Methods for optimizing and assessing diagnostic capability, demonstrated for collective Thomson scattering (invited)

H. Bindslev

Citation: Review of Scientific Instruments **70**, 1093 (1999);
View online: https://doi.org/10.1063/1.1149521
View Table of Contents: http://aip.scitation.org/toc/rsi/70/1
Published by the American Institute of Physics

Methods for optimizing and assessing diagnostic capability, demonstrated for collective Thomson scattering (invited)

H. Bindslev
JET Joint Undertaking, Abingdon, Oxfordshire OX14 3EA, United Kingdom

(Presented on 11 June 1998)

Commonly diagnostic systems are sensitive to a wide range of parameters, only a subset of which are *parameters of interest*. Optimal analysis and assessment of diagnostic capability must take account of the sensitivity to the remaining *nuisance parameters* which may not be accurately known. New least squares type methods of inference in the presence of uncertain nuisance parameters have been developed, along with methods for assessing diagnostic capability and identifying critical nuisance parameter uncertainties. These methods are of a general nature and suitable for complex and computationally demanding forward models. They are demonstrated by application to collective Thomson scattering for fast ion diagnosis in fusion plasmas. © *1999 American Institute of Physics.* [S0034-6748(99)71101-9]

I. INTRODUCTION

In magnetically confined fusion plasmas collective Thomson scattering (CTS) has particular capabilities for spatially localized diagnosis of the velocity distribution of fast ions. Parameters characterizing this quantity we refer to as *parameters of interest*.

CTS spectra are, however, also sensitive to a wide range of other plasma and system parameters, which are either of no direct interest, or adequately diagnosed by other systems. Sensitivity to these *nuisance parameters* encumber the inverse problem of inference about the parameters of interest and reduces the amount of information that can be extracted. The number of nuisance parameters in CTS generally makes the inverse problem ill posed without externally provided prior estimates for the nuisance parameters.

The problem of optimal inference in the presence of nuisance parameters is common to many fusion plasma diagnostics and indeed to diagnostic problems throughout the physical sciences,[1–3] yet mostly it is ignored because of the difficulty in accounting for the effects of the nuisance parameters.

Here new methods of optimal inference about parameters of interest in the presence of uncertain prior estimates of nuisance parameters are discussed. Particular attention is given to the LSN methods (least squares fitting accommodating nuisance parameters) which remain tractable for diagnostic systems where the data are fitted with complex and computationally demanding *forward models*. We then proceed to investigate the accuracies with which parameters of interest can be resolved including the effects of nuisance parameters. Finally methods are given for identifying which nuisance parameters or combinations of nuisance parameters most severely limit the diagnostic capability, and thus which externally provided parameters we would get most benefit out of improving the accuracies of. For new diagnostics this provides an effective tool for design optimization.

The methods of optimal inference and assessment of diagnostic capability are illustrated by CTS and include examples from the Joint European Torus (JET). The methods are, however, generally applicable to diagnostics with nuisance parameters and complex forward models. The present discussion is intended to provide sufficient insight for the reader to assess the usefulness of these methods and to apply them. Detailed technical discussions can be found in Refs. 4 and 5 which also provide reviews of the relevant literature. Introductory discussions of least squares fitting and the underlying principle of maximum likelihood estimation can be found in Refs. 6–10. These do not, however, include discussions of the problems introduced by nuisance parameters. This topic has recently received much attention in the specialist literature on theoretical statistics to which the review in Ref. 4 may serve as a useful introduction. Other reviews are given in Refs. 11–14.

II. SENSITIVITIES OF THE CTS SPECTRUM

Collective Thomson scattering is considered here as an example of a diagnostic system which is sensitive to a wide range of plasma and system parameters, only a small subset of which are of interest. Here we assume that parameters characterizing the fast ion velocity distribution are the *parameters of interest*. The modelling of the received CTS spectrum is based on the expressions given in Ref. 15 which provide a full electromagnetic model both of the microscopic fluctuations and of the scattering process. Contributions from density, flux, and field fluctuations are all accounted for, including the effects of their cross correlations.

We find that the received CTS spectrum is sensitive to the one-dimensional fast ion distribution along the resolved wave vector component of the microscopic fluctuations, $\mathbf{k}^{\delta} = \mathbf{k}^{s} - \mathbf{k}^{i}$: $f_{\alpha}(u) = \int f_{\alpha}(\mathbf{v})\, \delta(u - \hat{k}^{\delta} \cdot \mathbf{v})\, d^3\mathbf{v}$. Here $\mathbf{k}^{i}$ and $\mathbf{k}^{s}$ are respectively the wave vectors of the incident and scattered radiation at the scattering volume. The sensitivity of the received CTS spectrum to $f_{\alpha}(u)$ is illustrated in Fig. 1. A nearly complete list of parameters to which CTS is sensitive is given in Table I.

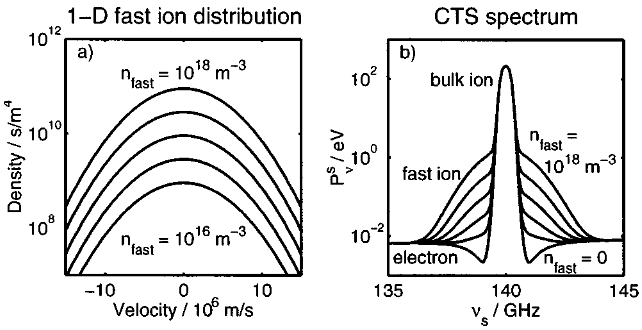

FIG. 1. Sensitivity of the CTS spectrum to the fast ion distribution. (a) A set of fast ion distributions projected onto the resolved direction in velocity space. (b) The resulting CTS spectra. The remaining plasma and system parameters are given in Table I.

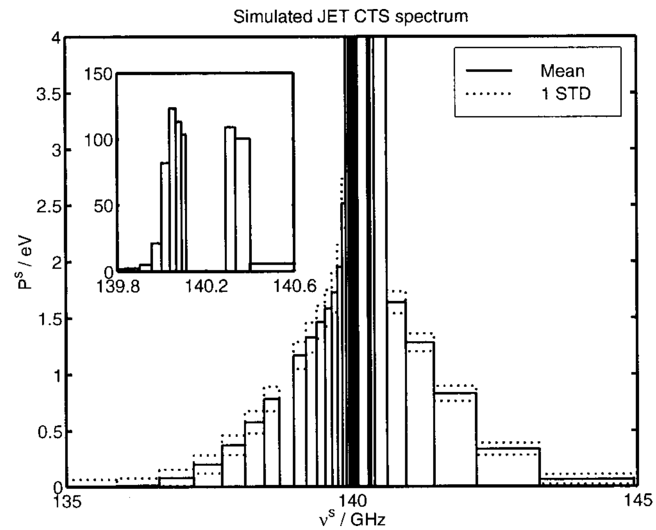

FIG. 2. CTS spectrum resolved into the 31 channels which are available to the CTS diagnostic at JET. The main parameters used to generate this synthetic spectrum are given in Table II with the fast proton population shown as the model in Fig. 3.

III. LEAST SQUARES FITTING AND NUISANCE PARAMETERS

Resolving the CTS spectrum in a finite set of channels, as available in the CTS diagnostic at JET, we might find the spectrum shown in Fig. 2. There is not sufficient information in such a spectrum to provide useful estimates of all the parameters on which the spectrum depends. It is necessary to introduce additional information from other measurements and possibly theoretical considerations. We refer to this information as the *prior information*. In the example considered here, only the N_θ parameters characterizing the fast ion distribution, $f_H(u)$, are *parameters of interest*. In keeping with standard terminology in statistics, the parameters of interest are the elements of the vector $\boldsymbol{\theta}$ (not to be confused with the scattering angle θ). The remaining N_ϕ parameters (e.g., θ, O_b, n_e, T_e, T_i, N_i, calibration coefficients) are the *nuisance parameters* which make up the elements of the vector $\boldsymbol{\phi}$. For convenience we also introduce the vector $\boldsymbol{\psi}=(\boldsymbol{\theta},\boldsymbol{\phi})$ containing all the parameters on which the CTS spectrum depends.

The following questions now arise when fitting model to data: (a) Do we treat nuisance parameters as fixed or free? (b) Do we incorporate prior information? (c) How do we account for uncertainties in the prior information?

Least squares fitting provides optimal inference when noise and other uncertainties are normally distributed. It is, however, the workhorse in data analysis even when the distribution describing the uncertainties is not normal or not exactly known. This is partly justified by the robustness of least squares, and partly by the ease with which the technique can be applied. In the subsequent we therefore concentrate on least squares fitting (LS) and a generalization of least squares fitting which accounts for uncertainties in prior information about Nuisance parameters (LSN). For the detailed discussion we assume uncertainties are adequately approximated by normal distributions. We emphasize though that LSN retains the robustness of LS.

So in answer to the above questions: With LS we can treat nuisance parameters as fixed, thereby using prior information but ignoring its uncertainties. If all nuisance parameters are fixed the fitting has N_θ degrees of freedom over which the χ^2 merit function is to be minimized. Alternatively we can treat the nuisance parameters as free, thus ignoring prior information. Treating all nuisance parameters as free would result in $N_\theta+N_\phi$ degrees of freedom, and in the case of CTS, the inverse problem (inference about parameters) would be singular. With LSN we can treat the nuisance parameters either as free or as fixed, and in both cases make full use of the prior information while taking the uncertainty in this information into account. With adequate prior information the CTS inverse problem is regular. LSN with all nuisance parameters free provides optimal inference in the widest range of situations. Treating nuisance parameters as fixed, LSN remains optimal provided the model of the data is linear over the range of the uncertainties in those nuisance parameters which we keep fixed. The very significant practical advantage of LSN with fixed as opposed to LSN with free nuisance parameters is the reduction in the dimension of the space in which χ^2 is to be minimized.

TABLE I. Plasma and system parameters which the received CTS spectrum is sensitive to.

Parameters		Typical JET value
B	$\lvert\mathbf{B}\rvert$, modulus of magnetic field	3.4 T
n_e	electron density	4×10^{19} m^{-3}
T_e	electron temperature	8 keV
N_i	n_i/n_e, norm. impurity (carbon) density	0.04
R_i	fuel ratio, $n_T/(n_D+n_T)$	0.02
T_i	bulk ion temperature	8 keV
f_H	fast ion distribution	Maxw. (T_H=200 keV)
n_H	fast ion density ($\int f_H(u)du$)	10^{18} m^{-3}
θ	scattering angle, $\angle(\mathbf{k}^i,\mathbf{k}^s)$	30°
ϕ	$\angle(\mathbf{k}^\delta,\mathbf{B})$	115°
ψ	asymmetry angle [a]	0°
P^i	incident beam power	400 kW
O_b	beam overlap	18.8 m$^{-1}/\lvert\sin(\theta)\rvert$
ν^i	frequency of incident radiation	140 GHz
C	calibration of detector channels	

[a]Azimuthal angle of $\mathbf{B}$ when $\hat{\mathbf{z}}\parallel\mathbf{k}^\delta$ and $\hat{\mathbf{x}}\parallel\mathbf{k}^i\times\mathbf{k}^s$.

TABLE II. Model values were used when generating the synthetic spectrum shown in Fig. 2. Prior values were used when inferring the fast ion velocity distribution. Column labelled error gives the error of the prior relative to the model while column σ prior gives the estimates of the uncertainty in the prior used for LSN.

Nuisance	Model	Prior	Error	σ prior
$\theta/1°$	25	24	4%	4%
$n_e/10^{19}\ \mathrm{m}^{-3}$	3.0	3.3	10%	10%
T_e/ keV	8.0	7.0	13%	13%
N_i	0.04	0.05	25%	25%
R_i	0.02	0.01	50%	50%
T_i/ keV	3.0	3.5	17%	17%

IV. SIMULATED CTS DATA ANALYSIS WITH LS AND LSN

Here we illustrate the power of LSN with fixed nuisance parameters when applied to the analysis of simulated CTS data. The spectrum shown in Fig. 2 was generated with the model data given in Table II and model fast ion distribution shown in Fig. 3. The remaining parameters are as given in Table I.

From the simulated spectrum in Fig. 2 inference was made about the fast ion distribution using LS and LSN. In both cases the scaling factor O_b was left free, reflecting the large uncertainty in the absolute spectral density of a measured spectrum. The remaining nuisance parameters were fixed. The parameters included in Table II were fixed at the values given in the column marked prior, which differ from the model values by realistic amounts. In the LSN inference it was further assumed that the standard deviations of the uncertainties in the prior values are as indicated in the column marked σ prior. The results of the LS and LSN inferences are shown in Fig. 3 along with the model distribution. The LS inferred distribution is clearly unsatisfactory with large kinks on the flanks of the distribution and unphysical values near the center. In contrast hereto, the LSN inferred distribution is quite satisfactory with excellent recovery of shape and a small error in the absolute scaling.

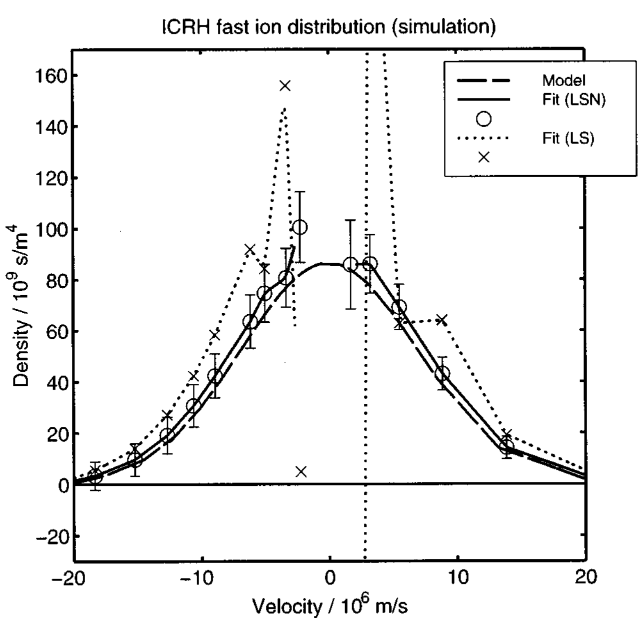

FIG. 3. Fast proton distributions. Model is as predicted by theory for typical ion cyclotron heating scenarios at JET. The two remaining curves indicate the distributions inferred from the simulated spectrum in Fig. 2 using LS and LSN, with O_b free and the remaining nuisance parameters fixed. For the nuisance parameters listed in Table II the prior values were used, with LSN assuming these values have the uncertainties listed as σ prior.

V. FOUNDATIONS OF LS AND LSN

To discuss the theoretical foundations of LS and LSN we need to define the tools and information which is at our disposal. First we must have a model of the expectation values of the measured data (in our example the calibrated spectral densities resolved into channels) as a function of plasma and system parameters. This is the *forward model*, $\mathbf{y}=K(\boldsymbol{\psi})$, where $\mathbf{y}$ is the vector of expectation values for the measured data (CTS spectrum), K is the operator or function representing the theoretical model and $\boldsymbol{\psi}=(\boldsymbol{\theta},\boldsymbol{\phi})$ is the set of all parameters on which the model of our data depends.

The information contained in the data is represented by the likelihood function $P(\tilde{\mathbf{y}}|\boldsymbol{\psi})$ which for given parameters $\boldsymbol{\psi}$ represents the probability density for measuring the data vector $\tilde{\mathbf{y}}=\mathbf{y}+\delta\tilde{\mathbf{y}}$. The stochastic nature of $\tilde{\mathbf{y}}$ is due to the noise in the data and the possibly stochastic nature of the data signal itself, as is the case for CTS. For given $\tilde{\mathbf{y}}$, $P(\tilde{\mathbf{y}}|\boldsymbol{\psi})$ represents the likelihood that $\boldsymbol{\psi}$ gave rise to the measured data. Additional information from other diagnostics or theory may be available for all parameters $\boldsymbol{\psi}$. We assume that this information can be represented by the probability density, $P(\boldsymbol{\psi})$, which is maximum at $\boldsymbol{\psi}=\tilde{\boldsymbol{\psi}}$. This distribution is referred to as the prior. Here we are taking a Bayesian approach to statistical inference.[16–18] The rationale behind LSN can also be presented from a frequentists point of view[4] in which case the prior information is regarded as the result of a measurement represented by a separate likelihood function $P(\tilde{\boldsymbol{\psi}}|\boldsymbol{\psi})$. Here we continue in the Bayesian picture.

Prior information and new information obtained in the measurement are combined through Baye's relation to give the *posterior*, $P(\boldsymbol{\psi}|\tilde{\mathbf{y}})=P(\tilde{\mathbf{y}}|\boldsymbol{\psi})P(\boldsymbol{\psi})/P(\tilde{\mathbf{y}})$, representing all the available information. The relations between prior, likelihood, and posterior are illustrated in Fig. 4 where a contour is plotted for each of the three distributions in a two dimensional parameter space. In this example the distributions are normal as indicated by their projections.

The inferred parameter values, obtained with LSN, are the values at the maximum of the posterior, provided the noise in the data and the prior are both normal. LS with all parameters free ignores the prior and thus yields the parameters at the maximum of the likelihood function assuming normal noise. This is illustrated in Fig. 5 where the distributions are plotted in a two-dimensional parameter space consisting of one nuisance parameter and one parameter of interest. Here the contours of the prior are parallel to the $\boldsymbol{\theta}$ axis, reflecting the fact that the prior contains no information about the parameter of interest.

LSN with $\boldsymbol{\phi}$ free returns the values of $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ at the maximum of the posterior. The uncertainty in the inferred values is represented by the posterior distribution. LSN with

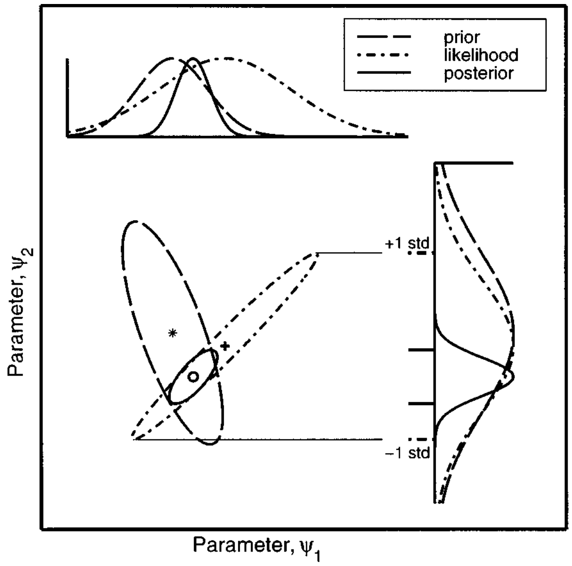

FIG. 4. The prior (dashed), likelihood (dash-dotted), and posterior (solid) distributions are each represented by a contour in a two dimensional parameter space. Here the distributions are normal as indicated by their projections which are shown as subfigures.

$\boldsymbol{\phi}$ fixed returns the value of $\boldsymbol{\theta}$ at the maximum of the projection of the posterior onto $\boldsymbol{\theta}$ space. Here the uncertainty is represented by the projection of the posterior onto $\boldsymbol{\theta}$ space. LS with $\boldsymbol{\phi}$ free returns the values of $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ at the maximum of the likelihood function. The uncertainty is represented by the likelihood distribution. The uncertainty in $\boldsymbol{\theta}$ is thus the projection of the likelihood function onto $\boldsymbol{\theta}$ space. LS with $\boldsymbol{\phi}$ fixed returns the value of $\boldsymbol{\theta}$ at the maximum of the

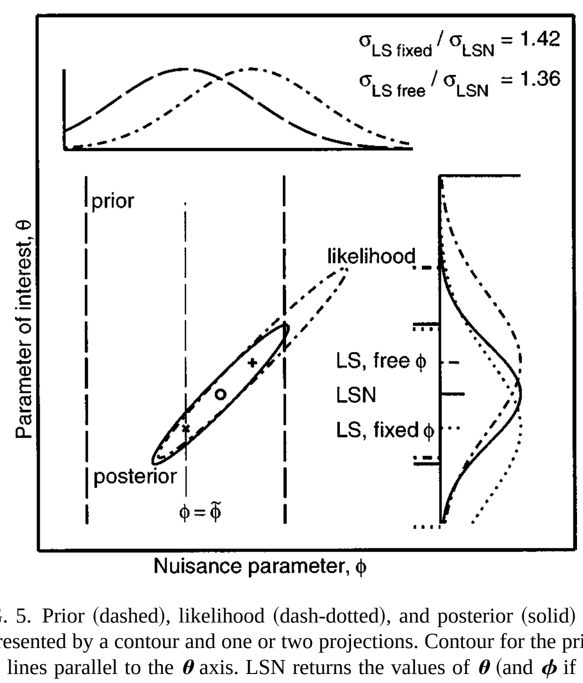

FIG. 5. Prior (dashed), likelihood (dash-dotted), and posterior (solid) each represented by a contour and one or two projections. Contour for the prior is two lines parallel to the $\boldsymbol{\theta}$ axis. LSN returns the values of $\boldsymbol{\theta}$ (and $\boldsymbol{\phi}$ if free) at the maximum (○) of the posterior, LS with $\boldsymbol{\phi}$ free returns the value of $\boldsymbol{\theta}$ at the maximum (+) of the likelihood function while LS with $\boldsymbol{\phi}$ fixed returns the value of $\boldsymbol{\theta}$ at the maximum (×) of the intersection of the likelihood function with the line $\boldsymbol{\phi}=\tilde{\boldsymbol{\phi}}$. Uncertainties in each of the three estimates of $\boldsymbol{\theta}$ are indicated by distributions and the corresponding 1 standard deviation points are marked.

intersection of the likelihood function with $\boldsymbol{\phi}=\tilde{\boldsymbol{\phi}}$. The uncertainty in this estimate is found by standard error propagation and displayed as a distribution. It is clear that LSN always gives the estimate with least uncertainty. LSN also accommodates prior information about parameters of interest. Referring to Fig. 4 it is evident that if the prior contains information about correlations between $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ then the difference between accuracies of LS and LSN estimates can be even more marked.

VI. DEFINITION OF LEAST SQUARES WITH NUISANCE PARAMETERS (LSN)

Here we give the definition of LSN, a generalization of least squares fitting which accounts for uncertainties in prior information about Nuisance parameters. To bring out the new features we first remind the reader of the definition of standard least squares (LS). In all cases the inferred parameter values are found by minimizing a χ^2 functional.

LS, $\boldsymbol{\phi}$ free:

$$\chi^2=\widetilde{\Delta \mathbf{y}}^T \boldsymbol{\alpha}^{(y)} \widetilde{\Delta \mathbf{y}} \quad (=-2\log\{P(\tilde{\mathbf{y}}|\boldsymbol{\psi})\})_{\text{normal}}$$

$$\widetilde{\Delta \mathbf{y}}=\tilde{\mathbf{y}}-\mathbf{K}(\boldsymbol{\theta},\boldsymbol{\phi}), \quad (\boldsymbol{\alpha}^{(y)})^{-1}=\mathbf{C}^{(y)}=\langle \delta\tilde{\mathbf{y}}\delta\tilde{\mathbf{y}}\rangle.$$

LS, $\boldsymbol{\phi}$ fixed: as LS, $\boldsymbol{\phi}$ free, only $\widetilde{\Delta \mathbf{y}}=\tilde{\mathbf{y}}-\mathbf{K}(\boldsymbol{\theta},\tilde{\boldsymbol{\phi}})$.

LSN, $\boldsymbol{\phi}$ free:

$$\chi^2=\widetilde{\Delta \mathbf{v}}^T \boldsymbol{\alpha}^{(v)} \widetilde{\Delta \mathbf{v}} \quad (=-2\log\{P(\boldsymbol{\psi}|\tilde{\mathbf{y}})\})_{\text{normal}}$$

$$\widetilde{\Delta \mathbf{v}}=\begin{pmatrix} \tilde{\mathbf{y}}-\mathbf{K}(\boldsymbol{\theta},\boldsymbol{\phi}) \\ \tilde{\boldsymbol{\phi}}-\boldsymbol{\phi} \end{pmatrix},$$

$$(\boldsymbol{\alpha}^{(v)})^{-1}=\mathbf{C}^{(v)}=\begin{pmatrix} \langle \delta\tilde{\mathbf{y}}\delta\tilde{\mathbf{y}}\rangle & \langle \delta\tilde{\mathbf{y}}\delta\tilde{\boldsymbol{\phi}}\rangle \\ \langle \delta\tilde{\boldsymbol{\phi}}\delta\tilde{\mathbf{y}}\rangle & \langle \delta\tilde{\boldsymbol{\phi}}\delta\tilde{\boldsymbol{\phi}}\rangle \end{pmatrix}.$$

LSN, $\boldsymbol{\phi}$ fixed (reduced form):

$$\chi^2=\widetilde{\Delta \mathbf{y}}^T \boldsymbol{\alpha}^{(\Delta y)} \widetilde{\Delta \mathbf{y}} \quad \left(=-2\log\left\{\int P(\boldsymbol{\psi}|\tilde{\mathbf{y}})\,d\boldsymbol{\phi}\right\}\right)_{\text{normal}}$$

$$\widetilde{\Delta \mathbf{y}}=\tilde{\mathbf{y}}-\mathbf{K}(\boldsymbol{\theta},\tilde{\boldsymbol{\phi}}),$$

$$\mathbf{C}^{(\Delta y)}=\mathbf{C}^{(y)}+K_{\boldsymbol{\phi}}\mathbf{C}^{(\phi)}K_{\boldsymbol{\phi}}^T-K_{\boldsymbol{\phi}}\langle \delta\tilde{\boldsymbol{\phi}}\delta\tilde{\mathbf{y}}\rangle-\langle \delta\tilde{\mathbf{y}}\delta\tilde{\boldsymbol{\phi}}\rangle K_{\boldsymbol{\phi}}^T.$$

Here $\{K_{\boldsymbol{\phi}}\}_{ij}=\partial y_i/\partial \phi_j$. The second equalities for each of the expressions for χ^2 holds when noise and uncertainties in nuisance parameters are normal.

VII. LSN: FITTING ONLY RELIABLE FEATURES

In Sec. V the rationale behind LSN was discussed in terms of the prior, likelihood, and posterior distributions. The rationale behind the reduced form of LSN (fixed $\boldsymbol{\phi}$) is, however, amenable to a simple alternative explanation. To simplify the discussion consider a situation where there are no correlations between uncertainties in the data on the one hand, and the prior for nuisance parameters on the other. In that case the covariance tensor entering the definition of χ^2 simplifies to $\mathbf{C}^{(\Delta y)}=\mathbf{C}^{(y)}+K_{\boldsymbol{\phi}}\mathbf{C}^{(\phi)}K_{\boldsymbol{\phi}}^T$. The effect of the sec-

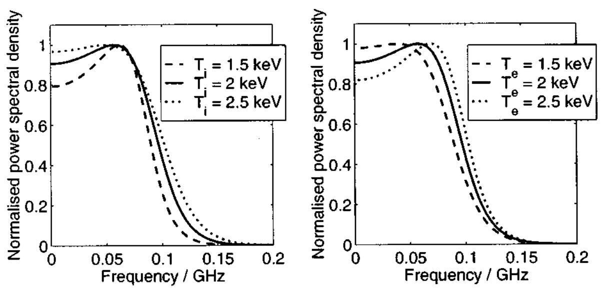

FIG. 6. Normalized CTS spectra for a range of ion and electron temperatures.

ond term is to decrease the weight given in the fitting to those features in the data which could be explained by incorrect values of nuisance parameters.

To illustrate this, consider the problem of inferring the ion temperature from a noisy CTS spectrum with imperfect knowledge of the electron temperature. The spectrum is sensitive to both the ion temperature and the electron temperature as illustrated in Fig. 6. The spectra have been normalized reflecting the fact that the absolute scaling of measured CTS spectra are generally very uncertain and thus contains little information.

The spectral variations associated with varying these two quantities are clearly not orthogonal, so an incorrect value of T_e would clearly affect the estimate of T_i if the inference was based on LS with T_e fixed. LSN with T_e fixed would however assume an effective uncertainty in the data represented by the covariance tensor

$$\mathbf{C}^{(\Delta y)} = \mathbf{C}^{(y)} + K_{T_e} (\sigma^{(T_e)})^2 K_{T_e}^T,$$

where $\sigma^{(T_e)}$ is the standard deviation of the uncertainty in prior information on T_e. Deviations between the data and the model of the functional form K_{T_e} would thus not be regarded as significant unless the deviation was large compared with $K_{T_e} \times \sigma^{(T_e)}$. Figure 7 shows a synthetic spectrum generated with $T_e = T_i = 2$ keV, and the fitted spectra resulting from LS and LSN assuming $T_e = 2.5$ keV.

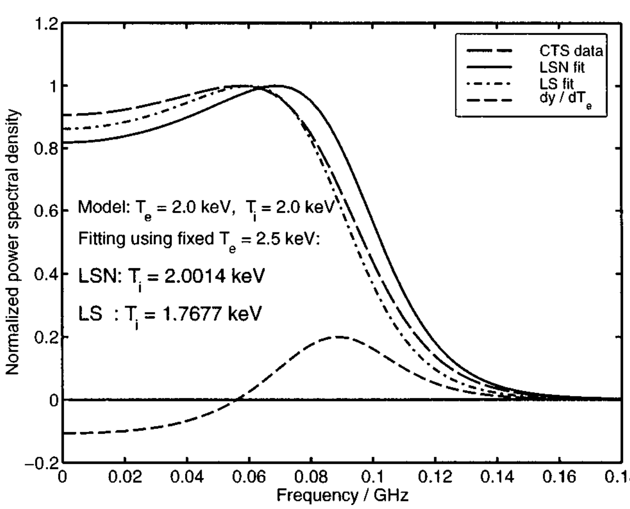

FIG. 7. Synthetic spectrum generated with $T_e = T_i = 2$ keV, and the fitted spectra resulting from LS and LSN assuming $T_e = 2.5$ keV. Also shown is the derivative, $K_{T_e} = \partial \mathbf{y} / \partial T_e$, of the spectrum with respect to T_e.

The LS fit appears to be better, but the inferred ion temperature (LS: $T_i = 1.77$ keV) is significantly more in error than that obtained with LSN (LSN: $T_i = 2.0014$ keV). Most of the deviation between the synthetic data and the LSN fit is of the form $\partial \mathbf{y} / \partial T_e = K_{T_e}$ which the χ^2, used in LSN, is less sensitive to.

VIII. RELATIONS BETWEEN UNCERTAINTIES IN INFERRED AND PRIOR VALUES

Referring to Fig. 4 it is clear that with perfect prior knowledge about a nuisance parameter, ϕ, the uncertainty in the inferred value of a parameter of interest is given by the intersection of the likelihood distribution with the line $\phi = \tilde{\phi}$ (which yields the minimum standard deviation $\sigma_{\min}$). With complete absence of prior information the posterior is equal to the likelihood, and the uncertainty in θ is given by the projection of the likelihood onto the θ axis (which results in the maximum standard deviation $\sigma_{\max}$). Assuming that the prior contains no information about the parameters of interest, these two limiting values for the uncertainty coincide with the uncertainties obtained with LS, treating ϕ as fixed in the first case and free in the second.

It can be shown[4] that the relation between the uncertainty in the LSN estimate of a parameter of interest and the uncertainty in the prior information about a nuisance parameter is a simple sigmoid

$$\text{LSN:} \quad (\sigma^{(\theta)})^2 = \frac{\sigma_{\min}^2 + \sigma_{\max}^2 (\sigma^{(\phi)}/\sigma_c)^2}{1 + (\sigma^{(\phi)}/\sigma_c)^2}$$

characterized by the minimum and maximum values, $\sigma_{\min}$ and $\sigma_{\max}$, of $\sigma^{(\theta)}$, and the critical value, σ_c, of the standard deviation of the prior, $\sigma^{(\phi)}$, at which

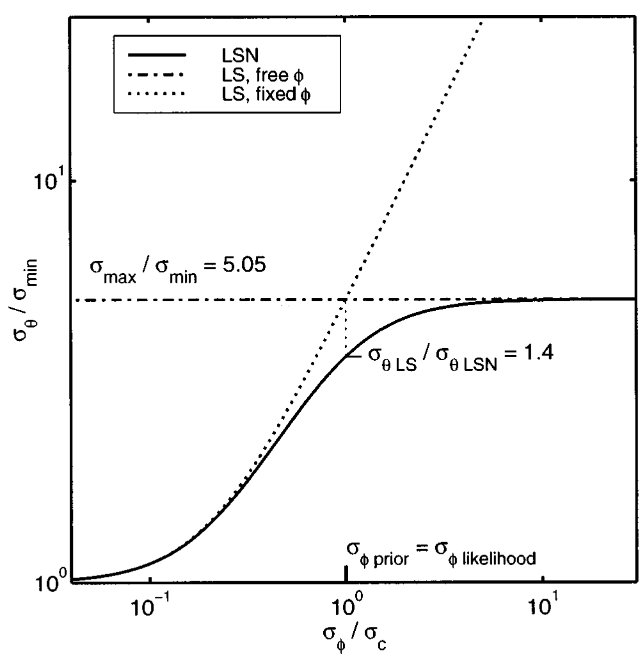

FIG. 8. Sigmoidal relation between the uncertainty in prior value of a nuisance parameter and the uncertainty in the LSN inferred value of a parameter of interest. Also plotted are the uncertainties in inferred values obtained using LS with ϕ fixed and with LS with ϕ free.

$(\sigma^{(\theta)})^2=(\sigma^2_{\min}+\sigma^2_{\max})/2$. The critical value of $\sigma^{(\phi)}$ is also the standard deviation of the projection of the likelihood onto the ϕ axis. The sigmoidal relation between uncertainties for LSN is plotted in Fig. 8 for the specific case of $\sigma_{\max}/\sigma_{\min}=5$. This ratio can take any value and depends on the likelihood function, i.e., on the sensitivity of the data to all parameters and on the noise in the data. Also plotted are the uncertainties in inferred values obtained using LS with ϕ fixed and with LS with ϕ free.

If the optimal version of LS were chosen the variance of the uncertainty of the inference would maximally be twice that of LSN. If the wrong version was chosen the increase in uncertainty can be arbitrarily large. With a large number of nuisance parameters and parameters of interest it is not a trivial task finding which linear combinations of nuisance parameters should be treated as fixed and which as free. LSN eliminates this problem, and is in addition always superior.

The simplicity of the sigmoidal relationship combined with the fact that it is characterize by just three parameters is very useful when identifying which nuisance parameter uncertainties most severely limit the accuracy with which a given parameter of interest is estimated. Parameters $\sigma_{\min}$, $\sigma_{\max}$, σ_c are readily computed (by evaluating $\sigma^{(\theta)}$ for three values of $\sigma^{(\phi)}$ and solving for $\sigma_{\min}$, $\sigma_{\max}$, and σ_c) for all combinations of $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ parameters. If the difference between $\sigma_{\min}$ and $\sigma_{\max}$ is small it follows that uncertainty in that element of $\boldsymbol{\phi}$, given the other uncertainties, does not mask information about the selected element in $\boldsymbol{\theta}$. In the case of a large difference, the value of $\sigma^{(\phi)}/\sigma_c$ is an indication of whether the magnitude of the uncertainty is critical or not.

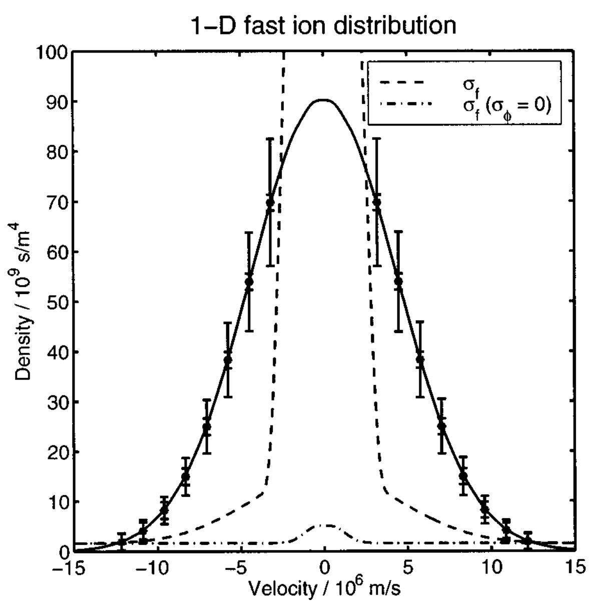

FIG. 9. Model fast proton distribution with uncertainties (1 std) in resolved nodes indicated by errorbars. The smaller errorbars indicate the uncertainty in nodes if there was no uncertainty in the prior. The dashed and dash-dotted curves also represent the errorbars. Parameters: τ (signal integration time) $=200$ ms, T_N (spectral noise density) = 200 eV, ν_n (half width of notch filter) = 100 MHz. The remaining parameters are as in Table I. Uncertainties in prior: $\sigma_\theta=1°$, $\sigma_\phi=1°$, $\sigma_\psi=1°$, $\sigma_{O_b}=O_b$, $\sigma_B=0.1\times B$, $\sigma_{n_e}=0.2\times n_e$, $\sigma_{T_e}=0.1\times T_e$, $\sigma_{N_i}=0.5\times N_i$, $\sigma_{R_i}=0.3\times R_i$, $\sigma_{T_i}=0.3\times T_i$.

TABLE III. Relative uncertainty in scaling factor, σ_F/F, for combinations of noise temperatures and standard (caption of Fig. 9) or no uncertainties in nuisance parameters.

σ_F/F	$T_N=200$ eV	$T_N=400$ eV
Standard σ_ϕ	0.19	0.20
$\sigma_\phi=0$	0.018	0.035

IX. ASSESSING THE DIAGNOSTIC POTENTIAL OF JET'S CTS

Here we illustrate the impact uncertainties in prior information can have on the accuracy of inferred values for parameters of interest, and demonstrate the usefulness of the sigmoidal relation between prior and posterior uncertainties when minimizing this impact. We consider the CTS spectrum we would obtain with the plasma and system parameters given in Table I. It is assumed here that the frequency resolution of the CTS spectrum is sufficient that it does not affect the diagnostic capability.

With the noise temperature and the uncertainties in prior indicated in the caption of Fig. 9 the inferred fast ion distribution would have the uncertainties in resolved nodes indicated by error bars in Fig. 9.

The errorbars represent the standard deviations of individual parameters of the fast ion distribution, i.e., the projection of the posterior onto parameter axes. Much of the uncertainty is correlated and due to uncertainty in the absolute scaling factor, F. This and other correlations can be brought out by principal component analysis of the posterior.[5,19]

The uncertainty in the scaling factor for combinations of two noise temperatures and standard or no uncertainties in nuisance parameters is given in Table III. Clearly the uncertainty in nuisance parameters give rise to most of this uncertainty.

To identify which nuisance parameters mask the information on the scaling factor we make use of the simple sigmoidal relationship between $\sigma^{(\theta)}$ and $\sigma^{(\phi)}$. The three parameters characterizing this relation are computed for each nuisance parameter. For most nuisance parameters the difference between $\sigma_{\min}$ and $\sigma_{\max}$ is small indicating little masking by this parameter. Uncertainties in electron density and beam overlap are significant contributors to the uncertainty in the scaling factor as indicated by the parameters in Table IV.

It is noteworthy that we can virtually eliminate all the masking of information on F by eliminating the uncertainty in either n_e or O_b. However the required improvements for

TABLE IV. Parameters defining sigmoidal relationships between uncertainties in estimated parameters F and principal component 21 on the one hand, and nuisance parameters θ, O_b, n_e, and T_i on the other. Here $T_N=200$ eV. The parameters are similar for $T_N=400$ eV.

	σ_F/F			σ_{21}		
	σ_ϕ/σ_c	$\sigma_{\min}$	$\sigma_{\max}$	σ_ϕ/σ_c	$\sigma_{\min}$	$\sigma_{\max}$
θ	0.35	0.19	0.19	0.35	1.17	4.6
O_b	4.7	0.024	0.19	4.7	1.89	1.90
n_e	0.55	0.024	0.39	0.55	1.89	1.92
T_i	2.915	0.18	0.19	2.9	1.63	1.93

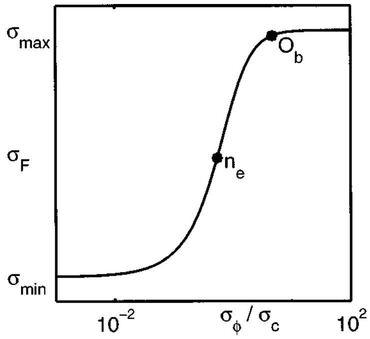

FIG. 10. Sigmoidal relationship between the uncertainty in the scaling factor, $\sigma^{(F)}$, and uncertainties, $\sigma^{(\phi)}$, in priors for n_e and O_b. Positions on sigmoid for current uncertainties are indicated with asterisks.

each of these quantities to reach this end are not equal. The values of σ_ϕ/σ_c indicate where on the sigmoid we are with the current level of uncertainty in the prior. These points are indicated on the plot of the sigmoid in Fig. 10. The uncertainty in n_e is in the critical range where changes in its uncertainty directly impact on the uncertainty in the scaling factor. The loss of information due to uncertainty in O_b, on the other hand, is nearly complete. Further uncertainty in O_b (beam overlap and other quantities affecting the absolute calibration) has little impact and considerable improvement is required to significantly increase the accuracy of F.

By similar investigations (see Table IV) we find that information on the width of the distribution, referred to here as principal component 21, is masked primarily by uncertainty in scattering angle, θ, and to a lesser extent by uncertainty in the bulk ion temperature, T_i. In these studies it must be kept in mind that if two nuisance parameters give rise to (nearly) parallel perturbations of the data (e.g., beam overlap and absolute calibration) they should be treated as one parameter or expanded in other parameters giving rise to orthogonal spectral perturbations.

X. SUMMARY

Many parameters. Commonly diagnostics depend on many parameters. Some are parameters of interest the remainder are nuisance parameters. This introduces the need for, possibly uncertain, prior information. Collective Thomson scattering (CTS) and charge exchange recombination spectroscopy (CXRS) are examples of such diagnostics.

Optimal analysis. Incorporates prior information and accounts for its uncertainties. LSN, a new least squares based method, does this.

Assessing diagnostic capability. Uncertainties in estimates of parameters of interest depend on uncertainties in the prior information. With optimal analysis, these uncertainties are given by the posterior which combines prior and new information. Principal component analysis of the posterior is an effective tool for identifying correlated uncertainties.

Optimizing diagnostic capability. Posterior versus prior uncertainties are related by simple three parameter sigmoids. This is useful when identifying (a) which uncertainties in the prior information mask what posterior information, and (b) which prior information is critical.

These methods have been applied to CTS at JET, JT60U and ITER. Recent CTS data at JET were analyzed using LSN.

[1] J. C. Spall and J. P. Garner, IEEE Trans. Aerosp. Electron. Syst. **26**, 992 (1990).
[2] B. Friedlander, IEEE Trans. Aerosp. Electron. Syst. **26**, 953 (1990).
[3] C. V. Kimbal, P. Lewicki, and N. I. Wijeyesekera, IEEE Trans. Signal Process. **43**, 2928 (1995).
[4] H. Bindslev, Maximum likelihood estimators for inverse problems with nuisance parameters, JET report JET−R(97)13 (1997).
[5] H Bindslev, Assessing and optimizing measuring capabilities of complex diagnostic systems, JET report JET−R(97)12 (1997).
[6] W. Ledermann, *Handbook of Applicable Mathematics. Vol. VI. Statistic Part A*, edited by E. Lloyd (Wiley, New York, 1984).
[7] R. J. Barlow, *Statistics* (Wiley, Chichester, 1989).
[8] K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis* (Academic, London, 1979).
[9] P. R. Bevington and D. K. Robinson, *Data Reduction and Error Analysis for the Physical Sciences,* 2nd ed. (McGraw−Hill, New York, 1994).
[10] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes,* 2nd ed. (Cambridge University Press, Cambridge, 1992).
[11] J. D. Kalbfleisch and D. A. Sprott, J. R. Statist. Soc. Ser. B **32**, 175 (1970).
[12] D. Basu, J. Am. Statist. Assoc. **72**, 355 (1977).
[13] O. E. Barndorff-Nielsen, Theor. Probab. Appl. **38**, 179 (1993).
[14] K. Y. Liang and S. L. Zeger, Stat. Sci. **10**, 158 (1995).
[15] H. Bindslev, J. Atmos. Terr. Phys. **58**, 983 (1996).
[16] G. E. P. Box and G. C. Tiao, *Bayesian Inference in Statistical Analysis* (Addison−Wesley, Reading, MA, 1973).
[17] C. R. Smith, and W. T. Grandy, Jr., *Maximum-Entropy and Bayesian Methods in Inverse Problems* (Reidel, Dordrecht, 1985).
[18] W. T. Grandy, Jr., *Maximum-Entropy and Bayesian Methods* (Reidel, Dordrecht, 1990).
[19] H. Bindslev, 8th Int. Symp. on Laser−Aided Plasma Diagnostics Doorwerth, The Netherlands 265 FOM (1997). (JET preprint JET−P(97)24).