ means that the diagnostician has overestimated the uncertainty ("conservative") whereas a value of $s_k > 1$ means that the error was underestimated (maybe by neglecting systematic error sources). The uncertainty scaling parameters s_k are often useful when within an Integrated Data Analysis (IDA) approach [3] the data from heterogeneous diagnostics have to be combined. If the analysis of the individual diagnostics data would comprise the correct description of the measurement and the physical model, and if all sources of measurement (statistical and systematic) errors are considered in the likelihood, then the scaling parameters s_k would not be needed. The nuisance parameters s_k can be estimated or marginalized.

The likelihood for the present data from profile measurements is assumed to be Gaussian with independent normally distributed uncertainties. Assuming independent uncertainties the total likelihood is the product over all likelihoods for N_k data $\vec{d_k}$ derived from diagnostic k with uncertainty scaling factor s_k

$$p(d_{ik}|\sigma_{ik},s_k,\vec{f},\vec{\lambda},\vec{\xi},E,\mathcal{I}) = \frac{1}{\sqrt{2\pi(s_k\sigma_{ik})^2}} \exp\left\{-\frac{(d_{ik}-S_{ik})^2}{2(s_k\sigma_{ik})^2}\right\} \tag{9}$$

where S_i is the exponential spline value calculated with parameter set $(\vec{f},\vec{\lambda},\vec{\xi},E)$.

The prior probabilities

The prior pdf, $p(\vec{f},\vec{\lambda},\vec{\xi},\vec{s}|E)$, constitutes information we have about the parameters independent of the measured data. The uncertainty scaling factors $\vec{s}$ used in the likelihood pdf adds to the specified parameter list. According to the product rule of Bayesian probability theory the prior can be split into the individual parts

$$p(\vec{f},\vec{\lambda},\vec{\xi},\vec{s}|E) \quad = \quad p(\vec{f}|E)\,p(\vec{\lambda}|E)\,p(\vec{\xi}|E)\,p(\vec{s}) \tag{10}$$

where the symbol $\mathcal{I}$ is omitted for practical reasons. The prior for the knot amplitudes, $p(\vec{f}|E)$, was chosen to be constant for positive values below a reasonable upper limit and zero elsewhere

$$p(\vec{f}|E) = \begin{cases} \frac{1}{V_f} = \prod_{j=1}^{E} \frac{1}{f_{\max}}\,; & 0 \leq f_j \leq f_{\max} \\ 0 \quad \text{elsewhere} & \end{cases} \tag{11}$$

The prior for the stiffness parameters was chosen to be Jeffrey's prior since λ is a scale parameter

$$p(\vec{\lambda}|E) = \begin{cases} \prod_{j=1}^{E-1} \ln\left(\frac{\lambda_{\max}}{\lambda_{\min}}\right) \frac{1}{\lambda_j}\,; & \lambda_{\min} \leq \lambda_j \leq \lambda_{\max} \\ 0 \quad \text{elsewhere} & \end{cases} \tag{12}$$

where the boundaries of λ are chosen to allow both liming cases of cubic splines and polygon interpolation. For numerical benefits it is useful to calculate in terms of the logarithm of lambda. The equivalent prior for the logarithm of lambda is a constant prior between the boundaries and zero elsewhere.

The prior for the knot positions assumes that the positions are ordered, that a minimum distance between neighboring positions is given and that there has to be at least one data point between neighboring positions. The end point positions are set to be at the plasma center $\xi_1 = 0$ and at the plasma edge $\xi_E = r_{\max}$, respectively. A noncommittal prior is given by the uniform prior taking into account the minimum spacing $\Delta\xi$ and the required ordering of the knot positions ($\xi_1 + \Delta\xi \leq \xi_2; \quad \xi_2 + \Delta\xi \leq \xi_3; \cdots ; \xi_{E-1} + \Delta\xi \leq \xi_E$) [4]. The prior on $\vec{\xi}$ is $p(\vec{\xi}|E, \Delta\xi) = Z^{-1} \prod_{k=2}^{E} \theta[\xi_{k-1} + \Delta\xi \leq \xi_k]$, where the function θ is one when its argument conditions are true and zero otherwise. The normalization integral

$$Z = \int_{\xi_1+\Delta\xi}^{\xi_E-(E-2)\Delta\xi} d\xi_2 \int_{\xi_2+\Delta\xi}^{\xi_E-(E-3)\Delta\xi} d\xi_3 \cdots \int_{\xi_{E-2}+\Delta\xi}^{\xi_E-\Delta\xi} d\xi_{E-1} \tag{13}$$

is easily calculated, resulting in

$$p(\vec{\xi}|E, \Delta\xi) \;=\; \frac{(E-2)! \prod_{k=2}^{E} \theta[\xi_{k-1} + \Delta\xi \leq \xi_k]}{[\xi_E - \xi_1 - (E-1)\Delta\xi]^{(E-2)}} \quad . \tag{14}$$

The denominator is simply the total volume of space of the $(E-2)$ parameters ξ_ν. The factorial in the numerator accounts for the ordering requirement. The minimum distance $\Delta\xi$ is chosen small enough to allow flexible profile structures and large enough to avoid position degeneration. Additionally, the prior for the knot positions is set to zero for all settings where no data point is between any two neighboring positions.

The prior for the uncertainty scaling factors $\vec{s}$ is chosen to be Jeffrey's prior

$$p(\vec{s}|E) = \begin{cases} \prod_{k=1}^{N_{\text{diag}}} \ln\left(\frac{s_{\max}}{s_{\min}}\right) \frac{1}{s_k}\,; & s_{\min} \leq s_k \leq s_{\max} \\ 0 \quad \text{elsewhere} & \end{cases} \tag{15}$$

because s is a scaling parameter. N_{diag} is the number of data sets from different diagnostics used. The boundaries $(s_{\min}, s_{\max})$ reflect the credibility we assign to the estimation of the scale of the uncertainties. If detailed prior information about the scale is present an alternative prior is given by the Gamma-distribution.

For comparison of models with different numbers of spline knots E and for model marginalization over E we need the prior $p(E)$. This prior is chosen to be uniform for all integer values of E between the minimum number, $E_{min} = 2$, and the maximum number, $E_{max} = \text{integer}[(\xi_E - \xi_1)/\Delta\xi] + 1$, namely $p(E) = [E_{max} - E_{min} + 1]^{-1}$ and zero elsewhere.

MCMC sampling of the posterior and number of knots

The posterior probability distribution (7) describes the full solution of our profile estimation problem. Single estimates of the profile, the profile gradient and its uncertainties can be derived from the maximum and variance of the posterior (maximum-a-posteriori, MAP solution) or from the mean value and variance of the marginal

$$p(\vec{f}|\vec{d},\vec{\sigma},E) = \int d\vec{\xi}\, d\vec{\lambda}\, d\vec{s}\, p(\vec{f},\vec{\lambda},\vec{\xi},\vec{s}|\vec{d},\vec{\sigma},E)\,. \tag{16}$$

The mean value and variance of (16), and of the marginals of $\vec{\lambda}$, $\vec{\xi}$, and $\vec{s}$ are estimated using a Markov Chain Monte Carlo (MCMC) technique.

The most probable number of knots is calculated applying the Bayesian theorem again

$$p(E|\vec{d},\vec{\sigma}) = \frac{p(E)\,p(\vec{d}|\vec{\sigma},E)}{p(\vec{d}|\vec{\sigma})} \tag{17}$$

where $p(E)$ is the prior on E specified above and $p(\vec{d}|\vec{\sigma})$ is a normalization constant which can be determined from $\sum_E p(E|\vec{d},\vec{\sigma}) = 1$. The marginal likelihood $p(\vec{d}|\vec{\sigma},E)$ quantifies the probability of the data $\vec{d}$ marginalized over the total parameter space:

$$p(\vec{d}|\vec{\sigma},E) = \int d\vec{f}\, d\vec{\lambda}\, d\vec{\xi}\, d\vec{s}\, p(\vec{d},\vec{f},\vec{\lambda},\vec{\xi},\vec{s}|\vec{\sigma},E) \tag{18}$$

$$= \int d\vec{\lambda} p(\vec{\lambda}|E) \int d\vec{\xi} p(\vec{\xi}|E) \int d\vec{s} p(\vec{s}) \quad I_f(\vec{\lambda},\vec{\xi},\vec{s}) \tag{19}$$

$$I_f(\vec{\lambda},\vec{\xi},\vec{s}) = \int d\vec{f} p(\vec{d}|\vec{f},\vec{\lambda},\vec{\xi},\vec{\sigma},\vec{s},E)\, p(\vec{f}|E) \tag{20}$$

The integral over the spline amplitudes $\vec{f}$ can be calculated analytically because the likelihood is Gaussian and can be written as

$$p(\vec{d}|\vec{f},\vec{\lambda},\vec{\xi},\vec{\sigma},\vec{s},E) = \frac{1}{\prod_{ik}\sqrt{2\pi(\sigma_{ik}s_k)^2}} \exp\left\{-\frac{1}{2}(\vec{d}-W\vec{f})^T\,\Sigma\,(\vec{d}-W\vec{f})\right\} \tag{21}$$

where Σ is the inverse covariance matrix with the diagonal elements $1/\sigma_{ik}^2$. Assuming that the prior $p(\vec{f}|E)$ is broader than the likelihood the integration bounds can be extend to $\pm\infty$ and the integral calculates to

$$I_f(\vec{\lambda},\vec{\xi},\vec{s}) = \frac{1}{V_f}(2\pi)^{\frac{E-N_d}{2}}\sqrt{\frac{\det(\Sigma)}{\det(Q)}}\exp\left\{-\frac{R}{2}\right\} \tag{22}$$

where V_f is the prior volume of the amplitudes $\vec{f}$. The matrix Q and the scalar R follow from a comparison of the coefficients of $\vec{f}$

$$Q = W^T\Sigma W \tag{23}$$

$$R = \vec{d}^T\Sigma\left\{1-WQ^{-1}W^T\Sigma\right\}\vec{d} \quad . \tag{24}$$

The integration over the stiffness parameters $\vec{\lambda}$, the knot positions $\vec{\xi}$, and the uncertainty scaling factors $\vec{s}$ is calculated using simple sampling as follows: First, a vector $\vec{\lambda}_l$ is sampled from the prior $p(\vec{\lambda}|E)$. Equivalent to sampling λ from Jeffrey's prior, the logarithm of λ is sampled from the uniform distribution. Second, a vector $\vec{s}_l$ is sampled from the prior $p(\vec{s}|E)$. Third, unsorted values $\vec{\xi}_l$ were sampled from the prior $p(\vec{\xi}|E)$. Then, the values $\xi_{i,l}$ were sorted in ascending order. The sorted values $\xi_{i,l}$ were accepted if at least one data point lays between neighboring knots and if all neighboring knots are more than the minimum distance separated, else a new set of values $\vec{\xi}_l$ is sampled. Then, $I_f(\vec{\lambda}_l,\vec{\xi}_l,\vec{s}_l)$ is calculated with the sampled $\vec{\lambda}_l$, $\vec{\xi}_l$, and $\vec{s}_l$-values. The value $p(\vec{d}|\vec{\sigma},E)$ and its simple sampling uncertainty were estimated from the mean value $\langle I_f\rangle$ and its variance.

The marginal posterior distribution of the profile $\vec{f}$ independent on the number of knots provided is given by

$$p(\vec{f}|\vec{d},\vec{\sigma}) = \sum_E p(\vec{f}|\vec{d},\vec{\sigma},E)\,p(E|\vec{d},\vec{\sigma}) \,. \tag{25}$$

The terms in the sum are calculated from (16) and (17).

RESULTS

The right panel of figure 1 shows data sets from three different experiments [2] and an estimate of the ion temperature profile. The solid line and the error bars represent the mean value and the ± 1 standard deviation of the marginal posterior probability distribution (25). The shape of the exponential spline is close to linearity for $r < 13$ cm and $r > 17$ cm, and close to a cubic spline in between. The error bars depict a large variability close to the plasma center and at the plasma edge where the data are sparse.

The left panel of figure 2 shows an estimate of the profile gradient calculated from the MCMC samples of the analytic derivative of the exponential spline. The gradient profile is well determined at the linear region where the number of data is large. The gradient error bars are larger close to the plasma center and in the plasma edge region where the data are sparse. Due to the methodologically inherent competition between flexibility and reliability the approach provides a reliable tool for gradient estimation.

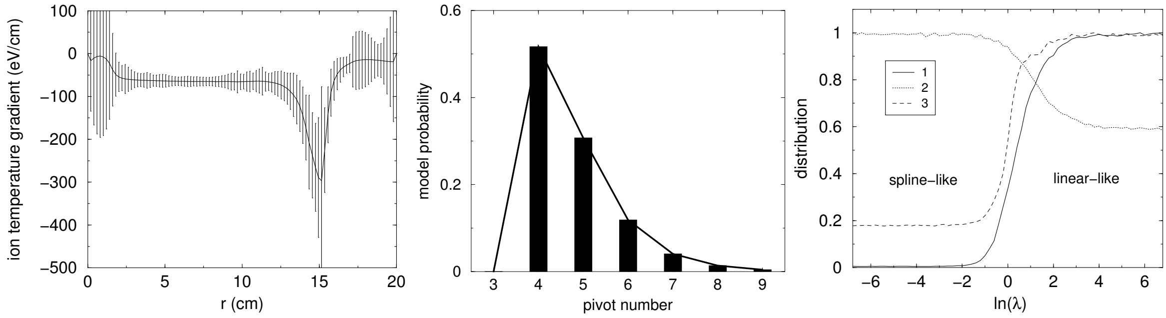

FIGURE 2. Ion temperature profile gradient marginalized over all number of knots. Model probability for different number of exponential spline knots. MCMC distribution of 3 stiffness parameters λ belonging to the 3 intervals between 4 knots. The exponential spline is linear like (polygon) for small values of λ and cubic-spline like for large values of λ, respectively.

The middle panel of figure 2 shows the marginal posterior of the number of spline knots E (17). 4 spline knots are sufficient to fit the data which is also confirmed by the marginal distribution of the deviance (not shown here).

The right panel depicts the probability distribution (not normalized) of the logarithm of three stiffness parameters belonging to the three intervals between 4 knots. The probability for λ_1, $p(\lambda_1|\vec{d},\vec{\sigma})$, is large for values $\lambda > 0$ resembling the linear behavior for $r < 13$ cm. The small values for λ are ruled out completely showing that the linear region can not be fitted reasonably by a cubic spline. In contrast to the first interval, λ_2 shows a completely different behavior. The second interval between knot 2 and 3 is best fitted by a cubic spline with small values for λ. The large values for λ are not ruled out completely showing that the pedestal region can reasonably be fitted also with a polygon mainly if the knot positions are sufficiently close. The edge region is dominated again by a linear behavior if the knot position is sufficiently distant from the gradient region. If the knot position is close to the gradient region spline-like features become more important.

In conclusion, the non-parametric exponential-spline approach for profile and profile gradient estimation provides a robust method for a useful balance between flexibility and reliability. Uncertainties of profiles and gradients is readily derived from data uncertainties. The DOF is determined by the significant information content of the data.

REFERENCES

1. V. Dose and R. Fischer. Function estimation employing exponential splines. In K.H. Knuth, A.E. Abbas, R.D. Morris, and J.P. Castle, editors, *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, volume Conf. Proc. 803, pages 67–71, Melville, NY, 2005. AIP.
2. H. Ehmler, Y. Turkin, C.D. Beidler, H. Maassberg, A. Dinklage, T. Klinger, and W7-AS Team. Experimental check of neoclassical predictions for the radial electric field in a stellarator. *Nucl. Fusion*, 43:L11–L13, 2003.
3. R. Fischer and A. Dinklage. Integrated data analysis of fusion diagnostics by means of the Bayesian probability theory. *Rev. Sci. Instrum.*, 75:4237–4239, 2004.
4. R. Fischer, K. M. Hanson, V. Dose, and W. von der Linden. Background estimation in experimental spectra. *Phys. Rev. E*, 61:1152, 2000.